

KOCAELİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

BİLİŞİM SİSTEMLERİ MÜHENDİSLİĞİ
ANABİLİM DALI

YÜKSEK LİSANS TEZİ

BEBEKLERİN BEDEN DİLİNİ POZ TABANLI GERÇEK
ZAMANLI ÇÖZME

MUHAMMET MÜCAHİT ENES YURTSEVER

KOCAELİ 2022

KOCAELİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

BİLİŞİM SİSTEMLERİ MÜHENDİSLİĞİ
ANABİLİM DALI

YÜKSEK LİSANS TEZİ

BEBEKLERİN BEDEN DİLİNİ POZ TABANLI GERÇEK
ZAMANLI ÇÖZME

MUHAMMET MÜCAHİT ENES YURTSEVER

Doç. Dr. Süleyman EKEN

Danışman, Kocaeli Üniv.

.....

Doç. Dr. Hakan GÜNDÜZ

Jüri Üyesi, Kocaeli Üniv.

.....

Assist. Prof. Dr. Akhtar JAMIL

Jüri Üyesi, Nat. Univ. of Comp. and Emerg. Sci.

.....

Tezin Savunulduğu Tarih: 09.06.2022

ETİK BEYAN VE ARAŞTIRMA FONU DESTEĞİ

Kocaeli Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırladığım bu tez/proje çalışmada,

- Bu tezin/projenin bana ait, özgün bir çalışma olduğunu,
- Çalışmamın hazırlık, veri toplama, analiz ve bilgilerin sunumu olmak üzere tüm aşamalarında bilimsel etik ilke ve kurallara uygun davrandığımı,
- Bu çalışma kapsamında elde edilen tüm veri ve bilgiler için kaynak gösterdiğimi ve bu kaynaklara kaynakçada yer verdiğimi,
- Bu çalışmanın Kocaeli Üniversitesi'nin abone olduğu intihal yazılım programı kullanılarak Fen Bilimleri Enstitüsü'nün belirlemiş olduğu ölçütlere uygun olduğunu,
- Kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- Tezin/Projenin herhangi bir bölümünü bu üniversite veya başka bir üniversitede başka bir tez/proje çalışması olarak sunmadığımı,

beyan ederim.

☐ Bu tez/proje çalışmasının herhangi bir aşaması hiçbir kurum/kuruluş tarafından maddi/alt yapı desteği ile desteklenmemiştir.

☒ Bu tez/proje çalışması kapsamında üretilen veri ve bilgiler KOU BAP tarafından 2293 no'lu proje kapsamında maddi/alt yapı desteği alınarak gerçekleştirilmiştir.

Herhangi bir zamanda, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçları kabul ettiğimi bildiririm.

(İmza)

Muhammet Mücahit Enes YURTSEVER

YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI

Fen Bilimleri Enstitüsü tarafından onaylanan lisansüstü tezimin/projemin tamamını veya herhangi bir kısmını, basılı ve elektronik formatta arşivleme ve aşağıda belirtilen koşullarla kullanıma açma izninin Kocaeli Üniversitesi'ne verdiğimi beyan ederim. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin/projemin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanımı bana ait olacaktır.

Tezin/projenin kendi özgün çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin/projenin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim kurulu tarafından yayınlanan **“Lisanüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge”** kapsamında tezim aşağıda belirtilen koşullar haricinde YÖK Ulusal Tez Merkezi/ Kocaeli Üniversitesi Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- ☐ Enstitü yönetim kurulu kararı ile tezimin/projemin erişime açılması mezuniyet tarihinden itibaren 2 yıl ertelenmiştir.
- ☐ Enstitü yönetim kurulu gerekçeli kararı ile tezimin/projemin erişime açılması mezuniyet tarihinden itibaren 6 ay ertelenmiştir.
- ☒ Tezim/projem ile ilgili gizlilik kararı verilmemiştir.

(İmza)

Muhammet Mücahit Enes YURTSEVER

ÖNSÖZ VE TEŞEKKÜR

Yüksek lisans eğitimim boyunca benden yardımlarını esirgemeyen, akademik ve sosyal anlamda bana örnek olan ve doğru yolu gösteren akademik danışmanım Doç. Dr. Süleyman EKEN hocama teşekkür ederim.

Aileme ve arkadaşlarıma desteklerini esirgemedikleri ve yanımda oldukları için teşekkür ederim.

Haziran – 2022

M. Mücahit Enes YURTSEVER



İÇİNDEKİLER

ETİK BEYAN VE ARAŞTIRMA FONU DESTEĞİ.....	i
YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI.....	ii
ÖNSÖZ VE TEŞEKKÜR.....	iii
İÇİNDEKİLER.....	iv
ŞEKİLLER DİZİNİ.....	v
TABLOLAR DİZİNİ.....	vi
SİMGELER VE KISALTMALAR DİZİNİ.....	vii
ÖZET.....	ix
ABSTRACT.....	x
1. GİRİŞ.....	1
1.1. Konunun Önemi.....	1
1.2. Araştırma Problemleri ve Katkıları.....	2
1.3. Tez Yapısı.....	4
2. İLGİLİ ÇALIŞMALAR.....	6
2.1. Poz Tahmini.....	6
2.1.1. Yetişkinler için Poz Tahmini.....	6
2.1.2. Bebekler için Poz Tahmini.....	9
2.1.3. Bebek İzleme için Endüstriyel Uygulamalar.....	11
2.2. Videolarda Gerçek Zamanlı Nesne Takibi.....	12
2.2.1. Nesne Takibinde Geleneksel Yöntemler.....	12
2.2.2. Derin Öğrenme Tabanlı Nesne Takibi.....	16
3. MATERYAL VE METOTLAR.....	20
3.1. OpenPose Tabanlı Poz Tahmini.....	20
3.2. AlphaPose Tabanlı Poz Tahmini.....	23
3.3. KAPAO Tabanlı Poz Tahmini.....	26
3.4. LSTM ile Bebek Aktivitesi Tanıma.....	28
3.5. Deep SORT ile Gerçek Zamanlı Bebek Takibi.....	28
4. DENEYSEL SONUÇLAR ve TARTIŞMA.....	30
4.1. BabyPose Veri Seti.....	30
4.1.1. Veri Ön İşleme Adımları.....	32
4.2. Çalışma Ortamı Hazırlıkları.....	33
4.2.1. GitHub Üzerinden Projenin Kopyalanması.....	34
4.2.2. OpenPose HPE ile Aktivite Tanıma ve Bağlılıkların Kurulumu.....	34
4.2.3. AlphaPose HPE ile Aktivite Tanıma ve Bağlılıkların Kurulumu.....	35
4.2.4. KAPAO HPE ile Aktivite Tanıma ve Bağlılıkların Kurulumu.....	35
4.3. Performans Metrikleri.....	35
4.4. Performans Sonuçları ve Tartışma.....	36
4.4.1. OpenPose HPE Tabanlı Derin Ağın Performans Değerlendirmesi.....	37
4.4.2. AlphaPose HPE Tabanlı Derin Ağın Performans Değerlendirmesi.....	39
4.4.3. KAPAO HPE Tabanlı Derin Ağın Performans Değerlendirmesi.....	41
5. SONUÇLAR VE ÖNERİLER.....	43
5.1. Çalışmanın Kısıtları.....	43
5.2. Gelecek Çalışmalar.....	43
KAYNAKLAR.....	45
KİŞİSEL YAYIN VE ESERLER.....	50
ÖZGEÇMİŞ.....	51

ŞEKİLLER DİZİNİ

Şekil 3.1. OpenPose'un Önerilen Mimarisi: İki dallı çok aşamalı CNN.....	21
Şekil 3.2. OpenPose çok aşamalı CNN mimarisi.....	22
Şekil 3.3. SPPE uygulamalarında birden fazla sınırlayıcı kutu tespit problemi	24
Şekil 3.4. Önerilen RMPE çerçevesinin boru hattı	25
Şekil 3.5. Önerilen SSTN mimarisi ve paralel SPPE ile eğitim stratejisi	25
Şekil 3.6. Önerilen KAPAO mimarisi	27
Şekil 3.7. Bebek aktivitelerinin sınıflandırılması için önerilen mimari.....	29
Şekil 4.1. BabyPose veri setinden bazı örnekler.....	32
Şekil 4.2. Veri ön işleme adımları	32
Şekil 4.3. OpenPose HPE için eğitilen LSTM modeline ait karmaşıklık matrisi	38
Şekil 4.4. BabyPose veri setine ait görsellerin OpenPose HPE ile etiketlenmesi.....	39
Şekil 4.5. AlphaPose HPE için eğitilen LSTM modeline ait karmaşıklık matrisi	40
Şekil 4.6. KAPAO HPE için eğitilen LSTM modeline ait karmaşıklık matrisi	42

TABLolar DİZİNİ

Tablo 3.1. OpenPose HPE ile etiketleme işleminde kullanılan parametreler.....	23
Tablo 3.2. AlphaPose HPE ile etiketleme işleminde kullanılan parametreler.....	26
Tablo 3.3. KAPAO HPE ile etiketleme işleminde kullanılan parametreler	27
Tablo 4.1. Bebek aktivite video görüntülerinin sınıflara göre dağılımı.....	31
Tablo 4.2. Oluşturulan BabyPose veri setinin detayları	31
Tablo 4.3. OpenPose HPE ile aktivite tanıma için gerekli kütüphaneler	34
Tablo 4.4. OpenPose HPE ile aktivite tanıma için eğitilen LSTM ağının özeti	38
Tablo 4.5. OpenPose HPE ile HAR için eğitilen derin ağların karşılaştırması	38
Tablo 4.6. AlphaPose HPE ile aktivite tanıma için eğitilen LSTM ağının özeti	40
Tablo 4.7. AlphaPose HPE ile HAR için eğitilen derin ağların karşılaştırması	40
Tablo 4.8. KAPAO HPE ile aktivite tanıma için eğitilen LSTM ağının özeti	41
Tablo 4.9. KAPAO HPE ile HAR için eğitilen derin ağların karşılaştırması	41



SİMGELER VE KISALTMALAR DİZİNİ

Kısaltmalar

AE	: Autoencoder (Otokodlayıcı)
API	: Application Programming Interface (Uygulama Programlama Arayüzü)
ASD	: Autism Spectrum Disorder (Otizm Spektrum Bozukluğu)
BLE	: Bluetooth Low Energy (Bluetooth Düşük Enerji)
CNN	: Convolutional Neural Networks (Evrişimli Sinir Ağları)
CSRT	: Channel and Spatial Reliability Tracking (Kanal ve Mekansal Güvenilirlik Takibi)
DL	: Deep Learning (Derin Öğrenme)
DSA	: Derin Sinir Ağ
FPS	: Frames per Second (Saniyedeki Kare Sayısı)
GAN	: Generative Adversarial Network (Çekişmeli Üretici Ağ)
GPU	: Graphics Processing Unit (Grafik İşlem Birimi)
HAR	: Human Activity Recognition (İnsan Aktivite Tanıma)
HOG	: Histograms of Oriented Gradients (Yönlendirilmiş Gradyanların Histogramı)
HPE	: Human Pose Estimation (İnsan Poz Tahmini)
HSV	: Hue, Saturation, Value (Renk, Doygunluk, Değer)
IoT	: Internet of Things (Nesnelerin İnterneti)
IoU	: Intersection over Union (Birlik Üzerinden Kesişme)
JSON	: JavaScript Object Notation (JavaScript Nesne Gösterimi)
KAPAO	: Keypoints and Poses as Objects (Nesne Olarak Anahtar Noktalar ve Pozlar)
KCF	: Kernelized Correlation Filters (Çekirdekli Korelasyon Filtreleri)
KDE	: Kernel Density Estimation (Çekirdek Yoğunluğu Tahmini)
LSTM	: Long Short-Term Memory (Uzun-Kısa Hafızalı Bellek)
mAP	: Mean Average Precision (Ortalama Kesinlik)
MHAD	: Berkeley Multimodal Human Action Database (Berkeley Çok Modlu İnsan Eylem Veritabanı)
MIL	: Multiple Instance Learning (Çoklu Örnek Öğrenmesi)
MOSSE	: Minimum Output Sum of Squared Error (Kare Hatasının Minimum Çıkış Toplamı)
MOT	: Multiple Object Tracking (Çoklu Nesne Takibi)
MPII	: Max Planck Institute for Informatics
MSE	: Mean Squared Error (Ortalama Karesel Hata)
NMS	: Non-maximum suppression (Maksimum Olmayan Bastırma)
PAF	: Part Affinity Fields (Parça Yakınlık Alanları)
PGPG	: Pose Guided Proposal Generator (Poz Kılavuzlu Öneri Oluşturucu)
RGB	: Red, Green, Blue (Kırmızı, Yeşil, Mavi)
RGB-D	: Red, Green, Blue – Depth (Kırmızı, Yeşil, Mavi – Derinlik)
RMPE	: Regional Multi-Person Pose Estimation (Bölgesel Çok İnsanlı Poz Tahmini)
RNN	: Recurrent Neural Networks (Tekrarlı Sinir Ağları)
SDTN	: Spatial De-Transformer Network (Uzamsal De-Transformer Ağ)

SMIL	: Skinned Multi-Infant Linear Model (Tenli Çoklu Bebek Lineer Modeli)
SNN	: Siamese Neural Networks (Siyam Sinir Ağları)
SORT	: Simple Online and Realtime Tracking (Basit Çevrimiçi ve Gerçek Zamanlı İzleme)
SOT	: Single Object Tracking (Tek Nesne Takibi)
SPPE	: Single Person Pose Estimation (Tek İnsan Poz Tahmini)
SSTN	: Symmetric Spatial Transformer Network (Simetrik Uzamsal Transformatör Ağı)
STN	: Spatial Transformer Network (Uzamsal Transformatör Ağı)
TD	: Typically Developing (Tipik Olarak Gelişen)
TLD	: Tracking-Learning-Detection (İzleme-Öğrenme-Algılama)
TTA	: Test-Time Augmentation (Test Zamanı Artırımı)
WHO	: World Health Organization (Dünya Sağlık Örgütü)



BEBEKLERİN BEDEN DİLİNİ POZ TABANLI GERÇEK ZAMANLI ÇÖZME

ÖZET

İnsan poz tahmini; insan-bilgisayar etkileşimleri, oyun, sanal gerçeklik, video gözetimi, spor analizi, tıbbi destek ve yardımda yaygın olarak kullanılmaktadır. Bu çalışmada uzmanlar tarafından ortaya konulan bebeklerin belli başlı beden hareketlerinin poz tabanlı gerçek zamanlı çözülerek ilgili eylemlerin tespit edilmesi ve anlamlandırılması amaçlanmıştır. Bununla birlikte bebeğin güvenliğinin sağlanması hedeflenmiştir. Bu amaçla video paylaşım platformlarından 6 adet bebek aktivitesi, ilgili anahtar kelimeler kullanılarak toplanmıştır. Bu videolardan aktiviteyi içeren 4 ile 10 saniye uzunluğunda alt klipler oluşturulmuştur. Bu kliplere ait video çerçeveleri kullanılarak gerçek zamanlı iskelet tespit sistemleri olan OpenPose, AlphaPose ve KAPAO ile görsel üzerindeki iskelete ait koordinat bilgileri üretilmiştir. Bu iskelet koordinat bilgileri kullanılarak hafıza tabanlı bir derin öğrenme modeli oluşturulmuş ve bu derin öğrenme yöntemi ile bebek aktiviteleri sınıflandırılmıştır. Bununla birlikte Deep SORT algoritmasıyla bebeğin gerçek zamanlı takibi gerçekleştirilmiştir. Oluşturulan derin öğrenme modeli ile birlikte bebeğin olası şikayetinin habercisi olan hareketler başarıyla tespit edilmiştir. Bununla birlikte kamera açısında yer alan bebeğin takibi Deep SORT algoritması ile yapılmıştır. Bu çalışma ile hafıza tabanlı derin öğrenme modeli geliştirilerek bebekler üzerinde poz tabanlı gerçek zamanlı aktivite tanıyan sistem literatüre kazandırılmıştır.

Anahtar Kelimeler: Aktivite Tanıma, AlphaPose, KAPAO, OpenPose.

POSE-BASED REAL-TIME DECODING OF BABIES' BODY LANGUAGE

ABSTRACT

Human pose estimation is widely used in human-computer interactions, gaming, virtual reality, video surveillance, sports analysis, medical support and assistance. In this study, it is aimed to determine and make sense of certain activities of babies, which were revealed by experts, by solving poses in real-time. In addition to that, it is aimed to ensure the safety of the baby. For this purpose, 6 baby activities from video-sharing platforms were collected using relevant keywords. Sub-clips of 4-10 seconds long containing the activity were created from these videos. By using the video frames of these clips, the coordinate information of the skeleton on the image was produced with the real-time HPE systems such as OpenPose, AlphaPose and KAPAO. Using this skeleton coordinate information, a memory-based deep learning model was created and baby activities were classified with this deep learning method. Moreover, real-time tracking of the baby was made with the Deep SORT algorithm. With the deep learning model created, the movements that are the harbinger of the baby's possible complaint have been successfully detected. The baby in the camera angle was also followed up using Deep SORT algorithm. With this study, a memory-based deep learning model has been developed and a pose-based real-time activity recognition system on babies has been brought to the literature.

Keywords: Activity Recognition, AlphaPose, KAPAO, OpenPose.

1. GİRİŞ

İletişim, insanların sosyalleşmesini sağlayan en önemli unsurların başında gelmektedir. Bu iletişim unsurları aracılığıyla insanlar arasında bir etkileşim başlar. İnsanlar etkileşimi iletişime dönüştürmek için bazen kelimelere bazen de beden diline başvurur. Canlı bir varlığın mimikleriyle, jestleriyle ve vücut hareketleri ile gerçekleştirdiği sözel olmayan iletişim biçimi; beden dilini ifade etmektedir. Beden dili, duygu ve düşünceleri karşı tarafa daha doğru ve net bir şekilde aktarabilmek için bir araçtır. Beden dilini oluşturan öğeler ise; yüz ve mimik hareketleri, kaş ve göz hareketleri, vücudun duruş şekli veya hareket biçimi, ellerin duruşu ve hareketi, ayakların duruşu ve hareketi ve karşı tarafa olan mesafedir (Rowlands, 2011). Beden dilini kullanarak gönderilen işaretler veya sinyaller karşı tarafın bilinçaltında algılandıktan sonra anlamlandırılmış olur. Aslında beden dili, duygu ve düşünceleri karşı tarafa daha doğru ve net bir şekilde aktarabilmek için bir yöntemdir. Beden dili kullanılarak karşı taraf daha iyi anlaşılabilir ve daha doğru bir şekilde etkilenebilir. İlk karşılaşmalarda beden dilinin çok büyük bir önemi olduğundan bahsedilir; çünkü çoğu durumda ilk izlenim önemlidir ve bunu en iyi beden dili ile göstermek mümkündür. Psikoloji, sosyoloji gibi bilim dallarında bu konu geniş ve kapsamlı bir şekilde ele alınmıştır (Müller ve diğ., 2013).

Beden dili, bebeklerin beden dili özeline indirgendiğinde ise ayrıca önem arz etmektedir. Bebeğin hareketlerine alışmak, özellikle konuşmadan önce çok önemlidir. Boston'daki Brazelton Enstitüsü Müdürü ve “Bebeğiniz Sizinle Konuşuyor” kitabı yazarı Dr. J. Kevin Nugent, bebeğin beden dilinin okunması ile birlikte onun daha mutlu olacağını, böylelikle ebeveynlerin bebeğin kişiliğinin ve mizacının nasıl şekillendiği hakkında bilgi sahibi olabileceğini belirtmiştir (Nugent, 2011). Ayrıca bebek profesyonelleri bazı yaygın bebek hareketlerini yorumlama konusunda Dr. Nugent gibi fikir ve ipuçları vermektedir. Küçük bedenleri anlamak için bu ipuçlarını kullanmak önemlidir. Bu ipuçları çeşitli nedenlerden dolayı kimi bebeklerde daha az kimilerinde ise ortalamanın üzerinde seyretnmektedir.

