Sony AI

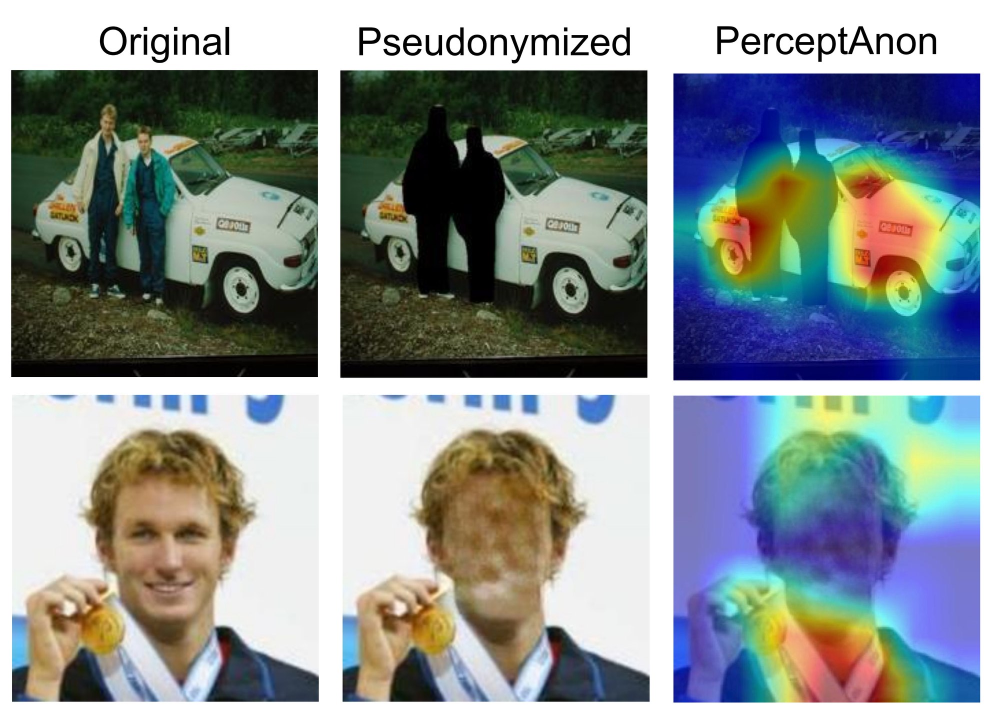
PerceptAnon: Exploring the Human Perception of Image Anonymization beyond Pseudonymization for GDPR

