Review on Human Pose Estimation using AI Fitness Tracker

Prince Gupta Merin Reji Rohit Vishwakarma

Department of Information Technology

Xavier Institute of Engineering

Abstract

Human pose estimation, a fundamental computer vision task, has witnessed significant advancements with the advent of deep learning techniques. This paper provides an overview of recent developments in human pose estimation, focusing on the application of convolutional neural networks (CNNs) and recurrent neural networks (RNNs). We explore the distinctions between 2D and 3D pose estimation, highlighting the challenges and solutions for each. The paper discusses the various techniques for single-person and multi-person pose estimation, as well as the trade-offs between part-based and holistic approaches. Additionally, we delve into the real-time and offline aspects of pose estimation and the significance of temporal analysis in videos. Lastly, we touch on the translation of 2D pose information into 3D space. This review underscores the practical implications of human pose estimation in diverse fields such as sports analysis, healthcare, surveillance, and entertainment. The dynamic nature of this field, fueled by ongoing research and technological advancements, promises to unlock new possibilities for human-computer interaction, augmented reality, and beyond. This abstract provides an overview of the paper's focus on deep learning in the context of human pose estimation and highlights the key aspects of the field. You can further tailor it to your specific paper or presentation needs.

Keywords:

Human pose estimation, camera, Artificial intelligence, Machine learning

1. Introduction

The field of computer vision explores human body position estimation, addressing challenges like complex body structures and varying appearances due to factors such as clothing and lighting. Applications like smart gym trainers aim to create posture estimation systems for users, providing feedback for workout improvement. Human pose recovery is crucial for applications like video indexing and human-computer interaction. Challenges include degrees of freedom, appearance variations, and occlusions. Methods, categorized as model-based or model-free, employ techniques such as 3D estimates and address challenges in dense environments. In human posture estimation, deep learning has revolutionized the approach. Using Deep Neural Networks (DNNs), joint regression is performed, capturing the entire context more efficiently than traditional models. A cascade of DNN-based posture predictors enhances joint localization precision, offering significant breakthroughs in the field. Recovering both 2D and 3D human poses from images is crucial for various applications, including action identification and motion capture. Convolutional Neural Networks (CNNs) have shown success in 2D pose estimation, handling challenges like occlusion. However, extending CNNs to 3D pose estimation faces difficulties due to a lack of depth information. Integrating both 2D and 3D pose information can improve precision. In the realm of computer vision, 3D pose estimation encounters challenges like a wider pose space and perspective projection issues. Model-based generative methods, such as the pictorial structural model (PSM), and discriminative methods are two prominent approaches. Discriminative methods, using deep neural networks like CNNs, have shown effectiveness in extracting features and addressing challenges in 2D pose estimation. Despite the prevalence of 3D pose estimation from depth images, the focus remains on 2D estimation due to its common occurrence in visual media. The proposed multi-task architecture, combining CNNs with regression and auxiliary tasks, enhances convergence and outperforms networks solely trained for regression tasks.

Gait and postural issues are on the rise due to factors like aging, cardiovascular problems, and neurological illnesses, leading to stability loss and an increased risk of falls. Traditional rehabilitation methods are time-consuming and face clinical assessment heterogeneity. Robotic-based rehabilitation, utilizing smart walkers like ASBGo, aims to enhance the quality of life for individuals with motor impairments. This research introduces a deep learning-based method for real-time, accurate, and lightweight full-body human position assessment through the ASBGo smart walker, addressing limitations in current smart walker systems for patient monitoring and rehabilitation assessment. Gait analysis is crucial for understanding and detecting pathological gait issues, impacting overall health. Previous studies explored various gait analysis tools, and this research proposes a unique technique that combines human position estimation with Convolutional Neural Networks (CNN) for categorizing normal and pathological gait. By utilizing CNNs for pose estimation, this approach generates skeletal images without sensors, mitigating challenges related to view angle, walking speed, and clothing. The study begins with an overview of the system design and demonstrates improved accuracy on the Human3.6m dataset compared to existing methods. Estimating 3D human pose from single images faces challenges such as occlusions and depth information loss. This paper introduces a novel approach, utilizing an overcomplete auto-encoder to project body joint positions into a high-dimensional space. A CNN-based mapping is learned from input images to this pose representation, enhancing accuracy by considering dependencies. This combined approach outperforms current state-of-the-art methods on the Human3.6m dataset.

