



**YOLO, FASTER R-CNN ve SSD MOBILE NET
KULLANARAK MARKET RAF ÜRÜNLERİNİN
ALGILANMASI VE TANIMLANMASI**

Alaa Mohamed Jamel Abdulqader MAWLAWİ

Danışman: Dr. Öğr. Üyesi Gülşah TÜRÜKLÜ ÖZYER
Yüksek Lisans Tezi

Bilgisayar Mühendisliği Ana Bilim Dalı

2024

(Her hakkı saklıdır.)

T.C.
ATATÜRK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANA BİLİM DALI

**YOLO, FASTER R-CNN ve SSD MOBILE NET KULLANARAK MARKET RAF
ÜRÜNLERİNİN ALGILANMASI VE TANIMLANMASI**

(Market Shelf Products Detection and Recognition Using Yolo, Faster R-CNN, and SSD
Mobile Net)

YÜKSEK LİSANS TEZİ

Alaa Mohamed Jamel Abdulqader MAWLAWİ

Danışman: Dr. Öğr. Üyesi Gülşah TÜMÜKLÜ ÖZYER

Erzurum
Aralık, 2024

KABUL VE ONAY TUTANAĞI

Alaa Mohamed Jamel Abdulqader MAWLAWİ tarafından hazırlanan “YOLO, FASTER R-CNN ve SSD MOBİLE NET KULLANARAK MARKET RAF ÜRÜNLERİNİN ALGILANMASI VE TANIMLANMASI” başlıklı çalışması 16/ 12 / 2024 tarihinde yapılan tez savunma sınavı sonucunda başarılı bulunarak jürimiz tarafından Bilgisayar Mühendisliği Ana Bilim Dalı, Bilgisayar Mühendisliği Bilim Dalında Yüksek Lisans tezi olarak kabul edilmiştir.

Jüri Başkanı:	Prof. Dr. Köksal ERENTÜRK <i>Atatürk Üniversitesi</i>	Aslı Islak İmzalıdır
Danışman:	Dr. Öğr. Üyesi Gülşah TÜMÜKLÜ ÖZYER <i>Atatürk Üniversitesi</i>	Aslı Islak İmzalıdır
Jüri Üyesi:	Dr. Öğr. Üyesi Işıl KARABEY AKSAKALLI <i>Erzurum Teknik Üniversitesi</i>	Aslı Islak İmzalıdır

Enstitü Yönetim
Kurulunun .../.../....
tarih ve sayılı
kararı.

Bu tezin Atatürk Üniversitesi Lisansüstü Eğitim ve Öğretim Yönetmeliği'nin ilgili maddelerinde belirtilen şartları yerine getirdiğini onaylarım.

Prof. Dr. Alper NUHOĞLU

Enstitü Müdürü

Aslı Islak İmzalıdır

Not: Bu tezde kullanılan özgün ve başka kaynaklardan yapılan bildiriş, çizelge, şekil ve fotoğrafların kaynak olarak kullanımı, 5846 sayılı Fikir ve Sanat Eserleri Kanunundaki hükümlere tabidir.

ETİK BİLDİRİM VE İNTİHAL BEYAN FORMU

Yüksek Lisans Tezi olarak Dr. Öğr. Üyesi Gülşah TÜMÜKLÜ ÖZYER danışmanlığında sunulan “YOLO, FASTER R-CNN ve SSD MOBİLE NET KULLANARAK MARKET RAF ÜRÜNLERİNİN ALGILANMASI VE TANIMLANMASI” başlıklı çalışmanın tarafımızdan bilimsel etik ilkelere uyularak yazıldığını, yararlanılan eserlerin kaynakçada gösterildiğini, Fen Bilimleri Enstitüsü tarafından belirlenmiş olan Turnitin Programı benzerlik oranlarının aşılmadığını ve aşağıdaki oranlarda olduğunu beyan ederiz.

Tez Bölümleri	Tezin Benzerlik Oranı (%)	Maksimum Oran (%)
Giriş	8%	30
Kuramsal Temeller	6%	30
Materyal ve Yöntem	9%	35
Araştırma Bulguları ve Tartışma	0%	20
Sonuç ve Öneriler	0%	20
Tezin Geneli	14%	25

Not: Yedi kelimeye kadar benzerlikler ile Başlık, Kaynakça, İçindekiler, Teşekkür, Dizin ve Ekler kısımları tarama dışı bırakılabilir. Yukarıdaki azami benzerlik oranları yanında tek bir kaynaktan olan benzerlik oranlarının %5'den büyük olmaması gerekir.

Beyan edilen bilgilerin doğru olduğunu, aksi halde doğacak hukuki sorumlulukları kabul ve beyan ederiz.

Tez Yazarı (Öğrenci)	Tez Danışmanı
Alaa Mohamed Jamel Abdulqader MAWLAWİ	Dr. Öğr. Üyesi Gülşah TÜMÜKLÜ ÖZYER
16.12.2024	16.12.2024
İmza: Aslı Islak İmzalıdır	İmza: Aslı Islak İmzalıdır

* Tez ile ilgili YÖKTEZ'de yayınlamasına ilişkin bir engelleme var ise aşağıdaki alanı doldurunuz.

Tezle ilgili patent başvurusu yapılması / patent alma sürecinin devam etmesi sebebiyle Enstitü Yönetim Kurulunun/.../.... tarih ve sayılı kararı ile teze erişim 2 (iki) yıl süreyle engellenmiştir.

Enstitü Yönetim Kurulunun/.../.... tarih ve sayılı kararı ile teze erişim 6 (altı) ay süreyle engellenmiştir.

TEŞEKKÜR

Danışmanım Dr. Öğr. Üyesi Gülşah TUMÜKLÜ ÖZYER'e, tez sürecim boyunca gösterdiği samimi destek, rehberlik ve yol gösterici yaklaşımı için en içten teşekkürlerimi sunarım. Akademik bilgi ve tecrübesiyle çalışmama yön verirken, karşılaştığım her sorunda çözüm odaklı yaklaşımı ve sabrı ile her zaman yanımda oldu.

Prof. Dr. Köksal ERENTÜRK'e, tezimin akademik derinliğini artırmaya yönelik değerli katkıları ve kıymetli önerileri için teşekkür ederim. Bilimsel bakış açısını ve geniş tecrübesini paylaşarak, çalışmamın ilerlemesinde önemli bir rol üstlendi. Kendisinin sağladığı destek, bu süreci daha verimli kılmamda etkili olmuştur.

Dr. Öğr. Üyesi Işıl Karabey Aksakallı'ya, tez çalışmamın değerlendirilmesi sırasında sunduğu değerli görüşler ve yapıcı eleştiriler için minnettarım. Kendisi, tezimin detaylarında dikkatli bir inceleme ve yönlendirme sağlayarak, çalışmamın titizlikle tamamlanmasına büyük katkıda bulunmuştur.

Doç. Dr. Barış ÖZYER'e, bilimsel perspektifimi geliştirmemi sağlayan değerli katkıları ve önerileri için teşekkür ederim. Tezimin içeriğine yönelik yaptığı değerlendirmeler, çalışmamın kalite ve bütünlüğünü artırmıştır.

Alaa Mohamed Jamel Abdulgader MAWLAVİ

ÖZET

YÜKSEK LİSANS TEZİ

YOLO, FASTER R-CNN ve SSD MOBİLE NET KULLANARAK MARKET RAF ÜRÜNLERİNİN ALGILANMASI VE TANIMLANMASI

Alaa Mohamed Jamel Abdulqader MAWLAWİ

Danışman: Dr. Öğr. Üyesi Gülşah TÜRÜKLÜ ÖZYER

Amaç: Bu çalışmanın amacı, market raflarındaki ürünlerin otomatik olarak tespit edilmesi ve tanımlanması için YOLOv11, YOLOv8, Faster R-CNN ve SSD MobileNet gibi derin öğrenme tabanlı modelleri kullanarak performans karşılaştırması yapmaktır. Bu teknolojinin, süpermarket envanter yönetimi ve müşteri deneyimi üzerindeki etkileri analiz edilmiştir.

Yöntem: Çalışmada, Erzurum ilindeki marketlerden elde edilen raf görüntülerinden oluşan bir veri seti kullanılmıştır. Görseller, Roboflow 3.0 aracılığıyla detaylı olarak etiketlenmiş ve YOLOv11, YOLOv8, Faster R-CNN ve SSD MobileNet modelleri eğitilmiştir. Model performansları mAP (Ortalama Doğruluk), kayıp eğrileri ve F1-güven, doğruluk-duyarlılık gibi metriklerle değerlendirilmiştir.

Bulgular: Deneysel sonuçlara göre YOLOv11 modeli en yüksek doğruluk oranını sağlarken, YOLOv8 modeli hız ve performans açısından dengeli bir çözüm sunmuştur. SSD MobileNet, düşük hesaplama maliyetiyle donanım kısıtlı ortamlarda başarılı performans sergilemiştir. Faster R-CNN ise özellikle küçük nesnelerin tespitinde daha yüksek başarı göstermiştir.

Sonuçlar: YOLOv11, doğruluk ve hız açısından üstün performans göstererek market raflarındaki ürünlerin tespitinde en uygun çözüm olarak belirlenmiştir. YOLOv8 modeli ise daha hızlı bir yapı sunarak, gerçek zamanlı uygulamalar için avantaj sağlamıştır. Bu bulgular, otomatik envanter yönetimi, raf boşluklarının tespiti ve müşteri deneyiminin iyileştirilmesi için YOLO tabanlı modellerin etkin kullanımını önermektedir.

Anahtar Kelimeler: YOLO, Faster RCNN, SSD MobileNet, Derin Öğrenme, Nesne Tespiti, mAP

Aralık 2024, 129 Sayfa

ABSTRACT

MASTER'S THESIS

MARKET SHELF PRODUCTS DETECTION AND RECOGNITION USING YOLO, FASTER R-CNN, AND SSD MOBILE NET

Alaa Mohamed Jamel Abdulqader MAWLAWI

Supervisor: Assist. Prof. Dr. Gülşah TÜRÜKÜ ÖZYER

Purpose: The aim of this study is to conduct a performance comparison of deep learning-based models, including YOLOv11, YOLOv8, Faster R-CNN, and SSD MobileNet, for the automated detection and identification of products on supermarket shelves. The study analyzes the impact of this technology on supermarket inventory management and customer experience.

Method: The study utilized a dataset consisting of shelf images obtained from markets in Erzurum. The images were meticulously labeled using the Roboflow 3.0 tool, and the models YOLOv11, YOLOv8, Faster R-CNN, and SSD MobileNet were trained. Model performances were evaluated using metrics such as mAP (Mean Average Precision), loss curves, and F1-confidence, precision-recall curves.

Findings: Experimental results show that the YOLOv11 model achieved the highest accuracy, while the YOLOv8 model provided a balanced solution in terms of speed and performance. SSD MobileNet demonstrated successful performance in hardware-constrained environments due to its low computational cost. Faster R-CNN, on the other hand, excelled in detecting small objects.

Conclusions: The YOLOv11 model exhibited superior performance in terms of accuracy and speed, making it the most suitable solution for detecting products on supermarket shelves. The YOLOv8 model offered a faster structure, making it advantageous for real-time applications. These findings suggest the effective use of YOLO-based models for automatic inventory management, shelf gap detection, and improving customer experience.

Keywords: YOLOv, Faster RCNN, SSD MobileNet, Deep Learning, Object Detection, mAP

December 2024, 129 pages

İÇİNDEKİLER

KABUL VE ONAY TUTANAĞI.....	i
ETİK BİLDİRİM VE İNTİHAL BEYAN FORMU	ii
TEŞEKKÜR	iii
ÖZET	iv
ABSTRACT	v
TABLolar DİZİNİ.....	ix
ŞEKİLLER DİZİNİ	x
KISALTMALAR VE SİMGELER DİZİNİ	xii
GİRİŞ.....	1
KURAMSAL TEMELLER.....	4
Görüntü İşleme.....	4
Derin Öğrenme Nedir?	5
Tekrarlayan Sinir Ağları (RNNs).....	6
Evrışimsel Sinir Ağları (CNN).....	7
Transfer Öğrenmesi.....	8
Market Ürünlerinin Görsel Tespiti	8
Özet	13
MATERYAL YÖNTEM.....	14
SSD Mobilenet v2	14
YOLOV-5	15
Faster R-CNN	15
YOLO V-8	16
YOLO V-11	17
Karışıklık Matrisi	17
Ortalama Doğruluk Değeri (mAP).....	19
Veri Seti Oluşturma ve Veri Etiketleme	19
ARAŞTIRMA BULGULARI	24
Etiket Dağılımı ve Sınır Kutusu Görselleştirme.....	24
YOLO V-5 Deneysel Sonuçlar	26
YOLO V-5 için karmaşıklık matrisi	26
YOLO V-5 için F-1 F1-güven eğrisi.....	27

YOLO V-5 için doğruluk-güven eğrisi, doğruluk-duyarlılık eğrisi ve duyarlılık-güven eğrisi grafikleri	28
YOLO V-5 için sonuç değerlendirmesi	30
YOLO V-8 Deneysel Sonuçlar	31
YOLO V-8 için karmaşıklık matrisi	31
YOLO V-8 için F1 F1-güven eğrisi.....	33
YOLO V-8 için doğruluk-güven eğrisi, doğruluk-duyarlılık eğrisi ve duyarlılık-güven eğrisi grafikleri	34
YOLO V-8 için kayıp grafikleri.....	35
YOLO V-8 için sonuç	36
YOLO V-11 Çıktıları	36
YOLO V-11 için karmaşıklık matrisi	36
YOLO V-11 için F1 F1-güven eğrisi.....	37
YOLO V-11 için doğruluk-güven eğrisi, doğruluk-duyarlılık eğrisi ve duyarlılık-güven eğrisi grafikleri	38
YOLO V-11 için kayıp grafikleri.....	40
YOLO V-11 için sonuç	40
YOLO V-5, YOLO V-8 ve YOLO V-11' in Karşılaştırılması	40
SSD MobilenetV2 Nesne Tanıma Deneysel Sonuçlar	42
Faster R-CNN Deneysel Sonuçları	46
Yolo V-11 Toplu İşlem Boyutu (Batch Size) 16-32-80 Kıyaslama.....	48
F1-skor eğrileri.....	48
Sınır kutusu karşılaştırması	49
Farklı toplu işlem boyutu değerleri için kesinlik-güven eğrilerinin karşılaştırılması	50
Farklı toplu işlem boyutu değerlerine göre kesinlik-duyarlılık eğrileri	51
Farklı toplu işlem boyutu değerlerine göre duyarlılık-güven eğrileri.....	51
Farklı toplu işlem boyutlarının model performansı üzerindeki etkilerinin karşılaştırılması	52
Yolo V-11 Epok 30-40-50 ve 60 Karşılaştırması	55
Epok 30-40-50 ve 60 için F1 skor eğrisi eğrilerinin karşılaştırması	55
Epok 30-40-50 ve 60 için kesinlik güven eğrilerinin karşılaştırılması	56
Epok 30-40-50 ve 60 için kesinlik-duyarlılık eğrilerinin karşılaştırması	58
Epok 30-40-50 ve 60 için duyarlılık-güven eğrilerinin karşılaştırması	59
Epok 30-40-50 ve 60 için model performansı üzerindeki etkilerinin karşılaştırılması... ..	60
Epok boyutu karşılaştırma sonucu	62

Yolo V-11 Görüntü Boyutu 640-1280 Karşılaştırması.....	63
Görüntü boyutu 640-1280 için F1 skor eğrisi eğrilerinin karşılaştırması	63
Görüntü boyutu 640-1280 için kesinlik-güven eğrilerinin karşılaştırılması	64
Görüntü boyutu 640-1280 için kesinlik-duyarlılık eğrilerinin karşılaştırması	64
Görüntü boyutu 640-1280 için duyarlılık-güven eğrilerinin karşılaştırması	65
Görüntü boyutu 640-1280 için model performansı üzerindeki etkilerinin karşılaştırılması	66
Görüntü boyutu karşılaştırma sonucu	67
Örnek Mobil Tabanlı Uygulama	69
Bu Çalışma Sonuçları İle Shen (2020) ve Torenzo (2023) Çalışma Sonuçlarının Karşılaştırılması	70
TARTIŞMA VE SONUÇ.....	76
Literatür Değerlendirme Sonuçları	76
Model Performansının Değerlendirilmesi.....	77
Eğitim Süresi, Kaynak Kullanımı ve Enerji Verimliliği.....	77
Uygulama Odaklı Katkılar ve Gelecek Çalışmalar	78
Sonuç.....	78
KAYNAKÇA	79
EKLER	82
EK-1. Veri Seti Örnek Görseller ve Sınır Kutuları	82
EK-2. Faster-Rcnn-SSDmobilenetv2 Kodları.....	86
EK-3. YOLO Kodları.....	107
ÖZGEÇMİŞ.....	115

ÇİZELGELER DİZİNİ

Çizelge 1. Sınıflandırma, Etiket Sayısı ve Fotoğraf Sayısı	21
Çizelge 2. YOLO V-5, YOLO V-8 ve YOLO V-11' in Metrik Karşılaştırması	41
Çizelge 3. SSD MobilenetV2 mAP Metrikleri.....	44
Çizelge 4. SSD MobileNetV2 ve Faster R-CNN Karşılaştırma Tablosi.....	48
Çizelge 5. Toplu İşlem Boyutu Hiper Parametre Çizelgesi	54
Çizelge 6. Toplu İşlem Boyutu Karşılaştırma Çizelgesi	54
Çizelge 7. Epok Boyut Hiper Parametre Çizelgesi	62
Çizelge 8. Epok Boyut Metrik Karşılaştırma Çizelgesi	62
Çizelge 9. Görüntü Boyutu Hiper Parametre Çizelgesi	68
Çizelge 10. Görüntü Boyutu Metrik Karşılaştırma Çizelgesi	68
Çizelge 11. Model Karşılaştırma Çizelgesi	74

ŞEKİLLER DİZİNİ

Şekil 1. Basit ve derin öğrenme sinir ağlarının katmanlı çalışma modeli.....	5
Şekil 2. Tekrarlayan sinir ağları modeli	6
Şekil 3. Evrimsel sinir ağı modeli	7
Şekil 4. CNN uygulama örneği	8
Şekil 5. SSD MobileNet.....	14
Şekil 6. YOLOv5 mimarisi	15
Şekil 7. Etiket programında görsellerin sınıflandırılması	21
Şekil 8. Tez çalışmasının akış diyagramı	23
Şekil 9. Etiket dağılımı ve sınır kutusu görselleştirmesi	26
Şekil 10. YOLO V-5 için karmaşıklık matrisi	27
Şekil 11. YOLO V-5 için normalize karmaşıklık matrisi	27
Şekil 12. YOLO V-5 için F1 skor eğrisi	28
Şekil 13. YOLO V-5 için doğruluk-güven eğrisi	29
Şekil 14. YOLO V-5 için duyarlılık-güven eğrisi	29
Şekil 15. YOLO V-5 için doğruluk-duyarlılık eğrisi	30
Şekil 16. YOLO V-5 için kayıp grafikleri.....	30
Şekil 17. YOLO V-8 için karmaşıklık matrisi	32
Şekil 18. YOLO V-8 için normalize karmaşıklık matrisi	33
Şekil 19. YOLO V-8 için F-1 skor eğrisi	33
Şekil 20. YOLO V-8 için doğruluk-güven eğrisi	34
Şekil 21. YOLO V-8 için duyarlılık-güven eğrisi	34
Şekil 22. YOLO V-8 için kesinlik-duyarlılık eğrisi	35
Şekil 23. YOLO V-8 için kayıp grafikleri.....	35
Şekil 24. YOLO V-11 için karmaşıklık matrisi	37
Şekil 25. YOLO V-11 için normalize karmaşıklık matrisi	37
Şekil 26. YOLO V-11 için F-1 skor eğrisi	38
Şekil 27. YOLO V-11 için kesinlik-güven eğrisi.....	38
Şekil 28. YOLO V-11 için duyarlılık-güven eğrisi.....	39
Şekil 29. YOLO V-11 için kesinlik-duyarlılık eğrisi	39
Şekil 30. YOLO V-11 için kayıp grafikleri.....	40
Şekil 31. SSD MobilenetV2 nesne tanıma performans sonuçları	43

Şekil 32. SSD MobilenetV2 örnek test sonuçları.....	46
Şekil 33. Faster R-CNN eğitim süreci kayıp fonksiyonları	47
Şekil 34. Yolo V-11 toplu işlem boyutu 16-32-80.....	48
Şekil 35. Sınır kutusu boyut-16, boyut -32 ve boyut -80	49
Şekil 36. kesinlik-güven eğrileri boyut-16, boyut -32 ve boyut -80	50
Şekil 37. kesinlik-duyarlılık eğrileri boyut -16, boyut -32 ve boyut -80.....	51
Şekil 38. duyarlılık-güven eğrileri boyut -16, boyut -32 ve boyut -80	52
Şekil 39. Model performans grafikleri boyut -16, boyut -32 ve boyut -80	53
Şekil 40. Epok 30-40-50 ve 60 için F1 skor eğrisi.....	56
Şekil 41. Epok 30-40-50 ve 60 için kesinlik-güven eğrileri	57
Şekil 42. Epok 30-40-50 ve 60 için kesinlik-duyarlılık eğrileri.....	58
Şekil 43. Epok 30-40-50 ve 60 için duyarlılık-güven eğrileri.....	60
Şekil 44. Epok 30-40-50 ve 60 için model performans eğrileri	61
Şekil 45. Görüntü Boyutu 640-1280 için F1 skor eğrisi	63
Şekil 46. Görüntü Boyutu 640-1280 için kesinlik-güven eğrileri.....	64
Şekil 47. Görüntü boyutu 640-1280 için kesinlik-duyarlılık eğrileri.....	65
Şekil 48. Görüntü Boyutu 640-1280 için duyarlılık-güven eğrileri	65
Şekil 49. Görüntü boyutu 640-1280 için model performans eğrileri	66
Şekil 50. Örnek mobil tabanlı uygulama ekran görüntüsü.....	69
Şekil 51. Faster-RCNN'e ilişkin sonuçlar	72
Şekil 52. YOLOv3'e ilişkin sonuçlar.....	72
Şekil 53. Faster-RCNN/YOLOv3 performansı.....	73
Şekil 54. Değiştirilmiş bir yöntemle elde edilen ürün tespit sonuçları	73
Şekil 55. Değiştirilmiş bir yöntemle elde edilen ürün tespit sonuçları	73

KISALTMALAR VE SİMGELER DİZİNİ

ACT	Ortalama İşlem Süresi
AP	Ortalama Hassasiyet
ASM	Aktif Şekil Modeli
CNN	Evrişimli Sinir Ağı
DA	Veri Artırma
DCNN	Derin Evrişimli Sinir Ağları
DL	Derin Öğrenme
F. R-CNN	Hızlı Menzil Tabanlı Evrişimli Sinir Ağı
FN	Yanlış Negatif
FP	Yanlış Pozitif
GPU	Grafik İşlem Birimi
LSTM	Uzun Kısa Süreli Bellek
mAP	Ortalama Hassasiyet (mAP)
ms	Milisaniye
OCR	Optik Karakter Tanıma
PCA	Temel Bileşen Analizi
RCNN	Menzil Tabanlı Evrişimli Sinir Ağları
RNN	Tekrarlayan Sinir Ağı
SSD	Tek Aşamalı Nesne Algılayıcı
TP	Doğru Pozitif
YOLO	You Only Look Once-Tek Bakışta Algılama

GİRİŞ

Görsel ürün tanıma, bir süpermarket ortamındaki bireysel ürünleri tespit etmeyi ve sınıflandırmayı amaçlayan güncel bir bilgisayarlı görme problemidir. Görsel ürün tanımlamanın potansiyel uygulamaları arasında otomatik ödeme sistemleri, gerçek zamanlı envanter yönetimi, raf diyagram uyumluluğu veya görme engelliler için yardımcı teknoloji yer alır. Son yıllarda, görsel ürün tanıma probleminin çözümünde derin öğrenme yaklaşımları yüksek başarı sağlamakta ve bu alanda çalışmalar yürütülmektedir (Shen 2020; Toranzo *et al.* 2023). Görsel ürün tanıma probleminde, derin öğrenme yaklaşımlarının uygulanmasında geniş ürün katalogları için veri toplanması ve etiketlenmesi önemli bir zorluk içermektedir. (Schmidhuber 2015). Bu zorluğun çözümüne yönelik olarak sentetik veriler kullanarak eğitim veri setlerinin otomatik üretilmesi yaklaşımı getirilmiştir (Schmidhuber 2015). Ancak, gerçek dünya verilerinin sentetik olarak oluşturulması ve bu verilerin hedef alanla uyumlu hale getirilmesi hâlâ önemli bir araştırma problemidir getirilmiştir (Zhu *et al.* 2019).

Market raflarında yer alan ürünlerin görsel olarak tanınmasında önemli problemlerden biri de ürün alt kategorilerinin görsel olarak bir birine çok benzemesidir. Örneğin bir içecek markasının aynı rafta hem farklı türleri (şekerli-şekersiz gibi) hem de farklı boyutları (1 litre-0.5 litre gibi) bulunmakta ve bu ürünler görsel olarak bir birine benzemektedir. Mevcut literatürde, genellikle farklı kategorilerin etiketlenmesi üzerinden model geliştirme çalışmaları yürütülse de, aynı kategorideki alt ürünlerin ayırımına yönelik spesifik veri setlerinin eksikliği dikkat çekmektedir (Khan *et al.* 2023).

Bu tezde kapsamında öncelikle, ürün alt kategori çeşitliliğini ele almak amacıyla, market raflarında bulunan ve görsel olarak büyük benzerlikler gösteren ürün alt kategorilerini içeren özel bir veri seti geliştirilmiştir. Bu veri seti, ürünlerin tür, boyut, ambalaj farkları gibi alt düzey farklılıklarını doğru bir şekilde tanıyabilecek derin öğrenme tabanlı modellerin eğitimi ve değerlendirilmesi için tasarlanmıştır. Ayrıca bu tez çalışmasıyla, YOLO, Faster R-CNN ve SSD MobileNet gibi nesne algılama modelleri kullanılarak süpermarket raflarındaki ürünlerin tespiti ve tanınması ele alınmıştır. Çalışmanın amacı, çeşitli modellerin performansını karşılaştırarak envanter yönetimi ve müşteri memnuniyeti gibi alanlarda uygulanabilir çözümler sunmaktır. Modellerin başarımları, ortalama doğruluk (mAP), kesinlik (precision), duyarlılık (recall), F1 skoru gibi metrikler kullanılarak değerlendirilmiştir.

Bu tez kapsamında yapılan analizlerin ve geliştirilen veri setinin potansiyel kullanım alanları ve faydaları şu şekilde özetlenebilir:

- Otomatik Envanter Yönetimi: Hızlı ve doğru ürün tanıma ile stok takibi ve raf optimizasyonu sağlanabilir.
- İş Gücü Verimliliği: Manuel envanter kontrolü yerine çalışanlar müşteri hizmetleri gibi değerli görevlere odaklanabilir.
- Müşteri Memnuniyeti: Güncel stok bilgileriyle ürün bulunabilirliği artar, alışveriş süreçleri hızlanır.

Bu çalışmada ayrıca, farklı modellerin güçlü ve zayıf yönleri karşılaştırılarak perakende sektöründe en uygun yapay zeka tabanlı çözümlerin seçimi üzerine önerilerde bulunulmuştur.

Bu tez çalışmasının katkıları aşağıda sunulmuştur:

1. Veri Seti Oluşturulması: Yerel marketlerden toplanan görüntüler ile yeni ve kapsamlı bir veri seti oluşturulmuştur. Veri seti aynı ürünün farklı türlerini içeren bir ürün yelpazesine sahiptir. Aynı rafta bulunan benzer ürünlerinin tespitinin analizi için uygun bir veri seti hazırlanmıştır.
2. Derin Öğrenme Modellerinin Karşılaştırılması: YOLOv11 ailesi, Faster R-CNN ve SSD MobileNet gibi modellerin ayrıntılı bir karşılaştırılması yapılmıştır. Yapılan deneysel çalışmalar, gerçek hayat uygulamaları için model seçiminde yol gösterici niteliktedir.
3. Performans Analizi: Yukarıda verilen modellerin, farklı donanım gereksinimleri ve farklı parametreler ile analiz edilmiştir. Büyük modellerin karmaşık problemler için daha uygun olduğu, hafif modellerin hızlı çözümler sunabildiği ile ilgili bulgular elde edilmiştir.
4. Perakende sektörü için uygulama önerileri: Elde edilen bulguların ışığında, süpermarket raflarındaki ürün tespitinin yanı sıra, bu teknolojilerin güvenlik, sağlık ve diğer alanlardaki potansiyel uygulamaları da tartışılmıştır. Tez kapsamında elde edilen bulgulardan mobil tabanlı bir uygulama geliştirilmiştir. Geliştirilen bu uygulama ile mobil ortamda market raflarından alınan görüntüler üzerinden nesne tespiti yapılabilmektedir.

Bu çalışmanın birinci bölümünde, tez konusu ve tez içerisinde yapılacak olan çalışmalardan bahsedilmiştir. İkinci bölümde ise kullanılacak olan modeller ve yapılan çalışma hakkında genel bilgilerin bulunduğu 'Kuramsal Temeller' kısmı yer almaktadır. Üçüncü bölümü, çalışmanın modeller ve hangi parametreler üzerine kurulu olduğu, kütüphanelerin

hazırlanışı ve sistemin hazırlanışı hakkında bilgilerin bulunduğu ‘Materyal ve Yöntem’ kısmını içermektedir. Dördüncü bölüm sistem ve model çıktılarının yer aldığı grafikleştirilmiş karşılaştırma sonuçlarının yer aldığı, modeller, ortalama doğruluk değeri ve toplam kayıp metrikleri üzerinden gerçekleştirilmiştir. Modellerin performansları; F1-güven eğrisi, doğruluk-güven eğrisi, doğruluk-duyarlılık eğrisi, duyarlılık-güven eğrisi ve başarı oranı gibi performans metrikleriyle ile nasıl karşılaştırıldığına dair çalışmanın sonuçlarının yer aldığı ‘Tartışma ve Sonuç’ kısmından oluşmaktadır.



KURAMSAL TEMELLER

Market ürünlerinin görsel olarak tanınması problemi, literatürde temel olarak bir bilgisayarlı görme problemi olarak ele alınmaktadır ve genellikle bir nesne tespiti sorunu olarak tanımlanmaktadır. Bu bölümde, öncelikle nesne tespiti problemiyle ilgili temel yaklaşımlar ele alınacak; ardından görüntü işleme, derin öğrenme ve nesne tespitinde yaygın olarak kullanılan Tekrarlayan Sinir Ağları (RNN), Evrimsel Sinir Ağları (CNN) ve transfer öğrenme gibi yöntemler hakkında genel bilgiler sunulacaktır. Son olarak, market ürünlerinin tanınmasıyla ilgili literatürde yer alan çalışmalar değerlendirilecektir.

Görüntü İşleme

Görüntü işleme, sayısal bir görüntünün analiz edilmesi, iyileştirilmesi, dönüştürülmesi ve yorumlanması süreçlerini içeren bir yaklaşımdır. Görüntü işleme, makine öğrenmesi, bilgisayarlı görme ve robotik alanında tanımlı görevler için kullanılabilen bir ön işleme aracı olarak da tanımlanabilir. Bu görevler, bölütleme, öznitelik çıkarma, optimizasyon, desen tanıma, iç boyama, sıkıştırma ve nesne izleme gibi çok çeşitli görevleri içerir (Muntarina *et al.* 2022). Görüntü işleme teknolojisi gürültüyü azaltma ve gereksiz bilgileri kaldırma işlemi, sonraki işlemler için yüksek kaliteli görüntüler sağlar. Bu süreç tipik olarak görüntünün bölütlenmesini, dönüştürülmesini ve iyileştirilmesini içermektedir. Son yıllarda derin öğrenme mimarilerinin gelişmesi ile görüntü işleme öznitelik çıkarımında daha az kullanılmaktadır (Muntarina *et al.* 2022) Görüntü işleme çoğunlukla görüntü iyileştirmede, gürültüyü yok etmekte, kontrast düzeltmede, keskinleştirmede kullanılmaktadır (Muntarina *et al.* 2022).

Genel bir görüntü işleme probleminin adımları şu şekildedir (Anonymous 2024):

- Görüntü iyileştirme: Gürültü azaltma, yumuşatma, keskinleştirme, kontrast düzeltme işlemleri veriye uygun olarak uygulanır.
- Öznitelik çıkarma: Görüntü verilerinin analizi ve işlenmesi, görüntüdeki değişmez özellikleri tanımlamak veya ayırt edici özellikleri çıkarmak için yapılır.
- Öznitelik sınıflandırması: Belirli görüntü özellikleri, bir özellik vektörüne veya uzaya eşlenir ve tanıma sonuçlarının doğru olması için belirli karar kuralları uygulanır.

- Tanıma sonuçları: Perakende ürünlerin kategorisini tahmin etmek için önceden eğitilmiş bir sınıflandırıcı kullanılır. Ürün keşfinin özü, doğru özellikleri çıkarabilmektir.

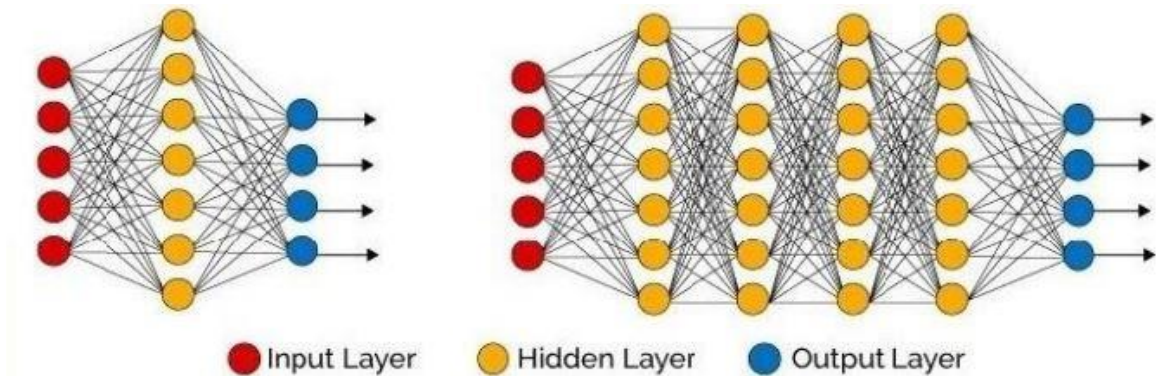
Derin Öğrenme Nedir?

Derin öğrenme, yapay zeka alanının bir alt dalıdır ve büyük veri setleri kullanarak problem çözümünde kullanılan bir makine öğrenmesi yöntemidir. Derin öğrenme yaklaşımları, yapay sinir ağlarına dayanmaktadır ve insan beynindeki sinir hücrelerinin çalışma prensibine dayanmaktadır. Geleneksel yapay sinir ağlarından farkı, katman sayısının fazla olması, etiketli ve etiketsiz büyük veriler üzerinde çalışabilmesi ve otomatik öznelik çıkarabilmesidir (Hinton 2006).

Derin öğrenmenin bazı temel özellikleri şunlardır:

1. **Çok katmanlı yapı:** Derin öğrenme modelleri, verilerin her katmanında daha soyut özellikleri öğrenen bir yapıya sahiptir. Bu, ağın derinliğine göre değişir ve genellikle birden fazla gizli katmandan oluşur. Şekil 1.'de basit ve derin öğrenme sinir ağlarının katmanlı modeli gösterilmektedir.
2. **Veri bağımlılığı:** Derin öğrenme, büyük miktarda veriye ihtiyaç duyar, çünkü daha fazla veri ile daha iyi sonuçlar elde edilir.
3. **Otomatik özellik çıkarımı:** Özellik mühendisliğine manuel müdahale gerektirmeden, derin öğrenme modelleri verilerden otomatik olarak anlamlı özellikler çıkarabilir.

