Digital Image Transformations Degrade Gaze Prediction Accuracy

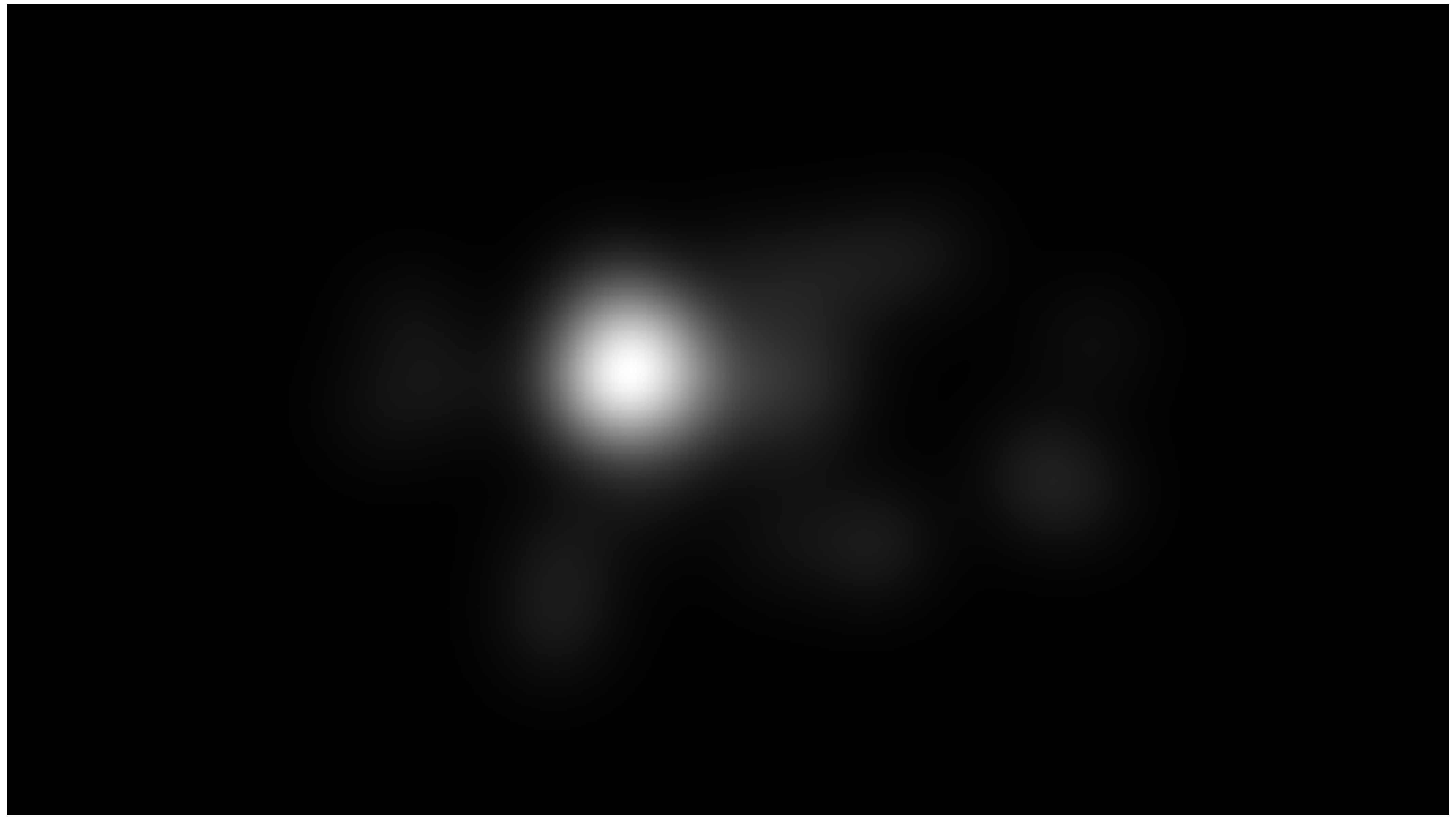
James Youngblood, Oct. 21, 2025

Outline

- Presentation
  - Introduction
  - Related Work
  - Background
  - Method
  - Results
- Discussion
- Final Examination

