CNN based Image Forgery Detection using pre-trained AlexNet Model

Amit Doegar[a], *Maitreyee Dutta*[a], *Gaurav Kumar*[b]

aDepartment of Computer Science and Engineering, NITTTR, Chandigarh,160019, India

bMagma Research and Consultancy Services, Ambala, 133006, India

Abstract:

Image forgery detection is an approach for detection and localization of forged component from a manipulated image. To find manipulation or tampering in the original image, an adequate number of features are required to classify the given image is either a forged or non-forged. To achieve this convolutional neural network (CNN) based pre-trained AlexNet model's deep features have been utilized which are efficient and effective, as compared to the existing state-of-the-art approaches on publicly available benchmark dataset MICC-F220. The experiment result shows that the proposed approach using a pre-trained AlexNet model based deep features with Support Vector Machine (SVM) classifier has achieved 93.94% accuracy.

Keywords- Image Forgery, Deep Learning, SVM, CNN, AlexNet Model

1. Introduction

Image forgery detection approach detects whether the image is manipulated or not. An adequate number of features are required to detect the given image is forged or not. Features based on convolutional Neural Network (CNN) models are effective features to classify the category of the image, due to the fact that existing approaches for feature extraction are based on handcraft features or feature engineering and are not invariant to various types of transformations, geometrical and post-processing operations. Moreover, feature engineering and feature extraction are key and time-consuming (https://ai.intel.com). CNN's today, as multiple layers of neurons for processing more complex features are deeper layers of the network. The greatest advantage of CNN and deep learning is that they can learn appropriate features by themselves automatically whereas designing features manually or through feature engineering is extremely hard. Nowadays CNN and deep learning based applications widely being used in computer vision and digital image processing (Dureja and Pahwa, 2018). The techniques on the forgery evaluation in the images by different researchers are classified into the passive and active approaches (Walia and Kumar, 2018)(Ansari et al., 2014)(Asghar et al., 2017) as shown in Fig. 1.

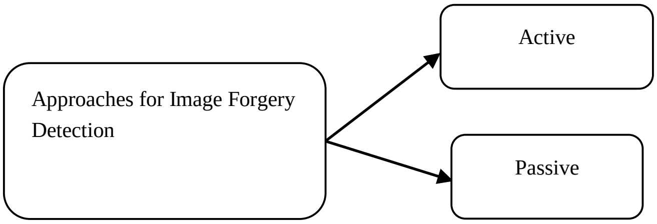

Fig. 1 - Image Forgery Detection Approaches.

In active approaches, images must be pre-embed either with a digital signature or watermarking, whereas in passive approaches no pre-embed information is required in the image.

The key goal of this manuscript is to perform the process for detection of image forgery using convolutional neural network, a subdomain of machine learning. The approach is based on pre-trained existing AlexNet Architecture on the publicly available benchmark dataset MICC-F220(Amerini et al., 2011), which encomprising of total 220 images, 110 forged images and 110 non forged images.

1.1. AlexNet Architecture and Layers

In the recent years, advances in deep learning, more concretely especially convolutional neural networks progressing at a dramatic pace. The architecture of a CNN determines how many layers it has, what each of these layers is doing, and how the layers are connected to each other. Choosing a good architecture is crucial to successful learning with a CNN. For our main training tasks, we have used the pre-trained CNN based AlexNet architecture. The network contains multiple layers with learnable parameters.

The concept of AlexNet model was proposed by (Krizhevsky et al., 2012). The AlexNet model consists of 25 layers. Table 1 shows the AlexNet model layers along with a description. The main layers of the AlexNet model are convolutional, pooling, fully connected and softmax along with activation function ReLU.

