

# DMobile-ELA: Digital Image Forgery Detection via Cascaded Atrous MobileNet and Error Level Analysis

Karma M.Fathalla<sup>1</sup>, Malak Sowelem<sup>2</sup>, Radwa Fathalla<sup>3</sup>

Computer Engineering Dept., College of Engineering, AAST, Alexandria, Egypt<sup>1,2</sup>  
Computer Science Dept., College of Computing and Information Technology, AAST, Alexandria, Egypt<sup>3</sup>

**Abstract**—With the current developments in technology, not only has digital media become widely available, the editing and manipulation of digital media has become equally available to everyone without any prior experience. The need for detecting manipulated images has grown immensely as it can now cause false information in news media, forensics, and daily life of common users. In this work, a cascaded approach DMobile-ELA is presented to ensure an image’s credibility and that the data it contains has not been compromised. DMobile-ELA integrates Error Level Analysis and MobileNet-based classification for tampering detection. It was able to achieve promising results compared to the state of the art on CASIAv2.0 dataset. DMobile-ELA has successfully reached a training accuracy of 99.79% and a validation accuracy of 98.48% in detecting image manipulation.

**Keywords**—Tampering detection; MobileNet; error level analysis; CASIAv2.0

## I. INTRODUCTION

Digital imaging use has recently prevailed in various domains, starting from social networks [1], through medical diagnostics [2] till reaching its use as digital forensics court evidence [3]. Coupled with its use in different critical fields, technology advancements has led to the ease of digital image manipulation and forgery[4].

Image tampering and forgery include a wide range of types such as copy-move, splicing, retouching and image morphing [5]. Copy-move forgery includes copying a piece of the same picture and moving it to cover another part of the image, while splicing involves copying a part of an image to place it in another image. Retouching often involves changes in shape, color and texture of image parts to improve its visual and technical quality, whereas image morphing perform images interpolation to create an image blend. Recently, Generative Adversial Networks (GANs) made it possible to create full face fake images and media using DeepFake technology[6].

The wide availability of editing and enhancement tools may encourage the malicious use of such tools in criminal acts. Such possibility raises public concern and demand for verifying the originality of the images. Hence, effective approaches are required for detecting image forgery [7]. Forgery detection revolves around the recognition of image manipulation and authenticity validation. Active techniques such as digital signatures and watermarking can be used, in addition to passive detection techniques [7].

In this study, a passive forgery detection approach DMobile-ELA is proposed to automatically detect copy-move and splicing image edits with high accuracy. A dilated

modified MobileNet architecture is presented to determine whether an image is authentic or tampered. Error Level Analysis (ELA) is used to preprocess the investigated image at different compression levels before being input to the Dilated-MobileNet. The proposed approach has the advantages of encompassing a light weight architecture suitable for mobile device use. Also, the atrous modification allows the network to capture larger spatial context; which increases its ability to reconstruct more complex edge structures. In addition, the adopted ELA preprocessing enables the detection of the tampered areas easily, due their characteristic aspects in the ELA representation.

This paper is organized as follows: a background on deep learning is given in Section II covering two of the most popular architectures VGG16 and ResNet-50 to allow further comparison. In Section III , a briefing on the related studies will be provided before presenting DMobile-ELA forgery detection system in Section IV. The experimental setup and results will be discussed in Section V. Finally, the conclusions will be drawn in Section VI.

## II. BACKGROUND

### A. Deep Neural Network

Deep learning or Deep neural network (DNN) belongs to the class of machine learning, which models high level abstractions in the data with multiple nonlinear transformations [8]. DNN is a subclass of neural networks requiring large volumes of data to increase the efficiency of the training processes. The term “deep” also known as hierarchical learning represents the large number of multiple hidden layers, which includes nonlinear processing units for the purpose of conversion and automatic feature extraction [8].

1) *Convolution neural networks*: Convolution Neural Networks (CNNs) can extract automatic discriminative features which have some invariance properties (e.g. translation invariance) [9]. It consists of three main layers which are convolution layers, pooling layers and fully connected layers [8].

