T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE DENİZ ÜZERİNDEKİ İNSANLARIN TESPİTİ

Murat TETİK

YÜKSEK LİSANS TEZİ
Elektronik ve Haberleşme Mühendisliği Anabilim Dalı
Haberleşme Programı

Danışman
Doç. Dr. Bülent BOLAT

Ocak, 2026

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE DENİZ ÜZERİNDEKİ İNSANLARIN TESPİTİ

Murat TETİK tarafından hazırlanan tez çalışması 05.01.2026 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Elektronik ve Haberleşme Mühendisliği Anabilim Dalı Haberleşme Programı **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Doç. Dr. Bülent BOLAT
Yıldız Teknik Üniversitesi
Danışman

Jüri Üyeleri

Doç. Dr. Bülent BOLAT, Danışman
Yıldız Teknik Üniversitesi _____

Prof. Dr. Umut Engin AYTEN, Üye
Yıldız Teknik Üniversitesi _____

Doç. Dr. Muhammet Ali KARABULUT, Üye
Milli Savunma Üniversitesi _____

Danışmanım Doç. Dr. Bülent BOLAT sorumluluğunda tarafımca hazırlanan Makine Öğrenmesi Yöntemleri ile Deniz Üzerindeki İnsanların Tespiti başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Murat TETİK

İmza

Aileme ve bana ilham veren tüm eğitmenlerime ithafen.

TEŞEKKÜR

Bu tez çalışması süresince bilimsel rehberliği, geri bildirimleri ve sürekli desteği için danışman hocam Doç. Dr. Bülent BOLAT'a teşekkür ederim.

Tez sürecinde motivasyonumu korumamı sağlayan aileme ve arkadaşlarıma minnettarım.

Murat TETİK

İÇİNDEKİLER

SİMGE LİSTESİ vii

KISALTMA LİSTESİ viii

ŞEKİL LİSTESİ x

TABLO LİSTESİ xi

ÖZET xii

ABSTRACT xiv

1 GİRİŞ 1

1.1 Araştırma Konusunun Önemi 1

1.2 Problemin Tanımı 3

1.3 Literatür İncelemesi 5

1.3.1 Geleneksel Görüntü İşleme ve Makine Öğrenmesi Yaklaşımları 5

1.3.2 Derin Öğrenme Tabanlı Yaklaşımlar 5

1.3.3 Literatürdeki Boşluklar ve Tezin Konumu 6

1.3.4 Tez Çalışmasının Hedefleri 6

2 DENİZDE İNSAN TESPİTİNİN ÖNEMİ VE YÖNTEMLER 7

2.1 Yöntem Kategorileri: Gelenekselden Derin Öğrenmeye 7

2.2 Geleneksel Yaklaşımlar 8

2.3 Makine Öğrenmesi Tabanlı Yaklaşımlar 9

2.3.1 Özellik Tabanlı Makine Öğrenmesi 9

2.3.2 Derin Öğrenme Tabanlı Tek ve İki Aşamalı Modeller 9

2.4 Derin Öğrenme Tabanlı Nesne Tespiti Algoritmaları 10

2.4.1 Bölge Tabanlı (İki Aşamalı) CNN Mimarileri 11

2.4.2 Tek Aşamalı Dedektörler 11

2.4.3 Güncel Karşılaştırmalar: Nitel Bir Çerçeve 12

2.4.4 Tez Kapsamı ile İlişkilendirme 12

2.5 Nesne Tespiti Performans Metrikleri 12

2.5.1 Kesişimlerin Birleşime Oranı (IoU) ve Varyantları 13
2.5.2 Kesinlik (Precision), Duyarlılık (Recall) ve F1-Skoru 13
2.5.3 Ortalama Doğruluk (mAP) 14
2.5.4 Bu Tezde Kullanılan Değerlendirme Protokolü 14

3 YOLO (YOU ONLY LOOK ONCE) MODELLERİ 15
3.1 YOLO Modellerinin Tarihsel Gelişimi 16
3.2 YOLOv8 Mimarisi ve Özellikleri 17
3.3 YOLOv9 ve YOLOv11 (Literatür Özeti) 18
3.4 YOLOv10 Mimarisi ve Özellikleri 18
3.5 YOLOv12 Mimarisi (Deneysel Çalışma) 18
3.6 Modellerin Karşılaştırmalı Özeti 19

4 MATERYAL VE METOT 20
4.1 SeaDronesSee Veri Seti 20
4.2 Donanım ve Yazılım Altyapısı 23
4.3 Veri Ön İşleme ve Artırma 24
4.4 Değerlendirme Metodolojisi 25

5 DENEYSEL SONUÇLAR VE TARTIŞMA 27
5.1 Eğitim Süreci ve Metodoloji 27
5.2 Eğitim Dinamikleri ve Kayıp Analizi 29
5.3 Test Sonuçları ve Karşılaştırma 31
5.4 Karmaşıklık Matrisi Değerlendirmesi 35

6 SONUÇ 40
6.1 Genel Değerlendirme ve Çıkarımlar 40
6.2 Denizde İnsan Tespiti İçin Model Önerisi 41
6.3 Kısıtlar ve Gelecek Çalışmalar 42

KAYNAKÇA 43

A DETAYLI DENEYSEL SONUÇLAR 47
A.1 Modellerin Eğitim Süreçleri 47
A.2 Modellerin Detaylı Performans Metrikleri (Test) 47
A.3 Sınıf Bazlı Değerlendirmeler (Test) 48
A.4 Hata Analizleri (Sınıf Bazlı Tahmin Özetleri) 48
A.5 Zorlayıcı Durumlar ve Küçük Nesne Analizi 49

TEZDEN ÜRETİLMİŞ YAYINLAR 51

SİMGE LİSTESİ

lr_0	Başlangıç öğrenme oranı
$\cup$	Birleşim
TN	Doğru Negatif (True Negative)
TP	Doğru Pozitif (True Positive)
R	Duyarlılık (Recall)
λ	Kayıp fonksiyonu ağırlığı
P	Kesinlik (Precision)
$\cap$	Kesişim
σ	Sigmoid aktivasyon fonksiyonu
FN	Yanlış Negatif (False Negative)
FP	Yanlış Pozitif (False Positive)

KISALTMA LİSTESİ

AdamW	Adam with Weight Decay (Ağırlık Sönümlü Adam Eniyileyici)
AMP	Automatic Mixed Precision (Otomatik Karışık Hassasiyet)
AP	Average Precision (Ortalama Hassasiyet)
C2f	CSP Bottleneck with 2 Convolutions - Fast
CNN	Convolutional Neural Network (Evrişimli Sinir Ağı)
COCO	Common Objects in Context
CPU	Central Processing Unit (Merkezi İşlem Birimi)
CSP	Cross Stage Partial (Çapraz Aşama Kısmi Bağlantılar)
CUDA	Compute Unified Device Architecture (Birleşik Aygıt Mimarisi)
DFL	Distribution Focal Loss (Dağılım Odak Kaybı)
FLOPs	Floating Point Operations (Kayan Nokta İşlemleri)
FPN	Feature Pyramid Network (Özellik Piramidi Ağı)
FPS	Frames Per Second (Saniye Başına Kare)
GPU	Graphics Processing Unit (Grafik İşlemci Birimi)
IoU	Intersection over Union (Kesişim Bölü Birleşim)
İHA	İnsansız Hava Aracı
mAP	mean Average Precision (Ortalama Hassasiyet)
NaN	Not a Number (Sayısal Değil)
NMS	Non-Maximum Suppression (Maksimum Olmayan Bastırma)
PAN	Path Aggregation Network (Yol Toplama Ağı)
PANet	Path Aggregation Network (Yol Toplama Ağı)
RAM	Random Access Memory (Rastgele Erişimli Bellek)
R-CNN	Region-based Convolutional Neural Network (Bölge Tabanlı Evrişimli Sinir Ağı)

SAHI	Slicing Aided Hyper Inference (Dilimleme Destekli Çıkarım)
SAR	Search and Rescue (Arama ve Kurtarma)
SPPF	Spatial Pyramid Pooling - Fast (Uzamsal Piramit Havuzlama - Hızlı)
YOLO	You Only Look Once (Sadece Bir Kez Bak)

ŞEKİL LİSTESİ

Şekil 1.1 WHO 2024 Küresel Boğulma İstatistikleri [5] 2
Şekil 1.2 a) Işık koşullarının tespite etkisi, b) Dalga yüksekliği ve deniz durumunun görsel kaliteye etkisi 4
Şekil 2.1 Deniz ortamında insan/nesne tespiti yöntemlerinin taksonomik sınıflandırması. 8
Şekil 2.2 Geleneksel makine öğrenmesi yaklaşımlarının temsili şemaları. (a) HOG öznitelik çıkarımı [16], (b) SVM ile marjin maksimizasyonu ve sınıflandırma [23], (c) Yüz tespiti için kullanılan Haar-benzeri öznitelikler [24], (d) Zayıf sınıflandırıcıların sıralı ağırlıklandırılmasına dayalı AdaBoost yapısı [17]. 9
Şekil 2.3 Deniz ortamında tek aşamalı (YOLO, SSD) ve iki aşamalı (Faster R-CNN) yaklaşımların şematik karşılaştırması [26]. 11
Şekil 2.4 Örnek Precision–Recall eğrisi ve mAP kavramsal gösterimi [32]. 14
Şekil 3.1 YOLO sürümlerinin evrimi ve hız/doğruluk eğilimi [33]. 15
Şekil 3.2 YOLO modellerinin kronolojik evrimi ve temel mimari değişimler. 16
Şekil 4.1 SeaDronesSee veri setinden örnek görüntüler (Farklı irtifa ve açılar). 21
Şekil 4.2 CVAT arayüzü üzerinden veri etiketleme örneği. 22
Şekil 4.3 Deniz ortamında tespit performansını zorlayan örnek durumlar: yansıma, dalga köpüğü ve küçük hedefler. 23
Şekil 5.1 YOLOv8x modeli eğitim ve doğrulama kayıp eğrileri. 29
Şekil 5.2 YOLOv10x modeli eğitim ve doğrulama kayıp eğrileri. 30
Şekil 5.3 YOLO12X modeli eğitim ve doğrulama kayıp eğrileri (NaN bölgeleri işaretli). 30
Şekil 5.4 YOLOv8x modeli test örnekleri. 33
Şekil 5.5 YOLOv10x modeli test örnekleri. 34
Şekil 5.6 YOLO12X modeli test örnekleri. 35
Şekil 5.7 YOLOv8x modeli normalize edilmiş karmaşıklık matrisi (test seti). 36
Şekil 5.8 YOLOv10x modeli normalize edilmiş karmaşıklık matrisi (test seti). 37
Şekil 5.9 YOLO12X modeli normalize edilmiş karmaşıklık matrisi (test seti). 38

TABLO LİSTESİ

Tablo 2.1 Deniz ortamında yaygın dedektör ailelerinin nitel karşılaştırması. 12
Tablo 3.1 Tez kapsamında eğitilen modellerin mimari bileşen özeti. 19
Tablo 4.1 Bu çalışmada kullanılan SeaDronesSee alt kümesi: görüntü ve örnek sayıları. 22
Tablo 4.2 Colab çalışma ortamı donanım ve yazılım özellikleri. 24
Tablo 4.3 Model bazında eğitim parametreleri. 24
Tablo 4.4 Değerlendirmede kullanılan temel metrikler. 25
Tablo 4.5 Değerlendirmede uygulanan eşik değerleri. 26
Tablo 4.6 Modeller arası temel hiperparametre ve yapılandırma karşılaştırması. 26
Tablo 5.1 Veri seti sınıf ve görüntü dağılımı. 28
Tablo 5.2 Modellerin eğitim hiperparametreleri karşılaştırması. 28
Tablo 5.3 Test seti genel performans metrikleri. 31
Tablo 5.4 Sınıf bazlı test sonuçları (mAP50-95). 32
Tablo 5.5 Modellerin doğru sınıflandırma oranları (karmaşıklık matrisi köşegen değerleri). 36
Tablo 5.6 Modellerin nesne kaçırma (Arka Plan Yanlış Negatif) oranları. . . 38
Tablo 6.1 Modellerin genel performans karşılaştırması. 40
Tablo 6.2 Operasyonel gereksinimlere göre model önerileri. 42
Tablo A.1 Modellerin eğitim süreci özetleri (Log kayıtlarına dayalı). 47
Tablo A.2 Genel test metrikleri ve hesaplama profili. 48
Tablo A.3 Sınıf bazlı mAP50 değerleri (Test seti). 48
Tablo A.4 YOLOv8x — Test seti sınıf bazlı yaklaşık TP/FN/FP değerleri. . 49
Tablo A.5 YOLOv10x — Test seti sınıf bazlı yaklaşık TP/FN/FP değerleri. . 49
Tablo A.6 YOLO12X — Test seti sınıf bazlı yaklaşık TP/FN/FP değerleri. . 49
Tablo A.7 Küçük nesne performans göstergesi (Buoy mAP50-95). 50

ÖZET

MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE DENİZ ÜZERİNDEKİ İNSANLARIN TESPİTİ

Murat TETİK

Elektronik ve Haberleşme Mühendisliği Anabilim Dalı
Yüksek Lisans Tezi

Danışman: Doç. Dr. Bülent BOLAT

Bu tez çalışması, yüksek yansımalar, dalga kaynaklı örtülmeler, uzak mesafedeki küçük nesne boyutu ve sınıf dengesizliği gibi zorlu deniz koşullarında insan tespiti performansını artırmak amacıyla, güncel YOLO mimarilerinin (YOLOv8x, YOLOv10x ve deneysel YOLO12X) SeaDronesSee veri seti üzerindeki etkinliğini karşılaştırmalı olarak incelemektedir. Çalışma kapsamında, üç büyük ölçekli model Google Colab A100-SXM4-80GB donanımı üzerinde 1536 piksel yüksek çözünürlükle eğitilmiş ve standart bir protokol (6.225 eğitim, 1.778 doğrulama, 890 test görüntüsü; toplam 63.081 etiketli nesne) altında değerlendirilmiştir. Elde edilen nihai test sonuçlarına göre; YOLOv8x modeli %92,3 mAP50 ve %87,6 duyarlılık değerleri ile özellikle insan sınıfında %94 doğru sınıflandırma oranı ve yalnızca %6 kaçırma oranı ile en kapsamlı tespit performansını sunmuştur. Maksimum Olmayan Bastırma gerektirmeyen mimariye sahip YOLOv10x modeli, düşük gecikme süresiyle benzer doğruluğu (%57,8 mAP50-95) sağlamış ancak şamandıra sınıfında %21 kaçırma oranıyla daha tutucu bir profil sergilemiştir. Deneysel YOLO12X ise eğitim sürecindeki sayısal kararsızlıklara ve gradyan patlamalarına rağmen %92,6 kesinlik skoru ile yanlış alarmların kritik olduğu senaryolar için yüksek kesinlik vaat etmiş; ancak şamandıra sınıfında %27 kaçırma oranı ve %70 doğru sınıflandırma başarısıyla en düşük performansı göstermiştir. Karmaşıklık matrisi analizlerinde, YOLOv8x'in tüm sınıflarda en yüksek doğru sınıflandırma oranlarına (%94 insan, %93 tekne, %80 şamandıra) ve en düşük nesne kaçırma oranlarına sahip olduğu belirlenmiştir. Tezin temel katkıları arasında tekrarlanabilir bir kıyaslama sunulması, dikkat mekanizması tabanlı YOLO12X için

kararlılık odaklı bir eğitim stratejisi geliştirilmesi ve farklı operasyonel senaryolar için model seçim rehberi oluşturulması yer almaktadır. Sonuçlar, modern YOLO tasarımlarının deniz güvenliği uygulamalarında hız, duyarlılık ve kesinlik arasında uygulanabilir dengeler sunduğunu göstermekte olup; arama-kurtarma operasyonları için YOLOv8x'in en güvenilir seçenek olduğu ortaya konmuştur.

Anahtar Kelimeler: Nesne tespiti, denizde insan tespiti, SeaDronesSee, küçük nesne performansı, NMS-free çıkarım.

ABSTRACT

HUMAN DETECTION AT SEA BY USING MACHINE LEARNING METHODS

Murat TETİK

Electronics and Communication Engineering
Master of Science Thesis

Supervisor: Assoc. Prof. Dr. Bülent BOLAT

This thesis comparatively investigates the effectiveness of state-of-the-art YOLO architectures (YOLOv8x, YOLOv10x, and experimental YOLO12X) on the SeaDronesSee dataset to enhance human detection performance under challenging maritime conditions such as high reflections, wave-induced occlusions, small object sizes at distant ranges, and class imbalance. Within the scope of this study, three large-scale models were trained at 1536-pixel high resolution on Google Colab A100-SXM4-80GB hardware and evaluated under a standardized protocol (6,225 training, 1,778 validation, 890 test images; 63,081 labeled objects in total). According to the final test results, the YOLOv8x model achieved 92.3% mAP50 and 87.6% recall, demonstrating the most comprehensive detection performance with a 94% correct classification rate and only 6% miss rate in the human class. The YOLOv10x model, featuring a Non-Maximum Suppression-free architecture, achieved similar accuracy (57.8% mAP50-95) with lower latency but exhibited a more conservative profile with a 21% miss rate in the buoy class. Despite numerical instabilities and gradient explosions during training, the experimental YOLO12X promised high precision for scenarios where false alarms are critical with a 92.6% precision score; however, it showed the lowest performance with a 27% miss rate and 70% correct classification accuracy in the buoy class. Confusion matrix analyses revealed that YOLOv8x achieved the highest correct classification rates across all classes (94% human, 93% boat, 80% buoy) and the lowest object miss rates. The key contributions of this thesis include providing a reproducible benchmark, developing a stability-oriented training strategy for

attention mechanism-based YOLO12X, and establishing a model selection guide for different operational scenarios. The results demonstrate that modern YOLO designs offer viable trade-offs between speed, recall, and precision in maritime security applications, with YOLOv8x identified as the most reliable choice for search and rescue operations.

