Detection and Localization of Image Tampering using Deep Residual UNET with Stacked Dilated Convolution

Ali Ahmad Aminu, Nwojo Nnanna Agwu and Adeshina Steve
aaminu3@gmail.com, nagwu@nileuniversity.edu.ng, steve.adeshina@nileuniversity.edu.ng
Nile University of Nigeria, Abuja, Nigeria

Summary
Image tampering detection and localization have become an active area of research in the field of digital image forensics in recent times. This is due to the widespread of malicious image tampering. This study presents a new method for image tampering detection and localization that combines the advantages of dilated convolution, residual network, and UNET Architecture. Using the UNET architecture as a backbone, we built the proposed network from two kinds of residual units, one for the encoder path and the other for the decoder path. The residual units help to speed up the training process and facilitate information propagation between the lower layers and the higher layers which are often difficult to train. To capture global image tampering artifacts and reduce the computational burden of the proposed method, we enlarge the receptive field size of the convolutional kernels by adopting dilated convolutions in the residual units used in building the proposed network. In contrast to existing deep learning methods, having a large number of layers, many network parameters, and often difficult to train, the proposed method can achieve excellent performance with a fewer number of parameters and less computational cost. To test the performance of the proposed method, we evaluate its performance in the context of four benchmark image forensics datasets. Experimental results show that the proposed method outperforms existing methods and could be potentially used to enhance image tampering detection and localization.

Key words:
Image Tampering, detection, localization, residual UNET, Dilated convolution

I. Introduction

Nowadays, digital images have become the core medium of communication due to their expressive potentials and ease of distribution. Consequently, they represent a common source of evidence in resolving everyday life controversies such as in criminal investigations and legal proceedings [1]. However, the advent of sophisticated image editing tools and the ease with which digital images can be altered to convey false or misleading information has questioned the reliability of visual information [2]. This difference between the importance of digital images on one hand and the doubts regarding their susceptibility to manipulations, on the other hand, calls for a means of validating their authenticity before using them in important settings.

Among the known image tampering methods, copy-move [3], splicing [4], and removal [1] are the most popular types of tampering [3]. Copy move is achieved by copying and pasting a portion of an image to another portion on the same image. This is usually done to create an object that never existed or to conceal an existing object in the image. Image splicing copies and paste portions from one image to another image to create a composite image. Object removal deletes an object from an original image followed by in-painting to create a tampered image.

To create more convincing tampering, often some post-processing operations such as median filtering, and JPEG compression are applied to copy moved regions or spliced regions to smooth the boundaries of forgery regions making them robust against tampering detection techniques. Examples of copy-move, splicing, and object removal alongside their original images and their corresponding ground-truth masks are illustrated in Fig. 1. It can be noticed that the tampered images cannot be easily distinguished from the authentic images visually even with careful human examination. Consequently, detecting image tampering has become increasingly difficult, which leads to the widespread of malicious image tampering.

To address this problem, a handful of techniques have been proposed recently for detecting and localizing image tampering, aiming at improving the state-of-the-art forgery detection methods. The earliest methods such as [10], [11], and [12] perform forgery detection by exploiting frequency domain features, Color Filter Array features, and local binary descriptors. Inspired by the performance of the Spatial Rich Model (SRM) in image steganalysis, many image tampering detection approaches [3] [15] [17] utilizing SRM features have been proposed, which produced excellent results. With the success of deep learning methods, specifically, CNN in many visual recognition tasks, recent studies [21] [22] [23] [24] in image forensics also seek to leverage the strength of deep learning

Manuscript received September 5, 2021
Manuscript revised September 20, 2021
https://doi.org/10.22937/IJCSNS.2021.21.9.28

Figure. 1. Examples of different types of image tampering. Columns 1, 2, and 3 show the original images, tampered images, and their ground-truth masks respectively. While row 1, 2, and 3 show the examples of removal, copy-move, and splicing manipulations respectively.

methods to solve the problem of detecting and localizing digital image tampering. These methods can automatically learn manipulations traces directly from data without the need for handcrafted features or human analysis by using a set of convolution kernels whose weights are learned via a neural network training technique known as back-propagation. Hence, they provide better performance than the earliest methods.

In recent times, a variety of studies utilizing deep neural networks have suggested that their performance could be improved by increasing the depth of the network [8-9]. But, as the network gets deeper, they become difficult to train due to the performance degradation problem. He et al [5] proposed the concept of residual learning, which uses identity mapping to speed up training, thus, overcoming this problem. To overcome the limitation of having large training data, Ronnebegger *et al* [6], proposed the UNET architecture. The UNET architecture heavily relies on data augmentation and allows signals to propagate easily from low levels to high levels of the network to improve image segmentation performance. To increase the receptive field size of convolution kernels, Yu *et al* [7] suggested the use of dilated convolutions. The Dilated convolution introduces a new parameter to the standard convolution, the dilation rate that widens the receptive field size of the convolution kernels enabling them to capture more image clues without loss of resolution and at a reduced computational cost.

