A Novel Machine Learning based Method for Deepfake Video Detection in Social Media

Presenter : Alakananda Mitra

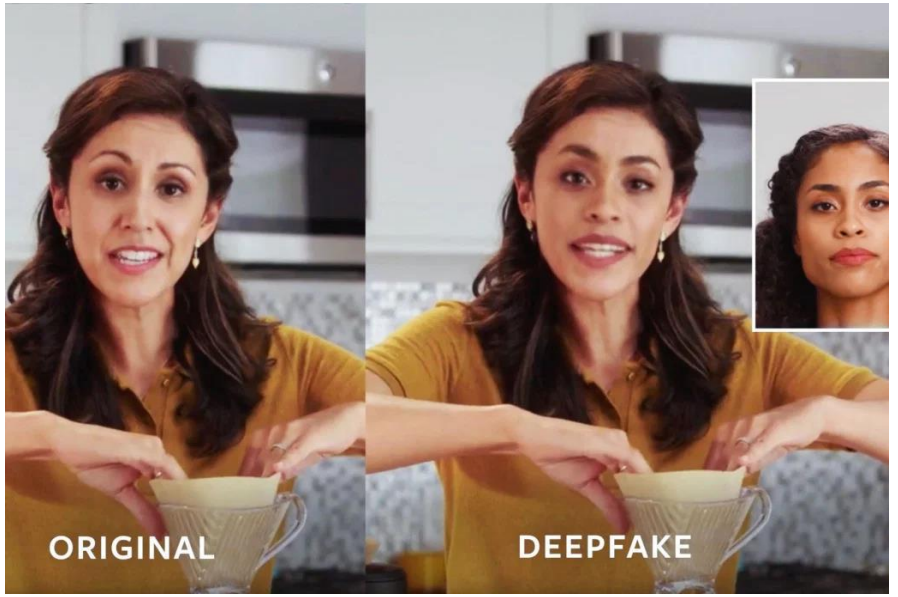
A. Mitra[1], S. P. Mohanty[2], P. Corcoran[3], and E. Kougianos[4]

University of North Texas, Denton, TX , USA.[1,2,4] and National University of Ireland, Galway, Ireland[3].

Email: AlakanandaMitra@my.unt.edu[1], saraju.mohanty@unt.edu[2], peter.corcoran@nuigalway.ie[3], and elias.kougianos@unt.edu[4]

Outline

- Introduction
- Motivation
- Deepfake Video Creation
- Existing Works in Detecting Deepfake Video & Issues
- Proposed Solution
- Results
- Conclusions & Future Work



source: Facebook

Introduction

Source: FaceForensics++

Motivation

- Sophisticated in nature

- DeepFake = Deep Learning + Fake
  - Funny
  - Threat to individual's identity, reputation & national security

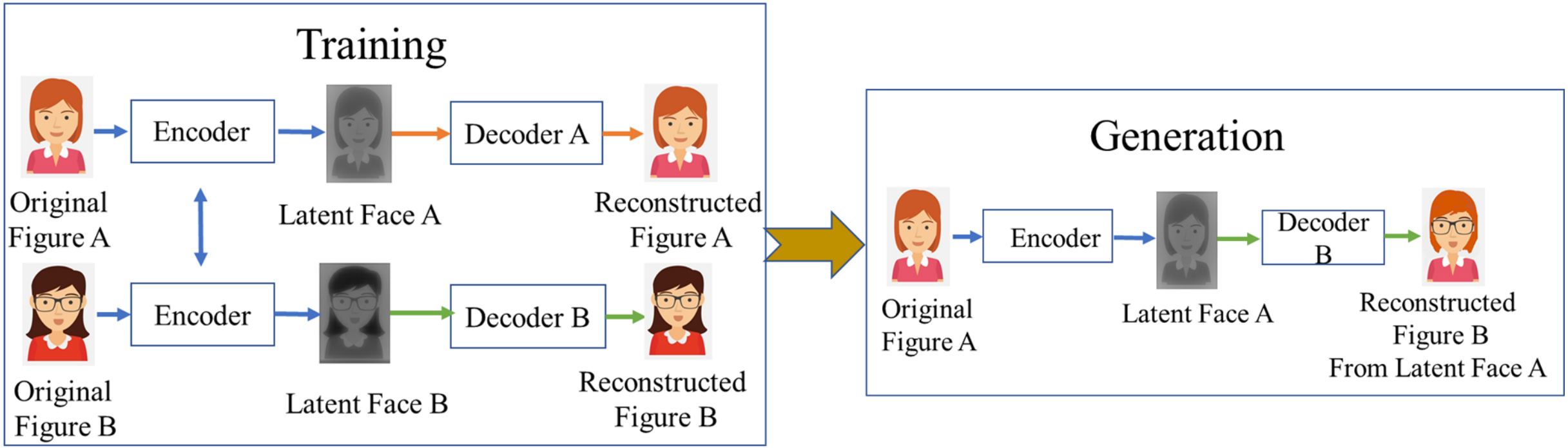
image: Freepik.com

- Social Media in People's lives
- Social Media Uploaded Videos
  - Compressed
  - Higher Compression → Higher Loss → Chances of Detection Low

