





Stronger Together? An Ensemble of CNNs for Deepfakes Detection

Angelica Gardner





Background

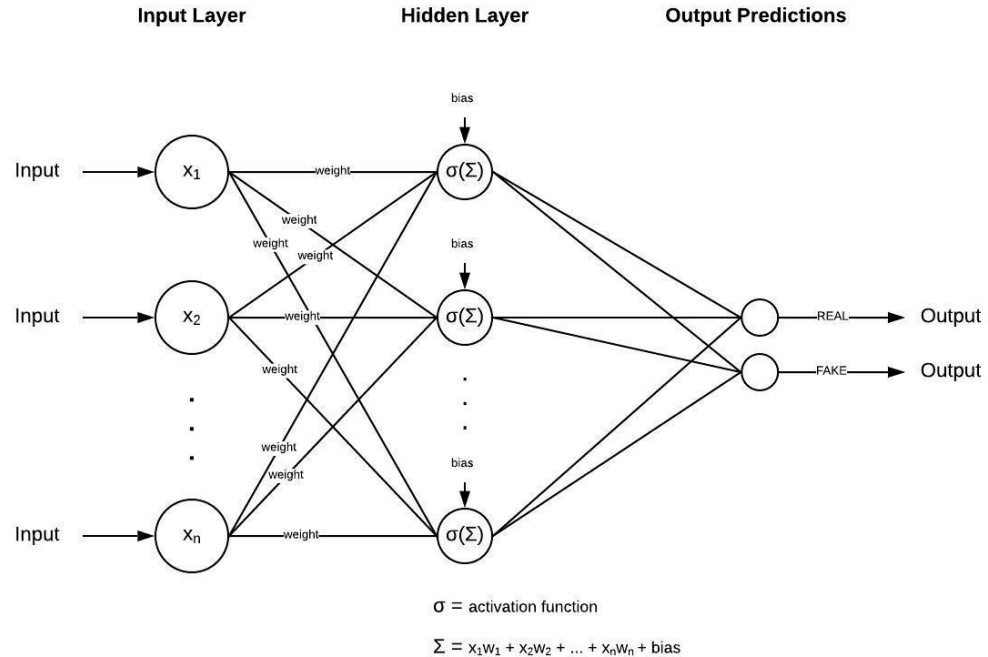
Deepfakes

- Face swap technique
- November 2017
- GitHub repository
- Colour inconsistencies, blurriness, etc.
- Improved visual quality
- Misused in vicious ways
- Propaganda purposes, fake news, attack public figures, revenge porn, etc.



Deepfakes detection

- Detection methods use ML models
- Artificial Neural Network (ANN)
- Subclass: Convolutional Neural Network (CNN)
- Input: Video
- Hidden layers: Calculations
- Output: Real / Fake





Related work

Ensemble learning

- Ensemble = a collection of ML models
- Improve accuracy performance
- Winning strategy
- Used in real-life situations

“This is how you win ML competitions: you take other peoples’ work and ensemble them together.”

- Vitaly Kuznetsov, NIPS 2014

Single models

Model name	CNN architecture	Published	Original	*
Capsule	VGG19	2019.10	94.4 - 97.6%	53.3 - 64.0%
DSP-FWA	ResNet50	2018.11	93.2 - 99.9%	64.6 - 81.1%
Ictu Oculi	VGG16	2018.06	98.0 - 99.0%	-
XceptionNet	Xception	2019.01	81.0 - 99.2%	48.2 - 53.9%

* Y. Li, X. Yang, P. Sun, H. Qi, S. Lyu, “Celeb-DF (v2): A New Dataset for DeepFake Forensics” arXiv preprint arXiv:1909.12962, 2020.