The objective of 2D human pose estimation from images is crucial for various applications, facing challenges like clothing variations, dynamic backgrounds, lighting changes, and occlusions. Deterministic convolutional neural networks, such as Hourglass networks, are accurate but lack uncertainty management. This research explores three probabilistic models for human pose prediction, utilizing DISCO networks and Bayesian SegNet frameworks. Proposed modifications show an improvement in total accuracy over base models on benchmarks like LSP and MPII Human Pose. Advancements in 3D pose reconstruction from videos involve integrating attention processes to enhance accuracy and maintain temporal coherence. This method, categorized as 2D-to-3D estimation, employs attention-based models and dilated convolutions to learn implicit dependencies. The experimental results demonstrate competitive precision, making it suitable for real-time applications like computer gaming and avatar animation retargeting. The approach is exemplified by constructing 3D avatars from various video sources and showcasing automated pose extraction and motion retargeting in performance-based animations. The research addresses challenges in human pose estimation, emphasizing the transition from marker-based to markerless pose estimation. While newer CNN-based methods excel in 2D pose estimation, 3D pose estimation using depth/3D data with CNNs faces constraints. The proposed 3D-CNN architecture provides a fully convolutional, detection-based solution for efficient 3D human pose estimation over the depth sensor's effective Field-of-View. This method enables global context use and extends to multi-person pose estimation, introducing a sequential network architecture for per-voxel likelihood maps and a unique multi-person technique for simultaneously estimating multiple human positions.

The study addresses the challenge of estimating 3D human pose from single 2D images by introducing an innovative approach that combines autoencoders and convolutional neural networks (CNNs) within a deep learning framework. This integration enhances accuracy without relying on computationally expensive optimization. Leveraging reliable 2D joint location heatmaps as input, the model demonstrates top-tier performance, especially with the ResNet architecture. Additionally, the method is extended to handle image sequences through LSTM-based architectures, showcasing improved outcomes. In summary, the paper contributes by merging conventional CNNs with autoencoders, leading to enhanced 3D human pose estimation performance. In another context, the research focuses on human posture estimation in computer vision, emphasizing the transition from marker-based to markerless pose estimation. While newer CNN-based methods excel in 2D posture estimation, 3D pose estimation using depth/3D data with CNNs faces constraints. The proposed 3D-CNN architecture offers a fully convolutional, detection-based solution for efficient 3D human pose estimation over the depth sensor's effective Field-of-View. This method enables global context use and extends to multi-person pose estimation, introducing a sequential network architecture for per-voxel likelihood maps and a unique multi-person technique for simultaneously estimating several human positions. Furthermore, the work addresses the challenge of 6D object posture estimation crucial in robotic activities and human-robot interaction. The proposed PoseCNN architecture, a Convolutional Neural Network, explicitly models dependencies, predicts object labels, 2D pixel coordinates for the center, and calculates object distance. To handle symmetries, a ShapeMatch-Loss function is introduced, emphasizing matching 3D forms. PoseCNN demonstrates cutting-edge performance on datasets like OccludedLINEMOD and introduces the YCB-Video dataset for further testing, contributing a robust CNN model, a unique loss function, and a rich dataset for 6D object pose estimation.

The research explores cutting-edge advancements in computer vision, addressing challenges across diverse applications. In the realm of 6D object pose estimation, the proposed PoseCNN introduces a novel Convolutional Neural Network (CNN) architecture, explicitly modeling dependencies to predict object labels, 2D pixel coordinates, and object distance. To overcome challenges with texture-free and symmetrical objects, a unique ShapeMatch-Loss function is devised, achieving state-of-the-art performance on datasets like OccludedLINEMOD. In the domain of 3D hand pose estimation, the study pioneers a real-time 3D CNN-based approach, effectively handling posture variations and self-occlusion by encoding the 3D hand point cloud with Directional Truncated Signed Distance Function (D-TSDF) values. This methodology outperforms prior approaches in terms of 3D feature learning, achieving real-time processing, and demonstrating robustness to changes in hand size and orientation. For 3D human pose estimation, the MargiPose model introduces a fresh perspective by employing 2D marginal heatmaps for joint location predictions, offering enhanced memory efficiency compared to volumetric heatmaps. With its state-of-the-art performance on datasets like MPI-INF-3DHP and competitive results on Human3.6M, MargiPose proves effective in 3D posture estimation. The incorporation of regularized soft-argmax further improves accuracy and visual coherence, showcasing the model's efficacy. The CNN architecture of MargiPose strategically uses axis permutation to address discrepancies between input and output domains. These methodologies collectively contribute innovative solutions to advance computer vision tasks, from object and hand pose estimation to efficient 3D human pose estimation.

