

Derin Öğrenmeye Dayalı 2 Boyutlu İnsan Poz Tahmin Modellerinin Karşılaştırılması

Comparison of Deep Learning Based 2D Human Pose Estimation Models

Cumhur TORUN^{1*}, Abdulkadir KARACI²

^{1*} Kastamonu Üniversitesi, Mühendislik ve Mimarlık Fakültesi Malzeme Bilimi ve Mühendisliği, Kastamonu, Türkiye

² Samsun Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi Yazılım Mühendisliği, Samsun, Türkiye

ÖZET

İnsan hareketlerinin analizi, bilgisayarlı görü ve yapay zekâ alanlarında büyük öneme sahip bir çalışma alanıdır. Özellikle eklem noktalarının tespiti, insan hareketlerinin ve duruşlarının dijital ortamda modellenmesi açısından kritik rol oynar. Bu alan, tıp, spor, rehabilitasyon, güvenlik, insan-bilgisayar etkileşimi gibi birçok disiplinde geniş bir kullanım alanına sahiptir. Eklem noktalarının doğru ve etkin bir şekilde belirlenmesi sayesinde, sporcuların performans değerlendirilmesi, hastaların rehabilitasyon süreçlerinin izlenmesi ve işaret dili gibi karmaşık hareketlerin dijital ortamda tanınması sağlanabilmektedir. Eklem noktalarını tespit etmek için geliştirilen çeşitli yazılım kütüphaneleri, farklı algoritmalar kullanarak hız, doğruluk ve kullanım kolaylığı açısından avantajlar sunmaktadır. MediaPipe, MoveNet, OpenPose, AlphaPose, Detectron2 ve HRNet gibi popüler kütüphaneler, bu alanda yaygın olarak kullanılmakta ve her birinin belirli avantaj ve dezavantajları bulunmaktadır. Bu çalışmada, kütüphaneler karşılaştırılarak insan hareketlerinin analizinde kullanılan yöntemler ve kullanılacakları alanlara yönelik değerlendirmeler yapılmıştır. MediaPipe Holistic ve MoveNet kütüphaneleri gerçek zamanlı uygulamalarda başarılı bulunurken, AlphaPose, ViTPose ve HRNet'in yüksek doğruluk gerektiren uygulamalar için daha etkili olduğu görülmüştür.

Anahtar Kelimeler: İnsan hareket analizi, eklem noktası tespiti, bilgisayarlı görü, poz tahmin modelleri

ABSTRACT

Human motion analysis is a field of study of great importance in computer vision and artificial intelligence. In particular, the detection of joint points plays a critical role in digitally modeling human movements and postures. This field has many applications in many disciplines such as medicine, sports, rehabilitation, security, and human-computer interaction. By accurately and efficiently identifying joint points, it is possible to evaluate athletes' performance, monitor patients' rehabilitation process, and recognize complex gestures such as sign language in a digital environment. Various software libraries developed to detect joint points offer advantages in terms of speed, accuracy, and ease of use by using different algorithms. Popular libraries such as MediaPipe, MoveNet, OpenPose, AlphaPose, Detectron2, and HRNet are widely used in this field, and each has certain advantages and disadvantages. In this study, libraries are compared and compared for analyzing human movement. evaluations of the methods used and the areas where they can be used It was done. MediaPipe Holistic and MoveNet libraries were found to be successful in real-time applications, whereas AlphaPose, ViTPose, and HRNet were found to be more effective for applications requiring high accuracy.

Keywords: Human motion analysis, joint point detection, computer vision, pose estimation models

Başvuru: 25.10.2024 Son Revizyon: 19.11.2024 Kabul: 06.12.2024

Doi: 10.51764/smutgd.1573626

^{1*}Sorumlu yazar: Kastamonu Üniversitesi, Mühendislik ve Mimarlık Fakültesi Malzeme Bilimi ve Mühendisliği, Kampüs, Kastamonu, Türkiye;
E-mail: cumhurtorun@gmail.com; ORCID: <https://orcid.org/0009-0000-9984-1384>

² E-mail: akaraci@gmail.com; ORCID: <https://orcid.org/0000-0002-2430-1372>

1. GİRİŞ

İnsan hareketinin analizi üzerine geçmişten günümüze kadar çeşitli çalışmalar yapılmıştır. Kinetik ve kinematik hareket analizi için birçok deney, yöntem ve araç geliştirilmiştir. Bilgisayarlı görü alanında eklem noktalarının tespiti, insan hareketlerinin ve duruşlarının dijital olarak anlaşılması ve modellenmesi açısından kritik bir öneme sahiptir. Bu tespit işlemi, insan vücudunun anatomik yapısına uygun olarak belirli anahtar noktaların (örneğin, eklemler ve uzuvlar) konumlarının hassas bir şekilde belirlenmesini içerir. Eklem noktalarının doğru ve etkin bir şekilde saptanması, çeşitli uygulama alanlarında derinlemesine analiz ve yenilikçi çözümler sunar. Özellikle tıp ve rehabilitasyon alanlarında, eklem noktalarının analizi sayesinde hastaların hareket bozuklukları teşhis edilebilir ve tedavi süreçleri izlenebilir (Wu vd, 2020). Spor bilimlerinde atletlerin performans değerlendirmeleri ve hareket optimizasyonu için kullanılırken, insan-bilgisayar etkileşimi kapsamında doğal ve sezgisel arayüzlerin geliştirilmesine katkı sağlar (Jafarzadeh vd, 2021). Ayrıca, güvenlik ve gözetim sistemlerinde anormal davranışların tespiti, artırılmış ve sanal gerçeklik uygulamalarında gerçekçi avatarların oluşturulması ve robotik alanında insan hareketlerinin taklit edilmesi gibi çeşitli disiplinlerde de yaygın olarak kullanılmaktadır (Gao vd, 2019) (Wu vd, 2020).

Bilgisayarlı görü alanında eklem noktalarının tespiti, işaret dili tanıma ve analizinde temel bir rol oynamaktadır. İşaret dili, el hareketleri, yüz ifadeleri ve beden duruşlarının kombinasyonu ile zengin bir iletişim biçimi sunar. Bu karmaşık hareketlerin doğru anlaşılması ve dijital ortama aktarılması için el ve parmak eklemlerinin hassas tespiti kritik öneme sahiptir. Eklem noktalarının tespiti, işaret dilinin bileşenlerinin ayrıştırılmasını ve hareketlerin zamansal ve mekânsal özelliklerinin analiz edilmesini sağlar. Bu sayede, derin öğrenme tabanlı modeller ve hareket yakalama teknikleri kullanılarak işaretlerin otomatik olarak tanınması ve sınıflandırılması mümkün hale gelir (Karacı vd, 2021). Ayrıca, bu yaklaşım işaret dili eğitimi, tercümesi ve işitme engelli bireylerle etkileşimde yeni teknolojik çözümlerin geliştirilmesine katkı sunar. Dolayısıyla, bilgisayarlı görü ile eklem noktalarının tespiti, işaret dili araştırmalarında ve uygulamalarında hem teorik hem de pratik açıdan önemli bir araştırma alanıdır.

Bilgisayarlı görü alanında eklem noktalarının tespiti için çeşitli yazılım kütüphaneleri geliştirilmiştir. Bu kütüphaneler, insan vücudu üzerindeki eklem noktalarını belirlemek için farklı algoritmalar ve mimariler kullanarak, hareket analizi ve tanıma görevlerinde önemli rol oynamaktadır. Her bir kütüphane, işlem hızı, doğruluk, platform uyumluluğu ve kullanım kolaylığı gibi farklı avantaj ve dezavantajlara sahiptir. Özellikle işaret dili kelimelerinin tespitinde, el ve parmak hareketlerinin hassas bir şekilde analiz edilmesi gerekmektedir. Bu nedenle, kullanılacak yazılım kütüphanesinin bu spesifik gereksinimleri karşılayabilmesi kritik öneme sahiptir. Bu çalışmada, işaret dili kelimelerinin tespiti için uygun olabilecek yazılım kütüphaneleri karşılaştırılmış ve performansları detaylı bir şekilde değerlendirilmiştir.