1.1. Konunun Önemi

ABD’de yapılan bir araştırmaya göre, beden hareketleriyle iletişim kurabilen bebekler ileriki yaşlarında daha geniş kelime haznesine sahip oluyor. Chicago Üniversitesi’nden bilim adamları (URL-1), farklı sosyal çevreden 50 kadar aileyi kapsayan araştırmalarında bebeklerin beden dilini mercek altına aldı. Araştırma 14 aylıkken iletişim kurmak için

geniş bir hareket yelpazesi kullanan bebeklerin, 4.5 yaşında daha zengin kelime hazinesine sahip olduklarını ortaya koydu. Araştırma çocukların beden dilinin ailenin sosyo-ekonomik düzeyine göre farklılıklar gösterdiğini de belirledi. Yayınlanan araştırmaya göre, ekonomik durumu iyi olan ailelerde 14 aylık çocuklar kendini ifade etmek için en az 24 farklı vücut hareketi yapabilirken, ekonomik geliri daha düşük sereyden ailelerin bebeklerinde vücut hareketi sayısı ortalama 13'te kalmaktadır.

Bebeğin beden dili size nasıl hissettiğini ve sizden neye ihtiyacı olduğunu söyleyebilir. Bebeğin beden dili yorgunluk, açlık, huzursuzluk, uyanık ve oynamaya hazır olma, dinlenmeye ihtiyacı var gibi çeşitli ipuçları verebilmektedir. Bununla birlikte her bebek, kendi işaretini geliştirebilmektedir. Bebeğin beden dili fark edilip ve ona cevap verildiğinde bebek kendini daha güvende hissetmektedir. Bebeğin bireysel ipuçları ve duyguları hakkında ne söylediklerini öğrenmek mümkündür. Ebeveyn ve bebek birbirini tanıdıkça bebeğin bireysel ipuçlarına da yanıt vermek kolaylaşır. Bu çalışma özelinde bebeklerin bazı hareketleri yorumlanacak olursa; göz kaşıma ve gerinme hareketi ile bebek yorgunluğunu ve uyku isteğini, parmak emme ile bebeğin acıkmış olduğu ya da diş kaşınması yaşadığını ve dizleri kırma, tekme sallama ile bebeğin mutsuzluk ve huzursuzluk içerisinde bulunduğunu çıkarımı yapılabilir. Bebeğin ilgili şikâyetlerinin tespiti ve bu şikâyetin giderilmesi; ebeveyn-bebek arasındaki bağı güçlenmesi, birbirleri arasındaki uyumun artması ve bebeğin karakteristik özelliklerinin doğru şekillenebilmesi adına büyük önem taşıdığı aşikardır.

1.2. Araştırma Problemleri ve Katkılar

Beden dili birçok canlının iletişimi açısından büyük öneme sahiptir. Bebeklerin de bedenleriyle/uzuvlarıyla yapmış oldukları hareketlerin de bir anlamı vardır. Bu hareketlere ve olası anlamlarına daha önceden değinilmiştir. Bu çalışmada bebek video görüntüleri kullanılarak onların beden dillerinin çözülmüştür. Bu sayede yeni çocuk sahibi ebeveynlere yardımcı olunacağı hususunda fayda sağlayacağı aşikardır.

Çalışma kapsamında önerilen aktivite tanıma yöntemi insan pozlarına dayanmaktadır. Literatürde çoğunlukla yetişkin insan görüntüleri üzerinden aktivite tanıma çalışmaları yapılmıştır. Dolayısıyla yetişkinlerin poz iskeletlerindeki anahtar nokta ve bu noktalar arasındaki uzunluklar ile bebeklerin poz iskeletlerindeki anahtar nokta ve bu noktalar

arasındaki uzunluklar farklılık göstermektedir. Poz tahminleme noktasında oluşan bu farka göre performans değerlendirilmesi yapılmıştır.

Literatürde sınıflandırılan aktiviteler; bükme, boks, kol çaprazlama, kalkma, postür, alkışlama, koşma, atlama, tekme, oturma, yürüme olarak karşımıza çıkmaktadır. Çalışma kapsamında bebeklerin beden eylemlerine karşılık gelen (gerinme, parmak emme, sırtını geriye doğru kendine çekme, düzenli tekme sallama, kafayı vurma, göz kaşıma) aktivite görüntülerinden oluşan bir veri seti oluşturulmuş ve eylemler bu bahsedilen altı sınıf içerisinde sınıflandırılmıştır. Kullanılan derin öğrenme modellerinde modelin başarısı büyük oranda girdi verisinin kalitesine bağlı olduğundan, altı adet aktiviteden eşit miktarda veri toplanmış ve üretilmiştir.

Çalışma kapsamında videodaki bebeklerin gerçek zamanlı olarak takibinin yapılması da ele alınan problemlerdendir. En popüler ve en yaygın kullanılan nesne izleme çerçevelerinden biri SORT'un (Simple Real Time Tracker, SORT) bir uzantısı olan Deep SORT'dur (Wojke ve diğ., 2017). Video çerçevesindeki bebeğin takibi Deep SORT algoritması ile yapılmıştır.

Literatürdeki çalışmalar incelendiğinde önerilen projenin amaçlarından birisi, poz tabanlı olarak bebeklerin beden dilini gerçek zamanlı çözerek onların aktivitelerini anlamlandırma yoluyla yeni ebeveynlere yardımcı olmaktır. Belirli aksiyonlardan yola çıkarak bebeklerin beden dili anlamlandırılmıştır. Geliştirilen derin öğrenme modeli, daha önceden bahsedilen altı adet bebek eylemini sınıflandırmaktadır. Bebeklerin hareketlerinden birtakım riskli durumların belirlenmesi şeklinde önerilen çalışma genişletilebilir bir yapıdadır.

Bu çalışmada bebek videoları çevrimiçi içerik paylaşım platformlarından toplanmış, toplanan videolardan altı adet aktivite referans alınarak aktiviteyi içeren kısa alt klipler oluşturulmuştur. Oluşturulan bu klipler, gerçek zamanlı iskelet tespit sistemleri olan OpenPose, AlphaPose ve KAPAO ile etiketlenmiştir. Etiketlenen girdi görsellerinde etiket bilgisi olarak görsel üzerinde tespit edilen eklemlerin koordinat bilgileri yer almaktadır. Bu koordinat bilgilerinden 19 özellik alınmıştır. Eksik etiketlenen görsellerin özellikleri, veri setinden çıkarılmıştır. Gerekli ön işlemlerden sonra bu etiketli veriler, makine öğrenimi için eğitim ve test olarak iki gruba ayrılmıştır. Makine öğrenmesi

modeli, Uzun kısa süreli bellek (Long Short-Term Memory, LSTM) kullanılarak oluşturulmuştur. LSTM modelleri, gelecekteki verileri tahmin edebilmek için bir dizi veriye ihtiyaç duyar. Bu nedenle, LSTM modelinin aktiviteleri tahmin etmesi için bir blok boyutu belirlenmelidir. Bu çalışmada blok boyutu 20 olarak ayarlanmıştır. Blok boyutu, sınıflandırıcının karar vermek için girdi videosundan yirmişer çerçeveyi değerlendirerek bir sonuç üreteceği anlamına gelir. Girdi videolarının saniyedeki kare sayısı 20-25 arasında seyrettiğinden, blok sayısının 20 olması, sınıflandırıcı tarafında yaklaşık olarak saniyede bir tahmin üretileceği anlamına gelmektedir.

Makine öğrenimi tabanlı aktivite tanıma ve nesne izleme uygulamalarıyla çalışan araştırmacıların çalışma kapsamında üretilen veri setinden yararlanması amaçlanmıştır. Aktivite tanıma alanında daha önce kaynak olarak bulunmayan bebek videolarını derleyerek, BabyPose adı verilen veri seti oluşturulmuştur. Bu veri seti ile birlikte aktivite tanıma alanında çalışan araştırmacılara farklı bir veri setinin sunulması hedeflenmiştir. Bu sayede var olan aktivite tanıyan modeller yeniden değerlendirilebilir ve başarıları artırılabilir.

1.3. Tez Yapısı

Takip eden ilgili çalışmalar bölümünde, poz tahmini için önerilen mimari ve yöntemlere değinilmiştir. Bu çalışmada önerilen HPE yöntemlerine göre diğer gerçek zamanlı iskelet tespit sistemlerine değinilmiştir. Poz tahminleme kütüphaneleri kendi kategorilerinde incelenmiştir. Takip eden alt bölümde gerçek zamanlı nesne takibi mimarilerine değinilmiş, Deep SORT ile gerçekleştirilen önceki çalışmalara değinilmiş ve mevcut çalışma ile karşılaştırılmıştır.

Materyal ve metotlar kısmında bu çalışmada kullanılan gerçek zamanlı insan iskelet sistemi tanıyan kütüphanelerden bahsedilmiştir. Yöntemin nasıl çalıştığından, daha tutarlı ve hızlı çalışması açısından yapılan iyileştirmelere değinilmiştir. Ardından bebek aktivite tanıma bölümünde, giriş kısmında kısaca bahsedilen önerilen yöntemin detayları verilmiştir. Takip eden son alt bölümde Deep SORT ile nesne takibinin işleyişinden bahsedilmiştir.

Deneysel sonuçlar ve tartışma kısmında BabyPose çalışma ortamının hazırlığı, veri setinin oluşturulması, etiketlenmesi, veri ön işleme adımları ve önerilen hafıza tabanlı

makine öğrenmesi modelinin performansının incelenmesi yapılmıştır. Sonuçlar kısmında ise çalışmanın kısıtlarından bahsedilmiş, gelecek çalışmalar ile daha tutarlı bir modelin nasıl oluşturulabileceği hakkında fikirlere yer verilmiştir.



2. İLGİLİ ÇALIŞMALAR

Bu bölümde poz tahmini ve gerçek zamanlı nesne takibi ile ilgili yapılan çalışmalar hakkında bilgiler verilmiş ve önerilen mimarının bu çalışmalardan farkına değinilmiştir.

2.1. Poz Tahmini

Önceki başlıklarda değinildiği üzere bebeklerin ilgili hareketlerini bir aktivite olarak değerlendirmek mümkündür. Bu aktiviteler de insan iskelet sistemi duruşundan tanınabilir. Bir insan iskeleti görseli, bir kişinin grafiksel bir biçimde oryantasyonunu temsil etmektedir. Kişinin pozunu belirten ise ilgili görsel üzerindeki bir dizi koordinattır. İskeletteki her bir koordinat bir parça (veya eklem veya anahtar noktası, keypoint) olarak bilinir. İki parça arasındaki geçerli bağlantı ise bir çift (veya uzuv) olarak tanımlanır. İnsan poz tahmini; insan-bilgisayar etkileşimleri, oyun, sanal gerçeklik, video gözetimi, spor analizi, tıbbi destek ve tıbbi yardımda yaygın olarak kullanılmaktadır. Bu kullanım da onu bilgisayarlı görü alanında oldukça popüler bir araştırma konusu haline getirmektedir (Insafutdinov ve diğ., 2016; Cao ve diğ., 2017; URL-2). İnsan poz tahmininde spor analizi için geliştirilen poz tahmini uygulaması HomeCourt (Ulin ve diğ., 2016), basketbol oyuncularının hareketlerini analiz etmek için poz tahminlemesini kullanmaktadır.

2.1.1. Yetişkinler için Poz Tahmini

İnsan poz tahmini, insan vücudunun parçalarını görüntülerden veya videolardan otomatik olarak bulmayı amaçlamaktadır. Görüntüde sadece bir kişinin bulunduğu veya kişinin pozisyonunun verildiği basit bir durumda görsel üzerinde boynun merkezi, sol/sağ dirsekler ve bacaklar ile sol/sağ omuzlar gibi eklem ve uzuvların tespit edilebilmesi için tek kişilik bir algoritmanın dikkate alınması yeterlidir. Görüntüdeki kişi sayısı ve kişilerin konumlarının bilinmediği daha karmaşık durumlarda çok kişili poz tahmin algoritmaları gerçekleştirilir (Luvizonve diğ., 2019). Bu algoritmalarda görüntü üzerindeki insanlara ait tüm uzuvlar tespit edilmeli ve anahtar/kilit noktaları ilişkilendirilmelidir. Tüm insan poz tahmini metodolojileri, (i) tek insan içeren poz tahmin yaklaşımları ve (ii) çok kişi içeren poz tahmin yaklaşımları olmak üzere ikiye ayrılır. Tek kişi içeren yaklaşımlar bir görüntüdeki belirli bir kişinin iskeletini algılar. Verilen pozisyon bilgisi sebebiyle tek kişi içeren yaklaşımların amacı, görsel üzerinde o alandaki anahtar nokta pozisyonunu

bulmaktır. Çok kiři ieren poz tahmin yaklařımlarının amacı ise kısıtsız bir problemi özmektir. Kiřilerin konumu ve sayısı, tek kiřilik sistemlerdeki gibi izole edilmiř deęildir.

Tek kiři ieren poz tahminleme yaklařımları, anahtar noktaları öngörme řekline göre (i) doğrudan regresyon tabanlı yaklařımlar (Newell ve dię., 2016) ve (ii) ısı haritası (heatmap) tabanlı yaklařımlar (Toshev ve Szegedy, 2014) olmak üzere iki tipte sınıflandırılmaktadır. Doğrudan regresyon yaklařımlarında ıkıř özellięi anahtar noktalara baęlanım (regression) amaçlı kullanılır. Isı haritası tabanlı yaklařımlarda ise video görselleri üzerinde ısı haritaları oluşturulur. Isı haritasındaki piksel deęeri, o konumdaki anahtar nokta var olma olasılıęına iřaret eder. İnsan uzvuna iřaret eden anahtar noktalar bu řekilde tespit edilmiř olur. Yakın zamanda regresyon ve ısı haritası tabanlı yaklařımlar haricinde daha farklı ve hızlı alıřmayı amaçlayan sistemler de geliřtirilmiřtir. Örnek olarak KAPAO, poz tespiti ve anahtar nokta tespitini ayrı ayrı yapar ve bu tespitleri birleřtirerek ısı haritasından baęımsız, farklı bir yöntemle poz tahminleme gerekleřtirir.

ok kiřili poz tahmin yaklařımları ise (i) yukarıdan ařaęıya (top-down) yaklařımlar ve (ii) ařaęıdan yukarıda (bottom-up) yaklařımlar olmak üzere temel olarak iki kategoride sınıflandırılabilir. Yukarıdan ařaęıya yaklařımlar, insan tespiti ve insan uzvunu belirten anahtar nokta tespiti (Insafutdinov ve dię., 2017) olmak üzere iki ana ařamadan oluşur. Ařaęıdan yukarıya yaklařımlarda (Dang ve dię., 2019) ise öncelikle anahtar nokta tespiti yapılır. Ardından bu anahtar noktaları belirten uzuvların hangi kiřiye ait olduęunu tespit eder ve gruplar (Liu ve dię., 2015; Yang ve Ramanan, 2012).

İnsan poz tahmini problemini özmek iin literatürde bazı yaklařımlar önerilmiřtir. El yapımı özellikler (hand-crafted features) ilk yapılan alıřmalarda kullanılmıřtır. Yönlendirilmiř Gradyanların Histogramı (Histograms of oriented gradients, HOG) (Eichner ve Ferrari, 2010) ve Edgelet (Herath ve Harandi, 2017) gibi el yapımı özellikler, uzuvların doğru řekilde belirlenmesinde yetersiz kalmıřtır. El yapımı özelliklerin aksine, derin öğrenme tabanlı yöntemler meta-verilerden daha yeterli özellikler elde edebilmektedir. Bu tür yöntemler, daha iyi sonuç ve büyük tolerans ile derin olmayan yöntemlerden (el yapımı ve grafik tabanlı olan yöntemler) daha iyi performans göstermektedir. Poz tahminleme alanında derin öğrenmenin kullanımı nispeten yeni olmasına karřılık bu konuda birok alıřma yapılmıřtır. El yapımı özellikler ile kurulan

poz tahminleme mekanizmalarının başarısızlığının aksine hem el yapımı hem de derin öğrenme temelli yaklaşımlardan en umut verici sonuçları birleştirmeye çalışarak aralarında avantaj-dezavantaj durumlarını da gözetken hibrit yaklaşımlar da mevcuttur (Toshev ve Szegedy, 2014).

El yapımı özellikler gibi klasik poz tahmini yaklaşımlarının başarı anlamında büyük sınırlamaları vardır. Bu sebeple poz tahmini uygulamaları evrimsel sinir ağları (Convolutional Neural Networks, CNN) tarafından büyük ölçüde yeniden şekillendirilmiştir. Toshev ve arkadaşlarının DeepPose (Toshev ve Szegedy, 2014) çalışması ile birlikte insan poz tahmini uygulamaları klasik yaklaşımlardan derin öğrenme tabanlı yaklaşımlara evrilmeye başladı. Son yıllarda yapılan iskelet duruş ve poz tahmini çalışmalarının büyük bir kısmında evrimsel sinir ağları ana yapı taşı olarak benimsenmiş ve bu evrimsel derin öğrenme modelleri, el yapımı ile grafik modellerinin yerini almıştır. Bu değişiklik ile değerlendirme testlerinde (benchmarks) önemli gelişmeler görülmüştür.

Literatürde bahsi geçen çalışmalarda insan poz tahmini birçok farklı alanda kullanılmıştır. Çalışmalarda görülen iki ana konu poz tahmini ve aktivite tanıma üzerine kurulan senaryolardır. Angelini ve arkadaşları (Angelini ve ark., 2018) duruş düzeyinde insan eylemi tanıma (human action recognition, HAR) için yeni bir 2D poz tabanlı algoritma olan ActionXPose'yi önermişlerdir. Kapsamlı simülasyonlar, ActionXPose algoritmasının aynı anda farklı veri kümeleri kullanılarak başarılı bir şekilde eğitilebileceğini göstermiştir. Duruşla ilgili HAR problemleri (i3DPost, KTH) için popüler veri kümelerinde başarılı bir performans göstermiş ve sonuçlar, ActionXPose baz alınacak şekilde diğer yöntemlerle kıyaslanmıştır. ActionXPose ile; bükme, boks, kol çaprazlama, kalkma, alkışlama, koşma, atlama, tekme atma, oturma ve yürüme gibi eylemler sınıflandırılmıştır. Osokin (Osokin, 2018) ise çalışmasında uç cihazlarda OpenPose tabanlı gerçek zamanlı aktivite tanıma performansı elde edebilmek amacıyla bu platformlar için optimize edilmiş bir HAR ağı önermiştir. Önerilen bu yeni derin ağ, işlemci üzerinde çalışmaktadır. Yadav ve arkadaşları (Yadav ve diğ., 2019) OpenPose tabanlı derin öğrenme algoritması kullanarak yoga asanalarını gerçek zamanlı olarak doğru bir şekilde tanıyan bir model inşa etmişlerdir. Çalışma kapsamında altı adet yoga asanası, on erkek ve beş kadın birey kullanılarak RGB web kamerasında kaydedilmiş ve

oluşturulan veri seti de paylaşılmıştır. CNN-LSTM hibrit derin öğrenme modeli kullanılmış ve sistemin tek video çerçevelerinde %99,04 test doğruluğuna ulaştığı belirtilmiştir. Noori ve arkadaşları (Noori ve diğ., 2019) RGB görüntülerden anatomik anahtar noktaları çıkarmak için açık kaynaklı gerçek zamanlı poz tahmin kütüphanesi OpenPose'yi kullanarak aktivite tanıma için gürbüz bir yaklaşım önermişlerdir. Bu tespit edilen kilit noktalarını, ardışık karelerdeki hareketleri de göz önünde bulundurarak gürbüz hareket özelliklerini elde etmişlerdir. Atlama, oturma, kalkma gibi on bir adet eylem içeren Berkeley Multimodal Human Action Database (MHAD) üzerinde önerdikleri sistemi test etmişler %92,4 genel doğruluğa ulaşan RNN-LSTM derin öğrenme modelini inşa etmişlerdir.