Kartik Patwari*, Chen-Nee Chuah, Lingjuan Lyu, Vivek Sharma*

ICML International Conference On Machine Learning

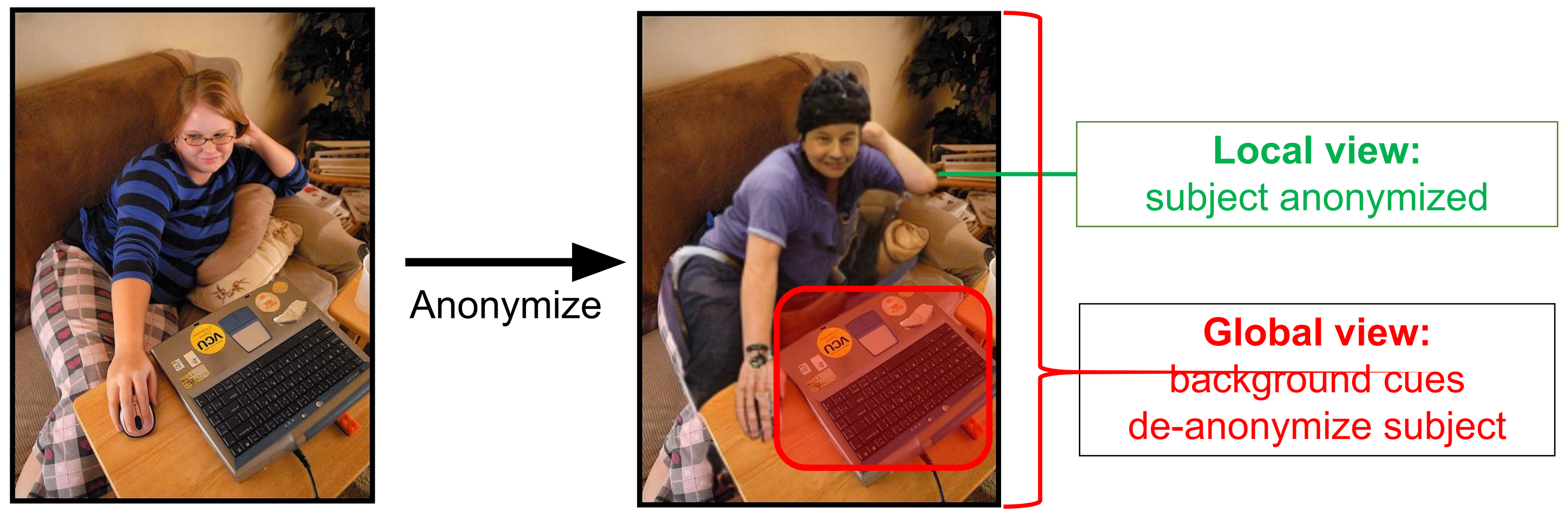
Task Definition

- ➢ General Data Protection Regulation (GDPR): "*the use of additional information can lead to identification of individuals*"
- ➢ PerceptAnon (**Percept**ion of **Anon**ymization) identifies and quantifies privacy-compromising cues in face/full body pseudonymized images.



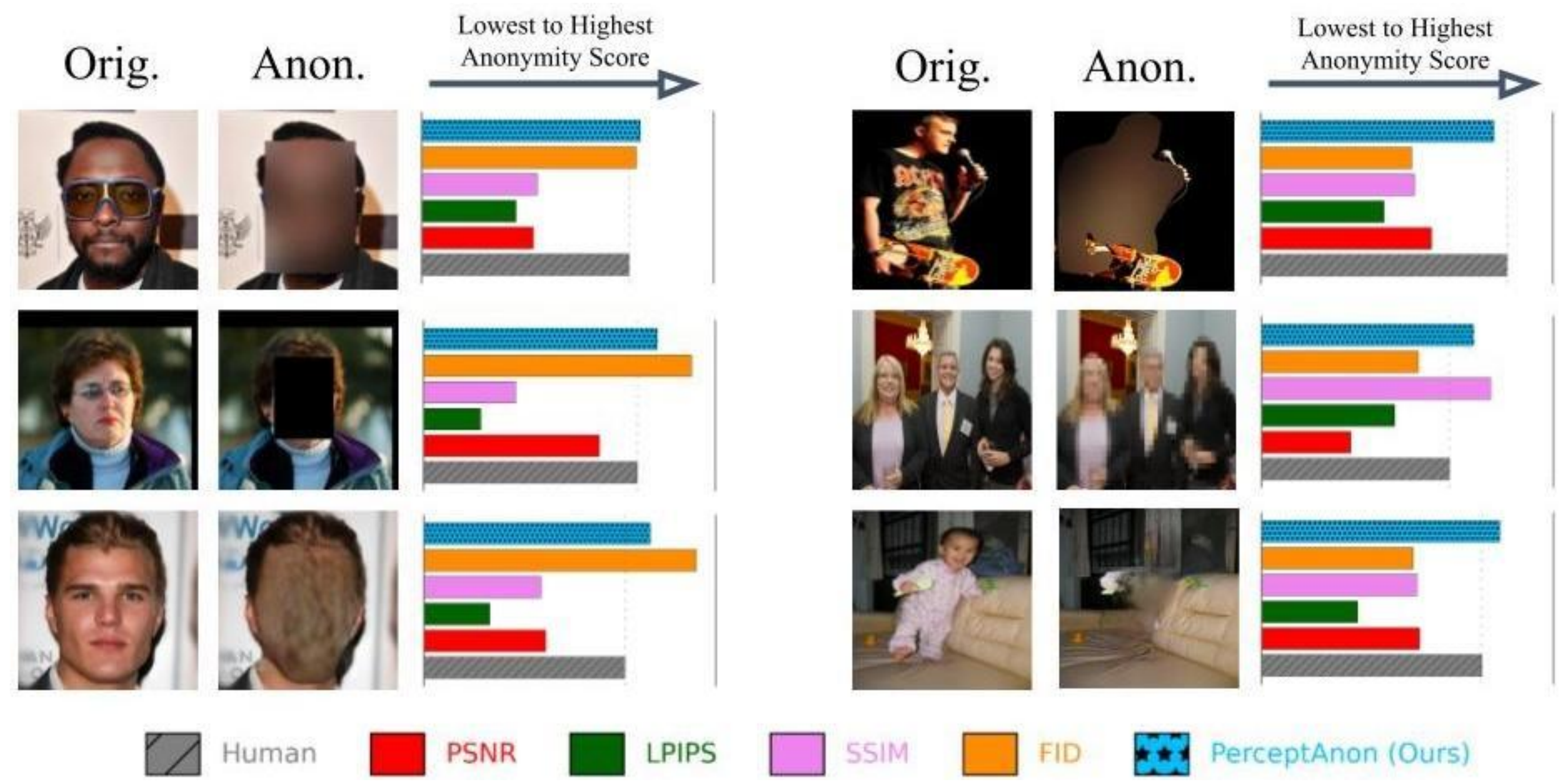
Motivation

Are images fully anonymized according to humans?



