Deepfake Detector

Team Faux Fighters

Cho, Anna Denq, Christopher Nelson, Reid (Jackson)





Problem Statement



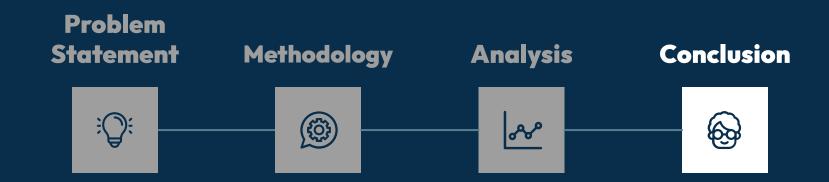
















Problem Statement



"Al is spook."

—Elon Musk, probably



PROBLEM

Generative AI makes it easy to deepfake.



PROBLEM

Generative AI makes it easy to deepfake.

Deepfake



Real



PROBLEM

Generative AI makes it easy to deepfake.

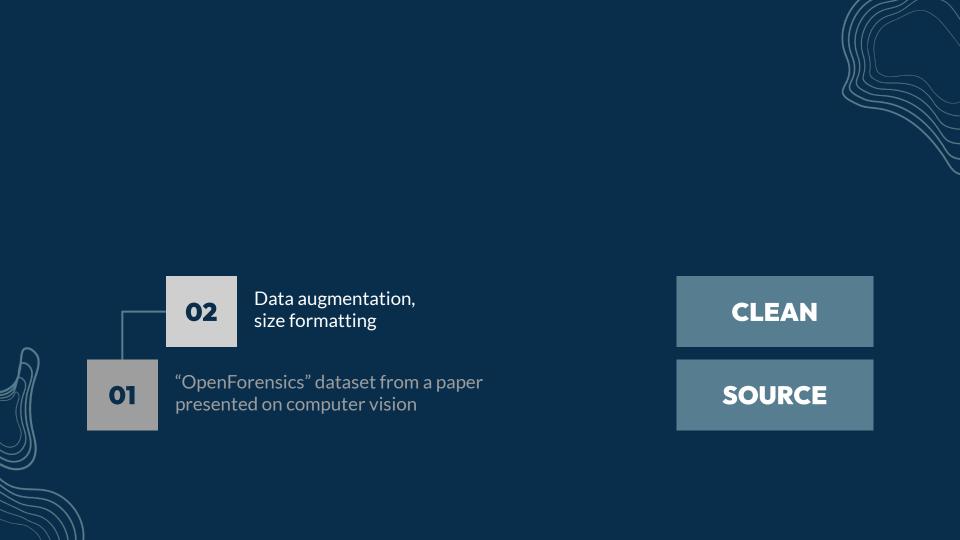
SOLUTION

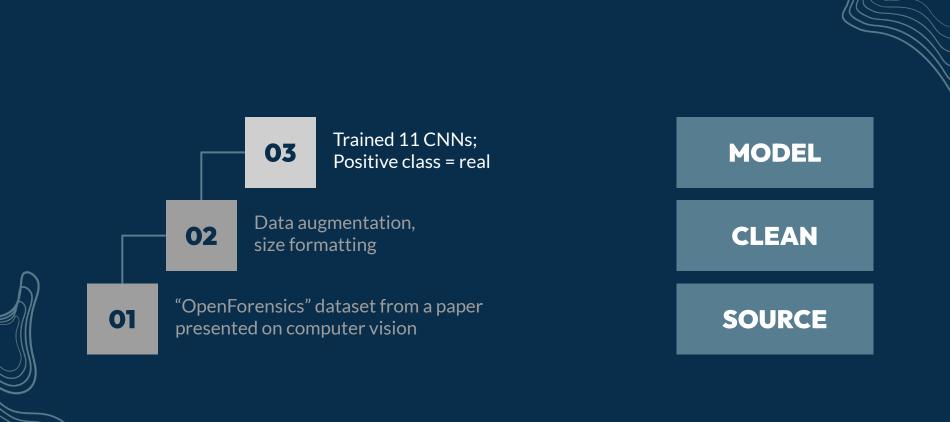
We have "out-of-box" deepfake detector.

