A Deep Learning and Machine Learning Approach for Image Classification of Tempered Images in Digital Forensic Analysis

Praveen Chitti[1], Dr. K. Prabhushetty[2], Dr. Shridhar Allagi[3]
Department of Electronics and Communication Engineering[1]
Jain College of Engineering, Visvesvaraya Technological University, Belagavi[1]
Department of Electronics and Communication Engineering[2]
Veerappa Nisty Engineering College, Shorpur
Visvesvaraya Technological University, Belagavi
Department of Computer Science and Engineering, KLE Institute of Technology, Hubballi[3]

***Abstract*—Multimedia images are the primary source of communication across social media and other websites. Multimedia security has gained the attention of modern researchers and has posed dynamic challenges such as image forensics, image tampering, and deep fakes. Malicious users tamper with the image embedding noise, leading to misinterpretation of the content. Identifying and authenticating the image by detecting the forgery operations performed on it is essential. In our proposed model, we detect the forged region using the machine learning model SVM in the first iteration and Convolution Neural Network in the second iteration with Discrete Cosine Transform (DCT) for feature extraction. The proposed model is tested with a Corel 10K dataset, and an average accuracy of 98% is obtained for all kinds of image operations, including scaling, rotation, and augmentation.**

Keywords—Support Vector Machine (SVM); Discrete Cosine Transform (DCT); Convolution Neural Network (CNN); Image Forensics and Image Forgery

I. INTRODUCTION

With the invention of high-speed networks and advanced storage technologies, the internet used in almost every domain has increased drastically. Significant technological innovations have occupied every area with intelligent technologies and devices. The use of multimedia in social networking and other applications has invited other challenges where intruders modify or forge the data for illicit usage. Digital transformations have opened up a new domain of image forensic analysis. The need for providing authentic documents with integrity is a much-needed research area.

Recent Information and Communication Technology (ICT) has a vast spectrum of applications imbibing from each machine to pools of networks by large organizations. Users over the globe are not limited to work applications but are used for various applications such as personal banking transactions, multimedia data sharing across social media platforms, etc. [1] Larger companies regularly conduct their every transaction and operation across the internet, and more extensive data migrations across the firms occur via the networks comprising lesser secured public internet.

The drastic inventions in the cloud and big data technology solutions have seen exponential usage in social networking and mobile communication applications, leading to more extensive organization information flow operations. This limited the organization's security architecture for handling manipulated and tampered data. The increased activity of users and organizations has led to the misuse of computers and network-related applications ranging from simple password hacks to unauthorized access to servers.

Increased misuse of network and computer-related applications has led to increased research in computer-related investigations. Traditionally auditing the logs was a simple way to trace fraud operations, and the advanced hacking techniques are so powerful that no traces are found in the records. Hence the need for an automated and intelligent way of forensics is necessary. Digital forensics has made advanced developments over the years. The primary reason for the development is intelligence in tools and techniques that use local hardware machines to perform complex auditing tasks more precisely. The difficult task of auditing the logs is automated using machine learning models.

The rapid increase in the digital content comprising images over the internet has challenged the retrieval efficiency of several applications. The content-based image retrieval (CBIR), the effective and modern method for retrieval of images from the web, has addressed the challenge to some extent. CBIR is defined as the operation for retrieving images over the more extensive database in an effective and timely manner using dominant features of stored images such as texture, color, shape, etc. The effectiveness and efficiency of any CBIR model rely entirely on the identified extracted feature subset, as these values are used in the computation of similarity among the stored and retrieved images. Fig. 1 shows the classic examples of image tempering instances.

The primary purpose of the research is to secure the machine learning-based models for forensic analysis in the training and testing phase of the proposed architecture. The primary model experiments with the Support Vector Machine and the transformation functions applied with feature reduction for training the classifier. In the next level, forensic analysis is performed on the images, which have the specific traces acquired by the operations, such as coding, acquisition, and preprocessing. To enhance the robustness of the model, the Convolution Neural Network (CNN) has been integrated for additional training of the model. CNN is the programming model that aids the machine in learning from the operational data provided. It is a subset of the class for a deep neural network that proactively addresses several image processing operations such as pattern matching and recognition. The model is treated to be complex and interconnected with several neurons in each of the layers.