Keywords: Object detection, maritime human detection, SeaDronesSee, small object performance, NMS-free inference.

1
GİRİŞ

Deniz ortamında insan tespiti; arama-kurtarma (SAR) operasyonları, kıyı güvenliği ve plaj gözetimi gibi hayati uygulamalarda kritik bir role sahiptir. Ancak bu görev; dinamik deniz yüzeyi, değişken aydınlatma koşulları, su yansımaları ve zorlu hava olayları nedeniyle bilgisayarlı görü alanındaki en karmaşık problemlerden biri olarak kabul edilmektedir [1]. Son yıllarda geliştirilen derin öğrenme tabanlı nesne tespit algoritmaları, özellikle YOLO (You Only Look Once) mimarisi, bu tür zorlu ortamlarda umut verici sonuçlar ortaya koymuştur [2, 3].

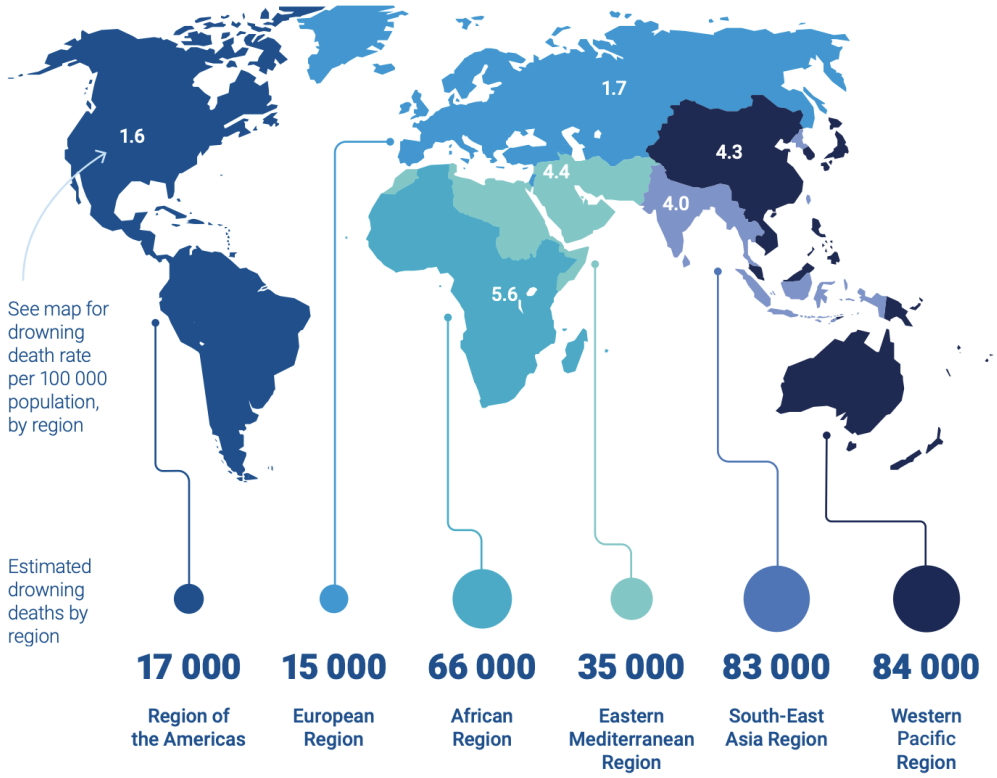
Bu tez çalışmasında, deniz ortamında insan, tekne ve şamandıra tespiti amacıyla, yüksek çözünürlüklü SeaDronesSee veri seti kullanılarak üç ileri seviye YOLO türevi (YOLOv8x, YOLOv10x ve deneysel/özel yapılandırılmış YOLO12X) eğitilmiş ve kapsamlı bir performans analizi gerçekleştirilmiştir. Çalışma kapsamında model eğitimi yalnızca *xlarge* ("x") ölçekli varyantlar üzerinde yürütülmüş; özellikle şamandıra (*Buoy*) ve uzaktaki insan (*Human*) örnekleri gibi küçük nesnelerin tespiti için 1536 piksel giriş çözünürlüğü tercih edilmiştir. YOLOv9 ve YOLOv11 sürümleri bu çalışmanın eğitim kapsamına dahil edilmemiş olup, ilerleyen bölümlerde yalnızca literatür taraması bağlamında referans olarak ele alınacaktır.

Çalışmanın temel odak noktası, gerçek deniz koşullarının yarattığı zorluklara rağmen, yüksek duyarlılık (precision) ve Ortalama Doğruluk (Mean Average Precision - mAP) metriklerinde dengeli bir başarı sağlayan model konfigürasyonlarını belirlemektir. Bu doğrultuda, veri seti seçimi ve yüksek çözünürlüklü eğitim stratejisinin küçük nesne tespit performansına olan etkileri ayrıntılı olarak tartışılacaktır.

1.1 Araştırma Konusunun Önemi

Denizlerde ve iç sularda her yıl yüz binlerce insan boğulma tehlikesiyle karşı karşıya kalmakta ve ne yazık ki bu vakaların önemli bir kısmı ölümle

sonuçlanmaktadır. Dünya Sağlık Örgütü'nün (WHO) Aralık 2024 tarihinde yayımladığı *Küresel Boğulma Önleme Durum Raporu* verilerine göre, dünya genelinde yılda yaklaşık 300.000 kişi boğulma nedeniyle hayatını kaybetmektedir. Bu istatistik, her saat başı 30'dan fazla insanın yaşamını yitirdiği anlamına gelmektedir [4, 5]. Söz konusu ölümlerin yaklaşık %57'sini çocuklar ve gençler oluştururken, vakaların %92'si düşük ve orta gelirli ülkelerde meydana gelmektedir [4].



Şekil 1.1 WHO 2024 Küresel Boğulma İstatistikleri [5]

Kazazedelerin hızlı ve doğru bir şekilde tespit edilmesi, hayat kurtarma operasyonlarının başarısında en belirleyici faktördür. Geleneksel yöntemler, yani insan gözlemine dayalı aramalar, geniş alanların taranmasında yetersiz kalabilmekte; ayrıca yorgunluk ve dikkat dağınıklığı gibi insani faktörlerden olumsuz etkilenmektedir [6]. Bu noktada otonom sistemler ve yapay zeka destekli algılama teknolojileri stratejik bir önem kazanmaktadır [7]. 2023–2025 dönemindeki güncel araştırmalar, deniz ortamına uygun şekilde optimize edilmiş YOLO tabanlı modellerin %90'ın üzerinde tespit doğruluğuna ulaşabildiğini göstermektedir [8, 9].

Deniz ortamında yürütülen arama-kurtarma faaliyetleri, karasal operasyonlara kıyasla çok daha değişken ve öngörülemez koşullara sahiptir. Bu süreci zorlaştıran

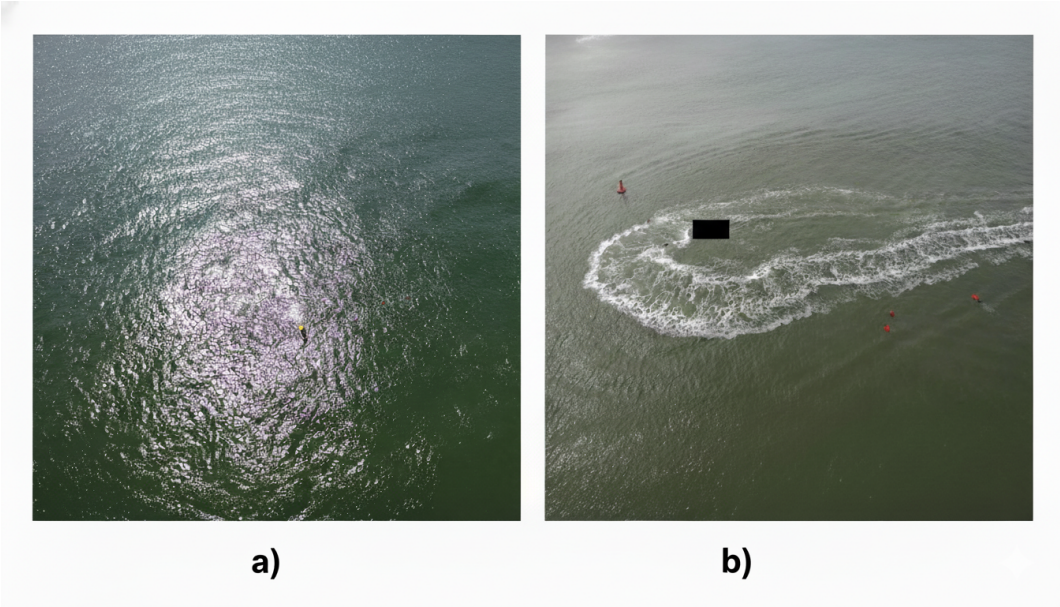
temel faktörler arasında geniş ve dinamik arama alanlarında konumlandırma güçlükleri, dalga yüksekliği, rüzgâr, yağış ve sis gibi anlık değişebilen meteorolojik parametreler ile düşük görünürlük ve zayıf kontrast sayılabilir. Suyla benzer renk tonlarına sahip, kısmen batmış insan siluetinin ayırt edilmesi oldukça güçtür. Bunun yanı sıra hipotermi riski nedeniyle müdahale süresinin son derece sınırlı olması, sınırlı insan kaynağı ve operasyonel araç erişim kısıtları da operasyonları zorlaştıran unsurlardandır. Yüzey yansımaları ve güneş parlamalarının yarattığı görsel gürültü [10] ile dalga köpüğü ve kırılmalarının yarattığı yanlış pozitif algılamalar da tespit sistemlerinin karşılaştığı önemli engellerdir.

Bu zorlukların üstesinden gelmek amacıyla İnsansız Hava Araçları (İHA) kullanımı giderek yaygınlaşmaktadır. İHA sistemleri; geniş alanları hızlı tarama, tehlikeli bölgelere erişim ve personel güvenliğini riske atmadan görev icrası gibi avantajlar sunar [11]. Ancak modern İHA platformlarının sahadaki etkinliği, taşıdıkları görüntüleme sensörlerinin ve gömülü yapay zeka modellerinin başarısıyla doğrudan ilişkilidir [12].

1.2 Problemin Tanımı

Deniz ortamında insan tespiti, klasik nesne tespit problemlerinden ayrılan özgün zorluklar içermektedir. Bu zorluklar; tespit hedefinin doğasından kaynaklanan güçlükler, görüntü işleme süreçlerindeki teknik engeller ve çevresel faktörlerin yarattığı değişkenlikler olmak üzere üç temel kategoride incelenebilir.

Tespit hedefinin doğasından kaynaklanan en temel sorun küçük hedef boyutudur. Havadan alınan görüntülerde denizde yüzen bir insan, genellikle görselin çok küçük bir kısmını (bazen sadece birkaç piksel) kaplamaktadır [1]. Bu çalışmada, söz konusu sorunu aşmak adına 1536 piksel gibi yüksek bir giriş çözünürlüğü ve küçük nesne senaryolarına özel veri artırma stratejileri benimsenmiştir. Kısmi görünürlük de önemli bir engel teşkil etmektedir; denizdeki bir insanın genellikle sadece baş ve omuz kısmı su yüzeyinde kaldığından, modelin öğrenebileceği görsel öznitelikler sınırlı kalmaktadır [7]. Özellikle koyu renkli kıyafetler ve karanlık su koşullarında, insan vücudu ile deniz yüzeyi arasındaki ayırt edilebilirlik minimum seviyeye inmektedir. Dalgaların hareketi, köpükler ve güneş yansımaları sürekli değişen bir arka plan yaratarak tespiti daha da güçleştirmektedir [12]. Kazazedelerin yüzme, batma veya can yeleğiyle hareketsiz durma gibi çok çeşitli pozisyonlarda bulunabilmesi modelin genelleme yeteneğini zorlarken, gün içerisindeki güneş açısı, bulut geçişleri ve su yüzeyindeki parlamalar görüntü kalitesini anlık olarak değiştirmektedir [13].



Şekil 1.2 a) Işık koşullarının tespite etkisi, b) Dalga yüksekliği ve deniz durumunun görsel kaliteye etkisi

Görüntü işleme açısından değerlendirildiğinde, deniz ortamı çeşitli teknik engellerle karakterize edilmektedir. Dalga tepeleri, köpükler veya yüzen döküntüler algoritmalar tarafından hatalı şekilde insan olarak sınıflandırılabilmekte ve yüksek yanlış pozitif oranlarına yol açabilmektedir [10]. Deniz yüzeyindeki görsel karmaşa, asıl tespit edilmesi gereken sinyali maskeleyerek küçük hedeflerin kaybolmasına neden olmakta ve düşük sinyal-gürültü oranı problemi ortaya çıkmaktadır. İHA'nın irtifasına bağlı olarak nesne boyutlarının değişkenlik göstermesi çok ölçekli özellik çıkarımını zorunlu kılarken [14], farklı kamera açılarının nesnelerin geometrisini değiştirmesi modelin çeşitli perspektiflere adapte olmasını gerektirmektedir. SAR operasyonlarında saniyelerin kritik öneme sahip olması nedeniyle gerçek zamanlı işleme gerekliliği de göz ardı edilemez bir faktördür. Bu bağlamda, NMS-free (Non-Maximum Suppression gerektirmeyen) mimariler gecikmeyi azaltmak için önerilmekte olup, literatürde YOLOv10 düşük gecikme ve yüksek doğruluk dengesiyle öne çıkmaktadır.

Çevresel faktörler, tespit algoritmalarının başarısını doğrudan etkilemektedir. Beaufort skalasında 4 ve üzeri deniz durumlarında dalga yüksekliğinin artmasıyla tespit başarısı düşmektedir [8]. Özellikle gün doğumu ve batımında oluşan düşük açılı ışık kamerada kör noktalara neden olabilmekte [15], yağmur ve sis gibi atmosferik olaylar ise görüşü kısıtlayarak optik sensörlerin yanı sıra termal sistemlerin önemini artırmaktadır. Şiddetli rüzgâr hem deniz yüzeyinde köpüklenmeyi artırmakta hem de İHA platformunun stabilizasyonunu bozmaktadır [11].

1.3 Literatür İncelemesi

Deniz ortamında insan tespiti, yıllar içerisinde geleneksel görüntü işleme tekniklerinden modern derin öğrenme yaklaşımlarına doğru önemli bir dönüşüm geçirmiştir. Bu bölümde, alandaki temel yaklaşımlar kronolojik bir perspektifle ele alınmakta ve mevcut literatürdeki boşluklar değerlendirilmektedir.

1.3.1 Geleneksel Görüntü İşleme ve Makine Öğrenmesi Yaklaşımları

Erken dönem çalışmalar, genellikle ön işleme (gürültü azaltma), segmentasyon (eşikleme) ve el yordamı ile özellik çıkarma (şekil, renk) adımlarına dayanmaktaydı. Westall ve ark. [11], İHA görüntüleri için renk uzayı analizi (RGB, HSV, YCbCr) ve zamansal izleme kullanan bir sistem önermiş, ancak dalga hareketlerinin yarattığı yanlış alarmlar yöntemin etkinliğini kısıtlamıştır. Benzer şekilde Fefilatyev ve ark. [12], ufuk çizgisi tespiti ve Kalman filtresi ile gemi takibi üzerine çalışmış, ancak insan gibi küçük hedeflerin tespitinde başarı sınırlı kalmıştır.

Geleneksel yöntemlerin ardından makine öğrenmesi algoritmaları bu alanda yaygınlaşmaya başlamıştır. Dalal ve Triggs [16] tarafından önerilen gradyan histogramları (HOG) ve Destek Vektör Makineleri (SVM) kombinasyonu, yaya tespitinde çığır açan bir yaklaşım olmuş ve denizcilik uygulamalarında da geniş çapta benimsenmiştir; bununla birlikte dinamik arka plan koşullarında performans düşüşü gözlemlenmiştir. Gomez ve ark. [14], Haar-benzeri özelliklerle gemi tespiti gerçekleştirmiş ve termal görüntülerde %70-85 aralığında hassasiyet değerleri raporlamıştır. Öte yandan Freund ve Schapire [17] tarafından geliştirilen AdaBoost yöntemi, zayıf sınıflandırıcıların kademeli olarak güçlendirilmesi prensibine dayanmakta olup, hesaplama yükünü azaltırken SVM ile kıyaslanabilir doğruluk oranları sağlamıştır.

1.3.2 Derin Öğrenme Tabanlı Yaklaşımlar

Derin öğrenme devrimiyle birlikte nesne tespitinde köklü bir paradigma değişimi yaşanmıştır. İlk olarak öne çıkan iki aşamalı dedektörlerden Faster R-CNN [18] ve Mask R-CNN [19], bölge öneri ağları aracılığıyla yüksek doğruluk oranları sunmuştur. Ne var ki bu mimariler, çift aşamalı yapılarından kaynaklanan hesaplama maliyetleri nedeniyle gerçek zamanlı uygulamalarda kullanım alanı sınırlı kalmıştır.

