

Received 24 September 2022, accepted 24 October 2022, date of publication 31 October 2022, date of current version 8 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3218322

RESEARCH ARTICLE

Clothes Retrieval Using M-AlexNet With Mish Function and Feature Selection Using Joint Shannon's Entropy Pearson's Correlation Coefficient

MARRYAM MURTAZA¹, MUHAMMAD SHARIF¹, (Senior Member, IEEE),
MUSSARAT YASMIN¹, MUHAMMAD FAYYAZ²,
SEIFEDINE KADRY^{3,4,5}, (Senior Member, IEEE),
AND MI YOUNG LEE⁶, (Member, IEEE)

¹Department of Computer Science, COMSATS University Islamabad, Wah Campus, Islamabad 47040, Pakistan

²Department of Computer Science, FAST—National University of Computer and Emerging Sciences, Chiniot-Faisalabad Campus, Islamabad 44000, Pakistan

³Department of Applied Data Science, Noroff University College, 4612 Kristiansand, Norway

⁴Artificial Intelligence Research Center (AIRC), College of Engineering and Information Technology, Ajman University, Ajman, United Arab Emirates

⁵Department of Electrical and Computer Engineering, Lebanese American University, Byblos 1701, Lebanon

⁶Department of Software, Sejong University, Seoul 30035, South Korea

Corresponding authors: Marryam Murtaza (marryam85@gmail.com) and Mi Young Lee (miylee@sejong.ac.kr)

This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant 2021R111A1A01055652.

ABSTRACT The online retrieval of clothes-related images is crucial because finding the exact items, like the query image from a large amount of data, is highly challenging. However, significant clothes image variations degrade visual search retrieval accuracy. Another problem with retrieval accuracy is the high dimensions of feature vectors obtained from pre-trained deep CNN models. This research aims to enhance clothes retrieval training and test accuracy by using two different means. Initially, features are extracted using the modified AlexNet (M-AlexNet) with slight modification. The ReLU activation function is replaced with a self-regularized Mish activation function because of its non-monotonic nature. The M-AlexNet with Mish is trained on the CIFAR-10 dataset using the SoftMax classifier. Another contribution is to reduce the dimensions of feature vectors obtained from M-AlexNet. The dimensions of features are reduced by selecting the top k -ranked features and removing some of the different features using the proposed Joint Shannon's Entropy Pearson Correlation Coefficient (JSE-PCC) technique to enhance the clothes retrieval performance. To calculate the efficacy of suggested methods, the comparison is performed with other deep CNN models such as baseline AlexNet, VGG-16, VGG-19, and ResNet50 on DeepFashion2, MVC, and the proposed Clothes Image Dataset (CID). Extensive experiments indicate that AlexNet with Mish attains 85.15%, 82.04%, and 83.65% accuracy on DeepFashion2, MVC, and 83.65% on CID datasets. Hence, M-AlexNet and the proposed feature selection technique surpassed the results with a margin of 5.11% on DeepFashion2, 1.95% on MVC, and 3.51% on CID datasets.

INDEX TERMS Clothes retrieval, feature ranking, JSE-PCC technique, Pearson's correlation, Shannon's entropy.

I. INTRODUCTION

In the current COVID-19 pandemic since 2019, rapid economic disasters created surprising challenges for the

The associate editor coordinating the review of this manuscript and approving it for publication was Maurizio Tucci.

fashion industry [1]. Moreover, with the shutter of physical e-commerce businesses, people's preferences for physical to online shopping are highly increased. Despite this, online clothing sites catch the attention of targeted customers through model posters to visualize clothes with different view angles or in the form of appealing video advertisements [2].

However, it is tough to retrieve desired clothes from various image collections, hence it is challenging to retrieve fast fashion products effectively [3]. To speed up the retrieval process, fashion recommendation plays a vital role in suggesting similar or mixed and matching items according to the customer's preferences.

Generally, the retrieval methods involve feature extraction, feature selection, and feature similarity measurement. Optimization at each stage is required to strengthen the performance of the retrieval process. An appropriate clothes feature descriptor is needed in the first feature extraction stage. Automatic features are extracted through different deep feature models like AlexNet, VGG, GoogleNet, ResNet, Xception series, etc. [4]. These models are proven to be very popular and influential in different domains of computer vision. The appropriate and effective linear/nonlinear activation functions can help boost these models' performance [5]. Therefore, the choice of a non-saturated activation function plays a considerable role in increasing the effectiveness of the networks. ReLU is one of the most imperative, and straightforward, and is a commonly used activation function that provides sparsity and is computationally effective as compared to other functions like sigmoid or Tanh functions [6]. AlexNet is one of the renowned CNN models for image classification tasks because it reduces the error rate to 15.3% [7].

The widely used AlexNet contains 25 layers which profoundly impact image classification with the function of ReLU. However, the problem with ReLU is that it suffers from dying ReLU and is non-differentiable at 0 [8]. To overcome this problem, many activation functions with non-monotonic nature have been introduced in this modern age, like Swish or Mish, that replace ReLU due to their utmost advantages over ReLU [5]. Similarly, in the second stage of the retrieval process, the optimal features of clothes are required. Appropriate feature selection techniques are used to calculate the most relevant features through ranking while reducing the dimensions from the feature space. Finally, in the last stage, the similarity score between target and gallery images is calculated using a distance threshold to find the candidate image. More training phases are required to reach the optimal value.

The motivation behind this research is to increase the training and test accuracy of the clothes retrieval process. The progress of retrieval not only depends on the feature extraction but also on the score of each extracted feature [9]. Initially, the features are extracted through the baseline AlexNet with a little bit of modification i.e., the ReLU layer is replaced with a self-regularized Mish activation function [10]. The AlexNet is used as a test model because of its simplicity and breakthrough results in various computer vision applications. Similarly, Mish has a property of regularization with maximum information storage capacity that helps to improve retrieval accuracy.

Another contribution of this research is to reduce the features relevant to the retrieval process. The problem with high dimension features increases the number of parameters that

helps to train well but is computationally expensive [11]. To reduce the size of high-dimensional feature vectors, the proposed feature selection operator JSE-PCC is designed to calculate the rank of each feature. Ranking of features also improves clothes retrieval performance. Finally, various feature-matching techniques are used to get the retrieval results. The experiments are conducted on DeepFashion2, MVC, and the proposed Clothing Image Dataset (CID). The comparison is performed with other pre-trained models to show the efficacy of the proposed model.

The significant involvement of this research article is discussed as under:

- (1) Extraction of features from AlexNet with selfregularized Mish instead of ReLU activation function.
- (2) Fine-tuning of M-AlexNet with different learning rates and dropout values on CIFAR-10 dataset.
- (3) Optimized the feature vectors obtained from M-AlexNet using proposed JSE-PCC technique for feature ranking and dimension reduction. Finally, the retrieval results are obtained by applying five different matching techniques.

The rest of the article is structured as follows. The relevant existing studies are outlined in Section 2. In Section 3, the feature extraction process using AlexNet with Mish activation function is discussed in detail. In detail, experiments are discussed and tabulated under Section 4. In Section 5, the comparison of the suggested model is performed with other off-the-shelf pre-trained deep CNN models. Finally, future directions and conclusion is drawn based on experimental results and analysis in Section 6.

II. RELATED WORK

Clothes retrieval is very important in e-commerce applications, online shopping, person re-identification, monitoring of person behavior, etc. Modern study generally hits this task through deep learning that endeavors outstanding breakthroughs in a viable way. Likewise, clothing retrieval reveals promising results in fashion outfits analysis using deep learning [12]. Understanding outfit attributes are very critical because of their compact and complex representation [13]. The unexpected need for clothes retrieval applications has triggered researchers to deal with some robust, efficient, and effective image retrieval methods. The focus of this literature study is only based on two stages, involved in clothes retrieval process i.e., feature extraction techniques, and feature selection/ optimization techniques, which are discussed as under.

A. EXISTING FEATURE EXTRACTION TECHNIQUES AND ROLE OF ACTIVATION FUNCTIONS

Initially, the feature extraction using pre-trained deep CNN architectures are the most superlative model for understanding the complex and dense contents of an image and have shown outstanding performance in detection, classification, segmentation, and image retrieval tasks [14], [15]. In the literature study, this manuscript only highlights the efforts and contributions of deep models in image retrieval applications.

Normally there are four families of deep convolutional networks with extensive parameterization which serve in the fashion image retrieval tasks naming AlexNet [7], GoogLeNet [16], VGG [17], and ResNet [18]. Depending on the problem domain, the performance of retrieval accuracy varies with extracted features. Similarly, the performance of feature extraction depends on the following factors appropriate selection of deep models, hyper-parameters tuning, selection of activation functions, etc. [19]. Here, we only discuss the role of activation functions to enhance retrieval accuracy. The modern study of Yuxin et. al. [5] indicates the importance of various activation functions in classification tasks and reveals that LeNet architecture with LeakyReLU outperforms ReLU and Swish activation functions with a 98.23% accuracy rate. Similarly, by using AlexNet it shows outstanding performance by using ReLU but the accuracy rate is lower when using PReLU and Swish, as compared to Mish activation function. According to Philip et. al. [20] Mish and Swish are the latest functions and perform better as compared to ReLU. Inspired by these functions, a novel self-optimized Phish activation function is introduced with the property of non-discontinuity in the differentiated graph.

Safa et al. [21] reduces 51% parameters of pre-trained LeNet-5 architecture just by replacing it with an improved, self-regularized activation function, SigmaH, and attains a 98.25% accuracy rate. Dabal et al. [22] perform a comparison of various nonlinear activation functions of Sigmoid, ReLU, LReLU, ELU, and SELU by varying learning rates on MNIST dataset. Shiv et al. [23] presented a comprehensive survey of different activation functions and provides a platform for different researchers keeping in view to effectively transform nonlinear inputs into more linear feature vectors. Zheng et al. [24] speed up and enhances the image classification process by introducing a proposed activation function FELU over ReLU, ELU, SLU, MPELU, and TReLU. Shiv et al. [25] shows highest face retrieval accuracy of 98.97% by using Average Biased ReLU function on PolyUNIR dataset. There are lots of other proven research evidence [26], [27], [28] that reveals improvements in accuracy rate just by replacing with appropriate activation functions, like Bekir et al. [29] shows 95% accuracy in MLP architecture when using Tanh functions, Cancan et al. [30] shows outstanding performance in Generative Antagonism Network when using the modern activation function of Mish.

B. EXISTING FEATURE SELECTION AND OPTIMIZATION TECHNIQUES

The extensive size of deep CNN can be trained on million and trillions of images, so the dimensions of extracted features are extremely high. To reduce feature dimensions, deep feature selection and feature enhancement are the most commonly used techniques. These techniques are applied to the straight-forward features obtained at last FC layers. For the retrieval tasks, FC layers are replaced to reduce dimensions using feature selection techniques like PCA, Linear Discriminant Analysis (LDA), SVM, etc. A large amount of connections

of neurons at an FC layer leads to the following limitations i.e., the absence of location of objects and local geometric invariance [31]. If an input is the patches of an image, then it has more capability to retain spatial information. Similarly, the features extracted from convolutional layers assemble spatial information and hence preserve the structural details; hence it is beneficial for instance-level retrieval tasks. Sum-pooling convolutional features (SPoC), BoW model, Vector of Aggregate Locally Descriptor (VLAD), etc. are the latest techniques to obtain compact descriptors.

Feature fusion [32] is another vanilla approach to better characterize an image and also reduces the feature dimensions. Zheng et al. [33] used Caffe and AlexNet model to automatically extract features for retrieval application and concludes that the deep feature fusion improves the retrieval results. Xishan et al. [34] exploit the hybrid multi-label CNN with SVM classifier to find out the relation between fine-grained clothes attributes with the corresponding spatial information. Alhassan et al. [35] used the feature fusion technique to merge color and texture using Hue-Saturation-Value (HSV) color moments and Gabor filters and got better accuracy in the top 10 and top 20 retrieval images. Alkhwilani et al. [36] produced image signatures using SIFT and SURF and enhanced the retrieval process using BoVW as a feature descriptor with an 88% precision rate on ALOI dataset. Alsmadi et al. [37] amalgamated color, shape, and texture features using color histogram, neutrosophic clustering algorithm, and GLCM. This combination has the strong capability to discriminate between features. Jabeen et al. [38] performed fusion using BovW and merged SURF and retina-inspired fast descriptors with an image retrieval precision rate of 86% on Corel-1000 dataset. Feature fusion can be performed using correlation analysis to calculate the correlation between feature vectors.

The most common correlation techniques are Discriminant Correlation Analysis (DCA), Canonical Correlation Analysis (CCA), Orthogonal Canonical Correlation Analysis (OCCA), and Locality preserving CCA (LPPCA), etc [39]. Haghghat et al. [40] used DCA to reduce Small Sample Size (SSS) problem and achieved the recognition rate of 99.60% on iris and fingerprint modalities. Other feature fusion operators are Weighted Averaging (WA) operator, Ordered Weighted Averaging (OWA) operator, Intuitionistic Fuzzy Weighted Averaging (IFWA) operator, Intuitionistic Fuzzy Ordered Weighted Averaging (IFOWA) operator, and Weighted Directed Graph [41], etc.

