HUMAN IDENTIFIER WITH MANNERISM USING DEEP LEARNING

Saathvik Krishnan[1], Shrijeeth S[2]

[1]Saathvik Krishnan, Dept. of computer science and engineering, Rajalakshmi Engineering College, Tamil Nadu, India

[2]Shrijeeth S, Dept. of computer science and engineering, Rajalakshmi Engineering College, Tamil Nadu, India

---***---

Abstract - *A human posture and tracking system in which we utilize a deep learning model to fit in data from humans and then train the model to recognize a certain personality utilizing media pipelines. It may be used as a security system in homes and businesses to prevent unwanted entrance. Pose Detection is a framework for creating multimodal audio, video, or any type of time series data. The Pose detection framework may be used to create an outstanding ML pipeline for inference models like TensorFlow and TFLite, as well as media processing routines. Pose estimation from the video is critical for augmented reality applications such as overlaying digital content and data on top of the physical world, signing recognition, full-body gesture management, and even quantifying physical exercises, which will form the basis for yoga, dance, and fitness applications. A large number of possible positions, variable degrees of mobility, occlusions and a range of looks or clothes make creating an assessment for fitness applications, particularly tough. In contrast to the Face Mesh and Pose detection Hand following pipelines, where we calculate the ROI from predicted key points, we expressly forecast two extra virtual key points for the human cause following that clearly depict the actual body rotation and measurement as the circle. So this application can be used for Home security applications and if built on a larger scale, it can be used in companies for high-profile security with extra facial recognition making it a more secure system.*

Key Words: **Media Pipelines, TensorFlow, Mobility, Occlusions, Pose detection, ROI, TFLite**

1.INTRODUCTION

People detection has long been a focal point of debate for many purposes in classical object detection. Using stance detection and pose tracking, computers can now read human body language thanks to recent advancements in machine learning algorithms. These detections' accuracy and hardware needs have now improved to the point where they are economically practical. High-performing real-time pose identification and tracking will deliver some of the most significant developments in computer vision, has a profound impact on the technology's progress. Computer vision tasks like human position estimate and tracking include finding, connecting, and following semantic key points. "Right shoulders," "left knees," or "vehicle's left brake lights" are a few examples of semantic keypoints. The accuracy of posture estimate has been limited by the large processing resources needed to execute semantic keypoint tracking in live video data.

Human Posture Tracking and Estimates is one of the major research areas of computer vision and deep learning. The reason for its importance is the abundance of applications of this type of technology in our daily lives. Human pose skeletons usually represent the orientation of a person graphically. That is, it can be used to identify the personality of Mannerism. In essence, a set of points coordinated can be used to show a person by linking those points. Here every part in the physical structure is called a hinge, or valid point. A valid connection between two parts is called a pair, but not all combinations of parts are valid pairs. Various approaches to tracking and estimating human posture have been introduced over the years. These methods usually first identify each part and then form a connection between them to create that particular pose.

2. LITERATURE REVIEW

L. Sigal et al, They have displayed data that was gathered using a hardware setup that can record synchronised video and 3D motion that is accurate. The resulting HUMANEVA databases have several participants carrying out a series of predetermined behaviours repeatedly. Approximately 40,000 frames of synchronised motion capture and multi-view video (producing over a quarter million total picture frames) and an additional 37,000 time instants of pure motion capture data were gathered at 60 Hz. In order to evaluate algorithms for 2D and 3D posture estimation and tracking, a common set of error metrics is provided. Additionally, they have discussed a baseline approach for 3D articulated tracking that employs a comparatively conventional Bayesian framework with optimization through sequential importance resampling and annealed particle filtering.They investigated various likelihood functions, previous models of human motion, and the implications of algorithm settings in the context of this fundamental algorithm. Their research indicates that image observation models and motion priors play significant roles in performance, and that Bayesian filtering often performs well in a multi-view laboratory setting where initialization is possible. The scientific community has access to the datasets and the software. This infrastructure will aid in the creation of fresh articulated motion and pose estimation algorithms, serve as a benchmark for assessing and contrasting novel approaches, and contribute to the advancement of the state-of-the-art in human pose estimation and tracking[1].

