Advanced Image Tampering Detection using Lightweight Deep Learning Models for Forensic Authentication

P. Prashanthi[1], B. Shanmukha[2], M. Deshwanth[2], M. Arun[2], M. Nikhil[2]

[1]Assistant Professor, [2]UG Scholar, [1,2]Department of Computer Science and Engineering
[1,2]Malla Reddy Engineering College and Management Sciences, Medchal, 501401, Hyderabad

ABSTRACT

In the current era of digital technology, images and videos are being used more and more as powerful pieces of evidence in several fields, including judicial processes, insurance fraud investigations, and social networking platforms. The intrinsic adaptability of altering tools specifically created for digital photos, particularly when there is no apparent indication of tampering, gives rise to issues regarding their legitimacy. The primary duty of an image forensics authority is to develop technological innovations capable of detecting instances of image manipulation and deception. So far, researchers have examined three primary groups of approaches for detecting modification or forgery: those that depend on descriptors of features, those that depend on inconsistent shadows, and those that depend on double JPEG compression. Image forgery detection poses a substantial challenge in many real-time applications, social media, and online information platforms. Conventional detection approaches that depend on recognizing indications of visual alterations are limited by predetermined assumptions such manually designed characteristics, dimensions, and brightness. This study introduces a decision-making approach for identifying image tampering through the utilization of fusion methodologies. The decision fusion is dependent on lightweight deep learning models, including SqueezeNet, MobileNetV2, and ShuffleNet. The fusion decision system is implemented in a two-phase approach. At first, the current weights of the effective deep learning models are used to evaluate the genuineness of the photographs. Moreover, the recalibrated weights are utilized to assess the results of image counterfeiting in relation to the current models. The experimental results demonstrate that the fusion-based decision technique surpasses the present state-of-the-art approaches in terms of accuracy.

Keywords: Image fusion, Lightweight models, Fusion, Support vector machine, Deep learning.

1. INTRODUCTION

In this digital era, images and videos are being used as influential sources of evidence in a variety of contexts like evidence during trials, insurance fraud, social networking, etc. The easy adaptability of editing tools for digital images, especially without any visual proof of manipulation, give rise to questions about their authenticity. It is the job of image forensics authorities to develop technological innovations that would detect the forgeries of images. There are three primary classes of manipulation or forgery detectors studies until now: those supported features descriptors, those supported inconsistent shadows and eventually those supported double JPEG compression.

With sophisticated software, it is easy to tamper the contents of the image to influence the opinions of others. Image forgery techniques are broadly classified into two categories namely copy-move and splicing. For copy-move forgery, elements of the image content area are traced and smudge inside a similar image, whereas for splicing forgery, parts of the image content smudge from alternative pictures. To reconstruct the trust in pictures, various image forgery detection techniques have been proposed over the past few years. Many previous studies have tried to extract totally different properties from the image to spot the copy-paste or splicing of forged areas, such as the lighting, shadows, sensing element noise, and camera reflections.

Researchers determined the credibility of the image wherever it is known either as authentic or forged. Currently, there are many techniques to spot forged regions that exploits the artefacts left by multiple JPEG compression and other techniques of image manipulation to sight the forged regions. Camera primarily based ways have additionally analyzed where the detection relies on demosaicing regularity or sensing element pattern noise wherever the irregularities of the sensing element pattern area unit extracted and compared for anomalies. Forged or manipulated pictures can mislead people and may threaten individuals' life. This paper aims to find the manipulated pictures by automating the method of feature extraction instead of feature engineering or feature extraction through the manual process. Deep learning to make use of highly correlated pixels in a vicinity, thus considering grouped native connections.

The motivation to use lightweight models in favour to prevent overfitting of the convolutional neural network (CNN) architectures and can be easily deployed on resource constrained hardware and can learn enriched representations. ShuffleNet makes more feature map channels for a given computation complexity budget, which helps to encode more information and is especially important to the efficiency of small networks. MobileNet, makes use of deep-separable convolutions and gains state-of-the-art results and demonstrated the effectiveness of MobileNet when applied to a broad range of tasks. SqueezeNe, optimizing the architecture for fast processing speed CNN system with 50×, fewer parameters than AlexNet and retains standard accuracy. The lightweight models can be deployed effectively on resource-restricted hardware and can learn enriched representation.