Uddin ve Torresen (Uddin ve Torresen, 2018) vücut iskeletini OpenPose ile çıkardıktan sonra gürbüz uzay-zamansal özellikler çıkarmışlardır. Termal videolardan elde edilen gürbüz özellikler, aktivite modelleme ve tanıma için tekrarlı sinir ağına (RNN) uygulanmıştır. Önerilen mimari, diğer RGB kameraların görsel olarak anlaşılabilir görüntüler üretmediği ortamlardaki insanları gözlemleyebilmek adına oldukça yararlıdır. Literatürde farklı aktivite türlerinin tanınmasına yönelik araştırma ve inceleme makaleleri de bulunmaktadır (Dang ve ark., 2020; Demrozi ve ark., 2020). Bildiğimiz kadarıyla, video akışlarında bebeklerin hareketlerini (sırtını eğme, kafaya vurma, bacaklarını tekmeleme, göz ovma, germe, parmak emme gibi) tanıyan bir çalışma bulunmamaktadır. Gerçekleştirilen bu çalışma ile birlikte bebekler için çalıştırılacak iskelet-aktivite tanıma modelleri için bir veri seti de literatüre kazandırılmıştır.

2.1.2. Bebekler için Poz Tahmini

Bebeklerin poz ve derinlik bilgilerini içeren RGB-D (Red, Green, Blue – Depth, Derinlik) verileri üzerinde yapılan çalışmalar, literatürde vücut modelleme, gelişimsel bozuklukların değerlendirilmesi gibi farklı amaçlar için de kullanılmıştır. Hesse ve arkadaşları (Hesse ve ark., 2017, Hesse ve ark., 2018) bebeklerin şeklini ve hareketini yakalayan ve müdahaleci olmayan, düşük maliyetli bir görüntü alma sistemi geliştirmişlerdir. Yetişkin vücut şeklini modelleme çalışmalarının ötesine geçerek gürültülü, düşük kaliteli ve eksik RGB-D verilerinden bir 3D Derili Çok Bebekli Doğrusal vücut modeli (3D Skinned Multi-Infant Linear body model, SMIL) öğrenmişlerdir. Klinik ortamında 37 bebek ile şekil ve hareket kaydettiler. Kamuya açık

değerlendirme veri setlerinin (kıyaslamaların) olmaması nedeniyle, farklı yaklaşımların karşılaştırılması şöyle dursun, standartlaştırılmış bir değerlendirme bile mümkün olmamaktadır. Hesse ve arkadaşları (Hesse ve ark., 2019), yakın zamanda piyasaya sürülen 3D SMIL gövde modelini kullanarak MINI-RGBD veri setini oluşturmuş ve bu veri setini açık kaynak olarak herkesin erişimine sunmuşlardır.

Hesse ve ark. (Hesse ve ark., 2018), bağlı işaretleyicilere veya sensörlere ihtiyaç duymadan yenidoğanların 3D eklem konumlarını ve kafa rotasyonlarını tespit etmek (yakalamak) için otomatik bir teknik oluşturmuşlardır. Hareketlerin aralığı, değişkenliği ve simetrisi ile ilgili her iki vücut tarafının baş, gövde, üst ve alt eklemlerinin hareket parametrelerini tanıtip motor davranışın değerlendirilmesi için objektif bilgiler sunmuşlardır. Bozuk motor gelişimi riski yüksek 5 bebeğin 6 kaydını analiz edip, sistemin bozukluklara işaret eden hareket özelliklerini nasıl vurguladığını, vurgulayabildiğini değerlendirmişlerdir. Marcoft ve ark. (Marcoft ve ark., 2015), yüksek motor sakatlık riski taşıyan yenidoğanlarda sinir sistemi gelişiminin daha iyi anlaşılmasına yol açabilecek, büyümekte olan bir disiplinler arası iş birliği alanını temsil edecek bir literatür taraması sunmuşlardır. Sciortino ve ark. (Sciortino ve ark., 2017) çocuklar üzerinde insan poz tahmini için farklı son teknoloji yaklaşımlarının performanslarını anlamak için bir değerlendirme veri kümesinde bu yaklaşımları karşılaştırmışlardır. Sciortino ve arkadaşları tarafından yapılan bu çalışmalar, klinik çalışmalarda bebek hareket risklerinin değerlendirmesinde kullanılmıştır.

Chambers ve ark. (Chambers ve ark., 2020) video tabanlı bir nöromotor test tekniği önermiştir. Çevrimiçi olarak keşfedilen 85 videoyu kullanarak, bebek hareketlerine ilişkin “normatif” bir referans kütüphanesi oluşturmuşlardır. Vücut uzuvlarının hareketlerini belirlemek için OpenPose kullandılar. Normatif veri tabanını kullanarak, her yenidoğanın sağlıklı bebeklerin normal hareketlerinden ne kadar saptığını ölçmek amacıyla tek bir puan olan Nave Gaussian Bayesian Surprise (Lonini ve ark., 2016) kullanırlar. Sistem daha sonra nöromotor risk miktarının bir doktor tarafından tanımlandığı (düşük, orta ve yüksek) bir klinik popülasyonda (N = 19) test edilmiş ve Bayesian Surprise'nin katılımcı grupları arasında farklılık gösterdiği keşfedilmiştir.

Putatunda (Putatunda, 2020), fiziksel ve motor bozuklukları olan çocuklarda öz bakım sorunlarını çözmek için Care2Vec adlı yeni bir derin öğrenmeye dayalı yaklaşım

önermiştir. Önerilen çalışma, Dünya Sağlık Örgütü (World Health Organization, WHO) tarafından geliştirilen kavramsal bir çerçeveye dayanmakta ve öz bakım davranışlarını içeren gerçek dünya veri setini kullanmaktadır. Kojovic ve ark. (Kojovic ve ark., 2021), Otizm spektrum bozukluğu (Autism spectrum disorders, ASD) olanlarla tipik olarak gelişen (typically developing, TD) olanlar arasında ayırım yapabilen bir makine öğrenimi modeli geliştirdiler. Bir çocuk (otizmlili veya TD'li) ve bir yetişkin arasındaki sosyal etkileşimleri içeren, otizmde erken gelişim üzerine daha geniş bir çalışmanın parçası olarak toplanan kayıtları kullanarak altın standart tanı muayenesi üzerinde derin bir sinir ağı eğittiler. Giriş videolarının boyutu, videonun tüm katılımcıları dahil olmak üzere iskelet anahtar noktaları çıkarmak için OpenPose teknolojisi kullanılarak düşürülmüştür. Daha sonra aktivite tanıma için CNN-LSTM mimarisini kullanmışlardır. Yukarıda görülebileceği gibi, bu sınırlı çalışmalar, bebek hareket risklerini değerlendirmek için klinik çalışmalarda kullanılmıştır.

2.1.3. Bebek İzleme için Endüstriyel Uygulamalar

Literatürde bebek bakımı veya gözetimi ile ilgili çok az araştırma bulunmaktadır. Ayrıca araştırmacılar, bebek izlemeden daha çok çocuk bakımı ve gözetimi ile ilgilenmektedirler. Ferron ve ark. (Ferron ve ark., 2019), çocukların bağımsız hareketlerinin gelişimlerine katkısına dayalı olarak, Bluetooth Düşük Enerji (Bluetooth Low Energy, BLE) ve akıllı telefon tabanlı çocukların yakınlığı ve güvenliğinin izlenmesi üzerinde çalışmışlardır. Gilmore (Gilmore, 2020), ebeveynlerin çocuklarının nerede olduğunu takip etmelerini sağlayan bir konum izleme aygıtı olan giyilebilir Jiobit'in eleştirel bir incelemesini yapmıştır. Cihazın coğrafi çit özelliklerini birleştirdiğini vurgulayarak, çeşitli teknolojik ve altyapı süreçleri aracılığıyla çocukları izlemek için bir yöntem önermişlerdir. Hüseyin ve ark. (Hussain ve ark., 2019), akıllı bir multimodal sistem kullanarak bebek davranışını izleyerek görüntü sensörü IoT (Internet of Things, Nesnelerin İnterneti) teknolojilerinde yeni bir paradigmaya öncülük etmiştir. Hasan ve Negulescu (Hasan ve Negulescu, 2020), bebek izleme giyilebilir teknolojileri alanındaki son gelişmeleri ve bunların gerçek zamanlı yaşamsal belirtileri tespit edebilme yeteneklerini araştırmış ve açıklamışlardır. Morales ve ark. (Morales ve ark., 2019), gerçek zamanlı olarak sıcaklık okumaları alan bir elektrikli cihaz ürettir. Bu cihaz, eşzamanlı olarak 0-12 aylık yenidoğanlardan

edindiđi ses seviyesini son kullanıcıya sunmakta ve ebeveynlere bebek takibi konusunda yardımcı olmaktadır.

Literatürdeki tüm çalışmalar incelendiğinde BabyPose'nin amacı, bebeklerin vücut dilini gerçek zamanlı olarak pozlardan hareketle deşifre ederek ve etkinliklerini anlamlandırarak yeni anne babalara yardımcı olmaktır. Belirli eylemlere dayanarak bebeklerin beden dili yorumlanabilmektedir.

2.2. Videolarda Gerçek Zamanlı Nesne Takibi

Video içerisinde nesne tespiti ve takibi üzerine uzun zaman aralığında birçok çalışma mevcuttur. Bu çalışmalarda kullanılan materyaller zaman içerisinde deęişiklik göstermiştir. Derin öğrenme öncesi modellerde genellikle el yapımı özellikler gibi üretmesi maliyet ve zaman alan özellikler ile tespit ve takip uygulamaları geliştirilmiştir. AlexNet gibi başarılı makine öğrenimi çalışmalarının sayısı literatürde arttıkça gerçek zamanlı nesne takibi ve tespiti problemleri, derin öğrenme tabanlı modeller ile çözülmeye çalışılmıştır. Bu bölümde nesne takibi özelinde geleneksel yöntemlere ve derin öğrenme tabanlı yöntemlere değinilmiştir.

2.2.1. Nesne Takibinde Geleneksel Yöntemler

Video izleme olarak bilinen bilgisayarlı görü alanı, videodaki hareketli nesnelerin konumunun tespit edilmesi ile ilgilidir. Video izleme, robotik, spor analizi ve video gözetimi gibi alanlarda çok sayıda kullanıma sahiptir. Tespit yoluyla izleme yöntemi, nesne izlemeye yönelik önemli bir yaklaşımdır. Nesne algılama algoritmaları, bir çerçevedeki nesneleri algılamak için tespit yoluyla izlemede kullanılır. Daha sonra bu nesneler, mevcut çerçevedeki nesneleri önceki çerçevelerdeki nesnelerle ilişkilendiren bir izleme algoritması kullanılarak izlenir. İzleme algoritması, her çerçevede tanınan nesnelere dayandığından, nesne tespiti için güvenilir bir mekanizmaya sahip olmak çok önemlidir.

Bilgisayarlar; düşünme, durumları analiz etme ve insanlar gibi karar mekanizması oluşturma hususunda her geçen gün daha başarılı bir hale gelmektedir. Makinenin bir fotoğrafı veya video çerçevesini anlaması ve önceki çerçeveleri değerlendirmek suretiyle bir karara varabiliyor olması, makine zekâsı alanındaki bu ilerlemenin bir nevi somut bir

göstergesidir. Buradaki kilit gelişme, makinenin görseller veya video çerçeveleri üzerindeki kapsayıcı desenleri tespit edip anlamlandırabilmesidir. Nesne tespitine nazaran üzerinde daha az çalışılan konulardan birisi de nesne izlemedir (object tracking). Bir videodaki nesneleri izlemek, geçici (static) görüntülerdeki nesneleri tespit etme ve zamansal (temporal) bilgileri analiz etme bilgilerini birleştirmeyi gerektirir. Nesne takibindeki amaç, görüntüdeki bir nesneye kitlenmek ve ilgili nesneyi çerçeveden çıkana kadar izlemektir.

Nesne takibi için önerilen geleneksel yöntemlerin birisi de optik akıştır (optical flow) (Horn ve Schunck, 1981). Optik akış, nesne ve kamera arasındaki göreceli hareketin neden olduğu ardışık video çerçeveleri arasındaki nesnelerin hareketidir. Optik akış her vektörün, noktaların birinci kareden ikinci kareye hareketini gösteren bir yer değiştirme vektörü olduğu bir 2 boyutlu vektör alanıdır. Optik akış ile video çerçevesindeki nesnenin yatay ve düşeydeki koordinatının yanı sıra, çerçevedeki objenin zaman bilgisi de kullanılır. Trafikte seyir halindeki araçların hızlarını tahmin etmek veya bir sonraki karedeki konumlarını tahminlemek için araçların çerçeveler arasındaki hareketini izlemek, optik akış ile mümkün olabilmektedir. Aynı şekilde basketbol, bowling, okçuluk gibi insan aktivitelerini de tanıyabilmek için nesne tanıma uygulamalarında zaman boyutuna ihtiyaç duyulur. Optik akış, bir nesnenin piksel yoğunluklarının ardışık kareler arasında değişmediğini ve komşu piksellerdeki takip edilen objelerin benzer harekete sahip olduğunu varsayar. Seyrek (sparse) ve yoğun (dense) optik akış olmak üzere iki temel implementasyonu bulunur. Seyrek optik akışta, çerçeve üzerindeki nesnenin kenarlarını ya da köşelerini gösteren birkaç akış vektörleri gösterilir. Yoğun optik akışta ise tüm çerçevenin optik akış vektörü gösterilir. Yoğun optik akışın seyrek modele göre daha yüksek doğruluğa sahip olmasına karşın daha yavaş çalışmaktadır.

Bilgisayarlı görü alanında en çok talep edilen özelliklerinden biri nesne takibidir. Ancak, bu aynı zamanda bilgisayarlı görünün gerçek zamanlı olarak başarılı çalışması gerektiğinden uygulanması birkaç açıdan meşakkatlidir; bir nesnenin sadece tanımlanması yetmemekte, bununla birlikte hareket halindeyken gerçek zamanlı olarak işlenebilmesi için yeterince hızlı bir şekilde tanımlanması gerekmektedir. Meanshift (Cheng, 1995) ve Camshift (Bradski, 1998) algoritmaları nesne takibi için basitleştirilmiş bir çözüm sunmaktadır. Başlıca yapısal verilerden ziyade renk verileriyle çalışan bu

metotlar, nesneleri doğru koşullar altında verimli bir şekilde izleyebilmektedir. Meanshift algoritması aslında kümeleme ile ilgili daha geniş bir istatistiksel kavramdır. Diğer kümeleme algoritmalarında olduğu gibi, meanshift algoritması da veri kümesinde yüksek konsantrasyondaki veri noktaları veya kümeleri olan noktaları aramaya çalışır. Algoritma, her veri noktasına bir çekirdek yerleştirir ve bir Çekirdek Yoğunluğu Tahmini (Kernel Density Estimation, KDE) gerçekleştirmek için bunları bir araya toplar. KDE, sırasıyla yüksek ve düşük veri yoğunluğuna sahip yerlere karşılık gelen tepelere sahip olur. Model, veri noktalarının bir kopyasını alır ve noktaları küçük bir miktar ile tek iterasyonda en yakın KDE zirvesine kaydırır. Model, noktalar hareketsiz kalıncaya kadar noktaları kaydırma adımlarına devam eder. Bu, verileri basit hale getirmeye ve kümeleri daha belirgin duruma getirmeye yardımcı olmaktadır.

Meanshift algoritmasında görsel, HSV renk uzayına dönüştürülür. Görsel HSV renk uzayına dönüştürüldüğünde histogramı elde edilebilir ve görsel üzerindeki keskin farklar (renk değişimi gibi) rahatlıkla tespit edilebilir. KDE yapılırken, bir gökkuşağı örneğinde ilgili renklerin olduğu kısımlarda zirve oluşacaktır. Noktaları kaydırma iterasyonlarında, renkler arasındaki ince solma ortadan kaldırılmış ve gökkuşağının her rengi daha belirgin hale gelmiş olur. Genel olarak toparlanacak olursa, gökkuşağındaki belirli bir rengin tanımlanması gerekiyorsa, meanshift algoritması ile histogram oluşturulduğundan bu işlem daha rahat görülür. Eğer rengin izlenmesi gerekiyorsa, bir gökkuşağı videosunun her bir karesi gerçek zamanlı olarak meanshift algoritması ile takip edilebilir. Meanshift, video çerçeveleri üzerinde nesne takibi yaparken nesnenin çerçeve içerisindeki mesafe bilgisini değerlendirmemektedir. Birkaç kare sonrasında takip edilen nesne uzaklaşabilir ya da daha yakın hale gelebilir. Tespit edilen nesne çerçevesi, Camshift algoritması ile güncellenir. Camshift algoritması öncelikle meanshift uygular. Meanshift ile yeteri kadar yakınsama gerçekleşince pencere boyutunu matematiksel olarak günceller. Ayrıca kendisine en uygun elipsin yönünü de hesaplar. Yine yeni ölçekli arama penceresi ve önceki pencere konumu ile ortalama kaymasını (meanshift) uygular. İşlem, gerekli doğruluk sağlanıncaya kadar devam eder.

Meanshift ve Camshift'in çalıştırılması için derin sinir ağları ya da diğer nesne takip yöntemlerinden farklı olarak herhangi bir ön eğitim sürecine ihtiyaç duyulmaz. Bir arabanın izlenmesi gerekiyorsa, modele arabaya ait farklı renk ve boyuttaki etiketli

yüzlerce görüntüsünü vermeye gerek yoktur. Algoritma, arabanın ilk renk girişini analiz eder ve onu takip etmeye başlar. Ancak, ortalama kaydırma ve cam kaydırma yalnızca renk analiziyle çalışır. İzlenen objenin renk ya da doku bilgisi fark edilir biçimde değişiyorsa, nesnenin takibi zorlaşacaktır. Aynı şekilde, bir görüntü veya videonun fazla renk kombinasyonu olan gürültülü bir arka plana sahipse, izlenen diğer nesneler ile çakışabilir. Dolayısıyla; meanshift ve Camshift, en iyi şekilde kontrollü ortamlarda çalışacaktır. Örneğin, bir depoda veya montaj hattında tek tip renkli ambalaj ürününün izlenmesi bu algoritmalar için en iyi çalışma ortamlarından birisi olabilir.

Geleneksel nesne takibi uygulamalarından bir diğeri de Kalman Filtresidir (Welch ve Bishop, 1995). Kalman filtresi, karesi alınmış hatanın ortalamasını (Mean squared error, MSE) en aza indirecek şekilde bir sürecin durumunu tahmin etmek için verimli bir hesaplamalı (özyinelemeli) araç sağlayan bir dizi matematiksel denklemdir. Filtre birkaç açıdan oldukça etkilidir: geçmiş, şimdiki ve gelecekteki durumların tahminlerini destekler ve modellenen sistemin kesin doğası bilinmemesine rağmen bunu yapabilir. Kalman filtresi, bir geri besleme kontrolü formunu kullanarak bir süreci tahmin eder. Filtre, belirli bir zamanda sürecin durumunu tahmin eder ve daha sonra gürültüye sahip ölçümleri geri bildirim olarak alır. Kalman filtreleri için denklemler; zaman güncelleme denklemleri ve ölçüm güncelleme denklemleri olmak üzere ikiye ayrılır. Zaman güncelleme denklemleri, bir sonraki zaman adımı için a priori tahmin elde etmek için mevcut durum ve hata kovaryans tahminlerini ileriye (zamanda) yansıtmaktan sorumludur. Ayrıca bir sonraki zaman adımı için öncelikli tahmin elde etmek için hata kovaryans tahminlerinden sorumludur. Geri beslemeden ise ölçüm güncelleme denklemleri sorumludur. Zaman güncelleme denklemleri aynı zamanda öngörücü denklemler olarak da düşünülebilirken, ölçüm güncelleme denklemleri düzeltici denklemler olarak düşünülebilir (Gunjal ve ark., 2018).

