



T.C.

BİLECİK ŞEYH EDEBALI ÜNİVERSİTESİ

LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

ELEKTRONİK VE BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

ELEKTRONİK VE BİLGİSAYAR MÜHENDİSLİĞİ DOKTORA PROGRAMI

**İKİ BOYUTLU TEKLİ GÖRÜNTÜDEN ÇEKİŞMELİ ÜRETİCİ AĞLAR
KULLANILARAK ÜÇ BOYUTLU NESNE ÜRETİMİ**

DOKTORA TEZİ

ZAFER SERİN

TEZ DANIŞMANI

PROF. DR. UĞUR YÜZGEÇ

İKİNCİ TEZ DANIŞMANI

PROF. DR. CİHAN KARAKUZU

BİLECİK, 2026

10591002

T.C.

BİLECİK ŞEYH EDEBALI ÜNİVERSİTESİ

LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

ELEKTRONİK VE BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

ELEKTRONİK VE BİLGİSAYAR MÜHENDİSLİĞİ DOKTORA PROGRAMI

**İKİ BOYUTLU TEKLİ GÖRÜNTÜDEN ÇEKİŞMELİ ÜRETİCİ AĞLAR
KULLANILARAK ÜÇ BOYUTLU NESNE ÜRETİMİ**

DOKTORA TEZİ

ZAFER SERİN

TEZ DANIŞMANI

PROF. DR. UĞUR YÜZGEÇ

İKİNCİ TEZ DANIŞMANI

PROF. DR. CİHAN KARAKUZU

BİLECİK, 2026

10591002

BEYAN

İki Boyutlu Tekli Görüntüden Çekişmeli Üretici Ağlar Kullanılarak Üç Boyutlu Nesne Üretimi başlıklı doktora tezinin hazırlık ve yazım aşamasında bilimsel araştırma ve etik kurallarına uyduğumu, [Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Faaliyetlerinde Üretken Yapay Zekâ Kullanımına Dair Etik Rehberine](#) uygun olarak tez/dönem projemi hazırladığımı, başkalarının eserlerinden yararlandığım bölümlerde bilimsel etik kurallarına uygun olarak atıfta bulunduğumu, kullandığım verilerde herhangi bir tahrifat yapmadığımı, çalışmamın herhangi bir kısmının başka bir tez/dönem projesi olarak sunulmadığını, aksinin tespit edilmesi durumunda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Zafer SERİN

.. / .. / 2026

İmza

ÖN SÖZ

Bu çalışmamın başından sonuna kadar bilgi ve deneyimleriyle bana yol gösteren, yapıcı eleştirileriyle çalışmamın gelişmesini ve bir üst seviyeye çıkmasını sağlayan danışmanım Sayın Prof. Dr. Uğur YÜZGEÇ'e ve değerli katkıları, yönlendirmeleri ve vizyonu ile çalışmaya yön veren ikinci danışmanım Sayın Prof. Dr. Cihan KARAKUZU'ya teşekkürlerimi sunarım.

Tez izleme ve savunma sürecinde yapıcı değerlendirme ve önerileriyle çalışmaya katkı sağlayan Sayın Dr. Öğr. Üyesi Burakhan ÇUBUKÇU'ya, Dr. Öğr. Üyesi Zekeriya KATILMIŞ'a, Dr. Öğr. Üyesi Emre GÜNGÖR'e, Dr. Öğr. Üyesi Hakan ÜÇGÜN'e, Dr. Öğr. Üyesi Vedat MARTTİN'e ve Prof. Dr. Devrim AKGÜN'e teşekkür ederim.

Bu çalışmanın her sayfasında benim emeğim olsa da beni asıl var eden ailemdir. Benim için her türlü güçlüğü göze alacaklarını ve ihtiyaç duyduğum her an tüm imkânlarını tereddütsüz önüme sereceklerini bilmenin verdiği güven bu tezin en büyük motivasyonudur. Ömrüm onların bu eşsiz varlığına olan borcumu ödemeye yetmeyecek kadar kısa olsa da aileme en derin minnetimle teşekkür ederim.

ÖZET

Bu çalışmada, tek iki boyutlu görüntüden üç boyutlu nesne üretimi için Varyasyonel Otomatik Kodlayıcı (*Variational Autoencoder* – VAE) ve Çekişmeli Üretici Ağ (*Generative Adversarial Network* – GAN) temelli 3D-VAE-GAN mimarisi kullanılmıştır. Temsil için Derin Sinir Ağları (*Deep Neural Networks* – DNN) ile uyumlu voksel (*volumetric pixel*) gösterimi seçilmiştir. Üretim başarımını artırmak için yeniden yapılandırma kaybı (*reconstruction loss*) olarak Ortalama Karesel Hata (*Mean Squared Error* – MSE), İkili Çapraz Entropi (*Binary Cross-Entropy* – BCE), Odak Kaybı (*Focal Loss* – FL) türevi olan Modülasyon Faktörü Kaybı (*Modulating Factor Loss* – MFL), Dice Kaybı (*Dice Loss* – DLS) ve Kesişim Üzeri Birleşim (*Intersection over Union* – IoU) Kaybı kıyaslanmış; yüzey doğruluğu için özgün bir Uyarlamalı Ağırlıklandırılmış IoU-Odak Kaybı (*Adaptive Weighted IoU-Focal Loss* – AWIF) önerilmiştir. Literatürdeki Sıkıştırma ve Uyarma (*Squeeze-and-Excitation* – SE) ile Evrişimsel Blok Dikkat Modülü (*Convolutional Block Attention Module* – CBAM) dikkat (*attention*) mekanizmalarına ek olarak frekans düzleminde çalışan özgün bir Karmaşık Faz Kaydırma Dikkati (*Complex Phase-Shifting Spatial Attention* – CPSA) geliştirilmiştir. GAN sorunları olan mod çökmesi (*mode collapse*), kaybolan gradyan (*vanishing gradient*) ve kararsız eğitimi aşmak için Wasserstein Çekişmeli Üretici Ağ (*Wasserstein Generative Adversarial Network* – WGAN) uygulanmıştır. Kırmızı-Yeşil-Mavi (*Red-Green-Blue* – RGB) kanallarına derinlik haritaları (*depth maps*) eklenmiştir. Mimari 2D mixup veri artırma (*data augmentation*), kodlayıcı (*encoder*) olarak Derin Artık Ağlar (*Deep Residual Networks* – ResNet), topluluk (*ensemble*) tahmini, Spektral Normalizasyon (*Spectral Normalization* – SN) ve Lipschitz kısıtına (*Lipschitz constraint*) uyum için Gradyan Cezalandırması (*Gradient Penalty* – GP) ile güçlendirilerek Spektral Wasserstein Dikkat Üretici Topluluk-3D (*Spectral Wasserstein Attention Generative Ensemble 3D* – SWAGE-3D) yapısı oluşturulmuştur. ShapeNet veri setindeki 13 sınıfta IoU, Chamfer Uzaklığı (*Chamfer Distance* – CD) ve F-Skoru (F-Score) başarımleriyle yapılan analizler, bu geliştirmelerin nicel (*quantitative*) ve nitel (*qualitative*) başarıyı belirgin şekilde artırdığını kanıtlamaktadır.

Anahtar Kelimeler: 3D-VAE-GAN, 3D yeniden yapılandırma, derinlik haritalama, dikkat mekanizması, kayıp fonksiyonu.

ABSTRACT

In this study, a 3D-VAE-GAN architecture based on Variational Autoencoder (VAE) and Generative Adversarial Network (GAN) was utilized for 3D object generation from a single two-dimensional image. A voxel (volumetric pixel) representation compatible with Deep Neural Networks (DNN) was selected. To improve generation performance, Mean Squared Error (MSE), Binary Cross-Entropy (BCE), Modulating Factor Loss (MFL) which is a derivative of Focal Loss (FL), Dice Loss (DLS), and Intersection over Union (IoU) Loss were compared as reconstruction losses; and a novel Adaptive Weighted IoU-Focal Loss (AWIF) was proposed for surface accuracy. In addition to the Squeeze-and-Excitation (SE) and Convolutional Block Attention Module (CBAM) attention mechanisms in the literature, a novel Complex Phase-Shifting Spatial Attention (CPSA) operating in the frequency domain was developed. To overcome GAN issues such as mode collapse, vanishing gradient, and unstable training, the Wasserstein Generative Adversarial Network (WGAN) was implemented. Depth maps were added to the Red-Green-Blue (RGB) channels. The architecture was fortified with 2D mixup data augmentation, Deep Residual Networks (ResNet) as the encoder, ensemble prediction, Spectral Normalization (SN), and Gradient Penalty (GP) to comply with the Lipschitz constraint, thereby establishing the Spectral Wasserstein Attention Generative Ensemble 3D (SWAGE-3D) structure. Analyses conducted on 13 classes from the ShapeNet dataset using IoU, Chamfer Distance (CD), and F-Score performance metrics prove that these enhancements significantly improve both quantitative and qualitative performance.

Keywords: 3D reconstruction, 3D-VAE-GAN, attention mechanism, depth mapping, loss function.

İÇİNDEKİLER

ÖN SÖZ.....	i
ÖZET.....	ii
ABSTRACT	iii
İÇİNDEKİLER.....	iv
TABLolar LİSTESİ.....	vi
ŞEKİLLER LİSTESİ.....	vii
KISALTMALAR VE SİMGELER LİSTESİ.....	ix
1. GİRİŞ.....	1
1.1. Literatür Taraması.....	4
1.2. Çalışmanın Katkıları.....	8
2. KULLANILAN SAYISAL VERİ, YÖNTEMLER ve ALGORİTMALAR.....	13
2.1. Kullanılan Veri Seti ve Görüntü Temsili	13
2.2. Veri Ön İşleme ve Artırma	19
2.2.1. SWAGE-3D veri ön işleme ve veri artırma işlemleri.....	19
2.3. CNN Mimarisi ve Temel Bileşenleri	20
2.4. ResNet-18 Mimarisi.....	23
2.5. AE Mimarisi ve Devrik Evrişim Temelleri	24
2.6. VAE Mimarisi.....	25
2.7. GAN Mimarisi	25
2.8. WGAN-GP ve SN	26
2.9. 3D-VAE-GAN Mimarisi	27
2.10. Yeniden Yapılandırma Kayıp Fonksiyonları.....	30
2.10.1. MSE kaybı.....	31
2.10.2. BCE kaybı.....	31
2.10.3. Focal kaybı.....	32
2.10.4. MFL kaybı	33
2.10.5. Dice kaybı.....	33
2.10.6. IoU kaybı.....	34
2.10.7. AWIF kaybı	34
2.11. Dikkat Mekanizmaları	39

2.11.1.	Kanal ve uzamsal dikkat mekanizmaları.....	40
2.11.2.	Öz dikkat mekanizmaları	42
2.11.3.	CPSA mekanizması	46
2.11.3.1.	CA bileşeni	48
2.11.3.2.	CFAM bileşeni	49
2.12.	SA Yöntemi	52
2.13.	SWAGE-3D Mimarisi	53
2.13.1.	RGB-D kodlayıcı	55
2.13.2.	Üretici omurga ve çekişmeli kararlılaştırma	56
2.13.3.	Öz dikkat modülü	57
2.13.4.	Topluluk stratejisi	58
2.13.5.	Bileşik kayıp fonksiyonları ve eğitim dinamikleri.....	60
3.	DENEYSEL KURULUM VE DEĞERLENDİRME	61
3.1.	Kullanılan Ortak Altyapı.....	61
3.2.	AWIF Deney Düzeneği ve Analizi.....	62
3.3.	CPSA Deney Düzeneği ve Analizi	63
3.4.	SWAGE-3D Deney Düzeneği ve Analizi	64
3.5.	Değerlendirme Metrikleri.....	65
3.5.1.	IoU metriği	66
3.5.2.	CD metriği.....	67
3.5.3.	F-Skoru metriği	68
4.	BULGULAR VE TARTIŞMA.....	71
4.1.	Nicel Bulgular	71
4.1.1.	AWIF nicel bulguları	71
4.1.2.	CPSA nicel bulguları.....	75
4.1.3.	SWAGE-3D nicel bulguları	80
4.2.	Nitel Bulgular	83
4.2.1.	AWIF nitel bulguları.....	83
4.2.2.	CPSA nitel bulguları	88
4.2.3.	SWAGE-3D nitel bulguları	91
4.3.	Çalışmanın Sınırlılıkları	95
5.	SONUÇLAR VE ÖNERİLER.....	97
	KAYNAKÇA	100

TABLÖLAR LİSTESİ

	Sayfa
Tablo 1.1. 3D nesne üretimi ve yeniden yapılandırma literatür özeti	5
Tablo 3.1. AWIF hiperparametreleri	63
Tablo 3.2. CPSA hiperparametreleri	64
Tablo 3.3. SWAGE-3D hiperparametreleri	65
Tablo 4.1. MSE, BCE, MFL, Dice, IoU ve AWIF kayıp fonksiyonlarının 13 farklı sınıf için elde ettiği IoU değerlerinin kıyaslanması.	72
Tablo 4.2. MSE, BCE, MFL, Dice, IoU ve AWIF'in 13 farklı sınıf için elde ettiği CD değerlerinin kıyaslanması.....	73
Tablo 4.3. Temel Mimari ve farklı dikkat mekanizmalarının (SE, CBAM, CPSA) 13 sınıf üzerindeki IoU başarımlarının SA ile birlikte karşılaştırılması	75
Tablo 4.4. Temel Mimari ile önerilen mimarinin çoklu tohum üzerinden istatistiksel başarımlar değerlendirilmesi ve eğitim kararlılığı karşılaştırması	76
Tablo 4.5. Uçak, ekran ve masa sınıfları için farklı mimarilerin hesaplama karmaşıklığı, verimlilik ve donanımsal kaynak kullanımı bakımından karşılaştırmalı analizi.....	77
Tablo 4.6. Önerilen bileşenlerin IoU metriğine dayalı detaylı ablasyon çalışması.....	79
Tablo 4.7. IoU metriği üzerinden SOTA mimariler ile CPSA+SA'nın kıyaslanması	79
Tablo 4.8. SWAGE-3D ve literatürdeki SOTA mimarilerin IoU metriği üzerinden kıyaslanması	81
Tablo 4.9. SWAGE-3D'nin IoU metriği üzerinden ablasyon çalışması	82
Tablo 4.10. SWAGE-3D'nin F-Skoru metriği üzerinden ablasyon çalışması	82
Tablo 4.11. SWAGE-3D'nin CD metriği üzerinden ablasyon çalışması	83

ŞEKİLLER LİSTESİ

	Sayfa
Şekil 1.1. Aynı at nesnesinin poligon ağı, nokta bulutu ve voksel temsilleri ile gösterimi (Poly Haven, 2022)	2
Şekil 2.1. Otomobil, masa ve telefon nesnelerinin 2D ve farklı açılardan 3D görünümüleri.....	14
Şekil 2.2. Eğitim ve test aşamalarında nesne bazlı kullanılan örnek sayıları	14
Şekil 2.3. 13 sınıfa ait 2D görüntüler (24-bit) ve karşılık gelen Depth Anything V2 tabanlı derinlik haritaları (8-bit).....	16
Şekil 2.4. 2D görüntüler ve karşılık gelen 3D nesneler	18
Şekil 2.5. Düşük, orta ve yüksek çözünürlüklü aslan nesnesi voksel temsilleri (Poly Haven, 2025).....	19
Şekil 2.6. ResNet-18 mimarisi ve katmanları	24
Şekil 2.7. Temel 3D-VAE-GAN mimarisi	30
Şekil 2.8. Örnek bir vokselize nesne.....	37
Şekil 2.9. Öz dikkat mekanizmasının sayısal örnekleme	45
Şekil 2.10. CPSA ve SA mekanizmaları eklenen temel 3D-VAE-GAN mimarisi.....	47
Şekil 2.11. Görüntülerin uzamsal ve frekans düzlemlerindeki temsillerine ilişkin örnekler.....	49
Şekil 2.12. CPSA mekanizmasının uçtan uca çalışma aşamaları	51
Şekil 2.13. SA tekniğinin örnek girdiler üzerindeki etkisi.....	53
Şekil 2.14. SWAGE-3D mimarisinin öbek şeması	54
Şekil 2.15. Bank, dolap, otomobil ve sandalye sınıfları için eğitimin 125. tur sonrası ulaştığı plato bölgesi	59
Şekil 3.1. IoU değerinin hacimsel yapılar üzerindeki görsel temsili	67
Şekil 3.2. 3D yeniden yapılandırma başarımının değerlendirilmesinde kullanılan CD metriğinin görsel temsili	68
Şekil 4.1. Farklı kayıp fonksiyonları için 100 turluk eğitim sürecinde tur başına düşen işlem sürelerinin karşılaştırılması	74
Şekil 4.2. AWIF ile 2D girdilerden üretilen 3D nesnelerin, gerçek nesnelerle üst üste karşılaştırması.....	84
Şekil 4.3. Lamba sınıfı için seçilen 2D girdi ve ilgili kayıp fonksiyonları tarafından üretilen 3D nesneler.....	85
Şekil 4.4. 13 sınıf için 2D girdi, kullanılan kayıp fonksiyonları ile üretilen 3D nesne ve gerçek 3D nesne.....	87

Şekil 4.5. CPSA ile 2D girdilerden üretilen 3D nesnelerin, gerçek nesnelerle üst üste karşılaştırması.....	89
Şekil 4.6. 13 sınıf için 2D girdi, kullanılan dikkat mekanizmaları ile üretilen 3D nesne ve gerçek 3D nesne.....	91
Şekil 4.7. Masa, sandalye ve kanepeler için SWAGE-3D başarımının karşılaştırılması	92
Şekil 4.8. RGB-D girdi, girdilere karşılık gelen gerçek 3D nesne ve bu girdilere karşılık 3D-R2N2, Pix2Mesh ve SWAGE-3D tarafından üretilen 3D nesneler	94

KISALTMALAR VE SİMGELER LİSTESİ

2D: İki boyutlu (Two-dimensional)

3D: Üç boyutlu (Three-dimensional)

3DGS: 3D Gauss saçılımı (3D Gaussian Splatting)

AE: Otomatik Kodlayıcı (Autoencoder)

AI: Yapay Zekâ (Artificial Intelligence)

ANN: Yapay Sinir Ağları (Artificial Neural Network)

AR: Artırılmış Gerçeklik (Augmented Reality)

AWIF: Uyarlamalı Ağırlıklandırılmış IoU-Odak Kaybı (Adaptive Weighted IoU-Focal Loss)

BCE: İkili Çapraz Entropi (Binary Cross-Entropy)

BN: Yığın Normalizasyonu (Batch Normalization)

CA: Kanal Dikkati (Channel Attention)

CBAM: Evrimsel Blok Dikkat Modülü (Convolutional Block Attention Module)

CD: Chamfer Uzaklığı (Chamfer Distance)

CFAM: Karmaşık Dikkat Frekans Maskesi (Complex Frequency Attention Mask)

CNN: Evrimsel Sinir Ağı (Convolutional Neural Network)

CPSA: Karmaşık Faz Kaydırma Dikkati (Complex Phase-Shifting Spatial Attention)

CPU: Merkezi İşlem Birimi (Central Processing Unit)

CV: Bilgisayarlı Görü (Computer Vision)

DiT: Yayılım Dönüştürücü (Diffusion Transformer)

DL: Derin Öğrenme (Deep Learning)

DLS: Dice Kaybı (Dice Loss)

DNN: Derin Sinir Ağları (Deep Neural Network)

FFT: Hızlı Fourier Dönüşümü (Fast Fourier Transform)

FL: Odak Kaybı (Focal Loss)

FN: Yanlış Negatif (False Negative)

FP: Yanlış Pozitif (False Positive)

GAN: Çekişmeli Üretici Ağ (Generative Adversarial Network)

GB: Gigabyte

GC-HG: Geometrik Tutarlılık-Yüksek Genelleme (Geometric Consistency-High Generalization)

GHz: Gigahertz

GP: Gradyan Cezalandırması (Gradient Penalty)

GPU: Grafik İşleme Birimi (Graphical Processing Unit)

IFFT: Ters Hızlı Fourier Dönüşümü (Inverse Fast Fourier Transform)

IoU: Kesişim Üzeri Birleşim (Intersection Over Union)

IWGAN: İyileştirilmiş Wasserstein Çekişmeli Üretici Ağ (Improved Wasserstein Generative Adversarial Network)

KL: Kullback-Leibler

LRM: Büyük Yeniden Yapılandırma Modeli (Large Reconstruction Model)

LSD-SLAM: Büyük Ölçekli Doğrudan Eş Zamanlı Konum Belirleme ve Haritalama (Large-Scale Direct Monocular Simultaneous Localization and Mapping)

MFL: Modülasyon Faktörü Kaybı (Modulation Factor Loss)

MHz: Megahertz

ML: Makine Öğrenmesi (Machine Learning)

MLP: Çok Katmanlı Algılayıcı (Multi Layer Perceptron)

MSE: Ortalama Karesel Hata (Mean Squared Error)

MSFR: Çok Ölçekli Öznitelik İyileştirme (Multi-Scale Feature Refinement)

NeRF: Nöral Işınım Alanları (Neural Radiance Fields)

RAM: Rastgele Erişimli Bellek (Random Access Memory)

ResNet: Derin Artık Ağlar (Deep Residual Networks)

RGB: Kırmızı-Yeşil-Mavi (Red-Green-Blue)

RGBA: Kırmızı-Yeşil-Mavi-Şeffaflık (Red-Green-Blue-Alpha)

RGB-D: Kırmızı-Yeşil-Mavi-Derinlik (Red-Green-Blue-Depth)

SA: Silüet Artırma (Silhouette Augmentation)

SAGAN: Öz Dikkatli Çekişmeli Üretici Ağ (Self-Attention Generative Adversarial Network)

SE: Sıkıştırma ve Uyarma (Squeeze-and-Excitation)

SN: Spektral Normalizasyon (Spectral Normalization)

SOTA: En Güncel ve En Başarılı Düzey (State of the Art)

SVD: Tekil Değer Ayırımı (Singular Value Decomposition)

SWAGE-3D: Spektral Wasserstein Dikkat Üretici Topluluk-3D (Spectral Wasserstein Attention Generative Ensemble-3D)

TN: Doğru Negatif (True Negative)

TP: Doğru Pozitif (True Positive)

VAE: Varyasyonel Otomatik Kodlayıcı (Variational Autoencoder)

VR: Sanal Gerçeklik (Virtual Reality)

VRAM: Video Rastgele Erişimli Bellek (Video Random Access Memory)

WGAN: Wasserstein Çekişmeli Üretici Ağ (Wasserstein Generative Adversarial Network)

α : Dengeleme Faktörü / Leaky ReLU negatif eğim değeri

β : Kaydırma eşiği / Momentum parametresi

γ : Odaklanma parametresi / Ölçeklendirme faktörü

ϵ : Matematiksel güvenlik / Yumuşatma katsayısı

λ : Düzenlileştirme / Ceza / Karışım katsayısı

μ : İstatistiksel ortalama

σ : Standart sapma / Sigmoid aktivasyon fonksiyonu

σ^2 : Varyans

$\odot$: Hadamard (eleman bazlı) çarpım operatörü

$\mathbb{R}$: Reel sayılar uzayı

$\mathcal{L}$: Kayıp (loss) fonksiyonu

W : Ağırlık matrisi

z : Gizli uzay vektörü

1. GİRİŞ

Tek 2D görüntüden 3D geometri üretmek neden zor bir problemdir? Bir insan tarafından tek perspektiften görülen 2D nesnenin arka planı, eksik kısımları ve 3D derinliği bilişsel ön bilgileri sayesinde kolayca tahmin edilebilmekte ve zihninde canlandırılabilir (Pizlo, 2008; Goldstein, 1989). Bununla birlikte bir makine için tek görüntüden derinlik ve hacim gibi bilgilerin çıkarılması, matematiksel açıdan doğası gereği eksik tanımlı (*ill-posed*) bir problem olarak ifade edilir (Hartley ve Zisserman, 2003). Çünkü 2D görüntü olarak ifade edilen bu yapı, aslında 3D uzayın 2D bir düzleme izdüşümüdür ve izdüşüm işlemi gerçekleştirilirken derinlik (z-ekseni) eksenindeki bilgilerin tamamı kaybedilir.

Yapay Zekâ (*Artificial Intelligence* – AI) çalışmaları makinelerin insan benzeri bir zekâ sergileyip sergileyemeyeceğine dair felsefi tartışmalarla başlamıştır (Turing, 1950). Günümüzde ise Otonom Sürüş, Robotik Navigasyon, Tıbbi Görüntüleme, Sanal Gerçeklik (*Virtual Reality* – VR) ve Artırılmış Gerçeklik (*Augmented Reality* – AR) gibi endüstriyel alanlarda bu sistemler tarafından çevrelerinin fiziksel olarak algılanması ve nesnelerin 3D yapılarının öğrenilmesine ihtiyaç duyulur. Geleneksel nesne tespiti algoritmalarında renk ve doku değişimlerinin sonuca doğrudan etki ettiği görülür (Gonzalez ve Woods, 2009). Derinlik bilgisi sayesinde bir makine tarafından sadece bir otomobil görüldüğü bilgisinin dışında o otomobile çarpılmaması için nasıl bir manevra yapılması gerektiği de öğrenilebilir. Mimari ve tasarım alanlarında da makinelerin 2D verileri kullanarak 3D nesneleri tümdengelimsel olarak üretebilmesinin hem literatürdeki eksiklikleri gidermesi hem de mühendislik uygulamalarında sağlayacağı faydalar bakımından büyük önem taşıdığı görülür.

Makine öğrenmesi (*Machine Learning* – ML) literatüründe mimariler temel olarak; verinin sınıf sınırlarını öğrenen ayırt edici (*discriminative*) ve verinin kendi dağılımını modelleyen üretici (*generative*) yaklaşımlar olmak üzere ikiye ayrılmaktadır (Ng ve Jordan, 2001). Üretici mimariler ile veri setinin istatistiksel özellikleri modellenerek yapay ancak gerçeğe oldukça yakın çıktılar üretilebilmektedir (Ruthotto ve Haber, 2021; Jebara, 2012). Bu mimarilerin başarılı olma sebeplerinden en önemlisi, yalnızca elde bulunan verinin analiz edilmekle kalmayıp aynı zamanda bu verinin altında yatan yapısal dağılımın öğrenilerek bu dağılımdan tamamen yeni, tutarlı ve özel örneklerin türetilmesidir. Varyasyonel Otomatik Kodlayıcı (*Variational Autoencoder* – VAE) ve Çekişmeli Üretici Ağ (*Generative Adversarial Network* – GAN) gibi mimarilerin geliştirilmesi ve popülerleşmesi ile birlikte üretici mimarilerde meydana gelen bu sıçrama Bilgisayarlı Görü (*Computer Vision* – CV) alanındaki en kritik problemlerden biri olan tek 2D görüntüden 3D nesneyi yeniden yapılandırma problemine yeni bir perspektif kazandırmıştır. 2D görüntülerden 3D nesne üretilmesini sağlayan 3D-VAE-GAN

mimarisi; VAE mimarisinin öznitelik temsil yeteneği ile GAN mimarisinin keskin ve net detay üretme özelliklerini birleştiren karma bir mimari olarak ifade edilebilir (Wu vd., 2016; Brock vd., 2016).

Derin Öğrenme (*Deep Learning* – DL) mimarileri aracılığıyla 3D nesnelerin başarılı şekilde üretilmesinde, bu mimarilerin 3D uzayı temsil etme biçimi hayati önem taşımaktadır. Literatürde 3D nesne temsili için voksel, nokta bulutu (*point cloud*) ve poligon ağı (*mesh*) formatları sık tercih edilmektedir (Achlioptas vd., 2018; Chen vd., 2020). Şekil 1.1’de aynı at nesnesinin poligon ağı, nokta bulutu ve voksel temsilleri ile gösterimi verilmiştir. Bu temsiller arasında voksel gösterimi, Evrişimsel Sinir Ağlarının (*Convolutional Neural Network* – CNN) 3D uzaya doğrudan uygulanabilmesini sağlaması bakımından elverişlidir. Nitekim literatürde düzensiz 3D nokta bulutu verilerinin düzenli öklidyen voksel ızgaralarına dönüştürülerek uçtan uca evrişimsel ağlarla başarılı bir şekilde işlenebileceğini gösteren öncü çalışmalar (Zhou ve Tuzel, 2018), voksel gösteriminin hacimsel öznitelik çıkarımındaki gücünü kanıtlamıştır. Bu tez çalışmasında, CNN mimarilerinin doğasına uygunluğu ve hacimsel doluluk bilgisini net bir şekilde ifade edebilmesi bakımından voksel temsili tercih edilmiştir.



Şekil 1.1. Aynı at nesnesinin poligon ağı, nokta bulutu ve voksel temsilleri ile gösterimi (Poly Haven, 2022)

Mevcut 3D-VAE-GAN mimarilerinin standart versiyonları incelendiğinde, temel olarak üç alanda yapısal kısıtlamalarının mevcut olduğu görülmektedir. İlk olarak, yeniden yapılandırma hatalarını hesaplamak için geleneksel olarak kullanılan Ortalama Karesel Hata (Mean Squared Error – MSE), İkili Çapraz Entropi (Binary Cross-Entropy – BCE), Dice ve Kesişim Üzeri Birleşim (Intersection over Union – IoU) gibi kayıp fonksiyonları, uzamsal

sürekliliği sağlamak ve yüzey voksellerinin oluşturulmasında yetersiz kalmakta; buna bağlı olarak sınırları bulanık, detaydan yoksun ve ince topolojik parçaları eksik nesneler üretilebilmektedir. Nitekim literatürdeki parça bütünlüğüne odaklanan çalışmalar da bu eksikliğe dikkat çekmektedir (Li vd., 2020; Wu vd., 2020). İkinci olarak, mevcut yaklaşımlarda evrişim katmanlarının yerel sınırlılıklarını aşmak için uzamsal düzlemde öz dikkat mekanizmaları önerilmiş olsa da (Zhang vd., 2019), uzamsal kısıtlamalar frekans düzleminin sunduğu küresel yapısal algıyı tam olarak karşılamakta yetersiz kalabilmektedir. Üçüncü olarak, geleneksel GAN mimarilerinde ayırt edici sınır ağının eğitilmesi sırasında Lipschitz süreklilik kısıtı etkin bir şekilde dayatılamamakta, buna bağlı olarak kaybolan gradyan veya gradyan patlaması gibi problemler ortaya çıkabilmektedir (Gulrajani vd., 2017). Ayrıca, tek 2D görüntünün doğrudan ağa RGB girdi olarak verilmesinin z-eksenindeki (derinlik) geometrik ipuçlarının yok olmasına ve mimari tarafından bu özniteliklere erişimin tamamen kaybedilmesine sebep olduğu görülür (Hartley ve Zisserman, 2003).

Bu tez çalışmasının temel amacı, 3D-VAE-GAN mimarisinde bulunan söz konusu eksikliklerin giderilmesi ve 2D'den 3D'ye dönüşüm kalitesini artırmak üzere bağımsız ancak kavramsal olarak birbirini destekleyen üç farklı yöntemsel iyileştirme sağlamaktır.

- Çalışmanın ilk aşamasında, hacimsel kesinliği artırmak amacıyla yeni bir yeniden yapılandırma kayıp fonksiyonu olarak AWIF önerilmiştir. Bu kayıp fonksiyonu ile nesnenin kenar yani sınır voksellerinin daha başarılı bir şekilde üretilmesi sağlanmaktadır.
- Çalışmanın ikinci aşamasında ise mimarinin uzamsal kısıtlamalara takılmadan yüksek frekanslı yapısal detayları da tespit edebilmesi için temel olarak frekans düzleminde çalışan ve küresel bütünlüğü sağlayan Karmaşık Faz Kaydırma Dikkati (*Complex Phase-Shifting Spatial Attention – CPSA*) mekanizması önerilmiştir. Önerilen bu mekanizma, 2D görüntünün doğrudan bir silüet olarak gösterilmesini sağlayan ve mimarinin nesne sınırlarına olan duyarlılığını daha da artıran Silüet Artırma (*Silhouette Augmentation – SA*) adlı veri artırma yöntemi ile desteklenmiştir.
- Çalışmanın üçüncü aşamasında ise mimarinin kararlılığının artırılması ve geometrik algılama kapasitesinin maksimuma çıkarılması amacıyla SWAGE-3D adı verilen, tümleşik geliştirmeler içeren yeni bir mimari yapı önerilmiştir. Bu mimari ile geleneksel 3D-VAE-GAN mimarisinde bulunan GAN kısmı Wasserstein GAN (WGAN) ile değiştirilmiştir. Ayrıca girdi verilerine ek bilgiler sağlamak ve geometrik ipuçlarını artırmak amacıyla Depth Anything V2 mimarisiyle çıkarılan derinlik haritaları ağa

dördüncü bir kanal olarak entegre edilmiştir. Dört kanallı bu girdiden kodlayıcı sinir ağının daha başarılı öznitelikler çıkararak gizli uzayı (*latent space*) daha etkin bir şekilde temsil etmesi için Derin Artık Ağlar (ResNet-18) mimarisi sisteme entegre edilmiştir (He vd., 2016).

Bu işlemlerin gerçekleştirilmesi için 3D nesne üretimi konusunda literatürdeki en temel veri setlerinden biri olan ShapeNet (Chang vd., 2015) seçilmiştir. IoU, Chamfer Uzaklığı (*Chamfer Distance* – CD) (Barrow vd., 1977) ve F-Skoru (Van Rijsbergen, 1979) gibi değerlendirme metrikleri raporlanmıştır. Gerçekleştirilen bu üç aşamalı iyileştirme ve yeniliğin, tek 2D görüntüden üretilen 3D nesne kalitesini mevcut başarımların metriklerinin üzerine taşıması otomatik üretim/yeniden yapılandırma literatürüne özgün bir yöntemsel katkı sunmaktadır.

1.1. Literatür Taraması

Tek 2D görüntüden 3D nesne üretimi ve yeniden yapılandırılması temel amacı akıllı etmenler tasarlamak olan AI (Russell ve Norvig, 2021) ve görsel verilerden anlamsal yapılar çıkarmayı hedefleyen CV (Szeliski, 2011) alanlarında, son yılların en aktif araştırma konularından biri haline gelmiştir. Öyle ki, yalnızca tek 2D görüntü kullanılarak insan vücudunun, yüzünün ve ellerinin 3D olarak yüksek hassasiyetle modellenenbildiği sistemlerin geliştirilmesi, bu alandaki algoritmik kapasitenin geldiği noktayı gözler önüne sermektedir (Pavlakos vd., 2019). Geleneksel parça tabanlı yaklaşımlardan (Funkhouser vd., 2004) ve matematiksel 3D şekil tanımlayıcılarından (Kazhdan vd., 2003) şekil sentezinde enerji tabanlı derin tanımlayıcı ağların kullanımına (Xie vd., 2018) ve oradan DL tabanlı üretici mimarilere (GAN, VAE, Difüzyon) kadar uzanan bu gelişim süreci ile problemin çözümüne yönelik çeşitli yöntemsel gelişmeler sunulmuştur (Ho vd., 2020). Özellikle güçlü 2D difüzyon mimarilerini yönlendirerek karmaşık metin girdilerinden yüksek kaliteli 3D nesneler üretebilen sistemlerin geliştirilmesi, bu alandaki üretim sınırlarını önemli ölçüde genişletmiştir (Poole vd., 2023). Nokta bulutu temsili kullanılarak karmaşık girdilerden saniyeler içinde 3D nesneler üretebilen sistemler de geliştirilmiştir (Nichol vd., 2022). Benzer şekilde, transformatör mimarilerinin gücünü doğrudan 3D poligon üretimine taşıyan güncel sistemler de yüksek kaliteli ve tutarlı yüzey inşasında literatürde büyük dikkat çekmektedir (Yan vd., 2022; Siddiqui vd., 2024). Nesnelerin geometrik detayları ile parçalar arası yapısal ilişkileri eş zamanlı modelleyerek yapı farkındalıklı üretim gerçekleştiren ağlar da geometrik tutarlılık açısından kritik bir role sahiptir (Wu vd., 2019). Bunların da ötesinde, tek 2D görüntüden sadece geometriyi değil nesnenin malzeme özelliklerini ve yüzey yansımalarını da tahmin ederek 3D modelleme gerçekleştiren yenilikçi mimariler, alandaki üretim sınırlarını fiziksel gerçekçilik boyutuna taşımıştır (Wang vd., 2025). Mevcut mimarilerin gelişiminin ve karşılaştıkları sınırlılıkların 3D veri temsil

yöntemleri, kayıp fonksiyonu optimizasyonları, dikkat mekanizmaları ve üretici ağ mimarileri ekseninde tematik olarak incelenmesi, literatürdeki yapısal eksikliklerin nesnel bir şekilde tespit edilebilmesi açısından kritik olarak ifade edilebilir.

Bu bağlamda, 3D nesne üretimi alanında literatürde gerçekleştirilen temel çalışmalar ve yöntemsel yaklaşımlar sınıflandırılarak Tablo 1.1’de özetlenmiştir.

Tablo 1.1. 3D nesne üretimi ve yeniden yapılandırma literatür özeti

Kategori	Referans	Önerilen Mimari	Temel Katkı
Tek Görüntüden 3D Yeniden Yapılandırma	(Choy vd., 2016)	3D-R2N2	Tek ve çoklu görünüm görüntülerden voksel tabanlı 3D nesne üretimi sağlanmış ve IoU metriği bir standart haline getirilmiştir.
	(Mandikal vd., 2018)	3D-LMNet	İki aşamalı bir eğitim süreci izlenerek nokta bulutu tabanlı 2D görüntülerden 3D nesne üretimi sunulmuştur.
	(Saito vd., 2019)	PIFu	Görüntü tabanlı öznetelik çıkarımı ile 3D örtük fonksiyon birleştirilerek tek görüntüden yüksek detaylı 3D insan modeli üretilmiştir.
	(Liu vd., 2024)	One-2-3-45++	Güçlü 2D görüntü difüzyon modelleri kullanılarak tek görüntüden çok kısa sürede 3D nesne üretilmesi amaçlanmıştır.
Voksel, Nokta Bulutu, Poligon ve Örtük Temsil Yöntemleri	(Wu vd., 2015)	3DShapeNets	Derin İnanç Ağları kullanılarak nesnelerin voksel tabanlı gösterimle ve olasılıksal dağılımlarla temsil edilmesi sağlanmıştır.
	(Fan vd., 2017)	PSGN	Geleneksel kayıp fonksiyonlarının yetersiz kalacağı öngörülerek nokta bulutu temsiliinde 3D nesne üretimi gerçekleştirilmiştir.
	(Wang vd., 2018)	Pixel2Mesh	Voksel tabanlı yaklaşımların yüksek bellek tüketimi sorununu aşmak üzere doğrudan poligon ağı üreten graf evrişimsel sinir ağı tabanlı mimari sunulmuştur.

Tablo 1.1. Devamı

	(Mescheder vd., 2019)	Occupancy Networks	Nesne yüzeyinin sürekli bir karar sınırı olarak modellendiği sürekli (<i>continuous</i>) bir 3D fonksiyon ve örtük (<i>implicit</i>) öğrenme yöntemi önerilmiştir.
	(Mildenhall vd., 2020)	NeRF	2D görüntülerden gerçekçi yeni açılar sentezleyen ve 3D sahne temsiline paradigma değiştiren Nöral Işınlam Alanları (<i>Neural Radiance Fields</i> – NeRF) yöntemi sunulmuştur.
3D-VAE-GAN ve GAN Tabanlı Üretim	(Wu vd., 2016)	3D-GAN / 3D-VAE-GAN	3D nesne üretimi için Hacimsel Evrişimsel Sinir Ağı kullanılan ve 200 boyutlu gizli bir vektörden voksel nesneler üreten DL tabanlı yeni bir yaklaşım sunulmuştur.
	(Smith ve Meger, 2017)	3D-IWGAN	3D-VAE-GAN'daki kaybolan gradyan ile mod çökmesi problemlerinin giderilmesi amacıyla Wasserstein uzaklık metriği kullanılmıştır.
	(Yu vd., 2024)	VAE-Cascade-GAN	VAE'nin olasılıksal modelleme yeteneği GAN'ın keskin detay koruma kapasitesiyle birleştirilerek hibrit yapı elde edilmiştir.
Kayıp Fonksiyonları ve Yüzey/Geometri Duyarlı Kayıplar	(Rahman ve Wang, 2016)	IoU Kaybı	Koordinatlardaki sayısal fark yerine tahmin edilen ve gerçek veri arasındaki örtüşme kalitesinin doğrudan optimize edilmesi hedeflenmiştir.
	(Milletari vd., 2016)	Dice Kaybı	Voksel tabanlı verilerdeki sınıf dengesizliğine karşı direnç gösterilmiş ve ince topolojik detayların korunması sağlanmıştır.
	(Lin vd., 2017b)	Focal Loss (FL)	Baskın arka plan piksellerinin ezberlenmesi önlenerek, öğrenilmesi zor olan nesne sınırlarına ve detaylara odaklanılması sağlanmıştır.
Dikkat ve Frekans Düzleminde Dikkat Yaklaşımları	(Hu vd., 2018)	SE Ağları	Kanalları tek değere sıkıştırarak elde edilen önem skorlarıyla öznitelikleri yeniden uyaran kanal bazlı dikkat mekanizması kazandırılmıştır.

Tablo 1.1. Devamı

	(Woo vd., 2018)	CBAM	Kanal ve uzamsal dikkati ardışık olarak birleştirerek sistemin “neye” ve “nereye” odaklanacağını öğrenmesi sağlanmıştır.
	(Zhang vd., 2019)	SAGAN	Öz Dikkat (Self-attention) tekniği kullanılarak uzun mesafeli uzamsal bağımlılıkların modellendiği yapı ile uzamsal kısıtlamalar aşılmıştır.
	(Wu vd., 2025)	Direct3D-S2	Yayılım Dönüştürücülerin (<i>Diffusion Transformer</i> – DiT) işlem yükünü hafifleten ve hesaplama sorunlarını çözen Simetrik Seyrek Dikkat mekanizmasını literatüre kazandırmıştır.
RGB-D ve Derinlik Destekli Yeniden Yapılandırma	(Engel vd., 2014)	LSD-SLAM	Monoküler kameralar için önerilen doğrudan yaklaşım ile RGB-D sensör verilerinin ve derinlik bilgisinin işlenmesi için etkili bir zemin hazırlanmıştır.
	(Gong vd., 2025)	GC-HG	Ağın geometriyi yanlış algılamasını engellemek için mimariye derinlik önbilgisi (<i>depth prior</i>) ve sözde üç-düzlemli gösterim entegre edilmiştir.
En Güncel ve En Başarılı Düzey (<i>State of the Art</i> – SOTA) Voksel Tabanlı Yöntemler	(Xie vd., 2019)	Pix2Vox	Tek görüntüden üretim yapan CNN tabanlı ve aynı nesnenin birden çok açıdan görüntüsünü alan bağlam farkındalığına sahip mimari önerilmiştir.
	(Ren vd., 2024)	XCube	Yüksek çözünürlüklü ve seyrek voksel ızgaralarını bellek taşırmadan işleyebilmek için tasarlanan büyük ölçekli 3D üretici mimari önerilmiştir.
	(Hong vd., 2024)	LRM	Tek 2D görüntüden yüksek çözünürlüklü 3D şekil üretimini saniyeler içinde tamamlayabilen transformatör tabanlı devasa mimari sunulmuştur.
	(Voleti vd., 2024)	SV3D	Tek 2D görüntü girdisi olarak yapay bir video üreten ve ardından bu videoyu 3D yapıya dönüştüren çok kipli SOTA yaklaşım sunulmuştur.

Literatürdeki çalışmalar genel olarak değerlendirildiğinde, üretici AI mimarilerinin 2D görüntülerden 3D nesne sentezleme konusunda büyük bir algoritmik olgunluğa ulaştığı ve

özellikle örtük fonksiyonlar ile yönlendirilen mimarilerin saniyeler içinde karmaşık 3D temsiller üretebildiği (Wang vd., 2021; Müller vd., 2022; Jun ve Nichol, 2023) görülmektedir (Gao vd., 2022). Ayrıca NeRF'e güçlü bir alternatif olarak geliştirilen ve gerçek zamanlı 3D işleme (*rendering*) imkânı sunan 3D Gauss saçılımı (*3D Gaussian Splatting* – 3DGS) gibi yeni nesil yöntemler, alanın görsel kalite standartlarını önemli ölçüde yükseltmiştir (Kerbl vd., 2023). Bu doğrultuda, tek 2D görüntüden tutarlı çoklu bakış açılı görüntüler sentezleyerek doğrudan 3D Gauss gösterimine geçiş yapan sistemler, tekil görüntülerden yüksek kaliteli üretim sınırlarını daha da ileri taşımıştır (Yang vd., 2025). Bununla birlikte özellikle voksel tabanlı 3D-VAE-GAN mimarileri incelendiğinde standart eğitim ve yüzey voksellerine az önem verilmesi gibi problemlerden kaynaklanan bulanık kenar üretimlerinin geleneksel kayıp fonksiyonlarıyla aşamadığı, sadece uzamsal dikkat mekanizmalarının kullanımının yüksek frekanslı yapısal detayları yakalamakta yetersiz kaldığı ve GAN eğitimindeki Lipschitz sürekliliğinin mutlak suretle sağlanamamasının optimizasyon kararsızlıklarına yol açtığı görülür. Ayrıca tek girdi kullanılan yaklaşımlarda z-ekseni (derinlik) bilgisinin ağı entegre edilmemesinin geometrik algıyı kısıtlayan temel bir darboğaz olduğu ifade edilebilir. Sonuç olarak literatürde tespit edilen bu yapısal ve algoritmik boşluklar ile eğitim kararlılığını garanti altına alan, frekans düzleminde öznitelik çıkarabilen, kenarları daha başarılı bir şekilde üretebilen ve hacimsel kesinliği artıran mimarilere ihtiyaç olduğu tespit edilmiştir. İlerleyen bölümlerde detaylandırılacak olan bu tez çalışması, literatürde gözlemlenen söz konusu eksiklikleri yöntemsel bir yaklaşımla gidermek amacıyla yapılmıştır.

1.2. Çalışmanın Katkıları

Bu tez çalışmasında tek 2D görüntü kullanılarak 3D nesne üretimi problemi vokselize temsil kullanılarak ele alınmıştır. 3D-VAE-GAN mimarisi üzerinde yapılan aşamalı geliştirmelerin katkıları detaylı bir şekilde incelenmiştir. Tek 2D görüntüden 3D nesne üretimi, görüntüdeki derinlik ve arka yüz gibi bilgilerin eksikliği nedeniyle doğası gereği belirsiz bir problem olarak ifade edilir. Buna bağlı olarak başarılı bir üretim için sadece üretici bir mimariye değil aynı zamanda yüzey yapısını koruyacak etkili kayıp fonksiyonlarına, 2D görüntüden daha başarılı öznitelikler çıkarabilecek dikkat mekanizmalarına, önemli özniteliklere daha yüksek ağırlık verilmesine, frekans düzleminin getirilerinden faydalanılmasına ve çekişmeli eğitimin kararlı hale getirilmesine ihtiyaç duyulmaktadır.

