Proceedings of the International Conference on Sustainable Communication Networks and Application (ICSCNA 2023)
IEEE Xplore Part Number: CFP23DW8-ART; ISBN: 979-8-3503-1398-7

Modelling of Firefly Algorithm with Densely Connected Networks for Near-Duplicate Image Detection System

Srinidhi Sundaram
Research Scholar Department of Information Technology, Annamalai University, Annamalai Nagar - 608002.
gayathri_research@yahoo.com

Kamalakkannan Somasundaram
Associate Professor, Department of Information Technology, School of Computing Sciences VISTAS (Vels Institute of Sciences Technology & Advanced Studies), Pallavaram, Chennai – 600 117.
kannan.scs@velsuniv.ac.in

S. Jothilakshmi
Associate Professor, Department of Information Technology, Annamalai University, Annamalai Nagar - 608002.
jothi.sekar@gmail.com

Sasikala Jayaraman, Associate Professor, Department of Information Technology, Annamalai University, Annamalai Nagar - 608002.
sasikala.au@gmail.com

P. Dhanalakshmi, Professor, Dept of Computer Science & Engineering, Annamalai University, Annamalai Nagar - 608002.
abi_dhana500@rediffmail.com

2023 International Conference on Sustainable Communication Networks and Application (ICSCNA) | 979-8-3503-1398-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICSCNA58489.2023.10370117

Abstract—**Near-duplicate image detection is the way of detecting and flagging images that are highly similar to each other but not identical. It is a crucial task in different fields, including search engines, content management, and copyright enforcement, as it helps in content and organization deduplication. To achieve this, techniques and algorithms are employed that calculate similarity metrics, compare features, or analyze image content to define the degree of resemblance between images. Near-duplicate image detection can include approaches such as feature extraction, machine learning, and perceptual hashing to effectively manage and identify similar images in large datasets, which offer benefits in content retrieval and storage optimization. Therefore, this study presents a new firefly algorithm with deep learning-based near-duplicate image detection (FFADL-NDID) technique. Initially, median filtering (MF) approach is used to preprocess the input images. The proposed FFADL-NDID technique exploits the robust feature extraction abilities of DenseNet, a pre-trained DL model for capturing complex visual patterns from database and query images. In addition, the FFA is applied to carry out the hyperparameter tuning, optimizing the system for superior performance. This synergistic fusion enhances the overall efficiency of near-duplicate image detection by successfully searching the hyperparameter space for optimal configurations. Finally, the FFADL-NDID framework applies Euclidean distance-based similarity matching processes, which detects the near-duplicate images significantly. The simulation analysis of the FFADL-NDID method is tested on multiple datasets and the outcomes show its promising performance over other DL models in terms of different measures.**

Keywords— Near duplicate image detection; Computer vision; Deep learning; Firefly algorithm; Hyperparameter tuning

I. INTRODUCTION

Due to the fast growth of mobile Internet and portable multi-media sensor technologies, the quantity of multimedia information is increasing at a high speed [1]. Therefore, there are numerous near-duplicate images (NDI) present among the extensive data. NDI is normally referred to as images that comprise similar objects or scenes, in different standpoints, camera settings illumination levels etc. [2], or images attained through reediting of the actual images however, not restricted to varying watermarking, cropping, contrast, tone, and rotating [3]. Currently, automated NDI pair identification by implementing pattern recognition and Computer Vision (CV) technologies is engaging extensive focus since it can be a high possible standard in the application of autonomous vehicle driving, misleading image recognition [4], management of hardware device storage, and image copyright violations identification. Standard hand-crafted local features-based techniques like the extensively implemented Fisher Vector [5], and scale-invariant feature transform (SIFT), attained the image-level features via combining approaches namely vector of locally aggregated descriptors (VLAD), histogram of oriented gradients (HOG) descriptor, and so on. These algorithms can be affected by the issues of confined representation capabilities and complex extraction stages [6].

Recently, deep learning (DL) methods like convolutional neural networks (CNN) have accomplished wide focus in the domain of CV [7]. By employing deep neural networks (DNNs) to describe images and incorporating various similarity measures for implementing this task, CNN-based algorithms could obtain excellent performance [8]. In this view, a few research workers tend to utilize feature extraction from CNNs rather than handcrafted features for the responsibilities of content-based image retrieval (CBIR) or NDI detection [9]. In this study, it is established that CNN-based features attain higher performance than the standard handcrafted features. Generally, the present CNN-based features are separated into two types namely local and global CNN features [10].

