

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

**VIDEO ÜZERİNDE DERİN ÖĞRENME İLE NESNE
SANSÜRLÜME**

YÜKSEK LİSANS TEZİ

Yerniyaz BAKHYTOV

**Bilgisayar Mühendisliği Anabilim Dalı
Bilgisayar ve Bilişim Mühendisliği Bilim Dalı**

TEMMUZ 2024

T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

VIDEO ÜZERİNDE DERİN ÖĞRENME İLE NESNE
SANSÜRLÜME

YÜKSEK LİSANS TEZİ

Yerniyaz BAKHYTOV

Bilgisayar Mühendisliği Anabilim Dalı
Bilgisayar ve Bilişim Mühendisliği Bilim Dalı

Tez Danışmanı: Prof.Dr. Cemil Öz

TEMMUZ 2024

Yerniyaz BAKHYTOV tarafından hazırlanan “Video üzerinde derin öğrenme ile nesne sansürlüme” adlı tez çalışması 22.07.2024 tarihinde aşağıdaki jüri tarafından oy birliği/oy çokluğu ile Sakarya Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Bilim Dalı’nda Yüksek Lisans tezi olarak kabul edilmiştir.

Tez Jürisi

Jüri Başkanı: **Prof.Dr. Cemil ÖZ** (Danışman)
Sakarya Üniversitesi

Jüri Üyesi: **Dr.Öğr.Üyesi İsmail ÖZTEL**
Sakarya Üniversitesi

Jüri Üyesi: **Doç.Dr. Süleyman UZUN**
Sakarya Uygulamalı Bilimler Üniversitesi



ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Sakarya Üniversitesi Fen Bilimleri Enstitüsü Lisansüstü Eğitim-Öğretim Yönetmeliğine ve Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesine uygun olarak hazırlamış olduğum “VİDEO ÜZERİNDE DERİN ÖĞRENME İLE NESNE SANSÜRLÜME” başlıklı tezin bana ait, özgün bir çalışma olduğunu; çalışmamın tüm aşamalarında yukarıda belirtilen yönetmelik ve yönergeye uygun davrandığımı, tezin içerdiği yenilik ve sonuçları başka bir yerden almadığımı, tezde kullandığım eserleri usulüne göre kaynak olarak gösterdiğimi, bu tezi başka bir bilim kuruluna akademik amaç ve unvan almak amacıyla vermediğimi ve 20.04.2016 tarihli Resmi Gazete ’de yayımlanan Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 9/2 ve 22/2 maddeleri gereğince Sakarya Üniversitesi’nin abonesi olduğu intihal yazılım programı kullanılarak Enstitü tarafından belirlenmiş ölçütlere uygun rapor alındığını, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun ortaya çıkması halinde doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi beyan ederim.

(22/07/2024)

Yerniyaz BAKHYTOV





Ülkeme ve aileme



TEŞEKKÜR

İlk önce bana bu tez araştırmasını yapmaya fırsat veren Allah'a ve Türkiye Cumhuriyeti'ne teşekkür etmek isterim. Ayrıca, bu süreçte bana destek veren aileme, tez danışmanım Cemil Öz ve tezi destekleyen Türkiye Bursları programına teşekkür ederim.

Yerniyaz BAKHYTOV



İÇİNDEKİLER

Sayfa

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ.....	v
TEŞEKKÜR	ix
İÇİNDEKİLER	xi
KISALTMALAR.....	xiii
SİMGELER	xv
TABLO LİSTESİ	xvii
ŞEKİL LİSTESİ	xix
ÖZET	xxi
SUMMARY.....	xxiii
1. GİRİŞ	1
1.1. Tezin Amacı	2
1.2. Literatür Araştırması	3
2. YAPAY ZEKA VE YOLO KÜTÜPHANESİ	7
2.1. Makine Öğrenmesi	7
2.2. Derin Öğrenme	8
2.3. Yapay Sinir Ağları.....	9
2.4. Evrişimli Sinir Ağları	10
2.5. Kayıp fonksiyonu	11
2.5.1. Stokastik Gradyan İnişi.....	11
2.5.2. Adam.....	12
2.5.2. AdamW.....	13
2.6. YOLO algoritması.....	13
3. MATERYAL ve YÖNTEM.....	15
3.1. Veri Setleri	15
3.2. Modelin Eğitilmesi	16
3.2.1. YOLOv5 modeli	16
3.2.2. YOLOv8 modeli	20
3.2.3. YOLOv9 modeli	22
3.3. Değerlendirme Metrikleri.....	23
3.4. Algoritma.....	24
4. BULGULAR VE TARTIŞMA	27
4.1. Gerekli Kütüphanelerin Kurulması	27
4.2. Modellerin eğitilmesi	27
4.2.1. Hiper-parametre optimizasyonu	30
4.3. Görüntüleri karşılaştırma.....	46
6. SONUÇ VE ÖNERİLER.....	49
KAYNAKLAR.....	51
ÖZGEÇMİŞ.....	55



KISALTMALAR

Adam	: Adaptive Moment Estimation
AdamW	: Adaptive Moment Estimation with Weight Decay
CBAM	: Convolutional Block Attention Module
CNN	: Convolution Neuron Network
CoT	: Contextual Transformer
CPU	: Central processing unit
CSP	: Cross-Stage Partial
ELAN	: Efficient Layer Aggregation Networks
FPN	: Feature Pyramid Networks
FFNN	: Feed-Forward Neural Network
FSPP	: Spatial Pyramid Pooling Module
IoU	: Intersection Over Union
mAP	: Mean Average Precision
MSAM	: Multi-Scale Spatial Attention Module
PAN	: Path Aggregation Network
R-CNN	: Region-based Convolutional Neural Network
RNN	: Recurrent Neural Network
SOTA	: State of the Art
SSD	: Single-Shot Detector
SGD	: Stokastik Gradyan İnişi
YOLO	: You Once Look Once



SİMGELER

- α : Öğrenme katsayısı
 w_t : Mevcut ağırlık
 θ_t : Adım t 'deki model parametreleri
 η : Öğrenme oranı
 λ : Ağırlık azalma faktörü





TABLO LİSTESİ

Sayfa

Tablo 4.1. Eğitilen modellerimizin sonuçları.	27
Tablo 4.2. Modellerinin farklı optimize ediciler ile sonuçları.	31





ŞEKİL LİSTESİ

Sayfa

Şekil 3.1. Sigara veri seti.	16
Şekil 3.2. YOLOv5 modelinin genel mimarisi.	17
Şekil 3.3. Bottleneck, SPPF ve Conv bloklarının mimarileri.	18
Şekil 3.4. C3 bloğunun mimarisi.	18
Şekil 3.5. YOLOv5 modelinin detaylı mimarisi.	19
Şekil 3.6. YOLOv8 algoritmasının mimarisi.	20
Şekil 3.7. YOLOv8 model mimarisinin ayrıntılı çizimi.	21
Şekil 3.8. C2f bloğunun mimarisi.	22
Şekil 3.9. GELAN'ın mimarisi.	23
Şekil 3.10. Algoritma.	25
Şekil 4.1. 50 epok boyunca YOLOv5 eğitimi ile IoU=0,50 olan mAP.	28
Şekil 4.2. 50 epok boyunca YOLOv8 eğitimi ile IoU=0,50 olan mAP.	29
Şekil 4.3. 50 epok boyunca YOLOv9 eğitimi ile IoU=0,50 olan mAP.	29
Şekil 4.4. Önerilen modelin sonuçları.	30
Şekil 4.5. YOLOv5 modeli SGD optimize ediciyle.	32
Şekil 4.6. YOLOv5 modeli Adam optimize ediciyle.	32
Şekil 4.7. YOLOv5 modeli AdamW optimize ediciyle.	33
Şekil 4.8. YOLOv8 modeli SGD optimize ediciyle.	33
Şekil 4.9. YOLOv8 modeli Adam optimize ediciyle.	34
Şekil 4.10. YOLOv8 modeli AdamW optimize ediciyle.	34
Şekil 4.11. YOLOv9 modeli SGD optimize ediciyle.	35
Şekil 4.12. YOLOv9 modeli Adam optimize ediciyle.	35
Şekil 4.13. YOLOv9 modeli AdamW optimize ediciyle.	36
Şekil 4.14. YOLOv5 modelin SGD optimize ediciyle hiper-parametre optimizasyonu grafikleri.	37
Şekil 4.15. YOLOv5 modelin Adam optimize ediciyle hiper-parametre optimizasyonu grafikleri.	38
Şekil 4.16. YOLOv5 modelin AdamW optimize ediciyle hiper-parametre optimizasyonu grafikleri.	39
Şekil 4.17. YOLOv8 modelin SGD optimize ediciyle hiper-parametre optimizasyonu grafikleri.	40
Şekil 4.18. YOLOv8 modelin Adam optimize ediciyle hiper-parametre optimizasyonu grafikleri.	41
Şekil 4.19. YOLOv8 modelin AdamW optimize ediciyle hiper-parametre optimizasyonu grafikleri.	42
Şekil 4.20. YOLOv9 modelin SGD optimize ediciyle hiper-parametre optimizasyonu grafikleri.	43
Şekil 4.21. YOLOv9 modelin Adam optimize ediciyle hiper-parametre optimizasyonu grafikleri.	44
Şekil 4.22. YOLOv9 modelin AdamW optimize ediciyle hiper-parametre optimizasyonu grafikleri.	45

Şekil 4.23. Optimize edilen YOLOv8 modelinin sonucu.	46
Şekil 4.24. YOLOv9 modelinin video üzerinde sansürleme sonuçları.	47



VİDEO ÜZERİNDE DERİN ÖĞRENME İLE NESNE SANSÜRLÜME

ÖZET

Dünya Sağlık Örgütü'ne göre sigara içenler sayısı 1,3 milyara ulaşmış ve bunların 8 milyondan fazlası sigaraya bağlı hastalıklardan ölmüştür. Her yıl bu sayılar sürekli artmaktadır. Bazı ülkeler sigaranın zararlarını önlemek ve sigara kullanımını azaltmak için açık alanlarda sigara içmeyi yasaklamıştır. Buna rağmen sigara içenler sayısı düşmemektedir. Sigara alışkanlığı genellikle erken yaşlarda, çocuk diye bileceğimiz yaşlarda başlamaktadır. Çünkü çevrelerinde, sosyal medya, izlenilen film ve videolarda sigara tüketimini özendirilecek formatta görmektedirler. Tüm bu faktörler sigaraya özentiyi tetikliyor, insanlar sigaraya başlıyor ve zamanla bağımlı hale geliyorlar. Bu çalışmanın temel amacı, günümüz internet videolarının çocuklar üzerindeki olumsuz etkisini azaltmak ve sigara içmeyle ilişkilendirilen nesneleri tanımanın doğruluğunu ve güvenilirliğini artırmak için bilgisayar görme ve derin öğrenme tekniklerini kullanarak güvenli ve eğitici dijital deneyimler oluşturmaktır. Böylece sigara kullanımı açısından daha iyi medya içerik analizini yaparak bu sahnelerin sansürlenmesini sağlayarak sigara tüketimine yönlendiren videoların etkisini azaltmak ve sigara karşıtı kampanyaları desteklemektir.

Çalışmanın önemli bir yönü, sigara algılama görevi için YOLOv5, YOLOv8 ve YOLOv9 modellerinin geliştirilmesi ve eğitilmesidir. YOLO, yüksek doğruluğu ve hızıyla bilinen, nesne algılama görevleri için popüler bir algoritmadır. Bu çalışmada, çeşitli aydınlatma koşulları ve arka planlar da dahil olmak üzere görüntülerdeki sigaraları doğru ve hızlı bir şekilde tespit edebilen bir sinir ağı uygulanmıştır.

Sigara tanıma doğruluğu, Gerçek dünya ortamlarında sigara tespit eden güvenilir izleme sistemleri oluşturmak, sigara ile mücadelede kritik öneme sahiptir. Çalışma da sigara ve kül tablasını içeren iki tane veri seti oluşturulmuştur. Veri setindeki sigara ile kül tablasının konumlarını Roboflow web sitesin üzerinde sınırlayıcı kutu olarak işaretlendi. YOLOv5, YOLOv8 ve YOLOv9 modellerinin performansların yükseltmek için hiper parametre optimizasyonunu kullanarak hazırlanan veri setleri üzerinde eğilmiştir. Modeller kesinlik, duyarlılık, Birleşim Üzerinden Kesişme ve genel ortalama kesinlik puanına göre değerlendirildi. Hiper parametre optimizasyonu, derin öğrenme modelleri için optimum performansa ulaşmada kritik bir rol oynamaktadır. Bu yüzden, çalışmada YOLO modellerine ince ayar yapmak için çeşitli optimizasyon teknikleri kullanılmıştır.

İyi sonuçları elde etmek için seçildiği modellerin hiper parametreleri SGD, Adam ve AdamW gibi çeşitli optimize edicileri kullanılarak optimize edildi. Optimizasyon sonucunda SGD optimizasyonu kullanan YOLOv9 modeli diğer modellerden daha yüksek performansa ulaştığı görülmektedir. Ayar dağılım grafikleri, bu kombinasyonun doğruluk ve yakınsama hızı açısından üstünlüğünü göstererek onu uygulamalarımız için tercih edilen seçenek haline getirdi.

Varsayılan parametre ile ilk eğitilmiş modellerin sonuçları: YOLOv5 %89'a, YOLOv8 %87'ye ve YOLOv9 %88'e ulaşmıştır. Hiper parametre optimizasyonundan

sonra SGD optimize edicinin kullanımının en etkili olduđu kanıtlanmaktadır ve diğerk optimize edicilerle karşılaştırıldığında üstün performans göstermektedir. YOLOv5'in performansı %89'da değişmeden kalırken, YOLOv8 ve YOLOv9'un performansı %93'e çıkmaktadır. YOLOv8 ve YOLOv9, SGD ile optimizasyondan sonra aynı mAP(50) değerine sahip olsa da, YOLOv9 mAP(50-95) metriğinde YOLOv8'den %2 doğruluk oranında daha iyi sonuç vermiştir.

Çalışmamızın önemli bir yönü kül tablası tespitinin görüntülere dâhil edilmesidir. Kül tablası, sigara içmenin önemli bir göstergesidir ve bunun tanınması, sigara içme olaylarının tespitinin doğruluğunu önemli ölçüde artırır. Modellerimize başarılı bir şekilde entegre edilen kül tablasını tespit etmeye yönelik yaklaşımlar geliştirilmiştir. Bu yaklaşım ile genel sonuçlar iyileştirilmiş ve tanıma sistemi daha kararlı ve başarılı hale getirilmiştir.

Sigara ve kül tablasını tespit etmek için bir algoritma geliştirilerek test edilmiştir; bu algoritma %96'lık bir doğruluğa ulaşarak daha iyi sonuçlar vermiştir. Bu algoritma, sigaraları ve kül tablasını başarıyla tanıyarak yanlış pozitif sayısını en aza indirerek eğitilen modellerin ilk sonucundan daha iyi performans göstermiştir.

Çalışmamızın sonuçları, en yeni derin öğrenme modellerinin ve hiper parametre optimizasyonunun kullanılmasının, görüntülerde sigara ve kül tablası tespitinin verimliliğini ve doğruluğunu önemli ölçüde artırabileceğini göstermektedir. Geliştirilen model ve algoritmaların kullanımı, tütünün halk sağlığı üzerindeki zararlı etkilerinin izlenmesi, analiz edilmesi ve önlenmesine yönelik etkili araçlar sunarak tütün kullanımıyla mücadeleye önemli katkılar sağlayacaktır.

SENSOR PROCESSING ON VIDEO WITH DEEP LEARNING

SUMMARY

According to the World Health Organization, the number of smokers reached 1.3 billion last year, and more than 8 million of them died from smoking-related diseases. These numbers are constantly growing every year, so to prevent this, several countries have introduced a ban on smoking in open areas. But still, the number of smokers does not decrease. Today, most smokers start smoking at a young age because they see people smoking in their environment, in films and videos on the Internet, and all this leads to the onset of smoking. Therefore, the main goal of the study is to create a safe and educational digital experience using computer vision and deep learning techniques to reduce the negative impact of modern Internet videos on children and improve the accuracy and reliability of smoking-related object recognition. This will significantly improve media content analysis and support anti-smoking campaigns.

The rise of digital media and the internet has presented new challenges in tobacco control. Children and adolescents are frequently exposed to content that glorifies smoking, which can influence their attitudes towards tobacco use. There is a critical need for technologies that can automatically detect and analyze media content related to smoking-related images to reduce these effects. Therefore, it aims to examine modern technologies in the field of computer vision that allow to automate the process of recognizing smoking-related objects in images and videos, and solutions such as automatic content filtering and providing age-appropriate recommendations are being developed to solve problems such as exposure to harmful content in the video and excessive screen time.

An important aspect of the work is the development and training of YOLOv5, YOLOv8 and YOLOv9 models for the cigarette detection task. YOLO is a popular algorithm for object detection tasks, known for its high accuracy and speed. YOLOv5 is an open-source object detection model that offers significant improvements in speed and accuracy over its predecessors. It is highly optimized for real-time detection tasks, making it suitable for applications requiring quick response times. Compared to previous versions, YOLOv8 is faster and more accurate, and requires fewer parameters to achieve its performance. It utilizes novel techniques in feature extraction and bounding box regression to achieve higher accuracy in object detection tasks. YOLOv9 is one of the latest versions of the YOLO algorithm, published in February 2024. A distinctive feature of YOLOv9 from previous versions is the use of the Generalized Efficient Layer Aggregation Network (GELAN) and Programmable Gradient Information (PGI). This project will create and tune a neural network that can accurately and quickly detect cigarettes in images, including various lighting conditions and backgrounds. Cigarette recognition accuracy is critical to creating reliable monitoring systems that can be used in real-world environments.

We compiled a comprehensive dataset of images featuring cigarettes and ashtrays. In the study, we collected two datasets for the cigarette detection task and the ashtray detection task. It contains 1200 images of people smoking and 368 images of ashtrays. Each image in the dataset is annotated to indicate the presence of a cigarette in the photo, in the form of bounding boxes around the cigarette. The dataset includes various scenes, lighting and poses of people smoking, as well as background options such as streets, cafes and houses, making it representative of different cigarette detection scenarios. Bounding boxes were added to the dataset to indicate the locations of objects of interest. First, the model was trained using different sizes of YOLOv5, YOLOv8 and YOLOv9. When comparing the results of YOLO models, YOLOv5 models have higher accuracy than YOLOv8 and YOLOv9. More precisely, the highest result was 89%. For YOLOv8, the highest accuracy is 87%. For YOLOv9, the accuracy is 88%. Then we trained YOLOv5, YOLOv8 and YOLOv9 models on the prepared dataset using hyperparameter optimization to fine-tune their performance. Models were evaluated based on accuracy, precision, recall, and mAP score. Hyperparameter optimization plays a critical role in achieving optimal performance for deep learning models. Therefore, in the study, we tried various optimization techniques to fine-tune the YOLO models.

To obtain good results, the hyperparameters of YOLOv5, YOLOv8 and YOLOv9 models were optimized using various optimizers such as SGD, Adam and AdamW. As a result of the optimization, it was seen that using the SGD optimizer with the YOLOv9 model allowed us to obtain the best results. Setting distribution plots demonstrated the superiority of this combination in terms of accuracy and convergence speed, making it the preferred choice for our applications.

Results of the first trained models with default parameters: YOLOv5 achieves 89%, YOLOv8 achieves 87%, and YOLOv9 achieves 88%. After hyperparameter optimization, using the SGD optimizer proves to be the most effective and shows superior performance compared to other optimizers. While the performance of YOLOv5 remains unchanged at 89%, the performance of YOLOv8 and YOLOv9 increases to 93%. Although YOLOv8 and YOLOv9 have the same mAP(50) value after optimization with SGD, YOLOv9 outperforms YOLOv8 by 2% in the mAP(50-95) metric.

An important aspect of our study was the inclusion of ashtray detection in the images. Ashtray is an important indicator of smoking, and its recognition significantly increases the accuracy of detecting smoking incidents. We developed approaches to detecting ashtrays that were successfully integrated into our models, improving the overall results and making the system more versatile.

In addition, a special algorithm for detecting cigarettes and ashtrays has been developed and tested; This algorithm produced impressive results, reaching an accuracy of 96%. This algorithm successfully recognized cigarettes and ashtrays, minimizing the number of false positives and outperforming the initial result of models YOLOv5, YOLOv8 and YOLOv9.

The results of our study show that the use of state-of-the-art deep learning models and hyperparameter optimization can significantly increase the efficiency and accuracy of cigarette and ashtray detection in images. The use of the developed models and algorithms provides significant contributions to the fight against tobacco use by providing effective tools for monitoring, analyzing and preventing the harmful effects of tobacco on public health.

The use of state-of-the-art deep learning models and hyperparameter optimization has significantly increased the efficiency and accuracy of cigarette and ashtray detection in images. The developed models and custom algorithm provide effective tools for monitoring, analyzing, and preventing the harmful effects of tobacco on public health. The integration of these technologies into digital media platforms can play a crucial role in reducing the exposure of smoking-related content to vulnerable populations, particularly children. The promising results of this study pave the way for further research and development in this field, with the potential to make a significant impact on public health and safety.





1. GİRİŞ

Sigara kullanımı ve bağımlılığından dolayı ortaya çıkan sağlık sorunları dünyanın en büyük sağlık sorunlarından biridir. Sigara içmek vücudun birçok organına zarar vermekte ve çok sayıda hastalığa sebep olmaktadır. Dünya Sağlık Örgütü'ne göre 2023 yılında dünya çapında yaklaşık 1,3 milyar insan sigara içiyor ve bunların 8 milyondan fazlası sigaraya bağlı hastalıklardan ölüyor. Bu nedenle sigarayı kontrol altına almak ve önlemek için etkili yöntemlerin geliştirilmesi ve sigaranın tehlikeleri konusunda farkındalık yaratılması önemlidir (Dünya Sağlık Örgütü, 2023).

Sigara içmeyi kontrol etmenin yollarından biri, halka açık yerlerde sigara içen kişileri izlemek, sigara içmeyi düzenleyen yasaların ihlallerini tespit etmek ve sigara içenlerin davranış ve alışkanlıklarını analiz ederek önleyici adımlar oluşturmaktır. Bu işlemler için görmeye dayalı zeki tespit ve rapor sistemi oluşturulmalıdır. Geliştirilecek bu sistem ile sigarayla ilgili psikoloji, sosyoloji ve tıp araştırmaların yapılması da mümkündür.

