Robust Detection of Deepfake Videos

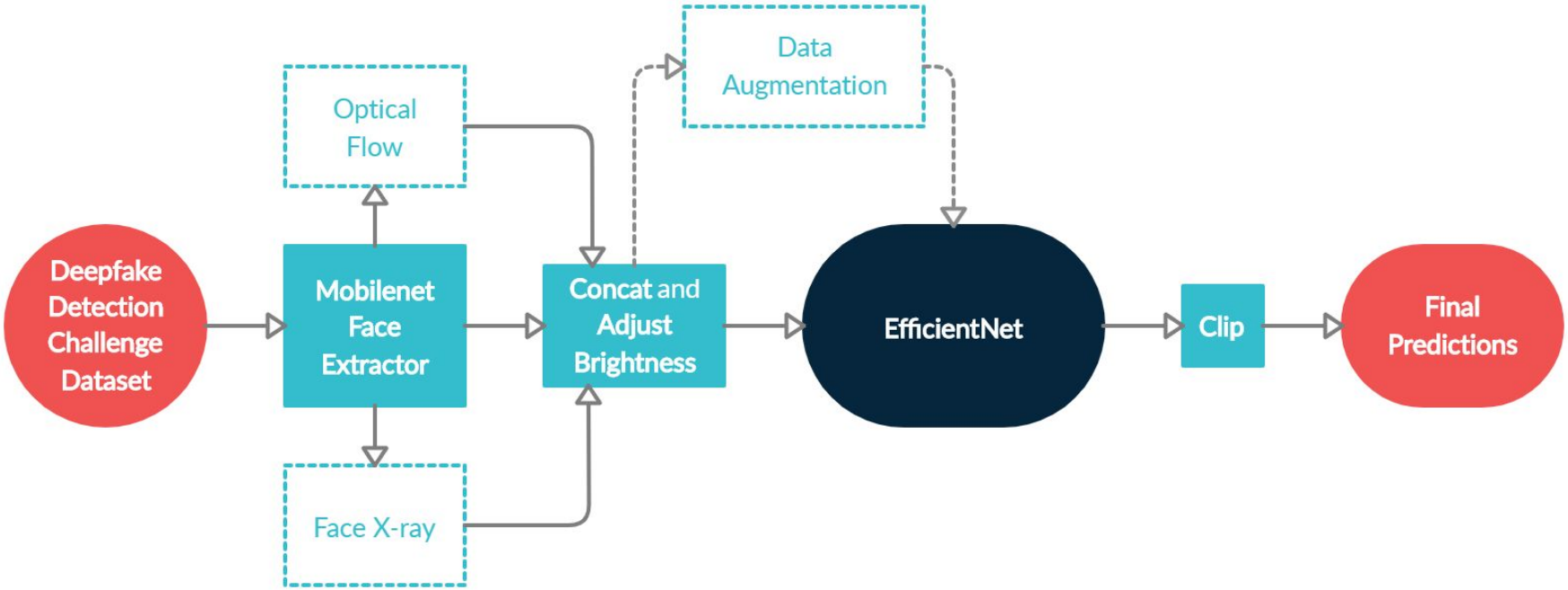
Pranay Pasula, Ujjwal Singhania, Yijiu Zhong

Advisors: Dawn Song,
Ruoxi Jia

Motivation

Source: https://ai.facebook.com/blog/deepfake-detection-challenge/

Pipeline

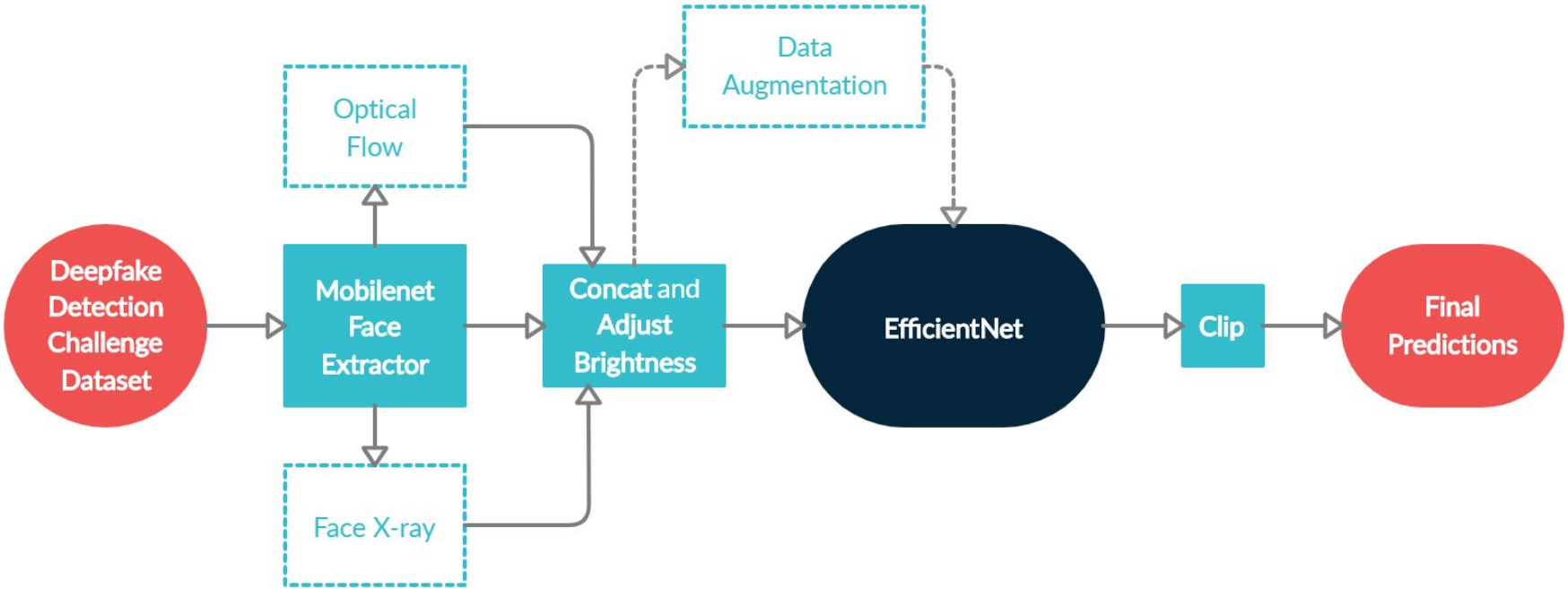

Data Augmentation
Optical Flow
Deepfake Detection Challenge Dataset
Mobilenet Face Extractor
Concat and Adjust Brightness
EfficientNet
Clip
Final Predictions
Face X-ray

