ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 11, Issue 12, December 2024

Models And Systems for Pattern Recognition in Human Identification Through Neural Networks

Himmatov Ibodilla, Akmal Akhatov, Premanand V

PhD student. Samarkand state university named after Sharof Rashidov, Samarkand, Uzbekistan
Sr. Professor, Vice-rector of Samarkand state university named after Sharof Rashidov, Samarkand, Uzbekistan
Assistant Professor. VIT Chennai Campus, Chennai, India

ABSTRACT: The article discusses the study of systems and models for personal identification based on human action recognition. This topic includes computer vision, deep learning, and motion analysis, which allow people to be identified or authenticated through their actions or behaviors. In addition, the main content of the article includes models for personal identification using facial recognition, voice recognition, fingerprint recognition, signature verification, and iris recognition. In addition, attention is paid to behavioral biometrics, multimodal identification, and document verification models. The article emphasizes the use of Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and 3D ResNet as key technologies and algorithms, which are highlighted for their role in motion recognition and personal identification. The need for human pose estimation and motion graph generation is also discussed, with emphasis on body position analysis and the use of neural networks to generate heat maps for key points. Based on this framework, the paper also analyzes 3D position estimation, solving problems such as occlusion caused by overlapping objects and changes in viewpoints. Applications and practical use cases are also shown, and these technologies can enhance security in public places such as airports, markets, and other crowded places. They are also useful in health monitoring and sports analysis. Regarding the opportunities and limitations, the paper reviews the current state of the technologies and emphasizes the need to develop and improve algorithms to increase their overall usability. The main idea of the paper is that analyzing human movements through neural networks allows for accurate and efficient identification. It also emphasizes the need to improve models and algorithms and improve data processing methods in the future.

I. INTRODUCTION

In the context of machine learning and deep learning, recognition models typically refer to models designed to perform recognition-related tasks. These tasks often involve identifying and verifying the identity of individuals or legal entities based on various data sources. To date, identification models are rapidly improving, and we will mention their most common types and those that are actively used today, as well as their main tasks.

The purpose of the research is to evaluate the human motion and identify the person, similar to the problem of arbitrary computer vision, even in motion detection, which does not change the content of the video in any problematic situations, but changes the external appearance and therefore affects the internal state of the gait. there are many factors that do. All these factors are divided into two groups: those affecting the image itself and those affecting only the condition. The factors of the first group really change the way you walk and are more difficult to recognize for both humans and computers. In the second group, there are conditions that change only the internal representation, and the way a person walks does not change. To the human eye, the two videos will show the same person, but to the computer, they will be two different image sequences with different characteristics. Today, the process of identifying a person by his actions is one of the most difficult and fast-growing areas compared to other identification technologies. So, first of all, it is necessary to consider the drawings that form human pose and unig graphs using neural networks and other tools.

Estimating human pose involves identifying key points like joints (shoulders, elbows, hips, knees, feet) and others (neck, head). This task can be approached in 2D or 3D, influencing its complexity and practical use.

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 11, Issue 12, December 2024

II. RECOGNITION MODELS IN PERSON IDENTIFICATION

Recognition models for personal identification are rapidly improving today. We present the following models as the most popular of these models:

Facial recognition models. The most popular models today are Face Recognition models, which are mainly designed to identify and verify individuals based on facial features, and serve as security systems for applications, access control, and facial authentication for mobile devices. FaceNet, VGGFace and OpenFace models are popular models in face recognition nowadays.

Fingerprint recognition models. Another popular model is Fingerprint recognition models, which also use fingerprint patterns to identify and verify individuals. These models are widely used in law enforcement, border control, and mobile device security. Popular models used in fingerprint recognition include the Minutiae-based fingerprint recognition system and deep fingerprint matching networks. these two models serve as the main factor in the security systems of the customs control systems, and these two recognition parameters are included in the common database in the citizen's passports.

Voice recognition models. these models are also being used positively in practice. Voice recognition models identify people based on voice characteristics. Application security systems include voice authentication for laboratories and voice assistants, which are critical for opening safes in banks. Common models include deploying depth speakers and convolutional neural networks (CNNs) for voice recognition.

Signature Verification Models. These models are becoming a well-developed recognition system in society. These models verify individuals based on their handwritten signatures. These models find applications in banking, document verification, and fraud detection, and popular models in this field include Siam Networks and Online Signature Verification models, which are proving effective today.

Biometric models of behavior. Behavioral biometric models analyze a person's specific behaviors, such as typing, walking, or mouse movements. These models are used in continuous authentication and fraud detection, and in determining behavior based on psychological questionnaires, as well as in determining the types of character in a person. Specific algorithms differ depending on the type of behavior being analyzed. Based on this model, today's behavioral biometric gait recognition systems are also rapidly developing.

Iris recognition models. Iris recognition models identify people by analyzing unique patterns in their irises. Applications include border control, airport security, and access control. Popular algorithms in this model include Daugman's Integro-Differential Operator and deep learning-based approaches.