This study presents a new firefly algorithm with deep learning-based near-duplicate image detection (FFADL-NDID) technique. Initially, median filtering (MF) approach is used to preprocess the input images. The proposed FFADL-NDID technique exploits the robust feature extraction abilities of DenseNet, a pre-trained DL model for capturing complex visual patterns from database and query images. In addition, the FFA is applied to carry out the hyperparameter tuning, optimizing the system for superior performance. This synergistic fusion enhances the overall efficiency of near-duplicate image detection by successfully searching the hyperparameter space for optimal configurations. Finally, the FFADL-NDID framework applies Euclidean distance (ED) based similarity matching processes, which detects the near-duplicate images significantly. The simulation evaluation of the FFADL-NDID method is examined on multiple datasets.

II. Related Works

Islam et al. [11] developed a Dual Shallow Siamese Network (DSSN) for the tasks. Initially, the networks were trained by employing negative and positive sets of images. Then, the trained method was utilized for detecting the dissimilarity between query and gallery images. The similarity measure of the Siamese network is further combined through weighted averaging. In [12], a spatial transformer-based CNN technique was introduced. Particularly, this analysis primarily offers a related CNN model. Additionally, a spatial transformer comparing CNN method is also presented by integrating a spatial transformer component to the related CNN algorithm. Mehta et al. [13] implemented an adapted DL algorithm for identifying NDI forensics attacks depending on removed wavelet Haar features. Dissimilar traditional DL methods, a wavelet decomposed preprocessing layer has been employed beforehand the DL techniques, and an SVM method was utilized in the classification phase of an adapted DL algorithm.

In [14], a new DL model depends on the flask that uniquely integrates the DL methods on CNN and Generative Adversarial Networks (GAN) was developed. This algorithm includes GAN and is verified with LSTM. Comparison of Pixel wise is executed utilizing KNN over the frame, Singular Value Decomposition (SVD) and K-means clusters. In [15], an efficient CBIR depends on CNN (IRB_CNN) method was introduced. Now, feature origination could be performed by employing pretrained CNN in the absence of fully connected (FC) layers. The training data enhancement technique was exploited to increase training images. The effectiveness of CNN has been improved through meta-heuristic optimization. Next, CNN as a classifier is utilized for classifying and estimating this developed CNN.

Fisichella [16] presented an innovative effective approach for NDI identification by applying a deep Siamese coding NN, which can be the ability to remove proficient features from images, advantageous for making LSH indices. Wide-ranging research implemented on 2 benchmark databases affirms the performance of developed deep Siamese coding network and forecast method. Phan et al. [17] suggested an efficacious technique for feature extraction on large video databases employing DL methods and implemented an incorporation of 3 techniques namely object detection with DL methods, automated speech recognition (ASR) and optical character recognition (OCR). This offers 3 network architectures designed from networks of Single Shot Detector MobileNetV2, Fast RCNN Inception ResNet_V2, and Fast RCNN ResNet.

III. Proposed Model

In this study, an automated near-duplicate image detection using the FFADL-NDID technique is designed. The primary goal of the FFADL-NDID method is to exploit the DL model for the identification of near-duplicate images effectually. The FFADL-NDID technique comprises four subprocesses namely MF-based pre-processing, ED-based similarity matching, FFA-based hyperparameter tuning, and DenseNet feature extraction process. Fig. 1 portrays the entire procedure of FFADL-NDID approach.

