Deep-Learning for Self-Learning in Yoga and Fitness: A Literature Review

Dhananjay Sharma[1], Harshil Panwar[2], Harshit Goel[3], and Rahul Katarya[4]

[1] Department of Computer Science and Engineering, Delhi Technological University, Delhi, India, dhanajaysharma_2k18co119@dtu.ac.in
[2] Department of Computer Science and Engineering, Delhi Technological University, Delhi, India, harshilpanwar_2k18co142@dtu.ac.in
[3] Department of Computer Science and Engineering, Delhi Technological University, Delhi, India, harshitgoel_2k18co143@dtu.ac.in
[4] Department of Computer Science and Engineering, Delhi Technological University, Delhi, India, rahuldtu@gmail.com

Abstract. With the rising levels of stress in the hustle and bustle of modern life, physical exercise has become imperative in order to maintain a sound body and soul. Yoga is one such art form that has received a lot of approval and has become well sought-after in recent times because of its physical, mental and spiritual benefits. Sports and other gym exercises also provide similar benefits. The paper analyses the significance of modern deep learning techniques in making cost-effective self-learning aids for the aforementioned activities.

Keywords: Open Pose, Pose Estimation, Posture Correction, Self-Learning, Yoga

1 Introduction

Yoga, calisthenics, sports and minor gym exercises have slowly started finding their way into everyone's daily routines due to their numerous benefits and minimal prerequisites. One major factor that determines the effectiveness of the aforementioned exercises is posture correctness. The correct posture will not only allow the user to reap maximum benefits but it will also prove to be instrumental in preventing injuries. Unfortunately, professional instructors and institutes providing such guidance to beginners are often very expensive and have a very tight schedule due to their popular nature. This paper aims to summarize the role and scope of deep learning in building assistive technology to provide efficient and inexpensive alternatives for the same.

Pose estimation [1] in deep learning is the sub domain associated with analyzing and approximating various key-points on a human being in order to estimate the pose. It broadly follows two approaches; the bottom-up approach where each joint is estimated and then connected to form a pose and the top-down approach where a human's bounding box is estimated and then joints are approximated within the region. Pose estimation today forms the base for many other research fields such as human activity recognition, pedestrian analysis for self-driving cars, etc. This paper analyzes the use of various pose estimation techniques in building posture correction technology for the various forms of exercise mentioned above.

Nowadays, most end-to-end models in the domain use libraries/hardware such as OpenPose [2], Kinect, etc along with convolutional neural networks (CNN) [3] to detect key-points on a human being. These pose analysis libraries are often paired with LSTMs [4] and/or CNNs to classify the exact pose being performed ("asana" in the case of yoga). In the case of video input, the combination of both proves to work the best as LSTMs help record the temporal relation between the frames while CNNs record the spatial data in every frame thus combining to produce fast and accurate real time results. Once the pose has been classified and the joint locations have been estimated using key points, various mathematical and trigonometric models are employed to perform a comparison between the obtained joint locations and ideal joint locations for a particular pose. This data is then used to generate feedback reports allowing users to fix their posture in real-time.

Thus, by utilizing deep learning techniques, many assistive solutions have been and can be developed to tackle the problem of ensuring posture correctness during exercises. These solutions are not only accurate, they provide real time feedback and are affordable enough to provide a better alternative to posh institutes and personal trainers.

2 Methodology

Fig. 1. General Yoga Self-Learning Pipeline

The general approach that has been adopted by most researchers in the domain comprises two major steps. Firstly, a pose detection algorithm is employed to detect key-points on a human being. A pose detection network first localizes human body joints and then group them into valid pose configuration.

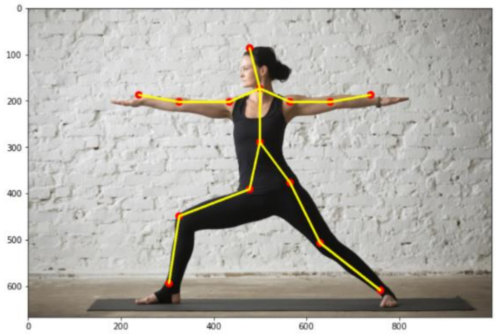

Fig. 2. Detected Joint Coordinates

2.1 Backbone Architecture

The backbone architecture used in these networks primarily range from AlexNet [5], to the recently developed ones such as Fast R-CNN [6], Mask R-CNN [7], feature pyramid networks (FPN) [8]. However, VGG [9] and ResNet [10] remain the popular choice.

