Person Identification using Deep Learning

Nand Kumar Bansode, Vikas Manhas, Reshav Kumar, Saurabh Shubham, Varun Nayal

Computer Engineering, AIT, Pune.

***Abstract*—**In the present scenario, digital data generation, data consumption becoming necessary due to advancement in technology. The business process are taking advantage of the available data. The human data processing becoming important in various types of applications like person authentication, verifications automatically by the machines. One of the application is to identify the person automatically by the machine.

Face recognition technology is available for use for couple of years. The face recognition technology is limited by the use of the restricted environment. In this paper, the method for person identification in unrestricted environment is presented using deep neural network. The face recognition and body part recognition these two important steps are used to identify the person.

***Keywords*--Face recognition, deep learning, Person Re-identification.**

1. INTRODUCTION

Identification of the individual person using various technologies becoming important due to the use of person identification in various applications like verification as airport, different unities, digital transactions, access to the restricted area or information.

The person identification problem has been studied for several years, but the human like performance for person recognition by the machine is not achieved. There are many challenges for the person identification such as size, color, orientation and occlusion. The face recognition, recently available for use in the restricted environment.

The person identification is done using face matching process. In this case, face images are stored in the face database. The unknown face image is matched with the face images available in the face database. The Face Recognition is implemented to person recognition but the constraints is the person should be close enough and also should front towards the camera. This process of face identification has limitations for real time face recognition application.

In surveillance application, person recognition becoming very important as video cameras are installed in different areas. Previous work related to the Identification of Person is done through Facial Recognition only and that in addition, when the person has to show himself in front of the camera with properly aligned face fronting camera. This approach was very tedious as each time person has to manually show himself in front of camera to mark himself present many areas. This produces large video data for the processing.

The person identification in surveillance video is challenging problem due to several issues like person orientation, scale, occlusion by other objects, lighting illumination etc. This paper the problem of person Identification using process of the person re identification is explored.

Person re-identification is the process of mapping images of the individual person captured from various cameras or in a different directions or in different situations or instances. Another way to define is allocating an identity (ID) to a person in multiple camera configuration. Generally the re-identification is limited to a minor duration and a small environment (area) covered by camera. Humans have that ability to recognize other persons by using descriptors based on the person's characteristics related to body like height, face, clothing, hair style and shade, locomotion(walk pattern), etc. and this seems to be an easy problem for humans but for a machine to solve this problem is extremely difficult.

In visual surveillance technique, it is very important to link or associate individual people across different camera orientations. Cross view individual person re-identification ensure automatic identification and structure of particular individual person-specific features or movements over huge expanded environment and it is important for surveillance used in many applications for example tracking people using multi-camera and in forensic search. Particularly, for doing person re-identification, one compares a query person (person to be identified) the image is captured by camera view against a database created of the many people captured in another view for creating a ranked list or array according to their comparison distance similarity index.

The most existing methods or approaches in order to perform ReID (re-identification) by changing visual appearance such as shape of the face, texture of the body and color of individual or multiple person's images. People's appearance is naturally limited because of the unavoidable ambiguities related to visual ability and untrust due to appearance

similarities among various different people and variations in appearance of the individual same person from unknown changes in human pose(a way of standing or sitting), point of view, lighting, occlusion effect(blockage) and changing background color. This is what motivates the need or requirement (demand) of obtaining additional visual information sources for person re-identification. Now Person Identification is not through only face recognition but using other body alignment features also. This paper divided into the following sections. The section 2 describes the literature survey required for understanding the research done in this areas. Section 3 describes person identification methods implementation using face recognition and body part recognition. Sections 4 explores result of the person identifications and conclusion is given in Sections 5

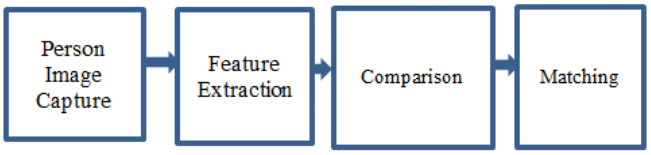

Figure 1.1 : Facial Recognition Technology

2. LITERATURE SURVEY

Person Re-identification is one important and difficult problem for that the research going on, in the areas of computer vision. The research in face recognition techniques and with the achievement and advancement in face recognition task, researchers are extending the face recognition task to person recognition task. It is being used in many countries to automate the monitoring process in surveillance system. In the previous work of person re-identification, researchers are considering this problem as facial recognition and tried to find solution in the same way.