D. Mehta et al, They have put forth a CNN-based method for estimating a 3D human body's posture from a single RGB photograph that overcomes the problem of models' poor generalizability when trained just on the incredibly little amount of publically accessible 3D pose data. Through the transfer of learnt features and generalisation to in-the-wild scenarios, they have demonstrated cutting-edge performance on known benchmarks using just the current 3D pose data and 2D pose data. Additionally, a novel training set for estimating human body posture from monocular photos of actual people was provided, with the ground truth being obtained via a multi-camera marker-less motion capture system. It enhances the diversity of poses, human look, clothes, occlusion, and views in the current corpora and allows for a wider range of augmentation.

Their 3D posture dataset exhibits superior in-the-wild performance than current annotated data, and is further enhanced in conjunction with transfer learning from 2D pose data. They also supplied a new benchmark that includes both indoor and outdoor situations. In essence, they proposed that transfer learning of representations is essential for universal 3D body position estimation, together with algorithmic and data contributions[2].

L. Bourdev et al, Poselets, a novel kind of component presented by them, are designed to form dense clusters both in the configuration space of keypoints and in the appearance space of picture patches. In this research, they created a novel poselet-based method for people detection. They simply employed 2D annotations, which are significantly simpler for novice human annotators, as opposed to that work, which used 3D annotations of keypoints. A two-layer feed-forward network that has its weights set using the maximum margin method has been trained. Then, based on experimentally obtained spatial keypoint distributions, the revised poselet activations are grouped into mutually coherent hypotheses. In order to offer a segmentation, shape masks are matched to picture edges and bounding boxes are predicted for each human hypothesis.With an average precision of 47.8 percent and 40.5 percent, respectively, on PASCAL VOC 2009, the resultant system is now the best performer on the job of persons detection and segmentation, to the best of their knowledge[4].

L. Pishchulin et al. they have compared human activity recognition methods that are holistic and pose-based on a large scale using the "MPII Human Pose" dataset. Additionally, they have examined the variables that affect the success and failure of holistic and pose-based approaches.

The data in this dataset was systematically gathered from YouTube videos using a taxonomy of common human activities that includes 410 activities. Around 25K photos, 40K annotated postures, and rich annotations on the test set are included. Over 1 million frames, 3D torso and head position, body component occlusions, and a video sample for each picture [5].

Z. Cao et al, They have provided a method for accurately identifying the 2D poses of several persons in a photograph. The method learns to link body parts with people in the image using nonparametric representations known as Part Affinity Fields (PAFs). No matter how many individuals are in the image, the architecture stores global context to enable a greedy bottom-up parsing phase that achieves real-time speed while maintaining excellent accuracy. Two branches of the same sequential prediction method are used in the architecture to jointly learn the locations of the parts and their associations. Our approach won the first COCO 2016 keypoints challenge and, in terms of performance and efficiency, outperforms the previous state-of-the-art result on the MPII Multi-Person benchmark[6].

Mykhaylo Andriluka et al. they has collected a comprehensive dataset with established classifications for more than 800 human activities. The images collected cover a wider range of human activities than previous datasets, including various leisure, professional, and home activities, and capture people from a wider range of angles. For each frame, we provided adjacent video frames to facilitate the use of motion information[7].

Leonid Pishchulin et al.they have discussed about the task of clearly estimating the human posture of multiple people in a real image. They proposed the division and labeling formulation of a set of body part hypotheses generated by a CNN-based parts detector. Its formulation, which is an instance of integer linear programming, implicitly performs non-maximum suppression of a set of sub-candidates and groups them to form a body part composition that respects geometric and appearance constraints increase[8].

N. Dalal et al, They investigate the issue of feature sets for reliable visual object recognition using a test case of linear SVM-based person detection. They examined current edge- and gradient-based descriptors and demonstrated experimentally that grids of HOG descriptors greatly exceed existing feature sets for human detection. They looked at how each stage of the computation affected performance and came to the conclusion that fine-scale gradients, fine orientation binning, somewhat coarse spatial binning, and excellent local contrast normalisation in overlapping descriptor blocks are all crucial for successful outcomes. The original MIT pedestrian database can now be separated almost perfectly using the new method, thus we present a more difficult dataset with over 1800 annotated human photos with a wide variety of poses and backdrops[9].

