

Human Pose Estimation and Skeleton-Based Action Recognition: A Systematic Review of 2D/3D Deep Learning Approaches

Kudratjon Zohirov*

*Department of Software and
hardware support of computer
systems*

*Karshi state technical
university*

Karshi, Uzbekistan

qzohirov@kstu.uz

*Corresponding author

Feruz Ruziboev*

*Department of Convergence
of Digital Technologies*

Tashkent University of

Information Technologies

Tashkent, Uzbekistan

f.roziboyev@tuit.uz,

*Corresponding author

Sardor Boykobilov

*Department of Software and
hardware support of computer
systems*

*Karshi state technical
university*

Karshi, Uzbekistan

sboyqobilov@kstu.uz

Mirjakhon Temirov

*Department of Convergence
of Digital Technologies*

Tashkent University of

Information Technologies

Tashkent, Uzbekistan

m_temirov@tuit.uz

Kamoliddin Ablakulov

*Department of Software and
hardware support of computer
systems*

*Karshi state technical
university*

Karshi, Uzbekistan

k.ablaqulov@kstu.uz

Elbek Jabborov

*Department of Information
Systems and Technologies,*

*Karshi State Technical
University*

Karshi, Uzbekistan

e.jabborov@kstu.uz

Abstract – This article presents a systematic review of modern 2D and 3D deep learning approaches used in the fields of Human Pose Estimation (HPE) and Skeleton-Based Action Recognition (SBAR). The research was conducted based on the SALSA methodology and the published works in leading scientific databases were analyzed. Within the framework of the review, approaches based on the detection of human joint points from image and video sequences, reconstruction of the skeletal structure, and modeling of actions in the spatial and spatio-temporal (ST) domains were compared. The article reviews the architectures of 2D and 3D HPE models, joint hiding in multi-person scenes, real-time requirements, and the possibilities of application in embedded devices. The spatial, spatiotemporal, and graphical features used in SBAR systems and their impact on computational complexity and energy efficiency are also analyzed. The performance of the models was compared based on evaluation criteria such as MPJPE (Mean Per Joint Position Error), AP (Average Precision), RMSE (Root Mean Square Error), and Pearson correlation. The results show that skeleton-based approaches are effective solutions for real-time and resource-constrained systems. However, choosing the optimal model and features requires a trade-off between accuracy, computational complexity, and energy efficiency. This paper identifies promising architectures, features, and hardware adaptation strategies for practical applications of HPE and SBAR systems in real-world environments.

Keywords – Human pose estimation; 2D pose estimation; 3D pose estimation; skeleton-based action recognition; deep learning;

multi-person scenes; real-time systems; computer vision; biomechanics; human-machine interface.

I. INTRODUCTION

Human motion analysis is a research area of great scientific and practical importance in many fields, such as biomechanics, sports medicine, clinical rehabilitation, robotics, and human-machine interfaces [1]. Human motion is a complex ST biomechanical process resulting from the coordinated activity of muscles, joints, and the skeletal system, and its accurate numerical modeling is one of the important tasks of modern research [2]. Therefore, the need for systems for analyzing human posture and movements in real-world conditions such as daily activities, sports training, rehabilitation, and safety monitoring is increasing. In recent years, the development of computer vision and deep learning technologies has brought the field of human motion analysis to a new level. High-resolution imaging devices and deep neural networks have led to the development of HPE and SBAR [3]. These approaches allow for the identification of human joint points from image or video sequences, reconstruction of the skeletal structure, and analysis of movement dynamics.

HPE is a key component of human motion analysis systems, and is aimed at determining the location of joint points from image or video sequences. In practice, HPE approaches are divided into 2D and 3D types: 2D HPE is characterized by its real-time performance by determining the pose in the image plane via (x, y) coordinates; 3D HPE is characterized by its high

accuracy for biomechanical analysis by reconstructing the spatial structure of the skeleton via (x, y, z) coordinates [4].

SBAR approaches analyze human motion based on the temporal changes of joint points and are one of the important directions in the field of Human Activity Recognition (HAR). These approaches are robust to background noise and appearance changes, and allow modeling motion in a ST graphical form [5].

Therefore, this paper systematically analyzes modern 2D and 3D deep learning approaches for HPE and SBAR. The research examines the architectures, skeletal topologies, extractable features of pose recognition models, and their potential for real-time and embedded applications.