Human posture estimation (HPE) plays a crucial role in computer vision applications, providing geometric and kinematic information from sensor-captured data like images and videos. Deep learning has significantly advanced HPE, especially in 2D tasks, surpassing traditional computer vision methods. Despite progress, challenges such as occlusion, limited training data, and depth ambiguity persist. In the context of 3D HPE, obtaining accurate annotations remains challenging, particularly in real-world environments. Optical skeletal motion capture, transitioning from marker suits to marker-free systems like Microsoft Kinect, has gained traction. This study introduces a real-time method for extracting consistent global 3D human pose from single RGB videos. Combining a real-time, fully-convolutional 3D body pose formulation with model-based kinematic skeleton fitting, the approach excels in accuracy across various scenarios, including outdoor scenes and low-quality recordings. In the realm of 3D HPE, the study addresses the challenge by utilizing the inherent connections between human joints. Inspired by previous work, a network is designed to take 2D joint positions as input and accurately predict 3D locations. The use of relational modules and innovative techniques like relational dropout and hierarchical relational networks enhances the model's performance, achieving state-of-the-art results on the Human 3.6M dataset, even in situations with missing joint information. The study collectively reflects the ongoing advancements and challenges in deep learning-based 2D and 3D HPE.

2. Literature Review

Grandel Dsouza et al in [1] presented a smart gym trainer employing human pose estimation in computer vision/graphics is the study of techniques, methods, and pre-trained models that recover the posture of an articulated body, which consists of joints and rigid components, using image-based observations.

Xavier Perez-Sala et al in [2] Define a comprehensive taxonomy for model-based methods to Human Pose Recovery, consisting of five major modules: appearance, perspective, spatial connections, temporal consistency, and behaviour.

Alexander Toshev et al in [3] suggested a Deep Neural Network (DNN)-based technique for human posture estimation. Pose estimation is expressed as a DNN-based regression issue for body joints. We demonstrate a cascade of such DNN regressors that produce high-precision posture estimations.

Sungheon Park et al in [4] The suggested strategy increases CNN performance by introducing two innovative ideas. First, we combined 2D pose estimation results with picture attributes to estimate a 3D posture. Second, we discovered that integrating information on relative locations with regard to several joints yields more accurate 3D poses than using only one root joint.

Sijin Li et al in [5] suggested a deep convolutional neural network for 3D human pose estimation from monocular photos. We train the network with two strategies: 1) A framework that trains pose regression and body part detectors simultaneously; 2) A pre-training technique that initializes the pose regressor with a network trained for body part detection.

Zhi-Qiang Liu et al in [6] A heterogeneous multi-task learning framework for human posture estimation from monocular images was suggested, utilizing a deep convolutional neural network. In specifically, we simultaneously learn a pose-joint regressor and a sliding-window body-part detector in a deep network.

Manuel Palermo et al in [7] A broad, real-time, full-body position prediction framework was examined using two RGB+D video streams with nonoverlapping views mounted on a smart walker for rehabilitation purposes.

Ali Rohan et al in [8] proposed a strategy where human position estimate is integrated with a CNN for classification of normal and abnormal gait of a human, with the ability to offer information about the observed anomalies from an extracted skeletal picture in real-time.

Bugra Tekin et al in [9] propose a Deep Learning regression architecture for structured prediction of 3D human pose from monocular pictures, which relies on an overcomplete auto-encoder to train a high-dimensional latent pose representation and account for joint dependencies.

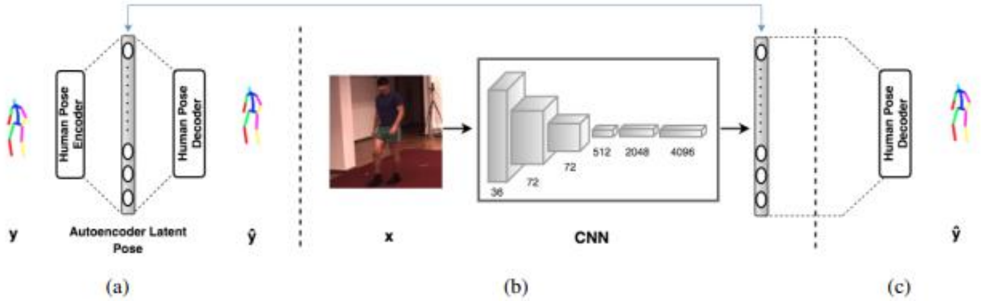

Figure 1: Our architecture for the structured prediction of the 3D human pose. (a) An auto-encoder whose hidden layers have a larger dimension than both its input and output layers is pretrained. In practice we use either this one or more sophisticated versions that are described (b) A CNN is mapped into the latent representation learned by the auto-encoder. (c) the latent representation is mapped back to the original pose space using the decoder.

Ilia Petrov et al in [10] Consider the challenge of human posture estimation with probabilistic convolutional neural networks. They use probabilistic deep learning frameworks to increase human posture estimation accuracy on established pose estimation benchmarks such as the MPII human pose and Leeds Sports posture (LSP) datasets.

Ruixu Liu et al in [11] shown a systematic design (from 2D to 3D) for how traditional networks and other types of restrictions may be included into the attention framework for learning long-range dependencies for the job of pose estimation.