Bu kısıtlamaya yanıt olarak tek aşamalı dedektörler geliştirilmiştir. Liu ve ark. [20] tarafından önerilen SSD (Single Shot Detector), tek geçişli mimarisiyle çıkarım hızını önemli ölçüde artırmış ve denizcilik uygulamalarında yaklaşık 45 FPS

değerlerine ulaşmıştır. Benzer şekilde, Redmon ve ark. [2] tarafından tanıtılan YOLO (You Only Look Once) mimarisi, hız ile doğruluk arasında kurduğu denge sayesinde bu alandaki en yaygın kullanılan modellerden biri haline gelmiştir. Sonraki sürümler olan YOLOv3 [21] ve YOLOv4 [3], çok ölçekli özellik piramitleri ve geliştirilmiş omurga ağları aracılığıyla özellikle küçük boyutlu nesnelerin tespitinde kayda değer iyileştirmeler sağlamıştır.

Son yıllarda ise transformer tabanlı modeller nesne tespiti alanında dikkat çekmeye başlamıştır. Carion ve ark. [22] tarafından geliştirilen DETR (Detection Transformer), nesne tespitini uçtan uca bir küme tahmini problemi olarak yeniden formüle etmiş ve yoğun sahnelerde geleneksel yöntemlerle rekabet edebilir sonuçlar ortaya koymuştur.

1.3.3 Literatürdeki Boşluklar ve Tezin Konumu

Mevcut literatür incelendiğinde, YOLO ailesinin en güncel sürümlerinin (v8, v10, v12) deniz ortamında sistematik olarak karşılaştırıldığı çalışmaların oldukça sınırlı olduğu görülmektedir. Bunun yanı sıra, yüksek çözünürlük stratejilerinin (1536 piksel) küçük hedeflerin tespitine etkisi ve NMS-free (Non-Maximum Suppression içermeyen) yapıların operasyonel avantajları üzerine kapsamlı araştırmalar bulunmamaktadır. Bu tez, SeaDronesSee veri seti üzerinde *xlarge* model konfigürasyonları ile yüksek çözünürlük parametrelerini birleştirerek söz konusu boşluğu doldurmayı ve alanda çalışan araştırmacılara pratik bir uygulama rehberi sunmayı hedeflemektedir.

1.3.4 Tez Çalışmasının Hedefleri

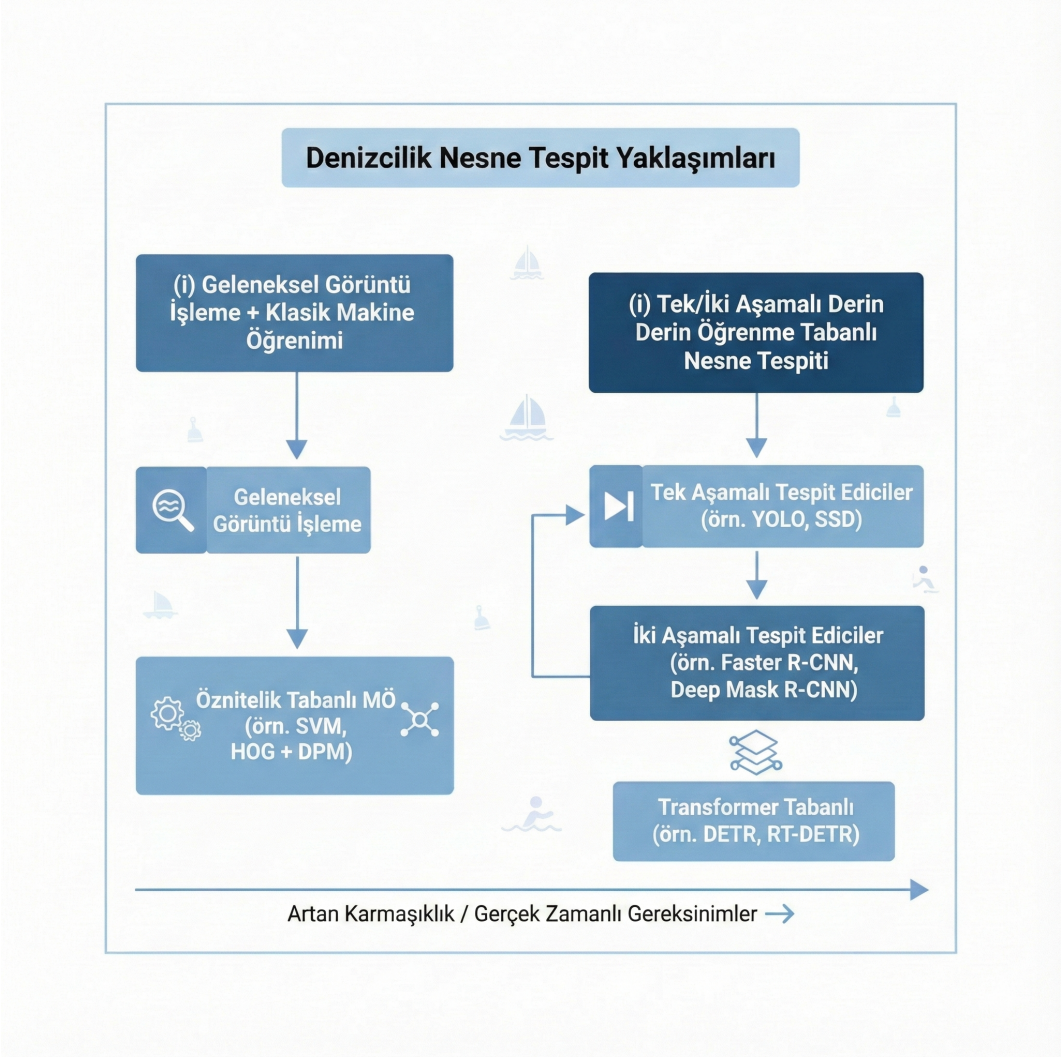
- Modellerin doğruluk (mAP, precision, recall) ve hız (gecikme) performanslarını karşılaştırmalı olarak değerlendirmek.
- Sınıf bazlı hata analizleri gerçekleştirmek.
- Küçük hedeflerin tespiti için yüksek çözünürlük ve veri artırma stratejilerinin etkinliğini incelemek.
- Eğitim kararlılığı (NaN hataları vb.) üzerine teknik bulgular ortaya koymak.
- Operasyonel gereksinimlere göre model seçim rehberi oluşturmak.

2
DENİZDE İNSAN TESPİTİNİN ÖNEMİ VE YÖNTEMLER

Bu bölümde, denizde insan tespitine yönelik temel uygulama alanları ile literatürdeki geleneksel ve modern yaklaşımların teorik çerçevesi sunulmaktadır. Özellikle 2022–2025 döneminde donanım (artan uçuş süreleri, çoklu sensör entegrasyonu) ve yazılım (gelişmiş tek aşamalı dedektörler, dikkat mekanizmaları, veriye dayalı stratejiler) alanlarında kaydedilen ilerlemeler; dalga kapanması (occlusion), yüzey yansımaları ve küçük hedef boyutu gibi deniz ortamına özgü zorlukların üstesinden gelmede kritik rol oynamaktadır.

2.1 Yöntem Kategorileri: Gelenekselden Derin Öğrenmeye

Deniz ortamındaki tespit yaklaşımları genel bir perspektifle iki ana kategoride incelenebilir: (i) Geleneksel görüntü işleme ve klasik makine öğrenmesi teknikleri, (ii) Tek veya iki aşamalı derin öğrenme tabanlı nesne tespit mimarileri. Bu taksonomi, veri karmaşıklığı ile gerçek zamanlı işlem gereksinimleri arasındaki dengeyi anlamak adına kavramsal bir zemin oluşturmaktadır.



Şekil 2.1 Deniz ortamında insan/nesne tespiti yöntemlerinin taksonomik sınıflandırması.

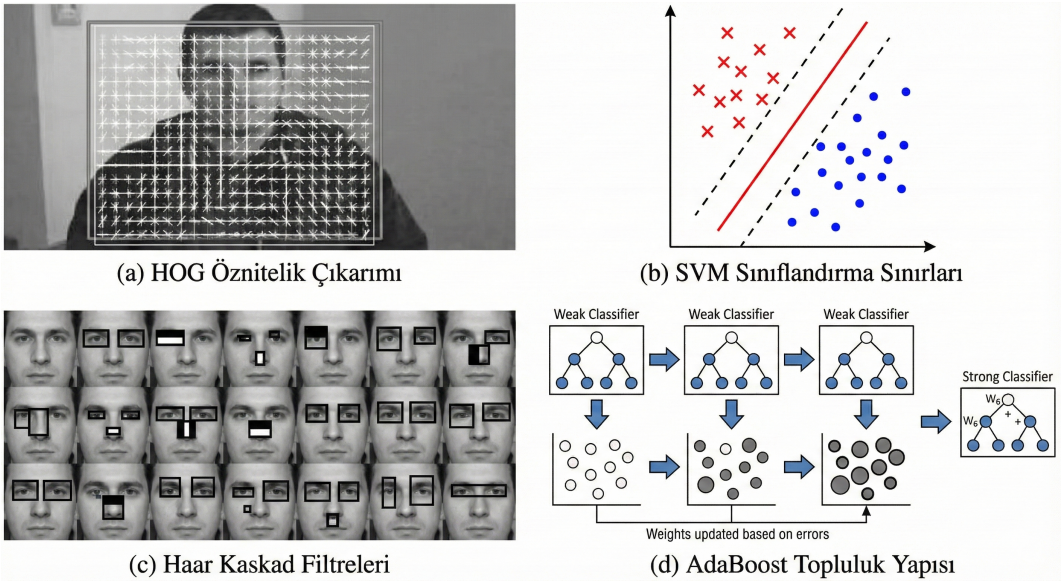
2.2 Geleneksel Yaklaşımlar

Erken dönem çalışmalar, ağırlıklı olarak ön işleme (gürültü azaltma, histogram eşitleme ile kontrast iyileştirme), ufuk çizgisi tespiti ve basit eşikleme/segmentasyon tekniklerine dayanmaktadır. Bu yöntemlerde sıklıkla kullanılan ufuk çizgisi tespiti ile bölge ilgisi daraltma (ROI reduction) mekanizmaları, gemi veya insan gibi küçük hedefleri ararken arka plan karmaşasını azaltmayı hedeflemiştir. Ancak bu teknikler, dinamik dalga yapıları ve güneş parlamaları (sun glint) karşısında yüksek yanlış pozitif oranları üretmekte ve genelleme kabiliyetleri sınırlı kalmaktadır.

2.3 Makine Öğrenmesi Tabanlı Yaklaşımlar

2.3.1 Özellik Tabanlı Makine Öğrenmesi

HOG (Histogram of Oriented Gradients) + SVM (Support Vector Machine), Haar-kaskad sınıflandırıcıları ve AdaBoost toplulukları; sınırlı işlem gücüne sahip donanımlarda düşük gecikme avantajı sunmaları nedeniyle uzun süre tercih edilmiştir. Ancak bu yöntemler; düşük kontrastlı, kameraya uzak veya dalgalar arasında kısmen kaybolmuş insan tespitinde belirgin doğruluk kayıpları yaşamaktadır. Özellik çıkarma (feature extraction) adımının ölçek değişimleri ve perspektif kaymaları karşısındaki dayanıksızlığı, modern senaryolarda bu yöntemlerin daha çok birer "referans noktası" (baseline) olarak değerlendirilmesine neden olmuştur.



(a) HOG Öznitelik Çıkarımı (b) SVM Sınıflandırma Sınırları

(c) Haar Kaskad Filtreleri (d) AdaBoost Topluluk Yapısı

Şekil 2.2 Geleneksel makine öğrenmesi yaklaşımlarının temsili şemaları. (a) HOG öznitelik çıkarımı [16], (b) SVM ile marjin maksimizasyonu ve sınıflandırma [23], (c) Yüz tespiti için kullanılan Haar-benzeri öznitelikler [24], (d) Zayıf sınıflandırıcıların sıralı ağırlıklandırılmasına dayalı AdaBoost yapısı [17].

2.3.2 Derin Öğrenme Tabanlı Tek ve İki Aşamalı Modeller

Derin öğrenme literatüründe mimariler, işlem adımlarına göre sınıflandırılır:

- Tek Aşamalı Dedektörler (YOLO, SSD): Tek bir sinir ağı geçişinde hem nesne konumunu hem de sınıf olasılığını tahmin ederler. Gerçek zamanlı uygulamalar için idealdirler.
- İki Aşamalı Mimariler (Faster/Mask R-CNN): Önce aday bölgeleri (Region Proposals) belirler, ardından sınıflandırma yaparlar. Kısmi görünürlükte daha hassas sonuçlar verseler de, geniş alan taraması gereken deniz gözetimi

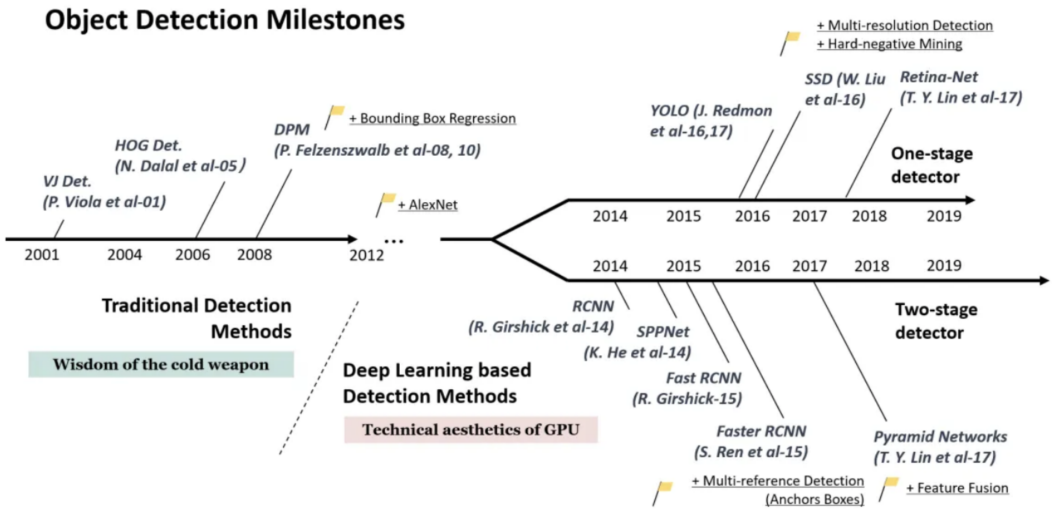
senaryolarında hesaplama maliyetleri darboğaz yaratabilmektedir.

Bu mimarilerin çoğu, çıktı aşamasında Non-Maximum Suppression (NMS) adı verilen bir son işleme adımına ihtiyaç duymaktadır. Nesne dedektörleri, tek bir nesne için birbirine yakın konumlarda çok sayıda örtüşen sınırlayıcı kutu (bounding box) üretme eğilimindedir. NMS algoritması, bu tekrarlayan tahminler arasından en yüksek güven skoruna sahip olanı seçerek diğerlerini baskılar ve böylece her nesne için yalnızca tek bir tespit sonucu elde edilmesini sağlar [25]. Bu süreç şu adımlarla gerçekleştirilir: öncelikle tüm tahminler güven skorlarına göre sıralanır, ardından en yüksek skorlu kutu seçilir ve bu kutuyla belirli bir IoU (Intersection over Union) eşiğinin üzerinde örtüşen diğer kutular elenir. Bu işlem, kalan tüm kutular için yinelemeli olarak tekrarlanır. Ancak NMS, hesaplama maliyeti eklemesi ve kalabalık sahnelerde yakın nesnelerin yanlışlıkla baskılanması gibi dezavantajlara sahiptir. Bu nedenle YOLOv10 gibi güncel mimariler, uçtan uca eğitim stratejileriyle NMS bağımlılığını ortadan kaldıran yapılar sunmaktadır.

Ayrıca, Transformer tabanlı modeller (DETR ve türevleri), "Set Prediction" yaklaşımıyla NMS bağımlılığını ortadan kaldırmayı vaat etse de, uzun eğitim süreleri ve yüksek veri gereksinimleri nedeniyle bu tezin çözüm stratejisi dışında bırakılmıştır.

2.4 Derin Öğrenme Tabanlı Nesne Tespiti Algoritmaları

Son yıllarda derin öğrenme, deniz arama-kurtarma operasyonlarında standart çözüm haline gelmiştir. Özellikle 2022–2025 yılları arasında geliştirilen; anchor-free (çapa kutusu bağımsız) başlıklar, NMS-free çıkarım yapıları ve çok ölçekli özellik füzyonu gibi yenilikler, zorlu deniz koşullarında kararlılığı artırmıştır.



Şekil 2.3 Deniz ortamında tek aşamalı (YOLO, SSD) ve iki aşamalı (Faster R-CNN) yaklaşımların şematik karşılaştırması [26].

2.4.1 Bölge Tabanlı (İki Aşamalı) CNN Mimarileri

Bölge tabanlı yaklaşımlar, nesne tespitini iki ardışık aşamada gerçekleştirmektedir. R-CNN ile başlayan bu süreç, Faster R-CNN [27] mimarisinde entegre edilen Bölge Öneri Ağı (RPN) sayesinde önemli bir hız kazanımı elde etmiştir. Bu mimariler deniz yüzeyindeki karmaşık yansımalar içinde kararlı sonuçlar üretebilseler de, acil müdahale gerektiren durumlarda işlem yükü belirgin bir dezavantaj oluşturmaktadır.

İki aşamalı mimarilerin bir diğer önemli temsilcisi olan Mask R-CNN [19], nesne tespitiyle eş zamanlı olarak piksel bazlı segmentasyon gerçekleştirmektedir. Suyun içinde yalnızca başı görünen kazazedeler için siluet bilgisi ayırt edici bir özellik sunsa da, bu ek hesaplama yükü İHA gibi kısıtlı donanıma sahip platformlar için zorlayıcı olmaktadır.