II. METHODOLOGY

This section describes the methodology for selecting and analyzing scientific literature related to HPE and SBAR approaches. The research was conducted based on the SALSA (Search, Appraisal, Synthesis and Analysis) methodology [7]. To identify scientific sources, a systematic search was conducted for articles indexed in leading scientific databases such as Scopus, IEEE Xplore, MDPI, PubMed, and ProQuest.

The literature selection process mainly considered studies published between 2010 and 2026. At the same time, exceptions were made for some fundamental works that are important in the development of HPE and deep learning models, based on the year of publication. The search process used keywords such as “human pose estimation”, “2D pose estimation”, “3D pose estimation”, “skeleton-based methods”, “deep learning”, “graph neural networks”, “transformer”, “computer vision”, and “human motion analysis” and their combinations. Initially, articles were reviewed based on title and annotation, and duplicate publications were excluded.

In the next step, the full texts of the selected papers were analyzed and their relevance to HPE and SBAR issues was assessed. The final selection was made based on the following criteria:

- studies aimed at identifying joint points from static images or video sequences;
- HPE approaches based on 2D and 3D deep learning models;
- skeleton-based ST modeling methods.

The following papers were excluded from the analysis:

- studies that did not adequately describe the architecture or evaluation methods;
- papers that focused only on data collection but did not propose a model or analysis.

The main characteristics of the selected studies, including the type of HPE, architecture used, number of nodes, datasets, and evaluation criteria, were summarized. A total of 44 scientific articles were analyzed within the framework of this review. The general scheme of the literature selection process is shown in Figure 1. The selected works were evaluated based on the criteria of methodological rigor, the presence of experimental evaluation, and the reproducibility of the results.

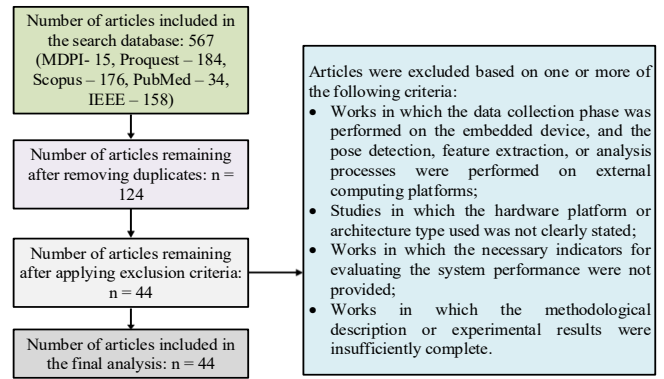


Fig. 1. Overall methodology for article selection.

III. DATA PREPARATION AND COLLECTION.

The data acquisition stage is one of the most important stages of HPE systems based on computer vision and machine learning approaches [3]. The decisions made and the technical solutions chosen in this process directly affect the accuracy, reliability, and practical effectiveness of the research. In particular, the camera settings, the quality and diversity of the datasets, as well as the normalization, noise reduction, and preprocessing processes are important for the effective extraction of human pose features and the training of models based on deep learning.

In addition, the variability of lighting conditions, the occlusion of body parts, the complexity of the background, and the morphological differences between subjects that occur in real environments significantly affect the generalization ability of HPE models. Figure 2 shows the main stages of the data acquisition process in HPE systems.

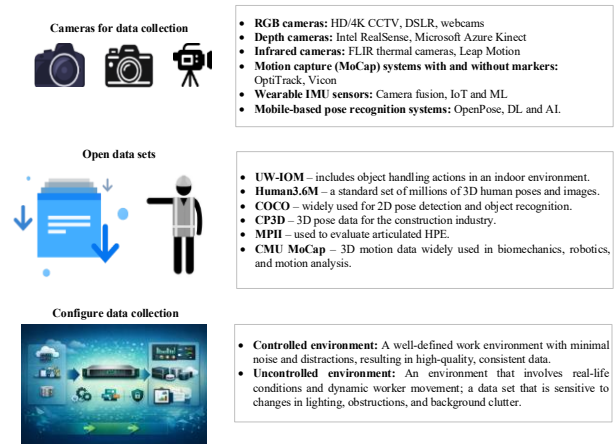


Fig. 2. Data collection stages for HPE.

The accuracy and efficiency of HPE systems directly depend on the technologies used in the data collection process. One of the most widely used methods in practice is RGB cameras, which provide high-resolution color images, are relatively inexpensive, and are easy to integrate into various systems.

Depth cameras, on the other hand, provide additional spatial information, allowing for more accurate analysis of the human body. However, their high cost and sensitivity to lighting conditions may limit their use in some real-world environments [8].