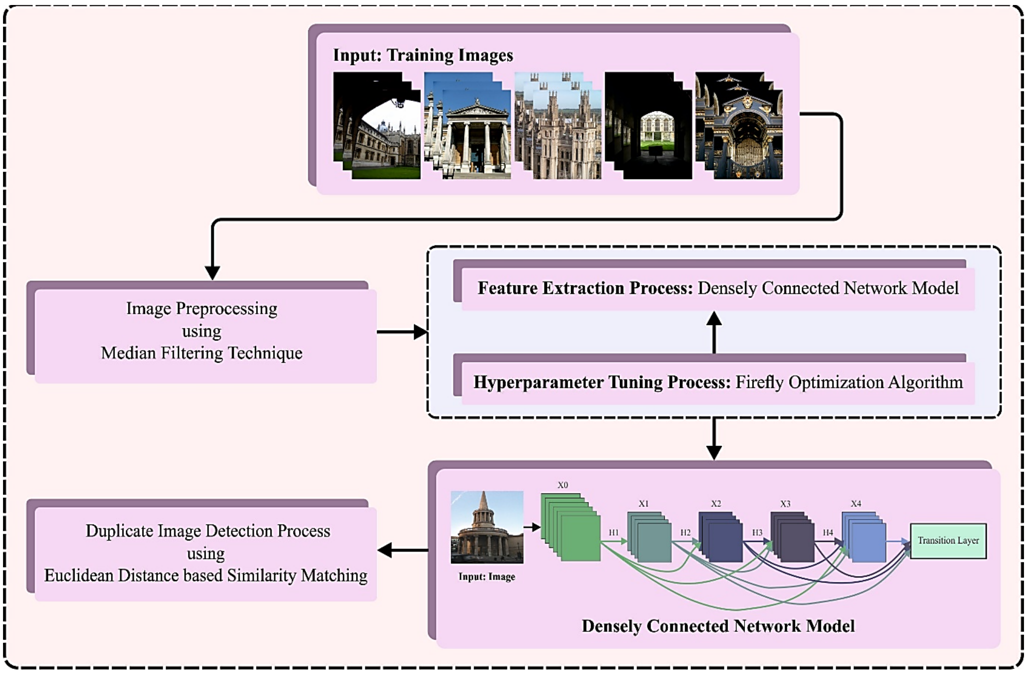


Fig. 1. Complete process of FFADL-NDID method

A. Image Pre-processing

Initially, the MF model is applied. It is a classical image preprocessing technique used to enhance image quality and diminish noise. It works by switching all the pixel values with the median values within the specified window or neighbourhood.

- Neighborhood Selection: For all the pixels in the image, a rectangular or square neighborhood (kernel or window) of definite size is centred on that pixel. The neighborhood size is a critical parameter that affects the preservation of image details and the degree of noise reduction.
- Sorting Pixel Values: Within the selected neighborhood, each pixel value is gathered and arranged in ascending sequence. In this sorted list, the median value is the middle value. If there is even number of values, the median is generally evaluated as the average of the two middle values.
- Replacement: The pixel value at the neighborhood centre is replaced by the estimated median value. This procedure is reiterated for all the pixels in an image.

B. Feature Extraction

For the extraction of useful feature vectors, the DenseNet model is used. It is a prevalent fact that the extraction of the discriminatory aspects becomes easy with the deepness of the network [18]. On the other hand, the initial problem is regarded as the vanishing gradient initiated by the increase of the network layer. Even though the structure of this network is not the same, the objective is to establish the shortest path between layers. DenseNet concatenating each layer by ensuring the maximal data communication between layers.

DenseNet primarily includes transition and dense blocks. The layout of 1 dense block. It is a four-layer dense block with the rate of growth of $k = 4. H(\cdot)$ and is a compound function of 3 successive processes namely 3×3 Conv, ReLU and BN layers. All the layers obtain input from the prior layer and pass it on its feature map to the following layers in feed- forward fashion.

In DenseNet, the features of l^{th} layer is represented by x_l:

$$x_l = H([x_0, x_1, \cdots x_{l-1}]) \quad (1)$$

Where $[x_0, x_1, \cdots x_{l-1}]$ represents the concatenation of feature map. The amount of output channels dramatically increases for each concatenating. The transition block is introduced to control the model complexity which changes the amount of channels by using 1×1 Conv and splits the width and length of the input. Backbone network of DenseNet is used to establish the linking of dissimilar layers and improves the flow of the feature; therefore, it has stronger feature learning capability.