2.4.2 Tek Aşamalı Dedektörler

Tek aşamalı dedektörler, bölge önerisi aşamasını atlayarak doğrudan tahmin yapmaları sayesinde önemli hız avantajı sağlamaktadır. Liu ve ark. [20] tarafından geliştirilen SSD (Single Shot MultiBox Detector), farklı ölçeklerdeki özellik haritalarını kullanarak tahmin yürütmektedir. Büyük hedeflerde başarılı sonuçlar verse de, şamandıra gibi küçük hedeflerin tespitinde modern yöntemlerin gerisinde kalabilmektedir.

Sınıf dengesizliği sorununa çözüm getiren RetinaNet [28], "Focal Loss" fonksiyonu

ile bu alandaki önemli bir açığı kapatmıştır. Deniz görüntülerinde arka planın (su) baskınlığına karşı, ağın zor örneklere (insan) odaklanmasını sağlayarak daha dengeli bir öğrenme süreci sunmaktadır.

YOLO (You Only Look Once) ailesi [2, 3], hız-doğruluk dengesindeki başarısıyla tek aşamalı dedektörler arasında öne çıkmaktadır. Özellikle v8, v10 ve v12 sürümlerinde yer alan "Task Aligned Assigner" (TAL) gibi gelişmiş atama stratejileri ve yüksek çözünürlüklü giriş desteği, bu mimariyi deniz uygulamaları için en güçlü aday konumuna getirmiştir.

2.4.3 Güncel Karşılaştırmalar: Nitel Bir Çerçeve

Farklı dedektör ailelerinin deniz ortamındaki karakteristik davranışları Tablo 2.1'da özetlenmiştir.

Tablo 2.1 Deniz ortamında yaygın dedektör ailelerinin nitel karşılaştırması.

Mimari Ailesi	Doğruluk	Çıkarım Hızı	Küçük Nesne Başarısı
Faster/Mask R-CNN	Yüksek	Düşük-Orta	Yüksek
SSD	Orta	Orta-Yüksek	Orta-Düşük
RetinaNet	Orta-Yüksek	Orta	Orta-Yüksek
YOLO (v8/v10/v12)	Yüksek	Çok Yüksek	Yüksek (Konfigürasyona bağlı)
DETR ve Türevleri	Orta-Yüksek	Orta	Orta

2.4.4 Tez Kapsamı ile İlişkilendirme

Bu tez çalışmasında; SeaDronesSee veri seti üzerinde (Eğitim: 6225, Doğrulama: 1778, Test: 890 görüntü), YOLOv8x, YOLOv10x ve deniz ortamı için özelleştirilmiş deneysel YOLO12X (xlarge) modelleri eğitilmiştir. Küçük hedef tespitini maksimize etmek için giriş çözünürlüğü 1536 piksel olarak belirlenmiş ve eğitimler NVIDIA A100 donanımında gerçekleştirilmiştir.

2.5 Nesne Tespiti Performans Metrikleri

Deniz gözetimi gibi kritik alanlarda başarı ölçümü, sadece doğruluğu değil, hataların türünü de (yanlış alarm vs. kaçırma) kapsamalıdır.

2.5.1 Kesişimlerin Birleşime Oranı (IoU) ve Varyantları

IoU (Intersection over Union), tahmin edilen kutu (A) ile gerçek kutu (B) arasındaki örtüşme başarısını ifade eder:

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|} \tag{2.1}$$

Standart IoU'nun sınırlılıklarını (örneğin kutular kesişmediğinde gradyanın kaybolması) aşmak için çeşitli varyantlar geliştirilmiştir. Rezatofighi ve ark. [29] tarafından önerilen GIoU (Generalized IoU), kutuları kapsayan en küçük dışbükey alanı (C) hesaba katarak standart IoU'yu genelleştirmektedir:

$$\text{GIoU} = \text{IoU} - \frac{|C \setminus (A \cup B)|}{|C|} \tag{2.2}$$

Zheng ve ark. [30] tarafından geliştirilen DIoU (Distance IoU) ise kutu merkezleri arasındaki Öklid mesafesini minimize etmeyi hedeflemektedir:

$$\text{DIoU} = \text{IoU} - \frac{\rho^2(b, b^{gt})}{c^2} \tag{2.3}$$

Bu denklemde ρ merkezler arası mesafeyi, c ise kapsayıcı kutunun köşegen uzunluğunu temsil etmektedir. Aynı araştırmacılar tarafından önerilen CIoU (Complete IoU), merkez uzaklığına ek olarak en-boy oranı tutarlılığını ($\alpha \cdot v$) da denkleme dahil ederek daha kapsamlı bir benzerlik ölçütü sunmaktadır:

$$\text{CIoU} = \text{IoU} - \frac{\rho^2(b, b^{gt})}{c^2} - \alpha \cdot v \tag{2.4}$$

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

Sınıflandırma performansı aşağıdaki temel metriklerle ölçülür:

$$\text{Kesinlik} = \frac{TP}{TP + FP}, \qquad \text{Duyarlılık} = \frac{TP}{TP + FN} \tag{2.5}$$

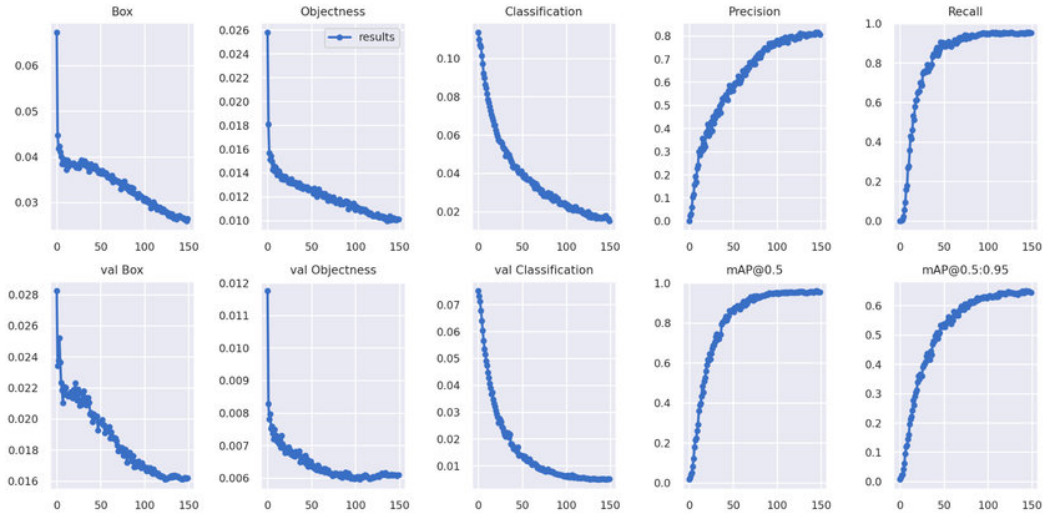
Deniz arama-kurtarma operasyonlarında, bir kazazedeyi gözden kaçırmanın maliyeti tolere edilemez olduğundan "Duyarlılık" (Recall) önceliklidir. Ancak operatör yorgunluğunu önlemek için dengeli bir F1-Skoru hedeflenir:

$$\mathrm{F1} = 2 \cdot \frac{\text{Kesinlik} \cdot \text{Duyarlılık}}{\text{Kesinlik} + \text{Duyarlılık}} \tag{2.6}$$

2.5.3 Ortalama Doğruluk (mAP)

Bu tezde COCO standardına uygun iki temel mAP metriği kullanılmıştır:

- mAP@50: IoU eşiği 0.50 olduğundaki ortalama doğruluk.
- mAP@50-95: IoU eşiği 0.50'den 0.95'e kadar artırıldığında elde edilen ortalama değer. Bu metrik, küçük nesnelerin (şamandıra, insan başı) konumlandırma hassasiyetini ölçmek için kritiktir [31].



Şekil 2.4 Örnek Precision–Recall eğrisi ve mAP kavramsal gösterimi [32].

2.5.4 Bu Tezde Kullanılan Değerlendirme Protokolü

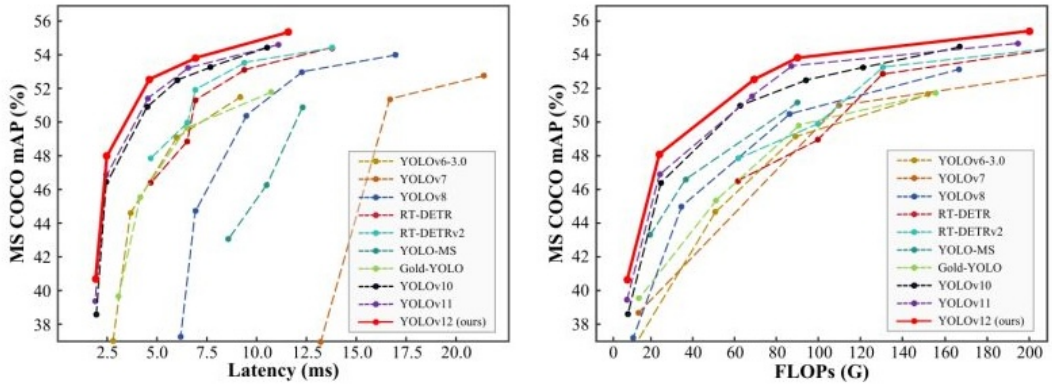
Modeller (YOLOv8x, YOLOv10x, YOLO12X), ayrılan test veri seti (890 görüntü) üzerinde şu parametrelerle değerlendirilmiştir:

- Güven Eşiği: 0.25 (Literatür standardı).
- IoU Eşiği (NMS için): 0.45.
- Görüntü Çözünürlüğü: 1536 piksel (Eğitimle tutarlı).
- Donanım: NVIDIA A100-SXM4-80GB.

3
YOLO (YOU ONLY LOOK ONCE) MODELLERİ

YOLO (You Only Look Once) ailesi, nesne tespiti problemini geleneksel yöntemlerin aksine çok aşamalı bir süreç (bölge önerisi ve sınıflandırma) olarak değil, tek bir regresyon problemi olarak ele almaktadır. Bu yaklaşımda, bir giriş görüntüsü tek bir ileri beslemeli sinir ağı üzerinden geçirilir; uzamsal olarak ayrılmış sınırlayıcı kutular ve bunlara ait sınıf olasılıkları eş zamanlı olarak tahmin edilir [2]. Bu tek aşamalı mimari yapısı, modelin görüntüye sadece bir kez bakarak tüm hedefleri tespit etmesine olanak tanımaktadır. Söz konusu özellik, arama-kurtarma operasyonları, otonom deniz araçları ve gerçek zamanlı gözetim gibi düşük gecikme süresi gerektiren kritik uygulamalar için belirleyici bir avantaj sağlamaktadır.

Bu tez çalışmasında, YOLO ailesinin tarihsel gelişimi özetlendikten sonra; 2023–2025 döneminin en güncel ve güçlü temsilcileri olan YOLOv8x, YOLOv10x ve bu çalışma kapsamında deniz ortamı için özel olarak optimize edilen deneysel YOLO12X modelleri ele alınacaktır. Bahsi geçen modeller, SeaDronesSee veri seti üzerinde eğitilmiş ve ayrıntılı performans analizlerine tabi tutulmuştur.



Şekil 3.1 YOLO sürümlerinin evrimi ve hız/doğruluk eğilimi [33].

3.1 YOLO Modellerinin Tarihsel Gelişimi

YOLO mimarisi, ilk ortaya çıkışından günümüze kadar hız-doğruluk dengesi, küçük nesne tespiti yeteneği ve eğitim kararlılığı konularında sürekli bir evrim geçirmiştir. Bu bölümde, tezde kullanılan modellere temel oluşturan önceki sürümlerin kilit yenilikleri özetlenmektedir.



Şekil 3.2 YOLO modellerinin kronolojik evrimi ve temel mimari değişimler.

İlk sürüm olan YOLOv1, giriş görüntüsünü $S \times S$ boyutunda bir ızgaraya bölerek çalışmaktadır. Her ızgara hücresi, B adet sınırlayıcı kutu ve bu kutulara ait güven skorlarını tahmin etmektedir [2]. İşlem hızı açısından devrim niteliğinde olsa da, uzamsal kısıtlamalar nedeniyle küçük nesnelerin veya yoğun hedef gruplarının tespitinde başarısı sınırlı kalmıştır.

"Daha İyi, Daha Hızlı, Daha Güçlü" sloganıyla sunulan YOLOv2, yığın normalleştirme, yüksek çözünürlüklü sınıflandırıcı eğitimi ve en önemlisi çapa kutuları kavramını getirmiştir [34]. Veri kümesi üzerindeki nesne boyutlarına göre K-Ortalamalar kümeleme yöntemiyle belirlenen bu kutular, eğitim kararlılığını önemli ölçüde artırmıştır. Ardından gelen YOLOv3, daha derin bir özellik çıkarıcı olan Darknet-53 omurgasını ve artık bağlantıları kullanmıştır. En kritik yeniliği, tahminlerin üç farklı ölçekte (13×13, 26×26, 52×52) yapılmasıdır [21]. Bu özellik piramidi yaklaşımı, YOLO'nun önceki sürümlerinde zayıf olduğu küçük nesne tespiti performansını ciddi oranda iyileştirmiştir.

YOLOv4, CSPDarknet53 omurgası, PANet boyun yapısı ve Uzamsal Piramit Havuzlama bloğu gibi modern bileşenleri entegre etmiştir. Ücretsiz iyileştirmeler (veri artırma, düzenlileştirme) ve özel iyileştirmeler (Mish aktivasyonu, dikkat modülleri) stratejileri ile hızdan ödün vermeden doğruluk artırılmıştır [3]. PyTorch tabanlı yapısı ile endüstri standardı haline gelen YOLOv5 ise Nano, Small, Medium, Large ve XLarge gibi ölçeklenebilir varyantlar sunmuştur [35]. Otomatik çapa kutusu belirleme özelliği ve mozaik veri artırma teknikleri, modelin yeni veri setlerine uyumunu kolaylaştırmıştır. YOLOv6 ve YOLOv7 sürümleri ise yeniden parametrelendirme tekniklerine odaklanmıştır. YOLOv7, E-ELAN yapısı

ile eğitim sırasında karmaşık, çıkarım sırasında ise sadeleştirilmiş bir ağ yapısı sunarak gradyan akışını optimize etmiştir [36].

3.2 YOLOv8 Mimarisi ve Özellikleri

YOLOv8; çapa kutusuz yapı, ayrıştırılmış tespit kafası ve C2f modülleri ile genellenebilirlik ve eğitim kararlılığını artırmayı hedefleyen modern bir mimaridir [37]. Bu tez kapsamında YOLOv8'in en yüksek kapasiteli xlarge (YOLOv8x) varyantı, küçük nesne duyarlılığını en üst düzeye çıkarmak için yüksek çözünürlükte (1536 piksel) eğitilmiştir.

Omurga yapısında yer alan C2f blokları, önceki C3 modüllerinin yerini alarak daha fazla ve hafif atlama bağlantısı ile gradyan akışını iyileştirmektedir. Evrişim-Yığın Normalleştirme-SiLU katmanları daha kararlı veri dağılımı sağlarken, Hızlı Uzamsal Piramit Havuzlama katmanı son aşamada bütünsel bağlam bilgisini yoğunlaştırarak özellik temsilini zenginleştirmektedir. Boyun kısmında PANet tabanlı çift yönlü yol (yukarıdan aşağıya ve aşağıdan yukarıya) yapısı kullanılarak çok ölçekli özellik birleştirme gerçekleştirilmektedir. Farklı ölçeklerdeki anlamsal ve uzamsal bilgiyi birleştiren bu yapı, deniz yüzeyindeki küçük nesnelerin tespiti için kritik öneme sahiptir.

Çapa kutusuz yaklaşımda, doğrudan nesne merkezi ve boyutları tahmin edilmekte; bu durum çapa kutusu hiperparametre optimizasyonu ihtiyacını ortadan kaldırmaktadır. Ayrıştırılmış sınıflandırma ve regresyon dalları, her iki görevin (konumlandırma ve tanımlama) çatışmadan optimize edilmesini sağlamaktadır. Toplam kayıp fonksiyonu, sınıflandırma (İkili Çapraz Entropi), kutu regresyonu (IoU/CIoU) ve Dağılım Odaklı Kayıp bileşenlerinden oluşmaktadır:

$$L_{\text{toplam}} = \lambda_{cls} L_{cls} + \lambda_{box} L_{box} + \lambda_{dfl} L_{dfl} \tag{3.1}$$

Denklem 3.1'de yer alan değişkenler şu şekilde tanımlanmaktadır: L_{toplam} toplam kayıp değerini, L_{cls} sınıflandırma kaybını (İkili Çapraz Entropi), L_{box} kutu regresyonu kaybını (CIoU), L_{dfl} ise Dağılım Odaklı Kaybı ifade etmektedir. λ_{cls}, λ_{box} ve λ_{dfl} terimleri sırasıyla bu kayıp bileşenlerinin ağırlık katsayılarını temsil etmektedir. Bu tezde Ultralytics varsayılan ağırlık katsayıları korunmuş, ekstra bir yeniden ölçekleme yapılmamıştır.

3.3 YOLOv9 ve YOLOv11 (Literatür Özeti)

Bu tez çalışmasında YOLOv9 ve YOLOv11 modelleri eğitilmemiştir. Literatürdeki yerleri şu şekildedir: YOLOv9 (Şubat 2024), bilgi kaybını önlemek için Programlanabilir Gradyan Bilgisi yaklaşımını sunmuştur [38]. YOLO11 (Eylül 2024) ise parametre verimliliği ve özellik birleştirmeye odaklanan bir ara sürümdür [39].

3.4 YOLOv10 Mimarisi ve Özellikleri

YOLOv10 (Mayıs 2024), Maksimum Olmayan Bastırma gerektirmeyen çıkarım tasarımı ile son işlem süresini en aza indirmeyi ve hız/doğruluk dengesini iyileştirmeyi hedeflemektedir [40]. Mimari yeniliklerin merkezinde Tutarlı Çift Atama stratejisi yer almaktadır.