Inspired by the residual network, UNET architecture, and dilated convolution, we proposed the Deep Residual UNET with Stacked Dilated Convolutions to solve the problem of image tampering detection and localization. The proposed network combines the strengths of the residual network, UNET architecture, and dilated convolution. To assess the performance of the proposed network, we evaluated its performance on four benchmark image manipulation datasets. Experimental results show that the proposed method outperforms existing methods and could be potentially used to enhance image tampering detection and localization. The new network differs from the conventional UNET in the following areas: 1) the use of two types of residual unit, one for building the encoding path of the network and the other for building the decoding path; 2) the use of dilated convolutions, instead of using only the standard convolution, we adopted dilated convolutions in our residual units to widen their receptive field size; 3) the bottleneck path connecting the encoding and decoding path of the network is built entirely from a stacked of dilated convolutions.

II. Related work

A handful of techniques have been proposed for detecting and localizing image tampering such as copy-move, image-splicing, object-removal, and content preserving manipulations such as median filtering and JPEG compression. In this section, we will briefly discuss some of the existing methods used for detecting and localizing digital image tampering. The traditional methods try to come up with better ways of representing image manipulations using handcrafted features. For instance, in [10] the authors suggest a forgery detection model that exploits subtle inconsistencies in the color illumination of the image. They used texture and edge-based features from the images and utilized a machine learning method for classification. To adapt to JPEG compression, which can reduce the characteristics of local correlation patterns, Li et al [11], proposed a method of image tampering detection using Color Filter Array (CFA) interpolation. The frequency characteristics of the posterior probability map are computed, combined, and then compared to a threshold to classify the image as either tampered or not. Carvalho et al [12], utilized several local image descriptors and color space models to reveal the traces introduced by splicing in image illuminant maps and achieved excellent performance on DSO-1 image splicing datasets.

Inspired by the success of the Spatial Rich Model (SRM) in many image steganalysis tasks [13], several image manipulation detection based on SRM have been proposed. Cozzolino et al [14] examine and show the robustness of SRM features in detecting image tampering. They

combined SRM features with CNN to perform image tampering localization. Sundus et al [15] also investigate and demonstrate the performance of SRM and Local Binary Pattern (LBP) in detecting multiple image tampering. They embedded LBP in SRM sub-models to capture detailed statistics of the quantized version of image noise residuals. The resulting features were used for classification using an ensemble classifier. Rao et al. [16] and [17] initialized the weights of their networks with SRM filter kernels to enhance image forgery detection and localization.

With the success of deep learning methods, specifically, CNN in many visual recognition tasks, recent studies in digital image tampering detections have employed deep learning methods to address the problem of image tampering detection and localization. Salloum et *al*. [18] utilized a multi-task fully convolutional network (MFCN) to localize image splicing attacks. MFCN used two learning tasks to learn the label of the surface and the boundaries of the spliced regions. In [19] an image splicing detection and localization approach using illumination maps and CNN is proposed. CNN was used to extract discriminative features from illumination maps and SVM was used for the classification task. Their approach was able to attain an accuracy of 95% when evaluated using some publicly available splicing datasets. The work of [20] proposed a hybrid LSTM and encoder-decoder network for pixel-wise manipulation localization using resampling and spatial domain features. More recently, in [25], a two-branched architecture and a fusion procession model was suggested for CMFD. The two branches were used to localize and identify copy-move forgeries via CNN and GAN respectively. Rao et al [26] proposed a two-branch CNN for splicing detection and localization based on SRM and CNN. Zhang and Ni [27] proposed a dense Unit with a cross-layer intersection for detection and localization of image forgeries. They initialized the weights of their network with high pass filters used in SRM and used a multi-stage training approach to speed up convergence.

III. Materials and Methods

We propose the Deep Residual UNET with stacked dilated convolutions, for the task of image tampering detection and localization. The proposed model adopts the conventional UNET [6] architecture, originally designed for semantic segmentation, with modifications in the encoding, decoding, and bottleneck paths of the network. The UNET architecture overcame the need for a large amount of training data by providing a network and training technique that heavily depends on the use of data augmentation. Moreover, the UNET model was designed to solve the training problem of neural networks by allowing high-resolution features from its encoding path combine with the corresponding up-sampled output in the decoding path [29]. This design created pathways for information propagation between the encoding and decoding path, similar to that of the residual unit, thus, allowing signals to propagate easily between the encoding and decoding path and hence facilitating the network training process. Therefore, we adopted UNET as one of the basic architecture of the proposed model. To capture global image tampering artifacts and reduce the computational burden of the proposed method, we enlarge the receptive field size of the convolutional kernels by adopting dilated convolutions [7, 30] in the residual units used in building the encoder and decoder blocks of the proposed network. We introduced the residual learning module in the encoder and decoder blocks to speed up the training process and facilitate information propagation between the lower layers and the higher layers which are often difficult to train.

