Deep Fake Detection

Noor Abdelhamed
Martin Bernardi
Mary Chris Go

Introduction

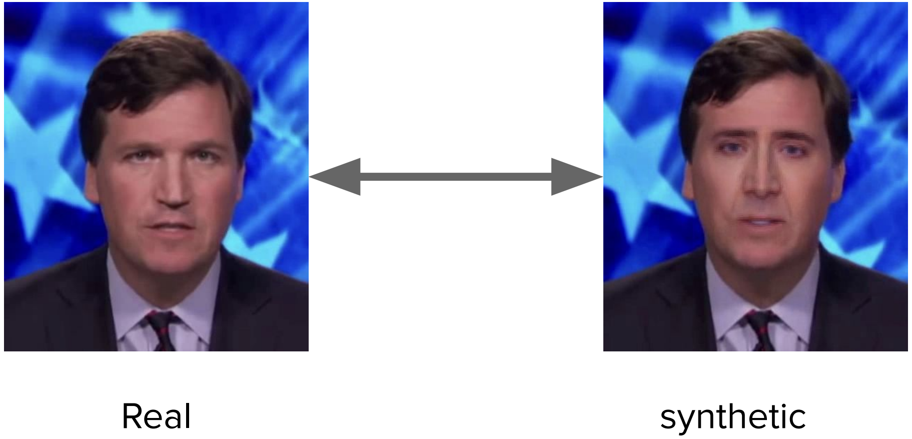
- Task 1 and 2
  - Siamese networks
  - **Block based approach**
- Task 3
  - Ensemble of CNNs
- Conclusions

Task 1, 2

Siamese Network

Why ?

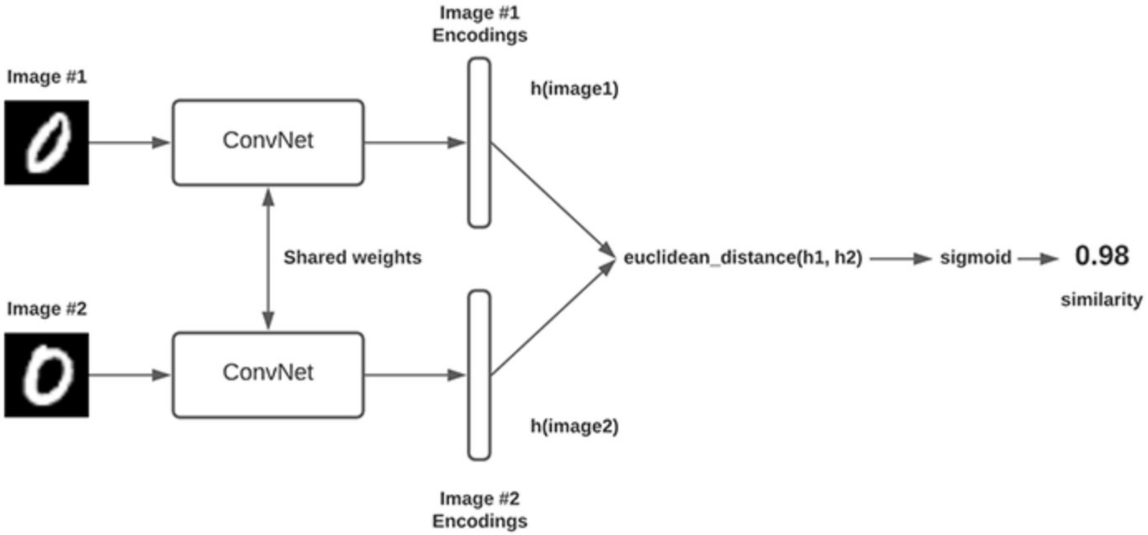
- Small Training Dataset
- Dataset Analysis
- Dissimilarity metric
- Feature Extraction
  - 48 embeddings



pyimagejobs
Powered by PyImageSearch.com

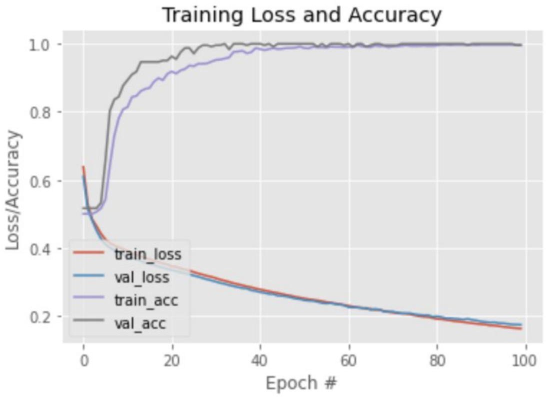
PyImageSearch: Siamese networks with Keras, TensorFlow, and Deep Learning

