Image Forgery Detection using CNN

Aadeesh Jain, Aditya Sharma, Kanishk Gupta, Ketan Likhi, Neha Mehra, Sonika Shrivastava, Divyansh Joshi

Department of Computer Engineering, Shri Govindram Seksaria Institute of Technology and Science, Indore, India

Corresponding author: Aadeesh Jain, Email: aadeeshjain.91a@gmail.com

In today's world, digital images are widely used in various domains such as; newspapers, scientific journals, magazines, and many other fields. Unfortunately, today's digital technology made it easy for digital images to be forged due to the availability of the low cost photo editing software like Adobe Photoshop. Thus, in order to recover people's trust towards digital images, it is important to develop new trustworthy techniques for digital images forgery detection. In this paper we present a novel fake image detection model where the acquisition method is based on an in-depth, process-based, convolutional neural network (CNN) for automatic learning and identifying independent presentations from color RGB input images. CNN is being used because of the advantages it provides as it performs feature engineering, i.e. feature extraction and feature selection, which was earlier performed using different statistical observations and methods.The research paper discusses transfer learning approach which has its own advantages, as it's a different approach from the custom CNN which was being used in earlier approaches.

Keywords: Deep Learning, CNN, Image Processing, Transfer Learning.

1 Introduction

Fake or tampered images shared on the Internet and social media may be misleading, cause emotional disturbance and affect public opinion and actions. And Studies have shown that people usually are least interested in checking the authenticity of image. Easy-to-use digital imaging software is ubiquitous. Driven by the rapid development of technology in the 21st century, the skills, cost, workload and the time required to create convincing visual fakes have been greatly

2021. In Raju Pal & Praveen K. Shukla (eds.), *SCRS Conference Proceedings on Intelligent Systems*, 293–304. Computing & Intelligent Systems, SCRS, India.
https://doi.org/10.52458/978-93-91842-08-6-29

reduced. These tampered images are propagated through popular social media applications, for instance 300 million images on facebook, 95 million images on instagram, and 140 million tweets, are being uploaded daily. With the advent of photo editing software, the number of processed images has greatly increased, which may be unknown.

These processed images are shared on social media with increasing frequency and complexity. Prediction of the exact daily uploads of forged image is still unknown. A large number of examples show that manipulated images have caused considerable damage on all levels consisting of the individuals, organizations, and society. The damage caused by manipulated or forged images is real. Studies have shown that manipulating images can distort the viewer's memory, further improve the fidelity of these images, and even affect decision-making behaviors such as voting.

Thus, in order to recover people's trust towards digital images, it is important to develop new trustworthy techniques for digital images forgery detection.

The Image forgery detection techniques has been divided into two approaches

- Active Apporach
- Passive Approach

The Active Approach is divided into 2 types

- Digital Watermarking
- Digital Signatures

The Passive Approach has also been further divided into 3 types:

- Image Splicing
- Copy Move
- Image Retouching

The paper mainly focuses on Image Splicing and Copy Move Forgery. The copy-move is defined by copying region of an image and pasting it in another place in the same image, generally to hide unwanted parts of the image. On the other hand, image splicing can be defined as copying a region from one image and pasting it in another image.Thus, detection of tampered regions is done through searching for very similar regions in copy-move images and completely odd regions in spliced images.

Following works have been done to detect forgery in images. Previously for feature extraction various computer vision algorithm were used among which major one is SURF(Speed Up Robust Feature) which is a descriptor used to recognise and locate the objects. Key role was extracting features from raw image. The latest works in the field of Image Forgery Detection has been done using Deep Learning, using CNN which could automatically learn feature representations, it was observed that the results obtained on the model gave satisfactory results.

Deep Learning methods outperforms traditional methods as there is no need of defining features and to do feature engineering, thus they give higher levels of accuracy and is most widely used.

Figure 1: Copy Move Forgery

The difficulty with the traditional computer vision based algorithms is that while performing feature extraction we need choose which features to look for in each given image, also when number of classification goes high or image clarity goes down then it is hard to cope up with these algorithms.

2 Related Work

An extensive amount of research has been conducted to mitigate the problem of identifying forged images and classifying images as forged or pristine [1]. Researchers proposed a new methodology for picture splicing detection. It includes a block discrete cosine transform (DCT) implementation for photos, and mixes the function extracted from the statistical moment of the feature function with the matrix function of each spatial domain and Markov transform capability in DCT so that it can Used for auxiliary vectors (SVM). They further developed the technique by referring to the Markov causal version implemented in each of the DCT and DWT domains and training the SVM classifier using the cross-region functions.