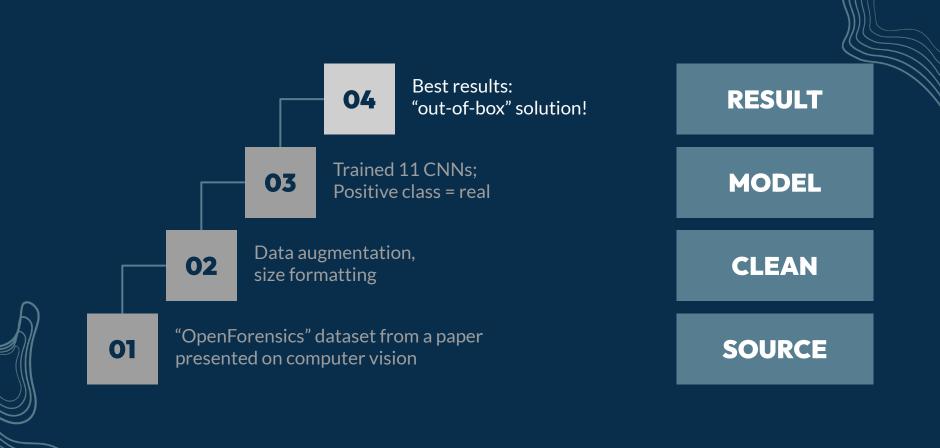










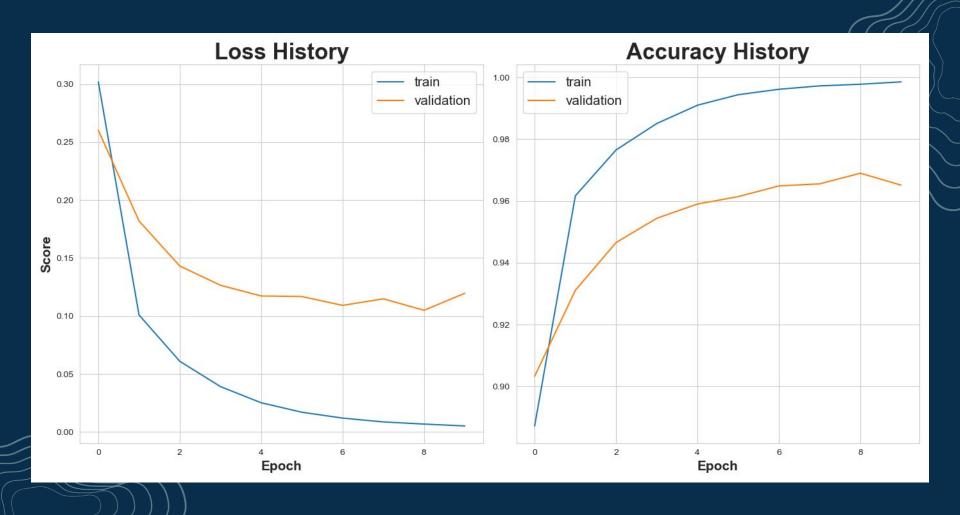


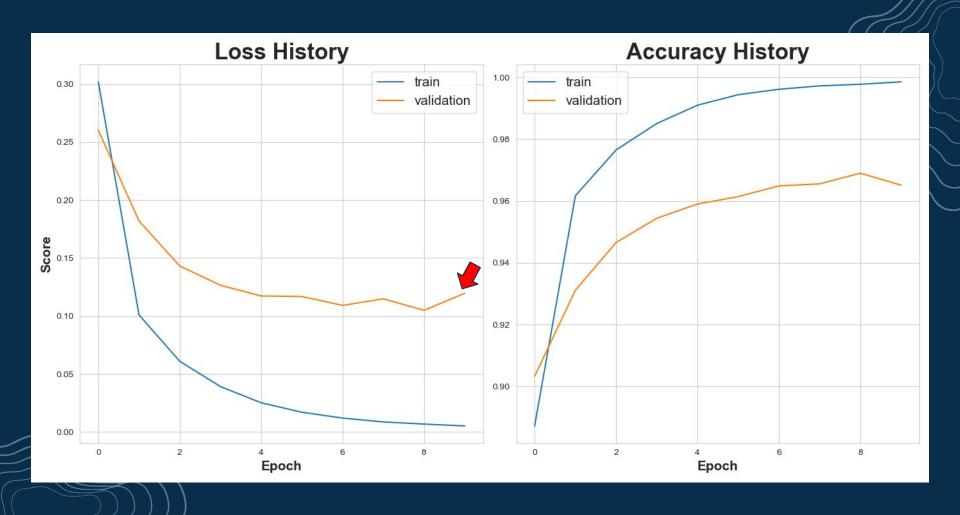
O3 ANALYSIS

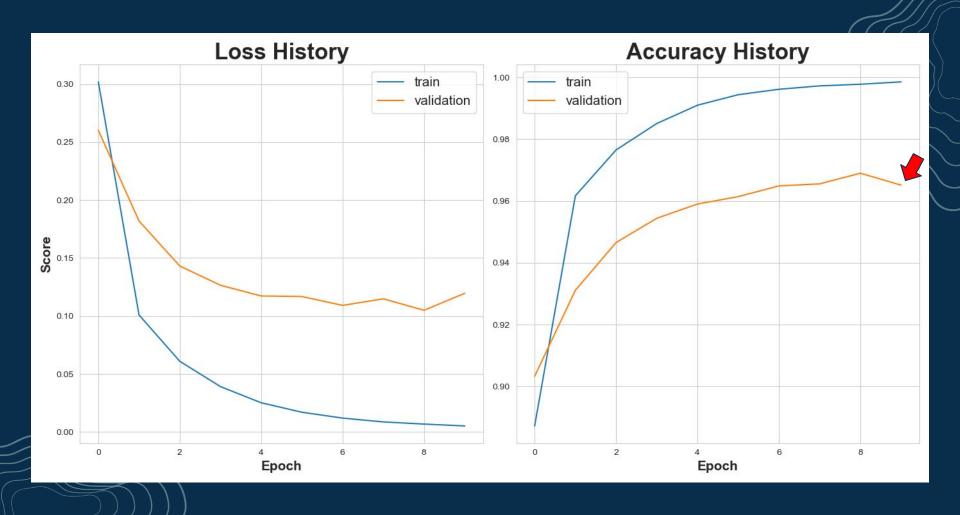


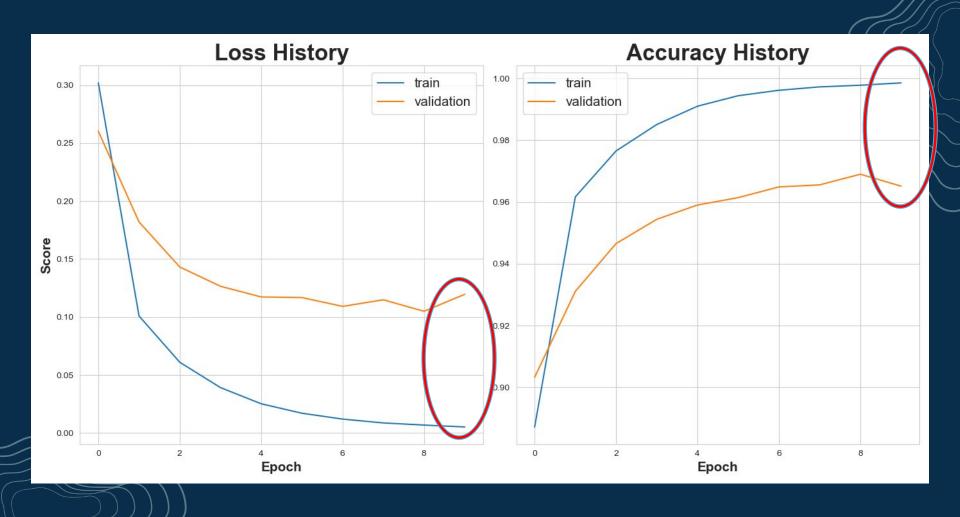






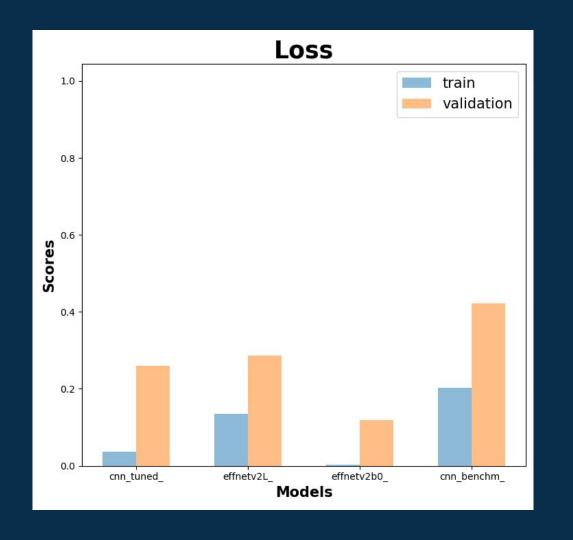




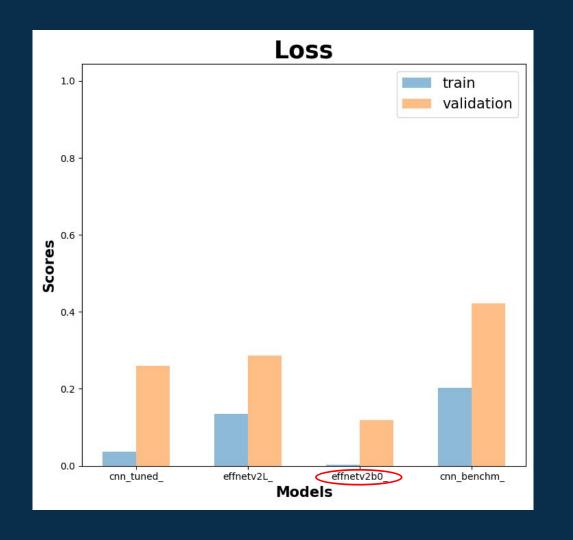




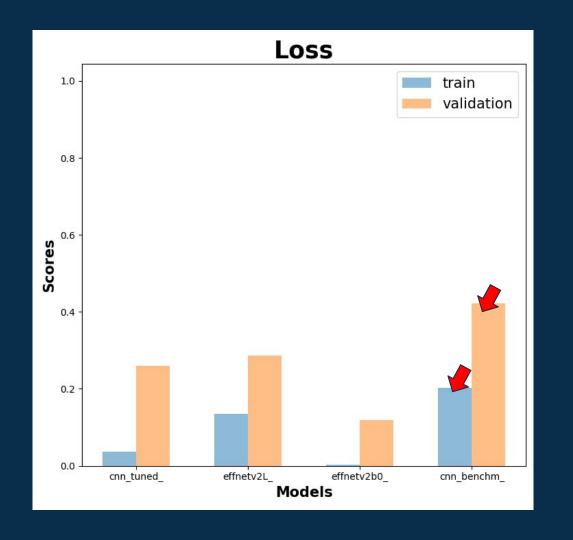




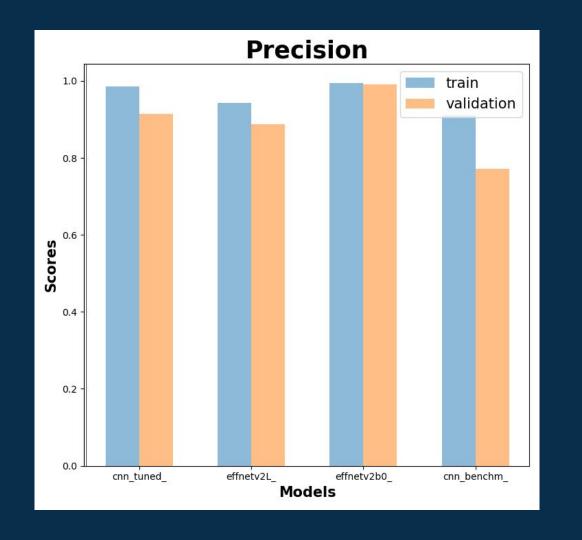




