How Private is Private?
A Comparative Study for Face De-Identification

Authors: Hui Wei[1,2], Hao Yu[1,2], Hui Kuurila-Zhang[2], Guoying Zhao[1,2]

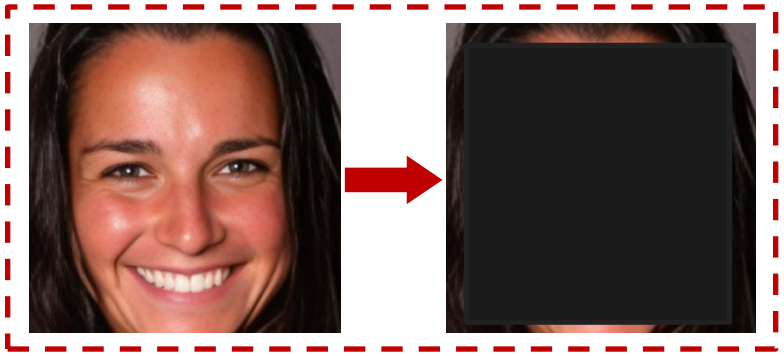
Affiliations: [1]ELLIS Institute Finland, Finland [2]Center for Machine Vision and Signal Analysis (CMVS), University of Oulu, Finland

Venue: NeurIPS 2026 · E&D Track

NEURAL INFORMATION PROCESSING SYSTEMS

What is privacy

Privacy IS NOT about keeping information secret, BUT about ensuring that personal information flows appropriately within its social context.



