

# A new framework for Person Re-identification: Integrated level feature pattern (ILEP)

V.Manimaran<sup>1\*</sup>, K.G.Srinivasagan<sup>2</sup>, S.Gokul<sup>3</sup>, I.Jeena Jacob<sup>4</sup> and S.Baburenagarajan<sup>5</sup>

<sup>1</sup>Department of Information Technology, National Engineering College  
Kovilpatti-628503, TamilNadu, India  
[e-mail:jeyamaran2002@gmail.com]

<sup>2</sup>Department of Information Technology, National Engineering College  
Kovilpatti-628503, TamilNadu, India  
[e-mail:kgsnec@rediffmail.com]

<sup>3</sup>Language Science and Technology, SaarlandUniversity  
Saarbrücken-66123, Germany  
[e-mail:sgokul151998@gmail.com]

<sup>4</sup>Department of Computer Science and Engineering, Gitam University  
Bengaluru -561203, Karnataka, India  
[e-mail:jeni.neha@gmail.com]

<sup>5</sup>Department of Computer Science and Engineering, PET Engineering College  
Vallioor-627117, TamilNadu, India  
[e-mail:babu\_sbr01@rediffmail.com]

\*Corresponding author:V.Manimaran

*Received July 20, 2021; accepted Septebmer 23, 2021; published December 31, 2021*

---

## Abstract

The system for re-identifying persons is used to find and verify the persons crossing through different spots using various cameras. Much research has been done to re-identify the person by utilising features with deep-learned or hand-crafted information. Deep learning techniques segregate and analyse the features of their layers in various forms, and the output is complex feature vectors. This paper proposes a distinctive framework called Integrated Level Feature Pattern (ILFP) framework, which integrates local and global features. A new deep learning architecture named modified XceptionNet (m-XceptionNet) is also proposed in this work, which extracts the global features effectively with lesser complexity. The proposed framework gives better performance in Rank1 metric for Market1501 (96.15%), CUHK03 (82.29%) and the newly created NEC01 (96.66%) datasets than the existing works. The mean Average Precision (mAP) calculated using the proposed framework gives 92%, 85% and 98%, respectively, for the same datasets.

---

**Keywords:** Person reidentification, LBP, HOG, Deep features, PCA, CNN, m-XceptionNet.

---

This work was supported by Research club of National Engineering College, Kovilpatti. The dataset, NEC01 with 200 images were collected from the concerned campus.

## 1. Introduction

**P**erson Re-identification (Person re-id) [4] is an inevitable activity for ensuring security in significant overpopulated areas or public places such as bus terminals, railway stations, airports, hospitals, universities, and malls. A good surveillance system with intelligence, multi-camera tracking [6], multi-camera activity analysis [7, 8] and crowd counting [9, 10] is required for locating a particular individual or a missed child/person among the large crowd. The objective of reidentification is achieved by utilising the pictures of the same person gathered from cameras placed at dissimilar points. The real-world applications may demand the identification of one or more people, by which the target set may also change. This is done by collecting a set of cropped frames from the video sequences of different cameras, retrieving relevant images from the set, and identifying needed persons in the different frames. The reidentification was first proposed by Alvin [1] in 1961. The same theory was further implemented to re-identify objects for diverse applications such as logic [2] and psychology [3]. Later, the researchers worked on implementing this concept for person reidentification [4]. Though many works were reported on this area, effective implementation became quite challenging because of reasons like (1) Wearing similar attires by different people, (2) Variations in body shape because of occlusion and illumination conditions, (3) Different images taken at different viewpoints, (4) The picture capturing done with a camera of very low frame rate, and (5) Lack of accuracy in algorithm during real-time video detection. Nevertheless, the traits like shape and cloth's colour make it easier to progress the person reidentification.

Person re-id was being implemented using various network structures. The proposed work implements re-id with an integrated framework called the ILFP framework functions in two stages. During the first stage, the local level features are extracted using Histogram of Oriented Gradients (HOG) and Local Binary Pattern (LBP). In the second stage, global level features are extracted using a new deep learning network called m-XceptionNet. Our framework integrates these two-level features by reducing the dimensionality of local-level features and further integrating it with the global-level feature using a fully connected layer. The present investigation created a new dataset, NEC01, by collecting the images from the college campus. The analysis is carried out with the datasets NEC01, CUHK03, and Market1501. The proposed aspects of the framework lead to large improvements over the earlier works because the hand-crafted features extract the local information, and the m-XceptionNet extracts the global information. This integrated-level feature helps to get the prominent information from the image to classify it precisely.

The organisation of the current work is as follows: section 2 presents the related work on person reidentification with various features and CNN, section 3 describes and elaborates the proposed architecture and framework, and section 4 demonstrates the experimental setup and elucidation of the results followed by concluding remarks.

## 2. Related Works

Research works on person re-id methods were done in different directions. Earlier re-id methods used hand-crafted features [11-15] of the images for feature extraction. Different distance metrics [16-21] were used to compare these hand-crafted features to attain output. A few researchers [22-24] worked both on feature extraction and distance metrics. The hand-crafted features may be based on colour, texture, shape or spatial layout. The most

commonly used feature of images is independent of the view, and the resolution of an image is colour [25, 26]. The color information was grabbed in the forms of color edge [27], color texture [28], color histograms [29, 30], color moments [31] and color correlograms [32]. Shape [33-35] is another feature that gets the intensity difference between the regions. Patterns in the image appearance and spatial definition are expressed as texture information [36-39]. Many researchers worked on similar measures after the extraction of the features. Some of the works related to metric learning for re-id are Local Fisher Discriminant Analysis [23], saliency weighted distances [40], Mahalanobis metric learning [19], Marginal Fisher Analysis [23], Locally Adaptive Decision Functions and attribute consistent matching [17].

Deep learning-based works gave another dimension in person reidentification. These works implemented person re-id with Convolutional Neural Network (CNN). CNN was used to retrieve more features from the images. CNN models were used in two different ways, either as classification models [41, 42] or as Siamese models [43-45] using image pairs. The person re-id cannot merit a lot of training data since the applications are done in real-time. Even datasets like VIPeR [11] provide only two images for identifying a person. In the CNN model, the feature extraction uses three overlapping parts of the given image by sending it through two convolutional layers and one fully-connected layer. The outputs from these layers are fused for the feature vector. The feature vectors of two images are fed into cosine similarity metrics for finding the match. This work was emended by Li et al. [41] by adding a patch-based matching layer used for multiplying the responses of convolution layers from incompatible horizontal stripes of two images. Ahmed et al. [46] worked on the Siamese model to improve its effectiveness by calculating the cross-input neighbourhood difference using one location of an image to the features in the neighbouring locations of another image. Deepening this CNN was done using PersonNet [47], a small convolutional filter. Long Short-Term Memory (LSTM) was proposed to incorporate the Siamese network to advance this method [48]. The image parts are processed sequentially to memorise the spatial connections. Variator et al. [49] proposed another advancement in CNN by adding a gating function after every convolutional layer, which gives better accuracy. Many state-of-art surveys were carried out on person reidentification methods [50-53]. Considering these surveys, the volume of works better comprehended in this research area and the effect of using local and global features are also understood in person re-id works. This knowledge motivated us to incorporate both local and global features into the framework. The main contributions of this paper are.

1. A novel framework named Integrated Level Feature Pattern (ILFP) is proposed for person re-id.
2. Global features of the images are extracted using the newly proposed m-XceptionNet.
3. Local features are extracted using spatial-oriented gradient and texture.
4. A new dataset, NEC01, is created and used for performance analysis along with the state-of-art datasets.