Manolis Vasileiadis et al in [12] suggested a new fully-convolutional, detection-based 3D-CNN architecture for 3D human posture estimation from 3D data. The architecture adheres to the sequential network architecture paradigm, generating per-voxel likelihood maps for each human joint from a 3D voxel-grid input, and is extended to multi-person 3D pose estimation via a bottom-up approach, allowing the algorithm to estimate multiple human poses at the same time without being affected by the number of people in the scene.

Isinsu Katircioglu et al in [13] proposed a Deep Learning regression architecture for structured prediction of 3D human pose from monocular pictures or 2D joint location heatmaps, which uses an overcomplete autoencoder to build a high-dimensional latent pose representation while accounting for joint dependencies.

Xiabing Liu et al in [14] developed a method to predict head posture using a convolutional neural network trained on synthetic head pictures, and framed head pose estimation as a regression issue. A convolutional neural network is trained to learn head characteristics and solve the regression problems.

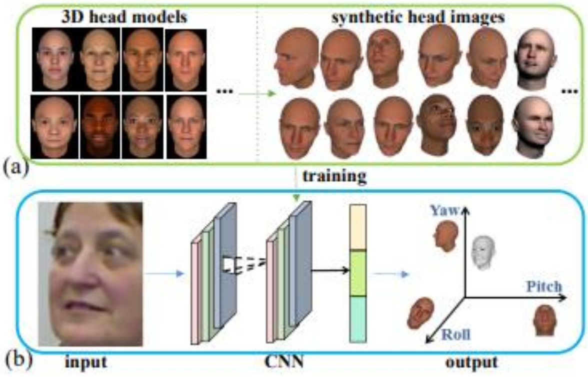

Figure 2: The framework of our method. (a) The synthetic head poses data for CNN training are generated from head models. (b) For a given head image, a CNN model is used to estimate the head pose, which is represented by three angles: yaw, pitch and roll.

Yu Xiang et al in [15] Estimating the 6D posture of known objects is critical for robots to interact with the physical environment. The task is difficult owing to the variety of items as well as the complexity of a scene generated by clutter and occlusions between them. In this paper, we provide PoseCNN, a novel Convolutional Neural Network for 6D object pose estimation.

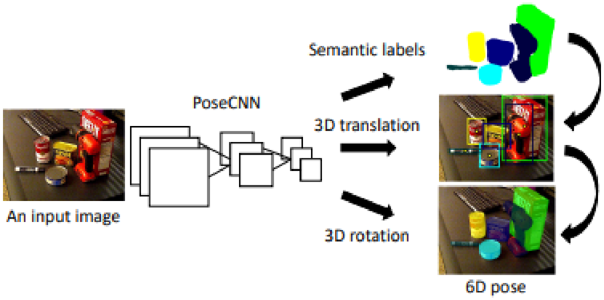

Figure 3: Proposed a novel PoseCNN for 6D object pose estimation, where the network is trained to perform three tasks: semantic labeling, 3D translation estimation, and 3D rotation regression.

Liuhao Ge et al in [16] Using threedimensional Convolutional Neural Networks (3D CNNs), we offer a simple but effective method for estimating hand posture in real time from single depth photos. Because of the lack of 3D spatial information, image-based characteristics retrieved by 2D CNNs are incompatible with 3D hand posture estimation.

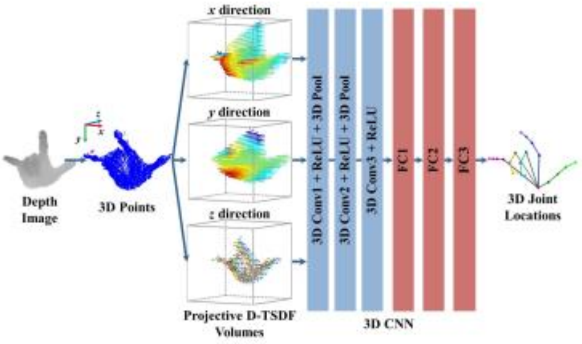

Figure 4: Overview of our proposed 3D CNN based hand pose estimation method. We generate the 3D volumetric representation of hand with projective D-TSDF from the 3D point cloud. 3D CNN is trained in an end-to-end manner to map the 3D volumetric representation to 3D hand joint relative locations in the 3D volume.

Aiden Nibali et al in [17] offer enhancements to 3D coordinate prediction which avoid the aforementioned unwanted features by predicting 2D marginal heatmaps under an upgraded soft-argmax scheme

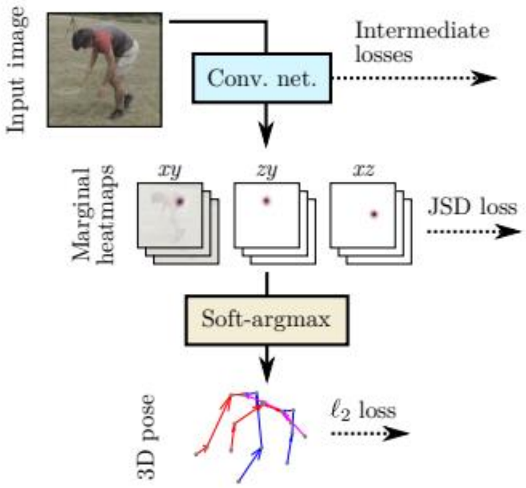

Figure 5: High-level system overview for MargiPose

Ce Zheng et al in [18] A systematic examination and comparison of contemporary deep learning-based methods for both 2D and 3D posture estimation was conducted, with input data and inference processes being used to offer a full assessment.