Introduction

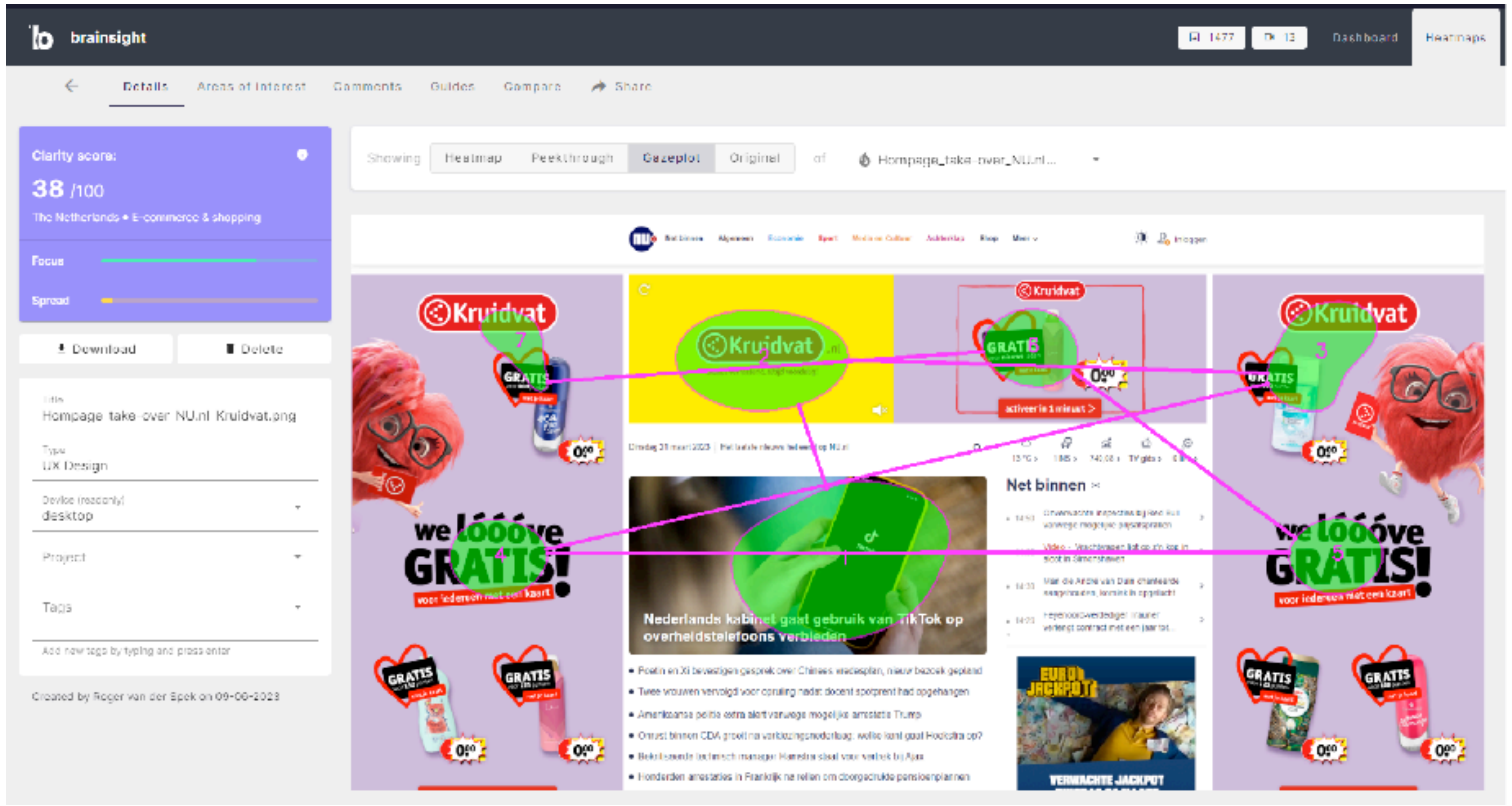
Image



Gaze Distribution

Introduction

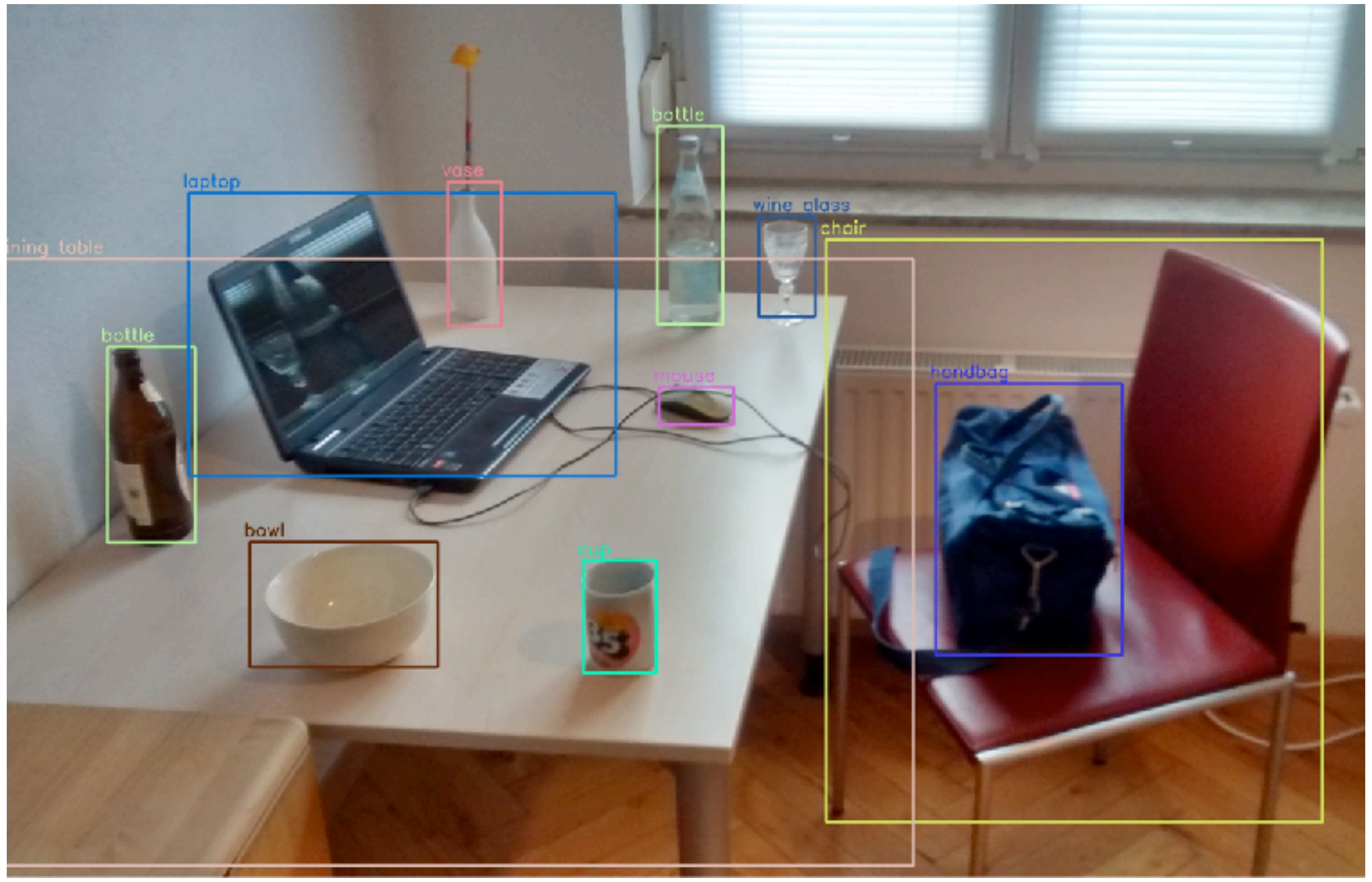
- User Interface
- Data Visualization
- Visual Media
- Computer vision and robotics



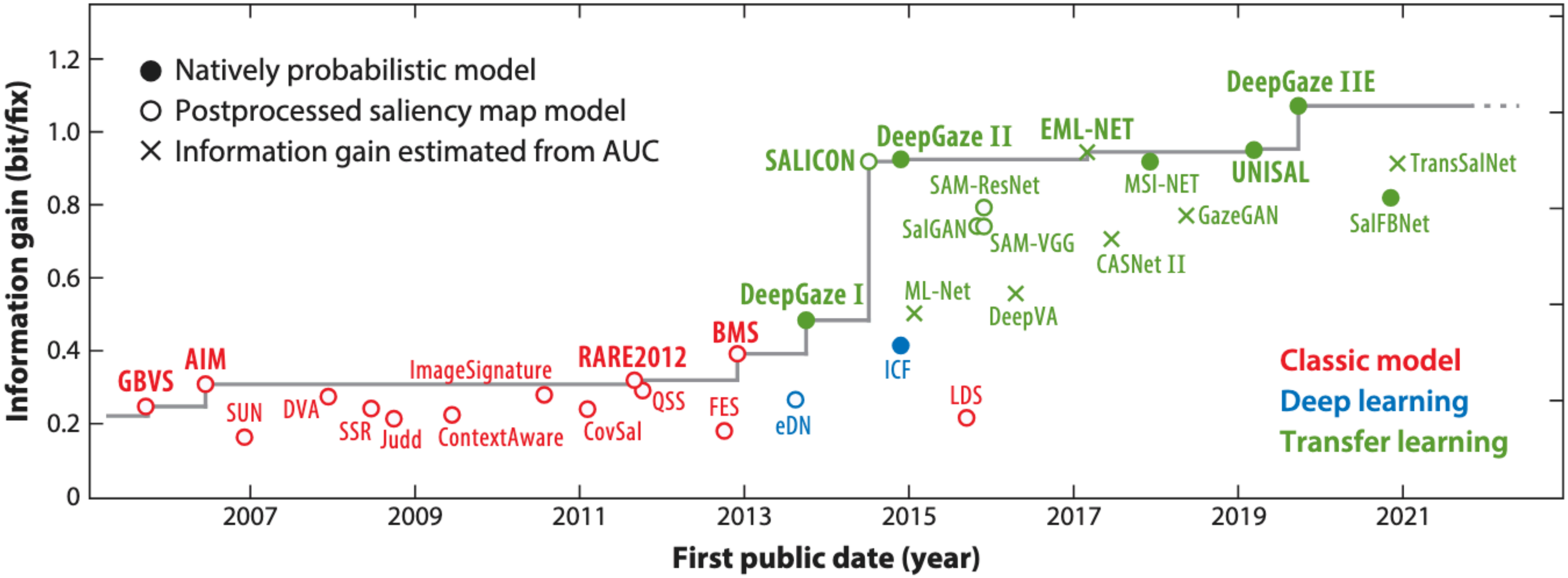
A plot of gaze behavior for a user interface

Introduction

Transfer learning (object detection to gaze prediction)