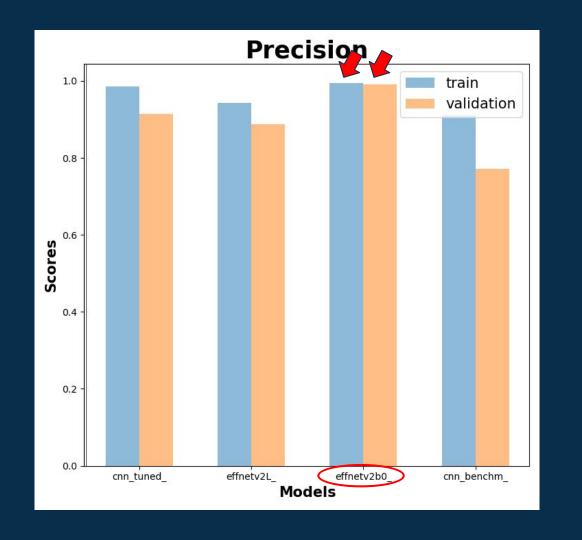




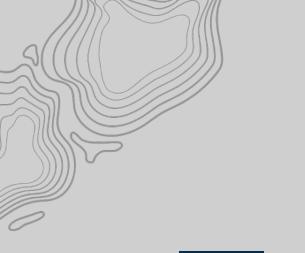
















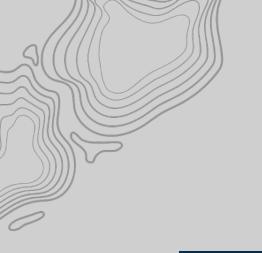






Benchmark

Experimental













Benchmark

Sequential CNN (Base) **Prec**: 0.849

Experimental













Benchmark

Sequential CNN (Base) **Prec**: 0.849

Experimental

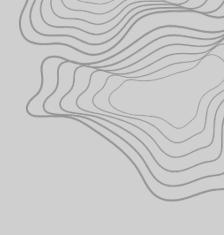
EfficientNetv2_L (TL)
