T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

GERÇEK ZAMANLI ANALİZ İÇİN UYARLANABİLİR HAREKET ÖĞRENME: BİR HİBRİT DERİN ÖĞRENME YAKLAŞIMI

Yunus AKKAYA

YÜKSEK LİSANS TEZİ
Veri Bilimi ve Büyük Veri Anabilim Dalı
Veri Bilimi ve Büyük Veri Programı

Danışman
Doç. Dr. Mert BAL

Haziran, 2025

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

GERÇEK ZAMANLI ANALİZ İÇİN UYARLANABİLİR HAREKET ÖĞRENME: BİR HİBRİT DERİN ÖĞRENME YAKLAŞIMI

Yunus AKKAYA tarafından hazırlanan tez çalışması 19.06.2025 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Veri Bilimi ve Büyük Veri Anabilim Dalı Veri Bilimi ve Büyük Veri Programı **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Doç. Dr. Mert BAL
Yıldız Teknik Üniversitesi
Danışman

Jüri Üyeleri

Doç. Dr. Mert BAL, Danışman
Yıldız Teknik Üniversitesi _____

Doç. Dr. Ramazan TEKERCİOĞLU, Üye
Yıldız Teknik Üniversitesi _____

Dr. Öğr. Üyesi Meryem UZUN PER, Üye
İstanbul Sağlık ve Teknoloji Üniversitesi _____

Danışmanım Doç. Dr. Mert BAL sorumluluğunda tarafımca hazırlanan Gerçek Zamanlı Analiz İçin Uyarlanabilir Hareket Öğrenme: Bir Hibrit Derin Öğrenme Yaklaşımı başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Yunus AKKAYA

İmza

Aileme

TEŞEKKÜR

Yüksek lisans sürecimde kendimi teknik anlamda ileriye taşımama yardımcı olmuş tüm hocalarıma, bu çalışmayı yapmamda bana destek ve yardımcı olan danışmanım Doç. Dr. Mert Bal'a ve bana her zaman destek olmuş olan aileme teşekkürlerimi sunarım.

Yunus AKKAYA

İÇİNDEKİLER

ŞEKİL LİSTESİ vii

TABLO LİSTESİ viii

ÖZET ix

ABSTRACT x

1 GİRİŞ 1

1.1 Çalışmanın Amacı ve Kapsamı 1
1.2 Veri Seti ve Ön İşleme 2
1.3 Çalışmanın Önemi ve Katkıları 3

2 LİTERATÜR 4

2.1 Görüntü İşleme Tarihçesi 4
2.2 Yolo ve Nesne Tespiti 4
2.3 OpenPose 5
2.4 Hareket Algılama 6

3 METODOLOJİ 8

3.1 Vuruş Tahminleme Modeli Eğitimi 8
3.2 Ön İşleme Adımları 8
3.3 Veri Seti Yapısı 9
3.4 Model Yapıları 10
3.4.1 ResNet (Residual Networks) 10
3.4.2 MobileNet 10
3.4.3 EfficientNet 12
3.4.4 GCN (Graph Convolutional Networks) 13
3.4.5 ST-GCN (Spatial-Temporal Graph Convolutional Networks) 14
3.4.6 ST-Transformer (Spatio-Temporal Transformer) 15
3.4.7 Uzun Kısa Süreli Bellek (LSTM) ile Hareket Tahmini . . . 16
3.5 Yolo ile Konum Tespiti - ROI (Region of Interest) 18
3.6 Hibrit Model Yapısı 19

3.6.1 EfficientNet Bileşeni 19
3.6.2 ST-GCN ve ST-Transformer Bileşeni 20
3.6.3 Hibrit Modelin Birleştirme Süreci 20
3.6.4 LSTM ile Sonraki Hareket Tahmini ve Uyarlanabilir Öğrenme 20

4 UYGULAMA 21
4.1 Veri Ön İşleme 21
4.2 Model Eğitimi 21
4.2.1 Başarı Metrikleri 22
4.2.2 Kayıp (Loss) ve Doğruluk (Accuracy) Fonksiyonları 22
4.2.3 ResNet 22
4.2.4 MobileNet 24
4.2.5 EfficientNet 26
4.3 ST-GCN ve ST-Transformer ile Görüntü Serisi İşleme 28
4.3.1 ST-GCN ile Görüntü Serisi İşleme 29
4.3.2 Çizge Evrişim Ağı (GCN) 29
4.3.3 ST-Transformer ile Görüntü Serisi İşleme 30
4.3.4 LSTM 30
4.3.5 ST-GCN ile Oluşturulan Hibrit Model Sonuçları 31
4.3.6 ST-Transformer ile Oluşturulan Hibrit Model Sonuçları . . . 32
4.3.7 Entegrasyon ve Video Analizi 34

5 SONUÇ 36

KAYNAKÇA 38

TEZDEN ÜRETİLMİŞ YAYINLAR 40

ŞEKİL LİSTESİ

Şekil 3.1 ResNet'te atlamalı bağlantı yapısı **10**
Şekil 3.2 MobileNet'te ayrılabilir evrişim akışı **11**
Şekil 3.3 EfficientNet temel blok yapısı **12**
Şekil 3.4 ST-GCN'de zamansal iskelet çizge yapısı **14**
Şekil 3.5 ST-Transformer işlem akışı **15**
Şekil 3.6 LSTM ile ardışık hareket tahmini **16**
Şekil 3.7 Hibrit model genel mimarisi: EfficientNet + ST-GCN/Transformer + LSTM **19**
Şekil 4.1 Ön İşlem Yapılmış Örnek Görüntüler [9] **21**
Şekil 4.2 ResNet Eğitimi **23**
Şekil 4.3 ResNet Tahmin Sonuçları **24**
Şekil 4.4 MobileNet Eğitimi **25**
Şekil 4.5 MobileNet Tahmin Sonuçları **26**
Şekil 4.6 EfficientNet Eğitimi **27**
Şekil 4.7 EfficientNet Tahmin Sonuçları **27**
Şekil 4.8 Mediapipe ile Çıkartılmış İskelet Verisi Örnekleri **28**
Şekil 4.9 ST-GCN Eğitimi **31**
Şekil 4.10 ST-GCN Başarı Metrikleri **31**
Şekil 4.11 ST-GCN Tahmin Matrisi **32**
Şekil 4.12 ST-Transformer Eğitimi **32**
Şekil 4.13 ST-Transformer Başarı Metrikleri **33**
Şekil 4.14 ST-Transformer Tahmin Matrisi **33**
Şekil 4.15 LSTM ile bir sonraki hareketin tahminine yönelik örnek kare dizisi. Model, geçmiş karelerden aldığı hareket bilgisini kullanarak bir sonraki pozisyonu öngörmektedir. **34**
Şekil 4.16 YOLO modeli ile boksörlerin maç anında tespit edilerek etiketlenmesi. Her sporcu farklı ID'lerle izlenmekte ve pozisyonları gerçek zamanlı olarak sınırlayıcı kutular ile işaretlenmektedir. [17] **35**

TABLO LİSTESİ

Tablo 4.1 CNN Tabanlı Modellerin Sınıflandırma Performans Karşılaştırması **28**
Tablo 4.2 Hibrit Modellerin Performans Karşılaştırması **34**

ÖZET

Gerçek Zamanlı Analiz İçin Uyarlanabilir Hareket Öğrenme: Bir Hibrit Derin Öğrenme Yaklaşımı

Yunus AKKAYA

Veri Bilimi ve Büyük Veri Anabilim Dalı
Yüksek Lisans Tezi

Danışman: Doç. Dr. Mert BAL

Zamanın hızla akışı ile birlikte dönem dönem bazı büyük toplumsal etki yaratan teknolojik gelişmelere şahit olunmaktadır. Geçmişte sanayi devrimi ve internet gibi dönüşümler toplumları derinden etkileyen gelişmelere örnek gösterilebilir. Günümüzde ise yapay zeka teknolojileri benzer çapta bir dönüşümü tetiklemektedir. Yapay zekanın geniş kullanım alanına rağmen mevcut uygulamaların çoğunlukla bilgisayar odaklı ve insan hareketini azaltmaya yönelik olduğu görülmektedir. Ayrıca, adil değerlendirme ve objektif karşılaştırma gerektiren alanlarda, insan kaynaklı önyargıyı azaltmaya yönelik yapay zeka uygulamalarının halen sınırlı olduğu söylenebilir. Bu çalışmada, daha fazla insan ve hareket odaklı olunması amacıyla, yapay zeka destekli görüntü işleme teknikleri spor müsabakaları ve antrenmanlarda değerlendirme aracı olarak kullanılmıştır. Çalışmada, boks müsabakalarındaki sporcuların hareketleri görüntü işleme teknikleri ve farklı derin öğrenme modelleri kullanılarak analiz edilmiş, farklı yaklaşımlar karşılaştırılarak objektif ve tutarlı bir değerlendirme sistemi geliştirilmiştir.

Anahtar Kelimeler: Hareket algılama, görüntü işleme, derin öğrenme.

YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

ABSTRACT

Adaptive Action Learning for Real-Time Analysis: A Hybrid Deep Learning Approach

Yunus AKKAYA

Department of Data Science and Big Data
Master of Science Thesis

Supervisor: Assoc. Prof. Mert BAL

With the rapid progression of time, we periodically witness significant technological advancements that create profound societal impacts. Historically, transformations such as the industrial revolution and the advent of the internet serve as prominent examples. Today, artificial intelligence (AI) technologies are triggering a comparable scale of transformation. Despite AI's broad spectrum of applications, current implementations tend to be predominantly computer-oriented, aiming to minimize human movement and involvement. Moreover, AI applications designed to reduce human biases in fair assessments and objective comparisons remain relatively limited. In this study, AI-driven image processing techniques were utilized as an evaluation tool in sports competitions and training sessions, aiming to be more human-centric and movement-oriented. Specifically, athletes' movements during boxing competitions were analyzed using image processing techniques and various deep learning models, with different approaches being compared to develop an objective and consistent evaluation system.

Keywords: Action detection, image processing, deep learning.

YILDIZ TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

1
GİRİŞ

1.1 Çalışmanın Amacı ve Kapsamı

Yapay zeka teknolojileri, sağlık, finans, eğitim ve otomasyon gibi birçok sektörde köklü dönüşümlere neden olarak kullanım alanlarını hızla genişletmektedir. Özellikle değerlendirme, analiz ve tahmin süreçlerinde sunduğu avantajlar sayesinde geleneksel yöntemlere kıyasla önemli üstünlükler sağlamaktadır. Ancak, insan faktörünün devrede olduğu değerlendirme süreçleri genellikle subjektif unsurlar ve önyargılarla zayıflamakta; bu durum adil ve tutarlı kararların alınmasını zorlaştırmaktadır. Yapay zekâ uygulamaları, bu sınırlılıkların üstesinden gelerek insan kaynaklı hataların azaltılmasını ve objektif analiz süreçlerinin oluşturulmasını mümkün kılmaktadır.

Diğer yandan, yapay zekânın iş yükünü azaltma konusundaki başarısı, insan kapasitesinin bazı yönlerden körelmesine neden olabilecek bir duraklama yaratma riski de taşımaktadır. Bu nedenle teknolojik gelişmelerin, insan yeteneklerini destekleyici ve geliştirici biçimde tasarlanması büyük önem arz etmektedir. Teknolojinin insanı ikame eden değil, onu tamamlayan bir unsur olarak kurgulanması, sürdürülebilir ve kalıcı faydaların elde edilmesini sağlayacaktır.

Bu çalışmanın temel amacı, yapay zeka ve görüntü işleme teknikleri kullanılarak sporcuların hareketlerini analiz eden, sınıflandıran ve sonraki hamleyi tahmin edebilen bütüncül bir sistem tasarlamaktır. Özellikle boks gibi hızlı ve dinamik yapıya sahip spor branşlarında, doğru analiz ve anlık geri bildirimler sunabilen hafif ve verimli yapay zeka tabanlı çözümlere duyulan ihtiyaç giderek artmaktadır. Geleneksel değerlendirme yöntemleri; yorgunluk, dikkat eksikliği ve öznel yargılar nedeniyle sporcu performansının sağlıklı bir şekilde ölçülmesini engelleyebilir. Bu bağlamda geliştirilen sistemin; hem antrenman süreçlerinde verimliliği artırması, hem de müsabakalarda adaletin sağlanmasına katkıda bulunması hedeflenmiştir.

Bu doğrultuda, çalışma üç temel aşamada yapılandırılmıştır:

- **Statik Görüntü Modellemesi:** İlk aşamada, Roboflow platformu üzerinden temin edilen ve çeşitli boks vuruşlarını içeren statik görüntü veri seti ile ResNet, MobileNet ve EfficientNet modelleri eğitilmiş; performansları karşılaştırılmıştır.
- **Hibrit Model Geliştirme:** En yüksek doğruluk oranına sahip olan EfficientNet modeli iskelet verisiyle desteklenmiş ve bu yapı ST-GCN ile ST-Transformer algoritmalarıyla birleştirilerek iki ayrı hibrit model oluşturulmuştur. Bu aşamada, özel olarak çekilmiş kısa videolardan oluşan bir veri seti kullanılmıştır.
- **Sonraki Hareket Tahmini:** Hibrit model çıktıları, ardışık hareket dizilerini değerlendirmek üzere LSTM (Long Short-Term Memory) modeline aktarılmış; bu sayede sporcuların bir sonraki hareketlerinin tahmin edilmesi mümkün hale gelmiştir.