YOLOv10, eğitim sırasında iki başlı bir yapı kullanmaktadır [40]. Birden çoğa başlık, her nesne için birden fazla aday tahmin üreterek eğitim sinyalini zenginleştirmektedir. Birden bire başlık ise her nesne için tek bir nihai tahmin üretmektedir. Çıkarım aşamasında yalnızca birden bire başlık kullanıldığından, Maksimum Olmayan Bastırma adımına gerek kalmamaktadır. Eşleştirme metriği aşağıdaki gibi formüle edilmektedir:

$$m(p, b) = p^{\alpha} \cdot \text{IoU}(b, b_{gt})^{\beta} \cdot s \tag{3.2}$$

Bu denklemde p sınıf olasılığını, b tahmin kutusunu, b_{gt} gerçek referans kutusunu ve s uzamsal öncelik puanını temsil etmektedir [40].

Literatürde YOLOv10'un verimliliğini artıran diğer unsurlar arasında parametre yükünü azaltan dar evrişimli hafif sınıflandırma başlığı, düşük verimli blokların Kompakt Ters Blok ile değiştirildiği derece rehberli blok seçimi ve hesaplama maliyetini düşürmek için dikkat mekanizmasının özelliklerin alt kümesine uygulandığı kısmi öz-dikkat yapısı bulunmaktadır [40]. Bu tezde, modelin en güçlü varyantı olan YOLOv10x eğitilmiştir.

3.5 YOLOv12 Mimarisi (Deneysel Çalışma)

Bu tezde, literatürde dikkat mekanizmalarına yönelen "YOLO12X" (xlarge) ön-eğitimli ağırlıkları kullanılarak deneysel bir çalışma yürütülmüştür [33, 41]. SeaDronesSee veri seti üzerinde 1536 piksel çözünürlükle yapılan eğitimlerde,

modelin yüksek kapasitesi ve mimari karmaşıklığı nedeniyle bazı kararlılık sorunları gözlemlenmiştir. Özellikle eğitim sürecinde periyodik olarak karşılaşılan Sayısal Değil (NaN) Kayıp fenomeni; öğrenme oranının düşürülmesi, ısınma periyodunun uzatılması, yığın boyutunun 1'e indirilmesi ve ağırlık sönümü değerinin artırılmasıyla kontrol altına alınmıştır.

YOLOv12 ile ilişkilendirilen temel mimari bileşenler şunlardır [33]: Alan Dikkat mekanizması, özellik haritasını bölgelere ayırarak dikkat mekanizmasını daha verimli ($O(n)$ karmaşıklığına yakın) uygulamakta ve denizdeki küçük hedeflerin ayrıştırılmasında potansiyel avantaj sağlamaktadır. R-ELAN yapısı, ELAN mimarisini artık bağlantılarla güçlendirerek derin ağlarda gradyan akışını kararlı hale getirmektedir [36]. Hızlı Dikkat mekanizması ise bellek erişimini optimize ederek dikkat hesaplamalarını hızlandırmaktadır [42].

3.6 Modellerin Karşılaştırmalı Özeti

Bu tezde eğitilen üç modelin (YOLOv8x, YOLOv10x, YOLO12X) temel mimari farkları Tablo 3.1'de özetlenmiştir.

Tablo 3.1 Tez kapsamında eğitilen modellerin mimari bileşen özeti.

Model	Omurga/Boyun	Kafa ve Öne Çıkan Özellikler
YOLOv8x	CSP-türev (C2f), PANet, SPPF	Çapa kutusuz, ayrıştırılmış kafa; Dağılım Odaklı Kayıp + IoU kaybı. Yüksek çözünürlük stratejisiyle uyumlu, kararlı bir temel model.
YOLOv10x	Verimlilik odaklı CSP, PANet birleştirme	Maksimum Olmayan Bastırma gerektirmeyen çıkarım sağlayan çift başlık (Birden Çoğa + Birden Bire). Düşük gecikme süresi ve tutarlı atama stratejisi.
YOLO12X	Dikkat ağırlıklı (R-ELAN, Alan Dikkat)	Yüksek kapasiteli deneysel yapı. Eğitim kararlılığı için hassas hiperparametre ayarı (düşük öğrenme oranı, uzun ısınma süresi) gerektirmiştir.

4
MATERYAL VE METOT

4.1 SeaDronesSee Veri Seti

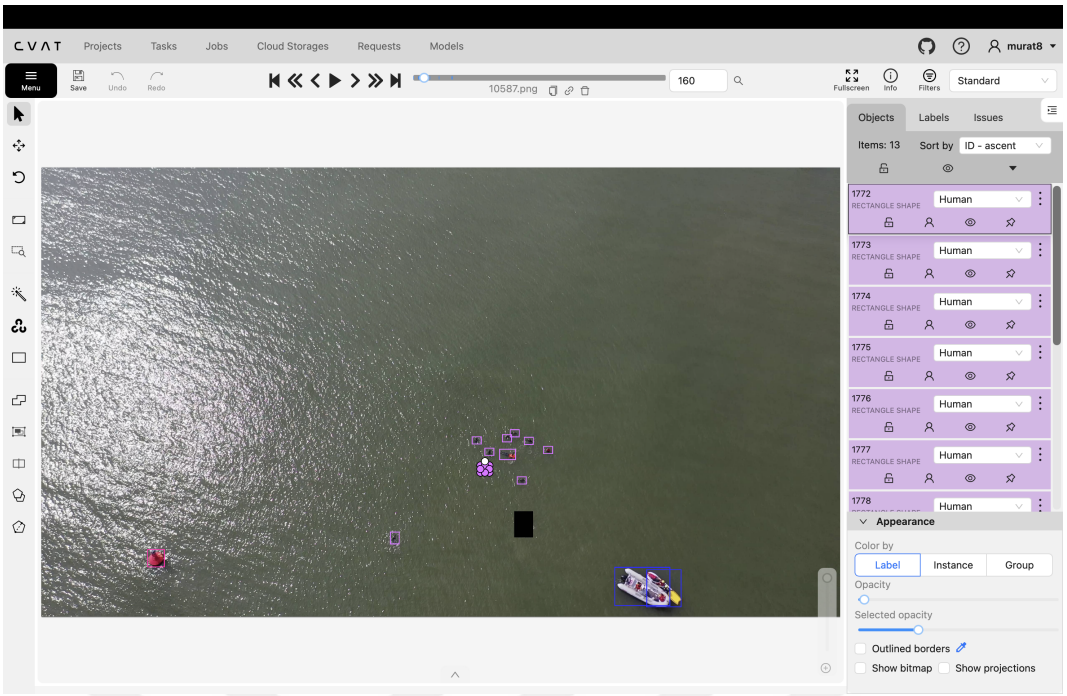
Bu çalışmada, denizde insan tespiti amacıyla YOLO ailesinin güncel sürümleri karşılaştırılırken temel veri kaynağı olarak SeaDronesSee veri seti kullanılmıştır. SeaDronesSee; insani arama-kurtarma senaryoları odaklı, açık su ortamında insan tespiti problemini ele alan ve İHA tabanlı deniz güvenliği uygulamalarına yönelik literatürdeki ilk büyük ölçekli kıyaslama veri setlerinden biridir [1].



Şekil 4.1 SeaDronesSee veri setinden örnek görüntüler (Farklı irtifa ve açılar).

WACV 2022'de sunulan SeaDronesSee; kara trafiği gibi nispeten iyi temsil edilen İHA veri setleri ile deniz tabanlı sahneler arasındaki boşluğu kapatmayı hedeflemektedir. Veri seti, farklı irtifalarda (yaklaşık 5–260 m), değişken kamera açılarında (0°–90°) ve çeşitli deniz/hava koşulları altında toplanmıştır. Amaç, su yüzeyi üzerindeki küçük ve kısmen örtülü insan tespiti problemini gerçekçi koşullarda modellemektir [1].

Orijinal SeaDronesSee etiket uzayı (yüzücü, can yelekli kişi, tekne vb.) bu çalışmada denetimli tespit görevini sadeleştirmek ve odağı netleştirmek amacıyla üç ana sınıfa indirgenmiştir: Human (sudaki veya kısmen görünen kişi; teknedeki kişiler dâhil olmak üzere her türlü insan varlığı), Boat (çeşitli tip ve boyutlardaki su araçları) ve Buoy (deniz şamandıraları ve işaretleyiciler). Etiketleme ve düzenleme süreçleri CVAT ve Microsoft Azure platformları üzerinde yürütülmüş, çıktılar YOLO formatına dönüştürülmüştür.



Şekil 4.2 CVAT arayüzü üzerinden veri etiketleme örneği.

SeaDronesSee verisi, bu tez kapsamında Tablo 4.1'te belirtilen şekilde bölünmüş ve tüm eğitim/test süreçleri bu sabit alt küme üzerinde gerçekleştirilmiştir.

Tablo 4.1 Bu çalışmada kullanılan SeaDronesSee alt kümesi: görüntü ve örnek sayıları.

Bölüm	Görüntü	Human	Boat	Buoy	Toplam Örnek
Eğitim	6,225	26,946	13,035	4,253	44,234
Doğrulama	1,778	7,637	3,619	1,245	12,501
Test	890	3,839	1,873	634	6,346
Toplam	8,893	38,422	18,527	6,132	63,081

Toplam etiket dağılımı (Human %60,9, Boat %29,4, Buoy %9,7) belirgin bir sınıf dengesizliğine işaret etmektedir. Bu nedenle eğitim stratejisinde özellikle *Buoy* sınıfı ve küçük nesneler için yüksek çözünürlük (1536×1536) ve zengin veri artırma (Mozaik, Mixup) yöntemleri tercih edilmiştir.

Deniz ortamında tespit performansını zorlayan başlıca fiziksel etkenler arasında küçük hedef boyutu (İHA irtifasına bağlı olarak bir insan veya şamandıranın görüntüde yalnızca birkaç piksellik alan kaplaması), dalga kaynaklı örtülme (hedeflerin dalga tepeleri ve çukurları arasında periyodik olarak kaybolup yeniden

belirmesi), parlama ve yansıma (düşük açılı güneş ışığının su yüzeyinde saturasyona neden olarak yanlış pozitifleri artırması) ve değişken deniz durumu (deniz durumu şiddetlendikçe oluşan köpük desenlerinin görsel gürültüyü artırması) sayılabilmektedir.



Şekil 4.3 Deniz ortamında tespit performansını zorlayan örnek durumlar: yansıma, dalga köpüğü ve küçük hedefler.

4.2 Donanım ve Yazılım Altyapısı

Bu çalışmada YOLOv8x, YOLOv10x ve deneysel YOLO12X modelleri, Google Colab Pro+ ortamında NVIDIA A100-SXM4-80GB GPU kullanılarak eğitilmiştir. Aşağıdaki veriler doğrudan deney günlüklerinden alınmıştır.

Tablo 4.2 Colab çalışma ortamı donanım ve yazılım özellikleri.

Bileşen	Özellik
İşlemci (CPU)	6 fiziksel / 12 mantıksal çekirdek ($\approx$ 2.2 GHz)
Grafik İşlemci (GPU)	NVIDIA A100-SXM4-80GB (CUDA 12.6)
GPU Belleği	79.3 GB
Sistem RAM	167.1 GB
Disk Alanı	235.7 GB
Depolama	Google Drive Entegrasyonu
Python	3.12.x
PyTorch	2.8.0+cu126
Ultralytics YOLO	v8.3.205 (YOLOv8x), v8.3.210 (YOLOv10x), v8.3.227 (YOLO12X)
CUDA Toolkit	Colab ön tanımlı

Her model için kullanılan eğitim parametreleri Tablo 4.3'te özetlenmiştir.

Tablo 4.3 Model bazında eğitim parametreleri.

Parametre	YOLOv8x	YOLOv10x	YOLO12X
Ön Eğitimli Ağırlık	yolov8x.pt	yolov10x.pt	yolo12x.pt
Giriş Boyutu (piksel)	1536	1536	1536
Planlanan Epoch	300	300	300
Gerçekleşen Epoch	103 (Erken Durdurma)	65 (Erken Durdurma)	En iyi ağırlık seçildi
Eniyileyici	AdamW	AdamW	AdamW
Başlangıç Öğrenme Oranı	0.003	0.003	0.0003
Ağırlık Sönümü	0.02	0.02	0.002
Yığın Boyutu	Otomatik	Otomatik	1
Isınma Epoch Sayısı	Varsayılan	Varsayılan	15 (Uzatılmış)
Mozaik	1.0	1.0	1.0
Mixup	0.2	0.2	0.3
Diğer Artırma	Yatay Çevirme=0.5	HSV ayarları	Kopyala-Yapıştır, Rastgele

YOLO12X eğitimi sırasında, özellikle 21-24. epoch aralığında doğrulama kayıplarında Sayısal Değil (NaN) oluşumu gözlemlenmiştir. Bu sorunu gidermek için öğrenme oranı 10 kat düşürülmüş ($0.003 \rightarrow 0.0003$), ısınma süresi uzatılmış, yığın boyutu 1'e indirilmiş ve bellek yönetimi ayarları optimize edilmiştir. Bu stratejilerle modelin eğitimi tamamlanabilmiş, ancak mAP50-95 performansı kararlılık sorunları nedeniyle diğer modellere kıyasla daha düşük kalmıştır.

4.3 Veri Ön İşleme ve Artırma

SeaDronesSee veri setindeki uzak insan ve şamandıra hedefleri piksel bazında çok küçük alan kaplamaktadır. Bu nedenle tüm eğitimlerde 1536 piksel giriş boyutu standartlaştırılmıştır. Bu yaklaşım, işlem yükünü artırsa da, küçük nesnelerin

geri çağırma oranını kritik düzeyde iyileştirmiştir. YOLOv8x ve YOLOv10x eğitimlerinde Ultralytics'in varsayılan dinamik artırma politikaları izlenmiş ve eğitim sonuna doğru mozaik kapatma uygulanmamıştır (erken durdurma devreye girdiği için). Ancak kararsızlık gösteren YOLO12X eğitiminde mozaik kapatma parametresi 40 olarak ayarlanarak, eğitimin son 40 epoch'unda mozaik veri artırma kapatılmış ve kararlılık hedeflenmiştir.

4.4 Değerlendirme Metodolojisi

Modellerin başarımı, Microsoft COCO standardına uygun metrikler ve gerçek zamanlı senaryo simülasyonları ile ölçülmüştür. Kullanılan temel metrikler Tablo 4.4'te tanımlanmıştır.

Tablo 4.4 Değerlendirmede kullanılan temel metrikler.

Metrik	Formül	Açıklama
Kesinlik (Precision)	$\frac{TP}{TP+FP}$	Doğru pozitif tahminlerin tüm pozitif tahminlere oranı
Duyarlılık (Recall)	$\frac{TP}{TP+FN}$	Doğru pozitif tahminlerin tüm gerçek pozitiflere oranı
F1-Skoru	$2 \cdot \frac{P \cdot R}{P+R}$	Kesinlik ve duyarlılık değerlerinin harmonik ortalaması
mAP50	-	IoU eşiği 0.50 iken sınıf başına ortalama hassasiyet
mAP50-95	-	IoU eşiği 0.50'den 0.95'e kadar 0.05 adımlarla hesaplanan ortalama hassasiyet; konumlandırma hassasiyetinin göstergesi

Tüm modellerin karşılaştırmalı testlerinde uygulanan sabit değerler Tablo 4.5'te sunulmuştur.

Tablo 4.5 Değerlendirmede uygulanan eşik değerleri.

Parametre	Değer	Açıklama
Güven Eşiği	0.25	Tahminlerin kabul edilmesi için minimum güven skoru
IoU Eşiği (NMS)	0.45	Maksimum Olmayan Bastırma için örtüşme eşiği
Son İşlem	Model bağımlı	YOLOv10x için NMS atlandı; diğerlerinde standart NMS uygulandı

Üç modelin temel hiperparametre ve yapılandırma farklılıkları Tablo 4.6'de karşılaştırmalı olarak sunulmuştur.

Tablo 4.6 Modeller arası temel hiperparametre ve yapılandırma karşılaştırması.

Model	Giriş Boyutu	Öğrenme Oranı	Ağırlık Sönümü	Yığın Boyutu	Mozaik
YOLOv8x	1536	0.003	0.02	Otomatik	1.0
YOLOv10x	1536	0.003	0.02	Otomatik	1.0
YOLO12X	1536	0.0003	0.002	1	1.0

5
DENEYSEL SONUÇLAR VE TARTIŞMA

Bu bölümde, SeaDronesSee veri setinin üç temel sınıfı (Human, Boat, Buoy) üzerinde eğitilen YOLOv8x, YOLOv10x ve deneysel YOLO12X modellerinin eğitim süreçleri, karşılaştırmalı performans analizleri ve nihai test sonuçları sunulmaktadır. Çalışma kapsamında literatürde incelenen diğer YOLO sürümleri (v9, v11 vb.) yerine; hesaplama gücü kısıtları ve mimari çeşitlilik (klasik, NMS-free ve dikkat odaklı) göz önüne alınarak belirlenen bu üç model üzerine odaklanılmıştır. Aşağıda sunulan tüm sayısal veriler ve grafikler, eğitim sırasında kaydedilen log dosyaları ve test aşamasında üretilen metrik raporlarına dayanmaktadır.