Alexander Toshev et al.they have suggested a method for estimating human poses based on Deep Neural Networks (DNN). Pose estimation is formulated as a DNN-based regression problem for body joints. They have presented a cascade of such DNN regressors, resulting in highly accurate pose estimates[15].

J. Tompson et al, they have proposed a unique architecture that consists of a productive "position refinement" model that is trained to guess the joint offset location within a circumscribed area of the picture. To increase accuracy in human joint position prediction, this refinement model is concurrently trained in cascade with a cutting-edge ConvNet model. We demonstrate that our detector beats all other methods on the MPII-human-pose dataset and that its variance approaches the variance of human annotations on the FLIC dataset[18].

Matthias Dantone et al.they used a two-tiered random forest as a common regressor. The first layer acts as an identifiable and independent classifier for body parts. The second layer can predict the location of joints by taking into account the estimated class distribution of the first layer and modeling the interdependence and general occurrence of parts. This leads to a frame of pose estimation that already considers the dependencies between body parts for general localization[24].

Yi Yang et al.the orientation is recorded by combining the templates of each part. They describe a general and flexible mixed model for capturing context-related co-occurrence relationships between parts and complement the standard spring model for encoding spatial relationships. In this paper they show that such a relationship can capture the idea of local rigidity. So they present experimental results on a standard pose estimation benchmark that suggests that their approach is the latest system for pose estimation[25].

Fangting Xia et al.they first train two fully convolutional neural networks (FCNs), the pose FCN and the part FCN, to provide initial estimates of the pose joint potential and the semantic partial potential. Then, the two types are fused with a fully connected conditional random field (FCRF) to use the new smoothness term for segment joints to create semantic and spatial consistency between parts and joints. To refine the part segment, the refined pose and the original part potential are integrated by the part FCN, and the pose skeleton feature serves as an additional regularization hint for the part segment. Finally, to reduce the complexity of FCRF, they guide human detection boxes, derive graphs for each box, and speed inference by 40 times[26].

Cheol-hwan Yoo et al. It has been demonstrated that converting an original picture into a high-dimensional (HD) feature is efficient for cacategorizingmages. In order to increase the face recognition system's capacity for discrimination, this research introduces a unique feature extraction technique that makes use of the HD feature space. The local binary pattern may be broken down into bit-planes, each of which contains scale-specific directional information of the facial picture, they have found. Each bit-plane has an illumination-robust property in addition to the intrinsic local structure of the face picture. They created an HD feature vector with enhanced discrimination by concatenating all of the deconstructed bit-planes.

The HD feature vector is subjected to orthogonal linear discriminant analysis, a supervised dimension reduction technique, in order to miniminimize computational complexity while maintaining the integrated local structural information. Numerous experimental findings demonstrate that under different lighting, stance, and expression changes, current classifiers with the suggested feature outperform those with other traditional features [30].

3. HUMAN BODY MODELLING

Today, there are several models for performing pose estimation. Here are some methods used to estimate the pose:

1. Open pose
2. Pose net
3. Blaze pose
4. Deep Pose
5. Dense pose
6. Deep cut

Model-based approaches are typically used to describe and derive poses for the human body and render 2D or 3D poses. Most methods use the rigid kinematic joints model. In this model, the human body is represented as a unit with joints and limbs that contains information about the structure and shape of the kinematic body. There are three types of models for modeling the human body:

3.1 TYPES OF HUMAN BODY MODELS

- Both 2D and 3D posture estimates are performed using a kinematic model, often known as a skeleton model. This flexible and intuitive human body model includes various joint positions and limb orientations to represent the anatomy of the human body. Therefore, skeletal posture assessment models are used to capture the relationships between different parts of the body. However, kinematic models have limitations in expressing texture or shape information.This can be seen in Fig.3.1.

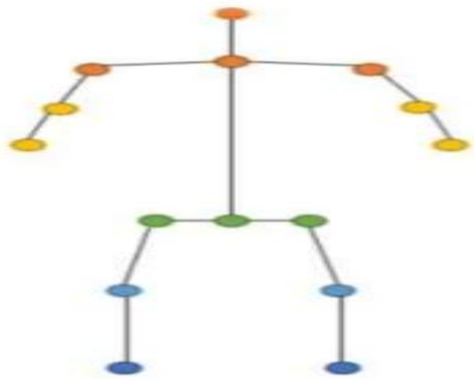

Fig.3.1. Kinematic Model

- A planar or contour-based model used to estimate a two-dimensional pose. A flat model is used to represent the shape and shape of the human body. In general, body parts are represented by many rectangles close to the contours of the human body. A typical example is the active shape model (ASM), which is used to obtain complete graphs of the human body and silhouette deformation using principal component analysis.

