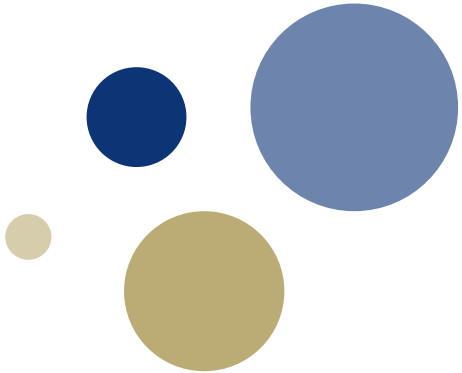
NTNU – Trondheim
Norwegian University of
Science and Technology



Automated Authentication of Audiovisual Contents: A Biometric Approach

Ali Khodabakhsh – COINS Winter School – Finse

Norwegian Biometrics Laboratory - NTNU
May 2019

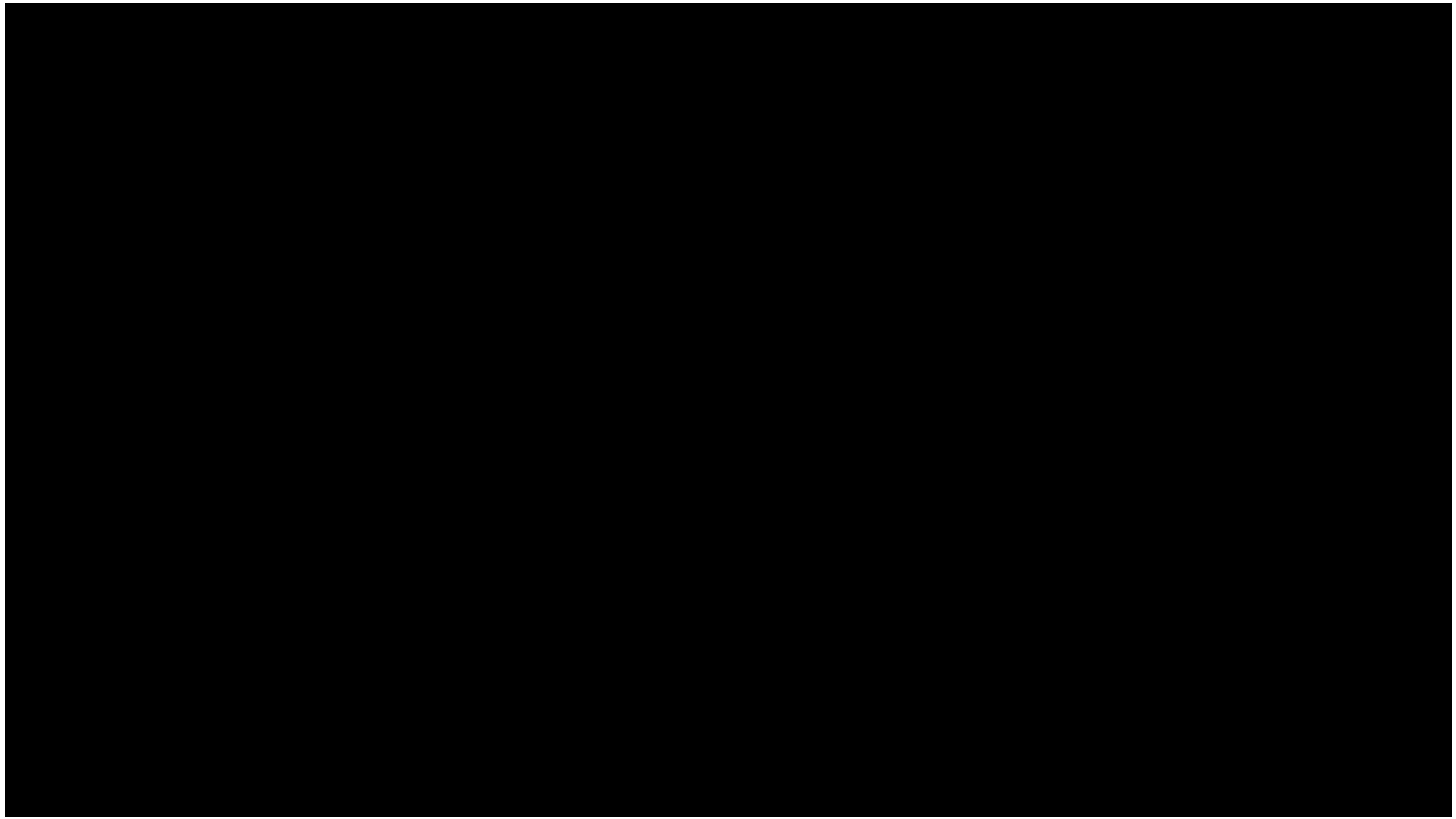
Background

- B.Sc. From University of Tehran
  - Electrical Engineering (Biomedical Engineering)
- M.Sc. From Özyeğin University
  - Computer Science (Speech Processing)
- PhD Candidate, NTNU (Since Sept. 2017)
  - Information Security (Biometrics)

Outline

- Introduction
  - Research Questions
- Research Progress
  - Taxonomy
  - Subjective Tests
  - Efficacy
- Future Work
  - In Progress