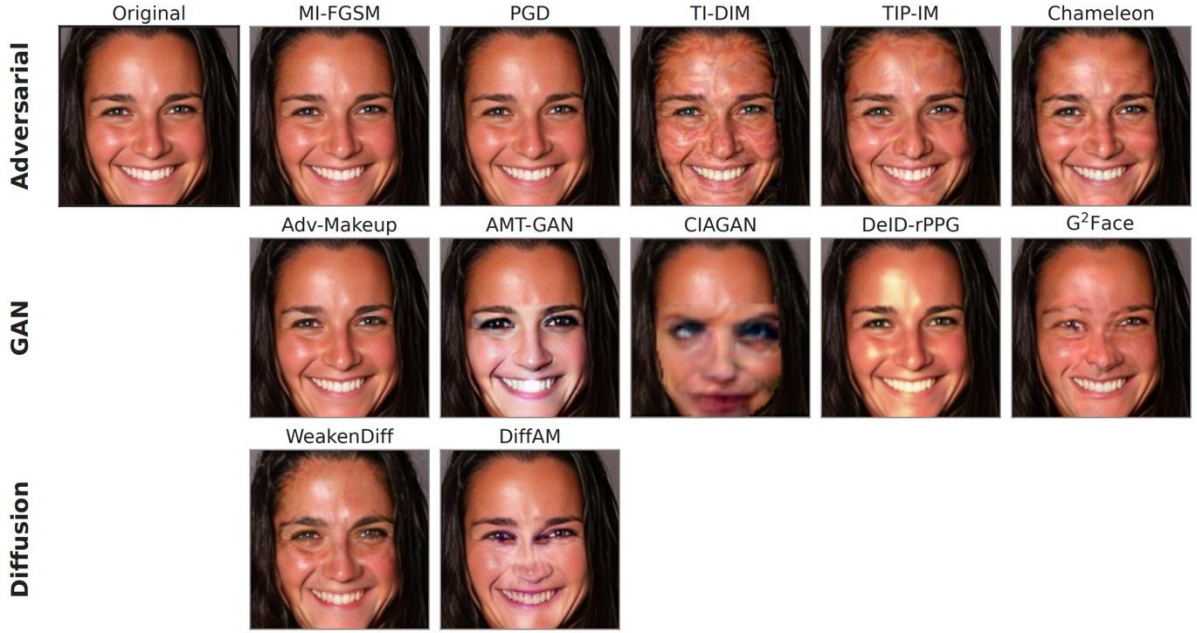
This is not privacy protection



This is privacy protection

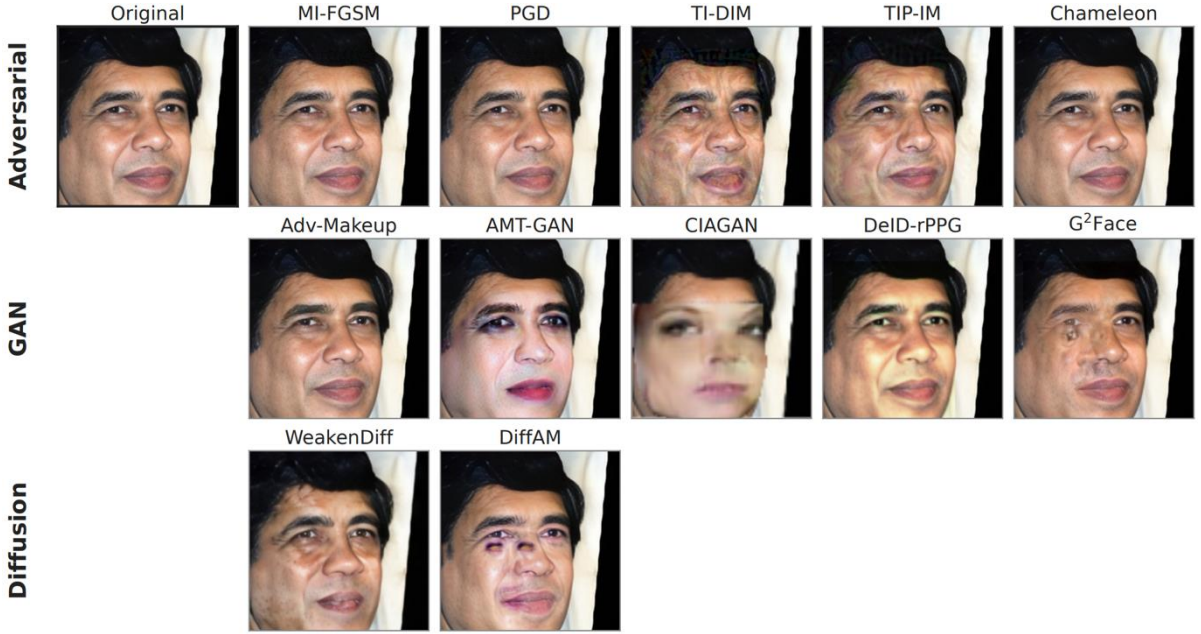
NEURAL INFORMATION PROCESSING SYSTEMS

Related work of face de-identification



NEURAL INFORMATION PROCESSING SYSTEMS

Related work of face de-identification



NEURAL INFORMATION PROCESSING SYSTEMS

How private is private?

Multi-objective:

- ❑ Identity suppression
- ❑ Utility preservation
- ❑ Quality maintenance

Multi-utility:

- ❑ Age
- ❑ Gender
- ❑ Ethnicity
- ❑ Landmark
- ❑ Macro-Expression
- ❑ Micro-Expression
- ❑ Gaze
- ❑ rPPG

Multi-dataset:

- ❑ LFW
- ❑ CelebA
- ❑ FFHQ
- ❑ AgeDB
- ❑ RFW
- ❑ AffectNet
- ❑ FairFace
- ❑ UTKFace
- ❑ GazeCapture
- ❑ ……(60+)

Multi-metric:

- ❑ Similarity
- ❑ Fooling Rate
- ❑ AUC
- ❑ Accuracy
- ❑ k-anonymity
- ❑ MSE
- ❑ Rank-N-T
- ❑ Success Rate
- ❑ Recall
- ❑ ……(20+)

It’s really hard to say!