Dushyant Mehta et al in [19] demonstrated the first real-time approach for capturing a human's entire global 3D skeleton posture in a stable and temporally consistent way utilizing a single RGB camera. Our approach integrates a novel convolutional neural network (CNN)-based posture regressor with kinematic skeleton fitting.

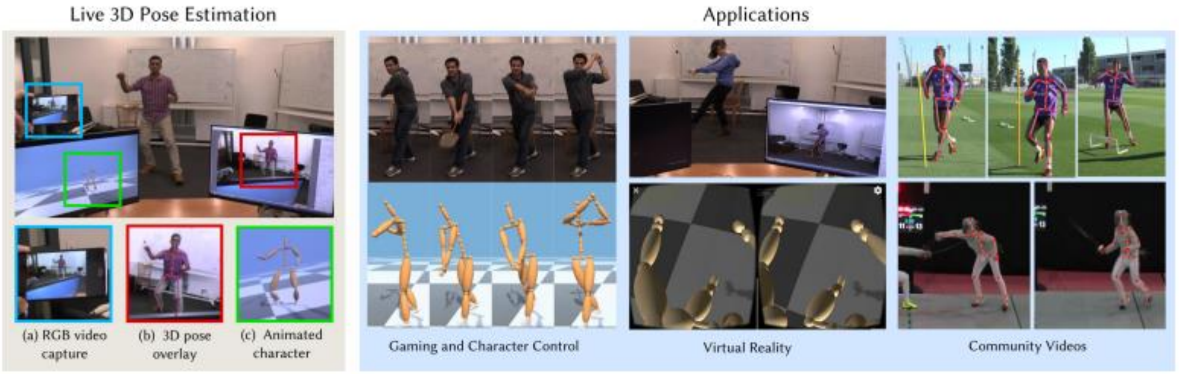

Figure 6: Recovered the full global 3D skeleton pose in real-time from a single RGB camera, even wireless capture is possible by streaming from a smartphone (left). It enables applications such as controlling a game character, embodied VR, sport motion analysis and reconstruction of community video (right). Community videos (CC BY) courtesy of Real Madrid C.F. [2016] and RUSFENCING-TV [2017]

Nojun Kwak et al in [20] present a unique 3D human posture estimation technique from a single image based on neural networks, using the form of relational networks to capture the relationships between distinct body components.

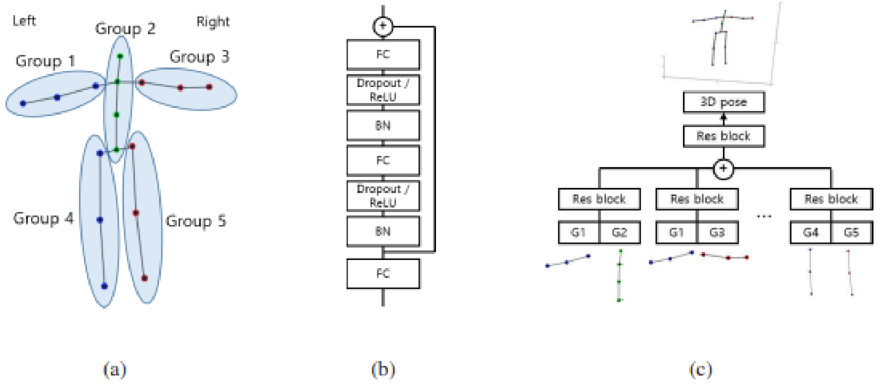

Figure 7: (a) Group configurations used in this paper. Then divided 16 2D input joints to no overlapping 5 groups each of which corresponds to left/right arms, left/right legs and a torso. (b) The residual module used in this paper. (c) The structure of the RN for 3D HPE. Features extracted from all pairs of groups are averaged to produce features for pose estimation. Each Resblock in the figure has the same structure shown in (b).