Prec: 0.822







EfficientNetv2_B0 **Prec**: 0.992







Sequential CNN (Base) **Prec**: 0.849

Benchmark

Experimental

EfficientNetv2_L (TL)
Prec: 0.822





CONCLUSION

TAKEAWAYS

BEST MODEL

EfficientNetv2_B0 Precision: 0.992



TAKEAWAYS

BEST MODEL

EfficientNetv2_B0 Precision: 0.992

UNCERTAINTIES

- Generalize to other types of deepfake image generators?
- Susceptible to image manipulations (eg. blur)?



TAKEAWAYS

BEST MODEL

EfficientNetv2_B0 Precision: 0.992

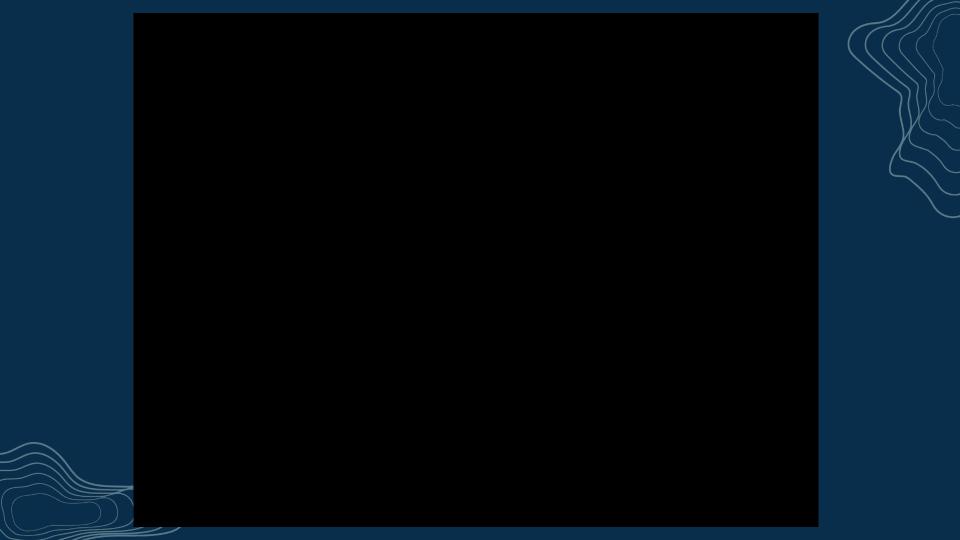
UNCERTAINTIES

- Generalize to other types of deepfake image generators?
- Susceptible to image manipulations (eg. blur)?

FUTURE

- Retrain larger EfficientNetv2 Model (eg. more compute resources)
- Address overfitting: (eg. data augmentation)

DEMOTIME



THANK YOU!

Questions?