Table 1 - AlexNet Model Layers and Description

Layer No.	Layer	Description
1.	Input Layer	227x227x3 first layer size
2.	Convolution Layer 1	96 filters of size 11x11x3 at stride 4 with pad 0
3.	ReLU	rectified linear units activation function
4.	Normalization	cross channel normalization layer
5.	Max Pooling	3x3 filter at stride 2
6.	Convolution Layer 2	256 filters of size 5x5 at stride 2 with pad 2
7.	ReLU	rectified linear units activation function
8.	Normalization	cross channel normalization layer
9.	Max Pooling	3x3 filter at stride 2
10.	Convolution Layer 3	384 filters of size 3x3 at stride 1 with pad 1
11.	ReLU	rectified linear units activation function
12.	Convolution Layer 4	384 filters of size 3x3 at stride 1 with pad 1
13.	ReLU	rectified linear units activation function
14.	Convolution Layer 5	256 filters of size 3x3 at stride 1 with pad 1
15.	ReLU	rectified linear units activation function
16.	Max Pooling	3x3 filter at stride 2
17.	Fully Connected (fc6)	4096 fully connected layer
18.	ReLU	rectified linear units activation function
19.	Dropout	50% dropout
20.	Fully Connected (fc7)	4096 fully connected layer
21.	ReLU	rectified linear units activation function
22.	Dropout	50% dropout
23.	Fully Connected (fc8)	1000 fully connected layer
24.	Softmax	softmax layer
25.	Classification Output	output classified 1000 classes

AlexNet model used the ReLU activation function which is non-saturating, and training performance is improved over sigmoid and tanh activation functions.

This paper is organized as follows. Section 2 contains the related work, section 3 introduced the proposed method, in section 4 proposed model is implemented, section 5 is about results, comparison with existing approaches, discussion and finally a conclusion.

2. Related Work

Most existing methods in the literature uses the extraction of explicit features, including statistical based, geometrical based, wavelet-based, block based, keypoint based, transformations based, texture based and so on. Most of the methods require hand-crafted or feature engineering. Most of the features have good results but not invariant to different types of geometrical operations and less robust to various types of image forgery.

To improve the accuracy of image forgery detection, some studies utilized machine learning, deep learning and CNN based approaches. (Hakimi, 2015)(Chen et al., 2015)(Y. Rao and Ni, 2017)(J. Zhang et al., 2016)(Y. Zhang et al., 2016)(Chen et al., 2017).

(Amerini et al., 2011) proposed approach for image forgery detection using Scale Invariant Features Transform(SIFT) features for the dataset MICC-F220 and MICC-F2000 and able to deal with affine geometric transformations. The False Positive Rate (FPR) and True Positive Rate (TPR) achieved is 8% and 100% respectively.

(Mishra et al., 2013) proposed approach for image forgery detection approach using speeded up robust features (SURF) and hierarchical agglomerative clustering (HAC) for the dataset MICC-F220.

(Fridrich et al., 2003) the approach is based on discrete cosine transform (DCT) features for each block and through lexicographical sorting of block-wise DCT coefficients forgery of the image is detected. This approach is only able to identify forgery with small variations in scaling and rotation. (Popescu and Farid, 2004) applied PCA on image blocks to reduce the dimension space and performed lexicographical sorting and robust to minor variations in the image due to lossy compression or additive noise.

(Diaa M. Uliyan et al., 2016) proposed image forgery detection approach based on combined features of Hessian points and a center-symmetric local binary pattern (CSLBP), which makes the features invariant to scale, translation and illumination, but not invariant under blur degradations. FPR and TPR obtained are 8% and 92% respectively.

(D.M. Uliyan et al., 2016) proposed image forgery detection approach based on Fourier and Gabor texture features and robust to blur artifact using two datasets, namely, Image data manipulation and MICC-F220. FPR and TPR obtained are 2.86% and 96.5% respectively.

(Hakimi, 2015) proposed splicing detection method based on LBP, PCA and SVM. In the proposed method, primarily the RGB image was converted into YCbCr color space and 16×16 non-overlapping blocks were formed on the basis of chrominance components. The features were extracted using LBP and DWT on all the blocks and subsequently PCA was also applied to increase the efficiency of the algorithm and these features given as input to SVM for classification.

