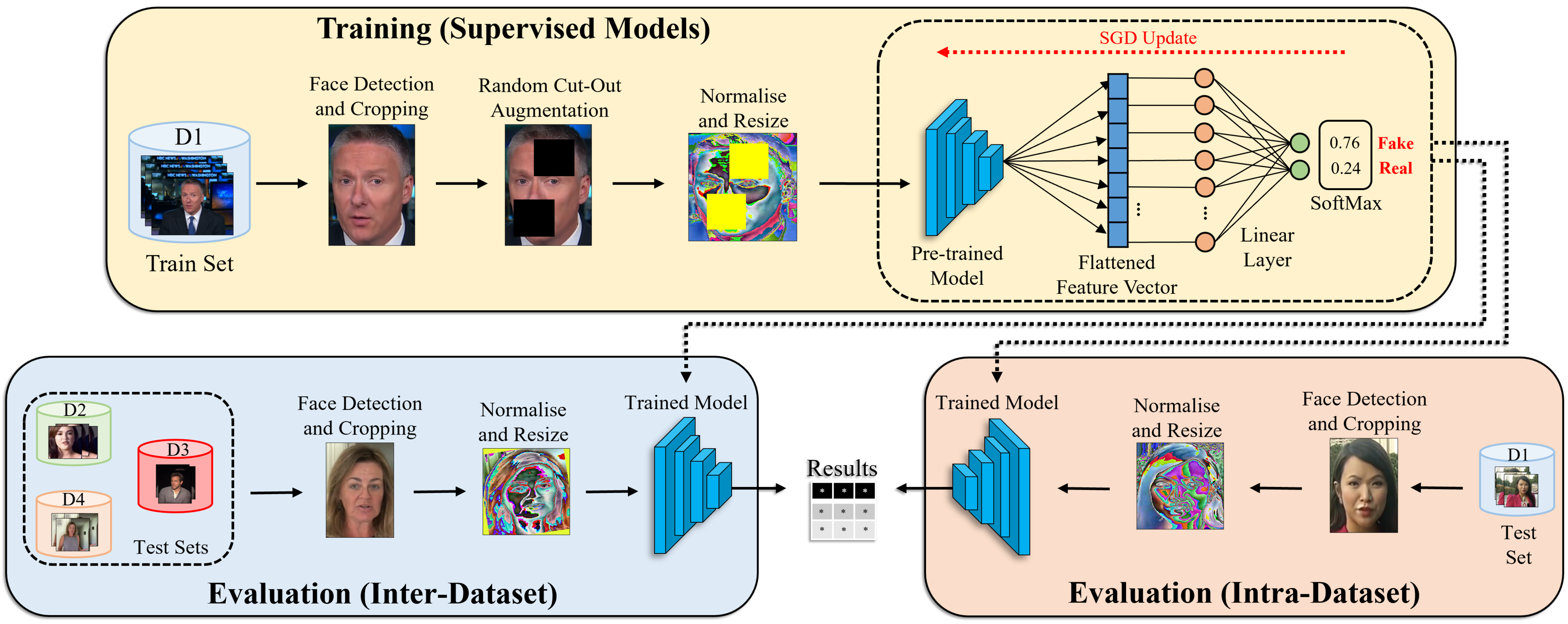
Deepfake Detection: A Comparative Analysis

Sohail Ahmed Khan and Duc-Tien Dang-Nguyen

Media Futures



Abstract

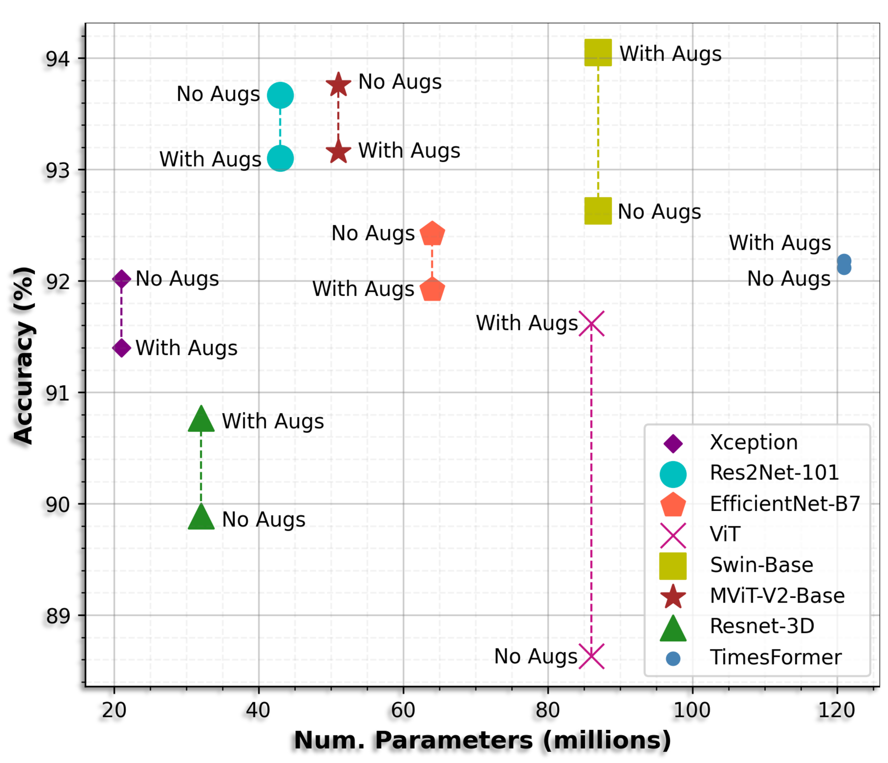
This paper present a comprehensive comparative analysis of supervised and self-supervised deep learning models for deepfake detection. We evaluate eight supervised deep learning architectures and two transformer-based models pre-trained using self-supervised strategies (DINO, CLIP) on four benchmarks (FakeAVCeleb, CelebDF-V2, DFDC, and FaceForensics++). Our analysis includes intra-dataset and inter-dataset evaluations, examining the best performing models, generalisation capabilities, and impact of augmentations. We also investigate the trade-off between model size and performance. Our main goal is to provide insights into the effectiveness of different deep learning architectures (transformers, CNNs), training strategies (supervised, self-supervised), and deepfake detection benchmarks. These insights will help guide the development of more accurate and reliable deepfake detection systems in the future, which are crucial in mitigating the harmful impact of deepfakes on individuals and society.