Bu kapsamda tezdeki metodoloji birbirinden bağımsız ama destekleyici üç eksen üzerinde yapılandırılmıştır. İlk olarak, voksel tabanlı temsillerdeki yüzey voksellerinin yeterince başarılı üretilmemesine karşın Uyarlamalı Ağırlıklandırılmış IoU-Odak Kaybı (*Adaptive Weighted IoU-Focal Loss* – AWIF) ismi verilen yeniden yapılandırma kayıp fonksiyonu

önerilmiştir. Buna bağlı olarak üretilen nesnenin hacimsel olarak örtüşmesi yanında yüzey bütünlüğünün de korunması hedeflenmiştir.

İkinci olarak, tek 2D görüntüden elde edilen temsilin içerdiği yapısal bilgi yetersiz kalabilmektedir. Bunun için kodlayıcı tarafında daha güçlü ve anlamlı öznitelik çıkarımı amaçlanmıştır. Uzamsal düzlemde çalışan klasik dikkat mekanizmalarının süreklilik kaybı ve alıcı alan kısıtlamaları gibi eksikliklerinin giderilmesi amacıyla frekans düzlemindeki faz ve genlik bileşenlerinden yararlanan CPSA dikkat mekanizması önerilmiştir. CPSA ile nesnenin genel yapısal formunun ve uzun mesafeli ilişkilerin öğrenilerek mimarinin daha duyarlı hale gelmesi amaçlanmıştır. Nesnenin genel formunun korunması amacıyla doku ve renk bilgisini oldukça baskılayan SA veri artırma tekniği de CPSA ile birlikte önerilmiştir.

Üçüncü olarak, aynı temel problem bütünleşik yeni bir mimari ile özellikle eğitim kararlılığı ekseninde değerlendirilmiştir. RGB-D girdi, ResNet-18 tabanlı kodlayıcı, WGAN-GP (*Gradient Penalty* – GP) entegrasyonu, Spektral Normalizasyon (*Spectral Normalization* – SN), öz dikkat bileşeni ve topluluk değerlendirme stratejileri en uygun bir şekilde bir araya getirilerek SWAGE-3D mimarisi oluşturulmuştur. Böylece tez çalışması AWIF, CPSA ve SWAGE-3D mimarileri kullanılarak aynı 3D yeniden yapılandırma problemini kayıp fonksiyonu, temsil öğrenme (Bengio vd., 2013) ve eğitim kararlılığı olmak üzere üç farklı teknik ekseninde ele alacak şekilde konumlandırılmıştır.

Bu bütüncül yaklaşım sonucunda yürütülen deneysel çalışmalarda AWIF’in farklı kayıp fonksiyonları ile kıyaslanması, CPSA’nın farklı dikkat mekanizmaları ile değerlendirilmesi ve SWAGE-3D’nin temel 3D-VAE-GAN ile literatürdeki diğer ilgili yöntemler karşısındaki başarımının analizi gerçekleştirilmiştir. Elde edilen bulgular IoU, CD ve F-Skoru gibi farklı metrikler üzerinden değerlendirilmiştir. Böylece sunulan katkıların yalnızca hacimsel örtüşme açısından değil yüzey geometrisi ve yapısal tutarlılık açısından da değerlendirilmesi sağlanmıştır.

Literatüre sunulan başlıca katkılar ve yenilikler aşağıda maddeler halinde özetlenmiştir. Sunulan her bir katkı ile problemin farklı bir sınırlılığı hedeflenmiştir.

- **Özgün Yeniden Yapılandırma Kayıp Fonksiyonu:** Voksel tabanlı 3D nesne yeniden yapılandırmasında yüzey ve sınır vokselleri yeterince başarılı üretilmemektedir. Bu problemin çözülmesi amacıyla; Yüzey Duyarlı Odak (*Surface-aware Focal*), IoU Rehberli Ağırlıklandırma (*IoU-guided Weighting*), Uyarlanabilir Zorluk Aşamalandırması (*Adaptive Difficulty Progression*) ve Sınır İyileştirme (*Boundary Enhancement*) bileşenlerini bütünleştiren karma yapıli özgün bir AWIF kayıp fonksiyonu

önerilmiştir. Voksel temsili kullanılan çalışmalarda 3D uzayın büyük bir bölümü boş voksellerden meydana gelmektedir. Buna bağlı olarak standart kayıp fonksiyonları tarafından nesnenin yüzey ve sınır voksellerine daha az odaklanılmaktadır. Böylece üretilen 3D nesnelerde yüzey bütünlüğü bozularak ince yapısal detaylar kaybolabilmektedir. AWIF ile yalnızca hacimsel örtüşme değil aynı zamanda yüzey yapısının korunması da hedeflenmektedir. Çalışma kapsamında; BCE, MSE, Modülasyon Faktörü Kaybı (*Modulating Factor Loss* – MFL), Dice ve IoU gibi geleneksel fonksiyonlar ile AWIF hem nicel hem de nitel olarak kıyaslanmıştır. Sonuç olarak AWIF'in özellikle yüzey yapısı bakımından daha başarılı sonuçlar elde ettiği görülmüştür. Bu katkı tezin genel mimarisinde yeniden yapılandırma kaybının önemini ön plana çıkararak üretici ağı geometrik tutarlılığa sahip 3D nesneler üretmesini sağlamaktadır.

- **Frekans Düzleminde Dikkat Mekanizması ve Silüet Artırma:** Mimarinin geleneksel uzamsal kısıtlamaları aşması ve tüm öznitelik haritası üzerinde küresel yapısal etkileşimi sağlayabilmesi amacıyla frekans düzleminde çalışan CPSA mekanizması geliştirilmiştir. CPSA ile 2D girdiden elde edilen öznitelik temsili güçlendirilmektedir. Tek 2D görüntüde derinlik ve 3D geometrik yapıya ilişkin bilgiler doğrudan bulunmadığı için bunların tahmin edilmesi gerekir (Yang vd., 2024a). Standart CNN mimarisine dayanan kodlayıcı ağlar yerel öznitelikleri başarılı bir şekilde çıkarabilse bile nesnenin genel geometrik formunu ve uzun mesafeli ilişkileri anlamada sınırlı kalabilmektedir. CPSA ile bu sınırlılığın azaltılması hedeflenerek frekans düzlemindeki faz (*phase*) ve genlik (*magnitude*) bileşenlerinden yararlanılmıştır. CPSA mekanizması SE, CBAM ve temel 3D-VAE-GAN mimarileri ile nicel ve nitel olarak kıyaslanmış böylece frekans düzleminde uygulanan dikkat mekanizmasının 3D nesne üretimi konusundaki katkısı gösterilmiştir. Bu yaklaşım mimarinin doku ve renk bağımlılıklarının azaltılarak doğrudan geometrik sınırların öğrenilmesini zorlayan SA tekniği ile desteklenmiştir. Özünde bir veri artırma tekniği olan SA bu problem özelinde özellikle küresel bütünlüğün korunması amacıyla kullanılmıştır. Önerilen bu bütüncül yaklaşım, 3D nesne yeniden yapılandırmasında küresel yapıları yakalama kapasitesini artırdığı nicel ve nitel olarak kanıtlanmıştır.
- **Tümleşik ve Kararlı Yeni Bir Mimari:** Geleneksel 3D-VAE-GAN mimarisindeki geometrik algı eksikliği ve optimizasyon kararsızlıklarının giderilmesi amacıyla SWAGE-3D mimarisi önerilmiştir. Bu mimari kapsamında literatüre sunulan yöntemsel yenilikler şunlardır:

- **Derinlik Destekli ResNet-18 ile RGB-D Kodlama:** Tek 2D görüntüde derinlik bilgisi bulunmadığı için mimari tarafından bu kritik bilgi öğrenilememektedir. Bu durum özellikle nesnenin hacimsel yapısının ve ilgili açıdan görünmeyen bölgelerinin tahmin edilmesini zorlaştırmaktadır. Bu eksikliğin giderilmesi amacıyla Depth Anything V2 mimarisi ile derinlik haritalarının ağı dördüncü bir kanal olarak dâhil edildiği bir RGB-D girdi stratejisi uygulanmıştır. Bu doğrultuda kodlayıcı olarak eklenen ResNet-18 mimarisi ortalama ile başlatılmış kanal genişletme yöntemiyle desteklenerek bu derinlik bilgisine uyarlanmış ve kararlı bir transfer öğrenme süreci sağlanmıştır. Başarımı artırmak, daha kararlı, gürültüye dayanıklı ve başarılı bir mimari elde edilmesi amacıyla 2D *mixup* veri artırma tekniği kullanılmıştır. Bu katkı SWAGE-3D mimarisinde girdi temsilini güçlendiren temel bileşenlerden biri olarak yer almaktadır.
- **Kararlı Çekişmeli Hacimsel Eğitim:** GAN mimarisi kullanılan çalışmalarda eğitim kararsızlığı probleminin önüne geçilmesi amacıyla WGAN-GP ve SN birlikte kullanılmıştır. GAN mimarilerinde mod çökmesi, kaybolan gradyan ve kararsız eğitim süreçleri önemli problemler olarak ifade edilir. Bu problemler özellikle yüksek boyutlu voksel üretiminde daha belirgin hale gelebilmektedir. Ayırt edici ağı tarafında GP ve SN'nin birlikte kullanımı ile Lipschitz kısıtının daha kararlı biçimde sağlanması hedeflenmiştir. Bu katkı tezin genel mimarisinde çekişmeli öğrenme sürecini kararlı hale getiren bir bileşen olarak konumlandırılmıştır.
- **Dikkat Destekli Üretici:** Dikkat mekanizmaları kullanılmayan mimarilerde girdiler arasındaki uzun mesafeli bağımlılıklar modellenememekte ve buna bağlı olarak daha kötü sonuçlar elde edilebilmektedir. Bu sorunun çözülmesi ve üretilen 3D nesnelerdeki yapısal tutarlılığın artırılması amacıyla üretici ağı bir öz dikkat (*self-attention*) mekanizması entegre edilmiştir (Vaswani vd., 2017). Bu katkı tezin genel mimarisinde bilgilerin birlikte değerlendirilerek daha başarılı sonuçlar üretilmesi bakımından önemlidir.
- **Kontrol Noktası Tabanlı Topluluk Çıkarımı:** DL mimarilerinde farklı turlarda (*epoch*) yapılan tahminler büyük farklılık gösterebilmekte ve

buradaki salınım yüksek olabilmektedir. Tek tura dayalı olarak karar verilmesi mimarının genel başarımını sınırlandırabilmekte veya mimarının belirli eğitim anlarına aşırı bağımlı hale gelmesine sebep olabilmektedir. Bu gibi problemlerin önüne geçmek ve bu değişimi azaltmak amacıyla tek tura değil IoU metriğine dayalı olarak tespit edilen ve öğrenmenin yavaşladığı plato noktalarına göre bir tahmin yapılması sağlanmıştır. Bu katkı SWAGE-3D mimarisinde test aşamasının tahmin güvenilirliğini artıran ve tur bazlı değişimi azaltan tamamlayıcı bir bileşen olarak konumlandırılmıştır.

Bu tez kitabının geri kalan kısmı şu şekilde düzenlenmiştir: İkinci bölümde görüntü temsili, kullanılan veri setleri, temel DL mimarileri, kayıp fonksiyonları ve dikkat mekanizmaları detaylandırılmaktadır. Üçüncü bölümde deneysel kurulum, analiz düzenekleri ve başarımlar ölçütleri sunulmaktadır. Dördüncü bölümde elde edilen nicel ve nitel bulgular tartışılarak analiz edilmekte, beşinci bölümde ise çalışmanın genel bir değerlendirilmesi yapılmaktadır.

2. KULLANILAN SAYISAL VERİ, YÖNTEMLER ve ALGORİTMALAR

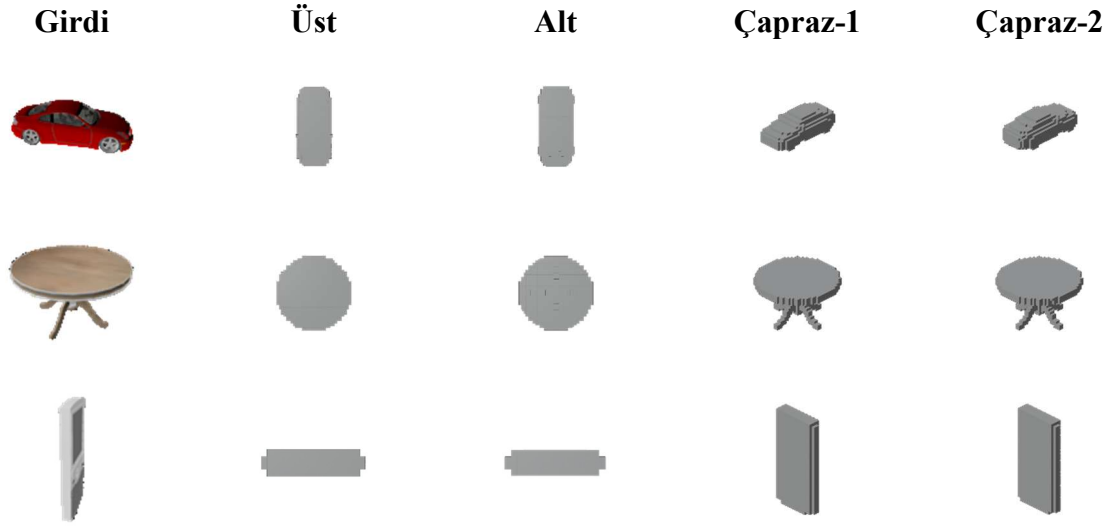
Bu bölümde, tek 2D görüntüden 3D nesne üretimi ve yeniden yapılandırılması probleminin çözülmesi amacıyla geliştirilen yöntemsel yaklaşımlar ve algoritmik yenilikler tüm detaylarıyla açıklanmaktadır. Kullanılan veri setleri, ön işleme adımları ve veri artırma teknikleri bu bölümde detaylandırılmıştır. Çalışmanın yöntemsel temeli, geleneksel 3D-VAE-GAN mimarisinde sıklıkla karşılaşılan geometrik tutarsızlık, mod çökmesi, eksik detaylar, hacimsel bütünlük eksikliği ve optimizasyon kararsızlığı gibi problemlerin aşılması için geliştirilen voksel tabanlı yeni üretici mimarilerden oluşur.

Bu kapsamda ilk olarak, literatürdeki mevcut kayıp fonksiyonlarının hacimsel veri üzerindeki sınırlılıkları analiz edilmiş ve yüzey geometrisine doğrudan odaklanarak yeniden yapılandırma hassasiyetini artıran AWIF kayıp fonksiyonu literatüre kazandırılmıştır. Ağın eğitim aşamasında özellikle renk ve doku gibi özelliklere fazla güvenmesi probleminin aşılması ve ağın yalnızca küresel geometriyi öğrenmesinin zorunlu kılınması amacıyla SA veri artırma tekniği uygulanmıştır. Geleneksel uzamsal dikkat modüllerinin eksikliklerinin giderilmesi amacıyla faz ve genlik bileşenlerini modüle eden, frekans düzlemi tabanlı CPSA mekanizması kodlayıcı ağı entegre edilmiştir. Derinlik destekli RGB-D kodlama, SN ve GP ile kararlı hale getirilmiş Wasserstein çekişmeli öğrenme altyapısının birlikte sunulduğu SWAGE-3D mimarisi önerilmiştir. Bu mimaride öz dikkat destekli bir üretici ve topluluk çıkarım stratejisi kullanılmıştır.

2.1. Kullanılan Veri Seti ve Görüntü Temsili

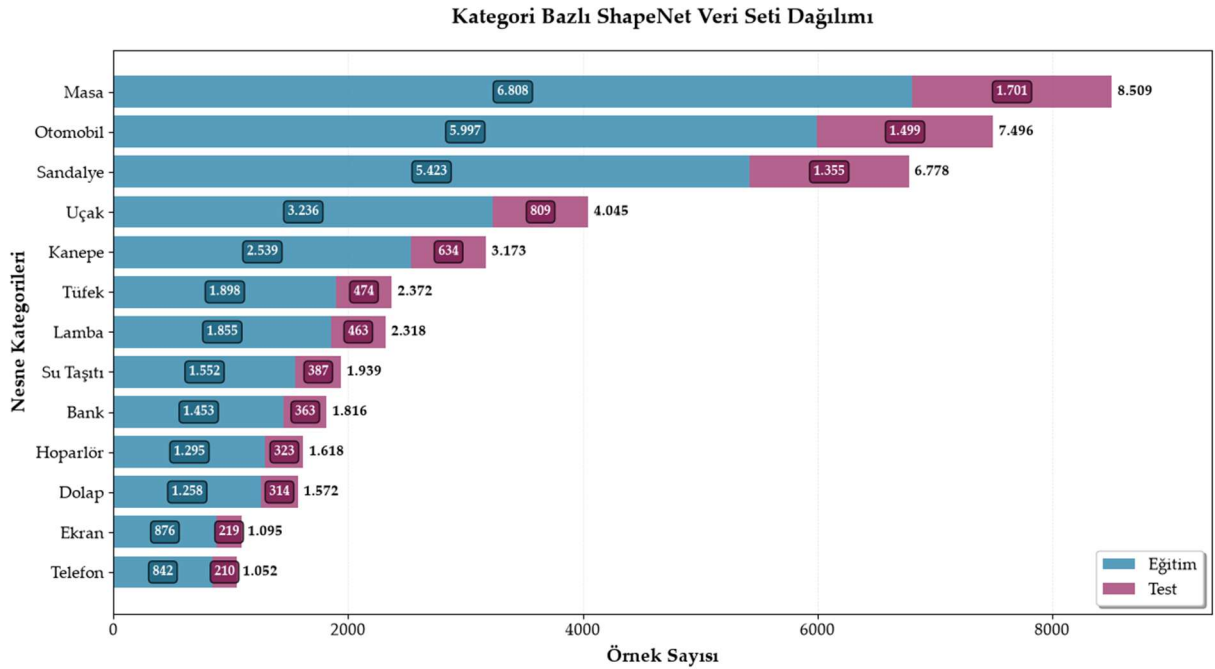
Çalışma kapsamında 3D nesne tanıma, yeniden yapılandırma ve üretici AI alanlarında literatürde yaygın kullanılması sebebiyle ShapeNet veri seti kullanılmıştır (Chang vd., 2015). Toplamda 55 farklı nesne kategorisi ve 51.300'den fazla 3D nesne bu veri setinde bulunmaktadır. Anlamsal olarak etiketlenmiş hiyerarşik yapısı sayesinde 3D nesnelerin çok açılı 2D görüntüleri ve bunlara karşılık gelen nesneler bir arada sunulmakla birlikte bu çalışmada yalnızca tek açı kullanılmıştır.

Güncel çalışmalarla tutarlı ve adil bir karşılaştırma yapılabilmesi amacıyla veri setinden 13 temel sınıf (uçak, bank, dolap, otomobil, sandalye, ekran, lamba, hoparlör, tüfek, kanep, masa, telefon ve su taşıtı) seçilmiştir. Ağın eğitimi ve başarımı, bu 13 sınıfa ait toplam 34.963 örnek üzerinden değerlendirilmiştir. Veri seti, literatürdeki standart yaklaşımlara uygun olarak %80 eğitim (27.972 örnek) ve %20 test (6.991 örnek) olmak üzere iki alt kümeye ayrılmıştır. Şekil 2.1'de otomobil, masa ve telefon sınıfları üzerinden sistemin referans aldığı gerçek verilerin temsilleri gösterilmiştir.



Şekil 2.1. Otomobil, masa ve telefon nesnelerinin 2D ve farklı açılardan 3D görüntüleri

Veri seti bölünürken her bir nesne örneği düzeyinde tek tek ayırma işlemi gerçekleştirilmiştir. Kategoriler arası veri dağılımı incelendiğinde dağılımın homojen olmadığı görülür. Şekil 2.2’de her bir kategoriden eğitim ve test aşamalarında kaçör örnek kullanıldığı detaylı bir şekilde gösterilmiştir. Şekilden görüleceği üzere en fazla örneğin masa, ortalama sayıda örneğin lamba ve en az örneğin ise telefon sınıfında bulunduğu gözlemlenmiştir.

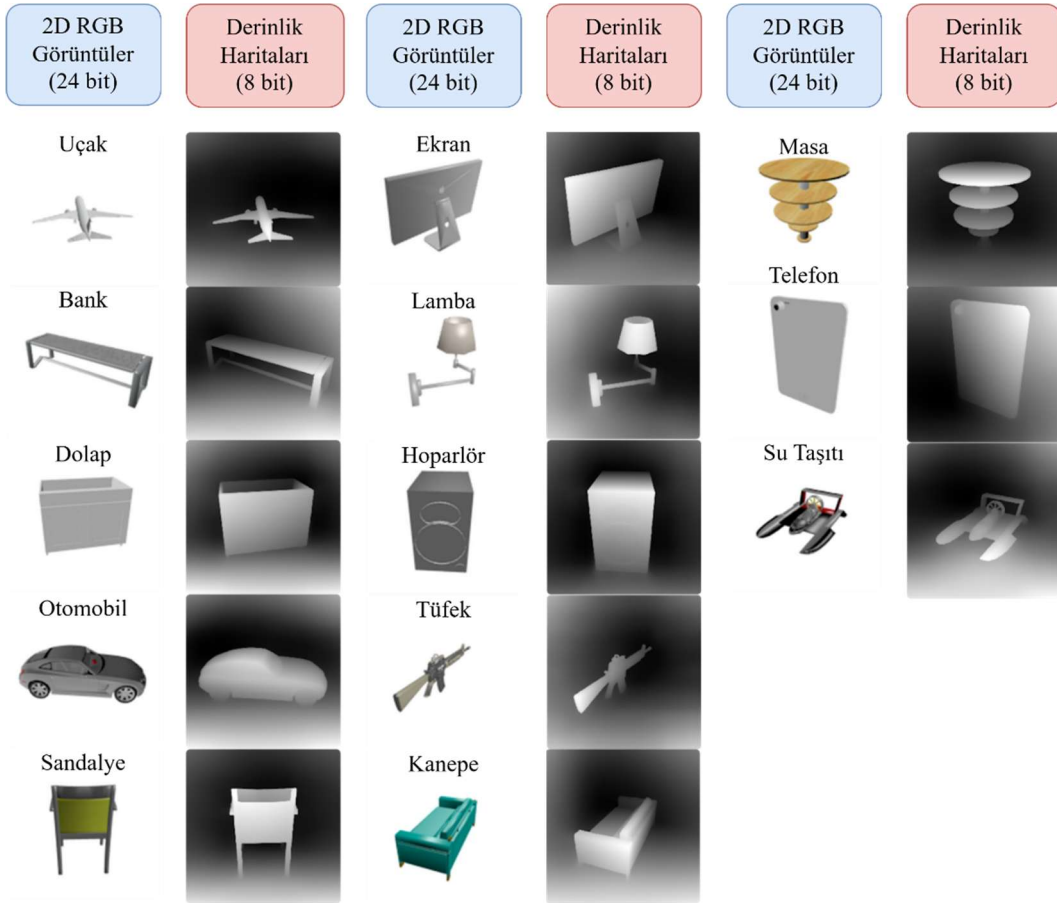


Şekil 2.2. Eğitim ve test aşamalarında nesne bazlı kullanılan örnek sayıları

Kullanılan orijinal 2D girdi görüntülerinin boyutunun 137x137x4 olarak veri setinde bulunduğu görülür. Önerilen AWIF ve CPSA yaklaşımlarında bu orijinal boyutlara sadık kalınmış ve doğrudan bu boyutlarda girdiler kullanılmıştır. SWAGE-3D mimarisinde ise kodlayıcı ağı omurgası ResNet-18 mimarisi tarafından oluşturulmaktadır. Bu mimaride varsayılan olarak 224x224x3 boyutunda girdi beklenmektedir. Bu yapısal uyumluluğun

sağlanması amacıyla görüntüler başlangıçta 224x224 piksel çözünürlüğüne getirilmiş ve alfa kanalı atılarak 3 kanallı bir yapıya dönüştürülmüştür.

SWAGE-3D mimarisinde geometrik algının güçlendirilmesi, aydınlatma değişimlerinden kaynaklanan sorunların hafifletilmesi ve nesne sınırlarının daha net belirlenmesi amacıyla standart RGB tensörlerine dördüncü bir kanal olarak derinlik bilgisi entegre edilmiştir. Bu işlem için Depth Anything V2 mimarisi kullanılarak her bir 2D görüntüden monoküler (tek 2D görüntüden derinlik bilgisi çıkarılması) derinlik haritaları (*depth maps*) çıkarılmıştır (Yang vd., 2024b). 8 bit gri tonlamalı (*grayscale*) olarak elde edilen bu haritaların RGB girdi tensörü ile birleştirilmesiyle 224x224x4 boyutlarında RGB-D tensörleri elde edilmiştir. ResNet-18 mimarisinin varsayılan olarak üç kanallı girdi kabul etmesine bir çözüm olarak mekansal derinlik bilgisi buna uygun dört kanallı bir tensör topolojisine dönüştürülmüştür. Veri ön işleme aşamasında bağımsız olarak normalize edilen RGB ve derinlik tensörleri kanal eksenini bozmadan birbirine eklenmiş ve tek girdi tensörü elde edilmiştir. Ağı bu yeni girdi boyutuna uyarlamak için ResNet-18 mimarisinin ilk evrişim katmanı dört giriş kanalı kabul edecek şekilde değiştirilmiş ve parametre başlatma sürecinde kısmi transfer öğrenme stratejisi kullanılmıştır. Ağırlıkları içeren tensörün ilk üç kanalı doğrudan ImageNet (Russakovsky vd., 2015) üzerinde önceden eğitilmiş ResNet-18 mimarisinin orijinal RGB filtrelerinden alınırken, sonradan eklenen derinlik kanalının ağırlıkları ise mevcut üç RGB kanalının uzamsal filtre ağırlıklarının aritmetik ortalaması alınarak uygulanmıştır. Bu işlem ile derinlik kanalının sıfırdan veya rastgele dağılımlarla başlatılması yerine; ağın RGB verisinden öğrendiği düşük seviyeli kenar ve köşe gibi öznitelikler korunarak doğrudan yeni mimariye aktarılmış ve böylece eğitim kararlılığı artırılarak yakınsama süreci iyileştirilmiştir. Şekil 2.3'te veri setinde bulunan her bir sınıftan birer örnek 24 bit 2D RGB görüntü ve bunlara karşılık Depth Anything V2 mimarisi ile çıkarılan 8 bit RGB-D derinlik haritaları gösterilmiştir.



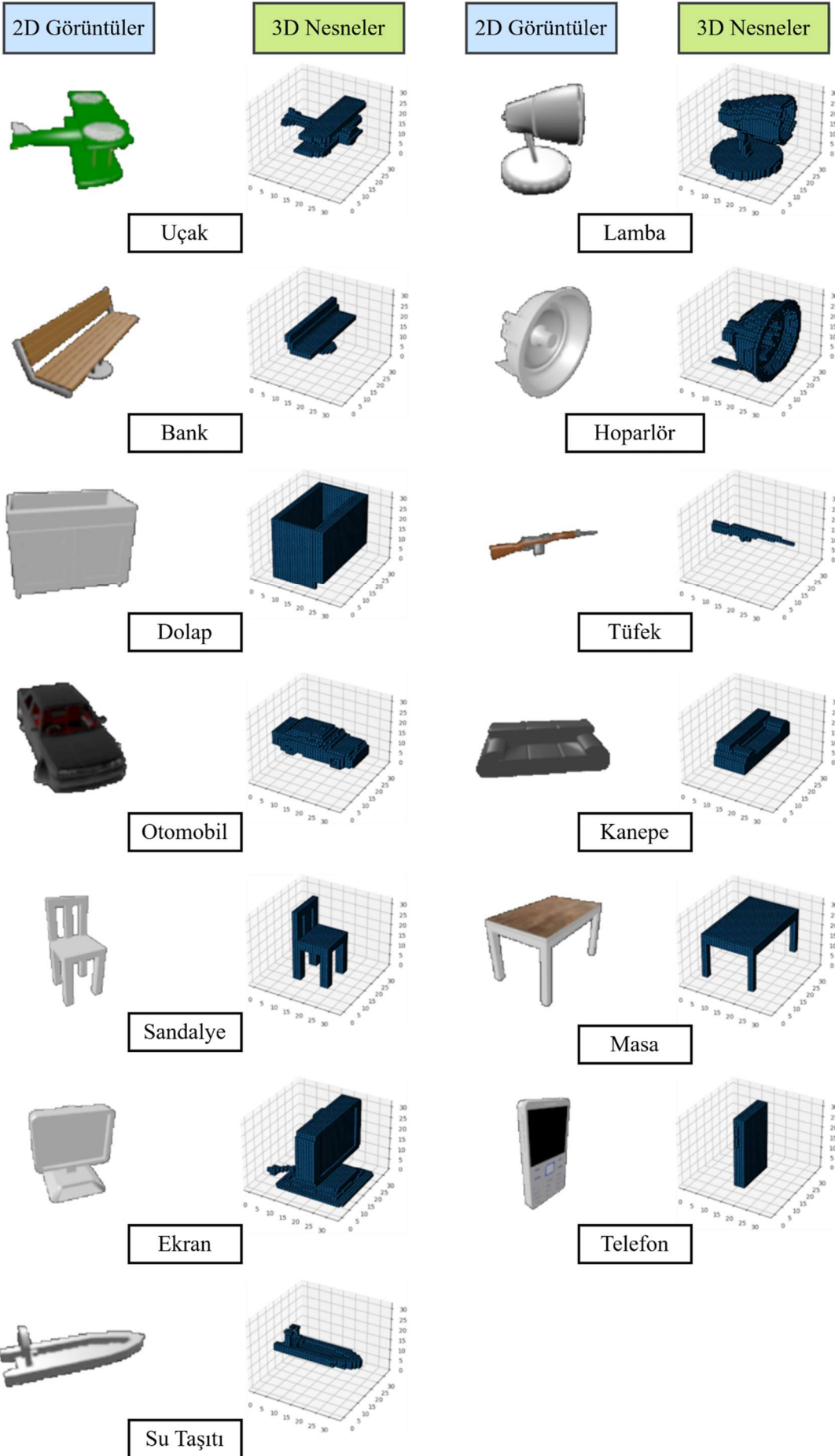
Şekil 2.3. 13 sınıfa ait 2D görüntüler (24-bit) ve karşılık gelen Depth Anything V2 tabanlı derinlik haritaları (8-bit)

Sistemin nihai çıktısı olarak üretilecek 3D nesnelerin hacimsel veri gösterimi için voksel temsili benimsenmiştir. Nokta bulutları veya çokgen ağları gibi alternatif 3D veri temsillerinin, yüzey geometrisini 3D uzayda yüksek potansiyel ile temsil etme yetenekleri bulunsada (Ben-Hamu vd., 2018; Guo vd., 2021) bu temsillerin topolojik olarak düzensiz yapıda olmaları nedeniyle, CNN gibi yapılarla doğrudan kullanılabilmesi oldukça güçleşmektedir. Öyle ki, literatürde nokta kümelerinin bu düzensiz yapısını standart evrişim işlemleri olmadan doğrudan işleyebilmek adına DL mimarilerinin geliştirilmesine ihtiyaç duyulmuştur (Qi vd., 2017). Voksel temsili ise 2D ortamda ifade edilen piksellerin 3D ortama genişletilmiş hali olarak ifade edilebilir. Bu temsil ile ilgili veri düzenli ve öklidyen bir ızgara formatında ifade edilebilir. Bu matematiksel düzenlilik ile DL mimarilerindeki tensör operasyonları 3D uzaya en az asimetrik sapma ile genelleştirilebilir.

Voksel kavramı aslında 3D uzay içindeki her bir ayrık koordinat noktasında bir hacmin var olduğunu (1) veya bir hacmin var olmadığını (0) ifade eden ikili (*binary*) bir sisteme karşılık gelir. Bu yaklaşım sayesinde sürekli uzaydaki ($\mathbb{R}^3$) sonsuz geometrik varyasyonların donanım üzerinde işlenebilecek bir hale getirilerek sonlu bir durum uzayına indirgenmesi sağlanır. Bu

topolojik indirgeme işlemi sonucunda mimarinin öğrenme durumuna doğrudan etki eden matris seyrekliği problemi oluşur. Literatür incelendiğinde 3D nesnelerin katı hacminin, kendisini çevreleyen sınırlayıcı kutunun (*bounding box*) yalnızca çok küçük bir yüzdesini (genelde %5 ile %10 arası) kapladığı görülür.

Mimarinin beslendiği verinin matematiksel kararlılığının sağlanması ve eğitim aşamasında ağırlık ölçek veya öteleme türlerine göre aşırı öğrenmeye kaymasının engellenmesi amacıyla ShapeNet veri setindeki nesneler öncelikle sınırlandırıcı kutu normalizasyonu işlemine tabi tutulmuştur. Bu işlem ile her bir nesnenin x, y ve z eksenlerindeki maksimum ve minimum uçları hesaplanmış ve nesnelerin ağırlık merkezinin doğrudan orijine kaydırılması sağlanmıştır. Buna ek olarak tüm nesneler önceden tanımlanmış standart birim hacim içerisine orantısal bütünlükleri bozulmadan yerleştirilmiştir. Yapılan bu işlemlerin sonucunda her bir nesnenin tam olarak 32x32x32 boyutlarında ve toplam olarak 32.768 bağımsız hacimsel birimden oluşan 3D tensörler olarak gösterimi sağlanır. Şekil 2.4'te veri setinden alınan 2D örnek görüntüler ve bunlara karşılık gelen 3D nesneler her bir sınıf için ayrı ayrı gösterilmiştir. Burada daha yüksek veya daha düşük bir çözünürlük seçilmesi mümkün olmakla birlikte seçilen bu değerin uygun aralıkta olduğu söylenebilir. Bu değerin daha da düşürülmesi nesnenin pek çok detayının kaybedilmesine, kenarlarının ve köşelerinin net olarak görülememesine sebep olmaktadır. Bu değerin daha da artırılması ise detayları artırmakla birlikte hesaplama yükünü kübik olarak artırmakta özellikle Grafik İşleme Birimi (*Graphical Processing Unit – GPU*) Video Rastgele Erişimli Belleğinin (*Video Random Access Memory – VRAM*) taşmasına sebep olabilmektedir. Şekil 2.5'te aynı aslan nesnesi düşük, orta ve yüksek çözünürlükte gösterilmiştir. Aradaki fark açık bir şekilde görülmektedir.



Şekil 2.4. 2D görüntüler ve karşılık gelen 3D nesneler



Şekil 2.5. Düşük, orta ve yüksek çözünürlüklü aslan nesnesi voksel temsilleri (Poly Haven, 2025)

2.2. Veri Ön İşleme ve Artırma

Büyük veri ve AI alanındaki çalışmalarda da sıklıkla vurgulandığı üzere, ML mimarilerinin başarısı ve güvenilirliği doğrudan kullanılan verinin kalitesi, çeşitliliği ve doğru işleme süreçleriyle ilişkilidir (Roh vd., 2021). Bu doğrultuda, DL mimarilerinde aşırı öğrenmenin önlenmesi ve genelleştirme yeteneğinin artırılması amacıyla rastgele kırpma, döndürme, renk sapması veya bölgesel maskeleme (DeVries ve Taylor, 2017; Shorten ve Khoshgoftaar, 2019) gibi standart veri artırma yöntemleri sıklıkla kullanılmaktadır. Ayrıca girdi görüntülerinin belirli alanlarının rastgele silinmesi yoluyla ağın nesne özneteliklerini ezberlemesini önleyen alternatif veri artırma stratejileri de literatürde yaygın kabul görmektedir (Zhong vd., 2020). Ancak bu yöntemler, girdi uzayındaki karmaşık karar sınırlarının yumuşatılması ve gürültüye dayanımının artırılmasında yetersiz kalabilmektedir. Bu nedenle, bu çalışmada geleneksel tekniklerin sınırlarını aşmak için önerilen mimarilere özel bir ön işleme ve veri artırma yöntemi izlenmiştir.

Önerilen AWIF kayıp fonksiyonunda izole bir şekilde başarıyı değerlendirmek ve kıyaslamak amaçlandığı için herhangi bir veri artırma uygulanmamıştır. Önerilen CPSA mekanizmasında ise detayları 2.12 bölümünde bahsedilen SA veri artırma tekniği kullanılmış ve detaylı bir ablasyon çalışması ile etkileri gösterilmiştir.

2.2.1. SWAGE-3D veri ön işleme ve veri artırma işlemleri

SWAGE-3D mimarisi, girdi olarak yalnızca RGB kanallı görüntüleri değil derinlik bilgisini de içeren 4 kanallı RGB-D tensörleri ile çalışacak şekilde geliştirilmiştir. Ön işleme aşamasında veri setindeki farklı çözünürlüklere sahip tüm 2D görüntüler ve derinlik haritaları

ResNet-18 mimarisi ile uyumlu olacak şekilde 224x224 boyutuna getirilmiştir. Gradyan akışının kararlılaştırılması amacıyla görüntülere standartlaştırma uygulanmıştır. 3D vokselize nesneler için ise bellek optimizasyonu ve mimari standartlar gereği 32x32x32 boyutları kullanılmıştır.

SWAGE-3D mimarisinin eğitim sürecinde geleneksel geometrik veri artırma yöntemleri yerine 2D *mixup* stratejisi kullanılmıştır (Zhang vd., 2018). Literatürde özellikle sınıflandırma problemlerinde etkili olan bu yöntem 3D yeniden yapılandırma için girdi gürültüsüne dayanımı artırmak amacıyla özel olarak uyarlanmıştır. Burada karıştırılan 2 görüntüden biri özellikle baskın hale getirilerek diğerinin anlamlı bir gürültü gibi davranması sağlanmıştır. Bu stratejinin matematiksel formülasyonu F. 1’de verilmiştir.

$$\tilde{X} = \lambda X_i + (1 - \lambda)X_j \quad (F. 1)$$

Burada:

$\tilde{X}$: Üretilen yeni 2D girdi tensörünü,

λ : Karışım oranını belirleyen katsayıyı,

X_i ve X_j : 2D veri setinden rastgele seçilen iki farklı girdi tensörünü ifade etmektedir.

Karışım katsayısı olan λ , $\alpha = 0,2$ seçimiyle uç değerlere yoğunlaşan bir Beta dağılımından ($\lambda \sim Beta(\alpha, \alpha)$) örneklenmektedir. Ağa aynı anda hem saf görüntülerin hem de karma görüntülerin gösterilmesinin hem VRAM darboğazına sebep olduğu hem de eğitim süresini artırdığı görülür. Buna bağlı olarak SWAGE-3D mimarisinde %50 ihtimalle bu teknik uygulanmıştır. Her bir yığın için bu ihtimal bağımsız olarak değerlendirilmektedir. Eğer bu teknik uygulanırsa ağ o yığında orijinal girdileri hiç görmez ve yalnızca tekniğin uygulandığı karma verilerle eğitilir. Aksi takdirde ise bu teknik uygulanmaz ve ağa o yığın için yalnızca orijinal veriler gösterilir.

2.3. CNN Mimarisi ve Temel Bileşenleri

Yüksek boyutlu verilerden hiyerarşik özniteliklerin çıkarılması ve karmaşık karar sınırlarının öğrenilmesi için literatürde temel bir standart olarak DL mimarileri (LeCun vd., 2015; Schmidhuber, 2015) ve özelinde CNN’ler kullanılmaktadır. Temelleri Neocognitron mimarisine (Fukushima, 1980) dayanan ve ilk modern mimari yapısı LeNet-5 (LeCun vd., 1998) ile ortaya konan bu ağlar, özellikle büyük ölçekli görsel veri setlerinde derin mimarilerin gücünü kanıtlayan AlexNet (Krizhevsky vd., 2012) çalışmasıyla günümüzdeki popüleritesine kavuşmuştur. Bu mimariler; bölgesel özniteliklerin tespitini sağlayan evrişim (*convolution*), ağa eğrisellik katan aktivasyon, uzamsal boyutu düşürerek hesaplama yükünü azaltan havuzlama

(*pooling*), gradyan akışını kararlılaştıran normalizasyon ve seyrelterek (*dropout*) (Srivastava vd., 2014) aşırı öğrenmeyi engelleyen düzenleme katmanlarından oluşur. Bu çalışmada önerilen 3D-VAE-GAN, AWIF, CPSA ve SWAGE-3D mimarilerinin kodlayıcı, üretici ve ayırt edici ağ omurgalarında bu temel CNN bileşenleri amaca uygun olarak özelleştirilmiş ve hacimsel veri üretimi probleminin doğasına göre adapte edilmiştir.

Evrişim işlemi, klasik IP’de kullanılan 3x3 boyutlu temel kenar bulma filtrelerinin çalışma mantığına benzer şekilde (Sobel ve Feldman, 1968); girdi matrisi üzerinde daha küçük boyutlu bir filtre matrisinin kaydırılması ve örtüşen elemanların çarpılıp toplanması ile bölgesel özniteliklerin çıkarılmasını sağlayan temel bir operatördür (Goodfellow vd., 2016). Bu işlemin amacı, pikselleri birbirinden bağımsız incelemek yerine aralarındaki uzamsal ilişkilerin korunmasıdır. Matematiksel olarak 2 boyutlu ayırık evrişim işlemi F. 2’de gösterildiği gibi tanımlanmaktadır.

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(i + m, j + n)K(m, n) \quad (F. 2)$$

Burada:

S : hesaplanmak istenen skaler sonucu,

I : Girdi matrisini,

K : Filtre matrisini,

i, j : Filtrenin girdi matrisi üzerinde o an konumlandığı referans noktasının koordinatlarını,

m, n : Filtre içindeki yerel gezinme değişkenlerini ifade etmektedir.

Bir evrişim katmanının uzamsal çözünürlüğü ve öğrenme kapasitesi temel olarak; ağırın algı alanını (*receptive field*) tanımlayan filtre boyutu, filtrenin kaydırma miktarını belirleyen adım büyüklüğü (*stride*) ve kenar bilgilerinin kaybolmasını önleyen dolgu (*padding*) hiperparametreleri ile kontrol edilir. Ayrıca, geleneksel tam bağılı mimarilerin aksine, evrişim işleminde filtrenin ağırlık matrisi tüm uzamsal alan boyunca sabit kalarak tekrar tekrar kullanılır. Ağırlık paylaşımı adı verilen bu mekanizma, öğrenilmesi gereken parametre sayısını üstel olarak azaltırken mimarinin öteleme değişmezliği kazanarak öznitelikleri konumdan bağımsız öğrenmesini sağlar.

DL mimarilerinde katman sayısı arttıkça verilerin istatistiksel dağılımının kontrolsüzce değişmesi, kaybolan veya patlayan gradyan problemlerine yol açmaktadır. Optimizasyon

yüzeyinin pürüzsüzleştirilmesi ve gradyan akışının kararlılaştırılması için normalizasyon teknikleri kritik bir role sahiptir. Bu tez çalışmasında geliştirilen mimarilerin kodlayıcı ve üretici ağlarında genel standart olarak Yığın Normalizasyonu (*Batch Normalization* – BN) kullanılmıştır (Ioffe ve Szegedy, 2015). BN tekniği, her kanal için istatistikleri yığın üzerinden hesaplayarak veriyi normalize eder ve ardından öğrenilebilir ölçek (γ) ile kaydırma (β) parametreleri aracılığıyla ağın doğrusal olmayan temsil yeteneğinin korumasını sağlar.

Öte yandan, önerilen SWAGE-3D mimarisindeki ayırt edici (eleştirmen) ağda, ağın 1-Lipschitz sürekli olmasını garanti altına almak amacıyla SN (Miyato vd., 2018) uygulanmıştır. SN, evrişim ve tam bağlı katmanların ağırlık matrislerini kendi spektral normuna (en büyük tekil değer) bölerek gradyanların matematiksel olarak 1'i aşmasını doğrudan engeller. SN uygulandığında ağırlık matrisinin normalize edilmiş hali F. 3'te verilmiştir.

$$w_{SN} = \frac{w}{\sigma_{SN}(w)} \quad (F. 3)$$

Burada:

$\sigma_{SN}(w)$: w ağırlık matrisinin spektral normunu ifade etmektedir.

Yüksek boyutlu matrislerde Tekil Değer Ayrışımı (*Singular Value Decomposition* – SVD) yapmak hesaplama maliyetini artıracak için ağırlık matrisinin spektral normu eğitim boyunca kuvvet iterasyonu yöntemi kullanılarak çok düşük bir matematiksel maliyet ile yaklaşık olarak hesaplanır.

Yapay Sinir Ağlarında (*Artificial Neural Network* – ANN) doğrusal olmayan karar sınırlarının öğrenilmesi aktivasyon fonksiyonları ile sağlanır. Bu tez çalışmasında geliştirilen mimarilerin farklı bileşenlerinde, ilgili katmanın görevine uygun olarak çeşitli aktivasyon fonksiyonları tercih edilmiştir. Kodlayıcı ve üretici ağların gizli katmanlarında, hesaplama maliyeti düşük olan ve pozitif eksende kaybolan gradyan problemini çözen ReLU (*Rectified Linear Unit*) fonksiyonu standart olarak kullanılmıştır (Nair ve Hinton, 2010). Çekişmeli ağların ayırt edici (eleştirmen) bileşenlerinde ise, gerçek-sahte ayrımının yapılabilmesi için verideki negatif gradyan bilgilerine ihtiyaç duyulduğundan, negatif değerleri tamamen sıfırlamayan Leaky ReLU ($\alpha = 0,2$) fonksiyonu tercih edilmiştir (Maas vd., 2013). Ayrıca, mimarilerin çıktı katmanlarında (özellikle voksel doluluk oranlarının belirlenmesinde), elde edilen değerlerin 0 ile 1 arasında bir olasılık dağılımına dönüştürülmesi amacıyla Sigmoid aktivasyon fonksiyonu kullanılmıştır (Goodfellow vd., 2016).

Geleneksel ayırmacı CNN mimarilerinde uzamsal boyutu düşürmek, hesaplama maliyetini

azaltmak ve öteleme değişmezliği sağlamak amacıyla maksimum veya ortalama havuzlama işlemleri standart olarak kullanılmaktadır. Ancak, 3D nesne üretimi gibi piksellerin ve voksellerin uzamsal koordinat bilgilerinin kritik öneme sahip olduğu üretici mimarilerde, havuzlama işlemi uzamsal bilgiyi geri döndürülemez şekilde yok ettiği için tercih edilmemektedir. Bu nedenle, bu çalışmada sunulan 3D-VAE-GAN ve SWAGE-3D mimarilerinde boyut küçültme işlemleri havuzlama yerine öğrenilebilir parametrelere sahip adımli evrişim (*strided convolution*) ile; boyut büyütme işlemleri ise uzamsal boyutları ve detayları geri kazandıran devrik evrişim (*transposed convolution*) katmanları ile gerçekleştirilmiştir.