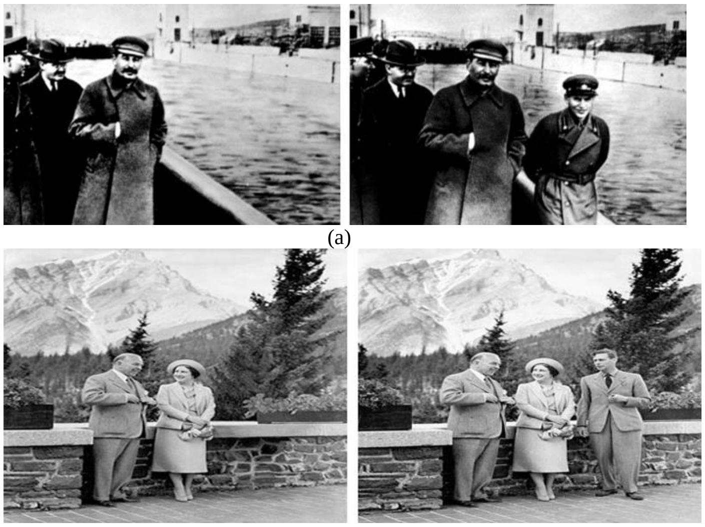


Fig. 1. Image Tempering Throughout the History [Courtesy: https://twistedsifter.com/2012/02/famously-doctored-photographs/].

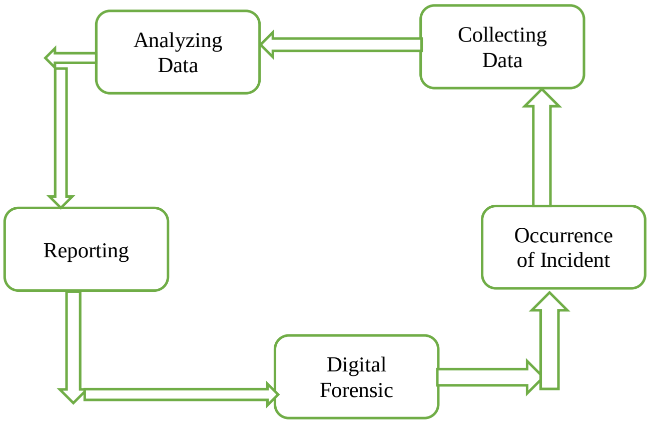


Fig. 2. Dimensions of Digital Forensics.

Fig. 2 provides the dimensions of digital forensics. Digital forensics involves the process of incident occurrence, analyzing collected data, and reporting the traces of anomalies in it.

II. Related Works

In [2] the work used pre-trained models for the lifecycle of forensic digital imaging tools. The work focused on the detection of a gun in a set of images. The work focused on the four pre-trained models, such as ImageNet-trained models: InceptionV3, Xception, ResNet, and VGG16, on the Adhoc realistic datasets without any of the fine-tuning of the models used. The results were on par with the realistic determination of guns in the images. The work was limited to a particular class and could be extended with a few more classes to make the model more realistic.

In [3][4] the authors discussed the various needs for using machine learning and deep learning models and their implications. The various experimentations in the article show the augmenting of the performance of traditional models with the integration of machine learning models and the development of the conceptual framework for digital forensics.

The work in [5][6] proposed the support vector machine learning model with DWT and PCA as feature reduction models. The proposed model has experimented with an Adhoc dataset comprising 10000 images, and the proposed work was compared against the decision tree and Bayesian model. The proposed architecture gave the optimized performance of an accuracy of 97% in a limited subset of training images.

In [7] they worked on detecting tampering on the images using the CFA artifacts. The model provided optimal results in the detection of forged region localizations. The model has experimented on the UCID dataset and several images pulled from various social media websites. The model used the scalar-based approach with forensic analysis by machine learning. The experimental results were optimized with this approach and limited to a few classes. The procedure failed when experimented with multi-class label datasets.

The model in [8] addressed image forgery on social media and other prominent websites. The work focused on copy-move forgery detection using block processing and extracted the features transformed from the blocks. The convolutional neural network was for detecting the forgery. The serial pairs of convolutions are used for extracting the features and then enable the classification of images as tampered and originals, including the consideration of transformation operations. For Experimentation, various trigonometric transforms for 1D and 2D are considered. The model gave the optimum accuracy but was limited to the specifically trained dataset and required a robust model to deal with all kinds of images.

