Deep Fake Detection

INTRODUCTION

What is **Deepfake**?

How was it made?

- **Origin**
- **Deep Learning**
- **positives?**

How do they work?

Step 1: extract Brie face

shared encoder → Carrey decoder

Step 3: insert fake Carrey face

Step 2: create fake Carrey face
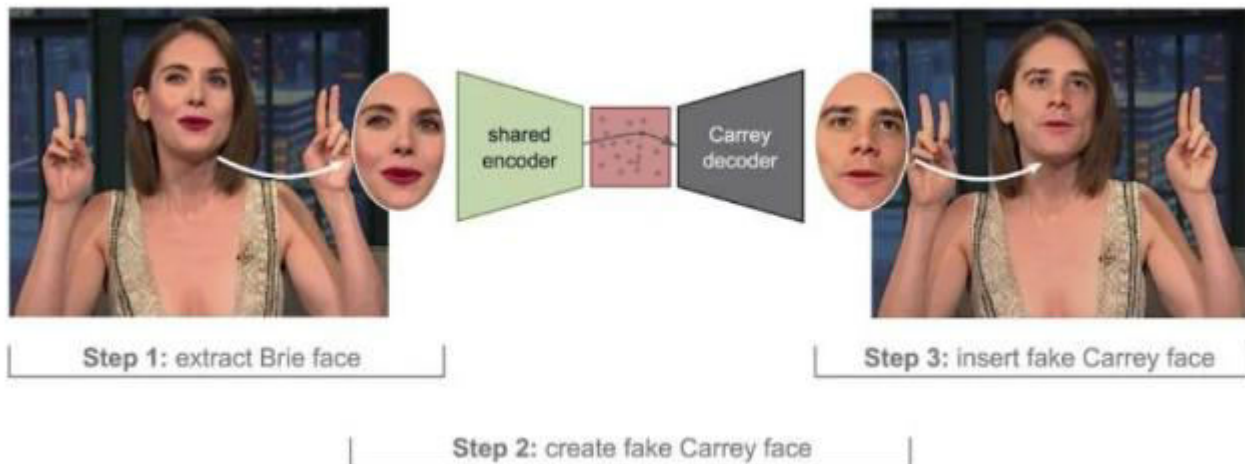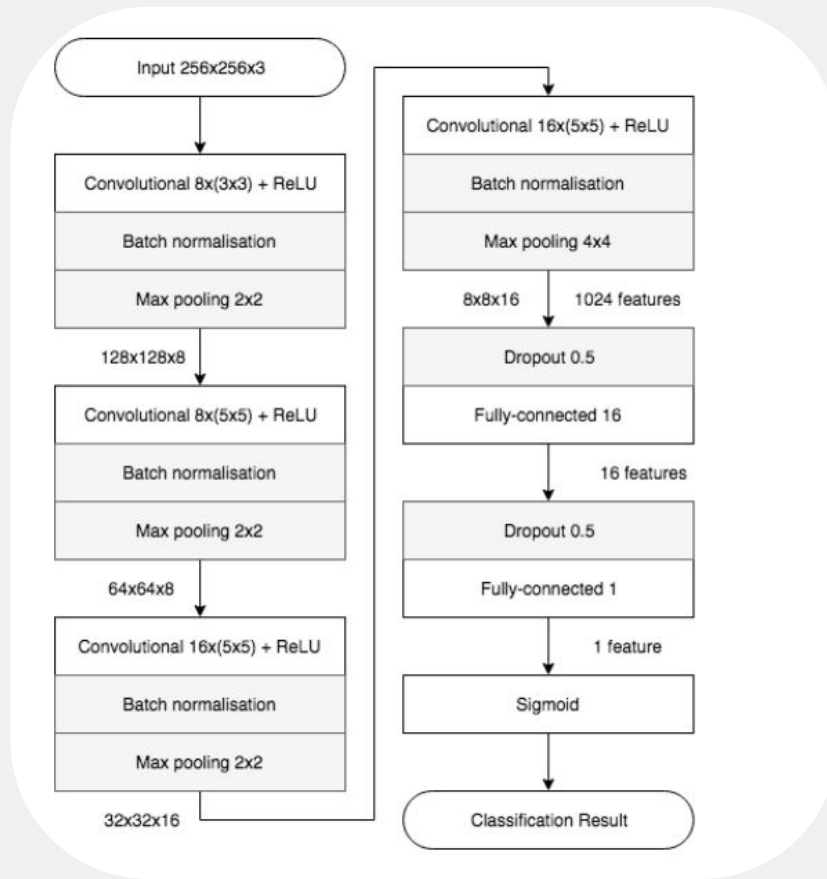
Why detection is important?

- **Defamation**
- **Misinformation**
- **Resolving conflicts**

How does detection work?

- **Meso-4**
- **MesoInception-4**

Meso-4

Dataset

- **Free available**
- **Custom**
- **Videos**

Setting up Environment

- **Google Colab**
- **Jetson**
- **Good Internet**

Meso4 on Jetson

- **Tensorflow**
- **Keras**
- **No TensorRt implementation yet**

Implemention

```python
class Meso4(Classifier):
    def __init__(self, learning_rate = 0.001):
        self.model = self.init_model()
        optimizer = Adam(lr = learning_rate)
        self.model.compile(optimizer = optimizer,
                            loss = 'mean_squared_error',
                            metrics = ['accuracy'])
```

Results?

```
.class 'numpy.ndarray'>
Predicted likelihood: 0.8994
Actual label: 1

Correct prediction: True
```