- Current anonymization techniques focus on **local** face/full body **pseudonymization.**
- This can leave privacy compromising background cues **globally**.
- Humans interpret images wholistically to identify and assess such cues.

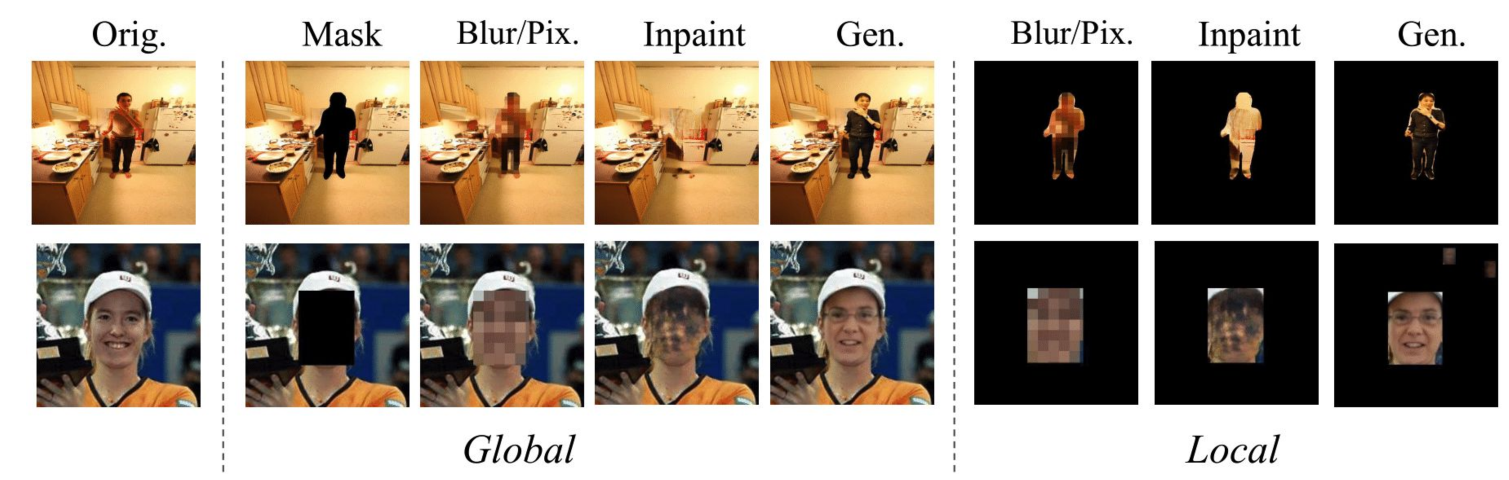
Existing global image metrics and their alignment with human perception



- Existing metrics do not correlate well with human perception.