2. LITERATURE SURVEY

Amerini et al. proposed a step forward in this direction by analyzing how a single or double JPEG compression can be revealed and localized using convolutional neural networks (CNNs). Different kinds of input to the CNN have been taken into consideration, and various experiments have been carried out trying also to evidence potential issues to be further investigated.

Xiao et al. proposed a splicing forgery detection method with two parts: a coarse-to-refined convolutional neural network (C2RNet) and diluted adaptive clustering. The proposed C2RNet cascades a coarse convolutional neural network (C-CNN) and a refined CNN (R-CNN) and extracts the differences in the image properties between un-tampered and tampered regions from image patches with different scales. Further, to decrease the computational complexity, an image-level CNN is introduced to replace patch-level CNN in C2RNet. The proposed detection method learns the differences of various image properties to guarantee a stable detection performance, and the image-level CNN tremendously decreases its computational time.

Zhang et al. studied the first stage; this paper utilized a Stacked Autoencoder model to learn the complex feature for each individual patch. For the second stage, this paper integrated the contextual information of each patch so that the detection can be conducted more accurately.

Goh et al. proposed a hybrid evolutionary framework to perform a quantitative study to evaluate all features in image tampering for the best feature set. Upon feature evaluation and selection, the classification mechanism must be optimised for good performance. Therefore, in addition to being able to determine an optimal set of features for a classifier, the hybrid framework can determine the optimal multiple classifier ensembles while achieving the best classification performance in terms of low complexity and high accuracy for image tampering detection.

Sutthiwan et al. proposed image statistical features are generated by applying Markovian rake transform to image luminance component. Markovian rake transform is the application of Markov process to difference arrays which are derived from the quantized block discrete cosine transform 2-D arrays with multiple block sizes. The efficacy of thus generated features has been confirmed over a

recently established large-scale image dataset designed for tampering detection, with which some relevant issues have been addressed and corresponding adjustment measures have been taken. The initial tests by using thus generated classifiers on some real-life forged images available in the Internet show signs of promise of the proposed features as well as the challenge encountered by the research community of image tampering detection.

He et al. proprosed a Markov based approach to detect this specific artifact. Firstly, the original Markov features generated from the transition probability matrices in DCT domain by Shi et al. is expanded to capture not only the intra-block but also the inter-block correlation between block DCT coefficients. Then, more features are constructed in DWT domain to characterize the three kinds of dependency among wavelet coefficients across positions, scales, and orientations. After that, feature selection method SVM-RFE is used to fulfill the task of feature reduction, making the computational cost more manageable. Finally, support vector machine (SVM) is exploited to classify the authentic and spliced images using the final dimensionality-reduced feature vector.

Change et al. proposed a novel forgery detection algorithm to recognize tampered inpainting images, which is one of the effective approaches for image manipulation. The proposed algorithm contains two major processes: suspicious region detection and forged region identification. Suspicious region detection searches the similarity blocks in an image to find the suspicious regions and uses a similarity vector field to remove the false positives caused by uniform area. Forged region identification applies a new method, multi-region relation (MRR), to identify the forged regions from the suspicious regions. The proposed approach can effectively recognize if an image is a forged one and identify the forged regions, even for the images containing the uniform background. Moreover, this paper proposed a two-stage searching algorithm based on weight transformation to speed up the computation speed.

Rhee et al. presented a short feature vector that is made up of three types of feature sets. The first set is defined by the variation to be the 3-D length in the gradient difference of the intensity values of the adjacent row and column line pairs in the image, respectively. The second set is defined by the variation in the coefficient difference of the Fourier transform to be the 3-D length in the adjacent line pairs. The last set is defined by the residual image between an image and its reconstructed image by the gradient based on solving Poisson's equation, which is also the 3-D length. Two of the sets are extracted in the spatial and spectral domains of an image, respectively, and the last set is extracted from the residual image. The totally formed 9-D feature vector is subsequently trained in the support vector machine classifier for MFD.

