MARMARA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

ÇOK SENARYOLU OLAY GÖRÜNTÜLERİNDE ANOMALİ TESPİTİ İÇİN TRANSFER ÖĞRENME VE TOPLULUK ÖĞRENMESİ YAKLAŞIMI

ZİYA ANIL ŞEN

YÜKSEK LİSANS TEZİ
Bilgisayar Mühendisliği
Anabilim Dalı

Bilgisayar Mühendisliği Yüksek Lisans Programı

DANIŞMAN

Dr. Öğr. Üyesi Ali SARIKAŞ

EŞ-DANIŞMAN

Doç. Dr. Kazım YILDIZ

İSTANBUL, 2025

MARMARA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

ÇOK SENARYOLU OLAY GÖRÜNTÜLERİNDE ANOMALİ TESPİTİ İÇİN TRANSFER ÖĞRENME VE TOPLULUK ÖĞRENMESİ YAKLAŞIMI

ZİYA ANIL ŞEN

523622013

YÜKSEK LİSANS TEZİ
Bilgisayar Mühendisliği
Anabilim Dalı

Bilgisayar Mühendisliği Yüksek Lisans Programı

DANIŞMAN

Dr. Öğr. Üyesi Ali SARIKAŞ

EŞ-DANIŞMAN

Doç. Dr. Kazım YILDIZ

İSTANBUL, 202

ÖNSÖZ

Tez çalışmam süresince desteklerini, rehberliklerini, zamanlarını, yardımlarını ve sabırlarını esirgemeyen değerli hocalarım Dr. Öğr. Üyesi Ali SARIKAŞ'a ve Doç. Dr. Kazım YILDIZ'a, bana olan inançları, koşulsuz sevgileri ve destekleri ile hayatımın her anında yanımda olan çok kıymetli Annem Hülya ŞEN ve Babam Fikret ŞEN'e, hep yanımda olan yanımdaymış hissi veren çok değerli arkadaşlarıma, teşekkürü borç bilirim.

Aralık, 2025 **Ziya Anıl ŞEN**

İÇİNDEKİLER

ÖNSÖZ ..... **iii**

ÖZET ..... **vi**

ABSTRACT ..... **viii**

SEMBOLLER ..... **x**

KISALTMALAR ..... **xi**

ŞEKİL LİSTESİ ..... **xiii**

1 GİRİŞ **1**

1.1 YAPAY ZEKANIN GELİŞİMİ 3

1.2 GEÇMİŞTE YAPILAN ÇALIŞMALAR 3

1.3 ÖNE ÇIKAN ÖZELLİKLER 11

2 MATERYAL VE YÖNTEM **12**

2.1 VERİ KÜMESİ 12

2.1.1 Veri Etiketleme *12*

2.1.2 Veri Arttırma *15*

2.1.3 Veri Çıktıları *16*

2.2 KULLANILAN MODELLER 16

2.2.1 YOLOV8 ..... *17*

2.2.2 U-NET *18*

2.2.3 FPN *20*

2.2.4 DeepLabV3+ *21*

2.2.5 PSPNet *22*

2.3 KULLANILAN OMURGALAR 23

2.3.1 EfficientNet *23*

2.3.2 MobileNet *24*

2.3.3 ResNet *24*

2.4 ÖNERİLEN MODEL 26

2.4.1 Topluluk Öğrenmesi Stratejisi *26*

2.4.2 Model Çeşitliliği ve Mimari Yapı *26*

2.4.3 Tahmin Birleştirme *26*

2.4.4 Topluluk Ağırlıklarının Optimizasyonu *27*

2.5 ANOMALİ TÜRLERİ 27

2.5.1 Çığ *28*

2.5.2 Ev Hırsızlığı *28*
2.5.3 Araç Kazaları *28*
2.5.4 Kavga – Dövüş *29*
2.5.5 Silahlı Saldırılar *29*
2.5.6 Toprak Kayması *30*
2.5.7 Market Soygunları *30*
2.5.8 Hortum *30*
2.5.9 Tsunami *31*
2.5.10 Orman Yangınları *31*
2.6 PERFORMANS METRİKLERİ 32
2.6.1 Doğruluk *32*
2.6.2 Kesinlik *32*
2.6.3 Geri Çağırma *32*
2.6.4 Ortalama Kesinlik Ortalaması *33*
2.6.5 F1 Skoru *33*
2.6.6 Jaccard Katsayısı *33*
2.7 KAYIP FONKSİYONU 34
2.7.1 İkili Çapraz Entropi Kayıp Fonksiyonu *34*
2.7.2 Odaklanmış Kayıp Fonksiyonu *35*
3 BULGULAR VE TARTIŞMA 36
3.1 İLK ÇALIŞMALAR 36
3.2 KAPSAMLI ÇALIŞMALAR 39
3.2.1 Gerçek Değer ve Tahmin Sonuçları *45*
4 SONUÇLAR 56
KAYNAKLAR 58

ÖZET

ÇOK SENARYOLU OLAY GÖRÜNTÜLERİNDE ANOMALİ TESPİTİ İÇİN TRANSFER ÖĞRENME VE TOPLULUK ÖĞRENMESİ YAKLAŞIMI

Bu çalışmada hem kamuya açık veri kümelerinden hem de özel olarak toplanan görüntülerden oluşan hibrit bir veri kümesi kullanılarak kapsamlı bir anomali tespit çerçevesi geliştirilmiştir. Özel veri kümesinin eklenmesi, mevcut kamuya açık veri kümelerindeki eksiklikleri gidermeyi ve gerçek dünya anomali senaryolarını daha doğru bir şekilde temsil etmeyi hedeflemiştir. İki ayrı veri kümesi üzerinden giden bu çalışmada, her veri kümesinden ayrı ayrı sonuçlar alınıp, adım adım gidilmiştir. İki veri kümesi de COCO formatında titizlikle etiketlenmiş ve çok sayıda anomali sınıfını kapsayan çok sınıflı bölütleme yaklaşımına uygun olacak şekilde, her bir görüntü için format dönüşümü ile bölütleme tabanlı eğitim amacıyla maske görüntüleri üretilmiştir.

Çalışmada, 17,000 görselli ve 6 anomali sınıflı olan ilk veri kümesi, beş farklı derin öğrenme mimarisi – YOLOv8, DeepLabV3+, U-Net, Feature Pyramid Network (FPN) ve PSPNet – çok sınıflı anomali tespiti ve bölütleme problemi kapsamında anomali tespiti için kullanılmıştır. Buradan gelen sonuçlara göre, iki farklı mimarinin diğerlerine göre daha etkili olduğu gözlemlenmiştir. Bu tez çalışmasının temel katkısı ise, daha etkili olduğu tespit edilen FPN ve DeepLabV3+ mimarileri üzerinde daha kompleks ve detaylı olan son veri kümesiyle gerçekleştirilen kapsamlı bir ablasyon çalışmasına odaklanmış olmasıdır. Bu ablasyon çalışmasında, her iki mimari için üç farklı omurga (İng. backbone) ağı – ResNet18, EfficientNet-b0 ve MobileNetV2 – ayrı ayrı değerlendirilmiş ve toplamda altı farklı model yapılandırması üzerinde sistematik deneyler gerçekleştirilmiştir.

Genel olarak tüm metrikler birlikte değerlendirildiğinde, DeepLabV3+–ResNet18 mimari–omurga kombinasyonunun (Model 4), çok sınıflı ve hibrit anomali veri kümesi üzerinde en dengeli, istikrarlı ve genellenebilir performansı sunduğu sonucuna ulaşılmıştır. FPN tabanlı modeller ise bazı lokal ve daha az karmaşık anomali türlerinde güçlü sonuçlar üretmiş olsa da genel bölütleme başarımı ve metrikler arası denge açısından DeepLabV3+ mimarisinin gerisinde kalmıştır. Bu bulgular, mimari ve omurga seçiminin anomali türüne duyarlı şekilde yapılmasının önemini vurgulamakta ve geliştirilecek gerçek zamanlı anomali tespit sistemleri

için DeepLabV3+–ResNet18 kombinasyonunu güçlü bir referans yapı olarak önermektedir.

Son olarak, tekil model yapılandırmalarının ötesine geçilerek, model çıktılarının birleştirilmesine dayalı Topluluk öğrenmesi yaklaşımları da değerlendirilmiştir. Bu kapsamda, en iyi performans gösteren modellerin farklı ağırlık başlangıçları ve bağımsız eğitim süreçleriyle elde edilen çıktıları kullanılarak, 15 ve 30 bağımsız modelden oluşan iki ayrı Topluluk öğrenmesi konfigürasyonu oluşturulmuştur. Elde edilen sonuçlar, Topluluk öğrenmesi yaklaşımının özellikle geri çağırma metriğinde belirgin iyileşmeler sağladığını ve kritik anomalilerin kaçırılma oranını azaltmada etkili olduğunu göstermiştir. Buna karşılık, doğruluk ve kesinlik metriklerinde Topluluk öğrenmesi boyutunun artmasıyla birlikte performans kazanımlarının sınırlı kaldığı ve belirli bir noktadan sonra doygunluğa ulaştığı gözlemlenmiştir. Bu bulgular, Topluluk öğrenmesi yöntemlerinin özellikle olay kaçırmanın yüksek maliyetli olduğu senaryolarda güçlü bir alternatif sunduğunu, ancak hesaplama maliyeti–performans dengesi gözetilerek optimal yapılandırmanın seçilmesi gerektiğini ortaya koymaktadır.

Son aşamada, tekil model yapılandırmalarının ötesine geçilerek topluluk öğrenmesi yaklaşımlarının çok sınıflı anomali tespiti üzerindeki etkisi ayrıntılı biçimde incelenmiştir. Bu kapsamda, en iyi performans gösteren model yapılandırmalarının farklı ağırlık başlangıçları ve bağımsız eğitim süreçleriyle elde edilen çıktıları birleştirilerek, 15 ve 30 ağırlıklandırılmış model konfigürasyonundan oluşan iki ayrı topluluk öğrenmesi tipi değerlendirilmiştir. Deneysel sonuçlar, topluluk öğrenmesi modellerinin doğruluk ve kesinlik metriklerinde tekil modellere kıyasla sınırlı veya daha düşük değerler üretmesine karşın, özellikle geri çağırma metriğinde tüm anomali türleri genelinde belirgin ve tutarlı iyileşmeler sağladığını ortaya koymuştur. Özellikle “Çığ”, “Silahlı Saldırılar”, “Toprak Kayması”, “Market Soygunları” ve “Orman Yangınları” gibi kritik ve kaçırılmasının yüksek maliyetli olduğu anomalilerde topluluk öğrenmesi yapılarının en yüksek geri çağırma değerlerine ulaştığı gözlemlenmiştir. Dice ve IoU metrikleri incelendiğinde ise topluluk öğrenmesi yaklaşımlarının, sınırları belirsiz ve zamansal değişkenlik içeren olaylarda göreli kazanımlar sağladığı, ancak genel bölütleme hassasiyeti açısından tekil, yüksek doğruluklu mimarilerin gerisinde kaldığı görülmüştür. Ayrıca 15 ve 30 konfigürasyondan oluşan topluluk öğrenmeleri arasında metrikler açısından belirgin bir fark oluşmaması, boyutun belirli bir noktadan sonra doygunluğa ulaştığını ve sınırlı performans artışı olduğunu göstermektedir. Bu bulgular, topluluk öğrenmesi yaklaşımlarının özellikle olay kaçırmama odaklı uygulamalarda güçlü bir alternatif sunduğunu, ancak gerçek zamanlı ve kaynak kısıtlı sistemlerde dikkatli yapılandırılması gerektiğini ortaya koymaktadır.

ABSTRACT

A TRANSFER LEARNING AND ENSEMBLE LEARNING APPROACH FOR ANOMALY DETECTION IN MULTI-SCENARIO EVENT IMAGES

In this study, a comprehensive anomaly detection framework is developed using a hybrid dataset composed of both publicly available datasets and privately collected images. The inclusion of the custom dataset aims to address the limitations of existing public datasets and to represent real-world anomaly scenarios more accurately. The study follows a two-dataset experimental design, where results are obtained independently from each dataset and analyzed in a step-by-step manner. Both datasets are meticulously annotated in COCO format, and segmentation masks are generated for each image through format conversion to enable segmentation-based training suitable for a multi-class anomaly segmentation approach covering a wide range of anomaly categories.

The first dataset, consisting of 17,000 images and six anomaly classes, is utilized for multi-class anomaly detection and segmentation using five different deep learning architectures—YOLOv8, DeepLabV3+, U-Net, Feature Pyramid Network (FPN), and PSPNet. Based on the results obtained from this stage, two architectures are observed to be more effective than the others. The primary contribution of this thesis is its focus on a comprehensive ablation study conducted on the more effective FPN and DeepLabV3+ architectures using a more complex and detailed final dataset. In this ablation study, three different backbone networks—ResNet18, EfficientNet-b0, and MobileNetV2—are independently evaluated for each architecture, resulting in systematic experiments across a total of six different model configurations.

When all evaluation metrics are considered jointly, the DeepLabV3+–ResNet18 architecture–backbone combination (Model 4) is found to provide the most balanced, stable, and generalizable performance on the multi-class hybrid anomaly dataset. Although FPN-based models produce strong results for certain localized and less complex anomaly types, they fall behind the DeepLabV3+ architecture in terms of overall segmentation performance and inter-metric balance. These findings emphasize the importance of selecting architecture and backbone combinations in an anomaly-sensitive manner and suggest the DeepLabV3+–

ResNet18 configuration as a strong reference model for real-time anomaly detection systems.

Finally, beyond single-model configurations, ensemble learning approaches based on the aggregation of model outputs are also evaluated. In this context, two different ensemble learning configurations composed of 15 and 30 independent models are constructed using outputs obtained from the best-performing models with different weight initializations and independent training processes. The results indicate that ensemble learning provides significant improvements particularly in recall, effectively reducing the rate of missed detections for critical anomalies. In contrast, performance gains in accuracy and precision remain limited as the ensemble size increases, eventually reaching a saturation point. These findings suggest that ensemble learning methods offer a strong alternative for scenarios where missing critical events is highly costly, while highlighting the necessity of selecting an optimal configuration by carefully balancing computational cost and performance.

In the final stage, the impact of ensemble learning approaches on multi-class anomaly detection is examined in greater detail. Two different ensemble types, composed of 15 and 30 weighted model configurations, are evaluated by aggregating outputs obtained from independent training processes with different weight initializations. Experimental results demonstrate that, although ensemble models yield limited or lower accuracy and precision compared to single models, they achieve consistent and significant improvements in recall across all anomaly types. Notably, ensemble configurations attain the highest recall values for critical anomalies such as avalanches, armed attacks, landslides, market robberies, and wildfires. Analysis of Dice and IoU metrics reveals that ensemble approaches provide relative gains in events characterized by ambiguous boundaries and temporal variability, while remaining less precise than high-accuracy single architectures in terms of overall segmentation quality. Furthermore, the absence of a significant performance difference between the 15- and 30-model ensemble configurations indicates a saturation effect, suggesting diminishing returns despite increased computational cost. These results demonstrate that ensemble learning approaches constitute a powerful alternative for recall-oriented anomaly detection applications, provided that careful configuration is applied in real-time and resource-constrained systems.

SEMBOLLER

° **:** Derece

~ **:** Değer Aralığı

$\boldsymbol{\alpha}$ **:** Sınıf Ağırlıklandırma Parametresi

$\boldsymbol{\gamma}$ **:** Odaklanma Parametresi

$\boldsymbol{\sigma}$ **:** Sigmoid fonksiyonu

KISALTMALAR

ACC : Accuracy

AP : Average Precision

ASPP : Atrous Spatial Pyramid Pooling

CAM : Class Activation Map

CNN : Convolutional Neural Networks

COCO : Common Objects In Context

CRF : Conditional Random Field

DCNN : Deep Convolutional Neural Networks

FN : False Negative

FP : False Positive

FPN : Feature Pyramid Network

IoU : Intersection over Union

mAP : Mean Average Precision

MIL : Multiple Instance Learning

PPM : Pyramid Pooling Module

PRE : Precision

PSPNet : Pyramid Scene Parsing Network

RE : Recall

ResNet : Residual Neural Network

TN **:** True Negative

TP **:** True Positive

YOLO **:** You Only Look Once

ŞEKİL LİSTESİ

Şekil 2.1. Etiketlenmiş anomali görselleri 15

Şekil 2.2. Roboflow Etiketleme Programı [40] 15

Şekil 2.3. Mimari – Omurga Benchmark Yapısı 17

Şekil 2.4. YOLOv8 Mimarisi [45] 18

Şekil 2.5. U-Net Mimarisi [46] 19

Şekil 2.6. FPN Mimarisi [53] 21

Şekil 2.7. PSPNET Mimarisi [58] 23

Şekil 2.8. Çığ Orijinal Görselleri 28

Şekil 2.9. Ev Hırsızlığı Orijinal Görselleri 28

Şekil 2.10. Araç Kazaları Orijinal Görselleri 29

Şekil 2.11. Kavga – Dövüş Orijinal Görselleri 29

Şekil 2.12. Silahlı Saldırı Orijinal Görselleri 29

Şekil 2.13. Toprak Kayması Orijinal Görselleri 30

Şekil 2.14. Market Soygunları Orijinal Görselleri 30

Şekil 2.15. Hortum Orijinal Görselleri 31

Şekil 2.16. Tsunami Orijinal Görselleri 31

Şekil 2.17. Orman Yangınları Orijinal Görselleri 31

Şekil 3.1. Modellerin eğitim kaybı sonuçları 45

Şekil 3.2. Çığ olayına ait gerçek (solda) ve tahmin (sağda) maskelerinin orijinal görüntülerinin üzerine bindirilmiş görüntüleri 46

Şekil 3.3. Soygun olayına ait gerçek (solda) ve tahmin (sağda) maskelerinin orijinal

görüntülerinin üzerine bindirilmiş görüntüleri 47

Şekil 3.4. Araç Kazaları olayına ait gerçek (solda) ve tahmin (sağda) maskelerinin orijinal görüntülerinin üzerine bindirilmiş görüntüleri 48

Şekil 3.5. Kavga - Dövüş olayına ait gerçek (solda) ve tahmin (sağda) maskelerinin orijinal görüntülerinin üzerine bindirilmiş görüntüleri 49

Şekil 3.6. Silah olayına ait gerçek (solda) ve tahmin (sağda) maskelerinin orijinal görüntülerinin üzerine bindirilmiş görüntüleri 50

Şekil 3.7. Toprak Kayması olayına ait gerçek (solda) ve tahmin (sağda) maskelerinin orijinal görüntülerinin üzerine bindirilmiş görüntüleri 51

Şekil 3.8. Soygun olayına ait gerçek (solda) ve tahmin (sağda) maskelerinin orijinal görüntülerinin üzerine bindirilmiş görüntüleri 52

Şekil 3.9. Hortum olayına ait gerçek (solda) ve tahmin (sağda) maskelerinin orijinal görüntülerinin üzerine bindirilmiş görüntüleri 53

Şekil 3.10. Tsunami olayına ait gerçek (solda) ve tahmin (sağda) maskelerinin orijinal görüntülerinin üzerine bindirilmiş görüntüleri 54

Şekil 3.11. Orman Yangını olayına ait gerçek (solda) ve tahmin (sağda) maskelerinin orijinal görüntülerinin üzerine bindirilmiş görüntüleri 55

TABLO LİSTESİ

Tablo 2.1. Anomali görüntü sayıları 12

Tablo 2.2. Veri Kümesi Detayları 13

Tablo 2.3. Veri arttırma işlemleri 16

Tablo 2.4. Omurgaların Parametre Sayıları 25

Tablo 2.5. Derin Öğrenme Mimarilerinde Kullanılan Omurgaların Varyasyonları ve Parametre Sayıları (M) 25

Tablo 3.1. İlk Veri Kümesinde En İyi Doğruluk Değerleri 36

Tablo 3.2. İlk Veri Kümesinde En İyi Kesinlik Değerleri 37

Tablo 3.3. İlk Veri Kümesinde En İyi Geri Çağırma Değerleri 38

Tablo 3.4. İlk Veri Kümesinde En İyi F1 Skor Değerleri 38

Tablo 3.5. Model Tanımları 39

Tablo 3.6. Hibrit Veri Kümesinin Doğruluk Değerleri 40

Tablo 3.7. Hibrit Veri Kümesinin Kesinlik Değerleri 41

Tablo 3.8. Hibrit Veri Kümesinin Geri Çağırma Değerleri 42

Tablo 3.9. Hibrit Veri Kümesinin Dice Katsayıları 43

Tablo 3.10. Hibrit Veri Kümesinin IoU Değerleri 44

1 GİRİŞ

Bu çalışmanın temel amacı, "anomali" olarak tanımlanan herhangi bir olayın—ki bu terim; adli ve kriminal vakaları (örneğin can ve mal güvenliğine yönelik saldırılar, şiddet içerikli olaylar, hırsızlık girişimleri, izinsiz girişler gibi) olduğu kadar, orman yangını gibi doğal afetleri de kapsamaktadır—gerçekleştiği anda mümkün olan en yüksek doğruluk oranıyla hızlı ve güvenilir biçimde tespit edilmesidir. Günümüzde hem şehir güvenliği hem çevresel izleme hem de akıllı ulaşım sistemleri gibi geniş bir yelpazede kullanılan video tabanlı gözetim sistemleri, insanlar ve çevre için potansiyel risk oluşturan olayların anlık olarak tespit edilmesine ihtiyaç duymaktadır [1-3]. Bu nedenle anomali tespiti konusu yalnızca akademik çevrelerin değil, güvenlik şirketleri, belediyeler, afet yönetim kurumları ve hatta bireysel kullanıcılar dahil olmak üzere çok geniş bir kitlenin ilgi odağı hâline gelmiştir.

