Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

O-LiPE: An Optimized Lightweight Human Pose Estimation System with Structured Pruning and Quantization for Real-Time Inference

Zeshan Ali
Department of Computer Science & Information Technology, The Superior University Lahore. Email: zeshan.ali.aiengineer@gmail.com

Khalid Hussain
Department of Computer Science & Information Technology, The Superior University Lahore. Email: khalidhussain.fsd@superior.edu.pk

Faiza
Department of Computer Science & Information Technology, The Superior University Lahore. Email: faizazarshal@gmail.com

Faiqa
Department of Computer Science & Information Technology, The Superior University Lahore. Email: faiqa.official.pk@gmail.com

ABSTRACT

The use of human poses in estimating human poses has gained tremendous importance in a number of computer vision systems, such as activity recognition and human-computer interaction, as well as intelligent surveillance systems. However, many state-of-the-art pose estimation models rely on complex deep neural networks with large computational requirements, making them difficult to deploy in real-time or resource-constrained environments. In order to overcome this constraint, this paper presents a light human pose estimation model that emphasizes enhancing computational efficiency with an acceptable prediction ability.The suggested method is an implementation of a backbone based on MobileNetV2 and depthwise separable deconvolution layers to minimize the model parameters. Besides, an attention refinement mechanism is included to improve the description of spatial features. In order to optimize the model further to be used in a real-time system, structured pruning and dynamic quantization techniques are used. The structured pruning operation eliminates all of the redundant channels in the network, thus significantly reducing the number of parameters, and quantization reduces the numeric accuracy of model weights to reduce memory usage.The model is trained using the COCO keypoint dataset with heatmap-based supervision and evaluated through both qualitative visualization and computational performance analysis. The experimental findings indicate that the optimized model can realize a significant reduction in the model size and attain real-time inference. The optimised and quantised network has around 0.36 million

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

parameters and has an inference rate of around 55 frames a second. These findings suggest that the suggested framework can be used as a practical solution when it comes to lightweight human pose estimation in real-time scenarios.

Introduction: Human pose estimation is a core task of computer vision that seeks to identify keypoints that describe various organs of the human body based on the visual information. It has now become a significant part of numerous applications in modern use, including activity recognition, human-computer interaction, sports analysis, augmented reality, and intelligent surveillance systems. Pose estimation systems help machines to comprehend the behavior and movement of humans in complex environments by approximating the position of body joints by comparing images or video frames with a known set of joint positions. The recent developments in deep learning have dramatically enhanced the accuracy of pose estimation models, especially the convolutional neural networks that acquire hierarchical spatial representations directly from image data.

A number of deep learning design structures have demonstrated impressive results on large-scale datasets in human pose estimation. High-resolution networks and multi-stage frameworks have shown high potential in the detection of keypoints with high rates of accuracy. Nevertheless, due to the large number of parameters and complicated network structures, such models usually demand considerable computing and memory. Consequently, most state of the art pose estimation algorithms cannot be easily implemented in real time systems or on devices with limited resources like mobile platforms, embedded systems, and edge computing systems [1],[2].

To counter this obstacle, recent studies have aimed at the creation of lightweight architectures that are both accurate in predictions and have an adequate level of computational efficiency. The use of efficient backbone networks like MobileNet has been extensively implemented in most vision tasks because they can simplify computational costs by use of depthwise separable convolutions [3].Besides lightweight architectures, model compression methods have also been pursued with the aim of decreasing the size and computation cost of deep neural networks. Pruning and quantization are two methods that can eliminate unnecessary parameters and lower numerical accuracy to allow smaller and faster models without cutting off performance short [4],[5].

The pruning techniques remove the less significant weights or channels of neural networks hence lowering the model complexity and time of inference. Structured pruning methods can be deployed exploitatively, as they eliminate whole filters or channels, and this reduction in logic cost does not incur any additional computational cost when running the model [6].Equally, quantization methods reduce the size of models by transforming floating-point weights into lower precision forms, enabling their execution with lower frequency and with less memory. These are especially useful in real time applications where computation efficiency is one of the major needs.

Based on these issues, this paper aims to develop an efficient human pose estimation system, with emphasis made on the design of a lightweight architecture and model compression. The suggested scheme combines a small backbone network with guided pruning and quantization to minimize the model complexity and retain a reasonably

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

high pose estimation. The proposed framework will offer a viable solution in case of real-time pose estimation in computationally-constrained environments by integrating these strategies.

Related Work:

The estimation of human pose has been a widely researched topic of computer vision, and many methods have been suggested to estimate the joints of the human body given an image and a video sequence. The initial work done in pose estimation had used the conventional computer vision algorithms like pictorial structures and graphical models, in approximating joint positions of the body. Such techniques tried to model the human body as a system of interacting components and determine the most likely structure by probabilistic models. These methods though, yielded valuable information about body structure modeling, were in many cases not successful in challenging scenes with occlusions, messy backgrounds, and great differences in the appearance and posture of human bodies [7].

CNNs have been seen to greatly enhance how pose estimation systems perform with the advent of deep learning. The CNN based structures are trained on hierarchical representations that are extracted directly out of image data, which facilitates stronger human joint detection. A number of deep neural network structures have been put forward to work on pose estimation by heatmap regression procedures. In these methods, the network forecasts a heatmap of each joint of the body, with the highest value corresponding to the probable location of the joint. The use of heatmap-based techniques in contemporary pose estimation systems is becoming a standard practice because it can be used to extract the spatial relationship among body parts [8]. A large number of deep learning pose estimation models either have a top-down approach or a bottom-up approach. In top-down methods, a human detector is used to detect each individual person in the image, then a pose estimation network is applied to a specific area. This is a strategy that tends to be more accurate since the network concentrates on individual persons at a given time. The two-stage pipeline, however, comes at the cost of more computations and is potentially a bottleneck in real-time performance. The bottom-up methods, on the other hand, identify all the body joints in an image at a time and subsequently cluster them into separate human poses. Even though bottom-up approaches help to work with several people effectively, there is a possibility that the grouping may add more complexity [9].

A number of popular pose estimation systems have shown good performance based on deep neural networks. OpenPose is one of the powerful bottom-up approaches that introduced the idea of part affinity fields to connect the detected body joints with single individuals in multi-person locales [1].This technique met real-time performance and was heavily used in other applications like human motion analysis and gesture recognition. The architecture is however, very computationally intensive, and has many parameters.

The use of high-resolution networks is another significant advance of pose estimation research. In High-Resolution Networks (HRNet), a high-resolution feature representation is preserved across the whole network rather than gradually diminishing the spatial resolution. A high precision of localizing the body joints is possible with

Online ISSN 3007-3197
Print ISSN 3007-3189

HRNet since it preserves detailed information about the spatial location of the body joints. Although it has good performance, the model has many parameters and needs much computational power; hence, the deployment in resource-limited devices is challenging [10].

The efficient visual recognition tasks have been faced with a number of lightweight architectures that are offered to cope with the computational challenges involved. MobileNet frameworks came up with depthwise separable convolutions, that mean that the number of parameters and floating-point operations that convolutional neural networks need is considerably lower. These light networks have been extensively used in mobile vision tasks as they have the capability to balance their computational capacity and their feature representation capacity [3]. Lightweight backbone networks have also been considered in the context of pose estimation to come up with efficient pose estimation models that can be deployed in real time.

Besides lightweight architectures, model compression has been extensively studied to minimise the size and computational demands of deep neural networks. Pruning is one of the most popular methods that can be applied to eliminate redundant parameters of a trained network. Initial studies have shown that neural networks could have many redundant weights that they could do away without, without a material impact on model accuracy [11]. The pruning techniques can be used at the weight level or at the structural level. The structured pruning techniques are especially useful due to the fact that they completely cut convolutional filters or channels and result in smaller networks that can be effectively run with inference [12]. Quantization is another significant model compression method that decreases the numerical accuracy of network weights and activations. Quantized networks instead of representing parameters with 32-bit floating point numbers, represent parameters with lower precision numbers like 8-bit integers. This lowers memory usage and speeds up inference by making computation on modern hardware platforms more efficient. Quantization has been effectively used in a range of deep learning models and allows them to be deployed to low-computational devices and mobile systems [13].

The combination of pruning and quantization to create compressed models has also been studied in recent research. These methods are commonly combined to deliver large-scale brevity in model structure and evaluation price at a tolerable performance. These optimization practices come in handy especially when implementing deep learning models to real-time applications where speed and memory efficiency is essential [14].

Although lightweight architectures and model compression methods have already been developed, even nowadays, it is a hard task to design efficient pose estimation systems that can preserve good prediction accuracy. Numerous pose estimation models with high accuracy are based on large networks with millions of parameters, and lightweight models can compromise prediction accuracy by being efficient. Consequently, the demand of pose estimation frameworks with effective integration of efficient architecture design as well as model compression schemes to realize real-time behavior without high computational needs has increased.

This challenge is met by the approach suggested in this work by combining a lightweight backbone architecture with the structured use of pruning and quantization

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

techniques. Through a combination of the strategies, the proposed framework can be used to considerably downsize model size and computational complexity without negative impacts on inference in real-time. The objective of this approach is to offer an effective solution when it comes to efficient human pose estimation within systems of limited resources.

Methodology:

In this segment, the suggested lightweight human pose estimation system created in this research is described. The main task of the suggested system is to create an effective pose estimation pipeline that will be able to make real-time inference but retain a reasonable prediction quality. The lightweight convolutional backbone network is combined with model compression methods in the framework to achieve massive model reduction and cut down on the complexity of the computation. The suggested system will include four key steps, such as dataset preprocessing, pose estimation network architecture, structured model pruning, and model quantization.