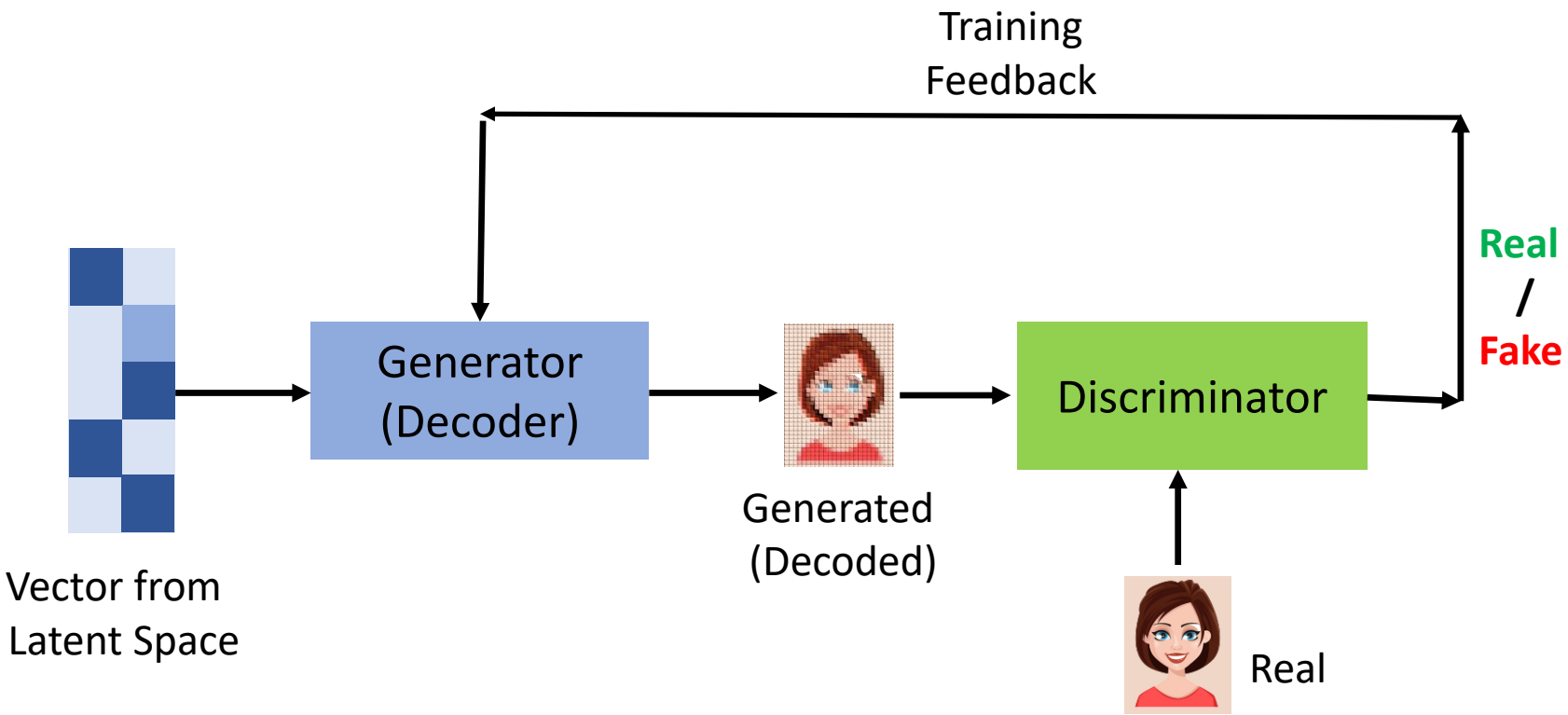
- OUR GOAL :

  A Model to Detect highly compressed videos in Social media

Deepfake Training & Generation (Autoencoders)



Generative Adversarial Network (GAN)



Related Works

Works	Dataset	Model Features	Remarks
Sabir, et al. [2019]	FaceForensics++	Bidirectional LSTM+DenseNet/ ResNet50	Classifier network complex.
Guera and Delp [2018]	HOHA	InceptionV3 + LSTM	No compression
Li, et al. [2018]	CEW	VGG16+LSTM+FC	Uncompressed
Afchur, et al. [2018]	From Internet	Mesonet structure	Less accuracy for high compression.
Nguyen, et al. [2019]	Four major datasets	VGG19+ Capsule Network	Less accuracy for high compression.
Matern, et al. [2019]	A combination of various sources.	Facial texture difference mostly eye and teeth. Logistic Regression.	Not for compressed video.

Issues in Existing Work & Solution

- ❑ Few Works on Compressed Videos
- ❑ For Highly Compressed Videos Accuracy is low
- ❑ Complex Solution

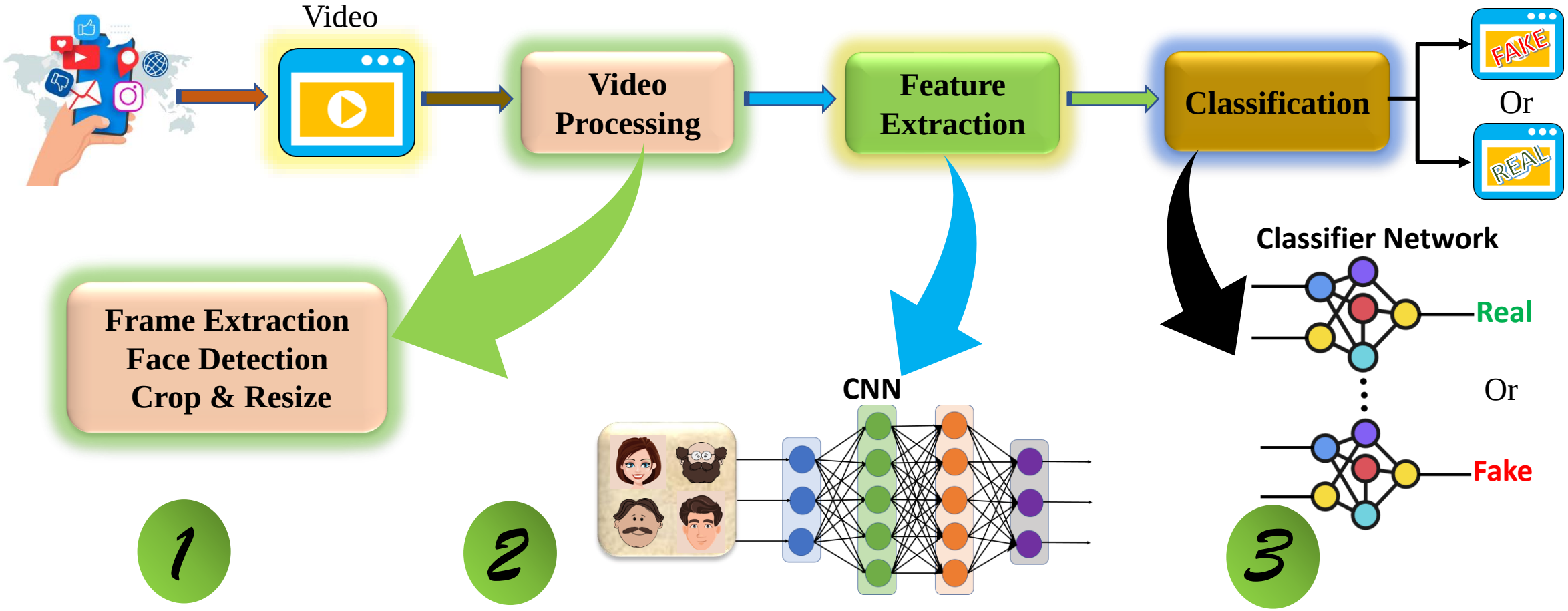
- ➢ **Needed:**
  - ✓ Simple Solution for Compressed Videos
  - ✓ Less Computation