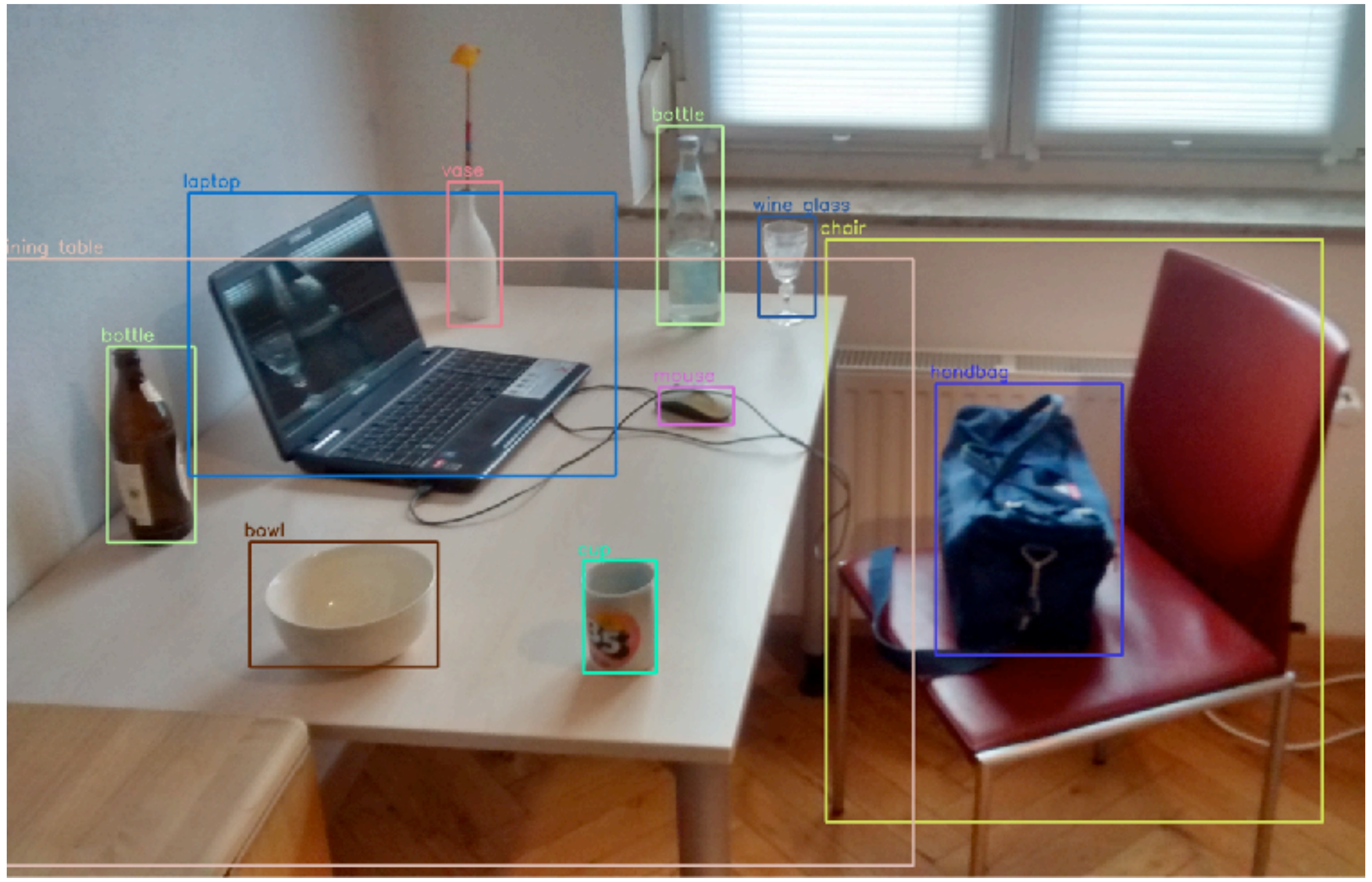


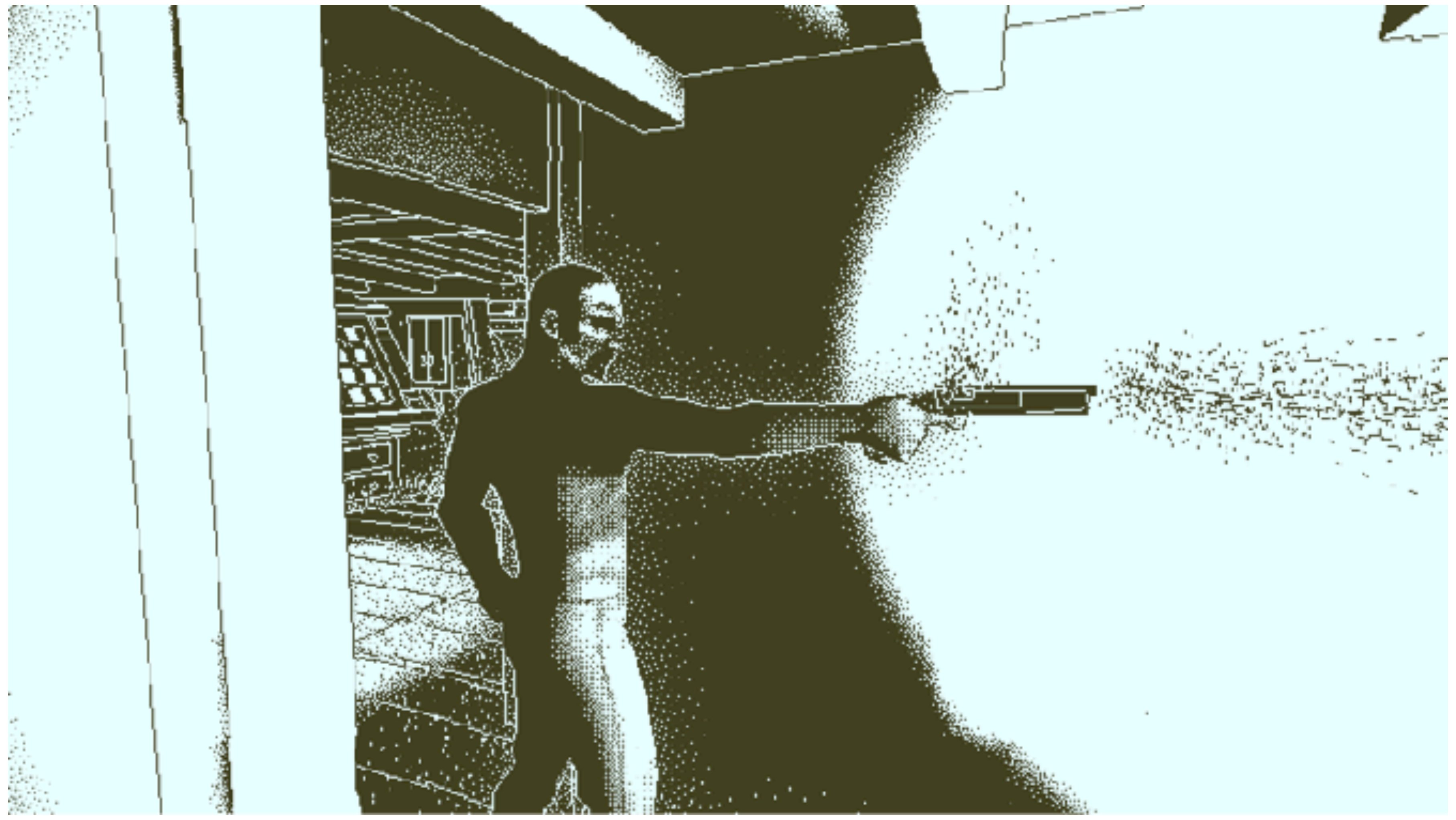
Introduction



Introduction

Will the model generalize?



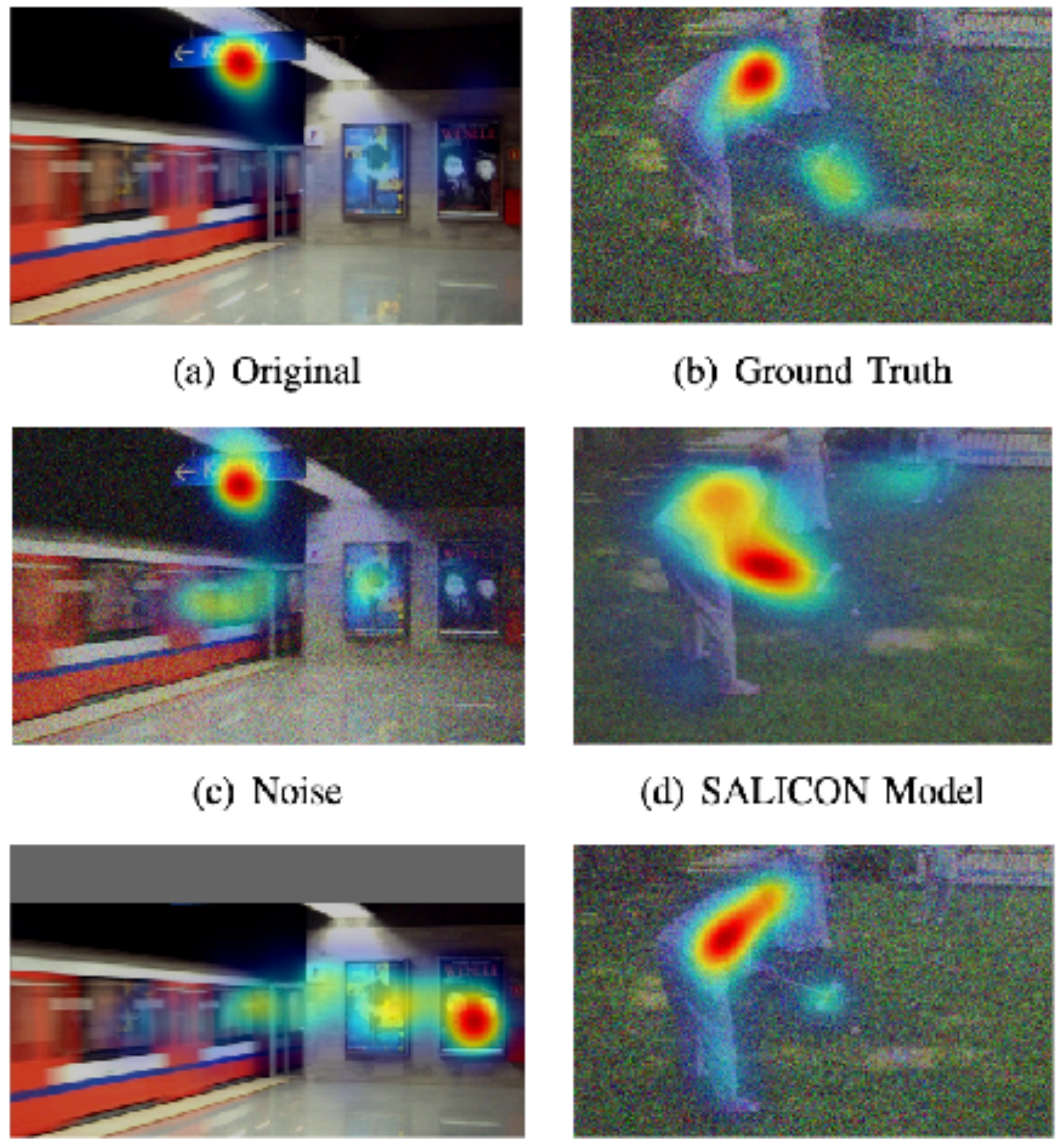


Related Work

How is Gaze Influenced by Image Transformations? Dataset and Model

Zhaohui Che, Ali Borji, *Member, IEEE,* Guangtao Zhai, *Senior Member, IEEE,* Xiongkuo Min, *Member, IEEE,* Guodong Guo, *Senior Member, IEEE* and Patrick Le Callet, *Fellow, IEEE*

***Abstract*—Data size is the bottleneck for developing deep saliency models, because collecting eye-movement data is very time-consuming and expensive. Most of current studies on human attention and saliency modeling have used high-quality stereotype stimuli. In real world, however, captured images undergo various types of transformations. Can we use these transformations to augment existing saliency datasets? Here, we first create a novel saliency dataset including fixations of 10 observers over 1900 images degraded by 19 types of transformations. Second, by analyzing eye movements, we find that observers look at different locations over transformed versus original images. Third, we utilize the new data over transformed images, called data augmentation transformation (DAT), to train deep saliency models. We find that label-preserving DATs with negligible impact on human gaze boost saliency prediction, whereas some other DATs that severely impact human gaze degrade the performance. These label-preserving valid augmentation transformations provide a solution to enlarge existing saliency datasets. Finally, we introduce a novel saliency model based on generative adversarial networks (dubbed GazeGAN). A modified U-Net is utilized as the generator of the GazeGAN, which combines classic "skip connection" with a novel "center-surround connection" (CSC) module. Our proposed CSC module mitigates trivial artifacts while emphasizing semantic salient regions, and increases model nonlinearity, thus demon-**



