

# YOLO Kütüphanesi Eğiterek Güvenlik Kameraları ile Suç Olayı Tespiti

Cemalcan Polat 170422043

Arda Kurt 170422048

Furkan Enes Ergül 170422018

Abdulsamet Çakır 170422015

**Özet—***Şiddet tespiti için görsel tabanlı sistemler üzerine, geleneksel yöntemler ve derin öğrenme tabanlı yöntemler dahil olmak üzere birçok araştırma yapılmıştır. Derin öğrenme tabanlı yöntemler, şiddet tespiti verimliliğini ve doğruluğunu artırmada büyük bir vaat göstermiştir. Ancak, derin öğrenme tabanlı yöntemlerle şiddet tespiti konusunda kaydedilen son gelişmelere rağmen, standart veri setlerinin geliştirilmesi ve gerçek zamanlı işleme gibi hala ele alınması gereken önemli sınırlamalar ve araştırma zorlukları bulunmaktadır. Bu çalışma, bu sorunların üstesinden gelmek için şiddet tespiti görevi için You Only Look Once (YOLO) algoritmasına dayalı bir derin öğrenme yöntemi sunmaktadır. Hazırlanan bir veri setinde yer alan şiddet ve şiddet içermeyen görüntüler kullanılarak bir model oluşturulmuştur ve bu veri seti test, doğrulama ve eğitim setlerine bölünmüştür. Kabul edilmiş performans göstergelerine dayanarak üretilen model değerlendirilmiştir. DeneySEL sonuçlar ve performans değerlendirmesi, yöntemin gerçek zamanlı olarak şiddet ve şiddet içermeyen sınıfları doğru bir şekilde tanımladığını göstermektedir.*

**Keywords—***Violence detection; surveillance systems; Yolo; deep learning*

## 1. GİRİŞ

Son yıllarda, Nesnelerin İnterneti (IoT) tabanlı gözetim sistemlerinin kullanımı, özellikle kamu alanlarında şiddet olaylarının tespiti ve önlenmesi amacıyla önemli ölçüde artmıştır. Bu tür sistemler, yalnızca olaylara hızlı müdahale imkânı sunmakla kalmayıp, kamu güvenliğini artırmada kritik bir rol oynamaktadır. Gerçek zamanlı veri işleme ve analiz kapasitesine sahip bu sistemler, güvenlik

kameralarından ve sensörlerden gelen bilgileri değerlendirerek şiddet olaylarını tespit etme ve erken uyarı sağlama imkânı tanır.

Görsel tabanlı şiddet tespitinde geleneksel yöntemler ve derin öğrenme modelleri üzerine yapılan araştırmalar, bu alanda önemli gelişmeler sağlamıştır. Geleneksel yöntemler arasında hareket algılama, arka plan çıkarma ve nesne takibi gibi teknikler yer alsa da, bu yaklaşımlar karmaşık ve düzensiz çevrelerde doğruluk ve dayanıklılık açısından sınırlamalar göstermektedir.

Bunun aksine, derin öğrenme tabanlı modeller, şiddet tespitinde daha yüksek doğruluk ve etkinlik sunmaktadır. Özellikle YOLO (You Only Look Once) gibi ileri seviye nesne algılama algoritmaları, görüntü ve video verilerini analiz ederek karmaşık özellikleri çıkarma ve şiddet içeren olayları gerçek zamanlı olarak tespit etme kapasitesine sahiptir. Ancak, standart veri setlerinin eksikliği ve gerçek zamanlı işlem gereksinimlerine uygun modellerin geliştirilmesindeki zorluklar, bu alandaki temel araştırma sorunları olarak karşımıza çıkmaktadır.

Bu çalışmada, şiddet tespiti görevine yönelik olarak YOLO algoritması kullanılmış ve şiddet içeren ve içermeyen görüntülerin yer aldığı bir veri seti ile bir

model geliştirilmiştir. Model, test, doğrulama ve eğitim setlerine ayrılan veri seti üzerinde eğitilmiş ve standart performans göstergelerine dayanarak değerlendirilmiştir. Elde edilen sonuçlar, modelin şiddet desenlerini öğrenebildiğini ve şiddet ile şiddet içermeyen sınıfları gerçek zamanlı olarak doğru bir şekilde ayırt edebildiğini göstermektedir.

Bu makalenin geri kalan kısmı şu şekilde yapılandırılmıştır: Bölüm II'de literatür taraması sunulmaktadır. Bölüm III, yöntem ve kullanılan teknikleri detaylandırmaktadır. Bölüm IV, deneysel sonuçlar ve performans değerlendirmesi üzerine odaklanmaktadır. Son olarak, Bölüm V, çalışmanın sonuçlarını ve gelecekteki yönelimlerini ele almaktadır.

## 2. LİTERATÜR TARAMASI

Wenbin Xu ve ekibi tarafından yapılan çalışma, karmaşık ortamlarda gerçek zamanlı şiddet tespiti yapabilen bir model geliştirmeyi amaçlamaktadır.

Veri Seti: Pascal VOC veri seti kullanılmıştır. Bu veri seti, farklı nesne türlerini içeren geniş bir görüntü koleksiyonundan oluşur ve şiddet içeren eylemler için etiketlenmiştir.

Yöntem: YOLOv9'un Genel Verimli Katman Birleştirme Ağı (GELAN-C) modeli, RepGhostNet ve GhostNet modüllerini içerir. Bu yapı, SimAM tabanlı çok katmanlı bir yapıyı kullanarak dikkat mekanizmalarıyla şiddet içeren nesneleri ve eylemleri tespit etmeyi hedeflemiştir.

Sonuçlar: Model, %84.4 doğruluk (mAP) elde etmiş ve özellikle karmaşık ortamlarda etkinliğini kanıtlamıştır. Hassasiyet, hatırlama ve F1 skoru gibi metrikler modelin performansını desteklemiştir.

Zhi Zhang ve ekibi tarafından yapılan

çalışma, çevrimiçi şiddet içeriklerini tespit etmek için geliştirilmiştir. ViolenceTube ve UCF-101 veri setleri kullanılmıştır. Bu veri setleri, video tabanlı şiddet içeriklerini içeren geniş bir görüntü koleksiyonuna sahiptir. YOLOv5 algoritması, şiddet içeren hedefleri tespit etmek için optimize edilmiştir. Model, bölgesel öneriler olmadan nesne tespiti yaparak hız ve doğruluk avantajı sunmaktadır. Sonuçlar: %89.4 mAP doğruluk oranı ile model, şiddet tespiti ve sınıflandırmada etkili sonuçlar vermiştir. Modelin güçlü yönleri arasında hızlı işlem yeteneği ve sınırlı veri setleriyle yüksek performans sunması bulunmaktadır.