2.2 Loss Functions

Loss functions are at the core of any machine-learning or deep learning model to learn from the dataset. In case of human pose estimation models, Cross-Entropy loss, Mean Absolute Error (MAE), Mean Squared Error (MSE) are most used loss functions.

MAE or L_1 loss function is measured as mean of absolute error between prediction and true value. Being outlier insensitive, this loss function is more robust.

$$L_1 = 1/n \sum_{i=1}^{n} |y_i - f(x_i)| \quad (1)$$

MSE or L2 loss function is mean of squared sum of errors between true and predicted value. This loss function penalizes outliers in a dataset.

$$L_2 = 1/n \sum_{i=1}^{n} \left(y_i - f(x_i)\right)^2 \quad (2)$$

For a classification model with probability of output between 0 and 1, cross-entropy loss is used for measuring performance. Similar to other loss function, cross-entropy loss increases for increasing deviation from true value.

$$Log_{loss} = -\left(y_i log\left(f(x_i)\right) + (1 - y_i) log\left(1 - f(x_i)\right)\right) \quad (3)$$

2.3 Evaluation Metrics

For performance comparison of human pose estimation models, several custom evaluation metrics have been developed.

Percentage of Correct Parts (PCP)

It focuses on correct prediction of limbs. A body limb is identified, if distance between two joint positions is less than half in comparison to true value between these key points. Hence due to scaling issues, it penalizes for shorter limbs.

Percentage of Detected Joints (PDJ)

This metric is based on distance between true and predicted joint within a threshold of torso diameter, thus achieving localization precision over former metric.

Percentage of Correct Key-Points (PCK)

It measures if true joint and predicted key point are within a certain threshold. Due to smaller head bone connections and torsos, it overcomes the shorter body limb problem.

The estimated body joint key point data or the frame itself is then passed through a classifier network that identifies the yoga pose or exercise being performed.
In the second step, the joint body key-points are then used to compute the joint angles using various trigonometric functions. These angle values are then compared to the angle values obtained from the instructor (the ideal values, in this scenario). This comparison helps generate a feedback score and detect flaws in the user's posture, if any.

$$m = (y2 - y1)/(x2 - x1) \tag{4}$$

$$angle = |\tan^{-1} m| \tag{5}$$

```
In [47]: diff_matrix[1]

Out[47]: [2.3509990606347797,
          9.782407031807287,
          6.314564091110093,
          12.329753344666454,
          33.92254327774141,
          14.272830843813622,
          20.16585144175515,
          3.465795525230117,
          3.783385236533917]
```

Fig. 3. Estimated Angle Differences

This feedback score is then utilized to generate a report to help users learn by themselves. These approaches have been summarized in table 2.

Table 1. Scoring Algorithm based on angle differences observed

Angle Difference	Score
<10	10
<15	9
<20	8
<25	7
>=25	Make a better Attempt

Table 2. Analysis of existing posture correction pipelines

Authors	Pose Estimation Techniques	Correction Techniques	Pose Classification Accuracy