Deepfake Detection Challenge Dataset (DFDC)*

- ~500GB (119,146 real/fake videos)
- Deepfakes created by multiple methods to test generalization

* https://deepfakedetectionchallenge.ai

Pipeline

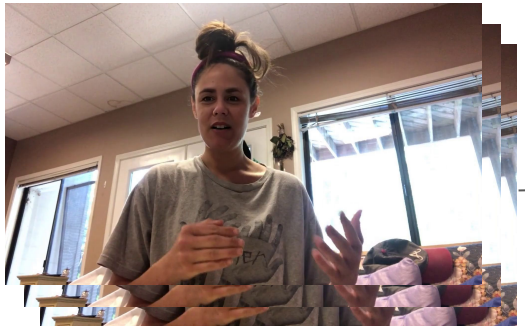

Data Augmentation
Optical Flow
Deepfake Detection Challenge Dataset
Mobilenet Face Extractor
Concat and Adjust Brightness
EfficientNet
Clip
Final Predictions
Face X-ray

Optical Flow

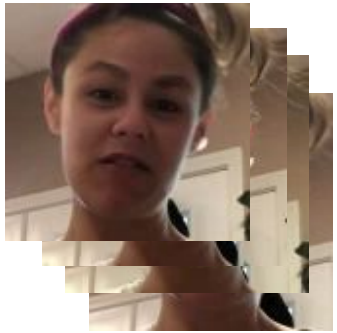
Optical flow denotes apparent motion of objects, surfaces and edges in a visual scene

Image Source: Nvidia Developer Blog

Optical Flow Data Pipeline

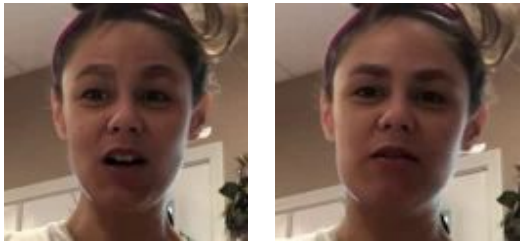


300 frames (1920 x 1080)

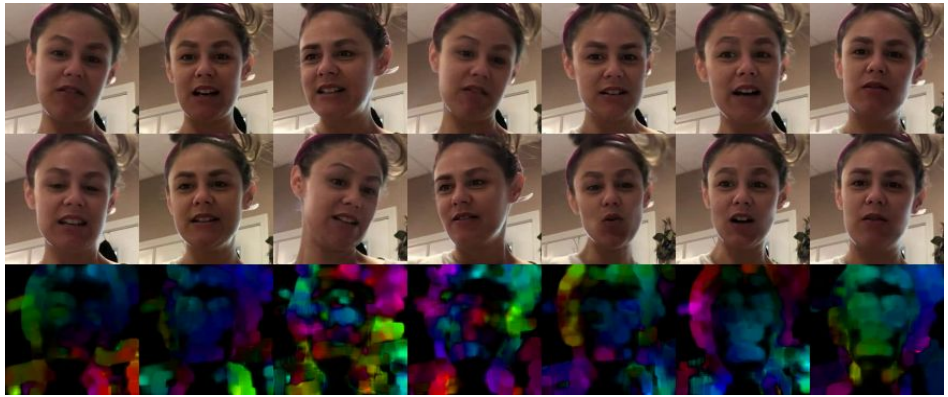


300 faces (250 x 250)

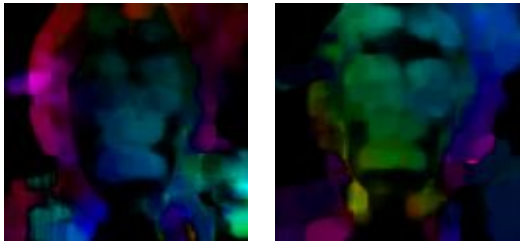
T



T + 1

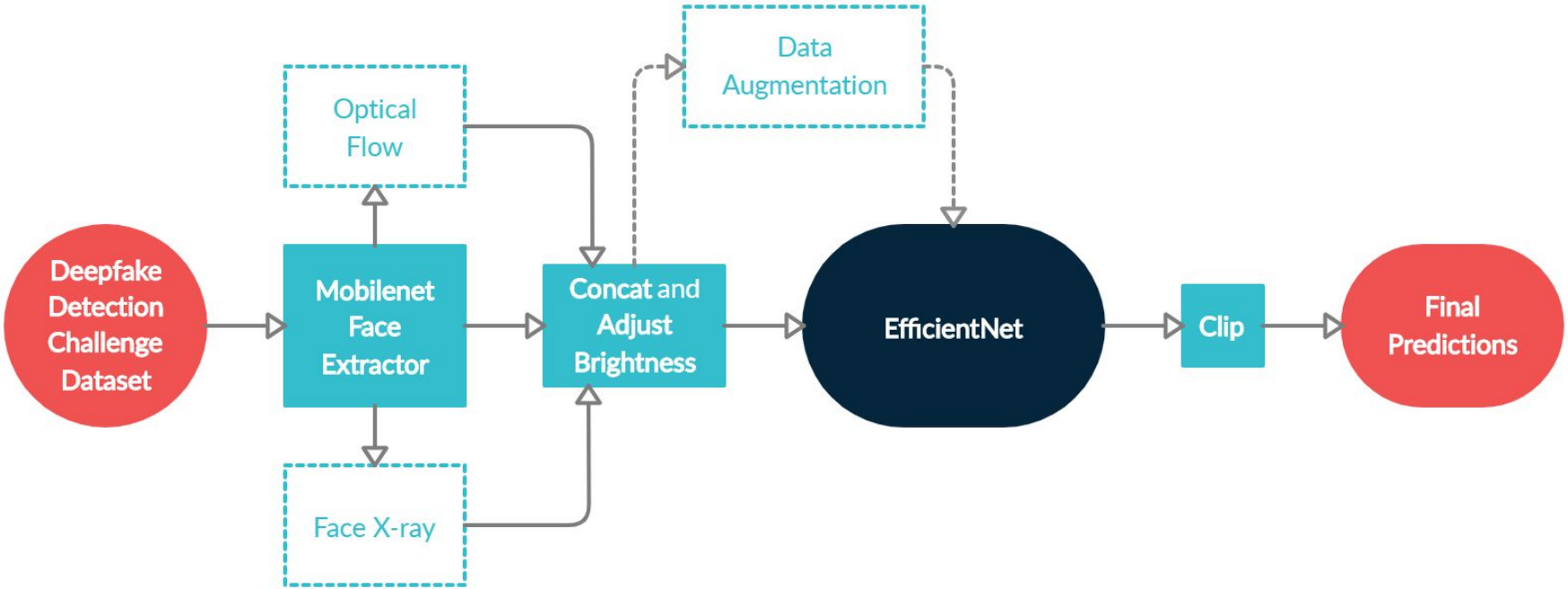


1 temporal face grid (875 x 375)