Evrişimli katmanlardan elde edilen hiyerarşik öznitelik haritaları, ağıın karar verme veya öznitelik sıkıştırma aşamasına geçebilmesi için düzleştirme (*flattening*) işlemi ile tek boyutlu bir öznitelik vektörüne dönüştürülür. Bu aşamadan sonra kullanılan tam bağı (*fully connected*) katmanlar, tüm girdi birimlerinin her bir çıktı birimine bağlandığı doğrusal bir yapı sunarak evrişimsel öznitelikler arasındaki küresel ilişkileri öğrenir (Goodfellow vd., 2016). Bu çalışmada önerilen mimarilerde tam bağı katmanlar; 2D kodlayıcı ağıından elde edilen özniteliklerin VAE gizli uzayındaki ortalama (μ) ve varyans (σ) parametrelerine haritalanmasında ve çekişmeli öğrenme sürecinde ayırt edici ağıın nihai gerçeklik skorunu üretmesinde temel bir rol oynamaktadır

2.4. ResNet-18 Mimarisi

ResNet-18 mimarisinin artık öğrenme (*residual learning*) kullanan ve 18 katmanlı bir yapıya sahip olduğu görülür (He vd., 2016). ResNet-18 ile klasik CNN mimarisinde katman sayısı arttıkça meydana gelen bozulma problemine bir çözüm sunulur. Bu problem eğitim hatasının önce doyunluğa ulaşması sonra ise aniden hızla artması anlamına gelir. Bu duruma başlangıçta aşırı öğrenme probleminin sebep olduğu düşünülse de aslında bununla alakası olmayan özünde katmanlı mimarilerdeki gradyan patlaması ve kaybolan gradyan problemlerinin sebep olduğu görülür. ResNet mimarileri ile temel olarak bu sorunun aşılması hedeflenir.

Geleneksel ağılarda birbirlerine ardışık olarak bağlanan katmanlar kullanılarak ağıın dönüşüm fonksiyonunu öğrenmesi amaçlanır. ResNet mimarisinde ise bu yaklaşım yerine katmanların bir asıl fonksiyonu değil bir artık fonksiyonu öğrenmesi hedeflenir. Buradaki matematiksel işlem F. 4'te gösterildiği şekilde kurulur.

$$H(x) = F(x) + x \quad (F. 4)$$

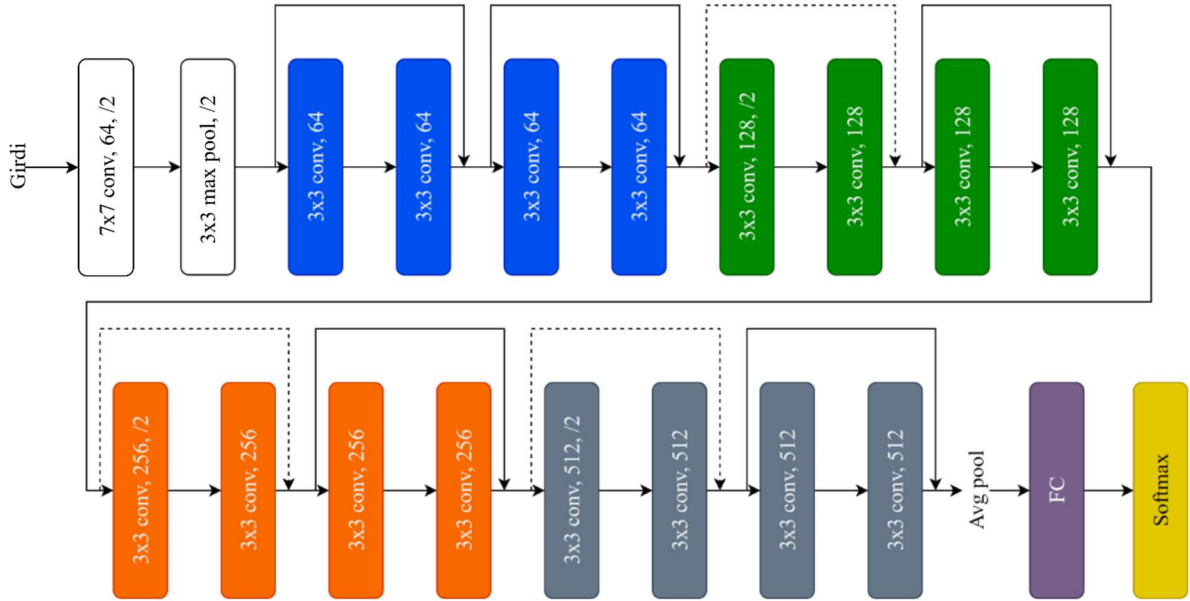
Burada:

x : İlgili bloğa giren orijinal girdi tensörünü,

$F(x)$: Bloğun içindeki evrişim katmanları tarafından öğrenilen artık eşleşmeyi,

$H(x)$: Bloğun ürettiği nihai çıktıyı ifade etmektedir.

Buradaki en iyi eşleme bir birim fonksiyon ise sinir ağının doğrusal olmayan katmanları vasıtası ile bunun sıfırdan öğrenilmesinin çok zor olduğu görülür. Ancak artık öğrenme yaklaşımında ağın tek yapması gereken $F(x)$ değerinin sıfıra doğru sürüklenmesidir. Böylece $H(x) = x$ olur ve önceki katmanlarda öğrenilmiş bilgiler kayıpsız olarak ileri aktarılır. Buna literatürde kısa yol bağlantısı (*skip connections*) adı verilir. Şekil 2.6’da ResNet-18 mimarisi ve her bir katmanı detaylı olarak gösterilmiştir. Şekildeki düz ve kesikli oklar ile kısa yol bağlantıları gösterilmiştir. Düz çizgiler ağ içindeki bilgi akışının hiçbir yapısal değişime uğramadan doğrudan ileriye taşındığını ifade ederken, kesikli çizgiler ise tensör boyutlarında matematiksel bir uyumsuzluk olduğunu ve bunu çözmek için adım değeri 2 olan 1×1 boyutunda evrişim uygulandığını gösterir.



Şekil 2.6. ResNet-18 mimarisi ve katmanları

2.5. AE Mimarisi ve Devrik Evrişim Temelleri

AE mimarisi giriş verilerini daha düşük boyutlu bir gizli uzay temsiline sıkıştıran bir kodlayıcı ve bu temsilden veriyi yeniden inşa eden bir kod çözücü ağından oluşmaktadır (Hinton ve Salakhutdinov, 2006). Literatürde boyutsallığı azaltmak için yaygın olarak kullanılmalarına karşın, tek 2D görüntüden 3D nesne üretimi bağlamında bu yapı, hiyerarşik bir öznetelik dönüşümü rolü üstlenir. Nitekim kodlayıcı ağların, görsel verilerdeki eksik veya görünmeyen kısımları tahmin etmeye zorlanarak çok güçlü bir anlamsal öznetelik öğrenme kapasitesine ulaştığı literatürde de kanıtlanmıştır (Pathak vd., 2016). Özellikle voksel tabanlı hacimsel şekil

üretimi çalışmalarında, genelleştirilmiş otomatik kodlayıcı mimarilerinin 3D nesne üretiminde etkili sonuçlar verdiği görülmektedir (Guan vd., 2020). Bu tez çalışmasında kullanılan mimarilerin temelini oluşturan bu düzende; kodlayıcı ağ 2D görüntülerdeki yapısal bilgileri darboğaz katmanında sıkıştırılmış bir vektör haline getirir. Kod çözücü ağ ise bu kompakt gizli vektörden yararlanarak hedef 3D voksel topolojisini sentezler.

Kod çözücü aşamasında, gizli uzaydaki düşük boyutlu verinin $32 \times 32 \times 32$ çözünürlüğünde 3D hacimsel bir yapıya genişletilebilmesi için devrik evrişim katmanlarından yararlanır. Devrik evrişim işlemi, standart evrişimin aksine girdi tensörünün boyutlarını hiyerarşik olarak artırarak mimarinin 2D izdüşümlerden 3D nesne koordinatlarını öğrenmesine olanak tanır (Dumoulin ve Visin, 2016). Böylelikle AE, 2D girdi uzayı ile 3D çıktı uzayı arasındaki geometrik geçişin temelini oluşturmaktadır.

2.6. VAE Mimarisi

VAE'ler standart AE mimarisinin deterministik doğasını olasılıksal bir düzleme taşıyarak veri üretim kapasitesini artıran güçlü üretici mimarilerdir (Kingma ve Welling, 2014). Tek 2D görüntüden 3D nesne üretimi probleminde VAE, gizli uzayın belirli bir dağılıma uymasını zorunlu kılarak mimarinin sadece eğitim verilerini ezberlemesini önlemektedir. Bu mimaride kodlayıcı ağ, giriş verisini doğrudan tek noktaya eşlemek yerine, gizli uzayda bir olasılık dağılımının parametrelerini (ortalama ve varyans) öğrenmektedir.

3D nesne inşası sürecinde VAE kullanımı, gizli uzayın sürekli ve düzenli bir yapıda tutulmasını sağlamaktadır. Eğitimde kullanılan yeniden yapılandırma kaybı (*reconstruction loss*) mimarinin girdi görüntüsüne bağlı kalmasını sağlarken, Kullback-Leibler (KL) ıraksama (Kullback ve Leibler, 1951) bileşeni ise gizli uzayı standart normal dağılıma yaklaştırarak geometrik temsiller arasında anlamlı geçişler yapılmasına olanak tanır. Bu tez çalışmasında VAE, karmaşık 3D nesnelerin düşük boyutlu ve sürekli bir temsilini öğrenmek amacıyla kullanılmıştır. Yeniden parametrelendirme hilesi (*reparameterization trick*) sayesinde gradyan akışı korunarak eğitilen bu yapı ile, gizli uzaydan örnekleme yapılarak daha önce görülmemiş ancak yapısal bütünlüğe sahip 3D voksel çıktıları üretilebilmektedir.

2.7. GAN Mimarisi

GAN mimarisi veri dağılımını öğrenerek gerçeğe en yakın yeni verileri üretmeyi amaçlayan, bir üretici ve bir ayırt edici ağı karşılıklı optimizasyonuna dayanan DL mimarisi olarak ifade edilir (Goodfellow vd., 2014). Başlangıçta Çok Katmanlı Algılayıcılarla (*Multilayer Perceptron* – MLP) tasarlanan bu temel yapının, CNN ile entegre edilerek görsel veri üretiminde kararlı ve ölçeklenebilir hale getirilmesi de sağlanmıştır (Radford vd., 2015). Ayrıca GAN

mimarilerinin belirli bir girdi verisine koşullandırılarak yönlendirilmesi, hedef odaklı üretim başarımını önemli ölçüde artırmaktadır (Mirza ve Osindero, 2014; Isola vd., 2017). Birebir eşleşen eğitim verilerinin bulunmadığı senaryolarda döngüsel tutarlılık kaybı kullanarak iki farklı alan arasında veri dağılımı transferini başlatan yaklaşımlar da (Zhu vd., 2017) üretici ağların esnekliğini ve genelleme kapasitesini önemli ölçüde genişletmiştir. Bununla birlikte, üretici ve ayırt edici ağlar arasındaki çekişmeli eğitim sürecinin doğası gereği barındırdığı kararsızlık ve mod çökmesi gibi temel sorunların çözümü için stabilizasyon ve optimizasyon teknikleri önerilmektedir (Salimans vd., 2016). Geleneksel GAN eğitimlerinde karşılaşılan kararsızlıkları gidermek ve daha yüksek çözünürlüklü sentezleme yapabilmek adına, mimarinin kademeli olarak büyütüldüğü ilerici eğitim stratejileri (Karras vd., 2018) ve bu stratejilerin üzerine inşa edilerek üretim sürecinde detayların kontrol edilebilirliğini sağlayan gelişmiş ağ mimarileri (Karras vd., 2019), yüksek kaliteli sentezleme başarılarıyla literatürde önemli bir yer tutmaktadır. Tek 2D görüntüden 3D nesne üretimi bağlamında GAN mimarisi, salt VAE tabanlı yaklaşımların ürettiği bulanık ve detaysız voksel çıktılarını keskinleştirmek ve geometrik gerçekçiliği artırmak için kritik bir rol üstlenmektedir (Wu vd., 2016).

Bu tez çalışmasında temel alınan 3D-VAE-GAN mimarisinde; kodlayıcı tarafından sıkıştırılan ve VAE aracılığıyla olasılıksal bir dağılıma oturtulan gizli uzay vektörü, üretici tarafından hedeflenen 3D voksel topolojisine dönüştürülür. Ayırt edici ağ ise, üretici tarafından sentezlenen bu 3D voksel yapılarının gerçek veri setindeki 3D nesnelere ne kadar benzediğini değerlendirmektedir. Eğitim süreci boyunca üretici, ayırt ediciyi yanıltabilecek kadar gerçekçi ve yüzey detaylarına sahip 3D nesneler sentezlemeyi öğrenirken; ayırt edici de üretilen sentetik 3D yapılar ile gerçek olanları ayırt etme yeteneğini geliştirmektedir.

Bu karşılıklı optimizasyon sürecinde ayırt edici ağ optimum başarıma ulaştığında, üretici ağın temel hedefi gerçek veri dağılımı ile üretilen sentetik veri dağılımı arasındaki Jensen-Shannon Iraksamasını sıfıra yaklaştırmaktır. Olasılık dağılımları arasındaki bu sapmayı ölçen Jensen-Shannon Iraksama metriği, kuramsal temelini bilgi teorisindeki Shannon entropisinden almaktadır (Lin, 1991).

2.8. WGAN-GP ve SN

GAN mimarilerinin özellikle yüksek boyutlu 3D voksel uzayında eğitilmesi sürecinde mod çökmesi ve gradyan kararsızlığı gibi kronik problemlerle karşılaşılmaktadır. Çekişmeli ağların kararlılığını sağlamak amacıyla kullanılan Wasserstein uzaklığı matematiksel temelini, olasılık dağılımları arasındaki geometrik mesafeyi optimize eden optimal taşıma teorisinden almaktadır (Villani, 2009). Bu tez çalışmasında, söz konusu eğitim kararsızlıklarını gidermek ve

Lipschitz kısıtını daha etkin bir şekilde sağlamak amacıyla standart GAN mimarisi yerine WGAN (Arjovsky vd., 2017) yaklaşımı ve GP bileşeni birlikte kullanılmıştır (Gulrajani vd., 2017).

WGAN-GP, ayırt edici (eleştirmen) ağı ağırlıklarını kırmak (*clipping*) yerine gradyanların normuna yönelik bir ceza terimi ekleyerek eğitimin daha geniş bir hiperparametre aralığında kararlı kalmasına olanak tanır. SWAGE-3D mimarisinde bu yapıya ek olarak, her katmandaki ağırlık matrislerinin spektral normunu sınırlayan spektral normalizasyon tekniği de entegre edilmiştir (Miyato vd., 2018). Bu bütünleşik stabilizasyon yaklaşımı, çekişmeli öğrenme sürecini dengeleyerek üretilen 3D nesnelerin geometrik tutarlılığını ve yüzey detaylarının korunmasını sağlamaktadır.

2.9. 3D-VAE-GAN Mimarisi

3D-VAE-GAN mimarisi 2D görüntülerden 3D nesneler üretmeyi amaçlayan ve bu tez çalışmasının omurgasını oluşturan temel referans mimaridir (Wu vd., 2016; Brock vd., 2016). Bu mimari, VAE'nin gizli uzaya eşleme yeteneği ile GAN'ın rekabetçi üretim gücünü birlikte kullanan ve bu iki yaklaşımı olasılıksal bir düzlemde bütünleştiren (Mescheder vd., 2017) kodlayıcı, üretici ve ayırt edici olmak üzere üç temel sinir ağından oluşmaktadır.

Mimarinin veri akışı incelendiğinde, sürecin 4 kanallı (RGBA) ve 137x137 çözünürlüğündeki 2D ön işlemden geçmiş görüntülerin kodlayıcı ağı girdi olarak verilmesiyle başladığı görülmektedir. Beş katmandan oluşan bu ağda tensörler yığın boyutu 32 olmak üzere [32,4,137,137] boyutlarında işleme alınmaktadır. Veriler sırasıyla 2D evrişim (Conv2D), 2D BN (BatchNorm2D) ve ReLU aktivasyon fonksiyonundan geçirilmektedir. Örneğin birinci katmanda uygulanan Conv2D (4,64,5,2,2) şeklindeki işlem 4 girdi kanalı, 5x5 uzamsal boyutundaki filtreler, 2 adım büyüklüğü ve 2 dolgu hiperparametreleri kullanılarak 64 adet çıktı kanalına (öznitelik haritasına) dönüştürüldüğünü ifade etmektedir. Beş evrişim katmanının ardından elde edilen uzamsal veriler düzleştirme işlemiyle tek boyutlu bir vektöre dönüştürülmekte ve ardından paralel olarak çalışan iki ayrı tam bağlı katmana aktarılmaktadır. FC1 ve FC2 olarak isimlendirilen bu katmanlar ile verinin gizli uzaydaki olasılıksal dağılımını tanımlayan ortalama (μ) ve varyans (σ^2) vektörleri hesaplanmaktadır. 2.6 bölümünde detaylandırılan yeniden parametrelendirme hilesi kullanılarak bu değerlerden 200 boyutlu ($z = 200$) gizli uzay vektörü elde edilmektedir.

Üretici ağı, kodlayıcıdan gelen bu 200 boyutlu z vektörünü girdi olarak alıp 3D nesneleri üretmekle görevlidir. Beş katmandan oluşan bu ağda veriler standart evrişimin aksine veriyi boyutsal olarak genişletme sağlayan 3D devrik evrişim (ConvTranspose3D), 3D BN ve ReLU

aktivasyon fonksiyonundan geçirilmektedir. Üretici ağı son katmanında sigmoid aktivasyon fonksiyonu kullanılarak ağ çıktılarının 0 ile 1 arasında olasılık değerlerine dönüştürülmesi sağlanmıştır. Bu değerin 1'e yaklaşması ile ilgili koordinata vksel ekleneceği, 0'a yaklaşması ile ise o koordinatın boş bırakılacağı belirlenmektedir. Bu doluluk oranına göre sahte 3D nesneler üretilmektedir. Ayrıca mimarinin üretkenlik kapasitesinin artırılması amacıyla, doğrudan standart normal dağılımdan rastgele örneklenen z vektörleri kullanılarak da veri setinde bulunmayan yeni 3D nesneler sentezlenebilmektedir.

Ayırt edici ağ ise girdi olarak veri setindeki $32 \times 32 \times 32$ boyutlu orijinal (gerçek) vokseller ile üreticinin sentezlediği aynı boyuttaki sahte vokselleri almaktadır. Beş katmanlı bu ağda 3D evrişim (Conv3D) ve 3D yığın normalleştirme işlemlerinin ardından aktivasyon fonksiyonu olarak ReLU yerine LeakyReLU (negatif eğim katsayısı $\alpha = 0,2$ ile) kullanılmıştır. Bunun temel mimari sebebi ayırt edici ağı gerçek-sahte ayrımını yapabilmesi için verideki negatif gradyan bilgilerine de ihtiyaç duyulması ve standart ReLU fonksiyonunun bu negatif değerleri sıfırlayarak bilgi kaybına yol açmasının kasıtlı olarak engellenmesidir. Tüm bu ileri yönlü ağ hesaplamaları tamamlandığında elde edilen çıktılar üzerinden ilgili kayıp fonksiyonları hesaplanmakta ve elde edilen hata değerleri geri yayılım algoritması (Rumelhart vd., 1986) ile ağlara iletilerek eğitim süreci devam ettirilmektedir.

Mimarinin uçtan uca eğitim süreci kodlayıcı, üretici ve ayırt edici ağlar arasında dönüşümlü bir optimizasyon prosedürü işletmektedir. Mimarinin istikrarlı bir şekilde yakınsaması için öncelikle ayırt edici ağ ardından ise kodlayıcı ve üretici ağlar güncellenmektedir. Mimarinin her bir bileşeni için tanımlanan kayıp fonksiyonları F. 5, F. 6 ve F. 7'de verilmiştir.

$$\mathcal{L}_{Kodlayici} = \mathcal{L}_{recon} + \mathcal{L}_{KL} \quad (F. 5)$$

$$\mathcal{L}_{Uretici} = \mathcal{L}_{adv} + \mathcal{L}_{recon} \quad (F. 6)$$

$$\mathcal{L}_{AyirtEdici} = \mathcal{L}_{gercek} + \mathcal{L}_{sahte} \quad (F. 7)$$

Burada:

$\mathcal{L}_{recon}$: Yeniden yapılandırma kaybını,

$\mathcal{L}_{KL}$: KL ıraksamasını,

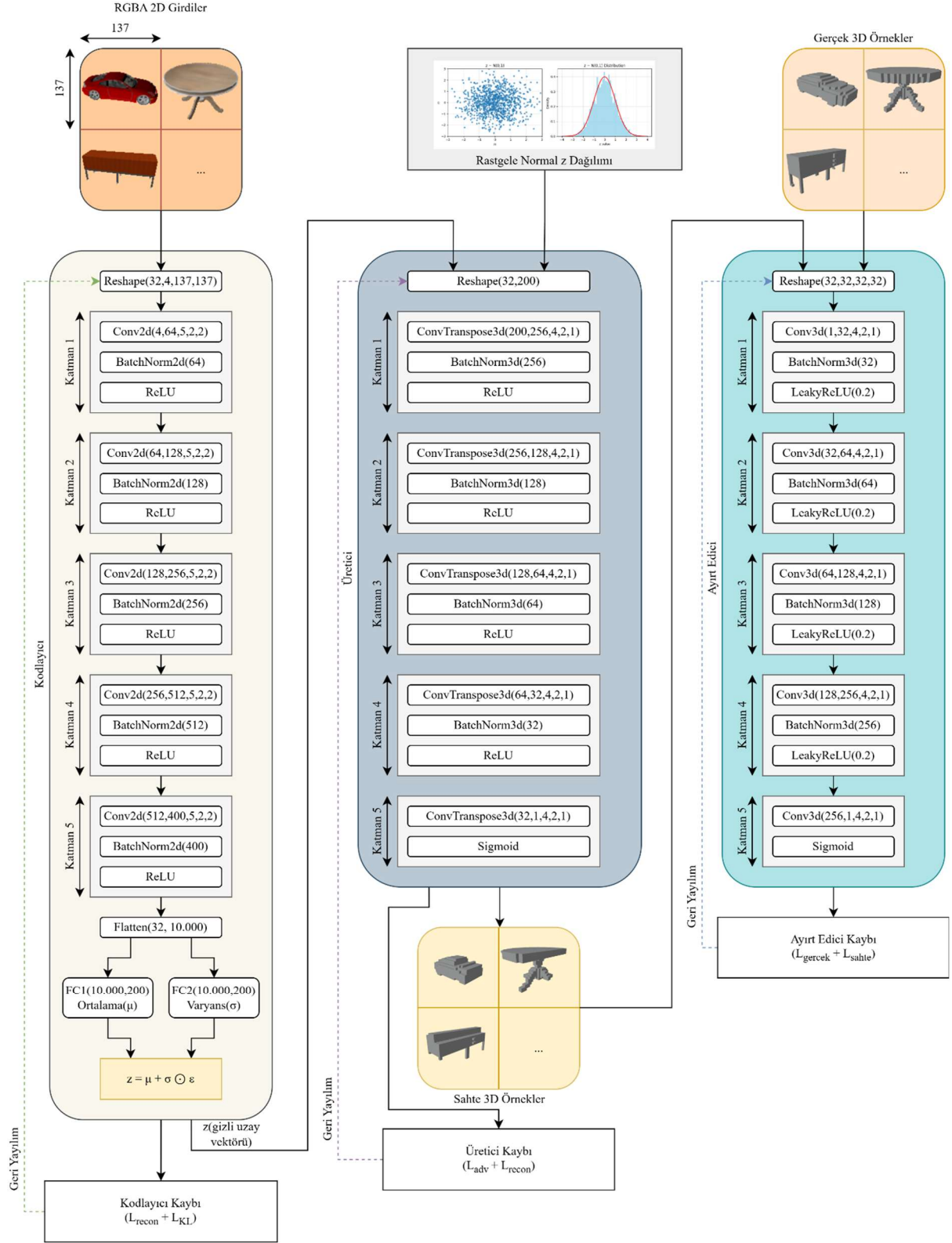
$\mathcal{L}_{adv}$: Üreticinin ayırt ediciyi yanıltma kaybını,

$\mathcal{L}_{gercek}$: Ayırt edici ağı gerçek örnekleri sınıflandırma hatasını,

$\mathcal{L}_{sahte}$: Ayırt edici ağı sahte örnekleri sınıflandırma hatasını ifade eder.

Bu formülasyonlarda verilen kayıp fonksiyonlarının açık halleri 2.6 ve 2.7 bölümlerinde verilmiştir. F. 6’da verilen $\mathcal{L}_{adv}$ terimi 2.7 bölümünde anlatılan üretici ağ kaybına ($\mathcal{L}_G$) karşılık gelmektedir. Ancak 3D-VAE-GAN mimarisinde üretici ağ aynı zamanda bir kod çözücü görevi üstlendiğinden ağın toplam hedefine ($\mathcal{L}_{Uretici}$) yapısal benzerliği sağlamak amacıyla yeniden yapılandırma kaybı ($\mathcal{L}_{recon}$) da entegre edilmiştir.

Şekil 2.7’de temel olarak 3D-VAE-GAN mimarisinin katman akışları ve bütünleşik öbek şeması sunulmuştur. Çalışmamızda uzamsal boyutları vurgulamak amacıyla tensörler 4 ve 5 boyutlu (yığın boyutu ve kanal bilgisi dahil olarak) olmalarına rağmen geleneksel literatürdeki 2D görüntü ve 3D nesne terminolojisi kullanılmıştır.



Şekil 2.7. Temel 3D-VAE-GAN mimarisi

2.10. Yeniden Yapılandırma Kayıp Fonksiyonları

DL mimarilerinde kayıp fonksiyonu seçiminin yakınsama hızı, optimizasyon hızı ve nihai başarımı üzerinde belirleyici bir role sahip olduğu görülür. DL eğitimindeki temel amaç mimarinin yaptığı tahminler ile gerçek (*ground-truth*) değerler arasındaki sapmanın

matematiksel olarak ölçülmesi ve bu sapmanın uygun bir optimizatör yardımıyla mümkün olan en küçük değere indirilmesidir (Chollet, 2018). Bu tez çalışmasında, 3D-VAE-GAN mimarisinin kodlayıcı ve üretici ağlarının kayıp fonksiyonlarında sıklıkla kullanılan standart yeniden yapılandırma kayıp fonksiyonlarının başarımları karşılaştırmalı olarak analiz edilmiştir. Buna ek olarak hacimsel verilerin doğasında bulunan yüzey kayıpları, geometrik tutarsızlık ve köşe detaylarının kaybı gibi problemlerin çözülmesi amacıyla yeni bir karma yeniden yapılandırma kayıp fonksiyonu olan AWIF literatüre kazandırılmış ve tezin temel katkısı olarak sunulmuştur.

2.10.1. MSE kaybı

MSE'nin ML mimarilerindeki temel kayıp fonksiyonlarından biri olduğu görülür (Bishop, 2006; Mitchell, 1997). Bu fonksiyon ile vokselize nesne yapılandırması kısmında mimarinin tahmin ettiği değerler ile gerçek değerler arasındaki farkın karesi alınarak tüm voksel konumları için toplam hata hesaplanmaktadır. Burada karesel bir işlem yapılmasının temel amacı negatif değerlerin tamamen ortadan kaldırılması ve büyük hataların orantısız olarak daha fazla cezalandırılmasıdır. 3D vokselize nesne yapılandırması bağlamında MSE kayıp fonksiyonunun formülü F. 8'deki gibi hesaplanır.

$$\mathcal{L}_{MSE} = \sum_{i=1}^{N_T} (y_i - \hat{y}_i)^2 \quad (\text{F. 8})$$

Burada:

$\hat{y}_i$: Tahmin edilen i . voksel değerini,

y_i : Veri setinde bulunan gerçek i . voksel değerini,

N_T : Toplam voksel sayısını ifade etmektedir.

MSE kendi başına ifade edildiğinde konveks bir yapıya sahip olduğu görülür, bu sayede tek küresel minimuma yakınsar ve matematiksel olarak kolayca türevlenebilir. Buna rağmen MSE'nin içinde aktivasyon fonksiyonları bulunan katmanlı bir sinir ağı mimarisinde kullanılması ağı ağırlıklarının kayıp yüzeyini bu konveks yapıdan çıkarır. MSE hacimsel optimizasyonda kararlı bir altyapı sunabilir. Ancak seyrek (*sparse*) 3D yapılar üzerindeki uç değerlere (*outliers*) karşı yüksek hassasiyet göstermesi ve 0 ve 1 olarak ifade edilen ikili voksel değerleri için optimum başarımları sağlayamaması MSE'nin yöntemsel sınırları olarak ifade edilebilir.

2.10.2. BCE kaybı

BCE'nin gerçek ve tahmin edilen olarak ifade edilen iki olasılık dağılımı arasındaki

mesafenin ölçüldüğü ve özellikle ikili sınıflandırma problemlerinde kullanılan standart bir kayıp fonksiyonu olduğu görülür. 3D vokselleştirme yapılandırması da temel olarak bu duruma doğrudan uymaktadır. Çünkü bir uzay koordinatında bir vokselin dolu olması 1, boş olması ise 0 değeri ile ifade edilir.

Seyrek 3D yapılarda mimarinin hatasını yüksek kayıp değerleriyle cezalandıran BCE fonksiyonu F. 9’da sunulmuştur. Üretici ağın son katmanında kullanılan sigmoid aktivasyon fonksiyonu ile çıktılar 0 ile 1 arasına sıkıştırılsa da farklı evrişim katmanlarından gelebilecek hatalı gradyan akışlarının hesaplama kararsızlığına yol açmaması için tahmin edilen değerler $[10^{-7}, 1 - 10^{-7}]$ aralığına kısıpılmıştır (*clamp*). Eğer üretici ağın çıktısı ($\hat{y}_i$) tam olarak 0 olursa F. 9’un ilk bileşeni $\log(0) = -\infty$ ve eğer bu çıktı tam olarak 1 olursa da aynı formülün ikinci bileşeni $\log(1 - 1) = \log(0) = -\infty$ olur. Bilgisayar bilimlerinde sonsuz değerler ile matematiksel işlem gerçekleştirilemeyeceği için kısıpama işlemi uygulanmıştır.

$$\mathcal{L}_{BCE} = - \sum_{i=1}^{N_T} [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (F. 9)$$

2.10.3. Focal kaybı

Focal, BCE altyapısı üzerine inşa edilen ve veri setindeki şiddetli sınıf dengesizliklerini çözmeyi amaçlayan gelişmiş bir kayıp fonksiyonu olarak ifade edilir (Lin vd., 2017b). Standart kayıp fonksiyonları veri setindeki sayıca fazla ama tespiti “kolay” olan örnekler odaklanarak toplam kaybı hızlı bir şekilde düşürme eğilimindedir. Örneğin hastalık teşhis sistemlerinde pozitif örneklerin ve 3D vokselize nesnelerde “boş” olan voksellerin tahmin edilmesinin diğer durumlara kıyasla daha kolay olduğu görülür. Bu durumda toplam başarı yapay olarak yüksek gösterilebilir. Focal kayıp odaklanma parametresi (γ_{FL}) ve dengeleme faktörü (α_{FL}) adı verilen bileşenler aracılığıyla ağır ve nadir örnekler odaklanılması zorunlu kılınır. Bu kayıpta ağır kolayca doğru bir şekilde tahmin ettiği örnekler $(1 - p_i)^{\gamma_{FL}}$ modülatörü ile sıfıra yaklaştırılırken, mimarinin yanıldığı veya emin olamadığı zor örnekler ise kayıp değeri yüksek tutularak gradyan akış yönü değiştirilmektedir. Dolu voksellerin boş voksellere oranla çok daha az olduğu 3D nesne üretimi problemlerinde Focal kayıp F. 10 ve F. 11’de verildiği şekilde hesaplanır.

$$h_i = y_i \hat{y}_i + (1 - y_i)(1 - \hat{y}_i) \quad (F. 10)$$

$$\mathcal{L}_{Focal} = - \sum_{i=1}^{N_T} \alpha_{FL,i} (1 - h_i)^{\gamma_{FL}} \log(h_i) \quad (F. 11)$$

Burada:

h_i : i. vokselle için mimarinin doğru tahmin olasılığını,

$\alpha_{FL,i}$: i. voksele ait dengeleme faktörünü,

γ_{FL} : Odaklanma parametresini ifade etmektedir.

2.10.4. MFL kaybı

MFL, standart Focal kayıp formülasyonunda yer alan sınıfa özgü dengeleme parametresinin elimine edildiği versiyon olarak ifade edilir. Sınıf dağılımlarının uzaysal olarak değişkenlik gösterdiği veya statik bir ön ağırlıklandırmanın mimarinin genelleme kapasitesini kısıtladığı senaryolarda uygulanır. Bu yapı, optimizasyon sürecini tamamen mimarinin anlık tahmin güvenilirliğini dinamik olarak ölçen modülasyon faktörüne dayandırır.

Sınıf dengeleme parametresinin denklemden çıkarılması hiperparametre arama uzayını daraltırken, gradyan ceza mekanizmasının doğrudan ağırlık hatalı veya düşük güvenle tahmin ettiği örneklerle yönelmesini zorunlu kılar. 3D vokselle üretim problemlerinde hacmin büyük kısmını oluşturan “boş” voksellerin yarattığı gradyan baskınlığı, yalnızca ağırlık zorlandığı “dolmuş” voksel sınırlarında üretilen yüksek hata sinyalleri ile bastırılır. Mimarinin yüksek güvenle doğru tahmin ettiği kolay örneklerde modülasyon faktörü sıfıra yaklaşarak toplam kayba katkıyı asimptotik olarak yok eder.

Bu çalışmada, MFL’nin sınıflar arası dinamik dengeleme avantajından yararlanılmış; ancak eğitim sırasındaki aşırı gradyan güncellemelerini engellemek ve öğrenme kararlılığını artırmak amacıyla tüm kayıp fonksiyonu küresel bir ölçekleme faktörü ($\lambda_{MFL} = 0,25$) ile normalize edilmiştir. Bu sabit çarpan, sınıf seçiciliğini değiştirmemekte, yalnızca adım büyüklüğünü kontrol etmektedir. MFL formülasyonu, F. 10’da tanımlanan h_i tahmin olasılık değeri kullanılarak F. 12’de verildiği şekilde hesaplanır.

$$MFL = -\lambda_{MFL}(1 - h_i)^{\gamma_{FL}} \log(h_i) \quad (F. 12)$$

2.10.5. Dice kaybı

Dice kaybı özellikle görüntü bölütleme (segmentasyon) alanında kullanılarak piksellerin (veya voksellerin) doğruluğunu tekil olarak ölçmek yerine, mimarinin tahmini ile gerçek veri maskesi arasındaki geometrik örtüşmeye (*overlap*) odaklanmaktadır (Milletari vd., 2016). Temelde bir benzerlik metriği olan Dice 0 ile 1 arasında bir değer almaktadır (Dice, 1945). 1 değeri kusursuz eşleşmeyi ifade ederken 0 değeri ise hiçbir eşleşme olmadığını ifade eder. DL mimarilerinde kayıp fonksiyonlarının doğası gereği düşürülmesi amaçlandığından eğitim sırasındaki hesaplama işlemi $1 - Dice$ formülü üzerinden yürütülmektedir. Optimizasyon

sırasındaki gradyan akışının sürekliliğinin sağlanması amacıyla bu çalışmada toplam tabanlı bir yaklaşım benimsenmiştir. F. 13'te Dice sayısı ve F. 14'te ise Dice kaybının matematiksel hesabı gösterilmiştir.

$$Dice = \frac{2 \sum_{i=1}^{N_T} y_i \hat{y}_i + \epsilon}{\sum_{i=1}^{N_T} y_i + \sum_{i=1}^{N_T} \hat{y}_i + \epsilon} \quad (F. 13)$$

$$\mathcal{L}_{Dice} = 1 - Dice \quad (F. 14)$$

Burada yer alan ϵ parametresi, her iki kümenin de boş olduğu durumlarda matematiksel istikrarı sağlayan yumuşatma katsayısı olarak ifade edilir. Ayrıca paydanın 0 olduğu durumlarda oluşacak sıfıra bölme hatası da bu katsayı ile önlenmiş olur. Bu yaklaşım az sayıda vokselin dolu olduğu seyrek nesnelerde bu boş voksellerin mimarinin başarısını baskılamasını engellemektedir.

2.10.6. IoU kaybı

IoU kaybı Dice kaybı ile benzer matematiksel temellere dayanan ancak hataları çok daha sert bir şekilde cezalandıran bir kayıp fonksiyonu olarak ifade edilir (Rahman ve Wang, 2016). Hesaplama bakımından aradaki temel fark IoU fonksiyonunun birleşim hesaplaması sırasında kesişim bölgesini toplam değerden çıkarmasıdır. Bu sayede mimarinin geometrik sınırları daha hassas öğrenmesi zorunlu kılınır. Optimizasyon süreci Dice'a benzer şekilde $1 - IoU$ formülasyonu ile hesaplanmaktadır. IoU formülasyonları F. 15 ve F. 16'da gösterilmiştir.

$$IoU = \frac{2 \sum_{i=1}^{N_T} y_i \hat{y}_i + \epsilon}{\sum_{i=1}^{N_T} Y_i + \sum_{i=1}^{N_T} \hat{y}_i - \sum_{i=1}^{N_T} y_i \hat{y}_i + \epsilon} \quad (F. 15)$$

$$\mathcal{L}_{IoU} = 1 - IoU \quad (F. 16)$$

2.10.7. AWIF kaybı

AWIF kaybı, özellikle 3D seyrek vokselize yapılarıdaki sınıf dengesizliği ve yüzey kalitesizliği problemlerini çözmek üzere literatüre kazandırılan karma bir kayıp fonksiyonudur. Standart kayıp fonksiyonlarının yetersiz kaldığı bu 3D ortam zorluklarını aşmak için AWIF, birbirini tamamlayan üç temel bileşenden meydana gelmekte ve gücünü bu bileşenlerin eşzamanlı çalışmasından almaktadır. Bunlar:

- **Adaptif Zorluk Değişimi (*Adaptive Difficulty Progression*):** DL mimarileri, eğitimin ilk aşamalarında yüksek frekanslı detayları öğrenmeye zorlanırsa gradyan patlaması yaşayabilir. Bu bileşen, müfredat öğrenimi (*curriculum learning*) prensibini uygular. Eğitimin başlarında ağırlıklı olarak nesnenin kaba hacmini (genel yapısını) öğrenmesine

izin verir; eğitim ilerledikçe zorluk çarpanını dinamik olarak artırarak mimariyi ince detayları ve zor vokselleri öğrenmeye zorlar.

- **IoU Rehberli Ağırlıklandırma (*IoU-Guided Weighting*):** Sabit ağırlıklı kayıp fonksiyonları, ağırlığın zaten iyi öğrendiği örnekler için gereksiz gradyanlar üretirken, zorlandığı örneklerin hata sinyallerini bastırabilir. Bu bileşen, mimarinin anlık başarımını (IoU skorunu) dinamik olarak ölçer. Eğer bir örneğin tahmini kolaysa (yüksek IoU), odaklanmayı azaltır; eğer örnek zorsa (düşük IoU), o spesifik örnek için hata ağırlığını artırarak mimarinin zor vakalara agresif bir şekilde odaklanmasını sağlar.
- **Yüzey Farkındalıklı Odak Bileşeni (*Surface-Aware Focal Component*):** 3D voksellerin ızgaralarının doğası gereği hacmin büyük bir kısmı (%90'dan fazlası) boşluktan oluşur. Standart kayıp fonksiyonları, gradyanları bu boş uzaya eşit olarak dağıtma eğilimindedir. Bu bileşen, 6-yönlü bağlantı algoritması kullanarak doğrudan nesnenin sınır (yüzey) voksellerini tespit eder. Mimariyi iç voksellerden ziyade dış geometriye odaklanmaya zorlayarak, 3D nesnelerin şekilsel bütünlüğünün ve ince yüzey detaylarının kaybolmasını engeller.

Adaptif zorluk değişimi bileşeni ile müfredat öğrenimi temel alınmıştır (Bengio vd., 2009). DL mimarileri eğitimin ilk aşamalarında karmaşık ve yüksek frekanslı detayları öğrenmeye zorlanırsa gradyan patlaması veya yerel minimumlara takılma gibi problemler meydana gelebilmektedir. Buna bağlı olarak F. 17 kullanılarak zorluk çarpanı değerinin eğitimin ilk %30'luk kısmında 0'dan 1'e çıkması ve sonrasında da 1 olarak sabit kalması sağlanmıştır. Bu formülasyon ile mimari eğitimin başında sadece kaba hacmi öğrenmeye odaklanır. Eğitim ilerledikçe bu zorluk çarpanı artar. F. 18 ve F. 19 bundan doğrudan etkilenir. Buna bağlı olarak kayıp fonksiyonu kolay vokselleri tolere etmez ve mimarinin giderek daha zor vokselleri öğrenmesi sağlanır.

$$d = \min\left(1, \frac{e}{e_{max} \cdot 0,3}\right) \quad (F. 17)$$

$$\gamma_{AWIF} = 2 + d \cdot 2 \quad (F. 18)$$

Burada:

d : İlgili iterasyondaki adaptif zorluk çarpanını ($0 \leq d \leq 1$),

e : Mimarinin o an bulunduğu turu,

e_{max} : Eğitim için belirlenen toplam tur sayısını,

γ_{AWIF} : Örnek için hesaplanan adaptif odak tolerans parametresini ifade eder.

Yüzey farkındalıklı odak bileşeni ile 6-yönlü bağlantı algoritması kullanılarak, üretilen sahte nesnelerin değil veri setinde bulunan gerçek nesnelerin yüzey yani sınır vokselleri tespit edilmektedir (Rosenfeld ve Kak, 1976). Vokselize nesneler dolu bir hacim kütleli olarak ifade edilebilir. Buna bağlı olarak da bu bileşen dolu bir vokselin bitişiğinde (altında, üstünde, sağında, solunda, önünde ve arkasında) herhangi bir boş voksel tespit ederse bunu doğrudan yüzey vokseli olarak etiketlemekte ve mimarinin özellikle bu voksellere ağırlık vermesi sağlanmaktadır. Bunun için F. 19’da verilen formülasyon kullanılır. Bu formülasyonda S ile ifade edilen yüzey vokselleri için çıkarılan yüzey maskesidir. Bu maske 2D ortamda sıklıkla kullanılan kenar bulma filtresi gibi çalışarak yüzey voksellerine 1 iç voksellere ve nesnenin dışındaki tamamen boş voksellere ise 0 değerinin atandığı bir tensördür. Yapılan analizlerde yüzey voksellerine verilen önem için optimum değerin 3 olduğu tespit edilmiş ve kullanılan bu yaklaşım ile mimari tarafından iç vokseller yerine dış geometriye odaklanması 4 kata kadar artırılmıştır. Bu işlemlerin ardından yüzey farkındalıklı odak F. 20 ve F. 21’de gösterildiği şekilde hesaplanır.

$$w_{s,i} = 1 + M_{surf,i} \cdot 3 \cdot d \quad (F. 19)$$

$$p_{t,i} = t_i \cdot p_i + (1 - t_i) \cdot (1 - p_i)$$

$$FL_i = -\alpha_j \cdot (1 - p_{t,i})^{Y_{AWIF,j}} \cdot [t_i \cdot \log(p_i) + (1 - t_i) \cdot \log(1 - p_i)] \quad (F. 20)$$

$$F_{enh,j} = \sum_{i=1}^{N_T} FL_i \cdot w_{s,i} \quad (F. 21)$$

Burada:

$w_{s,i}$: i. voksel için hesaplanan yüzey ağırlık katsayısını,

$M_{surf,i}$: i. vokselin yüzey maskesindeki değerini,

$F_{en,j}$: j. örnek için yüzey farkındalığı ile güçlendirilmiş toplam focal kaybını,

N_T : Bir örnekte bulunan toplam voksel sayısını (örneğin $32 \times 32 \times 32$ çözünürlük için bu sayıların doğrudan çarpımı),

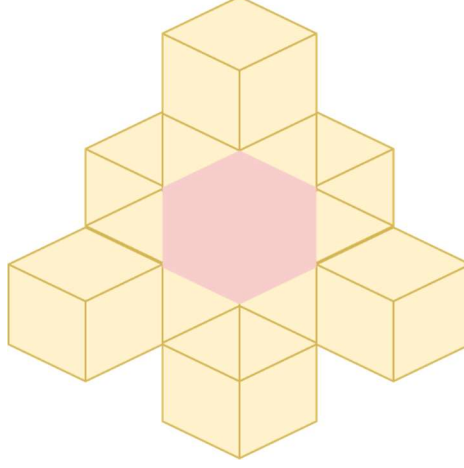
FL_i : i. vokselin sınırlandırılmış standart focal kaybını,

p_i : Mimarinin i. vokselin dolu olma ihtimaline dair ürettiği tahmin olasılığını,

α_j : F. 23’te hesaplanışı verilen ve IoU başarımına göre çapraz entropi kaybını ölçeklendiren adaptif simetrik ağırlık katsayısını,

t_i : i. vokselin veri setindeki gerçek etiketini ifade etmektedir.

Şekil 2.8’de örnek bir vokselize nesne 3D yapıda gösterilmiştir. Şekil incelenecek olursa sarı voksellerin tamamının bitişiğindeki en az 1 vokselin boş olduğu görülür. Dolayısıyla sarı vokseller yüzey vokselleridir. Buna karşın kırmızı vokselin bitişiğindeki voksellerin tamamı doludur ve buna bağlı olarak da kırmızı vokselin bir iç voksel olduğu açığa çıkar.



Şekil 2.8. Örnek bir vokselize nesne

Eğitim aşamasındaki her bir örneğin zorluk derecesi ve ağırlık o örnekteki mevcut başarımla değişkendir. Sabit ağırlıklı kayıp fonksiyonları zaten iyi öğrenilmiş örnekler için gereksiz gradyanlar üretirken kötü öğrenilmiş örneklerin hata sinyallerini bastırmaktadır. IoU rehberli ağırlıklandırma bileşeni bu sorunu çözmek için 3 alt bileşen içerir. İlk bileşende F. 22 kullanılarak yığın içindeki düşük IoU skoruna sahip örneklerin kayıp değeri yapay bir şekilde artırılır. İkinci bileşende fonksiyon F. 23’ü kullanarak başarı düştükçe α değerinin artmasını sağlar. Orijinal Focal kayıptaki asimetrik arka plan dengelemesinin aksine, bu çalışmada α katsayısı tüm voksellere simetrik olarak uygulanır. Böylece ağırlık yalnızca dolu voksellere değil tahmini zor olan nesnenin tüm hacimsel yapısına odaklanılması hedeflenmiştir. Üçüncü bileşende fonksiyon F. 24 ve F. 25’de verilen Dice kaybını ve Focal kaybı birleştirir. Bu bileşenler neticesinde nesne henüz genel olarak öğrenilememiş yani doğrudan IoU skoru düşük ise mimari tarafından hacimsel örtüşmenin maksimize edilmesi amacıyla Dice katsayısına daha yüksek bir değer atanır. Tam tersi durumda mimari nesneyi genel olarak öğrenmiş ve IoU skoru yüksek ise Dice kaybının katsayısı düşürülür ve sistem tamamen ince detayları ve kenar voksellerini düzeltmek için tasarlanan Focal kayba güvenir. Böylece eğitime herhangi bir müdahale yapılmadan ağırlık başarımına göre bu geçiş sağlanmış olur.

$$w_j = 1 + (1 - IoU_j) \cdot 2 \quad (F. 22)$$

$$\alpha_j = 0,25 + (1 - IoU_j) \cdot 0,75 \quad (F. 23)$$

$$\beta_j = (1 - IoU_j) \cdot 0,6 \quad (F. 24)$$

$$\lambda_j = IoU_j \cdot 0,4 + 0,6 \quad (F. 25)$$

Burada:

w_j : j. örneğin eğitim kümesi içindeki hata payını modüle eden örnek ağırlığını,

IoU_j : j. örnek için tahmin edilen hacim ile gerçek hacim arasındaki IoU skorunu,

α_j : j. örnek için hesaplanan adaptif simetrik ağırlık katsayısını,

β_j : j. örneğin Dice kaybı etkisini belirleyen katsayıyı,

λ_j : j. örneğin Focal kaybı etkisini belirleyen katsayıyı ifade etmektedir.