Typically, Person Re-identification (re-id) task is broken down into two main component. First component consist of extracting unique features from the body of person. It can be facial features or can be from other parts of body. Other features may consist the cloth type, their body shape, etc. So the methodology designed for extraction of features should extract unique features from body and it should not match with other person's features. The second component of person re-identification task is to learn similarities between these features. Once the features has been extracted, these features can be given as input to metric learning task. Features of same person will give similar features and distance between features will very less in n-dimensional space. To find similarities between features, researchers tried out different techniques such as Euclidean Distance, Manhattan Distance, and Cosine Similarities. Feature extraction and metric learning techniques are explained in next paragraph.

The Histogram of Gradients is used to extract global features from an image. It gives fixed size vector of descriptor. It first calculates vertical and horizontal gradients of pixels. Adjacent diagonal gradients are calculated by taking the resultant of these horizontal and vertical gradients. The gradients are more sensitive where the pixel intensity is changing gradually. It refers as the adjacent pixel is of any other object. Once the gradient magnitude and gradient direction is calculated, its histogram is created. These histograms represent a fixed size vector where size of vector is same as the number of bins used in histogram. Histogram of Gradients feature descriptor is one of the most important component in face detection task. There are many feature extraction techniques present, which can be used in person re-identification task according to its use cases. Some of the frequently used algorithms for feature extraction are SIFT, SURF, HOG, etc. In order to handle image frames of different resolution, Scale Invariant Feature Transform was discovered. One of the main problem in different resolution of images is that same type of image may give different features.

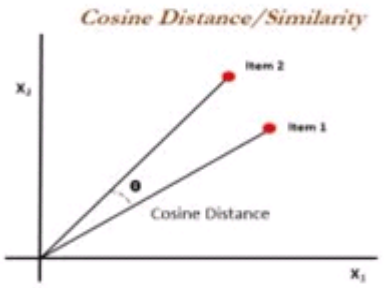

Figure 2.1 : Frequently used similarity technique.

Consider a situation where we extract circular feature from a part of any image. Now feature extractor will extract feature corresponding to circle. If the same image is zoomed in, same circle curvature will start getting lost and now new extracted features will not match to the previously extracted features. To handle this type of problem, Scale Invariant Feature Transform is used which is mostly invariant to the image scale. It will handle the case when people is standing too close to the camera. Although Scale Invariant Feature Transform gives, only local features that is not suitable input to the metric learning component. In order to find out global features from the images, Histogram of Gradients are used. With the evolution of deep learning, achievements in computer vision field improved drastically. With object classification and detection to localization and segmentation task, deep learning algorithms drive every field. The key aspects behind success of such algorithms are well-generalized data, network architecture and construction of loss function. These complex architecture and design of loss function is responsible for learning the complex pattern inside image. In analysis of image, Convolutional Neural Networks have shown many successful results and it is considered as main backbone for solving any image-based problems with deep learning. Convolutional Neural Networks are inspired by the Artificial Neural Networks and only difference is the learning methodologies of the weights parameter. CNNs are able to learn spatial information from the images. Weights parameter in Convolutional Neural Network are also known as filters or kernels. While training of Convolutional Neural Network, these filters are, being learnt which corresponds to minimization of loss function. In very recent works on person

Re-identification, which achieves results more than 80%, uses Convolutional Neural Networks for automatically selecting features of the object. There are many benchmark networks are present which were trained to classify images. These networks are trained on millions of images to classify them into thousands of classes. The weight learnt by these networks extract low-level features as well as high-level features from images. In addition, since they are trained on such large dataset, it can extract unique features

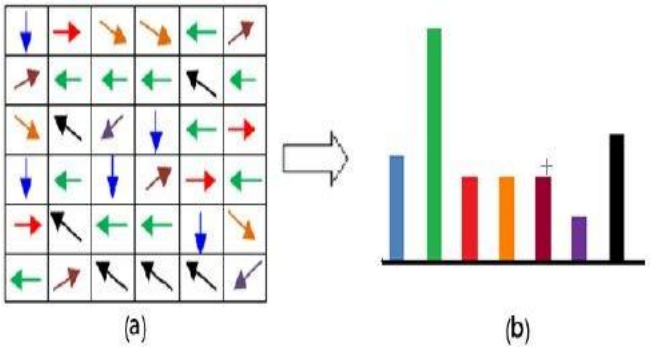

Figure 2.2 : Histogram of Gradient

from images. Therefore, most of the successful architecture of individual person re-identification is motivated by the architecture of Convolutional Neural Networks and benchmark networks are being used as a backbone for feature extraction task.