C. Hyperparameter Tuning

The FFA is used in this study for the hyperparameter selection of the DenseNet architecture. FFA is a bio- inspired metaheuristic approach inspired by the social behaviors of fireflies (FFs) that originate in the tropical zone [19]. Mainly, FF's produce different kinds of flashing patterns to find, communicate and search for their mating partners. The concept of FFA focuses on two significant problems, viz., how the attraction is formulated and how the intensity of light is to be varied. For simplicity, the attraction of FFs is evaluated by the brightness which can be further related to the objective function. For maximization problems, at a certain location, the brightness I of FFs can be considered as $(x) \propto f(x)$. But the attraction β is relative which implies it should be visualized by the other FFs in the region. Thus, it varies with the modification in the distance between the FFs. According to the basic laws of physics, the luminous intensity and attraction must be decreased with the increase in distance to the source that variation of light concentration and attraction must be monotonically reducing function. The cumulative effects of inverse square law and absorption are described as follows:

$$I = I_0 \exp(-\gamma r^2) \quad (2)$$

In Eq. (2), I denotes the light intensity, I_0 indicates the initial luminous intensity and γ shows the coefficient that accounts for different levels of light absorption factor. The attraction of FFs is proportional to the light intensity visualised by other FFs, hence the attraction β is expressed by the following equation:

$$\beta = \beta_0 \exp\ (-\gamma r^2) \quad (3)$$

In Eq. (3), β_0 denotes the attraction at distance $r = 0$ and γ shows the light absorption coefficient. The distance r_{ij} between i^{th} and j^{th} FFs positioned at X_i and X_j, correspondingly, is defined by the Euclidean norm, and movement of lesser bright i^{th} FFs towards bright FF j^{th} is defined as follows:

$$x_i = x_i + \beta_0^{-\gamma r_{ij}^2}(x_j - x_i) + \alpha\left(rand - \frac{1}{2}\right) \quad (4)$$

In Eq. (4), third and second term is due to randomization with the arbitrary vector parameters and attraction.

Algorithm 1: Pseudocode of FFA
Objective Function j(X), $X = (x_1, x_2, \dots x_d)$
Create populace of n FFs, X_i, $i = 1, 2, \dots, n$
Light concentration I_i at X_i is defined by employing $f(X_i)$
Determine the light absorption coefficient γ
While ($t <$ MaxGeneration)
For $i = 1{:}n$, each n FFs
For $j = 1{:}n$, each n FF (inner loop)
If ($I_i < I_j$),
Move i^{th} FFs towards j^{th} FF according to Eq. (4)
End if
Differ attraction with distance i through exp $[-\gamma\mu]$
End for j
End for i
Rank the FFs and find the global optimum solution
End while
Post- process the outcomes

D. Similarity-based Matching Process

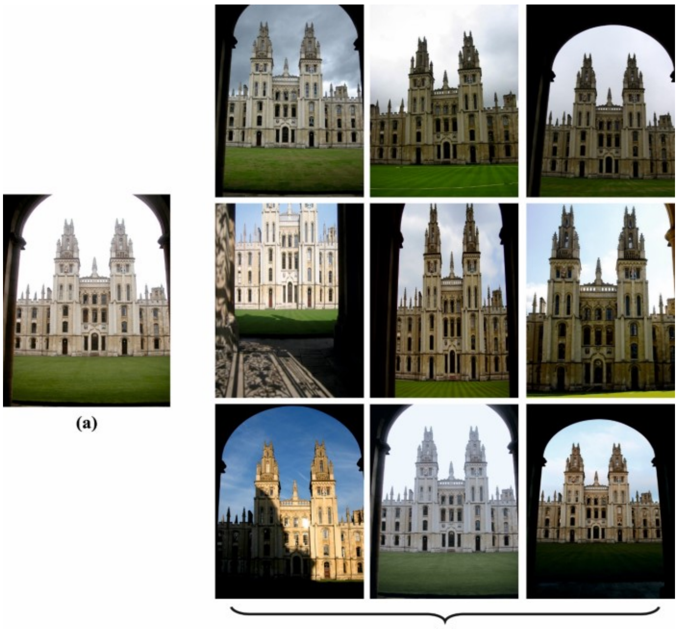
At the final stage, the ED-based similarity matching process is performed to detect the near duplicate images. In near-duplicate image detection, ED-based similarity matching plays a major role in recognizing images that show high amount of visual resemblance [20]. Here, feature vectors extracted from images are compared by quantifying the ED between them in multi-dimensional feature space. Once the ED between the feature vectors of two images is smaller, it specifies a large similarity, indicating that the images likely share common visual patterns or content. This method effectively distinguishes near-duplicate from diverse images by establishing a threshold for the ED, making it an effective and robust technique for tasks including content organization, content deduplication, and copyright enforcement in largescale image database. Its ability, simplicity, and efficiency to capture subtle visual correlation make it a cornerstone in the recognition field of near-duplicate image.