Architecture



Results & Adaptations

- Similarity Classification
  - Using same pairing as training dataset : 93%
  - Using random pairing from the same class: 53%
- Real/Fake Classifier
  - Fine Tune the network
    - Adapt the Siamese to single input and sigmoid output
    - Poor learning
  - Transfer Learning
    - Augment Siamese with MLP and freeze the weights (usage of 48 embedding feature vector)

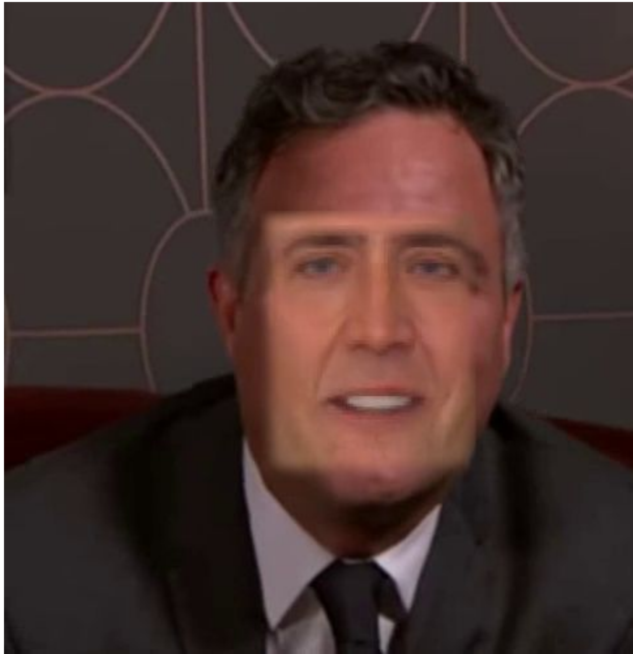


Task 1, 2

Block based approach

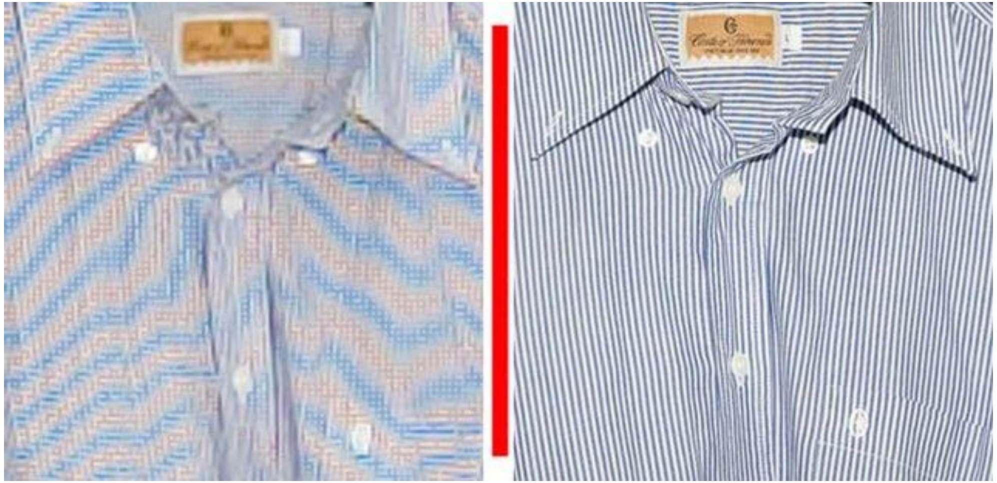
Reasoning

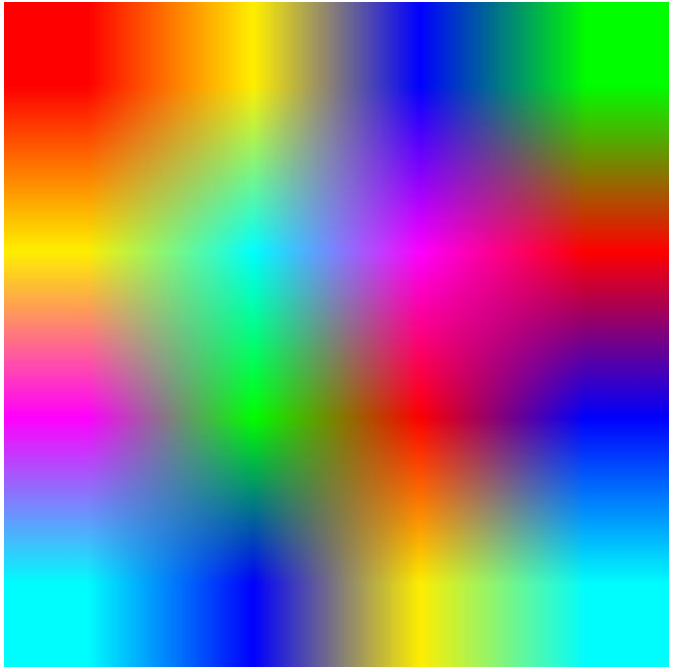
- Image modifications add almost invisible low level artifacts
  - Rescaling
  - Blurring
  - Brightness modification
  - JPEG compression
- The dataset is too small to focus in high level clues
  - Face too small for head
  - Unnatural generation of mouth
- Features used:
  - Discrete Fourier Transform
  - Histogram
  - Error Level Analysis (ELA)