5.1 Eğitim Süreci ve Metodoloji

Modellerin eğitimi, Google Colab platformu üzerinde tahsis edilen ve yüksek işlem kapasitesi sunan bir sanal makine üzerinde gerçekleştirilmiştir. Donanım altyapısı; NVIDIA A100-SXM4-80GB GPU, 167 GB sistem belleği ve 12 mantıksal CPU çekirdeğinden oluşmaktadır. Eğitim sürecinde aşırı öğrenmeyi engellemek ve modelin genelleme yeteneğini korumak adına erken durdurma mekanizması aktif edilmiş, maksimum epoch sayısı ise 300 olarak sınırlandırılmıştır.

SeaDronesSee veri seti, model başarısını objektif bir şekilde ölçebilmek amacıyla eğitim, doğrulama ve test olmak üzere üç bağımsız alt kümeye ayrılmıştır. Toplam 63. 081 etiketli nesne içeren veri setinin dağılımı Tablo 5.1'de verilmiştir. Human sınıfı %60,9 ile en baskın kategoriyi oluştururken, Buoy sınıfı yalnızca %9,7'lik bir paya sahiptir.

Tablo 5.1 Veri seti sınıf ve görüntü dağılımı.

Sınıf	Eğitim	Doğrulama	Test	Toplam	Oran (%)
Human	26.946	7.637	3.839	38.422	60,9
Boat	13.035	3.619	1.873	18.527	29,4
Buoy	4.253	1.245	634	6.132	9,7
Görüntü Sayısı	6.225	1.778	890	8.893	–

Üç model için kullanılan temel eğitim parametreleri Tablo 5.2'te karşılaştırmalı olarak sunulmuştur. Özellikle YOLO12X modelinin eğitimi sırasında gözlemlenen kayıp stabilizasyon sorunları ve Sayısal Değil (NaN) hataları nedeniyle, bu model özelinde öğrenme oranı düşürülmüş, ısınma periyodu uzatılmış ve yığın boyutu küçültülmüştür. Bu müdahaleler, dikkat mekanizması tabanlı mimarilerin yüksek çözünürlüklü görüntülerde eğitilirken karşılaşılan tipik zorluklara işaret etmektedir.

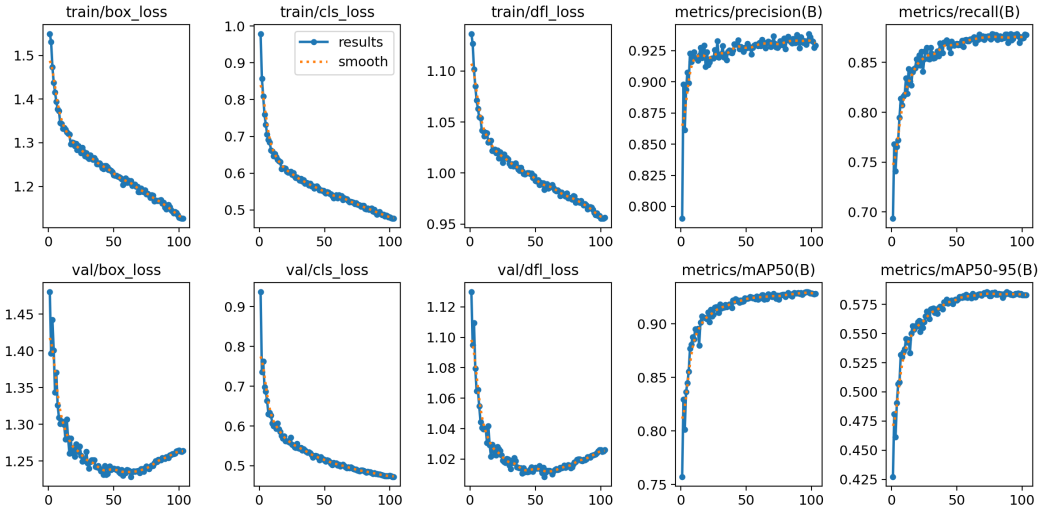
Tablo 5.2 Modellerin eğitim hiperparametreleri karşılaştırması.

Parametre	YOLOv8x	YOLOv10x	YOLO12X
Ön-eğitimli Ağırlık	yolov8x.pt	yolov10x.pt	yolo12x.pt
Görüntü Boyutu	1536 piksel	1536 piksel	1536 piksel
Yığın Boyutu	Otomatik	Otomatik	1 (Stabilizasyon için)
Maksimum Epoch	300	300	300
Sabır (Patience)	30	30	50
Eniyileyici	AdamW	AdamW	AdamW
Başlangıç Öğrenme Oranı	0,003	0,003	0,0003
Ağırlık Sönümü	0,02	0,02	0,002
Isınma Epoch Sayısı	3	3	15
Mozaik Artırma	1,0	1,0	1,0
Karışık Hassasiyet (AMP)	Evet	Evet	Evet

5.2 Eğitim Dinamikleri ve Kayıp Analizi

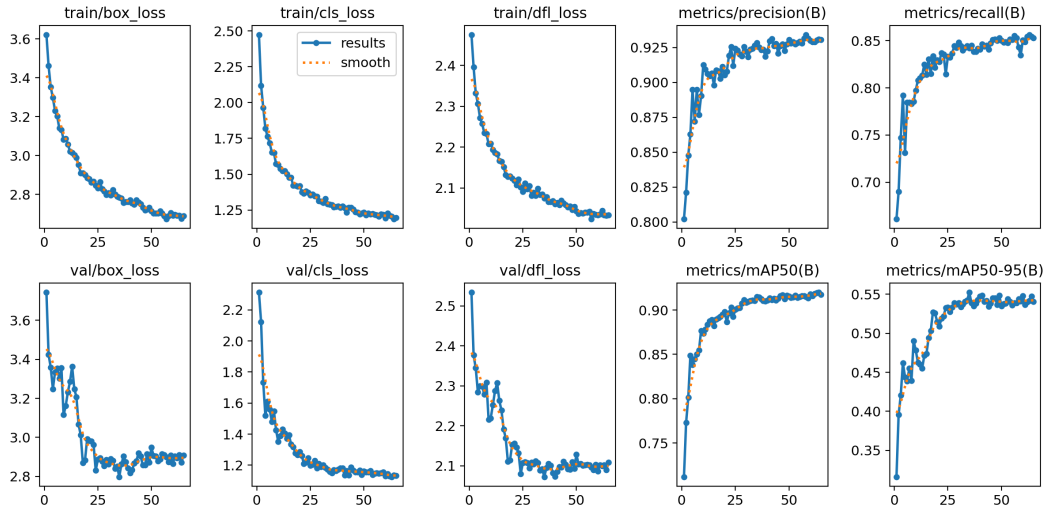
Modellerin eğitim süreçleri incelendiğinde, her üç mimarinin de kayıp fonksiyonlarını minimize etme konusunda farklı karakteristikler sergilediği görülmüştür. Bu farklılıklar, mimarilerin tasarım felsefelerini ve deniz ortamındaki zorlu koşullara uyum yeteneklerini yansıtmaktadır.

YOLOv8x, toplam 103 epoch süren eğitimini yaklaşık 12,2 saatte tamamlamış ve en iyi doğrulama skorlarına 73. epoch'ta ulaşmıştır. Kutu kaybı 1,55'ten 1,12'ye (%27 azalış), sınıflandırma kaybı ise 0,98'den 0,47'ye (%51 azalış) gerilemiştir. Epoch 73 itibarıyla doğrulama setinde mAP50-95: 0,586 değerine ulaşarak eğitim sürecindeki en kararlı ve yüksek performanslı model olmuştur. Bu sonuçlar, YOLOv8x'in ayrıştırılmış kafa yapısının ve C2f modüllerinin, deniz ortamındaki heterojen nesne boyutlarına başarıyla uyum sağladığını göstermektedir.



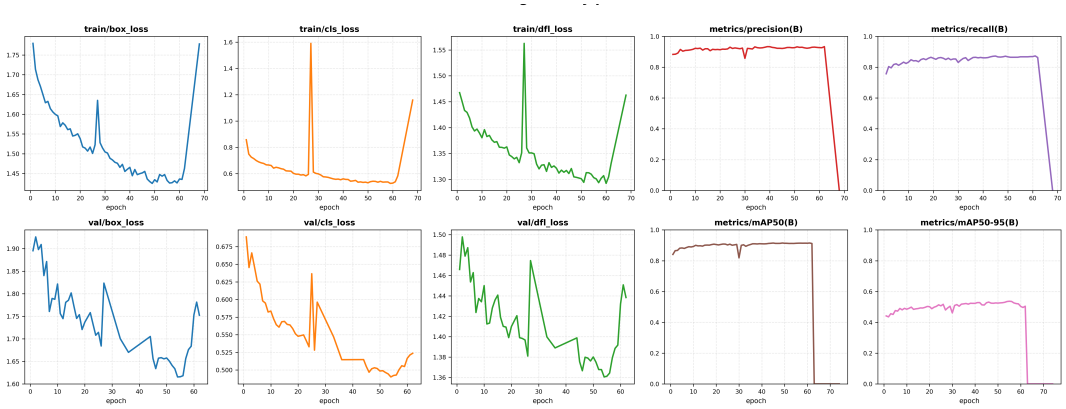
Şekil 5.1 YOLOv8x modeli eğitim ve doğrulama kayıp eğrileri.

Maksimum Olmayan Bastırma gerektirmeyen mimariye sahip YOLOv10x, daha hızlı yakınsama göstererek henüz 35. epoch'ta en iyi değerine ulaşmış ve toplam 65 epoch sonunda erken durdurma ile eğitimi tamamlamıştır. Başlangıç kutu kaybı (3,62), mimari farklılık nedeniyle YOLOv8x'e göre daha yüksek başlamış ancak epoch 65'te 2,68 seviyesine inmiştir. Doğrulama setinde mAP50-95: 0,552 skorunu elde etmiştir.



Şekil 5.2 YOLOv10x modeli eğitim ve doğrulama kayıp eğrileri.

YOLO12X eğitimi, sayısal kararsızlıklar ve kayıp fonksiyonunda NaN oluşumu nedeniyle zorlu bir süreç olmuştur. Optimize edilen parametrelere rağmen belirli periyotlarda (örneğin Epoch 21 ve 63 sonrası) gradyan patlamaları gözlemlenmiştir. Model, Epoch 46-56 aralığında stabil hale gelmiş ve en iyi performansını (mAP50-95: 0,535) bu aralıkta sergilemiştir. Diğer modellere kıyasla kayıp fonksiyonundaki düşüş oranı daha sınırlı kalmıştır (kutu kaybı: %20 düşüş). Bu deneysel süreç, dikkat mekanizması tabanlı mimarilerin yüksek çözünürlüklü ve küçük nesneli veri setlerinde eğitilirken özel stabilizasyon stratejilerine ihtiyaç duyduğunu açıkça ortaya koymaktadır. Özellikle Alan Dikkat mekanizmasının hesaplama yoğunluğu ve gradyan akışındaki hassasiyeti, bu zorlukların temel kaynağı olarak değerlendirilmektedir.



Şekil 5.3 YOLO12X modeli eğitim ve doğrulama kayıp eğrileri (NaN bölgeleri işaretli).

5.3 Test Sonuçları ve Karşılaştırma

Eğitilen modellerin nihai başarımları, eğitim setinden tamamen bağımsız olan 890 görüntülük test seti üzerinde değerlendirilmiştir. Test aşamasında güven eşiği 0,25 ve IoU eşiği 0,45 olarak ayarlanmıştır. Bu eşik değerleri, literatürdeki standart uygulamalarla tutarlı olup, modeller arası adil bir karşılaştırma yapılmasını sağlamaktadır.

Tablo 5.3'de modellerin genel performans metrikleri özetlenmiştir. YOLOv8x, mAP ve duyarlılık açısından en dengeli performansı sunarken, YOLO12X en yüksek kesinlik değerine ulaşmıştır. Bu sonuçlar, farklı mimarilerin farklı hata profillerine sahip olduğunu göstermektedir: YOLOv8x nesneleri kaçırmama konusunda üstünken, YOLO12X tespit ettiği nesnelerin doğruluğu konusunda öne çıkmaktadır.

Tablo 5.3 Test seti genel performans metrikleri.

Model	Kesinlik (P)	Duyarlılık (R)	F1-Skoru	mAP50	mAP50-95
YOLOv8x	0,919	0,876	0,897	0,923	0,577
YOLOv10x	0,914	0,831	0,871	0,897	0,578
YOLO12X	0,926	0,820	0,870	0,890	0,528

Duyarlılık metriği, gerçekte var olan nesnelerin ne kadarının tespit edildiğini gösterdiğinden, bir kazazedeyi kaçırmanın maliyetinin çok yüksek olduğu SAR senaryolarında en önemli gösterge olarak kabul edilmektedir. Bu perspektiften bakıldığında, YOLOv8x'in %87,6'lık duyarlılık değeri, rakiplerine göre (%83,1 ve %82,0) belirgin bir üstünlük sağlamaktadır. Öte yandan, kesinlik metriği yanlış alarmların kontrolünde önemlidir; sürekli yanlış pozitif üreten bir sistem operatör yorgunluğuna ve güven kaybına yol açabilir. YOLO12X'in %92,6'lık kesinlik değeri, bu model tarafından üretilen alarmların büyük çoğunluğunun gerçek nesne olduğunu göstermektedir.

Deniz ortamındaki nesnelerin tespiti, nesne boyutlarına ve dalga etkisine göre büyük farklılıklar göstermektedir. Tablo 5.4, her bir sınıf için modellerin başarısını detaylandırmaktadır.

Tablo 5.4 Sınıf bazlı test sonuçları (mAP50-95).

Model	Sınıf	Kesinlik	Duyarlılık	mAP50-95
YOLOv8x	Human	0,918	0,908	0,485
	Boat	0,945	0,914	0,723
	Buoy	0,894	0,808	0,525
YOLOv10x	Human	0,888	0,872	0,486
	Boat	0,942	0,880	0,695
	Buoy	0,911	0,741	0,553
YOLO12X	Human	0,926	0,881	0,469
	Boat	0,926	0,892	0,622
	Buoy	0,928	0,688	0,493

İnsan tespiti, küçük nesne boyutuna rağmen yüksek başarıyla gerçekleştirilmiştir. YOLOv8x, %90,8'lik duyarlılık değeriyle insanları kaçırmama konusunda en başarılı model olmuştur. Bu sonuç, deniz arama-kurtarma operasyonları için kritik öneme sahiptir çünkü her kaçırılan insan potansiyel bir can kaybı anlamına gelmektedir. YOLO12X ise %92,6 ile tespit ettiklerinin doğruluğu konusunda öne çıkmıştır; ancak %88,1'lik duyarlılık değeri, bazı insanların gözden kaçırıldığını göstermektedir.



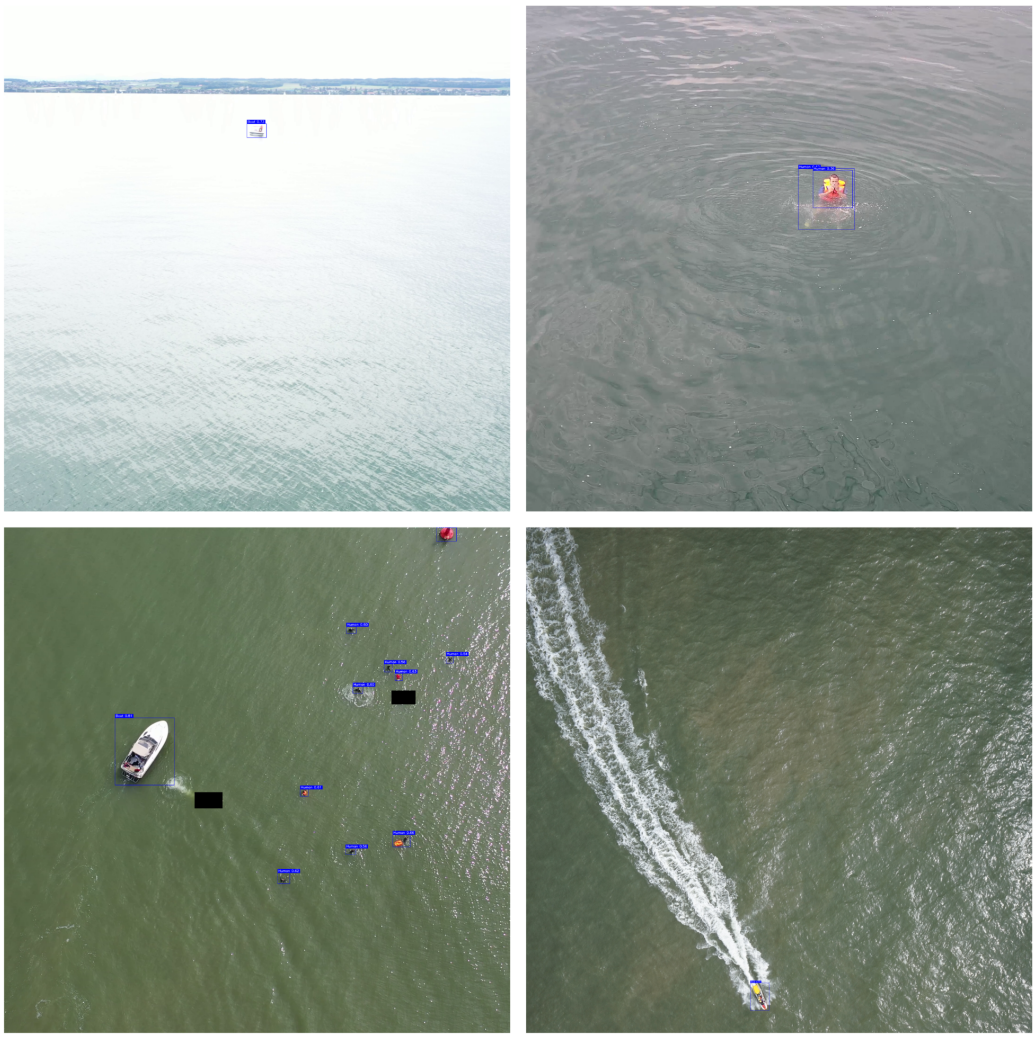
Şekil 5.4 YOLOv8x modeli test örnekleri.

