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

Guide: Dr. K. Indra Gandhi

Aravind J 2019115017
Krishnan S 2019115047
Pranay Varma 2019115067

INTRODUCTION

- 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.

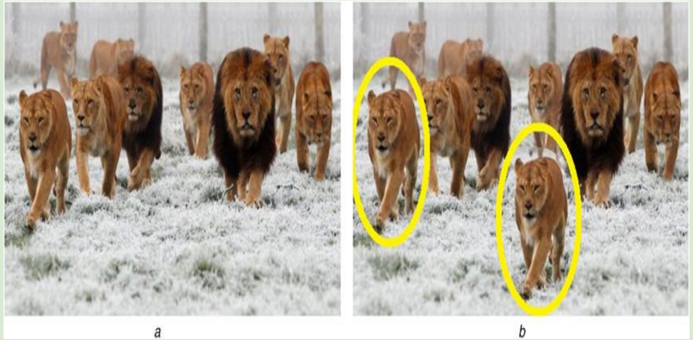
IMAGE TAMPERING METHODS

Three of the most common image manipulation techniques:

- **Splicing:** A region from an authentic image is copied into a different image.
- **Image Inpainting:** An image region is removed and the removed part is then filled in to complete the image.
- **Copy-move:** A specific region from the image is copy pasted within the same image.

IMAGE TAMPERING METHODS - Examples

Copy and Move



Splicing

Image Inpainting

PROBLEM STATEMENT & 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.

Project Domain

Related works and Constraints

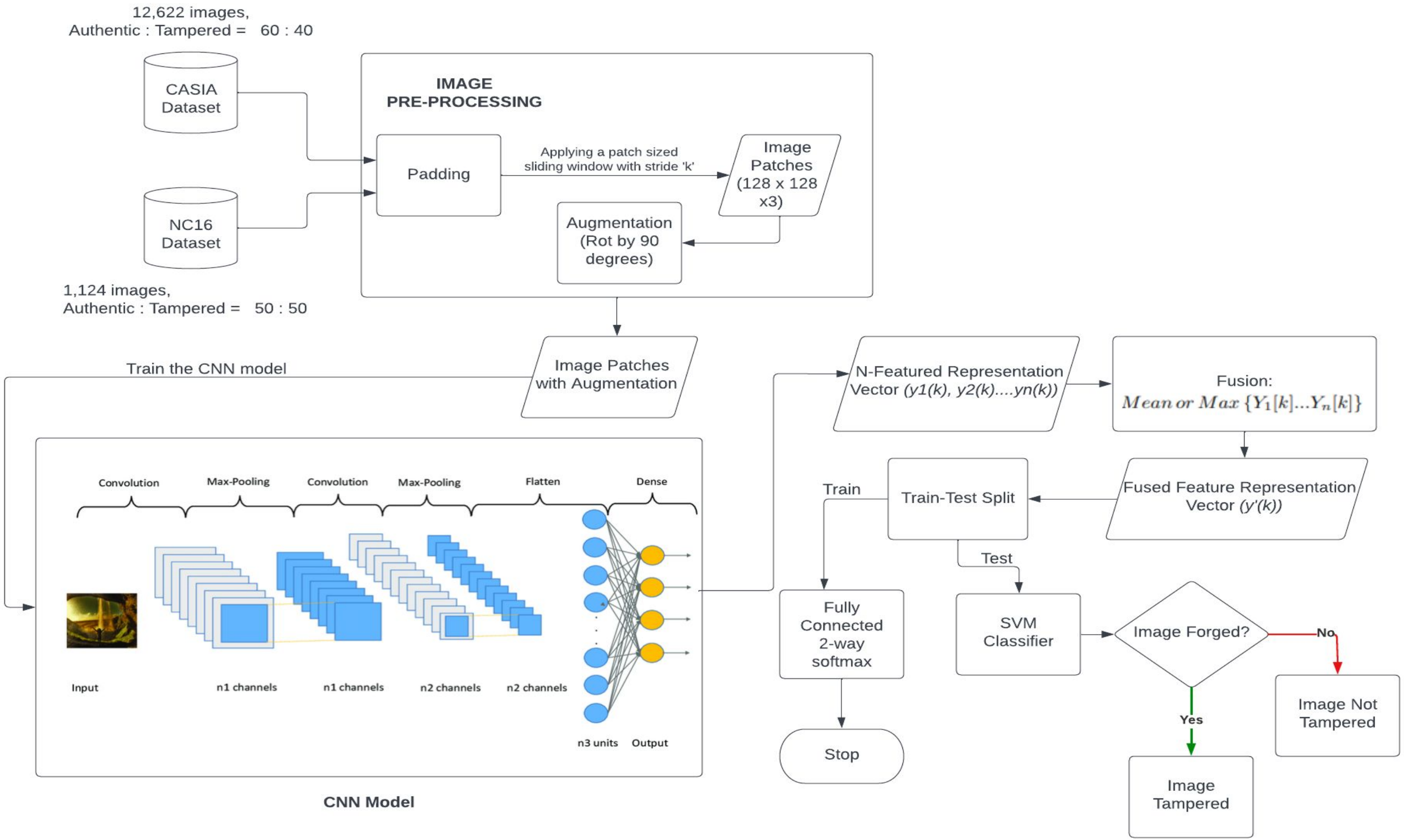
SNo	Title of the Paper	Year	Methodologies/Approach used	Pros	Cons
1	An Efficient CNN Model to Detect Copy-Move Image Forgery	2022	Using CNN feature extraction is done followed by max-pooling layer and then the classification stage is called to classify data.	The proposed architecture is computationally lightweight	The accuracy of classification decreases when the samples are challenging.
2	Copy Move and Splicing Image Forgery Detection using CNN	2022	Pre-processing and then error analysis and using CNN to predict output	More time efficient	The model does not easily generalize to datasets with different underlying distributions.
3	Forgery Classification via Unsupervised Domain Adaptation	2020	Generating more dataset images using image inpainting and copy and move and then using it to train the model.	Since dataset other than publicly available dataset is used, it will have high accuracy in realistic test data.	More processing power is needed.

