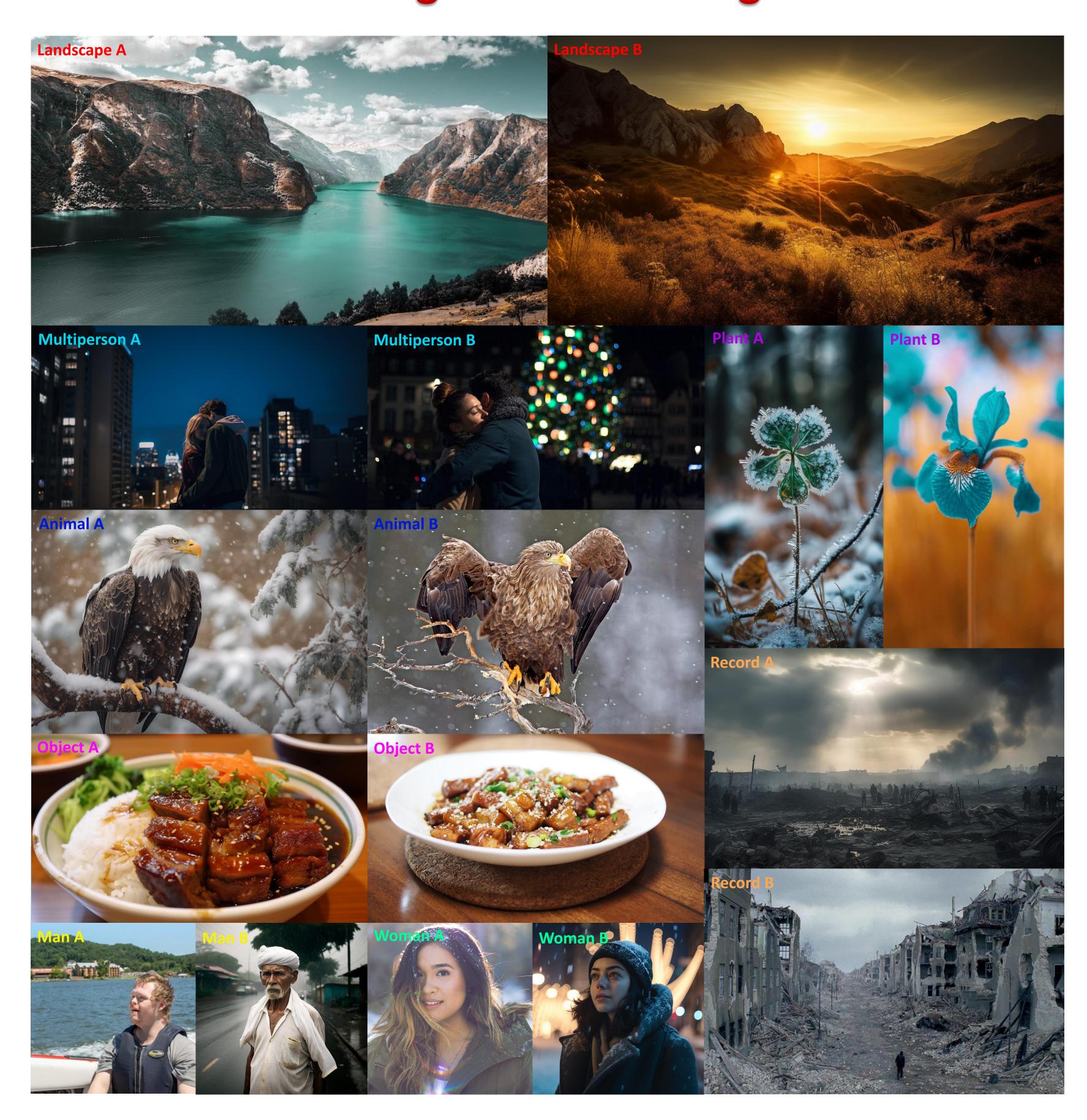# Seeing is not always believing: Benchmarking Human and Model Perception of AI-Generated Images

Zeyu Lu[1,2], Di Huang[2,3], Lei Bai[2], Jingjing Qu[2], Chengyue Wu[4], Xihui Liu[4], Wanli Ouyang[2]

[1]Shanghai Jiao Tong University, [2]Shanghai Artificial Intelligence Laboratory, [3]The University of Sydney, [4]The University of Hong Kong