The experiments in [9] discussed the significant risks and shortcomings of using CNN models in clinical applications. The research focused on noise discovered in medical images and its impact on deteriorating the model's performance. They proposed a defense mechanism to such noises by incorporating the sparsity of denoising methods performed inherently in the CNN models to enhance the model's accuracy and overall performance. The model gave an accuracy of up to 97%.

The model in [10] discussed the impacts of adversarial perturbations on CNN and its challenges. The model is designed on the new architecture that specifically increases the robustness of the adversarial effects by using feature denoising. The network design consists of blocks that denoise each feature using various nonlocal means and a combination of other filters. The model has experimented on Imagenet, and the model's performance is enhanced by embedding this method in machine learning models.

The work by authors in [11] addressed the problem with the authenticity of the images shared across social media and other websites. They proposed a digital image forensics model to identify the images' averaging and gaussian filtering. The model first normalized the image and computed the difference in array values for calculating the co-occurrence features. The model achieved higher accuracy for even small-size images with less resolution, but the results were minimal for the higher-size images.

The authors in [12] proposed a model based on the pixel-pair histogram (PPH) and coefficients of an autoregressive moving average model (ARMA). These features are extracted from various directions in an image for computing the median filters. Experimentations were done on multiple single and compound databases, and the model enhanced the accuracies of the models, especially for the JPEG formats and compressed versions of the images.

III. Proposed Methodology

The proposed methodology works in two layers. In the initial phase, the support vector machine is used with the RBF kernel in training the images. In the second layer, CNN is used for training the model for making the classification more robust. Fig. 3 shows the proposed model for forensic image detection.

A. *Dataset*

To test the proposed model's robustness, experimentations were conducted on a target set of 100000 images. The images span different categories, such as oceans, mountains, fruits, etc. Every image is scaled to 192 X 128. The dataset is spread across 100 categories, composing 100 images in each type. Fig. 4 shows the sample images in the dataset.

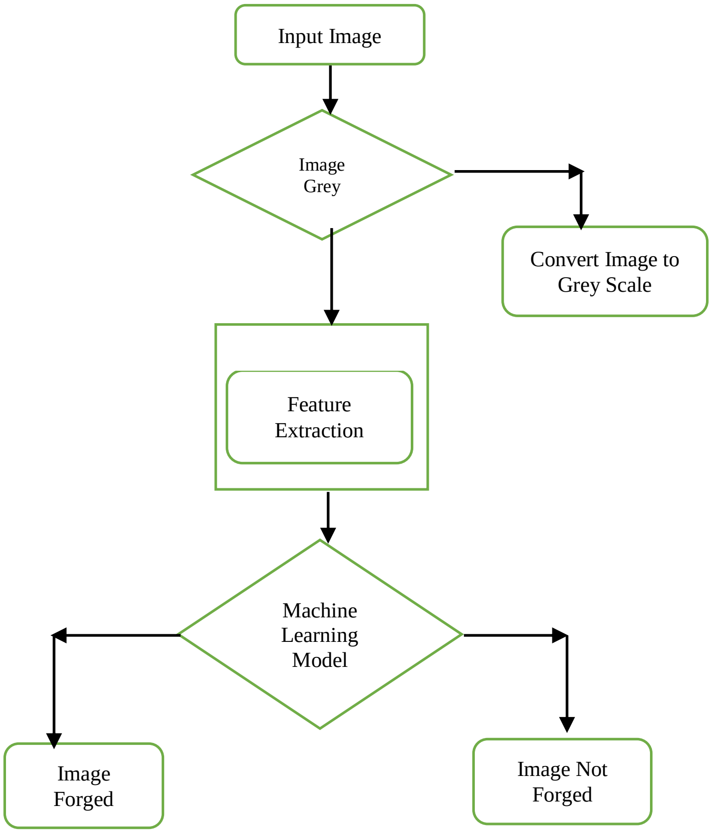


Fig. 3. Proposed Model for Forensic Image Detection.

Fig. 4. Sample Images in the Corel -10K Image Dataset.

Each image has been extracted with 89 features. These features are composed of four sets based on color histogram layout, histogram, co-occurrence, and color moments.