Addressed Research Question

Q. How to Detect Compressed Deepfake videos in Social Media?

- Challenges :
  - Compression causes Losses.
  - Fake attributes might be lost.
  - Detection might not be accurate
- Assumption:
  - Auto-encoder generated Deepfake videos

Proposed Method Overview



Method Details

- Transfer Learning
  - Feature Extractor → ResNet50, InceptionV3 & Xception – Pretrained on Imagenet (1000 Classes of 1,000,000 images)
    - Already Learned : General Features of Images
      - ✓ Edges
      - ✓ Corners
      - ✓ Colors
  - Training Time Saved
  - Accuracy Better

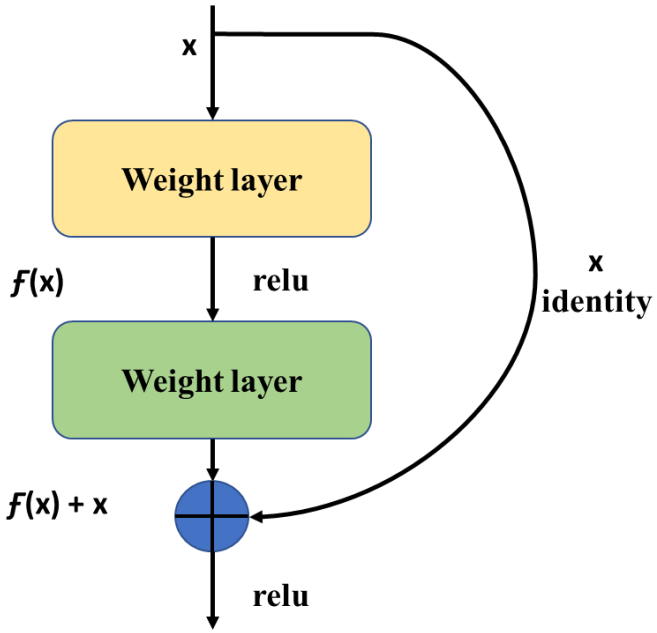
Training Strategy

- Replace the Last Layer of CNN
- Add Classifier Network
- Freeze the Feature Extractor
- Train Classifier Network with FF++ data
  - Last Layers Learn :
    - ✓ Fake & Real Features
- Fine Tune the whole Network End-to-End

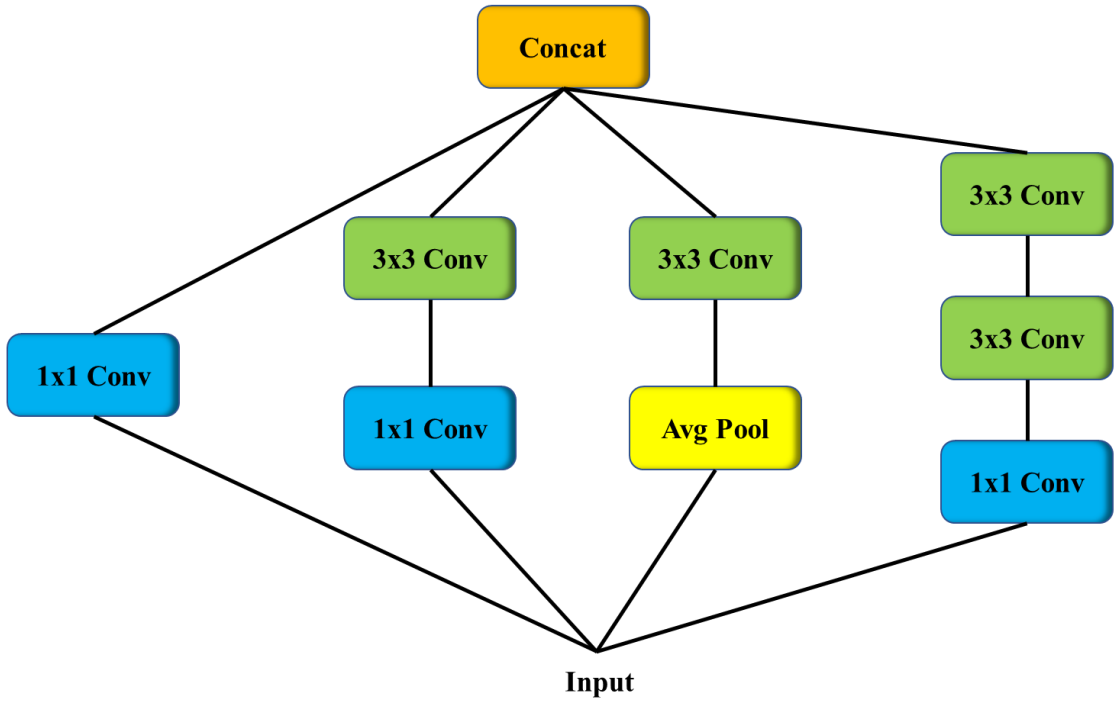
- Face Detected & Cropped
  - ✓ CNN learned Better Fake/Real Features

