

**T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**



**RAY VE ÇEVRESİNDEKİ NESNELERİN TESPİTİ VE MESAFE
ÖLÇÜMÜ İÇİN DERİN ÖĞRENME TABANLI
YAKLAŞIMLARIN GELİŞTİRİLMESİ**

Muhammed Amir ELMUHAMMEDCEBBEN

Yüksek Lisans Tezi

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

TEMMUZ 2025

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Bilgisayar Mühendisliği Anabilim Dalı

Yüksek Lisans Tezi

**RAY VE ÇEVRESİNDEKİ NESNELERİN TESPİTİ VE MESAFE
ÖLÇÜMÜ İÇİN DERİN ÖĞRENME TABANLI YAKLAŞIMLARIN
GELİŞTİRİLMESİ**

Tez Yazarı

Muhammed Amir ELMUHAMMEDCEBBEN

Danışman

Prof. Dr. İlhan AYDIN

TEMMUZ 2025

ELAZIĞ

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Bilgisayar Mühendisliği Anabilim Dalı

Yüksek Lisans Tezi

Başlığı:	Ray ve Çevresindeki Nesnelerin Tespiti ve Mesafe Ölçümü için Derin Öğrenme Tabanlı Yaklaşımların Geliştirilmesi
Yazarı:	Muhammed Amir ELMUHAMMEDCEBBEN
İlk Teslim Tarihi:	16.06.2025
Savunma Tarihi:	17.07.2025

TEZ ONAYI

Fırat Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına göre hazırlanan bu tez aşağıda imzaları bulunan jüri üyeleri tarafından değerlendirilmiş ve akademik dinleyicilere açık yapılan savunma sonucunda OYBİRLİĞİ ile kabul edilmiştir.

Danışman:	Prof. Dr. İlhan AYDIN Fırat Üniversitesi, Mühendislik Fakültesi	<i>İmza</i> Onayladım
Başkan:	Dr. Öğr. Üyesi Gülşah KARADUMAN Fırat Üniversitesi, Mühendislik Fakültesi	Onayladım
Üye:	Dr. Öğr. Üyesi İhsan TUĞAL Muş Alparslan Üniversitesi, Mühendislik ve Mimarlık Fakültesi	Onayladım

Bu tez, Enstitü Yönetim Kurulunun/...../20..... tarihli toplantısında tescillenmiştir.

İmza

Prof. Dr. Burhan ERGEN
Enstitü Müdürü

BEYAN

Fırat Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırladığım “Ray ve Çevresindeki Nesnelerin Tespiti ve Mesafe Ölçümü için Derin Öğrenme Tabanlı Yaklaşımların Geliştirilmesi” Başlıklı Yüksek Lisans Tezimin içindeki bütün bilgilerin doğru olduğunu, bilgilerin üretilmesi ve sunulmasında bilimsel etik kurallarına uygun davrandığımı, kullandığım bütün kaynakları atıf yaparak belirttiğimi, maddi ve manevi desteği olan tüm kurum/kuruluş ve kişileri belirttiğimi, burada sunduğum veri ve bilgileri unvan almak amacıyla daha önce hiçbir şekilde kullanmadığımı beyan ederim.

17.07.2025

Muhammed Amir ELMUHAMMEDCEBBEN



ÖNSÖZ

Demiryolu sektöründe güvenlik, otomasyon ve operasyonel verimlilik gereksinimlerinin giderek artması, akıllı algı sistemlerinin geliştirilmesini öncelikli bir araştırma konusu haline getirmiştir. Demiryolu ağlarının genişlemesi ve bu ağların akıllı ulaşım ekosistemleriyle bütünleşmesiyle birlikte, karmaşık gerçek dünya ortamlarını gerçek zamanlı olarak algılayabilme, anlayabilme ve bu ortamlara uygun şekilde tepki verebilme yeteneği, kritik bir ihtiyaç olarak öne çıkmaktadır. Bu tez çalışması, söz konusu ihtiyaca yönelik bir çözüm sunmak amacıyla gerçekleştirilmiştir. Çalışmada, dinamik ve güvenlik açısından kritik demiryolu operasyonlarının kısıtları altında çalışabilecek, nesne tespiti ve mesafe tahmini gerçekleştirebilen sağlam ve gerçek zamanlı bir görsel algı sisteminin tasarlanması ve uygulanması hedeflenmiştir.

Çalışmanın temelini oluşturan araştırma; derin öğrenme, anlamsal segmentasyon, nesne tespiti, stereo görme ve üç boyutlu (3B) mekânsal akıl yürütme gibi ileri düzey disiplinlerin kesişiminde konumlanmaktadır. Özellikle yapısal açıdan düzenli, ancak çevresel değişkenlik bakımından yüksek karmaşıklık gösteren demiryolu ortamlarında, bu alanların her biri özgün teknik zorluklar içermektedir. Bu zorluklar arasında, demiryoluna özgü açıklamalı veri kümelerinin yetersizliği, stereo görme sistemlerinden elde edilen RGB-D verilerinin senkronizasyonundaki teknik karmaşıklık ve gömülü sistemlerde gerçek zamanlı çıkarım gereksinimleri öne çıkmaktadır. Ancak, akıllı demiryolu izleme ve otomasyonuna yönelik pratik ve ölçeklenebilir bir çözüm geliştirme arzusu, bu çalışmanın her aşamasında temel motivasyon ve itici güç olmuştur.

Bu tez aynı zamanda akademik araştırma ile gerçek dünya uygulamaları arasında köprü kurma amacını taşıyan daha geniş çaplı bir çabanın parçasıdır. Disiplinli bir akademik çalışma ortamı, ileri düzey hesaplama kaynaklarına erişim imkânı ve ZED2 stereo kamera ile RailSem19 veri seti gibi gelişmiş araçların kullanıma sunulması sayesinde; gerçek dünya kısıtlarını gözetken, aynı zamanda demiryolu algı sistemlerine dair bilimsel bilgi birikimini de zenginleştiren çok görevli bir yapay zekâ çerçevesi tasarlanmış, uygulanmış ve değerlendirilmiştir.

Bu araştırma sürecinde, değerli danışmanım Sayın Prof. Dr. İlhan AYDIN'a en içten teşekkürlerimi sunarım. Kendisinin uzman rehberliği, yapıcı değerlendirmeleri ve sürekli desteği, bu çalışmanın her aşamasında yön gösterici olmuş; çalışmanın bilimsel niteliğini ve entelektüel derinliğini belirgin biçimde artırmıştır. Ayrıca, bu tezin yürütülmesinde gerekli olan akademik altyapıyı ve araştırma olanaklarını sağlayan Fırat Üniversitesi'ne ve Fırat Üniversitesi Fen Bilimleri Enstitüsü'ne teşekkür ederim. Tez çalışmasında kullanılan veri setini geliştiren RailSem19 araştırma ekibine de katkılarından ötürü şükranlarımı sunarım.

Son olarak, bu süreç boyunca her daim yanımda olan, sabırları ve destekleriyle bana güç veren aileme ve yakın arkadaşlarıma gönülden teşekkür ederim. Onların desteği, bu akademik yolculuğun her aşamasında en değerli motivasyon kaynağı olmuştur.

Muhammed Amir ELMUHAMMEDCEBBEN

ELAZIĞ, 2025

İÇİNDEKİLER

Sayfa

ÖNSÖZ.....	iv
İÇİNDEKİLER	v
ÖZET	vii
ABSTRACT	viii
ŞEKİLLER LİSTESİ	ix
TABLolar LİSTESİ	xi
EKLER LİSTESİ	xii
KISALTMALAR	xiii
1. GİRİŞ	1
2. İLGİLİ ÇALIŞMALAR	4
2.1. Demiryolu ve Ulaştırma Alanlarında Nesne Tespiti.....	4
2.2. Demiryolu ve Ulaştırma Alanlarında Anlamsal Segmentasyon	6
2.3. Çoklu Görevli Öğrenme (Multi-Task Learning - MTL) ve Model Birleştirme	9
2.4. Demiryolu ve Ulaştırma Alanlarında Stereo Görü, Mesafe Ölçümü ve Derinlik Tahmini	11
2.5. Sınırlılıklar (Limitations), Yayınlar (Publications) ve Araştırma Açıkları (Research Gaps).....	13
3. VERİ SETİ VE ÖN İŞLEME	16
3.1. Veri Seti Genel Bakışı	16
3.2. Veri Bölme Stratejisi	18
3.3. Ön İşleme ve Veri Artırımı.....	19
3.3.1. Nesne Tespiti.....	19
3.3.2. Anlamsal Segmentasyon	19
3.4. Etik ve Lisans Uygunluğu	20
4. MATERYAL VE METOT	21
4.1. Nesne Tespit Modelleri	21
4.1.1. YOLO'nun Evrimi	22
4.1.2. YOLOv8.....	23
4.1.3. YOLOv9.....	24
4.1.4. YOLOv11.....	26
4.1.5. YOLOv12.....	27
4.1.6. Faster R-CNN.....	29
4.1.7. Sahte Etiket Üretim Hattı	31
4.2. Anlamsal Segmentasyon Modelleri	31
4.2.1. UNet-Vanilla	32
4.2.2. UNet-Pretrained	33
4.2.3. DeepLabV3+	34
4.2.4. Öznitelik Piramidi Ağı (Feature Pyramid Network – FPN)	35
4.2.5. Piramit Sahne Ayırıştırma Ağı (Pyramid Scene Parsing Network – PSPNet)	37
4.3. Çoklu Görev (Multi-Task) Birleştirme Yaklaşımları	38
4.3.1. Segmentasyon Yönlendirmeli Geometrik Tespit	39
4.3.2. Çift Akışlı Tespit ve Segmentasyon Birleşmesi	40
4.3.3. Uygulama Çerçevesi	41
4.3.4. Stratejik Karşılaştırma.....	41

4.4. ZED2 Kullanılarak Mesafe Ölçümü ve Derinlik Entegrasyonu	42
4.4.1. ZED2 Kamerası ile Stereo Derinlik Tahmini	43
4.4.2. Nesne Konumlandırma İçin RGB-D Birleştirme	45
4.4.3. Sınırlayıcı Kutu Verilerinin 3B Metrik Uzaya Eşlenmesi	46
4.4.4. Gerçek Zamanlı Çoklu Görev İşlem Hattı ile Entegrasyon	47
4.4.5. Tez Hedefleriyle Uyum	48
5. DENEYSEL KURULUM	49
5.1. Donanım ve Yazılım Yapılandırması	49
5.2. Eğitim Yapılandırması	51
5.2.1. Nesne Tespiti Modelleri	52
5.2.2. Anlamsal Segmentasyon Modelleri	53
5.2.3. Derinlik Tahmini ve Kalibrasyon	54
5.3. Değerlendirme Metrikleri	54
5.3.1. Nesne Tespiti Metrikleri	55
5.3.2. Anlamsal Segmentasyon Metrikleri	56
5.3.3. Derinlik Tahmini ve Mesafe Ölçüm Metrikleri (ZED2 Kamera)	58
6. BULGULAR VE TARTIŞMA	61
6.1. Nesne Tespiti Sonuçları	61
6.2. Anlamsal Segmentasyon Sonuçları	65
6.3. Çoklu Görev Birleşim Analizi (Multi-Task Fusion Analysis)	73
6.4. ZED2 Entegrasyonu ve Mesafe Ölçümü Analizi	80
6.5. Tartışma ve Analiz	85
7. SONUÇLAR	88
7.1. Katkıların ve Sonuçların Özeti	88
7.2. Sınırlılıklar	90
7.3. Son Değerlendirme ve Sonuç	91
ÖNERİLER	92
KAYNAKLAR	93
EKLER	101
ÖZGEÇMİŞ	

ÖZET

Ray ve Çevresindeki Nesnelerin Tespiti ve Mesafe Ölçümü için Derin Öğrenme Tabanlı Yaklaşımların Geliştirilmesi

Muhammed Amir ELMUHAMMEDCEBBEN

Yüksek Lisans Tezi

FIRAT ÜNİVERSİTESİ
Fen Bilimleri Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı

Temmuz 2025, Sayfa: xiii + 100

Akıllı demiryolu sistemlerinin geliştirilmesi, giderek daha fazla sayıda görevi eşzamanlı ve gerçek zamanlı olarak gerçekleştirebilen görsel algı çerçevelerine bağlı hale gelmektedir. Bu tez, demiryolu ortamlarında güvenliği ve otomasyonu artırmak amacıyla nesne tespiti, anlamsal segmentasyon ve derinlik odaklı mesafe tahminini entegre eden modüler bir derin öğrenme sistemi önermektedir. 2B anlamsal kavrayışı 3B mekânsal akıl yürütme ile birleştirerek, sistem demiryolu operasyonlarına özel kapsamlı ve bağlama duyarlı sahne yorumlaması sağlamaktadır.

Segmentasyon ve tespit görevlerini desteklemek üzere RailSem19 veri seti, özgün bir açıklama sentezleme hattı aracılığıyla genişletilmiştir. Her alt görev için en gelişmiş modeller değerlendirilmiş; mAP, mIoU ve piksel bazında doğruluk gibi metrikler kullanılarak en iyi yapılandırmalar seçilmiştir.

İki çok görevli füzyon stratejisi geliştirilmiştir: biri segmentasyon çıktılarından yararlanan kural tabanlı bir çıkarım hattı, diğeri ise nesne düzeyinde ve piksel düzeyinde anlamsal bilgileri eşzamanlı olarak işleyen çift akışlı bir mimaridir. Çift akışlı model, karmaşık senaryolarda daha yüksek mekânsal hassasiyet, zamansal tutarlılık ve dayanıklılık göstermiştir.

Derinlik algısı, ZED2 kamera kullanılarak elde edilmiş ve gerçek zamanlı 3B konumlama mümkün kılınmıştır. Kalibre edilmiş derinlik haritalarının RGB görüntülerle hizalanması sayesinde, sistem çeşitli koşullarda ortalama %10'un altında hata oranıyla mesafe tahmini gerçekleştirebilmiş; yayaların, araçların ve engellerin güvenilir bir şekilde tespitini desteklemiştir.

Sistem, CUDA destekli donanım üzerinde verimli bir şekilde çalışarak gerçek zamanlı performansla birlikte yüksek doğruluk ve başarılı segmentasyon ölçütleri sergilemiştir. Bu sonuçlar, sistemin demiryolu izleme, tehlike tespiti ve altyapı denetimi gibi pratik uygulamalar için etkinliğini ve güvenilirliğini teyit etmektedir.

Bu çalışma, demiryolu otomasyonu için genişletilebilir bir algı çerçevesi sunmakta ve çok modlu füzyon, alan uyarlaması ve güvenlik-kritik uygulamalar alanlarında gelecekteki araştırmalar için sağlam bir temel oluşturmaktadır.

Anahtar Kelimeler: Derin Öğrenme, Anlamsal Segmentasyon, Nesne Tespiti, Stereo Görü, Çoklu Görevli Öğrenme, Mesafe Ölçümü.

ABSTRACT

Development of Deep Learning-Based Approaches for the Detection and Distance Measurement of Rails and Surrounding Objects

Muhammed Amir ELMUHAMMEDCEBBEN

Master's Thesis

FIRAT UNIVERSITY
Graduate School of Natural and Applied Sciences
Department of Computer Engineering

July 2025, Pages: xiii + 100

The development of intelligent railway systems increasingly depends on visual perception frameworks capable of executing multiple tasks concurrently and in real time. This thesis proposes a modular deep learning system integrating object detection, semantic segmentation, and depth-aware distance estimation to improve safety, situational awareness, and automation in rail environments. By combining 2D semantic understanding with 3D spatial reasoning, the system enables comprehensive and context-aware scene interpretation tailored to railway operations.

The RailSem19 dataset was extended via a novel annotation synthesis pipeline to support both segmentation and detection tasks. State-of-the-art models were benchmarked for each subtask, with top-performing configurations selected using metrics such as mean Average Precision (mAP), mean Intersection over Union (mIoU), and pixel-wise accuracy.

Two multi-task fusion strategies were developed: a rule-based inference using segmentation outputs, and a dual-stream architecture processing object- and pixel-level semantics concurrently. The dual-stream model showed improved spatial precision, temporal consistency, and robustness in complex scenarios.

Depth perception was achieved using a ZED2 stereo camera, enabling real-time 3D localization. Calibrated depth maps aligned with RGB images allowed the system to estimate distances with an average error below 10% across diverse conditions, supporting reliable detection of pedestrians, vehicles, and obstacles.

The system delivered reliable accuracy and real-time performance on CUDA-enabled hardware, demonstrating its applicability to tasks such as railway monitoring, hazard detection, and infrastructure inspection. This work introduces a flexible perception framework tailored for railway automation and establishes a foundation for future advancements in multi-modal fusion, domain adaptation, and deployment in safety-critical environments.

Keywords: Deep Learning, Semantic Segmentation, Object Detection, Stereo Vision, Multi-Task Learning, Distance Measurement.

ŞEKİLLER LİSTESİ

Sayfa

Şekil 3.1. Veri Seti Örnekleri	17
Şekil 4.1. YOLO Modellerinin Evrimi: mAP@0.95 ve Gecikme Karşılaştırması (T4 TensorRT FP16 Üzerinde) (Ultralytics, 2025)	22
Şekil 4.2. YOLOv8 Mimarisi	24
Şekil 4.3. GELAN Mimarisi: (a) CSPNet, (b) ELAN ve (c) önerilen GELAN. [77]	25
Şekil 4.4. YOLOv11 Mimarisi	27
Şekil 4.5. YOLOv12'de kullanılan R-ELAN	28
Şekil 4.6. YOLOv12 Mimarisi	29
Şekil 4.7. Faster R-CNN Mimarisi	30
Şekil 4.8. U-Net Mimarisi [78]	32
Şekil 4.9. DeepLabV3+ Mimarisi [79]	34
Şekil 4.10. FPN Mimarisi	36
Şekil 4.11. PSPNet Mimarisi	37
Şekil 4.12. ZED2 Kamera (stereolabs.com)	42
Şekil 4.13. ZED2 Kameranın Çalışma Prensibi [80]	44
Şekil 4.14. (a) Derinlik Tahmini. (b) Sol Görüntü ve Derinlik Haritası İlişkisi. [80]	45
Şekil 5.1. Google Colaboratory Logosu	49
Şekil 5.2. ZED Kameranın Kapsamlı API'sinin Çoklu Ortam Uyumluluğu	51
Şekil 6.1. Nesne Tespiti Modelleri Çıktı Örnekleri - 1	63
Şekil 6.2. Nesne Tespiti Modelleri Çıktı Örnekleri - 2	64
Şekil 6.3. Nesne Tespiti Modelleri Çıktı Örnekleri - 3	64
Şekil 6.4. Nesne Tespiti Modelleri Çıktı Örnekleri - 4	65
Şekil 6.5. Beş Segmentasyon Modeli İçin Eğitim ve Doğrulama IoU Eğrileri	68
Şekil 6.6. Anlamsal Segmentasyon Modelleri Çıktı Örnekleri - 1	69
Şekil 6.7. Anlamsal Segmentasyon Modelleri Çıktı Örnekleri - 2	70
Şekil 6.8. Anlamsal Segmentasyon Modelleri Çıktı Örnekleri - 3	71
Şekil 6.9. Anlamsal Segmentasyon Modelleri Çıktı Örnekleri - 4	72
Şekil 6.10. DeepLabV3+ Kullanarak Segmentasyon Tabanlı Geometrik Çıkarım Örneği - 1	74
Şekil 6.11. DeepLabV3+ Kullanarak Segmentasyon Tabanlı Geometrik Çıkarım Örneği - 2	74
Şekil 6.12. DeepLabV3+ Kullanarak Segmentasyon Tabanlı Geometrik Çıkarım Örneği - 3	75
Şekil 6.13. DeepLabV3+ Kullanarak Segmentasyon Tabanlı Geometrik Çıkarım Örneği - 4	75

Şekil 6.14. DeepLabV3+ Modeline Ait Eğitim ve Doğrulama Metriği Eğrileri (CE Loss, Dice Loss, mIoU, Pixel Accuracy, Mean Accuracy).....	77
Şekil 6.15. Çift Akışlı Tespit-Segmentasyon Birleşimi Örneği - 1	78
Şekil 6.16. Çift Akışlı Tespit-Segmentasyon Birleşimi Örneği - 2	79
Şekil 6.17. Çift Akışlı Tespit-Segmentasyon Birleşimi Örneği - 3	79
Şekil 6.18. DeepLabV3plus + YOLOv11 + ZED2 Örneği - 1	82
Şekil 6.19. DeepLabV3plus + YOLOv11 + ZED2 Örneği - 2	82
Şekil 6.20. DeepLabV3plus + YOLOv11 + ZED2 Örneği - 3	83
Şekil 6.21. DeepLabV3plus + YOLOv11 + ZED2 Örneği - 4	83
Şekil 7.1. YOLOv12-Normalize edilmiş Karışıklık Matrisi.....	103
Şekil 7.2. YOLOv12-F1 Eğrisi.....	104
Şekil 7.3. YOLOv12-Kesinlik-Duyarlılık Eğrileri	104
Şekil 7.4. YOLOv12-Sonuçları	105
Şekil 7.5. YOLOv11-Normalize edilmiş Karışıklık Matrisi.....	105
Şekil 7.6. YOLOv11-F1 Eğrisi.....	106
Şekil 7.7. YOLOv11-Kesinlik-Duyarlılık Eğrileri	106
Şekil 7.8. YOLOv11-Sonuçları	107
Şekil 7.9. YOLOv9-Normalize edilmiş Karışıklık Matrisi.....	107
Şekil 7.10. YOLOv9-F1 Eğrisi.....	108
Şekil 7.11. YOLOv9-Kesinlik-Duyarlılık Eğrileri	108
Şekil 7.12. YOLOv9-Sonuçları	109
Şekil 7.13. YOLOv8-Normalize edilmiş Karışıklık Matrisi.....	109
Şekil 7.14. YOLOv8-F1 Eğrisi.....	110
Şekil 7.15. YOLOv8-Kesinlik-Duyarlılık Eğrileri	110
Şekil 7.16. YOLOv8-Sonuçları	111

TABLÖLER LİSTESİ

	Sayfa
Tablo 6.1. Nesne Tespiti Sonuçları	62
Tablo 6.2. Eğitim (Training) Anlamsal Segmentasyon Sonuçları	67
Tablo 6.3. Doğrulama (Validation) Anlamsal Segmentasyon Sonuçları	67
Tablo 6.4. Test Anlamsal Segmentasyon Sonuçları.....	67
Tablo 6.5. DeepLabV3+ Modeline Ait Eğitim, Doğrulama ve Test Performansları	76



EKLER LİSTESİ

	Sayfa
Ek- 1: Temel Teknik Terimler Sözlüğü	101
Ek- 2: Nesne Tespiti Modellerine İlişkin Eğitim Kayıtları Ve Değerlendirme Çıktıları.	103



KISALTMALAR

2B	: İki Boyutlu (2D-Two-Dimensional)
3B	: Üç Boyutlu (3D- Three-Dimensional)
AP	: Ortalama Kesinlik (Average Precision)
ASPP	: Aralıklı Uzamsal Piramit Havuzlama (Atrous Spatial Pyramid Pooling)
C2PSA	: Uzamsal Dikkatli Aşamalı Kısmi Ağı (Cross-Stage Partial with Spatial Attention)
CNN	: Evrimsel Sinir Ağı (Convolutional Neural Network)
CIoU	: Birlik Üzerinde Tam Kesişim (Complete Intersection over Union)
CSPNet	: Aşamalı Kısmi Ağları (Cross-Stage Partial Networks)
CTAN	: Görevler Arası Dikkat Ağları (Cross-Task Attention Networks)
ECA	: Verimli Kanal Dikkati (Efficient Channel Attention)
ELAN	: Verimli Katman Toplama Ağları (Efficient Layer Aggregation Networks)
FCN	: Tam Evrişimli Ağı (Fully Convolutional Network)
FPN	: Özellik Piramidi Ağı (Feature Pyramid Network)
FPS	: Saniyedeki Kare Sayısı (Frames Per Second)
GELAN	: Genel Verimli Katman Toplama Ağı (Generalized Efficient Layer Aggregation Network)
GPU	: Grafik İşlem Birimi (Graphics Processing Unit)
IMU	: Atalet Ölçüm Birimi (Inertial Measurement Unit)
IoU	: Kesişim Bölü Birleşim (Intersection over Union)
MAE	: Ortalama Mutlak Hata (Mean Absolute Error)
MARE	: Ortalama Mutlak Göreli Hata (Mean Absolute Relative Error)
mAP	: Ortalama Ortalama Kesinlik (Mean Average Precision)
mIoU	: Ortalama Kesişim Bölü Birleşim (mean Intersection over Union)
MTL	: Çoklu Görevli Öğrenme (Multi-Task Learning)
PANet	: Yol Toplama Ağı (Path Aggregation Network)
PGI	: Programlanabilir Gradyan Bilgisi (Programmable Gradient Information)
PPM	: Piramit Havuzlama Modülü (Pyramid Pooling Module)
PSPNet	: Piramid Sahne Ayırıştırma Ağı (Pyramid Scene Parsing Network)
R-CNN	: Bölge Tabanlı Evrişimli Sinir Ağı (Region-based Convolutional Neural Network)
R-ELAN	: Artık Verimli Katman Toplama Ağları (Residual Efficient Layer Aggregation Networks)
RGB	: Kırmızı, Yeşil, Mavi (Red, Green, Blue)
RGB-D	: Kırmızı, Yeşil, Mavi – Derinlik (Red, Green, Blue - Depth)
RMSE	: Karekök Ortalama Kare Hata (Root Mean Square Error)
RoI	: İlgi Bölgesi (Region of Interest)
RPN	: Bölge Öneri Ağı (Region Proposal Network)
SCAM	: Sıralı Çapraz Dikkat Modülleri (Sequential Cross-Attention Modules)
SCI	: Bilimsel Atıf İndeksi (Science Citation Index)
SCIE	: Genişletilmiş Bilimsel Atıf İndeksi (Science Citation Index Expanded)
SGD	: Stokastik Gradyan İnişi (Stochastic Gradient Descent)
SPPF	: Hızlı Uzamsal Piramit Havuzlama (Spatial Pyramid Pooling Fas)
SSD	: Tek Aşamalı Çoklu Kutu Algılayıcı (Single Shot MultiBox Detector)
YOLO	: Sadece Bir Kez Bakarsın (You Only Look Once)

1. GİRİŞ

Demiryolu taşımacılığı, modern altyapının bel kemiğini oluşturarak hem insanlar hem de yükler için kentsel ve şehirlerarası ortamlarda güvenilir, enerji verimli ve maliyet etkin bir ulaşım yöntemi sunmaktadır. Küresel hareketlilik taleplerinin artmasıyla birlikte, demiryolu sistemlerinin dijitalleştirilmesi ve otomatikleştirilmesine yönelik paralel bir eğilim söz konusudur. Bu çabalar; işletme güvenliğini artırma, insan hatasını azaltma, kestirimci bakım sağlama ve tam ya da yarı otonom seyri destekleme gibi gereksinimlerle yönlendirilmektedir. Bu gelişmelerin merkezinde, farklı işletme koşulları altında gerçek zamanlı çevresel algılama gerçekleştirebilen akıllı görsel algı sistemlerinin geliştirilmesi yer almaktadır.

Otonom sürüş literatüründe kapsamlı şekilde çalışılmış olan karayolu tabanlı algı sistemlerinden farklı olarak, demiryolu ortamları özgün zorluklar sunmaktadır. Bu zorluklar arasında son derece doğrusal ve tekrar eden sahne geometrileri; seyrek ancak güvenlik açısından kritik öneme sahip ray kenarı personeli, altyapı bileşenleri veya enkaz gibi nesne sınıfları; ayrıca bitki örtüsü, gölgeler, yetersiz aydınlatma ve değişken hava koşulları gibi çevresel engeller bulunmaktadır. Ayrıca makaslar, sinyal direkleri, hemzemin geçitler ve katener hatları gibi demiryoluna özgü altyapı bileşenleri, geleneksel trafik sahnelerinden farklı özel görsel tanıma yetenekleri gerektirmektedir.

Sadece nesne tespiti (Object Detection) veya anlamsal segmentasyon (Semantic Segmentation) gibi tek görevli geleneksel bilgisayarla görme (Computer Vision) sistemleri, bu kadar karmaşık işletme alanları için yetersiz kalmaktadır. Bu sistemler genellikle görevler arası bağlamsal ipuçlarını bütünleştirme yetisinden yoksundur ve yalnızca 2B - İki Boyutlu (2D - Two-Dimensional) görüntü verilerine dayanmanın getirdiği sınırlamalara sahiptir. Bu nedenle, bu sistemler derinlik tahmini (Depth Estimation) yapamaz veya mekânsal yakınlık değerlendirmesi gerçekleştiremez; oysa bu yetenekler, engel tespiti, ray ihlali uyarıları ve tepki kontrolü gibi uygulamalar için hayati öneme sahiptir. Bu sınırlamaları aşmak amacıyla hem anlamsal segmentasyon hem de nesne tespiti gerçekleştirebilen ve stereo görüşten (Stereo Vision) elde edilen derinlik bilgilerini de bütünleştiren çok görevli derin öğrenme (Deep Learning) mimarilerine yönelik ilgi artmaktadır.

Bu tez çalışması, akıllı demiryolu sahne algılaması için derin öğrenme tabanlı bütünleşik bir çerçevenin geliştirilmesini araştırmaktadır. Önerilen sistem, üç temel yeteneği birleştirmektedir: sahnenin piksel düzeyinde ayrıştırılması için anlamsal segmentasyon, kritik varlıkların tanımlanması için nesne tespiti ve gerçek dünya mesafe ölçümü (Distance Measurement) için stereo tabanlı derinlik tahmini. Stereo kamera sisteminden elde edilen RGB-D (Red, Green, Blue - Depth) görüntülerinin entegrasyonu sayesinde, bu çerçeve çevresel farkındalığı artırmakta ve gerçek zamanlı uygulamalara uygun mekânsal temelli akıl yürütmeyi mümkün kılmaktadır.

Bu tezde ele alınan araştırma problemi, demiryoluna özgü nesneleri gerçek zamanlı kısıtlar altında eşzamanlı olarak segmentlere ayırabilen, tespit edebilen ve konumlandırabilen bir algı sisteminin tasarımı ve uygulanmasıdır. Sistem, gerçek dünyadaki karmaşık koşullarda güvenilir bir şekilde çalışmalı ve mobil veya gömülü uygulamalara uygun donanım yapılandırmalarıyla uyumlu olmalıdır. Bu hedefe ulaşmak için araştırma; çok görevli öğrenme teknikleri, gerçek zamanlı çıkarım stratejileri ve engel konumlandırması için geometrik son işlem hatlarını kapsamaktadır.

Bu araştırmanın dört temel amacı bulunmaktadır. Birincisi, yapılandırılmış demiryolu ortamlarına özgü, ileri düzey anlamsal segmentasyon ve nesne tespiti modellerini tasarlamak ve karşılaştırmaktır. İkincisi, stereo kamera (özellikle ZED2 sistemi) kullanılarak elde edilen RGB-D verisini bütünleştirerek derinlik farkındalığına sahip çıkarımı mümkün kılmaktır. Üçüncüsü, segmentasyon ve tespiti eşzamanlı olarak gerçekleştirirken gerçek zamanlı performansı koruyan çok görevli bir füzyon mimarisi önermek ve değerlendirmektir. Son olarak, sistemin etiketlenmiş veri kümeleri ve gerçek dünya video dizileri ile saha benzeri koşullarda doğrulamasını yapmaktır.

Bu araştırma, birkaç temel hipotez tarafından yönlendirilmiştir. Segmentasyon ve tespit için birlikte eğitilen çok görevli bir derin öğrenme modelinin, yalnızca tek bir göreve odaklanan modellere kıyasla doğruluk ve hesaplama verimliliği açısından demiryolu sahnelerinde üstün performans göstereceği öngörülmektedir. Ayrıca, stereo görüşün bütünleştirilmesiyle, yalnızca RGB (Red, Green, Blue) verisiyle çalışan modellere kıyasla nesne konumlandırma ve mesafe tahmini yeteneklerinin önemli ölçüde artacağı hipotezi öne sürülmektedir. Son olarak, optimize edilmiş işlem hatları ve hafif mimariler kullanılarak gerçek zamanlı çıkarımın doğruluk kaybı olmadan gerçekleştirilebileceği varsayılmaktadır.

Bu tez çalışması, yer üstü demiryolu ortamlarına odaklanmakta ve görsel algı için RGB ile RGB-D görüntülerini kullanmaktadır. Kullanılan birincil veri kümesi hem segmentasyon hem de tespit görevlerinin karşılaştırılması için etiketli görüntüler sunan RailSem19 veri kümesidir. Algı sisteminin GPU (Graphics Processing Unit) tabanlı iş istasyonlarında gerçek zamanlı çalışacak şekilde tasarlanmasına rağmen, gömülü platformlara konuşlandırılması gelecek çalışmalarda ele alınması planlanan bir yön olarak belirlenmiştir. Bu çalışmada, tek bir veri kümesinin kullanılması, uçtan uca örnek bazlı segmentasyon veya nesne takibinin gerçekleştirilmemesi ve etiketli gerçek derinlik haritalarının bulunmaması nedeniyle derinlik tabanlı işlemlerde sezgisel ayarlamalara dayanılması gibi bazı sınırlamalar da kabul edilmektedir.

Bu tezin katkıları; algoritmik, sistem düzeyinde ve akademik alanlarda yer almaktadır. İlk olarak, demiryolu alanına özgü olarak geliştirilmiş ve değerlendirilmiş özgün çok görevli derin öğrenme mimarileri sunulmaktadır. İkinci olarak, algı çerçevesine stereo görüşün entegre edilmesiyle, gerçek zamanlı derinlik tahmini ve iyileştirilmiş engel tespiti doğruluğu sağlanmıştır. Üçüncü olarak, bu çalışmanın çıktıları iki hakemli yayında yer almıştır: biri ulusal bir ulaştırma mühendisliği dergisinde kabul edilmiş, diğeri ise uluslararası bir SCI/SCIE (Science Citation Index/

Science Citation Index Expanded) indeksli dergiye sunulmuştur. Bu yayınlar, araştırmanın teknik titizliğini ve özgünlüğünü yansıtmaktadır. Son olarak, sistemin gerçek zamanlı çalışabilirliği ZED2 stereo kamera kullanılarak gerçekleştirilen deneysel doğrulamalarla ortaya konmuştur.

Bu tez aşağıdaki şekilde yapılandırılmıştır: 2. Bölüm, anlamsal segmentasyon, nesne tespiti, çok görevli öğrenme ve demiryolu algısı konularındaki mevcut yaklaşımların ayrıntılı bir literatür taramasını sunmaktadır. 3. Bölüm, deneylerde kullanılan veri kümelerini, etiketleme formatlarını ve ön işleme hatlarını açıklamaktadır. 4. Bölüm, önerilen yöntemleri; model mimarileri, çok görevli öğrenme stratejileri ve stereo görüşle entegrasyon dahil olmak üzere ortaya koymaktadır. 5. Bölüm, deneysel düzeni; donanım ve yazılım ortamı, eğitim yapılandırması ve değerlendirme metrikleriyle birlikte açıklamaktadır. 6. Bölüm, model performansı, füzyon çıktıları ve gerçek zamanlı konuşlandırma sonuçlarına ilişkin bulguları ve analizleri sunmaktadır. Son olarak, 7. Bölüm, sonuçların özetini ve az örnekli öğrenme, gömülü sistemlere konuşlandırma ve uzamsal-zamansal sahne anlayışı gibi gelecekteki araştırma yönlerini ortaya koymaktadır.

2. İLGİLİ ÇALIŞMALAR

Bu bölüm, akıllı demiryolu ortamları bağlamında görsel algı sistemlerine ilişkin mevcut literatürün sistematik ve eleştirel bir incelemesini sunmaktadır. İnceleme, bu tezin temel bileşenlerine karşılık gelen beş tematik başlık altında yapılandırılmıştır. Bölüm 2.1 ve Bölüm 2.2, ulaştırma sistemleri ve demiryolu odaklı senaryolardaki uygulamalarına özel bir vurgu yaparak, temel bilgisayarla görme görevleri olan nesne tespiti ve anlamsal segmentasyon konularını ele almaktadır. Bu bölümlerde, derin öğrenmeye dayalı yöntemlerin gelişimi izlenmekte; gerçek zamanlı çıkarım, yüksek çözünürlüklü anlamsal anlayış ve bağlamsal modelleme alanlarındaki güncel ilerlemelere dikkat çekilmektedir.

Bölüm 2.3, çok görevli öğrenme (MTL) ve model birleştirme stratejilerindeki gelişmeler ele alınmakta; özellikle demiryolu altyapısında yaygın olan kaynak kısıtlı kurulum koşulları altında model verimliliği, genelleme yetkinliği ve görevler arası bağımlılıkların artırılmasına olan katkıları vurgulanmaktadır. Bölüm 2.4 ise stereo görüş ve monoküler derinlik kestirimi yaklaşımlarını incelemekte; bu tekniklerin güvenlik açısından kritik demiryolu ortamlarında mekânsal çıkarım ve üç boyutlu algılamaya yönelik kullanımına odaklanmaktadır.

Bölüm 2.5, alandaki mevcut sınırlamalar sentezlenmekte, demiryolu odaklı görsel sistemlere dair literatür eleştirel bir bakışla değerlendirilmektedir. Ayrıca, bu tezin katkılarını motive eden temel bilgi boşlukları belirlenmektedir. Genel olarak bu bölüm, araştırma sorularının, yöntemsel tercihlerinin ve sonraki bölümlerde sunulan yeniliklerin kuramsal ve teknik bağlamda temellendirilmesini sağlayacak bir zemin oluşturmaktadır.

2.1. Demiryolu ve Ulaştırma Alanlarında Nesne Tespiti

Nesne tespiti görevi, son on yılda önemli ilerlemeler kaydetmiş olup, bu gelişmeler büyük ölçüde otonom sistemlere ve akıllı ulaşım çözümlerine yönelik artan talepler tarafından tetiklenmiştir. Demiryolu taşımacılığı bağlamında nesne tespiti hem sabit altyapı unsurlarının hem de yayalar, araçlar ve ray üzerindeki veya çevresindeki yabancı cisimler gibi dinamik engellerin tanımlanması yoluyla işletme güvenliğinin sağlanmasında hayati bir rol oynamaktadır. Genel ortamlardan farklı olarak demiryolu sahaları, yüksek doğruluk ve düşük gecikmeyle gerçek zamanlı çalışabilecek özel tespit stratejilerini gerektiren özgün görsel ve geometrik kısıtlamalar sunmaktadır.

Derin öğrenme tabanlı nesne tespitinde ilk büyük atılımlar, özellikle R-CNN (Region-based Convolutional Neural Network) ailesi olmak üzere iki aşamalı mimarilerle gerçekleşmiştir [1]. Fast R-CNN [2] ve Faster R-CNN [3], [4], bölge önerisi ve sınıflandırma için giderek daha verimli mekanizmalar sunmuş; Faster R-CNN, uçtan uca optimizasyonu mümkün kılan Bölge Öneri Ağı

(Region Proposal Network - RPN) bileşeniyle öne çıkmıştır [5]. Bu modeller, çok ölçekli nesnelerin yüksek doğrulukla tespit edilmesinin gerektiği senaryolarda oldukça etkili olmuştur. Ancak, modern GPU sistemlerinde dahi bu modellerin çıkarım gecikmeleri, demiryolu otomasyonu gibi zaman hassasiyetinin kritik olduğu alanlardaki gerçek zamanlı uygulamalar için optimal olmaktan uzaktır [6].

Gerçek zamanlı gereksinimlere yanıt olarak, SSD (Single Shot MultiBox Detector) [7], [8] ve YOLO (You Only Look Once) [9], [10] serisi gibi tek aşamalı dedektörler, daha verimli alternatifler olarak ortaya çıkmıştır. SSD, çoklu ölçeklerde evrimsel filtreler uygulayarak nispeten hızlı ve doğru tespit imkânı sunsa da dağınık ortamlarda küçük ya da kısmen gizlenmiş nesnelerin tespitinde zorlanabilmektedir. Buna karşılık, YOLO tabanlı modeller doğrudan gerçek zamanlı konuşlandırma için tasarlanmıştır [6]. Bu modeller, nesne sınıflarını ve sınırlayıcı kutuları tek bir ileri geçişte tahmin ederek, hız ve performans arasında özellikle güvenlik açısından kritik alanlarda cazip bir denge sunmaktadır.

YOLO mimarilerinin YOLOv3'ten YOLOv8'e evrimi, mimari düzeyde sürekli iyileştirmeleri ve performans artışlarını yansıtmaktadır. Özellikle YOLOv7, saniyede çerçeve bazında çalışırken %56,8'e varan ortalama kesinlik (AP) elde ederek hem SSD hem de Faster R-CNN'i dinamik senaryolarda geride bırakmış ve doğruluk-hız dengesinde alanın en iyisi olduğunu göstermiştir [6], [11]. YOLOv8, dikkat mekanizmaları ve uyarlanabilir evrim modüllerinin entegrasyonu sayesinde, küçük ölçekli veya kısmen gizlenmiş nesnelerin tespitinde performans artışı sağlayarak özellikle istasyonlar, geçitler ve makas bölgeleri gibi karmaşık demiryolu alanları için daha güçlü hale gelmiştir [12], [13].

Daha yakın tarihlerde geliştirilen deneysel mimariler olan YOLOv11 [14], [15] ve YOLOv12 [16], [17], hesaplama grafiği optimizasyonu, dikkat yönlendirmeli özellik birleştirme ve çözünürlüğe duyarlı tahmin başlıkları gibi yenilikler sunmuştur. Hâlen akademik ve endüstriyel araştırmalarda test edilmekte olan bu modeller, doğruluk ve çıkarım hızı açısından umut verici iyileştirmeler göstermektedir. Bu tez kapsamında, YOLOv11 modeli, demiryolu odaklı veri kümeleri üzerinde hem ortalama kesinlik (mAP@0.5) hem de tespit sağlamlığı bakımından en yüksek performansı sergilemiş; önceki sürümleri geride bırakarak yapılandırılmış ancak seyrek nesne içeren ortamlarda daha iyi genelleme yeteneği ortaya koymuştur.

Karayolu temelli akıllı ulaşım sistemlerinde yaygınlaşmaları artmasına rağmen, kentsel trafik sahneleri için geliştirilen nesne tespit modelleri demiryolu ortamlarına doğrudan uygulanabilir değildir. Bunun temel nedeni, demiryolu altyapısına özgü alan ve açığa bağlı zorluklardır. Özellikle uzayıp giden ve tekrar eden ray geometrileri, nesnelerin dağılımında görülen sınıf dengesizlikleri ve kritik altyapı unsurlarının karmaşık mekânsal yerleşimleri, bu alanda özelleştirilmiş model mimarileri ile dikkatle tasarlanmış eğitim stratejilerini zorunlu kılmaktadır. Ayrıca, gerçek dünyadan elde edilen demiryolu verilerinin genellikle sınırlı sayıda örnek içermesi,

bu verilerin yalnızca belirli sınıfları kapsayacak şekilde etiketlenmiş olması ve farklı çevresel koşullar altında toplanmış bulunması, mevcut nesne tespit modellerinin doğrudan demiryolu ortamına aktarılmasını daha da güçleştirmektedir.

Bu alan farkını kapatmak amacıyla bazı çalışmalar demiryoluna özgü tespit stratejileri geliştirmiştir. Örneğin, bu çalışma yenilikçi bir şekilde olay tespiti ile yabancı cisimlerin şiddetine ilişkin bilgileri birleştiren bir erken uyarı yaklaşımı önermektedir. Geliştirilen yöntem, gelişmiş YOLOv5 modeli kullanılarak cisimleri tespit edebilmekte ve iyileştirilmiş demiryolu tespit yöntemi aracılığıyla hat alanını çıkarabilmektedir. Tehlikeli cisimler belirlenmekte ve engelin konumu ile şiddetine göre uyarı seviyesi tayin edilmektedir [18]. Bir diğer dikkat çekici katkı ise Guindel ve arkadaşlarının [19] çalışmasıdır. Bu çalışmada, nesne konumu ve yönelimi hakkında eşzamanlı çıkarım yapan çok görevli sistemlerin avantajları ortaya konulmuş; bu yaklaşım, özellikle ray-araç etkileşimlerinin anlaşılması açısından önemlidir.

Bu çalışmalar kayda değer ilerlemeler sunmakla birlikte, literatürde hâlen demiryolu ortamlarının yapısal ve operasyonel kısıtlarına özel olarak optimize edilmiş bütünlüklü bir nesne tespit çerçevesi eksiktir. Mevcut modellerin büyük çoğunluğu ya karayolu sahneleri için uyarlanmış [20] ya da yalnızca engel tespiti [21], [22], [23] veya ray segmentasyonu [24] gibi dar kapsamlı problem tanımlarıyla sınırlı kalmıştır. Bu nedenle, sahne türleri arasında genelleyebilen, ancak hesaplama verimliliğinden ve gerçek zamanlı çalışabilirlikten ödün vermeyen bütüncül ve alana uyarlanmış tespit sistemlerine duyulan ihtiyaç açıktır.

Bu tez kapsamında geliştirilen nesne tespit bileşeni, bu ihtiyaca yanıt olarak YOLOv11 modelini daha geniş kapsamlı birçok görevli algı sistemine entegre etmektedir. Bu sistem, demiryolu sahnelerinin görsel özelliklerine özel olarak tasarlanmış olup, nesne düzeyinde çıkarımı anlamsal segmentasyon ve derinlik farkındalığına sahip konumlandırma ile birlikte gerçekleştirmektedir. Bu görevlerin birleşik bir mimari içerisinde bütünleştirilmesi, otonom ve yarı otonom demiryolu platformları için ölçeklenebilir ve yorumlanabilir görsel algıya doğru önemli bir adımı temsil etmektedir.

2.2. Demiryolu ve Ulaştırma Alanlarında Anlamsal Segmentasyon

Anlamsal segmentasyon, bir görüntüdeki her bir piksele kategorik bir etiket atama süreci olup, otonom sistemler için görsel algı görevlerinde temel bir unsur haline gelmiştir. Ulaşım ortamlarında, özellikle demiryolu alanlarında, anlamsal segmentasyon yoğun sahne anlayışı ve altyapı bileşenlerinin ayrıntılı şekilde tanımlanmasını mümkün kılarak kritik bir rol oynamaktadır. Nesne tespitinden farklı olarak, yalnızca sınırlayıcı kutular ve kaba nesne konumları sağlayan tespit sistemlerinin aksine, anlamsal segmentasyon raylar, peronlar, ray kenarı sinyalleri ve görüşü engelleyen nesneler gibi sınıfların piksel düzeyinde sınırlarının çizilmesine olanak tanır. Bu ayrıntı

seviyesi, ray doluluk tahmini, engel kaçınma ve ilgi bölgesi önceliklendirmesi gibi ardıl uygulamalar için hayati önem taşır [25], [26].

Derin öğrenme tabanlı modern anlamsal segmentasyonun temeli, geleneksel tam bağlantılı katmanlar yerine evrişimsel işlemler kullanan ve böylece mekânsal olarak hizalanmış tahmin haritaları üretebilen Tam Evrişimli Ağı (Fully Convolutional Network - FCN) ile atılmıştır [27], [28]. FCN'ler, uçtan uca eğitilebilen segmentasyon ağları açısından paradigma değişimi yaratmış olsa da, bağlamsal bilgi birleştirme ve çok ölçekli özellik temsili konularında zayıf kalmış; bu da karmaşık ulaşım sahnelerinde etkinliklerini sınırlamıştır. Bu sınırlama, hiyerarşik özellik çıkarımı ve bağlam farkındalıklı çözümleme bileşenlerini entegre eden daha güçlü mimarilerin geliştirilmesine yol açmıştır.

Bu alanda öne çıkan en etkili modellerden biri, başlangıçta biyomedikal görüntü analizi [29] için geliştirilen UNet'tir [30], [31], [32], [33]. UNet, simetrik bir kodlayıcı-çözücü (encoder-decoder) mimarisi sunmuş ve karşılıklı katmanlar arasında atlama bağlantıları kullanarak, modelin ince mekânsal özellikleri korurken bağlamsal bilgileri geri kazanmasına olanak tanımıştır. Basitliği ve uyarlanabilirliği sayesinde UNet ve türevleri, kentsel ve ulaşım odaklı sahne segmentasyonu görevlerinde yaygın olarak kullanılmaktadır. Bu çalışmada iki sürüm değerlendirilmiştir: temel bir UNet (UNet-Vanilla) ve ImageNet üzerinde önceden eğitilmiş ResNet34 kodlayıcı kullanan transfer öğrenmesi (Transfer learning) tabanlı bir varyant (UNet-Pretrained). Temel model, yapılandırılmış demiryolu sahnelerinde kabul edilebilir bir genelleme gösterirken; önceden eğitilmiş varyant daha yüksek doğruluk ve sağlamlık elde etmiş, kodlayıcı düzeyinde aktarım öğrenmesinin faydasını doğrulamıştır [34], [35].

Özellikle karmaşık veya görsel olarak belirsiz ortamlarda bağlamsal farkındalığı artırmak amacıyla, sonrasında geliştirilen mimariler küresel bağlam modülleri ve çok ölçekli özellik birleştirme yaklaşımlarını içermiştir. Örneğin, Piramid Sahne Ayrıştırma Ağı (Pyramid Scene Parsing Network - PSPNet), farklı alımsal alanlarda (receptive field) bağlamsal bilgiyi toplayan piramit havuzlama modülünü entegre ederek, heterojen sahnelerde sınıf ayrımını iyileştirmiştir [33], [36]. Bir diğer önemli adım olan DeepLabV3+, aralıklı uzamsal piramit havuzlama (Atrous Spatial Pyramid Pooling - ASPP) ve çözümleyici iyileştirme yolu (decoder refinement path) ile alımsal alanı artırmış ve segmentasyon sınırı hassasiyetini geliştirmiştir. DeepLabV3+, özellikle ResNet50 gibi yüksek kapasiteli omurga ağlarla birleştirildiğinde performans ve hesaplama maliyeti arasında denge kurma yeteneği sayesinde, ulaşım odaklı anlamsal segmentasyonda yaygın bir standart haline gelmiştir [37], [38], [39], [40]. Ayrıca, MobileNetV2 gibi hafif omurgalarla eşleştirildiğinde DeepLabV3+, gerçek zamanlı performansını koruyarak %84,3 ortalama piksel doğruluğuna (mPA) ve 41,8 FPS (Frames Per Second)'ye ulaşmakta; bu da farklı kullanım senaryolarına uyarlanabilirliğini ortaya koymaktadır [41].

Bu tez kapsamında gerçekleştirilen karşılaştırmalı deneylerde DeepLabV3+, çoğu segmentasyon metriğinde rekabetçi sonuçlar vermiştir; ancak ortalama kesişim birlik (mIoU) skorunda, Feature Pyramid Network (FPN) tarafından marjinal olarak geride bırakılmıştır. Başlangıçta nesne tespiti için geliştirilen FPN, çok ölçekli özellikleri birleştirmek için yanıl bağlantılar ve yukarıdan aşağıya yolları etkin şekilde kullanır; bu da onu demiryolu sahnelerinde mekânsal olarak dağılmış ve geometrik olarak düzensiz yapıların tespiti için özellikle uygun kılar [33], [42], [43]. Bu çalışmada değerlendirilen ResNet34 tabanlı FPN konfigürasyonu, en yüksek mIoU ve piksel düzeyinde doğruluk skorlarını elde etmiş; raylar gibi uzatılmış nesneler ile ince ayrıntılara sahip altyapı sınıflarının farklı görsel koşullar altında segmentasyonunda etkinliğini göstermiştir.