NEURAL INFORMATION PROCESSING SYSTEMS

How we do it

- A new benchmark: UtilFace
- A new metric: HiFD

NEURAL INFORMATION PROCESSING SYSTEMS

UtilFace: a new benchmark

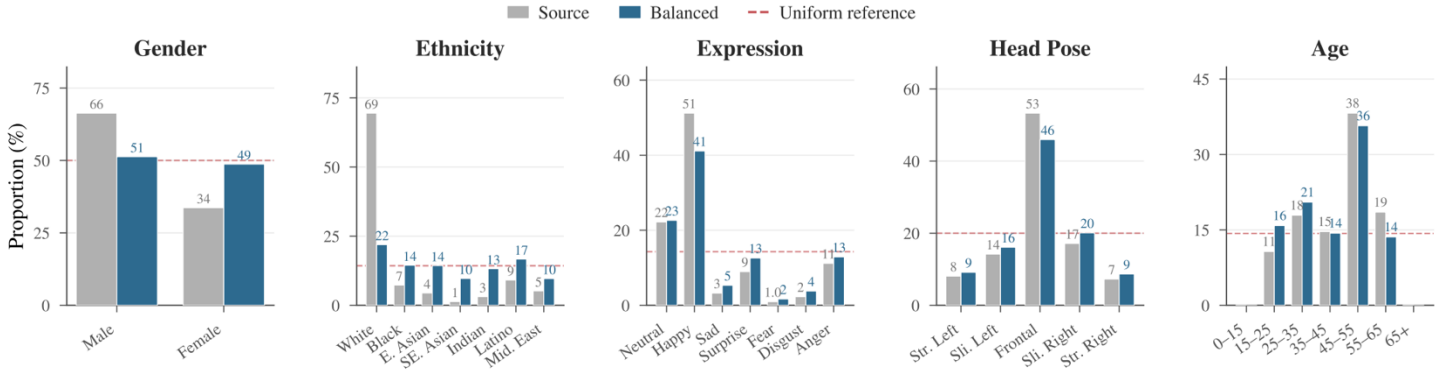
Current challenge: No existing dataset is built for de-identification: each carries different annotations, and none covers all of them.

Dataset	Modality	#Images/Videos	#IDs	L1: Macro-Cues					L2: Micro-Cues		L3
				Age	Gender	Ethnicity	Macro-Exp.	Landmark	Micro-Exp.	Gaze	rPPG
LFW [26]	Image	13,233	5,749	✗	✗	✗	✗	✗	✗	✗	✗
CelebA [51]	Image	202,599	10,177	✗	✓	✗	✓	✓	✗	✗	✗
FFHQ [34]	Image	70,000	–	✗	✗	✗	✗	✓	✗	✗	✗
AgeDB [62]	Image	16,488	568	✓	✓	✗	✗	✗	✗	✗	✗
RFW [78]	Image	40,000	12,000	✗	✗	✓	✗	✗	✗	✗	✗
AffectNet [61]	Image	450,000	–	✗	✗	✗	✓	✓	✗	✗	✗
FairFace [33]	Image	108,501	–	✓	✓	✓	✗	✗	✗	✗	✗
UTKFace [96]	Image	23,708	–	✓	✓	✓	✗	✓	✗	✗	✗
GazeCapture [39]	Image	2,445,504	1,450	✗	✗	✗	✗	✗	✗	✓	✗
PubFig [40]	Image	58,797	200	✗	✗	✗	✗	✗	✗	✗	✗
FERET [65]	Image	14,126	1,199	✗	✗	✗	✗	✗	✗	✗	✗
VGGFace2 [6]	Image	3,310,000	9,131	✗	✗	✓	✗	✓	✗	✗	✗
CASIA-WebFace [88]	Image	494,414	10,575	✗	✗	✗	✗	✗	✗	✗	✗
DFME [97]	Video	7,526 clips	671	✗	✗	✗	✗	✗	✓	✗	✗
PURE [72]	Video	60 videos	10	✗	✗	✗	✗	✗	✗	✗	✓