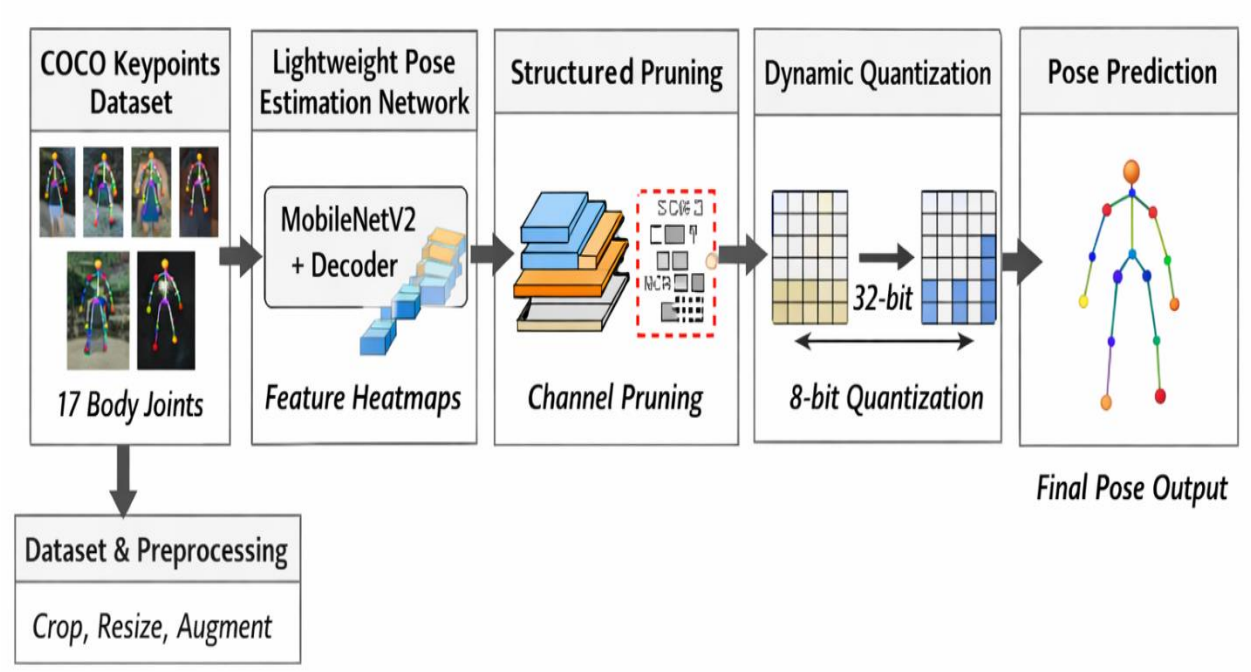


Fig 1. Proposed lightweight human pose estimation framework integrating MobileNetV2 backbone, structured pruning, and dynamic quantization for efficient pose prediction

Dataset and Preprocessing

The COCO 2017 keypoint dataset is used as the training dataset in the proposed model, and it is one of the most popular datasets in the human pose estimation task. The data available consists of images that have been labeled using spatial coordinates that represent various human body joints. The data set has 17 keypoints on the bodies of every single person, which are significant body parts like nose, eyes, shoulders, elbows,

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

wrists, hips, knees, and ankles. Large-scale datasets like COCO have enabled training of deep learning pose estimation systems to a great extent since they offer a variety of human poses and visual conditions [15].

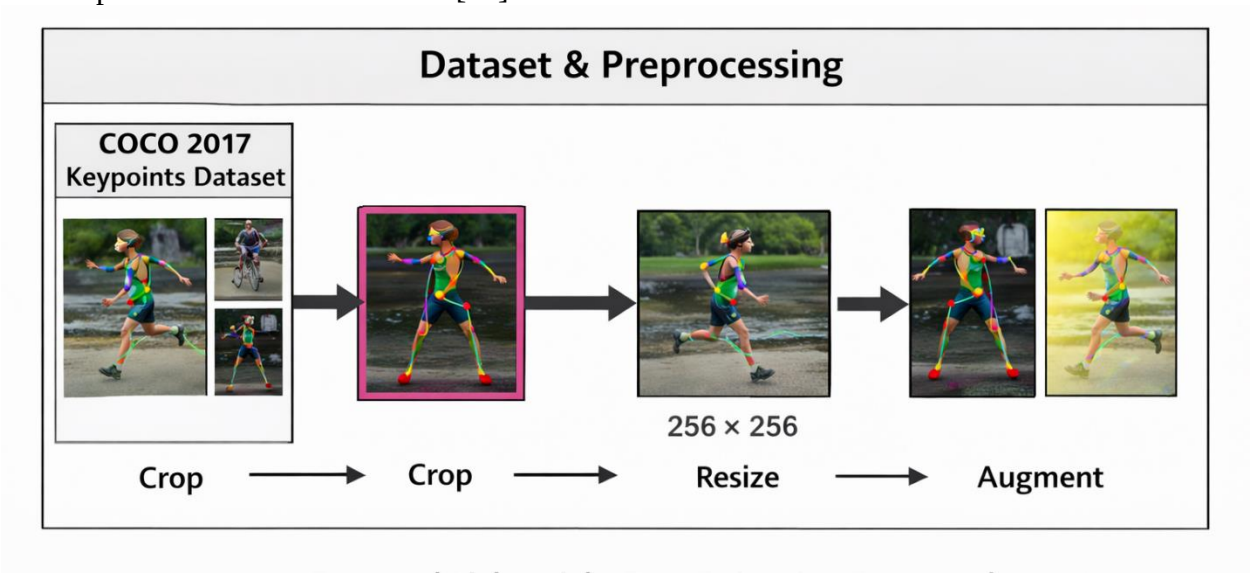


Fig 2. Data preprocessing pipeline for the COCO keypoint dataset including human cropping, image resizing (256 × 256), and data augmentation.

In this study, a subset of the COCO dataset was utilized to improve training efficiency. Approximately **30,000 images were used for training**, while **2,000 images were reserved for validation** to monitor model performance during training.

The selection of images with human instances is carried out by means of the given annotation files during the preprocessing stage. We then crop the region of interest, which contains the person, using the bounding box of each human instance. Cutting the picture of the human body assists the network in concentrating on pertinent visual data and also minimizing the impact of peripheral data.

All cropped images are rescaled to a fixed spatial resolution of 256 x 256 pixels giving all inputs to the neural network the same size. Image normalization is achieved on the basis of generally used mean and standard deviation values to stabilize the training process. Moreover, the techniques of data augmentation like horizontal flipping are used in training to enhance the capability of the model to changes in human pose and viewpoint. It has been demonstrated that data augmentation enhances the ability of deep learning models to generalize [16]. In order to train the network to localize the keypoints, ground truth heatmaps are created per body joint. The spatial probability distribution of each joint is represented by a Gaussian distribution of annotated joint location. The use of heatmap-based supervision has gained significant popularity in contemporary pose estimation systems since it enables neural networks to acquire spatial relations of body joints in a more efficient way than direct coordinate regression techniques [17].

Pose Estimation Network Architecture.

The pose estimation network in this work is created in such a way that it is

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

computationally efficient but has an effective feature extraction ability. The architecture employs MobileNetV2 as the underlying network in extracting features because it is lightweight and boasts of the effective convolutional processes.

MobileNetV2 proposes depthwise separable convolutions and inverted residual blocks, which are much less parameter and floating-point operations intensive than the traditional convolutional neural networks [9]. This architecture has permeated through mobile vision applications since it allows efficient features to be extracted at lower computational requirements.

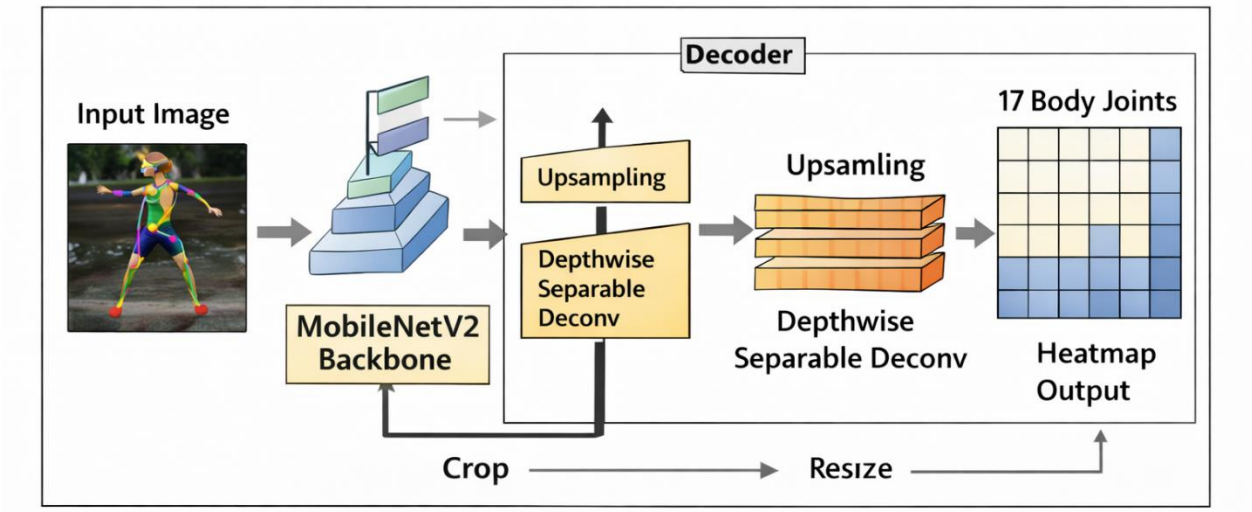


Fig 3. Proposed pose estimation network architecture for human keypoint prediction

The backbone network derives high-level feature representations of an input image. The resulting feature maps are then fed into a lightweight decoder network, which is tasked with producing joint heatmaps. The decoder has upsampling layers, which gradually enhance the spatial resolution of the feature maps, to generate heatmaps, respectively, per human body joint.