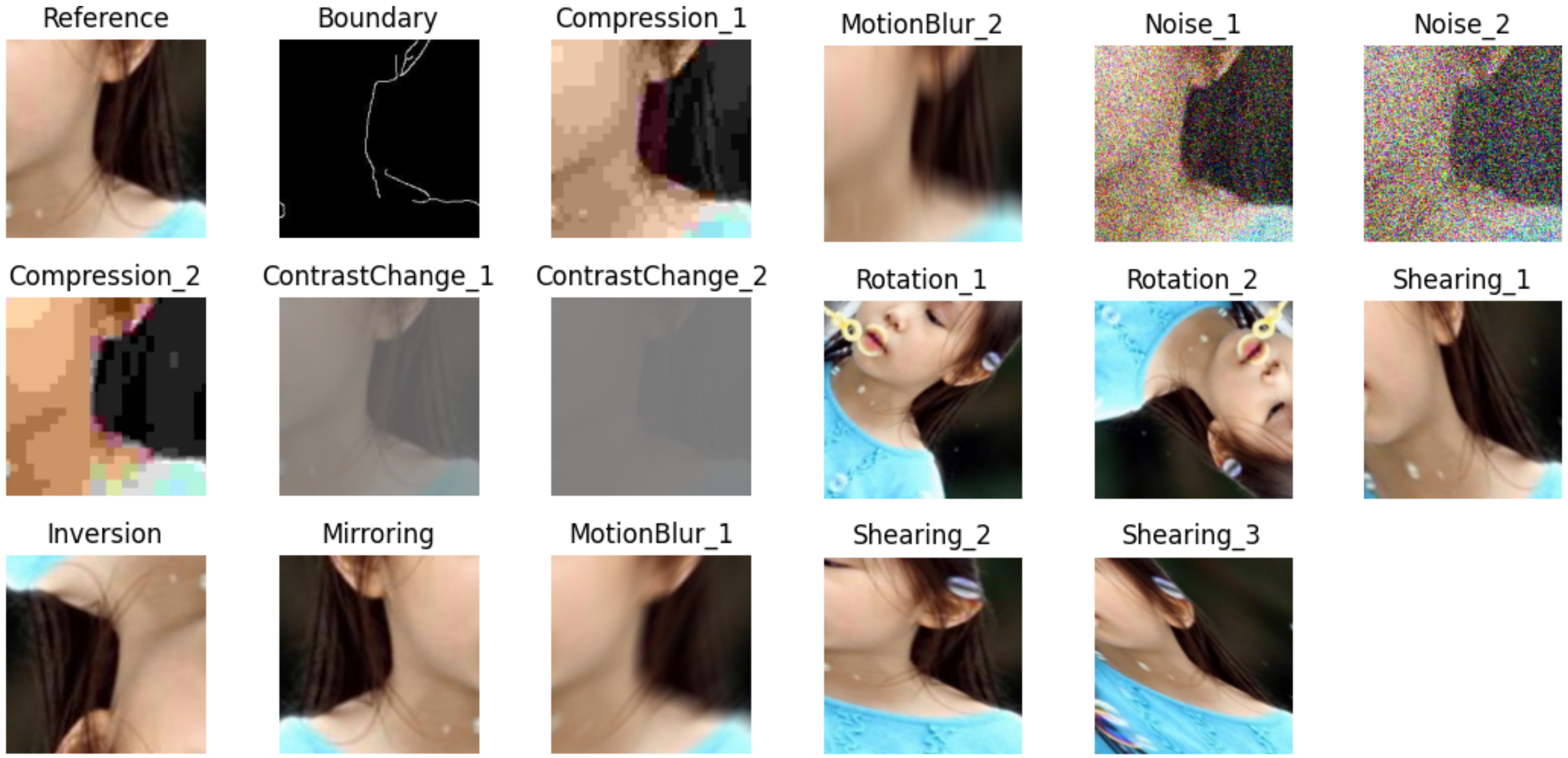
(a) Original (b) Ground Truth

(c) Noise (d) SALICON Model

Related Work

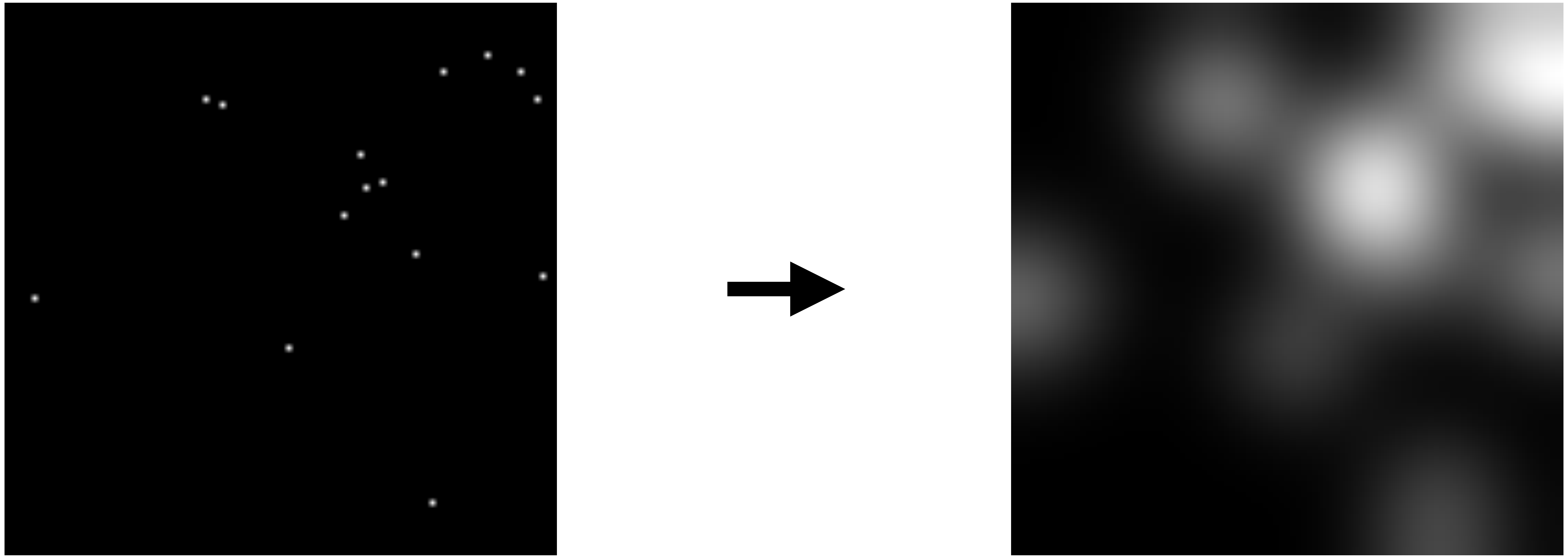
Cropping_1

Cropping_2

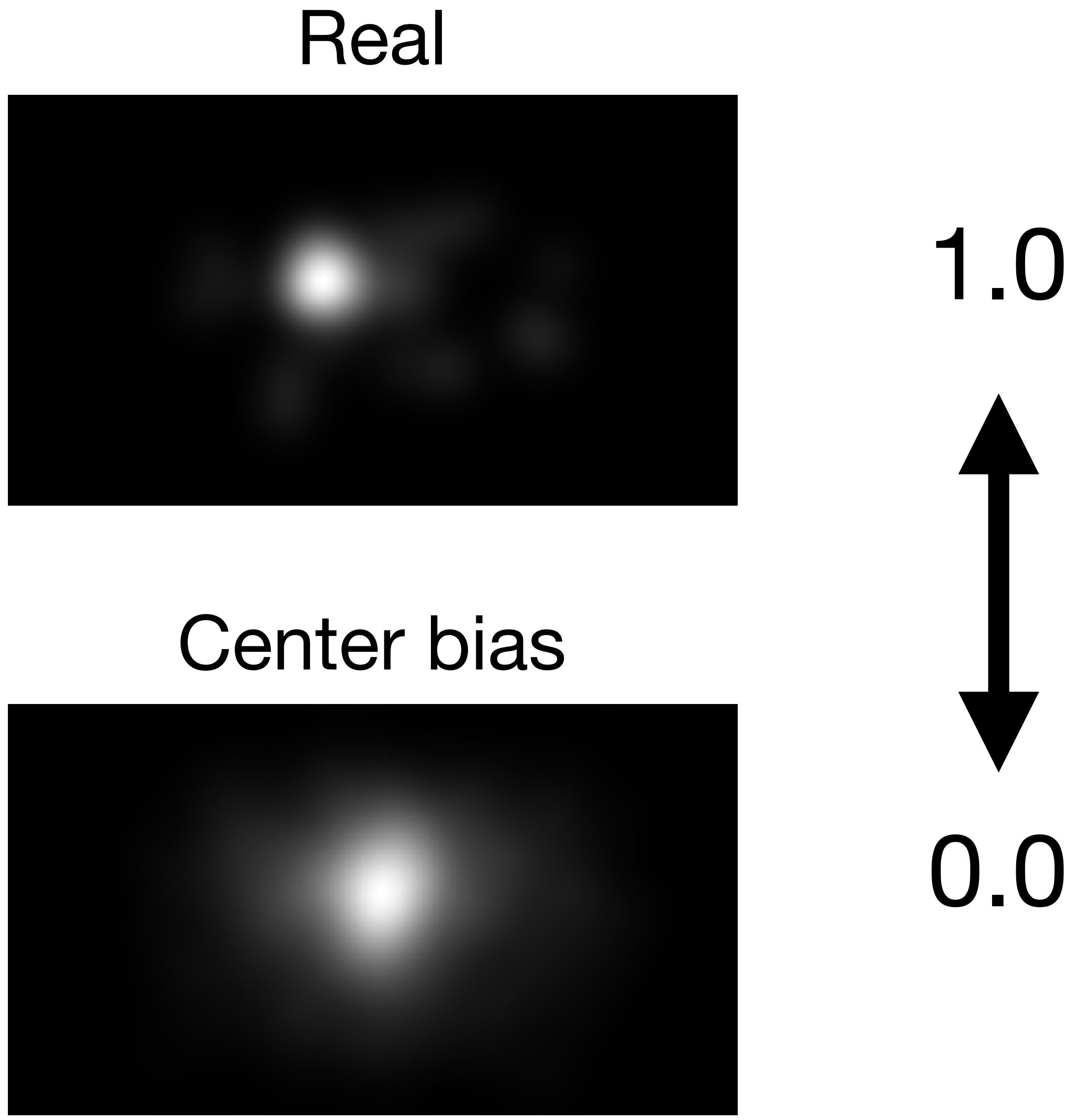


Background

Saliency maps



Background



Background

MIT/Tuebingen Saliency Benchmark

Leaderboard CAT2000

Click on a metric name to sort that metric. Models evaluated as probabilistic models are shown with green background. The performance under the metric which a model has been trained on is shown in italic. The code for evaluation models can be found here.

Name	Published	Code	IG [?]	AUC [?]	sAUC [?]	NSS [?]	CC [?]	KLDiv [?]	SIM [?]	Date tested
Gold Standard	*Gaussian kernel density estimate using all fixations of an image with uniform mixture component. Crossvalidated over subjects. Joint performance.*		*0.8026*	0.9159	0.7865	2.7429	0.9685	0.0893	0.8657	First tested 2023-07-06 Last tested 2023-07-06
Gold Standard (leave-one-subject-out)	*Gaussian kernel density estimate using all fixations of an image with uniform mixture component. Crossvalidated over subjects. Leave-one-subject-out performance*		*0.4730*	0.8840	0.6930	2.4878				First tested 2023-07-05 Last tested 2023-07-05
DeepGaze IIE	A. Linardos, M. Kümmerer, O. Press, M. Bethge: DeepGaze IIE: Calibrated prediction in and out-of-domain for state-of-the-art saliency modeling [ICCV 2021]	python (pytorch)	*0.1893*	0.8692	0.6677	2.1122	0.8189	0.3448	0.7060	First tested 2023-07-05 Last tested 2023-07-05 densities from authors
Forebrain | Attentive AI (v1)			0.1134	0.8598	0.6432	1.9664	0.7676	0.4114	0.6714	First tested 2024-07-27 Last tested 2024-07-27 predictions from authors
DeepGaze II	Matthias Kümmerer, Thomas S. A. Wallis, Leon A. Gatys, Matthias Bethge. Understanding Low- and High-Level Contributions to Fixation Prediction [ICCV 2017]	python (tensorflow)	*0.0839*	0.8640	0.6498	1.9619	0.7950	0.3815	0.6865	First tested 2021-04-09 Last tested 2023-07-04