Reasoning: Rescaling

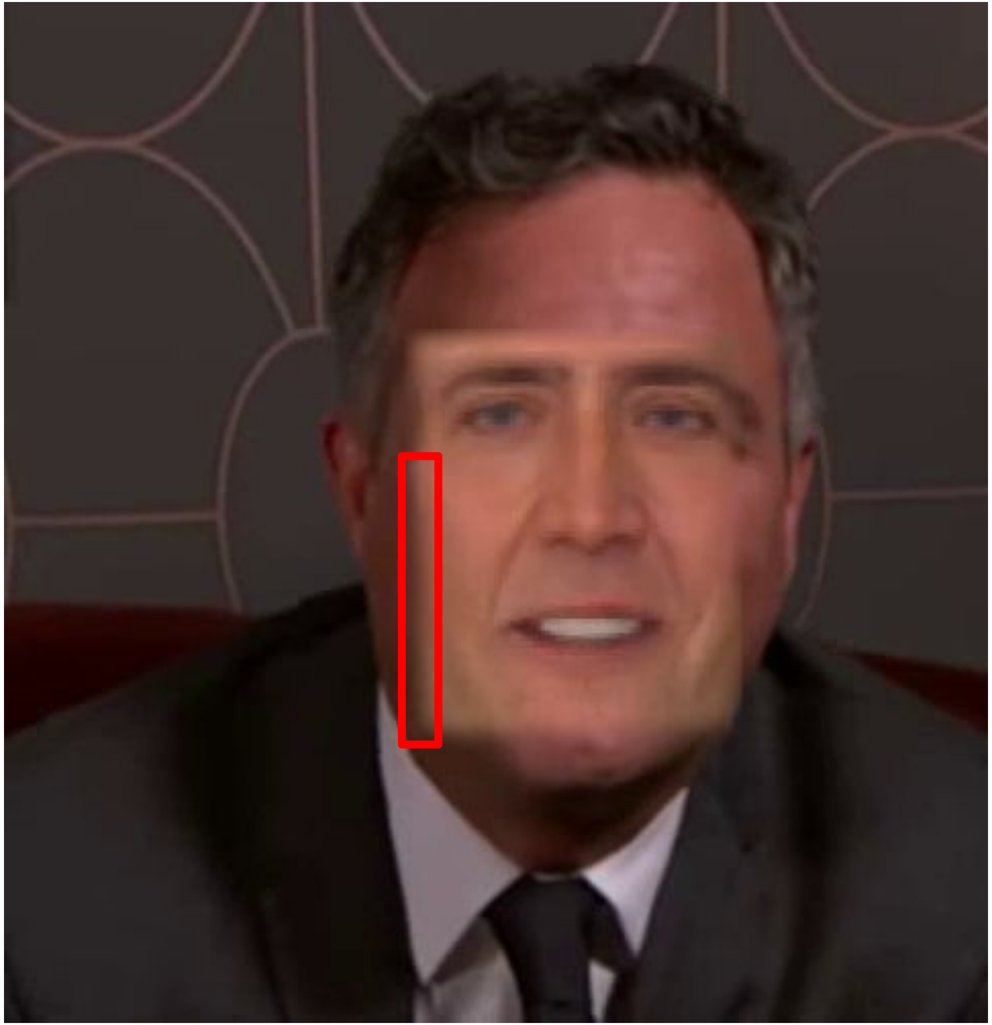
- Face is rescaled when positioned over the fake video
- Bilinear interpolation? Aliasing?
- Seen in fourier domain