Recently, with the advancement in development of deep learning algorithms, there are mostly three kinds of network applied in person re-identification. These are classification network, Siamese network, triplet networks. Classification networks treat this problem as one versus all or yes no classification problem. It directly extract features from the input images with the help of superior performance of deep Convolutional Neural Networks on large-scale dataset. In recent work by Xiao[2], they jointly trained a classification model on multiple domains by using a domain based dropout method

to improve the performance.

In case of one versus all classification, the network give probabilities of all the labels present which sum up to one. In case of yes and no type classification, the network will give probability of identification for a particular label, i.e. probability of yes and probability of no which will sum up to one. Classification networks modelled and trained on the basis of end-to-end fashion or one can also use metric learning to find out the probabilities of the similar classes obtained .

The loss function is made such that it will minimize the distance between probe image and positive image features and simultaneously it will increase the distance between probe image and negative image. Loss function consist of a margin value, which denotes the threshold of similarity up to which a network can consider. Based on work of G.C. Wang et.al.[4], they developed a point-to-set triplet for the image-to-video person re identification. After excellent achievement in person re-identification, some of the few works made initial attempts to solve challenges like occluded person.

Siamese networks [8] are a neural network architecture. In general, a convolutional neural networks loss function is defined such that it learns the similarities and patterns in the images but in Siamese network, the model is trained instead learning to classify its inputs, the neural network provided with training to differentiate between two inputs. Siamese networks make use of contrastive loss function. It consists of two identical neural networks also known as sister network.

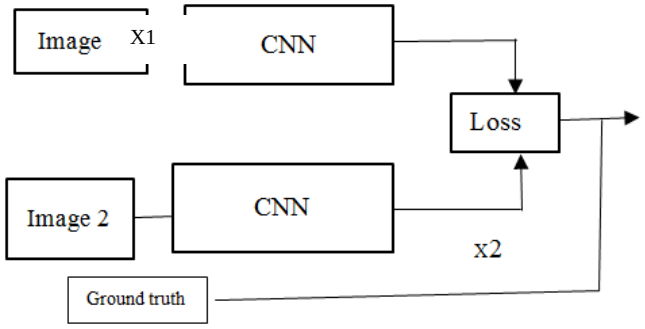

Figure 2.3 : Siamese Network

Each network shares exact same filter. Image pair are fed as input to the both network and last layer output s are send to the contrastive loss function that is last part of architecture. It take image pairs as input and learns similarities between them. Siamese networks are the most trending and give the optimum performance in face recognition task. The neighborhood differences [3] of an image pair to learn a metric indicating whether these are different images of same person. It was one of the major performance increment observed in the field of person re-identification. Triplet networks [7] are the extension version of Siamese networks. It make use of triple loss function, which is used to calculate gradients and update the network parameter accordingly. While training process, the network take three images as input namely anchor image, positive image, and negative image.

W. S. Zheng [5] suggested a solution in terms of partial person re-identification. They aimed to match partial image with the gallery of full body image. They only focused on the matching between body parts, which are not occluded and full-body parts. Some of the critical problems are still need to be solved like the output of the occluded person detection model. This case arises when a person standing behind other person and in person detector model, it will detect both partial image and consider both partial image as a single image. The work by Cheng et.al.[6], have done it manually and it seems quite unrealistic in practice. Another suggested approach is patch based matching method. It works to some extent but it need a large amount of calculations. Some of these problems were solved. Jiaxuan Zhuo et.al. [1], proposed a Convolutional Neural Network which differs from the approach of Cheng's solution as they directly compute the matching between occluded person images and full-body person images and propose an Attention Framework of Person Body(AFPB) framework that automatically focuses on the person body by watching various occluded person data generated by an Occlusion Simulator. The Attention Framework of Person Body(AFPB) includes two main components, i.e. Occlusion Simulator (OS) and multi-task losses. The Occlusion Simulator (OS) aims to generate artificial occluded person data. These data are used to simulate occlusion cases using full-body person data The Convolutional Neural Network architecture explored by Zhuo et.al.[6] perform best on ResNet-50 as a backbone architecture. It surpasses the entire previously presented model. Most frequently used datasets in person re-identification problem are CUHK and Market 1501.

$$(1-Y)\frac{1}{2}(D_W)^2 + (Y)\frac{1}{2}\{max(0, m - D_W)\}^2$$

Contrastive Loss Function

(D_w : Euclidean Distance between Features)

Person Re-ID research initiated with multi-camera image tracking. Several important re-identification directions have been developed since then. Some of the famous approach that is driving this research field are: Multi-camera image tracking, multi-camera tracking with explicit re-identification, video based re-identification, end to end deep learning model for re-identification task. All recent works achieving state of the art performance are based on deep learning methods.