IV. Performance Validation

To examine the near-duplicate image recognition analysis of the FFADL-NDID technique, a detailed set of simulations was performed on three datasets namely Oxford5k [21], the Holidays [22], and the Paris6k datasets [23]. Fig. 2 demonstrates the sample images.

Fig. 2. Sample Images

Fig. 3 shows the sample recognition results of the FFADL-NDID technique. Fig. 3a illustrates the input query image. Followed by, Fig. 3b depicts the near-detected images generated by the FFADL-NDID technique. The figures highlighted that the FFADL-NDID technique properly detected the near-duplicated images.



(a)

(b)

Fig. 3. Images a) Query b) Near Detected

Table 1 represents an overall near-image detection performance of the FFADL-NDID technique with existing DL models on three distinct datasets [24, 25]. The experimental results stated that the FFADL-NDID technique offers enhanced performance with maximal values of mAP.

TABLE I

MAP OUTCOME OF FFADL-NDID METHODOLOGIES WITH OTHER DL APPROACHES UNDER THREE DATASETS

Methodology	Oxford5k (mAP)	Holidays (mAP)	Paris6k (mAP)
max-pooling	52.40	71.10	53.75
VLAD-CNN	55.80	83.60	58.30
SPoC	58.90	80.20	59.01
R-MAC	66.90	85.20	83.00
CroW	68.40	85.10	76.50
NDIDS-CMS	71.50	88.60	77.20
FFADL-NDID	90.02	93.16	91.20

Fig. 4 exemplifies a comparative mAP outcome of the FFADL-NDID method with existing systems on the Oxford5k database. The results indicate that the max-pooling, VLAD-CNN, and SPoC models obtain reduced mAP values of 52.40%, 55.80%, and 58.90% respectively. Along with that, the R-MAC, CroW, and NDIDS-CMS models offer slightly improved mAP values of 66.90%, 68.40%, and 71.50% respectively. However, the FFADL-NDID technique gains maximum performance with higher mAP of 90.02%.

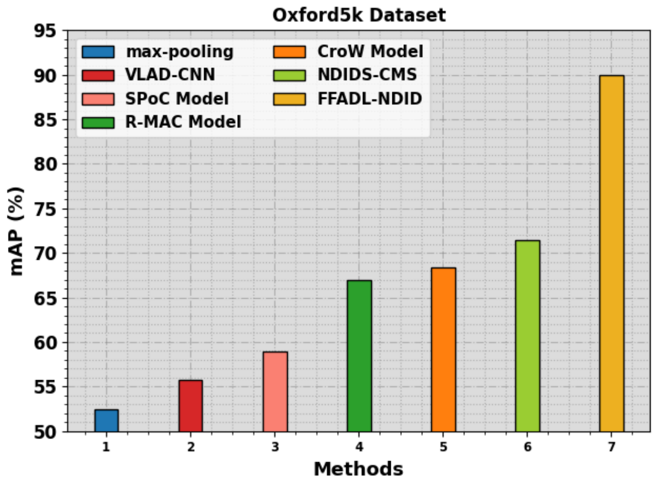


Fig. 4. mAP outcome of FFADL-NDID algorithm on Oxford5k dataset

The ROC results of the FFADL-NDID technique with current approaches on the Oxford5k dataset are reported in Fig. 5. The simulated values show that the FFADL-NDID method reaches improved ROC values over other models.