UNISAL	R. Droste, J. Jiao, J.A. Noble: Unified Image and Video Saliency Modeling. ECCV 2020 (arXiv)	python (pytorch)	0.0321	0.8604	0.6684	1.9359	0.7399	0.4703	0.6633	First tested 2024-02-28 Last tested 2024-02-28

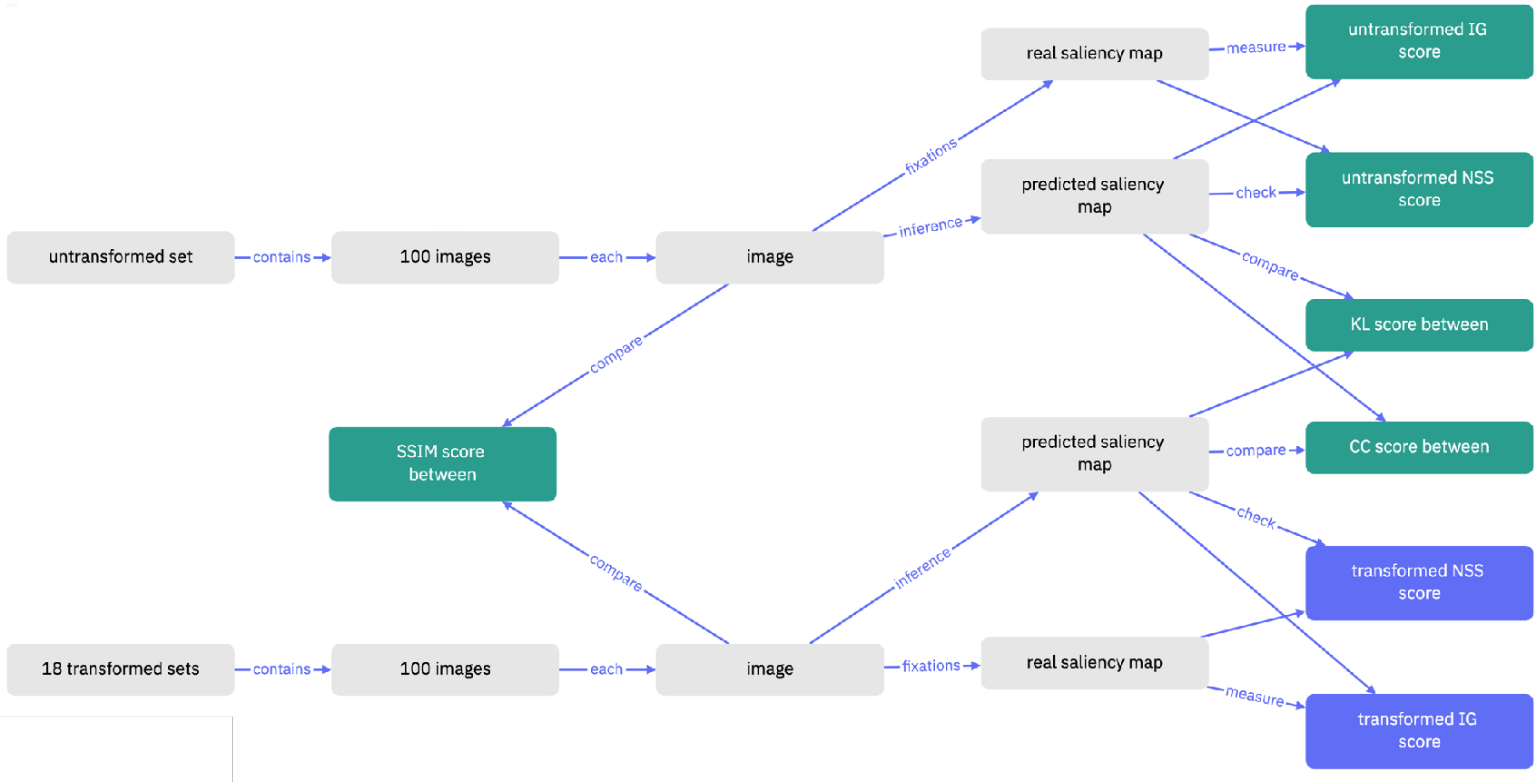
Background

Metric	Quick take-aways
Area under ROC Curve (AUC)	Historically the most commonly-used metric for saliency evaluation. Invariant to monotonic transformations. Driven by high-valued predictions and largely ambivalent of low-valued false positives. Currently saturating on standard saliency benchmarks [14], [15]. Good for detection applications.
Shuffled AUC (sAUC)	A version of AUC that compensates for dataset bias by scoring a center prior at chance. Most appropriate in evaluation settings where the saliency model is not expected to account for center bias. Otherwise, has similar properties to AUC.
Similarity (SIM)	An easy and fast similarity computation between histograms. Assumes the inputs are valid distributions. More sensitive to false negatives than false positives.
Pearson's Correlation Coefficient (CC)	A linear correlation between the prediction and ground truth distributions. Treats false positives and false negatives symmetrically.
Normalized Scanpath Saliency (NSS)	A discrete approximation of CC that is additionally parameter-free (operates on raw fixation locations). Recommended for saliency evaluation.
Earth Mover's Distance (EMD)	The only metric considered that scales with spatial distance. Can provide a finer-grained comparison between saliency maps. Most computationally intensive, non-local, hard to optimize.
Kullback-Leibler divergence (KL)	Has a natural interpretation where goal is to approximate a target distribution. Assumes input is a valid probability distribution with sufficient regularization. Mis-detections are highly penalized.
Information Gain (IG)	A new metric introduced by [45], [46]. Assumes input is a valid probability distribution with sufficient regularization. Measures the ability of a model to make predictions above a baseline model of center bias. Otherwise, has similar properties to KL.