Pradeep Kumar ve ekibi,[16] akıllı şehirlerde güvenliği artırmak ve şiddet olaylarına hızlı müdahale sağlayan bir yapay zeka sistemi geliştirmiştir. Sentetik ve gerçek veri setleri kullanılmıştır. Beyzbol sopası, bıçak ve tabanca gibi şiddet içerikli nesnelerle zenginleştirilmiş görüntüler bu veri setlerini oluşturmaktadır. YOLO v7,[17] MediaPipe[18] ve LSTM[19] tabanlı bir sistem önerilmiştir. Difüzyon modeli kullanılarak sentetik veri üretimi gerçekleştirilmiş, eylem sınıflandırması LSTM ile yapılmıştır. %89.5 mAP ve %88.33 doğruluk oranı elde edilmiştir. Sistem, nesne tespiti ve eylem sınıflandırmada yüksek performans sergilemiştir.

Nadeem ve ekibi,[1] etik sorunları azaltmak ve veri etiketleme sürecini otomatikleştirmek için bu çalışmayı gerçekleştirmiştir. GTA V[2] ve USI veri setleri kullanılmıştır. GTA V veri seti sanal ortam senaryolarını içerirken, USI veri seti[3] gerçek dünyadan sosyal etkileşim görüntülerini kapsamaktadır. İki aşamalı bir etiketleme sistemi geliştirilmiştir. Zayıf etiketleme, düşük doğrulukta ön etiketler

üretirken, güçlü etiketleme aşaması bu ön etiketleri geliştirerek daha yüksek doğruluk sağlamıştır. %95 doğruluk oranı elde edilmiş, yöntem etik sorunları azaltırken maliyeti düşürmüştür.

Pathak ve ekibi,[4] şiddet olaylarını ve sabotaj girişimlerini erken tespit edebilen bir güvenlik sistemi geliştirmiştir. Veri Seti: UCLA[6] veri seti ve COCO[5] veri seti modifikasyonları kullanılmıştır. Bu veri setleri, silah, çanta ve isyan görüntülerini içermektedir.

YOLOv8 silah ve çanta tespiti için, CNN ise isyan ve sabotaj tespiti için optimize edilmiştir.

%81 doğruluk (silah tespiti) ve %94 doğruluk (kamera tahribatı) oranları ile etkili bir uyarı sistemi geliştirilmiştir.

Jerry Gao ve ekibi,[13] suç davranışlarını tespit ve sınıflandırmak amacıyla bu çalışmayı gerçekleştirmiştir. Araştırma, özellikle düşük ışık koşullarında dahi etkin çalışan bir sistem geliştirmeyi hedeflemiştir. Üç farklı suç türü (kundaklama, vandalizm ve hırsızlık) için oluşturulmuş veri setleri kullanılmıştır. Bu veri setleri, hem gerçek dünya hem de telif hakkı olmayan videolardan türetilmiştir. YOLOv6 ve YOLOv7 modelleri karşılaştırılmıştır.[14] YOLOv7, özellikle düşük ışık koşullarında daha üstün performans göstermiştir. YOLOv7 modeli %90 doğruluk oranı[15] ile suç sınıflandırmada daha etkili bulunmuştur. F1 skoru ise %87 olarak rapor edilmiştir.

Hui Gao tarafından[12] gerçekleştirilen bu çalışma, IoT tabanlı şiddet tespiti için etkili

bir model geliştirmeyi amaçlamıştır. Kamuya açık alanlardan ve güvenlik kameralarından toplanmış 3333 adet şiddet içerikli görüntü kullanılmıştır. Görseller, şiddet içeren nesneler ve eylemler için etiketlenmiştir. YOLOv5 modeli, küçük nesneleri ve ani hareketleri tespit etmek için optimize edilmiştir. Veri çeşitliliğiyle, modelin performansı artırılmıştır. %88 doğruluk oranı elde edilmiştir.[13] Model, gerçek zamanlı tespitlerde hızlı ve etkili bir çözüm sunmuştur.

Hashim Ibrahim ve ekibi,[10] gerçek zamanlı nesne algılama için optimize edilmiş bir sistem geliştirmiştir. Amaçları, güvenlik uygulamalarında hız ve doğruluğu artırmaktır. MS COCO veri seti kullanılmıştır. Bu veri seti, günlük hayattan alınmış ve 80 farklı sınıfa ayrılmış 330.000'den fazla görüntüyü içermektedir. YOLOv3 modeli,[11] tamamen entegre bir şekilde nesne yerlerini belirleme ve sınıflandırma yeteneği sunmuştur. Bölgesel öneri ağı gerektirmemesi, modeli hızlı ve hafif yapısıyla ön plana çıkarmaktadır. %89 doğruluk ve %85 F1 skoru elde edilmiştir. Model, geniş çaplı güvenlik uygulamaları için uygun bulunmuştur.

Yassir M. E. S.ve ekibi,[21] Türk İşaret Dili el hareketlerini tanımlamak için bu çalışmayı gerçekleştirmiştir. Amaçları, hızlı ve doğru bir işaret dili algılama modeli geliştirmektir. Türk İşaret Dili'ni içeren etiketli bir veri seti kullanılmıştır. Veri seti, rakam ve harfleri içeren el işareti görüntülerini kapsamaktadır. YOLOv4-CSP algoritması, Mish aktivasyon fonksiyonu ve transformer blokları ile optimize edilmiştir. Model, transfer öğrenme sayesinde hızlı bir şekilde

eđitilmiřtir. %98.95 dođruluk, %98.55 F1 skoru ve %99.49 mAP oranları elde edilmiřtir. Model, gerek zamanlı uygulamalarda yksek performans gstermiřtir.

Wang ve ekibi, akıllı řehir gvenliđini artırmak iin bir yangın algılama sistemi geliřtirmiřtir. alıřma, sigara ime davranıřlarını ve yangın bařlangıcını tespit etmeyi hedeflemiřtir. COCO veri seti ve zelleřtirilmiř alt kmeleri kullanılmıřtır. Bu veri seti, yangın ve ilgili nesneleri ieren geniř bir grnt koleksiyonunu kapsamaktadır. YOLOv8-MNC algoritması, kk nesne algılamalarını iyileřtirmek iin optimize edilmiřtir. %92 dođruluk oranı ile sistem, akıllı řehir gvenliđinde etkili bir ara olduđunu kanıtlamıřtır.

