ISSN 1816-6075 (Print), 1818-0523 (Online)
Journal of System and Management Sciences
Vol. 12 (2022) No. 6, pp. 454-467
DOI:10.33168/JSMS.2022.0628

Deep Learning to Detect Image Forgery Based on Image Classification

Mamdouh M. Gomaa, Eman R. Mohamed, Alaa M. Zaki and Alaa Elnashar

Computer Science Department, Minia University, Minia, Egypt

mamdouh.gomaa@mu.edu.eg, alaa_zaki@mu.edu.eg, alaa.elnashar@mu.edu.eg, emanreda995@gmail.com

Abstract. Nowadays, Digital images can be seen in magazines, newspapers, hospitals, shopping malls, on the Internet, and among other places. As technology advances, at the same time, the trust in images is decreasing day by day because the easy to forgery in these images. One of the major topics for the researcher is the detection of forgery in images, and copy- move (CMFD) is one of main types of image forgery. The majority of CMFD algorithms now in use relies on key-point or block approaches, individually or merges of them. Many deep convolutional neural network (CNN) methods have recently been used in image classification and image forensics to outperform more conventional techniques. In this paper, we proposed a new method for image forgery detection using a CNN. CASIAV1, CASIAV2, and Columbia datasets are used in the proposed method. The pre-trained (CNN) is used to extract dense features from the test images for Support vector machine (SVM) and K-Nearest neighbor (KNN) classification. The model is designed, implemented and tested. From the experimental results we can observe that the accuracy is 98.22% for CASIAV1, 97.02% for CASIAV2, and 85.1% for the Columbia dataset.

Keywords: image forgery detection, support vector machine (SVM), deep learning (DL), convolutional neural networks (CNN), k-nearest neighbor (KNN)

1. Introduction

Since they are considered as a key source of information, electronic images have recently become increasingly important. It is fairly simple to make a forged image using different tools. When an image is used as proof in a legal court, the originality of the image becomes extremely important. Image manipulation, often known as image editing, which can define as any modification made to digital images through the use of any software. Passive and active authentication are the two different types of authentications (Derroll and Divya 2015). Active type is achieved through the use of several methods, like digital signatures and cryptography. Active authentication requires that the original content of the test image be available for comparison with the test image (Vartak and Deshmukh 2014). Different payload partitioning methods and data-embedding algorithms are connected to active authentication, which is used to secure colored image steganography. The image's original content is not available during passive authentication. This type of authentication is frequently used to identify forgery images as the image's original content of the image under examination is unavailable. As fragments of evidence, forgery detection in digital images is a crucial issue. It is carried out by looking for peculiar features, characteristics, or diseases. The various techniques that have been employed to create forgery images can be divided into three groups: copy- move (CMF), image resampling and spliced images.

To hide or add features to an image, CMF involves copying and pasting a part of the image, regardless of its size and shape, in a different position. Since the forged part is produced from the same original image, its basic characteristics, such as texturing and brightness, are identical to those of the real part, making this type of forgery difficult to recognize.

The structure of this paper is as follows, the relevant work is shown in the next section. A proposed method in Section 3. Section 4 describes the results and presents a discussion and analysis of the results. Section 5 gives the conclusion.

2. Related work

Digital image became the primary source of information in today's modern world. The field of research known as digital forensics is dependent on blind investigative techniques. Digital image forensic methods work by detecting image forgeries without having any prior information of the original content. CMFD algorithms are available in two types, deep learning and traditional algorithms. Traditional algorithms are those that depend solely on the consistency of the image's statistical properties (Farid 2009).

The two primary types of CMFD are keypoints algorithms and block-based. Block-based, which vary in their approach and algorithm, divide the images into circular or overlapping rectangular blocks. Then, using different feature extraction methods, the features from each block are extracted. Block-based algorithms