Multimodal identification models. Multimodal identification models combine information from various databases, such as face and voice recognition or fingerprint and iris recognition. the main reason for this involves the formation of a two- or three-level protection system. which fully satisfies the need to sufficiently strengthen reliability and security in any system, and these models offer high accuracy and robustness in identity verification. fusion techniques such as late fusion and early fusion are commonly used for these models.

Document review models. Document verification models are based on all official documents that authenticate individuals. They are used in border control, financial services and KYC (Know Your Customer) processes. These models may include optical character recognition (OCR) and verification of document security features.
Biometric Fraud Detection Models. These models are designed to detect and prevent fraud attacks where an adversary tries to trick the system with fake biometrics. Fraud detection models often use neural networks to determine the authenticity of biometric data.

The main components of the system include factors common to these recognition systems, and they are as follows:
Data collection and preparation. It is very important to collect a complete data set of human activities. These data sets usually include video recordings or sequences of actions performed by different individuals. Data preprocessing techniques are used to extract relevant features such as optical flow, pose estimation, or motion vectors. It is effective to use data augmentation techniques to ensure the diversity, stability, and security of the data set.

Neural network architectures. Various neural network architectures are used for task-specific motion detection and person identification. The most popular architectures are CNN (convolutional neural networks) and RNN (recurrent neural networks), and 3D CNN. In this case, CNN excels at extracting spatial features from video frames, while RNN is suitable for modeling temporal relationships in action sequences, and 3D CNNs are also designed to process spatio-temporal data and adapts to motion detection tasks [1].

Model design and training. Neural networks are designed to process captured features and make predictions about the actions and individuals involved. During training, the model learns to recognize actions and individuals from a given set of data.

Fusion feature. Summarizing data from different platforms and databases is a widely used process, such as RGB frames and optical flow, for identity identification. For this, two-stream networks combining spatial and temporal information are often used.
Popular models of identity based on motion recognition [2]:

Two-stream networks. A two-stream CNN processes RGB frames and optical streams in separate streams before combining them for prediction. This approach makes efficient use of spatial and temporal information.
Long short-term memory networks (LSTM). LSTMs are also recurrent neural networks that excel at capturing temporal relationships in action sequences. They are valuable in modeling movements with long-term dependencies.

3D residual networks (3D ResNets). 3D ResNets extend the concept of residual networks to a 3D convolutional architecture. They are well suited for motion detection tasks due to their ability to capture spatio-temporal features.

The need for evaluation and placement is also a key factor, where model evaluation measures the performance of motion detection models in recognizing a person in a specific test dataset. For example, it can be in the form of accuracy, recall, or average mean accuracy (MAP). is evaluated using Cross-validation can be used to assess model generalizability [3]. Based on the model deployment process, the deployed models can be used in real-world applications, including live video streams or pre-recorded videos for human identification. In turn, we face difficulties in the process of biometric identification. As challenges, i.e., variations in lighting, viewpoints, and occlusions, among the disruptive factors, are privacy and security concerns in applications such as tracking for accurate motion detection and identification.

Deep Learning Frameworks. TensorFlow, PyTorch, and Keras as the most popular Deep Learning-based libraries are commonly used to implement and train neural networks in this domain. The applications created for these tasks find applications in motion detection-based personal identification tracking, health monitoring, human-computer interaction, sports analytics, and more.

Neural network algorithms are fundamental in recognition systems, including image recognition, speech recognition, and various other pattern recognition tasks. These algorithms require the use of artificial neural networks, modeled after the structure and functions of natural biological neural networks. Here are some of the key neural network algorithms used in recognition systems [4].

Multilayer perceptrons (MLPs). Multilayer perceptrons are also being used as a very efficient form of neural network, consist of layers of interconnected neurons or units. These networks process input data sequentially through multiple hidden layers to produce an output. FNNs are commonly used for various recognition tasks, including image classification and speech recognition.

Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). Convolutional Neural Networks are specialized artificial intelligence models primarily designed for processing images and spatial data. They use convolutional layers to extract spatial hierarchies of features from input data. These networks excel in tasks such as image classification, object detection, and facial recognition. Thanks to their spatial analysis-based technology, CNNs are widely applied in various fields that require vision and image processing. Recurrent Neural Networks (RNNs) are also widely used. RNNs have been providing effective solutions in speech recognition, natural language processing, and gesture recognition.

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 11, Issue 12, December 2024

Long Short-Term Memory Networks (LSTMs) and Gated Recurrent Unit (GRU). LSTMs are a type of RNN designed to solve the vanishing gradient problem, allowing them to obtain efficient solutions for extracting long-term relationships in sequences. LSTMs are widely used in areas such as speech recognition, machine translation, and sentiment analysis. GRUs, similar to LSTMs, are used to extract long-term relationships in sequential data. They are simpler than LSTMs and are used in applications such as natural language processing and speech recognition.

Autoencoders. Autoencoders are neural networks used for unsupervised learning and dimensionality reduction. They are employed in feature extraction and data compression for various recognition tasks.