Tekne sınıfında, büyük boyutları ve belirgin konturları nedeniyle en yüksek başarım (mAP > 0,70) elde edilmiştir. YOLOv8x ve YOLOv10x benzer yüksek performans göstermiştir. Teknelerin diğer sınıflara göre daha kolay tespit edilmesi, modellerin çok ölçekli özellik birleştirme mekanizmalarının büyük nesnelerde etkin çalıştığını doğrulamaktadır.



Şekil 5.5 YOLOv10x modeli test örnekleri.

Şamandıra sınıfı, en zorlu tespit kategorisi olarak öne çıkmıştır. YOLOv10x, mAP50-95 (0,553) skorunda en iyi sonucu verirken; YOLOv8x, duyarlılık (0,808) açısından, yani şamandıraları gözden kaçırmama konusunda rakiplerine büyük fark atmıştır. YOLO12X'in şamandıra duyarlılığının yalnızca %68,8'de kalması, dikkat mekanizmalarının küçük ve düşük kontrastlı nesnelerde henüz yeterli performansı sağlayamadığını göstermektedir. Bu durum, şamandıraların genellikle dalga köpükleriyle karışan küçük boyutlu ve düşük kontrastlı nesneler olmasıyla açıklanabilir.



Şekil 5.6 YOLO12X modeli test örnekleri.

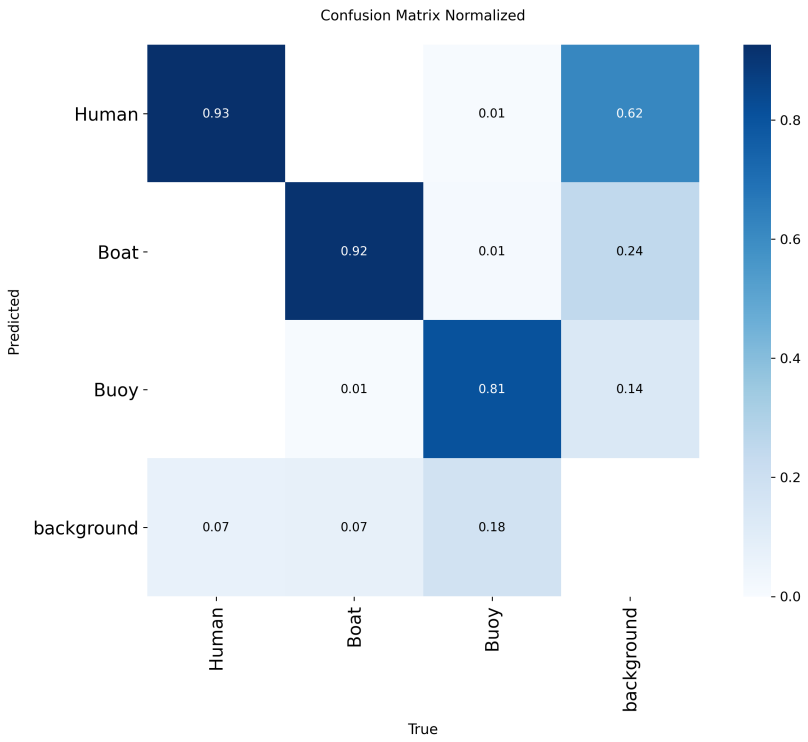
5.4 Karmaşıklık Matrisi Değerlendirmesi

Modellerin sınıflandırma performansını ve özellikle nesneleri arka planla karıştırma durumlarını analiz etmek amacıyla normalleştirilmiş karmaşıklık matrisleri incelenmiştir. Karmaşıklık matrisi, bir modelin hangi sınıfları birbiriyle karıştırdığını ve hangi nesneleri tespit edemediğini görsel olarak ortaya koyan güçlü bir analiz aracıdır. Modellerin doğru sınıflandırma oranları Tablo 5.5'da, nesne kaçırma oranları ise Tablo 5.6'da özetlenmiştir.

Tablo 5.5 Modellerin doğru sınıflandırma oranları (karmaşıklık matrisi köşegen değerleri).

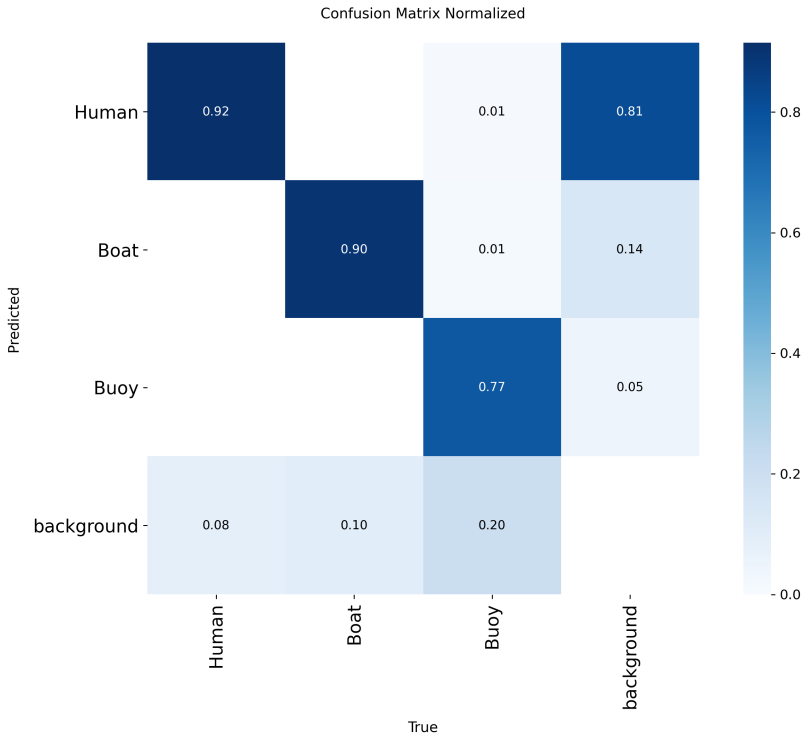
Sınıf	YOLOv8x	YOLOv10x	YOLO12X
Human (İnsan)	%94	%92	%90
Boat (Tekne)	%93	%92	%90
Buoy (Şamandıra)	%80	%78	%70

YOLOv8x, insan tespitinde %94 doğru sınıflandırma oranıyla en başarılı model olmuştur. Arka plan hatası (kaçırma) sadece %6'dır; bu oran, her 100 insandan yalnızca 6'sının tespit edilemediği anlamına gelmektedir. Şamandıra tespitinde ise %80 başarı yakalamış, kaçırma oranını %18'de tutarak rakiplerine belirgin bir üstünlük sağlamıştır. Bu sonuçlar, YOLOv8x'in hem büyük hem de küçük nesnelerde dengeli bir performans sergilediğini ortaya koymaktadır.



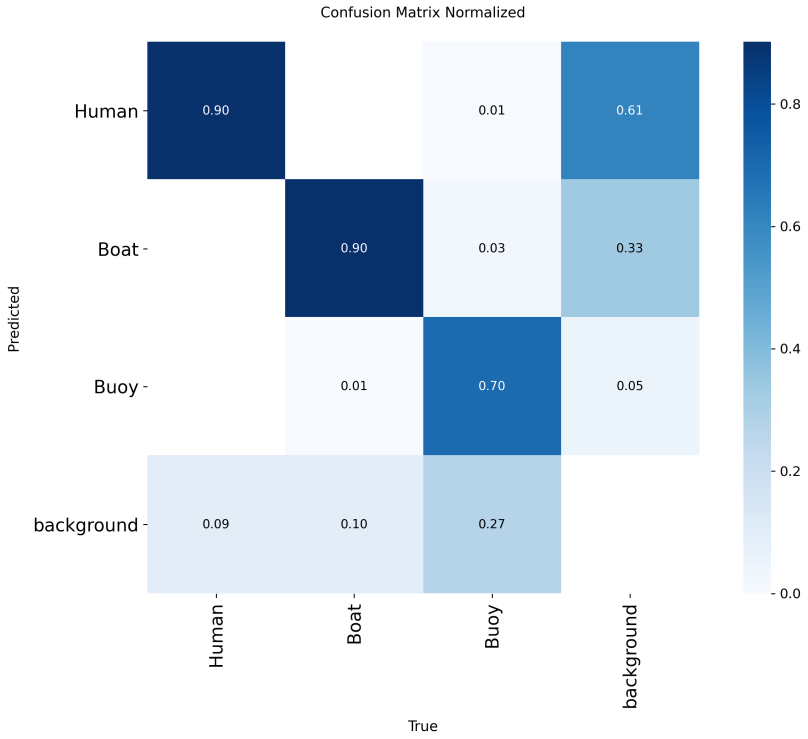
Şekil 5.7 YOLOv8x modeli normalize edilmiş karmaşıklık matrisi (test seti).

YOLOv10x'in en zayıf olduğu alan şamandıra tespitidir. Şamandıraları doğru bilme oranı %78'de kalmış, %21 oranında şamandıraları arka plan olarak etiketleyip kaçırmıştır. Bu durum, Maksimum Olmayan Bastırma gerektirmeyen mimarisinin küçük nesnelerde aşırı tutucu davrandığını göstermektedir. Birden bire başlık yapısının, düşük güven skorlu küçük nesneleri elemesi bu sonucun muhtemel nedenidir.



Şekil 5.8 YOLOv10x modeli normalize edilmiş karmaşıklık matrisi (test seti).

YOLO12X, insan tespitinde %90 doğru sınıflandırma oranıyla üçüncü sırada yer almıştır. Şamandıra tespitinde %70 başarı oranıyla en düşük performansı sergilemiş ve %27 oranında şamandırayı kaçırmıştır. Alan Dikkat mekanizmasının küçük nesneleri bölgesel olarak işlemesi, beklenen avantajı sağlayamamış ve YOLOv8x'in çok ölçekli özellik birleştirme yaklaşımının gerisinde kalmıştır.



Şekil 5.9 YOLO12X modeli normalize edilmiş karmaşıklık matrisi (test seti).

Arama-kurtarma operasyonlarında en kritik hata türü olan nesneyi kaçırma oranları Tablo 5.6'da özetlenmiştir. Bu metrik, gerçekte var olan bir nesnenin model tarafından tespit edilememesi durumunu ifade etmekte olup, SAR operasyonlarında doğrudan can kaybı riskiyle ilişkilendirilmektedir.

Tablo 5.6 Modellerin nesne kaçırma (Arka Plan Yanlış Negatif) oranları.

Sınıf	YOLOv8x	YOLOv10x	YOLO12X
Human (İnsan)	%6 (En İyi)	%7	%9
Boat (Tekne)	%6 (En İyi)	%7	%10
Buoy (Şamandıra)	%18 (En İyi)	%21	%27 (En Kötü)

Bu veriler ışığında, YOLOv8x'in arama-kurtarma görevleri için en güvenilir seçenek olduğu sonucuna varılmıştır. Tüm sınıflarda en yüksek doğru sınıflandırma oranlarına ve en düşük kaçırma oranlarına sahip olan bu model, kritik

SAR senaryolarında can kaybı riskini minimize etme potansiyeli taşımaktadır. YOLOv10x, Maksimum Olmayan Bastırma gerektirmeyen yapısının sağladığı hız avantajına rağmen, özellikle küçük nesnelerde düşük duyarlılık değerleriyle sınırlı kalmakta ve bu durum modelin SAR uygulamalarındaki kullanılabilirliğini azaltmaktadır. YOLO12X ise deneysel bir model olarak dikkat mekanizması tabanlı mimarilerin potansiyelini ortaya koymasına rağmen, eğitim sürecindeki stabilizasyon zorlukları ve tüm sınıflarda en düşük doğru sınıflandırma oranları nedeniyle operasyonel kullanım için kapsamlı optimizasyon gerektirmektedir.

6
SONUÇ

Bu tez çalışması kapsamında; insansız hava araçlarından elde edilen görüntüler üzerinden deniz ortamında arama-kurtarma faaliyetlerini desteklemek amacıyla, derin öğrenme tabanlı güncel nesne tespit modellerinin performansı karşılaştırmalı olarak incelenmiştir. SeaDronesSee veri seti temel alınarak YOLOv8x, YOLOv10x ve YOLO12X modelleri; Human (İnsan), Boat (Tekne) ve Buoy (Şamandıra) sınıfları üzerinde eğitilmiş ve test edilmiştir. Çalışma, modellerin yalnızca teorik mimari üstünlüklerini değil, dalga hareketi, yüzey yansımaları ve küçük nesne boyutu gibi gerçek dünya zorlukları karşısındaki pratik başarımını da ortaya koymaktadır.

6.1 Genel Değerlendirme ve Çıkarımlar

Modellerin eğitimi ve test süreçleri, standardizasyonu sağlamak adına özdeş donanım altyapısı (NVIDIA A100) ve sabit veri bölünmesi (Eğitim: 6.225, Doğrulama: 1.778, Test: 890 görüntü) altında gerçekleştirilmiştir. Deneysel sonuçların özeti Tablo 6.1'da sunulmuştur.

Tablo 6.1 Modellerin genel performans karşılaştırması.

Model	mAP50	mAP50-95	Kesinlik	Duyarlılık	Güçlü Yön
YOLOv8x	%92,3	%57,7	%91,9	%87,6	Dengeli performans
YOLOv10x	%89,7	%57,8	%91,4	%83,1	Düşük gecikme
YOLO12X	%89,0	%52,8	%92,6	%82,0	Yüksek kesinlik

YOLOv8x, %92,3 mAP50 ve %87,6 duyarlılık değerleri ile genel performansta öne çıkmıştır. Özellikle insan sınıfındaki %94'lük doğru sınıflandırma başarısı ve

yalnızca %6'lık kaçırma oranı, modeli nesne kaçırmama konusunda en güvenilir seçenek haline getirmiştir. Bu sonuç, deniz arama-kurtarma operasyonlarında her kazazedenin tespit edilmesinin kritik öneme sahip olduğu düşünüldüğünde, YOLOv8x'in SAR uygulamaları için birincil tercih olması gerektiğini açıkça ortaya koymaktadır.

YOLOv10x, Maksimum Olmayan Bastırma gerektirmeyen mimarisi sayesinde son işlem süresini minimize ederek en düşük gecikmeyi sunmuştur. Genel başarımı (mAP50-95: %57,8) rekabetçi olsa da, küçük nesnelerde (şamandıra) %21'e varan kaçırma oranıyla daha tutucu bir profil sergilemiştir. Tutarlı Çift Atama stratejisinin hızlı yakınsama sağlamasına rağmen, birden bire başlık yapısının düşük güven skorlu küçük nesneleri elemesi, modelin SAR senaryolarındaki kullanılabilirliğini sınırlamaktadır.

YOLO12X, %92,6 kesinlik değeriyle yanlış alarm üretmeme konusunda en başarılı model olmuştur. Ancak eğitim sürecinde gözlemlenen stabilite sorunları, gradyan patlamaları ve rakiplerine göre düşük kalan duyarlılık (%82,0) değeri, genel mAP50-95 başarısını %52,8 seviyesinde sınırlamıştır. Dikkat mekanizması tabanlı mimarilerin yüksek çözünürlüklü ve küçük nesneli veri setlerinde eğitilirken özel stabilizasyon stratejilerine ihtiyaç duyduğu bu çalışmayla ampirik olarak doğrulanmıştır.

Tez kapsamında belirlenen hipotezlerin değerlendirilmesi şu şekildedir: Maksimum Olmayan Bastırma gerektirmeyen mimarilerin gecikmeyi azaltacağı hipotezi, YOLOv10x'in işlem süresindeki belirgin düşüş ile doğrulanmıştır. Daha yeni mimarinin her zaman daha iyi sonuç vereceği hipotezi ise YOLO12X'in genel mAP skorunda YOLOv8x'in gerisinde kalmasıyla çürütülmüştür. Deniz koşullarının tespiti zorlaştıracağı hipotezi, tüm modellerin şamandıra sınıfındaki düşük performansı ve artan arka plan hataları ile teyit edilmiştir.

6.2 Denizde İnsan Tespiti İçin Model Önerisi

Elde edilen bulgular ışığında, operasyonel gereksinimlere göre model seçim stratejisi Tablo 6.2'da özetlenmiştir.

Tablo 6.2 Operasyonel gereksinimlere göre model önerileri.

Senaryo	Önerilen Model	Gerekçe
Arama-Kurtarma (SAR) Operasyonları	YOLOv8x	%87,6 duyarlılık, %6 insan kaçırma oranı
Canlı İzleme ve Otonom Seyir	YOLOv10x	Düşük gecikme, yüksek FPS değeri
Uzun Süreli Kıyı Güvenliği	YOLO12X	%92,6 kesinlik, düşük yanlış alarm oranı

Arama-kurtarma operasyonlarında hayat kurtarma odaklı senaryolarda tek bir kazazedeyi dahi kaçırmanın maliyeti tolere edilemez. YOLOv8x, %87,6'lık genel duyarlılığı, insan sınıfındaki %94'lük doğru sınıflandırma başarısı ve tüm sınıflarda en düşük kaçırma oranları ile bu görev için en uygun mimaridir. Canlı izleme ve otonom seyir sistemlerinde düşük gecikme süresinin kritik olduğu ve anlık manevra gerektiren durumlarda, Maksimum Olmayan Bastırma gerektirmeyen yapısı ve yüksek kare hızı değeri nedeniyle YOLOv10x tercih edilmelidir. Sistemin sürekli hatalı uyarı vererek operatör yorgunluğuna neden olmaması gereken uzun süreli kıyı güvenliği senaryolarında ise %92,6'lık yüksek kesinlik değeri nedeniyle YOLO12X alternatif olarak değerlendirilebilir.