Her ne kadar literatürde bu konuya ilişkin çok sayıda araştırma yapılmış ve yöntem önerilmiş olsa da mevcut çalışmaların önemli bir bölümü belirli türdeki anomali olaylarına odaklanmakta veya sınırlı kapsamlı veri kümeleri kullanmaktadır [4-5]. Bu durum, farklı anomali senaryolarının aynı sistem içerisinde doğru ve tutarlı bir şekilde tespit edilmesini güçleştirmektedir. Bu bağlamda, bu çalışmayı önceki araştırmalardan ayıran en önemli özellik, kapsamının genişliği, metodolojik derinliği ve ele aldığı anomali kategorilerinin çeşitliliğidir. Çalışmanın çok yönlü bir tespit kabiliyeti geliştirme hedefi, doğal olarak veri çeşitliliğini ve sistem performans analizinin karmaşıklığını da artırmaktadır.

Anomali kavramının çok boyutlu yapısı, tek bir yöntem ya da mimari ile tüm olay türlerinin doğru şekilde temsil edilmesini zorlaştırmaktadır. Örneğin şiddet içeren bir olayın görsel dinamikleri, bir doğal afetin ya da teknik bir arızanın görsel özelliklerinden tamamen farklıdır. Bu nedenle çalışmada kullanılan veri kümesinin kapsamlı ve çeşitli olması kritik bir önem taşımaktadır [6]. Bu amaçla oluşturulan veri kümesi, yalnızca Kaggle [7] gibi platformlardan elde edilen mevcut açık kaynak veri kümelerini değil, aynı zamanda harici kaynaklardan toplanan, farklı çevresel koşulları, kamera açılarını, ışık seviyelerini ve sahne bağlamlarını içeren geniş bir görsel materyal havuzunu da kapsamaktadır. Bu çeşitlilik, gerçek hayatta karşılaşılabilecek senaryoların daha gerçekçi bir şekilde temsil edilmesine olanak sağlamaktadır. Aynı zamanda veri kümesinin zenginliği, eğitilen modellerin genelleme kapasitesinin artmasına önemli ölçüde katkıda bulunmaktadır.

Bu çalışmanın dikkat çeken bir diğer yönü, kapsamlı ve çok boyutlu karşılaştırmalı analiz sürecidir. Literatürdeki birçok çalışmanın aksine, burada tek bir modelin başarımını incelemekle sınırlı kalınmamış, farklı mimari yapılara sahip geniş bir model yelpazesi sistematik olarak değerlendirilmiştir. Bu değerlendirme süreci, farklı derin öğrenme mimarilerinin çeşitli anomali türleri karşısındaki performanslarını karşılaştırma imkânı sunmakta ve her bir mimarinin güçlü ve zayıf yönlerinin somut olarak ortaya konmasına yardımcı olmaktadır. Bu yönüyle çalışma, yalnızca sonuç odaklı bir performans raporu değil, aynı zamanda gelecekte yapılacak araştırmalar için yol gösterici bir teknik çerçeve sunmaktadır.

Buna ek olarak, çalışmada yapılan detaylı literatür incelemesi, alanda daha önce uygulanan yöntemleri, kullanılan veri kümelerini, deneysel protokolleri ve karşılaşılan temel problemleri kapsamlı bir biçimde ele almaktadır. Bu inceleme sayesinde mevcut literatürde yer alan eksiklikler, sınırlılıklar ve metodolojik karşıtlıklar belirlenmiş; böylelikle bu araştırmanın katkılarının hangi noktada yenilik sunduğu daha net bir şekilde ortaya konmuştur.

Bu araştırmayı önceki çalışmalardan belirgin şekilde ayrıştıran temel unsur, sahip olduğu kapsamlı, çok sınıflı ve ince detaylara dayalı benchmark sürecidir. Çalışmada ele alınan anomali sınıflarının sayısının fazlalığı, tek bir modele bağımlı kalınmaması ve çeşitli omurga (İng. backbone) yapılarının aynı veri kümesi üzerinde adil kriterlerle değerlendirilmesi, araştırmanın özgünlük katsayısını artırmaktadır. Bu yaklaşım, hem geniş kapsamlı hem de derinlikli bir analiz gerektiren anomali tespiti gibi kompleks bir probleme daha bütüncül bir çözüm sunmaktadır.

Sonuç olarak, bu çalışma; kapsamlı veri kümesi oluşturma stratejisi, çoklu model mimarileri arasında yapılan geniş ölçekli karşılaştırmalı analizleri ve deneysel süreç boyunca korunan metodolojik titizliğiyle literatüre önemli katkılar sağlamaktadır. Bu nitelikler bir araya geldiğinde çalışma, yalnızca mevcut sistemlerin geliştirilmesine değil, aynı zamanda gelecekte daha esnek, daha hızlı, daha güvenilir ve gerçek zamanlı anomali tespit sistemlerinin tasarlanmasına da ışık tutmaktadır. Bu nedenle araştırma, hem akademik hem de uygulamalı alanlarda önemli bir referans niteliği taşımaktadır.

1.1 Yapay Zekanın Gelişimi

Yapay zekâ; insanların kendi kontrolünde yürüttükleri işlemleri, bilgisayar sistemleri tarafından taklit edilmesini hedefleyen geniş bir araştırma alanıdır. Yapay zekanın temel amacı; insanların sahip olduğu öğrenme, problem çözme, akıl yürütme, karar verme ve adaptasyon gibi yeteneklerin geliştirilmesini sağlamaktır. Günümüzde yapay zekâ; akademik araştırmalardan sanayi uygulamalarına, günlük yaşama entegre dijital hizmetlerden kritik güvenlik teknolojilerine kadar geniş bir yelpazede etkisini göstermekte ve dijital dönüşümün temel itici gücü hâline gelmektedir.

"Yapay zekâ" kavramının kökenleri 20. Yüzyılın ortalarına uzanmaktadır. Bu dönemde bilgisayar biliminin henüz yeni şekillendiği yıllarda araştırmacılar, makinelerin sadece hesaplamadan ibaret olmayıp, daha gelişmiş düşünebilme yetisine sahip olan araçlar olabileceği ihtimali üzerinde durmaya başlamıştır. Bu ihtimale cevap verebilen ilk örneklerden biri, Alan Turing'ın 1950 yılında ortaya koyduğu "Taklit Oyunu" (Turing Testi) dur. Bu çalışma, yalnızca yapay zekâ kavramının teorik temelini oluşturmakla kalmamış; aynı zamanda insan–makine etkileşimi üzerine yapılan araştırmaların yönünü de belirlemiştir [8].

1.2 Geçmişte Yapılan Çalışmalar

Öngün ve arkadaşlarının çalışmasında, kalabalık ortamlardan elde edilen gözetim görüntülerindeki anormal olayların insan davranışları içerisindeki tespitine odaklanılmaktadır. Bu amaçla, video karelerinin yoğunluk haritaları Sonlu-Zaman Lyapunov Üsleri (FTLE) yöntemi kullanılarak oluşturulmakta ve ardından kalabalık hareketi analiz edilerek davranışlar sınıflandırılmaktadır. Ancak, bu yaklaşımın gerçek zamanlı tespit sağlamadığı özellikle vurgulanmaktadır [9].

Ke ve ekibi, arka plan karmaşıklığı ile tıkanıklığın önemli zorluklar oluşturduğu kalabalık ve dinamik sahnelerde gerçek dünya insan eylemlerinin tespiti için bir yöntem önermektedir. Yazarlar, nesne alma, el sallama veya düğmeye basma gibi 110 gerçek olaydan oluşan yaklaşık 20 dakikalık video görüntüleri toplamış ve metodolojilerini bu veriler üzerinde değerlendirmiştir. Önerilen yaklaşım, etkileşimli bölütleme araçları ile tek bir eğitim örneğinden uzamsal-zamansal olay şablonlarının oluşturulmasını ve bu şablonların parçalanarak şekil ve akış tanımlayıcılarının birleşimiyle video dizileriyle eşleştirilmesini içermektedir. Model, tıkanıklığa ve aktör değişkenliğine karşı sağlamlık

göstermekte ve literatürdeki tüm kare akış tutarlılığına dayalı yöntemlerden daha iyi performans sergilemektedir. Kesinlik–geri çağırma analizi, özellikle karmaşık sahnelerde önerilen parça tabanlı yöntemin üstünlüğünü doğrulamaktadır. Bununla birlikte, sistemin en önemli sınırlılığı, olay sınıfı başına tek bir örneğe dayanmasıdır. Gelecekteki çalışmaların, çoklu örneklem modellemesi ve otomatik parça seçimi yoluyla genelleme kapasitesini artırmayı hedeflediği belirtilmektedir [10].

Zhou ve ekibinin çalışmasında, AnomalyNet adı verilen birleşik bir derin öğrenme çerçevesi önerilmektedir. Bu çerçeve, özellik öğrenme, seyrek temsil ve sözlük öğrenimini uyumlu bir sinir ağı mimarisinde birleştirmektedir. Model; hareket ve görünüm özellikleri arasındaki uzamsal-zamansal tutarlılığı koruyan bir hareket füzyon bloğu, ayırt edici özellik öğrenimini güçlendiren bir özellik transfer bloğu ve adaptif ISTA algoritmasını yeniden formüle eden yenilikçi bir LSTM tabanlı kodlama bloğundan oluşmaktadır. Bu yeni LSTM tasarımı, klasik optimizasyon yöntemleri ile sinir ağlarını birleştirerek diferansiyel programlamaya özgün bir bakış açısı sunmaktadır. Gerçekleştirilen kapsamlı deneyler, AnomalyNet'in anormal olay tespiti ve görüntü yeniden yapılandırma görevlerinde, özellikle video gözetimi bağlamında, son teknoloji yöntemlerle rekabet edebilecek veya onları aşabilecek düzeyde performans sergilediğini göstermektedir [11].

Medioni ve arkadaşları, düşük seviyeli görsel verileri yüksek seviyeli sembolik senaryolarla sistematik biçimde ilişkilendiren bir senaryo modelleme yöntemi önermektedir. Model; görüntü özellikleri, hareketli nesne özellikleri ve senaryolardan oluşan üç katmanlı bir soyutlama yapısına dayanmaktadır. Bu yapı basit görünmekle birlikte, tıkanıklıklar, eksik tespitler ve sahne karmaşıklığı gibi gerçek dünya koşulları nedeniyle uygulamada zorluklar ortaya çıkmaktadır. Bu nedenle sistem; tespit, izleme ve davranış çıkarımı modüllerini entegre etmekte ve kısa süreli izleme kayıplarını telafi edecek mekanizmalar içermektedir. Ayrıca, daha karmaşık aktivitelerin tanınmasına imkân tanıyan sembolik akıl yürütme süreçlerini de desteklemektedir. Devam eden çalışmaların, dinamik ortamlarda doğruluk ve çıkarım güvenilirliğini artırmak amacıyla belirsizlik modelleme, stabilizasyon teknikleri ve güven tahmini üzerinde yoğunlaştığı belirtilmektedir [12].

Al-Gayyim tarafından yapılan çalışmada, ev ortamında toplanan insan aktivite verileri kullanılarak anomali tespiti gerçekleştirilmektedir. Tespit modeli, Geri Beslemeli Sinir Ağı (FFNN) ile Parçacık Sürü Optimizasyonu (PSO) kombinasyonu kullanılarak geliştirilmiştir. Elde edilen sonuçlar, modelin %97,95 doğruluk oranı ve 0.016 mutlak hata ile oldukça

başarılı performans sergilediğini göstermektedir [13].

Taheri Tajar ve ekibinin çalışmasında, karşılaştırmalı analizlerde de kullanılan YOLO mimarisi temel alınmaktadır. Tiny-YOLOv3 modelinden tanıma performansına katkısı düşük olan belirli tam bağlantılı ve evrişimsel katmanlar çıkarılarak hafif bir sinir ağı elde edilmiştir. Ayrıca ardışık evrişimsel katmanlar birleştirilmekte ve basit arka planlara sahip görüntülerde toplu normalleştirmenin atlanabileceği gösterilmektedir. Modifiye edilen ağ, orijinal modele benzer doğrulukta tanıma performansı sunarken, özellikle NVIDIA Jetson Nano üzerinde TensorRT kullanılarak 22 FPS'ye ulaşılmakta ve gerçek zamanlı çalışabilirlik korunmaktadır. Bu sonuçlar, optimize edilmiş sığ ağların kaynak kısıtlı cihazlarda verimli gerçek zamanlı nesne tespiti için uygun bir çözüm sunduğunu göstermektedir [14].

Boyrazli ve arkadaşları, çeşitli video veri kümeleri üzerinde anormal insan davranışlarının tespiti için bir model geliştirmektedir. Normal ve anormal senaryolar ayrı ayrı analiz edilmekte; beş video üzerinden hesaplanan metriklerin ortalaması yaklaşık %70 doğruluk, %70 duyarlılık, %93 kesinlik ve %77 F1-skoru olarak raporlanmaktadır [15].

Xiang ve ekibi, davranış tanıma ve anomali tespiti için tamamen denetimsiz bir çerçeve önermektedir. Bu çerçeve, Dinamik Bayes Ağları karışımı, spektral kümeleme ve çevrimiçi olabilirlik oranı testini bir arada kullanmaktadır. Yöntem, yeni veya belirsiz davranışları tespit etmede denetimli yaklaşımlardan daha başarılı sonuçlar sunmakta ve farklı gözetim senaryolarında etkili olduğunu göstermektedir. Bununla birlikte, kalabalık sahnelerin işlenmesi ve değişen görsel bağlamlara uyum sağlama konuları sınırlılıklar arasında yer almakta olup, gelecekteki çalışmaların uyarlanabilir öğrenme ve gelişmiş olay tespiti üzerine odaklanacağı belirtilmektedir [16].

Pennisi ve ekibi tarafından önerilen model, Görsel Özellik Çıkarımı ve Görüntü Bölütleme tekniklerinin birleştirilmesiyle oluşturulmakta ve dört farklı video veri kümesi üzerinde değerlendirilmiştir. Deneysel sonuçlar, modelin ilgili veri kümelerinde sırasıyla %90, %86, %87 ve %87 doğruluk oranlarına ulaştığını göstermektedir [17].

Sultani ve arkadaşları, zayıf etiketli veri kullanarak anomali tespiti için derin çoklu örnek öğrenme (MIL) tabanlı bir çerçeve önermektedir. Video düzeyinde etiketlenen veriler kullanılarak, modelin anormal segmentlere daha yüksek anomali puanı vermesi sağlanmakta; aynı zamanda seyreklik ve zamansal düzgünlük zorunlu kılınmaktadır. Ayrıca

çalışma, 13 farklı anomali türünü içeren 1.900 düzenlenmemiş gözetim videosundan oluşan büyük ölçekli bir veri kümesi sunmaktadır. Deneyler, önerilen yaklaşımın mevcut yöntemlerden daha yüksek performans sergilediğini göstermektedir [18].

Sabokrou ve ekibinin bir çalışmasında, kalabalık sahnelerde anormal davranışların algılanması için tam evrişimsel sinir ağları (FCN) kullanılmaktadır. Model gözetimsiz bir yapıya dönüştürülerek hem anomali tespiti hem de lokalizasyon gerçekleştirilebilmektedir. Şekil ve zaman boyutlu verilerin birlikte işlenmesi sayesinde hız ve doğruluk açısından güçlü sonuçlar elde edilmektedir [19].

Aynı araştırma grubunun diğer çalışmasında videolar, örtüşmeyen "cubic patch" birimlerine bölünmekte; her patch için hem yerel hem küresel betimleyiciler çıkarılmaktadır. Bu betimleyiciler gözetimsiz biçimde kullanılarak gerçek zamanlı anomali tespiti ve lokalizasyon gerçekleştirilmektedir. Yöntem, özellikle zaman çizgisi boyunca anormal davranışların anlık algılanmasında etkili sonuçlar sunmaktadır [20].

Zhang ve ekibi, çok kalabalık sahnelerde bireysel izleme yerine kısa dönemli optik akışlardan yararlanarak "trajektori şekil betimleyicileri (shape contexts)" oluşturmaktadır. K-NN tabanlı benzerlik ölçümleri ile oluşturulan istatistiksel model sayesinde uzamsal-zamansal alanda anomali tespiti ve lokalizasyon yapılmaktadır. Bu yaklaşım, perspektif bozulmalarını ve kalabalık kaynaklı karışıklıkları minimize etmektedir [21].

Lv ve arkadaşları tarafından önerilen yöntemde, yoğun kalabalık sahnelerde birey takibi yapılmamakta; bunun yerine video kareleri grid yapısına bölünmekte ve her grid için bir temsilci parçacık seçilmektedir. Parçacıklar arasındaki itici kuvvetlerden bir "force flow matrisi" oluşturulmakta, sparse coding ve görsel sözlük teknikleriyle yeniden yapılandırma yapılmakta ve yeniden yapılandırma hatalarından anomali tespiti gerçekleştirilmektedir. Bu yöntem, takip yapılmasının zor veya imkânsız olduğu durumlarda oldukça etkili sonuçlar sunmaktadır [22].

Doung ve arkadaşları tarafından hazırlanan derleme makalesi, video gözetiminde anomali tespiti için kullanılan derin öğrenme tabanlı yöntemleri sistematik biçimde incelemektedir. Öznitelik çıkarımı, ön işleme, öğrenme paradigması (süpervize, yarı-süpervize, gözetimsiz) ve kullanılan veri kümeleri gibi çeşitli boyutlarda kapsamlı bir sınıflandırma yapılmakta; ayrıca mevcut zorluklar ve gelecekteki araştırma yönelimleri tartışılmaktadır [23].

Ji-Woon Lee ve ekibi tarafından önerilen üç aşamalı derin öğrenme mimarisi, gözetim videolarında anomalilerin tespiti için tasarlanmaktadır. İlk aşamada ön-eğitimli bir CNN ile karelerden öznitelikler çıkarılmakta, ikinci aşamada BiLSTM ve multi-head attention bileşenleriyle zamansal ilişkiler modellenmekte, son aşamada ise anomali karar mekanizması çalışmaktadır. Bu mimari hem uzamsal hem zamansal bilgiyi kullanarak daha doğru ve güvenilir tespit sağlamaktadır [24].

Qaraqe ve ekibinin güncel çalışmasında, kalabalık davranış analizinde yalnızca bireysel eylemler değil, aynı zamanda toplu davranış, kalabalık yoğunluğu ve potansiyel tehlike düzeyi de modellenmektedir. Özellikle büyük kalabalıklar ve karmaşık sahneler için geçerli olan yaklaşım, anomali tespitinin ötesine geçerek topluluk düzeyinde risk analizine odaklanmaktadır [25].

Jebur ve arkadaşları tarafından yapılan başka bir çalışmada, farklı anomaliler için ayrı ayrı yeniden eğitim gerektirmeyen, genel amaçlı bir anomali tespit modeli önerilmektedir. Bu başarım, transfer öğrenme, model birleşimi (model fusion) ve çoklu görev öğrenme (multi-task learning) tekniklerinin bir arada kullanılmasıyla sağlanmaktadır. Böylece sistem, saldırı, hırsızlık veya vandalizm gibi farklı anomalileri tek bir modelle tanıyabilmekte ve yeniden eğitim ihtiyacı azalmaktadır [26].

Zhang ve ekibinin çalışmasında, değişken zaman ölçeklerinde (kısa, orta, uzun dönem) video bilgisini yakalayabilen bir model önerilmektedir. Uzamsal-zamansal tutarlılığın korunması yoluyla daha rafine anomali tespiti yapılmakta ve yapılan deneyler, yöntemin yaygın benchmark veri kümeleri üzerinde önceki yöntemleri geride bıraktığını göstermektedir [27].

Chidananda ve ekibi, CCTV ve gözetim videolarında anomali tespiti için VidAnomalyNet adında özel bir derin öğrenme mimarisi önermektedir. CNN tabanlı bu mimari; yangın, kaza ve soygun gibi çeşitli anomali türlerini sınıflandırabilmekte ve MobileNetV1 gibi ön-eğitimli ağlarla birlikte kullanıldığında yaklaşık %96 doğruluğa ulaşmaktadır [28].