Lamba et al. proposed a discrete fractional wavelet transform-based scheme for identification of duplicated regions in the image. The test image is split into overlapping image blocks with fixed dimensions. Then, on each image block, discrete fractional wavelet transform is employed for the extraction of their features. All the feature vectors are systematized in lexicographical manner followed by the block matching and block filtering steps to obtain the replicated blocks, if any. The proposed method can detect single and multiple duplicated regions successfully.

Lin et al. proposed detecting tampered images by examining the double quantization effect hidden among the discrete cosine transform (DCT) coefficients. This paper is the only one to date that can automatically locate the tampered region, while it has several additional advantages: fine-grained detection at the scale of DCT blocks, insensitivity to different kinds of forgery methods (such as alpha matting and inpainting, in addition to simple image cut/paste), the ability to work without fully decompressing the JPEG images, and the fast speed. Experimental results on JPEG images are promising.

3. PROPOSED SYSTEM

The architecture of the proposed decision fusion is based on the lightweight deep learning models as shown in Figure 1. The lightweight deep learning models chosen are SqueezeNet, MobileNetV2, and ShuffleNet. The proposed system is implemented in two phases i.e., with pre-trained and fine-tuned deep learning models. In the pre-trained model's implementation, regularization is not applied, and the pre-trained weights are used and for the fine-tuned implementation, regularization is applied to detect image forgery. Each phase consists of three stages namely, data pre-processing, classification, and fusion. In the data pre-processing stage, the image in the query is pre-processed based on the dimensions required by the deep learning models. SVM is used for the classification of the image as forged or non-forged. Initially, we discuss the lightweight deep learning models and then the strategy used for the regularization is discussed in the further sections.

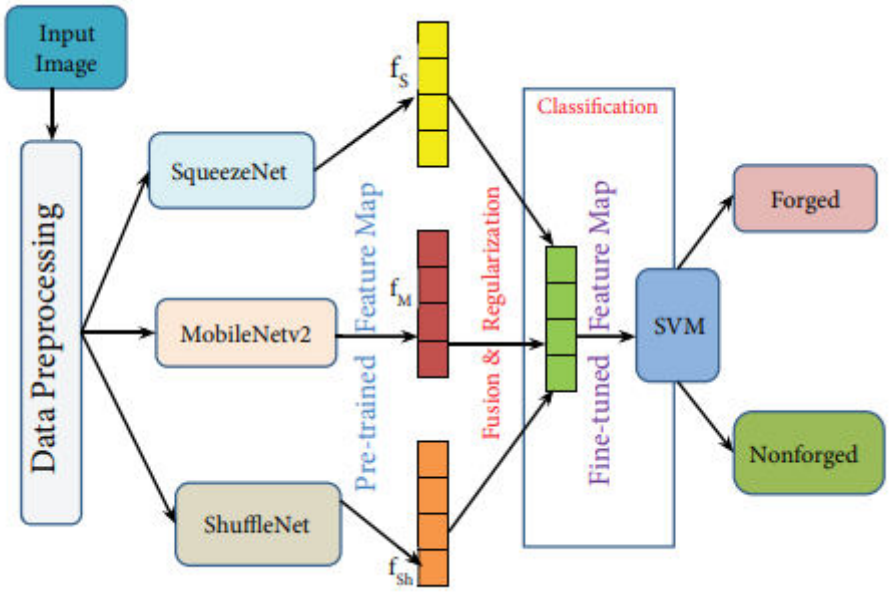

Figure 1: Fusion based decision model for forgery detection.

Data preprocessing

In this stage, the image in a query that needs to be identified whether it is forged or not is subjected to preprocessing. The height and width of the image required for SqueezeNet is 227×227. The height and width of the image required for MobileNetV2 is 224×224. The height and width of the image required for ShuffleNet is 224×224. The input image is pre-processed first based on the dimensions required for each of the models. Each model then takes the input image to produce feature vector in further stages.

ShuffleNet

It is a CNN that is also trained on the ImageNet dataset with 50 layers deep and can classify the images up to 1000 categories. Table 1. Parameters of lightweight deep learning models. (Depth represents the largest number of sequential convolutional or fully connected layers on a path from the input layer to the output layer, parameter represents the total number of learnable parameters in each layer and image input size represents the required input image size).