Geliştirilen sistem, yalnızca boks branşıyla sınırlı kalmayıp; farklı spor dallarına entegre edilebilecek esnek bir mimariye sahiptir. Bu özelliğiyle sporcu gelişimi, antrenman optimizasyonu ve hakem destek sistemleri gibi pek çok alana katkı sunabilecek potansiyele sahiptir.

1.2 Veri Seti ve Ön İşleme

Bu çalışmada standard görüntü sınıflandırma ve vuruş kategorisi tahmin işlemi için roboflow üzerinden bulunmuş olan ve bir boksörün torbaya karşı yaptığı teknik vuruşları içeren farklı açılardan çekilmiş fotoğraflardan oluşan bir veri seti kullanılmıştır. Daha sonra ST-GCN ve ST-Transformer modelleri ile hazırlanacak zamansal akışa bağlı tespit için ise sıralı görüntüler gerektiğinden, onun için ayrıca bir video veri seti hazırlanmıştır.

Ön işleme aşamalarında, sabit görüntüleri içeren veri seti üzerinde gerekli yerlerin kırpılması, bazı filtreler aracılığı görüntüsel gürültüler eklenmesi ve çeşitli veri artırımı yöntemleri ile model performansı artırılması sağlanmıştır.

Test ortamı olarak, videoların kare kare işleneceği ve her bir karede sporcuların konumları belirlenecek şekilde YOLO tabanlı model kullanılarak sporcuların sınır kutularının (bounding box) belirleneceği ve her bir sporcunun hareketlerinin izleneceği bir ortam oluşturulmuştur. Daha sonra, sınırlar içerisindeki bölgeler (ROI - Region of Interest) derin öğrenme modellerine giriş olarak verilmiştir. Görüntülerin normalize edilmesi, yeniden boyutlandırılması ve veri artırma teknikleri (augmentation) kullanılarak modelin performansı gözlenmiştir.

1.3 Çalışmanın Önemi ve Katkıları

Çalışmanın spor ve hareket analizi süreçlerine getirdiği objektiflik ve gerçek zamanlı değerlendirme yeteneği ile geleneksel yöntemlerle kıyaslandığında, insan hatalarının ve önyargılarının azaltılması mümkün hale gelirken, sporcuların performansının sistematik bir şekilde takip edilmesi ve geliştirilmesi sağlanmaktadır. Ayrıca geliştirilen modelin geniş bir uygulama potansiyeline sahip olması, farklı spor branşlarında ve eğitim süreçlerinde de kullanılabilirliğini artırmakta, böylece genel spor bilimine ve antrenman yöntemlerine katkı sağlamaktadır.

2
LİTERATÜR

2.1 Görüntü İşleme Tarihçesi

Bilgisayar ile görüntü işleme çalışmaları tarihte 1960'lara dayanır. Ilk başlarda yalnızca bilgisayara bir kamera bağlayıp görüntüler üzerinde analiz yapma işleriyle başlayan çalışmalar daha sonra kenar algılama algoritmalarının kullanılması, nesnelerin ve insanların görüntülerin tespit edilmesi için oluşturulan algoritmalar ve evrişimsel sinir ağları (CNN) olarak adlandırdığımız derin öğrenme yapılarının ortaya çıkışı ile teknik anlamda oldukça detay kazandı [1]. Günümüzde ise bu tarz modellerin bir çok farklı görevi yerine getirmek üzere bir çok farklı varyasyonu bulunmaktadır. Örneğin nesne tespiti için kullanılan Yolo modeli bu alanda oldukça başarı sağlamış ve farklı türde çalışmalar için de kapsam kazanmıştır [2, 3].

2.2 Yolo ve Nesne Tespiti

Gerçek zamanlı nesne algılama alanında önemli bir gelişim gösteren YOLO modeli, her versiyonunda hız ve doğruluk açısından kayda değer ilerlemeler sağlamıştır. İlk çalışmaları 2015 yılına dayanan ve Joseph Redmon ve ekibi tarafından geliştirilen YOLO'nun ilk versiyonu, nesne algılama işlemini bir regresyon problemi olarak ele alarak görüntüyü belirli bir ızgara sistemine böler ve her hücrede tek bir nesne algılamayı hedefler [2]. Bu yaklaşım, geleneksel nesne algılama yöntemlerine göre büyük bir hız kazanımı sunmuştur.

2016 yılında YOLOv2, aynı zamanda "YOLO9000" olarak adlandırılan versiyon, algılama doğruluğunu artırmak amacıyla daha derin bir CNN (Convolutional Neural Network) yapısı olan Darknet-19'u kullanarak önceki sürümde yaşanan sınırlamaların üstesinden gelmeye çalışmıştır [3].

2024 Şubat ayında yayınlanan YOLOv9 ile, derin sinir ağlarında bilgi kaybı sorununu ele almak için Programlanabilir Gradyan Bilgisi (PGI) önerisi ve nesne algılama görevlerinde oldukça iyi sonuçlar elde eden Genelleştirilmiş Verimli

Katman Toplama Ağı (GELAN) adı verilen yeni bir hafif ağ mimarisinin tasarımı yer almaktadır [4].

2.3 OpenPose

İnsan iskelet yapısını anlamak için kullanılan modeller arasında OpenPose büyük bir etki oluşturmuştur [5]. OpenPose, görüntüdeki bireylerle vücut parçalarını ilişkilendirmeyi öğrenmek için parametrik olmayan bir temsil (nonparametric representation) olan Parça Yakınlık Alanları (Part Affinity Fields, PAFs) yöntemini kullanmaktadır. Yöntem şu adımlardan oluşmaktadır:

· Vücut parçası konumlarının 2D güven haritaları (2D confidence maps) ve Parça Yakınlık Alanları'nın 2D vektör alanları (2D vector fields of part affinity fields, PAFs) setini tahmin etmek.

· Güven haritaları (confidence maps) ve PAF'ler açgözlü çıkarım (greedy inference) ile ayrıştırarak görüntüdeki tüm kişilerin 2D anahtar noktalarını çıkarmak.

· OpenPose, vücut bölgelerinin konum ve yönelimini kodlayan PAF'leri tahmin etmek için CNN yapısı kullanılmaktadır. PAF'ler daha sonra aday parça tespitleri arasındaki ilişkilendirmeyi hesaplamak için kullanılır ve çoklu kişilerin tam vücut duruşlarını oluşturmak için açgözlü gevşeme algoritması (greedy relaxation algorithm) kullanılır.

Çalışmada kullanılan yöntemler arasında, görüntülerde insan vücudu, ayak, el ve yüz anahtar noktalarını birlikte tespit eden gerçek zamanlı çoklu kişi sistemi olan OpenPose, ve vücut anahtar noktası tespiti doğruluğunu artıran yeni bir ayak anahtar noktası tespit algoritmasının tanıtılması bulunmaktadır.

Yöntemler:

· PAF ve güven haritalarını kullanarak parça tespiti ve ilişkilendirmesinin ortak öğrenilmesi.

· Çoklu kişi ayrıştırması için aşgözlü ayrıştırma algoritması.

· ReLU katmanları yerine PReLU kullanımı ve momentumlu SGD yerine Adam optimizasyonu.

· MPII, COCO ve COCO+foot dahil olmak üzere çeşitli veri setlerinde eğitim yapılması.

Tahminleri iyileştirmek için her aşamada ara denetim ile iteratif tahmin mimarisi kullanmak.

2.4 Hareket Algılama

Hareket algılama ve insan iskeletinin tanımlanması için en uyarlanabilir alanlardan biri sportif faaliyetlerdir. Özellikle çizge tabanlı modeller bu alanda oldukça yüksek başarı sağladığı görülmüştür. Örnek olarak; bu alanda yapılan bir çalışmada çizge evrişim ağları, çoklu zaman evrişimi (multi-scale time convolution) ve dikkat katmanları (attention layers) birlikte kullanmıştır. Sporcuların dijital görüntüleri üzerinden temel egzersizlerin sınıflandırması üzerine çalışılmıştır [6]. Burada; ışık ve açı gibi problemlerin görüntü sınıflandırılması üzerindeki olumsuz etkisi düşürülmeye çalışılmıştır. Veri seti olarak NTU-RGB+D120 ismi ile bahsedilen, 108 deneğin katıldığı ve 32 farklı ayar kullanılarak 114.490 eylem görüntüsünü içeren bir veri seti ile (538 verinin ulaşılabilir olmadığı belirtilmiş) çalışma gerçekleştirilmiştir. Veri setinde değerlendirme yapmak üzere CS (Cross Subjects) ve SS (Cross Set Number) olarak iki ayar türünden bahsedilmiştir. Bu kapsamda, literatürdeki diğer kullanılan algoritmalara kıyasla en yüksek olarak, CS ortamında %5,95 ve SS ortamında %4,73 oranında bir iyileştirme sağlanmıştır. Bu çalışmadan yola çıkılarak TCN tabanlı zaman katmanları ve dikkat katmanlarının çizge tabalı modeller ile birlikte kullanımında, fiziksel hareket tahminleme işleminde başarının arttığı gözlemlenebilmektedir.

Vücut duruş tahmini ve sınıflandırılması ile ilgili etki yaratan çalışmalardan biri de AlphaPose olarak sunulan yapıdır. Bu çalışma, hem hız hem de doğruluk açısından mevcut son teknoloji yöntemlere göre önemli iyileştirmelerle gerçek zamanlı olarak doğru tüm vücut poz tahmini ve takibini birlikte gerçekleştiren bir sistem olan AlphaPose'u sunmaktadır. Sunulan metot simetrik integral regresyonu olarak tanımlanan yöntem ile özellikle yüz ve el anahtar noktalarının tespitinde oldukça iyi çalıştığı gözlemlenmektedir. Önerilen sistem, COCO-wholebody, COCO, PoseTrack ve Halpe-FullBody poz tahmini veri setinde hem hız hem de doğruluk açısından önemli bir gelişme göstermektedir. Özellikle birden fazla kişinin duruş analizini yapmayı hedefleyen bu çalışma PoseTrack datasetini bu amaçla doğrulama için kullanmıştır. Önerilen yöntem, hızlı ve hassas yerelleştirme için Simetrik İntegral Anahtar Nokta Regresyonu (Symmetric Integral Keypoint Regression, SIKR), gereksiz insan tespitlerini ortadan kaldırmak için Parametrik Poz Non-Maximum Bastırma (Parametric Pose Non-Maximum Suppression, P-NMS) ve poz tahmini ile takibi birlikte gerçekleştirmek için Poz Farkında Kimlik Gömme (Pose Aware Identity Embedding) gibi birkaç yeni

teknik kullanmaktadır. Yazarlar, yeni bir simetrik integral kayıp fonksiyonu (symmetrical integral loss function) ve temeli ResNet50 veya ResNet152 olan bir FastPose ağı önermektedirler. Ayrıca, deformasyonel evrişim katmanı (deformable convolutional layer) ve çoklu alan bilgi damıtma (multi-domain knowledge distillation) yöntemlerini de benimsemektedirler [7].

Başka bir çalışma ise, hareket tanıma doğruluğunu artırmak için nedensellik ve zamansal-uzamsal çizge evrişim ağı (spatiotemporal graph convolution network) yapısını birleştirerek antrenmanlarda duruş analizine yönelik geliştirilmiş bir derin öğrenme yöntemi önermektedir. Kullanılan yöntem, eklem koordinat dizisine göre eklemlerin nedensel katsayılarını hesaplamakta ve ardından bir nedensel katsayı matrisi (causal coefficient matrix) oluşturmaktadır. Bu matris daha sonra çizge evrişim ağına uygulanmış ve kemik verilerinden derin özellikler elde edilmiştir.

Yöntem, eklemler arasındaki nedensel ilişkileri hesaplamak için yakınsak çapraz haritalama (Convergent Cross Mapping, CCM) kullanmakta ve CCM'ye dayalı olarak kenar ağırlıkları tahsis etmektedir. Kenar ağırlıkları daha sonra çizge komşuluk matrisine yerleştirilerek çizge evrişim ağını güçlendirmektedir. Önerilen yöntemin dört kamuya açık, hareket veri seti üzerinde, mevcut ana akım yöntemlerini geride bıraktığı gösterilmiştir [8].

3
METODOLOJİ

3.1 Vuruş Tahminleme Modeli Eğitimi

Çalışmanın ilk aşamasında, teknik boks vuruşlarını sınıflandırmaya yönelik olarak Roboflow üzerinden temin edilen BoxPunch Detector veri seti kullanılmıştır [9]. Veri seti, farklı açılardan çekilmiş ve çeşitli yumruk hareketlerini içeren toplam 744 görüntüden oluşmaktadır. Görseller, jab, cross, hook ve uppercut gibi farklı yumruk türlerine göre etiketlenmiş olup, %80 eğitim, %10 doğrulama ve %10 test şeklinde ayrılmıştır.

Veriler 640x640 piksel boyutuna yeniden boyutlandırılmış ve sınırlayıcı kutular (bounding boxes) ile etiketlenmiştir. Sınıflandırma için ResNet, MobileNet ve EfficientNet modelleri kullanılmış; bu modellerin doğruluk, kayıp ve hesaplama süreleri değerlendirilerek en iyi performans gösteren EfficientNet modeli seçilmiştir.