2. LİTERATÜR İNCELEMESİ

Bu çalışmada, literatür taraması yapılırken Google Scholar, IEEE Xplore, ScienceDirect ve Springer gibi önde gelen veri tabanları kullanılmıştır. Arama sırasında 'insan poz tahmini', 'derin öğrenme', 'bilgisayarlı görü', 'MediaPipe', 'OpenPose', 'AlphaPose', 'Detectron2' ve 'HRNet' gibi anahtar kelimeler kullanılmıştır. Araştırmaya dahil edilen çalışmalar, 2021-2024 yılları arasında yayımlanmış, yüksek etki faktörüne sahip dergilerde ve önemli konferanslarda sunulan makalelerden seçilmiştir. Seçim kriterleri olarak, insan poz tahmini alanındaki güncel ve özgün çalışmaların yanı sıra uygulama alanlarında kullanılan modellerin performanslarını değerlendiren çalışmalar dikkate alınmıştır.

Kim vd. (2023), evde yalnız yaşayan yaşlı bireyler için bir mobil robot kullanarak poz tahminine odaklanan bir çalışma yapmıştır. MediaPipe'in hızlı ve hafif 2D poz tahmin yeteneklerini kullanarak eklem açılarını tahmin eden bir model geliştirmişlerdir. Bu sistem, derin öğrenme yöntemleri ve optimizasyon teknikleriyle 3D poz tahmininde karşılaşılan zorlukları çözmek için kullanılmıştır. Çalışmada, eklem açılarının doğru bir şekilde tahmin edilmesi için hareket verileri analiz edilmiş ve video görüntüleri kullanılarak pratik senaryolarda test edilmiştir. Dill vd. (2023) çalışması, MediaPipe Pose'un doğruluğunu değerlendirmeye odaklanarak, özellikle fiziksel egzersizlerin analizi için potansiyelini araştırmaktadır. Araştırma, dört katılımcının çeşitli egzersizleri yaparken iki kamera ve hareket yakalama sistemi kullanılarak gerçekleştirilen deneyler üzerine kurulmuştur. Elde edilen bulgular, MediaPipe Pose'un doğruluğunun hem kameranın bakış açısına hem de egzersiz türüne göre önemli ölçüde değiştiğini ortaya koymaktadır. Optimal bakış açıları yüksek doğruluk sağlarken, öz-örtüşme veya elverişsiz açılarda önemli doğruluk kayıpları gözlemlenmiştir. Çalışmada, çoklu kamera entegrasyonu ve biyomekanik modellerin doğruluğu artırabileceği öne sürülmektedir. Garg vd. (2023) yaptığı çalışmada yoga duruşlarının sınıflandırılmasına yönelik MediaPipe kütüphanesi gibi araçlar kullanılarak vücut anahtar noktalarının tespiti ve iskeletleştirme işlemleri gerçekleştirilmiştir. Sonuçlar, iskeletleştirilmiş görüntülerle eğitilen derin öğrenme modellerinin, iskeletleştirilmemiş görüntülere kıyasla daha yüksek doğruluk oranlarına ulaştığını göstermektedir. Bora vd.

(2023), Assamese İşaret Dili'ni tanıyan bir sistem geliştirmek için makine öğrenimi tekniklerini kullanarak el hareketlerini tanımlamaya yönelik bir yaklaşım sunmuştur. Çalışmada Assamese el hareketlerinin iki ve üç boyutlu görüntülerinin yer aldığı 2094 veri noktasından oluşan bir veri seti hazırlanmış ve MediaPipe görüntülerdeki işaret noktalarını tespit etmek için kullanılmıştır. Dokuz statik kelimeyi içeren bu veri seti, bir ileri beslemeli yapay sinir ağı modeliyle eğitilmiş ve %99 doğruluk oranı elde edilmiştir. Singh vd. (2021), MediaPipe Holistic çözümü, yüz, el ve vücut pozlarını eşzamanlı olarak gerçek zamanlı tahmin edebilen bir sistemin tasarımında kullanmıştır. Bu çalışmada, özellikle insan-bilgisayar etkileşimi ve güvenlik uygulamaları için hızlı ve doğruluk oranı yüksek bir çözüm sunulmuştur. Çalışmada, gerçek zamanlı video analiziyle insan hareketleri yakalanmış ve analiz edilmiştir.

Parashar vd. (2023), MediaPipe ve MoveNet'i entegre ederek yoga pozlarının tespitini ve sınıflandırmasını gerçekleştirmişlerdir. Bu modellerin kombinasyonu, yoga pozlarını doğru şekilde iskeletleştirme ve anahtar noktaları tespit etme konusunda önemli bir iyileşme sağlamıştır. Model, sağlık ve fitness alanlarında geniş uygulama potansiyeline sahip olup, eğitim verisi olarak geniş ve çeşitli yoga veri setleri kullanılmıştır. Jo vd. (2022), MoveNet, OpenPose ve PoseNet gibi poz tespiti modellerini karşılaştırmışlardır. Çalışma sonuçlarına göre, MoveNet'in özellikle mobil cihazlarda hız ve hesaplama verimliliği açısından öne çıktığı tespit edilmiştir. Ayrıca, OpenPose'un daha fazla ayrıntılı anahtar nokta tespiti yapabildiği ve MoveNet'in daha hızlı ve daha az kaynak kullanımıyla daha iyi performans gösterdiği tespit edilmiştir.

Liv vd. (2024), yaşlıların sağlık durumları takibi ve acil durum müdahalesi için derin öğrenmeye dayalı gerçek zamanlı düşme tespiti yapmışlardır. Mevcut poz tanıma yöntemlerinin büyük model boyutları, zayıf zamanlılık ve çoklu hedef pozları doğru şekilde tanıma yetersizliği gibi sorunlarına çözüm getirmek amacıyla, hafif ve geliştirilmiş bir OpenPose tabanlı gerçek zamanlı düşme tespit algoritması önermişlerdir. Bu algoritma, VGG-19 yerine hafif bir MobileNet ağı kullanarak model boyutunu ve hesaplama yükünü önemli ölçüde azaltmaktadır. Deneysel sonuçlar, algoritmanın başarılı bir gerçek zamanlı performans sergileyerek çoklu kişi poz tespiti için uygulama gereksinimlerini karşıladığını göstermektedir. Arkushin vd. (2023), HamNoSys işaret dili notasyonunu işaret hareket dizilerine dönüştüren ilk yöntem olan Ham2Pose'u sunmuştur. HamNoSys, dil bağımsızlığı sunan evrensel bir işaret dili transkripsiyon sistemidir ve bu çalışma, farklı işaret dillerine uygun genel bir çözüm sunmayı hedeflemektedir. Araştırma, HamNoSys metnini ve referans bir poz çerçevesini alarak adım adım doğru işaret hareketini üretecek şekilde bir model tasarlamıştır. Model, OpenPose ile elde edilen anahtar noktalar üzerinde eğitim gerçekleştirmiştir. Cao vd. (2017) yaptığı çalışmada, OpenPose kullanılarak insan vücut pozlarının çoklu kişiyle eş zamanlı tespiti üzerine odaklanılmaktadır. Yapılan sistemde, Part Affinity Fields (PAFs) adlı bir teknik kullanarak insan vücudundaki parçalar arasındaki ilişkileri öğrenmiş ve bu parçaları doğru bir şekilde bireylere atanmıştır. Badiola-Bengoa vd. (2021) tarafından yapılan çalışmada, insan vücut pozlarının spor ve egzersizlerde izlenmesi amacıyla OpenPose'un kullanıldığı çalışmaların sistematik bir incelemesi yapılmıştır. Çalışmada, spor ve egzersiz uygulamalarında HPE'nin (Human Pose Estimation) nasıl geliştirildiği ve OpenPose'un bu bağlamda kullanımının nasıl optimize edilebileceği üzerine odaklanılmıştır.