Sentry-Questionnaire : Can you identify which ones are AI-generated images?
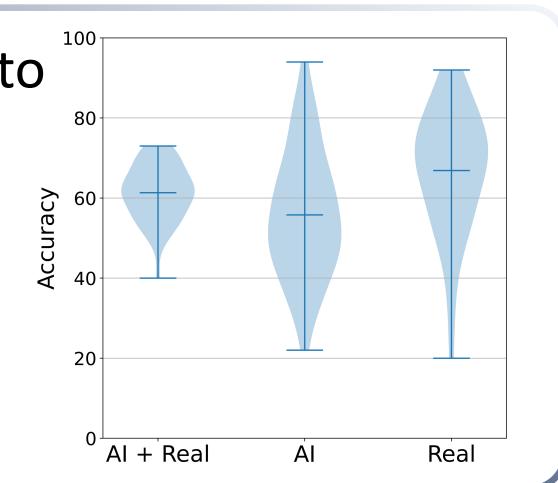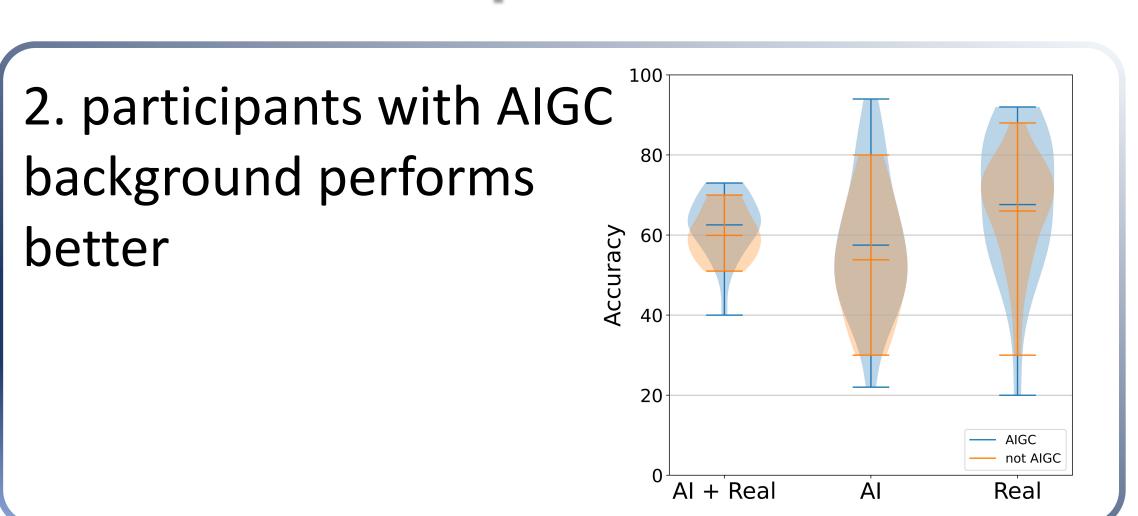