In the existing retrieval methods, each stage of the Query-Based Image Retrieval (QBIR) system needs enhancement to further improve the accuracy of the retrieval process. The individual deep CNN approaches revealed encouraging results, but they are still encountered with conventional issues [42], for example, discriminative feature representation [43], lower retrieval accuracy rate [44], etc. The obtained features are suitable for large variations in clothes fine-grained features, minor variations in apparel styles, and more prevalent feature representation capabilities of a

given image. In addition, a feature fusion effectively characterizes the image information for the retrieval process [45] but still there needs to design a robust method that effectively retrieves similar clothing products with maximum accuracy output. The accuracy of the suggested retrieval method is improved using the feature selection technique for dimensionality reduction, optimal feature selection, and compact representations. For retrieval of similar clothes, the focus is to select the optimal feature subsets from obtained features based on their maximum correlation values. Furthermore, the goal is to offer distinctive feature representation which helps to contribute to gallery search optimization and improves the retrieval rate at different ranks.

III. PROPOSED METHOD

This section provides a novel method to improve the training and test accuracy of the apparel retrieval algorithm. It is comprised of three modules 1) feature extraction, 2) feature selection and 3) feature matching. The first module includes feature extraction using M-AlexNet. Initially, the features are extracted through the baseline AlexNet with a little bit of modification i.e., the ReLU layer is replaced with a self-regularized Mish activation function [10] and is pre-trained on CIFAR-10 dataset. AlexNet is used as a test model because of its simplicity and breakthrough results in various computer vision applications. Similarly, Mish has a property of regularization with maximum information storage capacity that helps to improve retrieval accuracy. In the second module, an appropriate set of features are selected using the proposed JSE-PCC technique through feature ranking.

The progress of retrieval not only depends on the feature extraction but also on the score and size of each extracted feature [9]. Usually, the problem with high dimension features is that it increases the number of parameters that helps to train well but is computationally expensive [11]. To reduce the size of high-dimensional feature vectors, the proposed feature selection operator JSE-PCC is designed to calculate the rank of each feature. Ranking of features also improves apparel retrieval performance. Similarly, in the last module, feature matching techniques are applied to get the retrieval results. The proposed algorithm is evaluated on various existing as well as proposed CID datasets.

The major involvement is discussed as under:

1. Initially, the features are extracted using M-AlexNet with Mish activation function and pre-training the model on CIFAR-10 dataset.
2. Optimized the feature vectors obtained from MAlexNet using JSE-PCC technique for feature ranking and dimension reduction.
3. Retrieval results are obtained by applying five different matching techniques. The performance of the suggested model is analyzed on DeepFashion2, MVC, and the proposed CID dataset.

Fig. 1 demonstrates the functionality of each module graphically.

A. M-ALEXNET WITH MISH ACTIVATION FUNCTION AND PRE-TRAINING THE MODEL ON CIFAR-10 DATASET

AlexNet is the foremost CNN architecture that competes with ILSVRC challenge 2012 [7] and showed cutting-edge performance in image classification and recognition tasks. It is extensively used in transfer learning [46]. In this paper, AlexNet architecture is used for feature extraction. The motivation for selecting this model is because of its simple structure, and the proposed strategy is tested on its straightforward architecture. The baseline AlexNet shows outstanding performance by minimizing the top-5 error rate up to 15.3%.

The original AlexNet architecture contains 25 layers comprised of five conv layers, three FC layers, three pooling layers, seven ReLU layers, and two dropouts and SoftMax layers. In this research, our first method provides a modification to the original model by just replacing the ReLU with a self-regularized Mish activation function to test the clothes retrieval performance. ReLU is a frequently used function because of its simple and fast convergence rate of the training process but ReLU has the dying ReLU problem in which some of the neurons die because it discards the negative values and is non-differentiable at zero [47]. Mish has the property of self-regularization that not only overcomes the problem of ReLU and dying ReLU but also encourage the information storage capacity due to its non-monotonic nature. Another advantage of using Mish is that it is unbounded above means it is differentiable at any value and also introduces regularization because of the property of bounded below[48].

Fig. 2 portrays the architectural view of M-AlexNet model. The input accepted is identical to that of the previous AlexNet i.e., $227 \times 227 \times 3$. As shown in Fig. 2 the initial layer accepts the input named I, then used the first convolutional layer C1. After this, the ReLU layer is replaced with Mish layer M, and P1 is the pooling layer. The second Conv layer is represented as C2, followed by the Mish activation function, and normalizes the activation using Cross Channel Normalization layer CN, then P2 is the pooling layer. The third Conv layer is represented as C3, followed by modified layer M which is replaced by ReLU layer.

Similarly, C4 is the fourth convolutional layer, then the modified layer M. C5 is the fifth convolutional layer followed by pooling layer P5. The FC layers are FC6, FC7, and FC8 but in the modified version of AlexNet only FC6 features are considered in the retrieval process. The output layer SoftMax layers are removed as this is used for retrieval application instead of classification.

In short, in the original architecture, there were 7 ReLU layers which we have replaced into 7 Mish activation functions to increase the information capacity. The features are extracted at FC6, FC7, and FC8 layers but the experimental results, discussed in section 4, show that the FC6 features are more robust and show high efficacy in the retrieval result [33], [34]. In the proposed method, only FC6 features are optimized using the feature selection technique

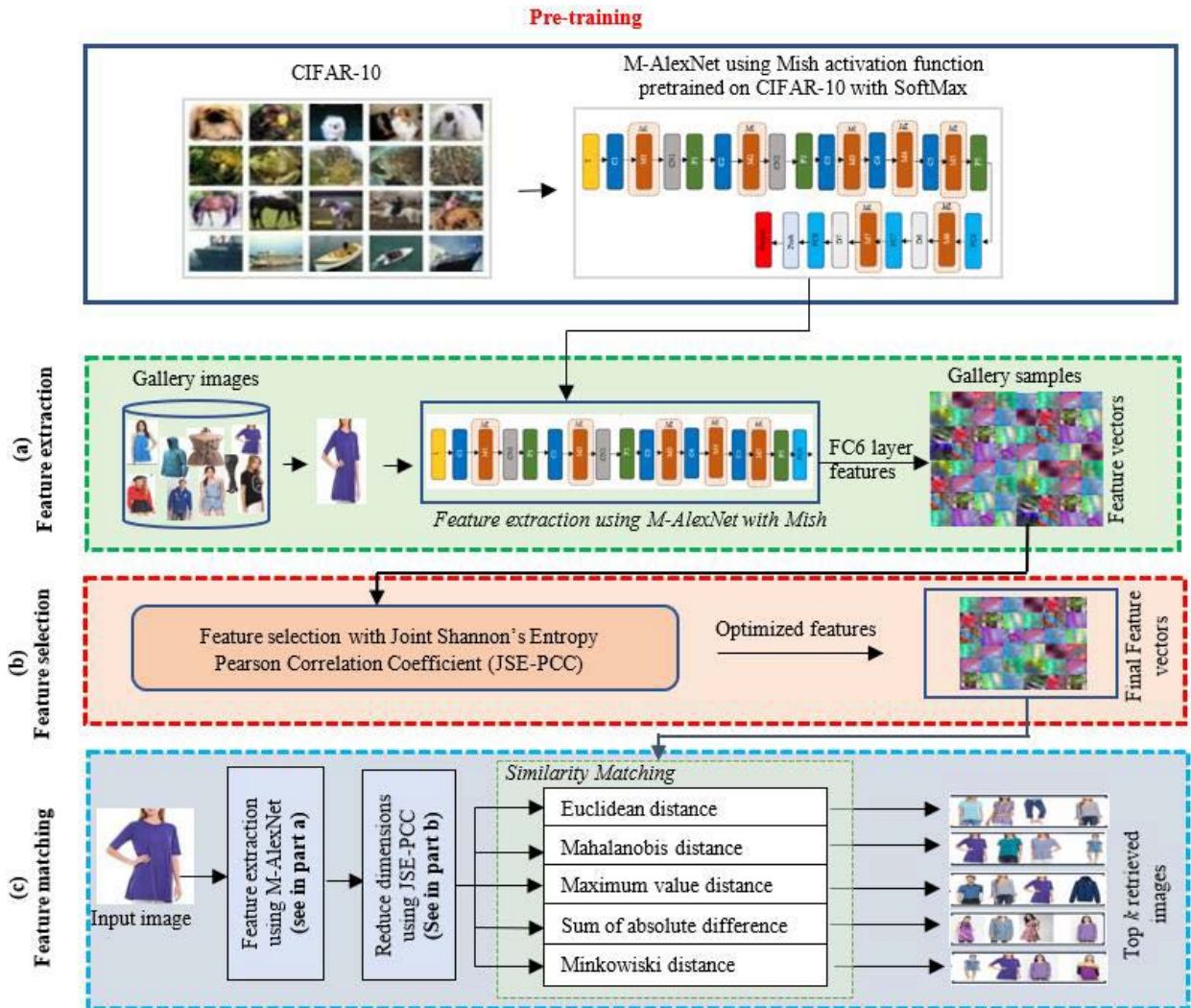


FIGURE 1. The proposed framework of clothes retrieval includes pre-training of AlexNet with Mish Activation Function on CIFAR-10 Dataset. The proposed model is further consisting of three segments (a) Feature extraction phase is useful to calculate the feature vectors using fewer layers of M-AlexNet (b) Feature selection is performed to compute the optimal feature vector using the proposed JSE-PCC Technique and (c) Feature matching techniques are applied between the gallery and input image to obtain the retrieval result.

for reducing the feature dimensions. The parameters of M-AlexNet with learnable are depicted in Table 1.

The training of the M-AlexNet with Mish activation function on the limited size of the dataset is not large enough to support the deep learning models. Therefore, the M-AlexNet is initially trained on the CIFAR-10 dataset which comprises of 10 classes and there are 6000 images per class, so the entire size of the dataset is 60000 which we split into 50000 training images and 10000 test images [35]. Here the model is pre-trained using the k-fold cross-validation method and training is performed on 500 images per category using the SoftMax classifier.

1) FEATURE EXTRACTION USING PRE-TRAINED M-ALEXNET WITH MISH AND VISUALIZATION OF EXTRACTED FEATURES
After transforming the M-AlexNet model into the pre-trained model on the CIFAR-10 dataset, the features from clothes images are extracted from different clothes image datasets.

The original model contains 25 layers, but M-AlexNet uses 17 layers for feature extraction by deleting the last two FC layers and the SoftMax layer in the retrieval process. The feature vectors are calculated from the FC6 layer. The FC7 and FC8 don't take part in the retrieval process because the performance of the FC6 layer is superior as compared to the FC7 and the FC8 layers.

Processing the clothes features is a nonlinear retrieval problem, and it is extremely challenging to learn complex deep learning models. The visualization of high-level features of clothes input images obtained at FC layers of different models i.e., M-AlexNet, baseline AlexNet, VGG-16, VGG-19, and ResNet50 is envisioned in Fig. 3, where the deep feature visualization gives insight that how features are specifically learned at each layer and which features take part in image retrieval. As shown in Fig. 3, part (a), the FC6 features of M-AlexNet provide a better understanding of features, as compared to the rest of the deep CNN models. Similarly,

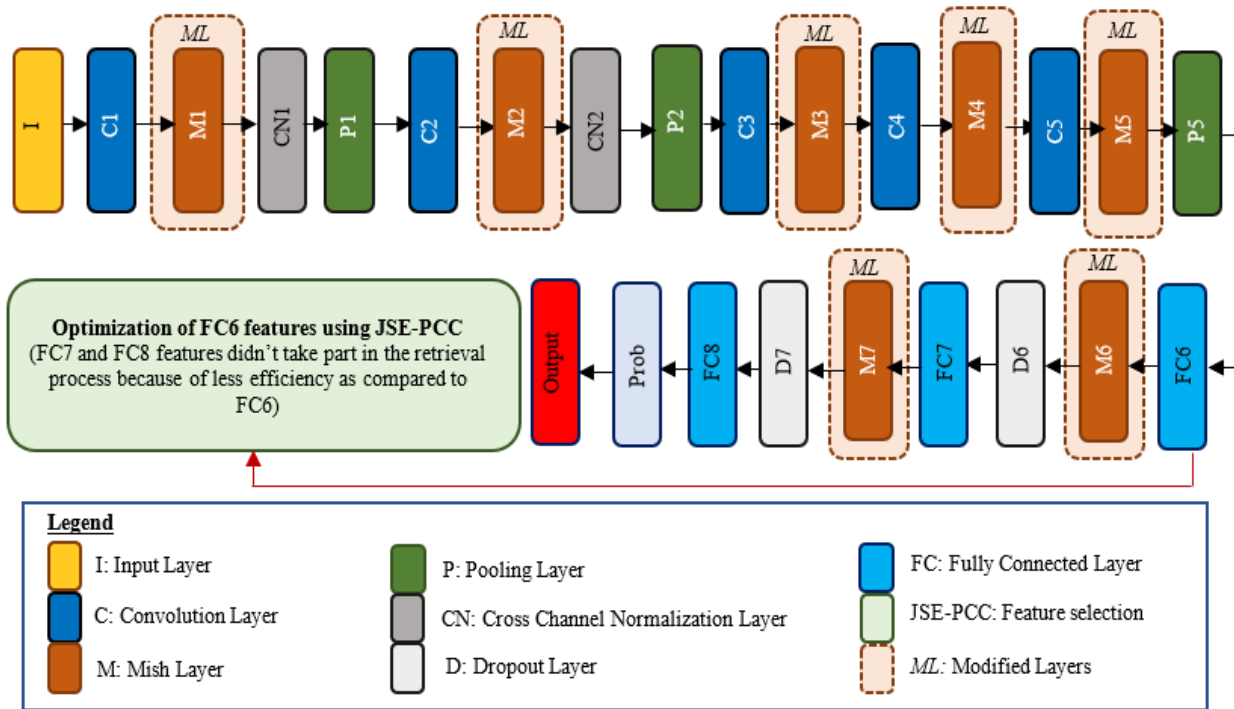


FIGURE 2. Architectural view of AlexNet with Mish activation function.

in part (b), the baseline architecture of the AlexNet model provides more robust features as compared to M-AlexNet. The clothes retrieval applications don't require detailed information and generally rely on the style and color features of the clothes, which are best described through the M-AlexNet rather than the baseline AlexNet architecture. The FC6 features of VGG-16 and ResNet50 in parts (c) and (e) are difficult to interpret hence affecting the retrieval performance. Similarly, VGG-19 in part (d) is better in comparison to VGG-16 model.

