International Journal of Advanced Computer Technology
ISSN: 2319-7900 www.ijact.org Volume-X, Issue-V, October 2021

SRU-NET: SOBEL RESIDUAL U-NET FOR IMAGE MANIPULATION DETECTION

Baoxiang Jiang, School of Electronic Science and Engineering, Xiamen University, Xiamen, China;
Jingbo Xia*, School of Information Science and Technology, Xiamen University Tan Kah Kee College, Zhangzhou, China;
Yanting Wang, School of Electronic Science and Engineering, Xiamen University, Xiamen, China;
*** Correspondence author**： jbxiad@sina.com;

Abstract:- *Recently, most successful image manipulation detection methods have been based on convolutional neural networks (CNNs). Nevertheless, Existing CNN methods have limited abilities. CNN-based detection networks tend to extract signal features strongly related to content. However, image manipulation detection tends to extract weak signal features that are weakly related to content. To address this issue, We propose a novel Sobel residual neural network with adaptive central difference convolution, an extension of the classical U-Net architecture, for image manipulation detection. Adaptive central differential convolution can capture the essential attributes of an image by gathering intensity and gradient information. Sobel residual gradient block can capture forgery edge discriminative details. Extensive experimental results show that our method can significantly improve the accuracy of localising the forged region compared with the state-of-the-art methods.*

Keywords: Image Manipulation Detection, Sobel Residual, Central Differential Convolution.

1 Introduction

With the continuous development of image editing technology, one can easily correct and correct the image content or manipulate the image generation process. The authenticity and challenge of the image have seriously affected people's trust in news reports and authenticity in the military economy. Researchers divide image content into three categories in the existing research scope: splicing, copy-move, and removal. Examples of these manipulations are shown in Figure 1. Splicing, copying a certain area in one image to another image to generate a new image. Copy-move, the same image Copy and paste part of the area to other locations in the figure. Removal, Remove the content from the image. Existing image manipulation detection methods have tried to use the same feature extraction methods to explore traces of image manipulation. Feature extraction methods for image operation detection can be divided into two categories: traditional detection methods based on Feature Extraction and detection methods based on convolutional neural networks (CNN). Copy-move, the same image Copy and paste part of the area to other locations in the figure. Removal, Remove the content from the image. Existing image manipulation detection methods have tried to use some feature extraction methods to explore traces of image manipulation. Feature extraction methods for image operation detection can be divided into two categories: traditional detection methods based on feature extraction and detection methods based on convolutional neural networks (CNN). Traditional feature extraction methods focus on the image generation process's statistical information and physical characteristics, such as colour interpolation, sensor noise, and other processing signals. Numerous algorithms for tampering passive detection have been proposed put forward [1-3]. However, the traditional image manipulation detection technology is only designed for a certain image attribute, so the final detection rate is low and lacks robustness. In recent years, with the continuous development of deep learning technology, especially the excellent performance of Convolutional Neural Network (CNN) [4, 22] represented by AlexNet in feature extraction, some researchers have adopted end-to-end networks[5-9], this network treats image manipulation detection as image segmentation or object detection task. The tamper detection technology based on the convolutional neural network uses the multi-layer structure of the deep learning network and powerful feature learning capabilities to achieve tamper detection that does not depend on the single attribute of the image, which makes up for the lack of applicability of traditional image tamper detection technology based on feature extraction shortcomings. All images have inherent properties by imaging processing. These properties are inherent to the image's content as image fingerprints, and they can be used to distinguish an image from other images. When splicing fake images, the tampered and un-tampered areas come from different source images. When removing and copying fake pictures, the edges of the operation area are inconsistent. However, the CNN detection network tends to strongly extract signal features related to the content, so the existing detection methods based on convolution cannot effectively extract the details of image fingerprint features and inconsistent edges generated during image manipulation. In summary, Existing CNN-based detection methods have not reached expectations for the extraction of image fingerprint features that are not related to image content. To solve this problem, we propose a novel network called SRU-Net for forgery detection. Our main contributions are summarised as follows: 1) A residual connected architecture named SRU-Net that achieves manipulation region segmentation. 2) An Adaptive Centre Different Convolution captures the image's essential attributes by gathering intensity and gradient information. 3)Residual Sobel Gradient Block, which enhances the ability to express manipulation region edge details. Experiments show that our method outperforms other state-of-the-art methods on the four benchmark datasets.