### 3. Proposed ILEP work with m-XceptionNet

The proposed work implements person reidentification using the integrated-level framework with joint local and global features. The texture information in terms of HOG and LBP are used as local features. These features are optimised using Principle Component Analysis (PCA). The global features are extracted by using m-XceptionNet. Both local and global feature sets are combined in the fully connected layer called the dense layer, which further produces the classification output based on these features.

#### 3.1 Spatial-oriented Gradient and Texture

The images of the person reidentification system may not employ many unique features because they may be taken from far apart and in different illumination and contrast. In order to make it accountable for the changes that happen in contrast and illumination, the strengths of gradient measures should be normalised locally. During this process, the cells are grouped to form bigger spatially connected blocks. HOG descriptor [38] is calculated by forming the vector of normalised cell histogram components by utilising the block regions. The calculation is done in an overlapping manner to acquire the contribution from each cell repeatedly. Similarly, texture information also can contribute to the feature well by incorporating LBP [54].

##### 3.1.1. Descriptor extraction using HOG

The images used for person reidentification may vary in direction, position, illumination, and contrast. These variations should be normalised in spatial domain and direction. They can be normalised by extracting the descriptors in scale space and by normalising the rotation by detecting the extrema of scale-space, assigning the orientation, and extracting the descriptor. Fig. 1 illustrates a patch with their corresponding gradient directions of the given input image.

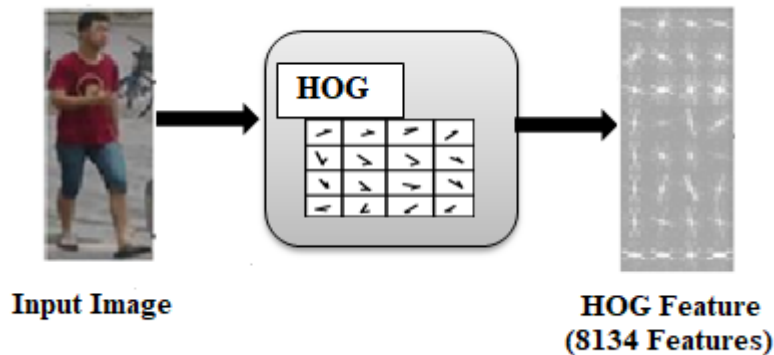


Fig. 1. Patch in HOG with their corresponding gradient directions

Gaussian smoothing is applied for testing the gradients by using several discrete derivative masks. Several smoothing scales had undergone testing. The One-dimensional vector  $[-1, 0, 1]$  works best as the mask for testing. If larger masks are used, performance gets decreased. When analysing the strengths of gradient vectors, it differs with local variations and contrast change. Hence local contrast normalisation helps us to get a proper descriptor block.

### 3.1.2. Descriptor extraction using LBP

Local Binary Pattern (LBP) is proved to be an efficient local texture feature. The LBP descriptor helps us to retrieve the inter-relationships between the pixels. In-person re-id, this information helps a lot to get the unique feature because the major information that can be retrieved from the image is the relationship between the pixels. Fig. 2 gives the formation of the LBP feature vector.

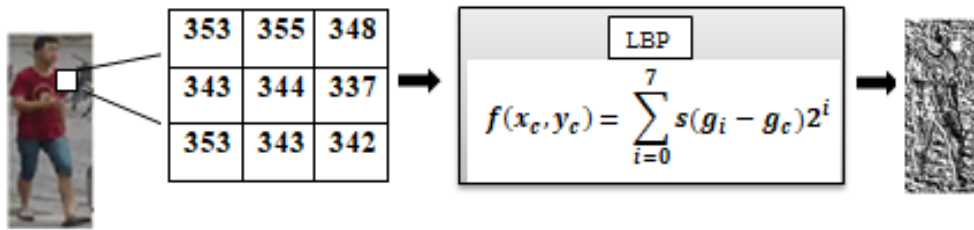
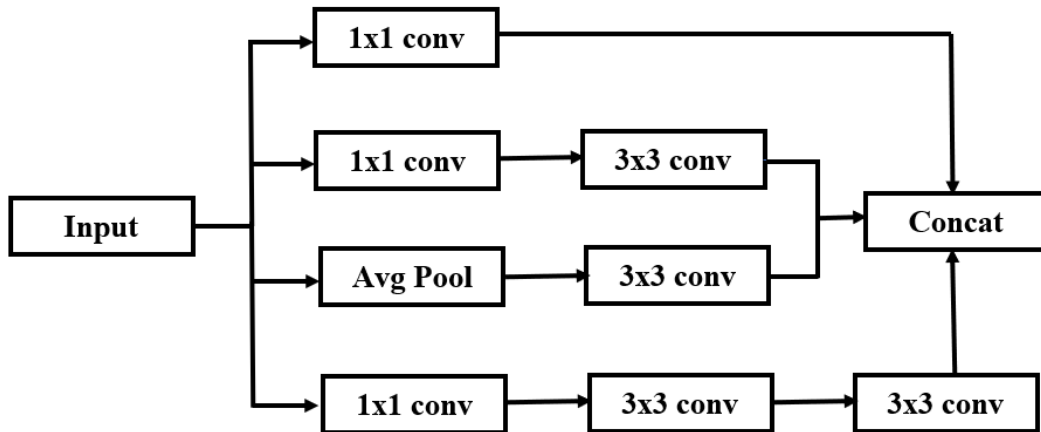


Fig. 2. LBP Feature formation of the image

The LBP calculation is done in a 3 x 3 window. For each neighbouring pixel, eight neighbouring pixels exist. The neighbouring pixels are substituted with 1 when that pixel value is bigger than the mid pixel and 0 otherwise. This process gives 8 binary numbers in the neighbouring positions. It is multiplied by  $2^i$ , where  $i$  is the corresponding position. The sum of these values is replaced in the mid pixel position, and the process is done throughout the image to get the LBP descriptor.

### 3.2. Global Feature extraction using m-XceptionNet

Different variations were proposed when the researchers used Convolutional Neural Network (CNN) according to the applications. The design of CNN started with LeNet [42], which has some convolution layer and max-pooling layer. Convolution layers were used to extract the features and sub-sampling in the spatial domain using the max-pooling layer. AlexNet architecture was proposed in 2012 analysed the change of adding many convolution layers between max-pooling layers [55]. Later, more deeper architectures like VGGNet [56] were proposed. During this time, a new architecture style was introduced by Szegedy et al. [57]. They proposed an inception module, as shown in Fig. 3. Stack of modules was used in different inception-based architectures, whereas convolution layers were stacked in the normal CNN architectures. The inception modules are called convolutional feature extractors that can learn richer features using fewer parameters, as depicted in Fig. 3.



**Fig. 3.** Inception module in InceptionV3

The inception module was initially used for deep learning using the 1 x 1 convolution layer and then different spatial correlations for different output channels. This kind of inception module was handled in GoogLeNet [57], Inception V3 [58] and Inception-ResNet [5].

The next advancement in the deep learning framework was separable convolution which learns deeper than the inception module. In this, spatial convolution is done for every input channel independently, following a 1x1 convolution. Separable convolution works better than the normal inception module due to non-linearity and change in the order of operation. In 2017, another framework called XceptionNet was proposed by Francois completely based on depthwise separable convolution layers. This framework decouples feature maps of CNNs, which contains the mapping of spatial and cross-channel correlations.

The Xception architecture comprises three flows: entry, middle and exit. The feature extraction phase of the Xception network has 36 convolutional layers modelled into 14 modules with linear residual connections except for the first and last modules. Based on this architecture, this paper proposes a new network called m-XceptionNet. The proposed m-XceptionNet is given in Fig. 4. The input image with 128x128 is fed into two convolution layers to get an intermediate output with 64 feature maps. This is fed into two separable convolutions layers and a max-pooling layer. The feature vector is added with the output of the convolution layer. Subsequently, the output of this layer is fed into three sepconv layers. The output image is fed into two sepconv layers and one max-pooling layer. These steps will generate the output with 1024 features maps and is fed into two sepconv layers and one global average pooling layer to produce 2048 feature vectors.