TABLE I. DATA ACQUISITION METHODS USED FOR HPE.

Data acquisition	Data acquisition source and description	Ref.
RGB cameras	IP cameras used for acquiring 2D images for CV – based pose estimation	[9-13]
Depth cameras	Microsoft Kinect and Intel RealSense depth cameras for collecting 3D depth maps and performing pose estimation	[14-15]
Infrared cameras	FLIR infrared cameras for motion detection and tracking	[16-17]
Marker-based and markerless motion capture systems	Vicon, OptiTrack, and Qualisys systems, which use infrared cameras and reflective markers for precise motion capture	[18]
Motion capture (MoCap)	OpenPose, MediaPipe, and Detectron-AI-based, markerless pose estimation systems	[9], [19-22]
IMU sensors	Camera fusion, multi-IMU, IoT, and ML technologies	[12], [23-24]
Mobile device-based pose estimation systems	Smartphone cameras integrated with OpenPose, edge computing, DL, artificial intelligence (AI), and 2D/3D pose estimation models	[18], [25]

Table I lists the main data collection technologies used in HPE research and the corresponding data sources.

A. Fundamentals of HPE and Joint Keypoint-Based Motion Analysis.

Human movement is a complex biomechanical process that occurs as a result of the coordinated activity of muscles, joints, and the skeletal system. This process is closely related to human daily activities, sports training, rehabilitation, and other physiological functions [26]. Digital analysis of human movements is one of the important research areas in the fields of computer vision and biomechanics.

Human movement analysis based on computer vision usually includes three main steps: feature extraction from images, matching these features in successive frames, and performing high-level analysis. This process is aimed at determining the posture of the human body, the kinematic relationships between joints, and the dynamics of movement [2].

The main element of the HPE process is the joint points. Joint points are coordinates representing important anatomical points in the human body, which serve as the basis for building a skeletal model and representing movements digitally [29]. These points typically represent joint centers such as the head, neck, shoulder, elbow, wrist, hip, knee, and ankle. In 2D HPE systems, joint points are defined by (x, y) coordinates in the image plane, while in 3D HPE approaches, these points are evaluated in spatial (x, y, z) coordinates. As a result, the motion state of the human body is reconstructed in the form of a skeleton. The accuracy of joint points is one of the main factors determining the overall performance of HPE systems. Especially in scenes with many people, the mutual occlusion of joints, low spatial resolution, changing lighting conditions, and the presence of complex backgrounds complicate the pose detection process [30].

Modern deep learning-based HPE models mainly use heatmap or regression approaches to detect joint points. In heatmap-based methods, a probability map is generated for each joint, while regression models directly predict the joint coordinates. The choice depends on computational resources, accuracy requirements, and real-time needs.

Skeletal data formed on the basis of joint points creates a convenient structure for ST analysis. In this case, joints are represented as graph nodes, and anatomical connections are represented as edges. In recent years, 2D HPE approaches have been widely used in real-time systems. These models are computationally lightweight, and algorithms such as OpenPose,

MediaPipe Pose, MoveNet, and HRNet allow for fast pose detection using simple RGB cameras [3]. Therefore, they are effectively used in mobile devices and embedded systems.

3D HPE approaches, on the other hand, allow for the reconstruction of the full spatial structure of the human skeleton and provide high accuracy for biomechanical analysis [5]. However, these models require high computational complexity and hardware resources. Therefore, while 2D HPE is more commonly used in real-time and resource-constrained systems, 3D HPE is preferred in applications requiring high accuracy. The number of joints and tasks in the main skeletal models used in the reviewed studies are listed in Table II.

TABLE II. NUMBER OF JOINT KEYPOINTS USED IN HUMAN POSE AND MOTION ANALYSIS.

Model	Number of joints	Task type
OpenPose (COCO-18)	18	2D Pose Estimation
OpenPose (BODY-25)	25	Pose Estimation + HAR
BlazePose (MediaPipe)	33	Pose Estimation + HAR
MediaPipe Pose	33	2D/3D-lite Pose Estimation
VideoPose3D	17	3D Pose Estimation
GAST-Net	17	3D Pose Estimation
Root-GAST-Net	17	3D Pose Estimation
ST-GCN	25	SBAR
TSHDC	17	3D Pose Estimation
SimpleNet	17	2D → 3D Pose Estimation
SMPL	24	Parametric Body Model
SMPL-X	55	Full-body Pose Modeling
Azure Kinect Body Tracking	32	RGB-D Motion Capture
Multi-Sensor Fusion	17–25	Multi-sensor 3D Pose Estimation

B. Feature Extraction on Embedded Devices

In the HPE process, skeletal data extracted from image or video sequences provide important information about the human body posture, spatial location of joints, and movement dynamics. These data are the main input data for automatic human movement analysis systems. The pose detection process usually includes the initial image processing stage and subsequent motion analysis or classification stages.