Sharif ve arkadaşlarının yaptığı çalışma, derin öğrenme tabanlı kalabalık anomali tespiti alanındaki yöntemleri sistematik olarak incelemektedir. Kullanılan veri kümeleri, metodolojik yaklaşımlar, performanslar ve gerçek zamanlılık, genelleme, çeşitlilik ve etiketleme gibi hâlen çözüm bekleyen problemler detaylı şekilde tartışılmaktadır. Bu çalışma, literatürdeki ilerlemeyi ve mevcut boşlukları ortaya koyan önemli bir kaynak

niteliği taşımaktadır [29].

Liu ve ekibi tarafından yapılan çalışmada, video anomali tespiti problemi yeniden inşa temelli yaklaşımlar yerine gelecek kare tahmini çerçevesinde ele alınmaktadır. Bu yaklaşımda, normal davranışları temsil eden bir derin ağ, bir sonraki video karesini mümkün olan en iyi şekilde tahmin edecek biçimde eğitilmekte; tahmin edilen kare ile gerçek kare arasındaki fark, anomali skoruna dönüştürülmektedir. Model, yalnızca görünüş (yoğunluk, gradyan) hatasını değil, aynı zamanda optik akış üzerinden zamansal tutarlılık kısıtlarını da içermekte ve böylece hareket bilgisini de gözetmektedir. Deneysel bulgular, tahmin hatasının normal örnekler için düşük, anormal olaylar için ise belirgin biçimde yüksek gerçekleştiğini ve yöntemin önceki pek çok taban çizgiye kıyasla daha duyarlı bir anomali tespiti sunduğunu göstermektedir [30].

Cho ve arkadaşlarının çalışmasında, gözetimsiz video anomali tespiti için normalleştirici akış (normalizing flow) ve örtük iki-yollu otoenkoder (ITAE) bileşenlerini birleştiren bir çerçeve önerilmektedir. İki ayrı kodlayıcı, görünüş ve hareket özelliklerini açıkça ayırmak yerine örtük biçimde modellemekte; tek bir kod çözücü bu bilgileri birleştirerek normal sahnelerin dağılımını öğrenmektedir. Normal örneklerin karmaşık dağılımını daha doğru modelleyebilmek için, ITAE tarafından üretilen özellikler üzerinde normalleştirici akışa dayalı bir yoğunluk modeli kurulmakta ve olasılık (likelihood) temelli anomali skoru elde edilmektedir. Bu sayede, etiketlenmemiş gözetim videoları üzerinden yalnızca normal davranışları öğrenen sistem, eğitim dağılımının dışında kalan örnekleri anomali olarak işaretleyebilmektedir [31].

Huang ve ekibinin önerdiği çerçevede, video anomali tespiti için optik akış yeniden inşası ile silinmiş kare (erased frame) tahmini ortak bir gözetimsiz mimaride birleştirilmektedir. Model, bir yandan normal hareket örüntülerini yansıtan optik akış alanlarını yeniden yapılandırmakta, diğer yandan bazı bölgeleri silinmiş kareleri doğru biçimde tahmin etmeye çalışmaktadır. Bu iki görev hem düşük seviyeli hareket bilgisi hem de daha üst seviye yapısal bağlamın aynı anda öğrenilmesini sağlamaktadır. Deneyler, özellikle karmaşık sahnelerde, bu birleşik kayıp tasarımının yalnızca tek bir yeniden inşa veya tahmin görevi kullanan modellere kıyasla daha yüksek doğruluk ve daha kararlı çakışma (trade-off) eğrileri ürettiğini ortaya koymaktadır [32].

Khan ve arkadaşlarının çalışmasında, kalabalık sahnelerde çerçeve tabanlı anomali tespiti

için ince ayarlı bir AlexNet mimarisi kullanılmaktadır. Önerilen derin ağ, izleme senaryolarından elde edilen görüntü kareleri üzerinde eğitilmekte; birden fazla evrişim katmanı, havuzlama ve toplu normalleştirme adımı ile ayırt edici kalabalık desenleri öğrenilmektedir. Bu özellikler, normal ve anormal kalabalık davranışını (koşma, panik, dağılma vb.) ayırt etmeye yönelik bir sınıflandırıcıya aktarılmakta ve kare düzeyinde anomali etiketleri üretilmektedir. Çalışmada raporlanan sonuçlar, modelin çeşitli kalabalık veri kümelerinde yüksek doğruluk ve tatmin edici genelleme performansı sunduğunu göstermektedir [33].

Sardar Waqar Khan ve ekibi ise, trafik gözetim videolarında (VTSS) kaza gibi anormal olayların tespiti için derin öğrenme tabanlı bir yaklaşım önermektedir. Bu çalışmada, araç kazalarını içeren görüntülerden oluşan bir VAID (Vehicle Accident Image Dataset) oluşturulmakta ve bir CNN modeli bu veri kümesi ile eğitilmektedir. Eğitilen model, gerçek trafik gözetim videoları üzerinde kayan pencere ve 9 olling prediction stratejisi ile çalıştırılmakta; her bir zaman dilimi için anomali kararı verilmektedir. Deneysel analizler, yönteminin yüksek doğrulukla kazaları tespit edebildiğini ve gerçek trafik senaryolarında uygulanabilir bir çözüm sunduğunu ortaya koymaktadır [34].

Yang ve arkadaşlarının geliştirdiği yöntem, kalabalık hareketlerini grafik tabanlı bir yapı üzerinden modelleyerek anomali tespiti gerçekleştirmeyi hedeflemektedir. Önerilen mimaride, kalabalık içindeki hareket vektörleri ve bölgeler, grafın düğüm ve kenarları şeklinde temsil edilmekte; graf yapısı zaman içinde “evolving graph” yaklaşımıyla dinamik olarak güncellenmektedir. Bu grafik üzerinde tanımlanan derin öğrenme tabanlı ağ, normal kalabalık hareket kalıplarını öğrenmekte; bu kalıplardan belirgin şekilde sapan düğümler ve alt yapılar anomali göstergesi olarak işaretlenmektedir. The Visual Computer dergisinde sunulan deneysel sonuçlar, yöntemin farklı kalabalık veri kümelerinde güçlü doğruluk ve lokalizasyon performansı sergilediğini göstermektedir [35].

Habeb ve ekibinin çalışmasında, Vision Transformer (ViT) temelli bir kodlayıcı ile uzamsal-zamansal ilişkiyi (STR attention) yakalayan bir dikkat bloğu birleştirilerek, büyük ve heterojen video veri kümeleri üzerinde gözetimsiz anomali tespiti yapılmaktadır. Ön-eğitimli ViT, karelerden küresel düzeyde uzamsal özellikler çıkarmakta; ardından STR bloğu, nesneler arası zamansal ilişkileri ve etkileşimleri daha iyi modellemektedir. Bu hibrit yapı, klasik CNN tabanlı modellerin sınırlı görüş alanı (receptive field) problemini hafifletmekte ve hem lokal hem de global bağlamı birlikte değerlendirmektedir. Çalışmada

raporlanan sonuçlar, özellikle çeşitli sahne türlerini içeren büyük veri kümelerinde, önerilen çerçevenin mevcut pek çok dikkat tabanlı yöntemi geride bıraktığını göstermektedir [36].

Muna ve arkadaşları tarafından önerilen modelde, video anomali tespiti için attention odaklı yinelenen (recurrence) mekanizmalarla zenginleştirilmiş bir Vision Transformer gömülü mimarisi kullanılmaktadır. Önerilen çerçeve, ilk aşamada ViT ile karelerden uzamsal özellikleri çıkarmakta; devamında, dikkat güdümlü tekrarlayan katmanlar, zaman içindeki uzun menzilli bağımlılıkları modellemektedir. Böylece hem kare içi uzamsal ilişkiler hem de kareler arası zamansal dinamikler ortak bir dikkat yapısı altında öğrenilmektedir. Array dergisinde sunulan bulgular, bu yaklaşımın özellikle karmaşık, uzun süreli senaryolarda önceki ViT tabanlı çözümlere göre daha kararlı ve yüksek performanslı bir anomali tespiti sunduğunu ortaya koymaktadır [37].

Elmetwally ve çalışma arkadaşları, gerçek zamanlı video akışlarında düzensiz olayları tespit etmek için I3D-ResNet50 tabanlı bir özellik çıkarıcı ile derin Çoklu Örnek Öğrenme (deep MIL) yaklaşımını entegre etmektedir. Düzenli ve düzensiz videolar, MIL çerçevesinde negatif ve pozitif paketler (bag) olarak ele alınmakta; her bir video segmenti, bu paketlerin bir örneği olarak temsil edilmektedir. Tam bağlantılı bir sinir ağı aracılığıyla her segmente anomali skorları atanmakta ve bu skorların dağılımı üzerinden video düzeyinde anomali kararı verilmektedir. Çalışma, modelin hem doğruluk hem de zaman açısından gerçek zamanlı gözetim uygulamaları için uygun bir çözüm sunduğunu göstermektedir [38].

Son olarak, Mahareek ve ekibi tarafından hazırlanan kapsamlı derleme çalışması, gözetim videolarında anomali tespiti için önerilen yöntemleri hem klasik hem de derin öğrenme temelli yaklaşımlar bağlamında sistematik biçimde özetlemektedir. Çalışma; öznitelik çıkarımı, hareket modelleme, öğrenme paradigması (tam süpervizyon, zayıf etiketleme, gözetimsiz öğrenme) ve kullanılan benchmark veri kümeleri açısından ayrıntılı bir sınıflandırma sunmaktadır. Ayrıca gerçek zamanlılık, sınırlı etiketli veri, genelleme ve açıklanabilirlik gibi güncel zorluklar sistematik olarak tartışılmakta; gelecekteki araştırmalar için transformer, grafik tabanlı modeller ve difüzyon temelli yaklaşımlar gibi yönelimler vurgulanmaktadır. Bu bağlamda söz konusu derleme, alandaki yöntemsel çeşitliliği ve araştırma boşluklarını bütüncül bir bakış açısıyla ortaya koyan önemli bir referans niteliği taşımaktadır [39].

1.3 Öne Çıkan Özellikler

Günümüzde anomalilerin tespiti üzerine yapılan çalışmaların büyük bir bölümü, belirli bir senaryoya veya tek bir olay türüne odaklanan modeller geliştirmeye yönelmiştir. Literatürde yer alan bu çalışmaların çoğu, yalnızca kavga, yangın, hırsızlık, kaza ya da belirli bir davranış biçimi gibi tekil anomalileri algılamayı hedefleyen dar kapsamlı çözümler sunmaktadır. Dolayısıyla bu yaklaşımlar, gerçek dünyadaki çok boyutlu ve çok kaynaklı anomali senaryolarını aynı anda ele almakta yetersiz kalabilmektedir.

Bu tezin temel farkı, anomalilerin birbirinden bağımsız ve ayrı modellerle analiz edilmesi yerine, çoklu görev **(**multi-task**)** mimarisini merkeze alan bir yaklaşım benimsemesidir. Bu mimari sayesinde sistem, aynı anda birden fazla anomalinin semantik özelliklerini öğrenebilmekte; görsel verilerdeki davranışsal, çevresel ve nesne temelli ipuçlarını ortak bir temsil üzerinde birleştirerek daha kapsamlı bir değerlendirme yapabilmektedir. Böylece tek bir model, farklı anomali türlerini eş zamanlı olarak algılayabilmekte; bu da hem hesaplama maliyetlerini düşürmekte hem de tespit performansını artırmaktadır.

Ayrıca, çoklu görev yaklaşımının sunduğu en önemli avantajlardan biri, görevler arasındaki bilgi paylaşımıdır. Literatürdeki tek-anomali modelleri yalnızca kendi problem alanlarına özgü özellikleri öğrenirken; bu tezde önerilen mimari, farklı anomalilerin ortak özelliklerini de yakalayarak genelleme kabiliyetini güçlendirmektedir. Bu durum, özellikle düşük örnek sayısına sahip anomali türlerinde performansın belirgin şekilde artmasını sağlamaktadır. Böylece sistem hem geniş kapsamlı olay çeşitliliğini hem de veri dengesizliğini başarılı bir şekilde yönetebilen daha dayanıklı bir yapıya kavuşmaktadır.

Sonuç olarak, bu tezde sunulan çalışma, klasik tekli-anomali tespit yaklaşımlarının ötesine geçerek gerçek dünyadaki çok boyutlu anomalileri bütüncül bir bakış açısıyla değerlendirebilen, ölçeklenebilir ve yüksek doğrulukta çalışan bir yapay zekâ çözümü ortaya koymaktadır. Bu yönleriyle çalışma, literatürdeki mevcut yöntemlere kıyasla hem metodolojik hem de pratik açıdan önemli bir yenilik sunmakta; çoklu senaryo analizi gerektiren uygulamalar için daha işlevsel bir alternatif oluşturmaktadır.

2 MATERYAL VE YÖNTEM

2.1 VERİ KÜMESİ

2.1.1 Veri Etiketleme

Çalışmada kullanılan veri kümesi, toplam 25.069 görsel içermektedir. Veri kümesindeki anomali (olay) türlerinin görüntü sayılarına göre dağılımı, Tablo 2.1'de detaylı olarak gösterilmektedir.

Tablo 2.1. Anomali görüntü sayıları

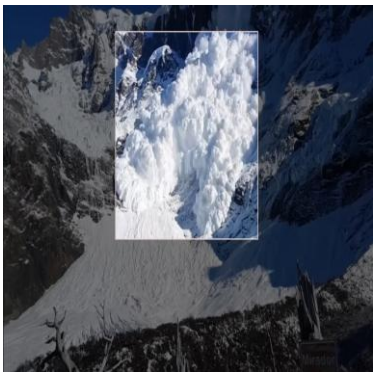
Anomali	Kamuya Açık Veri Kümesi (Adet)	Özel Veri Kümesi (Adet)	Toplam Görüntü Sayısı
Çığ	-	2,053	2,053
Ev Hırsızlığı	1,217	2,122	3,339
Araç Kazaları	2,170	2,558	4,728
Kavga-Dövüş	1,183	-	1,183
Silahlı Saldırılar	-	5,269	5,269
Toprak Kayması	-	2,043	2,043
Market Soygunları	1,113	-	1,113
Hortum	-	2,019	2,019
Tsunami	-	1,591	1,591
Orman Yangınları	1,731	-	1,731

Veri kümesinde bulunan görseller, eğitim, doğrulama (İng. validation) ve test kümeleri olmak üzere üç ana gruba ayrılmıştır. Bu ayrım için sırasıyla %87 eğitim, %9 doğrulama ve %4 test oranları belirlenmiştir. Belirlenen bu oranlara uygun olarak, her bir anomali türü için ayrı ayrı dağıtılan görüntü sayıları Tablo 2.2'de sunulmuştur.

Tablo 2.2. Veri Kümesi Detayları

Olay	Eğitim Kümesi (%87)	Doğrulama Kümesi (%9)	Test Kümesi (%4)	Toplam Görüntü Sayısı
Çığ (Avalanche)	1,791	171	85	2,053
Ev Hırsızlığı (Burglary)	2,862	319	158	3,339
Araç Kazaları (Crash)	4,137	394	197	4,728
Kavga-Dövüş (Fighting)	981	135	67	1,183
Silahlı Saldırılar (Gun)	4,611	439	219	5,269
Toprak Kayması (Landslide)	1,788	170	85	2,043
Market Soygunları (Robbery)	927	123	63	1,113
Hortum (Tornado)	1,767	168	84	2,019
Tsunami	1,392	133	66	1,591
Orman Yangınları (Wildfire)	1,521	143	67	1,731

Özel veri kümesi oluşturulurken iki ana kaynaktan yararlanılmıştır: (a) İnternet kaynaklı videolar ve (b) kamera çekimi ile elde edilen yerel kayıtlar. İnternetten elde edilen videolar çerçevelere ayrılarak ve saniye başına 5 kare alınarak görüntülere dönüştürülmüş ve etiketlenmiştir. Kamera çekimi ile elde edilen videolar ise bilgisayara aktarıldıktan sonra, internet kaynaklı videolarda uygulanan aynı yöntemle çerçevelere ayrılmış ve etiketleme süreci tamamlanmıştır Her anomali olayını temsil eden örnek görseller, Şekil 2.1'de sunulmuştur.



(a) Çığ



(b) Hırsızlık



(c) Araç Kazası



(ç) Kavga



(d) Silahlı Saldırı



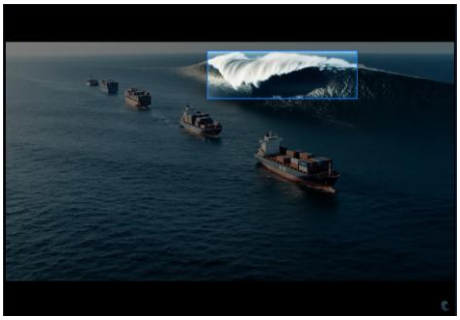
(e)Toprak Kayması



(f) Soygun



(g) Hortum



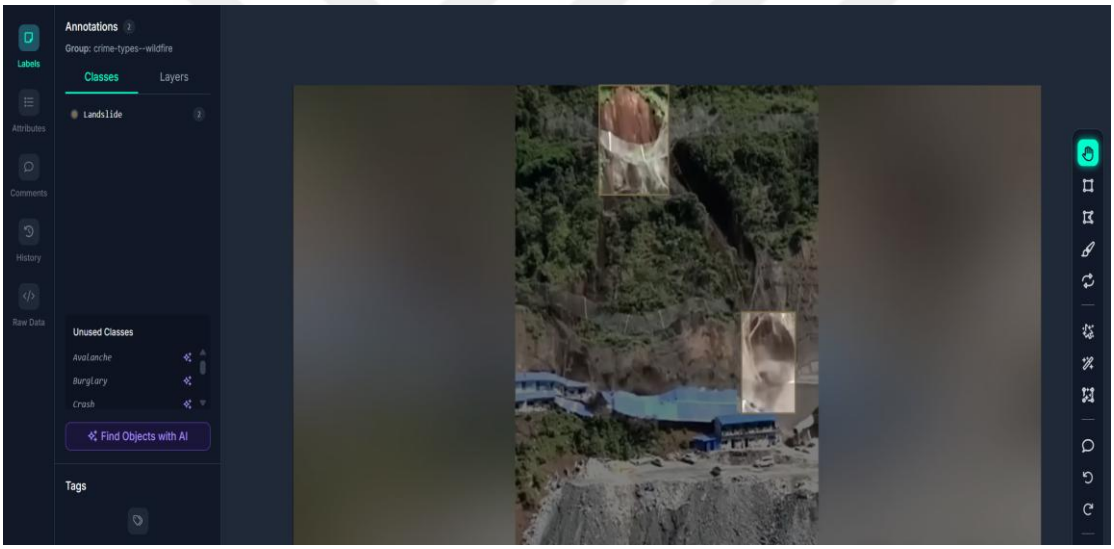
(ğ) Tsunami



(h) Orman Yangını

Şekil 2.1. Etiketlenmiş anomali görselleri

Görsellerin her biri, sınırlayıcı kutular (İng. bounding box) kullanılarak etiketlendirilmiştir. Uygulanan bu etiketleme yöntemine ait bir örnek Şekil 2.2'de gösterilmektedir.



Şekil 2.2. Roboflow Etiketleme Programı [40]

Etiketleme sonrasında tüm etiketlenmiş görseller, tek klasör içinde COCO [41, 42] formatına uygun olarak çıkartılmaya hazır hale getirilmiştir.

2.1.2 Veri Arttırma

Veri kümesinin daha uygun ve sorunsuz bir şekilde modelde eğitilebilmesi için tüm görseller, 320 x 320 boyutuna getirilmiştir. Daha sonrasında eğitilen her görsel, 3 katı sayıya çıkartılmıştır. Görsellerde uygulanan veri arttırma sistemi, Tablo 2.3'te belirtilmiştir.

