NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

School of Computer Science and Engineering
College of Engineering

Forgery Localization in Images

Using Machine Learning Techniques

Student: Nur Dilah Binte Zaini Supervisor: Assoc. Prof Deepu Rajan

Project Objectives:

This project aims to build a general low-dimensional feature set which is utilized to detect tampered regions in the post JPEG-compressed images. The regions in images are tampered with information-preserving operations to conceal artifacts left by information-changing operations. Forgeries in images have been evolved to deepfakes in the recent years. Based on the pool of selected features, specific features are then chosen with additional features to identify post JPEG-compressed deepfake images.

Image Forgery Detection

Methodology

- Convert tampered image in spatial domain to residual domain (MFR)
- Division of MFR into image blocks
- Utilize joint statistics of image blocks with neighboring blocks to build **18-dimensional universal feature set**
- Train an ensemble of decision tree classifiers with a bagging model using the feature set
- Detect tampered image blocks of post-JPEG compressed images using proposed detector

Results

Experiments are performed for different sizes of blocks

Table 1: Average accuracy and F1-scores for different sizes of image blocks with quality factor of 70 & 30

Size of non-overlapping blocks	*Q*	Average Accuracy	Average F1-score
128 x 128 **(1000 image blocks)**	70	90.8	90.2
64 x 64 **(10000 image blocks)**		79.1	81.2
32 x 32 **(10000 image blocks)**		74.1	80.6
128 x 128 **(1000 image blocks)**	30	73.6	79.2
64 x 64 **(10000 image blocks)**		73.5	70.2
32 x 32 **(10000 image blocks)**		73.8	66.6

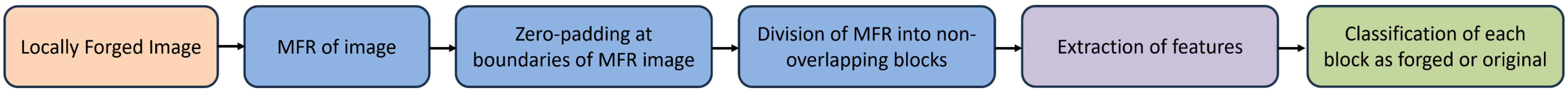
Image Datasets

BOWS2

BOSSBase

IEEE IFS-TC DB

Public DB



Locally Forged Image → MFR of image → Zero-padding at boundaries of MFR image → Division of MFR into non-overlapping blocks → Extraction of features → Classification of each block as forged or original

Figure 1: Overall workflow of proposed image forgery detection

DeepFake Detection

Methodology

- Extract features from MFR image patches of each facial part to build each **7-dimensional feature set**
- Each feature set is used to train individual random forest classifiers to produce the confidence scores
- The confidence scores are used to train a Decision Tree classifier to predict class of the image (late fusion)

Results

Table 2: Accuracy scores and F1-scores for 3 ML models

Model	Classification Accuracy Score	F1 - score
Decision Tree Classifier	0.61	0.69

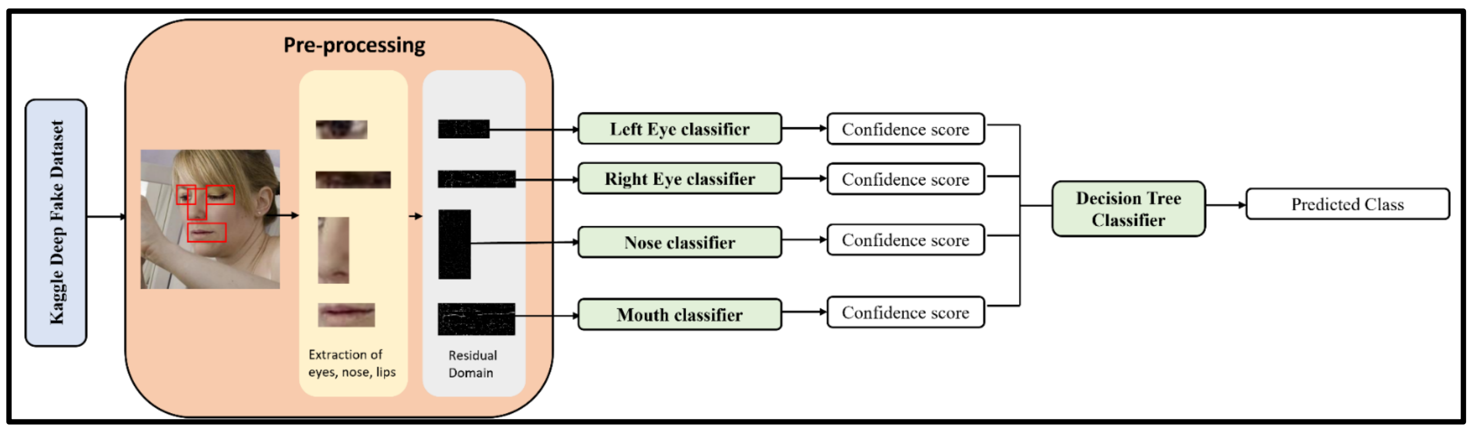


Figure 2: Overall workflow of proposed deepfake detection

Image Dataset

Kaggle DeepFake Dataset