A. Proposed residual units for the encoder and decoder block

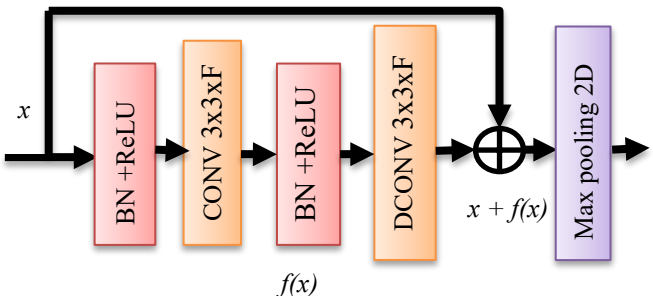


Figure 2. Residual Unit for Encoder Block of the Proposed Network. Dconv corresponds to a dilated convolution and F represents the number of filters used in the block.

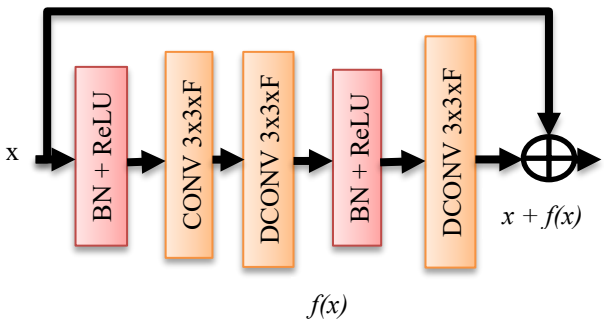


Figure 3. Residual Unit for the Decoder block (D) of the proposed network.

The core unit in the proposed network is proposed based on the residual module in ResNet [5] and the dilated convolutions in [7]. A limitation of the Conventional UNET is that the lowest level of the network has a relatively small receptive field [31]. This prevents the network from

extracting the necessary image clues or traces needed for image recognition applications. Thus, to overcome this, we utilized dilated convolution in building the residual units that constitute the building blocks of the encoding and decoding path of the proposed model and a stacked of dilated convolutions in building the bridge/bottleneck that connects the encoding and decoding path of the proposed network. As shown in **Figures** 2 and 3, compared to the conventional UNET architecture, our design replaces the two 3x3 standard convolutions in each block of the encoding path with a residual unit built from one standard convolution and one dilated convolution, each preceded by a Batch Normalization (BN) and ReLU operations. The 2x2 standard convolution and the two 3x3 standard convolutions in each block of the decoding path are replaced by one standard convolution and two dilated convolutions, each preceded by a BN and ReLU operations. The residual learning module and dilated convolutions solve the problem of vanishing gradient and encourage feature propagations, expands the receptive field size of convolution kernels, and save computational resources, respectively.

Let x_i be the input of the ^{i}th encoder block, the output of the encoder block can be formulated as

$$y_i = \quad P(x_i + f(x_i)) \tag{1}$$

where p*(.)* denotes the pooling operations and *f(.)* represents the residual learning function.

B. Deep Residual UNET with stacked Dilated Convolutions

Here we present the Deep Residual UNET with Stacked Dilated Convolution, a network for image tampering detection and localization that uses the conventional UNET as a backbone. The proposed network combines the advantages of the residual network, UNET, and dilated convolution. These combinations lead to the following benefits: 1) the identity mapping between the encoding and decoding path of the network and within the residual unit will help propagates gradients to higher layers which are the most difficult to train due to the problem of vanishing gradient. Hence speeding up the network training and improving its performance; 2) the UNET architecture will ease the training of the network as it allows signals to transmit easily between the encoding and decoding path and does not require a huge amount of data; 3) the dilated convolution will expand the receptive field size of the network kernels enabling the network to capture more image tampering clues even at the lowest level of the network.