Self-Organizing Maps (SOMs). SOMs are also a type of unsupervised neural network that is convenient for clustering and is effectively used for visualization tasks. They can be used for recognizing patterns and grouping similar data points together.

Radial Basis Function Networks (RBFNs). RBFNs are used for classification and regression tasks. They consist of an input layer, a hidden layer with radial basis functions, and an output layer. RBFNs are often applied in speech recognition and function approximation.

Residual Neural Networks (ResNets). ResNets are specifically developed to tackle the issue of vanishing gradients that occurs in extremely deep neural networks. By introducing shortcut connections, ResNets enable smoother gradient flow during backpropagation, ensuring that deeper layers can be trained effectively. These networks are widely recognized for their ability to improve performance in complex tasks while maintaining computational efficiency. They have skip connections that allow for the training of deep networks, making them suitable for complex recognition tasks.

Attention Mechanisms. Attention mechanisms are often incorporated into neural network architectures to focus on specific parts of the input data. This approach allows the network to dynamically weigh different parts of the input based on their importance to the task at hand. Attention mechanisms play a crucial role in natural language processing applications, including machine translation, document summarization, and text generation, by enabling more accurate and context-aware processing of information.

These neural network algorithms can be combined and customized to meet the requirements of specific recognition systems, and they continue to be a driving force behind advancements in fields such as computer vision, speech processing, and natural language understanding.

III. ESTIMATION OF HUMAN POSE

Human pose detection is the most important tool for human recognition and identification. The number of subjects is relatively small, but it has a large variation of views: a total of 124 subjects were recorded from 11 distinct viewing angles, ranging from 0° to 180° in increments of 18°. The video recordings vary in length from 3 to 5 seconds, featuring a frame rate of approximately 24 frames per second (fps) and a resolution of 360×280 pixels. These recordings provide a comprehensive dataset, capturing diverse perspectives and ensuring sufficient detail for analysis. All trajectories are recorded directly and closed to all people in the same room. In addition to the diversity of perspectives, as in the first collection, different clothing and various carrying scenarios are included, such as wearing a coat or carrying a bag. In total, each individual has 10 video sequences captured from each viewpoint: 6 sequences depict normal walking without any added conditions, 2 sequences show walking while wearing open clothing, and 2 sequences involve walking while carrying a bag. This setup ensures a diverse range of conditions for comprehensive analysis.The CASIA dataset is popular because of its variability in visibility. However, there are very few cases where a deep neural network can be trained to recognize any kind of gait without overfitting.

Therefore, the spatial random field (MRF) model was used to assess the internship between the joints and the main articulation points of the human body in order to ensure the correlation of the main points. As a more efficient method than this method, a new architecture called "conv-deconv" with very deep sequences with large receptive fields has also been developed for direct pose matching using heatmaps [5]. Later, it was found that the interpolation between conv-explorer pairs was introduced to avoid gradient loss, and then a deeper neural network could learn and analyze the

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 11, Issue 12, December 2024

practical aspect between the keypoint detection area. This means that Deep Neural Network can be implemented. Thus, we see that there is an increased interest in human state detection and the need to develop new algorithms to boost the invariance of deep CNN at various levels by learning convolutional filters at different feature resolutions using PRM [6]. Based on this, several other types of neural networks can be used to estimate human pose, and we can see that modern DNNs for estimating human pose are still not fully developed in modeling the structure of the human body to find and generalize the main joint points, so based on existing methods, neural networks must be sufficiently developed to model the pose nodes, i.e. joints. For this purpose, in order to have enough external factors for the main points of the pose, one of the main issues is to eliminate the congestion in the recognition scene, the ambiguities caused by the background noise, or the cases where several parts of the body block each other, as the main weakness in this process. Therefore, based on the pose estimation method, multi-pose estimation (MPII) relies on several iterations of pose estimation methods at different scales to averaging the results of many methods and techniques to significantly increase efficiency, even by a small margin [7]. This shows the need to find effective solutions to solve the main task of work and system priorities in modeling [8].

In this study, we aim to divide the main task into two main types, namely, assessing the position of a single person and assessing the position of each person in a crowd.

Detecting a person's pose is a major challenge in computer vision because humans are inherently deformed, meaning they can be in different positions, so there are many variations in pose.

In human pose assessment images or videos, human joints (also considering key points - elbows, wrists, etc.) are appropriate for the situation [9]. This is a set of poses that are all articulated poses. Images or videos can be used to perform tasks that can control the movement of human joints.