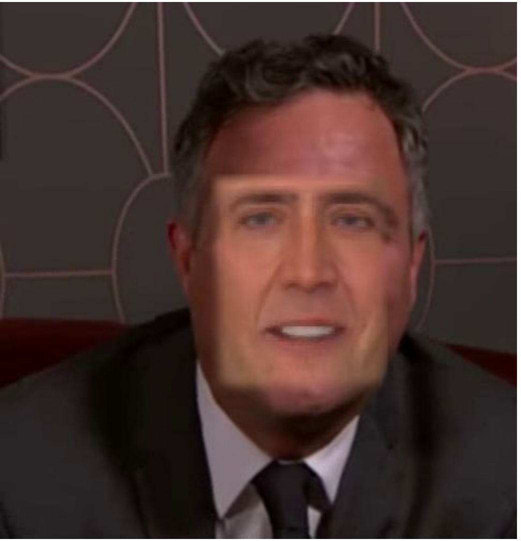
Reasoning: Blurring

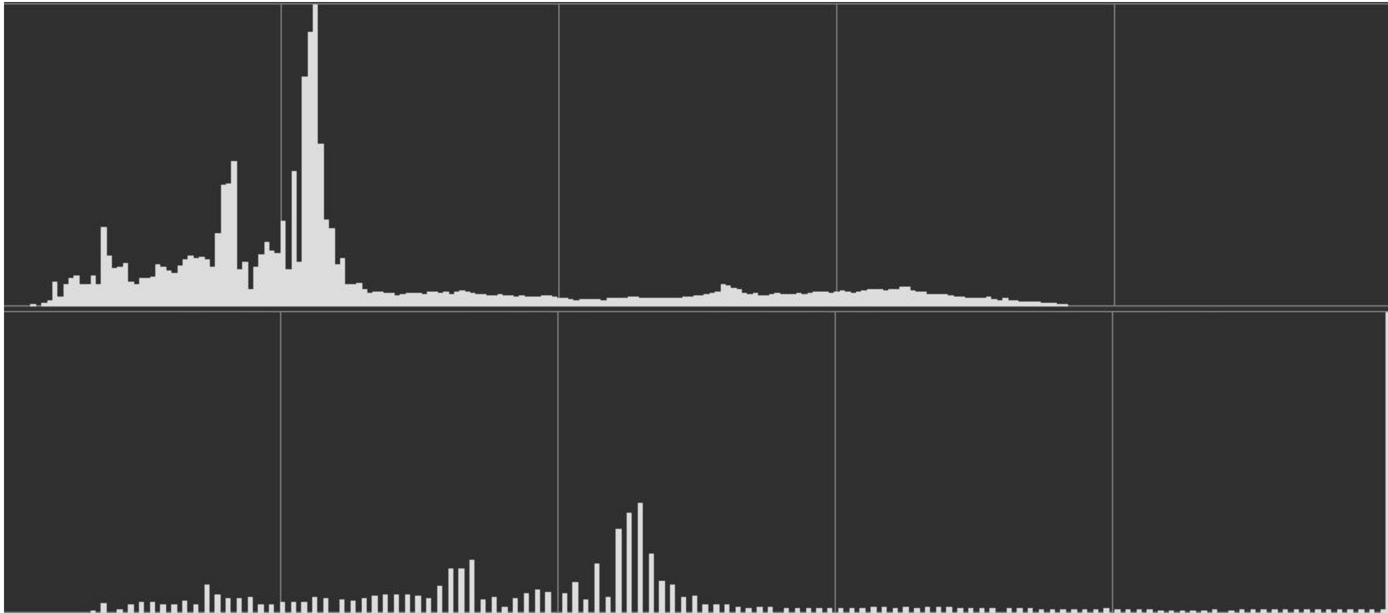
- Gaussian blurring in edges?



Reasoning: Brightness modification

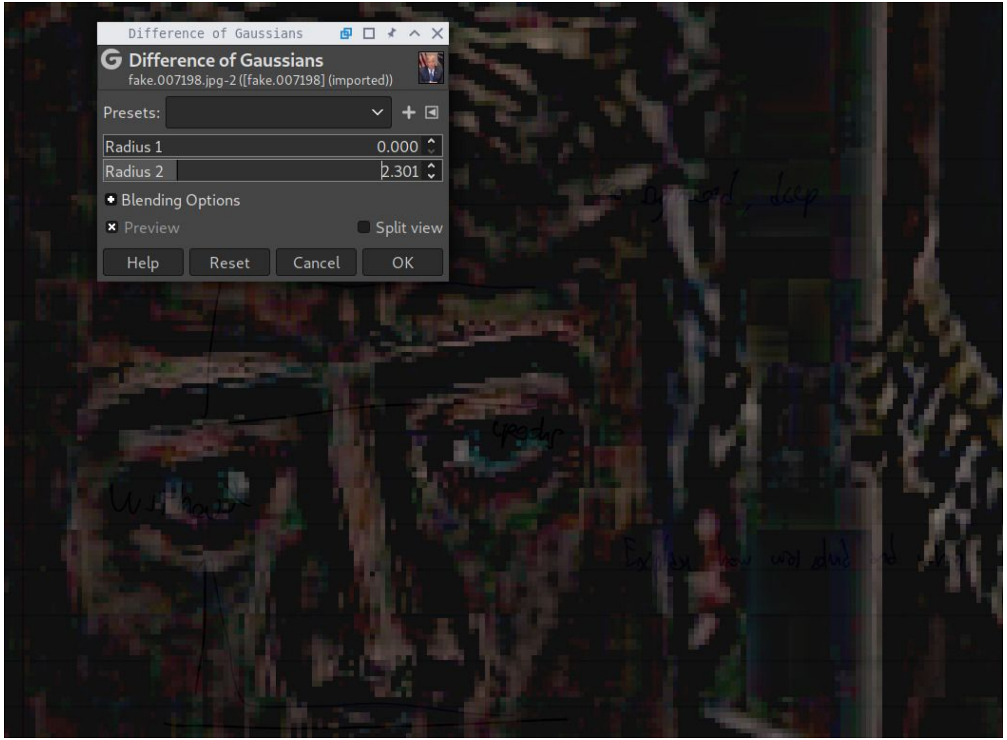
- Brightness has to be modified to match video
- Visible in histogram