implemented a variety of feature extraction methods involving invariant moments, frequency transformations, and changes to the intensity, texture, and color of the image (Warif et al., 2016). The last step is to identify related blocks based on their features using a matching technique (nanda et al., 2014). To identify forgery images, many techniques include the Discrete Wavelet Transform (DWT), Principal Component Analysis (PCA), Singular Value Decomposition (SVD), Discrete Cosine Transform (DCT), Dyadic Wavelet Transform (DyWT) and their combinations are employed (Warif et al., 2016).The first algorithm, developed by Fridrich et al. (2018), utilized DCT for CMFD. In this approach, the image is split into overlapping blocks, and for each block, the DCT is determined. Local Binary Patterns (LBPs) were used for CMFD by Ahmet Boz et al. (2016) in addition to DCT. Each LBP block is applied to the DCT. To find related blocks, the lexicographic feature matrix is sorted. The method in (Kang et al., 2012) used both DCT and SVD. After the image has been separated into blocks, each block's features are extracted using the DCT. The SVD is used after the DCT to increase noise resistance and enable dimensionality reduction. The algorithms in (Leng et al., 2010; Zhang et al., 2010) are applied to improve the DCT domain's ability for discrimination.

Non-block-based CMFD algorithms, on the other hand, use features that are taken from the entire image (Elaskily et al., 2017). Using invariant keypoint approaches, local features like edges are retrieved from the image. The most often used keypoint descriptors for CMFD that have shown good accuracy are the Speeded Up Robust Features (SURF) and the Scale Invariant Feature Transform (SIFT) (Sadeghi et al., 2017), but the detection of small-size tampered regions is a problem for this early method. To find the candidate matches, it is used a Euclidean distance. The removal and injection of SIFT keypoint problem was solved using a CMFD approach in (Costanzo et al., 2014). With SIFT keypoints deleted, three new detectors—SVM detector, keypoints-to-corner ratio detector and chi-square distance detector are employed to identify forged images. The findings demonstrated that the detectors were effective against forging of copy-moves by hiding both fake keypoint injection and keypoint removal.

The authors in (hansda et al., 2022) used hybrid method to identify Copy-Move images. This method built on some steps, first, The Spatio-structured SIFT (S-SIFT) technique is used to find matched key-points. Then, use a two nearest neighbor matching approach to find the final forged areas.

Recently, other fields have started using deep learning algorithms. One of these fields is the detection of forgery images. A popular artificial neural network for classifying and recognizing images and objects is the CNN. This fields used deep features extracted from an image or blocks of images using CNN=based architecture has been more profound than the existing techniques. A set of feature maps is generated at each stage. Several authors have suggested using deep learning

to detect image forgeries. Wang et al. in (2015), a technique based on CNN is presented for automatically building hierarchical representations of color images, in addition to image splicing, this algorithm is created to detect copy-move forgeries. 30 high-pass filters from a Spatial Rich Model (SRM) are utilized as the first layer's initialization instead of the random filter trials used by default in CNNs (Rao and Ni 2016). Ouyang et al. in (2017), a different approach based on CNN is suggested for identifying copy-move forgeries. A pre-trained model used by this method, which is derived from a large database as ImageNet, then the network's structure is adjusted slightly by using a small number of training examples of copy-move forgeries, when using computer-generated tampered images, this method performed well, but when used in an real situation in CMF, the performance is extremely bad.

Muzaffer and Ulutas in (2019), After the features are retrieved using a CNN architecture built on the AlexNet model, a feature matrix is produced, and feature matching is completed, They calculated distance between two vectors, and it is compared by a predefined threshold after the feature matrix has been lexicographically sorted. This method provided low computational time.

Rao and Ni 2016 suggested CNN model for image splicing and copy-move detection. To be able to determine the effects caused by tampering operations, the first convolution layer of the CNN is employed for preprocessing. On labelled path samples taken from the training images, the CNN was trained. Following that, test images were used to apply this pre-trained CNN, and an SVM classifier was used to identify any tampered convolutional layers. CMF is considered the most common type of image forgery and this method is easy to employ and difficult to detect. Because it has a small effect on the image, the complexity of detecting it is limited. The convolutional neural network-based method for detecting forgery images is introduced in this paper.