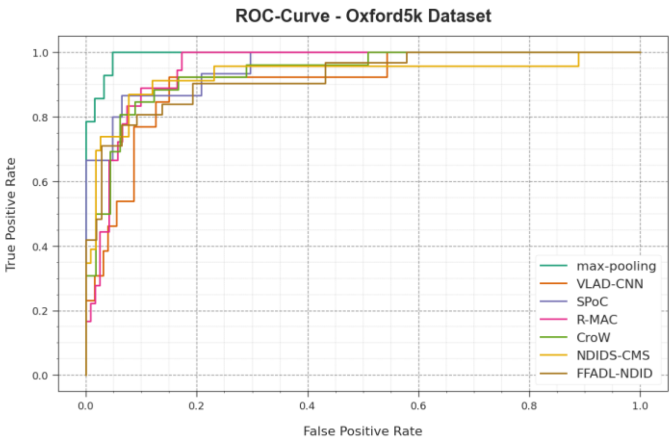


Fig. 5. ROC curve of FFADL-NDID algorithm on Oxford5k database

Fig. 6 shows a relative mAP output of the FFADL-NDID methodology with recent systems on the Holidays database. The simulated values specify that the max-pooling, VLAD-CNN, and SPoC systems get decreased mAP values of 71.10%, 83.60%, and 80.20%. Also, the R-MAC, CroW, and NDIDS-CMS techniques give moderately increased mAP values of 85.20%, 85.10%, and 88.60% individually. However, the FFADL-NDID model achieves excellent performance with greater mAP of 93.16% correspondingly.

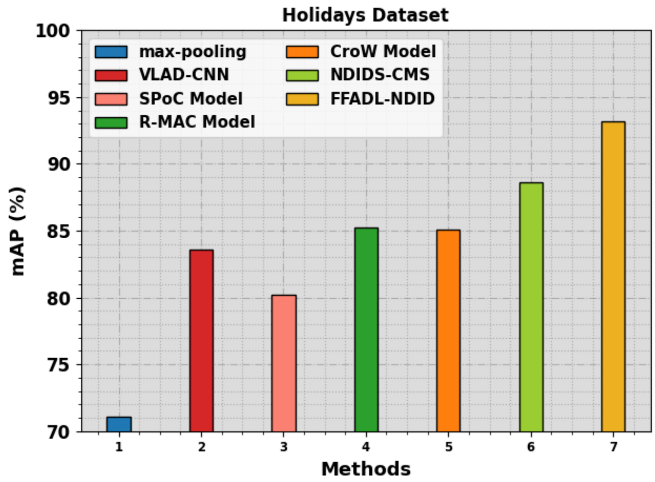


Fig. 6. mAP outcome of FFADL-NDID algorithm on Holidays database

The ROC analysis of the FFADL-NDID methodology with recent algorithms on the Holidays database is demonstrated in Fig. 7. The simulated values displayed that the FFADL-NDID system achieves enriched ROC values over other techniques.

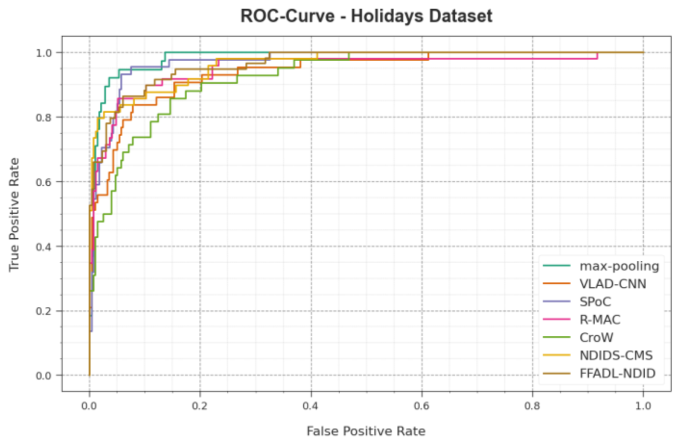


Fig. 7. ROC curve of FFADL-NDID algorithm on Holidays database

Fig. 8 represents a comparison mAP analysis of the FFADL-NDID method with recent techniques on the Paris6k database. The simulated values show that the max-pooling, VLAD-CNN, and SPoC methods obtain reduced mAP values of 53.75%, 58.30%, and 59.01% individually. Moreover, the R-MAC, CroW, and NDIDS-CMS methodologies offer slightly enhanced mAP values of 83.00%, 76.50%, and 77.20%. Also, the FFADL-NDID model attains superior performance with better mAP of 91.20% correspondingly.

