A DEEP-LEARNING BASED APPROACH FOR DETECTING SPLICING AND COPY-MOVE IMAGE FORGERIES

A PROJECT REPORT
SUBMITTED BY

Aravind J 2019115017
Krishnan S 2019115047
Pranay Varma 2019115067

Submitted to the faculty of
INFORMATION TECHNOLOGY

DEPARTMENT OF INFORMATION SCIENCE AND TECHNOLOGY
COLLEGE OF ENGINEERING, GUINDY
ANNA UNIVERSITY
CHENNAI 600 025

ABSTRACT:

Digital picture usage has increased at a never-before-seen rate in our day and age, due to the proliferation of gadgets like smartphones and tablets. Furthermore, the development of user-friendly image manipulation software that is available at reasonable prices, has made the manipulation of such content easier than ever. Some of these images are tampered in such a way that it is absolutely impossible for the human eye to detect.

Three of the most common image manipulation techniques are:
i) **Splicing**: In splicing a region from an authentic image is copied into a different image.

ii) **Image-inpainting:** In image inpainting an image region is removed and the removed part is then filled in to complete the image.

ii) **Copy-move:** A specific region from the image is copy pasted within the same image.

PROBLEM STATEMENT AND OBJECTIVES

The goal of this study is to detect **Splicing and Copy-Move** forgeries in images using **CNN, self-consistency learning** and **unsupervised domain adaptation** and analyse how the performance of image forgery detection varies based on the test sample difficulty and the deep-learning model used.

LITERATURE SURVEY

1. **An Efficient CNN Model to Detect Copy-Move Image Forgery**
(K. M. Hosny, A. M. Mortda, M. M. Fouda and N. A. Lashin) in IEEE Access, vol. 10, pp. 48622-48632, 2022, doi: 10.1109/ACCESS.2022.3172273.

This research paper is about the copy-move image foregery, where a part of an image is copied and placed in the same imaeg to produce forgery image. In this paper an accurate CNN architecture is proposed for effective detection of copy-move forgery. It is lightweight and presents fast and accurate testing process.

2.**Copy Move and Splicing Image Forgery Detection using CNN**
(Devjani Mallick, Mantasha Shaikh, Anuja Gulhane, Tabassum Maktum)
ITM Web Conf. 44 03052 (2022) DOI: 10.1051/itmconf/20224403052
This paper present a method using CNN with three different models like Error Level Analysis to identify copy move and splicing picture forgeries.The pre-processing methodology is used in the suggested method to acquire the images at a specific compression rate and they are used to train the model.

3. **Syn2Real:Forgery Classification via Unsupervised Domain Adaptation** (Kumar, Akash, Arnav Bhavsar, and Rajesh Verma) - *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops*. 2020.

This research suggests using the copy-move forgery technique and deep semantic image inpainting to produce a synthetically faked dataset. When it was tested on more accurate data, models developed on these datasets performed significantly worse. This issue has been addressed by using unsupervised domain adaption networks to map the feature space from our artificially generated dataset and identify copy-move fraud in new domains. On the CASIA and CoMoFoD datasets the F1 score was also improved, bringing it to 80.3% and 78.8%, respectively. In situations where the classification of data is absent, this approach can be useful. Since dataset other than publicly available dataset is used, it will have high accuracy in realistic test data.

4. **Fighting Fake News: Image Splice Detection via Learned Self-Consistency** (Huh, Minyoung, Andrew Liu, Andrew Owens, and Alexei A. Efros.) - In *Proceedings of the European conference on computer vision (ECCV)*, pp. 101-117. 2018.