Bir görüntüden sigara tespiti, görüntüdeki sigaraların konumlandırılması ve sınıflandırılmasını içeren bilgisayarlı görme görevidir. Bu görev, bir görüntüdeki tüm nesneler için sınırlayıcı kutuların ve sınıf etiketlerinin bulunmasından oluşan nesne algılamamanın bir alt görevidir. Nesne tespiti, güvenlik, tıp, robotik ve eğlence gibi çok çeşitli alanlarda kullanılmaktadır. Bilgisayarlı görmenin en hızlı büyüyen uygulama alanında birisidir.

Nesne tespiti bölge temelli ve tek geçişli olarak iki ana kategoriye ayrılmaktadır. Bölge temelli yöntemler de önce nesneleri içerebilecek aday bölgeler belirlenir ve sonrasında bu alanlar CNN'ler kullanarak sınıflandırır. Bu tür yöntemlerin örnekleri R-CNN ve Faster R-CNN'dir (Girshick, 2015). Tek geçişli yöntemler, sınırlayıcı kutuları ve sınıf etiketlerini tahmin etmek için bağlantı kutularını veya merkez noktaları kullanarak, ağı tek geçişinde nesne algılamayı gerçekleştirir. Bu tür yöntemlere örnek olarak YOLO, SSD ve RetinaNet verilebilir (Arkin ve ark, 2023).

1.1. Tezin Amacı

Bu tez çalışmada, video görüntülerinde sigara tespiti yapmak ve tespiti yapılan sigara bölgesini sansürlemektir. Sigara tanımayı iyileştirmek için kül tablası ve duman görüntüsü gibi destekleyici nesneler kullanılabilir. Bu çalışmada sınıflandırma başarısına kül tablasının katkısı da incelenmiştir. Sınıflandırma ve tespit için YOLOv5, YOLOv8 ve YOLOv9 gibi modern derin öğrenme modelleri kullanarak görüntülerdeki sigara ve kül tablasını tespit etmede etkili yöntemler geliştirmek ve uygulamaktır (Seddik ve ark, 2024). Bu çalışma, görüntü ve videolarda sigara içmeyle ilişkili nesneleri tanıma sürecini otomatikleştirmeyi sağlayan bilgisayarlı görme alanındaki modern teknolojileri incelemeyi amaçlamaktadır ve videodaki zararlı içeriklere maruz kalma gibi sorunları çözmek için otomatik içerik filtreleme ve yaşa uygun öneriler sunma gibi çözümler geliştirilmektedir.

Çalışmada sigara ve sigarayla ilişkili nesneleri tespit etmek konusunda daha önce yapılmış çalışmalarda kullanılan mevcut yöntemler ile algoritmaları analiz edilecektir. Günümüzde bilgisayarlı görme ve yapay zekâ teknolojileri hızla gelişiyor ve uygulamaya dönüşüyor. Bu alandaki gelişmeler tütün kullanımı kontrolünün etkinliğini önemli ölçüde artırabilmektedir (Santiago ve ark, 2022). Mevcut yaklaşımlara yönelik araştırmalar, en iyi çözümün belirlenmesine ve daha doğru ve güvenilir tanıma sistemlerin oluşturmada gerekli eksikliklerin belirlenmesi ve iyileştirilmesine yardımcı olacaktır.

Çalışmanın önemli bir yönü, sigara algılama görevi için YOLOv5, YOLOv8 ve YOLOv9 modellerinin geliştirilmesi ve eğitilmesidir. YOLO, yüksek doğruluğu ve hızıyla bilinen, nesne algılama görevleri için popüler bir algoritmadır. Ayrıca YOLO modelleri farklı aydınlatma koşulları ve arka planlarına sahip görüntülerdeki sigaralara nesnesini doğru ve hızlı bir şekilde tespit edebilen bir sinir ağıdır. Bu yönü ile sigara tanımada, gerçek dünya ortamlarında kullanılabilecek güvenilir izleme sistemleri oluşturmak için kullanımı ön plana çıkmaktadır.

Çalışmada sonuçları iyileştirmek için kül tablasının da veri seti oluşturulmuştur. Kül tablası, sigara içmenin önemli bir göstergesidir ve tanınması, sigara tamamen görünmese bile, sigara içme olaylarının tespiti başarısını önemli ölçüde arttırmaktadır.

1.2. Literatür Araştırması

Sigara içmek dünyada önde gelen ölüm ve hastalık nedenlerinden biridir, bu nedenle halka açık yerlerde sigara içilmesini tespit etmek ve önlemek için etkili yöntemler geliştirmek önemlidir. Son yıllarda, bu sorunu çözmek için derin evrişimli sinir ağlarının ve diğer bilgisayarlı görme tekniklerinin kullanımına ilişkin birçok araştırma ortaya çıkmıştır. Bundan dolayı önce literatürdeki benzer çalışmalardan birkaç tanesine bakılacak, bunların yaklaşımları, yöntemleri, sorunları ve sonuçları karşılaştırılacaktır.

Sigara içilmesini tespit eden sistemlere örnek olarak Eye-Smoker verilebilir. Bilgisayar görmesine ve sinir ağlarına dayanan Eye-Smoker sistemi, görüntüde sigara içilmesini tespit etmek için yenilikçi bir yaklaşım sunar. Eye-Smoker, sigara içilip içilmediğini tespit için sigara içen kişileri burun bölgesini kullanır. YOLOv3 mimarisi ile görüntüdeki sigara ve burun deliğine göre sigara içilip içilmediği kararı verilir. Sistem, sigara içilip içilmediğini doğru bir şekilde tespit etmek için el hareketleri, duman ve insan duruşları gibi özelliklerinde başarıya etkisini inceler. Eğitim verisi olarak, video ve web kamerası görüntülerinden etiketlenen burun, sigara ve el gibi nesneler kullanılmıştır. Sistem halka açık alanlarda sigara kullanımını tespit etmek için bir avantaj sunar. Sistemin dezavantajı ise görüntüde burun görünmüyorsa, sigara içme tespitinin başarılı olmamasıdır (Macalisang ve ark, 2020).

Konuyla ilgili bir diğer çalışma ise, halka açık yerlerde sigara içen kişileri tespit etmek için derin CNN'e kullanılan bir çalışmadır. Bu çalışma az sayıda eğitim verisi kullanılarak gerçekleştirilmiştir. Gerçekleştirilen istemde, sigara içme belirtilerinin sıklıkla görüldüğü yüz ve el bölgesine odaklanılmakta, bu bölgelerdeki hareketliliğe göre sigara içme faaliyeti olup olmadığına karar verilmektedir (Khan ve ark, 2022).

Santiago ve ark. (2022) sigarayı tespit etmek için YOLO-Sigara sistemini oluşturmuşlardır. Çalışmalarında kullandıkları modelin ana mimarisi, yüksek nesne algılama hızıyla bilinen YOLO'nun geliştirilmiş bir versiyonu olan YOLOv5'dir. YOLO-Sigara sisteminin en önemli özelliklerinden biri küçük sigara nesnelerinin algılamasında yüksek derecede doğruluk vermesidir. Bunu başarmak için model, küçük hedeflerin tespit doğruluğunu artıran yeni bir FSPP modülünü içerir. Ayrıca MSAM mekanizmasının kullanılması, modelin görüntünün önemli bölümlerine odaklanma yeteneğini geliştirir ve bu da algılama doğruluğunun iyileştirilmesine

yardımcı olur. Hesaplama karmaşıklığını azaltmak ve tespit sürecini hızlandırmak için model parametreleri incelenir. Bu model, sigara içen kişilerin tespit doğruluğunda orijinal YOLOv5 modeline göre üstünlük göstermekte ve gerçek dış mekân koşullarında yüksek performansa ulaşmaktadır (Santiago ve ark, 2022).

Zhang ve ark. (2018) araştırmasında, SmokingNet adı verilen evrişimli bir sinir ağıni temel alan bir sigara içme tespit modelini açıklamaktadır. Bu model, görüntüler aracılığıyla video içeriğindeki sigara içilmesini otomatik olarak algılar. Sigara dumanı algılama algoritmalarına dayalı olarak sigara içmeyi tespit eden geleneksel yöntemlerin aksine, SmokingNet, sigara dumanını tespit etmeye gerek kalmadan yalnızca insanın sigara içme hareketi bilgilerini ve sigara görüntüsü özelliklerini kullanarak sigara içme görüntülerini tespit edebilir. Gerçek zamanlı izleme için yüksek doğruluk ve üstün performans göstermektedir.

Wang ve arkadaşlarının çalışmasında (2023), sağlık ve güvenliği artırmak amacıyla hastaneler, okullar ve istasyonlar gibi halka açık yerlerde sigara içme davranışını tespit etmek için geliştirilmiş bir YOLOv5 tabanlı algoritma sunmaktadır. Algoritma, C3 modülünü CoT modülüyle değiştirerek ve CBAM dikkat mekanizmasını SPPF yapısından önce entegre ederek, özellik çıkarımını iyileştirerek algılamayı geliştirir. Geliştirilmiş algoritma, deneysel karşılaştırmalar yoluyla tespit performansını önemli ölçüde arttırmaktadır, doğruluğu %61,3'ten %62,9'a yükselterek etkinliğini ve gerçek zamanlı tespit yeteneklerini göstermektedir.

Ding ve arkadaşlarının çalışmasında (2023), karmaşık sahnelerdeki küçük ve orta ölçekli hedefler için düşük tespit doğruluğunu ele alan, halka açık yerlerde sigara içme tespiti için geliştirilmiş bir YOLOv5 algoritması önermektedir. Geliştirmeler arasında, daha iyi veri geliştirme için Mosaic-9'un kullanılması, hedef konumlara odaklanmayı geliştirmek için CBAM tanıtılması, daha iyi semantik tanıma için geleneksel üst örnekleme transpoze evrişimle değiştirilmesi ve daha iyi anlamsal tanıma için SPD-Conv modülünün dahil edilmesi yer alır. Düşük çözünürlüklü ve küçük hedef görüntüler. Deneysel sonuçlar, sigara içme hedeflerinin tespit doğruluğunda %3'lük bir iyileşme göstermektedir.

Peng ve ark. (2023) araştırmasında, ekipman hasarını önlemeyi ve güvenliği sağlamayı amaçlayan, trafo merkezi personeli tarafından sigara içildiğini tespit etmek için gelişmiş bir algoritmayı tartışıyor. GhostNetV2-YOLOv5'i temel alan bu

yeni yaklaşım, algılama doğruluğunu ve hızını iyileştirerek orijinal YOLOv5'i geliştirir. Spesifik olarak, toplam mAP %2,58'lik bir artış ve tahmin hızında 1,61 kat artış elde eder.

Bu çalışmalar karşılaştırıldığında hepsinin sigarayı tespit etmek için derin öğrenmeyi kullandıkları ancak farklı yaklaşım ve teknikleri kullandıkları görülmektedir. Eye-Smoker ve YOLO-Cigarette, sigara içme tespiti için YOLO modelinin çeşitlerini kullanırken, diğeri kendi derin öğrenme modelini kullanır. Tüm bu çalışmalar, sonuçlarında yüksek doğruluk olduğunu ortaya koyuyor ve sigara tespitinde derin öğrenmenin kullanılmasının etkinliğini vurguluyor.





2. YAPAY ZEKA VE YOLO KÜTÜPHANESİ

2.1. Makine Öğrenmesi

Makine öğrenme, bir makineye, bir algoritmaya, deneyime veya verilere dayanarak insan gibi düşünmesinin ve davranmasının öğretildiği bir yöntemler sınıfıdır. Makine öğrenimi programlama ile aynı şey değildir. Programcı makine için bir algoritma oluşturur: istenen sonuca yol açacak net bir eylem dizisi belirler. Modeli eğiten ML uzmanı, makine için bir program yazmaz. Verileri iletir ve sonuç olarak ne elde etmek istediğini açıklamaya çalışır. Algoritmanın ulaşması gereken belirli bir cevabı yoktur, yalnızca sorulan soruyu yanıtlayan bir modelin nasıl oluşturulacağını bilir. Makine öğreniminin amacı modele kendi başına bir çözüm bulmayı öğretmektir (Mitchell, 2006).

Makine öğreniminin ilkesi, insanlar gibi düşünen bilgisayarlar ve programlar oluşturmanıza olanak tanır. Ancak aynı zamanda insanlardan farklı olarak yorulmazlar, daha az hata yapabilirler, her türlü veriyle çalışabilirler ve bunları tarafsız bir şekilde değerlendirebilirler. Bu, yapay zekanın olanakları için devasa bir alan açıyor.

Makine öğrenmesi birçok hizmet ve program oluşturmakta kullanılmaktadır. Günlük yaşamda kullandığımız en basit araçlardan (navigatörler, yüz tanıma özellikli kameralar, öneri hizmetleri) endüstri veya güvenlikte ihtiyaç duyulan karmaşık araçlara kadar. Örneğin, bir yayın hizmeti çeşitli filmler sunduğunda ve bir trafik kamerası kalabalığın içindeki bir suçluyu tanıdığında, makine öğrenme teknikleri ile gerçekleştirilmektedir. Verilerin alınmasından, sınıflandırılmasından ve doğru bir tahmin veya sonuca varılmasından sorumludur.

Aşağıda makine öğreniminin altı temel adımla nasıl çalıştığı gösterilmektedir (Chapman ve ark, 1999):

1. Veri Seçmek: Bu aşamada modeli eğitmek için kullanılacak veriler seçilir ve hazırlanır. Buna çeşitli kaynaklardan veri toplanması, gürültü ve hatalardan arındırılması ve analiz edilmesi kolay bir formata dönüştürülmesi de dahildir.

2. Verileri analiz etmek: Burada verileri analiz eden ve kalıpları tanımlayan bir matematiksel model oluşturulur. Bu adım, doğrusal regresyon, karar ağacı veya sinir ağı gibi eldeki göreve en uygun makine öğrenimi algoritmasının seçilmesini içerir.
3. Modeli doğrulamak: Model, bir test veri seti kullanılarak doğruluk ve kesinlik açısından test edilir. Bu süreç modeldeki kusurların belirlenmesine ve geri bildirim alınmasına yardımcı olur. Doğrulama genellikle daha güvenilir sonuçlar için çapraz doğrulama tekniklerinin kullanılmasını içerir.
4. Modeli ayarlamak: Model maksimum doğruluk ve performans için optimize edilmiştir. Bu adım, öğrenme hızı, karar ağacı derinliği veya bir katmandaki nöron sayısı gibi hiper parametrelerin ayarlanmasını ve aşırı uyumu önlemek için düzenleme tekniklerinin uygulanmasını içerir.
5. Modeli kullanmak: Model hazır ve optimize edildiğinde gerçek problemleri çözmek için kullanılmaya başlar. Örneğin model, talebi tahmin etmek, görüntüleri sınıflandırmak veya sahte işlemleri gerçek zamanlı olarak tespit etmek için kullanılabilir.
6. Modeli test etmek: Bu aşamada model, sağlamlığından ve doğruluğundan emin olmak için düzenli olarak test edilir ve yeni veriler üzerinde değerlendirilir. Test, zaman içinde model performansında meydana gelen değişiklikleri tanımlamaya yardımcı olur ve modelin güncel ve güvenilir olmasını sağlar.

2.2. Derin Öğrenme

Derin öğrenme, makine öğrenmesinin bir alt kümesidir ve çok katmanlı yapay sinir ağlarının kullanımına dayanır. Derin öğrenme ile geleneksel makine öğrenimi arasındaki temel fark, sinir ağlarının verilerden özellik hiyerarşilerini otomatik olarak çıkarma yeteneğidir. Bu özellikle büyük miktarda veriyi işlerken ve karmaşık sorunları çözerken faydalıdır. Derin öğrenmenin temelini oluşturan sinir ağları bir giriş katmanı, birden fazla gizli katman ve bir çıkış katmanından oluşur (Lecun ve ark, 2015). Ağın her katmanı, verilerden giderek daha soyut özelliklerin çıkarılmasından sorumludur; bu, sınıflandırma, regresyon ve tahmin problemlerini etkili bir şekilde çözmenize olanak tanır.

Derin öğrenme ile makine öğrenmesi arasında birkaç temel fark vardır (Nguyen ve ark, 2019):

1. Modellerin yapısı. Geleneksel makine öğrenmesi sınırlı sayıda katmana sahip basit modeller kullanırken, derin öğrenme çok katmanlı sinir ağlarını kullanır. Derin ağlar, modellerin verilerden otomatik olarak özellikler çıkarmasına olanak tanıyarak manuel özellik mühendisliği ihtiyacını en aza indirir.

2. Veri gereksinimleri. Makine öğrenimi genellikle daha az eğitim verisi gerektirir. Yüksek doğruluk elde etmek için genellikle birkaç bin veya on binlerce örnek yeterlidir. Derin öğrenme ise aksine, bu verileri işlemek için büyük miktarda eğitim verisi (yüzbinlerce veya milyonlarca örnek) ve önemli bilgi işlem kaynakları gerektirir.

3. Veri işleme. Makine öğreniminde, alanın derinlemesine anlaşılmasını gerektiren manuel özellik mühendisliğine önemli bir vurgu vardır. Derin öğrenme, özellikleri ham verilerden otomatik olarak çıkararak veri hazırlama sürecini basitleştirir ve modellerin karmaşık bağımlılıkları bulmasına olanak tanır.

Derin öğrenme, karmaşık problemleri çözebilme ve büyük miktarda veriyi işleyebilme yeteneği nedeniyle çeşitli alanlarda yaygın olarak kullanılmaktadır. Bilgisayarlı görme alanında derin öğrenme, görüntü tanıma ve sınıflandırmayı, nesne algılamayı ve görüntü bölütlemeyi mümkün kılar. Örneğin, sürücüsüz arabalar, güvenlik sistemleri ve X-ışını ve MRI analizi gibi tıbbi teşhisler, sistemlerinin doğruluğunu ve güvenilirliğini artırmak için derin sinir ağlarını kullanır.

2.3. Yapay Sinir Ağları

Yapay sinir ağlarını insan beynini gibi çalışmaktadır. İnsan beynindeki nöron birbirine oldukça bağlı bir karmaşık ağ oluşturur ve insanların bilgiyi işlemesine yardımcı olmak için birbirlerine elektrik sinyalleri gönderir. Benzer şekilde bir yapay sinir ağı da problemleri çözmek için etkileşime giren yapay nöronlardan oluşur. Yapay nöron, birden fazla girdi olgusunu tek bir çıktıya dönüştüren bir fonksiyondur. Bunu nasıl yapacağı, olguların (ağırlıkların) önemine ve nöronun kendi ayarlarına bağlıdır. Bu tür nöronlardan oluşan bir sinir ağı, girdi verilerini analiz edebilen ve kararlar verebilen matematiksel bir modeldir (Jain ve ark, 1996). Buna göre, yapay sinir ağları, matematiksel hesaplamaları gerçekleştirmek için bilgisayar sistemlerini kullanan programlar veya algoritmalarıdır.

Temel sinir ağı, birbirine bağlı üç yapay nöron katmanı içerir:

1. Giriş katmanı, dış dünyadan gelen bilgiler girdi katmanından yapay sinir ağına girer. Giriş düğümleri verileri işler, analiz eder veya sınıflandırır ve bir sonraki katmana aktarır.

2. Gizli katmanlar, giriş katmanından veya diğer gizli katmanlardan girdi alır. Yapay sinir ağları çok sayıda gizli katmana sahip olabilir. Her gizli katman, bir önceki katmanın çıktısını analiz eder, işler ve bir sonraki katmana aktarır.

3. Çıkış katmanı, tüm verilerin yapay sinir ağı tarafından işlenmesinin nihai sonucunu verir. Bir veya daha fazla düğüme sahip olabilir. Örneğin, bir ikili sınıflandırma problemini çözerken, çıktı katmanında "1" veya "0" sonucunu üretecek bir çıktı düğümü bulunur. Ancak çoklu sınıflandırma durumunda çıktı katmanı birden fazla çıktı düğümünden oluşabilir.

Mimari, işletim özellikleri ve uygulama alanları göre sinir ağları birkaç türlere ayrılmaktadır. Aynı zamanda FFNN, CNN ve RNN tür sinir ağları en çok yaygındır. Bunların içinde CNN, görüntü sınıflandırma, nesne tanıma, tahmin, doğal dil işleme ve diğer görevler için kullanılır.

2.4. Evrişimli Sinir Ağları

Evrişimli Sinir Ağları (CNN), bilgisayarların görüntüleri ve videoları görmesine ve anlamasına yardımcı olan özel türde sinir ağlarıdır. Bu tür ağlar, evrişimli ağlar adı verilen çok sayıda katmana sahiptir. Bunlar, CNN'lerin karmaşık özellikleri öğrenmesine ve görsel materyallerin içeriği hakkında daha doğru tahminler yapmasına olanak tanır (Long ve ark, 2015). Evrişimli sinir ağları, yüz tanıma, otopilot, tıbbi prototip oluşturma ve doğal dil işleme gibi çok alanlarda kullanılmaktadır.

Evrişimli sinir ağının çalışma prensibi aşağıdaki gibidir (Matthew ve Fergus, 2013):

1. Evrişim Katmanı: bu katman evrişim işleminin gerçekleştiği yerdir. Filtre adı verilen ve boyutu (n,n) olan bir kayan pencere, giriş özelliği boyunca hareket eder. Hareket sayısı belirtilen filtre sayısına göre belirlenir. Ortaya çıkan her şablon (n,n,d) formuna sahiptir; burada d , giriş özelliğinin derinliğidir.

2. Aktivasyon Katmanı: evrişim katmanını genellikle evrişim sonuçlarına doğrusal olmayan bir aktivasyon fonksiyonu uygulayan bir aktivasyon katmanı takip eder. Bu,

ağın daha karmaşık fonksiyonları öğrenmesine ve bunlara yaklaşmasına olanak tanıyan modele doğrusal olmama özelliğini ekler.

3. Havuzlama Katmanı: giriş verilerinin boyutsallığını azaltmak için kullanılır; bu, aşırı uyumun azaltılmasına yardımcı olur ve eğitimi hızlandırır.

4. Tam Bağlantı Katmanı: evrişimli bir sinir ağının sonunda, genellikle, çıkarılan özelliklere göre nesneleri sınıflandırmak için kullanılan, tamamen bağlı bir sinir ağı bulunur.

