Deepfake Detection with Deep Learning

Marten Thompson

School of Statistics
University of Minnesota

December 17, 2020

Product is Nothing New

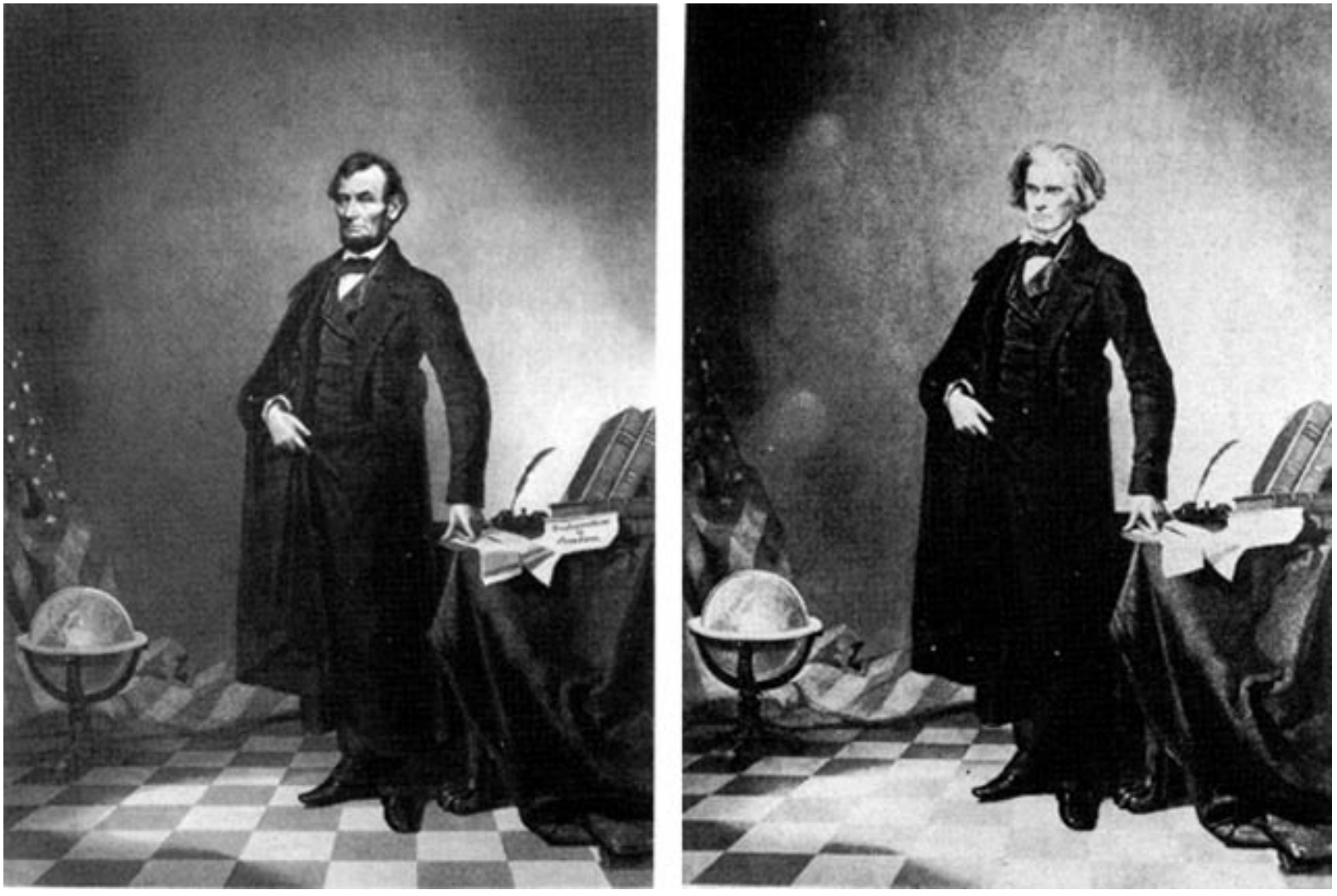


Figure: Face swapping at the highest office [4]