J. Palanimeera, K. Ponmozhi [11]	KNNs [12], SVMs [13], Logistic Regression [14] and Naive Bayes [15] are used on data set of 12 joints obtained from 17 detected key points.	Only pose classification approaches mentioned.	0.9902 (KNN), 0.9817 (SVM), 0.7347 (NB), 0.8461 (LR)
Munkhjargal Gochoo, Tan-Hsu Tan, Shih-Chia Huang, Tsedevdorj Batjargal, Jun-Wei Hsieh, Fady S. Alnajjar, Yung-Fu Chen [16]	Introduces an IOT based approach using a very deep CNN along with a low-resolution infrared sensor. This infrared sensor is based a wireless sensor network (WSN). WSN is used on the client side to scan the user's movements and then employ a deep CNN to detect the yoga poses.	Only pose classification approaches mentioned.	0.9945 (mean all axes)
Maybel Chan Thar, Khine Zar Ne Winn, Nobuo Funabiki [17]	Open Pose is used along with a CNN to detect yoga poses accurately.	Tan-inverse formula is used to estimate the joint angles using joint key points. These angles are compared to ideal joint angle values of a particular pose. The output pose is displayed using a greed-red gradient to display degree of correctness in posture	0.9777
Fazil Rishan, Binali De Silva, Sasmini Alawathugoda, Shakeel Nijabdeen, Lakmal Rupasinghe, Chethana Liyanapathirana [18]	Open pose and mask-RCNN are used as key point estimators in video frames. Time distributed CNNs are used to capture spatial features. An LSTM is then used to capture temporal and spatial changes based on the features. A SoftMax layer then uses these	Only pose classification approaches mentioned.	0.9991 (Open Pose), 0.9996 (Mask-RCNN)

	changes to display final pose probabilities.		
Manisha Verma, Sudhakar Kumawat, Yuta Nakashima, Shanmuganathan Raman [19]	A novel fine grain hierarchical classification method based on visual similarity of poses is used. Many famous CNN architectures along with a modified dense net 201 with hierarchical connections is proposed. Increased classification accuracy and exemplar performance is observed on an 82-pose dataset.	Only pose classification approaches mentioned.	0.9347 (Var 1 on L3)
Rehnhao Huang, Jiqing Wang, Haowei Lou, Haodang Lu, Bofei Wang [20]	Kinect combined with Open Pose is used to improve upon previous results.	Joint angles are calculated by finding the tan inverse of the value of the vector difference between the two vectors joining to make a joint. These angles are used for comparison with ideal values.	None
Ian Gregory, Samuel Mahatmaputra Tedjojuwono [21]	Open pose is used to detect 18 key points from multiple angles.	Cosine-distance formula is used to obtain joint angles. The difference between the user's joint angles and instructor's joint angles are then used to provide a feedback score.	None
Edwin W. Trejo, Peijiang Yuan [22]	Kinect v2 is paired with the adaboost algorithm to implement a yoga pose classifier for 6 poses for up to 6 people. Adaboost selects the final dataset for training in order to obtain max accuracy.	Uses joint angle comparisons based on joint coordinates obtained with the help of Kinect v2.	0.9478 (DB-3 Mean)

Grandel Dsouza, Deepak Maurya, Anoop Patel [23]	CNN trained on different body part images is used.	Comparison of shoulder and joint angles with that of an athlete's are displayed on a graph over time.	None
Yuxin Hou, Hongxun Yao, Haoran Li, Xiaoshuai Sun [24]	Convolutional Pose Machine [25] and Faster RCNN [26] are used to detect a person and map their position.	An action guided network is used to rate how accurately a pose is being performed.	None
Amit Nagarkoti, Revant Teotia, Amith K. Mahale, Pankaj K. Das [27]	Real time two dimensional multi person Pose Estimation is performed. The proposed approach is based on Part Affinity Fields.	For comparison, Dynamic Time Warping (DTW) between each frame of the user and trainer is done. For output feedback used affine transformations to solve camera problems.	None
Jianbo Wang, Kai Qiu, Houwen Peng, Jianlong Fu, and Jianke Zhu [28]	Region based networks are used to detect humans in an image and isolate them. A custom single person tracking algorithm based off of Siamese tracking [29] is used. A custom pose estimation algorithm is built to analyze both spatial and temporal key points.	Based on the features obtained, a classifier identifies all bad poses so athletes can work on their posture.	None

Further, we have also analyzed recent advancements in the field of pose detection. Deep-pose [30] approaches the problem as a CNN-based regression problem. It refines roughly estimated pose using stream of regressors and images are cropped around predicted joints and fed to next stage, thus allowing feature learning for finer scales. Convolutional Pose Machines (CPM) consists of series of convolutional networks which yields belief maps for each predicted key-point. Hence, feature representation and implicit spatial information are learned at the same time, with supervision at each stage to solve the problem of vanishing gradients. Unlike conventional tech-nique i.e., down sampling high-resolution feature maps and then recovering it later, High-Resolution Net (HRNet) [31] maintains a high-resolution representation throughout, thus comprising of parallel high low-resolution networks with information exchange across multiple resolution networks.