An important step in this process is the detection of joint points and the calculation of their spatial coordinates. Key points representing the human skeleton are extracted from each frame, and various geometric and kinematic indicators describing the movement are formed based on them. In practical research, these features are mainly determined based on spatial coordinates and their changes over time, and in some approaches, ST features are used together.

The choice of skeletal features directly affects the overall accuracy of the system. There are many studies in the scientific literature that have achieved high accuracy rates in skeletal-based motion recognition systems, confirming the practical effectiveness of this approach [31]. However, the system efficiency varies depending on the HPE model used, the number of selected joints, and the complexity of the motion. Another important factor for systems running on embedded devices is

computational efficiency. The feature extraction process must meet real-time requirements, not cause excessive computational burden, and provide the ability to work on different hardware platforms.

Table III categorizes the main features used in skeleton-based HPE and SBAR systems into spatial, temporal, and graphical domains.

TABLE III. FEATURES EMPLOYED IN THE REVIEWED STUDIES.

Feature	Abb.	Domain	Use in models
Joint coordinates	JC	Spatial	OpenPose, HRNet, AlphaPose
2D pose coordinates	(x,y)	Spatial	COCO, MPII
3D pose coordinates	(x,y,z)	Spatial	Human3.6M, VideoPose3D
Joint distance	JD	Spatial	Pose analysis
Bone vector	BV	Spatial	Kinematic modeling
Joint angle	JA θ	Spatial	Biomechanics, HAR
Joint velocity	JV	ST	Motion analysis
Joint acceleration	JA	ST	HAR
Temporal difference	ΔJ	Temporal	Frame motion
Trajectory displacement	TD	Temporal	Motion trajectory
Skeleton embedding	SE	Graph	PoseFormer
Attention weights	AW	Graph	Transformer
Skeleton graph node feature	GN	Graph	ST-GCN
Graph adjacency matrix	AM	Graph	GCN

According to the analysis results, the most commonly used features belong to the spatial and ST domains, which allow for effective representation of human pose and movement

dynamics. While features based on graphical structures have high expressiveness in modeling complex movements, their computational complexity is high.

TABLE IV. FEATURE EXTRACTION ON EMBEDDED DEVICES.

Ref	Main features	Device / Platform
[19], [23]	JC, JV, ΔJ , TCF, PSE, (x,y), (x,y,z)	Embedded platform (Jetson TX2, IMU sensors)
[9], [10], [13], [15]	JC, HR, KCS, JD, JA θ , (x,y)	RGB / depth camera systems
[32], [11], [20], [21]	JC, HR, KCS, PN, JN, BV, BO, BL	GPU platforms (V100, RTX series)
[33], [36], [37]	JC, HR, AW, (x,y)	High-performance GPU (RTX 3090–4090, A100)
[14], [35]	JC, JD, (x,y), (x,y,z)	RGB-D / LiDAR sensors
[17]	JC, HR, AW, PSE	Event camera systems
[24], [12]	JC, JA θ , JV, JO, LD, (x,y,z)	Multi-sensor systems (IMU + camera)
[25], [18]	JC, KCS, JA θ , BAC	Smartphone-based systems
[34]	JA θ , JD, AW, PSE	Hybrid sensing (GPU + EMG sensors)

Table IV presents a comparative analysis of the features and hardware platforms used in the HPE and SBAR studies conducted on embedded and limited computing devices. According to the results of the analysis, the most commonly used features were JC, (x, y), (x, y, z), HR, KCS, JA θ , JD, and AW and PSE. Figure 3 shows the most commonly used features in the reviewed studies.

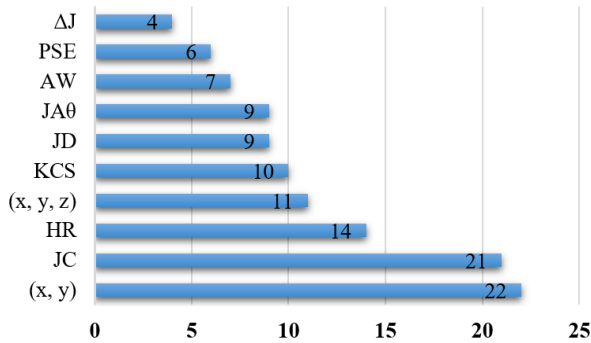


Fig. 3. The ten most frequently used features in the reviewed studies.