Reasoning: JPEG compression

- Face and rest of video compressed different amount of times with different quality.
- Artifacts are different in real and fake parts of image
- Easier to observe in gradient image



Difference of Gaussians

fake.007198.jpg-2 ([fake.007198] (imported))

Presets:

Radius 1 0.000

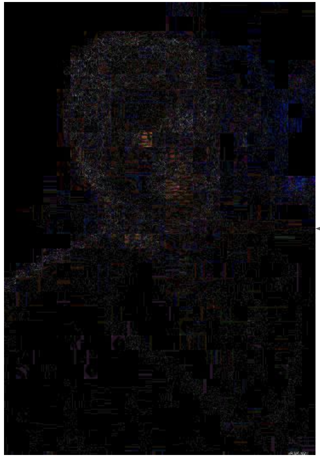
Radius 2 2.301

☒ Preview

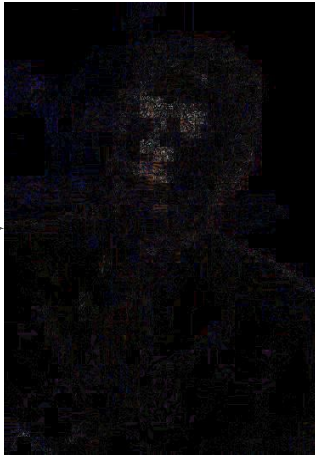
□ Split view

Reasoning: JPEG compression

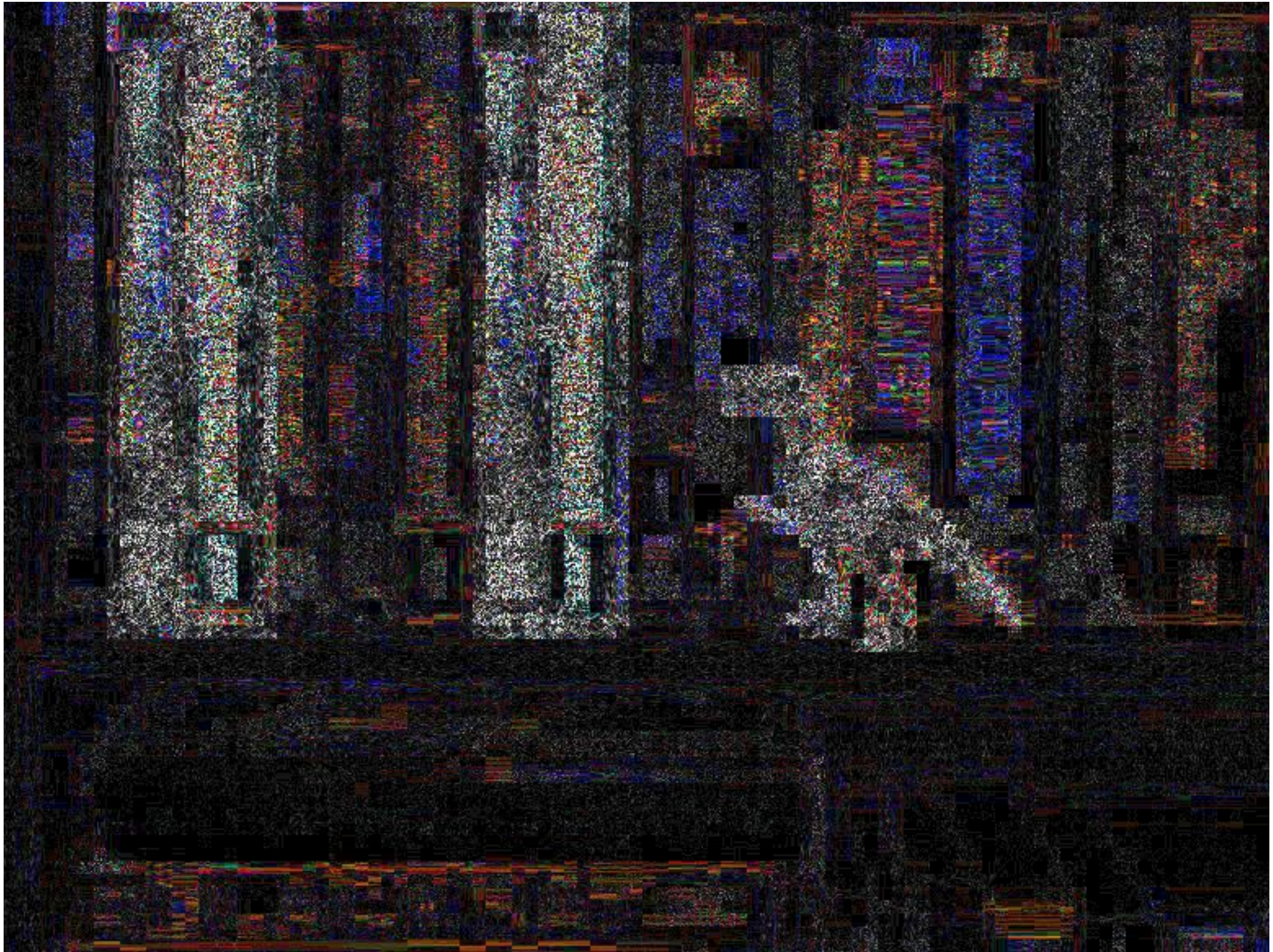
- Error Level Analysis (ELA)
- Detecting different compression levels in the same image
- Based in compressing the image again and checking the differences



real



Fake



ML Classifiers on ELA Features