CNNs as Feature Extractor

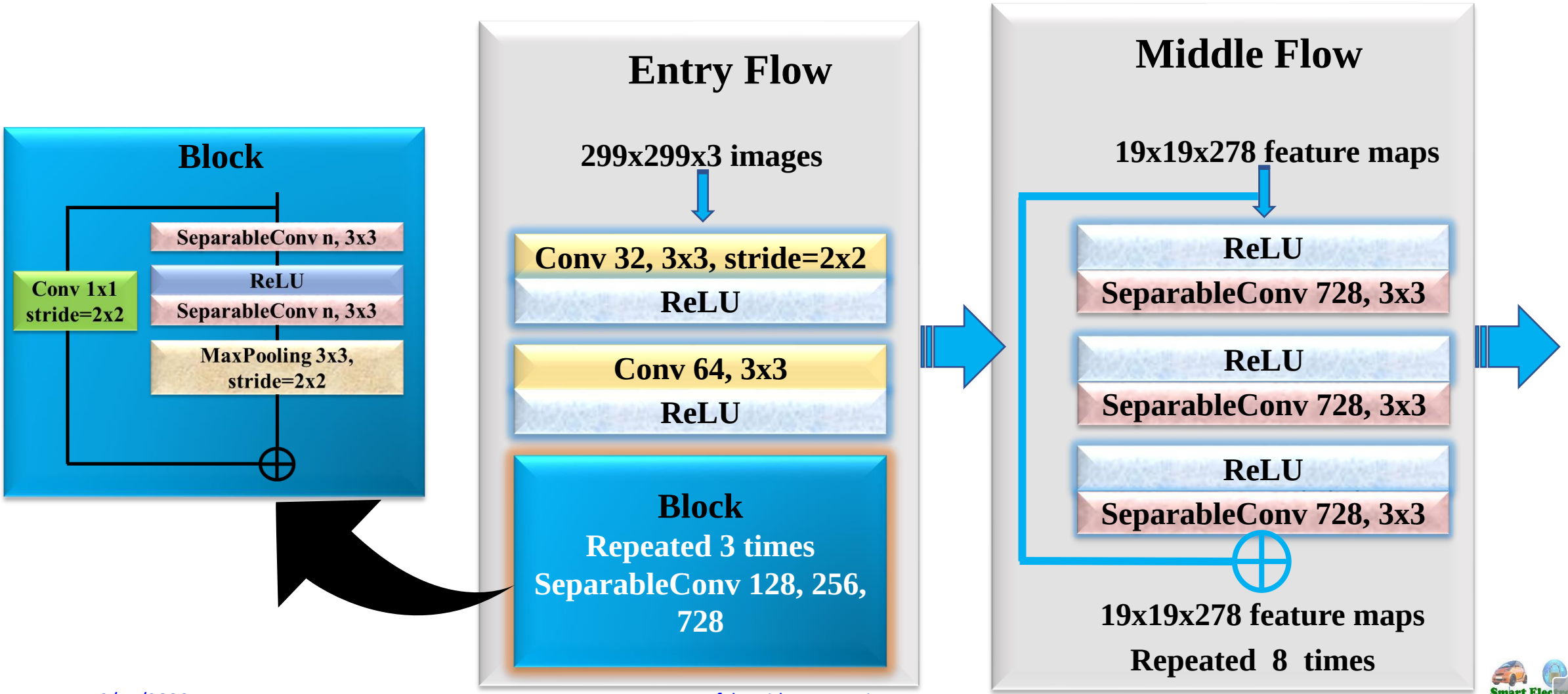
ResNet50



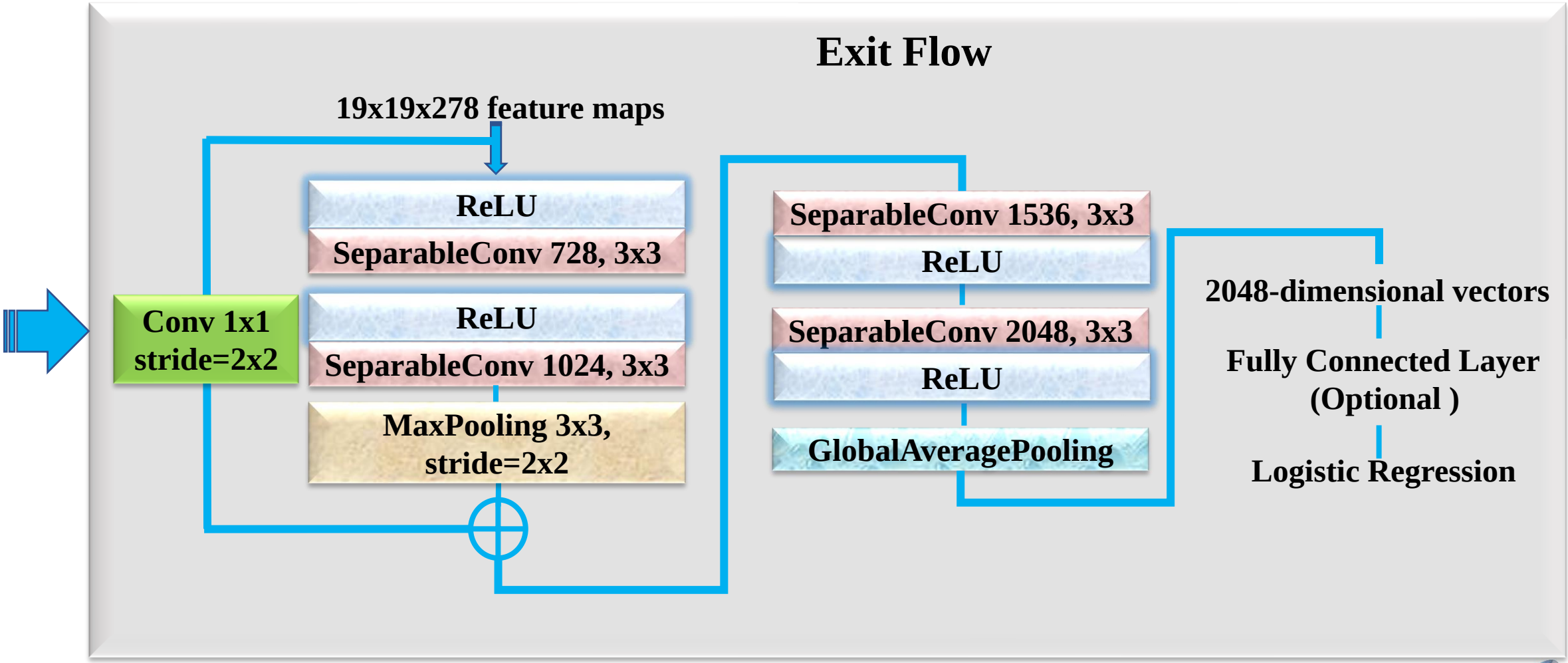
Inception Module from InceptionV3



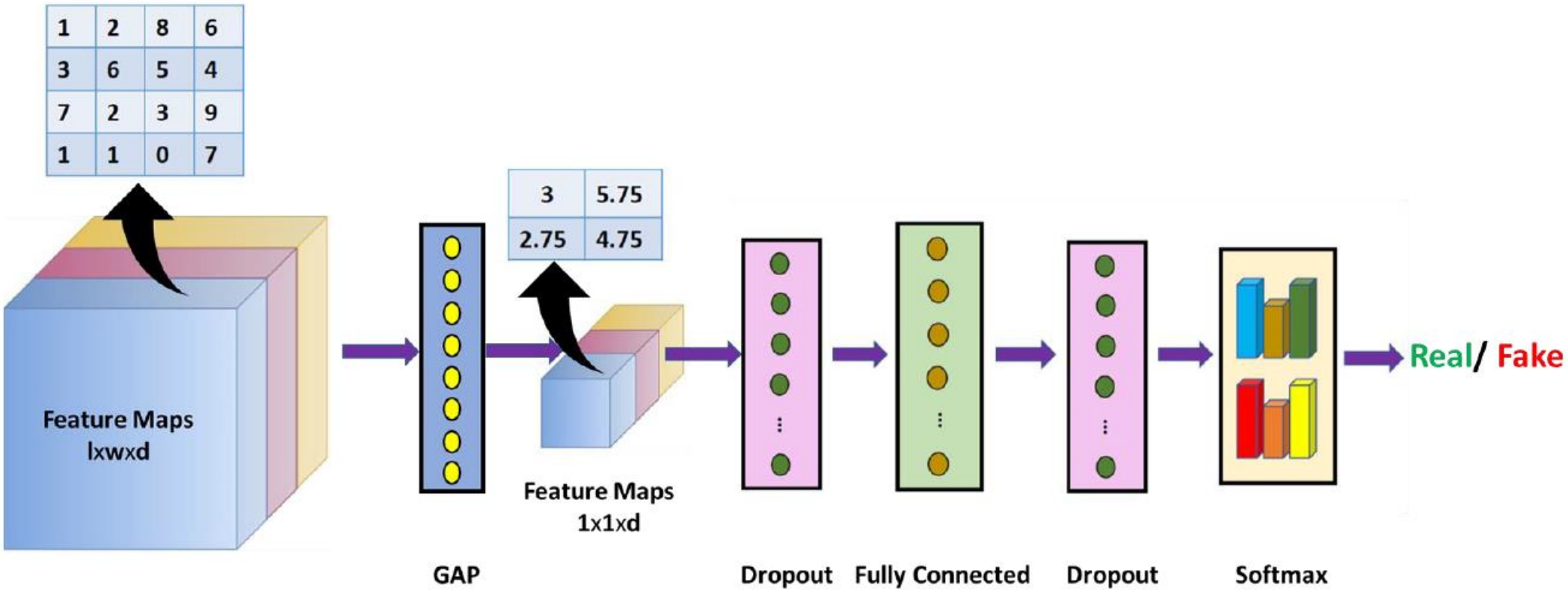
Final Feature Extractor- Xception Net (Contd..)