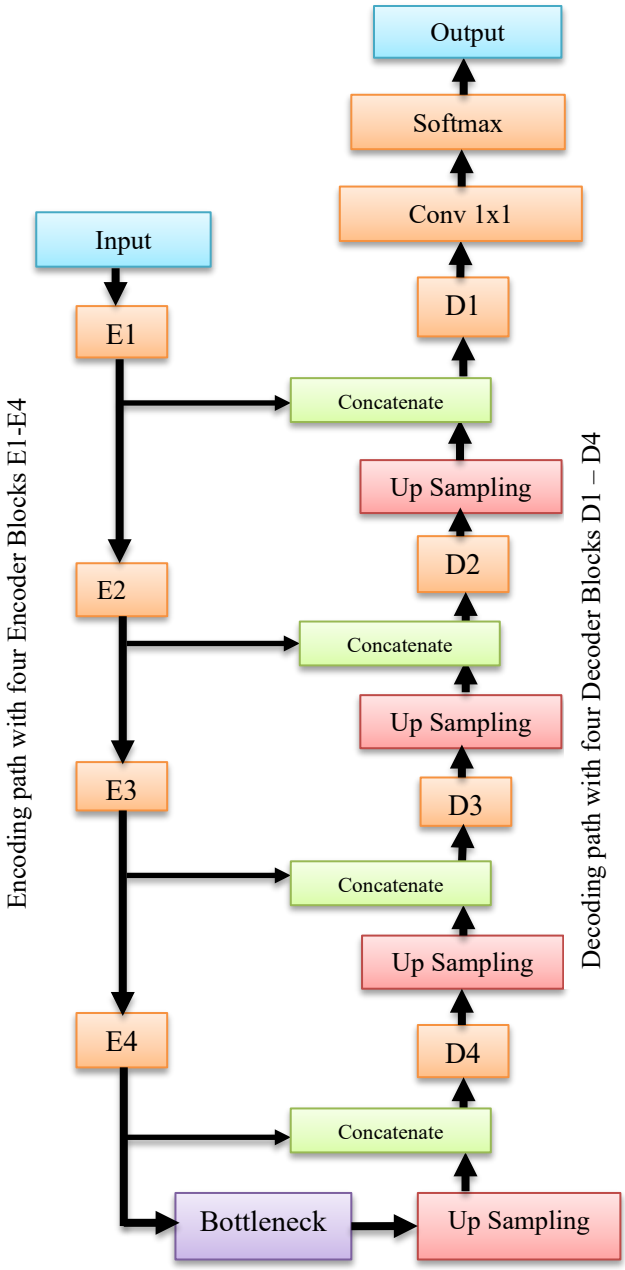


Figure 4. The architecture of the proposed network for detection and localization of image tampering. The details are omitted for simplicity. The encoding path consists of four encoder blocks (E1, E2, E3, and E4) and the decoding path has four decoder blocks (D1, D2, D3, and D4). The two paths are connected by the bridge/ bottleneck which consists of five dilated convolutions.

As shown in **Figure 4**, the proposed model comprises of three parts: 1) the encoding path, responsible for encoding the input image into compact form; 2) the decoding or expanding part responsible for restoring the size of feature maps; 3) the bridge or bottleneck path which connects the

encoding and decoding paths. The encoding path consists of four encoder blocks (**Figure 2**), each block of the encoding path is a residual unit built from one standard convolution and one dilated convolution, each preceded by Batch Normalization (BN) and ReLU layers. The four encoder blocks in the encoding path E1, E2, E3, and E4 used 32, 64, 64, 64 filters, respectively to filter and encode the input image into compact form. The decoding path consists of four decoder blocks, each preceded by an up-sampling layer and a concatenation operation for recovering the encoded information and combining it to the corresponding encoding path respectively. Each block of the decoding path is a residual unit constructed from one standard and two dilated convolutions, each preceded by BN and ReLu layers. The four decoder blocks D1, D2, D3, and D4 used 32, 64, 64, 64 filters to filter and recover the encoded input image. The bottleneck connecting the encoding and decoding path comprises five dilated convolutions with 1, 2, 4, 8, 16 dilation rates respectively.

The encoding path takes input Image X, extracts image tampering features, and down-samples the feature maps which serve as the input to the bottleneck block. The output of every dilated convolution in the bottleneck is added up and the resulting feature map serves as input to the decoding path, which is responsible for up-sampling the feature maps to original image size. After the last layer of the decoding path, our network uses a 1x1 convolution layer and a softmax layer to map each feature map for the detection and localization of image tampering. In total, the proposed model uses 26 convolution layers.

C. Implementation detail

The proposed model is implemented using Tensor-Flow on a machine with NVidia GeForce GTX 1050 GPU and 8GB RAM. All input images were first resized to 256 and augmented to enlarge the datasets and reduce the imbalance between the positive classes and negative classes before using them for our experiments.