In another paper [2], researchers examined several block-based methods to detect the copy-move forgery. Bayram et al. showed their time complexity and robustness in the results. They discussed Discrete Cosine Transform (DCT),Fourier Mellin Transform (FMT) and Principal Component Analysis (PCA). The results were good on any JPEG image, but the algorithm is limited to non-rotated or scaled objects. However, the efficiency of copy-move forgery was increased with help of techniques like counting bloom filters which is applicable when resolution or the quality of the image was high.

Figure 2: Image Splicing

An alternative method of detecting and locating stitched images was evolved by identifying local noise inconsistencies caused by post-processing or camera sensors [3]. Local Binary Pattern (LBP) and Controlled Pyramid Transform (SPT) have been used to detect the deformation of texture attributes in fake images, and the CASIA dataset has achieved the latest recognition performance so far. Previously for feature extraction various computer vision algorithm were used [4]. In one of such algorithms, They used the hit point technique after the feature extraction process through SIFT and SURF. To detect the splicing, they extracted the edges of the integral image of the Y, Cb and Cr image components. Then the feature vector is passed to the SVM classifier. The results show that the extraction of SURF features is more efficient than the extraction of SIFT.

Previously an algorithm for digital image forgery detection based on shadow detection of the spliced object was presented by Ying et al [5]. They based their algorithm on the fact that a shadow wouldn't change the surface texture, thus if two adjacent areas (with and without shadow) had different texture, then the image could very likely be forged. The algorithm used Local Binary Pattern (LBP) from shadow areas and adjacent non-shadow areas. The energy and entropy extracted from the features histograms proved to be the most discriminating.

In recent years [6], neural networks such as Convolutional Neural Network (CNN), BeliefNetwork and the Auto Encoder have been shown to be able to extract complex mathematical relationships in a variety of sensory inputs and

successfully study their royal presentations. It can perform a various activities including image classification, speech recognition, etc.

Yuan Rao and his team [7] have proposed a new method of acquiring photography that allows automatic reading of feature presentations based on an in-depth learning framework. Their main work focuses on the following:

1. They trained a CNN supervised learning model to learn the sequence features of deviation activities such as spinning and copying-navigation with labeled labels from training photos.

2. Features in image are extracted with a pre-trained CNN model using sliding-window to scan the entire image.

3. Finally, SVM segregation is trained in accordance with the introduction of the accepted aspect of image separation (true / counterfeit).

3 Method

We proposed two approaches for training and testing of our CNN model. The first approach does patch sampling on the dataset and the patches are then supplied to the CNN for training and evaluation. In the second approach proposed we perform, data augmentation and resize the image and supply it for training of the CNN and for the testing.

3.0.1 With Patch Sampling

- Training Algorithm
  - The image to be given as an input to the CNN, first needs to be pre-processed.
  - For any forged image we need to identify the part from where it is forged before feeding to CNN, so the pre-processing steps first include creating masks of a given image for this purpose.
  - We use that mask to sample the patches from fake image along the boundary of the spliced region in such a way so as to ensure at least a 25 percent contribution from both forged part and unforged part of the image, patches of size 64X64 are taken.
  - For forged images we take patches from the sampling done in the previous step.

  - For pristine images we take patches randomly but roughly the number is same as forged images patches.
  - The labelled images are given as input to the CNN for its training.
  - Soft-max classifier is applied after the CNN to classify whether image is forged or pristine.
  - The result is then compared with ground reality of image and weights and biases are adjusted during back-propagation
  - Steps 5-8 are repeated over multiple epochs, till the optimum value parameters is achieved.

- Testing Algorithm
  - The image to be tested is divided into patches of specific size(64X64).
  - They are then given as input in the CNN.
  - The CNN identifies whether the patch is forged or pristine and gives the result from softmax classifier.
  - All the patches along with with the result(as predicted by CNN)are stored.
  - If the percentage of forged patches cross a certain value then the whole image is forged otherwise it is considered unforged.

3.0.2 Without Patch Sampling

The above mentioned approach is computationally expensive, a different approach to train the model is,

- Training Algorithm
  - The images are resized to 224*224 standard input size for VGG16,
  - If the image is not same as the target size, images are interpolated accordingly following best approach.
  - All the images are then converted to RGB images.
  - Image augmentation is applied, where different translations, rotations are applied.
  - The images are then supplied in batches, and the CNN is trained first by freezing all the layers and tuning only the newly connected fully connected layer.

  - After that last block of VGG16 is made trainable and again training of CNN is done.

- Testing Algorithm
  - The images are resized to 224*224 standard input size for VGG16,
  - If the image is not same as the target size, images are interpolated accordingly following best approach.
  - All the images are then converted to RGB images.
  - Images are then supplied to CNN, and CNN decides whether the image is forged or pristine, a threshold is decided, to distinguish for forged or pristine image.