Derin öğrenme, özellikle **görüntü işleme**, **ses tanıma**, **doğal dil işleme (NLP)** ve **otonom araçlar** gibi birçok uygulama alanı bulunmaktadır.



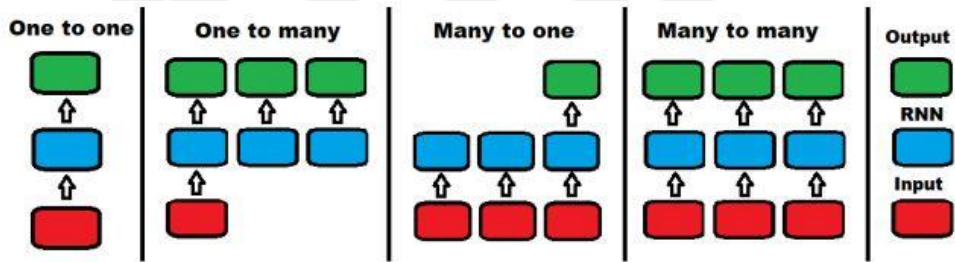
Şekil 1. Basit ve derin öğrenme sinir ağlarının katmanlı çalışma modeli (Saleem 2022)

Tekrarlayan Sinir Ağları (RNNs)

Tekrarlayan sinir ağı (RNN), düğümler arasındaki zamansal bağlantı dizisi boyunca yönlendirilmiş bir grafik oluşturan bir tür yapay sinir ağıdır. LSTM (Uzun Kısa Süreli Bellek), dizi tahmini için RNN modellerinin en ünlü yaklaşımı olarak kabul edilebilir (Hochreiter ve Schmidhuber, 1997).

Aşağıdaki durumlar da tercih edilebilecek sinir ağı modellerindendir. Şekil 2'de tekrarlayan sinir ağları model şeması verilmiştir.

- Birden Bire: Tek bir çıktı ile eşleşen bir girdi (Görüntü sınıflandırma)
- Birden Çoğa: Bir dizi çıktı ile eşleşen bir girdi (Görüntü altyazısı- metin görüntüsünden sözcükler ayırt etme)
- Çoktan Bire: Bir dizi girdiden tek bir çıktı (Duyarlılık Analizi- birden çok kelimedenden ikili çıktı)
- Çoktan çoğa: Bir dizi girdi, bir dizi çıktı (Video Sınıflandırması- videoyu karelere bölme ve her kareyi ayrı etiketleme)



Şekil 2. Tekrarlayan sinir ağları modeli (Anonymous 2024)

RNN'lerin temel çalışma prensibi, girdi verilerini sırayla işlerler. Her adımda da önceki adımın çıktısını (gizli durumu) kullanırlar. Bu, ardışık verilerin bağlantısını korumalarına yaramaktadır. Önceki bilgilere dayalı tahminler yapmalarına olanak tanımaktadır.

Mimari olarak bir RNN, bir giriş katmanı, bir gizli (veya durum) katman ve bir çıkış katmanından oluşur. Her adımda, gizli katmanın durumu güncellenir ve girdiyle birleştirilir, böylece her adımın önceki bilgilere dayanarak tahminlerde bulunmasına olanak sağlanır.

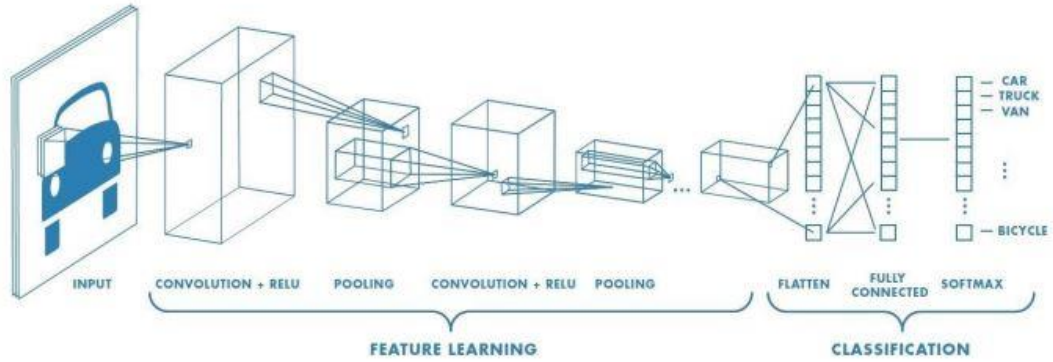
RNN'ler, giriş verileri ile hedef çıktılar arasındaki ilişkileri öğrenmek üzere eğitilir. Bu eğitim süreci, RNN'ye verilerdeki kalıpları anlama yeteneği kazandıran geri yayılım ve diferansiyel hesaplamaları içerir.

Evrişimsel Sinir Ağları (CNN)

CNN, bir görüntüdeki farklı nesnelere ayırt etmeye yardımcı olan bir derin öğrenme algoritmasıdır. Öncü bir sinir ağı olan CNN, hayvanların görme merkezlerinden ilham aldı.

Evrişimsel sinir ağları (CNN'ler veya ConvNet'ler), çitlerle çevrili sinir ağlarına sahip en yaygın kullanılan derin öğrenme modelleridir. 2012 yılında Krizhevsky, büyük ölçekli görüntü verilerini tahmin etmede CNN modellerinin önemini vurguladı (Krizhevsky *et al.* 2012). CNN, giriş verilerinden (2D) elde edilen özellikleri (çıkı) geliştirir. Katmanlar özellikle 2 boyutlu görüntülerin işlenmesini kolaylaştırır (Doğan ve Türkoğlu 2019).

CNN, manuel özellik çıkarma adımlarına olan ihtiyacı ortadan kaldırır ve böylece görüntüleri sınıflandırmak için manuel özellik belirleme ihtiyacını ortadan kaldırır. CNN bu özellikleri doğrudan görüntüden elde ederek daha doğru nesne sınıflandırma sonuçları elde eder (Şekil 3.). Özellik çıkarmanın manuel adımı yerine, daha fazla gizli katman kullanılır ve ne kadar çok katman kullanılırsa, elde edilen sonuçlar o kadar çeşitli ve ayrıntılı olur, bu da işlem süresini artırır. Tüm bu özellikler, söz konusu modeli bu çalışmada tercih edilen sinir ağı modeli haline getirmektedir.



Şekil 3. Evrişimsel sinir ağı modeli (Elgamily *et al.* 2024)

CNN, standart çok katmanlı sinir ağları gibi, bir veya daha fazla evrişimsel katmandan, bir alt örnekleme katmanından ve bir veya daha fazla bağlantı katmanından oluşur. CNN'in avantajı, aynı sayıda gizli birime sahip, tamamen bağlı bir ağa göre daha az eğitim ve parametre gerektirmesidir.

CNN uygulamaları esas olarak görüntü işlemede kullanılmaktadır. Ancak biyotıp, konuşma tanıma gibi doğal dil işleme süreçlerinin yanı sıra birçok farklı alanda incelenmektedir. Özellikle görüntü işleme alanında son teknoloji başarılarına imza atılmıştır. 2015 yılında çok katmanlı CNN'ler, baş aşağı yüzler de dahil olmak üzere geniş açılardan yüzleri tespit etme yeteneğini göstermiştir. Ağ, çeşitli açı ve yönlerde yüze sahip 200.000

görüntü ve yüzüz ek 20 milyon görüntü içeren bir veritabanı üzerinde eğitilmiştir. CNN algoritmaları farmasötik arařtırmalarda da kullanılmaktadır. Örnek bir uygulama Şekil 2.4'de gösterilmektedir. Kısaca özetlemek gerekirse CNN uygulamaları genel olarak řu řekilde çalıřır.

- Yüz Tanımlama
- Belgeleri Analiz Etme
- İklimi Tanımlama
- Gri Alanlar
- Reklamlar
- Tarihi Eserlerin Ayrıştırılması



Şekil 4. CNN uygulama örneđi (Anonymous 2024)

Transfer Öğrenmesi

Transfer öğrenmesi için iki ana başlık vardır. Bunlardan biri Gelişmiş Model Yaklaşımıdır. Diğeri ise ön eğitilmiş model yaklaşımı.

Önceden eğitilmiş model yaklaşımı, son yıllarda yapay zeka ve makine öğrenmesi alanlarında çığır açan bir yaklaşımdır. Bu yaklaşım, büyük veri kümeleri üzerinde önceden eğitilmiş derin öğrenme modelleri üzerine kuruludur ve daha sonra bu modelleri daha özel görevler için optimize eder. Bu çalışmada da veri seti olarak önceden eğitilmiş kaynak model seçilmiştir.

Bu yaklaşımın temeli öncelikle veriye dayalı öğrenmedir. Önceden eğitilmiş modeller genellikle çok büyük veri kümeleri üzerinde eğitilmiş derin öğrenme modelleridir. Bu modeller, ortak veri temsillerini öğrenerek kapsamlı öğrenme bilgisi kazanır.

Market Ürünlerinin Görsel Tespiti

Bu makale Wei *et al.* (2020), perakende ürün tanıma alanındaki derin öğrenme uygulamalarını ve karşılaşılan zorlukları ele alarak mevcut teknikleri ve gelecekteki yönelimleri kapsamlı bir şekilde incelemektedir. Görüntü tabanlı ürün tanıma, otomatik ödeme,

stok takibi ve görme engellilere yardım gibi çeşitli alanlarda büyük uygulama potansiyeline sahip, bilgisayarla görme alanında zorlayıcı bir konudur. Çalışma, derin öğrenmenin bu alandaki başarısını ve geleneksel yöntemlere kıyasla daha yüksek doğruluk sağlayan özellik öğrenme yeteneğini vurgulamaktadır. Makale, marketlerde ürün tanımının karşılaştığı ölçeklendirme, veri kısıtlaması, alt sınıf (intraclass) çeşitliliği ve sistem esnekliği gibi dört temel zorluğu analiz etmektedir. Bu zorluklarla başa çıkmak için kullanılan derin öğrenme teknikleri, CNN tabanlı özellik çıkarımı, veri artırma yöntemleri, ince detay sınıflandırması ve bir örnekle öğrenme (one-shot learning) yaklaşımlarını içermektedir. Ayrıca, çalışma bu alan için kullanılabilecek halka açık veri setlerini detaylı olarak tanıtarak, perakende ürün tanıma araştırmalarına rehberlik etmektedir. Bu çalışmada CNN tabanlı yöntemlerin 3.000'den fazla sınıfa ait ürünleri tanıma konusunda %52.16 mAP ile etkili olduğunu, ancak perakende sektöründeki daha karmaşık durumlar için hala iyileştirme gerektiğini göstermiştir. Sonuç olarak, derin öğrenme yöntemlerinin perakende ürün tanıma için büyük bir potansiyel sunduğu, ancak hala veri sınırlamaları, esneklik sorunları ve sınıflandırma doğruluğunun geliştirilmesi gereken alanlar olduğu belirtilmektedir. Çalışma, yeni araştırma yönlerini tanımlayarak, bu alandaki ilerlemeler için kapsamlı bir kaynak sağlamayı amaçlamaktadır.

Şikic *et al.* (2024) Bu makale, süpermarket raflarındaki ürünlerin eksikliğini (out-of-stock, OOS) tespit etmek için derin öğrenme tabanlı yeni bir yöntem önermektedir. OOS, raflarda eksik ürünlerin bulunmadığı durumları ifade eder ve bu, müşteri memnuniyetsizliği ve satış kayıpları gibi ciddi sorunlara yol açabilir. Çalışma, iki aşamalı bir eğitim süreci ve yanlış tespitleri ortadan kaldırmak için özel bir son işlem tekniği içeren yenilikçi bir yaklaşımı tanıtmaktadır. Araştırmada kullanılan OOS veri seti, tamamen boş OOS durumlarını ve yeni tanımlanan ön yüz OOS sınıfını içermektedir. Yöntemde, orijinal veri setine dayalı olarak görüntü artırma teknikleri uygulanmış ve bu teknikler, tamamen boş ve ön yüz OOS sınıflarını tespit etmek için bir nesne algılama modelini eğitmekte kullanılmıştır. Model, YOLOv5, YOLOv7 ve EfficientDet gibi modern nesne algılama algoritmaları kullanılarak eğitilmiş ve iki aşamalı eğitim sürecinde orijinal veri setiyle ince ayar yapılmıştır. Ayrıca, son işlem tekniği, model tarafından tespit edilen OOS bölgelerinin boyut oranlarına göre doğrulanmasını sağlamış ve yanlış tespitlerin ortadan kaldırılmasına katkıda bulunmuştur. Deneysel sonuçlar, önerilen yöntemin, mevcut OOS tespit yöntemlerini %3.9 gibi bir farkla geride bıraktığını göstermiştir. Normal OOS sınıfı için ortalama doğruluk (AP) %86.3, ön yüz OOS sınıfı için ise %83.7 olarak elde edilmiştir. İki aşamalı eğitim süreci, modellerin genel performansını %5.3'e kadar artırmış ve son işlem tekniği, doğruluğu %2.2'ye kadar yükseltmiştir. Çalışma ayrıca, modelin düşük güçlü cihazlarda çalışabilmesi için optimize edilmiştir. Bu yöntem, gerçek zamanlı perakende ortamlarında raf eksikliği sorunlarının daha etkili bir şekilde yönetilmesine olanak sağlayarak,

satış kayıplarını ve müşteri memnuniyetsizliğini azaltma potansiyeline sahiptir. Gelecekte, derinlik algılama entegrasyonu ile sonuçların daha da geliştirilmesi planlanmaktadır.

Bu çalışma Biasio (2019), derin öğrenme ve görüntü işleme tekniklerini kullanarak perakende raf analitiği için otomatik bir sistem geliştirmeyi amaçlamaktadır. Araştırma, Padova Üniversitesi ve uluslararası bir müşteri şirketi arasında gerçekleştirilen iş birliği çerçevesinde, derin öğrenmeye dayalı yenilikçi bir yazılım çözümü geliştirilmesini içermektedir. Çalışmada kullanılan veri seti, müşteri iş süreçlerine uygun olacak şekilde özel olarak tasarlanmış ve ürünlerin raf görüntülerinde piksel düzeyinde tanımlanmasını sağlayacak şekilde elle etiketlenmiştir. Mask R-CNN algoritması, raf görüntülerindeki ürünlerin segmentasyonunu sağlamak ve her bir ürünün raf alanını hassas bir şekilde hesaplamak amacıyla tercih edilmiştir. Çalışma, Mask R-CNN'in mevcut iş süreçlerine entegrasyonu ile stratejik önem taşıyan metrikler ve KPI'lar sunan bir sistem modeli oluşturmuştur. Bu sistem, bulut tabanlı SAP platformu ile uyumlu bir şekilde tasarlanarak, müşteri markasının satış noktalarındaki etkisini analiz etmek için kullanılabilir hale getirilmiştir. Tez kapsamında, özellikle otomatik etiketleme tekniklerinin derin öğrenme tahminleri üzerindeki etkilerini araştırmak için deneysel testler gerçekleştirilmiştir. Sistemin performansı, hassasiyet ve doğruluk arasındaki optimal dengeyi sağlamak için ince ayar yapılmış ve geliştirilmiştir. Sonuçlar, önerilen sistemin müşteri markasının görünürlüğünü değerlendirme ve perakende ortamındaki etkisini artırma konusunda etkili bir araç olduğunu göstermektedir.

Santra *et al.* (2022) bu çalışmada, süpermarket raflarında sergilenen perakende ürün çeşitlerinin oldukça benzer görsellerinin sınıflandırılması için oldukça yeni bir çözüm (detaylı sınıflandırma) önermişlerdir. Önerilen şema, ürün görüntülerinden nesne düzeyinde ve parça düzeyinde ipuçlarını eşzamanlı olarak yakalamaktadır. Nesne düzeyinde ürün resim etiketleri, yenilikçi bir yeniden yapılandırma sınıflandırma ağı (RC-Net) tarafından yakalanmaktadır. Marka açıklaması olmadan parça düzeyinde modelleme, önemli noktalar etrafındaki ürün görselinin ayırt edici kısımlarını tanımlamaktadır. Karakteristik parçaların bu sürekli dizileri, ürünü benzersiz şekilde tanımlayan evrişimli bir LSTM kullanılarak kodlanmaktadır. Son olarak, parça düzeyinde ve nesne düzeyinde modeller, ürünü açıkça tanımlayan daha kaba ve daha ayrıntılı ürün açıklamalarını açıklamaktadır. Bu iki katmanlı mimari, süpermarket raflarındaki perakende ürün çeşitlerini tespit etmek için R-CNN'ye dahil edilmiştir. Dahili bir iş veri seti ve üç kıyaslama veri seti üzerinde kapsamlı deneyler gerçekleştirilmiştir. Önerilen şema neredeyse tüm değerlendirmelerde rakip yöntemlerden daha iyi performans göstermektedir.

Qiao *et al.* (2017) Süpermarketlerde ürün tanıma ile ilgili olarak bu çalışmada, süpermarket görsellerinde ve diğer doğal görsellerde nesne önerisi üretme sorununu incelemektedir. Görüntülerdeki nesnelerin boyutlarını tahmin etmenin, özellikle nesne boyutlarının genellikle dar bir aralığa düştüğü süpermarket görüntüleri için nesne önerileri oluşturmak için yararlı olduğu savunulmaktadır. Bu nedenle nesne teklifini oluşturmadan önce görüntüden nesne boyutlarını tahmin etmeyi önermişlerdir. Görüntü nesnelerinin boyutlarını tahmin etmek için önerilen yöntem "ScaleNet" adı verilmektedir. ScaleNet'in etkinliğini doğrulamak için üç süpermarket veri seti oluşturulmuştur. Bunlardan ikisi test amaçlı kullanılan gerçek veri setidir. Diğer ise eğitim amaçlı kullanılan sentetik bir veri setidir. Başka bir deyişle, bir boyut tahmin aşaması ekleyerek en uygun nesnelere önermek için önceki yöntemleri geliştirmiştir. Ortaya çıkan bu yöntemde, süpermarket veri kümesindeki önceki en iyi yöntemlerden çok daha iyi performans gösterdiği sonucuna ulaşılmıştır. Ayrıca nesne teklifi yaklaşımımızın diğer doğal görüntüler üzerinde çalıştığını ve MS COCO veri kümesindeki önceki en iyi nesne teklifi yöntemlerinden daha iyi performans gösterdiğini de raporlamışlardır. Süpermarket veri kümeleri, sanal süpermarketler ve diğer sentetik veri kümelerini oluşturmaya yönelik araçlar kamuya açık hale getirilecektir.

Kuo *et al.* (2021) çalışmalarında perakende ürün tespitinde nesne tespitini gerçekleştirmek için maskelenmiş R-CNN ve veri artırma (DA) kullanmışlardır. Bu yöntemin en temel amacı üst üste binen nesnelerin ya da benzer renkteki bitişik nesnelerin neden olduğu hataları en aza indirmektir. Deneysel sonuçlar, önerilen yöntemin atıştırmalık veri seti üzerinde %98,92 mAP gibi iyi bir tanıma oranına sahip olduğunu ve tanıma görevinde mevcut en iyi yöntemlerden daha başarılı olduğunu göstermektedir.

Hsia *et al.* (2022) bu çalışmada, kullanıcılar ve bilgisayarlar arasındaki etkileşimi yönetmek için bilgisayar teknolojisinin nasıl kullanıldığını araştıran insan-bilgisayar etkileşimine (HCI) odaklanıyor gibi görünmektedir. Nesne tanıma sistemlerini geliştirmeye ve iyileştirmeye yönelik derin öğrenme (DL) teknikleri de tartışılmaktadır. Özellikle daha hızlı menzile tabanlı evrişimli sinir ağı (Faster R-CNN) yerine maskeli R-CNN kullanılarak görüntüdeki önemsiz ayrıntılar derin öğrenme özelliği çıkarımına ve nesne tespitine engel olmuştur. Amaç bunu önlemektir. Bu yaklaşımın sonuçları, özellikle daha hızlı olan R-CNN ile karşılaştırıldığında, örtüşen ve benzer renklere sahip nesnelerin daha doğru ve verimli bir şekilde algılanmasını sağlamaktadır. Bu çalışmanın akıllı perakende mağazaları için üstün çözünürlük ve gerçek zamanlı işlem sağlayacağı da belirtilmektedir. Bu tür araştırmalar, nesne tanıma teknolojisinin geliştirilmesinde ve ticari uygulamalarda kullanılmasında önemli bir rol oynayabilme ihtimali yüksektir. Bu çalışmada, görüntülerdeki önemsiz bilgilerin özellik

çıkarmı ve DL doğruluğunu etkilemesi önlenmek için daha hızlı R-CNN yerine veri artırma (DA) tekniği ile maskelenmiş R-CNN kullanılmıştır. Deneysel sonuçlar, önerilen algoritmanın, üst üste binen ve benzer renkli nesnelere neden olduğu sorunları çözdüğünü göstermektedir.

Bu çalışmada Toranzo *et al.* (2023), süpermarketlerdeki envanter ve depo yönetimi süreçlerinde yaşanan zorluklara çözüm bulmak amacıyla YOLOv5 nesne tespit modeli tabanlı bir uygulama geliştirilmiştir. Süpermarketlerde genellikle ürünlerin durumu ve stok seviyeleri etkin bir şekilde izlenememektedir, bu da yanlış ürün dağıtımı, hatalı yerleştirme ve ürünlerin durumunun (bozuk ya da hasarlı) gözden kaçırılması gibi sorunlara yol açmaktadır. Çalışmanın amacı, bilgisayarla görü tabanlı bir uygulama geliştirerek şişelenmiş ve kutulanmış ürünlerin tespit edilmesi, sayılması ve durumlarının doğrulanmasıdır. Farklı nesne tespit modelleri, performans ve metrikler açısından karşılaştırılmış ve YOLOv5, üstün sonuçları nedeniyle seçilmiştir. Model, Peru süpermarketlerinde yaygın olarak bulunan şişelenmiş ve kutulanmış ürünlerden oluşan özel bir veri setiyle eğitilmiştir. Veri seti, hem süpermarket ortamında çekilen (in situ) hem de ideal koşullarda çekilen (in vitro) görüntülerden oluşmaktadır. Ayrıca, veri seti artırma teknikleri uygulanarak modelin farklı açılardan ürünleri tanıma kapasitesi artırılmıştır. Model, ürünlerin iyi veya kötü durumda olup olmadığını belirlemek için tasarlanmış ve bu durum farkını algılayabilmek için uygun hiperparametre ayarlamaları yapılmıştır. Uygulama, kullanıcılara bir görüntü yükleyerek veya cihaz kameralarıyla çekilen fotoğraflar aracılığıyla ürünlerin tespitini sağlar. Bu görüntülerde, ürünlerin isimleri, sayıları ve tespit edilen durumları belirtilir. Ayrıca, mağaza çalışanlarının günlük olarak ürün kayıtlarını tutabilmeleri için bir veri tabanı oluşturulmuştur. Çalışmada YOLOv5s modelinin SGD optimizasyon algoritmasıyla birlikte kullanıldığı ve mAP@0.5 değerinin %93.09, mAP@0.5:0.95 değerinin ise %89.04 olduğu belirtilmiştir. Adam algoritmasıyla yapılan karşılaştırmalı deneylerde ise SGD'nin daha genel sonuçlar verdiği ve aşırı uyum sorunlarını azalttığı gözlemlenmiştir. Sonuç olarak, bu çalışma, süpermarketlerdeki ürün tespiti ve durum doğrulama sorunlarını çözmek için etkili bir yaklaşım sunmaktadır. Gelecekte, ürünlerdeki hasar bölgelerinin daha ayrıntılı tespit edilmesi amacıyla Generative Adversarial Networks (GAN) teknolojisinin entegre edilmesi ve model metriklerinin daha da geliştirilmesi hedeflenmektedir.

Bu tez Shen (2020), yoğun nesne algılama ve tespiti üzerine odaklanarak, market raflarındaki ürünlerin otomatik tespiti ve sayılması için derin öğrenme tabanlı CenterNet modelini incelemektedir. Çalışmanın temel amacı, market raflarında yoğun ve küçük nesnelere algılanmasında insan hatalarını ve zaman kaybını azaltmak için etkili bir yöntem geliştirmektir. Bu bağlamda, market raflarında yer alan ürünleri içeren "Bottle" veri seti oluşturulmuş ve bu veri seti üzerinde CenterNet, Faster-RCNN ve YOLOv3 modelleri karşılaştırılmıştır.

Araştırmada, özellikle geometrik bilgilerin algılama üzerindeki etkisi incelenmiş ve renk bilgisinin ek katkı sağlamakla birlikte geometrik bilgilere kıyasla daha az etkili olduğu görülmüştür. Perspektif dönüşümü gibi teknikler kullanılarak, modelin güçlü açılara sahip görüntülerdeki performansı iyileştirilmiştir. CenterNet'in ResNet-101 tabanlı versiyonu, mAP ve AP50 metriklerinde diğer modellerden daha yüksek performans sergilemiştir. YOLOv3 hızlı ve etkili bir model olarak değerlendirilirken, Faster-RCNN doğruluk açısından yüksek sonuçlar sağlamıştır. Sonuçlar, geliştirilen yöntemin market raflarındaki ürünlerin algılanmasında endüstriyel gereksinimleri karşılayabilecek doğruluğa ulaştığını göstermektedir. Gelecekte, daha karmaşık geometrik dönüşümlerin ve renk uzaylarındaki varyasyonların incelenmesiyle model performansının artırılması planlanmaktadır.

Özet

Yapay sinir ağları, makine öğrenmesinde sıklıkla kullanılan bir öğrenme yöntemidir. İnsan beyninin sinir yapısı dikkate alınarak modellenmiştir. Yapay sinir ağları deneyimlerden öğrenmektedir. Geleneksel algoritmalarından farklı olarak yapay sinir ağları tarafından öğrenilen modellerin insanlar tarafından anlaşılması zor olabilmektedir. Çok katmanlı yapay sinir ağları derin öğrenmede kullanılmaktadır. Literatürde derin öğrenme ağı ilk olarak Ivakhnenko and Lapa (1966) tarafından oluşturulmuştur. Bu yöntem o zamandan beri büyük ölçüde gelişmiştir. Ancak bu gelişme ile birlikte oldukça yüksek hesaplama gücüne olan ihtiyaç ortaya çıkmıştır. Grafik işlem birimi (GPU) donanım kapasitesi arttıkça günlük hayatta kullanımı da hızlanmaktadır. Evrimsel sinir ağı, bilgisayarlı görme, görüntü işleme, ses işleme, doğal dil işleme gibi birçok alanda kullanılan bir derin öğrenme yöntemidir. Bireysel katmanlardan oluşan bu model, bir görüntü üzerinde çalıştırıldığında kenarlar, çizgiler, eğriler gibi yerel özelliklerden, bir nesneyi diğer nesnelere ayıran genel özelliklere kadar birçok özelliği belirlemektedir.

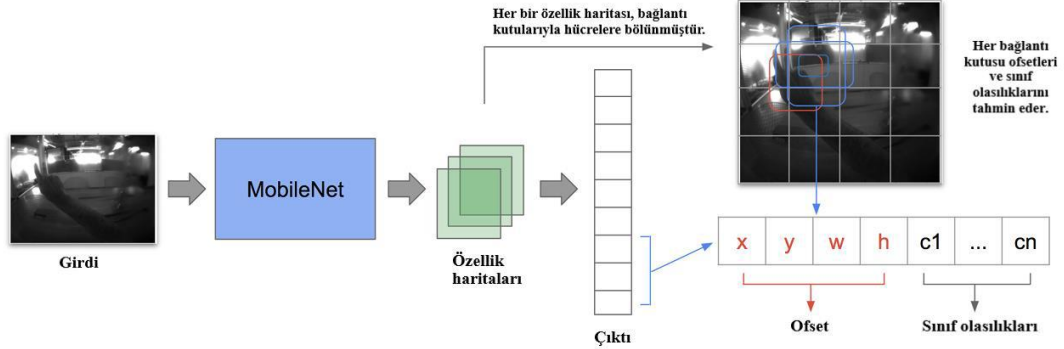
Genel olarak konuyu ele aldığımızda, özellik çıkarımı görüntü işleme teknolojisiyle sağlanmaktadır. Birçok ön işleme işleminden sonra farklı sınıflandırma yöntemleri uygulanır. Bu sonuçla yüksek performans elde edilmektedir. Bu işlemler çok zaman aldığı için yazılımı lisanslayarak sınıflandırma yöntemini mevcut verilere uygulamak için hazırlık yapmak gerekebilmektedir. Bu tür zaman, malzeme veya çaba harcamalarını azaltmak için ya da ortadan kaldırmak için, özellikleri kendi kendine öğrenebilen ve bu özelliklere göre sınıflandırma yapabilen evrimsel sinir ağları gibi yöntemlerin kullanılması daha uygun olabilir.

MATERYAL YÖNTEM

Tezin bu bölümde nesne tespiti için kullanılan yöntemler açıklanmıştır. Literatürdeki nesne tespiti ve market ürünlerinin tespiti ile ilgili en ileri seviye çalışmalar incelenmiş ve SSD Mobilenet ve YOLO modelleri analizler için kullanılmıştır. Ayrıntılı bir analiz için YOLO modelinin farklı versiyonları çalışmaya katılmış ve bunları farkları açıklanmıştır. Sonraki bölümde performans analizinde kullanılan metriklerin tanımı ve denklemleri verilmiştir. Tez kapsamında geliştirilen veri setinin ayrıntıları ve veri etiketleme işleminin nasıl yapıldığı bir sonraki bölümde açıklanmıştır. Bölümün sonunda tez kapsamında gerçekleştirilen deney ve analizlerin akış şeması sunulmuştur.

SSD Mobilenet v2

SSD Mobilenet nesne tespiti uygulamaları için geliştirilmiş bir derin öğrenme mimarisidir. İki temel bileşenden oluşur; SSD ve MobileNet. SSD (Single Shot Multibox Detector) bir nesne tespit modelidir ve çoklu sınır kutuları tahmini yapar. SSD, gerçek zamanlı nesne tespiti için oldukça uygun bir modeldir. MobileNet bileşeni mobil uygulamalar için geliştirilmiş CNN tabanlı bir derin öğrenme modelidir.



Şekil 5. SSD MobileNet (Pandey ve ark., 2017)

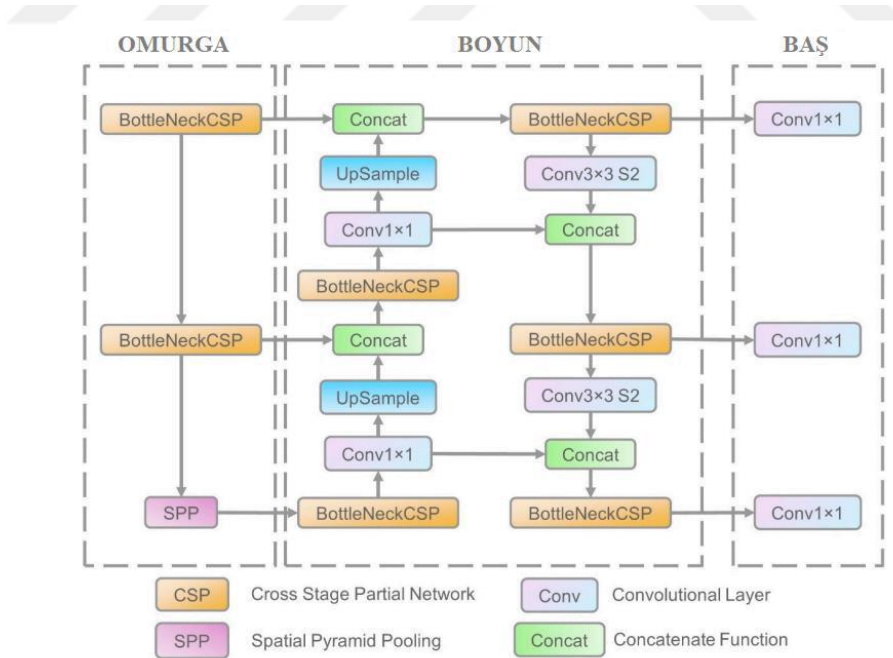
MobileNet V2, MobileNet mimarisinin geliştirilmiş bir versiyonudur. Mobil cihaz üzerinde görüntü işleme uygulamalarını çalıştırmak için geliştirilmiş bir yazılım **yazılım mimarisidir** (Çetiner, 2022). MobileNet, 28 katman ve 4253864 parametreden oluşan evrişimli bir sinir ağıdır. MobileNet V2 ise 53 katmana ve 3,4 milyon parametreye sahiptir. Bu değerler aynı derinlikteki ağlarla karşılaştırıldığında parametre sayısı önemli ölçüde artmaktadır (Şafak ve Barışçı, 2023).

Sonuç olarak, SSD MobileNet v2 kombinasyonu, hızlı nesne tespiti ve tanıma yapmak için kullanılan, mobil cihazlar ve sınırlı hesaplama gücüne sahip sistemler için ideal bir modeldir. Bu model, görüntülerdeki nesnelere tanımlamak ve sınıflandırmak için düşük gecikme süresiyle çalışacak şekilde tasarlanmıştır. Şekil 3.1’de SSD MobilNet Modelin yapısı sunulmaktadır (Çelebi 2023).

YOLOV-5

YOLOv5 (Zhang 2023), Ultralytics'in Kurucusu ve CEO'su Glenn Jocher tarafından 2020 yılında geliştirilmiş derin öğrenme temelli bir nesne tespit modelidir. YOLO (You Only Look Once) modelinin beşinci sürümü olan bu model, orijinal YOLO modelinin bir devamıdır fakat farklı bir geliştirici ekibi tarafından geliştirilmiştir. YOLOv5, PyTorch kütüphanesi kullanılarak geliştirilmiş ve GitHub üzerinde açık kaynak olarak yayımlanmıştır.

YOLOv4 Eğitim süreci Darknet kütüphanesinde YOLOv5 modeli takip edilerek gerçekleştirilmiştir. Pytorch kütüphanesinde uygulanmaktadır. Ek olarak YOLOv5 tarafından oluşturulan ağırlık dosyası YOLOv4'e taşınabilmektedir. Model boyutu daha küçüktür. Bu şekilde YOLOv5, GPU'da YOLOv4'e dönüştürülebilir. Daha hızlı çalışabilen ve daha yüksek çözünürlüğe sahip bir modeldir. Sonuç olarak YOLOv5 modeli doğruluk ve hız açısından ileri bir modeldir (Şahin 2023). Şekil 3.2’de model mimarisi sunulmuştur.



Şekil 6. YOLOv5 mimarisi (Şahin 2023)

Faster R-CNN

Faster R-CNN, gerçek zamanlı nesne tespiti için geliştirilen ve derin öğrenme tabanlı mimariler arasında önemli bir yere sahip bir modeldir. Bu model, nesne tespiti görevinde hız ve

doğruluk açısından büyük bir ilerleme sağlamıştır. Faster R-CNN, iki temel bileşeni birleştirir: Bölge Öneri Ağı (Region Proposal Network-RPN) ve Fast R-CNN. Bölge Öneri Ağı, görüntü üzerinde potansiyel nesne bölgelerini hızlı bir şekilde belirler ve bu bölgeleri nesne sınıflandırması ve konumlandırma için Fast R-CNN bileşenine iletir. Bu yapı, geleneksel bölge öneri yöntemlerinin neden olduğu yavaşlığı ortadan kaldırarak modeli çok daha hızlı hale getirir (Ren 2017).