Tablo 2.3. Veri arttırma işlemleri

Yöntem	Değer
Rotation	-90° ~ +90°
Saturation (Doygunluk)	-25% ~ +25%
Exposure (Parlaklık Farkı)	-10% ~ +10%

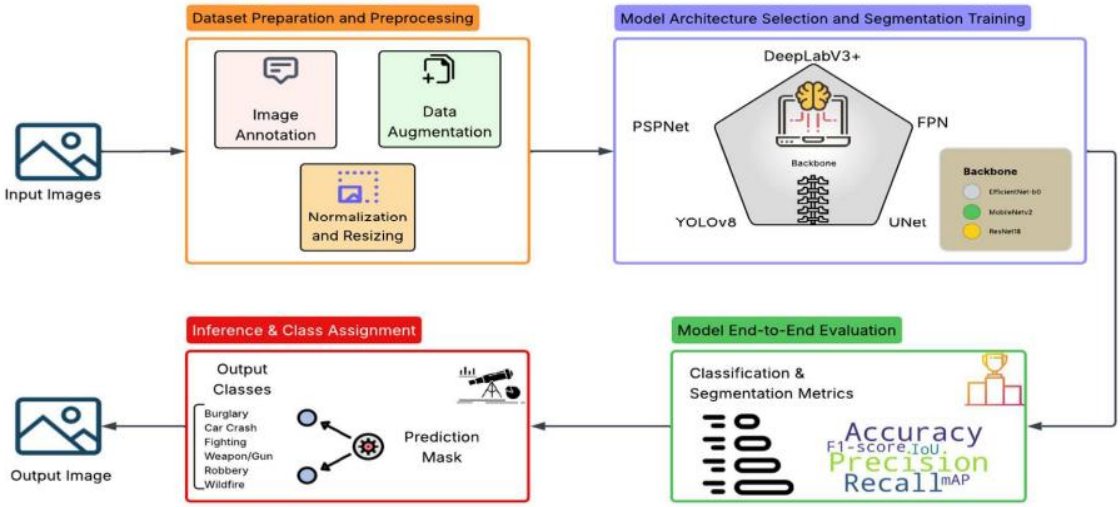
2.1.3 Veri Çıktıları

Modelin eğitilebilmesi için, görsellerin etiketli (İng. bounding box) hallerine ek olarak maskeli görüntüleri (bölütleme maskeleri) de gerekmektedir. Veri kümesi, ilk aşamada COCO formatında dışa aktarıldıktan sonra, özel bir ek yazılım kullanılarak etiketlemeler COCO'dan maske formatına dönüştürülmüştür.

Maskeli görüntüleri elde etmek amacıyla yazılan kod; girdi olarak tüm görsellerin bulunduğu dosyayı almaktadır. Klasöre erişimin ardından, her görüntü tek tek işlenerek etiketlenen kısım tespit edilir ve maske görüntüsüne çevrilir. Elde edilen tüm maskeli görüntüler ayrı bir klasöre kaydedilmektedir. Bu işlem, model eğitimi için zorunlu olan maske klasörünün dosya yolunu (İng. path) oluşturmuştur.

2.2 KULLANILAN MODELLER

Bu tezde yapılan karşılaştırmalı değerlendirme (İng. Benchmark); Şekil 2.3'te görüldüğü gibi, 5 mimari ve onlara bağlı olarak çalışan 3 omurga (İng. Backbone) aracılığıyla gerçekleştirilmiştir.

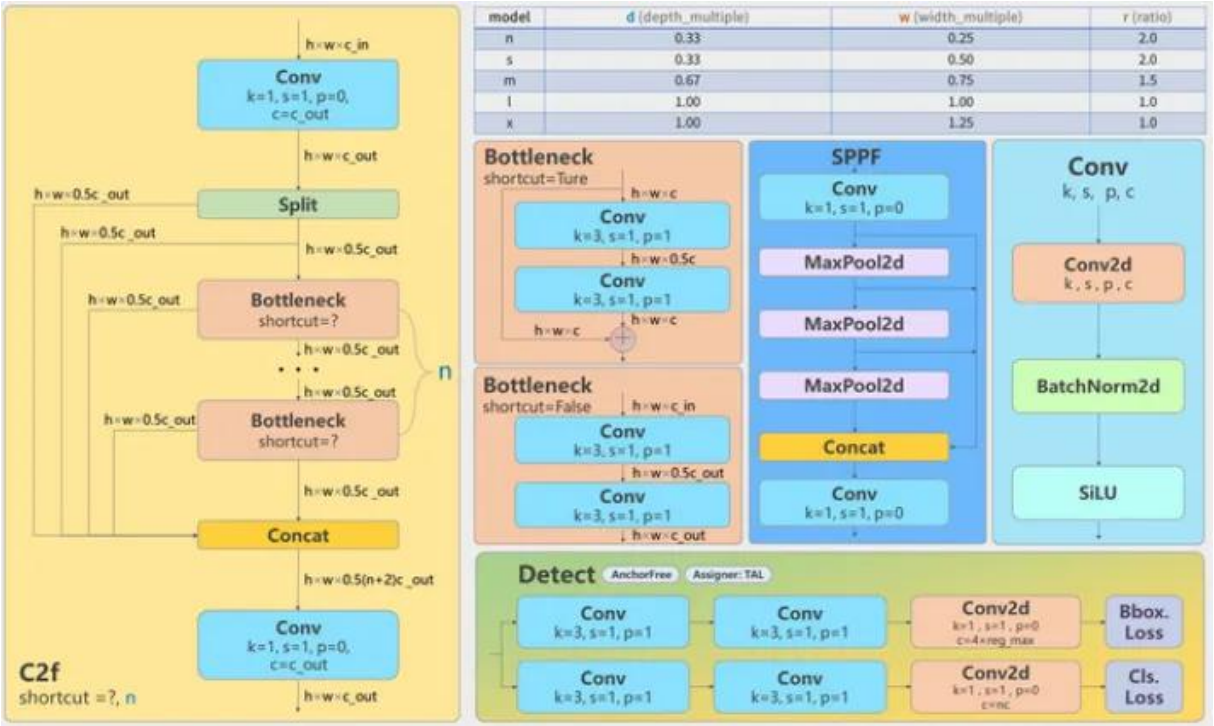


Şekil 2.3. Mimari – Omurga Benchmark Yapısı

Kullanılan mimariler YoloV8, U-Net, FPN, DeepLabV3+, PSPNET olmakla birlikte, kullanılan omurgalar EfficientNet-b0, MobileNetV2, Resnet18'dir. Söz konusu mimariler ve omurgalar için detaylı açıklamalara sıradaki bölümden erişilebilir.

2.2.1 YOLOV8

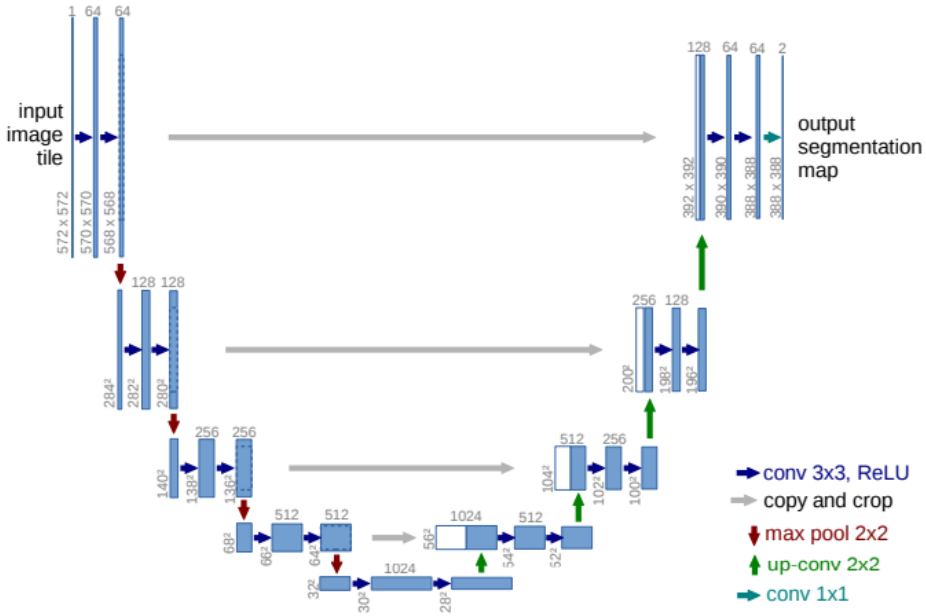
YOLOv8 (You Only Look Once v8), gerçek zamanlı nesne tespiti alanında YOLO ailesinin devamı niteliğinde geliştirilen ve tek aşamalı algılama yaklaşımını modern bir mimariyle birleştiren ileri seviye bir tespit modelidir. Bu yapı, yüksek doğruluk ve hızlı çıkarım gerektiren görevler için optimize edilmiştir. YOLO mimarisinin temel prensipleri ilk olarak Redmon ve arkadaşları tarafından sunulan YOLO modelleri ile ortaya konmuş olup [43], daha sonraki sürümlerde gerçekleştirilen iyileştirmelerle özellikle anchor-free yapılar, kayıp fonksiyonlar ve çok ölçekli tespit yetenekleri önemli ölçüde geliştirilmiştir. YOLO ailesinin evrimsel ilerleyişi YOLOv3 ile önemli bir dönüşüm yaşamış ve mimari modern derin öğrenme teknikleriyle yeniden şekillenmiştir [44]. YOLOv8'in resmi teknik raporu olmamakla birlikte mimari özelliklerine ilişkin kapsamlı açıklamalar Ultralytics tarafından sağlanmaktadır [45]. Bu yönüyle YOLOv8, anlık analiz, güvenlik sistemleri ve anomali tespiti gibi zaman kritik uygulamalarda öne çıkan güçlü bir modeldir. Önerilen YOLO mimarisi, Şekil 2.4'te görülmektedir.



Şekil 2.4. YOLOv8 Mimarisi [45]

2.2.2 U-NET

U-Net, özellikle tıbbi görüntü bölütlemesi alanında çığır açan bir mimari olarak literatürde yer almakta olup, kodlayıcı–kod çözücü yapısı ve benzersiz atlama bağlantıları sayesinde sınırlı veri koşullarında dahi yüksek başarım elde edebilmektedir. İlk kez Ronneberger ve arkadaşları tarafından tanıtılan mimari [46], bölütleme görevlerinde hem yerel detayların korunmasını hem de derin katmanlarda öğrenilen semantik bilgilerin etkili bir şekilde birleştirilmesini mümkün kılar. Atlama bağlantıları (İng. skip connections) yapıları sayesinde uzamsal bilgi kaybı önemli ölçüde azaltılarak daha hassas bölütleme maskeleri üretilebilmektedir. Bu özellikleri ile U-Net, yalnızca tıbbi görüntülerde değil, genel amaçlı bölütleme problemlerinde de temel referans modellerden biri hâline gelmiştir. Mimari, Şekil 2.5'te görülmektedir.



Şekil 2.5. U-Net Mimarisi [46]

Base U-Net, kodlayıcı–kod çözücü tabanlı simetrik yapısı ile semantik bölütleme alanında temel referans mimarilerden biri olarak kabul edilmektedir. İlk olarak Ronneberger ve arkadaşları tarafından önerilen bu mimari, kodlayıcı bölümünde evrişim ve havuzlama katmanları aracılığıyla görüntünün uzamsal boyutlarını kademeli olarak azaltırken, yüksek seviyeli anlamsal özelliklerin öğrenilmesini sağlar. Kod çözücü bölümünde ise bu özellikler yukarı örnekleme işlemleri ile yeniden yüksek çözünürlüklü temsillere dönüştürülerek piksel seviyesinde sınıflandırma gerçekleştirilir. Kodlayıcı ve kod çözücü arasındaki skip connection yapıları, erken katmanlardaki uzamsal bilgilerin korunmasına olanak tanıyarak bölütleme maskelerinin sınır doğruluğunu önemli ölçüde artırmaktadır. Bu özellikleri sayesinde Base U-Net, sınırlı veri kümelerinde dahi etkili sonuçlar üretebilen güçlü bir temel mimari olarak literatürde geniş kabul görmüştür.

Inception U-Net, Base U-Net mimarisini, çok ölçekli özellik çıkarımını mümkün kılan Inception modülleri ile zenginleştiren gelişmiş bir bölütleme yaklaşımıdır. Inception modülleri, farklı çekirdek boyutlarına sahip paralel evrişim katmanları aracılığıyla aynı anda hem yerel hem de geniş bağlamsal bilgilerin öğrenilmesini sağlar [47]. Bu yapı, nesne boyutlarının ve şekillerinin büyük farklılıklar gösterdiği bölütleme problemlerinde modelin temsil gücünü önemli ölçüde artırmaktadır. Inception U-Net, U-Net'in kodlayıcı–kod çözücü yapısını ve skip connection mekanizmasını korurken, çok ölçekli temsiller sayesinde daha zengin ve ayırt edici özellik haritaları üretmektedir [48]. Bu yönüyle mimari, karmaşık sahnelerde ve düzensiz

yapıların bulunduğu görüntülerde daha kararlı ve genellenebilir sonuçlar sunmaktadır.

Residual U-Net, U-Net mimarisine residual bağlantıları entegre ederek derin ağların daha stabil ve verimli bir şekilde eğitilmesini amaçlayan bir modeldir. Residual bağlantılar, giriş bilgisinin katman çıkışına doğrudan eklenmesi prensibine dayanmakta ve bu sayede gradyan akışını kolaylaştırarak derin ağlarda sıkça karşılaşılan gradyan kaybolması problemini azaltmaktadır. Bu yaklaşım, U-Net mimarisine uygulandığında, ağın daha derin hâle getirilmesine olanak tanırken eğitim sürecinin daha hızlı ve kararlı bir şekilde yakınsamasını sağlar [49]. Residual U-Net, özellikle yüksek çözünürlüklü görüntülerde ve karmaşık yapısal detayların bulunduğu bölütleme görevlerinde Base U-Net'e kıyasla daha güçlü performans sergilemektedir.

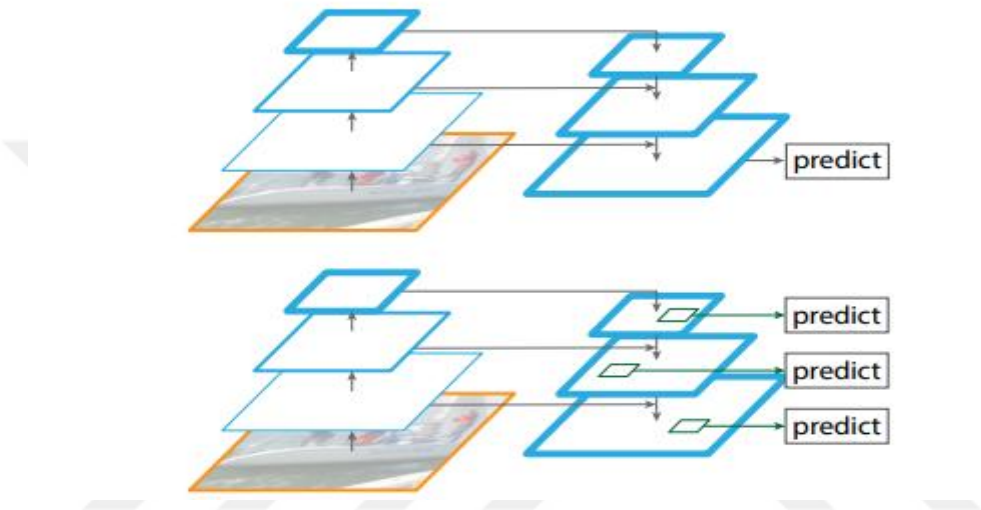
Attention U-Net, bölütleme sürecinde modelin yalnızca ilgili bölgeler üzerine odaklanmasını sağlamak amacıyla dikkat (İng. attention) mekanizmalarını U-Net mimarisiyle birleştiren ileri seviye bir yaklaşımdır. Bu mimaride, skip connection yollarına entegre edilen attention kapıları, kod çözücü aşamasına aktarılacak özelliklerin önem derecesini dinamik olarak belirler. Böylece anlamsal olarak önemsiz arka plan bölgeleri baskılanırken, hedef yapılar daha belirgin hâle getirilir. Attention mekanizması, özellikle küçük nesnelerin tespiti ve düşük kontrastlı bölgelerin bölütlenmesi gibi zorlu senaryolarda modelin duyarlılığını artırmaktadır [50]. Bu yapı sayesinde Attention U-Net, yanlış pozitif oranlarını azaltarak daha hassas ve güvenilir bölütleme çıktıları üretmektedir.

Dense U-Net, DenseNet mimarisinin temel prensiplerini U-Net yapısına entegre ederek katmanlar arası yoğun bağlantılar (dense connections) kullanan bir bölütleme modelidir. Bu bağlantılar sayesinde her katman, kendisinden önceki tüm katmanların çıktılarını girdi olarak alır ve böylece özelliklerin yeniden kullanımı maksimum seviyeye çıkarılır [51]. Dense bağlantılar, bilgi akışını güçlendirirken parametre verimliliğini artırmakta ve modelin aşırı öğrenme riskini azaltmaktadır. Dense U-Net, özellikle ince sınır detaylarının doğru biçimde yakalanmasının kritik olduğu bölütleme problemlerinde daha zengin özellik temsilleri sunarak Base U-Net'e kıyasla daha yüksek doğruluk elde edebilmektedir [52]. Bu yönüyle mimari, modern bölütleme uygulamaları için güçlü ve ölçeklenebilir bir alternatif sunmaktadır.

2.2.3 FPN

Çok ölçekli nesne tespiti problemlerini çözmek amacıyla geliştirilmiş ve modern tespit modellerinin çoğunda temel bileşen hâline gelmiş bir üst-yapı mimarisidir. Lin ve arkadaşlarının çalışmasında tanıtılan model [53], CNN (Convolutional Neural Networks)'in

farklı katmanlarından gelen hem yüksek çözünürlüklü hem de semantik açıdan zengin özellikleri birleştirerek çok ölçekli bir piramit yapısı oluşturur. Top-down ve lateral bağlantılar sayesinde her seviyedeki özellik haritası hem bağlamsal hem de ince detay bilgilerini taşır. Bu durum, modelin hem küçük boyutlu nesneleri algılama duyarlılığını yükseltmekte hem de büyük ölçekli nesnelerde tespit performansını geliştirmektedir. FPN (Feature Pyramid Network), bölütleme ve tespit modellerinde standart hâline gelmiş ve güncel mimarilerin çoğuyla uyumlu esnek bir yapı sunmaktadır. FPN mimarisi Şekil 2.6'da görülebilmektedir.



Şekil 2.6. FPN Mimarisi [53]

2.2.4 DeepLabV3+

DeepLabV3+, aslında DeepLab'ın farklı ve ileri sürümüdür. Sürümlerin çıkış sıraları, sırasıyla; DeepLab – DeepLabV2 – DeepLabV3 – DeepLabV3+ şeklinde olmuştur. DeepLabV3+, DeepLab'dan çok daha başarılı bölütleme sonuçları verir.

DeepLab [54], piksel düzeyinde anlamsal bölütlemenin doğruluğunu artırmak amacıyla derin evrişimli sinir ağları (DCNN) [55] ile tam bağlı Koşullu Rastgele Alanların (CRF) bütünleşik kullanımını temel alan bir yöntem olarak önerilmiştir. Önerilen model, DCNN mimarilerinin bölütleme süreçlerinde karşılaşılan yetersiz uzamsal duyarlılık problemini gidermek üzere CRF yapısını sisteme entegre etmekte ve böylece nesne sınırlarının daha keskin ve doğru biçimde tespit edilmesini mümkün kılmaktadır. DeepLab yaklaşımı, PASCAL VOC-2012 [56] veri kümesi üzerinde gerçekleştirilen deneylerde %71.6 IOU değerine ulaşarak dönemin en başarılı sonuçlarından birini elde etmiştir. Ayrıca yöntem, hesaplama verimliliğini artıran optimizasyon teknikleri sayesinde gerçek zamanlı uygulamalara uygun performans sergilemektedir.

DeepLabv3+, semantik bölütleme alanında yüksek doğruluk sunan, Atrous konvolüsyon ve kodlayıcı–kod çözücü yaklaşımını birleştiren gelişmiş bir mimaridir. Chen ve arkadaşları tarafından geliştirilen bu model [57], Atrous Spatial Pyramid Pooling (ASPP) modülü aracılığıyla farklı ölçeklerdeki bağlamsal bilgilerin aynı anda yakalanmasını sağlar. Atrous konvolüsyon, çözünürlüğün korunmasını mümkün kılarken, geniş bir algılama alanı sunar.

DeepLabV3+ ise V3 mimarisinin güçlü yönlerini korurken eksik kalan kısımlarını tamamlamak üzere geliştirilmiş bir genişletilmiş versiyondur. Bu model, V3'te yer almayan bir kod çözücü modülü ekleyerek yüksek çözünürlüklü bölütleme haritalarının daha doğru biçimde yeniden oluşturulmasını sağlar. Kod çözücü aşaması, özellikle kenar bölgelerinde ve ince yapılı objelerde belirgin iyileştirmeler sunar. Aynı zamanda, düşük seviyeli özellik haritalarının yüksek seviyeli bağlam bilgisiyle etkili şekilde birleştirilmesi sayesinde hem detay hem de bağlam bilgisini dengeli biçimde işleyebilir. Böylece DeepLabV3+ hem hesaplama verimliliğini koruyan hem de bölütleme doğruluğunu belirgin şekilde artıran bir mimari yaklaşım ortaya koymaktadır.