(Chen et al., 2015) proposed the modified version of CNN to detect cut and paste forgery. A filter layer was added before the first convolutional layer to take an image as its input and output the Median Filtering Residual (MFR) of the image. The proposed method learned hierarchical features representation automatically with low false rate and high detection rate.

(Yuan Rao and Ni, 2017) stated automated hierarchical feature representations learning model to detect splicing and copy-move forgeries. They proposed the CNN model with 8 convolutional layers and a fully connected layer with a 2-way classifier.

(J. Zhang et al., 2016) proposed CNN based models to detect copy-move forgery. In the fundamental models with two forms, Siamese and pseudo-siamese, there were 3 convolutional layers with 2 max-pooling layers and 2 fully connected layers with a softmax layer.

(Y. Zhang et al., 2016) presented the two-stage deep learning approach using Stacked Autoencoder (SAE) model for the detection of forged images. (Zhou et al., 2017) presented the CNN model with blocking strategy for image forgery detection. Firstly, the image was divided into blocks using tight blocking and marginal blocking. Then, the blocks were inputted into the rich model Convolutional Neural Network (rCNN). At last, the pooling was performed, followed by the classification of the input image based on the feature vectors using the SVM classifier. (Chen et al., 2017) presented the image splicing detection using Camera Response Function and deep learning. The CNN model was trained on edge patches of authentic image and the forged image. The edge patches were extracted from the image and then the features were extracted from these patches and classified using CNN to localize the spliced region.

3. Proposed Method

The proposed approach has been implemented to detect and recognize whether the digital image under investigation is forged or not using the CNN based pre-trained AlexNet model on the publicly available benchmark MICC-F220 dataset images (Amerini et al., 2011). It is observed that the performance of the deep learning features based on AlexNet model is quite satisfactory. In this approach, a number of input corresponds to the number of images to perform on pre-trained AlexNet based convolutional operations and pooling with Relu activation function to extract the deep features. SVM classifier is trained with the extracted deep features from the pre-trained AlexNet model and compared the result with the six different state-of-the-art approaches for the MICC-F220 dataset. In this experiment publicly available MICC-F220 dataset has been used for image level forgery detection. Labels are marked manually. The images of the dataset are pre-processed and resized to 227x227 as per the first input layer of the AlexNet model and features get extracted from the fully connected f7 layer. To reduce the effect of random samples for the deep features, the average classification accuracy is computed with five iterations over the images in the dataset.

4. Implementation

This section presents the implementation of the proposed approach using the hardware as Intel(R) Core™ i7-5500U CPU with 2.40 GHz, 16 GB RAM and software as Ubuntu 16.04 with Matlab release R2018a and compared the proposed approach with the existing state-of-the-art approaches.

4.1. Dataset

In this section, MICC-F220 (Amerini et al., 2011) publically available benchmark dataset is used for the experimental result. This dataset consists of 110 non-forged and 110 forged with 3 channels i.e. color images of size 722 × 480 to 800 × 600 pixels with 10 different combinations of geometrical and transformations attacks to the original image as shown in Fig. 2. and Fig. 3. This dataset is used for the detection of forged images where cloned or copy-move forgery is carried out.

4.2. Classifier

SVM is used as a classifier. SVM is popular and efficient (Mushtaq and Mir, 2014) for binary classification. Performance of the proposed approach is evaluated at image level by calculating the performance metrics as Precision, False Positive Rate (FPR), Recall also known as True Positive Rate (TPR), F-measure and Accuracy along with the execution time.