NEURAL INFORMATION PROCESSING SYSTEMS

UtilFace: a new benchmark

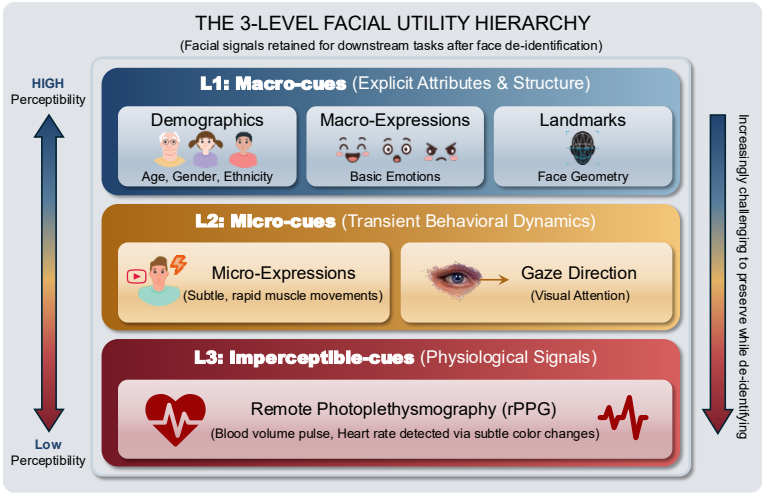
Stage	Operation	#Images	#Identities
1. Aggregation	Merge 4 source datasets	$\sim$ 69.55M	$\sim$ 3.12M
2. Cleaning + Enhancement	Identity-aware filtering; GFPGAN / CodeFormer	$\sim$ 20.58M	$\sim$ 701,810
3. Quality filter	NIQE < 5.0	$\sim$ 8.69M	$\sim$ 640,035
4. Balanced sampling	Hierarchical stratified sampling	99,928	2,069



HiFD: a new metric

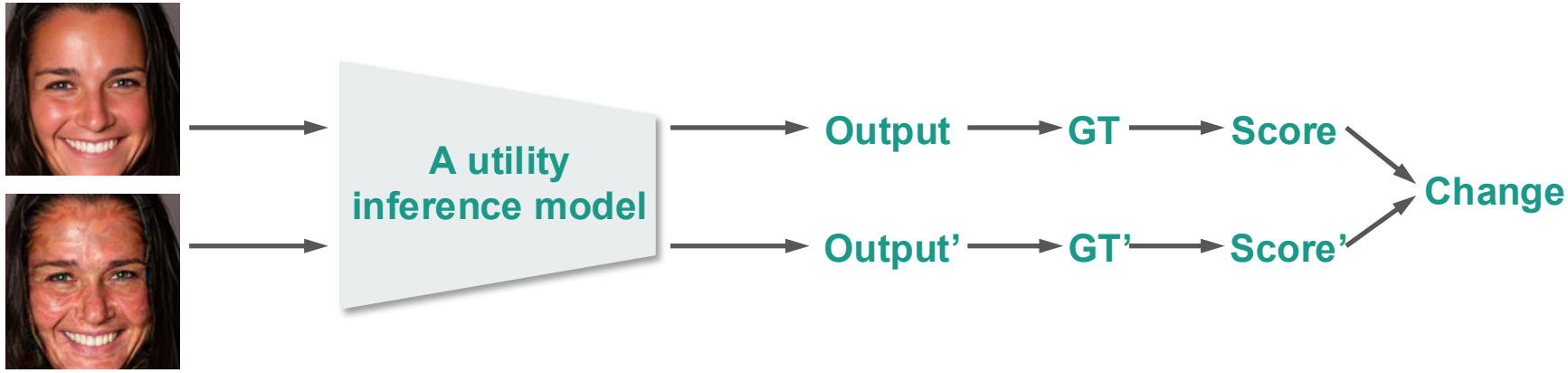
Our insight: facial signals exhibit a natural hierarchy of perceptual granularity.





HiFD: a new metric

A hierarchical, consistency-based metric.