We consider the human condition as having two-dimensional and three-dimensional methods. In 2D, each joint in the RGB image is assigned a pair of two-dimensional coordinates (x, y) to represent its position. Using 3D methods (such as a Position Estimator), this 2D position is then estimated and mapped to the 3D space, providing a corresponding set of coordinates in the form of (x, y, z) within the RGB image, accounting for depth and spatial positioning [10]. We should pay attention to the fact that it is important to see the state of a person - to process the information describing the graphic form of the movement

IV. CREATING A GRAPHIC FORM OF MOVEMENT

The core of the issue lies in identifying the essential features of human motion when observing an individual's actions in order to address the challenges of "person recognition" and the demonstration of movements [10]. Leveraging these insights, the fast identification function aids in swiftly recognizing potentially suspicious individuals in crowded public spaces, such as airports, bus terminals, shopping areas, and other densely populated environments. This capability allows for the rapid implementation of appropriate security measures based on the investigation's findings. Additionally, it facilitates a detailed examination of images captured by surveillance cameras through the use of advanced neural network algorithms, enabling more accurate and comprehensive analysis.

One of the primary techniques for constructing a motion graph is the graph formation method, where image skeletons are created by connecting points that correspond to joints in three-dimensional space (Figure 1). This approach enables the analysis of human behavior by identifying and defining specific actions, providing valuable insights into the psychological state of a person during movement. By mapping and examining these joint-based skeletons, it becomes possible to gain a deeper understanding of how individuals interact and perform various tasks.

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 11, Issue 12, December 2024

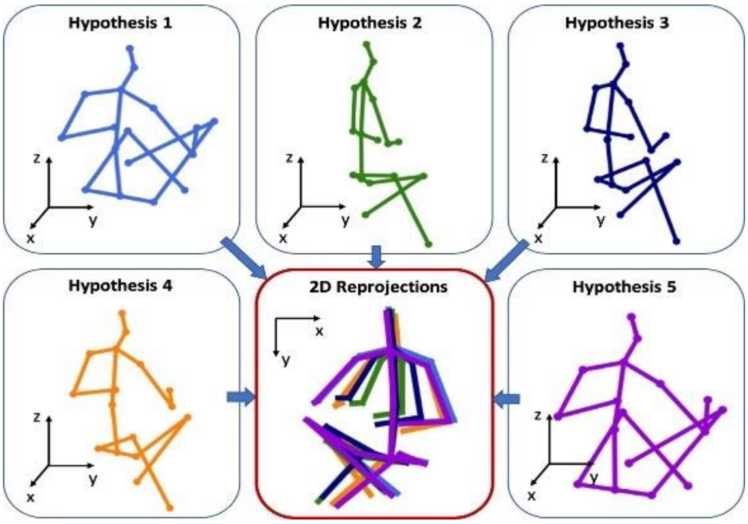

Figure 4.1. A three-dimensional model of the human skeleton in different types of motion

The main idea of this method is to isolate and find "key points on the body" in the image, that is, the main parts of the body such as the joint, shoulder, ankle, knee, wrist, etc., combined with the connected skeleton [11]. The most important thing is to find a set of data an important part of research. An effective method of recognizing the images taken from the camera and determining the human reference points was developed, and the stage of showing the human skeleton separately was carried out based on the reference points. As a result, various possible configurations for identifying key reference points on the human body were designed and illustrated (Figure 4.2). These configurations highlight the critical locations on the body that are essential for analysis and understanding, offering different approaches to effectively determine and map these important points for further examination.

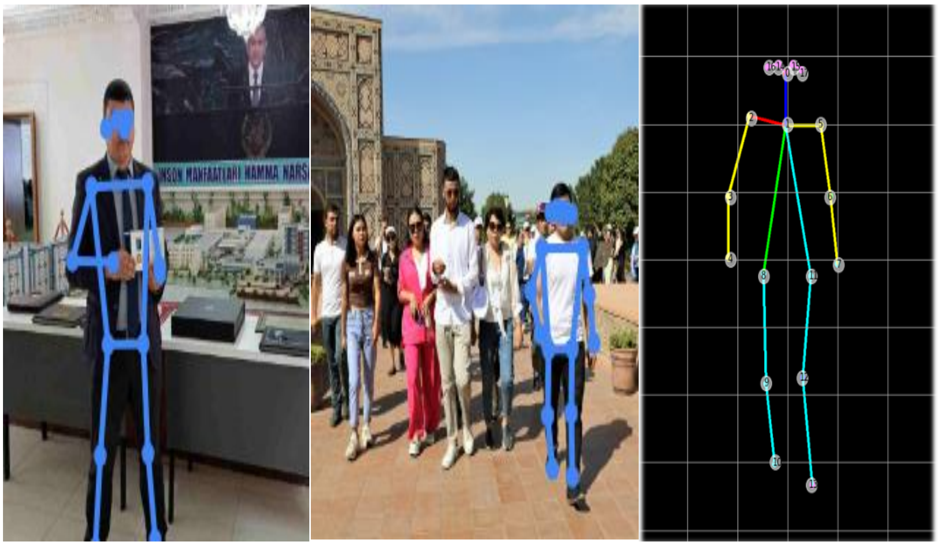

Figure 4.2. Determining the important points of the human body using fulcrums.