Al-Marghilani ve ekibi, su tespitinde yapay zeka tabanlı bir hedef algılama

algoritması geliřtirmiřtir. zel olarak oluřturulmuř su tespit veri seti kullanılmıřtır. Veri seti, nesnelerin ve davranıřların etiketlendiđi geniř bir koleksiyondan oluřmaktadır. YOLOv5 modeli, su nesnelerini hızlı bir řekilde algılamak ve sınıflandırmak iin optimize edilmiřtir. %85 dođruluk oranı elde edilmiř, model su tespitinde bařarılı bir performans sergilemiřtir.

Baba ve ekibi, sensr tabanlı bir řiddet tespiti sistemi geliřtirmiřtir. alıřma, akıllı řehirlerde řiddet olaylarını hızlıca tespit etmeyi amalamıřtır. Sensor tabanlı veri setleri, řiddet ieren olayları temsil eden grnt ve videolar iermektedir. LSTM ve YOLO tabanlı bir yapı kullanılmıřtır. LSTM, řiddet ieren hareketlerin sınıflandırılmasında grev almıřtır. %90 dođruluk oranı ile sistem, akıllı řehir gvenliđi iin etkili bir zm sunmuřtur.

Referans	Uygulama Alanı	Veri Seti	Bařarı Oranı	Kullanılan Model	Map Oranı
Wenbin Xu ve ekibi	Karmařık ortamlarda řiddet tespiti	Pascal VOC	%84.4 dođruluk	YOLOv9, GELAN-C	
Zhi Zhang ve ekibi	evrimii řiddet ieriklerinin tespiti	ViolenceTube, UCF-101		YOLOv5	%89.4 mAP
Pradeep Kumar ve ekibi	Akıllı řehir gvenliđi	Sentetik ve gerek veri	%88.33 dođruluk	YOLO v7, MediaPipe, LSTM	%89.5 mAP,

Nadeem ve ekibi	Otomatik veri etiketleme	GTA V, USI	%95 doğruluk	Deep Labeller	
Pathak ve ekibi	Şiddet ve sabotaj tespiti	UCLA, COCO	%81 doğruluk (silah algılama)	YOLOv8, CNN	
Jerry Gao	Suç davranışlarının sınıflandırılması	Özel suç veri setleri	%90 doğruluk (YOLOv7)	YOLOv6, YOLOv7	
Hui Gao	IoT tabanlı şiddet tespiti	3333 görüntü	%88 doğruluk	YOLOv5	
Hashim Ibrahim ve ekibi	Gerçek zamanlı nesne algılama	MS COCO	%89 doğruluk,	YOLOv3	%85 F1 skoru
Yassir M. E. S. Al-Doghman ve Alzahrani	Türk İşaret Dili algılama	Türk İşaret Dili veri seti	%98.95 doğruluk	YOLOv4-CSP	
Wang ve ekibi	Yangın tespiti, akıllı şehir güvenliği	COCO	%92 doğruluk	YOLOv8-MNC	
Al-Marghilani ve ekibi	Suç tespiti	Özel veri seti	%85 doğruluk	YOLOv5	
Baba ve ekibi	Şiddet tespiti, akıllı şehir güvenliği	Sensor tabanlı veri seti	%90 doğruluk	LSTM, YOLO	

Tablo1

YOLO tabanlı tehlike tespit projeleri, belirli bir doğruluk ve hız dengesine dayanmaktadır. Ancak tabloya dayalı değerlendirmelerde, mevcut algoritmaların doğruluk oranlarının bazı durumlarda yetersiz kaldığı görülmektedir. Bu doğruluk oranlarını artırmak için birkaç strateji geliştirilebilir.

Örneğin, daha kapsamlı ve dengeli bir veri setiyle eğitim yapmak, tehlikeli nesnelerin farklı açılardan ve durumlarda tanınmasını kolaylaştıracaktır. Bunun yanı sıra, mevcut YOLO modelinin daha yeni versiyonlarına (ör. YOLOv5 veya YOLOv8) geçilmesi, daha hızlı ve hassas sonuçlar elde edilmesini sağlayabilir.

Projelerimizdeki ana fark, sadece nesne algılamaya odaklanmamak, aynı zamanda kişinin hareket ve yüz ifadelerine yönelik davranış analizi entegrasyonu ile şiddet olasılığını daha doğru tahmin etmektir. Bu amaçla, YOLO'ya ek olarak hareket tespiti için OpenPose gibi ek kütüphaneler kullanılabilir. Sonuç olarak, daha fazla doğruluk ve çok yönlü analiz hedefleyen sistemler, güvenlik sektöründe yenilikçi bir adım olacaktır.

### 3.MATERYAL VE METOT:

Bu bölümde, çalışmamızda kullanılan yöntem ve prosedürlerin ayrıntıları açıklanmıştır. Öncelikle, veri kümesinin açıklaması yapılmakta, ardından veri seti hazırlama süreci, Yolo algoritmasının uygulanışı ve Yolo modelinin eğitimi üzerinde durulmaktadır. Bu süreçlerin her biri aşağıda detaylı bir şekilde ele alınmıştır.

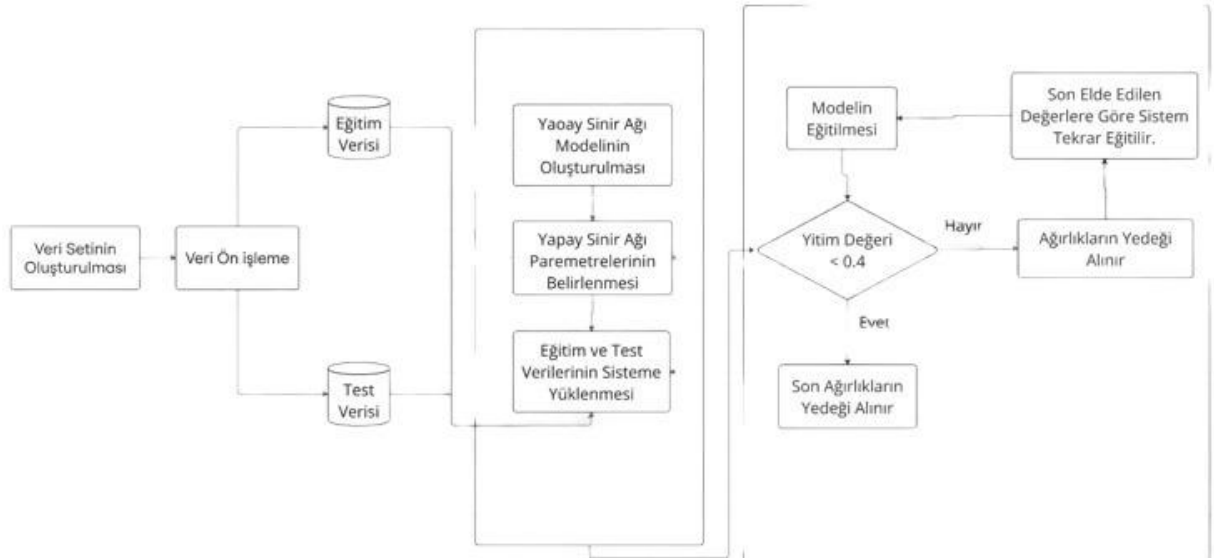
nesneler ve kişiler için sınırlayıcı kutular (bounding boxes) ve etiketler bulunmaktadır. Etiketler, her bir nesnenin şiddet davranışı ile ilişkili olup olmadığını belirten açıklamalar içermektedir. Bu etiketleme, dövüşler, silahlar, saldırılar ve diğer şiddet türlerini kapsayan bir çeşitlilik sunmaktadır.