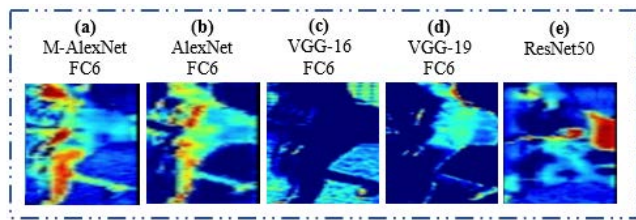


FIGURE 3. Comparison of deep features visualization of M-AlexNet and pre-trained AlexNet, VGG-16, VGG-19, and ResNet50 models.

B. FEATURES OPTIMIZATION USING JSE-PCC

The main problem of deep models is that they only accept a massive amount of data to learn features. The extracted features using the deep models are huge in number considering the case of AlexNet, VGG-16, and VGG-19, the dimension of the feature vector for each model is 4096 against each image. If there are say N images in the dataset then the size of the feature vectors is like $N \times 4096$, similarly, the amount of the feature vector for ResNet50 is 1000 which is extremely

huge in number. This drastically huge number of features slows down the training phase of the retrieval process. The massive number of features increases the curse of dimensionality which results in the degradation of the performance in the retrieval results. To overcome this problem, a suitable feature selection technique is required. In this paper, a joint operator based on Shannon's entropy and Pearson Correlation Coefficient named as JSE-PCC technique is designed which is discussed in the subsequent sections.

1) THE JSE-PCC TECHNIQUE

The ranking of retrieved images is calculated by computing the score of each feature vector. In retrieval applications, entropy is a degree of ranking features based on the relevance of input and gallery images and is a process of evaluating specific patterns in the data. Similar images indicate similar patterns between the image and are represented with low entropy values. The low entropy means the similarity between the two features is obvious because of maximum knowledge. The entropy of uncertainty rises before getting knowledge while information can be obtained after performing random experiments [52]. Retrieval-based feature selection using entropy measures diminishes the problem of high dimensional feature space and reduces the computational cost. The problem with Shannon's entropy is that it is not large enough to select the best-retrieved features because it doesn't evaluate all the features and selects only random variables. To avoid the above-stated problem, Pearson's Correlation Coefficient is used along with the entropy value that has more capability to reduce feature space with accurate feature selection.

TABLE 1. Layer-level configuration of AlexNet with Mish architecture.

	Input	Output	Kernel, Filter	Stride	Padding
Input Layer	227×227×3	227×227×96	96, 11×11	1	1
Conv1	227×227×96	55×55×96	96, 11×11	[4,4]	[0,0,0,0]
Mish1	55×55×96	55×55×96	with scale 0.01		
CN1	55×55×96	55×55×96			
Pool1	55×55×96	27×27×96	3×3	[2,2]	[0,0,0,0]
Grouped Conv2	27×27×96	27×27×256	2×128, 5×5	[1,1]	[2,2,2,2]
Mish2	27×27×256	27×27×256	with scale 0.01		
CN2	27×27×256	27×27×256	with 256 channels		
Pool2	27×27×256	13×13×256	3×3	[2,2]	[0,0,0,0]
Conv3	13×13×256	13×13×384	384, 3×3	[1,1]	[1,1,1,1]
Mish3	13×13×384	13×13×384	with scale 0.01		
Grouped Conv4	13×13×384	13×13×384	2×192, 3×3	[1,1]	[1,1,1,1]
Mish4	13×13×384	13×13×384	with scale 0.01		
Grouped Conv5	13×13×384	13×13×256	2×128, 3×3	[1,1]	[1,1,1,1]
Mish5	13×13×256	13×13×256	with scale 0.01		
Pool5	13×13×256	6×6×256	3×3	[2,2]	[0,0,0,0]
FC6	6×6×256	1×1×4096			

Pearson’s correlation is the linear relationship between the input and target feature subset, depending on the value of covariance and standard deviation. Pearson’s value between the input and gallery image of clothes is ranging between -1 and 1 , representing the strength of the relationship indicated as strongly positive, strong negative, medium positive, medium negative, and weak or no correlation exists between the variables. If all the data points are close to the regression line, then it shows a high correlation represented with $+1/-1$ similarly, the zero value indicates that there is a weak /no relationship between the values and so on.

Initially, the entropy is employed to reduce the dimensions of each feature vector by eliminating the redundant pattern in the data while PCC finds the correlation between the query and gallery images. The zero-correlation value between the target and gallery image indicates that both images are dissimilar, so the whole feature vector of the gallery image is removed and not considered in the matching phase. Similarly, an approximate $+1/-1$ correlation shows the strength of the relationship hence, PCC lessens the number of feature vectors while entropy value reduces the dimensions of each feature vector.

The process of the optimized feature selection technique in which the size of N feature vectors is initially reduced

to N' using the low entropy value as shown in Fig. 4, parts (a) and (b), and then is reduced to M number of vectors using the PCC threshold value as demonstrated in Fig. 4, parts (c) and (d).

Mathematical Formulation: Suppose F_i is a feature vector of each gallery image with the set of values f_1, f_2, \dots, f_N . The list of all notations is enlisted in Table 2. The N represents the size of each extracted feature vector as shown in Fig. 4, part (a). Here the goal is to reduce the size of N into M . Initially the dimension of each feature vector is reduced into f'_1, f'_2, \dots, f'_N by finding the redundant or similar pattern from a single feature space through Shannon’s entropy. The mean and spread of the data are measured using the covariance formula as in Eq. (1).

$$cov(x_i, y_j) \leftarrow sum((x_i - x') \times (y_j - y')) \quad (1)$$

TABLE 2. List of notations.

Abbreviation	Description
F	Random feature or a variable
N	Size of each feature vector
f_N	The finite set of values of random variable F
$H(F)$	Entropy function of random feature
PCC	Pearson's Correlation Coefficient
N	Number of the data points
$\sum V_i$	Sum of V_i scores
$\sum V_j$	Sum of c scores
$\sum V_i V_j$	Sum of the product of V_i and V_j scores
$\sum V_i^2$	Sum of squared of V_i scores
$\sum V_j^2$	Sum of squared of V_j scores
$P(x_i)$	Probability of feature score of x_i
$\partial(x_i)$	The standard deviation of feature points x_i

Suppose “P” is the probability, provided that $\mathcal{P}(f_1), \mathcal{P}(f_2), \dots, \mathcal{P}(f_N)$ is the probability of each finite set of values of a feature vector. The value of each probability of a specific number should be ≥ 0 , provided that $\mathcal{P}(f_k) \geq 0$ where $k = 1, 2, \dots, N$ and $\sum_{k=1}^N \mathcal{P}(f_k) = 1$, then the information of f_k is represented as in Eq. (2) [37]:

$$I(f_k) = -\log_2 \mathcal{P}(f_k) \quad (2)$$

The \log is taken because the best approach is to convert all the results into a sum which is only possible through the \log function. The property of the sum of a \log is:

$$\log(ab) = \log(a) + \log(b) \quad (3)$$

The probability function is used to calculate the score of each feature. The features with a high covariance value indicate a

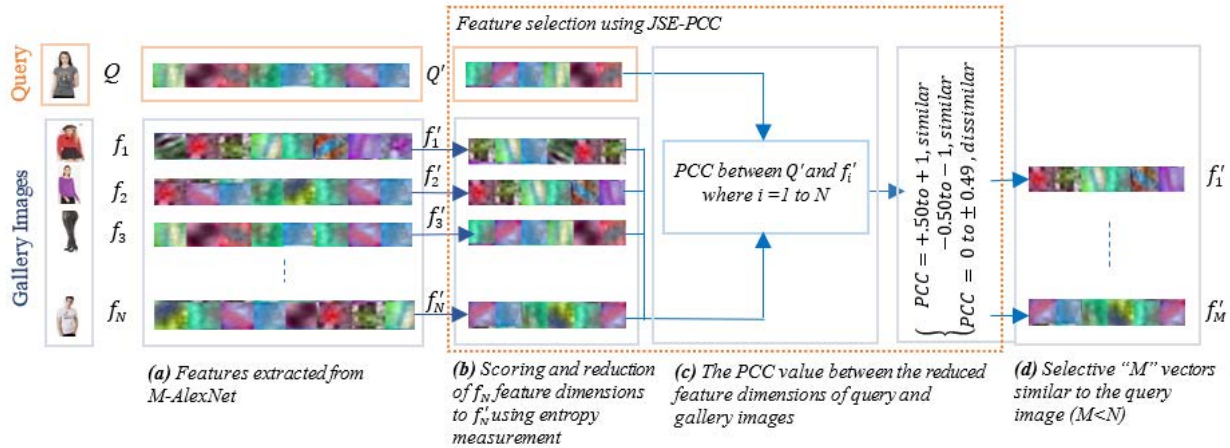


FIGURE 4. The process of optimized feature selection using the JSE-PCC technique (a) The Feature Vectors of size N extracted from M-AlexNet architecture is Represented as f_N (b) Depicts that the variables of each Feature Vector are ranked and reduced from f_N to f'_N (c) Shows the correlation between the reduced query image vector and the gallery images. The threshold defines the strength of correlation between images (d) Selects only those Feature Vectors which have high positive or negative correlation (high degree of similarity with the query vector) while discards rest of the vectors (highly dissimilar with the query vector).

high probability. The probability of high covariance feature f_i computed as indicated in Eq. (4).

$$P(x_i) \leftarrow \frac{cov(x_i, y_j)}{N} \tag{4}$$

In the retrieval process, the feature dimensions of query and gallery images are reduced using Shannon’s entropy and are computed as in Eq. (4). The sum of all the probability values is approximately equal to one. The feature with the lower entropy value is dropout because of the similar and redundant feature pattern as shown in Fig. 4, part (b). The entropy function $H(f)$ is expressed as the average information of the distribution of random features f provided in Eq. (5), and can be expressed as [38]:

$$H(\mathcal{F}) = - \sum_{k=1}^{\Phi} \mathcal{P}(f_k) \log_2 \mathcal{P}(f_k) \tag{5}$$

The PCC is applied to the reduced feature dimensions to eliminate some of the unrelated vectors. Entropy reduces the size of each vector while PCC eliminates some of the vectors while preserving the rest of the feature vectors for matching.

PCC finds the strength of the relationship between the query and gallery images. The only features with approximate zero values indicate no relationship and hence dropout from the retrieval process. The criteria for dropout of the dissimilar vectors are specified in the threshold function as defined in Fig. 4, part (c). The high correlation values are close to the regression line indicating the high similarity between the images. Let V_i and V_j represent the two variables, then the correlation between the two data points can be calculated through the formula of Pearson’s Correlation as in Eq. (6) [39]:

$$PCC = \frac{N \sum V_i V_j - \sum V_i \sum V_j}{\sqrt{(N \sum V_i^2 - (\sum V_i)^2) - (N \sum V_j^2 - (\sum V_j)^2)}} \tag{6}$$

The remaining resultant similar feature vectors are shown in Fig. 4, part (d). All the steps of the JSE-PCC technique are enlisted in Algorithm 1.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this portion, a detailed description of the experimental setting and obtained results is provided. Several sets of experiments for image retrieval are performed on the proposed method. Initially, the performance of FC6, FC7, and FC8 feature layers is evaluated. In the second experiment, ReLU, Swish, and Mish activation functions are initially assessed to improve the training and test accuracy and provide a comparison between the original and M-AlexNet on Deep-Fashion2, MVC, and CID datasets. The third experiment is conducted on the reduced feature dimensions using JSE-PCC technique. In the last experiment, different matching techniques are employed to assess the performance of clothes retrieval. Ultimately, the assessment of the intended method is performed with the original AlexNet, VGG16, VGG19, and ResNet50 models.