3. Proposed Method

Three fundamental steps make up the proposed method: feature learning (Mask extraction and patch sampling), feature extraction, and classification. The CNN model is first trained using the labelled patch samples from the training images. The boundaries of the cloned patches act as the boundaries of the positive patch samples in forged images, while the boundaries of the negative patch samples are randomly selected from the original images, CNN might then concentrate on local artefacts and learn a hierarchical representation for the forgery image as a result of tampering operations. The entire image is scanned in the second step using a patch-sized sliding window, and then patch-based features are extracted for the image using a pre-trained CNN. Through feature fusion, the patch-based features are combined to create the discriminating feature for an image, which is then used to train either an SVM or a KNN to identify image forgery. Figure 1 shows the proposed system architecture. In Algorithm 1, the forgery detection is described. The steps of the

algorithm are: First, use image patches that closely resemble the distribution of the images the network will use to train the CNN. Both the modified and unmodified parts of the related image are present in the training patches. Then, after the network's last convolutional layer, feature fusion is used to extract the features from the images. Finally, use SVM and KNN classifier on the extracted features for the final classification, the complexity of the main algorithm is O (n^2).

Algorithm 1 Forgery Detection Algorithm

Input: Forged and Original Images I (R, G, B)
Output: Detected result whether the image is forged or not

```
procedure
  For Each original image component (R, G, B) do
      Convert image to grayscale image
      For each forgery image
          Get forgery image using original's name
          Convert forgery image to grayscale image
      Get the difference of the two-grayscale (original and forgery) image
      Generate mask by make background black and tampered area white
          Save mask
      End for
  End for
  For Each forgery image
      For Each mask
          If mask equal to forgery
              Extract Number of patches determined by the size of the image's forged area
              Save patches
          End if
      End for
  End for
  For Each original image
          Generate random patches
          Save patches
  End for
  Train CNN model using patches
  Extract Features by CNN model
  Apply SVM/KNN to classify (Forged or not)
end procedure
```

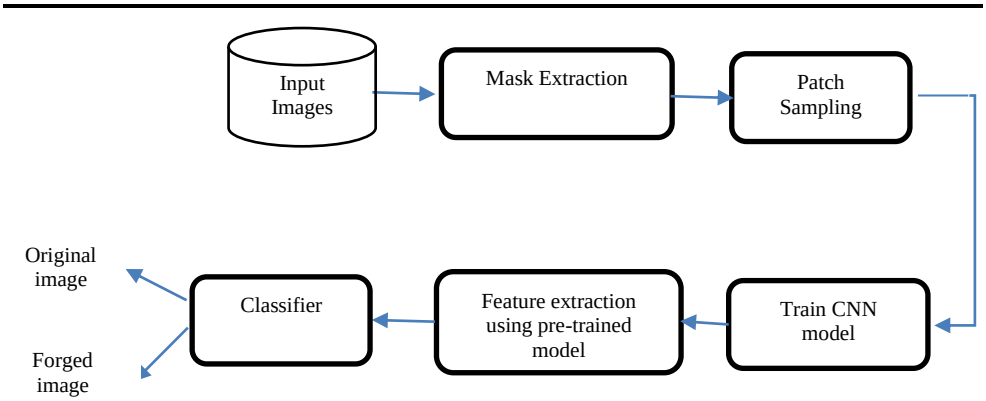

Fig. 1: Proposed system architecture

3.1. Mask Extraction

Masking means, defined a small pieced of the image as forgery part. Firstly, we convert images to grayscale, then, we are able to tell the difference between authentic and forgery images. Finally, the background is made black and the forgery area white. This process is applied to CASIA v1 and CASIA v2 datasets. The extraction of the mask is shown in Fig. 2.

(a) Original image (b) Forged image (c) Mask image

Fig. 2. Mask extraction process example

3.2. Patch Sampling

To train a discriminating CNN Model, large and representative data samples are typically required. So that, to prepare the positive samples (forgery), we generate the patch for colored images. The number of positive patches generated from a forgery image, which is determined by the size of the image's forgery area. The same number of patches is generated at random from the original images in the training image set for negative samples, however. The CNN's training data set is made up of the sampled patches. Some label preserving transformations are used in CNN training to avoid overfitting and improve generalization capacity.

3.3. Feature Extraction (Train the CNN Model)

The input and output of a CNN consist of several convolutional layers, which are represented as feature maps. Convolution, non-linear activation, and pooling are the three steps that each convolutional layer in CNN goes through to produce feature maps. A softmax classifier follows certain fully connected layers at the end of CNN.