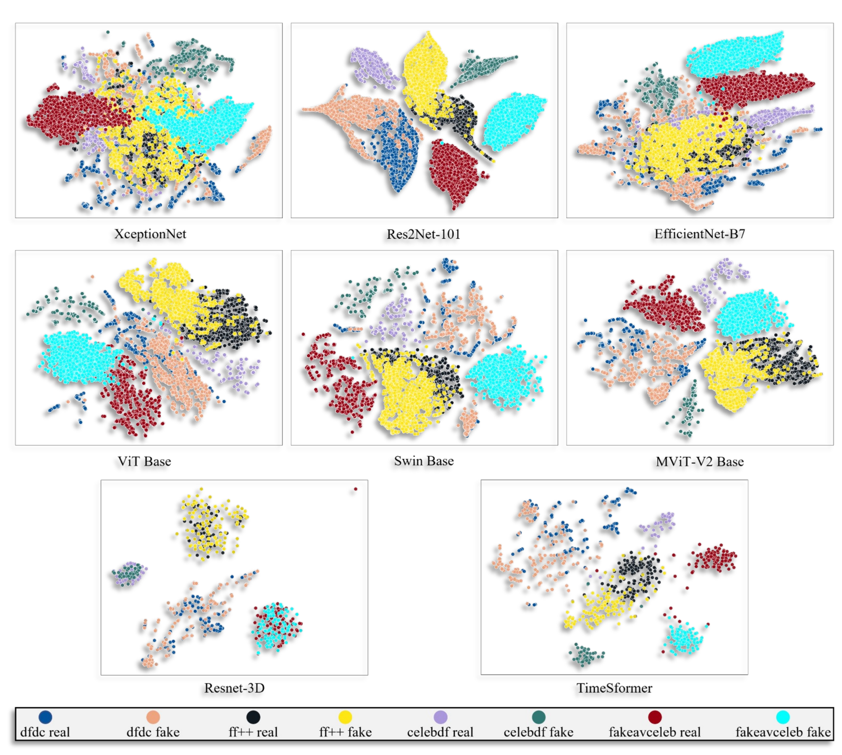
Research question

- Which architecture is most effective in detecting in-distribution deepfakes
- Which architecture posses the highest generalisation capability?
- Self-supervised vs supervised pretraining? Which is better?
- Which dataset is the most challenging for the models to learn?
- Which dataset offers best generalisation capability?
- How helpful are the image augmentations?

Conclusion

- Through experimentation and analysis of the achieved results we conclude that models possessing the capability of processing multi-scale features (Res2Net-101, MViT-V2, and Swin Transformer) achieve better overall performance in intra-dataset comparison.
- Results show that transformer models posses better generalisation capability as compared to the CNN models.
- For deepfake detection, the Self-supervised pretraining strategy (DINO) is found to be better than the supervised pretraining strategy.
- DFDC dataset is the most challenging dataset to learn for the models, whereas FaceForensics++ dataset is ranked as second most challenging dataset.
- FaceForensics++ dataset offers best generalisation capability.
- Image augmentations do not always help.





PARTNERS

amedia, Bergens Tidende, HIGHSOFT, IBM, NRK, Schibsted, TV 2, vizrt, Universitetet i Stavanger, Universitetet i Oslo, NORCE, nb.no, WOLFTECH, Faktisk.

HOST

UNIVERSITY OF BERGEN

FUNDED BY

This research is funded by SFI MediaFutures partners and the Research Council of Norway (grant number 309339).

Forskningsrådet, Norwegian Centre for Research-based Innovation