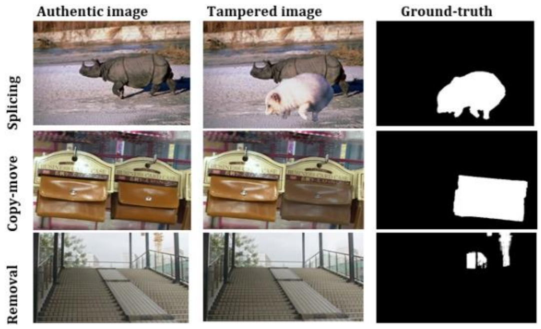

Figure 1. Examples of tampered images that have undergone different tampering techniques. From top to bottom are the examples showing manipulations of splicing, copy-move and removal.

2 RELATED WORKS

Manipulation Detection and Localisation mainly includes two methods: (1) Use the image feature extraction method to extract the image's statistical information and physical characteristics to detect the tampered region. Use the image, including dual JPEG compression [1], CFA [2], local noise analysis [3], and using noise inconsistencies for blind image forensics[10]. This research method usually focuses on complex manual feature construction, but it is difficult to determine which features should be extracted for many tasks. (2) DNN features that are fully implicitly learned instead of manual features. Rao et al. [5] used a convolutional neural network to detect digital images for the first time. This method uses CNN to learn feature-level representation from input RGB colour images automatically.

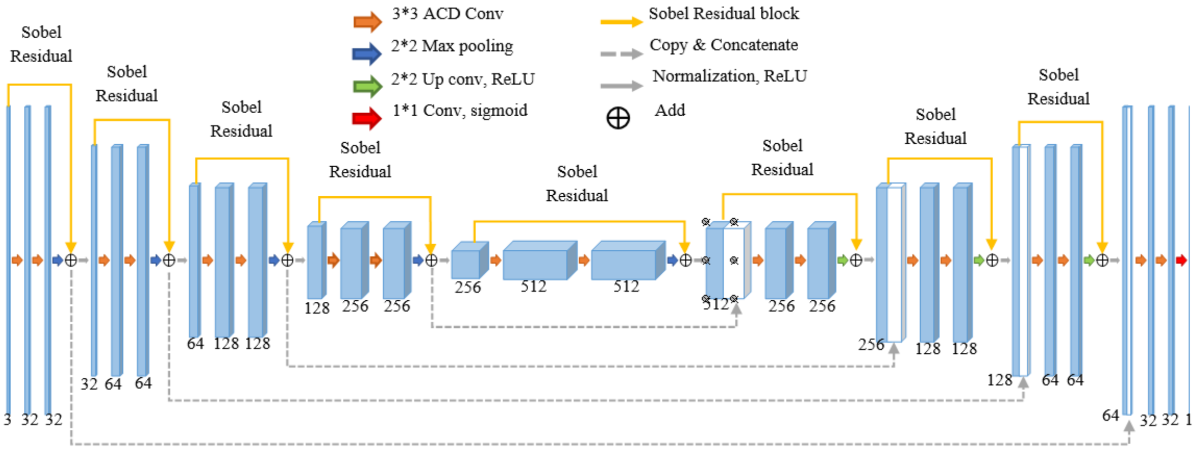

Figure 2. Overview of the Sobel Residual U-Net. The number on the box represents the number of features.