In order to be efficient at the decoding step, depthwise separable deconvolution layers would help to rebuild higher resolution feature maps with minimal computational cost. Architectures with lightweight pose estimation, which make use of efficient convolutional operations, have been demonstrated to achieve a reasonable ratio between computational efficiency and prediction [18]

The last prediction layer gives a prediction of a set of heatmaps, which are agreed upon or dissenting towards the 17 joints in the body of the COCO dataset. The heatmaps indicate the distribution of the probability of a specific joint location. The joint coordinates predicted are produced by finding the peak value in each heatmap.

The Mean Squared Error (MSE) loss function is used to train the network with predicted and ground truth heatmaps. MSE-based heatmap regression has been widely applied in pose estimation systems since it is effective at penalizing deviation of the actual spatial distributions of body joints [19] when using it as ground truth.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Structured Model Pruning

Despite the fact that lightweight backbone networks greatly minimize the complexity of computation, deep neural networks still have redundant parameters, which do not add much to the overall prediction. In order to optimize the model more to be suitable in real time, structured model pruning is applied on the trained network.

Structured pruning removes redundant channels from convolutional layers while preserving the structural integrity of the network. Unlike unstructured pruning, this method eliminates entire filters or neurons, resulting in a more compact architecture that can be efficiently executed during inference

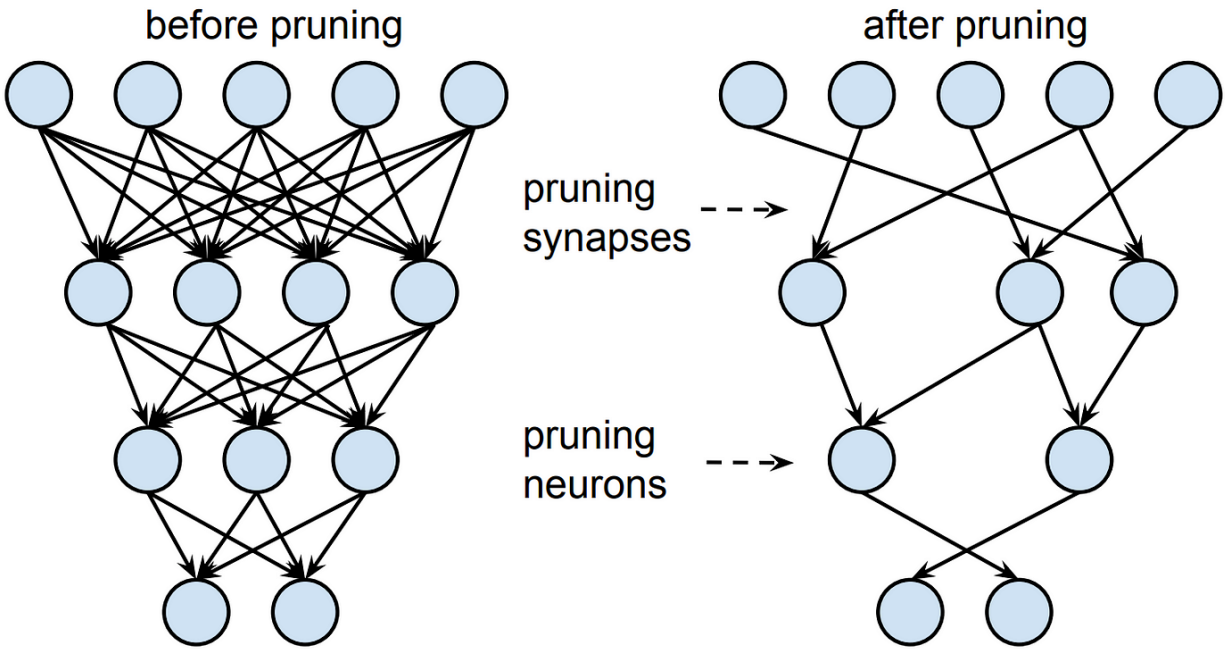


Figure 4: Conceptual illustration of structured pruning showing the removal of redundant neurons and connections to create a more compact neural network architecture.

As illustrated in Figure 4, pruning reduces the number of neurons and connections in the network while maintaining the essential feature extraction capability. This approach significantly reduces model complexity and computational cost.

Pruning algorithms eliminate redundant parameters in neural networks and retain their predictive power. Initial studies showed that neural networks tend to be over-weighted on a number of redundant weights that do not have a significant impact on the accuracy of the model [11]. This idea is furthered by structured pruning, which is pruning on a convolutional filter / channel basis instead of individual weights. The result of this method is smaller models and more efficient to compute when used in inference [12].

The magnitude-based structured pruning is employed in this work in the trained pose estimation model. The significance of every convolutional channel is measured by the strength of the weights. Other channels whose weight is less important are eliminated

Online ISSN 3007-3197
Print ISSN 3007-3189

in the network. It has been demonstrated that channel-level pruning is an effective method of reducing the complexity of the model and still achieving reasonable performance in deep neural networks [20]
The trained network is pruned by a ratio of 40%, i.e., around forty percent of the channels are dropped out of the model. This greatly decreases the amount of trainable parameters and arithmetic operations necessary during inference.

Model Quantization
The model is then optimized, and dynamic quantization is applied after pruning to minimize memory usage and increase inference speed. Quantization is a process that decreases the numerical accuracy of the parameters of a neural network by quantizing 32 bit floating point weights to lower-precision data like 8-bit integers.
Additional arithmetic operations with lower precision use fewer computing resources and can be implemented more efficiently on current hardware platforms. The quantization thus allows the inference to be made faster, and the memory footprint of the model is also lower.
Inference is performed using dynamic quantization without modifying the network structure that was trained. The method allows the model to retain its weights and to enjoy the benefits of reduced storage requirements and increased degrees of computational efficiency. It has been established that pruning and quantization can work in conjunction to compress deep neural networks without much impact to their predictive performance. [20].

Experimental Results
This chapter shows the experimental analysis of the suggested lightweight human pose estimation framework. The experiments were carried out to examine the efficacy of the suggested model with regard to training performance, parameter reduction, computational effectiveness, and inference speed. The dataset was the COCO 2017 keypoint dataset, with which the model was trained and implemented in the PyTorch deep learning platform.
The suggested system combines a lightweight backbone that uses MobileNetV2 and systematic pruning and dynamic quantization models. The purpose of these techniques is to simplify the model and still be able to predict the human body joints. Both quantitative and qualitative measures of the results of the experiments are assessed through quantitative indicators and qualitative pose prediction visual representations.

Training performance analysis
Training loss curve
The training and validation loss values over a number of epochs were used to monitor the training process. The loss curves show a progressive decrease in the values of losses with training, which corresponds to the fact that the model managed to learn spatial relations of image features with the positions of body joints.
The training behavior of the proposed LiPE model was monitored by analyzing the loss values during the optimization process. The training loss curve provides insight into how effectively the model learns spatial representations of human body joints

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

throughout the training epochs.

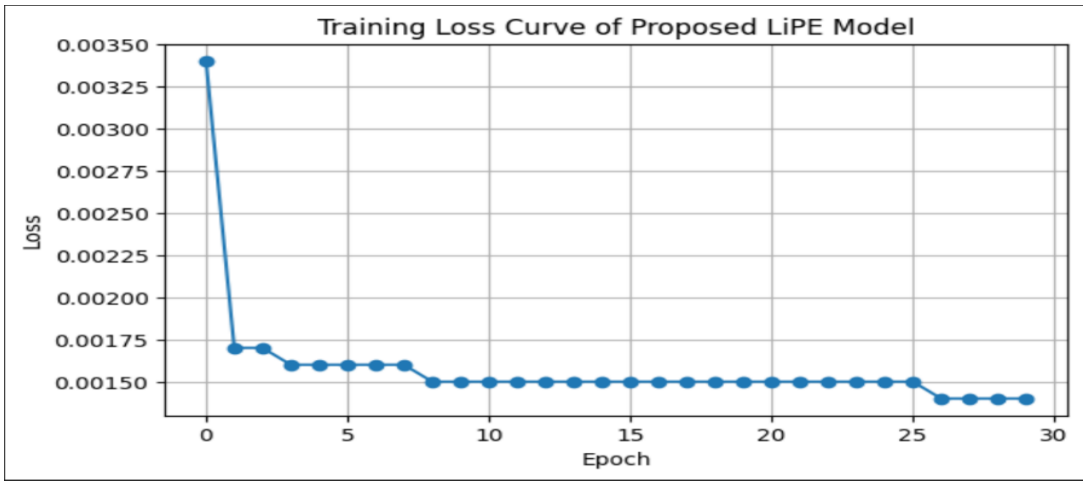


Figure 5: Training loss curve of the proposed LiPE model showing the gradual reduction of loss values during the training process.

As illustrated in Figure 4, the training loss decreases rapidly during the initial epochs and gradually stabilizes as training progresses. This behavior indicates that the network successfully learns meaningful feature representations for human pose estimation.

Training loss vs validation loss curve

In order to further test the ability of the proposed model to generalize, training and validation loss values were tracked throughout the training epochs. The comparison of these curves can be used to identify the occurrence of over fitting in the model during the training process.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about



Figure 6: Training and validation loss curves showing the convergence behavior of the proposed LiPE pose estimation model.

Both training and validation loss values decrease steadily and reach steady levels as illustrated in Figure 5. The thin line between the curves means that the model has no over fitting and generalizes well.

Pruning

Pruning Concept

Pruning was structured to eliminate the unnecessary channels in the network structure. The pruning was very effective in minimizing the model parameters, without affecting the performance of the model in pose estimation.