These cutting-edge algorithms, while exceptional in analyzing the human pose, have not yet been utilized for self-learning and posture correction in the field of exercise. The new state of the art pipelines would be able to detect humans and poses more accurately and efficiently, thus, generating a much more relevant feedback report than their predecessors. These new algorithms and networks have been summarized in table below.

Table 3. Comparison of Pose Estimation Algorithms

Model	Year	Backbone Architecture	Approach	Loss Function	Evaluation Metrics	Training Datasets
DeepPose	2014	AlexNet	Top-down	L2	PDJ, PCP	LSP, FLIC
DeeperCut [32]	2016	ResNet	Bottom-Up	L1, Cross-Entropy	AP, mAP, AUC, PCKh@0.5	MPII, COCO, LSP
Convolutional Pose Machines	2016	VGG	Top-down	L2	PCKh@0.1,PCKh@0.2, PCKh@0.5	FLIC, LSP, MPII
Rmpe: Regional Multi-Person Pose Estimation [33]	2017	VGG, ResNet	Top-down	L2	mAP	MPII, COCO
DensePose [34]	2018	FCN, Mask-RCNN	Bottom-up	Cross-Entropy	AP, PCP, GPS	COCO
High-Resolution Network (HRNet)	2019	ResNet	Bottom-up	L2	AP, mAP, PCKh@0.5	MPII, COCO
Human Pose Estimation for Real-World Crowded	2019	ResNet	Top-down	L2	AP, OKS	CrowdPose, JTA

Scenarios [35]						

3 Result and Discussion

In table 2, the paper employing a novel IOT method in order to reduce privacy related issues associated with the use of Kinect and other camera devices makes stellar progress in the self-learning and posture correction domain. It averages an F1 score of 0.93 along all axes and an accuracy of 99.45 percent. The paper conducting a comparison study between open pose and mask-Rcnn as pose analysis modules also achieves excellent results. Open pose as the pose module gives an accuracy of 99.91 percent whereas the mask RCNN algorithm surpasses that result and achieves a 99.96 percent accuracy on the test dataset. The paper employing the convolutional pose machine and faster RCNN also achieved a mean accuracy precision value of 99.9 percent. Lastly the paper using regional based networks also improves upon its predecessor on the same dataset by achieving an F1 score of 0.71. As other papers have not proposed a classification algorithm, accuracy metrics are not available for them.

For each recent human pose estimation model, parameters such as backbone architecture, loss function have been weighed in Table 3. The majority of models opt for Res-Net as the backbone architecture due to its ability to negate the vanishing gradient problem, hence providing a deeper model with greater accuracy. Apart from Average Precision (AP), Mean Average Precision (mAP) many of the models prefer PCKh@0.5 as a good evaluation metric since it directly deals with human images. For training purposes, apart from the standard cross-entropy loss, most models are shifting towards Least Square Errors(L2) even though it is sensitive towards outliers. Last but not the least, in these recent works COCO and MPII have been, by far, default choices of Dataset with richer training data.

4 Future Work

There is potential for the introduction of numerous improvements and improvisations along every step of the pipeline. Firstly, a novel IOT method was proposed to capture user input in the place of a traditional camera or a Kinect to tackle privacy issues that their predecessors were not even aware of. Further improving this approach or finding better and safer methods is one track future authors could research along.

Pose detection networks, still have to overcome many challenges. To this day, multi person pose detection, dynamic activity tracking etc. have not yet reached their saturation points. Thus, improving upon this part of the pipeline is also a track researchers could entertain.

Moreover, the aforementioned pose analysis networks have successfully surpassed their predecessors in terms of accuracy, efficiency and raw speed. Still, these networks have not yet been tested in the domain of self-learning and posture correction for exercise and fitness. Thus, by utilizing the potential of these models, future researchers could help make accurate and truly real time posture classification and analysis possible.

Lastly, by using Natural Language Processing (NLP), rather than tree-based hard coded correction output mechanisms the pipeline could be perfected in a more end to end manner. The use of NLP could allow the model to give a personalized and more relevant feedback instead of relying on a few pre-coded statements. This will help create impactful products that the end users could truly benefit from.