The analysis results show that computational complexity and real-time requirements are the decisive factors in the selection of features for HPE and SBAR systems running on embedded devices. Although features in the graphics domain provide high accuracy, their computational complexity is high and requires powerful GPUs or specialized hardware platforms to run effectively. Therefore, the selection of features in embedded systems is based on a balance between accuracy and computational efficiency.

IV. HPE ON EMBEDDED SYSTEMS.

HPE is one of the important tasks in computer vision and biomedical engineering. Artificial intelligence-based algorithms are widely used to perform this task, as they allow detecting complex spatial and ST motion patterns in image and video data [44].

However, the implementation of such algorithms often requires high computing power. Therefore, many studies have used resource-intensive computing platforms such as servers or high-performance workstations. However, this situation limits the possibility of using HPE solutions in everyday life.

This problem is especially relevant for real-time applications, including intelligent rehabilitation systems, motion analysis, human-computer interaction, and surveillance systems. Therefore, in recent years, research has been focused on energy-efficient and compact hardware platforms, especially embedded devices.

One of the effective approaches widely used in embedded systems is to perform model inference on the device itself. This method allows for real-time operation, and the model training process is typically performed on high-performance computing systems.

TABLE V. ARCHITECTURES EMPLOYED FOR HPE.

Model	Complete Model Description
OpenPose	Multi-person 2D human pose estimation framework
BlazePose	Lightweight real-time pose estimation model
MediaPipe Pose	Real-time pose tracking framework by Google
HRNet	High-Resolution Network for accurate pose estimation
AlphaPose	Top-down multi-person pose estimation framework
MoveNet	Fast and efficient PE model for edge devices
VideoPose3D	Temporal convolutional network for 3D pose estimation
GAST-Net	Graph Attention ST Network for pose estimation
ST-GCN	ST Graph Convolutional Network for action recognition
SMPL	Parametric human body modeling framework

Table V lists the main model architectures used in HPE systems. The table shows that in recent years, along with convolutional neural networks, temporal and graph-based models have also been actively developed. Although these models provide high accuracy, their computational complexity requires additional optimization for use in embedded systems.

TABLE VI. PERFORMANCE METRICS OF EMBEDDED AND SENSOR-BASED HPE SYSTEMS.

Ref.	Device	Joint Keypoints	Performance (mm)	Model
[26]	RGB camera	17	MPJPE = 44.9	GAST-Net
[27]	-	17	MPJPE = 34.5	VideoPose3D
[27]	Smartphone	17	RMSE = 46.8 ± 58.2	BlazePose, OpenPose
[38]	RGB camera	17	MPJPE = 97	MediaPipe Pose
[39]	RGB camera	17	MPJPE = 44.9	Root-GAST-Net
[40]	RGB camera	18	MPJPE = 100	OpenPose
[24]	Totalcapture, IMU	17	MPJPE = 20.7	SimpleNet
[35]	Lidar sensor	17	MPJPE = 82	OpenPose
[16]	Jetson TX2	17	MPJPE = 48.4	TSHDC (MSA + THDC)
[9]	RGB camera	17	MPJPE = 52.6	ST-GCN
[41]	RGB camera	18	MPJPE = 100	OpenPose
[42]	LiDAR, IMU	16	RMSE = 107.5	Heterogeneous Multi-Sensor Fusion
[43]	Azure Kinect RGB-D	32	RMSE = 28.4	Azure Kinect

Table VI presents the performance indicators of embedded and sensor-based HPE systems. The results show that the system accuracy depends significantly on the type of device and sensor configuration used. While RGB camera-based approaches are widely used in practice, multimodal systems combining IMU, LiDAR, and RGB-D sensors have achieved higher accuracy through lower MPJPE or RMSE values. In general, effective implementation of HPE algorithms in embedded systems

requires a balance between accuracy, computational complexity, and energy efficiency. Despite the high accuracy of modern deep learning models, their application in real-time and on devices with limited resources requires model simplification and hardware-adapted optimization methods..