Burada ilk bakışta adaptif zorluk değişimi ve IoU rehberli ağırlıklandırma bileşenlerinin aynı şeyi yaptığı düşünülebilir ancak bu yanlış bir varsayım olacaktır. Öncelikle adaptif zorluk değişimi tamamen zaman odaklı bir bileşendir yani referans olarak sadece eğitimin o anki tur sayısını temel alır. IoU rehberli ağırlıklandırma ise tamamen başarımla odaklıdır. Zamanla herhangi bir ilişkisi bulunmaz. Mimarının o anki iterasyonda ürettiği tahmini nesne ile gerçek nesne arasındaki örtüşme miktarını temel alır. Bir diğer önemli nokta ise adaptif zorluk değişimi o anki iterasyonda bulunan tüm örneklerle aynı zorluk çarpanının uygulanmasını sağlar. IoU rehberli ağırlıklandırma ise o iterasyondaki her bir özel örnek için tek tek yeniden değerlendirilir. Bunu bir örnekle açıklamak gerekirse o anki kümede çok kolay tahmin edilebilecek bir küp nesnesi ve çok zor tahmin edilebilecek ince yapılı bir uçak nesnesi olduğunu düşünelim. Eğitimin 50. turuna gelindiğinde adaptif zorluk değişimi bileşeni her iki nesnenin de hata ağırlığını eğitim ilerledi diyerek artırır. Buna karşın IoU rehberli ağırlıklandırma bileşeni bu iki nesneye ayrı ayrı bakar. Küpün IoU değeri zaten yüksek olduğu için hata ağırlığını azaltır buna karşın uçağın IoU değeri düşük olduğu için sadece uçağa özel parametreleri artırır. Tüm bu bileşenlerin birleştirilmesiyle elde edilen nihai AWIF kayıp fonksiyonu F. 26'da ve AWIF'in adım adım çalışması Algoritma 2.1'de sözde kod olarak gösterilmiştir. AWIF formülasyonunda IoU bileşeni ağırlıklandırma yapısını öğrenmesini domine etmemesi için 100 sabiti ile ölçeklendirilmiştir. Temel hacimsel öğrenme Dice ve Focal bileşenleri üzerinden sağlanırken IoU bileşeni bir mikro düzenleyici ceza faktörü olarak konumlandırılmıştır.

$$\mathcal{L}_{AWIF} = \sum_{j=1}^B w_j \cdot [\lambda_j \cdot F_{enh,j} + \beta_j \cdot (1 - D_j) \cdot N + (1 - IoU_j) \cdot 100] \quad (F. 26)$$

Burada:

$\mathcal{L}_{AWIF}$: İlgili küme üzerinden hesaplanan nihai AWIF kaybını temsil etmektedir.

B : İlgili iterasyondaki yığın boyutunu,

D_j : j . örnek için tahmin edilen ve gerçek vokseller üzerinden hesaplanan Dice katsayısını ifade etmektedir.

Algoritma 2.1. AWIF sözde kodu.

Uyarlamalı Ağırlıklandırılmış IoU-Odak Kaybı (AWIF)
Girdi: Tahmin Vokselleri $P \in [0, 1]^{B \times N}$, Hedef Vokseller $T \in \{0, 1\}^{B \times N}$ Parametreler: Güncel tur e , Maksimum tur e_{max} , Yumuşatma $\epsilon = 1e^{-6}$ Çıktı: Toplam AWIF Kaybı $\mathcal{L}_{AWIF}$
1. $d \leftarrow \min\left(1, \frac{e}{e_{max} \times 0,3}\right)$ 2. $\mathcal{L}_{AWIF} \leftarrow 0$ 3. Döngü Başlangıcı: Yığındaki her bir $p, t \in (P, T)$ örneği için 4. $IoU \leftarrow \frac{\sum(p \cdot t) + \epsilon}{\sum p + \sum t - \sum(p \cdot t) + \epsilon}$ 5. $M_{surf} \leftarrow \mathbb{I}(\sum_v \mathbb{I}(t > 0,5 \wedge shift(t, v) < 0,5) > 0)$ // 6-Yönlü Yüzey Tespiti 6. $\alpha \leftarrow 0,25 + (1 - IoU) \times 0,75$ 7. $\gamma_{AWIF} \leftarrow 2,0 + d \times 2,0$ 8. $P_t \leftarrow t \cdot p + (1 - t) \cdot (1 - p)$ 9. $\mathcal{L}_{bce} \leftarrow -[t \cdot \log(p) + (1 - t) \cdot \log(1 - p)]$ 10. $FL \leftarrow \alpha \cdot (1 - p_t)^{\gamma_{AWIF}} \cdot \mathcal{L}_{bce}$ 11. $w_s \leftarrow 1,0 + M_{surf} \times 3,0 \times d$ 12. $F_{enh} \leftarrow \sum(FL \cdot w_s)$ 13. $D \leftarrow \frac{2 \sum(p \cdot t) + \epsilon}{\sum p + \sum t + \epsilon}$ 14. $\beta \leftarrow (1 - IoU) \times 0,6$ 15. $\lambda \leftarrow IoU \times 0,4 + 0,6$ 16. $w \leftarrow 1,0 + (1 - IoU) \times 2,0$ 17. $\mathcal{L}_{AWIF} \leftarrow \mathcal{L}_{AWIF} + (w \cdot (\lambda \cdot F_{enh} + \beta \cdot (1 - D) \cdot N + (1 - IoU) \times 100,0))$ 18. Döngü Sonu 19. Döndür $\mathcal{L}_{AWIF}$

2.11. Dikkat Mekanizmaları

DL mimarilerinde 2D görüntülerden 3D nesne üretimi doğası gereği ciddi bir bilgi eksikliği ve belirsizlik barındıran zorlu bir problem olarak ifade edilir. Geleneksel evrişim katmanları ve bunların yaptığı işlemler yalnızca yerel alıcı alanlar içerisinde kullanıldıkları için öznitelik haritalarındaki küresel ilişkileri yakalamada yetersiz kalabilmektedir. 3×3 veya 5×5 'lik bir filtre görüntü üzerinde kaydırılırken bu filtrenin kapsadığı alandaki öznitelikleri tarayabilmekte bu alanın dışında kalan diğer öznitelikler ile bu alanın ilişkisini öğrenememektedir.

Bu yapısal kısıtlamayı aşmak için literatürde dikkat mekanizmaları geliştirilmiştir. Bu mekanizmalar insan görsel sistemini temel alarak ilgili DL mimarisinin girdi verisindeki tüm

bilgilere eşit ağırlık vermesi yerine göreve yönelik olarak en önemli özniteliklere odaklanılmasını sağlamaktadır (Li vd., 2025). Böylece gereksiz ve önemsiz gürültüler baskılanır. DL literatüründe bu mekanizmalar işlem yaptıkları veri boyutlarına ve matematiksel temellerine göre kanal ve uzamsal dikkat mekanizmaları ile öz dikkat mekanizmaları olarak iki ana kategoriye ayrılmaktadır.

2.11.1. Kanal ve uzamsal dikkat mekanizmaları

Kanal ve uzamsal dikkat mekanizmaları aslında temel olarak evrişim işlemine dayalı mekanizmalar olarak da ifade edilebilir. Standart evrişim işlemlerinin ürettiği öznitelik haritalarını girdi olarak alarak bu haritalarda bulunan önemli bölgeleri, alanları veya kanalları doğrusal olmayan fonksiyonlar kullanarak yeniden ölçeklendirmeyi sağlar. Literatürdeki en bilinen örnekleri Sıkıştırma ve Uyarma (*Squeeze-and-Excitation* – SE) ve Evrişimsel Blok Dikkat Modülü (*Convolutional Block Attention Module* – CBAM) dikkat mekanizmalarıdır.

SE ağı, dikkat mekanizmalarının yalnızca kanal boyutunda uygulanmasına odaklanan temel bir mimari olarak ifade edilir (Hu vd., 2018). Geleneksel evrişim katmanlarında her bir kanalın eşit öneme sahip olması varsayımını reddeden bu mimari bazı filtrelerin (kanalların) diğerlerinden daha önemli bilgiler taşıdığını öne sürer. Bu mimari sıkıştırma ve uyarma olmak üzere iki aşamada çalışır.

Sıkıştırma aşamasında $C \times H \times W$ boyutlarındaki girdi öznitelik haritası küresel ortalama havuzlama işlemine tabi tutulur. Böylece tensör daraltılır ve $C \times 1 \times 1$ boyutlarına sahip bir z vektörüne dönüştürülür. Yapılan bu işlem F. 27’de matematiksel olarak gösterilmiştir.

$$z_c = \frac{1}{H \cdot W} \sum_{i=1}^H \sum_{j=1}^W X_c(i, j) \quad (\text{F. 27})$$

Burada:

z_c : Sıkıştırma işlemi sonucunda elde edilen vektörün c . kanalına ait değeri,

X_c : Girdi öznitelik haritasının c . kanalını temsil eden 2D uzamsal matrisi,

$X_c(i, j)$: X_c matrisinin i . satır ve j . sütununda bulunan aktivasyon değerini ifade etmektedir.

Uyarma aşamasında ise elde edilen bu z vektörü kanallar arasındaki ilişkileri öğrenmek üzere iki tam bağlı katmandan oluşan bir MLP’ye girdi olarak verilir. ReLU ve sigmoid aktivasyon fonksiyonları kullanılarak her bir kanal için 0 ile 1 arasında değişen bir önem katsayısı elde edilmiş olur. İsminden de anlaşılacağı üzere bu değer ne kadar yüksek ise ilgili

öznitelik kanalının mimari için o kadar önemli olduğu anlaşılır. Son olarak hesaplanan bu katsayılar orijinal öznitelik haritası ile eleman eleman çarpma işlemine tabi tutularak ağırlıklandırılması gerçekleştirilir. Böylece sonuç için hangi kanalların önemli hangilerinin önemsiz olduğu tespit edilmiş olur. Yapılan bu hesaplamaların matematiksel formülasyonları F. 28 ve F. 29'da gösterilmiştir.

$$s = \sigma_{sig}(W_2 \delta(W_1 z)) \quad (F. 28)$$

$$\tilde{X}_c = s_c \cdot X_c \quad (F. 29)$$

Burada:

s : Her bir kanal için önem katsayısını barındıran ölçeklendirme vektörünü,

σ_{sig} : Sigmoid fonksiyonunu,

W_2 : İkinci tam bağlı katmana ait kanalları orijinal boyutuna geri getiren ağırlık matrisini

δ : ReLU fonksiyonunu,

W_1 : İlk tam bağlı katmana ait boyut indirgeme işlemi yapan ağırlık matrisini,

$\tilde{X}_c$: Dikkat mekanizması tarafından ağırlıklandırılarak güncellenmiş yeni öznitelik haritasının c . kanalını,

s_c : s vektörü içindeki c . kanala ait skaler önem katsayısını,

X_c : Orijinal girdi öznitelik haritasının c . kanalını ifade etmektedir.

SE mekanizmasının yalnızca kanallara odaklanma zafiyetini gidermek CBAM mimarisi önerilmiştir (Woo vd., 2018). Bu mimari yapı ile neye (kanal) dikkat edilmesi gerektiği sorusuna ek olarak nereye (uzamsal) dikkat edilmesi gerektiğinin de cevabı verilmiştir.

İlk aşamada kanal dikkati, SE ağından farklı olarak bilgi kaybını da önlemeye odaklanmış ve bunun için küresel ortalama havuzlama ile aynı anda maksimum havuzlama da uygulanmıştır. Bu işlemlerin çıktısı ortak bir MLP yapısından geçirilerek toplanmış ve sigmoid fonksiyonu ile kanal ağırlıkları elde edilmiştir. Bu işlem F. 30'da matematiksel olarak ifade edilmiştir.

$$M_c(F) = \sigma_{sig} \left(MLP(AvgPool(F)) + MLP(MaxPool(F)) \right) \quad (F. 30)$$

Burada:

F : CBAM mekanizmasına giren $C \times H \times W$ boyutlarındaki orijinal girdi öznitelik haritasını,

M_C : Hesaplanan $C \times 1 \times 1$ boyutlarındaki kanal dikkat maskesini ifade etmektedir.

Elde edilen kanal ağırlıklı öznitelik haritası ikinci aşama olan uzamsal dikkat bileşenine iletilir. Burada kanal eksenini boyunca yine ortalama ve maksimum havuzlama işlemleri yapılarak 2D yeni bir öznitelik haritası elde edilir. Bu harita 7×7 boyutunda geniş bir evrişim filtresinden geçirilerek uzamsal önem maskesi oluşturulur. Böylece hangi kanalın önemli olduğunun yanında hangi konumun da önemli olduğu bilgisi elde edilmiştir. F. 31’de bu işlemin matematiksel formülasyonu gösterilmiştir.

$$M_s(F') = \sigma(f^{7 \times 7}([AvgPool(F'); MaxPool(F')])) \quad (F. 31)$$

Burada:

M_s : Uzamsal dikkat aşamasında hesaplanan $1 \times H \times W$ boyutlarındaki 2D uzamsal önem maskesini,

F' : Orijinal girdinin kanal dikkat maskesi ile çarpılması sonucu elde edilen ve uzamsal dikkat bileşenine girdi olarak verilen ara öznitelik haritasını,

$f^{7 \times 7}$: Geniş alansal bağlamı yakalamak için kullanılan 7×7 boyutundaki evrişim işlemini ifade etmektedir.

CBAM mimarisi hem kanal hem de yerel uzamsal ilişkileri kurmada başarılı olmasına karşın uyguladığı uzamsal dikkat işlemi halen evrişim çekirdeklerinin boyutuyla sınırlıdır. Arasında mesafe bulunan piksellerin birbirleriyle olan küresel ve tekrarlı ilişkilerini yakalayamaması bu mimarinin karmaşık 3D geometrileri üretmedeki kapasitesini sınırlandırmaktadır.

2.11.2. Öz dikkat mekanizmaları

Öz dikkat kavramı literatürde daha önceki yıllarda bulunmasına karşın günümüzdeki modern matematiksel formuna ilk olarak 2017 yılında Vaswani ve arkadaşları tarafından ulaştırılmıştır. 2017 yılında yayımlanan ve DL literatüründeki en önemli dönüm noktalarından birisi olan “Attention is All You Need” başlıklı makale ile literatüre kazandırılmıştır (Vaswani vd., 2017). Bu çalışmada o zamana kadar kullanılan geleneksel mimarilerin paralelleştirmeye imkân tanımayan uzun mesafe bağımlılıklarını öğrenmede yetersiz kalmaları üzerinde durulmuştur. Tüm öznitelik çıkarma aşaması dikkat mekanizması üstüne inşa edilerek *Transformer* mimarisi tanıtılmıştır. Bu yaklaşım ile elemanların birbirleri ile olan ilişkileri uzaklıklarından bağımsız ve eşzamanlı olarak hesaplanabilmektedir. Bu sayede işlemler paralelleştirilerek GPU kullanımı maksimize edilmektedir. Daha sonra CV alanı da dahil pek

çok DL mimarisi bu yaklaşımı benimsemiştir.

Öz dikkat mekanizması temel olarak girdi verisinde bulunan piksel, voksel veya herhangi bir sıralı verinin aynı veri seti içerisinde bulunan diğer tüm elemanlar ile olan anlamsal ve geometrik ilişkisini öğrenen bir yöntem olarak ifade edilebilir. Klasik evrişim katmanları öznitelik çıkarımını yalnızca tanımlanmış filtre boyutunun kapsadığı alan ile sınırlı tutarken öz dikkat mekanizması girdi uzayının tamamını kapsar. Bu durum sanki tüm girdi boyutunda bir filtre gibi düşünülebilir. Öz dikkat tarafından girdi öznitelik haritası doğrusal dönüşümlerden geçirilerek üç temel matris türetilmektedir.

- **Sorgu (Query – Q):** Ağın ilişkisini aradığı hedef elemanı temsilidir.
- **Anahtar (Key – K):** Girdideki diğer tüm elemanların kimliğini ve özniteliklerini temsil eder.
- **Değer (Value – V):** Girdi elemanlarının taşıdığı asıl bilgidir.

Matematiksel açıdan bu üç matris F. 32, F. 33 ve F. 34'te gösterildiği şekilde hesaplanmaktadır.

$$Q = XW_Q \quad (F. 32)$$

$$K = XW_K \quad (F. 33)$$

$$V = XW_V \quad (F. 34)$$

Burada:

X : Girdi öznitelik haritasını,

W_Q, W_K ve W_V : Öğrenilebilir ağırlık matrislerini ifade etmektedir.

Q ve K matrislerinin geometrik benzerliğini ölçmek ve dikkat skorlarını hesaplamak için eleman elemana çarpma işlemi uygulanır. Kaybolan gradyan probleminin önüne geçilmesi amacıyla elde edilen sonuçlar Q ve K 'nın boyutunun kareköküne bölünerek ölçeklendirme işlemi yapılır. Daha sonra sonuçlar softmax fonksiyonundan geçirilerek 0 ile 1 arasında olasılık değerleri elde edilir. Son durumda küresel dikkat haritası olarak isimlendirilen bu çıktı V matrisi ile çarpılarak verinin son çıktısı oluşturulur. Bu işlemler matematiksel olarak F. 35 formülasyonu ile gösterilmiştir.

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_K}}\right)V \quad (F. 35)$$

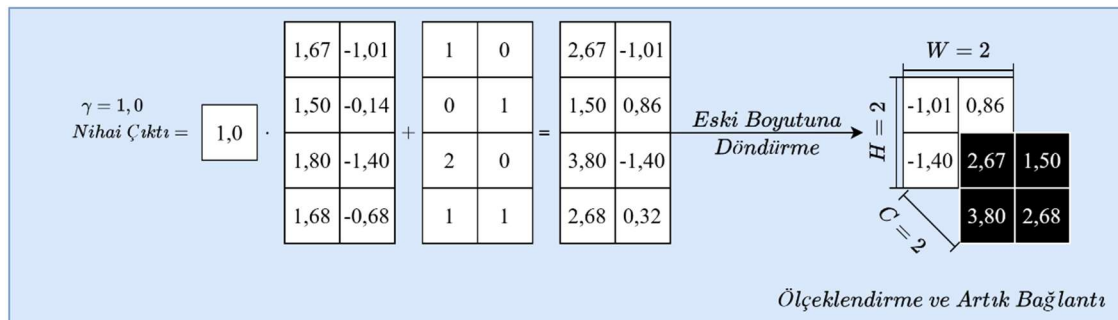
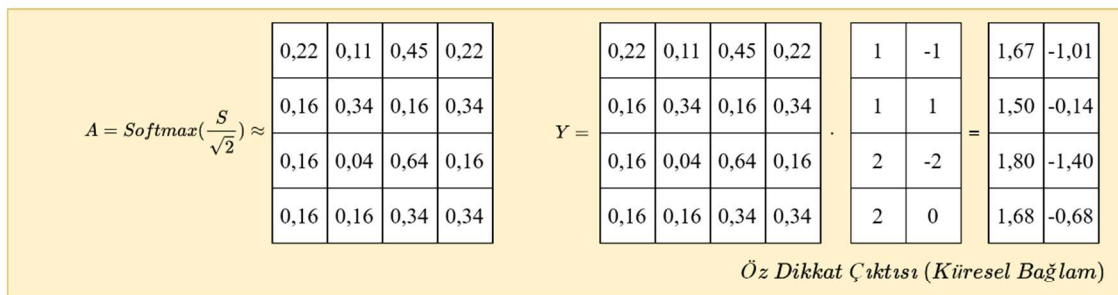
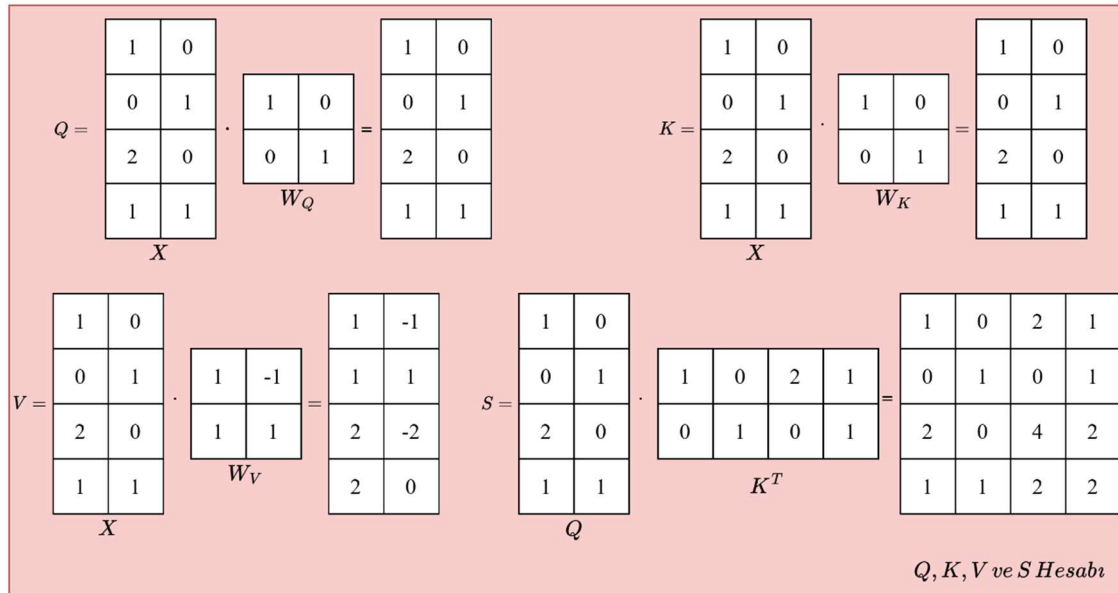
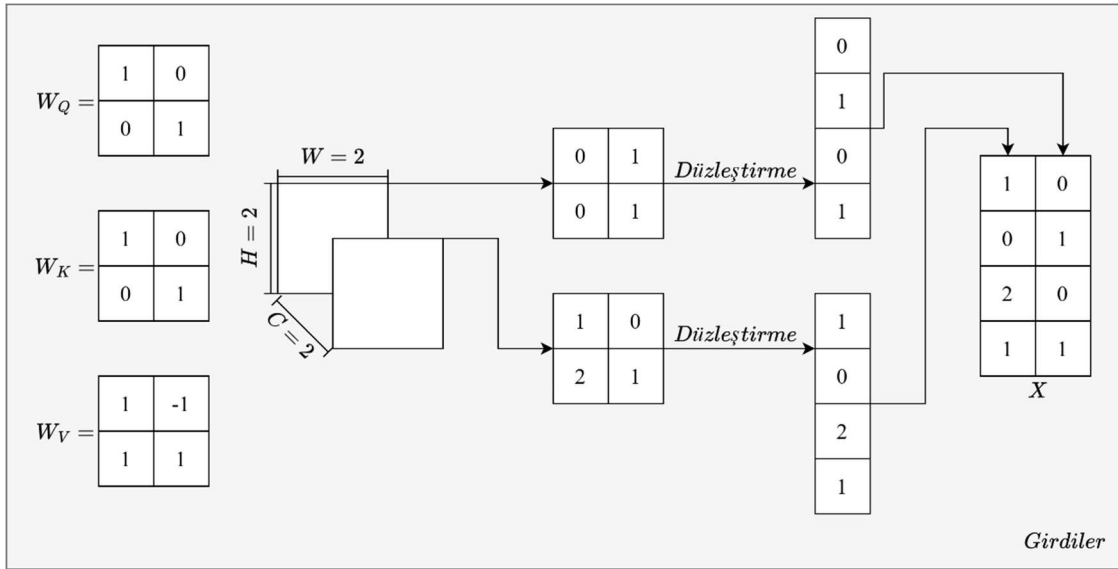
Burada:

d_K : Anahtar (K) vektörlerinin boyutunu ifade etmektedir.

CV mimarilerinde bu öz dikkat mekanizması entegre edilirken $C \times H \times W \times D$ boyutlarındaki 3D öznitelik haritası matematiksel işlemlere uygun hale getirilmek için $C \times N$ formatına getirilir. Buradaki N değeri H, W ve D değerlerinin çarpımına eşittir. Bu işlem sonucu hesaplanan dikkat çıktısı en son eski boyutlarına döndürülerek bir γ_{SA} parametresi ile ölçeklenir. Son olarak girdi üzerine bir artık bağlantı olarak eklenerek bir verinin konumdan bağımsız olarak piksellerinin veya voksellerinin aralarında bulunan ilişkiler öğrenilmiş olur. Yapılan bu işlemlerin matematiksel formülasyonu F. 36'da verilmiştir.

$$Y = \gamma_{SA} \cdot Attention(Q, K, V) + X \quad (F. 36)$$

Şekil 2.9'da kısıtlı bir öznitelik uzayı üzerinden öz dikkat mekanizmasının matematiksel işleyişi adım adım gösterilmiştir. İlk blokta (Girdiler aşaması) farazi olarak tanımlanan öğrenilebilir ağırlık matrisleri ve kanal sayısı 2, uzamsal çözünürlüğü ise 2×2 olan bir öznitelik haritasının kanal bazlı düzleştirilerek tensör çarpımına hazır hale getirilmesi gösterilmiştir. İkinci blokta (Q, K, V ve S Hesabı) girdilerin bu ağırlık matrisleriyle çarpılarak dönüştürüldüğü yeni anlamsal uzaylar ve pikseller arası ham benzerliği ifade eden dikkat skorları (S) hesaplanmıştır. Üçüncü blokta (Öz Dikkat Çıktısı – Küresel Bağlam) elde edilen çıktı, literatürde Küresel Bağlam (*Global Context*) matrisi olarak isimlendirilen ve uzaydaki diğer tüm piksellerin ilgili piksel üzerindeki ortak etkisini yansıtan sentezlenmiş bir öznitelik haritasıdır. Bu durum birinci piksel üzerinden örneklenecek olursa orijinal girdideki $[1, 0]$ değeri o pikselin kendi fiziksel koordinatına ait yerel kimliğini ifade etmektedir. Dikkat işlemi sonucunda hesaplanan $[1, 67, -1, 01]$ vektörü ise artık doğrudan o piksele ait bir özellik değildir. Bu vektör uzaydaki diğer piksellerin (2., 3. ve 4. pikseller) taşıdığı özelliklerin 1. piksel üzerinde oluşturduğu matematiksel ve geometrik yönlendirmedir. Elde edilen yeni değerler (örneğin -1,01 gibi negatif bir sayı) doğrudan fiziksel bir renk veya derinlik değil, ağın keşfettiği soyut öznitelik değerleridir. Ancak bu ham matris mimaride doğrudan yeni bir öznitelik haritası olarak kullanılamaz. Piksellerin yalnızca bu küresel etkiye göre güncellenmesi, literatürde aşırı düzleştirme olarak ifade edilen probleme yol açarak görüntüdeki ince yerel detayların kaybolmasına ve tüm piksellerin birbirine benzemesine sebep olacaktır. Bu nedenle küresel yönlendirme ile yerel kimlik arasındaki denge sağlanmalıdır. Dördüncü blokta (Ölçeklendirme ve Artık Bağlantı) bu dengenin sağlanması için küresel bağlam çıktısı örnek bir $\gamma = 1,0$ katsayısı ile ölçeklendirilmiş ve orijinal girdi (X) bir artık bağlantı olarak matrise eklenmiştir. Son adımda ise elde edilen nihai öznitelik haritası, bir sonraki evrişim katmanına aktarılmak üzere orijinal yapısına ($2 \times 2 \times 2$) geri döndürülerek yeniden boyutlandırılmıştır.

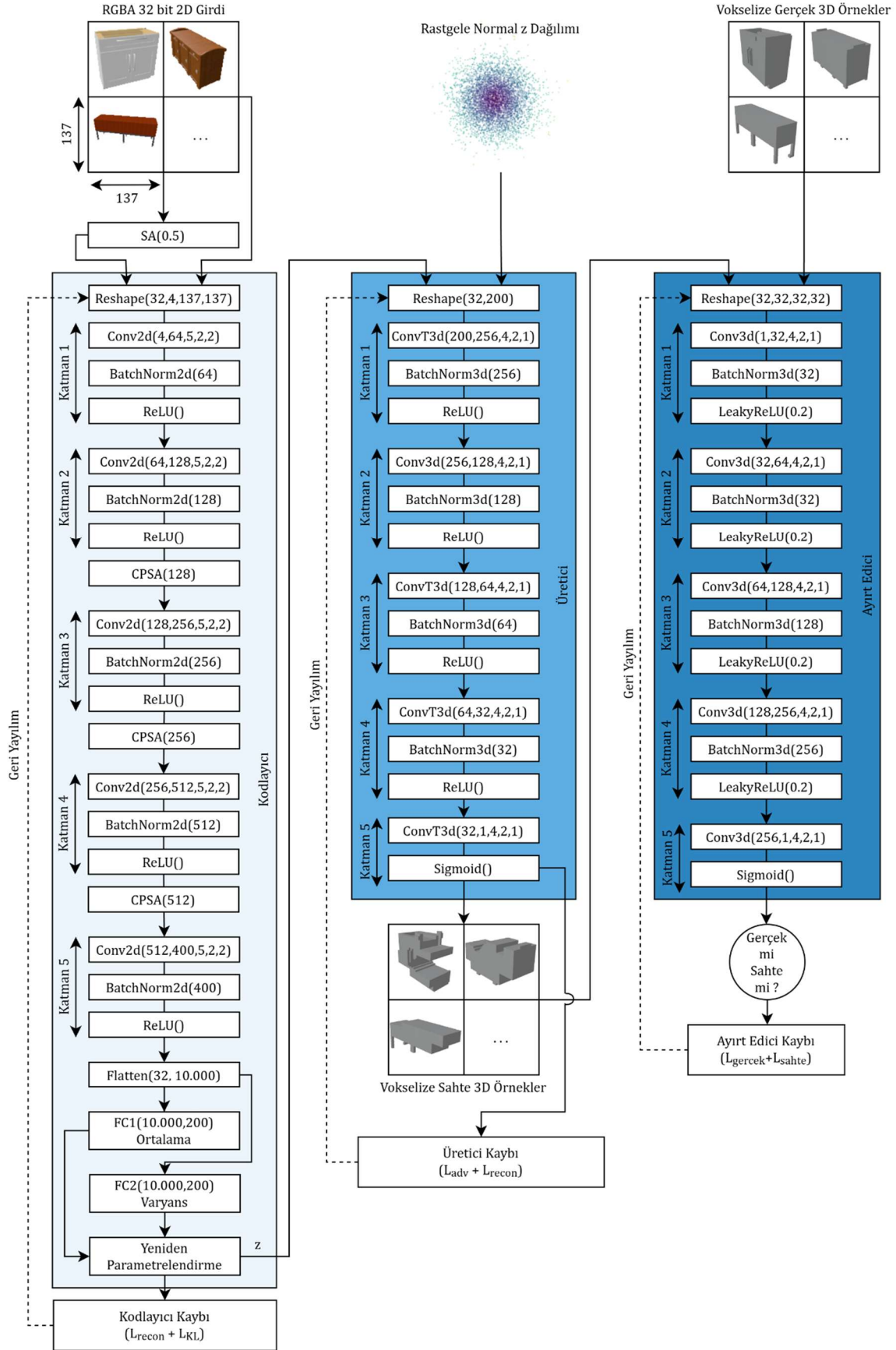


Şekil 2.9. Öz dikkat mekanizmasının sayısal örnekleme

2.11.3. CPSA mekanizması

SE ve CBAM gibi klasik dikkat mekanizmaları her ne kadar farklı bağlamsal öznitelikleri yakalamaya çalışsalar da özellikle uzamsal düzlemde işlem yapmaları ve yerel ilişkilere odaklanmaları bakımından küresel bağlamı yakalamada yapısal darboğazlara sebep olabilmektedir. Bu çalışmada, bahsedilen bu eksiklikleri gidermek amacıyla evrişim katmanlarının doğrudan tespit edemediği uzak mesafeli ilişkileri doğrudan frekans düzleminde çözümleyen CPSA adında yeni bir dikkat mekanizması önerilmiş ve literatüre kazandırılmıştır.

3D-VAE-GAN gibi DL mimarilerinin ilk katmanlarında görüntüde bulunan kenar, köşe ve renk geçişleri gibi çok temel ve düşük seviyeli öznitelikler çıkarılmaktadır. CPSA gibi karmaşık bir dikkat mekanizmasının bu katmanlarda uygulanması henüz anlamsal bir bütünlük kazanmayan gürültülü veriyi işlemek anlamına gelmektedir. Ayrıca frekans düzleminde erken aşamada yapılacak bir maskeleyme nesnenin ana hatlarını oluşturan önemli yüksek frekanslı bileşenleri sönmüştürme ihtimali taşımaktadır. Literatürde Çok Ölçekli Öznitelik İyileştirme (*Multi-Scale Feature Refinement – MSFR*) amacıyla yukarıdan aşağıya hiyerarşik bağlantıların ardışık olarak kullanıldığı görülmektedir (Lin vd., 2017a). Benzer şekilde, dikkat mekanizmalarının da tek katman yerine ağırlıklı farklı derinliklerine dağıtılması, mimarinin hem yerel hem de küresel özellikleri daha iyi kavramasını sağlamaktadır. Orta katmanlarda mimarinin nesnenin parçalarını öğrenmeye başladığı bilinir. Burada CPSA kullanılması yerel frekans gürültülerinin temizlenmesine katkıda bulunur. Derin katmanlarda ise mimari tarafından artık nesnenin küresel yapısı kavranmıştır. Burada uygulanan CPSA mekanizması nesnenin genel silüetini ve formunu arka plandan ayırarak netleştirilmesini sağlar. Düzleştirme işlemi öncesi uygulanan CPSA ise kodlayıcı tarafından üretilen en yüksek anlamlı verinin tam bağlı katmanlara girmeden son kez filtrelenmesini ve gizli uzaya nesnenin sadece saf geometrik bilgisini aktarmayı sağlar. Burada anlatılanlara dayanarak CPSA mekanizması mimaride 2, 3 ve 4. katmanlardan sonra yerleştirilmiştir. Tüm mimarinin en başında ise 2D girdilere detayları bölüm 2.12’de verilen SA veri artırma tekniği %50 olasılıkla uygulanmıştır. Adil ve eşit bir kıyaslama gerçekleştirmek için SE ve CBAM mekanizmaları da tamamen aynı konumlara yerleştirilmiştir. Şekil 2.10’da CPSA’nın ve SA’nın tam olarak konumlandırılması 3D-VAE-GAN mimarisi üzerinde detaylı bir şekilde gösterilmiştir.



Şekil 2.10. CPSA ve SA mekanizmaları eklenen temel 3D-VAE-GAN mimarisi

CPSA'nın en önemli özelliklerinden biri doğrudan frekans alanında uygulanmasına bağlı olarak nesnenin küresel yapısal bütünlüğünü kavraması ve 2D'den 3D'ye dönüştürme problemlerinde geometrik tutarlılığı sağlamasıdır. Ancak, CPSA her ince detayı kusursuz bir şekilde çözememektedir. Özellikle tüfek gibi ince yapılı sınıflarda frekans düzlemindeki küresel filtreleme yeteneği, mevcut voksel çözünürlüğü kısıtlamalarına bağlı kaldığı için sınırlı olabilmektedir. Dolayısıyla CPSA'nın asıl katkısı, yerel çok ince detaylardan ziyade, nesnenin küresel formunun yüksek doğrulukla yapılandırılmasıdır. CPSA temel olarak iki ana bileşenden meydana gelir. İlk bileşen daha önceden de zaten kullanılan ve çıkarılan öznitelik haritaları üzerinde çalışan klasik bir kanal dikkatidir (*Channel Attention – CA*). İkinci aşama ise asıl yeniliğin olduğu frekans alanında çalışan Karmaşık Dikkat Frekans Maskesi (*Complex Frequency Attention Mask – CFAM*) bileşenidir. CPSA en genel hali ile F. 37 - F. 39 arasında matematiksel olarak ifade edilen işlemleri gerçekleştirir.

$$Y = CA(X) \odot X \quad (F. 37)$$

$$M_{freq} = CFAM(Y) \quad (F. 38)$$

$$Output = Y \odot M_{freq} \quad (F. 39)$$

Burada:

X : Girdi öznitelik haritasını,

Y : Kanal dikkati uygulanmış ara öznitelik haritasını,

M_{freq} : Frekans düzleminde türetilen uzamsal dikkat maskesini,

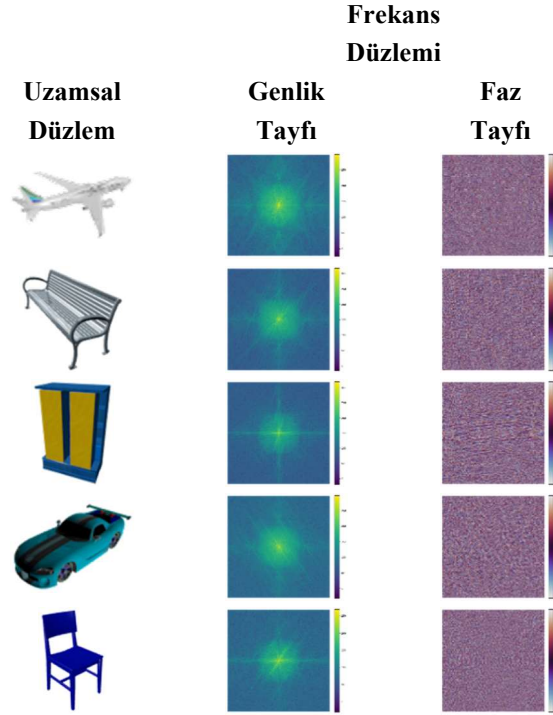
$Output$: CPSA'nın nihai çıktısını ifade etmektedir.

2.11.3.1. CA bileşeni

Bu bileşen temel olarak 2.11.1 bölümünde anlatılan SE ve CBAM mekanizmalarının uyguladığı kanal dikkatini uygular. X ile ifade edilen öznitelik haritasındaki kanallar arasında bulunan ilişkileri inceler ve hangi kanalların daha fazla yapısal bilgi taşıdığı belirlenir. Girdi haritası öncelikle küresel ortalama havuzlama ve maksimum havuzlama işlemleriyle 1×1 uzamsal boyutlarına sıkıştırılır. Elde edilen küresel bağlam vektörü bir MLP'ye girdi olarak verilir ve sigmoid aktivasyon fonksiyonu ile kanal ağırlıkları elde edilir. Bu ağırlıklar orijinal harita ile eleman elemana çarpılarak mimarinin anlamsal olarak en zengin kanallara odaklanması sağlanır.

2.11.3.2. CFAM bileşeni

Önerilen CPSA mimarisinin kalbi CFAM bileşeni tarafından oluşturulmaktadır. Bu bileşen yapılan işlemleri uzamsal düzlemde çıkarıp doğrudan frekans düzleminde gerçekleştirerek literatüre özgün bir katkı sunmaktadır. Uzamsal düzlemdeki görüntüler bilindiği üzere pikseller üzerinden ifade edilirken, frekans düzleminde veriler genlik tayfı (*magnitude spectrum*) ve faz tayfı (*phase spectrum*) olmak üzere çok daha soyut iki yapının birleşimiyle ifade edilir. Şekil 2.11’de örnek olarak verilen 5 görüntü verisi üzerinden uzamsal düzlem ve frekans düzlemi gösterilmiştir. Genlik tayfı görüntüdeki baskın ve yoğun frekansları temsil eder. Yüksek genliğe sahip bir bölüm keskin geçişleri ve kenarları, düşük genliğe sahip bir bölüm ise homojen alanları belirler. Faz tayfı ise bu frekans bileşenlerinin uzamsal konum bilgisini taşır. Sadece faz bilgisi görüntünün ana iskeletini korumaya yeterken, sadece genlik bilgisi görüntüyü tamamen tanınmaz hâle getirir.



Şekil 2.11. Görüntülerin uzamsal ve frekans düzlemlerindeki temsillerine ilişkin örnekler

Çalışmamızda önerilen CFAM bileşeninin adım adım işleyişi şu şekildedir:

Adım 1 Spektral Dönüşüm: CA mekanizması uygulanmış Y öznitelik haritası 2D Hızlı Fourier Dönüşümü (*Fast Fourier Transform 2D – FFT2D*) kullanılarak uzamsal düzlemde frekans düzlemine geçirilir (Cooley ve Tukey, 1965). F. 40’da verilen matematiksel formülasyon ile gerçekleştirilen bu işlem mimarisinin yerel pikseller yerine küresel tekrarlanan (periyodik) yapılara odaklanmasını sağlar.

$$F_Y = FFT_{2D}(Y) \quad (F. 40)$$

Adım 2 Karmaşık Modülasyon: Spektral Dönüşüm işleminin sonunda elde edilen F_Y karmaşık (*complex*) bir sayıdır. Bu sayının gerçek ($Re(F_Y)$) ve sanal ($Im(F_Y)$) kısımları ayrılarak birleştirilir ve özel olarak tasarlanılan Modülatör Ağına (*Modulator Network*) girdi olarak verilir. 3 adet 2D evrişim, 2 adet BN ve ReLU katmanlarından oluşan bu alt ağın temel görevi frekans düzlemindeki genlik ve fazı adaptif bir şekilde ayarlayacak karmaşık bir modülasyon maskesi (M_{mod}) tahmin etmektir. Bu işlem F. 41’de matematiksel olarak gösterilmiştir. Tahmin edilen bu maske F_Y ile çarpılarak karmaşık faz kaydırması olarak isimlendirdiğimiz işlem gerçekleştirilir. Bu durum F. 42’de matematiksel olarak gösterilmiştir.

$$M_{mod} = \text{Modulator} \left((Re(F_Y); Im(F_Y)) \right) \quad (\text{F. 41})$$

$$F_{Y,modulated} = F_Y \odot M_{mod} \quad (\text{F. 42})$$

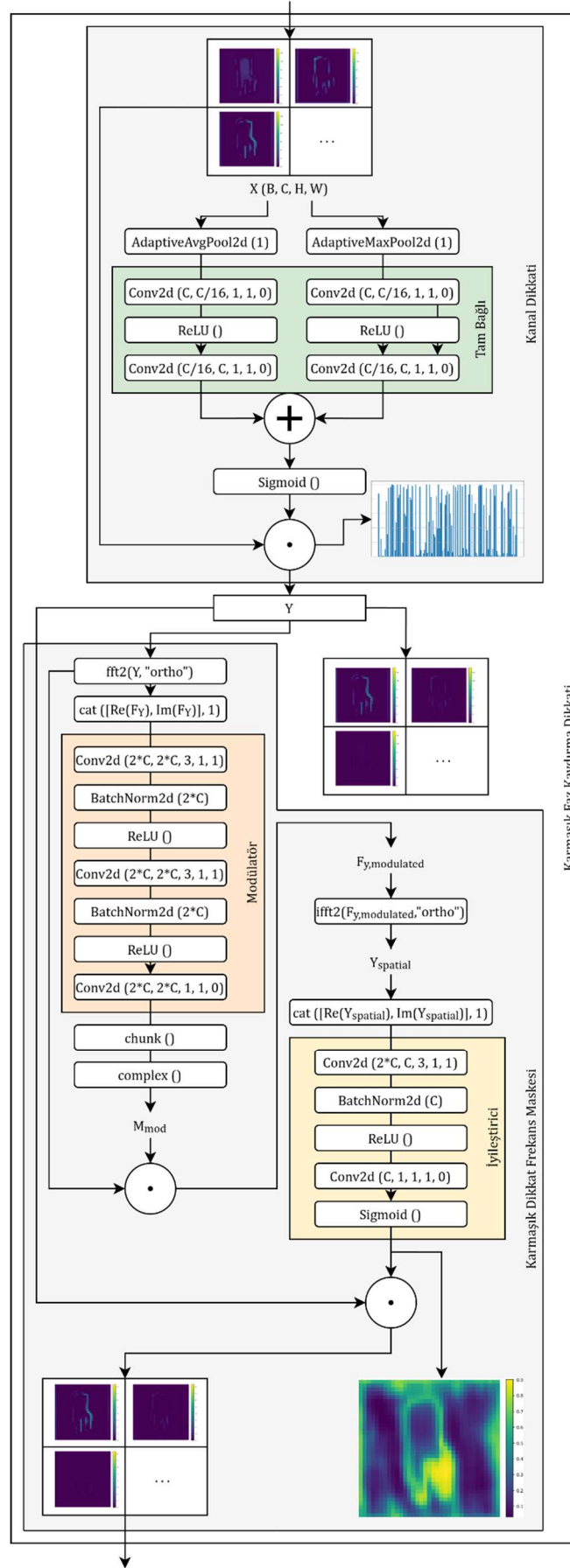
Adım 3 Uzamsal Geri Dönüşüm: Frekans düzleminde uygulanan bu filtreleme işleminin uzamsal düzlemdeki karşılığını bulmak için modüle edilen veriye 2D Ters Hızlı Fourier Dönüşümü (*Inverse Fast Fourier Transform 2D – IFFT2D*) uygulanarak uzamsal düzleme geri dönüş sağlanır. F. 43’te yapılan bu işlem matematiksel olarak gösterilmiştir.

$$Y_{spatial} = \text{IFFT}_{2D}(F_{Y,modulated}) \quad (\text{F. 43})$$

Adım 4 İyileştirme: Elde edilen yeni uzamsal öznitelik haritası da karmaşık bir yapıya sahip olduğundan gerçek ve sanal kısımları yine ayrılarak bir İyileştirici Ağ (*Refiner Network*) girdi olarak verilmiştir. Modülatör ağı ile benzer bir mimari yapıya sahip olan bu ağ karmaşık öznitelik haritasını tek kanallı ve anlamsal bir uzamsal maskeye dönüştürür. Sigmoid fonksiyonu ile 0-1 aralığına normalize edilen bu yapı sonucu nihai frekans-tabanlı dikkat maskesi olan M_{freq} değerini oluşturur. Bu durum matematiksel olarak F. 44’te gösterilmiştir.

$$M_{freq} = \sigma_{sig} \left(\text{Refiner} \left((Re(Y_{spatial}); Im(Y_{spatial})) \right) \right) \quad (\text{F. 44})$$

Elde edilen bu maske kanal dikkati uygulanmış ara öznitelik haritası ile F. 39’da verildiği üzere çarpılarak işlem tamamlanır. Şekil 2.12’de baştan sona CPSA’nın çalışma mekanizması bir öbek şema ile gösterilmiştir. Algoritma 2.2’de ise CPSA’nın adım adım yaptığı işlemler bir sözde kod olarak sunulmuştur. Önerilen bu yapı ile 3D vokselize nesneler üretilirken hem kanallar arası ilişkiler hem de klasik evrişim işlemlerinin ulaşamadığı küresel ilişkilerin eş zamanlı öğrenilebilmesi sağlanmıştır.



Şekil 2.12. CPSA mekanizmasının uçtan uca çalışma aşamaları

Algoritma 2.2. CPSA sözde kodu.