Table 1. The comparison between the base model and the pruned model.

Model	Number of Parameters
Base Model	674449
Pruned Model	385505

This represents a **parameter reduction of approximately 42.8%**, which significantly decreases the computational complexity of the model.

Once structured pruning was applied to the trained LiPE model, the number of network parameters were analyzed to measure the reduction of the network by channel elimination. This is a comparison of the efficiency achieved through eliminating unnecessary model components.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

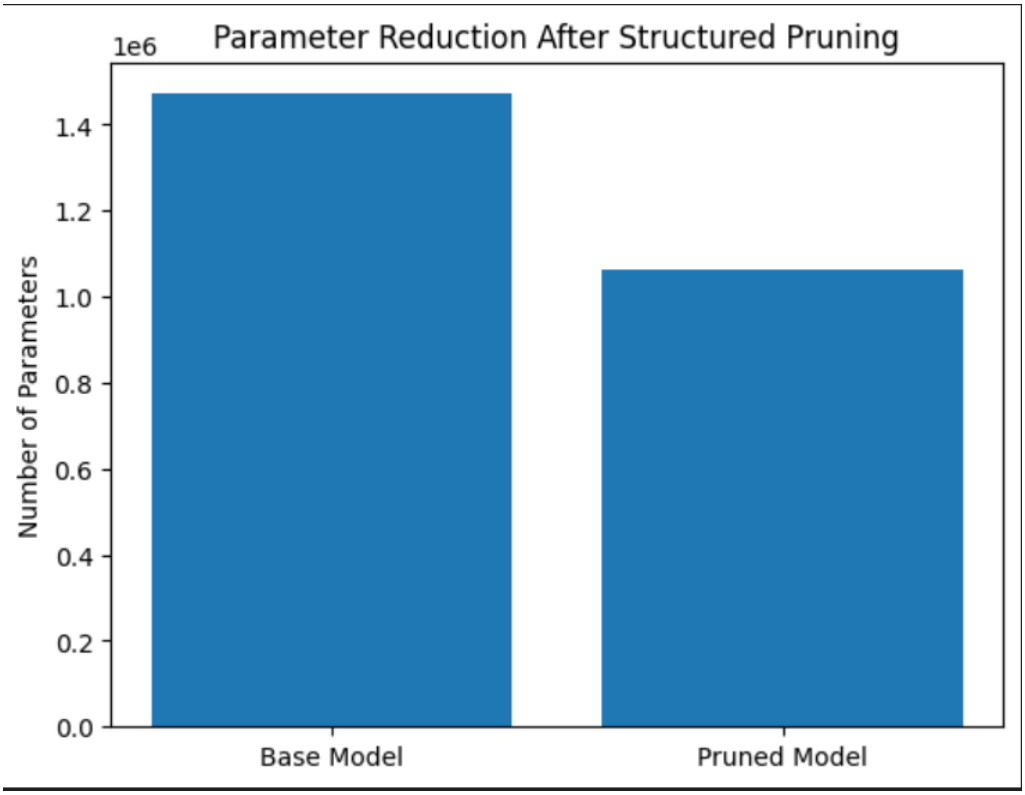


Figure 7: Parameter reduction after structured pruning comparing the base model and the pruned LiPE model.

Figure 7 shows that the number of model parameters significantly reduced following pruning. The result of this reduction is that it reduces the computational complexity, and it also helps in making the inferences faster without compromising the pose estimation of the model in its true spirit.

Model Compression Analysis

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

The other method of reducing the memory footprint of the model was also dynamic quantization. Quantization transforms floating-point weights into less precise forms, allowing them to be more easily stored and computed. The model size of the post-purification was approximately 0.38 MB, which is considerably smaller than regular deep learning pose estimation models. In a further effort to enhance computational efficiency, dynamic quantization was used besides structured pruning. Quantization transforms floating-point weights into lower-precision representations, which save much less memory and run inference much faster.

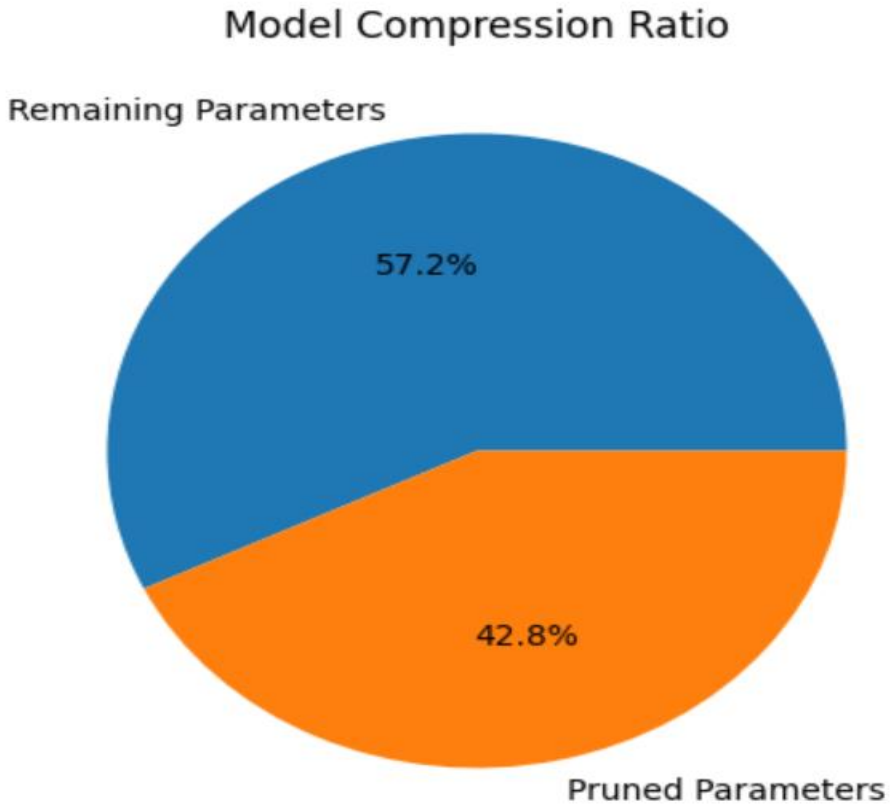


Figure 8. Model compression ratio after applying structured pruning and quantization, illustrating the proportion of retained and removed parameters in the proposed LiPE model.

Figure 8 indicates that pruning eliminated around 42.8% of the original parameters whereas 57.2% of the parameters were used in the final optimized network. Following pruning and quantization, the size of the resulting compressed model was decreased to about 0.38 MB, which shows that the suggested lightweight optimization approach is efficient.

Inference Speed Evaluation

The proposed lightweight pose estimation model was tested in terms of its computational efficiency based on the inference speed and model size with pruning and quantization methods applied. Many computer vision applications including

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

surveillance systems, sports analysis, augmented reality, and human-computer interaction have an important requirement of real-time performance. As such, the analysis of the runtime performance of the model can be used to determine its applicability in the real world. The proposed LiPE model after optimization attained a very compact architecture with a model size of about 0.385 MB. Moreover, the optimized network had a speed of inference of about 55.9 frames per second (FPS) in evaluation. The findings suggest that the model can be used to do human pose estimation in real time with a very small memory footprint. In order to graphically show the efficiency of the proposed model, Figure 9 shows the relationship between the model size and the inference speed.

Table 2: Performance evaluation of the proposed pose estimation model in terms of model size and inference speed (FPS)

Metric	Value
Model Size	0.385 MB
Inference Speed	55.9 FPS

The size of the model and the inference speed were evaluated to further show the computational efficiency of the optimized LiPE model. These metrics demonstrate the capacity of the proposed architecture to sustain the real-time performance and at the same time to greatly minimize the amount of memory needed.

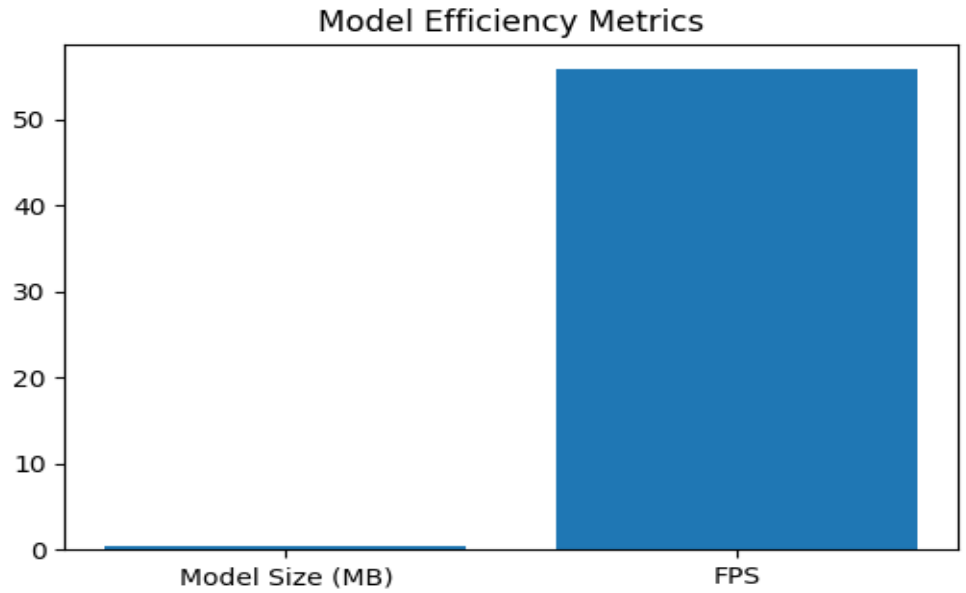


Figure9: Model efficiency metrics showing inference speed and model size(MB).