V. CONCLUSION

This systematic review analyzes state-of-the-art 2D and 3D deep learning approaches used in the fields of HPE and SBAR. The results show that the performance of HPE and SBAR systems directly depends on the choice of data acquisition methods, skeletal topology, and spatial and ST features. The computational simplicity of 2D HPE models makes them a convenient solution for real-time operation and application in resource-constrained embedded devices, while 3D HPE, while providing high accuracy in biomechanical and clinical analysis, requires large computational resources. Graph-based approaches effectively model complex motion patterns, but their computational complexity limits their application in embedded systems, while multimodal approaches increase system complexity while increasing accuracy. In general, it is important to balance accuracy, computational complexity, and energy efficiency when implementing HPE and SBAR solutions in embedded systems, and future research should focus on simplifying model architectures, reducing redundant parameters, quantization, and exploring hardware-level optimization techniques. This review identifies promising scientific and engineering directions for effective implementation of HPE and SBAR systems in real-time and practical environments.

REFERENCES

- [1] Alderink, G.; Öunpuu, S. Biomechanics of Human Motion and Its Clinical Applications: Instrumented Gait Analysis. *Bioengineering* 2025, 12, 1076. <https://doi.org/10.3390/bioengineering12101076>
- [2] He, C.; Fu, H.; Ma, C.Z.-H. Biomechanics and Motion Analysis: From Human Performance to Clinical Practice. *Bioengineering* 2025, 12, 1114. <https://doi.org/10.3390/bioengineering12101114>
- [3] Samkari, E.; Arif, M.; Alghamdi, M.; Al Ghamdi, M.A. Human Pose Estimation Using Deep Learning: A Systematic Literature Review. *Mach. Learn. Knowl. Extr.* 2023, 5, 1612-1659. <https://doi.org/10.3390/make5040081>
- [4] Sun, R.; Lin, Z.; Leng, S.; Wang, A.; Zhao, L. An In-Depth Analysis of 2D and 3D Pose Estimation Techniques in Deep Learning: Methodologies and Advances. *Electronics* 2025, 14, 1307. <https://doi.org/10.3390/electronics14071307>
- [5] Nguyen, H.-C.; Nguyen, T.-H.; Scherer, R.; Le, V.-H. Deep Learning for Human Activity Recognition on 3D Human Skeleton: Survey and Comparative Study. *Sensors* 2023, 23, 5121. <https://doi.org/10.3390/s23115121>
- [6] De Marchi, Mirco & Turetta, Cristian & Pravadelli, Graziano & Bombieri, Nicola. (2024). Combining 3D Human Pose Estimation and IMU Sensors for Human Identification and Tracking in Multi-Person Environments. *IEEE Sensors Letters*. PP. 1-4. 10.1109/LSENS.2024.3400614.
- [7] Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Info Libr J.* 2009 Jun;26(2):91-108. doi: 10.1111/j.1471-1842.2009.00848.x. PMID: 19490148.
- [8] Guo, Y.; Gao, T.; Dong, A.; Jiang, X.; Zhu, Z.; Wang, F. A Survey of the State of the Art in Monocular 3D Human Pose Estimation: Methods, Benchmarks, and Challenges. *Sensors* 2025, 25, 2409. <https://doi.org/10.3390/s25082409>