Karmaşık Faz Kaydırma Dikkati (CPSA)

Girdi: Öznitelik Haritası $X \in \mathbb{R}^{B \times C \times H \times W}$

Çıktı: İyileştirilmiş Öznitelik Haritası $Y \in \mathbb{R}^{B \times C \times H \times W}$

1. // Kanal Dikkati (CA) Aşaması
 2. $w \leftarrow CA(X)$
 3. $X_{ca} \leftarrow X \odot w$
 4. // Spektral Dönüşüm (FFT) Aşaması
 5. $F \leftarrow FFT2D(X_{ca}) \in \mathbb{C}^{B \times C \times H \times W}$
 8. Karmaşık Modülasyon (Faz Kaydırma) Aşaması
 9. $F_{in} \leftarrow [Re(F); Im(F)]$
 10. $M \leftarrow Modulator(F_{in}) \in \mathbb{C}^{B \times C \times H \times W}$
 11. $F_{mod} \leftarrow F \odot M$
 12. // Uzamsal Geri Dönüşüm (IFFT) ve İyileştirici (Refiner) Aşaması
 13. $X_{ifft} \leftarrow IFFT2D(F_{mod}) \in \mathbb{C}^{B \times C \times H \times W}$
 14. $X_{in} \leftarrow [Re(X_{ifft}); Im(X_{ifft})]$
 15. $S \leftarrow Sigmoid(Refiner(X_{in}))$
 16. // Çıktının Hesaplanması
 17. $Y \leftarrow X_{ca} \odot S$
 18. **Döndür** Y
-

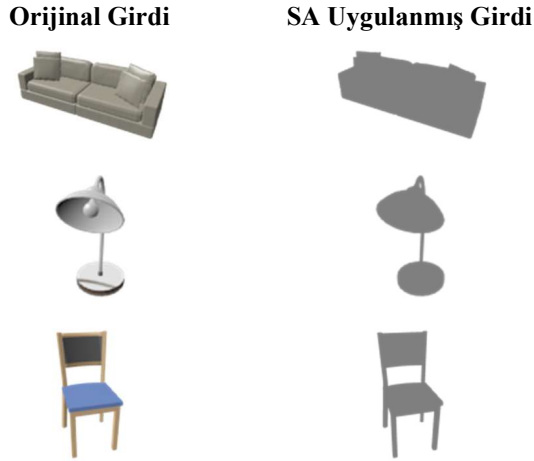
2.12. SA Yöntemi

DL mimarilerinin eğitim verisini ezberlemesini engellemek ve farklı veri dağılımlarına karşı bu mimarilerin genelleştirme yeteneğini artırmak amacıyla veri artırma teknikleri kritik bir öneme sahiptir. Özellikle 3D nesne yapılandırması gibi problemler yüksek boyutlu çıktılar ürettiği için bu durumun önemi daha da artmaktadır. 2D görüntülerden 3D nesne üretiminin en büyük zorluklarından biri kullanılan mimarilerin ilgili nesnenin topolojisini değil renk, doku ve aydınlatma gibi yüzeysel özniteliklere aşırı bağımlı hale gelmesidir. Bu durum literatürde Doku Önyargısı (*Texture Bias*) olarak isimlendirilir (Geirhos vd., 2019). Bu sorun öyle bir hal alır ki ilgili mimari tarafından nesnenin genel görünümü yok sayılarak çok özel bir detay öğrenilmeye çalışılır. Bir örnek üzerinden açıklamak gerekirse, masa nesnesinin genel hatları oluşturulmadan ayağındaki bir desen oluşturulmaya çalışılabilir. Bu çalışmada, CPSA'nın doğrudan nesnenin genel biçimine odaklanmasını sağlamak için SA adı verilen yeni bir veri artırma tekniği geliştirilmiş ve literatüre kazandırılmıştır. Böylece CPSA'nın küresel bütünlüğü koruma özelliğinin daha da perçinlenmesi amaçlanmıştır.

SA tekniğinin temel yapısı girdi tensörünün kanallarına yapılan seçimsel bir maskeleme işlemine dayanmaktadır. Literatürdeki pek çok standart mimari yalnızca RGB girdileri kabul ederken bu çalışmada 2.1 bölümünde bahsedildiği üzere, SWAGE-3D mimarisi haricinde önerilen mimarilerde RGBA veri formatı kullanılmıştır. SA işlemi uygulandığında girdi tensörünün renk ve doku bilgisini taşıyan ilk üç kanalı yani RGB kanalları doğrudan

sıfırlanmakta geriye yalnızca nesnenin uzamsal sınırlarını ve doluluk/boşluk bilgisini barındıran şeffaflık kanalı bırakılmaktadır. Bu işlem sonucu girdi karmaşık dokulara sahip bir yapıdan ağır geometrik çıkarım yapmasını oldukça kolaylaştıran saf bir silüet maskesine dönüştürülmektedir.

Mimarinin eğitim aşamasında uygulanan bu işlem, veri setindeki her bir girdiye rastgele %50 ihtimalle uygulanmaktadır. Bu oranın %50 olarak seçilmesi mimari kararlılık için çok önemlidir. Bu işlem eğitim verisinin tamamına uygulanacak olur ise mimarinin standart renkli görüntüleri hiç görmemesine sebep olarak ağır renk/doku ipuçlarından elde edebileceği ikincil bilgiler tamamen yok olur. Bu mimariye yarardan daha çok zarar verir. Bu nedenle eğitim verisinin yarısına uygulanan bu yaklaşım ile mimari bir yandan orijinal görüntülerdeki zengin görsel öznitelikleri öğrenirken, diğer yandan karşılaştığı rastgele silüetler ile nesnenin saf geometrisini öğrenmektedir. Şekil 2.13'te SA tekniğinin örnek üç nesne üzerinde uygulanması sonucu elde edilen sonuç gösterilmiştir.



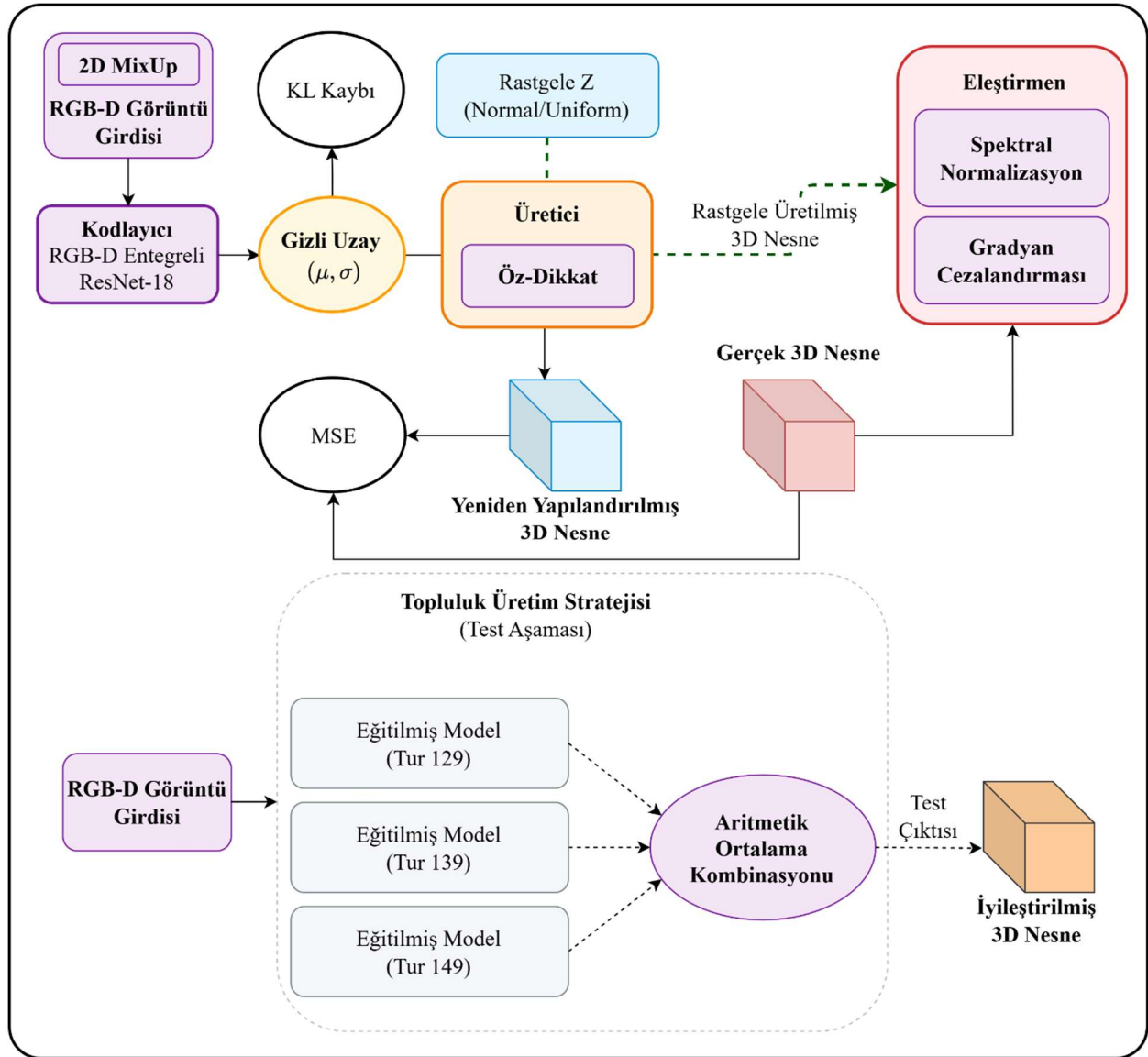
Şekil 2.13. SA tekniğinin örnek girdiler üzerindeki etkisi

Geliştirilen SA tekniği format farklılıkları nedeniyle SWAGE-3D mimarisinin eğitim sürecine dâhil edilmemiştir. Benzer şekilde, AWIF kayıp fonksiyonunun yüzey farkındalığı yeteneklerini dış etkenlerden izole bir şekilde ölçebilmek amacıyla SA uygulanmamıştır. Bu çalışmada SA tekniği özellikle CPSA ile birlikte kullanılmıştır.

2.13. SWAGE-3D Mimarisi

Önceki bölümlerde tek 2D görüntüden 3D nesne üretimi probleminin farklı yönlerine odaklanan iki yöntemsel katkı sunulmuştur. AWIF tarafından voksel tabanlı yeniden yapılandırma yüzey ve sınır voksellerinin daha etkili öğrenilmesi hedeflenmektedir. CPSA tarafından ise frekans düzleminden yararlanarak küresel yapısal temsilin güçlendirilmesi amaçlanmaktadır. Bu bölümde ise bu iki katkıdan farklı olarak girdi temsili ve çekişmeli eğitim kararlılığı üzerine odaklanan ve bütünleşik geliştirmeler içeren SWAGE-3D mimarisi sunulmaktadır.

SWAGE-3D özellikle klasik 3D-VAE-GAN gibi mimarilerin zafiyetlerin analiz edildiği ve buna bağlı olarak tümleşik geliştirmeler içeren tamamlayıcı bir mimari olarak tasarlanmıştır. SWAGE-3D derinlik destekli öznetelik çıkarımı, ResNet-18 gibi gelişmiş bir kodlayıcı sinir ağı, 2D *mixup* veri artırma tekniği, tayfsal düzenlenmiş WGAN öğrenimi, dikkat tabanlı üretici ve kontrol noktası tabanlı topluluk stratejilerini birlikte sunan uçtan uca bütünsel bir eğitim yaklaşımı sunmaktadır. Şekil 2.14'te SWAGE-3D mimarisinin öbek şeması ve Algoritma 2.3'te de SWAGE-3D mimarisinin sözde kodu verilmiştir. Devamındaki alt başlıklarda önerilen bu mimarinin işleyişi açıklanmıştır.



Şekil 2.14. SWAGE-3D mimarisinin öbek şeması

Algoritma 2.3. SWAGE-3D mimarisi sözde kodu

Spektral Wasserstein Dikkat Üretici Topluluk-3D (SWAGE-3D)
Girdi: Eğitim verisi $\mathcal{D}$, yığın boyutu m, eleştirmen iterasyonu n_{critic}, hiperparametreler λ_{GP}, λ_{recon}, β_V Çıktı: Ağırlıklı Topluluk Modeli G_{ens} 1. C ağı ağırlıklarına spektral normalizasyon uygula; 2. Döngü Başlangıcı: $epoch = 1 \dots n_{epochs}$ 3. Döngü Başlangıcı: $\mathcal{D}$ içindeki her yığın $(x_{rgb}, x_{depth}, x_{real})$ için 4. $x_{in} \leftarrow \text{concat}(x_{rgb}, x_{depth})$ 5. // Eleştirmen (C) Optimizasyonu 6. Döngü Başlangıcı: $t = 1 \dots n_{critic}$ 7. $z \sim \mathcal{N}(0, I)$ 8. $\mu, \log \sigma^2 \leftarrow E(x_{in})$ 9. $z_{vae} \leftarrow \mu + \exp\left(\frac{\log \sigma^2}{2}\right) \odot \epsilon_n, \epsilon_n \sim \mathcal{N}(0, I)$ 10. $x_{fake}^{vae} \leftarrow G(z_{vae})$ // Öz dikkat mekanizması içerir. 11. $x_{fake}^{prior} \leftarrow G(z)$ 12. $\mathcal{L}_C \leftarrow C(x_{fake}^{prior}) - C(x_{real}) + \lambda_{GP}(\mathcal{P}_{GP}^{prior})$ 13. $\theta_C \leftarrow \theta_C - \alpha_{LR} \nabla_{\theta_C} \mathcal{L}_C$ 14. Döngü Sonu 15. // Üretici (G) ve Kodlayıcı (E) Optimizasyonu 16. $\mathcal{L}_G \leftarrow -C(x_{fake}^{prior}) + \lambda_{recon} \text{MSE}(x_{fake}^{vae}, x_{real})$ 17. $\mathcal{L}_E \leftarrow \lambda_{recon} \text{MSE}(x_{fake}^{vae}, x_{real}) + \beta_V D_{KL}(\mathcal{N}(\mu, \sigma^2) \mathcal{N}(0, I))$ 18. $\theta_G \leftarrow \theta_G - \alpha_{LR} \nabla_{\theta_G} \mathcal{L}_G$ 19. $\theta_E \leftarrow \theta_E - \alpha_{LR} \nabla_{\theta_E} \mathcal{L}_E$ 20. Döngü Sonu 21. Döngü Sonu 22. // Başarım Odaklı Topluluk Entegrasyonu 23. 129, 139 ve 149. turlardan $\{G_1, G_2, G_3\}$ ve $\{E_1, E_2, E_3\}$ modellerini yükle; 24. Döngü Başlangıcı: Test setindeki her (x_{in}^{test}) yığını için 25. Döngü Başlangıcı: Her bir model $i = 1 \dots 3$ için 26. $\mu_i, \log \sigma_i^2 \leftarrow E_i(x_{in}^{test})$ 27. $x_{fake,i} \leftarrow G_i(\mu_i)$ 28. Döngü Sonu 29. // Voksel tahminlerinin aritmetik ortalamasını alarak normalize et 30. $y_{ens} \leftarrow \frac{1}{3} \sum_{i=1}^3 x_{fake,i}$ 31. Döngü Sonu 32. Döndür y_{ens}

2.13.1. RGB-D kodlayıcı

Literatürde RGB-D verilerinin eşzamanlı yüzey haritalama ve geometrik algıdaki büyük başarısını kanıtlayan öncü çalışmalardan (Newcombe vd., 2011) yola çıkarak; SWAGE-3D mimarisinin ilk aşaması geleneksel 3 kanallı RGB girdi sınırlamasını aşarak mimariye doğrudan geometrik bilgi sağlayacak ve ek öznetelikler öğrenilmesini sağlayacak 4 kanallı RGB-D girdi

alan kodlayıcı yaklaşımıdır. Bu aşama da girdi olarak kullanılan $224 \times 224 \times 3$ boyutlarındaki RGB görüntülerden Depth Anything V2 mimarisi kullanılarak monoküler derinlik haritaları çıkarılmış ve bu haritalar tensöre 4. bir kanal olarak eklenerek $224 \times 224 \times 4$ boyutlarında zenginleştirilmiş bir girdi uzayı oluşturulmuştur.

Öznitelik çıkarımı için ImageNet veri seti üzerinde önceden eğitilmiş ResNet-18 omurgası kullanılmıştır (Deng vd., 2009). ResNet-18 seçimi rastgele değil bu konuda yapılan çalışmalarda hem zamansal açıdan hem de başarımleri açısından elverişli olması nedeniyledir (Serin vd., 2024). Serin ve arkadaşları tarafından yapılan çalışmada KL ıraksama değeri baz alınmış ve 4 farklı ön-eğitilmiş ağ ile tam bağlı katmana sahip CNN ve tam bağlı katmana sahip olmayan CNN kodlayıcı sinir ağında kullanılarak kıyaslanmıştır. Elde edilen ortalama KL ıraksama değerlerine göre (daha düşük KL ıraksama değeri daha düzenli ve sürekli bir gizli uzay anlamına gelir) en başarılı sonucu RegNet16, ondan sonra ResNet-18, daha sonra ise EfficientNetB0 vermiştir. Buna karşın aynı çalışmada zamansal açıdan tam tersi bir sonuç elde edilerek en kısa eğitim süresi EfficientNetB0 ile daha sonra ResNet-18 ile daha sonra ise RegNet16 ile elde edilmiştir. Bu sonuçlara dayanarak en uygun ön-eğitilmiş ağ olarak bu çalışmada ResNet-18 seçilmiş ve hem daha başarılı bir gizli uzay temsili hem de zamansal açıdan kabul edilebilir bir süre hedeflenmiştir. RegNet16 en başarılı sonucu vermesine karşın zamansal açıdan büyük bir dezavantaja sahip olduğu görülmüştür. Standart ResNet-18 mimarisinin ilk evrişim katmanı önerilen yeni 4 kanallı girdiyi kabul edecek şekilde değiştirilmiştir. Transfer öğrenmenin (*transfer learning*) avantajlarını kaybetmemek ve ağıın başarımleri korumak için orijinal 3 RGB kanalının ağırlıkları sabit tutulurken yeni eklenen derinlik kanalının ağırlıkları RGB ağırlıklarının matematiksel ortalaması alınarak başlatılmıştır. Bu durum F. 45'te matematiksel olarak gösterilmiştir. Bu başlatma stratejisi derinlik kanalının öğrenilmiş öznitelik dağılımlarını bozmadan kararlı bir başlangıç yapmasını sağlamaktadır.

$$W_{depth} = \frac{1}{3}(W_R + W_G + W_B) \quad (F. 45)$$

Burada:

W_{dep} : Derinlik kanalının hesaplanan başlangıç ağırlık matrisini,

W_R, W_G, W_B : Önceden eğitilmiş ağıın sırasıyla kırmızı, yeşil ve mavi kanallarına ait ağırlık matrislerini ifade etmektedir.

2.13.2. Üretici omurga ve çekişmeli kararlılaştırma

SWAGE-3D mimarisinin üretici ve ayırt edici ağıları 3D nesne üretimi probleminde

sıklıkla karşılaşılan eğitim kararsızlığı ve gradyan kaybolması problemlerini engellemek amacıyla gradyan cezalandırma destekli WGAN-GP formülasyonu üzerine inşa edilmiştir. Detayları 2.8 bölümünde verildiği üzere eleştirmen ağı 1-Lipschitz süreklilik kısıtını sağlamak için GP kullanılabilmesine karşın yüksek boyutlu voksel uzayında bu durum ağı kapasitesini kısıtlayabilmektedir. Bunu çözmek amacıyla SWAGE-3D mimarisinde karma bir kararlılaştırma yöntemi izlenmiştir. Eleştirmen ağı tüm katmanlarına küresel bir şekilde Lipschitz denetimi sağlayan SN eklenmiştir. Bu durum ağı doğrudan doğasına, istese de bu koşulu esnetme imkânı tanımamaktadır. Bununla birlikte yerel olarak ağ halen bu kısıttan kaçmaya çalışabilmektedir. Olası en uygun yol izlenerek küresel Lipschitz koşulu için SN, yerel Lipschitz koşulu için ise GP kullanılmıştır. Orijinal WGAN-GP mimarisinde GP katsayısı olarak 10 değeri seçilmiştir. Bu değer SN kullanılmadığında uygun bir değer olmasına karşın SN ile birlikte kullanımda ağ gerçek 3D nesnelerin geometrisini öğrenmeyi ve sahte ile gerçeği ayırt etmeyi bırakır. Sadece bu devasa cezadan kaçarak gradyanı her noktada 1'e sabitlemeyi amaçlar. Bu durumda mimari çok basit ve kalitesiz nesneler üretmeye başlamaktadır. Bu durumu önlemek için GP katsayısı olarak SWAGE-3D'de 1 değeri uygulanmış ve aradaki denge sağlanmıştır. Bu ikili kısıtlama yaklaşımı, eleştirmen ağı ifade gücü ile kararlılık arasındaki dengeyi kurarak, modelin asimptotik yakınsamasını güvence altına almaktadır.

2.13.3. Öz dikkat modülü

Geleneksel 3D evrişim katmanları yalnızca yerel alanlar üzerinde işlem yapabilme kapasitesine sahip olduklarından, üretilen voksel uzayındaki uzak mesafe uzamsal bağımlılıkları yakalamada yapısal olarak yetersiz kalmaktadırlar. 3D nesne üretimi probleminde nesnenin küresel tutarlılığını sağlamak hayati bir öneme sahiptir. SWAGE-3D mimarisinde bu yapısal sınırlılığı aşmak ve üretilen hacimsel verilerde yapısal bütünlüğü artırmak amacıyla üretici ağı içerisine bir öz dikkat modülü entegre edilmiştir.

2.11.2 bölümünde teorik altyapısı ve matematiksel hesaplaması detaylandırılan öz dikkat mekanizması, SWAGE-3D mimarisinde üretici ağı rastgele bir konuma değil hesaplama karmaşıklığı ile uzamsal temsil gücü arasında dengenin sağlandığı yerde konumlandırılmıştır. Klasik öz dikkat mekanizmasının bellek ve işlem maliyeti girdi eleman sayısının karesi ile orantılı şekilde artmaktadır. 3D vokselize ortamlarda ise bu maliyet 2D görüntülere kıyasla kübik olarak artar. Öz dikkat modülü üretici ağı 32x32x32 gibi yüksek çözünürlüklü son katmanlarına doğru entegre edilirse ortaya çıkacak devasa dikkat matrisi donanımsal bellek sınırlarını aşabilir ve ağı eğitilebilirliğini tamamen ortadan kaldırabilirdi. Buna karşın ağı uzamsal olarak $1 \times 1 \times 1$ uzamsal boyutlarına sahip ilk katmanlarına uygulanması ise henüz hiçbir geometrik bağlam ve anlam oluşmadığı için öz dikkat mekanizmasının anlamını yitirmesine ve tamamen

işlevsiz bir hale gelmesine sebep olurdu.

Bu kısıtlamaları aşmak için SWAGE-3D mimarisinde öz dikkat modülü üretici ağı tam merkezine, uzamsal çözünürlüğün orta seviyelere ulaştığı ve temel geometrik özniteliklerin belirginleşmeye başladığı üçüncü evrişim bloğunun sonrasına yerleştirilmiştir. Bu mimari tercih sayesinde ağ, bir uçağın iki kanadı veya bir sandalyenin karşılıklı ayakları gibi birbirlerine uzak ancak yapısal olarak simetrik olması gereken bölgeler arasındaki küresel ilişkileri bellek israfına yol açmadan ve hesaplama maliyetini çok arttırmadan en verimli noktada üretme yeteneği kazanmıştır.

SWAGE-3D mimarisinin tasarım aşamasında öz dikkat mekanizması ağı eğitim kararlılığını ve verimliliği korumak amacıyla bilinçli olarak kodlayıcı veya eleştirmen ağlar yerine yalnızca üretici ağda kullanılmıştır. 2.13.1 bölümünde bahsedildiği üzere kodlayıcı sinir ağına ön eğitilmiş ResNet-18 omurgası kullanılmaktadır. ResNet mimarisinin sahip olduğu derin artık bağlantılar 2D girdi görüntülerindeki hiyerarşik ve küresel öznitelikleri çıkarma konusunda oldukça yeteneklidir. Bu yapıya fazladan bir dikkat modülü eklemek transfer öğrenme sürecini bozma ve mimariyi aşırı öğrenme problemine itme gibi riskler barındırır.

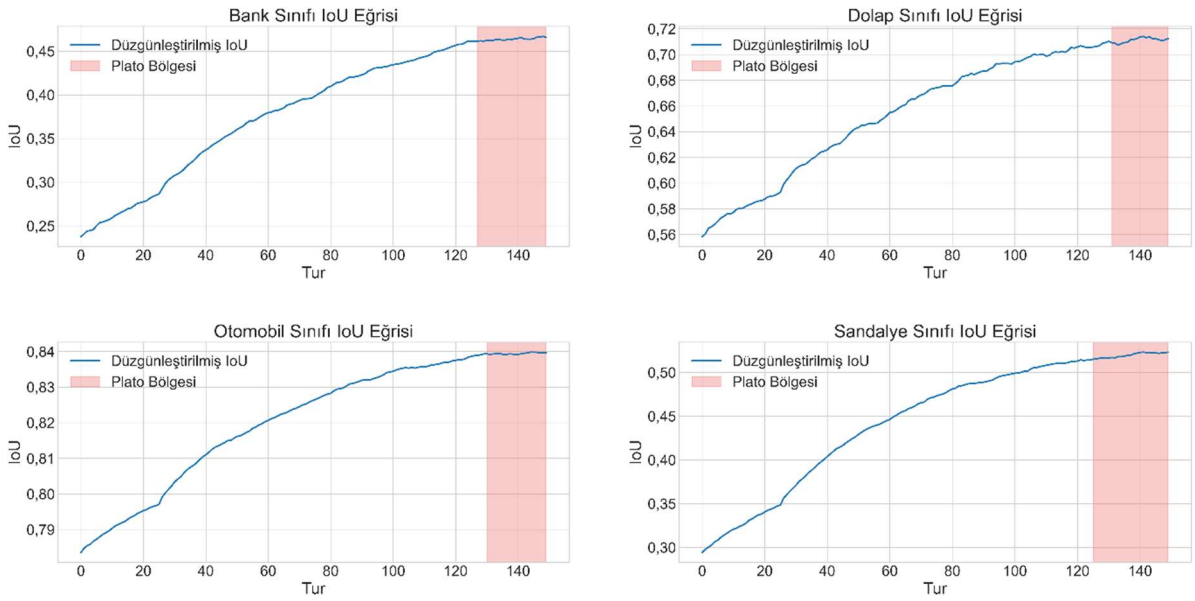
Ayırt edici ağda dikkat modülünün kullanılmaması ise doğrudan Nash dengesini gözetmek içindir (Nash, 1950). SWAGE-3D'nin eleştirmen ağına 1-Lipschitz süreklilik kısıtını sağlamak ve eğitimi kararlaştırmak için 2.13.2 bölümünde bahsedildiği üzere zaten SN uygulanmıştır. Yüksek uzamsal boyuta sahip bir 3D eleştirmen sinir ağına yüksek hesaplama maliyetine sahip bir öz dikkat modülü eklemek yalnızca donanımsal darboğaz oluşturmakla kalmayıp aynı zamanda eleştirmen ağı kapasitesini üreticiye oranla orantısız olarak artırarak tüm mimarinin dengesini bozacaktır. Bu nedenle öz dikkat modülü ağı en çok ihtiyaç duyduğu 3D nesnenin yapısal geometrisini sıfırdan oluşturmakla görevli üretici ağına merkezine konumlandırılmıştır.

2.13.4. Topluluk stratejisi

DL üretici mimarilerinde gradyan dinamiklerine bağlı olarak özellikle eğitim aşamasının sonuna doğru başarımların dalgalanmaları yaşanmaktadır. Çekişmeli eğitim kullanan mimarilerde daha da şiddetlenen bu dalgalanmayı dizginlemek ve sonuçların güvenilirliğini artırmak amacıyla test aşamasında IoU metriğine bağlı olan bir topluluk stratejisi kullanılmıştır.

Eğitim aşaması gözlemlendiğinde, pek çok sınıf için IoU değerinin özellikle 125. turdan sonra kararlılık platosuna geldiği ve artık gelişim göstermediği tespit edilmiştir. Şekil 2.15'te bank, dolap, otomobil ve sandalye sınıfları için 125. tur sonrası meydana gelen bu plato bölgesi kırmızı bir alan olarak gösterilmiştir. Bu şekilden de anlaşılabileceği üzere plato bölgesinde IoU

artışı yavaşlamış ve yatay bir ilerleme göstermeye başlamıştır. Buna bağlı olarak tek son tur değerine güvenmek yerine bu plato bölgesinden seçilen 129, 139 ve 149. turlara ait mimari kontrol noktaları aynı anda yüklenerek bir topluluk oluşturulmuştur (Huang vd., 2017). Bu kontrol noktalarına dayanarak 3D voksel tahmini, her bir mimarinin ürettiği voksel olasılık değerlerinin aritmetiksel ortalaması alınarak F. 46'daki gibi sentezlenmiştir. Aynı 2D görüntü 129, 139 ve 149. turlarda eğitilmiş 3 farklı mimariye (bunların her biri ayrı ayrı kendi kodlayıcı, üretici ve ayırt edici sinir ağlarına sahiptir) verilir. Bu işlem sonucu sistemin elinde aynı nesneye ait 3 farklı 3D voksel tahmini bulunur. Bunların aritmetiksel ortalaması alınarak salınının azaltılması amaçlanmıştır.



Şekil 2.15. Bank, dolap, otomobil ve sandalye sınıfları için eğitimin 125. tur sonrası ulaştığı plato bölgesi

$$\hat{V}(x) = \frac{1}{N} \sum_{i=1}^{N_{TM}} G_i(x) \quad (F. 46)$$

Burada:

$\hat{V}(x)$: Nihai 3D voksel tahminini,

x : 2D girdi verisini,

$G_i(x)$: i. mimarinin ürettiği bağımsız 3D tahmini,

N_{TM} : Topluluk içindeki toplam model sayısını (bu çalışma için 129, 139 ve 149. turlar olmak üzere $N = 3$) ifade etmektedir.

Birleştirme işlemi sonrası elde edilen sürekli olasılık değerleri sabit bir voksel doluluk eşiği olan $\tau = 0,5$ kullanılarak ikili 3D nesnelere dönüştürülür. Bu mekanizma ile tekil

mimarilerin rastgele hataları önlenerek SWAGE-3D mimarisinin genelleştirme kapasitesi artırılmıştır.

2.13.5. Bileşik kayıp fonksiyonları ve eğitim dinamikleri

SWAGE-3D mimarisinin eğitim süreci kodlayıcı, üretici ve eleştirmen ağlarının koordineli bir şekilde güncellendiği bir kayıp yapısına dayanmaktadır. VAE bileşeni mimarinin 2D'den 3D'ye eşleşmeyi öğrenmesini ve gizli uzayı standart normal dağılıma uymaya zorlar. WGAN bileşeni ise üretilen nesnelerin yapısal gerçekçiliğini artırmaya odaklanır. Üretici ağ yalnızca eleştirmen ağı kandırmak üzere değil aynı zamanda orijinal girdinin topolojik yapısına uygun kalmak için hem VAE hem de GAN bileşenlerine birlikte uyum sağlar. Kodlayıcı, üretici ve eleştirmen ağlarının toplam kayıp formülasyonları F. 47, F. 48 ve F. 49'da verilmiştir.

$$\mathcal{L}_E = \lambda_{rec}\mathcal{L}_{rec} + \lambda_{KL}\mathcal{L}_{KL} \quad (\text{F. 47})$$

$$\mathcal{L}_G = \lambda_{rec}\mathcal{L}_{rec} - \mathbb{E}_{z \sim P_z}[C(G(z))] \quad (\text{F. 48})$$

$$\mathcal{L}_{C_{WGAN}} = \mathbb{E}_{z \sim P_z}[C(G(z))] - \mathbb{E}_{x \sim P_{data}}[C(x)] + \lambda_{gp}GP \quad (\text{F. 49})$$

Dinamik bir denge kurmak amacıyla, üretici ağı her bir ağırlık güncellemesine karşılık, eleştirmen ağı 5 kez güncellenmektedir. Yapılan deneysel çalışmalar sonucunda dengeleme parametreleri $\lambda_{rec} = 1,0$ ve $\lambda_{KL} = 1,0$ olarak belirlenmiştir.

3. DENEYSEL KURULUM VE DEĞERLENDİRME

Bu bölümde çalışma kapsamında önerilen yöntemlerin başarımlarını nesnel olarak ölçmek amacıyla kurgulanan deneysel altyapı sunulmaktadır. Yapılan çalışmalar 3 farklı yöntemsel aşamada gerçekleştirilmiş olup her bir aşamanın kendine özgü veri ön işleme adımları, hiper parametreleri ve kıyaslama stratejileri detaylandırılmıştır. Tüm deneylerde ortak olarak kullanılan altyapı ve değerlendirme metrikleri detaylandırılmıştır. AWIF, CPSA ve SWAGE-3D tarafından aynı temel problem olan tek 2D görüntüden buna karşılık gelen 3D nesnenin üretilmesine odaklanılmaktadır. AWIF tarafında özellikle yeniden yapılandırma kayıp fonksiyonlarının etkisi, CPSA tarafında dikkat mekanizmalarının kodlayıcı temsili üzerindeki etkisi ve SWAGE-3D tarafında ise daha zengin öznitelik çıkarımı ile eğitim kararlılığı incelenmiştir. Buna bağlı olarak yapılan deneyler arasında giriş boyutu, eğitim süresi, veri artırma yaklaşımı ve mimari bileşenler açısından farklılıklar bulunmaktadır. Bu farklılıklar yöntemlerin aynı deney düzeneğinde test edilen doğrudan mimari varyantlar değil aynı araştırma problemine farklı teknik açılardan katkı sunan yaklaşımlar olmalarındandır.

3.1. Kullanılan Ortak Altyapı

Bu çalışma kapsamındaki tüm ağ mimarilerinin geliştirilmesi, eğitim ve test süreçleri ortak bir bilgisayar üzerinde gerçekleştirilmiştir. Donanımsal olarak Intel Core i7 – 4790 3,50 GHz Merkezi İşlem Birimi (*Central Processing Unit – CPU*), 2x8 GB (Toplam 16 GB) 1333 MHz Rastgele Erişimli Bellek (*Random Access Memory – RAM*) ve 6 GB VRAM'e sahip NVIDIA GTX 980 Ti FE (*Founders Edition*) GPU kullanılmıştır. Yazılımsal geliştirme ortamı olarak Python 3.11, DL işlemleri için ise PyTorch 2.6 kullanılmıştır (Paszke vd., 2019). Paralel hesaplama ve sinir ağları için CUDA 12.6 altyapısı kullanılmıştır (Nickolls vd., 2008). Kullanılan veri setinin ve veri ön işleme adımlarının detayları bölüm 2.1 ve 2.2'de verilmiştir. Önerilen tüm mimariler için nicel ve nitel değerlendirmeler ilgili metrikler ve ölçümler üzerinden yapılmıştır.

Bu çalışmada sunulan mimarilerin (AWIF, CPSA ve SWAGE-3D) tamamı, literatürdeki çoklu bakış açısına (*multi-view*) dayalı yaklaşımların aksine matematiksel olarak eksik tanımlı (*ill-posed*) bir problem olan tek bakış açılı (*single-view*) 3D Yeniden Yapılandırma gerçekleştirmektedir. Buna ek olarak mimarilerin eğitimi kategori bazlı bir şekilde ayrı ayrı yapılmıştır. Tüm sınıfların tek veri setinde birleştirildiği karma eğitim stratejisi kesinlikle kullanılmamıştır. Hem eğitim hem de test aşamaları tek bakış açılı bir girdi ve tek kategori üzerinden gerçekleştirilmiştir.

Uygulanan bu yöntemsel kısıtlamalar otonom sistemler ve robotik algı gibi eş zamanlı

çoklu görünüm almanın pratik olmadığı gerçek dünya problemleri düşünülerek uygulanmıştır. Literatürdeki çoklu bakış açısına sahip mimariler (Dai vd., 2017; Ji vd., 2017) nesnenin görünmeyen yüzeylerini veriden doğrudan eşleştirme gibi yöntemlere başvurmaktadır. Özellikle çoklu bakış açılı yaklaşımlarda, DL tabanlı uçtan uca derinlik tahmini ile yapılandırılmamış görüntülerden geometrik bilgi çıkarılması da sağlanmıştır (Yao vd., 2018). Bu çalışmada ise mimarinin yalnızca tek yönden elde edilen kısıtlı gözlem altında nesnenin geometrisini ve topolojisini öğrenerek başarılı sonuçlar üretmesi hedeflenmiştir. Bu üretimin yapılabilmesi için gizli uzayda hedefe yönelik önsel bir bilginin olması matematiksel bir zorunluluktur. Farklı topolojilere sahip 13 farklı kategorinin aynı ağda eğitilmesi gizli uzay dağılımı seyrelterek tekil bakış açısından üretim başarımını düşürmektedir. Buna bağlı olarak kategori bazlı eğitim stratejisi benimsenmiştir.

3.2. AWIF Deney Düzenegi ve Analizi

Çalışma kapsamında önerilen AWIF fonksiyonunun başarımı 3D-VAE-GAN mimarisi üzerinde tasarlanan denetimli deneylerle test edilmiştir. Deneyler AWIF'in 3 boyutlu yeniden yapılandırma başarısını standart kayıp fonksiyonları (MSE, BCE, MFL, Dice, IoU) ile eğitilen taban (*baseline*) mimarilerle nesnel bir şekilde kıyaslayacak biçimde kurgulanmıştır.

AWIF kapsamında kullanılan zorluk faktörü literatürdeki geleneksel yaklaşımlardan farklı kurgulanmıştır. Burada α parametresi sınıflara paylaştırılmak yerine ağı genel güven skoruna bağlı bir küresel sönümleyici görevi üstlenmiştir. Bu deneysel tercihteki amaç mimarinin yüksek hata payı ile tahmin ettiği örneklere odaklanmasını sağlamaktır.

AWIF deneyleri 100 tur, 32 mini-yığın boyutu ile gerçekleştirilmiştir. Mimarinin bileşen eğitimlerinde Adam optimizasyon algoritması ($\beta_1 = 0,5, \beta_2 = 0,5$) kullanılmış olup üretici için 0,0025, kodlayıcı için 0,0001 ve ayırt edici için 0,001 sabit öğrenme oranları tercih edilmiştir (Kingma ve Ba, 2015). AWIF'e ait önemli hiper parametreler Tablo 3.1'de listelenmiştir.

Mimarinin yeniden yapılandırma başarımı IoU ve CD metrikleri ile değerlendirilmiştir. Voksel ikili hale getirme eşiği 0,5 alınmış olup, özellikle yüzey hassasiyetinin doğrulanması ve AWIF'in buradaki başarımının ölçülmesi için CD önemli bir metrik olarak değerlendirilmiştir. AWIF bilindiği üzere kendi içerisinde IoU metriğini de barındırmaktadır. Buna bağlı olarak yalnızca IoU metriği kullanılarak AWIF'in başarımı değerlendirilecek olursa literatürde kendi kendini doğrulama (*self validation*) olarak isimlendirilen problem ile karşılaşabilmektedir. Bunun önüne geçmek için ek olarak CD metriğinin kullanılmasının çok önemli olduğu görülür. Ayrıca CD metriği özellikle yüzey başarımını ölçmeyi sağlar ve bu da AWIF'in en temel geliştirilme amaçlarından biri olan yüzey doğruluğunun ölçülmesi için çok sağlam bir temel

oluşturur. Tüm deneylerin sonucunda giriş görüntüsü, gerçek 3D nesne ve üretilen nesne karşılaştırmalı bir şekilde gösterilmiş aynı zamanda zamansal açıdan AWIF'in sonuca etkileri gösterilmiştir.

Tablo 3.1. AWIF hiperparametreleri

Parametre	Değer	Açıklama
Eğitim:		
Tur	100	Eğitimdeki maksimum tur sayısı
Yığın Boyutu	32	Eğitim için mini-yığın boyutu
Öğrenme Oranı (Üretici)	0,0025	Üretici için Adam öğrenme oranı
Öğrenme Oranı (Kodlayıcı)	0,0001	Kodlayıcı için Adam öğrenme oranı
Öğrenme Oranı (Ayırt Edici)	0,001	Ayırt edici için Adam öğrenme oranı
Adam Beta Parametreleri	(0,5; 0,5)	Adam için momentum parametreleri (Sabit)
Ayırt Edici Eşiği	0,8	Ayırt edici için eğitim eşiği
Yumuşak Etiketleme	True	Kararlı eğitim için etiket yumuşatma
Mimari:		
Gizli Uzay Boyutu (z)	200	Gizli uzay vektörünün boyutu
Voksel Çözünürlüğü	(32, 32, 32)	3D voksel ızgarası boyutu
Girdi Görüntü Boyutu	(137, 137)	2D girdi görüntü çözünürlüğü
Leaky ReLU α	0,2	Leaky ReLU için negatif eğim değeri
Gizli Uzay Dağılımı	Normal	Gizli uzay örnekleme için dağılım tipi
Kayıp Fonksiyonu:		
AWIF Dinamik α	0,25 – 1,0	IoU ile ters orantılı simetrik odak çarpanı
AWIF Adaptif γ	2,0 – 4,0	Eğitim süresince artan odaklanma katsayısı
Zorluk Katsayısı (d)	0,0 – 1,0	Eğitimin ilk %30'unda doğrusal artan çarpan
Dice Yumuşatması	1×10^{-6}	Sayısal kararlılık parametresi
AWIF Yüzey Ağırlığı	4,0	Yüzey voksellerine uygulanan ceza katsayısı
Örnek Ağırlık Katsayısı	1,0 – 3,0	IoU'su düşük örneklere uygulanan ağırlık
Değerlendirme:		
IoU Eşiği	0,5	Metrikler için voksel ikili hale getirme eşiği

3.3. CPSA Deney Düzenegi ve Analizi

Bu alt bölümde kodlayıcı ağına entegre edilen CPSA mekanizmasının 3D yeniden yapılandırma başarımı için kurulan altyapı ve değerlendirmeler detaylandırılmıştır. Deneyler CPSA modülünün frekans düzleminde uyguladığı dikkat maskesinin uzamsal doğruluğa etkisini ölçmek amacıyla tasarlanmıştır. CPSA'nın başarısının nesnel bir şekilde değerlendirilmesi için literatürde yaygın olarak kullanılan SE ve CBAM dikkat mekanizmaları ile ve hiçbir dikkat mekanizması içermeyen standart mimari ile kıyaslama yapılmıştır.

CPSA deneyleri 100 tur ve 32 mini-yığın boyutu ile gerçekleştirilmiştir. Mimarinin geometrik formları renk veya dokudan bağımsız olarak öğrenebilme kapasitesini test etmek amacıyla, eğitim aşamasında görüntünün RGB kanallarını sıfırlayarak yalnızca alfa kanalı üzerinden bilgi almayı sağlayan SA veri artırma tekniği %50 olasılıkla uygulanmıştır.

Ağ bileşenlerinin eğitiminde kullanılan Adam optimizasyon algoritmasının parametreleri $\beta_1 = 0,5, \beta_2 = 0,5$ kullanılmış, üretici için 0,0025, kodlayıcı için 0,0001 ve ayırt edici için 0,001 sabit öğrenme oranları tercih edilmiştir. Tablo 3.2’de CPSA için kullanılan hiperparametreler listelenmiştir.

Tablo 3.2. CPSA hiperparametreleri

Parametre	Değer	Açıklama
Eğitim:		
Tur	100	Eğitimdeki maksimum tur sayısı
Yığın Boyutu	32	Eğitim için mini-yığın boyutu
Öğrenme Oranı (Üretici)	0,0025	Üretici için Adam öğrenme oranı
Öğrenme Oranı (Kodlayıcı)	0,0001	Kodlayıcı için Adam öğrenme oranı
Öğrenme Oranı (Ayırt Edici)	0,001	Ayırt edici için Adam öğrenme oranı
Adam Beta Parametreleri	(0,5; 0,5)	Adam için momentum parametreleri (Sabit)
Ayırt Edici Eşiği	0,8	Ayırt edici için eğitim eşiği
Yumuşak Etiketleme	True	Kararlı eğitim için etiket yumuşatma
Veri Artırımı	Silhouette Aug.	%50 olasılıkla RGB kanalı silinmesi
Mimari:		
Gizli Uzay Boyutu (z)	200	Gizli uzay vektörünün boyutu
Voksel Çözünürlüğü	(32, 32, 32)	3D voksel ızgarası boyutu
Girdi Görüntü Boyutu	(137, 137)	2D girdi görüntü çözünürlüğü
Leaky ReLU α	0,2	Leaky ReLU için negatif eğim değeri
Gizli Uzay Dağılımı	Normal	Gizli uzay örnekleme için dağılım tipi
Dikkat Mekanizması	CPSA	Kullanılan dikkat mekanizması
Kayıp Fonksiyonu:		
Yeniden Yapılandırma	MSE	Standart Ortalama Kare Hatası
Değerlendirme:		
IoU Eşiği	0,5	Metrikler için voksel ikili hale getirme eşiği

CPSA’nın başarımı IoU ve CD metrikleri üzerinden değerlendirilmiştir. Voksel ikili hale getirme eşiği olarak 0,5 seçilmiştir. Deney aşamasında, CPSA’nın diğer algoritmalarla kıyasla eğitim süresi ölçülmüş ve elde edilen sonuçlar grafikler üzerinden sunulmuştur. Ayrıca detaylı bir ablasyon çalışması yapılarak SA, CA ve CFAM bileşenlerinin ayrı ayrı sonuçlara etkileri tespit edilmiştir.

3.4. SWAGE-3D Deney Düzeneği ve Analizi

Bu alt bölümde SWAGE-3D mimarisinin başarımlarını analizine yönelik kurgulanan deneysel altyapı detaylandırılmıştır. Deneyler WGAN-GP ve SN entegrasyonu ile eğitilen mimarinin 3D yeniden yapılandırma sürecindeki kararlılığını ve uygulanan topluluk test stratejisinin başarısını tarafsız bir şekilde ölçmeyi hedeflemektedir.

SWAGE-3D mimarisinin eğitimi 150 tur ve 32 mini-yığın boyutu ile gerçekleştirilmiştir. Veri ön işleme adımında kodlayıcı ağa 224x224 çözünürlüğünde RGB-D formatında 4 kanallı

görüntüler girdi olarak verilmiş ve 2D *mixup* veri artırma stratejisi %50 olasılıkla uygulanmıştır.

Ağ optimizasyonunda WGAN dinamiğine uyum için Adam algoritmasının momentum parametreleri ($\beta_1 = 0,5; \beta_2 = 0,9$) değiştirilmiştir. Üretici ve eleştirmen arasındaki oyunu dengelemek amacıyla üretici ağı her bir güncellemesine karşılık eleştirmen ağ 5 kez ($n_{critic} = 5$) güncellenmektedir. SN kullanıldığı için gradyan cezalandırma katsayısı ($\lambda_{GP} = 1$) alınmıştır. Eğitim sürecinde öğrenme oranı zamanlayıcısı kullanılarak 60, 90 ve 120. turlarda öğrenme oranları yarıya ($\gamma = 0,5$) düşürülmüştür. SWAGE-3D mimarisine ait hiper parametreler Tablo 3.3'te listelenmiştir.