Çalışmanın devamında ise zamansal özelliklerin analizinin de yapılabilmesi amacıyla ST-GCN ve ST-Transformer modellerini de içeren hibrit yapı, 295 kısa video içeren ve temel Teknik vuruşları barındıran bir veri seti ile eğitilmiştir. Belirlenmiş her bir vuruş (Jab, Cross, Hook, Uppercut) ve bir de ek olarak vuruş içermeyen görüntüler içeren 5 sınıf ve her bir sınıfta ortalama 59 kısa video bulunan bir veri seti kullanılmıştır.Bu video veri seti, manuel olarak bu çalışma için çekilmiş ve hazırlığı yapılmıştır.

3.2 Ön İşleme Adımları

Veri setinin etkin bir şekilde işlenebilmesi ve model performansının artırılabilmesi amacıyla aşağıdaki ön işleme adımları uygulanmıştır:

- **Kırpma (Cropping):** Arka plan bilgisi çıkarılarak yalnızca boksörlerin bulunduğu bölge korunmuştur.

- **Filtreleme:** Görüntü kalitesini artırmak amacıyla çeşitli görüntü işleme filtreleri uygulanmıştır.
- **Gürültü Azaltma:** Rastgele gürültüler filtrelenerek modelin dikkatini önemli bölgelere odaklaması sağlanmıştır.
- **Kontrast Ayarı:** Detayların belirginleştirilmesi için kontrast optimizasyonu yapılmıştır.
- **Grileştirme (Grayscale):** Görüntüler RGB formatından gri tonlamaya dönüştürülerek işlem yükü azaltılmıştır.
- **Normalizasyon:** Tüm görüntüler aynı boyuta ölçeklenmiş ve pikseller [0, 1] aralığına normalize edilmiştir.

3.3 Veri Seti Yapısı

Çalışmada kullanılan iki farklı veri seti bulunmaktadır:

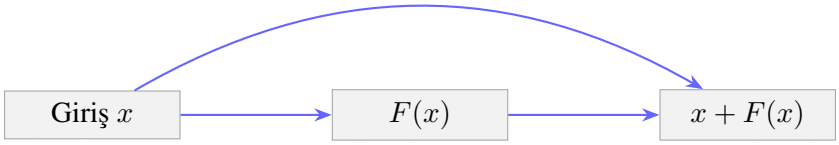
- **Fotoğraf Veri Seti:** Roboflow-Box Punch veri seti, 744 etiketli görüntü içermektedir. Görseller farklı vuruş türlerini içermekte olup sınıflar: *jab, cross, hook, uppercut* ve *nopunch* şeklindedir.
- **Video Veri Seti:** Özel olarak oluşturulan bu veri seti, her biri yaklaşık 2-3 saniyelik 295 kısa videodan oluşmaktadır. Her sınıf için yaklaşık 59 video örneği mevcuttur. Bu veri seti ST-GCN ve ST-Transformer modellerinin eğitimi için kullanılmıştır.

Görüntüler, 640x640 piksel boyutuna yeniden boyutlandırılmış ve her bir görüntüde boksörlerin gerçekleştirdiği farklı yumruk türleri (örneğin, jab, hook, uppercut) etrafında sınırlayıcı kutular (bounding boxes) ile ayrıntılı olarak etiketlemeler yapılmıştır. Veri seti, modelin genelleme yeteneğini artırmak amacıyla çeşitli aydınlatma koşulları, arka plan çeşitliliği gibi faktörleri içermektedir. Ayrıca, veri artırma teknikleri uygulanarak her bir eğitim örneği için üç çıktı üretilmiştir; bu teknikler arasında yatay ve dikey çevirme, kırpma, döndürme, parlaklık ayarlamaları ve gürültü ekleme bulunmaktadır.

3.4 Model Yapıları

3.4.1 ResNet (Residual Networks)

ResNet, 2015 yılında Microsoft Research tarafından geliştirilen ve derin sinir ağlarının eğitiminde karşılaşılan "kaybolan gradyan" problemini çözmeyi amaçlayan bir mimaridir [10]. ResNet'in en temel yeniliği, "residual block" adı verilen yapı ile gelen sinyali doğrudan bir sonraki katmana aktaran atlamalı bağlantılar (skip connections) kullanmasıdır. Bu sayede, modelin yalnızca öğrenilmesi gereken artık (residual) fonksiyonu öğrenmesi sağlanır.



Şekil 3.1 ResNet'te atlamalı bağlantı yapısı

Bu bağlantılar sayesinde model, her katmanda öğrenmesi gereken dönüşümü doğrudan öğrenmek yerine, sadece girdiye eklenmesi gereken farkı öğrenir. Bu yapı aşağıdaki gibi formüle edilir:

$$y = \mathcal{F}(x, \{W_i\}) + x \tag{3.1}$$

Burada x giriş vektörü, $\mathcal{F}(x)$ ise residual fonksiyondur. Katman derinliği arttıkça bile, bu atlamalı bağlantılar sayesinde gradyan akışı bozulmaz ve eğitim stabil kalabilir.

Artık (Residual) bloklar, özellikle derin ağlarda optimizasyon sürecini kolaylaştırır ve modelin genelleme performansını artırır. Sadece öğrenme kabiliyetini değil, aynı zamanda modelin parametrik verimliliğini de artırmaktadır.

Bu çalışmada ResNet18 mimarisi, karşılaştırma modellerinden biri olarak kullanılmış ve statik yumruk sınıflandırma görevlerinde 89% doğrulukla oldukça tatmin edici sonuçlar vermiştir. Ancak bazı benzer sınıflar arasında karışıklıklar da gözlenmiştir.

3.4.2 MobileNet

MobileNet, mobil ve gömülü cihazlarda verimlilik sağlamak üzere tasarlanmış bir CNN mimarisidir [11]. Temel özelliği, geleneksel evrişim katmanlarını iki aşamalı bir işlem olan "derinlik ayrılabilir evrişimler" (depthwise separable convolutions)

ile değiştirmesidir. Bu yapı, derinlik evrişimi (her giriş kanalına ayrı filtre) ve noktasal evrişim (1x1 evrişim) işlemlerini ardışık olarak uygular. Hesaplama yükünü ve parametre sayısını ciddi oranda azaltarak hem hız hem de enerji verimliliği açısından büyük avantaj sağlayabilmektedir.



Şekil 3.2 MobileNet'te ayrılabilir evrişim akışı

Bu yapı iki adımdan oluşur:

- **Derinlik Bazlı Evrişim (Depthwise Convolution)**: Her giriş kanalına ayrı ayrı bir evrişim filtresi uygulanır. Bu işlem uzamsal filtreleme sağlar.
- **Noktasal Evrişim (Pointwise Convolution)**: Elde edilen ara çıktılar, kanallar arası ilişkiyi modellemek amacıyla 1x1 evrişim katmanı ile birleştirilir.

Bu ayrım, klasik evrişim yapısına kıyasla yaklaşık şu kadar parametre azaltımı sağlar:

$$\frac{D_K \cdot D_K \cdot M \cdot N}{D_K \cdot D_K \cdot M + M \cdot N} \approx \frac{1}{N} + \frac{1}{D_K^2} \tag{3.2}$$

Burada D_K filtre boyutu, M giriş kanal sayısı ve N çıkış kanal sayısıdır. Örneğin, 3x3 filtre ve 32 kanallı bir yapı için %80–90 oranında hesaplama tasarrufu sağlanabilir.

MobileNet ayrıca, parametre verimliliği sağlayan iki hiperparametre ile ölçeklenebilir:

- **Genişlik Çarpanı (Width Multiplier)** (α): Kanal sayısını azaltarak modelin genişliğini kontrol eder.
- **Çözünürlük Çarpanı (Resolution Multiplier)** (ρ): Giriş çözünürlüğünü azaltarak işlem maliyetini düşürür.

Bu çalışmada kullanılan MobileNet mimarisi, düşük donanım gereksinimlerine rağmen yüksek doğruluk sağlayarak %96 başarıya ulaşmıştır. Ancak kroşe ve direkt vuruşlar arasında küçük çaplı karışıklıklar gözlenmiştir.

3.4.3 EfficientNet

EfficientNet, hem model doğruluğunu hem de verimliliğini artırmayı hedefleyen bir evrişimsel sinir ağı ailesidir [12]. Sistematik bir ölçekleme stratejisi olan bileşik ölçekleme (compound scaling) ilkesine dayanır. Bu yapıda temel bir model olan EfficientNet-B0'ı otomatik makine öğrenimi (AutoML) yöntemi ile optimize ettikten sonra, bu yapıyı derinlik, genişlik ve çözünürlük boyutlarında birlikte ölçekleyerek daha büyük modeller (B1–B7) üretilir. Mimarisinde özel bir blok yapısı olan Mobil Ters Dar Boğaz Katmanı (Mobile Inverted Bottleneck Convolution) yer alır; bu bloklar, düşük hesaplama maliyetiyle yüksek doğruluk sağlayan, derinlik ayrılabilir evrişimleri ve sıkıştırma-genişletme mekanizması olarak adlandırılan iki yaklaşımı birleştirmektedir.



Şekil 3.3 EfficientNet temel blok yapısı

Geleneksel ölçekleme yaklaşımları; derinlik, genişlik veya çözünürlük artırımı gibi tek boyutlu stratejilere dayanır. EfficientNet ise bu soruna daha dengeli ve teorik temelli bir çözüm olan **compound scaling** yöntemini sunar. Bu yöntem, modelin derinliğini (d), genişliğini (w) ve giriş çözünürlüğünü (r) birlikte ölçekler:

$$\begin{aligned} \text{depth: } & d = \alpha^\phi \\ \text{width: } & w = \beta^\phi \\ \text{resolution: } & r = \gamma^\phi \\ \text{subject to: } & \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \quad (\text{FLOPs constraint}) \end{aligned} \tag{3.3}$$

Burada ϕ ölçekleme faktörü, α, β, γ ise "grid search" ile belirlenen sabitlerdir. Bu yapı sayesinde, modelin boyutu artırılırken parametre/fayda oranı dengede tutulur.

EfficientNet mimarisi, temel yapı taşı olarak MobileNetV2'den türetilen **MBConv bloklarını** kullanır. Bu bloklar aşağıdaki bileşenleri içerir:

- **Ters Artık Bağlantı (Inverted residual connection)**: Dar $\rightarrow$ geniş $\rightarrow$ dar kanal geçişi.
- **Derinlik Ayrık Evrişim (Depthwise separable convolution)**: Hesaplama yükünü azaltır.
- **Sıkıştırma ve Uyarma (Squeeze-and-Excitation) mekanizması**: Kanal

bazlı dikkat (attention) uygular.

- **Swish aktivasyon fonksiyonu**: ReLU yerine daha akışkan "gradyan" geçişi sağlar.

Bu yapı, özellikle sınırlı donanımda çalışan görsel sınıflandırma sistemleri için yüksek performans sunar. EfficientNet mimarisi, başlangıçta B0 modeli olarak tanımlanmış ve ardından B1–B7 varyantlarıyla farklı kaynak gereksinimlerine göre genişletilmiştir.

Bu çalışmada, veri miktarı ve hız gereksinimleri göz önüne alınarak EfficientNet-B0 modeli tercih edilmiştir. Statik yumruk sınıflandırma görevinde %96 doğruluk oranı ile en başarılı mimari olmuş; bazı sınıflarda %100 başarı sağlamıştır. Bu başarı, hem modelin parametrik verimliliği hem de detay yakalama becerisiyle açıklanabilir.

3.4.4 GCN (Graph Convolutional Networks)

GCN, çizge verileri üzerinde çalışan ve sınıflandırma, bağlantı tahmini gibi görevlerde kullanılan, düğümler arası ilişkileri öğrenerek çizge yapılar üzerinde evrişimsel işlem gerçekleştiren bir mimaridir [13]. Her bir düğümün özelliklerini ve komşularının bilgilerini birleştirerek, düğüm temsil bilgilerini günceller. Bu güncellemeler çok katmanlı olarak uygulanarak, düğümleri sınıflandırma veya bağıntılar üzerinde tahmin yapma gibi görevlerde kullanılabilir. Sosyal ağlar, biyolojik bağlantılar veya insan iskelet yapılarında çalışma yapmak için tercih edilebilmektedir.

Bir çizge $G = (V, E)$, N düğüm ve $A \in R^{N \times N}$ bitişiklik matrisi ile tanımlanır. Düğümlerin özellikleri $X \in R^{N \times F}$ matrisinde tutulur. Temel GCN katmanı aşağıdaki şekilde formüle edilir:

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right) \tag{3.4}$$

Burada:

- $\tilde{A} = A + I$: Bitişiklik matrisine birim matris eklenerek kendine bağ (self-loop) sağlanır.
- $\tilde{D}$: $\tilde{A}$'nın derece matrisidir.
- $H^{(l)}$: l. katmandaki düğüm temsil matrisidir.

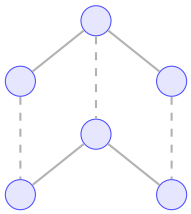
- $W^{(l)}$: Öğrenilen ağırlık matrisidir.
- σ: Aktivasyon fonksiyonu (genellikle ReLU).

Bu yapı, normalizasyon sayesinde öğrenmeyi kararlı hale getirir ve mesaj yayılımı (message passing) mekanizması ile düğüm ve komşularının öznitelik bilgilerini harmanlar.