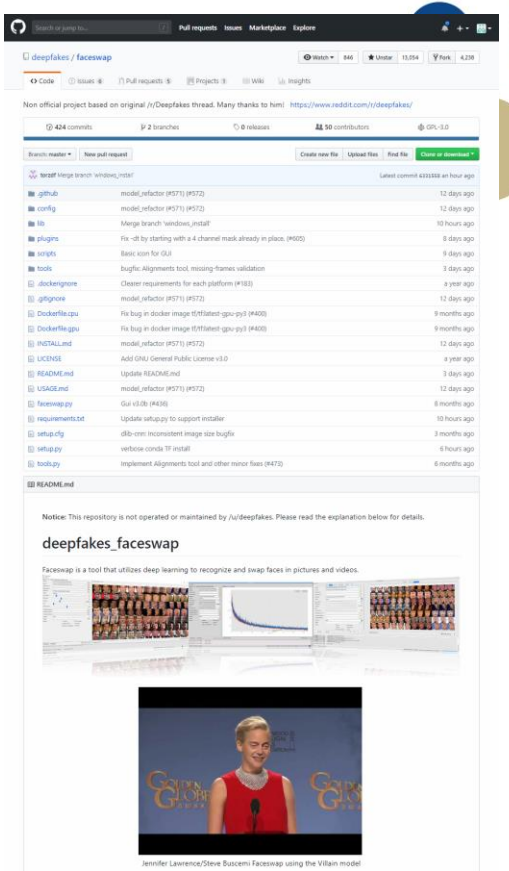
INTRODUCTION



https://youtu.be/cQ54GDm1eL0

Fake Faces

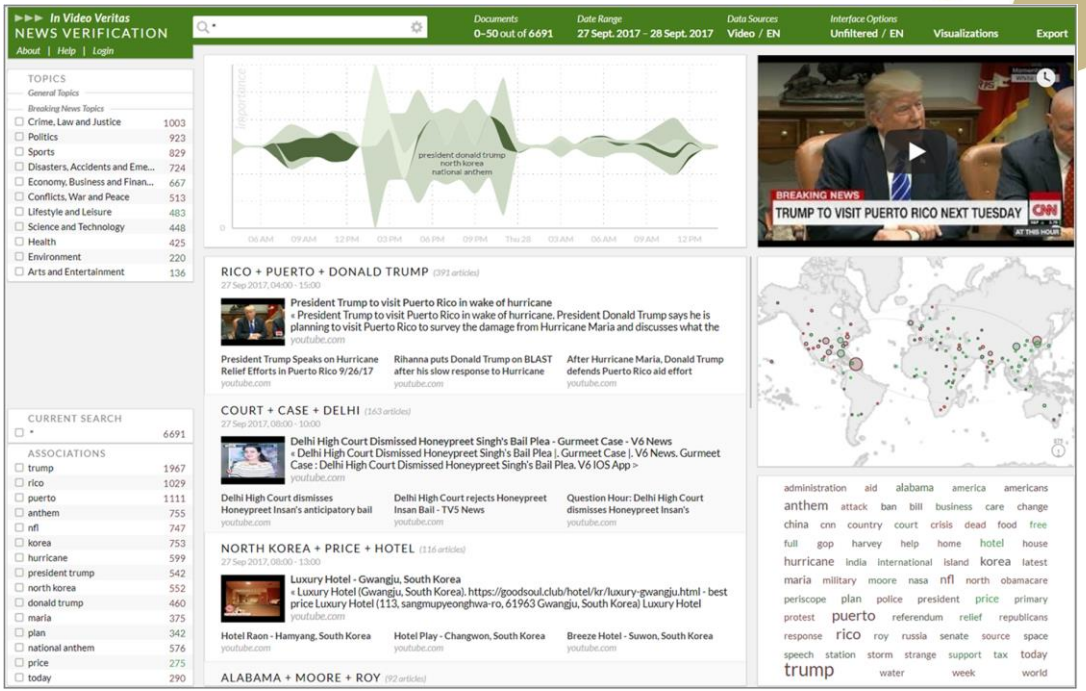
- Rapid advancements
  - Pinscreen (2016), Face2face (2016), WaveNet (2016), Synthesizing obama (2017), DeepFake (2017), Vid2vid (2018), HeadOn (2018), Deep video portraits (2018), …
- Entering age of Video-realism
  - Automatic, Cheap or Free, Publicly Available
- Vulnerability
  - Unreliable sources and echo chambers



https://github.com/deepfakes/faceswap

Automatic Detection

- Context-based
  - Source
  - Network
  - Metadata
- Content-based
  - Linguistic