Table 1. Models description.

Models	Depth	Parameter (millions)	Image input size
SqueezeNet	18	1.24	227 × 227
MobileNetV2	53	3.5	224 × 224
ShuffleNet	50	1.4	224 × 224

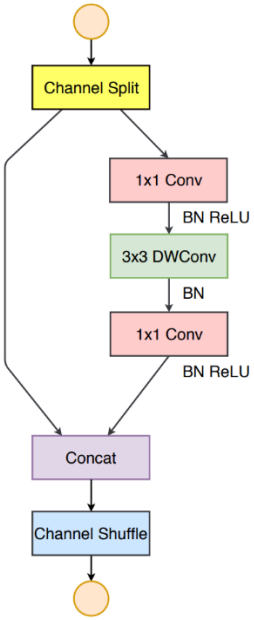

Figure 2: ShuffleNet.

Fusion model and regularization

The proposed system is first implemented with lightweight deep learning models using pretrained weights for the image forgery detection, afterward, the proposed system is implemented as a fusion of the decision of lightweight models as discussed in the previous section. Initially, the input image is passed to the lightweight models to obtain their feature maps respectively. The feature map from the SqueezeNet is denoted as f_s, the feature map from the MobileNetV2 is denoted as f_m , the feature map from the ShuffleNet is denoted as f_{sh} . For the fusion model, the pretrained lightweight deep learning model's output feature mapping f_p is used. This feature map f_p is a combination of the feature maps obtained from the lightweight models as shown in Equation (1).

$$f_p = f_s + f_m + f_{sh} \quad (1)$$

The fusion model uses feature map f_p as a local descriptor for an input patch to extract the features of the image. The image for the fusion model is represented as a function $Y_{fusion} = f_{(x)}$where x is the patch in the input image. For a test image size $m \times n$, a sliding window of size p×p is used to compute the local descriptor Y_{fusion} is computed as shown in the equation (2) where $Y_1, Y_2, , Y_3$ represents the descriptors of the patches of the image obtained from the deep learning models. It is obtained as a concatenation of all the input patches xi and the new image representation is given by equation (3) where s is the size of the stride used for transforming the input patch, this new image representation f_{fusion} is used as the feature map for the classification by the SVM as forged or nonforged.

$$Y_{fusion} = [Y_1 + Y_2 + \ldots + Y_T] \quad (2)$$

$$f_{fusion} = \frac{m-w}{s} + 1 * \frac{n-w}{s} + 1 \quad (3)$$

For fine tuning of the parameters of the fusion model, the initialization of the weight kernels is used as shown in Equation (4). In this equation Wf represents the weights of the fusion model, W_s represents the weights of the SqueezeNet model, W_m represents the weights of the MobileNetV2 model and W_{sh} represents the weights of the ShuffleNet model. The weight of the fusion model W_f is initialized as shown in Equation (5). The initialization of the weights acts as a regularization term and facilitates the fusion model to learn the robust features of detecting the forgery rather than the complex image representations.

$$W_f = [W_{sj}\ W_{mj}\ W_{shj}]\ \text{j} = 1, 2, 3 \quad (4)$$

$$W_f = [{W_f}^{4k-2}\ {W_m}^{4k-2}\ {W_{sh}}^{4k}] \text{ where } k = [[j + 1] mod 11] + 1 \quad (5)$$

4. RESULTS AND DISCUSSION

Dataset

The dataset used for the experiment is benchmark publicly available MICC-F220 of 110 non-forged images and 110 forged images with 3 channels i.e., color images of size 722 × 480 to 800 × 600 pixels. As shown in Figure 3, Figures 3a–3j are forged images with 10 different combinations of geometrical and transformations attacks and Figure 3k is the non-forged image. From the dataset 154 images are chosen randomly for training purposes and remaining for testing purpose.