2.2.1.1. OpenCV ile Nesne Takibi

OpenCV, nesne veya hareket izlemenin ihtiyaçları ve amaçları için özel olarak optimize edilmiş çok sayıda gömülü (yerleşik) algoritmaya sahip olduğundan, nesne takibi aracı olarak popüler bir yöntemdir. Belirli OpenCV nesne izleyicileri, BOOSTING, MIL, KCF, CSRT, MedianFlow, TLD, MOSSE ve GOTURN izleyicilerini içerir. Bu izleyicilerin her biri farklı hedefler için en iyisidir. Örneğin, CSRT, kullanıcı daha yüksek bir nesne izleme

doğruluğu gerektirdiğinde ve daha yavaş FPS çıktısını tolere edebildiğinde en iyisidir. Bir OpenCV nesne izleme algoritmasının seçimi, o belirli izleyicinin avantaj ve dezavantajlarına ve faydalarına bağlıdır. KCF obje takibi, CSRT ile karşılaştırıldığında yüksek doğrulukta çalışmaz fakat nispeten daha yüksek FPS sağlar. MOSSE ile obje takibi oldukça hızlı olmasına karşın genel doğruluğu KCF ile nesne takibinden bile daha düşüktür. Buna karşın, OpenCV ile çalışılması gerekiyorsa, en hızlı nesne izleme yöntemi MOSSE olacaktır. GOTURN izleyici, OpenCV ile derin öğrenme tabanlı nesne takibi için tek modeldir. GOTURN'un orijinal uygulaması Caff'e'dedir. OpenCV takip API (Application Programming Interface, API)'sine uyarlanmıştır.

2.2.1.2. MATLAB ile Nesne Takibi

MATLAB, OpenCV ve derin öğrenme tabanlı diğer yöntemlere kıyasla implementasyon kısmı biraz daha farklı olan bir sayısal bilgi işlem platformudur, ancak yine de görsel izleme görevleri için iyi bir seçim olacaktır. MATLAB'deki Computer Vision Toolbox (URL-3), tek bir nesneyi izlemek veya daha karmaşık bir izleme sisteminde yapı taşları olarak kullanmak için sürekli uyarlanabilir Meanshift, CAMShift ve Kanade-Lucas-Tomasi (KLT) gibi video izleme algoritmalarını barındırmaktadır.

2.2.2. Derin Öğrenme Tabanlı Nesne Takibi

Nesne takibi, genellikle bir video olmak üzere bir dizideki farklı nesnelerin yörüngesinin takibinden ibarettir. Son yıllarda Derin Öğrenmenin (Deep Learning, DL) yükselişi ile bu soruna çözüm sağlayan algoritmalar derin modellerin temsil gücünden faydalanmıştır. Son yıllarda eğitim veri setlerindeki artış ve önerilen derin öğrenme modelleri ile elde edilen nesne tespit ve takip başarısı ile birlikte derin öğrenme tabanlı nesne takibi uygulamaları da popülerlik kazanmıştır. Derin Sinir Ağlarının (DSA) gücü, girdiye ait zengin özellikleri öğrenme ve onlardan karmaşık ve soyut özellikler çıkarma yeteneklerinde yatar. Derin öğrenme yöntemleri; uzaysal desen çıkarma, nesne tespiti, görüntü sınıflandırma, ses sinyallerinin işlenmesi, metinlerin işlenmesi gibi görevlerin çoğunda yüksek performansa ulaşabilen modeller oluşturabildiğinden, tek nesne takibi (single object tracking, SOT) ve çoklu nesne takibi (multiple object tracking, MOT) uygulamalarında da derin ağların kullanımı popüler hale gelmiştir. MOT; yayalar, arabalar, hayvanlar ve cansızlar gibi bir veya daha fazla kategoriye ait nesneleri tanımlamak ve izlemek için videoları analiz etmeyi amaçlayan bir bilgisayarlı görü

görevidir. Nesne takibi algoritmaları, tespit edilen nesneyi içine alacak şekilde bir kare ya da dikdörtgen çizen nesne tespiti algoritmalarından farklı olarak, sınıf içi nesneleri ayırt etmek için her kutuya da bir hedef kimliği atar ve bu şekilde farklı sınıflara ait nesneleri gruplandırmış olur (Ciaparrone ve ark., 2020).

Çoğu MOT algoritması, tespit yoluyla izleme (tracking-by-detection) adı verilen bir yaklaşımı içerir. Tespit yoluyla izleme yöntemi, olası tespitleri elde etmek için tüm görüntü çerçevelerine uygulanan bağımsız bir detektörü ve ardından tespit setinde çalıştırılan bir izleyiciyi içerir. Böylelikle, izleyici veri ilişkilendirmesi gerçekleştirmeye çalışır (örneğin, tam yörüngeleri elde etmek için algılamaları birbirine bağlar). Video girdilerinden çıkarılan tespitler, onları bağlamak ve aynı hedefi içeren sınırlayıcı kutulara özdeş kimlikler atamak suretiyle izleme sürecini yönlendirmek için kullanılır. MOT algoritmalarında toplu izleme ve çevrimiçi izleme metotları olmak üzere iki yöntem mevcuttur. Toplu izleme algoritmaları, belirli bir karedeki bir nesnenin kimliğini çıkarırken gelecekteki video karelerinden gelen bilgileri kullanır. Toplu izleme algoritmaları, nesneyle ilgili yerel olmayan bilgileri kullanır. Bu metodoloji, daha iyi bir izleme kalitesi sağlar. Toplu izleme algoritmaları ilerideki çerçevelere erişirken, çevrimiçi izleme algoritmaları belirli bir çerçeveye ilgili sonuçlara varmak için yalnızca mevcut ve geçmiş bilgileri kullanır.

MOT gerçekleştirmek için çevrimiçi izleme yöntemleri, mevcut çerçeveye sınırlı kalan çevrimiçi yöntemlerin sınırlaması nedeniyle genellikle toplu yöntemlerden daha kötü performans gösterir. Navigasyon veya otonom sürüş gibi nesnelerin izlenmesini gerektiren gerçek zamanlı problemler, gelecekteki video karelerine erişime sahip olmadığından çevrimiçi izleme yöntemleri hala geçerli bir çözümdür. Çoğu MOT algoritması, algoritmalar değişikçe sabit kalan temel bir adım kümesi içerir. MOT algoritmalarının çoğu; tanımlama ve tespit (designation or detection), hareket (motion) ve recall olmak üzere üç genel adımı takip eder. Tanımlama aşamasında ilgilenilen hedefler not edilir ve vurgulanır. Algoritma, hedef sınıflara ait nesneleri tanımlamak için giriş çerçevelerini analiz eder. Algoritmanın bir parçası olarak algılamaları gerçekleştirmek için sınırlayıcı kutular kullanılır. Özellik çıkarma algoritmaları, görünüm ve etkileşim özelliklerini çıkarmak için tespitleri analiz eder. Çoğu durumda, izlenen her bir hedefin sonraki konumlarını tahmin etmek için bir hareket tahmincisi kullanılır.

Özellik tahminleri, tespit çiftleri arasındaki benzerlik puanlarını hesaplamak için kullanılır. Bu puanlar daha sonra aynı hedefe ait tespitleri ilişkilendirmek için kullanılır. Kimlikler benzer tespitlere atanır ve çiftlerin parçası olmayan algılamalara farklı kimlikler uygulanır. Bazı nesne izleme modelleri, bu adımları birbirinden ayrı kullanılarak oluşturulurken, diğerleri adımları birleştirir ve birlikte kullanır. Algoritma işlemedeki bu farklılıklar, bazılarının diğerlerinden daha doğru olduğu benzersiz modeller oluşturmaktadır.

CNNler, nesne takibi için en çok kullanılan ve güvenilir ağ olmaya devam etmektedir. Bununla birlikte, çoklu mimariler ve algoritmalar da araştırılmaktadır. Bu algoritmalar arasında RNN'ler, Otomatik Kodlayıcılar (Autoencoder, AE), Çekişmeli üretici ağ (Generative adversarial network, GAN), Siyam Sinir Ağları (SNN'ler) ve özel sinir ağları bulunur. Nesne tespiti modelleri, videolarda her bir kare için uygulandığında nesneleri izlemek için kullanılabilse de bu işlem, hesaplama açısından sınırlayıcı ve dolayısıyla nesne izlemeyi gerçekleştirmek için oldukça verimsiz bir yöntemdir. Bunun yerine, nesne algılama bir kez uygulanmalıdır ve ardından nesne izleyici birinciden sonraki her kareyi işleyebilir. Bu, nesne izlemeyi gerçekleştirmek için hesaplama açısından daha etkili ve daha az hantal bir süreçtir (Leal-Taixé, 2014).

2.2.2.1. Deep SORT ile nesne takibi

Deep SORT (Wojke ve ark., 2017), başarılı bir nesne izleme modelidir ve en yaygın kullanılan nesne izleme çerçevelerinden biridir. Nesnelerin görünüş bilgileri, Deep SORT modeline entegre edilmiştir ve bu, modelin performansını büyük ölçüde artırmıştır. Entegrasyon nedeniyle, takip edilen nesnelerin uzun süre boyunca üst üste nesne gelmesi durumunda (occlusion sorunu), uzun süre boyunca izlenebilir, bu da modelin nesneler üzerinde sürekli tespit numarasını (ID) değiştirmesinin önüne geçer. Deep SORT ile Kalman Filtresi ile yapılan her bir tespit için, gerekli tüm durum bilgisine sahip bir “İz” oluşturulur. Bu noktadan sonra, Kalman filtresinden izlenen yeni sınırlayıcı kutular (bounding box) oluşturulunca bir sonraki sorun, yeni tespitleri yeni tahminlerle ilişkilendirmekte yatmaktadır. Bağımsız olarak işlendiklerinden burada yeni tespitleri, yeni tahminlere atama sorunu ortaya çıkmaktadır. Bu sorunun çözülmesi için, ilişkiyi ölçmek için bir mesafe metriği ve verileri ilişkilendirmek için verimli bir algoritma gerekir.

Deep SORT arařtırmacıları, Kalman filtresinden gelen belirsizlikleri dahil etmek için kare Mahalanobis mesafesini (squared Mahalanobis distance) kullanmaya karar vermiřlerdir. Bu mesafeyi eřik deęer olarak tutmak, gerek iliřkiler hakkında iyi bir fikir verebilir. Bu metrik, iki daęılım arasındaki mesafenin etkin bir řekilde ldğnden dolayı klid mesafesinden daha doęru sonu retmektedir. Bu noktaya kadar Deep SORT'ta; nesne tespiti yapan bir algoritma, bu tespitleri takip eden Kalman Filtresi ve filtrenin kaırdıęı takipler ve tespit ile takipleri iliřkilendiren Macar algoritması (Hungarian algorithm) mevcuttur. Kalman filtresi verimli alıřmasına raęmen; kapanmalar (bir nesnenin, arkadaki nesneyi kapatması), farklı bakıř aıları vb. gibi gerek dnya senaryolarının oęunda bařarısız olabilmektedir. Bu handikapların nne geebilmek adına Deep SORT yazarları, nesnenin "grnmne" dayalı bařka bir mesafe metrięi tanıtmıřtır. Bu amala yazarlar, nce veri kmesi zerinde bir sınıflandırıcı oluřturdular. Sınıflandırıcıyı olduka iyi bir doęruluk elde edene kadar eęittiler ve ardından son sınıflandırma katmanını ıkardılar. Klasik bir mimariye gre, tek bir znitelik vektr reten ve sınıflandırılmayı bekleyen dense (yoęun) bir katman kalır. Bu zellik vektr de nesnenin "grnm tanımlayıcısı" olur. Bu alıřmada arařtırmacılar, nceden eęitilmiř bir iliřkilendirme metrięi aracılıęıyla grnm bilgilerini ieren bir SORT (Cao ve ark., 2017) uzantısını nermiřlerdir. Bu alıřma ile, nesnelerin daha uzun kapanma dnemleri takip edilebilmekte ve bu da SORT'u son teknoloji evrimii izleme algoritmalarına karřı gl bir rakip haline getirmektedir. Bu eklemelere karřın algoritmanın uygulanması halen basittir ve gerek zamanlı olarak alıřmaktadır.

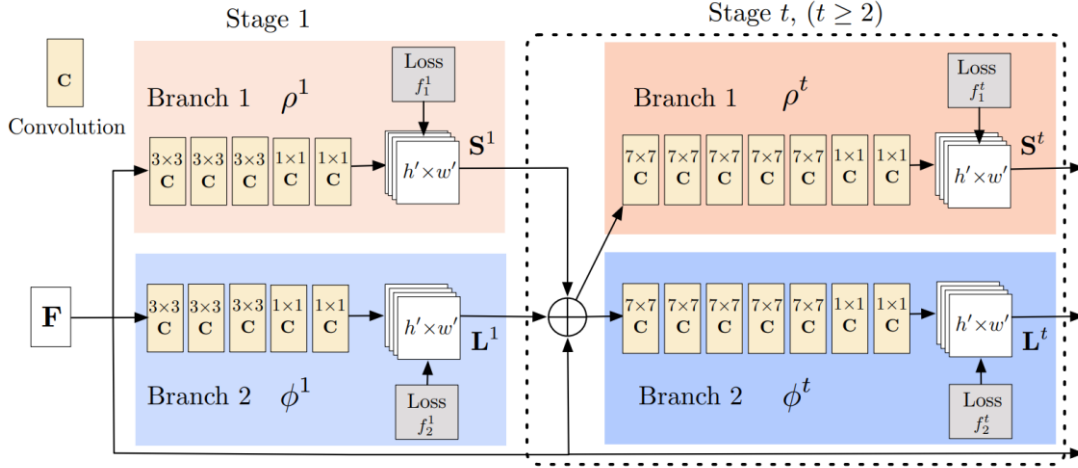
3. MATERYAL VE METOTLAR

Bebeklerin beden hareketlerinin; OpenPose, AlphaPose ve KAPAO tabanlı açık kaynak insan iskelet tespit sistemleri ile birlikte çözülmesi hususunda internet üzerinden çeşitli içerik platformlarından elde edilmesi suretiyle BabyPose adını verdiğimiz veri seti oluşturulmuştur. BabyPose veri seti kullanılarak video çerçevelerinden OpenPose, AlphaPose ve KAPAO ile vücut eklem ve uzuv bilgilerini içeren özellikler çıkarılmıştır. Bu özellikler, aktivite tanıma amacıyla oluşturulan LSTM derin öğrenme modeline girdi olarak verilmiştir. LSTM tekrarlı sinir ağı kullanılarak bebeklerin aktiviteleri sınıflandırılmıştır. Bu bölümde BabyPose veri setinden özneteliklerin çıkarıldığı, HPE sistemleri hakkında bilgiler verilmiş ve oluşturulan derin öğrenme modeline değinilmiştir. HPE çıktılarının oluşturulan derin öğrenme modeline girdilerinden bahsedilmiş, modelin daha başarılı sonuç vermesi adına çıkarılan özelliklere değinilmiştir. Ardından gerçek zamanlı nesne takip sistemi ile bebeklerin video çerçevesi içerisinde yapılan takibinin detaylarına yer verilmiştir.

3.1. OpenPose Tabanlı Poz Tahmini

OpenPose, Carnegie Mellon Üniversitesi araştırmacıları tarafından geliştirilmiş açık kaynak kodlu poz tahminleme kütüphanesidir (Cao ve ark., 2017). OpenPose, çok kişili insan pozu tahmini için en yaygın olarak kullanılan aşağıdan yukarıya yaklaşımlardan biridir. OpenPose, diğer aşağıdan yukarıya yöntemlerine benzer olarak görüntüdeki her bir kişiye ait anahtar noktaları tespit eder ve ardından parçaları ayrı bireylere atayarak tespit işlemini yapar. OpenPose işleyişi için iki farklı kurulum mevcuttur. İlk yayında kurulan mimaride OpenPose ağı, ilk birkaç katmanı kullanarak bir görüntüden özellikler toplayarak başlar. Özellikler daha sonra birbirine paralel çalışan iki evrimsel katman dalına gönderilir. İlk dal, her biri insan pozu iskeletinin farklı bir bölümünü temsil eden 18 güven haritasını tahminlemektedir. İkinci dal, uzuvlar arasındaki ilişkinin derecesini tanımlamak için kullanılan bir dizi 38 Uzuv İlişki Alanlarını (Part Affinity Fields, PAF) tahminler. Her dalın tahminleri, birbirini izleyen aşamalarda rafine edilir. Parça güven haritaları kullanılarak parça çiftleri arasında ikili grafikler oluşturulur. İki parçalı grafiklerdeki daha zayıf bağlantılar, PAF değerleri kullanılarak kırılır (Cao ve ark., 2017). İnsan konumu iskeletleri, yukarıda özetlenen prosedürler kullanılarak görüntüdeki

her bir kişiye göre atanabilmektedir. OpenPose için önerilen ilk mimari, Şekil 3.1’de verilmiştir.



Şekil 3.1. OpenPose'un Önerilen Mimarisi: İki dallı çok aşamalı CNN (Cao ve ark., 2017)

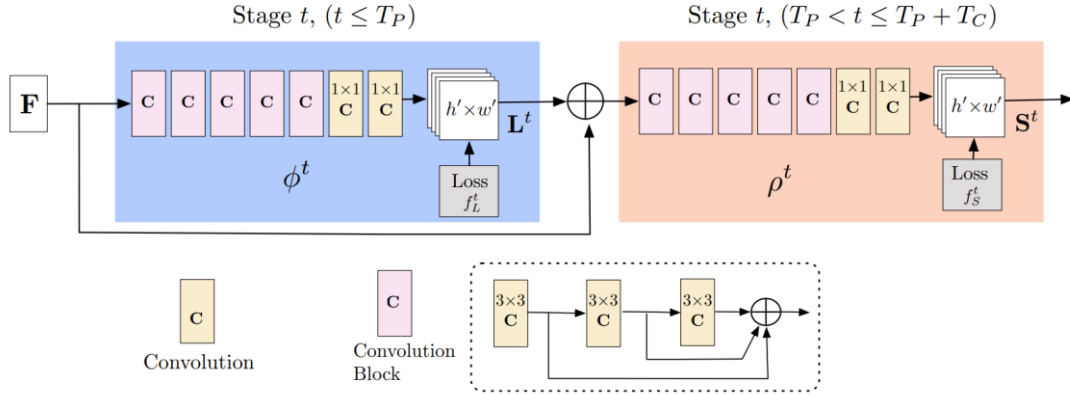
Şekil 3.1’de OpenPose ilk yayındaki (Cao ve ark., 2017) OpenPose mimarisi verilmiştir. İlk daldaki her aşama güven haritalarını (S^t) tahmin etmekte ve ikinci daldaki her aşama PAF'leri (L^t) tahmin etmektedir. Her aşamadan sonra, iki daldan gelen tahminler, görüntü özellikleriyle birlikte bir sonraki aşama için birleştirilir. Mimaride her bir görüntü için 18 eklem bölgesi koordinatları (COCO çıktısı) elde edilir. Bu yaklaşım, COCO 2016 anahtar nokta yarışmasında birinci gelmiştir. OpenPose ile elde edilen iskelet çıktı bilgisi, Denklem (3.1)’de verilmiştir.

$$A = P_0^{(x)}, P_0^{(y)}, P_0^{(acc)}, \dots, P_{17}^{(x)}, P_{17}^{(y)}, P_{17}^{(acc)} \quad (3.1)$$

Burada A , ilgili çerçeve için 18 anahtar nokta koordinatlarını ve doğruluk oranını tutan 54 özellik barındırmaktadır. $P_0^{(x)}, P_0^{(y)}$ İlgili anahtar noktanın görsel üzerindeki (x, y) koordinatıdır. $P_0^{(acc)}$ Değeri ise tespit edilen her bir uzuv için verilir. Bu değer, tespit edilen uzvun tespit güven değeridir. Ne kadar doğrulukla uzvun tespit edildiği noktasında fikir verebilmektedir.