İki model arasındaki bir diğer önemli fark, DeepLabV3+'ın kodlayıcı kısmında sıklıkla *Xception* tabanlı derinleştirilmiş ayrıklaştırılmış konvolüsyonlar (depthwise separable convolutions) kullanmasıdır. Bu yapı, modelin hem hızını hem de parametre verimliliğini artırırken performans kaybı yaşanmamasını sağlar. Sonuç olarak DeepLabV3+, yalnızca V3'ün bağlamsal algılama kapasitesini geliştirmekle kalmamış, aynı zamanda endüstriyel uygulamalarda ve gerçek zamanlı sistemlerde tercih edilebilecek daha optimize edilmiş bir mimari sunmuştur.

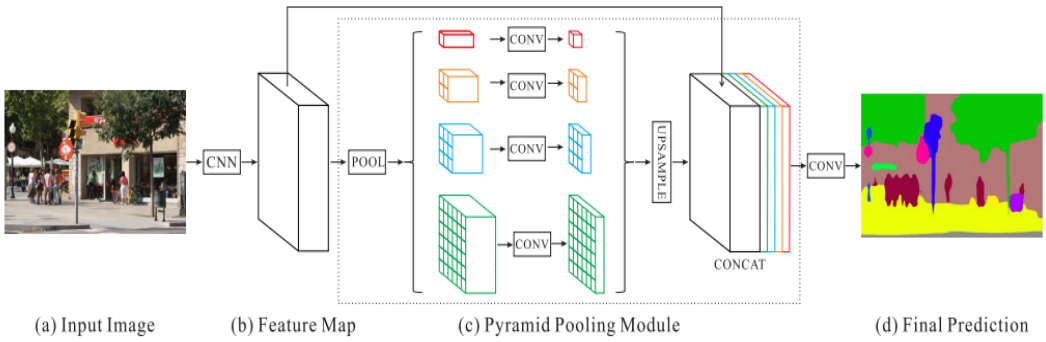
DeepLabv3+'ın önceki modellerden ayrıldığı en önemli nokta, eklenen kod çözücü modülünün bölütleme çıktısının kenar bölgelerinde daha hassas ve düzgün sonuçlar üretmesidir. Bu mimari hem küresel bağlamı hem de yerel detayları dengeli bir şekilde işleyerek karmaşık sahnelerde başarılı sonuçlar sunmaktadır.

2.2.5 PSPNet

PSPNet, sahne anlama (scene parsing) ve semantik bölütleme problemlerine yönelik geliştirilmiş güçlü bir mimaridir. Zhao ve arkadaşları tarafından önerilen bu model [58], Pyramid Pooling Module (PPM) adı verilen çok ölçekli havuzlama mekanizmasıyla görüntünün hem yerel hem de küresel bağlamını etkili biçimde yakalar.

Çok seviyeli havuzlama işlemleri, Şekil 2.7'de de görüldüğü üzere modelin görüntüdeki geniş

sahne ilişkilerini anlamlandırmasına olanak tanırken, aynı zamanda yerel ince detaylar da korunur. Bu yapı sayesinde PSPNet, karmaşık ve geniş sahnelerde bölütleme doğruluğunu önemli ölçüde artırmış ve literatürde birçok takip eden çalışmaya yön vermiştir.



Şekil 2.7. PSPNET Mimarisi [58]

2.3 KULLANILAN OMURGALAR

2.3.1 EfficientNet

EfficientNet, derin öğrenme mimarilerinde doğruluk–verimlilik dengesini optimize etmeye yönelik bileşik ölçekleme (compound scaling) stratejisini temel alan modern bir ağ ailesidir. Tan ve Le tarafından önerilen bu yaklaşım [59], model derinliği, genişliği ve çözünürlüğünü eş zamanlı olarak dengeli bir biçimde artırarak hem daha kompakt hem de daha yüksek doğrulukta çalışan modeller elde edilmesini sağlar. EfficientNet'in temel bileşenleri olan MBConv blokları ve squeeze-and-excitation mekanizması, modelin önemli özelliklere odaklanmasını artırarak temsil gücünü geliştirmektedir. Daha ileri bir sürüm olan EfficientNetV2'nin ortaya koyduğu hız ve eğitim verimliliği artışları [60], EfficientNet mimarisinin ölçeklenebilirliğini ve pratik kullanım değerini daha da güçlendirmiştir. Ayrıca EfficientNet tabanlı mimariler, nesne tespiti gibi görevlerde Scaled-YOLOv4 gibi modellerde omurga olarak kullanılarak performans artışı sağlamış ve literatürde yaygın uygulama alanı bulmuştur [61]. Bu yönleriyle EfficientNet, hem sınırlı hesaplama kaynağına sahip cihazlarda hem de yüksek doğruluk gerektiren büyük ölçekli görüntü işleme görevlerinde tercih edilen etkili ve modern bir çözüm olarak öne çıkmaktadır.

2.3.2 MobileNet

MobileNet, özellikle mobil cihazlar ve gömülü sistemler gibi sınırlı hesaplama kapasitesine sahip ortamlarda verimli çalışmak üzere tasarlanmış hafif bir derin öğrenme mimarisidir. Howard ve arkadaşlarının geliştirdiği ilk MobileNet sürümü [62], derinlik ayrılabilir evrişim (depthwise separable convolution) operasyonunu temel alarak geleneksel evrişimsel sinir ağlarına kıyasla hesaplama maliyetini önemli ölçüde azaltmıştır. MobileNetV2 ile tanıtılan tersine çevrilmiş artık bloklar (inverted residual blocks) ve lineer darboğaz yapıları [63], modelin doğruluk–verimlilik dengesini daha da iyileştirmiş; MobilenetV3 ise mimari aramasına dayalı optimizasyon süreçleri sayesinde (NAS) bazı görevlerde ek performans kazançları sağlamıştır [64]. Bu modellerin düşük parametre sayısı, hızlı çıkarım süreleri ve enerji verimlilikleri, MobileNet'i nesne tespiti, bölütleme ve anomali analizi gibi geniş bir uygulama alanında ideal bir yapı hâline getirmektedir. Ayrıca MobileNet tabanlı hafif modeller, sınıf aktivasyon haritalarını üreten CAM tekniğiyle birlikte kullanıldığında görselleştirme ve açıklanabilirlik açısından ek avantajlar sunmaktadır [65].

2.3.3 ResNet

ResNet (Residual Neural Network), derin sinir ağlarının eğitiminde karşılaşılan gradyan kaybolması (vanishing gradient) problemini çözerek derin öğrenme alanında büyük bir kırılma noktası yaratmış bir mimaridir. He ve arkadaşları tarafından geliştirilen residual bağlantı mekanizması [66], ağın çok daha derin katmanlarla stabil bir şekilde eğitilmesine olanak tanır. Residual bloklar, giriş sinyalini kısayol yolları üzerinden doğrudan blok çıkışına aktararak gradyan akışının kaybolmadan iletilmesini mümkün kılar. Bu yaklaşımın ilerleyen versiyonlarda kimlik eşleşmeleriyle (identity mappings) daha da iyileştirildiği görülmektedir [67]. ResNet'in genişletilmiş türevleri arasında yer alan WideResNet [68], ağın genişliğini artırarak daha az derin ama daha yüksek performans sunan alternatifler ortaya koymuş; ResNeXt ise gruplandırılmış konvolüsyonları kullanarak daha güçlü bir temsil kapasitesi geliştirmiştir [69]. Hem sağlam genel performansı hem de modern modeller için oluşturduğu temel altyapı sayesinde ResNet, görüntü sınıflandırma, nesne tespiti ve bölütleme gibi birçok bilgisayarlı görü görevinin standart omurga mimarisi hâline gelmiştir.

Tablo 2.4. Omurgaların Parametre Sayıları

Omurga	Parametre Sayısı (M)
ResNet18	11
MobileNet-V2	2
EfficientNet-b0	4

Tablo 2.4'te, tercih edilen omurga versiyonları verilmiştir. Söz konusu omurgaların tüm versiyonları da Tablo 2.5'te verilmiştir. En düşük parametreli versiyonlarının seçilmesinin nedeni, model eğitirken zaman tasarrufu sağlamaktır. Parametre sayısının az olmasıyla birlikte, model sonuçları daha hızlı gelecek olup, gereken metrik sonuçlarına daha hızlı erişilmesi amaçlanmıştır.

Tablo 2.5. Derin Öğrenme Mimarilerinde Kullanılan Omurgaların Varyasyonları ve Parametre Sayıları (M)

Omurga	Varyasyon	Parametre Sayısı (M)
RESNET	ResNet18	11
	ResNet34	21
	ResNet50	23
	ResNet101	42
	ResNet152	58
MOBILENET	MobileNetV2	2
EFFICIENTNET	EfficientNet-b0	4
	EfficientNet-b1	6
	EfficientNet-b2	7
	EfficientNet-b3	10
	EfficientNet-b4	17
	EfficientNet-b5	28
	EfficientNet-b6	40

	EfficientNet-b7	63

2.4 ÖNERİLEN MODEL

2.4.1 Topluluk Öğrenmesi Stratejisi

Bu çalışmada, tek bir modelin yanlılık (İng. bias) ve varyans hatalarını minimize etmek, segmentasyon doğruluğunu artırmak amacıyla farklı mimari ve omurgalara sahip modellerden oluşan bir topluluk (İng. ensemble) yapısı kurgulanmıştır.

2.4.2 Model Çeşitliliği ve Mimari Yapı

Topluluk modeli, iki farklı segmentasyon mimarisi (DeepLabV3+ ve FPN) ile üç farklı omurganın (ResNet18, EfficientNet-B0, MobileNetV2) kombinasyonundan oluşan toplam 6 alt modelden meydana gelmektedir.

- **DeepLabV3+:** Atrous Spatial Pyramid Pooling (ASPP) modülü sayesinde çok ölçekli bağlamsal bilgiyi yakalamada etkindir.
- **FPN:** Farklı çözünürlükteki özellik haritalarını birleştirerek küçük nesnelerin tespitinde yüksek performans sergiler.

Bu çeşitlilik, modellerin görüntüdeki nesnelere farklı perspektiflerden bakmasını sağlamıştır.

2.4.3 Tahmin Birleştirme

Alt modellerin çıktılarını birleştirmek için Olasılık Tabanlı Ortalama yöntemi kullanılmıştır. Her bir M_i modeli, giriş görüntüsü x için bir çıktı üretir. Bu çıktılar, normalleştirilmiş olasılık haritalarına dönüştürülmek üzere Denklem (2.1)'de görülen Sigmoid aktivasyon fonksiyonundan geçirilir:

$$P_i(x) = \sigma\big(M_i(x)\big) \tag{2.1}$$

Nihai topluluk tahmini, her bir modelin ürettiği olasılık haritalarının aritmetik ortalaması alınarak Denklem (2.2)'deki gibi hesaplanır:

$$y = \frac{1}{N}\sum_{i=1}^{N} P_i(x) \tag{2.2}$$

Burada $N = 6$ olup, her modelin nihai karara katkısı eşittir.

2.4.4 Topluluk Ağırlıklarının Optimizasyonu

Topluluğu oluşturan her bir modelin (M_i) tahmin yeteneği ve hata karakteristiği farklılık göstermektedir. Bu nedenle, basit aritmetik ortalama yerine, her modele performansıyla doğru orantılı bir ağırlık (w_i) atanan Ağırlıklı Ortalama (Weighted Averaging) yöntemi tercih edilmiştir. Nihai topluluk tahmini Denklem (2.3) ile hesaplanmaktadır:

$$y = \sum_{i=1}^{N} w_i \cdot P_i(x), \qquad \sum_{i=1}^{N} w_i = 1 \tag{2.3}$$

Söz konusu ağırlıkların ($w_1, w_2, \dots, w_N$) optimizasyon sürecinde, arama uzayını temsil eden olasılık simpleksi üzerinden rastgele ve tutarlı örnekler türetmek amacıyla Dirichlet dağılımından yararlanılmıştır. Dirichlet dağılımı, toplamları bire eşit olan ve negatif değer almayan değişken kümeleri üretmek için kullanılan çok değişkenli bir olasılık yoğunluk fonksiyonudur. Bu çalışmada, her bir alt modelin başlangıçta eşit olasılıkla temsil edildiği tekdüze bir dağılım elde etmek için yoğunluk fonksiyonunun parametre vektörü $\alpha = (1, \dots, 1)$ olarak belirlenmiş ve ağırlık kümeleri Denklem (2.4)'teki olasılık yoğunluk fonksiyonu uyarınca örneklenmiştir:

$$f(w_1, \dots, w_N; \alpha_1, \dots, \alpha_N) = \frac{1}{B(\alpha)} \prod_{\{i=1\}}^{N} w_i^{\{\alpha_i - 1\}} \tag{2.4}$$

2.5 ANOMALİ TÜRLERİ

Tez kapsamında kullanılan 10 anomali olayın hepsine, aşağıdaki bölümde görüntü örnekleriyle birlikte bahsedilmiştir.

2.5.1 Çığ

Çığ (İng. Avalanche) görüntülerinin veri kümesi, özel olup, içerikler Şekil 2.8'de görüldüğü gibi internetten, çeşitli filmler ve dizilerin kesitlerinden toplanmıştır.



Şekil 2.8. Çığ Orijinal Görselleri

2.5.2 Ev Hırsızlığı

Hırsızlık (İng. Burglary) için kullanılan veri kümesi, hibrit yapıda olup, hem halka açık kaynaklardan elde edilen hem de özel olarak toplanan görüntülerden oluşmaktadır. Şekil 2.9'da bir örneği görülen halka açık veri kümesi görselleri, Roboflow [40] platformundan toplanmıştır. Özel görüntüler ise, çeşitli mekânların kamera kayıtlarından elde edilmiştir.



Şekil 2.9. Ev Hırsızlığı Orijinal Görselleri

2.5.3 Araç Kazaları

Araç kazaları (İng. Crash) için kullanılan veri kümesi de hibrit yapıya sahip olup hem halka açık hem de özel görüntüler içermektedir. Halka açık veri kümesi görüntüleri, Kaggle'dan [7] toplandı, özel görüntüler ise, kamera ile çekim sonucu elde edilmiştir. Şekil 2.10'da örnekler görülebilmektedir.



Şekil 2.10. Araç Kazaları Orijinal Görselleri

2.5.4 Kavga – Dövüş

Dövüş (İng. Fighting) için kullanılan veri kümesi, Şekil 2.11'deki gibi halka açık veri kümesi olup, Roboflow'dan [40] toplanmıştır.



Şekil 2.11. Kavga – Dövüş Orijinal Görselleri

2.5.5 Silahlı Saldırılar

Silahlı Saldırı (İng. Gun) için gerekli veri kümesi, tamamıyla özel veri kümesi olup, Şekil 2.12'deki örnekler gibi internet üzerinden paylaşılan videolardan görüntüler elde edilmiştir.



Şekil 2.12. Silahlı Saldırı Orijinal Görselleri

2.5.6 Toprak Kayması

Toprak kayması (İng. Landslide) görüntüleri özel veri kümesi olup, internet üzerindeki videolardan elde edilmiştir. Örnekler Şekil 2.13'teki gibidir.



Şekil 2.13. Toprak Kayması Orijinal Görselleri

2.5.7 Market Soygunları

Soygun (İng. Robbery) için kullanılan veri kümesi de Şekil 2.14'teki gibi halka açık yapıya sahiptir ve içerikler Roboflow'dan [40] toplanmıştır.



Şekil 2.14. Market Soygunları Orijinal Görselleri

2.5.8 Hortum

Hortum (İng. Tornado) için kullanılan görseller, Şekil 2.15'te görüleceği üzere özel veri kümesidir. İnternette insanların yaptığı kayıtlar üzerinden yüklenen videolar kullanılarak görüntüler alınmıştır.



Şekil 2.15. Hortum Orijinal Görselleri

2.5.9 Tsunami

Tsunami veri kümesi için kullanılan görseller, özel veri kümesidir. Çeşitli filmler ve dizilerin kesitlerinden ekran kaydı alınıp, videolar görüntülere çevrilmiştir. Çeşitli örnekler Şekil 2.16'da gösterilmiştir.



Şekil 2.16. Tsunami Orijinal Görselleri

2.5.10 Orman Yangınları

Orman yangınları (İng. Wildfire) veri kümesi için kullanılan görseller, Şekil 2.17'deki gibi halka açık veri kümesi olup, Kaggle'dan [7] toplanmıştır.



Şekil 2.17. Orman Yangınları Orijinal Görselleri

2.6 PERFORMANS METRİKLERİ

2.6.1 Doğruluk

Bir nesnenin doğru sınıflandırılmasını ifade eder ve modelin ürettiği tüm tahminler içerisinde ne kadarının gerçeğe uygun olduğunu gösterir. Bu metrik, sınıflandırma performansının genel bir özetini sunarak modelin ne ölçüde isabetli sonuçlar verdiğine dair temel bir değerlendirme yapılmasına olanak tanır. Dolayısıyla hem doğru pozitiflerin hem de doğru negatiflerin toplam tahminler içerisindeki oranını yansıtan bu ölçüt, modelin genel güvenilirliğini anlamada kritik bir role sahiptir. Söz konusu hesaplama yöntemi (2.5) numaralı formülde detaylı biçimde sunulmaktadır [70].

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.5}$$

2.6.2 Kesinlik

Model tarafından pozitif olarak etiketlenen örneklerin ne kadarının gerçekten pozitif olduğunu ortaya koyan bir ölçüttür. Bu metrik, özellikle yanlış pozitiflerin kritik sonuçlar doğurabileceği uygulamalarda modelin pozitif tahminlerinin güvenilirliğini değerlendirmek açısından büyük önem taşır. Precision değeri ne kadar yüksekse, modelin pozitif olarak işaretlediği örneklerin o kadar büyük bir kısmı gerçeği yansıtmaktadır. Bu nedenle, modelin yanlış alarm verme eğilimini analiz etmede temel bir performans ölçütü olarak kabul edilir. Formülasyonu (2.6) numaralı denklemde verilmiştir [71].

$$Pre = \frac{TP}{TP + FP} \tag{2.6}$$

2.6.3 Geri Çağırma

Bir modelin veri içerisindeki gerçek pozitif örnekleri ne ölçüde doğru bir şekilde tespit edebildiğini gösteren kritik bir istemdir. Geri Çağırma, modelin pozitif sınıfa ait örnekleri yakalama başarısını ölçer. Başka bir ifadeyle, modelin gözden kaçırdığı (yanlış negatif) örneklerin oranını anlamaya yardımcı olur. Gerçek pozitiflerin, gerçek pozitif ve yanlış negatif toplamına oranı alınarak hesaplanan bu metrik, özellikle kaçırılmasının yüksek maliyet doğurduğu anomali tespiti, hastalık teşhisi veya güvenlik uygulamalarında büyük önem taşımaktadır. Bu hesaplamaya ilişkin formül (2.7) numaralı denklemde sunulmaktadır [71].

$$Re = \frac{TP}{TP + FN} \tag{2.7}$$

2.6.4 Ortalama Kesinlik Ortalaması

Algılama modelinin performansını daha kapsamlı ve dengeli bir şekilde değerlendirmeye olanak tanıyan bir ölçüttür. Tek başına kesinlik veya geri çağırma değerlerine bakmak genellikle yeterli değildir; bu nedenle F1-skoru gibi birleşik metrikler, modelin hem pozitif örnekleri doğru tahmin etme yeteneğini hem de bu örnekleri eksiksiz yakalama başarısını tek bir çatı altında birleştirir. Bu ölçüt, özellikle sınıflar arası dengesizlik bulunan veri kümelerinde modelin gerçek performansını daha doğru bir şekilde ortaya koyar. Söz konusu hesaplama yaklaşımı denklemde verilmiştir.

$$mAP = \frac{1}{N}\sum_{k=1}^{N}(AP)^k \tag{2.8}$$

Denklem (2.8)'de; N tespit edilecek sınıf sayısını, $(AP)^k$ ise k sınıfı için ortalama başarı değerini temsil etmektedir [72].