  - Audiovisual



https://www.invid-project.eu/invid-multimodal-analytics-dashboard/

Existing Approaches

- CNN- and RNN-based detectors for:
  - Face Swap [1]
  - Face2Face [2,3]
  - DeepFake [4,5]
- Challenges
  - Lack of diverse large datasets
  - Incidental Manipulations (Compression, Overlay, etc)
  - Generalizability

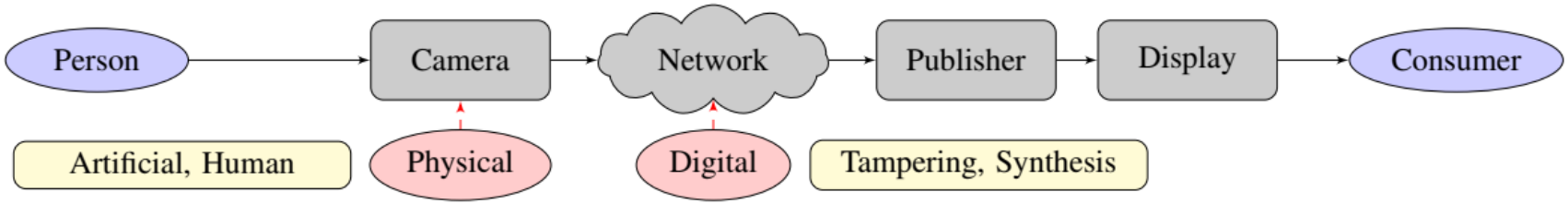
Research Questions

- Are existing fakes believable? (Done)
- Can we detect them with video-based methods? (In progress)
- Can we detect them with speech-based methods?
- Can we combine modalities?
- Can we detect using cross-modality methods?