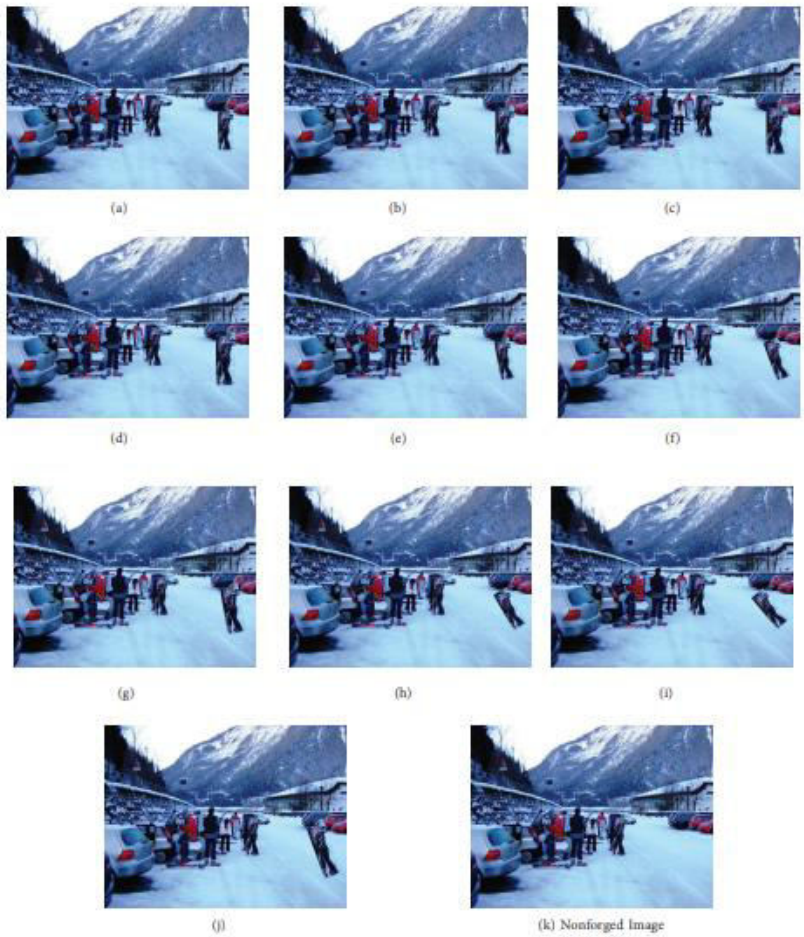

Figure 3: Dataset with 10 different combinations of geometrical and transformation attacks; (a–j), forged; (k), non-forged images.

Baseline modules

The baseline models that are used for the comparison of the fusion model are summarized as follows.

1) Upload MICC-F220 Dataset: using this module we will upload dataset to application
2) Pre-process Dataset: using this module we will read all images and then normalize their pixel values and then resize them to equal size
3) Generate & Load Fusion Model: using this module we will train 3 algorithms called SqueezeNet, MobileNetV2 and ShuffleNet and then extract features from it to train fusion model. All algorithms prediction accuracy will be calculated on test data
4) Fine Tuned Features Map with SVM: using this module we will extract features from all 3 algorithms to form a fusion model and then fusion data get trained with SVM and then calculate its prediction accuracy.
5) Run Baseline SIFT Model: using this module we will extract SIFT existing technique features from images and then train with SVM and get its prediction accuracy
6) Accuracy Comparison Graph: using this module we will plot accuracy graph of all algorithms
7) Performance Table: using this module we will display all algorithms performance table.