The input images and their corresponding ground truth masks are used to train the proposed network. He-normal initialization was used to initialize the weights of the convolution layers. We trained the proposed model end to end in each experiment for 100 epochs with a batch of 16 images for each training iteration to minimize the cross-entropy loss function (1) using the adam optimizer with a learning rate of 0.001 and the default values of the moments ($\beta1 = 0.9$, $\beta2 = 0.999$, and $\varepsilon = 10^{-7}$).

Let W be the parameter vector associated with the image tampering localization task, the cross-entropy loss can be formulated as:

$$L(W) = \frac{-1}{M}\sum_{m}^{M}\sum_{n}^{N} 1(y^m = n)\log\,(y^m = n|y^m\,W) \quad (2)$$

Where M and N represent the total number of pixels and the number of class, y denote the input pixels and $1(.)$ is an indicator function which equals 1 if m = n, otherwise 0.

We minimizing the cross-entropy loss using the Adam optimizer with all the training samples to learn the network's optimal set of parameters. Using these learned parameters, the network can predict whether a given image is tampered with or not from the test samples.

D. Datasets

We validate the proposed method with current methods from the literature on Casia v2.0 [32-33], Columbia uncompressed [34], Nist Nimble 2016 [35], and MICC F2000 [36] image forensics datasets. For each dataset, we first split the whole images into three subsets; training, testing, and validation subsets before data augmentation to enlarge the number of samples.

TABLE I.
SHOWING THE TRAIN/TEST/VALIDATION SPLIT BEFORE AND AFTER AUGMENTATION

Datasets	**Training**	**Testing**	**Validation**
Casia v2.0			
Before Augmentation	1146	359	287
After Augmentation	5393	1543	876
Columbia			
Before Augmentation	215	69	54
After Augmentation	971	337	243
Nist Nimble 2016			
Before Augmentation	755	236	189
After Augmentation	3971	994	876
MICC F2000			
Before Augmentation	1210	379	303
After Augmentation	5261	1497	1317

E. Performance Evaluation Metrics

After training the proposed model, its performance was evaluated using three standard evaluation metrics. Pixel-wise f1-score and AUC score were used to evaluate the model's localization performance and the pixel-wise accuracy is used to evaluate the model's performance for detecting image splicing, copy-move, and object removal. These metrics were also utilized by the baseline models [18][21][27] and many states of the art techniques for detecting and localization of image forgeries [20, 26] hence, we also used them to easily compare the performance of the proposed method with the existing methods from the literature. For each dataset used in our experiments, we report the detection performance of the proposed model in terms of accuracy and the localization performance in terms of f1-score and AUC score as shown in Table II.

IV. Results and Discussions

Table II summarises the localization and detection results obtained by the proposed method on four image forensics datasets. From the table of results, we can observe that our method achieves high detection performance (99.39% in MICC-F2000 and 97.47% in casiav2.0) and high localization performance with an average F1-score and AUC score respectively more than 0.70 and 0.888 on the four benchmark image forensics datasets.

TABLE II.
PROPOSED METHOD EXPERIMENTAL RESULTS ON FOUR BENCHMARK DATASETS

Datasets	F1 score	AUC	Accuracy
CasiaV2.0	0.7033	0.8886	97.47%
Columbia	0.7540	0.9228	91.33%
MICC F2000	**0.8239**	**0.9358**	**99.39%**
Nist Nimble 2016	0.7526	0.9279	95.07%

To further validate the performance of the proposed model, we compare its results with that of the existing state-of-the-art methods which include [18] [21] [27] [37], and [38]. The results of these methods are replicated from the original papers as we could not access and run their codes. The results of the proposed method in comparison with previous methods from the literature are presented in Table III. The numbers highlighted in bold show the best result obtained for that given dataset. The "–" sign indicates the given method did not use those metrics in evaluating their model. As shown in Table III, it can be observed that the proposed method outperformed the conventional approaches like [37] and [38] that relied on handcrafted features by far in terms of f1-score and AUC score on all three datasets. This is because it can automatically learn and extract image manipulation fingerprints directly from the input images without the need for complex preprocessing steps and hand design features that may introduce noise, which may interfere with image tampering artifacts. Moreover, the methods in [37] and [38] are tailored toward detecting a specific type of tampering operation, hence they may not have the necessary image clues needed for localization, thus, limiting their performance.