Faster R-CNN'in en önemli katkılarından biri, tamamen uçtan uca eğitim yapılabilmesidir. Bölge önerileri ve sınıflandırma aynı ağ içinde optimize edildiği için, hem bölge önerilerinin doğruluğu hem de nesne tespiti performansı önemli ölçüde artar. Modelin mimarisi, derin öğrenmenin güçlü özellik çıkarma kapasitesini kullanarak hem küçük hem de büyük nesnelerin tespitinde yüksek başarı sağlar. Ayrıca, bu model, bölge önerilerinin hesaplanmasında kullanılan kayar pencere ve çapraz çözünürlük gibi yenilikçi yöntemlerle, farklı ölçeklerdeki nesnelere algılamak için yüksek esneklik sunar (Ren 2017).

Makale, Faster R-CNN'in, özellikle Bölge Öneri Ağı ile tespit hızını artırırken doğruluktan ödün vermediğini vurgulamaktadır. Bu mimari, RPN'nin potansiyel nesne bölgelerini filtrelemek için sınıflandırma skorlarını kullanmasını ve ardından bu bölgelerin Fast R-CNN tarafından işlenmesini sağlar. Sonuç olarak Faster R-CNN, özellikle büyük ölçekli veri setlerinde hızlı ve doğru nesne tespiti yapabilen bir model olarak, bilgisayarla görme alanında önemli bir kilometre taşı olmuştur (Ren 2017).

YOLO V-8

YOLOv8 (Chen 2023), Ultralytics tarafından geliştirilen ve nesne algılama alanında ileri teknoloji sunan bir modeldir. Önceki sürümlerine kıyasla daha yüksek doğruluk, hız ve esneklik sunarak bilgisayarla görme görevlerinde önemli bir avantaj sağlar. YOLOv8'in mimarisi, modern derin öğrenme prensiplerine dayanır ve kullanıcıların nesne algılama, sınıflandırma ve segmentasyon gibi görevleri daha verimli bir şekilde gerçekleştirmesine olanak tanır.

Model, optimize edilmiş konvolüsyonel katmanlar ve özellik piramidi ağlarını kullanarak daha iyi özellik çıkarımı yapar ve bu sayede hem küçük hem de büyük nesnelere yüksek doğruluk sağlar. YOLOv8'in düşük gecikme süresi, gerçek zamanlı uygulamalarda üstün performans sunmasına olanak tanır. Ayrıca, modelin kullanıcı dostu arayüzü, kolayca eğitilebilir olması ve transfer öğrenimi desteği sayesinde geniş bir veri kümesi üzerinde hızlıca uyarlanabilir.

Geliştirilmiş mimarisi sayesinde YOLOv8, eğitim sırasında öğrenme sürecini optimize ederken, çıkarım aşamasında da hız ve hassasiyeti dengeler. Model, endüstriyel otomasyon, sağlık sektörü ve tarım gibi çeşitli alanlarda uygulanabilme potansiyeline sahiptir. Özetle, YOLOv8, doğruluk ve hız açısından önceki sürümleri geride bırakarak nesne algılama teknolojilerinde önemli bir referans noktası haline gelmiştir.

YOLO V-11

YOLOv11 (Sharma 2024), diğer YOLO sürümleri gibi derin öğrenme temelli bir nesne tespit modelidir. Gelişmiş mimarisi, modern derin öğrenme yaklaşımlarını bir araya getirerek, hem karmaşık görevleri hızlı bir şekilde çözme hem de yüksek doğruluk oranları sunma hedefiyle tasarlanmıştır. Bu model, nesne algılama ve sınıflandırma süreçlerinde yeni bir standart belirlerken, segmentasyon gibi daha detaylı analizler için de oldukça uygun bir yapıya sahiptir.

Önceki YOLO modellerine göre, YOLOv11'in yapısında kullanılan daha derin ve daha etkili özellik çıkarım katmanları, modelin küçük ve büyük nesnelere aynı hassasiyetle algılayabilmesini sağlamaktadır. Ayrıca, veri setlerinde bulunan çeşitliliği daha iyi kavrayabilmesi için geliştirilen yeni nesil veri artırma teknikleri, modelin farklı senaryolarda daha tutarlı sonuçlar vermesine yardımcı olur. YOLOv11, optimize edilmiş hesaplama yapısıyla hem güçlü bilgisayarlarda hem de düşük donanım kapasitelerine sahip cihazlarda üstün bir performans sunar.

Model, özellikle gerçek zamanlı uygulamalarda hız ve doğruluk dengesini korumasıyla dikkat çeker. Yoğun nesne dağılımına sahip görüntülerde bile hatasız tespitler yapabilen YOLOv11, büyük veri setlerini işlerken verimliliğinden ödün vermemektedir. Eğitim sürecinde kullandığı optimize edilmiş algoritmalar, modelin hızlı öğrenme yeteneğini artırır ve kullanıcıların özel veri kümeleri üzerinde yüksek doğrulukla çalışmasını mümkün kılar.

Günümüzde YOLOv11, güvenlik sistemlerinden tıbbi görüntü analizi ve akıllı şehir uygulamalarına kadar pek çok farklı alanda kullanılmaktadır. Özellikle, yüksek doğruluk ve düşük gecikme süresi gerektiren uygulamalarda, YOLOv11, performansı ve esnekliği ile rakiplerini geride bırakmayı başarmaktadır. Bu model, bilgisayarla görme teknolojilerinin geleceğini şekillendiren önemli bir araç haline gelmiştir.

Karışıklık Matrisi

Karışıklık matrisi, sınıflandırma problemlerinde modelin performansını ölçmek için kullanılan bir tablodur. İki veya daha fazla sınıfın tahmin sonuçlarını özetleyen bir matristir. Dört ana bileşenden oluşur:

- **Doğru Pozitif (TP):** Doğru pozitif. Modelin doğru şekilde pozitif sınıfa atadığı örnekler.
- **Doğru Negatif (TN):** Doğru negatif. Modelin doğru şekilde negatif sınıfa atadığı örnekler.
- **Yanlış Pozitif (FP):** Yanlış pozitif. Negatif sınıfa ait bir örneğin yanlışlıkla pozitif sınıfa atanması.
- **Yanlış Negatif (FN):** Yanlış negatif Pozitif sınıfa ait bir örneğin yanlışlıkla negatif sınıfa atanması.

Bu dört bileşen yardımıyla karışıklık matrisi şu şekilde tanımlanır:

	Pozitif Tahmin	Negatif Tahmin
Doğru Pozitif	TP	FN
Doğru Negatif	FP	TN

Doğruluk (Accuracy): Denklem 1’de verilmiştir.

$$\text{Doğruluk} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Modelin doğru tahminlerinin tüm tahminlere oranını verir.

Kesinlik (Precision): Denklem 2’de verilmiştir.

$$\text{Kesinlik} = \frac{TP}{TP + FP} \quad (2)$$

Pozitif olarak tahmin edilenlerin ne kadarının gerçekten pozitif olduğunu gösterir.

Duyarlılık (Recall): Denklem 3’da verilmiştir.

$$\text{Duyarlılık} = \frac{TP}{TP + FN} \quad (3)$$

Gerçek pozitiflerin ne kadarının doğru bir şekilde tahmin edildiğini gösterir.

F1- Skor: Denklem 4’de verilmiştir.

$$\text{F1 - Skor} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

Kesinlik ve Duyarlılık arasındaki dengeyi ölçen bir değerdir.

Bu denklemler yardımıyla modelin doğruluğu, kesinliği, duyarlılığı ve diğer performans ölçütleri hesaplanır.

Ortalama Doğruluk Değeri (mAP)

Ortalama hassasiyeti (AP) hesaplamak için, hassasiyet ve geri çağırma işlevlerinin fonksiyonlarını anlamamız gerekmektedir. Hassasiyet, tahminin doğruluğunu hesaplar ve geri çağırma fonksiyonu, tahminin ne kadarının geri alındığını hesaplar.

Her bir sınıf için, modelin farklı eşik değerlerindeki hassasiyet ve geri çağırma değerlerini kullanarak bir hassasiyet-geri çağırma eğrisi oluşturulur.

Her bir sınıf için, hassasiyet-geri çağırma eğrisi altında kalan alan, o sınıf için AP (average precision) değerini verir. AP'nin denklemi Denklem 5'de verilmiştir. Bu, modelin o sınıftaki performansının bir özetidir.

$$AP = \int_0^1 P(R) dR \quad (5)$$

Son adım olarak, tüm sınıflar için hesaplanan AP değerlerinin ortalaması alınarak mAP değeri elde edilir. Bu, modelin tüm sınıflar üzerindeki genel performansının bir özetidir (Denklem 6).

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (6)$$

Burada N , sınıf sayısını ve AP_i , i sınıf için hesaplanan AP değerini ifade eder.

Ortalama doğruluk değeri, modelin farklı sınıfları ne kadar iyi ayırt ettiğini ve tespit ettiğini genel bir ölçü olarak sunar. Yüksek bir ortalama doğruluk değeri, modelin tüm sınıflarda iyi bir performans gösterdiğini gösterirken, düşük bir ortalama doğruluk değeri modelin bazı sınıflarda zayıf performans gösterdiğine işaret edebilir.

mAP(0.5), yalnızca IoU'nun 0.5 olduğu bir eşığe göre hesaplanır. Bu, modelin tahmin ettiği sınırlayıcı kutunun, gerçek kutu ile en az %50 örtüşme sağladığında başarılı kabul edildiği anlamına gelir.

Veri Seti Oluşturma ve Veri Etiketleme

Bu tez çalışmasında Erzurum market zincirlerinden nesne tespiti yapabilmek için raflardaki ürünler kullanılarak veri seti oluşturulmuştur. Bu veri setinin temel amacı market raflarında yer alan benzer ürünlerinin tespitinin analiz edilmesidir. Bu amaçla beş farklı içecek markasının alt kategorileri seçilerek 47 sınıftan oluşan bir veri seti oluşturulmuştur. Veri seti oluşturulurken görüntüler, ürünlerin birlikte olduğu raf görüntülerinden oluşmaktadır. Toplam 456 görüntü alınmış ve bu görüntülerde toplam 7725 etiket (nesne işaretlenmiştir).

Veri setinin etiketlenmesi için, kullanımı kolay ve açık kaynaklı bir etiketleme aracı olan 'Roboflow 3.0 Object Detection (Fast)' kullanılmıştır. Ekran çıktıları Şekil 3.4.'de çıktılar verilmiştir. Bu araç sayesinde, her bir görüntüdeki ürünlerin elle işaretlenmesi ve kategorilere ayrılması sağlanmıştır. Etiketleme süreci sırasında, her bir ürünün konumu, türü ve diğer ayırt edici özellikleri dikkatlice belirlenmiş ve etiketlenmiştir. Bu işlem, veri setinin doğruluğunu ve tutarlılığını artırmak için titizlikle gerçekleştirilmiş olup, makine öğrenimi modellerinin eğitiminde kullanılacak yüksek kaliteli bir veri seti oluşturulmasına olanak tanımıştır. Toplam 456 görüntü alınmış ve bu görüntülerden toplam 7725 etiket (nesne) işaretlenmiştir.

Veri Seti Sayısal Değerleri:

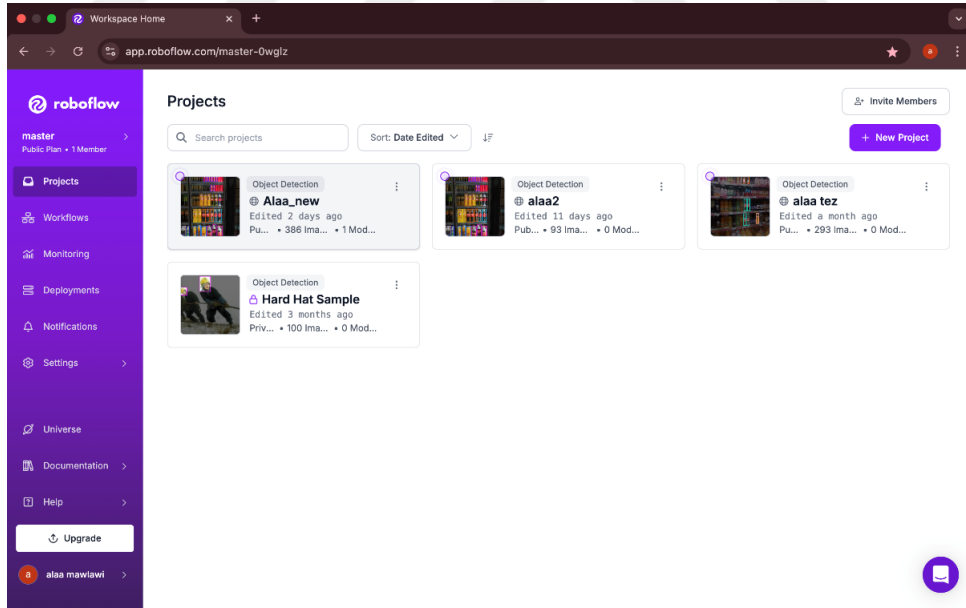
Toplam 414 adet eğitim görüntüsü ve 414 etiket dosyası kullanılmıştır.

42 adet doğrulama (validation) görüntüsü bulunmaktadır.

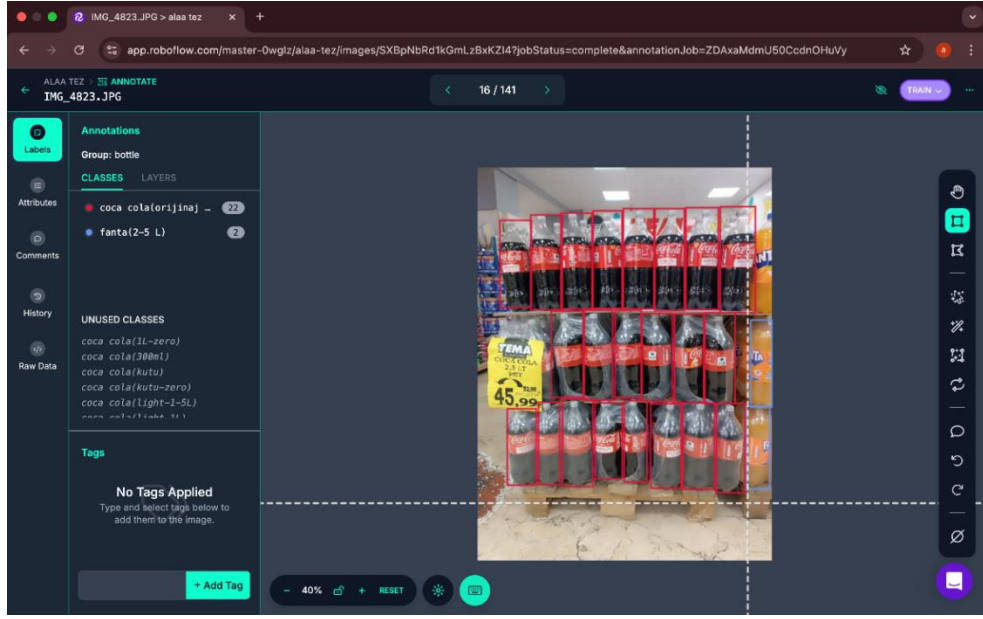
Etiket Sayıları:

Eğitim setinde, market ürünlerinin görsellerinde aynı fotoğraf üzerinde birden fazla nesne bulunduğu için 7050 adet etiket (nesne) yer almaktadır.

Doğrulama setindeki 42 fotoğrafta ise toplam 675 adet etiket (nesne) bulunmaktadır.



Şekil 7. Etiket programında görsellerin sınıflandırılması (devamı)



Şekil 7. Etiket programında görsellerin sınıflandırılması

Sınıflandırmamız, sınıfa düşen etiket sayısı ve sınıf içerisinde yer alan fotoğraf sayıları Tablo 1’de verilmiştir:

Çizelge 1. Sınıflandırma, Etiket Sayısı ve Fotoğraf Sayısı

Sınıf Adı	Etiket Sayısı	Fotoğraf Sayısı
FANTA-2 L-	102	58
FANTA-300 ml-	38	20
Pepsi-1-5L-	112	72
Pepsi-1L-	102	60
Pepsi-2-5L-	12	6
Pepsi-2L-	136	64
Pepsi-kutu-	30	16
Sprite-2L-	66	58
coca cola -2L-	108	52
coca cola -300 ml - zero-	6	6
coca cola -300 ml-	62	28
coca cola -kutu-	266	48
coca cola-1L-zero-	128	96
coca cola-kutu-	518	62
coca cola-kutu-zero-	388	104
coca cola-light-1-5L-	80	34
coca cola-light-1L-	56	22
coca cola-oriijinal tat-1-5L-	364	116
coca cola-oriijinal tat-1L-	248	118

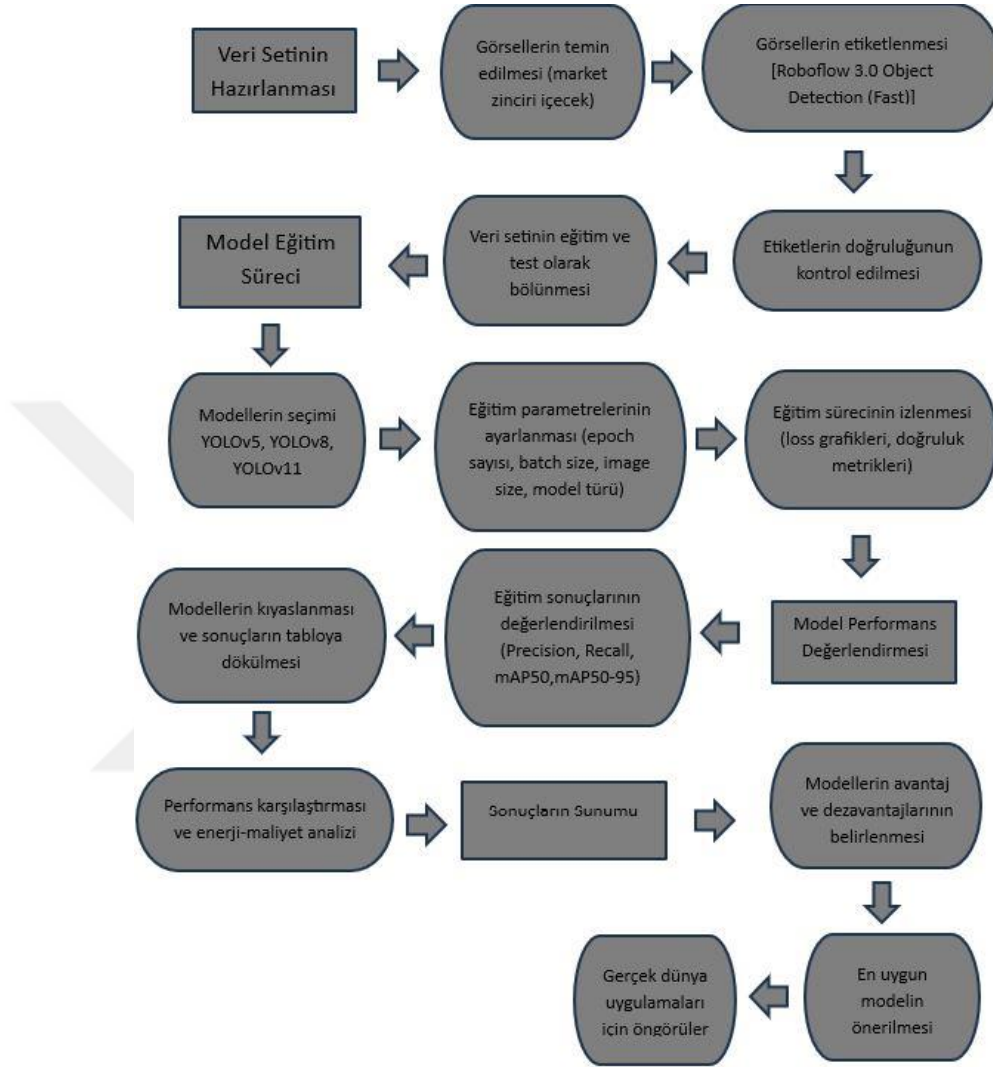
Çizelge 1. (devamı)

coca cola-orijinal tat-2-5L-	820	156
coca cola-sise-	146	54
coca cola-zero-2-5L-	74	10
coca cola-zero-1-5L-	156	92
cola-turka-	24	6
cola-turka-sise	44	12
fanta-1-5 L-	228	106
fanta-1L-	284	108
fanta-2-5 L-	816	146
fanta-kutu-	232	28
pepsi -kutu-	106	38
pepsi -kutu- lime-	114	34
pepsi -kutu- zero-	208	60
pepsi -zero-1L-	42	26
pepsi-sise-	26	14
sprite-1 L-	58	28
sprite-1 L- zero-	30	18
sprite-1-5 L-	46	24
sprite-2-5 L-	294	130
sprite-kutu - zero-	16	8
sprite-kutu-	62	18
uludag	104	34
uludag-1-5L-	30	8
uludag-az kalori-2-5L-	56	30
uludag-az sekerli 1-5L-	58	30
uludag-az sekerli 2-5L-	52	46
uludag-sekersiz-	50	24
uludag-sise-	52	14

Bu çalışmada, derin öğrenme modellerinin performansını değerlendirmek amacıyla sistematik bir süreç izlenmiştir. İlk olarak, veri setinin hazırlanması aşamasında market zincirine ait içecek görselleri temin edilip Roboflow 3.0 Object Detection aracı kullanılarak etiketlenmiştir. Etiket doğruluğu kontrol edildikten sonra, veri seti eğitim ve test verisi olarak ikiye bölünmüştür. YOLOv5, YOLOv8 ve YOLOv11 gibi farklı model türleri seçilmiş ve eğitim parametreleri ayarlanarak eğitim süreci başlatılmıştır. Eğitim boyunca elde edilen eğitim sonuçları değerlendirilmiştir. Modellerin avantaj ve dezavantajları belirlenerek en uygun model önerilmiş ve sonuçlar detaylı performans karşılaştırması ve enerji-maliyet analizi ile

desteklenmiştir. Gerçek dünya uygulamaları için yapılan öngörüler doğrultusunda bu çalışma, farklı modellerin etkinliğini kıyaslamak ve optimum çözüm önerisi sunmak adına kapsamlı bir çerçeve sunmaktadır.

Aşağıda Şekil 8’de bu çalışmanın akış diyagramı yer almaktadır.



Şekil 8. Tez çalışmasının akış diyagramı

ARAŞTIRMA BULGULARI

Bu çalışmada farklı derin öğrenme modellerinin performanslarını analiz etmek amacıyla sistematik bir süreç izlenmiştir. İlk olarak, YOLOv5, YOLOv8 ve YOLOv11 gibi nesne tanıma modelleri kullanılarak geniş kapsamlı bir karşılaştırma yapılmıştır. Araştırmaya SsdMobileNetv2 ve Faster R-CNN modelleri de destek vermiş olup çalışmanın kapsama alanını genişletmiştir. Araştırmanın temel aşamaları; veri setinin oluşturulması, görsellerin etiketlenmesi, eğitim süreçlerinin gerçekleştirilmesi ve elde edilen sonuçların değerlendirilmesi üzerine yapılandırılmıştır.

Her bir modelin doğruluk, duyarlılık, kesinlik ve ortalama doğruluk (mAP) gibi performans metrikleri detaylı bir şekilde incelenmiştir. Ayrıca eğitim süresinin, enerji tüketiminin ve donanım kullanımının etkileri de karşılaştırmalı olarak ele alınmıştır.

Araştırma sürecinde elde edilen bulgulara göre, YOLOv8 modeli duyarlılık değeri açısından en yüksek performansı sergilerken, YOLOv5 modeli kesinlik oranında öne çıkmıştır. Öte yandan YOLOv11, kesinlik ve duyarlılık arasında en dengeli performansı sunarak F1 skoru açısından lider konuma ulaşmıştır. Bununla birlikte, her bir modelin eğitim süreci boyunca kayıp fonksiyonlarının azaldığı ve model performansının istikrarlı bir şekilde iyileştiği gözlemlenmiştir. Eğitim süreci boyunca kaydedilen grafikler, modellerin öğrenme kabiliyetlerini ve veriye uyum sağlamalarını net bir şekilde ortaya koymaktadır.

Sonuç olarak, araştırmada elde edilen bulgular, projelerin ihtiyaçlarına göre model seçiminin belirlenmesinin önemini vurgulamaktadır. Enerji ve donanım kısıtlamalarının bulunduğu senaryolarda YOLOv5 öne çıkarken, yüksek doğruluk gerektiren uygulamalarda YOLOv8 daha uygun bir seçenek olarak değerlendirilebilir. Dengeli performans ve genel tutarlılık arayışında ise YOLOv11 modeli güçlü bir alternatif sunmaktadır. Bu analizler ışığında, her bir modelin avantajları ve dezavantajları detaylı olarak ortaya konmuş, gerçek dünya uygulamaları için kapsamlı öngörüler sağlanmıştır.

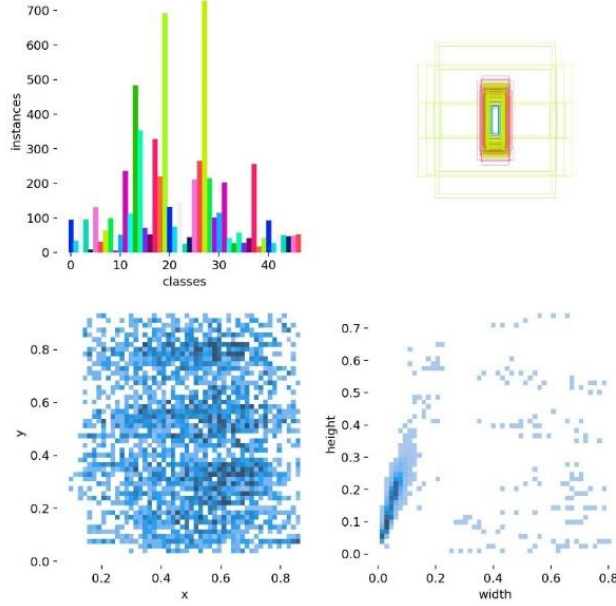
Etiket Dağılımı ve Sınır Kutusu Görselleştirmesi

Tez kapsamında oluşturulan veri seti 47 sınıftan oluşmaktadır. Elde edilen görüntülerde yaklaşık 7000 nesne el ile sınır kutularını alınmış ve etiketlenmiştir. Veri setinin kapsamlı bir analizi için bu etiketlerin görsellere nasıl yayıldığı ve sınır kutularının boyutlarının ne oranda

değişiklik gösterdiği görselleştirilmiştir. İlgili görsel Şekil 4.1.'te sunulmuştur. Yapılan analizler ile varılan sonuçlar şu şekildedir:

- Sol üst etiket dağılımı: Sınıf dağılımının eşit olmadığı kanısına varabiliyoruz. Bazı sınıflar diğerlerine kıyasla çok daha fazla örneğe sahip. Bu durum, modelin bu sınıfları daha iyi öğrenmesine neden olabilirken, az örneğe sahip sınıflar için modelin düşük performans göstermesi olasılığı doğurabilir. Bu sınıf dengesizliği, veri artırma (data augmentation) yöntemleriyle giderilebilir. Market raflarında böyle dengesizliklerin olması veri setini de etkilediğini görmekteyiz.
- Sağ üst sınır kutusu dağılımı: Çoğu sınır kutusu benzer boyutlara ve merkez konumlara sahip olduğu görülüyor. Bu, nesnelerin görüntüde sıkça belirli alanlarda yoğunlaştığını ve boyutlarının genelde sabit bir aralıkta kaldığını gösteriyor.
- Sol Alt: x-y Koordinatları Dağılımı: Nesnelerin görüntü içindeki bu dağılımı, gerçek dünya verilerinde çekim özelliklerini yansıtır. Ancak, yoğunluk yalnızca merkezde yoğunlaşıyorsa, modelin görüntülerin kenarlarında yer alan nesnelere algılamakta zorlanabileceği düşünülebilir.
- Sağ alt: genişlik-yükseklik dağılımı: Veri kümesindeki nesnelerin çoğu dar ve yüksek yapıda görünüyor, bu da nesnelerin dikdörtgen şeklinde olduğunu düşündürür. Ancak genişlik ve yükseklik dağılımında aşırı bir değişkenlik olmaması, modelin farklı boyutlardaki nesnelere için genelleme yapmasını kolaylaştırabilir.

Veriler, genelde tutarlı bir dağılıma sahip olsa da, kenarlardaki nesnelere veya küçük boyutlu sınır kutusu için ek düzenlemeler yapılması, modelin daha dengeli bir performans göstermesini sağlar.



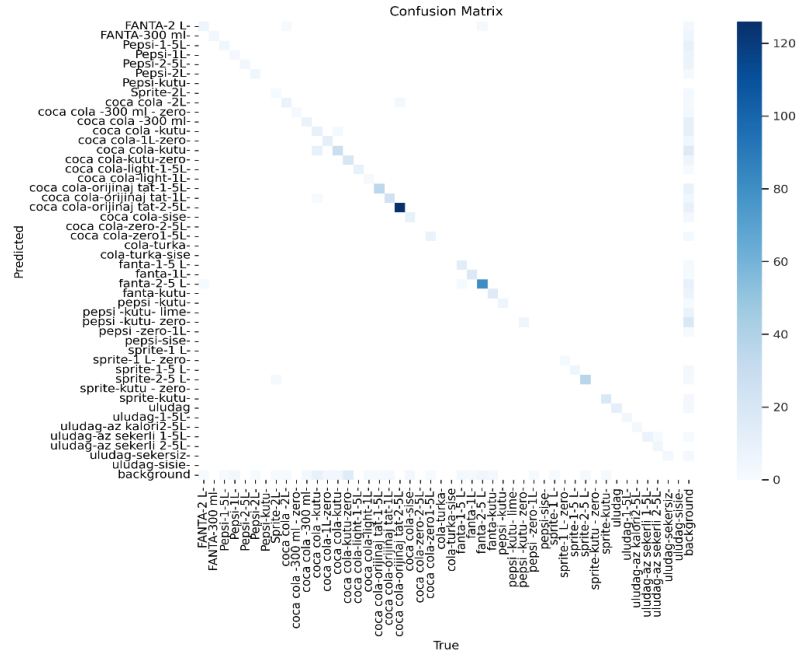
Şekil 9. Etiket dağılımı ve sınır kutusu görselleştirmesi

YOLO V-5 Deneysel Sonuçlar

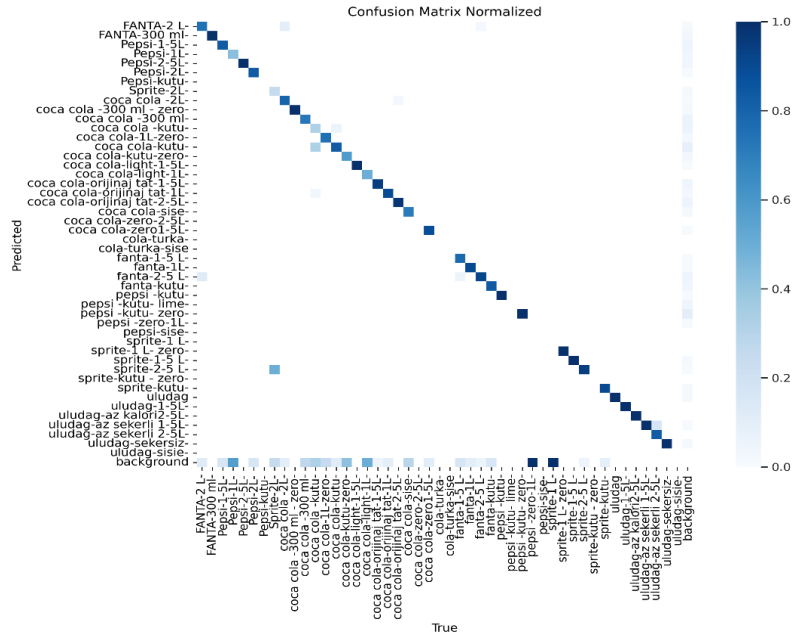
YOLO V-5 için karmaşıklık matrisi

YOLOv5 modeline ait karmaşıklık matrisleri Şekil 10'de sunulmuştur. Modelin genel performansının oldukça iyi olduğu görülmektedir. İlk görsel olan ham karmaşıklık matrisi, modelin doğru tahminlerini ve olası hatalarını ham değerlerle göstermektedir. Burada köşegen boyunca yoğunlaşan koyu mavi renkler, modelin sınıfları büyük ölçüde doğru bir şekilde tahmin ettiğini ifade etmektedir. Ancak, köşegen dışında yer alan düşük yoğunluklu alanlar, modelin bazı sınıflar arasında karışıklık yaşadığını göstermektedir. Bu durum, özellikle benzer özelliklere sahip sınıflar arasında doğal bir hata olasılığına işaret etmektedir. Örneğin, aynı marka içeceklerin farklı türlerinin karıştırılması gibi durumlar bu karışıklığa neden olabilir.

İkinci görsel olan Şekil 11'de normalize edilmiş karmaşıklık matrisi ise sınıflar arasındaki tahmin oranlarını ortaya koymaktadır. Normalizasyon sayesinde, her sınıfın kendi içindeki doğru tahmin oranı daha belirgin hale gelmiştir. Köşegen boyunca görülen yüksek yoğunluklu mavi alanlar, modelin doğruluk oranının oldukça yüksek olduğunu göstermektedir. Yanlış tahmin oranlarının düşük seviyede olması, modelin veri dengesizliklerinden fazla etkilenmediğini ve genelleme yeteneğinin güçlü olduğunu ifade etmektedir. Genel olarak, modelin performansı oldukça iyi bir seviyededir ve doğru tahmin oranları yüksek olduğu için sınıflandırma görevinde başarılı bir performans sergilediği görülmektedir. Ancak, bazı sınıflar arasındaki karışıklığın nedenleri detaylıca analiz edilerek, ek veri veya model optimizasyonu ile bu karışıklık daha da azaltılabilir.