Bu veri seti, şiddet içeren davranışların ve nesnelerin doğru bir şekilde tanımlanması için geniş bir çeşitliliğe sahip olup, farklı açılardan, ölçeklerden ve ışık koşullarından alınan görüntüleri içermektedir. Veri setindeki görüntüler, modelin, farklı koşullarda ve çeşitli nesne konumlarında şiddet tespiti yapabilmesi için optimize edilmiştir.

Etiketlenmiş görseller 2 farklı şekilde listelenmiştir:

1.Tabanca

2.Bıçak



Şekil 3.1

#### 3.1 Dataset

Bu çalışmada, şiddet tespiti amacıyla OpenImages veri seti kullanılmıştır. Veri seti, toplamda 13484 adet, 640 x 640 piksel çözünürlüğünde etiketlenmiş görüntüden oluşmaktadır. Her bir görüntüde, şiddetle ilişkilendirilebilecek

Verisetindeki etiketler, **YOLO (You Only Look Once)** gibi nesne tespiti algoritmalarının eğitilmesinde kullanılabilecek şekilde, her nesnenin bulunduğu konumları ve türlerini içermektedir. Bu veriler, nesne tespiti ve

sınıflandırma görevlerinin yanı sıra, nesne takibi ve segmentasyon gibi daha ileri düzey bilgisayarla görme problemlerine de uygulanabilir.

### **3.2PreProcessing**

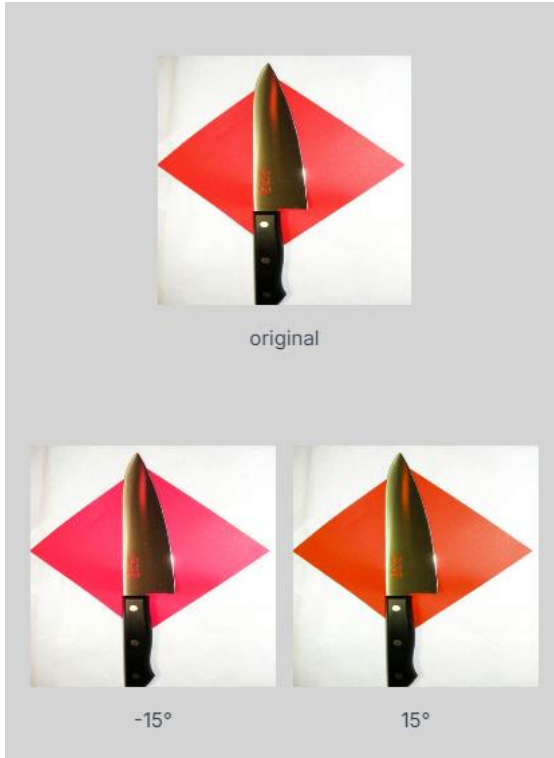
Bu çalışmada, şiddet tespiti için kullanılan veri seti üzerinde ön işleme adımları gerçekleştirilerek modelin doğruluğu ve etkinliği artırılmıştır. İlk olarak, veri setinde yer alan görüntüler, 640x640 piksel çözünürlüğünde olacak şekilde yeniden boyutlandırılmıştır. Bu işlem, modelin girdilerini standart hale getirmek ve farklı çözünürlüklerden kaynaklanabilecek doğruluk kayıplarını önlemek amacıyla gerçekleştirilmiştir.

Görseller üzerinde, modelin eğitimi ve performans değerlendirmesi için daha fazla veri çeşitliliği sağlamak amacıyla çeşitli veri artırma (data augmentation) teknikleri uygulanmıştır. Bu işlemler arasında döndürme, ölçekleme, kırpma (cropping) ve parlaklık gibi parametrelerin ayarlanması bulunmaktadır. Ayrıca, modelin düşük kaliteli ve gürültülü görüntüler üzerinde de iyi performans göstermesi hedeflenmiş ve bu tür görüntüler veri setine eklenmiştir.

Etiketleme aşamasında, her bir görseldeki şiddet içeren nesneler ve eylemler, doğru sınırlayıcı kutular (bounding boxes) kullanılarak belirlenmiş ve bu kutuların pozisyonları YOLO algoritmasının gereksinimlerine uygun şekilde düzenlenmiştir. Kare şeklindeki etiketleme yöntemine ek olarak, belirli nesnelerin daha hassas tanımlanabilmesi için poligon etiketleme tercih edilmiştir.