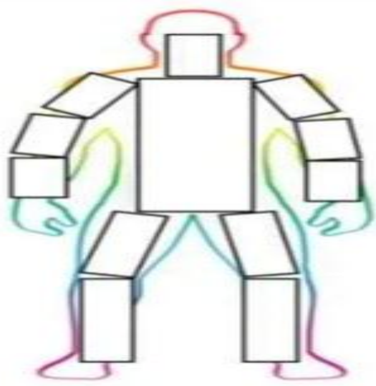

Fig.3.2. Planar or Contour based Model

- A volumetric model is utilised to estimate 3D posture. There are a variety of common 3D human body models that may be used to estimate 3D human position using deep learning. For example, GHUM& GHUML(ite) are fully trainable end-to-end deep learning pipelines that simulate statistical and articulated 3D human body form and position using a high-resolution dataset of full-body scans of over 60'000 human configurations.

Fig.*3.3.Volumetric Model*

3.2 CHALLENGES FACED DURING MODELLING

The body's look joins alter dynamically owing to various types of clothing, arbitrary occlusion, occlusions due to viewing angle, and backdrop contexts, making human position estimation a difficult operation. Pose estimation must be resistant to difficult real-world variables like illumination and weather. As a result, image processing algorithms have a hard time identifying fine-grained joint coordinates. It's particularly tough to keep track of minor, scarcely visible joints.

4. VIDEO PERSON POSE TRACKING

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

4.1 HOW DOES POSE ESTIMATION WORK

Position estimation uses a person's or object's pose and orientation to estimate and monitor their location. Pose estimation, on the other hand, allows programs to estimate the spatial locations ("poses") of a body in an image or video. Most posture estimators are two-step frameworks that first identify human bounding boxes before estimating the pose within each box. 17 distinct key points may be detected using human pose estimation (classes). Each key point has three integers (x, y, v), with x and y indicating the coordinates and indicating whether or not the key point is visible.

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 08 | Aug 2022 www.irjet.net p-ISSN: 2395-0072

4.2 POSE ESTIMATION WITH DEEP LEARNING

In posture estimation tasks, deep learning approaches (image segmentation or object identification) have resulted in considerable advancements and performance benefits.

Model for detecting people/poses (BlazePose Detector)

The detector we utilized is based on our own lightweight Blaze Face model, which is used as a surrogate for a human detector in Face Detection. It predicts two more virtual key points that accurately characterize the human body's center, rotation, and size as a circle. The radius of a circle circumscribing the entire body, as well as the inclination angle of the line joining the shoulder and hip midpoints, are all predicted.

Landmark Model Pose (BlazePose GHUM 3D)

In Posture detection, we employed a landmark model to forecast the placement of 33 pose landmarks.

5. PROPOSED WORK

A deep learning method is developed in this research to recognise a person based on his or her walking style or mannerism. To begin, the BlazePose Model is utilised to extract 33 skeletal key points from a video stream of 5 seconds. The essential elements are then entered into the model algorithm, which predicts the likelihood of the individual walking. For a testing purpose, data of 3 people are collected and trained by Deep Learning Model.

5.1 Data Collection

Datasets for various people are acquired by taking 30 5-6 seconds recordings of each individual, which are then put through the BlazePose Detector to obtain the 44 skeleton essential points. The facial region's key points are eliminated, and the remaining key points are put to a CSV file with the class label Person's unique identification. This information is then loaded into a Deep Learning Model, which classifies each individual.