Fig. 2 - original image

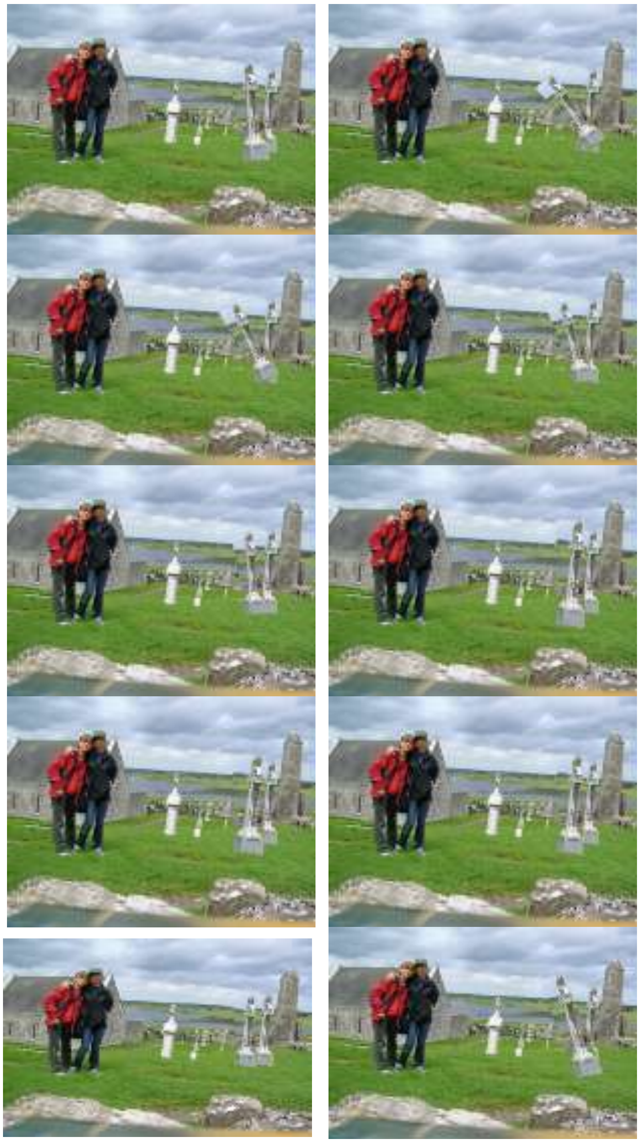

Fig. 3 - 10 different combinations of geometrical and transformations Attacks

4.3. Proposed Approach

The proposed approach is carried out in 2 stages, one involves the training of SVM classifier using a pre-trained AlexNet model based on deep features. The flow diagram of the proposed approach is shown as in Fig. 4.

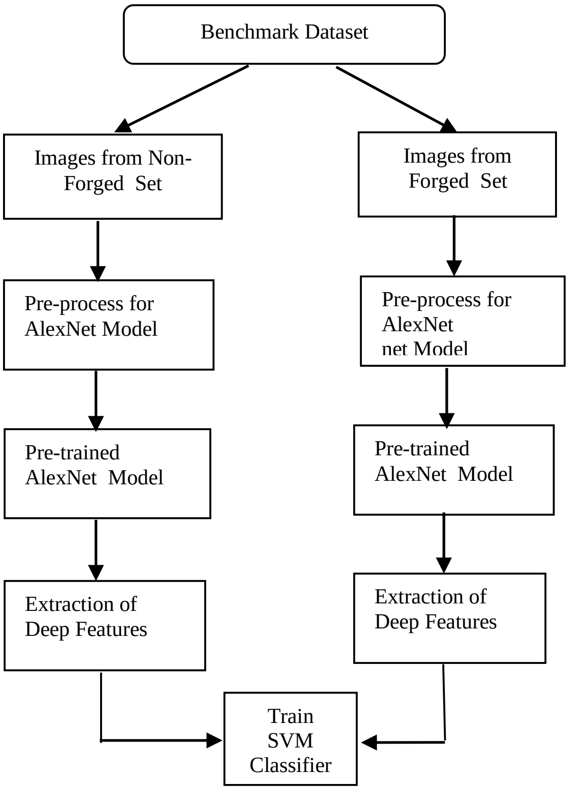
Benchmark Dataset
Images from Non-Forged Set
Images from Forged Set
Pre-process for AlexNet Model
Pre-process for AlexNet net Model
Pre-trained AlexNet Model
Pre-trained AlexNet Model
Extraction of Deep Features
Extraction of Deep Features
Train SVM Classifier

Fig. 4 - Block diagram to train the SVM classifier using AlexNet Model based Deep Features

Second is the testing step, test images are given as input to the classifier to choose whether the given image is forged or not as shown in Fig. 5.