Bu çalışmada, GCN doğrudan kullanılmamış ancak ST-GCN ve ST-Transformer modelleri için temel teorik altyapı olarak değerlendirilmiştir. Zira bu modeller, GCN yapılarını hem uzamsal hem de zamansal boyutta genişletmektedir.

3.4.5 ST-GCN (Spatial-Temporal Graph Convolutional Networks)

ST-GCN, özellikle insan hareketi tanıma gibi zaman ve mekân boyutlarını içeren çizge verileri üzerinde çalışan bir modeldir. Bu model, hem mekânsal (spatial) hem de zamansal (temporal) çizge evrişimlerini birleştirerek, zaman içinde değişen çizge yapılarındaki desenleri öğrenir. Örneğin, insan iskelet hareketlerini temsil eden çizge yapılarında, ST-GCN hem eklemler arasındaki mekânsal ilişkileri hem de bu ilişkilerin zaman içindeki evrimini modelleyerek, hareket tanıma görevlerinde yüksek performans gösterir [14].



Şekil 3.4 ST-GCN'de zamansal iskelet çizge yapısı

ST-GCN, her zaman adımında uzamsal çizge evrişimi ($\mathcal{G}_s$) uygulayarak eklemler arasındaki ilişkileri öğrenir. Ardından zaman boyutunda tek boyutlu evrişim ($\mathcal{G}_t$) işlemleriyle, hareketin zamansal değişimini modelleyerek daha zengin bir temsil üretir.

Bir uzamsal evrişim adımı, aşağıdaki şekilde formüle edilir:

$$f_{\text{out}}(v_i) = \sum_{v_j \in \mathcal{B}(v_i)} \frac{1}{Z_{ij}} W(l(v_j)) f_{\text{in}}(v_j) \tag{3.5}$$

Burada:

- $\mathcal{B}(v_i)$: v_i'nin komşu düğümleri
- Z_{ij}: Normalizasyon faktörü
- $l(v_j)$: Komşu düğümün topolojik etiket fonksiyonu (ör. mesafeye göre)
- W: Öğrenilen ağırlıklar

Zamansal boyutta, aynı eklem noktasının ardışık karelerdeki değerleri arasında tek boyutlu evrişim uygulanır:

$$f_t^{(i)} = \sigma(\text{Conv1D}(f_{t-k:t+k}^{(i)})) \tag{3.6}$$

Bu yapının en büyük avantajı, hem lokal (mekânsal) hem de zamansal yapıları birlikte modelleyerek karmaşık hareket örüntülerini etkili biçimde yakalayabilmesidir. Ayrıca, uzamsal kenarların topolojik etiketleme yöntemi sayesinde, eklem önemi model tarafından doğrudan öğrenilebilmektedir.

Bu çalışmada ST-GCN, EfficientNet ile birlikte hibrit modelin ikinci bileşeni olarak kullanılmış ve zaman içindeki hareket desenlerini öğrenerek %87 sınıflandırma başarısı elde etmiştir.

3.4.6 ST-Transformer (Spatio-Temporal Transformer)

ST-Transformer, klasik çizge tabanlı yöntemlerin ötesine geçerek, uzamsal ve zamansal bilgileri çok başlı dikkat (multi-head attention) mekanizması ile işleyen bir mimaridir. Bu yapı, özellikle ardışık hareketlerin bağlamını anlamada ve uzun menzilli bağımlılıkları modellemede ST-GCN'e göre daha güçlü bir kapasite sunar. Transformer tabanlı yaklaşımlar, dikkat ağırlıkları üzerinden bilgi akışı sağlayarak eklem noktalarının zaman içindeki önem derecesini öğrenir. Bu sayede, bir sporcunun hareket sekansları arasındaki bağlam ilişkisi daha doğru şekilde analiz edilebilir [15].



Şekil 3.5 ST-Transformer işlem akışı

Zamansal dikkat mekanizması aşağıdaki gibi tanımlanabilir:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3.7}$$

Burada:

- Q, K, V: Sırasıyla "Query", "Key" ve "Value" matrisleridir; her biri giriş dizisinden türetilir.
- d_k: "Key" vektörlerinin boyutudur (ölçekleme için).

ST-Transformer, giriş olarak iskelet verilerini alır: $X \in R^{T \times J \times C}$, burada T zaman, J eklem sayısı ve C her eklemin öznitelik boyutudur (örneğin konum, hız). Bu yapı genellikle 2 boyutlu matris şeklinde düzleştirilerek Transformer katmanlarına verilir.

Bu çalışmada ST-Transformer, EfficientNet ile birlikte hibrit model yapısında alternatif ikinci bileşen olarak değerlendirilmiş ve %85 sınıflandırma başarısı elde etmiştir.

3.4.7 Uzun Kısa Süreli Bellek (LSTM) ile Hareket Tahmini

Modelin son aşamasında, daha önce sınıflandırılmış olan hareket sekansları bir zaman serisi problemi olarak ele alınmış ve LSTM yaklaşımı ile ardışık hareket dizileri üzerinden bir sonraki hareketin tahmin edilmesi hedeflenmiştir. LSTM, RNN'lerin uzun vadeli bağımlılıkları öğrenmekte zorlanmasına çözüm getirme amaçlı kullanılan özel bir yapıdır [16]. Hücre durumu ve geçit mekanizmaları ile geçmiş bilgiyi uzun süre tutabilmekte ve zaman dizisi içindeki örüntüleri daha verimli şekilde öğrenebilmektedir.



Şekil 3.6 LSTM ile ardışık hareket tahmini

Bir LSTM hücresi, üç temel kapı mekanizması içerir:

- **Unutma Kapısı (Forget Gate)**: Önceki hücre bilgisinin ne kadarının tutulacağını belirler.
- **Giriş Kapısı (Input Gate)**: Yeni gelen bilginin ne kadarının hücre durumuna ekleneceğine karar verir.
- **Çıkış Kapısı (Output Gate)**: Hücre durumundan çıkışa ne kadar bilginin aktarılacağını kontrol eder.

Bu mekanizmalar aşağıdaki denklemlerle temsil edilir:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{3.8}$$
$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{3.9}$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{3.10}$$
$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \tag{3.11}$$
$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{3.12}$$
$$h_t = o_t \odot \tanh(C_t) \tag{3.13}$$

Burada:

- x_t: Mevcut giriş vektörü
- h_{t-1}: Önceki zaman adımından gelen gizli durum
- C_t: Hücre durumu
- σ: Sigmoid aktivasyon fonksiyonu
- $\odot$: Eleman bazında çarpma işlemi

Bu çalışmada, EfficientNet + ST-GCN/ST-Transformer hibrit modelinin çıktıları, her karedeki hareket sınıfları olarak LSTM katmanına giriş olarak verilir. Amaç, 5 ardışık kareye (t-4 ila t) bakarak, t+1 zaman adımındaki hareketin (yani bir sonraki vuruş türünün) ne olacağını tahmin etmektir.

Modelin çıktısı:

$$\hat{y}_{t+1} = \text{LSTM}(x_{t-4:t}) \rightarrow \text{FC} \rightarrow \text{Softmax} \tag{3.14}$$

LSTM mimarisi, hareket sekanslarını anlamada önemli başarı sağlamış ve bu çalışmada yaklaşık %75 başarı oranı elde edilmiştir. Bu başarı oranı, her sporcunun vuruş karakteristiklerine ve video dizilerinin uzunluğuna bağlı olarak farklılık göstermektedir.

LSTM yapısının kullanımı, sistemi yalnızca hareket tanıyıcı değil, aynı zamanda davranış tahmin edici ve potansiyel olarak gerçek zamanlı uyarıcı bir yapıya

dönüştürmektedir. Bu, sanal antrenör ve benzeri sistemlerin uygulanabilmesi açısından fayda sağlamaktadır.

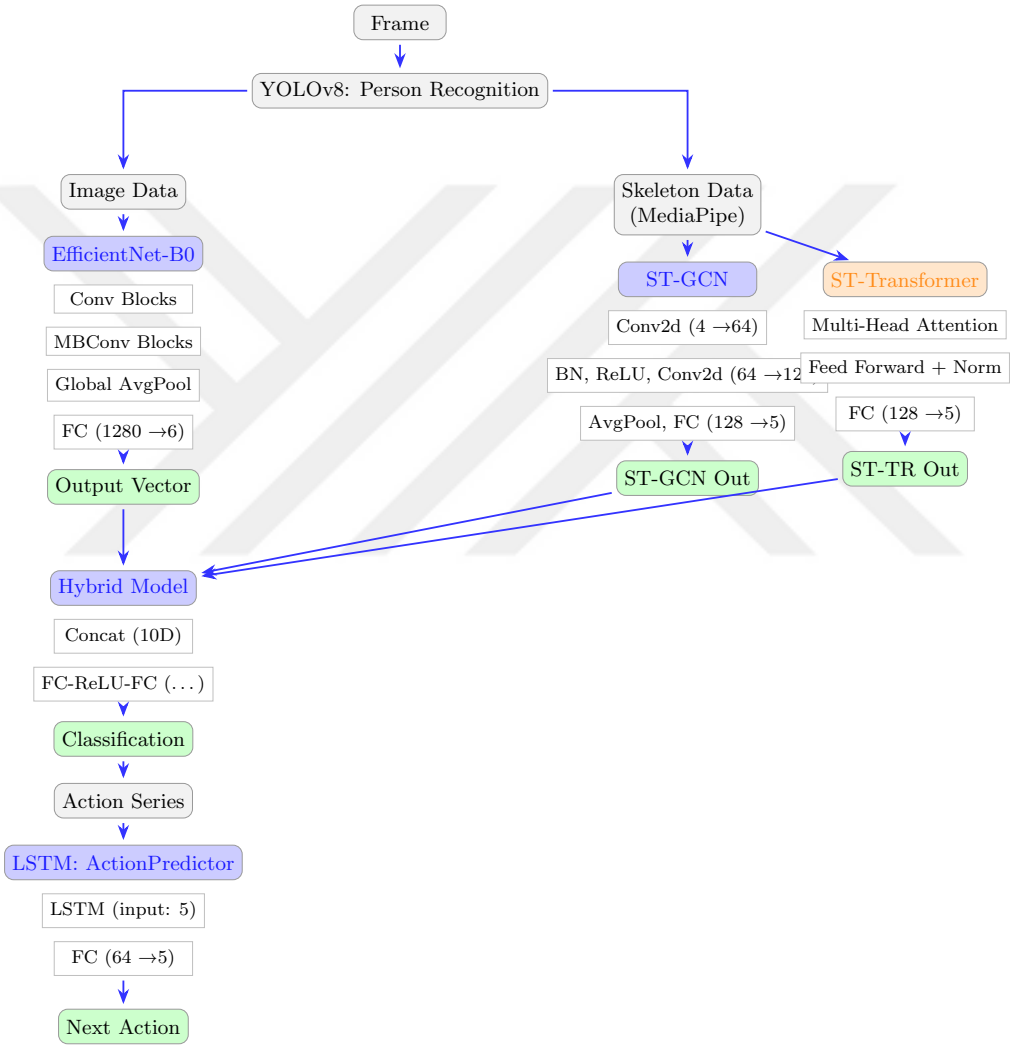
3.5 Yolo ile Konum Tespiti - ROI (Region of Interest)

Eğitim yapılan vuruş tahmin modelinin test edileceği videolar içerisinde, sporcuların görüntüdeki konumlarının tespiti için YOLO (You Only Look Once) modeli kullanılmıştır. YOLO, gerçek zamanlı nesne tespiti için optimize edilmiş bir derin öğrenme modelidir ve her bir video karesi üzerinde hızlı ve doğru tespitler gerçekleştirmektedir. İşlem aşamaları şu şekildedir:

- 1. Video Karelerinin Analizi:
  - Video, karelere ayrılarak her bir kare bağımsız olarak işlenir.
  - Her bir kare, YOLO modeline giriş olarak verilmeden önce, modelin gerektirdiği formatta (örneğin, 416x416 piksel) yeniden boyutlandırılır ve normalize edilir.
- 2. Nesne Tespiti:
  - YOLO modeli, video karelerinde kişilerin bulunduğu alanları belirlemek için uygulanır.
  - Model, her bir nesne için tahmini sınırlayıcı kutular (bounding boxes) ve sınıf etiketleri üretir.
- 3. Konumların Koordinatlarının Belirlenmesi: YOLO çıktısı:
  - x, y: Sınır kutusunun merkez koordinatları.
  - w, h: Genişlik ve yükseklik.
  - Güven skoru ve sınıf olasılığı.
- 4. Koordinatların Hesaplanması:
  - Sol üst köşe: $(x - w/2, y - h/2)$
  - Sağ alt köşe: $(x + w/2, y + h/2)$
- 5. Görselleştirme:
  - ROI alanları kare üzerinde gösterilir.
- 6. Temas Analizi:
  - IoU (Intersection over Union) $= \frac{A_{ortak}}{A_{birleim}}$
  - IoU > 0.5 durumunda temas var kabul edilir.

3.6 Hibrit Model Yapısı

Yapılan test ve karşılaştırmalar sonucu, başarı ve işlem yükü avantajları dolayısı ile tercih edilen EfficientNet modeli ile birlikte seri hareket analizleri için ST-GCN veya ST-Transformer, gelecek adım tahmini için ise LSTM (Long Short-Term Memory) modellerini de içeren hibrit bir model geliştirilmiştir. Hibrit model; görüntü tabanlı özellik çıkarımı, iskelet verileri üzerinden uzamsal ve zamansal özelliklerin elde edilmesi ve hareket dizilerinin zaman serisi analizi olmak üzere üç temel bileşeni içermektedir.