Product is Much More Convincing

Deepfake: audio, video, and image content altered by deep learning tools to replace the original subject with another

A myriad of machine learning tools: GANs, auto-encoders

Constant improvement by hobbyists, open source developers, and private enterprises

Obvious potential for personal and professional harm

See example

Deepfake Creation

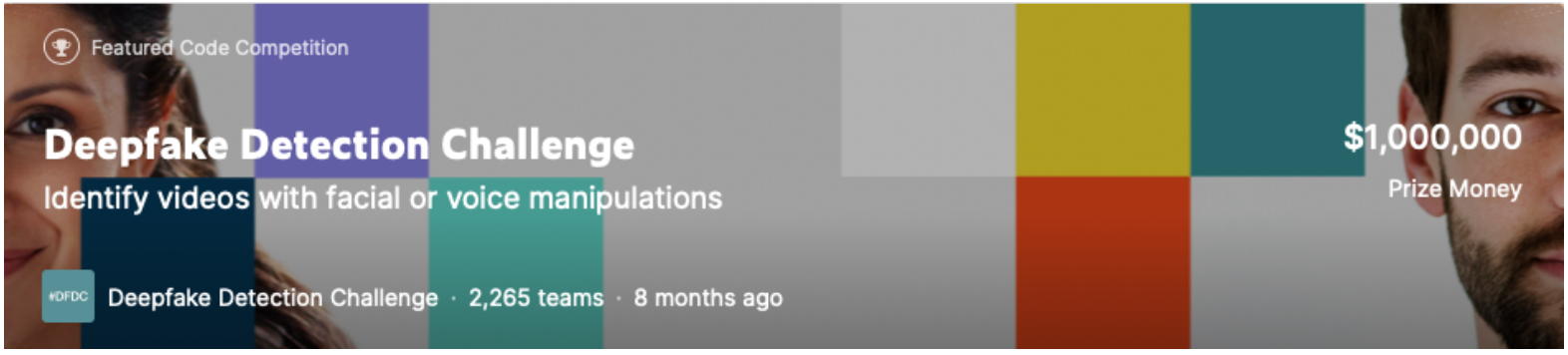
Deepfake software is

- Effective
- Automatic
- Accessible

Options include

- Opensource software: DeepFakeLab [5]
- Mobile applications: ZAO [7], FaceApp [3]
- Online market

The Deepfake Detection Challenge (DFDC)



Created by: Amazon Web Services, Facebook, Microsoft, and ThePartnership on AI's Media Integrity Steering Committee

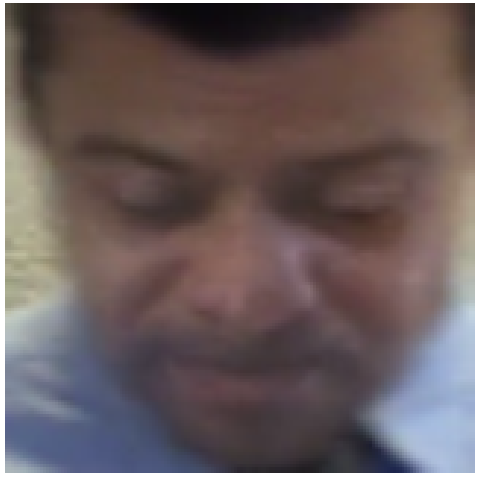
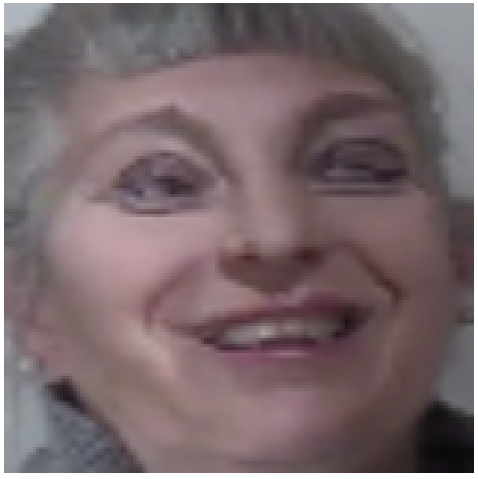
The DFDC Data I