Proposed metric

New dataset for face/full-body anonymity

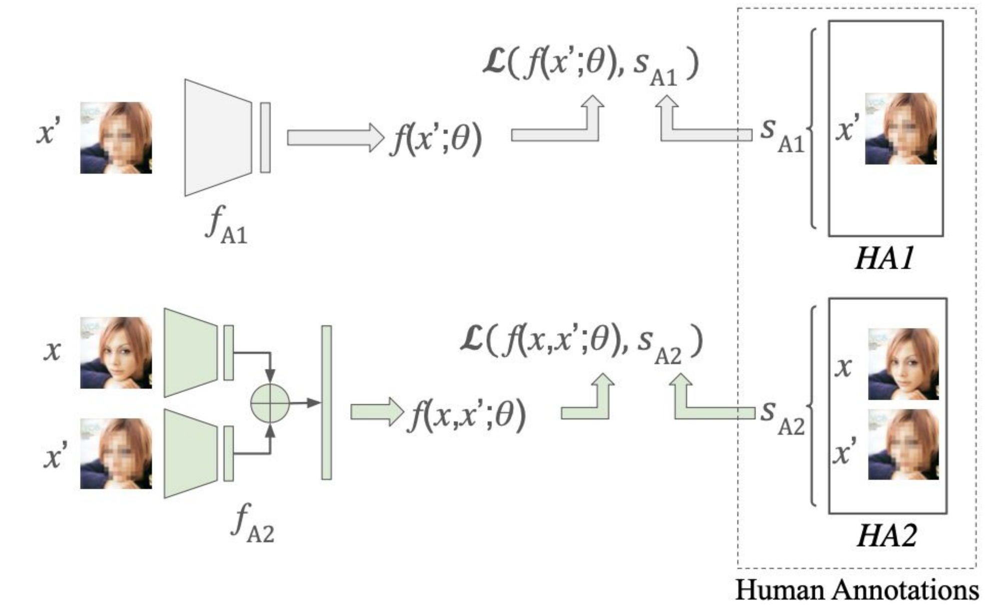


Annotations score 1-10 on degree of anonymity achieved. Two human annotation setups:
- *HA1:* annotators only see anonymized image
- *HA2:* annotators original and anonymized image pairs

PerceptAnon metric

- *HA1:* CNN with anonymized image input (x')
- *HA2:* Siamese network with original-anonymized image pair input (x,x')

Trained using both classification (CE) and regression (MSE) loss



Main Results

Comparison of different metrics on our dataset train/test splits

HA1

Train/Test Setup	Metrics	PSNR	MSE	LPIPS	SSIM	FID	PerceptAnon (Ours)
All	ρ	-0.7011	0.7011	0.7675	-0.8358	0.6578	**0.8817**
	τ	-0.5018	0.5018	0.5544	**-0.7601**	0.4667	0.7119
LOOV-VOC	ρ	-0.7448	0.7448	0.8244	-0.8185	0.6995	**0.8603**
	τ	-0.5437	0.5437	0.6288	-0.6289	0.5095	**0.6570**
LOOV-COCO	ρ	-0.771	0.771	0.805	-0.7702	0.733	**0.8643**
	τ	-0.5649	0.5649	0.6051	-0.5712	0.5385	**0.6845**
LOOV-LFW	ρ	-0.7354	0.7354	0.7574	-0.7615	0.7289	**0.8278**
	τ	-0.5256	0.5256	0.5487	-0.5509	0.5141	**0.6353**
LOOV-CelebA	ρ	-0.6239	0.6239	0.7301	-0.7321	0.6634	**0.8478**
	τ	-0.4407	0.4407	0.5151	-0.518	0.4594	**0.6549**
Task-Person	ρ	-0.7313	0.7313	0.7909	-0.75	0.6858	**0.8831**
	τ	-0.524	0.524	0.5929	-0.5452	0.4971	**0.7120**
Task-Face	ρ	-0.7547	0.7547	0.7906	-0.7838	0.7447	**0.8774**
	τ	-0.5528	0.5528	0.5887	-0.5825	0.547	**0.6940**