A. EXPERIMENTAL SETUP

The experimental environment involved desktop system with windows 10, Intel @Core i7-7700 CPU @ 3.60 GHz, RAM 16-GB, NVIDIA GeForce GTX1070, and matlab2020b. For testing the proposed framework of apparel retrieval, k-fold cross-validation is applied (k=10), which is believed to be a universal practice for the assessment of models.

B. DATASETS

Three datasets, namely DeepFashion2, MVC, and proposed CID datasets are selected for performance assessment. Variations in viewpoint, pose, occlusion, and illumination changes make these datasets more challenging. DeepFashion2 is a large-scale dataset [56]. It comprises of 491k images

Algorithm 1 Joint Shannon’s Entropy Pearson Correlation Coefficient JSE-PCC Based Feature Selection Technique

Input: $F = \{f_1, f_2, \dots, f_N\}$ (Feature vector of gallery image with a set of values, see Fig. 4)

N (Size of each feature vector)

Output: $F' = \{f'_1, f'_2, \dots, f'_M\}$ (Set of reduced features, ordered by relevance)

Provided that $M < N$

Begin

Step 1: **while** $i \in x_i \&\& j \in y_j$ no. of features

do//calculate the entropy of each extracted feature vector as in Fig. 4, part (a)

$$\begin{aligned} x' &= \text{mean}(x_i); & y' &= \text{mean}(y_j); \\ \text{cov}(x_i, y_j) &\leftarrow \text{sum}((x_i - x') \times (y_j - y')) // \text{covariance of single feature vector} \\ \text{cov}(x_i, y_j) &\leftarrow \text{argmax}(\text{cov}(x_i, y_j)) \\ P(x_i) &\leftarrow \frac{\text{cov}(x_i, y_j)}{N} \\ \text{entropy} &\leftarrow \text{sum}(x(:, i) \cdot \log(x(:, i))) \\ \text{entropy} &\leftarrow -1 \times \text{entropy} \end{aligned}$$

end while

Step 2: **for** each f_k // Score and reduce the size of N

feature vectors into N' having low entropy value, as in Fig. 4, part (b)

if $\text{sum}(\text{entropy}(f_k) \approx 1)$ **then**

$f_k.\text{drop}(\text{argmin}(\text{entropy}(x_k)))$

else go to step 3

end if

end for

Step 3: **while** $k \in x_k \&\& l \in y_l$ **do**// calculate the PCC between the gallery and input image, as in Fig. 4, part (c)

$$\begin{aligned} \sigma(x_k) &\leftarrow \text{sum}(\text{sqrt}(x_k - x')) \\ \sigma(y_l) &\leftarrow \text{sum}(\text{sqrt}(y_l - y')) \\ \text{PCC} &\leftarrow \frac{\text{cov}(x_k, y_l)}{\text{sqrt}(\sigma(x_k) \times \sigma(y_l))} \end{aligned}$$

end while

Step 4: **if** $(\text{PCC} \geq -0.49 \&\& \text{PCC} \leq 0.49)$ **then**

//Dropout the dissimilar vectors, as shown in Fig. 4, part (d)

$F' \leftarrow y_l.\text{drop}$

end if

End

with rich annotations using a bounding box and improves sparse landmarks. The dataset contains 13 categories and has minimal annotation accuracy. The types of images in this dataset include customers, commercials, shopping stores, etc. The main challenge of this dataset includes landmark estimation and apparel retrieval. The MVC dataset contains 161,260 images with a resolution size of 1920×2240 [57]. It contains 264 attributes with duplicate images covering four (left, right, front, and back) views. In this work, 2047 and 632 images are picked for training and testing respectively.

The proposed apparel retrieval is also evaluated on the proposed CID dataset. The dataset contains 130,000 colored images but the proposed method is tested on

66,349 images for experiments that are good enough to train the deep learning model. There is a total of 13 classes namely blouses, coats, dresses, hoodies, jackets vests, jeans, pants, shirts, shorts, skirts, sweaters, trousers, and t-shirts. Each class comprises 35,735 men’s images and 30,614 women’s images with different resolution sizes. Table 3 shows the statistics of proposed CID dataset.

TABLE 3. Statistics of proposed CID dataset for proposed apparel retrieval algorithm.

Class	Men	Women	Total Images	Resolution
Blouse	0	4,000	4,000	224×224×3
Coat	2,536	2,464	5,000	224×224×3
Dresses	2,314	3,015	5,329	224×224×3
Hoodie	1,628	3,057	4,685	224×224×3
Jackets	3,699	1,304	5,003	224×224×3
Vests				
Jeans	5,201	1,128	6,329	224×224×3
Pants	3,299	2,722	6,021	224×224×3
Shirts	3,221	1,157	4,378	224×224×3
Shorts	4,133	1,673	5,806	224×224×3
Skirts	0	3,689	3,689	224×224×3
Sweaters	2,236	2,464	4,700	224×224×3
Trousers	3,258	2,145	5,403	224×224×3
T-Shirts	4,210	1,796	6,006	224×224×3
Total	35,735	30,614	66,349	

The experiments are conducted with a resolution size of $224 \times 224 \times 3$. The proposed dataset provides both streets as well as shop images. This dataset intends to provide a benchmark to practically assess the progress of different modern computer vision techniques that rely on a big amount of data for fashion understanding. The dataset is not annotated with labels as it requires domain expertise.

1) DATA PREPARATION

Since the acquired images are not standardized because they are collected from various online sources. To normalize the dataset data cleaning is performed as preprocessing steps. Initially, the images are renamed, and their size is adjusted to $224 \times 224 \times 3$. Finally, the images are converted into the same .jpg format. The resultant pre-processed images are depicted in Fig. 5. Extensive experiments are carried out on training and test images with the k-fold cross-validation method.



FIGURE 5. Sample images of ALCID dataset.

TABLE 4. Performance analysis of fully connected layers of baseline AlexNet, VGG-16, and VGG-19 models on DeepFashion2 dataset.

Models	FC layers	Accuracy (%)	F1 Score (%)	Sensitivity (%)	Specificity (%)	FPR (%)	Training time (s)
AlexNet	FC6	83.93	87.61	85.84	98.93	0.016	84.0
	FC7	78.70	87.58	76.13	87.65	0.0138	84.6
	FC8	75.95	76.09	70.89	77.22	0.0151	86.9
VGG16	FC6	76.19	86.21	86.33	78.69	0.0159	561
	FC7	77.08	85.38	86.25	78.53	0.0163	509
	FC8	72.39	79.13	85.91	97.51	0.0165	478
VGG19	FC6	70.68	85.66	75.38	72.62	0.0158	466
	FC7	68.32	84.05	74.61	69.09	0.0162	569
	FC8	62.91	74.01	74.90	63.98	0.0155	603

C. RETRIEVAL EVALUATION PROTOCOL

In the entire collection of experiments, the performance of suggested image retrieval is calculated using Recall, mean average precision (mAP), F-score, Accuracy, and training time. These evaluation measurements performed a significant role in the analysis of experimental results. For retrieval tasks, precision is defined as the proportion of retrieved images that match the given query over the retrieved images. Similarly, Recall is the ratio of retrieved images that match the given query over the relevant images. For each query image, mAP is the mean of the whole AP of all the queries. The indicators formulas are derived in Eqs. (7), (8), (9), and (10) [42].

$$Precision = \frac{|\{relevant\ images\} \cap \{retrieved\ images\}|}{|\{retrieved\ images\}|} \tag{7}$$

$$Recall = \frac{|\{relevant\ images\} \cap \{retrieved\ images\}|}{|\{relevant\ images\}|} \tag{8}$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \tag{9}$$

N represents total images in the database and AP is the average precision that gives a superior knowledge of the proposed model in sorting the results of the query image. The collective values of precision and Recall represent overall accuracy of image retrieval called F-score or F-measure, which indicates overall effectiveness of the image retrieval that can be defined as:

$$F - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{10}$$

D. EXPERIMENT 1: RETRIEVAL PERFORMANCE ON DIFFERENT FC LAYERS OF BASELINE MODELS ON DEEPFASHION2 DATASET

In the experimental setting, the proposed method is evaluated by extracting the features using baseline AlexNet, VGG16, and VGG19 models with an image size of 224 × 224 × 3. The features are extracted from three FC layers FC6, FC7, and FC8, and the results showed that as compared to FC7 and FC8, FC6 is more dependable and gives an abundant feature set with the highest retrieval accuracy and is more reliable in terms of accuracy and training time as shown in Table 4.

The table shows that as compared to VGG16 and VGG19, the FC6 layer features of baseline AlexNet shows an accuracy of 83.93%.

E. EXPERIMENT 2: RESULTS OF TRAINING AND TEST ACCURACY USING RELU, SWISH, AND MISH ON CIFAR-10 DATASET

The second experiment is based on the assessment of various activation functions on CIFAR-10 dataset. Here, the training is performed only on 500 images per class using SoftMax classifier. The dataset splits into training and test sets using the k-fold cross-validation method. Initially, the performance of ReLU, Swish, and Mish using AlexNet architecture is tuned concerning different learning rates and estimates the gap between the training and test accuracy on SGD optimizer. The ideal value of the learning rate is very crucial in deep learning. Here the effect of different learning rates is investigated to evaluate the efficiency and training time of Mish over Swish and ReLU in Fig. 6. Fig. 6, part (A) depicts the accuracy between the training and test sets and shows that all the activation functions perform better on a 0.1 learning rate, out of which Mish leads as compared to other activation functions and minimizes the accuracy gap up to 0.4. Meanwhile, the training and test loss curve of Mish, Swish, and ReLU activation functions is portrayed in Fig. 6, part (B).

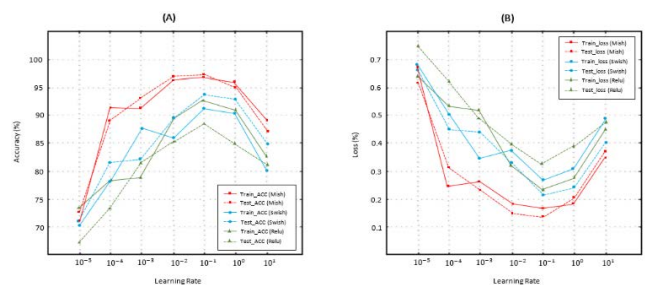


FIGURE 6. Accuracy and loss of different activation functions using AlexNet on CIFAR-10 dataset. (A) Training and test accuracy vs different learning rates. Similarly, (B) is the training and test loss vs different learning rates. The consistent line is the results of training data while the dotted line shows the results of test data.

TABLE 5. Training and test accuracy and loss of ReLU, Swish, and Mish under different learning rates.

Activation Function	Learning Rate	Training Acc. (%)	Test Acc. (%)	Training-Test Acc. (%)	Training Loss (%)	Test Loss (%)	Execution time
ReLU	0.00001	74.7	67.8	6.9	0.645	0.751	15 min 51 s
	0.0001	78.8	73.6	5.2	0.539	0.622	18 min 3 s
	0.001	78.9	82.0	3.1	0.527	0.491	18 min 54 s
	0.01	89.9	85.1	4.8	0.321	0.399	20 min 43 s
	0.1	93.5	88.4	5.1	0.246	0.337	23 min 22 s
Swish	0.00001	69.8	72.4	2.6	0.677	0.662	14 min 32 s
	0.0001	77.8	82.6	4.8	0.513	0.450	16 min 2 s
	0.001	88.5	82.7	5.8	0.351	0.448	16 min 55 s
	0.01	85.8	89.7	3.9	0.383	0.331	18 min 4 s
	0.1	91.6	94.0	2.4	0.281	0.221	22 min 38 s
Mish	0.00001	71.1	73.5	2.4	0.682	0.620	18 min 1 s
	0.0001	91.9	89.1	2.8	0.250	0.319	20 min 30 s
	0.001	91.5	93.3	1.8	0.254	0.246	23 min 12 s
	0.01	97.0	97.8	0.8	0.193	0.151	23 min 45 s
	0.1	97.6	98.0	0.4	0.188	0.143	24 min 5 s

The error is estimated using the cross-entropy loss function. The training and test accuracy and loss function with their estimated time is recorded in Table 5. Observing the values in the Table, it predicts that Mish has the minimum gap between the training and test accuracy and minimizes the test loss up to 0.143%, indicating in boldface. Similarly, due to the high complexity of Mish over Relu, Mish is not efficient in terms of training time.