Fig. 3 presents the proposed CNN model's architectural layout. The proposed CNN's eight convolutional layers, two pooling layers, and fully linked layer are presented. The CNN's input volume is made up of patches with a size of 128×128×3. While other layers have 16 kernels with size of 3×3, the first and second convolutional layers each have 30 kernels with size of 5×5. Neurons are activated by Rectified Linear Units (ReLU), causing them to selectively react to useful input signals. Following the first and fifth convolutional layers, the input is spatially resized using a non-overlapping max-pooling with size 2×2 filter. This is done because the max-pooling operation improves in the retention of additional texture data and enhances convergence performance. Finally, the 400 features are extracted and passed to the fully connected layer with zero "dropout" of neurons. Unlike other typical CNN architectures, which use two or more fully connected layers, our network's end uses only one fully connected layer. This is because overfitting is a common problem with fully connected layers, especially with small training set, as it is in our task, and the fully connected layer often has too many parameters to train.

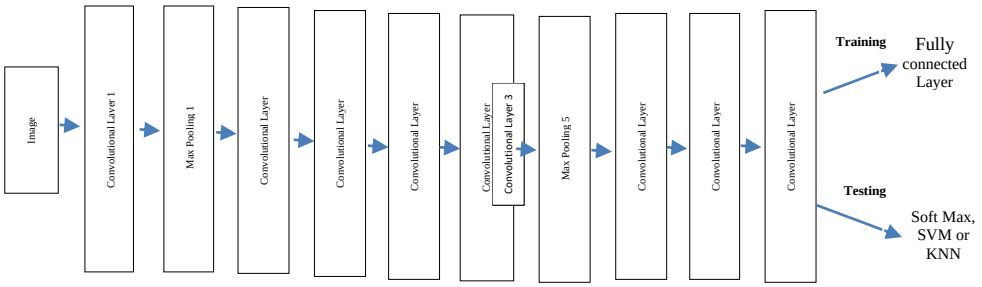

Fig. 3: The architecture of the 10-layer CNN

3.4. Image Classification

After the fully connected layer, we classified the images using three distinct classifiers; the SoftMax classifier in the CNN model, KNN and SVM classifier. In the classification process we used the 400 features in the fully connected layer which generated from the CNN model.

4. Experimental Results and Discussion

The outcomes produced when employing the proposed method are evaluated in this section. The experiments were executed on a Laptop have Windows 10 with Core i7, 2.6 GHz, a Graphics Card 1650Ti, 8GB DDR3 RAM and implemented in Spyder 3.3.6 (python 3.7).

4.1. Datasets

On three open benchmark datasets for forgery detection, all of our experiments are performed, i.e., CASIA v1.0, CASIA v2.0 (Casia Tide) and Columbia (Ng et al., 2004). 12,614 color images are included in the CASIA v2.0 dataset of size 384×256/256×384 with 7,491 original and 5,123 forgeries. The Columbia dataset contain 1,845 grayscale image blocks of size 128×128 has 933 original and 912 forgeries. The 1,721 color images in the CASIA v1.0 dataset, with size, 84×256 include 800 authentic copies and 921 forgeries. Fig. 4 displays a few examples from datasets.

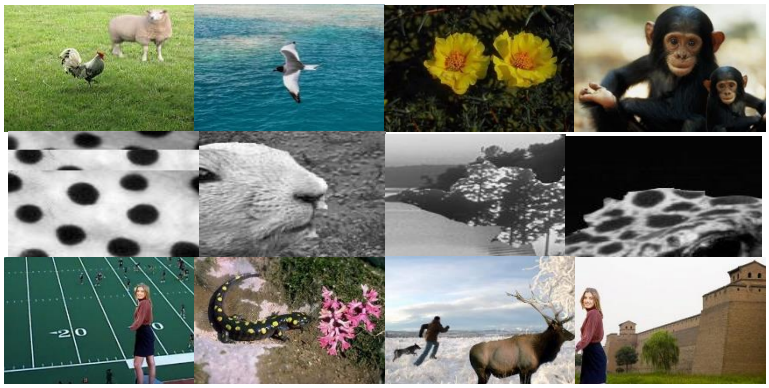

Fig. 4: Some examples of image forgery image dataset. Forged versions of images from CASIA v2.0 dataset in the first row. Forged versions of images from Columbia dataset in the second row. Forged versions of images from CASIA v1.0 dataset in the third row.