6.3 Kısıtlar ve Gelecek Çalışmalar

Bu çalışmanın sonuçları; kullanılan donanım, SeaDronesSee veri seti ve seçilen hiperparametre uzayı ile sınırlıdır. Sonuç olarak bu tez; derin öğrenme tabanlı nesne tespit modellerinin deniz güvenliğinde yüksek potansiyele sahip olduğunu, ancak en iyi model kavramının operasyonel önceliklere göre değişkenlik gösterdiğini bilimsel verilerle kanıtlamıştır. YOLOv8x'in dengeli performansı, YOLOv10x'in hız avantajı ve YOLO12X'in kesinlik odaklı yapısı, farklı kullanım senaryolarında birbirini tamamlayan seçenekler sunmaktadır.

KAYNAKÇA

[1] L. A. Varga, B. Kiefer, M. Messmer A. Zell, "SeaDronesSee: A Maritime Benchmark for Detecting Humans in Open Water," *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2022, ss. 2260–2270. doi: 10.1109/WACV51458.2022.00374.

[2] J. Redmon, S. Divvala, R. Girshick A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, ss. 779–788. doi: 10.1109/CVPR.2016.91.

[3] A. Bochkovskiy, C.-Y. Wang H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," 2020. arXiv: 2004.10934 [cs.CV].

[4] World Health Organization, "Global Status Report on Drowning Prevention," WHO, Geneva, Switzerland, Report, 2024.

[5] World Health Organization. "Drowning Deaths Decline Globally but the Most Vulnerable Remain at Risk." (2024), erişim adresi: https://www.who.int/news/item/13-12-2024-drowning-deaths-decline-globally-but-the-most-vulnerable-remain-at-risk (erişim tarihi 26/05/2025).

[6] E. Lygouras, N. Santavas, A. Taitzoglou, K. Tarchanidis, A. Mitropoulos A. Gasteratos, "Unsupervised Human Detection with an Embedded Vision System on a Fully Autonomous UAV for Search and Rescue Operations," *Sensors*, c. 19, no. 16, s. 3542, 2019. doi: 10.3390/s19163542.

[7] R. Mendonça, P. Santana, F. Marques, A. Lourenço, J. Silva J. Barata, "Kelpie: A Heterogeneous Multi-Robot SAR System for an Aerial-Aquatic Environment," *Proceedings of the IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, 2020, ss. 299–304. doi: 10.1109/ICARSC49921.2020.9096114.

[8] Y. Chen, H. Liu W. Zhang, "LFN-YOLO: Precision Underwater Small Object Detection via a Lightweight Reparameterized Approach," *Frontiers in Marine Science*, c. 12, s. 1 513 740, 2025. doi: 10.3389/fmars.2025.1513740.

[9] X. Li, J. Wang M. Zhou, "Maritime Small Object Detection Algorithm in Drone Aerial Images Based on Improved YOLOv8," *Scientific Reports*, c. 14, s. 25 542 148, 2024. doi: 10.1038/s41598-024-25542148.

[10] Y.-L. Chung, L.-H. Lin K.-C. Huang, "Efficient Shadow Detection of Color Aerial Images Based on Successive Thresholding Scheme," *IEEE Transactions on Geoscience and Remote Sensing*, c. 50, no. 2, ss. 679–690, 2012. doi: 10.1109/TGRS.2011.2165069.

[11] P. Westall, J. J. Ford, P. O'Shea S. Hrabar, "Vision-Based UAV Maritime Search and Rescue Using Point Target Detection," *Proceedings of the Australasian Conference on Robotics and Automation*, 2007.

[12] S. Fefilatyev, D. Goldgof, M. Shreve C. Lembke, "Detection and Tracking of Ships in Open Sea with Rapidly Moving Buoy-Mounted Camera System," *Ocean Engineering*, c. 54, ss. 1–12, 2012. doi: 10.1016/j.oceaneng.2012.06.028.

[13] M. A. Maruf S. Azim, "Person Detection in Maritime Search and Rescue Operations," *Proceedings of the International Conference on Information Technology Systems and Innovation (ICITSI)*, 2022, ss. 1–6. doi: 10.1109/ICITSI56531.2022.9970853.

[14] F. Gomez-Donoso, A. Escalona-Morán, F. J. Simó-Ten M. Cazorla, "Haar-like Features for Object Detection in Maritime Environments," *Proceedings of the Robot 2015: Second Iberian Robotics Conference*, 2015, ss. 537–548. doi: 10.1007/978-3-319-27149-1_41.

[15] A. Toet M. A. Hogervorst, "Progress in Color Night Vision," *Optical Engineering*, c. 51, no. 1, s. 010 901, 2012. doi: 10.1117/1.OE.51.1.010901.

[16] N. Dalal B. Triggs, "Histograms of Oriented Gradients for Human Detection," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, c. 1, 2005, ss. 886–893. doi: 10.1109/CVPR.2005.177.

[17] Y. Freund R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," *Journal of Computer and System Sciences*, c. 55, no. 1, ss. 119–139, 1997. doi: 10.1006/jcss.1997.1504.

[18] S. Ren, K. He, R. Girshick J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, c. 39, no. 6, ss. 1137–1149, 2017. doi: 10.1109/TPAMI.2016.2577031.

[19] K. He, G. Gkioxari, P. Dollár R. Girshick, "Mask R-CNN," *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, ss. 2961–2969. doi: 10.1109/ICCV.2017.322.

[20] W. Liu ve diğ., "SSD: Single Shot MultiBox Detector," *Proceedings of the European Conference on Computer Vision (ECCV)*, 2016, ss. 21–37. doi: 10.1007/978-3-319-46448-0_2.

[21] J. Redmon A. Farhadi, "YOLOv3: An Incremental Improvement," 2018. arXiv: 1804.02767 [cs.CV].

[22] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov S. Zagoruyko, "End-to-End Object Detection with Transformers," *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020, ss. 213–229. doi: 10.1007/978-3-030-58452-8_13.

[23] C. Cortes V. Vapnik, "Support-Vector Networks," *Machine Learning*, c. 20, no. 3, ss. 273–297, 1995. doi: 10.1007/BF00994018.

[24] P. Viola M. Jones, "Rapid Object Detection Using a Boosted Cascade of Simple Features," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, c. 1, 2001, ss. 511–518. doi: 10.1109/CVPR.2001.990517.

[25] A. Neubeck L. Van Gool, "Efficient Non-Maximum Suppression," *Proceedings of the 18th International Conference on Pattern Recognition (ICPR)*, c. 3, 2006, ss. 850–855. doi: 10.1109/ICPR.2006.479.

[26] V.-K. Nguyen, "Large-Scale Object Detection for Social Media User's Visual Privacy Protection," Master's Thesis, University of Rennes 1, Eyl. 2020.

[27] S. Ren, K. He, R. Girshick J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, c. 28, 2015, ss. 91–99.

[28] T.-Y. Lin, P. Goyal, R. Girshick, K. He P. Dollár, "Focal Loss for Dense Object Detection," *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, ss. 2980–2988. doi: 10.1109/ICCV.2017.324.

[29] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid S. Savarese, "Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, ss. 658–666. doi: 10.1109/CVPR.2019.00075.

[30] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye D. Ren, "Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression," *Proceedings of the AAAI Conference on Artificial Intelligence*, c. 34, 2020, ss. 12 993–13 000. doi: 10.1609/aaai.v34i07.6999.

[31] R. Padilla, W. L. Passos, T. L. B. Dias, S. L. Netto E. A. B. da Silva, "A Comparative Analysis of Object Detection Metrics with a Companion Open-Source Toolkit," *Electronics*, c. 10, no. 3, s. 279, 2021. doi: 10.3390/electronics10030279.

[32] O. T. Oladipupo, O. G. Abimbola A. O. Afolabi, "Sign-to-Speech Model for Sign Language Understanding: A Case Study of Nigerian Sign Language," *International Journal of Advanced Computer Science and Applications (IJACSA)*, c. 12, no. 11, 2021. doi: 10.14569/IJACSA.2021.0121103.

[33] Y. Tian, C. Ye D. Doermann, "YOLOv12: Attention-Centric Real-Time Object Detectors," 2025. arXiv: 2502.12524 [cs.CV].

[34] J. Redmon A. Farhadi, "YOLO9000: Better, Faster, Stronger," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, ss. 7263–7271. doi: 10.1109/CVPR.2017.690.

[35] G. Jocher ve diğ. "YOLOv5 by Ultralytics." (2020), erişim adresi: https://github.com/ultralytics/yolov5 (erişim tarihi 15/01/2024).

[36] C.-Y. Wang, A. Bochkovskiy H.-Y. M. Liao, "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, ss. 7464–7475. doi: 10.1109/CVPR52729.2023.00721.

[37] G. Jocher, A. Chaurasia J. Qiu. "Ultralytics YOLOv8." (2023), erişim adresi: https://github.com/ultralytics/ultralytics (erişim tarihi 15/01/2024).

[38] C.-Y. Wang, I.-H. Yeh H.-Y. M. Liao, "YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information," *Proceedings of the European Conference on Computer Vision (ECCV)*, 2024. doi: 10.1007/978-3-031-72751-1_1.

[39] G. Jocher, "YOLO11: Improved Efficiency and Accuracy in Real-Time Object Detection," *Proceedings of the YOLO Vision 2024*, Madrid, Spain, Eyl. 2024.

[40] A. Wang ve diğ., "YOLOv10: Real-Time End-to-End Object Detection," *Proceedings of the Conference on Neural Information Processing Systems (NeurIPS)*, 2024.

[41] Ultralytics. "YOLO12: Attention-Centric Object Detection." (2025), erişim adresi: https://docs.ultralytics.com/models/yolo12/ (erişim tarihi 20/02/2025).

[42] T. Dao, D. Y. Fu, S. Ermon, A. Rudra C. Ré, "FlashAttention: Fast and Memory-Efficient Exact Attention with IO-Awareness," *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, c. 35, 2022, ss. 16 344–16 359.

A DETAYLI DENEYSEL SONUÇLAR

Bu ekte, tez çalışması kapsamında fiilen eğitilip test edilen üç modelin (YOLOv8x, YOLOv10x ve deneysel YOLO12X) gerçek eğitim ve test loglarından elde edilen detaylı sonuçları raporlanmaktadır. Çalışma kapsamında literatürde adı geçen diğer YOLO varyantları (v8n/s/m/l, v9, v11 vb.) eğitilmemiştir; dolayısıyla bu bölümde sadece deney gerçekleştirilen modellere ait veriler sunulmaktadır.

A.1 Modellerin Eğitim Süreçleri

Aşağıdaki bilgiler, eğitim sırasında kaydedilen log dosyalarından derlenmiştir:

- YOLOv8x: EarlyStopping mekanizması 30 epoch sabırla çalışmış, model en iyi değerine 73. epoch'ta ulaşmış ve eğitim toplam 103 epoch sürmüştür. Toplam eğitim süresi: 12.199 saat.
- YOLOv10x: EarlyStopping 30 epoch sabırla devreye girmiş, en iyi epoch 35 olarak belirlenmiş ve eğitim 65 epoch sonunda tamamlanmıştır. Toplam eğitim süresi: 12.236 saat.
- YOLO12X: Eğitim stabilitesini sağlamak adına düşük öğrenme oranı (lr_0) ve $batch = 1$ kullanılmıştır. Doğrulama metriklerinde dönemsel `NaN` (Not a Number) hataları gözlemlenmiş; nihai model, `NaN` içermeyen en iyi pencereden seçilen ağırlıklarla belirlenmiştir.

Tablo A.1 Modellerin eğitim süreci özetleri (Log kayıtlarına dayalı).

Model	Çalışan Epoch	En İyi Epoch	Batch	imgsz	lr0	Optim.	Süre (saat)
YOLOv8x	103	73	Auto	1536	0.003	AdamW	12.199
YOLOv10x	65	35	Auto	1536	0.003	AdamW	12.236
YOLO12X	(Düzensiz)	(best.pt)	1	1536	0.0003	AdamW	(—)

Not: YOLO12X modelinde yaşanan kararsızlıklar nedeniyle standart bir yakınsama epoch değeri raporlanamamış; hiperparametreler stabilizasyon öncelikli olarak ayarlanmıştır.

A.2 Modellerin Detaylı Performans Metrikleri (Test)

Aşağıdaki metrikler, test veri seti (890 görüntü) üzerinde Ultralytics kütüphanesinin `model.val(split="test")` fonksiyonu ile üretilmiştir (Değerlendirme

eşikleri: $conf = 0.25$, $iou = 0.45$).

Tablo A.2 Genel test metrikleri ve hesaplama profili.

Model	Precision	Recall	mAP50	mAP50-95	Param (M)	GFLOPs	FPS (≈)
YOLOv8x	0.919	0.876	0.923	0.577	68.13	257.4	45.9
YOLOv10x	0.914	0.831	0.897	0.578	29.40	160.0	52.6
YOLO12X	0.926	0.820	0.890	0.528	59.05	198.5	31.6

FPS değerleri, log dosyalarındaki ortalama işlem sürelerinden ($Pre-process + Inference + Post-process$) türetilmiştir:

- YOLOv8x: $1.4 + 17.1 + 3.3 \approx 21.8$ ms/görüntü
- YOLOv10x: $0.9 + 17.8 + 0.3 \approx 19.0$ ms/görüntü
- YOLO12X: $1.9 + 28.8 + 1.0 \approx 31.7$ ms/görüntü

A.3 Sınıf Bazlı Değerlendirmeler (Test)

Modellerin sınıf bazlı mAP50 performansları Tablo A.3'de, mAP50-95 performansları ise metin içinde özetlenmiştir.

Tablo A.3 Sınıf bazlı mAP50 değerleri (Test seti).

Model	Human	Buoy	Boat
YOLOv8x	0.936	0.883	0.951
YOLOv10x	0.910	0.845	0.937
YOLO12X	0.916	0.817	0.936

Sınıf bazlı mAP50-95 Değerleri:

- YOLOv8x: Human 0.485, Buoy 0.525, Boat 0.723
- YOLOv10x: Human 0.486, Buoy 0.553, Boat 0.695
- YOLO12X: Human 0.469, Buoy 0.493, Boat 0.622

A.4 Hata Analizleri (Sınıf Bazlı Tahmin Özetleri)

Bu bölümde sunulan değerler; test veri setindeki sınıf sayıları (Human=3839, Boat=1873, Buoy=634) ve modellerin Precision/Recall metrikleri kullanılarak yaklaşık olarak hesaplanmıştır. Hesaplama Yöntemi: $TP = Recall \times Support$, $FN = Support - TP$, $FP \approx TP \times (\frac{1}{Precision} - 1)$.

Tablo A.4 YOLOv8x — Test seti sınıf bazlı yaklaşık TP/FN/FP değerleri.

Sınıf	TP (Doğru)	FN (Kaçan)	FP (Yanlış Alarm)
Human	3486	353	312
Boat	1712	161	100
Buoy	512	122	61

Tablo A.5 YOLOv10x — Test seti sınıf bazlı yaklaşık TP/FN/FP değerleri.

Sınıf	TP (Doğru)	FN (Kaçan)	FP (Yanlış Alarm)
Human	3348	491	422
Boat	1648	225	102
Buoy	470	164	46

Tablo A.6 YOLO12X — Test seti sınıf bazlı yaklaşık TP/FN/FP değerleri.

Sınıf	TP (Doğru)	FN (Kaçan)	FP (Yanlış Alarm)
Human	3382	457	270
Boat	1671	202	134
Buoy	436	198	34

Analiz ve Gözlemler:

- Human: YOLOv8x en yüksek TP sayısına (en az kaçırma) sahiptir. YOLO12X ise en düşük FP sayısına (en az yanlış alarm) sahiptir.
- Boat: Üç model de büyük boyutlu ve belirgin konturlu teknelerde yüksek başarı göstermiştir.
- Buoy: Şamandıra sınıfında FP sayıları genel olarak düşüktür ancak FN (kaçırma) oranları yüksektir. YOLOv10x, mAP50-95 skorunda (IoU hassasiyeti) önde olsa da, tespit sayısı (Recall/TP) bakımından YOLOv8x'in gerisindedir.

A.5 Zorlayıcı Durumlar ve Küçük Nesne Analizi

SeaDronesSee veri setinde "Şamandıra" (Buoy) sınıfı, küçük nesne tespiti performansının en önemli göstergesi (proxy) olarak kabul edilmiştir. Tablo A.7, modellerin bu sınıftaki mAP50-95 başarılarını özetlemektedir.

Tablo A.7 Küçük nesne performans göstergesi (Buoy mAP50-95).

Model	Buoy mAP50-95
YOLOv10x	0.553
YOLOv8x	0.525
YOLO12X	0.493

Yorum: YOLOv10x, küçük nesnelerin hassas lokalizasyonunda (yüksek IoU) en kararlı sonucu vermiştir. YOLOv8x dengeli bir profil çizerken, YOLO12X yüksek kesinliğine rağmen küçük nesneleri yakalama konusunda daha sınırlı kalmıştır.

TEZDEN ÜRETİLMİŞ YAYINLAR

Konferans Bildirisi

1. M. Tetik ve B. Bolat, "Makine öğrenmesi yöntemleri ile deniz üzerindeki insanların tespiti," *GRAD COLLOQUIUM'24 Artificial Intelligence Bildiri Özet Kitabı*, İstanbul, Türkiye: Yıldız Teknik Üniversitesi Yayınevi, 2024, s. 57.