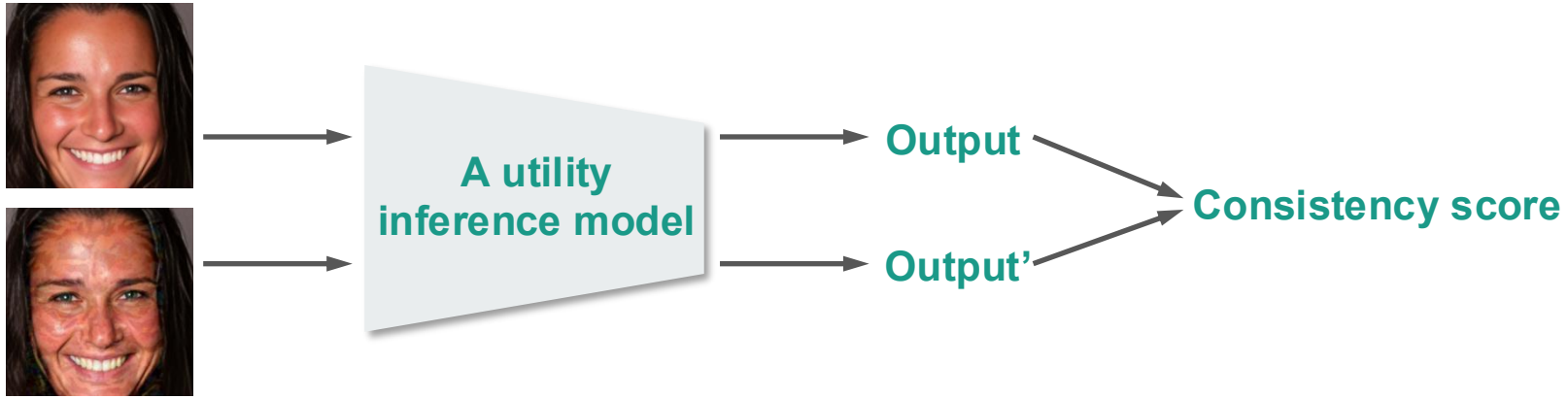


Current metrics

NEURAL INFORMATION PROCESSING SYSTEMS

HiFD: a new metric

A hierarchical, consistency-based metric.



Our metric

NEURAL INFORMATION PROCESSING SYSTEMS

HiFD: a new metric

A hierarchical, consistency-based metric.

$$\textbf{HiFD} = \left(w_P + w_Q + \sum_{\ell=1}^{L} w_\ell \right) \cdot \left(\frac{w_P}{\bar{P} + \epsilon} + \frac{w_Q}{Q + \epsilon} + \sum_{\ell=1}^{L} \frac{w_\ell}{U_\ell + \epsilon} \right)^{-1}$$