TABLE I. Feature Set Description

Feature	Dimensions	Description
Color Histogram	32	HSV is divided into 32 subspaces. 8 for H and 4 for S.
Color Histogram layout	32	Image is divided into four sub-images
Color Moments	9	3 for H, S, and V
Co-occurrence Texture	16	co-occurrence in 4 directions is computed (horizontal, vertical, and two diagonal directions)

B. *Feature Extraction and Feature Reduction*

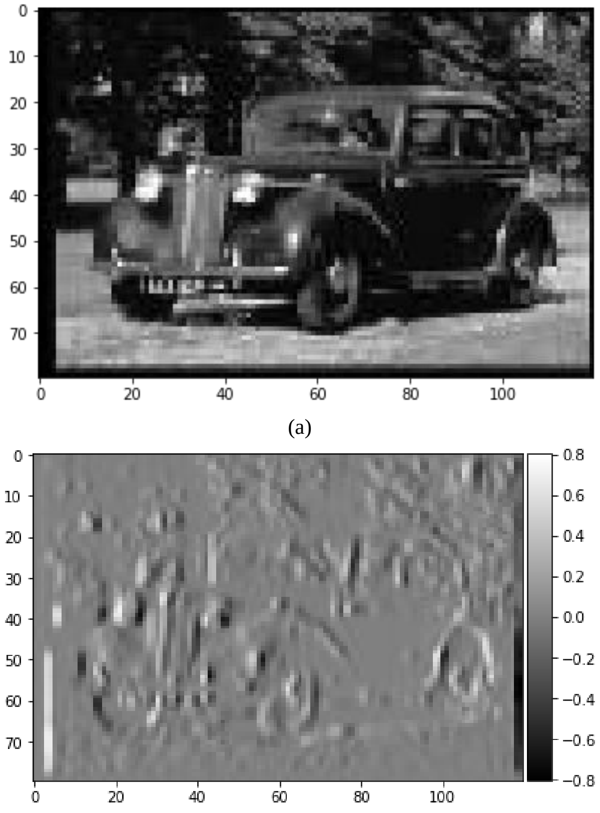
```
Algorithm: Feature Extraction
Input: Image I
Output: 89 Features i1, i2,….. in
Step 1: Convert RGB to HSV Image.
Step 2: Split the input image into H, S, and V sub-image spaces
Step 3: Canny Operator is used for edge detection and segmentation
Step 4: Discrete Cosine Transform (DCT) is used for feature extraction
Step 5: 89 feature sets are extracted from the image
```

The DCT coefficient entities signify the spatial frequency components in an image, and every pattern of the blocks in an image is augmented with various magnitudes. The extracted features depend on the prominent edge, and the values extracted signify the image's lowest frequencies.

Fig. 5 (a) shows the original image and the extracted features from HSV subspaces in images. Fig. 5 (b) shows the sample extracted feature matrix.

In computing the DCT coefficients matrix, the major component for the spectral coefficients of images represent the lower frequency sections and higher frequency sections with amplitude of areas across the image. Since the most dominating area of interest is around the lower frequency, we discard the values closer to 0. The coefficients of DCT are reduced w.r.t to (1).

$$D\_C\ (u,v,s)=F(u,v)\ u,v=1,2,3,\cdots\cdots.s\ \ 1<s\leqslant 8\ \cdots \qquad (1)$$



(a)

(b)

Fig. 5. (a) Original Image. (b) Image Feature Extraction.

Where F(u,v) Represents the DCT coefficients of image block I. D_C (u,v,s) Represent the reduced DCT Coefficient. S represents the reduction scale for the image.

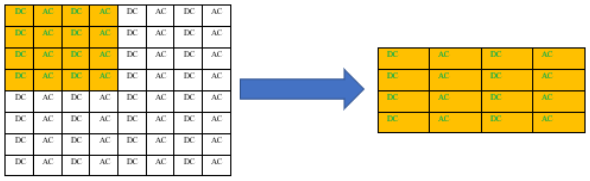
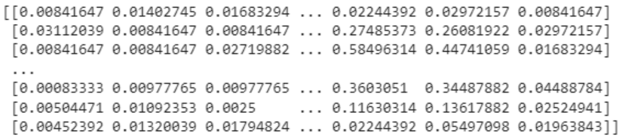


Fig. 6. DCT Reduced Coefficients from 8 X 8 to 4 X 4.

