

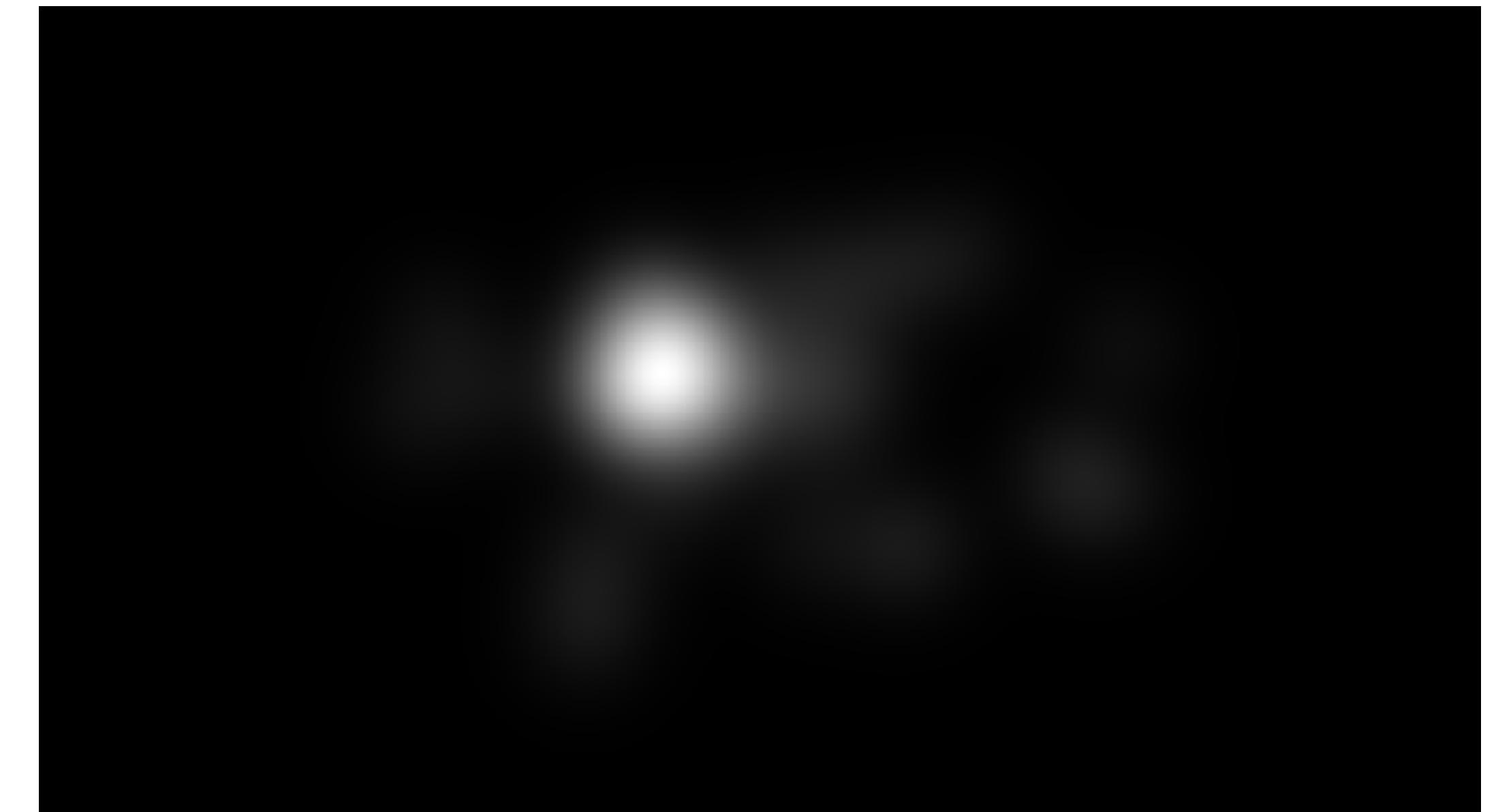
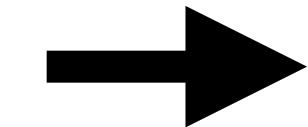
Digital Gaze Transformations Degrade Gaze Prediction Accuracy