NEURAL INFORMATION PROCESSING SYSTEMS

What HiFD sees

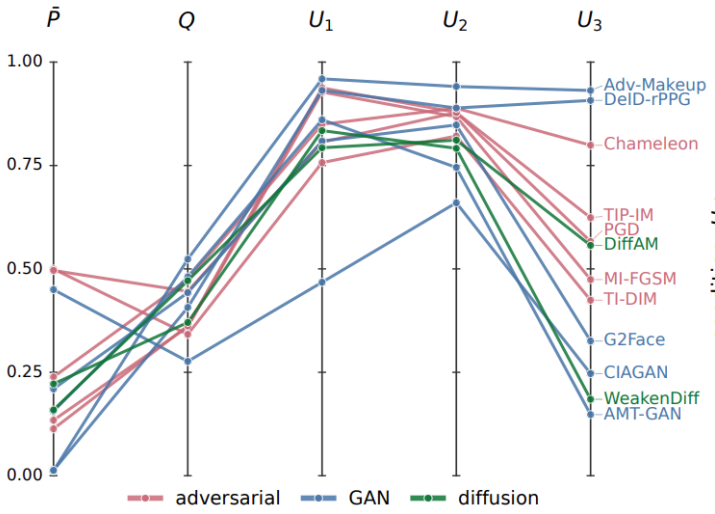
	Original	PGD	Chameleon	Adv-Makeup	CIAGAN	G²Face	WeakenDiff
		 $\bar{P}$: 0.11/Q: 0.36	 $\bar{P}$: 0.24/Q: 0.47	 $\bar{P}$: 0.01/Q: 0.52	 $\bar{P}$: 0.45/Q: 0.28	 $\bar{P}$: 0.21/Q: 0.44	 $\bar{P}$: 0.22/Q: 0.37
L1 - U_1 score **Macro cues** age • gender • ethnicity expression • landmark	1.00 **reference**	subtle noise and intact 0.94	visibly noisy but works 0.85	looks identical 0.96	geometry broken 0.47	mostly preserved 0.81	mostly preserved 0.83
L2 - U_2 score **Micro cues** gaze • micro-expression	1.00 **reference**	signals stable 0.88	signals stable 0.89	signals stable 0.94	signals drifts 0.66	signals stable 0.85	signals stable 0.79
L3 - U_3 score **Imperceptible cues** rPPG	1.00 **reference**	rPPG weakened 0.57 😲	rPPG stable 0.80 😲	rPPG intact 0.93	rPPG destroyed 0.25	rPPG destroyed 0.33	rPPG destroyed 0.18 😲
HiFD composite score	-----	**0.32** **Bottleneck**: privacy = 0.11	**0.51** **Bottleneck**: privacy = 0.24	**0.06** 😲 **Bottleneck**: privacy = 0.01	**0.37** **Bottleneck**: L3 = 0.25	**0.40** **Bottleneck**: privacy = 0.21	**0.33** **Bottleneck**: L3 = 0.18