The early convolution layers of the architecture are used for extracting local low-level features from the raw input while the deeper convolution layers of CNN are used for combining features together to generate global high-level features. The pooling layers are used to down sample the dimensionality of the extracted feature. The fully connected layers form an ANN network where each neuron in the

previous layer is connected to all the neurons in the current layer. The total number of fully connected neurons in the final layer determines the number of classes [8].

The advantages of CNNs include that they are well suited for end-to-end learning that generates automatic features from the raw data without any a priori feature selection. Moreover, CNNs scale well to large datasets. The disadvantages of CNNs include the large amount of training data, the long training time compared to simpler models, and the large number of hyper parameters to be learned. Two of the most famous CNN are VGG-16 [10] and ResNet-50 [11], which will be described briefly below.

a) *Visual graphic group net (VGG Net) model:* This net was developed by the technicians at the Visual Graphics Group from the Oxford and is in pyramid shape. The model consists of the bottom layers which are wide and the top layers are deep. There are two versions of VGG which are VGG-16 and VGG-19. VGG-16 is a combination 13 convolutional layers and Three fully connected layers as shown in . The VGG-19 is a much deeper network with 16 convolutional layers and three fully connected layers.

b) depicts VGG16 layered architecture.

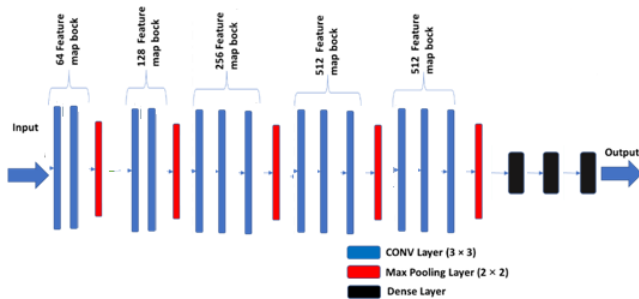


Fig. 1. VGG16 layered architecture.

c) *ResNet model:* It is a type of deep network based on residual learning. There are different versions of ResNet which are ResNet-18, ResNet-34, ResNet-50, ResNet-101 and ResNet-152. All of them have the same building units or residual blocks and formed by stacking the building residual blocks over each other. Any ResNet starts with block has a structure as shown in The residual blocks; are divided into two types which are Identity shortcut and Projection shortcut. The first block shown in Fig. 2 is the identity shortcut bottleneck block which is composed of a sequence of convolution layers of kernel size  $(1 \times 1)$  and stride = 1 connected to a convolution layer with kernel  $(3 \times 3)$  and stride = 1 followed by a convolution layer followed by kernel  $(1 \times 1)$  and stride = 2. This block is used when the input and output of feature map are the same. The other block shown in Fig. 2 is the projection shortcut bottleneck block which has the same sequence of layers with a newly added convolution layer in the projection shortcut which has a kernel size of  $(1 \times 1)$  with stride = 2. It is applied when shortcuts go across the feature map of two sizes. In the two blocks all the convolution layers are followed by

batch normalization and RELU activation function. The difference between different ResNet versions are the number of stacked residual blocks. For example, ResNet-50 which has 16 residual blocks and ends with fully connected layer as shown in Fig. 2.

### III. RELATED WORK

Due to the recent advancements in computer vision and the growing need for forgery detection, resources for targeted algorithms are vastly diverse in their approaches and practices.

Originally, the leading method in identifying tampered and non-tampered images was Support Vector Machine (SVM), as seen in [12], [13] and [14]. Shen et al. [12] were able to achieve quite high accuracies using the datasets CASIAv1.0 and CASIAv2.0 reaching 98% and 97% respectively. TF-GLCM method was proposed, which combines textural features extraction with grey level co-occurrence matrices. This method was directed at spliced images in particular. They used calculated textural features as components in feature vectors in order to recognize genuine and spliced images employing SVM as the classifier.

Similarly, Han et al. [13] used SVM to classify spliced images but after extracting features using the Markov method. They presented three types of Markov feature vectors and achieved accuracies up to 97.86% for CASIA v1 and 97.33% for CASIA v2 even with a small range of features.