James Youngblood, Oct. 21, 2025

Outline

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

Introduction

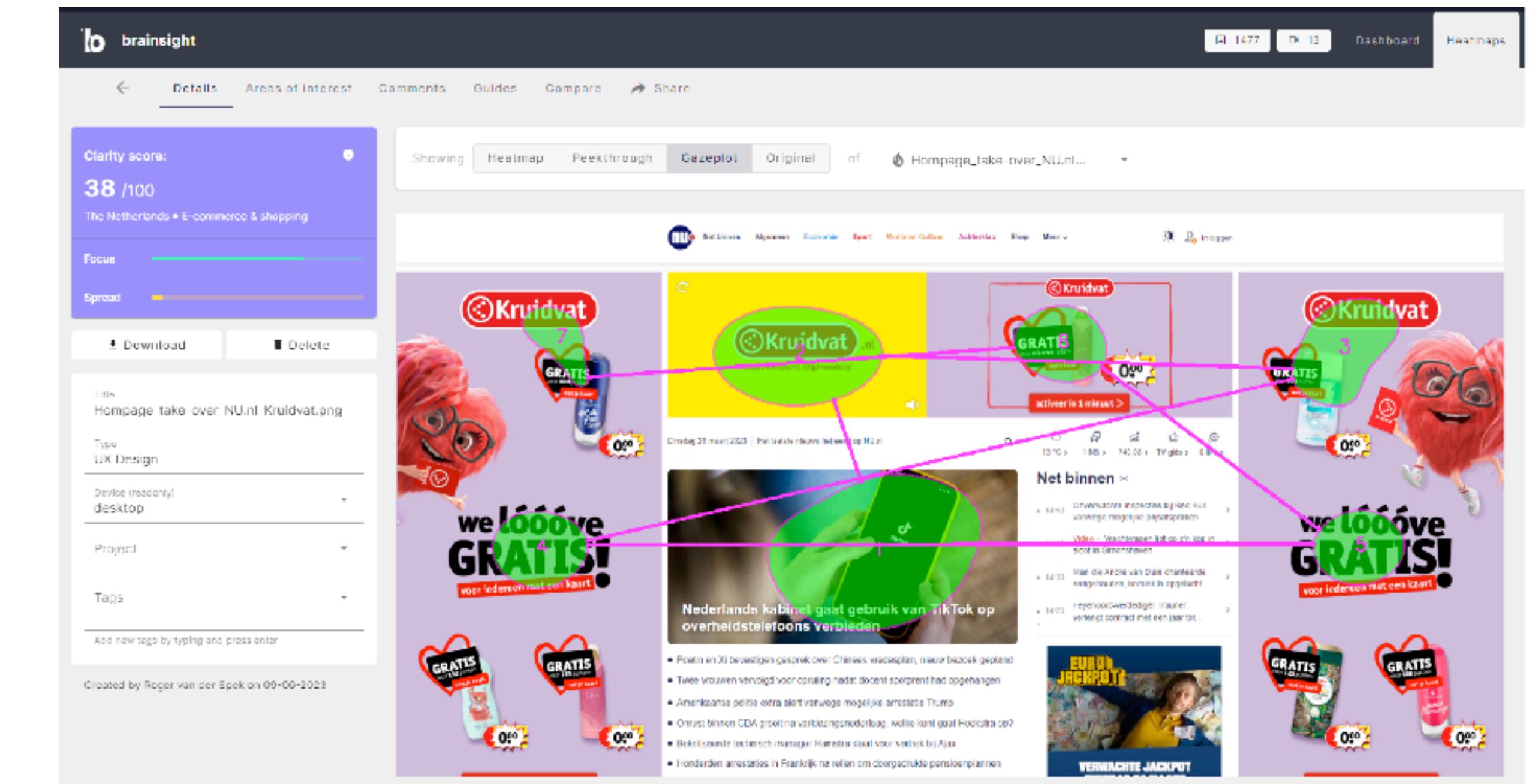


Image