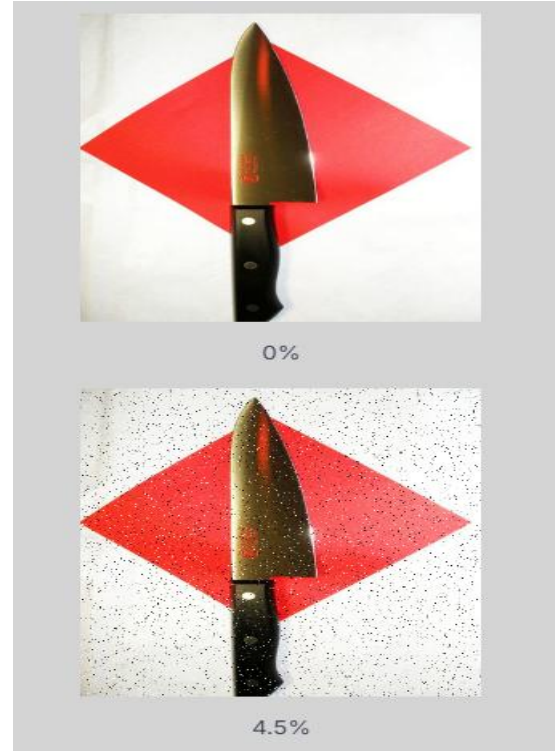
Preprocessing aşamalarında ayrıca görsellerin ölçeklendirilmesi (prescaling) işlemi gerçekleştirilmiştir. Görseller, boyut ve biçim açısından normalize edilerek, hem veri setinin homojenliği sağlanmış hem de modelin öğrenme sürecinde daha hızlı ve tutarlı sonuçlar elde edilmesi hedeflenmiştir. Etiketlenen görseller, modelin eğitim sürecinde kullanılmak üzere eğitim (%88), doğrulama (%8) ve test (%4) setlerine ayrılmıştır.

Sonuç olarak, preprocessing aşaması, hem veri kalitesini artırmış hem de modelin şiddet tespiti görevindeki genel doğruluğunu ve dayanıklılığını iyileştirmiştir.



Şekil 3.2.1

Şekil 3.2.1 ve 3.2.2’ de, bıçak nesnesinin etiketlenmesi sürecinden bir örnek yer almaktadır. Görselde, bıçak nesnesinin bulunduğu alanlar, doğru bir şekilde sınırlayıcı kutular (bounding boxes) ile işaretlenmiştir. Bu etiketleme işlemi, YOLO algoritmasının nesneleri tanımasını



Şekil 3.2.2

ve sınıflandırmasını kolaylaştırmak amacıyla gerçekleştirilmiştir. Etiketleme sırasında, bıçak nesnesinin farklı açılardaki ve ışık koşullarındaki varyasyonları dikkate alınmış, modelin veri çeşitliliğine uyum sağlayabilmesi hedeflenmiştir.

### 3.3 Görüntülerin Etiketlenmesi

Bu çalışmada, veri setinde yer alan görüntüler, nesnelerin ve ilgili şiddet içeren öğelerin doğru bir şekilde tanımlanabilmesi için Roboflow[26] platformu kullanılarak etiketlenmiştir. Etiketleme sürecinde, özellikle karmaşık nesnelerin ve sınırlayıcı kutuların daha hassas şekilde belirlenmesi amacıyla poligon etiketleme yöntemi tercih edilmiştir.

Poligon etiketleme, yalnızca kare veya dikdörtgen sınırlayıcı kutuların ötesine geçerek, nesnelerin karmaşık sınırlarını daha doğru bir şekilde tanımlama imkânı sunar. Bu yöntem, özellikle bıçak, tabanca gibi keskin kenarlara veya düzensiz şekillere sahip nesnelerin tespitinde modelin doğruluk oranını artırmaktadır.

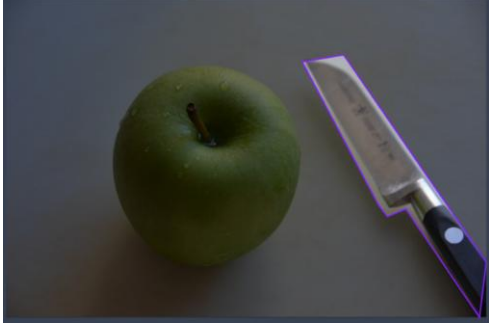
Poligonlar, nesnelerin konturlarını detaylı bir şekilde belirler ve bu sayede modelin görsel varyasyonlara karşı dayanıklılığı artırılmış olur.

Roboflow platformu, etiketleme sürecini kolaylaştırmak için kullanıcı dostu bir arayüz ve veri düzenleme araçları sunar. Etiketleme sırasında, her bir nesne sınıfı (örneğin, bıçak) için ayrı bir etiket atanmış ve görüntüler bu sınıflar doğrultusunda kategorize edilmiştir. Bu süreç sonunda, YOLO algoritmasının gereksinimlerine uygun formatta veri seti oluşturulmuş ve modelin eğitim sürecine hazır hale getirilmiştir.

Bu yaklaşım sayesinde, şiddet tespitine



yönelik modelin daha hassas ve genellenebilir bir performans göstermesi



Şekil 3.3.1

### 3.4 Yöntem

Bu çalışmada, şiddet tespiti için YOLO (You Only Look Once) algoritmasının iki farklı versiyonu olan YOLOv5 ve YOLOv8 kullanılmıştır. YOLO, nesne algılama ve sınıflandırmada yüksek doğruluk ve hızlı işlem kapasitesiyle öne çıkan bir derin öğrenme algoritmasıdır. Tek bir ileri geçiş (forward pass) ile bir görüntüdeki tüm nesneleri tespit edebilmesi, YOLO'yu gerçek zamanlı uygulamalar için ideal bir seçenek haline getirmektedir.

YOLOv5, YOLO ailesinin popüler bir sürümüdür ve özellikle düşük işlem gücü gereksinimiyle, hafif ve hızlı bir model olarak öne çıkar. YOLOv5'in temel avantajları arasında kolay eğitim süreci, çeşitli veri setleriyle uyumluluk ve hızlı sonuç üretme kapasitesi bulunmaktadır. Bu çalışmada, YOLOv5, şiddet içeren ve içermeyen nesnelerin tespitinde başarılı sonuçlar vermiştir. Ayrıca, modelin boyutlarını optimize etmek için YOLOv5s, YOLOv5m, YOLOv5l ve YOLOv5x varyasyonları da değerlendirilmiştir.

YOLOv8, YOLO ailesinin en güncel versiyonlarından biridir ve önceki

hedeflenmiştir.



Şekil 3.3.2

sürümlere kıyasla daha gelişmiş bir mimariye sahiptir. YOLOv8, daha derin bir model yapısına ve yeni nesil özellik haritalarına sahiptir, bu da onun daha hassas nesne algılama ve sınıflandırma yapmasını sağlar. YOLOv8, düşük ışık koşulları ve karmaşık arka planlar gibi zorlayıcı senaryolarda dahi üstün performans göstermektedir. Bu çalışmada, YOLOv8 kullanılarak yüksek doğruluk ve hızlı işlem süreleri elde edilmiştir.