Song vd. (2024) masa tenisi teknik hareketlerini tanımaya yönelik iki model geliştirmiştir. Çalışma, dinamik ve statik karmaşık ağ analizlerini topluluk algılama algoritmalarıyla birleştirerek oyuncuların teknik ve taktiksel örüntülerini değerlendirmeyi amaçlamaktadır. AlphaPose ile 37 elit oyuncunun 8015 yüksek çözünürlüklü videosunu içeren iki veri seti oluşturulmuştur. YOLOv8-AlphaPose tabanlı yapay zeka algoritmasının tanıma ve tahmin performansının çalışmadaki diğer algoritmalara (LTSM, biGRU, CNN) göre başarılı olduğu görülmüştür. Areerob vd. (2024), kişisel koruyucu ekipman (PPE- Personal Protective Equipment) tespitini, kalabalık sahnelerde iyileştirmek amacıyla dikkat tabanlı YOLOv7 ve insan duruş tahmini için AlphaPose'u kullanmışlardır. Çalışma, PPE'nin doğru konumlandırılıp konumlandırılmadığını doğrulamak için insan poz tahminini kullanarak yanlış tespitleri ortadan kaldırmış ve tespit doğruluğunu artırmıştır. AlphaPose'un, kalabalık sahnelerde yüksek doğruluk sağlaması ve mevcut yöntemlerin sınırlamalarını aşması hedeflenmiştir. Fang vd. (2022) yaptığı çalışmada, AlphaPose kullanılarak vücut, yüz, el ve ayak anahtar noktalarının eşzamanlı tespiti ve izlenmesi üzerine odaklanılmaktadır. Önerilen yöntem, COCO-WholeBody ve Halpe-FullBody gibi geniş veri setlerinde test edilmiştir ve mevcut yöntemlerden daha hızlı ve doğru sonuçlar sağlamıştır. Zhu vd. (2020) yaptığı çalışma; insan vücudunun 133 anahtar noktasını (vücut, yüz, el, ayak) içeren tam vücut poz tahminini ele alır. AlphaPose sistemine dayalı olarak, özellikle eller ve yüz gibi daha ince ayrıntıları tespit etmeyi hedefler. Çalışma, YOLO insan algılayıcısını kullanarak yüksek verimlilik ve doğruluk sağlamıştır ve MSCOCO2017 veri setinde test edilmiştir.

Bibin (2024) yaptığı çalışmada; Detectron2, insan hareketlerinin tanınması için LSTM modeliyle birleştirilmiştir. Video analizinde, her bir karedeki anahtar noktalar Detectron2 ile tespit edilmiş ve LSTM modeli bu hareketleri analiz etmiştir. Sonuç olarak, model %91'in üzerinde doğruluk sağlamış ve gerçek zamanlı hareket tanımadaki etkili olmuştur. Bu çalışma, spor analizi ve güvenlik gibi alanlarda kullanılacak bir sistem önermektedir. Lee vd. (2024) yaptığı çalışmada yanlış duruş alışkanlıklarından olan baş öne duruşunu (FHP-Front Head Position) tespit eden bir sistem geliştirmişlerdir. İnsan duruşunu standart 2D görüntülerden tahmin etmeye yönelik, mevcut

modellerin çoğu, FHP'nin tanısı için hayati öneme sahip olan C7 omuru ve kraniyovertebral açığı ölçmemektedir. Bu eksiklik, FHP tespit sistemlerinin geliştirilmesinde önemli bir zorluk teşkil etmektedir. Detectron 2 kullanarak geliştirilen sistem ile gerçek zamanlı FHP teşhisi yapılabilmektedir. Davoudi Khaskoli vd. (2024), süper düşük çözünürlüklü görüntülerde nesne tespiti için Detectron2 kullanarak verimli bir yöntem geliştirmişlerdir. Bu çalışmada, kişilerin tespit edilmesi amacıyla sentezlenmiş veri setleri oluşturulmuş ve Faster R-CNN ile RetinaNet modelleri eğitilmiştir. Modeller, COCO değerlendirme metrikleriyle test edilmiş ve RetinaNet'in daha yüksek doğruluk sağladığı gösterilmiştir. Yöntem, çeşitli çözünürlüklerde yüksek performans sergileyerek düşük çözünürlüklü senaryolar için etkili bir çözüm sunmuştur. Hernandez vd. (2021) yaptığı çalışmada, Detectron2, insan-robot etkileşimi bağlamında, özellikle rehabilitasyon süreçlerinde uygulanmıştır. Bu çalışmada, omuz ve dirsek açılarının tespit edilmesi için kullanılmıştır. Detectron2, üst ekstremitte rehabilitasyon hareketlerinin değerlendirilmesinde başarılı sonuçlar elde etmiş olsa da OpenPose gibi modellerle karşılaştırıldığında daha düşük doğruluk gösterdiği tespit edilmiştir.

Li vd. (2024), HRNet'in karmaşık yapısını hafifleterek sınırlı kaynaklara sahip cihazlarda daha verimli kullanılmasını amaçlamışlardır. Özellikle tıbbi rehabilitasyon ve insan-bilgisayar etkileşimine uygulamışlardır. Çalışmada, Convolutional Block Attention Module (CBAM) kullanılarak ağırlıklı korunurken yüksek doğruluk elde edilmiştir. Model, COCO 2017 ve MPII veri setlerinde test edilmiştir ve yüksek performans göstermiştir. Wang vd. (2024) HR-xNet adlı yeni bir yüksek çözünürlüklü poz tahmin algoritması geliştirmişlerdir. HR-xNet'te, HRNet'in özellik çıkarımı değiştirilmiş ve modelin kaynak tüketimi azaltılmıştır. HR-xNet ile düşük seviyeli detaylı özellikler ve yüksek seviyeli anlamsal özellikler birleştirilerek, küçük anahtar noktalar ve değişken insan pozları için kaybolan anahtar noktaların tespit edilmesi sağlanmıştır. Bao vd. (2023) yaptığı çalışmada hareketlerin çok hızlı olduğu skyjump spor dalındaki sporcuların hareketleri analiz edilmiştir. Sporcuların poz tespitinde HRNet'e, Efficient Channel Attention (ECA) modülünü ekleyerek daha doğru anahtar nokta tespiti yapılmıştır. Zhou vd. (2022) yaptığı çalışma, HRNet'in ağırlıklı hesaplama maliyetini azaltmayı amaçlamıştır. X-HRNet, özellikle daha küçük kişiler ve karmaşık sahnelerde ölçek farklarını daha iyi yönetmek için Spatially Unidimensional Self-Attention (SUSA) modüllerini kullanır. Çalışma, COCO val2017 veri seti üzerinde yapılan testlerde, diğer hafif ağırlıklara kıyasla rekabetçi performans sağlamıştır.