TABLE III.
COMPARISON OF FI SCORE AND AUC SCORE ON THREE DATASETS WITH THE STATE OF THE ART METHODS

Datasets	Methods	F1 score	AUC
CasiaV2.0	ELA[38]	0.2140	0.6130
	CFA[37]	0.2070	0.5220
	MFCN[18]	0.5410	-
	Zhou[21]	0.4080	0.7950
	Zhang[27]	0.6830	-
	Proposed	**0.7033**	**0.8886**
Columbia	ELA[38]	0.4700	0.5810
	CFA[37]	0.4670	0.7200
	MFCN[18]	0.6120	-
	Zhou[21]	0.6970	0.8580
	Zhang[27]	**0.9307**	-
	Proposed	0.7540	**0.9228**
Nist Nimble 16	ELA[38]	0.2360	0.4290
	CFA[37]	0.1740	0.5010
	MFCN[18]	0.5710	-
	Zhou[21]	0.7220	**0.9370**
	Zhang[27]	0.5240	-
	Proposed	**0.7526**	0.9279

We also compared our results with recent deep learning methods from the literature. The proposed method outperforms the method of Zhou[21], MFCN[18], and Zhang[27] on casiav2.0 datasets and obtained the highest f1-score and AUC score respectively on the Nist Nimble and Columbia datasets as shown in Table III, which indicates the robustness of the proposed method. The proposed method achieved better performance than the other deep learning methods due to the use of a larger receptive field size which allows the network to capture more image manipulation clues than the other methods. In the proposed network, the residual blocks used in both the encoding and decoding path of the network were constructed from dilated convolutions. This has increased the receptive field size of the proposed network at the same time reducing computation burden and hence facilitating convergence. Moreover, the residual unit has aided in propagating image manipulation signals from lower layers to higher layers which are the most difficult to train, thus, speeding up the network training process and improving performance.

A. Qualitative Evaluation

We illustrate some qualitative results (proposed network sample output) obtained from our network on casiav2.0, Columbia, and Nist Nimble datasets respectively in **Figures**. 5, 6, and 7. **Figures**. 5 and 6 show the performance of the model in detecting either splicing or copy-move while **Figure**. 7 demonstrates the performance of the model in detecting copy-move, splicing, and removal. As shown in the visualization results, the predicted masks of the proposed network could perfectly locate the tampered region within the tampered images in all three datasets. This indicates the effectiveness of our method in localizing splicing, copy-move, and object removal image manipulations.

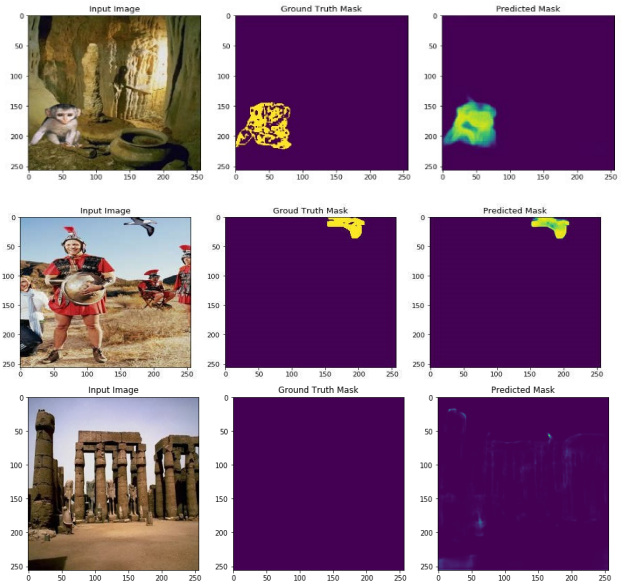
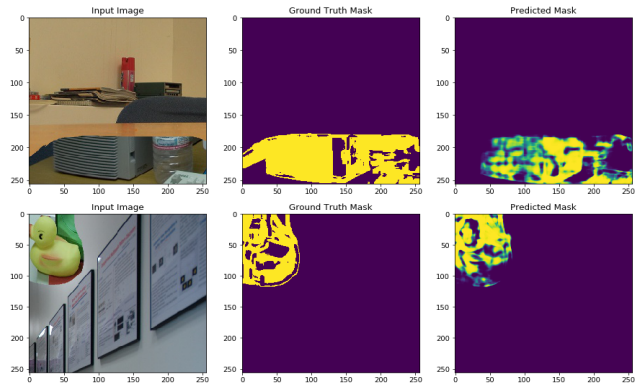


Figure 5. Visualization results on casiav2.0 dataset. The first and second columns are the input image and ground truth mask, while the third column is the network output Row 1 and 2 show the results when the input images have been tampered with, while row 3 shows the result for untampered image



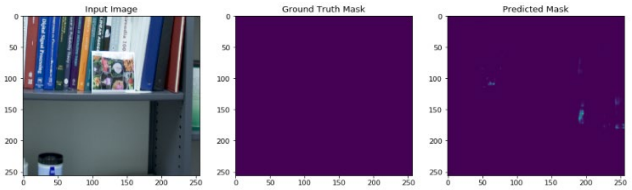


Figure 6. Visualization results for the Columbia dataset. The first and second columns are the input image and ground truth mask, while the third column is the network output. Row 1 and 2 show the results when the input images have been tampered with, while row 3 shows the result for untampered image