Dropout is a regularization technique and is a remarkably effective approach. Dropout rates require additional hyper-parameters for tuning to get optimal performance. By fixing the learning rate to 0.1, and providing the same implementation settings mentioned above, Mish, Swish, and ReLU activation functions are evaluated under different dropout rates and proved that Mish is shown to have constant progress over different activation functions. Fig. 7 part (A) shows that it minimizes the gap between training and test accuracy up to 0.3% at a 0.5 dropout rate. Similarly, the loss curve of different dropouts is depicted in Fig. 7, part (B). The smooth curve of loss functions of Mish is simple to optimize and has better capability to generalize the model. It minimizes the test

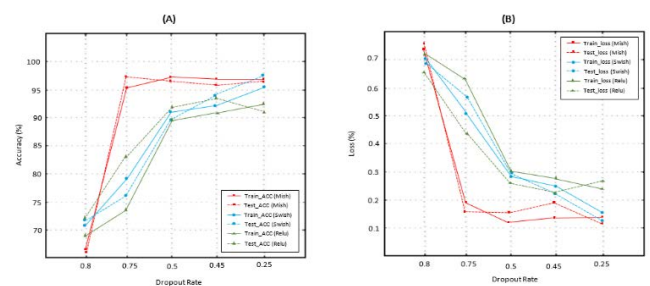


FIGURE 7. Accuracy and loss of different activation functions using AlexNet on CIFAR-10 dataset. (A) Training and test accuracy vs different dropout rates. Similarly, (B) is the training and test loss vs different dropout rates. The consistent line is the results of training data while the dotted line shows the results of test data.

loss up to 0.164. In short, Mish outperforms as compared to other activation functions, but again it takes more processing time as compared to ReLU and Swish. The training and test accuracy and loss function under varied dropout rate is indicated in Table 6. In Table 6, the bold values depict the highest values.

TABLE 6. Training and test accuracy and loss of ReLU, Swish, and Mish under different dropout rates.

Activation Function	Dropout Rate	Training Acc. (%)	Test Acc. (%)	Training-Test Acc. (%)	Training Loss (%)	Test Loss (%)	Execution time
ReLU	0.8	68.9	73.7	4.8	0.721	0.662	18 min 7 s
	0.75	73.8	83.9	10.1	0.643	0.428	18 min 43 s
	0.5	88.9	93.4	4.5	0.311	0.362	24 min 56 s
	0.45	91.2	94.3	3.1	0.280	0.243	24 min
	0.25	93.5	91.3	2.2	0.248	0.284	25 min 36 s
Swish	0.8	71.1	73.6	2.5	0.711	0.680	23 min 32 s
	0.75	79.3	76.5	2.8	0.512	0.582	28 min 48 s
	0.5	91.8	89.9	1.9	0.287	0.298	30 min 61 s
	0.45	92.9	94.9	2.0	0.258	0.234	32 min 3 s
	0.25	95.9	97.1	1.2	0.150	0.131	32 min 43 s
Mish	0.8	68.9	66.5	2.4	0.741	0.752	21 min 2 s
	0.75	96.0	97.5	1.5	0.196	0.163	20 min 45 s
	0.5	97.7	97.4	0.3	0.132	0.164	24 min 18 s
	0.45	97.5	96.2	1.3	0.145	0.191	24 min
	0.25	96.5	96.0	0.5	0.146	0.128	32 min 5 s

F. EXPERIMENT 3: RESULTS OF PROPOSED PRE-TRAINED M-ALEXNET

In this experiment, the performance of M-AlexNet, already pre-trained on CIFAR10 dataset is now evaluated on three different clothes image datasets namely, DeepFashion2, MVC, and CID datasets. Initially, the feature vectors are calculated from the entire FC layers of both network models. The retrieval performance is evaluated on all three fully connected layers to judge which FC layer can produce effective clothing features. As shown in Table 7, the top k images have a high mAP at FC6 layer on DeepFashion2 dataset. The FC6 layer offered abundant low-level features as compared to the high-level layers of FC7 and FC8. The low-level features are useful to extract fine-grained features of clothes provided that the image description is more realistic, accurate, and less generalized as compared to other fully connected layers. The AlexNet with Mish is improved from baseline AlexNet by 1.22% and shows a mean average precision of 85.15%. Similarly, the M-AlexNet on MVC dataset is improved by 2.2%

on FC6 layer, 5.5% on FC7 layer, and 2.75% on FC8 layer as compared to the baseline AlexNet model as indicated in Table 8. The retrieval performance of the M-AlexNet is graphically revealed using the average precision graph of FC6, FC7, and FC8 layers on the DeepFashion2 dataset as demonstrated in Fig. 8 parts (a), (b), and (c). In Fig. 9, Parts (a), (b), and (c) highlight the precision improvement of FC6 over FC7 and FC8 layers of AlexNet and M-AlexNet on MVC dataset.

On the proposed ALCID dataset, M-AlexNet is improved by 4.24% on the FC6 layer, 6.07% on FC7, and 6.66% on the FC8 layer, as depicted in Table 9. Fig. 10, shows the improvement of AlexNet in the FC6 layer on ALCID dataset.

The proposed outcomes reveal that M-AlexNet achieves better accuracy of 83.65% on CID as compared to MVC dataset which has AP of 82.04%. On the other way, DeepFashion2 leads with an accuracy of 85.15% as compared to MVC and CID datasets

TABLE 7. The mAP of Top “k” retrieval results of AlexNet and M-AlexNet at FC6, FC7, and FC8 layers on DeepFashion2 dataset.

	AlexNet (mAP %)						M-AlexNet (mAP %)					
	Top 10	Top 20	Top 30	Top 40	Top 50	AP	Top 10	Top 20	Top 30	Top 40	Top 50	AP
FC6	88.61	85.32	84.29	82.11	79.30	83.93	90.39	88.68	87.13	80.19	79.38	85.15
FC7	86.33	82.19	78.36	75.98	70.62	78.70	89.14	83.62	81.13	77.91	73.26	81.01
FC8	82.91	78.36	77.19	72.68	68.59	75.95	86.29	81.38	76.58	76.01	69.83	78.02



FIGURE 8. Average precision vs number of Images retrieved. (a) Shows the precision result of FC6 layer of AlexNet and M-AlexNet on DeepFashion2 dataset. Similarly (b) and (c) show the precision result of FC7 and FC8 layers of AlexNet and M-AlexNet on DeepFashion2 dataset respectively.

TABLE 8. The mAP of top “k” retrieval results of AlexNet and modified AlexNet at FC6, FC7, and FC8 layers on MVC dataset.

	AlexNet (mAP %)						M-AlexNet (mAP %)					
	Top 10	Top 20	Top 30	Top 40	Top 50	AP	Top 10	Top 20	Top 30	Top 40	Top 50	AP
FC6	88.65	85.94	80.32	74.38	69.91	79.84	88.61	87.37	82.19	79.39	72.65	82.04
FC7	88.48	85.36	79.99	72.68	62.64	77.83	89.66	87.91	86.55	79.42	73.11	83.33
FC8	85.32	83.69	80.68	69.92	62.39	76.40	86.11	84.21	80.96	73.54	70.91	79.15



FIGURE 9. Average precision vs number of images retrieved. (a) Shows the precision result of FC6 layer of AlexNet and M-AlexNet on MVC dataset. Similarly (b) and (c) show the precision result of FC7 and FC8 layers of AlexNet and M-AlexNet on MVC dataset respectively.

TABLE 9. The mAP of top “k” retrieval results of AlexNet and M-AlexNet at FC6, FC7, and FC8 layers on ALCID dataset.

	AlexNet (mAP %)						M-AlexNet (mAP %)					
	Top 10	Top 20	Top 30	Top 40	Top 50	AP	Top 10	Top 20	Top 30	Top 40	Top 50	AP
FC6	94.22	91.68	85.39	82.61	80.10	86.80	95.91	93.28	91.65	89.99	84.39	91.04
FC7	87.98	82.31	80.11	79.32	75.68	81.08	91.86	90.92	88.62	85.14	79.23	87.15
FC8	85.15	81.98	77.62	70.13	68.26	76.63	90.45	88.26	82.39	80.01	75.32	83.29

G. EXPERIMENT 4: RETRIEVAL PERFORMANCE OF OPTIMIZED FEATURE SELECTION BASED ON JSE-PCC TECHNIQUE

In this experiment, the retrieval performance is evaluated by reducing the dimensions of each feature vector and with

a smaller number of feature vectors using the proposed JSE-PCC operator on the selected datasets. Initially, the features of clothes are extracted using M-AlexNet, and then JSE-PCC technique is applied to reduce the dimensions. As stated earlier, the size of each extracted feature vector is



FIGURE 10. Average precision vs number of images retrieved. (a) Shows the precision result of FC6 layer of AlexNet and M-AlexNet with Mish on ALCID dataset. Similarly (b) and (c) show the precision result of FC7 and FC8 layers of AlexNet and M-AlexNet on ALCID dataset respectively.

reduced using the entropy value while the number of feature vectors is reduced using PCC value. In this paper instead of 4096 feature vectors, the top 2000 high-rank features are selected in M-AlexNet, AlexNet, VGG16, VGG19, and the top 800 features are selected in the case of ResNet50 as shown in Table 10.

TABLE 10. Top selected feature dimensions with low entropy values at fully connected layer of AlexNet with Mish, AlexNet, VGG-16, VGG19, and ResNet-50.

Models	FC Layer	Top selected features with low entropy values
M-AlexNet	FC6	2000
AlexNet	FC6	2000
VGG-16	FC6	2000
VGG-19	FC6	2000
ResNet50	FC	800

The vectors which have no relation will not take part in the retrieval process. As shown in Table 11, the correlation is calculated between the test and gallery images. The exactly zero value indicates that there is no relation between the query and gallery images and is entirely dissimilar. The values between $-0.4 \leq PCC < 0$ are represented with weak negative relation, the values between $-0.7 \leq PCC < 0.4$ show medium negative, and the values between $-1 \leq PCC < 0.4$ indicate the strength of strong negative relation between the images. Similarly, in another direction, the values between $0 \leq PCC < 0.4$ indicate a weak positive relation, the values between $0.5 \leq PCC < 0.8$ show the medium positive while the values between $0.7 \leq PCC < 1$ are represented as a strong positive relationship between the query and gallery images.

The strength of the relation between images is graphically visualized in Fig. 11. The higher positive values represent a high degree of correlation, meaning the similarity between images. Similarly, the minimum negative values indicate the minimum correlation and chances of dissimilarity between the query and gallery images. Fig. 11, part (a) shows the medium positive correlation between the sweater as a query

image and the coat from the gallery image because both categories have the same style and approximately similar material but belong to a different class. Part (b) shows a strong positive relationship if both the target and gallery images are the same because of a high degree of similarity, and part (c) provides medium negative relation between hoodie and T-shirt. The reason for medium negative relations between different class categories is to share some common properties like color or texture etc. Finally, part (d) indicates no relationship between the blouse and skirt because both belong to a different class as well as don't share any common properties among them.






Pearson correlation is a superlative approach to quantify the strength of correlation between variables of interest while discarding the rest of the feature vectors. In the retrieval process, some of the feature vectors are reduced, which have no relationship and hence improve the retrieval results. The retrieval performance of M-AlexNet on DeepFashion2 dataset is 85.15% which is 5.11% improved using JSE-PCC operator and 1.95% improved on the MVC dataset. Similarly, the mAP of M-AlexNet on CID dataset is 83.65% which is further enhanced by 3.51% using JSE-PCC operator and with the AP of 87.16%, as shown in Table 12. The size of entire feature vectors is reduced by considering only the top k-ranked feature vectors that have low entropy values. Similarly, only those feature vectors take part in the retrieval process which have exactly/ strong positive relationship between the images.

H. EXPERIMENT 5: RETRIEVAL PERFORMANCE USING DIFFERENT MATCHING TECHNIQUES

Effective image retrieval needs efficient similarity metric techniques for indexing database images based on minimum resemblance differences among query and indexed images. In this experimental analysis, various similarity measure techniques are used to calculate the performance of clothes image retrieval [59].

In this paper, five different matching techniques are employed to assess the robustness of the suggested method i.e., Euclidean Distance (ED), Mahalanobis Distance (MD), Maximum Value Distance (MVD), Sum of Absolute Difference (SAD), and Minkowski Distance (MiD) as

TABLE 11. The pearson correlation between target and gallery Image of ALCID dataset. The zero correlation indicates that there is no relation between the image. The values exactly/close to the +1 show that the target image either belongs to the same class (strongly positive) or the same type of different class (medium positive). Similarly, the values exactly/close to -1 indicate the perfect negative or medium negative relationship between the images.