Initially, an image of size 128 x 128 with 3 channels (red, green, blue) is given as input to the network. The convolution operation is applied on the image with 32 filters, each with the filter size of 3 x 3, the activation function of RELU and the stride of 2 pixels. The resultant output dimension is 63 x 63 x 32. Then, another convolution is done with 64 filters, encompassing a filter size of 3 x 3 and a stride of 1 pixel. Later, batch normalisation is applied.

Following this, separable convolution of 128 filters is applied on the output twice with a filter size of  $3 \times 3$ , followed by max pooling. Further convolution of output after batch normalisation layer is performed with 128 filters. This output is connected to the output of max pooling using the residual connection. This forms a single block in the proposed network. This block is repeated three times and followed by a similar block, which yields 2048 features after global average pooling.

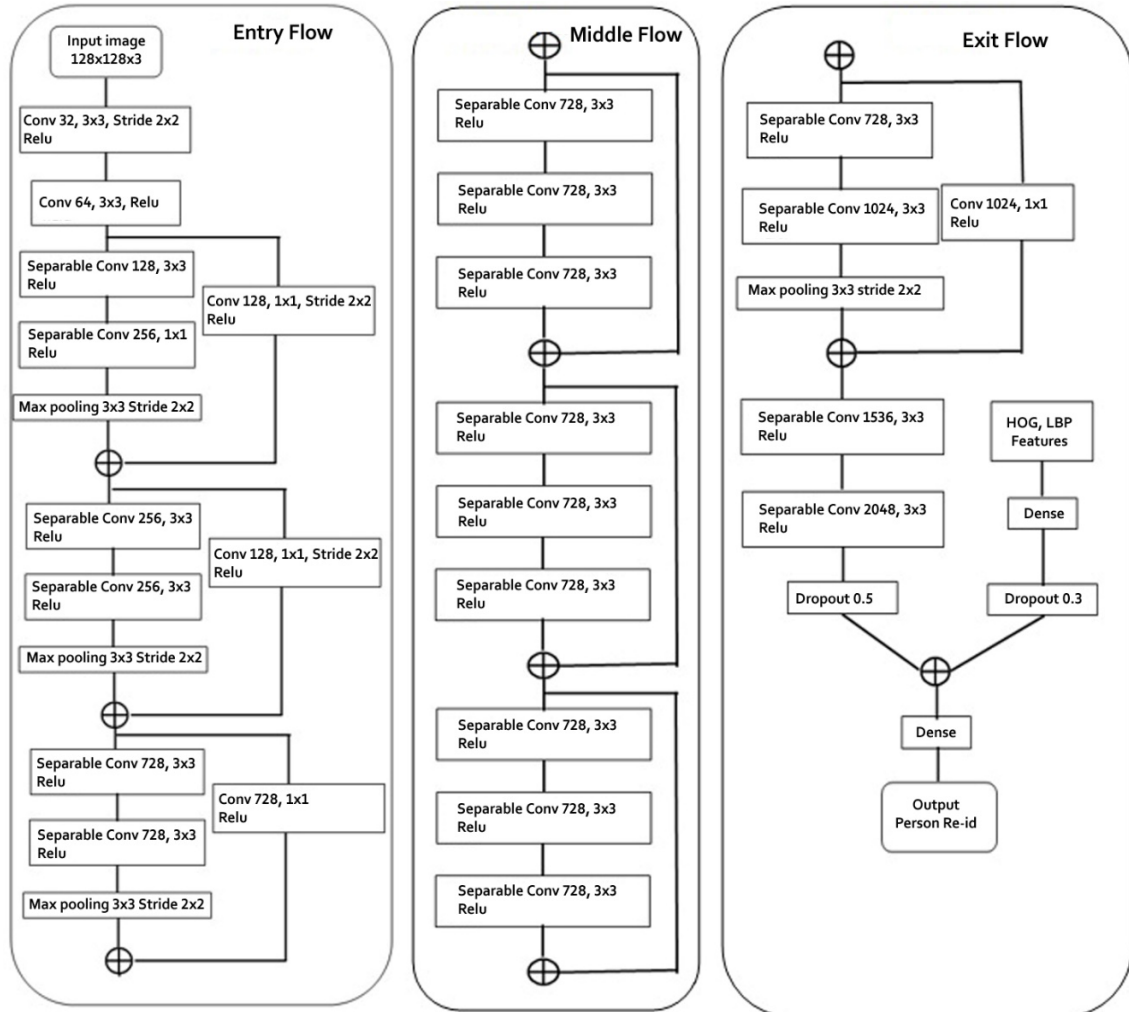


Fig. 4. Proposed m-XceptionNet

The proposed m-XceptionNet performs well because of depth-wise separable convolution levels used in this architecture. It extracts the information in spatial and depth dimensions. The depth-wise separable convolution can be broken down into two operations, Depth-wise convolution and Point-wise convolution. Fig. 5 shows Depth-wise and Point-wise convolution operations. Depth-wise convolution is the channel-wise  $n \times n$  spatial convolution. If 3 channels are used in depth-wise convolution, then three  $n \times n$  spatial convolutions are obtained. When Point-wise convolution is applied,  $1 \times 1$  convolution is used to change the dimension.

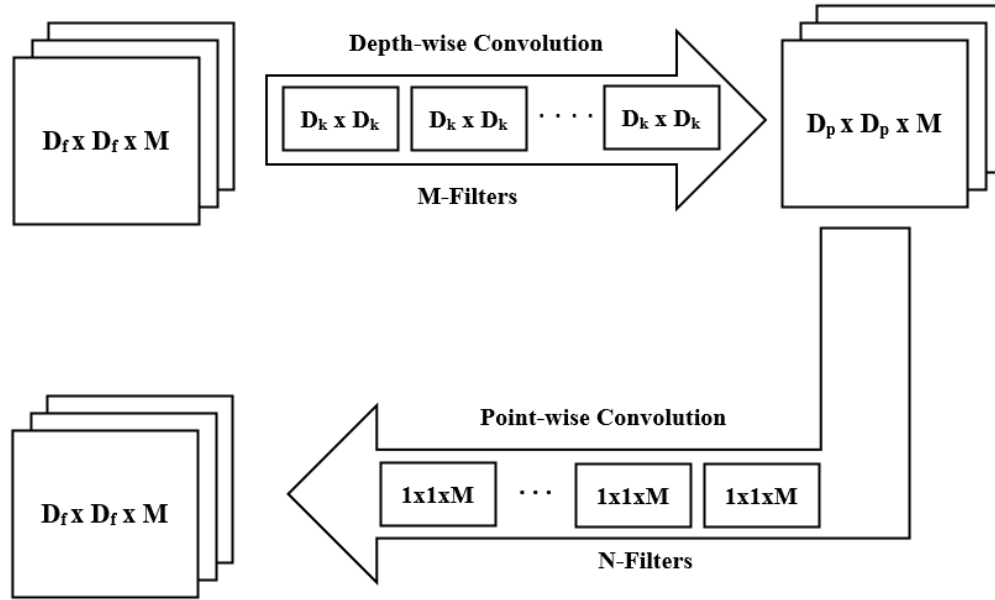


Fig. 5. Depth-wise and Point-wise convolution

Here,  $D_f$  refers to the number of pixels on the image,  $D_p$  refers to resultant dimension after depthwise convolution is applied and  $D_k$  refers to the size of the filter. In depth wise convolution operation, the convolution is applied to a single channel at a time. Here, the input size is  $D_f \times D_f \times M$ , the filter size is  $D_k \times D_k \times 1$ , and the number of filters is  $M$ . This produces output with size  $D_p \times D_p \times M$  with  $M \times D_p^2 \times D_p^2$  operations. In pointwise operation, a  $1 \times 1$  convolution operation is applied on the  $M$  channels. The input size as  $D_p \times D_p \times M$  filter size is  $1 \times 1 \times M$  and the number of filters is  $N$ . This produces output with size  $D_p \times D_p \times N$  with  $M \times D_p^2 \times N$  operations. Thus, the total number of multiplications in depth-wise separable convolution is as given.

$$\text{Depthwise separable conv} = M \times D_p^2 \times D_p^2 + M \times D_p^2 \times N = M \times D_p^2 (D_k^2 + N) \dots (1)$$

With a reduced number of multiplications than the convolution operation, the depth-wise separable convolutions can greatly reduce the computation time involved in the process.