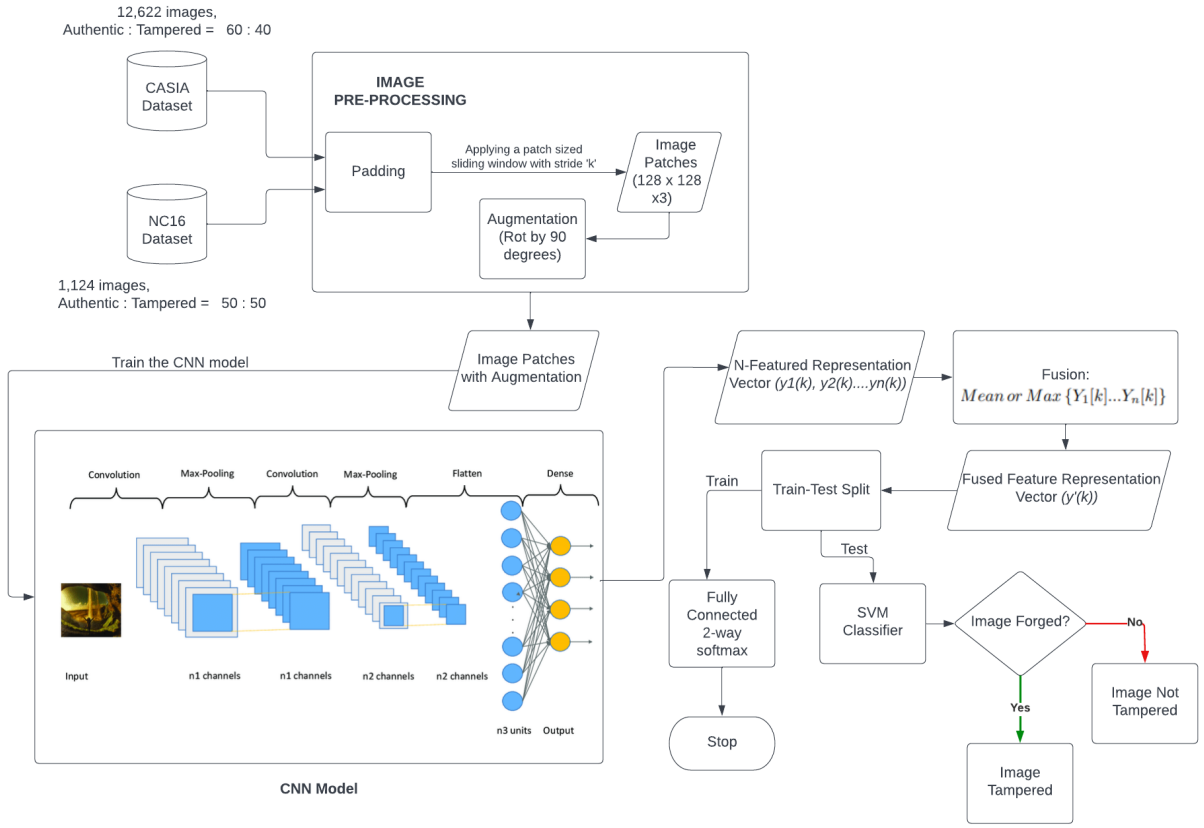
The proposed algorithm uses the automatically recorded photo EXIF metadata as supervisory signal for training a model to determine whether an image is self-consistent — that is, whether its content could have been produced by a single imaging pipeline. This self-consistency model has been used for detecting and localizing image splices.

5. **A Deep Learning Approach to Detection of Splicing and Copy-Move Forgeries in Images** (Y. Rao and J. Ni) - In the proceedings of IEEE International Workshop on Information Forensics and Security (WIFS), 2016, pp. 1-6, doi: 10.1109/WIFS.2016.7823911.