2019 Yılında yayınlanan yeni bir çalışma (Cao ve ark., 2019) ile ilgili çalışma OpenPose adını almış ve İki dallı çok aşamalı CNN mimarisi değiştirilmiştir. Önerilen yeni mimaride OpenPose, tam görüntüyü vücut uzuvlarını tanımak için güven haritalarını ve uzuv ilişkilendirmesi için Parça Yakınlık Alanlarını (Part Affinity Fields, PAF) ortaklaşa

tahmin etmek için bir CNN'ye girdi olarak kullanır. Vücut uzuv adaylarını ilişkilendirmek için, ayrıştırma aşaması bir dizi ikili eşleştirme gerçekleştirir. Son olarak tespit edilen bu vücut uzuvları, görüntüdeki tüm insanlar için tam vücut duruşlarında bir araya getirilerek iskelet yapısını meydana getirir (Cao ve ark., 2019). Yeni mimariye ait genel işleyiş şeması Şekil 3.2'de verilmiştir.



Şekil 3.2. OpenPose çok aşamalı CNN mimarisi (Cao ve ark., 2019).

Şekil 3.2'deki yeni OpenPose mimarisi ile mimarideki ilk aşama grubu, PAF'leri (L^t) tahmin ederken, son grup, (S^t) güven haritalarını tahminlemektedir. Her aşamanın tahminleri ve bunlara karşılık gelen görüntü özellikleri, sonraki her aşama için birleştirilir. Orijinal yaklaşımdaki (Cao ve ark., 2017) 7 çekirdek boyutlu evrişimli katmanlar, 3 adet 3x3 boyutlu evrişimli katmanlar ile değiştirilmiştir.

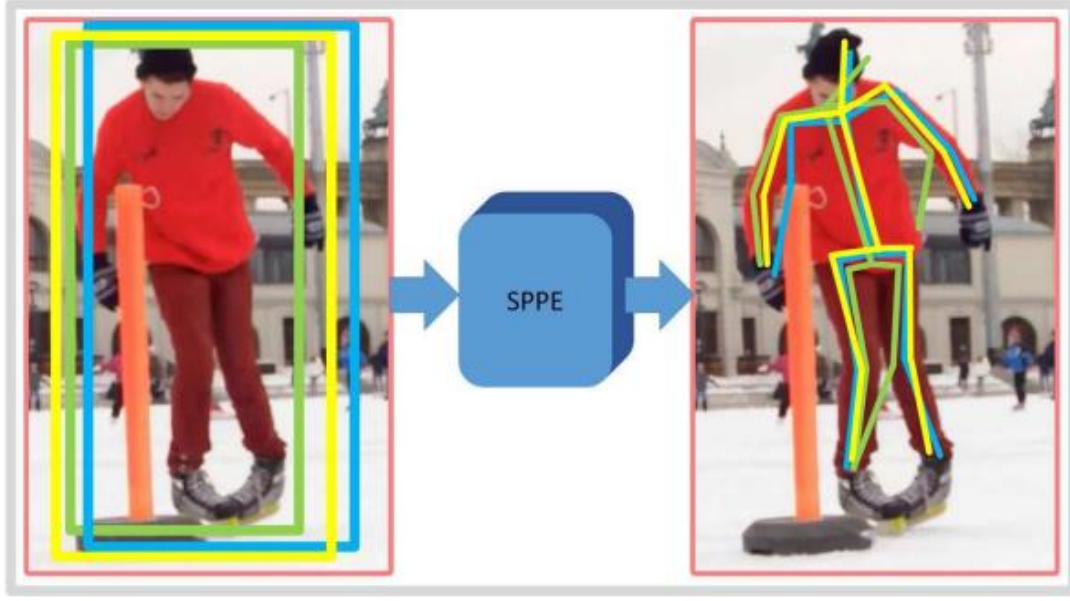
Bu çalışmada her bir video karesi için güncel OpenPose çalışması içerisinde yer alan BODY_25 (25 uzuv çıktısı veren OpenPose modeli) ile toplanan videolar etiketlenmiştir. BODY_25 modeliyle her bir görüntü için 25 uzva ait; x, y ve doğruluk bilgisi olmak üzere toplamda 75 özellik çıkarılır. En iyi performansı ve verimi alabilmek adına, Tablo 3.1'de yer alan parametreler ile OpenPose çalıştırılmıştır. 20-25 özellikleri ayaklar, ayak parmakları gibi özelliklerdir. Bu özelliklerin ilgili çalışmamızda aktivite sınıflandırma hususunda bir yararı olmayacağı için çıkarılmıştır. Tablo 3.1'de video çerçevelerinin OpenPose ile etiketlenmesi için kullanılan parametrelere değinilmiştir. OpenPose; özel parametreler ile daha doğru sonuçlar üretse de, GPU belleğini oldukça fazla tüketmektedir. BabyPose veri setinin etiketlenmesi için belirlenen OpenPose parametreleri, GPU belleğinin 10 gigabyte kapasitesi göz önünde bulundurularak belirlenmiştir.

Tablo 3.1. OpenPose HPE ile etiketleme işleminde kullanılan parametreler

OpenPose Parametre	Parametre Açıklaması
--net_resolution -1x368	-1x368 ile 656x368 boyutunda girdiler alınır. Boyut artırılırsa daha fazla GPU belleğe ihtiyaç duyulur.
--scale_number 4	Ortalama ölçek sayısını belirtir.
--disable_blending	Anahtar noktaları veya ısı haritalarını siyah bir arka planda oluşturur.
--display 0	Görüntü modu. 0 ile etiketleme işlemi gösterilmez, daha hızlı çalışır.
--keypoint_scale 3	Tespit edilen uzuvlara ait x ve y koordinatlarını (0, 1) arasında normalize eder.
--write_images *dest_folder*	Tespit edilmiş iskeletli görüntüleri belirtilen dosyaya kaydeder

3.2. AlphaPose Tabanlı Poz Tahmini

AlphaPose (Fang ve ark., 2017), Fang ve arkadaşları tarafından geliştirilen çok kişili bir poz tahmini modelidir. İlgili çalışmada yazarlar, hatalı sınırlayıcı kutuların (bounding box) çokça bulunduğu etiketli verilerde poz tahminini kolaylaştırmak için yeni bir bölgesel çok kişili poz tahmini (Regional Multi-person Pose Estimation) çerçevesi önermişlerdir. Herhangi bir ortamdaki birçok insanın poz tahminlemesini gerçekleştirebilmek için genelde iki yaklaşım bulunmaktadır. Bu yaklaşımlardan ilki, iki aşamalı çerçevedir. Bu yaklaşımda model önce insan sınırlayıcı kutuları tespit eder ve her kutunun içindeki pozu birbirinden bağımsız olarak tahminler. Diğer bir yöntem olan parça tabanlı çerçevede ise, öncelikle vücut parçalarını bağımsız olarak tespit eder ve algılanan vücut parçalarını birden fazla insan pozu oluşturmak için birleştirir. Her iki yaklaşımın kendisine göre avantajı ve dezavantajı bulunmaktadır. İlk yaklaşımda, poz tahmininin doğruluğu, tespit edilen sınırlayıcı kutuların doğruluğuna büyük ölçüde bağlıdır. İkinci çerçevede ise, yan yana ya da çok yakın insanların tespit edilen uzuv bilgileri üst üste binmekte, insan-uzuv eşleştirmesi güçleşmektedir. Şekil 3.3'te araştırmacılar, iki aşamalı çerçeve yönteminin tek bir poz için birbirinden bağımsız birden fazla sınırlayıcı çerçeve tespit ettiğini, bunun da tek bir insan üzerinde birbirinden farklı poz tahminlerinin bulunmasına yol açtığını göstermişlerdir. Şekilde soldaki görsel tespit edilen sınırlayıcı kutuları; sağdaki görsel ise bu sınırlayıcı kutular içerisindeki tespit edilen insan pozlarını gösterir. Her bir sınırlayıcı kutu birbirinden bağımsız işlendiği için, gelenek geleneksel yaklaşımlarda bir insan için birden çok poz tahmini gerçekleşir (Fang ve ark., 2017).

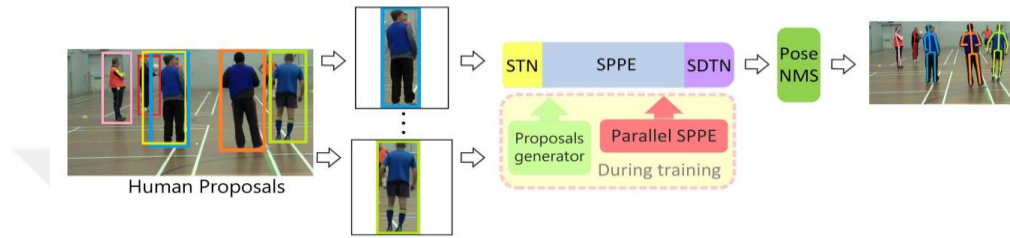


Şekil 3.3. SPPE uygulamalarında birden fazla sınırlayıcı kutu tespit problemi (Fang ve ark., 2017)

SPPE (Single Person Pose Estimation) uygulamalarında çoklu sınırlayıcı kutu tahminleme probleminin dışında, sınırlayıcı kutu yerleştirme hatası sorunları da mevcuttur. Sınırlayıcı kutuların IoU (Intersection over Union) $> 0,5$ ile doğru kabul edildiği durumlarda bile, tespit edilen insan pozları yine de yanlış olabilmektedir. SPPE, verilen her sınırlayıcı kutu için bir poz ürettiğinden, fazladan tespitler, fazlalık pozlar ya da yanlış poz tahminleri ile sonuçlanır. IoU, belirli bir veri kümesindeki bir nesne tespit modelinin doğruluğunu ölçmek için kullanılan bir değerlendirme metriğidir.

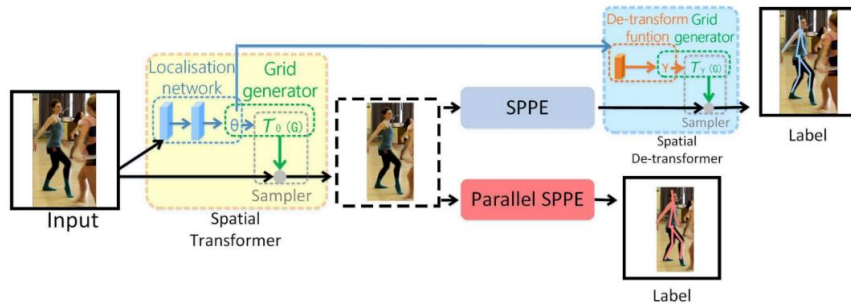
AlphaPose, COCO veri setinde 70+ mAP (mean Average Precision) ve MPII (Max Planck Institute for Informatics) veri setinde 80+ mAP (82.1 mAP) elde eden ilk açık kaynaklı sistemdir. Bahsedilen yukarıdan aşağıya insan poz tahmini yaklaşımlarının (iki aşamalı çerçeve) problemlerini çözmek adına araştırmacılar, hatalı bir sınırlayıcı kutudan yüksek kaliteli tek kişilik bir bölge çıkarmak için Simetrik Uzamsal Transformatör Ağı'nı (Symmetric Spatial Transformer Network, SSTN) önermişlerdir. Ardından bu çıkarılan bölgede poz tahmini için SPPE kullanılır. Tahmini insan pozunu orijinal görüntü koordinat sistemine geri döndürmek için ise, önerilen Uzamsal De-Transformer Ağı (SDTN) kullanılır. Son olarak, gereksiz poz kesintileri sorununu ele almak için bir parametrik poz Maksimum-olmayan bastırma (NMS) tekniği kullanılır. Ayrıca yazarlar, SPPE ve SSTN ağlarını daha iyi eğitmeye yardımcı olabilecek eğitim örneklerini artırmak

için bir Poz Kılavuzlu Öneri Oluşturucu (PGPG) sunar. Şekil 3.4'te AlphaPose çalışmasında önerilen metotlar, bir boru hattı olacak şekilde görselleştirilmiştir. Simetrik STN, SPPE'den önce ve sonra eklenen STN ve SDTN'den oluşur. STN, insan önerilerini alır ve SDTN, poz önerilerini oluşturur. Paralel SPPE, eğitim aşamasında ekstra bir düzenleyici işlevi görür. Son işlemde, gereksiz poz tahminlerini ortadan kaldırmak için parametrik Poz NMS gerçekleştirilir. Araştırmacılar geleneksel model eğitiminden ayrı olarak, SSTN+SPPE modülünü PGPG tarafından oluşturulan görüntülerle eğitmişlerdir.



Şekil 3.4. Önerilen RMPE çerçevesinin boru hattı (Fang ve ark., 2017)

Max Jaderberg ve diğerleri (Jaderberg ve ark., 2015) tarafından tanıtılan Uzamsal Transformatör Ağları (SDN), bir modelin öteleme, ölçekleme, döndürme ve kırpma gibi uzamsal dönüşümlere karşı uzamsal değişmezliğini arttırmanın popüler bir yoludur. SDN'ler mevcut derin ağlara ya girdinin hemen ardından ya da derin katmanlarda eklenebilir. Girdileri standart, beklenen bir poza dönüştürerek uzamsal değişmezliği elde eder, böylece daha iyi sınıflandırma performansına ulaşılmasını sağlarlar. SDN'ler, standart geri yayılım kullanılarak uçtan uca eğitebilir. AlphaPose çalışmasında da poz tahminleri üzerinde daha önceden bahsedilen lokalizasyon ve sınırlayıcı kutu hatalarının önüne geçmek amacıyla SDN'ler kullanılmıştır. Şekil 3.5'te çalışmada kullanılan simetrik SDN mimarisi verilmiştir. SDTN, yerelleştirme ağı tarafından oluşturulan bir θ parametresini alır ve geri-dönüşüm için γ değerini hesaplar (Fang ve ark., 2017).



Şekil 3.5. Önerilen SSTN mimarisi ve paralel SPPE ile eğitim stratejisi (Fang ve ark., 2017)

AlphaPose çalışmasında insan tespiti için VGG-tabanlı SSD-512 (Liu ve ark., 2016) nesne tespit modeli kullanılmıştır. Üstün performansı nedeniyle SPPE için de yığılmış kum saati ağını (stacked hourglass network) (Newell ve ark., 2016) kullanmışlardır. Tablo 3.2’de video çerçevelerinin AlphaPose ile etiketlenmesi için kullanılan parametrelere değinilmiştir.

Tablo 3.2. AlphaPose HPE ile etiketleme işleminde kullanılan parametreler

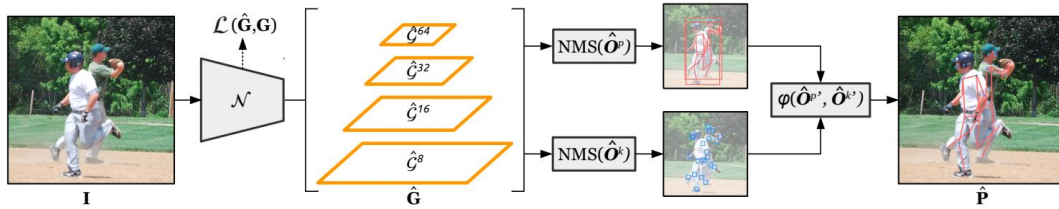
AlphaPose Parametre	Parametre Açıklaması
--sp	Windows cihazlarda verilmesi önerilen parametre.
--vis_fast	Daha hızlı tahminleme gerçekleştirir.
--format open	Çıktı formatını OpenPose formatında ayarlar.
--fast_inference False	Tahminlemeyi yavaşlatarak daha tutarlı çıktılar elde etmek için verilir.
--save_video	İskelete sahip AlphaPose çıktısını yerel diske kaydeder.

3.3. KAPAO Tabanlı Poz Tahmini

McNally ve arkadaşları tarafından geliştirilen KAPAO (McNally ve ark., 2021), ısı haritası içermeyen yeni bir anahtar nokta tespit yöntemini tanıtmış ve bunu tek aşamalı çok kişili insan pozu tahminine uygulamıştır. Genel yaklaşımda, anahtar nokta nesnelerini ve poz nesnelerini eşzamanlı olarak tespit eder. Her iki nesne temsilinin faydalarından yararlanmak için basit bir eşleştirme algoritması kullanarak sonuçları birleştirirler. Poz nesnelerini algılama sayesinde, kişi algılamayı ve anahtar nokta tahminini birleştirerek, çok kişili insan pozu tahminine yüksek verimli tek aşamalı bir yaklaşım sunmuşlardır.

Literatürde anahtar nokta konumlarını tahmin etmek için en yaygın yöntem, hedef anahtar nokta koordinatlarında ısı haritaları olarak adlandırılan hedef alanlar oluşturmayı içermektedir. Isı haritaları temelli yaklaşımlar çokça kullanılsa da bilinen birkaç dezavantajı vardır. İlk olarak, bu yöntemler niceleme hatasından muzdariptir; bir anahtar nokta tahmininin kesinliği, ısı haritasının uzamsal çözünürlüğü ile sınırlıdır. Bu nedenle daha büyük ısı haritaları avantajlıdır, ancak ek örnekleme işlemleri ve daha yüksek çözünürlüklerde maliyetli işlemler gerektirir. İkinci problem ise, aynı sınıftan iki anahtar nokta birbirine yakın olan bir durum ile karşılaşıldığında, örtüşen ısı haritası sinyalleri tek bir anahtar nokta ile karıştırılabilir.

Isı haritalarını kullanmamanın bir sonucu olarak KAPAO, özellikle TTA (test-time augmentation) kullanılmadığında, doğruluk ve çıkarım hızı açısından literatürdeki son tek aşamalı insan pozu tahmin modellerine kıyasla daha avantajlıdır. KAPAO, ortalama 54,4 ms gecikme süresi ile TTA olmadan Microsoft COCO Keypoints doğrulama setinde 70,6'lık bir AP'ye ulaşmıştır. KAPAO ile tek aşamalı çok kişili insan pozu tahmini için yeni bir yaklaşım, kilit nokta nesnelerini ve insan poz nesnelerini aynı anda tespit ederek ve her iki nesne temsilinin güçlü yönlerinden yararlanmak için algılamaları birleştirerek geliştirilmiştir. Isı haritalarını kullanan diğer son teknoloji metotlara kıyasla KAPAO, Microsoft COCO Keypoints doğrulama setinde diğer yöntemleri geride bırakmıştır. Şekil 3.6'da KAPAO için önerilen mimari verilmiştir. KAPAO; bir RGB görüntüsünü (I), tahminlenen poz nesnelerini ($\hat{O}^p$) ve anahtar nokta nesnelerini ($\hat{O}^k$) içeren bir çıktı ızgarasına ($\hat{G}$) haritalamak için eğitilmiş bir Dense Detection Network (DDN) ağını (N) kullanır. NMS, aday tespitleri ($\hat{O}^k$ ve $\hat{O}^p$) elde etmek için kullanılır. Bu aday tespitler ile de eşleştirme algoritması (ϕ) kullanılarak nihai insan pozu tahminleri ($\hat{P}$) elde edilmiş olur (McNally ve ark., 2021). Tablo 3.3'te video çerçevelerinin KAPAO ile etiketlenmesi için kullanılan parametrelere değinilmiştir.