### 3.3. Principle Component Analysis (PCA)

PCA is a method that descends dimensionality and was done by understanding the correlation between the variables. Though the dimensionality reduction is done more or less, this algorithm tries to preserve the variations between the variables to a greater extent. The covariance matrix is calculated for the variables for which the eigenvectors are found. Thus, by completely transforming the type of variables to a different set, the variation is achieved. These eigen vectors are known as the principal components, which are orthogonal. Ordering



of the principal components is done to showcase the variation of the variables. The first principal component will give maximum variation, which can be achieved using the original variables. This reduction in variables helps the algorithm to utilise the variables effectively with lesser time.

Spatial-oriented texture and gradient using LBP and HOG extract the local features. The global feature of the image is also important since the local feature misses the correlation between the pixels of the entire image. This is implemented by deep learning the raw image. Thus the proposed work focuses on retrieving both local and global features and ascends the accuracy of this work.

### 3.4. ILFP Framework

The detailed proposed ILFP framework is described as shown in Fig. 6. The image is fed into m-XceptionNet, LBP framework and HOG algorithm. The m-XceptionNet produces 2048 features that give the global vicinity of the image.

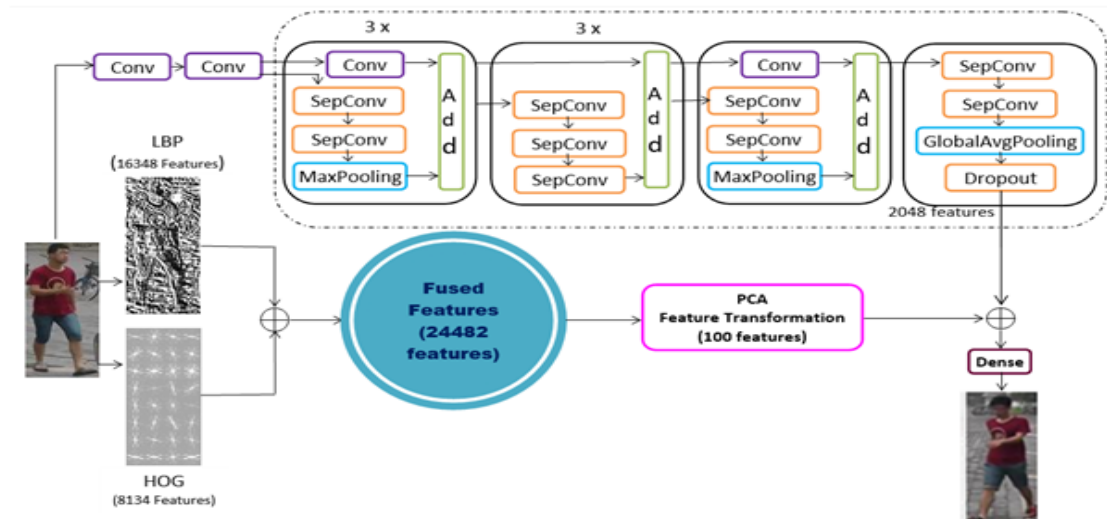


Fig. 6. ILFP Framework

These features are merged with the features obtained from PCA transformed combined features of LBP and HOG. LBP produces 16348 features, and the HOG algorithm gives 8134 features. LBP ensures the local texture-based features. HOG gives the texture information in that specified space. The integrated local information is prepared by fusing LBP and HOG features, which have 24482 features. These features are reduced to 100 features by PCA. The reduced feature combined with the global features forms the feature descriptor for the image. The proposed network identifies the person using the combination of both these features.

## 4. Experimental Results and Discussion

Person reidentification has a wide scope in the computer vision society in the current decade with various applications like video surveillance, pedestrian search, multi-camera tracking and behaviour analysis. Our experiments are executed on Google Colab with Nvidia GPU. Our network is trained in batches with size 32. The total epochs are set to 200 for Market1501, CUHK03 and NEC01 datasets. Adam optimiser is used with the base learning rate, which is

initialised to 0.001. The images are re-sized to 128 x 128 with normalisation, and it has 3 channels (RGB). This proposal has used two dropout layers with a dropout rate of 0.5 and 0.3, respectively.

#### 4.1. Datasets

There are quite a few numbers of publicly available datasets that various person reidentification algorithms have tested. Out of all benchmark datasets, the most used CUHK03 and Market1501 datasets have been employed with our system to evaluate its performance. In addition, we have employed our dataset NEC01. **Table 1** emphasises the details of datasets experimented on in the proposed work.

**Table 1.** Person Re-identification Datasets

Dataset	Year	No. of Boxes	No of Identity	No of Cameras	Detector	Scene	Crop Size	Evaluation Method
CUHK03	2014	13, 164	1, 467	5	DPM / Hand	Indoor	Varied	CMC
Market1501	2015	32, 688	1, 501	6	DPM / Hand	Outdoor	128x64	CMC+mAP
NEC01	2020	8053	1000	3	DPM / Hand	Indoor/ Outdoor	128x64	CMC+mAP

CUHK03 is a specific setup for each person images with three different partitions for the gathered images in a well-known recent dataset was provided by the Chinese University of Hong Kong (CUHK). It provides the bounding boxes detected from deformable part models (DPM) with manual labelling. CUHK03 has 1, 467 identities which are collected from 5 pairs of cameras and holds 13, 164 images. The dataset is stored in .mat file contains three cells as detected, labelled and testset. Detected means the bounding boxes, which are estimated by the pedestrian detector. Labelled means the bounding boxes that are labelled as human, and testset contains the testing protocols. **Fig. 7** shows the sample pair of images taken from CUHK03 dataset.



**Fig. 7.** Few Sample images of CUHK03

The Market1501 dataset is a dataset for person reidentification collected in front of a supermarket in Tsinghua University. Five cameras with high-resolution and one low-resolution camera are used. Cameras are arranged in overlapping with different cameras. This dataset contains 32, 668 annotated bounding boxes of 1, 501 identities. Each annotated identity is present in at least two cameras so that a cross-camera search can be performed.

The Market1501 dataset is annotated using the following rules. For each detected bounding box to be annotated, a bounding box with ground truth is annotated containing the

pedestrian. Then, the ratio of the overlapping area to the union area is calculated for the detected and hand-drawn bounding boxes. If the ratio is larger than 50%, the Deformable Part Model (DPM) bounding box is marked as good. If the ratio is lesser than 20%, the bounding box is marked as a distractor. If not, it is marked as junk, and the image has zero influence on the reidentification accuracy. **Fig. 8** shows the pair of sample images from the Market1501 dataset.