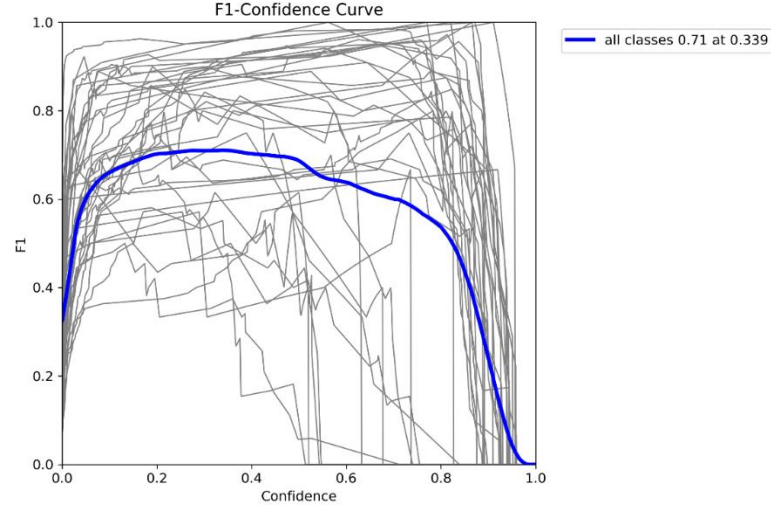
Şekil 10. YOLO V-5 için karmaşıklık matrisi



Şekil 11. YOLO V-5 için normalize karmaşıklık matrisi

YOLO V-5 için F1-güven eğrisi

Şekil 12 YOLOv5 modelinin güven eşiği ile F1 skorunun ilişkisini göstermektedir. Grafikte kalın mavi çizgi, tüm sınıfların F1 skorlarının ortalamasını temsil etmektedir. Mavi çizginin en yüksek noktası, modelin optimum performans gösterdiği eşiği işaret eder. Bu noktada F1 skoru 0.71 olarak hesaplanmış ve güven eşiği değeri yaklaşık 0.339 olarak

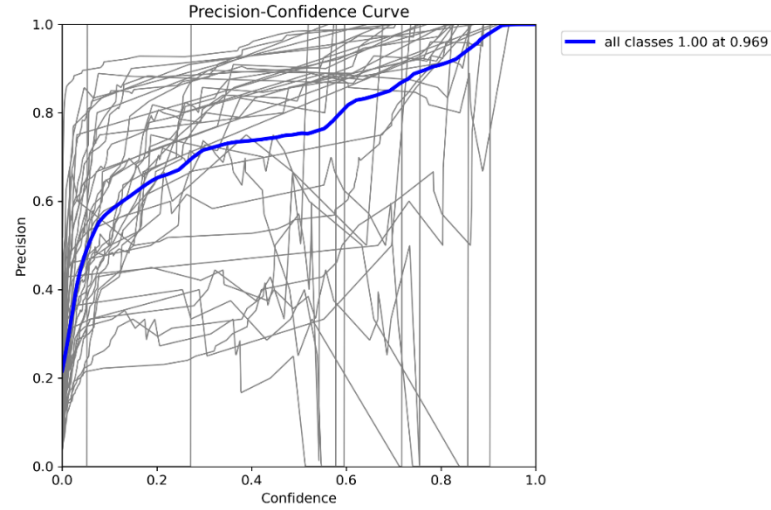


Şekil 12. YOLO V-5 için F1 skor eğrisi

Şekil 12, YOLOv5 modelinin güven eşiği ile F1 skorunun ilişkisini göstermektedir. Grafikte kalın mavi çizgi, tüm sınıfların F1 skorlarının ortalamasını temsil etmektedir. Mavi çizginin en yüksek noktası, modelin optimum performans gösterdiği eşiği işaret eder. Bu noktada F1 skoru 0.71 olarak hesaplanmış ve güven eşiği değeri yaklaşık 0.339 olarak belirlenmiştir. Bu, modelin hem doğruluk hem de geri çağırma açısından dengeli bir performans sergilediği güven seviyesidir. Daha düşük bir eşik değeri, modelin daha fazla tahmin yapmasına olanak tanırken yanlış pozitiflerin sayısını artırabilir. Yüksek bir eşik değeri ise yalnızca çok emin olunan tahminlerin yapılmasını sağlar ancak bu durum geri çağırma oranını düşürebilir.

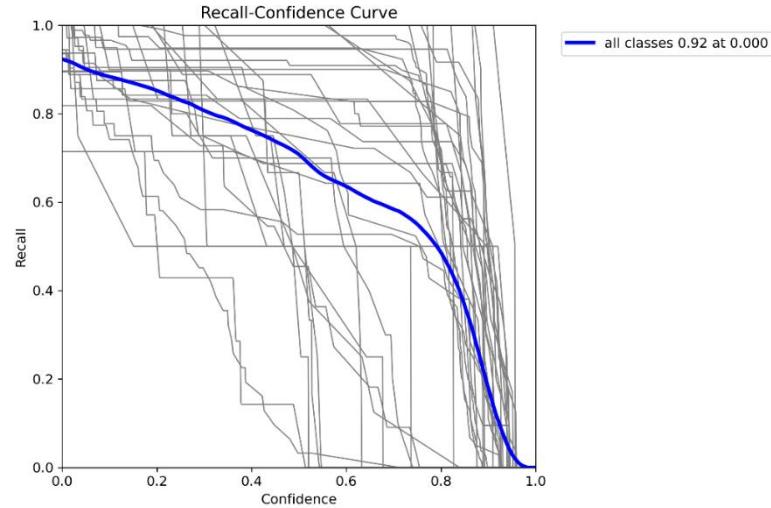
YOLO V-5 için doğruluk-güven eğrisi, doğruluk-duyarlılık eğrisi ve duyarlılık-güven eğrisi grafikleri

Güven eşiği (Confidence threshold) yükseldikçe, model yalnızca daha emin olduğu tahminleri yapmaya başlar ve bu durum doğruluk oranını artırır. Şekil 13’de görüldüğü üzere, modelin en yüksek doğruluk oranına güven eşiğinin 0.969 olduğu noktada ulaştığını ortaya koymaktadır. Bu, modelin doğru tahmin yapma kabiliyetinin yüksek olduğunu ve yanlış pozitiflerin oldukça az olduğunu ifade eder. Ancak, yüksek güven eşiği kullanıldığında model daha az tahmin yapar, bu da geri çağırma oranının düşmesine yol açar.



Şekil 13. YOLO V-5 için doğruluk-güven eğrisi

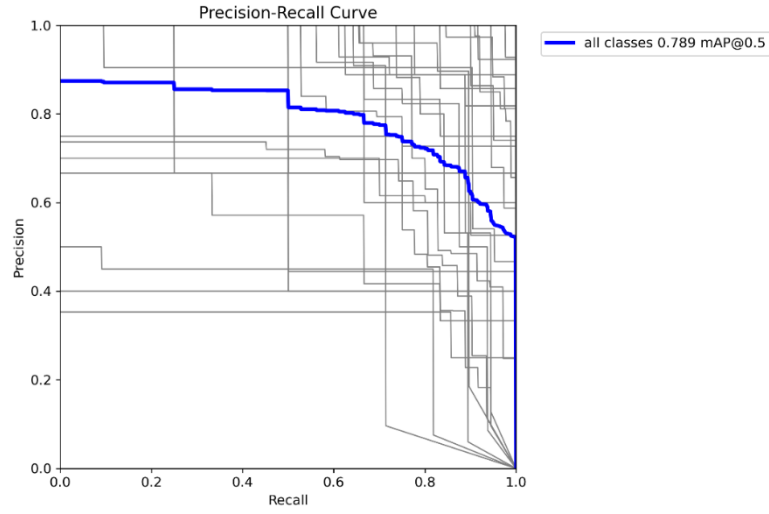
Güven eşiği yükseldikçe, model yalnızca daha emin olduğu tahminleri yapmaya başlar ve bu durum doğruluk oranını artırır. Modelin en yüksek doğruluk oranına güven eşiğinin 0.969 olduğu noktada ulaştığını ortaya koymaktadır. Bu, modelin doğru tahmin yapma kabiliyetinin yüksek olduğunu ve yanlış pozitiflerin oldukça az olduğunu ifade eder. Ancak, yüksek güven eşiği kullanıldığında model daha az tahmin yapar, bu da geri çağırma oranının düşmesine yol açar.



Şekil 14. YOLO V-5 için duyarlılık-güven eğrisi

Şekil 14 geri çağırma oranının güven eşiğine göre nasıl değiştiğini göstermektedir. Güven eşiği değeri arttıkça geri çağırma oranının azaldığı görülmektedir. Düşük güven eşiği seviyelerinde model, daha fazla tahmin yapar ve bu durum geri çağırma oranını artırır. Grafik, tüm sınıflar için maksimum geri çağırma oranının 0.92 olduğunu ve bunun düşük güven seviyelerinde elde edildiğini göstermektedir. Ancak, bu durum aynı zamanda yanlış pozitif tahminlerin de artabileceği anlamına gelir. Güven eşiğinin yüksek olduğu durumlarda ise model

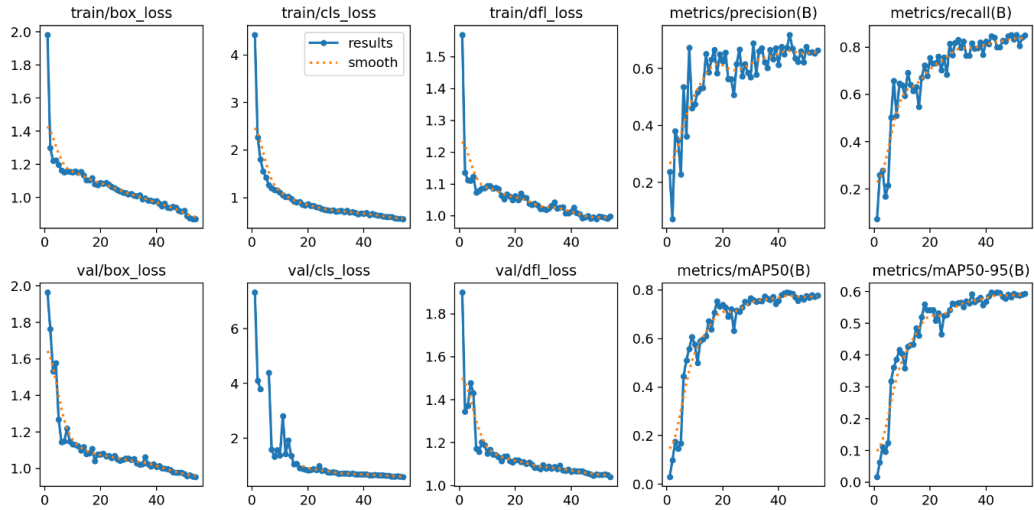
yalnızca çok emin olduğu tahminleri yapar, bu da geri çağırma oranını düşürse bile daha güvenilir tahminler elde edilmesini sağlar.



Şekil 15. YOLO V-5 için doğruluk-duyarlılık eğrisi

Şekil 15, doğruluk ve duyarlılık arasındaki ilişkiyi görselleştirmektedir. Tüm sınıflar için ortalama doğruluk değeri 0.789 olarak hesaplanmıştır, bu da modelin genel performansının oldukça yüksek olduğunu göstermektedir. Grafik, doğruluk oranının geri çağırma arttıkça düşme eğiliminde olduğunu göstermektedir. Bu, modelin daha fazla tahmin yaptığında bazı yanlış pozitiflerin meydana gelebileceğini ifade eder

YOLO V-5 için sonuç değerlendirmesi



Şekil 16. YOLO V-5 için kayıp grafikleri

Şekil 16'de YOLOv5 modeli ile yapılan eğitim ve doğrulama sürecinin metriklerini görselleştirmektedir. Grafikler, modelin farklı kayıp fonksiyonlarını ve doğruluk ölçümlerini boyunca nasıl geliştirdiğini ve optimize ettiğini göstermektedir.

Eđitim ve dođrulama kaybı

1. train/box_loss-val/box_loss:

Modelin sınır kutusu koordinatlarını optimize etme kaybı, hem eđitim (train) hem de dođrulama (val) sürecinde sürekli olarak azalmıřtır. Bu durum, modelin nesne sınırlarını daha dođru bir řekilde öğrenebildiđini ve genelleme kabiliyetinin iyi olduđunu göstermektedir.

2. train/cls_loss-val/cls_loss:

Modelin sınıflandırma kaybı, her iki süreçte de önemli ölçüde azalmıřtır. Bu, modelin farklı sınıfları ayırt etme yeteneđinin artıđını ve öğrenmenin etkili bir řekilde gerçekteřtiđini ifade eder.

3. train/df_l_loss-val/df_l_loss:

Distribution Focal Loss (DFL), modelin kesin nesne tahmini yapma kabiliyetini geliřtiren bir metriđi temsil eder. Grafikler, bu kaybın eđitim ve dođrulama süreçlerinde sürekli azaldıđını, modelin nesne konumlarında yüksek dođruluk sađladıđını göstermektedir.

Dođruluk ve performans metrikleri

1. metrics/precision-metrics/recall:

Kesinlik ve duyarlılık deđerleri, modelin hem dođru pozitif oranını hem de geri çağırma oranını artırdıđını göstermektedir. Epok (Epoch) ilerledikçe bu metriklerin daha kararlı hale gelmesi, modelin hem hassas hem de kapsayıcı bir performans sergilediđini iřaret eder.

2. metrics/mAP50-metrics/mAP50-95:

Ortalama Dođruluk, modelin genel performansını deđerlendiren bir metriktir. mAP@50 deđeri yaklaşık 0.8'e ulařırken, mAP@50-95 deđeri 0.6 seviyesine ulařmıřtır. Bu, modelin farklı IoU eřiklerinde yüksek performans gösterdiđini kanıtlamaktadır.

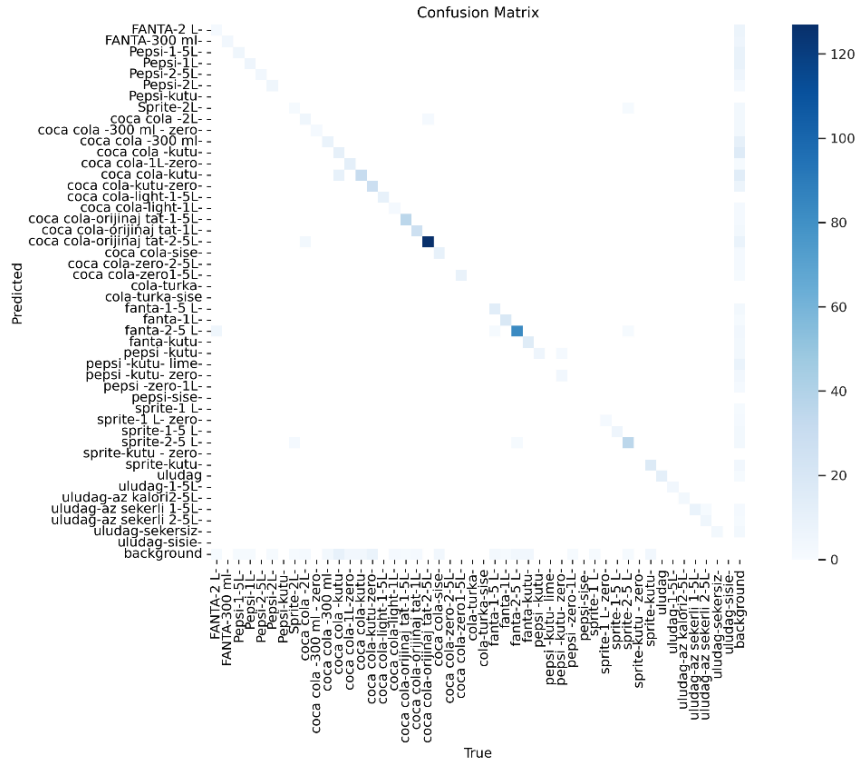
YOLO V-8 Deneysel Sonuçlar

YOLO V-8 için karmařıklık matrisi

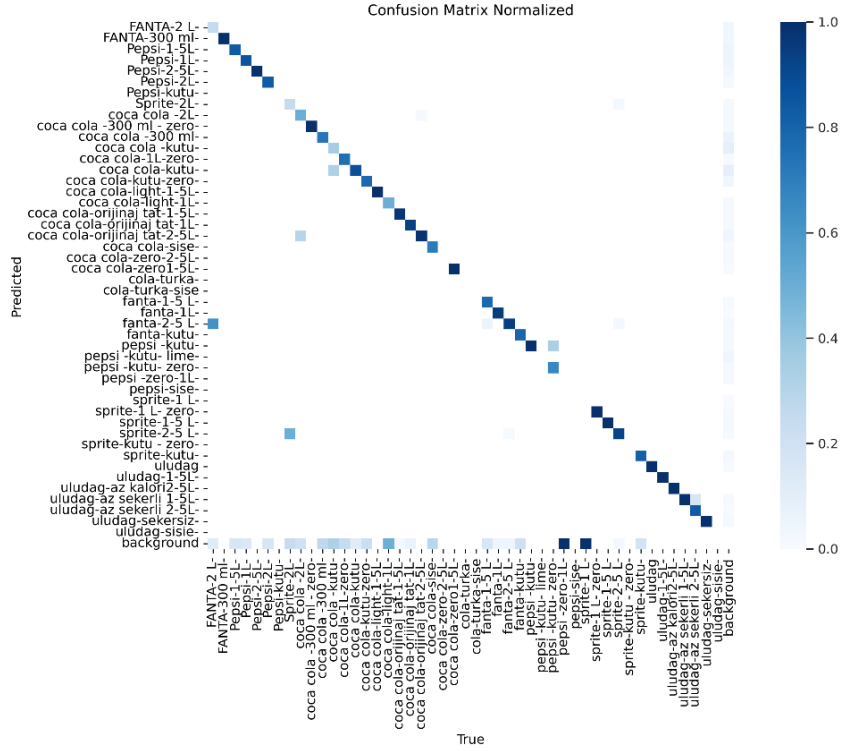
Yolo V-8 modeli için elde edilmiř karmařıklı matrisleri řekil 27 ve řekil 28'da gösterilmiřtir. İlk karmařıklık matrisinde, modelin genel dođruluk performansının oldukça yüksek olduđu gözlemlenmektedir. Dođru sınıflandırmalar diyagonal eksen boyunca yoğunlařmıř olup, modelin çođunlukla dođru tahminler yaptıđı görölmektedir. Ancak bazı sınıflar arasında karmařıklıklar belirgin bir řekilde mevcuttur. Özellikle aynı marka ve farklı boyutlara sahip

ürünler arasında yanlış tahminler yoğunlaşmaktadır. Bu durum, veri setindeki sınıflar arasındaki görsel benzerliklerden kaynaklanabilir ve modelin bu tür örneklerde zorlandığını göstermektedir.

İkinci matriste, normalize edilmiş veriler üzerinden yapılan analizlerde, sınıflandırma doğruluğu daha belirgin bir şekilde görülmektedir. Modelin güçlü olduğu sınıflar, normalize edilmiş değerlerde net bir şekilde vurgulanmıştır. Ancak, düşük normalize edilmiş değerlerin bulunduğu sınıflar, modelin bu örneklerdeki karışıklığının devam ettiğini göstermektedir. Özellikle, sınıflar arasındaki görsel benzerliklerin etkisi daha açık bir şekilde gözlemlenmiştir. Normalize edilmiş metrikler, yanlış sınıflandırma oranlarını daha iyi anlamamıza olanak tanımış ve iyileştirme gereken sınıfları netleştirmiştir. Genel olarak, modelin sınıflar arasında iyi bir doğruluk seviyesi sağladığı, ancak daha fazla veri çeşitliliği ve özellik iyileştirmesi ile performansın artırılabilceği sonucuna varılmıştır.



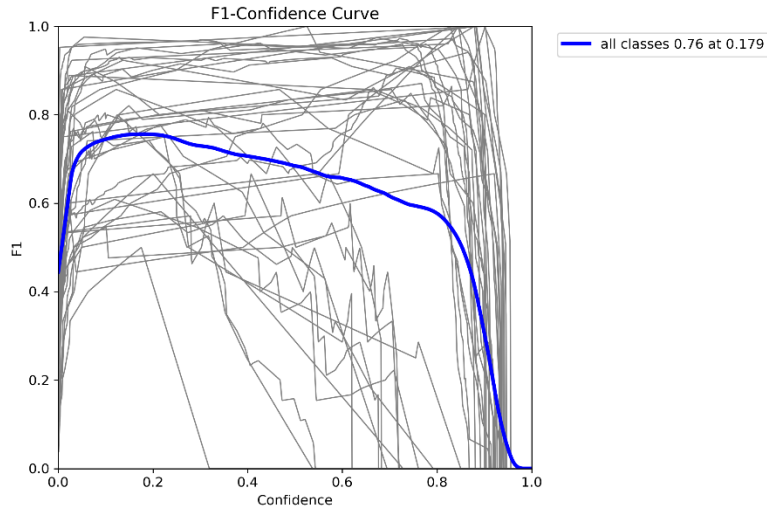
Şekil 17. YOLO V-8 için karmaşıklık matrisi



Şekil 18. YOLO V-8 için normalize karmaşıklık matrisi

YOLO V-8 için F1 F1-güven eğrisi

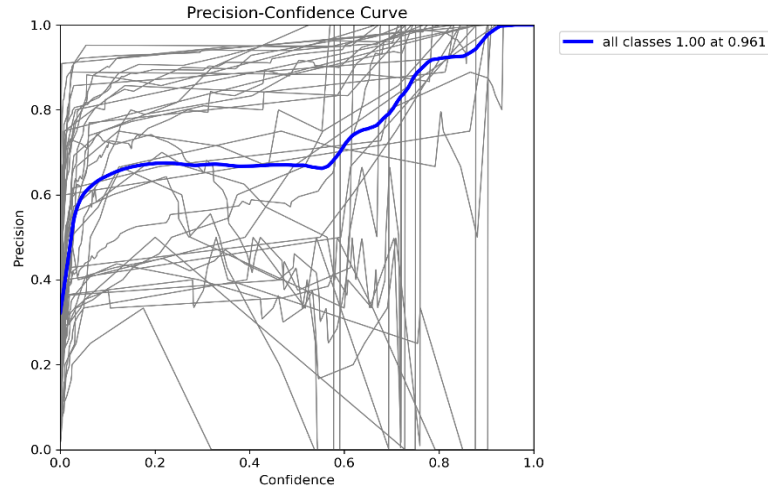
F1-güven eğrisinde, tüm sınıflar için F1 skoru ortalama olarak 0.76 değerine ulaşmıştır. Bu, modelin genel doğruluk ve geri çağırma dengesi açısından iyi bir performans sergilediğini göstermektedir. Şekil 19’de görüldüğü üzere eğri, güven seviyesi 0.179 noktasında en yüksek değerine ulaşmış ve bu güven seviyesinde modelin optimal bir şekilde çalıştığı sonucuna varılmıştır. Güven seviyesi arttıkça F1 skorunda bir düşüş gözlemlenmekte, bu da daha yüksek güven seviyelerinde modelin tahminlerinin daha seçici olduğunu gösterir



Şekil 19. YOLO V-8 için F-1 skor eğrisi

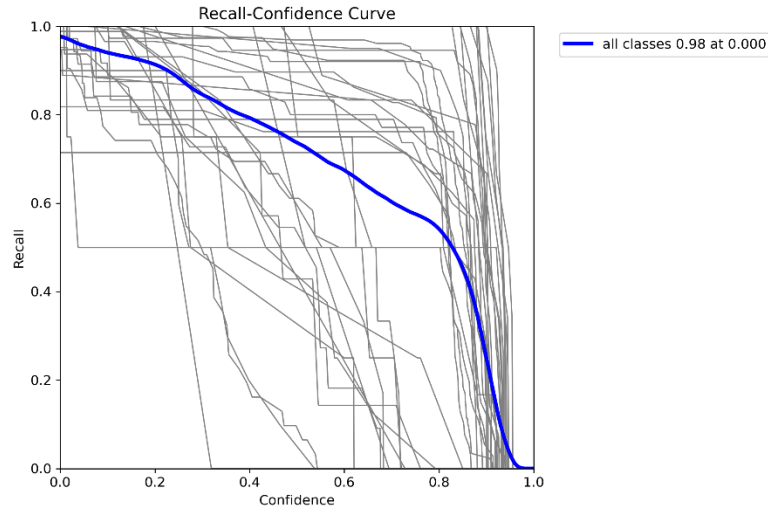
YOLO V-8 için doğruluk-güven eğrisi, doğruluk-duyarlılık eğrisi ve duyarlılık-güven eğrisi grafikleri

Şekil 20’de de görüldüğü gibi kesinlik-güven eğrisinde, tüm sınıflar için modelin hassasiyeti 0.961 güven seviyesinde 1.00 değerine ulaşmıştır. Bu, modelin bu güven seviyesinde yaptığı tahminlerin tamamının doğru olduğunu kesin bir şekilde göstermektedir. Daha düşük güven seviyelerinde hassasiyetin daha düşük olduğu ancak güven seviyesi arttıkça doğruluğun hızla iyileştiği gözlemlenmiştir. Genel olarak, modelin yüksek güven seviyelerinde oldukça tutarlı ve hatasız sonuçlar verdiği net bir şekilde söylenebilir.



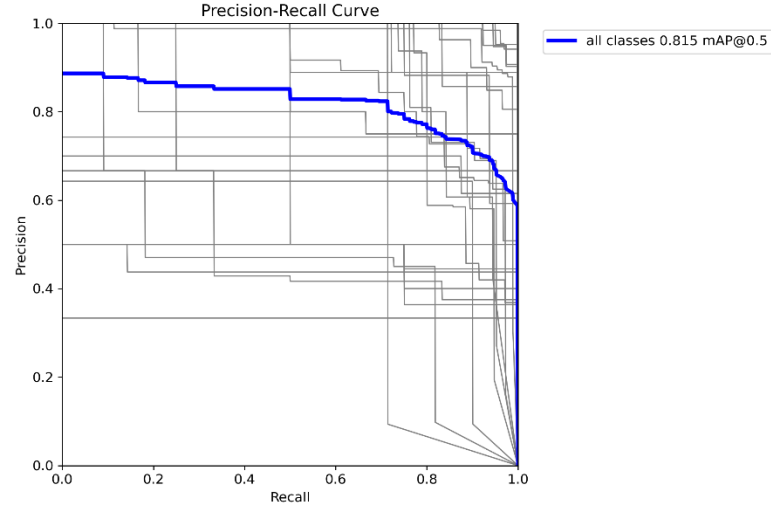
Şekil 20. YOLO V-8 için doğruluk-güven eğrisi

Şekil 21’de gösterilen duyarlılık-güven eğrisinde, tüm sınıflar için duyarlılık değeri, 0.000 güven seviyesinde 0.98 olarak ölçülmüştür. Bu, modelin düşük güven seviyelerinde oldukça yüksek bir kapsayıcılığa sahip olduğunu göstermektedir. Ancak, güven seviyesi arttıkça geri çağırma değerinde bir düşüş gözlemlenmektedir. Bu durum, modelin daha seçici hale geldikçe bazı doğru tahminleri göz ardı ettiğini net bir şekilde ortaya koymaktadır.



Şekil 21. YOLO V-8 için duyarlılık-güven eğrisi

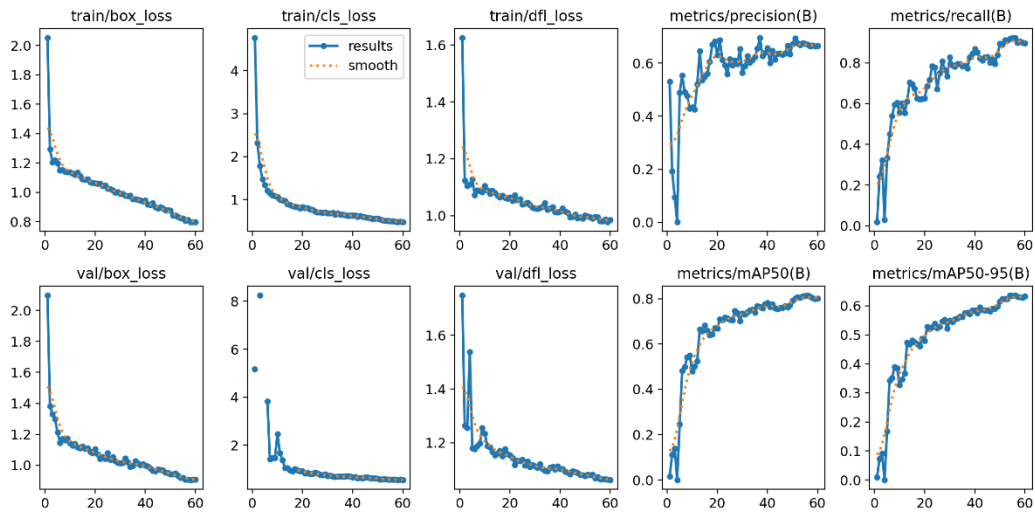
Şekil 22’de gösterilen kesinlik-duyarlılık eğrisinde, tüm sınıflar için ortalama ortalama doğruluk değeri 0.815 olarak hesaplanmıştır. Bu, modelin genel performansının oldukça iyi olduğunu göstermektedir. Eğri, modelin yüksek kesinlik ve duyarlılık dengesi sağladığını, ancak yüksek duyarlılık değerlerine yaklaştıkça kesinlik değerinde düşüş yaşandığını açıkça ortaya koymaktadır. Bu, modelin daha fazla doğru tahmin yaparken yanlış pozitif oranının da arttığını ifade eder.



Şekil 22. YOLO V-8 için kesinlik-duyarlılık eğrisi

YOLO V-8 için kayıp grafikleri

YOLO V-8 için Kayıp Grafikleri Şekil 23’de gösterilmiştir. Box loss, class loss ve DFL loss hem eğitim hem de doğrulama için sürekli bir düşüş sergilemiş, bu da modelin her iki süreçte de başarılı bir öğrenme gerçekleştirdiğini göstermektedir.



Şekil 23. YOLO V-8 için kayıp grafikleri

Kesinlik, duyarlılık, ve ortalama doğruluk değeri metrikleri açısından değerlendirildiğinde, modelin performansının her geçen epokta istikrarlı bir şekilde arttığı

gözlemlenmiştir. $mAP@0.5$ değeri yüksek bir seviyeye ulaşmış, bu da modelin doğru sınıflandırma ve nesne tespiti açısından güçlü bir performans sergilediğini kanıtlamaktadır. $mAP@0.5:0.95$ değerinde de tutarlı bir artış mevcuttur, bu da modelin hem küçük hem de büyük nesnelere tespit etme yeteneğinin güçlü olduğunu ifade etmektedir.

Sonuç olarak, grafikler modelin eğitim sürecinin verimli bir şekilde tamamlandığını, kayıpların minimize edildiğini ve metriklerde yüksek başarı sağlandığını net bir şekilde göstermektedir. Model, hem doğruluk hem de kapsayıcılık açısından beklentilerin üzerinde bir performans sergilemiştir.

YOLO V-8 için sonuç

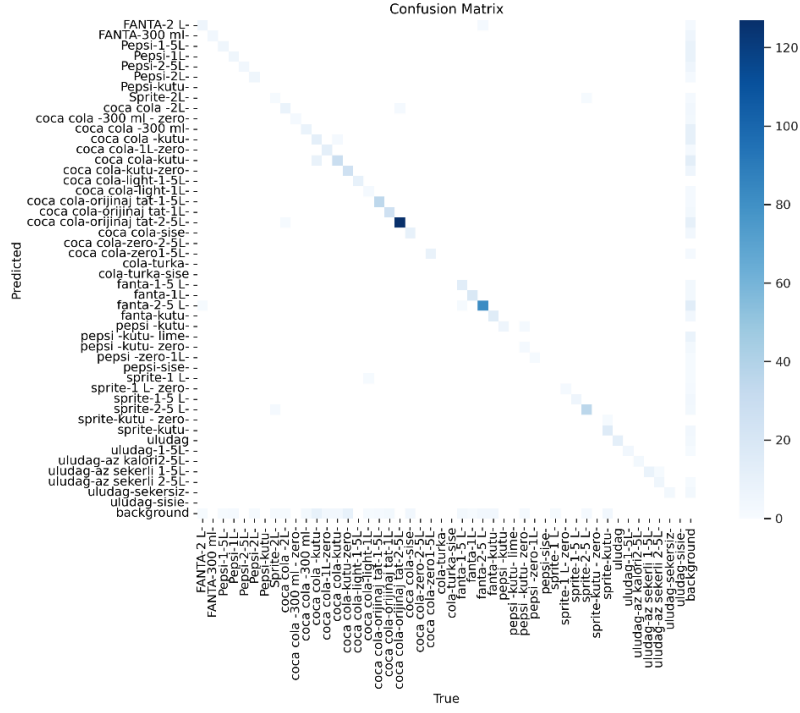
YOLOv8 modeli, $mAP@0.5$ değeri 0.815 ve F1 skoru 0.76 ile yüksek doğruluk sergilemiştir. Kesinlik 0.961 güven seviyesinde 1.00'e ulaşmış, duyarlılık ise 0.98 olarak ölçülmüştür. Eğitim ve doğrulama süreçlerinde box, class ve DFL kayıpları istikrarlı şekilde azalmış, modelin veriye iyi uyum sağladığı ve aşırı öğrenmeden kaçındığı görülmüştür. Karışıklık matrisinde genelde doğru tahmin yapılmış, ancak benzer sınıflar arasında düşük oranlı karışıklıklar gözlemlenmiştir. Küçük nesnelere yüksek performans sağlayan model, gerçek zamanlı uygulamalar ve hassas tahminler için güçlü bir aday olduğunu kanıtlamıştır.

YOLO V-11 Çıktıları

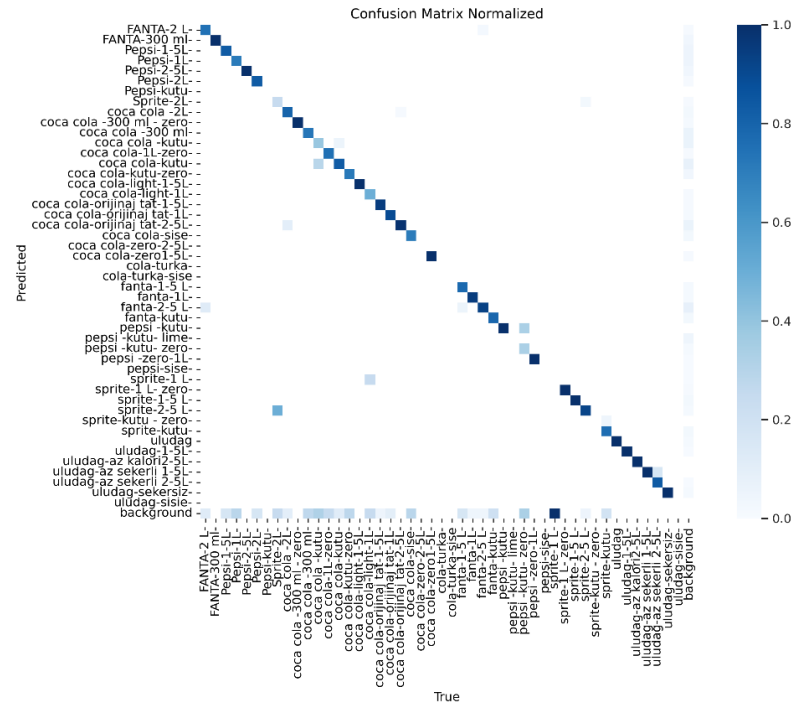
YOLO V-11 için karmaşıklık matrisi

YOLOv11 karışıklık matrisleri Şekil 24 ve Şekil 25'da gösterilmiştir. Grafikler incelendiğinde modelin sınıflandırma başarısının oldukça iyi olduğu net bir şekilde görülmektedir. Doğru tahminler, matrisin diyagonal çizgisinde yoğunlaşmıştır. Bazı sınıflar arasında karışıklıklar bulunsun da, bunlar oldukça sınırlıdır ve modelin temel yeteneklerini zayıflatmamaktadır. Arka plan verilerinin yüksek hassasiyetle ayrılmış olması, YOLOv11'in sahne ile nesne ayrımı konusundaki başarısını kanıtlamaktadır.

Modelin performansı, benzer özelliklere sahip sınıflarda dahi tatmin edici bir ayırım gücüne sahip olduğunu göstermektedir. Bununla birlikte, özellikle görsel benzerliklerin fazla olduğu sınıflarda zaman zaman tahminlerde sapmalar gözlemlenmiştir. Bu sapmalar, verilerin daha kapsamlı ve dengeli bir şekilde düzenlenmesiyle minimize edilebilir. YOLOv11, nesne tanıma alanında hem doğruluk hem de tutarlılık açısından ileri seviye sonuçlar ortaya koymuş ve zorlu veri setlerinde bile etkileyici bir performans göstermiştir. Model, gerçek zamanlı analiz için güçlü bir aday olarak değerlendirilebilir.