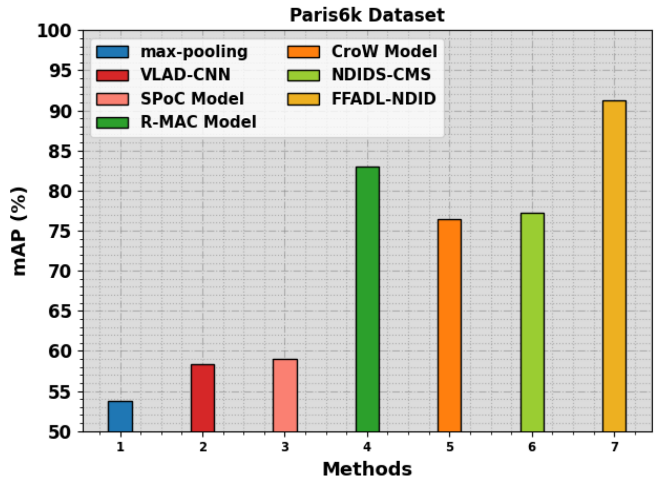


Fig. 8. mAP outcome of FFADL-NDID algorithm on Paris6k database

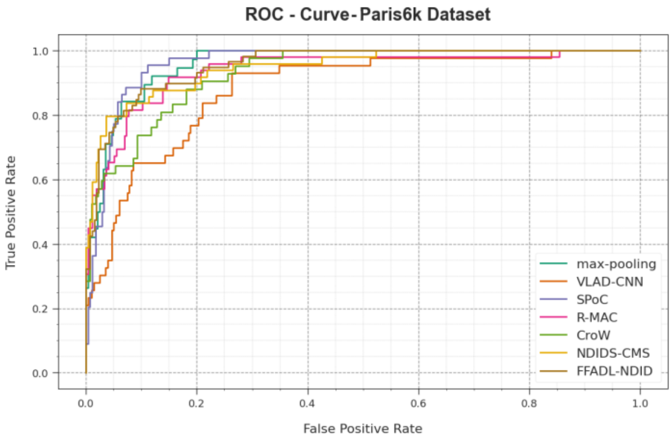


Fig. 9. ROC curve of FFADL-NDID algorithm on Paris6k database

The ROC evaluation of the FFADL-NDID method with current methods on the Paris6k database is shown in Fig. 9. The simulated values represented that the FFADL-NDID method attains increased ROC values over other systems. These simulated values highlighted the enhanced near-image detection performance of the FFADL-NDID technique.

V. CONCLUSION

In this study, an automated near-duplicate image detection using the FFADL-NDID technique is designed. The primary goal of FFADL-NDID method is to exploit the DL model for the identification of near-duplicate images. The FFADL-NDID technique comprises four subprocesses namely MF-based pre-processing, ED-based similarity matching, FFA-based hyperparameter tuning, and DenseNet feature extraction process. The proposed FFADL-NDID technique exploits the robust feature extraction abilities of DenseNet, a pre-trained DL model for capturing complex visual patterns from database and query images. In addition, the FFA is applied to carry out the hyperparameter tuning, optimizing the system for superior performance. This synergistic fusion enhances the overall efficiency of near-duplicate image detection by successfully searching the hyperparameter space for optimal configurations. Finally, the FFADL-NDID framework applies ED-based similarity matching processes, which detect the near-duplicate images significantly. The simulation evaluation of the FFADL-NDID method is examined on multiple datasets and the outcomes show its promising performance over other DL models in terms of different measures.

REFERENCES

[1] Net, F., Folia, M., Casals, P. and Gómez, L., 2023, August. Transductive Learning for Near-Duplicate Image Detection in Scanned Photo Collections. In *International Conference on Document Analysis and Recognition* (pp. 3-17). Cham: Springer Nature Switzerland.

[2] Bakhteev, O., Antiplagiat, D.C., Kuznetsova, R., Khazov, A., Ogaltsov, A., Safin, K., Gorlenko, T., Suvorova, M., Ivahnenko, A., Botov, P. and Chekhovich, Y., 2021. Near-duplicate handwritten document detection without text recognition. In *Komp'juternaja Lingvistika i Intellektual'nye Tehnologii* (pp. 47-57).

[3] Staino, A., Suwalka, A., Mitra, P. and Basu, B., 2022. Real-time detection and recognition of railway traffic signals using deep learning. *Journal of big data analytics in transportation*, *4*(1), pp.57-71.

[4] Sundaram, S., Kamalakkannan, S. and Jayaraman, S., 2023. Seagull optimization-based near-duplicate image detection in large image databases. *The Imaging Science Journal*, pp.1-13.

[5] Singh, A., Chhabra, S., Majumdar, P., Singh, R. and Vatsa, M., 2022, May. Mannet: A Large-Scale Manipulated Image Detection Dataset And Baseline Evaluations. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1780-1784). IEEE.