3,400 paid actors across 128,000 videos

25 terabytes of videos disseminated as 300 frame clips

85% altered by one of five popular deepfake methods

The DFDC Data II: Good, Bad, Ugly


The DFDC Data III: Opportunities

Wanted to leverage temporal artifacts

Considered both 2D and 3D data

The DFDC Data IV: Challenges

Inconsistent/missing application of deepfake model on data labeled fake

Obscured content and non-human subjects

All realistic in practice

Data Pipeline

Implemented locally, `compute.cla`, and Google Colab

Out-of-the-box solution for 2D data

Wrote and stored 3D numpy arrays, read in batches by custom generator

Models: CNN

Variations of basic convolution units, ending in fully connected layers

2D and 3D Models

- **one**, two, and three convolutions before pooling
- 8, 16, and **32** filters
- one, two, **three**, or four repetitions of these
- **one**, two, or three fully connected layers before the final output node
- trained with dropout with probabilities ranging from zero to **0.5**

MC 2 Hybrid Model

- two 3D convolutions, max pooling
- $\{$two 2D convolutions, max pooling$\} \times 2$
- fully connected (64)

Models: ResNets

Repeated residual units, ending in fully connected layers
2D and 3D models

- two or three convolutions per block
- filter size 16, . . ., 128
- stepping up the number of filters between blocks
- application ReLU activation only after the last convolution
- all ended in a global average pooling layer
- low performance

Models: Transfer Learning

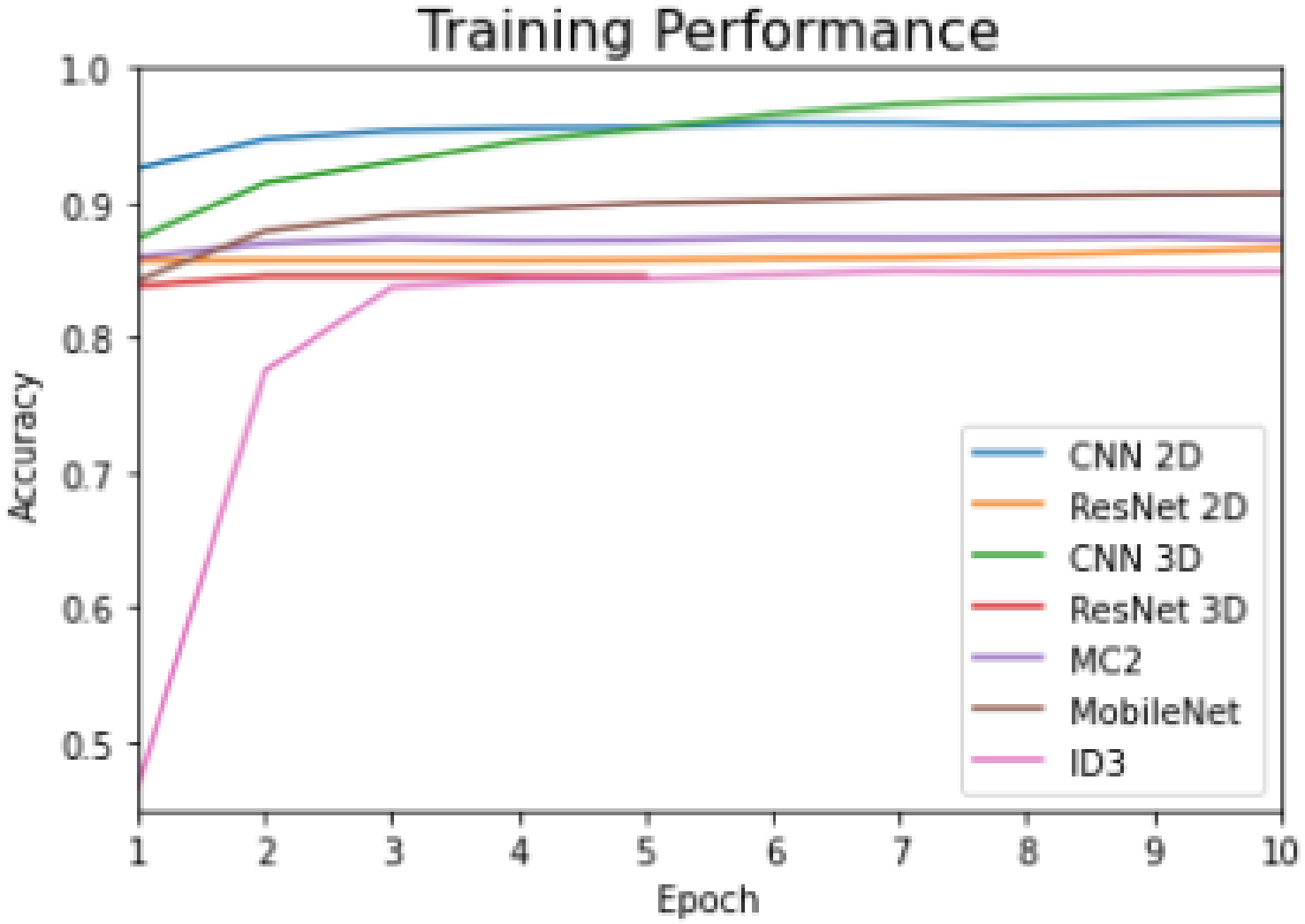
2D: MobileNetV2 [6]