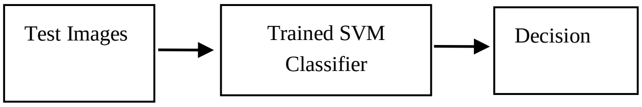

Fig. 5 - To verify whether the image is forged or not

The approach can be summed up as follows

- For the training step, images are selected from the forged and non-forged image set.

- Images are pre-processed as per the input layer of the first layer of AlexNet Model.
- Pre-trained CNN based AlexNet model is used to extract deep features.
- Feature vectors comprising of 4096 deep features used to train the SVM Classifier.
- Test images are chosen from the dataset randomly and feed into the trained SVM classifier to predict the image is forged or not. Results are calculated for the given dataset and mentioned in the next section.

5. Result and Discussion

There are various evaluation criteria to find out the accuracy of the experimental results. In this paper, the performance metrics of the CNN based pre-trained AlexNet model's deep features using SVM classifier is evaluated using the confusion matrix as shown in table 2 and table 3. The metrics calculated are accuracy, true positive rate (recall), false positive rate, precision, f-measure and execution time.

Table 2 – Confusion Matrix

Actual	Predicted Forged	Predicted Non-Forged
Forged	(True Positive) TP	(False Negative) FN
Non-Forged	(False Positive) FP	(True Negative) TN

TP - Forged Image detected as forged
FN - Forged Image detected as non-forged
FP - Non-Forged Image detected as forged
TN - Non-Forged Image detected as non-forged

False Positive Rate (FPR) = FP/(FP+TN) (1)
True Positive Rate (TPR) or Recall = TP/(TP+FN) (2)
Precision = TP/(TP+FP) (3)
F-Measure = 2 * ((Precision x Recall) / (Precision + Recall)) (4)
Accuracy = (TP+TN)/(TP+TN+FP+FN) (5)

Table 3 - Confusion Matrix for the Test Data Set

Test Dataset	Forged Predicted	Non-Forged Predicted	Accuracy
Forged	50%	0%	93.94%
Non-Forged	6.06%	43.94%	

It is observed that with MICC-F220 dataset the accuracy is 93.94% with Recall or TPR rate is 100%, Precision is 89.19%, F-measure is 94.28% and the average execution time for the approach is 4.86 seconds.

A comparison of the proposed approach with the state-of-the-art approaches for the MICC-F220 dataset is given as per the table 4.

Table 4 FPR, TPR and time (seconds) for different methods on MICC- F220 dataset

Approach	FPR, %	TPR, %	Time, s
(Amerini et al., 2011)	8	100	4.94
(Mishra et al., 2013)	3.64	73.64	2.85
(Fridrich et al., 2003)	84	89	294.69
(Popescu and Farid, 2004)	86	87	70.97
(Diaa M. Uliyan et al., 2016)	8	92	NA
(D.M. Uliyan et al., 2016)	2.86	96.5	NA
Proposed Approach	12.12	100	4.86

The performance of forgery detection approach is evaluated at the image level. The error is measured by FPR where some non forged images detected as forged and TPR where the forged image is detected as forged.

