University of Notre Dame

Grains of Saliency: Optimizing Saliency-based Training of Biometric Attack Detection Models

Colton Crum, **Samuel Webster** (presenting), Adam Czajka
contact: {ccrum,swebster}@nd.edu, https://cvrl.nd.edu

COMPUTER VISION @ND

Abstract

In response to the lack of research in saliency optimization as well as the expense of collecting high-fidelity human saliency, we explore several granularities of salience information based on saliency collected from humans. By applying these levels of saliency to iris and face presentation attack detection (PAD) tasks, we demonstrate that increased generalization capabilities of PAD can be achieved by using simple yet effective saliency post-processing techniques across several different Convolutional Neural Networks.

Salience Granularity
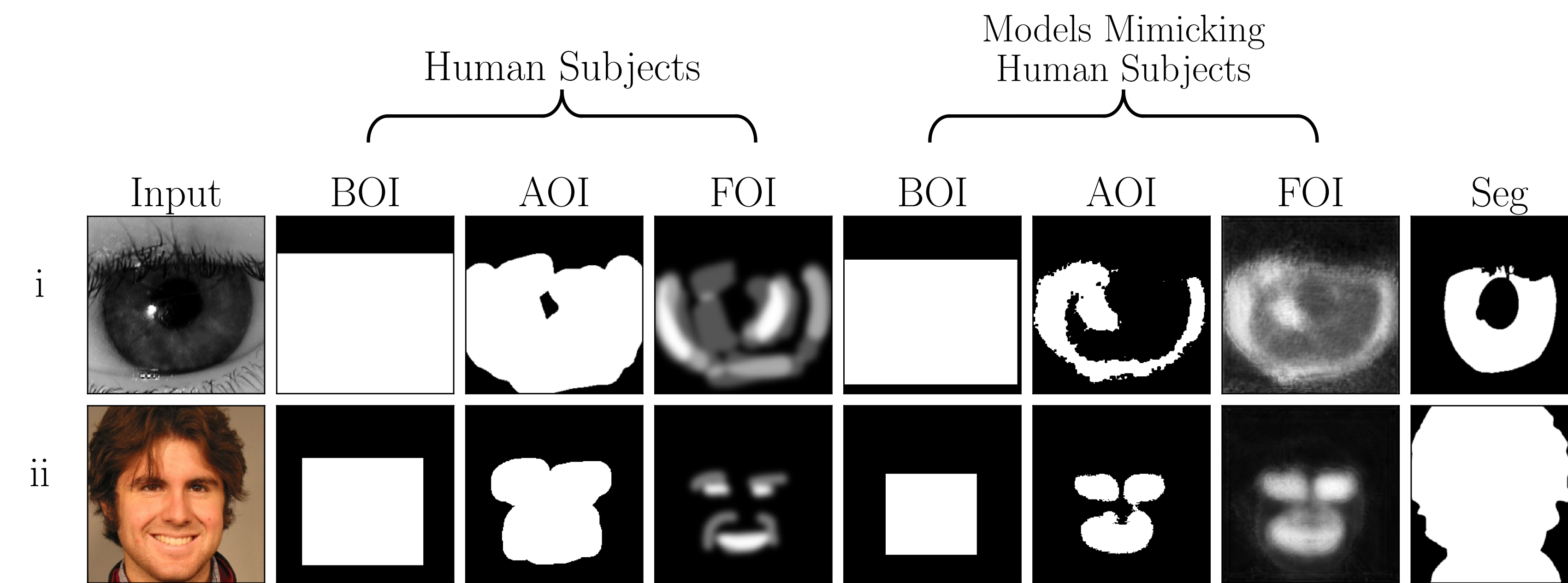
Explored Granularities:

Features of Interest (FOI): Directly sourced from human/mimicking model
Area of Interest (AOI): Generated by binarizing FOI saliency
Boundary of Interest (BOI): Generated by minimally enclosing AOI saliency
Segmentation (Seg): Sourced from SOTA iris/face segmenter

Conclusion

- **RQ1**: **Area of Interest (AOI)** Saliency is the optimal granularity for iris-PAD.