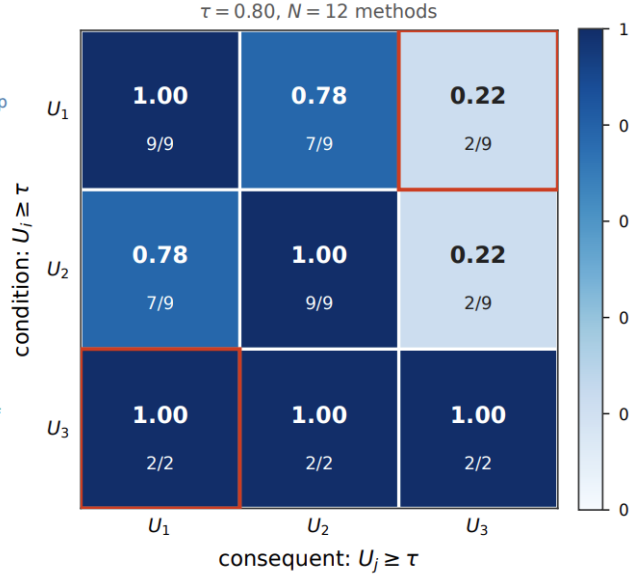
NEURAL INFORMATION PROCESSING SYSTEMS

What HiFD sees

Method	$\bar{P}\uparrow$	$Q\uparrow$	L1: Macro-cues					$U_1\uparrow$	L2: Micro-cues		$U_2\uparrow$	L3: rPPG		$U_3\uparrow$	HiFD↑
			Age	Gen.	Eth.	MExp	Lmk		Gaze	ME		BVP	HR		
(a) Adversarial perturbation															
MI-FGSM	0.134	0.360	0.975	0.962	0.916	0.951	0.834	0.928	0.936	0.799	0.868	0.471	0.478	0.474	0.343
PGD	0.113	0.362	0.978	0.968	0.935	0.964	0.843	0.938	0.940	0.817	0.878	0.577	0.555	0.566	0.321
TI-DIM	**0.500**	0.342	0.880	0.826	0.656	0.739	0.684	0.757	0.921	0.722	0.821	0.424	0.425	0.424	0.509
TIP-IM	0.496	0.446	0.927	0.882	0.701	0.794	0.728	0.806	0.944	0.812	0.878	0.643	0.605	0.624	**0.607**
Chameleon	0.239	0.474	0.915	0.897	0.780	0.839	0.816	0.850	0.961	0.816	0.889	0.821	0.777	0.799	0.508
(b) GAN-based generation															
Adv-Makeup	0.012	**0.524**	**0.987**	**0.980**	**0.942**	**0.971**	**0.917**	**0.960**	**0.966**	**0.916**	**0.941**	**0.964**	**0.898**	**0.931**	0.056
AMT-GAN	0.156	0.481	0.966	0.911	0.737	0.868	0.821	0.861	0.938	0.553	0.745	0.132	0.164	0.148	0.282
CIAGAN	0.450	0.276	0.894	0.549	0.291	0.529	0.075	0.468	0.748	0.572	0.660	0.225	0.269	0.247	0.369
DeID-rPPG	0.013	0.407	0.980	0.971	0.890	0.963	0.854	0.932	0.940	0.838	0.889	0.940	0.875	0.907	0.061
G^2Face	0.210	0.443	0.953	0.898	0.712	0.817	0.665	0.809	0.903	0.793	0.848	0.324	0.328	0.326	0.400
(c) Diffusion-based generation															
WeakenDiff	0.222	0.371	0.951	0.922	0.680	0.865	0.754	0.834	0.924	0.659	0.792	0.169	0.201	0.185	0.332
DiffAM	0.159	0.471	0.959	0.923	0.698	0.802	0.583	0.793	0.908	0.715	0.811	0.538	0.575	0.557	0.394