5.2 Data Preprocessing

Following data collection, 90 frames (90 rows) of the same class label are chosen and concatenated. During inference, the same technique is followed. Then the dataset is split into classes and labels during training period. Now this data is ready to be fed into the Deep Learning Model. The data is stored in CSV file as mentioned in Figure.5.1.

left_hip_y	right_hip_	right_hip_	left_knee	left_knee	right_kne	right_kne	left_ankle	left_ankle	right_ankl	right_ankl	left_heel_	left_heel_	right_hee	right_hee	left_foot_	left_foot_	right_foot	right_foot	label
0.453551	0.494962	0.444653	0.526164	0.490234	0.525621	0.478821	0.519397	0.557448	0.521103	0.550991	0.515923	0.571441	0.518633	0.561613	0.530337	0.569109	0.530585	0.569329	Person-1
0.412099	0.519563	0.407466	0.543836	0.481151	0.533318	0.464418	0.530957	0.548769	0.520557	0.546346	0.528248	0.569725	0.518611	0.557013	0.533866	0.583162	0.527755	0.569513	Person-1
0.396093	0.523957	0.390333	0.543096	0.475937	0.520832	0.466808	0.535401	0.553601	0.515906	0.549659	0.533977	0.571169	0.517354	0.559996	0.534851	0.591849	0.505449	0.588138	Person-1
0.397378	0.522109	0.382876	0.544901	0.477607	0.51926	0.460306	0.537881	0.553943	0.510699	0.54621	0.537087	0.56976	0.511997	0.554028	0.534621	0.593014	0.502276	0.595613	Person-1
0.383984	0.519505	0.385401	0.542732	0.478571	0.520979	0.462804	0.536996	0.561711	0.510968	0.560987	0.535714	0.573368	0.520395	0.564727	0.534757	0.593738	0.509556	0.591905	Person-1
0.392657	0.518249	0.394151	0.54245	0.485152	0.523258	0.463784	0.536106	0.570438	0.524022	0.564951	0.534834	0.581663	0.524905	0.575886	0.534587	0.596639	0.520456	0.59501	Person-1
0.416686	0.517144	0.407212	0.542387	0.496044	0.523291	0.482738	0.536731	0.575683	0.52748	0.57573	0.535456	0.586478	0.526588	0.588632	0.536764	0.59931	0.525963	0.598311	Person-1
0.432684	0.519523	0.430716	0.541696	0.512152	0.523356	0.503348	0.5364	0.579631	0.528881	0.583905	0.535761	0.587701	0.528746	0.59372	0.536781	0.600819	0.527117	0.603924	Person-1
0.441287	0.522379	0.439344	0.540304	0.517635	0.525391	0.51373	0.536638	0.582745	0.52929	0.58611	0.536025	0.590618	0.529322	0.595988	0.53852	0.602426	0.527067	0.60802	Person-1
0.442008	0.524425	0.441609	0.540216	0.518273	0.527186	0.515931	0.53624	0.584275	0.528992	0.58646	0.535272	0.592216	0.528699	0.595573	0.540374	0.603947	0.527908	0.607613	Person-1
0.444094	0.526078	0.444405	0.541397	0.519753	0.52849	0.518878	0.536252	0.583974	0.530297	0.590402	0.535036	0.591224	0.530096	0.600964	0.54075	0.603787	0.528189	0.610604	Person-1
0.443498	0.525997	0.444268	0.541171	0.519779	0.529388	0.520402	0.536209	0.584485	0.53145	0.593077	0.534827	0.590621	0.531262	0.603631	0.540799	0.604251	0.528921	0.615416	Person-1
0.446099	0.526214	0.447607	0.541617	0.521656	0.529131	0.521907	0.536225	0.583902	0.531698	0.594999	0.534836	0.590029	0.531765	0.606539	0.540384	0.604852	0.528586	0.618658	Person-1
0.445611	0.526121	0.44684	0.541481	0.526084	0.528859	0.524346	0.536472	0.583528	0.531662	0.598457	0.534982	0.589758	0.531611	0.609357	0.540031	0.60519	0.528825	0.623554	Person-1
0.445302	0.525843	0.446587	0.543486	0.527287	0.52869	0.526844	0.537195	0.582885	0.531824	0.600368	0.535497	0.586248	0.531912	0.611265	0.53995	0.605148	0.528781	0.6258	Person-1
0.44451	0.525828	0.446012	0.545053	0.526077	0.528579	0.528412	0.538068	0.578036	0.531623	0.601107	0.536489	0.577125	0.532187	0.611356	0.539803	0.60464	0.528571	0.626041	Person-1
0.444157	0.525446	0.44548	0.545774	0.525058	0.528484	0.528156	0.53876	0.577194	0.531373	0.602206	0.537003	0.576783	0.532084	0.612366	0.539853	0.604317	0.528483	0.626176	Person-1
0.44358	0.525185	0.444606	0.545752	0.523899	0.528405	0.526374	0.539152	0.577166	0.531304	0.601257	0.53765	0.578033	0.53213	0.611625	0.54055	0.604274	0.528392	0.626539	Person-1
0.443426	0.525022	0.444775	0.545203	0.525462	0.528061	0.528769	0.53915	0.57803	0.53122	0.601879	0.537896	0.579879	0.532211	0.611515	0.540527	0.606105	0.528319	0.627336	Person-1