Table 1: Summary of existing work on Human pose estimation

Sr. No.	Author name	Year	Algorithm and model used	Accuracy
1	Grandel Dsouza[1]	2020	CNN , Deep Leaning, COCO model, MPII model	--
2	Xavier Perez-Sala[2]	2014	Spatial Models, Motion Models, Temporal Models	--
3	Alexander Toshev[3]	2014	DNN-based pose regression.	0.5
4	Sungheon Park[4]	2016	CNN, Caffe framework	96 %
5	Sijin Li	2014	DconvMP	--
6	Zhi-Qiang Liu	2014	CNN, Heterogeneous multi-task framework	94.3 %
7	Manuel Palermo	2021	Spatio-temporal CNN,	73.10%
8	ALI ROHAN	2020	CNN, Linear discriminant analysis (LDA)	97.30%
9	Pascal Fua	2016	DNN, CNN	--
10	Anton Konushin	2018	CNN, DISCO networks, Bayesian SegNet	92 % & 88.2 %
11	Ruixu Liu	2021	CNN, GAN, LSTM, TCN	--

12	Manolis Vasileiadis	2019	CNN	98 %
13	Isinsu Katircioglu	2018	CNN, DNN, RNNs, LSTM, KDE	35.5mm , 97 %
14	Xiabing Liu	2016	CNN, 3DS Max Script, VRAY	95.50%
15	Dieter Fox	2018	PoseCNN,	93 %
16	Daniel Thalmann	2017	3D CNNs, TSDF	--
17	Aiden Nibali	2018	MargiPose	94 %
18	CE ZHENG	2023	CNN, GAN, Graph-PCNN, RSN	--
19	Dushyant Mehta	2017	CNN	82.50%
20	Sungheon Park	2018	Relational Networks, Relational dropout	--

3. Dataset Used

Sr No.	Dataset	Specification
1	Custom Dataset [1]	Contains a couple photos with stickmen on them to help identify the posture.
2	A Buffy Stickmen Dataset [2]	The package comprises 748 annotated video frames from the fifth season of BTVS, spread across 5 episodes.
3	FLIC [3]	comprises includes 4000 training and 1000 test pictures taken from major Hollywood films.
4	Leeds Sports Dataset [3]	They include 11000 training and 1000 testing photos.
5	Human 3.6m dataset [4]	Contains 3.6 million human postures and photos recorded by a high-speed motion capture device.
6	Human 3.6m dataset [5]	A high-speed motion capture system captures 3.6 million human postures and their related photos.
7	Buffy Stickmen [6]	The package of 748 annotated video frames over 5 episodes of the fifth season of BTVS.
8	ETHZ Stickmen [6]	549 photos from the 2008 PASCAL VOC travel release. The set comprises mostly of amateur images with inadequate illumination and image quality.
9	Leed Sport Pose [6]	They include 11000 training and 1000 testing photos.
10	Synchronic Activities Stickmen [6]	This dataset focuses on circumstances in which numerous people do an activity simultaneously.
11	Frames Labeled In Cinema [6]	comprises 4000 training and 1000 test photographs taken from significant Hollywood films.
12	We Are Family [6]	The dataset comprises 525 photographs retrieved from the Internet using Google's image search seeks such as "family photo", "rock band", "group photo", "music band", and "team photo".
13	Walker Dataset [7]	It analyzes RGB+D pictures from the walker cameras to ground-truth (GT) keypoint data (referred to as skeleton) from the Xsens system.

14	GAIT Dataset [8]	1.2 million high-resolution photos from the ImageNet collection were presented.
15	Human 3.6m dataset [9]	Provides 3.6 million human positions and photos collected by a high-speed motion capture device.
16	LSP Dataset [10]	They include 11000 training and 1000 testing photos.
17	Human 3.6m dataset [11]	Contains 3.6 million human postures and photos recorded by a high-speed motion capture device.
18	HumanEva Dataset [11]	The HumanEva-I dataset comprises seven calibrated video sequences (4 grayscale and 3 color) that are synced with 3D body positions recorded by a motion capture device.
19	ITOP Invariant Top View dataset [12]	Includes recordings of 20 persons executing 15 action sequences each, as collected by two depth cameras.
20	EVAL dataset [12]	Includes 24 recordings of three different people executing increasingly difficult tasks.
21	Personalized Depth Tracker dataset [12]	Includes 20 recordings of five distinct participants (3 males and 2 females) doing more difficult movements.
22	CMU PanopticStudio 3D PointCloud dataset [12]	Includes about 6 hours of recordings divided over 54 single and multi-person segments.
23	Human3.6m [13]	Contains 3.6 million human postures and photos recorded by a high-speed motion capture device.
24	HumanEva Dataset [13]	The HumanEva-I dataset comprises seven calibrated video clips (4 grayscale and 3 colors) synced with 3D body positions captured by a motion capture device.
25	KTH Multiview Football II [13]	This dataset includes photos of professional football players during an Allsvenskan league match. It is divided into two sections: one with ground truth posture in 2D and one with ground reality posture in both 2D and 3D.
26	Leeds Sports Pose (LSP) [13]	They consist of 11000 training and 1000 testing pictures.
27	Custom HeadPose Dataset [14]	It has 74,000 head postures generated from 37 headmodels.
28	YCB-Video dataset [15]	Accurate 6D postures of 21 items from the YCB collection were observed in 92 films totaling 133,827 frames.
29	MSRA dataset [16]	MSRA Hands is a dataset used for hand tracking. The Creative Interactive Gesture Camera from Intel captures the right hands of six individuals in total.
30	NYU dataset [16]	The NYU-Depth V2 data collection consists of video sequences that represent different interior situations captured by both the RGB and Depth cameras on the Microsoft Kinect.
31	MPII Human Pose [17]	A famous 2D posture estimation dataset consisting of still frames from YouTube videos. Each picture comprises at least one human subject with an arm description of a 16-joint skeleton in 2D.
32	Human3.6M [17]	Contains 3.6 million human postures and photos recorded by a high-speed motion capture gadget.
33	MPI-INF-3DHP [17]	It captures 8 performers conducting 8 actions with 14 camera perspectives. It comprises of about 1.3 million frames taken by 14 cameras.
34	Max Planck Institute for Informatics (MPII) Human Pose Dataset [18]	comprises around 25,000 photos with over 40,000 people who have labelled bodily joints.