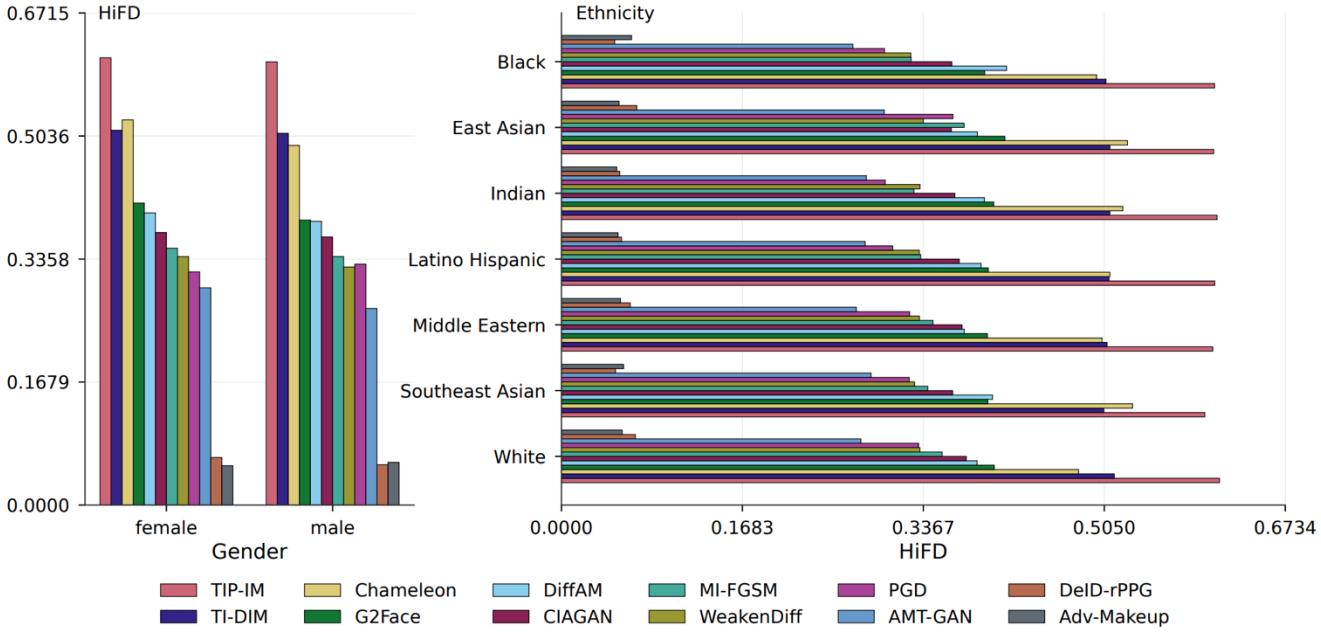
What HiFD sees





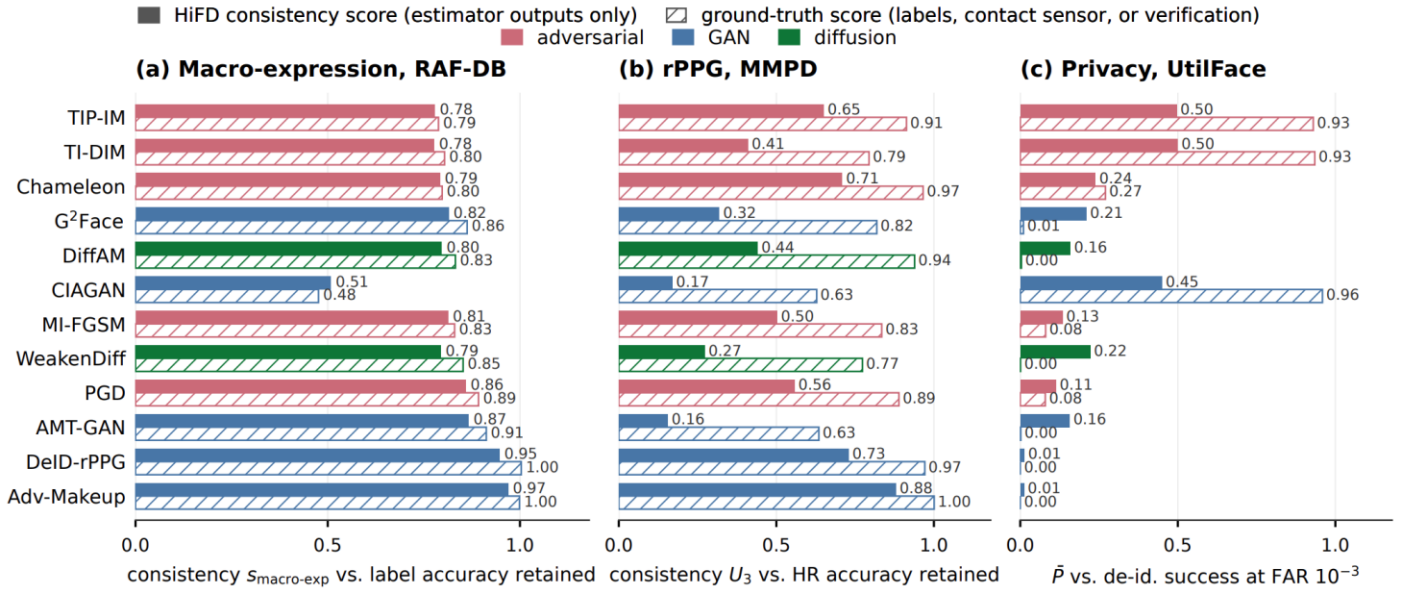
NEURAL INFORMATION PROCESSING SYSTEMS

What HiFD sees



NEURAL INFORMATION PROCESSING SYSTEMS

Does HiFD really work?



NEURAL INFORMATION PROCESSING SYSTEMS

How private is private?

- No current method is simultaneously private and utility-preserving across all three levels.
- The best method reaches only 0.61. Nine of twelve score below 0.3 on privacy against modern recognizers.

NEURAL INFORMATION PROCESSING SYSTEMS

Thanks for your attention!

Project Page: https://cv-ac.github.io/hifd/

Email to us: hui.wei@oulu.fi, guoying.zhao@oulu.fi

ellis INSTITUTE FINLAND

UNIVERSITY OF OULU