Person Re-identification task is approached by one shot learning mechanism and Siamese network always give promising result in case of one shot learning. Liang Zheng et.al. [6] proposed very clear and explained survey of re-identification task. Most of the work in person re-identification is done in collaboration of Liang Zheng. His major contribution in re-identification task is making market-1501 dataset. He continuously maintain and upgrade market-1501 dataset. He also supported this research field by maintaining record of state of the art performances and creating new evaluation methods. Typical Person Re-identification (re-id) work mainly consist of two steps: feature image classification and instance retrieval. The first step is to extract a robust and distinctive feature representation that is invariant to the challenges such as illumination, viewpoint, and occlusion, etc. The second step learns metrics or subspaces for better matching such that distances of the same class are closer than those of the different ones. Recently, with the development of deep learning, there are three kinds of network frameworks applied in person re-id, i.e., classification networks , to obtain 3D models. They reported enhanced performance on the positive and negative samples. In person re-identification, Some hybrid version of deep convolutional Neural Networks is also used. Several end-to-end deep Siamese convolutional neural network architectures have been proposed for human re-identification. These comparisons were done only at final level of architecture. X. zangh et al. [5] proposed a Siamese Long Short Term Memory (LSTM) architecture for human re-identification. Xiao et.al [9], they also make use of LSTM architecture. They tried to increase accuracy by transferring the information to person re-identification task. They trained identity classification and attribute recognition from deep convolutional neural network to learn person information. They extended the architecture of LSTM by a special gate. This Siamese networks [8] are a neural network architecture. In general, a convolutional neural networks loss function is defined such that it learns the similarities and patterns in the images but in Siamese network, instead learning to classify its inputs, the neural network

$$Loss = \sum_{i=1}^{N} \left[\| f_i^a - f_i^p \|_2^2 - \| f_i^a - f_i^n \|_2^2 + \alpha \right]_+$$

Triplet Loss

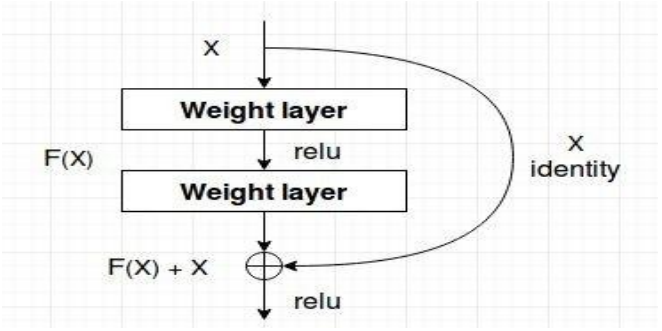

Figure 2.4 : Residual Network

learns to differentiate between two inputs. Siamese networks make use of contrastive loss function. It consists of two identical neural networks also known as sister network extended component is used to pay attention to certain special part of recurrent units. Y Zhang et.al.[1], proposed extended version of model distillation, i.e. Deep Mutual Learning. Model distillation is consists of teacher student network to meet requirement of low memory usage and faster execution. In deep mutual learning, an ensemble of students learn collaboratively and teach each other throughout the training process. Zhedong Zheng et.al [8] ,proposed proper analysis of two popular CNN in re-identification i.e. verification and identification network. They tried to make full usage of re-identification annotations by learning discriminative embedding and a similarity measurement at the same time. Schumann, A. et.al[7] do another similar type of work. They provide automatic prototype discovery for domain perceptive person re-identification. The separate re-identification model trained for each of the discovered prototype domains. The trained model is used to probe image to select automatically the model of the closet prototype domain. They created a deep domain embedding and after training they performed unsupervised clustering on embeddings. Another improvement measure has been done when dataset is subject to occlusion. Many alignment network is tried to cope out with problem of occlusion. A similar work of Liang Zheng et.al [10]

proposed pose invariant embedding for deep person re-identification. Their work improve the alignment feature even for varying pose to make the re-identification task more robust and accurate. It is generated through affine transformations. They also introduced Pose Box Fusion to reduce the pose estimation error and information loss during pose box construction. Wang T. et.al[2,11], explored work in video processing using image sequencing and to preserve space time information. Tesfaye, Y. et.al.[12] tried to solve the problem of tracking in multiple non-overlapping cameras.