Tablo 3.3. SWAGE-3D hiperparametreleri

Parametre	Değer	Açıklama
Eğitim:		
Tur	150	Eğitimdeki maksimum tur sayısı
Yığın Boyutu	32	Eğitim için mini-yığın boyutu
Öğrenme Oranı (Üretici)	0,0025	Üretici için Adam öğrenme oranı
Öğrenme Oranı (Kodlayıcı)	0,0001	Kodlayıcı için Adam öğrenme oranı
Öğrenme Oranı (Ayırt Edici)	0,001	Ayırt edici için Adam öğrenme oranı
Adam Beta Parametreleri	(0,5; 0,9)	Adam için momentum parametreleri (Sabit)
WGAN Eleştirmen İterasyonu	5	1 üretici güncellemesi için eleştirmen adımı
Gradyan Cezalandırması (λ)	1,0	WGAN-GP katsayısı
LR Zamanlayıcısı	MultiStepLR	60, 90 ve 120. turlarda LR düşüşü
Veri Artırımı	Mixup	%50 ihtimalle sadece mixup ($\alpha = 0,2$)
Mimari:		
Gizli Uzay Boyutu (z)	200	Gizli uzay vektörünün boyutu
Voksel Çözünürlüğü	(32, 32, 32)	3D voksel ızgarası boyutu
Girdi Görüntü Boyutu	(4, 224, 224)	4 kanallı RGB-D girdi çözünürlüğü
Kodlayıcı Ağı	Uyarlanmış ResNet-18	4 kanal destekli ResNet-18 ve ön eğitilmiş ağırlıkları
Dikkat Katmanı	Katman 3 sonrası	Öz dikkat bloğu yerleşimi
Spektral Normalizasyon	True	Eleştirmen katmanlarında aktif olan SN
Değerlendirme:		
Topluluk Turları	129, 139, 149	Çoklu model değerlendirme noktaları
IoU Eşiği	0,5	Metrikler için voksel ikili hale getirme eşiği

Mimarinin test aşamasında standart tekil değerlendirmenin ötesine geçilerek topluluk stratejisi kullanılmıştır. 129, 139 ve 149. turlardaki model ağırlıkları kullanılarak ortalama alma ile topluluk modellerinin başarımları ölçülmüştür. SWAGE-3D'nin değerlendirilmesi için IoU, CD ve F-Skoru metrikleri kullanılmıştır. Her bir bileşenin sonuca etkisini gösteren detaylı bir ablasyon çalışması gerçekleştirilmiştir.

3.5. Değerlendirme Metrikleri

Bu alt bölümde özellikle çalışmanın nicel bakımdan kıyaslanması için kullanılan IoU,

CD ve F-Skoru metrikleri detaylı bir şekilde açıklanmış hesaplamalarında kullanılan matematiksel formülasyonlar ve görsel açıdan karşılıkları verilmiştir.

3.5.1. IoU metriği

Literatürde Jaccard İndeksi olarak da bilinen IoU, nesne tespiti ve semantik segmentasyon çalışmalarında kullanılan temel başarımlar değerlendirme metriklerinden biridir (Jaccard, 1912). Bu metrik, temel olarak iki küme veya hacim arasındaki örtüşme miktarını, kesişim alanlarının birleşim alanlarına oranı üzerinden ölçer. IoU değeri 0 ile 1 arasında değişmekte olup 1 değeri mükemmel bir örtüşmeyi, 0 değeri ise hiçbir örtüşmenin olmadığını ifade eder.

Bu çalışmada IoU metriği üç farklı işlevle kullanılmıştır. Bunlar:

Bağımsız Kayıp Fonksiyonu: Geleneksel kayıp fonksiyonlarına alternatif bir optimizasyon hedefi olarak kullanılmıştır.

AWIF Kayıp Bileşeni: Önerilen AWIF kayıp fonksiyonu içerisinde, adaptif ağırlıklandırma ve ceza terimlerinin hesaplanmasında entegre edilmiştir.

Başarım Değerlendirme Metriği: Eğitim ve test aşamalarında, mimarilerin başarısını nesnel olarak ölçmek amacıyla ana metrik olarak kullanılmıştır.

3D vksel yeniden yapılandırma bağlamında IoU, her bir vkselin ikili değeri (0 veya 1) üzerinden hesaplanır. Üretici ağı çıktılarını $[0,1]$ aralığında doluluk olasılıkları üretmektedir. Bu olasılık değerleri, herhangi bir yumuşatma veya yapısal son işlem uygulanmaksızın, sabit bir $\tau = 0,5$ eşik değeri kullanılarak ikili hale getirme işlemine tabi tutulur ve ikili vksel ızgaralarına dönüştürülür. F. 50’de IoU değerinin hacimsel yapılar için matematiksel olarak nasıl hesaplandığı gösterilmiştir.

$$IoU(V^{pred}, V^{gt}) = \frac{|V^{pred} \cap V^{gt}|}{|V^{pred} \cup V^{gt}|} = \frac{\sum_i (V_i^{pred} \cdot V_i^{gt})}{\sum_i (V_i^{pred} + V_i^{gt} - V_i^{pred} \cdot V_i^{gt})} \quad (F. 50)$$

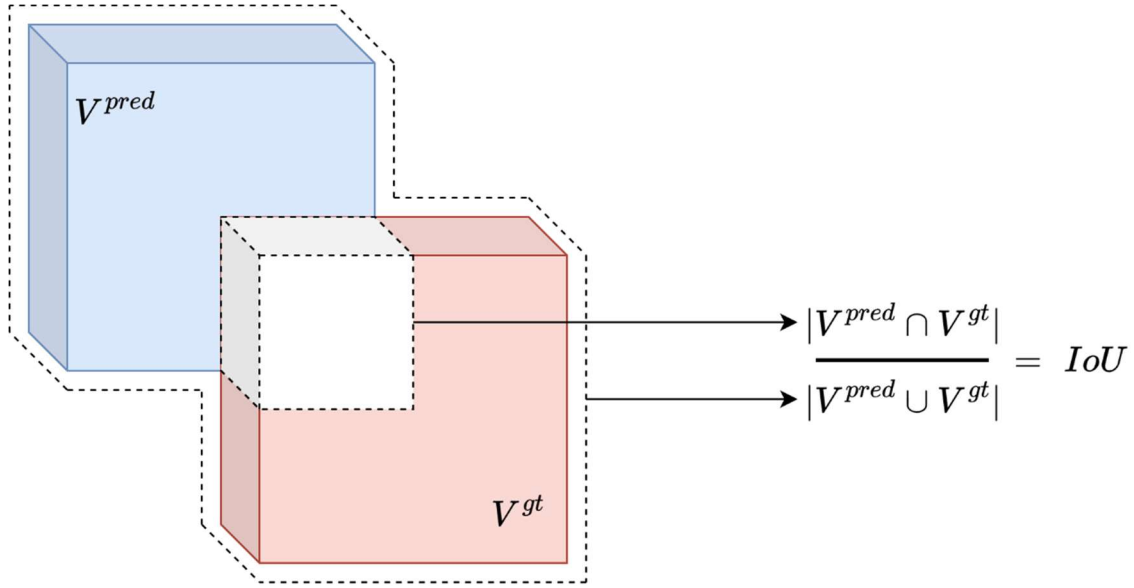
Burada:

V^{pred} : tahmin edilen vkselize nesneyi,

V^{gt} : gerçek vkselize nesneyi ifade etmektedir.

F. 50’de verilen formülasyonda pay kısmı tahmin edilen ve gerçek hacimlerin kesişimini, payda kısmı ise bu iki hacmin birleşimini temsil etmektedir. Sınıflandırma gibi problemlerde genel olarak 0,5 değeri kritik eşik olarak kabul edilir. Bu değer üzerindeki tahminler doğru

pozitif olarak adlandırılırken altında değerler ise yanlış pozitif olarak adlandırılır. Şekil 3.1’de IoU değerinin nasıl hesaplandığı ve görsel olarak 3 boyutlu hacimsel yapılarda nasıl temsil edildiği gösterilmiştir.

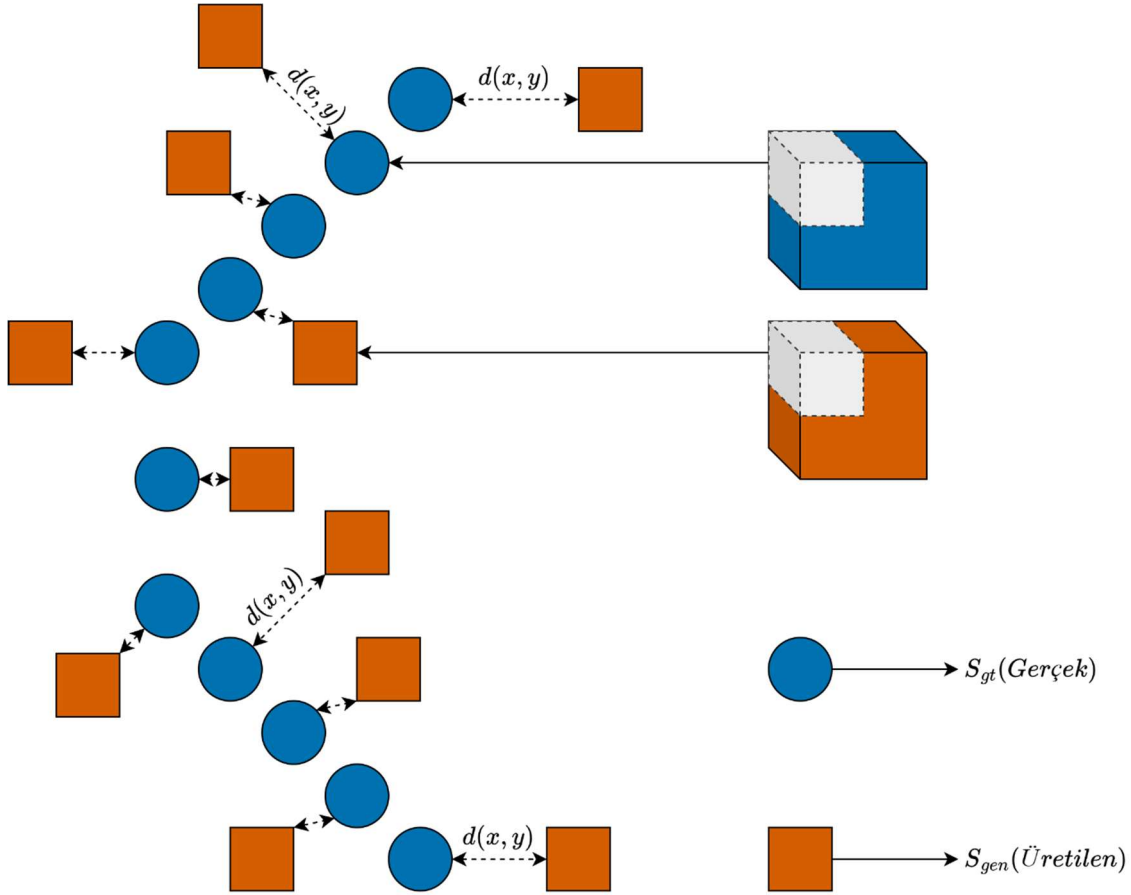


Şekil 3.1. IoU değerinin hacimsel yapılar üzerindeki görsel temsili

3.5.2. CD metriği

IoU hacimsel örtüşmeyi ölçmede etkili bir araç olsa da özellikle seyrek vöksel ızgaralarında yeniden yapılandırılan yüzeylerin geometrik pürüzsüzlüğünü ve yüzeySEL doğruluğunu tam olarak yansıtamayabilir. Tek metrik kullanımının sınırlamalarını aşmak ve daha sağlam bir değerlendirme sunmak amacıyla çalışmaya CD metriği de dahil edilmiştir. IoU’dan farklı olarak CD, üretilen yüzey ile gerçek yüzey üzerindeki noktalar arasındaki ortalama en yakın komşu mesafesini ölçerek geometrik doğruluğu nicelleştirir. Mimarilerin çıktıları hacimsel vöksel ızgaraları formatında olduğundan, CD hesaplaması öncesinde vöksel verilerini nokta bulutlarına dönüştürmek için bir yüzey çıkarma işlemi uygulanmıştır. Bu işlemde, 6’lı komşuluk hiyerarşisi içerisinde en az bir boş komşusu bulunan vökseller yüzey vökseli olarak tanımlanmış ve bu vöksellerin merkez koordinatları yüzey nokta kümelerini oluşturmak üzere dışa aktarılmıştır. Şekil 3.2’de 3D yeniden yapılandırma kalitesinin değerlendirilmesi için kullanılan CD metriği görselleştirilmiştir. Bu şekilde üretilen noktalar (kırmızı küpler) ile gerçek noktalar (mavi küreler) arasındaki çift yönlü en yakın komşu eşleştirmesi gösterilmektedir. Bu metrik yeniden yapılandırılan şeklin hem hassasiyetini (*precision*) hem de tamlığını (*recall*) değerlendirmek üzere her iki kümedeki her bir nokta için mesafelerin karelerinin ortalamasını hesaplar. CD metriğinin matematiksel olarak hesaplanışı F. 51’de verilmiştir (Fan vd., 2017).

$$d_{CD}(S_{gen}, S_{gt}) = \frac{1}{|S_{gen}|} \sum_{x \in S_{gen}} \min_{y \in S_{gt}} \|x - y\|_2^2 + \frac{1}{|S_{gt}|} \sum_{y \in S_{gt}} \min_{x \in S_{gen}} \|y - x\|_2^2 \quad (F. 51)$$



Şekil 3.2. 3D yeniden yapılandırma başarımının değerlendirilmesinde kullanılan CD metriğinin görsel temsili

3.5.3. F-Skoru metriği

IoU ve CD gibi metrikler 3D yeniden yapılandırma mimarilerinin başarımını sırasıyla hacimsel örtüşme ve yüzey sapması açısından değerlendiren literatürdeki en temel metriklerdir. Bununla birlikte nesnelerin yapısal sınırlarında meydana gelen sınıflandırma dengesizliklerinin giderilmesinde yetersiz kalabilirler. Bir vokselle ızgarasında genel olarak boş voksellerin dolu voksellerden daha fazla olduğu bilinir. Bu dengesizliğin altında mimarinin gerçek başarısını ölçmek ve vokselle doluluğuna dayalı analiz için F-Skoru (Knapitsch vd., 2017) metriğine ihtiyaç duyulmaktadır.

F-Skoru hassasiyet (*precision*) ve tamlık (*recall*) olarak isimlendirilen 2 bileşenden meydana gelir. Bu bileşenlerin matematiksel hesabı ise F. 52 ve F. 53'te verildiği şekildedir.

$$Precision = \frac{TP}{TP + FP} \quad (F. 52)$$

$$Recall = \frac{TP}{TP + FN} \quad (F. 53)$$

Burada:

Doğru Pozitif (True Positive – TP): Hem mimari tarafından üretilen hem de gerçek nesnede dolu olan vokselleri ifade eder. Mimarinin tam olarak doğru bildikleridir.

Yanlış Pozitif (False Positive – FP): Mimarinin dolu olarak tahmin ettiği ancak gerçekte boş olan voksellerdir. Bu durum 3D uzayda nesnenin etrafındaki hacimsel taşkınlıkları ifade eder.

Yanlış Negatif (False Negative – FN): Mimarinin boş olarak tahmin ettiği ancak gerçekte dolu olan voksellerdir. Bu durum 3D uzayda nesnenin eksik parçalarını ve üretilmeyen ince detaylarını ifade eder.

Verilen bu ifadelerde özellikle standart sınıflandırma problemleri ve karmaşıklık matrisleri için sıklıkla kullanılan **Doğru Negatif (True Negative – TN)** değerinin olmadığı görülür. 3D nesne yapılandırılması problemlerinde TN değeri gerçekte boş olması gereken bir voksele mimarinin de boş demesine karşılık gelir. Bu durum 3D ortamda felaketle sonuçlanır, çünkü 32^3 çözünürlüğe sahip bir vokselize ortamda 32.768 adet voksel hücresi bulunur. Bu ızgara içine bir nesne yerleştirildiğinde bu ızgaranın çok küçük bir bölümünü kaplayacak ve diğer voksel hücrelerinin boş kalması gerecektir. Burada eğer mimari hiçbir şey öğrenemez ve ekrana tamamen boş voksellerden oluşan bir yapı üretirse nesne üretilmediği için TP değeri 0 olur ancak 3D uzayın çok büyük bir bölümü olan boşlukları doğru tahmin ettiği için TN değeri çok yüksek olur ve başarı oranı devasa olarak artar.

Hassasiyet, üretilen voksellerin ne kadarının gerçekten nesneye ait olduğunu ifade etmektedir. Yüksek hassasiyet mimarinin ürettiği hacmin temiz ve gürültüden uzak olmasına karşılık gelirken, düşük hassasiyet nesnenin etrafında dağınık voksellerin bulunduğunu gösterir. Tamlık ise gerçek nesnenin ne kadarlık bir kısmının mimari tarafından başarıyla oluşturulabildiğini ölçer. Yüksek tamlık nesnenin parçalarının başarılı bir şekilde oluşturulmasına karşılık gelirken düşük tamlık mimarinin uç kısımları üretilmediğini gösterir. F-skoru ise birbirleriyle zıt yönlü çalışan bu iki metriğin harmonik ortalaması olarak F. 54'te verildiği şekilde hesaplanır (Knapitsch vd., 2017).

$$F_1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (F. 54)$$

Burada aritmetiksel ortalama yerine harmonik ortalama kullanılması matematiksel bir

zorunluluktur. Eđer mimari 3D uzayın tamamını dolu voksellerle ifade etmeye kalkarsa bu vokseller gerek nesnenin tamamını da kapsayacağı için Tamlık değeri %100'e yani 1 değeriine ulaşır. Ancak bu yapının içerisinde gerek nesnenin oranı ok küçük kalacağı için Hassasiyet neredeyse 0 olacaktır. Harmonik ortalama ile değlerden biri sıfıra yaklaşırsa sonuç aşağı çekilir ve mimarının bu tarz hacimsel oynamalarla metrikleri manipüle etmesi engellenir. Buna baėlı olarak yüksek bir F-Skoru mimarının hem gürültü üretmediğini hem de eksik para bırakmadığını ifade eder.

4. BULGULAR VE TARTIŞMA

Bu bölümde, önerilen AWIF, CPSA ve SWAGE-3D mimarilerinin nicel ve nitel değerlendirmeleri yapılarak, literatürdeki diğer mimariler ile kıyaslanmış ve elde edilen sonuçlar açık bir şekilde paylaşılmıştır. Kullanılan metrikler ve zaman analizleri sayısal ve görsel olarak gösterilmiştir. 3D nesne yeniden yapılandırma problemleri özellikle iki başlık altında nicel ve nitel değerlendirmeler olarak incelenmektedir. Sayısal bir metrik, değer veya ölçüm ile gösterilebilen ve kanıtlanabilen yaklaşımlar nicel olarak ifade edilirken, daha çok öznel ve görsel incelemeye dayanan yaklaşımlar ise nitel olarak ifade edilir. Bu problemlerde görsel çıktılar da özellikle önem arz ettiğinden literatürde de böyle bir değerlendirme yapıldığı görülür.

4.1. Nicel Bulgular

Bu alt bölümde AWIF, CPSA ve SWAGE-3D mimarilerinin nicel değerlendirmeleri gösterilmiştir. AWIF'in IoU ve CD metrikleri temelinde literatürdeki klasik kayıp fonksiyonlarından çoğu durumda daha başarılı olduğu görülmüştür (Fan vd., 2017). Zamansal açıdan ise AWIF'in dezavantajlı olduğu tespit edilmiştir. CPSA ve SA mimarilerinin ise IoU metriği üzerinden yapılan değerlendirmelerde diğer dikkat mekanizmalarına oranla pek çok nesnede daha başarılı sonuçlar ürettiği görülmüştür. Yapılan detaylı ablasyon çalışmaları ile CPSA'nın her bir bileşeninin sonuca etkisi incelenmiş ve olumlu etkileri görülmüştür. Her ne kadar CPSA temel başarımlar metriğinde başarımlar gösterse de kullandığı parametre sayısı, ortalama tur süresi ve kullanılan VRAM bakımından dezavantajlı olduğu görülür. Literatürdeki SOTA mimariler ile 3D-VAE-GAN mimarisine entegre edilen CPSA'lı önerdiğimiz mimarinin kıyaslamalarında, CPSA'nın rekabetçi potansiyeli açık bir şekilde görülmüştür. SWAGE-3D mimarisi IoU metriği üzerinden doğrudan SOTA mimariler ile kıyaslanmıştır. Kimi SOTA mimarilerden daha başarılı olduğu görülürken kimilerinden geri kaldığı da tespit edilmiştir. Yapılan detaylı ablasyon çalışması ile SWAGE-3D'nin her bir bileşeninin etkisi ayrı ayrı incelenmiş ve sonuca etkileri IoU, CD ve F-Skoru metrikleri üzerinden ayrı ayrı gösterilmiştir.

4.1.1. AWIF nicel bulguları

AWIF kayıp fonksiyonunun 3D nesne yeniden yapılandırması için sonuçlara olumlu etkisi açık bir şekilde görülmüştür. Tablo 4.1'de MSE, BCE, MFL, Dice ve IoU kayıp fonksiyonları ile AWIF doğrudan IoU metriği üzerinden 13 farklı sınıf kategorisi ile kıyaslanmıştır. Elde edilen sonuçlara göre AWIF bank, sandalye, ekran, lamba, hoparlör, masa nesnelerinde ve en nihayetinde ortalama IoU değerinde en başarılı sonuçları elde etmiştir. Buna göre IoU metriği için 6 sınıfta AWIF en başarılı sonuca ulaşmıştır. Bununla birlikte dolap, otomobil, kanepe ve telefon olmak üzere 4 sınıfta ise AWIF en başarılı rakibi ile çok yakın bir

seyir halindedir. Uçak, tüfek ve deniz taşıtı olmak üzere 3 sınıfta ise AWIF'in rakiplerine göre bir miktar geride kaldığı görülür.

Tablo 4.1. MSE, BCE, MFL, Dice, IoU ve AWIF kayıp fonksiyonlarının 13 farklı sınıf için elde ettiği IoU değerlerinin kıyaslanması.

Kategori	MSE	BCE	MFL	Dice	IoU	AWIF
Uçak	0,564265	0,578588	0,549472	0,590513	0,594328	0,58835
Bank	0,409332	0,421452	0,395356	0,428001	0,428973	0,43444
Dolap	0,640407	0,642637	0,638125	0,665938	0,667527	0,66599
Otomobil	0,823156	0,823587	0,813017	0,825907	0,825330	0,82350
Sandalye	0,427731	0,434126	0,421850	0,428429	0,438922	0,46132
Ekran	0,441221	0,445252	0,430082	0,444636	0,441140	0,45383
Lamba	0,352328	0,374906	0,341729	0,389625	0,389219	0,40279
Hoparlör	0,598765	0,603136	0,594574	0,616322	0,612769	0,62534
Tüfek	0,549898	0,550343	0,516043	0,581693	0,582201	0,57318
Kanepe	0,623140	0,626587	0,611887	0,635661	0,631351	0,63452
Masa	0,463989	0,479774	0,459438	0,479661	0,480346	0,49079
Telefon	0,661006	0,665704	0,638411	0,603938	0,603873	0,66420
Deniz Taşıtı	0,525574	0,525003	0,499248	0,545811	0,539751	0,53780
Ortalama	0,544678	0,551623	0,531479	0,556626	0,556595	0,56585

Not: En yüksek IoU değerleri kalın punto ile vurgulanmış olup, yüksek değerler daha iyi sonuçlara işaret etmektedir,

Tablo 4.2’de AWIF kayıp fonksiyonunun MSE, BCE, MFL, Dice ve IoU kayıp fonksiyonları ile CD metriği üzerinden kıyaslanması gösterilmiştir. CD metriğinde elde edilen sonuçların IoU’dan daha başarılı olduğu görülür. Bu durum AWIF’in özellikle yüzey detaylarını oluşturmada ne kadar başarılı olduğunu çok açık bir şekilde gösterir, çünkü CD metriği doğrudan bunu ifade etmektedir. Tablo incelenecek olursa uçak, bank, dolap, sandalye, ekran, lamba, hoparlör, masa ve deniz taşıtı olmak üzere 13 sınıfın toplam 9 tanesinde AWIF’in en başarılı sonucu elde ettiği ve dolayısıyla ortalamada da en iyi CD değerine ulaştığı görülür. IoU metrik sonuçlarına benzer şekilde CD metriğinde de AWIF otomobil, tüfek ve kanepe nesnelerinde nispeten geri kalmakla birlikte çok daha rekabetçi olduğu açıktır. Bu sonuçlar AWIF’in kendi içerisinde doğrudan IoU’yu barındırması bakımından ayrıca önemlidir çünkü kendi kendine doğrulama olmadan da AWIF’in çok başarılı olduğu açıktır. Hatta AWIF tarafından CD metriğinde IoU metriğinden bile daha başarılı sonuçlar elde edilmiştir. Ayrıca AWIF’in temel öneri hedefi olan yüzey voksellerinin başarılı üretimi için de CD metriği doğrudan temel değerlendirme kriteri olarak ifade edilebilmektedir.

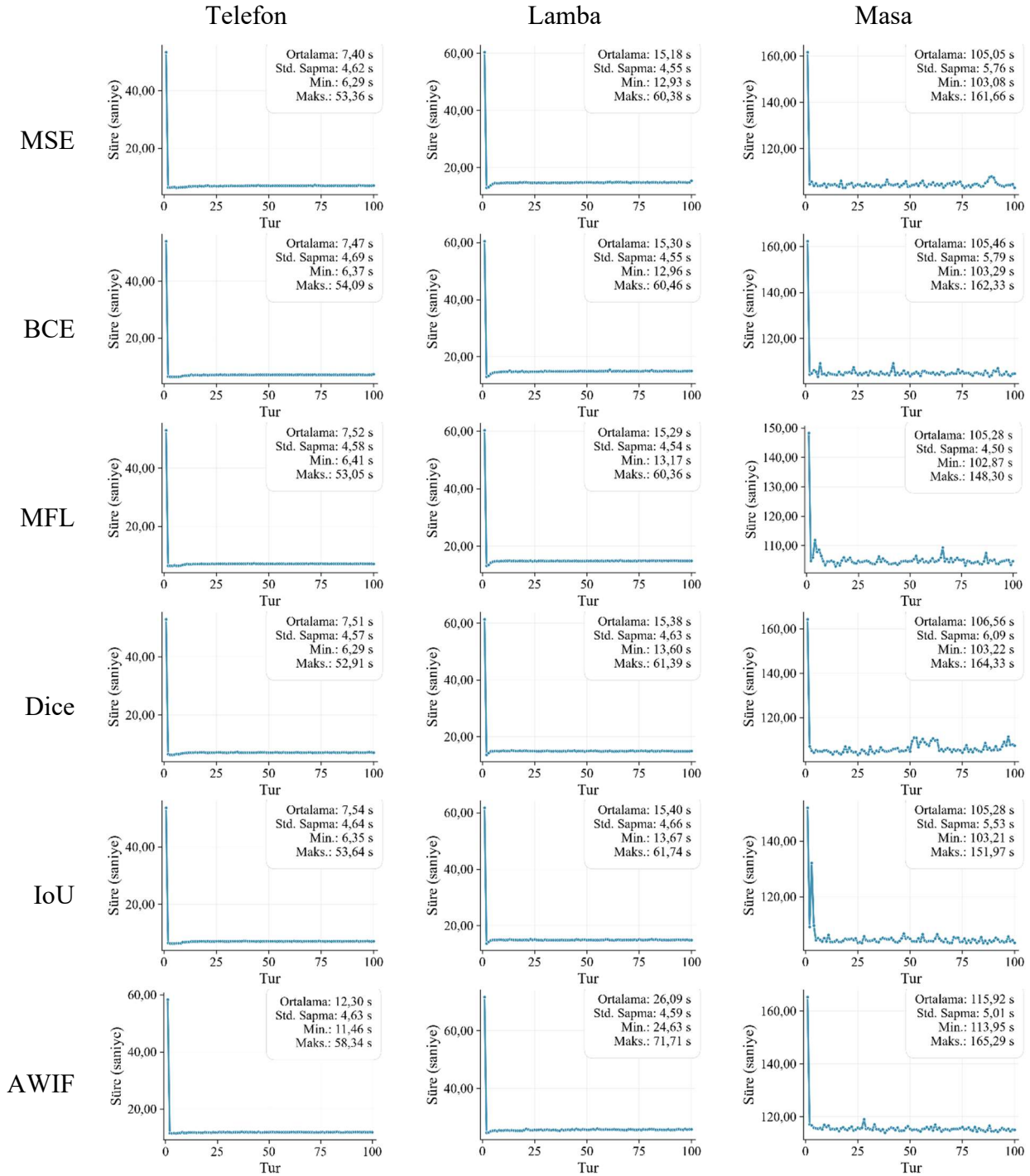
Tablo 4.2. MSE, BCE, MFL, Dice, IoU ve AWIF'in 13 farklı sınıf için elde ettiği CD değerlerinin kıyaslanması

Kategori	MSE	BCE	MFL	Dice	IoU	AWIF
Uçak	0,984244	0,916718	1,021677	0,904728	0,891985	0,860143
Bank	1,728554	1,637221	1,873113	1,763475	1,764908	1,625989
Dolap	2,515144	2,510579	2,537321	2,480013	2,482540	2,474661
Otomobil	1,512389	1,515267	1,515064	1,508156	1,503972	1,522696
Sandalye	2,459823	2,355357	2,595006	2,642065	2,590358	2,310026
Ekran	2,330144	2,273204	2,384319	2,362561	2,411664	2,271689
Lamba	3,409435	2,784925	3,707430	3,452508	3,778220	2,535157
Hoparlör	2,934796	2,882219	2,969982	2,945704	2,942546	2,812236
Tüfek	0,876811	0,842049	0,965724	0,763547	0,761422	0,777484
Kanepe	2,056374	2,021194	2,090775	2,052680	2,036030	2,039920
Masa	2,373276	2,233184	2,496486	2,495317	2,478882	2,169657
Telefon	1,492048	1,417542	1,572306	1,887291	1,893260	1,480254
Deniz Taşıtı	1,251947	1,205912	1,380930	1,197292	1,222094	1,150510
Ortalama	1,994230	1,891952	2,085395	2,035026	2,058299	1,848494

Not: En düşük CD değerleri kalın punto ile vurgulanmış olup, düşük değerler daha iyi sonuçlara işaret etmektedir,

Şekil 4.1’de, veri setinde en çok bulunan masa, ortalama sayıda bulunan lamba ve en az sayıda bulunan telefon nesneleri üzerinden kayıp fonksiyonlarının saniye cinsinden ne kadar sürede tamamlandıkları gösterilmiştir. 100 tur yapılan eğitimde saniyenin ortalama, standart sapma, minimum ve maksimum değerleri şekilde açık bir şekilde raporlanmıştır. Şekil açıkça gösterir ki MSE, BCE, MFL, Dice ve IoU kayıp fonksiyonlarında örnek sayısına göre süre artmakla birlikte ortalamada çok yakın değerlerde eğitim tamamlanmıştır. Bununla birlikte AWIF’in zaman açısından bir dezavantaj oluşturduğu görülür. Bu durumun AWIF içerisinde bulunan özellikle yüzey voksellerini hesaplama ve ağırlıklandırma işleminden kaynaklandığı öngörülmektedir. AWIF kayıp fonksiyonu ortalama olarak her bir tur için telefon sınıfında yaklaşık 4,8, lamba sınıfında yaklaşık 10,7 ve masa sınıfında ise yaklaşık olarak 10,6 saniye daha uzun sürmektedir. Şu açıktır ki bu değerler tur bazında düşük görünmekle birlikte 100 turluk bir eğitimde gerçekten ciddi süre artışlarına karşılık gelir. İlk turun istisnasız bir şekilde diğer turlara göre devasa bir farkla tüm kayıp fonksiyonları için uzun sürdüğü görülmektedir. Bunun sebebi olarak veri ön bellekleme, GPU ve kütüphane ısınması, kaynak tahsisi, veri seti hazırlığı gibi durumlar ön görülmekle birlikte özellikle *cudnn* kütüphanesinin *benchmark* modunun aktif edilmesi bunun en büyük nedeni olarak ön görülmektedir (Chetlur vd., 2014). Bu işlem başlangıç aşamasında, önerilen mimari ve girdi boyutlarına göre pek çok evrişim algoritmasını tek tek dener en hızlı olanı bularak sonraki aşamalarda bu evrişim işlemi ile devam eder. Bu deneme aşaması ilk turların uzun sürmesine neden olmaktadır. Ek olarak masa nesnesi için tüm kayıp fonksiyonlarında süre bakımından bir dalgalanma görülür. Bu durumun masa nesnesinin veri setinde en çok bulunan nesne olması sonucu eğitimin uzun sürmesi ve buna bağlı olarak da

sistem gürültüsü, ısıl darboğaz ve RAM veri akış hızının yavaşlaması gibi durumların sebep olduğu düşünülmektedir.



Şekil 4.1. Farklı kayıp fonksiyonları için 100 turluk eğitim sürecinde tur başına düşen işlem sürelerinin karşılaştırılması

Elde edilen sonuçlara göre AWIF'in başarımlı ve üretim kalitesi bakımından diğer kayıp fonksiyonlarından daha başarılı olduğu çok açık bir şekilde görülmektedir. IoU ve CD metriklerindeki pozitif değerler bunu doğrudan kanıtlar niteliktedir. Bununla birlikte özellikle zamansal açıdan AWIF'in ek bir yük getirdiği açıktır. Buna bağlı olarak zaman kısıtı olan

problemlerde AWIF tercih edilmesinin çok doğru olmayacağı görülür ancak önemli olan daha iyi bir sonuç elde etmekse AWIF'in başarısı ortadadır. Bu durumda probleme özgü tercih yapılması tavsiye edilir.

4.1.2. CPSA nicel bulguları

CPSA dikkat mekanizması ve SA veri artırma tekniği entegre edilen 3D-VAE-GAN mimarisinin nicel bulguları bu kısımda verilmiştir. Tablo 4.3'te 13 farklı sınıf için IoU metriği üzerinden detaylı bir kıyaslama yapılmıştır. Burada:

Temel Mimari: Hiçbir ekleme içermeyen orijinal 3D-VAE-GAN mimarisini,

Temel Mimari + SA: Orijinal 3D-VAE-GAN mimarisi üzerine yalnızca SA veri artırma tekniğinin eklendiği mimariyi,

SE+SA: Orijinal 3D-VAE-GAN mimarisi üzerine SE dikkat mekanizmasının ve SA veri artırma tekniğinin birlikte eklendiği mimariyi,

CBAM+SA: Orijinal 3D-VAE-GAN mimarisi üzerine CBAM dikkat mekanizmasının ve SA veri artırma tekniğinin birlikte eklendiği mimariyi,

CPSA + SA (Önerilen): Orijinal 3D-VAE-GAN mimarisi üzerine CPSA dikkat mekanizmasının ve SA veri artırma tekniğinin birlikte eklendiği mimariyi ifade etmektedir.

Elde edilen metrik değerler incelenecek olursa CPSA+SA mimarisinin açık bir şekilde en başarılı sonuçları elde ettiği görülür. Lamba ve tüfek sınıfları hariç toplam 13 sınıfın 11'inde CPSA+SA mimarisi en başarılı sonuçları elde etmiştir. Bu durum frekans düzleminde uygulanan dikkat mekanizmasının 3D nesne yeniden yapılandırma problemlerindeki başarısını açık bir şekilde gösterir. Ortalama olarak da CPSA+SA mimarisi en başarılı sonucu vermiştir. Adil bir kıyaslama için diğer mimarilere de SA veri artırma tekniği eklenmesine karşın CPSA'nın başarısı açıktır. Ek olarak lamba ve tüfek nesneleri için de CPSA+SA mimarisinin çok rekabetçi olduğu görülür, aradaki fark çok minimaldir. Bu durum önerilen mimarinin başarımını doğrudan kanıtlar.

Tablo 4.3. Temel Mimari ve farklı dikkat mekanizmalarının (SE, CBAM, CPSA) 13 sınıf üzerindeki IoU başarımlarının SA ile birlikte karşılaştırılması

Kategori	Temel Mimari	Temel Mimari+SA	SE+SA	CBAM+SA	CPSA+SA (Önerilen)
Uçak	0,575980	0,582330	0,589495	0,589028	0,611608
Bank	0,416848	0,423741	0,422960	0,417870	0,428522
Dolap	0,641112	0,637036	0,630690	0,665157	0,672622

Tablo 4.3. Devamı

Otomobil	0,817119	0,819607	0,819141	0,801314	0,820698
Sandalye	0,458249	0,465507	0,446671	0,466675	0,468890
Ekran	0,418954	0,425813	0,429476	0,433793	0,468800
Lamba	0,381543	0,400790	0,413315	0,418204	0,415884
Hoparlör	0,605094	0,616506	0,613985	0,621789	0,626232
Tüfek	0,551763	0,547947	0,555059	0,536750	0,550325
Kanepe	0,632300	0,641757	0,632332	0,630248	0,644645
Masa	0,522163	0,518084	0,513315	0,518846	0,535415
Telefon	0,650187	0,653251	0,675090	0,683888	0,688977
Deniz Taşıtı	0,531614	0,534351	0,540153	0,546879	0,561956
Ortalama	0,554071	0,558978	0,560129	0,563880	0,576506

Not: En yüksek IoU değerleri kalın punto ile vurgulanmış olup, yüksek değerler daha iyi sonuçlara işaret etmektedir.

Tablo 4.4'te Temel Mimari ve CPSA+SA mimarisi topoloji bakımından karmaşık olan uçak, ortalama karmaşık olan ekran ve az karmaşık olan masa sınıfları üzerinden farklı tohum değerleri ile kıyaslanmış ve mimarinin kararlılığı test edilmiştir. Elde edilen sonuçlara göre uçak ve masa sınıfı için CPSA+SA mimarisinin kararlılığı temel mimariye göre çok daha iyidir. Bununla birlikte ekran sınıfı için temel mimarinin daha kararlı olduğu görülür. Bu 3 sınıf için eğitim süresi incelendiğinde ise CPSA+SA mimarisinin eğitim süresini artırdığı açıktır.

Tablo 4.4. Temel Mimari ile önerilen mimarinin çoklu tohum üzerinden istatistiksel başarımlar değerlendirilmesi ve eğitim kararlılığı karşılaştırması

Kategori	Model	Ortalama IoU (ortalama $\pm$ std. s)	Eğitim Süresi (saniye/tur)
Uçak	Temel Mimari	0,6072 $\pm$ 0,0084	27,62 $\pm$ 0,03
	CPSA+SA (Önerilen)	0,6308 $\pm$ 0,0040	49,40 $\pm$ 0,01
Ekran	Temel Mimari	0,4242 $\pm$ 0,0104	8,25 $\pm$ 0,01
	CPSA+SA (Önerilen)	0,4326 $\pm$ 0,0130	14,25 $\pm$ 0,02
Masa	Temel Mimari	0,5064 $\pm$ 0,0046	60,92 $\pm$ 0,05
	CPSA+SA (Önerilen)	0,5270 $\pm$ 0,0023	107,22 $\pm$ 0,04

Not: Daha iyi değerler kalın punto ile vurgulanmıştır,

Tablo 4.5'te hesaplama karmaşıklığı, verimlilik ve donanım kaynak kullanımının değerlendirilmesi için ilgili değerler sunulmuştur. Tablodan görülebileceği üzere temel mimari, SE+SA ve CBAM+SA mimarilerinin parametre sayıları birbirine yakın olmakla birlikte en düşük parametreye sahip olan temel mimaridir. Bununla birlikte önerdiğimiz CPSA+SA mimarisinin içerisinde karmaşık işlemler ve bileşenler bulunmasına bağlı olarak parametre sayısı diğerlerinin 2 katından fazladır. Uçak, ekran ve masa sınıflarının tamamında bu durum görülür. Ortalama tur süreleri ise bu analiz içinde Tablo 4.4 ile benzer şekilde görülmektedir. Test zamanında ise bir örneğin üretilme süresi verilmiştir. CPSA+SA mimarisinin karmaşıklığına

bağlı olarak sürenin burada da arttığı görülür ve en iyi değeri temel mimari sunmuştur. VRAM kullanımı ise en az temel mimaride olmakla birlikte önerilen mimarinin VRAM kullanımı parametreyle benzer şekilde diğer mimarilerin 2 katından fazladır. Verilen değerler incelendiğinde, önerilen mimarinin parametre sayısı, hesaplama karmaşıklığı ve donanım gereksinimleri açısından temel mimariye kıyasla ek yükler getirdiği görülmektedir. Ancak bu durum salt bir dezavantaj olarak değil, bir başarımlı-maliyet dengesi çerçevesinde değerlendirilmelidir. CPSA+SA mimarisinin küresel yapısal tutarlılığı sağlama yeteneği test (çıkartım) süresinin yaklaşık 11,4 ms'den 18,0 ms'ye çıkmasına ve VRAM kullanımının da yaklaşık 1021 MB'den 2405 MB'ye yükselmesine sebep olmuştur. Dolayısıyla, frekans alanında elde edilen bu başarımlı kazanımı, donanımsal maliyet ve hesaplama süresindeki artış ile doğrudan ilişkilidir.

Tablo 4.5. Uçak, ekran ve masa sınıfları için farklı mimarilerin hesaplama karmaşıklığı, verimlilik ve donanımsal kaynak kullanımı bakımından karşılaştırmalı analizi

Kategori	Mimari	Parametre Sayısı (M)	Ortalama Tur Süresi (s)	Test Süresi (ms/örnek)	Maksimum VRAM (MB)
Uçak	Temel Mimari	22,24	27,66	11,38	1021,3
	SE+SA	22,28	27,85	11,53	1058,6
	CBAM+SA	22,28	28,83	12,23	1092,7
	CPSA+SA (Önerilen)	54,63	49,39	17,91	2404,6
	Temel Mimari	22,24	8,25	11,42	1021,4
Ekran	SE+SA	22,28	8,33	11,63	1056,7
	CBAM+SA	22,28	8,61	12,17	1092,5
	CPSA+SA (Önerilen)	54,63	14,23	17,94	2404,1
	Temel Mimari	22,24	60,97	11,42	1021,8
Masa	SE+SA	22,28	61,65	11,58	1057,1
	CBAM+SA	22,28	63,96	12,36	1092,7
	CPSA+SA (Önerilen)	54,63	107,25	18,05	2405,2

Not: Daha iyi değerler kalın punto ile vurgulanmıştır, Tabloda yer alan birimlerden M: milyon (parametre sayısı), s: saniye, ms: milisaniye ve MB: Megabayt'ı ifade etmektedir,

Tablo 4.6'da CPSA mimarisi için uçak, ekran ve masa sınıfları üzerinde IoU metriğine dayalı detaylı bir ablasyon çalışması sunulmuştur. Burada:

Temel Mimari: SA, CA ve CFAM bileşenlerinin hiçbirinin kullanılmadığı en temel mimariyi,

Temel Mimari + SA: Temel mimari üzerine SA veri artırımı tekniği uygulanmış

mimariyi,

CPSA (SA Olmadan): Temel mimari üzerine CPSA eklendiği ancak SA'nın uygulanmadığı mimariyi,

CA+SA: CFAM'nin kullanılmadığı ancak CA bileşeninin ve SA'nın uygulandığı mimariyi,

CFAM + SA: CA kullanılmadığı ancak CFAM bileşeninin ve SA'nın uygulandığı mimariyi,

CPSA (Sadece Faz) + SA: CFAM bileşeninin kısıtlandığı, frekans maskesinin genliğinin 1'e eşitlenerek ağırlık yalnızca tahmin edilen sanal ve gerçek değerlerin ark tanjantından türetilen faz açıları kullanarak modülasyon uyguladığı ve SA'nın kullanıldığı mimariyi,

CPSA (Sadece genlik) + SA: CFAM bileşeninin faz bilgisinin 0 kabul edilerek iptal edildiği, ağırlık yalnızca modülatörün tahmin ettiği sinyal şiddeti (genlik) üzerinden frekansları filtrelediği ve SA'nın uygulandığı mimariyi,

CPSA + SA (Önerilen): Önerilen tam mimariyi, CA ve kısıtlanmamış CFAM kullanan, tam frekans dikkati ve SA'nın eş zamanlı kullanıldığı mimariyi temsil eder.

Tablodaki sonuçlar incelenecek olursa SA'nın sonuca olumlu etkisi açık bir şekilde görülür. CPSA'nın hiç dahil edilmediği orijinal temel mimari üzerinde dahi SA uçak ve ekran sınıflarında olumlu bir etki sağlamış masa sınıfında ise minimal geride kalmıştır. SA olmadan CPSA uygulandığında ise temel mimariden her 3 sınıf içinde daha başarılı sonuçlar elde edildiği görülür, yalnızca SA uygulanan durum ile kıyaslandığında da uçak ve masa sınıflarında daha iyi sonuçlar elde edildiği ancak ekran sınıfı için daha geride kaldığı görülür. CA ve SA'nın birlikte uygulanması ise sonuçlara olumlu etki etmiştir. CFAM + SA mimarisi ise uçak ve ekran sınıfları için tabloda sunulan en başarılı sonuçları elde etmişlerdir. Bu durumda CA uygulanmadan doğrudan CFAM kullanılmasının da etkili olacağı öngörülebilir. Yalnızca faz açıları kullanarak yapılan modülasyonun ve SA'nın ise temel mimarilere karşı daha başarılı olduğu görülür. Yalnızca genlik üzerinden frekansların filtrelendiği mimaride ise uçak ve ekran sınıfları için bir miktar düşüş, masa sınıfı için ise bir miktar artış görülür. CPSA + SA olarak tam önerilen mimarinin ise uçak ve ekran sınıfları için temel mimarilerden daha başarılı olduğu açıktır bununla birlikte CFAM + SA'dan bir miktar geri kaldığı görülür. Masa sınıfı için ise CPSA + SA en başarılı sonucu elde etmiştir.

Tablo 4.6. Önerilen bileşenlerin IoU metriğine dayalı detaylı ablasyon çalışması

Yapılandırma	SA	CA	CFAM	Uçak	Ekran	Masa
Temel Mimari				0,6188	0,4111	0,5126
Temel Mimari + SA	✓			0,6232	0,4349	0,5097
CPSA (SA olmadan)		✓	✓	0,6256	0,4219	0,5156
CA + SA	✓	✓		0,6291	0,4417	0,5249
CFAM + SA	✓		✓	0,6381	0,4538	0,5248
CPSA (Sadece Faz) + SA	✓	✓	Faz	0,6330	0,4421	0,5164
CPSA (Sadece Genlik) + SA	✓	✓	Genlik	0,6279	0,4416	0,5196
CPSA + SA (Önerilen)	✓	✓	Tam	0,6326	0,4394	0,5255

Not: Daha iyi değerler kalın punto ile vurgulanmıştır,

Tablo 4.7’de literatürdeki 3D-R2N2 (Choy vd., 2016), 3D-VAE-GAN (Wu vd., 2016), DRC (Tulsiani vd., 2017), ONet (Mescheder vd., 2019), AtlasNet (Groueix vd., 2018) ve Pixel2Mesh (Wang vd., 2018) gibi SOTA mimariler ile CPSA + SA mimarisi IoU metriği üzerinden kıyaslanmıştır. Elde edilen sonuçlara göre uçak, otomobil, lamba, masa ve deniz taşıtı olmak üzere 13 sınıfın 5 tanesinde ve ortalama da önerilen mimari en başarılı sonuçları elde etmiştir. Hoparlör sınıfı için 3D-R2N2, tüfek sınıfı için DRC, bank, dolap, sandalye, kanepe ve telefon sınıfları için ONet, ekran sınıfı için se Pixel2Mesh mimarileri en başarılı sonuçları elde etmiştir. Bu sonuçlara göre CPSA + SA’nın rekabetçi olduğu açık bir şekilde görülür.