Method

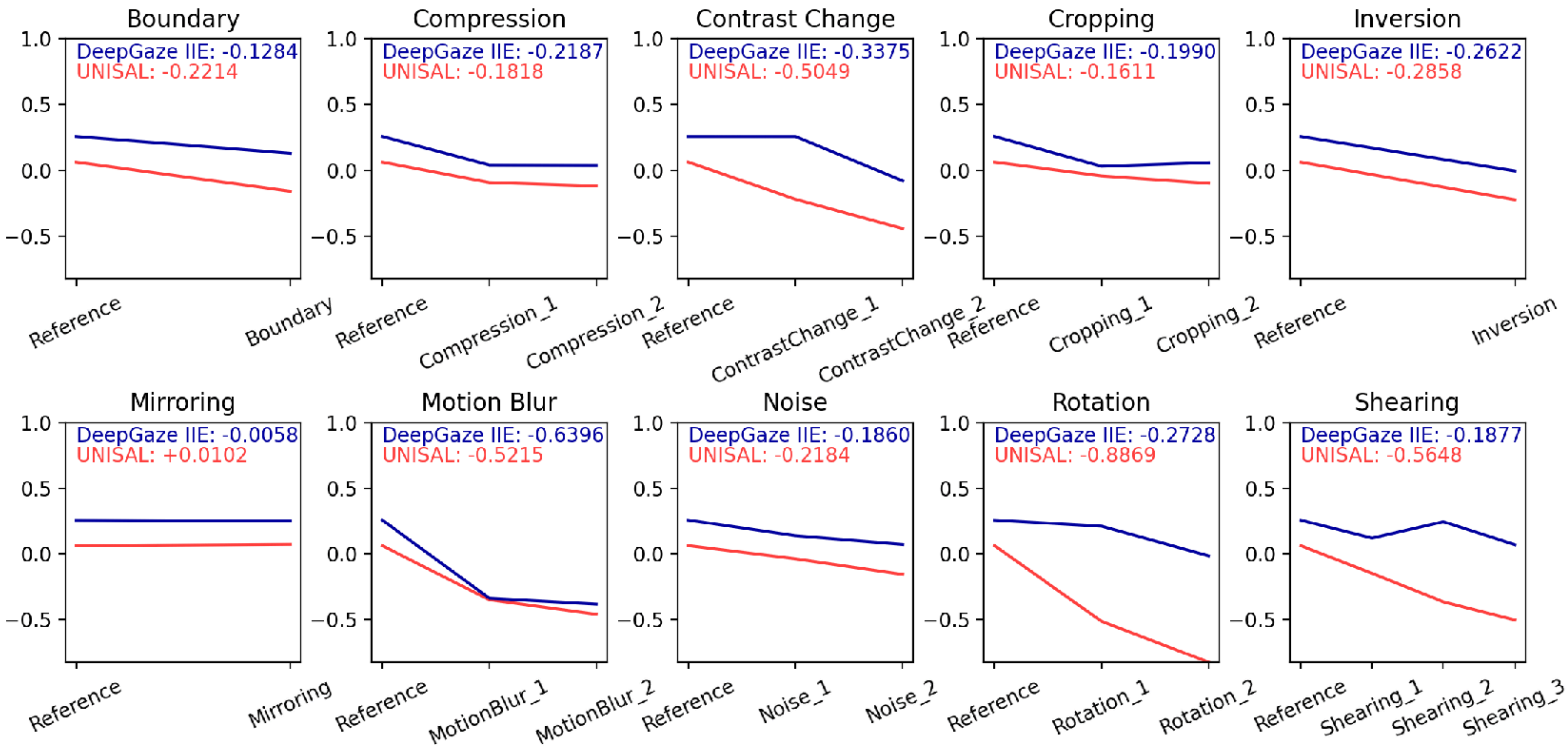
image sets
split
19 sets
group
10 similar groups
collect
performance degradation for group
contain
100 images
each
image
inference
predicted saliency map
sum
fixations
center bias
real saliency map
NSS and IG
plot
NSS and IG
NSS and IG
reference scores
range
performance percentage

Method

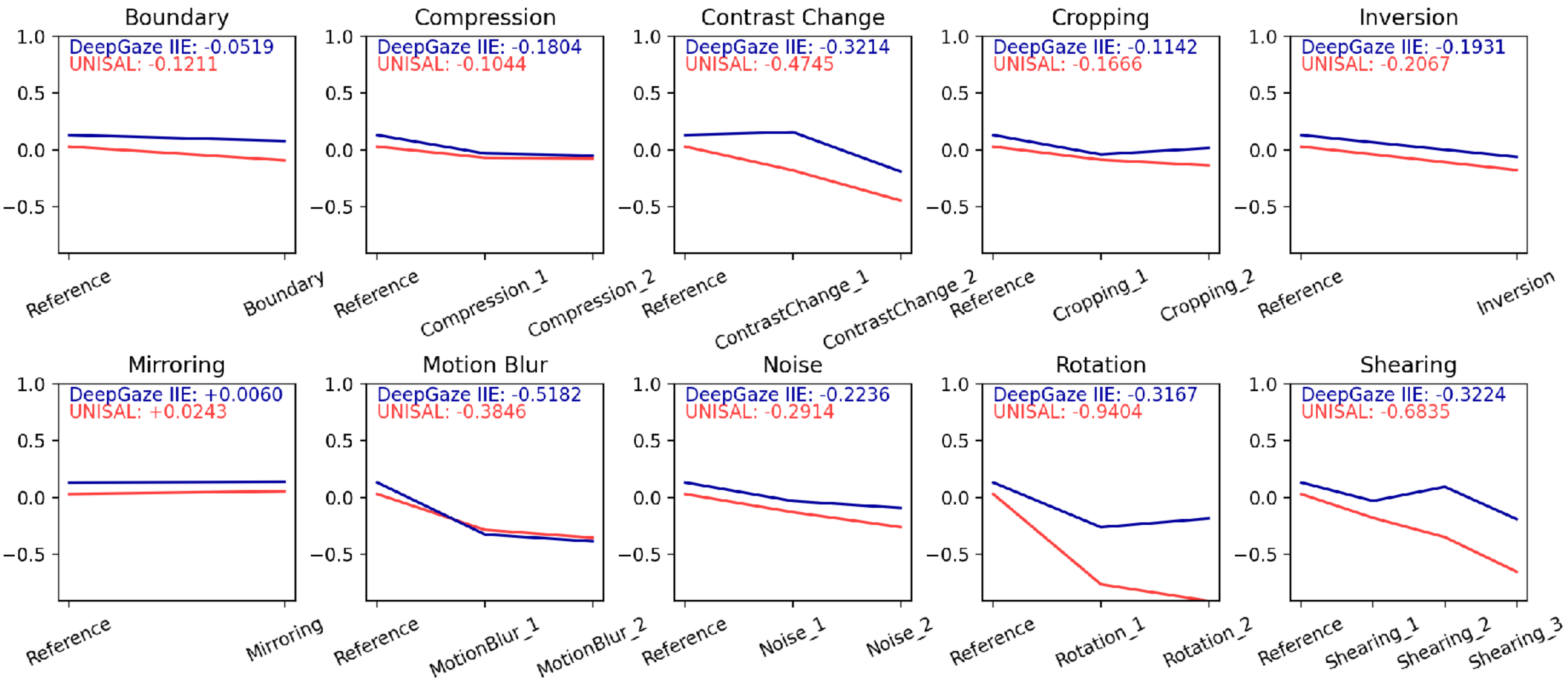


untransformed set
contains
100 images
each
image
fixations
real saliency map
measure
untransformed IG score
inference
predicted saliency map
check
untransformed NSS score
compare
KL score between
compare
SSIM score between
compare
predicted saliency map
compare
CC score between
check
transformed NSS score
inference
18 transformed sets
contains
100 images
each
image
fixations
real saliency map
measure
transformed IG score