4.2. Evaluation Metrics

Based on detection accuracy, the installed CNN's effectiveness is evaluated. The accuracy is determined as follows:

$$Accuracy = \frac{No.of\ correctly\ detected\ images}{Total\ No.of\ images} \times 100 = \frac{(T_N + T_P)}{(T_P + F_p + T_N + F_N)} \times 100 \quad (1)$$

Where the number of forgery images that are really detected as forgeries is known as True Positive (T_P). The number of true images that are incorrectly identified as forgeries is known as the False Positive rate (F_p). The number of forgery images that are incorrectly identified as authentic images is known as the False Negative (F_N). The number of authentic images that are actually identified as being authentic images is known as True Negative (T_N).

The performances are evaluated using the precision and recall measures. The precision is the ratio of the authentic forgery images/pixels to all other known images/pixels. The formula for precision is as follows.

$$Precision = \frac{T_p}{(T_p + F_p)} \tag{2}$$

Recall shows the percentage of all forgeries images/pixels that were properly identified. The formula for recall is as follows.

$$Recall = \frac{T_p}{(T_p + F_N)} \tag{3}$$

A complete criterion is the F1 score, which combines precision and recall. F1 is calculated as follows.

$$F1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{4}$$

The closer that F1, precision and recall are to 1, this means the better of performance.

4.3. Performance of the Proposed Method

Within this section, the evaluation outcomes for the proposed deep CMFD algorithm are provided in detail. The CASIA v1 and v2 datasets as well as the Columbia dataset have all been used with the suggested method. Train CNN using image patches that closely resemble the distribution of the images the network will use. Both the modified and unmodified parts of the related image are present in the training patches.

First, we trained the CNN by using the CASIA v1 and v2 datasets with a batch size of 128 images and Columbia dataset with a batch size of 32 images.

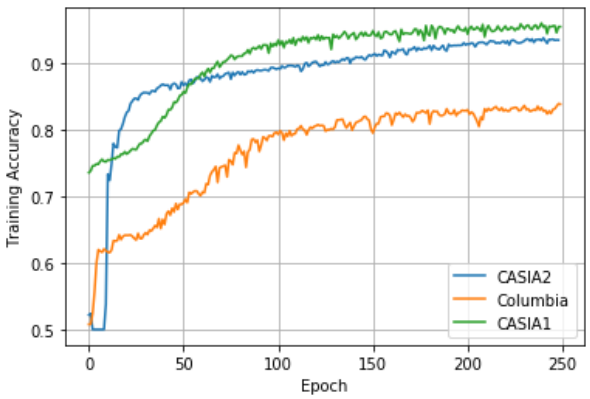

Fig. 5: Accuracy of the CNN for CASIA v1.0, CASIA v2.0 and Columbia dataset with 250 epochs

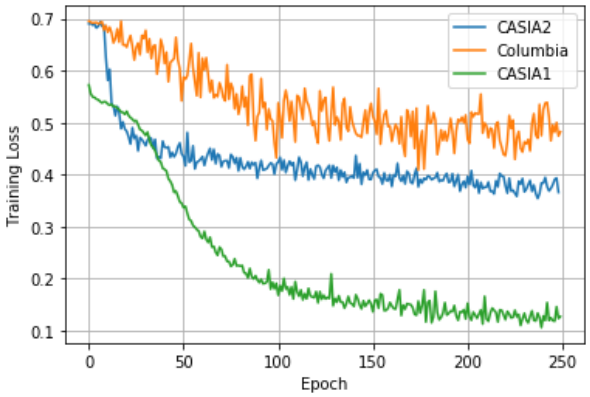

Fig. 6: Loss of the CNN for CASIA v1.0, CASIA v2.0 and Columbia dataset with 250 epochs

Fig. 5 and Fig.6 illustrate the training with 250 epochs for the CASIA v1, CASIA v2, and Columbia datasets, respectively.