A new, large-scale dataset for fake image detection

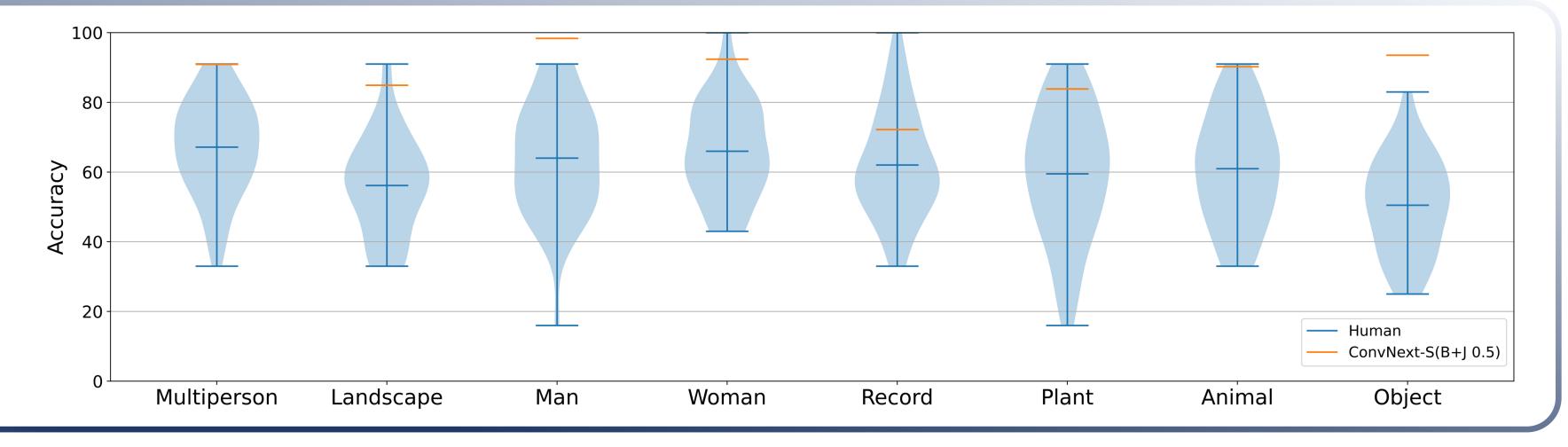| Dataset	CC3M-Train	StyleGAN3-Train	SD-V1.5Real-dpms-25	IF-V1.0-dpms++-25	StyleGAN3	ImageNet-Test	CelebA-HQ-Train	CC3M-Val	SD-V2.1-dpms-25	SD-V1.5-dpms-25	SD-V1.5Real-dpms-25	IF-V1.0-dpms++-10	IF-V1.0-dpms++-25	IF-V1.0-dpms++-50	IF-V1.0-ddim-50	IF-V1.0-ddpms-50	Cogview2	Midjourney	StyleGan3
Category	**Train**					**Validate**													
	R	R	F	F	F	R	R	R	F	F	F	F	F	F	F	F	F	F	F
Generator	-	-	Diff.	Diff.	GAN	-	-	-	Diff.	Diff.	Diff.	Diff.	Diff.	Diff.	Diff.	Diff.	AR	Unk.	GAN
Numbers	1M	87K	1M	1M	87K	100K	24K	15K	15K	15K	15K	15K	15K	15K	15K	15K	22K	5.5K	60K
This work	✗	✗	✓	✓	✓	✗	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

HPBench: A Benchmark for Human Perception

1. humans struggle significantly to distinguish real photos from AI-generated ones, with a misclassification rate of 38.7%

2. participants with AIGC background performs better

3. The average performance of human is far worse than the performance of the model

4. Statics of judgement criteria from participants correctly identifying fake images
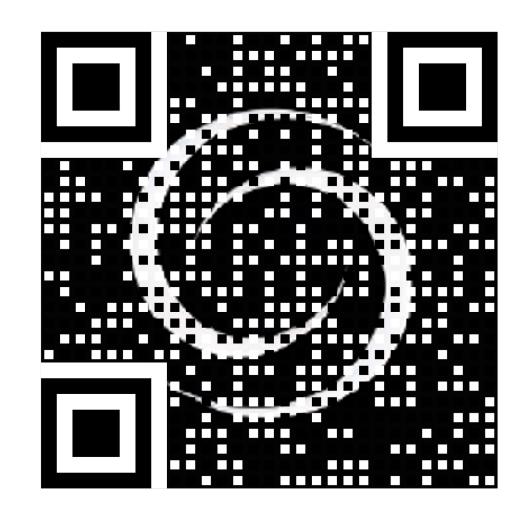
Category	Detail	Smooth	Blur	Color	Shadow & Light	Daub	Rationality	Intuition
Number	332	205	142	122	95	59	57	169
Percent	28%	17%	12%	10%	8%	5%	5%	14%