Zhang et al. [6] proposed a two-stage deep learning method based on convolutional neural networks to learn tamper Sampling features. The first stage uses an autoencoder model to learn each tamper feature, and the second stage integrates each tamper—contextual information of features for more accurate detection. BAPPY et al. [7] were inspired by a two-stage design algorithm and used a hybrid CNN-LSTM model to capture the distinguishing features between tampered regions and non-tampered regions. LSTM (Long Short Term Memory network model). To learn more features of image tampering, Zhou et al. [8] proposed a dual-stream Faster-RCNN network and trained it end-to-end to detect a given tampered image area. The network can accurately locate the tampered area and mark the type of tampering, such as whether it is copied and paste tampering. ManTra-Net [9] uses a self-supervised learning method to learn features from 385 types of tampering, and the tampering location problem is solved as a local Hardly any two DNN methods use the same network architecture, and most methods focus on a specific type of forgery. Central Difference Convolutional, [11]Propose a new convolution operator for live detection, called central differential convolution(called CDC), which can capture the inherent details of face images by the aggregating intensity and gradient information. Central difference convolution has two steps of sampling and aggregation, like vanilla convolution.

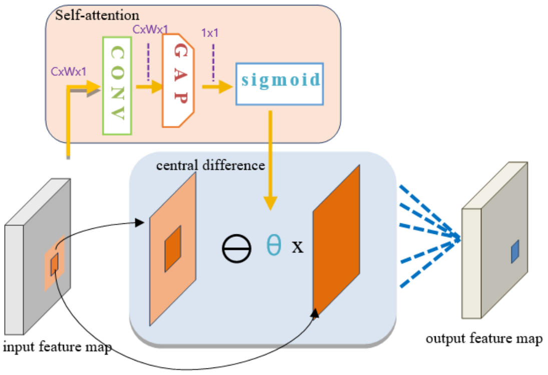

Figure 3. Adaptive central difference convolution.

The sampling step is similar to vanilla convolution; while the aggregation step is different, central difference convolution prefers to aggregate the centre-oriented gradient of sampled values. Aggregation steps of the central difference operator: The features sampled by each convolution kernel are subtracted from a certain proportion θ of the feature extracted from its centre. The inadequacy of central differential convolution is determined the θ by experimental methods requires considerable time and effort.

3 Method

In this section, We first provide an overview of the SRU-Net in section 3.1; the details of each module are introduced in section 3.2 and section 3.3. SRU-Net was developed to help localise the forged region using Sobel residual block and adaptive central differential convolution. The network architecture of SRU-Net is shown in Figure 2. It is an end-to-end image essence segmentation network and can directly localise the forged region. The differences between our SRU-Net and U-Net are two-fold. First, we replaced part of the U-Net convolution with an adaptive central differential convolution(ACDC), which can capture the essential attributes of the image by the aggregating intensity and gradient information. Second, we added the Sobel residual gradient block, which provides edge cues for image manipulation by extracting rich edge gradient information.

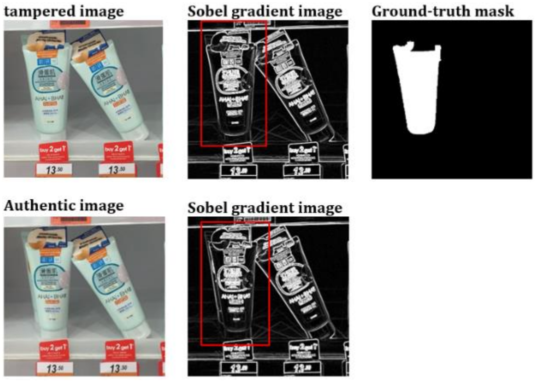

Figure 4. Sobel gradient magnitude difference between Authentic and tampered image.