5 Conclusion

Thus, this study manages to showcase the increasing importance of yoga, calisthenics, gyms and general fitness in the daily life of an average worker. Furthermore, the paper neatly summarizes the relevant works and benchmarks that have been set by past researchers aiming to develop an end-to-end pipeline capable of identifying the pose being performed by a user and suggesting corrective measures accordingly. The paper also mentions the algorithms, networks and devices required for the same.

Moreover, the paper also manages to summarize all the cutting-edge pose detection networks that have been created in the past few years. These networks, while much better than their predecessors, have not yet been applied in the field of posture correction and self-learning and posture correction in exercise and fitness.

Lastly the paper also breaks down the pipeline in a systematic format and manages to point out the scope for improvement and improvisation along each step of the pipeline for future researchers to pick up on.

References

1. T. L. Munea, Y. Z. Jembre, H. T. Weldegebriel, L. Chen, C. Huang and C. Yang, "The Progress of Human Pose Estimation: A Survey and Taxonomy of Models Applied in 2D Human Pose Estimation," in IEEE Access, vol. 8, pp. 133330-133348, 2020, doi: 10.1109/ACCESS.2020.3010248.
2. Cao, Zhe & Hidalgo, Gines & Simon, Tomas & Wei, Shih-En & Sheikh, Yaser. (2018). OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields.
3. Yamashita, R., Nishio, M., Do, R.K.G. *et al.* Convolutional neural networks: an overview and application in radiology. *Insights Imaging* 9, 611–629 (2018).
4. Hochreiter, Sepp & Schmidhuber, Jürgen. (1997). Long Short-term Memory. Neural computation. 9. 1735-80. 10.1162/neco.1997.9.8.1735.
5. Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In Advances in Neural Information Processing Systems. Curran Associates, Inc..
6. Girshick, Ross. "Fast r-cnn." Proceedings of the IEEE international conference on computer vision. 2015.
7. He, Kaiming & Gkioxari, Georgia & Dollár, Piotr & Girshick, Ross. (2017). Mask R-CNN
8. Lin, Tsung-Yi, et al. "Feature pyramid networks for object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

9. VGG - Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).
10. ResNet - He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
11. J. Palanimeera, K. Ponmozhi, Classification of yoga pose using machine learning techniques, Materials Today: Proceedings, Volume 37, Part 2, 2021, Pages 2930-2933, ISSN 2214-7853
12. Guo, Gongde & Wang, Hui & Bell, David & Bi, Yaxin. (2004). KNN Model-Based Approach in Classification.
13. Evgeniou, Theodoros & Pontil, Massimiliano. (2001). Support Vector Machines: Theory and Applications. 2049. 249-257. 10.1007/3-540-44673-7_12.
14. Peng, Joanne & Lee, Kuk & Ingersoll, Gary. (2002). An Introduction to Logistic Regression Analysis and Reporting. Journal of Educational Research - J EDUC RES. 96. 3-14. 10.1080/00220670209598786.
15. Rish, Irina. (2001). An Empirical Study of the Naïve Bayes Classifier. IJCAI 2001 Work Empir Methods Artif Intell. 3.
16. M. Gochoo *et al.*, "Novel IoT-Based Privacy-Preserving Yoga Posture Recognition System Using Low-Resolution Infrared Sensors and Deep Learning," in *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 7192-7200, Aug. 2019, doi: 10.1109/JIOT.2019.2915095.
17. M. C. Thar, K. Z. N. Winn and N. Funabiki, "A Proposal of Yoga Pose Assessment Method Using Pose Detection for Self-Learning," *2019 International Conference on Advanced Information Technologies (ICAIT)*, 2019, pp. 137-142, doi: 10.1109/AITC.2019.8920892.
18. F. Rishan, B. De Silva, S. Alawathugoda, S. Nijabdeen, L. Rupasinghe and C. Liyanapathirana, "Infinity Yoga Tutor: Yoga Posture Detection and Correction System," *2020 5th International Conference on Information Technology Research (ICITR)*, 2020, pp. 1-6, doi: 10.1109/ICITR51448.2020.9310832.
19. M. Verma, S. Kumawat, Y. Nakashima and S. Raman, "Yoga-82: A New Dataset for Fine-grained Classification of Human Poses," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020, pp. 4472-4479, doi: 10.1109/CVPRW50498.2020.00527.
20. R. Huang, J. Wang, H. Lou, H. Lu and B. Wang, "Miss Yoga: A Yoga Assistant Mobile Application Based on Keypoint Detection," 2020 Digital Image Computing: Techniques and Applications (DICTA), 2020, pp. 1-3, doi: 10.1109/DICTA51227.2020.9363384.
21. I. Gregory and S. M. Tedjojuwono, "Implementation of Computer Vision in Detecting Human Poses," 2020 International Conference on Information Management and Technology (ICIMTech), 2020, pp. 271-276, doi: 10.1109/ICIMTech50083.2020.9211145.
22. E. W. Trejo and P. Yuan, "Recognition of Yoga Poses Through an Interactive System with Kinect Device," 2018 2nd International Conference on Robotics and Automation Sciences (ICRAS), 2018, pp. 1-5, doi: 10.1109/ICRAS.2018.8443267.
23. G. Dsouza, D. Maurya and A. Patel, "Smart gym trainer using Human pose estimation," 2020 IEEE International Conference for Innovation in Technology (INOCON), 2020, pp. 1-4, doi: 10.1109/INOCON50539.2020.9298212.
24. Y. Hou, H. Yao, H. Li and X. Sun, "Dancing like a superstar: Action guidance based on pose estimation and conditional pose alignment," 2017 IEEE International Conference on Image Processing (ICIP), 2017, pp. 1312-1316, doi: 10.1109/ICIP.2017.8296494.
25. Wei, Shih-En & Ramakrishna, Varun & Kanade, Takeo & Sheikh, Yaser. (2016). Convolutional Pose Machines.