Implementation

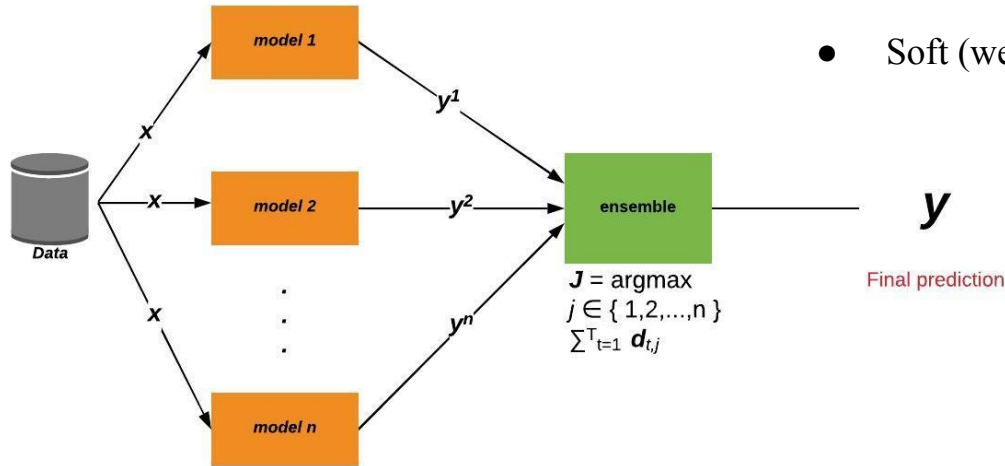
Single models

- Same configurations and training settings
- Pre-trained models
- Transfer learning: train pre-trained models some more on other data to use for the same or a similar problem
- Save new versions to use for ensemble and evaluate their performances

Model name	Research paper	GitHub repo
Capsule	Capsule-Forensics: Using Capsule Networks to Detect Forged Images and Videos	nii-yamagishilab/ Capsule-Forensics-v2
DSP-FWA	Exposing DeepFake Videos By Detecting Face Warping Artifacts	danmohaha/ DSP-FWA
Ictu Oculi	In Ictu Oculi: Exposing AI Generated Fake Face Videos by Detecting Eye Blinking	danmohaha/ WIFS2018_In_Ictu_Oculi
XceptionNet	FaceForensics++: Learning to Detect Manipulated Facial Images	ondyari/ FaceForensics

Ensembles, 1 / 2

- Stacking
- Single models make predictions
- Ensemble combines the predictions
- Hard (majority) voting
- Soft (weighted) voting



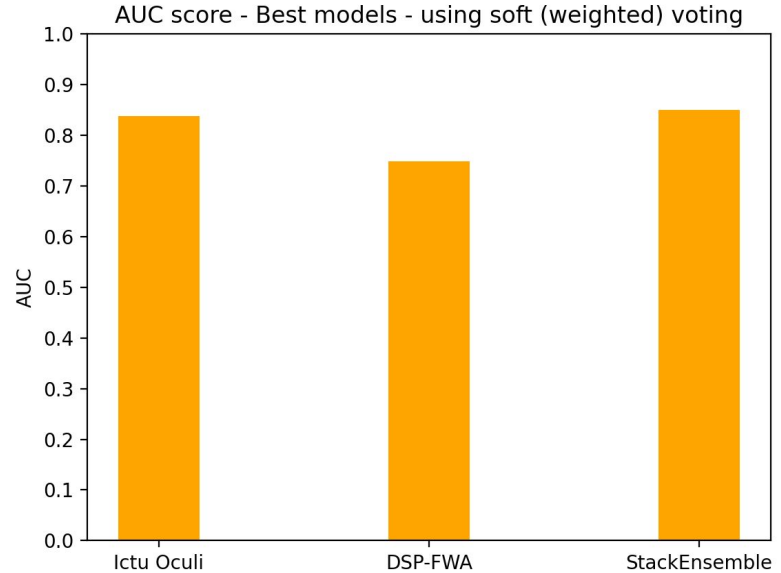
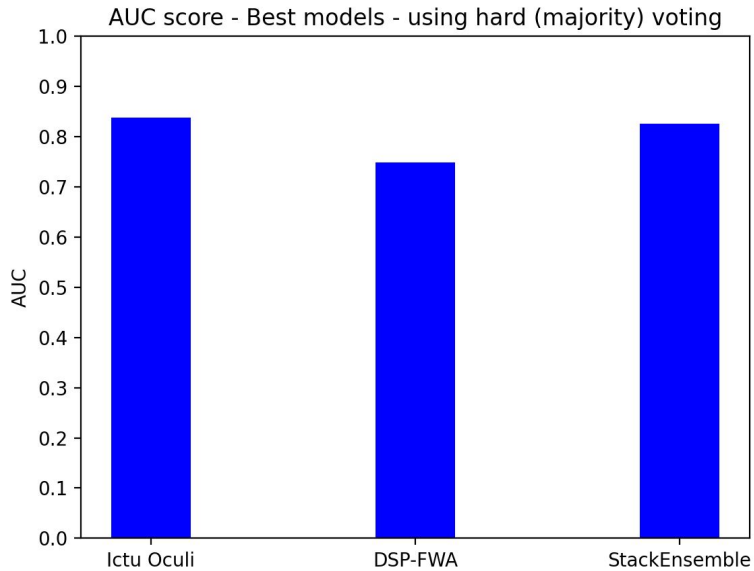
Ensembles, 2 / 2