3.2 Adaptive Centre Different Convolution

Convolution operators are the basis of convolutional neural networks. Convolution operators can achieve con-tent feature extraction. In image tampering detection, the image fingerprint features of the tampered area are weakly related to the image's content. Ordinary convolution cannot effectively extract image fingerprint features, and central differential convolution can capture the inherent details of temper images by aggregating intensity and gradient information. For this reason, we selected the central differential convolution operator to improve the network's ability to represent the details of image fingerprints. Centre Different Convolution(CDC) is performed in two stages, i.e., sampling and aggregation[11], 1)sampling local receptive field region over the input feature map x; 2) aggregation of sampled the centre-oriented gradient of sampled values. For each location p0 on the output feature maps:

$$y(p_0)=\underbrace{\sum_{p_n\in R} w(p_n)\cdot x(p_0+p_n)}_{\text{vanilla convolution}}+\theta\cdot\underbrace{\left(-x(p_0)\cdot\sum_{p_n\in R} w(p_n)\right)}_{\text{central difference term}} \qquad (1)$$

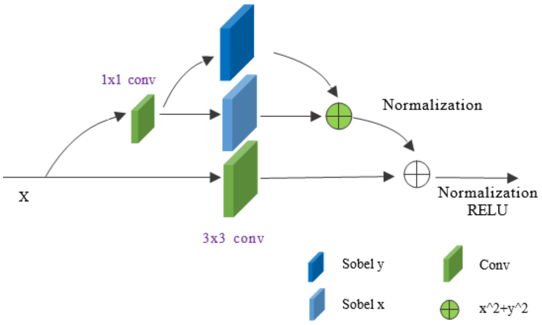

Figure 5. Residual Sobel Gradient Block

Although the CDC can improve the network's ability to represent the detailed information of image fingerprints, the CDC needs an operator to set the difference parameter θ manually. It is difficult to experimentally confirm the θ parameter for the CDC of each layer, so the θ parameter of each layer can only be set to the same value. It consider-ably limits the central difference of convolution fingerprints feature extraction. For this reason, we propose an adaptive central differential convolution(called ACDC), which learns the θ parameters of each ACDC layer by adding a self-attention path.

Table 1. Training and testing split for four datasets.

Datasets	NIST16	CASIA	Columbia	COVER
Training	404	5123	-	75
Testing	160	921	180	25

Adaptive Centre in the Different Convolution automatically updates the differential operator of each central differential convolution sampling area during the network training process. As illustrated in Figure 3, We divide the original convolutional network into two channels and share the input feature map. In the upper channel, the feature map first passes through the convolution layer to obtain the feature map of C*W*1, performs a global average pooling (pooling size is C*W) on it to obtain a 1 * 1 parameter, and finally sends it to the sigmoid for processing Normalise to obtain the difference parameter θ. Central differential convolution can use local zero-order intensity information and first-order gradient differential information. The first-order gradient differential information is conducive to capturing the detailed pattern inherent in the picture. However, the local zero-order intensity information is also important in the network, so we use the increased attention module to automatically learns the importance of local zero-order intensity information and first-order gradient difference information. The ablation study is in Section 4.2 to show the superior performance of ACDC for the image manipulation detection tasks.

3.3 Sobel Residual module

Image manipulation cause discontinuities in the correlation in the edge pixels of the tampered region. Hence, their edge differences between the real region and the tampered region. Figure 4 shows significant differences in edge gradient amplitude response between the real and forgery pictures.

Table 2. F1 score comparison on four benchmarks.

Methods	NIST16	COVER	Columbia	CASIA
NOI1 [15]	0.285	0.269	0.574	0.263
ELA [16]	0.236	0.222	0.47	0.214
J-LSTM [9]	0.57	-	0.612	**0.541**
RGB-N [5]	0.722	0.437	0.697	0.408
Ours	**0.843**	**0.544**	**0.748**	0.448

It provides an idea for designing the residual Sobel gradient block, which captures the splicing forged edge clues. We use the Sobel operation to calculate the edge gradient magnitude. The following convolution can obtain the horizontal and vertical edge gradients:

$$F_H(x) = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} \otimes x, F_V(x) = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} \otimes x \qquad (2)$$