Bu yüksek kapasiteli modellere ek olarak, gömülü sistemler ve kaynak kısıtlı ortamlar için Fast-SCNN [44], [45] ve BiSeNet [46], [47], [48] gibi hafif, gerçek zamanlı segmentasyon mimarileri önerilmiştir. Bu modeller, sadeleştirilmiş omurga yapıları ve agresif aşağı örnekleme stratejileri kullanarak yüksek çıkarım hızlarına ulaşmakta, çoğu zaman saniyede 100 kareyi aşabilmektedir [49]. Ancak, gerçek zamanlı performanslarına rağmen bu modeller, hassas sınır koruma ve sınıf düzeyinde ayrıştırma gerektiren senaryolarda genellikle yetersiz kalmakta; her iki gereksinim de güvenlik açısından kritik demiryolu uygulamaları için vazgeçilmezdir. Bu nedenle, bu tür modellerin demiryolu alanındaki uygulanabilirliği sınırlıdır. DDRNet, çift çözünürlüklü bir mimari kullanarak 102 FPS hızında %77,4 ortalama IoU (Intersection over Union) başarımı elde eden ve bu sayede gerçek zamanlı yol sahnesi algılama uygulamaları için uygun bir model olarak öne çıkmaktadır [20]. Öte yandan, dönüştürücü tabanlı modülleri entegre eden SegFormer [50], makul çıkarım süreleriyle yüksek segmentasyon doğruluğu sunarak özellikle otonom sürüş bağlamında avantaj sağlayan etkileyici bir orta yol sunmaktadır [51].

Segmentasyon modellerinin demiryolu ortamlarında konuşlandırılması, aynı zamanda bir dizi alana özgü zorluğu da beraberinde getirir. Bu zorluklar arasında aşırı sınıf dengesizliği (örneğin arka plan ile nadir nesneler arasındaki fark), bitki örtüsü veya hava koşullarına bağlı engellemeler ve etiketlenmiş veri setlerinin sınırlı çeşitliliği yer almaktadır. Bu sorunları hafifletmek amacıyla ağırlıklı kayıp fonksiyonları, veri artırma (augmentation) ve yardımcı kenar tabanlı denetim [48] gibi teknikler önerilmiştir. Ancak, bu yöntemler hâlâ yeterince araştırılmamış olup, bu tezde deneysel analizlerin temelini oluşturan RailSem19 [52] gibi demiryolu odaklı kıyaslamalar üzerinde kapsamlı şekilde doğrulanmamıştır.

Her ne kadar genel ulaşım senaryolarında anlamsal segmentasyon konusunu inceleyen çalışmalar artmakta olsa da, demiryollarının incelikli gereksinimlerine odaklanan araştırmalar görece az sayıdadır. Mevcut çalışmalar genellikle yalnızca tek görevli segmentasyona odaklanmakta, sık ya da hazır (off-the-shelf) ağlar kullanmakta ve nesne tespiti ya da mekânsal konumlandırma gibi tamamlayıcı görevlerle bütünleştirme içermemektedir. Bu durum, özellikle

güvenli seyrüsefer için bütünleşik sahne anlayışının gerekli olduğu otonom demiryolu sistemleri bağlamında önemli bir araştırma boşluğunu temsil etmektedir.

Bu tez çalışması, bu boşluğu sistematik olarak yüksek performanslı segmentasyon modellerini çok görevli bir çerçeveye entegre ederek ele almaktadır. Bu yaklaşım, stereo kamera sisteminden elde edilen RGB-D girdilerini kullanarak, derinlik farkındalıklı çıkarımla anlamsal anlayışı zenginleştirmekte ve piksel düzeyinde hem anlamsal olarak zengin hem de geometrik olarak temellendirilmiş sahne analizine olanak tanımaktadır. Bu yöntem, konuşlandırılabilir ve gerçek zamanlı algı sistemleri geliştirmeye yönelik uzun vadeli hedefle uyumludur ve akıllı demiryolu altyapısı için somut bir katkı sunmaktadır.

2.3. Çoklu Görevli Öğrenme (Multi-Task Learning - MTL) ve Model Birleştirme

Çoklu görevli öğrenme (MTL), derin öğrenme alanında güçlü bir paradigma olarak ortaya çıkmış ve paylaşılan temsilleri kullanarak tek bir modelin birden fazla ilişkili görevi eşzamanlı olarak öğrenip gerçekleştirmesine olanak tanımıştır. Bilgisayarla görme bağlamında, anlamsal segmentasyon, nesne tespiti, derinlik tahmini ve kenar tespiti gibi görevlerde başarıyla uygulanmış; genelleme, verimlilik ve performansta dikkate değer iyileşmeler sağlamıştır. Özellikle kaynak kısıtlı ya da gecikmeye duyarlı uygulamaların yer aldığı otonom demiryolu algısı gibi akıllı ulaşım sistemleri çerçevesinde, MTL işlem yüklerini birleştirmenin yanı sıra bağlamsal sahne anlayışını ve operasyonel sağlamlığı artırma potansiyeli sunmaktadır.

MTL'nin temelinde, görevler arası fayda sağlayan özelliklerin öğrenildiği temsili paylaşım ilkesi yer almaktadır. En yaygın mimari strateji olan sert (hard) parametre paylaşımında, giriş verisini işleyen paylaşımlı bir omurga (backbone) kullanılırken, her görev için ayrı özgül (task-specific) başlıklar (heads) alt görevleri gerçekleştirir [53]. Alternatif olarak, her görev için ayrı modellerin korunduğu ancak parametre benzerliğini teşvik eden yumuşak (soft) paylaşım yaklaşımları da mevcuttur [54]. Hibrit yaklaşımlar arasında, paylaşımlı kodlayıcı-çözücü (encoder-decoder) tasarımlar, hesaplamanın yeniden kullanımını görev uzmanlaşmasıyla dengelediği için yaygınlaşmıştır. Bu duruma örnek olarak [55] çalışmada tanımlanan, bilgi akışını teşvik ederken görevler arası paraziti engelleyen ara çapraz bağlantılarla yapılandırılmış çift görevli çözümleyiciler gösterilebilir.

Daha yeni uyarlanabilir MTL çerçeveleri ise dinamik özellik paylaşımı üzerine odaklanmıştır. [54] çalışmada sunulan model, görev uyumluluğuna bağlı olarak katman bazlı paylaşımı belirleyen öğrenilebilir ilkeler kullanmaktadır. Bu tür dinamik paylaşım stratejileri, görevler arası gereksinim farklılıklarını ele almak ve eğitim verimliliğini artırmak açısından oldukça etkilidir. Dikkat mekanizmaları da bu alanda ilgi görmüştür; Sıralı Çapraz Dikkat Modülleri (Sequential Cross-Attention Modules - SCAM) [56] ve Görevler Arası Dikkat Ağları (Cross-Task Attention Networks - CTAN) [57] gibi yöntemler, anlamsal hizalama ve görev

önceliğine bağlı olarak görevler arası bilgi yönlendirmesini seçici biçimde gerçekleştirir. [56] ve [57] gibi çalışmalarda bu mimarilerin, görevler arası sinerjileri kullanırken görev özgü öğrenme yollarını koruyabildiği gösterilmiştir [58].

Bununla birlikte, MTL'nin uygulanması, özellikle uzamsal ya da istatistiksel özellikleri oldukça farklı görevler söz konusu olduğunda, kolay değildir. İyi belgelenmiş zorluklardan biri negatif transferdir; burada bir görev, çakışan gradyanlar veya orantısız öğrenme sinyalleri nedeniyle diğer bir görevin öğrenimini olumsuz etkileyebilir. Bu sorunu hafifletmek için gradyan yansıtma [59] ve kalibrasyon gibi optimizasyon düzeyinde teknikler geliştirilmiştir [60], [61]. [62] çalışmada sunulan MoCoGrad yöntemi, geçmiş görev çakışmalarına dayalı olarak gradyan katkılarını dinamik şekilde ayarlayarak daha dengeli ve kararlı bir yakınsama sağlamaktadır. Bu tür yöntemler, özellikle segmentasyon gibi yoğun (dense) öngörü görevleri ile nesne tespiti gibi seyrek (sparse) görevleri bir araya getirme durumlarında kritik önem taşır.

Anlamsal segmentasyon ve nesne tespitinin birleştirilmesi, hem yoğun hem de örnek düzeyinde sahne anlayışı gerektiren görsel sistemlerde artan ilgi görmektedir. Buna klasik bir örnek, Faster R-CNN üzerine paralel bir segmentasyon dalı ekleyen Mask R-CNN'dir [63], [64]. Daha yeni hibrit yapılar arasında ise kodlayıcı-çözücü temelli HybridNets yer almaktadır. Bu mimariler, tespit ve segmentasyonu tek bir işlem hattında birleştirerek hem çıkarım verimliliğini hem de bağlamsal akıl yürütmeyi iyileştirir [55]. Ancak bu çalışmaların çoğu, nesne dağılımı ve uzamsal dinamikleri demiryolu ortamlarından oldukça farklı olan yol tabanlı otonom sürüş senaryolarına odaklanmaktadır.

Demiryolu algısı bağlamında, alana uyarlanmış MTL modelleri tasarlamaya yönelik çalışma sayısı oldukça azdır. Dikkate değer bir istisna, [65] çalışmadır. Bu çalışmada, demiryolu temelli engel tespiti ve altyapı ayrıştırmasına yönelik olarak özelleştirilmiş çift çözümleyicili bir mimari önerilmiştir. Çerçeve, görev özgü çözümleyicilerle desteklenen paylaşımlı bir kodlayıcı kullanmakta ve görevler arası özellik ayrımını artırmak için sıkıştırma-uyarma (squeeze-and-excitation) modüllerinden faydalanmaktadır. Bu tür yaklaşımlar umut verici olsa da, çok görevli sistemlerin demiryolu alanında sistematik olarak doğrulanması hâlâ sınırlıdır.

Bu tez çalışması, çoklu görevli çerçeve içinde iki tamamlayıcı birleştirme stratejisi sunarak bu temel kavramlar üzerine inşa edilmiştir. Birinci strateji, yalnızca segmentasyon temelli bir çıkarım hattı olup, raylar ve yükseltilmiş yüzeyler gibi sınıfa özgü maskelerden engel konumlarını sezgisel olarak türetmekte ve böylece sınırlayıcı kutu etiketlemeye olan ihtiyacı ortadan kaldırmaktadır. İkinci strateji ise çift akışlı bir mimaridir: DeepLabV3+ anlamsal segmentasyonu, YOLOv11 ise nesne tespitini paralel biçimde gerçekleştirir. Her iki modelin çıktıları sonradan işlenerek mekânsal olarak hizalanır ve çevreye dair birleşik bir anlayış elde edilir. “Multi-Task Deep Learning Framework for Real-Time Semantic Segmentation and Object Detection in Railway Environments” başlıklı çalışmada tanıtılan bu strateji, konuşlandırma esnekliği sağlamakta ve

piksel düzeyindeki bilgi ile örnek düzeyindeki bilgiyi birleştirerek sahne yorumlanabilirliğini artırmaktadır.

Demiryolu uygulamalarında çoklu görevli öğrenme pek çok avantaj sunar. Algı modüllerinin birleştirilmesiyle hesaplama tekrarı azaltılır, görevler arası denetim yoluyla bağlamsal öğrenme geliştirilir ve RGB ile derinlik verisi gibi tamamlayıcı modelliklerin entegrasyonu için doğal bir zemin sunar. Ayrıca, etiketli verinin sınırlı olduğu senaryolarda, görevler arası paylaşımlı denetim genelleme yetisini artırabilir ve aşırı öğrenmeyi önleyebilir. Ancak, bu tür sistemlerin uygulanması bazı zorlukları da beraberinde getirir. Görev özgü gereksinimlerin yönetilmesi mimari karmaşıklığı önemli ölçüde artırır ve segmentasyon (genellikle yoğun) ile tespit (genellikle seyrek) görevleri arasındaki dengesizlikler optimizasyon davranışlarında dengesizliğe yol açabilir. Bu durum, demiryolu sahnelerinde özellikle belirgindir; çünkü segmentasyon haritasında raylar gibi geniş arka plan bölgeleri hâkimken, yayalar veya döküntüler gibi küçük ve nadir nesneler tespit etiketlerinde yetersiz temsil edilebilir [21].

Bu bağlamda, çok görevli modellerin etkili şekilde eğitilmesi; kayıp ağırlıklandırma stratejileri, omurga paylaşım politikaları ve görev yakınsamasını dengeleyecek eğitim takvimleri gibi unsurların dikkatle tasarlanmasını gerektirir [65], [66], [67]. Ayrıca, bu tezde olduğu gibi stereo görme ve derinlik farkındalıklı çıkarım entegre edildiğinde, özellikle gerçek zamanlı çıkarım hatlarında geometrik özelliklerin anlamsal çıktılarla hizalanması daha da karmaşık hâle gelir.

Özetle, çoklu görevli öğrenme ve model birleştirme, akıllı demiryolu algı sistemlerinin geliştirilmesi açısından umut vadeden yönelimler sunmaktadır. Dikkatli bir şekilde uygulandığında, bu yaklaşımlar hem anlamsal hem de örnek düzeyinde sahne anlayışını eşzamanlı olarak modellemeyi mümkün kılar, hesaplama verimliliğini artırır ve karmaşık operasyonel ortamlarda dayanıklılığı güçlendirir. Stereo tabanlı mekânsal akıl yürütme ile birleştirildiğinde, bu çerçeveler modern otonom ve yarı otonom demiryolu platformlarının algı gereksinimlerini karşılamak için son derece uygundur.

2.4. Demiryolu ve Ulaştırma Alanlarında Stereo Görü, Mesafe Ölçümü ve Derinlik Tahmini

Nesne tespiti ve anlamsal segmentasyon gibi iki boyutlu görsel algı görevlerinde önemli ilerlemeler kaydedilmiş olmakla birlikte, yalnızca RGB görüntülerine dayalı geleneksel sistemler, geometrik akıl yürütme ve mekânsal farkındalık gerektiren senaryolarda yapısal sınırlamalarla karşı karşıya kalmaktadır. Özellikle güvenliğin ön planda olduğu ulaşım uygulamalarında, demiryolu gibi kritik alanlarda çevredeki nesnelerin derinliğinin ve mesafesinin doğru bir şekilde algılanması büyük önem taşımaktadır. Bu tür bir mekânsal farkındalık, gerçek zamanlı tehlike tespiti, engel önleme ve proaktif karar alma süreçlerinin etkin bir şekilde yürütülmesini sağlar. Nesnelerin yakınlık düzeyinin doğru biçimde tahmin edilmesi, algı sistemlerinin zamanında uyarı

vermesine olanak tanımakta ve bu sayede frenleme ya da rota değişikliği gibi düzeltici eylemlerin başlatılması mümkün olmaktadır. Böylelikle, operasyonel güvenlik ve sistemin genel güvenilirliği önemli ölçüde artırılmaktadır.

Stereo görme, daha pahalı ve donanım açısından yoğun olan LiDAR gibi aktif teknolojilere alternatif olarak, derinlik tahmini için pasif ve müdahalesiz bir çözüm olarak ortaya çıkmıştır. Senkronize sol ve sağ kamera görüntüleri arasındaki farkı (disparite) analiz ederek, stereo sistemler 3B - Üç Boyutlu (3D - Three-Dimensional) mekânsal bilgiyi kodlayan yoğun, piksel düzeyinde derinlik haritaları üretebilir. Aktif sensörlerin aksine, stereo görme daha geniş çevresel koşullarda çalışabilir ve nokta bulutu kaydı ya da dışsal kalibrasyon gerektirmez [68]. Bu yönüyle, görüşe dayalı ulaşım sistemleri için ölçeklenebilir ve entegre edilebilir bir yaklaşım sunar.

Son yıllarda, stereo tabanlı derinlik algılama hem otonom sürüşte hem de robotik navigasyonda önemli bir ivme kazanmıştır. Örneğin, [69] çalışmada, ardışık derinlik tahminleri arasında zamansal tutarlılığı sağlamak için akış güdümlü bellek birimleri önerilmiştir. Bu yöntem, dinamik dış ortamlarda mekânsal süreklilik ve dayanıklılığı artırmıştır. Benzer şekilde, [70] çalışmada, LiDAR ve RGB verilerinden elde edilen anlamsal önbilgilerin entegrasyonu ile stereo eşleme doğruluğu ve sahne segmentasyon performansının özellikle örtüşen ya da düşük dokulu bölgelerde belirgin şekilde iyileştirildiği gösterilmiştir.

Demiryolu algısı bağlamında, stereo görmenin derin öğrenme çerçevelerine entegrasyonu hâlâ gelişmekte olan ancak umut vadeden bir alandır. Örneğin, [71] çalışmada, evrişimli sinir ağları ile stereo disparite tahminini birleştiren hibrit bir mimari önerilmiş, bu sayede demiryolu altyapısının anlamsal yerleştirilmesi ve derinlik tahmini sağlanmıştır. Bu yaklaşım, %0,8 ortalama keskinlik (mAP) ve %10'un altında derinlik tahmin hatası ile stereo görmenin, nesne yakınlığının kritik güvenlik parametresi olduğu sınırlı demiryolu ortamlarında etkili olduğunu ortaya koymuştur.

Ampirik yöntemlerin yanı sıra, stereo tabanlı algı modellerinin geliştirilmesini hızlandırmak için sentetik ve hibrit veri kümeleri de kullanılmaktadır. Örneğin, ESRORAD veri kümesi, 3B sınırlayıcı kutularla etiketlenmiş gerçek dünya ve simülasyon karışımı demiryolu görüntüleri içermekte olup, derinlik farkındalığına sahip tespit algoritmalarının denetimli eğitimi için olanak sağlar [72].

Benzer şekilde, RailEnV-PASMVS [73] ve TrainSim [74] gibi simülasyon platformları, derinlik haritası sentezi için foto-gerçekçi, stereo kalibreli raylı senaryolar sunar. Ancak, sentetik verilere aşırı bağımlılık, özellikle değişken ışık ve hava koşullarında, gerçek dünyaya genelleştirme yetisini kısıtlayabilen alan önyargılarını beraberinde getirebilir.

Bu tezde önerilen algı boru hattının temel bileşeni olarak ZED2 stereo kamera sistemi kullanılmaktadır. ZED2, senkronize RGB stereo çiftlerini yakalamakta ve yerleşik stereo eşleme algoritmaları ile gerçek zamanlı derinlik bilgisi üretmektedir. ZED SDK, piksel düzeyinde derinlik

haritalama, kamera konum tahmini ve mekânsal filtreleme gibi yerleşik yetenekler sunarak ulaşım algı görevleri için uygun bir platform sağlar [75]. Bu çalışmada, ZED2 sisteminden elde edilen derinlik çıktısı, anlamsal segmentasyon ve nesne tespiti çıktıları ile birleştirilerek, tespit edilen ya da segmentlenen nesnelerin kamera çerçevesine göre gerçek dünya mesafesi tahmin edilmektedir. Bu sayede, yakın alan ve uzak alan nesneleri arasında ayırım yapabilen, riske duyarlı sahne yorumları oluşturulabilmektedir.

Bu entegrasyonun özgül bir uygulaması, yüksek lisans araştırmam kapsamında Demiryolu Mühendisliği Dergisi'nde yayımlanan “Demiryolu Ortamında Nesne Tespiti için Derin Öğrenme Yöntemlerinin Geliştirilmesi ve Zed Kamerası ile Mesafe Ölçümü” başlıklı makalede sunulmaktadır. Bu çalışmada, stereo görüntülerden elde edilen derinlik haritaları, segmentasyona dayalı engel konumlandırma ve sınıflandırma süreçlerini geliştirmek amacıyla kullanılmaktadır. Ortaya çıkan çerçeve, harici sensörlere veya yoğun derinlik etiketlemelerine ihtiyaç duymadan bağlamsal algılama imkânı sağlamaktadır. Bu tür yetenekler, insan, hayvan ve ekipman gibi seyrek ancak güvenlik açısından kritik nesne sınıflarının bulunduğu demiryolu ortamlarında özellikle değerlidir.

Bununla birlikte, stereo görmenin bazı sınırlamaları da vardır. Doku yoksun bölgelerde, düşük ışık koşullarında ya da örtüşme sınırlarında disparite tahmin kalitesi düşebilir ve bu durum gürültülü ya da eksik derinlik haritalarına yol açabilir. Yansıtıcı yüzeyler ve çevresel parazitler de doğruluğu olumsuz etkileyebilir. Bu zorlukların üstesinden gelmek için, son çalışmalar stereo görmeyi LiDAR, radar veya GPS gibi diğer modelliklerle birleştiren hibrit yaklaşımlar önermektedir. Ancak bu çok modlu sistemler genellikle yüksek donanım gereksinimleri ve karmaşık kalibrasyon süreçleri nedeniyle, hafif ya da gömülü uygulamalar için uygun olmayabilir [76].

Buna rağmen, stereo görme, raylı sistemlerde derinlik tahmini için etkili ve ölçeklenebilir bir yaklaşımdır. Mevcut 2B evrişimli ağlarla uyumluluğu, pasif algılama yapısı ve gerçek zamanlı işlem kapasitesi, demiryolu otomasyonu için cazip bir çözüm haline getirir. Bu tezde geliştirilen çok görevli öğrenme çerçevesine stereo görmenin entegrasyonu, yoğun anlamsal ipuçlarının derinlik farkındalığı ile birleşerek, akıllı demiryolu uygulamalarına özgü 3B sahne anlayışının pratikte uygulanabilirliğini ortaya koymaktadır.

2.5. Sınırlılıklar (Limitations), Yayınlar (Publications) ve Araştırma Açıkları (Research Gaps)

Önceki derleme, akıllı ulaşım sistemleri alanında nesne tespiti, anlamsal segmentasyon, çok görevli öğrenme ve stereo görme konularında kaydedilen önemli ilerlemeleri ortaya koymuştur. Ancak bu teknikler karayolu tabanlı otonom sürüşte olgunlaşmış olsalar da, demiryolu ortamlarına adaptasyonları hâlâ tamamlanmış değildir. Yapısal ve güvenliğin kritik olduğu raylı sistemlerde

uygulanabilir, sağlam ve gerçek zamanlı algı sistemlerinin geliştirilmesini engelleyen çeşitli kalıcı sınırlılıklar ve yeterince araştırılmamış alanlar mevcuttur.

Mevcut çalışmaların temel sınırlamalarından biri, yaygın olarak kullanılan nesne tespiti modellerinin alanlar arası aktarılabilirliğidir. [11] ve [12] çalışmalarda sunulan YOLOv7 ve YOLOv8 gibi algoritmalar, kentsel araç ortamlarında güçlü performans sergilemelerine rağmen, demiryolu altyapısını tanımlayan seyrek, tekrarlayan ve doğrusal geometrilere genellemede zorlanmaktadır. Ray kenarındaki ekipmanlar, sinyal direkleri ve yayalar gibi kritik nesne sınıfları nadiren görülmekte ve sıklıkla kısmen örtülmektedir; bu da tespit doğruluğunda ve güven skorlarında önemli bir düşüşe neden olmaktadır. Bu sınırlamalar, ayrıntılı sınıf düzeyinde etiketlenmiş demiryolu veri kümelerinin azlığı nedeniyle daha da derinleşmektedir.

Benzer şekilde, DeepLabV3+, PSPNet ve FPN gibi anlamsal segmentasyon modelleri büyük ölçekli kentsel veri kümelerinde yüksek doğruluk göstermelerine rağmen, demiryolu ortamlarında sağlamlıkları yeterince araştırılmamıştır. Sınıf dengesizliği, nesne sınırlarının düşük çözünürlüğü ve görsel olarak karmaşık sahneler gibi demiryoluna özgü zorluklar, model performansını azaltabilmektedir. UNet gibi kodlayıcı-çözücü mimarileri, mekânsal çözünürlükte iyileşme sunmakta ve önceden eğitilmiş kodlayıcılarla birlikte kullanıldığında umut vaat etmektedir; ancak, bu modellerin bağlamsal akıl yürütme yetenekleri, karmaşık sahnelerde genellikle yetersiz kalmaktadır.

Mimari entegrasyon açısından, çok görevli öğrenme, tespit ve segmentasyon görevlerini ortak şekilde optimize ederek hesaplama tekrarını azaltma ve model sinerjisini artırma konusunda teorik olarak sağlam bir çözüm sunmaktadır. Ancak, demiryolu algısında bu yöntemin uygulanabilirliği, gradyan çatışması, görev baskınlığı ve yoğun ve seyrek çıktılar arasında denge kurma gibi henüz çözülmemiş pratik sorunlarla sınırlıdır. AdaShare [55], PCGrad [61] ve MoCoGrad [62] gibi gelişmiş optimizasyon stratejileri, genel amaçlı çok görevli ayarlarda umut vaat etse de, stereo tabanlı ya da derinlik destekli demiryolu uygulamalarındaki etkinlikleri büyük ölçüde araştırılmamıştır.

Stereo görme yoluyla gerçekleştirilen derinlik tahmini, özellikle ZED2 stereo kamera gibi ileri düzey cihazlar kullanıldığında, sistemlere üç boyutlu mekânsal akıl yürütme kapasitesi kazandırmaktadır. Ancak, stereo eşleme algoritmaları çevresel koşullardan önemli ölçüde etkilenebilmektedir. Düşük dokulu yüzeyler, yoğun gölgeler, yansıtıcı malzemeler ve yetersiz aydınlatma gibi faktörler, bu algoritmaların performansını olumsuz yönde etkileyebilmektedir. Söz konusu çevresel zorluklar, dış mekân demiryolu koridorlarında oldukça yaygın bir şekilde gözlemlenmektedir. Buna ek olarak, kamuya açık, yüksek kaliteli ve derinlik etiketlerine sahip demiryolu veri kümelerinin yetersizliği, stereo tabanlı algı sistemlerinin hem standardizasyonunu hem de karşılaştırmalı değerlendirmesini zorlaştırmakta; dolayısıyla bu sistemlerin uygulamaya hazır olma düzeyini belirsiz hâle getirmektedir.

Tüm bu sınırlamalar, bazı önemli araştırma boşluklarını gözler önüne sermektedir. Birincisi, tespit, segmentasyon ve mesafe tahminini demiryolu ortamları için özel olarak optimize edilmiş birleşik bir mimaride entegre eden algı çerçevelerine net bir ihtiyaç vardır. İkincisi, stereo görme, LiDAR’a erişilebilir ve ölçeklenebilir bir alternatif sunsa da, raylı sistemler gibi yapılandırılmış alanlar için tasarlanmış gerçek zamanlı çok görevli boru hatlarına nadiren entegre edilmektedir. Üçüncüsü, özellikle RGB-D verilerini içeren alanlara özgü veri kümeleri üzerinde kapsamlı kıyaslamalar yapılması, bu modellerin gerçek dünya koşulları altındaki uygulanabilirliğini doğrulamak açısından kritik öneme sahiptir.

Bu tezde sunulan araştırma, bu boşluklara doğrudan yanıt vererek anlamsal segmentasyon, nesne tespiti ve stereo tabanlı derinlik akıl yürütmesini birleştiren çok görevli derin öğrenme mimarilerini önermekte ve değerlendirmektedir. İlk katkı, “Demiryolu Ortamında Nesne Tespiti için Derin Öğrenme Yöntemlerinin Geliştirilmesi ve Zed Kamerası ile Mesafe Ölçümü (Developing Deep Learning Methods for Object Detection and Using Zed Camera for Distance Measurement in Railway Environment)” adlı çalışmada tanımlanmıştır. Bu çalışmada, stereo kaynaklı derinlik haritalarıyla desteklenen bir tespit boru hattı sunularak mekânsal yerleştirme iyileştirilmiştir. Bu temelin üzerine inşa edilen ikinci çalışma olan “Demiryolu Ortamlarında Gerçek Zamanlı Anlamsal Segmentasyon ve Nesne Tespiti için Çok Görevli Derin Öğrenme Çerçevesi (Multi-Task Deep Learning Framework for Real-Time Semantic Segmentation and Object Detection in Railway Environments)” ise, DeepLabV3+ ve YOLOv11 modellerini ZED2 tabanlı derinlik çıkarımıyla senkronize eden ikili akışlı bir füzyon modeli sunmakta; böylece mekânsal farkındalığa sahip bütüncül bir sahne anlayışı elde edilmektedir. Bu iki çalışma birlikte, bu tezin deneysel ve yöntemsel çekirdeğini oluşturmaktadır.

Sonuç olarak, literatür, karayolu otomasyonu için geliştirilen algı sistemlerinin, demiryolu ortamlarının yapısal ve operasyonel karmaşıklıklarına yeterince uyarlanmadığı parçalı bir tabloyu ortaya koymaktadır. Bu tez, ölçeklenebilir, derinlik duyarlı ve gerçek zamanlı görsel algı modellerini, demiryolu altyapı izleme ve güvenlik otomasyonu gibi alanlara özgü gereksinimlere göre uyarlayarak bu boşluğu kapatmayı amaçlamaktadır. Bu doğrultuda yalnızca akademik söyleme katkı sağlamakla kalmayıp, aynı zamanda yeni nesil akıllı ulaşım sistemleri için pratik araçların geliştirilmesine de katkıda bulunmaktadır.

3. VERİ SETİ VE ÖN İŞLEME

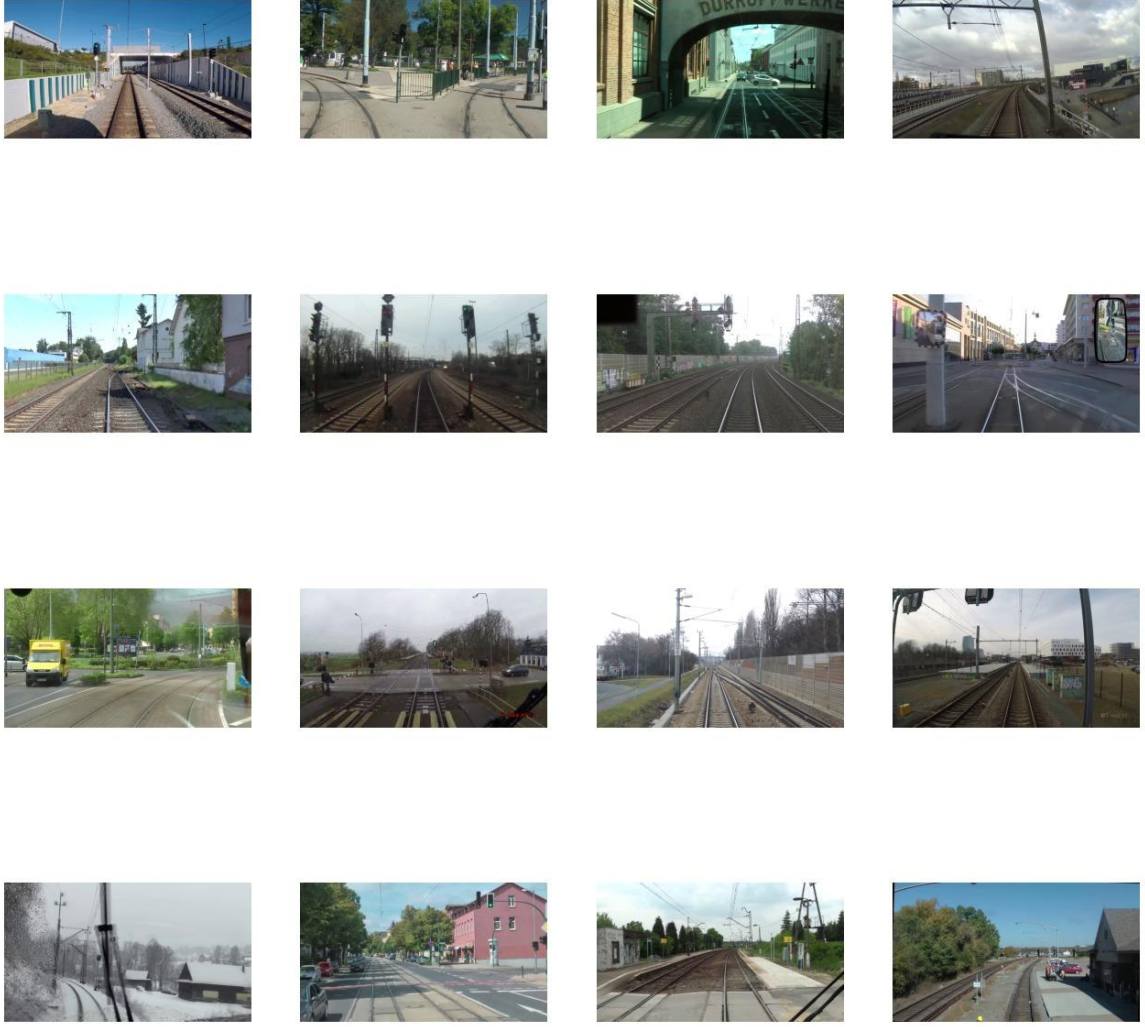
Bu bölümde, çalışmada önerilen derin öğrenme modellerinin eğitimi ve değerlendirilmesinde temel alınan veri setleri, veri bölme yöntemleri, ön işleme süreçleri ve etik hususlar ayrıntılı bir biçimde ele alınmaktadır. Demiryolu ortamlarında görsel algıya özgü sorunlar arasında yer alan sınıf dengesizliği, yapısal seyreklik ve çevresel değişkenlik gibi faktörler dikkate alındığında; sağlam ve genellenebilir bir model performansı elde edebilmek amacıyla veri setinin titizlikle seçilmesi ve uygun biçimde ön işlenmesi büyük önem arz etmektedir.

Bölüm 3.1'de, RailSem19 veri seti anlamsal zenginliği, açıklama formatı ve hem anlamsal segmentasyon hem de nesne tespiti görevlerine sağladığı katkılar açısından kapsamlı bir biçimde incelenmiştir. Bölüm 3.2'de, eğitim, doğrulama ve test alt kümeleri arasında veri sızıntısını önlemek ve adil bir değerlendirme süreci sağlamak amacıyla benimsenen veri bölme stratejileri açıklanmıştır. Bölüm 3.3, her görev türüne özgü olarak uygulanan ön işleme teknikleri ile veri artırma stratejilerini ayrıntılandırmakta; girdi verilerinin standartlaştırılması, sınıf dengesizliklerinin azaltılması ve modelin eğitim kararlılığının artırılması yönünde alınan kararları vurgulamaktadır. Son olarak, Bölüm 3.4'te veri setinin etik kullanımı ve lisans koşullarına uygunluğu ele alınmakta; veri kullanımının kurumsal araştırma etiği çerçevesinde yürütüldüğü güvence altına alınmaktadır.

Bu bölümde açıklanan tüm adımlar, izleyen derin öğrenme deneyleri için güvenilir, tekrarlanabilir ve sistematik bir veri işleme hattı oluşturulmasına katkı sağlamaktadır.

3.1. Veri Seti Genel Bakışı

Bu çalışmada, demiryolu ortamlarında anlamsal sahne anlayışı ve görsel algı görevlerini desteklemek amacıyla tasarlanmış, kamuya açık ve alana özgü bir kıyaslama veri seti olan RailSem19 kullanılmıştır. İlk olarak Zendel ve arkadaşları tarafından 2020 yılında sunulan RailSem19 [52], Almanya'daki çeşitli kentsel ve banliyö demiryolu güzergâhlarında faaliyet gösteren raylı taşıtlardan öne bakan açılardan toplanmış 8.500 yüksek çözünürlüklü RGB görüntü içermektedir. Veri seti; mevsimsel değişkenlik, örtüşmeler, değişken aydınlatma koşulları ve karmaşık altyapı geometrileri gibi çeşitli koşulları yansıtarak, derin öğrenme modellerinin gerçekçi dağıtım senaryolarında değerlendirilmesi için uygun bir ortam sunmaktadır. Şekil 3.1, RailSem19 veri setine ait örnek sahneleri görsel olarak sunmakta ve veri setinin çeşitliliği ile zorluk düzeyine dair temsili bir özet sağlamaktadır.



Şekil 3.1. Veri Seti Örnekleri

RailSem19'daki her bir görüntü, 19 ön tanımlı sınıfı kapsayan yoğun piksel seviyesinde anlamsal etiketlerle açıklanmıştır. Bu sınıflar arasında demiryolu hatları, yollar, taşıtlar, bitki örtüsü, yayalar, trafik işaretleri ve üst yapılar yer almaktadır. Etiketleme şeması, karayolu temelli veri setlerinden transfer öğrenmeye olanak tanıyacak şekilde Cityscapes sınıflandırma sistemi temel alınarak oluşturulmuş, ancak demiryollarına özgü ek kategorilerle genişletilmiştir. Anlamsal maskeler PNG formatında sunulmakta olup, sınıf değerleri tutarlı indekslerle eşleştirilmiştir.

Bu tez kapsamında veri seti, iki temel bilgisayarla görme görevi için uyarlanmıştır: anlamsal segmentasyon ve nesne tespiti. Segmentasyon görevi için tüm 19 özgün kategori korunmuş, böylece modellerin piksel bazlı sınıf sınırlarını öğrenebilmesi için yalnızca minimum ön işleme gerekmiştir. Buna karşın, nesne tespiti görevinde anlamsal maskelerin sınırlayıcı kutu (bounding box) açıklamalarına dönüştürülmesi gerekmiştir. Bu amaçla, bazı birleştirilmiş anlamsal sınıflar için kural tabanlı bir maske-kutu dönüşüm boru hattı uygulanmış, bağlantılı bileşenler çıkarılmış

ve demiryolu operasyonel güvenliğiyle ilişkili nesne kategorileri belirlenmiştir. Özellikle şu beş tespit sınıfı tanımlanmıştır:

- Ray Hattı (Track): Raylar, makaslar ve altyapı elemanlarını içerir.
- Trafik İşaret ve Sinyalleri: Yönlendirici ve düzenleyici görsel öğeleri kapsar.
- Araçlar: Tramvaylar, trenler ve kesişen karayolu araçlarını içerir.
- Engeller (Obstacles): Raylar üzerindeki ya da yakınındaki enkaz, nesne veya yabancı materyalleri temsil eder.
- İnsan (Person): Ray hattı üzerinde veya yakınında bulunan yayaları kapsar.

Sınırlayıcı kutular, OpenCV kütüphanesinin kontur algılama algoritmaları kullanılarak elde edilmiş; ardından sınıf üyeliği ve minimum alan eşikleri temelinde filtrelenerek açıklamaların doğruluğu ve kalitesi güvence altına alınmıştır. Bu işlem, RailSem19 veri setinin özgün yapısında sınırlayıcı kutu açıklamaları bulunmamasına rağmen, veri setinin nesne tespiti amaçlı eğitim süreçlerinde kullanılabilir hale getirilmesini sağlamıştır.

Mekânsal farkındalığı artırmak amacıyla, deneysel sistem mimarisine ZED2 stereo kamera görüntüleri entegre edilmiştir. ZED2 sistemi, stereo ayrıklık (disparity) temelli derinlik haritaları ile senkronize RGB stereo görüntü çiftleri sağlayarak, segmentasyon ve nesne tespiti süreçlerinde derinlik destekli analizlerin gerçekleştirilmesine olanak tanımaktadır. Her ne kadar bu gerçek zamanlı veriler doğrudan RailSem19 veri setinin bir parçası olmasa da, uygun bir biçimde dönüştürülerek veri formatı ile uyumlu hale getirilmiş ve sonraki bölümlerde ayrıntılı olarak açıklanan çıkarım ve dağıtım aşamalarında kullanılmıştır.

3.2. Veri Bölme Stratejisi

Model mimarileri ve görevler arasında adil karşılaştırma yapılabilmesini sağlamak için sağlam bir veri bölme stratejisi benimsenmiştir. RailSem19 veri seti, katmanlı örnekleme (stratified sampling) yöntemi kullanılarak karşılıklı dışlayıcı üç alt kümeye ayrılmıştır:

- Eğitim Seti (Training Set): %70 (5.950 örnek).
- Doğrulama Seti (Validation Set): %15 (1.275 örnek).
- Test Seti (Test Set): %15 (1.275 örnek).

Nadir sınıfların, özellikle güvenlik açısından kritik etiketlerin (örneğin yayalar ve sinyaller) tüm alt kümelerde benzer dağılıma sahip olması için katmanlama uygulanmıştır. Semantik sızıntıyı önlemek ve genelleme başarımını artırmak amacıyla, aynı mekânsal ya da zamansal diziden gelen görüntülerin yalnızca bir alt kümeye atanmasına dikkat edilmiştir.

Tüm veri bölmeleri tutarlılık açısından doğrulanmıştır. Bu kapsamda; maske-görüntü hizalaması, sınıf etiketlerinin bütünlüğü ve açıklamaların tamlığı denetlenmiştir. Özellikle görsel

olarak örtüşebilen sınıflarda (örneğin sinyaller ve direkler, raylar ve kaldırımlar) etiketleme hataları veya belirsizlikler açısından segmentasyon maskeleri elle gözden geçirilmiştir.

3.3. Ön İşleme ve Veri Artırımı

Anlamsal segmentasyon ve nesne tespiti boru hatlarının farklı gereksinimleri dikkate alınarak, eğitim ve değerlendirme aşamaları için ayrı ön işleme adımları uygulanmıştır.

3.3.1. Nesne Tespiti

YOLO ve Faster R-CNN gibi nesne tespiti modelleri, daha büyük giriş boyutları ve sınır kutusu açıklamaları gerektirmektedir. RGB görüntüler, 640×640 piksel boyutuna yeniden boyutlandırılmış ve Ultralytics YOLO ile Faster R-CNN yapılandırmalarıyla uyumlu hale getirilmiştir. Sınır kutuları, daha önce açıklanan maske-kutu dönüşüm işlem hattı kullanılarak çıkarılmış; (x, y, genişlik, yükseklik) değerleri, YOLO formatına uygun şekilde normalize edilerek kaydedilmiştir. Uygulanan artırmalar şunlardır:

- Mozaik artırımı (Mosaic augmentation): Dört rastgele görüntünün birleştirilmesiyle mekânsal bağlam çeşitliliği.
- HSV renk artırımı (HSV color augmentation): Renk değişkenliğine karşı dayanıklılığın artırılması.
- Rastgele çevirme ve yeniden ölçekleme (Random flipping and scaling): Bakış açısı ve nesne boyutlarındaki farklılıkların simülasyonu.

Artırma işlemleri, Ultralytics ve Albumentations API'leri kullanılarak uygulanmış; sınır kutuları, dönüştürülen görüntülerle otomatik olarak eşleştirilmiştir. Ön işleme işlem hattı, artırılmış örneklerin görsel olarak incelenmesi ve hata kontrolü yoluyla doğrulanmıştır.

3.3.2. Anlamsal Segmentasyon

Tüm segmentasyon modelleri için, RGB görüntüler ve karşılık gelen maskeler sabit 256×256 piksel çözünürlüğe yeniden boyutlandırılmıştır. Görüntüler, önceden eğitilmiş kodlayıcı omurgalarıyla uyumluluğu sağlamak amacıyla ImageNet ortalama ve standart sapma değerleriyle normleştirilmiştir. Girdiler, kayan noktalı PyTorch tensörlerine dönüştürülmüş ve $[0, 1]$ aralığına ölçeklenmiştir. Maskeler, sınıf sınırlarının bozulmaması için en yakın komşu enterpolasyonu (nearest-neighbor interpolation) kullanılarak yeniden boyutlandırılmıştır.

Segmentasyon maskeleri, 64-bit tamsayı tensörlerine dönüştürülmüş ve geçersiz pikseller için ayrılmış 255 indeks değeri kullanılarak yapılandırılmıştır. Bu, eğitim sırasında çapraz entropi kayıp fonksiyonunun (cross-entropy loss function) doğru şekilde uygulanabilmesi açısından gereklidir. Uygulanan artırmalar şunlardır:

- Rastgele yatay çevirme (Random horizontal flipping) (olasılık = 0.5).
- Rastgele kırpma ve doldurma (Random cropping and padding) ile mekânsal çeşitlilik.
- Renk titreşimi (Color jittering): Parlaklık, kontrast ve doygunluk üzerinde oynanarak ışık değişimlerinin simülasyonu.

Bu artırmalar Albumentations kütüphanesi kullanılarak uygulanmış ve görüntüler ile maskelere senkronize şekilde gerçekleştirilmiştir.

3.4. Etik ve Lisans Uygunluğu

Bu çalışmada kullanılan RailSem19 veri seti, Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0) kapsamında lisanslanmıştır. İlgili lisans, veri setinin ticari olmayan akademik araştırmalarda kullanılmasına olanak sağlamakta; uygun atıf yapılması ve oluşturulan türev çalışmaların benzer lisans koşulları altında paylaşılması şartını getirmektedir. Bu tez kapsamında gerçekleştirilen tüm veri kullanımı, söz konusu lisans hükümleri ile tam uyum içinde yürütülmüştür. Veri seti üzerinde geliştirilen tüm türevsel uygulamalar yalnızca akademik araştırma amaçlı olup, herhangi bir ticari faaliyeti içermemektedir. Geliştirilen yöntemler, yalnızca lisansın izin verdiği sınırlar dahilinde değerlendirilmiş ve kullanılmıştır.

RailSem19 veri seti, kişisel olarak tanımlanabilir herhangi bir bilgi içermemektedir. Veri setini oluşturan görüntüler, kamuya açık raylı sistem altyapılarında elde edilmiş olup; bireylerin yüz verileri, biyometrik göstergeleri veya özel mülklere ait tanımlayıcı öğeler içermemektedir. Bu yönüyle çalışma, hem Fırat Üniversitesi Bilimsel Araştırma ve Yayın Etiği Kurulu yönergelerine hem de ulusal ve uluslararası etik araştırma ilkelerine tam olarak uygundur. Deneysel süreçlerde insan katılımcılar, hayvanlar veya hassas nitelikte dış kaynak verileri kullanılmamıştır.

Tez kapsamında sunulan tüm modelleme çalışmaları, deneysel analizler, görselleştirme çıktıları ve yazılımsal bileşenler; akademik dürüstlük, bilimsel şeffaflık ve tekrarlanabilirlik ilkeleri doğrultusunda tasarlanmış, belgelendirilmiş ve açıkça tanımlanmıştır. Çalışmaya özgü katkılar, örneğin maskeleyme-kutu dönüşüm modülleri ve çoklu veri füzyonunu gerçekleştiren birleşik işlem hatları, ilgili bölümlerde ayrıntılı biçimde açıklanmış ve yazılım altyapısı kapsamında erişilebilir şekilde sunulmuştur.

4. MATERYAL VE METOT

Bu bölüm, çalışmada önerilen çok görevli algı çerçevesinin teknik temelini ve mimari tasarımını sunmaktadır. Yöntemin genel amacı, karmaşık demiryolu ortamlarında nesne tespiti, anlamsal segmentasyon ve mesafe tahmini görevlerini entegre bir biçimde ve gerçek zamanlı olarak gerçekleştirebilen bir sistem geliştirmektir. Bu doğrultuda, Bölüm 4.1, kullanılan nesne tespiti modellerini tanıtmakta; YOLO ailesine ait son teknoloji mimariler ile iki aşamalı bir algılayıcı olan Faster R-CNN modelini kapsamaktadır.

Bölüm 4.2, demiryolu yapısal bileşenlerinin piksel düzeyinde anlaşılmasına yönelik değerlendirilen anlamsal segmentasyon ağlarını ayrıntılandırmakta; hem hafif yapılı hem de bağlamsal açıdan zengin kodlayıcı-çözücü (encoder-decoder) tasarımlarına yer vermektedir. Bölüm 4.3, tespit ve segmentasyon çıktılarının bütünlük bir işlem hattı içinde birleştirilmesine yönelik incelenen çok görevli birleşim stratejilerini açıklamakta; kural tabanlı yaklaşımlar ile çift akışlı füzyon mekanizmalarını vurgulamaktadır. Bölüm 4.4 ise, üç boyutlu mekânsal çıkarım için derinlik odaklı akıl yürütmeyi mümkün kılan stereo görmenin ZED2 kamerası aracılığıyla entegrasyonunu açıklamakta; RGB-D verisinin kalibrasyonu, işlenmesi ve hizalanmasına odaklanmaktadır. Bu bileşenlerin tümü, demiryolu otomasyonuna özgü zorluklara yanıt verebilecek şekilde tasarlanmış, gerçek zamanlı çalışmaya uygun, sağlam ve modüler bir algı sistemi inşa etmek için yöntemsel temeli oluşturmaktadır.

4.1. Nesne Tespit Modelleri

Nesne tespiti, bu tez kapsamında geliştirilen görsel algı hattının temel bileşenlerinden birini oluşturur ve demiryolu ortamlarında karmaşık güvenlik-kritik varlıkların tanımlanması, konumlandırılması ve sınıflandırılmasını mümkün kılar. Yaya, araç, ray üzerindeki engeller ve altyapı elemanları gibi nesnelerin doğru ve gerçek zamanlı olarak tespit edilmesi, demiryolu taşımacılığı alanında operasyonel güvenliğin korunması, durumsal farkındalık sağlanması ve erken tehlike müdahalesi için otomatik sistemlerin geliştirilmesinde kritik öneme sahiptir.

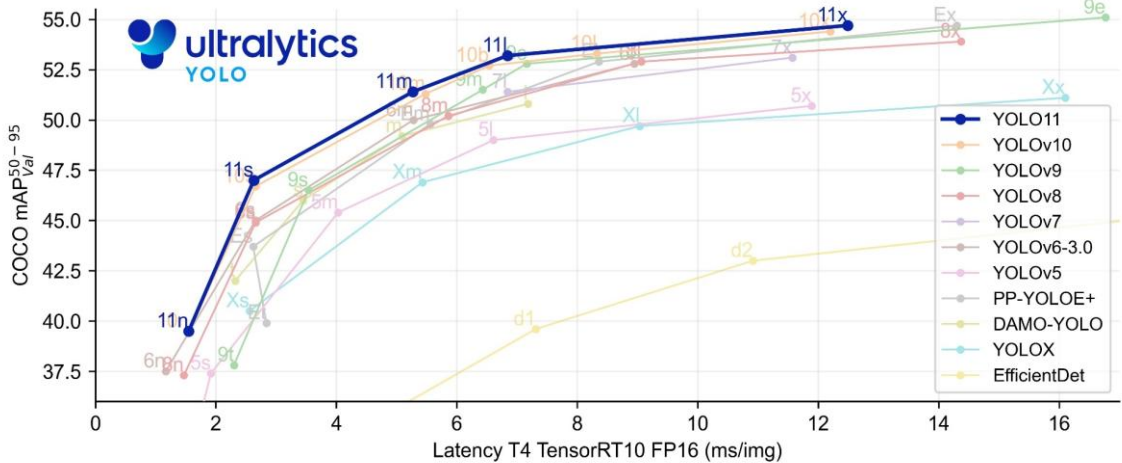
Bu amaçla, çeşitli operasyonel koşullarda sağlam tespit performansı elde etmek için beş ileri düzey nesne tespit mimarisi bu çalışmada uygulanmış, eğitilmiş ve değerlendirilmiştir: YOLOv8, YOLOv9, YOLOv11, YOLOv12 ve Faster R-CNN. Bu modeller, yalnızca çağdaş tespit paradigmalarını temsil etmeleri ve mimari çeşitlilikleri nedeniyle değil, aynı zamanda bu tezin çoklu görev ve derinlik-farkındalıklı füzyon hedeflerini destekleme kapasiteleri nedeniyle seçilmiştir. Her bir model, RailSem19 veri setine dayalı özel veri ön işleme, sahte Anotasyon üretimi ve alan-özü (domain-specific) eğitim stratejileriyle demiryolu bağlamına uyarlanmıştır.

4.1.1. YOLO'nun Evrimi

YOLO serisi, ortaya çıktığı günden bu yana önemli bir evrim geçirmiş; mimari iyileştirmeler ve algoritmik yenilikler yoluyla gerçek zamanlı nesne tespitindeki yetkinliğini sürekli olarak artırmıştır. Başlangıçta YOLO, nesne tespitini tek bir regresyon problemi olarak ele alan yenilikçi bir yaklaşım sunarak uçtan uca eğitim ve gerçek zamanlı çıkarımı mümkün kılmıştır. Takip eden sürümler olan YOLOv2 ve YOLOv3, bağlayıcı kutular (anchor boxes), yığın normalizasyonu (batch normalization) ve özellik piramidi ağları (FPN) gibi yenilikleri entegre ederek tespit doğruluğunu ve modelin sağlamlığını artırmıştır.

YOLOv4'ün tanıtımı, Aşamalı Kısmi Ağları (Cross-Stage Partial Networks - CSPNet) ve Mish aktivasyon fonksiyonları gibi stratejileri bünyesine katarak öznetelik aktarımı ve doğrusal olmayan öğrenme kapasitesini önemli ölçüde geliştirmiş ve seride önemli bir dönüm noktası olmuştur. Bağımsız olarak Ultralytics tarafından geliştirilen YOLOv5 ise, otomatik bağlayıcı kutu öğrenimi, mozaik veri artırımı ve Verimli Katman Toplama Ağları (Efficient Layer Aggregation Networks - ELAN) omurgası gibi tekniklerle performansı daha da optimize etmiş; doğruluk ve çıkarım hızında önemli kazanımlar sağlamıştır.