Since the DC coefficients among the adjacent blocks of images are redundant and the AC coefficients are small, these values are ignored. DCT transforms the correlation among the adjacent 8 X 8 blocks of images and is quite dominant. Hence smoothing these edges is necessary. The operation is performed using DCT coefficient values, which will uniformly distribute the frequency across the images. The reduced and smoothened coefficients of DCT are computed as of (2). The sample features extracted using DCT are shown in Fig. 7.

$$D_C = \begin{cases} F(u,v) \;\; (u,v) = 1 \\ F(u,v) + \frac{F(1,1)}{8} \;\; (u,v = 2, \ldots . s) \end{cases} (1 < s \leq 8) \qquad (2)$$



```
[[0.00841647 0.01402745 0.01683294 ... 0.02244392 0.02972157 0.00841647]
 [0.03112039 0.00841647 0.00841647 ... 0.27485373 0.26081922 0.02972157]
 [0.00841647 0.00841647 0.02719882 ... 0.58496314 0.44741059 0.01683294]
 ...
 [0.00083333 0.00977765 0.00977765 ... 0.3603051  0.34487882 0.04488784]
 [0.00504471 0.01092353 0.0025     ... 0.11630314 0.13617882 0.02524941]
 [0.00452392 0.01320039 0.01794824 ... 0.02244392 0.05497098 0.01963843]]
```

Fig. 7. Feature Matrix using DCT.

IV. Machine Learning Models

In the first phase, the support vector machine with the configuration in Table II is used to train the model. Support Vector Machine (SVM) is majorly used in classification, and recognition tools are embedded to avoid computational complexities. Our work proposes a methodology for detecting the forged image, which might comprise any subpart if the image has been added, removed, or altered. The SVM is used for identifying the similar neighboring regions of an image by matching with other blocks of the image computed. To identify a forged part in an image, the features are extracted w.r.t to HSV, texture, pixel value, and the edges of several regions in an image. The process works in two iterations. In the first iteration, the model is trained with the images without forged parts. The second iteration tests the model with the sample set of images containing the forged part. The SVM model is used to classify the images into two classes: forged or genuine. The model initially identifies the edges and decides the decision boundaries in the training phase. This information will be used for generalizing the images with higher dimensions. A decision space (support vectors) for the set of images is generated using the larger space of trained images which separates the objects belonging to a different class.

All the experiments were conducted on Ubuntu 22.0 LTS 64-bit Operating System with 24GB Ram, intel core i5 @3.40Ghz with python 3.0. The entire dataset was divided in the ratio of 70:30 for the training and testing dataset. In the initial phase, the training is done with a support vector machine with rbf kernel, as mentioned in Table II.

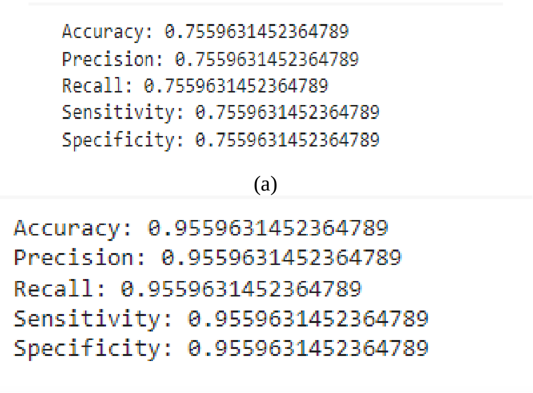
TABLE II. Configuration of SVM Model

Model	**Kernel**	**γ**
Support Vector Machine	rbf	1/n

To make the model more robust, it is trained with the CNN model as specified in Table III in the second phase.

TABLE III. CNN Model Configuration

Image Layer	**Filters**	**Feature map size**	**Size for filter**	**Strides**	**Padding**
1st CNN Layer	664	224*224*64	3*3*3	1	1
2nd CNN Layer	64	224*224*64	3*3*64	1	1
Max Pooling	1	112*112*64	2*2	2	0



```
Accuracy: 0.7559631452364789
Precision: 0.7559631452364789
Recall: 0.7559631452364789
Sensitivity: 0.7559631452364789
Specificity: 0.7559631452364789
```

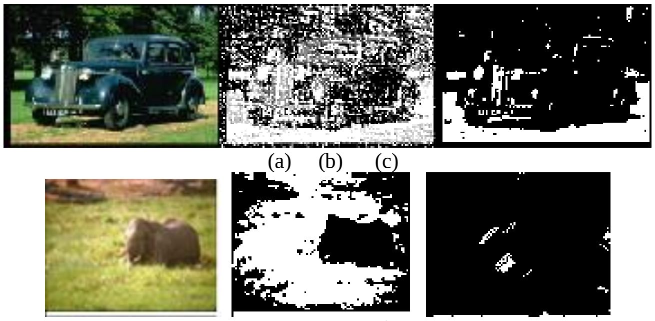
(a)

```
Accuracy: 0.9559631452364789
Precision: 0.9559631452364789
Recall: 0.9559631452364789
Sensitivity: 0.9559631452364789
Specificity: 0.9559631452364789
```

(b)

Fig. 8. (a) Performance Metric of SVM (b) Performance Metric of CNN.