- [9] Cheng, Z.; Zhang, Y.; Zhang, L. Quantitative Evaluation and Optimization of Museum Fatigue Using Computer Vision Human Pose Estimation. *Sensors* 2026, 26, 729. <https://doi.org/10.3390/s26020729>
- [10] Elshami, N.E.; Salah, A.; Abdellatif, A.; Mohsen, H. Toward Robust Human Pose Estimation Under Real-World Image Degradations and Restoration Scenarios. *Information* 2025, 16, 970. <https://doi.org/10.3390/info16110970>
- [11] Shi, Y.; Liu, Z.; Jiang, B.; Dai, T.; Lian, Y. A Fine-Tuning Method via Adaptive Symmetric Fusion and Multi-Graph Aggregation for Human Pose Estimation. *Symmetry* 2025, 17, 1098. <https://doi.org/10.3390/sym17071098>
- [12] Lee, D.; Kim, J. HDPose: Post-Hierarchical Diffusion with Conditioning for 3D Human Pose Estimation. *Sensors* 2024, 24, 829. <https://doi.org/10.3390/s24030829>
- [13] Jalal, A.; Akhtar, I.; Kim, K. Human Posture Estimation and Sustainable Events Classification via Pseudo-2D Stick Model and K-ary Tree Hashing. *Sustainability* 2020, 12, 9814. <https://doi.org/10.3390/su12239814>
- [14] Jia, W.; Wang, H.; Chen, Q.; Bao, T.; Sun, Y. Analysis of Kinect-Based Human Motion Capture Accuracy Using Skeletal Cosine Similarity Metrics. *Sensors* 2025, 25, 1047. <https://doi.org/10.3390/s25041047>
- [15] Xing, H.; Yang, J. WPL-Based Constraint for 3D Human Pose Estimation from a Single Depth Image. *Sensors* 2022, 22, 9040. <https://doi.org/10.3390/s22239040>
- [16] Reale, G.; Giovannini, S.; Iacovelli, C.; Castiglia, S.F.; Picerno, P.; Zauli, A.; Rabuffetti, M.; Ferrarin, M.; Maccauro, G.; Caliendo, P. Actigraphic Measurement of the Upper Limbs for the Prediction of Ischemic Stroke Prognosis: An Observational Study. *Sensors* 2021, 21, 2479. <https://doi.org/10.3390/s21072479>
- [17] Furmonas, J.; Liobe, J.; Barzdenas, V. Analytical Review of Event-Based Camera Depth Estimation Methods and Systems. *Sensors* 2022, 22, 1201. <https://doi.org/10.3390/s22031201>
- [18] Dill, S.; Ahmadi, A.; Grimmer, M.; Haufe, D.; Rohr, M.; Zhao, Y.; Sharbafi, M.; Hoog Antink, C. Accuracy Evaluation of 3D Pose Reconstruction Algorithms Through Stereo Camera Information Fusion for Physical Exercises with MediaPipe Pose. *Sensors* 2024, 24, 7772. <https://doi.org/10.3390/s24237772>
- [19] Zhang, L.; Dai, S.; She, L.; Huo, S. Human Pose Intelligent Detection Algorithm Based on Spatiotemporal Hybrid Dilated Convolution Model. *Electronics* 2025, 14, 4798. <https://doi.org/10.3390/electronics14244798>
- [20] Yoon, J.-h.; Kwon, S.-k. Robust Human Pose Estimation Method for Body-to-Body Occlusion Using RGB-D Fusion Neural Network. *Appl. Sci.* 2025, 15, 8746. <https://doi.org/10.3390/app15158746>
- [21] Cruz, M.H.d.I.; Solache, U.; Luna-Álvarez, A.; Zagal-Barrera, S.R.; Morales López, D.A.; Mujica-Vargas, D. CNN 1D: A Robust Model for Human Pose Estimation. *Information* 2025, 16, 129. <https://doi.org/10.3390/info16020129>
- [22] Zakir, A.; Salman, S.A.; Takahashi, H. SOCA-PRNet: Spatially Oriented Attention-Infused Structured-Feature-Enabled PoseResNet for 2D Human Pose Estimation. *Sensors* 2024, 24, 110. <https://doi.org/10.3390/s24010110>
- [23] Kim, M.; Lee, S. Fusion Poser: 3D Human Pose Estimation Using Sparse IMUs and Head Trackers in Real Time. *Sensors* 2022, 22, 4846. <https://doi.org/10.3390/s22134846>
- [24] Chen, M.; Tan, G. 3D Human Pose Estimation Based on Wearable IMUs and Multiple Camera Views. *Electronics* 2024, 13, 2926. <https://doi.org/10.3390/electronics13152926>
- [25] Moreira, R.; Teixeira, S.; Fialho, R.; Miranda, A.; Lima, L.D.B.; Carvalho, M.B.; Alves, A.B.; Bastos, V.H.V.; Teles, A.S. Validity Analysis of Monocular Human Pose Estimation Models Interfaced with a Mobile Application for Assessing Upper Limb Range of Motion. *Sensors* 2024, 24, 7983. <https://doi.org/10.3390/s24247983>