Şekil 3.6. Önerilen KAPAO mimarisi (McNally ve ark., 2021)

Tablo 3.3. KAPAO HPE ile etiketleme işleminde kullanılan parametreler

KAPAO Parametre	Parametre Açıklaması
--weights kapao_l_coco.pt	KAPAO-L modelinin kullanılması adına verilen parametre
--conf-thres-kp 0.05	KAPAO-L ağırlıkları
--kp-obj	Anahtar noktaları video üzerinde çizer
--start 0	İlk video çerçevesinden başlar
--end -1	Video sonuna kadar HPE yapar
--csv	HPE sonuçlarını csv dosyasına yazar
--face	Yüz anahtar noktalarını çizer

3.4. LSTM ile Bebek Aktivitesi Tanıma

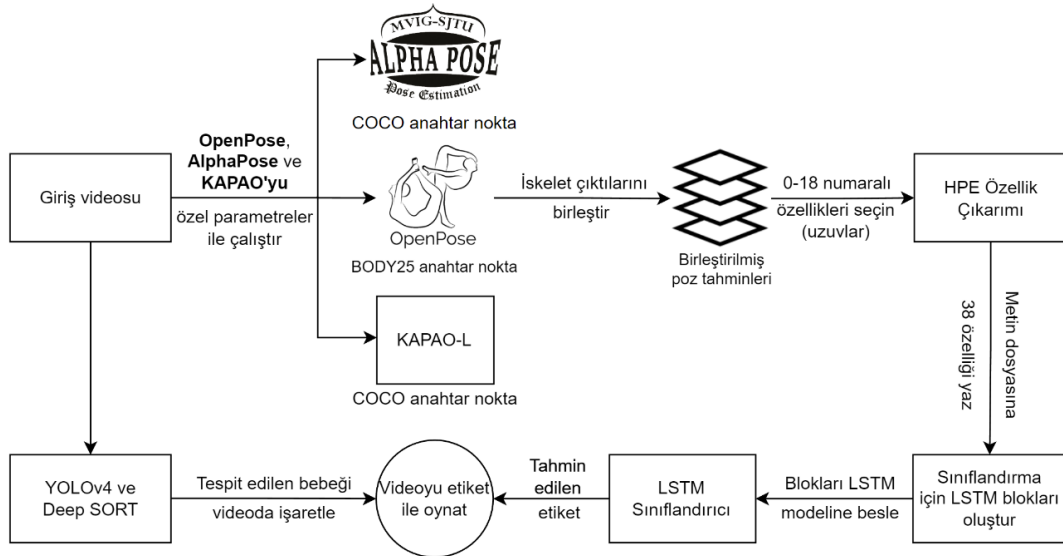
CNN'ler, bir tür ileri beslemeli yapay sinir ağıdır. Izgara benzeri bir yapıda ifade edilebilen verileri işlemek için kullanılırlar (Gu ve ark., 2018). CNN'ler, görüntüler gibi uzamsal verilerle ilgili problemlerin çözümünde yaygın olarak kullanılır. Fakat CNN'ler, videolar gibi zamansal, sıralı verileri analiz etmek için uygun değildir. Geleneksel sinir ağlarının “hafızası” yoktur. Her yeni girdi için daha önceki sonuçlar hakkında bir bilgi içermez ve geçmişteki bilginin karar yapısına etkisi yoktur. Bu durum, sıralı verileri içeren (borsa tahmini, aktivite sınıflandırma gibi) işlerde önemli bir kusurdur ve RNN'ler bu problemi gidermek amacıyla özel olarak oluşturulmuştur. Tekrarlayan sinir ağları, ağ içinde döngüler oluşturarak bilginin kalmasını sağlayan bir yapı içerisinde çalışır. Bu döngüler, bir değişkenin mevcut değerinin daha sonraki bir zaman adımında kendi değeri üzerindeki etkisini gösterir. RNN döngüleri açılabilir, bu da tekrarlanan bir sinir ağı dizisine neden olur (Graves, 2013). Bu çalışmada, 2 Yoğun (dense) katmanlı 3 Seviyeli Yığılmış (stacked) LSTM ve 2 Yoğun Katmanlı 1 LSTM hücresinden oluşan LSTM modelleri, her bir HPE modeli için oluşturulmuştur.

3.5. Deep SORT ile Gerçek Zamanlı Bebek Takibi

CNN'lere dayalı nesne algılama teknikleri, son zamanlarda doğruluk açısından klasik nesne algılama yaklaşımlarından daha iyi performans göstermiştir. Nesne algılama doğruluğu arttıkça, çoklu nesne takibi için algılama yoluyla izleme yöntemlerinin kullanımı da kolaylaşmaktadır.

Tespit yoluyla izleme (tracking-by-detection) yöntemi, nesne izlemeye yönelik önde gelen bir yaklaşımdır. Nesne algılama algoritmaları, bir çerçevedeki nesneleri algılamak için tespit yoluyla izlemede kullanılır. Bu nesneler daha sonra mevcut çerçevedeki nesneleri önceki çerçevelerdeki nesnelerle ilişkilendiren bir izleme algoritması kullanılarak izlenir. İzleme algoritması, her çerçevede tespit edilen nesnelere dayalı olduğundan, nesne tespiti için güvenilir bir mekanizmaya sahip olmak çok önemlidir. CNN'lere dayalı nesne algılama teknikleri, son zamanlarda doğruluk açısından klasik nesne algılama yaklaşımlarından daha iyi performans göstermektedir. Nesne tespit doğruluğu arttıkça, çoklu nesne takibi için algılama yoluyla izleme yöntemlerinin kullanımı kolaylaşmıştır.

Bu çalışmanın bebek takibi hususunda ana odak noktası, gerçek zamanlı bir izleme algoritması olan Deep SORT'tur. SORT, gelecekteki nesne konumlarını tahmin etmek için bir Kalman filtresi ve çerçeveden çerçeveye bağlantılar oluşturmak için Macar tekniğini (Hungarian algorithm) kullanan, amaca yönelik olarak basit bir izleme algoritmasıdır. Deep SORT, izleme aşamasında herhangi bir 'görünüm bilgisi' kullanmayan SORT algoritmasının bir uzantısı olarak önerilmiştir (Wojke ve ark., 2017). SORT ve Deep SORT'u karşılaştırmak, görsel tanımlayıcıların kullanımının izleme performansını ve buna bağlı olarak genel sistem performansını nasıl etkilediğini ortaya çıkaracaktır. Önerilen çalışmamızda Deep SORT ile birlikte nesne tespiti için YOLOv4 kullanılmıştır. Şekil 3.7'de çalışma kapsamında önerilen mimari verilmiştir.



Şekil 3.7. Bebek aktivitelerinin sınıflandırılması için önerilen mimari

4. DENEYSEL SONUÇLAR ve TARTIŞMA

Bu bölümde ilk olarak 6 kategorideki bebek aktivitelerine ait video görüntülerinin internet üzerindeki halka açık içerik paylaşım platformlarından toplanması suretiyle BabyPose veri setinin oluşturulmasına değinilmiştir. Ardından internet ortamından elde edilen bebek video görüntülerinin ön işleme, etiketleme ve sınıflandırılmasında kullanılan çalışma ortamından bahsedilmektedir. Bu çalışma ortamının nasıl kurulacağı anlatılmakta, kullanılan HPE modellerinin kurulumlarından bahsedilmektedir. Veri seti üzerinde ön işleme adımlarının icra edilmesinin ardından derin öğrenme ağına bu bilgiler verilmiş ve çeşitli yapılarda LSTM ağ modelleri eğitilmiştir. Eğitilen derin öğrenme modellerinin performans sonuçları, belirtilen performans metrikleri ile değerlendirilmiştir. Son olarak önerilen mimarinin bebek aktivitelerini tanıma hususunda ne derecede başarılı olduğu değerlendirilmiştir.

4.1. BabyPose Veri Seti

İnsan iskelet sistemi tahminleyici sistemler, uzun süredir araştırılmaktadır. Ancak, insanların kaynak video/görsel içerisindeki konumları, göreceli uzaklıkları ve boyutları değişkenlik gösterdiğinden, bu pozlar için genel bir veri kümesi oluşturmak zorlu bir süreçtir. Tahmin sorununu çözmek adına veri kümeleri boyut ve karmaşıklık açısından büyüme eğilimindedir. Derin öğrenme temelli yaklaşımlar ise eğitim için görüntü sayısı yetersiz olduğundan bu veri kümeleri için uygun değildir. MSCOCO, MPII, LSP, FLIC, PoseTrack ve AI Challenger, daha karmaşık durumlarda daha fazla fotoğraf içeren derin öğrenme tabanlı algoritmalarda sıklıkla kullanılan veri kümelerini göstermek için kullanılabilir. LSP ve FLIC veri tabanları küçüktür ve yalnızca birkaç aktivite türü içerir. LSP veri seti, spor sahnesinden fotoğraflar içerir. Hollywood filmleri FLIC veri tabanlarının temelini oluşturmaktadır. MSCOCO ve AI Challenger gibi en yeni veri kümeleri daha fazla kategori içerir ve boyut olarak daha büyüktür.

HPE sistemleri ile aktivite tanıma, yetişkinler ve bebekler arasındaki görünüm ve poz farklılıkları nedeniyle performans açısından yetersizdir. Bu çalışmada, HPE sistemlerini bebekler için özelleştirdik. İnsan aktivite verileri çevrimiçi video veri tabanlarında bulunabilir. Çalışma kapsamında ortak konum etiketleri olan bebeklerin görüntülerinden/hareketlerinden oluşan, BabyPose adını verdiğimiz bir veri seti oluşturduk. Veri setinde bulunan video klip sayısı, Tablo 4.1’de verilmiştir.

Tablo 4.1. Bebek aktivite video görüntülerinin sınıflara göre dağılımı

Aktivite	Video klibi sayısı
Sırtı geriye çekme (arching back)	22
Parmak emme (sucking fingers)	32
Tekme sallama (kicking legs)	23
Kafa vurma (head banging)	26
Göz kaşıma (rubbing eyes)	30
Gerinme (stretching)	23

Veri setini oluştururken 'göz ovuşturan bebek', 'parmak emen bebek' gibi arama anahtar kelimeleri kullandık. Tablo 4.1'de gösterildiği gibi 6 kategoride toplamda 145 video klip toplanmıştır. Bu video kliplerden video kareleri çıkarılarak her biri OpenPose açık kaynak iskelet tespit sistemi ile etiketlenerek derin öğrenme modeli kurulmuştur. Veri kümesinden bazı örnekler Şekil 4.1'de verilmiştir.

Tablo 4.2. Oluşturulan BabyPose veri setinin detayları

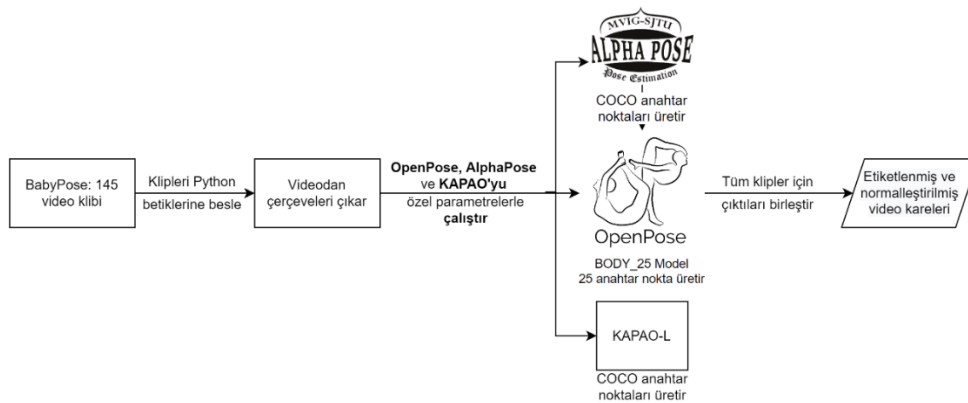
Konu	Uygulamalı makine öğrenmesi, Beden dilini anlama
Belirli konu alanı	Bebek aktivite tanıma
Verinin tipi	Resim, video, metin dosyası, ön işleme için Python betiği
Verinin toplanması	Veriler; tekme sallama, gerinme gibi bebek aktivitelerini tanımlayan anahtar kelimeleri kullanılarak çevrimiçi video veri tabanlarından toplanmıştır.
Veri formatı	Ham (çevrimiçi video kaynaklarından taranan ham video klipler) Analiz edilmiş (aktiviteleri açıklayan, oluşturulmuş alt klipleri belirtir) Filtrelenmiş (derin öğrenme girişi için uzuv koordinatlarını içeren JSON dosyaları)
Veri toplama detayları	Ham veriler: Bebeğin aktivitelerini tanımlayan bazı anahtar kelimeler, video klipleri elde etmek için kullanıldı. Videolar; Youtube, DailyMotion veya Vimeo gibi çevrimiçi video paylaşım platformlarından elde edilmiştir.
Verinin konumu	Veriler, Kocaeli Üniversitesi Teknoloji Fakültesi Bilişim Sistemleri Mühendisliği Bölümü Büyük Veri ve Bulut Bilişim Laboratuvarı'nda toplanmıştır.
Veri erişilebilirliği	Veri seti Github deposunda bulunabilir: https://github.com/meyurtsever/BabyPose



Şekil 4.1. BabyPose veri setinden bazı örnekler

4.1.1. Veri Ön İşleme Adımları

Her bir HPE modeli, farklı şablonda çıktılar üretmektedir. Bu çıktıların her biri; video kareleri için saniyede 20 kare olmak üzere alınmalı, her bir video için üretilen ve uzuv koordinat bilgilerini içeren JSON (JavaScript Object Notation) dosyaları her bir video için birleştirilmeli ve son olarak bu bilgiler, derin öğrenme modelinin eğitimi ve daha sonra tahminleme yapmak için normalize edilmeli, eğitim ve test olmak üzere veri seti bölünmelidir. BabyPose kapsamında veri ön işleme adımları bu temel algoritmaları takip etmektedir. Veri ön işleme adımları, Şekil 4.2’de özetlenmiştir. Veri setindeki tüm klipler, HPE modeline beslenir. HPE modeli çıktıları, çeşitli Python betiklerine beslenir. Python betiklerinde ilgili çıktılar, her bir klip için bir JSON çıktısı olarak düzenlenir.



Şekil 4.2. Veri ön işleme adımları

BabyPose proje dizininde her bir HPE modeli için AlphaPose, OpenPose ve KAPAO olacak şekilde ayrı klasörler mevcuttur. Klasörler içerisinde modele özgü veri ön işleme için oluşturulan Python betiklerinin yanı sıra; tahminleme için oluşturulan JSON çıktılarını içeren klasör, eğitilen LSTM modellerinin Python dosyalarını içeren klasörler, ilgili HPE için oluşturulan eğitim/test veri seti ve bebek aktivite tanıma için kullanılacak en iyi derin öğrenme modelinin kaydı bulunmaktadır.

JSON_to_txt Python betikleri ile OpenPose için BODY_25; KAPAO ve AlphaPose için COCO anahtar noktası çıktıları, her bir video klibi için bir metin dosyası çıktısı olacak şekilde düzenlenir. JSON çıktılar metin formatına dönüştürülürken her bir karenin yalnızca x ve y konumları, video karesi sırasında gerçekleştirilmekte olan aktivite ve karelerin dizilimi korunur. Bu bilgiler, ilgili aktivite sınıfı numarası 2D konum ile bir eğitim ve test veri seti oluşturmak için kullanılır.

Merge_all_txt Python betikleri, her bir HPE için mevcuttur. Tüm klipler için üretilmiş text dosyalarının hepsini, aktivite bilgisini (etiket) de içerecek şekilde birleştirir. Ardından bunlardan eğitim ve test veri setlerini oluşturur.

LSTM Models klasörlerinde hazırlanan veri setleri kullanılarak eğitilen LSTM modelleri mevcuttur. Her bir modelde en iyi modeli oluşturmak için K-Fold (k=10) eğitim yapılmıştır. Derin ağın eğitim detayları Performans Sonuçları ve Tartışma kısmında verilmiştir.

Predict activity Python betiği, gerçek zamanlı izleme için OpenCV'nin yanı sıra YOLOv4 ve Deep SORT ile videoda tahmin edilen aktivite sınıfının görselleştirmesini (implementasyon) içerir. Burada izlenen adım, genel mimarinin bir özetidir. Girdi olarak verilen video klibinden ilgili HPE kullanılarak uzuv koordinat bilgileri çıkarılır, çıktı koordinatları bir dosyada birleştirilir, koordinatlar normalize edilir ve LSTM ağına verilerek tahmin elde edilir. Elde edilen tahmin, gerçek zamanlı olarak video üzerinde gösterilir.

4.2. Çalışma Ortamı Hazırlıkları

Bu çalışmaya ait tüm kodlar tamamen Python'da yazılmıştır ve TensorFlow çerçevesine büyük ölçüde bağımlıdır. Kod özellikle Python 3.x, TensorFlow 1.5.x veya üstünü ve

NumPy modülünü gerektirir. AlphaPose ve KAPAO HPE modelleri için PyTorch modülü gereklidir. PyTorch GPU modelleri, hızlı etiketleme açısından tavsiye edilmektedir. Testler, belirtilen özelliklere sahip bir makinede yapılmıştır: CPU: Core i5-8400, RAM: 2x8GB DDR4 3000MHz, GPU: Nvidia RTX 3080 10GB, İşletim sistemi: Ubuntu 18.04 LTS ve Windows 10 20H2.

4.2.1. GitHub Üzerinden Projenin Kopyalanması

Projenin kaynak kodlarının ve kullanılan veri setinin sistemli, versiyon geçmişi ile birlikte tutulması ve çalışma ortamının kolaylıkla oluşturulabilmesi açısından BabyPose adını verdiğimiz proje GitHub platformuna yüklenmiştir. Popüler bir sürüm kontrol ve kaynak kod yönetim sistemi olan Git (URL-4) yardımı ile BabyPose projesi yerel diske kopyalanır. Ardından gdown paketi ile YOLOv4-tiny modeli, Google Drive platformundan indirilip BabyPose projesinin ana dizinine çıkarılır. YOLO modelinin boyutu fazla olduğundan, projenin haricinde verilmiştir. Projede kullanılan YOLOv4 modeli, büyük boyutu göz önüne alınarak Google Drive platformuna yüklenmiştir. Gdown Python kütüphanesi (URL-5) kullanılarak YOLO modeli, BabyPose ana dizinine indirilir ve çıkartılır.

4.2.2. OpenPose HPE ile Aktivite Tanıma ve Bağlılıkların Kurulumu

OpenPose ile poz tahmini yapabilmek için Anaconda kullanılarak sanal bir ortam oluşturulur. Platform aktif edildikten sonra tahminleyici, araştırmacıların sağladıkları rehber (URL-6) vasıtası ile kurulur. OpenPose, poz tahmini noktasında projenin temel yaklaşımını oluşturmaktadır.

Tablo 4.3. OpenPose HPE ile aktivite tanıma için gerekli kütüphaneler

Kütüphane	Sürüm
Keras	2.4.3
Keras-Preprocessing	1.1.2
Tensorflow	2.3.0
OpenCV-Python	4.4.0.44

Proje dahilinde derin öğrenme modeline yeni bir veri besleyip çıktı almak için OpenCV, Keras ve modelin yüklenmesi için TensorFlow modülleri de gereklidir. İlgili modüllerin kurulumu ile ilgili sürüm bilgileri, Tablo 4.3’de verilmiştir. Gerekli tüm gereksinimler

sağlandıktan sonra üzerinde aktivite tanıma yapılacak video, ana dizindeki demo klasörüne atılır. BabyPose ana dizininde Python betikleri çağrılarak demo çalıştırılır.