YOLOv6 ve YOLOv7 bu gelişim sürecini sürdüren sürümler olmuş; ölçeklenebilirlik ve verimlilik üzerine odaklanmıştır. YOLOv6, hafif yapısı ve dağıtım esnekliğiyle öne çıkarken, YOLOv7; RepVGG omurgasının kullanımı ve Verimli Kanal Dikkati (Efficient Channel Attention - ECA) mekanizmasının entegrasyonu gibi yeniliklerle farklı donanım platformlarında performansı daha da artırmıştır.



Şekil 4.1. YOLO Modellerinin Evrimi: mAP@0.95 ve Gecikme Karşılaştırması (T4 TensorRT FP16 Üzerinde) (Ultralytics, 2025)

Şekil 4.1, farklı YOLO sürümlerinin ortalama hassasiyet ($mAP@0.95$) ve çıkarım gecikmesi değerlerini karşılaştırarak, hız-doğruluk dengesine dayalı model seçimi için nicel bir temel sunmaktadır. Evrimi boyunca YOLO serisi, gerçek zamanlı performansa odaklanmayı sürdürmüş ve bu yönüyle hızlı çıkarım gerektiren uygulamalar, örneğin demiryolu nesne tespiti gibi alanlar için cazip bir seçenek olmaya devam etmiştir. Bununla birlikte, bu çalışmada da ortaya konduğu üzere, en uygun modelin seçimi, uygulama alanının özgül gereksinimlerine göre uyarlanmış performans ölçütlerinin kapsamlı bir değerlendirmesini gerektirmektedir.

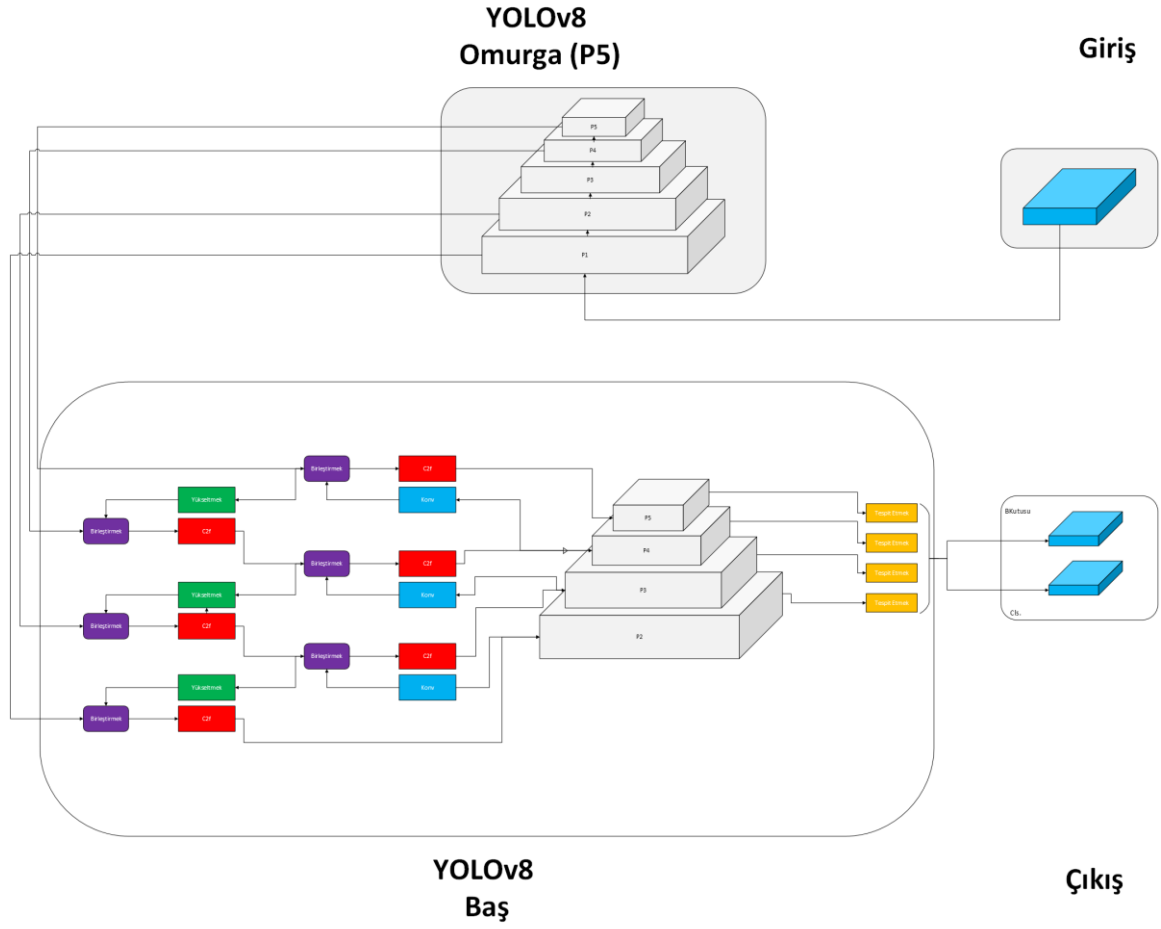
4.1.2. YOLOv8

YOLOv8, tek aşamalı nesne tespit algoritmalarından oluşan YOLO serisinin önemli bir evrimsel adımını temsil etmekte olup, tespit doğruluğunu ve çıkarım verimliliğini artırmaya yönelik mimari ve işlevsel yenilikler içermektedir. Tasarımının merkezinde, Aşamalı Kısmi Ağları (CSP) bağlantılar kullanan CSPDarknet53 omurgası yer almakta olup; bu yapı, gradyan fazlalığını azaltmakta ve öznitelik aktarımını iyileştirerek modelin hiyerarşik öznitelik çıkarımı kapasitesini artırmaktadır.

YOLOv8'in "neck" bileşeni, ölçekler arası etkili öznitelik birleştirmeyi mümkün kılan, değiştirilmiş bir Yol Toplama Ağı (Path Aggregation Network - PANet) mimarisi kullanmaktadır. Bu tasarım, sinyaller, raylar ve araçlar gibi boyutları büyük ölçüde değişkenlik gösteren nesnelerin bulunduğu demiryolu ortamlarında, farklı ölçeklerde nesne tespit yeteneğini önemli ölçüde güçlendirmektedir.

YOLOv8'i seleflerinden ayıran en belirgin özelliklerden biri, önceden tanımlanmış bağlayıcı kutulara (anchor boxes) olan ihtiyacı ortadan kaldıran, bağlayıcısız (anchor-free) tespit başlığına sahip olmasıdır. Bu yenilik, modelin nesne merkezlerini doğrudan tahmin etmesini sağlayarak eğitim sürecini sadeleştirmekte ve özellikle düzensiz şekilli ya da kısmen örtülmüş nesnelerin tespitinde doğruluk oranını artırmaktadır; bu da demiryolu senaryolarında sıkça karşılaşılan bir durumdur.

Şekil 4.2, YOLOv8 mimarisinin omurga, boyun ve tespit başlığı bileşenlerini göstererek, modelin çok ölçekli öznitelik işleme kapasitesine dayalı nesne tespiti sürecini şematik olarak ortaya koymaktadır.



Şekil 4.2. YOLOv8 Mimarisi

Ayrıca, YOLOv8 eğitim sürecinde Mosaic, MixUp ve Copy-Paste gibi ileri düzey veri artırma tekniklerini entegre etmektedir. Bu stratejiler, modelin çeşitli görsel koşullara karşı dayanıklılığını artırmakta ve demiryolu sistemlerinin dinamik ve karmaşık ortamları da dâhil olmak üzere, farklı senaryolara genelleme yeteneğini geliştirmektedir.

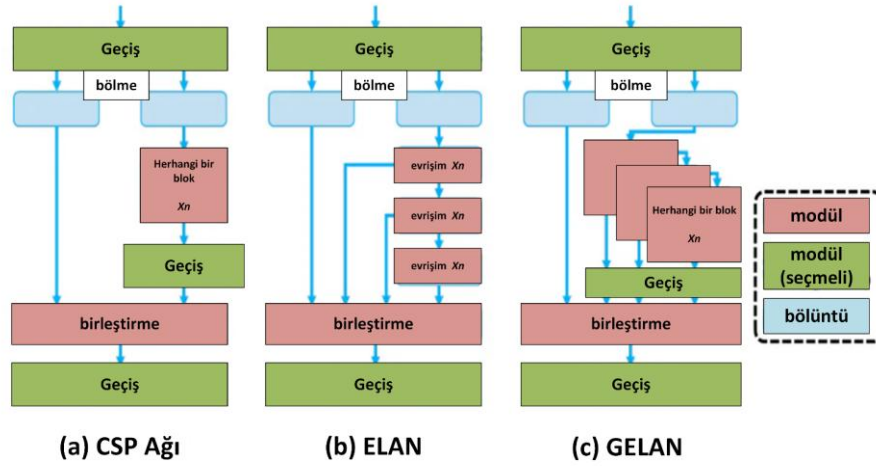
Bu çalışmada, YOLOv8, demiryolu uygulamalarıyla ilişkili beş tanımlı tespit kategorisini içeren, yarı otomatik etiketlenmiş RailSem19 veri seti üzerinde ince ayar yapılarak eğitilmiştir. Model, gerçek zamanlı uygulamalarda kayda değer bir performans sergileyerek, demiryolu ortamlarında nesne tespiti görevleri için güçlü bir başlangıç noktası (baseline) olduğunu ortaya koymuştur.

4.1.3. YOLOv9

YOLOv9, tek aşamalı nesne tespit algoritmalarından oluşan YOLO serisinin önemli bir evrimsel aşamasını temsil etmekte olup, özellikle demiryolu ortamlarında sıklıkla karşılaşılan küçük ve kısmen örtülmüş nesnelerin tespit doğruluğunu artırmaya yönelik gelişmiş mimari

iyileştirmeler içermektedir. Modelin tasarımının merkezinde, ELAN yapısının geliştirilmiş bir versiyonu olan Genel Verimli Katman Toplama Ağı (Generalized Efficient Layer Aggregation Network – GELAN) yer almaktadır. GELAN, daha esnek ve verimli bir öznitelik toplama mekanizması sunmakta; tersinir (reversible) fonksiyonlar aracılığıyla gradyan bilgisinin korunmasını ve etkin biçimde kullanılmasını sağlayarak, derin ağ işlemleri sırasında bilgi kaybını en aza indirmektedir.

Şekil 4.3, YOLOv9’un temel yapı taşı olan GELAN’ın mimari evrimini, sırasıyla CSPNet, ELAN ve önerilen GELAN yapılarının katman birleşim stratejileri bağlamında karşılaştırmalı olarak sunmaktadır.



Şekil 4.3. GELAN Mimarisi: (a) CSPNet, (b) ELAN ve (c) önerilen GELAN. [77]

GELAN’a tamamlayıcı olarak, YOLOv9 ayrıca Programlanabilir Gradyan Bilgisi (Programmable Gradient Information – PGI) modülünü içermektedir. Bu modül, ağ boyunca temel gradyan bilgilerinin korunmasını ve yayılmasını geliştirerek modelin eğitim kararlılığına ve yakınsama (convergence) hızına katkı sağlamakta; bu da demiryolu gibi karmaşık ve değişken koşullara sahip ortamlarda önemli bir avantaj sunmaktadır.

Tespit performansını daha da artırmak amacıyla, YOLOv9 yüksek frekanslı ve düşük frekanslı öznitelikleri eşzamanlı olarak işleyen çift yollu (dual-path) bir mimari benimsemektedir. Bu tasarım, modelin hem uzak sinyaller ya da yayalar gibi küçük nesnelerin ince detaylarını yakalayabilmesini hem de büyük nesneler için bağlamsal bilgiyi koruyabilmesini sağlamaktadır.

Buna ek olarak, YOLOv9 çıkarım (inference) sırasında çekirdek ağırlıklarının uyarlayabilen dinamik evrişim katmanları kullanmaktadır. Bu özellik, modelin farklı nesne görünüşleri ve çevresel koşullara uyum sağlamasına olanak tanımakta; aydınlatma, hava durumu ve kısmi

örtölmelerin tespit doğruluğunu etkileyebileceđi dinamik demiryolu senaryolarında büyük önem taşımaktadır.

Bu çalışmada YOLOv9, demiryolu uygulamalarıyla ilişkili beş tanımlı tespit kategorisini içeren, yarı otomatik etiketlenmiş RailSem19 veri seti üzerinde ince ayar yapılarak eğitilmiştir. Model, özellikle yaya ve uzak sinyal gibi küçük nesne sınıflarında doğruluk açısından gelişmiş sonuçlar sergilemiş ve gerçek zamanlı demiryolu nesne tespiti görevleri için sağlam bir aday olarak öne çıkmıştır.

4.1.4. YOLOv11

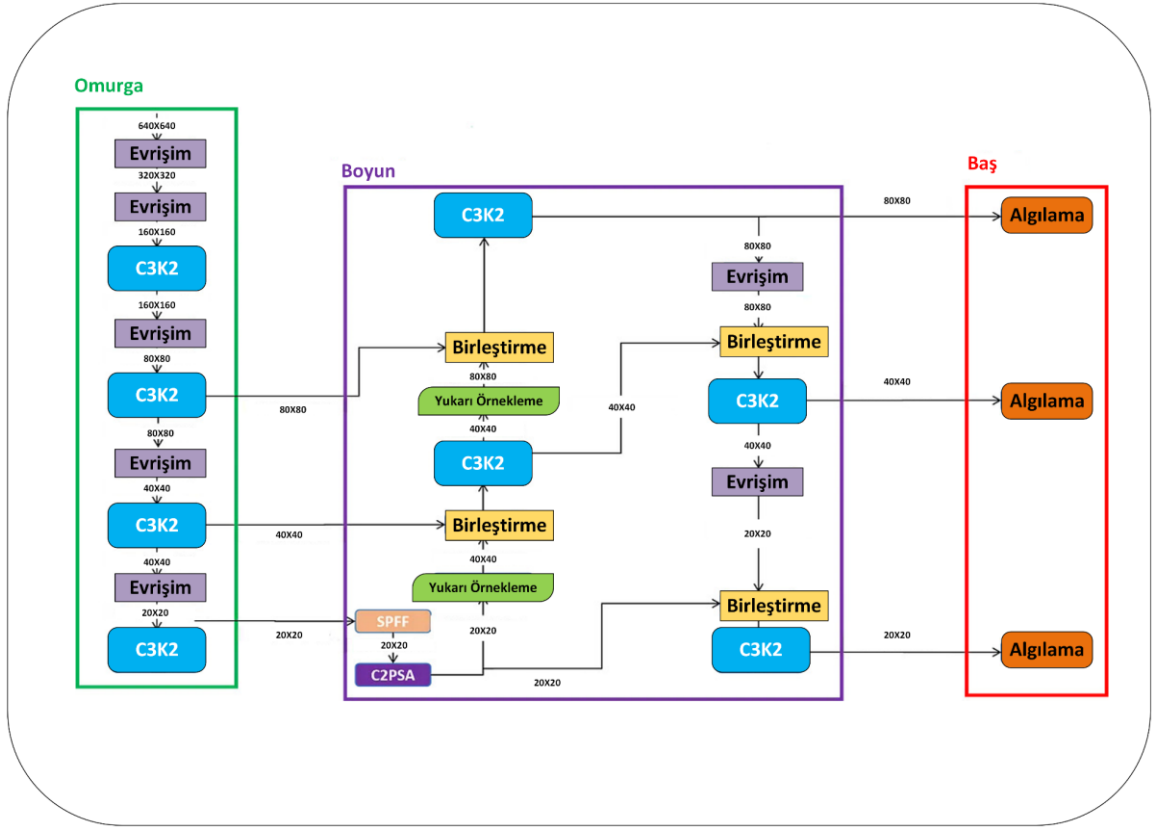
YOLOv11, tek aşamalı nesne tespit algoritmalarından oluşan YOLO serisinin önemli bir ilerlemesini temsil etmekte olup; tespit doğruluđu, mimari verimlilik ve karmaşık demiryolu ortamlarına uyum açısından geliştirilmiş yetenekler sunmaktadır. Modelin tasarımının merkezinde, önceki sürümlerde kullanılan C2f bloğunun yerine geçen C3k2 bloğunun entegrasyonu yer almaktadır. Bu değışiklik, yüksek performansı korurken hesaplama karmaşıklığını azaltmakta ve çok ölçekli öznitelik çıkarımını daha verimli hale getirmektedir.

YOLOv11'in omurgasında, farklı alma alanlarından (receptive field) bağlamsal bilgi yakalamak amacıyla çok ölçekli havuzlama işlemleri gerçekleştiren Hızlı Uzamsal Piramit Havuzlama (Spatial Pyramid Pooling Fast - SPPF) modülü yer almaktadır. Bu yetenek, demiryolu sahnelerinde karşılaşılan uzak sinyaller veya küçük engeller gibi farklı boyutlardaki nesnelerin tespiti açısından özellikle değerlidir.

Mekânsal dikkati artırmak amacıyla, YOLOv11'de Uzamsal Dikkatli Aşamalı Kısmi Ağı (Cross-Stage Partial with Spatial Attention - C2PSA) modülü tanıtılmıştır. Bu mekanizma, öznitelik haritalarını anlamlı bölgelere odaklanarak iyileştirmekte ve böylece dinamik demiryolu ortamlarında yaygın olan küçük ve örtölmüş nesnelerin tespitinde modelin başarımını artırmaktadır.

YOLOv11'in "neck" bileşeni, farklı çözünürlüklerden gelen özniteliklerin yukarı örnekleme ve birleştirme işlemleri yoluyla bütünleştirilmesini sağlar; ardından C3k2 ve C2PSA modülleri uygulanır. Bu tasarım, farklı çözünürlük seviyelerinden gelen özniteliklerin etkili bir şekilde kaynaştırılmasını sağlayarak değışken koşullar altında modelin dayanıklılığını artırmaktadır.

Şekil 4.4, YOLOv11 modelinin genel mimari yapısını ve C3k2, SPPF ile C2PSA gibi bileşenlerin ağı üzerindeki konumlarını ayrıntılı şekilde göstermektedir.



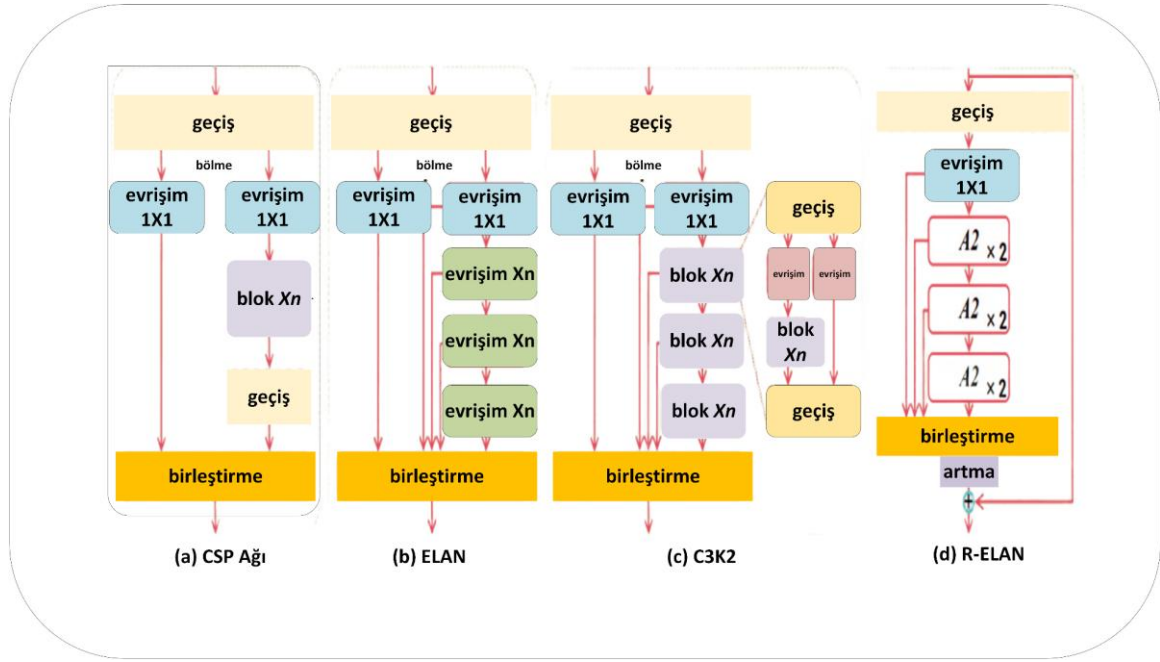
Şekil 4.4. YOLOv11 Mimarisi

Bu çalışmada, YOLOv11; tespit doğruluğu ile hesaplama verimliliği arasındaki üstün denge nedeniyle birincil nesne tespit modeli olarak seçilmiştir. Doğal bağlayıcı kutular içermeyen RailSem19 veri seti üzerinde, özel olarak geliştirilen yarı otomatik etiketleme (pseudo-labeling) yöntemiyle ince ayar yapılmıştır. YOLOv11, değerlendirilen modeller arasında en yüksek ortalama ortalama kesinlik oranı (mAP@0.5) elde ederek üstün bir performans sergilemiştir. Modelin daha önce görülmemiş test ortamlarında da güçlü genelleme yetenekleri sergilemesi, özellikle yayalar ve engeller gibi güvenlik açısından kritik sınıfların tespitinde dikkat çekmiştir. Bu özellikler, YOLOv11'i gerçek zamanlı, çok görevli demiryolu tespit sistemleri için optimal bir tercih haline getirmektedir.

4.1.5. YOLOv12

YOLOv12, gerçek zamanlı nesne tespiti alanında önemli bir ilerlemeyi temsil etmekte olup, dikkat (attention) merkezli bir mimari sunarak evrişimli sinir ağlarının (CNN) hesaplama verimliliği ile dikkat mekanizmalarının ifade gücü arasında bir denge kurmaktadır. Tasarımının merkezinde, hesaplama karmaşıklığını azaltmak amacıyla öznitelik haritalarını parçalara ayıran Alan Dikkati (Area Attention - A²) modülü ile öznitelik toplama ve eğitim kararlılığını artıran Artık

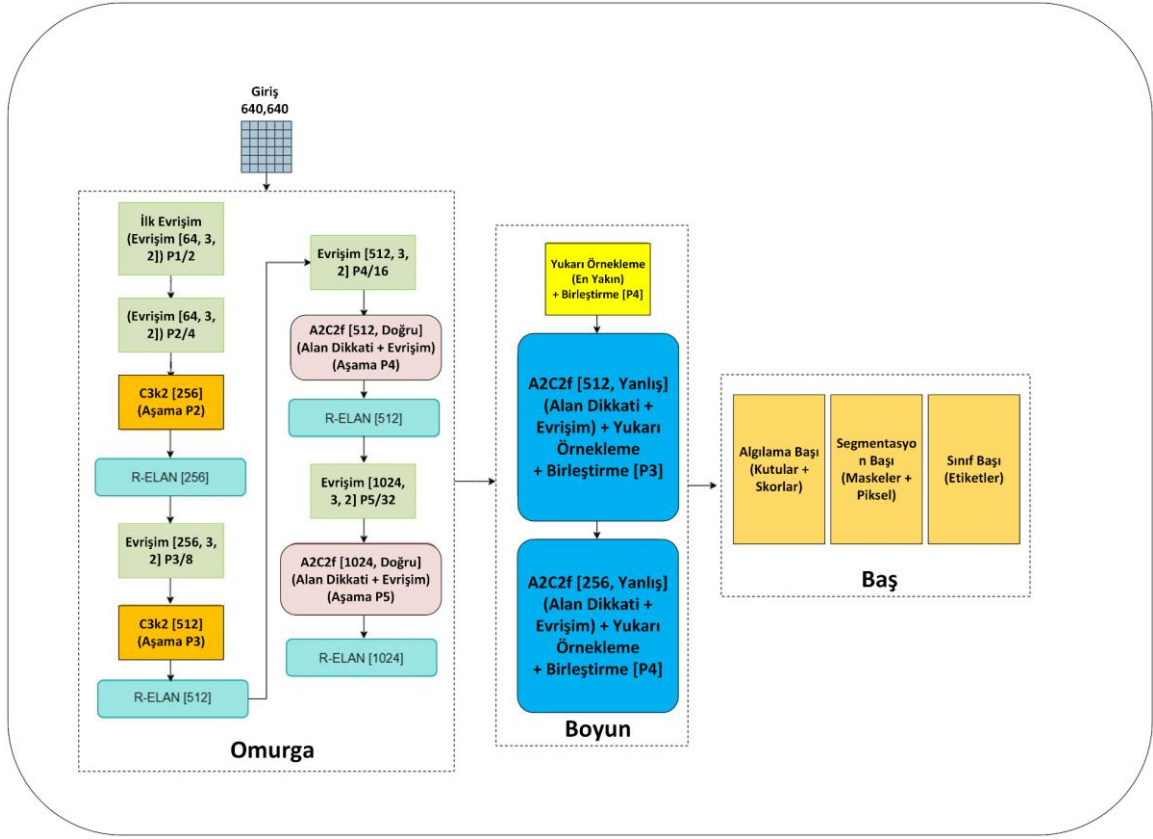
Verimli Katman Toplama Ağları (Residual Efficient Layer Aggregation Networks - R-ELAN) yer almaktadır. Buna ek olarak, FlashAttention'ın entegrasyonu bellek erişimini optimize ederek çıkarımı (inference) daha da hızlandırmaktadır. Şekil 4.5, YOLOv12 mimarisinde kullanılan R-ELAN bloklarının yapısal düzenini ve model içindeki konumlarını ayrıntılı biçimde göstermektedir.



Şekil 4.5. YOLOv12'de kullanılan R-ELAN

Bu yeniliklere rağmen, ampirik değerlendirmeler YOLOv12'nin doğruluk veya eğitim verimliliği açısından selefleri, örneğin YOLOv11, karşısında tutarlı bir üstünlük sağlamadığını ortaya koymaktadır. Gerçek zamanlı işlemenin kritik olduğu demiryolu ortamlarında, YOLOv12'nin sunduğu doğrulukta marjinal artışlar, özellikle YOLOv11 gibi modellerin daha düşük gecikmeyle üstün performans sergilediği durumlarda, artan hesaplama gereksinimlerini haklı çıkarmayabilir. Bu nedenle, YOLOv12 umut vadeden yapısal iyileştirmeler sunsa da, demiryolu nesne tespiti görevlerindeki pratik uygulanabilirliği, önceki modellerin performanslarıyla karşılaştırıldığında dikkatli bir şekilde değerlendirilmelidir.

Şekil 4.6, YOLOv12 mimarisinin temel bileşenlerini ve öznitelik akış yapısını katmanlar düzeyinde sunarak, modelin genel mimari düzenini görsel olarak açıklamaktadır.

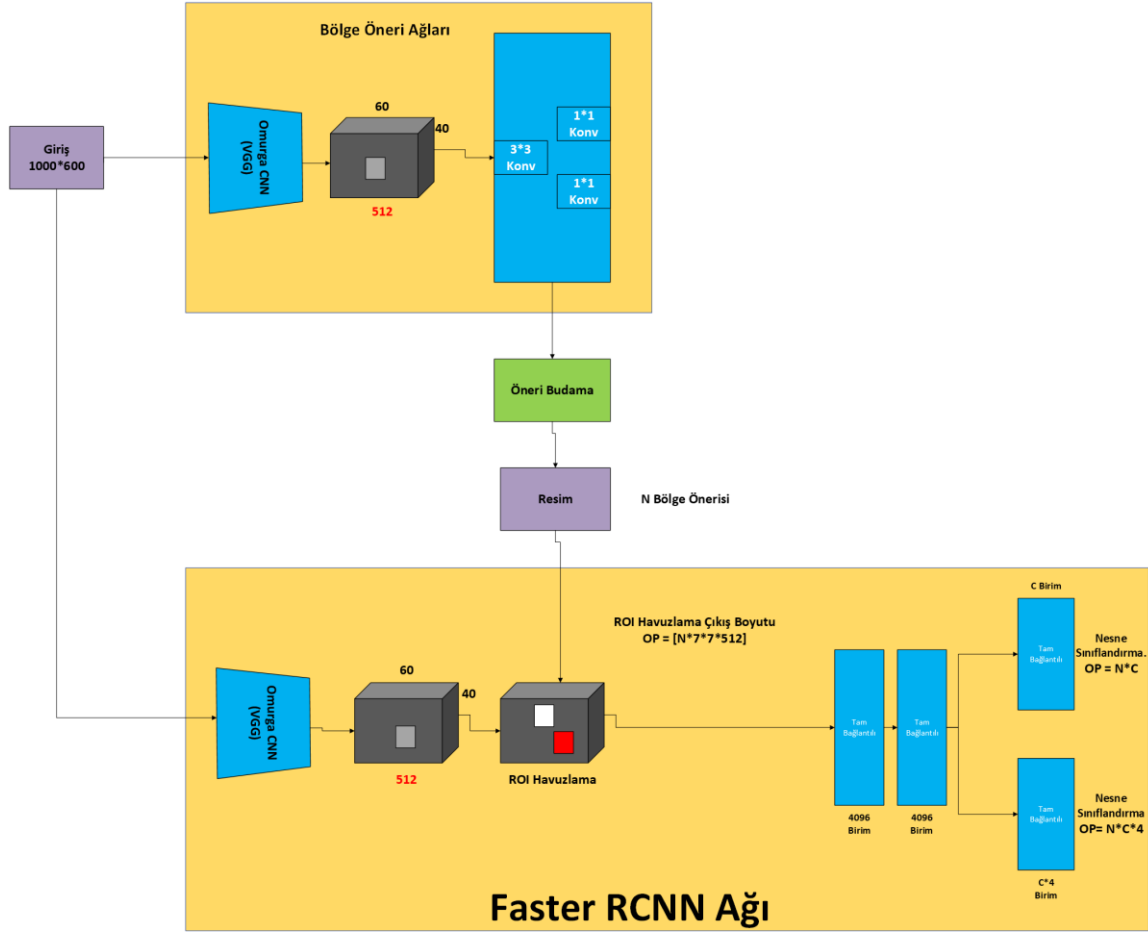


Şekil 4.6. YOLOv12 Mimarisi

4.1.6. Faster R-CNN

Karşılaştırmalı değerlendirme kapsamında, iki aşamalı bir nesne tespit mimarisi olan Faster R-CNN modeli de çalışmaya dâhil edilmiştir. Bu model, ilk aşamada Bölge Öneri Ağı (Region Proposal Network – RPN) aracılığıyla olası nesne konumlarını temsil eden bölge önerileri üretmekte; ikinci aşamada ise bu bölgeler, tam bağlantılı katmanlar üzerinden sınıflandırılmakta ve sınırlayıcı kutu regresyonu uygulanarak nesne konumları hassas biçimde tahmin edilmektedir. Uygulamada, modelin omurgasında COCO veri kümesi üzerinde önceden eğitilmiş bir ResNet50 kullanılmış; ardından RailSem19 veri setinin sahte etiketli alt kümesi üzerinde aktarım öğrenmesi (fine-tuning) gerçekleştirilmiştir.

Şekil 4.7, Faster R-CNN mimarisinin temel bileşenlerini ve iki aşamalı nesne tespit sürecinin işleyişini ayrıntılı biçimde görselleştirmektedir.



Şekil 4.7. Faster R-CNN Mimarisi

Bununla birlikte, demiryolu nesne tespiti bağlamında gerçekleştirilen ampirik değerlendirmeler, Faster R-CNN mimarisinin bu alandaki gereksinimleri karşılamada optimalin altında bir performans sergilediğini ortaya koymuştur. Model, uzatılmış eğitim süreleri ve göreceli olarak düşük doğruluk oranları nedeniyle test edilen modeller arasında en düşük etkililiğe sahip yapı olarak öne çıkmıştır. Bu performans sınırlamaları büyük ölçüde, mimarinin Bölge Öneri Ağı (Region Proposal Network – RPN) ve ardından gelen İlgi Bölgesi (Region of Interest – RoI) havuzlama katmanlarında yer alan hesaplama karmaşıklığından kaynaklanmaktadır. Söz konusu aşamalar, yüksek düzeyde işlem gücü ve bellek tüketimi gerektirdiğinden, dinamik ve zaman kısıtlı demiryolu ortamlarında gerçek zamanlı işleme koşullarına uygunluk göstermemektedir.

Ayrıca, demiryolu uygulamalarında sıkça karşılaşılan küçük boyutlu ya da kısmen örtülmüş nesnelerin tespitinde yaşanan zorluklar, modelin doğruluk performansını olumsuz yönde etkilemiştir. Tüm bu bulgular, Faster R-CNN'in teorik olarak sunduğu avantajlara rağmen, özel amaçlı ve zaman hassasiyeti gerektiren uygulamalarda pratik etkinliğinin sınırlı olduğunu

göstermektedir. Bu durum, yüksek doğruluk ile hesaplama verimliliği arasında denge kurabilen alternatif modellere duyulan ihtiyacı bir kez daha vurgulamaktadır.

4.1.7. Sahte Etiket Üretim Hattı

RailSem19 veri seti yerel sınırlayıcı kutu etiketleri içermediğinden, veri setinin anlamsal segmentasyon maskelerinden türetilmiş yüksek kaliteli tespit anotasyonları üretmek amacıyla özel bir sahte etiketleme hattı geliştirilmiştir. Hat aşağıdaki adımları içerir:

- Maske Filtreleme (Mask Filtering): Bölüm 3.1’de tanımlanan ilgili tespit kategorilerine göre segmentasyon maskeleri filtrelenmiştir.
- Kontur Çıkartımı (Contour Extraction): İkili maskeler, bağlı bileşen analizi ve OpenCV ile kontur çıkartımı kullanılarak işlenmiştir.
- Geometrik Filtreleme (Geometric Filtering): Konturlar, minimum alan, en-boy oranı ve ray bölgelerine yakınlığa göre filtrelenmiştir.
- Sınırlayıcı Kutuya Dönüşüm (Bounding Box Conversion): Geçerli konturlar dikdörtgen sınırlayıcı kutulara dönüştürülmüş ve YOLO ile COCO formatlarına normalleştirilmiştir.
- Anotasyon Serileştirme (Annotation Serialization): Son sahte etiketler, YOLO ile uyumlu .txt ve Faster R-CNN ile uyumlu .json dosyaları olarak saklanmıştır.

Bu işlem hattı, tüm tespit modellerinde tutarlı anotasyon kullanımını sağlamış ve segmentasyon ile tespit çıktılarının anlamsal ve uzamsal hizalanmasını mümkün kılan bütüncül bir öğrenme çerçevesi oluşturmuştur.

Bu nesne tespit modelleri bir araya geldiğinde, demiryolu sahnelerinin kısıtları ve zorluklarına özel olarak uyarlanmış, gerçek zamanlı ve yüksek doğruluklu tespit yeteneklerinin güçlü bir yelpazesini sunar. Bu modellerin çalışmaya dahil edilmesi, çağdaş tespit yaklaşımlarının sistematik değerlendirilmesine olanak sağlamış ve ilerleyen bölümlerde tartışılacak çok görevli ve derinlik odaklı görsel akıl yürütme çerçevelerine entegrasyonları için zemin hazırlamıştır.

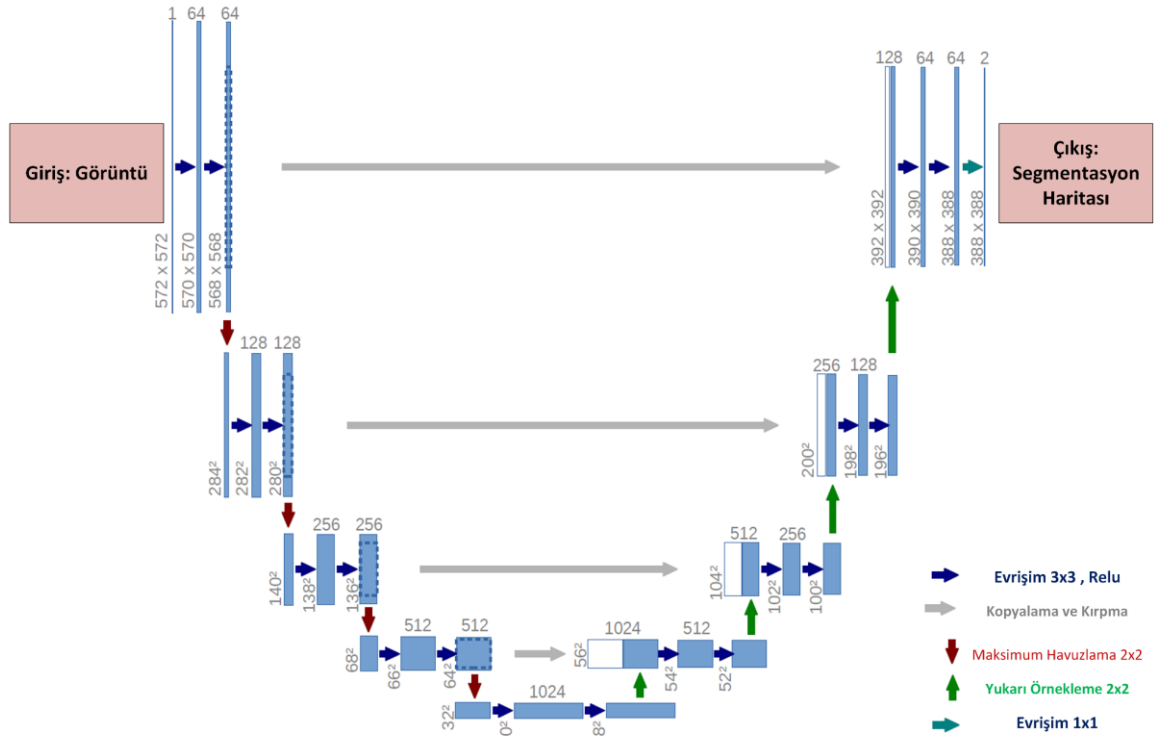
4.2. Anlamsal Segmentasyon Modelleri

Anlamsal segmentasyon, bu çalışmada geliştirilen görsel algı sisteminin temel taşlarından birini oluşturmakta olup, demiryolu ortamlarındaki karmaşık sahnelerin yoğun ve piksel düzeyinde sınıflandırılmasını sağlar. Nesne tespitinden farklı olarak, yalnızca ayrıık nesneler için kaba sınırlayıcı kutular sunmakla yetinmeyen anlamsal segmentasyon, her bir pikseli semantik bir kategoriyle etiketleyerek daha ayrıntılı bir mekânsal anlayış sunar. Bu düzeyde detay, çarpışma önleme, ray doluluğu tespiti ve altyapı izleme gibi görevlerde karar alma süreçlerini doğrudan etkileyen güvenlik-kritik ulaşım alanlarında hayati önem taşır.

Bu bağlamda modern segmentasyon tekniklerinin yeteneklerini değerlendirmek amacıyla, beş derin öğrenme tabanlı mimari uygulanmış, eğitilmiş ve sistematik olarak değerlendirilmiştir: UNet-Temel (Vanilla), UNet-Pretrained, DeepLabV3+, Özellik Piramidi Ağı (Feature Pyramid Network - FPN) ve Piramit Sahne Ayırıştırma Ağı (Pyramid Scene Parsing Network - PSPNet). Bu modeller, hafif kodlayıcı-çözücü yaklaşımlardan çok ölçekli ve bağlamsal farkındalığa sahip sofistike mimarilere kadar uzanan geniş bir tasarım felsefesi yelpazesini temsil etmektedir. Her bir model, karşılaştırmalarda tutarlılığı ve adaleti sağlamak amacıyla, Bölüm 3'te açıklanan standart ön işleme ve veri artırma yöntemleriyle RailSem19 veri seti üzerinde eğitilmiştir.

4.2.1. UNet-Vanilla

Klasik U-Net mimarisi, ilk olarak biyomedikal görüntü segmentasyonu amacıyla geliştirilmiş olup, bu çalışmada demiryolu ortamlarında yürütülen anlamsal segmentasyon görevlerinde temel bir karşılaştırma modeli (baseline) olarak kullanılmaktadır. U-Net'in tasarımı, ardışık aşağı örnekleme işlemleriyle hiyerarşik öznetelikler çıkaran simetrik bir kodlayıcı-çözücü (encoder-decoder) yapısı ile karakterize edilir; çözümleyici yol (decoder), uzamsal çözünürlüğü yukarı örnekleme yoluyla yeniden inşa ederken, atlama bağlantıları (skip connections) ince detayların korunmasına olanak tanır. Bu mimariye ilişkin genel yapısal bileşenleri ve veri akışını gösteren bir şematik temsil Şekil 4.8'de sunulmuştur.



Şekil 4.8. U-Net Mimarisi [78]

Bu araştırmada kullanılan UNet-Vanilla modeli, önceden eğitilmiş ağırlıklar kullanılmadan, standart evrişim bloklarıyla uygulanmış ve böylece doğrudan ve yorumlanabilir bir temel yapı sağlanmıştır. Mimarisinin sadeliğine rağmen, UNet-Vanilla modeli RailSem19 veri seti üzerinde eğitim sırasında hızlı yakınsama (convergence) göstermiş ve makul bir performans sergilemiştir. Bununla birlikte, sınırlı bağlamsal bilgi yakalama kapasitesi ve yüzeysel öznitelik temsilleri, modelin belirsiz nesne sınırlarını ayırt etmede ve karmaşık arka plan içeren sahnelerde etkinliğini azaltmış; bu durumlar, demiryolu ortamlarında sıklıkla karşılaşılan zorluklardandır.

UNet-Vanilla'nın performansı, zorlu senaryolarda segmentasyon doğruluğunu artırmak için daha derin öznitelik hiyerarşilerini ve bağlamsal ilişkileri kullanabilen gelişmiş mimarilerin gerekliliğini ortaya koymaktadır. Bu çalışmada daha sonra ele alınan modeller, öznitelik çıkarımı ve bağlam farkındalığı açısından bu sınırlamaları aşmayı hedefleyen daha sofistike mekanizmaları içermektedir.

4.2.2. UNet-Pretrained

Bu çalışmada, semantik segmentasyon görevlerinde öznitelik çıkarım kapasitesini artırmak ve modelin genelleme yetisini geliştirmek amacıyla UNet mimarisi, önceden ImageNet veri kümesi üzerinde eğitilmiş bir ResNet34 tabanlı kodlayıcı ile yeniden yapılandırılmıştır. Geliştirilen UNet-Pretrained mimarisi, orijinal U-Net yapısının simetrik kodlayıcı-çözücü (encoder-decoder) düzenini korumakta; yalnızca kodlayıcı kısmı, transfer öğrenme yaklaşımıyla ResNet34 ağı ile değiştirilmiştir. Bu entegrasyon sayesinde, özellikle düşük ve orta düzeydeki görsel özniteliklerin daha etkin bir biçimde çıkarılması mümkün hale gelmiş; modelin eğitim süreci hızlanmış ve sinyal direkleri ile yayalar gibi azınlık sınıflara ait doğruluk oranlarında belirgin iyileşmeler gözlemlenmiştir.

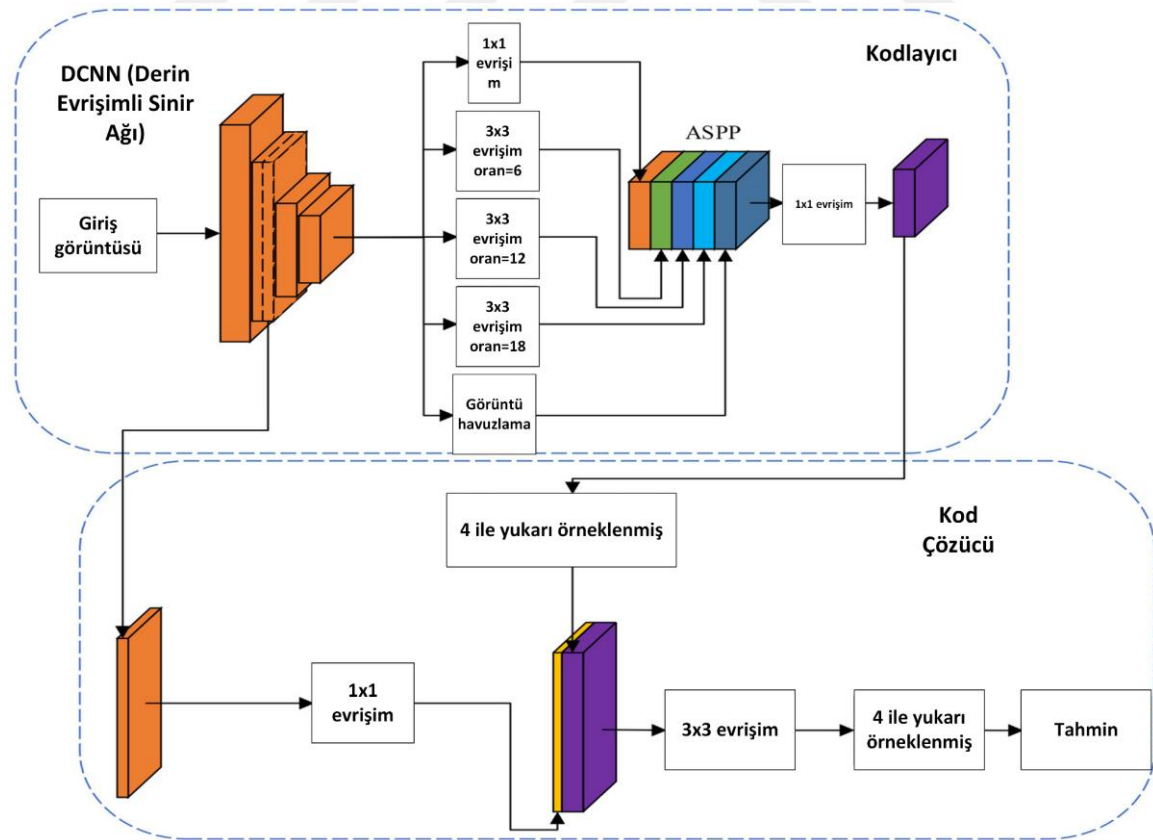
ResNet34 mimarisinde yer alan artık bağlantılar (residual connections), derin sinir ağlarında sıkça karşılaşılan gradyan sönümlenmesi (vanishing gradient) problemini azaltmakta ve böylece modelin daha derin ve ayırt edici öznitelik temsilleri öğrenebilmesini sağlamaktadır. Bu mimari iyileştirme, demiryolu sahnelerinde sık karşılaşılan aydınlatma farklılıkları, karmaşık arka plan öğeleri ve nesnelerin kısmi örtüşmesi gibi zorluklara karşı modelin dayanıklılığını önemli ölçüde artırmaktadır.

Elde edilen deneysel sonuçlar, UNet-Pretrained mimarisinin, temel UNet yapısına kıyasla anlamlı ölçüde daha yüksek bir performans sergilediğini ortaya koymakta; gerçek zamanlı semantik segmentasyon uygulamaları açısından hesaplama verimliliği ile doğruluk arasında dengeli ve uygulanabilir bir çözüm sunduğunu göstermektedir. Bu bulgular, önceden eğitilmiş sınıflandırma modellerinin tam evrişimli segmentasyon mimarilerine entegrasyonunun etkinliğini ortaya koymakta; ayrıca, etiketli veri miktarının sınırlı olduğu uzmanlaşmış alanlarda transfer öğrenme yaklaşımının segmentasyon başarımını artırma potansiyelini açıkça gözler önüne sermektedir.

4.2.3. DeepLabV3+

DeepLabV3+, nesne sınırlarını daha hassas biçimde belirlemek amacıyla kodlayıcı-çözücü (encoder-decoder) yapısı DeepLabV3 mimarisiyle birleştirilerek geliştirilmiş, son teknoloji ürünü bir anlamsal segmentasyon modelidir. Bu model, özellikle karmaşık sahne yapısına sahip demiryolu ortamlarında segmentasyon doğruluğunu artırmak üzere tasarlanmıştır. Modelin kodlayıcı bölümü, çok ölçekli bağlamsal bilgiyi yakalayabilmek için genişletilmiş evrişimler (dilated convolutions) kullanmakta; çözümleyici bölümü ise, iki aşamalı doğrusal yukarı örnekleme (bilinear upsampling) ve atlama bağlantıları (skip connections) aracılığıyla ince uzamsal ayrıntıların geri kazanılmasını sağlamaktadır.

Modelin temel bileşenlerinden biri olan Aralıklı Uzamsal Piramit Havuzlama (Atrous Spatial Pyramid Pooling – ASPP) modülü, farklı genişleme oranlarına sahip boşluklu evrişimler (atrous convolutions) uygulayarak çoklu ölçeklerde bağlamsal bilgi yakalayabilmektedir. Bu yapı sayesinde model, farklı boyut ve biçimlerdeki nesnelerin doğru şekilde ayrıştırılmasını mümkün kılmakta ve çok ölçekli nesne segmentasyonu problemini etkin biçimde çözebilmektedir. Şekil 4.9, DeepLabV3+ modelinin kodlayıcı-çözücü yapısı ile Aralıklı Uzamsal Piramit Havuzlama (ASPP) modülünün çok ölçekli bağlamsal bilgi yakalama mekanizmasını görsel olarak özetlemektedir.



Şekil 4.9. DeepLabV3+ Mimarisi [79]

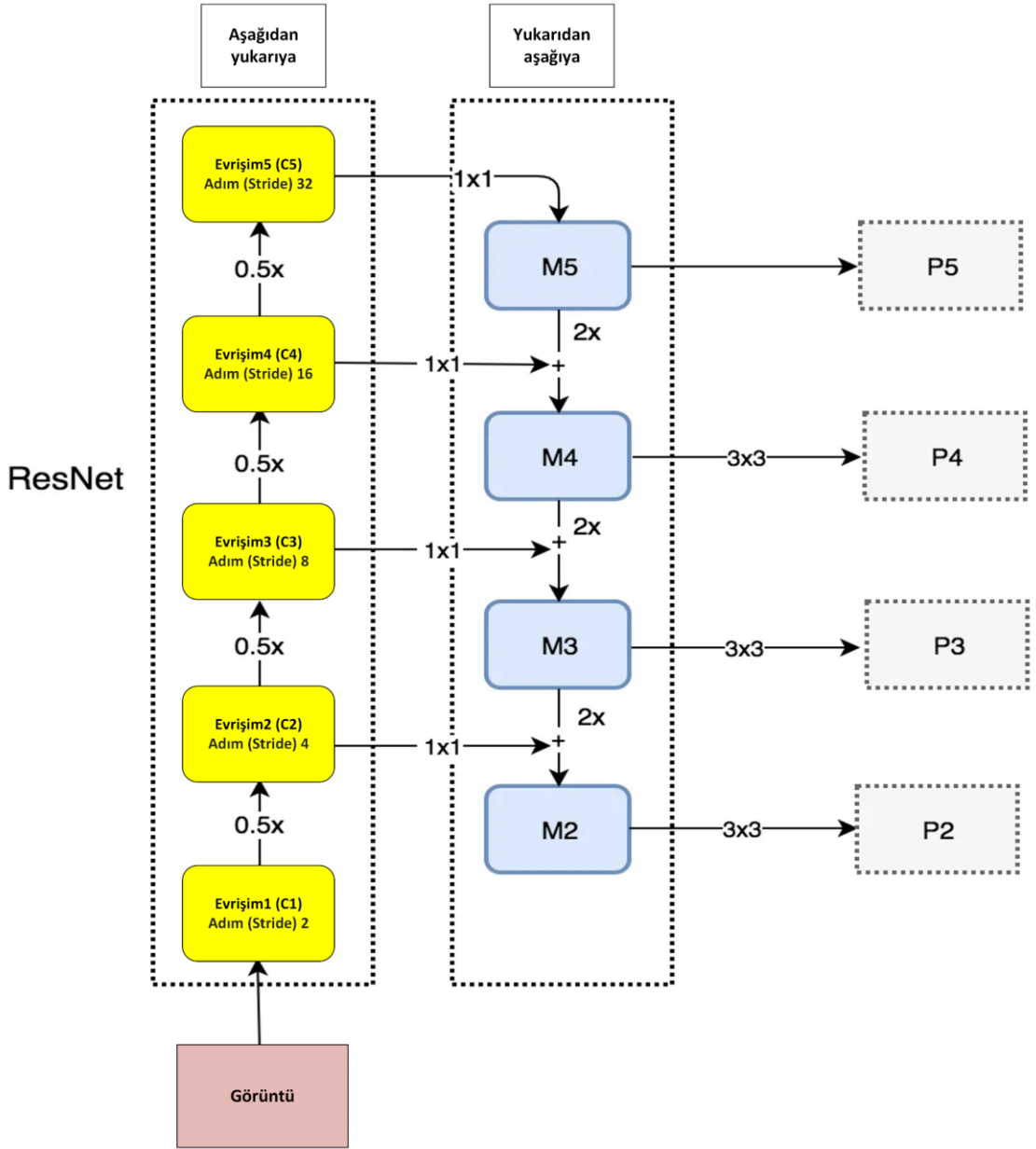
Bu çalışmada, DeepLabV3+ modeli ImageNet üzerinde önceden eğitilmiş ResNet50 omurgası ile uygulanmıştır. Bu yaklaşım, transfer öğreniminden yararlanarak hem öznitelik çıkarım yeteneklerini artırmış hem de eğitim süresini azaltmıştır. Model, çeşitli demiryolu sahnelerinde ortalama kesişim-birleşim oranı (mIoU) ve genel piksel doğruluğu açısından üstün performans göstermiştir. Özellikle raylar ile kaldırımlar gibi görsel açıdan birbirine yakın sınıfların ayırımında öne çıkmıştır. Derin arka plan karmaşıklığına rağmen gösterdiği güçlü performans ve hesaplama verimliliği ile uzamsal doğruluk arasında kurduğu denge, DeepLabV3+'ı Bölüm 4.3'te açıklanan çok görevli füzyon mimarisi için birincil segmentasyon omurgası olarak tercih edilir kılmıştır.