**Fig. 8.** Few Sample images of Market1501

Most of the benchmarked dataset is available with images of persons who do not belong to a particular region. So, it is planned to test native images with the proposed system. This paper utilises the institution's surveillance cameras fixed at various locations, and the dataset collection is done between Jan 2020 and Sep 2020. Images are captured using 3 disjoint cameras. It consists of 8053 images of both students and faculty members. **Fig. 9** gives the sample pair of images from NEC01 dataset.



**Fig. 9.** Few Sample images of NEC01

## 4.2. Performance Evaluation Metrics

Evaluation of the proposed system with rank- $i$  accuracy and mean average precision (mAP) is employed and tested with the dataset discussed in the earlier section. Here, rank- $i$  accuracy estimates the similarity between the query image and the subsequent gallery images. The similarity values are sorted in ascending order and organised in a table. Top- $k$  accuracy is estimated by checking the top- $k$  gallery images which contain the matched query image. If the correctly matched image is found in the  $i^{\text{th}}$  position of the retrieved list, then rank- $i$  is set as true. Rank can be verified at rank-1, 5, 10 etc.

Performance analysis is done by using Mean average precision (mAP) and Rank-1 accuracy. First, the Precision-Recall curve is drawn. The area under this curve is known as Average precision for that query. The average of all these APs is known as mAP. This measure is meant to be precise because it combines both the precision and recall of an algorithm. Let the query image be matched with 7 images from the gallery, and matched images are positioned at 1<sup>st</sup>, 4<sup>th</sup>, 5<sup>th</sup> and 6<sup>th</sup>, as shown in **Fig. 10**. The precision is marked below each matched figure. Average precision for query image is calculated as the average of precisions

of matched images. Thus, for the given image the average precision can be calculated as  $(1 + 0.5 + 0.6 + 0.66)/4$ . So, AP = 0.69 for this query image.

Similarly, average precision is computed for all query images. Mean Average Precision (mAP) is the average of all Average Precision estimated.



Fig. 10. Illustration of Average Precision Calculation

Besides, any classification process is evaluated with accuracy and loss of training and test dataset. These value for each dataset is plotted as a graph in Fig. 11, 12 and 13.

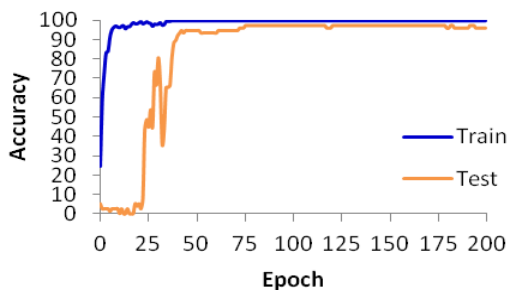
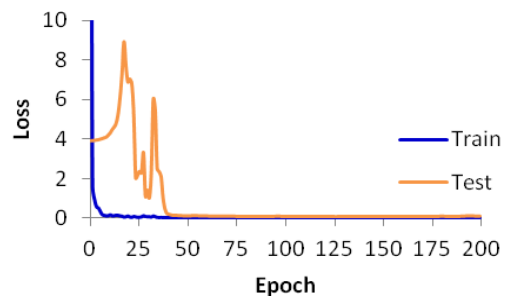


Fig. 11. a) Market1501 training and testing accuracy



b) Market1501 training and testing Loss

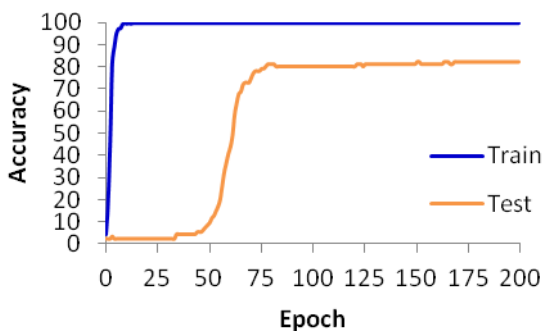
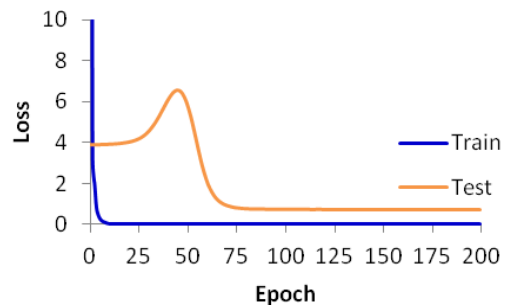
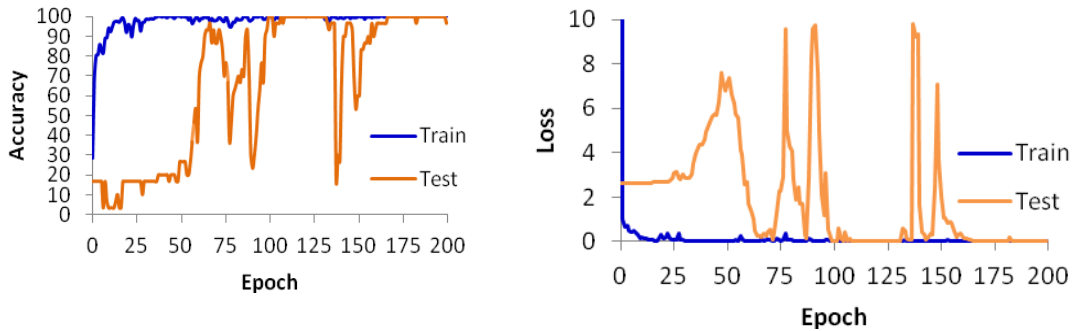


Fig. 12. a) CUHK03 training and testing accuracy



b) CUHK03 training and testing Loss



**Fig. 13. a)** NEC01 training and testing accuracy

**b)** NEC01 training and testing Loss

The present investigation recorded the results of different datasets discussed on various features ranging from 50 to 200. Experimentations were done for various existing methods as well. **Table 2** gives the comparison of different metrics with reference to the number of features on different datasets, and **Table 3** gives the comparison of different methods on different datasets. The highest Rank-1 accuracy is achieved on Market 1501 with 150 features. The mAP values reached the highest for NEC01 with 100 features. There is a decline in metric values after reaching 100 features almost on all datasets.

The state-of-art methods adapted to the dataset dealt showed the highest value of Rank-1 accuracy and mAP for combined HOG, LBP, and PCA with 100 features on the NEC01 dataset. For all other datasets also, this combination won with the highest Rank-1 accuracy and mAP metric. The graphical representation of both the results is depicted in **Fig. 14** and **Fig. 15**, respectively.

In Rank1 metric, HOG features with this network have produced 2.72%, 4.15%, and 4.75% than LBP for Market1501, CUHK03 and NEC01. The mAP metric also HOG performed with an increase of 2.3%, 6.17% and 1.11% than LBP.

HOG and LBP used the gradient around a pixel. So, we tried a combination of HOG and LBP features which obtained better results than HOG or LBP separately. 3.19%, 4.73% and 7.74 higher than LBP in terms of Rank1 and 1.16%, 0.56% and 2.85% increase in the HOG Rank1 metric on Market1501, CUHK03 and NEC01, respectively. Similarly, combined HOG and LGP results are better than LBP and HOG in the mAP metric also.