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 11, Issue 12, December 2024

V. HUMAN POSE ESTIMATION USING DEEPPOSE

One of the pioneering studies in predicting a person's pose within an image using deep neural networks is DeepPose. This work introduced a novel approach that leverages deep learning techniques to accurately estimate and predict the position and orientation of key body parts, marking a significant advancement in the field of pose estimation and computer vision. DeepPose laid the foundation for many subsequent developments in human pose recognition, enabling more precise and reliable analysis of human body movements from images. Many consider this problem to be a regression problem. DeepPose is an estimation of a person's pose from an image based on given joints, that is, when determining a person's pose, cutting out the areas where a person is present, ecomputing and standardizing the joint coordinates based on the pixelated areas relative to the center of the joints solves the problems (Figure 5.1).

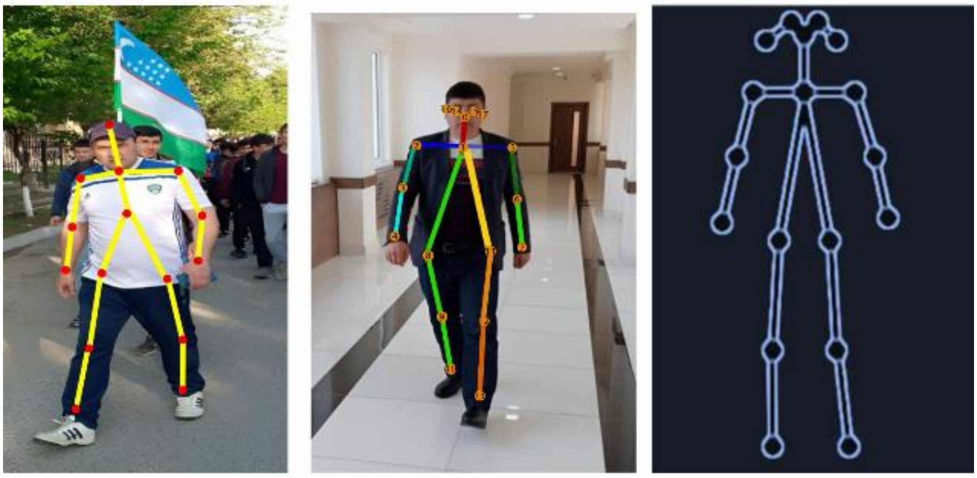

Figure 5.1. Creating a skeleton using a normalized pose

Human pose estimation is a crucial area of research within computer vision, with a wide array of practical applications across various industries. While advancements in Deep Learning and the availability of large-scale datasets have significantly enhanced the accuracy and efficiency of 2D pose estimation, 3D pose estimation still encounters difficulties. These challenges are largely attributed to the limited availability of comprehensive 3D datasets. In recent years, several approaches have been introduced to tackle this problem, and while they have achieved some level of success, there remains considerable potential for further improvement and innovation in this field [12].

The process of estimating the continuity of an image and its corresponding trace or trajectory is repeatedly performed through both top-down and bottom-up processing. This iterative approach enables the network to capture high-level features early in the process, which is then refined during deeper stages of feature estimation. Additionally, it preserves the spatial organization of elements, facilitating the precise localization of connections and enhancing the efficiency of the overall analysis. Overall, the performance analysis indicates that the hourglass neural network model can be optimized by utilizing multiple smaller filters rather than a single large filter. For instance, replacing a 6x6 filter with two 4x4 filters can enhance efficiency. Additionally, incorporating a 1x1 filter to reduce the number of pixels during convolution also contributes to better performance. As a result, this architecture proves to be more efficient when using filters of 4x4 size or smaller. Furthermore, to prevent excessive GPU memory consumption, it's advisable to input 64x64 resolution images into the network rather than high-resolution ones. It also allows for high-quality image extraction and spot-sensing extraction without affecting image processing.

VI. FORMING A POSE HEATMAP USING THE MSS-NET MODEL

The primary objective of the multi-scale control network (MSS-net) hourglass model is to generate a set of heatmaps, where each heatmap represents the likelihood or probability of the position of each key body point, such as elbows, wrists, ankles, knees, and so on. To utilize the MSS network effectively, these heatmaps are typically compared with

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 11, Issue 12, December 2024

those of the key body points, which are created using a 2D Gaussian computation technique. This comparison helps in refining the estimation of the body parts' locations, enhancing the accuracy of the model. Building on the multi-scale control network (MSS-net) and multi-scale regression network (MSR-net), which are controlled through a discontinuous structure loss, the entire network workflow is optimized using a keypoint masking learning approach. This method fine-tunes the network to better reflect the key points, as illustrated in (Figure 6.1). By incorporating this learning scheme, the network can more effectively focus on the essential features, improving overall accuracy and performance in detecting and estimating human pose.