Tablo 4.7. IoU metriği üzerinden SOTA mimariler ile CPSA+SA’nın kıyaslanması

Kategori	3D-R2N2	3D-VAE-GAN	DRC	ONet	AtlasNet	Pixel2Mesh	CPSA + SA (Önerilen)
Uçak	0,513	0,420	<u>0,571</u>	<u>0,571</u>	0,493	0,420	0,612
Bank	0,421	0,340	<u>0,453</u>	0,485	0,431	0,323	0,429
Dolap	<u>0,716</u>	0,600	0,635	0,733	0,257	0,664	0,673
Otomobil	<u>0,798</u>	0,760	0,755	0,737	0,282	0,552	0,821
Sandalye	0,466	0,360	<u>0,469</u>	0,501	0,328	0,396	<u>0,469</u>
Ekran	0,468	0,400	0,419	<u>0,471</u>	0,457	0,490	0,469
Lamba	0,381	0,320	<u>0,415</u>	0,371	0,261	0,323	0,416
Hoparlör	0,662	0,590	0,609	<u>0,647</u>	0,296	0,599	0,626
Tüfek	0,544	0,540	0,608	0,474	<u>0,573</u>	0,402	0,550
Kanepe	0,628	0,570	0,606	0,680	0,354	0,613	<u>0,645</u>
Masa	<u>0,513</u>	0,330	0,424	0,506	0,301	0,395	0,535
Telefon	0,661	0,680	0,413	0,720	0,543	0,661	<u>0,689</u>
Deniz Taşıtı	0,513	0,480	<u>0,556</u>	0,530	0,355	0,397	0,562
Ortalama	0,560	0,491	0,533	<u>0,571</u>	0,352	0,479	0,577

Not: En yüksek IoU değerleri kalın punto ile, en yüksek ikinci IoU değerleri de altı çizili olarak verilmiştir, Yüksek değerler daha iyi sonuçlara işaret etmektedir,

Önerilen CPSA dikkat mekanizması ve SA veri artırma tekniğinin nicel bulguları genel

olarak değerlendirildiğinde sonuca olumlu etkileri açık bir şekilde görülmektedir. Yapılan ablasyon çalışmaları ile önerilen mimarinin bileşenlerini tek tek incelenmiş ve başarımların net bir şekilde görülmüştür. Bununla birlikte, CPSA'nın getirdiği başarımların doğrudan bir başarımlar-maliyet dengesi oluşturmaktadır. Çıkarım süresinin 18,0 ms seviyelerine ve VRAM kullanımının 2405 MB'ye çıkması bu mimarinin maliyetidir. Bu artışlar bir zafiyet olarak gizlenmemeli mimarinin frekans düzleminde yaptığı karmaşık küresel ilişkileri çözme yeteneğinin (CFAM ve Modülatör ağı gibi) donanımsal karşılığı olarak kabul edilmelidir. Zaman ve bellek kısıtı olmayan, maksimum geometrik doğruluğun ve yapısal bütünlüğün hedeflendiği durumlarda CPSA+SA ideal bir tercih iken, donanım kaynaklarının ve sürenin kritik olduğu gerçek zamanlı senaryolarda bu başarımlar-maliyet dengesi mutlaka göz önünde bulundurulmalıdır. Zamandan bağımsız olarak en doğru ve geometrik açıdan tutarlı, başarılı 3D nesneler üretmek söz konusu olduğunda CPSA+SA'nın tercih edilmesi tavsiye edilir ancak zaman kısıtı söz konusu ise önerilen mimarinin dezavantajlı olduğu görülür. Literatürdeki SOTA mimariler ile yapılan kıyaslama da önerilen mimarinin en başarılı sonuçları elde edememekle birlikte rekabetçi olduğu görülmüştür.

4.1.3. SWAGE-3D nicel bulguları

SWAGE-3D mimarisi Tablo 4.8'de literatürdeki SOTA mimariler ile IoU metriği üzerinden kıyaslanmıştır. TMVNet (Peng vd., 2022) tüm nesneler ve dolayısıyla ortalama olarak da en başarılı sonucu vererek önerdiğimiz mimariden ve diğer mimarilerden çok daha başarılı sonuçlar elde etmiştir. Hoparlör ve kanepe nesnelerinde ise TMVNet ve Pix2Vox-A (Xie vd., 2019) benzer sonuçları elde etmişlerdir. Önerdiğimiz mimari 3D-R2N2 ve ONet ile rekabetçi başarımlar gösterirken, 3DVAEGAN, DRC ve Pixel2Mesh mimarilerinden daha başarılı sonuçlar elde etmiştir. Bununla birlikte TMVNet'in ve Pix2Vox-A'nın açık bir şekilde gerisinde kalmaktadır. Mimarimizin buradaki eksikliğini açık bir şekilde kabul etmekle birlikte SWAGE-3D'nin hem eğitim hem de test aşamasında yalnız ve yalnızca tek açıdan görüntü ile eğitildiği unutulmamalıdır. Adil bir kıyaslama yapabilmek için 3DVAEGAN ve SWAGE-3D özelinde aynı veri seti kullanılmış yalnızca SWAGE-3D için 4. kanal olarak derinlik bilgisi eklenmiştir. Sonuçlar incelendiğinde 3DVAEGAN'ın 4.1.1 ve 4.1.2'de verilen sonuçlardan bir miktar sapma elde ettiği görülür. Bunun sebebi olarak SWAGE-3D için önerilen derinlik destekli ResNet-18 mimarisi için 4 kanallı RGBA görüntülerin RGB olarak 3 kanallı formata çevrilmesi, sonradan derinlik bilgisi eklenmesi ve buna bağlı olarak şeffaflık bilgisinin yok olduğu görülür. 3DVAEGAN için ise orijinal 3 kanallı RGB veri seti kullanılmıştır ve diğer mimarilerde bulunan alfa kanalı bu verilerde doğrudan bulunmamaktadır. Ek olarak, yine SWAGE-3D'nin ResNet-18'e uyumlu olması için girdiler 224x224 formatına getirilmiştir. Çözünürlük artırma işleminin

görüntülerin bulanıklaşmasına ve kenarların bozulmasına sebep olduğu düşünülmektedir. 3DVAEGAN için de aynı veri seti kullanıldığından temel mimari bu bozulma ile başa çıkamamış olabilir, diğer mimarilerde ise herhangi bir boyut değişikliği yapılmadan doğrudan 137x137x4'lük görüntüler kullanılmıştır.

Tablo 4.8. SWAGE-3D ve literatürdeki SOTA mimarilerin IoU metriği üzerinden kıyaslanması

Kategori	3D-R2N2	3DVAEGAN	DRC	ONet	TMVNet	Pixel2Mesh	Pix2Vox-A	SWAGE-3D
Uçak	0,51	0,42	0,57	0,57	0,69	0,42	0,68	0,57
Bank	0,42	0,34	0,45	0,48	0,65	0,32	0,61	0,40
Dolap	0,71	0,60	0,63	0,73	0,85	0,66	0,79	0,72
Otomobil	0,79	0,76	0,75	0,73	0,87	0,55	0,85	0,83
Sandalye	0,46	0,36	0,46	0,50	0,72	0,39	0,56	0,46
Ekran	0,46	0,40	0,41	0,47	0,59	0,49	0,53	0,41
Lamba	0,38	0,32	0,41	0,37	0,53	0,32	0,44	0,42
Hoparlör	0,66	0,59	0,60	0,64	0,71	0,59	0,71	0,69
Tüfek	0,54	0,54	0,60	0,47	0,78	0,40	0,61	0,46
Kanepe	0,62	0,57	0,60	0,68	0,70	0,61	0,70	0,66
Masa	0,51	0,33	0,42	0,50	0,66	0,39	0,60	0,52
Telefon	0,66	0,68	0,41	0,72	0,80	0,66	0,77	0,73
Deniz Taşıtı	0,51	0,48	0,55	0,53	0,68	0,39	0,59	0,54
Ortalama	0,56	0,49	0,53	0,57	0,71	0,48	0,65	0,57

Not: En yüksek IoU değerleri kalın punto ile vurgulanmış olup, yüksek değerler daha iyi sonuçlara işaret etmektedir.

Veri setinden seçilen uçak, bank, otomobil, lamba, hoparlör ve telefon sınıfları üzerinden SWAGE-3D için üç farklı metrik ile detaylı bir ablasyon çalışması yapılmıştır. Sonuçlar Tablo 4.9-Tablo 4.11’de verilmiştir. Ablasyon işleminde aşağıda listelenen mimariler üst üste eklenen bileşenlerden meydana gelerek SWAGE-3D’yi oluşturmaktadır.

- Temel Mimari (RGB):** Doğrudan RGB girdilerle eğitilen 3D-VAE-GAN mimarisini,
- +Derinlik Haritası:** Temel mimari üzerine derinlik haritası eklenen mimariyi,
- +WGAN-GP:** Derinlik haritası üzerine WGAN-GP entegrasyonu içeren mimariyi,
- +Öz Dikkat:** WGAN-GP entegrasyonu üzerine eklenen öz dikkat mekanizması içeren mimariyi,
- +Topluluk:** Öz dikkat eklenen mimari üzerine topluluk değerlendirmesi eklenmiş mimariyi ifade etmektedir.

Tablo 4.9 incelenecek olursa IoU metriği temel alındığında önerilen mimarinin her bir bileşeninin temel mimariden daha iyi başarımlar gösterdiği açık bir şekilde görülür. Sınıf bazlı sonuçlar incelendiğinde, en başarılı sonuçları veren mimarilerin farklılık gösterdiği tespit edilmiştir. Uçak ve otomobil sınıflarında en yüksek başarımlar, derinlik haritası ve WGAN-GP

içeren mimari ile elde edilmiştir. Lamba ve telefon sınıflarının yanı sıra genel ortalama da ise derinlik haritasının girdiye 4. kanal olarak eklendiği mimari öne çıkmaktadır. Bank ve hoparlör sınıflarında da topluluk yaklaşımı haricindeki diğer eklemeleri içeren mimarinin en iyi sonuçları verdiği görülmektedir. Tablodan görüleceği üzere özellikle derinlik kanalı eklenmesi mimarinin başarısını belirgin bir şekilde artırmıştır. Özellikle lamba sınıfında meydana gelen gelişim oldukça dikkat çekicidir. Ortalama değerler incelendiğinde de özellikle derinlik kanalı içeren yapının en başarılı sonucu verdiği görülür.

Tablo 4.9. SWAGE-3D'nin IoU metriği üzerinden ablasyon çalışması

Yapılandırma	Uçak	Bank	Otomobil	Lamba	Hoparlör	Telefon	Ortalama
Temel Mimari (RGB)	0,557	0,380	0,828	0,340	0,677	0,729	0,585
+ Derinlik Haritası	0,570	0,385	0,825	0,431	0,689	0,730	0,605
+ WGAN-GP	0,581	0,387	0,829	0,395	0,689	0,727	0,601
+ Öz Dikkat	0,568	0,392	0,812	0,410	0,691	0,717	0,598
+ Topluluk	0,567	0,392	0,816	0,412	0,688	0,715	0,598

Not: En yüksek IoU değerleri kalın punto ile vurgulanmış olup, yüksek değerler daha iyi sonuçlara işaret etmektedir,

Tablo 4.10 incelenirse F-Skoru metriğinde de IoU metriği ile benzer sonuçlar elde edildiği görülmüştür. Lamba ve telefon sınıfları için ayrıca ortalama olarak da en başarılı sonuç temel mimari üzerine eklenen derinlik haritası ile elde edilmiştir. Buradan da derinlik haritasının sonuca olumlu etkisi açık bir şekilde görülür. Uçak ve otomobil sınıfları için ise derinlik haritası ve üzerine eklenen WGAN-GP entegrasyonu ile en başarılı sonuca ulaşılmıştır. Bank ve hoparlör sınıfları için ise derinlik haritası, WGAN-GP ve üzerine eklenen öz dikkat mekanizması ile en yüksek başarımlar elde edilmiştir. Bununla birlikte otomobil ve telefon sınıfları için doğrudan temel mimarinin en başarılı sonuca ulaştığı görülür ancak ortalama değerler söz konusu olduğunda SWAGE-3D'nin bileşenlerinin etkisi açık bir şekilde temel mimariden üstündür.

Tablo 4.10. SWAGE-3D'nin F-Skoru metriği üzerinden ablasyon çalışması

Yapılandırma	Uçak	Bank	Otomobil	Lamba	Hoparlör	Telefon	Ortalama
Temel Mimari (RGB)	0,715	0,550	0,906	0,506	0,807	0,843	0,721
+ Derinlik Haritası	0,725	0,554	0,904	0,598	0,815	0,843	0,740
+ WGAN-GP	0,734	0,555	0,906	0,563	0,816	0,841	0,736
+ Öz Dikkat	0,724	0,562	0,896	0,580	0,817	0,834	0,735
+ Topluluk	0,723	0,561	0,899	0,579	0,815	0,833	0,735

Not: En yüksek F-Skoru değerleri kalın punto ile vurgulanmış olup, yüksek değerler daha iyi sonuçlara işaret etmektedir,

Tablo 4.11 incelenecek olursa CD metriği temel alındığında SWAGE-3D'nin başarımları

çok daha keskin bir şekilde görülmektedir. Uçak ve lamba sınıfları için ve ayrıca ortalama olarak tüm bileşenleri dahil edilmiş SWAGE-3D mimarisi ile en başarılı sonuca ulaşılmıştır. Bank ve hoparlör sınıfları için ise topluluk bileşeni hariç olmak üzere öz dikkat mekanizmasına kadar eklenen bileşenler ile en başarılı sonuca ulaşıldığı görülür. Telefon sınıfı için ise özellikle derinlik haritası ve WGAN-GP entegrasyonunun başarısı görülmektedir. Otomobil sınıfı için ise doğrudan temel mimari en başarılı sonucu elde etmiştir.

Tablo 4.11. SWAGE-3D'nin CD metriği üzerinden ablasyon çalışması

Yapılandırma	Uçak	Bank	Otomobil	Lamba	Hoparlör	Telefon	Ortalama
Temel Mimari (RGB)	0,955	1,618	0,214	10,015	0,925	0,423	2,358
+ Derinlik Haritası	0,885	1,597	0,219	4,157	0,874	0,407	1,357
+ WGAN-GP	0,874	1,825	0,219	5,841	0,841	0,399	1,667
+ Öz Dikkat	0,845	1,469	0,233	4,252	0,839	0,413	1,342
+ Topluluk	0,826	1,499	0,227	4,051	0,872	0,423	1,316

Not: En düşük CD değerleri kalın punto ile vurgulanmış olup, düşük değerler daha iyi sonuçlara işaret etmektedir,

Elde edilen sonuçlar genel olarak değerlendirildiğinde SWAGE-3D mimarisinin temel mimariden daha başarılı olduğu, ancak SOTA mimarilere karşı belirli dezavantajlarını koruduğu görülmektedir. SWAGE-3D'nin tümleşik yapısı altında yer alan bileşenlerin katkıları incelendiğinde, bu katkıların tüm metriklerde aynı düzeyde ve doğrusal bir artış oluşturmadığı görülmektedir. Ablasyon sonuçları, özellikle derinlik haritalarının mimariye ek geometrik bilgi sağlayarak CD ve F-Skoru metriklerinde başarıyı belirgin biçimde artırdığını göstermektedir. WGAN-GP ve SN bileşenleri ise doğrudan yüksek IoU artışından ziyade çekışmeli eğitimin kararlılığına ve üretilen yapıların geometrik tutarlılığına katkı sağlamaktadır. Benzer şekilde öz dikkat ve topluluk tabanlı çıkarım stratejileri bazı metriklerde sınırlı artışlar üretmekle birlikte, yüzey kalitesi, yapısal bütünlük ve çıkarım kararlılığı açısından tamamlayıcı katkılar sunmaktadır. Bu nedenle SWAGE-3D bileşenlerinin etkisi, tek metrik üzerinden değil; IoU, F-Skoru ve CD metriklerinin birlikte değerlendirilmesiyle yorumlanmalıdır.

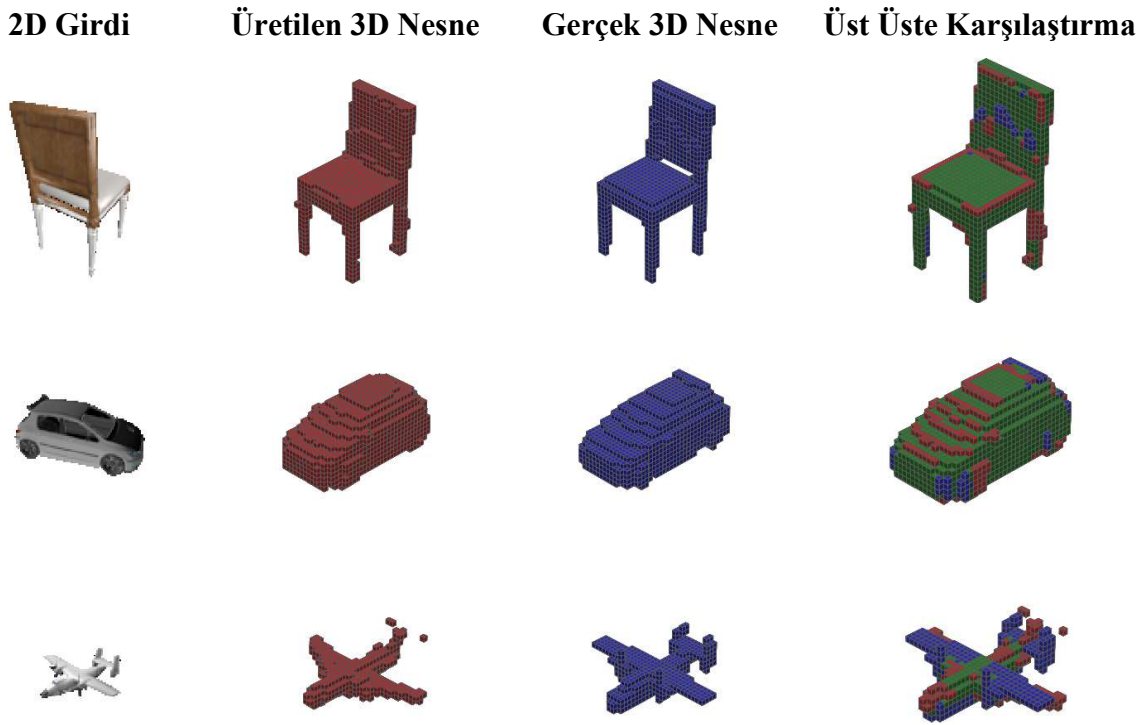
4.2. Nitel Bulgular

Literatürde nitel bulgular daha çok gözleme ve öznel değerlendirmelere dayanır. Görüntü üretme, nesne üretme ve ses üretme gibi problemlerde özellikle tercih edilmektedir. Bu değerlendirmeler her ne kadar öznel olsa da kenarlar, köşeler ve detayların açık bir şekilde görülebileceği açıktır.

4.2.1. AWIF nitel bulguları

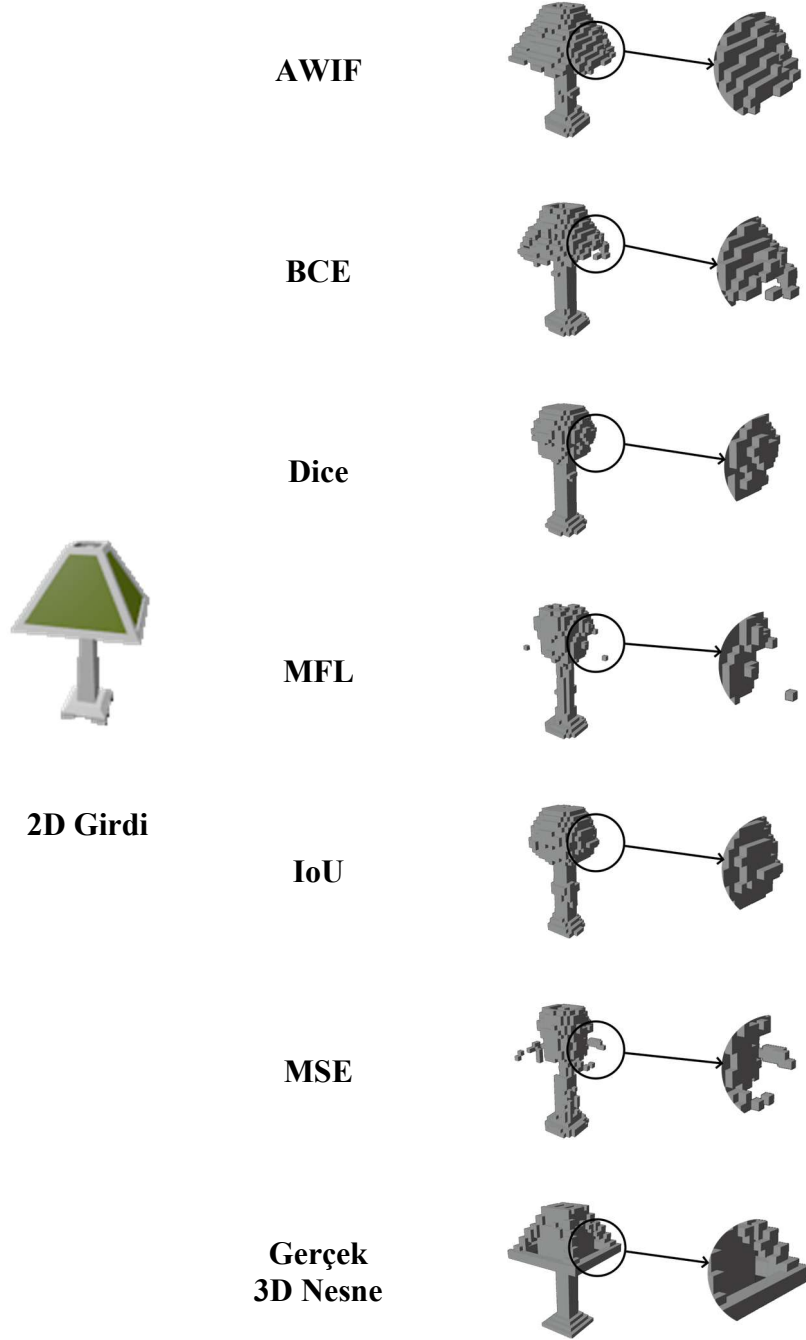
AWIF kayıp fonksiyonunun nitel değerlendirmeleri özellikle kenar, köşe ve yüzey

voksellerinin başarımını kanıtlar niteliktedir. Şekil 4.2’de sandalye, otomobil ve uçak sınıflarının 2D girdileri, bunlara karşılık üretilen 3D nesneler (kırmızı vokseller) ve gerçek 3D nesneler (mavi vokseller) gösterilmiştir. Üst üste karşılaştırma kısmı ise üretilen ve gerçek nesneleri doğrudan merkezde bir araya getirerek yapılan tahminin etkilerini görsel bir şekilde göstermeyi amaçlar. Burada yeşil vokseller doğrudan tam isabet bilinen voksellere karşılık gelir. Adil bir gösterim için sayısal metriklere göre başarılı olunan sandalye, ortalama başarılı olunan otomobil ve az başarılı olunan uçak sınıfları seçilmiştir. Şekiller incelenecek olursa ilgili sandalye nesnesinin AWIF ile yüzey voksellerinin başarılı bir şekilde üretildiği görülür. Otomobil nesnesi için ise genel hatlarıyla başarılı bir üretim yapılmakla birlikte otomobilin arkasında bulunan rüzgarlığın eksik kaldığı görülmektedir. Uçak nesnesinde ise daha çok genel bir uçak dağılımına yönelerek özellikle kanatlar ve kuyruk tarafında AWIF’in hata yaptığı görülmektedir.



Şekil 4.2. AWIF ile 2D girdilerden üretilen 3D nesnelerin, gerçek nesnelerle üst üste karşılaştırması

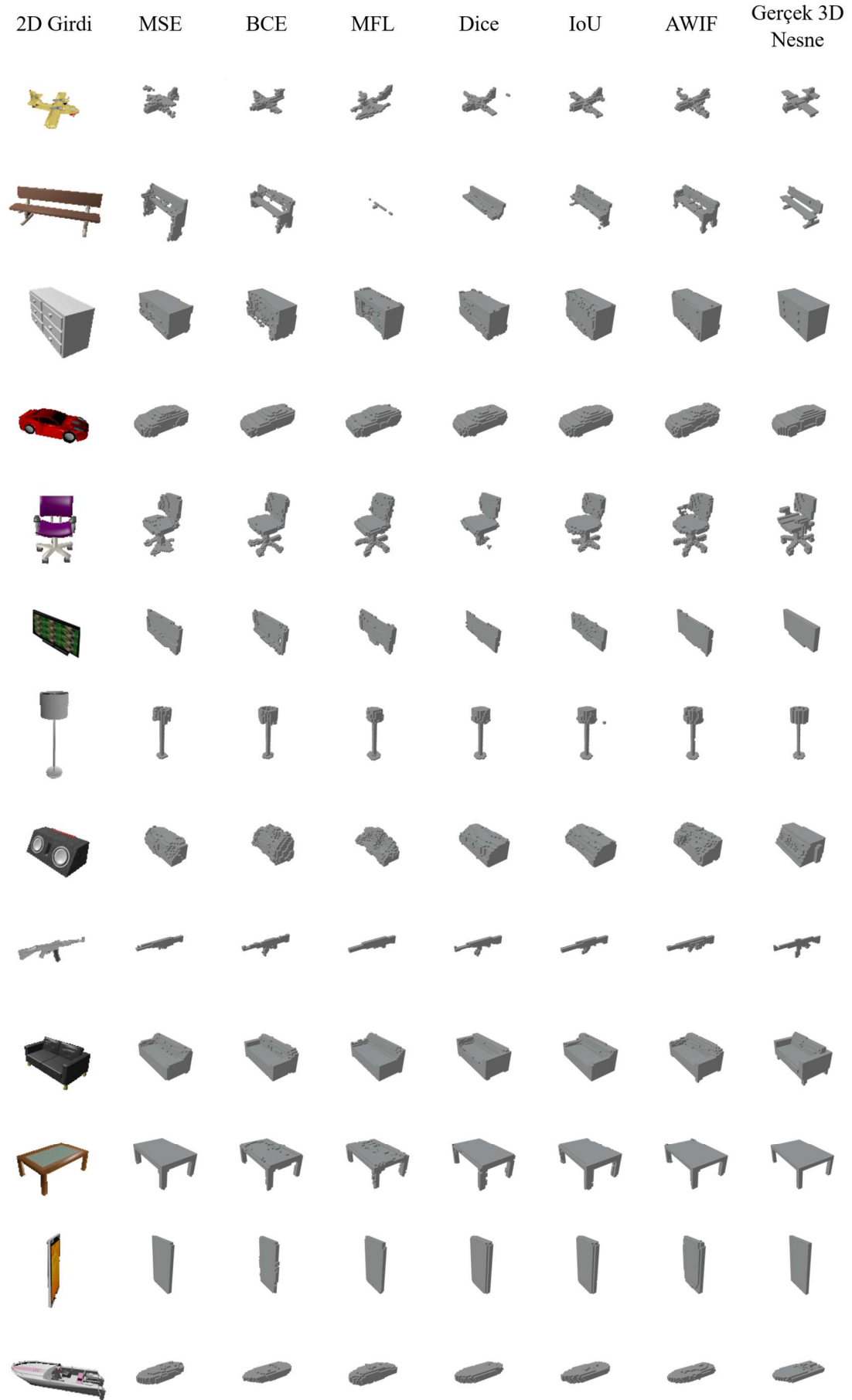
Şekil 4.3’te ise 2D girdiye karşılık özellikle kenar ve yüzey voksellerine odaklanılarak AWIF, BCE, Dice, MFL ve IoU kayıp fonksiyonları ile üretilen 3D nesneler gösterilmiştir. Şekil incelenecek olursa Dice, MFL, IoU ve MSE’nin yüzey voksellerini üretme konusundaki yetersizliği görülür. AWIF ve BCE ile ise yüzey vokselleri diğer kayıp fonksiyonlarına göre daha başarılı bir şekilde oluşturulmuştur. Bunlar arasında AWIF’in daha başarılı olduğu görülür, çünkü uç voksellerde BCE eksik kısımlar bırakırken AWIF bunları da tamamlayabilmiştir. Verilen gerçek 3D nesne ile kıyaslama yapıldığında AWIF’in başarısı görülür.



Şekil 4.3. Lamba sınıfı için seçilen 2D girdi ve ilgili kayıp fonksiyonları tarafından üretilen 3D nesneler

Şekil 4.4'te, çalışmada kullanılan veri setinin 13 sınıfı için seçilen birer örnek gösterilmiştir. Burada 2D girdi, bu 2D girdiye karşılık 3D-VAE-GAN mimarisi temel alınarak kullanılan kayıp fonksiyonları ile üretilen 3D nesne ve ilgili 2D girdiye karşılık gelen ve veri setinde bulunan gerçek 3D nesne gösterilmiştir. Şekil incelenecek olursa, otomobil için Şekil 4.2'de kaçırılan rüzgarlık tarafının bu görselde yalnızca AWIF tarafından başarılı bir şekilde üretildiği görülür. Sandalye için ise her ne kadar iki tutamaç bulunması gerekse de diğer kayıp fonksiyonları tarafından tek tutamacın dahi üretilmediği görülürken, AWIF tarafından bir tarafta bulunan tutamaç başarılı bir şekilde üretilmiştir. Uçak için ise tüm kayıp fonksiyonlarının

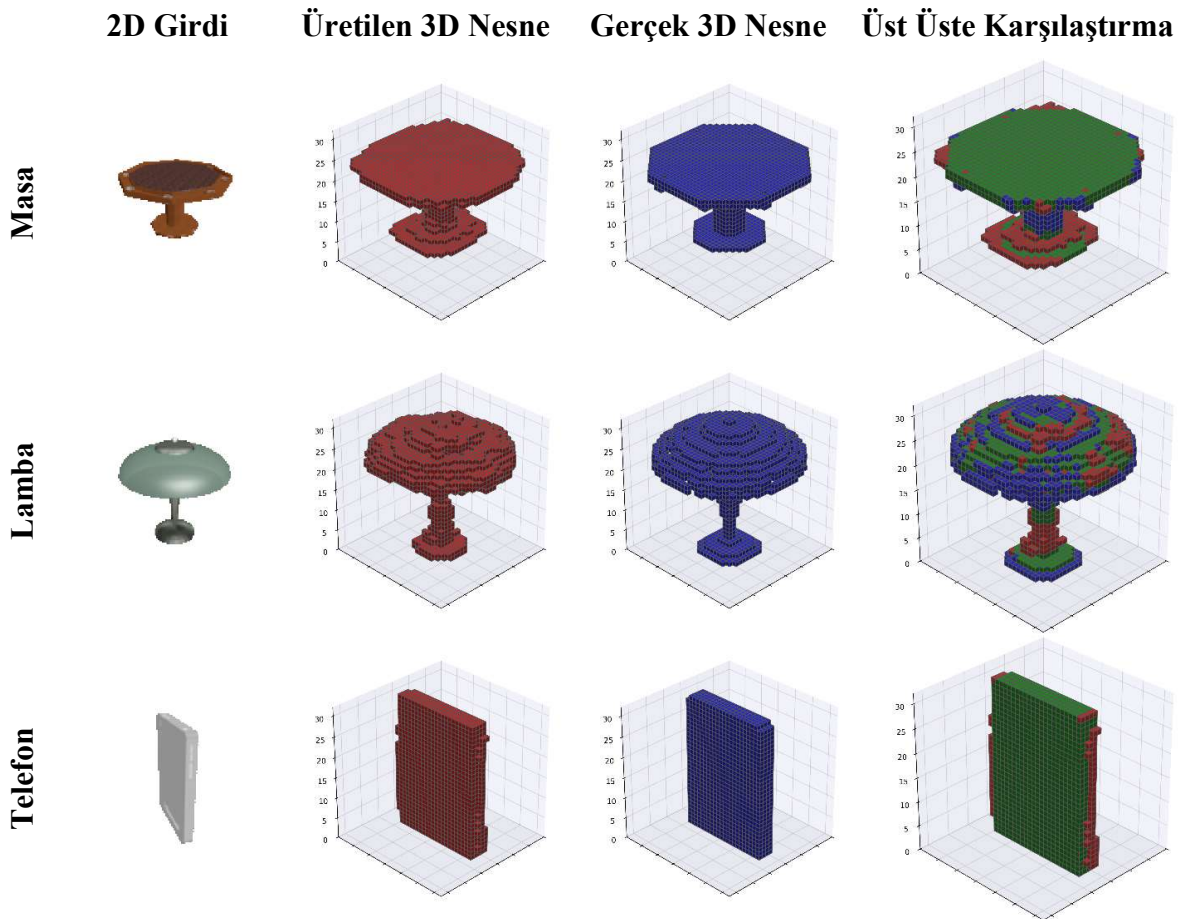
başarısız olduğu görülür, ancak IoU kayıp fonksiyonu kuyruk tarafını nispeten başarılı bir şekilde üretebilmiştir. Bank için özellikle MFL kayıp fonksiyonu ve yine tüm kayıp fonksiyonlarının pek çok detayı kaçırdığı görülmekle birlikte en azından ana geometrik formun AWIF tarafından üretildiği söylenebilir. Dolap için BCE ve MFL kayıplarının yetersiz olduğu görülmekle birlikte yüzey pürüzsüzlüğü konusunda AWIF başarılıdır. Otomobil için tüm kayıp fonksiyonları ana iskeleti yakalamayı başarmıştır. Ekran için diğer kayıp fonksiyonları detayları kaçırırken AWIF'in stand kısmını ve üst kısmı diğerlerine kıyasla üretebildiği görülür. Lamba için kenarlarda bulunan detaylar AWIF ve BCE tarafından yakalanmıştır, bununla birlikte bu nesne için 2D girdide kenarların pürüzsüz olduğu görülür ancak üretilmesi istenen asıl 3D nesnenin kenarlıkları vardır. Bu nedenle bu örnek bir uç değer olarak değerlendirilebilir. Hoparlör için Dice ve IoU kayıp fonksiyonlarının ana formu oluşturabildiği görülür. Tüfek için ön tutamaç kısmı BCE, Dice ve IoU tarafından oluşturulmakla birlikte özellikle arka tutamaç kısmının genelde kaçırıldığı BCE tarafından az miktar oluşturulabildiği görülür. Kanepe için tüm kayıp fonksiyonları tarafından ana formun oluşturulabildiği görülür. Masa için BCE ve MFL tarafından pürüzsüzlüğün bozulduğu ancak diğer kayıp fonksiyonlarının nispeten başarılı üretimler gerçekleştirdiği görülür. Telefon için tüm kayıp fonksiyonları tarafından ana form korunmuştur. Deniz taşıtı için ise genel formun oluşturulduğu görülmekle birlikte MFL'nin eklemeler yaptığı görülür.

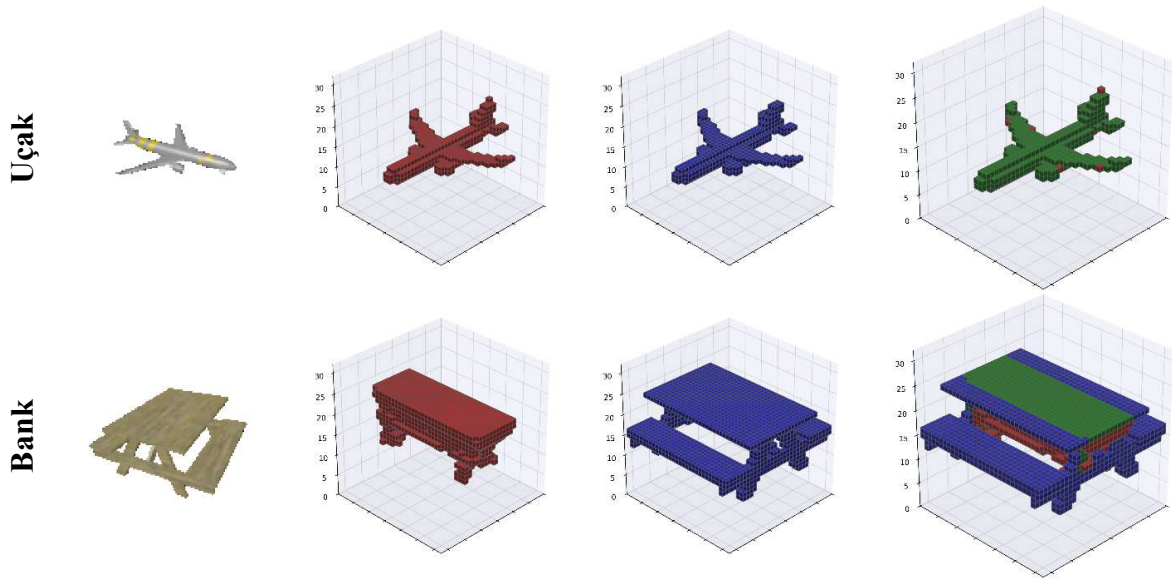


Şekil 4.4. 13 sınıf için 2D girdi, kullanılan kayıp fonksiyonları ile üretilen 3D nesne ve gerçek 3D nesne

4.2.2. CPSA nitel bulguları

Önerilen CPSA dikkat mekanizmasının ve SA veri artırma yönteminin birlikte kullanımına dayanan nitel bulgular bu bölümde detaylandırılmıştır. Şekil 4.5'te veri setinden seçilen masa, lamba, telefon, uçak ve bank sınıfları için kullanılan 2D girdi, bu 2D girdiye karşılık gelen ve veri setinde bulunan gerçek 3D nesne (mavi vokseller) ve 2D girdiye karşılık CPSA+SA mimarisi tarafından üretilen 3D nesne (kırmızı vokseller) gösterilmiştir. Ayrıca merkezde üst üste getirilerek doğru tahmin edilen vokseller yeşil ile işaretlenerek belirginleştirilmiştir. Şekil incelenecek olursa masanın başarılı bir şekilde oluşturulduğu görülmekle birlikte ayaklarında eksikler bulunmaktadır. Lamba için ise üst tarafta bulunan oval yapıdaki eksikler görülür. Telefon için başarılı bir üretim gerçekleştirilmiştir. Uçak için oldukça başarılı bir üretim gerçekleştirilmiştir neredeyse tüm detaylar CPSA tarafında yakalanmıştır. Bank için ise tam tersi şekilde sadece merkez konumda bir eşleşme yakalanmakla birlikte oturma kısımlarında ve ayaklardaki eksikler görülür.



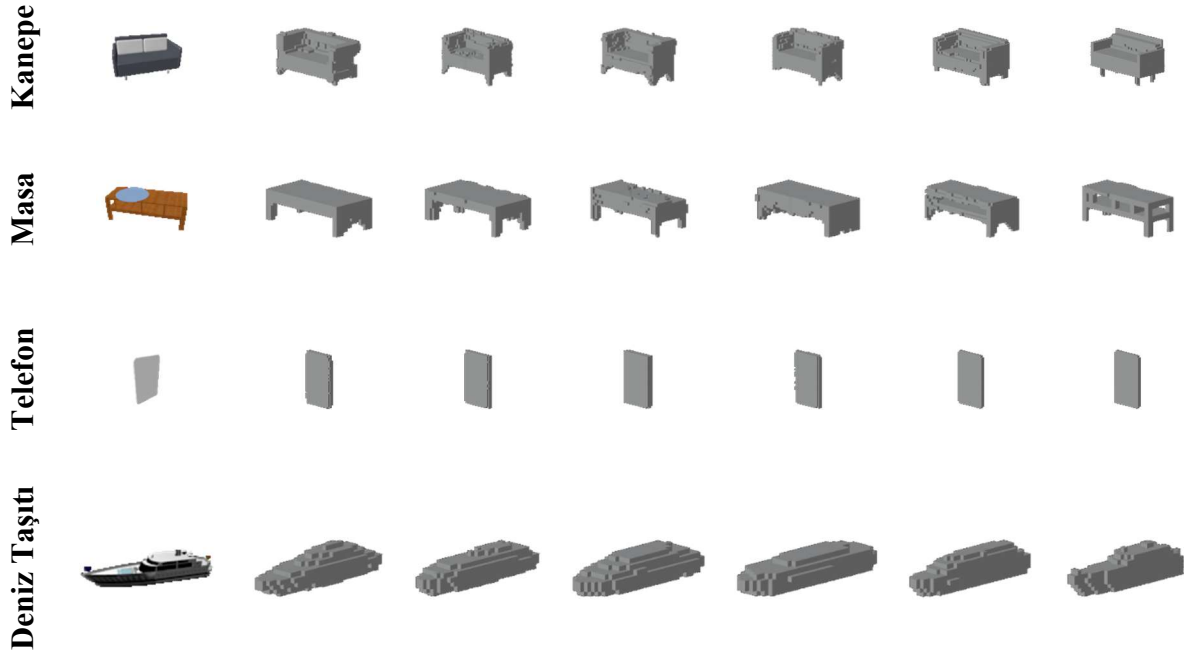


Şekil 4.5. CPSA ile 2D girdilerden üretilen 3D nesnelerin, gerçek nesnelerle üst üste karşılaştırması

Şekil 4.6’da 13 sınıf için seçilen örnek 2D girdi, bunlara karşılık gelen gerçek 3D nesne ve kıyaslama için kullanılan dikkat mekanizmaları tarafından ilgili 2D girdiye karşılık üretilen 3D nesneler gösterilmiştir. Şekil incelenecek olursa CPSA+SA’nın uçak sınıfında başarısı çok açıktır. Özellikle uçağın motor kısımlarını başarılı bir şekilde oluşturabilen tek mimari önerilendir. Bank için ise SA veri artırma tekniğinin dezavantajı görülür çünkü bankın ayaklarının arasının boş olması gerekirken özellikle SA kullanan mimariler bu aralığı da birleştirmiştir. Burada SA’nın doku özelliklerini yok etmesinin bir dezavantajı görülür. Bununla birlikte önerilen mimari bankın oturma kısmını pürüzsüz bir şekilde oluşturmuştur. Dolap için genel olarak tüm mimariler başarılı bir üretim gerçekleştirmiştir. Otomobil için CBAM+SA harici mimariler başarılı bir üretim gerçekleştirirken bu mimarinin genel dağılıma uyarak kamyonet tarzı bir üretim gerçekleştirdiği görülür. İlgili girdiye göre üretilen bu yapı oldukça farklıdır. Sandalye için CPSA+SA’nın yüksek frekanslı detayları yakalama başarımı bir kez daha açıkça görülür. Özellikle sandalyenin ayak kısmı başarılı bir şekilde oluşturulmuş, ayrıca oturma kısmındaki kavis ve tutamaçlar önerilen mimari tarafından diğer mimarilere kıyasla daha başarılı bir şekilde üretilmiştir. Ekran için yine önerilen mimarinin pürüzsüz üretimi görülür, stand kısmı da başarılı bir şekilde oluşturulmuş olsa da oval kısım yerine daha keskin bir üretim yapıldığı görülür. Lamba için alt ve üst kısımdaki bağlantının yalnızca önerilen mimari tarafından başarılı bir şekilde korunduğu görülür. Hoparlör için genel oval yapı pek çok mimari tarafından yakalanmakla birlikte CBAM + SA’nın genel dağılıma uyarak yapıldığı görülür. Tüfek için önerilen mimarinin başarımı bir kez daha açığa çıkar. Tüfeğin ön tutamaç ve arka tutamaç kısımları yalnızca önerilen mimari tarafından yakalanmış ve üretilmiştir. Bununla birlikte, CPSA’nın tüfek namlusu gibi çok ince detayların üretilmesinde voksel çözünürlüğü sınırlarından dolayı kısıtlı kaldığı da görülmüştür. Bu durum, CPSA’nın küresel geometriyi algılamadaki

gücüne karşın, düşük voksel çözünürlüğüne bağlı olarak çok ince alt parçaların inşasında bazı darboğazlar yaşayabildiğini doğrulamaktadır. Kanepe için tüm mimariler başarılı bir üretim gerçekleştirmekle birlikte özellikle oturma kısmındaki pürüzler göze çarpmaktadır. Masa için arada bulunan boşluk yalnızca önerilen mimari tarafından yakalanmakla birlikte boşlukta bulunan dikey bölümler eksik üretilmiştir. Telefonun genel olarak başarılı bir şekilde üretildiği görülür. Deniz taşıtı için genel hatlar tüm mimariler tarafından başarılı bir şekilde üretilse de detaylardaki eksiklikler görülür.

	2D Girdi	TM	TM+SA	SE + SA	CBAM + SA	CPSA + SA	Gerçek 3D Nesne
Uçak							
Bank							
Dolap							
Otomobil							
Sandalye							
Ekran							
Lamba							
Hoparlör							
Tüfek							

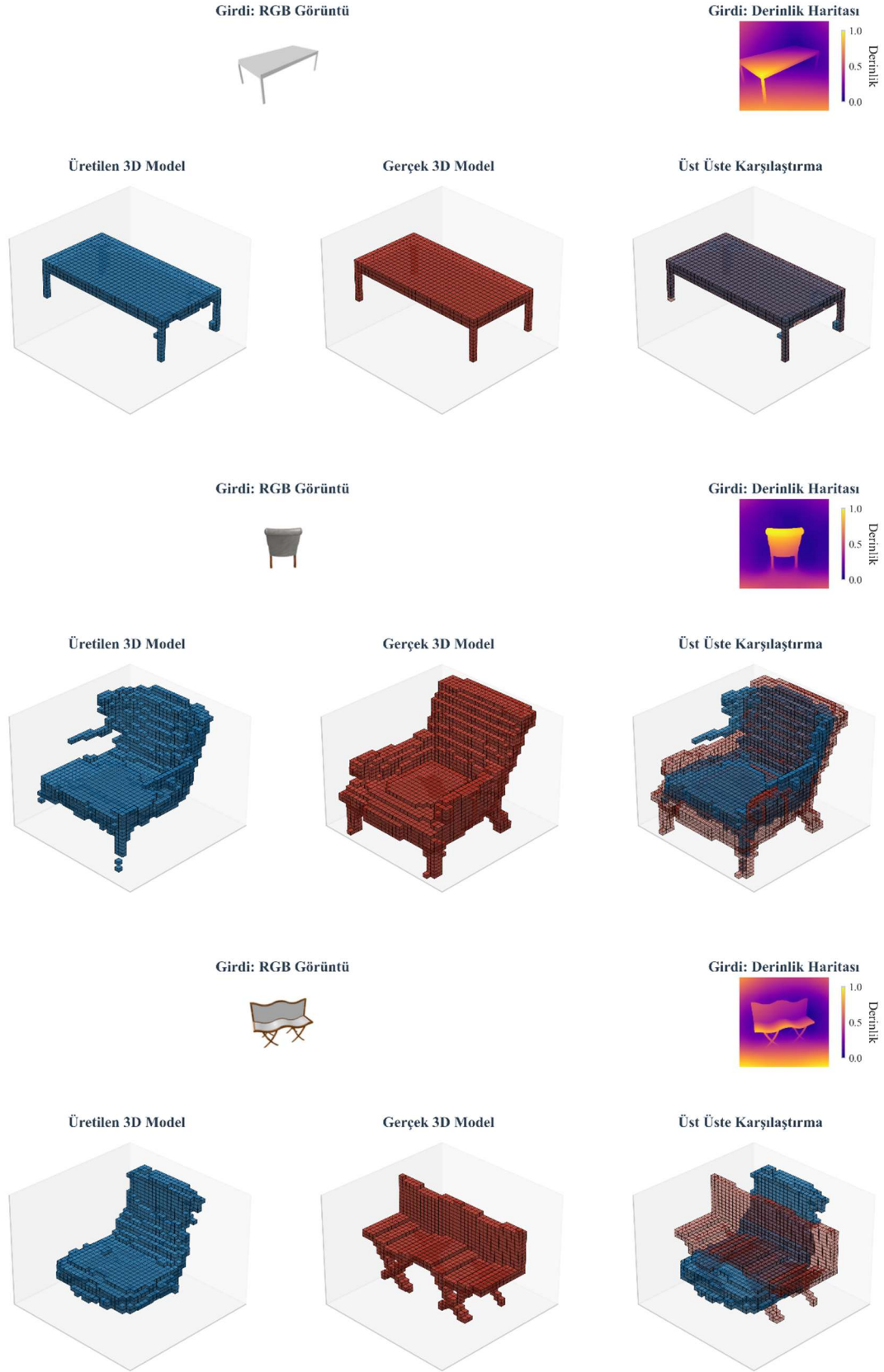


Şekil 4.6. 13 sınıf için 2D girdi, kullanılan dikkat mekanizmaları ile üretilen 3D nesne ve gerçek 3D nesne

CPSA+SA mimarisi genel olarak değerlendirildiğinde özellikle küresel yapının başarılı bir şekilde yeniden üretildiği görülür. Literatürdeki rakibi diğer dikkat mekanizmaları ile kıyaslandığında ise bunlarla açık bir şekilde rekabetçi olduğu ve hatta daha başarılı sonuçlar üretebildiği görülür.