Evrişimli sinir ağları, yapay zekâ alanında, özellikle bilgisayarlı görme ve görüntü işlemede önemli bir rol oynamaktadır. Onların yardımıyla yüz tanıma, otomatik görüntü açıklaması, araba ve dronların otomatik sürüşü, tıbbi görüntü analizi gibi karmaşık sorunları çözülmektedir.

Evrişimli sinir ağlarının performansını optimize etmek için kayıp fonksiyonu, düzenlileştirme ve normalleştirme gibi çeşitli yöntemler kullanılır. Eğitim sürecini ve genel ağ performansını iyileştirmek için ön eğitim, transfer öğrenimi ve veri artırma teknikleri de kullanılır.

2.5. Kayıp fonksiyonu

Bir derin öğrenme modelini eğitmek, belirli bir girdi-çıkı çiftleri kümesi için model tahminlerini gerçek çıktıya yaklaştırmakla ilgilidir. Modelin sonuçları gerçek sonuçlara yakın olduğu zaman model eğitimiz biter. Model tahminlerinin gerçek çıktılara ne kadar yakın olduğunu anlamak için kayıp fonksiyonu olarak bilinen bir matematiksel fonksiyona ihtiyacı var. Bu kayıp fonksiyonu, modelinizin belirli bir eğitim adımında ne kadar iyi performans gösterdiğini gösterir (Goodfellow ve ark, 2016).

Veri kümesi boyutu, gürültü seviyeleri ve seyrek özellikler gibi faktörler dikkate alınarak bir optimize edicinin dikkatli seçimi, eğitim verimliliğini ve derin öğrenme modellerinin nihai performansını önemli ölçüde etkiler.

2.5.1. Stokastik Gradyan İnişi

SGD bir gradyan inişinin çeşididir. Modelin parametrelerini daha sık güncellemeye çalışır. Burada model parametreleri, her eğitim örneğindeki kaybın hesaplanmasından sonra değiştirilir. Dolayısıyla, veri kümesi 1000 satır içeriyorsa

SGD, model parametrelerini Gradyan İnişin’de olduğu gibi bir kez yerine veri kümesinin bir döngüsünde 1000 kez güncelleyecektir. SGD aşağıdaki denklem ile hesaplanır:

$$w_{t+1}=w_t - a \frac{\partial L}{\partial w_t} \quad (2.1)$$

burada, Mevcut gradiyenti ($\partial L/\partial w_t$), öğrenme katsayısı (a) ile çarparak mevcut ağırlığı (w_t) güncellemektedir (Seyyarer ve ark, 2020).

SGD’nin birçok önemli avantajı vardır. Basitliği ve verimliliği onu sinir ağlarının eğitimi için popüler kılmaktadır. Ağırlıkların bireysel örneklerle göre güncellenmesiyle SGD daha hızlı yakınsar ve yineleme başına düşük hesaplama maliyetine sahiptir. Büyük veri kümeleriyle iyi çalışır ve yerel minimumlardan kaçınmaya yardımcı olarak çözüm alanının daha derinlemesine araştırılmasını kolaylaştırır ve model performansını artırır.

2.5.2. Adam

Adam yine stokastik gradyan inişine dayalı bir optimizasyon algoritmasıdır. Her parametre için uyarlanabilir öğrenme oranlarını hesaplayan bir yöntemdir. Hem momentuma benzer şekilde geçmiş gradyanların azalan ortalamasını hem de RMS-Prop ve Adadelta’ya benzer şekilde geçmiş kare gradyanların azalan ortalamasını saklar. Böylece her iki yöntemin avantajlarını birleştirir. Aşağıdaki denklemlerde optimizasyon algoritmasının nasıl hesaplanacağı gösterilmiştir:

$$w_{t+1}=w_t - \frac{\alpha}{\sqrt{\hat{S}_t} + \epsilon} * \hat{V}_t \quad (2.2)$$

$$\hat{S}_t = \frac{S_t}{1 - \beta_1^t} \quad (2.3)$$

$$\hat{V}_t = \frac{V_t}{1 - \beta_2^t} \quad (2.4)$$

$$S_t = \beta_1 S_{t-1} + (1 - \beta_1) \left[\frac{\partial L}{\partial w_t} \right]^2 \quad (2.5)$$

$$V_t = \beta_2 V_{t-1} + (1 - \beta_2) \left[\frac{\partial L}{\partial w_t} \right]^2 \quad (2.6)$$

burada S ve V başlangıçta 0, $\beta=0.95$ ve ϵ ise 10^{-6} olarak tercih edilmektedir (Kingma ve Lei Ba, 2015).

2.5.2. AdamW

AdamW optimize edici, Adam algoritmasının geliştirilmiş bir versiyonudur ve sınır ağırlarını eğitmek için kullanılır. AdamW ve Adam arasındaki temel fark, AdamW'nin ağırlıkları düzenlemek için L2 düzenlemesi yerine ağırlık azalmasını kullanmasıdır. Bu, daha iyi model performansına ve daha verimli eğitime olanak tanır.

$$\theta_t = \theta_{t-1} - \eta \left(\frac{\hat{S}_t}{\sqrt{\hat{V}_t} + \epsilon} + \lambda \theta_{t-1} \right) \quad (2.7)$$

burada, θ_t – adım t 'deki model parametreleri, η – öğrenme oranı, ϵ sifıra bölünmeyi önlemek için küçük bir sayıdır, genellikle $\epsilon=10^{-8}$, λ – ağırlık azalma faktörü (Loshchilov ve Hutter, 2019).

AdamW optimizier, geleneksel Adam'a göre önemli avantajlar sunar. L2 düzenlemesinden farklı olarak AdamW, ağırlıkların boyutunu doğrudan azaltan, aşırı uyumu önleyen ve genel model performansını artıran ağırlık azalmasını kullanır. Adam'ın uyarlanabilir öğrenme ve momentum gibi, istikrarlı ve hızlı öğrenme sağlayan faydalarını korur. Entegrasyon kolaylığı ve minimal hiper parametre ayarı, AdamW'ü çeşitli makine öğrenimi ve derin öğrenme görevleri için uygun hale getirir.

2.6. YOLO algoritması

YOLO, Washington Üniversitesi'nde Joseph Redmon ve Ali Farhadi tarafından geliştirilen en popüler ve etkili nesne algılama yöntemlerinden biridir. YOLO, yüksek hızı ve doğruluğu nedeniyle sıklıkla kullanılmaktadır. 2016 yılında algoritmanın yeni versiyonu YOLOv2 çıkmıştır. Toplu normalleştirme, bağlantı çerçeveleri ve boyut kümeleri sunarak orijinal modeli geliştirmektedir (Redmon ve ark, 2016). 2018'de ortaya çıkan YOLOv3, daha verimli bir omurga ağı, çoklu bağlantı çerçeveleri ve uzamsal piramit havuzu kullanarak modelin performansını daha da geliştirdi (Redmon ve ark, 2018). YOLOv4 2020'de ortaya çıktı ve döşemeli veri artırma, yeni çapasız algılama kafası ve yeni kayıp işlevi gibi yenilikler getirdi (Bochkovskiy ve ark, 2020). YOLOv5, model performansını daha da geliştirdi ve hiper parametre optimizasyonu, entegre deney izleme ve popüler dışa aktarma formatlarına otomatik dışa aktarma gibi yeni özellikler ekledi (Jayakumar ve ark,

2024). YOLOv6, Meituan tarafından 2022'de yayınlandı ve şirketin birçok otonom dağıtım robotunda kullanılmaktadır (Li ve ark, 2022). YOLOv7, COCO anahtar noktaları veri setine poz tahmini gibi ek görevler ekledi (Wang ve ark, 2022). YOLOv8, 2023 yılında ortaya çıkmış. Gelişmiş, türünün en iyisi SOTA modeli olan YOLOv8, performansı, esnekliği ve verimliliği artırmak için yeni özellikler ve iyileştirmeler sunarak önceki sürümlerin başarısını temel alıyor. YOLOv8, algılama, segmentasyon, poz tahmini, izleme ve sınıflandırma dahil olmak üzere çok çeşitli yapay görme görevlerini destekler. Bu çok yönlülük, kullanıcıların çeşitli uygulamalarda ve alanlarda YOLOv8'in yeteneklerinden yararlanmasına olanak tanır (Ultralytics YOLO Docs). YOLOv9, PGI ve GELAN gibi inovasyon teknikleri sunarak gerçek zamanlı nesne algılamada önemli bir ilerlemeye katkı sağlamaktadır.



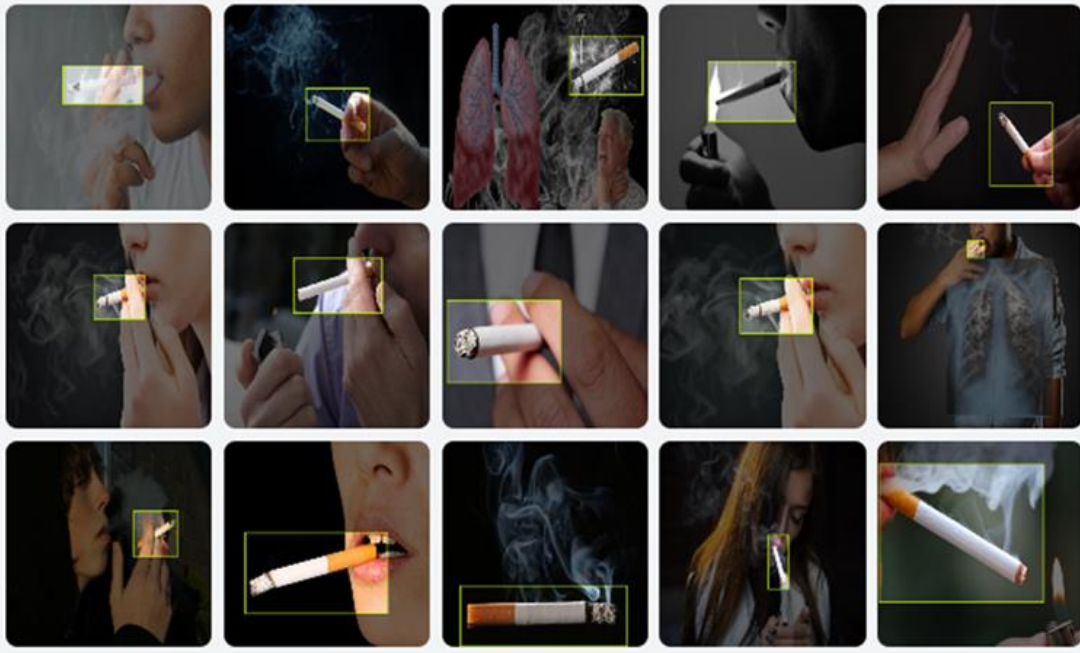
3. MATERYAL ve YÖNTEM

3.1. Veri Setleri

Sınıflandırma için iyi bir veri seti gereklidir. Bu çalışmada kendi veri setimizi oluşturduk. Veri seti oluşturmada sigara nesnesi olan fotoğraflar ve video frameleri analiz edilmiştir. Sınıflandırma da, gerek eğitim, gerek ise test görsellerin kalitesi ve örnek uzayı önemlidir ve modelin eğitim verimliliğini doğrudan etkilemektedir. Veri tabanı için farklı aydınlatma koşullarındaki sigara görüntülerini ve olası sigara nesnesi örneklerini içeren bir görüntü veri tabanının oluşturulmuştur. Veri setini toplamak ve görüntülerdeki nesneleri etiketlemek için Roboflow Universe web uygulaması kullanılmıştır (Ciaglia ve ark, 2022).

Roboflow Universe, görüntü veri kümelerini ve modellerini paylaşmaya yönelik bir platformdur. Kullanıcılar bu sayfada, sınıflandırma, nesne algılama ve segmentasyon gibi görüntü tanıma görevleri için veri kümelerini yükleyebilir, paylaşabilir ve var olan veri setlerini inceleyebilirler. Platform çeşitli veri formatlarını destekler. Veri setleri üzerinde veri artırma, açıklama ekleme ve ön işleme yapma için araçlar sunar. Roboflow Universe, kaliteli verilere ve modellere erişimi sağlayarak, bilgisayar görmesi alanında çalışan araştırmacılara ve geliştiricilere önemli bir kaynak teşkil eder (Ciaglia ve ark, 2022).

Veri setinde toplanan resimlerde sigara nesnesi ve sigara içen kişilere ait 1200 görüntü bulunmaktadır. Veri seti üç bölüme ayrılmıştır; birinci kısımda 840 görüntü içeren eğitim seti (70%), ikinci kısım 180 görüntü içeren doğrulama seti (15%) ve son bölümde ise 180 görüntü içeren bir test seti (15%) bulunmaktadır. Veri kümesindeki her bir görüntüye, sigaranın varlığını belirten bir sınırlayıcı çerçeve ve açıklama eklenmiştir. Veri seti, sigara içen insanların çeşitli sahnelerini, farklı ışıklandırmalarını, pozlarını ve farklı sokaklar, kafeler, evler gibi arka plan varyasyonları da yer almaktadır.



Şekil 3.1. Sigara veri seti.

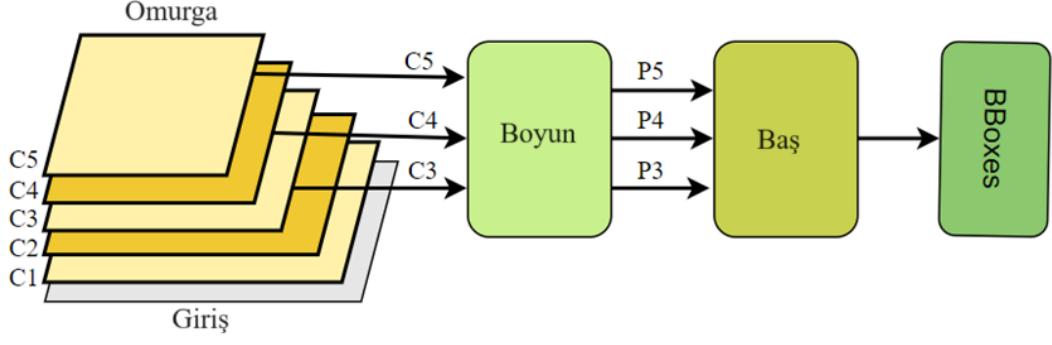
Sigara veri setinin yanında, ikinci bir veri seti daha oluşturulmuştur. Bu veri seti kül tablalarını içermektedir. Veri setinde 368 kül tablası görüntüsü yer almaktadır. Bu veri seti sigara tespitin iyileştirilmesi için kullanılacaktır.

3.2. Modelin Eğitilmesi

Görüntü üzerindeki sigara nesnelerini tespit etmek için YOLOv5, YOLOv8 ve YOLOv9 derin öğrenme modelleri kullanılmıştır. YOLO derin öğrenme modelleri bilgisayarlı görme araştırmalarında kullanılan güncel algoritmadır. Bu algoritmalar nesne tanımda yüksek doğruluğa ve hıza sahiptir. Bu bölümde sırası ile YOLOv5, YOLOv8 ve YOLOv9 derin öğrenme modelleri verilmiştir.

3.2.1. YOLOv5 modeli

YOLOv5, yüksek doğruluk oranıyla bilinen ve gerçek zamanlı nesne tespiti için yaygın olarak kullanılan YOLO algoritmasının bir versiyonudur. YOLOv5 modeli aşağıdaki Şekil 3.2’de gösterildiği gibi giriş, omurga, boyun ve tahmin olmak üzere dört ana bölümden oluşur (Jayakumar ve ark, 2024).

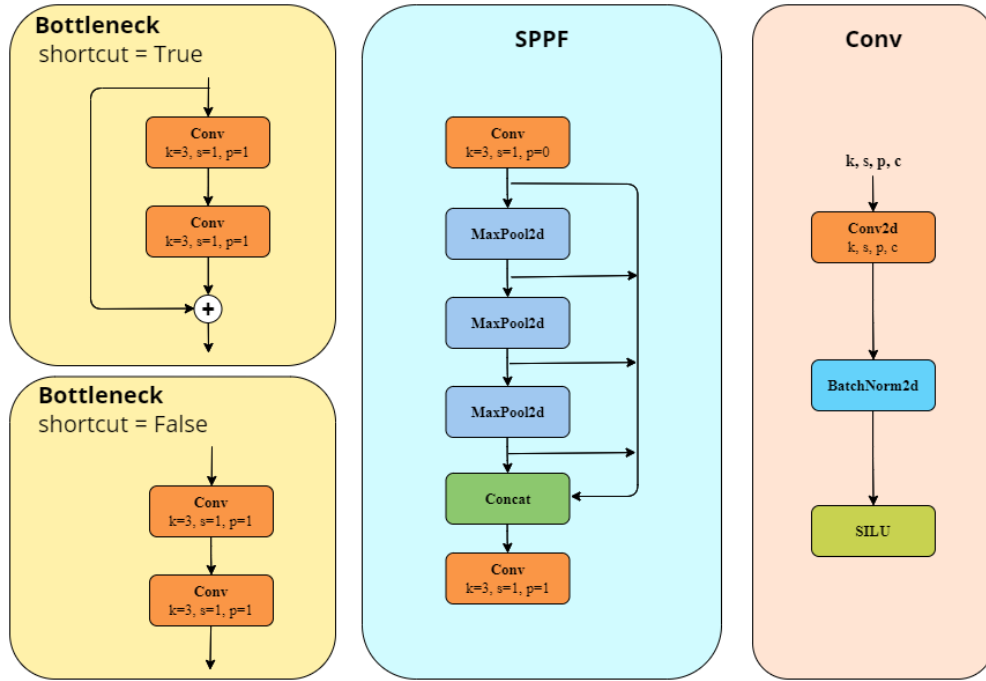


Şekil 3.2. YOLOv5 modelinin genel mimarisi.

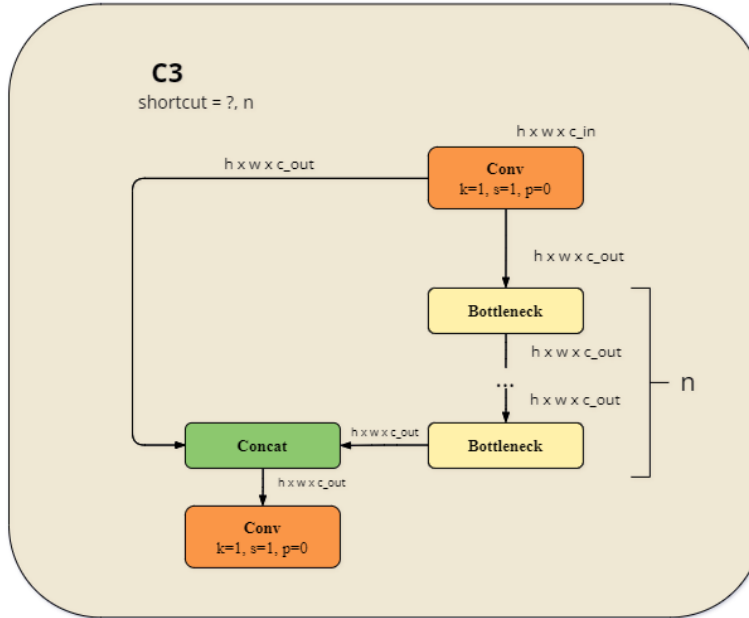
1. Giriş; temel olarak Mozaik veri geliştirmeyi, görüntü boyutu işlemeyi ve uyarlanabilir bağlantı çerçevesi hesaplamasını içerir. Ağın giriş görüntü boyutu 640*640'tır. Bu aşama genellikle bir görüntü ön işleme aşamasını içerir, yani giriş görüntüsü ağın giriş boyutuna göre ölçeklendirilir ve normalleştirilir. Ağ eğitimi aşamasında YOLOv5, rastgele ölçeklendirme, rastgele kırpmaya ve rastgele düzenleme yoluyla görüntüleri ve birleştirmeyi işlemek için Mozaik veri geliştirmeyi kullanarak veri setini büyük ölçüde zenginleştirir. Özellikle rastgele ölçeklendirme çok sayıda küçük hedef ekler, bu da ağı daha sağlam hale getirir ve küçük hedeflerin tespitine yardımcı olur.

2. Omurga, girişe uygulanan görüntülerden özellikler çıkarılmasından sorumludur. Omurga, farklı çözünürlük seviyelerine sahip birkaç evrişimli bloktan oluşur. Her blok evrişim katmanlarını, normalizasyon katmanlarını ve aktivasyon fonksiyonlarını içerir. YOLOv5'in temel bir özelliği, giriş özelliklerini iki parçaya bölen, bunları paralel olarak işleyen ve daha sonra bunları birleştiren BottleNeck, C3 bloklarının kullanılmasıdır, BottleNeck bloğu Şekil 3.3'te gösterilmiştir. C3 bloğu, doğruluk kaybı olmadan model parametrelerinin sayısını önemli ölçüde azaltmaktadır. Böylece, hesaplama karmaşıklığını ve bilgi kaybını azaltır. Şekil 3.4'te C3 bloğu gösterilmiştir. Her katmanda, maksimum havuzlama veya birden büyük adımlı evrişimli katmanlar gibi işlemler kullanılarak çözünürlük azaltılır. Her bloğun çıktı özellikleri, FPN ve PAN'ı içeren modelin Boyun kısmına iletilir. FPN (Lin ve ark, 2018), farklı boyutlardaki nesnelerin tanınmasını geliştirmek için farklı çözünürlük seviyelerindeki özellikleri birleştirir; PAN (Liu ve ark, 2017) ise öğrenmeyi geliştirmek için bilgi yolları ekler. Özellik çıkarmaya odaklanan omurganın aksine boyun, farklı ölçeklerde model performansını artırmak için bunları birleştirmek ve

işlemekten sorumludur. Böylece, omurga temel ve üst düzey özellikleri çıkarır ve boyun, doğru nesne tespiti için bu özellikleri optimize eder ve birleştirir.



Şekil 3.3. Bottleneck, SPPF ve Conv bloklarının mimarileri.