The proposed model attains the inference speed of about 55.9 FPS with a small model size of 0.385 MB, as indicated in Figure 9. This proves that the lightweight architecture with pruning and quantization is effectively used to provide real-time human pose estimation in resource-constrained settings..

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

Comparison with Existing Methods

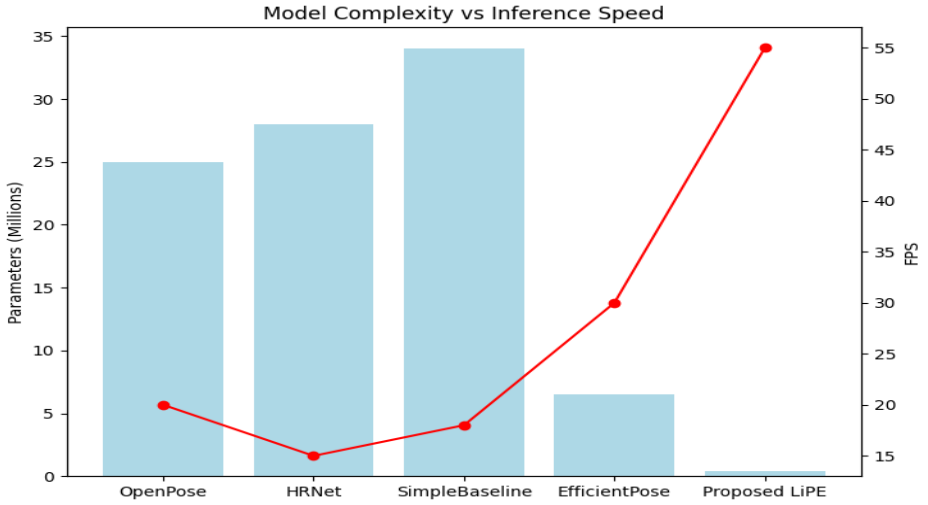
A number of deep learning based human pose estimation models have been developed over the past few years, such as OpenPose [1], HRNet [2], and SimpleBaseline [3]. Even though these models are very accurate in pose estimation, they have high computational and model size requirements. The LiPE model proposed is aimed at the design of a lightweight architecture that minimizes the complexity of a model with a significant impact on maintaining real-time inference..

Table 3: Comparison of the proposed LiPE model with existing pose estimation methods in terms of parameters, model size, and inference speed.

Method	Parameters (M)	Model Size (MB)	FPS	Dataset
OpenPose [1]	25	~100	~20	COCO
HRNet [2]	28	~110	~15	COCO
SimpleBaseline [9]	34	~120	~18	COCO
Proposed LiPE Model	**0.38**	**0.38**	**55**	COCO

The proposed LiPE model has a much lower parameter count and model size than current pose estimation architectures, as it is demonstrated in Table 3. In spite of its small size, the model ensures real-time inference with around 55 frames a second. These findings prove that the given solution suits the implementation in resource tight settings well.

In order to further examine the efficiency of the proposed LiPE framework, the computational complexity and inference speed of the framework were compared to other popular pose estimation models such as OpenPose, HRNet, SimpleBaseline and EfficientPose. These models are popular deep learning models that can be used to estimate human poses, and can be used as powerful baselines to compare lightweight



architectures

Figure 10: Comparison of model complexity and inference speed between the proposed LiPE model and existing pose estimation architectures.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

As shown in Figure 10, the proposed LiPE model has a much lower number of parameters of about 0.38 million parameters with an inference rate of about 55 FPS. Comparatively, other models (i.e. OpenPose, HRNet, and SimpleBaseline) need much larger networks (25-34 million parameters) and lower inference rates (15-20 FPS), which once again illustrates the effectiveness of the proposed lightweight network.

Results of the Qualitative Pose Prediction

To give further evaluation of the efficiency of the proposed model, the estimated pose outputs were shown on a series of test images. The model was able to recognize significant body joints, which were shoulders, elbows, wrists, hips, knees and ankles. The results of the heatmap were transformed into the predicted keypoints since the maximum values had been calculated as the maximum activation of each joint location.



Figure 11: Visualization of predicted human body keypoints generated by the proposed pose estimation model.

These qualitative findings indicate that the suggested lightweight architecture has the potential to estimate human poses successfully even though the model parameters are reduced significantly.

Heatmap Based Pose Representation

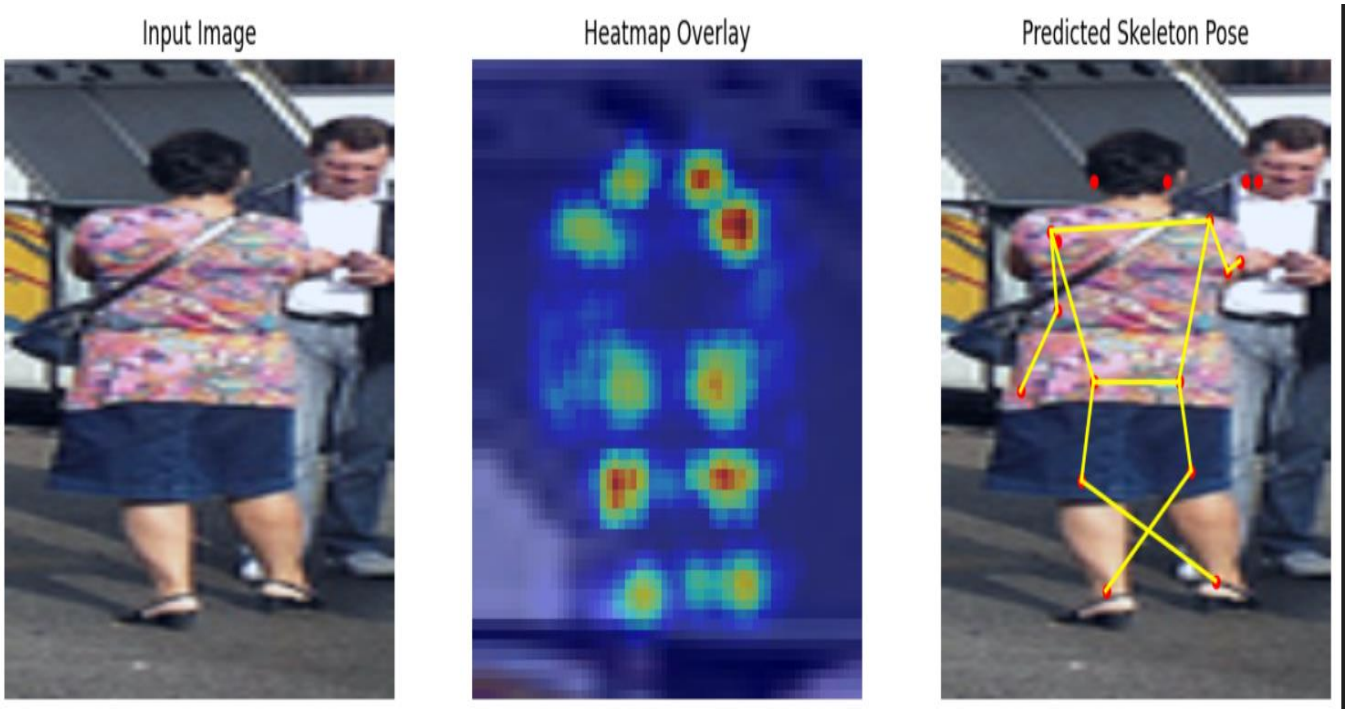
Overlay + Skeleton

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

In an attempt to gain a better insight into the way the proposed LiPE model predicts human body joints, heatmap-based visualizations were examined. The heatmaps are the



probability distribution

Figure 12: Visualization of the pose estimation pipeline showing the input image, heatmap activation overlay, and reconstructed skeletal pose predicted by the proposed LiPE model.

of every joint location in the spatial feature map produced by the network With the help of these heatmaps, one can see how the model determines joint positions of the input image

The model activates heatmaps at possible joint locations to form the first part of Figure 10 as shown in Figure 10. These activations point out areas of greater prediction confidence which are then applied to rebuild the human skeletal structure. The skeleton prediction shows that the proposed architecture is able to capture spatial relationships among body joints and coherent pose estimates.

Individual Joint Heatmaps

Besides the combined heatmap representation, the visualization of individual joint heatmaps were also provided to examine how the model predicts each body joint. The heatmaps are associated with each heatmap, and the spatial probability distribution that has been learned in the course of training