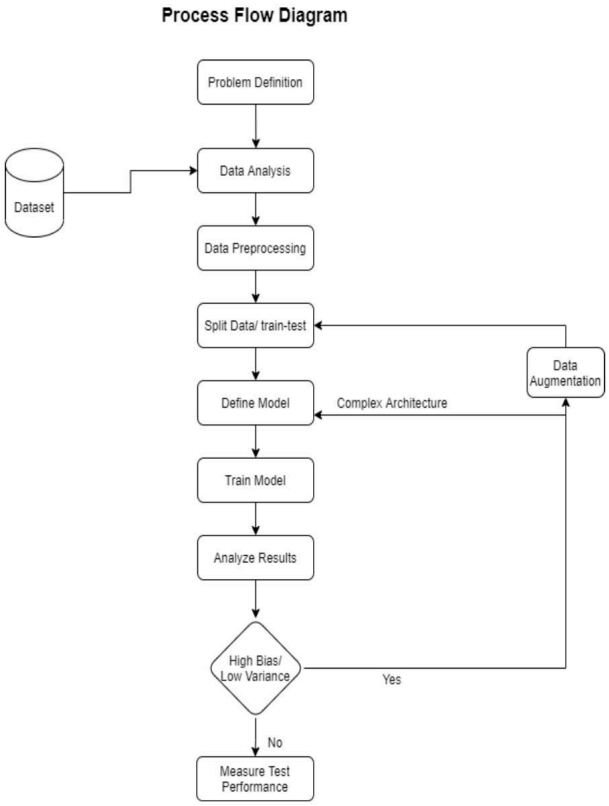


Figure 3: Process-Flow Diagram

System Architecture In this paper a transfer learning based model is used.We are using VGG16 model which was trained on IMAGENET.

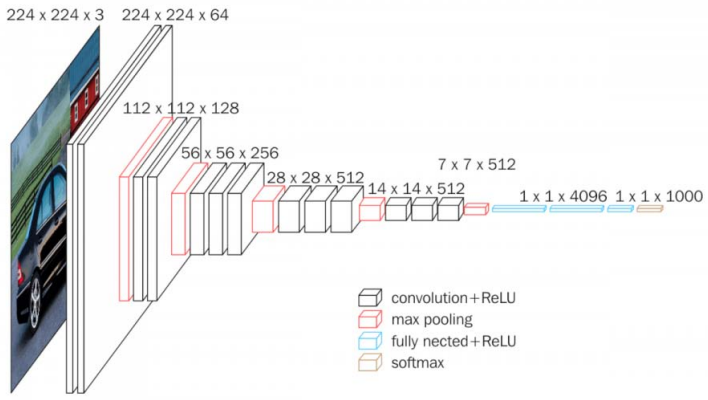


Figure 4: VGG16 Architecture

4 Results

A. Image Dataset

The training of the CNN model is done primarily using 2 public data-sets for forgery detection, i.e., CASIA v1.0 and CASIA v2.0 [8]. dataset contains images in JPEG format with 1,725 coloured images of dimensions 384×256 pixels, among which, dataset contains 925 forged images and the remaining 800 are pristine images. The CASIA v2.0 dataset contains images in JPEF, BMP, TIFF formats with 12,614 coloured images with dimensions ranging from 240 × 160 to 900 × 600 pixels, among which 7,491 pristine and 5,123 forged images.The data-sets contain copy-move and spliced forged images.

The dataset split was as 80% training and 20% testing.

B. Performance Analysis

Following results are observed on the testing images when the model is tested on some RGB images: Testing has been done with different types of images,some of them are pristine and others have forgeries like copy-move and image splicing. Here, the green boundary is representing that the image is predicted pristine and red boundary represents that the image if predicted forged.

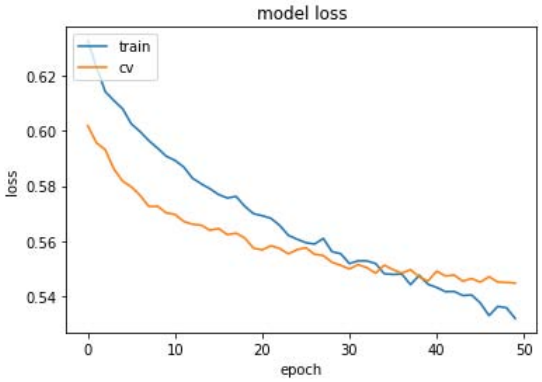


Figure 5: Loss vs Epoch

The continuous decrease in the validation loss as well as training loss is signifying that the model was converging, but logs show that model started to over-fit as it's loss stopped decreasing after certain number of epochs.