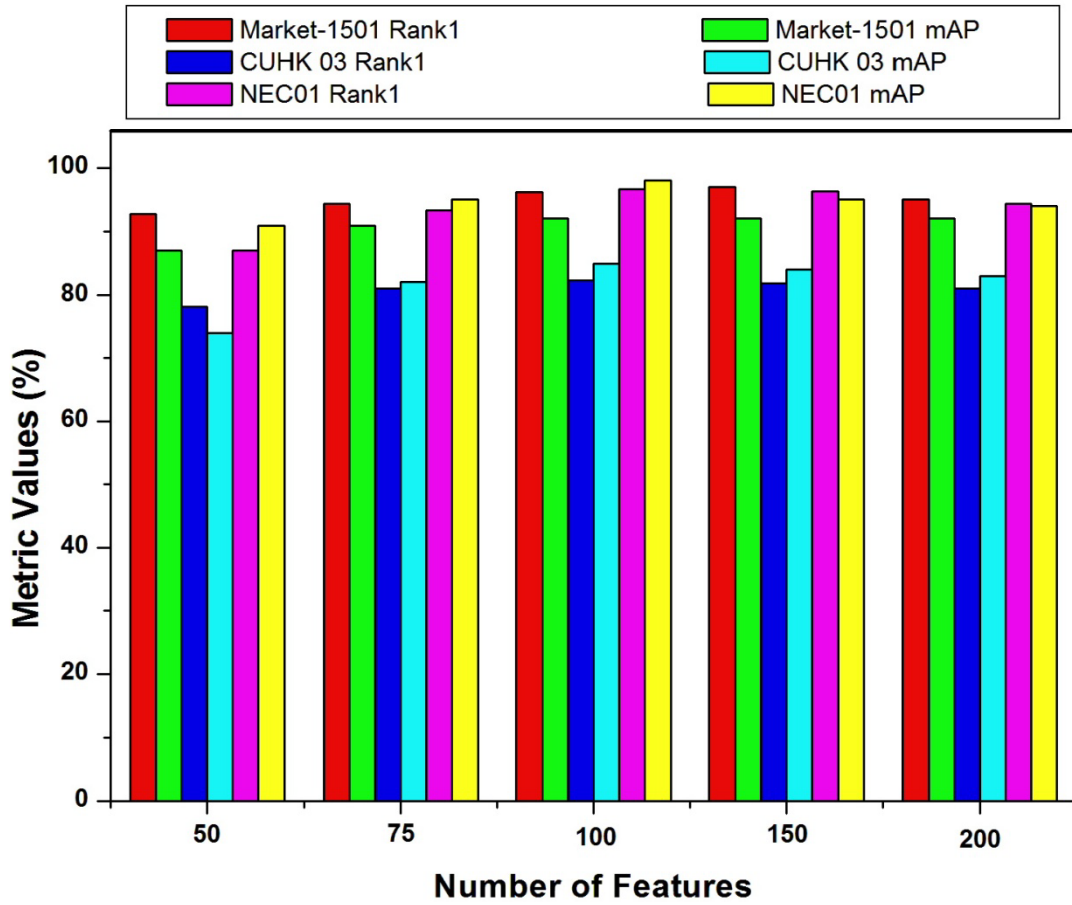
The proposed ILFP framework with 100 features yielded the highest result than all methods adapted earlier with our system on datasets. 1.47%, 0.96% and 2.67% higher than HOG and LBP feature combination in Rank1 metric. 1.1%, no change and 2.08% increase in mAP metric, respectively.

**Table 2.** Comparison of Different Metrics with reference to Number of Feature on Different datasets

Number of features	Market-1501		CUHK03		NEC01	
	Rank 1	mAP	Rank 1	mAP	Rank 1	mAP
50	92.72	87	78.15	74	87	91
75	94.31	91	80.99	82	93.27	95
100	96.15	92	82.29	85	96.66	98
150	96.92	92	81.82	84	96.32	95
200	95.02	92	81.02	83	94.29	94

**Table 3.** Comparison of Different Methods on Different datasets

Method	Market-1501		CUHK03		NEC01	
	Rank 1	mAP	Rank 1	mAP	Rank 1	mAP
m-XceptionNet + HOG	93.67	89	81.06	81	91.54	91
m-XceptionNet + LBP	91.19	87	77.83	76	87.39	90
m-XceptionNet + HOG + LBP	94.76	91	81.51	82	94.15	95
ILFP Framework (150 features)	95.87	91	81.74	85	94.98	96
ILFP Framework (100 features)	96.15	92	82.29	85	96.66	98



**Fig. 14.** Comparison of Different Metrics with reference to Number of Feature on Different datasets

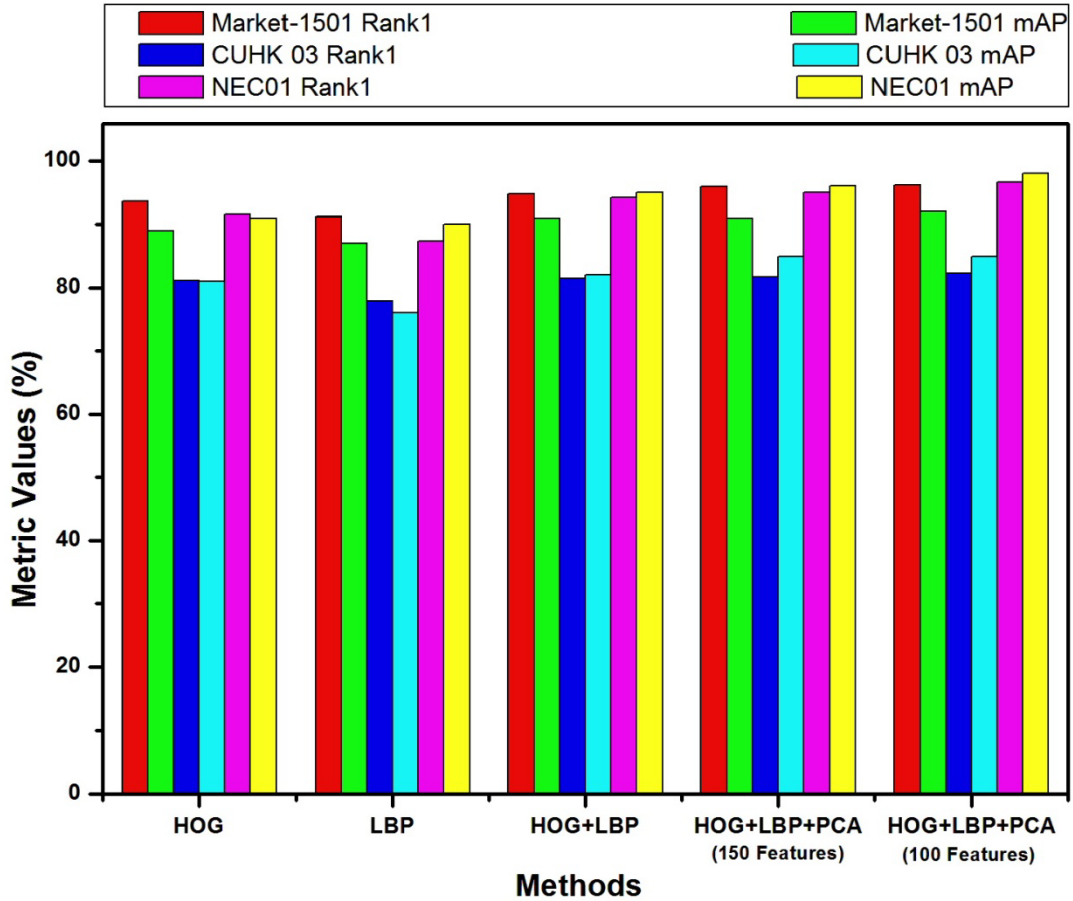


Fig. 15. Comparison of Different Methods on Different datasets

## 5. Conclusion

The present investigation proposed an Integrated Level Feature Pattern (ILFP) framework for person reidentification. The ILFP framework assimilates both local features and global features for re-identifying the persons. The spatial-oriented texture information in terms of local-level features is extracted by applying LBP and HOG. The fused features are optimised into 100 features using PCA. On the other hand, to obtain the global features, we propose a new CNN model as m-XceptionNet. This integrated framework extracts both pixel-level and overall features, which helps to get better performance over the other existing works. ILFP framework performed different features ranges of 50, 100, 150 and 200. In that, our ILFP with 100 features provides better Rank1 accuracy as 96.15%, 82.29% and 96.66% for Market1501, CUHK03, NEC01, respectively. The mean average precision (mAP) accuracy is highly appreciable as 92%, 85%, 98% for Market 1501, CUHK03, NEC01, respectively. Also, the proposed ILFP is compared with different methods, and the performance is plotted.