Results explain that the test accuracy increases to 91% with 250 epochs for the CASIA v1 and v2 dataset and reach 80% for the Columbia dataset with 250 epochs, while the loss recedes and settles at 35% with 250 epochs for the CASIA v1 and v2 dataset and reach 50% with 250 epochs for Columbia dataset.

After being trained, the CNN employs the pre-trained model to determine feature representation after the final convolution layer, then, features enter to SVM or KNN classifier.

Table 1: Confusion matrix - CASIA v2.0 using SVM

CASIA v2.0	Predicted Auth.	Predicted Tamp.
Actual Auth.	1426	72
Actual Tamp.	17	1008

Table 2: Confusion matrix - CASIA v1.0 using SVM

CASIA v1.0	Predicted Auth.	Predicted Tamp.
Actual Auth.	3784	85
Actual Tamp.	39	1283

Table 3: Confusion matrix - Columbia using SVM

Columbia	Predicted Auth.	Predicted Tamp.
Actual Auth.	161	26
Actual Tamp.	30	152

Table 4: Confusion matrix - CASIA v2.0 using KNN

CASIA v2.0	Predicted Auth.	Predicted Tamp.
Actual Auth.	1447	62
Actual Tamp.	13	1001

Table 5: Confusion matrix - CASIA v1.0 using KNN

CASIA v1.0	Predicted Auth.	Predicted Tamp.
Actual Auth.	3786	49
Actual Tamp.	43	1313

Table 6: Confusion matrix - Columbia using KNN

Columbia	Predicted Auth.	Predicted Tamp.
Actual Auth.	171	28
Actual Tamp.	27	143

Table 1 and table 4 show the confusion matrix of CASIA v2.0 dataset when applying SVM or KNN. The best SVM hyperparameters that we trained on were Regularization parameter (C) = 100 and gamma parameter (γ) = 0.001. The previous settings resulted in a classification accuracy of F1 score =96.47 %. The best KNN hyperparameters that we trained on were n_neighbors= 15 and weights= uniform. The previous settings resulted in a classification accuracy of F1 score=97.02%.

Table 3 and table 6 show the confusion matrix of Columbia dataset using SVM or KNN. The optimal SVM parameters chosen after the grid search were C = 1 and γ= 0.01, resulting in an accuracy of F1 score= 84.82%. The best KNN hyperparameters that we trained on were n_neighbors= 15 and weights= uniform. The previous settings resulted in a classification accuracy of F1 score=85.1%.

Table 7: Comparison Accuracy of the proposed method with other methods on different datasets

		Dataset	
Method	CASIA v1.0	CASIA v2.0	Columbia
CNN	95.55%	93.54%	83.90%
Yuan [20]	97.61%	96.47%	84.82%
Proposed	98.22%	97.02%	85.1%

Table 2 and table 5 show the confusion matrix of CASIA v1.0 using SVM or KNN. The optimal SVM parameters chosen after the grid search were C = 100 and γ=0.0001, resulting in an accuracy of F1 score= 97.61%. The best KNN hyperparameters that we trained on were n_neighbors= 17 and weights= distance. The previous settings resulted in a classification accuracy of F1 score=98.22%.

Table 7 provides an overview of the comparison's results. The results show that when compared to CNN and SVM, the proposed deep-learning-based CMFD method performs well.

5. Conclusion

Through this paper, we have emphasized the importance of determining how to determine if an image is forged or original. The method we proposed for detecting

image forgeries is based on CNN. Our method trained using labelled patch samples that are drawn without the forged boundaries in forged images. The test images are then processed by the pre-trained CNN to extract dense features, which are then input into the KNN classification algorithm. The proposed CNN-KNN algorithm improves previous image forgery detection methods, according to extensive tests on three publicly available datasets.