35	Microsoft Common Objects in Context (COCO) Dataset [18]	It includes almost 330,000 photos and 200,000 tagged people with key points, with every individual labelled with 17 joints.
36	PoseTrack Dataset [18]	PoseTrack2017 [1] comprises 514 video clips with 16,219 posture annotations, divided into 250 training, 50 validation, and 214 testing sequences. PoseTrack2018 [2] includes 1,138 video clips with 153,615 posture annotations, organized into three categories: training (593), validation (170), and testing (375).
37	MPI-INF3DHP [19]	It captures 8 performers conducting 8 actions with 14 camera perspectives. It comprises of about 1.3 million frames taken by 14 cameras.
38	Human3.6m [19]	Contains 3.6 million human postures and photos recorded by a high-speed motion capture device.
39	Human3.6m [20]	A high-speed motion-capturing system captures 3.6 million human postures and their related photos.

4. Conclusion

In conclusion, human pose estimation has emerged as a pivotal and dynamic field within computer vision, with far-reaching implications across various domains. This technology, empowered by deep learning and advanced algorithms, has unlocked a multitude of possibilities, from revolutionizing healthcare and sports analysis to enhancing human-computer interaction and immersive technologies. Through the systematic estimation of key body points and joints, it has become possible to decode the spatial arrangement of the human body accurately. This, in turn, enables the analysis of human movements, gestures, and activities in both 2D and 3D dimensions. The applications are diverse, including action recognition, motion capture, gesture-based control, and immersive gaming experiences. However, challenges persist in achieving consistent accuracy, particularly when dealing with complex poses, occlusions, and real-time requirements. The reliance on extensive and diverse training data, the optimization of deep learning models, and the adaptation to various environments and scenarios are critical endeavors in the pursuit of enhanced pose estimation. Furthermore, ethical considerations surrounding privacy and data usage in public spaces are paramount, and responsible development and deployment are essential for the acceptance and trust of these systems. As the field continues to evolve, human pose estimation promises to reshape how we interact with technology, understand human movements, and develop innovative applications. The journey toward seamless, real-time, and robust pose estimation is a testament to the profound impact technology can have on our understanding of the human body and its limitless potential in various industries and aspects of our daily lives.

References

[1] G. Dsouza, D. Maurya and A. Patel, "Smart gym trainer using Human pose estimation," 2020 IEEE International Conference for Innovation in Technology (INOCON), Bangluru, India, 2020, pp. 1-4, doi: 10.1109/INOCON50539.2020.9298212. Available: https://ieeexplore.ieee.org/document/9298212

[2] Perez-Sala, X.; Escalera, S.; Angulo, C.; Gonzàlez, J. A Survey on Model Based Approaches for 2D and 3D Visual Human Pose Recovery. Sensors 2014, 14, 4189-4210. https://doi.org/10.3390/s140304189

[3] A. Toshev and C. Szegedy, "DeepPose: Human Pose Estimation via Deep Neural Networks," 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 2014, pp. 1653-1660, doi: 10.1109/CVPR.2014.214
Available: https://ieeexplore.ieee.org/document/6909610

[4] Park, S., Hwang, J., and Kwak, N., “3D Human Pose Estimation Using Convolutional Neural Networks with 2D Pose Information”, <i>arXiv e-prints</i>, 2016. doi:10.48550/arXiv.1608.03075. Available: https://arxiv.org/abs/1608.03075

[5] Li, Sijin & Chan, Antoni. (2014). 3D Human Pose Estimation from Monocular Images with Deep Convolutional Neural Network. 9004. 332-347. 10.1007/978-3-319-16808-1_23 Available: http://visal.cs.cityu.edu.hk/static/pubs/conf/accv14-3dposecnn.pdf