Results

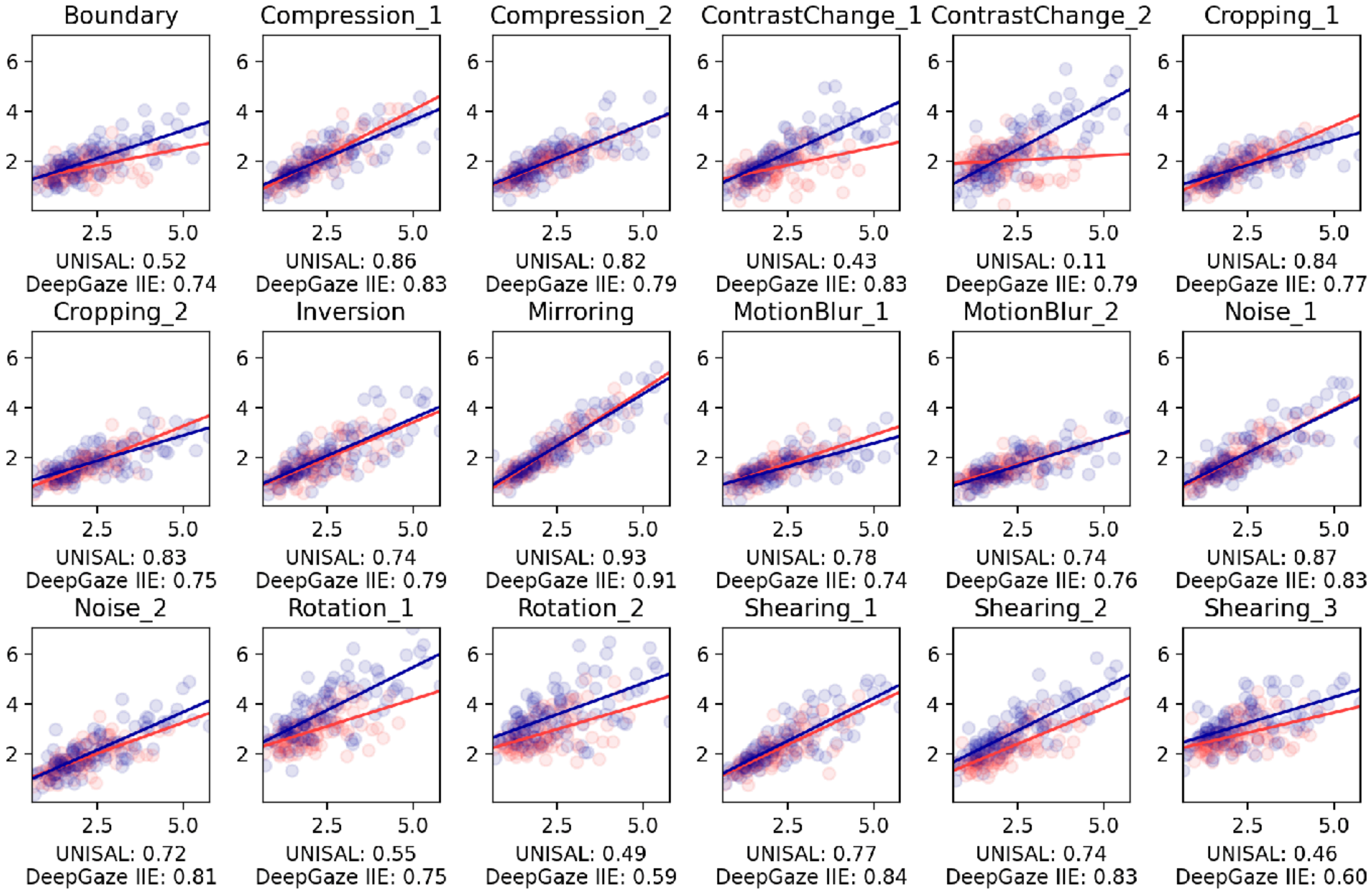

Boundary
DeepGaze IIE: -0.1284
UNISAL: -0.2214
Compression
DeepGaze IIE: -0.2187
UNISAL: -0.1818
Contrast Change
DeepGaze IIE: -0.3375
UNISAL: -0.5049
Cropping
DeepGaze IIE: -0.1990
UNISAL: -0.1611
Inversion
DeepGaze IIE: -0.2622
UNISAL: -0.2858
Mirroring
DeepGaze IIE: -0.0058
UNISAL: +0.0102
Motion Blur
DeepGaze IIE: -0.6396
UNISAL: -0.5215
Noise
DeepGaze IIE: -0.1860
UNISAL: -0.2184
Rotation
DeepGaze IIE: -0.2728
UNISAL: -0.8869
Shearing
DeepGaze IIE: -0.1877
UNISAL: -0.5648
Reference
Boundary
Compression_1
Compression_2
ContrastChange_1
ContrastChange_2
Cropping_1
Cropping_2
Inversion
Mirroring
MotionBlur_1
MotionBlur_2
Noise_1
Noise_2
Rotation_1
Rotation_2
Shearing_1
Shearing_2
Shearing_3
1.0
0.5
0.0
−0.5

Results



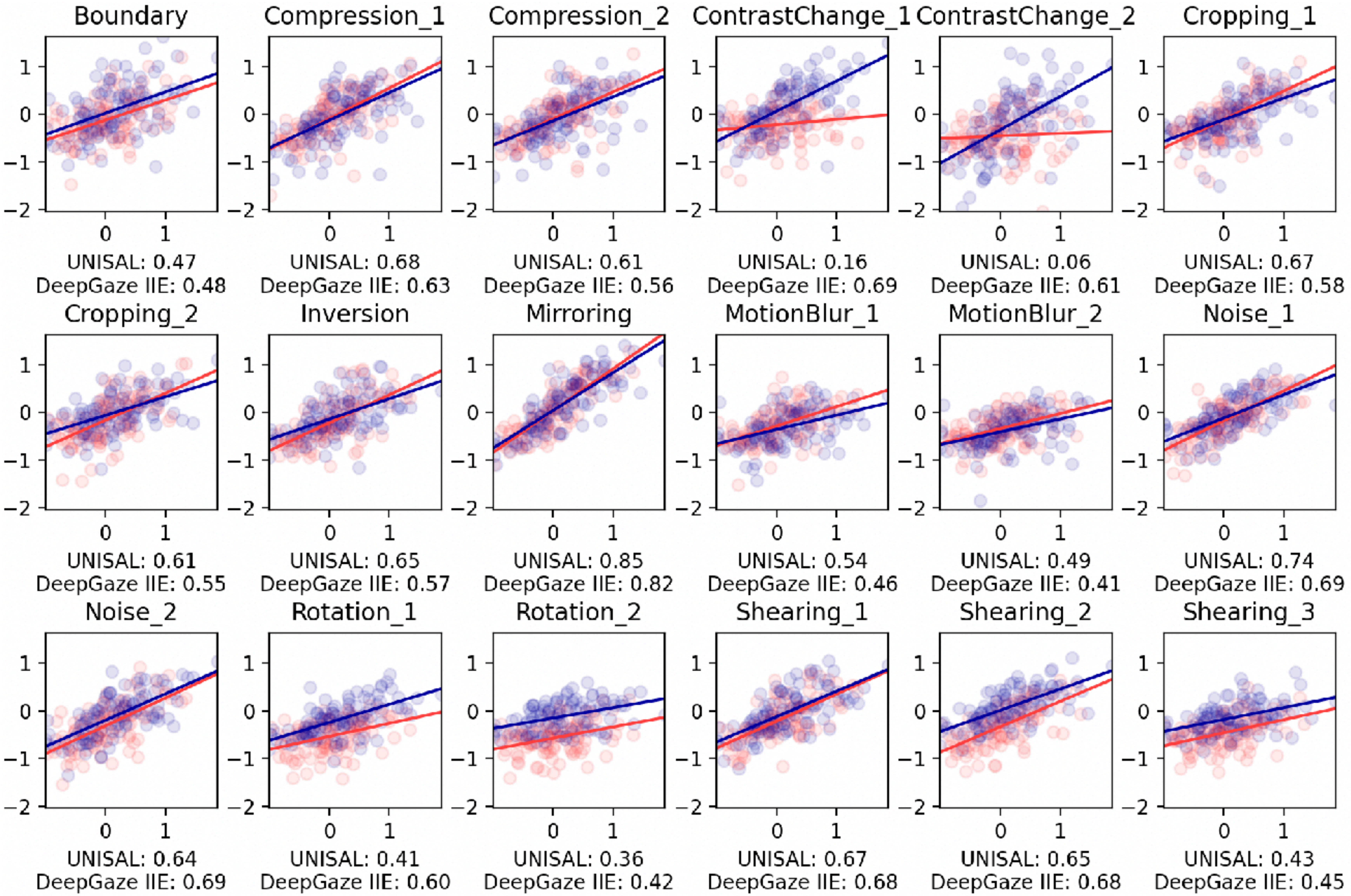
Boundary
DeepGaze IIE: -0.0519
UNISAL: -0.1211
Reference
Boundary
Compression
DeepGaze IIE: -0.1804
UNISAL: -0.1044
Reference
Compression_1
Compression_2
Contrast Change
DeepGaze IIE: -0.3214
UNISAL: -0.4745
Reference
ContrastChange_1
ContrastChange_2
Cropping
DeepGaze IIE: -0.1142
UNISAL: -0.1666
Reference
Cropping_1
Cropping_2
Inversion
DeepGaze IIE: -0.1931
UNISAL: -0.2067
Reference
Inversion
Mirroring
DeepGaze IIE: +0.0060
UNISAL: +0.0243
Reference
Mirroring
Motion Blur
DeepGaze IIE: -0.5182
UNISAL: -0.3846
Reference
MotionBlur_1
MotionBlur_2
Noise
DeepGaze IIE: -0.2236
UNISAL: -0.2914
Reference
Noise_1
Noise_2
Rotation
DeepGaze IIE: -0.3167
UNISAL: -0.9404
Reference
Rotation_1
Rotation_2
Shearing
DeepGaze IIE: -0.3224
UNISAL: -0.6835
Reference
Shearing_1
Shearing_2
Shearing_3
1.0
0.5
0.0
−0.5