References

Boz, A. & Bilge, H. Ş. (2016). Copy-move image forgery detection based on LBP and DCT. *24th Signal Processing and Communication Application Conference (SIU)*, 16–19

CASIA image tampering detection evaluation database (CASIA TIDE) v1.0 and v2.0, http://forensics.idealtest.org

Costanzo, A., Amerini, I., Caldelli, R., & Barni, M. (2014). Forensic analysis of SIFT Keypoint removal and injection. *IEEE Trans Inf For Secur*, 9(9), 1450–1464

Derroll, D., Divya, B. (2015). Image authentication techniques and advances survey, COMPUSOFT. *Int J Adv Comput Technol*, 4(4)

Elaskily, M. A. Aslan, H. K. Abd El-Samie, F. E. Elshakankiry, O. A. Faragallah, O. S., & Dessouky, M. M. (2017). Comparative study of copy-move forgery detection techniques. *Intl Conf on Advanced Control Circuits Systems (ACCS) Systems & Intl Conf on New Paradigms in Electronics & Information Technology (PEIT)*, Alexandria, Egypt

Farid, H. (2009). Image forgery detection a survey. *IEEE Signal Process Mag*, 26(2):16–25

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., Chen, T. (2018). Recent advances in convolutional neural networks. *Pattern Recogn*, 77, 354–377

Hansda, R., Nayak, R., & Balabantaray, B. K. (2022). Copy-move image forgery detection using phase adaptive spatio-structured SIFT algorithm. *SN COMPUT. SCI*. 3, 46. DOI:10.1007/s42979-021-00903-2

Kang, X., Lin, G., Chen, Y., Zhang, E., & Duan, G. (2012). Detecting tampered regions in digital images using discrete cosine transform and singular value decomposition. *Int J Digit Content Technol Appl (JDCTA)*

Leng, L., Zhang, J., Khan, MK., Chen, X., & Alghathbar, K. (2010). Dynamic weighted discrimination power analysis: A novel approach for face and palm print recognition in DCT domain. *Int J Phys Sci*, 5(17), 2543– 2554

Leng, L., Zhang, J., Xu, J., Khan, MK., & Alghathbar, K. (2010). Dynamic weighted discrimination power analysis in DCT domain for face and Palm print recognition. *International conference on information and communication technology convergence (ICTC)*. DOI:10.1109/ictc.2010.5674791

Muzaffer, G. & Ulutas, G. (2019). A new deep learning-based method to detection of copy-move forgery in digital images. In *2019 scientific meeting on electrical-electronics & biomedical engineering and computer science*. EBBT, 1–4. DOI:10.1109/EBBT.2019.8741657

Nanda, W., Diane, N., Xingming, S., (2014). Moise, F. K.: Survey of partition-based techniques for copy-move forgery detection. T*he scientific world journal*, 2014:Article ID 975456

Ng, T. -T., Chang, S. -F., & Sun, Q. (2004). A data set of authentic and spliced image blocks. Columbia University, ADVENT Technical Report, 203

Ouyang, J., Liu, Y., & Liao, M. (2017). Copy-move forgery detection based on DeepLearning. 10th international congress on image and signal processing. *BioMedical engineering and informatics (CISP-BMEI)*. DOI:10.1109/cisp-bmei.2017.8301940

Rao, Y. &Ni, J. (2016). A deep learning approach to detection of splicing and copy-move forgeries in images. *IEEE international workshop on information forensics and security (WIFS)*

Sadeghi, S., Dadkhah, S., Jalab, H., Mazzola, G., Uliyan, D. (2017). State of the art in passive digital image forgery detection: copy-move image forgery. *Pattern Anal Applic*, 21(2), 291–306

Vartak, R. & Deshmukh, S. (2014). Survey of digital image authentication techniques. *Int J Res Advent Technol*, 2(7)

Warif, N. B. A., Wahab, A. W. A., Idris, M. Y. I., Ramli, R., Salleh, R., Shamshirband, S., & Choo, K. –K. R. (2016). Copy-move forgery detection: survey, challenges and future directions. *J Netw Comput Appl*, 75:259–278

Warif, N. B. A., Wahab, A. W. A., Idris, M. Y. I., Ramli, R., Salleh, R., Shamshirband, S., Choo, K. –K. R. (2016). Copy-move forgery detection: Survey, challenges and future directions. *J Netw Comput Appl*, 75:259–278

Wang, P., Wei, Z., & Xiao, L. (2015). Pure spatial rich model features for digital image steganalysis. *Multimedia Tool Appl*, 75(5), 2879–2912

Yuan, R. & Jiangqun, N. (2016). A deep learning approach to detection of splicing and copy-move forgeries in images. *IEEE International Workshop on Information Forensics and Security (WIFS)*