- [26] J. Liu et al., "A Graph Attention Spatio-temporal Convolutional Network for 3D Human Pose Estimation in Video," 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 2021, pp. 3374-3380, doi: 10.1109/ICRA48506.2021.9561605.
- [27] H. Bai, "A Training Method For VideoPose3D with Ideology of Action Recognition," 2021 International Conference on Signal Processing and Machine Learning (CONF-SPML), Stanford, CA, USA, 2021, pp. 169-174, doi: 10.1109/CONF-SPML54095.2021.00041.
- [28] S. Mroz et al., "Comparing the Quality of Human Pose Estimation with BlazePose or OpenPose," 2021 4th International Conference on Bio-Engineering for Smart Technologies (BioSMART), Paris / Créteil, France, 2021, pp. 1-4, doi: 10.1109/BioSMART54244.2021.9677850.
- [29] Chung, J.-L.; Ong, L.-Y.; Leow, M.-C. Comparative Analysis of Skeleton-Based Human Pose Estimation. *Future Internet* 2022, 14, 380. <https://doi.org/10.3390/fi14120380>
- [30] Li, Y.; Shi, Q.; Song, J.; Yang, F. Human Pose Estimation via Dynamic Information Transfer. *Electronics* 2023, 12, 695. <https://doi.org/10.3390/electronics12030695>
- [31] Feng, M.; Meunier, J. Skeleton Graph-Neural-Network-Based Human Action Recognition: A Survey. *Sensors* 2022, 22, 2091. <https://doi.org/10.3390/s2206209>
- [32] Zhang, W.; Shi, Y.; Lin, J. AFJ-PoseNet: Enhancing Simple Baselines with Attention-Guided Fusion and Joint-Aware Positional Encoding. *Electronics* 2025, 14, 3150. <https://doi.org/10.3390/electronics14153150>
- [33] Tu, H.; Qiu, Z.; Yang, K.; Tan, X.; Zheng, X. HP-YOLO: A Lightweight Real-Time Human Pose Estimation Method. *Appl. Sci.* 2025, 15, 3025. <https://doi.org/10.3390/app15063025>
- [34] Du, G.; Zhu, H.; Ding, Z.; Huang, H.; Bie, X.; Jiang, F. Meta-Transfer-Learning-Based Multimodal Human Pose Estimation for Lower Limbs. *Sensors* 2025, 25, 1613. <https://doi.org/10.3390/s25051613>
- [35] Kovács, L.; Bódis, B.M.; Benedek, C. LidPose: Real-Time 3D Human Pose Estimation in Sparse Lidar Point Clouds with Non-Repetitive Circular Scanning Pattern. *Sensors* 2024, 24, 3427. <https://doi.org/10.3390/s24113427>
- [36] Choi, J.-Y.; Ha, E.; Son, M.; Jeon, J.-H.; Kim, J.-W. Human Joint Angle Estimation Using Deep Learning-Based Three-Dimensional Human Pose Estimation for Application in a Real Environment. *Sensors* 2024, 24, 3823. <https://doi.org/10.3390/s24123823>
- [37] Liao, J.; Xu, J.; Shen, Y.; Lin, S. THANet: Transferring Human Pose Estimation to Animal Pose Estimation. *Electronics* 2023, 12, 4210. <https://doi.org/10.3390/electronics12204210>
- [38] Kim, J.-W.; Choi, J.-Y.; Ha, E.-J.; Choi, J.-H. Human Pose Estimation Using MediaPipe Pose and Optimization Method Based on a Humanoid Model. *Appl. Sci.* 2023, 13, 2700. <https://doi.org/10.3390/app13042700>
- [39] El Kaid, A.; Brazey, D.; Barra, V.; Baina, K. Top-Down System for Multi-Person 3D Absolute Pose Estimation from Monocular Videos. *Sensors* 2022, 22, 4109. <https://doi.org/10.3390/s22114109>
- [40] Yan, G.; Yan, H.; Yao, Z.; Lin, Z.; Wang, G.; Liu, C.; Yang, X. Monocular 3D Multi-Person Pose Estimation for On-Site Joint Flexion Assessment: A Case of Extreme Knee Flexion Detection. *Sensors* 2024, 24, 6187. <https://doi.org/10.3390/s24196187>
- [41] Bao, W.; Ma, Z.; Liang, D.; Yang, X.; Niu, T. Pose ResNet: 3D Human Pose Estimation Based on Self-Supervision. *Sensors* 2023, 23, 3057. <https://doi.org/10.3390/s23063057>
- [42] Patil, A.K.; Balasubramanyam, A.; Ryu, J.Y.; Chakravarthi, B.; Chai, Y.H. An Open-Source Platform for Human Pose Estimation and Tracking Using a Heterogeneous Multi-Sensor System. *Sensors* 2021, 21, 2340. <https://doi.org/10.3390/s21072340>
- [43] Eichler, N.; Hel-Or, H.; Shimshoni, I. Spatio-Temporal Calibration of Multiple Kinect Cameras Using 3D Human Pose. *Sensors* 2022, 22, 8900. <https://doi.org/10.3390/s22228900>
- [44] Zou, Y.; Zhang, W.; Weng, W.; Meng, Z. Multi-Vehicle Tracking via Real-Time Detection Probes and a Markov Decision Process Policy. *Sensors* 2019, 19, 1309. <https://doi.org/10.3390/s19061309>