Recent approaches are now leaning towards more complex neural network architectures, especially the ones to be able to detect more than one type of tampering rather than only splicing which was previously the case. One prominent study by Rao et al. [17] introduces a new CNN designated for the detection of copy-move and splicing forgeries. It utilizes high-pass filters to calculate residual maps in a special rich model (SRM) to capture any subtle pattern that is produced when image manipulation happens. The used CNN extract features from the test images, and a feature fusion method is then applied to acquire the final key features that are fed to SVM for classification. This method was able to achieve 98.04% for CASIA v1 and 97.83% for CASIA v2.

An interesting approach was presented by Sudiatmika et al. [5], who utilized the idea of error-level analysis (ELA) in conjunction with CNNs to create a more universal tool for detecting various types of forgery. Sudiatmika proposed normalizing the images before pursuing ELA calculation and feeding the resulting images to a VGG16 network. Sudiatmika et al. reached 92.2% accuracy on CASIA v2.0.

Kuznetsov [18] took a slightly different tactic in detecting forgeries using VGG network. The adopted method did not use the entire image for classification but rather small patches that are identified by either being forged on original expanding the training pool. He used a sliding window method to analyse each fragment of the image regarding its authenticity. This approach achieved very good results reaching 97.8% accuracy, 97.1% precision and 96.8% recall.

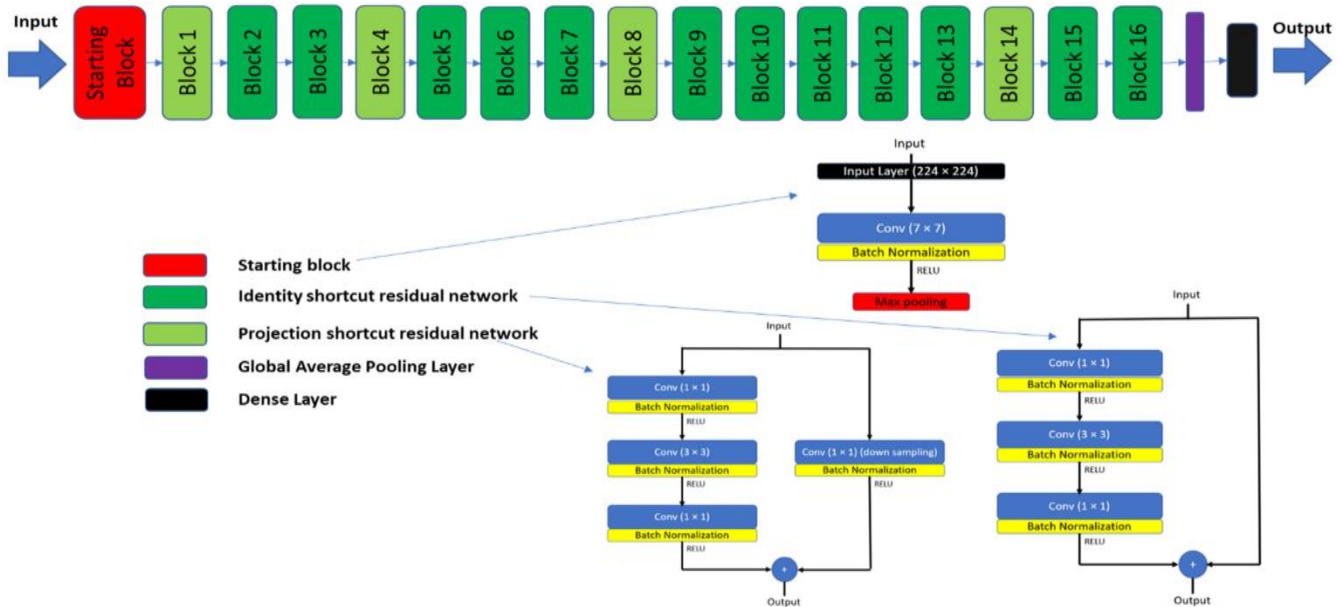


Fig. 2. ResNet layered architecture.

A modified ResNet architecture was used by Nath and Naskar [19] to automatically extract features, followed by a dense Artificial Neural network (ANN) for classification. The yielded results that exceeded 96% on sampled CASIA v2.0 to balance the classes.

Ding et al. [20] proposed a dual channel U-Net (DCU-Net), which accepts two inputs- the original tampered image and the residual tampered image. The residual image is generated by high pass filters to obtain the edges. The experimental results were shown on Casia2.0 and Columbia datasets, where the accuracy reached 97.93% and 97.27 % respectively.