MPBench: A Benchmark for model Perception

Model	Training Dataset	ImageNet-Test	CelebA-HQ-Train	CC3M-Val	Average Acc.	SD-V2.1-dpm-25	SD-V1.5-dpm-25	SD-V1.5Real-dpm-25	IF-V1.0-dpm++-10	IF-V1.0-dpm++-25	IF-V1.0-dpm++-50	IF-V1.0-ddim-50	IF-V1.0-ddpm-50	Cogview2	Midjourney	StyleGAN3	Average Acc.	Total Average Acc.
		Real				**Fake**												**R+F**
						Diff.	Diff.	Diff.	Diff.	Diff.	Diff.	Diff.	Diff.	AR	Unk.	GAN		
ConvNext-S(B+J 0.1)	Dataset Setting A: SD-V1.5Real-dpms-25 (1M) CC3M-Train (1M)	95.3	99.9	99.9	98.3	48.6	99.9	85.1	92.9	57.6	55.9	41.1	72.6	55.7	41.6	37.4	62.5	70.2
ConvNext-S(B+J 0.5)		95.9	99.9	99.9	98.5	53.5	100	83.3	91.1	50.2	49.6	35.3	66.9	54.9	44.7	35.7	60.4	68.6
ResNet50(B+J 0.1)		93.0	95.3	95.6	94.6	71.7	71.8	98.5	52.1	29.0	23.6	47.9	11.8	40.3	9.1	44.3	55.1
ResNet50(B+J 0.5)		93.2	95.5	95.9	94.8	70.6	70.4	98.5	52.1	23.6	26.1	21.6	46.2	10.6	36.6	7.2	42.1	53.4
CLIP-ViT-L(LC)		49.6	87.0	75.4	70.6	73.3	86.6	97.6	93.9	77.8	71.4	84.1	90.9	86.6	88.7	86.1	85.1	82.0
ConvNext-S(B+J 0.1)	Dataset Setting B: IF-V1.0-dpms++-25 (1M) CC3M-Train (1M)	87.6	99.9	99.9	95.8	2.2	34.5	2.5	99.7	99.9	99.9	11.1	66.4	19.2	8.9	10.1	41.3	52.9
ConvNext-S(B+J 0.5)		87.8	99.9	99.9	95.8	3.9	34.1	3.9	99.6	99.9	99.8	18.5	79.2	25.8	8.1	8.0	44.1	55.2
ResNet50(B+J 0.1)		89.4	95.8	95.0	93.4	37.5	56.5	20.0	84.0	95.6	91.7	39.7	69.4	45.3	15.7	6.8	54.3	60.3
ResNet50(B+J 0.5)		90.8	95.6	94.5	93.6	41.6	58.8	21.7	82.3	95.2	91.3	47.0	79.7	56.1	18.3	6.8	54.4	62.8
CLIP-ViT-L(LC)		81.5	83.3	93.0	85.9	38.2	18.5	13.1	80.4	79.7	70.9	61.1	77.7	76.6	33.7	32.8	52.9	60.0
ConvNext-S(B+J 0.1)	Dataset Setting C: StyleGAN3 (87K) StyleGAN3-Train (87K)	58.7	81.0	62.7	67.4	44.9	42.7	39.4	40.7	42.3	42.2	35.7	39.9	53.2	41.6	69.9	44.7	49.6
ConvNext-S(B+J 0.5)		67.7	92.6	71.2	77.1	33.1	33.7	32.1	33.6	36.5	34.2	35.0	35.4	48.5	42.6	35.4	44.3	44.3
ResNet50(B+J 0.1)		31.5	14.3	39.1	28.3	67.7	62.7	68.1	65.3	71.0	67.0	58.8	60.9	60.9	81.9	90.9	68.6	60.0
ResNet50(B+J 0.5)		70.6	29.5	63.4	54.5	37.3	34.9	40.0	43.3	43.9	41.9	34.8	38.4	47.8	58.1	81.3	45.6	47.5
CLIP-ViT-L(LC)		37.6	85.9	70.2	64.5	85.7	92.9	94.8	95.3	89.5	84.0	87.5	93.0	84.6	81.4	61.9	86.4	81.7
ConvNext-S(B+J 0.1)	Dataset Setting D: SD-V1.5Real-dpms-25 (460K) IF-V1.0-dpms++-25 (460K) StyleGAN3 (87K) CC3M-Train (1M) StyleGAN3-Train (87K)	95.4	99.9	99.9	98.4	39.9	95.1	99.9	99.7	99.8	99.7	43.1	90.4	48.3	32.4	99.8	77.1	81.6
ConvNext-S(B+J 0.5)		97.3	99.9	99.9	99.1	52.8	92.9	99.9	99.5	99.9	99.3	46.4	91.7	48.5	20.9	99.8	83.0	83.0
ResNet50(B+J 0.1)		77.3	93.6	91.7	87.5	83.9	90.4	96.8	91.2	90.6	86.7	52.3	82.2	56.4	49.4	80.8	78.2	80.2
ResNet50(B+J 0.5)		84.5	94.7	92.1	90.4	84.4	89.4	96.1	88.9	89.4	85.6	52.4	83.8	55.4	45.8	69.5	76.4	79.4
CLIP-ViT-L(LC)		50.4	96.0	79.8	75.4	66.9	87.8	93.6	96.1	85.9	78.9	84.9	92.6	87.4	79.8	62.8	83.3	81.6

Project Page

Questionnaire