Dense optical flow between T and T + 1 frames

Pipeline

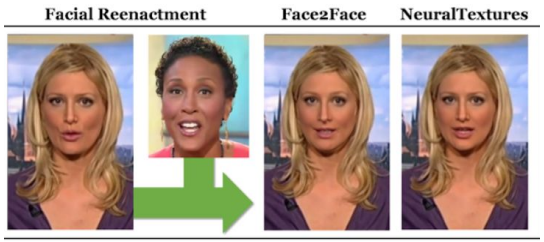

Data Augmentation
Optical Flow
Deepfake Detection Challenge Dataset
Mobilenet Face Extractor
Concat and Adjust Brightness
EfficientNet
Clip
Final Predictions
Face X-ray

Is there a commonality between different deepfake generation techniques?

Deepfake Generation

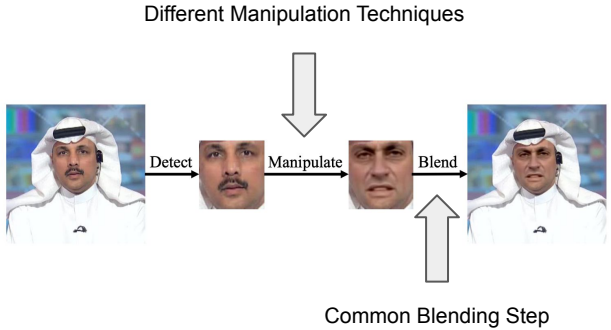
Facial Identity Manipulation

Facial Expression Manipulation



(Rossler et al., 2019)

Common Blending Step



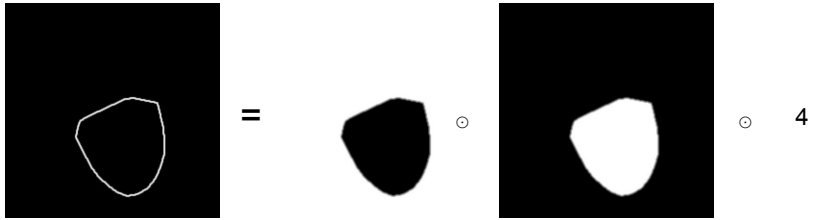
(Li et al., 2019)

Face X-ray Detection

Face X-ray Definition

Calculate $4 \cdot v \cdot (1 - v)$

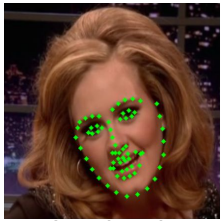
Definition



Face X-ray = 1 - mask ⊙ mask ⊙ 4

(Li et al., 2019)