Ensemble #	# of single models	Members	Combination
1	2	Best-performing	Hard (majority)
2	2	Best-performing	Soft (weighted)
3	2	Smallest file size	Hard (majority)
4	2	Smallest file size	Soft (weighted)
5	4	All	Hard (majority)
6	4	All	Soft (weighted)

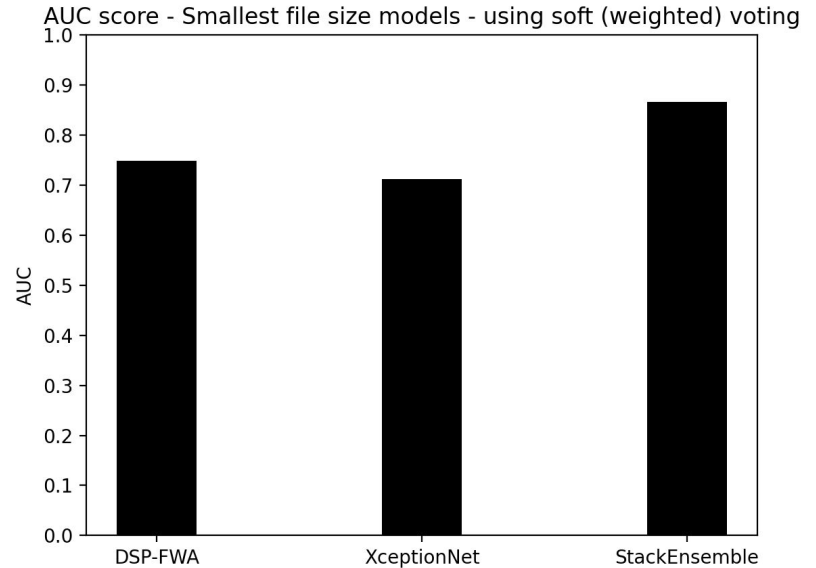
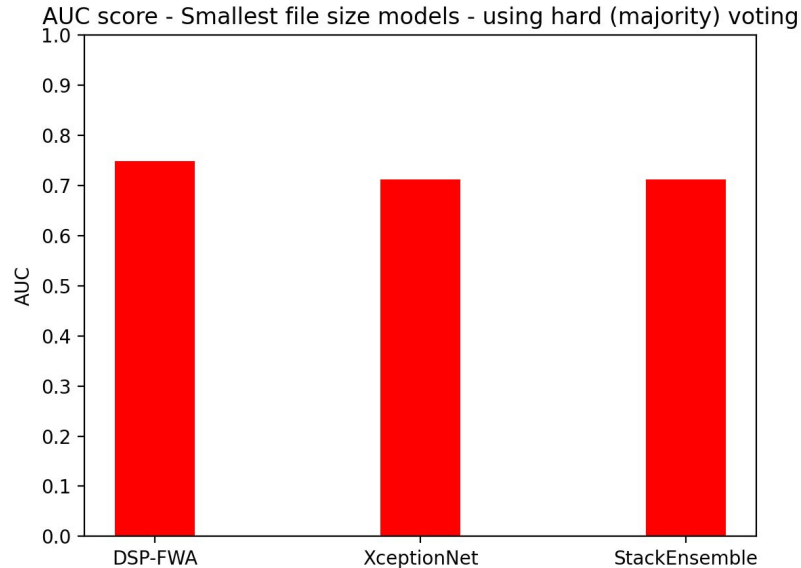