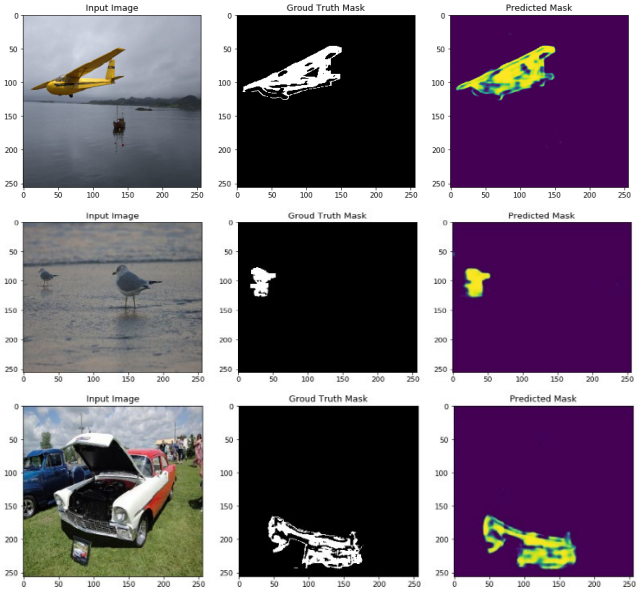


Figure 7. Visualization results for NIST nimble dataset. The first and second columns are the input image and ground truth mask, while the third column is the network output. Row 1 and 2 show the results for splicing and copy-move tampering, while row 3 shows the result for object removal tampering.

V. Conclusions

In this paper, we have proposed a novel method for image tampering detection and localization based on dilated convolution, residual network, and UNET architecture. The proposed method combines the benefits of dilated convolution, residual network, and UNET architecture. we enlarge the receptive field size of the convolutional kernel of the proposed network by adopting dilated convolutions in the residual units used in building the encoding and decoding paths of the network enabling it to capture global image tampering clues. These features allow us to build a simple, yet powerful network that can detect and localize different types of image tampering. In contrast to existing deep learning methods, having a large number of layers and many network parameters, which are often difficult to train,

the proposed method can achieve excellent performance with fewer layers and less computational cost.

Acknowledgement

This work is supported by the National Information Technology Development Agency (NITDA) of Nigeria.

References

[1] A. Piva, "An overview on image forensics," *ISRN Signal Processing*, (2013).

[2] P. Kakar, "Passive Approaches for Detecting Digital Image Forgery", Ph.D. Thesis, Nanyang Technological University, 2012.

[3] X. Qiu, H. Li, W. Luo, and J. Huang "A universal image forensics strategy based on steganalytic model," *In Proceedings of the 2nd ACM workshop on Information hiding and multimedia security,* (pp. 165-170), ACM (2014 Jun 11).

[4] H. Farid, "A survey of image forgery detection," *IEEE Signal Process Mag* 2:16–25, (2009).

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in CVPR, 2016, pp. 770–778.

[6] O. Ronneberger, F. Philipp, and B. Thomas, "U-net: Convolutional networks for biomedical image segmentation," In *International Conference on Medical image computing and computer-assisted intervention*, pp. 234-241. Springer, Cham, 2015.

[7] F. Yu, and K. Vladlen, "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv: 1511.07122* (2015).

[8] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in CVPR, 2015, pp. 1–9.

[9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv: 1409.1556, 2014.

[10] L. Baby, and A. Jose, "Detection of Splicing in Digital Images Based on Illuminant Features," *International Journal of Innovation and Scientific Research*, *11*(2351-8014), p.4, 2014

[11] L. Li L, J. Xue, X. Wang, and L. Tian, "A robust approach to detect digital forgeries by exploring correlation patterns," Pattern Anal Appl 1–15, 2013.

[12] T. Carvalho, F. A. Faria, H. Pedrini, R. Da S. Torres, and A. Rocha, ''Illuminant-based transformed spaces for image forensics,'' *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 4, pp. 720–733, Apr. 2016

[13] M. Goljan, F. Jessica, and C. Rémi, "Rich model for steganalysis of color images," In *2014 IEEE International Workshop on Information Forensics and Security (WIFS)*, pp. 185-190. IEEE, 2014.

[14] D. Cozzolino and L. Verdoliva. Single-image splicing localization through autoencoder-based anomaly detection. In *WIFS*, 2016.