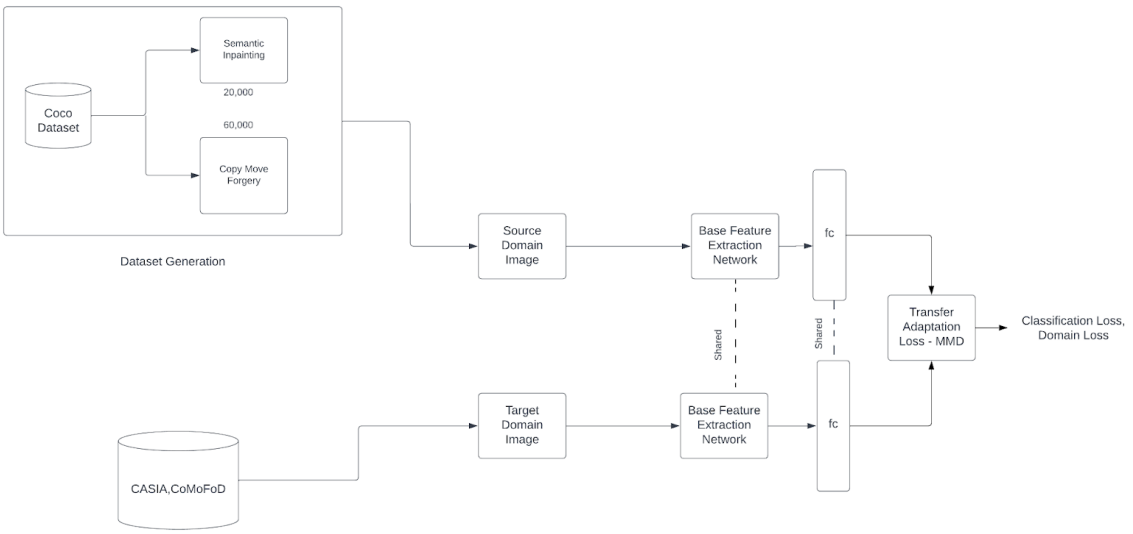
The proposed method makes use of convolutional neural networks (CNNs), which automatically build hierarchical representations from input RGB colour images. The CNN is made specifically for applications like copy-move detection and picture splicing. CNN is a patch descriptor here, which is pre-trained based on the labeled patch samples. The pre-trained CNN is then used to extract dense features from the test images, and a feature fusion technique is incorporated to obtain the final discriminative features for SVM classification. SVM's rbf model is used. It outperforms many state of the art models, in terms of speed and accuracy

ARCHITECTURE DIAGRAM

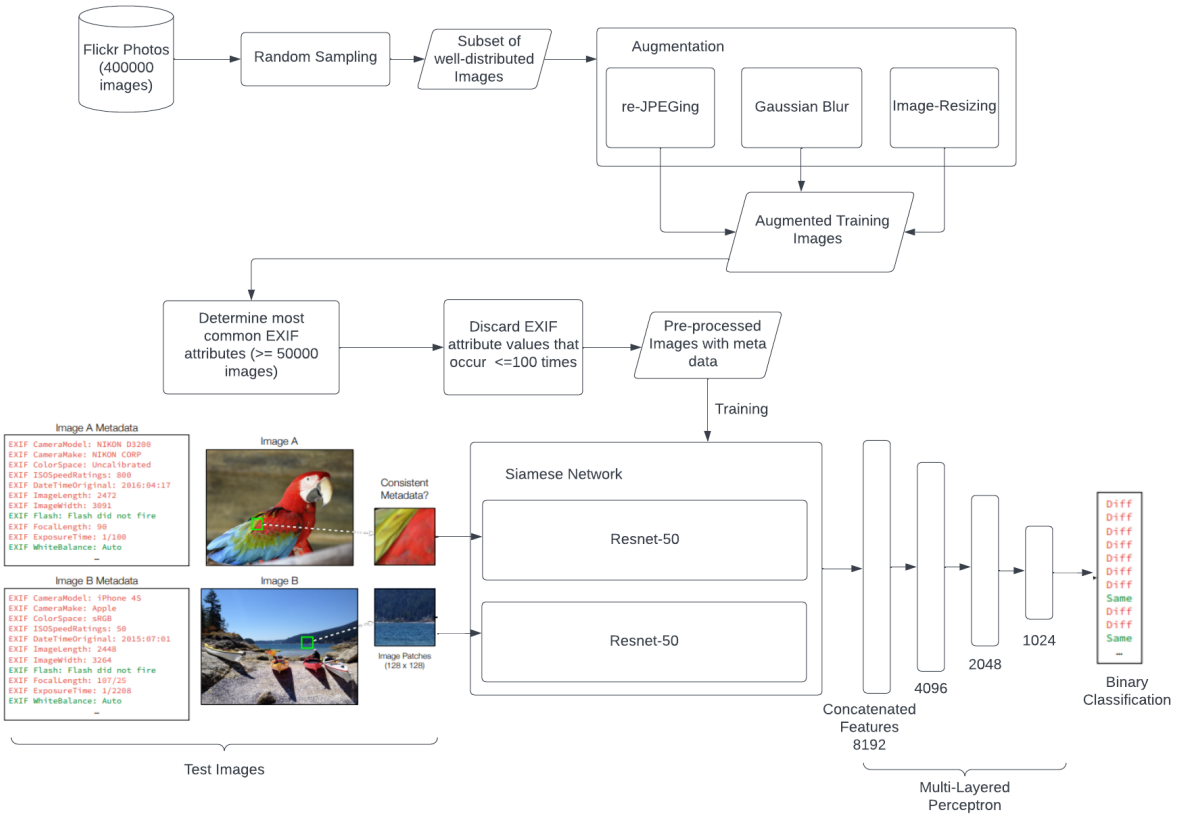
ARCHITECTURE DIAGRAM FOR CNN BASED FORGERY DETECTION