Face Crop



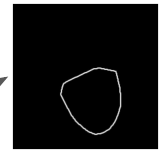
Landmark Detection

Data Generation Pipeline

HRNet*

Face X-ray prediction

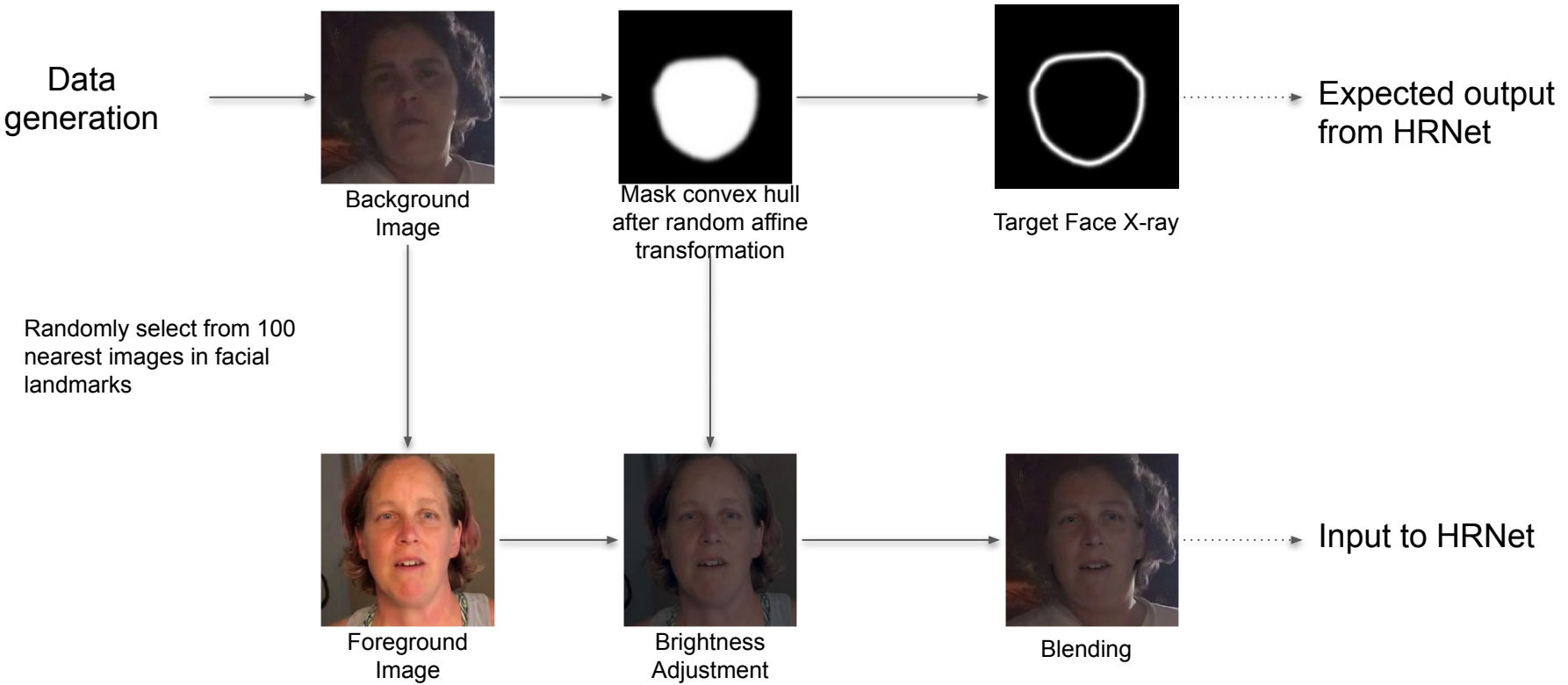
If fake



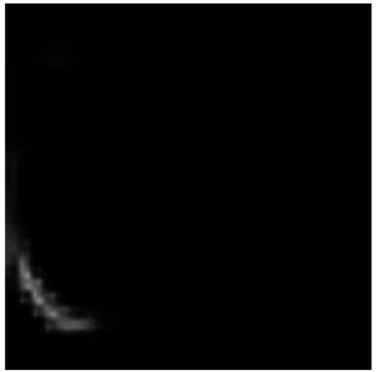
If real

* Wang et al., 2019

Face X-ray Data Generation Pipeline



Face X-ray prediction results



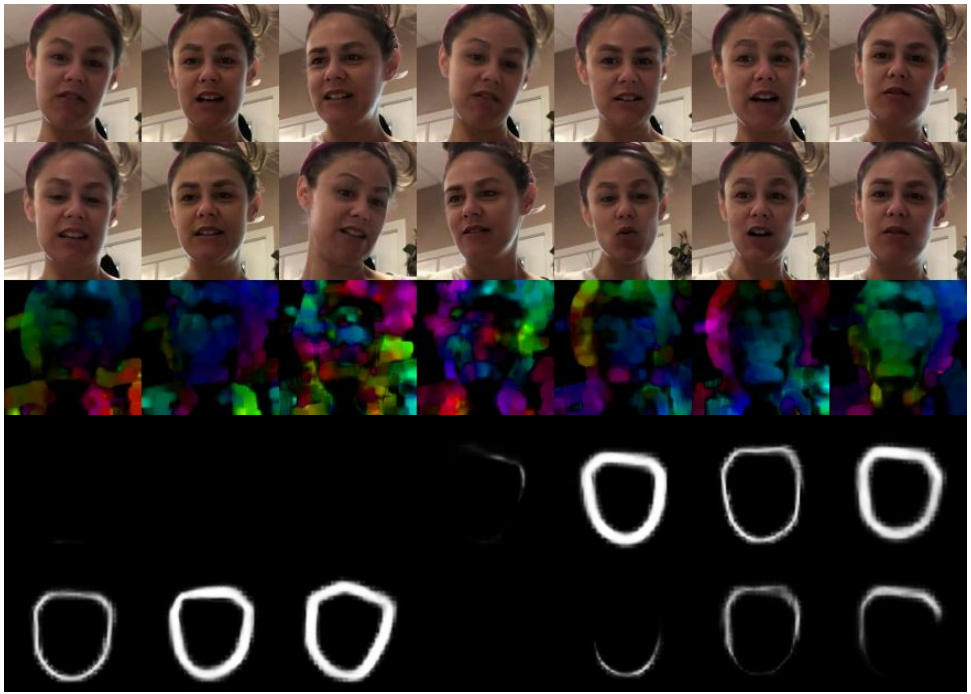
Real image

Image blended using Gaussian blur

Image blended using unknown method

- Result is not ideal
  - Different blending methods
  - Resolution
- However, provides us insights about blending boundary.

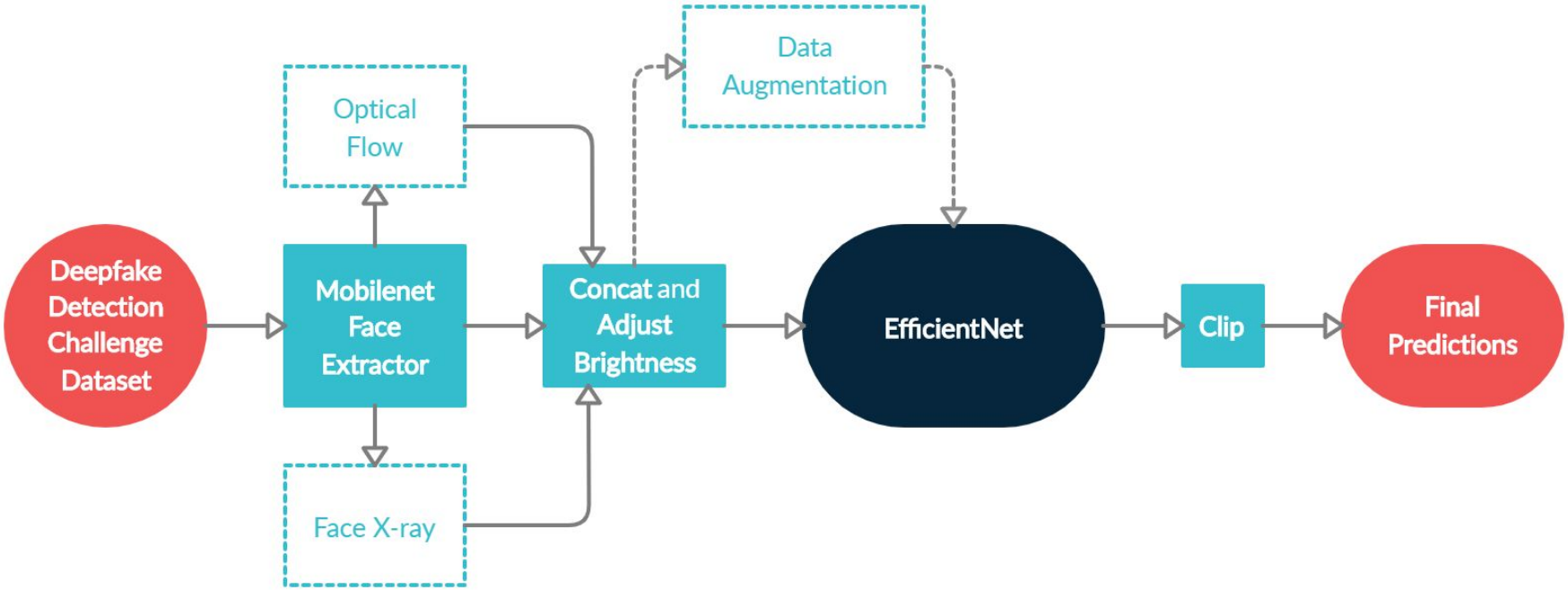
Face X-ray + Optical Flow



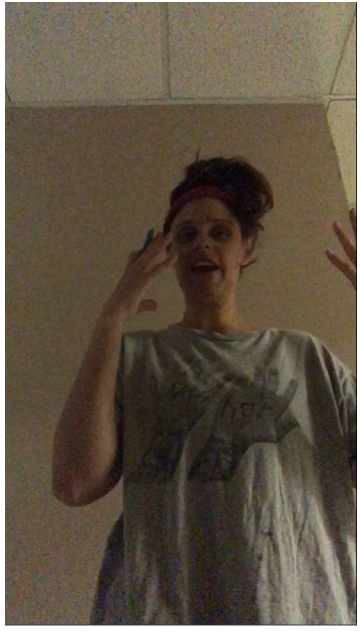
New Training Image:

- **14 face frames** (upper row is time = t and bottom row is time = t + 1)
- **7 optical flow frames** (between each t and t + 1 frame)
- **14 face x-ray frames** (one for each face frame)

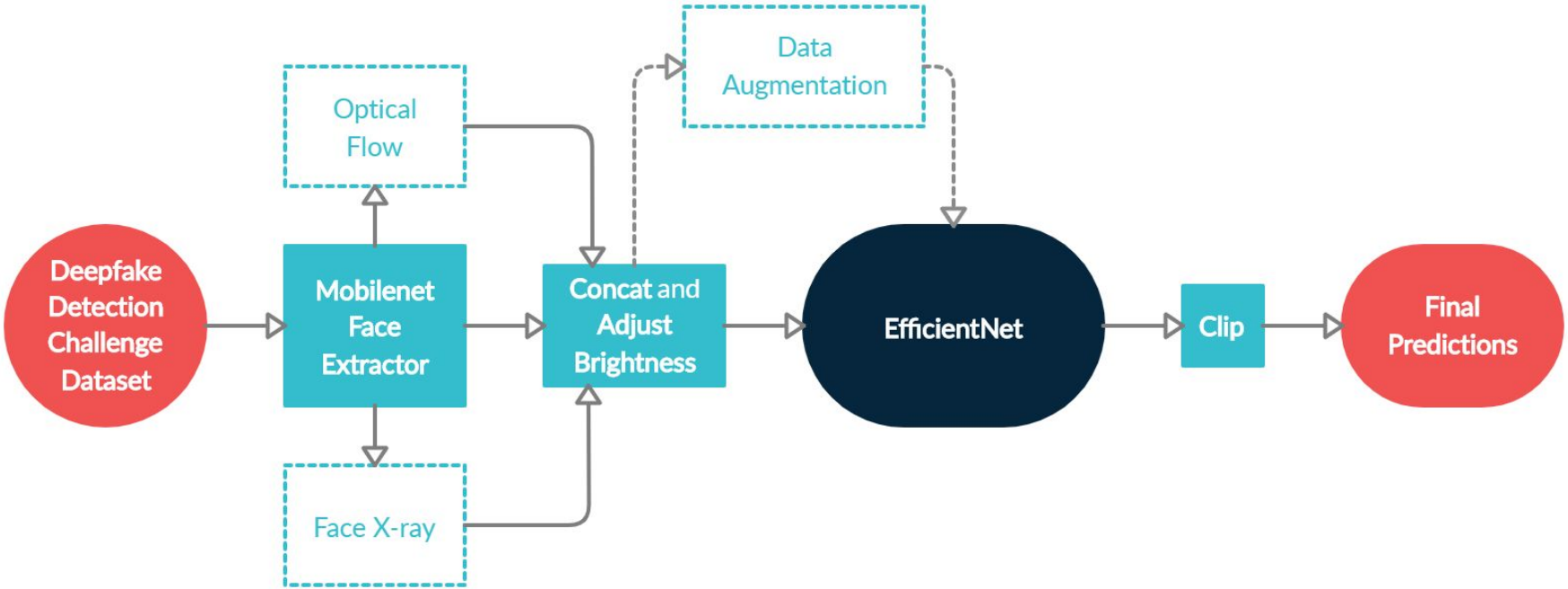
Pipeline


Data Augmentation
Optical Flow
Deepfake Detection Challenge Dataset
Mobilenet Face Extractor
Concat and Adjust Brightness
EfficientNet
Clip
Final Predictions
Face X-ray

Face extraction, concatenation, and brightness adjustment



Pipeline

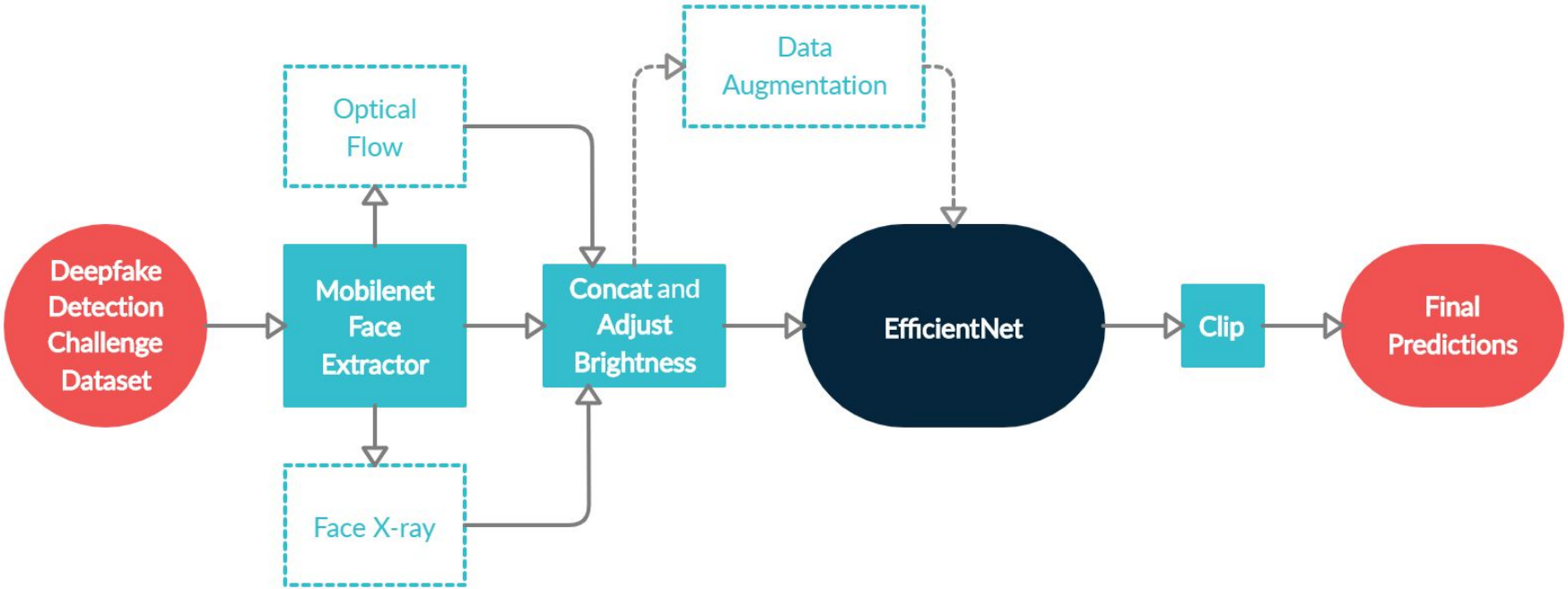

Data Augmentation
Optical Flow
Deepfake Detection Challenge Dataset
Mobilenet Face Extractor
Concat and Adjust Brightness
EfficientNet
Clip
Final Predictions
Face X-ray