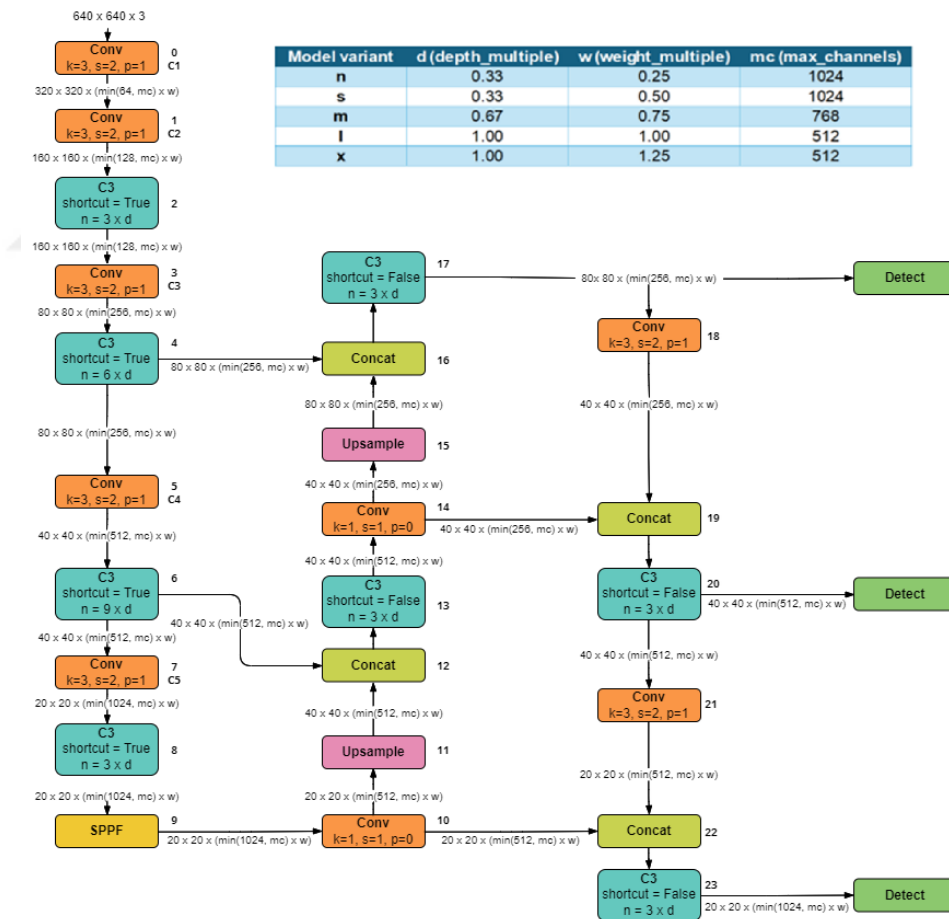


Şekil 3.4. C3 bloğunun mimarisi.

3. Boyun, yukarıda anlatıldığı gibi FPN modülünü ve PAN modülünü birleştirir. YOLOv5'in FPN modülü esas olarak yukarıdan aşağıya örnekleme yoluyla farklı katmanların özellik harita bilgilerinin füzyonunu birleştirir. PAN, FPN bazında aşağıdan yukarıya bir yön ekler ve FPN'deki alt katmandaki bilgileri tekrar üst katmana aktarır.

4. Tahmin, hedef nesnenin kategori olasılığını, nesne puanını ve o nesne için sınırlayıcı kutunun konumunu içeren bir vektörü verir. Algılama ağı, üç algılama katmanından oluşur. Farklı boyutlardaki hedef nesneleri tespit etmek için farklı boyutlardaki özellik haritaları kullanılır. Her tespit katmanına karşılık gelen vektörü çıkarır. Son olarak orijinal görüntüdeki hedefin tahmin sınır kutusunu ve kategorisini oluşturur ve işaretler.

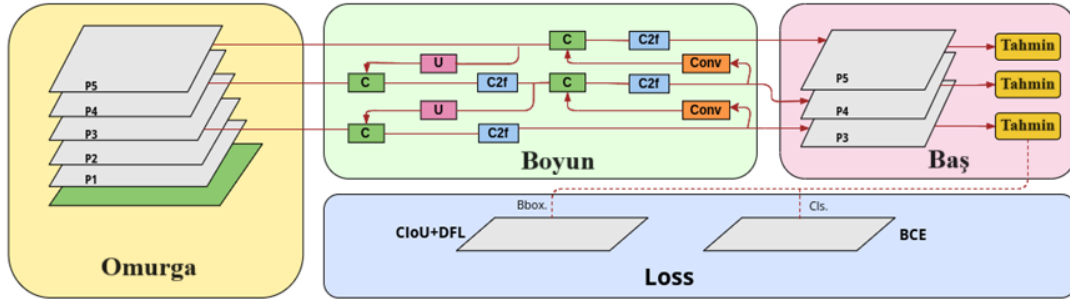
YOLOv5 modelinin farklı boyutlardaki detaylı mimarisi Şekil 3.5'te gösterilmiştir.



Şekil 3.5. YOLOv5 modelinin detaylı mimarisi.

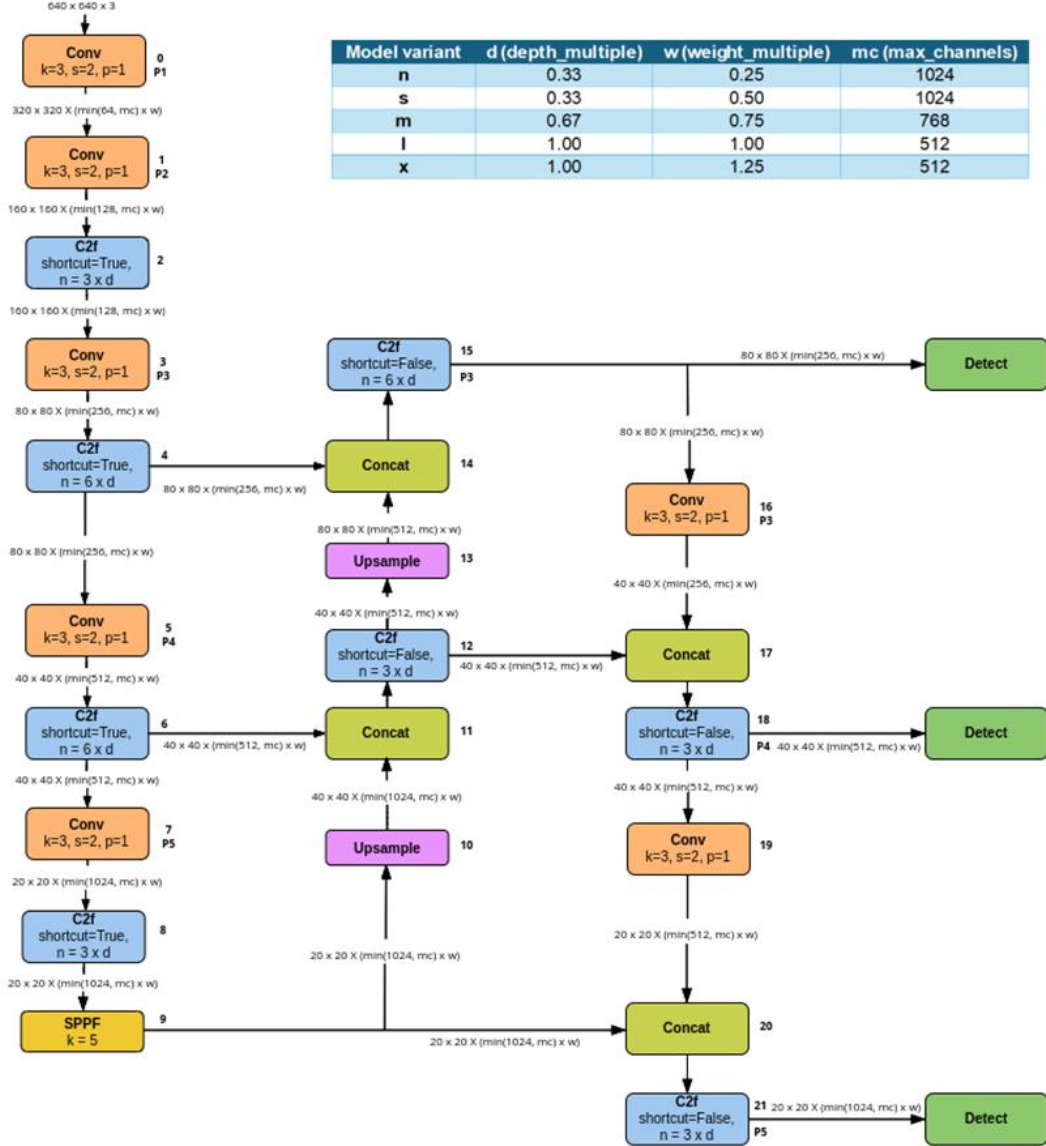
3.2.2. YOLOv8 modeli

Önceki sürümlerle karşılaştırıldığında, YOLOv8 daha hızlı ve doğru sonuç vermesinin yanında bunu daha az parametre ile gerçekleştirmektedir. YOLOv8, hız, boyut ve doğruluğa büyük önem verilerek oluşturulduğu için çeşitli görsel yapay zekâ işleri için iyi bir seçimdir (Kumari ve ark, 2023). YOLOv8 modelinin mimarisi Şekil 3.6'te gösterildiği gibi omurga, boyun ve baş kısımlarından oluşmaktadır.



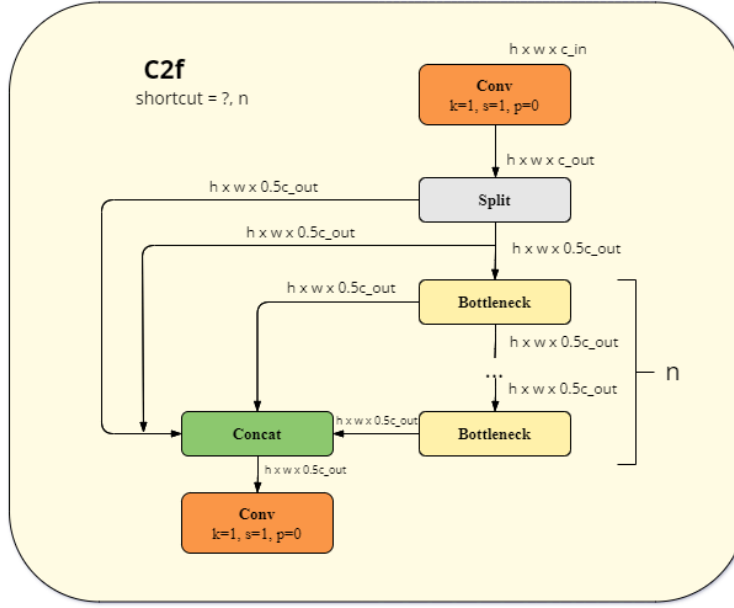
Şekil 3.6. YOLOv8 algoritmasının mimarisi.

1. Modelin omurgası, özellik haritasını iki parçaya bölmek için CSP (Wang ve ark, 2020) mimarisini kullanır. İlk bölüm, evrişim işlemlerini kullanır, ikinci bölüm ise önceki bölümün çıktılarını birleştirir. CSP mimarisi, CNN'lerin öğrenme yeteneğini geliştirir ve modelin hesaplama maliyetini azaltır. YOLOv8 modelinin ayrıntılı mimarisi Şekil 3.7'de gösterilmiştir.



Şekil 3.7. YOLOv8 model mimarisinin ayrıntılı çizimi.

YOLOv8, YOLOv5'teki C3 modülünü ve YOLOv7'nin ELAN konseptini birleştirerek modelin daha zengin gradyan akış bilgisinin elde etmesini sağlayan C2f modülünü sunar (Jiang ve ark, 2022). C3 modülü 2 Conv'ten ve n BottleNeck'ten oluşursa C2f modülü, Şekil 3.8'te gösterildiği gibi Split ve Concat yoluyla bağlanan 2 Conv ve n tane BottleNeck'ten oluşmaktadır. Burada Conv bloğu, YOLOv5'te anlattığımız gibi Conv2d, BN ve SiLU'dan oluşur (Ju ve Cai, 2023). YOLOv5'ten farklı C3 modülü yerine C2f modülünü kullanmasıdır.



Şekil 3.8. C2f bloğunun mimarisi.

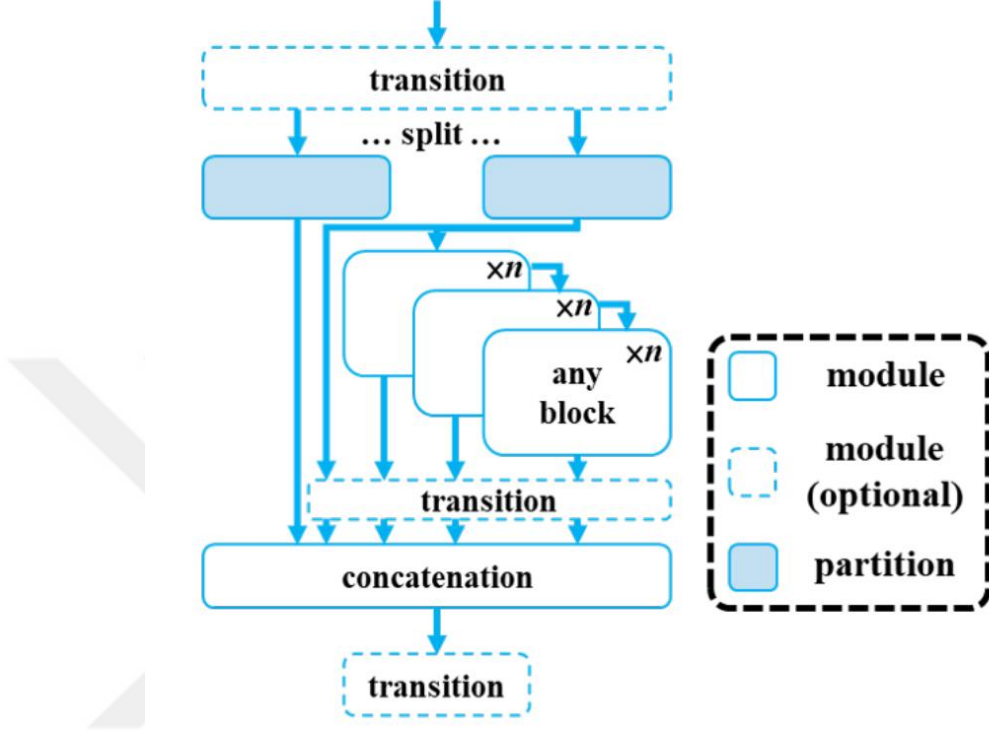
2. Boyun, omurgadaki üç farklı derinlikten gelen özellikleri birleştirmektedir. Bu birleştirilmiş özellikler daha sonra Başa iletilir. Burada da bilgi kaybını azaltmak için aynı YOLOv5 gibi PAN (Liu ve ark, 2017) ve FPN (Lin ve ark, 2018) yapılarını içermektedir. FPN alt düzey özellik haritasındaki özellik bilgilerini geliştirerek yukarıdan aşağıya doğru örnekleme gerçekleştirirken, PAN, üst düzey özellik haritasından ek bilgiler yakalayarak aşağıdan yukarıya doğru alt örnekler alır. Bu iki özellik çıktısı daha sonra farklı boyutlardaki görüntüler için doğru tahminler sağlamak amacıyla birleştirilir.

3. Modelin başı, farklı sınıflandırma ve regresyon görevlerine ayrılmış üç tespit modülünden oluşmaktadır. Bu ayrıştırma tekniği orijinal olarak YOLOX ve YOLOv6'da özellikle çapasız algılama için önerilmiştir (Terven ve ark, 2023). YOLOv5 modelinden farkı hem sınırlayıcı kutu konumuna hem de sınıflandırmaya odaklanan değiştirilmiş bir kayıp fonksiyonunun kullanılmasıdır.

3.2.3. YOLOv9 modeli

YOLOv9, 2024 yılı Şubat'ta yayınlanmış YOLO algoritmasının son versiyonlarının biridir. YOLOv9'in önceki versiyonlardan özelliği Genelleştirilmiş Verimli Katman Toplama Ağı (GELAN) ile Programlanabilir Gradyan Bilgisi (PGI) kullanılmasıdır. PGI, birleşmeyi engelleyen temel faktörlerden biri olan veri kaybını etkili bir şekilde azaltan bir kontrol mekanizmasıdır. GELAN, ağ bilgilerinin toplanmasını geliştirmek

için CSPNet ve ELAN'ı entegre ederek tasarlanmıştır, böylece nesne tespiti açısından farklı hesaplama bloklarında ve derinlik ayarlarında güçlü ve kararlı bir performans verecektir. Ayrıca bilgi kaybını da azaltacaktır. Şekil 3.9'da GELAN'ın mimarisi gösterilmiştir (Wang ve ark. 2024).



Şekil 3.9. GELAN'ın mimarisi.

3.3. Değerlendirme Metrikleri

Modelin performansı, kesinlik, duyarlılık, Birleşim Üzerinden Kesişme ve genel ortalama kesinlik (mAP) gibi ölçümler kullanılarak değerlendirilecektir.

Birleşim Üzerinden Kesişme (IoU), iki kutunun örtüşme derecesini tanımlamak için kullanılan bir terimdir. Örtüşme bölgesi ne kadar büyük olursa, IoU da o kadar büyük olur (Everingham ve ark, 2009). IoU esas olarak nesne algılamayla ilgili uygulamalarda kullanılır.

Denklem 3.1 ile temsil edilen IoU, hedef kutusu (B_{gt}) ile tahmin kutusunun (B) karşılaştırılmasını içerir.

$$IoU = \frac{|B \cap B^{gt}|}{|B \cup B^{gt}|} \quad (3.1)$$

Kesinlik (Precision), modelin olumlu örnekleri doğru şekilde tahmin etme yeteneğini değerlendirmek için kullanılır. Yapılan tüm olumlu tahminlerin doğru olumlu tahminlerin oranını hesaplar (Davis ve Goadrich, 2006). Daha yüksek bir kesinlik puanı, daha az sayıda yanlış pozitif anlamına gelir; bu da modelin gerçek pozitifleri doğru şekilde tanımlamada daha etkili olduğunu gösterir. Kesinlik hesaplama formülü Denklem 3.2'de verilmiştir.

$$Precision = \frac{TP}{TP + FP} \quad (3.2)$$

Duyarlılık (Recall), bir modelin toplam gerçek olumlu örneklerden tüm olumlu örneklerini doğru bir şekilde tanımlama yeteneğinin bir ölçüsüdür. Gerçek pozitif tahminlerin tüm gerçek pozitif vakalar içindeki oranını hesaplar (Davis ve Goadrich, 2006). Daha yüksek bir hatırlama puanı, daha az sayıda yanlış negatif anlamına gelir; bu da modelin tüm olumlu örnekleri belirlemede daha etkili olduğunu gösterir. Duyarlılık hesaplanmasının formülü Denklem 3.3'te verilmiştir.

$$Recall = \frac{TP}{TP + FN} \quad (3.3)$$

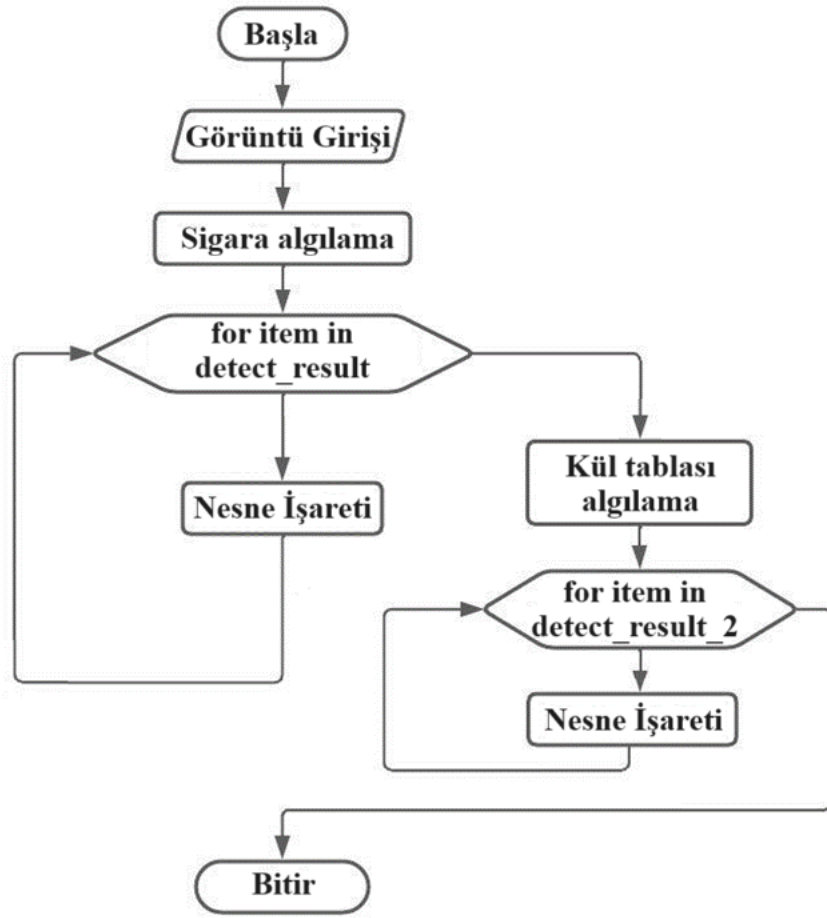
Genel ortalama kesinlik (mAP), modelin etkinliğini nesne algılama doğruluğu açısından değerlendiren bir ölçümdür. Farklı hatırlama seviyelerindeki kesinlik puanlarının ortalamasını temsil eder. Denklem 3.4 kullanılarak hesaplanır (Padilla ve ark, 2020).

$$mAP = \frac{1}{N} \sum (precision \text{ at each recall level}) \quad (3.4)$$

mAP, IoU eşiği 0,5 olarak ayarlandığında ortalama hassasiyet değerini ifade eden mAP(50) olarak temsil edilir. Ek olarak mAP(50-95), 0,5 ile 0,95 arasındaki IoU eşik aralığını belirtmek için kullanılır.

3.4. Algoritma

Çalışmada sigara tespitin doğruluk oranını yükseltmek için kendi algoritmamızı geliştirdik. Aşağıdaki Şekil 3.10'de gösterildiği gibi algorithmada ilk olarak sigara tespiti yapılacaktır, sonra eğitilen ikinci model ile kül tablasın arayacaktır, eğer kül tablası olursa bulduğu nesneyi sigara olarak tespit edecektir, böylece kül tablasın bularak ilk modelin doğruluk oranını yükseltecektir.



Şekil 3.10. Algoritma.

Algoritmada nesne tespiti için iki tane model kullanılmaktadır. Birinci model sigarayı tespit edecektir, ikinci model kül tablasını bulacaktır. Modeller YOLO algoritmalarının kullanılarak yapılmıştır.