2.6.5 F1 Skoru

Kesinlik ve geri çağırma değerlerinin harmonik ortalaması olarak tanımlanan birleşik bir performans ölçütüdür. Bu metrik, iki değerin dengesini tek bir sayı üzerinden değerlendirme imkânı sunarak modelin genel tespit başarısını daha bütüncül bir şekilde yansıtır. Değer aralığı 0 ile 1 arasında olup, 1'e yakın sonuçlar modelin daha yüksek doğruluk ve tutarlılıkla çalıştığını gösterir. Özellikle sınıf dengesizliğinin bulunduğu veri kümelerinde, model performansını objektif biçimde analiz etmek için sıklıkla tercih edilen güvenilir bir metriktir. F1-Skoru'nun matematiksel hesaplama yöntemi (2.9) numaralı denklemde sunulmuştur [73].

$$F1 = 2 * \frac{Pre * Re}{Pre + Re} \tag{2.9}$$

2.6.6 Jaccard Katsayısı

Jaccard Katsayısı (İng. Intersection over Union, IoU), nesne tespiti ve görüntü bölütleme görevlerinde model tarafından tahmin edilen bölge ile gerçek (İng. ground truth) bölge arasındaki örtüşme oranını ölçen temel bir performans metriğidir. Bu değer, tahmin edilen alan

ile gerçek alanın kesişiminin, bu iki alanın birleşimine oranlanmasıyla hesaplanır. IoU, yalnızca doğru sınıflandırmayı değil, aynı zamanda tahminin konum ve sınır doğruluğunu da dikkate alarak model performansına ilişkin daha kapsamlı bir değerlendirme sunar. Bu nedenle, özellikle hassas sınır tespiti gerektiren uygulamalarda yaygın olarak kullanılmaktadır. IoU denklem gösterimi (2.10) numaralı denklemde gösterilmiştir [74].

$$IoU = \frac{TP}{TP + FP + FN} \tag{2.10}$$

IoU değeri 0 ile 1 arasında değişmekte olup, yüksek IoU değerleri modelin tahminlerinin gerçeğe daha yakın olduğunu göstermektedir. Nesne tespiti literatüründe, bir tahminin doğru kabul edilmesi için genellikle belirli bir IoU eşik değeri (örneğin 0.5) kullanılır. Bu eşik değerine bağlı olarak doğru ve yanlış tespitler belirlenmekte; böylece Precision ve Recall metrikleri hesaplanmaktadır. Ayrıca farklı IoU eşiklerinde elde edilen Precision–Recall eğrilerinin alanı alınarak ortalama doğruluk ölçütü olan mean Average Precision (mAP) değeri elde edilmektedir. IoU'nun matematiksel ifadesi aşağıda verilen formül ile tanımlanmaktadır.

2.7 KAYIP FONKSİYONU

Kayıp fonksiyonu (İng. loss function), model tarafından üretilen tahminler ile gerçek etiketler arasındaki farkı nicel olarak ölçen ve eğitim sürecinde parametrelerin güncellenmesini yönlendiren temel bir bileşendir. Derin öğrenme tabanlı görüntü bölütlemesi ve nesne tespiti problemlerinde, kayıp fonksiyonunun seçimi; modelin öğrenme başarısını, sınıflar arası dengeyi ve nadir olayların tespit edilebilirliğini doğrudan etkilemektedir. Bu çalışmada, piksel bazlı sınıflandırma hatalarını ölçmek ve sınıf dengesizliği problemini ele almak amacıyla ikili çapraz entropi kayıp fonksiyonu (İng. Binary cross-entropy loss, BCE) ve odaklanmış kayıp fonksiyonu (İng. Focal loss) kullanılmıştır.

2.7.1 İkili Çapraz Entropi Kayıp Fonksiyonu

İkili çapraz entropi kayıp fonksiyonu (BCE), ikili sınıflandırma ve ikili bölütleme problemlerinde yaygın olarak kullanılan temel bir kayıp fonksiyonudur. BCE, her bir piksel için modelin tahmin ettiği olasılık değeri ile gerçek etiket arasındaki farkı logaritmik olarak ölçer. Bu yaklaşım, yanlış sınıflandırılan pikselleri daha yüksek oranda cezalandırarak modelin doğru sınıf olasılıklarına yakınsamasını sağlar. Ancak, sınıf dengesizliğinin yoğun olduğu veri kümelerinde, baskın sınıfların modele hâkim olması nedeniyle BCE kaybının sınırlı kalabildiği

bilinmektedir. BCE kayıp fonksiyonunun matematiksel ifadesi Denklem (2.11)'de verilmiştir:

$$L_{\{BCE\}} = -\left(\frac{1}{N}\right)\sum_{i=1}^{N}[\, y_i log(p_i) + (1 - y_i) log(1 - p_i)] \tag{2.11}$$

2.7.2 Odaklanmış Kayıp Fonksiyonu

Odaklanmış kayıp fonksiyonu, özellikle sınıf dengesizliğinin belirgin olduğu durumlarda, modelin kolay örnekler yerine zor örneklere odaklanmasını sağlamak amacıyla geliştirilmiş bir kayıp fonksiyonudur. BCE kaybına eklenen ayarlama terimi sayesinde, doğru sınıflandırılmış ve yüksek güvene sahip örneklerin katkısı azaltılırken, yanlış sınıflandırılmış örneklerin kayıp değerine olan etkisi artırılmaktadır. Bu özellik, nadir görülen anomalilerin daha etkin bir şekilde öğrenilmesine katkı sağlamaktadır. Odaklanmış kayıp fonksiyonunun ($L_{\{Focal\}}$) matematiksel ifadesi Denklem (2.12)'de sunulmaktadır:

$$L_{\{Focal\}} = \sum_{i=1}^{N}[\alpha\,(1 - p_i)^{\gamma}(\, y_i) log(\, p_i)] \tag{2.12}$$

3 BULGULAR VE TARTIŞMA

3.1 İLK ÇALIŞMALAR

Önceki çalışmada [75], yaklaşık 17.000 görselden oluşan ve altı farklı anomali türünü içeren veri kümeleri kullanılmıştır. Bu veri kümeleri üzerinde, beş farklı mimari ve üç farklı omurga (İng. backbone) ile model eğitimi gerçekleştirilmiştir. Elde edilen her bir metrikteki en iyi sonuçlar sırasıyla Tablo 3.1, Tablo 3.2, Tablo 3.3 ve Tablo 3.4'te kaydedilmiştir.

Tablo 3.1'de sunulan doğruluk sonuçları incelendiğinde, EfficientNet omurgasının farklı mimarilerle birlikte çoğu anomali türünde en yüksek performansı sağladığı görülmektedir. Özellikle "Silahlı Saldırılar" sınıfında DeepLabV3+–EfficientNet kombinasyonu %99,2 doğruluk oranı ile en başarılı sonuçları vermiştir. "Ev Hırsızlığı" ve "Araç Kazaları" olaylarında ise FPN mimarisinin EfficientNet omurgasıyla birlikte öne çıktığı dikkat çekmektedir. Buna karşılık, "Orman Yangınları" sınıfında U-Net mimarisinin MobileNet omurgası ile daha başarılı olduğu görülmekte olup, bu durum olay türüne bağlı olarak mimari–omurga uyumunun performans üzerinde belirleyici olduğunu göstermektedir.

Tablo 3.1. İlk Veri Kümesinde En İyi Doğruluk Değerleri

Anomali	**En İyi Sonuçlar (Doğruluk Skoruna Göre)**		
	Mimari	**Omurga**	**Skor**
Ev Hırsızlığı (Burglary)	FPN	EfficientNet	0,974
Araç Kazaları (Crash)	FPN	EfficientNet	0,907
Kavga-Dövüş (Fighting)	DeepLabV3+	EfficientNet	0,904
Silahlı Saldırılar (Gun)	DeepLabV3+	EfficientNet	0,992
Market Soygunları (Robbery)	U-Net	EfficientNet	0,941
Orman Yangınları (Wildfire)	U-Net	MobileNet	0,885

Tablo 3.2'de yer alan kesinlik (precision) değerleri, farklı anomali türleri için yanlış pozitiflerin minimize edilmesinde mimari ve omurga seçiminin kritik rol oynadığını ortaya koymaktadır.

“Ev Hırsızlığı” sınıfında FPN–EfficientNet kombinasyonu en yüksek kesinlik değerini sağlarken, “Silahlı Saldırılar” sınıfında DeepLabV3+ mimarisinin ResNet omurgası ile daha başarılı olduğu görülmektedir. “Araç Kazaları” olayında PSPNet mimarisinin öne çıkması, bu sınıf için çok ölçekli bağlam bilgisinin daha etkili olduğunu düşündürmektedir. Genel olarak değerlendirildiğinde, kesinlik metriğinde en iyi sonuçların farklı omurgalar arasında dağıldığı ve tek bir mimarinin tüm olay türleri için baskın olmadığı gözlemlenmektedir.

Tablo 3.2. İlk Veri Kümesinde En İyi Kesinlik Değerleri

Anomali	**En İyi Sonuçlar (Kesinlik Skoruna Göre)**		
	Mimari	**Omurga**	**Skor**
Ev Hırsızlığı (Burglary)	FPN	EfficientNet	0,928
Araç Kazaları (Crash)	PSPNET	EfficientNet	0,806
Kavga-Dövüş (Fighting)	DeepLabV3+	EfficientNet	0,825
Silahlı Saldırılar (Gun)	DeepLabV3+	ResNet	0,935
Market Soygunları (Robbery)	FPN	ResNet	0,813
Orman Yangınları (Wildfire)	U-Net	MobileNet	0,817

Tablo 3.3’te sunulan geri çağırma sonuçları incelendiğinde, özellikle olayların eksiksiz yakalanmasının kritik olduğu senaryolarda DeepLabV3+ ve U-Net mimarilerinin ön plana çıktığı görülmektedir. “Silahlı Saldırılar” sınıfında DeepLabV3+–EfficientNet kombinasyonu %95,3 geri çağırma değeri ile en yüksek performansı sergilemiştir. “Ev Hırsızlığı” ve “Market Soygunları” olaylarında ise U-Net mimarisinin ResNet omurgasıyla birlikte daha fazla pozitif örneği doğru şekilde tespit edebildiği gözlemlenmektedir. Bu bulgular, U-Net mimarisinin özellikle sınırlı veri içeren veya sınırları belirgin olay türlerinde avantaj sağladığını göstermektedir.

Tablo 3.3. İlk Veri Kümesinde En İyi Geri Çağırma Değerleri

Anomali	En İyi Sonuçlar (Geri Çağırma Skoruna Göre)		
	Mimari	**Omurga**	**Skor**
Ev Hırsızlığı (Burglary)	UNet	ResNet	0,891
Araç Kazaları (Crash)	DeepLabV3+	EfficientNet	0,792
Kavga-Dövüş (Fighting)	FPN	EfficientNet	0,847
Silahlı Saldırılar (Gun)	DeepLabV3+	EfficientNet	0,953
Market Soygunları (Robbery)	UNet	ResNet	0,821
Orman Yangınları (Wildfire)	UNet	EfficientNet	0,811

Tablo 3.4'te verilen F1 skorları, doğruluk ve geri çağırma arasındaki dengeyi yansıtarak genel model performansının daha bütüncül bir değerlendirmesini sunmaktadır. "Silahlı Saldırılar" sınıfında DeepLabV3+–EfficientNet kombinasyonu, en yüksek F1 skoru ile bu olay türünde en dengeli performansı sağlamıştır. "Ev Hırsızlığı" ve "Kavga-Dövüş" sınıflarında ise FPN mimarisinin EfficientNet omurgasıyla birlikte istikrarlı ve yüksek F1 değerleri ürettiği görülmektedir. "Orman Yangınları" olayında U-Net mimarisinin EfficientNet omurgasıyla öne çıkması, bu mimarinin hem kesinlik hem de geri çağırma arasında daha dengeli bir yapı sunduğunu göstermektedir.

Tablo 3.4. İlk Veri Kümesinde En İyi F1 Skor Değerleri

Anomali	En İyi Sonuçlar (F1-Skoruna Göre)		
	Mimari	**Omurga**	**Skor**
Ev Hırsızlığı (Burglary)	FPN	EfficientNet	0,900
Araç Kazaları (Crash)	DeepLabV3+	EfficientNet	0,727
Kavga-Dövüş (Fighting)	FPN	EfficientNet	0,799

Silahlı Saldırılar (Gun)	DeepLabV3+	EfficientNet	0,931
Market Soygunları (Robbery)	FPN	EfficientNet	0,791
Orman Yangınları (Wildfire)	UNet	EfficientNet	0,742

3.2 KAPSAMLI ÇALIŞMALAR

Yapılan ölçümler sonucunda da görüldüğü üzere, tüm mimari – omurga eşleşmeleri arasında öne çıkan mimariler FPN ve DeepLabV3+ olmuştur. Böylelikle bu mimariler ve omurgalar kullanılarak en iyi sonuçların alınması hedeflenmiştir.

Önceki ölçümlere ek olarak, veri kümesine 4 anomali türü daha eklenip, toplam 10 anomali türü içeren nihai veri kümesi üzerinden ablasyon çalışması yapılmıştır. Mimari ve omurga izolasyonlarının sağlanması için oluşturulan 6 farklı model yeni veri kümesi üzerinde değerlendirilmiştir.

Tablo 3.5'te, ablasyon çalışması kapsamında mimari ve omurga izolasyonunu sağlamak amacıyla oluşturulan altı farklı modelin yapısal bileşenleri sunulmuştur. Bu modellerde, önceki deneysel sonuçlar doğrultusunda geri çağırma metriğinde öne çıkan FPN ve DeepLabV3+ mimarileri, farklı derinlik ve hesaplama maliyetlerine sahip ResNet18, EfficientNet-b0 ve MobileNetV2 omurgaları ile eşleştirilmiştir. Bu yapılandırma sayesinde, mimari ve omurga kombinasyonlarının özellikle olayların eksiksiz tespiti açısından model performansına olan etkisi sistematik olarak incelenmiştir. Oluşturulan model kümesi, daha sonraki tablolarda sunulan performans karşılaştırmalarının adil ve tutarlı bir biçimde gerçekleştirilmesine olanak sağlamaktadır.

Tablo 3.5. Model Tanımları

Model	**Mimari**	**Omurga**
Model 1	FPN	ResNet18
Model 2	FPN	EfficientNet-b0
Model 3	FPN	MobileNetV2

Model 4	DeepLabV3+	ResNet18
Model 5	DeepLabV3+	EfficientNet-b0
Model 6	DeepLabV3+	MobileNetV2

Gerçekleştirilen ablasyon çalışmasında her bir model, NVIDIA GEFORCE RTX 5090 grafik işlemci birimi (GPU) kullanılarak eğitilmiştir. Eğitilen modellerin performans değerleri, sırasıyla doğruluk, kesinlik, geri çağırma, dice katsayısı ve IoU metrikleri için Tablo 3.6, Tablo 3.7, Tablo 3.8 ve Tablo 3.9'da kaydedilmiştir.

Tablo 3.6 incelendiğinde, "Çığ", "Hortum" ve "Tsunami" anomali türlerinde en yüksek doğruluk değerlerinin Model 4 (DeepLabV3+ – ResNet18) tarafından elde edildiği görülmektedir. "Ev Hırsızlığı" ve "Araç Kazaları" olaylarında ise Model 1 (FPN – ResNet18) ve Model 6 (DeepLabV3+ – MobileNetV2) benzer ve en yüksek doğruluk değerlerini üretmiştir. "Silahlı Saldırılar" ve "Toprak Kayması" sınıflarında Model 4 öne çıkarken, "Orman Yangınları" için en iyi doğruluk değeri Model 3 (FPN – MobileNetV2) tarafından elde edilmiştir. Bu sonuçlar, doğruluk metriğinde DeepLabV3+ mimarisinin özellikle ResNet18 omurgası ile daha istikrarlı sonuçlar sunduğunu göstermektedir.

Tablo 3.6. Hibrit Veri Kümesinin Doğruluk Değerleri

Anomali	**Sonuçlar (Doğruluk Skoruna Göre)**							
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Ens. N = 15**	**Ens. N = 30**
Çığ	0,975	0,972	0,974	0,976	0,974	0,971	0,942	0,941
Ev Hırsızlığı	0,967	0,962	0,962	0,965	0,962	0,963	0,954	0,954
Araç Kazaları	0,913	0,910	0,912	0,912	0,912	0,913	0,905	0,905
Kavga-Dövüş	0,894	0,889	0,891	0,894	0,888	0,891	0,883	0,885
Silahlı Saldırılar	0,984	0,981	0,981	0,985	0,982	0,982	0,963	0,963

Toprak Kayması	0,981	0,980	0,981	0,983	0,981	0,979	0,954	0,954
Market Soygunları	0,930	0,922	0,924	0,930	0,924	0,921	0,901	0,902
Hortum	0,975	0,974	0,973	0,976	0,974	0,974	0,964	0,964
Tsunami	0,980	0,976	0,978	0,980	0,980	0,977	0,977	0,976
Orman Yangınları	0,883	0,881	0,887	0,880	0,874	0,881	0,833	0,837

Tablo 3.7'de sunulan kesinlik sonuçlarına göre, "Çığ" ve "Tsunami" anomali türlerinde en yüksek kesinlik değerleri Model 6 (DeepLabV3+ – MobileNetV2) tarafından elde edilmiştir. "Kavga-Dövüş", "Silahlı Saldırılar" ve "Hortum" olaylarında ise Model 4 (DeepLabV3+ – ResNet18) en başarılı sonuçları vermiştir. "Toprak Kayması" sınıfında Model 3 (FPN – MobileNetV2) öne çıkarken, "Ev Hırsızlığı", "Araç Kazaları", "Market Soygunları" ve "Orman Yangınları" için en yüksek kesinlik değerlerinin Model 1 (FPN – ResNet18) tarafından sağlandığı görülmektedir. Bu dağılım, kesinlik metriğinde mimari–omurga kombinasyonlarının olay türüne duyarlı olduğunu göstermektedir.

Tablo 3.7. Hibrit Veri Kümesinin Kesinlik Değerleri

Anomali	**Sonuçlar (Kesinlik Skoruna Göre)**							
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Ens. N = 15**	**Ens. N = 30**
Çığ	0,964	0,968	0,967	0,968	0,963	0,969	0,874	0,871
Ev Hırsızlığı	0,868	0,851	0,850	0,860	0,852	0,842	0,758	0,756
Araç Kazaları	0,715	0,717	0,712	0,708	0,706	0,698	0,634	0,636
Kavga-Dövüş	0,806	0,799	0,807	0,817	0,799	0,791	0,729	0,735
Silahlı Saldırılar	0,919	0,897	0,899	0,921	0,900	0,895	0,756	0,757
Toprak Kayması	0,939	0,925	0,944	0,938	0,928	0,928	0,773	0,775
Market Soygunları	0,793	0,748	0,747	0,788	0,747	0,72	0,627	0,629

Hortum	0,933	0,927	0,919	0,938	0,928	0,933	0,848	0,848
Tsunami	0,957	0,948	0,953	0,950	0,956	0,957	0,906	0,901
Orman Yangınları	0,765	0,771	0,788	0,743	0,729	0,759	0,619	0,626

Tablo 3.8'deki geri çağırma sonuçları değerlendirildiğinde, "Çığ" anomalisinde en yüksek geri çağırma değerinin Model 1 (FPN – ResNet18) ve Model 4 (DeepLabV3+ – ResNet18) tarafından paylaşıldığı görülmektedir. "Araç Kazaları", "Kavga-Dövüş" ve "Market Soygunları" sınıflarında Model 6 (DeepLabV3+ – MobileNetV2) en iyi performansı sergilemiştir. "Silahlı Saldırılar", "Toprak Kayması", "Hortum" ve "Tsunami" olaylarında ise Model 4 (DeepLabV3+ – ResNet18) en yüksek geri çağırma değerlerine ulaşmıştır. "Orman Yangınları" için de yine Model 4 öne çıkmakta olup, bu durum DeepLabV3+ mimarisinin olay kaçırmama açısından avantaj sağladığını göstermektedir.

Tablo 3.8. Hibrit Veri Kümesinin Geri Çağırma Değerleri

Anomali	**Sonuçlar (Geri Çağırma Skoruna Göre)**							
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Ens. N = 15**	**Ens. N = 30**
Çığ	0,972	0,96	0,967	0,972	0,97	0,958	0,995	0,995
Ev Hırsızlığı	0,875	0,856	0,859	0,866	0,856	0,873	0,939	0,943
Araç Kazaları	0,651	0,606	0,647	0,657	0,663	0,686	0,786	0,784
Kavga-Dövüş	0,823	0,816	0,809	0,806	0,811	0,838	0,936	0,933
Silahlı Saldırılar	0,943	0,944	0,944	0,957	0,946	0,950	0,992	0,993
Toprak Kayması	0,938	0,946	0,932	0,951	0,952	0,937	0,993	0,992
Market Soygunları	0,750	0,760	0,782	0,760	0,785	0,819	0,915	0,917
Hortum	0,931	0,933	0,935	0,935	0,935	0,927	0,981	0,981
Tsunami	0,938	0,927	0,931	0,946	0,940	0,925	0,983	0,984

Orman Yangınları	0,797	0,773	0,773	0,822	0,820	0,793	0,932	0,927

Tablo 3.9'daki Dice skorları incelendiğinde, "Çığ", "Silahlı Saldırılar", "Toprak Kayması" ve "Hortum" anomali türlerinde en yüksek F1 skorlarının Model 4 (DeepLabV3+ – ResNet18) tarafından elde edildiği görülmektedir. "Araç Kazaları" sınıfında Model 6 (DeepLabV3+ – MobileNetV2) en iyi sonucu verirken, "Ev Hırsızlığı" ve "Market Soygunları" için Model 1 (FPN – ResNet18) öne çıkmıştır. "Tsunami" olayında Model 4 ve Model 5 (DeepLabV3+ – EfficientNet-b0) eşdeğer en yüksek Dice skoruna ulaşmıştır. "Orman Yangınları" için ise Model 3 (FPN – MobileNetV2) en dengeli performansı sunmuştur.