The related work presents studies that either use traditional learning or deep learning approaches. With the increasing volumes of media and the advancements of editing technologies, traditional models will not provide adequate solution to the problem [21]. On the other hand, the used deep learning architectures are computationally intensive reducing their applicability on mobile real time applications [22]. In addition, further performance enhancement is needed to handle the problem.

#### IV. DMOBILE-ELA PROPOSED MODEL

A cascaded model is proposed to analyze whether images are tampered or authentic. The flow of the process model is shown in Fig. 3. The images are preprocessed applying ELA, then passed to a Dilated Mobile Net for classification.

##### A. Error Level Analysis

Error Level Analysis (ELA) is a concept that measures and visualizes the difference between an image and a re-compressed version of the same image which emphasizes certain parts that have been altered during previous edits. ELA measures the amount of error based on 8x8, relying on two main conditions applicable to JPEG images:

- A JPEG is said to be original if all 8x8 blocks have a similar error pattern. Therefore, the 8x8 pixel block can be said to have attained local minima.
- A JPEG is said to be manipulated if any 8x8 block has a higher error pattern and an 8x8 pixel block is not at its local minima.

In general, the computation of Error Level Images (ELIs ) follows the formulation in Eq (1)....

$$\begin{aligned} I_0 - I_{rc1} &= ELI_1 \\ I_0 - I_{rc2} &= ELI_2 \\ I_0 - I_{rc3} &= ELI_3 \end{aligned} \quad (1)$$

where  $I_0$  denote the original image and  $I_{rc1}$  represent a recompressed image at a given rate of compression. ELI is generated through pixel-wise difference of the two images.

The resultant ELI conveys the different quality levels within an image through varying intensities. For example if the image is forged, the added regions will be compressed at a different rate than the remaining original image. Such variation will be reflected through a distinct error pattern as the forged regions will be quantized through a non-linear ratio. Thus, can be used to localize the tampered areas. In this study, three different levels of image compression were examined, namely 10%, 50%, and 90% compression.

Fig. 3 depicts the localized error pattern of the spliced person, when applying three level of compression. The images clarify the potential of ELA in locating tampered regions. As can be seen from Fig. 3, higher compression rates better localize the tampered region. A difference Error Level Image (ELI) is produced between compression rates 50% and 90% to eliminate details and detect changes. The resultant image is shown in Fig. 4. The difference image is input to the Dilated-Mobile Net.





Fig. 3. ELIs at different compression rates (a) 10%, (b) 50% and (c) 90% emphasizing the spliced person (marked in red).



Fig. 4. Difference image generation from the available ELIs.

### B. Multiscale Dilated-MobileNet

A dilated or atrous MobileNet deep learning architecture is developed for classifying images into authentic and tampered. A detailed description of the architecture will be given below.

1) *Multiscale dilation* : Multiple dilated filters are applied to the input difference ELI. The aim of applying filters with various dilation rates is to increase the receptive field of the filters. Expansion of the receptive field help in considering all the relevant regions in an image and capturing all important information [23].

The dilation process inserts zeros depending on the dilation rate, hence increasing the receptive field of the filter while maintaining the number of parameters to be learnt. For example, with a dilation rate of 2 the receptive field of 3x3 filter is expanded to 5x5 convolution. Similarly, a dilation rate of three enlarges the filter to 7x7. The output of the multiscale dilation convolution is concatenated and input the first layer of

of the MobileNet.

2) *Light weight convolution* : MobileNet is a light weight architecture known for its applicability on mobile devices [24]. It is characterized by fewer parameters, small convolution filters 3x3 and hence lower computation demand compared to other CNN architectures. MobileNet architecture employ Depthwise Separable Convolution (DSCConv) Layer (shown in Fig. 5) instead of the standard convolution layer.

During depthwise separable convolution, each channel is convolved with each filter separately. The process is divided into depthwise convolution (3x3 depthwise convolution, batch normalization BN and RELU), followed by pointwise convolution (1x1 convolution, batch normalization BN and RELU). Splitting the convolution task into two steps speed up the computation task by a factor that reaches  $\frac{1}{f^2} + \frac{1}{n}$ , where  $f$  is the filter (kernel) size assuming squared dimensions and  $n$  is the number of filters (corresponding output channels). In addition, DSCConv helps maintain a shallower network than traditional CNNs with competitive accuracies.