4.2.3. AlphaPose HPE ile Aktivite Tanıma ve Bağlılıkların Kurulumu

AlphaPose kullanılarak bebek aktivite tanıma uygulamasının gerçekleştirilebilmesi için Anaconda yardımı ile yeni bir sanal bir ortam oluşturulur. Platform aktif edildikten sonra AlphaPose, araştırmacıların sağladıkları rehber (URL-7) vasıtası ile kurulur. BabyPose projesi kapsamında AlphaPose v0.3.0 sürümü kullanılmıştır. Gerekli olan Torch sürümü, CUDA destekli kurulmuştur. Demonun hızlı çalışabilmesi adına imkân varsa GPU destekli Torch sürümü tavsiye edilmektedir.

Proje dahilinde derin öğrenme modeline yeni bir veri besleyip çıktı almak için OpenCV, Keras ve modelin yüklenmesi için TensorFlow modülleri gereklidir. Bu modüllerin tümü, AlphaPose klasörü altında requirements.txt klasöründe toplanmıştır. Torch gereksinimi, requirements kısmından çıkarılmıştır. Her sistemde donanımına uygun Torch sürümü kurulmalıdır. Torch'un CUDA ile uyumlu tüm sürümlerinin kurulum kodlarından (URL-8) sisteme uygun sürüm kurulmalıdır.

4.2.4. KAPAO HPE ile Aktivite Tanıma ve Bağlılıkların Kurulumu

KAPAO ile bebek aktivite tanıma uygulaması için Anaconda yardımı ile Python 3.6 sürümü kullanacak yeni bir sanal bir ortam oluşturulur. Platform aktif edildikten sonra KAPAO, araştırmacıların sağladıkları rehber (URL-9) takip edilerek kurulur. Gerekli olan Torch sürümü, CUDA destekli kurulmuştur. Demonun hızlı çalışabilmesi adına imkân varsa GPU destekli Torch sürümü tavsiye edilmektedir.

KAPAO için gerekli olan tüm isterler sağlandıktan sonra aktivite tanıma yapılacak video, ana dizin içerisindeki demo klasörüne yerleştirilir. KAPAO dizinine girilir. Komut satırında konumu ve adı belirtilmek suretiyle aktivite tanıma için oluşturulmuş betik dosyasına parametreler geçirilir ve işlem başlatılır.

4.3. Performans Metrikleri

Çalışma kapsamında LSTM modellerinin başarısını veya başarısızlığını değerlendirmek için doğruluk ve çapraz entropi dahil olmak üzere iki ölçüm kullanılmıştır. Doğruluk

hesabı, Denklem (4.1)'de verilmiştir. Doğruluk, karışıklık matrisinden elde edilir ve aşağıdaki denklemle gösterilir:

$$\text{Doğruluk} = \frac{TP + TN}{TP + FP + TN + FN} \quad (4.1)$$

Denklem (4.1)'de; TP (True Positive), modelin pozitif sınıfı doğru bir şekilde öngördüğü bir sonuçtur. Benzer şekilde TN (True Negative) gerçek bir negatif, modelin negatif sınıfı doğru bir şekilde tahmin ettiği bir sonuçtur. Yanlış pozitif (FP, False Positive), modelin pozitif bir sınıfı yanlış tahminlediği bir sonuçtur. Son olarak yanlış bir negatif (FN, False Negative), modelin negatif sınıfı yanlış tahminlediği sonuçtur.

Bu çalışmadaki kayıp fonksiyonu çapraz entropidir (CE). CE formülü, denklem (4.2)'de verilmiştir. Çapraz entropi, iki olasılık dağılımı arasındaki farkı ölçen ve entropiye dayanan bilgi teorisinden bir ölçüdür. P ve Q'dan gelen olayların olasılığı, çapraz entropiyi belirlemek için aşağıdaki gibi kullanılabilir:

$$CE(P, Q) = - \sum_{i=1}^{N_c} p(x_i) \log(q(x_i)) \quad (4.2)$$

Burada N_c sınıf sayısını belirtmektedir.

4.4. Performans Sonuçları ve Tartışma

Bu bölüm, önerilen mimarinin uygulama ayrıntılarını ve sonuçlarını özetlemektedir. Çalışma kapsamında modeller, 40 epochs (adım) ve 8/16 batch (grup) boyutları ile eğitilmiştir. Çalışmada katmanlar için sigmoid aktivasyon fonksiyonu seçilmiştir. Eğitim ve doğrulama kaybını tahmin etmek için adam optimizier ve kategorik çapraz entropi kaybı (categorical cross-entropy) kullanılmıştır. Ardından bu modeller doğruluk ölçütleri açısından değerlendirilmiştir. Her bir HPE için modeller, K-Fold çapraz doğrulama ile eğitilmiştir. K-Fold sonucu en iyi performans gösteren model seçilip erken durdurma (early stopping) ile birlikte tüm veri setinde eğitilerek final model elde edilmiştir. Takip

eden alt bölümlerde her bir HPE modeli için eğitilen LSTM ağ modellerinin performans sonuçları, seçilen final modelin karmaşıklık matrisi ile birlikte verilmiştir.

4.4.1. OpenPose HPE Tabanlı Derin Ağın Performans Değerlendirmesi

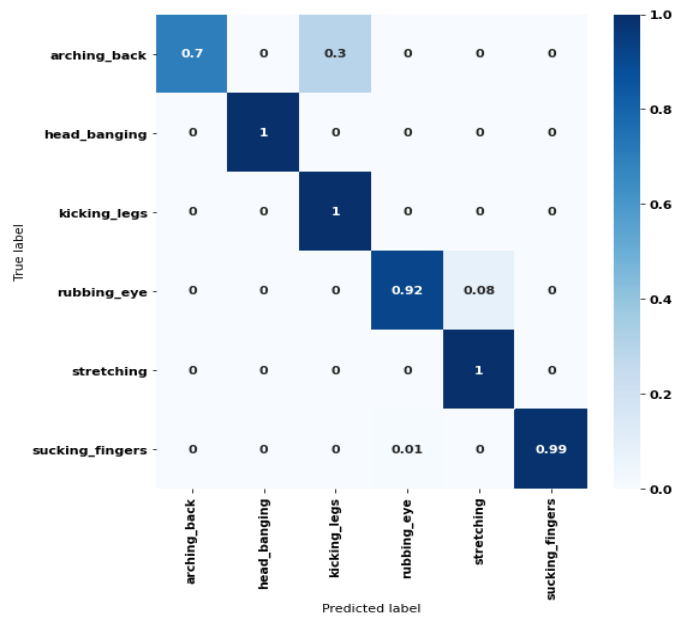
OpenPose HPE geliştiricileri, BODY_25 iskelet tespit çıktısının kullanımını önermişlerdir. Predict activity openpose Python dosyası, içerisinde belirli parametreleri ile OpenPose HPE uygulamasını başlatır. Takip parametrelerle OpenPose BODY_25 iskelet çıktısı olarak her video için bir metin dosyası olacak şekilde anahtar nokta çıktıları elde edilir: `-net resolution -1x368`, `-scale number 4`, `-scale gap 0.25`, `-disable blending`, `-display 0`, and `-keypoint scale 3`. JSON'lar text formatına çevrilirken her bir kare için yalnızca x ve y konumları, kare sırasında icra edilmekte olan aktivite ve karelerin sırası korunmuştur. Bu, ilgili aktivite sınıfı numarası ve bir dizi ortak 2D konum ile bir veri tabanı oluşturmak için kullanılır. BODY_25 modelinde 19-25 arasındaki anahtar noktalar, ayak parmak verisi olduğundan bu anahtar noktalar ihmal edilmiştir. Ardından, tüm çerçeve kümelerinin eğitim/test bölünmesi %80 ve %20 oranlarında yapılmıştır. Bundan sonra, n-adımları (pencere genişliği) numaralandıran bir çerçeve zaman serisi boyunca 19 eklem 2D konumu, çerçeve serisi için bir sınıf etiketi ile birlikte LSTM'ye beslenir. LSTM, 20 adımlı hafıza ile tasarlanmıştır. Bunun için her bir X_eğitim ve X_test verisi yeniden boyutlandırılmıştır. LSTM final model özeti, Tablo 4.4'te verilmiştir. Modelin ezberlemesini (overfitting) önlemek amacıyla L2 düzenleyici (regulizer) ve dropout katmanı kullanılmıştır. En iyi modeli oluşturmak için K-Fold (k=10 için) çapraz doğrulama yöntemi uygulanır. Son olarak, tahmin edilen aktivite sınıfı, gerçek zamanlı izleme için OpenCV ve YOLOv4 + Deep SORT ile video üzerinde görselleştirilir. OpenPose HPE ile aktivite tanıma için eğitilen tüm LSTM modellerinin performans sonuçları, Tablo 4.5'de verilmiştir. Görüleceği üzere, 1 LSTM ve 2 Yoğun katmana sahip LSTM ağı, en iyi performansı sergileyen ağ olmuştur. Şekil 4.3'te en iyi performansa sahip LSTM modelinin karmaşıklık matrisi verilmiştir. Karmaşıklık matrisi incelendiğinde; gerinme ve tekme sallama aktiviteleri haricinde model, diğer aktiviteleri başarıyla tanımıştır. Şekil 4.4'te ise OpenPose HPE'nin BabyPose veri seti üzerinde yaptığı iskelet tespitlerinin birkaç örneği paylaşılmıştır. Görüldüğü üzere OpenPose; çalışma kapsamında sınıflandırılması beklenen altı aktivite için de doğru etiketleme yapabilmektedir.

Tablo 4.4. OpenPose HPE ile aktivite tanıma için eğitilen LSTM ağının özeti

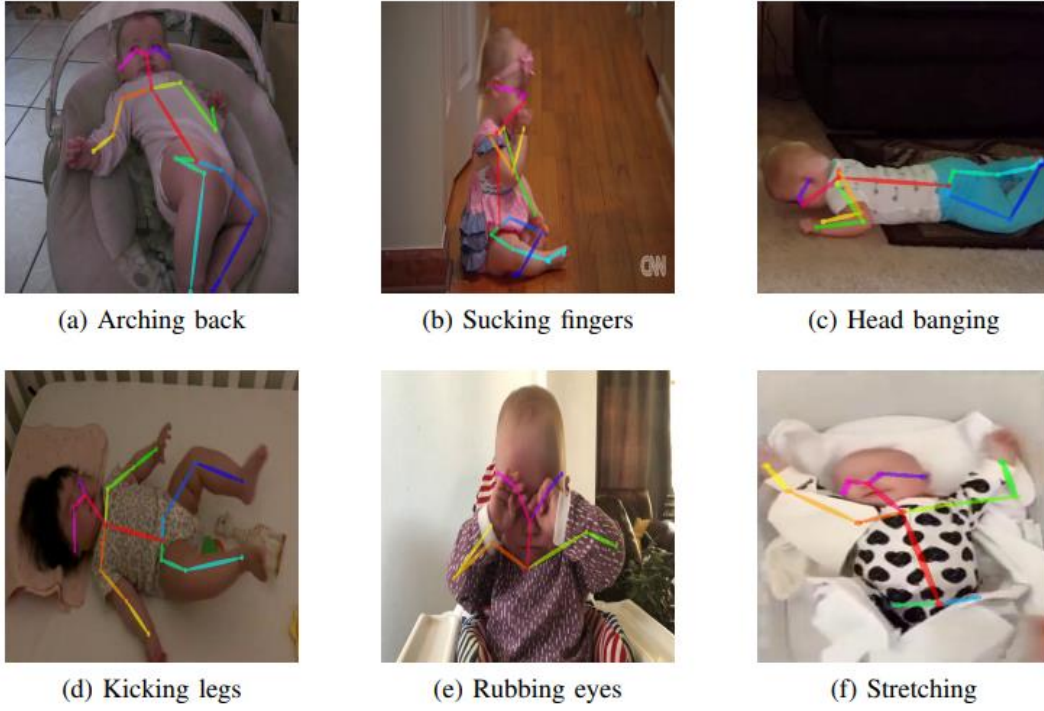
Katman Tipi	Çıktı Boyutu	Parametre Sayısı
lstm_1 (LSTM)	(None, 20, 16)	3520
lstm_2 (LSTM)	(None, 20, 32)	6272
lstm_3 (LSTM)	(None, 16)	3136
dense_1 (Dense)	(None, 32)	544
dropout_1 (Dropout)	(None, 32)	0
dense_2 (Dense)	(None, 32)	1056
dense_3 (Dense)	(None, 6)	198

Tablo 4.5. OpenPose HPE ile HAR için eğitilen derin ağların karşılaştırması

Model	Grup Boyutu	10-Fold Ortalama Kayıp	10-Fold Ortalama Doğruluk	Final Model Doğrulama Kaybı	Final Model Doğrulama Doğruluğu
3-seviyeli katmanlı LSTM ile 2 Yoğun Katman (32, 32)	16	0,4926	0,8384	0,2686	0,9447
3-seviyeli katmanlı LSTM ile 2 Yoğun Katman (32, 32)	8	0,4185	0,8721	0,2864	0,9414
1 LSTM hücresi ile 2 Yoğun Katman (32, 32)	16	0,1901	0,9564	0,1436	0,9698
1 LSTM hücresi ile 2 Yoğun Katman (32, 32)	8	0,1674	0,9691	0,0712	0,9983



Şekil 4.3. OpenPose HPE için eğitilen LSTM modeline ait karmaşıklık matrisi



Şekil 4.4. BabyPose veri setine ait görsellerin OpenPose HPE ile etiketlenmesi

4.4.2. AlphaPose HPE Tabanlı Derin Ağın Performans Değerlendirmesi

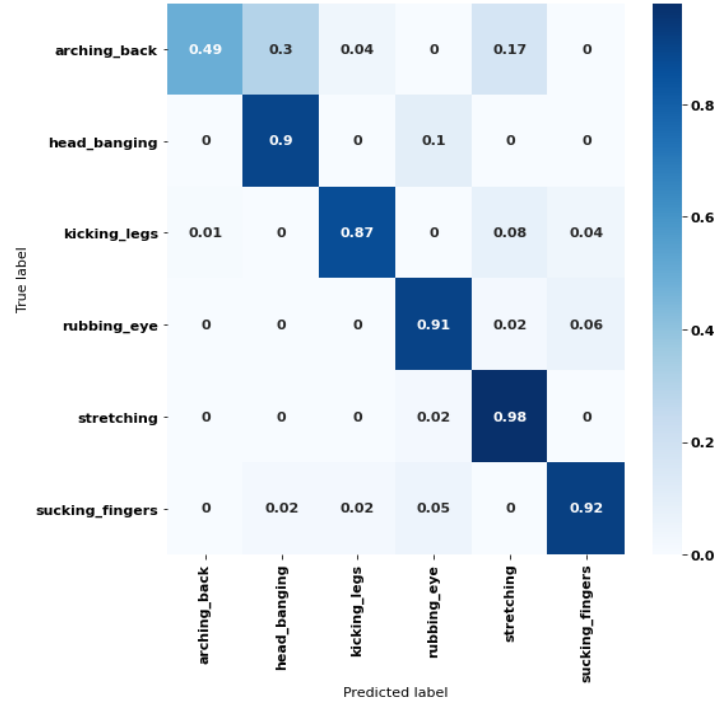
BabyPose dizininde yer alan AlphaPose dizini içerisindeki predict activity Python betiği, AlphaPose HPE'yi belirli parametrelerle çağırarak iskelet tespitini gerçekleştirir. BabyPose veri setinin etiketlenmesi de bu surette yapılmıştır. AlphaPose, takip eden parametreler ile çalıştırılmıştır: --sp, --vis_fast, --format open, --fast_inference False, --save_video. AlphaPose, OpenPose formatında çıktı alabilmektedir. Bu özellik de veri ön işleme aşamasında kolaylık sağlamaktadır. AlphaPose HPE çıktıları birleştirildikten sonra aktivite tanıma için LSTM modelleri eğitilmiştir. Aynı şekilde tüm LSTM ağları, K-Fold (k=10) ile eğitilmiş, en iyi performansa sahip fold, tüm veri setinde tekrardan eğitilerek en iyi model elde edilmiştir. Tablo 4.6, AlphaPose HPE ile aktivite tanıma için eğitilen final LSTM modelinin özetini göstermektedir. Tablo 4.7, AlphaPose HPE ile aktivite tanıma için oluşturulan LSTM ağ modellerinin performans karşılaştırmasını içermektedir. Görüldüğü üzere AlphaPose için 1 LSTM ve 2 Yoğun katmanlı model, en iyi başarıyı göstermiştir. Şekil 4.5'te AlphaPose HPE tabanlı aktivite tanıma modelinin karmaşıklık matrisi verilmiştir. Görüldüğü üzere model, gerinme aktivitesini doğru sınıflandırmakta problem yaşamaktadır. Model diğer aktiviteleri sınıflandırmakta ise başarılıdır.

Tablo 4.6. AlphaPose HPE ile aktivite tanıma için eğitilen LSTM ağının özeti

Katman Tipi	Çıktı Boyutu	Parametre Sayısı
lstm_1 (LSTM)	(None, 32)	8832
dense_1 (Dense)	(None, 32)	1056
dropout_1 (Dropout)	(None, 32)	0
dense_2 (Dense)	(None, 32)	1056
dense_3 (Dense)	(None, 6)	198

Tablo 4.7. AlphaPose HPE ile HAR için eğitilen derin ağların karşılaştırması

Model	Grup Boyutu	10-Fold Ortalama Kayıp	10-Fold Ortalama Doğruluk	Final Model Doğrulama Kaybı	Final Model Doğrulama Doğruluğu
3-seviyeli katmanlı LSTM ile 2 Yoğun Katman (32, 32)	16	0,7312	0,7568	0,6019	0,7923
3-seviyeli katmanlı LSTM ile 2 Yoğun Katman (32, 32)	8	0,6414	0,8036	0,4615	0,8944
1 LSTM hücresi ile 2 Yoğun Katman (32, 32)	16	0,3202	0,9143	0,2599	0,9304
1 LSTM hücresi ile 2 Yoğun Katman (32, 32)	8	0,3499	0,9042	0,3241	0,9232



Şekil 4.5. AlphaPose HPE için eğitilen LSTM modeline ait karmaşıklık matrisi

4.4.3. KAPAO HPE Tabanlı Derin Ağın Performans Değerlendirmesi

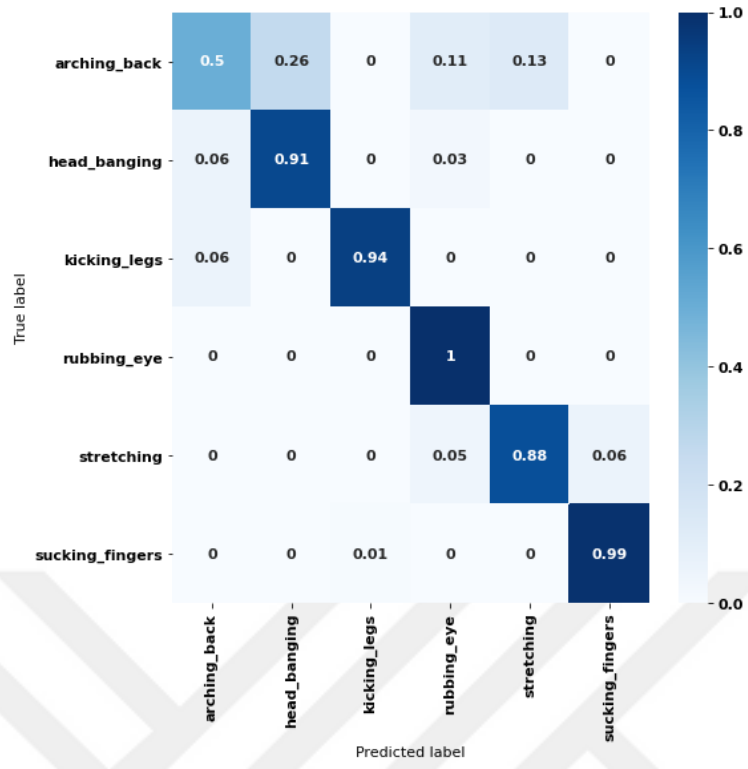
BabyPose dizininde yer alan KAPAO dizini içerisindeki predict activity Python betiği, KAPAO HPE'yi belirli parametrelerle çağırarak veri seti üzerinde iskelet tespiti gerçekleştirir. KAPAO, iskelet tespitlerini CSV olarak çıkarır. Bu verileri ön işlemek ve eğitim ya da tahminleme için kullanmak adına ilgili KAPAO dizininin altında ön işleyici ve tahminleyici Python betikleri bulunmaktadır. KAPAO ile iskelet çıktıları COCO formatında alınmış; tahminleme ve etiketlemede KAPAO-L ağırlıkları kullanılmıştır. Tablo 4.8, KAPAO HPE ile aktivite tanıma için eğitilen final LSTM model özetini içermektedir. Tablo 4.9, KAPAO HPE ile aktivite tanıma adına oluşturulan LSTM ağ modellerinin performans karşılaştırmasını içermektedir. KAPAO-L modeli için 1 LSTM hücresi ve 2 Yoğun katmandan oluşan model, en iyi performansı göstermiştir. Şekil 4.6'da önerilen KAPAO HPE ile aktivite tanıma LSTM modelinin karmaşıklık matrisi verilmiştir.