Chen vd. (2024), yaşlı bireylerin evde hareketsizliklerini izlemek için bir RGB-D kamera tabanlı sistem geliştirmişlerdir. Özellikle düşük ışık ve TV ışığı değişimleri gibi zorlu koşullarda hareketsizliği tespit etmeyi hedeflemişlerdir. ViTPose ve RAFT yöntemleri kullanılarak hareket algılama gerçekleştirilmiş ve sistem, çevresel değişimlere karşı yüksek doğruluk ve hassasiyet sergilemiştir. Model, çeşitli senaryolarda test edilerek zorlu ev ortamlarında güçlü bir performans ortaya koymuştur. Chen vd. (2024) yaptığı çalışmada ViTPose'i temel alan FanPose isimli bir insan poz tahmin modeli geliştirmiştir. Modelin başarısını arttırmak için ilk olarak giriş verisindeki tüm parçalar arasında daha güçlü bağlantılar kurabilen gelişmiş bir dikkat mekanizması kullanılmıştır. İkinci olarak, anahtar nokta yerelleştirme doğruluğunu artırmak amacıyla geleneksel Gauss çekirdekleri yerine Laplacian çekirdekleri kullanılarak modelin tanıma hassasiyeti iyileştirilmiştir. Xu vd. (2023), Vision Transformer mimarisini kullanarak vücut poz tahmini yapan ViTPose++ modelini geliştirmişlerdir. Top-Down ve Bottom-Up tarama yöntemlerine uyarlanabilen model, 20M ila 1B parametre arasında ölçeklenebilir ve hem insanlar hem de hayvanlar için eklem noktası tespiti sağlayabilmektedir. Model, MAE pre-training ile transfer öğrenmesini geliştirir ve çok görevli öğrenmede başarılıdır. Üstek vd. (2023), akıllı havalimanlarında şiddetli gerçek zamanlı tespit etmek amacıyla ViTPose kullanarak insan vücut duruşlarını analiz eden iki aşamalı bir şiddet tespit çerçevesi sunmuştur. ViTPose ile eklem noktaları tespit edilmiş ve ardından bir CNN-BiLSTM sınıflandırıcı, bu eklemlerden elde edilen mekânsal ve zamansal bilgileri analiz ederek şiddet içeren davranışları sınıflandırmıştır. Sistem, yüksek doğrulukla şiddetli ve şiddetsiz davranışları ayırt edebilmiştir.

Yukarıda özetlenen önceki çalışmalarda da görüldüğü üzere, insan poz tahmini ve hareket analizleri için farklı kütüphaneler ve modeller kullanılmıştır. MediaPipe, MoveNet, OpenPose, AlphaPose, Detectron2 ve HRNet gibi kütüphaneler, insan vücudunun anahtar noktalarının doğru ve hızlı bir şekilde tespit edilmesi için yaygın olarak kullanılmaktadır. Bu kütüphaneler, özellikle sağlık, güvenlik ve spor alanlarında önemli avantajlar sunmakla birlikte, her birinin kendine özgü avantaj ve dezavantajları bulunmaktadır. Bu çalışmada, insan poz tahmini ve hareket analizinde kullanılan kütüphanelerin avantaj ve dezavantajları detaylı bir şekilde incelenmiştir. Kütüphanelerin kullanım alanlarına uygunluğu analiz edilmiştir.

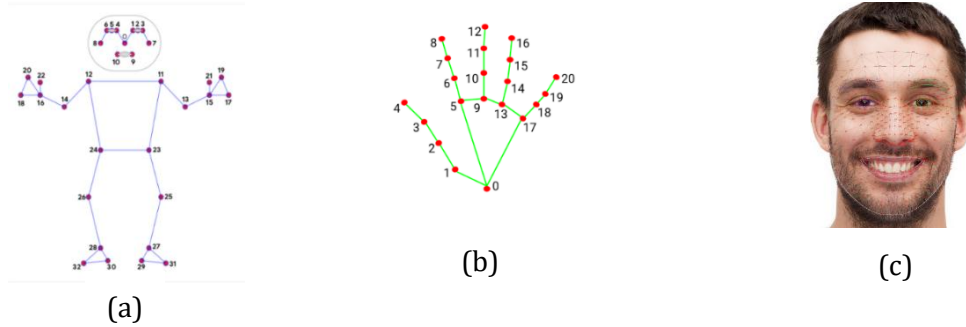
3. POZ TAHMİN YÖNTEMLERİ

3.1. Poz Tahmininde Kullanılan Metotlar

3.1.1. Media Pipe

MediaPipe, Google tarafından geliştirilen, insan vücudu, yüz ve eller gibi çeşitli algılama görevlerini gerçekleştiren açık kaynaklı bir bilgisayarla görme kütüphanesidir. Özellikle gerçek zamanlı poz tahmini ve el izleme uygulamalarıyla öne çıkar. Bu kütüphane, düşük gecikmeli çalışabilmesi sayesinde mobil cihazlar ve gömülü sistemler gibi sınırlı kaynaklara sahip ortamlarda bile yüksek performans sağlar. Derin öğrenme algoritmalarını kullanarak, insan vücudunun 2D ve 3D poz tahmininde önemli bir doğruluk sunar ve özellikle insan-bilgisayar etkileşimi, sağlık, fitness ve güvenlik gibi alanlarda geniş bir kullanım bulur (Google AI, 2024). *MediaPipe*, hızlı ve optimize edilmiş yapısıyla kullanıcıların gerçek zamanlı algılama ve analiz yapmasına olanak tanır. Kütüphane, insan hareketlerinin yakalanması, hareket bozukluklarının teşhisi ve eğitim uygulamaları gibi alanlarda yaygın olarak kullanılır.

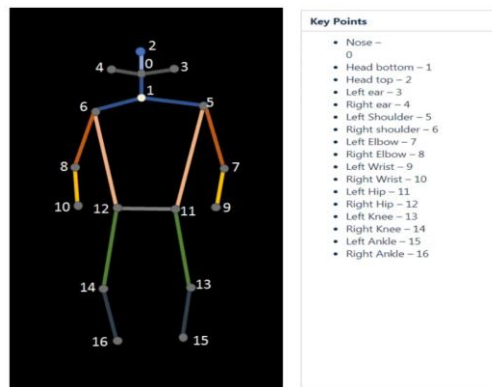
MediaPipe ile tespit edilebilecek yer işaretleri Şekil 1'de gösterilmiştir. Vücut üzerindeki temel eklem noktalarını tespit etmek için *MediaPipe*'in 'Pose' çözümü kullanılabilir. Pose çözümü, kişinin 33 yer işaretini tespit edilebilir. *MediaPipe*'in el algılama çözümü 'Hands' ile bir görüntüdeki eller ve el eklemleri algılanabilir. *MediaPipe*, bir el üzerindeki 21 anahtar noktayı tespit edebilir. Bu anahtar noktaların koordinatları, işaret dili tespitinde kullanılabilir. *MediaPipe*'in 'Face' çözümü, yüz üzerindeki 468 yer işaretini tespit edebilir. *MediaPipe* Holistic, bu 3 modeli birleştirerek kullanılabilir.



Şekil 1. MediaPipe Pose (a), Hand (b) ve Face (c) ile tespit edilen anahtar noktaları (Google AI, 2024)

3.1.2. MoveNet

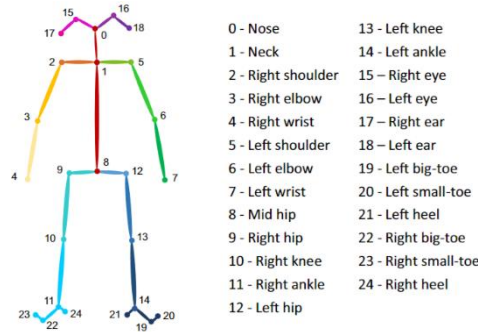
MoveNet, Google tarafından geliştirilen ve gerçek zamanlı insan vücut poz tahmini için kullanılan bir derin öğrenme modelidir. Model, özellikle sınırlı donanım kaynaklarına sahip mobil cihazlar ve gömülü sistemlerde yüksek performanslı ve düşük gecikmeli poz tahmini sağlamak üzere optimize edilmiştir. *MoveNet*, Şekil 2'de gösterilen burun, gözler, kulaklar, omuzlar, dirsekler, el bilekleri, kalçalar, dizler ve ayak bilekleri dahil olmak üzere, insan vücudundaki 17 eklem noktasını tespit edebilir (Grover vd, 2022). Modelin *MoveNet* Lightning ve *MoveNet* Thunder adında iki farklı versiyonu vardır. *Lightning* versiyonu, daha hızlı ve hafif bir modeldir; bu nedenle mobil ve düşük güçlü cihazlar için uygundur. *Thunder* versiyonu ise daha yüksek doğruluk sağlamak için tasarlanmıştır, bu yüzden daha fazla hesaplama gücü gerektirir (TensorFlow, 2024).