[6] Zhang, C., Liu, G. and Xiao, X., 2023. Cross-media correlation learning for web video event mining with integrated text semantics and network structural information. *Neural Computing and Applications*, pp.1-17.

[7] Wang, Z., Zuo, C. and Deng, D., 2022, June. Txtalign: efficient near-duplicate text alignment search via bottom-k sketches for plagiarism detection. In *Proceedings of the 2022 International Conference on Management of Data* (pp. 1146-1159).

[8] Conrad, R. and Narayan, K., 2023. Instance segmentation of mitochondria in electron microscopy images with a generalist deep learning model trained on a diverse dataset. *Cell Systems*, *14*(1), pp.58-71.

[9] Cheng, H., Wang, P. and Qi, C., 2021. CNN retrieval based unsupervised metric learning for near-duplicated video retrieval. *arXiv preprint arXiv:2105.14566*.

[10] Zancato, L., Achille, A., Liu, T.Y., Trager, M., Perera, P. and Soatto, S., 2023. Train/test-time adaptation with retrieval. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 15911-15921).

[11] Islam, S.M., Joardar, S. and Sekh, A.A., 2023. DSSN: dual shallow Siamese network for fashion image retrieval. *Multimedia Tools and Applications*, *82*(11), pp.16501-16517.

[12] Zhang, Y., Zhang, S., Li, Y. and Zhang, Y., 2021. Single-and cross-modality near duplicate image pairs detection via spatial transformer comparing CNN. *Sensors*, *21*(1), p.255.

[13] Mehta, P., Singh, M.K. and Singha, N., 2022. Near-duplicate image detection based on wavelet decomposition with modified deep learning model. *Journal of Electronic Imaging*, *31*(2), pp.023017-023017.

[14] Panneerselvam, K., Mahesh, K., Josephine, V.L. and Kumar, A.R., 2023. Effective and Efficient Video Compression by the Deep Learning Techniques. *Computer Systems Science & Engineering*, *45*(2).

[15] Ahmad, F. and Ahmad, T., 2021. Content based image retrieval system based on deep convolution neural network model by integrating three-fold geometric augmentation. *Optical Memory and Neural Networks, 30*, pp.236-249.

[16] Fisichella, M., 2022. Siamese coding network and pair similarity prediction for near-duplicate image detection. *International Journal of Multimedia Information Retrieval*, *11*(2), pp.159-170.

[17] Phan, T.C., Phan, A.C., Cao, H.P. and Trieu, T.N., 2022. Content-Based Video Big Data Retrieval with Extensive Features and Deep Learning. *Applied Sciences*, *12*(13), p.6753.

[18] Yang, L., Chen, G. and Ci, W., 2023. Multiclass objects detection algorithm using DarkNet-53 and DenseNet for intelligent vehicles. *EURASIP Journal on Advances in Signal Processing*, *2023*(1), p.85.

[19] Arora, S. and Singh, S., 2017. Node localization in wireless sensor networks using butterfly optimization algorithm. *Arabian Journal for Science and Engineering*, *42*, pp.3325-3335.

[20] Liu, Z. and Yu, X., 2022. Infringing Video Retrieval via Decoupled Representations and Mutual Information Maximization.

[21] Philbin, J.; Chum, O.; Isard, M.; Sivic, J.; Zisserman, A. Object retrieval with large vocabularies and fast spatial matching. In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; pp. 1–8.

[22] Jegou, H.; Douze, M.; Schmid, C. Hamming embedding and weak geometric consistency for large scale image search. In Proceedings of the European Conference on Computer Vision, Marseille, France, 12–18 October 2008; pp. 304–317.

[23] Philbin, J.; Chum, O.; Isard, M.; Sivic, J.; Zisserman, A. Lost in quantization: Improving particular object retrieval in large scale image databases. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008; pp. 1–8.

[24] Zhang, Y.; Zhang, S.; Li, Y.; Zhang, Y. Single- and Cross-Modality Near Duplicate Image Pairs Detection via Spatial Transformer Comparing CNN. Sensors 2021, 21, 255. https://doi.org/10.3390/s21010255

[25] Zhou, Z., Lin, K., Cao, Y., Yang, C.N. and Liu, Y., 2020. Near-duplicate image detection system using coarse-to-fine matching scheme based on global and local CNN features. *Mathematics*, *8*(4), p.644.