[6] Li, S., Liu, Z.-Q., and Chan, A. B., “Heterogeneous Multi-task Learning for Human Pose Estimation with Deep Convolutional Neural Network”, <i>arXiv e-prints</i>, 2014. doi:10.48550/arXiv.1406.3474. Available: https://arxiv.org/abs/1406.3474

[7] Palermo, M., Moccia, S., Migliorelli, L., Frontoni, E., and Santos, C. P., “Real-Time Human Pose Estimation on a Smart Walker using Convolutional Neural Networks”, <i>arXiv e-prints</i>, 2021. doi:10.48550/arXiv.2106.14739. Available: https://arxiv.org/abs/2106.14739

[8] A. Rohan, M. Rabah, T. Hosny and S. -H. Kim, "Human Pose Estimation-Based Real-Time Gait Analysis Using Convolutional Neural Network," in IEEE Access, vol. 8, pp. 191542-191550, 2020, doi: 10.1109/ACCESS.2020.3030086. Available: https://ieeexplore.ieee.org/document/9220146

[9] Tekin, B., Katircioglu, I., Salzmann, M., Lepetit, V., and Fua, P., “Structured Prediction of 3D Human Pose with Deep Neural Networks”, <i>arXiv e-prints</i>, 2016. doi:10.48550/arXiv.1605.05180. Available: https://arxiv.org/abs/1605.05180

[10] Petrov, Ilia & Shakhuro, Vlad & Konushin, Anton. (2018). Deep Probabilistic Human Pose Estimation. IET Computer Vision. 12. 10.1049/iet-cvi.2017.0382. Available: https://www.researchgate.net/publication/322914112_Deep_Probabilistic_Human_Pose_Estimation

[11] Liu, R., Shen, J., Wang, H., Chen, C., Cheung, S.-. ching ., and Asari, V. K., “Enhanced 3D Human Pose Estimation from Videos by using Attention-Based Neural Network with Dilated Convolutions”, <i>arXiv e-prints</i>, 2021. doi:10.48550/arXiv.2103.03170. Available: https://arxiv.org/abs/2103.03170

[12] Manolis Vasileiadis, Christos-Savvas Bouganis, Dimitrios Tzovaras, Multi-person 3D pose estimation from 3D cloud data using 3D convolutional neural networks, Computer Vision and Image Understanding, Volume 185, 2019, Pages 12-23, ISSN 1077-3142, https://doi.org/10.1016/j.cviu.2019.04.011.

[13] Isinsu Katircioglu, Bugra Tekin, Mathieu Salzmann, Vincent Lepetit, Pascal Fua. Learning Latent Representations of 3D Human Pose with Deep Neural Networks. International Journal of Computer Vision, 2018, 126 (12), pp.1326-1341. ff10.1007/s11263-018-1066-6ff. ffhal-02509358f. Available : https://link.springer.com/article/10.1007/s11263-018-1066-6

[14] X. Liu, W. Liang, Y. Wang, S. Li and M. Pei, "3D head pose estimation with convolutional neural network trained on synthetic images," 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 2016, pp. 1289-1293, doi: 10.1109/ICIP.2016.7532566. Available : https://ieeexplore.ieee.org/document/7532566

[15] Xiang, Y., Schmidt, T., Narayanan, V., and Fox, D., “PoseCNN: A Convolutional Neural Network for 6D Object Pose Estimation in Cluttered Scenes”, <i>arXiv e-prints</i>, 2017. doi:10.48550/arXiv.1711.00199. Available : https://arxiv.org/abs/1711.00199

[16] L. Ge, H. Liang, J. Yuan and D. Thalmann, "3D Convolutional Neural Networks for Efficient and Robust Hand Pose Estimation from Single Depth Images," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 5679-5688, doi: 10.1109/CVPR.2017.602. Available : https://ieeexplore.ieee.org/document/8100085

[17] A. Nibali, Z. He, S. Morgan and L. Prendergast, "3D Human Pose Estimation With 2D Marginal Heatmaps," 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2019, pp. 1477-1485, doi: 10.1109/WACV.2019.00162. Available : https://ieeexplore.ieee.org/document/8658906

[18] Zheng, C., “Deep Learning-Based Human Pose Estimation: A Survey”, <i>arXiv e-prints</i>, 2020. doi:10.48550/arXiv.2012.13392. Available : https://arxiv.org/abs/2012.13392

[19] Mehta, D., “VNect: Real-time 3D Human Pose Estimation with a Single RGB Camera”, <i>arXiv e-prints</i>, 2017. doi:10.48550/arXiv.1705.01583. Available : https://arxiv.org/abs/1705.01583

[20] Park, S. and Kwak, N., “3D Human Pose Estimation with Relational Networks”, <i>arXiv e-prints</i>, 2018. doi:10.48550/arXiv.1805.08961. Available : https://arxiv.org/abs/1805.08961