Tablo 4.8. KAPAO HPE ile aktivite tanıma için eğitilen LSTM ağının özeti

Katman Tipi	Çıktı Boyutu	Parametre Sayısı
lstm_1 (LSTM)	(None, 32)	8576
dense_1 (Dense)	(None, 32)	1056
dropout_1 (Dropout)	(None, 32)	0
dense_2 (Dense)	(None, 32)	1056
dense_3 (Dense)	(None, 6)	198

Tablo 4.9. KAPAO HPE ile HAR için eğitilen derin ağların karşılaştırması

Model	Grup Boyutu	10-Fold Ortalama Kayıp	10-Fold Ortalama Doğruluk	Final Model Doğrulama Kaybı	Final Model Doğrulama Doğruluğu
3-seviyeli katmanlı LSTM ile 2 Yoğun Katman (32, 32)	16	0.7261	0.7535	0.3614	0.9220
3-seviyeli katmanlı LSTM ile 2 Yoğun Katman (32, 32)	8	0.6816	0.7970	0.3789	0.8816
1 LSTM hücresi ile 2 Yoğun Katman (32, 32)	16	0.3115	0.9202	0.2266	0.9540
1 LSTM hücresi ile 2 Yoğun Katman (32, 32)	8	0.2743	0.9295	0.2362	0.9610



Şekil 4.6. KAPO HPE için eğitilen LSTM modeline ait karmaşıklık matrisi

5. SONUÇLAR VE ÖNERİLER

Bu çalışmada 2D video tabanlı poz tahmini ile bebeklerin vücut dillerini çözmek için, kurulan derin öğrenme modeliyle bebeklerin aktivitelerini tanıyıp sınıflandıran yeni bir sistem önerilmiştir. Çalışma kapsamında popüler üç adet iskelet tespit kütüphanesi kullanılarak bu yöntemlerin performans karşılaştırması yapılmıştır. Çalışma kapsamında iskelet anahtar noktalarını öğrenen bir LSTM ağı önerilmiştir. Sinir ağı modelimiz, herkese açık videolardan elde edilmiştir. Her bir HPE yapısı için oluşturulan derin öğrenme modeli, bebek aktivitelerini yüzde 90 ve üzerinde model tahmin doğruluğu ile doğru bir şekilde tahmin edebilmektedir. Geliştirilen sistem ile takip edilen hedeflere ulaşılmıştır: (i) Bebek pozu iskeletlerinin tahminlenmesi adına bir poz veri seti oluşturulması ve çeşitli bebek eylemlerinden oluşan veri kümesinin etiketlenmesi. (ii) Önerilen sistem ile bebeklerin beden dilinin çözülmesi ve bu konuda ebeveynlere yardımcı olunması. Bebeklerin gerçek zamanlı olarak izlenebilirliğinin sağlanması ve bebek güvenliği konusunda farkındalık yaratılması.

5.1. Çalışmanın Kısıtları

Bebeklerin poz tahmini, derin öğrenmeye özgü olmayan çeşitli sorunlara sahiptir: Güvenilir bir temel gerçek (doğru etiketli veri) içeren pozlarının toplanması, herhangi bir poz tahminleyicisini değerlendirmek için çok önemli bir kriterdir. Bu tür bilgilerin elde edilmesi genellikle büyük miktarda ham verinin toplanmasını gerektirir. Gelecekte daha fazla veri toplamaya ve daha farklı bebek aktivitelerine odaklanmaya devam edeceğiz.

5.2. Gelecek Çalışmalar

Literatürde farklı poz tahmincilerinin çapraz karşılaştırması için poz hatası yaygın olarak verilirken, çıkarım süresi ve bellek ayak izi gibi diğer önemli özellikler rutin olarak belgelenmez. Bu, bu tür kriterlerin çok önemli olduğu durumlarda (örneğin, sınırlı kaynaklara sahip gerçek zamanlı uygulamalar) tam bir karşılaştırma yapmayı imkânsız hale getirir. Sahne dinamiklerinin yanı sıra aydınlatma, mevsimsel ve bakış açısı değişikliklerinin tümü, bir sahnenin görsel görünümü üzerinde önemli bir etkiye sahip olabilir ve görsel ipuçları yoluyla yerini bulmayı zorlaştırabilir. Gelecek çalışmalarda bu hususlara dikkat edilecektir.

Son yıllarda mikro işlemciler tarafında hızla gelişmekte olan işlem gücüne dayanarak, bebek aktivite tanıma işleminin Raspberry Pi, NVIDIA Jetson Nano gibi mikro işlemciler üzerinde gerçek zamanlı olarak çalıştırılması hedeflenmektedir. Son yıllarda nesne takip modellerinin yüksek kare saniyelerde gerçek zamanlı çalışabildiği de göz önüne alındığında, mikro işlemciler üzerinde gerçek zamanlı nesne tespit, aktivite tanıma gibi uygulamaların konuşlandırılması artık daha da mümkün hale gelmiştir. İleriki çalışmalarda bu gelişmelerin göz önünde bulundurulması suretiyle geliştirilecek aktivite tanıma modellerinin uç cihazlar üzerinde konuşlandırılması planlanmaktadır.



KAYNAKLAR

- Angelini, F., Fu, Z., Long, Y., Shao, L., Naqvi, S. M. (2018). ActionXPose: A Novel 2D Multi-view Pose-based Algorithm for Real-time Human Action Recognition. *arXiv preprint arXiv:1810.12126*.
- Bradski, G. R. (1998). Computer vision face tracking for use in a perceptual user interface.
- Cao, Z., Hidalgo, G., Simon, T., Wei, S. E., Sheikh, Y. (2019). OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. *IEEE transactions on pattern analysis and machine intelligence*, 43(1), 172-186.
- Cao, Z., Simon, T., Wei, S. E., Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7291-7299.
- Chambers, C., Seethapathi, N., Saluja, R., Loeb, H., Pierce, S. R., Bogen, D. K., Kording, K. P. (2020). Computer vision to automatically assess infant neuromotor risk. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(11), 2431-2442.
- Cheng, Y. (1995). Mean shift, mode seeking, and clustering. *IEEE transactions on pattern analysis and machine intelligence*, 17(8), 790-799.
- Ciaparrone, G., Sánchez, F. L., Tabik, S., Troiano, L., Tagliaferri, R., Herrera, F. (2020). Deep learning in video multi-object tracking: A survey. *Neurocomputing*, 381, 61-88.
- Dang, L. M., Min, K., Wang, H., Piran, M. J., Lee, C. H., Moon, H. (2020). Sensor-based and vision-based human activity recognition: A comprehensive survey. *Pattern Recognition*, 108, 107561.
- Dang, Q., Yin, J., Wang, B., Zheng, W. (2019). Deep learning based 2d human pose estimation: A survey. *Tsinghua Science and Technology*, 24(6), 663-676.
- Demrozi, F., Pravadelli, G., Bihorac, A., Rashidi, P. (2020). Human activity recognition using inertial, physiological and environmental sensors: A comprehensive survey. *IEEE Access*, 8, 210816-210836.
- Eichner, M., Ferrari, V. (2010). We are family: Joint pose estimation of multiple persons. *In European conference on computer vision*, 228-242
- Fang, H. S., Xie, S., Tai, Y. W., Lu, C. (2017). Rmpe: Regional multi-person pose estimation. *In Proceedings of the IEEE international conference on computer vision*, 2334-2343.

- Ferron, M., Leonardi, C., Massa, P., Schiavo, G., Murphy, A. L., Farella, E. (2019). A walk on the child side: Investigating parents' and children's experience and perspective on mobile technology for outdoor child independent mobility. *In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1-12.
- Gilmore, J. N. (2020). Securing the kids: Geofencing and child wearables. *Convergence*, 26(5-6), 1333-1346.
- Graves, A. (2013). Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850*, 2013.
- Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Chen, T. (2018). Recent advances in convolutional neural networks. *Pattern recognition*, 77, 354-377.
- Gunjal, P. R., Gunjal, B. R., Shinde, H. A., Vanam, S. M., Aher, S. S. (2018). Moving object tracking using kalman filter. *In 2018 International Conference On Advances in Communication and Computing Technology (ICACCT)*, 544-547. IEEE.
- Hasan, M. N. U., Negulescu, I. (2020). Wearable technology for baby monitoring: a review. *J. Text. Eng. Fash. Technol*, 6(112.10), 15406.
- Herath, S., Harandi, M., Porikli, F. (2017). Going deeper into action recognition: A survey. *Image and vision computing*, 60, 4-21.
- Hesse, N., Bodensteiner, C., Arens, M., Hofmann, U. G., Weinberger, R., Sebastian Schroeder, A. (2018). Computer vision for medical infant motion analysis: State of the art and rgb-d data set. *In Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 0-0.
- Hesse, N., Pujades, S., Black, M. J., Arens, M., Hofmann, U. G., Schroeder, A. S. (2019). Learning and tracking the 3D body shape of freely moving infants from RGB-D sequences. *IEEE transactions on pattern analysis and machine intelligence*, 42(10), 2540-2551.
- Hesse, N., Pujades, S., Romero, J., Black, M. J., Bodensteiner, C., Arens, M., Sebastian Schroeder, A. (2018). Learning an infant body model from RGB-D data for accurate full body motion analysis. *In International Conference on Medical Image Computing and Computer-Assisted Intervention*, 792-800, Springer, Cham.
- Hesse, N., Sebastian Schroeder, A., Müller-Felber, W., Bodensteiner, C., Arens, M., G Hofmann, U. (2017). Markerless motion analysis for early detection of infantile movement disorders. *In Embec & Nbc 2017*, 197-200. Springer, Singapore.
- Horn, B. K., Schunck, B. G. (1981). Determining optical flow. *In Techniques and Applications of Image Understanding, International Society for Optics and Photonics*, 281, 319-331.

- Hussain, T., Muhammad, K., Khan, S., Ullah, A., Lee, M. Y., Baik, S. W. (2019). Intelligent baby behavior monitoring using embedded vision in IoT for smart healthcare centers. *Journal of Artificial Intelligence and Systems*, 1(1), 110-124.
- Insafutdinov, E., Andriluka, M., Pishchulin, L., Tang, S., Levinkov, E., Andres, B., Schiele, B. (2017). Artrack: Articulated multi-person tracking in the wild. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, 6457-6465.
- Insafutdinov, E., Pishchulin, L., Andres, B., Andriluka, M., Schiele, B. (2016). Deepercut: A deeper, stronger, and faster multi-person pose estimation model. *In European Conference on Computer Vision*, 34-50. Springer, Cham.
- Jaderberg, M., Simonyan, K., Zisserman, A. (2015). Spatial transformer networks. *Advances in neural information processing systems*, 28.
- Kojovic, N., Natraj, S., Mohanty, S. P., Maillart, T., Schaer, M. (2021). Using 2D Video-based Pose Estimation for Automated Prediction of Autism Spectrum Disorders in Preschoolers. *medRxiv*.
- Leal-Taixé, L. (2014). Multiple object tracking with context awareness. *arXiv preprint arXiv:1411.7935*.
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., Berg, A. C. (2016). Ssd: Single shot multibox detector. *In European conference on computer vision*, 21-37. Springer, Cham.
- Liu, Z., Zhu, J., Bu, J., Chen, C. (2015). A survey of human pose estimation: the body parts parsing based methods. *Journal of Visual Communication and Image Representation*, 32, 10-19.
- Lonini, L., Shawen, N., Scanlan, K., Rymer, W. Z., Kording, K. P., Jayaraman, A. (2016). Accelerometry-enabled measurement of walking performance with a robotic exoskeleton: a pilot study. *Journal of NeuroEngineering and Rehabilitation*, 13(1), 1-10.
- Luvizon, D. C., Tabia, H., Picard, D. (2019). Human pose regression by combining indirect part detection and contextual information. *Computers & Graphics*, 85, 15-22.
- Marcroft, C., Khan, A., Embleton, N. D., Trenell, M., Plötz, T. (2015). Movement recognition technology as a method of assessing spontaneous general movements in high risk infants. *Frontiers in neurology*, 5, 284.
- McNally, W., Vats, K., Wong, A., McPhee, J. (2021). Rethinking Keypoint Representations: Modeling Keypoints and Poses as Objects for Multi-Person Human Pose Estimation. *arXiv preprint arXiv:2111.08557*.

- Morales, L. V. M., Piratova, K. X. C., Martinez, M. A. R. (2019). Real-time temperature and audio baby monitoring using IoT technologies. *In 2019 Congreso Internacional de Innovación y Tendencias en Ingeniería (CONIITI)*, 1-6.
- Müller, C., Cienki, A., Fricke, E., Ladewig, S., McNeill, D., Tessendorf, S. (2013). Body-language-communication. *An international handbook on multimodality in human interaction*, 1(1), 131-232.
- Newell, A., Yang, K., Deng, J. (2016). Stacked hourglass networks for human pose estimation. *In European conference on computer vision*, 483-499. Springer, Cham.
- Newell, A., Yang, K., Deng, J. (2016). Stacked hourglass networks for human pose estimation. *In European conference on computer vision*, 483-499. Springer, Cham.
- Noori, F. M., Wallace, B., Uddin, M. Z., Torresen, J. (2019). A Robust Human Activity Recognition Approach Using OpenPose, Motion Features, and Deep Recurrent Neural Network. *In Scandinavian Conference on Image Analysis*, 299-310. Springer, Cham.
- Nugent, K. (2011). Your baby is speaking to you: A visual guide to the amazing behaviors of your newborn and growing baby. *HarperCollins*.
- Osokin, D. (2018). Real-time 2D multi-person pose estimation on CPU: lightweight OpenPose. *arXiv preprint arXiv:1811.12004*.
- Putatunda, S. (2020). Care2Vec: a hybrid autoencoder-based approach for the classification of self-care problems in physically disabled children. *Neural Computing and Applications*, 32(23), 17669-17680.
- Rowlands, M. J. (2011). Body language: Representation in action. *MIT Press*.
- Sciortino, G., Farinella, G. M., Battiato, S., Leo, M., Distanto, C. (2017). On the estimation of children's poses. *In International conference on image analysis and processing*, 410-421. Springer, Cham.
- Toshev, A., Szegedy, C. (2014). Deeppose: Human pose estimation via deep neural networks. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, 1653-1660.
- Uddin, M. Z., Torresen, J. (2018). A Deep Learning-Based Human Activity Recognition in Darkness. *In 2018 Colour and Visual Computing Symposium (CVCS)*, 1-5. IEEE.
- Ulin, L., Insafutdinov, E., Tang, S., Andres, B., Andriluka, M., Gehler, P. V., Schiele, B. (2016). Deepcut: Joint subset partition and labeling for multi person pose estimation. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, 4929-4937.
- URL-1: <https://voices.uchicago.edu/babylab/> (Ziyaret Tarihi: 20 Ekim 2021)

- URL-2: <https://www.homecourt.ai/> (Ziyaret Tarihi: 29 Ekim 2021)
- URL-3: <https://www.mathworks.com/help/vision/tracking-and-motion-estimation.html/> (Ziyaret Tarihi: 29 Ekim 2021)
- URL-4: <https://git-scm.com/> (Ziyaret Tarihi: 27 Mayıs 2022)
- URL-5: <https://github.com/wkentaro/gdown> (Ziyaret Tarihi: 27 Mayıs 2022)
- URL-6: <https://github.com/CMU-Perceptual-Computing-Lab/openpose#installation> (Ziyaret Tarihi: 27 Mayıs 2022)
- URL-7: <https://github.com/MVIG-SJTU/AlphaPose/tree/pytorch> (Ziyaret Tarihi: 27 Mayıs 2022)
- URL-8: <https://pytorch.org/get-started/previous-versions/> (Ziyaret Tarihi: 27 Mayıs 2022)
- URL-9: <https://github.com/wmcnally/kapao#setup> (Ziyaret Tarihi: 27 Mayıs 2022)
- Welch, G., Bishop, G. (1995). An introduction to the Kalman filter.
- Wojke, N., Bewley, A., Paulus, D. (2017). Simple online and realtime tracking with a deep association metric. *In 2017 IEEE international conference on image processing (ICIP)*, 3645-3649. IEEE.
- Wojke, N., Bewley, A., Paulus, D. (2017). Simple online and realtime tracking with a deep association metric. *In 2017 IEEE international conference on image processing (ICIP)*, 3645-3649. IEEE.
- Yadav, S. K., Singh, A., Gupta, A., Raheja, J. L. (2019). Real-time Yoga recognition using deep learning. *Neural Computing and Applications*, 31(12), 9349-9361.
- Yang, Y., Ramanan, D. (2012). Articulated human detection with flexible mixtures of parts. *IEEE transactions on pattern analysis and machine intelligence*, 35(12), 2878-2890.

KİŞİSEL YAYIN VE ESERLER

Yurtsever, M. M. E., Özcan, M., Taruz, Z., Eken, S., Sayar, A. (2022). Figure search by text in large scale digital document collections. *Concurrency and Computation: Practice and Experience*, 34(1), e6529.

Yurtsever, M. M. E., Eken, S. (2022). BabyPose: Real-Time Decoding of Baby's Non-verbal Communication using 2D Video-based Pose Estimation, *IEEE Sensors Journal*, 22(14), DOI: 10.1109/JSEN.2022.3183502.



ÖZGEÇMİŞ

Muhammet Mücahit Enes YURTSEVER; ilk, orta ve lise öğrenimini Kocaeli’de tamamladı. 2017 yılında başladığı Kocaeli Üniversitesi Mühendislik Fakültesi Bilgisayar Mühendisliği Bölümü’nden 2020 yılında mezun oldu. Takip eden eğitim-öğretim yılında Kocaeli Üniversitesi Bilişim Sistemleri Mühendisliği Anabilim Dalı’nda yüksek lisans eğitimine başladı. 2022 Yılında Kocaeli Üniversitesi’nde Araştırma Görevlisi olarak göreve başladı. Halen Kocaeli Üniversitesi Bilişim Sistemleri Mühendisliği Bölümü’nde Araştırma Görevlisi olarak görev yapmaktadır.