Despite the advantages of DSCConv, the small sized convolution filters may reduce the goodness of the captured filters. Hence, the atrous filters with varying dilation rates offers a promising solution to this issue.

3) *Fully connected classification*: A dense layer of fully connected neurons is utilized to produce the final classification of whether the investigated image is authentic or tampered.

Overview of the network architecture is presented in Fig. 6, depicting the multiscale dilation and DSCConv layers.

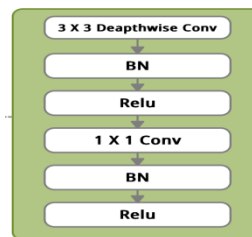


Fig. 5. Depthwise separable convolution layer structure.

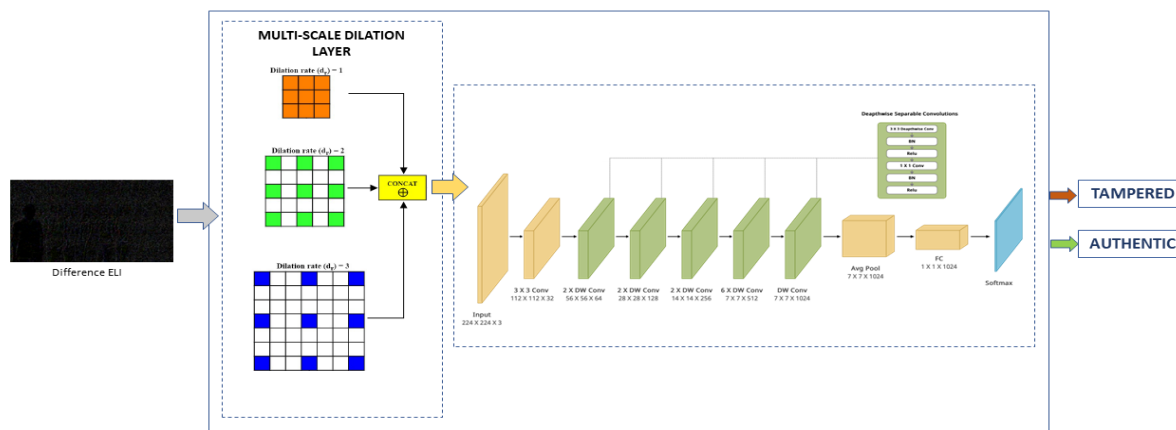


Fig. 6. Multiscale dilated MobileNet forgery classification on different ELI.

## V. RESULTS AND DISCUSSION

A briefing of the CASIAV2.0 dataset used in our experiments is given, followed by the performance measures applied to validate the performance of the proposed approach. The devised experimental setup is described for reproducibility of results. Then, the achieved results are presented and compared to recent forgery detection systems.

### A. Dataset Description

CASIAv2.0.[25] dataset is used in the following experiments to validate the performance of DMobile-ELA. CASIAv2.0 is a benchmark dataset created by Dong et al. [25] at the Institute of Automation, the Chinese Academy of Sciences, with the purpose of aiding the research and development of image tampering detection methods. It contains 5123 tampered images and 7491 authentic images. Tampered images contain both copy-move and splicing altered images, at 3295 and 1828, respectively. This dataset is a successor of CASIAv1.0 which only included spliced images. Fig. 7 displays samples of authentic and tampered images from the dataset.

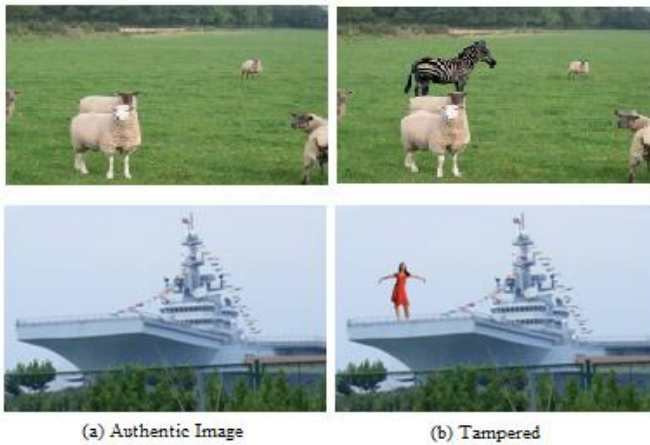


Fig. 7. Sample authentic and tampered images from CASIAV2.0 dataset.