Data augmentation

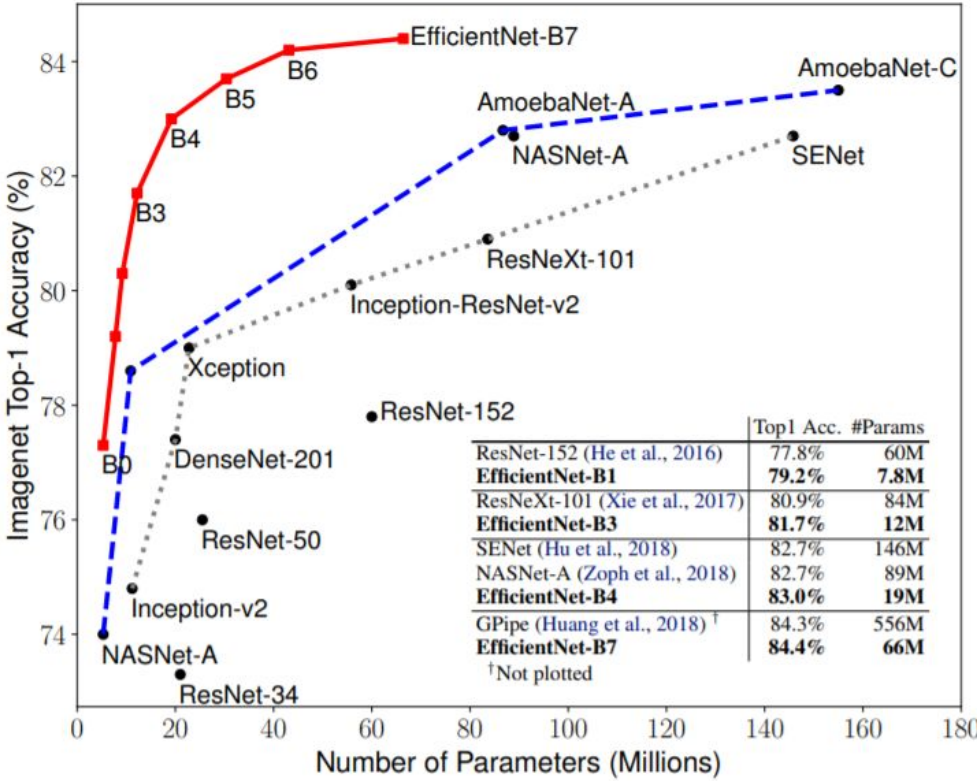
Original | HSV shift | JPEG compression | Elastic transform | Horizontal flip

Not shown: Gaussian blur and downscale + upscale

Pipeline


Data Augmentation
Optical Flow
Deepfake Detection Challenge Dataset
Mobilenet Face Extractor
Concat and Adjust Brightness
EfficientNet
Clip
Final Predictions
Face X-ray