SNo	Title of the Paper	Year	Methodologies/Approach used	Pros	Cons
4	Fighting Fake News: Image Splice Detection via Learned Self-Consistency	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.	The proposed method obtains state-of the-art performance on several image forensics benchmarks, despite never seeing any manipulated images at training.	i) The model is not well-suited to finding very small splices. ii) Over- and underexposed regions are sometimes flagged by the model to be inconsistent because they lack any meta-data signal.
5	A Deep Learning Approach to Detection of Splicing and Copy-Move Forgeries in Images	2016	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.	Outperforms many state of the art models, in terms of speed and accuracy	-

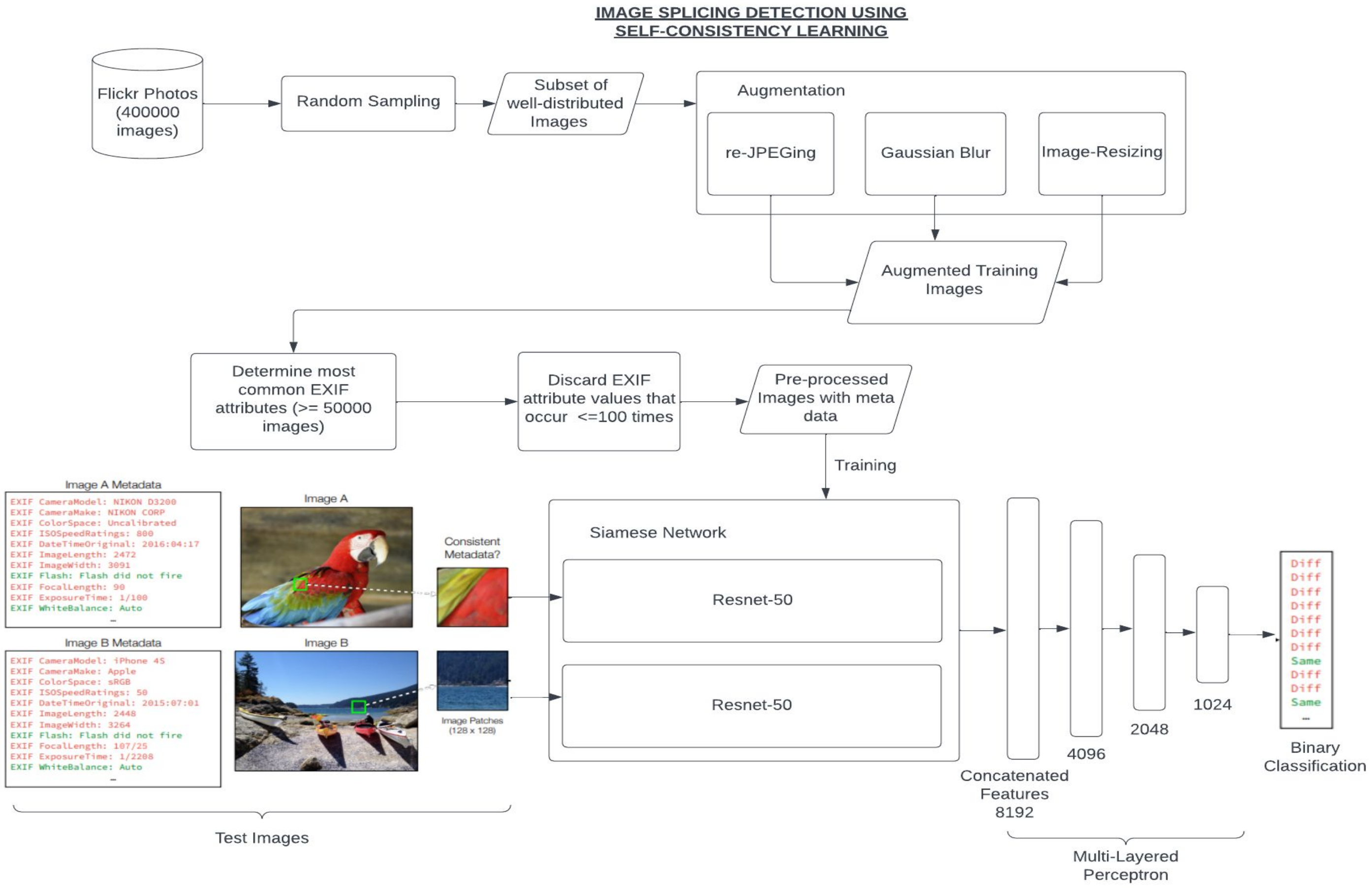
DATA SET

- **CASIA V2 Dataset:** CASIA V2 is a dataset for **forgery classification**. It contains **12,616** images among which **7492** are authentic and **5124** are forged. Tampering done in this dataset is easier to recognize by humans.
- **Media Forensics Challenge Dataset (NC16):** The images in this dataset are significantly more difficult to recognize. Contains **1,124 images** with a **50-50** distribution.
- **Common Objects in Context(COCO):** It contains **328,000** images of everyday objects and humans. The dataset contains annotations you can use to train machine learning models to recognize, label, and describe objects.
- **Copy-move forgery detection(CoMoFoD):** It contains **260** forged image sets in two categories (small **512x512**, and large **3000x2000**). Images are grouped in 5 groups according to applied manipulation: translation, rotation, scaling, combination and distortion.