Results



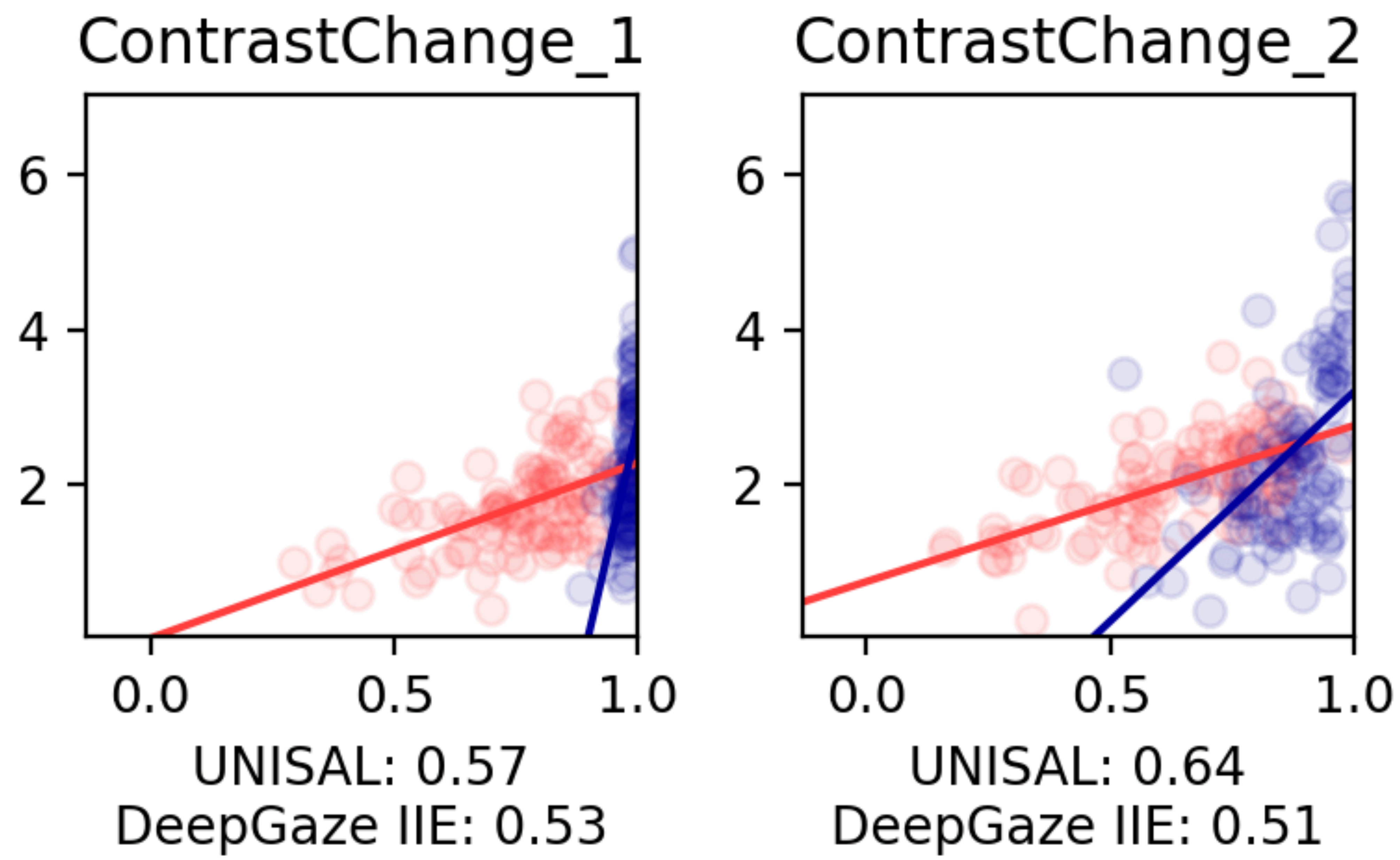
Boundary
UNISAL: 0.52
DeepGaze IIE: 0.74
Compression_1
UNISAL: 0.86
DeepGaze IIE: 0.83
Compression_2
UNISAL: 0.82
DeepGaze IIE: 0.79
ContrastChange_1
UNISAL: 0.43
DeepGaze IIE: 0.83
ContrastChange_2
UNISAL: 0.11
DeepGaze IIE: 0.79
Cropping_1
UNISAL: 0.84
DeepGaze IIE: 0.77
Cropping_2
UNISAL: 0.83
DeepGaze IIE: 0.75
Inversion
UNISAL: 0.74
DeepGaze IIE: 0.79
Mirroring
UNISAL: 0.93
DeepGaze IIE: 0.91
MotionBlur_1
UNISAL: 0.78
DeepGaze IIE: 0.74
MotionBlur_2
UNISAL: 0.74
DeepGaze IIE: 0.76
Noise_1
UNISAL: 0.87
DeepGaze IIE: 0.83
Noise_2
UNISAL: 0.72
DeepGaze IIE: 0.81
Rotation_1
UNISAL: 0.55
DeepGaze IIE: 0.75
Rotation_2
UNISAL: 0.49
DeepGaze IIE: 0.59
Shearing_1
UNISAL: 0.77
DeepGaze IIE: 0.84
Shearing_2
UNISAL: 0.74
DeepGaze IIE: 0.83
Shearing_3
UNISAL: 0.46
DeepGaze IIE: 0.60

Results



Boundary
UNISAL: 0.47
DeepGaze IIE: 0.48
Compression_1
UNISAL: 0.68
DeepGaze IIE: 0.63
Compression_2
UNISAL: 0.61
DeepGaze IIE: 0.56
ContrastChange_1
UNISAL: 0.16
DeepGaze IIE: 0.69
ContrastChange_2
UNISAL: 0.06
DeepGaze IIE: 0.61
Cropping_1
UNISAL: 0.67
DeepGaze IIE: 0.58
Cropping_2
UNISAL: 0.61
DeepGaze IIE: 0.55
Inversion
UNISAL: 0.65
DeepGaze IIE: 0.57
Mirroring
UNISAL: 0.85
DeepGaze IIE: 0.82
MotionBlur_1
UNISAL: 0.54
DeepGaze IIE: 0.46
MotionBlur_2
UNISAL: 0.49
DeepGaze IIE: 0.41
Noise_1
UNISAL: 0.74
DeepGaze IIE: 0.69
Noise_2
UNISAL: 0.64
DeepGaze IIE: 0.69
Rotation_1
UNISAL: 0.41
DeepGaze IIE: 0.60
Rotation_2
UNISAL: 0.36
DeepGaze IIE: 0.42
Shearing_1
UNISAL: 0.67
DeepGaze IIE: 0.68
Shearing_2
UNISAL: 0.65
DeepGaze IIE: 0.68
Shearing_3
UNISAL: 0.43
DeepGaze IIE: 0.45

Results



ContrastChange_1
6
4
2
0.0
0.5
1.0
UNISAL: 0.57
DeepGaze IIE: 0.53
ContrastChange_2
6
4
2
0.0
0.5
1.0
UNISAL: 0.64
DeepGaze IIE: 0.51

Results

Name	Published	Code	IG [?]	AUC [?]	sAUC [?]	NSS [?]	Our NSS:
Gold Standard	*Gaussian kernel density estimate using all fixations of an image with uniform mixture component. Crossvalidated over subjects. Joint performance.*		*0.8026*	0.9159	0.7865	2.7429	
Gold Standard (leave-one-subject-out)	*Gaussian kernel density estimate using all fixations of an image with uniform mixture component. Crossvalidated over subjects. Leave-one-subject-out performance*		*0.4730*	0.8840	0.6930	2.4878	
DeepGaze IIE	A. Linardos, M. Kümmerer, O. Press, M. Bethge: DeepGaze IIE: Calibrated prediction in and out-of-domain for state-of-the-art saliency modeling [ICCV 2021]	python (pytorch)	*0.1893*	0.8692	0.6677	2.1122	2.4429
Forebrain | Attentive AI (v1)			0.1134	0.8598	0.6432	1.9664	
DeepGaze II	Matthias Kümmerer, Thomas S. A. Wallis, Leon A. Gatys, Matthias Bethge. Understanding Low- and High-Level Contributions to Fixation Prediction [ICCV 2017]	python (tensorflow)	*0.0839*	0.8640	0.6498	1.9619	
UNISAL	R. Droste, J. Jiao, J.A. Noble: Unified Image and Video Saliency Modeling. ECCV 2020 (arXiv)	python (pytorch)	0.0321	0.8604	0.6684	1.9359	2.1563

Image Sources

- Che et al. - “How is Gaze Influenced by Image Transformations? Dataset and Model”
  Slides 3, 8, 9, 11.
  https://arxiv.org/abs/1905.06803

- Our work.
  Slides 3, 10, 11, 14, 15, 16, 17, 18, 19, 20.
  https://codeberg.org/soundeffects/transformations_and_gaze_prediction

- Brainsight - “Gaze Plots: optimizing visual hierarchy with insights from viewing behaviour”
  Slide 4.
  https://www.brainsight.app/post/gaze-plots-optimizing-visual-hierarchy

- Wikipedia - “Object Detection”
  Slides 5 and 7.
  https://en.wikipedia.org/wiki/Object_detection

- Kummerer et al. - “Predicting Visual Fixations”
  Slide 6.
  https://www.annualreviews.org/content/journals/10.1146/annurev-vision-120822-072528

- Screenshot from “Return of the Obra Dinn”
  Slide 7.
  https://obradinn.com/

- MIT/Tuebingen Saliency Benchmark
  Slide 12, 21.
  https://saliency.tuebingen.ai/results_CAT2000.html

- Bylinskii et al. - “What do different evaluation metrics tell us about saliency models?”
  Slide 13.
  https://arxiv.org/abs/1604.03605

Discussion

Final Examination