Şekil 3.7 Hibrit model genel mimarisi: EfficientNet + ST-GCN/Transformer + LSTM

3.6.1 EfficientNet Bileşeni

EfficientNet-B0 modeli, düşük hesaplama yüküyle etkin bir görüntü tabanlı özellik çıkarımı sunar. Giriş görüntüleri (224x224x3 boyutunda) ImageNet ortalama ve standart sapması kullanılarak normalize edilmiştir. EfficientNet-B0, konvolüsyon

ve MBConv bloklarından oluşur. Modelin çıktısı, son katmandan elde edilen 1280 boyutlu bir özellik vektörüdür. Son tam bağlantılı katman, 6 sınıflı ("bag", "cross", "hook", "jab", "no punch", "uppercut") bir olasılık dağılımı üretmiş ve bu dağılımdan kullanılmayacak olan sınıf ("bag") çıkarılarak 5 sınıfa indirgenmiştir ve hibrit modele girdi olarak sağlanmıştır.

3.6.2 ST-GCN ve ST-Transformer Bileşeni

ST-GCN, skeleton verisindeki eklem noktaları arasındaki uzamsal ilişkileri çizge tabanlı bir konvolüsyon ağıyla modellemektedir. ST-Transformer ise Transformer mimarisini kullanarak eklem noktaları ve zaman adımları arasındaki uzamsal-zamansal ilişkileri öğrenmektedir. Skeleton verisi, MediaPipe Pose ile 33 eklem noktası olarak çıkarılmış ve (T, 33, 3) boyutunda kaydedilmiştir. Veri ön işleme sırasında, her bir skeleton dizisi 6 frame'e sabitlenmiş ve eklemler arasındaki açılar hesaplanarak veri (T, 33, 4) boyutuna genişletilmiştir. Hibrit model, skeleton verisi için ST-GCN veya ST-Transformer, görüntü verisi için ise EfficientNet kullanılarak oluşturulmuştur. Eğitim iki aşamada gerçekleştirilmiştir: İlk aşamada yalnızca son katmanlar, ikinci aşamada ise tüm katmanlar eğitilmiştir.

3.6.3 Hibrit Modelin Birleştirme Süreci

EfficientNet ve ST-GCN/ ST-Transformer modellerinin ürettiği iki ayrı olasılık vektörü (her biri 5 boyutlu) birleştirilmektedir. Birleştirilen bu 10 boyutlu vektör, bir ara tam bağlantılı katmana gönderilmekte ve ReLU aktivasyonu ile işlenmektedir. İki dropout katmanı (%20 oranında) kullanılarak aşırı öğrenmenin önüne geçilmekte ve son olarak, 5 sınıflı bir olasılık vektörü oluşturmak için başka bir tam bağlantılı katman kullanılmaktadır.

3.6.4 LSTM ile Sonraki Hareket Tahmini ve Uyarlanabilir Öğrenme

Hibrit modelden elde edilen hareket tahminleri diziler halinde, sonraki hareketin tahmin edilmesi amacıyla LSTM modülüne aktarılır. Son 5 hareket girdi olarak alınır ve iki katmanlı ve 64 gizli birim içeren LSTM ile gizli durum ve hücre durumu kullanılarak sonraki hareket tahmin edilir. Yeni hareketlerin sürekli olarak modele eklenmesi ve çapraz entropi kaybının minimize edilmesi ile modelin sürekli bir öğrenme halinde olması sağlanmaktadır.

4
UYGULAMA

4.1 Veri Ön İşleme

Müsabaka sırasında vuruşların hangi kategoriye ait olduğunu sınıflandırmak için farklı yapıda bir kaç model eğitimi gerçekleştirilmiştir. Bunlar içinde ResNet, MobileNet yapıları ile sınıflandırma yapılmış ve öncelikle sabit resimler üzerinde test edilmiştir.

Kullanılan veri setindeki görüntüler öncelikle kırpma, boyutlandırma ve filtreleme gibi ön işleme adımlarından geçirilerek eğitim ve test verisi olarak düzenlenmiştir.



Şekil 4.1 Ön İşlem Yapılmış Örnek Görüntüler [9]

4.2 Model Eğitimi

ResNet18, EfficientNet ve MobileNet kullanılarak Nvidia T4 GPU üzerinde Pytorch kullanılarak 100 epoch, 0.001 öğrenme katsayısı ile "Adam" optimizasyon fonksiyonu kullanılarak görüntü sınıflandırma modelleri eğitilmiş ve sonuçları karşılaştırılmıştır.

Eğitim ve test veri setleri eldeki veriler üzerinden 0.8 eğitim, 0.2 test olarak kullanılmıştır. Çeşitli metrikler kullanılarak (Accuracy, recall, precision, f1-score) başarı değerleri ölçülmüştür.

Daha sonra bu eğitilen modeller video üzerinde dinamik bir ortamda kullanılarak vuruşların tanınması amacıyla entegre edilmiştir. Canlı videoda ise doğrulama

ve başarı metriği ölçümü, alınan kareler üzerinden etiketleme ve karşılaştırma yapılarak ölçülmüştür.

4.2.1 Başarı Metrikleri

Hareketlerin sınıflandırılması ve sporcuların vuruşlarının analiz edilmesi amacıyla dört farklı derin öğrenme modeli (ResNet, MobileNet ve EfficientNet) test edilmiş ve performansları karşılaştırılmıştır. Modellerin doğruluk, kayıp (loss) ve sınıf bazlı metrikler açısından elde ettiği sonuçlar detaylı olarak incelenmiştir.

4.2.2 Kayıp (Loss) ve Doğruluk (Accuracy) Fonksiyonları

Sınıflandırma problemlerinde modelin çıktıları ile gerçek etiketler arasındaki farkı ölçmek için kullanılan çapraz entropi kayıp fonksiyonu (cross-entropy loss) şu şekilde ifade edilir:

$$\mathcal{L}_{CE} = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \tag{4.1}$$

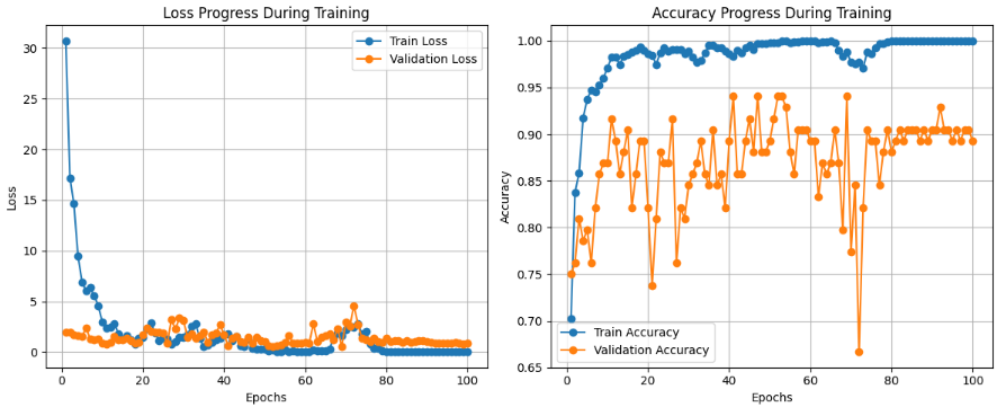
Burada, C sınıf sayısı, y_i gerçek etiket değeri ve $\hat{y}_i$ model tarafından tahmin edilen olasılıktır.

$$\text{Accuracy} = \frac{\sum_{i=1}^{N} \mathbf{1}(\hat{y}_i = y_i)}{N} \times 100\% \tag{4.2}$$

4.2.3 ResNet

ResNet modeli, karmaşıklık matrisi analizine göre %89 genel doğruluk sağlamış ve özellikle "Bag" ve "No Punch" gibi sınıflarda yüksek bir başarı elde etmiştir. Bununla birlikte, "Jab" ve "Cross" gibi benzer özelliklere sahip sınıflarda karışıklıklar yaşanmış ve modelin sınıflar arasındaki farklılıkları tam olarak öğrenemediği gözlemlenmiştir. Eğitim ve doğrulama süreçleri incelendiğinde, modelin doğrulama doğruluğunda dalgalanmalar yaşadığı ancak genel olarak dengeli bir performans gösterdiği görülmüştür

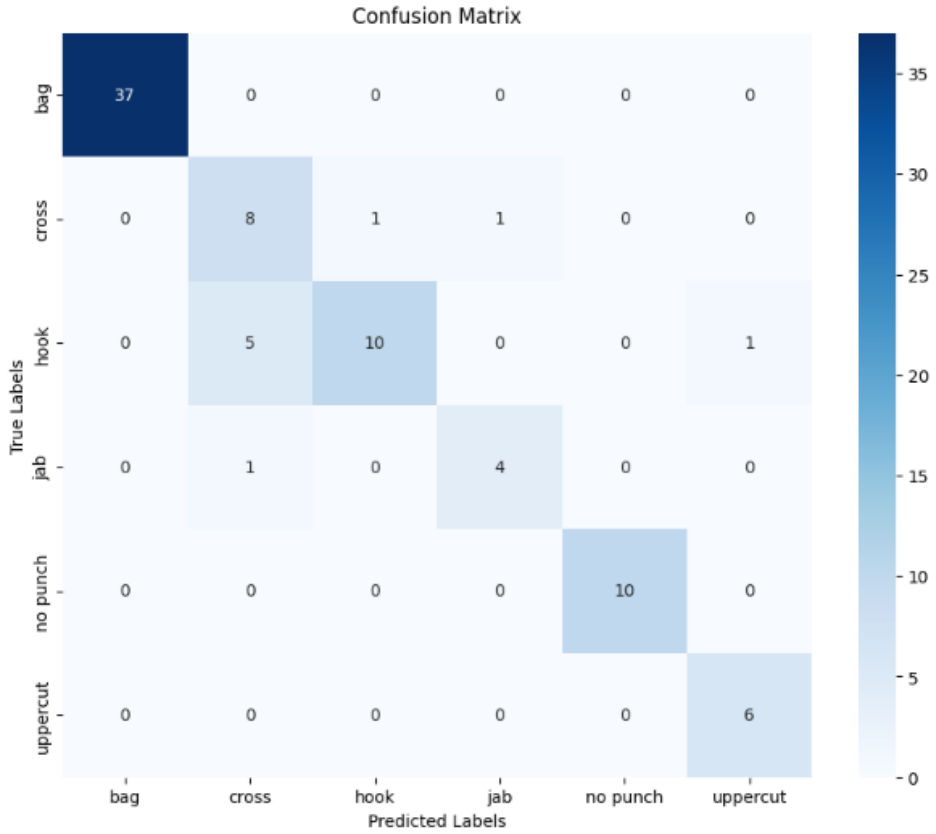
4.2.3.1 Eğitim



Şekil 4.2 ResNet Eğitimi

4.2.3.2 Tahmin Sonuçları

	precision	recall	f1-score	support
bag	1.00	1.00	1.00	37
cross	0.57	0.80	0.67	10
hook	0.91	0.62	0.74	16
jab	0.80	0.80	0.80	5
no punch	1.00	1.00	1.00	10
uppercut	0.86	1.00	0.92	6
accuracy			0.89	84
macro avg	0.86	0.87	0.86	84
weighted avg	0.91	0.89	0.89	84

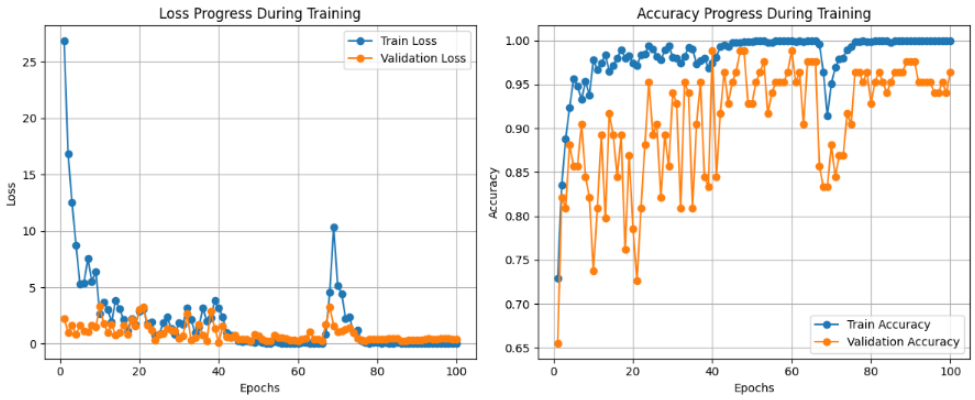


Şekil 4.3 ResNet Tahmin Sonuçları

4.2.4 MobileNet

MobileNet, düşük maliyetli ve hızlı bir yapı sunmasının yanında oldukça da güçlü bir performans göstermiştir. %96 genel doğruluk ile dikkat çeken MobileNet, "Bag," "Cross" ve "Uppercut" gibi sınıflarda mükemmel sonuçlar elde etmiştir. "Hook" ve "Jab" sınıflarında sınırlı karışıklıklar gözlemlenmiştir. Eğitim ve doğrulama doğruluğu arasındaki farkın az olması, modelin genelleme yeteneğinin iyi olduğunu ve aşırı öğrenme eğiliminin minimal seviyede olduğunu göstermiştir.