Wei, L. et.al. [13] proposed explicitly leverages the local and global cues in human body to generate a discriminative and robust representation. Qiqi Xiao et.al[14], proposed a new metric learning loss with hard sample mining called margin sample mining loss (MSML). Continuing on looking for other metric learning approaches. Zhang, D et.al. [15] developed a quadruplet loss, which can lead to the model output with a larger inter-class variation.

Zhang, X. et.al. [16], proposed method for comparison of image with video, video sequence represents addtional temporal information that utilized to improve the performance of re-identification system. Chen, W. et.al. [17] explored Person re-identification (re-ID) task is considered as an important task in computer vision.[18]. MSML is a new metric learning loss with hard sample mining, which can attain better accuracy compared with other metric learning losses, such as triplet loss Most of the work in person re-identification is done in collaboration of Liang Zheng[8]. His major contribution in re-identification task is making market-1501 dataset.

Liu, H. et.al. [20] , used very strange approach to improve the accuracy and robustness of re-identification task.

an over cross-data set transfer learning approach has been developed by Peixi Peng et.al. [21] to learn a discriminative representation.

In work of Ying-Cong Chen et.al.[22], proposed view specific person re-identification frame work from the feature augmentation point of view. The understanding of person re-identification, we come to know about the global features and local features. The triplet framework for re-identification. The proposed CNN model consist multiple channels to jointly learn both global and local features. The model is trained using minimizing loss function. We all know that identification is based on appearance [23]. The appearance-based methods of person re-identification. Chromatic content, the spatial arrangement of colors into stable region. All information is learned from different body parts and then the weight assigned to different body parts. New model based on the video by analyzing the space-time cues available in the video or image sequence frame. We learned about new time-space representation by encoding multiple granularities of spatio-temporal in form of time series[24]. The re-identification is done by calculating time shift dynamic time warping. In order to identify the person from multi-camera Ma, X et.al[25] presented an approach for matching of signature based on interest point descriptors collected on short video sequence. This approach use time-space images using person tracking and re-identification. Now to solve any real life problem CNN is very much common so person re-identification can be done using CNN the features is extracted from top layers of pre-trained CNN on large annotated dataset. The model by fine tuning by conducting on pedestrian dataset. In that they refer new label by combining different attribute label and use then form additional classification loss function. This loss function helps to give more person specific information and yielding more discriminative features. After this apply normal CNN will give very great result[26]. The metric between two image sets. Using two images may be same or not we try to find the similarities between the images and it may possible that quality of image is not good but quality aware network is build which comforts this problem. This approach has two branches in first extract features and second one predict quality score for each sample allow information to flow between steps[27]. The feature extraction and matching are two main steps in re-identification of a person. Different pose will create difficulties in the process. The model learn from global features and different local features and assign weight to them. In most of the approaches neglected adapting the feature selection or learn model over time[28]. This type of problem are addressed a temporal model adapting scheme with human in loop. The similarity-dissimilarity learning method which can be trained in by incremental fashion[29].

3. IMPLEMENTATION

The input to the person indetification is taken from the video camera. The application divides this footage into numerous frames and serially these frames are used for identifying the persons captured in the footage. Frames generated through main application are provided as an input to the 2 sub-modules that we are using for identifying the person i.e., Face Recognition Module [FRM] and Body Recognition Module [BRM].

This module is composed of series of several things such as detecting face images in all pictures. The second part, focus on each face, third, picking the unique features of the face that distinguish this face from other people's faces, finally, compare these captured unique features to all the people's we already know to determine the label of the detected face of person. All this phases are put in a pipeline where each phase of face recognition is solved and the result is passed to next phase. This means we have to chain together various ML algorithms. The process flow diagram in shown in figure 3.1.

Step 1: Finding all the Faces This is the first step or phase in our pipeline. In this phase, we have to find the areas of the

image we have to pass on to the next phase. We are going to use Histogram of Oriented Gradients [HOG] method for face detection