Gaze Distribution

Introduction

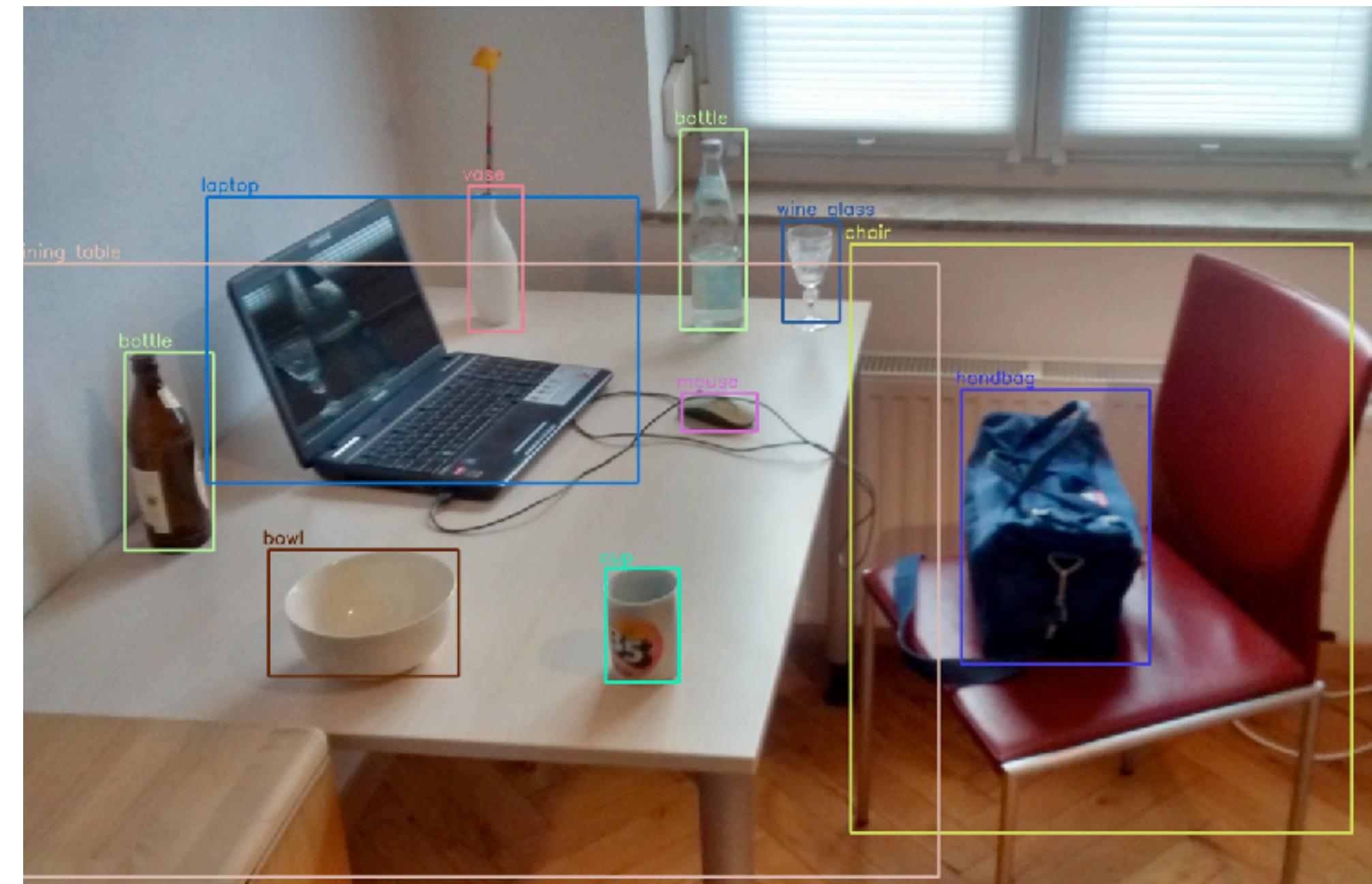
- User experience
- Data Visualization
- Visual Media
- Computer vision and robotics