As future research, local features using Scale Invariant Feature Transform (SIFT) and Speeded Up Robust Features (SURF), Binary Robust Invariant Scalable Keypoints (BRISK) feature detector and descriptor can be detected. The global features are extracted using m-XceptionNet for the ILFP framework to improve the effectiveness of the person re-id process.

## References

- [1] A. Plantinga, "Things and persons," *The Review of Metaphysics*, vol.14, no.3, pp. 493–519, 1961. [Article \(CrossRef Link\)](#)
- [2] N. B. Cocchiarella, "Sortals, natural kinds and re-identification," *Logique et analyse*, vol. 20, pp. 439–474, 1977. [Article \(CrossRef Link\)](#).
- [3] A. O. Rorty, "The transformations of persons," *Philosophy*, vol. 48, no. 185, pp. 261–275, 1973. [Article \(CrossRef Link\)](#)
- [4] S. Gong, M. Cristani, S. Yan, and C. C. Loy, *Person reidentification*, Springer, 2014. [Article \(CrossRef Link\)](#)
- [5] F. Mamalet and C. Garcia, "Simplifying ConvNets for Fast Learning," in *Proc. of International Conference on Artificial Neural Networks*, Springer, pp. 58–65, 2012. [Article \(CrossRef Link\)](#)
- [6] J. Berclaz, F. Fleuret, and P. Fua, "Multi-camera tracking and atypical motion detection with behavioral maps," in *Proc. of 10<sup>th</sup> European Conference on Computer Vision*, pp. 112-125, 2008. [Article \(CrossRef Link\)](#)
- [7] X. Wang, K. Tieu, and W. Grimson, "Correspondencefree multi-camera activity analysis and scene modelling," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2008. [Article \(CrossRef Link\)](#)
- [8] X. Wang, K. T. Ma, G.-W. Ng, and W. E. L. Grimson, "Trajectory analysis and semantic region modeling using a nonparametric bayesian model," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2008. [Article \(CrossRef Link\)](#)
- [9] W. Ge and R. T. Collins, "Marked point processes for crowd counting," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2009. [Article \(CrossRef Link\)](#)
- [10] A. Chan and N. Vasconcelos, "Bayesian poisson regression for crowd counting," in *Proc. of IEEE 12<sup>th</sup> International Conference on Computer Vision*, pp. 545 –551, 2009. [Article \(CrossRef Link\)](#)
- [11] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localised features," in *Proc. of 10<sup>th</sup> European Conference on Computer Vision*, pp. 262-275, 2008. [Article \(CrossRef Link\)](#)
- [12] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person reidentification by symmetry-driven accumulation of local features," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010. [Article \(CrossRef Link\)](#)
- [13] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person reidentification by local maximal occurrence representation and metric learning," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2015. [Article \(CrossRef Link\)](#)
- [14] R. Zhao, W. Ouyang, and X. Wang, "Learning mid-level filters for person reidentification," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2014. [Article \(CrossRef Link\)](#)
- [15] I. Kviatkovsky, A. Adam, and E. Rivlin, "Color invariants for person reidentification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, pp. 1622–1634, July 2013. [Article \(CrossRef Link\)](#)
- [16] L. Ma, X. Yang, and D. Tao, "Person reidentification overcamera networks using multi-task distance metric learning," *IEEE Transactions on Image Processing*, vol.23, pp.3656-3670, August 2014. [Article \(CrossRef Link\)](#)
- [17] Z. Li, S. Chang, F. Liang, T. S. Huang, L. Cao, and J. R. Smith, "Learning locally-adaptive decision functions for person verification," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2013. [Article \(CrossRef Link\)](#)
- [18] W.-S. Zheng, S. Gong, and T. Xiang, "Reidentification by relative distance comparison," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 3, pp. 653-668, 2013. [Article \(CrossRef Link\)](#).



- [19] M. Koestinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2012. [Article \(CrossRef Link\)](#)
- [20] R. Zhao, W. Ouyang, and X. Wang, "Person reidentification by salience matching," in *Proc. of 2013 IEEE International Conference on Computer Vision*, 2013. [Article \(CrossRef Link\)](#).
- [21] G. Lisanti, I. Masi, A. Bagdanov, and A. D. Bimbo, "Person reidentification by iterative re-weighted sparse ranking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, pp. 1629–1642, August 2015. [Article \(CrossRef Link\)](#)
- [22] S. Khamis, C. Kuo, V. Singh, V. Shet, and L. Davis, "Joint learning for attribute-consistent person reidentification," in *Proc. of European Conference on Computer Vision workshop*, pp 134-146, 2014. [Article \(CrossRef Link\)](#)
- [23] F. Xiong, M. Gou, O. Camps, and M. Sznai, "Person reidentification using kernel-based metric learning methods," in *Proc. of European Conference on Computer Vision workshop*, pp 1-16, 2014. [Article \(CrossRef Link\)](#)
- [24] Z. Zhang, Y. Chen, and V. Saligram, "A novel visual word co-occurrence model for person reidentification," in *Proc. of European Conference on Computer Vision workshop*, pp 122-133, 2014. [Article \(CrossRef Link\)](#)
- [25] Rui, Y, Huang, T and Chang, S, "Image Retrieval: Current Techniques, Promising Directions and Open Issues," *Journal of Visual Communication and Image Representation*, vol. 10, pp.39–62, March1999. [Article \(CrossRef Link\)](#)
- [26] Michael J. Swain and Dana H. Ballard, "Colour indexing," *International Journal on Computer Vision*, vol. 7, pp. 11-32, November1991. [Article \(CrossRef Link\)](#)
- [27] Gevers, T and Stokman, H, "Classifying colour edges in video into shadow-geometry, highlight, or material transitions," *IEEE Transactions on Multimedia*, vol. 5, no. 2, pp. 237-243, July 2003. [Article \(CrossRef Link\)](#)
- [28] Guan, H and Wada, S, "Flexible colour texture retrieval method using multiresolution mosaic for image classification," in *Proc.of 6<sup>th</sup> International Conference on Signal Processing*, vol. 1, pp. 612-615, 2002. [Article \(CrossRef Link\)](#)
- [29] Shi, Dong-cheng, Xu, Lan, Han, Ling-yan, "Image retrieval using both color and texture features," *The Journal of China Universities of Posts and Telecommunications*, vol. 14, Supplement 1, pp 94-99, October 2007. [Article \(CrossRef Link\)](#).
- [30] Ouyang, A and Tan, Y, "A novel multi-scale spatial-colour descriptor for content-based image retrieval," in *Proc.of 7<sup>th</sup> International Conference on Control, Automation, Robotics and Vision*, vol. 3, pp. 1204-1209, 2002. [Article \(CrossRef Link\)](#)
- [31] Yu, H, Li, M, Zhang, H and Feng, J, "Colour texture moments for content-based image retrieval," in *Proc.of International Conference on Image Processing*, Rochester, New York, USA, vol. 3, pp. 929-932, 2002. [Article \(CrossRef Link\)](#)
- [32] Vadivel, A, SuralShamik and Majumdar, AK, "An integrated colour and intensity co-occurrence matrix," *Pattern Recognition Letters*, vol. 28, no.8, pp. 974–983, January 2007. [Article \(CrossRef Link\)](#)
- [33] Belongie, S, Malik, J and Puzicha, J, "Shape matching and object recognition using shape contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no.4, pp.509-522, April 2002. [Article \(CrossRef Link\)](#)
- [34] Scott, GJ, MathewKlaric, M, Davis, CH and Che-Ren, S, "Entropy-Balanced Bitmap Tree for Shape-Based Object Retrieval From Large-Scale Satellite Imagery Databases," *IEEE Transactions on Geoscience and Remote Sensing*, vol.49, no.5, pp.1603-1616, May 2011. [Article \(CrossRef Link\)](#)
- [35] Shechtman, E and Irani, M, "Space-time behavior-based correlation," in *Proc. of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005. [Article \(CrossRef Link\)](#)
- [36] Manjunath, BS and Ma, WY, "Texture Features for browsing and Retrieval of Image Data," *IEEE Transactions on Pattern Analysis Machine Intelligence*, vol.18, no.8, pp.837-842, August 1996. [Article \(CrossRef Link\)](#)