Şekil 2. MoveNet, Detectron2, VitPose, AlphaPose, HRNeT tarafından tespit edilebilen vücut anahtar noktaları (Anuj, 2024)

3.1.3. OpenPose

OpenPose, birden fazla kişinin 2D vücut, yüz, el ve ayak anahtar noktalarını (Şekil 3) gerçek zamanlı olarak tespit edebilen açık kaynaklı bir poz tahmin sistemidir. OpenPose'un temel yöntemi, Part Affinity Fields (PAFs) kullanarak vücut parçaları arasındaki ilişkileri öğrenmek ve bu parçaları doğru şekilde bireylere atamaktır. Sistem, her bir kişi için anahtar noktaların konumunu tahmin etmeye dayalı bir yaklaşım kullanır. OpenPose, hem tek kişilik hem de çok kişilik poz tespiti yapabilir ve bu süreçte, CNN türevi olan VGG-19 modeli gibi evrişimli sinir ağlarını kullanarak yüksek doğrulukta sonuçlar elde eder (Gineshidalgo, 2024).



Şekil 3. OpenPose (Temel Model) tarafından tespit edilebilen vücut anahtar noktaları (Zhang vd., 2023)

3.1.4 Detectron2

Detectron2, Facebook AI Research (FAIR) tarafından geliştirilen ve PyTorch tabanlı bir nesne algılama ve segmentasyon kütüphanesidir. Kütüphane, bilgisayarla görü alanında nesne tespiti, örnek tabanlı segmentasyon, semantik segmentasyon ve insan poz tahmini gibi çeşitli görevleri yerine getirmek için kullanılır. İnsan vücudunun eklem noktalarını tespit etmek için, vücut parçalarının konumunu ve oryantasyonunu belirleyen modeller içerir. *Detectron2*, özellikle Mask R-CNN üzerine inşa edilmiş olup, insan vücudundaki 17 anahtar noktayı (örneğin omuzlar, dirsekler, dizler) tespit eder ve bu noktalar üzerinden poz tahmini gerçekleştirir. Tespit edilen vücut eklem noktalarını kullanarak, her eklem için doğru pozisyonunu ve yön bilgisini (örneğin, dirseklerin açısı, dizlerin bükülme durumu) belirleyebilir. Böylece hareketlerin ve vücut duruşunun ayrıntılı bir şekilde analiz edilmesini sağlar (Chatterjee vd., 2021). Gerçek zamanlı uygulamalarda yüksek doğruluk ve hız sunan bu sistem, spor analizi, insan-robot etkileşimi ve video analizinde yaygın olarak kullanılır (Lyttonhao, 2024).

3.1.5. ViTPose

ViTPose, Vision Transformer (ViT) tabanlı bir insan poz tahmin modelidir ve insan pozlarını yüksek doğrulukla tespit etmek için ViT mimarisini kullanır. Geleneksel CNN yaklaşımlarından farklı olarak, *ViTPose* geniş veri setleri üzerinde eğitilmiştir ve Transformer yapısının sunduğu geniş bağlam bilgisi sayesinde özellikle karmaşık sahnelerde başarılı performans gösterir. Temel model, insan vücudundaki 17 eklem noktasını tespit edebilir (Xu vd., 2022).

3.1.6. AlphaPose

AlphaPose, insan vücudu poz tahmini alanında kullanılan açık kaynaklı bir kütüphanedir ve özellikle çoklu kişilerin eklem noktalarını yüksek doğrulukla tespit etmek amacıyla geliştirilmiştir. *AlphaPose*, derin öğrenme tabanlı bir yöntem olarak, karmaşık insan pozlarının tespitinde üstün performans sunar ve bu süreçte hem vücut eklemlerinin konumlarını hem de bu eklemler arasındaki ilişkileri dikkate alarak poz tahmini gerçekleştirir. Kütüphane, insan eklemlerinin yüksek çözünürlükte tahmin edilmesini sağlayan "Spatio-Temporal Affinity Fields" gibi yenilikçi teknikler kullanır (Fang vd., 2022). *AlphaPose*, çoklu kişi tespitinde gerçek zamanlı performans sunarak geniş bir uygulama yelpazesinde kullanılabilir; bu uygulamalar arasında video analitiği, insan-bilgisayar etkileşimi ve spor analizleri yer alır. Sağladığı yüksek doğruluk ve esneklik, onu hem akademik araştırmalar hem de endüstriyel projeler için tercih edilen bir araç haline getirmektedir.

3.1.7. HRNet

HRNet (High-Resolution Network), insan pozunu tespiti konusunda yüksek çözünürlüklü temsil öğrenmesi sağlayan ve derin öğrenme tekniklerini kullanan bir modeldir. *HRNet*'in en büyük avantajı, yüksek çözünürlüklü temsili modelin her aşamasında koruyarak, özellikle insan poz tespitinde yüksek doğruluk ve hassasiyet sunmasıdır. Bu model, MS-COCO gibi geniş veri setlerinde test edilmiş ve derinlikli öğrenme tabanlı yaklaşımlar arasında öne çıkmıştır (Wang vd., 2020). *HRNet*, insan poz tahmini yanında semantik segmentasyon ve yüz tespiti gibi farklı bilgisayarlı görü problemlerinde de başarıyla kullanılmış ve literatürde bu alanda yeni bir standart oluşturmuştur

(Sun vd., 2019). Sağladığı yüksek doğruluk ve verimlilik, HRNet'i hem akademik araştırmalarda hem de endüstriyel uygulamalarda tercih edilen bir model haline getirmiştir.

3.2. Poz Tahmininde Kullanılan Tarama Modları

İnsan vücudu eklem noktalarının tespiti için kullanılan Top-Down ve Bottom-Up yöntemleri, iki farklı yaklaşımla poz tahmini yapmaktadır. Top-Down yönteminde önce görüntüdeki bireyler tespit edilir, ardından her birey için poz tahmini yapılır. Bu yöntem yüksek doğrulukla çalışır ve karmaşık pozlarda başarılıdır. Fakat yüksek hesaplama maliyeti ve kalabalık sahnelerde yavaş performans gösterir. Bottom-Up yönteminde ise tüm eklem noktaları aynı anda algılanır, ardından bu eklem noktalarına göre kişi oluşturulur. Bu yöntem hızlı ve verimli olup, kalabalık sahnelerde iyidir. Ancak daha düşük doğruluk ile tahmin yapmasına ve karmaşık pozlarda hatalar yapmasına neden olur (Fang vd, 2021). Çalışmanın, Bulgular ve Tartışma bölümünde poz tahmininde kullanılan metotların hangi tarama modlarını kullandığı sunulmaktadır.

4. BULGULAR VE TARTIŞMA

Tablo 1'de AlphaPose, ViTPose, Detectron2, MoveNet, OpenPose, MediaPipe Holistic, ve HRNet modelleri eklem noktası sayısı, tarama modu, programlama dili, eklem noktası üst üste binme oranı ve gerçek zamanlı performans (FSP) gibi kriterler açısından karşılaştırılmıştır.