Şekil 24. YOLO V-11 için karmaşıklık matrisi

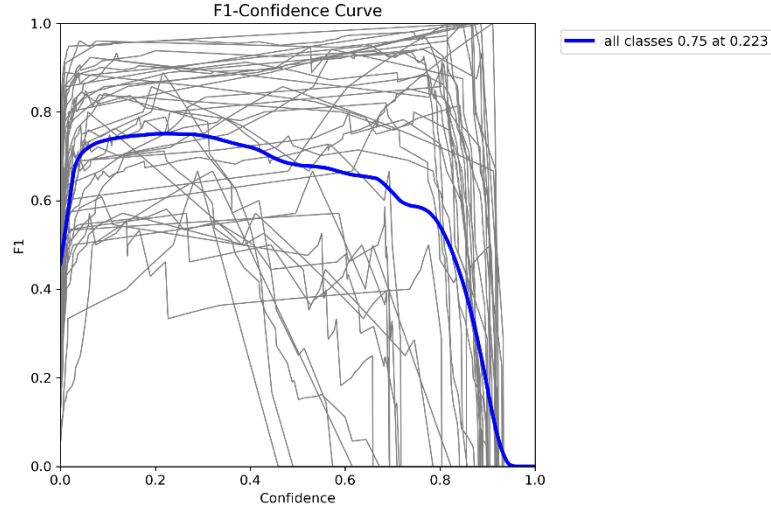


Şekil 25. YOLO V-11 için normalize karmaşıklık matrisi

YOLO V-11 için F1-güven eğrisi

YOLOv11'in F1-güven eğrisi Şekil 26'de gösterilmiştir. Eğri incelendiğinde, tüm sınıflar için F1 skoru 0.75 ile tatmin edici bir seviyeye ulaşmıştır. Bu en yüksek değer, güven seviyesi 0.223'te elde edilmiştir ve bu nokta modelin hem doğruluk hem de geri çağırma dengesi açısından en ideal performansı sergilediği aralığı temsil etmektedir. Güven seviyesi

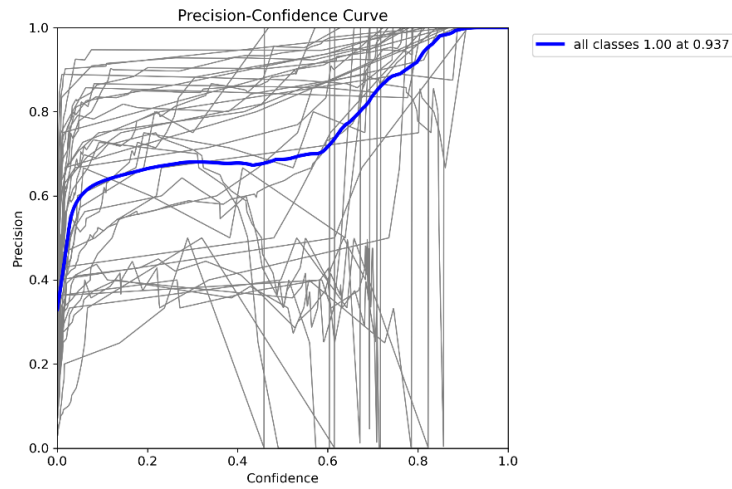
artıkça F1 skorunda kademeli bir azalma gözlemlenmiş, bu da modelin seçiciliğinin artmasıyla birlikte bazı doğru tahminlerin elenebildiğini göstermektedir.



Şekil 26. YOLO V-11 için F-1 skor eğrisi

YOLO V-11 için doğruluk-güven eğrisi, doğruluk-duyarlılık eğrisi ve duyarlılık-güven eğrisi grafikleri

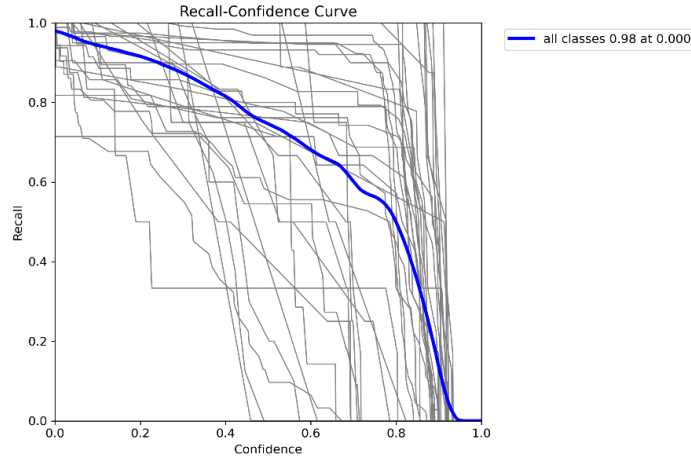
Şekil 27’de gösterilen YOLOv11’in kesinlik-güven eğrisi incelendiğinde, model tüm sınıflar için 0.937 güven seviyesinde %100 hassasiyet değerine ulaşmıştır. Bu, modelin bu güven seviyesinde yaptığı tahminlerin tamamen doğru olduğunu göstermektedir. Daha düşük güven seviyelerinde hassasiyet daha düşüktür, ancak güven arttıkça doğruluk hızla iyileşmektedir.



Şekil 27. YOLO V-11 için kesinlik-güven eğrisi

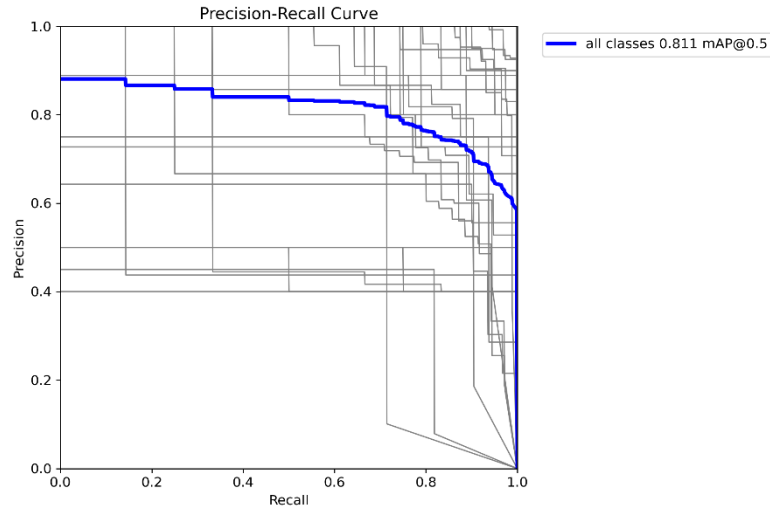
Şekil 28’de gösterilen YOLOv11’in duyarlılık-güven eğrisi incelendiğinde, model 0.000 güven seviyesinde %98 duyarlılık oranına ulaşmıştır. Bu, modelin düşük güven seviyelerinde neredeyse tüm doğru örnekleri kapsayabildiğini göstermektedir. Ancak, güven seviyesi arttıkça duyarlılık değerinde düşüş yaşanmaktadır. Bu durum, modelin daha seçici hale

geldikçe bazı doğru tahminleri göz ardı ettiğini ifade eder. YOLOv11, düşük güven seviyelerinde geniş kapsayıcılık sağlayarak, yüksek duyarlılık gerektiren uygulamalarda etkili bir performans sunmaktadır.



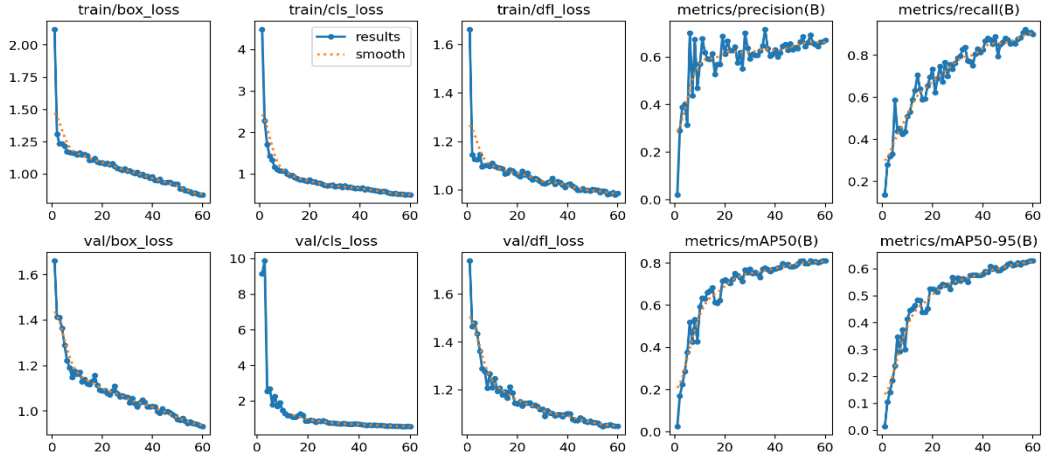
Şekil 28. YOLO V-11 için duyarlılık-güven eğrisi

Şekil 29’da gösterilen YOLOv11’in kesinlik-duyarlılık eğrisi incelendiğinde, modelin $mAP@0.5$ değeri 0.811 olarak hesaplanmıştır. Bu, modelin hem doğruluk hem de kapsayıcılık açısından dengeli bir performans sergilediğini göstermektedir. Kesinlik yüksek duyarlılık seviyelerinde dahi istikrarlı bir şekilde korunmuş, ancak duyarlılık %100’e yaklaştığında hassasiyet düşüşe geçmiştir.



Şekil 29. YOLO V-11 için kesinlik-duyarlılık eğrisi

YOLO V-11 için kayıp grafikleri



Şekil 30. YOLO V-11 için kayıp grafikleri

YOLOv11'in eğitim ve doğrulama grafikleri Şekil 30, box loss, class loss ve DFL loss değerleri hem eğitim hem de doğrulama aşamalarında düzenli olarak azalmış ve minimum seviyelere ulaşmıştır. Bu, modelin veriye uyum sağladığını ve aşırı öğrenme göstermediğini gösterir.

Hassasiyet metrikleri, her epokta belirgin bir iyileşme göstermiş ve modelin doğruluk odaklı performansını kanıtlamıştır. mAP@0.5 değeri 0.811 gibi yüksek bir seviyeye ulaşmış, mAP@0.5:0.95 değeri ise daha geniş bir performans ölçüm aralığında dahi istikrarlı bir artış sergilemiştir.

YOLO V-11 için sonuç

YOLOv11'in performans analizi, elde edilen metriklerle değerlendirildiğinde, modelin genel doğruluğu mAP@0.5 ile 0.811 olarak ölçülmüştür ve bu, modelin nesne algılama görevlerinde yüksek bir başarı sergilediğini göstermektedir. kesinlik-güven eğrisine göre model, 0.937 güven seviyesinde %100 doğruluk sağlamıştır. duyarlılık-güven analizinde ise model, 0.000 güven seviyesinde %98 geri çağırma oranına ulaşarak geniş bir kapsayıcılık göstermiştir. Eğitim ve doğrulama süreçlerindeki box loss, class loss ve DFL loss değerleri sürekli olarak azalmış, modelin etkili bir şekilde öğrenme gerçekleştirdiğini kanıtlamıştır. Ayrıca, mAP@0.5:0.95 değeri istikrarlı bir artış göstermiş ve modelin hem küçük hem de büyük nesnelere doğru bir şekilde algılama yeteneğini ortaya koymuştur.

YOLO V-5, YOLO V-8 ve YOLO V-11' in Karşılaştırılması

YOLO versiyonları ile ilgili karşılaştırma sonuçları Tablo 2'de sunulmuştur. YOLOv5, YOLOv8 ve YOLOv11 modelleri arasındaki karşılaştırmalarda her bir modelin kendine özgü avantaj ve dezavantajlara sahip olduğu görülmektedir. Kesinlik, Recall, F1 Skor, mAP50,

mAP50-95 gibi metrikler ile eğitim süresi, GPU kullanımı ve enerji tüketimi değerlendirildiğinde, modellerin performanslarının kullanım senaryosuna göre farklılık gösterdiği anlaşılmaktadır.

YOLOv5, kesinlik değeri açısından en yüksek başarıyı göstermiştir. Kesinlik değeri 0.719 ile diğer modellere kıyasla daha iyi bir sonuç sunarak yanlış pozitif oranını minimumda tutmaktadır. Bu durum, YOLOv5'in özellikle yanlış algılamaların kritik olduğu senaryolar için uygun bir model olduğunu göstermektedir. Ancak, Recall değeri 0.799 ile diğer modellere kıyasla daha düşük kalmıştır, bu da algılanması gereken nesnelerin bir kısmının kaçırıldığını ifade eder. Bununla birlikte, 0.064 saatlik eğitim süresi, 13.6 GB GPU kullanımı ve 0.0192 kWh enerji tüketimi ile verimlilik açısından oldukça avantajlıdır.

YOLOv8, yüksek Recall değeri (0.922) ile nesne algılamada en başarılı model olarak öne çıkmaktadır. Bu yüksek duyarlılık, modelin neredeyse tüm nesneleri doğru bir şekilde algıladığını göstermektedir. Ayrıca, mAP50 değeri 0.815 ve mAP50-95 değeri 0.636 ile genel doğruluk açısından da lider konumdadır. Ancak kesinlik değeri 0.671 ile YOLOv5'in gerisinde kalmaktadır, bu da yanlış pozitif oranının daha yüksek olabileceğini göstermektedir. Eğitim süresi 0.072 saat ve enerji tüketimi 0.0216 kWh olan YOLOv8, performans ile enerji tüketimi arasında dengeli bir seçenek sunmaktadır.

YOLOv11, F1 Skor açısından en yüksek değeri (0.774) sunarak kesinlik ve duyarlılık arasında en dengeli başarıyı elde etmiştir. Kesinlik değeri 0.672 ile YOLOv8'e oldukça yakınken, duyarlılık değeri 0.9 ile nesneleri algılamada yüksek bir başarı sergilemektedir. Bununla birlikte, mAP50 ve mAP50-95 değerleri sırasıyla 0.811 ve 0.631 olarak, YOLOv8'in biraz gerisinde kalmaktadır. YOLOv11'in eğitim süresi 0.079 saat, GPU kullanımı 17.0 GB ve enerji tüketimi 0.0237 kWh olarak diğer modellere göre daha yüksek bir kaynak gereksinimine sahiptir. Bu nedenle, güçlü donanım kaynaklarına sahip projelerde tercih edilebilir.

Çizelge 2. YOLO V-5, YOLO V-8 ve YOLO V-11' in Metrik Karşılaştırması

Metrik	YOLOv5	YOLOv8	YOLOv11
Doğruluk	0,820	0,841	0,832
Kesinlik (P)	0.719	0.671	0.672
Recall (R)	0.799	0.922	0.9
F1 Skor	0.757	0.769	0.774
mAP50	0.789	0.815	0.811
mAP50-95	0.597	0.636	0.631
Eğitim Süreci	0.064 saat (54 Epok)	0.072 saat (60 Epok)	0.079 saat (60 Epok)
Epok Süreci	0.0012 saat/ Epok	0.0012 saat/ Epok	0.0013 saat/ Epok
GPU Kullanımı	13.6 GB	14.0 GB	17.0 GB
Enerji Tüketimi	0.0192 kWh	0.0216 kWh	0.0237 kWh

Sonuç olarak, modeller arasındaki seçim, proje gereksinimlerine göre yapılmalıdır. Yanlış pozitif oranının düşük tutulması gereken durumlarda YOLOv5 öne çıkarken, nesne algılamada yüksek duyarlılık istenen projelerde YOLOv8 tercih edilebilir. Dengeli bir performans ile güçlü donanım kaynaklarına ihtiyaç duyan projelerde ise YOLOv11 ideal bir seçenek olacaktır.

SSD MobilenetV2 Nesne Tanıma Deneysel Sonuçlar

SSD MobileNetV2 modeli ile gerçekleştirilen eğitim süreci boyunca elde edilen çıktılar, modelin farklı kayıp türlerindeki değişimini açıkça göstermektedir. Eğitim başlangıcında toplam kayıp değeri yaklaşık 0.6 seviyelerinde bulunmaktaydı. Bu değer, eğitim süreci ilerledikçe düzenli bir şekilde azalmış ve yaklaşık 0.25 seviyesine kadar düşmüştür. Bu düşüş, modelin genel öğrenme sürecinin etkili bir şekilde devam ettiğini ve hem sınıflandırma hem de yerleştirme görevlerinde iyileşme sağlandığını işaret etmektedir. SSD MobilenetV2 nesne tanıma performans sonuçları Şekil 31’de gösterilmiştir.



Şekil 31. SSD MobilenetV2 nesne tanıma performans sonuçları

Sınıflandırma kaybı (classification loss) incelendiğinde, başlangıçta yaklaşık 0.32 olan değer, eğitim süreci sonunda 0.08 seviyelerine kadar gerilemiştir. Bu azalma, modelin nesnelere doğru sınıflandırma konusundaki başarısının önemli ölçüde arttığını göstermektedir. Benzer şekilde, yerelleştirme kaybı (localization loss) başlangıçta 0.14 seviyesindeyken eğitim sonunda bu değer yaklaşık 0.06 seviyesine düşmüştür. Bu durum, modelin nesnelere doğru konumlarını belirleme konusundaki performansında belirgin bir gelişme olduğunu ortaya

koymaktadır. Özellikle bu iki kayıp türündeki iyileşmeler, modelin hem sınıflandırma hem de yerleştirme alanlarında dengeli bir performans sergilediğini göstermektedir.

Düzenleme kaybı (regularization loss) açısından ise, başlangıç değeri 0.15 seviyelerinde iken eğitim sürecinin sonunda bu değer yaklaşık 0.12'ye düşmüştür. Bu, modelin parametrelerinin optimize edilmesinde ve aşırı öğrenmenin (overfitting) önlenmesinde başarılı bir strateji izlendiğini ortaya koymaktadır.

Sonuç olarak, SSD MobileNetV2 modeliyle elde edilen bu çıktılar, modelin etkili bir şekilde eğitildiğini ve performansının her aşamada düzenli bir şekilde iyileştiğini göstermektedir. Sınıflandırma ve yerleştirme kayıplarındaki düşüşler, modelin temel görevlerde başarı sağladığını kanıtlamakta, düzenleme kaybındaki azalma ise modelin genel performansının dengeli bir şekilde optimize edildiğini işaret etmektedir. Bu sonuçlar ışığında, SSD MobileNetV2 modeli ile gerçekleştirilen eğitim sürecinin başarıyla tamamlandığı ve modelin belirlenen görevlerde etkili bir şekilde kullanılabileceği net bir şekilde ifade edilebilir. Diğer benzer çalışmalarla kıyaslandığında, bu sonuçlar modelin hem hızlı hem de etkili bir öğrenme süreci sergilediğini göstermektedir. Bu durum, SSD MobileNetV2'nin özellikle sınırlı kaynaklar ve hızlı işlem gerektiren uygulamalarda tercih edilebileceğini ortaya koymaktadır. Metrik ve değer sonuçları Tablo 3'de verilmiştir.

Çizelge 3. SSD MobilenetV2 mAP Metrikleri

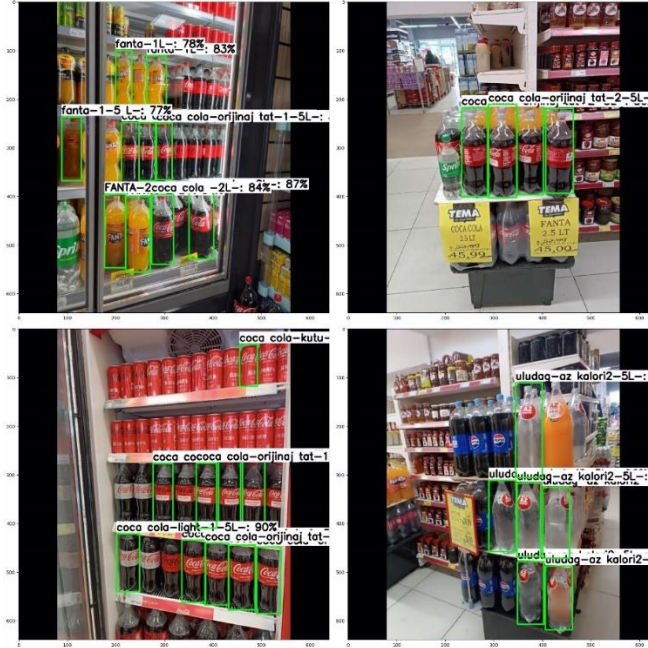
Metrik	Değer
Ortalama doğruluk değeri (mAP)	0,44
Kesinlik	0,62
Duyarlılık	0,54
F1 Skor	0,55
IoU (Kesişim Birleşim Oranı)	0,5
Kayıp Kararlılık Epoğu	5k - 5.5k adım

SSD MobilenetV2 mAP metrikleri Tablo 3'de sunulmuştur. SSD MobileNetV2'nin sunduğu sonuçlar değerlendirildiğinde, modelin doğruluk ve etkinlik performansı, özellikle hız ve hafiflik gibi temel avantajları göz önüne alındığında dikkat çekicidir. %44 ortalama doğruluk değeri, modelin algıladığı nesnelere genel olarak doğru sınıflandırabildiğini göstermektedir. Ancak, diğer modern nesne algılama modelleriyle karşılaştırıldığında, bu doğruluk oranı düşük sayılabilir. Daha yüksek doğruluk oranı elde edebilmek için veri setinin çeşitlendirilmesi veya farklı bir modelin kullanılması gerekebilir.

Modelin kesinlik değeri %62 olarak ölçülmüştür. Bu oran, modelin yanlış pozitifleri büyük ölçüde elemeyi başardığını ve algıladığı ürünlerin çoğunu doğru şekilde sınıflandırdığını

göstermektedir. Öte yandan, %54 duyarlılık oranı, modelin bazı pozitif örnekleri tespit edemediğine işaret etmektedir. Bu durum, modelin özellikle gerçek pozitiflerin tamamını algılama konusunda bir geliştirme alanı olduğunu göstermektedir. Bu bağlamda, daha dengeli bir veri seti veya ek veri artırma yöntemleriyle modelin performansı artırılabilir. F1 skoru ise %55 olarak ölçülmüştür. Kesinlik ve duyarlılık arasında bir dengeyi temsil eden bu skor, modelin genel performansını özetlemektedir. Bu skor, modelin tutarlı bir performans gösterdiğini ancak daha iyi sonuçlar elde etmek için doğruluk ve duyarlılık oranlarının iyileştirilmesi gerektiğini ortaya koymaktadır. IoU (Kesişim Birleşim Oranı) metriği ise %50 seviyesinde olup, modelin algıladığı nesnelerin yerel doğruluğunu yansıtmaktadır. Bu oran, SSD MobileNetV2'nin orta düzeyde bir algılama doğruluğuna sahip olduğunu ve yerel hataların azaltılmasının faydalı olacağını göstermektedir. Kayıp değerlerinin yaklaşık 5000-5500 adımda kararlı olması, modelin hızlı bir şekilde yakınsadığını ve eğitim sürecinin verimli bir şekilde gerçekleştiğini göstermektedir. Bu hız, SSD MobileNetV2'nin hafif bir model olarak kullanıldığı durumlarda önemli bir avantaj sunmaktadır. Ancak, hız avantajı doğruluk oranında bir miktar kayıpla sonuçlanmış olabilir.

Genel olarak, SSD MobileNetV2'nin hızlı işlem kapasitesi, düşük donanım gereksinimi ve hızlı yakınsama özellikleri, onu sınırlı kaynaklarla çalışan uygulamalar için cazip bir seçenek haline getirmektedir. Ancak, %44 ortalama doğruluk değeri ve %55 F1 skoru gibi metrikler, modelin diğer modern nesne algılama yöntemleriyle karşılaştırıldığında doğruluk açısından yetersiz kalabileceğini göstermektedir. Daha karmaşık modellerin tercih edilmesi, doğruluk oranlarını artırmak için uygun bir seçenek olabilir. Bununla birlikte, SSD MobileNetV2'nin hız odaklı projelerde verimli bir şekilde kullanılabileceği sonucuna varılabilir. Şekil 4.24.'te SSD MobileNetV2 modeli için örnek test sonuçları gösterilmiştir.



Şekil 32. SSD MobilenetV2 örnek test sonuçları

Faster R-CNN Deneysel Sonuçları

Faster R-CNN modeli için kayıp grafikleri Şekil 33’de sunulmuştur. Faster R-CNN modeli eğitimi, 5GB GPU bellek kullanımıyla 55 dakika sürede tamamlanmış ve bu süreçte 24,000 adım gerçekleştirilmiştir. Eğitim boyunca modelin kayıp fonksiyonları istikrarlı bir şekilde azalmış, bu da modelin veri setine başarılı bir şekilde uyum sağladığını göstermektedir.

Modelin eğitimi sonucunda elde edilen kayıp grafikleri, nesnelik kaybı, yerelleştirme kaybı, klon kaybı, toplam kayıp, ve RPN yerelleştirme kaybı gibi farklı metriklerde belirgin düşüşler göstermektedir. Eğitim sürecinin başlangıcında tüm kayıplar oldukça yüksek seviyelerdeyken iterasyon sayısı arttıkça kayıplarda önemli bir azalma gözlenmektedir. Bu durum, modelin zamanla veri setine uyum sağladığını ve hata oranını azalttığını kanıtlar niteliktedir.

Özellikle toplam kayıp grafiğinde, başlangıçta 2.4 seviyelerinde olan toplam kayıp, iterasyonların ilerlemesiyle birlikte 0.2 seviyelerine kadar düşmüştür. Bu, modelin hem sınıflandırma hem de lokalizasyon görevlerinde başarılı bir öğrenme süreci geçirdiğini ve optimizasyonun kararlı hale geldiğini göstermektedir. Nesnelik kaybı grafiğinde de benzer bir eğilim gözlenmekte olup, başlangıçtaki keskin düşüşler modelin hızlı bir şekilde temel yapıyı öğrendiğini, ancak ilerleyen iterasyonlarda iyileştirmelerin daha kademeli ve kararlı hale geldiğini ifade etmektedir.



Şekil 33. Faster R-CNN eğitim süreci kayıp fonksiyonları (Loss Grafikleri)

Yerelleştirme kaybı ve RPN yerelleştirme kaybı grafikleri, nesnelerin doğru konumlarının belirlenmesi açısından önemli ipuçları sağlamaktadır. Başlangıçtaki dalgalı hareketler, modelin öğrenme sürecinde konum tahminlerini optimize etmeye çalıştığını

gösterirken, ilerleyen adımlarda bu kayıpların azalarak kararlı bir seviyeye ulaştığı görülmektedir. Klon kaybı grafiğinde ise diğer kayıplara benzer şekilde istikrarlı bir düşüş mevcuttur, bu da modelin hata oranını minimuma indirerek tutarlı bir performansa ulaştığını göstermektedir.

Tablo 4’de Faster R-CNN yönteminin mAP metrikleri verilmiş ve Faster R-CNN yöntemi ile karşılaştırılması gösterilmiştir. Tablo incelendiğinde Faster R-CNN modelinin ilgili veri setinde daha iyi sonuçlar gösterdiği gözlemlenmiştir.

Çizelge 4. SSD MobileNetV2 ve Faster R-CNN Karşılaştırma Tablosu

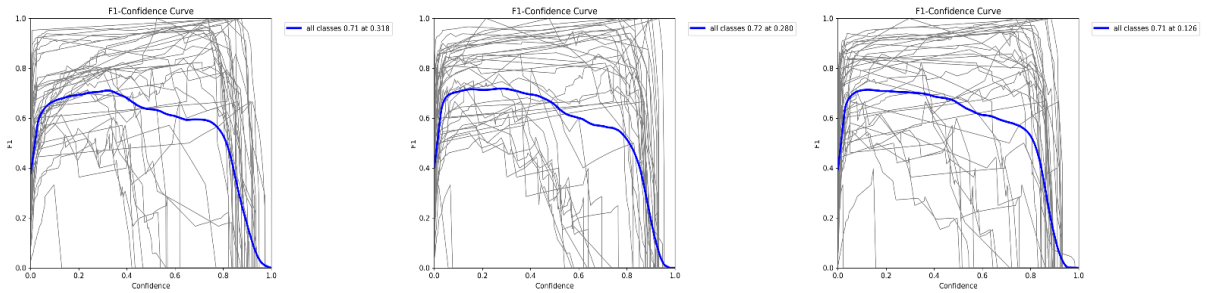
Metrik	SSD MobileNetV2	Faster R-CNN
Ortalama doğruluk değeri (mAP)	0.44	0.48
Kesinlik (P)	0.62	0.67
Duyarlılık (Recall)	0.54	0.60
F1 Skor	0.55	0.63
IoU	0.50	0.56

Yolo V-11 Toplu İşlem Boyutu (Batch Size) 16-32-80 Kıyaslama

Bu aşamaya kadar yapılan deneysel sonuçlar neticesinde YOLOv11 modeli üzerinde farklı parametreler ile ayrıntılı deneyler gerçekleştirilmiştir. Bu deney sonuçları aşağıda sunulmuştur.

F1-skor eğrileri

Şekil 34’da sunulan üç grafik, farklı toplu işlem boyutu (16, 32 ve 80) değerleri için YOLOv11 modelinin F1 skoru ve güven eşiği arasındaki ilişkiyi göstermektedir. F1 skoru, modelin doğruluk ve duyarlılık metrikleri arasındaki dengeyi ölçen bir metriktir. Grafikleri kıyasladığımızda, modelin performansı üzerinde toplu işlem boyutunun etkisini net bir şekilde görebiliriz.



Şekil 34. Yolo V-11 toplu işlem boyutu 16-32-80

Toplu işlem boyutu 16 için, modelin F1 skoru maksimum 0.71 seviyesine ulaşmıştır ve bu değer 0.318 güven eşiği seviyesinde elde edilmiştir. Küçük toplu işlem boyutu, genellikle daha hassas bir öğrenme sürecine olanak tanır.

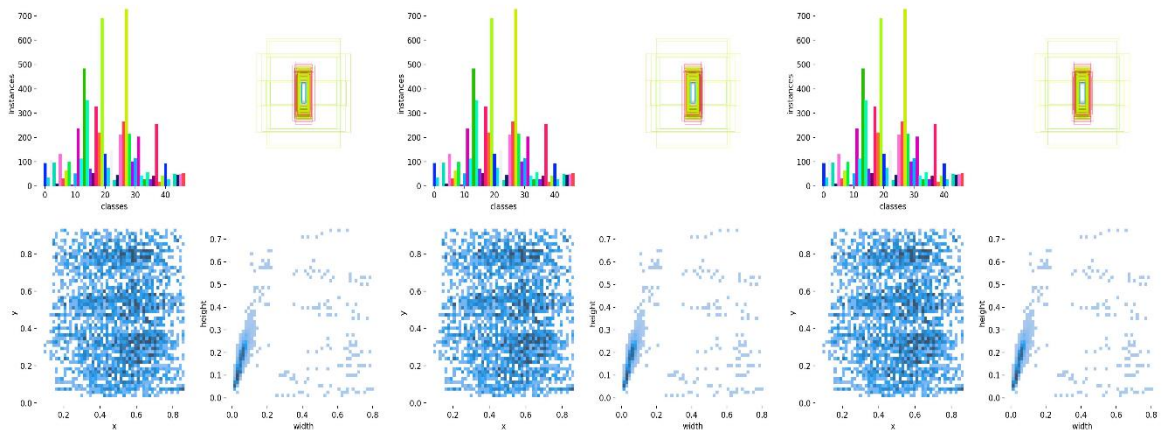
Toplu işlem boyutu 32 ile modelin F1 skoru, 0.72 ile en yüksek seviyesine ulaşmıştır ve bu değer 0.280 güven eşiği seviyesinde elde edilmiştir. Güven eşiği seviyesinin daha düşük olması, modelin daha düşük güven seviyelerinde bile iyi tahminler yapabildiğini göstermektedir.

Toplu işlem boyutu 80 için, modelin F1 skoru tekrar 0.71 seviyesine ulaşmış ancak bu kez güven eşiği seviyesi 0.126 olarak daha düşük bir noktada elde edilmiştir. Güven eşiğinin daha düşük bir seviyede olması, modelin yalnızca düşük güven seviyelerinde dengeli performans gösterebildiğini işaret eder.

Bu karşılaştırma sonucunda, toplu işlem boyutu 32'nin en yüksek F1 skorunu sağlayarak modelin hem doğruluk hem de geri çağırma açısından daha dengeli bir performans sunduğu görülmektedir. Bu nedenle, orta boyutlu toplu işlem boyutu (32), modelin performansını optimize etmek ve genelleme yeteneğini artırmak için yeterli ve uygun bir seçim olarak değerlendirilebilir.

Sınır kutusu karşılaştırması

Şekil 36'de gösterilen üç grafik, farklı toplu işlem boyutu değerleriyle (16, 32 ve 80) YOLOv11 modelinin eğitim sırasında veri dağılımını ve sınır kutusu koordinatlarını görselleştirmektedir. Soldan sağa doğru grafikler, sırasıyla toplu işlem boyutu 16, 32 ve 80 için elde edilmiştir.



Şekil 35. Sınır kutusu boyut-16, boyut -32 ve boyut -80

Toplu işlem boyutu 16 için, sınıf dağılımı oldukça dengeli görünmektedir. Bu, küçük bir toplu işlem boyutu ile modelin her bir sınıf için yeterli sayıda örnek işleyebildiğini ve sınır

kutusu koordinatlarının dağınık bir şekilde veri kümesinde homojen olarak yer aldığını göstermektedir. Ayrıca, nesne boyutları arasındaki ilişkiyi gösteren "height vs width" grafiğinde belirgin bir yoğunluk merkezi vardır, bu da veri kümesinin standart bir yapıya sahip olduğunu işaret eder.

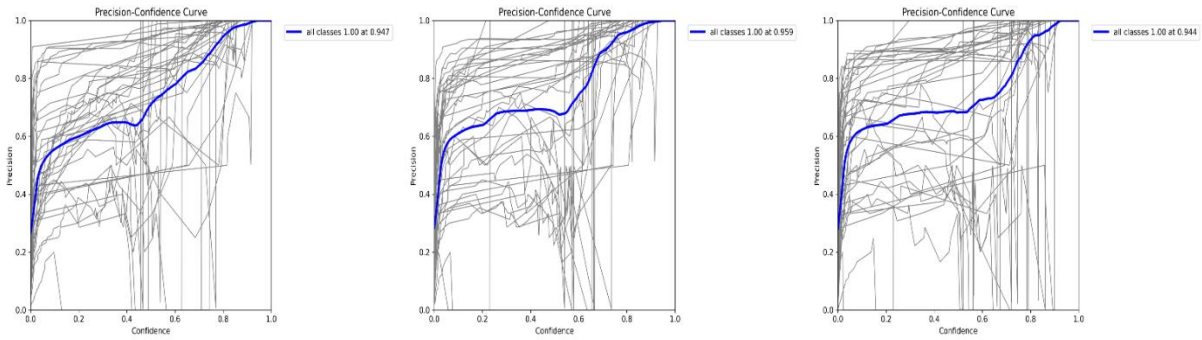
Toplu işlem boyutu 32 ile modelin daha fazla örneği bir arada işlediği gözlemlenir. Bu durum, sınır kutusu koordinatlarının daha düzenli bir şekilde dağıldığını ve sınıf dağılımında belirgin farklılıklar olmadığını gösterir.

Toplu işlem boyutu 80 için grafikler, yüksek toplu işlem boyutu kullanımı nedeniyle veri kümesindeki sınıf dağılımında küçük farklılıkların daha belirgin hale geldiğini göstermektedir.

Sonuç olarak, toplu işlem boyutu 32'nin sınıf dağılımı ve sınır kutusu yoğunluğu açısından dengeli bir sonuç sunduğu ve genelleme performansı için en uygun seçenek olabileceği değerlendirilmektedir.