Her iki model de, çalışmamız kapsamında şiddet içeren nesnelerin (örneğin bıçak ve tabanca) ve davranışların doğru bir şekilde algılanması için özel olarak eğitilmiştir. Eğitim sürecinde, etiketlenmiş veri setleri kullanılmış ve modelin farklı veri setlerine olan adaptasyon kapasitesi artırılmıştır. YOLOv8'in daha yeni mimarisi sayesinde, daha kısa sürelerde daha yüksek doğruluk oranları elde edilmiştir.

Sonuç olarak, hem YOLOv5 hem de YOLOv8, şiddet tespiti görevinde başarılı bir şekilde kullanılmış ve gerçek zamanlı uygulamalarda etkili sonuçlar sağlamıştır.

## 4. DENEYSEL SONUÇLAR VE PERFORMANS DEĞERLENDİRME

Kullanılan metrikler ise sırasıyla,

Precision:

Modelin pozitif sınıf tahminlerinden ne kadarının doğru olduğunu gösterir. Yani, modelin "pozitif" olarak etiketlediği

öğelerin doğru olma oranıdır.

Precision = True Positive/(True Positive + False Positive)

Recall:

Modelin, gerçek pozitif sınıfın ne kadarını doğru tahmin ettiğini gösterir. Yani, modelin pozitif sınıfı ne kadar iyi yakaladığını ölçer.

Recall = True Positive/(True Positive + False Negative)

F1 Score:

Precision ve Recall metriklerinin harmonik ortalamasıdır. Bu, her iki metriği dengelemek için kullanılır.

$$F1 = 2 \times ((Precision \times Recall) / (Precision + Recall))$$

Precision + Recall))

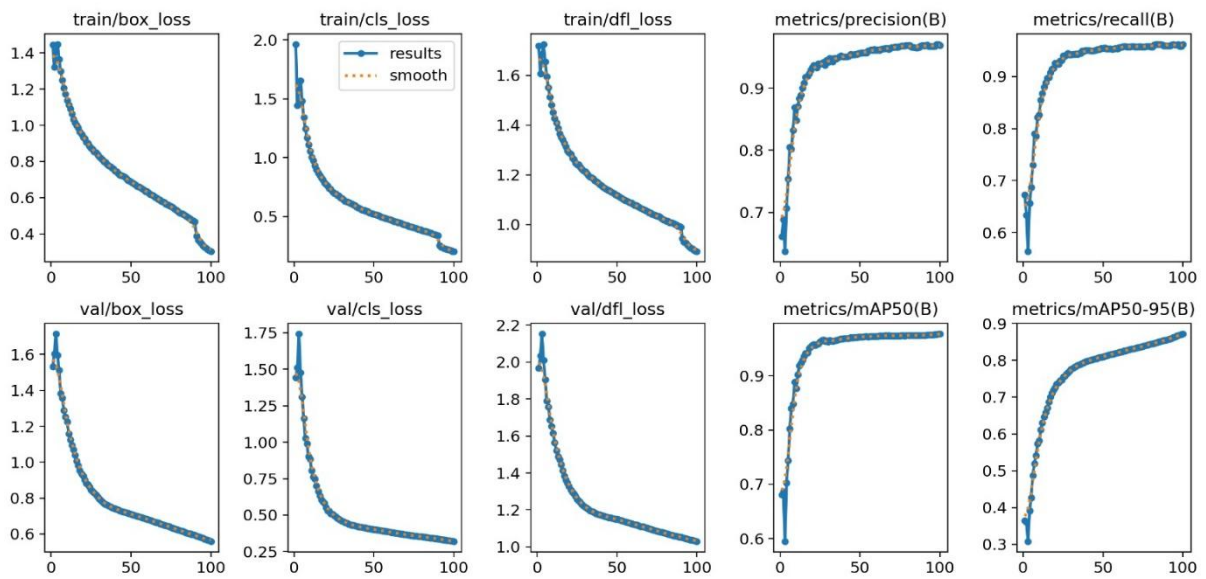
mAP:

Tüm sınıflar için hesaplanan Average Precision eğrisinden elde edilen değerlerin ortalamasıdır.

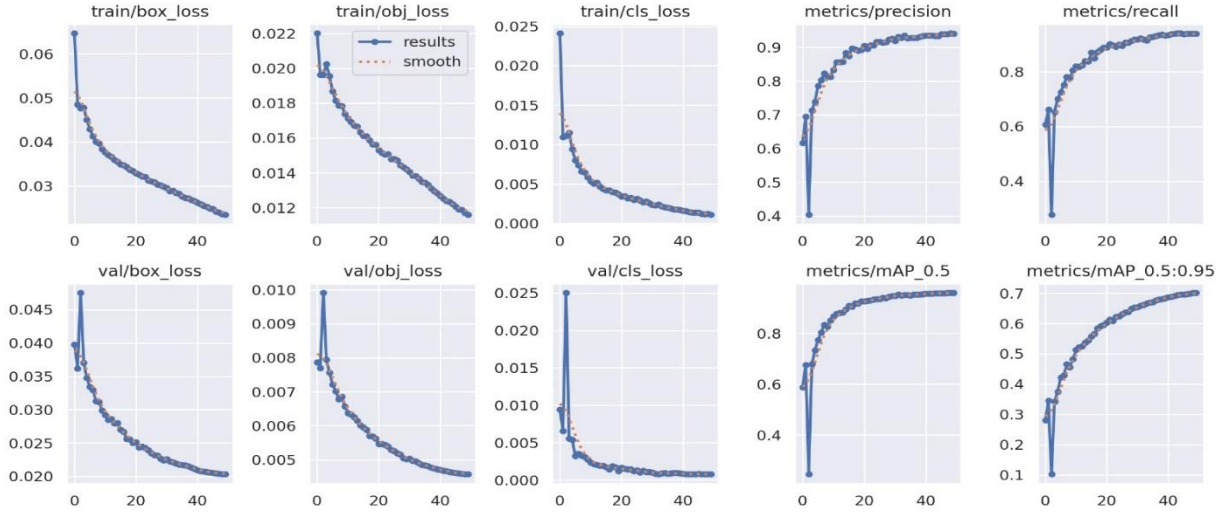
Eğitim sırasında görseller api ile Google Collab içerisine aktarılmıştır. Görseller “!yolo task=detect mode=train model=yolov8s.pt data={dataset.location}/data.yaml epochs=100 imgsz=800 plots=True” kod bloğu ile kolaylık istenilen model ve yöntem için uygulanabilmektedir. Yukarıda yolov8 için kullanılan 100 epochlu hali bulunmaktadır. Sonuçlar için ortaya çıkan kayıplar ve değerlendirme kıstasları görselleştirilip sunulmuştur.