### B. Performance Measures

Four performance measures are used to evaluate DMobile-ELA and allow its comparison with recent studies. The used measures are Accuracy (Acc), Precision (P), Recall (R) and  $F1_{score}$ . The computation of these measures relies on the confusion matrix given in Fig. 8.

		Predicted	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN

Fig. 8. Confusion matrix outlining true positive (TP), false negative (FN), false positive (FP) and true negative (TN).

The measures are calculated according to the following Eq. (2) to (5).

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$P = \frac{TP}{TP+FP} \quad (3)$$

$$R = \frac{TP}{TP+FN} \quad (4)$$

$$F1 = 2 \frac{P \times R}{P+R} \quad (5)$$

### C. Experimental Setup

The performance of DMobile-ELA is analyzed and compared to variable counterparts systematically. First, the performance of DMobileNet structure is contrasted to VGG16, ResNet-50 and MobileNet standard architectures. Also, the impact of transfer (pretrained on ImageNet [26]) learning or retraining from scratch is investigated. In addition, the effect of ELA on performance is elucidated through a comparison between models' performance with and without ELA. Finally, DMobile-ELA performance is compared against recent related studies.

The resolution of input images was adapted to the largest quadratic value that the MobileNet network supported which was 224x224. The default settings we used for assessing each model are splitting into 80% training and 20% validation sets, running the training for 10 epochs, and using 0.0001 learning rate. The model utilized Adam-optimizer while maintaining a batch size of 16.

### D. DMobile-ELA Performance Results

Forgery detection accuracy is measured for VGG16, ResNet-50, MobileNet and DMobileNet on the original image without ELA preprocessing. The results are shown in Fig. 9. Also, the performance of tuning pretrained models versus retraining of the models is tested. The results show that DMobile Net attains the highest accuracy, while VGG16 scores the lowest accuracy. Another observation is that retraining is better suited to the problem under study, as there is an evident performance gap that reaches around 15% in case of ResNet-50.

The accuracy of the models with ELA preprocessing is depicted in Fig. 10. From the shown accuracies, it can be seen that ELA aided the models to score higher accuracies than without ELA with differences ranging from 5% to 7% in case of retrained models. For pre-trained ResNet-50, the improvement arrived at 13%. Retrained models still presents higher accuracies compared to pre-trained models. Overall, retrained DMobile-ELA records the highest accuracy of 98.48%. The difference in accuracy between DMobileNet and MobileNet is around 3%, which is a considerable difference given that the number of parameter to be learned is the same.



Fig. 9. Accuracy of CNN models without ELA preprocessing.

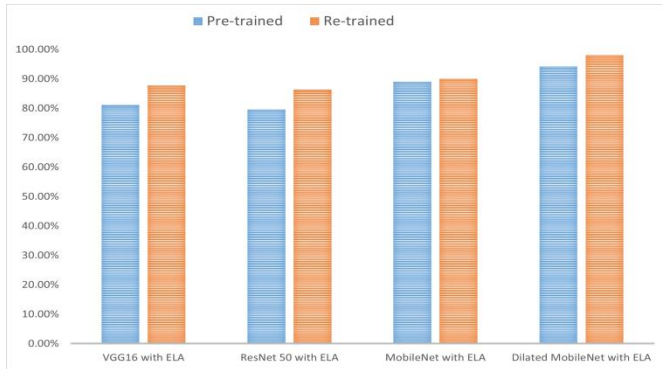


Fig. 10. Accuracy of CNN models with ELA preprocessing.

The training and validation accuracies learning curves for 10 epochs are shown in Fig. 11. The validation curve follows smoothly the training curve in the last three epochs diminishing the possibility of overfitting.