Bu bulgular, DeepLabV3+ modelinin demiryolu ortamlarında yürütülen anlamsal segmentasyon görevlerindeki etkinliğini ortaya koymakta ve karmaşık yapıları hassas bir şekilde ayırt etme becerisiyle ileri düzey demiryolu tespit sistemlerinin geliştirilmesine katkı sunduğunu göstermektedir.

4.2.4. Öznitelik Piramidi Ağı (Feature Pyramid Network – FPN)

Öznitelik Piramidi Ağı (FPN), çok ölçekli özniteliklerin temsiliyetini artırmak amacıyla, yüksek düzeydeki anlamsal bilgileri düşük düzeydeki uzamsal ayrıntılarla birleştiren gelişmiş bir yukarıdan-aşağıya mimari olarak geliştirilmiştir. İlk olarak nesne tespiti amacıyla tasarlanmış olan FPN, özellikle yapısal karmaşıklığın ve nesne ölçeklerindeki çeşitliliğin belirgin olduğu demiryolu gibi karmaşık ortamlarda, anlamsal segmentasyon görevlerinde de yüksek başarı sergilemiştir.

Şekil 4.10, Öznitelik Piramidi Ağı (FPN) mimarisinin, çok ölçekli özniteliklerin entegrasyonunu sağlayan yukarıdan aşağıya bilgi akışını ve farklı seviyelerdeki anlamsal ile uzamsal bilgilerin birleşimini göstermektedir.



Şekil 4.10. FPN Mimarisi

Bu çalışmada, FPN mimarisi, hiyerarşik öznitelik çıkarımı konusundaki güçlü yapısıyla bilinen ResNet34 kodlayıcı kullanılarak uygulanmıştır. Mimari, giriş görüntüsünü aşamalı olarak aşağı örnekleyerek kaba anlamsal öznitelikleri yakalayan bir alt-yukarı yol (bottom-up pathway) ile başlar. Ardından, bu öznitelikler yukarı örneklenerek ince uzamsal ayrıntıların geri kazanıldığı yukarıdan-aşağıya bir yol (top-down pathway) ile tamamlanır. Alt-yukarı ve yukarıdan-aşağıya yolların karşılık gelen seviyeleri arasında kurulan yanıl bağlantılar (lateral connections), düşük çözünürlüklü ancak semantik olarak zengin özniteliklerin, yüksek çözünürlüklü fakat semantik olarak zayıf özniteliklerle kaynaştırılmasını sağlar. Bu kaynaşım, ağırlıklı ölçeklerde bağlamsal bilgidir yararlanırken uzamsal doğruluğu da korumasına olanak tanır.

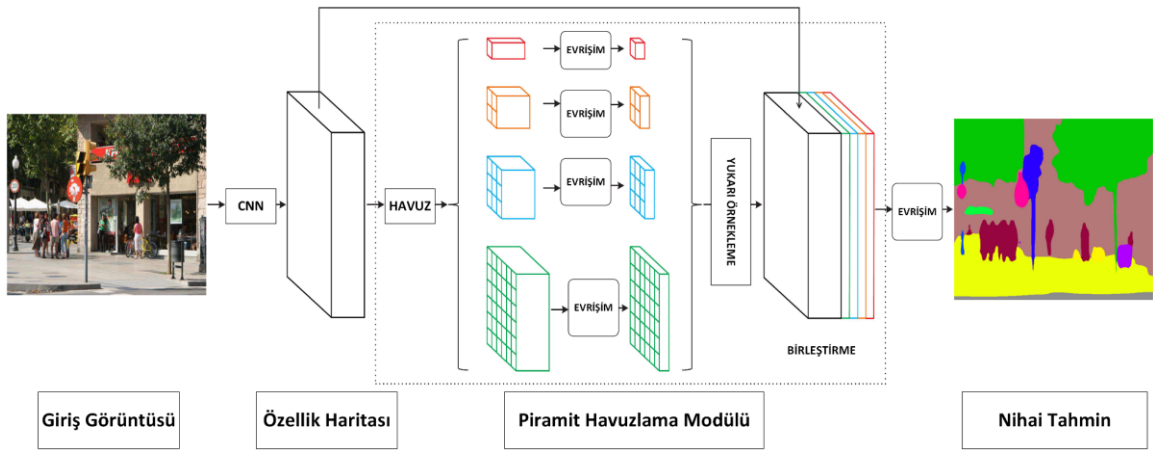
Bu çalışmada uygulanan FPN, özellikle demiryolu ortamında raylar, bariyerler ve yaya yolları gibi yapısal öğelerin sınırlarını ayırt etmede, ortalama kesişim-birleşim oranı (mIoU) ve piksel doğruluğu açısından üstün performans sergilemiştir. Hem küresel bağlamı hem de yerel ayrıntıları etkili biçimde yakalama kapasitesi sayesinde, FPN, ulaşım uygulamalarında segmentasyon görevleri için sağlam ve güvenilir bir tercih olarak öne çıkmaktadır. Bu tür uygulamalarda, hassas sınır tespiti ve bağlamsal farkındalık kritik öneme sahiptir.

Bu bulgular, FPN'nin anlamsal segmentasyon alanındaki etkinliğini vurgulamakta; karmaşık demiryolu sahnelerinin analizi ve yorumlanması için gerekli olan doğru ve bağlamsal olarak anlamlı segmentasyon haritalarını üretme konusundaki yetkinliğini ortaya koymaktadır.

4.2.5. Piramit Sahne Ayrıştırma Ağı (Pyramid Scene Parsing Network – PSPNet)

Piramit Sahne Ayrıştırma Ağı (PSPNet), hem yerel hem de küresel bağlamsal bilgiyi yakalamaya yönelik olarak tasarlanmış gelişmiş bir anlamsal segmentasyon modelidir. Bu modelin temel bileşeni olan Piramit Havuzlama Modülü (Pyramid Pooling Module – PPM), çoklu uzamsal ölçeklerden bağlamsal bilgiyi bir araya getirerek, ağına çeşitli mekânsal hiyerarşiler ve üst üste gelen yapılarla karakterize edilen karmaşık sahneleri daha iyi anlamasını sağlamaktadır.

Şekil 4.11, PSPNet modelinin Piramit Havuzlama Modülü (PPM) aracılığıyla çoklu uzamsal ölçeklerde bağlamsal bilgiyi nasıl birleştirdiğini ve böylece karmaşık sahne yapılarının etkin bir şekilde ayrıştırılmasını nasıl sağladığını göstermektedir.



Şekil 4.11. PSPNet Mimarisi

Bu çalışmada, PSPNet modeli, güçlü öznitelik temsilleri elde edilmesini sağlamak amacıyla ImageNet üzerinde önceden eğitilmiş bir ResNet50 omurgası ile uygulanmıştır. Farklı ölçeklerde bağlamsal bilgi yakalamayı sağlayan PPM, özellikle raylar ve kaldırımlar gibi birbirine yakın konumda bulunan sınıfların ayırt edilmesinde demiryolu ortamları için önemli bir avantaj

sunmuştur. Her ne kadar ortalama kesişim-birleşim oranı (mIoU) açısından DeepLabV3+ ve FPN modellerine kıyasla az da olsa daha düşük skorlar elde etmiş olsa da, PSPNet karmaşık arka planlara ve sınıf dengesizliklerine karşı yüksek dayanıklılık sergilemiştir. Modelin, üst üste gelen nesneler ve uzak altyapı unsurlarını içeren sahnelerde etkili sonuçlar verdiği gözlemlenmiştir. Sınıf tahminlerinde küresel tutarlılığı koruma yetisi, PSPNet'in demiryolu sahne analizleri gibi gerçek dünya uygulamaları için uygunluğunu ortaya koymaktadır.

Bu bulgular, PSPNet'in anlamsal segmentasyon görevlerindeki potansiyelini vurgulamakta; özellikle çok ölçekli bağlamsal bilginin doğru sahne anlayışı için kritik olduğu ortamlarda etkin bir model olduğunu göstermektedir.

Bu çalışmada değerlendirilen mimari topluluğu, anlamsal segmentasyon stratejilerinin çeşitli tasarım alanlarını temsil etmektedir. UNet varyantları, hızlı prototipleme ve kaynak kısıtlı uygulamalar için uygun, verimli ve erişilebilir başlangıç modelleridir. Öte yandan, DeepLabV3+ ve PSPNet gibi mimariler gelişmiş bağlam modelleme mekanizmalarını barındırarak yüksek varyansa sahip görsel alanlarda dayanıklılığı artırmaktadır. FPN ise dengeli mimari karmaşıklığı sayesinde, segmentasyon doğruluğu ve yapısal hassasiyet açısından en iyi performansı göstermiştir.

Bu mimarilerin Bölüm 6'da gerçekleştirilen karşılaştırmalı değerlendirmesi, model karmaşıklığı, sınır hassasiyeti ve gerçek zamanlı uygulanabilirlik arasındaki dengeyi ortaya koymuştur. Bu bulgular, önerilen çift akışlı füzyon sisteminde segmentasyon modülü olarak DeepLabV3+'ın seçilmesini doğrudan desteklemekte ve hafif çıkarım ile çok ölçekli anlamsal çıkarım yeteneklerini birleştiren hibrit mimarilerin gelecekteki araştırmalar için umut vadettiğini göstermektedir.

4.3. Çoklu Görev (Multi-Task) Birleştirme Yaklaşımları

Demiryolu ortamlarına yönelik akıllı algı sistemlerinin geliştirilmesi, birden fazla görsel yorumlama görevinin eş zamanlı olarak ele alınmasını zorunlu kılmaktadır. Bu görevler arasında, sahnenin genel yapısını anlamaya yönelik anlamsal segmentasyon ve potansiyel tehlike unsurlarının hassas konumlandırılması amacıyla gerçekleştirilen nesne tespiti öne çıkmaktadır. Geleneksel bilgisayarla görme yaklaşımları genellikle bu görevleri birbirinden bağımsız şekilde ele almakta; ancak bu tür tek görevli yaklaşımlar, gerçek zamanlı güvenlik ve algısal bütünlüğün sağlanmasının kritik olduğu karmaşık ve riskli demiryolu ortamlarında yetersiz kalmaktadır.

Bu doğrultuda, bu tez kapsamında nesne tespiti ile anlamsal segmentasyonu bütüncül ve tutarlı bir görsel akıl yürütme çerçevesi içerisinde bir araya getirmeyi amaçlayan iki tamamlayıcı çoklu görev birleştirme stratejisi önerilmiş ve uygulanmıştır. Geliştirilen bu stratejiler, sahip oldukları mimari yapı, kaynak gereksinimi ve çalışma felsefesi açısından birbirinden farklılık göstermekte olup; hesaplama verimliliği ile algısal hassasiyet arasında çeşitli ödünleşim seçenekleri sunmaktadır.

Söz konusu birleştirme stratejilerinin, önerilen demiryolu algı sistemi bünyesine entegre edilmesi yalnızca görsel durumsal farkındalığı artırmakla sınırlı kalmamış; aynı zamanda 4.4 numaralı bölümde ayrıntılı olarak ele alınan stereoskopik girdi ile birlikte kullanıldığında derinlik farkındalıklı akıl yürütmeyi de mümkün kılmıştır.

4.3.1. Segmentasyon Yönlendirmeli Geometrik Tespit

İlk birleştirme stratejisi, yalnızca anlamsal segmentasyon çıktılarından yararlanarak, nesne tespitini sonradan gerçekleştirilen geometrik çıkarımlarla yaklaşık olarak elde etmeye dayanmaktadır. Bu yaklaşım, özel olarak eğitilmiş bir nesne tespit modelinin kullanılmadığı; düşük işlem gücüne sahip ya da gecikmeye duyarlı sistemlerin gereksinim duyduğu senaryolar için özellikle elverişlidir. Önerilen yöntem, çeşitli çevresel koşullarda güvenilir piksel düzeyinde sınıflandırma doğruluğu ve üstün sınır belirleme kapasitesi sunması nedeniyle, DeepLabV3+ tabanlı bir segmentasyon mimarisi üzerine inşa edilmiştir.

Modelin çıkarım sürecini takiben, DeepLabV3+ tarafından üretilen anlamsal segmentasyon haritası, hedeflenen kritik nesne sınıflarına karşılık gelen ikili maskeleri elde etmek amacıyla filtrelenmektedir. Bu sınıflar arasında raylar, insanlar ve olası fiziksel engeller yer almaktadır. Elde edilen ikili maskeler; genişletme (dilation), erozyon (erosion) ve kontur çıkarımı gibi morfolojik işlemlerden geçirilerek bağlantılı bileşenler tespit edilmekte, her bir bileşene karşılık gelen sınırlayıcı diktörtgenler ise kontur bilgisi ve minimum alan kriterleri doğrultusunda hesaplanmaktadır. Bu sınırlayıcı diktörtgenler, daha sonra sözde tespit çıktıları (pseudo-detections) olarak kullanılmakta ve görsel açıdan, geleneksel nesne tespit algoritmalarının ürettiği sonuçlara oldukça benzer çıktılar sunmaktadır.

Yalnızca segmentasyon temelli bu tespit yaklaşımı, her ne kadar nesne düzeyindeki örtüşmeleri çözme ya da güven skorları tahmin etme gibi gelişmiş işlemlere sahip olmasa da, iki temel avantajı bünyesinde barındırmaktadır. İlk olarak, tespit edilen nesne sınırlarının, ilgili anlamsal sınıflandırma bölgeleriyle hizalı ve mekânsal olarak tutarlı olması sayesinde, çapa tabanlı (anchor-based) yaklaşımlarda sıkça karşılaşılan bölünme (fragmentation) sorunlarını azaltmaktadır. İkinci olarak, yalnızca segmentasyon modelinin tek bir ileri geçişine (forward pass) dayanması ve son işlem adımlarının oldukça düşük hesaplama maliyetine sahip olması nedeniyle, yöntem yüksek derecede verimlilik sunmaktadır.

Bu strateji, özellikle nesne görünümünün çeşitliliğinin düşük, ancak mekânsal düzenin belirgin olduğu yapısal demiryolu sahnelerinde etkili sonuçlar sağlamaktadır. Öte yandan, segmentasyon haritalarında gözlemlenen gürültüye karşı yüksek duyarlılık ve kalabalık ya da örtüşen nesnelerin bulunduğu sahnelerde yaşanan performans düşüşleri, yöntemin genel uygulanabilirliğini kısıtlayan önemli sınırlılıklar arasında yer almaktadır. Bu nedenle, sınırlı işlem kaynaklarına sahip sistemler ile gerçek zamanlı yanıt gereksinimi bulunan uygulamalarda, hafifliği

ve hesaplama açısından verimliliği sayesinde öne çıkan uygulanabilir bir çözüm olarak değerlendirilmektedir.

4.3.2. Çift Akışlı Tespit ve Segmentasyon Birleştirilmesi

Tek akışlı segmentasyon çıkarımının sınırlamalarını aşmak ve hem nesne tespiti hem de segmentasyon yaklaşımlarının güçlü yönlerinden eşzamanlı olarak yararlanmak amacıyla, paralel çift akışlı ve birleştirme temelli alternatif bir strateji önerilmektedir. Bu mimari, 4.1 bölümünde en yüksek performansı sergileyen tespit modeli YOLOv11 ile 4.2 bölümünde tanımlanan segmentasyon omurgası DeepLabV3+'ın entegrasyonuna dayanmaktadır. Her iki model, aynı girdi görüntüsünü birbirinden bağımsız ancak eşzamanlı biçimde işleyerek tamamlayıcı nitelikte çıktılar üretir: YOLOv11 modeli nesnelere ait güven skorlarıyla birlikte sınırlayıcı kutular sunarken, DeepLabV3+ modeli yoğun anlamsal segmentasyon maskeleri üretmektedir.

Birleştirme işlemi, çıkarım düzeyinde gerçekleştirilmekte olup, mekânsal hizalama ve anlamsal doğrulama adımlarını içeren özel bir son işlem modülü aracılığıyla yürütülmektedir. Bu bağlamda, tespit edilen sınırlayıcı kutular, anlamsal segmentasyon haritası üzerine projekte edilmekte ve her bir tespit, içeriksel tutarlılığı temelinde değerlendirilmektedir. Örneğin, YOLOv11 tarafından “insan” olarak sınıflandırılan bir sınırlayıcı kutunun, segmentasyon haritasında “ray” sınıfına karşılık gelen bir bölgeyle örtüşmesi durumunda, sistem ya mekânsal olasılık temelli bir çıkarımda bulunmakta ya da potansiyel bir anomaliyi işaretleyebilmektedir. Bu çapraz mod doğrulama yaklaşımı, özellikle yanlış pozitiflerin veya mekânsal tutarsızlıkların operasyonel risk oluşturabileceği güvenlik kritik uygulamalarda, karar alma sürecinin güvenilirliğini ve doğruluğunu anlamlı ölçüde artırmaktadır.

Ayrıca, önerilen mimari, her bir tespitin piksel düzeyindeki derinlik bilgisi ile birleştirilmesi yoluyla 3B koordinatlarının tahmin edilmesini mümkün kılan stereoskopik derinlik haritalarıyla modüler entegrasyonu da desteklemektedir. Bu entegrasyon, sistemin yalnızca iki boyutlu görüntü düzleminde değil, aynı zamanda üç boyutlu mekânsal bağlamda da anlamlı çıkarımlar yapabilmesine olanak tanır. Söz konusu yetenek, çarpışma tahmini, mesafeye dayalı uyarı sistemleri ve dinamik risk değerlendirmesi gibi güvenlik odaklı ve durumsal farkındalık gerektiren uygulamalarda gelecekte gerçekleştirilecek genişletmeler açısından kritik bir rol oynamaktadır.

Çift akışlı tasarımın temel avantajlarından biri, yüksek düzeyde modülerlik ve işlevsel ayrıştırma yeteneğidir. Segmentasyon ve tespit modülleri, birbirlerinden bağımsız olarak optimize edilebilir, güncellenebilir veya tamamen yeni modellerle değiştirilebilir. Bu esneklik, sistemin gelişen donanım altyapılarına uyum sağlamasını ve yeni algoritmaların tüm işlem hattını baştan tasarlamaya gerek kalmaksızın entegre edilmesini mümkün kılmaktadır. Ayrıca, her bir görev için ayrı işlem yollarının korunması, çok görevli öğrenme süreçlerinde sıklıkla karşılaşılan görevler arası girişim riskini ortadan kaldırmaktadır. Bu mimari yapı, ortak optimizasyon stratejilerinde

gözlemlenen gradyan çatışmaları ve özellik paylaşımından kaynaklanan rekabet problemlerini önleyerek, her görev için daha kararlı ve güvenilir bir öğrenme süreci sağlamaktadır.

4.3.3. Uygulama Çerçevesi

Her iki birleştirme stratejisi de, genişletilebilirlik ve gerçek zamanlı dağıtım gereksinimlerini karşılayacak şekilde modüler bir çıkarım çerçevesi içerisinde, PyTorch altyapısı kullanılarak uygulanmıştır. Segmentasyon yönlendirmeli tespit yaklaşımı için, OpenCV tabanlı görselleştirme rutinleriyle entegre çalışan ve son işlem aşamasında geliştirilen sezgisel algoritmalarla desteklenen birleşik bir anlamsal işleme boru hattı geliştirilmiştir.

Çift akışlı birleştirme stratejisinde ise, çok iş parçacıklı giriş akışları aracılığıyla paralel çıkarım gerçekleştirilmiş, birleştirme mantığı senkronize son işlem betikleri kullanılarak uygulanmış ve böylece eş zamanlı veri bütünlüğü sağlanmıştır.

Her iki boru hattı da, orijinal girdi görüntüleri, segmentasyon maskeleri, sınırlayıcı kutular ve derinlik projeksiyonlarını bütüncül biçimde görselleştirebilen gerçek zamanlı bir görselleştirme katmanı ile desteklenmiştir. Geliştirme süreci boyunca bu görselleştirme modülleri, hem model çıktılarının nitel değerlendirmesini gerçekleştirmek hem de birleştirme tutarlılığını doğrulamak amacıyla etkin biçimde kullanılmıştır. Her iki stratejiye ait örnek çıktı kareleri, başarılı ve başarısız örnekler de dâhil olmak üzere, 6. Bölüm'de sunulmuş ve kapsamlı bir şekilde analiz edilmiştir.

4.3.4. Stratejik Karşılaştırma

Segmentasyon yönlendirmeli tespit stratejisi, verimlilik, yapısal sadelik ve tam anlamsal hizalamayı önceliklendirerek, sınırlı hesaplama kaynaklarına sahip uç cihazlarda dağıtımına uygunluk açısından öne çıkmaktadır. Bu yönüyle, kaynak kısıtlı senaryolarda düşük gecikmeli ve güvenilir algı sistemlerinin gereksinimlerini karşılamak için ideal bir çözüm sunmaktadır. Buna karşılık, çift akışlı mimari; algısal zenginlik, görev bazlı uzmanlaşma ve derinlik farkındalıklı çıkarım yeteneklerini merkeze alarak, yarı otonom ve tam otonom demiryolu sistemlerinde sağlam ve kapsamlı algı modüllerinin temelini oluşturacak niteliktedir.

Bu iki strateji, çalışmanın temelini oluşturan çift yönlü tasarım yaklaşımını somutlaştırmaktadır. İlk strateji, dağıtılabilirlik ile metodolojik titizlik arasında dengeli ve sade bir yöntem benimserken; ikinci strateji, çok modlu sensör sistemlerinin ölçeklenebilir entegrasyonunu mümkün kılan gelişmiş bir yapısal esneklik üzerine odaklanmaktadır. “Bulgular ve Tartışma” bölümünde sunulan karşılaştırmalı analizler, bu stratejilerin doğruluk, işlem gecikmesi ve gerçek dünya demiryolu ortamlarındaki uyarlanabilirlik gibi kritik performans kriterlerine ilişkin deneysel düzeyde ayrıntılı ve kapsamlı içgörüler sunmaktadır.

4.4. ZED2 Kullanılarak Mesafe Ölçümü ve Derinlik Entegrasyonu

Bu tezde sunulan temel yeniliklerden biri, önerilen demiryolu sahne anlama çerçevesine derinlik algısının entegre edilmesidir. Geleneksel monoküler RGB tabanlı görüntüleme sistemleri, nesnelerin iki boyutlu görüntü düzleminde sınıflandırılması ve konumlandırılması görevlerinde başarılı performans sergilese de, nesneye olan mesafeyi doğrudan tahmin edebilme yeteneğinden yoksundur. Oysa, nesne uzaklığının doğru biçimde belirlenmesi, özellikle akıllı demiryolu sistemlerinde güvenliğe duyarlı uygulamalar açısından kritik bir gereksinimdir. Engellerden kaçınma, yakınlık temelli uyarı üretimi ve öngörüye dayalı çarpışma önleme gibi görevlerin tamamı, yalnızca semantik değil, aynı zamanda hassas üç boyutlu mekânsal farkındalık gerektirir.

Bu eksikliği gidermek amacıyla, bu çalışma gerçek zamanlı stereo görüşü ZED2 stereo kamera aracılığıyla sistem mimarisine entegre ederek, sahne anlayışını yalnızca anlamsal düzeyde değil, aynı zamanda metrik derinlik bilgisiyle de zenginleştirmektedir. Böylece sistem, hem semantik hem de geometrik bağlamda güçlendirilmiş çok boyutlu bir algılama yeteneği kazanmakta; bu da, güvenlik kritik demiryolu uygulamaları için daha sağlam ve öngörülü karar mekanizmalarının inşasına zemin hazırlamaktadır.

Şekil 4.12, bu tezde derinlik ölçümü ve üç boyutlu sahne anlayışının gerçekleştirilmesinde kullanılan ZED2 stereo kameranın yapısını görsel olarak sunmaktadır.



Şekil 4.12. ZED2 Kamera (stereolabs.com)

Stereolabs tarafından geliştirilen ZED2 stereo kamera, pasif stereo görüntüleme prensibine dayalı olarak çalışan ve yüksek çözünürlüklü RGB-D (renk + derinlik) verisini gerçek zamanlı olarak üretebilen gelişmiş bir algılayıcı sistemdir. Cihaz, her biri global shutter özelliğine sahip iki senkronize görüntü sensörü kullanmakta olup, 12 cm'lik sabit taban açıklığı ile derinlik algılama doğruluğunu artıracak şekilde tasarlanmıştır. ZED2, 2K çözünürlüğe kadar görüntü kaydı desteği sunarak hem detaylı hem de yüksek kaliteli veri üretimine olanak tanımaktadır.

Cihazın en öne çıkan teknik özelliklerinden biri, stereo görüntü çiftleri arasındaki ayrıklığı (disparity) hesaplayarak yoğun derinlik tahmini gerçekleştirebilmesidir. Bu işlem, özellikle gerçek zamanlı uygulamalarda yüksek doğruluk ve hız elde etmek amacıyla optimize edilmiş yarı-küresel eşleştirme (Semi-Global Matching, SGM) algoritması aracılığıyla yürütülmektedir. Bu yapı, derinlik verisinin hem nesne tespiti hem de mesafe ölçümü bağlamında güvenilir bir biçimde kullanılmasını mümkün kılmakta, böylece demiryolu gibi güvenlik-kritik alanlarda uygulanabilirliği artırmaktadır.

ZED2 ayrıca, yerleşik atalet ölçüm birimi (Inertial Measurement Unit - IMU), manyetometre ve barometre gibi sensörlerle donatılmıştır. Bu donanımlar, sensör füzyonu ve uzamsal-zamansal durum kestirimi gibi uygulamalar için cihazı oldukça uygun hale getirmektedir. Kameranın işlevselliği, Stereolabs tarafından sağlanan ZED SDK aracılığıyla daha da genişletilmiştir. Bu yazılım geliştirme kiti; görüntü düzeltme, ayrıklık haritası çıkarımı, nokta bulutu (point cloud) üretimi ve kesintisiz RGB-D veri entegrasyonu gibi temel işlevleri destekleyerek, ZED2'yi gerçek zamanlı 3B algılama çözümleri için güçlü ve esnek bir araç haline getirmektedir.

4.4.1. ZED2 Kamerası ile Stereo Derinlik Tahmini

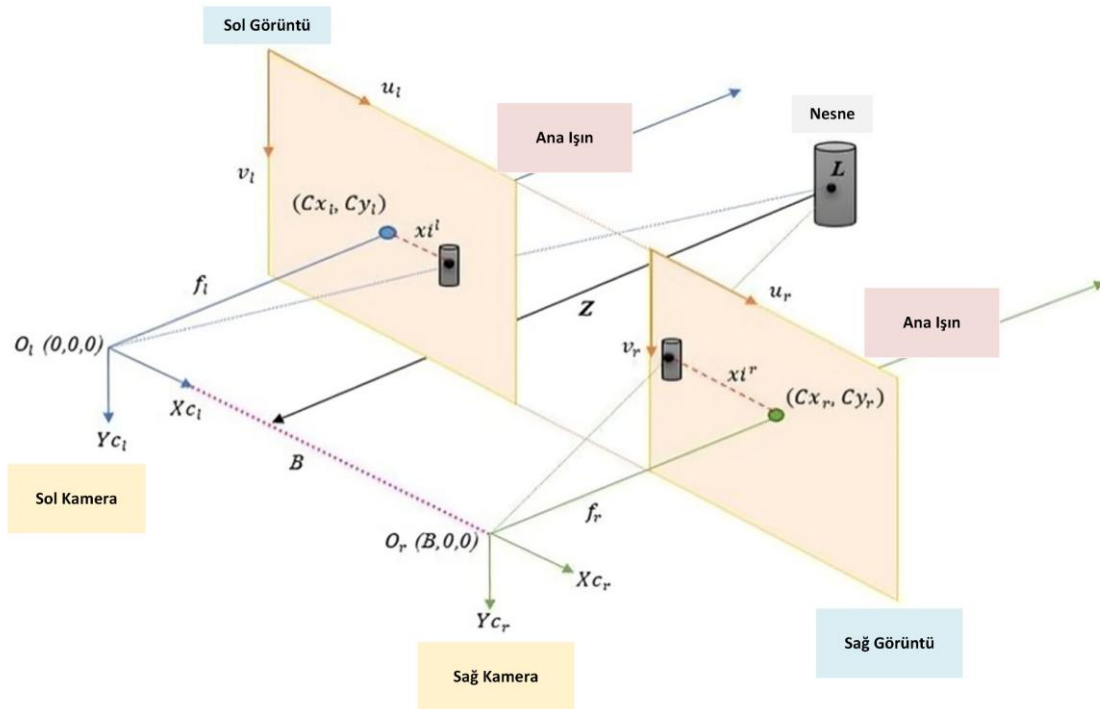
Stereo görüş, bilgisayarlı görü sistemlerinde derinlik algısının temel taşlarından birini oluşturmaktadır. İnsan gözünün binoküler yapısından esinlenerek geliştirilen bu yöntem, yatay düzlemde konumlandırılmış iki görüntü sensöründen elde edilen verilerle üç boyutlu mekânsal yapının yeniden inşasını mümkün kılar. Stereolabs tarafından geliştirilen ZED2 stereo kamera, bu ilkeyi kullanarak senkronize biçimde sol ve sağ görüntüleri yakalar; ardından ayrıklık (disparity) tahmini ve geometrik üçgenleme (triangulation) yoluyla sahne derinliğini çıkarır. Bu tür sistemlerde yüksek doğrulukta derinlik tahmini, kamera kalibrasyonu (Calibration), görüntü düzetimi (Rectification), ayrıklık hesaplama ve derinlik dönüşümünü (Depth Estimation) içeren sağlam bir işlem hattı gerektirir. Bu işlem adımlarının tümü, ZED Yazılım Geliştirme Kiti (SDK) tarafından entegre biçimde sunulmaktadır.

Derinlik tahmin süreci, kapsamlı bir stereo kalibrasyon adımıyla başlar. Bu adımda, her iki kameranın içsel parametreleri (odak uzunlukları f_x, f_y , ana nokta koordinatları C_x, C_y ve mercek bozulma katsayıları k_1, k_2, k_3, p_1, p_2) ile dışsal parametreleri (rotasyon matrisi R ve çeviri vektörü T) tahmin edilir. Bu parametreler, stereo çiftin geometrik yapısını tanımlar. ZED2 kamerası fabrika çıkışında kalibre edilmiş olarak sunulsa da, ZED SDK gerçek zamanlı kendini kalibrasyon yetenekleriyle çalışma sırasında parametreleri dinamik olarak güncelleyerek yeniden projeksiyon hatasını en aza indirir.

Kalibrasyonun ardından, stereo görüntü düzetimi gerçekleştirilir. Bu işlem, sol ve sağ görüntüleri epipolar geometriye uygun şekilde hizalayarak karşılık gelen piksellerin yatay

doğrultuda aynı çizgi üzerinde bulunmasını sağlar. Bu hizalama, eşleşen piksel arayışını tek boyutlu hale getirerek hem hesaplama verimliliğini artırır hem de eşleşme doğruluğunu yükseltir. Düzeltilmiş görüntü çiftleri daha sonra, ZED SDK'nin özel blok-eşleme (block-matching) algoritmasıyla işlenir ve piksellere özgü ayrıklık değerleri hesaplanır. Bu algoritma; alt piksel düzeyinde iyileştirme, örtüşme (occlusion) yönetimi ve güven tabanlı filtreleme gibi gelişmiş teknikler içererek düşük dokulu veya karmaşık bölgelerdeki belirsizlikleri azaltır.

Şekil 4.13 ve Şekil 4.14, ZED2 stereo kamerasının derinlik tahmini sürecinde izlediği temel adımları; stereo kalibrasyon, görüntü düzetimi, ayrıklık hesaplama ve derinlik çıkarımı aşamalarını şematik olarak göstermektedir.



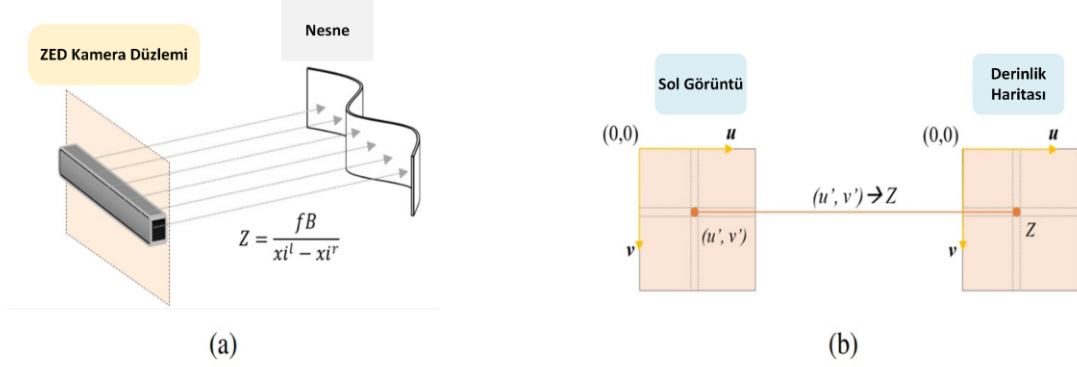
Şekil 4.13. ZED2 Kameranın Çalışma Prensibi [80]

Elde edilen ayrıklık haritası, stereo kameralar arasındaki bilinen taban açıklığı B ve odak uzaklığı f kullanılarak, klasik üçgenleme formülü aracılığıyla metrik ölçekli bir derinlik haritasına dönüştürülür:

$$Z = \frac{f \cdot B}{d} \quad (4.1)$$

Burada Z sahnedeki bir noktanın derinliğini, f odak uzaklığını, B kamera taban açıklığını ve d ayrıklık değerini ifade eder. Bu dönüşüm sonucunda her pikselin, kameraya olan gerçek mesafeyi temsil ettiği metrik doğrulukta bir derinlik görüntüsü elde edilir. Söz konusu derinlik haritası, RGB

görüntü ile uzamsal olarak hizalanmıştır ve bu sayede sahnenin eş zamanlı bir RGB-D temsili sağlanır. Bu bütünleşik yapı, nesne tespiti, örnek düzeyinde segmentasyon ve 3B sahne kavrayışı gibi üst düzey algılama görevlerinde derinlik bilgisinin etkin biçimde kullanılmasına olanak tanır.



Şekil 4.14. (a) Derinlik Tahmini. (b) Sol Görüntü ve Derinlik Haritası İlişkisi. [80]

Şekil 4.14, ZED stereo kameralarının üç boyutlu mekânsal verileri elde etme yöntemini göstermekte olup, sol kamera referans koordinat sistemi olarak alınmıştır. Sol kameranın görüntü düzleminde (u', v') piksel koordinatlarına sahip bir nokta verildiğinde, bu koordinatlar kullanılarak karşılık gelen derinlik değeri Z , Şekil 4.14.b'de gösterilen derinlik haritası üzerinden elde edilmektedir.

Bu tez kapsamında, ZED2 kamera sistemi 1280×720 (720p) çözünürlükte çalışacak şekilde yapılandırılmıştır; bu ayar, mekânsal çözünürlük ile gerçek zamanlı performans arasında etkili bir denge sunar. Tipik demiryolu sahnelerinde yapılan deneysel değerlendirmeler, stereo derinlik işlem hattının saniyede 15 kareyi (FPS) aşan bir işlem hızına ulaştığını ve nesne konumlandırma ile mesafe farkındalıklı risk analizi gibi derin öğrenme tabanlı alt görevler için yeterli uzamsal doğruluk sağladığını göstermektedir. Bu doğrultuda, ZED2 kamerası, demiryolu altyapı izleme ve otonom raylı sistemler gibi karmaşık ve dinamik dış mekân ortamlarında yoğun derinlik algılaması için güvenilir ve verimli bir çözüm sunmaktadır.

4.4.2. Nesne Konumlandırma İçin RGB-D Birleştirme

Tespit edilen nesnelerin üç boyutlu konumlarının doğru şekilde belirlenebilmesi amacıyla, YOLO tabanlı nesne tespit sisteminden elde edilen iki boyutlu sınırlayıcı kutular, ZED2 stereo kamera aracılığıyla elde edilen derinlik verisiyle zenginleştirilmiştir. Nesne tespitinin ardından, sınırlayıcı kutuya karşılık gelen İlgi Bölgesi (ROI), hizalanmış derinlik haritası üzerinde tanımlanmakta ve bu alan içerisindeki derinlik değerleri analiz edilmek üzere çıkarılmaktadır.

Derinlik verilerinde özellikle nesne kenarlarında ve yansıtıcı yüzeylerde sıkça karşılaşılan gürültü ve ölçüm tutarsızlıklarını azaltmak amacıyla, çıkarılan derinlik değerlerine medyan filtreleme işlemi uygulanmaktadır. Bu gürültü giderme adımı, aykırı değerleri bastırarak derinlik tahmininin kararlılığını artırmaktadır. Elde edilen medyan derinlik değeri, ilgili nesnenin kameraya olan tahmini mesafesi olarak atanır.

Sınırlayıcı kutuların, seyrek, bozulmuş ya da güvenilmez derinlik ölçümleri içeren bölgelerle çakıştığı durumlarda, uyarlanabilir ROI iyileştirme yöntemi devreye sokulmaktadır. Bu aşamada, anlamsal segmentasyon çıktıları kullanılarak derinlik ortalaması yalnızca tespit edilen nesne sınıfına ait piksellerle (örneğin, “insan” veya “araç”) sınırlandırılmakta ve böylece kısmen görünür ya da örtülmüş hedeflerin konumlandırma doğruluğu artırılmaktadır. Bu semantik kısıtlama, arka plan parazitlerinin etkisini azaltmakta ve hedefe özgü derinlik tahminini hassaslaştırmaktadır.

Gerçekleştirilen deneyler, önerilen RGB-D birleştirme yaklaşımının, değişken doğal aydınlatma koşulları, altyapı kaynaklı kısmi örtülmeler ve heterojen zemin geometrileri gibi zorlukların hâkim olduğu demiryolu ortamlarında mekânsal güvenilirlik ve konumlandırma dayanıklılığını anlamlı ölçüde artırdığını ortaya koymaktadır. Anlamsal ve geometrik ipuçlarının bütünleşik olarak kullanımı, güvenlik-kritik otonom demiryolu sistemlerinde dağıtılabilir sağlam bir algılama hattının temel bileşeni olarak öne çıkmaktadır.

4.4.3. Sınırlayıcı Kutu Verilerinin 3B Metrik Uzaya Eşlenmesi

Tespit edilen nesnelerle ilişkilendirilen derinlik değerlerinin ardından, her bir nesnenin merkez noktası (centroid), iki boyutlu görüntü koordinatlarından üç boyutlu metrik uzaya projekte edilerek konumsal hassasiyet artırılmıştır. Bu dönüşüm, ZED2 stereo kamera sistemine ait içsel kamera matrisinden yararlanılarak iğne deliği (pinhole) kamera modeli temelinde gerçekleştirilmiştir. Bu model, 2B piksel koordinatlarını ve ilgili derinlik değerlerini kullanarak uzamsal (x, y, z) koordinatlarına dönüştürür. Bu bağlamda:

- z , kameradan nesneye olan mesafeyi (derinlik değerini),
- x ve y , optik merkeze göre yatay ve dikey fiziksel konumu (genellikle metre cinsinden) ifade eder.

Elde edilen 3B koordinatlar, yalnızca görselleştirme amacıyla değil, aynı zamanda çarpışma riski tahmini ve yakınlık tabanlı güvenlik analizleri gibi kritik alt görevlerde de kullanılmıştır. Bu uzamsal veriler, OpenGL ve OpenCV kütüphaneleri kullanılarak gerçek zamanlı biçimde işlenmiş; her bir tespit edilen nesne için sınırlayıcı kutular, sınıf etiketleri ve tahmini mesafe bilgileri RGB görüntüsü üzerine bindirilmiştir. Bu eşzamanlı RGB-D görselleştirme, sistemin değerlendirme aşamasında yorumlanabilirliği artırmış ve insan operatör destekli gözetim sistemlerinde potansiyel uygulama alanları sunmuştur.

Ayrıca, nesnelere ait 3B merkez noktalarının ardışık kareler boyunca izlenmesiyle, zamansal hareket rotaları çıkarılmıştır. Bu izleme süreci, zamansal tutarlılığın denetlenmesini sağlamış ve yörünge tabanlı davranış çıkarımını mümkün kılmıştır. Özellikle güvenlik-kritik demiryolu uygulamalarında, anomali tespiti, hareket tahmini ve dinamik risk modelleme gibi ileri düzey görevler açısından önemli bir altyapı sağlamıştır.

4.4.4. Gerçek Zamanlı Çoklu Görev İşlem Hattı ile Entegrasyon

Bu çalışmada geliştirilen çoklu görevli çıkarım mimarisi, ZED2 temelli RGB-D görüntüleme modülü ile bütünleşik bir yapıda tasarlanmıştır. Her bir giriş çerçevesi alındığında sistem, çok iş parçacıklı (multi-threaded) bir yürütme modeli kullanarak görsel veriyi eşzamanlı biçimde üç farklı görev modülüne yönlendirmektedir: anlamsal segmentasyon, nesne tespiti ve derinlik tahmini. Bu modüller birbirinden bağımsız olarak çalışmakta ve her biri kendi görevine özgü optimize edilmiş anlamsal ve geometrik çıktı kümeleri üretmektedir.

Çıkarım aşamasını takiben, nihai işlem modülü bu çıktılar arasında senkronizasyon işlemini gerçekleştirmektedir. Özellikle, nesnelere ait sınırlayıcı kutular (bounding boxes) ve segmentasyon maskeleri, ZED2 sensörü tarafından sağlanan derinlik haritası ile uzamsal olarak hizalanmakta; böylece her bir nesneye ilişkin sınıf etiketi, konumsal sınırlar ve metrik derinlik bilgisi içeren açıklamalı birleşik RGB-D çıktılar oluşturulmaktadır. Bu bütünleşik çıktı yapısı, hem insan yorumlayıcılar hem de otonom karar verme sistemleri açısından çevresel algının bütüncül ve tutarlı bir biçimde temsiline olanak tanımaktadır.

Sistem mimarisi, gerçek zamanlı çalışma koşullarını destekleyecek şekilde kapsamlı biçimde optimize edilmiştir. CUDA tabanlı GPU hızlandırması, çok işlemcili görev zamanlayıcıları ve verimli bellek yönetimi stratejileri ile işlem hacmi yüksek görevlerin zamanında ve kararlı biçimde yürütülmesi sağlanmıştır. Ayrıca, dinamik çalışma koşullarına adaptasyon amacıyla, toplu işlem boyutları (batch size) ve tampon bellek derinlikleri çalışma zamanı sırasında otomatik olarak ayarlanabilmektedir. Sistem entegrasyonu, ZED Python API'si ile PyTorch temelli çıkarım motorunun birlikte kullanılması yoluyla gerçekleştirilmiş olup, gerçekleştirilen test senaryolarında uçtan uca (end-to-end) işlem gecikmesi büyük oranda 200 milisaniyenin altında kalacak biçimde sınırlandırılmıştır.

Önerilen modüler mimari, genişletilebilirlik ve esneklik açısından yüksek düzeyde uyarlanabilirlik sunmaktadır. Mimari yapının modüler tasarımı, sistemin yeniden yapılandırılmasına gerek kalmaksızın farklı bileşenlerin entegrasyonuna olanak tanımaktadır. Bu sayede, gelecekte ek uzamsal filtrelerin, yeni görev modüllerinin veya alternatif derinlik algılayıcılarının sisteme dâhil edilmesi mümkün hâle gelmektedir.

Söz konusu tasarım yaklaşımı, sistemin hem donanım teknolojilerindeki gelişmelere hem de farklı uygulama senaryolarının gerektirdiği operasyonel koşullara ölçeklenebilir şekilde uyum

sağlamasını temin etmektedir. Bu durum, önerilen yapının uzun vadeli sürdürülebilirlik, yeniden kullanılabilirlik ve sistem seviyesinde esneklik açısından avantajlı bir çözüm sunduğunu göstermektedir.

4.4.5. Tez Hedefleriyle Uyum

Bu tez kapsamında geliştirilen sistem mimarisi içerisinde derinlik farkındalıklı işlem kabiliyetlerinin bütünleştirilmesi, çalışmanın temel araştırma hedeflerinden biri olan kapsamlı, gerçek zamanlı ve çok görevli bir demiryolu algılama çerçevesinin oluşturulması amacını doğrudan desteklemektedir. Stereo görüş teknolojisinin sistem bileşenlerine entegre edilmesiyle, yalnızca iki boyutlu anlamsal çıktılara dayalı sınırlı bir algı düzeyinden, üç boyutlu uzamsal farkındalıkla zenginleştirilmiş, yüksek bilinç düzeyine sahip bir çevresel algı sistemine geçiş sağlanmıştır. Bu yapısal dönüşüm, özellikle güvenlik açısından kritik uygulamalarda bağlama duyarlı ve yüksek güvenilirlikte karar mekanizmalarının geliştirilmesine olanak tanımaktadır.

Sistem mimarisine entegre edilen derinlik verisi, aşağıda sıralanan güvenlik-kritik senaryolarda karar destek süreçlerinin sağlamlaştırılmasına katkı sunmaktadır:

- Ray hattı üzerinde veya çevresinde tespit edilen yabancı cisimlerin konumlarının ve metrik uzaklıklarının hassas biçimde belirlenmesi suretiyle, dinamik çarpışma önleme sistemlerinin etkinleştirilmesi;
- Demiryolu çevresine izinsiz giriş yapan yaya ya da kara taşıtlarının algılanması durumunda, konuma ve mesafeye duyarlı uyarı sistemlerinin otomatik olarak devreye sokulması;
- Sinyalizasyon üniteleri, makas mekanizmaları ve platformlar gibi altyapıya ait kritik yapısal bileşenlerin üç boyutlu uzayda konumlandırılması aracılığıyla, dijital ikiz sistemleriyle ileri düzeyde entegrasyon stratejilerinin araştırılmasına olanak tanınması.

İki boyutlu anlamsal segmentasyon ve nesne tespit çıktılarının üç boyutlu uzamsal temsillere dönüştürülebilmesi, demiryolu sahne algılama alanında önemli bir metodolojik paradigma değişimini temsil etmektedir.

Bu dönüşüm, yalnızca görsel temsillerin zenginleştirilmesiyle sınırlı kalmamakta; aynı zamanda mekânsal çıkarım, risk analizi ve operasyonel planlama gibi üst düzey görevlerin bilgi temelli bir biçimde yürütülmesini de mümkün kılmaktadır.

ZED2 stereo kamera sistemine dayalı RGB-D veri işleme hattı aracılığıyla gerçekleştirilen anlamsal bilgi ile geometrik farkındalık arasındaki bu bütünsel füzyon, yeni nesil demiryolu otomasyon sistemlerinin uygulanabilirliğini, ölçeklenebilirliğini ve çevresel bağlama uyarlanabilirliğini artıran akıllı görüntüleme çözümlerinin geliştirilmesi için sağlam ve genişletilebilir bir teknolojik temel oluşturmaktadır.

5. DENEYSEL KURULUM

Bu bölüm, önerilen çok görevli algı sisteminin geliştirildiği, eğitildiği ve değerlendirildiği deneysel çerçeveyi ortaya koymaktadır. Gerçek zamanlı demiryolu sahne anlayışı bağlamında çeşitli nesne tespiti, anlamsal segmentasyon ve derinlik tahmini modellerinin performansını, verimliliğini ve uygulanabilirliğini titizlikle değerlendirmek amacıyla deneyler planlanmıştır.

Bölüm 5.1, uygulama sürecinde kullanılan donanım ve yazılım yapılandırmalarını tanımlamakta; hesaplama platformları, geliştirme ortamları ve destekleyici kütüphaneler hakkında bilgi vermektedir. Bölüm 5.2, hem tespit hem de segmentasyon ağları için uygulanan eğitim yapılandırmalarını ayrıntılı şekilde sunmakta; başlangıç parametreleri, kayıp fonksiyonları, optimizasyon stratejileri ve çalışma süresine ilişkin hususları kapsamaktadır. Bölüm 5.3, model etkinliğini ölçmek için kullanılan değerlendirme metriklerini tanıtmaktadır; bu metrikler arasında kesinlik (precision), duyarlılık (recall), ortalama ortalama kesinlik (mAP), Birleşim Üzerinden Kesişim (IoU), kare hızı (frame rate) ve derinlik tahmin doğruluğu yer almaktadır. Bu bileşenler bir araya geldiğinde, model performansının nicel olarak değerlendirilmesine ve bütünleşik algı sisteminin pratik uygulanabilirliğinin doğrulanmasına yönelik sağlam bir deneysel temel oluşturmaktadır.

5.1. Donanım ve Yazılım Yapılandırması

Önerilen çok görevli algılama çerçevesinin ölçeklenebilirliğini, modülerliğini ve pratik uygulanabilirliğini titizlikle değerlendirmek amacıyla tüm deneyler, hem yüksek performanslı yerel bir çalışma istasyonu hem de Google Colab Pro+ üzerinden erişilen bulut tabanlı GPU kaynaklarını içeren heterojen bir hesaplama ortamında gerçekleştirilmiştir. Bu çift platformlu yaklaşım, hızlı prototipleme ve yinelemeli deneysel çalışmaları yüksek hacimli model eğitimi ve gerçek zamanlı çıkarım doğrulama gereksinimleriyle dengelemek amacıyla stratejik olarak seçilmiştir. Şekil 5.1, deneylerde kullanılan bulut tabanlı hesaplama ortamı Google Colaboratory platformunun logosunu göstermektedir.



Şekil 5.1. Google Colaboratory Logosu

Yerel hesaplama altyapısı, NVIDIA GeForce RTX 2070 Max-Q GPU (8 GB VRAM), Intel Core i7-10750H CPU (6 çekirdek, 12 iş parçacığı, 2.60 GHz temel frekans) ve 16 GB DDR4 RAM ile donatılmış bir geliştirme istasyonundan oluşmaktaydı. Bu yapılandırma, entegre RGB-D segmentasyon ve nesne tespiti hatlarının gecikmeye duyarlı koşullar altında, özellikle demiryolu algılama sistemlerinde karşılaşılan gerçek dünya kısıtlarını yansıtan durumlarda, verimli bir şekilde dağıtılmasını ve gerçek zamanlı çalıştırılmasını mümkün kılmıştır.

Derin evrişimsel mimariler (örneğin DeepLabV3+ ve PSPNet) içeren hesaplama açısından yoğun eğitim prosedürleri için, ek deneyler Google Colab Pro+ üzerinde gerçekleştirilmiştir. Bu platform, dinamik olarak NVIDIA Tesla T4 ve A100 GPU'lara erişim sağlamış; 16 GB ile 52 GB arasında değişen sistem belleği sunmuştur. CUDA hızlandırmalı altyapılar üzerine inşa edilen Colab ortamı, yüksek hacimli model eğitimi, otomatik kaynak tahsisi ve veri kümeleri, kontrol noktaları ve ara sonuçların takibi için Google Drive ile bütünleşik veri yönetimi sağlamıştır.

