NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

School of Computer Science and Engineering
College of Engineering

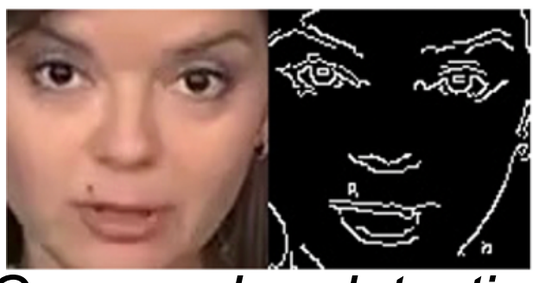
Classification of Deepfakes and 'Beautified' Media

Student: Addi Debashree Supervisor: Liu Yang

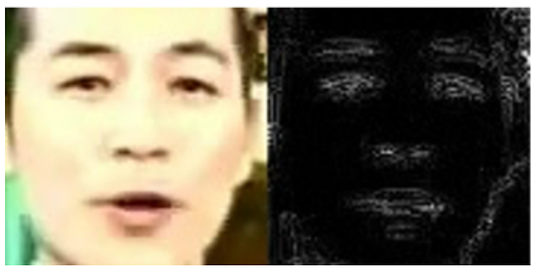
Problem statement: Many deepfake detection models exist which detect fake and real digital media. With beautified media becoming more common, there is a need to develop a model which detects these digitally manipulated media accurately into real, fake and 'beautified' categories.

Real, fake and beautified cropped face images from FaceForensics++ dataset

Project Objective: The objective of this paper is to present experiments using the CNN model: Xception with a large dataset of labelled 'real', 'fake' and 'beautified' videos to give a highly accurate classification result.



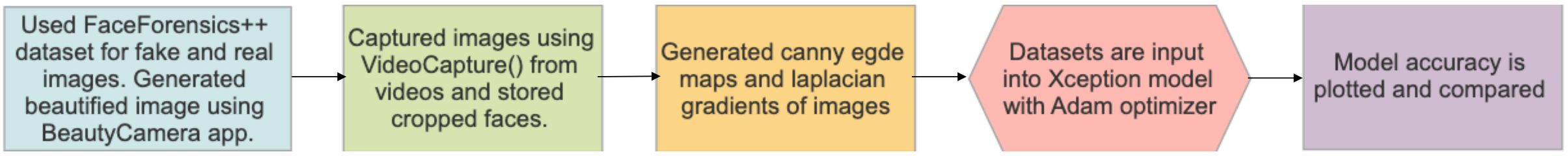
Canny edge detection



Laplacian gradient

Data pre-processing: Following extraction of cropped face images, Canny edge algorithm and Laplacian gradient was used for edge detection.

Pipeline:



Results: Best model accuracy of ~89% is given by Xception CNN trained with dataset of images concatenated with their Laplacian gradient. Confusion matrix is plotted with 100 images from each category as test dataset. It shows model performs best for beautified dataset and worst for fake dataset.

Dataset preprocessing	Accuracy
None (only image dataset)	0.8573
Edge map (image concatenated with its edge map)	0.8077
Laplacian (image concatenated with its Laplacian gradient)	0.8962