Final Feature Extractor- Xception Net



Classifier Network



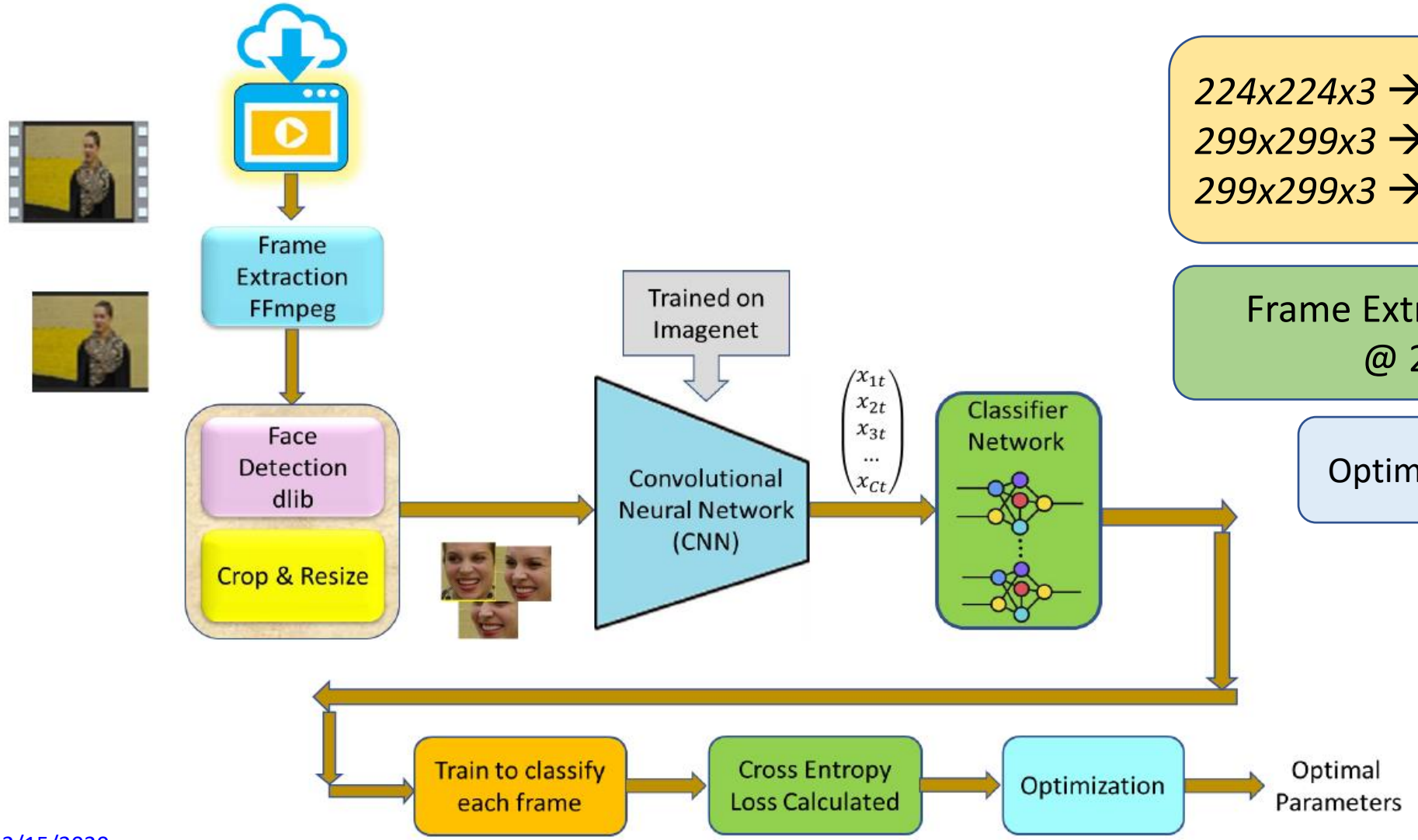
Dataset

- FaceForensics++
  - Deepfake Videos – 1000 Videos
  - Unaltered Videos – 1000 Videos

- Dataset Division :
    - Train : Valid – 80:20
    - Test – 100 Videos

- Compression Level for Training : c=23
  for Testing : c=23 & c=40

Training

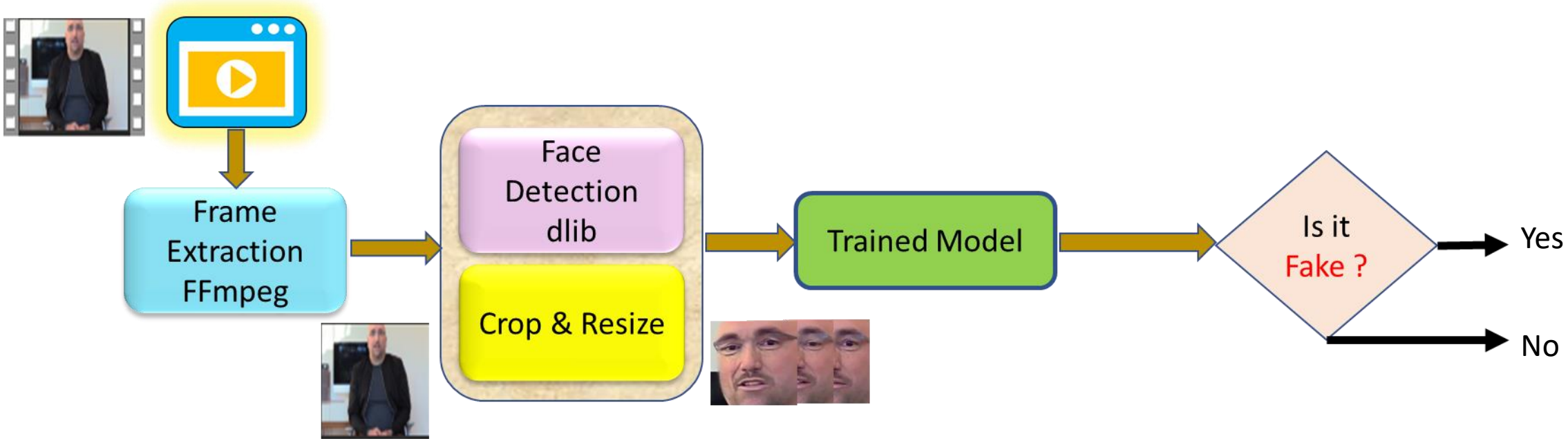


224x224x3 → ResNet50
299x299x3 → InceptionV3
299x299x3 → Xception Net

Frame Extraction Rate @ 24fps