- **RQ2**: **No**, optimal saliency does not generalize across biometric modalities.

- **RQ3**: **Yes**, models mimicking human saliency provide better generalizing saliency.

- **RQ4**: **No**, saliency is best sourced from human or human-inspired methods.

Research Questions

- **RQ1**: What is the optimal granularity of human saliency maps for saliency-based training of models detecting biometric spoofs?

- **RQ2**: Does the optimal level of granularity generalize across different biometric PAD modalities?

- **RQ3**: Does training with saliency sourced from models trained to mimic human saliency lead to better generalization?

- **RQ4**: Can saliency be sourced from domain-specific segmentation models instead of humans?

Results

Iris Presentation Attack Detection

Source of Saliency	ResNet	DenseNet	Inception	Average
	Backbones Used in Saliency-Based Training			
Human Subjects				
Boundary of Interest (BOI)	0.886±0.015	0.903±0.010	0.873±0.023	0.887±0.016
Area of Interest (AOI)	0.909±0.006	0.921±0.013	0.900±0.005	0.910±0.008
Features of Interest (FOI)	0.908±0.005	0.895±0.018	0.890±0.015	0.898±0.013
Models Mimicking Human Subjects				
Boundary of Interest (BOI)	0.939±0.008	0.933±0.016	0.953±0.007	0.942±0.010
Area of Interest (AOI)	**0.956±0.006**	**0.962±0.005**	**0.962±0.013**	0.960±0.008
Features of Interest (FOI)	0.945±0.007	0.955±0.003	0.958±0.007	0.953±0.006
Segmentation Models				
Iris Segmentations	0.894±0.010	0.884±0.004	0.878±0.022	0.885±0.012
None				
Baseline	0.875±0.013	0.893±0.019	0.889±0.006	0.886±0.010