- [37] Nouredine, A, "Computational Perceptual Features for Texture representation and Retrieval," *IEEE Transactions on Image Processing*, vol.20, no.1, pp.236-246, January 2011. [Article \(CrossRef Link\)](#)
- [38] Ojala, T, Pietikainen, M and Harwood, D, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, January1996. [Article \(CrossRef Link\)](#)
- [39] Heikkila, M, Pietikainen, M and Schmid, C, "Description of interest regions with local binary patterns," *Pattern Recognition*, vol. 42, no. 3, pp. 425-436, March 2009. [Article \(CrossRef Link\)](#)
- [40] N. Martinel, C. Micheloni, and G. Feresti, "Saliency weighted features for person re-identification," in *Proc. of ECCV Workshop on Visual Surveillance and Re-identification*, pp.191-208, 2014. [Article \(CrossRef Link\)](#)
- [41] W. Li, R. Zhao, T. Xiao, and X. Wang, "Deepreid: Deep filter pairing neural network for person reidentification," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 152–159, 2014. [Article \(CrossRef Link\)](#)
- [42] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84-90, 2017. [Article \(CrossRef Link\)](#)
- [43] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587, 2014. [Article \(CrossRef Link\)](#)
- [44] F. Radenović, G. Toliás, and O. Chum, "CNN image retrieval learns from bow: Unsupervised fine-tuning with hard examples," in *Proc. of European Conference on Computer Vision*, pp.3-20, 2016. [Article \(CrossRef Link\)](#)
- [45] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 815–823, 2015. [Article \(CrossRef Link\)](#)
- [46] E. Ahmed, M. Jones, and T. K. Marks, "An improved deep learning architecture for person reidentification," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3908–3916, 2015. [Article \(CrossRef Link\)](#)
- [47] L. Wu, C. Shen, and A. v. d. Hengel, "Personnet: Person reidentification with deep convolutional neural networks," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, *arXiv:1601.07255*, 2016. [Article \(CrossRef Link\)](#)
- [48] R. R. Varior, B. Shuai, J. Lu, D. Xu, and G. Wang, "A siamese long short-term memory architecture for human reidentification," in *Proc. of European Conference on Computer Vision*, 2016. [Article \(CrossRef Link\)](#)
- [49] T. D'Orazio and G. Cicirelli, "People reidentification and tracking from multiple cameras: a review," in *Proc. of 19<sup>th</sup> IEEE International Conference on Image Processing*, pp. 1601–1604, 2012. [Article \(CrossRef Link\)](#)
- [50] A. Bedagkar-Gala and S. K. Shah, "A survey of approaches and trends in person reidentification," *Image and Vision Computing*, vol. 32, no. 4, pp. 270–286, 2014. [Article \(CrossRef Link\)](#)
- [51] R. Satta, "Appearance descriptors for person reidentification: a comprehensive review," *arXiv preprint, arXiv:1307.5748*, July, 2013. [Article \(CrossRef Link\)](#)
- [52] X. Wang, "Intelligent multi-camera video surveillance: A review," *Pattern recognition letters*, vol. 34, no. 1, pp. 3–19, January2013. [Article \(CrossRef Link\)](#)
- [53] Navneet Dalal and Bill Triggs, "Histograms of Oriented Gradients for Human Detection," in *Proc. of IEEE computer society conference on Computer Vision and Pattern Recognition*, vol.1, pp. 886-893, 2005. [Article \(CrossRef Link\)](#)
- [54] Y. LeCun, L. Jackel, L. Bottou, C. Cortes, J. S. Denker, H. Drucker, I. Guyon, U. Muller, E. Sackinger, P. Simard, et al., "Learning algorithms for classification: A comparison on handwritten digit recognition," *Neural networks: the statistical mechanics perspective*, pp. 261-276, 1995. [Article \(CrossRef Link\)](#)

- [55] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint, arXiv:1409.1556*, September 2014. [Article \(CrossRef Link\)](#)
- [56] S. Ioffe and C. Szegedy, "Batch normalisation: Accelerating deep network training by reducing internal covariate shift," in *Proc. of 32<sup>nd</sup> International Conference on Machine Learning*, pp. 448–456, 2015. [Article \(CrossRef Link\)](#)
- [57] C. Szegedy, V. Vanhoucke, S.Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," *arXiv preprint arXiv:1512.00567*, 2015. [Article \(CrossRef Link\)](#)
- [58] C. Szegedy, S. Ioffe, and V. Vanhoucke, "Inception-v4, inception-resnet and the impact of residual connections on learning," *arXiv preprint arXiv:1602.07261*, 2016. [Article \(CrossRef Link\)](#).



**V.MANIMARAN** is working as an Assistant Professor (Senior Grade) in the Department of Information Technology at National Engineering College, Kovilpatti, Tamilnadu, India. He has received his B.Tech., (IT) degree from Anna University and M.E., (CSE) degree from Manonmaniam Sundaranar University. His area of interest are Image processing and Data Science. He has published more than 5 Journal and 15 International conference publications. He is a Life Member in Indian Society for Technical Education.



**K.G.SRINIVASAGAN** received his B.E and M.E degree both in Computer Science and Engineering from Madurai Kamaraj University, Madurai, India and Bharathiar University, Coimbatore, India in 1989 and 1998 respectively. He received his PhD degree from the Manonmaniam Sundaranar University in 2009. Currently he is working as a Dean and Head in the Department of Information Technology. His area of research is in Image processing and Network security. He has published more than 25 Journal and International conference publications.



**S.GOKUL** has completed his Bachelor of Engineering in Computer Science and Engineering in 2020 at Anna University, Guindy. He is currently pursuing his Masters at Saarland University, Germany. His area of specialization includes Computer Vision, Image processing and visualization, Machine learning algorithms, Deep learning architectures, Pattern Recognition.



**I. JEENA JACOB** has received her Bachelor of Engineering in Computer Science and Engineering in 2004, received her Master of Engineering in Computer Science and Engineering in 2006 and received her Doctor of Philosophy in 2015. Her research areas are Pattern Recognition and Machine Learning. Currently she is working as an Associate Professor in GITAM University, Bengaluru Campus, Karnataka, India.



**S.BABU RENGHA RAJAN** is presently working as Professor in the Computer Science and Engineering Department at PET Engineering College, Vallioor Tamilnadu, India. Graduated in “Computer Engineering” at National Engineering College, Kovilpatti. Completed his Masters in Computer Science at Birla Institute of Technology and Science, Pilani. He also received his PhD degree from the Manonmaniam Sundaranar University in 2017. Areas of interest are Data Mining, Text Mining, Software Quality and Software Engineering. He is a Fellow Member of Institution of Engineers (India) and Life Member in Indian Society for Technical Education.