4.2.4.1 Eğitim

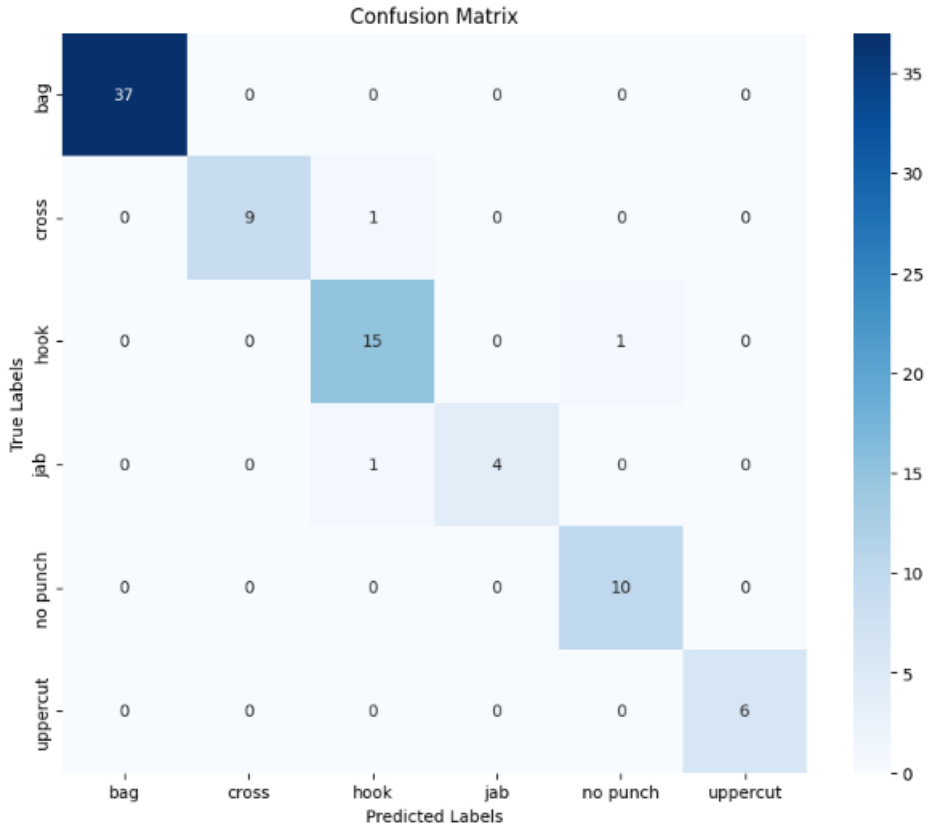


Şekil 4.4 MobileNet Eğitimi

4.2.4.2 Tahmin Sonuçları

Epochs

	precision	recall	f1-score	support
bag	1.00	1.00	1.00	37
cross	1.00	0.90	0.95	10
hook	0.88	0.94	0.91	16
jab	1.00	0.80	0.89	5
no punch	0.91	1.00	0.95	10
uppercut	1.00	1.00	1.00	6
accuracy			0.96	84
macro avg	0.97	0.94	0.95	84
weighted avg	0.97	0.96	0.96	84

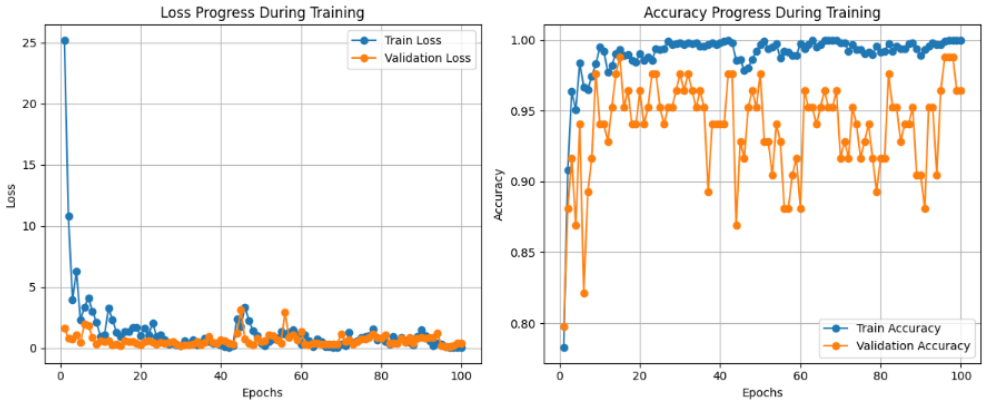


Şekil 4.5 MobileNet Tahmin Sonuçları

4.2.5 EfficientNet

EfficientNet modeli, hem doğruluk hem de sınıf bazlı metriklerde en iyi sonuçları elde ederek bu çalışmadaki en başarılı model olarak öne çıkmıştır. %96 genel doğruluk sağlayan EfficientNet, "Bag," "Cross" ve "Uppercut" sınıflarında %100 başarı elde etmiştir. "Hook" ve "Jab" sınıflarında küçük çaplı hatalar yapılmıştır. Eğitim ve doğrulama süreçlerinde kayıpların hızlı bir şekilde azaldığı ve doğrulamanın genelde dengeli bir şekilde ilerlediği görülmüştür. EfficientNet'in yüksek doğruluğu ve dengeli öğrenme süreci, bu çalışmada en iyi performans gösteren model olmasını sağlamıştır.

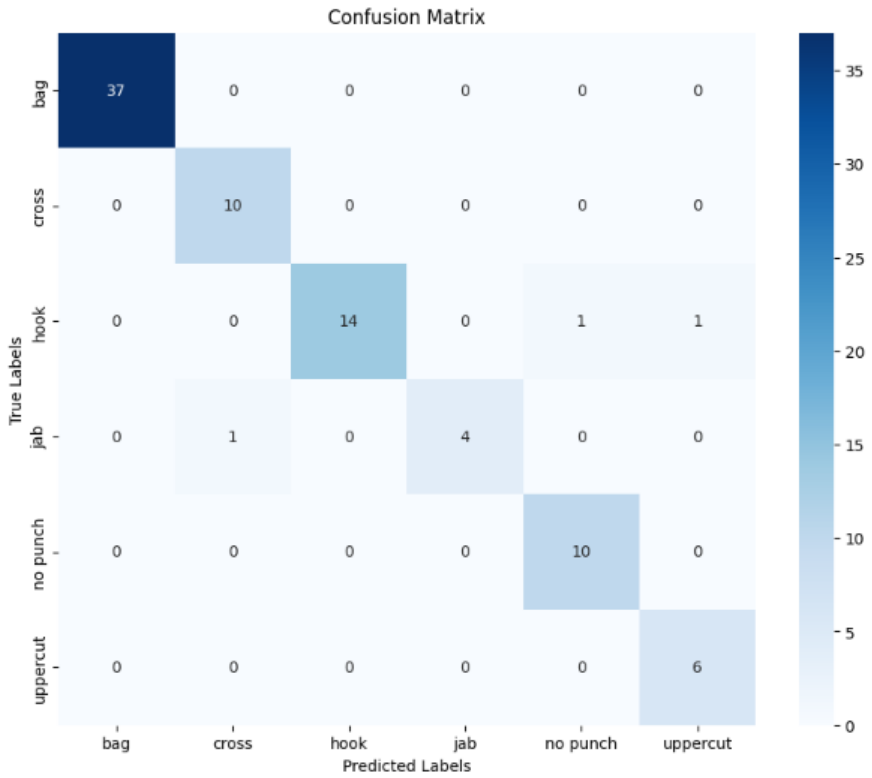
4.2.5.1 Eğitim



Şekil 4.6 EfficientNet Eğitimi

4.2.5.2 Tahmin Sonuçları

	precision	recall	f1-score	support
bag	1.00	1.00	1.00	37
cross	0.91	1.00	0.95	10
hook	1.00	0.88	0.93	16
jab	1.00	0.80	0.89	5
no punch	0.91	1.00	0.95	10
uppercut	0.86	1.00	0.92	6
accuracy			0.96	84
macro avg	0.95	0.95	0.94	84
weighted avg	0.97	0.96	0.96	84



Şekil 4.7 EfficientNet Tahmin Sonuçları

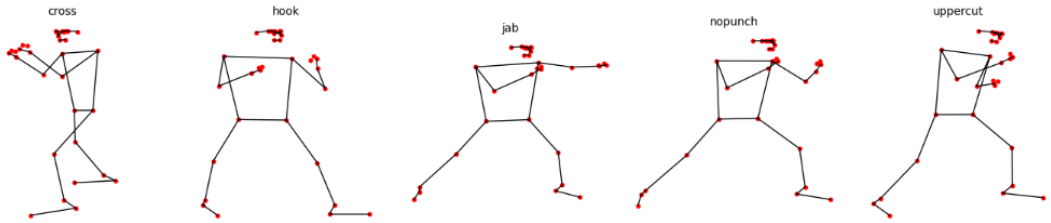
Tablo 4.1 CNN Tabanlı Modellerin Sınıflandırma Performans Karşılaştırması

Model	Macro Precision (%)	Macro Recall (%)
ResNet18	86	87
MobileNet	97	94
EfficientNet	95	95
Roboflow Pretrained	90.4	87.6

Bu çalışmada kullanılan üç farklı CNN tabanlı modelin sonuçları, dış kaynaklı Roboflow platformunda eğitilmiş bir model ile karşılaştırılmıştır [9]. Roboflow modeli, %97.6 oranındaki genel başarı (mAP@50) değeriyle yüksek performans sergilese de, sınıflar arası dengeyi ölçen ortalama kesinlik (macro precision) ve duyarlılık (recall) metriklerinde, EfficientNet ve MobileNet modelleri daha yüksek sonuçlar elde etmiştir.

4.3 ST-GCN ve ST-Transformer ile Görüntü Serisi İşleme

İlk aşamada kullanılan model yapılarında ve veri setinde, herhangi bir zaman serisi olmaksızın tek kare için tek etiket yöntemi kullanılarak sınflandırma işlemi yapılmış ve alınan video karelerinden vuruş esnasındaki tek bir görüntü üzerinden tahmin yapılmıştır. Fotoğraf ya da video karesi bazlı düşündüğümüzde bu işlem gayet başarılı sonuç verse de canlı bir akış esnasında analiz yaparken seri halinde birkaç görüntü almak ve bu şekilde bir sınıflandırma yapmak daha doğru sonuç verebilir. Bu sebeple bu aşamaya kadar kullandığımız modellere ek olarak çizge tabanlı bir sistem yaklaşımı olan ST-GCN ve ST-Transformer yapıları da entegre edilmiştir. İki farklı hibrit yaklaşım ile yalnızca görüntü sınıflandırma işleminin ötesinde seri halinde bir analiz de gerçekleştirilmiştir.



Şekil 4.8 Mediapipe ile Çıkartılmış İskelet Verisi Örnekleri

ST-GCN, iskelet verisindeki eklem noktaları arasındaki uzamsal ilişkileri çizge tabanlı bir evrişim ağıyla modellemektedir. ST-Transformer ise Transformer mimarisini kullanarak eklem noktaları ve zaman adımları arasındaki

uzamsal-zamansal ilişkileri öğrenmektedir. İskelet verisi, MediaPipe Pose ile 33 eklem noktası olarak çıkarılmış ve (T, 33, 3) boyutunda kaydedilmiştir. Veri ön işleme sırasında, her bir iskelet dizisi 6 kareye sabitlenmiş ve eklemler arasındaki açılar hesaplanarak veri (T, 33, 4) boyutuna genişletilmiştir. Hibrit model, iskelet verisi için ST-GCN veya ST-Transformer, görüntü verisi için ise EfficientNet kullanılarak oluşturulmuştur. Eğitim iki aşamada gerçekleştirilmiştir: İlk aşamada yalnızca son katmanlar, ikinci aşamada ise tüm katmanlar eğitilmiştir.

Çalışmanın devamında seri halinde bir veri setine ihtiyaç olduğundan vuruş türlerinin kısa videolar halinde etiketlenmiş olduğu ayrı bir veri seti çalışmaya katılarak hem önceki görüntü sınıflandırma modelinden gelen sonuçları hem de seri analizini kapsayacak şekilde hibrit bir model geliştirilmiştir. Bu sayede bir anlık görüntüyü değil bir kaç kareyi kapsayan, tam bir hareket sınıflandırma işlemi gerçekleştirilmiştir.

4.3.1 ST-GCN ile Görüntü Serisi İşleme

ST-GCN (Spatio-Temporal Graph Convolutional Networks), zaman içinde değişen insan iskelet hareketlerini analiz etmek için çizge temelli bir yaklaşımdır. Bu model, bir dizi video karesinden çıkarılmış iskelet verilerini hem uzamsal hem de zamansal bağlamda işleyerek sporcuların yaptığı hareketlerin türünü belirlemeye çalışır.

Her bir video, OpenPose ile iskelet verisine dönüştürülmüş ve bu veriler ST-GCN modeline girdi olarak verilmiştir. Model, uzuvlar arası bağlantıları temsil eden kenarları ve zaman boyunca düğümleri takip ederek, boksörlerin hareket örüntülerini analiz etmiş ve sınıflandırma işlemi gerçekleştirmiştir.

