

Makale Adı	Distributed Intelligent Video Surveillance for Early Armed Robbery Detection based on Deep Learning																														
Makale Linki	<a href="#">(PDF) Distributed Intelligent Video Surveillance for Early Armed Robbery Detection based on Deep Learning</a>																														
Makale Konusu	Silahlı soygunlarda erken uyarı sistemi																														
Kullanılan Veri Setleri	SOHAS Weapon( <a href="#">Link</a> ) , Arka plan için COCO veri setleri, UCF-Crime veri seti kullanılmıştır. Veri seti için bakabileceğimiz kaynaklar: <a href="https://www.itemfix.com/">https://www.itemfix.com/</a>																														
Kullanılan yöntemler	YOLO ile sahnede silah olup olmadığı tespiti, silah varsa 3DCNN ile sahnede kesin soygun var mı yok mu tespiti yapılmıştır.																														
Özet	<p>Daha önce eşi benzeri olmadığı durumlar için false positive( Gerçekte soygun yok ama var diye tahmin edilmiş) durumuyla karşılaşılma olasılığı yüksektir.</p> <p>Çalışmada 3DCNN modeline video frameleri gönderilmiş silah olanlara soygun olmayanlara normal denilmiştir. 16,799 resim ile eğitilmiştir. 3 sistemden oluşmaktadır: IoT-based, sound-based, and vision-based. Kendi geliştirdikleri momentum algoritması ile false positive oranını azaltmışlar.</p> <p>Momentum algoritması her sınıfın son 6 güvenini içeren bir confidence(güven) queue (kuyruk)u oluşturur. Kuyruğa bir eleman eklendiğinde değerler toplanır. Bu toplam eşik değerle karşılaştırılır. Eğer toplam eşik değerden büyükse tespitin geçerli olduğu kabul edilir.</p> <p><b>Weapon Detection on the Edge:</b> (güvenlik kameraları, IoT cihazları veya düşük güçlü bilgisayarlar "edge device" olabilir.) SAHOS veri setindeki silah ve bıçaklar “cropping, rotating, generating mosaics, and varying the brightness” veri arttırma teknikleri ile çoğaltılmıştır. COCO veri setindeki çatal, uzaktan kumanda, cep telefonu, kitap ve diş fırçası gibi nesneleri içeren arka plan görüntüleri de dahil edilmiştir. Train %70, validation %20, test %10 olarak belirlenmiştir.</p> <table><tr><th>Dataset/Class</th><th>Gun</th><th>Knife</th><th>Background</th><th>Total</th></tr><tr><td>SW</td><td>3,611</td><td>3,468</td><td>0</td><td>7,079</td></tr><tr><td>COCO</td><td>0</td><td>0</td><td>2,641</td><td>2,641</td></tr><tr><td>Combined</td><td>3,611</td><td>3,468</td><td>2,641</td><td>9,720</td></tr><tr><td>Augmented</td><td>3,611</td><td>3,468</td><td>0</td><td>7,079</td></tr><tr><td>Total</td><td>7,222</td><td>6,936</td><td>2,641</td><td>16,799</td></tr></table> <p>GPU Tesla A100 ile YOLO’nun farklı versiyonları denenmiştir. Mean Average Precision (mAP), area under the curve (AUC), accuracy (Acc), precision (P), recall (R), and F1-score (F1s) gibi başarı metrikleri kullanılmıştır.</p> $g(x, y) = \alpha \cdot f(x, y) + \beta$ <p>Her bir karenin ön işleme için yukarıdaki formül kullanılmıştır. Alfa parlaklığı beta kontrastı simgelemektedir.</p> <p>Arka plan kaldırma adımı OPENCV’de “BackgroundSubtractorMOG” algoritması ile hareketli ve statik bölge ayrımı yapılır.</p>	Dataset/Class	Gun	Knife	Background	Total	SW	3,611	3,468	0	7,079	COCO	0	0	2,641	2,641	Combined	3,611	3,468	2,641	9,720	Augmented	3,611	3,468	0	7,079	Total	7,222	6,936	2,641	16,799
Dataset/Class	Gun	Knife	Background	Total																											
SW	3,611	3,468	0	7,079																											
COCO	0	0	2,641	2,641																											
Combined	3,611	3,468	2,641	9,720																											
Augmented	3,611	3,468	0	7,079																											
Total	7,222	6,936	2,641	16,799																											