RESEARCH PROGRESS

Taxonomy

- **Defining categories of fake faces**
- **Providing grounds for collection of datasets**



Taxonomy

		Visual	Auditory	Animation
Physical	Artificial	Androids - Screens	Biomechanical - Loudspeaker	-
	Human	Twins - Prosthetic makeup	Impersonation	Impersonation
Digital	Record-based	Image-based (Tampering)	Unit-selection (Tampering)	Cloning
	Model-based	CG (Synthesis)	SSS (Synthesis)	Autonomous

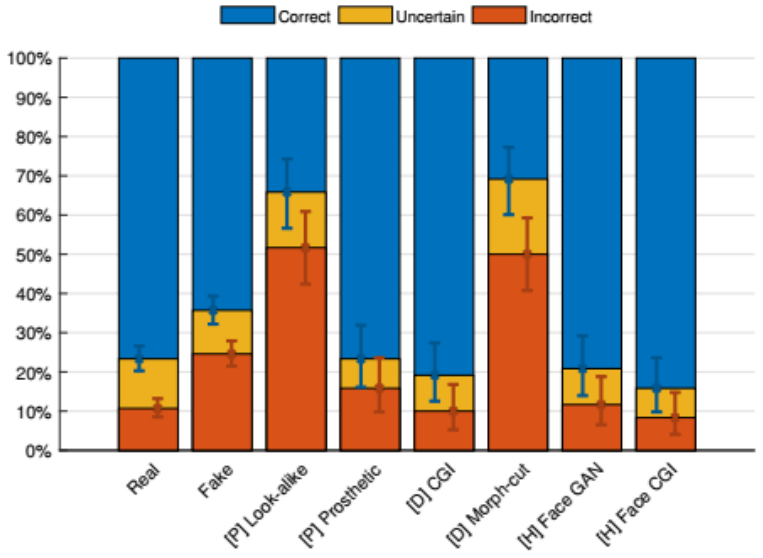
- Combination
- Obfuscation

Subjective Tests

- Assess vulnerability of humans
- Assess effectiveness of fakes
- Effects of training
- Effects of having a biometric reference
- Effects of knowing of the individual
- Common clues
- …

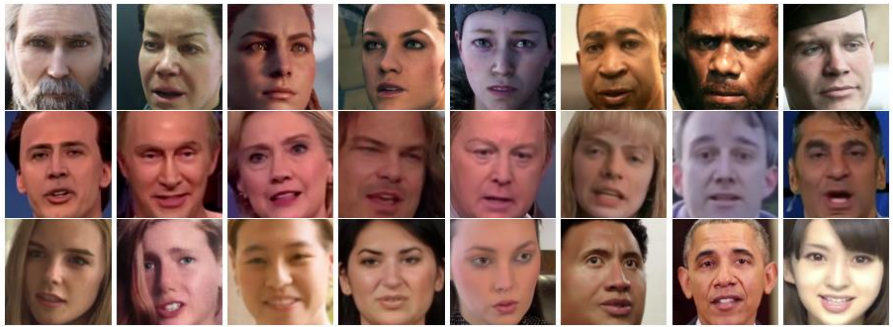
Subjective Tests

- Two types of effective fakes identified
- Head and face are the most used clues
- Training results in less mistakes
- Biometric Reference and Knowing of the target lowers uncertainty
- Older people make more mistakes and are less uncertain



Efficacy

- Evaluation of 6 detectors
  - 1 Texture-based
  - 5 Deep Learning (Transfer Learning)
- Applying the state-of-the-art to data collected in the wild
  - 150 Videos from YouTube
  - 3 Different Categories
  - Available for download at http://ali.khodabakhsh.org/ffw/