Tablo 3.9. Hibrit Veri Kümesinin Dice Katsayıları

Anomali	**Sonuçlar (Dice Skoruna Göre)**							
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Ens. N = 15**	**Ens. N = 30**
Çığ	0,968	0,964	0,967	0,970	0,966	0,963	0,931	0,929
Ev Hırsızlığı	0,871	0,853	0,855	0,863	0,854	0,857	0,839	0,839
Araç Kazaları	0,682	0,657	0,678	0,682	0,684	0,692	0,702	0,702
Kavga-Dövüş	0,815	0,807	0,808	0,812	0,805	0,814	0,820	0,822
Silahlı Saldırılar	0,931	0,920	0,921	0,938	0,922	0,922	0,858	0,859
Toprak Kayması	0,939	0,935	0,938	0,944	0,940	0,932	0,869	0,870
Market Soygunları	0,771	0,754	0,764	0,774	0,766	0,766	0,744	0,746
Hortum	0,932	0,930	0,927	0,937	0,931	0,930	0,910	0,910
Tsunami	0,947	0,937	0,942	0,948	0,948	0,941	0,943	0,941
Orman Yangınları	0,780	0,772	0,781	0,780	0,772	0,776	0,744	0,747

Tablo 3.10'da verilen IoU değerlerine göre, "Çığ", "Silahlı Saldırılar", "Toprak Kayması", "Hortum" ve "Tsunami" anomali türlerinde en yüksek IoU skorları Model 4 (DeepLabV3+ –

ResNet18) tarafından elde edilmiştir. “Araç Kazaları” ve “Market Soygunları” sınıflarında ise Model 6 (DeepLabV3+ – MobileNetV2) en başarılı sonuçları üretmiştir. “Ev Hırsızlığı” için Model 1 (FPN – ResNet18) öne çıkarken, “Kavga-Dövüş” ve “Orman Yangınları” olaylarında yine Model 1 en yüksek IoU değerini sağlamıştır. Bu sonuçlar, genel bölütleme başarısı açısından DeepLabV3+–ResNet18 kombinasyonunun baskın bir performans sergilediğini ortaya koymaktadır.

Tablo 3.10. Hibrit Veri Kümesinin IoU Değerleri

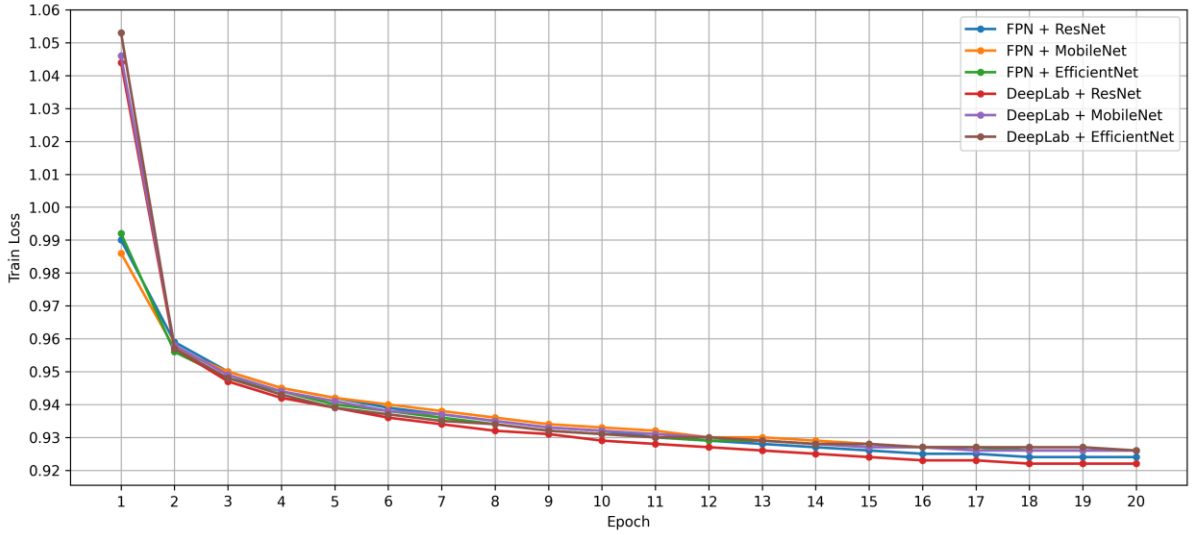
Anomali	**Sonuçlar (IoU Skoruna Göre)**							
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Ens. N = 15**	**Ens. N = 30**
Çığ	0,938	0,931	0,937	0,941	0,935	0,930	0,870	0,868
Ev Hırsızlığı	0,772	0,744	0,746	0,759	0,746	0,751	0,723	0,724
Araç Kazaları	0,517	0,489	0,513	0,517	0,519	0,529	0,541	0,541
Kavga-Dövüş	0,688	0,677	0,678	0,683	0,673	0,686	0,695	0,698
Silahlı Saldırılar	0,871	0,852	0,855	0,884	0,856	0,856	0,752	0,753
Toprak Kayması	0,885	0,879	0,884	0,895	0,887	0,874	0,769	0,771
Market Soygunları	0,628	0,605	0,619	0,631	0,621	0,621	0,592	0,596
Hortum	0,873	0,870	0,864	0,881	0,872	0,870	0,834	0,835
Tsunami	0,900	0,883	0,891	0,902	0,900	0,889	0,893	0,888
Orman Yangınları	0,640	0,629	0,640	0,640	0,629	0,634	0,593	0,596

Ablasyon için kullanılan 6 mimari – omurga çifti için elde edilen eğitim kaybı değerlerinin grafiği Şekil 3.1’de verilmiştir.

Topluluk öğrenmesi yaklaşımının farklı sayıda bağımsız model çıktısı kullanılarak oluşturulmuş iki ayrı konfigürasyonunu temsil etmektedir. Bu çalışmada N parametresi, aynı mimari yapı korunarak farklı ağırlık başlangıçları ve bağımsız eğitim süreçleri ile elde edilen

model çıktılarının sayısını ifade etmektedir. Metrik sonuçlar üzerindeki etkisini incelemek amacıyla, Ensemble yapısı 15 ve 30 bağımsız model çıktısı kullanılarak ayrı ayrı değerlendirilmiştir. Bu karşılaştırma, ağırlık başlatma farklılıklarının ve model çeşitliliğinin Ensemble performansına olan katkısını analiz etmeyi amaçlamaktadır.

Elde edilen sonuçlar, bağımsız model sayısının artırılmasının özellikle geri çağırma gibi kapsayıcılığı ölçen metriklerde daha kararlı ve yüksek değerlere ulaşılmasını sağladığını göstermektedir. Buna karşın, doğruluk ve kesinlik metriklerinde belirli bir noktadan sonra performans artışının sınırlı kaldığı ve Ensemble boyutunun büyümesiyle birlikte marjinal kazanımların azaldığı gözlemlenmiştir. Bu bulgu, Ensemble büyüklüğünün artırılmasının her zaman doğrusal bir performans artışı sağlamadığını da ortaya koymaktadır.



Şekil 3.1. Modellerin eğitim kaybı sonuçları

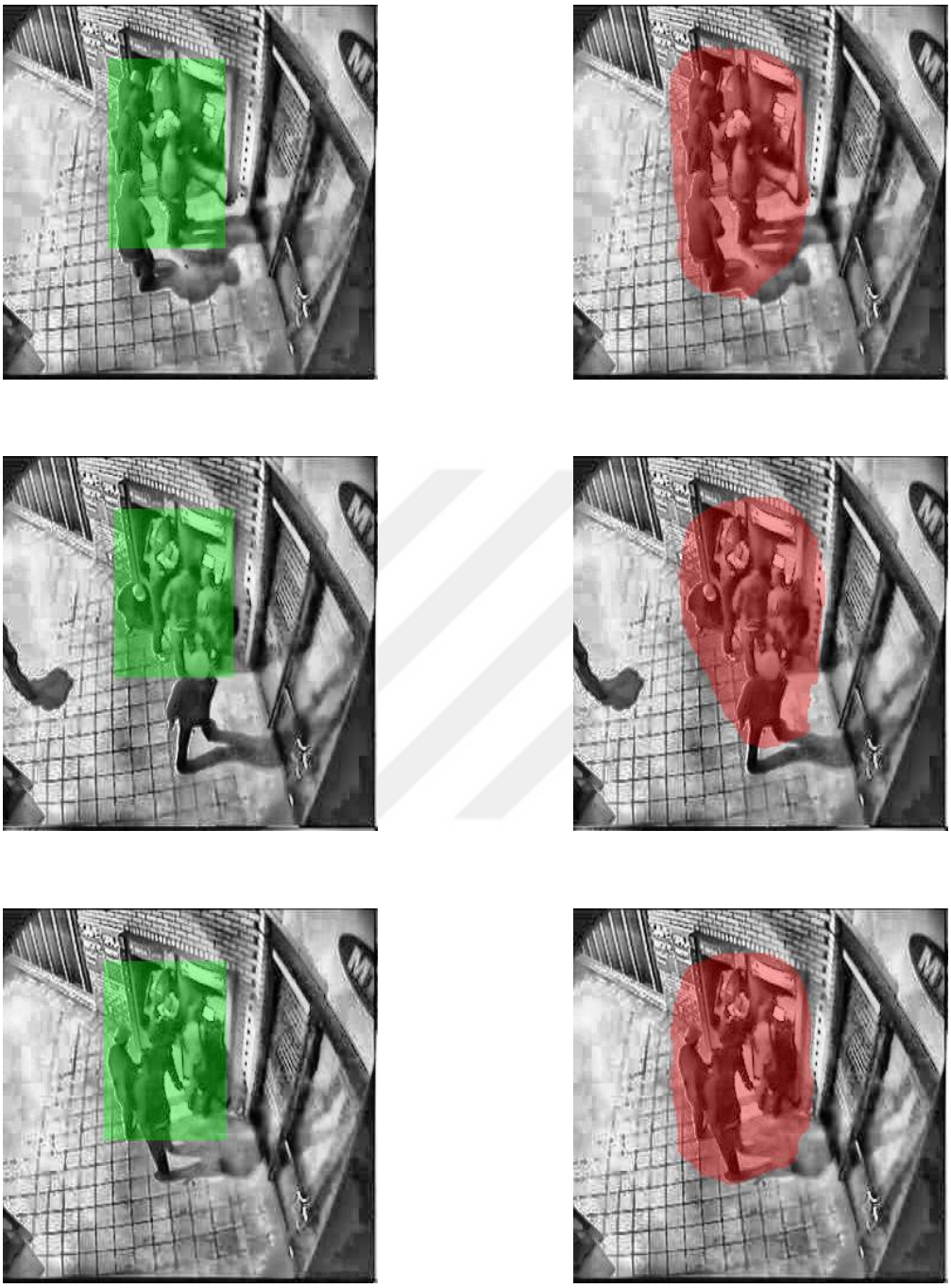
3.2.1 Gerçek Değer ve Tahmin Sonuçları

Şekil 3.2 ila Şekil 3.11'de sunulduğu üzere, eğitilen modeller arasından performans ölçütleri açısından en başarılı altı model seçilmiş ve bu modeller için rastgele örneklenen gerçek (İng. ground truth) ve tahmin çıktıları görsel olarak karşılaştırılmıştır. Bu görseller, modellerin yalnızca sayısal metrikler üzerinden değil, uzamsal tutarlılık, sınır doğruluğu ve hedef olay temsili açısından da değerlendirilmesine olanak sağlamaktadır. Böylece, elde edilen sonuçlar model performansının nitel boyutunu ortaya koyarak, farklı mimarilerin ve kayıp fonksiyonlarının görsel çıktı kalitesi üzerindeki etkilerinin daha açık biçimde analiz edilmesini

mümkün kılmaktadır.



Şekil 3.2. Çığ olayına ait gerçek (solda) ve tahmin (sağda) maskelerinin orijinal görüntülerinin üzerine bindirilmiş görüntüleri



Şekil 3.3. Soygun olayına ait gerçek (solda) ve tahmin (sağda) maskelerinin orijinal görüntülerinin üzerine bindirilmiş görüntüleri



Şekil 3.4. Araç Kazaları olayına ait gerçek (solda) ve tahmin (sağda) maskelerinin orijinal görüntülerinin üzerine bindirilmiş görüntüleri



Şekil 3.5. Kavga - Dövüş olayına ait gerçek (solda) ve tahmin (sağda) maskelerinin orijinal görüntülerinin üzerine bindirilmiş görüntüleri



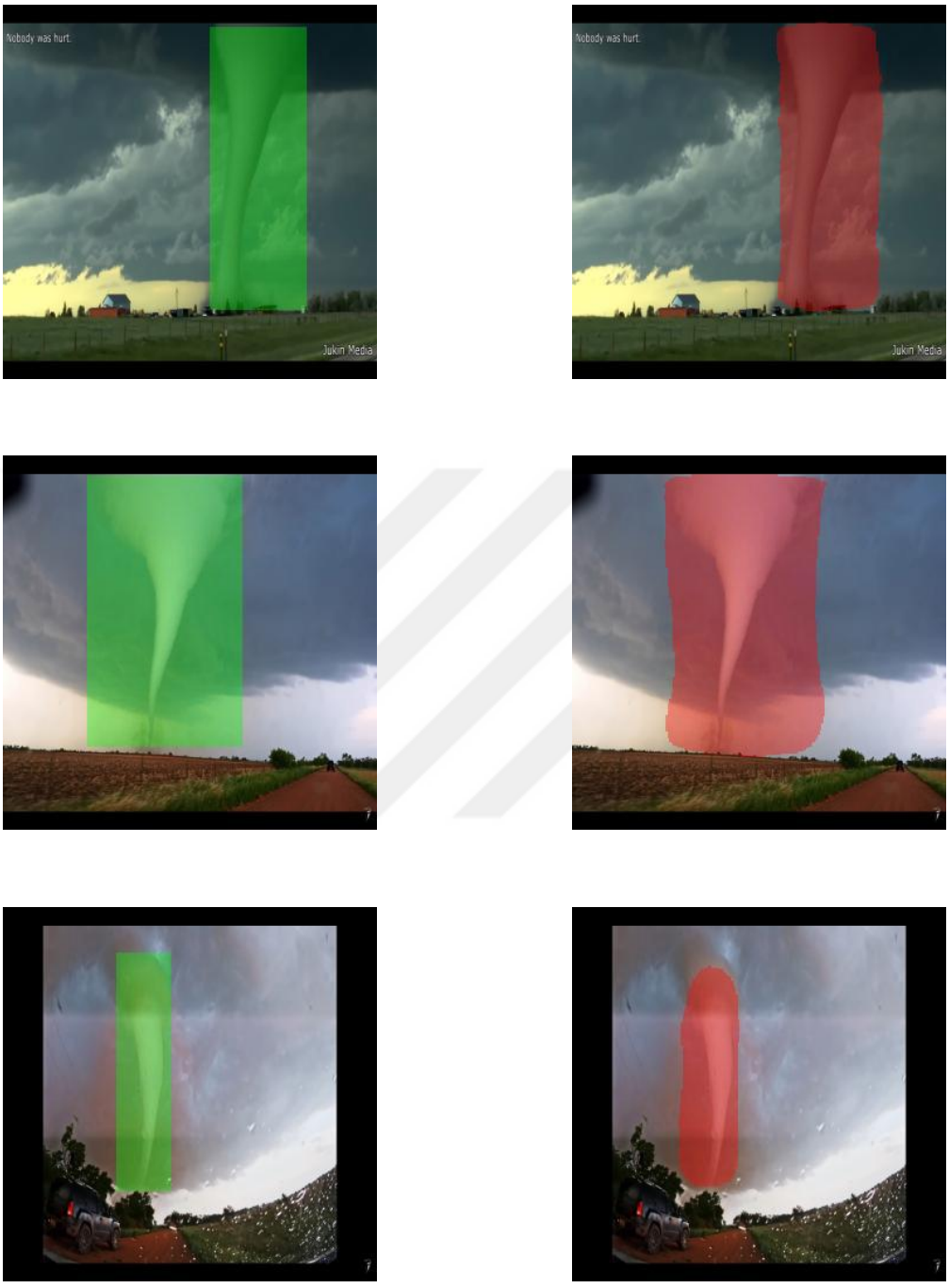
Şekil 3.6. Silah olayına ait gerçek (solda) ve tahmin (sağda) maskelerinin orijinal görüntülerinin üzerine bindirilmiş görüntüleri



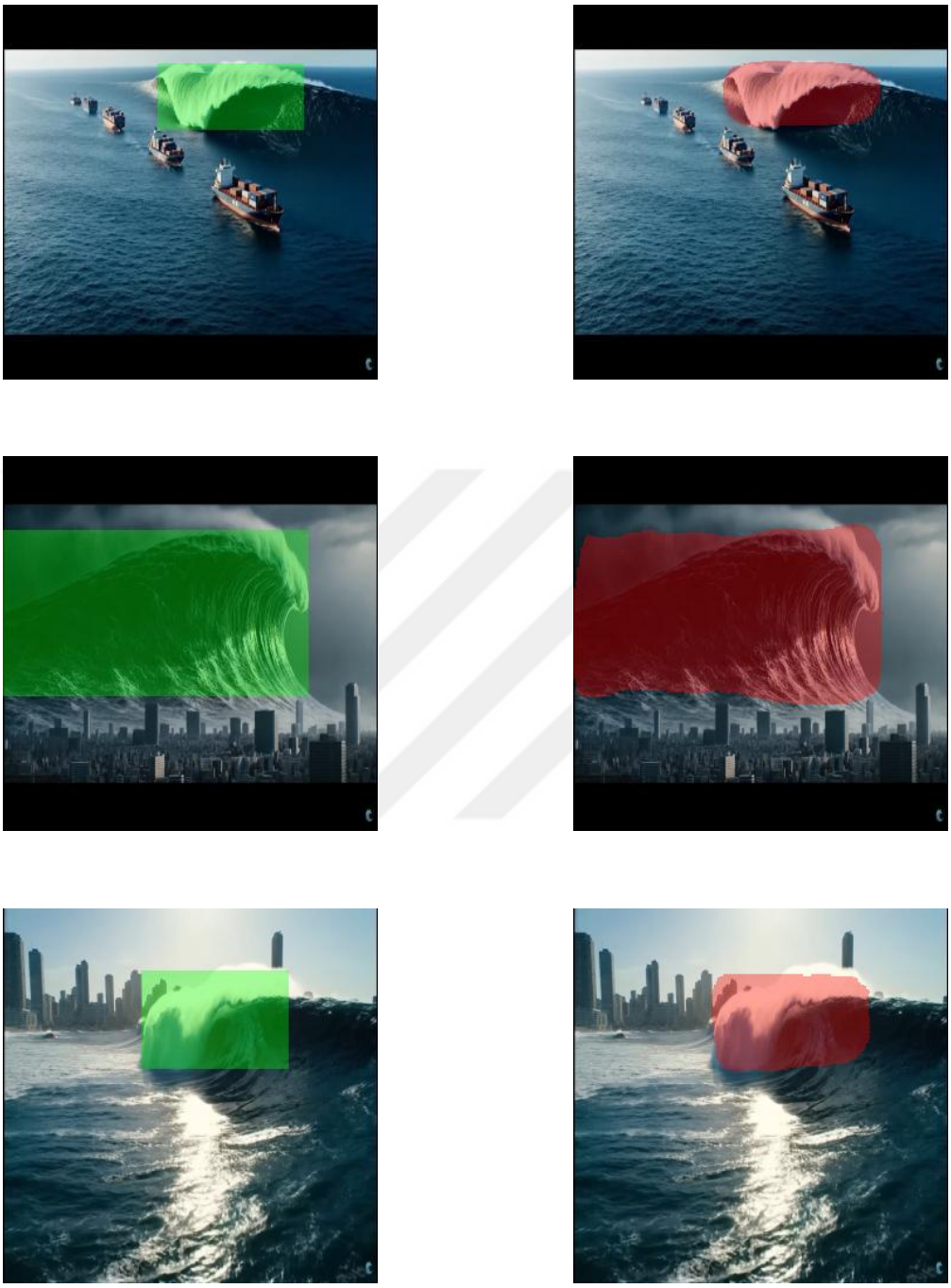
Şekil 3.7. Toprak Kayması olayına ait gerçek (solda) ve tahmin (sağda) maskelerinin orijinal görüntülerinin üzerine bindirilmiş görüntüleri



Şekil 3.8. Soygun olayına ait gerçek (solda) ve tahmin (sağda) maskelerinin orijinal görüntülerinin üzerine bindirilmiş görüntüleri



Şekil 3.9. Hortum olayına ait gerçek (solda) ve tahmin (sağda) maskelerinin orijinal görüntülerinin üzerine bindirilmiş görüntüleri



Şekil 3.10. Tsunami olayına ait gerçek (solda) ve tahmin (sağda) maskelerinin orijinal görüntülerinin üzerine bindirilmiş görüntüleri



Şekil 3.11. Orman Yangını olayına ait gerçek (solda) ve tahmin (sağda) maskelerinin orijinal görüntülerinin üzerine bindirilmiş görüntüleri

4 SONUÇLAR

Elde edilen deneysel sonuçlar, model mimarisi ve omurga ağı seçiminin çok sınıflı anomali bölütleme performansı üzerinde belirgin etkiler yarattığını ortaya koymaktadır. Bununla birlikte, tüm anomali türleri ve tüm değerlendirme metrikleri için her zaman üstün olan tek bir yapılandırmanın bulunmadığı gözlemlenmiştir. Anomali sınıfı başına düşen eğitim verisi miktarının, özellikle geri çağırma ve örtüşme tabanlı metrikler (Dice ve IoU) üzerinde belirleyici bir rol oynadığı; veri miktarının artmasıyla birlikte modellerin olayları eksiksiz yakalama başarısının arttığı görülmüştür. Bu bulgular, sınıf dağılımı ve veri dengesinin, mimari tercihler kadar kritik olduğunu göstermektedir.