4. BULGULAR VE TARTIŞMA

4.1. Gerekli Kütüphanelerin Kurulması

Bu tez çalışmasında; Ultralytics YOLOv8.2.8, Python-3.10.6, torch-2.0.1+cu117 kütüphane ve yazılımları, NVIDIA GeForce RTX 3060, 12GB grafik işlemci, 8 CPUs, 31.9 GB RAM'e sahip bilgisayar kullanılmıştır. Video görüntüleri üzerindeki sigara nesnesini tanıma ve sansürleme için OpenCV görüntü işleme kütüphanesi kullanılmıştır.

4.2. Modellerin eğitilmesi

Bu çalışmada nesne tanıma için YOLOv5, YOLOv8 ve YOLOv9 modelleri kullanılmıştır. Tablo 4.1'de varsayılan parametreler ile eğitilmiş farklı boyutlardaki YOLOv5, YOLOv8 ve YOLOv9 modellerinin sonuçları verilmiştir. YOLOv5 modelinin YOLOv8 ile YOLOv9 modellerinden daha yüksek doğruluk gösterdiği görülmektedir. YOLOv5 %89 ile en yüksek doğruluk oranını verirken, YOLOv8 %87 ve YOLOv9 %88 doğruluk oranını vermektedir. Yani YOLOv5 modeli, YOLOv8'den %2'ye ve YOLOv9'dan %1'e daha iyi performans göstermektedir.

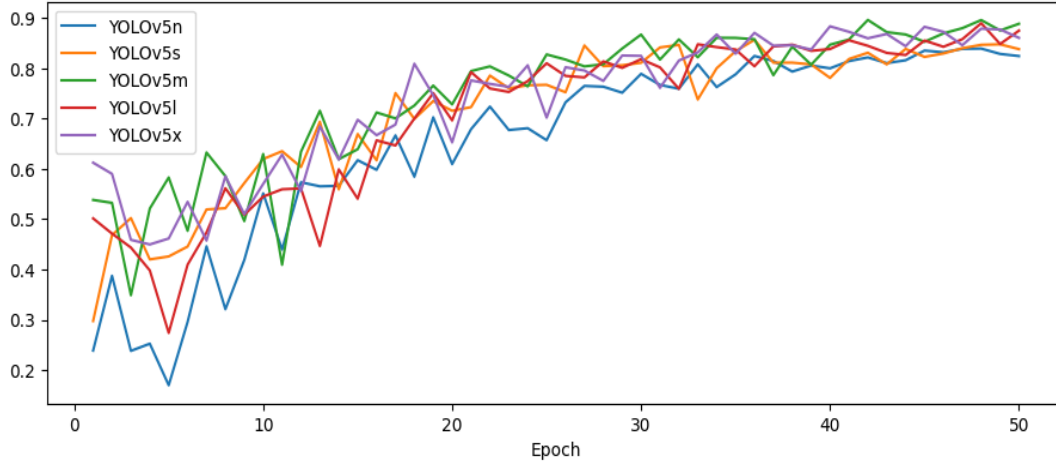
Tablo 4.1. Eğitilen modellerimizin sonuçları.

Model	Parametre	mAP(50)	mAP(50-95)
YOLOv5n	2508659	0,83	0,41
YOLOv5s	9122579	0,85	0,40
YOLOv5m	25065715	0,89	0,43
YOLOv5l	53164115	0,89	0,42
YOLOv5x	97200371	0,88	0,41
YOLOv8n	3011043	0,82	0,38
YOLOv8s	11135987	0,87	0,42

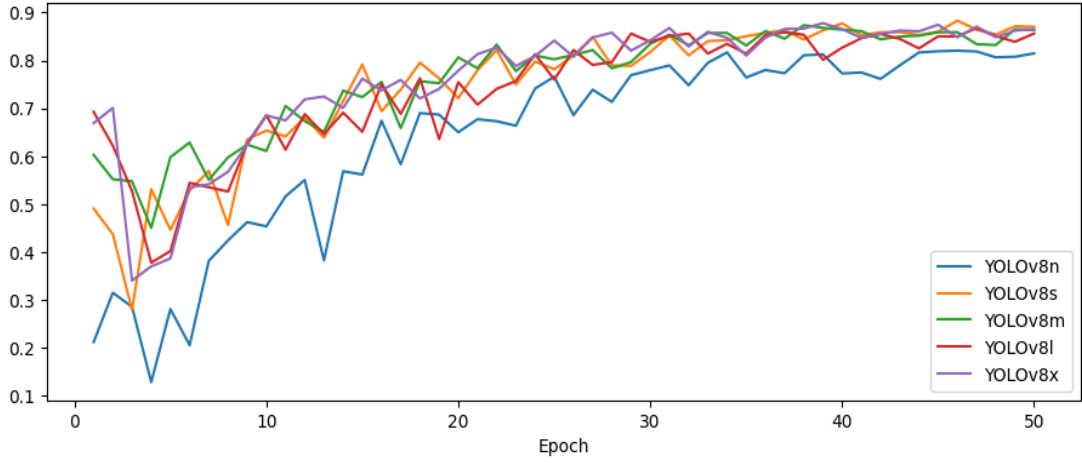
Tablo 4.1. (Devamı): Eğitilen modellerimizin sonuçları.

Model	Parametre	mAP(50)	mAP(50-95)
YOLOv8m	25856899	0,85	0,42
YOLOv8l	43630611	0,87	0,42
YOLOv8x	68153571	0,87	0,42
YOLOv9t	1970979	0,80	0,37
YOLOv9s	7167475	0,84	0,40
YOLOv9m	20013715	0,88	0,42
YOLOv9c	25320019	0,88	0,42
YOLOv9e	57377171	0,87	0,42

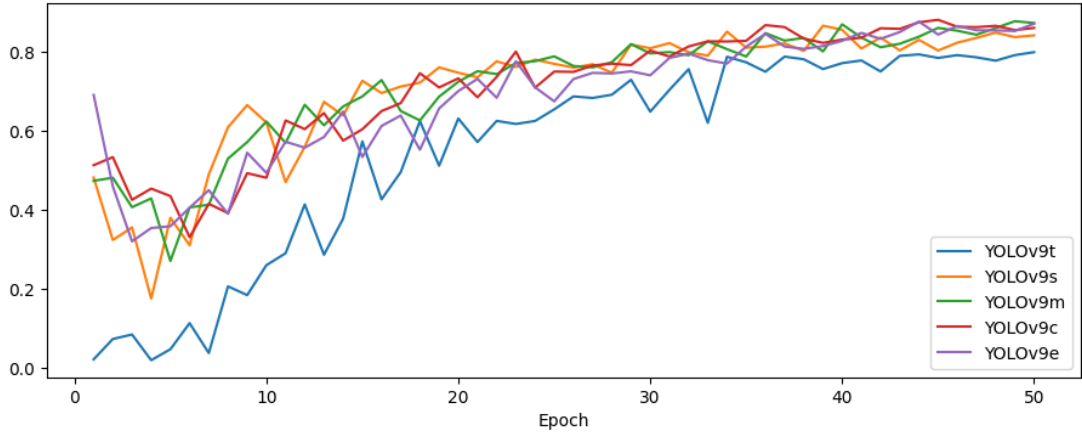
Şekil 4.1 - 4.3, deneysel ortamında eğitilen YOLOv9, YOLOv8 ve YOLOv5 modellerinin mAP(50) performansını göstermektedir.



Şekil 4.1. 50 epok boyunca YOLOv5 eğitimi ile IoU=0,50 olan mAP.

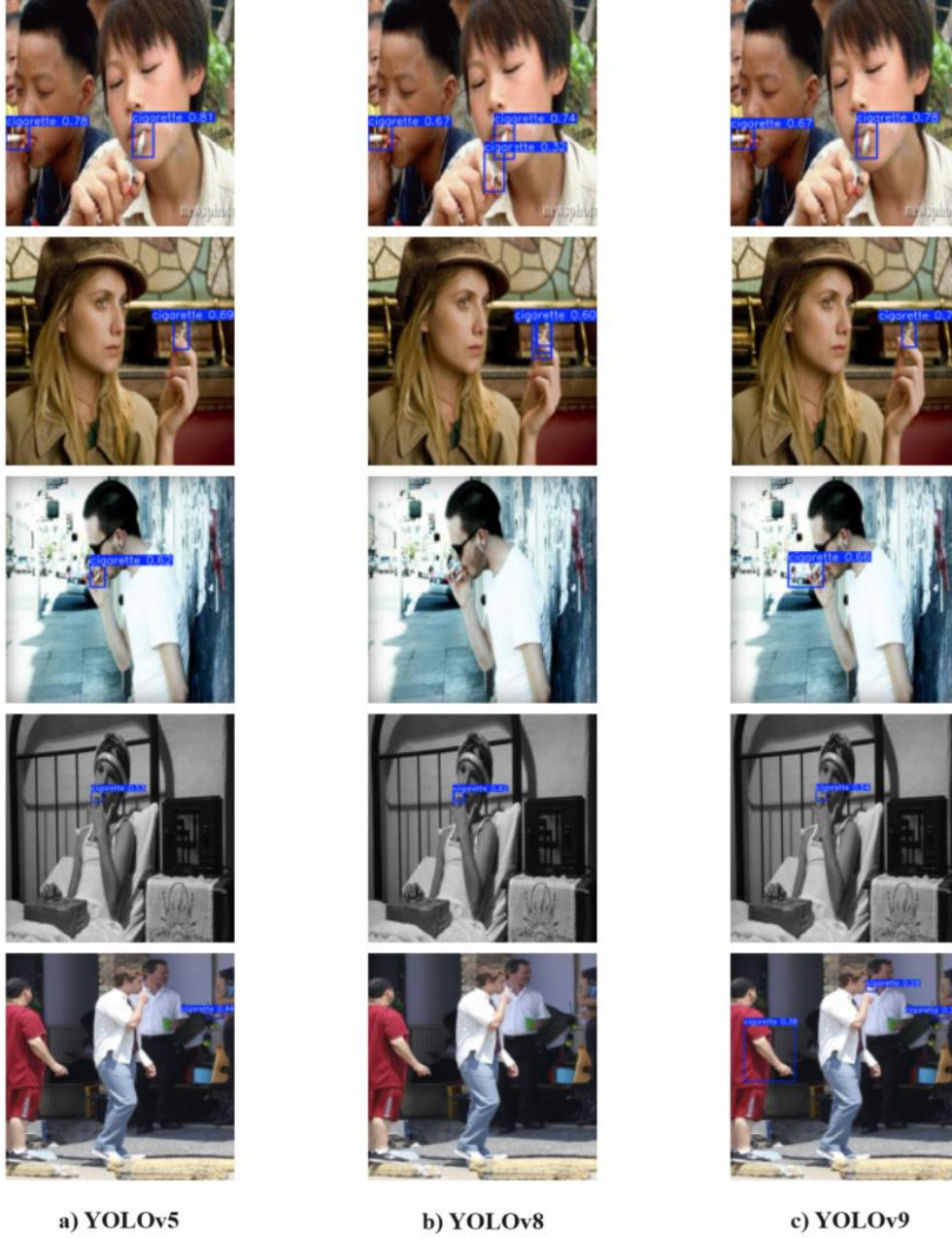


Şekil 4.2. 50 epok boyunca YOLOv8 eğitimi ile IoU=0,50 olan mAP.



Şekil 4.3. 50 epok boyunca YOLOv9 eğitimi ile IoU=0,50 olan mAP.

Aynı veri seti ile eğitilen modellerin, test sonuçlarından 5 tanesi Şekil 4.4'te sunulmuştur. YOLOv5 modelinin en iyi sonuç verdiği görülmektedir. Bu model 4 tane görüntüdeki nesneleri doğru olarak tespit etmiş, ama 5. Görüntüdeki sigaralardan bir tanesini tanımamıştır. YOLOv8 ise 3. ve 5. görüntülerdeki sigaraları tespit edememiş, 2. ve 4. Görüntülerdeki bir sigara nesnesini birkaç defa işaretlemiştir. YOLOv9 ise 5 görüntüdeki sigara nesnelerini tespit etmiş, ancak beşinci görüntüde 1 tane sigara olmayan nesneyi sigara olarak etiketlemiştir.



Şekil 4.4. Önerilen modelin sonuçları.

4.2.1. Hiper-parametre optimizasyonu

Çalışmada, tanıma doğruluğunu ve hızını artırmak için YOLOv5, YOLOv8 ve YOLOv9 modellerinin hiper parametreleri çeşitli optimize ediciler kullanarak optimize edilmiştir. Hiper parametre optimizasyonu, derin öğrenme modellerinin performansını iyileştirmede önemli adımlardan biridir (Wahyudi ve ark. 2023). Doğru optimize edici seçimi, yakınsama hızını ve nihai sonucun kalitesini önemli ölçüde etkilemektedir.

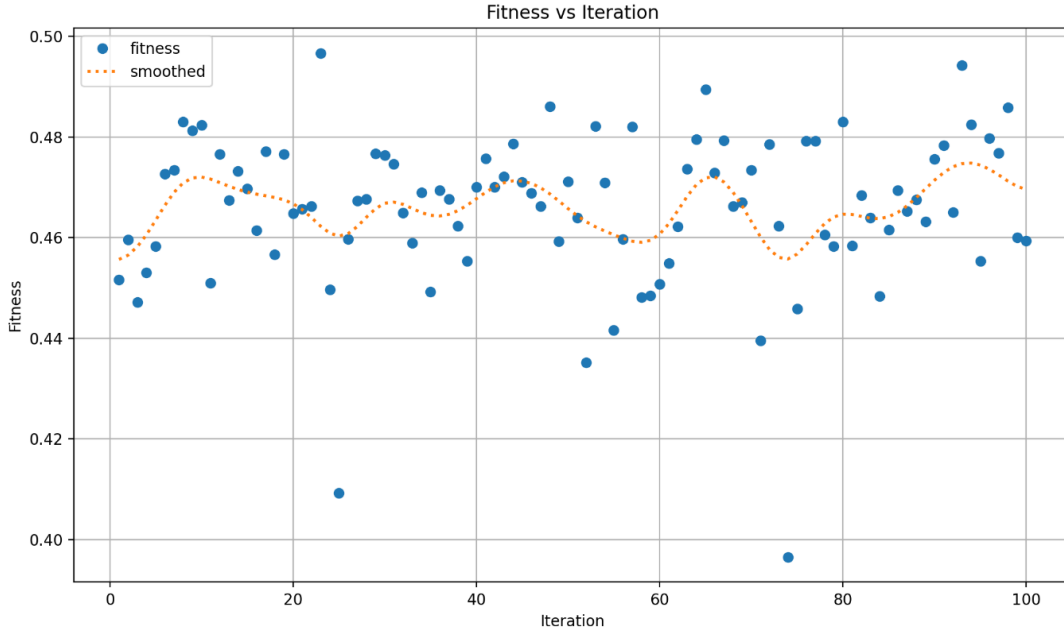
Sonuçları iyileştirmek için YOLOv5, YOLOv8 ve YOLOv9'da kullanılan standart optimize edicileri SGD, Adam ve AdamW gibi daha verimli yöntemlerle değiştirilmiştir.

Hiper parametre optimizasyonu, görüntülerdeki nesne tanıma doğruluğunu önemli ölçüde artırılmasını, sonuçların iyileştirilmesini ve sigaraların daha güvenilir ve verimli bir şekilde algılanmasını mümkün kılmıştır. Hiper parametre optimizasyonunun sonuçları Tablo 4.2'de verilmiştir. Bu sonuçlara göre yüksek performansı SGD optimize ediciyi kullanan YOLOv9 modeli vermektedir. YOLOv8 ile YOLOv9'ın mAP(50) değerleri aynı olup %93 dür. YOLOv9'un mAP(50-95) değeri %51 olup YOLOv8'den %2 daha iyi performans göstermiştir. Hiper parametre optimizasyonu uygulanmasından sonra, doğruluk oranı YOLOv9'da %88'den %93'e, YOLOv8'de %87'den %93 yükselmiştir. YOLOv5'in doğruluk oranı değişmemiş, yani %89 olan önceki doğruluk oranında sonuç vermiştir.

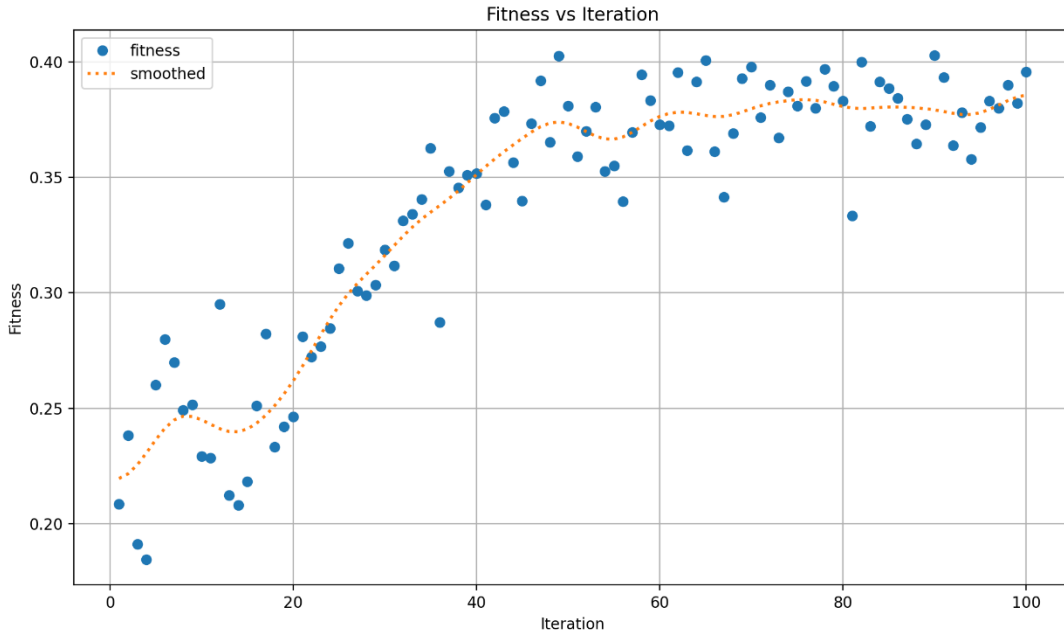
Tablo 4.2. Modellerinin farklı optimize ediciler ile sonuçları.

Model	Optimize edici	mAP(50)	mAP(50-95)	Recall	Precision
YOLOv9	SGD	0,93	0,51	0,87	0,94
YOLOv9	Adam	0,89	0,44	0,81	0,91
YOLOv9	AdamW	0,87	0,42	0,79	0,92
YOLOv8	SGD	0,93	0,49	0,88	0,96
YOLOv8	Adam	0,89	0,43	0,81	0,86
YOLOv8	AdamW	0,85	0,4	0,76	0,85
YOLOv5	SGD	0,89	0,43	0,84	0,95
YOLOv5	Adam	0,79	0,36	0,71	0,80
YOLOv5	AdamW	0,76	0,34	0,67	0,82

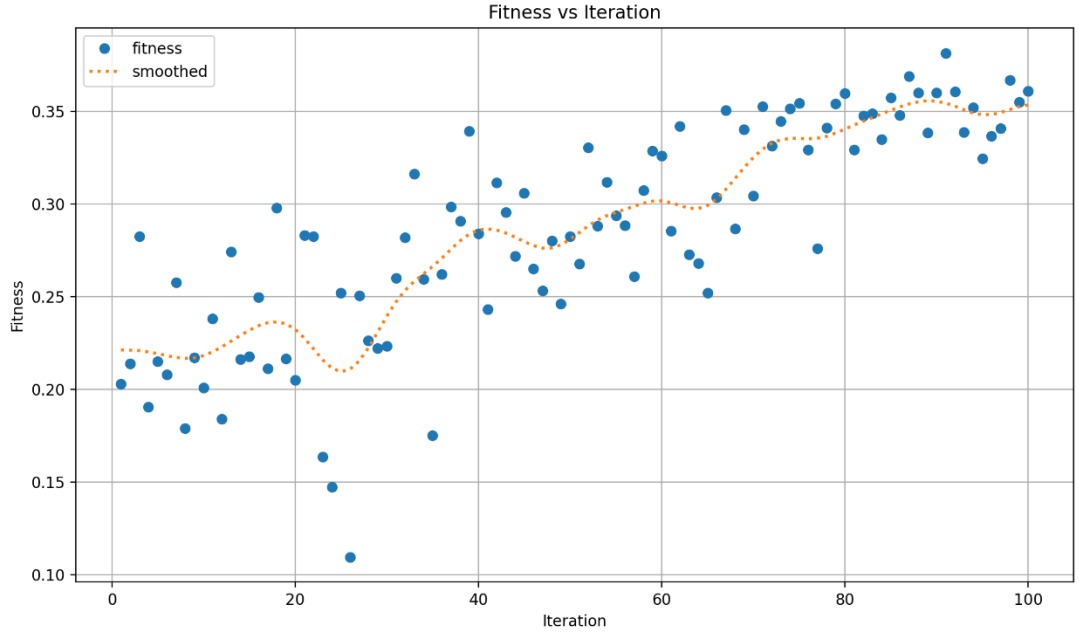
YOLOv5, YOLOv8 ve YOLOv9 modellerinin SGD, Adam ve AdamW optimize edicilerle 100 jenerasyon üzerinde yapılan optimizasyonların her jenerasyondaki doğruluk oranı Şekil 4.5 - 4.13'te gösterilmiştir.



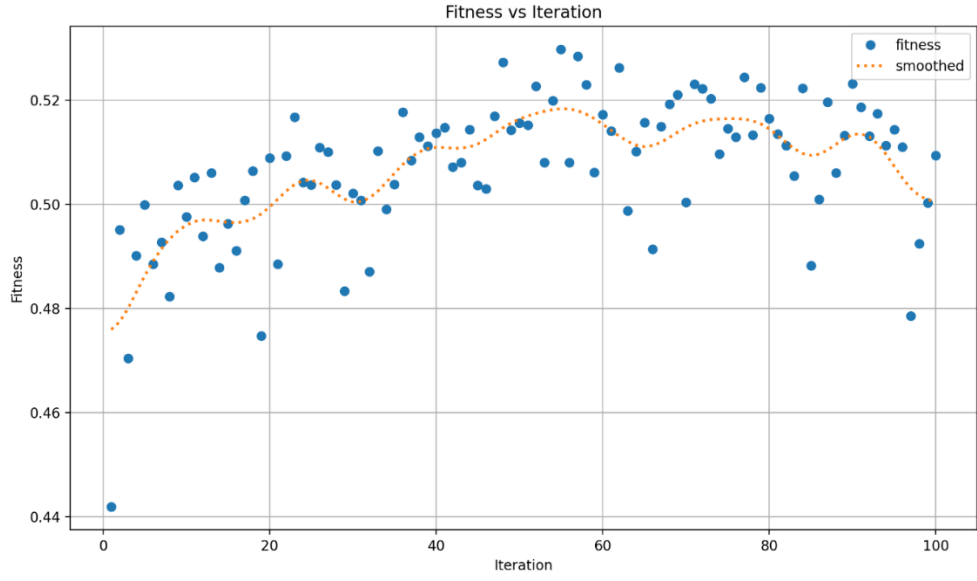
Şekil 4.5. YOLOv5 modeli SGD optimize ediciyle.