Optimizer Adam

Detecting A Video



Algorithm 1: How to Detect Deepfake Video?

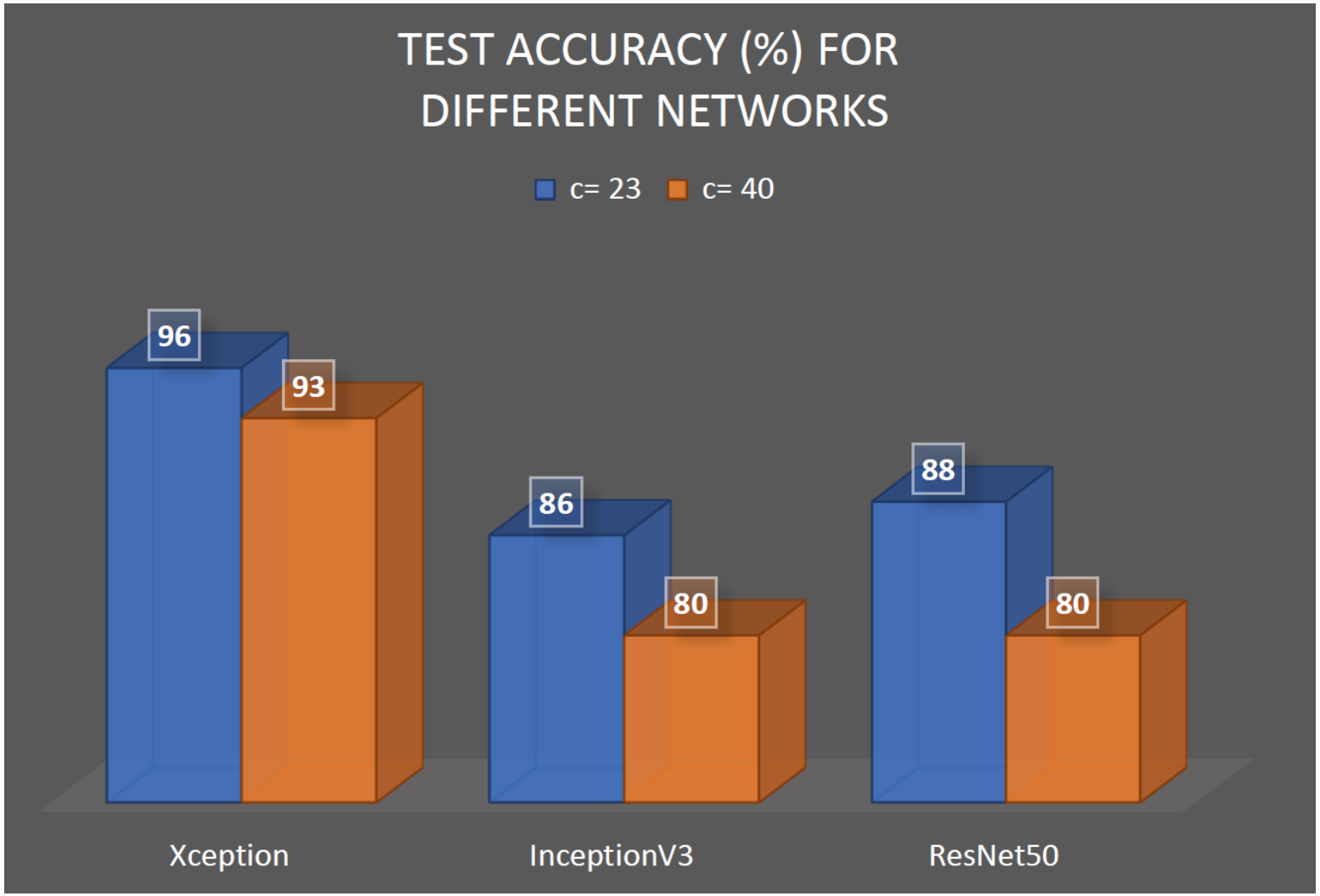
1. Extract frames from the test video
2. Store frames in a data frame
3. Detect and Crop face from each frame
4. Resize each image
5. Save them in another data frame
6. Load the saved model
7. **for** an image in the range of resized images **do**
8. Check for authenticity
9. **if** image is real **then**
10. **continue**
11. **else**
12. Consider the video Fake
13. **break**