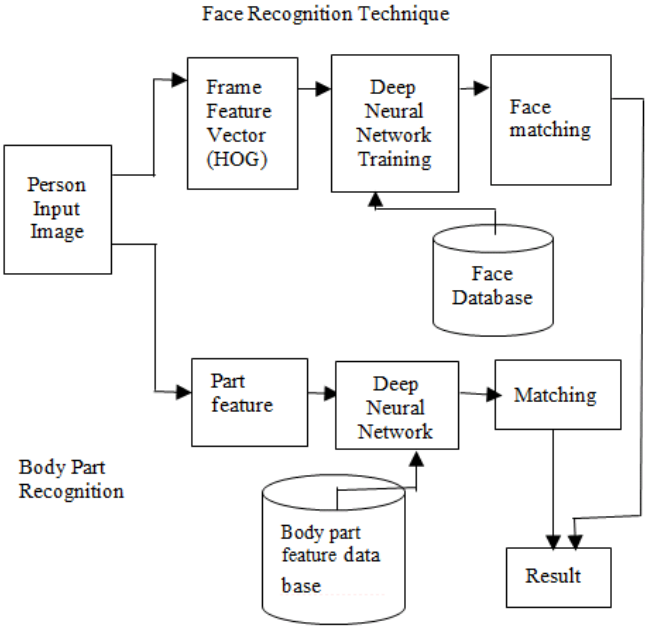

Figure 3.1 : Flow Diagram Face

Figure 3.2 Person Identification flow diagram

In order to detect faces in an image, the face frame is converted into black and white. In the image a single pixel one at a time is checked. In addition, for that each single pixel, the surrounding pixels are also checked. The current pixel is compared with its surrounding pixel and then an arrow is drawn in order to show in which direction our image is getting darker. This process is repeated for every single pixel in the image and finally each pixel is replaced by an arrow, which is known as gradients because they show the direction of the flow from black area to white area in the whole image. However, this method of saving the gradient for every pixel results in many details and there is chance to error in recognizing the overall pattern from details. It would be good to see the basic flow of lightness or darkness at broader level that ensure the basic pattern of the image. So, the image is divided into small squares of 16*16 pixels. In each square the gradient points in eight directions are located such as top, bottom, right left, etc. is counted. Then we will replace that square with the arrow direction whose count is more i.e. strongest one amongst others. After doing all this , a simple face image representation which encodes the basic structure of a face. To find a face, search a part of our image that is resemble to known patterns that were extracted from training faces. This technique easily helps in finding faces in any image.

Step 2: Posing and Projecting Faces

In this step, the face orientation is considered for processing.

To overcome this problem, we will use face landmark estimation algorithm. There are 68 specific points that are present on every face which are known as landmarks. We will train a ML algorithm, which will be able to find these landmarks on any face given as input. Now we know where eyes is and where is mouth, we will apply basic image transformation like rotation, scaling and shearing so that mouth and eyes are centered.

Step 3: Encoding Faces

In this step, the face frame image is submitted to the deep neural network to extract basic feature of the face image Using these measures, the known face with the closest measurements is determined. These basic measurements are not defined initially. Deep neural network learning itself figures out which measurements are important and then the face encoding measurements are extracted automatically.

The face frame image is used as input to the deep Convolutional Neural Network. The deep convolution network is trained using face images as training dataset. The deep convolution neural network process the face image through several internal network nodes and finally output is generated by the output nodes of the deep neural network. These output is denoted as 128 measurements for each face.

The training process of the deep convolution neural network works as described below:

1. In the training process, all images which are used in the training set are stored as face dataset in a database. These images are also called as gallery images.
2. The deep neural network model is developed during the this training process.
3. The test images of the one known person(this image is present in the training dataset) is loaded for testing of the trained model.
4. The test images of the unknown person(this image is not present in the dataset) is also loaded for testing of the trained model.
5. The 128 face encoding measurement is generated for all training images. Similarly 128 face encoding also generated test images in step no 3(known image) and 4(ununkown image).
6. The distance from the test image (known) is less when compared with training image distance.
7. The distance from the test image (known) is more when compared with training images distance.
8. The less distance image is closed approximation for the test image and identified as person identification.

Step 4: Finding the person's name from the encoding

In this last step, we have to find the person in our database knowledge base i.e., database of known people who has the embeddings closest to our test image. We will use Support Vector Machine [SVM] classifier. We need to train a classifier that can extract embedding from a new test image and results in name of the known person who has closest embedding.

Body Recognition Module [BRM]

The Body recognition module takes frame as input from main application. After that the following operation are done one is crop body from the image and other Identification of body crop body from the image: The whole frame may contain more than one person in frame. So first of all we need to crop all the bodies of person from a given frame so that it is easy to identify the person. For cropping of body from a given frame. Every Bounding box contain five value associated with itself : x,y,w,h and confidence. The x and y value gives information about the center of box with respect to the grid cell, w and h represents width and height of the box relative to the given frame and finally we have confidence which is how much it is confident about predicted class this calculated based on the conditional probability with every other class.