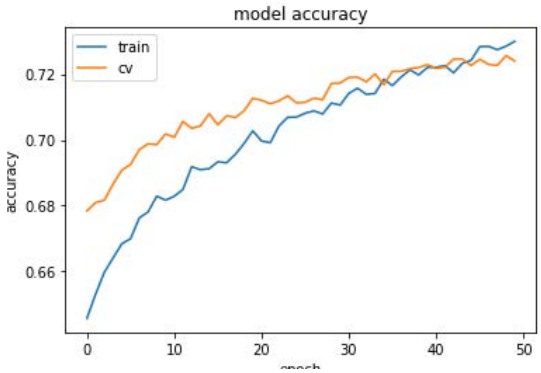


Figure 6: Accuracy vs Epoch

The continuous increase in the validation accuracy as well as training accuracy is signifying that the model is continuously improving.

Following are the different test cases:

4.1 Test Case - 1

Input Image and Prediction:

Figure 7: Pristine Image

Testing Result :
Predicted - 'Authentic'
Actual - 'Authentic'
Prediction Probability - 0.1087583

4.2 Test Case - 2

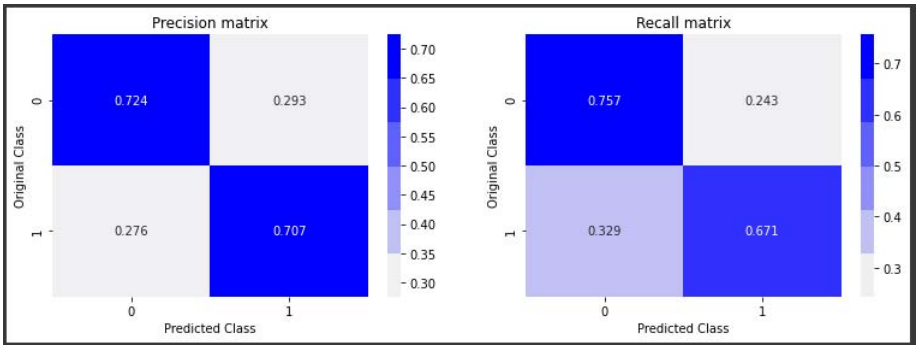
Input Image and Prediction:

Figure 8: Fake Image

Testing Result :
Predicted - 'Fake'
Actual - 'Fake'
Prediction Probability - 0.61155885

Table 1: Comparison Table

	VGG16	ResNet50
Patch Sampling	71.66	68.5
Without Patch Sampling	70.77	67.4



5 Discussion

In this paper, from the proposed method it can be concluded that this model can significantly detect tampered images.

The transfer learning approach has its own advantages, as it's a different approach from the custom CNN which was used in earlier approaches, the model in transfer learning is pre-trained on Imagenet dataset which had around 14,197,122 images. Removing the VGG16's top model, and adding our own fully connected layer,and training the model on CASIA dataset gave a respectable accuracy.

Whilst the overall accuracy of the model was promising, we've also learnt that the success rate can be further improved in future, provided the dataset prepared is investigated more, and covers different aspects of forgery. Further work can be done in, designing a better algorithm which can predict forgeries in different types of images which can be tampered using techniques like image recolouring, image retouching along with copy-move and image splicing forgery detection,

2021. In Raju Pal & Praveen K. Shukla (eds.), *SCRS Conference Proceedings on Intelligent Systems*, 293–304. Computing & Intelligent Systems, SCRS, India.
https://doi.org/10.52458/978-93-91842-08-6-29

also predicting forgery in text documents and videos.

References

[1] Kuznetsov, A. (2019). Digital image forgery detection using deep learning approach.*Journal of Physics: Conference Series*, 1368(3):1368 032028.

[2] Agarwal, R. et al. (2020). Image Forgery Detection and Deep Learning Techniques: A Review. In *4th International Conference on Intelligent Computing and Control Systems*, 1096-1100.

[3] Shi, Y. Q., Chen, C. and Chen, W. (2007). A natural image model approach to splicing detection. In *Proceedings of the 9th workshop on Multimedia & security*, 51–62.

[4] William, Y., Safwat, S. and SalemRobust, M. A. M. (2019). Image Forgery Detection Using Point Feature Analysis. In *Federated Conference on Computer Science and Information Systems*, 373-380

[5] Ying, Z. et al. (2016). Image Region Forgery Detection: A Deep Learning Approach. In *Proceedings of the Singapore Cyber-Security Conference*, 14:1–11.

[6] LeCun, Y. et al. (1998). Gradient-based learning applied to document recognition.In *Proceedings of the IEEE*, 86(11):2278–2324.

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

[8] Kaggle DataSet : https://www.kaggle.com/sophatvathana/casia-dataset