Classifier	Testing Score	AUC
PCA+SVM	55%	45%
LR	54%	59%
RF	52%	51%
GBoosting	59%	61%
AdaBoosting	45%	48%

On the evaluation set of Task 1

Architecture

- 32x32 blocks
- DCT and DCT of histogram as features

Training:

- Divide image in blocks, select blocks with skin color
- Extract features
- Classify as a real or fake block

Evaluation

- Divide image in blocks
- Extract features
- Classify all blocks as real or fake
- Average of score of all blocks is the score for the image

Results

- No improvement when adding ELA features

Scores for each dataset:

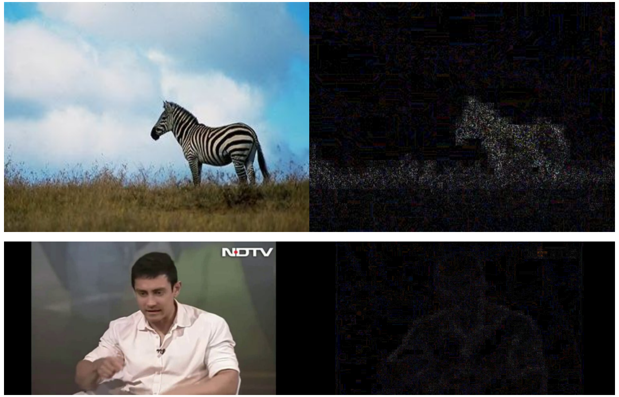
- Task 1, training: 83.1%
- Task 1, testing: 65.1%
- Task 2, testing: 56.6%