Results

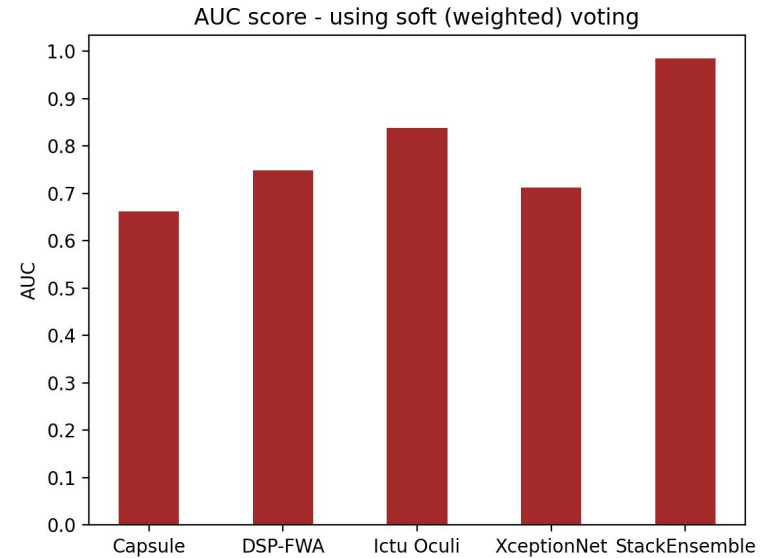
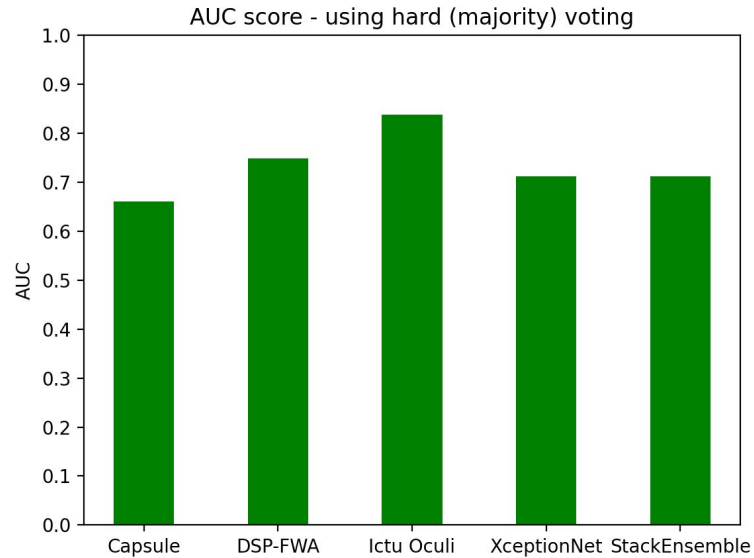
Members: Best-performing



Members: Smallest file sizes



Members: all





Conclusion

Proposed solution

Proposed Ensemble Solution	98.4%
Capsule	66.1%
DSP-FWA	74.7%
Ictu Oculi	83.8%
XceptionNet	71.1%

- Ensembles are important
- Crucial to choose the right ensemble approach
- Correspond with the results from Li et al.
- A continuing need to improve deepfakes detection methods
- Best-performing single model: Ictu Oculi
- Eye blinking is poorly reproduced in fake videos

Thank you!