Architecture for Copy Move Detection - Unsupervised Domain Adaptation



PROPOSED ARCHITECTURE FOR IMAGE SPLICING DETECTION



MODULE DESCRIPTION

i) IMAGE PATCH EXTRACTOR (Common to both CNN and self-consistency learning)

INPUT: input_path, output_path, patches_per_image, no_of_rotations, stride
OUTPUT: Rotated image patches

- For each image in authentic and tampered class, a patch sized sliding window of size (128 x 128 x 3) is applied and the window is slides based on the stride value.
- Extract patches from the image by sliding the window, till the threshold patches has reached.
- Image are augmented by rotating 90, 180 and 270 degrees
- Save the patches of size 128 x 128 in a separate directory for both the classes

ii) CNN Model Training

INPUT: Augmented Image Patches
OUTPUT: Fused Feature Vector

● The individual patches are passed through a series of filters (convolution layers and max pooling) which enable feature extraction.
● At each successive layer, the filters increase in complexity and learn to detect more complex features.
● The output of every layer serves as an input to the next
● The N featured vector representation undergoes mean/max fusion leading to a fused vector, which is then taken up by the classifier.

iii) Classifier for CNN model

INPUT: Fused feature vector from CNN model
OUTPUT: Classification of the test image, whether it's tampered or not

● The fused feature representation from CNN is splitted into the training and testing dataset for the classifier.
● 80% of the data is used for the training phase, which is connected to a fully connected 2-way softmax and 20% of the data is used for testing, which is connected to the SVM classifier

iv) Exif attribute processor (self-consistency learning)

INPUT: A set of various exif attributes related to the metadata
OUTPUT: FIltered set with rarely occurring attributes removed

● The exif metadata is the basis for determining whether two patches correspond to the same image or not.
● As there are many attributes associated with exif metadata for an image, a list of the most common ones is created by considering those that occur in at least 50,000 images of the dataset.
● For these attributes, values that occur less than 100 times are removed/not to be considered for predictions.

v) Dataset Generation (Domain Adaptation for Copy Move Forgery)

INPUT: Coco Dataset
OUTPUT: Dataset with over 80,000 artificially tampered images

● The COCO dataset serves as a base for the generation of artificially tampered images using the methods of copy move and object removal/image inpainting forgery.
● Around 20,000 inpainted images are created, with 60,000 images through copy move.