Identification of body**:**

After detecting, the body of person next step is to identify the body. To identify the person we refer the architecture shown below:

This architecture compare two image and tells us how much they are different from each other less the difference means images are probably same. The first image is cropped image and the second image is from the known body part database (which contain folder of every person and each folder contain some images of respective person images) then this two images compared with each other and distance is calculated. The steps through which the image goes are given below:

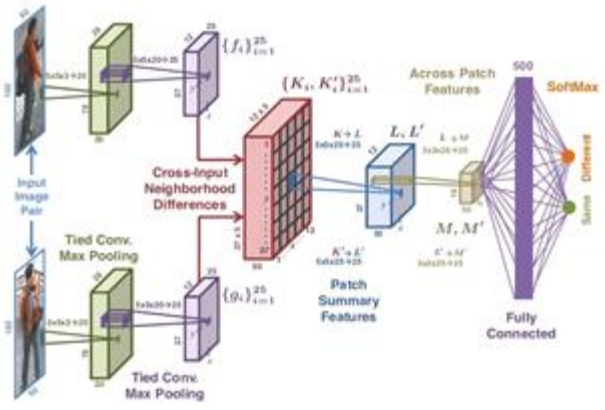

Figure 3.2 : Model Architecture

First of all, The given frame image into 60 X 160 RGB image. Then this 60 x 160 x 3 passed through 20 learned features of size 5 x 5 x 3. The resulting frame passed through max pooling layer which almost half the length and width of the frame and then resultant frame again passed through another convolutional layer of max pooling of 25 learned filters which result the image of 12 x 37 x 25.

The two tied convolution layers provide a set of 25 feature maps for each input image from these, we get the relationship between two images. Then calculate the difference of 25 neighborhood of 5 x 5 of strid of size 1. It is observed that the difference is not symmetric so we calculate from each other from both side (f-g and g-f). Which result into 50 neighborhood difference maps each of size 12 x 37 x 5 x 5. Then we need to summarize the data we have so we consider a stride of size 5 which result into 12 x 37 x 50. Then then we again pass it through convolution layer of stride of size 1 which reduce length and width by factor 2. After this the data is flatten and passed through neural network

Architecture in which it is passed through number of hidden layer and as a result, the difference between two image is given. Based on the distance we calculate the person in which the given cropped image matches most and accuracy is returned. In this way the person re-identification is done.

4. RESULT

The person identification using deep learning is experimented with our model using a video footage, which is a combination of video clips of ten to fifteen seconds duration. These video clips are broken down into large number of frames that need to labelled. We take each frame from testing dataset and compares it with the knowledge base dataset. Label which gets highest probability of similarity index is been assigned to that testing frame and the result i.e., frame with output label is being written and saved. Our frames consists of single as well as multiple persons that needs to be labelled. When experimented with our own dataset, that consists of 690 frames that results into 741 cropped images of persons, we are able to correctly classify 578 number of cropped images out of those 741-cropped images. That results into an accuracy percentage of approx. 78 %.

5. CONCLUSION

In this paper, a deep architecture for person identification is presented. The implementation of the person identification using two modules i.e., FRM and BRM. The combined result of both the modules i.e. FRM and BRM is considered as the result. The result demonstrate the effectiveness of our method by performing testing on our own local dataset.

REFERENCES

[1]Zhang, Y., Xiang, T., Hospedales, T. M., & Lu, H. (2018). Deep mutual learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4320-4328).

[2] Wang, T., Gong, S., Zhu, X., & Wang, S. (2016). Person re-identification by discriminative selection in video ranking. IEEE transactions on pattern analysis and machine intelligence , 38 (12), 2501-2514.

[3] Radenović, F., Tolias, G., & Chum, O. (2016, October). CNN image retrieval learns from BoW: Unsupervised fine-tuning with hard examples. In European conference on computer vision (pp. 3-20). Springer, Cham.

[4] Varior, R. R., Haloi, M., & Wang, G. (2016, October). Gated Siamese convolutional neural network architecture for human re-identification. In European conference on computer vision (pp. 791-808). Springer, Cham.

[5] Varior, R. R., Shuai, B., Lu, J., Xu,D.,& Wang,G.(2016,October).A Siamese long short-term memory architecture for human re-identification. InEuropean conference on computer vision (pp. 135-153). Springer, Cham.

[6] Zheng, L., Yang, Y., & Hauptmann, A. G. (2016). Person re-identification: Past, present and future. arXiv preprint arXiv:1610.02984 .

[7]Schumann,A.,Gong,S.,& Schuchert, (2017,September). Deep learning proto type domains for person re-identification. In 2017 IEEE International Conference on Image Processing (ICIP) (pp. 1767-1771). IEEE.