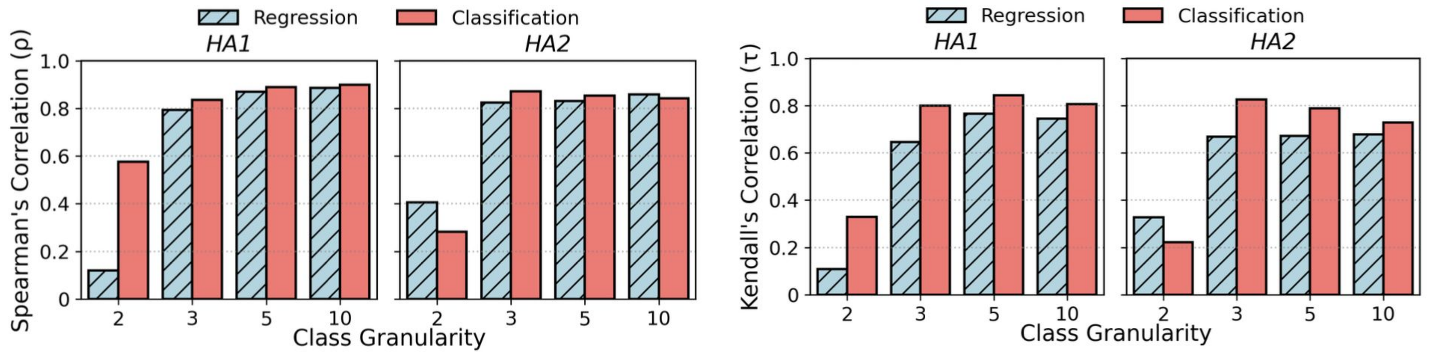
HA2

Train/Test Setup	Metrics	PSNR	MSE	LPIPS	SSIM	FID	PerceptAnon (Ours)
All	ρ	-0.7631	0.7631	0.7622	-0.7655	0.6444	**0.8421**
	τ	-0.5434	0.5434	0.5385	-0.5448	0.4456	**0.6477**
LOOV-VOC	ρ	-0.7833	0.7833	0.7869	-0.7971	0.6203	**0.8218**
	τ	-0.575	0.575	0.5694	-0.5827	0.4338	**0.6211**
LOOV-COCO	ρ	-0.7941	0.7941	0.7851	-0.785	0.6478	**0.8404**
	τ	-0.5842	0.5842	0.5713	-0.5739	0.4559	**0.6456**
LOOV-LFW	ρ	-0.7551	0.7551	0.7137	-0.7358	0.7032	**0.8462**
	τ	-0.5243	0.5243	0.4683	-0.5001	0.4536	**0.6495**
LOOV-CelebA	ρ	-0.7157	0.7157	0.7082	-0.7542	0.679	**0.8250**
	τ	-0.4875	0.4875	0.4753	-0.5354	0.4569	**0.6270**
Task-Person	ρ	-0.7757	0.7757	0.7833	-0.7997	0.6408	**0.8320**
	τ	-0.5647	0.5647	0.5668	-0.5872	0.4477	**0.6328**
Task-Face	ρ	-0.7435	0.7435	0.6956	-0.7623	0.6756	**0.8590**
	τ	-0.5154	0.5154	0.4537	-0.5387	0.4454	**0.6675**

- PerceptAnon has stronger correlation to human perception than existing metrics.

Analysis of PerceptAnon

How to best align PerceptAnon with respect to human perception?

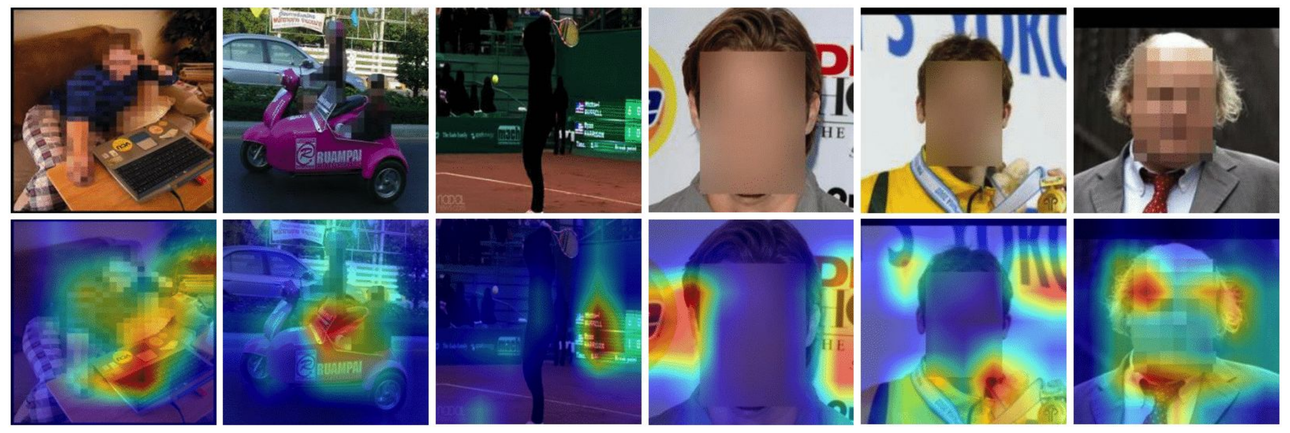


PerceptAnon's faithfulness to human perception under:
- Varying class granularities
- Regression vs. classification training strategy

Correlation between *HA1* and *HA2*

Model	*HA1*	*HA2*	$\rho(HA1, HA2)$
ResNet18	0.8752	0.8345	0.9050
ResNet50	0.8912	0.8346	0.8848

PerceptAnon GRAD-CAM visualizations



- PerceptAnon focuses on remaining potential privacy compromising cues

Future work

- ➢ Extend PerceptAnon to consider different characteristics like medical images.
- ➢ Develop anonymization techniques that consider full anonymity including residual privacy compromising cues in anonymized image background.

Conclusion

- ➢ Image anonymity often equates to pseudonymity; no current metric addresses global image privacy, including residual background cues.
- ➢ We propose novel annotation and evaluation setups to study and understand image privacy from a human perspective.
- ➢ We introduce **PerceptAnon**, a learning-based metric that better aligns with human perception of global image privacy.
- ➢ PerceptAnon not only considers original-anonymized image pairs but also sole anonymized images, mimicking human perception.

Contact:
- kpatwari@ucdavis.edu
- viveksharma@sony.com

Project

Github

Reference:

10 Misunderstanding Related To Anonymisation, European Union, EDPS. 2021