A plot of gaze behavior for a user interface

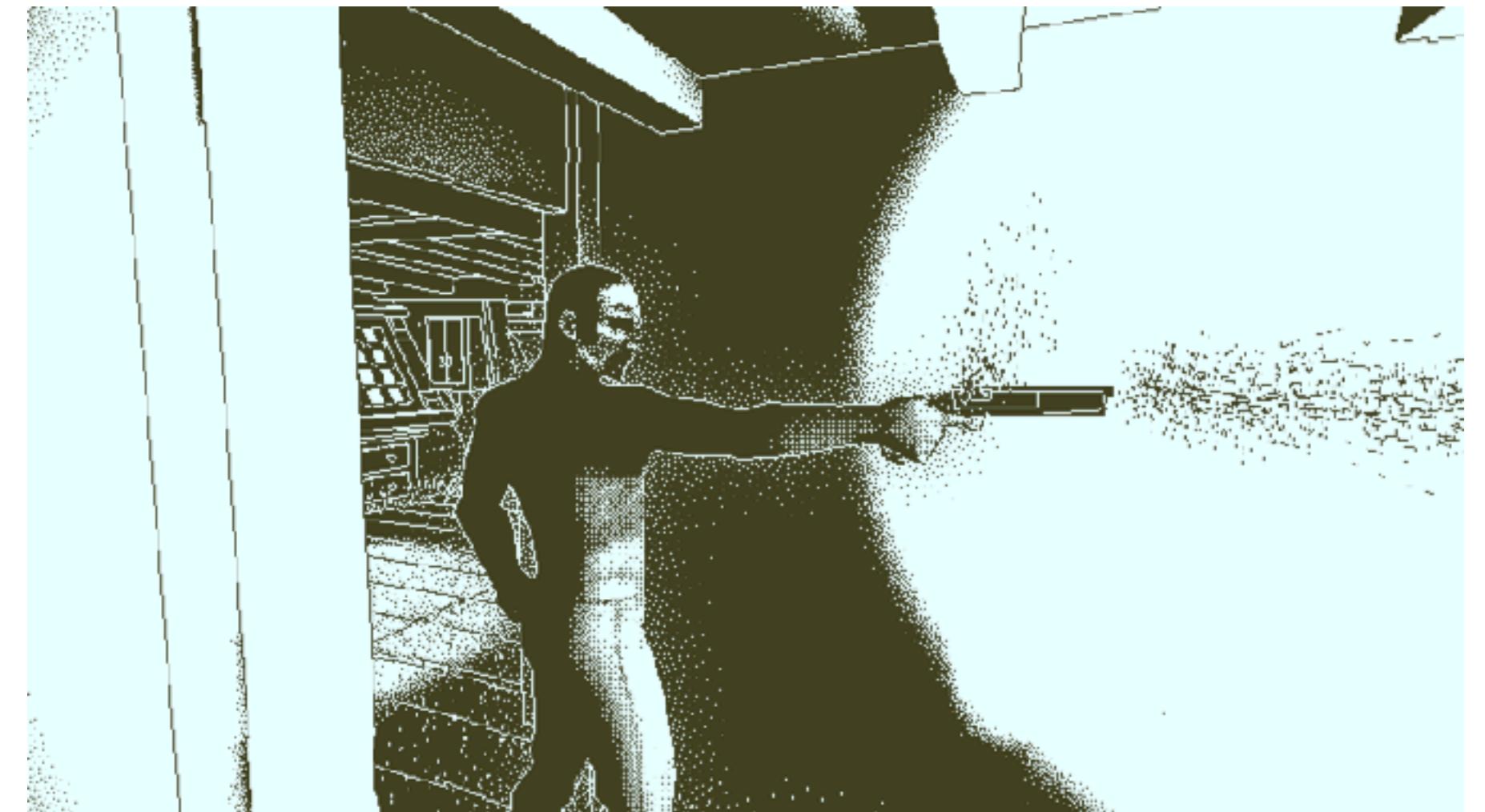
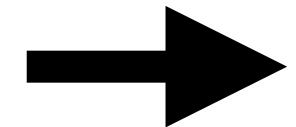