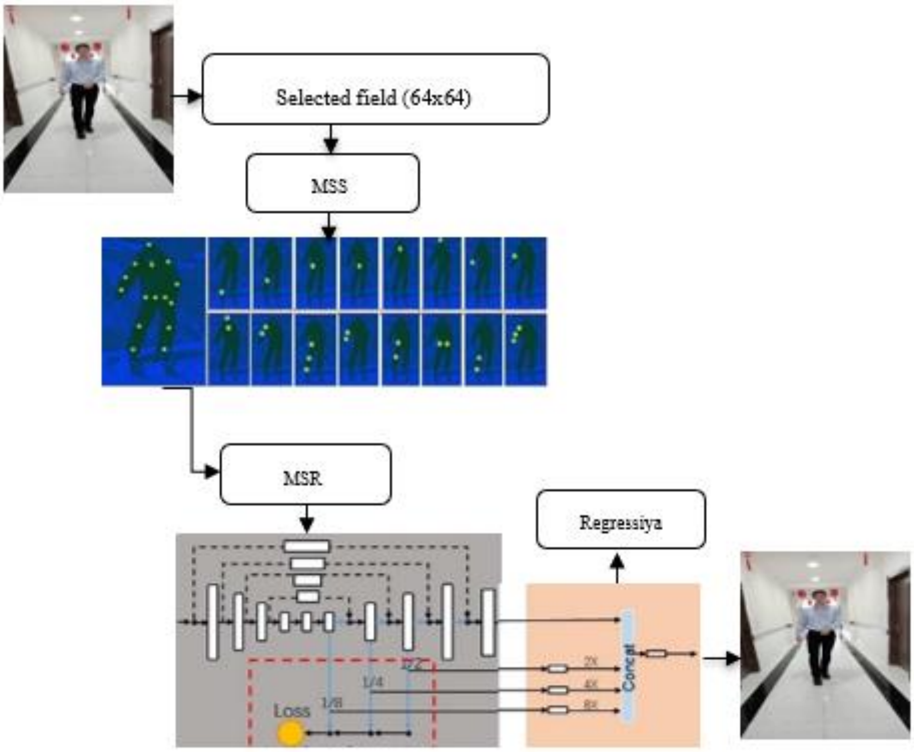

Figure 6.1. Key point masking scheme using MSS-net and MSR-net networks.

The main disadvantage of the non-neural network-based hourglass model is that the heat map of each key point is evaluated independently, so the relationship between key points is not taken into account [13]. In other words, the structural correspondence between the defined reference points is not optimized. This allows for the introduction of intermediate tracking to better define proximity and connectivity between key body points to ensure structural consistency in the pose estimation process, and to introduce structural losses between the MSS-net hourglass models it serves as targets. It is used to perform redundancy error removal from the structure and also to globally track at the final stage of the entire process flow, all heatmaps of key points across various scales in the MSR (Multi-Scale Regression) network are generated [14]. These heatmaps represent the probability distributions of the positions of key body points, such as elbows, knees, and wrists, at different resolution levels. This comprehensive output helps to accurately capture human poses by aggregating information from multiple scales, leading to a more refined and precise pose estimation. Thus, as a final result, it is possible to determine a globally consistent pose configuration.

VII. DEVELOPMENT OF A 3D POSE ESTIMATION SCHEME FOR HUMAN POSE

An additional problem when evaluating multiple individuals in a monocular image is occlusion caused by nearby people. The main challenges in estimating 3D poses of multiple people with different appearances are obstacles, as shown in Figure 7.1. In this case, there is more space for poses, more complexity in occlusion and interposition. In addition, most

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 11, Issue 12, December 2024

of the existing methods are based on two-stage schemes, which have performance issues. Although one-step methods have been proposed to solve this problem, but we can see that they have not yet achieved significant improvement, and this creates the requirement for several people to develop a 3D pose formation scheme at the same time, and this process is carried out by developing the following scheme it is possible to solve (Figure 7.1).