4.2.3. SWAGE-3D nitel bulguları

SWAGE-3D mimarisinin tümleşik yapısı ve pek çok başarılı bileşeni birlikte sunması nitel değerlendirmelerde de görülmüştür. Şekil 4.7’de veri setinden seçilen 3 adet sınıf için girdi olarak verilen RGB görüntü ve derinlik haritası gösterilmiştir. İlgili şekilde bu girdilere karşılık SWAGE-3D tarafından üretilen 3D nesneler, 2D girdilere karşılık gelen gerçek 3D nesne ve üst üste karşılaştırma açık bir şekilde görülür. Sonuçların açık bir şekilde gösterilebilmesi bakımından SWAGE-3D’nin en başarılı olduğu masa, ortalama başarılı olduğu sandalye ve başarısız olduğu kanepe sınıflarından örnekler seçilmiştir. Burada başarıdan kasıt örnek bazlı olup tüm genel mimariyi yansıtmaz. Şekil incelenecek olursa masanın başarılı bir şekilde üretildiği çok ufak voksel fazlalıkları bulunduğu görülür. Sandalye için ise özellikle ayak ve tutamaç kısımlarında eksikler bulunduğu SWAGE-3D tarafından tutamaç kısımlarının oluşturulmaya başlandığı ancak aradaki bağlantının başarılı bir şekilde kurulamadığı görülür. Kanepe için ise görüleceği üzere en kötü sonuçlar elde edilmiştir. Özellikle ayak kısımları ve kanepenin genişliği SWAGE-3D tarafından doğrudan yanlış bir şekilde üretilmiş yalnızca merkez bölgesinde bir voksel topluluğunun doğru bulunduğu görülür ancak bu toplulukta konumsal olarak hatalıdır.



Şekil 4.7. Masa, sandalye ve kanepeler için SWAGE-3D başarımının karşılaştırılması

Şekil 4.8’de RGB ve derinlik kanalı ayrı ayrı gösterilmiştir. RGB girdilere karşılık gelen ve veri setinde hali hazırda bulunan gerçek 3D nesneler, ilgili 2D RGB girdilere karşılık 3D-

R2N2, Pix2Mesh ve SWAGE-3D tarafından üretilen 3D nesneler gösterilmiştir. Şekil incelenecek olursa uçak için SWAGE-3D'nin başarımı görülür, nitekim uçak kokpitinin üst tarafında bulunan ayrıntı önerilen mimari tarafından başarılı bir şekilde üretilmiştir. Bank için ise mimarilerin başarısız olduğu söylenebilir, özellikle bankın ayak kısımları başarılı bir şekilde üretilmemiştir. SWAGE-3D tarafından tutamaç kısımları üretilse dahi konumsal olarak yanlıştır. Dolap için tüm mimarilerin ana hattı başarılı bir şekilde ürettiği görülmekle birlikte detaylarda eksiklikler görülür. Otomobil için her 3 mimari de ana hattı başarılı bir şekilde üretmiştir. Sandalye için oturma yeri kalınlığı SWAGE-3D tarafından daha nitelikli üretilmekle birlikte ana hattı 3 mimaride başarılı bir şekilde üretmiştir. Ekran için özellikle stand kısmının SWAGE-3D tarafından başarılı bir şekilde ürettiği görülür, diğer mimarilerin bu kısmı kaçırdığı açıktır. Lamba için 3D-R2N2 ve SWAGE-3D düzgün ve pürüzsüz bir üretim gerçekleştirirken Pix2Mesh'in özellikle lambanın bağlantı kısmında eksikliği görülür. Hoparlör için genel hatlar ve topoloji tüm mimariler tarafından üretilmişken özellikle kolon kısımlarında bulunan detayların mimariler tarafından başarılı bir şekilde üretilmediği görülür. Tüfek için SWAGE-3D ve Pix2Mesh tarafından tutamaç kısımları başarılı bir şekilde üretilmiştir. Kanepe için genel hatlarıyla tüm mimariler başarı göstermekle birlikte SWAGE-3D'nin özellikle oturma kısmındaki nispeten daha pürüzsüz üretimi görülür. Masa için ayak kısımları önerilen mimari tarafından başarılı bir şekilde üretilmiştir. Telefon için her 3 mimaride başarılı olmakla birlikte SWAGE-3D çok daha pürüzsüz ve doğru köşeli bir yapı üretmiştir. Deniz taşıtı için özellikle üst bölümün SWAGE-3D tarafından başarılı bir şekilde üretildiği ancak detaylarda eksiklik bulunduğu görülür.



Şekil 4.8. RGB-D girdi, girdilere karşılık gelen gerçek 3D nesne ve bu girdilere karşılık 3D-R2N2, Pix2Mesh ve SWAGE-3D tarafından üretilen 3D nesneler

4.3. Çalışmanın Sınırlılıkları

Tez çalışmasında önerilen yaklaşımlar ile tek 2D görüntüden 3D nesne üretimi konusunda başarılı sonuçlar elde edilmiştir. Bununla birlikte çalışmanın bazı sınırlılıkları da bulunmaktadır. İlk olarak deneysel çalışmaların tamamında $32 \times 32 \times 32$ voksel çözünürlüğünün tercih edilme nedeni, elde edilen geometrik başarımla hesaplama maliyeti arasındaki dengeyi korumaktır. Her ne kadar voksel çözünürlüğü arttıkça temsil kapasitesi yükselse de bellek kullanımı ve hesaplama maliyeti kübik olarak artış göstermektedir. Bu noktada 32^3 seçimi, mimarinin temel geometrik formları öğrenmesi için yeterli bir zemin sunarken, donanım kaynaklarının verimli kullanılmasını sağlamaktadır. Daha yüksek çözünürlüklerin (örneğin 64^3) doğrudan daha iyi sonuçlar vereceği varsayılmamalıdır. Bu tür bir artışın beraberinde getireceği devasa bellek yükü ve eğitim süresi artışı göz önünde bulundurulmalıdır. Bu değer nesnenin genel formunun öğrenilmesi ve gözle ayırt edilebilmesi bakımından yeterli olmakla birlikte çok ince detayların gösteriminde sınırlı kalabilmektedir. Bu sınırlılık özellikle sandalye ayakları, uçak kanatları, lamba gövdesi, tüfek namlusu gibi ince ve uzun yapısal bileşenlere sahip sınıflarda daha belirgin olarak görülmüştür. Voksel temsilinin nispeten düşük çözünürlüğü nedeniyle bu tür ince parçalar ya kısmen kaybolmakta ya da çok daha kalın ve kaba şekilde ifade edilerek detayların seçilememesine sebep olmaktadır. Bu durumda IoU gibi metriklerde kabul edilebilir sonuçlar elde edilse dahi yüzey geometrisi ve görsel bütünlük açısından sınırlılıklar oluşabilmektedir.

İkinci olarak çalışmada kullanılan veri seti ve yapılan deneysel çalışmaların tamamı ShapeNet veri seti üzerinde gerçekleştirilmiştir. ShapeNet bu alanda yapılan çalışmalardaki standart ve en çok tercih edilen veri setlerinden biridir. Bununla birlikte ShapeNet'te bulunan görüntülerde çoğunlukla kontrollü arka plan, düzenli kamera açısı ve temiz nesne görüntüleri bulunur. Bu nedenle önerilen yöntemlerin gerçek dünya görüntüleri üzerinde, farklı ışık koşullarında, karmaşık arka planlarda veya kamera gürültüsünün arttığı durumlarda nasıl davranacağı test edilmemiştir. Bu durum özellikle gerçek dünya görüntüleri ve sentetik veri arasındaki alan kayması açısından önemli bir sınırlılık olarak ifade edilir.

Üçüncü olarak SWAGE-3D mimarisinde RGB girdilere derinlik kanalı eklenerek mimari RGB-D girdiler ile beslenmiştir. Derinlik bilgisi mimariye kritik bir geometrik ipucu sağlamakla birlikte bu derinlik haritaları gerçek sensör ölçümlerinden değil Depth Anything V2 tabanlı mimariden elde edilmiştir. Buna bağlı olarak derinlik haritalarında olası tahmin hataları, ölçek belirsizlikleri veya nesne sınırlarında oluşabilecek bozulmalar bulunabilir. Dolayısıyla derinlik destekli giriş temsili mimarinin başarımlarını artıran önemli bir bileşen olmakla birlikte bu

bileşenin doğruluğu kullanılan derinlik haritası çıkarıcı mimariye de doğrudan bağlıdır.

Dördüncü olarak GAN tabanlı mimariler doğası gereği eğitim kararlılığına duyarlıdır. Bu tez çalışmasında önerilen SWAGE-3D mimarisinde WGAN-GP ve SN gibi yaklaşımlar ile eğitim sürecinin kararlı hale getirilmesi amaçlanmıştır. Bununla birlikte GAN mimarilerinde mod çökmesi, gradyan kararsızlığı, hiperparametre duyarlılığı ve eğitim süresinin kullanılan donanıma bağımlılığı halen bulunabilmektedir. Özellikle 3D nesne üretimi gibi yüksek boyutlu çıktı uzayına sahip problemlerde yığın boyutu, tur sayısı, üretici-ayırt edici dengesi ve GPU bellek kapasitesi mimarinin başarımı üzerinde doğrudan etkili olabilmektedir. Kullanılan bu bileşenler, yapılar ve hiperparametreler de sınırlılık olarak ifade edilebilir.

Beşinci olarak yapılan önermelerin tamamında hem eğitim hem de test aşamaları sınıf bazlı olarak ayrı ayrı gerçekleştirilmiş ve her iki aşamada da yalnızca tek açıdan girdiler kullanılmıştır. Bu durumda kısıtlı kaynak ile maksimum başarımlar hedeflenmekle birlikte bir 2D görüntünün tek açıdan verisi o görüntüde bulunmayan kısımları için ek bilgi eksikliğine sebep olabilmektedir. Ayrıca sınıf bazlı eğitim yapılması eğitilen mimarilerin yalnızca o sınıf özelinde başarılı çıkarımlar yapabilmesini sağlar ve genelleme yeteneğinin bir miktar kaybolacağı söylenebilir. Bu durumlarda çalışmamızın sınırlılıkları olarak ifade edilebilir.

Son olarak daha yüksek çözünürlük ile teorik olarak daha ayrıntılı ve detaylı bir 3D nesne ortaya koyulmaktadır. Buna karşın pratik olarak bu durum daha yüksek bellek kullanımı, daha uzun eğitim süresi ve daha zor optimizasyon süreci anlamına gelmektedir. Gelecek çalışmalarda, 64^3 gibi daha yüksek çözünürlüklü voksel temsillerinin kullanımı bir seçenek olarak durmaktadır. Ancak bu geçiş, sadece çözünürlüğü artırmak olarak değil daha yüksek temsil kapasitesinin getirdiği karmaşık optimizasyon süreçleri ve katlanan donanım maliyetleri ile birlikte bir denge arayışı olarak tartışılmalıdır. Literatürdeki benzer çalışmalar ve SWAGE-3D deneyleri göstermektedir ki çok yüksek çözünürlüklerde sabit tur sayıları altında başarımın artmaması, hatta optimizasyonun zorlaşması nedeniyle metriklerde düşüş yaşanması olasıdır. Bu nedenle, yüksek çözünürlük hedeflenirken çok ölçekli üretici mimariler veya seyrek voksel yapıları gibi alternatiflerin, bu hesaplama maliyetini dengelemek üzere kullanılması planlanmaktadır. Bu nedenle gelecekteki çalışmalarda daha yüksek çözünürlüklü voksel temsilleri, çok ölçekli üretici mimariler, alternatif 3D temsil yöntemleri ve gerçek dünya görüntüleri üzerinde daha kapsamlı analizler ele alınabilir.

5. SONUÇLAR VE ÖNERİLER

Yapılan bu tez çalışmasında, 2D görüntülerden 3D nesneler üretimi amacıyla AWIF olarak adlandırılan yeni bir yeniden yapılandırma kayıp fonksiyonu, CPSA olarak adlandırılan frekans düzleminde işlem gerçekleştiren yeni bir dikkat mekanizması ve SWAGE-3D adıyla tümleşik geliştirmeler içeren yeni bir mimari önerilmiştir. IoU, CD ve F-Skoru metrikleri kullanılarak bu mimarilerin başarımı kıyaslanmış ve değerlendirilmiştir.

Yapılan eğitim ve test aşamaları incelendiğinde, önerilen mimarilerinin her birinin başarımı açık bir şekilde gösterilmiştir. Özellikle tüm önerilerin üzerine kurulduğu temel 3D-VAE-GAN mimarisi ile yapılan kıyaslamalar bunu açık bir şekilde göstermektedir. Her bir mimarinin kendi rakibi olan mekanizmalar ile yapıldığı kıyaslamada ise yine her birinin rekabetçi olduğu görülmüştür. Bununla birlikte literatürdeki SOTA mimariler ile kıyaslanabilmelerine karşın eksiklikleri bulunduğu da görülmüştür.

AWIF'in diğer yeniden yapılandırma kayıp fonksiyonlarına göre özellikle yüzey voksellerini daha başarılı bir şekilde ürettiği genel hacimsel doğruluğu ve topolojiyi de olumlu etkilediği görülmüştür. Bununla birlikte zamansal açıdan külfetli olduğu tespit edilmiştir. Kullanılan yüzey bulma algoritmasının karmaşıklığının buna sebep olduğu düşünülmektedir.

CPSA'nın klasik uzamsal dikkat mekanizmalarından çok daha etkili bir yöntem olduğu tespit edilmiştir. Özellikle nesnenin küresel yapısının frekans düzleminde uygulanan dikkat mekanizmaları ile çok daha kolay tespit edilebileceği ve ilgili mimarilerin buna göre daha doğru kararlar verebileceği gösterilmiştir. SA veri artırma tekniğinin özellikle 3D yeniden yapılandırma problemlerinde tercih edilebilecek güçlü bir veri artırma tekniği olduğu tespit edilmiştir. Bununla birlikte klasik sınıflandırma problemlerinde kullanılması başarımı düşürebilir, çünkü 3D yeniden yapılandırma alanında şu an için genel topolojinin üretimi önem arz etmekte iken sınıflandırmada dokusal bilgiler daha fazla önem arz edebilir. CPSA+SA'nın birlikte kullanımı oldukça başarılı sonuçlar verebilmesine karşın zamansal açıdan ve kullanılan parametre sayısı açısından sakınca barındırmaktadır. Bu bağlamda probleme özgü bir tercih yapılması tavsiye edilir.

SWAGE-3D'nin tümleşik yapısı altında yer alan bileşenlerin katkıları incelendiğinde, ablasyon sonuçlarının tüm metriklerde homojen bir artış yerine tamamlayıcı ve metriğe özgü etkiler sunduğu görülmüştür. Özellikle derinlik bilgisinin bu tarz problemlerde ağırlık ek bir geometrik önsel bilgi sağlamasının, CD ve yüzey hassasiyeti üzerinde oldukça belirgin bir iyileşme sağladığı gözlemlenmiştir. WGAN-GP ve öz dikkat mekanizması gibi bileşenler ise sayısal metriklerdeki dalgalanmalara karşın, üretilen nesnelerin yapısal kalitesi ve hiyerarşik

tutarlılığı açısından kritik rol oynamaktadır.

Ablasyon çalışmalarının önemi ve sonuca etkisi açık bir şekilde görülmüştür. Özellikle birden çok geliştirme ve bileşen içeren mimarilerdeki her bir parçanın başarımının ayrı ayrı görülmesi hem literatür bakımından hem de mimarinin başarımı açısından elzemdir. Böylece hangi bileşenin daha etkili hangisinin daha az etkili olduğu görülebilir ve sonraki çalışmalarda buna göre bir tercih yapılması sağlanabilir.

Tek metrik yerine birden çok metrik kullanımının sonuçları tarafsız ve adil bir şekilde kıyaslama bakımından çok önemli olduğu tespit edilmiştir. Farklı metrikler üzerinden doğrulama yapılarak önerilen mimarilerin hangi kısımlarda eksik oldukları ve genel başarımları görülmüştür.

Bu alanda çalışacak araştırmacıların GAN dengesini çok doğru ve dengeli bir şekilde kurması tavsiye edilir. Özellikle standart GAN mimarisinin dengesi çok kolay bir şekilde bozulabilmekte ve ağırlardan biri diğerini baskılayabilmektedir. Bu gibi durumlarda WGAN-GP gibi alternatif mimariler tercih edilebilir. 3D-VAE-GAN mimarisi doğası gereği karmaşık bir yapıya sahiptir, bu nedenle çalışmanın başında en doğru kurgunun yapılarak bu doğrultuda ilerlenmesi tavsiye edilir. Bu mimarinin eğitimi içinde tek yaklaşım bulunmamakla birlikte girdiler ve çıktılar hangi sıra ile izlenecek, nerede ayrımlar gerçekleştirilecek, gradyan akışları nerede kesilip devam ettirilecek bunların sistematik bir şekilde planlanarak mimarinin kurulması ilerleyen aşamalarda kolaylık sağlayacaktır. Güncel bir GPU veya bulut sistem kullanılması özellikle 3D nesne yapılandırması alanında büyük fayda sağlayacaktır. IoU gibi metriklerin eşik değerlerinin doğru bir şekilde seçilmesi, probleme ve tasarlanan mimariye göre kullanılması sonuçları olumlu etkileyecektir. Geliştirme ortamının kurulumu çok titiz bir şekilde yönetilmelidir. Sürüm uyumsuzlukları, CUDA sürüm farklılıkları ve GPU uyumsuzlukları gibi sorunların elimine edilmesi amacıyla en başta ortam doğru bir şekilde kurulmalı ve mümkünse buna sadık kalınmalıdır.

Gelecek çalışmalarda 3D nesne temsili için sadece voksel değil nokta bulutu ve poligon gibi yöntemlerin de kullanılması planlanmaktadır. Sadece ShapeNet veri seti değil ModelNet gibi alternatif veri setleri karma bir şekilde kullanılarak daha doğru ve başarılı sonuçlar elde edilmesi planlanmaktadır. Geliştirilen yöntemlerin doğrudan canlı uygulamalara çevrilmesi ve endüstri tarafından kullanılması için çalışmalar planlanmaktadır. Derinlik haritasının sonuca olumlu etkisi görüldüğü için bu alandaki diğer mimariler ile de çıkarımlar yapılarak bunlar arasında bir kıyaslama yapılabilir. Özellikle kübik maliyet sebebiyle 3D-VAE-GAN mimarilerinin yüksek çözünürlüklere ulaşması oldukça zordur buna bağlı olarak difüzyon

modelleri ve örtük temsiller gibi yöntemlerin kullanılması hem çok daha gerçekçi sonuçlar elde edilmesini sağlamakta hem de hesaplama maliyeti bakımından makul olabilmektedir (Chabra vd., 2020). Bu bağlamda difüzyon modelleri ve örtük temsiller gibi yöntemler gelecek çalışmalarda tercih edilebilirler.

KAYNAKÇA

- Achlioptas, P., Diamanti, O., Mitliagkas, I., & Guibas, L. (2018). *Learning Representations and Generative Models for 3D Point Clouds*. J. Dy, A. Krause (Ed.). 35th International Conference on Machine Learning, Stockholm, İsveç. s. 40-49.
- Arjovsky, M., Chintala, S., & Bottou, L. (2017). *Wasserstein Generative Adversarial Networks*. D. Precup, Y. W. Teh (Ed.). 34th International Conference on Machine Learning, Sidney, Avustralya. s. 214-223.
- Barrow, H. G., Tenenbaum, J. M., Bolles, R. C., & Wolf, H. C. (1977). *Parametric Correspondence and Chamfer Matching: Two New Techniques for Image Matching*. R. Reddy (Ed.). 5th International Joint Conference on Artificial Intelligence, Cambridge, Massachusetts, ABD. s. 659-663.
- Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8), 1798-1828.
- Bengio, Y., Louradour, J., Collobert, R., & Weston, J. (2009). *Curriculum Learning*. A. Danyluk, L. Bottou (Ed.). 26th Annual International Conference on Machine Learning, Montreal, Kanada. s. 41-48.
- Ben-Hamu, H., Maron, H., Kezurer, I., Avineri, G., & Lipman, Y. (2018). Multi-chart generative surface modeling. *ACM Transactions on Graphics*, 37(6), Article 215.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer, ABD.
- Brock, A., Lim, T., Ritchie, J. M., & Weston, N. (2016). Generative and discriminative voxel modeling with convolutional neural networks. *arXiv preprint*, arXiv:1608.04236.
- Chabra, R., Lenssen, J. E., Ilg, E., Schmidt, T., Straub, J., Lovegrove, S., & Newcombe, R. (2020). *Deep Local Shapes: Learning Local SDF Priors for Detailed 3D Reconstruction*. A. Vedaldi vd. (Ed.). Computer Vision ECCV 2020, Springer International Publishing, İsviçre. s. 608-625.
- Chang, A. X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., & Yu, F. (2015). ShapeNet: An information-rich 3D model repository. *arXiv preprint*, arXiv:1512.03012.
- Chen, Z., Tagliasacchi, A., & Zhang, H. (2020). *BSP-Net: Generating Compact Meshes via Binary Space Partitioning*. C. Liu vd. (Ed.). IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, Washington, ABD. s. 45-54.
- Chetlur, S., Woolley, C., Vandermersch, P., Cohen, J., Tran, J., Catanzaro, B., & Shelhamer, E. (2014). cuDNN: Efficient primitives for deep learning. *arXiv preprint*, arXiv:1410.0759.
- Chollet, F. (2018). *Deep learning with Python*. Manning Publications, ABD.
- Choy, C. B., Xu, D., Gwak, J., Chen, K., & Savarese, S. (2016). *3D-R2N2: A Unified Approach for Single and Multi-View 3D Object Reconstruction*. B. Leibe vd. (Ed.). Computer Vision - ECCV 2016, Springer International Publishing, İsviçre. s. 628-644.
- Cooley, J. W., & Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 19(90), 297-301.
- Dai, A., Nießner, M., Zollhöfer, M., Izadi, S., & Theobalt, C. (2017). BundleFusion: Real-time globally consistent 3D reconstruction using on-the-fly surface reintegration. *ACM Transactions on Graphics*, 36(3), 1-18.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). *ImageNet: A Large-Scale Hierarchical Image Database*. D. Forsyth vd. (Ed.). IEEE Conference on Computer Vision and Pattern Recognition, Miami, Florida, ABD. s. 248-255.
- DeVries, T., & Taylor, G. W. (2017). Improved regularization of convolutional neural networks with cutout. *arXiv preprint*, arXiv:1708.04552.
- Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26(3), 297-302.
- Dumoulin, V., & Visin, F. (2016). A guide to convolution arithmetic for deep learning. *arXiv preprint*, arXiv:1603.07285.
- Engel, J., Schöps, T., & Cremers, D. (2014). *LSD-SLAM: Large-Scale Direct Monocular SLAM*. D. Fleet vd. (Ed.). Computer Vision-ECCV 2014, Springer International Publishing, İsviçre. s. 834-849.
- Fan, H., Su, H., & Guibas, L. J. (2017). *A Point Set Generation Network for 3D Object Reconstruction from a Single Image*. J. Rehg vd. (Ed.). IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Hawaii, ABD. s. 605-613.
- Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4), 193-202.

- Funkhouser, T., Kazhdan, M., Shilane, P., Min, P., Kiefer, W., Tal, A., & Dobkin, D. (2004). Modeling by example. *ACM Transactions on Graphics*, 23(3), 652-663.
- Gao, J., Shen, T., Wang, Z., Chen, W., Yin, K., Li, D., & Fidler, S. (2022). *Get3D: A Generative Model of High Quality 3D Textured Shapes Learned from Images*. *Advances in Neural Information Processing Systems*, 35, 31841-31854.
- Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F. A., & Brendel, W. (2019). *ImageNet-Trained CNNs Are Biased Towards Texture; Increasing Shape Bias Improves Accuracy and Robustness*. Y. Bengio, Y. LeCun (Ed.). *International Conference on Learning Representations*, New Orleans, Louisiana, ABD. s. 1-14.
- Goldstein, E. B. (1989). *Sensation and perception*. Wadsworth, ABD.
- Gong, H., Wang, P., Ma, Y., & Zhang, Y. (2025). GC-HG Gaussian splatting single-view 3D reconstruction method based on depth prior and pseudo-triplane. *Algorithms*, 18(12), 761.
- Gonzalez, R. C., & Woods, R. E. (2009). *Digital image processing*. Pearson Education, Hindistan.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press, ABD.
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., & Bengio, Y. (2014). *Generative Adversarial Networks*. Z. Ghahramani vd. (Ed.). *Advances in Neural Information Processing Systems*, Montreal, Kanada. s. 2672-2680.
- Groueix, T., Fisher, M., Kim, V. G., Russell, B. C., & Aubry, M. (2018). *A Papier-Mâché Approach to Learning 3D Surface Generation*. S. Dickinson vd. (Ed.). *IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, Utah, ABD. s. 216-224.
- Guan, Y., Jahan, T., & van Kaick, O. (2020). *Generalized Autoencoder for Volumetric Shape Generation*. C. Liu vd. (Ed.). *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Seattle, Washington, ABD. s. 268-269.
- Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. (2017). *Improved Training of Wasserstein GANs*. I. Guyon vd. (Ed.). *Advances in Neural Information Processing Systems*, Long Beach, Kaliforniya, ABD. s. 5767-5777.
- Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., & Bennamoun, M. (2021). Deep learning for 3D point clouds: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(12), 4338-4364.
- Hartley, R., & Zisserman, A. (2003). *Multiple view geometry in computer vision*. Cambridge University Press, İngiltere.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep Residual Learning for Image Recognition*. R. Zabih vd. (Ed.). *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, Nevada, ABD. s. 770-778.
- Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, 313(5786), 504-507.
- Ho, J., Jain, A., & Abbeel, P. (2020). *Denoising Diffusion Probabilistic Models*. H. Larochelle vd. (Ed.). *Advances in Neural Information Processing Systems*, Sanal, Kanada. s. 6840-6851.
- Hong, Y., Zhang, K., Gu, J., Bi, S., Zhou, Y., Liu, D., & Tan, H. (2024). *LRM: Large Reconstruction Model for Single Image to 3D*. Y. Bengio, Y. LeCun (Ed.). *International Conference on Learning Representations*, Viyana, Avusturya. s. 1-18.
- Hu, J., Shen, L., & Sun, G. (2018). *Squeeze-and-Excitation Networks*. S. Dickinson vd. (Ed.). *IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, Utah, ABD. s. 7132-7141.
- Huang, G., Li, Y., Pleiss, G., Liu, Z., Hopcroft, J. E., & Weinberger, K. Q. (2017). *Snapshot Ensembles: Train 1, Get M for Free*. Y. Bengio, Y. LeCun (Ed.). *International Conference on Learning Representations*, Toulon, Fransa. s. 1-13.
- Ioffe, S., & Szegedy, C. (2015). *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. F. Bach, D. Blei (Ed.). *32nd International Conference on Machine Learning*, Lille, Fransa. s. 448-456.
- Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2017). *Image-to-Image Translation with Conditional Adversarial Networks*. J. Rehg vd. (Ed.). *IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, Hawaii, ABD. s. 1125-1134.
- Jaccard, P. (1912). The distribution of the flora in the alpine zone. 1. *New Phytologist*, 11(2), 37-50.
- Jebara, T. (2012). *Machine learning: Discriminative and generative*. Springer, ABD.

- Ji, M., Gall, J., Zheng, H., Liu, Y., & Fang, L. (2017). *SurfaceNet: An End-to-End 3D Neural Network for Multiview Stereopsis*. K. Ikeuchi vd. (Ed.). IEEE International Conference on Computer Vision, Venedik, Ītalya. s. 2307-2315.
- Jun, H., & Nichol, A. (2023). Shap-E: Generating conditional 3D implicit functions. *arXiv preprint*, arXiv:2305.02463.
- Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2018). *Progressive Growing of GANs for Improved Quality, Stability, and Variation*. Y. Bengio, Y. LeCun (Ed.). International Conference on Learning Representations, Vancouver, Kanada. s. 1-14.
- Karras, T., Laine, S., & Aila, T. (2019). *A Style-Based Generator Architecture for Generative Adversarial Networks*. A. Yuille vd. (Ed.). IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, Kaliforniya, ABD. s. 4401-4410.
- Kazhdan, M., Funkhouser, T., & Rusinkiewicz, S. (2003). *Rotation Invariant Spherical Harmonic Representation of 3D Shape Descriptors*. L. Kobbelt vd. (Ed.). Symposium on Geometry Processing, Aachen, Almanya. s. 156-164.
- Kerbl, B., Kopanas, G., Leimkühler, T., & Drettakis, G. (2023). 3D Gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4), Article 139.
- Kingma, D. P., & Ba, J. (2015). *Adam: A Method for Stochastic Optimization*. Y. Bengio, Y. LeCun (Ed.). International Conference on Learning Representations, San Diego, Kaliforniya, ABD. s. 1-15.
- Kingma, D. P., & Welling, M. (2014). *Auto-Encoding Variational Bayes*. Y. Bengio, Y. LeCun (Ed.). International Conference on Learning Representations, Banff, Kanada. s. 1-14.
- Knapitsch, A., Park, J., Zhou, Q.-Y., & Koltun, V. (2017). Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics*, 36(4), Article 78.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *ImageNet Classification with Deep Convolutional Neural Networks*. F. Pereira vd. (Ed.). Advances in Neural Information Processing Systems, Lake Tahoe, Nevada, ABD. s. 1097-1105.
- Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1), 79-86.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324.
- Li, J., Niu, C., & Xu, K. (2020). *Learning Part Generation and Assembly for Structure-Aware Shape Synthesis*. V. Conitzer, F. Sha (Ed.). AAAI Conference on Artificial Intelligence, New York, ABD. s. 11362-11369.
- Li, Z., Li, W., Wang, T., Wang, Z., Wu, J., Wang, H., & Guo, C. (2025). MoCA: Mixture-of-components attention for scalable compositional 3D generation. *arXiv preprint*, arXiv:2512.07628.
- Lin, J. (1991). Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*, 37(1), 145-151.
- Lin, T.-Y., Dollar, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017a). *Feature Pyramid Networks for Object Detection*. J. Rehg vd. (Ed.). IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Hawaii, ABD. s. 2117-2125.
- Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017b). *Focal Loss for Dense Object Detection*. K. Ikeuchi vd. (Ed.). IEEE International Conference on Computer Vision, Venedik, Ītalya. s. 2980-2988.
- Liu, M., Shi, R., Chen, L., Zhang, Z., Xu, C., Wei, X., & Su, H. (2024). *One-2-3-45++: Fast Single Image to 3D Objects with Consistent Multi-View Generation and 3D Diffusion*. K. Saenko vd. (Ed.). IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, Washington, ABD. s. 10072-10083.
- Maas, A. L., Hannun, A. Y., & Ng, A. Y. (2013). *Rectifier Nonlinearities Improve Neural Network Acoustic Models*. S. Dasgupta, D. McAllester (Ed.). 30th International Conference on Machine Learning, Atlanta, Georgia, ABD. s. 3.
- Mandikal, P., Navaneet, K. L., Agarwal, M., & Babu, R. V. (2018). *3D-LMNet: Latent Embedding Matching for Accurate and Diverse 3D Point Cloud Reconstruction from a Single Image*. R. Fisher vd. (Ed.). British Machine Vision Conference, Newcastle, Īngiltere. s. 1-12.
- Mescheder, L., Nowozin, S., & Geiger, A. (2017). *Adversarial Variational Bayes: Unifying Variational Autoencoders and Generative Adversarial Networks*. D. Precup, Y. W. Teh (Ed.). International Conference on Machine Learning, Sidney, Avustralya. s. 2391-2400.

- Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S., & Geiger, A. (2019). *Occupancy Networks: Learning 3D Reconstruction in Function Space*. A. Yuille vd. (Ed.). IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, Kaliforniya, ABD. s. 4455-4465.
- Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2020). *NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis*. A. Vedaldi vd. (Ed.). European Conference on Computer Vision, Glasgow, Birleşik Krallık. s. 405-421.
- Milletari, F., Navab, N., & Ahmadi, S.-A. (2016). *V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation*. C. Schmid vd. (Ed.). 2016 Fourth International Conference on 3D Vision (3DV), Stanford, Kaliforniya, ABD. s. 565-571.
- Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. *arXiv preprint*, arXiv:1411.1784.
- Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill, ABD.
- Miyato, T., Kataoka, T., Koyama, M., & Yoshida, Y. (2018). *Spectral Normalization for Generative Adversarial Networks*. Y. Bengio, Y. LeCun (Ed.). International Conference on Learning Representations, Vancouver, Kanada. s. 1-14.
- Müller, T., Evans, A., Schied, C., & Keller, A. (2022). Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics*, 41(4), Article 102.
- Nair, V., & Hinton, G. E. (2010). *Rectified Linear Units Improve Restricted Boltzmann Machines*. J. Fürnkranz, T. Joachims (Ed.). 27th International Conference on Machine Learning, Hayfa, İsrail. s. 807-814.
- Nash, J. F. (1950). Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, 36(1), 48-49.
- Newcombe, R. A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A. J., & Fitzgibbon, A. (2011). *KinectFusion: Real-Time Dense Surface Mapping and Tracking*. G. Klinker vd. (Ed.). 2011 10th IEEE International Symposium on Mixed and Augmented Reality, Basel, İsviçre. s. 127-136.
- Ng, A., & Jordan, M. (2001). *On Discriminative vs. Generative Classifiers: A Comparison of Logistic Regression and Naive Bayes*. T. Dietterich vd. (Ed.). Advances in Neural Information Processing Systems, Vancouver, Kanada. s. 841-848.
- Nichol, A., Jun, H., Dhariwal, P., Mishkin, P., & Chen, M. (2022). Point-E: A system for generating 3D point clouds from complex prompts. *arXiv preprint*, arXiv:2212.08751.
- Nickolls, J., Buck, I., Garland, M., & Skadron, K. (2008). Scalable parallel programming with CUDA: Is CUDA the parallel programming model that application developers have been waiting for? *Queue*, 6(2), 40-53.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., & Chintala, S. (2019). *PyTorch: An Imperative Style, High-Performance Deep Learning Library*. H. Wallach vd. (Ed.). Advances in Neural Information Processing Systems, Vancouver, Kanada. s. 8024-8035.
- Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., & Efros, A. A. (2016). *Context Encoders: Feature Learning by Inpainting*. R. Zabih vd. (Ed.). IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, Nevada, ABD. s. 2536-2544.
- Pavlakos, G., Choutas, V., Ghorbani, N., Bolkart, T., Osman, A. A., Tzionas, D., & Black, M. J. (2019). *Expressive Body Capture: 3D Hands, Face, and Body from a Single Image*. A. Yuille vd. (Ed.). IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, Kaliforniya, ABD. s. 10975-10985.
- Peng, K., Islam, R., Quarles, J., & Desai, K. (2022). *TMVNet: Using Transformers for Multi-View Voxel-Based 3D Reconstruction*. K. Grauman vd. (Ed.). IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, Louisiana, ABD. s. 222-231.
- Pizlo, Z. (2008). *3D shape: Its unique place in visual perception*. MIT Press, ABD.
- Poly Haven (2022). *Horse Statue 01*. [Erişim: 06.05.2026, https://polyhaven.com/a/horse_statue_01]
- Poly Haven (2025). *Lion Head*. [Erişim: 06.05.2026, https://polyhaven.com/a/lion_head]
- Poole, B., Jain, A., Barron, J. T., & Mildenhall, B. (2023). *DreamFusion: Text-to-3D Using 2D Diffusion*. Y. Bengio, Y. LeCun (Ed.). International Conference on Learning Representations, Kigali, Ruanda. s. 1-15.
- Qi, C. R., Su, H., Mo, K., & Guibas, L. J. (2017). *PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation*. J. Rehg vd. (Ed.). IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Hawaii, ABD. s. 652-660.

- Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint*, arXiv:1511.06434.
- Rahman, M. A., & Wang, Y. (2016). *Optimizing Intersection-over-Union in Deep Neural Networks for Image Segmentation*. G. Bebis vd. (Ed.). Advances in Visual Computing, Springer International Publishing, İsviçre. s. 234-244.
- Ren, X., Huang, J., Zeng, X., Museth, K., Fidler, S., & Williams, F. (2024). *XCube: Large-Scale 3D Generative Modeling Using Sparse Voxel Hierarchies*. K. Saenko vd. (Ed.). IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, Washington, ABD. s. 4209-4219.
- Roh, Y., Heo, G., & Whang, S. E. (2021). A survey on data collection for machine learning: A big data-AI integration perspective. *IEEE Transactions on Knowledge and Data Engineering*, 33(4), 1328-1347.
- Rosenfeld, A., & Kak, A. C. (1976). *Digital picture processing*. Academic Press, ABD.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533-536.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., & Fei-Fei, L. (2015). ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3), 211-252.
- Russell, S. J., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4. Baskı). Pearson, ABD.
- Ruthotto, L., & Haber, E. (2021). An introduction to deep generative modeling. *GAMM-Mitteilungen*, 44(2), e202100008.
- Saito, S., Huang, Z., Natsume, R., Morishima, S., Kanazawa, A., & Li, H. (2019). *PIFu: Pixel-Aligned Implicit Function for High-Resolution Clothed Human Digitization*. K. Lee vd. (Ed.). IEEE/CVF International Conference on Computer Vision, Seul, Güney Kore. s. 2304-2314.
- Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., & Chen, X. (2016). *Improved Techniques for Training GANs*. D. Lee vd. (Ed.). Advances in Neural Information Processing Systems, Barselona, İspanya. s. 2234-2242.
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85-117.
- Serin, Z., Yüzgeç, U., & Karakuzu, C. (2024). *Pre-Trained Variational Autoencoder Approaches for Generating 3D Objects from 2D Images*. M. N. Seyman (Ed.). 2nd International Congress of Electrical and Computer Engineering, Springer Nature, İsviçre. s. 87-101.
- Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), 60.
- Siddiqui, Y., Alliegro, A., Artemov, A., Tommasi, T., Sirigatti, D., Rosov, V., & Nießner, M. (2024). *MeshGPT: Generating Triangle Meshes with Decoder-Only Transformers*. K. Saenko vd. (Ed.). IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, Washington, ABD. s. 19615-19625.
- Smith, E. J., & Meger, D. (2017). *Improved Adversarial Systems for 3D Object Generation and Reconstruction*. J. Peters vd. (Ed.). Conference on Robot Learning, Mountain View, Kaliforniya, ABD. s. 87-96.
- Sobel, I., & Feldman, G. (1968). *A 3x3 isotropic gradient operator for image processing*. Stanford Artificial Intelligence Project.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1), 1929-1958.
- Szeliski, R. (2011). *Computer vision: Algorithms and applications*. Springer, İngiltere.
- Tulsiani, S., Zhou, T., Efros, A. A., & Malik, J. (2017). *Multi-View Supervision for Single-View Reconstruction via Differentiable Ray Consistency*. J. Rehg vd. (Ed.). IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Hawaii, ABD. s. 2626-2634.
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433-460.
- Van Rijsbergen, C. (1979). *Information Retrieval: Theory and Practice*. B. Randell vd. (Ed.). Joint IBM/University of Newcastle Upon Tyne Seminar on Data Base Systems, Newcastle, İngiltere. s. 1-14.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Polosukhin, I. (2017). *Attention Is All You Need*. I. Guyon vd. (Ed.). Advances in Neural Information Processing Systems, Long Beach, Kaliforniya, ABD. s. 5998-6008.
- Villani, C. (2009). *Optimal transport: Old and new*. Springer, Almanya.

- Voleti, V., Yao, C.-H., Boss, M., Letts, A., Pankratz, D., Tochilkin, D., & Jampani, V. (2024). *SV3D: Novel Multi-View Synthesis and 3D Generation from a Single Image Using Latent Video Diffusion*. A. Leonardis vd. (Ed.). Computer Vision - ECCV 2024, Springer Nature, İsviçre. s. 439-457.
- Wang, H., Wang, Z., Long, X., Lin, C., Hancke, G., & Lau, R. W. (2025). *MAGE: Single Image to Material-Aware 3D via the Multi-View G-Buffer Estimation Model*. J. Zhang vd. (Ed.). Computer Vision and Pattern Recognition Conference, Nashville, Tennessee, ABD. s. 10985-10995.
- Wang, N., Zhang, Y., Li, Z., Fu, Y., Liu, W., & Jiang, Y.-G. (2018). *Pixel2Mesh: Generating 3D Mesh Models from Single RGB Images*. V. Ferrari vd. (Ed.). Computer Vision - ECCV 2018, Springer International Publishing, İsviçre. s. 55-71.
- Wang, P., Liu, L., Liu, Y., Theobalt, C., Komura, T., & Wang, W. (2021). NeuS: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *arXiv preprint*, arXiv:2106.10689.
- Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. (2018). *CBAM: Convolutional Block Attention Module*. V. Ferrari vd. (Ed.). European Conference on Computer Vision, Münih, Almanya. s. 3-19.
- Wu, J., Zhang, C., Xue, T., Freeman, W. T., & Tenenbaum, J. B. (2016). Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling. *arXiv preprint*, arXiv:1610.07584.
- Wu, R., Zhuang, Y., Xu, K., Zhang, H., & Chen, B. (2020). *PQ-Net: A Generative Part Seq2Seq Network for 3D Shapes*. C. Liu vd. (Ed.). IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, Washington, ABD. s. 829-838.
- Wu, S., Lin, Y., Zhang, F., Zeng, Y., Yang, Y., Bao, Y., & Yao, Y. (2025). Direct3D-S2: Gigascale 3D generation made easy with spatial sparse attention. *arXiv preprint*, arXiv:2505.17412.
- Wu, Z., Wang, X., Lin, D., Lischinski, D., Cohen-Or, D., & Huang, H. (2019). SAGNet: Structure-aware generative network for 3D-shape modeling. *ACM Transactions on Graphics*, 38(4), Article 91.
- Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., & Xiao, J. (2015). *3D ShapeNets: A Deep Representation for Volumetric Shapes*. S. Sclaroff vd. (Ed.). IEEE Conference on Computer Vision and Pattern Recognition, Boston, Massachusetts, ABD. s. 1912-1920.
- Xie, H., Yao, H., Sun, X., Zhou, S., & Zhang, S. (2019). *Pix2Vox: Context-Aware 3D Reconstruction from Single and Multi-View Images*. K. Lee vd. (Ed.). IEEE/CVF International Conference on Computer Vision, Seul, Güney Kore. s. 2690-2698.
- Xie, J., Zheng, Z., Gao, R., Wang, W., Zhu, S.-C., & Wu, Y. N. (2018). *Learning Descriptor Networks for 3D Shape Synthesis and Analysis*. S. Dickinson vd. (Ed.). IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Utah, ABD. s. 8629-8638.
- Yan, X., Lin, L., Mitra, N. J., Lischinski, D., Cohen-Or, D., & Huang, H. (2022). *ShapeFormer: Transformer-Based Shape Completion via Sparse Representation*. K. Grauman vd. (Ed.). IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, Louisiana, ABD. s. 6239-6249.
- Yang, L., Kang, B., Huang, Z., Xu, X., Feng, J., & Zhao, H. (2024a). *Depth Anything: Unleashing the Power of Large-Scale Unlabeled Data*. K. Saenko vd. (Ed.). IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, Washington, ABD. s. 10371-10381.
- Yang, L., Kang, B., Huang, Z., Zhao, Z., Xu, X., Feng, J., & Zhao, H. (2024b). *Depth Anything V2*. *Advances in Neural Information Processing Systems*, 37, 21875-21911.
- Yang, S., Zhang, H., Ren, J., Tang, Z., Zhao, M., & Liu, Y. (2025). *Zero-1-to-3DGS: A Single Image to 3D Gaussian by Consistent Multi-View Generation*. G. Valenzise vd. (Ed.). IEEE International Conference on Multimedia and Expo, Nantes, Fransa. s. 1-6.
- Yao, Y., Luo, Z., Li, S., Fang, T., & Quan, L. (2018). *MVSNet: Depth Inference for Unstructured Multi-View Stereo*. V. Ferrari vd. (Ed.). European Conference on Computer Vision, Münih, Almanya. s. 767-783.
- Yu, M.-S., Jung, T.-W., Yun, D.-Y., Hwang, C.-G., Park, S.-Y., Kwon, S.-C., & Jung, K.-D. (2024). A variational autoencoder cascade generative adversarial network for scalable 3D object generation and reconstruction. *Sensors*, 24(3), 751.
- Zhang, H., Cisse, M., Dauphin, Y. N., & Lopez-Paz, D. (2018). *Mixup: Beyond Empirical Risk Minimization*. Y. Bengio, Y. LeCun (Ed.). International Conference on Learning Representations, Vancouver, Kanada. s. 1-14.
- Zhang, H., Goodfellow, I., Metaxas, D., & Odena, A. (2019). *Self-Attention Generative Adversarial Networks*. K. Chaudhuri, R. Salakhutdinov (Ed.). 36th International Conference on Machine Learning, Long Beach, Kaliforniya, ABD. s. 7354-7363.

- Zhong, Z., Zheng, L., Kang, G., Li, S., & Yang, Y. (2020). *Random Erasing Data Augmentation*. V. Conitzer, F. Sha (Ed.). AAAI Conference on Artificial Intelligence, New York, ABD. s. 13001-13008.
- Zhou, Y., & Tuzel, O. (2018). *VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection*. S. Dickinson vd. (Ed.). IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Utah, ABD. s. 4490-4499.
- Zhu, J.-Y., Park, T., Isola, P., & Efros, A. A. (2017). *Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks*. K. Ikeuchi vd. (Ed.). IEEE International Conference on Computer Vision, Venedik, İtalya. s. 2242-2251.