ARCHITECTURE DIAGRAM FOR CNN BASED FORGERY DETECTION

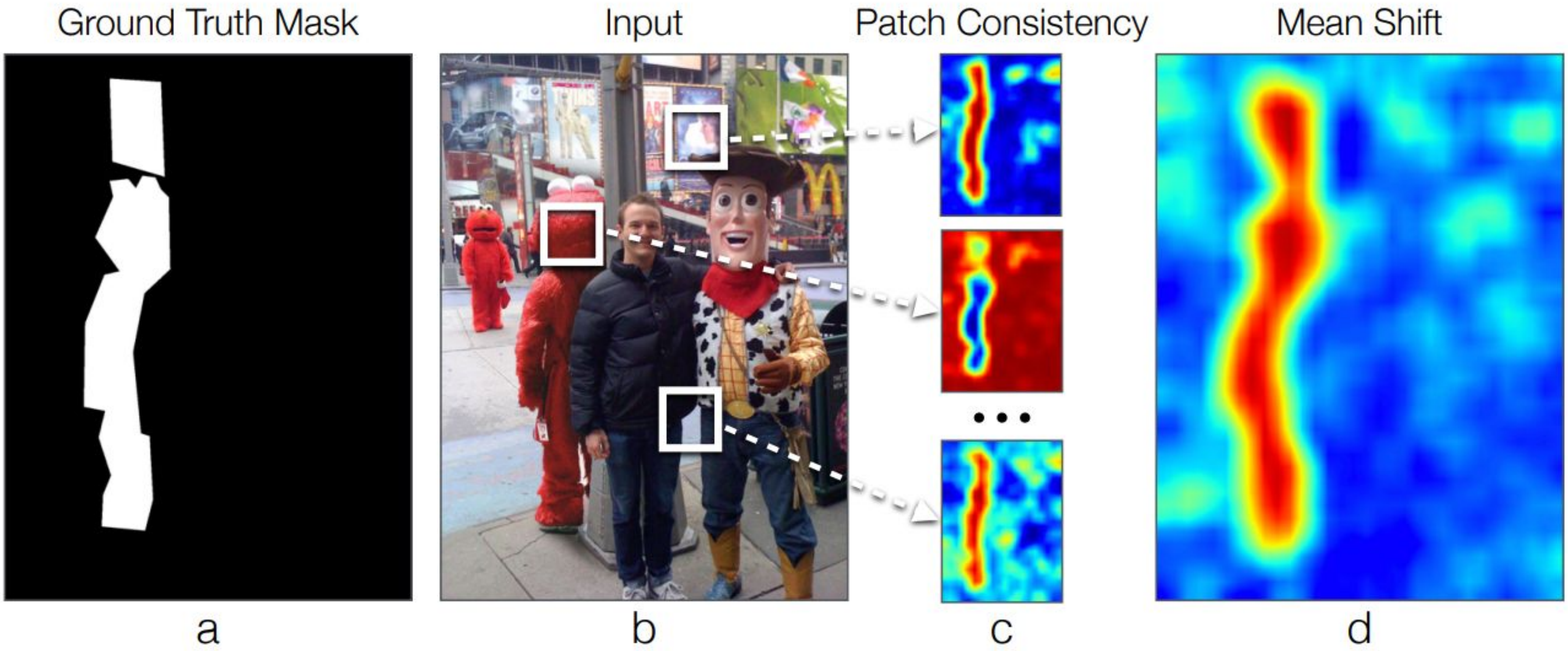


PROPOSED ARCHITECTURE FOR IMAGE SPLICING DETECTION

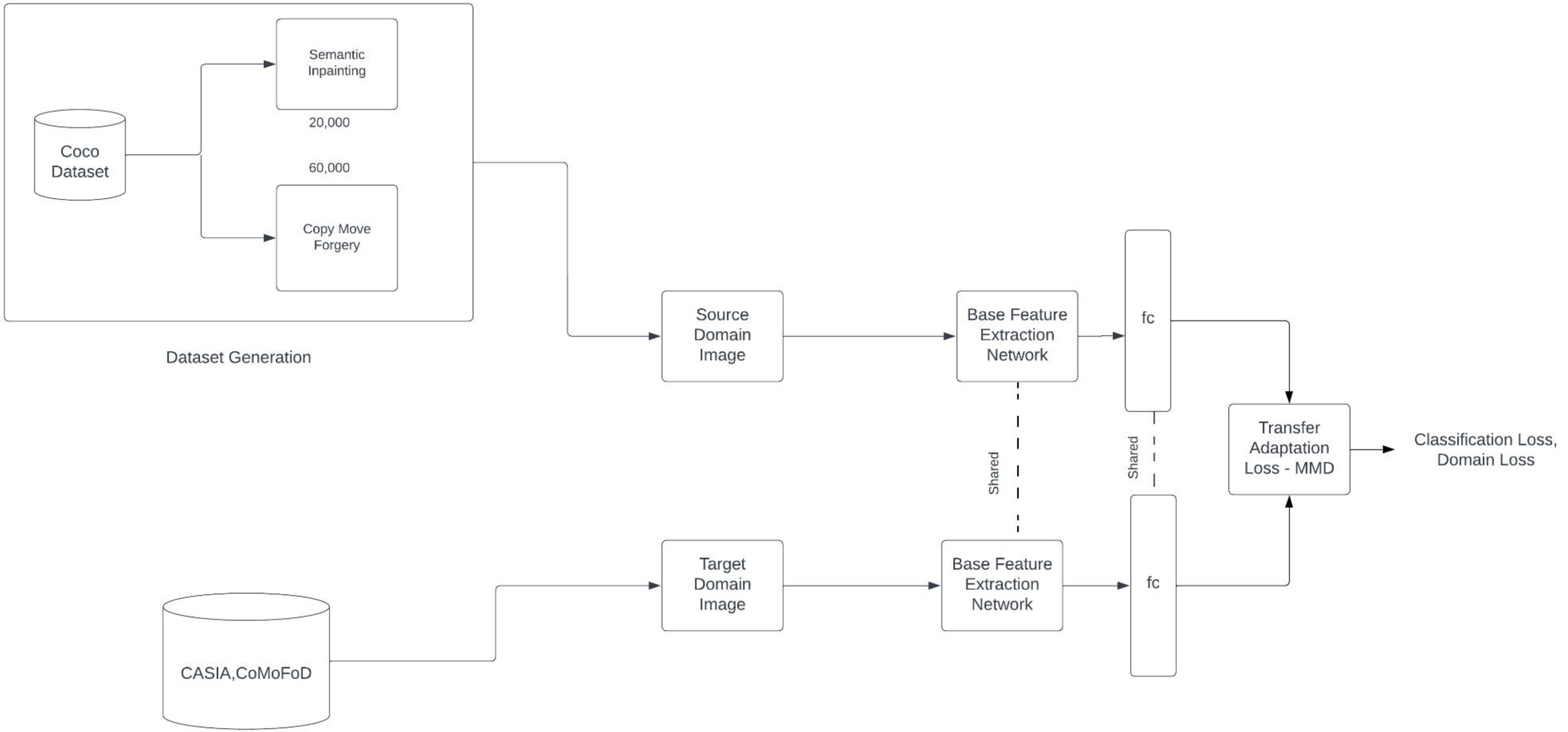


EXIF Attribute comparison for a spliced image

Consistency Matrix



Architecture for Copy Move Detection - Unsupervised Domain Adaptation



MODULE DESCRIPTION

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.

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.

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.

Exif attribute processor (self-consistency learning)

INPUT: A set of image patches from patch extractor
OUTPUT: Filtered set with rarely occurring attributes removed

- Exif metadata is extracted from the image patches
- 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.

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.

Base Feature Extraction, fc Layer Based Classification

INPUT : Images from both source and target domains
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.

TOOLS AND LIBRARIES

colab

Keras

TensorFlow

PyTorch

Streamlit

scikit learn

Implementation so far

- Exploratory dataset analysis of **CASIA2** and **COCO** dataset

- Augmenting the dataset with different augmenting techniques like **Image rotation, image resizing, applying grayscale features** and s**hifting the image**.

- Implementation of the patch extractor module for both authentic and tampered images in CASIA2.

- Extracted patches of size **128 x 128** saved for both authentic and tampered classes.

EXPECTED DELIVERABLES

- A web app developed using **streamlit/Flask** where users can upload an image to detect image forgery.
- Detection of 2 types of image forgeries - **Copy-Move** and **Splicing**
- Map the regions where the image is tampered.
- Analysis of the performance of the different deep-learning models on varied test-sample difficulty.

REFERENCES

- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.

THANK YOU!