Tablo 3.6 ila Tablo 3.10'da sunulan sonuçlar incelendiğinde, tekil modeller arasında DeepLabV3+ mimarisinin ResNet18 omurgasıyla oluşturulan Model 4'ün, doğruluk, geri çağırma, Dice ve IoU metriklerinin büyük bir bölümünde istikrarlı ve güçlü performans sergilediği görülmektedir. Özellikle "Silahlı Saldırılar", "Toprak Kayması", "Hortum" ve "Tsunami" gibi karmaşık sahne yapısına sahip olay türlerinde Model 4'ün yüksek geri çağırma ve IoU değerlerine ulaşması, bu mimarinin olay kaçırmama ve bölütleme hassasiyeti açısından güvenilir bir yapı sunduğunu ortaya koymaktadır. Buna karşılık, bazı sınırlı alan ve daha küçük nesne içeren anomalilerde (örneğin "Ev Hırsızlığı" ve "Market Soygunları") FPN tabanlı modellerin daha rekabetçi kesinlik değerleri üretebildiği gözlemlenmiştir.

Topluluk Öğrenmesi yaklaşımlarını temsil eden 15 ve 30 tekrarlı sonuçların ise, tekil modellere kıyasla metrikler arasında farklı bir denge sunduğunu göstermektedir. Doğruluk ve kesinlik metriklerinde topluluk öğrenmesi modellerinin, özellikle veri dağılımı görece dengeli ve sınırları belirgin anomalilerde tekil modellere kıyasla daha düşük değerler ürettiği görülmektedir. Bununla birlikte, geri çağırma sonuçları incelendiğinde topluluk öğrenmesi modellerinin belirgin bir üstünlük sağladığı dikkat çekmektedir. Özellikle "Çığ", "Silahlı Saldırılar", "Toprak Kayması", "Market Soygunları" ve "Orman Yangınları" gibi olay türlerinde topluluk öğrenmesi modellerinin geri çağırma değerlerinin tüm tekil modellerin üzerine çıkması, bu yaklaşımın olayları kaçırmama konusunda daha agresif ve kapsayıcı bir tespit stratejisi benimsediğini göstermektedir.

Dice ve IoU metrikleri değerlendirildiğinde ise Topluluk öğrenmesi modellerinin, tekil modellere kıyasla genellikle daha düşük veya benzer seviyede sonuçlar ürettiği, ancak özellikle "Araç Kazaları" ve "Kavga-Dövüş" gibi sınırları daha belirsiz olaylarda göreli iyileşmeler

sağladığı gözlemlenmiştir. Farklı tekrar sayıları ile denenen topluluk öğrenmesi modelleri arasındaki farklar incelendiğinde, tekrar sayısının artması bazı anomalilerde geri çağırma açısından marjinal iyileşmeler sunduğu, ancak genel olarak iki topluluk öğrenmesi yapısının arasında dramatik bir fark oluşmadığı görülmektedir.

Sonuç olarak, bu tez çalışması; tekil mimariler arasında DeepLabV3+ tabanlı yapıların bölütleme hassasiyeti açısından öne çıktığını, topluluk öğrenmesi yaklaşımlarının ise özellikle kritik olayların kaçırılmaması gereken senaryolarda geri çağırma odaklı güçlü bir alternatif sunduğunu ortaya koymaktadır. Çalışmadan elde edilen sonuçlar, kullanım amacına göre yüksek hassasiyetli tekil modeller ile geniş kapsamlı topluluk modelleri arasında stratejik bir tercih yapılması gerektiğine işaret etmektedir. Bu yaklaşım, çok sınıflı anomali tespit sistemlerinin geliştirilme sürecinde dikkate alınması gereken temel yapı taşlarını sunmaktadır.

KAYNAKLAR

[1] Kukade, J., & Panse, P. (2024). Advanced Deep Learning Model for Anomaly Detection Based Video Surveillance System. *International Journal of Intelligent Systems and Applications in Engineering*, *12*(5s), 477–485.

[2] *Deep learning based anomaly detection in real-time video | Multimedia Tools and Applications*. (n.d.). Retrieved December 2, 2025, from https://link.springer.com/article/10.1007/s11042-024-19116-9?utm_source=chatgpt.com

[3] Samaila, Y. A., Sebastian, P., Singh, N. S. S., Shuaibu, A. N., Ali, S. S. A., Amosa, T. I., Mustafa Abro, G. E., & Shuaibu, I. (2024). Video anomaly detection: A systematic review of issues and prospects. *Neurocomputing*, *591*, 127726. https://doi.org/10.1016/j.neucom.2024.127726

[4] Wastupranata, L. M., Kong, S. G., & Wang, L. (2024). Deep Learning for Abnormal Human Behavior Detection in Surveillance Videos—A Survey. *Electronics*, *13*(13), 2579. https://doi.org/10.3390/electronics13132579

[5] Duong, H.-T., Le, V.-T., & Hoang, V. T. (2023). Deep Learning-Based Anomaly Detection in Video Surveillance: A Survey. *Sensors*, *23*(11), 5024. https://doi.org/10.3390/s23115024

[6] Shrestha, B., Pant, D. R., Gautam, G., & Shrestha, B. (2025). *Meta Learning Approach for Adaptive Anomaly Detection from Multi Scenario Video Surveillance* (2025051613). Preprints. https://doi.org/10.20944/preprints202505.1613.v1

[7] *UCF Crime Dataset*. (n.d.). Retrieved December 2, 2025, from https://www.kaggle.com/datasets/odins0n/ucf-crime-dataset

[8] Tan, H. (2025). Yapay Zekâ ile İnsan Yakınsamasını Felsefe Düşünürlerinin Yaklaşımları Üzerinden Eleştirel Bir Değerlendirme. *HUMANITAS – Uluslararası Sosyal Bilimler Dergisi*, 13(25), 306–333. https://doi.org/10.20304/humanitas.1602451

[9] Öngün, C. (2014). *Anomaly detection for video surveillance in crowded environments* / [Master Thesis, Middle East Technical University]. https://open.metu.edu.tr/handle/11511/23774

[10] Ke, Y., Sukthankar, R., & Hebert, M. (2007). Event Detection in Crowded Videos. *2007*

IEEE 11th International Conference on Computer Vision, 1–8. https://doi.org/10.1109/ICCV.2007.4409011

[11] Zhou, J. T., Du, J., Zhu, H., Peng, X., Liu, Y., & Goh, R. S. M. (2019). AnomalyNet: An Anomaly Detection Network for Video Surveillance. *IEEE Transactions on Information Forensics and Security*, *14*(10), 2537–2550. https://doi.org/10.1109/TIFS.2019.2900907

[12] Medioni, G., Cohen, I., Bremond, F., Hongeng, S., & Nevatia, R. (2001). Event detection and analysis from video streams. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*(8), 873–889. https://doi.org/10.1109/34.946990

[13] Al-Gayyim, G. M. H. (2024). *ANOMALY DETECTION FROM BIG DATA OF MOTION SENSORS USING DEEP LEARNING*. https://hdl.handle.net/20.500.14619/3409

[14] Taheri Tajar, A., Ramazani, A., & Mansoorizadeh, M. (2021). A lightweight Tiny-YOLOv3 vehicle detection approach. *J. Real-Time Image Process.*, *18*(6), 2389–2401. https://doi.org/10.1007/s11554-021-01131-w

[15] Boyrazlı, H. K., & Çınar, A. (2021). Anomaly Detection in Crowded Scenes With Machine Learning Algorithms. *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi*, *12*(2), 229–235. https://doi.org/10.24012/dumf.849923

[16] Xiang, T., & Gong, S. (2008). Video Behavior Profiling for Anomaly Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*(5), 893–908. https://doi.org/10.1109/TPAMI.2007.70731

[17] Pennisi, A., Bloisi, D. D., & Iocchi, L. (2016). Online real-time crowd behavior detection in video sequences. *Computer Vision and Image Understanding*, *144*, 166–176. https://doi.org/10.1016/j.cviu.2015.09.010

[18] Sultani, W., Chen, C., & Shah, M. (2018). Real-World Anomaly Detection in Surveillance Videos. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6479–6488. https://doi.org/10.1109/CVPR.2018.00678

[19] Sabokrou, M., Fayyaz, M., Fathy, M., Moayedd, Z., & klette, R. (2017). *Deep-Anomaly: Fully Convolutional Neural Network for Fast Anomaly Detection in Crowded Scenes* (arXiv:1609.00866). arXiv. https://doi.org/10.48550/arXiv.1609.00866

[20] Sabokrou, M., Fathy, M., Hosseini, M., & Klette, R. (2015). *Real-Time Anomaly Detection and Localization in Crowded Scenes* (arXiv:1511.06936). arXiv. https://doi.org/10.48550/arXiv.1511.06936

[21] Zhang, X., Yang, S., Zhang, X., Zhang, W., & Zhang, J. (2018). *Anomaly Detection and Localization in Crowded Scenes by Motion-field Shape Description and Similarity-based Statistical Learning* (arXiv:1805.10620). arXiv. https://doi.org/10.48550/arXiv.1805.10620

[22] Lv, P., Liu, S., Xu, M., & Zhou, B. (2018). *Abnormal Event Detection and Location for Dense Crowds using Repulsive Forces and Sparse Reconstruction* (arXiv:1808.06749). arXiv. https://doi.org/10.48550/arXiv.1808.06749

[23] Duong, H.-T., Le, V.-T., & Hoang, V. T. (2023). Deep Learning-Based Anomaly Detection in Video Surveillance: A Survey. *Sensors*, *23*(11), 5024. https://doi.org/10.3390/s23115024

[24] Lee, J.-W., & Kang, H.-S. (2024). Three-Stage Deep Learning Framework for Video Surveillance. *Applied Sciences*, *14*(1), 408. https://doi.org/10.3390/app14010408

[25] Qaraqe, M., Yang, Y. D., Varghese, E. B., Basaran, E., & Elzein, A. (2024). Crowd behavior detection: Leveraging video swin transformer for crowd size and violence level analysis. *Applied Intelligence*, *54*(21), 10709–10730. https://doi.org/10.1007/s10489-024-05775-6

[26] Jebur, S. A., Hussein, K. A., Hoomod, H. K., Alzubaidi, L., Saihood, A. A., & Gu, Y. (2024). *A Scalable and Generalized Deep Learning Framework for Anomaly Detection in Surveillance Videos* (arXiv:2408.00792). arXiv. https://doi.org/10.48550/arXiv.2408.00792

[27] Zhang, Y., Akdag, E., Bondarev, E., & With, P. H. N. D. (2025). *MTFL: Multi-Timescale Feature Learning for Weakly-Supervised Anomaly Detection in Surveillance Videos* (arXiv:2410.05900). arXiv. https://doi.org/10.48550/arXiv.2410.05900

[28] Chidananda, K., & Kumar, A. P. (2024). VidAnomalyNet: An Efficient Anomaly Detection in Public Surveillance Videos Through Deep Learning Architectures. *International Journal of Safety and Security Engineering*, *14*, 953–966. https://doi.org/10.18280/ijsse.140326

[29] Sharif, M. H., Jiao, L., & Omlin, C. W. (2022). *Deep Crowd Anomaly Detection: State-of-the-Art, Challenges, and Future Research Directions* (arXiv:2210.13927). arXiv.

https://doi.org/10.48550/arXiv.2210.13927

[30] Liu, W., Luo, W., Lian, D., & Gao, S. (2018). Future frame prediction for anomaly detection—A new baseline. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 6536–6545). https://doi.org/10.1109/CVPR.2018.00684

[31] Cho, M. A., Kim, T., Kim, W. J., Cho, S., & Lee, S. (2022). Unsupervised video anomaly detection via normalizing flows with implicit latent features. *Pattern Recognition, 129*, 108703. https://doi.org/10.1016/j.patcog.2022.108703

[32] Huang, H., Zhao, B., Gao, F., Chen, P., Wang, J., & Hussain, A. (2023). A novel unsupervised video anomaly detection framework based on optical flow reconstruction and erased frame prediction. *Sensors, 23*(10), 4828. https://doi.org/10.3390/s23104828

[33] Khan, A. A., Nauman, M. A., Shoaib, M., Jahangir, R., Alroobaea, R., Alsafyani, M., Binmahfoudh, A., & Wechtaisong, C. (2022). Crowd anomaly detection in video frames using fine-tuned AlexNet model. *Electronics, 11*(19), 3105. https://doi.org/10.3390/electronics11193105

[34] Khan, S. W., Hafeez, Q., Khalid, M. I., Alroobaea, R., Hussain, S., Iqbal, J., Almotiri, J., & Ullah, S. S. (2022). Anomaly detection in traffic surveillance videos using deep learning. *Sensors, 22*(17), 6563. https://doi.org/10.3390/s22176563

[35] Yang, M., Feng, Y., Rao, A. S., Rajasegarar, S., Tian, S., & Zhou, Z. (2024). Evolving graph-based video crowd anomaly detection. *The Visual Computer, 40*(1), 303–318. https://doi.org/10.1007/s00371-023-02783-4

[36] Habeb, M. H., Salama, M., & Elrefaei, L. A. (2024). Enhancing video anomaly detection using a transformer spatiotemporal attention unsupervised framework for large datasets. *Algorithms, 17*(7), 286. https://doi.org/10.3390/a17070286

[37] Muna, U. M., Biswas, S., Zarif, S. A. A. M., Deori, P. J., Tajwar, T., & Shatabda, S. (2025). Vision transformer embedded video anomaly detection using attention driven recurrence. *Array, 27*, 100471. https://doi.org/10.1016/j.array.2025.100471

[38] Elmetwally, A., Eldeeb, R., & Elmougy, S. (2024). Deep learning based anomaly detection in real-time video. *Multimedia Tools and Applications, 84*(11), 9555–9571.

https://doi.org/10.1007/s11042-024-19116-9

[39] Mahareek, E. A., Elsaid, E. K., El-Desouky, N. M., & El-dahshan, K. A. (2024). Survey: ANOMALY DETECTION IN SURVEILLANCE VIDEOS. *International Journal of Theoretical and Applied Research*, *3*(1), 328–342. https://doi.org/10.21608/ijtar.2024.229847.1073

[40] Roboflow. (2024). Roboflow: Computer Vision Annotation and Dataset Management Platform. https://roboflow.com

[41] Lin, T., et al. (2014). Microsoft COCO: Common Objects in Context. Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part V, 740–755. https://doi.org/10.1007/978-3-319-10602-1_48

[42] Jain, S., et al. (2022). Object Detection Using COCO Dataset. 2022 International Conference on Cyber Resilience (ICCR), 1–4. https://doi.org/10.1109/ICCR56254.2022.9995808

[43] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. 2016 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 779–788. https://doi.org/10.1109/CVPR.2016.91

[44] Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. arXiv:1804.02767. https://doi.org/10.48550/arXiv.1804.02767

[45] Ultralytics. (2023). YOLOv8 Documentation. https://docs.ultralytics.com/

[46] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

[47] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going Deeper with Convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1–9. https://doi.org/10.1109/CVPR.2015.7298594

[48] Guo, Y., Liu, Y., Georgiou, T., & Lew, M. S. (2018). A Review of Semantic Segmentation Using Deep Neural Networks. *International Journal of Multimedia Information Retrieval*, 7,

87–93. https://doi.org/10.1007/s13735-018-0141-2

[49] Zhang, Z., Liu, Q., & Wang, Y. (2018). Road Extraction by Deep Residual U-Net. *IEEE Geoscience and Remote Sensing Letters*, 15(5), 749–753. https://doi.org/10.1109/LGRS.2018.2802944

[50] Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B., & Rueckert, D. (2018). Attention U-Net: Learning Where to Look for the Pancreas. *Medical Image Analysis*, 53, 55–68. https://doi.org/10.1016/j.media.2019.01.007

[51] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4700–4708. https://doi.org/10.1109/CVPR.2017.243

[52] Jegou, S., Drozdzal, M., Vazquez, D., Romero, A., & Bengio, Y. (2017). The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 11–19.
https://doi.org/10.1109/CVPRW.2017.156

[53] Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature Pyramid Networks for Object Detection. 2017 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 936–944. https://doi.org/10.1109/CVPR.2017.106

[54] Chen, L., et al. (2014). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. *arXiv*. arXiv:1412.7062. https://doi.org/10.48550/arXiv.1412.7062

[55] Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). Imagenet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems (NeurIPS 2012)*, 25.

[56] Everingham, M., et al. (2010). The PASCAL Visual Object Classes (VOC) Challenge. International Journal of Computer Vision, 88, 303–338. https://doi.org/10.1007/s11263-009-0275-4

[57] Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. *arXiv*. arXiv:1802.02611. https://doi.org/10.48550/arXiv.1802.02611

[58] Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid Scene Parsing Network. 2017 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2881–2890. https://doi.org/10.1109/CVPR.2017.660 AI Summer School 2025

[59] Tan, M., & Le, Q. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. arXiv:1905.11946. https://doi.org/10.48550/arXiv.1905.11946

[60] Tan, M., & Le, Q. (2021). EfficientNetV2: Smaller Models and Faster Training. *Proceedings of the 38th International Conference on Machine Learning (ICML).* https://arxiv.org/abs/2104.00298

[61] Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. M. (2020). Scaled-YOLOv4: Scaling Cross Stage Partial Network. arXiv:2011.08036. https://arxiv.org/abs/2011.08036

[62] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv:1704.04861. https://doi.org/10.48550/arXiv.1704.04861

[63] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4510–4520. https://doi.org/10.1109/CVPR.2018.00474

[64] Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q. V., & Adam, H. (2019). Searching for MobileNetV3. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 1314–1324. https://arxiv.org/abs/1905.02244

[65] Zhou, B., Khosla, A., Lapedriza, Á., Oliva, A., & Torralba, A. (2016). Learning Deep Features for Discriminative Localization. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2921–2929. https://doi.org/10.1109/CVPR.2016.319

[66] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image

Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. https://doi.org/10.1109/CVPR.2016.90

[67] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Identity Mappings in Deep Residual Networks. *Proceedings of the European Conference on Computer Vision (ECCV)*, 630–645. https://doi.org/10.1007/978-3-319-46493-0_38

[68] Zagoruyko, S., & Komodakis, N. (2016). Wide Residual Networks. *Proceedings of the British Machine Vision Conference (BMVC).* https://arxiv.org/abs/1605.07146

[69] Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated Residual Transformations for Deep Neural Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1492–1500. https://doi.org/10.1109/CVPR.2017.11

[70] Esri. (2024). How Compute Accuracy for Object Detection Works. *ArcGIS Pro Documentation – Image Analyst*. https://pro.arcgis.com/en/pro-app/latest/tool-reference/image-analyst/how-compute-accuracy-for-object-detection-works.htm

[71] Google. (2024). Accuracy, Precision, and Recall. *Google Developers – Machine Learning Crash Course*. https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall

[72] V7 Labs. (2024). Mean Average Precision (mAP) Explained. *V7 Labs Blog*. https://www.v7labs.com/blog/mean-average-precision

[73] Roboflow. (2024). F1 Score Explained. *Roboflow Blog*. https://blog.roboflow.com/f1-score/

[74] V7 Labs. (2024). Intersection over Union (IoU): A Practical Guide. *V7 Labs Blog*. https://www.v7labs.com/blog/intersection-over-union-guide

[75] Şen, Z. A., Sarıkaş, A., & Yıldız, K. (2025). Multi-Scenario Anomaly Detection with a Hybrid Dataset: A Benchmark and Comparative Study. *ASYU 2025 – Akıllı Sistemlerde Yenilikler ve Uygulamaları Konferansı*. https://doi.org/10.1109/ASYU67174.2025.11208327