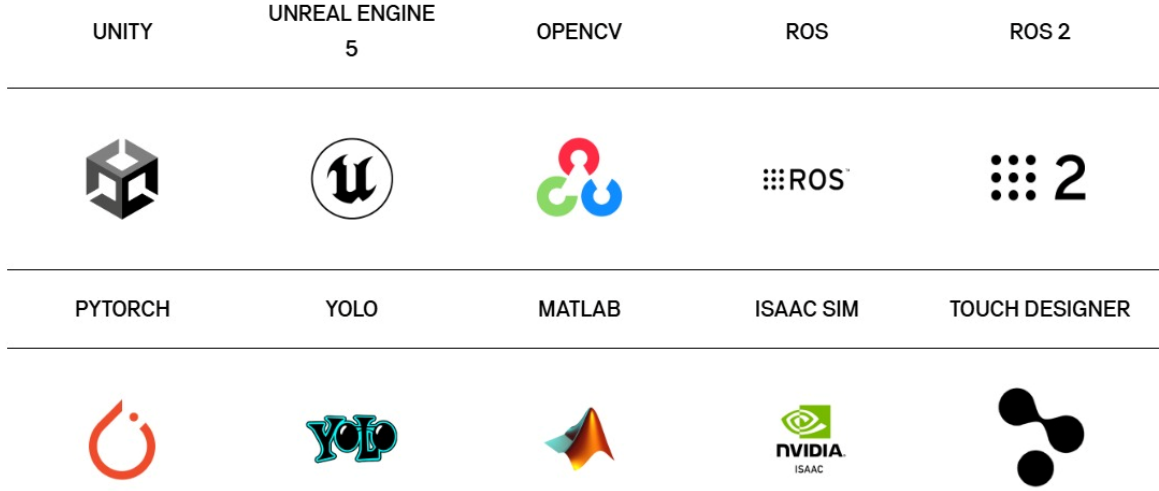
Yazılım yığını Python 3.10 kullanılarak geliştirilmiş olup, model inşası, eğitimi ve değerlendirmesi için birincil derin öğrenme çerçevesi olarak PyTorch 2.0.1 tercih edilmiştir. GPU hızlandırması, tüm sinir ağı mimarileri için gradyan hesaplaması ve bellek yönetimi açısından optimize edilmiş rutinler sunan CUDA 11.8 ve cuDNN kütüphanesi aracılığıyla etkinleştirilmiştir. Veri ön işleme, artırma ve dönüştürme hatları, her görev türünün modalite gereksinimlerine özel olarak uyarlanmış Albumentations ve Torchvision kütüphaneleri kullanılarak uygulanmıştır.

YOLOv8, YOLOv9, YOLOv11 ve YOLOv12'yi içeren nesne tespiti modülleri, Ultralytics açık kaynaklı çerçeve kullanılarak hayata geçirilmiştir. Bu platform, genişletilebilir eğitim arayüzleri, özelleştirilebilir artırma iş akışları ve entegre değerlendirme araçları sunmuştur. Ayrıca, Facebook AI Research (FAIR) tarafından geliştirilen Detectron2 çerçevesi, Faster R-CNN mimarisinin uygulanması ve karşılaştırmalı değerlendirilmesi için kullanılmıştır. Görevler arası değerlendirmeyi kolaylaştırmak amacıyla, anlamsal segmentasyon maskelerini COCO uyumlu sınırlayıcı kutu açıklamalarına dönüştürmek üzere özel ön işleme betikleri geliştirilmiştir.

Anlamsal segmentasyon modelleri arasında yer alan UNet'in temel ve önceden eğitilmiş varyantları, DeepLabV3+, Feature Pyramid Network (FPN) ve PSPNet, modüler PyTorch bileşenleri kullanılarak geliştirilmiştir. Bu süreçte, önceden eğitilmiş kodlayıcı omurgalarının entegrasyonuna ve yapılandırılabilir kod çözücü mimarilerinin kullanımına imkân tanıyan Segmentation Models PyTorch (SMP) kütüphanesinden bazı modüller uyarlanmış; böylece modeller arasında karşılaştırmalı analizlerin gerçekleştirilmesi kolaylaştırılmıştır.

Derinlik algısı ve stereo görüş işlevselliği, Stereolabs tarafından sağlanan ZED SDK v4.0 ve ilgili PyZED Python API aracılığıyla gerçekleştirilmiştir. Bu araçlar, ZED2 stereo kamera ile stereo görüntü edinimi, disparite kestirimi, iç ve dış kamera kalibrasyonu ve nokta bulutu üretimini desteklemiştir. Gerçek zamanlı görselleştirme, son işleme ve çerçeve düzeyinde açıklamalı bindirmeler ile derinlik farkındalıklı nesne konumlandırma mekanizmaları, OpenCV 4.8 ve

OpenGL kullanılarak geliştirilmiştir. Şekil 5.2, ZED2 kameranın kapsamlı API'sinin, farklı programlama dilleri ve platformlarla çoklu ortam uyumluluğunu şematik olarak göstermektedir.



Şekil 5.2. ZED Kameranın Kapsamlı API'sinin Çoklu Ortam Uyumluluğu

Bu entegre donanım-yazılım ekosistemi, hem kontrollü hem de dinamik çevresel koşullar altında tekrarlanabilir ve ölçeklenebilir çok görevli öğrenme deneylerinin yürütülmesini mümkün kılmıştır. Mimari olarak modüler tasarıma ve gerçek zamanlı performansa yapılan vurgu, özellikle demiryolu odaklı algılama sistemleri bağlamında, dağıtımaya yönelik çağdaş araştırma standartlarıyla hizalanmıştır.

5.2. Eğitim Yapılandırması

Tüm deneysel senaryolar boyunca yöntemsel tutarlılığı sağlamak ve titiz, karşılaştırılabilir performans değerlendirmelerini mümkün kılmak amacıyla birleşik ve tekrarlanabilir bir eğitim çerçevesi geliştirilmiştir. Bu çerçeve, hem anlamsal segmentasyon hem de nesne tespiti görevlerine tutarlı bir yapıda uygulanmış; ancak görevler arası yapısal farklılıklar dikkate alınarak hiperparametre ayarları, optimizasyon stratejileri ve veri ön işleme protokolleri açısından kontrollü esneklik sağlanmıştır.

Model yapılandırmaları, mevcut literatürde yaygın kabul gören yerleşik en iyi uygulamalara (best practices) dayandırılmış olup, aynı zamanda RailSem19 veri kümesinin özgül yapısal ve içeriksel özelliklerine uyum sağlayacak şekilde uyarlanmıştır. Bu yaklaşım, modellerin hem deneysel geçerliliğini hem de görev bağlamına uygunluğunu güvence altına almaktadır.

Buna ek olarak, demiryolu izleme sistemlerinde pratik konuşlanma (deployment) senaryolarının gerektirdiği zaman kısıtlı işlem koşulları göz önünde bulundurulmuş ve gerçek

zamanlı çıkarım (inference) yeteneğinin uygulanabilirliğine özel önem verilmiştir. Bu doğrultuda geliştirilen çerçeve, yalnızca akademik doğrulama açısından değil, aynı zamanda saha uygulamaları açısından da işlevsel bir temel sunmaktadır.

5.2.1. Nesne Tespiti Modelleri

Bu çalışmanın nesne tespiti bileşeni, son teknoloji bir dizi modeli değerlendirmiştir: YOLOv8, YOLOv9, YOLOv11, YOLOv12 ve Faster R-CNN. Her model, mimari tasarımı ve operasyonel paradigmasına uygun biçimde yapılandırılmış ve eğitilmiştir. RailSem19 veri setinde yerel sınırlayıcı kutu (bounding box) açıklamaları bulunmadığından, tüm nesne tespiti modelleri, Bölüm 4'te ayrıntılı olarak açıklanan sahte etiketlenmiş sınırlayıcı kutular (pseudo-labels) üzerinde eğitilmiştir.

YOLO tabanlı mimarilerin eğitimi, Ultralytics çerçevesi kullanılarak gerçekleştirilmiştir. Giriş görüntüleri, çözünürlük ve hesaplama yükünü standartlaştırmak amacıyla 640×640 piksel boyutuna yeniden ölçeklendirilmiştir. Uygun olduğu durumlarda, modeller çapasız (anchor-free) tespit mekanizmalarıyla çalıştırılmıştır. Eğitim sürecinde mozayik artırma, HSV tabanlı renk dalgalanması, rastgele ölçekleme ve yatay çevirme gibi çeşitli veri artırma teknikleri uygulanmıştır. Bu artırma stratejileri, özellikle demiryolu altyapı sahnelerinde karşılaşılan zorlu görsel koşulların etkisini azaltmak ve gizlenme, değişken aydınlatma ile karmaşık arka plan dokuları gibi faktörlere karşı modelin genelleme yeteneğini artırmak amacıyla tercih edilmiştir.

Optimizasyon, momentum katsayısı 0.937 ve ağırlık bozunumu (weight decay) 5×10^{-4} olarak ayarlanmış Stokastik Gradyan İnişi (Stochastic Gradient Descent - SGD) algoritması ile 50 epok (epochs) boyunca gerçekleştirilmiştir. Öğrenme oranı (learning rate) başlangıçta 1×10^{-3} olarak belirlenmiş ve kosinüs azalma (cosine annealing) öğrenme oranı çizelgesi uygulanmıştır. Kayıp fonksiyonu olarak, nesne varlığı, sınıflandırma ve konumlandırma bileşenlerini tek bir formülasyonda bütünleştiren ve hizasız sınırlayıcı kutuları ile oran tutarsızlıklarını cezalandıran Birlik Üzerinde Tam Kesişim (Complete Intersection over Union - CIoU) kaybı tercih edilmiştir. Bu sayede tespit doğruluğu artırılmıştır.

Faster R-CNN modeli, Detectron2 kütüphanesi kullanılarak ResNet-50 omurgası ve çok ölçekli özellik temsiline olanak tanıyan Feature Pyramid Network (FPN) ile birlikte uygulanmıştır. Açıklama verileri COCO standardına uygun hale getirilmiş ve bu sayede eğitim hattı ile uyumluluk sağlanmıştır. Model, hem sınıflandırma hem de sınırlayıcı kutu regresyon hedeflerini içeren çok görevli standart kayıp fonksiyonu ile eğitilmiştir. Eğitim, 2'lik yığın boyutu (batch size) ile 15.000 iterasyon boyunca sürdürülmüş; performans değerlendirmesi ise, %50 eşleşme oranı (IoU) eşiğinde ortalama ortalama kesinlik (mAP@0.5) ve %50 ila %95 aralığında (mAP@0.5:0.95) hesaplanarak yapılmıştır.

5.2.2. Anlamsal Segmentasyon Modelleri

Bu bölümde, UNet (temel ve önceden eğitilmiş varyantları), DeepLabV3+, Feature Pyramid Network (FPN) ve Pyramid Scene Parsing Network (PSPNet) gibi son teknoloji anlamsal segmentasyon mimarilerinin performansları, standartlaştırılmış bir deneysel çerçeve altında değerlendirilmiştir. Böylece performans metriklerinin geçerliliği ve karşılaştırılabilirliği sağlanmıştır. Tüm modeller, PyTorch derin öğrenme çerçevesi kullanılarak uygulanmış ve ya sıfırdan eğitilmiş ya da ImageNet üzerinde önceden eğitilmiş kodlayıcı omurgalarıyla ince ayar yapılmıştır. Özellikle, UNet-Pretrained ve FPN için ResNet34, DeepLabV3+ ve PSPNet için ise ResNet50 kodlayıcı olarak kullanılmıştır.

Deneyler arasında tutarlılığı korumak amacıyla, tüm giriş görüntüleri ve karşılık gelen yer doğruluğu içeren segmentasyon maskeleri 256×256 piksel çözünürlüğe yeniden boyutlandırılmıştır. Bu çözünürlük, mekânsal ayrıntının korunması ile hesaplama açısından yönetilebilirlik arasında optimal bir denge sağlamıştır. Doğruluk maskeleri, iki boyutlu etiket haritaları biçiminde kodlanmış; tanımsız ya da belirsiz alanlara 255 değeri atanarak bu bölgelerin eğitim sırasında gradyan tabanlı optimizasyondan hariç tutulması sağlanmıştır.

Model eğitimi, başlangıç öğrenme oranı 1×10^{-3} olan Adam optimizasyon algoritmasıyla gerçekleştirilmiştir. Yakınsama davranışındaki doğrusal olmayanlıkları dengelemek ve aşırı uyumu (overfitting) önlemek amacıyla ReduceLROnPlateau öğrenme oranı ayarlayıcısı kullanılmıştır. Bu çizelge, doğrulama kaybı üst üste üç epok boyunca iyileşmediğinde öğrenme oranını 0.1 katsayısıyla azaltmıştır. Temel hedef fonksiyonu, çok sınıflı segmentasyon görevleri için uygun olan kategorik çapraz entropi kaybıdır.

Tüm modeller, CUDA destekli grafik işlem birimleri (GPU'lar) üzerinde, tek hassasiyetli kayan nokta (FP32) hesaplamaları kullanılarak eğitilmiştir. Eğitim süreci, 4 örnekten oluşan yığın (batch) boyutu ve 40 epokluk bir eğitim periyodu ile gerçekleştirilmiştir. Model kontrol noktaları, doğrulama veri kümesinde elde edilen en yüksek ortalama Intersection over Union (mIoU) değerine göre periyodik olarak kaydedilmiş; böylece yalnızca doğruluk açısından en başarılı sürümlerin korunması sağlanmıştır.

Model değerlendirmesinde temel nicel metrikler olarak ortalama Intersection over Union (mIoU) ve Piksel Doğruluğu (Pixel Accuracy, PA) kullanılmıştır. Bu metrikler her epok sonunda hesaplanmış ve modelin performansı zamana bağlı olarak sürekli izlenmiştir.

Modelin genelleme kapasitesini artırmak ve farklı sahne yapıları, ışık koşulları ve görsel varyasyonlara karşı dayanıklılığını geliştirmek amacıyla, eğitim sürecine çeşitli veri artırma (data augmentation) stratejileri entegre edilmiştir. Uygulanan dönüşümler arasında rastgele yatay çevirme, rastgele kırpma işlemleri ile parlaklık, kontrast ve doygunluk üzerinde rastgele değişiklikler içeren renk dalgalanması (color jittering) yer almaktadır. Tüm bu artırma işlemleri,

eğitim sırasında dinamik olarak Albumentations kütüphanesi aracılığıyla uygulanmış ve bu sayede hem verimli hem de tekrarlanabilir dönüşümler elde edilmiştir.

5.2.3. Derinlik Tahmini ve Kalibrasyon

ZED2 stereo kamera sistemi kendi entegre derinlik tahmin hattını sunsa da, segmentasyon ve nesne tespiti modüllerinde kullanılan RGB giriş akışı ile hizalamayı sağlamak amacıyla ek kalibrasyon ve tutarlılık doğrulama prosedürleri uygulanmıştır. ZED SDK'nin kalibrasyonu, üretici tarafından sağlanan resmi araçlar kullanılarak gerçekleştirilmiş; böylece stereo görüntü çiftleri arasında subpiksel hizalama sağlanmış ve kameranın içsel matris parametrelerinin kestirimi iyileştirilmiştir.

ZED2 sisteminin ürettiği derinlik ölçümlerinin güvenilirliğini ve doğruluğunu değerlendirmek amacıyla, deneysel sahnede mesafeleri bilinen nesnelerin konumlandırıldığı ampirik referans veriler toplanmıştır. Bu amaçla, standart demiryolu rayları boyunca önceden kalibre edilmiş nesne yerleşimleri kullanılmıştır. Stereo kameradan elde edilen derinlik değerleri, bu gerçek dünya ölçümleriyle nicel olarak karşılaştırılarak sistemin hassasiyeti analiz edilmiştir. Özellikle düşük dokulu alanlarda ve uzak mesafelerde gözlemlenen yüksek frekanslı gürültü ile kararsızlıkları azaltmak amacıyla, ham disparite haritalarına disparite filtreleme ve medyan yumuşatma gibi görüntü işleme teknikleri uygulanmıştır.

Stereo derinlik tahmin modülü, kare başına işlem hızı, uçtan uca gecikme süresi ve mekânsal derinlik doğruluğu açısından sistematik olarak değerlendirilmiştir. Nesne tespiti ve anlamsal segmentasyon hatlarıyla entegrasyon, eşzamanlı çok iş parçacıklı bir akış mimarisi aracılığıyla sağlanmıştır. Sistem yapılandırması ve eğitim sırasında, RGB-D veri işleme hattının çıkarım gecikmesini kare başına 200 milisaniyenin altında tutması için özel optimizasyonlar gerçekleştirilmiş; bu, Bölüm 1'de belirtilen gerçek zamanlı performans gereksinimleriyle uyumlu olarak gerçekleştirilmiştir.

5.3. Değerlendirme Metrikleri

Güvenlik açısından kritik öneme sahip demiryolu altyapısı gibi ortamlarda, sağlam derin öğrenme sistemlerinin geliştirilmesinde doğru ve kapsamlı değerlendirme süreci hayati bir unsur olarak öne çıkmaktadır. Bu çalışmada ele alınan nesne tespiti, anlamsal segmentasyon ve ZED2 stereo kamera kullanılarak gerçekleştirilen derinlik tahmini gibi görevlerin her biri, kendilerine özgü değerlendirme metrikleri gerektirmektedir. Bu bölümde, geliştirilen modellerin söz konusu üç görev alanındaki başarımını nicel olarak değerlendirmek amacıyla kullanılan yöntemler ayrıntılı şekilde sunulmaktadır. Belirlenen değerlendirme ölçütleri, hem bilgisayarla görme alanında kabul görmüş standartları hem de demiryolu uygulamalarına özgü ihtiyaçları dikkate alarak

belirlenmiştir. Özellikle gerçek zamanlı tepki süresi, tespit doğruluğu ve mekânsal konumlama hassasiyeti gibi uygulamaya özel gereksinimler, değerlendirme sürecinin temel odak noktalarını oluşturmaktadır.

5.3.1. Nesne Tespiti Metrikleri

Demiryolu ortamlarında nesne tespiti; sinyallerin, engellerin, raylı taşıtların ve işaret levhalarının tanımlanmasında kritik bir rol oynar. Değerlendirme metriklerinin bu nedenle sadece genel doğruluğu değil, aynı zamanda kısmi örtüşme, hareket bulanıklığı ve değişken aydınlatma gibi zorlu koşullar altındaki model davranışını da yansıtması gerekmektedir.

Kesinlik (Precision), Duyarlılık (Recall) ve F1-Skoru

TP, FP ve FN sırasıyla doğru pozitifler, yanlış pozitifler ve yanlış negatifleri temsil etsin. Kesinlik, duyarlılık ve harmonik ortalaması F1-Skorunu şu şekilde tanımlanır:

$$Kesinlik = \frac{TP}{TP+FP} \quad (5.1)$$

$$Duyarlılık = \frac{TP}{TP+FN} \quad (5.2)$$

$$F1 - Skoru = 2 * \frac{Kesinlik * Duyarlılık}{Kesinlik + Duyarlılık} \quad (5.3)$$

Demiryolu güvenliği uygulamalarında, özellikle tehlikeli engellerin atlanmaması adına yüksek duyarlılık (recall) hayati önem taşırken, yanlış alarmları azaltmak adına yüksek kesinlik (precision) de gereklidir.

Nesne Üzerine Birleşim (IoU - Intersection over Union)

Her bir tespit edilen nesne için, yer gerçeği (Ground Truth) ile IoU değeri hesaplanır:

$$IoU = \frac{|Prediction \cap Ground Truth|}{|Prediction \cup Ground Truth|} \quad (5.4)$$

IoU değeri genellikle 0.5'in üzerinde ise tespit doğru kabul edilir.

Ortalama Ortalama Kesinlik (mAP - mean Average Precision)

Yerelleştirme doğruluğunu ve dayanıklılığı değerlendirmek amacıyla, COCO değerlendirme protokolünde tanımlandığı şekilde, farklı IoU eşiklerinde ($mAP@0.5$ ve $mAP@[0.5 : 0.95]$) ortalama ortalama kesinlik kullanılır. mAP formel olarak şu şekilde tanımlanır:

$$mAP = \frac{1}{N} \sum_{i=1}^N \int_0^1 p_i(r) dr \quad (5.5)$$

Burada $p_i(r)$, i -inci sınıf için duyarlılığın fonksiyonu olarak kesinliği; N ise sınıf sayısını gösterir. $mAP@0.5$ ve $mAP@[0.5 : 0.95]$ birlikte raporlanır ve farklı tespit eşiklerinde dengeli bir değerlendirme sunar.

Çıkarım Süresi ve Gerçek Zamanlı Performans

Tren üstü görüş sistemlerinin gerçek zamanlılık kısıtları göz önünde bulundurularak, ortalama çıkarım süresi (inference time) ve saniyedeki kare sayısı (FPS) değerlendirme donanımı üzerinde ölçülmüştür. 10–15 FPS altında çalışan sistemlerin, hızlı hareket eden lokomotiflerde pratik kullanımı sınırlı olabilir.

5.3.2. Anlamsal Segmentasyon Metrikleri

Anlamsal segmentasyon, demiryolu sahnesinin ray, balast, bitki örtüsü, altyapı unsurları ve güvenlik bölgeleri gibi anlamsal olarak anlamlı bölgelere yoğun, piksel düzeyinde sınıflandırılmasını sağlayarak algılama sisteminde kritik bir rol oynar. Doğru segmentasyon, engel konumlandırma, ray izleme ve açıklık tahmini gibi aşağı akış modülleri için hayati öneme sahiptir.

Segmentasyon performansını değerlendirmek için model davranışının tamamlayıcı yönlerini yansıtan üç temel nicel kriter kullanılmıştır: En İyi Birleşim Üzerinden Kesişim (Best Intersection over Union - IoU), En İyi Doğruluk (Best Accuracy) ve En İyi Kayıp (Best Loss). Bu metrikler, eğitim ve doğrulama aşamalarında hesaplanmakta ve en iyi model kontrol noktalarının belirlenmesi ile model varyantlarının karşılaştırılmasında kullanılmaktadır.

Birleşim Üzerinden Kesişim (IoU) – En İyi IoU (Best - IoU)

IoU, diğer adıyla Jaccard İndeksi, semantik segmentasyon görevlerinde model performansını değerlendirmek amacıyla en yaygın kullanılan ve yorumlanması en açık kıyaslama metriklerinden biridir. Belirli bir sınıf c için IoU, model tarafından tahmin edilen segmentasyon maskesi ile karşılık gelen gerçek (ground truth) segmentasyon maskesi arasındaki kesişim ve birleşim oranı olarak tanımlanır:

$$IoU_c = \frac{|Prediction_c \cap Ground Truth_c|}{|Prediction_c \cup Ground Truth_c|} \quad (5.6)$$

Ortalama IoU (mIoU), tüm anlamsal sınıflar üzerinden ortalama alınarak elde edilir:

$$mIoU = \frac{1}{N} \sum_{c=1}^N IoU_c \quad (5.7)$$

Burada N , sınıf sayısını temsil eder. En İyi IoU, doğrulama sürecinde elde edilen en yüksek mIoU skorunu ifade eder ve modelin tüm sınıflar arasında aşırı ve yetersiz segmentasyon dengesini

en iyi sağladığı durumu gösterir. Demiryolu bağlamında, özellikle ray, peron kenarı ve bitki örtüsü istilasası gibi kritik sınıflar için sınıf başına IoU analizine ayrıca önem verilir.

IoU, segmentasyon görevlerinde en anlaşılır ve yaygın olarak benimsenmiş kıyaslama metriğidir. Mekânsal hizalama ve sınıf sınırlarının doğruluğu gibi, demiryolu algılama sistemlerinde hayati öneme sahip özelliklere karşı oldukça hassastır.

Piksel Düzeyinde Doğruluk – En İyi Doğruluk (Best Accuracy)

Piksel düzeyinde doğruluk, doğru sınıflandırılan piksellerin toplam piksel sayısına oranını ölçer ve şu şekilde hesaplanır:

$$Pixel Accuracy = \frac{\sum_{c=1}^N TP_c}{\sum_{c=1}^N (TP_c + FP_c + FN_c)} \quad (5.8)$$

Burada TP_c , ilgili sınıf için doğru pozitiflerin sayısını; FP_c ve FN_c ise sırasıyla yanlış pozitif ve yanlış negatif örneklerin sayısını ifade etmektedir. Söz konusu metrik, modelin genel doğruluğu hakkında genel bir fikir sunmakla birlikte, sınıf dağılımının dengesiz olduğu durumlarda yanıltıcı sonuçlara yol açabilir. Özellikle arka plan veya balast gibi belirli sınıfların piksel sayılarının, diğer sınıflara kıyasla belirgin biçimde fazla olduğu demiryolu sahnelerinde bu durum daha belirgin hale gelmektedir.

En İyi Doğruluk, doğrulama sürecinde elde edilen en yüksek piksel doğruluğunu ifade eder. Küçük veya nadir görülen sınıflar açısından IoU kadar ayırt edici olmasa da, hızlı tutarlılık kontrolleri ve kaba performans karşılaştırmaları için faydalıdır.

Kayıp Minimizasyonu – En İyi Kayıp (Best Loss)

Segmentasyon modellerinin eğitim amacı, veri setinin yapısına ve sınıf dağılımına bağlı olarak kategorik çapraz entropi (categorical cross-entropy), Dice kaybı (Dice loss) veya focal loss gibi piksel düzeyinde kayıp fonksiyonları içerebilir. Bu çalışmada, çapraz entropi kaybı kullanılmıştır:

$$\mathcal{L}_{CE} = - \sum_{i=1}^N y_i \log(\hat{y}_i) \quad (5.9)$$

Burada y_i , piksel i için gerçek sınıf etiketini; $\hat{y}_i$ ise tahmin edilen olasılığı temsil eder. En İyi Kayıp, eğitim süresince doğrulama verisi üzerinde gözlemlenen en düşük kaybı ifade eder. Bu değer, modelin yakınsama (convergence) ve genelleme (generalization) kabiliyetini gösterir.

Ancak, özellikle dengesiz veri kümelerinde, kayıp tek başına IoU veya piksel doğruluğu gibi değerlendirme metrikleriyle her zaman doğrudan ilişkili olmayabilir. Bu nedenle, En İyi Kayıp, En İyi IoU ve En İyi Doğruluk ile birlikte, nihai model kontrol noktalarının seçilmesi ve performans dengelerinin değerlendirilmesinde kullanılır.

5.3.3. Derinlik Tahmini ve Mesafe Ölçüm Metrikleri (ZED2 Kamera)

Doğru derinlik tahmini, özellikle engel konumlandırma, açıklık değerlendirmesi ve güvenli yol tahmini gibi görevlerde, otonom ve yarı otonom demiryolu sistemlerinde mekânsal farkındalığın sağlanması için kritik öneme sahiptir. Bu çalışmada, derinlik bilgisi, stereo eşleştirme algoritmaları aracılığıyla iki gözlü (binoküler) görüntü çiftlerinden üç boyutlu sahne geometrisini yeniden yapılandıran ZED2 stereo kamera kullanılarak elde edilmiştir. Derinlik tahmin hattının performansını titizlikle değerlendirmek için, farklı çalışma koşulları ve derinlik aralıklarında hem mutlak hem de göreceli hata özelliklerini yakalayan, iyi bilinen nicel metrikler seti kullanılmıştır.

Demiryolu bağlamında derinlik tahmini değerlendirmesi, yalnızca tek tek piksel bazında derinlik tahminlerinin doğruluğunu değil, aynı zamanda sistemin farklı çevresel senaryolardaki (örneğin açık hat, tünel, istasyon peronu) ve geniş mesafe aralıklarındaki dayanıklılığını da ele almalıdır. Aşağıdaki alt bölümlerde, bu çalışmada benimsenen metrikler detaylandırılmıştır.

Ortalama Mutlak Hata (MAE) ve Karekök Ortalama Kare Hata (RMSE)

Ortalama Mutlak Hata (Mean Absolute Error - MAE) ve Karekök Ortalama Kare Hata (Root Mean Square Error - RMSE), tahmin edilen derinlik değerleri ile karşılık gelen gerçek derinlik ölçümleri arasındaki mutlak sapmayı nicelendirmenin standart metrikleridir. Piksel i için tahmin edilen derinlik $\hat{D}_i$, gerçek derinlik ise D_i olarak gösterildiğinde metrikler şu şekilde tanımlanır:

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{D}_i - D_i| \quad (5.10)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{D}_i - D_i)^2} \quad (5.11)$$

MAE, ortalama tahmin hatasının yorumlanabilir bir ölçüsünü sunarken; RMSE, özellikle kritik mesafe aralıklarında başarısızlık durumlarını tespit etmek için büyük hatalara daha fazla ağırlık verir.

Ortalama Mutlak Göreceli Hata (MARE)

Mutlak metrikler doğrudan hata niceliği için faydalı olsa da, stereo görüş sistemlerinin mesafe arttıkça derinlik tahminindeki belirsizliğin artması gibi bilinen bir sınırlamayı hesaba katmazlar. Bu nedenle, Ortalama Mutlak Göreceli Hata (Mean Absolute Relative Error - MARE) de hesaplanmıştır:

$$MARE = \frac{1}{N} \sum_{i=1}^N \left| \frac{\hat{D}_i - D_i}{D_i} \right| \quad (5.12)$$

Bu metrik, tahmin edilen derinliklerin gerçek değerlerden göreceli sapmasını değerlendirir ve özellikle demiryolu uygulamalarında, erken karar alma için 30 metrenin üzerindeki sinyal veya engel tespiti gibi uzak mesafe derinlik tahmininin doğruluğu açısından önemlidir.

Mesafe Bazlı Hata Ayırıştırması

Pratik demiryolu ortamlarında, derinlik doğruluğunun operasyonel önemi mesafeye bağlı olarak önemli ölçüde değişir. Örneğin, yakın mesafe tespiti (<10 m) çarpışma önleme için kritikken, orta (10–30 m) ve uzun mesafe (>30 m) tahminleri sinyal tespiti ve öngörücü planlama için önemlidir. Bu amaçla, derinlik tahmin performansı üç mesafe aralığına göre ayrıştırılmıştır:

- Yakın mesafe: 0–10 metre
- Orta mesafe: 10–30 metre
- Uzak mesafe: >30 metre

Her aralık için MAE ve RMSE doğruluğu raporlanır. Bu ayrıştırılmış değerlendirme, derinlik arttıkça performanstaki azalmayı ortaya koyar ve operasyonel koşullar altında sistem sınırlamalarını aydınlatır.

3B Konumlandırma Hatası (Localization Error)

Piksel bazlı derinlik değerlendirmesinin ötesinde, tespit edilen nesnelerin üç boyutlu konumlandırmasının doğruluğu da incelenir. Bir nesnenin görüntü koordinatlarındaki merkezi ve karşılık gelen derinlik harci kullanılarak stereo üçgenleme ile nesnenin 3B uzaydaki konumu hesaplanır. Tahmin edilen ve gerçek 3B koordinatlar $X_{pred} = (x_p, y_p, z_p)$ ile $X_{gt} = (x_g, y_g, z_g)$ arasındaki Öklidyen mesafe, 3B konumlandırma hatası olarak hesaplanır:

$$Localization Error = ||X_{pred} - X_{gt}||_2 \quad (5.13)$$

Bu metrik, engeller veya demiryolu altyapısı gibi nesnelerin doğru mekânsal yerleştirilmesinin karar alma ve güvenlik sınırları üzerinde etkili olduğu uygulamalarda hayati öneme sahiptir.

Bu metrikler dizisi, ZED2 kameranın demiryolu bağlamındaki derinlik tahmin yeteneklerinin kapsamlı ve çok yönlü bir değerlendirmesini sunar. Mutlak ve görelî hata ölçümlerinin, ayrıştırılmış performans analizinin ve mekânsal konumlandırma metriklerinin kombinasyonu sayesinde, akademik standartlarla ve gerçek dünya operasyonel gereksinimleriyle uyumlu titiz ve uygulamaya yönelik bir değerlendirme sağlanmıştır.

Bu kapsamlı değerlendirme çerçevesi, geliştirilen algı sisteminin titiz, yorumlanabilir ve güvenlik odaklı bir analizine imkân tanımaktadır. Değerlendirme sürecinde kullanılan metrikler, hem akademik literatürde kabul görmüş standartları hem de gerçek dünya demiryolu uygulamalarının pratik ve güvenlik odaklı gereksinimlerini yansıtacak biçimde özenle seçilmiştir.

Bu çift yönlü yaklaşım, sistemin yalnızca kuramsal olarak sağlam değil, aynı zamanda sahadaki operasyonel kısıtlar ve güvenlik açısından kritik koşullarla da uyumlu olmasını garanti altına almaktadır.



6. BULGULAR VE TARTIŞMA

Bu bölüm, akıllı demiryolu algısı için önerilen çok görevli derin öğrenme çerçevesinin değerlendirilmesiyle elde edilen deneysel sonuçların kapsamlı bir analizini sunmaktadır. Analiz, sistemin temel bileşenleri etrafında yapılandırılmıştır: nesne tespiti, anlamsal segmentasyon, çok görevli füzyon ve stereo tabanlı mesafe ölçümü.

Bölüm 6.1, çeşitli nesne tespiti modellerinin standart değerlendirme metrikleri ve gerçek dünya senaryolarındaki performansını tartışmaktadır. Bölüm 6.2, raylı altyapı bileşenlerinin ayrıntılı bir şekilde yakalanmasında kullanılan anlamsal segmentasyon mimarilerinin etkinliğini incelemektedir. Bölüm 6.3, doğruluk, verimlilik ve sahne bütünlüğü açısından iki farklı çok görevli füzyon stratejisinin değerlendirmesini sunmaktadır. Bölüm 6.4 ise, ZED2 tabanlı derinlik algılamanın entegrasyonunu ve bu bileşenin mekânsal kavrayışa katkısını araştırmaktadır.

Son olarak, Bölüm 6.5 bu bulguları eleştirel bir tartışma çerçevesinde sentezlemekte; temel eğilimleri, ödünleşimleri (trade-off) ve sınırlılıkları belirlemekte, ayrıca elde edilen sonuçların demiryolu ortamlarında gerçek dünya uygulamaları açısından taşıdığı daha geniş etkiler üzerine değerlendirmelerde bulunmaktadır. Bu analizlerin bütünlüğü, sistemin operasyonel kısıtlar altında doğru, gerçek zamanlı ve derinlik farkındalığına sahip sahne yorumlama kapasitesini doğrulamaktadır.

6.1. Nesne Tespiti Sonuçları

Bu bölümde, demiryolu ortam algısı bağlamında çeşitli son teknoloji derin öğrenme mimarileri kullanılarak gerçekleştirilen nesne tespiti performansına ilişkin kapsamlı bir değerlendirme sunulmaktadır. Çalışmanın temel amacı, gecikme, doğruluk ve hesaplama verimliliği gibi kritik kısıtların bulunduğu demiryolu izleme sistemlerinde, söz konusu modellerin gerçek zamanlı uygulamalara uygunluğunu analiz etmektir. Bu doğrultuda YOLOv8, YOLOv9, YOLOv11, YOLOv12 ve Faster R-CNN olmak üzere beş farklı nesne tespiti modeli uygulanmış; bu modeller, hedeflenen operasyonel alana uygunluğu nedeniyle tercih edilen RailSem19 veri seti üzerinde eğitilmiştir.

Bu modeller, mimari karmaşıklık, çıkarım hızı ve konumlandırma doğruluğu arasında farklı dengeleri temsil etmeleri amacıyla seçilmiştir. Tüm eğitim deneyleri, yöntemsel tutarlılığı sağlamak ve adil karşılaştırma yapabilmek için tek tip ve kontrollü koşullarda gerçekleştirilmiştir. Tespit görevi, demiryolu operasyonları açısından güvenlik açısından kritik beş nesne kategorisine odaklanmıştır: raylar, araçlar, trafik işaretleri, yayalar ve engeller. Sınır kutusu anotasyonları, Bölüm 4'te açıklanan sahte etiketleme (pseudo-labeling) süreci ile piksel düzeyindeki segmentasyon maskelerinden türetilmiştir.

YOLO sürümleri (v8, v9, v11, v12), öğrenme oranı ayarlaması için kosinüs azalan zamanlamayla birlikte stokastik gradyan inişi (SGD) kullanılarak Ultralytics çerçevesinde 50 epok (epoch) boyunca eğitilmiştir. Buna karşılık, Faster R-CNN, Detectron2 çerçevesinde adım bazlı bozunma zamanlaması (stepwise schedule) kullanan AdamW optimize edici ile 10 epok boyunca eğitilmiştir. Tüm modeller 640×640 piksel giriş çözünürlüğüyle eğitilmiş olup, toplu işlem boyutları (batch size) kullanılabilir GPU belleğine (NVIDIA RTX 2070) göre dinamik olarak ayarlanmıştır.

Değerlendirme metrikleri arasında, 0.5 IoU eşik değerinde ortalama kesinlik ($mAP@0.5$), 0.5 ile 0.95 arası IoU eşiklerinde ortalama kesinlik ($mAP@0.5:0.95$), kesinlik (precision), duyarlılık (recall) ve F1 skoru yer almaktadır. Tablo 6.1’de özetlendiği üzere, YOLOv11 genel olarak en iyi performansı göstermiş; 0.50412 $mAP@0.5$ ve 0.32043 $mAP@0.5:0.95$ değerlerine ulaşmıştır. Model ayrıca 0.63409 kesinlik, 0.46402 duyarlılık ve 0.53672 F1 skoru elde etmiştir. YOLOv9 her ne kadar en yüksek $mAP@0.5$ (0.52141) değerini raporlamış olsa da, düşük F1 skoru ve nesne kategorileri arasında tutarsız tespit performansı, onu yüksek güvenilirlik gerektiren gerçek zamanlı uygulamalar için daha az uygun hale getirmiştir.

Tablo 6.1. Nesne Tespiti Sonuçları

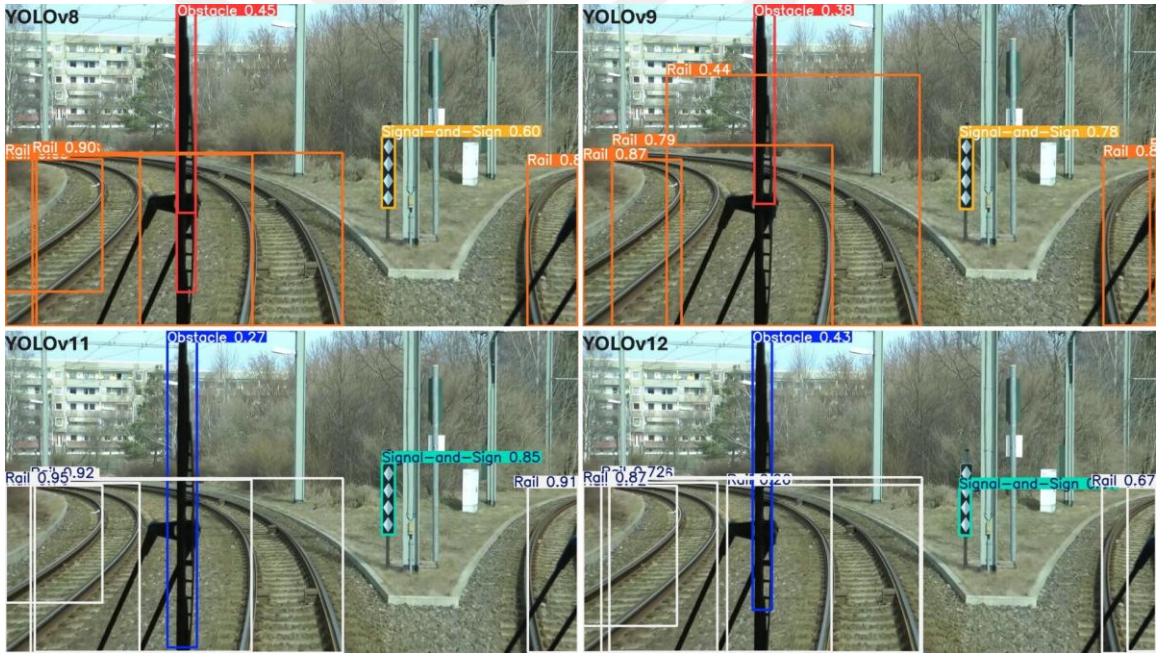
Model	Kesinlik (Precision)	Duyarlılık (Recall)	F1-Skoru	mAP50	mAP50- 95	Eğitim Süresi
YOLOv12	0.49300	0.28890	0.36500	0.36500	0.22000	3S 24d
YOLOv11	0.63409	0.46402	0.53672	0.50412	0.32043	2S 52d
YOLOv9	0.68196	0.46181	0.55017	0.52141	0.33722	2S 41d
YOLOv8	0.67661	0.45112	0.54002	0.49569	0.31709	3S 14d
FasterRCNN	N/A	N/A	N/A	0.03950	0.01760	2S 45d

YOLOv8, dengeli performansı sayesinde kaynak kısıtlı ortamlar için güçlü bir aday olarak öne çıkmıştır. Model, 0.49569 $mAP@0.5$, 0.31709 $mAP@0.5:0.95$ ve 0.54002 F1 skoru ile doğruluk ve hesaplama verimliliği arasında etkili bir denge sağlamıştır. Buna karşın, YOLOv12 düşük gecikmeli görevler için optimize edilmiş olmasına rağmen tespit doğruluğu ve duyarlılık açısından yetersiz kalmış, bu da karmaşık görsel koşullar altında sağlam tespit kabiliyetinden ödün verildiğini göstermektedir. Faster R-CNN, sınırlı karmaşıklığa sahip sahnelerde yüksek kesinlik sergilemesine rağmen, küçük veya kısmen örtülmüş nesneleri tespitte büyük zorluk yaşamış ve farklı demiryolu görüntülerine genellemede başarısız olarak yalnızca 0.0395 $mAP@0.5$ değeri elde etmiştir.

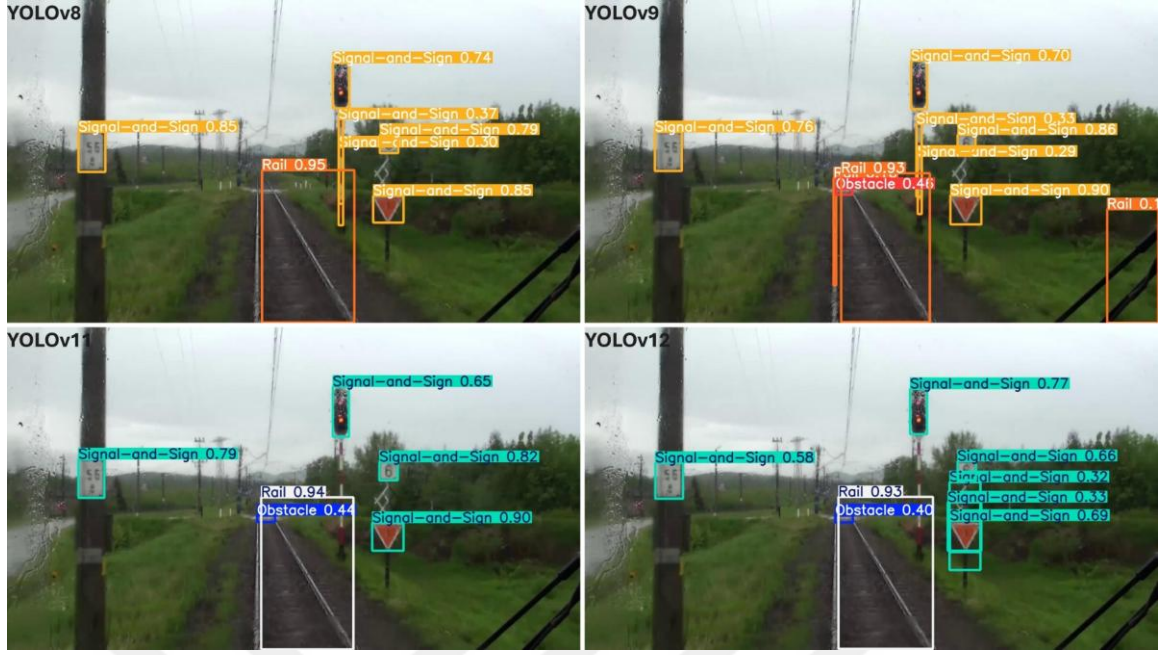
Model davranışlarını daha ayrıntılı biçimde değerlendirmek amacıyla kapsamlı tanılama (diagnostic) analizleri gerçekleştirilmiştir. Eğitim sürecine ait günlük kayıtları ve değerlendirme

çıktıları, Ek 2'de sunulmakta olup, bu kapsamda kesinlik-duyarlılık (precision-recall) eğrileri, F1 skoru zaman serileri, etiket korelasyon matrisleri ve karışıklık (confusion) matrisleri yer almaktadır (Şekil 7.1 ile Şekil 7.16 arasında sunulan görseller). Bu analizler, özellikle veri kümesinde düşük örnekleme sıklığına sahip sınıfların (örneğin yayalar ve ray kenarı engelleri) tespitinde, YOLOv11 modelinin diğer mimarilere kıyasla tutarlı biçimde üstün performans sergilediğini ortaya koymuştur.

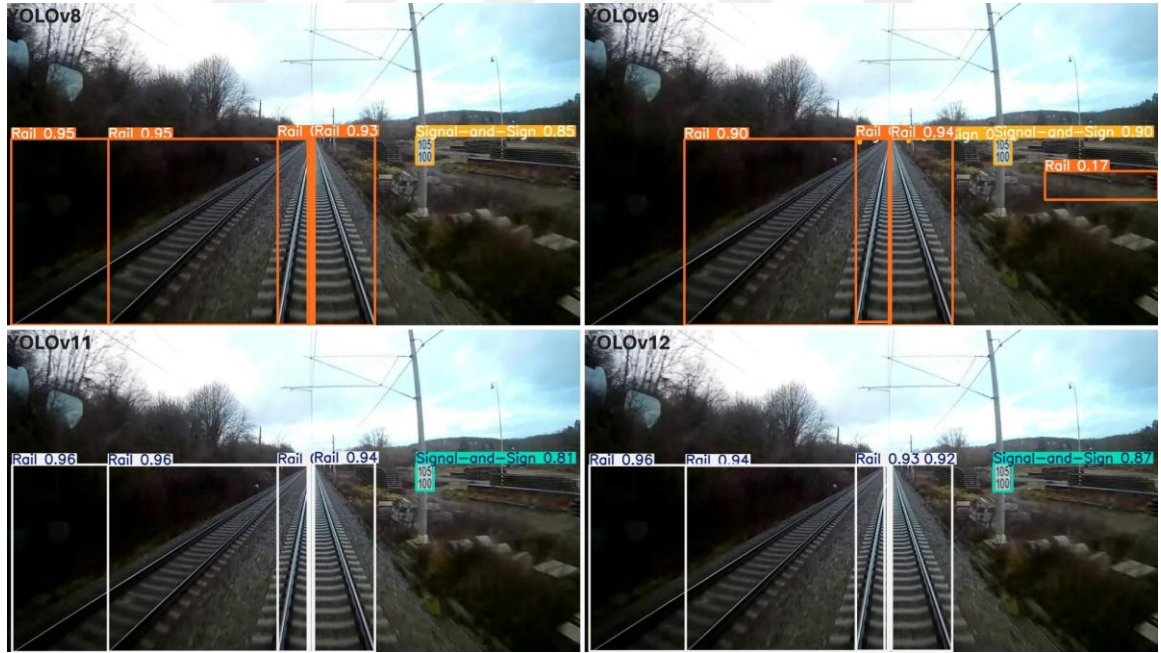
Ayrıca, çıkarım çıktılarının görsel incelemesi (Şekil 6.1 ile Şekil 6.4 arasındaki şekiller), YOLOv11 ve YOLOv8 modellerinin hareket bulanıklığı, arka plan karmaşası ve değişken aydınlatma gibi görsel zorlukların bulunduğu senaryolarda dahi yüksek düzeyde yapısal sağlamlık (robustness) sergilediğini desteklemektedir. Bu bulgular, yalnızca sayısal metriklerle değil, aynı zamanda nitel sahne analiziyle de modellerin genelleme ve dayanıklılık yeteneklerini ortaya koymakta; özellikle güvenlik-kritik demiryolu uygulamaları için algoritmik tercihlerin gerekçelendirilmesine katkı sunmaktadır.



Şekil 6.1. Nesne Tespiti Modelleri Çıktı Örnekleri - 1



Şekil 6.2. Nesne Tespiti Modelleri Çıktı Örnekleri - 2



Şekil 6.3. Nesne Tespiti Modelleri Çıktı Örnekleri - 3



Şekil 6.4. Nesne Tespiti Modelleri Çıktı Örnekleri - 4

Gerçek zamanlı çıkarım kabiliyeti de kritik biçimde incelenmiştir. Gerçekleştirilen kıyaslama sonuçları, YOLOv11 ve YOLOv8'in hedef donanım (NVIDIA RTX 2070) üzerinde saniyede 35 (FPS) kareyi aşan çıkarım hızlarını sürdürebildiğini ve dolayısıyla gerçek zamanlı demiryolu izleme uygulamalarının operasyonel gerekliliklerini karşıladığını göstermiştir. Buna karşılık, Faster R-CNN ortalama 10 FPS'nin altında bir performans sergilemiş ve gecikmeye duyarlı uygulamalar için pratik olmaktan uzak kalmıştır. YOLOv12 her ne kadar daha yüksek FPS değerlerine ulaşmış olsa da, yetersiz tespit doğruluğu, onu yoğun nesne dağılımına veya düzensiz geometrilere sahip karmaşık sahnelerde sınırlı bir kullanım alanına indirgemıştır.

Sonuç olarak, deneysel bulgular, tek aşamalı tespit mimarileri arasında özellikle YOLOv11 ve YOLOv8 modellerinin, gerçek zamanlı demiryolu algılama sistemlerinde tespit doğruluğu, model sağlamlığı ve hesaplama verimliliği açısından en uygun dengeyi sunduğunu ortaya koymaktadır. YOLOv11 modeli, Bölüm 4.3'te ele alınan anlamsal segmentasyon görevleriyle entegrasyon bakımından en umut vadeden seçenek olarak öne çıkmaktadır. Elde edilen bu sonuçlar, YOLO ailesinin gelişiminin sürdüğünü ve demiryolu altyapısı izleme gibi sınırlı ve güvenlik açısından kritik uygulama alanları için uygunluğunu koruduğunu göstermektedir.

6.2. Anlamsal Segmentasyon Sonuçları

Bu bölümde, önerilen çok görevli sahne anlayış sisteminin anlamsal segmentasyon bileşenine ilişkin deneysel bulgular ayrıntılı bir şekilde sunulmaktadır. Bu bağlamda amaç, çeşitli derin öğrenme mimarilerinin karmaşık demiryolu ortamlarında sınıf düzeyindeki ayırım

kapasitesini ve genelleme yeteneklerini hem nicel hem de nitel açıdan değerlendirmektedir. Bu değerlendirme kapsamında, UNet-Vanilla, önceden eğitilmiş ResNet34 tabanlı UNet (UNet-Pretrained), ResNet50 kodlayıcılı DeepLabV3+, Özellik Piramit Ağı (Feature Pyramid Network – FPN) ve Piramit Sahne Ayırıştırma Ağı (Pyramid Scene Parsing Network – PSPNet) olmak üzere beş farklı mimari yapı, RailSem19 veri seti üzerinde 40 epok süresince eğitilmiştir. Modeller, eğitim, doğrulama ve test olmak üzere üçlü veri bölünmesi temel alınarak kapsamlı şekilde değerlendirilmiştir.

Her bir modelin başarımı, çapraz entropi kaybı (Loss), piksel düzeyinde doğruluk (Accuracy) ve sınıf başına ortalama Intersection over Union (IoU) olmak üzere üç temel metrik üzerinden incelenmiştir. Tablo 6.2, Tablo 6.3 ve Tablo 6.4'te sırasıyla eğitim, doğrulama ve test kümesi sonuçları özetlenmiştir.

Eğitim kümesi üzerinde en yüksek IoU değeri, %68.3 ile FPN modeli tarafından elde edilmiştir. FPN'nin çok ölçekli üstten alta öznitelik birleştirme yapısı, özellikle ince yapılı raylar, sinyal direkleri ve peron kenarlarının doğru şekilde ayrıştırılmasında etkili olmuş; aynı zamanda %91.9 doğruluk oranı ile sahne içi yapısal tutarlılığı güçlü biçimde modelleyebilmiştir. Buna karşılık, doğrulama ve test kümelerinde FPN'nin IoU değerlerinin sırasıyla %50.6 ve %50.6 olarak daha sınırlı kalması, bu modelin veri çeşitliliği karşısında belirli bir aşırı öğrenme eğilimi sergilediğini göstermektedir.