Eklem noktası sayısı, modellerin tespit edebildiği insan vücudu eklemlerinin sayısını ifade eder ve el, yüz, vücut gibi bölümlere göre hassasiyet gösterir. AlphaPose, VitPose, Detectron2, MoveNet ve HRNet modelleri temel olarak COCO veriseti ile eğitilmiştir. COCO verisetinde Şekil 2'de gösterildiği gibi 17 vücut anahtar noktası bulunmaktadır. OpenPose modeli, COCO formatındaki 17 anahtar noktanın ötesine geçerek, daha fazla anahtar noktayı tespit eden bir model sunar. Body_25 adı verilen bu genişletilmiş format, insan vücudunun daha detaylı bir iskelet yapısını sağlar (Fang vd, 2022). MediaPipe Holistic çözümünde 3 model birleştirilmiştir. Bu sayede insan vücudunda tespit edilebilen anahtar nokta sayısı diğer modellere göre daha fazladır. MediaPipe Pose, temel vücut noktalarını tespit eder. Tespit edilen bilek ve yüz noktalarına göre görüntü yeniden boyutlandırılır. Tespit edilen bilek ve yüz noktalarından itibaren görüntü yeniden boyutlandırılır. El ve yüz anahtar noktaları, MediaPipe Hand ve MediaPipe Face kullanılarak yeniden boyutlandırılan görüntü içinde tespit edilmeye çalışılır. MediaPipe Holistic, yeniden boyutlandırma işlemi, nesnenin çerçeveler arasında önemli ölçüde hareket etmediğini varsayar ve mevcut çerçevedeki nesne bölgesine bir kılavuz olarak önceki çerçeveden elde edilen tahmini kullanır. Bu da tepki süresinin kılmasını sağlar. Aynı zamanda, çerçevedeki bir kişinin sol ve sağ ellerinin ya da vücut bölümlerinin karışmasını önleyerek modelin gövde ve parçaları arasında anlamsal tutarlılığı korur.

Tarama modu, Top-Down ve Bottom-Up olmak üzere ikiye ayrılır. Top-Down'da önce birey tespit edilir, ardından poz tahmini yapılır. Bottom-Up'da tüm eklem noktaları aynı anda bulunur ve kişilere atanır. AlphaPose, ViTPose ve MoveNet modelleri Top-Down tarama modunu, OpenPose ise Bottom-Up tarama modunu kullanır. Detectron2 ve MediaPipe Holistic modelleri iki tarama modu arasında geçiş yapabilir. Bu da yapılacak çalışmanın özelliğine göre tarama modunu değiştirerek uygulamanın performansının artmasına fayda sağlar.

Programlama dili, modellerin performans ve kullanım kolaylığı açısından önemlidir. Tüm modeller Python programlama dilini desteklemektedir. Python'un geniş kullanıcı tabanı, zengin kütüphane ekosistemi ve makine öğrenimi alanındaki yaygın kullanımı, bu modellerin hızlı prototipleme, kolay entegrasyon ve geniş kapsamlı uygulamalarda kullanılmasını mümkün kılmaktadır. Derin öğrenme ve bilgisayarla görme kütüphaneleri olan TensorFlow, PyTorch ve OpenCV gibi popüler framework'lerin Python tabanlı olması, bu modellerin geliştirilme ve uygulanma süreçlerini hızlandırmaktadır. Ayrıca, Python'un öğrenimi kolay bir dil olması, araştırmacıların ve mühendislerin karmaşık algoritmaları daha hızlı bir şekilde test etmelerine ve uygulamalarına olanak sağlamaktadır. Geniş topluluk desteği, güçlü dökümantasyon kaynakları ve veri işleme, analiz ile görselleştirme araçlarının Python'da yaygın olması, vücut anahtar nokta tahmin modellerinin etkin bir şekilde kullanılmasını ve gerçek zamanlı uygulamalara entegrasyonunu kolaylaştırmaktadır. Bu bağlamda, Python'un sağladığı bu esneklik ve gücü, derin öğrenme tabanlı insan poz tahmini projelerinde geniş bir kullanım alanı yaratmaktadır.

Eklem noktası üst üste binme değeri, kalabalık sahnelerde veya eklem noktalarının üst üste geldiği durumlarda anahtar noktaların doğru atanıp atanmadığını gösterir. Örneğin, bir kişinin kolları üst üste geldiğinde, dirseklerden bileklere doğru bağlantının tespit edilmesi modelin başarısını gösterir. Üst üste binme değeri yüksek olduğunda modeldeki üst üste binen eklem noktalarının tespitinin başarısız olduğu, düşük olduğunda ise üst üste inen eklem noktalarının tespitinin başarılı olduğunu gösterir.

FPS, modellerin saniyede işlediği kare sayısını belirler ve gerçek zamanlı uygulamalarda kritik bir ölçüttür. Yüksek FPS, saniyede daha fazla görüntü karesinin işlenmesini sağlayarak, akıcı ve doğru hareket algılamayı mümkün kılar. Bu durum, özellikle spor analizi, sağlık uygulamaları ve insan-bilgisayar etkileşimi gibi gerçek zamanlı geri bildirim

gerektiren alanlarda büyük avantaj sağlar. Düşük FPS, hareketlerin kesik kesik algılanmasına, gecikmeye ve doğruluk kaybına yol açarak hem uygulamanın performansını hem de kullanıcı deneyimini olumsuz etkileyebilir. Yüksek FPS, bu nedenle vücut anahtar noktalarının tahmini modellerinin çeşitli uygulama alanlarında etkin kullanımını belirleyen önemli bir faktördür.

Karşılaştırması yapılan modeller arasında MediaPipe Holistic, el ve parmak eklemlerini yüksek doğrulukla tespit edebilmesi ve düşük donanım gereksinimleri sayesinde işaret dili tanıma uygulamalarında öne çıkmaktadır. Bu model, mobil cihazlar gibi sınırlı kaynaklarda dahi hızlı ve gerçek zamanlı performans sağlayarak işaret dili uygulamaları için büyük bir potansiyele sahiptir. AlphaPose, vücut, el, yüz ve ayak gibi çok sayıda eklem noktasını hassas bir şekilde tespit edebildiği için daha karmaşık işaret dili ifadelerini tanımada avantaj sunmaktadır. OpenPose, genişletilmiş iskelet yapısı (Body_25) sayesinde, özellikle işaret dili ifadelerinde el hareketlerinin ve parmak pozisyonlarının ayrıntılı analizinde güçlü bir araç olarak değerlendirilebilir. Bu bağlamda, işaret dili tanıma uygulamaları için seçilecek modelin hem hareketlerin doğruluğu hem de uygulama ortamının kaynak kısıtlamaları göz önüne alınarak belirlenmesi önemlidir.

Tablo 1. Modellerin Karşılaştırılması

Model	Eklem Noktası Sayısı	Tarama Modu	Programlama Dili	Eklem Noktası Üst Üste Binme	FPS	Donanım Gereksinimi	Kullanım Alanları
AlphaPose	17	Top-Down	Python, C++	Orta	10-15	GPU (NVIDIA), CPU	Spor analizi, hareket tespiti (Hu, 2024)
ViTPose	17 (temel model)	Top-Down	Python	Yüksek	5-10	Yüksek performanslı GPU (NVIDIA)	Poz tahmini, insan-bilgisayar etkileşimi (Yu, 2024)
Detectron2	17	Top-Down Bottom-Up	Python	Yüksek	15-20	GPU (NVIDIA), CPU	Nesne ve insan poz tespiti, güvenlik uygulamaları (Abu Awwad, 2024)
MoveNet	17	Top-Down	Python	Orta	30	Düşük-orta GPU, mobil cihaz	Gerçek zamanlı poz tahmini, mobil uygulamalar (Parle, 2024)
OpenPose	25	Bottom-Up	Python, C++	Yüksek	10-15	Yüksek performanslı GPU (NVIDIA)	Çok kişili poz tespiti, insan hareket analizi (Gineshidalgo, 2024)
MediaPipe Holistic	543 (el, yüz, vücut)	Top-Down Bottom-Up	Python	Düşük	30	Mobil cihazlar, CPU	İşaret dili tanıma, insan hareket analizi, artırılmış gerçeklik (Singh, 2021)
HRNet	17	Bottom-Up	Python	Yüksek	15-20	Yüksek performanslı GPU (NVIDIA)	Karmaşık poz tespiti, insan-bilgisayar etkileşimi, tıp (Lin, 2024)