4.3.2 Çizge Evrişim Ağı (GCN)

ST-GCN modelinde, çizge verileri üzerinde özelliklerin çıkarılması için kullanılan temel işlemin formülü şu şekildedir:

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right) \tag{4.3}$$

Burada:

- $\tilde{A} = A + I$; A çizge komşuluk matrisi, I ise birim matrisi,
- $\tilde{D}$, $\tilde{A}$'nın derece matrisidir,
- $H^{(l)}$, l'inci katmandaki düğüm özelliklerini,

- $W^{(l)}$, öğrenilebilir ağırlık matrisini,
- σ, aktivasyon fonksiyonunu ifade eder.

Bu yapı, insan iskeletinin konum bilgilerini kullanarak düğüm ve kenar haritaları oluşturur; ST-GCN ise zaman serisi bilgisini de entegre ederek birkaç kareyi birlikte analiz eder.

4.3.3 ST-Transformer ile Görüntü Serisi İşleme

ST-GCN modeline alternatif olarak kullanılan bir diğer yapı olan ST-Transformer, zaman serilerindeki uzun menzilli bağımlılıkları daha iyi öğrenebilmek için dikkat (attention) mekanizması kullanmaktadır. Bu modelde, iskelet noktaları arasındaki ilişki doğrudan çizge formunda değil, çok başlı dikkat katmanları üzerinden modellenmiş ve zamansal bilgi daha esnek biçimde yakalanmıştır.

ST-Transformer ile yapılan denemelerde modelin hem doğruluk hem de F1-Score açısından ST-GCN'ye çok yakın bir başarıda olduğu gözlemlenmiştir.

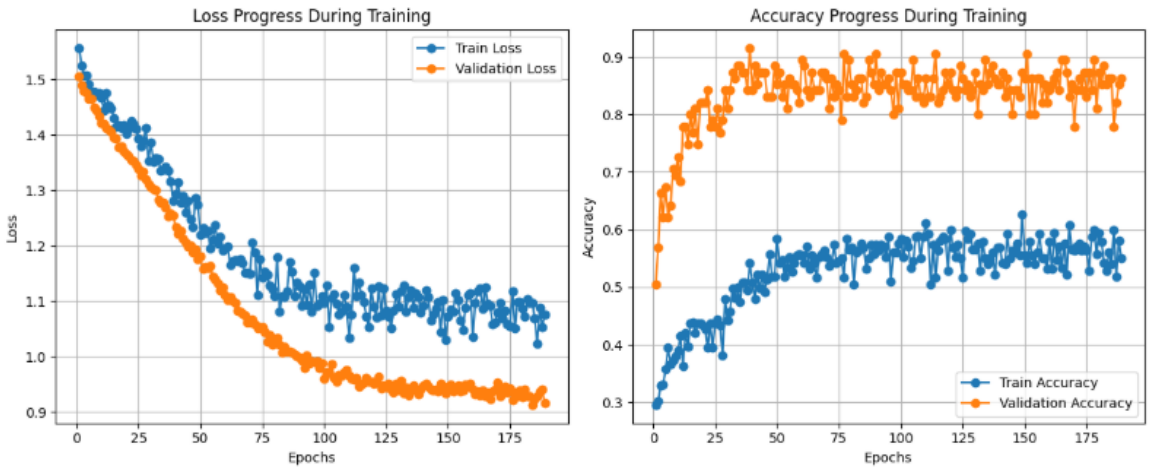
4.3.4 LSTM

Sistemin nihai aşamasında, hibrit modellerin (EfficientNet + ST-GCN / ST-Transformer) çıkardığı hareket sınıflandırma sonuçları bir zaman dizisi olarak ele alınmış ve bu dizilerden bir sonraki hareketin tahmini amaçlanmıştır. Bu bağlamda LSTM (Long Short-Term Memory) mimarisi kullanılarak geçmişte yapılan hareketlere göre bir sonraki olası hareket sınıfı belirlenmiştir.

Model, ardışık olarak alınan 5 hareket dizisini girdi olarak almış ve 6. adımı (yani bir sonraki hamleyi) tahmin edecek şekilde eğitilmiştir.

4.3.5 ST-GCN ile Oluşturulan Hibrit Model Sonuçları

4.3.5.1 Eğitim



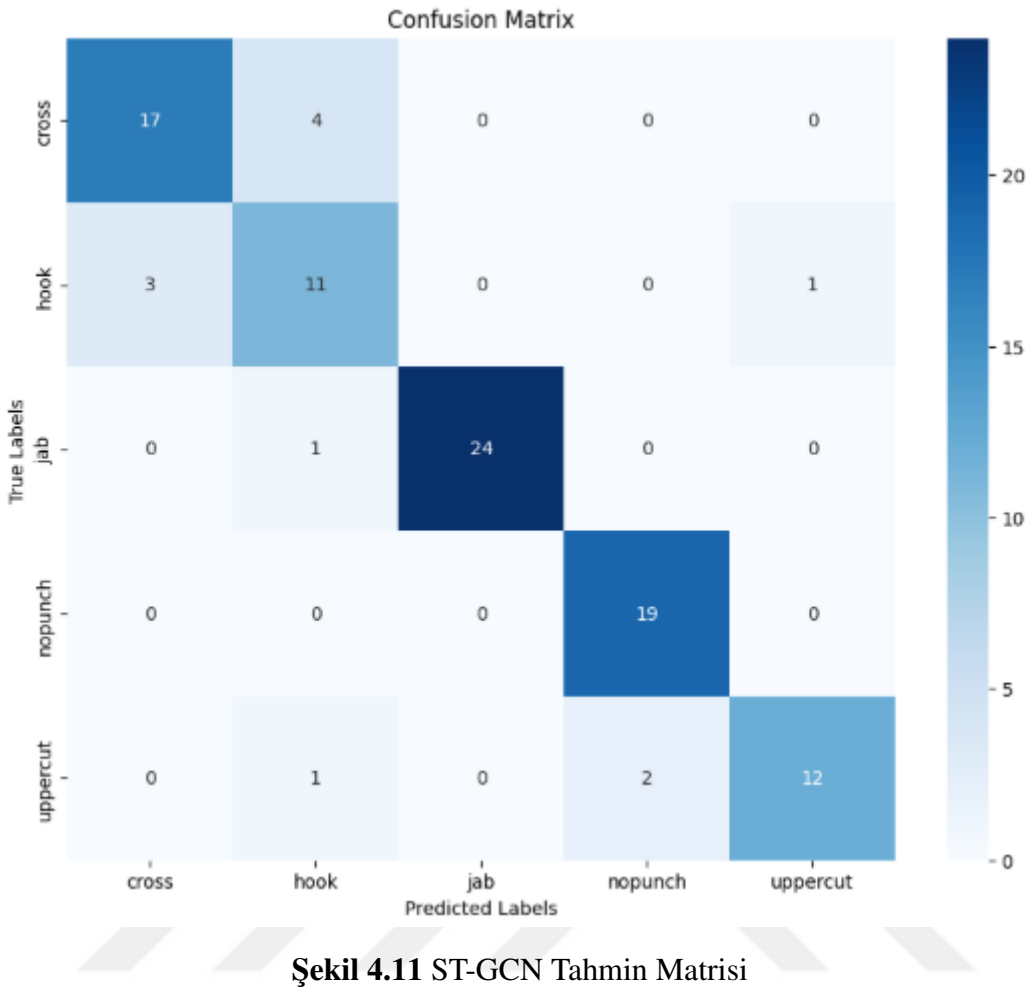
Şekil 4.9 ST-GCN Eğitimi

4.3.5.2 Tahmin Sonuçları

Classification Report:

	precision	recall	f1-score	support
cross	0.85	0.81	0.83	21
hook	0.65	0.73	0.69	15
jab	1.00	0.96	0.98	25
nopunch	0.90	1.00	0.95	19
uppercut	0.92	0.80	0.86	15
accuracy			0.87	95
macro avg	0.86	0.86	0.86	95
weighted avg	0.88	0.87	0.87	95

Şekil 4.10 ST-GCN Başarı Metrikleri



Şekil 4.11 ST-GCN Tahmin Matrisi

4.3.6 ST-Transformer ile Oluşturulan Hibrit Model Sonuçları

4.3.6.1 Eğitim



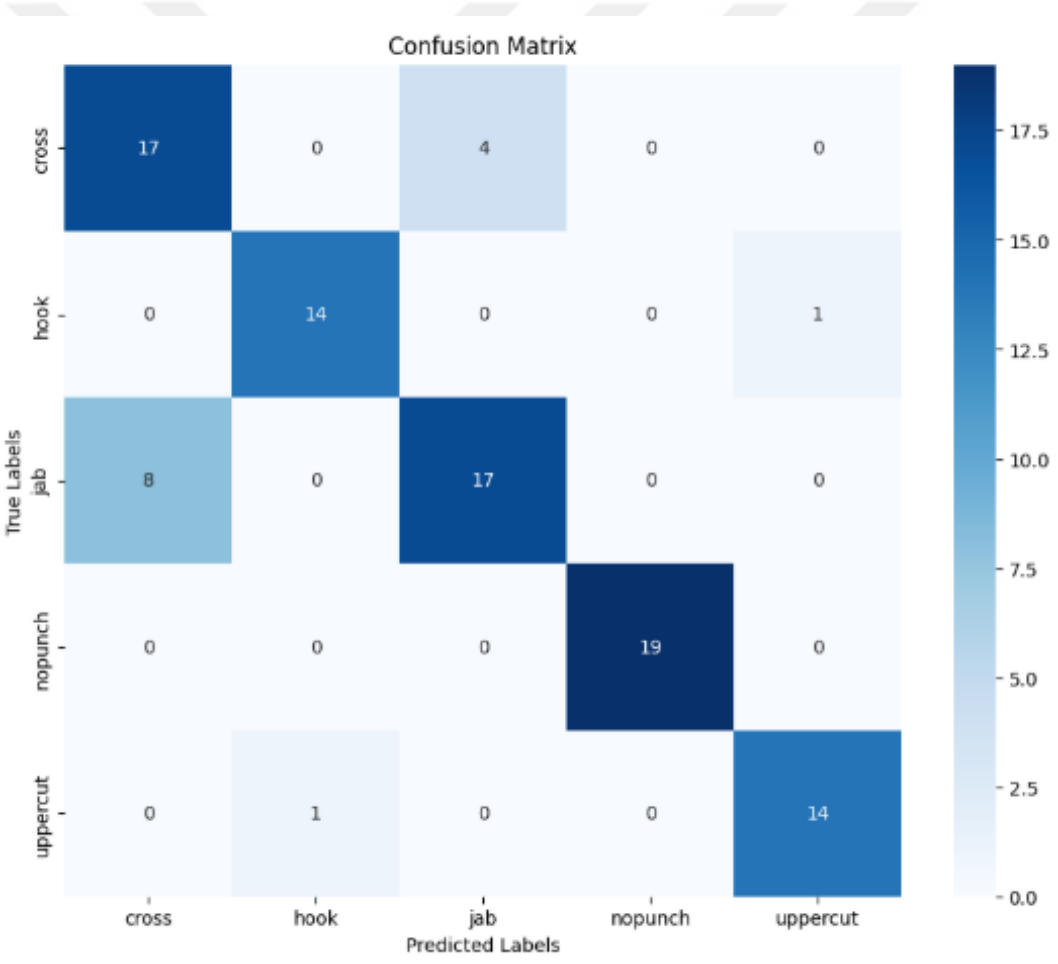
Şekil 4.12 ST-Transformer Eğitimi

4.3.6.2 Tahmin Sonuçları

Classification Report:

	precision	recall	f1-score	support
cross	0.68	0.81	0.74	21
hook	0.93	0.93	0.93	15
jab	0.81	0.68	0.74	25
nopunch	1.00	1.00	1.00	19
uppercut	0.93	0.93	0.93	15
accuracy			0.85	95
macro avg	0.87	0.87	0.87	95
weighted avg	0.86	0.85	0.85	95

Şekil 4.13 ST-Transformer Başarı Metrikleri



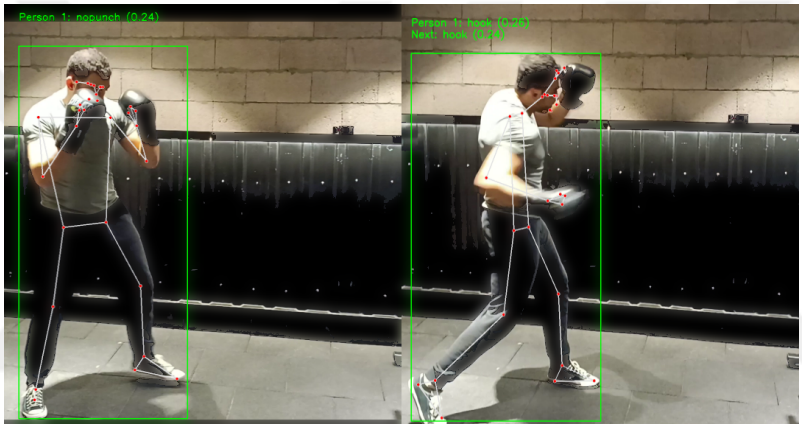
Şekil 4.14 ST-Transformer Tahmin Matrisi

Tablo 4.2 Hibrit Modellerin Performans Karşılaştırması

Model	Sınıflandırma Doğruluğu (%)	Tahmin Başarımı (%)
ST-GCN + EfficientNet	87	75
ST-Transformer + EfficientNet	85	75

4.3.7 Entegrasyon ve Video Analizi

Test amaçlı kullanılan videolarda sporcuların hareketlerini tespit etmek ve vuruşlarını sınıflandırmak için YOLO ve Mediapipe araçları kullanılmıştır. YOLO modeli, video karelerinde sporcuların konumlarını belirlemek için kullanılırken, Mediapipe ise sporcuların vücut iskelet noktalarını çıkararak hareketlerin daha detaylı analiz edilmesini sağlamıştır.