[8] Zheng, Z., Zheng, L., & Yang, Y. (2018). A discriminatively learned cnn embedding for person re-identification. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) , 14 (1), 13.

[9] Xiao, Q., Cao, K., Chen, H., Peng, F., & Zhang, C.(2016). Cross-domain knowledge transfer for person re-identification. ArXiv preprint arXiv:1611.06026 .

[10] Zheng, L., Huang, Y., Lu, H., & Yang, Y. (2019). Pose invariant embedding for deep person re-identification. IEEE Transactions on Image Processing .

[11] Zhang, W., Hu, S., & Liu, K. (2017). Learning compact appearance representation for video-based person re-identification. ArXiv preprint arXiv: 1702.06294 .

[12] Tesfaye, Y. T., Zemene, E., Prati, A., Pelillo, M., & Shah, M. (2017). Multi-target tracking in multiple non-overlapping cameras using constrained dominant sets. ArXiv preprint arXiv:1706.06196 .

[13] Wei, L., Zhang, S., Yao, H., Gao, W., & Tian, Q. (2017, October). Glad: Global-local-alignment descriptor for pedestrian retrieval. In Proceedings of the 25th ACM international conference on Multimedia (pp. 420-428). ACM.

[14] Xiao, Q., Luo, H., & Zhang, C.(2017).Margin sample mining loss: A deep learning based method for person re-identification. ArXiv preprint arXiv:1710.00478 .

[15] Zhang, D., Wu, W., Cheng, H., Zhang, R., Dong, Z., & Cai, Z. (2018). Image-to-video person re-identification with temporally memorized similarity learning. IEEE Transactions on Circuits and Systems for Video Technology , 28 (10), 2622-2632.

[16] Zhang, X., Luo, H., Fan, X., Xiang, W., Sun, Y., Xiao, Q. & Sun, J. (2017). Aligned re-id: Surpassing human-level performance in person re-identification. ArXiv preprint arXiv:1711.08184 .

[17] Chen, W., Chen, X., Zhang, J., & Huang, K. (2017). Beyond triplet loss: a deep quadruplet network for person re-identification. InProceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 403-412).

[18] Chen, Y., Wang, N., Zhang, Z. (2018, April).Dark rank: Accelerating deep metric learning via cross sample similarities transfer. In Thirty-Second AAAI Conference on Artificial Intelligence .

[19] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[20]Liu, H., Feng, J., Jie,Z.,Jayashree, K.,Zhao, B.,Qi,M., .& Yan,S. (2017). Neural person search machines. In Proceedings of the IEEE International Conference on Computer Vision (pp. 493-501).

[21]Peng, P., Xiang, T., Wang, Y., Pontil, M., Gong, S., Huang, T., & Tian, Y. (2016). Unsupervised cross-dataset transfer learning for person re-identification. InProceedings of the IEEE conference on computer vision and pattern recognition (pp. 1306-1315).

[22]Chen, Y. C., Zhu, X., Zheng, W. S., & Lai, J. H. (2017). Person re-identification by camera correlation aware feature augmentation. IEEE transactions on pattern analysis and machine intelligence , 40 (2), 392-408.

[23]Liao, S., Hu, Y., Zhu, X., & Li, S. Z. (2015). Person re-identification by local maximal occurrence representation and metric learning. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2197-2206).

[24] Cheng, D., Gong, Y., Zhou, S., Wang, J., & Zheng, N. (2016). Person re-identification by multi-channel parts-based cnn with improved triplet loss function. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1335-1344).

[25] Ma, X., Zhu, X., Gong, S., Xie,X. ,Hu,J. ,Lam,K. M.,& Zhong, Y.(2017).Person re-identification by unsupervised video matching. Pattern Recognition , 65 , 197-210.

[26] Hamdoun, O., Moutarde, F., Stanciulescu, B., & Steux, B. (2008, September). Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences. In 2008 Second ACM/IEEE International Conference on Distributed Smart Cameras (pp. 1-6).

[27] Matsukawa, T., & Suzuki,E. (2016,December). Person re-identification using CNN features learned from combination of attributes. In2016 23rd International Conference on Pattern Recognition (ICPR) (pp. 2428-2433).

[28] Liu, Y., Yan, J., & Ouyang, W. (2017). Quality aware network for set-to-set recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 5790-5799).

[29] McLaughlin, N., Martinez del Rincon, J., & Miller, P. (2016). Recurrent convolutional network for video-based person re-identification. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1325-1334).