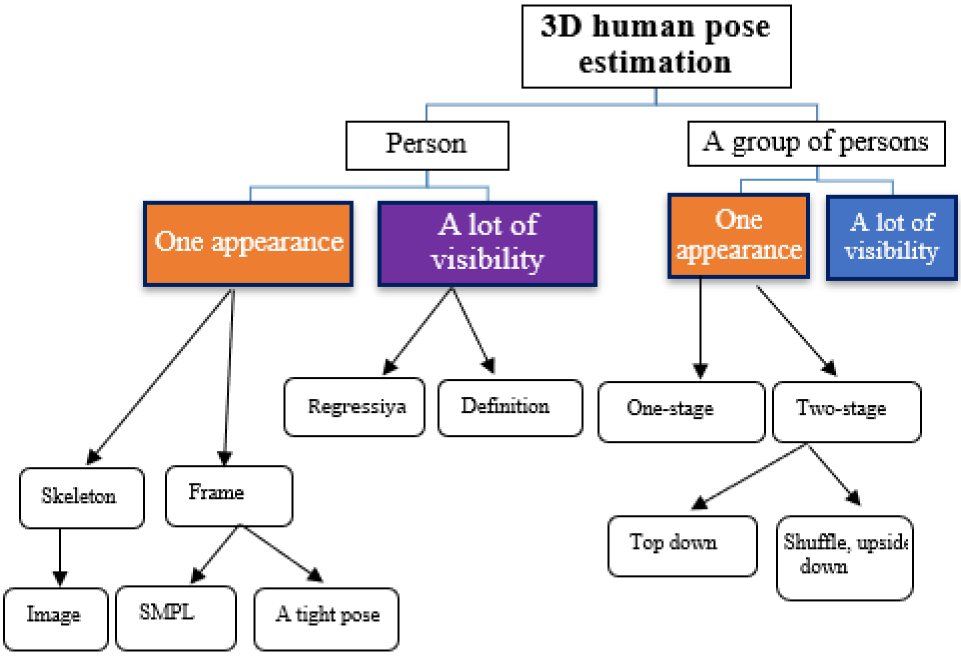

Figure 7.1. 3D pose formation scheme

As you can see from the structure created in the above figure, in the estimation of a group of persons in the 3D pose estimation of a person, in the Two-stage section of the single-view section, a CNN neural network is usednumber of images to be included in the input section layer is at least 8, and the number of outputs in the output layer is at least 2 if it is used and improved and processed based on algorithms, it will effectively solve the problem of forming a human movement pose, as well as create an opportunity to solve the problem of sensing the captured images.

VIII. RESULTS

Pose estimation results are evaluated conforming to norms PCK (Percent Correct Keypoint) metric [15], that provides information and reporting of keypoint detection percentage, enhancement metrics, and adjusted or normalized measurement of the distance from the ground level. The FLIC is calibrated based on the percentage of difference between the detected key points and the actual key point locations, following normalization to a fraction of the default PCK size. In the case of MPII, these discrepancies are adjusted and normalized to represent the percentage of head size, which is expressed as PCKh, and the resulting efficiency can be seen to improve by 0.3%. This approach ensures that the error measurements are relative to a specific reference point, providing a more accurate evaluation of the model's performance.

Table 1.
The results from FLIC with a PCK value of 0.3% are an indicator of the accuracy level of keypoint detection based on correct keypoints.

Model name	Year	Wrist part	Elbow part
CVPR	2015	93.1	92.5
CVPR	2016	97.7	95.1
ECCV	2016	99.1	97.1
A new model developed	2023	99.4	97.4

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 11, Issue 12, December 2024

The enhancement of the developed representation, taking into account the structural aspects within the MSS network and the MSR network in our approach, leads to more refined and accurate results. By considering the specific architecture and relationships between different components, our method optimizes the representation, improving the overall performance and efficiency in human pose estimation. This approach allows for better integration of multi-scale features, resulting in a more precise understanding of the human body's movement and posture. Summarizing the FLIC results in Table 1 above, we can see that our results are 99.4% for the elbow and 97.4% for the wrist. It is worth noting that the elbows and wrists are difficult to localize in the FLIC dataset.

We can see in Figure 6 some very complex examples of overcrowding and severe congestion. In this case, we estimate each individual's position relative to the boundary area specified from the MPII dataset. The approach we present is capable of extracting intricate postures for each individual without misidentifying them with the poses of others, even in situations involving occlusion.

The conclusions of the article show that with the help of neural network algorithms, it is possible to achieve effective and high accuracy in evaluating the human pose and recognizing its important features. Refinement of the regression process using the CNN and hourglass method has been proven to be particularly effective in achieving high accuracy. However, further research is needed to evaluate the accuracy and generalizability of these methods for sensing extraction of images from different datasets and video.

In addition, the article emphasizes the importance of rapid reproduction of images in the state of the picture and from different perspectives and angles, as this allows for effective assessment of the person's pose. Development of accurate and reliable recognition methods for human pose detection and future behavior assessment and person identification, rapid search and identification of people in public and crowded areas, and public is important in ensuring safety.

IX. CONCLUSION

In conclusion, in training the neural network, a multi-scale estimation model is applied to the relevant datasets using the SGD optimizer for 300 iterations. In this study, we employ multiple hourglass modules for both training and testing. The neural network training is divided into three phases: (1) MSS-Net training, (2) MSR-Net training, and (3) combined training of MSS-Net and MSR-Net, during which we apply base point masking. We adopt the same data augmentation techniques as in the original hourglass model, including the image rotation factor, i.e., a +/-30 degree rotation angle, and a scaling factor of 0.25 during the training phase. Due to limitations in CPU memory, we process input images cropped and resized to 256x256 pixels. Initially, MSS-Net is trained for 150 iterations with a starting learning rate of 5e-4. If performance does not improve after 8 iterations, the learning rate is decreased by a factor of 5. Next, we train MSR-Net with the pre-set parameters of MSS-Net for 75 iterations. Finally, the complete network is trained for 75 iterations, with cycle point masking adjusted throughout the process.