Where $\otimes$ represents the deep convolution operation, x represents the input feature map, F_H represents is horizontal Sobel operation, and F_V represents the vertical Sobel operation. As shown in Figure 5, our SR block uses an advanced jump connection structure to aggregate learnable convolution features and gradient amplitude information, enhancing the ability to express fine-grained spatial details. We add the number of channels of the feature map changed by the Sobel operation feature map through 1x1 convolution and the subsequent residual addition to keeping the same channel number, the output of the Sobel residual path is defined as:

$$y_f = \mathrm{N}\left(F_V\left(F_1(x,\{W_i\})\right)^2 + F_H\left(F_1(x,\{W_i\})\right)^2\right) \qquad (3)$$

Where x represents the input features maps, and F1 represents 1x1 Conv denotes the normalisation layer. Then after passing through the normalisation layer and the Relu layer, perform the addition operation. The Sobel residual block is defined as Eq.(4)

$$y = \varphi\left(\mathrm{N}\ \left(F\left(x,\{W_i\}\right) + y_f\right)\right) \qquad (4)$$

Where φ is denotes as a Relu layer. The SR block provides edge cues for image manipulation by extracting rich edge gradient information.

Table 3. AUC comparison on four benchmarks.

Methods	NIST16	COVER	Columbia	CASIA
NOI1 [15]	0.487	0.587	0.545	0.612
ELA [16]	0.429	0.583	0.581	0.613
J-LSTM [9]	0.764	0.614	-	-
RGB-N [5]	0.937	0.817	0.858	0.795
MT-Net [7]	0.795	0.819	0.824	0.817
Ours	**0.912**	**0.867**	**0.892**	**0.842**

4 Experiments and Results

4.1 Performance Evaluation Metrics

Experimental Datasets: For evaluating the performance of the proposed method on image manipulation detection, we analysed and evaluated three public datasets, i.e., CASIA[12], COLUMB [13], NIST'16 [14], COVER[15]. The CASIA data set provides forged images such as stitching and copy movement of various targets and applies post-processing such as filtering and blurring. The CASIA 1.0 dataset includes 921 forgery images, and CASIA 2.0 dataset includes 5123 forgery images; we generated the ground truth masks by thresholding the difference between forgery and original images. The COLUMB dataset contains 180 forgery images and focuses on splicing based on uncompressed images, and the corresponding ground-truth masks are provided. The NIST16 dataset contains tampering methods such as splicing, copy-move, and deletion, including 564 tampered images. The COVER dataset contains 100 forgery images, and this data includes only splicing forgery. The NIST16 dataset and COVER dataset also provide ground-truth masks. We compare with other approaches on the same training and testing split; all of the experimental data is listed in Table 1. On CASIA, we use CASIA 2.0(5123 images) for training and CASIA 1.0(921 images) for testing. On COLUMB, Columbia is only used for testing the model trained on our synthetic dataset. On NIST16, we randomly divided into 404 images for the training set and 160 images for the testing set. Similarly, on COVER, 75 sets of images are chosen as the training set, and 25 sets as the test set.

Evaluation criteria

We evaluate the model's performance by calculating the pixel-level F1 score and area Under the model's receiver operating characteristic curve (AUC) on the test dataset. F1 score is the most commonly used evaluation criteria for image manipulation detection.

Table 4. Comparison of SRU-Net variants evaluated with F1 metric on four benchmarks

Methods	NIST16	COVER	Columbia	CASIA
RU-Net [15]	0.529	0.287	0.585	0.226
RU-NET(CDC)	0.669	0.425	0.692	0.318
RU-NET(ACDC)	0.754	0.484	0.743	0.383
SR-UNET	0.843	0.544	0.824	0.448

Implementation Details

We implement our framework in PyTorch, using an NVIDIA 3080TI GPU. We trained all our models to minimise the cross-entropy loss function and stochastic gradient descent as the optimiser in this work. For N samples, the cross-entropy loss is computed as:

$$L_{\text{loss}} = \frac{1}{N}\sum_i \left\| m_i \log\left(p_i\right) + \left(1-m_i\right)\log\left(1-p_i\right)\right\|_1 \qquad (5)$$

Prediction mask pi and ground-truth mask mi corresponding to the ith input xi. The learning rate is 0.01, and the mini-batch size is four images. The learning rate decays by 0.1 every five steps, and on CDC, we choose centre difference parameter θ is 0.7 for comparing with ACDC.