[15] F. Sundus, M. H. Yousaf, and F. Hussain. "A generic passive image forgery detection scheme using local binary pattern with rich models." *Computers & Electrical Engineering* 62 (2017): 459-472.

[16] Y. Rao, and J. Ni, "A deep learning approach to detection of splicing and copy-move forgeries in images," In *IEEE International Workshop on Information Forensics and Security (WIFS)* (pp. 1-6), IEEE, 2016.

[17] W. Yue, W. AbdAlmageed, and P. Natarajan, "ManTra-Net: Manipulation tracing network for detection and localization of image forgeries with anomalous features," In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9543-9552. 2019.

[18] R. Salloum, Y. Ren, and C. Kuo, "Image splicing localization using a multi-task fully convolutional network," (mfcn). *arXiv* **2017**.

[19] T. Pomari, G. Ruppert, E. Rezende, A. Rocha, and T. Carvalho, "Image splicing detection through illumination inconsistencies and deep learning," In *2018 25th IEEE International Conference on Image Processing (ICIP)* (pp. 3788-3792). IEEE, 2018.

[20] J. H. Bappy, C. Simons, L. Nataraj, B. S. Manjunath, and A. K. RoyChowdhury, ''Hybrid LSTM and encoder-decoder architecture for detection of image forgeries,'' IEEE Trans. Image Process., vol. 28, no. 7, pp. 3286–3300, Jul. 2019.

[21] P. Zhou, X. Han, V. I. Morariu, and L. S. Davis, "Learning rich features for image manipulation detection," In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1053-1061. 2018.

[22] M. Boroumand, and J. Fridrich, "Deep learning for detecting processing history of images*," Society for Imaging Science and Technology,* (2018).

[23] Y. Chen, K. Xiangui, Q. S. Yun, and Z. Jane Wang, "A multi-purpose image forensic method using densely connected convolutional neural networks," *Journal of Real-Time Image Processing* 16, no. 3 pp. 725-740, (2019).

[24] Y. Zhan, Y. Chen, Q. Zhang, and X. Kang, "Image forensics based on transfer learning and convolutional neural network," In *Proceedings of the 5th ACM Workshop on Information Hiding and Multimedia Security* (pp. 165-170), 2017.

[25] Y. Abdalla, M.T. Iqbal, and M.Shehata, "Copy-Move Forgery Detection and Localization Using a Generative Adversarial Network and Convolutional Neural-Network," *Information*, *10*(9), p.286, (2019).

[26] Y. Rao, J. Ni, and H. Zhao, "Deep Learning Local Descriptor for Image Splicing Detection and Localization," *IEEE Access* 8 (2020): 25611-25625.

[27] R. Zhang, and J. Ni. "A Dense U-Net with Cross-Layer Intersection for Detection and Localization of Image Forgery." In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2982-2986. IEEE, 2020.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," In *European conference on computer vision*, pp. 630-645. Springer, Cham, 2016.

[29] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual u-net," *IEEE Geoscience and Remote Sensing Letters* 15, no. 5 (2018): 749-753.

[30] M. Holschneider, R. Kronland-Martinet, J. Morlet, and P.H. Tchamitchian, "A real-time algorithm for signal analysis with the help of the wavelet transform," In *Wavelets: Time-Frequency Methods and Phase Space. Proceedings of the International Conference*, 1987.

[31] S. Vesal, N. Ravikumar, and A. Maier, "A 2D dilated residual U-Net for multi-organ segmentation in thoracic CT." *arXiv preprint arXiv: 1905.07710* (2019).

[32] J. Dong, W. Wang, and T. Tan. Casia image tampering detection evaluation database 2010. http://forensics.idealtest.org.

[33] J. Dong, W. Wang, and T. Tan. Casia image tampering detection evaluation database. In *ChinaSIP*, 2013.

[34] Y.F. Hsu, and S. Chang "Detecting Image Splicing Using Geometry Invariants And Camera Characteristics Consistency," in International Conference on Multimedia and Expo (ICME), Toronto, Canada, July 2006

[35] Nist nimble 2016 datasets. https://www.nist.gov/itl/iad/mig/ nimble-challenge-2017-evaluation/.

[36] I. Amerini, L. Ballan, R. Caldelli, A. Del Bimbo, G. Serra. "A SIFT-based forensic method for copy-move attack detection and transformation recovery", IEEE Transactions on Information Forensics and Security, vol. 6, issue 3, pp. 1099-1110, 2011.

[37] P. Ferrara, T. Bianchi, A. De Rosa, and A. Piva. Image forgery localization via fine-grained analysis of CFA artifacts. *TIFS*, 2012.

[38] N. Krawetz. A picture's worth... *Hacker Factor Solutions*, 2007.