EfficientNet performance vs SotA classifiers

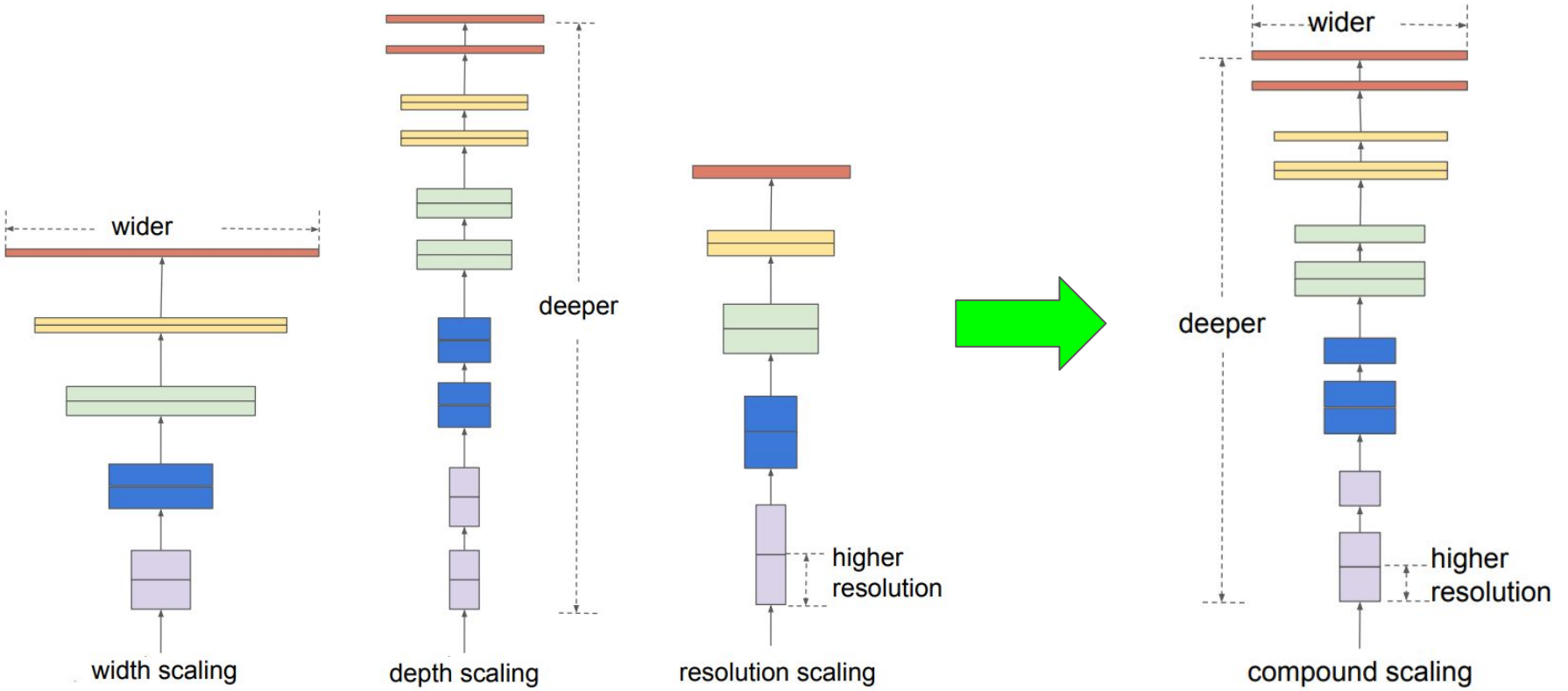


	Top1 Acc.	#Params
ResNet-152 (He et al., 2016)	77.8%	60M
EfficientNet-B1	**79.2%**	**7.8M**
ResNeXt-101 (Xie et al., 2017)	80.9%	84M
EfficientNet-B3	**81.7%**	**12M**
SENet (Hu et al., 2018)	82.7%	146M
NASNet-A (Zoph et al., 2018)	82.7%	89M
EfficientNet-B4	**83.0%**	**19M**
GPipe (Huang et al., 2018) †	84.3%	556M
EfficientNet-B7	**84.4%**	**66M**

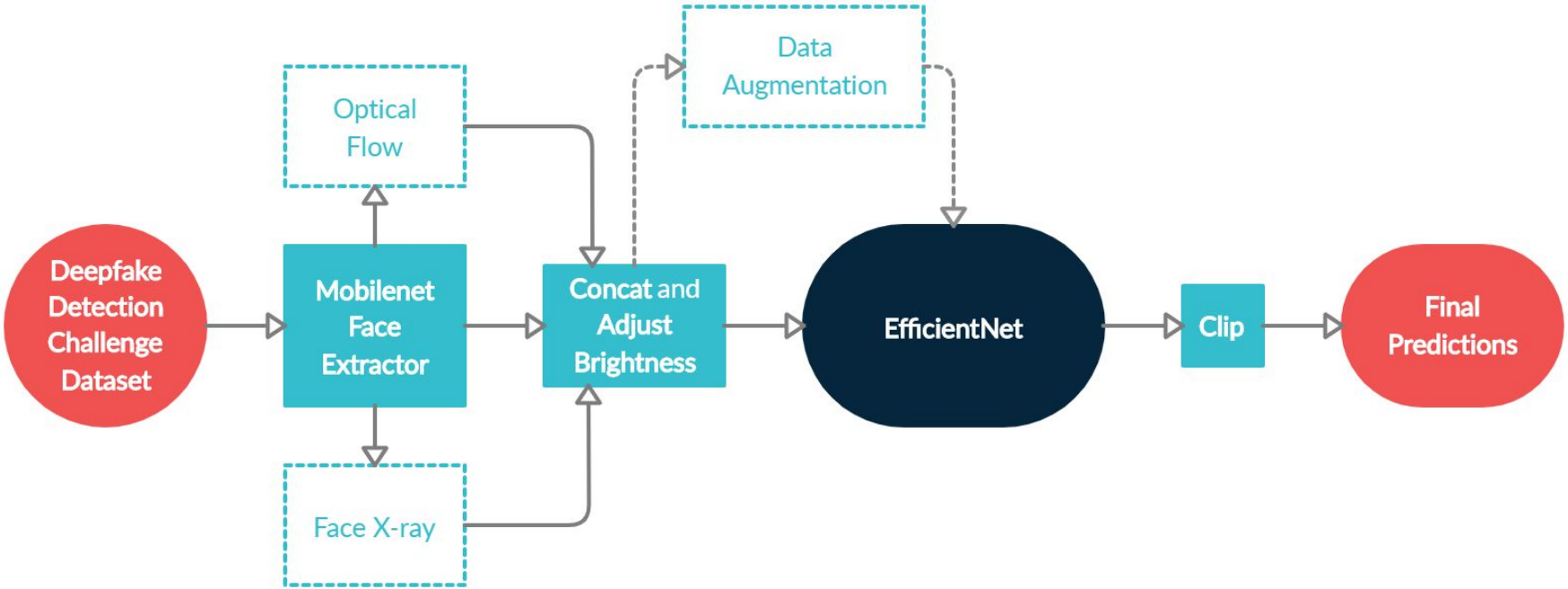
†Not plotted

[Tan et al., 2019]

EfficientNet - compound scaling

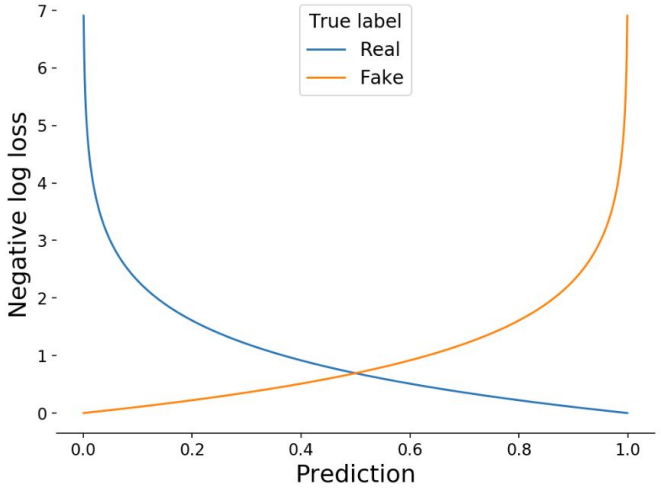


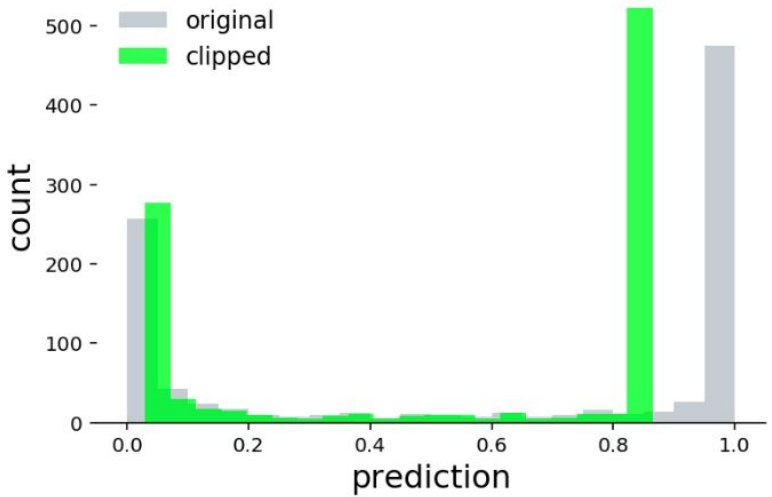
Pipeline


Data Augmentation
Optical Flow
Deepfake Detection Challenge Dataset
Mobilenet Face Extractor
Concat and Adjust Brightness
EfficientNet
Clip
Final Predictions
Face X-ray