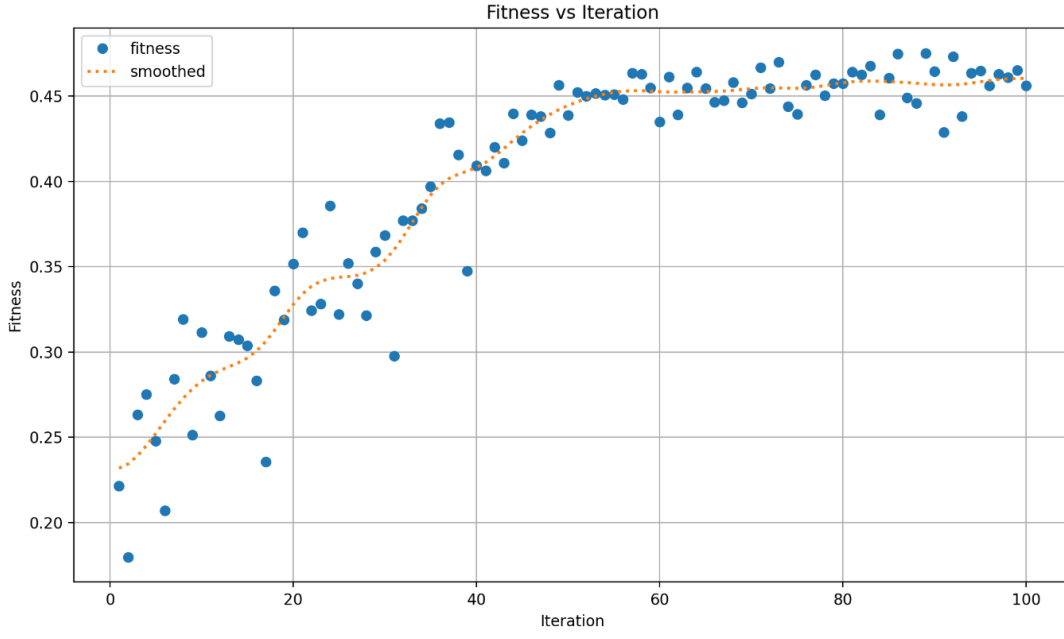
Şekil 4.6. YOLOv5 modeli Adam optimize ediciyle.



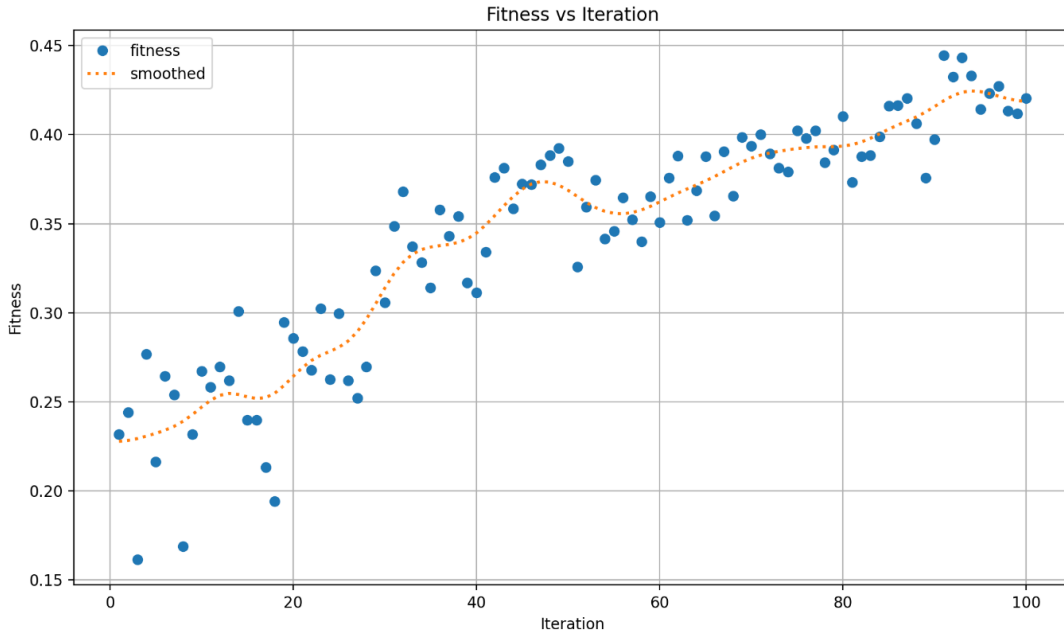
Şekil 4.7. YOLOv5 modeli AdamW optimize ediciyle.



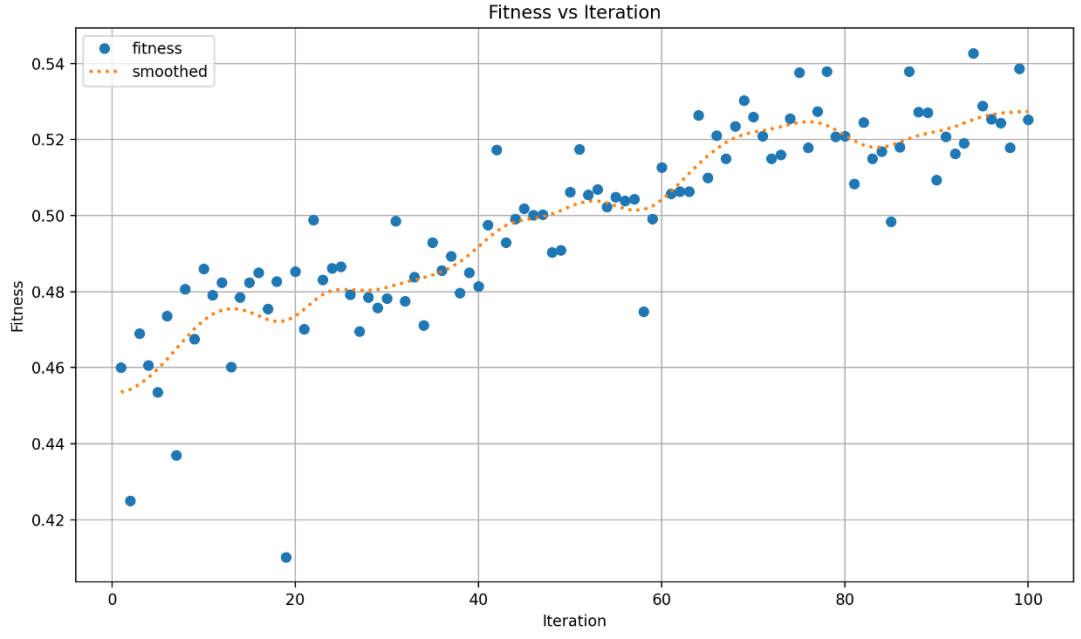
Şekil 4.8. YOLOv8 modeli SGD optimize ediciyle.



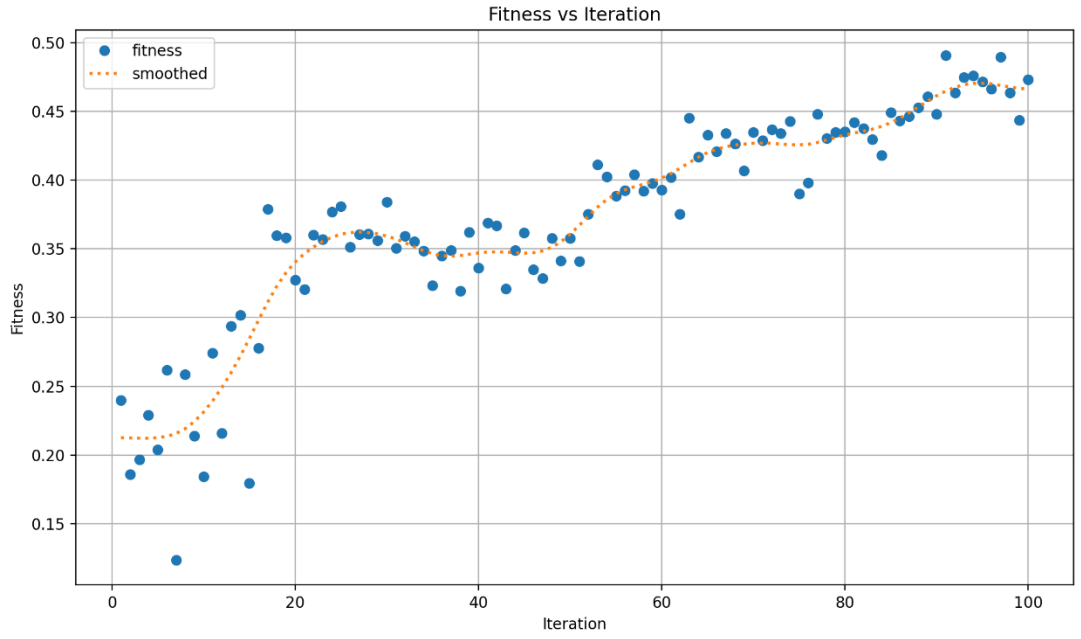
Şekil 4.9. YOLOv8 modeli Adam optimize ediciyle.



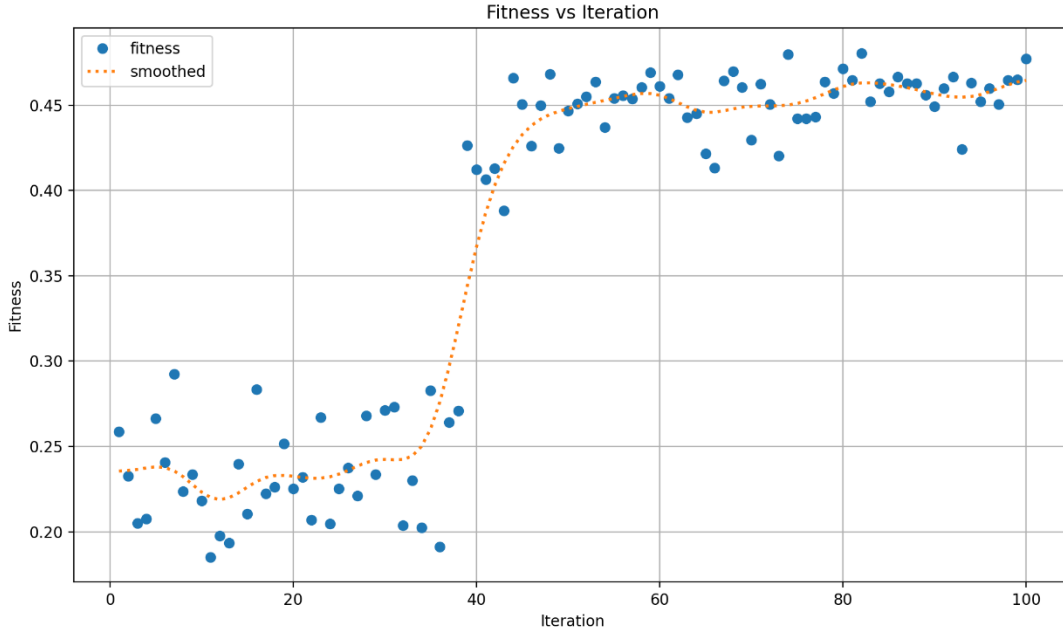
Şekil 4.10. YOLOv8 modeli AdamW optimize ediciyle.



Şekil 4.11. YOLOv9 modeli SGD optimize ediciyle.

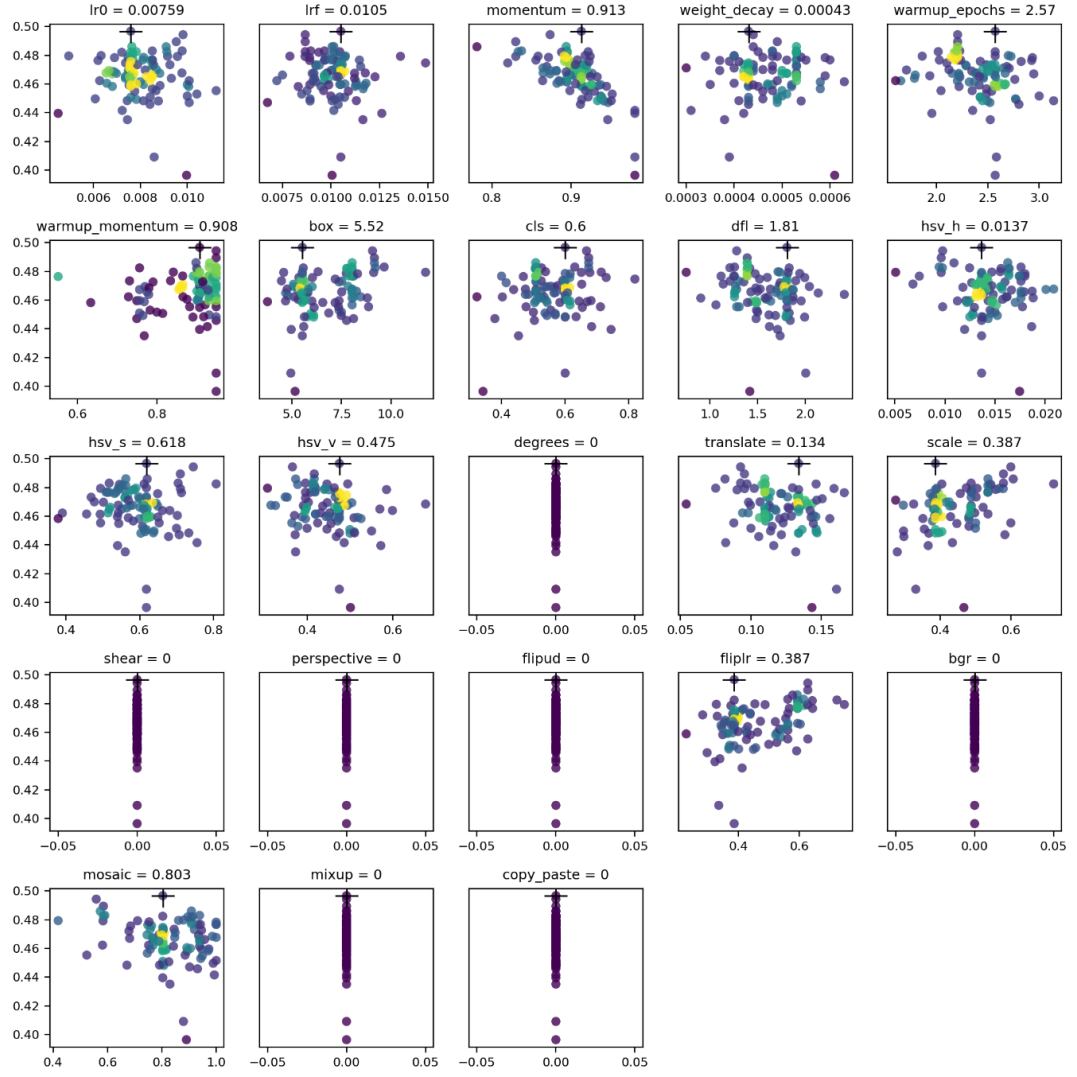


Şekil 4.12. YOLOv9 modeli Adam optimize ediciyle.

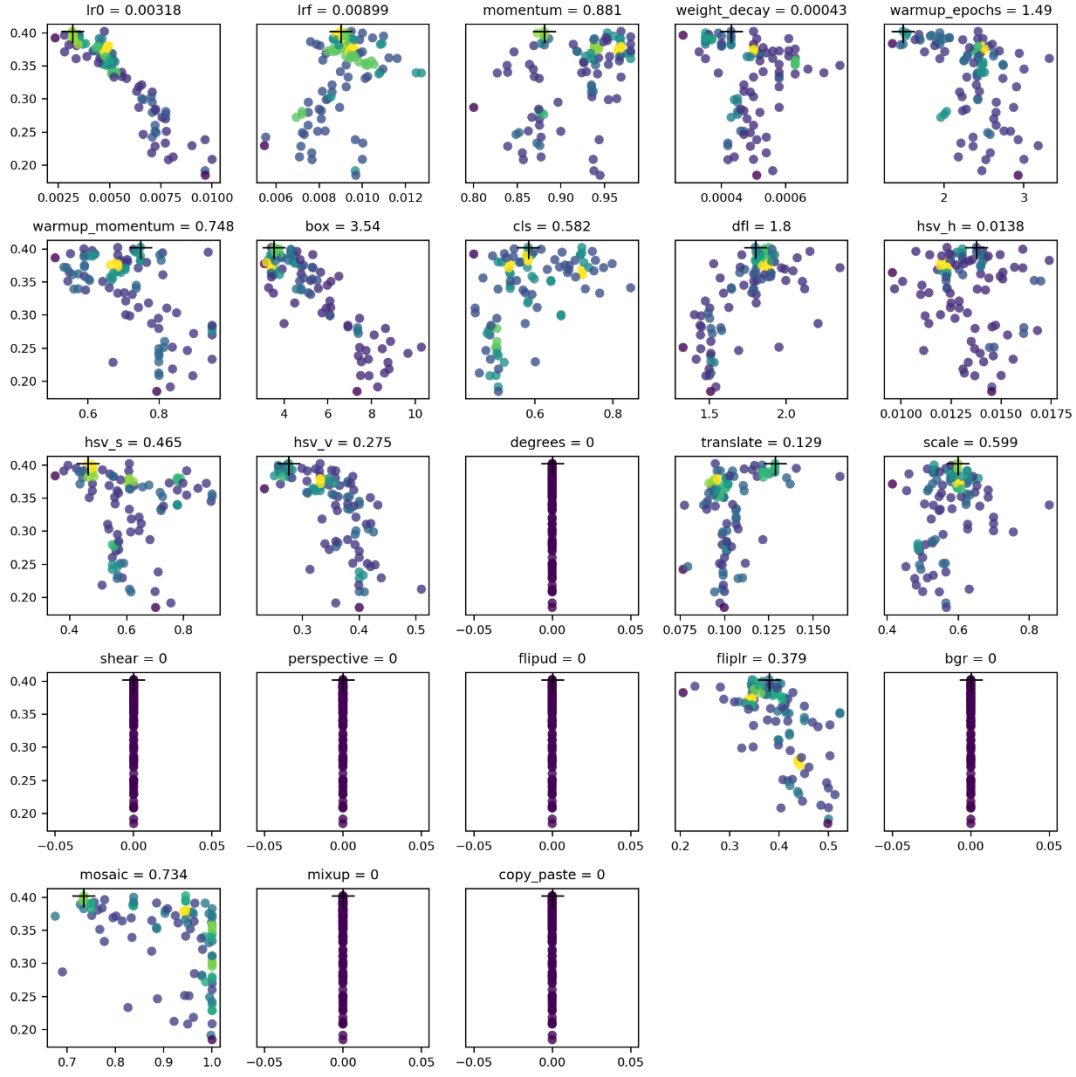


Şekil 4.13. YOLOv9 modeli AdamW optimize ediciyle.

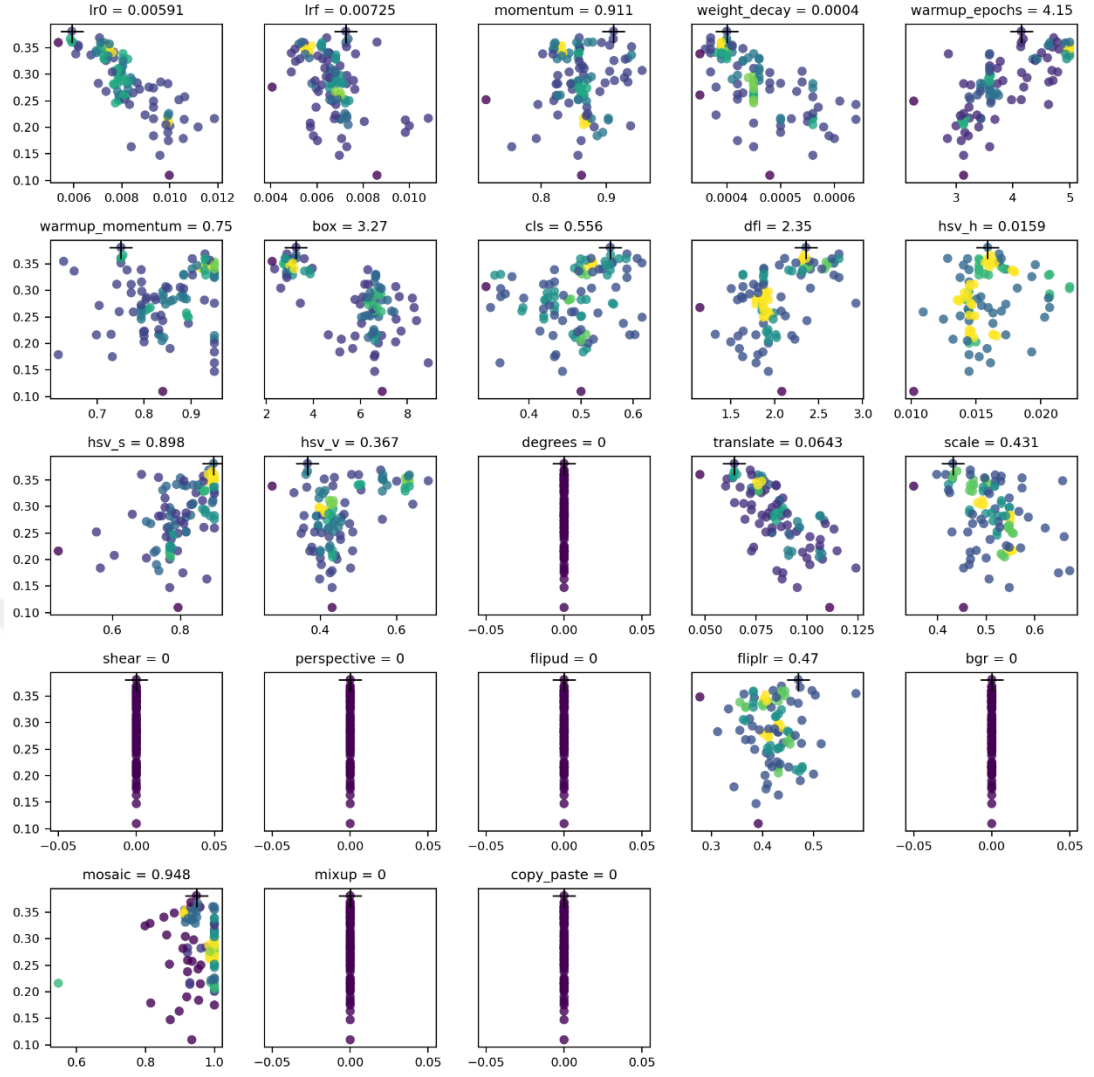
Optimizasyonun sonuçlarını görselleştirmek için, her optimize edicinin doğruluk, yakınsama oranı ve kayıp gibi temel model ölçümleri üzerindeki etkisini görselleştirmek amacıyla dağılım grafikleri oluşturulmuştur. Dağılım grafiklerini ile analiz, farklı optimize edicilerin ve parametrelerin model performansı üzerindeki etkisini keşfetmenize olanak tanıyarak, en iyi sigara algılama sonuçlarını elde etmek için iyi ayarları seçmeye yardımcı olmaktadır. YOLOv5, YOLOv8 ve YOLOv9 modellerinin farklı optimize edicilerle dağılım grafikleri Şekil 4.14 – 4.22 gösterilmektedir.