UNet-Vanilla modeli, doğrulama kümesinde %51.2 ve test kümesinde %50.8 IoU değeri ile en yüksek genelleme başarısını göstermiştir. Bu başarı, simetrik encoder-decoder mimarisi ve atlama bağlantılar (skip connections) sayesinde sağlanan yapısal bütünlük ile açıklanabilir. Modelin önceden eğitilmiş bir kodlayıcı içermemesi, veri dağılımındaki dengesizliklere karşı daha esnek bir öğrenme süreci yürütmesine katkı sağlamıştır.

DeepLabV3+ modeli, eğitim kümesinde %65.3 IoU ve %91.2 doğruluk ile güçlü bir yakınsama profili sergilemiş; ancak doğrulama ve test kümelerinde sırasıyla %50.2 ve %49.8 IoU değerleriyle orta düzeyde bir genel performans ortaya koymuştur. Özellikle Atrous Spatial Pyramid Pooling (ASPP) modülünün sağladığı çoklu bağlam yakalama yeteneği, geniş yapıli sahnelerde avantaj sağlamış; ancak dar ve örtölmüş nesnelerde ayırım gücü sınırlı kalmıştır.

UNet-Pretrained, ImageNet üzerinde önceden eğitilmiş ResNet34 kodlayıcısının katkısıyla eğitim aşamasında hızlı yakınsama göstermiş ve %61.4 IoU ve %90.4 doğruluk oranı elde etmiştir. Doğrulama ve test kümelerinde sırasıyla %50.0 ve %50.0 IoU değerleri ile dengeli fakat sınırda performans sergilemiştir. Modelin eğitim süresinin yalnızca 69 dakika olması, hesaplama verimliliği açısından önemli bir avantajdır.

PSPNet, üç veri kümesi üzerinde de en düşük segmentasyon başarımını göstermiştir. Eğitim, doğrulama ve test kümelerinde sırasıyla %59.5, %47.5 ve %47.8 IoU değerleri elde eden model, özellikle az örneklemlili sınıflarda yeterli ayırım gücü sunamamıştır. Piramit havuzlama

mekanizması genel bağlamsal farkındalık sağlasa da, sınıf içi karışmalar ve genelleme zorlukları, modelin başarımını sınırlamıştır.

Tablo 6.2. Eğitim (Training) Anlamsal Segmentasyon Sonuçları

Model	En İyi Kayıp (Loss)	En İyi Doğruluk (Accuracy)	En İyi IoU (IoU)
UNet-Vanilla	0.255	0.912	0.648
UNet- Pretrained	0.281	0.904	0.614
DeepLabV3+	0.245	0.912	0.653
FPN	0.221	0.919	0.683
PSPNet	0.318	0.885	0.595

Tablo 6.3. Doğrulama (Validation) Anlamsal Segmentasyon Sonuçları

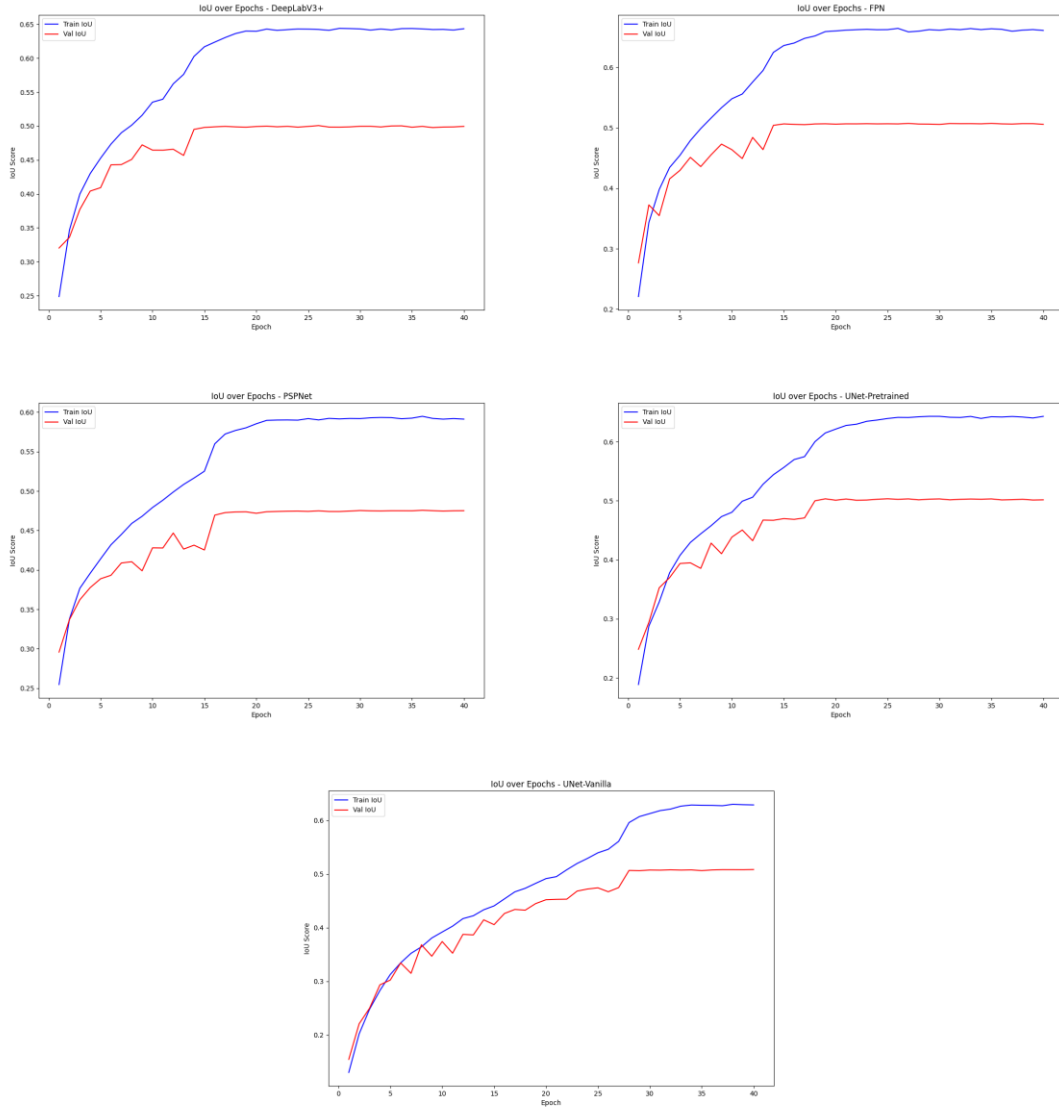
Model	En İyi Kayıp (Loss)	En İyi Doğruluk (Accuracy)	En İyi IoU (IoU)
UNet-Vanilla	0.465	0.850	0.512
UNet- Pretrained	0.481	0.847	0.500
DeepLabV3+	0.474	0.846	0.502
FPN	0.501	0.845	0.506
PSPNet	0.499	0.833	0.475

Tablo 6.4. Test Anlamsal Segmentasyon Sonuçları

Model	En İyi Kayıp (Loss)	En İyi Doğruluk (Accuracy)	En İyi IoU (IoU)
UNet-Vanilla	0.4794	0.8471	0.5086
UNet- Pretrained	0.5098	0.8444	0.5002
DeepLabV3+	0.4973	0.8427	0.4985
FPN	0.5232	0.8426	0.5054
PSPNet	0.5108	0.8319	0.4780

Tüm modeller, 40 epok boyunca kararlı bir şekilde yakınsamış; eğitim ve doğrulama kayıpları sistematik olarak azalmış, IoU (Intersection over Union) değerleri ise genel olarak artış

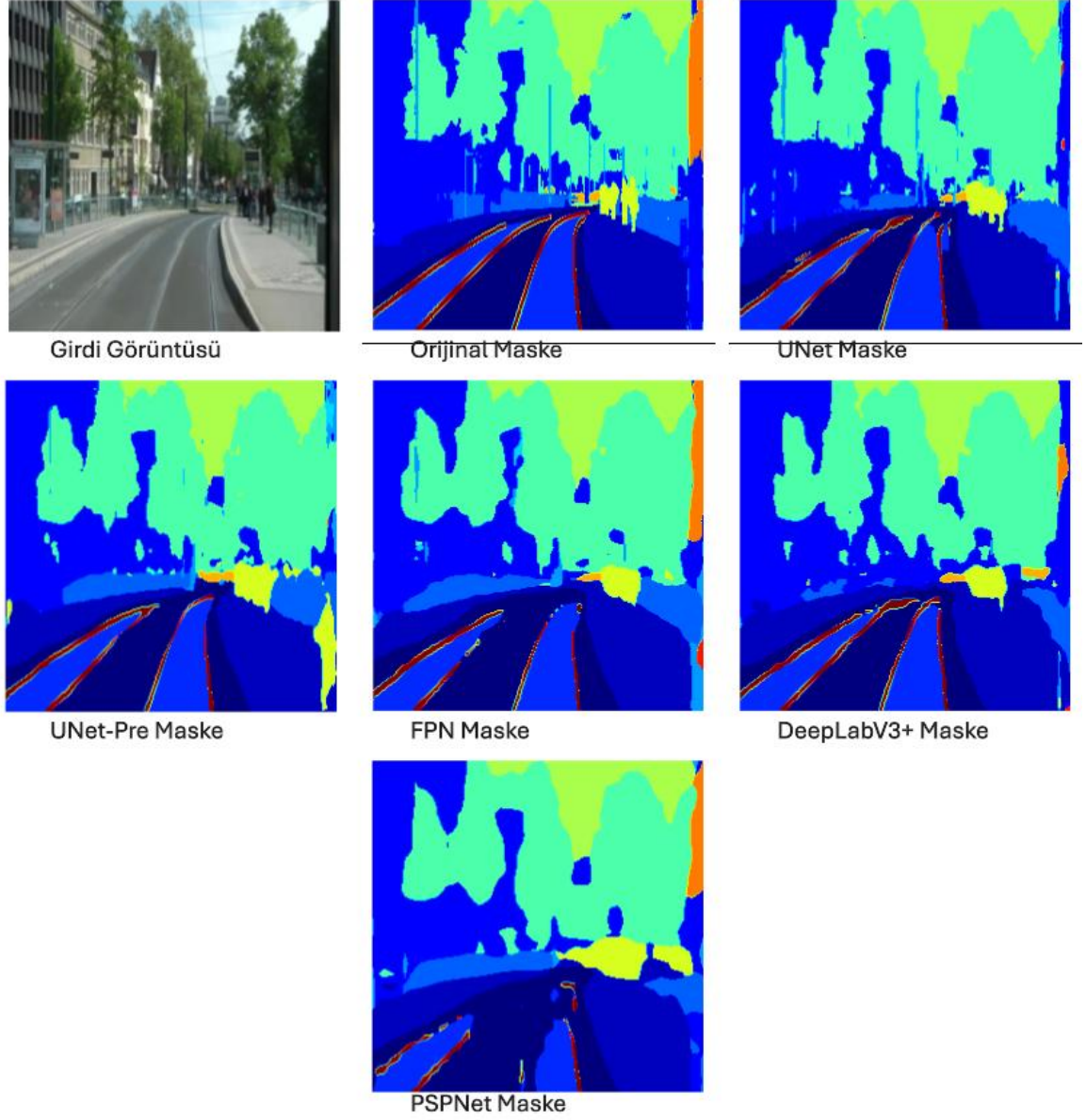
göstermiştir. Bu öğrenme dinamikleri, Şekil 6.5'te sunulan birleşik IoU eğrileri aracılığıyla görsel olarak desteklenmektedir. Her bir modelin eğitim ve doğrulama IoU değerleri aynı eksen üzerinde sunulmuş, böylece modellerin genel performans eğilimleri doğrudan karşılaştırılabilir hale getirilmiştir. Eğitim süreleri incelendiğinde, UNet-Pretrained ve FPN modelleri 69 dakikalık süreleriyle en verimli mimariler olurken, UNet-Vanilla modeli 107 dakikalık süresiyle en uzun eğitimi gerektirmiştir. Bu bulgu, mimari karmaşıklığın yalnızca doğruluk üzerinde değil, aynı zamanda hesaplama maliyeti üzerinde de belirleyici bir faktör olduğunu göstermektedir.



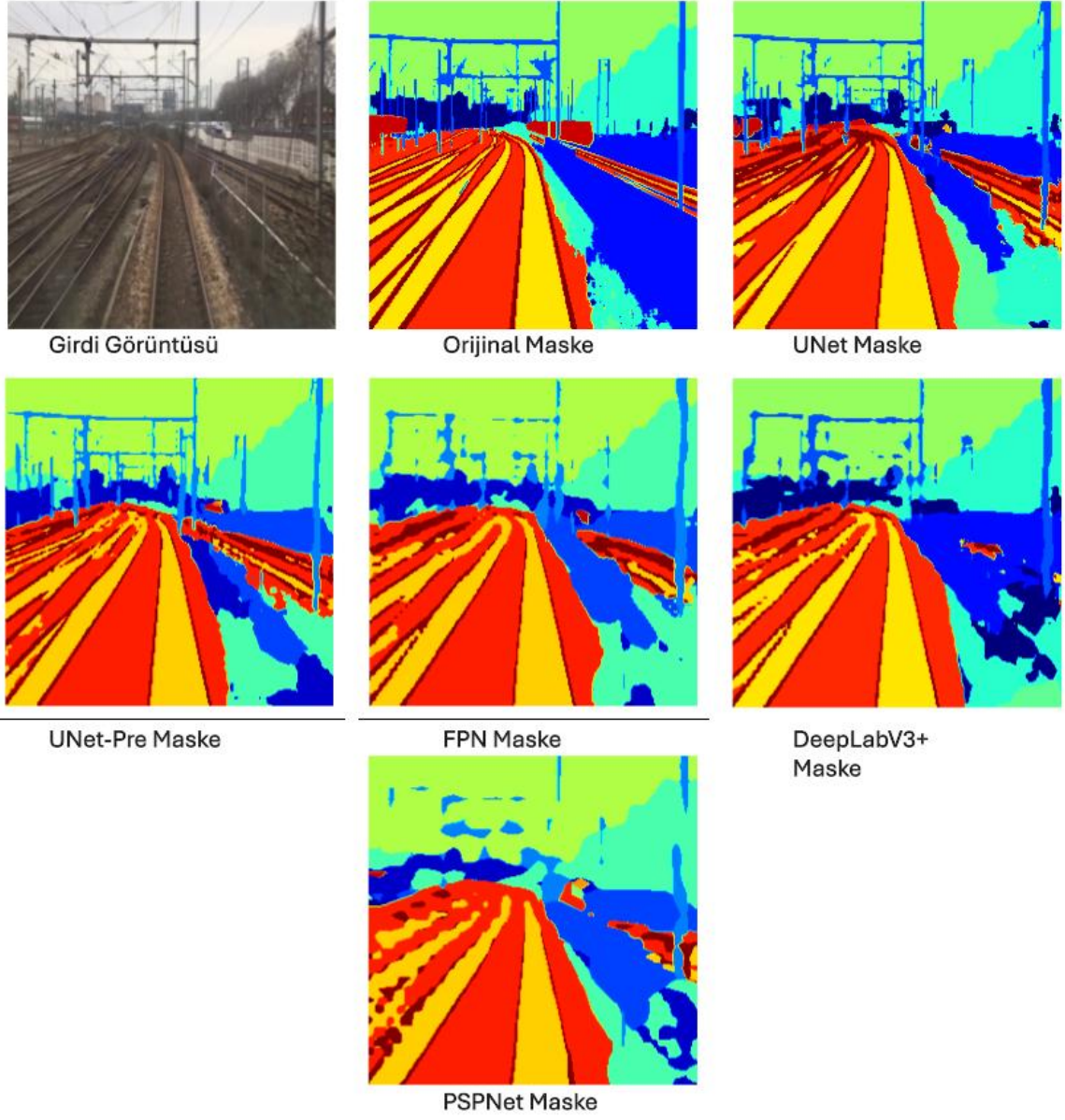
Şekil 6.5. Beş Segmentasyon Modeli İçin Eğitim ve Doğrulama IoU Eğrileri

Niteliksel değerlendirme, test kümesinden seçilen temsilci örneklerde tahmin maskelerinin gerçek görüntüler üzerine bindirilmesi yoluyla gerçekleştirilmiştir. Şekil 6.6 ile Şekil 6.9 arasındaki

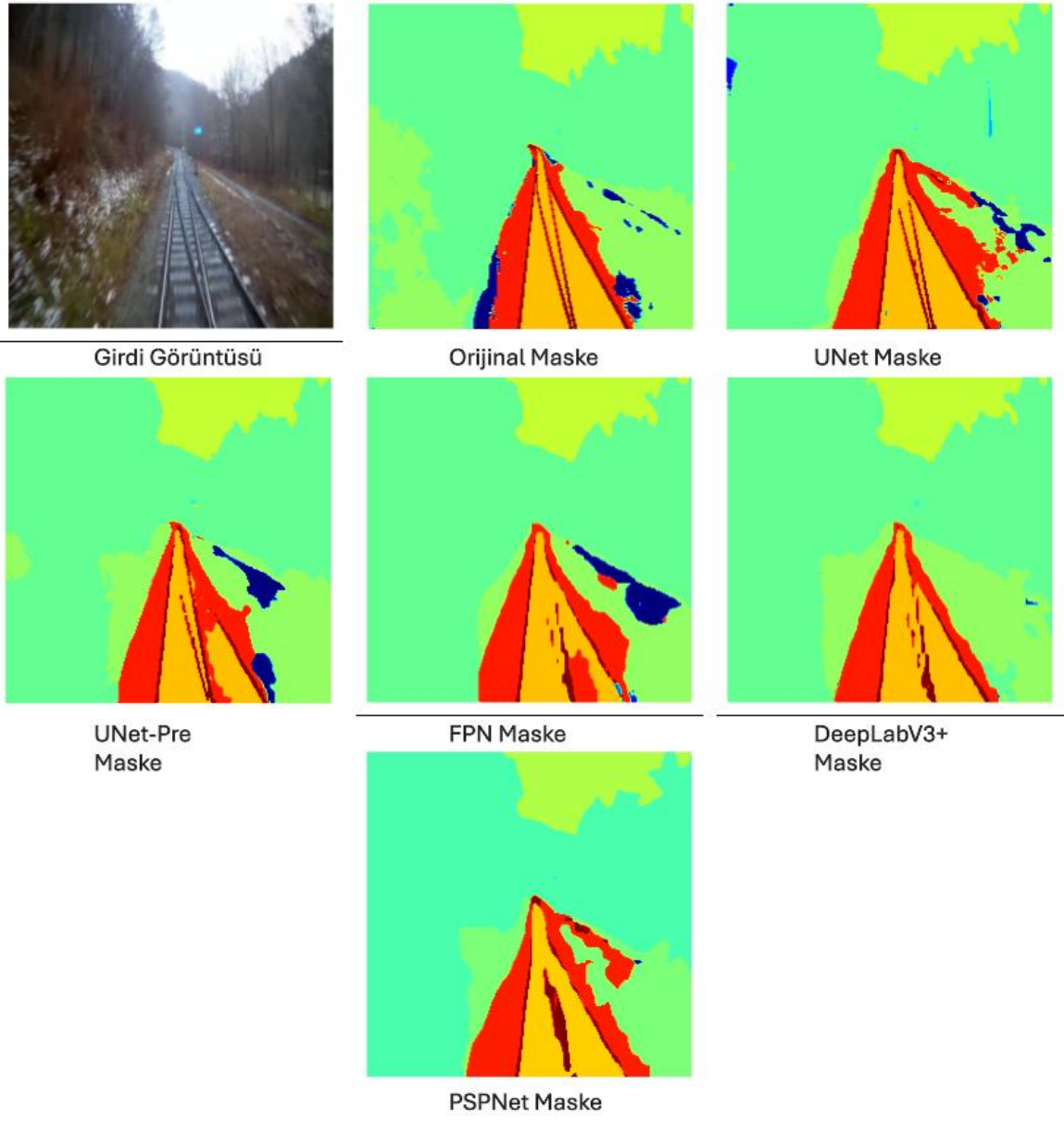
şekillerde görüldüğü üzere, FPN ve UNet-Pretrained modelleri; raylar, peron kenarları, işaret levhaları ve yayalar gibi sahne öğeleri üzerinde anlamlı, keskin ve yapısal olarak tutarlı segmentasyon çıktıları üretmiştir. Bu durum, modellerin gerçek zamanlı izleme sistemlerinde kullanılabilirliğini desteklemektedir.



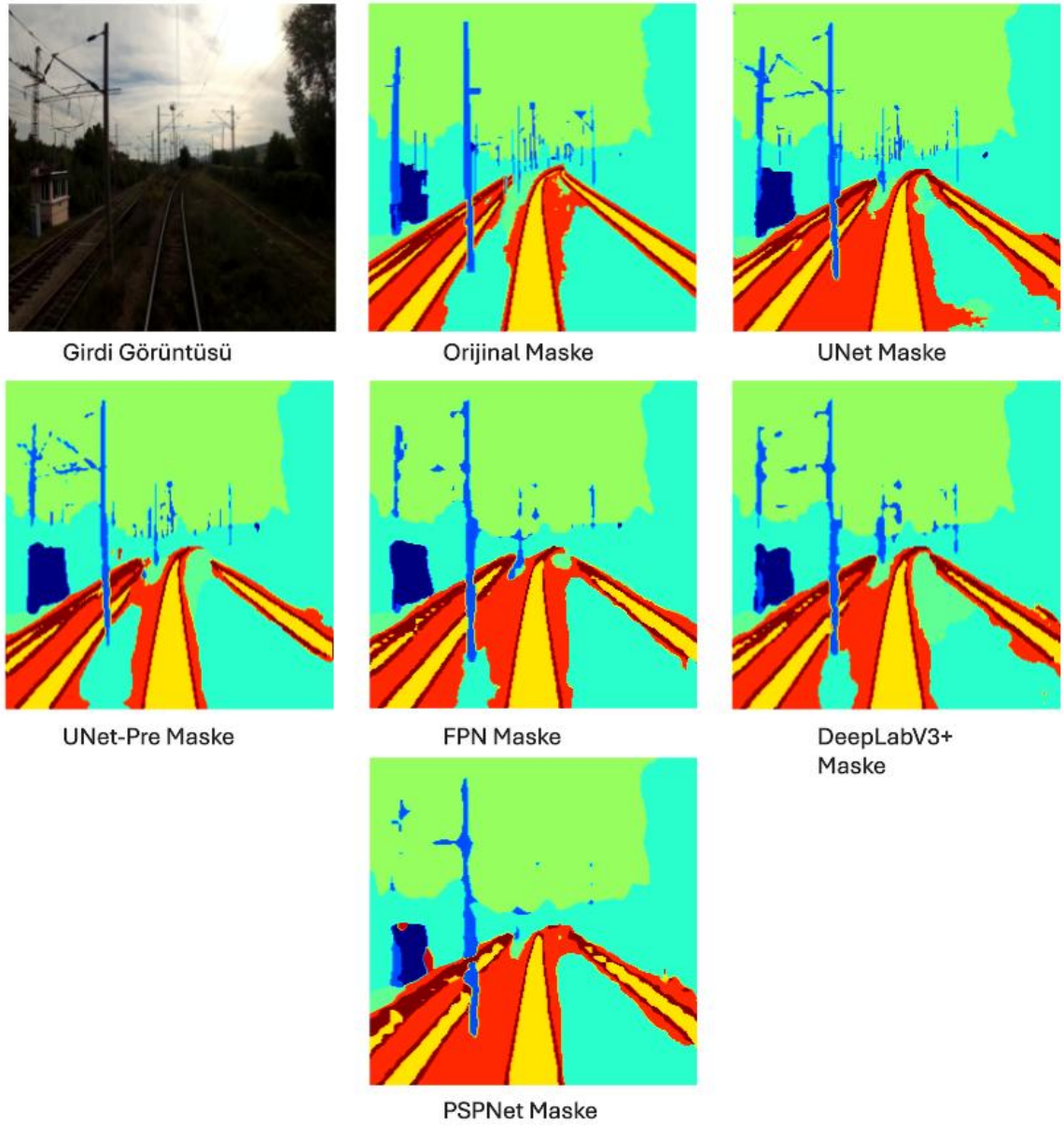
Şekil 6.6. Anlamsal Segmentasyon Modelleri Çıktı Örnekleri - 1



Şekil 6.7. Anlamsal Segmentasyon Modelleri Çıktı Örnekleri - 2



Şekil 6.8. Anlamsal Segmentasyon Modelleri Çıktı Örnekleri - 3



Şekil 6.9. Anlamsal Segmentasyon Modelleri Çıktı Örnekleri - 4

Sonuç olarak, segmentasyon görevleri bağlamında FPN ve UNet-Pretrained mimarileri, doğruluk, genelleme yeteneği ve hesaplama verimliliği açısından dengeli bir performans profili sunmaktadır. Bu mimariler, özellikle güvenlik açısından kritik olan uygulamalarda, zamansal tutarlılığı koruyabilme ve gerçek zamanlı çıkarım gereksinimlerini karşılayabilme yetenekleri sayesinde, çok görevli (multi-task) öğrenme sistemlerine entegrasyon açısından güçlü adaylar olarak değerlendirilmektedir. Tezin ilerleyen bölümlerinde, söz konusu segmentasyon mimarilerinin nesne tespiti ve stereo tabanlı derinlik kestirimi bileşenleri ile nasıl bütünleştirildiği ayrıntılı biçimde ele alınacaktır.

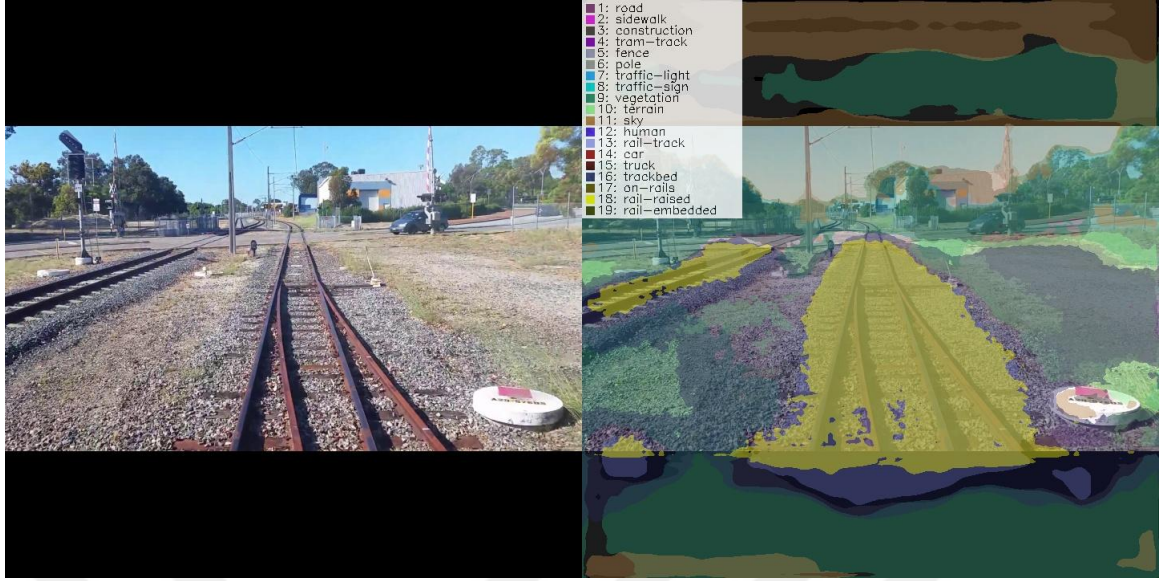
6.3. Çoklu Görev Birleşim Analizi (Multi-Task Fusion Analysis)

Birden fazla görsel algı görevini tek bir hesaplama çerçevesi içinde bütünleştirmek, akıllı demiryolu izleme sistemlerinin geliştirilmesinde kritik bir tasarım zorluğu oluşturmaktadır. Operasyonel bağlamda, bu tür sistemlerin, gerçek zamanlı olarak hem yoğun anlamsal sahne anlayışını gerçekleştirmesi hem de dinamik varlıkları doğru şekilde tespit edip konumlandırması gerekmektedir. Nesne tespiti veya anlamsal segmentasyon gibi göreve özel modeller, ayrı ayrı değerlendirildiklerinde güçlü performans sergilemiş olsalar da, bu modellerin koordinasyonsuz paralel kullanımı genellikle yinelenen hesaplamalara, artan kaynak gereksinimlerine ve sınırlı anlamsal birlikte çalışabilirliğe yol açmaktadır. Bu sınırlamaları aşmak adına, bu çalışma, yapılandırılmış demiryolu ortamlarına özgü, bütünleşik ve alan optimizasyonlu bir algı mimarisi çerçevesinde nesne tespiti ile anlamsal segmentasyonu uyumlu hale getirmeye yönelik iki tamamlayıcı çoklu görev birleşim stratejisini incelemektedir.

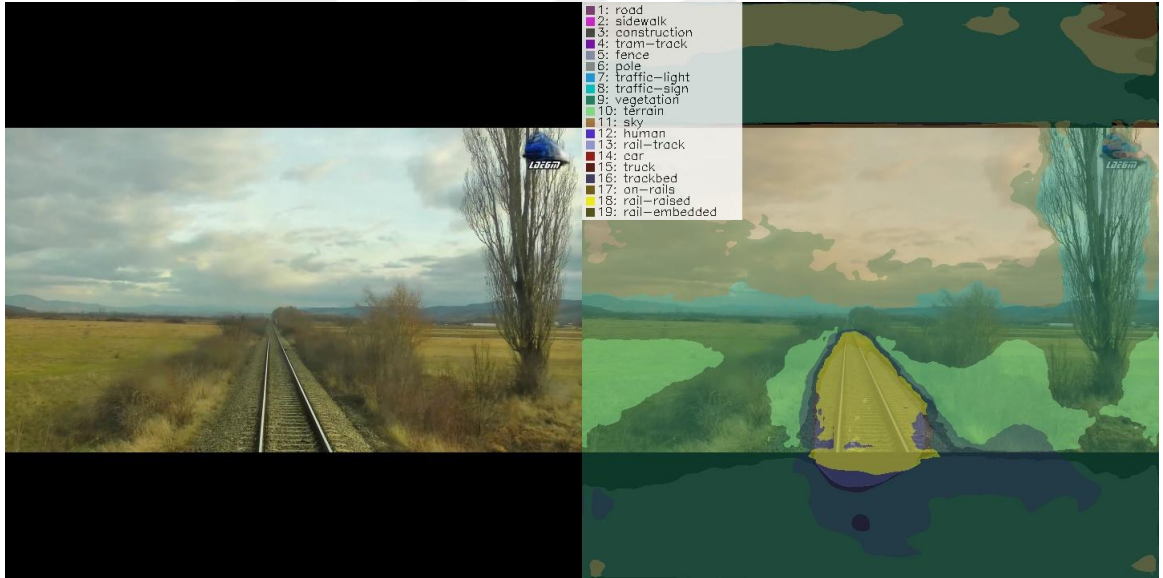
İlk strateji olan Segmentasyon Tabanlı Geometrik Çıkarım, yüksek performanslı bir anlamsal segmentasyon ağı olan ResNet50 omurgalı DeepLabV3+ modelinin çıktılarını, doğrudan bir tespit başlığına ihtiyaç duymaksızın nesne düzeyinde bilgi çıkarmak amacıyla yeniden yapılandırmaktadır. Bu yaklaşım, güvenlik açısından kritik kabul edilen sınıflar (örneğin insan, araç ve engel) için ikili maskeler üretmektedir. Daha sonra bu maskeler, morfolojik dönüşümler, bağlı bileşen analizi ve kontura dayalı sınırlayıcı kutu çıkarımı gibi adımları içeren bir işlem hattı aracılığıyla işlenmektedir. Anlamsal maskeleri sahte tespit ipuçları olarak değerlendiren bu yöntem, ayrı bir nesne tespiti modeline olan gereksinimi ortadan kaldırarak hem çıkarım süresinde gecikmeyi hem de bellek kullanımını önemli ölçüde azaltmaktadır.

Ancak, segmentasyon temelli yaklaşımın bazı içsel sınırlamaları bulunmaktadır. Özellikle nesne odaklı özel bir özellik kodlamasının eksikliği, düşük kontrast, örtülme veya sahne karmaşasının yüksek olduğu zorlu görsel koşullarda sınır belirleme hatalarına karşı duyarlılığı artırmaktadır. Ayrıca, anlamsal segmentasyonun doğası gereği sınıf temelli olması, modelin aynı sınıfa ait çakışan birden fazla nesneyi ayırt etme yeteneğini sınırlamaktadır. Bu durum, özellikle istasyonlar ve hemzemin geçitler gibi yoğun trafiğin bulunduğu demiryolu ortamlarında sıkça karşılaşılan bir sorundur. Bu nedenle, söz konusu sezgisel yöntemin kullanılabilirliği, yalnızca sınırlı nesne ayrımı ve mekânsal-zamansal tutarlılık gerektiren senaryolarla sınırlı kalmaktadır.

Şekil 6.10. ile Şekil 6.13. arasında yer alan görseller, DeepLabV3+ modeli kullanılarak elde edilen segmentasyon tabanlı geometrik çıkarım örneklerini göstermekte olup, modelin nesne sınırlarını belirlemedeki performansını ve bu yaklaşıma bağlı olarak ortaya çıkan sınırlamaları detaylandırmaktadır.



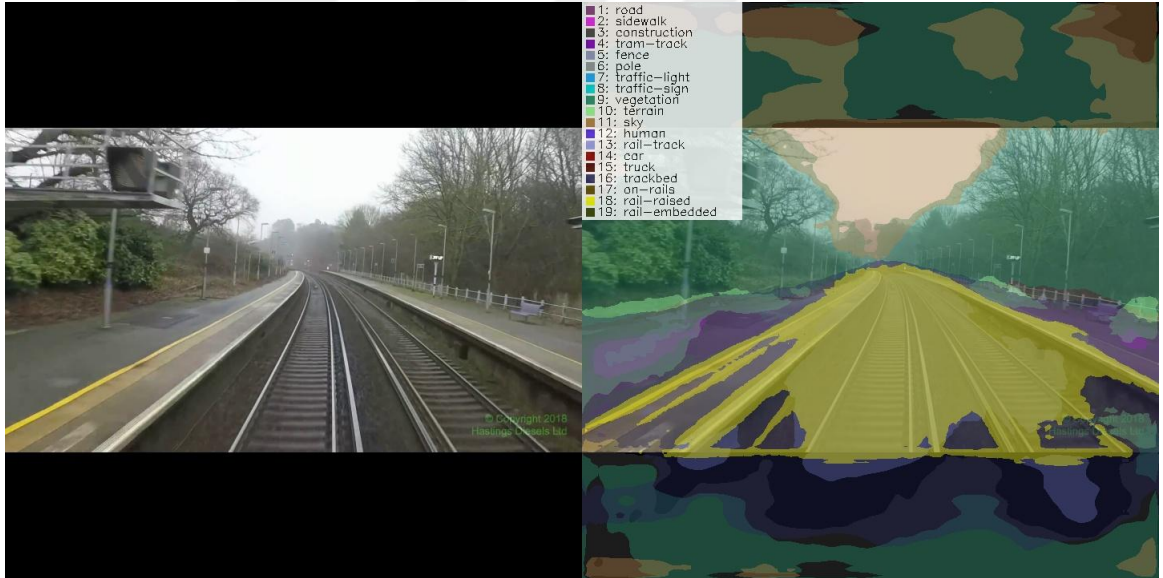
Şekil 6.10. DeepLabV3+ Kullanarak Segmentasyon Tabanlı Geometrik Çıkarım Örneği - 1



Şekil 6.11. DeepLabV3+ Kullanarak Segmentasyon Tabanlı Geometrik Çıkarım Örneği - 2



Şekil 6.12. DeepLabV3+ Kullanarak Segmentasyon Tabanlı Geometrik Çıkarım Örneği - 3



Şekil 6.13. DeepLabV3+ Kullanarak Segmentasyon Tabanlı Geometrik Çıkarım Örneği - 4

Segmentasyon tabanlı çıkarımın doğasında bulunan sınırlamaların üstesinden gelmek ve sahne anlayışının sağlamlığını daha da artırmak amacıyla, DeepLabV3+ anlamsal segmentasyon modeli ile YOLOv11 nesne tespit ağı entegre edilerek daha gelişmiş bir çift akışlı füzyon mimarisi uygulanmıştır. Bu yapılandırma, her iki yöntemin birbirini tamamlayıcı güçlü yönlerinden faydalanacak şekilde özel olarak tasarlanmıştır: DeepLabV3+'ın piksel düzeyindeki semantik hassasiyeti ile YOLOv11'in nesne düzeyindeki konumlandırma doğruluğu. Her iki model de bağımsız eğitim süreçlerinden geçirildikten sonra aynı girdi çerçevesini paralel olarak işlemek

üzere konuşlandırılmıştır. Segmentasyon akışı, özellikle raylar olmak üzere statik demiryolu altyapı unsurlarını yüksek yapısal tutarlılıkla öne çıkarırken; tespit akışı ise tren, araç ve yaya gibi dinamik ve güvenlik açısından kritik nesneleri sınırlayıcı kutular (bounding box) aracılığıyla belirlemiştir.

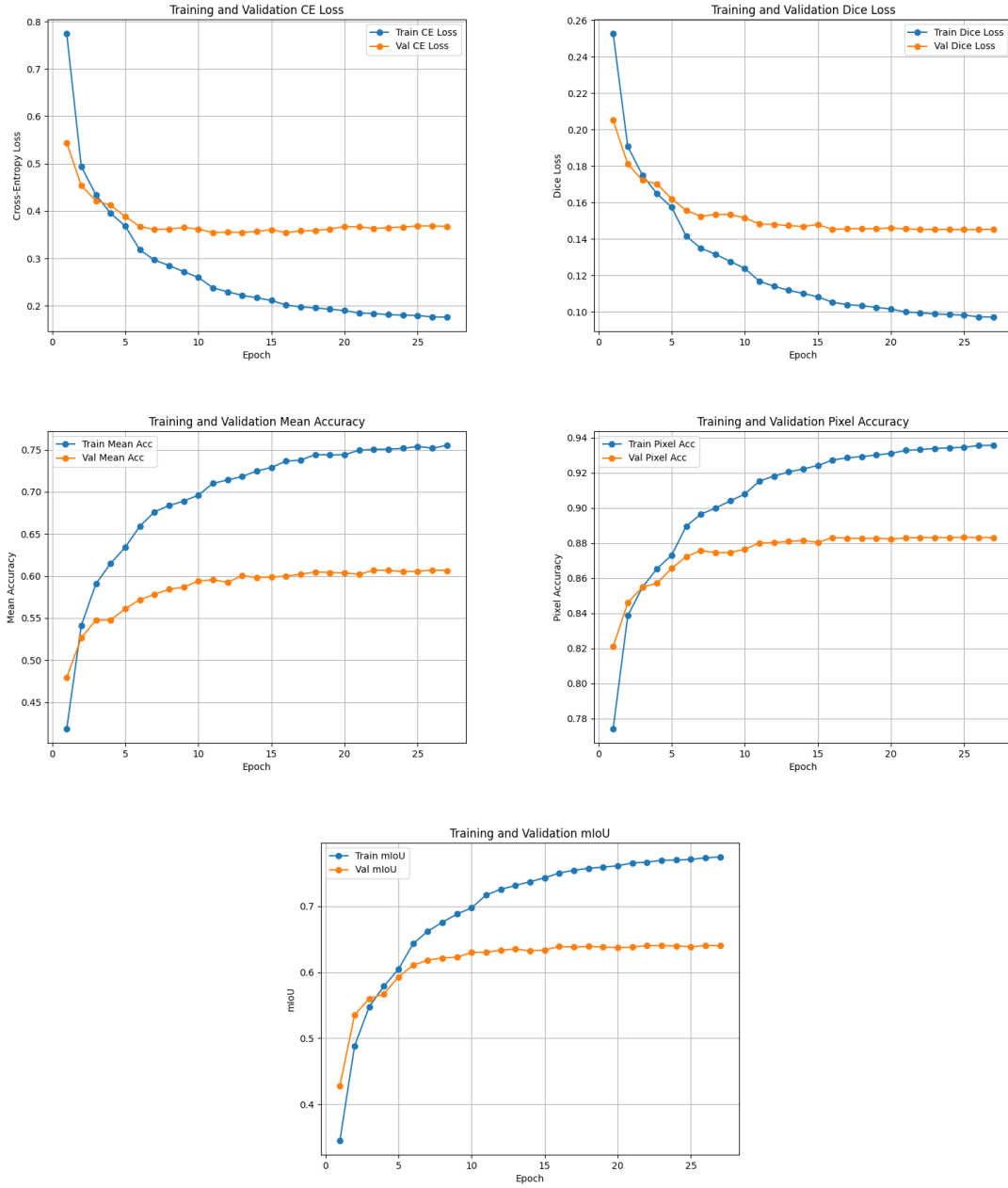
Bu araştırma aşamasında getirilen önemli bir iyileştirme, DeepLabV3+ modelinin geliştirilmiş hiperparametrelerle ve uzatılmış eğitim süresiyle 30 epoch boyunca yeniden eğitilmesi olmuştur. Bu sayede hem eğitim kararlılığı hem de segmentasyon doğruluğunda kayda değer bir artış sağlanmış, başlangıçtaki 40 epoch'luk yapılandırmaya kıyasla ortalama Birleşim Üzerinden Kesişim (mIoU) değerinde belirgin bir iyileşme elde edilmiştir.

Şekil 6.14'te, yeniden eğitilen DeepLabV3+ modeline ait beş farklı değerlendirme metriği olan çapraz entropi kaybı (ce_loss), dice kaybı (dice_loss), ortalama doğruluk (mean accuracy), ortalama kesişim-birleşim oranı (mIoU - mean Intersection over Union) ve piksel doğruluğu (pixel accuracy), hem eğitim hem de doğrulama kümeleri için ayrı ayrı sunulan birleşik bir grafik üzerinde gösterilmektedir. Bu grafik, modelin öğrenme sürecine ilişkin çok boyutlu ve kapsamlı bir değerlendirme yapılmasına imkân tanımaktadır. Özellikle çapraz entropi kaybı (ce_loss) ile dice kaybı (dice_loss) değerlerinde gözlemlenen düzenli azalma, modelin giderek daha kararlı bir biçimde öğrenme gerçekleştirdiğini ortaya koymaktadır. Öte yandan, doğruluk temelli metrikler arasında yer alan ortalama doğruluk (mean accuracy) ve piksel doğruluğu (pixel accuracy) değerlerinde gözlenen genel artış, modelin aşırı öğrenmeden (overfitting) kaçınarak güçlü bir genelleme yeteneği kazandığını göstermektedir.

Elde edilen sonuçların nicel bir özeti Tablo 6.5'te sunulmaktadır. Buna göre, eğitim aşamasında mIoU değeri %77,45'e, piksel doğruluğu %93,57'ye ulaşmıştır. Doğrulama ve test aşamalarında elde edilen mIoU değerleri sırasıyla %64,05 ve %64,84 olup; bu durum, modelin farklı veri kümelerinde istikrarlı bir performans sergilediğini ortaya koymaktadır. Özellikle test verisindeki başarı, modelin gerçek dünya senaryolarında da uygulanabilirliğini göstermektedir. Bu sonuçlar, DeepLabV3+ modelinin yeniden eğitilmesinin yalnızca eğitim verisi üzerinde değil, aynı zamanda görülmemiş veriler üzerindeki başarımını da artırdığını ortaya koymaktadır.

Tablo 6.5. DeepLabV3+ Modeline Ait Eğitim, Doğrulama ve Test Performansları

Metrik	Eğitim	Doğrulama	Test
CE Loss	0.1763	0.3677	0.3532
Dice Loss	0.0971	0.1453	0.1444
mIoU	0.7745	0.6405	0.6484
Pixel Accuracy	0.9357	0.8831	0.8871
Mean Accuracy	0.7552	0.6064	0.6001



Şekil 6.14. DeepLabV3+ Modeline Ait Eğitim ve Doğrulama Metriği Eğrileri (CE Loss, Dice Loss, mIoU, Pixel Accuracy, Mean Accuracy)

Bu bağlamda, her iki akışın (segmentasyon ve nesne tespiti) çıktılarının tutarlı bir şekilde bütünleştirilmesini sağlamak amacıyla, çıkarım sonrası bir füzyon modülü geliştirilmiştir. Anlamsal segmentasyon çıktıları, sahne içindeki mekânsal ilişkilerin yorumlanmasına olanak tanıyan zengin bir bağlamsal yapı sunmakta olup, özellikle ray alanının tamamının sarı renkle sınırlanması yoluyla görsel olarak tanımlanmasında kullanılmıştır. Aynı anda, YOLOv11 tabanlı nesne tespit modeli tarafından sahnede tanımlanan hareketli nesneler kırmızı sınırlayıcı kutular ile

gösterilmiş ve böylece dinamik varlıkların konumları işaretlenmiştir. Tespit edilen bir nesnenin (örneğin bir tren veya kara taşıtı), segmentasyon sonucu elde edilen ray bölgesi ile çakışması durumunda — bu çakışma piksel düzeyinde örtüşme analizi aracılığıyla belirlenmiştir — ilgili sınırlayıcı kutu yeşil renkle vurgulanarak olası bir ray işgali durumu görsel biçimde ifade edilmiştir. Bu yaklaşım, semantik ve geometrik bilgileri bütünleştiren ek bir mantıksal katman oluşturarak, yalnızca sahnede algılama yapılmasını değil, aynı zamanda mekânsal düzlemde akıl yürütülmesini mümkün kılmış; böylece ham veriler, anlamlı ve eyleme dönüştürülebilir içgörülere dönüştürülmüştür.



Şekil 6.15. Çift Akıslı Tespit-Segmentasyon Birleşimi Örneği - 1



Şekil 6.16. Çift Akıslı Tespit-Segmentasyon Birleşimi Örneği - 2



Şekil 6.17. Çift Akıslı Tespit-Segmentasyon Birleşimi Örneği - 3

Bu füzyon sürecine ait görsel örnekler, çeşitli sahneler ve aydınlatma koşullarında elde edilen başarılı entegrasyonları göstermek üzere Şekil 6.15 ile Şekil 6.17 arasındaki şekillerde sunulmuştur. Elde edilen çıktılar, yalnızca daha yüksek yorumlanabilirlik sağlamakla kalmayıp; çarpışma riski değerlendirmesi, dijital ikizler için sahne açıklaması (annotasyon) ve operatör destek

sistemleri gibi aşığı akış görevlerine de katkı sağlamaktadır. Önemle vurgulanmalıdır ki, bu çift akışlı mimari gerçek zamanlı uygulanabilirliğini koruyarak, CUDA destekli sistemlerde çerçeve başına toplam çıkarım gecikmesini 100 milisaniyenin altında tutmuştur.

Standart nesne tespiti ve segmentasyon ölçütleriyle yapılan nicel değerlendirme, çift akışlı mimarinin yalnızca segmentasyona dayalı temel yapıya kıyasla üstün performans sergilediğini doğrulamıştır. Özellikle, birleştirilmiş sistem nesne tespiti için daha yüksek ortalama keskinlik (mAP), segmentasyon doğruluğu için daha iyi ortalama IoU (mIoU) ve CUDA hızlandırması altında saniyede 30 kareyi (FPS) aşan gerçek zamanlı çıkarım hızı sunmuştur. Bu performans artışları, özellikle örtümler, değişken aydınlatma koşulları ve yüksek nesne yoğunluğu gibi olumsuz durumlarda belirgin olmuştur. Niteliksel analiz de bu bulguları desteklemiş; görselleştirmeler, yalnızca segmentasyon temelli çıktılara kıyasla daha keskin nesne konturları, azaltılmış mekansal hizalama hataları ve daha tutarlı örtüşmeler ortaya koymuştur.

Çift akışlı tasarımın ek bir avantajı ise modülerliği ve genişletilebilirliğidir. Her bir akış bağımsız olarak çalıştığından, genel sistemi bozmadan güncellenebilir veya değiştirilebilir; bu da gelecekteki model geliştirmeleri veya donanım özelinde yapılan iyileştirmelerle kolayca bütünleşmesini sağlar. Bu gevşek bağlı mimari, görevler arası etkileşim ve yetersiz ortak optimizasyon gibi uçtan uca çok görevli öğrenmeye özgü yaygın sorunlardan kaçınırken, çok modlu algı için esnek bir temel sunar. Özellikle, sistemin stereo ayrıklık haritaları, LiDAR nokta bulutları veya atalet sensörü verileri gibi yardımcı veri kaynaklarını da entegre edecek şekilde genişletilmesi mümkündür.

Özetle, segmentasyon tabanlı geometrik çıkarım, durağan veya düşük riskli demiryolu ortamları için hesaplama açısından hafif bir temel sağlarken, çift akışlı birleşim mimarisi gerçek zamanlı sahne anlayışı için çok daha sağlam ve anlamsal olarak zengin bir çözüm sunmaktadır. Bu yaklaşım, örnek düzeyindeki hassasiyet ile bağlamsal farkındalığı dengeleyerek, modern demiryolu izleme sistemleri için gerekli olan akıllı ve güvenlik odaklı algıyı mümkün kılar. İzleyen bölümde, bu birleşim çerçevesi stereo görüş yoluyla derinlik çıkarımı ile genişletilecek; böylece tam sahne 3B anlayışı ve zamansal nesne takibi sağlanacaktır.

6.4. ZED2 Entegrasyonu ve Mesafe Ölçümü Analizi

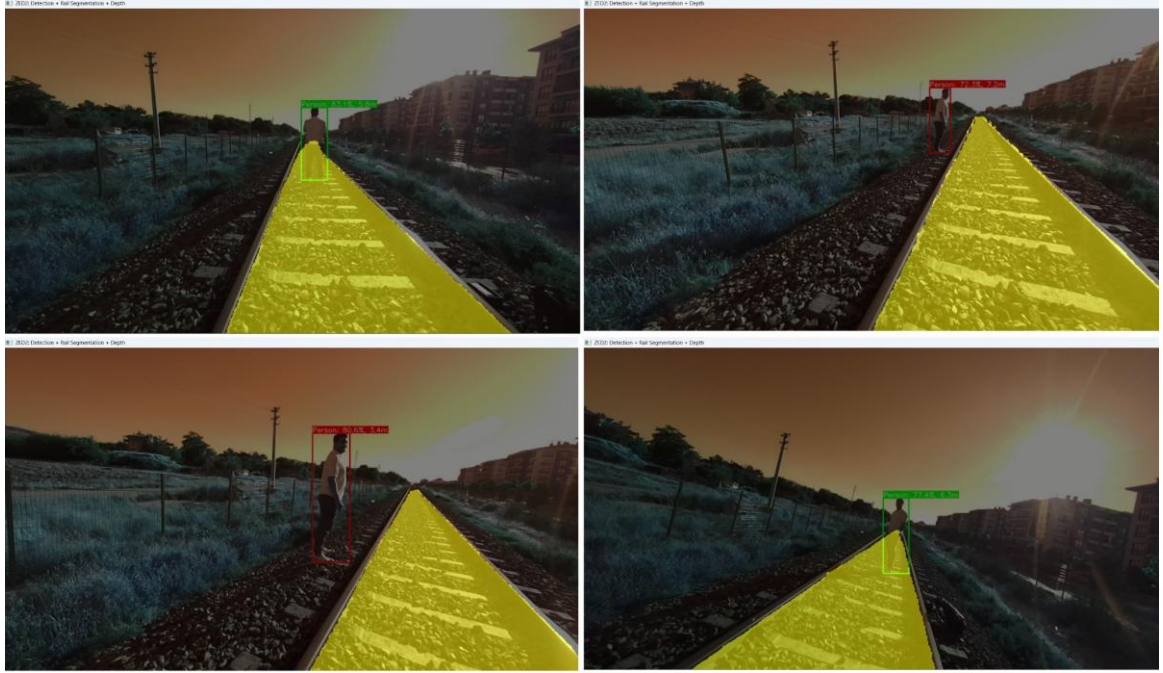
Bu araştırmanın temel katkılarından biri, önerilen görsel algı çerçevesine gerçek zamanlı stereo derinlik algısının entegre edilmesidir. Bu entegrasyon, ZED2 stereo kameranın kullanımıyla gerçekleştirilmiş olup, iki boyutlu görsel işleme hattının yeteneklerini önemli ölçüde genişletmiştir. Böylece, otonom demiryolu ortamlarında operasyonel güvenlik için kritik öneme sahip olan metrik ölçekli mekânsal çıkarım ve derinliğe dayalı karar alma süreçleri mümkün hale gelmiştir. Geleneksel monoküler görsel sistemler, anlamsal yorumlama ve nesne tespiti açısından etkili olsalar da, doğru derinlik tahmini yapma kapasitesinden yoksundurlar. Bu durum, özellikle

demiryolu otomasyonu gibi dinamik ve güvenlik açısından kritik alanlarda uygulanabilirliklerini sınırlandırmaktadır. Stereo görüşün dâhil edilmesi bu sınırlamayı ortadan kaldırmakta; piksel düzeyinde yoğun derinlik tahminleri sunarak iki boyutlu sahne analizinden üç boyutlu durumsal farkındalığa geçişi mümkün kılmaktadır.