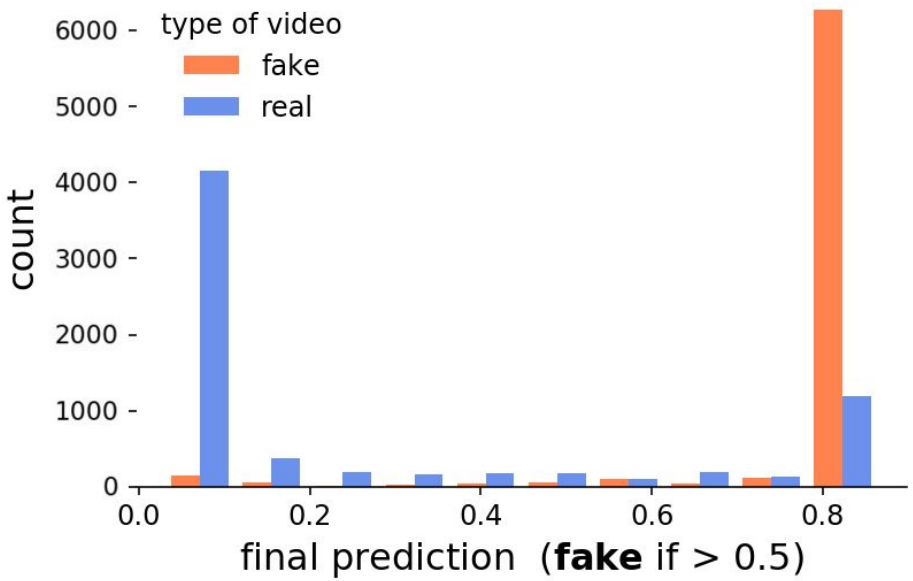
Clipping predictions

$$\text{LogLoss} = -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)\right]$$





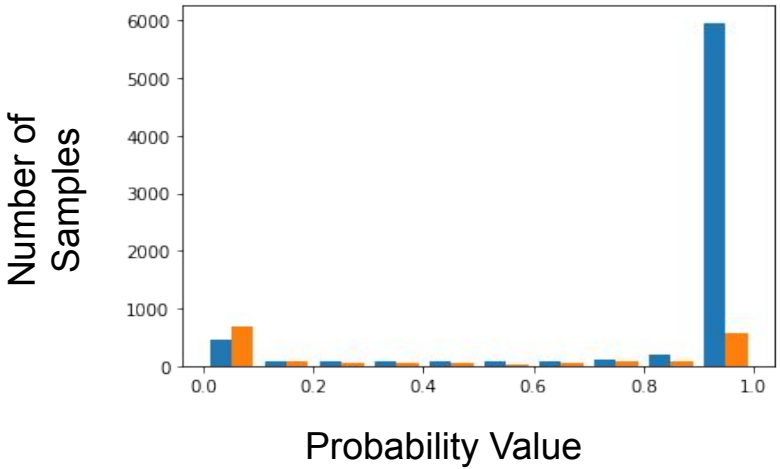
Final predictions (without x-ray or optical flow)



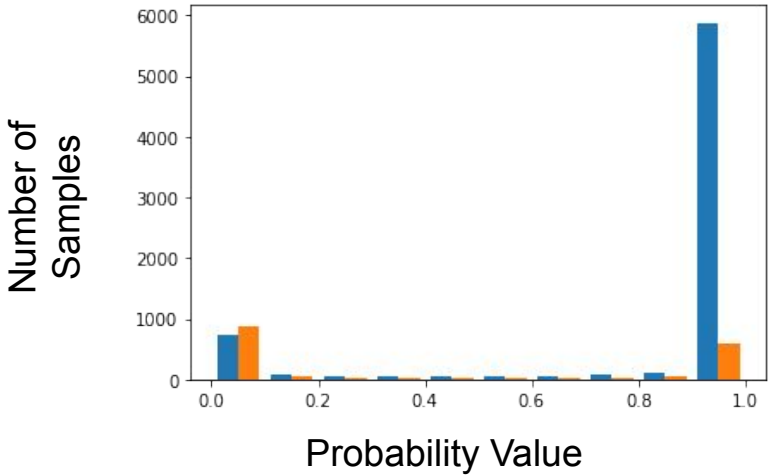
- Accuracy: 0.89
- F1 Score: 0.90
- Fake Detection Rate: 0.98
- Real Detection Rate: 0.78

Face X-ray + Optical Flow vs Optical Flow - Results

- Recall: 0.896
- Precision: 0.890
- F1 Score: 0.893
- Fake Detection Rate: 0.896
- Real Detection Rate: 0.535
- Accuracy: 0.827



- Recall: 0.86
- Precision: 0.90
- F1 Score: 0.88
- Fake Detection Rate: 0.86
- Real Detection Rate: 0.59



Discussion

- Strong fake detection rate
  - Usually more important to detect deepfakes than to detect real videos
- Recurrent failure modes
  - Facial hair
  - Glasses
  - Multiple people
- Effect of image resolution on optical flow output and face x-ray
  - Tradeoff between performance and granularity of optical flow information
  - HRNet bad performance on low resolution grids

Future Work

- Optical flow on x-ray images
- Sound as synergistic modality
- Explore better optical flow calculation
  - PWC-Net by Nvidia - SoTA on Sintel final pass

References

Amerini, Irene, Leonardo Galteri, Roberto Caldelli, and Alberto Del Bimbo. "Deepfake Video Detection through Optical Flow based CNN." In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 0-0. 2019.

Li, Lingzhi, Jianmin Bao, Ting Zhang, Hao Yang, Dong Chen, Fang Wen, and Baining Guo. "Face X-ray for More General Face Forgery Detection." arXiv preprint arXiv:1912.13458(2019).

Rossler, Andreas, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Nießner. "Faceforensics++: Learning to detect manipulated facial images." In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1-11. 2019.

Sun, Deqing, Xiaodong Yang, Ming-Yu Liu, and Jan Kautz. "Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8934-8943. 2018.

Tan, Mingxing, and Quoc V. Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." *arXiv preprint arXiv:1905.11946* (2019).

Wang, Jingdong, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu et al. "Deep high-resolution representation learning for visual recognition." *arXiv preprint arXiv:1908.07919* (2019).