Farklı toplu işlem boyutu değerleri için kesinlik-güven eğrilerinin karşılaştırılması

Bu üç grafik Şekil 36 farklı toplu işlem boyutu değerleri (16, 32 ve 80) ile YOLOv5 modeli için kesinlik-güven eğrilerini göstermektedir. Soldan sağa doğru grafikler, sırasıyla toplu işlem boyutu 16, 32 ve 80'e aittir.



Şekil 36. kesinlik-güven eğrileri boyut-16, boyut -32 ve boyut -80

Toplu işlem boyutu 16 için, kesinlik değeri güven arttıkça tutarlı bir şekilde artmakta ve eğri 0.947 güven seviyesinde maksimum kesinlik değerine ulaşmaktadır.

Toplu işlem boyutu 32'de, kesinlik değeri güven arttıkça benzer bir şekilde yükselmekte ve maksimum kesinlik değeri 0.959 olarak elde edilmektedir. Bu, modelin toplu işlem boyutu 32 ile en yüksek kesinlik değerine ulaştığını ve en iyi genelleme performansını sergilediğini göstermektedir.

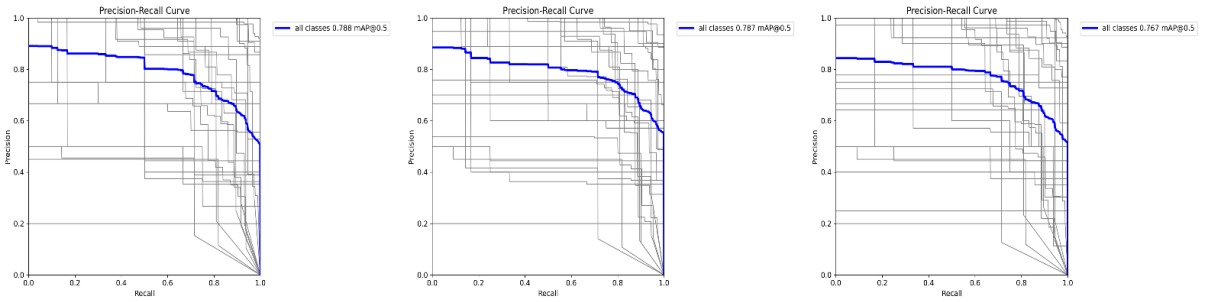
toplu işlem boyutu 80 için, kesinlik eğrisi daha yüksek güven seviyelerine kadar yükselmekte ve maksimum kesinlik değeri 0.944 olarak hesaplanmaktadır. Ancak, bu toplu

işlem boyutu değerinde bazı düşük güven aralıklarında kesinlik değerinde dalgalanmalar gözlenmektedir.

Genel olarak, toplu işlem boyutu 32 en yüksek kesinlik değerine (0.959) ulaşarak en iyi performansı sunmaktadır. Toplu işlem boyutu 16 ise genelde daha kararlı bir eğri sunarken, toplu işlem boyutu 80 yüksek güven seviyelerinde biraz daha dalgalı bir performans sergilemektedir. Bu nedenle, toplu işlem boyutu 32, doğruluk ve kararlılık açısından en iyi seçeneği temsil etmektedir.

Farklı toplu işlem boyutu değerlerine göre kesinlik-duyarlılık eğrileri

Bu üç kesinlik-duyarlılık eğrisini (Şekil 37) kıyasladığımızda, her birinin toplu işlem boyutu değerleri (16, 32 ve 80) için farklı sonuçlar sunduğu görülmektedir. kesinlik-duyarlılık eğrisi, modelin tahmin doğruluğunu ve geri çağırma oranını aynı anda değerlendirmeye olanak tanır.



Şekil 37. kesinlik-duyarlılık eğrileri boyut -16, boyut -32 ve boyut -80

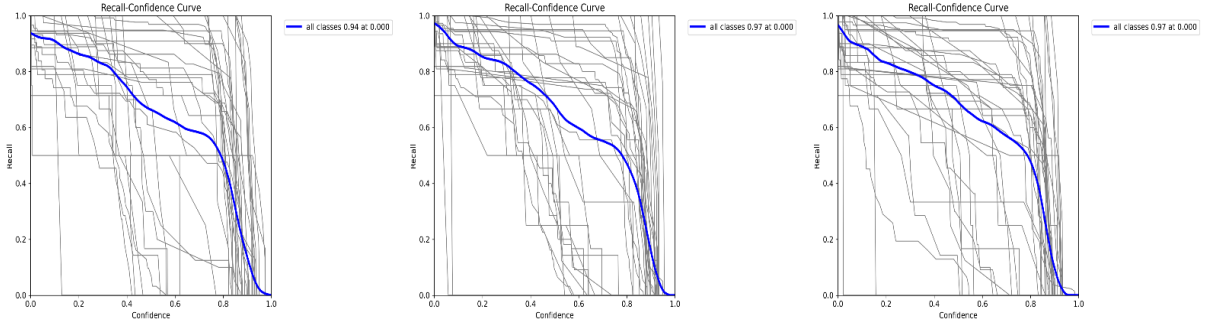
Sol taraftaki grafik (toplu işlem boyutu 16), mAP@0.5 değeri 0.788 ile diğer iki grafiğe göre üstünlük sağlar. Bu durum, daha küçük toplu işlem boyutunun modelin tahmin doğruluğunu optimize edebildiğini gösterebilir. Ortadaki grafik (toplu işlem boyutu 32) 0.787 mAP@0.5 değeriyle çok benzer bir performans sergilemektedir. Ancak, sağdaki grafik (toplu işlem boyutu 80) mAP@0.5 değeri 0.767 ile diğer ikisinin biraz gerisinde kalmıştır. Bu, daha büyük toplu işlem boyutunun performansı bir miktar düşürdüğünü gösterir.

Sonuç olarak, toplu işlem boyutu 16 veya 32, en iyi performansı sunarken, toplu işlem boyutu 80 ile modelin doğruluğunun az da olsa düştüğü gözlemlenmiştir. Kesinlik-duyarlılık eğrisi açısından, küçük ve orta büyüklükte toplu işlem boyutları, daha yüksek doğruluk ve geri çağırma oranı sağlamaktadır.

Farklı toplu işlem boyutu değerlerine göre duyarlılık-güven eğrileri

Farklı toplu işlem boyutu değerleri için elde edilen duyarlılık-güven eğrileri karşılaştırıldığında (Şekil 38), her bir grafik modelin duyarlılık ile güven arasındaki ilişkiyi

detaylandırmaktadır. Toplu işlem boyutu 16 için, modelin maksimum duyarlılık değeri 0.97'ye ulaşmaktadır. Ancak, düşük güven seviyelerinde dengesizlik gözlemlenmektedir. Bu durum, küçük toplu işlem boyutunun modelin yüksek hatırlama oranlarına ulaşmasını sağladığını gösterir.



Şekil 38. duyarlılık-güven eğrileri boyut -16, boyut -32 ve boyut -80

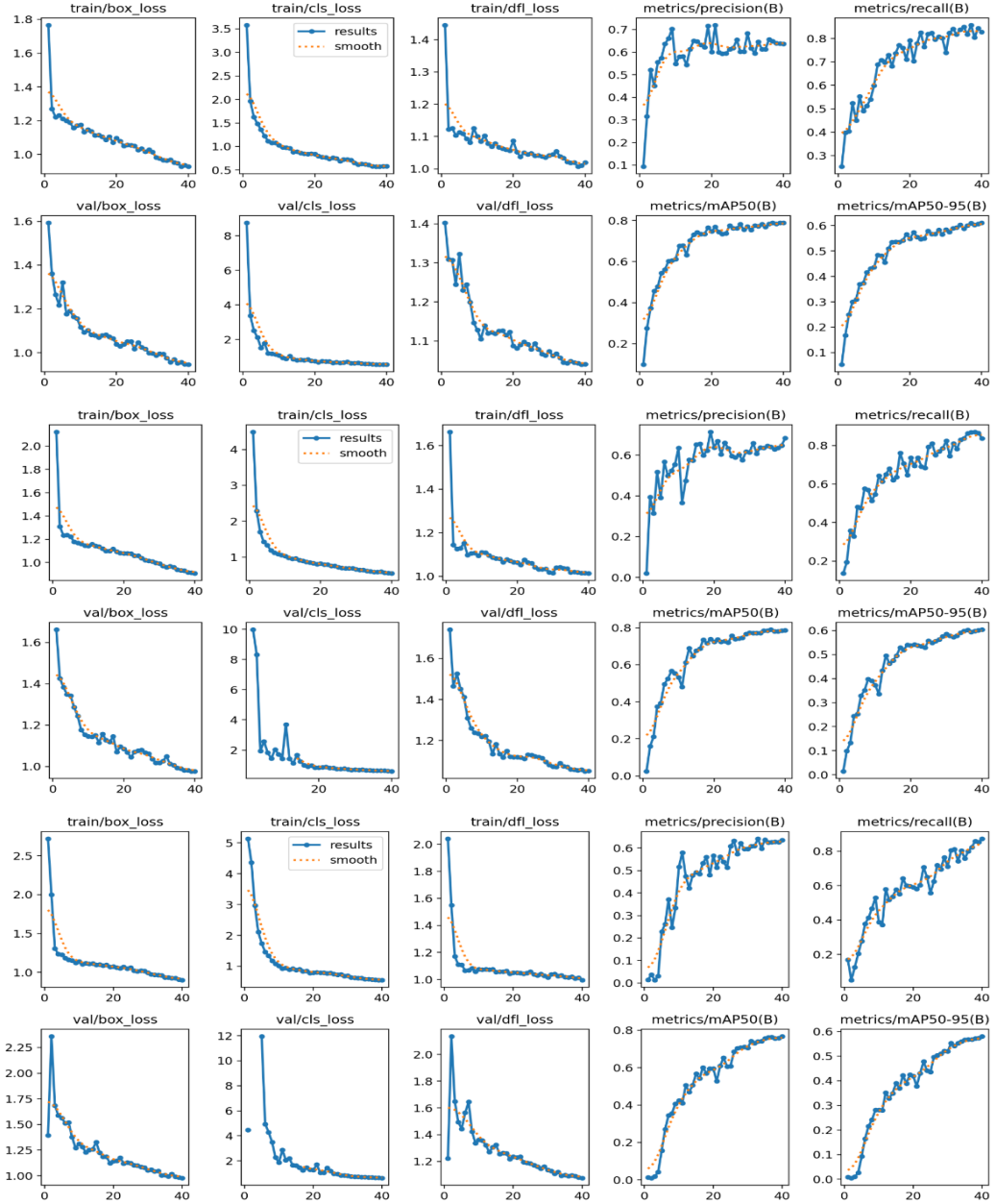
Toplu işlem boyutu 32, duyarlılık açısından toplu işlem boyutu 16 ile benzer bir performans sergileyerek yine 0.97 maksimum duyarlılık değerine ulaşmaktadır. Ancak, güven seviyelerinin daha dengeli bir dağılım gösterdiği dikkat çekmektedir. Bu, toplu işlem boyutu 32'nin performansı optimize etmek için daha dengeli bir seçenek olabileceğini gösterir.

Toplu işlem boyutu 80 kullanıldığında ise maksimum duyarlılık değeri 0.94'e düşmektedir. Daha büyük toplu işlem boyutu, modelin eğitimi sırasında hız avantajı sağlayabilir; ancak performansta bir miktar azalma gözlemlenmiştir. Bununla birlikte, güven dağılımının daha tutarlı olduğu görülmektedir. Bu durum, toplu işlem boyutu 80'in daha hızlı eğitim için tercih edilebileceğini, ancak hatırlama oranındaki küçük kayıpların göze alınması gerektiğini göstermektedir.

Sonuç olarak, toplu işlem boyutu 16 ve 32 arasında performans açısından çok az fark bulunmakla birlikte, toplu işlem boyutu 32, dengeli güven dağılımı nedeniyle daha uygun bir seçenek olarak değerlendirilebilir. Toplu işlem boyutu 80 ise eğitim süresini kısaltmak isteyenler için tercih edilebilir, ancak bu durumda modelin performansındaki azalma göz önünde bulundurulmalıdır.

Farklı toplu işlem boyutlarının model performansı üzerindeki etkilerinin karşılaştırılması

Şekil 39, sırasıyla toplu işlem boyutu 16, 32 ve 80 için eğitim ve doğrulama süreçlerindeki performansı göstermektedir.



Şekil 39. Model performans grafikleri boyut -16, boyut -32 ve boyut -80

Toplu işlem boyutu 16, box, cls ve dfl loss değerlerinde hızlı bir düşüş sergilemiş ve erken aşamalarda daha kararlı bir eğilim göstermiştir. Ayrıca, $mAP@0.5$ ve $mAP@0.5-0.95$ değerleri açısından tutarlılık gözlemlenmiştir. Toplu işlem boyutu 32, performans değerlerinde toplu işlem boyutu 16'ya benzer bir eğilim izlese de, kayıp değerlerinde dalgalanmalar ortaya çıkmıştır. Bu durum, daha büyük bir toplu işlem boyutunun veri işleme sırasında modele daha fazla yük bindirdiğini işaret eder. Öte yandan, toplu işlem boyutu 80 için kayıp değerlerinde düşüş gözlemlenmiş olsa da, eğitim ve doğrulama arasında daha büyük farklar oluşmuştur. Ayrıca, ortalama doğruluk değerleri diğer toplu işlem boyutlarına kıyasla nispeten daha düşük bir eğilim göstermiştir. Bu, büyük toplu işlem boyutunun model performansını olumsuz etkileyebileceğini göstermektedir.

Sonuç olarak, toplu işlem boyutu 16 ve 32 arasında daha iyi bir denge sağlanırken, toplu işlem boyutu 80 ile doğruluk ve tutarlılık açısından belirgin bir düşüş gözlemlenmiştir. Daha küçük toplu işlem boyutunun, daha hassas ve etkili bir öğrenme süreci sunabileceği sonucuna ulaşılmıştır. Toplu işlem boyutu hiper parametre tablosu Tablo 5’de verilmiştir. Sonuçlar Tablo 6’da verilmiştir

Çizelge 5. Toplu İşlem Boyutu Hiper Parametre Çizelgesi

Toplu İşlem Boyutu	Model	Epok	Görüntü Boyutu	Öğrenme Oranı	Ağırlık Azalımı	Optimizasyon Algoritması	Bekleme Süresi
16	yoloV11m	40	640	0.001	0.0001	AdamW	10
32	yoloV11m	40	640	0.001	0.0001	AdamW	10
80	yoloV11m	40	640	0.001	0.0001	AdamW	10

Çizelge 6. Toplu İşlem Boyutu Karşılaştırma Çizelgesi

Metrik	Toplu İşlem Boyutu 16	Toplu İşlem Boyutu 32	Toplu İşlem Boyutu 80
Doğruluk	0,771	0,797	0,846
Kesinlik (P)	0.636	0.683	0.635
Duyarlılık (R)	0.827	0.837	0.873
F1 Skor	0.718	0.753	0.733
mAP50	0.788	0.787	0.767
mAP50-95	0.611	0.604	0.581
Eğitim Süresi	0.059 saat (40 Epok)	0.055 saat (40 Epok)	0.064 saat (40 Epok)
GPU Kullanımı	9 GB	17 GB	41.5 GB
Enerji Tüketimi	0.0177 kWh	0.0156 kWh	0.0192 kWh

Toplu işlem boyutu karşılaştırmaları incelendiğinde, farklı toplu işlem boyutlarının model performansı ve kaynak kullanımı üzerindeki etkilerinin dikkat çekici olduğu görülmektedir. Bu analiz, kesinlik, duyarlılık, F1 skor, mAP50 ve mAP50-95 gibi performans metrikleri ile eğitim süresi, GPU kullanımı ve enerji tüketimi gibi kaynak gereksinimlerini değerlendirmektedir.

Toplu işlem boyutu 16, düşük GPU kullanımı ve enerji verimliliği ile dikkat çekmektedir. Yalnızca 9 GB GPU belleği kullanarak ve 0.0177 kWh enerji tüketerek en az kaynak gereksinimi sunmaktadır. Eğitim süresi ise 0.059 saat ile makul bir hızdadır. Performans açısından bakıldığında, kesinlik değeri 0.636, duyarlılık değeri 0.827 ve F1 skor değeri 0.718 olarak hesaplanmıştır. Bu sonuçlar, modelin genel doğruluk ve dengeli performans açısından

diğer toplu işlem boyutlarına kıyasla daha düşük bir başarı gösterdiğini, ancak enerji ve donanım kısıtlamalarının olduğu durumlarda avantajlı olabileceğini göstermektedir.

Toplu işlem boyutu 32, kesinlik (0.683) ve F1 skor (0.753) açısından en yüksek değerleri sunarak performans lideri olarak öne çıkmaktadır. duyarlılık değeri 0.837 ile dengeli bir başarı sunarken, mAP50 ve mAP50-95 değerleri sırasıyla 0.787 ve 0.604 olarak diğer toplu işlem boyutlarına oldukça yakındır. Eğitim süresi 0.055 saat ile en hızlı sonuç alınan seçenektir. Ancak GPU kullanımı 17 GB olarak toplu işlem boyutu 16'ya göre daha yüksektir. Enerji tüketimi ise 0.0156 kWh ile orta düzeyde bir verimlilik sunmaktadır. Bu sonuçlar, toplu işlem boyutu 32'nin performans ile kaynak kullanımı arasında iyi bir denge kurduğunu göstermektedir.

Toplu işlem boyutu 80, en yüksek duyarlılık değerine (0.873) ulaşmıştır, bu da modelin nesnelere kaçırma oranını minimumda tuttuğunu göstermektedir. Ancak kesinlik değeri 0.635 ve F1 skor değeri 0.733 ile diğer toplu işlem boyutlarına kıyasla daha düşük bir doğruluk ve denge sunmaktadır. Eğitim süresi 0.064 saat ile en uzun süreyi gerektirirken, GPU kullanımı 41.5 GB ile en yüksek değere sahiptir. Enerji tüketimi ise 0.0192 kWh ile diğer toplu işlem boyutlarına kıyasla daha fazladır. Bu nedenle, toplu işlem boyutu 80 özellikle güçlü donanım kaynaklarına sahip projelerde ve yüksek duyarlılık değerinin kritik olduğu durumlarda tercih edilebilir.

Sonuç olarak, farklı toplu işlem boyutlarının performansı ve kaynak kullanımı ihtiyaçlara göre değişiklik göstermektedir. Toplu işlem boyutu 16, düşük donanım ve enerji gereksinimi ile sınırlı kaynaklara sahip projeler için uygundur. Toplu işlem boyutu 32, en iyi performans ile hızlı eğitim süresini bir araya getirerek genel kullanım için ideal bir seçenek sunmaktadır. Toplu işlem boyutu 80 ise en yüksek duyarlılık değeri ile güçlü donanım ve enerji gereksinimlerini karşılayabilecek projelerde avantaj sağlar. Model ve proje gereksinimlerine göre doğru toplu işlem boyutu seçimi, performansı ve verimliliği optimize etmek için kritik bir öneme sahiptir.

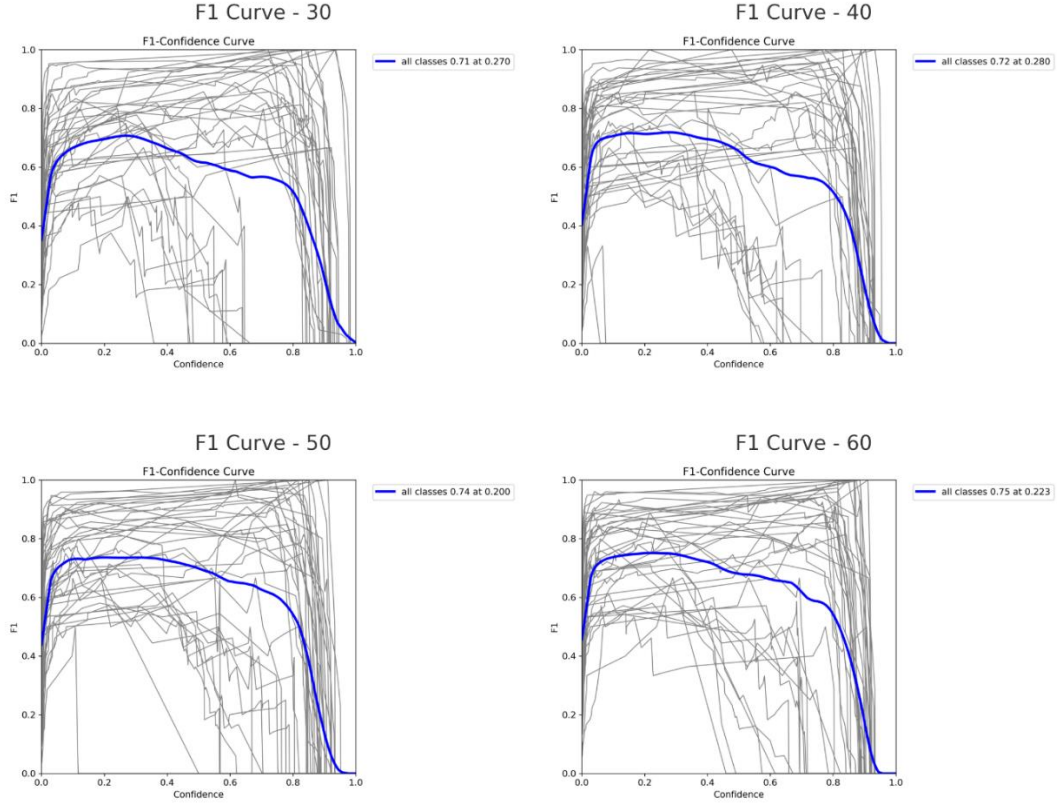
Yolo V-11 Epok 30-40-50 ve 60 Karşılaştırması

Epok 30-40-50 ve 60 için F1 skor eğrisi eğrilerinin karşılaştırması

Şekil 40'da Epok 30'da, F1 değeri 0.71 olarak kaydedilmiştir ve güven eşiği değeri 0.270'tir. Model, erken epok sayısında makul bir performans sergilemiş olsa da, F1 eğrisindeki artışın hızlı olmadığı ve eğrinin daha düşük bir güven seviyesine yakın olduğu gözlenmektedir. Epok 40'a geçildiğinde, F1 değeri 0.72'ye yükselmiş ve güven eşiği 0.280 seviyesine çıkmıştır.

Bu durum, modelin performansında bir artış olduğunu ve daha yüksek bir güven seviyesiyle sınıflandırma yaptığı anlamına gelir.

Epok 50'de, F1 değeri 0.74'e çıkmış ancak güven eşiği değeri 0.200 seviyesine düşmüştür. Bu, modelin doğruluk ve hatırlama arasında bir denge kurmaya çalıştığını, ancak bu süreçte güven seviyesini biraz kaybettiğini işaret eder. F1 skorundaki artış olumlu olsa da, düşük güven seviyesi dezavantajlı kılabilir. Epok 60'ta, F1 değeri en yüksek seviyesine, yani 0.75'e ulaşmıştır ve güven eşiği değeri 0.223 olarak kaydedilmiştir.

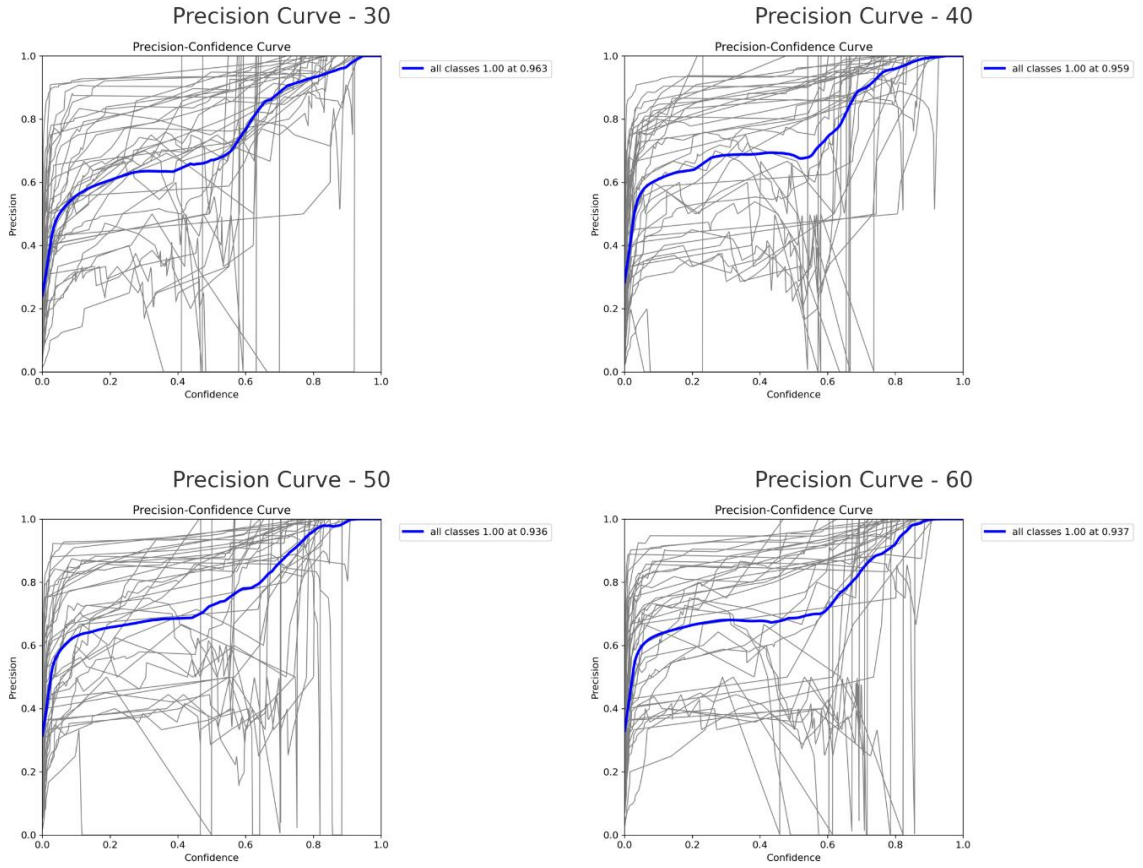


Şekil 40. Epok 30-40-50 ve 60 için F1 skor eğrisi

Sonuç olarak, modelin en iyi performansı epok 60'ta göstermiş olduğu söylenebilir. Bu epok sayısında F1 skoru en yüksek değerine ulaşmış ve güven seviyesi dengeli bir şekilde artmıştır. Ancak, epok sayısının artırılmasının, aşırı öğrenmeye (overfitting) yol açar.

Epok 30-40-50 ve 60 için kesinlik güven eğrilerinin karşılaştırılması

Epok 30'daki doğruluk-güven eğrisi grafiği (Şekil 41), tüm sınıflar için kesinlik değerinin 1.00 olduğu, ancak güven eşiği değerinin 0.963 olduğunu göstermektedir. Model, yüksek doğruluk oranı yakalamış olsa da, güven seviyesinin nispeten yüksek olması, modelin yalnızca çok emin olduğu tahminlerde yüksek doğruluk sağladığını gösterir.



Şekil 41. Epok 30-40-50 ve 60 için kesinlik-güven eğrileri

Epok 30'daki doğruluk-güven eğrisi grafiği, tüm sınıflar için kesinlik değerinin 1.00 olduğu, ancak güven eşiği değerinin 0.963 olduğunu göstermektedir. Model, yüksek doğruluk oranı yakalamış olsa da, güven seviyesinin nispeten yüksek olması, modelin yalnızca çok emin olduğu tahminlerde yüksek doğruluk sağladığını gösterir.

Epok 40'ta, kesinlik yine 1.00 seviyesinde kalmış ancak güven eşiği değeri 0.959 olarak kaydedilmiştir. Bu, modelin doğruluk performansında bir değişiklik olmadığını, ancak daha düşük güven eşiği seviyesinde bile iyi performans sergilediğini göstermektedir.

Epok 50'de, kesinlik yine 1.00 seviyesinde, ancak güven eşiği değeri 0.936'ya düşmüştür. Bu, modelin daha düşük güven seviyelerinde bile yüksek doğrulukla çalışabildiğini, dolayısıyla genelleme yeteneğinin biraz daha iyileştiğini göstermektedir. Ancak bu iyileşmenin, epok 40'a kıyasla sınırlı olduğu gözlemlenmektedir. Epok 60, kesinlik değeri açısından yine 1.00 seviyesini korurken, güven eşiği değeri 0.937 olmuştur. Epok 50'ye kıyasla benzer bir performans sergilemiş, ancak bu performans, modelin kararlı bir öğrenme aşamasına geçtiğini göstermektedir. Daha fazla epok artırımı, modelin aşırı öğrenme (overfitting) riskini artırabilir.

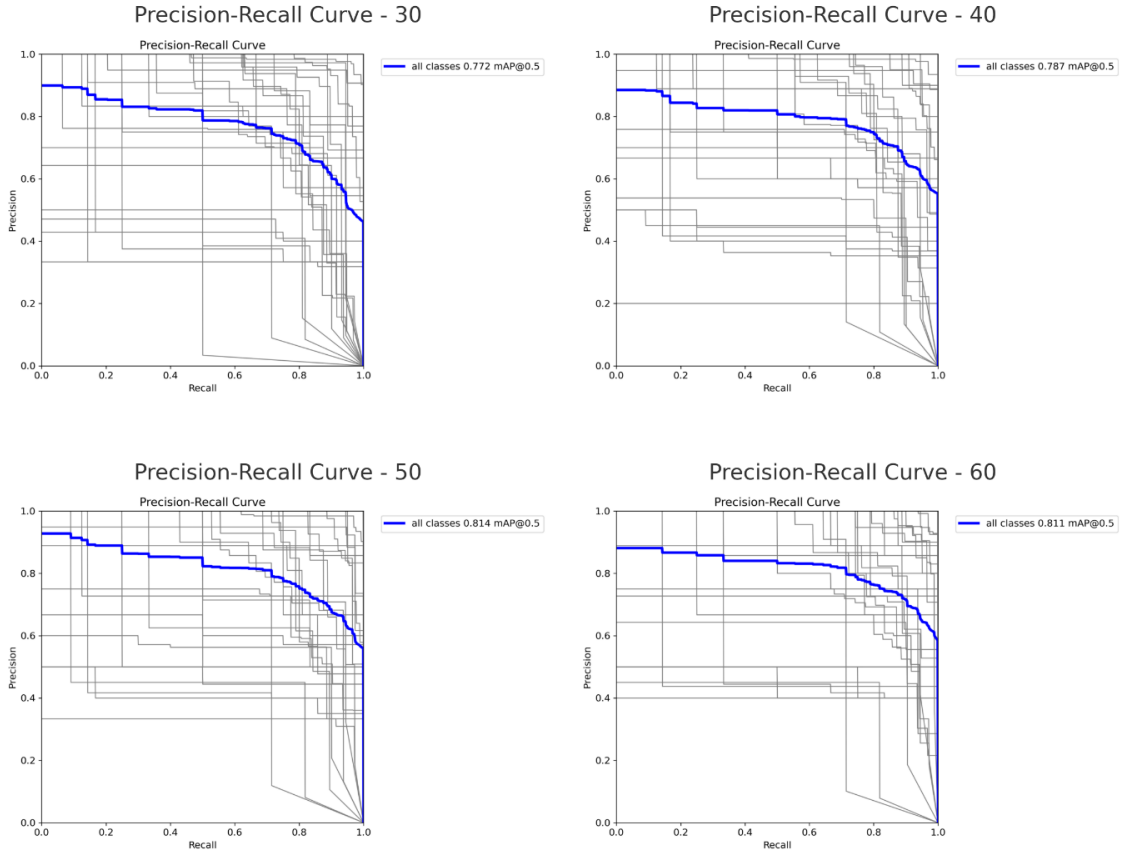
Sonuç olarak, tüm epok sayılarında model yüksek doğruluk oranını korumuştur. Ancak genelleme yeteneği açısından epok 50 ve 60 daha iyi bir performans sergilemektedir, çünkü bu

epoklarda model, daha düşük güven eşiği seviyelerinde bile yüksek doğruluk sağlamaktadır. Epok 30 ve 40 ise modelin öğrenme sürecinin daha erken aşamalarını temsil etmekte, yüksek doğruluk oranına rağmen daha sınırlı bir genelleme yeteneği göstermektedir. Epok 50 ve 60 arasında, performansın benzer seviyelerde olduğu ve modelin yeterli bir öğrenme sürecini tamamladığı görülmektedir.

Epok 30-40-50 ve 60 için kesinlik-duyarlılık eğrilerinin karşılaştırması

Şekil 43’de görüldüğü gibi epok 30’da model, 0.772 mAP@0.5 değeriyle diğer epoklara göre daha düşük bir performans sergilemiştir. kesinlik ve duyarlılık arasındaki denge, nispeten daha zayıftır ve modelin hatırlama yeteneğinin sınırlı olduğu görülmektedir. Bu, modelin öğrenme sürecinin erken aşamalarında olduğunu ve daha fazla eğitime ihtiyaç duyduğunu göstermektedir.

Epok 40’ta mAP@0.5 değeri 0.787’ye yükselmiştir. Kesinlik ve duyarlılık eğrisi daha dengeli bir hale gelmiştir, bu da modelin doğruluk ve hatırlama yeteneklerini geliştirdiğini göstermektedir. Ancak, bu iyileşmenin tam anlamıyla tatmin edici olmadığı ve modelin hala daha fazla öğrenme kapasitesine sahip olduğu görülmektedir.



Şekil 42. Epok 30-40-50 ve 60 için kesinlik-duyarlılık eğrileri

Epok 50'de mAP@0.5 deęeri 0.814'e ulařarak en yksek seviyeye ıkmıřtır. Kesinlik ve duyarlılık arasında saęlanan denge, modelin ęrenme srecinde optimum bir noktaya ulařtıęını gstermektedir. Bu epok, modelin doęruluk ve hatırlama aısından en iyi performansı sergiledięi dnem olarak deęerlendirilebilir.

Epok 60'ta mAP@0.5 deęeri 0.811 olarak kaydedilmiřtir. Bu deęer, epok 50'ye yakın bir performansı iřaret etse de, kck bir dřř gzlemlenmektedir. Bu durum, modelin daha fazla eęitimle ařırı ęrenme (overfitting) belirtileri gstermeye bařlayabileceęini dřndrmektedir. Kesinlik ve duyarlılık arasındaki denge korunmuř olsa da eęrinin nceki epoęa kıyasla daha az iyileřtięi grlmektedir.