5. SONUÇLAR

MoveNet ve MediaPipe Holistic, gerçek zamanlı performans gerektiren uygulamalarda öne çıkmaktadır. MoveNet, düşük donanım gereksinimleri ve yüksek FPS ile mobil cihazlarda hızlı egzersiz takibi ve sağlık uygulamaları için idealdir. MediaPipe Holistic ise vücut, yüz ve el hareketlerini aynı anda takip edebilmesi sayesinde işaret dili tanıma gibi uygulamalarda büyük avantaj sağlar. Hem üst hem de alt vücut hareketlerini doğru bir şekilde algılayarak çok yönlü çözümler sunar ve düşük donanım gereksinimiyle geniş bir kullanım alanına sahiptir. Daha karmaşık ve yüksek doğruluk gerektiren uygulamalarda, özellikle spor analizi, fizyoterapi ve tıp alanlarında, AlphaPose,

ViTPose ve HRNet gibi modeller tercih edilmelidir. AlphaPose ve HRNet, çok kişili sahnelerde yüksek doğrulukla çalışırken, ViTPose, Vision Transformer mimarisi sayesinde bireysel egzersiz ve insan-bilgisayar etkileşimi projelerinde etkili bir performans sunar. Yüksek doğruluk ve işlevsellik gerektiğinde bu modellerin kullanımı önerilmektedir. Detectron2, nesne tanıma ve poz tahmini görevlerini aynı anda gerçekleştirebilen, daha genel amaçlı bir modeldir. Güvenlik uygulamaları, nesne algılama, spor ve sağlık analizlerinde yer bulabilse de karmaşık sahnelerde düşük FPS değeri nedeniyle gerçek zamanlı performansta dezavantaj oluşturabilir. İşaret dili tanıma uygulamalarında ise her modelin farklı avantajlar sunduğu görülmektedir. MediaPipe Holistic, el ve parmak eklemlerini yüksek doğrulukla tespit etme yeteneği ve düşük donanım gereksinimi ile mobil ve gerçek zamanlı işaret dili uygulamaları için idealdir. MediaPipe Pose'un bilek noktasını tespit ederek elin konumunu belirlemesi ve ardından el üzerindeki anahtar noktaları MediaPipe Hand çözümü ile yüksek doğrulukla analiz etmesi, bu kütüphaneyi işaret dili tanıma sistemleri için öne çıkarır. AlphaPose, el ve yüz hareketlerinin ayrıntılı analizini gerektiren karmaşık işaret dili ifadelerinde yüksek doğruluk sağlar ve sabit sistemlerde ya da yüksek işlem gücüne sahip ortamlarda etkili bir çözüm sunar. OpenPose, genişletilmiş iskelet yapısı sayesinde çok kullanıcı senaryolarında el ve parmak hareketlerini hassas bir şekilde analiz edebilir.

Sonuç olarak, gerçek zamanlı ve mobil uyumlu projeler için MoveNet veya MediaPipe Holistic önerilirken, yüksek doğruluk gerektiren projeler için AlphaPose, ViTPose veya HRNet tercih edilmelidir. Projenin gereksinimlerine göre uygun modeli seçmek hem verimlilik hem de doğruluk açısından istenilen sonuçları elde etmeye yardımcı olacaktır.

Teşekkür

Bu çalışma, Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) tarafından 124E379 numaralı hibe kapsamında desteklenmiştir. Yazarlar destekleri için TÜBİTAK'a teşekkürlerini sunarlar.

Çıkar Çatışması Beyanı

Makale yazarları aralarında herhangi bir çıkar çatışması olmadığını beyan ederler.

Yazar Katkısı Beyanı

Yazar-1: Fikir, Orijinal Taslak Oluşturma, Yazım, Revize, Metot Oluşturma, Verilerin Düzenlenmesi, Sonuç ve Tartışma

Yazar-2: Revize, Verilerin Düzenlenmesi, Sonuç ve Tartışma

KAYNAKLAR

- Abu Awwad, Y., Rana, O., & Perera, C. (2024). Anomaly detection on the edge using smart cameras under low-light conditions. *Sensors*, 24(3), 772.
- Areerob, P., Matangkasombut, T., Monnikhof, K. O., & Kumwilaisak, W. (2024, May). Crowded Scene PPE Detection Using Attention Based YOLOv7 and Alpha Pose. In 2024 21st International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON) (pp. 1-6). IEEE.
- Arkushin, R. S., Moryossef, A., & Fried, O. (2023). Ham2pose: Animating sign language notation into pose sequences. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 21046-21056).
- Anuj Shah. (2024, Eylül 10). Posetrack data set summary. https://medium.com/@anuj_shah/posetrack-data-set-summary-9cf61fc6f44e
- Badiola-Bengoia, A., & Mendez-Zorrilla, A. (2021). A systematic review of the application of camera-based human pose estimation in the field of sport and physical exercise. *Sensors*, 21(18), 5996.
- Bao, W., Niu, T., Wang, N., & Yang, X. (2023). Pose estimation and motion analysis of ski jumpers based on ECA-HRNet. *Scientific Reports*, 13(1), 6132.
- Bibin Sebastian. (2024, Eylül 10). Human Action Recognition using Detectron2 and LSTM. <https://learnopencv.com/human-action-recognition-using-detectron2-and-lstm/>
- Bora, J., Dehingia, S., Boruah, A., Chetia, A. A., & Gogoi, D. (2023). Real-time assamese sign language recognition using mediapipe and deep learning. *Procedia Computer Science*, 218, 1384-1393.