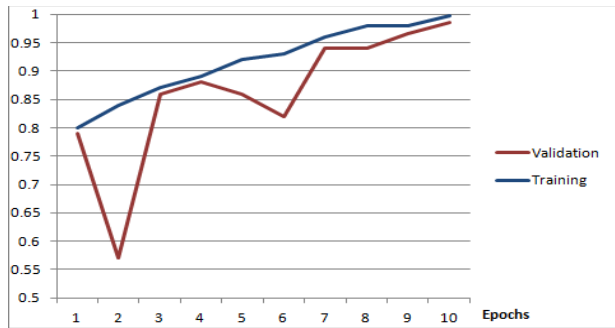


Fig. 11. Validation and training accuracy learning curves of DMobile – ELA.

TABLE I. PERFORMANCE COMPARISON BETWEEN DMOBILE-ELA AND RECENT RELATED STUDIES

Approach	Performance Measures			
	Acc	P	R	F1 <sub>score</sub>
Alahmadi et al. [14]	0.9645	0.9669	0.9415	0.9540
Ding et al. [20]	0.9793	0.8772	0.8893	0.8667
Kanwal et al. [16]	0.9759			
Niyishaka et al. [15]	0.9459	0.9000	0.9900	0.9300
A Kuznetsov [18]	0.9780	0.9710	0.9680	
Proposed DMobile-ELA	0.9848	0.9781	0.9862	0.9821

TABLE II. NUMBER OF PARAMETERS COMPARISON BETWEEN DMOBILE-ELA AND RECENT RELATED STUDIES

Approach	Parameters
Ding et al. [20]	17.2M
A Kuznetsov [18]	138M
Proposed DMobile-ELA	3.4M

Table I details the measures for assessing DMobile-ELA against some of the recent studies. The metrics show that the proposed DMobile-ELA surpasses its counterparts. It presents superior P, R and F1<sub>score</sub> than Ding et al. [20] with a gap of around 0.1 in all these measures. Also, it scores higher accuracy, P and F1<sub>score</sub> compared to Niyishaka et al. [15] with differences of around 0.04, 0.07 and 0.05 respectively. Kanwal et al. [16] and Alahmadi et al. [14] offer solutions with comparable accuracy. Similarly, Kuznetsov [18] presents competitive performance in terms of all metrics. However, Kuznetsov used VGG16 for detection, which is a computationally demanding architecture. Table II outlines the number of parameters to be learned for each backbone architecture. The given numbers highlight the favorable low computation demand of DMobileNet as a light weight architecture.

## VI. CONCLUSION

In this study, a forgery detection approach named DMobile-ELA is proposed. It integrates dilated MobileNet and Error Level Analysis (ELA), which leads to a lightweight high performing solution. The conducted experiments confirmed the success of DMobile-ELA in forgery detection, emphasizing the advantageous effect of ELA on performance. In addition, the experiments indicated the higher suitability of model retraining to the problem of forgery detection. Retrained DMobile-ELA performance reached Acc, P, R and F1<sub>score</sub> of 0.9848, 0.9781, 0.9862 and 0.9821 respectively on CASIAv2.0 dataset. Further improvements can be applied such as integrating different preprocessing procedures and merging textural features. Also, forgery types other than copy-move and splicing can be investigated to increase the applicability scope of the proposed approach.

## REFERENCES

- [1] F. Marcon, C. Pasquini, and G. Boato, "Detection of Manipulated Face Videos over Social Networks: A Large-Scale Study," J. Imaging, vol. 7, no. 10, 2021.
- [2] M. M. Eltoukhy, M. Elhoseny, K. M. Hosny, and A. K. Singh, "Computer aided detection of mammographic mass using exact Gaussian-Hermite moments," J. Ambient Intell. Humaniz. Comput., 2018.
- [3] A. Ross, S. Banerjee, and A. Chowdhury, "Security in smart cities: A brief review of digital forensic schemes for biometric data," Pattern Recognit. Lett., vol. 138, pp. 346–354, 2020.
- [4] W. Luo, Z. Qu, F. Pan, and J. Huang, "A survey of passive technology for digital image forensics," Front. Comput. Sci. China, vol. 1, no. 2, pp. 166–179, 2007.
- [5] K. D. Kadam, S. Ahirrao, and K. Kotecha, "Efficient Approach towards Detection and Identification of Copy Move and Image Splicing Forgeries Using Mask R-CNN with MobileNet V1," Comput. Intell. Neurosci., vol. 2022, p. 6845326, 2022.
- [6] M. Westerlund, "The Emergence of Deepfake Technology: A Review," Technol. Innov. Manag. Rev., vol. 9, pp. 39–52, 2019.