Time Complexity **O**(n) small

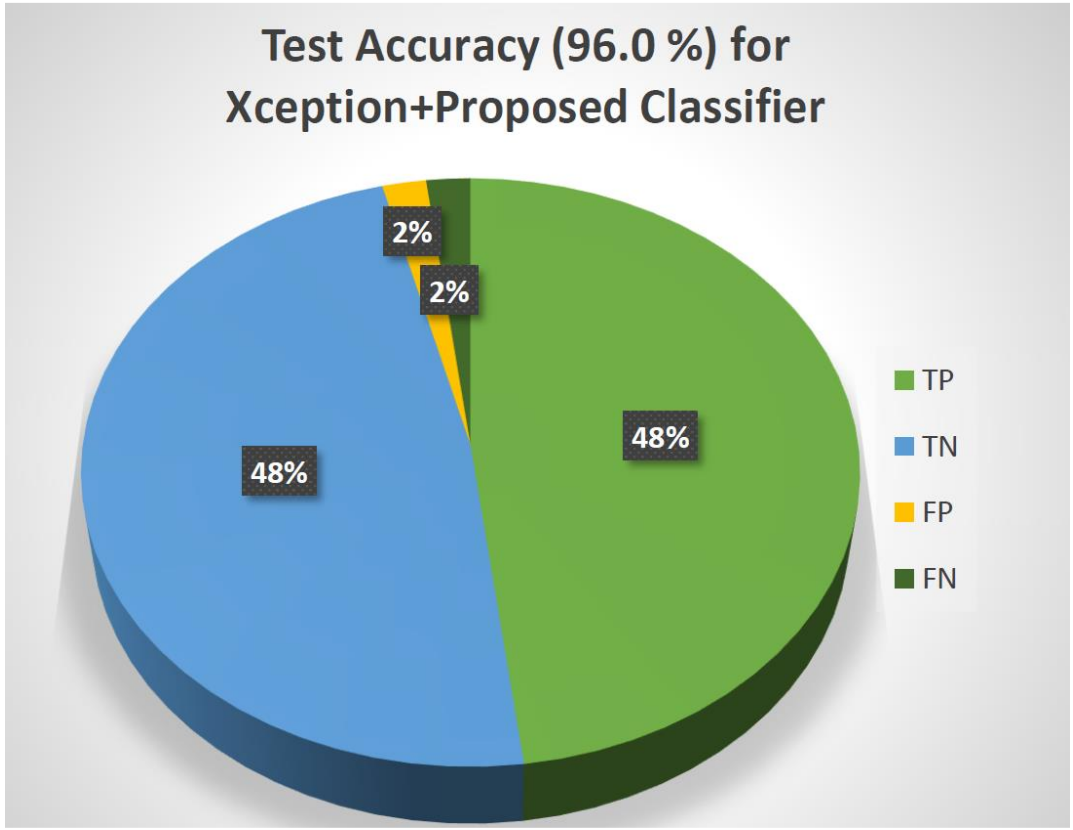
where n is number of frames

Best Case → **First frame FAKE**
Worst Case → **REAL / Last frame FAKE**

Results


TEST ACCURACY (%) FOR DIFFERENT NETWORKS
c= 23
c= 40
96
93
86
80
88
80
Xception
InceptionV3
ResNet50