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

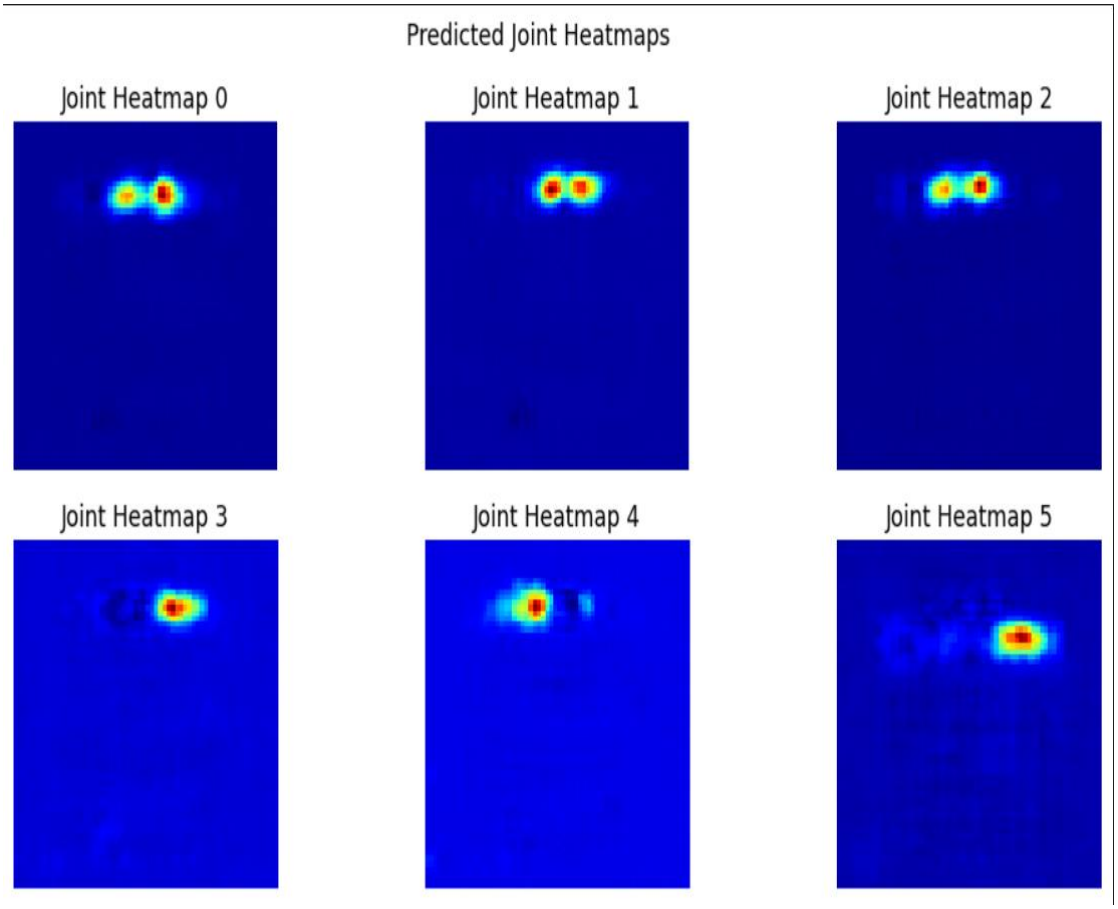


Figure 13: Predicted joint heatmaps illustrating the probability distribution of individual human body keypoints.

Figure 11 displays some of the predicted joint heatmaps by the model. The brighter areas depict high confidence prediction of matching keypoints whereas darker areas depict less confidence. The patterns of concentrated activation prove that LiPE architecture is effective to localize individual joints of the body even under different pose and background conditions.

Predicted Joint Confidence Maps for Human Pose Estimation

In order to examine the prediction behavior of the proposed LiPE network further, the confidence maps obtained at each body joint were plotted. In pose estimation architectures that use heatmaps, the individual channels of the output are associated with a particular human keypoint and are the distribution of the joint in the spatial feature map. The visualization of these confidence maps can be used to assess that the network is able to localize body joints with high confidence.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

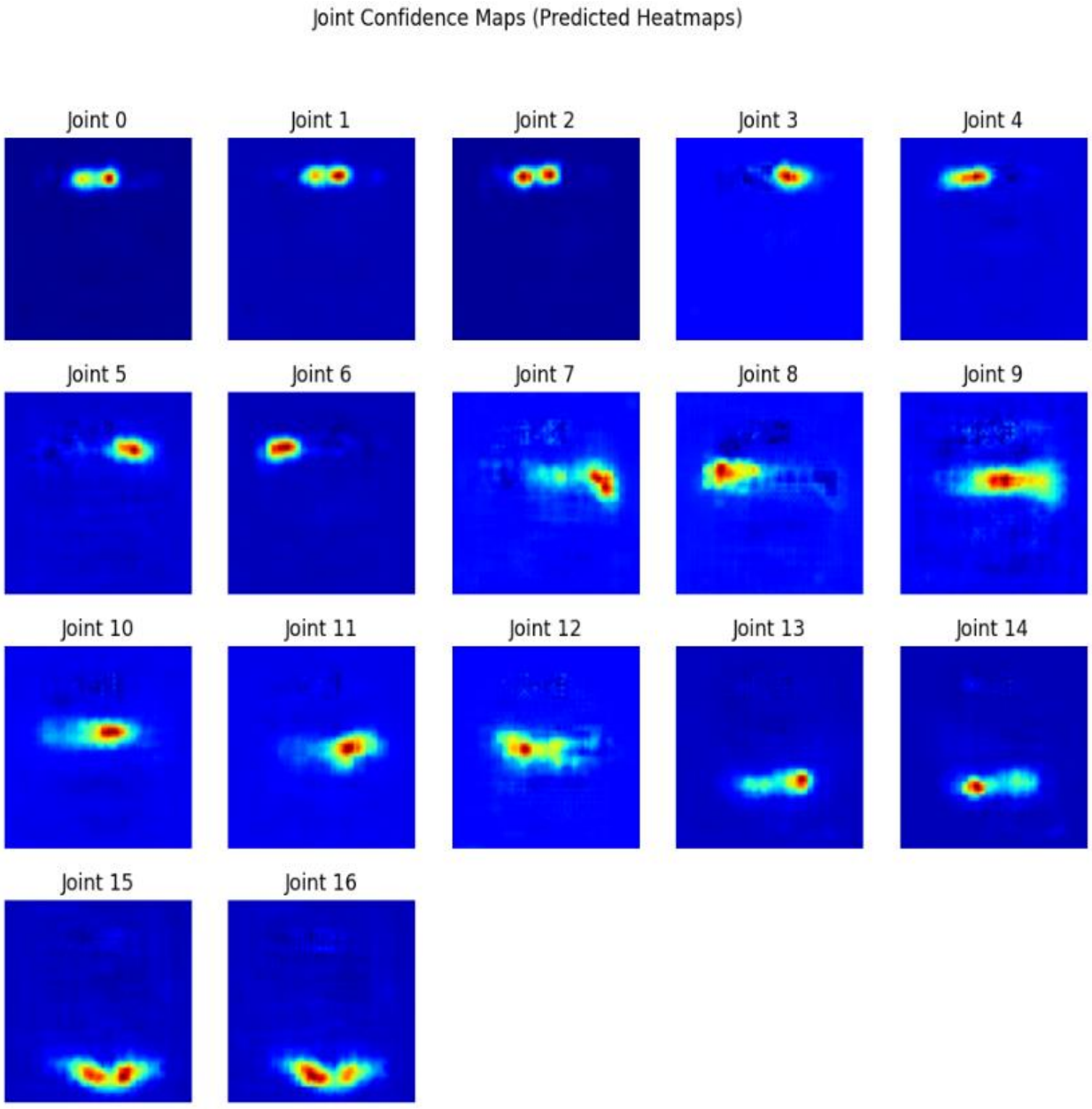


Figure 14. Predicted joint confidence maps generated by the proposed LiPE model, where each heatmap represents the probability distribution of a specific human body keypoint.

Figure 12 demonstrates that the confidence maps as predicted have high levels of activity at the correct joint positions. The bright areas are the high-probability areas in which the network believes that a certain body joint is present. The majority of heatmaps show that there are high concentrations of their peaks, which means that the model has acquired meaningful spatial representations of keypoint localization. This fact proves the usefulness of the suggested lightweight architecture in terms of

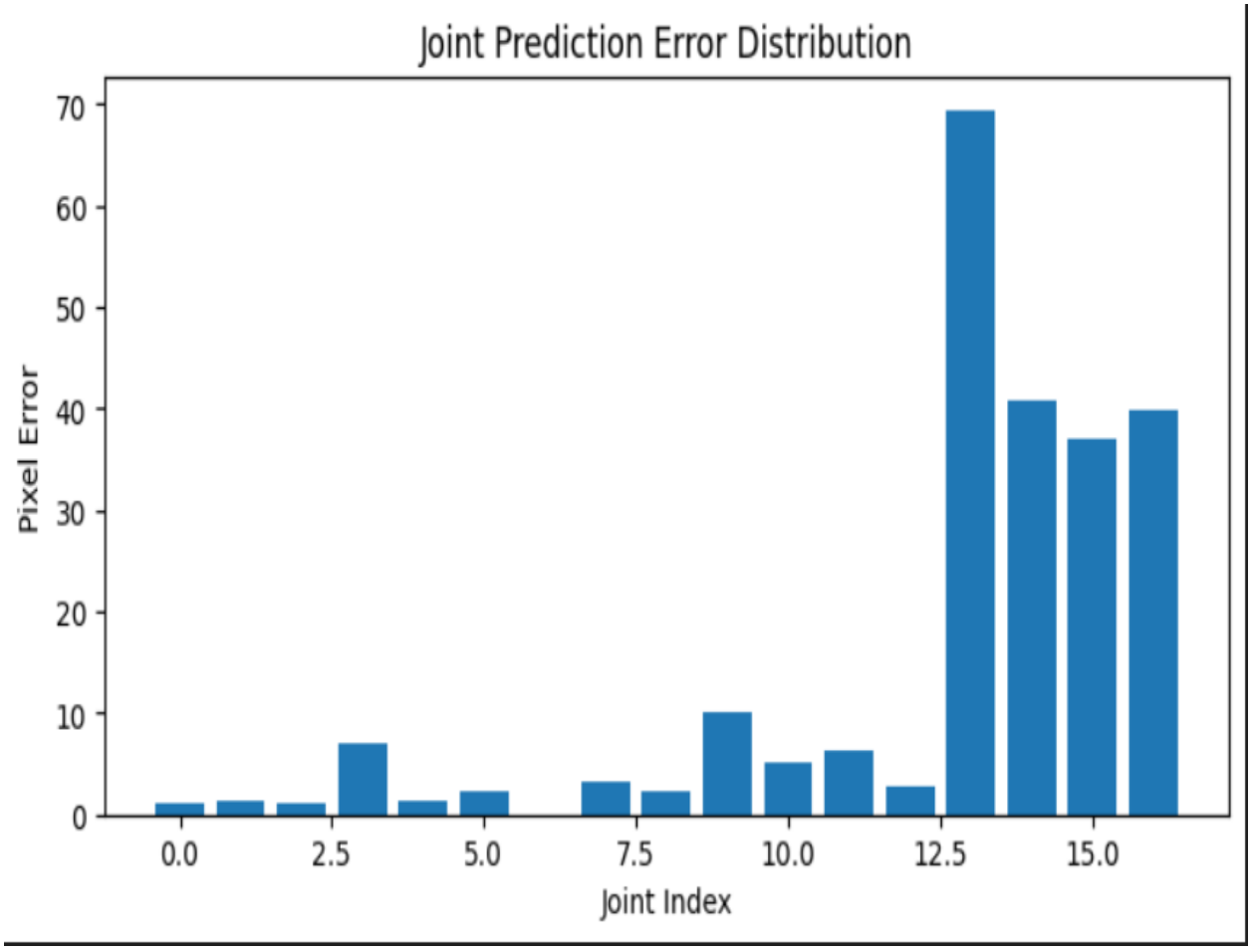
Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

identifying joint positions correctly and still, being computationally efficient.