- [7] P. Sharma, M. Kumar, and H. Sharma, "Comprehensive analyses of image forgery detection methods from traditional to deep learning approaches: an evaluation," *Multimed. Tools Appl.*, 2022.
- [8] S. Dargan, M. Kumar, M. R. Ayyagari, and G. Kumar, "A Survey of Deep Learning and Its Applications: A New Paradigm to Machine Learning," *Arch. Comput. Methods Eng.*, vol. 27, no. 4, pp. 1071–1092, 2020.
- [9] J. Heaton, *Artificial Intelligence for Humans: Deep learning and neural networks*. Heaton Research, Incorporated, 2015.
- [10] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition." *arXiv*, 2014.
- [11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," 2016, pp. 770–778.
- [12] X. Shen, Z. Shi, and H. Chen, "Splicing image forgery detection using textural features based on the grey level co-occurrence matrices," *IET Image Process.*, vol. 11, pp. 44–53, 2017.
- [13] J. G. Han, T. H. Park, Y. H. Moon, and I. K. Eom, "Efficient Markov feature extraction method for image splicing detection using maximization and threshold expansion," *J. Electron. Eng.*, vol. 25, pp. 21–30, 2016.
- [14] A. Alahmadi, M. Hussain, H. Aboalsamh, G. Muhammad, G. Bebis, and H. Mathkour, "Passive detection of image forgery using DCT and local binary pattern," *Signal, Image Video Process.*, vol. 11, no. 1, pp. 81–88, 2017.
- [15] P. Niyishaka and C. Bhagvati, "Image splicing detection technique based on Illumination-Reflectance model and LBP," *Multimed. Tools Appl.*, vol. 80, no. 2, pp. 2161–2175, 2021.
- [16] N. Kanwal, A. Girdhar, L. Kaur, and J. S. Bhullar, "Digital image splicing detection technique using optimal threshold based local ternary pattern," *Multimed. Tools Appl.*, vol. 79, no. 19, pp. 12829–12846, 2020.
- [17] Y. Rao and J. Ni, "A deep learning approach to detection of splicing and copy-move forgeries in images," in *2016 IEEE International Workshop on Information Forensics and Security (WIFS)*, 2016.
- [18] A. Kuznetsov, "Digital image forgery detection using deep learning approach," *J. Phys. Conf. Ser.*, vol. 1368, pp. 28–32, 2019.
- [19] S. Nath and R. Naskar, "Automated image splicing detection using deep CNN-learned features and ANN-based classifier," *Signal, Image Video Process.*, vol. 15, no. 7, pp. 1601–1608, 2021.
- [20] H. Ding, L. Chen, Q. Tao, Z. Fu, L. Dong, and X. Cui, "DCU-Net: a dual-channel U-shaped network for image splicing forgery detection," *Neural Comput. Appl.*, vol. 35, no. 7, pp. 5015–5031, 2023.
- [21] A. L'Heureux, K. Grolinger, H. F. Elyamany, and M. A. M. Capretz, "Machine Learning With Big Data: Challenges and Approaches," *IEEE Access*, vol. 5, pp. 7776–7797, 2017.
- [22] V. Isuyama and B. Albertini, "Comparison of Convolutional Neural Network Models for Mobile Devices," in *Anais do XX Workshop em Desempenho de Sistemas Computacionais e de Comunicaao*, 2021, pp. 73–83.
- [23] W. Luo, Y. Li, R. Urtasun, and R. Zemel, "Understanding the Effective Receptive Field in Deep Convolutional Neural Networks," in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 2016, pp. 4905–4913.
- [24] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *arXiv Prepr.*, vol. arXiv:1704, 2017.
- [25] J. Dong, W. Wang, and T. Tan, "CASIA Image Tampering Detection Evaluation Database," in *2013 IEEE China Summit and International Conference on Signal and Information Processing*, 2013, pp. 422–426.
- [26] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.