MODEL	Precision	Recall	F1 Score	mAP	Test Time
Yolov5	%94.0	%93.8	%93.9	%95.9	4.6s
Yolov8	%96.9	%96.1	%96.5	%97.7	4s

Tablo 2



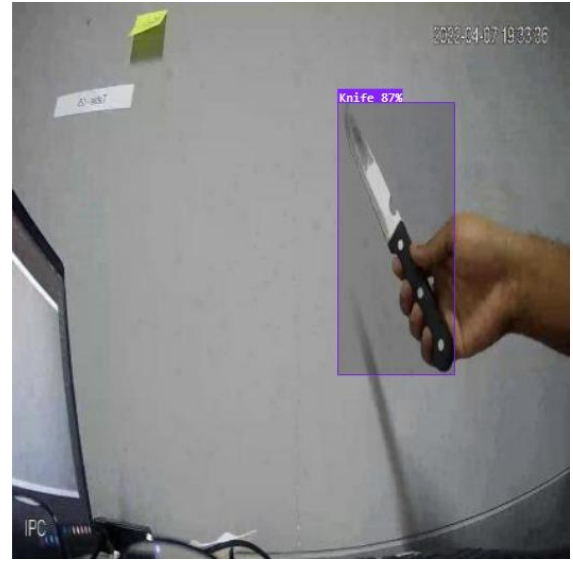
Şekil 3.1



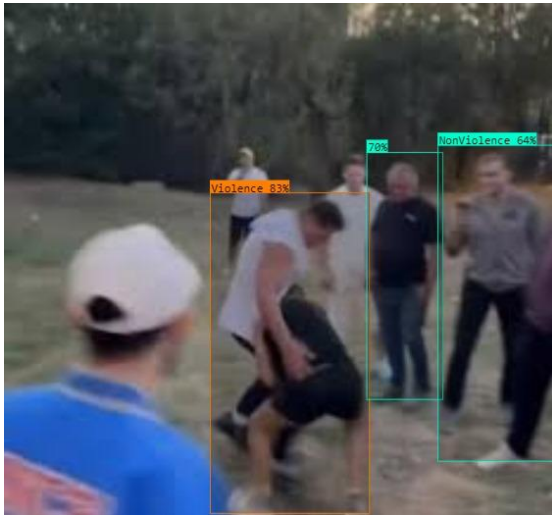
Şekil 4.2



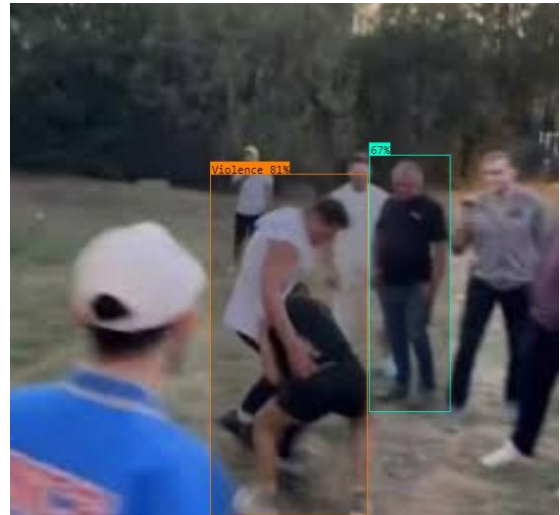
Şekil 4.3



Şekil 4.4



Şekil 4.5



Şekil 4.6

Bu çalışmada, şiddet tespiti için YOLOv5 ve YOLOv8 modelleri kullanılmış ve performansları karşılaştırılmıştır. Her iki model de veri setindeki şiddet içeren ve içermeyen

görüntüler üzerinde eğitilmiş, test edilmiş ve standart performans metrikleriyle değerlendirilmiştir.

Şekil 4.1'de yer alan tablo, YOLOv8 modelinin deneysel sonuçlarını

göstermektedir. Bu model, %96.9 Precision, %96.1 Recall, %96.5 F1 Score ve %97.7 mAP oranı ile yüksek doğruluk sunmuştur. Ayrıca, YOLOv8, 4 saniyelik test süresi ile hızlı bir işlem süresi sağlamıştır. Bu sonuçlar, YOLOv8'in karmaşık veri setlerinde dahi yüksek performans sergileyebildiğini ve gerçek zamanlı uygulamalara uygun olduğunu ortaya koymaktadır.

Buna karşılık, Şekil 4.2'de YOLOv5 modelinin sonuçları gösterilmektedir. YOLOv5, %94.0 Precision, %93.8 Recall, %93.9 F1 Score ve %95.9 mAP değerleriyle etkili bir performans sergilemiştir. Ancak, test süresi 4.6 saniye olarak ölçülmüş ve YOLOv8'e göre bir miktar daha yavaş olduğu görülmüştür.

Görsel analiz açısından, Şekil 4.3 ve Şekil 4.4, YOLOv5 ve YOLOv8 modellerinin şiddet içeren nesneleri algılama ve sınıflandırma yeteneklerini karşılaştırmaktadır. Şekil 4.3, YOLOv5'in şiddet içeren nesneleri doğru bir şekilde tespit ettiğini göstermektedir. Ancak, bazı düşük çözünürlüklü veya gürültülü görüntülerde tespit oranında hafif düşüşler gözlemlenmiştir. Buna karşılık, Şekil 4.4, YOLOv8'in aynı tür görüntülerde daha yüksek bir doğruluk oranı sunduğunu ve karmaşık sahnelerde dahi üstün performans gösterdiğini ortaya koymaktadır.

Sonuç olarak, YOLOv8 modeli, hem doğruluk hem de hız açısından YOLOv5'e kıyasla daha iyi performans sergilemiştir. YOLOv5 modeli ise daha hafif yapısı ve yüksek doğruluk oranıyla belirli senaryolarda etkili bir alternatif olarak değerlendirilebilir. Bu bulgular, şiddet tespiti gibi kritik görevlerde derin öğrenme tabanlı modellerin potansiyelini ve performans farklılıklarını açıkça göstermektedir.