- frozen 20 layer CNN trained on ImageNet
- removed final layers, replaced with two fully connected layers and final classification node
- highest performance

3D: I3D [1]

- frozen 3D inception net trained on Kinetics data
- removed final layers, replaced with two fully connected layers and final classification node
- unremarkable performance

Results I


Training Performance
Accuracy
Epoch
1.0
0.9
0.8
0.7
0.6
0.5
1
2
3
4
5
6
7
8
9
10
CNN 2D
ResNet 2D
CNN 3D
ResNet 3D
MC2
MobileNet
ID3

Results II

Model	Accuracy
MobileNetV2	0.915
CNN 2D	0.884
I3D	0.855
CNN 3D	0.853
ResNet 2D	0.849
MC2	0.849
ResNet 3D	0.843

2D methods consistently outperform 3D architectures

Discussion

Confirm difficulty in DFDC

Temporal information was not leveraged successfully

MobileNetV2 and I3D may be sensitive to pre-training [2]

References I

Carreira, J., and Zisserman, A.

Quo vadis, action recognition? a new model and the kinetics dataset.

In *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017), pp. 6299–6308.

de Lima, O., Franklin, S., Basu, S., Karwoski, B., and George, A.

Deepfake detection using spatiotemporal convolutional networks.

arXiv preprint arXiv:2006.14749 (2020).

FaceApp.

`https://www.faceapp.com`, (accessed 12/10/2020).

Güera, D., and Delp, E. J.

Deepfake video detection using recurrent neural networks.

In *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)* (2018), IEEE, pp. 1–6.

References II

Petrov, I., Gao, D., Chervoniy, N., Liu, K., Marangonda, S., Umé, C., Jiang, J., RP, L., Zhang, S., Wu, P., et al.

Deepfacelab: A simple, flexible and extensible face swapping framework.

arXiv preprint arXiv:2005.05535 (2020).

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C.

Mobilenetv2: Inverted residuals and linear bottlenecks.

In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2018), pp. 4510–4520.

ZAO.

`https://zaodownload.com`, (accessed 12/10/2020).