Generalization performance in **iris-PAD** across varying model architectures, sources of saliency, and salience granularities, reported over **3 independent runs**.
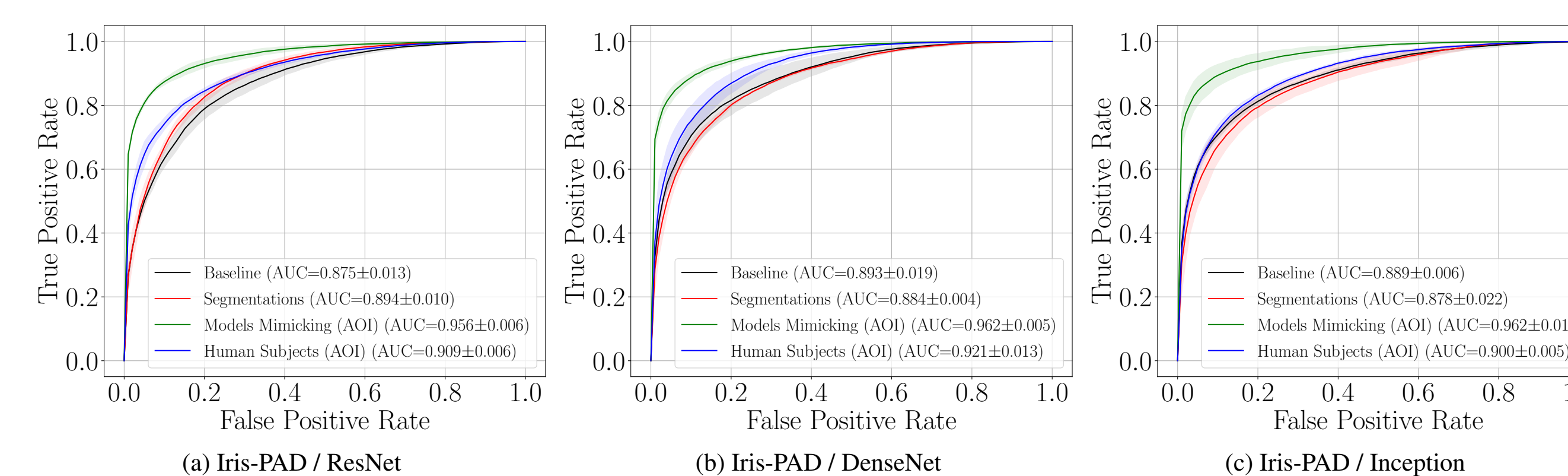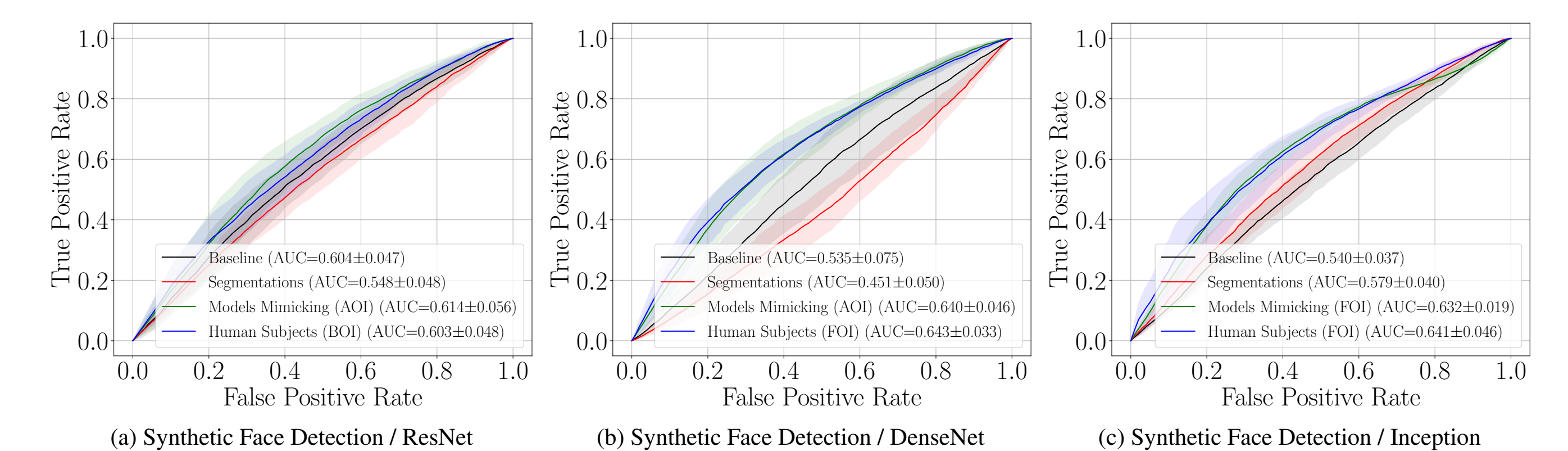

Mean ROC curves and bands representing standard deviations (along the True Positive Rate axis) for all backbones used in saliency-based training with varied configurations of saliency for **iris-PAD**.

Synthetic Face Detection

Source of Saliency	ResNet	DenseNet	Inception	Average
	Backbones Used in Saliency-Based Training			
Human Subjects				
Boundary of Interest (BOI)	0.604±0.048	0.546±0.059	0.617±0.062	0.589±0.056
Area of Interest (AOI)	0.579±0.035	0.577±0.045	0.639±0.029	0.598±0.036
Features of Interest (FOI)	0.590±0.023	0.643±0.033	**0.641±0.046**	0.629±0.027
Models Mimicking Human Subjects				
Boundary of Interest (BOI)	0.584±0.031	0.583±0.054	0.539±0.034	0.569±0.040
Area of Interest (AOI)	**0.614±0.056**	**0.640±0.046**	0.608±0.071	0.621±0.058
Features of Interest (FOI)	0.600±0.025	0.619±0.033	0.632±0.019	0.617±0.026
Segmentation Models				
Face Segmentations	0.548±0.048	0.451±0.050	0.579±0.040	0.526±0.046
None				
Baseline	0.572±0.047	0.535±0.075	0.540±0.037	0.549±0.053

Generalization performance in **synthetic face detection** across varying model architectures, sources of saliency, and salience granularities, reported over **5 independent runs**.

Mean ROC curves and bands representing standard deviations (along the True Positive Rate axis) for all backbones used in saliency-based training with varied configurations of saliency for **synthetic face detection**.