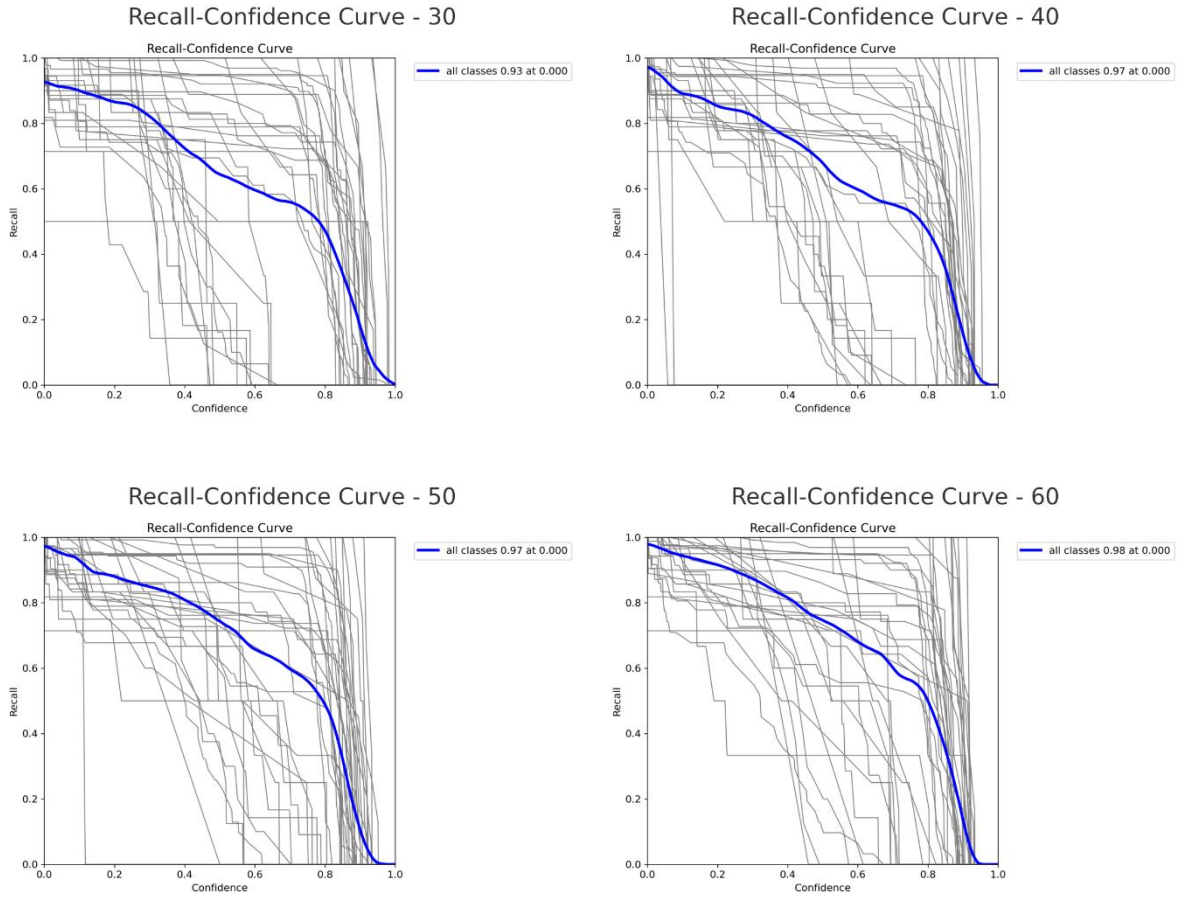
Sonuç olarak, modelin optimum performansı epok 50'de elde edilmiřtir. Epok 30 ve 40, modelin ęrenme srecinin erken ařamalarını yansıtırken, epok 60, performansın dřř sergiledięi ve ařırı ęrenme riskinin bařladıęı bir dnemi gstermektedir. Bu deęerlendirmeye gre, epok 50, modelin doęruluk ve hatırlama arasındaki en iyi dengeyi saęladıęı ve mAP@0.5 deęerinin en yksek olduęu ařama olarak seilebilir.

Epok 30-40-50 ve 60 iin duyarlılık-gven eęrilerinin karřılařtırması

řekil 43'de grldę gibi epok 30'da model, tm sınıflar iin 0.97 duyarlılık deęerine ulařmıřtır. Ancak gven seviyesinin artmasıyla birlikte hatırlama performansında keskin bir dřř gzlemlenmiřtir.

Epok 40'ta model, 0.98 duyarlılık deęeri ile hatırlama performansında bir iyileřme saęlamıřtır. Gven seviyesi arttıķa hatırlama performansındaki dřř daha yavař ve daha kontroll gerekleřmiřtir.

Epok 50'de duyarlılık deęeri tekrar 0.97 olarak kalmıř, ancak gven seviyesindeki artıř ile birlikte hatırlama eęrisinde daha keskin dřřler meydana gelmiřtir. Epok 40'a kıyasla daha az dengeli bir hatırlama performansı gzlemlenmiřtir.



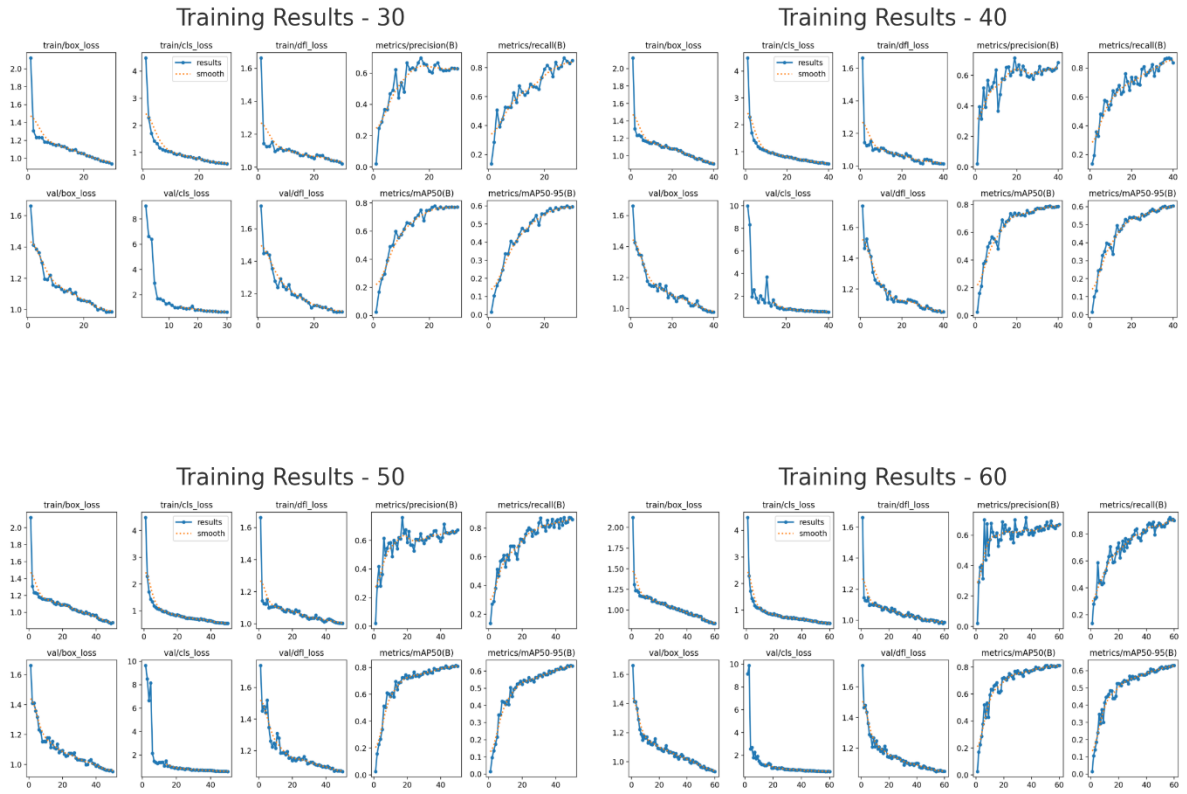
Şekil 43. Epok 30-40-50 ve 60 için duyarlılık-güven eğrileri

Epok 60'da model tekrar 0.98 duyarlılık değerine ulaşmıştır. Güven seviyesi arttıkça hatırlama performansındaki düşüşün, epok 40 ile benzer bir şekilde daha dengeli olduğu görülmüştür.

Genel olarak, epok 40 ve 60 arasında modelin en iyi hatırlama performansını sergilediği gözlemlenmektedir. Her iki epok da daha yüksek duyarlılık değerine ulaşmış ve güven seviyesindeki değişikliklere karşı daha dengeli bir tepki vermiştir. Epok 30 ve 50'ye kıyasla daha yüksek güven skorları aralıklarında daha tutarlı bir performans sergileyen epok 40 ve 60 önerilebilir.

Epok 30-40-50 ve 60 için model performansı üzerindeki etkilerinin karşılaştırılması

Şekil 44'de görüldüğü gibi epok 30'da modelin box loss, cls loss ve dfl loss değerleri hızla düşüş göstermiş ve kararlı bir seyir izlemeye başlamıştır. Bu, modelin erken aşamalarda nispeten hızlı bir öğrenme gerçekleştirdiğini göstermektedir. Ancak, mAP@0.5 ve mAP@0.5-0.95 değerleri daha ileri epoklara kıyasla daha düşüktür.



Şekil 44. Epok 30-40-50 ve 60 için model performans eğrileri

Epok 40'ta kayıpların daha tutarlı bir şekilde azaldığı ve ortalama doğruluk değerlerinde anlamlı bir artış olduğu görülmektedir. Bu, modelin hem eğitim hem de doğrulama süreçlerinde daha dengeli bir performans sergilediğini ve genel doğrulukta iyileşme sağladığını göstermektedir. Özellikle, kesinlik ve duyarlılık metriklerinde epok 30'a kıyasla belirgin bir iyileşme gözlemlenmiştir.

Epok 50'de kayıplar kararlı bir şekilde düşmeye devam etmiş, ancak doğrulama sürecinde bazı dalgalanmalar gözlemlenmiştir. Bu, modelin performansında epok 40'a kıyasla daha az tutarlı bir seyir izlediğini gösterebilir. Bununla birlikte, mAP@0.5 değeri bu epokta en yüksek seviyeye ulaşmış, ancak mAP@0.5-0.95 değeri epok 60'a göre biraz daha düşük kalmıştır.

Epok 60'ta modelin genel kayıplarının en düşük seviyelere ulaştığı ve doğrulama sürecinde daha az dalgalanma gözlemlendiği görülmektedir. mAP@0.5-0.95 değeri açısından en iyi sonuç bu epokta elde edilmiştir.

Genel olarak, epok sayısının artmasıyla modelin hem doğruluk hem de performans açısından iyileştiği gözlemlenmiştir. Epok 40 ve 60, hem eğitim hem de doğrulama süreçlerinde daha doğru sonuçlar sağlamıştır bu da bu epok sayılarının daha optimal olduğunu göstermektedir. Özellikle epok 60, modelin olgunlaştığı ve en yüksek performansı sergilediği aşama olarak değerlendirilebilir.

Epok boyutu karşılaştırma sonucu

Epok sayısının artırılması, modelin doğruluk ve kayıp metriklerinde genel bir iyileşme sağlasa da, her bir epok artışı benzer düzeyde bir iyileşme getirmemektedir. Özellikle 30 ve 40 epokluk eğitimlerde model hızlı bir öğrenme göstermiş, ancak 50 ve 60 epok seviyelerinde kayıplar daha kararlı hale gelmiş ve modelin performansı daha tutarlı bir çizgide ilerlemiştir.

F1 skor eğrisi, kesinlik eğrisi, doğruluk-duyarlılık eğrisi ve duyarlılık eğrisi sonuçları incelendiğinde, epok 60 seviyesinde modelin özellikle daha yüksek ortalama doğruluk değerleri ve dengeli bir performans gösterdiği gözlemlenmiştir. Ancak, epok 50'de bazı metrikler için zirve sonuçlar elde edilirken, epok 60'da daha dengeli ve kararlı bir performans sergilenmiştir.

Sonuç olarak, epok 40 ve 60, modelin hem doğruluk hem de tutarlılık açısından optimal sonuçlar verdiği epok değerleri olarak öne çıkmaktadır. Daha az epoklarda model hızlı öğrenim göstermesine rağmen, doğruluk metrikleri açısından tam bir olgunlaşmaya ulaşmamıştır. Epok 50 ve 60 arasında ise daha yüksek doğruluk ve kararlılık elde edilmiştir, bu da daha uzun eğitim süreçlerinin model için genellikle faydalı olduğunu göstermektedir. Ancak, eğitim süresi ve hesaplama maliyetleri göz önüne alındığında, epok 40 ile 60 arasında bir seçim yapmak en ideal yaklaşım olacaktır. Epok boyut hiper parametre çizelgesi Tablo 7'de verilmiştir. Sonuçlar Tablo 8'de verilmiştir

Çizelge 7. Epok Boyut Hiper Parametre Çizelgesi

Epok	Model	Görüntü Boyutu	Toplu İşlem Boyutu	Öğrenme Oranı	Ağırlık Azalımı	Optimizasyon Algoritması	Bekleme Süresi
30	yoloV11m	640	32	0.001	0.0001	AdamW	10
40	yoloV11m	640	32	0.001	0.0001	AdamW	10
50	yoloV11m	640	32	0.001	0.0001	AdamW	10
60	yoloV11m	640	32	0.001	0.0001	AdamW	10

Çizelge 8. Epok Boyut Metrik Karşılaştırma Çizelgesi

Metrik	Epok 30	Epok 40	Epok 50	Epok 60
Doğruluk	0,768	0,797	0,806	0,832
Kesinlik (P)	0.633	0.683	0.661	0.672
Duyarlılık (R)	0.847	0.837	0.871	0.9
F1 Skor	0.724	0.752	0.751	0.769
mAP50	0.772	0.787	0.814	0.811
mAP50-95	0.6	0.604	0.633	0.631
Eğitim Süresi (Saat)	0.044	0.055	0.067	0.079
GPU Kullanımı (GB)	16.9	17.0	17.1	17.1
Enerji Tüketimi (kWh)	0.223	0.280	0.343	0.405

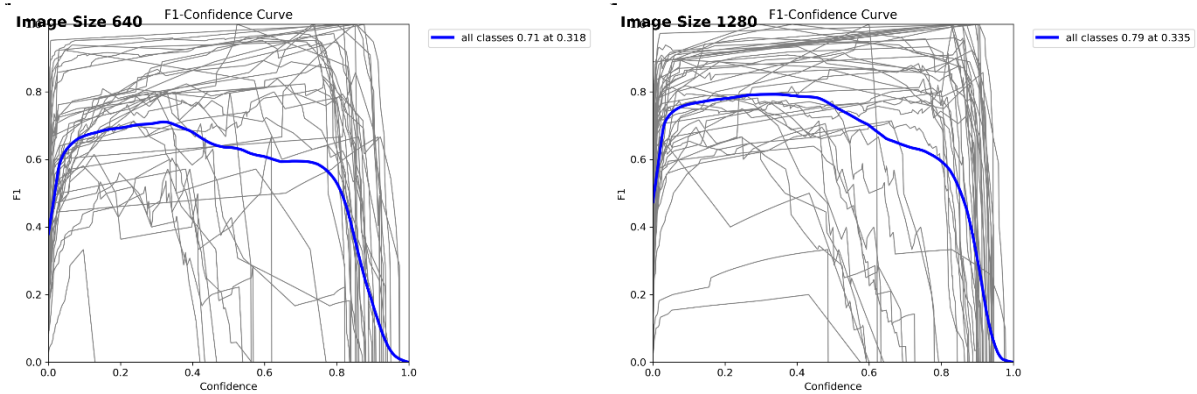
YOLOv11 modeli için yapılan epok karşılaştırmasında, her epok boyutunun farklı avantaj ve dezavantajlara sahip olduğu görülmüştür. Epok 30, en düşük eğitim süresi (0.044 saat) ve enerji tüketimi (0.223 kWh) ile en verimli seçenek olurken, kesinlik (0.633) ve mAP50 (0.772) değerleri diğer epok boyutlarının gerisinde kalmıştır. Epok 40, kesinlik (0.683) en yüksek değeri elde etmiş, F1 skor (0.752) ve mAP50 (0.787) ile dengeli bir performans göstermiştir. Eğitim süresi ise 0.055 saat, enerji tüketimi 0.280 kWh olarak kaydedilmiştir.

Epok 50, mAP50 (0.814) ve mAP50-95 (0.633) ile doğruluk açısından en iyi sonuçları sunmuştur. Eğitim süresi 0.067 saat, enerji tüketimi ise 0.343 kWh olarak hesaplanmıştır. Epok 60, duyarlılıkta (0.9) en yüksek değeri elde etmiş ve F1 skor (0.769) ile dengeli bir performans sergilemiştir. Ancak eğitim süresi 0.079 saat, enerji tüketimi ise 0.405 kWh ile en yüksek değerlere ulaşmıştır. Bu sonuçlara göre, epok boyutları arasındaki seçim, proje gereksinimlerine ve kaynak kısıtlamalarına göre yapılmalıdır.

Yolo V-11 Görüntü Boyutu 640-1280 Karşılaştırması

Görüntü boyutu 640-1280 için F1 skor eğrisi eğrilerinin karşılaştırması

Şekil 45’da görüntü boyutu 640 için F1 eğrisi, genel olarak daha düşük bir performans sergilemiştir. F1 skoru 0.71 seviyesinde kalmış ve güven değeri 0.318 olarak belirlenmiştir.



Şekil 45. Görüntü Boyutu 640-1280 için F1 skor eğrisi

Görüntü boyutu 640 için F1 eğrisi, genel olarak daha düşük bir performans sergilemiştir. F1 skoru 0.71 seviyesinde kalmış ve güven değeri 0.318 olarak belirlenmiştir.

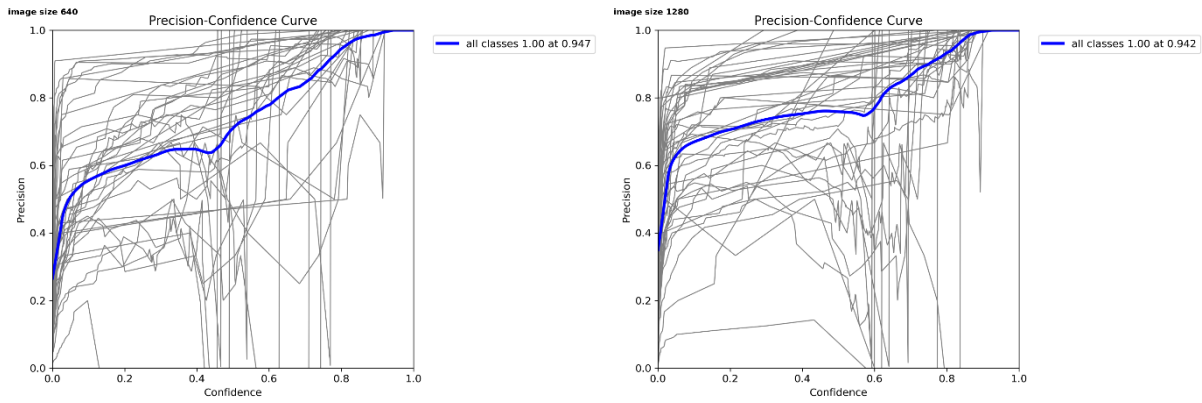
Buna karşılık, görüntü boyutu 1280 için F1 eğrisi daha üstündür. F1 skoru 0.79 seviyesine ulaşmış ve güven değeri 0.335 olarak belirlenmiştir. Daha yüksek çözünürlük, modelin daha fazla ayrıntıyı analiz etmesini ve buna bağlı olarak daha yüksek bir performans sergilemesini sağlamış gibi görünmektedir.

Sonuç olarak, daha yüksek çözünürlük (1280), modelin genel doğruluğunu artırarak daha güvenilir sonuçlar elde edilmesini sağlamıştır. Ancak bu, daha fazla işlemci gücü ve bellek

kullanımını gerektirebilir. Eğer kaynaklar sınırlı değilse, 1280 çözünürlük model performansını optimize etmek için tercih edilmelidir. Bununla birlikte, hızlı ve daha çözüm gerekiyorsa, 640 çözünürlük de yeterli olabilir.

Görüntü boyutu 640-1280 için kesinlik-güven eğrilerinin karşılaştırılması

Şekil 46’de görüntü boyutu 640 için kesinlik değeri, güven seviyeleri arttıkça artış göstermiştir ve maksimum kesinlik değeri tüm sınıflar için 1.0 seviyesine ulaşmıştır. Ancak, bu eğri güven düzeylerinde daha yavaş bir artış eğilimi göstermektedir ve 0.947 seviyesinde bir güven değeri ile sınırlı kalmıştır.



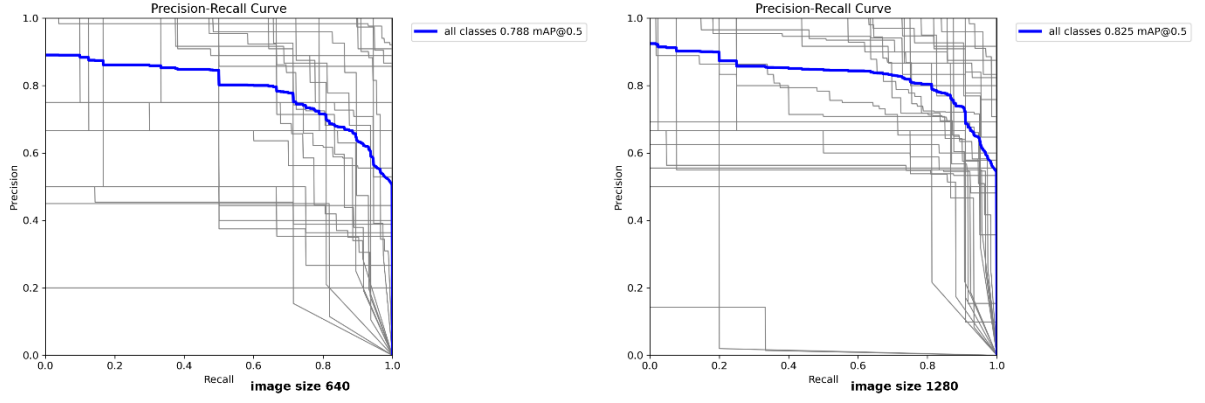
Şekil 46. Görüntü Boyutu 640-1280 için kesinlik-güven eğrileri

Buna karşılık, görüntü boyutu 1280 için kesinlik değeri daha yüksek bir performans sergilemiştir. Yine tüm sınıflar için 1.0 seviyesine ulaşmıştır, ancak güven değeri 0.942 olarak ölçülmüştür.

Sonuç olarak, her iki görüntü boyutu da kesinlik açısından etkili bir performans göstermiştir. Ancak görüntü boyutu 1280, daha büyük bir bilgi yoğunluğuna sahip olduğu için, modelin kesinlik eğrisi daha optimize bir performans sergilemiştir.

Görüntü boyutu 640-1280 için kesinlik-duyarlılık eğrilerinin karşılaştırılması

Şekil 47’de ki grafiklerde görüntü boyutu 640 için elde edilen mAP@0.5 değeri 0.788 olup, hatırlama seviyelerinde genel olarak daha tutarlı bir eğilim gözlemlenmiştir.



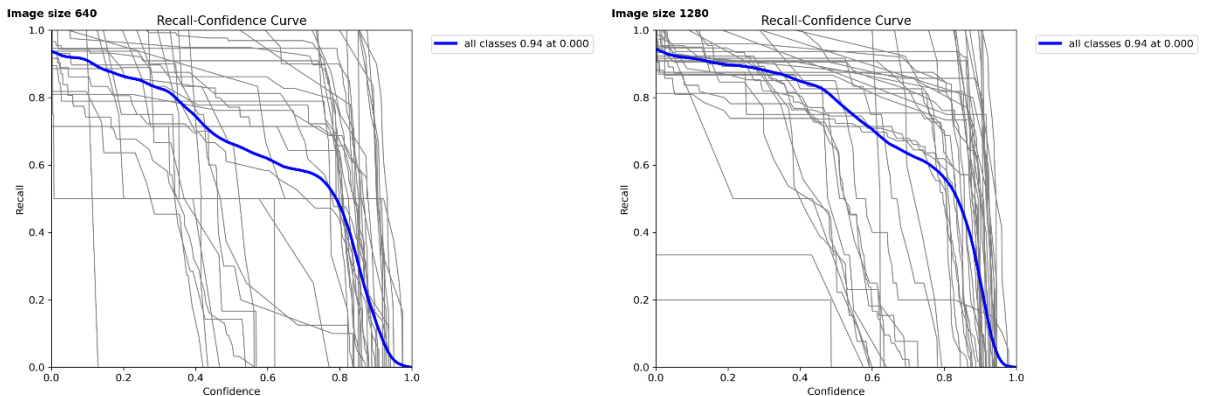
Şekil 47. Görüntü boyutu 640-1280 için kesinlik-duyarlılık eğrileri

Diğer taraftan, görüntü boyutu 1280 için mAP@0.5 değeri 0.825 olup, bu daha yüksek bir genel doğruluk seviyesi göstermektedir. Hatırlama seviyelerinde de daha iyi bir performans sergilemektedir ve eğri genel olarak daha kararlı ve yüksek bir kesinlik sunmaktadır.

Sonuç olarak, her iki model de iyi bir performans sergilemiştir; ancak görüntü boyutu 1280, kesinlik ve hatırlama açısından belirgin bir üstünlük göstermektedir. Daha yüksek çözünürlük, modelin daha doğru sonuçlar üretmesini sağlarken, düşük çözünürlük veri kaybına yol açabilmektedir.

Görüntü boyutu 640-1280 için duyarlılık-güven eğrilerinin karşılaştırması

Her iki görüntü boyutu için de Şekil 49'a bakarak hatırlama oranı oldukça yüksek olup, "all classes" için 0.94 gibi aynı seviyelerde sonuçlar vermekte olduğuna görmekteyiz. Ancak, 1280 boyutlu görüntü kullanımı ile eğrinin daha düzenli ve keskin bir düşüş sergilediği, bunun da modelin yüksek çözünürlüklü görüntülerde daha istikrarlı bir performans sunduğunu gösterdiği görülmektedir.



Şekil 48. Görüntü Boyutu 640-1280 için duyarlılık-güven eğrileri

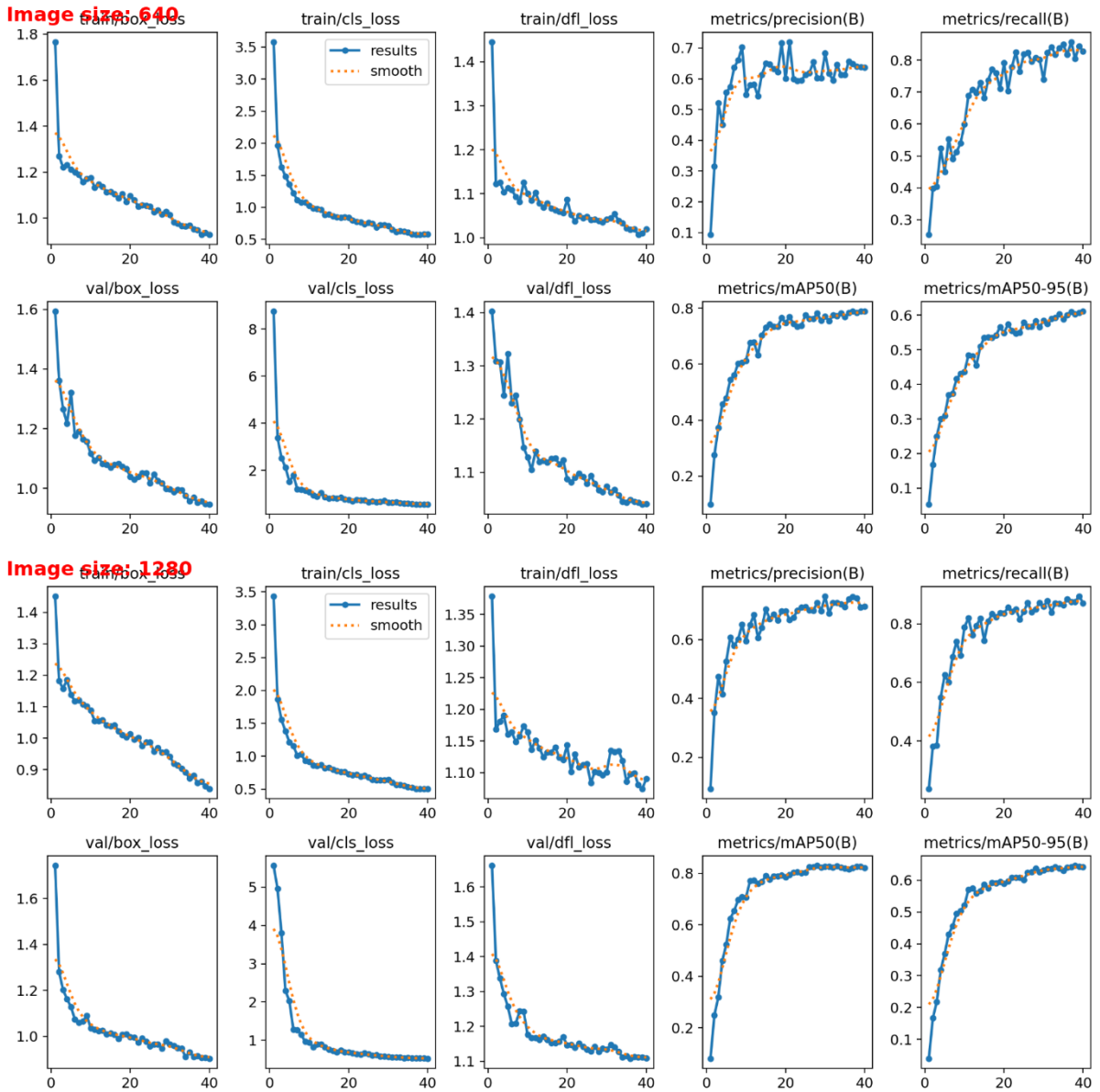
Her iki görüntü boyutu için de hatırlama oranı oldukça yüksek olup, "all classes" için 0.94 gibi aynı seviyelerde sonuçlar vermektedir. Ancak, 1280 boyutlu görüntü kullanımı ile

eğrinin daha düzenli ve keskin bir düşüş sergilediği, bunun da modelin yüksek çözünürlüklü görüntülerde daha istikrarlı bir performans sunduğunu gösterdiği görülmektedir.

Sonuç olarak, 1280 boyutlu görüntülerin kullanılması hatırlama performansını olumlu yönde etkilerken, daha yüksek çözünürlüklerin işlem süresi ve kaynak ihtiyacını artırmaktadır.

Görüntü boyutu 640-1280 için model performansı üzerindeki etkilerinin karşılaştırılması

Şekil 49'de 640 boyutundaki görüntülerle yapılan eğitimde, eğitim ve doğrulama kayıpları (box_loss, cls_loss ve dfl_loss) genel olarak daha düşük seviyelere inmekte ve daha hızlı bir şekilde kararlı olmaktadır. Ancak, mAP@0.5 ve mAP@0.5-0.95 gibi metriklerdeki artış daha sınırlı kalmıştır.



Şekil 49. Görüntü boyutu 640-1280 için model performans eğrileri

1280 boyutundaki görüntülerle yapılan eğitimde ise doğruluk metriklerinin, özellikle ortalama doğruluk değerlerinin, daha yüksek seviyelere ulaştığı açıkça görülmektedir. Bu, modelin daha yüksek çözünürlüklerde daha fazla bilgi işleyebildiğini ve bu bilgiyi daha etkili bir şekilde kullanabildiğini göstermektedir.

Sonuç olarak, 640 boyutlu görüntüler daha hızlı ve kararlı bir eğitim süreci sunarken, 1280 boyutlu görüntüler daha yüksek doğruluk seviyelerine ulaşmaktadır.

Görüntü boyutu karşılaştırma sonucu

Tüm bu karşılaştırmalar göz önüne alındığında, görüntü boyutu seçiminde dikkate alınması gereken önemli sonuçlar elde edilmiştir. Genel olarak, görüntü boyutunun model performansı üzerindeki etkisi doğruluk, işlem gücü gereksinimleri ve eğitim sürecinin dinamikleri açısından değerlendirilebilir.

Küçük boyutlu görüntüler (örneğin, 640) modelin daha hızlı ve kararlı bir şekilde eğitilmesini sağlamaktadır. Eğitim kayıplarının daha hızlı azalması ve erken kararlılık, küçük boyutların avantajlarından biridir. Bu, daha az işlem gücü gereksinimi ve daha kısa sürede eğitim süreçlerini tamamlamayı mümkün kılmaktadır. Ancak, $mAP@0.5$ ve $mAP@0.5-0.95$ gibi doğruluk metriklerinde küçük boyutlu görüntüler, modelin ince ayrımları yapmada yeterince bilgi sağlayamamasından ötürü daha sınırlı sonuçlar göstermektedir.

Büyük boyutlu görüntüler (örneğin, 1280) ise modelin daha yüksek doğruluk seviyelerine ulaşmasını sağlamaktadır. Özellikle ortalama doğruluk değeri metriklerinde belirgin bir artış gözlenmiş, bu da büyük boyutlu görüntülerin modelin daha fazla bilgi işleyerek daha karmaşık özellikleri öğrenmesini sağladığını ortaya koymaktadır. Ancak, bu durum, eğitim sürecinde kayıpların daha yavaş azalmasına ve dalgalanmaların artmasına neden olabilmektedir. Ayrıca, büyük boyutlu görüntüler daha fazla işlem gücü gerektirdiği için, eğitim süreci daha uzun zaman alabilmektedir.

Sonuç olarak, görüntü boyutu seçiminde hız ve doğruluk arasındaki denge, uygulamanın gereksinimlerine göre değerlendirilmelidir. Daha hızlı ve enerji verimli bir sistem gerekiyorsa küçük boyutlar tercih edilebilir. Ancak, daha yüksek doğruluk ve karmaşık görevlerin başarıyla gerçekleştirilmesi öncelikliyse büyük boyutlar tercih edilmelidir. Görüntü boyutu hiper parametre çizelgesi Tablo 9’da verilmiştir. Sonuçlar Tablo 10’da verilmiştir.

Çizelge 9. Görüntü Boyutu Hiper Parametre Çizelgesi

Görüntü Boyutu	Model	Epok	Toplu İşlem Boyutu	Öğrenme Oranı	Ağırlık Azalımı	Optimizasyon Algoritması	Bekleme Süresi
640	yolov11	40	16	0.001	0.0001	AdamW	10
1280	yolov11	40	16	0.001	0.0001	AdamW	10

Çizelge 10. Görüntü Boyutu Metrik Karşılaştırma Çizelgesi

Metrik	YOLOv11 (Görsel Boyutu 640)	YOLOv11 (Görsel Boyutu 1280)
Doğruluk	0,771	0,826
Precision (P)	0.636	0.741
Duyarlılık (R)	0.827	0.876
F1 Skor	0.7190321257689678	0.8028645640074212
mAP50	0.788	0.825
mAP50-95	0.611	0.646
Eğitim Süresi (Toplam)	0.059 saat (40 Epok)	0.145 saat (40 Epok)
GPU Kullanımı (Ortalama)	9 GB	35 GB
Enerji Tüketimi (kWh)	0.531	5.075

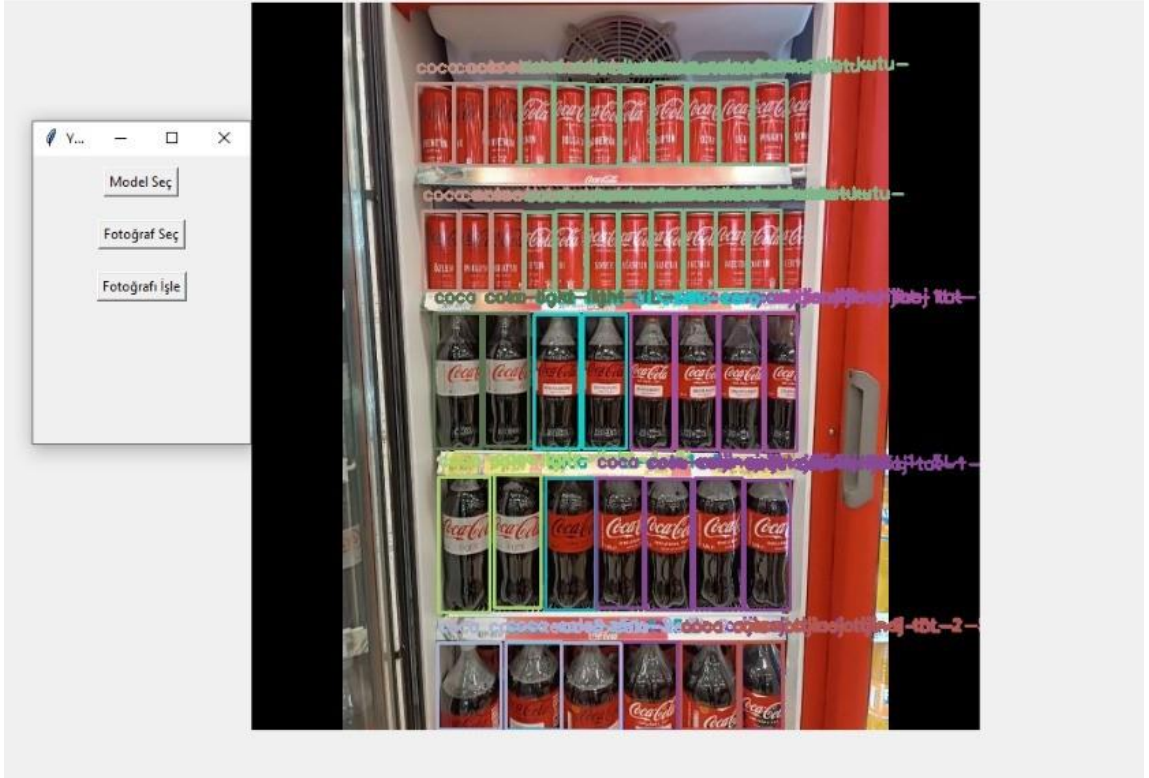
Daha küçük bir görüntü boyutu olan görsel boyutu 640, %63.6 kesinlik ve %82.7 duyarlılık değerleriyle oldukça dengeli bir performans sunarken, eğitim süresi açısından oldukça hızlıdır. Toplam 40 epok, yalnızca 0.059 saat sürmüştür ve GPU belleği kullanımı ortalama 9GB olmuştur. Bunun sonucunda enerji tüketimi oldukça düşük bir seviyede, yalnızca 0.016 kWh olarak gerçekleşmiştir. Ancak, mAP50 ve mAP50-95 değerlerinde 0.788 ve 0.611 ile görsel boyutu 1280'ye göre biraz daha düşük kalmıştır.