Efficacy

- Known Fakes

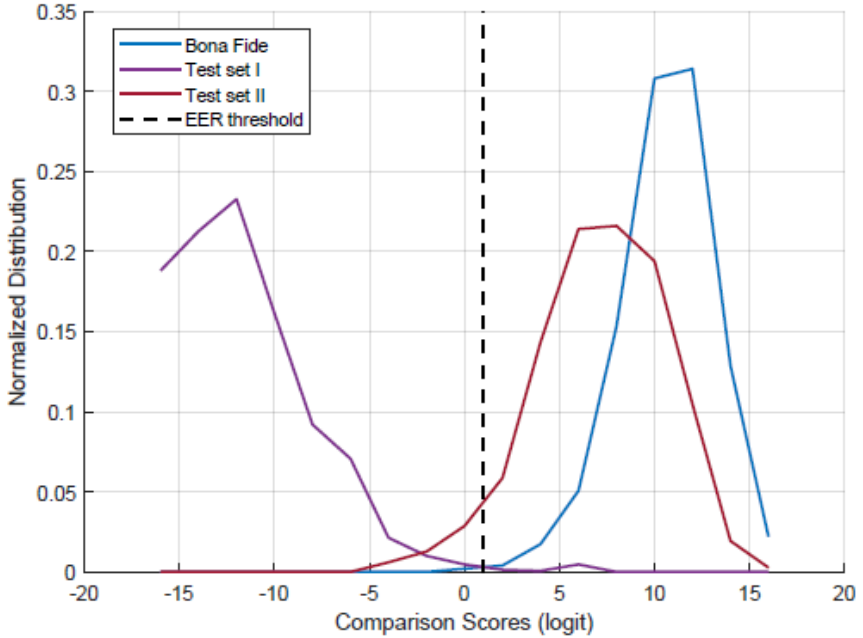
	APCER	BPCER	EER
LBP	3.80% ± 0.99%	2.87% ± 0.86%	3.33%
AlexNet	7.80% ± 1.38%	1.73% ± 0.67%	3.73%
VGG19	2.47% ± 0.80%	0.47% ± 0.35%	1.40%
ResNet	2.27% ± 0.77%	0.47% ± 0.35%	1.40%
Xception	2.47% ± 0.80%	0.13% ± 0.19%	1.07%
Inception	0.67% ± 0.42%	0.47% ± 0.35%	0.53%

- Unknown Fakes

	APCER	BPCER	EER
LBP	89.00% ± 1.62%	2.87% ± 0.86%	48.73%
AlexNet	91.47% ± 1.44%	1.73% ± 0.67%	32.13%
VGG19	90.73% ± 1.50%	0.47% ± 0.35%	29.40%
ResNet	89.53% ± 1.58%	0.47% ± 0.35%	30.33%
Xception	93.20% ± 1.30%	0.13% ± 0.19%	26.87%
Inception	91.93% ± 1.41%	0.47% ± 0.35%	27.47%

Generalizability

	Full CGI	Image Manipulation	
		FakeApp	Other
AlexNet	32.60%	28.80%	34.37%
VGG19	28.00%	31.20%	28.60%
ResNet	28.80%	28.37%	34.40%
Xception	23.60%	25.20%	31.20%
Inception	23.40%	27.40%	31.40%



IN PROGRESS AND FUTURE WORK

In Progress

- Generalizability
  - A balanced dataset of 6 categories
    - Deepfake, Morph-cut, CGI, Lookalike and Prosthetic makeup, FaceSwap, + Face2face
  - Cross category evaluation
  - Comparison with subjective performance
  - A new loss function
- Encyclopedia of fake faces and voices
  - Reference for all known types of fakes + datasets

Future Work

- Incorporating temporal data
- Repeat for speech
- Encoding Decoding theory of Communication

Extended Subjective Test

- Browser-based:
  - Simulates real-life encounter
  - Takes ~40 minutes
- Includes:
  - Feedback and training

http://test.khodabakhsh.org

Thank you for your attention. Questions? Comments? Suggestions?

Ali Khodabakhsh (ali.Khodabakhsh@ntnu.no)

Test: http://test.khodabakhsh.org/
Dataset: http://www.khodabakhsh.org/ffw

References

- [1] Zhou, Peng, et al. "Two-stream neural networks for tampered face detection." *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2017.
- [2] Rössler, Andreas, et al. "FaceForensics++: Learning to Detect Manipulated Facial Images." *arXiv preprint arXiv:1901.08971*(2019).
- [3] Afchar, Darius, et al. "Mesonet: a compact facial video forgery detection network." *2018 IEEE International Workshop on Information Forensics and Security (WIFS)*. IEEE, 2018.
- [4] Güera, David, and Edward J. Delp. "Deepfake Video Detection Using Recurrent Neural Networks." *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. IEEE, 2018.
- [5] Korshunov, Pavel, and Sébastien Marcel. "DeepFakes: a New Threat to Face Recognition? Assessment and Detection." *arXiv preprint arXiv:1812.08685* (2018).