$$momentum_c = \sum_{i=0}^n q_c(t-i) \cdot k^i$$

Momentum formülü ile tespit edilen silahların güveni ağırlıklı bir şekilde toplanmıştır.

k: Kullanıcı tarafından birim aralıkta tanınan parametre

c: Tek seferde işlenen kategori (bıçak veya tabanca)

t: Zaman

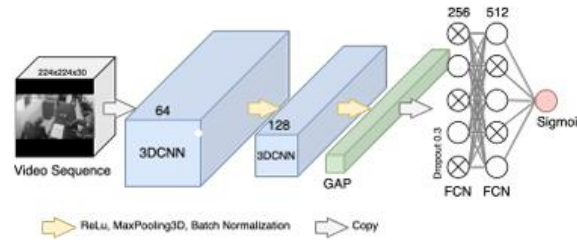
qc: Silah kategorisi için frame'in güven seviyesi

n frame dizisi verildiğinde formül kategori başına bir momentum puanı üretir ve bu puan Clouddaki Armed Robbery Detection modeline frame gönderilecek mi gönderilmeyecek mi belirlenir. (Tabanca için 1.05, bıçak için 0.7). Makale de n=5, k=0.5 seçilmiş ve false positivelere azalttığı görülmüştür.

**Armed Robbery Detection on the Cloud:** Edgede yapılan silah ön tespitini doğrulamak için kullanılır. UCF-Crime veri seti ve sosyal medyadan toplanarak oluşturulan veri seti kullanılmıştır. UCF-Crime veri setinden "shooting, robbery, normal" kategorileri çıkartılmıştır. Bu veri seti anomaliyi kısa periyotta gösteren uzun videolardan oluştuğu için anomali durumunun başlangıç ve bitiş yerleri manuel etiketlenmiştir. Sosyal medyadan toplanan verilerde "assault, armed, robbery" olarak etiketlenmiştir.

Keşifsel veri analizinde (EDA) video 224x224 hale getirilmiş ve Opencv ile grayscale(gri tonlama) uygulanmıştır. Örnekler 600 frameden (20 saniye) fazla olmamalı fazla ise birkaç parçaya ayrılmalıdır. Son durumda 309 video ve "robbery, normal" 2 tane etiket oluşturulmuştur. Overfit olmaması için

**Albumentations** kütüphanesinden "HorizontalFlip, ShiftScaleRotate, RandomBrightnessContrast, CLAHE, Affine, and RandomResizedCrop" fonksiyonları ile her iki etikettenden 1000er tane video oluşturulmuştur.



3DCNN Mimarisi:

2 tane GlobalAveragePooling3D katmanı, 2 tane Fully Connected Network katmanı, tüm katmanlarda RELU aktivasyon fonksiyonu, son katmanda Binary Cross Entropy (BCE) and Sigmoid fonksiyonu birleştirilmiştir.

$$BCE = -\frac{1}{T} \sum_{i=0}^T y_i \cdot \log(f(\hat{y}_i)) + (1 - y_i) \cdot \log(1 - f(\hat{y}_i)) \quad (3)$$