	Blouse	Coat	Dresses	Hoodie	Jackets & Vests	Jeans	Pants	Shirts	Shorts	Skirts	Sweater	Trouser	T-shirts
	0.88	-0.8	-0.6	0.75	0.5	0	-0.1	0.8	-0.1	-0.01	-0.9	0	0.89
	-0.01	0	-0.1	-0.35	-0.001	0.91	0.98	-0.2	0.65	0.59	-0.11	0.35	-0.56
	0.78	0.05	-0.25	-0.4	-0.21	-0.65	0	0.89	-0.1	-0.36	0.25	-0.11	0.75
	-0.03	-0.1	-0.7	0	-0.65	-0.01	0.32	-0.5	0.91	-0.5	0	-0.9	-0.11
	-0.9	0.95	0.2	0.66	-0.86	-0.01	-0.3	-0.6	0	-0.06	0.79	0	-0.72

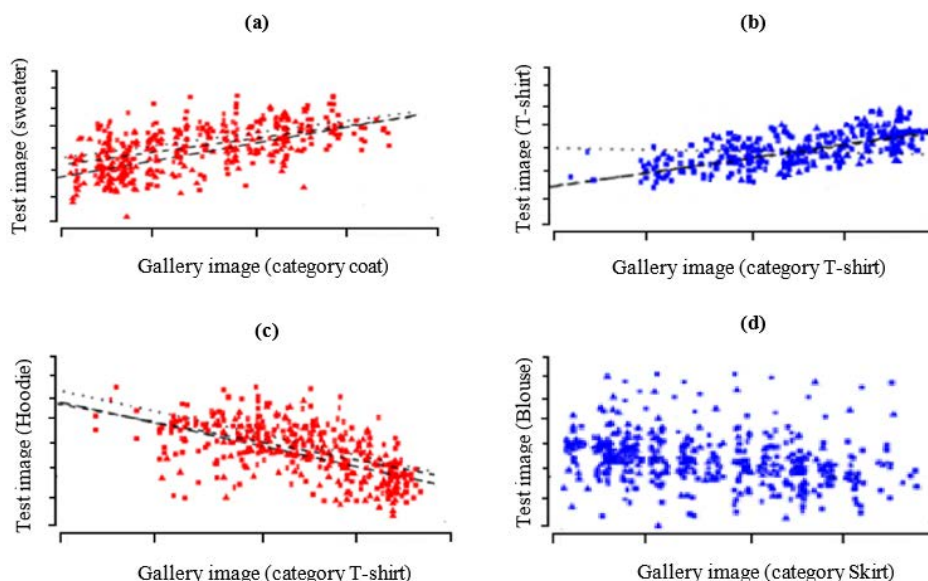


FIGURE 11. The correlation between test and gallery images (a) Shows the medium positive relation between sweater and coat because of approximately same style but belongs to different class (b) illustrate the strong positive relationship because both query and gallery images belong to the same class (c) Shows the medium negative relation between different classes because both hoodie and t-shirt may have same features like colors (d) Shows no relation because blouse and skirts belongs to extremely different class.

TABLE 12. The mAP of top “k” image retrieval using M-AlexNet with optimized features using JSE-PCC technique on DeepFashion2, MVC, and ALCID datasets.

Datasets	Top Rank Results (%)					AP (%)
	10	20	30	40	50	
DeepFashion2	91.18	93.65	90.21	89.38	86.91	90.26
MVC	86.32	88.92	84.38	82.99	77.38	83.99
CID	89.41	92.59	88.68	84.46	80.69	87.16

depicted in Table 13. The highest accuracy is attained using ED on Deepfashion2, MVC, and CID datasets. Moreover, as compared to MVC and CID, DeepFashion2 performs better. The graphical results of the above-mentioned matching techniques are shown in Fig. 12. The experimental results

show that the ED has excellent performance as compared to other metrics. It shows effective results with a high average precision and accuracy rate. The pictorial representation of the retrieval results of the entire matching techniques is shown in Fig.13

TABLE 13. Accuracy of proposed image retrieval method using different distance measure techniques on ALCID dataset.

Distance Metrics	Datasets		
	DeepFashion2	MVC	CID
	(%)	(%)	(%)
ED	90.26	83.99	87.16
MD	81.26	69.35	81.99
MVD	89.35	79.56	80.67
SAD	86.34	81.92	85.23
MiD	79.48	76.22	75.39

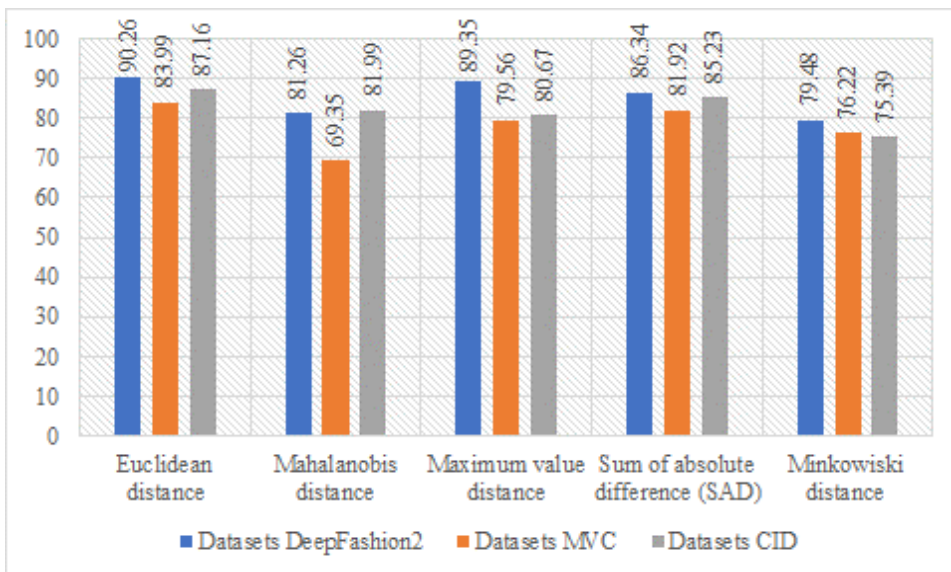


FIGURE 12. Graphical results of different matching techniques.

I. COMPARATIVE ANALYSIS WITH EXISTING PRE-TRAINED DEEP CNN MODELS

Processing the deep features of clothes is an extremely critical problem. The clothes features are obtained from fully connected layers using four different existing CNN models that are used in the above experiments i.e., AlexNet, VGG-16, VGG-19, and ResNet50. Initially, the performance of Mish, Swish, and ReLU activation functions is evaluated by tuning the network with different learning rates and dropout values. Fig. 14 shows the overall performance of tuning dropout values of activation functions by fixing the learning rate to 0.1.

The motivation of this research is to improve the training and test accuracy of different models. Here, AlexNet is used as a test model in which the ReLU function is swapped with the Mish activation function. Different experiments show that Mish performs better as compared to ReLU. Extensive

experiments conducted by various researchers discovered that Mish has more potential and better performance as compared to other activation functions, so this is the reason why Mish is the most recommended function nowadays [10], [44], [45], [46], [47]. Table 14 shows the comparison of M-AlexNet with and without dropout technique with other existing approaches, which shows that M-AlexNet with a 0.5 dropout rate performs better with improved training and test accuracy up to 0.003%.

Similarly, the comparison of four different deep CNN models with reduced feature dimensions is performed based on precision and recall values. The top selected features are used for the retrieval of clothes images. Table 15 reveals the relative results of other existing techniques with M-AlexNet and M-AlexNet + JSE-PCC operator. It is exciting to observe that the obtained results of the suggested method outperform other existing approaches. We compare the accuracy of four

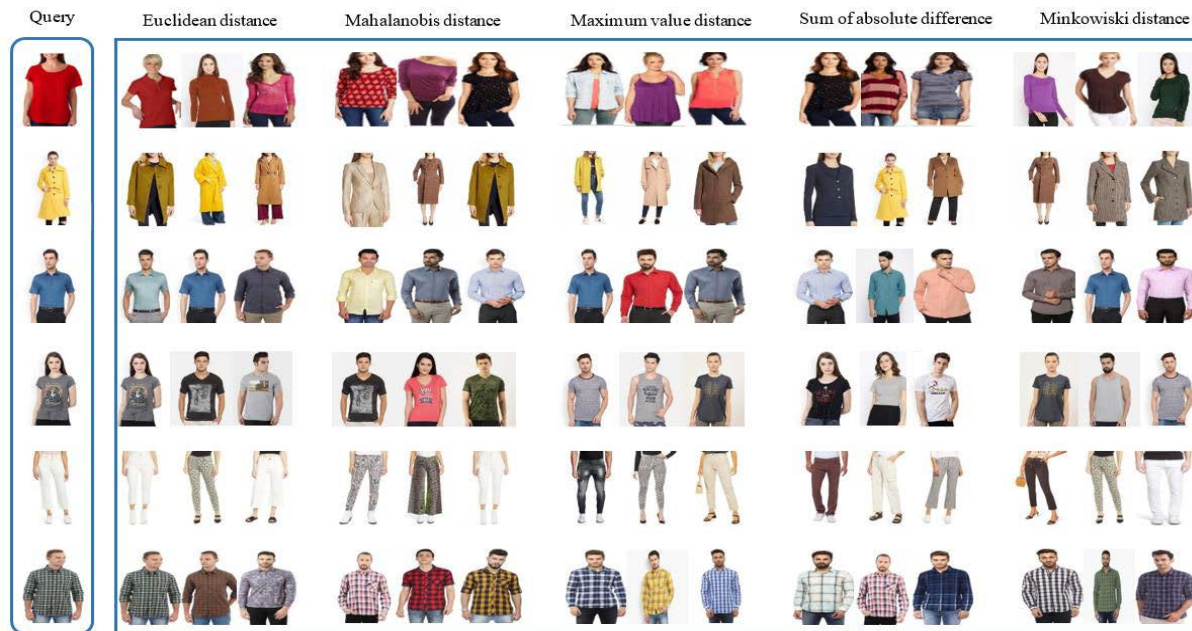


FIGURE 13. Snapshots of clothes image retrieval based on Euclidean distance, Mahalanobis distance, maximum value distance, Sum of absolute difference (sad), and Minkowski distance.

TABLE 14. Comparison of AlexNet with different activation functions, optimizer, learning rate, and dropout values. The bold value indicates better training and test accuracy.

Dataset	AV	Optimizer	LR	Dropout Rate	Train_ACC (%)	Test_ACC (%)	Train-Test_ACC (%)
CIFAR-10[5]	ReLU	SGD	0.001	xxx	0.8720	0.604	0.268
	Swish	SGD	0.001	xxx	0.8734	0.6724	0.201
	Mish	SGD	0.001	xxx	0.8738	0.7398	0.134
CIFAR-100[48]	ReLU	SGD	0.01	0.25	0.6651	0.4191	0.246
	Mish	SGD	0.01	0.25	0.6137	0.4097	0.204
	Swish	SGD	0.01	0.25	0.6125	0.4175	0.195
CIFAR-10 [48]	ReLU	SGD	0.01	0.25	0.9155	0.7565	0.159
	Swish	SGD	0.01	0.25	0.8786	0.7656	0.113
	Mish	SGD	0.01	0.25	0.8676	0.7626	0.105
VOT 2018 [44]	Mish	SGD_GCC	0.01	xxx	0.726	0.511	0.215
OTB50 [44]	Mish	SGD_GCC	0.01	xxx	0.80	0.602	0.198
Malaria dataset [49]	ReLU	Nadam	0.002	0.5	0.8164	0.6914	0.125
	Mish	Nadam	0.002	0.5	0.8271	0.7251	0.102
	ReLU	RMSprop	0.001	0.5	0.7971	0.6611	0.136
	Mish	RMSprop	0.001	0.5	0.8092	0.6962	0.113
	ReLU	SGD	0.002	0.5	0.7607	0.6307	0.130
ImageNet2012 [7]	Mish	SGD	0.002	0.5	0.7742	0.6652	0.109
	ReLU	SGD	0.01	0.5	-	0.847	-
Plant Village dataset [30]	Swish	Adam	0.001	0.5	0.91	0.92	0.01
Fish school feeding behavior dataset[50]	ReLU	Ranger	0.0001	0.5	0.85	0.8	0.05
ALCID (Ours)	Mish	SGD	0.1	xxx	0.976	0.98	0.004
ALCID (Ours)	Mish	SGD	0.1	0.5	0.977	0.974	0.003

different deep CNN models to estimate the performance of the suggested methodology on top 10, 20, 30, 40, and 50 retrieval results.

The AlexNet is the best classifier that shows 83.8% precision and 84.51% recall rate on CID dataset. The VGG16, VGG19, and ResNet50 models also show good precision and

TABLE 15. Performance comparison of AlexNet with Mish and feature optimization using JSE-PCC technique with existing pre-trained Deep CNN models.