Fig. 8 shows the performance metrics computed for the support vector machine and the CNN. Fig. 9 gives the different forged images detected.



(a) (b) (c)

(d) (e) (f)

Fig. 9. Results of Proposed Model (a) Input Image (b) Threshold Image (c) Classified as Not Forged (d) Input Image (e) Threshold Image (f) Classified as Forged Image.

V. Conclusion

In recent research, many researchers proposed various methods for forgery detection, which can be further categorized into supervised and unsupervised classification models. The two models experimented with, in our work, support vector machine and the convolution neural network have performed well, with an accuracy of 75% and 95%. The CNN model wrongly classified the few forged images to non-forged. These misclassifications were handled correctly in the SVM. Hence the model proposed is trained in two iterations with respect to SVM and CNN. The combined classifier gave an accuracy of over 98%. The proposed model can be further enhanced with an optimal feature reduction mechanism, and using unsupervised models may augment the model's performance.

References

[1] https://ciosea.economictimes.indiatimes.com/blog/how-ai-will-transform-digital-forensics-in-2022-and-beyond/91141155.

[2] Del Mar-Raave, J. R., Bahşi, H., Mršić, L., & Hausknecht, K. (2021). A machine learning-based forensic tool for image classification - A design science approach. Forensic Science International: Digital Investigation, 38, 301265. https://doi.org/10.1016/j.fsidi.2021.301265.

[3] Qadir, A. M., & Varol, A. (2020, June). The Role of Machine Learning in Digital Forensics. 2020 8th International Symposium on Digital Forensics and Security (ISDFS). https://doi.org/10.1109/isdfs49300.2020.9116298.

[4] Solanke, A. A. (2022, July). Explainable digital forensics AI: Towards mitigating distrust in AI-based digital forensics analysis using interpretable models. Forensic Science International: Digital Investigation, 42, 301403. https://doi.org/10.1016/j.fsidi.2022.301403.

[5] Qadir, S., & Noor, B. (2021, May 20). Applications of Machine Learning in Digital Forensics. 2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2). https://doi.org/10.1109/icodt252288.2021.9441543.

[6] Monika, & Passi, A. (2021, August 1). Digital Image Forensic based on Machine Learning approach for Forgery Detection and Localization. Journal of Physics: Conference Series, 1950(1), 012035. https://doi.org/10.1088/1742-6596/1950/1/012035.

[7] Singh, G., & Singh, K. (2020, March). Digital image forensic approach based on the second-order statistical analysis of CFA artifacts. Forensic Science International: Digital Investigation, 32, 200899. https://doi.org/10.1016/j.fsidi.2019.200899.

[8] Al_Azrak, F. M., Sedik, A., Dessowky, M. I., El Banby, G. M., Khalaf, A. A. M., Elkorany, A. S., & Abd. El-Samie, F. E. (2020, March 2). An efficient method for image forgery detection based on trigonometric transforms and deep learning. Multimedia Tools and Applications, 79(25–26), 18221–18243. https://doi.org/10.1007/s11042-019-08162-3.

[9] Robust convolutional neural networks against adversarial attacks on medical images. (2021, December). Pattern Recognition, 132, 108923. https://doi.org/10.1016/j.patcog.2022.108923.

[10] Cihang Xie, Yuxin Wu, Laurens van der Maaten, Alan L. Yuille, Kaiming He; Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 501-509.

[11] Agarwal, S., & Jung, K. H. (2021, January 31). Image Forensics using Optimal Normalization in Challenging Environment. 2021 International Conference on Electronics, Information, and Communication (ICEIC). https://doi.org/10.1109/iceic51217.2021.9369794.

[12] Gao, H. (2020, January 21). Detection of median filtering based on ARMA model and pixel-pair histogram feature of difference image. SpringerLink. Retrieved September 7, 2022, from https://link.springer.com/article/10.1007/s11042-019-08340-3?error=cookies_not_supported&code=4b1663fa-e0a8-49e4-bd2b-bc3865481542.