6. Conclusion

This paper proposed a image forgery detection approach using CNN based pre-trained AlexNet model to extract deep features, without investing much time in training. The proposed approach also exploits the SVM as a classifier. Compared to the previous work on MICC-F220 dataset, the best accuracy of image forgery detection achieved is 93.94%. In this paper, MICC-F220 dataset encomprising of 220 images of forged and non-forged images are classified using SVM Classifier. Performance of the deep features extracted from a pre-trained AlexNet based model is quite satisfactory, even in the presence of rotational and geometrical transformation and also compared the results of the given approach with the existing state-of-the-art approaches. In the future, we plan to work on various benchmark image forgery datasets and to compare the performance with the existing approaches.

References

Amerini I, Ballan L, Member S, Caldelli R, Bimbo A Del, Serra G. A SIFT-Based Forensic Method for Copy – Move Attack Detection and Transformation Recovery 2011;6:1099–110.

Ansari MD, Ghrera SP, Tyagi V. Pixel-Based Image Forgery Detection: A Review. IETE J Educ 2014;55:40–6. doi:10.1080/09747338.2014.921415.

Asghar, K., Habib, Z., & Hussain, M. (2017). Copy-move and splicing image forgery detection and localization techniques: a review. *Australian Journal of Forensic Sciences*, *49*(3), 281-307

Chen, C., McCloskey, S., & Yu, J. (2017, July). Image Splicing Detection via Camera Response Function Analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 5087-5096).

Chen, J., Kang, X., Liu, Y., & Wang, Z. J. (2015). Median filtering forensics based on convolutional neural networks. IEEE Signal Processing Letters, 22(11), 1849-1853.

Dureja A., Pahwa P. (2018), Image Retrieval Techniques: A survey, International Journal of Engineering & Technology. Vol 7, No 1.2, 215-219

Fridrich, A. J., Soukal, B. D., & Lukáš, A. J. (2003). Detection of copy-move forgery in digital images. In *in Proceedings of Digital Forensic Research Workshop*.

Hakimi, F., Hariri, M., & GharehBaghi, F. (2015, November). Image splicing forgery detection using local binary pattern and discrete wavelet transform. In *Knowledge-Based Engineering and Innovation (KBEI), 2015 2nd International Conference on* (pp. 1074-1077). IEEE

https://ai.intel.com. n.d.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).

Mishra, P., Mishra, N., Sharma, S., & Patel, R. (2013). Region duplication forgery detection technique based on SURF and HAC. *The Scientific World Journal*, *2013*

Mushtaq, S., & Mir, A. H. (2014, November). Forgery detection using statistical features. In *Computational Intelligence on Power, Energy and Controls with their impact on Humanity (CIPECH), 2014 Innovative Applications of* (pp. 92-97). IEEE.

Popescu, A. C., & Farid, H. (2005). Exposing digital forgeries in color filter array interpolated images. *IEEE Transactions on Signal Processing*, *53*(10), 3948-3959.

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

Uliyan DM, Jalab HA, Abdul Wahab AW. Copy move image forgery detection using Hessian and center symmetric local binary pattern. ICOS 2015 - 2015 IEEE Conf Open Syst 2016:7–11

Uliyan, D. M., Jalab, H. A., Wahab, A. W. A., Shivakumara, P., & Sadeghi, S. (2016). A novel forged blurred region detection system for image forensic applications. Expert Systems with Applications, 64, 1-10.

Walia, S., & Kumar, K. (2018). Digital image forgery detection: a systematic scrutiny. *Australian Journal of Forensic Sciences*, 1-39

Zhang, J., Zhu, W., Li, B., Hu, W., & Yang, J. (2016, November). Image copy detection based on convolutional neural networks. In Chinese Conference on Pattern Recognition (pp. 111-121). Springer, Singapore.

Zhang, Y., Goh, J., Win, L. L., & Thing, V. L. (2016, January). Image Region Forgery Detection: A Deep Learning Approach. In *SG-CRC* (pp. 1-11)

Zhou, J., Ni, J., & Rao, Y. (2017, August). Block-Based Convolutional Neural Network for Image Forgery Detection. In *International Workshop on Digital Watermarking* (pp. 65-76). Springer, Cham