Ref.	Year	Models	Datasets	Matching Technique	Precision (%)	Recall (%)
[67]	2021	AlexNet	INRIA	Stereo matching	79.3	--
[68]	2022	VGG16	METU	ED	35.6	--
[69]	2022	VGG19	COVID-19 X-ray images	Cosine distance	84.33	--
[70]	2021	VGG19	CIFAR-10	ED	85.01	65.31
[69]	2022	ResNet50	COVID-19 X-ray images	Cosine distance	89.37	--
[71]	2021	ResNet50	ArtImages	ED	77.4	75.2
[68]	2022	ResNet50	METU	ED	75.7	--
	2022	AlexNet (Ours)	DeepFashion2	ED	86.93	86.32
			CID	ED	83.80	84.51
	2022	VGG16 (Ours)	CID	ED	78.45	76.38
	2022	VGG19 (Ours)	CID	ED	78.69	80.13
	2022	ResNet50 (Ours)	CID	ED	78.01	72.39
	2022	M-AlexNet (Ours)	DeepFashion2	ED	90.15	88.96
			MVC	ED	82.04	85.26
			CID	ED	90.04	88.65
	2022	M-AlexNet + JSE-PCC (Ours)	DeepFashion2	ED	92.19	91.26
			MVC	ED	83.89	89.39
			CID	ED	86.67	91.42

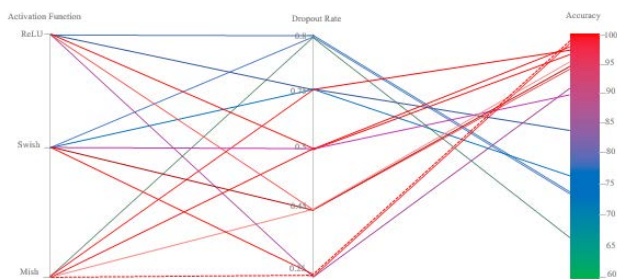


























FIGURE 14. Comparison of Mish with Swish and ReLU activation functions by tuning on different dropout values. The dotted red line indicates the highest value obtained from Mish under the dropout rate of 0.5.

recall rates. The M-AlexNet with Mish competes with the accuracy of AlexNet with a 92.15% average precision and 88.95% recall rate on DeepFashion2 dataset. Similarly, if the number of feature vectors and size of each vector is reduced

using JSE-PCC operator, maximum precision of 92.19% on DeepFashion2 and 91.42% recall value on the CID dataset is attained as compared to other deep CNN models. The subjective clothes retrieval comparisons of M-AlexNet and JSE-PCC technique with other state-of-the-art methods are shown in Table 16.

In the Table 16, the proposed M-AlexNet and JSE-PCC technique shows cutting edge performance using Euclidean distance as the retrieval results based on the query image not only focus on the color, and shape but also considers the fine detailed information of an object as compared to the results provided by the Mahalanobis distance. The state-of-the-art retrieval results like sketch based, only considers the input as sketches, the AsymNet shows better performance by handling occlusion, color, and styles. The GeM pooling and attention mechanism handle multiple views and shape but not colors. Similarly, the divide and conquer approach for deep metric

TABLE 16. Retrieval results comparison of M-AlexNet and JSE-PCC technique with existing state-of-the-art methods.

Ref.	Methods	Distance Measures	Query Image	Retrieval Results
[72]	Sketch based clothing image retrieval	ED		
				
[73]	Video2Shop retrieval using AsymNet,	LSTM hidden states		
				
[74]	Generalized mean (GeM) pooling and attention mechanism	ED		
				
[75]	Divide and conquer approach for deep metric learning	K-separate distance metrics		
				
[76]	Global and item feature extraction using Semantic Fusion Network	ED		
				
	Proposed M-AlexNet and JSE-PCC	ED		
		MD		

learning and global item feature extraction using semantic fusion network didn't provide the specific retrieval results and consider the finegrained features globally which fails to limit the searching criteria so, the proposed M-AlexNet and JSE-PCC technique provide better retrieval results as compared to state-of-the-art techniques.

V. DISCUSSION AND COMPARISONS

The experimental findings show that the performance of clothes image retrieval depends on deep features ranking. The impact of modifying the AlexNet architecture for reducing the network capacity to fewer layers and changing the ReLU

to Mish activation function improves the retrieval accuracy. Furthermore, the features extracted from modified AlexNet are reduced to minimize the complexity of the retrieval process. The size of each feature vector is reduced using the entropy indicator. The entropy is a good measure to calculate the top *k* values and rank the features in descending order. The number of features is reduced using Pearson's coefficient. Pearson's coefficient already discards some of the features which are entirely dissimilar from the query image. Reducing the size and number of feature dimensions using the proposed JSE-PCC technique helps in the retrieval process in such a way that already most of the features which are entirely

different don't take part in the retrieval process. Similarly, the feature values are already ranked. It reduces the overhead of the retrieval process during matching.

VI. CONCLUSION

In this paper, we have demonstrated that the proposed M-AlexNet can extract a more significant number of features holding maximum information capacity, improving the training and test accuracy of clothes retrieval. Furthermore, to further refine the clothes retrieval accuracy, a more robust JSE-PCC technique is introduced to diminish the irrelevant features while ranking the rest of the features, significantly improving the precision and recall rate. Extensive experiments prove that M-AlexNet and JSE-PCC outperform state-of-the-art approaches and show cutting-edge performance on DeepFashion2 and proposed CID datasets.

These results indicate that the proposed M-AlexNet and JSE-PCC technique is particularly beneficial in online clothes retrieval and recommendation systems, where results accuracy is the primary concern. In such applications, extracted features play a vital role due to unexpected clothes related issues like minor variations between styles. The improved AlexNet architecture using Mish can be employed to extract these minor cloth-related features. Here, the JSE-PCC technique is beneficial to reduce and rank the large number of features extracted from M-AlexNet.

Future research will focus on post-ranking between the query image and the top k retrieval vectors obtained from all five matching techniques into a single retrieval result. This technique will encourage the development of adequate clothes image retrieval and deliver a practical approach for intelligent retrieval methods in the future

REFERENCES

- [1] M. McMaster, C. Nettleton, C. Tom, B. Xu, C. Cao, and P. Qiao, "Risk management: Rethinking fashion supply chain management for multinational corporations in light of the COVID-19 outbreak," *J. Risk Financial Manage.*, vol. 13, no. 8, p. 173, Aug. 2020.
- [2] H. Zhang, Y. Sun, L. Liu, X. Wang, L. Li, and W. Liu, "ClothingOut: A category-supervised GAN model for clothing segmentation and retrieval," *Neural Comput. Appl.*, vol. 32, no. 9, pp. 4519–4530, May 2020.
- [3] S. Park, M. Shin, S. Ham, S. Choe, and Y. Kang, "Study on fashion image retrieval methods for efficient fashion visual search," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 316–319.
- [4] W. Chen, Y. Liu, W. Wang, E. Bakker, T. Georgiou, P. Fieguth, L. Liu, and M. S. Lew, "Deep learning for instance retrieval: A survey," 2021, *arXiv:2101.11282*.
- [5] Y. Sun, "The role of activation function in image classification," in *Proc. Int. Conf. Commun., Inf. Syst. Comput. Eng. (CISCE)*, May 2021, pp. 275–278.
- [6] A. F. Agarap, "Deep learning using rectified linear units (ReLU)," 2018, *arXiv:1803.08375*.
- [7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012, pp. 84–90.
- [8] L. Lu, Y. Shin, Y. Su, and G. Karniadakis, "Dying ReLU and initialization: Theory and numerical examples," 2019, *arXiv:1903.06733*.
- [9] C. Ning, Y. Di, and L. Menglu, "Survey on clothing image retrieval with cross-domain," *Complex Intell. Syst.*, vol. 8, pp. 1–14, May 2022.
- [10] D. Misra, "Mish: A self regularized non-monotonic activation function," 2019, *arXiv:1908.08681*.
- [11] S. Bouktif, A. Fiaz, A. Ouni, and M. A. Serhani, "Optimal deep learning LSTM model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches," *Energies* vol. 11, no. 7, p. 1636, Jun. 2018.
- [12] K. Lin, H.-F. Yang, K.-H. Liu, J.-H. Hsiao, and C.-S. Chen, "Rapid clothing retrieval via deep learning of binary codes and hierarchical search," in *Proc. 5th ACM Int. Conf. Multimedia Retr.*, Jun. 2015, pp. 499–502.
- [13] Y. Lin, P. Ren, Z. Chen, Z. Ren, J. Ma, and M. de Rijke, "Explainable outfit recommendation with joint outfit matching and comment generation," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 8, pp. 1502–1516, Aug. 2020.
- [14] L. Jiao, F. Zhang, F. Liu, S. Yang, L. Li, Z. Feng, and R. Qu, "A survey of deep learning-based object detection," *IEEE Access*, vol. 7, pp. 128837–128868, 2019.
- [15] M. Murtaza, M. Sharif, M. Yasmin, and S. Kadry, "A novel approach of boundary preservative apparel detection and classification of fashion images using deep learning," *Math. Methods Appl. Sci.*, Mar. 2022, doi: 10.1002/mma.8197.
- [16] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9.
- [17] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
- [18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [19] B. Ding, H. Qian, and J. Zhou, "Activation functions and their characteristics in deep neural networks," in *Proc. Chin. Control Decis. Conf. (CCDC)*, Jun. 2018, pp. 1836–1841.
- [20] P. Naveen, "Phish: A novel hyper-optimizable activation function," TechRxiv, 2021, doi: 10.36227/techrxiv.17283824.v2.
- [21] S. Bouguezzi et al., "Improved architecture for traffic sign recognition using a self-regularized activation function: SigmaH," *Vis. Comput.*, vol. 38, pp. 3747–3764, 2022.
- [22] D. Pedamonti, "Comparison of non-linear activation functions for deep neural networks on MNIST classification task," 2018, *arXiv:1804.02763*, doi: 10.48550/arXiv.1804.02763.
- [23] S. R. Dubey, S. K. Singh, and B. B. Chaudhuri, "Activation functions in deep learning: A comprehensive survey and benchmark," *Neurocomputing*, vol. 503, pp. 92–108, 2022, doi: 10.1016/j.neucom.2022.06.111.
- [24] Z. Qiumei, T. Dan, and W. Fenghua, "Improved convolutional neural network based on fast exponentially linear unit activation function," *IEEE Access*, vol. 7, pp. 151359–151367, 2019.
- [25] S. R. Dubey and S. Chakraborty, "Average biased ReLU based CNN descriptor for improved face retrieval," *Multimedia Tools Appl.*, vol. 80, no. 15, pp. 23181–23206, Jun. 2021.
- [26] B. A. Harshanand and A. K. Sangaiah, "Comprehensive analysis of deep learning methodology in classification of leukocytes and enhancement using swish activation units," *Mobile Netw. Appl.*, vol. 25, no. 6, pp. 2302–2320, Dec. 2020.
- [27] L. Skovajsova and I. Mokris, "Dimension reduction in text document retrieval by Hebbian neural network and nonlinear activation functions," in *Proc. 4th IEEE Int. Symp. Logistics Ind. Informat.*, Sep. 2012, pp. 59–64.
- [28] R. Kapoor, D. Sharma, and T. Gulati, "State of the art content based image retrieval techniques using deep learning: A survey," *Multimedia Tools Appl.*, vol. 80, no. 19, pp. 29561–29583, Aug. 2021.
- [29] B. Karlik and A. Vehbi, "Performance analysis of various activation functions in generalized MLP architectures of neural networks," *Int. J. Artif. Intell. Expert Syst.*, vol. 1, no. 4, pp. 111–122, 2015.
- [30] C. Yang, Z. Yang, S. Liao, Z. Hong, and W. Nai, "Triple-GAN with variable fractional order gradient descent method and mish activation function," in *Proc. 12th Int. Conf. Intell. Hum.-Mach. Syst. Cybern. (IHMSC)*, vol. 1, Aug. 2020, pp. 244–247.
- [31] W. Chen, Y. Liu, W. Wang, E. Bakker, T. Georgiou, P. Fieguth, L. Liu, and M. S. Lew, "Deep learning for instance retrieval: A survey," 2021, *arXiv:2101.11282*, doi: 10.48550/arXiv.2101.11282.
- [32] Z.-P. Fan, G.-M. Li, and Y. Liu, "Processes and methods of information fusion for ranking products based on online reviews: An overview," *Inf. Fusion*, vol. 60, pp. 87–97, Aug. 2020.
- [33] Z.-W. Yuan and J. Zhang, "Feature extraction and image retrieval based on AlexNet," in *Proc. SPIE*, Aug. 2016, Art. no. 100330.
- [34] X. Zhang, J. Jia, K. Gao, Y. Zhang, D. Zhang, J. Li, and Q. Tian, "Trip outfits advisor: Location-oriented clothing recommendation," *IEEE Trans. Multimedia*, vol. 19, no. 11, pp. 2533–2544, Nov. 2017.