26. Ren, Shaoqing & He, Kaiming & Girshick, Ross & Sun, Jian. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence. 39. 10.1109/TPAMI.2016.2577031.
27. A. Nagarkoti, R. Teotia, A. K. Mahale and P. K. Das, "Realtime Indoor Workout Analysis Using Machine Learning & Computer Vision," 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019, pp. 1440-1443, doi: 10.1109/EMBC.2019.8856547.
28. Jianbo Wang, Kai Oiu Houwen Peng, Jianlong Fu, and Jianke Zhu. 2019. AI Coach: Deep Human Pose Estimation and Analysis for Personalized Athletic Training Assistance. In Proceedings of the 27th ACM International Conference on Multimedia (MM ’19).
29. Bertinetto, Luca & Valmadre, Jack & Henriques, Jao & Vedaldi, Andrea & Torr, Philip. (2016). Fully-Convolutional Siamese Networks for Object Tracking. 9914. 850-865. 10.1007/978-3-319-48881-3_56.
30. A. Toshey and C. Szegedy, “DeepPose: Human Pose Estimation via Deep Neural Networks,” 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp, 1653-1660, doi: 10.1109/CVPR.2014.214.
31. Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., Liu, W., & Xiao, B. (2020). *Deep High-Resolution Representation Learning for Visual Recognition*.
32. Insafutdinov, Eldar & Pishchulin, Leonid & Andres, Bioern & Andriluka, Mykhaylo & Schiele, Bernt. (2016). DeeperCut: A Deeper, Stronger, and Faster Multi-person Pose Estimation Model. 9910. 34-50. 10.1007/978-3-319-46466-4_3.
33. Fang, Hao-Shu & Xie, Shuqin & Tai, Yu-Wing & Lu, Cewu (2017). RMPE: Regional Multi-person Pose Estimation. 2353-2362. 10.1109/ICCV.2017.256.
34. R. A. Güler, N. Neverova and I. Kokkinos, “DensePose: Dense Human Pose Estimation in the Wild,” 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 7297-7306, doi: 10.1109/CVPR.2018.00762.
35. T. Golda, T. Kalb, A. Schumann and J. Beyerer, "Human Pose Estimation for Real-World Crowded Scenarios," 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2019, pp. 1-8, doi: 10.1109/AVSS.2019.8909823.