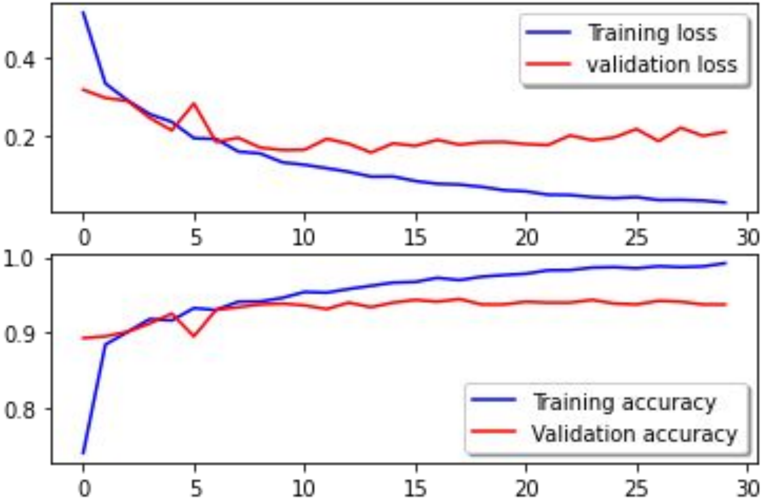
Task 3

Ensemble of CNNs

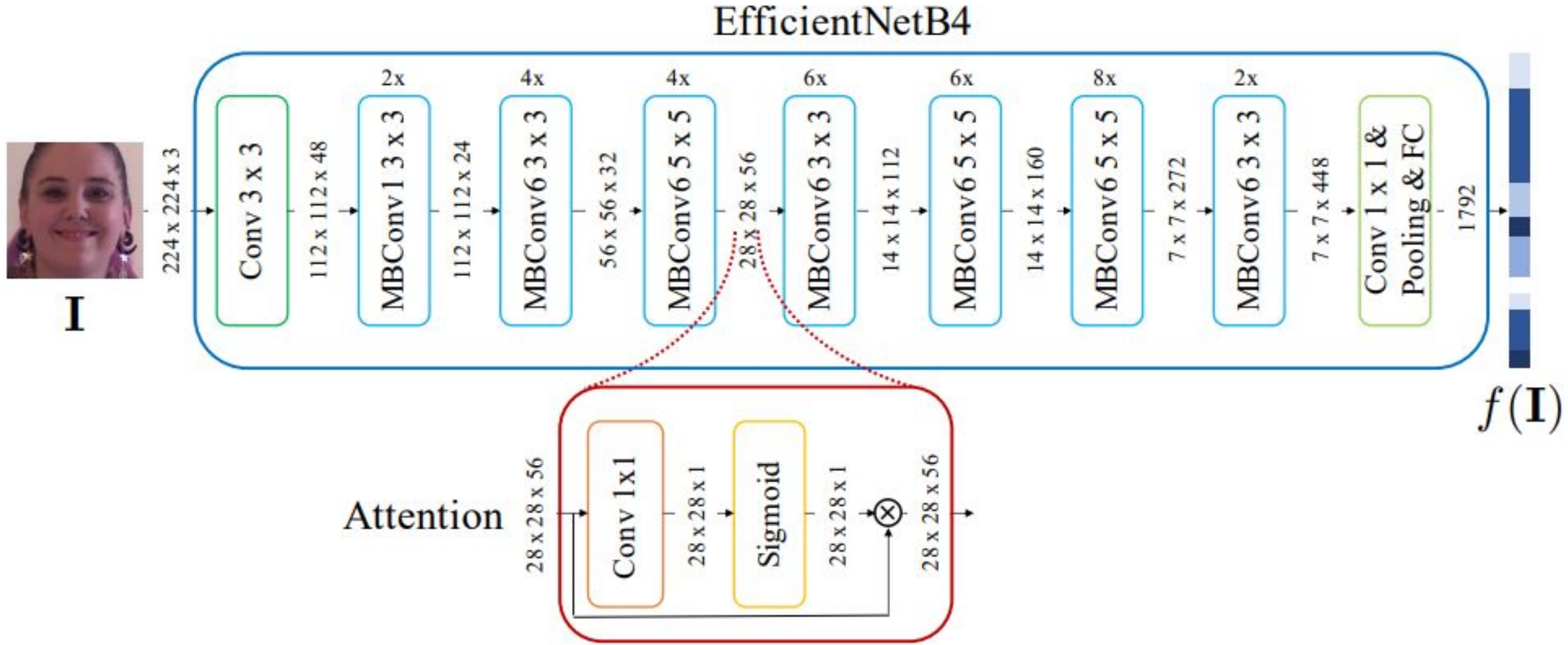
Trial # 1: ELA and Deep Learning

- network trained by CASIA dataset
- CASIA test set: 98% accuracy
- Evaluation set: 53.17%





Trial # 2: Ensemble of CNNs



Bonettini, Nicolò & Cannas, Edoardo & Mandelli, Sara & Bondi, Luca & Bestagini, Paolo & Tubaro, Stefano. (2020). Video Face Manipulation Detection Through Ensemble of CNNs.