Test Accuracy

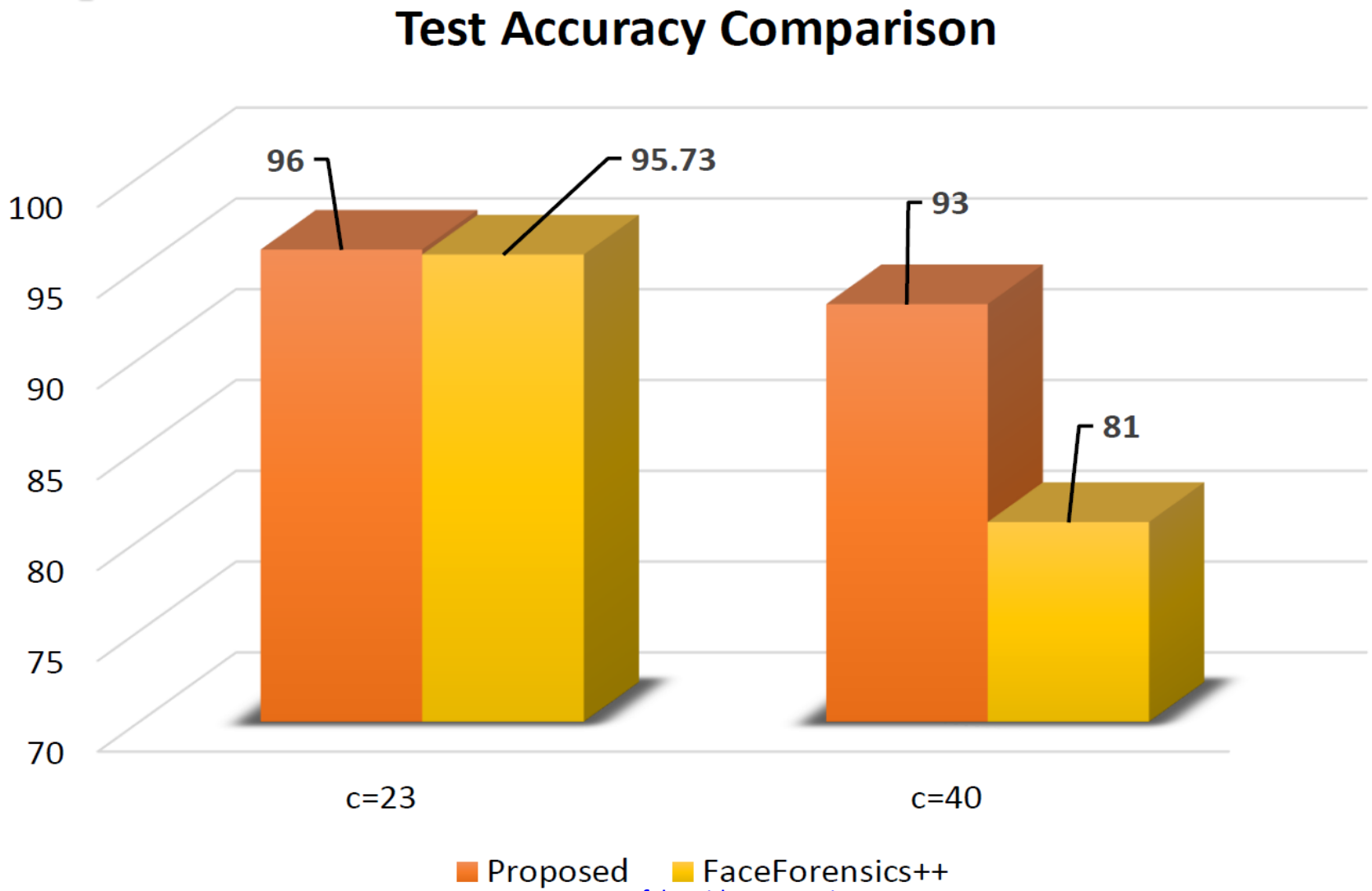


$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

Accuracy = 96% for c=23

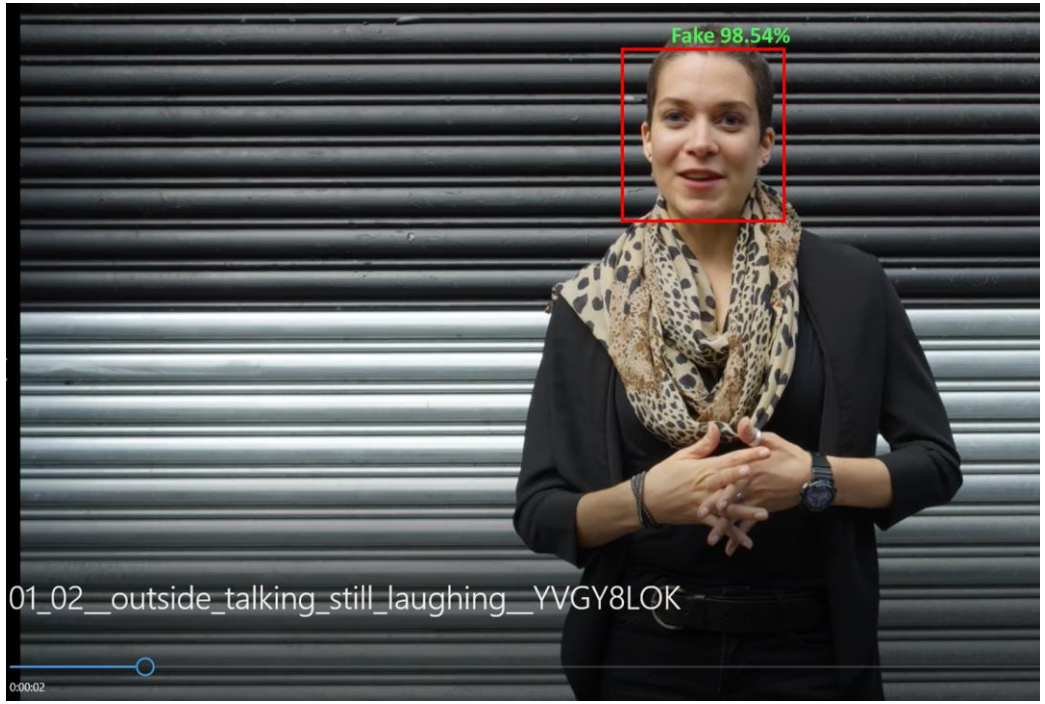
Accuracy = 93% for c=40

Comparisons with Related Work



Experimental Results-1

Fake Video

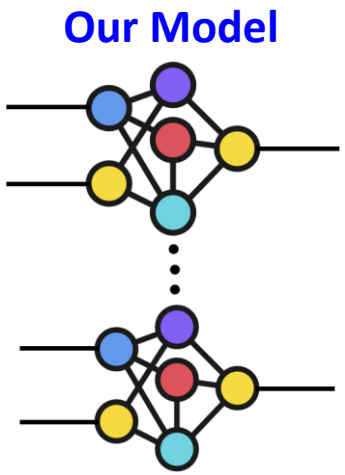


Real Video

Experimental Results-2

Test Video

Fake Video



Conclusions & Future Work

- Our proposed method detects Deepfake Video in Social Media
  - Neural Network-based Method
    - Xception Net as Feature Extractor
    - Simple Classifier Network
  - High Accuracy
  - Algorithm with Less Computation
- Deepfake Video Detection using Key video Frame
- Achieved Higher Accuracy & Lesser Computation
- Detection of Deepfake Images created by GAN

Thank you !