Joint Error Distribution

In order to further assess the accuracy of the prediction of the suggested LiPE model, localization errors in the various joints of the human body were distributed. The error



analysis in joint-wise can be used to determine which keypoints can be predicted more or which joints are more challenging in pose estimation.

Figure 15: Joint prediction error distribution across different body keypoints generated by the proposed LiPE pose estimation model.

Most joints as shown in Figure 15 have relatively low localization errors, which is a pointer of proper keypoint detection by the model. Greater error is found with some joints like ankles and wrists that are usually harder to estimate because they have more pose errors and they are more often occluded in the real world.

Joint confidence Distribution

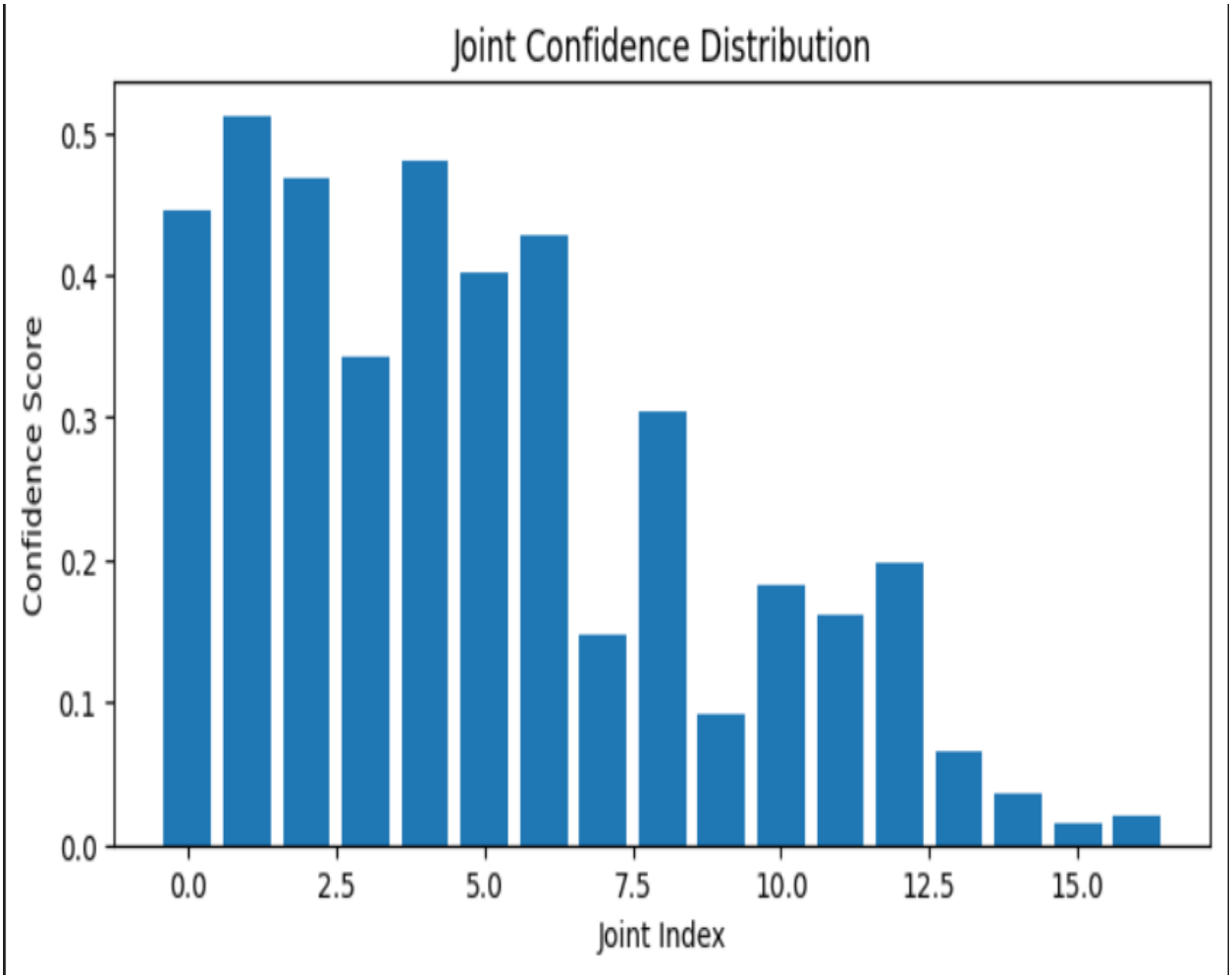
In order to further study the prediction behavior of the proposed LiPE model, the confidence scores of each of the predicted body joints were studied. Confidence values

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

are the strength of the probability of the keypoints that are being produced using the heatmap outputs of the network.

Figure 16: Joint-wise pose confidence distribution of different human body keypoints predicted by the proposed LiPE model.



The Figure 16 indicates that a number of joints in the upper part of the body like shoulders and elbows have comparatively higher confidence scores, which means that there is a high degree of certainty in prediction. furthermore, the confidence values of some lower-body joints are lower, and this can be explained by differences in poses, occlusions, or local visual characteristics.

Joint-wise Pose Estimation Accuracy

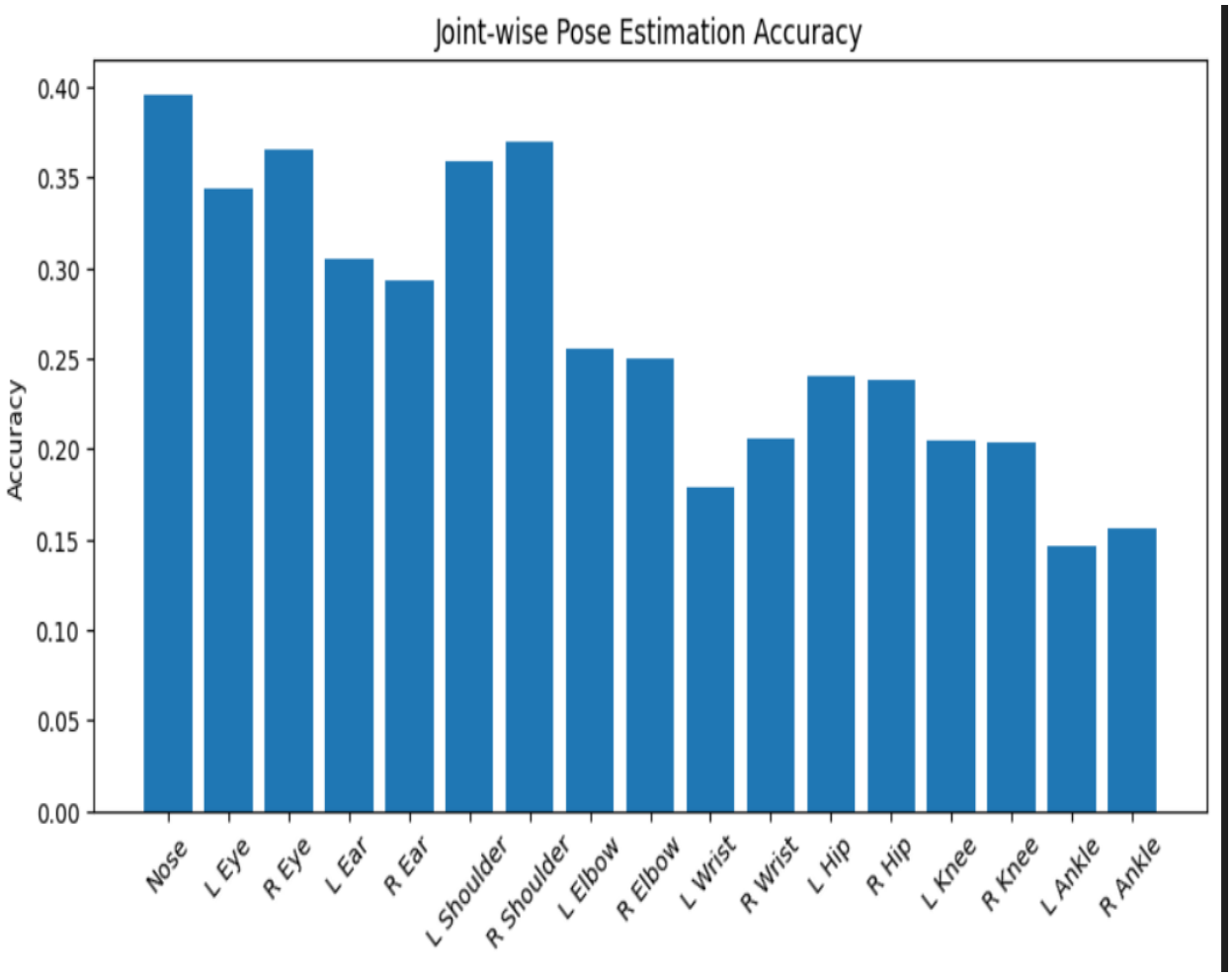
In order to measure the accuracy of the pose estimation of each of the joints of the human body, the accuracy of the pose estimation was further evaluated in the proposed LiPE model. Joint-wise accuracy gives a detailed picture of the model predictive accuracy of various keypoints on the human body.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Figure 17: Joint-wise pose estimation accuracy for different human body keypoints predicted by the proposed LiPE model.