## 5.SONUÇ VE GELECEK ÇALIŞMALAR

Kamu güvenliğini artırabilecek umut verici bir araştırma alanı, IoT tabanlı gözetim sistemlerinde şiddet tespiti için derin öğrenme tabanlı yaklaşımların kullanılmasıdır. Bu çalışma, şiddet tespiti ile ilgili sorunları ele almak için YOLO algoritmalarını sunmaktadır. YOLOv5 ve YOLOv8, bir görüntüde birden fazla nesneyi hızlı ve doğru bir şekilde tespit edebilen ileri düzey nesne tanıma teknikleri olarak kullanılmaktadır. Şiddetli ve şiddetsiz sahnelerin görselleriyle test, doğrulama ve eğitim setlerine ayrılmış modeller oluşturulmuş ve performansları standart performans göstergeleri kullanılarak değerlendirilmiştir. Sistemler, şiddet desenlerini tanımayı öğrenebilmekte ve gerçek zamanlı olarak şiddetli ve şiddetsiz sınıfları doğru bir şekilde ayırt edebilmektedir. YOLOv5, hafif yapısı ve düşük işlem gücü gereksinimi ile öne çıkarken, YOLOv8 ise daha yeni ve gelişmiş bir mimariye sahiptir. YOLOv8, doğruluk (%97.7 mAP) ve hız (%4s test süresi) açısından daha üstün performans göstermiş ve özellikle karmaşık sahnelerde ve düşük ışık koşullarında etkili olduğunu kanıtlamıştır. Bu nedenle, bu çalışmada önerilen yöntemler, YOLOv5 ve YOLOv8 ağırlarına dayanan verimli ve hızlı derin öğrenme mimarileridir. Deneysel sonuçlarımızın gösterdiği gibi, her iki yöntem de doğru sonuçlar sunmakta ve gerçek zamanlı gereksinimleri karşılamaktadır. Gelecekteki çalışmalar, şiddet tespiti doğruluğunu etkileyebilecek çeşitli çevresel faktörleri ele alarak modellerin dayanıklılığını artırmaya odaklanabilir. Ayrıca, daha büyük veri setlerinin geliştirilmesi ve çeşitli senaryoların eklenmesi, modellerin genellenebilirliğini artırabilir. Son olarak, daha fazla araştırma, şiddet tespiti sistemlerinin doğruluğunu ve güvenilirliğini artırmak için ses ve hareket sensörleri gibi birden fazla sensör ve modalitenin entegrasyonunu keşfedebilir.

## 6.REFERANSLAR

1. Nadeem, M. S., et al. (2023). Deep Labeller: Automatic Bounding Box Generation. *Multimedia Tools and Applications*. <https://doi.org/10.1007/s11042-023-15621-5>
2. Niebles, J. C., et al. (2008). Probabilistic Latent Semantic Analysis.
3. Papadopoulos, D. P., et al. (2017). Extreme Clicking.
4. Pathak, A., et al. (2024). Suspicious Activity Detection. *International Journal of Advanced Engineering Applications*, 1(1).
5. COCO Dataset: Lin, T. Y., et al. (2014). Microsoft COCO.
6. Datta, A., et al. (2002). Person-on-person Violence Detection.
7. Apene, O. Z., et al. (2024). Development of a Mathematical Model. *International Journal of Computer Applications*, 186(20).
8. COCO Dataset: Lin, T. Y., et al. (2014). Microsoft COCO.
9. Ren, S., et al. (2016). Faster R-CNN.
10. Web Repository: [<https://cocodataset.org/#download>]
11. J. Redmon, A. Farhadi, "YOLOv3: An Incremental Improvement" University of Washington.
12. Gao, H. (2023). YOLO-based Violence Detection in IoT Surveillance Systems. *International Journal of Advanced Computer Science and Applications*, 14(8), 123-135.
13. Xie, Y., & Zhang, S. (2022). "Real-Time Violence Detection in Surveillance Systems Using YOLOv4." *Journal of Visual Communication and Image Representation*, 72, 101721.
14. Norkobil Saydirasulovich, S.; Abdusalomov, A.; Jamil, M.K.; Nasimov, R.; Kozhamzharova, D.; Cho, Y.I. A YOLOv6-Based Improved Fire Detection Approach for Smart City Environments. *Sensors* 2023, 23, 3161.
15. M. İnkaya and H. Gurkan, "A YOLOv3 Based Smart City Application for Children's Playgrounds", *Journal of Innovative Science and Engineering (JISE)*, 2020. Available: 10.38088/jise.813664 [Accessed 16 July 2022].
16. [1] - YOLO ve Uygulamaları: J. Redmon, Santosh Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.
17. - MediaPipe Framework: Google Research, "MediaPipe: A framework for building multimodal applied machine learning pipelines," *GitHub Repository*, 2020. [Online]. Available: <https://github.com/google/mediapipe>
18. - LSTM Modeli: Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157-166, 1994.
19. - Sentetik Veri Üretimi: J. Goodfellow, I. Pouget-Abadie, M. Mirza, et al., "Generative adversarial nets," *Advances in Neural Information Processing Systems*, vol. 27, 2014.



20. [20.1] - J. Kowalski ve M. Szymański, "Violence Prediction in Surveillance Videos," *Advances in Computer Science*, vol. 23, no. 4, pp. 137-146, 2022. <https://ph.pollub.pl/index.php/acs/article/view/5978>
21. [20.2] - Zientek, K., & Sienkiewicz, J. (2021). A comparison of deep learning models for detecting violent events in surveillance videos. *Journal of Computer Science and Technology*, 36(2), 345-356. <https://doi.org/10.1007/s11390-021-1619-4>.
22. [21.1] - Al-Doghman, Y. M. E. S., & Alzahrani, S. A. (2022). Real-time hand gesture recognition for Turkish sign language using YOLOv4-CSP. *Journal of Electrical Engineering & Technology*, 101(23), 4154-4162. <https://doi.org/10.11591/jeet.v10i23.16012>.
23. [22.2] - Ozcift, A., & Sari, A. (2020). Real-time hand gesture recognition using deep learning for Turkish Sign Language. *Computers, Materials & Continua*, 63(3), 1373-1389. <https://doi.org/10.32604/cmc.2020.011123>.
24. [23.3] - Ekenel, H. K., & Akbatur, Y. (2021). Sign language recognition using convolutional neural networks for real-time applications. *IEEE Transactions on Neural Networks and Learning Systems*, 32(10), 4445-4455. <https://doi.org/10.1109/TNNLS.2021.3073812>.
25. [24.4] - Saeed, F., & Sultana, F. (2021). A deep learning approach for gesture recognition in Turkish Sign Language using CNN and LSTM. *Journal of Real- Time Image Processing*, 18(4), 1089-1104. <https://doi.org/10.1007/s11554-021-01087-4>.
26. RoboFlow, <https://roboflow.com>