Fig. 5.1. Sample Data for Person 1

5.3 Model Architecture

After Data Preprocessing, the data is fed into the Deep Learning Model for training. This model was trained using the given data for 100 epochs using Tensorflow Keras API. Categorical Cross Entropy (Softmax Loss) loss function is used with an Adam Optimizer with a learning rate of 0.0001. This model produced an accuracy of 96.88% in the Training Set and 94.12% in Test Set. A Deep Learning Model is created as shown in Figure.5.2.

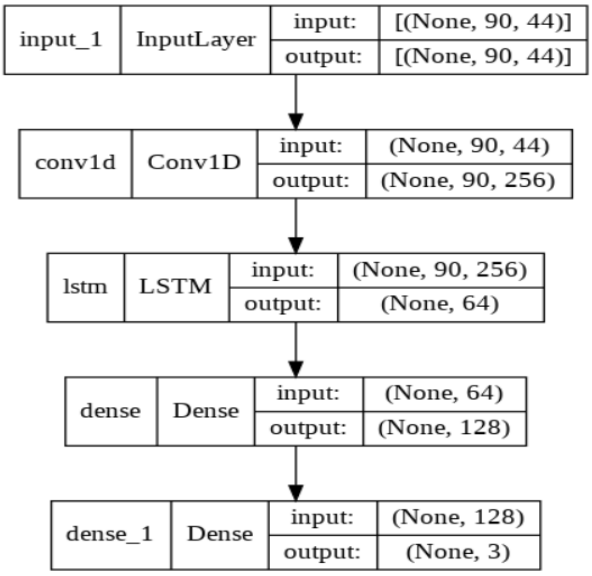

Fig 5.2. Deep Learning Model Architecture

5.4 Overall Working

First the key points for 90 frames from video streams of 3 persons walking is taken by BlazePose Model. Using these key points, a Deep Learning Model is trained. After training, model is deployed as an application. The video stream is read from the camera continuously and the application 90 frames from this video stream and extracts the key points from those 90 frames. These key points are the preprocessed and are concatenated together as a single array. This array is then fed into the Deep Learning Model. The Deep Learning Model predicts the probability of Person walking.

6. RESULTS

There are lots of methods to identify a human or a person. Some of them are fingerprint, face recognition, voice recognition, etc. These methods are called as biometrics. Biometrics refers to the identification of humans by their characteristics or traits. "Footstep" a new biometric has been proposed by a group of researchers in 1997. In their research, they stated that the footstep of a person has a unique property and it is different for a different person. So it could be an identity to detect a person. The main benefit of footsteps over the known biometrics is that this is a behavioral fact of humans and it is difficult to replicate by others.

So using this footstep biometrics we can differentiate the walking pattern form from person to person. Moreover, we will be able to find or identify a human or a person by their walking pattern. There are various methodologies proposed and implemented to recognize persons based on their footsteps. One of the methodologies uses sensors. There are different kinds of sensors to get information about footsteps like pressure sensors. That could be placed in the floor or could be in the person's foot ware. But implementing this is costly and requires a lot of preprocessing, power and energy. Another pattern can also be the walking sound of a person. A study shows that it is possible to identify a person accurately just by hearing their walking sound. And the accuracy rate is about 66%. This technique requires high-quality noise canceling microphones to be deployed in the Detection System. In this paper, we propose a most efficient and cost effective way to detect person based on their walking style. We have collected video samples for each person (30 videos each 5 seconds). Body Key points of each person's is being extracted from each frame of the video using pretrained BlazePose Model. First, the key points for 90 frames from video streams of 3 persons walking is taken by BlazePose Model. Using these key points, a Deep Learning Model is trained. After training, the model is deployed as an application. The video stream is read from the camera continuously and the application 90 frames from this video stream and extracts the key points from those 90 frames. These key points are preprocessed and are concatenated together as a single array. This array is then fed into the Deep Learning Model. The Deep Learning Model predicts the probability of a Person walking. This methodology produces accuracy on test set upto 94%. This model uses LSTM to learn walking manner across 90 frames and Conv1D is used to identify features in each frame.