Stereolabs tarafından geliştirilen ZED2 kamera, sabit 12 santimetrelik taban uzaklığına sahip çift global shutter sensör içermektedir. Bu yapılandırma, kalibre edilmiş stereo eşleme yoluyla hassas binoküler ayrıklık (disparity) tahminine olanak tanır. Sistem, senkronize görüntü çiftleri yakalar ve ZED SDK'nin derinlik tahmin boru hattını kullanır. Bu boru hattı, yarı-küresel eşleme (semi-global matching) (SGM), kenar farkındalıklı filtreleme ve güven skoruna dayalı iyileştirme yöntemlerini birleştirerek gerçek zamanlı yoğun derinlik haritaları üretir. Bu tez kapsamında ZED2, yüksek mekânsal çözünürlük ve genişletilmiş derinlik aralığına öncelik veren ULTRA modunda çalıştırılmıştır. Bu mod, demiryolu ortamlarına özgü orta ve uzun mesafelerdeki sabit altyapı ve hareketli nesnelerin tespit ve takibi açısından kritik önemdedir.

Stereo derinlik boru hattının mevcut çoklu görev algı çerçevesiyle sıkı entegrasyonunu sağlamak amacıyla, bu yapı YOLOv11 nesne tespit modeli ve DeepLabV3+ anlamsal segmentasyon modeli çıktılarıyla birleştirilmiştir. RGB çerçevesinde bir nesne tespit edildiğinde, sınırlayıcı kutunun ağırlık merkezi, karşılık gelen derinlik haritasına yansıtılarak yerel ayrıklık değeri elde edilir. ZED SDK tarafından sağlanan içsel kalibrasyon parametreleri ve stereo üçgenleme modelleri kullanılarak, bu ayrıklık değeri gerçek dünya (X, Y, Z) koordinatlarına dönüştürülür. Böylece sistem, nesnelerin fiziksel uzaydaki konumlarını yüksek zamansal ve mekânsal doğrulukla belirleyebilir; bu da tehlike tahmini, engel takibi ve yakınlık temelli risk analizi gibi görevleri destekler.

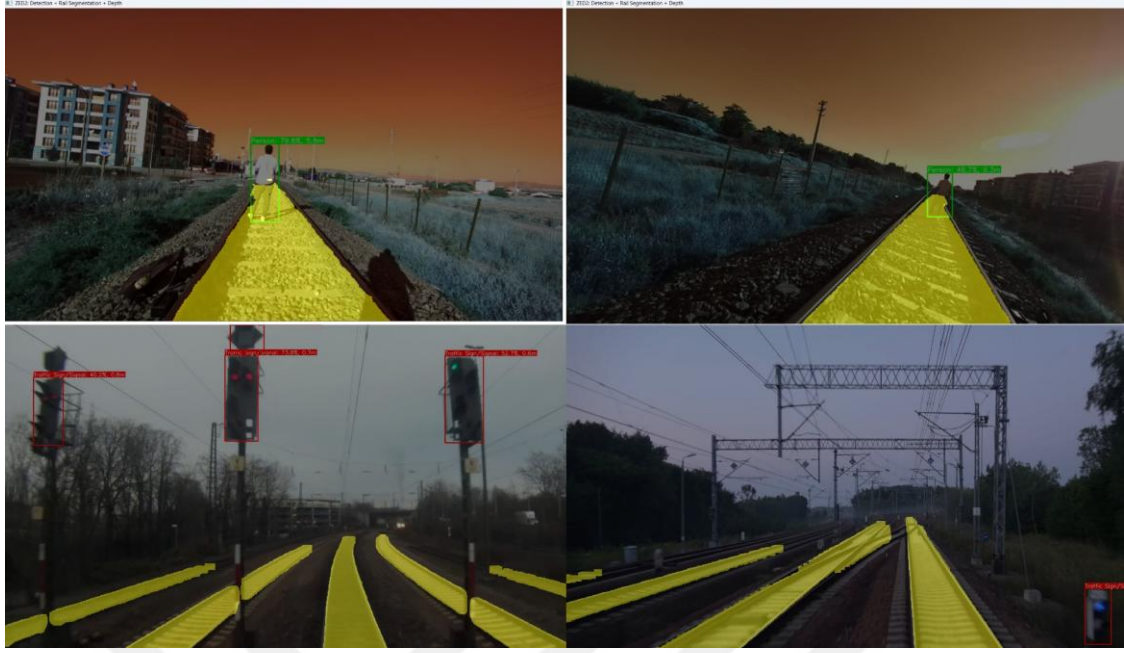
Şekil 6.18 ile Şekil 6.21 arasında sunulan görseller, önerilen çoklu görev algılama çerçevesi kapsamında DeepLabV3+, YOLOv11 ve ZED2 stereo derinlik kamerasının entegrasyonu ile elde edilen nesne tespiti, anlamsal segmentasyon ve üç boyutlu mesafe ölçümü sonuçlarını örneklemektedir.



Şekil 6.18. DeepLabV3plus + YOLOv11 + ZED2 Örneği - 1



Şekil 6.19. DeepLabV3plus + YOLOv11 + ZED2 Örneği - 2



Şekil 6.20. DeepLabV3plus + YOLOv11 + ZED2 Örneği - 3



Şekil 6.21. DeepLabV3plus + YOLOv11 + ZED2 Örneği - 4

Sistemin çıktı görselleştirmesi, OpenCV ve OpenGL'in birlikte kullanımıyla uygulanmıştır. 2B bindirmelerde (overlay), derinlik farkındalıklı sınırlayıcı kutular ve anlamsal segmentasyon maskeleri canlı RGB video akışına üst üste bindirilmiş; mesafe aralıklarını temsil eden renk kodlu açıklamalarla zenginleştirilmiştir. Üç boyutlu görselleştirme için ise OpenGL kullanılarak derinlik

ipuçları içeren sahne geometrisi render edilmiştir. Bu görselleştirme yetenekleri, özellikle durumsal farkındalığın hızlı ve net bir şekilde iletilmesinin gerekli olduğu operasyonel bağlamlarda büyük avantaj sağlamaktadır.

Derinlik tahmin modülünün doğruluğunu değerlendirmek amacıyla, 1 ila 20 metre arasında bilinen mesafelerde konumlandırılmış referans işaretleyiciler kullanılarak kontrollü deneyler gerçekleştirilmiştir. Sistemin ortalama mutlak derinlik hatası %10'un altında kalmış; en yüksek doğruluk, 3 ile 10 metre arasındaki mesafelerde elde edilmiştir. Bu mesafe aralığı, demiryolu izleme senaryolarında tipik olarak tehdit bölgesiyle örtüşmektedir. Görsel dokunun zayıf olduğu veya yetersiz aydınlatılmış alanlarda doğrulukta bir miktar düşüş gözlemlenmiş olup, bu durum pasif stereo sistemlerin yapısal sınırlamalarıyla uyumludur. Bununla birlikte, anlamsal segmentasyonun sisteme entegrasyonu sayesinde, arka plan bölgelerinden kaynaklanan gürültü bastırılmış ve derinlik temelli tespitlerin mekânsal doğruluğu önemli ölçüde artırılmıştır.

Stereo derinlik tahmini tarafından eklenen hesaplama gecikmesi, gerçek zamanlı çalışmaya uygun sınırlar içerisinde kalmıştır. Nesne tespiti, anlamsal segmentasyon, derinlik yansıtımı ve görselleştirme dâhil olmak üzere tüm çıkarım boru hattı, CUDA hızlandırması altında kare başına yaklaşık 180–200 milisaniyelik ortalama işlem süresine ulaşmıştır. Bu da saniyede 15 kareyi aşan etkili bir kare hızına karşılık gelmektedir. Bu performans seviyesi, yerinde (edge) konuşlandırılmış demiryolu izleme sistemlerinin canlı yayın işleme gereksinimlerini karşılamaktadır.

Niteliksel değerlendirme sonuçları, stereo görüş entegrasyonunun sağladığı avantajları doğrulamıştır. Kısmi örtülmelerin bulunduğu, dar ya da düşük profilli nesnelerin yer aldığı ve heterojen arka planların gözlemlendiği karmaşık sahnelerde, derinlik modülünün kullanımı, nesne ayrımını iyileştirmiş ve mekânsal tutarlılığı artırmıştır. Örneğin, istasyon altyapısı tarafından kısmen örtülen yayalar dahi üç boyutlu olarak doğru şekilde konumlandırılabilmiş; bu durum, engel önceliklendirme ve risk modelleme süreçlerinin daha etkin bir şekilde yürütülmesini sağlamıştır. Buna ek olarak, derinlik eklentisi, yalnızca iki boyutlu temsillerde çakışma eğiliminde olan ön plan ve arka plan nesnelerinin, örneğin uzaktaki bir sinyal direğiyle çakışan yakın bir yayanın, birbirinden ayrılabilmesini mümkün kılmış ve böylece anlamsal yorumlamaların güvenilirliğini artırmıştır.

Sonuç olarak, ZED2 stereo kamerasının algı çerçevesine entegrasyonu, sağlam ve mekânsal olarak akıllı demiryolu sahne anlayışı yönünde önemli bir ilerleme sunmaktadır. Derinlik farkındalıklı nesne tespiti, anlamsal segmentasyon ve üç boyutlu konumlandırmayı mümkün kılarak sistem, monoküler görüşün sınırlamalarını aşmakta ve kestirimci, proaktif güvenlik mekanizmaları için bir temel oluşturmaktadır. Geometrik ve anlamsal modalitelerin bu birleşimi, yorumlanabilirliği artırmakta, belirsizlikleri azaltmakta ve çarpışma önleme, ray izinsiz giriş uyarıları ve riske dayalı karar verme gibi kritik işlevleri desteklemektedir. Bu bağlamda, bu tezde

geliştirilen stereo destekli algı hattı, yeni nesil otonom demiryolu platformları için pratik ve genişletilebilir bir çözüm sunmaktadır.

6.5. Tartışma ve Analiz

Önceki bölümlerde sunulan deneysel sonuçlar, önerilen algı çerçevesinin nesne tespiti, anlamsal segmentasyon, çok görevli füzyon ve stereo tabanlı derinlik tahmini dâhil olmak üzere performans ve entegrasyon yeteneklerini bütüncül bir şekilde ortaya koymaktadır. Bu bölümde, elde edilen bulgular sentezlenmekte; sistemin gözlemlenen davranışları, altında yatan ödünleşimler, konuşlandırma hususları ve gerçek dünya demiryolu otomasyonu için daha geniş etkileri eleştirel bir değerlendirmeye incelenmektedir.

Nesne tespiti deneylerinden elde edilen sonuçlara göre, YOLOv11, kesinlik, çıkarım gecikmesi ve alana özgü görsel karmaşıklıklara karşı dayanıklılık bakımından en dengeli mimari olarak öne çıkmıştır. YOLOv8, mAP@0.5 ve precision gibi standart metriklerde marjinal olarak daha yüksek puanlar elde etse de, YOLOv11; yoğun örtülme, sınıf dengesizliği ve karmaşık demiryolu arka planları gibi zorlu koşullarda daha üstün bir dayanıklılık sergilemiştir. Bu avantaj, büyük ölçüde, optimize edilmiş özellik toplama modülleri ve geliştirilmiş uzamsal piramit havuzlama katmanlarına atfedilebilir; bu da görsel açıdan dağınık ya da dinamik sahnelerde daha iyi konumlandırma ve sınıflandırma sağlamıştır. Daha hafif mimarili olan YOLOv12 ise tespit doğruluğunda belirgin bir düşüş sergileyerek, yalnızca parametre sayısını azaltmanın temsil gücünü korumak için yeterli olmadığını ortaya koymuştur. Öte yandan, Faster R-CNN kontrollü ortamlarda yüksek sınıflandırma doğruluğu sunsa da, çıkarım gecikmesi gerçek zamanlı gömülü uygulamalar için engelleyici düzeydedir. Bu bulgular, gecikmeye duyarlı otonom sistemlerde çapa (anchor) içermeyen, tek aşamalı dedektörlerin savunulduğu güncel literatürle örtüşmektedir.

Anlamsal segmentasyon alanında farklı mimari tasarımlar değerlendirilmiş ve bu tasarımların çeşitli performans avantajları ortaya konulmuştur. Özellik Piramit Ağı (Feature Pyramid Network, FPN), çok ölçekli özellikleri etkili biçimde birleştirme yeteneği sayesinde, karmaşık sahnelerin ayrıntılı çözümlemesinde başarılı olmuş ve en yüksek ortalama Birleşim Üzerinden Kesişim (mean Intersection over Union, mIoU) skorunu elde etmiştir. Bununla birlikte, önceden eğitilmiş ResNet34 kodlayıcısıyla yapılandırılan UNet türevi mimari, büyük ölçekli doğal görüntü kütlelerinden sağlanan aktarım öğrenmesi (transfer learning) avantajlarıyla benzer bir performansı daha yüksek eğitim kararlılığı ve daha hızlı yakınsama süreci ile sağlamıştır. DeepLabV3+ modeli ise toplam performans metriklerinde ilk sırada yer almamış olmakla birlikte, boşluklu (atrous) uzamsal piramit havuzlama ve sınır iyileştirme mekanizmaları sayesinde raylar, makaslar ve üst hatlar gibi dar ve altyapı açısından kritik sınıfları ayırt etme konusunda üstün bir performans sergilemiştir. Öte yandan, Piramit Sahne Ayrıştırma Ağı (Pyramid Scene Parsing Network, PSPNet), özellikle uzun ve düşük dokulu kategorilerde nesne sınırlarını belirleme

açısından yetersiz kalmıştır. Elde edilen bulgular, segmentasyon görevlerinde yalnızca genel skorların değerlendirilmesinin yeterli olmadığını; bunun yanı sıra yapısal özellikler ve uygulama alanına özgü ölçütlerin de göz önünde bulundurulması gerektiğini göstermektedir. Bu durum, özellikle güvenlik açısından kritik öneme sahip demiryolu ortamlarında daha da belirgin hale gelmektedir.

Çok görevli füzyon mimarilerinin değerlendirilmesi, sistem karmaşıklığı ile algısal doğruluk arasındaki ödünleşimlere dair ek içgörüler sunmuştur. Anlamsal çıktıları dolaylı tespit ipuçları olarak yeniden yorumlayan segmentasyon tabanlı geometrik çıkarım boru hattı, hesaplama açısından verimli bir çözüm sunmuştur. Bu yaklaşım sistem karmaşıklığını ve çıkarım süresini azaltmakla birlikte; örtülmeler, çakışan örnekler ve düşük güven skorlu sınırlarda sınırlı başarı göstermiştir. Buna karşın, YOLOv11 ile DeepLabV3+'ın birleştirildiği çift akışlı füzyon mimarisi, örnek düzeyinde nesne tespiti ile piksel düzeyinde anlamsal bağlamı sinerjik şekilde bütünleştirmiştir. Bu yapılandırma, niceliksel olarak üstün sonuçlar üretmiş ve niteliksel olarak daha tutarlı sahne yorumları sağlamıştır. Çıkarım sonrası uygulanan mekânsal hizalama ve sınıfa duyarlı filtreleme gibi işlemler, yanlış pozitifleri azaltmış ve mekânsal konumlandırmayı iyileştirmiştir; bu da özellikle insan-karar döngüsü içeren sistemler ve otonom müdahale senaryoları için uygun bir yapı sunmaktadır.

ZED2 stereo görüş sisteminin entegrasyonu, gerçek zamanlı derinlik algısı sağlayarak mekânsal anlayışa önemli bir boyut kazandırmıştır. Derinlik bilgisi, nesne tespiti ve segmentasyon çıktılarıyla birleştirilmiş; bu sayede doğru mesafe tahminine dayalı üç boyutlu (3B) sahne rekonstrüksiyonları elde edilmiştir. Ampirik doğrulamalar, 1 ila 20 metre arasındaki mesafelerde ortalama mutlak hatanın %10'un altında kaldığını ortaya koymuştur. En yüksek doğruluk ise 3 ila 10 metre arasındaki mesafelerde gözlemlenmiş; bu durum, söz konusu aralığın kısa ve orta menzilli demiryolu gözetim görevleriyle en yakından ilişkili mesafeler olması bakımından önem arz etmektedir. Yakınlık temelli nesne önceliklendirme yeteneği, riske duyarlı karar alma süreçleri açısından ciddi katkı sağlamıştır. Bununla birlikte, stereo görüş sistemleri düşük dokulu ya da yüksek yansıtıcılığa sahip yüzeylerde ayrıklık (disparite) tahmininde kararsızlık göstermiştir. Ayrıca, derinlik işleme hattı, özellikle çift akışlı görsel boru hatları ile entegre edildiğinde, gerçek zamanlı performansın sürdürülebilmesi için etkin bellek yönetimi ve görev zamanlaması gerektiren kayda değer bir hesaplama yükü oluşturmıştır.

Genel sistem performansı, GPU hızlandırılmalı koşullarda gerçek zamanlılık sınırları içinde kalmayı başarmıştır. Kare başına çıkarım süresi sürekli olarak 90 milisaniyenin altında gerçekleşmiş ve bu da tespit, segmentasyon, derinlik projeksiyonu ve görselleştirme gibi bileşenleri içeren algı yığını açısından saniyede 11 kareyi aşan bir işlem hızı anlamına gelmiştir. Elde edilen bu performans düzeyi, gerçek zamanlı demiryolu denetimi, erken uyarı sistemleri ve izleme uygulamaları gibi alanlarda ihtiyaç duyulan gecikme gereksinimleriyle uyumlu niteliktedir.

Özellikle füzyon mimarisinin, bileşenlerin tekil performansını düzenli olarak aşması dikkat çekicidir. Aydınlatma değişkenliği, hareket bulanıklığı ve arka plan kalabalığı gibi çevresel bozulmalara karşı gösterilen gelişmiş dayanıklılık, çok modlu algılama yaklaşımlarının daha bağlam farkındalıklı ve dayanıklı sahne yorumları sunduğu yönündeki temel tezi desteklemektedir.

Buna rağmen sistemin bazı sınırlamaları bulunmaktadır. RailSem19 veri seti, hava koşulları ve sahne çeşitliliği açısından zengin bir içeriğe sahip olmakla birlikte; gece operasyonları, yüksek hızlı nesne hareketleri ve nadir güvenlik açısından kritik olaylar gibi uç senaryoları yeterince kapsamamaktadır. Bu durum, sistemin az temsil edilen bu senaryolarda genelleme yeteneğini sınırlandırmaktadır. Ayrıca, entegre sistemin kaynak tüketimi, özellikle GPU bellek kullanımı ve enerji ihtiyacı, NVIDIA Jetson Nano ya da Raspberry Pi gibi düşük güçlü gömülü platformlara konuşlandırılmasını güçleştirmektedir. Bu sınırlamaların giderilebilmesi için model sıkıştırma tekniklerinin (örneğin, budama, niceleme, bilgi damıtımı) daha derinlemesine incelenmesi; bunun yanı sıra kenar bilişim ortamlarına uygun hafif omurga ağları ile dinamik çıkarım stratejilerinin geliştirilmesi gerekmektedir.

Sonuç olarak, önerilen algı çerçevesi, gerçek zamanlı demiryolu uygulamaları için birleşik, modüler ve derinlik farkındalıklı bir sahne anlayışı sistemi sunma hedefine başarıyla ulaşmıştır. Anlamsal segmentasyon, nesne tespiti ve stereo derinlik tahmini gibi tamamlayıcı bilgisayarla görme görevlerinin bütünsel biçimde entegre edilmesi, sistemin hem teknik uygulanabilirliğini hem de işlevsel güvenilirliğini önemli ölçüde artırmıştır. Bu yapı; engel tespiti, yakınlık temelli uyarı sistemleri ve altyapı izleme gibi güvenlik açısından kritik işlevlerin yürütülmesini yüksek doğrulukla destekleyerek, sahadaki operasyonel risklerin azaltılmasına katkı sağlamıştır.

Geliştirilen sistem mimarisi, yalnızca mevcut ihtiyaçlara yanıt vermekle kalmayıp; zamansal izleme, 3B sahne yeniden yapılandırması ve çok etmenli işbirlikçi algı gibi ileri düzey araştırma alanlarına da uyarlanabilir yapısıyla açık bir potansiyel sunmaktadır. Elde edilen bulgular, akıllı ulaşım sistemleri literatürüne anlamlı bir katkı sağlayarak, otonom demiryolu platformları ve gömülü algı teknolojileri odağında yürütülecek gelecekteki çalışmalara sağlam ve genişletilebilir bir temel teşkil etmektedir.

7. SONUÇLAR

Bu son bölüm, gerçek zamanlı demiryolu sahne anlayışı için önerilen çok görevli derin öğrenme çerçevesine ilişkin temel bulguları ve katkıları sentezleyerek araştırmanın genel bir değerlendirmesini sunmaktadır. Bir önceki bölümde ortaya konan sonuçlara dayanarak, bu kısım; nesne tespiti, anlamsal segmentasyon, derinlik tahmini ve çok görevli füzyon alanlarında elde edilen ampirik başarıları bir araya getirmekte ve bunların akıllı demiryolu algı sistemlerinin gelişimine olan toplu etkisini vurgulamaktadır.

Bölüm 7.1, tezin başlıca katkılarını ve çıktılarının özetini sunmakta; geliştirilen çerçevenin etkinliği, verimliliği ve gerçek dünya koşullarındaki uygulanabilirliğini öne çıkarmaktadır. Bölüm 7.2, deneysel süreç ve sistem geliştirme sırasında karşılaşılan yöntemsel ve operasyonel sınırlılıkları ele almakta; iyileştirme alanlarını ve gerçek dünya uygulamalarına özgü zorlukları kabul etmektedir.

Son olarak, Bölüm 7.3 ileriye dönük bir değerlendirme sunarak, gelecekteki araştırmalara ve sistem iyileştirmelerine yönelik olası yönleri tartışmaktadır. Bu öneriler özellikle güvenlik kritik taşıma ortamlarında genellenebilirliğin, sağlamlığın ve ölçeklenebilirliğin artırılmasına odaklanmaktadır. Bu tartışmalar bütününde, tez; akıllı ulaşım sistemleri bağlamında konumlandırılmakta ve demiryolu otomasyonunda sürekli yenilik için sağlam bir temel teşkil etmektedir.

7.1. Katkıların ve Sonuçların Özeti

Bu tez, demiryolu ortamlarında görsel algılama için kapsamlı, modüler ve gerçek zamanlı bir derin öğrenme çerçevesi sunmaktadır. Emniyet açısından kritik demiryolu altyapılarını izleme ihtiyacını karşılamayı hedefleyen bu sistem; anlamsal segmentasyon, nesne tespiti ve stereo görüş tabanlı derinlik tahminini birleşik bir iş hattında entegre etmektedir. ZED2 stereo kamera ile birlikte son teknoloji görsel mimarilerden faydalanarak, sistem dinamik, karmaşık ve yapısal olmayan çalışma koşullarında yüksek düzeyde mekânsal farkındalık ve anlamsal anlama kabiliyeti sunmaktadır.

Bu algılama sisteminin geliştirilmesi ve değerlendirilmesini desteklemek amacıyla, RailSem19 veri seti üzerinde beş farklı nesne tespit mimarisi (YOLOv8, YOLOv9, YOLOv11, YOLOv12 ve Faster R-CNN) ile beş anlamsal segmentasyon modeli (UNet-Vanilla, UNet-Pretrained, DeepLabV3+, Feature Pyramid Network (FPN) ve Pyramid Scene Parsing Network (PSPNet)) uygulanmış ve sistematik olarak kıyaslanmıştır. Tespit modelleri arasında YOLOv11, (mAP@0.5 = 0.50412), çıkarım verimliliği ve görsel örtüşme ile altyapıya özgü bozulmalara karşı dayanıklılık arasında en iyi dengeyi sergilemiştir. Segmentasyon tarafında ise FPN ve UNet-

Pretrained yüksek nicel performans gösterirken, DeepLabV3+ sınır koruma ve yapısal tutarlılık açısından üstünlük sağlayarak çok görevli füzyon için özellikle uygun olduğunu kanıtlamıştır.

Anlamsal ve örnek düzeyde algılamayı bütünleştirmek amacıyla iki farklı füzyon stratejisi geliştirilmiştir: (1) segmentasyon çıktısına dayalı hafif, kurallara dayalı bir çıkarım hattı ve (2) DeepLabV3+ ile YOLOv11 mimarilerinin entegre edildiği çift akışlı (dual-stream) bir sinir ağı yapısı. Geliştirilen çift akışlı mimari, yalnızca sezgisel yöntemle elde edilen sonuçlara kıyasla nicel performans açısından değil, aynı zamanda sahne tutarlılığı ve semantik bütünlük bakımından da üstünlük sergilemiştir. Bu yapı, semantik olarak zengin segmentasyon çıktıları ile hassas nesne konumlandırmasını aynı çerçevede birleştirerek daha güvenilir sahne yorumlamaları sunmaktadır. Ayrıca, önerilen yaklaşım gömülü sistemlerde dağıtım için gerekli olan kare hızı (frame rate) gereksinimlerini karşılayabilecek düzeyde çalışabilmektedir.

Sisteme entegre edilen ZED2 stereo kamera, gerçek zamanlı derinlik tahmini ve metrik ölçekli nesne konumlandırması açısından kritik bir yetenek sağlamıştır. ZED SDK aracılığıyla kullanılan eşleşme haritalama (disparity mapping) ve üçgenleme (triangulation) teknikleri sayesinde, sistem çeşitli test senaryoları boyunca ortalama %10'un altında mutlak derinlik hatası (Mean Absolute Error – MAE) ile sonuçlar üretmiştir. Bu doğruluk düzeyi, özellikle ray izleme (track-following), engel tespiti ve kaçınma gibi orta menzilli demiryolu uygulamaları açısından yüksek pratik değere sahiptir.

Bu derinlik tahmini kabiliyeti, sistemin emniyet açısından kritik işlevleri yerine getirebilmesini mümkün kılmaktadır. Bunlar arasında, derinlik tabanlı risk modelleme, yakınlık uyarılarının oluşturulması ve dinamik nesne önceliklendirmesi gibi ileri seviye görevler yer almaktadır. Sonuç olarak, geliştirilen birleşik yapı hem görsel algılama doğruluğunu hem de mekânsal farkındalığı artırarak, akıllı demiryolu izleme sistemleri için bütüncül ve uygulanabilir bir çözüm sunmaktadır.

Bu tezin temel katkıları şu şekilde özetlenebilir:

- Anlamsal segmentasyon, nesne tespiti ve stereo temelli derinlik tahminini birleştiren çok görevli ve gerçek zamanlı bir algılama mimarisinin tasarımı ve uygulanması.
- RailSem19 veri seti üzerinde, altyapıya özgü semantik sınıflar ve zorlayıcı görsel koşullar dikkate alınarak, son teknoloji segmentasyon ve tespit modellerinin kapsamlı biçimde kıyaslanması.
- Semantik açıdan zengin ve mekânsal olarak tutarlı bir algılama sağlayan sezgisel segmentasyon temelli bir iş hattı ile sağlam çift akışlı füzyon mimarisini içeren iki füzyon stratejisinin geliştirilmesi.
- ZED2 platformu aracılığıyla stereo derinlik tahmininin gerçek zamanlı entegrasyonu; bu sayede nesne odaklı 3B konumlandırma ve mesafeye dayalı muhakeme işlevlerinin desteklenmesi.

- CUDA hızlandırmalı çıkarım, modüler iş hattı mimarisi, senkronize model yürütme ve gerçek zamanlı görselleştirme gibi gömülü sistemlere yönelik optimizasyonların gerçekleştirilmesi.

Bu katkılar, demiryolu odaklı bilgisayarlı görü alanında önemli bir ilerlemeyi temsil etmekte ve akıllı ulaşım sistemlerinde kullanılabilecek ölçeklenebilir, sağlam ve gerçek zamanlı algılama sistemleri için sağlam bir temel sunmaktadır.

7.2. Sınırlılıklar

Geliştirilen sistem birçok görev ve operasyonel bağlamda güçlü performans sergilese de, deneyler sırasında bazı sınırlılıklar tespit edilmiştir.

- İlk olarak, RailSem19 veri seti çevresel çeşitliliği sağlasa da, nadir ve emniyet açısından kritik uç durumlar (ör. gece operasyonları, düşük görüş koşulları [sis, kar], sıra dışı nesne sınıfları [vahşi hayvanlar, sahipsiz bagajlar, bakım personeli]) açısından yeterli temsile sahip değildir. Bu nedenle, sistemin bu tür senaryolarda genelleme yeteneği doğrulanmamıştır ve daha geniş veri kümeleri veya sentetik veri üretimi ile bu eksikliklerin giderilmesi gerekmektedir.
- İkinci olarak, ZED2 stereo görüş sistemi genel olarak güvenilir derinlik haritaları üretse de, düşük dokulu yüzeyler, yansıtıcı materyaller ve sert ışık geçişleri gibi özel koşullarda doğruluğu azalmaktadır. Bu durum, eşleşme tahminlerinde gürültüye ve 3B konumlandırma güvenilirliğinde düşüşe yol açmaktadır. Bu etkilerin azaltılması için LiDAR veya time-of-flight sensörleri gibi alternatif ya da yedekleyici sensörlerin entegrasyonu önerilmektedir.
- Üçüncü olarak, modüler ve genişletilebilir olsa da, çift akışlı füzyon mimarisi ciddi hesaplama ve bellek yükü oluşturmaktadır. Her ne kadar çıkarım gecikmesi gerçek zamanlı algılama için kabul edilebilir sınırlar içinde kalsa da (yaklaşık 90 ms/kare), kapalı çevrim uygulamalar veya sınırlı kaynaklara sahip uç cihazlara dağıtım için daha agresif optimizasyonlar gereklidir. Ayrıca, çoklu modelin eşzamanlı çalıştırılması sırasında bellek yönetimi gömülü sistemlerde önemli bir zorluk oluşturmaktadır.
- Son olarak, sistem yalnızca açıklamalı video dizileri ve kontrollü stereo kamera girdileri kullanılarak yarı-simüle bir ortamda değerlendirilmiştir. Gerçek dünya koşullarında bir dağıtım henüz gerçekleştirilmemiştir. Bu nedenle, sistemin mekanik titreşimler, sıcaklık değişimleri, sensör kalibrasyonundaki sapmalar ve çevresel gürültü gibi gerçek operasyonel etkenler altında göstereceği performans henüz tam anlamıyla doğrulanmamıştır. Bu eksikliğin giderilmesi, kapsamlı saha testlerinin yanı sıra uzun vadeli sağlamlık değerlendirmeleri ile mümkündür.

7.3. Son Değerlendirme ve Sonuç

Sonuç olarak, bu tez demiryolu ortamlarında gerçek zamanlı sahne anlayışı için kapsamlı, çok görevli bir derin öğrenme çerçevesi sunmaktadır. Anlamsal segmentasyon, nesne tespiti ve stereo tabanlı mesafe tahmini entegrasyonu yoluyla sistem; engel tespiti, altyapı haritalama ve emniyet açısından kritik risk modelleme gibi zorlukları ele almaktadır. Her ne kadar sensör dayanıklılığı, alan genellemesi ve gömülü dağıtım konularında bazı zorluklar sürse de, bu araştırma sonuçları; dayanıklı, ölçeklenebilir ve akıllı algılama sistemlerinin geliştirilmesi için sağlam bir temel sunmakta ve demiryolu bilgisayarlı görüşünde mevcut literatüre anlamlı katkılar sağlamaktadır. Bu katkılar, otonom ulaşım, akıllı altyapı ve güvenlik odaklı yapay zekâ sistemleri gibi daha geniş hedeflerle de örtüşmektedir.



ÖNERİLER

Bu araştırmanın sunduğu temel katkılar üzerine inşa edilecek bazı gelecek geliştirme alanları şunlardır:

1. Az Örnekli ve Alana Uyumlu Öğrenme:

Az temsil edilen veya yeni senaryolarda genelleme yeteneğini artırmak için, gelecekteki çalışmalar az örnekli öğrenme, alan uyarlama ve sentetik veri üretimi tekniklerine odaklanmalıdır. Meta-öğrenme, özdenetimli ön eğitim ve düşmanca alan aktarımı gibi yöntemler; yeni altyapı türlerine, hava koşullarına ve nadir nesne sınıflarına yönelik hızlı model adaptasyonunu mümkün kılabilir.

2. Tam 3B Sahne Rekonstrüksiyonu:

Bu tez, nesne düzeyinde 3B konumlandırmaya odaklanmış olsa da, gelecek çalışmalar Eşzamanlı Konumlandırma ve Haritalama (Simultaneous Localization and Mapping – SLAM) veya Hareketten Yapılandırma (Structure-from-Motion – SfM) gibi yoğun 3B rekonstrüksiyon yöntemlerini entegre ederek sürekli ortam modellemesi, ray hizalaması ve mekânsal-zamansal muhakeme imkânı sağlayabilir.

3. Çok Modlu Sensör Füzyonu:

LiDAR, GNSS ve IMU gibi ek sensörlerin entegrasyonu, görsel olarak bozulmuş koşullarda sistem dayanıklılığını artıracak ve kritik arıza modlarında yedeklilik sağlayacaktır. Bu sensörlerin stereo görüş çıktılarıyla füzyonu, yüksek doğrulukta çevresel temsil ve küresel konumlandırma sağlayarak karmaşık ve yapısal olmayan ortamlarda daha gelişmiş otonomi sunacaktır.

4. Model Sıkıştırma ve Gömülü Sistem Optimizasyonu:

NVIDIA Jetson Orin veya Xavier NX gibi uç platformlara dağıtım için, budama, niceleme, bilgi damıtma ve dinamik sinir ağları gibi sıkıştırma stratejilerinin araştırılması önerilmektedir. TensorRT ve ONNX Runtime gibi optimizasyon araçları kullanılarak gecikme azaltılabilir, güç tüketimi düşürülebilir ve sınırlı donanım koşullarında model performansı korunabilir.

5. Gerçek Dünya Testleri ve Sistem Doğrulaması:

Operasyonel olgunluğa ulaşmanın son adımı, sistemin trenlere monte edilmiş platformlar, otonom denetim araçları veya rayların üstünden geçebilen İHA'lar gibi gerçek dünya demiryolu senaryolarında dağıtılmasıdır. Bu dağıtımlar, sistemin uçtan uca performansının, kalibrasyon kaymalarına karşı dayanıklılığının ve çevresel etkilere karşı dirençliliğinin değerlendirilmesini mümkün kılacaktır.

KAYNAKLAR

- [1] X. Xie, G. Cheng, J. Wang, X. Yao, and J. Han, "Oriented R-CNN for Object Detection," 2021. Accessed: Jun. 02, 2025. [Online]. Available: <https://github.com/jbwang1997/>
- [2] R. Girshick, "Fast R-CNN," 2015. Accessed: Jun. 02, 2025. [Online]. Available: <https://github.com/rbgirshick/>
- [3] A. Thoke and S. Rai, "Exploring Faster R-CNN Algorithms for Object Detection," *2024 1st International Conference on Software, Systems and Information Technology, SSITCON 2024*, 2024, doi: 10.1109/SSITCON62437.2024.10796389.
- [4] R. Gavrilescu, C. Zet, C. Fosallau, M. Skoczylas, and D. Cotovanu, "Faster R-CNN:an Approach to Real-Time Object Detection," *EPE 2018 - Proceedings of the 2018 10th International Conference and Expositions on Electrical And Power Engineering*, pp. 165–168, Dec. 2018, doi: 10.1109/ICEPE.2018.8559776.
- [5] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Trans Pattern Anal Mach Intell*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: 10.1109/TPAMI.2016.2577031.
- [6] E. Arani, S. Gowda, R. Mukherjee, O. Magdy, S. Kathiresan, and B. Zonooz, "A Comprehensive Study of Real-Time Object Detection Networks Across Multiple Domains: A Survey," Aug. 2022, Accessed: May 15, 2025. [Online]. Available: <https://arxiv.org/pdf/2208.10895>
- [7] P. Sengaphone, J. De Leon, G. Augusto, J. Gonzaga, J. Chua, and L. Gan Lim, "Implementation of Single Shot Multibox Detector (SSD) Algorithm for Object Detection," *ICBIR 2024 - 2024 9th International Conference on Business and Industrial Research, Proceedings*, pp. 702–707, 2024, doi: 10.1109/ICBIR61386.2024.10875914.
- [8] G. More, O. Patil, O. More, M. More, S. Suryavanshi, and M. Mali, "Comparison of Object Detection Algorithms CNN, YOLO and SSD," *Int. J. Sci. R. Tech*, vol. 2024, no. 11, p. 1, Accessed: May 15, 2025. [Online]. Available: www.ijstjournal.com
- [9] Y. Chen, M. C. Goorden, F. J. Beekman, and J. Du, "Understanding of Object Detection Based on CNN Family and YOLO," *J Phys Conf Ser*, vol. 1004, no. 1, p. 012029, Apr. 2018, doi: 10.1088/1742-6596/1004/1/012029.
- [10] T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object detection using YOLO: challenges, architectural successors, datasets and applications," *Multimed Tools Appl*, vol. 82, no. 6, pp. 9243–9275, Mar. 2023, doi: 10.1007/S11042-022-13644-Y/TABLES/7.
- [11] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," pp. 7464–7475, Jul. 2022, doi: 10.1109/cvpr52729.2023.00721.

- [12] R. Varghese and M. Sambath, "YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and Robustness," *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems, ADICS 2024*, 2024, doi: 10.1109/ADICS58448.2024.10533619.
- [13] M. Sevi and İ. Aydın, "Detection of Foreign Objects Around the Railway Line with YOLOv8," *Computer Science*, no. IDAP-2023, pp. 19–23, Oct. 2023, doi: 10.53070/BBD.1346317.
- [14] R. Khanam and M. Hussain, "YOLOv11: An Overview of the Key Architectural Enhancements," Oct. 2024, Accessed: May 15, 2025. [Online]. Available: <https://arxiv.org/pdf/2410.17725>
- [15] M. Al and R. Alif, "YOLOv11 for Vehicle Detection: Advancements, Performance, and Applications in Intelligent Transportation Systems," Oct. 2024, Accessed: Jun. 02, 2025. [Online]. Available: <https://arxiv.org/pdf/2410.22898>
- [16] R. Khanam and M. Hussain, "A Review of YOLOv12: Attention-Based Enhancements vs. Previous Versions," Apr. 2025, Accessed: Jun. 02, 2025. [Online]. Available: <https://arxiv.org/pdf/2504.11995>
- [17] Y. Tian, Q. Ye, and D. Doermann, "YOLOv12: Attention-Centric Real-Time Object Detectors," Feb. 2025, doi: 10.0.
- [18] Z. Zhang, P. Chen, Y. Huang, L. Dai, F. Xu, and H. Hu, "Railway obstacle intrusion warning mechanism integrating YOLO-based detection and risk assessment," *J Ind Inf Integr*, vol. 38, p. 100571, Mar. 2024, doi: 10.1016/J.JII.2024.100571.
- [19] C. Guindel, D. Martin, and J. M. Armingol, "Fast joint object detection and viewpoint estimation for traffic scene understanding," *IEEE Intelligent Transportation Systems Magazine*, vol. 10, no. 4, pp. 74–86, Dec. 2018, doi: 10.1109/MITS.2018.2867526.
- [20] Y. Hong, H. Pan, W. Sun, S. Member, and Y. Jia, "Deep Dual-resolution Networks for Real-time and Accurate Semantic Segmentation of Road Scenes," Jan. 2021, Accessed: May 15, 2025. [Online]. Available: <https://arxiv.org/pdf/2101.06085>
- [21] D. Ristić-Durrant, M. Franke, and K. Michels, "A Review of Vision-Based On-Board Obstacle Detection and Distance Estimation in Railways," *Sensors 2021, Vol. 21, Page 3452*, vol. 21, no. 10, p. 3452, May 2021, doi: 10.3390/S21103452.
- [22] F. U. Rahman, M. T. Ahmed, M. M. Hasan, and N. Jahan, "Real-Time Obstacle Detection Over Railway Track using Deep Neural Networks," *Procedia Comput Sci*, vol. 215, pp. 289–298, Jan. 2022, doi: 10.1016/J.PROCS.2022.12.031.
- [23] A. Jenefa, A. Ande, T. Mounikuttan, M. Anuj, G. Makros, and G. Rejoice, "Real-Time Rail Safety: A Deep Convolutional Neural Network Approach for Obstacle Detection on

- Tracks,” *ICSPC 2023 - 4th International Conference on Signal Processing and Communication*, pp. 101–105, Mar. 2023, doi: 10.1109/ICSPC57692.2023.10125284.
- [24] S. Lee, B. Kim, and H. Lee, “Semantic segmentation models with frozen weights for railway track detection,” *Sci Prog*, vol. 107, no. 4, Oct. 2024, doi: 10.1177/00368504241304204/ASSET/7B9E51D1-59A5-41A5-BF64-15D02840D4F0/ASSETS/IMAGES/LARGE/10.1177_00368504241304204-TABLE6.JPG.
- [25] X. Ye, S. Dong, K. Hu, Y. Zhang, X. Xiong, Y. Zhang, and T. Sheng, “Classification and extraction method of hidden dangers along railway lines based on semantic segmentation network,” *Int J Remote Sens*, Dec. 2024, doi: 10.1080/01431161.2024.2411069; PAGE:STRING:ARTICLE/CHAPTER.
- [26] J. Wang, H. Zhai, Y. Yang, N. Xu, H. Li, and D. Fu, “A Review of Intrusion Detection for Railway Perimeter Using Deep Learning-based Methods,” *IEEE Access*, 2024, doi: 10.1109/ACCESS.2024.3510746.
- [27] J. Long, E. Shelhamer, and T. Darrell, “Fully Convolutional Networks for Semantic Segmentation,” 2015.
- [28] E. Shelhamer, J. Long, and T. Darrell, “Fully Convolutional Networks for Semantic Segmentation,” *IEEE Trans Pattern Anal Mach Intell*, vol. 39, no. 4, pp. 640–651, Apr. 2017, doi: 10.1109/TPAMI.2016.2572683.
- [29] G. Du, X. Cao, J. Liang, X. Chen, and Y. Zhan, “Medical Image Segmentation based on U-Net: A Review.,” *Journal of Imaging Science & Technology*, vol. 64, no. 2, pp. 1–12, Mar. 2020, doi: 10.2352/J.IMAGINGSCI.TECHNOL.2020.64.2.020508.
- [30] K. L. Sailaja, P. R. Kumar, G. Nikhitha, and V. A. Siddhartha, “Detection of Railway Tracks from Satellite images using U-Net Architecture,” *2024 5th International Conference for Emerging Technology, INCET 2024*, 2024, doi: 10.1109/INCET61516.2024.10593407.
- [31] A. Naseer and A. Jalal, “Holistic Scene Recognition through U-Net Semantic Segmentation and CNN,” 2024, Accessed: May 15, 2025. [Online]. Available: <https://www.researchgate.net/publication/385074174>
- [32] M. Sevi and I. Aydin, “Improving Unet segmentation performance using an ensemble model in images containing railway lines”, doi: 10.55730/1300-0632.4014.
- [33] M. Netrakar, D. Upare, C. Tejas, T. Y. Baddi, and R. Benni, “A Comparative Analysis of U-Net, PSPNet, AND FPNet: Deep Learning Techniques for Image Segmentation,” *2024 IEEE Conference on Engineering Informatics, ICEI 2024*, 2024, doi: 10.1109/ICEI64305.2024.10912308.

- [34] X. Zhou, P. Shi, B. Sheil, and S. Suryasentana, "Knowledge-based U-Net and transfer learning for automatic boundary segmentation," *Advanced Engineering Informatics*, vol. 59, p. 102243, Jan. 2024, doi: 10.1016/J.AEI.2023.102243.
- [35] W. Ye, J. Ren, C. Li, W. Liu, Z. Zhang, and C. Lu, "Intelligent Detection of Surface Defects in High-Speed Railway Ballastless Track Based on Self-Attention and Transfer Learning," *Struct Control Health Monit*, vol. 2024, no. 1, p. 2967927, Jan. 2024, doi: 10.1155/2024/2967927.
- [36] C. Zhang, J. Zhao, and Y. Feng, "Research on Semantic Segmentation Based on Improved PSPNet," *Proceedings - 2023 International Conference on Intelligent Perception and Computer Vision, CIPCV 2023*, pp. 1–6, 2023, doi: 10.1109/CIPCV58883.2023.00012.
- [37] Y. Liu, X. Bai, J. Wang, G. Li, J. Li, and Z. Lv, "Image semantic segmentation approach based on DeepLabV3 plus network with an attention mechanism," *Eng Appl Artif Intell*, vol. 127, p. 107260, Jan. 2024, doi: 10.1016/J.ENGAPPAI.2023.107260.
- [38] Y. Weng, J. Yang, C. Zhang, J. He, C. Peng, L. Jia, and H. Xiang, "An improved DeepLabv3 + railway track extraction algorithm based on densely connected and attention mechanisms," *Sci Rep*, vol. 15, no. 1, p. 2556, Dec. 2025, doi: 10.1038/s41598-024-84937-5.
- [39] Q. Zeng, L. Zhang, Y. Wang, X. Luo, and Y. Chen, "DeepLab-Rail: semantic segmentation network for railway scenes based on encoder-decoder structure," <https://doi.org/10.1117/1.JEI.33.4.043038>, vol. 33, no. 4, p. 043038, Aug. 2024, doi: 10.1117/1.JEI.33.4.043038.
- [40] Y. Zheng and M. Cai, "Wheel-rail contact area image segmentation method based on Deeplabv3+," Aug. 2024, doi: 10.21203/RS.3.RS-4792093/V1.
- [41] Z. Ren, L. Wang, T. Song, Y. Li, J. Zhang, and F. Zhao, "Enhancing Road Scene Segmentation with an Optimized DeepLabV3+," *IEEE Access*, 2024, doi: 10.1109/ACCESS.2024.3521597.
- [42] B. Dingyuan, G. Baoqing, R. Tao, Z. Xingfang, S. Tao, W. Yu, and L. Tao, "F2RAIL: panoptic segmentation integrating Fpn and transFormer towards RAILway," *Applied Intelligence*, vol. 55, no. 4, pp. 1–21, Feb. 2025, doi: 10.1007/S10489-024-06158-7/FIGURES/13.
- [43] X. Ni, P. Fieguth, Z. Ma, B. Shi, Y. Qiu, Y. Chen, and H. Liu, "Superpixel-Guided Multi-Type Rail Segmentation via Contextual Information Aggregation," *IEEE Transactions on Intelligent Transportation Systems*, 2024, doi: 10.1109/TITS.2024.3397509.
- [44] L. J. Chen, J. M. Zou, Y. Zhou, G. Z. Hao, and Y. Wang, "Real-Time Semantic Segmentation of Maritime Navigation Scene Based on Improved Fast-SCNN," *Lecture*

- Notes in Electrical Engineering*, vol. 1394, pp. 346–362, 2025, doi: 10.1007/978-981-96-3977-9_38.
- [45] R. P. K. Poudel, S. Liwicki, and R. Cipolla, “Fast-SCNN: Fast Semantic Segmentation Network,” *30th British Machine Vision Conference 2019, BMVC 2019*, Feb. 2019, Accessed: May 15, 2025. [Online]. Available: <https://arxiv.org/pdf/1902.04502>
 - [46] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, “BiSeNet: Bilateral Segmentation Network for Real-time Semantic Segmentation,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11217 LNCS, pp. 334–349, Aug. 2018, doi: 10.1007/978-3-030-01261-8_20.
 - [47] M. Fan, S. Lai, J. Huang, X. Wei, Z. Chai, J. Luo, and X. Wei, “Rethinking BiSeNet for Real-Time Semantic Segmentation,” 2021. Accessed: May 15, 2025. [Online]. Available: <https://github.com/>
 - [48] Q. Xu, Y. Ma, J. Wu, and C. Long, “Faster BiSeNet: A Faster Bilateral Segmentation Network for Real-time Semantic Segmentation,” *Proceedings of the International Joint Conference on Neural Networks*, vol. 2021-July, Jul. 2021, doi: 10.1109/IJCNN52387.2021.9533819.
 - [49] H. Mohammed Taha and G. M Hadi, “Semantic Segmentation for Self-driving Cars,” in *5TH INTERNATIONAL CONFERENCE ON COMMUNICATION ENGINEERING AND COMPUTER SCIENCE (CIC-COCOS’24)*, Cihan University-Erbil, 2024, pp. 99–103. doi: 10.24086/cocos2024/paper.1537.
 - [50] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, “SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers,” *Adv Neural Inf Process Syst*, vol. 34, pp. 12077–12090, Dec. 2021.
 - [51] M. Siam, S. Elkerdawy, M. Jagersand, and S. Yogamani, “Deep Semantic Segmentation for Automated Driving: Taxonomy, Roadmap and Challenges,” *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, vol. 2018-March, pp. 1–8, Jul. 2017, doi: 10.1109/ITSC.2017.8317714.
 - [52] O. Zendel, M. Murschitz, M. Zeilinger, D. Steininger, S. Abbasi, and C. Beleznaï, “RailSem19: A dataset for semantic rail scene understanding,” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, vol. 2019-June, pp. 1221–1229, Jun. 2019, doi: 10.1109/CVPRW.2019.00161.
 - [53] L. Zhang, Q. Yang, X. Liu, and H. Guan, “Rethinking Hard-Parameter Sharing in Multi-Domain Learning,” *Proc (IEEE Int Conf Multimed Expo)*, vol. 2022-July, Jul. 2021, doi: 10.1109/ICME52920.2022.9859706.

- [54] X. Sun, R. Panda, R. Feris, and K. Saenko, “AdaShare: Learning What To Share For Efficient Deep Multi-Task Learning,” *Adv Neural Inf Process Syst*, vol. 2020-December, Nov. 2019, Accessed: May 16, 2025. [Online]. Available: <https://arxiv.org/pdf/1911.12423>
- [55] R. Kawakami, R. Yoshihashi, S. Fukuda, S. You, M. Iida, and T. Naemura, “Cross-connected Networks for Multi-task Learning of Detection and Segmentation,” *Proceedings - International Conference on Image Processing, ICIP*, vol. 2019-September, pp. 3636–3640, May 2018, doi: 10.1109/ICIP.2019.8803687.
- [56] S. Kim, H. Choi, and D. Min, “Sequential Cross Attention Based Multi-task Learning,” Sep. 2022, Accessed: May 17, 2025. [Online]. Available: <https://arxiv.org/pdf/2209.02518>
- [57] S. Kim, T. G. Purdie, and C. McIntosh, “Cross-Task Attention Network: Improving Multi-Task Learning for Medical Imaging Applications,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 14393, pp. 119–128, Sep. 2023, doi: 10.1007/978-3-031-47401-9_12.
- [58] R. Shu and H. Nakayama, “An Empirical Study of Adequate Vision Span for Attention-Based Neural Machine Translation,” *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pp. 1–10, Dec. 2016, doi: 10.18653/v1/w17-3201.
- [59] H. Chai, Z. Yin, Y. Ding, L. Liu, B. Fang, and Q. Liao, “A Model-Agnostic Approach to Mitigate Gradient Interference for Multi-Task Learning,” *IEEE Trans Cybern*, vol. 53, no. 12, pp. 7810–7823, Dec. 2023, doi: 10.1109/TCYB.2022.3223377.
- [60] G. Shi, Q. Li, W. Zhang, J. Chen, and X. M. Wu, “Recon: Reducing Conflicting Gradients from the Root for Multi-Task Learning,” *11th International Conference on Learning Representations, ICLR 2023*, Feb. 2023, Accessed: May 17, 2025. [Online]. Available: <https://arxiv.org/pdf/2302.11289>
- [61] R. Alison, W. Jonathan, and D. Suhartono, “Analyzing the Effects of Combining Gradient Conflict Mitigation Methods in Multi-Task Learning,” *CommIT (Communication and Information Technology) Journal*, vol. 18, no. 1, pp. 99–107, Apr. 2024, doi: 10.21512/COMMIT.V18I1.8905.
- [62] H. Chai, Z. Liu, Y. Tong, Z. Yao, B. Fang, and Q. Liao, “Towards Task-Conflicts Momentum-Calibrated Approach for Multi-task Learning,” in *Proceedings - International Conference on Data Engineering*, IEEE Computer Society, 2024, pp. 939–952. doi: 10.1109/ICDE60146.2024.00077.
- [63] Y. Zhang, H. Xu, J. Fan, M. Qi, T. Liu, and J. Wang, “Joint Semantic Segmentation and Object Detection Based on Relational Mask R-CNN,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 13393 LNCS, pp. 506–521, 2022, doi: 10.1007/978-3-031-13870-6_43;WEBSITE:WEBSITE:DL-SITE;CTYPE:STRING:BOOK.