- [35] A. K. Alhassan and A. A. Alfaki, "Color and texture fusion-based method for content-based image retrieval," in *Proc. Int. Conf. Commun., Control, Comput. Electron. Eng. (ICCCCEE)*, Jan. 2017, pp. 1–6.
- [36] M. Alkhwilani, M. Elmogy, and H. Elbakry, "Content-based image retrieval using local features descriptors and bag-of-visual words," *Int. J. Adv. Comput. Sci. Appl.*, vol. 6, no. 9, pp. 212–219, 2015.
- [37] M. K. Alsmadi, "An efficient similarity measure for content based image retrieval using memetic algorithm," *Egyptian J. Basic Appl. Sci.*, vol. 4, no. 2, pp. 112–122, Jun. 2017.
- [38] S. Jabeen, Z. Mehmood, T. Mahmood, T. Saba, A. Rehman, and M. T. Mahmood, "An effective content-based image retrieval technique for image visuals representation based on the bag-of-visual-words model," *PLoS ONE*, vol. 13, no. 4, Apr. 2018, Art. no. e0194526.
- [39] X. Yang, W. Liu, W. Liu, and D. Tao, "A survey on canonical correlation analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 6, pp. 2349–2368, Dec. 2019.
- [40] M. Haghghat, M. Abdel-Mottaleb, and W. Alhalabi, "Discriminant correlation analysis for feature level fusion with application to multimodal biometrics," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 1866–1870.
- [41] H. Zhang, H. Xu, X. Tian, J. Jiang, and J. Ma, "Image fusion meets deep learning: A survey and perspective," *Inf. Fusion*, vol. 76, pp. 323–336, Dec. 2021.
- [42] J. Amin, M. Sharif, M. Yasmin, H. Ali, and S. L. Fernandes, "A method for the detection and classification of diabetic retinopathy using structural predictors of bright lesions," *J. Comput. Sci.*, vol. 19, pp. 153–164, Mar. 2017.
- [43] P. Perera, V. I. Morariu, R. Jain, V. Manjunatha, C. Wigington, V. Ordonez, and V. M. Patel, "Generative-discriminative feature representations for open-set recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11814–11823.
- [44] H. Müller, W. Müller, D. M. Squire, S. Marchand-Maillet, and T. Pun, "Performance evaluation in content-based image retrieval: Overview and proposals," *Pattern Recognit. Lett.*, vol. 22, no. 5, pp. 593–601, 2001.
- [45] S. Bai, P. Tang, P. H. S. Torr, and L. Jan Latecki, "Re-ranking via metric fusion for object retrieval and person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 740–749.
- [46] Y. Xu, B. Zhao, Y. Zhai, Q. Chen, and Y. Zhou, "Maize diseases identification method based on multi-scale convolutional global pooling neural network," *IEEE Access*, vol. 9, pp. 27959–27970, 2021.
- [47] L. Lu, Y. Shin, Y. Su, and G. E. Karniadakis, "Dying ReLU and initialization: Theory and numerical examples," *Commun. Comput. Phys.*, vol. 28, no. 5, pp. 1671–1706, 2020, doi: 10.4208/cicp.OA-2020-0165.
- [48] D. Misra, "Mish: A self regularized non-monotonic activation function," 2019, *arXiv:1908.08681*, doi: 10.48550/arXiv.1908.08681.
- [49] H. Zhu, L. Yang, J. Fei, L. Zhao, and Z. Han, "Recognition of carrot appearance quality based on deep feature and support vector machine," *Comput. Electron. Agricult.*, vol. 186, Jul. 2021, Art. no. 106185.
- [50] P. K. Sethy, N. K. Barpanda, A. K. Rath, and S. K. Behera, "Deep feature based Rice leaf disease identification using support vector machine," *Comput. Electron. Agricult.*, vol. 175, Aug. 2020, Art. no. 105527.
- [51] A. Krizhevsky, "Learning multiple layers of features from tiny images," Univ. Toronto, Toronto, ON, Canada, Tech. Rep. TR-2009, 2009.
- [52] L. Sun, L. Wang, W. Ding, Y. Qian, and J. Xu, "Feature selection using fuzzy neighborhood entropy-based uncertainty measures for fuzzy neighborhood multigranulation rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 1, pp. 19–33, Jan. 2020.
- [53] F. Afza, M. A. Khan, M. Sharif, S. Kadry, G. Manogaran, T. Saba, I. Ashraf, and R. Damaševićius, "A framework of human action recognition using length control features fusion and weighted entropy-variances based feature selection," *Image Vis. Comput.*, vol. 106, Feb. 2021, Art. no. 104090.
- [54] M.-Q. Tran, M. Elsisli, and M.-K. Liu, "Effective feature selection with fuzzy entropy and similarity classifier for chatter vibration diagnosis," *Measurement*, vol. 184, Nov. 2021, Art. no. 109962.
- [55] Y. Liu, Y. Mu, and K. Chen, "Daily activity feature selection in smart homes based on Pearson correlation coefficient," *Neural Process. Lett.*, vol. 51, pp. 1771–1787, Jan. 2020.
- [56] Y. Ge, R. Zhang, X. Wang, X. Tang, and P. Luo, "DeepFashion2: A versatile benchmark for detection, pose estimation, segmentation and re-identification of clothing images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5337–5345.
- [57] K.-H. Liu, T.-Y. Chen, and C.-S. Chen, "MVC: A dataset for view-invariant clothing retrieval and attribute prediction," in *Proc. ACM Int. Conf. Multimedia Retr.*, Jun. 2016, pp. 313–316.
- [58] T. Krishna, K. McGuinness, and N. O'Connor, "Evaluating contrastive models for instance-based image retrieval," in *Proc. Int. Conf. Multimedia Retr.*, 2021, pp. 471–475.
- [59] J. Ma, X. Jiang, A. Fan, J. Jiang, and J. Yan, "Image matching from handcrafted to deep features: A survey," *Int. J. Comput. Vis.*, vol. 129, pp. 23–79, Aug. 2020.
- [60] X. Liu and X. Di, "TanhExp: A smooth activation function with high convergence speed for lightweight neural networks," *IET Comput. Vis.*, vol. 15, no. 2, pp. 136–150, Mar. 2021.
- [61] F. Zhang and X. Wang, "Object tracking in Siamese network with attention mechanism and Mish function," *Acad. J. Comput. Inf. Sci.*, vol. 4, no. 1, pp. 75–81, 2021.
- [62] K. Sriporn, C.-F. Tsai, C.-E. Tsai, and P. Wang, "Analyzing malaria disease using effective deep learning approach," *Diagnostics*, vol. 10, no. 10, p. 744, Sep. 2020.
- [63] L. Yang, H. Yu, Y. Cheng, S. Mei, Y. Duan, D. Li, and Y. Chen, "A dual attention network based on efficientNet-B2 for short-term fish school feeding behavior analysis in aquaculture," *Comput. Electron. Agricult.*, vol. 187, Aug. 2021, Art. no. 106316.
- [64] R. Sarkar, S. Julai, S. Hossain, W. T. Chong, and M. Rahman, "A comparative study of activation functions of NAR and NARX neural network for long-term wind speed forecasting in Malaysia," *Math. Problems Eng.*, vol. 2019, pp. 1–14, Mar. 2019.
- [65] X. Wang, H. Ren, and A. Wang, "Smish: A novel activation function for deep learning methods," *Electronics*, vol. 11, p. 540, 2022, doi: 10.3390/electronics11040540.
- [66] B. Zheng and Z. Wang, "PATS: A new neural network activation function with parameter," in *Proc. 5th Int. Conf. Comput. Commun. Syst. (ICCCS)*, May 2020, pp. 125–129.
- [67] S. Gkelios, A. Sophokleous, S. Plakias, Y. Boutalis, and S. A. Chatzichristofis, "Deep convolutional features for image retrieval," *Expert Syst. Appl.*, vol. 177, Dec. 2021, Art. no. 114940.
- [68] H. Alshowaish, Y. Al-Ohali, and A. Al-Nafjan, "Trademark image similarity detection using convolutional neural network," *Appl. Sci.*, vol. 12, no. 3, p. 1752, Feb. 2022.
- [69] V. Bhandi, "COVID-19 X-ray image retrieval using deep convolutional neural networks," *Amer. J. Eng. Res.*, vol. 9, no. 7, pp. 47–55, 2020.
- [70] R. Khwildi, A. Ouled Zaid, and F. Dufaux, "query-by-example HDR image retrieval based on CNN," *Multimedia Tools Appl.*, vol. 80, no. 10, pp. 15413–15428, Apr. 2021.
- [71] M. Parola, A. Nannini, and S. Poleggi, "Web image search engine based on LSH index and CNN Resnet50," 2021, *arXiv:2108.13301*.
- [72] H. Lei, Y. Yi, Y. Li, G. Luo, and M. Wang, "A new clothing image retrieval algorithm based on sketch component segmentation in mobile visual sensors," *Int. J. Distrib. Sensor Netw.*, vol. 14, no. 11, 2018, Art. no. 1550147718815627.
- [73] Z.-Q. Cheng, X. Wu, Y. Liu, and X.-S. Hua, "Video2Shop: Exact matching clothes in videos to online shopping images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4048–4056.
- [74] Y. Deng, X. Lin, R. Li, and R. Ji, "Multi-scale gem pooling with N-pair center loss for fine-grained image search," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2019, pp. 1000–1005.
- [75] A. Sanakoyeu, V. Tschernetzki, U. Buchler, and B. Ommer, "Divide and conquer the embedding space for metric learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 471–480.
- [76] A.-A. Liu, T. Zhang, D. Song, W. Li, and M. Zhou, "FRSFN: A semantic fusion network for practical fashion retrieval," *Multimedia Tools Appl.*, vol. 80, no. 11, pp. 17169–17181, May 2021.



MARRYAM MURTAZA received the B.S. and M.S. degrees from COMSATS University Islamabad, Wah Campus, Pakistan, in 2008 and 2011, respectively, where she is currently pursuing the Ph.D. degree. Her research interests include computer vision, pattern recognition, artificial intelligence, medical image processing, and machine learning. She has so far 12 research publications in highly impact factor, international journals. She got three research productivity awards from COMSATS University. She was also engaged as a Lecturer in different universities, from 2011 to 2021. She has supervised more than 30 undergraduate students.



MUHAMMAD SHARIF (Senior Member, IEEE) received the Ph.D. degree. He is currently an Associate Professor at COMSATS University Islamabad, Wah Campus, Pakistan. He worked for one year at Alpha Soft U.K.-based Software House, in 1995. He is also an OCP in the Developer Track. He has been the teaching profession, since 1996. His research interests include medical imaging, biometrics, computer vision, machine learning, and agriculture plants. He is being awarded

COMSATS Research Productivity Award, from 2011 to 2017. He served in TPC for IEEE FIT (2014–2019) and currently serving as an Associate Editor for IEEE ACCESS, a guest editor of special issues, and a reviewer for well-reputed journals. He also headed the department, from 2008 to 2011, and achieved the targeted outputs. He has more than 285 research publications in IF, SCI, and ISI journals and in national and international conferences, and obtained more than 550 impact factor. He has supervised/co-supervised ten Ph.D. (CS) and more than 90 M.S. (CS) theses to date.



MUSSARAT YASMIN received the Ph.D. degree. She is currently an Associate Professor at COMSATS University Islamabad, Wah Campus, Pakistan. Her area of specialization is image processing. She has been in education field, since 1993. She has so far more than 60 research publications in IF, SCI, and ISI journals and in national and international conferences. A number of undergraduate projects are complete under her supervision. She is currently supervising five Ph.D. (CS)

students and a good number of M.S. students. She is an active member of university research group “Computer Vision and Intelligent Systems.” She is the Gold Medalist in M.S. (CS) from IQRA University, Pakistan. She has been getting COMSATS Research Productivity Award, since 2012. Her research interests include neural networks, algorithms design and analysis, machine learning, and image processing.



MUHAMMAD FAYYAZ received the Ph.D. degree in computer science from COMSAT University Islamabad, Wah Campus, Pakistan, in September 2021. He is currently working as an Assistant Professor at the Department of Computer Science, FAST—National University of Computer and Emerging Sciences (NUCES), Chiniot-Faisalabad Campus, Pakistan. Before this, he worked as a Lecturer at the Department of Computer Science, CUI, from September 2014 to

January 2022. He remained a very active research member of fields related to pedestrian analysis, computer vision, pattern recognition, and machine/deep learning. He served as a TPC member, the organizing co-chair, and a reviewer for many leading international conferences and journals.



SEIFEDINE KADRY (Senior Member, IEEE) received the bachelor’s degree from Lebanese University, in 1999, the joint M.S. degree from the University of Reims, France, and the EPFL, Lausanne, in 2002, the Ph.D. degree from Blaise Pascal University, France, in 2007, and the H.D.R. degree from the University of Rouen Normandy, in 2017. He is currently a Full Professor in data science with the Noroff University College, Norway. He is also an ABET Program

Evaluator in computing and an ABET Program Evaluator in engineering technology. His current research interests include data science, education using technology, system prognostics, stochastic systems, and probability and reliability analysis.



MI YOUNG LEE (Member, IEEE) received the M.S. and Ph.D. degrees from the Department of Image and Information Engineering, Pusan National University. Currently, she is a Research Professor at Sejong University and provide research services at the Intelligent Media Laboratory (IMLab). She is broadly working in artificial intelligence, computer vision, image processing, and energy informatics. Her particular research interests include video summarization,

movies data analysis, electrical energy forecasting, and video retrieval. She has published several novel contributions in these areas in reputed journals and peer-reviewed conference proceedings, including IEEE ACCESS, *Sensors* (MDPI), *Multimedia Tools and Applications* (Springer), and International Joint Conference on Neural Networks, in 2020. She has carried out several research projects successfully and is a principal investigator of several ongoing research projects under the supervision of Korean Government and has filed more than 13 patents during her career.

...