Trial # 2: Ensemble of CNNs

- Why EfficientNetB4
    - number of parameters
    - run time
    - classification performance
        - top 1 performance in ImageNet dataset (83.8%)

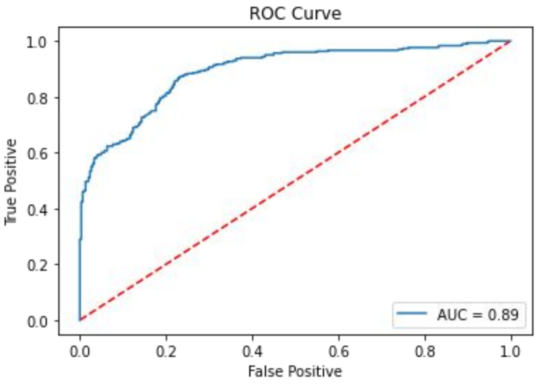
Trial # 2: Ensemble of CNNs

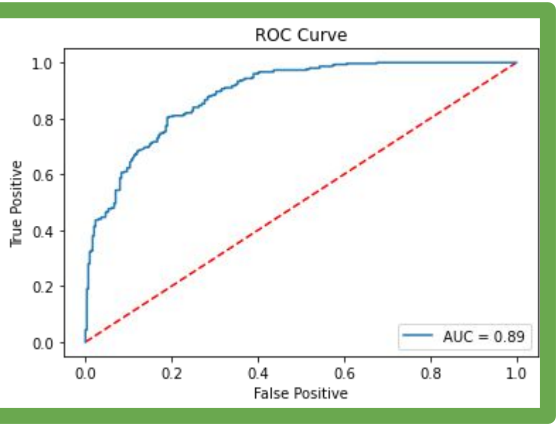
- Takes inspiration from the family of EfficientNet models

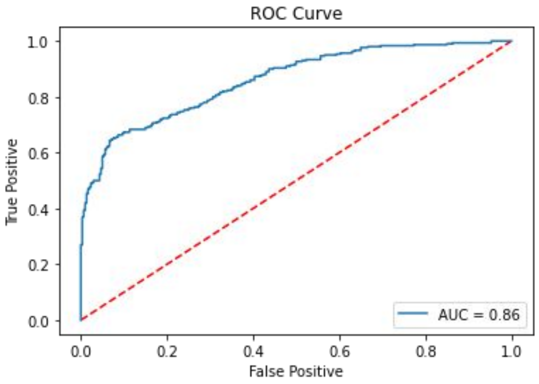
2 main concepts:

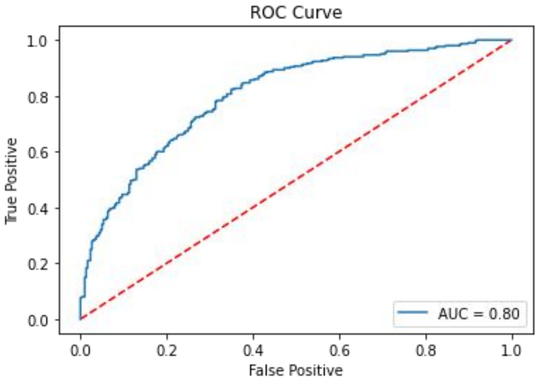
- Attention Mechanism
  - generates human comprehensible inference of the model
- Tripet siamese training strategy
  - extract features from data to achieve better classification performance.

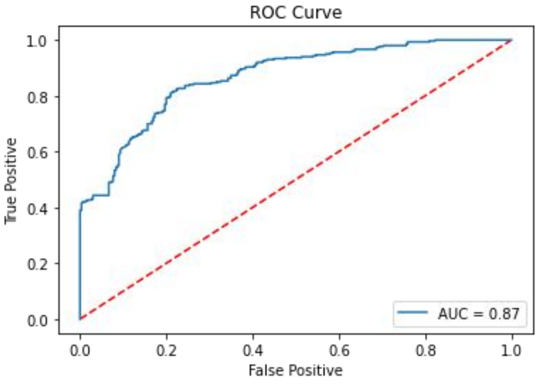
Results : DFDC dataset











Conclusions

Conclusion

Task 1 and 2

- Look at image artifacts, generic for any kind of fake image
- Siamese network to take advantage of the pairs of images present in the dataset

Task 3

- Fusion of ensemble CNNs is better than a single CNN
- Training set plays a big part in determining if your model will perform well in a specific evaluation set
- Take advantage of motion videos
- Try more models to fuse

References

- Bonettini, Nicolò & Cannas, Edoardo & Mandelli, Sara & Bondi, Luca & Bestagini, Paolo & Tubaro, Stefano. (2020). **Video Face Manipulation Detection Through Ensemble of CNNs**.
- Alin C. Popescu and Hany Farid. **Statistical Tools for Digital Forensics**
- Lilei Zheng Ying Zhang, and Vrizlynn L.L. Thing. **A survey on image tampering and its detection in real-world photos**
- Neal Krawetz. **A Picture's Worth... Digital Image Analysis and Forensics**
- Adrian Rosebrock. **Siamese networks with Keras, TensorFlow, and Deep Learning**