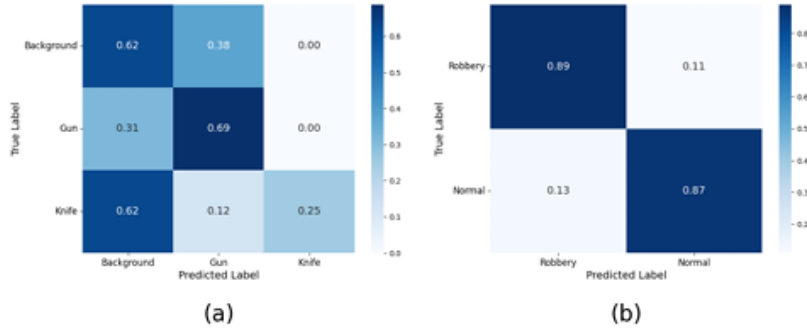
$$f(s_i) = \frac{1}{1 + e^{-s_i}} \quad (4)$$

Deneyisel Sonuçlar:

**Weapon Detection:** Veri 640×480 normalize edilip gri tonlama uygulanmıştır. %70-%20-%10 train, val, test şeklinde ayrılmıştır. YOLOv5, YOLOv7, YOLOs, and YOLOv8 modelleri test edilmiştir. Son olarak YOLOv5s seçilmiştir. Sonra kendi oluşturdukları veri setinde bazı ön işleme adımları ile test edilmiştir. Brightness and contrast (B&C), applying background removal (BR), compressing the model

with TensorRT (TRT), and implementing Momentum(M), frames black and white (B&W) dönüştürülmüştür.

#### Armed Robbery Detection:



a: YoloV5s

b: 3DCNN

#	Processes	Accuracy	F1-score	AUC	FPS
1	B&C	0.48	0.41	0.56	2.21
2	B&C, BR	0.5	0.49	0.57	2.16
3	B&C, BR, TRT	0.75	0.74	0.74	4.43
4	B&C, BR, TRT, B&W	0.62	0.61	0.61	4.37

Kendi oluşturdukları veri setinde YoloV5s performansları

#	FPS	Seconds	Accuracy	F1-score	AUC
1	15	2	0.88	0.87	0.88
2	6	5	0.82	0.53	0.68
3	5	6	0.82	0.58	0.71
4	1	30	0.6	0.2	0.47
5	0.5	60	0.57	0.19	0.45
6	0.25	120	0.67	0.38	0.58

Cloudda 3DCNN için en iyi performans testi

Sonuç: Bu sistem Jetson Nano karta Tensor RT ile entegre edilerek cloudda 30 fps alınmıştır. Ayrıca soygun tespit ettiğinde whatsapp veya telegramdan fotoğraf da çekip göndermektedir.

#### Makaleden Çıkarımlar

Bizim projemize uyarlayacak olursak. Hem RGB görüntüler hem termal görüntüler kullanılabilir veya ikisi de aynı anda kullanılarak proje geliştirilebilir. Bizim için en büyük zorluk bunu gerçek zamanlı yapmak olduğunu düşünüyorum. Rapor oluşturma kısmında ise silahlı veya bıçaklı bir görüntü tespit edildiğinde image to text dil modeline veya Vision Language modeline API ile gönderilerek raporun oluşturulabileceğimizi düşünüyorum. Alarm sistemini de malzemeleri tedarik ederek ekip deki donanım alanında bilgili kişilerle yapabiliriz.

Benzeyen arka plan görüntüleri ekledik ancak modele bu görüntüleri label olarak verirse mi daha iyi çıkarım yapar yoksa sadece arka plan olarak verirse mi? Örneğin bu çalışma da çatal, uzaktan kumanda, cep telefonu, kitap ve diş fırçası gibi nesneler arka plan olarak verilmiş ancak etiketli verip sadece silah ve bıçak nesnesi tahmin edildiğinde çıktı olarak almak test edilmemiştir.

Araştırmalarım doğrultusunda bu konu hakkında farklı çıkarımlar mevcuttur. İstemediğimiz nesneleri arka plan olarak vermek hem nesne tespiti

	hem görüntü sınıflandırmasında daha iyi performans sağladığını anlıyoruz ama tabi yine de test edilmesi gerekiyor.
<b>Makaleyi Okuyan</b>	Hakan KOÇAK