- [64] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” *IEEE Trans Pattern Anal Mach Intell*, vol. 42, no. 2, pp. 386–397, Mar. 2017, doi: 10.1109/TPAMI.2018.2844175.
- [65] B. Pang, Q. Zhang, M. Chai, and H. Wang, “An Efficient Network for Obstacle Detection in Rail Transit Based on Multi-Task Learning,” *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, pp. 1795–1800, 2023, doi: 10.1109/ITSC57777.2023.10421820.
- [66] H. Pu, G. Wang, T. Song, P. Schonfeld, and H. Zhang, “Multi-Task Deep Learning Methods for Determining Railway Major Technical Standards,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 6, pp. 5904–5918, Jun. 2024, doi: 10.1109/TITS.2023.3338146.
- [67] S. Wang, Y. Li, and S. Feng, “A Multi-task Learning Approach for Weather Classification on Railway Transportation,” *2018 International Conference on Intelligent Rail Transportation, ICIRT 2018*, Jul. 2018, doi: 10.1109/ICIRT.2018.8641680.
- [68] A. M. Selver, E. Ataç, B. Belenlioglu, S. Dogan, and Y. E. Zoral, “Visual and LIDAR Data Processing and Fusion as an Element of Real Time Big Data Analysis for Rail Vehicle Driver Support Systems,” <https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-5225-3176-0.ch003>, pp. 40–66, Jan. 2018, doi: 10.4018/978-1-5225-3176-0.CH003.
- [69] C. Eom, H. Park, and B. Ham, “Temporally Consistent Depth Prediction with Flow-Guided Memory Units,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 11, pp. 4626–4636, Sep. 2019, doi: 10.1109/TITS.2019.2942096.
- [70] C. Hu, H. Liu, B. Huang, L. Huang, Z. Gao, and Y. Zhao, “A Novel Cross-Modal Scene Recognition Algorithm Leveraging Semantic Information,” *IEEE Vehicular Technology Conference*, 2024, doi: 10.1109/VTC2024-SPRING62846.2024.10683497.
- [71] V. A. Fedorov, “Recognizing Railway Infrastructure Using CNN and Stereoscopic Vision,” *Proceedings - 2024 International Russian Smart Industry Conference, SmartIndustryCon 2024*, pp. 13–18, 2024, doi: 10.1109/SMARTINDUSTRYCON61328.2024.10516208.
- [72] R. Khemmar, A. Mauri, C. Dulompont, J. Gajula, V. Vauchey, M. Haddad, and R. Boutteau, “Road and Railway Smart Mobility: A High-Definition Ground Truth Hybrid Dataset,” *Sensors 2022, Vol. 22, Page 3922*, vol. 22, no. 10, p. 3922, May 2022, doi: 10.3390/S22103922.
- [73] A. Broekman and P. J. Gräbe, “RailEnV-PASMVS: A perfectly accurate, synthetic, path-traced dataset featuring a virtual railway environment for multi-view stereopsis training and reconstruction applications,” *Data Brief*, vol. 38, p. 107411, Oct. 2021, doi: 10.1016/J.DIB.2021.107411.

- [74] G. D'Amico, M. Marinoni, F. Nesti, G. Rossolini, G. Buttazzo, S. Sabina, and G. Lauro., "TrainSim: A Railway Simulation Framework for LiDAR and Camera Dataset Generation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 12, pp. 15006–15017, Feb. 2023, doi: 10.1109/TITS.2023.3297728.
- [75] T. Padberg, J. Heikkonen, and R. Kanth, "Study on Stereo AI Based Zed-2i Camera," *Lecture Notes in Networks and Systems*, vol. 933 LNNS, pp. 46–56, 2024, doi: 10.1007/978-3-031-54256-5_5/TABLES/2.
- [76] A. Zhukov, A. Rivero, J. Benois-Pineau, A. Zemmari, and M. Mosbah, "A Hybrid System for Defect Detection on Rail Lines through the Fusion of Object and Context Information," *Sensors 2024, Vol. 24, Page 1171*, vol. 24, no. 4, p. 1171, Feb. 2024, doi: 10.3390/S24041171.
- [77] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information," Feb. 2024, doi: 10.1007/978-3-031-72751-1_1.
- [78] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9351, pp. 234–241, 2015, doi: 10.1007/978-3-319-24574-4_28.
- [79] X. Zhang, H. Bian, Y. Cai, K. Zhang, and H. Li, "An improved tongue image segmentation algorithm based on Deeplabv3+ framework," *IET Image Process*, vol. 16, no. 5, pp. 1473–1485, Apr. 2022, doi: 10.1049/IPR2.12425; WEBSITE:WEBSITE:IETRESEARCH; JOURNAL:JOURNAL:17519667;WGROU:STRING:PUBLICATION.
- [80] L. E. Ortiz, E. V. Cabrera, and L. M. Gonçalves, "Depth data error modeling of the ZED 3D vision sensor from stereolabs," *Electronic Letters on Computer Vision and Image Analysis*, vol. 17, no. 1, pp. 1–15, 2018, doi: 10.5565/REV/ELCVIA.1084.

EKLER

EK- 1: TEMEL TEKNİK TERİMLER SÖZLÜĞÜ

Bu ek, bu tezde sunulan yöntemleri ve katkıları destekleyen temel terimlerin kısa tanımlarını içermektedir. Söz konusu kavramlar Yapay Zekâ, Makine Öğrenmesi ve Bilgisayarla Görü alanlarından kaynaklanmakta olup, demiryolu sahne analizine yönelik önerilen algılama sisteminin yapısını ve işlevini anlamak için gereklidir.

Yapay Zekâ (Artificial Intelligence - AI)

Yapay Zekâ (YZ): İnsan zekâsı gerektiren görevleri yerine getirebilen algoritmalar ve sistemler geliştirmeye odaklanan bir çalışma alanıdır. Bu görevler arasında akıl yürütme, karar verme, görsel algı ve otonom kontrol yer alır. Yapay zekâ, bu tezde modellenen akıllı davranışların kavramsal temelini oluşturmaktadır.

Makine Öğrenmesi (Machine Learning - ML)

Makine Öğrenmesi (MÖ): YZ'nin, verilerden desenleri öğrenebilen ve açıkça programlanmaksızın tahminler ya da kararlar alabilen algoritmalara odaklanan alt alanıdır. Bu çalışmada, etiketlenmiş demiryolu veri kümeleri kullanılarak nesne tanıma ve sahne anlayışı görevleri için denetimli makine öğrenmesi yöntemleri kullanılmaktadır.

Derin Öğrenme (Deep Learning - DL)

Derin Öğrenme (DÖ): Büyük veri kümelerinden otomatik olarak öznitelik çıkarımı yapabilen ve bu öznitelikleri hiyerarşik olarak temsil eden çok katmanlı derin sinir ağlarını kullanan makine öğrenmesinin özel bir alanıdır. Bu tezde, nesne tespiti ve anlamsal segmentasyon gibi görevlerde evrimsel sinir ağları (CNN) gibi derin öğrenme mimarileri kullanılmaktadır.

Bilgisayarla Görü (Computer Vision)

Bilgisayarla Görü: Dijital görüntü ve videolardan anlamlı bilgi çıkarımını mümkün kılan bir YZ alt alanıdır. Görüntü sınıflandırma, nesne tespiti ve sahne yeniden yapılandırma gibi çok çeşitli görevleri kapsamaktadır. Bu tezin temel katkıları, demiryolu ortamlarında gerçek zamanlı bilgisayarla görüş alanında konumlanmaktadır.

Nesne Tespiti (Object Detection)

Nesne Tespiti: Görüntü içerisindeki önceden tanımlanmış nesne sınıflarını tanıma ve konumlandırma sürecidir. Nesne tespiti modelleri, sistemlerin yayalar, araçlar ve demiryolu altyapı elemanları gibi varlıkları tanımasını sağlayan sınıf etiketleri ve sınırlayıcı kutular üretir. Bu

işlevsellik, demiryolu izleme sistemlerinde tehlike tespiti ve sahne yorumlama açısından merkezi önemdedir.

Anlamsal Segmentasyon (Semantic Segmentation)

Anlamsal Segmentasyon: Görüntüdeki her piksele bir anlamsal etiket atanarak sahnenin bileşiminin yoğun (piksel düzeyinde) bir şekilde anlaşılmasını sağlayan sınıflandırma işlemidir. Anlamsal segmentasyon, raylar, peronlar ve sinyaller gibi yapısal ve bağlamsal unsurların belirlenmesini mümkün kılarak sonraki karar alma aşamaları için önemli bağlam sunar.

Mesafe Ölçümü (Distance Measurement)

Mesafe Ölçümü: Kamera ile gözlemlenen bir nesne arasındaki mekânsal mesafeyi belirlemeye yönelik geometrik bir kestirim sürecidir. Doğru mesafe ölçümü, demiryolu uygulamalarında yakınlık temelli risk değerlendirmesi, engel önleme ve gerçek zamanlı güvenlik uyarıları için kritik öneme sahiptir.

Derinlik Tahmini (Depth Estimation)

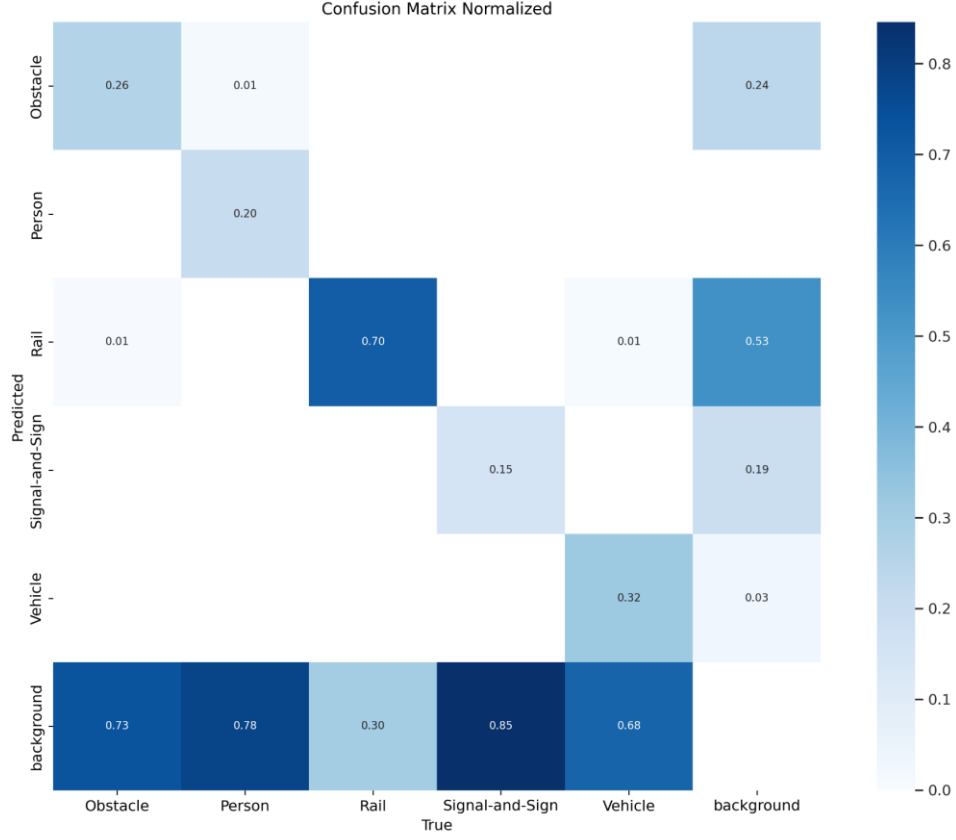
Derinlik Tahmini: Sahnedeki nesnelerin göreceli derinliğini belirlemeye yönelik bir bilgisayarla görü tekniğidir. Bu tezde derinlik tahmini, kalibre edilmiş bir kamera çiftinden elde edilen görüntüler arasındaki farkı (disparite) kullanarak sahne geometrisini çıkaran stereo görüş yöntemiyle gerçekleştirilmiştir.

Stereo Görüş (Stereo Vision)

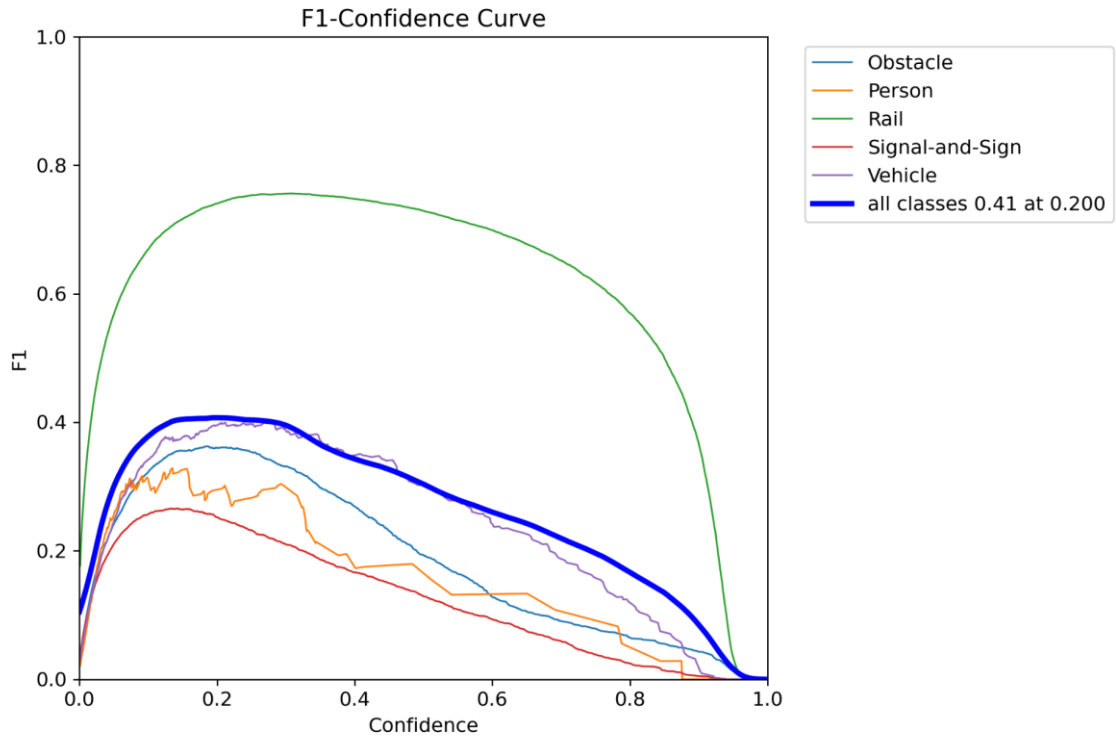
Stereo Görüş (Stereoskopik Görüş): İki yatay olarak yer değiştirilmiş kamera kullanarak insanın çift gözlü görme yetisini taklit eden bir derinlik algılama yaklaşımıdır. Sol ve sağ görüntüler arasındaki disparite hesaplanarak yoğun derinlik haritaları üretilebilir ve algılanan nesnelerin üç boyutlu konumlandırılması sağlanır. Bu çalışmada, gerçek zamanlı derinlik algısı ve mekânsal farkındalık sağlamak amacıyla ZED2 stereo kamera kullanılmaktadır.

EK- 2: NESNE TESPİTİ MODELLERİNE İLİŞKİN EĞİTİM KAYITLARI VE DEĞERLENDİRME ÇIKTILARI.

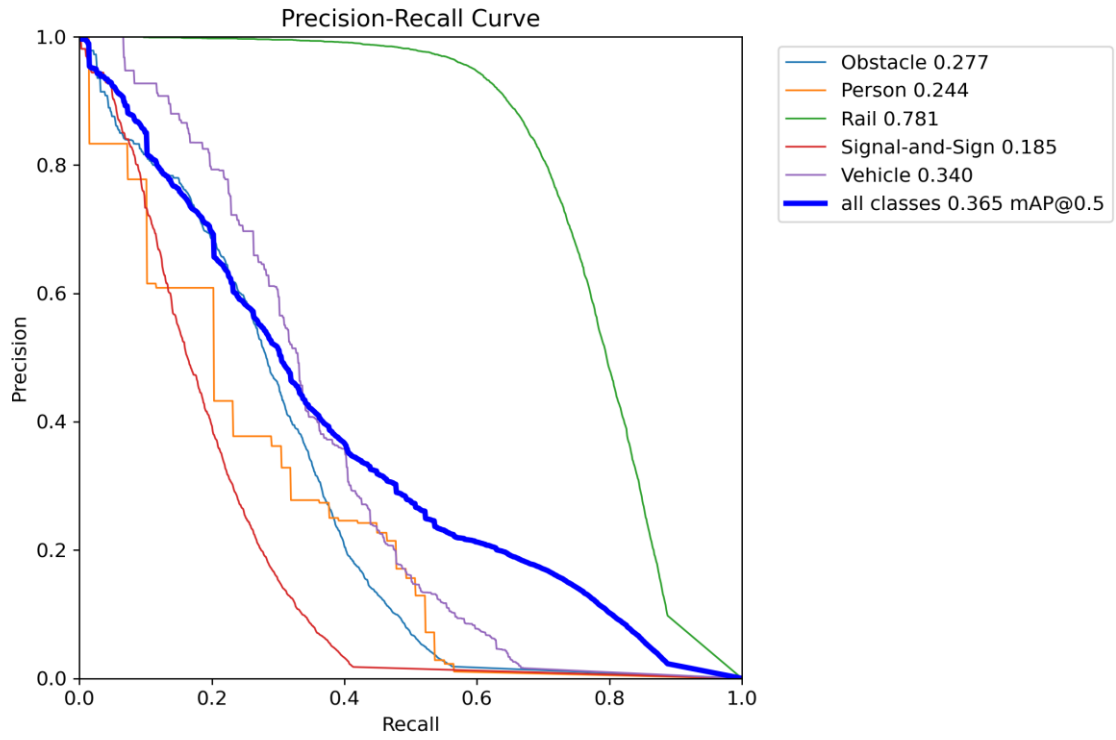
YOLOv12



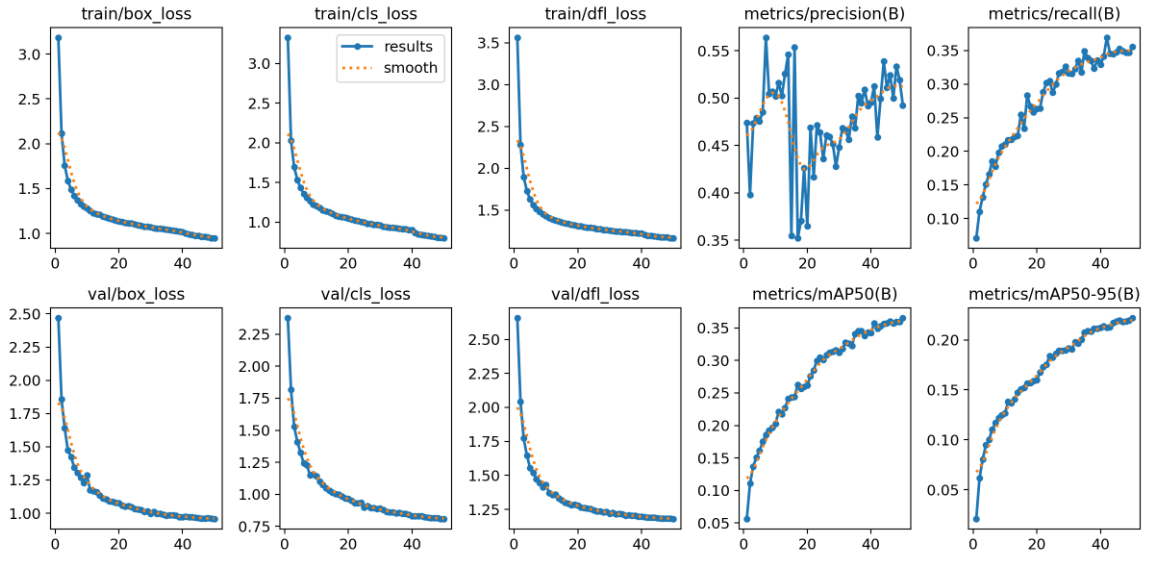
Şekil 7.1. YOLOv12-Normalize edilmiş Karışıklık Matrisi



Şekil 7.2. YOLOv12-F1 Eğrisi

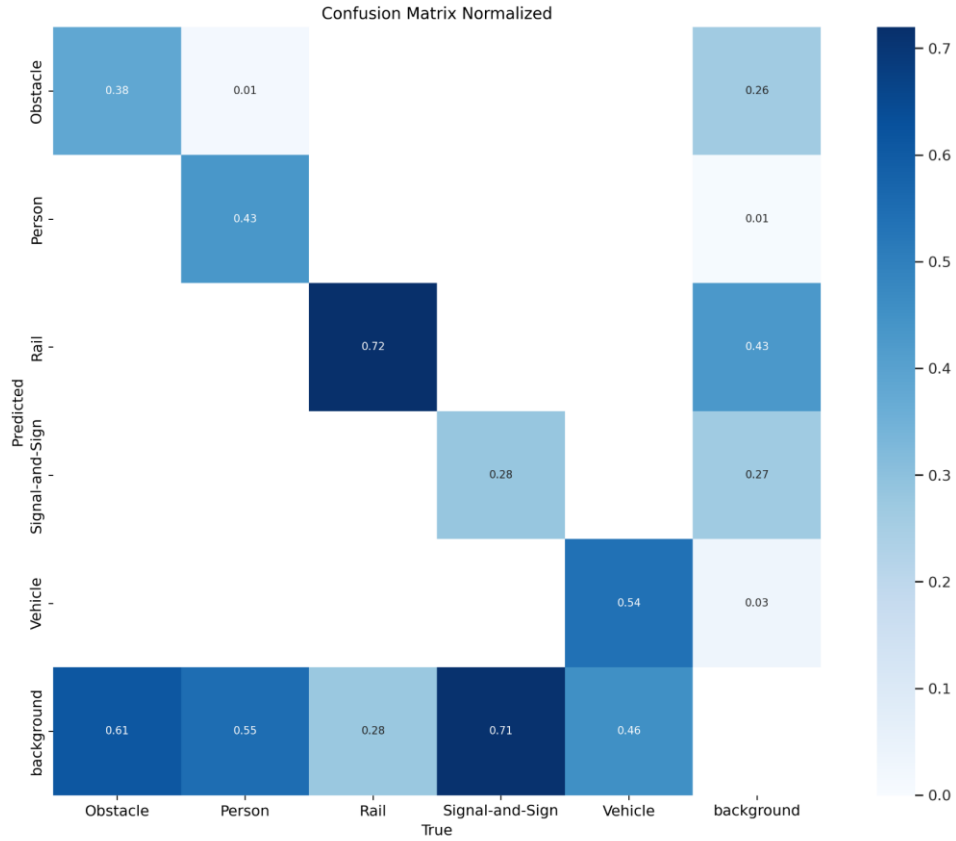


Şekil 7.3. YOLOv12-Kesinlik-Duyarlılık Eğrileri

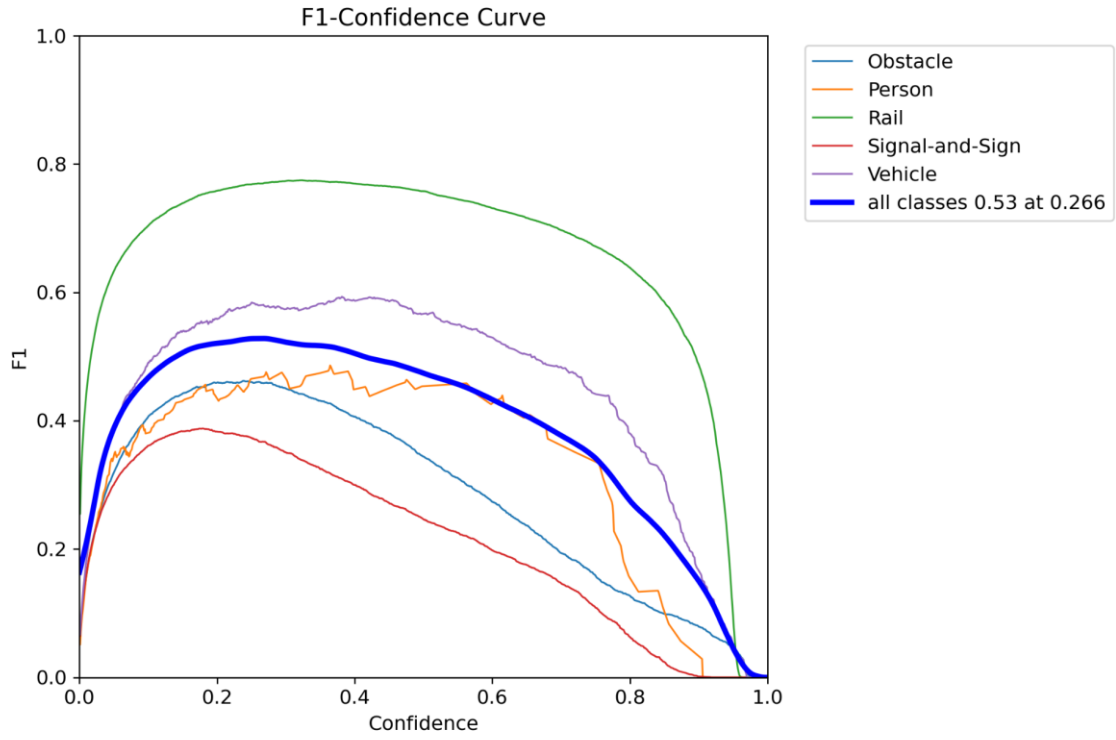


Şekil 7.4. YOLOv12-Sonuçları

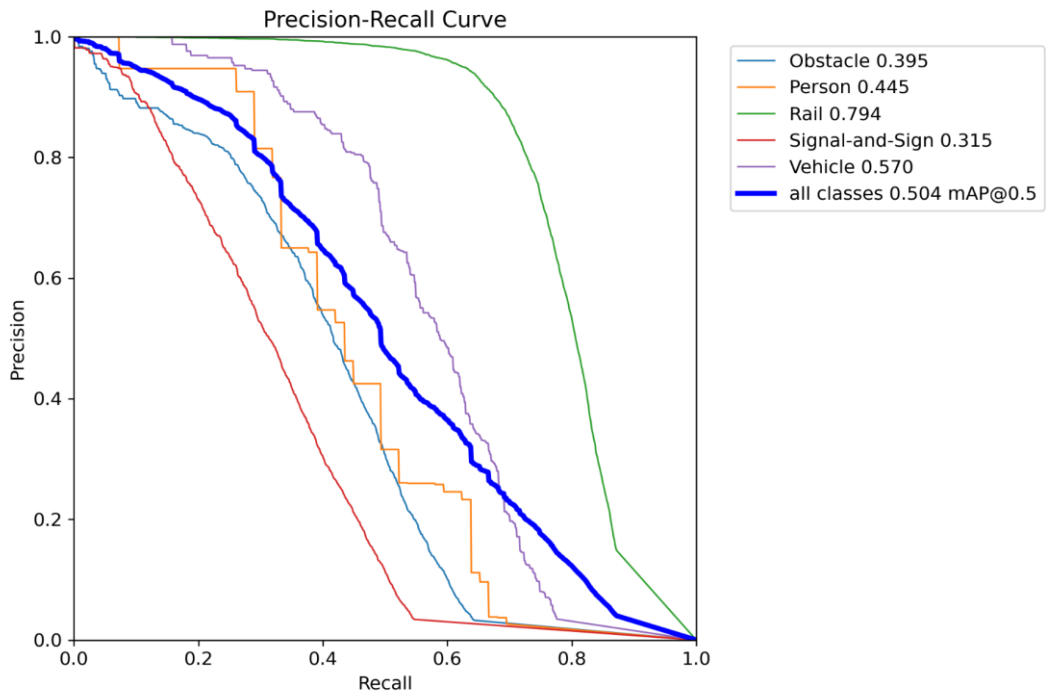
YOLOv11



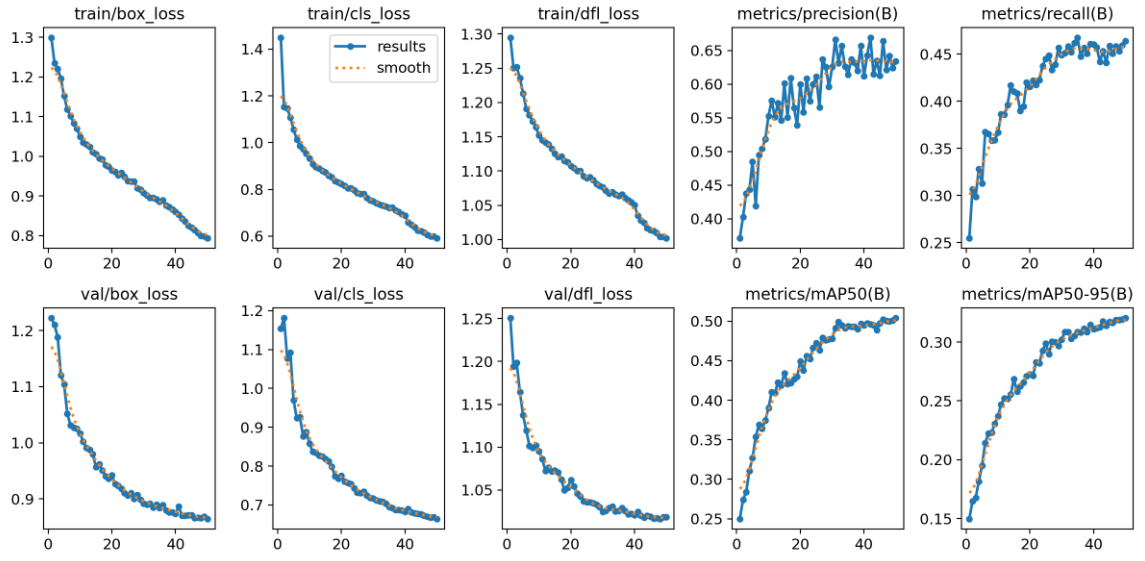
Şekil 7.5. YOLOv11-Normalize edilmiş Karışıklık Matrisi



Şekil 7.6. YOLOv11-F1 Eğrisi

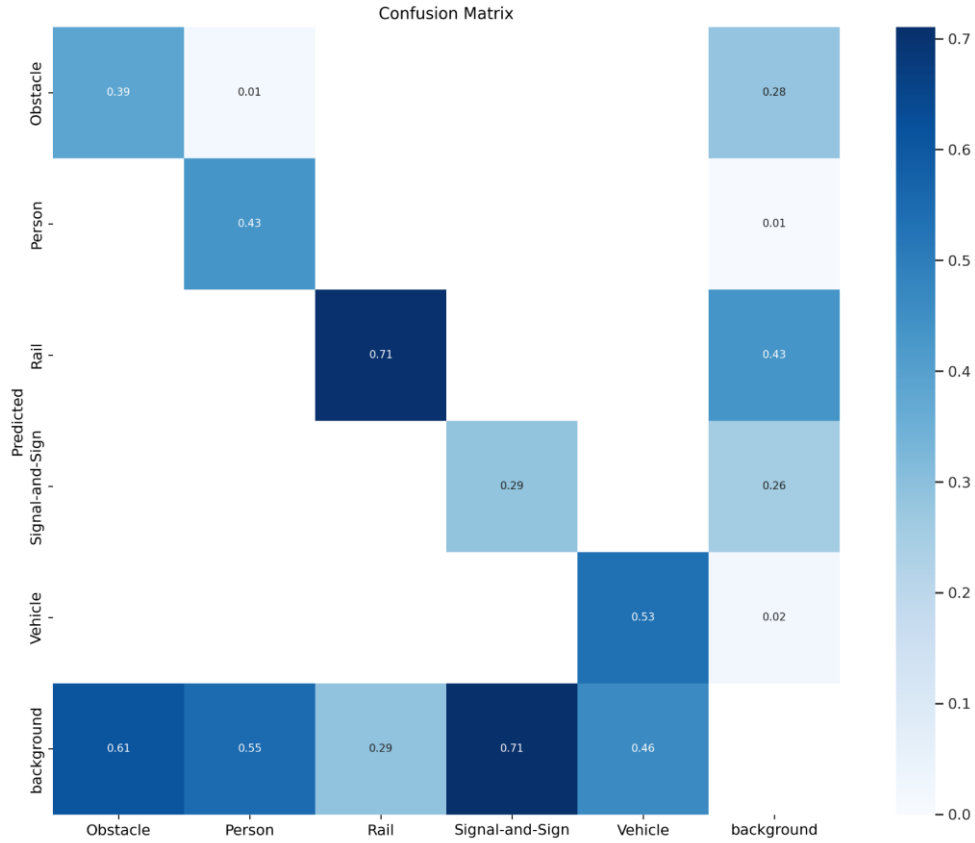


Şekil 7.7. YOLOv11-Kesinlik-Duyarlılık Eğrileri

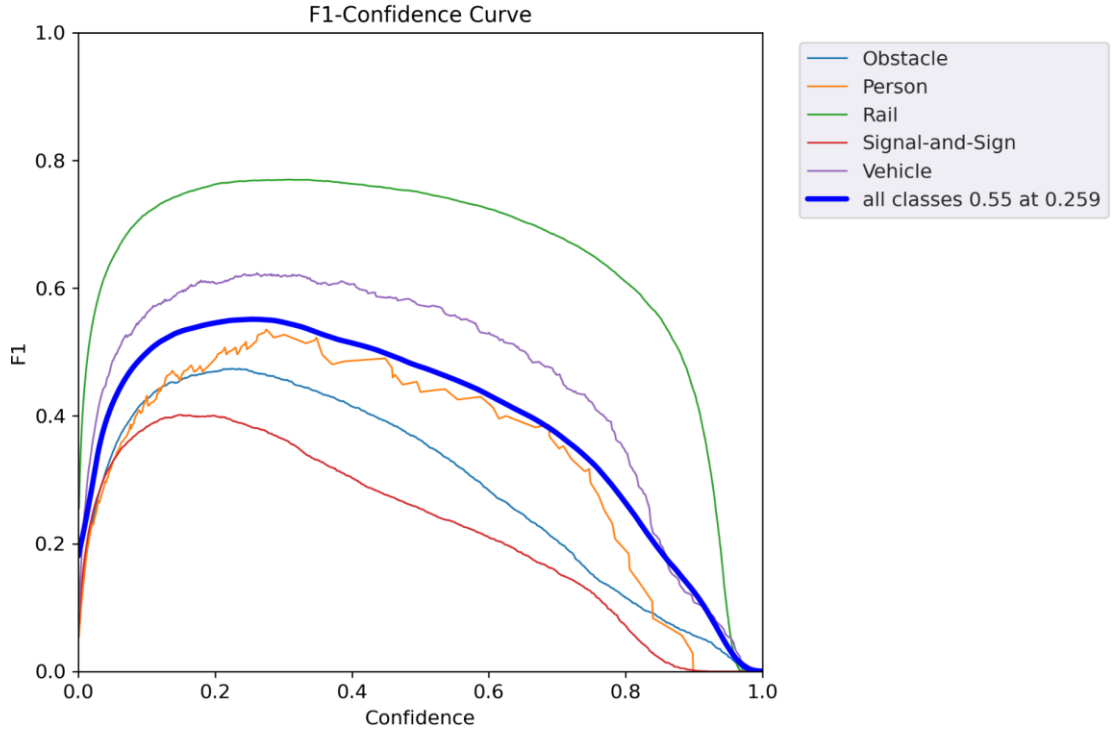


Şekil 7.8. YOLOv11-Sonuçları

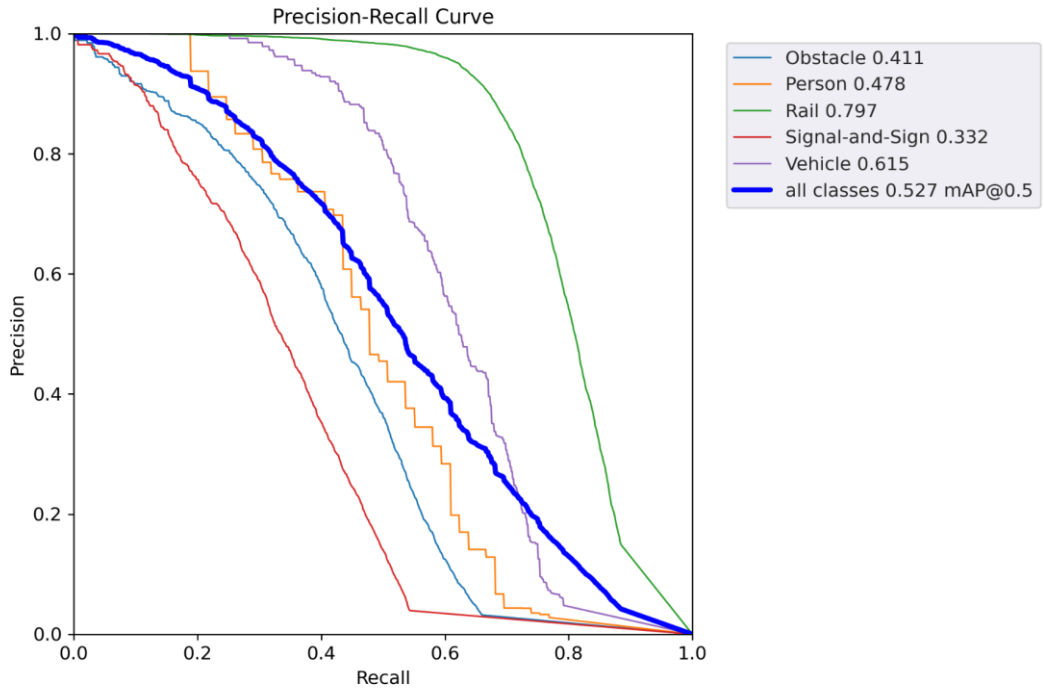
YOLOv9



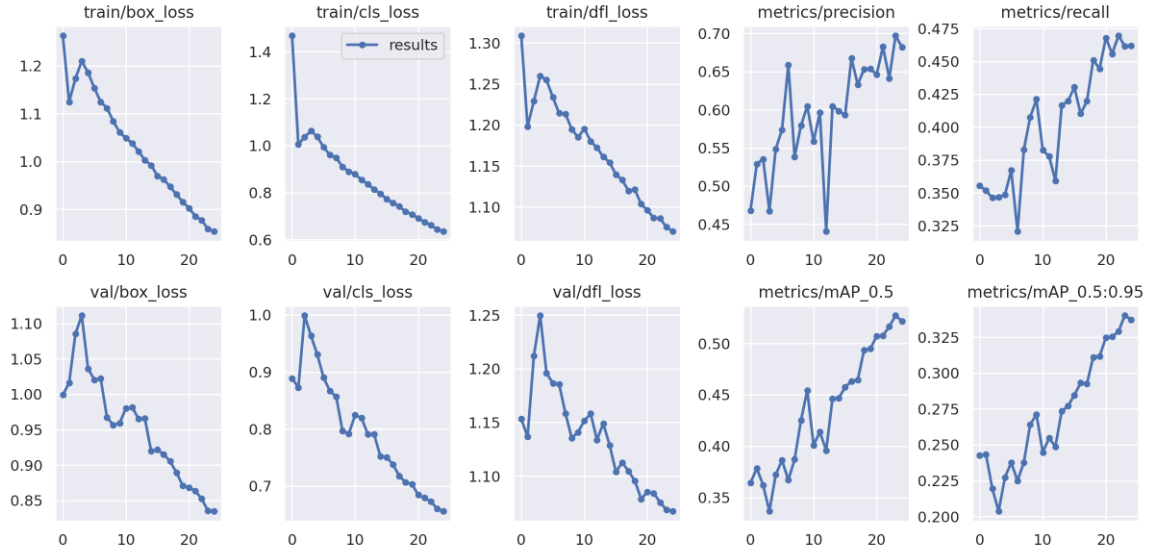
Şekil 7.9. YOLOv9-Normalize edilmiş Karışıklık Matrisi



Şekil 7.10. YOLOv9-F1 Eğrisi

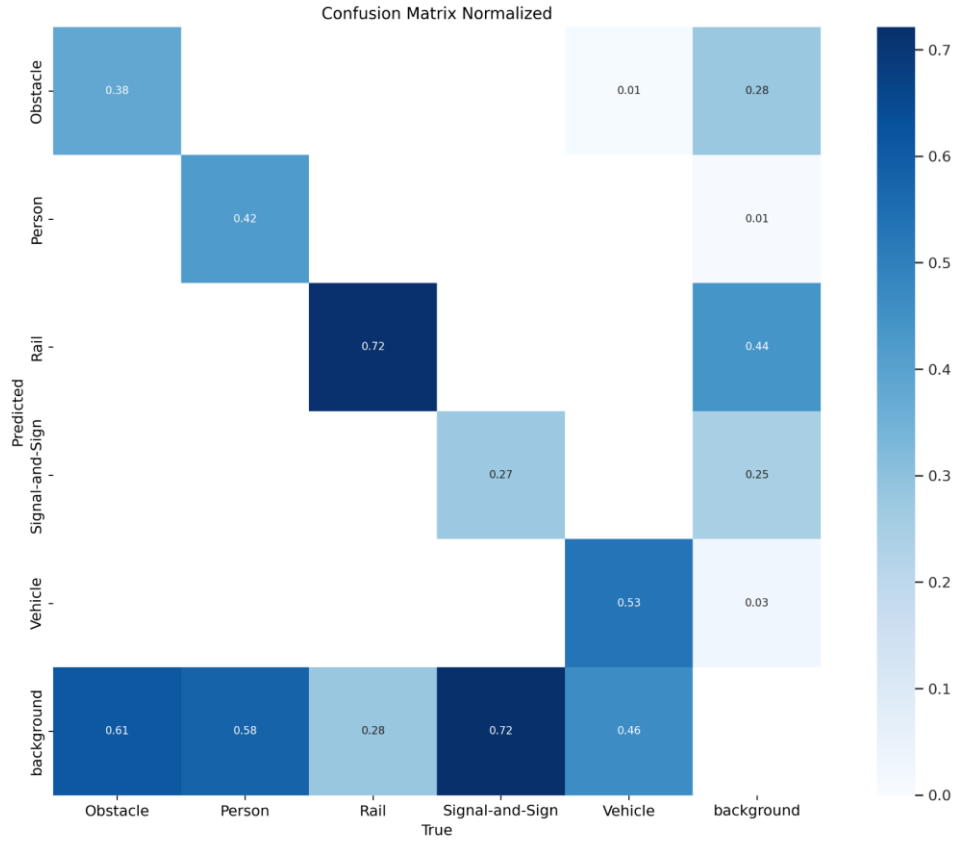


Şekil 7.11. YOLOv9-Kesinlik-Duyarlılık Eğrileri

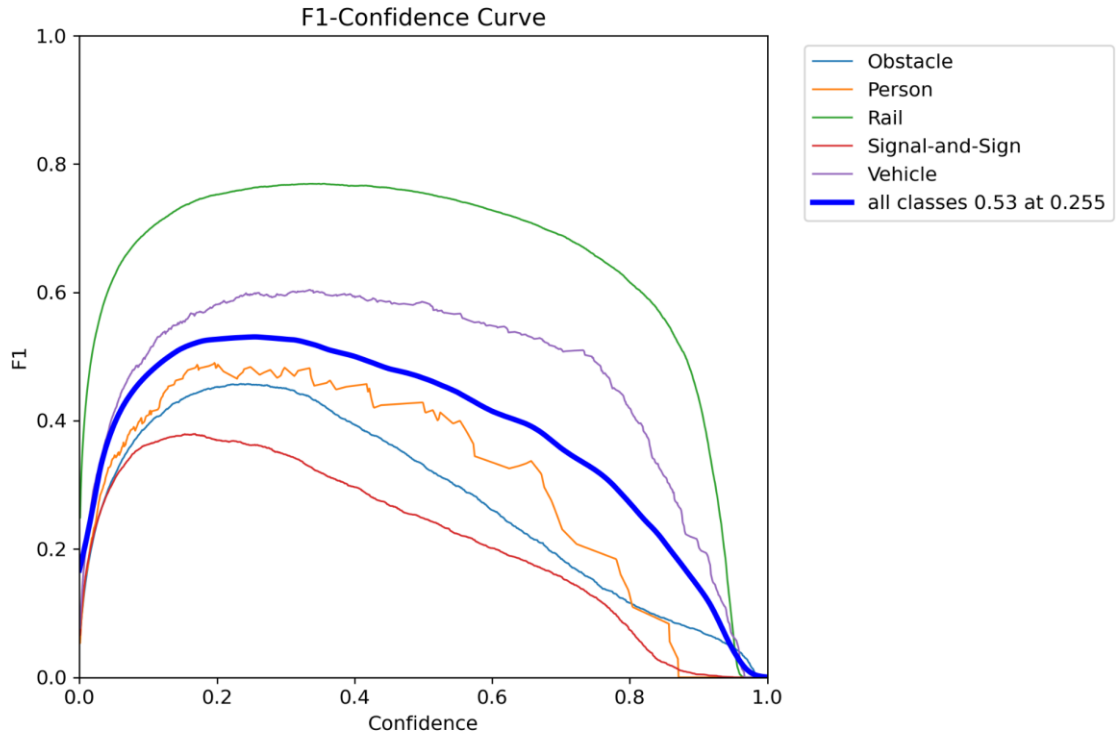


Şekil 7.12. YOLOv9-Sonuçları

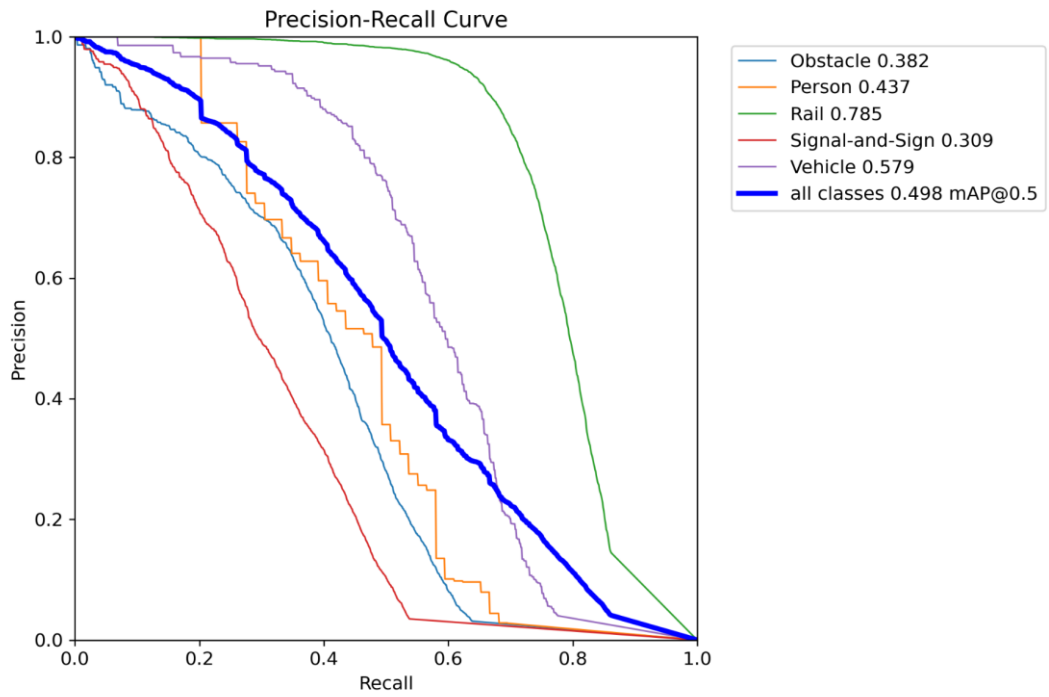
YOLOv8



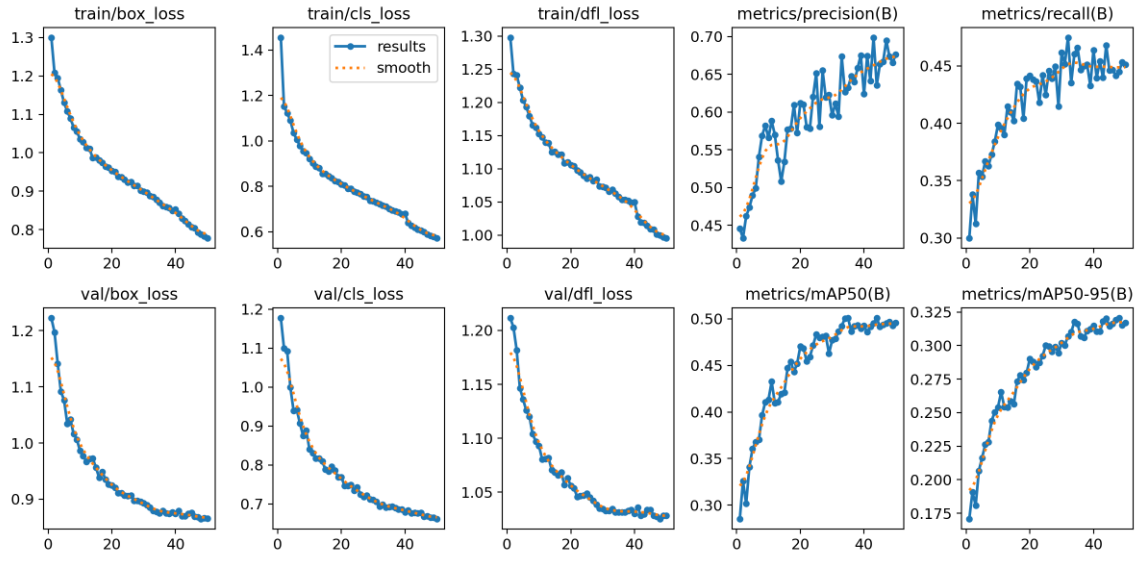
Şekil 7.13. YOLOv8-Normalize edilmiş Karışıklık Matrisi



Şekil 7.14. YOLOv8-F1 Eğrisi



Şekil 7.15. YOLOv8-Kesinlik-Duyarlılık Eğrileri



Şekil 7.16. YOLOv8-Sonuçları

ÖZGEÇMİŞ

Muhammed Amir ELMUHAMMEDCEBBEN

KİŞİSEL BİLGİLER

[Redacted Personal Information]

ARAŞTIRMACI BİLGİLERİ

Öğrenci Orcid ID : 0000-0003-1398-3002
Danışman Orcid ID : 0000-0001-6880-4935

EĞİTİM BİLGİLERİ

[Redacted Education Information]

ARAŞTIRMA DENEYİMİ

- ✓ Programlama Dilleri: Python, Java
- ✓ Teknik Beceriler: Makine Öğrenimi, Derin Öğrenme, Bilgisayarlı Görü
- ✓ Sosyal Beceriler: Araştırma Becerileri, Takım Çalışması, Problem Çözme Becerileri

İŞ DENEYİMİ

[Redacted Work Experience Information]

AKADEMİK FAALİYETLER

Makaleler:

1. M. A. Elmuhammedcebben, İ. Aydın, “Demiryolu Ortamında Nesne Tespiti için Derin Öğrenme Yöntemlerinin Geliştirilmesi ve Zed Kamerası ile Mesafe Ölçümü,” Demiryolu Mühendisliği, sy. 22, Temmuz 2025.