The dataset was collected and is divided into 3 parts or splits (Training, Validation, and Test Set). Softmax Loss and Categorical Accuracy are used as metrics to check the performance of the model. The Training split had a Softmax Loss of 0.1873 with an accuracy of 96.88%. Validation Split had a Softmax Loss of 0.2521 with an accuracy of 99.98%. Test Split had a Softmax Loss of 0.2368 with an accuracy of 94.12%. Deep Learning Model Softmax loss and accuracy for different Data Splits is given in Table 1.

Split	Loss	Accuracy
Training	0.1873	96.88%
Validation	0.2521	99.98%
Test	0.2368	94.12%

Table. 6.1 Loss and Accuracy for various splits

7. CONCLUSIONS

To summarize, we have developed a Deep Learning Model for detecting a person from his/her walk mannerism. This model predicts the probability of the person walking from 3 persons the model was trained on. By experimenting, on various Hyper Parameters, we got Training Accuracy of 96.88% and Test Accuracy of 94.12% for 3 persons. Therefore, in this paper we have proposed a new method for detecting people based on their walking style. There can be more research done on this paper, for example, a more advanced and complicated Deep Learning Model can be trained to obtain a even more higher accuracy. Research can also be done to improve this methodology so that it can be applied to more persons which can then be used in various sectors as a security measure.

REFERENCES

[1] L. Sigal, A. O. Balan, M. J. Black, Humaneva: Synchronized videoand motion capture dataset and baseline algorithm for evaluation of ar-ticulated human motion, International Journal of Computer Vision 87(2010) 4.

[2] D. Mehta, H. Rhodin, D. Casas, P. Fua, O. Sotnychenko, W. Xu,C. Theobalt, Monocular 3d human pose estimation in the wild usingimproved CNN supervision, in: International Conference on 3D Vision,2017, pp. 506–516.

[3] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, T. Serre, HMDB: a largevideo database for human motion recognition, in: IEEE InternationalConference on Computer Vision, 2011.

[4] L. D. Bourdev, J. Malik, Poselets: Body part detectors trained using 3dhuman pose annotations, in: IEEE International Conference on Com-puter Vision, 2009, pp. 1365–1372.

[5] L. Pishchulin, E. Insafutdinov, S. Tang, B. Andres, M. Andriluka, P. V.Gehler, B. Schiele, Deepcut: Joint subset partition and labeling for multiperson pose estimation, in: IEEE Conference on Computer Vision andPattern Recognition, 2016, pp. 4929–4937.

[6] Z. Cao, T. Simon, S. Wei, Y. Sheikh, Realtime multi-person 2d poseestimation using part affinity fields, in: IEEE

Conference on ComputerVision and Pattern Recognition, 2017, pp. 1302–1310.

[7] M. Andriluka, U. Iqbal, E. Insafutdinov, L. Pishchulin, A. Milan, J. Gall,B. Schiele, Posetrack: A benchmark for human pose estimation andtracking, in: IEEE Conference on Computer Vision and Pattern Recog-nition, 2018, pp. 5167–5176.

[8] L. Pishchulin, M. Andriluka, B. Schiele, Fine-grained activity recogni-tion with holistic and pose based features, in: German Conference onPattern Recognition, Springer, 2014, pp. 678–689.

[9] N. Dalal, B. Triggs, Histograms of oriented gradients for human detec-tion, in: IEEE Conference on Computer Vision and Pattern Recognition,volume 1, IEEE, 2005, pp. 886–893.

[10] L. Bourdev, S. Maji, T. Brox, J. Malik, Detecting people using mutuallyconsistent poselet activations, in: European Conference on ComputerVision, Springer, 2010, pp. 168–181.