4.2 Compared Detection Methods and Result

We comparing our proposed approach with following baseline models: error level analysis (ELA)[16], colour filter array (CFA) [2], noise inconsistency (NOI) [10], M-FCN[17], J-LSTM[18], RGB-N[8], ManTra-Net[9]. ELA, An error level analysis method [16], analyses the JPEG compression qualities and calculates its error level to determine whether the picture has manipulation. CFA[2] detects image tampering by analysing the image's colour filter traces during the interpolation process. NOI1[10] using high pass wavelet coefficients to model local noise. MFCN[17], a multi-task fully convolutional network has two branches; one is used to predict the image tampering regions, the other is used to predict the edge of the tampering regions. J-LSTM[18], an LSTM based network, jointly train patch level tampered edge classification and pixel-level tampered region segmentation. RGB-N[8], a dual-stream Faster R-CNN network a dual-stream Faster R-CNN network, combines the RGB stream and the noise stream obtained through SRM convolution to detect the tampered area of the image. ManTra-Net[9] uses a self-supervised learning method to learn features from 385 tamper types, and the tamper location problem is solved as a local outlier detection problem. Table 2 describes the comparison of F1 scores between our method and the baseline. Table 3 shows the comparison of AUC. From the results in Table 2 and Table 3, it can be found that our method has significant improvements on the NIST16 and COVERAGE datasets compared to the state-of-the-art methods. In particular, the F1 score of our model is on the NIST16 and COVERAGE datasets. Increased by 0.121 and 0.107, respectively, the growth rates were 16.7% and 15.3%. In terms of AUC performance, the performance of our model on the four data sets increased by more than 10%.

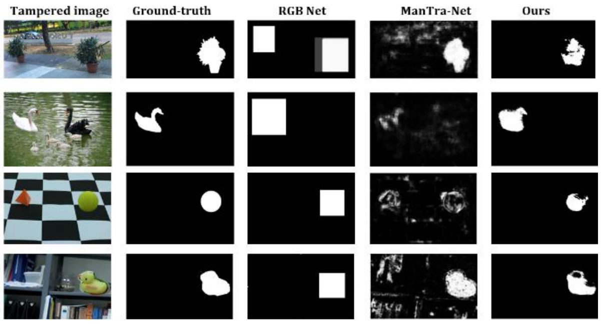

Figure 6. Qualitative visualisation of localisation results.

Ablation studies

We explored the contribution of each proposed component to the final performance. Our baseline uses the RU-Net structure. First, we explore (1) the influence of central differential convolution on image tampering detection, (2) the difference between adaptive central differential convolution and central differential convolution with fixed differential values (3) The role of residual link with Sobel. Among them, we selected the best-performing 0.7 for comparison through experiments. The comparison of different variants is shown in Table 5. Using adaptive central differential convolution instead of ordinary convolution and our proposed link with Sobel residuals have significantly improved performance on all data sets. We think this is Because adaptive central differential convolution is better than ordinary convolution operators for extracting essential image attributes, and the extraction of edge gradient information is helpful to detect the tampered area of the image.

Qualitative Comparison

In Figure 6, the detection results of SRU-Net and RGB-N [5] and ManT-Net [7]. As shown in Figure 6, SRU-Net produces good manipulation technology classification performance and is better than other models. The manoeuvrability and positioning performance. It is because adaptive central differential convolution can better extract the essential attributes of the image. By extracting edge gradient information, it is helpful to ensure more accurate segmentation results under various tampering attacks.

5 Conclusion

In this paper, we propose SRU-Net, which extracts the essential attributes of the image through an adaptive central difference convolution operator, and uses the Sobel residual link to obtain edge gradient information to realise image operation detection in a variety of tampering methods. SRU-NET is superior to the most advanced models. This method is accurate and robust in general operation detection and positioning, which shows that acquiring edge gradient information helps capture the basic information in operation positioning.