● Semantic Inpainting helps the model to learn edge discrepancies when the objects are removed.
● Copy-Move tampered images improve the focus of the network to recognize similar patches.

vi) Base Feature Extraction, fc Layer Based Classification

INPUT: Images from both source and target domains. Around 80,000 pairs in total
OUTPUT: A binary result indicating whether a given image has been forged or not

● A method called Deep Domain Confusion (DDC) is used here.
● Using domain confusion loss, DDC learns the mapping of the source domain. It minimizes the distance between the source and target distributions via Maximum Mean Discrepancy (MMD) loss.
● The architecture separately learns the discriminative features needed to classify via supervised learning using source images and labels and features required to classify the domain of the image.
● The network aims to learn a representation that could easily be transferable across various domains.
●Images from both domains are passed through convolution layers before the fc layer aids in the classification of images as tampered or real.

ALGORITHMS/PSEUDO CODE FOR THE MODULES IMPLEMENTED

IMAGE PATCH EXTRACTOR

INPUT: input_path, output_path, patches_per_image, no_of_rotations, stride
OUTPUT: Rotated image patches

i) Start
ii) For each image in Dataset_Name/authentic
- Apply a patch-sized sliding window of size (128 x 128 x3)
- Slide the window by the stride value
- Extract patches from the image by sliding the window, till the threshold patches has reached.
- Rotate the patches by 90, 180 & 270 degrees.
- Save the patches of size 128 x 128 in Output/patches_with_rot/Authentic

iii) For each image in Dataset_Name/tampered

- Apply a patch-sized sliding window of size (128 x 128 x3)
- Slide the window by the stride value
- Extract patches from the image by sliding the window, till the threshold patches has reached.
- Rotate the patches by 90, 180 & 270 degrees.
- Save the patches of size 128 x 128 in Output/patches_with_rot/tampered

iv) Stop

IMPLEMENTATION :

- Exploratory dataset analysis of CASIA2 and COCO dataset.
- Augmenting the dataset with different augmenting techniques like Image rotation, image resizing, applying grayscale features and shifting the image.
- Implementation of the patch extractor module for both authentic and tampered images in CASIA2.
- Extracted patches of size 128 x 128 were saved for both authentic and tampered classes.

REFERENCES

[1] G. Muzaffer and G. Ulutas, "A new deep learning-based method to detection of copy-move forgery in digital images," 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT), 2019, pp. 1-4, doi: 10.1109/EBBT.2019.8741657.

[2] Sreelakshmy I J and J. Anver, "An improved method for copy-move forgery detection in digital forensic," 2016 Online International Conference on Green Engineering and Technologies (IC-GET), 2016, pp. 1-4, doi: 10.1109/GET.2016.7916684.

[3] M. N. Nazli and A. Y. A. Maghari, "Comparison between image forgery detection algorithms," 2017 8th International Conference on Information Technology (ICIT), 2017, pp. 442-445, doi: 10.1109/ICITECH.2017.8080040.

[4] K. M. Hosny, A. M. Mortda, M. M. Fouda and N. A. Lashin, "An Efficient CNN Model to Detect Copy-Move Image Forgery," in IEEE Access, vol. 10, pp. 48622-48632, 2022, doi: 10.1109/ACCESS.2022.3172273.

[5] M. Elmaci, A. N. Toprak and V. Aslantas, "A Comparative Study on the Detection of Image Forgery of Tampered Background or Foreground," 2021 9th International Symposium on Digital Forensics and Security (ISDFS), 2021, pp. 1-5, doi: 10.1109/ISDFS52919.2021.9486363.

[6] R. Abecidan, V. Itier, J. Boulanger and P. Bas, "Unsupervised JPEG Domain Adaptation for Practical Digital Image Forensics," 2021 IEEE International Workshop on Information Forensics and Security (WIFS), 2021, pp. 1-6, doi: 10.1109/WIFS53200.2021.9648397.

[7] T. Zhao, X. Xu, M. Xu, H. Ding, Y. Xiong and W. Xia, "Learning Self-Consistency for Deepfake Detection," 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 15003-15013, doi: 10.1109/ICCV48922.2021.01475.

[8] R. Li, W. Cao, S. Wu and H. -S. Wong, "Generating Target Image-Label Pairs for Unsupervised Domain Adaptation," in IEEE Transactions on Image Processing, vol. 29, pp. 7997-8011, 2020, doi: 10.1109/TIP.2020.3009853