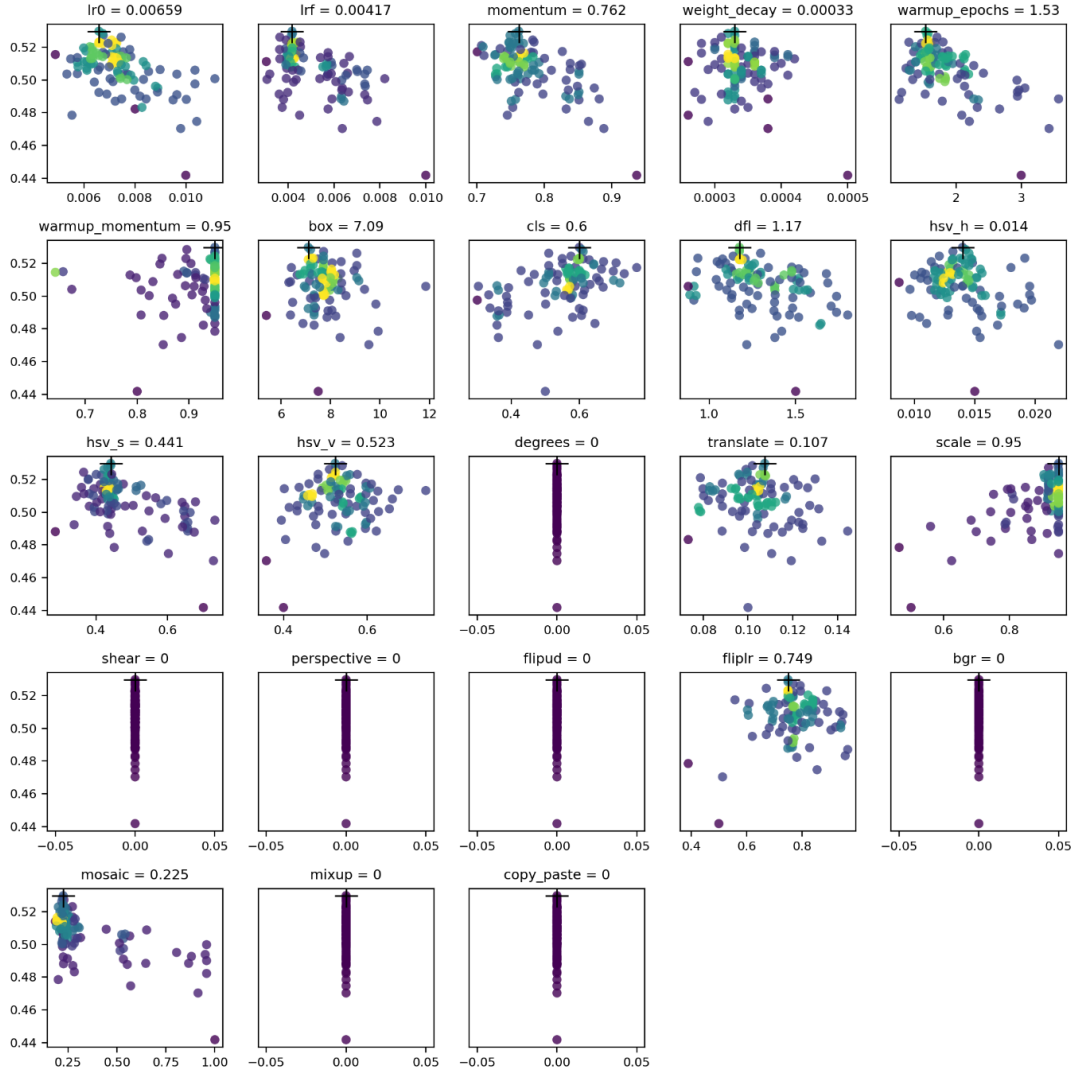
Şekil 4.14. YOLOv5 modelin SGD optimize ediciyle hiper-parametre optimizasyonu grafikleri.



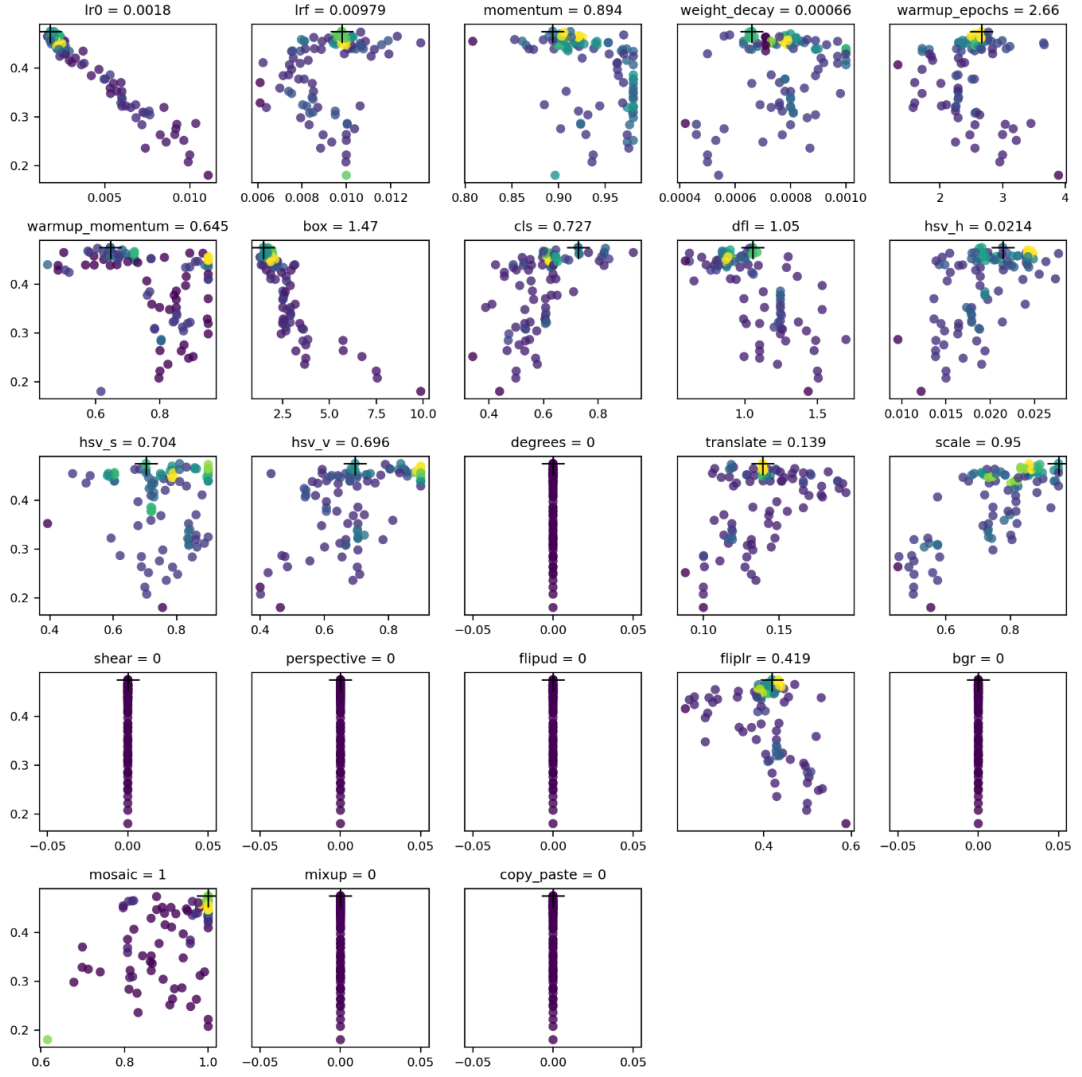
Şekil 4.15. YOLOv5 modelin Adam optimize ediciyle hiper-parametre optimizasyonu grafikleri.



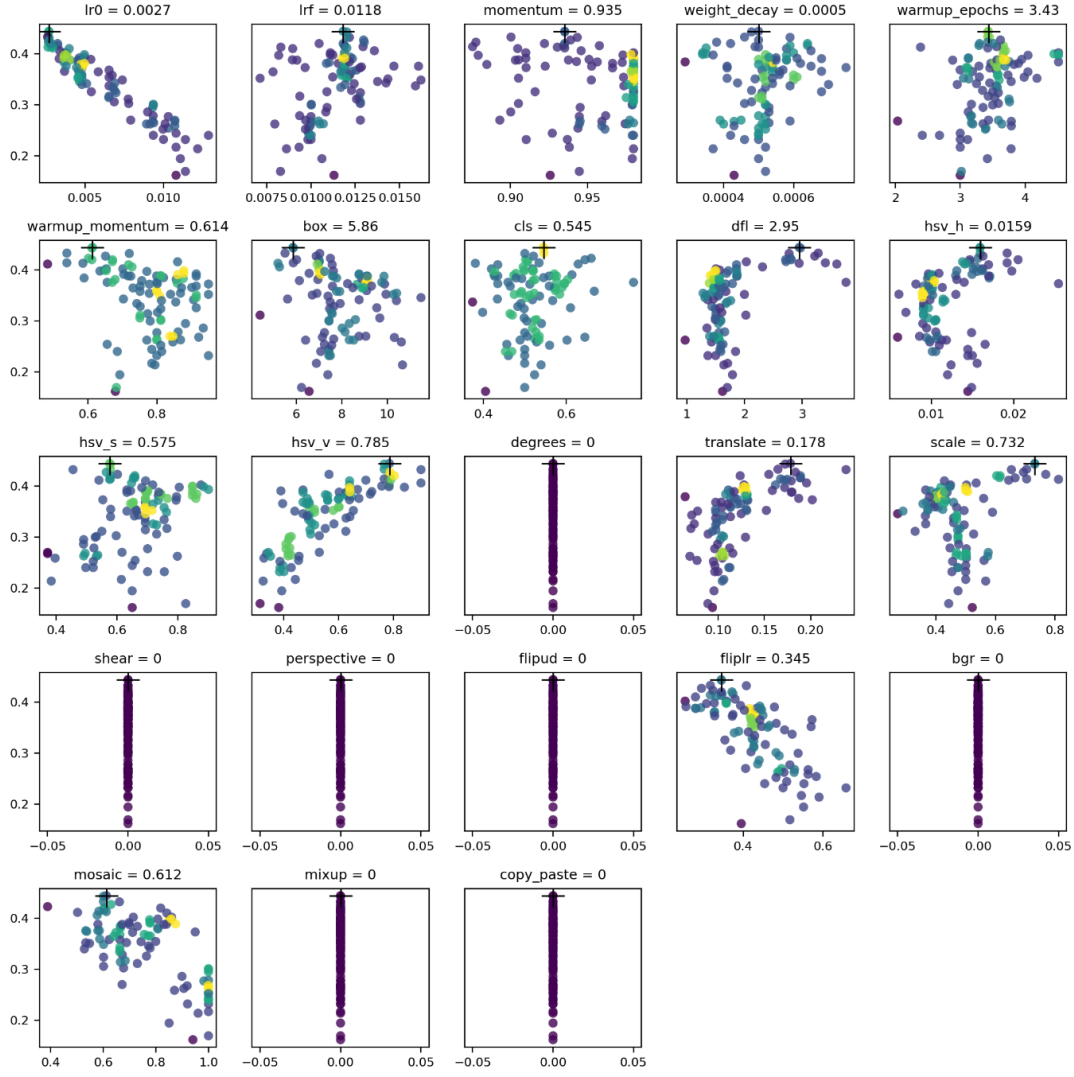
Şekil 4.16. YOLOv5 modelin AdamW optimize ediciyle hiper-parametre optimizasyonu grafikleri.



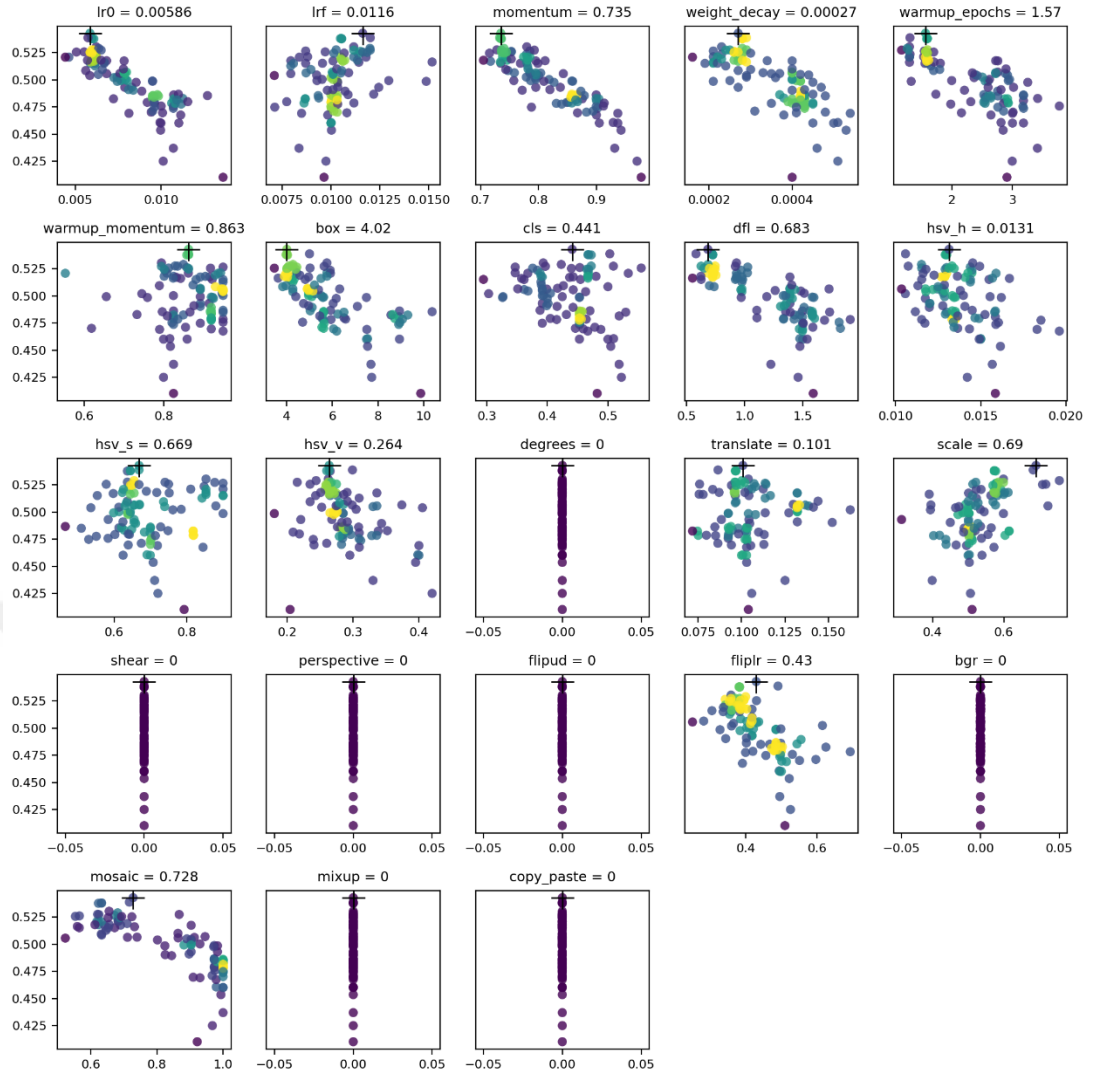
Şekil 4.17. YOLOv8 modelin SGD optimize ediciyle hiper-parametre optimizasyonu grafikleri.



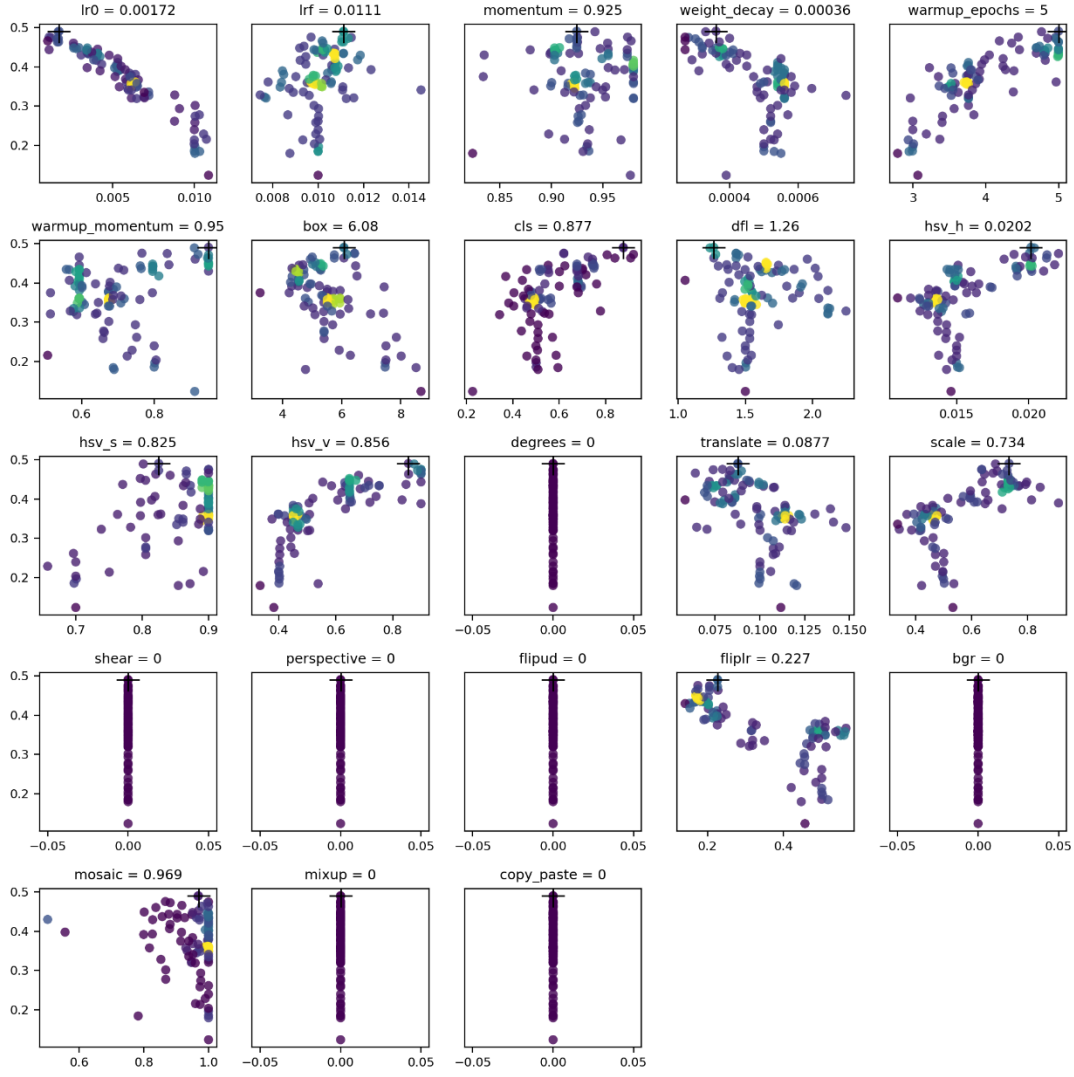
Şekil 4.18. YOLOv8 modelin Adam optimize ediciyle hiper-parametre optimizasyonu grafikleri.