- 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 (pp. 7291-7299).
- Chatterjee, R., Roy, S., Islam, S. H., & Samanta, D. (2021, August). An AI Approach to Pose-based Sports Activity Classification. In 2021 8th International Conference on Signal Processing and Integrated Networks (SPIN) (pp. 156-161). IEEE.
- Chen, L., & Fisher, R. B. (2024). Miso: Monitoring inactivity of single older adults at home using rgb-d technology. *ACM Transactions on Computing for Healthcare*, 5(3), 1-19.
- Chen, M., & Tan, G. (2024, July). FANpose: 2D human pose estimation with fully attentional networks under vision transformer baselines. In Third International Symposium on Computer Applications and Information Systems (ISCAIS 2024) (Vol. 13210, pp. 879-884). SPIE.
- Davoudi Kashkoli, M., Javied, A., Barrera-Animas, A. Y., & Davila Delgado, J. M. (2024). A synthetic data approach for object detection in super low resolution images. Proceedings of the 2024 International Conference on Innovation in Artificial Intelligence (ICIAI '24), 86-91.
- Dill, S., Rösch, A., Rohr, M., Güney, G., De Witte, L., Schwartz, E., & Hoog Antink, C. (2023, September). Accuracy Evaluation of 3D Pose Estimation with MediaPipe Pose for Physical Exercises. In Current Directions in Biomedical Engineering (Vol. 9, No. 1, pp. 563-566). De Gruyter.
- Fang, H. S., Li, J., Tang, H., Xu, C., Zhu, H., Xiu, Y., ... & Lu, C. (2022). Alphapose: Whole-body regional multi-person pose estimation and tracking in real-time. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(6), 7157-7173.
- Fang, Y., Han, Z., Hu, Z., & Wang, Z. (2021, December). Human Posture Estimation. In 2021 3rd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI) (pp. 220-234). IEEE.
- Garg, S., Saxena, A., & Gupta, R. (2023). Yoga pose classification: a CNN and MediaPipe inspired deep learning approach for real-world application. *Journal of Ambient Intelligence and Humanized Computing*, 14(12), 16551-16562.
- Gao, Q., Liu, J., Ju, Z., & Zhang, X. (2019). Dual-hand detection for human-robot interaction by a parallel network based on hand detection and body pose estimation. *IEEE Transactions on Industrial Electronics*, 66(12), 9663-9672.
- Gineshidalgo. (2024, Ekim 20). OpenPose: Real-time multi-person keypoint detection library for body, face, hands, and foot estimation. <https://github.com/CMU-Perceptual-Computing-Lab/openpose>
- Google AI. (2024, Eylül 10). MediaPipe Solutions guide. <https://ai.google.dev/edge/mediapipe/solutions/guide>
- Grover, A., Arora, D., & Grover, A. (2022, December). Human pose estimation using deep learning techniques. In Proceedings of the 4th International Conference on Information Management & Machine Intelligence (pp. 1-6).
- Hernández, Ó. G., Morell, V., Ramon, J. L., & Jara, C. A. (2021). Human pose detection for robotic-assisted and rehabilitation environments. *Applied Sciences*, 11(9), 4183.
- Hu, M., Zhang, M., & Yu, K. (2024). Design of sports training information analysis system based on a multi-target visual model under sensor-scale spatial transformation. *PeerJ Computer Science*, 10, e2030.
- Jafarzadeh, P., Virjonen, P., Nevalainen, P., Farahnakian, F., & Heikkonen, J. (2021, October). Pose estimation of hurdles athletes using openpose. In 2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME) (pp. 1-6). IEEE.
- Jo, B., & Kim, S. (2022). Comparative analysis of OpenPose, PoseNet, and MoveNet models for pose estimation in mobile devices. *Traitement du Signal*, 39(1), 119.
- Karacı, A., Akyol, K., & Turut, M. U. (2021). Real-Time Turkish Sign Language Recognition Using Cascade Voting Approach with Handcrafted Features. *Applied Computer Systems*, 26(1), 12-21.
- Kim, J. W., Choi, J. Y., Ha, E. J., & Choi, J. H. (2023). Human pose estimation using mediapipe pose and optimization method based on a humanoid model. *Applied sciences*, 13(4), 2700.
- Lee, H., Oh, B., & Kim, S. C. (2024). Recognition of Forward Head Posture Through 3D Human Pose Estimation With a Graph Convolutional Network: Development and Feasibility Study. *JMIR Formative Research*, 8(1), e55476.

- Li, R., Yan, A., Yang, S., He, D., Zeng, X., & Liu, H. (2024). Human Pose Estimation Based on Efficient and Lightweight High-Resolution Network (EL-HRNet). *Sensors*, 24(2).
- Li, N., Wang, Y., Liu, F., & Huang, W. (2024, August). Real-time multitarget fall detection based on OpenPose. In *Seventh International Conference on Advanced Electronic Materials, Computers, and Software Engineering (AEMCSE 2024)* (Vol. 13229, pp. 751-755). SPIE.
- Lin, H., Chen, H., & Lin, J. (2024). Deep neural network uncertainty estimation for early oral cancer diagnosis. *Journal of Oral Pathology & Medicine*, 53(5), 294-302.
- Lyttonhao. (2024, Ekim 15). Detectron2. <https://ai.meta.com/tools/detectron2/>
- Mishra, A. K., Sahoo, D., Shubhankar, I., & Samal, I. YogaSiddhi: AI-Powered Pose Analysis using MoveNet for Yoga Refinement. *International Journal of Computer Applications*, 975, 8887.
- Parashar, D., Mishra, O., Sharma, K., & Kukker, A. (2023). Improved Yoga Pose Detection Using MediaPipe and MoveNet in a Deep Learning Model. *Revue d'Intelligence Artificielle*, 37(5).
- Parle, A., Shinde, R., Chougule, R., & Agrawal, S. (2024, April). YogaWise: Enhancing Yoga with Intelligent Real Time Tracking using TensorFlow MoveNet. In *2024 International Conference on Cognitive Robotics and Intelligent Systems (ICC-ROBINS)* (pp. 498-505). IEEE.
- Singh, A. K., Kumbhare, V. A., & Arthi, K. (2021, June). Real-time human pose detection and recognition using mediapipe. In *International conference on soft computing and signal processing* (pp. 145-154). Singapore: Springer Nature Singapore.
- Sun, K., Xiao, B., Liu, D., & Wang, J. (2019). Deep high-resolution representation learning for human pose estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5693-5703).
- Song, H., Li, Y., Fu, C., Xue, F., Zhao, Q., Zheng, X., ... & Liu, T. (2024). Using complex networks and multiple artificial intelligence algorithms for table tennis match action recognition and technical-tactical analysis. *Chaos, Solitons & Fractals*, 178, 114343.
- TensorFlow. (2024, Ekim 1). MoveNet: Ultra fast and accurate pose detection model. <https://www.tensorflow.org/hub/tutorials/movenet?hl=tr>
- Üstek, İ., Desai, J., Torrecillas, I. L., Abadou, S., Wang, J., Fever, Q., ... & Tsoydos, A. (2023, August). Two-Stage Violence Detection Using ViTPose and Classification Models at Smart Airports. In *2023 IEEE Smart World Congress (SWC)* (pp. 797-802). IEEE.
- Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., ... & Xiao, B. (2020). Deep high-resolution representation learning for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 43(10), 3349-3364.
- Wang, Y., Wang, R., Shi, H., & Liu, D. (2024). MS-HRNet: multi-scale high-resolution network for human pose estimation. *The Journal of Supercomputing*, 1-23.
- Wu, M. Y., Ting, P. W., Tang, Y. H., Chou, E. T., & Fu, L. C. (2020). Hand pose estimation in object-interaction based on deep learning for virtual reality applications. *Journal of Visual Communication and Image Representation*, 70, 102802.
- Wu, Q., Xu, G., Zhang, S., Li, Y., & Wei, F. (2020, July). Human 3D pose estimation in a lying position by RGB-D images for medical diagnosis and rehabilitation. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp. 5802-5805). IEEE.
- Xu, Y., Zhang, J., Zhang, Q., & Tao, D. (2022). Vitpose: Simple vision transformer baselines for human pose estimation. *Advances in Neural Information Processing Systems*, 35, 38571-38584.
- Xu, Y., Zhang, J., Zhang, Q., & Tao, D. (2023). Vitpose++: Vision transformer for generic body pose estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Yu, N., Ma, T., Zhang, J., Zhang, Y., Bao, Q., Wei, X., & Yang, X. (2024, October). Adaptive Vision Transformer for Event-Based Human Pose Estimation. In *Proceedings of the 32nd ACM International Conference on Multimedia* (pp. 2833-2841).

- Zhang, F., Juneau, P., McGuirk, C., Tu, A., Cheung, K., Baddour, N., & Lemaire, E. (2021, June). Comparison of OpenPose and HyperPose artificial intelligence models for analysis of hand-held smartphone videos. In 2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA) (pp. 1-6). IEEE.
- Zhang, M., Zhou, Y., Xu, X., Ren, Z., Zhang, Y., Liu, S., & Luo, W. (2023). Multi-view emotional expressions dataset using 2D pose estimation. *Scientific Data*, 10(1), 649.
- Zhou, Y., Wang, X., Xu, X., Zhao, L., & Song, J. (2022, July). X-hrnet: Towards lightweight human pose estimation with spatially unidimensional self-attention. In 2022 IEEE international conference on multimedia and expo (ICME) (pp. 01-06). IEEE.
- Zhu, H., Jie, C., & Jiang, S. (2020). Multi-Person Full Body Pose Estimation. arXiv preprint arXiv:2008.10060.