[11] Y. Chen, Z. Wang, Y. Peng, Z. Zhang, G. Yu, J. Sun, Cascaded pyra-mid network for multi-person pose estimation, in: IEEE Conference onComputer Vision and Pattern Recognition, 2018, pp. 7103–7112.

[12] B. Xiao, H. Wu, Y. Wei, Simple baselines for human pose estimationand tracking, in: European Conference on Computer Vision, 2018, pp.472–487.

[13] K. Sun, B. Xiao, D. Liu, J. Wang, Deep high-resolution representationlearning for human pose estimation, in: IEEE Conference on ComputerVision and Pattern Recognition, 2019, pp. 5693–5703.

[14] A. Jain, J. Tompson, M. Andriluka, G. W. Taylor, C. Bregler, Learninghuman pose estimation features with convolutional networks, in: Inter-national Conference on Learning Representations, 2014.

[15] A. Toshev, C. Szegedy, Deeppose: Human pose estimation via deepneural networks, in: IEEE Conference on Computer Vision and PatternRecognition, 2014, pp. 1653–1660.

[16] J. Tompson, A. Jain, Y. LeCun, C. Bregler, Joint training of a convo-lutional network and a graphical model for human pose estimation, in:Advances in Neural Information Processing Systems, 2014, pp. 1799–1807.

[17] A. Jain, J. Tompson, Y. LeCun, C. Bregler, Modeep: A deep learningframework using motion features for human pose estimation, in: AsianConference on Computer Vision, 2014, pp. 302–315.

[18] J. Tompson, R. Goroshin, A. Jain, Y. LeCun, C. Bregler, Efficient ob-ject localization using convolutional networks, in: IEEE Conference onComputer Vision and Pattern Recognition, 2015, pp. 648–656.

[19] S. Wei, V. Ramakrishna, T. Kanade, Y. Sheikh, Convolutional pose ma-chines, in: IEEE Conference on Computer Vision and Pattern Recogni-tion, 2016, pp. 4724–4732.

[20]J. Katona, "A review of human–computer interaction and virtual reality research fields in cognitive InfoCommunications," Applied Sciences, vol. 11, no. 6, p. 2646, (2021)

[21]M. Andriluka, L. Pishchulin, P. Gehler and B. Schiele, "2D Human Pose Estimation: New Benchmark and State of the Art Analysis," 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 3686-3693, doi: 10.1109/CVPR.2014.471.(2014)

[22]D. L. Quam, "Gesture recognition with a DataGlove," IEEE Conference on Aerospace and Electronics, vol. 2, pp. 755–760, (1990)

[23]https://arxiv.org/abs/1911.07424

[24]M. Dantone, J. Gall, C. Leistner and L. Van Gool, "Human Pose Estimation Using Body Parts Dependent Joint Regressors," 2013 IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3041-3048, doi: 10.1109/CVPR.2013.391.(2013)

[25]Y. Yang and D. Ramanan, "Articulated pose estimation with flexible mixtures-of-parts," CVPR 2011, 2011, pp. 1385-1392, doi: 10.1109/CVPR.2011.5995741.(2011)

[26]https://arxiv.org/abs/1708.03383

[27]K. Khan, M. Mauro, P. Migliorati and R. Leonardi, "Head pose estimation through multi-class face segmentation," 2017 IEEE International Conference on Multimedia and Expo (ICME), 2017, pp. 175-180, doi: 10.1109/ICME.2017.8019521.(2017)

[28]Y. Ibuki and T. Tasaki, "Improvement of Segmentation Accuracy for Pose Estimation by Selecting an Instance," 2021 IEEE/SICE International Symposium on System Integration (SII), 2021, pp. 823-824, doi: 10.1109/IEEECONF49454.2021.9382764.(2021)

[29] Dr. Bapurao Bandgar, 2021, Implementation of Image Processing Tools for Real-Time Applications, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 10, Issue 07 (July 2021),

[30] Cheol-Hwan Yoo, Seung-Wook Kim, June-Young Jung, Sung-Jea Ko,High-dimensional feature extraction using bit-plane decomposition of local binary patterns for robust face recognition,Journal of Visual Communication and Image Representation,Volume 45,2017,