When multiple position detection is the identification and estimation of the positions of different individuals within a single image or video frame tests are executed, only those scores above the 82% threshold are selected for effective pose output image resolution synthesis. An accuracy value of 87% and above is also determined empirically from the test set and ensures a high performance rate. It should be noted that this refinement of the test might lower the efficiency of the pose estimation test, as the process also takes into account variations in the input data.

The application of advanced analysis techniques through neural networks enables fast processing of these acquired images. In the future, this will facilitate the identification of a person’s psychological traits by utilizing algorithms and neural networks to shape behavior and reveal their psychological disposition and by determining the psychology of a person, it allows the formation of biometrics of human behavior.

Although neural network-based image processing and recognition of human landmarks in images are helpful, further research is needed to evaluate the generalizability of these methods to different datasets and to extracting images from background and external factors that contribute to pose detection. It is also necessary to solve problems related to pre-processing of images and improvement of models.

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 11, Issue 12, December 2024

Neural networks for individual recognition based on human activity detection are a powerful technology with wide-ranging applications. Continual research and innovation in neural network architectures and training techniques are driving improvements in accuracy and robustness, making this field both exciting and promising for the future. Identity models play an important role in improving security, providing user-friendly authentication, and simplifying identity verification processes in various industries. They are driven by artificial intelligence and deep learning techniques to make accurate and reliable identification decisions based on various data sources.

REFERENCES

[1]. Akhatov, A.R. & Djumanov, O.I. 2006. Mechanisms of images visualization in the system of micro-objects recognition and classification. In proceedings of the Fourth World Conference on Intelligent Systems for Industrial Automation, 21-22 November, 2006. 206-210. Tashkent, Uzbekistan.
[2]. Akhatov, A. Nazarov, F. Rashidov, A. 2021. Increasing data reliability by using bigdata parallelization mechanisms. International conference on information science and com-munications technologies. 4,5,6 November. ICISCT 2021(IEEE), art. no. 9670387 DOI: 10.1109/ICISCT52966.2021.9670387.
[3]. Akhatov, A. Nazarov, F. Rashidov, A.2022. Mechanisms of Information Reliability In big data and Blockchain Technologies. International conference on information science and communications technologies. 4,5,6 November. ICISCT 2021(IEEE), art. no. 9670052 DOI: 10.1109/ ICISCT 52966.2021.9670052.
[4]. Andriluka, M. Pishchulin, L. Gehler, P. Schiele, B. 2014. 2D human pose estimation: New benchmark and state of the art analysis. In: CVPR. 3686–3693.
[5]. Axatov, A.R. & Ximmatov, I.Q. 2020. Foydalanuvchilarni biometrik autentifikatsiya turlari asosida haqiqiyligini tasdiqlash usullarinning samaradorligi. Innovatsion yondashuvlar ilm-fan taraqqiyoti kaliti sifatida: yechimlar va istiqbollar, 8-10 oktyabr 2020 y. 20-26. Jizzax: Uzbekistan.
[6]. Bourdev, L. & Malik, J. 2009. Poselets: Body part detectors trained using 3D human pose annotations. In: ICCV. 1365–1372.
[7]. Bulat, A. Tzimiropoulos, G. 2016. Human pose estimation via convolutional part heatmap regression. ECCV 717–732.
[8]. Chang, M. Qi, H. Wang, X. Cheng, H. Lyu, S. 2015. Fast online upper body pose estima-tion from video. In: BMVC. 104.1–104.12. Swansea, England.
[9]. Cherian, A. Mairal, J. Alahari, K. Schmid, C. 2014. Mixing body-part sequences for hu-man pose estimation. In: CVPR. pp.2361–2368.
[10]. Chu, X. Yang, W. Ouyang, W. Ma, C. Yuille, A.L. Wang, X. 2017. Multi-context attention for human pose estimation. In: 2017 IEEE Conference on Computer Vision and Pat-tern Recognition (CVPR). pp.5669–5678.
[11]. Liu, Z. Zhu, J. Bu, J. Chen, C. 2015. A survey of human pose estimation. Journal of Visu-al Communication and Image Representation 32, pp. 10–19.
[12]. Newell, A. Yang, K. & Deng, J. 2016. Stacked Hourglass Networks for Human Pose Esti-mation// 26 Jul 2016. pp.3-4.
[13]. Sapp, B. & Taskar, B. 2013. Multimodal decomposable models for human pose estima-tion. In: CVPR. pp. 3674–3681.
[14]. Ximmatov, I.Q. 2020. Advantages of biometric gait recognition. Important factors in evaluation of gait analysis systems. Scientific journal of SamSu. ISSN 2091-5446, 2020, vol-3 (121), pp. 104-107. Samarkand: Uzbekistan.
[15]. Ximmatov, I.Q. 2020. Important factors in evaluation of gait analysis systems and ad-vantages of biometric gait recognition. Innovatsion va zamonaviy axborot texnologiya-larini ta'lim, fan va boshqaruv sohalarida qo'llash istiqbollari. 14-15 may, 2020 y. pp. 262-267. Samarkand: Uzbekistan.