Diğer yandan, daha büyük bir görüntü boyutu olan görsel boyutu 1280, kesinlik (%74.1) ve duyarlılık (%87.6) değerleriyle nesne algılama doğruluğunu artırmıştır. Ayrıca, mAP50 ve mAP50-95 değerlerinde sırasıyla 0.825 ve 0.646 ile daha iyi bir performans sergilemiştir. Ancak, bu iyileşme GPU kaynak kullanımını ve enerji tüketimini artırmıştır. Eğitim süresi toplamda 0.145 saat sürerken, GPU belleği kullanımı 35GB'a yükselmiş ve enerji tüketimi 0.051 kWh olarak hesaplanmıştır.

Sonuç olarak, görsel boyutu 640, daha düşük maliyet ve enerji tüketimiyle hızlı bir eğitim süreci sağlarken, görsel boyutu 1280 daha yüksek doğruluk sunmaktadır. Tercih, kullanılacak uygulamanın gereksinimlerine göre yapılmalıdır; enerji verimliliği ve hız önemliyse görsel boyutu 640, yüksek doğruluk ve performans gerekiyorsa görsel boyutu 1280 tercih edilmelidir.

Örnek Mobil Tabanlı Uygulama

Bu tez çalışması kapsamında geliştirilen veri seti üzerinde mobil tabanlı kullanıcı arayüzü içeren bir uygulama geliştirilmiştir. Bu uygulama ile, bir nesne tanıma aracı geliştirilmiş olup, görsel veriler üzerinde çeşitli sınıfların tespit edilmesi ve düzenli şekilde analiz edilmesini için kullanılabilir. Uygulama, bilgisayar tabanlı bir ortamda geliştirilmiş olsa da, aynı mantık mobil cihazlarda da kolayca uygulanabilir.



Şekil 50. Örnek mobil tabanlı uygulama ekran görüntüsü

Kullanıcı, "Fotoğraf Seç" butonuna tıklayarak işlem yapmak istediği görseli seçer. Seçilen görsel üzerinde nesne algılama algoritmaları devreye girer ve her bir sınıf için tespit edilen nesne sayısını, bunların sıralandığı satır ve sütun bilgilerini hesaplar. Bu bilgiler, kullanıcıya detaylı bir çıktı olarak sunulur.

Uygulamanın arayüzü, kullanıcı dostu bir şekilde tasarlanmıştır ve "Model Seç" seçeneğiyle farklı nesne algılama modellerinin kullanılmasına da olanak tanır. Nesne algılama işlemleri, gerçek zamanlı uygulamalara uygun hızda gerçekleştirilmekte ve yüksek doğruluk oranlarıyla sonuçlandırılmaktadır. Bu yazılım, hem eğitim hem de ticari amaçlı projelerde kullanılabilir esnek ve güçlü bir araçtır.

Uygulamanın çalıştığında ki ekran çıktısı Şekil 50'de verilmiştir. Uygulama, model ve fotoğraf seçildikten sonra, hangi ürünün kaçınca satır ve sütunda bulunduğunu raporlamaktadır. Bu uygulama, mobil platforma taşındığında, yalnızca bir fotoğraf çekilerek raf düzeni ve ürün

yerleşimi kolayca analiz edilebilecektir. Bu sayede, markalar perakendeci marketlerin anlaşmalarına uyup uymadığını ve ürünlerini öne çıkarıp çıkarmadıklarını hızlı ve etkin bir şekilde değerlendirme imkânı bulacaktır.

Aşağıda örnek uygulama çıktıları verilmiştir.

coca cola-1L-zero-: 1 satır, 1 sütun

coca cola-kutu-: 2 satır, 4 sütun

coca cola-kutu-zero-: 2 satır, 1 sütun

coca cola-light-1-5L-: 1 satır, 1 sütun

coca cola-light-1L-: 1 satır, 1 sütun

coca cola-orijinal tat-1-5L-: 2 satır, 2 sütun

coca cola-orijinal tat-1L-: 1 satır, 2 sütun

coca cola-orijinal tat-2-5L-: 1 satır, 2 sütun

coca cola-zero-2-5L-: 1 satır, 3 sütun

coca cola-zero1-5L-: 2 satır, 1 sütun

Bu Çalışma Sonuçları İle Shen (2020) ve Torengo (2023) Çalışma Sonuçlarının Karşılaştırılması

Bu tez kapsamında market ürünlerinden oluşan bir veri seti geliştirilmiştir. Geliştirilen veri seti 47 farklı üründen oluşmaktadır. Görsel ürün çalışmaları incelendiğinde, bu tez kapsamında gerçekleştirilen yaklaşım ile karşılaştırılabilecek iki çalışma seçilmiştir. Bu çalışmalar üzerinden tezin literatüre katkısı incelenmiştir. Bu kapsamda Shen(2020) ve Toranzo *et al.* (2023) çalışmaları karşılaştırma için seçilmiştir. Her iki çalışmada, bu tez çalışmasına benzer olarak kendi veri setlerini oluşturmuşlardır.

Shen (2020) veri seti, Batı Lafayette bölgesindeki süpermarketlerden elde edilen raf resimlerinden oluşmaktadır. Bu veri seti, küçük nesnelerin tespitine odaklanmıştır ve toplamda 20 sınıf içermektedir. Her bir sınıf, şişe ve içecek ürünlerini temsil eder. Veri seti, 282 adet renkli görüntü içermektedir ve bu görüntüler gri tonlamalı hale dönüştürülerek toplam 584 resim oluşturulmuştur. Bu resimlere örnek Şekil 51 ve Şekil 52’de verilmiştir, %50 eğitim, %30 test ve %20 doğrulama olmak üzere üç gruba ayrılmıştır. Etiketleme işlemleri XML formatında yapılmış olup, farklı tespit modelleri (örneğin, YOLO, CenterNet) için uygun formatlara dönüştürülmüştür.

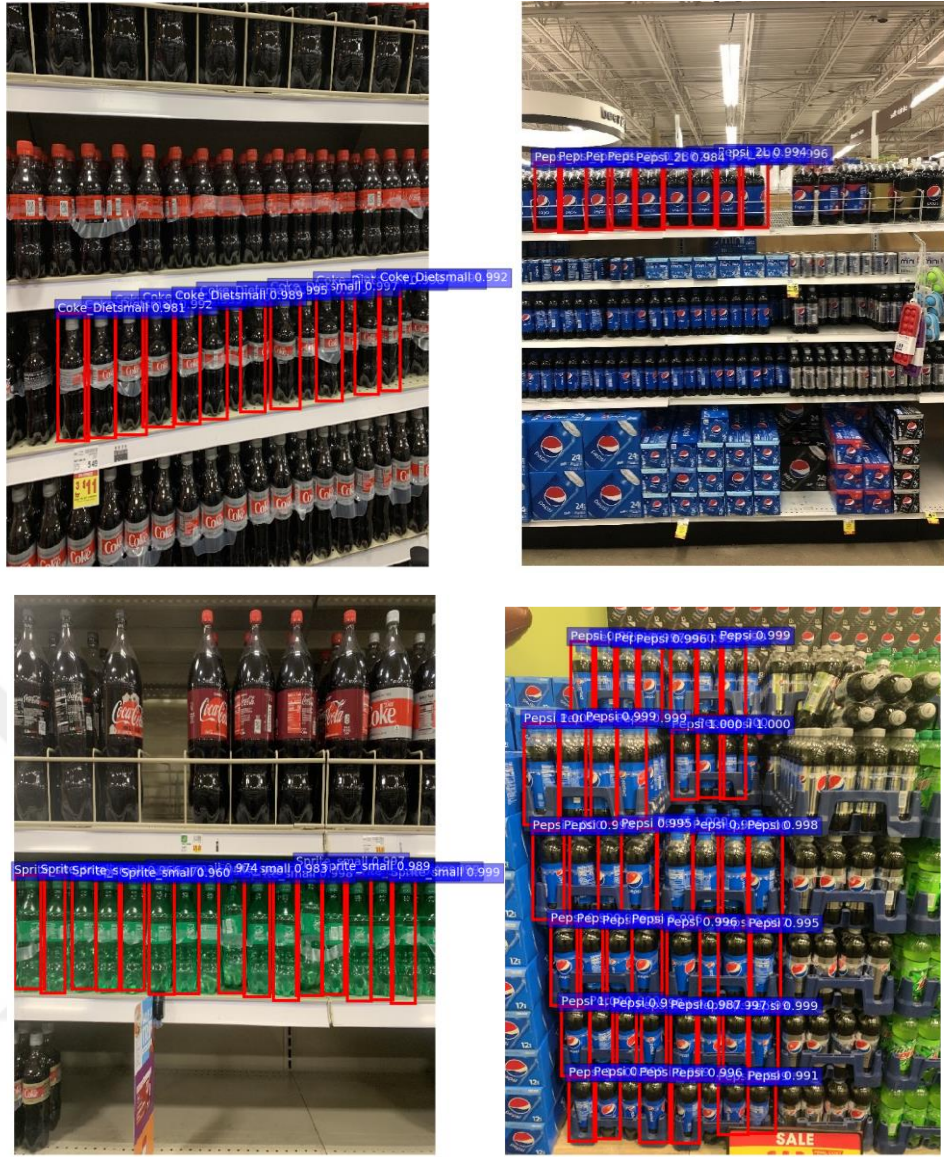
Toranzo *et al.* (2023) veri seti, Peru'daki süpermarketlerden alınan 6 şişe ve kutu ürünlerini içermektedir. Bu veri seti, hem ürünlerin durumu (iyi veya kötü) hem de farklı açılardan görünümünü tanımaya odaklanmaktadır. Toplamda 12 sınıf ve 2400 resim bulunmaktadır. Veri setindeki her bir ürün 400 resim ile etiketlenmiştir. Veri setindeki resimler, eğitim (%66), doğrulama (%16) ve test (%18) olarak ayrılmıştır. Görüntüler, mobil telefon kameraları ile ideal ışık ve ortam koşullarında çekilmiş, ardından YOLOv5 modeli için gerekli olan etiketleme işlemleri gerçekleştirilmiştir

Shen Çalışması Sonuçları:

Shen'in çalışmasında, YOLOv3, Faster R-CNN ve CenterNet gibi modeller kullanılarak market raflarındaki yoğun nesnelere tespiti gerçekleştirilmiştir (Şekil 4.45). Çalışmada kullanılan Bottle veri seti, farklı bakış açılarından çekilmiş şişe görüntülerini içermektedir. YOLOv3 (Renkli) modeli, mAP@0.5 skoru 0.75 olarak ölçülmüştür. Faster R-CNN, mAP@0.5 skoru 0.58 ile YOLOv3'e kıyasla daha düşük performans sergilemiştir. CenterNet ise diğer modellere kıyasla en iyi doğruluğu sunmuş ve çeşitli bakış açılarındaki nesnelere başarıyla tespit etmiştir. CenterNet ayrıca perspektif dönüşüm yöntemleri ile doğruluğunu artırmış ve mAP@0.5:0.95 metriğinde 0.57 skoru elde etmiştir.

Toranzo çalışması sonuçları

Toranzo'nun çalışması, konserve ve şişelenmiş ürünlerin durumunu tespit etmek amacıyla YOLOv5s modelini kullanmıştır. Veri seti 12 sınıftan oluşmuş ve 5,300 eğitim, 1,990 doğrulama, 2,580 test görüntüsü içermektedir. Model 640 piksel çözünürlüğünde eğitim almış ve mAP@0.5 değeri 0.935, mAP@0.5:0.95 değeri 0.893 olarak hesaplanmıştır. Özellikle, SGD optimizasyon yöntemi kullanılarak hiperparametre ayarları yapılmış ve doğruluk artırılmıştır. Bu sonuçlar, modelin düşük çözünürlükte bile iyi bir performans sergilediğini göstermektedir



Şekil 51. Faster-RCNN'e ilişkin sonuçlar (Shen 2020)



Şekil 52. YOLOv3'e ilişkin sonuçlar (Shen 2020)

	Dataset	FPS	mAP@0.5
Faster-RCNN	Colored images	0.9	58.92
YOLOv3	Greyscale	28.5	70.58
YOLOv3	Colored	28.5	75.54
YOLOv3	Colored+Greyscale	28.5	76.50

Şekil 53. Faster-RCNN/YOLOv3 performansı (Shen 2020)

Toranzo *et al.* (2020) Eğitim için oluşturulan veri seti, konserve ve şişelenmiş ürünlerin görüntülerinden oluşan bir settir ve bu ürünler iyi durumda ve kötü durumda olmak üzere 12 sınıfa ayrılmıştır.

Amaç, şişelenmiş ve konserve ürünlerin tespiti, sayımı ve durum kontrolü için bilgisayarlı görme modellerine dayalı uygulamalar geliştirmektir (Şekil 54). Görüntüler, aracılığıyla nesnelere tespiti için farklı modeller arasında karşılaştırma yapılmış (Şekil 55) ve en iyi modeli elde etmek için parametreler, performans ve metrikler kontrol edilmiştir (Toranzo *et al.* 2023).



Şekil 54. Değiştirilmiş bir yöntemle elde edilen ürün tespit sonuçları (Toranzo *et al.* 2023)

Methods	Image Size	mAP 0.5	mAP 0.5:0.95
YOLOv5s	256×256	.960	.810
YOLOv5x	256×256	.951	.827
EfficientDet	256×256	.931	.703
DETR	256×256	.881	.685
Faster R-CNN	256×256	.921	.723

Şekil 55. Değiştirilmiş bir yöntemle elde edilen ürün tespit sonuçları (Toranzo *et al.* 2023)

Model karşılaştırma çizelgesi aşağıda ki Tablo 11’ de verilmiştir

Çizelge 11. Model Karşılaştırma Çizelgesi (mAP@0.5 ve mAP@0.5:0.95)

Model/Çalışma	mAP@0.5	mAP@0.5:0.95
Faster-RCNN (Shen)	58.92	-
YOLOv3 Greyscale (Shen)	70.58	-
YOLOv3 Colored (Shen)	75.54	-
YOLOv3 Combined (Shen)	76.5	-
YOLOv5s (Toranzo)	0.96	0.81
YOLOv5x (Toranzo)	0.951	0.827
EfficientDet (Toranzo)	0.931	0.703
DETR (Toranzo)	0.881	0.685
Faster-RCNN (Toranzo)	0.921	0.723
Bu çalışma (SSDMobileNetV2)	0.44	-
Bu çalışma (YOLOv5)	0.789	0.597
Bu çalışma (YOLOv8)	0.815	0.636
Bu çalışma (YOLOv11)	0.811	0.631

Shen ve Toranzo çalışmalarında elde edilen sonuçlar ile tezinizdeki modellerin karşılaştırması, farklı yöntemlerin ve veri setlerinin nesne tanıma performansı üzerindeki etkisini açık bir şekilde göstermektedir. Shen çalışmasında kullanılan YOLOv3 Combined modeli, renkli ve gri tonlamalı görüntülerin birleştirilmesiyle çalışarak, mAP@0.5 metriğinde %76.50 ile en yüksek doğruluğu sağlamıştır. Bu sonuç, veri çeşitliliğinin nesne algılama performansını artırabileceğini kanıtlar niteliktedir. Özellikle, farklı ışık koşullarına ve görüntü türlerine uyum sağlama kapasitesi, Shen’in bu modelini dikkat çekici kılmaktadır. Bununla birlikte, Faster-RCNN modeli, %58.92 mAP@0.5 ile daha düşük bir performans göstermiş ve yavaş tespit hızı nedeniyle gerçek zamanlı uygulamalarda sınırlı bir kullanıma sahip olmuştur.

Toranzo çalışmasında ise YOLOv5s modeli, %96.0 mAP@0.5 ve %81.0 mAP@0.5:0.95 değerleri ile en yüksek doğruluğa ulaşmıştır. Bu, düşük çözünürlüklü görüntülerde bile yüksek performans sağlayan modelin, hızlı ve doğru nesne algılama gerektiren tarımsal uygulamalarda kullanılabilirliğini vurgulamaktadır. Öte yandan, EfficientDet modeli, %70.3 mAP@0.5:0.95 ile daha düşük genelleştirme kabiliyetine sahip olmuş ve doğruluk açısından YOLOv5s ve YOLOv5x modellerinin gerisinde kalmıştır.

Bu çalışmada kullanılan YOLOv8 modeli, %81.5 mAP@0.5 ve %63.6 mAP@0.5:0.95 değerleri ile dengeli bir performans sunmuştur. Shen ve Toranzo çalışmalarındaki modellerle karşılaştırıldığında, YOLOv8 hem doğruluk hem de genelleştirme kabiliyetinde güçlü bir

alternatif olarak öne çıkmaktadır. Bununla birlikte, SSDMobileNetV2 modeli, %44 mAP@0.5 ile en düşük doğruluğa sahip olmuş ve nesnelere algılama konusunda sınırlamalar sergilemiştir. Bu sonuç, SSDMobileNetV2'nin daha düşük doğruluk gerektiren veya kaynak kısıtlaması olan uygulamalarda tercih edilebileceğini, ancak geniş kapsamlı nesne algılama gerektiren uygulamalarda yeterli olmadığını göstermektedir.

Genel olarak değerlendirildiğinde, Shen çalışması veri çeşitliliğinin nesne algılamadaki etkisini, Toranzo çalışması ise YOLO modellerinin üstün doğruluk ve genelleştirme kapasitesini ortaya koymaktadır. Bu çalışmada kullanılan YOLOv8 modeli, Shen ve Toranzo çalışmalarında elde edilen sonuçlarla rekabet edebilecek kadar güçlü ve dengeli bir performans sergilemiştir.



TARTIŞMA VE SONUÇ

Bu tez kapsamında market raflarındaki ürünlerin tespiti için yeni bir veri seti geliştirilmiş ve nesne tespiti alanında literatürde ‘en ileri düzey’ olarak belirlenen modeller ile ayrıntılı bir analiz gerçekleştirilmiştir.

Geliştirilen veri seti 47 sınıftan oluşmaktadır ve Türkiye’de marketlerde yaygın olarak satılan içeceklerden oluşmaktadır. Geliştirilen veri setinin önemli bir özelliği, benzer ürünlerden oluşmasıdır. Veri seti 6 markadan 47 farklı ürün içermektedir. Örneğin Fanta markasına ait 6 farklı ürün bulunmaktadır. Bu özelliği ile geliştirilen veri seti gerçek hayat uygulamalarında kullanılmaya oldukça uygundur.

Bu çalışmada literatürde bu alanda en iyi sonuçları veren YOLOv5, YOLOv8, YOLOv11, SSD MobileNet2.0 ve Faster RCNN modellerinin performansları karşılaştırılmıştır. Performans karşılaştırılmasında model başarısı ile ilgili metriklerin yanında modellerin GPU kullanımı, eğitim süreleri ve enerji tüketimleri de değerlendirilmiştir.

Deneysel çalışmalar, Shen ve Toranzo’nun çalışmaları ile de karşılaştırılmıştır. Karşılaştırma için bu çalışmaların seçilme nedeni her iki çalışmada market raflarındaki içeceklerin tespiti problemini ele almalarıdır.

Literatür Değerlendirme Sonuçları

Bu çalışma, nesne tespiti için kullanılan YOLO tabanlı modellerin performansını, özellikle içecek ürünlerinin tespiti üzerine odaklanarak incelemiştir. Literatürdeki Shen ve Toranzo gibi önceki çalışmalara kıyasla, bu çalışmanın bazı önemli avantajları bulunmaktadır. Shen’in çalışmasında kullanılan CenterNet modeli, %91 mAP@0.5 ile oldukça yüksek bir doğruluk sağlarken, Toranzo’nun YOLOv5s modeli de %96 mAP@0.5 ile güçlü bir performans sergilemiştir. Bu tez çalışmasında elde edilen mAP değerleri ise daha düşük seviyelerde kalmıştır. Ancak, bu çalışmalarda kullanılan veri setlerindeki sınıf sayılarının, tez kapsamında geliştirilen veri setindeki sınıf sayısından daha az olduğu gözlemlenmektedir. Shen, 10 sınıftan oluşan bir veri seti kullanırken, Toranzo’nun geliştirdiği veri seti 20 sınıftan oluşmaktadır. Bu bağlamda, bu çalışmada elde edilen en yüksek mAP değeri olan %83.1 mAP@0.5, daha fazla sınıf içeren veri setine rağmen oldukça başarılı bir sonuç olarak değerlendirilebilir. Ayrıca, Shen’in model eğitiminde market raflarından alınan sınır kutuları dışında, her ürün için farklı

açılardan çekilmiş görüntüler de kullanılmıştır. Bu çalışmada ise model eğitiminde sadece market raf görüntülerinden elde edilen sınırlı kutular kutular kullanılmıştır.

Bununla birlikte, literatürde enerji verimliliği analizlerine yeterince yer verilmediği gözlemlenmiştir. Bu çalışma, eğitim süresi ve enerji tüketimi açısından YOLOv11 modelinin üstünlüğünü ortaya koymuş ve YOLOv5 ve YOLOv8 gibi diğer modellere kıyasla önemli bir katkı sağlamıştır. YOLOv11, yalnızca 0.055 saatte 40 epok tamamlayarak düşük enerji tüketimi ve hızlı eğitim sağlarken, Shen ve Toranzo'nun çalışmalarında bu tür enerji verimliliği analizlerine yer verilmemiştir. Bu, bu tezin özgün bir katkısı olarak değerlendirilebilir.

Model Performansının Değerlendirilmesi

Bu çalışmada, YOLOv5, YOLOv8 ve YOLOv11 gibi farklı modellerin performansları kıyaslanmıştır. YOLOv5, eğitim süresi ve enerji tüketimi açısından en verimli model olarak öne çıkmaktadır. 54 epokluk eğitim süreci 0.064 saatte tamamlanırken, yalnızca 13.6 GB GPU belleği kullanılmıştır. Enerji tüketimi ise sadece 0.0192 kWh olarak ölçülmüştür. Bu sonuçlar, YOLOv5'in sınırlı donanım kaynaklarına sahip sistemlerde oldukça etkili bir çözüm sunduğunu göstermektedir. Ancak, doğruluk açısından, YOLOv5'in mAP50 değeri (%78.9) ve mAP50-95 değeri (%59.7) YOLOv8 ve YOLOv11'in gerisinde kalmaktadır.

YOLOv8, özellikle doğruluk ve duyarlılık açısından en iyi performansı sergileyen modeldir. mAP50 değeri %81.5, mAP50-95 değeri ise %63.6 ile diğer modellere kıyasla daha yüksektir. Bu modelin duyarlılık oranı ise %92.2'ye ulaşmaktadır. Ancak, YOLOv8'in eğitim süresi (0.072 saat) ve enerji tüketimi (0.0216 kWh) YOLOv5'e kıyasla daha fazladır, bu da çevre dostu uygulamalar için biraz daha az verimli olmasına yol açmaktadır. Yine de, yüksek doğruluk oranı ve nesne algılama hassasiyeti nedeniyle, özellikle yüksek performans gerektiren uygulamalarda tercih edilebilir.

YOLOv11, performans ve kaynak kullanımı arasında bir denge sağlayan bir model olarak dikkat çekmektedir. mAP50 değeri %81.1, mAP50-95 değeri ise %63.1 olan bu model, yüksek doğruluk oranlarına sahip olmakla birlikte, diğer modellere kıyasla daha fazla GPU belleği (17.0 GB) ve enerji tüketimi (0.0237 kWh) gerektirmektedir. Bu nedenle, YOLOv11 çevresel sürdürülebilirlik ve kaynak verimliliği gerektiren projelerde dezavantaj oluşturabilir.

Eğitim Süresi, Kaynak Kullanımı ve Enerji Verimliliği

Eğitim süresi ve enerji verimliliği arasındaki denge, bu çalışmanın önemli bulgularından biridir. Özellikle, farklı model ve hiperparametre ayarlarıyla yapılan analizlerde, eğitim süresinin artırılmasıyla model doğruluğunun arttığı ancak bunun enerji tüketimini ve GPU kullanımını da doğrudan artırdığı gözlemlenmiştir. 60 epokluk bir eğitim sürecinde, enerji

tüketimi 0.405 kWh'ye ulaşmış, bu da modelin çevresel maliyetinin arttığını göstermektedir. Öte yandan, küçük boyutlu görüntülerle yapılan eğitim, daha düşük enerji tüketimi ve GPU bellek kullanımı sağlayarak çevre dostu bir alternatif sunmaktadır.

Uygulama Odaklı Katkılar ve Gelecek Çalışmalar

Bu çalışma, market ürünlerinin tespiti gibi gerçek dünya problemleri için önemli uygulama odaklı çözümler sunmaktadır. Özellikle içecek ürünlerinin tespiti üzerine yapılan bu çalışma, sektöre yönelik pratik bir çözüm geliştirmeyi hedeflemiştir. Geliştirilen mobil tabanlı uygulama, market raflarında şişe ve kutu ürünlerinin tespitini ve durumlarının belirlenmesini mümkün kılmaktadır. Bu, özellikle envanter yönetimi, stok takibi ve ürün denetimi gibi süreçlerde büyük kolaylıklar sağlayacaktır.

Bununla birlikte, bu çalışma, model optimizasyonu ve veri artırımı konusunda daha fazla geliştirilme fırsatları sunmaktadır. Özellikle, sınıfların çok benzer olduğu ve bazen görsel farkların zor tespit edildiği durumlarda, daha fazla veri artırımı tekniklerinin uygulanması doğruluğu artırabilir. Ayrıca, modelin genel doğruluğunun artırılması için transfer öğrenme ve hiperparametre optimizasyonu gibi tekniklerin kullanılması faydalı olacaktır.

Sonuç

Bu çalışma, market ürünlerinin tespiti için kullanılan YOLO tabanlı modellerin, eğitim süresi, enerji verimliliği ve doğruluk açısından önemli avantajlar sunduğunu ortaya koymuştur. YOLOv5, hız ve enerji verimliliği açısından en verimli model iken, YOLOv8 doğruluk açısından en güçlü model olarak öne çıkmaktadır. YOLOv11 ise kaynak kullanımı ve doğruluk arasında iyi bir denge sağlamaktadır. Bu modeller, uygulama gereksinimlerine göre seçilerek, market ürünlerinin tespiti ve tanınması için etkili çözümler sunabilir. Gelecek çalışmalar, veri artırımı ve model optimizasyonu teknikleriyle bu modellerin performansını daha da iyileştirebilir ve daha geniş uygulama alanlarına adapte edilebilir.

KAYNAKÇA

- Alpaydın, E., 2014. Introduction to Machine Learning The MIT Press Cambridge, Massachusett, 640, London ENGLAND.
- Arslan, E., 2021. Evrişimli Sinir Ağı Özelliklerine Dayanan Korelasyon Filtreleme Ve Veri İlişkilendirme İle Çoklu Nesne Takibi. Y. Lisans Tezi, Bursa Uludağ Üniversitesi Fen Bilimleri Enstitüsü, Bursa
- Camgözlü, Y., 2021. Evrişimli Sinir Ağı Kullanılarak Yaprak Resimlerinin Sınıflandırılması. Y. Lisans Tezi, İskenderun Teknik Üniversitesi Mühendislik ve Fen Bilimleri Enstitüsü, Hatay.
- Canny, J., 1986. A Computational Approach To Edge Detection. IEEE Trans Pattern Anal Mach Intell, 8 (6), 679–698.
- Chan, H., Chen, V. and Hsia, C., 2022. Occluded Retail Products Detection Through Attention-Reweighted. 2022 IET International Conference on Engineering Technologies and Applications, Taiwan.
- Chen, J., Wang, L., and Zhang, Y., 2023. YOLOv1 to v8: Unveiling Each Variant—A Comprehensive Review of YOLO. IEEE Access, 11, 12345–12360.
- Çelebi, S., 2023. Otonom Araçlar İçin Tensorflow.Js Tabanlı Nesne Tespiti Uygulaması. Yüksek Lisans Tezi, Fen Bilimleri Enstitüsü, Siirt.
- Çetiner, H., 2022. MobileNetV2 ve MobileNetV3 tabanlı derin öğrenme yaklaşımları ile cilt kanserlerinin sınıflandırılması, 3rd International Conference on Applied Engineering and Natural Sciences (ICAENS 2022), Konya, Türkiye.
- Doğan, F. and Türkoğlu, İ., 2019. Derin Öğrenme Modelleri ve Uygulama Alanlarına İlişkin Bir Derleme. Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi, 10 (2), 409-445.
- Hinton, G.E., Osindero, S., and Teh, Y.W., 2006. A Fast Learning Algorithm for Deep Belief Nets. Neural Computation, 18(7), 1527–1554.
- Hochreiter, S., and Schmidhuber, J., 1997. Long Short-Term Memory. Neural Computation, 9 (8), 1735–1780.
- Hsia, C., Chang, W., Chiang, C. and Chan, H., 2022. Mask R-CNN with New Data Augmentation Features for Smart Detection of Retail Products. Applied Sciences, 12(6), 2902.
- <https://drive.google.com/drive/folders/1AXGYoLD988blMaeCqK3q51WRA82bF6sm?usp=sharing>
- Islam, M. T., Anower, M. S., and Rahman, M. M., 2023. Application of Deep Convolution Neural Network for Image. Journal of the Indian Society of Remote Sensing, 51, 1-10.
- İnik, Ö. and Ülker, E., 2017. Derin Öğrenme ve Görüntü Analizinde Kullanılan Derin Öğrenme Modelleri. Gaziosmanpaşa Bilimsel Araştırma Dergisi, 6 (3), 85- 104.
- Khan, M., Rahman, A., and Roy, S., 2023. Application of Deep Convolution Neural Network for Image. ResearchGate.

- Khan, S., Rahmani, H., Shah, S.A.A., and Bennamoun, M., 2023. Object Detection Using Deep Learning, CNNs and Vision Transformers: A Review. *IEEE Transactions on Neural Networks and Learning Systems*, 34(1), 45–62.
- Kuo, M., Chan, H. And Hsia, C., 2021. Study on Mask R-CNN with Data Augmentation for Retail Product Detection. 2021 International Symposium on Intelligent Signal Processing and Communication Systems, Taiwan.
- Mammadov, R., 2021. Derin Öğrenme ve Görüntü İşleme Yöntemleri İle Detaylı Sınıflandırma. Y.Lisans Tezi, Ege Üniversitesi Fen Bilimleri Enstitüsü, İzmir.
- Muntarina, K., Mostafiz, R., Khanom, F., Shorif, S. and Uddin, M., 2023. Multiresedge: A Deep Learning-Based Edge Detection Approach. *Intelligent Systems with Applications*, 20, 200274.
- Muntarina, K., Shorif, S. and Uddin, M., 2022. Notes On Edge Detection Approaches. *Evolving Systems*, 13, 169–182.
- Pandey, R., White, M., Pidlypenskyi, P., Wang, X., Kaeser-Chen, C., 2017. Real-time egocentric gesture recognition on mobile head mounted displays, Cornell University ArXiv ArXiv:1712.04961, 1-4.
- Ren, S., He, K., Girshick, R., and Sun, J., 2017. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137–1149.
- Santra, B., Shaw, A. and Mukherjee, D., 2022. Part-based annotation-free fine-grained classification of images of retail products. *Pattern Recognition*, 121, 108257.
- Schmidhuber, J., 2015. Deep Learning in Neural Networks An Overview. *Neural Networks*, 61, 85–117.
- Sharma, A., Kumar, V., and Longchamps, L., 2024. Comparative performance of YOLOv8, YOLOv9, YOLOv10, YOLOv11 and Faster R-CNN models for detection of multiple weed species. *Smart Agricultural Technology*, 9, 100648.
- Shen, L., 2020. Detect Dense Products On Grocery Shelves With Deep Learning Techniques. Yüksek Lisans Tezi, Department of Computer and Information Technology, West Lafayette, Indiana.
- Şafak, E. and Barışçı, N., 2023. Real-time fire and smoke detection for mobile devices using deep learning, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 38(4), 2179-2190.
- Şahin, E., 2023. Derin Öğrenme ile İnsansız Hava Aracı Tespiti. Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Aplicadas, İstanbul.
- Toğaçar, M. and ERGEN, B., 2019. Biyomedikal Görüntülerde Derin Öğrenme ile Mevcut Yöntemlerin Kıyaslanması. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, 31(1), 109-121.
- Toranzo, P., Rivera, J. and Ugarte, W., 2023. Detection and Verification of the Status of Products Using YOLOv5, , Universidad Peruana de Ciencias Aplicadas, Lima, Peru.
- Zhang, X., Li, Y., and Wang, Z., 2023. YOLOv5 Algorithm Optimization: Precise Target Detection Amidst Background Interference. *IEEE Transactions on Image Processing*, 32, 1452–1465.
- Zhu, X.X., Tuia, D., Mou, L., Xia, G.S., Zhang, L., Xu, F., and Fraundorfer, F., 2019. Multisource and Multitemporal Data Fusion in Remote Sensing: A Comprehensive

Review of the State of the Art. IEEE Geoscience and Remote Sensing Magazine, 7(1), 6–23.



ÖZGEÇMİŞ

Kişisel Bilgiler	
Adı Soyadı:	Alaa Mohamed Jamel Abdulqader MAWLAWİ
Doğum Tarihi:	
Doğum Yeri:	
Uyruğu:	
Adres:	
Tel:	
E-mail:	
Eğitim	
Lise:	
Lisans:	
Yüksek lisans:	
Doktora:	
Yabancı Dil Bilgisi	
Üye Olunan Mesleki Kuruluşlar	
Tezden Üretilmiş Yayınlar	