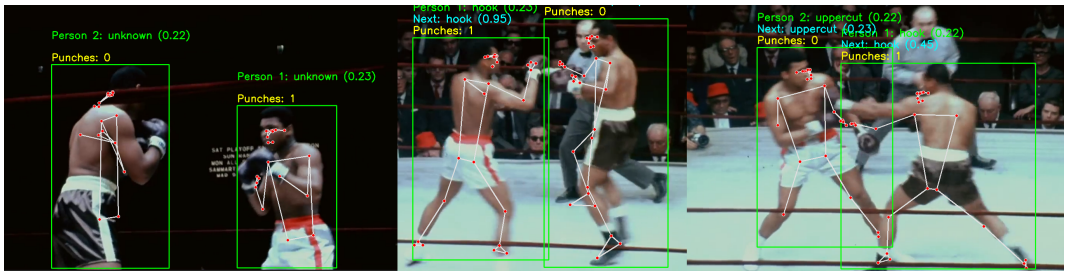
Şekil 4.15 LSTM ile bir sonraki hareketin tahminine yönelik örnek kare dizisi. Model, geçmiş karelerden aldığı hareket bilgisini kullanarak bir sonraki pozisyonu öngörmektedir.

4.3.7.1 Sporcuların Tanımlanması: YOLO ile Bounding Box Tespiti

Video akışı sırasında her bir kare YOLO modeline verilmiş ve model, karedeki iki sporcuyu sınırlayıcı kutular (bounding box) kullanarak tanımlamıştır. YOLO, her bir sporcuyu farklı ID'lerle etiketlemiş ve sporcuların pozisyonlarını kare bazında takip etmiştir. Bounding box verileri, her sporcunun x-y koordinatlarını ve boyutlarını belirlemek için kullanılmıştır. Bu işlem, sporcuların birbirinden ayırt edilmesi ve hangi sporcunun hareket yaptığı bilgisinin çıkarılması için temel oluşturmuştur.

4.3.7.2 Hareketlerin Çözümlemesi: Mediapipe ile İskelet Çıkartımı

YOLO tarafından belirlenen bounding box'lar içerisindeki sporcuların hareketlerini detaylandırmak amacıyla Mediapipe kullanılmıştır. Mediapipe, her bir sporcu için



Şekil 4.16 YOLO modeli ile boksörlerin maç anında tespit edilerek etiketlenmesi. Her sporcu farklı ID'lerle izlenmekte ve pozisyonları gerçek zamanlı olarak sınırlayıcı kutular ile işaretlenmektedir. [17]

33 iskelet noktası çıkartmış ve bu noktalar kullanılarak sporcuların vücut duruşları ve hareketleri analiz edilmiştir [18]. Özellikle, el, dirsek ve omuz noktaları, sporcuların vuruş yapıp yapmadığını anlamak için bilgi sağlamıştır.

4.3.7.3 Vuruş ve Temas Tespiti

Vuruş anlarının belirlenmesi için Mediapipe tarafından çıkarılan iskelet noktaları kullanılarak sporcuların kollarının ve ellerinin hareketleri incelenmiştir. Bir sporcu vuruş yaparken elinin konumu, rakibin bounding box'ı ile çakıştığında bu durum bir temas olarak tanımlanmıştır. YOLO ve Mediapipe çıktılarına dayanarak, temasın hangi tür vuruşa (örneğin, jab, hook, uppercut) karşılık geldiği belirlenmiştir. Bu süreçte, temas tespit edilen kareler sınıflandırma modeline verilmiş ve vuruşun türü tespit edilmiştir. YOLO ve Mediapipe verilerinden elde edilen kareler, önceden eğitilmiş derin öğrenme modeli ile analiz edilerek vuruş türleri sınıflandırılmıştır.

5
SONUÇ

Bu çalışmada, dinamik hareketlerin yer aldığı sportif karşılaşmaların veya antrenman süreçlerinin gerçek zamanlı olarak analiz edilmesi, müsabaka değerlendirmelerinin yapılması, kişisel hareket kalıplarının öğrenilmesi ve antrenman süreçlerine katkı sağlanması amacıyla bir sistem geliştirilmiştir. Boks sporu temel alınarak geliştirilen bu sistemde; yapılan model karşılaştırmaları sonucunda en başarılı bulunan yapılar hibritlenmiş, statik görüntü verileri için EfficientNet, iskelet-temporal veriler için ST-GCN ve ardışık hareket tahmini için LSTM entegre edilerek kapsamlı bir yapı oluşturulmuştur. Bu yapı, boks branşında yapılan deneylerde sırasıyla %96 statik görüntü sınıflandırma doğruluğu, %87 hareket tanıma başarısı ve yaklaşık %75 ardıl hareket tahmini doğruluğu elde etmiştir.

Sadece statik ya da iskelet veriler kullanılarak yapılan analizlerde, çevresel değişkenler ve zamansal bağlamın eksikliği nedeniyle genelleme başarısı düşmektedir. Özellikle tek kare üzerinden yapılan değerlendirmeler eğitim verisinde yüksek doğruluk sağlarken, video testlerinde bu başarı ciddi oranda azalmaktadır. Sadece iskelet verisiyle yapılan analizlerde ise ortam değişkenlerine duyarlılık yetersiz kalmaktadır. Bu nedenle, çevresel ve zamansal faktörleri birlikte dikkate alan hibrit bir yaklaşım geliştirilerek, bu sınırlamalar aşılmıştır.

Modelin gerçek dünya koşullarına uyarlanabilirliği göz önünde bulundurularak, maç senaryolarını taklit eden kısa videolarla test edilmiş, kare bazlı yumruk tespiti için YOLO ve MediaPipe entegrasyonu ile optimize edilmiştir. Bu sayede sistem, akademik veri kümelerinin ötesinde genel geçer uygulamalara adapte olabilecek yapıya kavuşturulmuştur.

Öte yandan bazı sınırlılıklar da mevcuttur. Kullanılan veri seti, ilk denemeler için yeterli olsa da derin öğrenme modelleri açısından hem boyut hem çeşitlilik bakımından sınırlıdır. Ayrıca LSTM'nin ardıl hareket tahmin başarımı, sporcunun kişisel yumruk tarzlarına ve hareket çeşitliliğine bağlı olarak değişkenlik

göstermektedir. Kullanılan sekans uzunluğunun (5 kare) artırılması veya Transformer tabanlı sıralı modellerin denenmesiyle, uzun süreli bağımlılıkların daha iyi yakalanarak tahmin performansının artırılması mümkündür.

LSTM entegrasyonu, yalnızca tahmin performansını artırmakla kalmayıp, gelecekte sporcuya özgü uyarlanabilir geri bildirim sistemlerinin geliştirilmesine de zemin hazırlamaktadır. %75 civarındaki tahmin başarısı, test verisinin yapısına ve sporcunun hareket çeşitliliğine bağlı olarak değişse de; sanal antrenman partneri teknolojisiyle entegre edilerek, sporcunun gelişiminin sürdürülebilirliği desteklenebilir.

Uygulamada sistem, bir boks müsabakasında sanal hakem olarak görev alabilir; yumruk temaslarını %87 doğrulukla tespit edebilir ve yumruk türlerini gerçek zamanlı sınıflandırarak hakem kararlarında insan hatasını minimize edebilir. Antrenman süreçlerinde ise sistem, sporcunun yumruk dizilerini analiz ederek kişiselleştirilmiş geri bildirim sağlayabilir ve antrenörlerin teknik iyileştirmelerini daha objektif şekilde yapmasına olanak tanır.

Bu çalışma, EfficientNet, ST-GCN/ST-Transformer ve LSTM yapılarını entegre ederek gerçek zamanlı hareket tanıma ve ardıl hareket tahmini yapabilen hibrit bir derin öğrenme sistemi sunmaktadır. Bu sistem, karmaşık sporlarda farklı derin öğrenme yöntemlerinin bir arada kullanılmasıyla yapay zekâ destekli sportif analizlerin daha objektif, ölçülebilir ve veri temelli hâle gelebileceğini ortaya koymuştur. Bu çalışmanın çıktıları, antrenman analizinden hakem karar destek sistemlerine kadar birçok uygulama alanında değerlendirilebilir.

Bu tez çalışması, derin öğrenme ve iskelet tabanlı modelleme yaklaşımlarının spor hareketlerinin sınıflandırılması ve tahmini süreçlerinde ne derece etkili olabileceğini göstermiştir. Geliştirilen sistem, boks gibi dinamik ve hızlı karar verilmesi gereken bir branşta hem mevcut vuruş türlerini yüksek doğrulukla sınıflandırabilmiş hem de ardışık hareketlerden bir sonraki hamleyi başarıyla öngörebilmiştir. Çalışmanın sonuçları, spor teknolojileri alanında yapay zekâ destekli çözümlerin yaygınlaştırılmasına katkı sağlamakta ve daha adil, ölçülebilir ve veriye dayalı karar mekanizmalarının geliştirilmesine olanak sunmaktadır.

Gelecek çalışmalarda sistemin farklı spor branşlarında (örneğin dövüş sporları, tenis, futbol) çoklu sporcu analizlerini destekleyecek şekilde genişletilmesi, giyilebilir sensör verileriyle entegre edilmesi ve pekiştirmeli öğrenme yaklaşımları ile kişiye özel geri bildirim mekanizmalarının geliştirilmesi hedeflenmektedir.

KAYNAKÇA

[1] Y. LeCun, Y. Bengio G. Hinton, "Deep learning," *Nature*, c. 521, no. 7553, ss. 436–444, 2015. doi: 10.1038/nature14539.

[2] J. Redmon, S. Divvala, R. Girshick A. Farhadi, "You only look once: Unified, real-time object detection," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, ss. 779–788.

[3] J. Redmon A. Farhadi, "YOLO9000: better, faster, stronger," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, ss. 7263–7271.

[4] C.-Y. Wang, A. Bochkovskiy H.-Y. M. Liao, "YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information," *arXiv preprint arXiv:2402.13616*, 2024.

[5] Z. Cao, T. Hidalgo, T. Simon, S.-E. Wei Y. Sheikh, "OpenPose: Realtime multi-person 2D pose estimation using part affinity fields," *IEEE transactions on pattern analysis and machine intelligence*, c. 43, no. 1, ss. 172–186, 2019.

[6] Y. Sun Y. Li, "A Deep Learning Method for Intelligent Analysis of Sports Training Postures," *Computational Intelligence and Neuroscience*, c. 2022, ss. 1–9, 2022. doi: 10.1155/2022/2442606.

[7] H.-S. Fang ve diğ., "AlphaPose: Whole-Body Regional Multi-Person Pose Estimation and Tracking in Real-Time," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, c. 45, no. 6, ss. 7157–7171, 2023. doi: 10.1109/TPAMI.2022.3222784.

[8] X. Yan, M. Wang, R. Zhang F. Li, "An Improved Deep Learning Method for Intelligent Analysis of Sports Training Posture Based on Causal Inference and Graph Convolution," *IEEE Access*, c. 11, ss. 16 932–16 945, 2023. doi: 10.1109/ACCESS.2023.3241234.

[9] *BoxPunch Detector Dataset on Roboflow Universe*, https://universe.roboflow.com/markmcquade/boxpunch-detector, Erişim tarihi: 12 Şubat 2025, 2025.

[10] K. He, X. Zhang, S. Ren J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE conference on computer vision and pattern recognition*, ss. 770–778, 2016.

[11] A. G. Howard ve diğ., "MobileNets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[12] M. Tan Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," *Proceedings of the 36th International Conference on Machine Learning*, PMLR, 2019, ss. 6105–6114.

[13] T. N. Kipf M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[14] S. Yan, Y. Xiong D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," *Proceedings of the AAAI Conference on Artificial Intelligence*, c. 32, 2018.

[15] C. Plizzari, M. Cannici M. Matteucci, "Skeleton-based action recognition via spatial and temporal transformer networks," *Computer Vision and Image Understanding*, c. 208, s. 103 219, 2021.

[16] S. Hochreiter J. Schmidhuber, "Long short-term memory," *Neural computation*, c. 9, no. 8, ss. 1735–1780, 1997.

[17] *Muhammad Ali vs. Zora Folley, 1967*, https://www.youtube.com/watch?v=VFFDe9FQL3s, Erişim tarihi: 12 Şubat 2025, 2025.

[18] C. Lugaresi ve diğ., "MediaPipe: A Framework for Building Perception Pipelines," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019, ss. 0–0.

TEZDEN ÜRETİLMİŞ YAYINLAR

Konferans Bildirisi

1. Y. Akkaya ve M. Bal, “Hibrit Derin Öğrenme Modeli ile Sanal Hakem Sistemi,” *Grad Symposium 2025: Mühendislikte Sürdürülebilir Çözümler*, YTÜ Davutpaşa Kampüsü, İstanbul, 15 Mayıs 2025. (Poster sunumu)