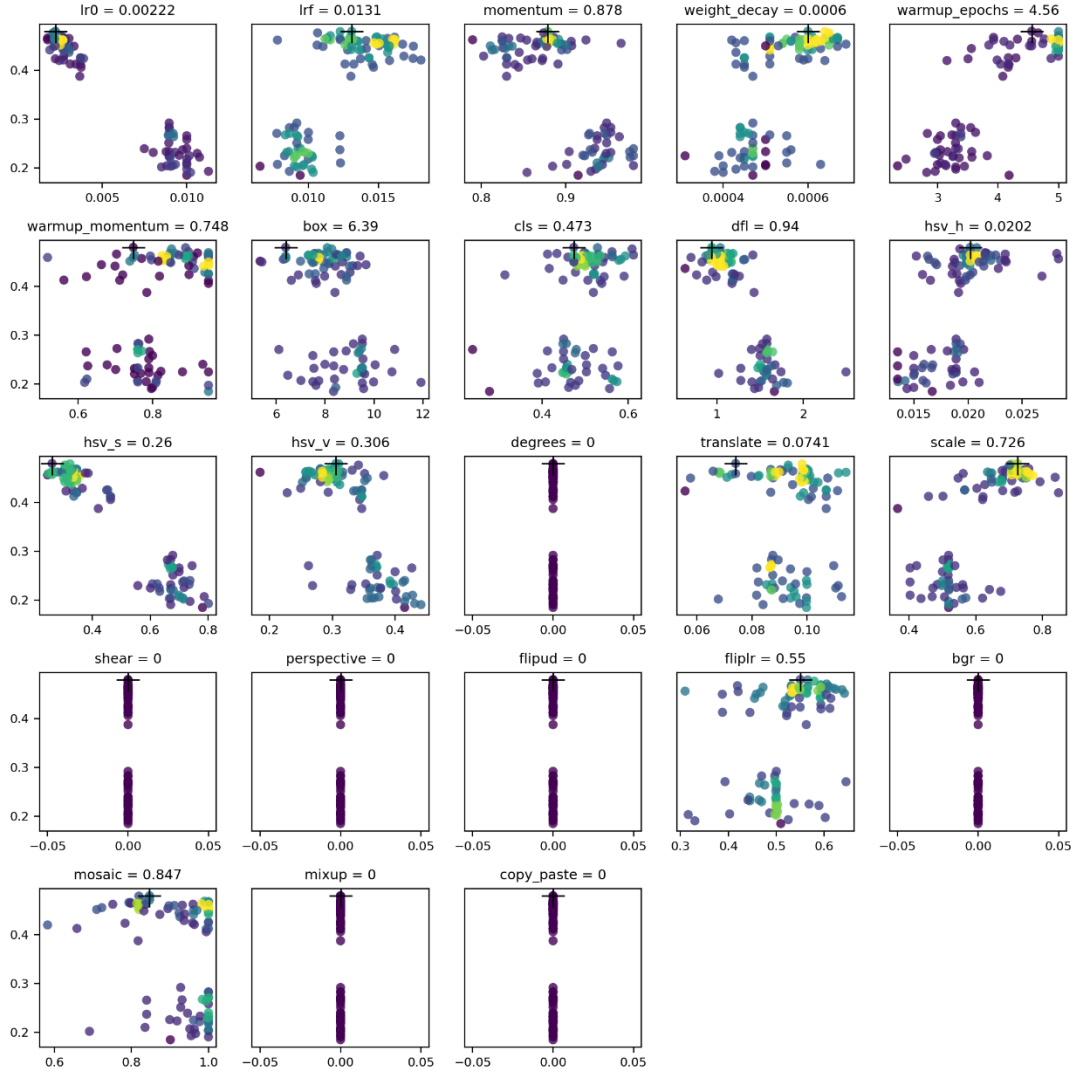
Şekil 4.19. YOLOv8 modelin AdamW optimize ediciyle hiper-parametre optimizasyonu grafikleri.



Şekil 4.20. YOLOv9 modelin SGD optimize ediciyle hiper-parametre optimizasyonu grafikleri.



Şekil 4.21. YOLOv9 modelin Adam optimize ediciyle hiper-parametre optimizasyonu grafikleri.



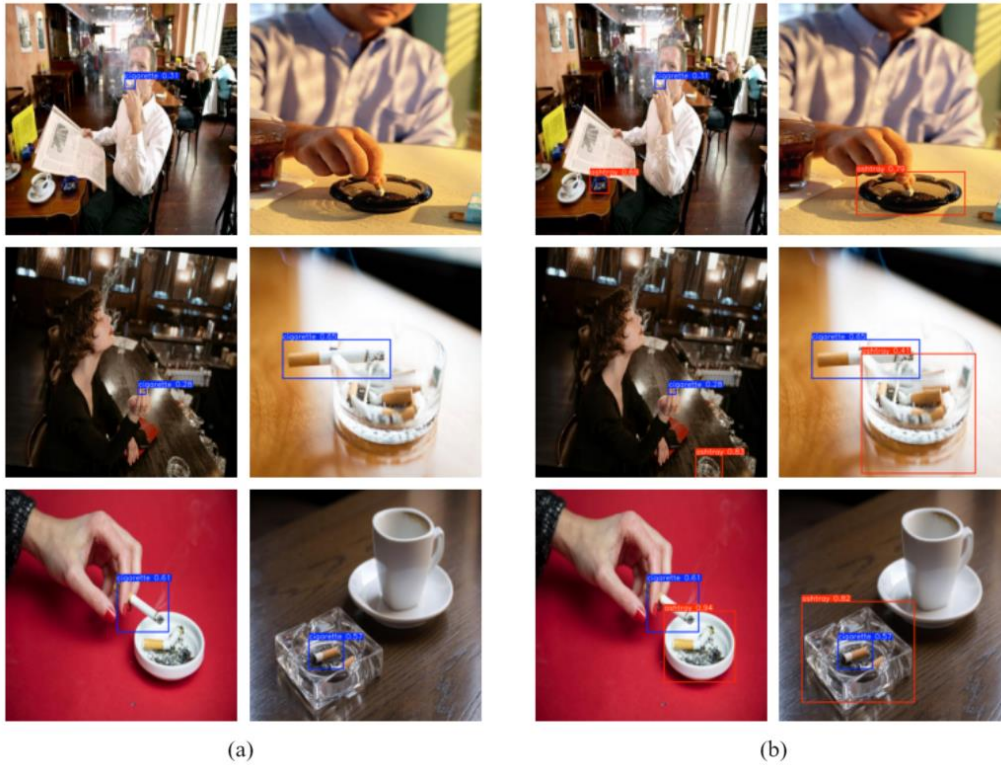
Şekil 4.22. YOLOv9 modelin AdamW optimize ediciyle hiper-parametre optimizasyonu grafikleri.

Hiper parametre kombinasyonlarının dikkatli bir şekilde ayarlanması ve test edilmesinden sonra, SGD optimize ediciyle sahip YOLOv9 versiyonu diğer versiyonlardan daha iyi sonuçlar vermiştir. Büyük hacimli verilerle ilgili problemlerin çözümünde kararlılığı ve verimliliği ile bilinen SGD optimizieri (Seyyarer ve ark, 2020), modelin yüksek tanıma doğruluğuna ve daha iyi genelleme yeteneğine ulaşmasını sağlamıştır. SGD'nin kullanımı, gerçek dünya uygulamalarında modelin güvenilirliğini artırmaktadır. Kayıp fonksiyonunun hızlı yakınsamasını ve en aza indirilmesini sağlamaktadır (Li ve ark, 2023).

4.3. Görüntüleri karşılaştırma

Modeller sadece sigara nesneleri için eğitildikten sonra sonuçlar analiz ederek eksiklikleri araştırılmıştır. Modeller, sigaranın tamamının görünmemesi ya da kül tablasındaki sigaraların küçük ve karmaşık arka plan içerisinde olmasından dolayı tespit edilmesinde zorluk yaşamaktadır. Bu yüzden, çalışmada sigara tanımayı artırmak için kül tablasının da tespit edilmesinin etkileri de incelenmiştir.

Şekil 4.23'te kül tablasının etkisinin incelendiği modellerden yüksek performans veren YOLOv9 modellerinin sonuçları gösterilmektedir. Burada görüntüler ikiye ayrılmıştır. Birincide kül tablası olmadan sadece sigara tespit edilmektedir. Ekincide ise sigara tanımada kül tablası tanınmanın etkisi sunulmaktadır. Şekil 4.23'te modelin, kül tablası tanıma eklenmediğinde, bazen kül tablasındaki sigaraları algılayamadığı görülmektedir. Bu sorun kül tablası tanımayı eklediğimizde iyileştirilmiştir. Yani, yanlış pozitif sayısı düşürülmüş, bunun sonucunda doğruluk oranı 93%'ten 96% yükselmiştir.



Şekil 4.23. Optimize edilen YOLOv8 modelinin sonucu.

Model eğitilip test edildikten sonra, video üzerinde sigara ve kül tablasını tespit edilmiş ve sansürlenmiştir. Şekil 4.24, videodaki sigara sahnelerini filtrelemek için YOLOv9 üzerine kurulu bir sigara ve kül tablasını algılama modelinin çıktısını

göstermektedir. Birinci satırdaki resimlerde, sigara içilen sahnelerin görülebildiği filmlerden orijinal fotoğraflar, Sol üst resimde, bir fincan kahvenin ve kül tablasının yanında sigara içen bir adam görüyoruz. Sol orta resimde, biri diğerine sigara uzatan iki kişiyi ve sol alt resimde, sandalyede oturup kulaklıkla müzik dinlerken sigara içen bir adam göstermektedir.

İkinci satırda aynı sahneler YOLOv9 modeli kullanılarak filtrelendikten sonra gösterilir. Bu resimlerde sigara ve kül tablası gizlenmiş ve sahnelerde sigara içme unsurları sansürlenmiştir. Örneğin sağ üst karede elinde sigara bulunduran kişinin elindeki sigara ve masadaki kül tablası sansürlenmiştir. Örnek sonuçların verildiği bu şekil, eğitilen modelin sigara içilen sahnelerdeki sigaraları başarılı ve etkili bir şekilde tespit edip kaldırabildiğini göstermektedir. Böylece filmlerdeki sigara kullanımını ve alışkanlığını yaygınlaştırma etkisi olan sahnelerin izleyenler için uygun ve zararsız hale getirilmesi sağlanmaktadır.



Şekil 4.24. YOLOv9 modelinin video üzerinde sansürleme sonuçları.



6. SONUÇ VE ÖNERİLER

Bu tez çalışmasında, video içerisinde yer alan, sigara nesnelerini tespit edip bu sahnelerdeki sigara nesnesini sansürleyen YOLOv5, YOLOv8 ve YOLOv9 modelleri oluşturulmuş, bu modellerin performansını artırmak için hiper parametre optimizasyonu ve farklı optimize ediciler kullanılmıştır. Video görüntülerdeki sigara ve kül tablası tespit edilip sansürlenmesi sağlanmış, sansürleme başarısı iyileştirme çalışmaları gerçekleştirilmiştir.

Çalışmada birinci aşamada, bilgisayarlı görme alanındaki modern teknolojilerin analizi ve bunların nesne tanıma problemlerine uygulanması gerçekleştirilmiştir. Nesne algılamada popüler ve etkili olan YOLO algoritması kullanımı tercih edilmiştir. YOLOv5, YOLOv8 ve YOLOv9 modellerinin sigara nesnesi tanımadaki başarısı incelenmiştir. Bu modellerin yüksek oranda doğruluk verdiği ve hızlı çözüm sundukları görülmüştür.

Başarı oranlarını iyileştirme çalışmaları yürütülmüş, bu kapsamda YOLOv5, YOLOv8 ve YOLOv9 modellerinin parametreleri, SGD, Adam ve AdamW gibi çeşitli optimize ediciler kullanılarak optimize edilmiştir. Optimizasyonun modellerin başarılarında olumlu etkisi olduğu görülmüştür. SGD optimizasyonunun YOLOv9 modeliyle kullanımında en iyi sonuçlar elde edilmiştir. Ayar dağılım grafiklerinde verildiği üzere, doğruluk ve yakınsama hızı açısından öne çıktığı görülmüş ve uygulamada kullanımı sağlanmıştır.

Çalışmanın en önemli katkısından birisi sigara nesnesi tanımadaki kül tablası tanıma özelliğinin dâhil edilmesidir. Kül tablası, sigara içmenin önemli bir göstergesidir ve tanınması, sigara tamamen görünmese bile, sigara içme olaylarının tespitinin doğruluğunu önemli ölçüde artırmaktadır. Genel sonuçları iyileştiren ve sistemi çok yönlü ve sağlam hale getiren kül tablası tanımlama yaklaşımları modellerimize başarılı bir şekilde uygulanmıştır.

Optimizasyondan sonra, YOLOv5'in performansı %89'da kalırken, YOLOv8 ve YOLOv9'un performansı %93'e çıkmaktadır. YOLOv8 ve YOLOv9, SGD ile optimizasyondan sonra aynı mAP(50) değerine sahip olsa da, YOLOv9 mAP(50-95)

metriğinde YOLOv8'i %2 oranında geride bırakmaktadır. Son olarak, YOLOv9 modelin kullanarak oluşturduğumuz algoritma %96 doğrulukla en iyi sonucu elde ederek etkinliğini ve ilerideki uygulamalara yönelik potansiyelini ortaya koymuştur.

Sigara nesnesi ve sigara içme fiilinin otomatik algılanması, sigara ile ilgili yasalarının uygulanmasına olumlu yönde önemli katkılar sunacaktır. Halka açık kapalı mekanlara entegre edilmesi durumuna sigara içilmesini azaltma yönünde caydırıcı olacak, sigara kullanmayan kişilerin maruz kalacağı pasif sigara içme problemini azaltacaktır. Bu sistemler sigarayla mücadelede görevli devlet memurlarına yardımcı ve kanıt sağlayıcı bir sistem olarak kullanılması mümkündür. Sistem medya içeriklerini analiz etmede ve çocuk, genç başta olmak üzere sigara içmeyi özendiren içeriklerden korunmalarına katkı sağlayacaktır.

Çalışmamızda, modellerin sonuçları iyileştirmek ve en iyi sonucu elde etmek için hiper parametreleri optimize etmenin ve uygun optimize edicileri seçmenin önemi vurgulamaktadır. YOLOv9 modelinin SGD optimize ediciyle kombinasyonu en iyi sonuçları göstermiştir. Bu yöntemlerin nesne algılama görevlerinde kullanılmasının başarılı sonuçlar ortaya koyduğu gösterilmiştir. Bu modeller için hiper parametreleri optimize etmek, etkili izleme ve analiz sistemleri oluşturmada yüksek doğruluk ve tanıma güvenilirliği elde etmeyi sağlamıştır.

Ortaya çıkan sonuçlar YOLO derin öğrenme modellerinin ve bu modellerde hiper parametre optimizasyonunun kullanılmasının, görüntülerde sigara ve kül tablası tespitin verimliliğini ve doğruluğunu önemli ölçüde artırdığını göstermektedir. Doğru model ve optimize edici mimari seçiminin yanı sıra kül tablası tespiti gibi ek yöntemlerin entegrasyonunun, sigara içimini kontrol etmek ve halk sağlığını korumak için başarılı bir tanıma ve sansürleme sistemi oluşturulmasını sağlamıştır. Duman nesnesinin tanınması sonuçların daha iyileştirilmesini sağlayabilir.

KAYNAKLAR

- Arkin, E., Yadikar, N., Xu, X., Aysa, A. ve Ubul, K. (2023). A survey: object detection methods from CNN to transformer. *Multimedia Tools and Applications*.
- Bochkovskiy, A., Wang, C. ve Liao, M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv preprint arXiv: 2004.10934*.
- Chapman, P., Kerber, R., Clinton, J., Khabaza, T., Reinartz, T., Wirth, R. (1999). The CRISP-DM Process Model. *Discussion Paper*.
- Ciaglia, F., Zuppichini, F., Guerrie, P., McQuade, M. ve Solawetz, J. (2022). Roboflow 100: A Rich, Multi-Domain Object Detection Benchmark. *arXiv preprint arXiv: 2211.13523*.
- Davis, J. ve Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. *23rd international conference on Machine learning*.
- Dataset. "Cigarette Computer Vision Project", 2022. Eriřim adresi: <https://universe.roboflow.com/smoke-tamtu/cigarette-6ubdv>.
- Ding, Q., Dong, X., Guo, W., Zheng W. ve Pan, Y. (2023). Smoking Detection Algorithm Based On Improved YOLOv5. *2023 IEEE 7th Information Technology and Mechatronics Engineering Conference (ITOEC)*.
- Everingham, M., Gool, L., Williams, C., Winn. J. ve Zisserman, A. (2009). The PASCAL Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision*.
- Girshick, R. (2015). Fast R-CNN. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 1440-1448.
- Goodfellow, I., Bengio, Y. ve Couville, A. (2016). *Optimization for Training Deep Models*. Deep Learning
- Huang, T., Cheng, M., Yang, Y., Lv, X., Xu, Y. (2022). Tiny Object Detection based on YOLOv5. *5th International Conference on Image and Graphics Processing*.
- Jain, A. K., Mao, J. ve Mohiuddin K.M. (1996). *Artificial Neural Networks: A Tutorial*. IBM Almaden Research Center.
- Jayakumar, D. ve Peddakrishna, S. (2024). Performance Evaluation of YOLOv5-based Custom Object Detection Model for Campus-Specific Scenario. *International Journal of Experimental Research and Review*.
- Jiang, K., Xie, T., Yan, R., Wen, X., Li, D., Jiang, H., Jiang, N., Feng, L., Duan, X. ve Wang, J. (2022). An Attention Mechanism-Improved YOLOv7 Object Detection Algorithm for Hemp Duck Count Estimation. *Internet and Computers for Agriculture*.

- Ju, R. ve Cai, W. (2023). Fracture Detection in Pediatric Wrist Trauma X-ray Images Using YOLOv8 Algorithm. arXiv preprint arXiv:2304.05071v5.
- Khan, A., Khan, S., Hassan, B. ve Zheng, Z. (2022). CNN-Based Smoker Classification and Detection in Smart City Application. Sensors.
- Kingma, P. D ve Lei Ba J. (2015). ADAM: A Method For Stochastic Optimization. arXiv preprint arXiv: 1412.6980.
- Kumari, S., Gautam A., Basak. S. ve Saxena, N. (2023). YOLOv8 Based Deep Learning Method for Potholes Detection. 2023 IEEE International Conference on Computer Vision and Machine Intelligence (CVMI).
- Lecun, Y., Bengio, Y. ve Hinton, G. (2015). Deep learning.
- Li, Z., Liu, Z. ve Wang, X. (2023). On-Board Real-Time Pedestrian Detection for Micro Unmanned Aerial Vehicles Based on YOLO-v8. 2023 2nd International Conference on Machine Learning, Cloud Computing and Intelligent Mining (MLCCIM).
- Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., Ke, K., Li, Q., Cheng, M., Nie, W., Li, Y., Zhang, B., Liang, Y., Zhou, L., Xu, X., Chu, X. ve Wei, X. (2022). YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications. arXiv preprint arXiv: 2209.02976.
- Lin, T., Dollar, P., Girshick, R., He, K., Hariharan, B. ve Belongie, S. (2018). Feature Pyramid Networks for Object Detection. arXiv preprint arXiv:1803.01534.
- Liu, S., Qi, L., Qin, H., Shi, J. ve Jia, J. (2017). Path Aggregation Network for Instance Segmentation. arXiv preprint arXiv:1612.03144.
- Long, J., Shelhamer, E. ve Darrell. T. (2015). Fully Convolutional Networks for Semantic Segmentation. The Computer Vision Foundation.
- Loshchilov, I. ve Hutter, F. Decoupled Weight Decay Regularization. arXiv preprint arXiv: 1711.05101.
- Macalisang, J. R., Merencilla, N. E., Ligayo D. (2020). Eye-Smoker: A Machine Vision-Based Nose Inference System of Cigarette Smoking Detection using Convolutional Neural Network. 7th International Conference on Engineering Technologies and Applied Sciences (ICETAS).
- Matthew, D., Fergus, R. (2013). Visualizing and Understanding Convolutional Networks. arXiv: 1311.2901.
- Mitchell, T. M. (2006). The Discipline of Machine Learning. CMU-ML-06-108.
- Munawar, R., Jocher, G., Chaurasia. A. (2023). Ultralytics YOLO Docs. <https://docs.ultralytics.com/#where-to-start>
- Nguyen, G., Dlugolinsky, S., Bobák, M., Tran, V., López, A. G., Heredia, I., Malík, P. ve Hluchý L. (2019). Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey. Artificial Intelligence Review (pp. 77–124). <https://doi.org/10.1007/s10462-018-09679-z>
- Padilla, R., Netto, S. ve Eduardo, A. (2020). A Survey on Performance Metrics for Object-Detection Algorithms. 2020 International Conference on Systems, Signals and Image Processing (IWSSIP).

- Peng, J., Wang, C., Li, Y. ve Chen, H. (2023). Substation Personnel Smoking Detection Based On GhostNetV2-YOLOv5. 2023 6th International Symposium on Autonomous Systems (ISAS).
- Redmon, J., Divvala, S., Girshick, R. ve Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. arXiv preprint arXiv: 1506.02640.
- Redmon, J. ve Farhadi, A. (2018). YOLOv3: An Incremental Improvement. arXiv preprint arXiv: 1804.02767.
- Santiago, C., Reyes, M. ve Tria, D. (2022). Deep Convolutional Neural Network for Detection of Cigarette Smokers in Public Places: A Low Sample Size Training Data Approach. 2022 International Conference on Decision Aid Sciences and Applications (DASA).
- Seddik, E., Hewieg, M. ve Afify, R. (2024). Artificial Intelligence (AI) Powered Detection and Identification of Pressure Vessel External Damage – YOLO-based Application. Research Square.
- Seyyarer, E., Ayata, A., Uçkan, T. ve Karcı A. (2020). Derin Öğrenmede Kullanılan Optimizasyon Algoritmalarının Uygulanması Ve Kiyaslanması. Anatolian Journal of Computer Sciences.
- Terven, J., Cordova-Esparza, D. ve Romero-Gonzalez J. (2023). A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. Machine Learning and Knowledge Extraction.
- Wahyudi, D., Soesanti, I., ve Nugroho, H. (2023). Optimizing Hyperparameters of YOLO to Improve Performance of Brain Tumor Detection in MRI Images. 2023 6th International Conference on Information and Communications Technology (ICOIACT).
- Wang, C., Bochkovskiy, A. ve Liao, M. (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv preprint arXiv: 2207.02696.
- Wang, C., Liao, M., Wu, Y., Chen, P., Hsieh, J. ve Yeh, I. (2020). CSPNet: A New Backbone that can Enhance Learning Capability of CNN. CVPR, 2020 open access.
- Wang, C., Zheng, T., Sun, F. ve Lia, H. (2023). A Smoking Detection Algorithm Based on Improved YOLOV5. 2023 IEEE 3rd International Conference on Power, Electronics and Computer Applications (ICPECA).
- Wang, C., Yeh, I. ve Liao, M. (2024). YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. arXiv preprint arXiv: 2402.13616.
- World Health Organization, “Tobacco” 2023. Erişim adresi: <https://www.who.int/news-room/fact-sheets/detail/tobacco>.
- Zhang, D., Jiao, C. ve Wang, S. (2018). Smoking Image Detection Based on Convolutional Neural Networks. 2018 IEEE 4th International Conference on Computer and Communications (ICCC).



ÖZGEÇMİŞ

Ad-Soyad Yerniyaz BAKHYTOV

ÖĞRENİM DURUMU:

- Lisans** : 2021, Al-Farabi Kazakh National University, Bilgisayar Mühendisliği.

TEZDEN TÜRETİLEN ESERLER:

- Yerniyaz B. ve Cemil Ö. (2024, June). Cigarette Detection in Images Based on YOLOv8. *Sakarya University Journal of Computer and Information Sciences (SAUCIS)*, Sakarya, Turkey.