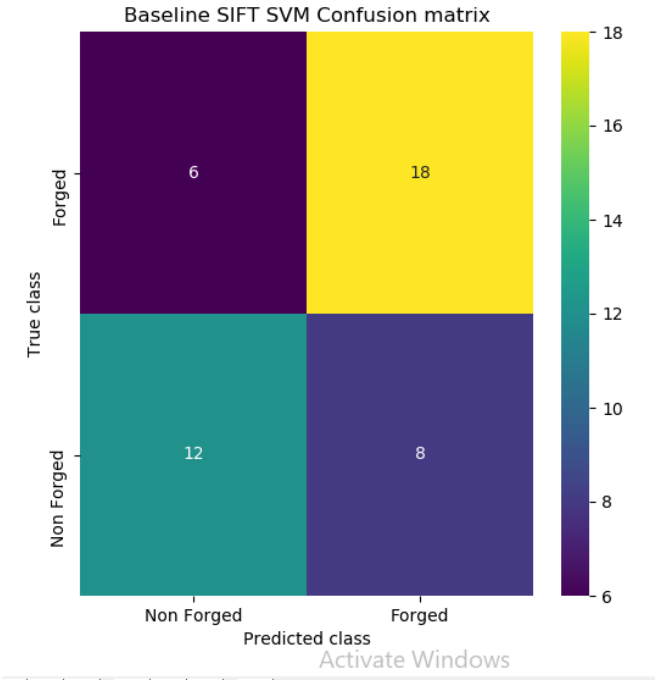

Figure 4: Confusion matrixes of fusion model and baseline SIFT SVM.

Table 2: Performance comparision.

Method	Accuracy	Precision	Recall	FSCORE
Existing SIFT SVM	68.1	67.9	67.5	67.5
Only SqueezeNet	79.5	81.1	79.5	79.2
Only ShuffleNet	56.8	62.7	56.8	51.1
Only MobileNetV2	81.8	82.9	81.8	81.6

Proposed Fusion Model SVM	95.4	95	96.1	95.3

5. CONCLUSION

Image forgery detection helps to differentiate between the original and the manipulated or fake images. In this work, a decision fusion of lightweight deep learning-based models is implemented for image forgery detection. The idea was to use the lightweight deep learning models namely SqueezeNet, MobileNetV2, and ShuffleNet and then combine all these models to obtain the decision on the forgery of the image. Regularization of the weights of the pretrained models is implemented to arrive at a decision of the forgery. The experiments carried out indicate that the fusion-based approach gives more accuracy than the state-of-the-art approaches. In the future, the fusion decision can be improved with other weight initialization strategies for image forgery detection.

REFERENCES

[1] Amerini, T. Uricchio, L. Ballan, and R. Caldelli, "Localization of JPEG Double Compression Through Multi-Domain Convolutional Neural Networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017, pp. 1865-1871, doi: 10.1109/CVPRW.2017.233.

[2] B Xiao, Y Wei, X Bi, W Li, J Ma. “Image splicing forgery detection combining coarse to refined convolutional neural network and adaptive clustering”, Information Sciences, Volume 511, Pages 172-191, 2020, ISSN 0020-0255, https://doi.org/10.1016/j.ins.2019.09.038.

[3] Zhang Y, Goh J, Win LL, Thing VL. Image region forgery detection: a deep learning approach. SG-CRC 2016; 2016: 1-11.

[4] Goh J, Thing VL. A hybrid evolutionary algorithm for feature and ensemble selection in image tampering detection. International Journal of Electronic Security and Digital Forensics 2015; 7 (1): 76-104

[5] Sutthiwan P, Shi YQ, Zhao H, Ng TT, Su W. Markovian rake transform for digital image tampering detection. In: Shi YQ, Emmanuel S, Kankanhalli MS, Chang S-F, Radhakrishnan R (editors). Transactions on Data Hiding and Multimedia Security VI. Lecture Notes in Computer Science, Vol. 6730. Berlin, Germany: Springer; 2011, pp. 1-17

[6] He Z, Lu W, Sun W, Huang J. Digital image splicing detection based on Markov features in DCT and DWT domain. Pattern Recognition 2012; 45 (12): 4292-4299.

[7] Chang IC, Yu JC, Chang CC. A forgery detection algorithm for exemplar-based inpainting images using multi-region relation. Image and Vision Computing 2013; 31 (1): 57-71.

[8] Rhee KH. Median filtering detection based on variations and residuals in image forensics. Turkish Journal of Electrical Engineering & Computer Science 2017; 25 (5): 3811-3826.

[9] Lamba AK, Jindal N, Sharma S. Digital image copy-move forgery detection based on discrete fractional wavelet transform. Turkish Journal of Electrical Engineering & Computer Science 2018; 26 (3): 1261-1277.

[10] Lin Z, He J, Tang X, Tang CK. Fast, automatic, and fine-grained tampered JPEG image detection via DCT coefficient analysis. Pattern Recognition 2009; 42 (11): 2492-2501.