References

[1]. T. Bianchi, A. De Rosa, and A. Piva. Improved duct coefficient analysis for forgery localisation in jpeg images. InICASSP, 2011. 2, 3.

[2]. M. Goljan and J. Fridrich. Cfa-aware features for steganalysis of colour images. In SPIE/IS&T Electronic Imaging, 2015.1, 3.

[3]. D. Cozzolino, G. Poggi, and L. Verdoliva. Splicebuster: A new blind image splicing detector. In WIFS, 2015.

[4]. Fridrich, J., Soukal, D. and Lukas, J. (2003) Detection of Copy-Move Forgery in Digital Images. Proceedings of Digital Forensic Research Workshop, Cleveland, August 2003, 55-61.

[5]. Y. Rao and J. Ni. A deep learning approach to detection of splicing and copy-move forgeries in images. In WIFS, 2016.3.

[6]. ZHANG Ying, GOH J, WIN L Shukla, et al. Image region forgery detection: A Deep Learning Approach[M]. MATHUR A and ROY CHOUDHURY R.

Proceedings of the Singapore Cyber-Security Conference. Amsterdam: IOS Press, 2016: 1–11.

[7]. Bappy M J H, Roy-Chowdhury A K, Bunk J, et al. Exploiting Spatial Structure for Localising Manipulated Image Regions[C]// International Conference on Computer Vision (ICCV), 2017. IEEE Computer Society, 2017.

[8]. Zhou, Peng, Han, Xintong, Morariu, Vlad I. Learning Rich Features for Image Manipulation Detection[J].

[9]. Y. Wu, W. AbdAlmageed, and P. Natarajan, "Mantra net: Manipulation tracing network for detection and localisation of image forgeries with anomalous features," in CVPR, 2019, pp. 9543–9552.

[10]. B. Mahdian and S. Saic, "Using noise inconsistencies for blind image forensics," Image and Vision Computing, vol. 27, no. 10, pp. 1497–1503, 2009.

[11]. Yu, Zitong, et al. "Searching central difference convolutional networks for face anti-spoofing." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.

[12]. J. Dong, W. Wang, and T. Tan, "CASIA image tampering detection evaluation database," in ChinaSIP, 2013, pp. 422–426.

[13]. Columbia image splicing detection evaluation dataset," http://www.ee.columbia.edu/ln /dvmm/ downlod /AuthSpliced DataSet /AuthSpliced DataSet.htm,2004.

[14]. "Nist 2016 datasets," https://www. nist.gov/sites/ default /files/ documents /2016/11/30/should_i_believe or not.pdf.

[15]. B. Wen, Y. Zhu, R. Subramanian, T. Ng, X. Shen, and S. Winkler, "Coverage a novel database for copy-move forgery detection," in ICIP, Sep. 2016, pp. 161–165.

[16]. Krawetz, N., Solutions, HF: A picture's worth. Hacker Factor Solutions 6(2), 2(2007)

[17]. R. Salloum, Y. Ren, and C.-C. J. Kuo. Image splicing localisation using a multi-task fully convolutional network(mfcn). arXiv preprint arXiv:1709.02016, 2017. 3, 6.

[18]. J. H. Bappy, A. K. Roy-Chowdhury, J. Bunk, L. Nataraj, and B. Manjunath. Exploiting spatial structure for localising manipulated image regions. In ICCV, 2017. 1, 3, 6.

[19]. M. Zampoglou, S. Papadopoulos, and Y. Kompatsiaris, "Large-scale evaluation of splicing localisation algorithms for web images," Multimedia Tools Appl., vol.76, no. 4, pp. 4801–4834, 2017.

[20]. O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention, pages 234–241. Springer, 2015.

[21]. K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages770–778, 2016.

[22]. Arvind Mewada, and Rupesh Kumar Dewang. "Research on False Review Detection Methods: A state-of-the-art review." Journal of King Saud University-Computer and Information Sciences (2021).