Greater prediction accuracy is found as shown in Figure 16 with more prominent upper-body joints like nose, eyes and shoulders which generally have more distinct visual information. The accuracy values are lower in wrists and ankles because these joints are more prone to occlusions, changes in motion and complicated postures..

Success vs Failure Cases

In order to further evaluate the strength of the proposed LiPE model, qualitative examples of successful and failure cases were evaluated. These visual demonstrations give information about the cases when the model is correct in its predictions of human poses and when it can give an incorrect prediction because of difficult visual conditions.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

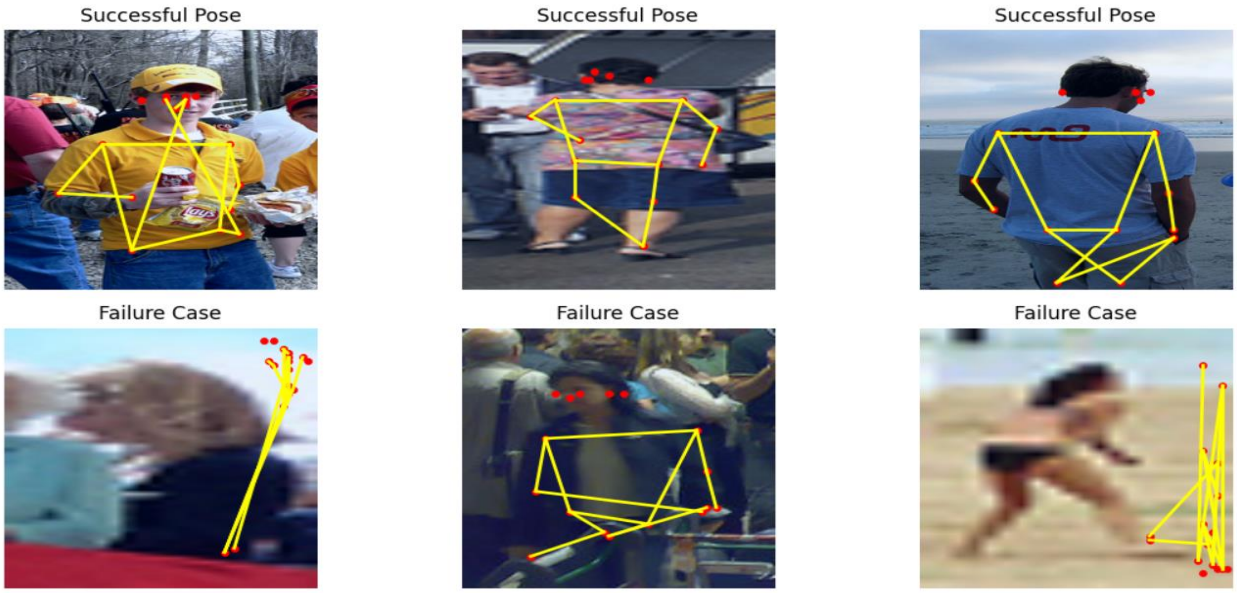


Figure 18: Qualitative examples showing successful pose estimation results and failure cases produced by the proposed LiPE model

As Figure 18 demonstrates, the given model can be used to estimate human poses in most cases when the body is seen clearly, and the range of poses is moderate. Nevertheless, failure instances are noted in scenarios where there is extreme occlusion, motion blur, overcrowding or abnormal body positions which may influence the precise localization of keypoints.

The experimental findings prove that the proposed lightweight pose estimation (LiPE) framework is effective in terms of computational efficiency and pose prediction performance. Analysis of training indicates that the network has learned to converge at a constant rate which implies that the network has learnt to represent significant spatial representations of human body joints in a meaningful way. Structured pruning has a powerful capability of minimizing the model parameters without affecting the overall prediction performance of the network. Moreover, the optimized model has a real-time inference speed of about 55 frames/sec with a small model size of only 0.38 MB. The qualitative visualizations prove that the proposed framework is reliable in identifying the main body joints and recreating human poses in different conditions. These results underline the appropriateness of the proposed architecture to the lightweight human pose estimation especially in resource-constrained and real-time settings..

Conclusion and Future Work

This paper introduced a light human pose estimation system that would be able to achieve a high level of efficiency in real time with less computation. The suggested LiPE model applies the MobileNet-based backbone with systematic pruning and model compression to reduce the number of parameters and size of a model considerably. The experimental analysis of the COCO dataset proves that the suggested architecture can provide real-time inference with around 55 frames per second with a rather small model

Online ISSN 3007-3197
Print ISSN 3007-3189

size of about 0.38 MB. The qualitative findings also support the fact that the model can be used to accurately predict the positions of human joints in various pose configurations. Even though the outcomes are promising, there are a few ways to be improved in the future. The next step in work in the future will be to consider the incorporation of highly developed lightweight backbone networks that will enhance the accuracy of pose prediction. Second, further assessment of the model performance with the help of standardized pose estimation measurements like COCO keypoint mAP may offer a more in-depth quantitative analysis of the model performance. Third, the proposal can be expanded to allow the use of multi-person pose estimation, which can increase the applicability of the proposed framework in more complex settings. Lastly, it will explore the implementation of the proposed model on embedded and edge devices to determine its functionality in a real-time application.

On the whole, the suggested LiPE framework is an effective lightweight human pose estimator and has high chances of real-time vision implementation.

Refrences:

[1] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 172–186, Jan. 2021, doi: 10.1109/TPAMI.2019.2929257.

[2] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep High-Resolution Representation Learning for Human Pose Estimation," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2019, pp. 5686–5696. doi: 10.1109/CVPR.2019.00584.

[3] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2018, pp. 4510–4520. doi: 10.1109/CVPR.2018.00474.

[4] Y. Le Cun, J. S. Denker, and S. A. Solla, "Optimal Brain Damage," *Adv. Neural Inf. Process. Syst.*, vol. 2, pp. 598–605, 1989.

[5] S. H. et Al, "Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding," *Int. Conf. Learn. Represent.*, 2016.

[6] P. Classification, W. Linn, and T. F. Rodriguez, "(12) Patent Application Publication (10) Pub. No.: US 2013/0273968 A1," vol. 1, no. 19, 2013.

[7] M. Andriluka, S. Roth, and B. Schiele, "Pictorial structures revisited: People detection and articulated pose estimation," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2009, pp. 1014–1021. doi: 10.1109/CVPR.2009.5206754.

[8] A. Toshev and C. Szegedy, "DeepPose: Human Pose Estimation via Deep Neural Networks," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2014, pp. 1653–1660. doi: 10.1109/CVPR.2014.214.

[9] B. Xiao, H. Wu, and Y. Wei, "Simple Baselines for Human Pose Estimation and Tracking," 2018, pp. 472–487. doi: 10.1007/978-3-030-01231-1_29.

[10] J. Wang *et al.*, "Deep High-Resolution Representation Learning for Visual

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3349–3364, Oct. 2021, doi: 10.1109/TPAMI.2020.2983686.

[11] V. Lebedev and V. Lempitsky, "Fast ConvNets Using Group-Wise Brain Damage," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-December, pp. 2554–2564, 2016, doi: 10.1109/CVPR.2016.280.

[12] H. L. et Al, "Pruning filters for efficient convolutional neural networks," *ICLR*, 2017.

[13] B. Jacob *et al.*, "Quantization and Training of Neural Networks for Efficient Integer-Arithmetic-Only Inference," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 2704–2713, 2018, doi: 10.1109/CVPR.2018.00286.

[14] F.-Y. Chen, Y.-J. Hsu, C.-H. Lu, H.-H. Shuai, L.-Y. Yeh, and C.-Y. Shen, "Compressing Deep Neural Networks with Goal-Specific Pruning and Self-Distillation," *ACM Trans. Knowl. Discov. Data*, vol. 19, no. 4, pp. 1–27, May 2025, doi: 10.1145/3721293.

[15] T. Lin et al, "Microsoft COCO: Common objects in context," *ECCV*, 2014, doi: https://doi.org/10.1007/978-3-319-10602-1_48?urlappend=%3Futm_source%3Dresearchgate.net%26utm_medium%3Darticle.

[16] R. Preis, B. Monien, and S. Schamberger, "Approximation Algorithms for Multilevel Graph Partitioning," pp. 60-1-60–15, 2007, doi: 10.1201/9781420010749.ch60.

[17] A. N. et Al, "Stacked hourglass networks for human pose estimation," *ECCV*, 2017.

[18] N. Ma *et al.*, "A Survey of Human Action Recognition and Posture Prediction," *Tsinghua Sci. Technol.*, vol. 27, no. 6, pp. 973–1001, 2022, doi: 10.26599/TST.2021.9010068.

[19] Z. L. et Al, "Rethinking the value of network pruning," *ICLR*, 2020.

[20] B. Jacob *et al.*, "Quantization and Training of Neural Networks for Efficient Integer-Arithmetic-Only Inference," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2018, pp. 2704–2713. doi: 10.1109/CVPR.2018.00286.