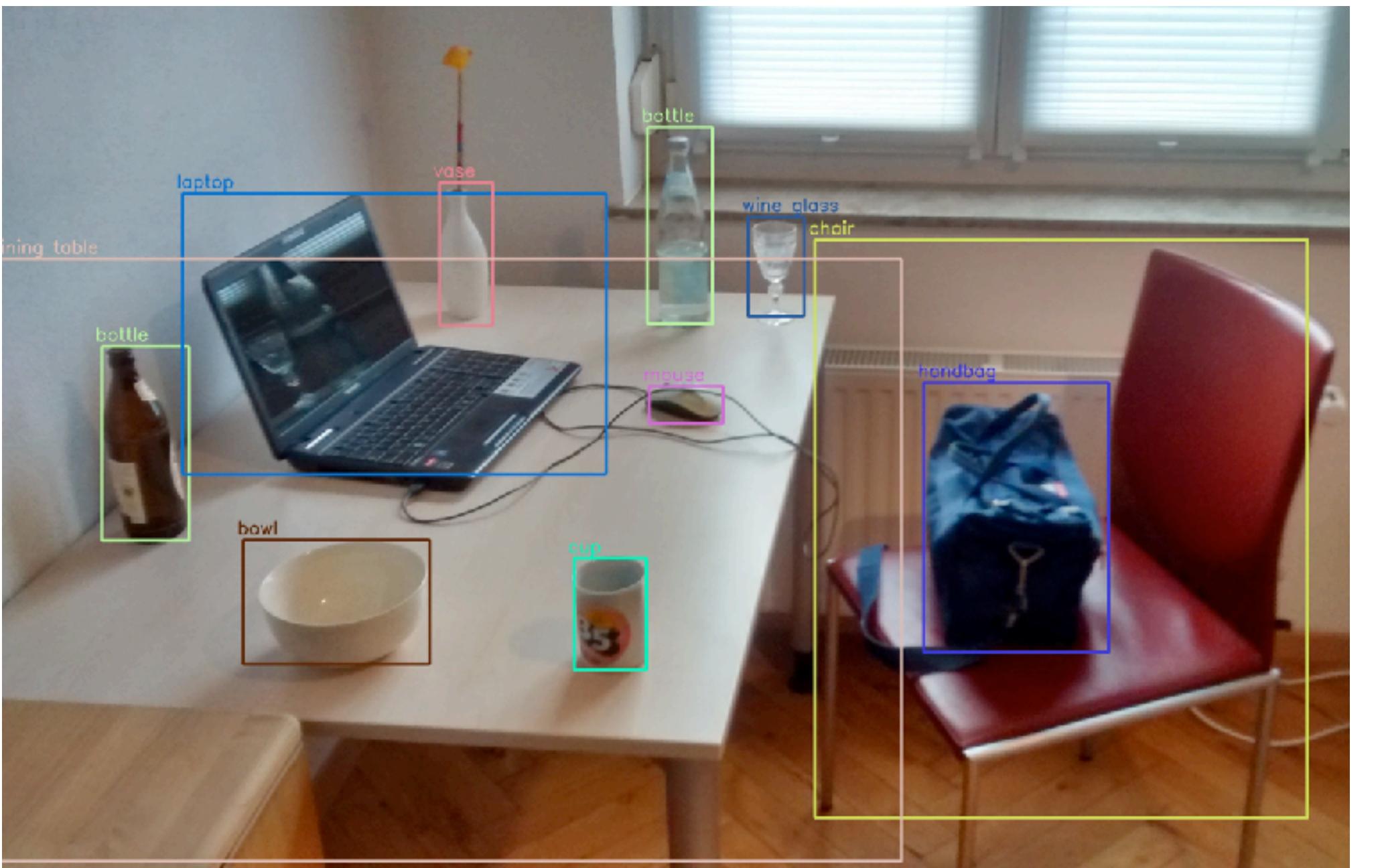
Introduction

Transfer learning (object detection to gaze prediction)



Introduction

Will the model generalize?

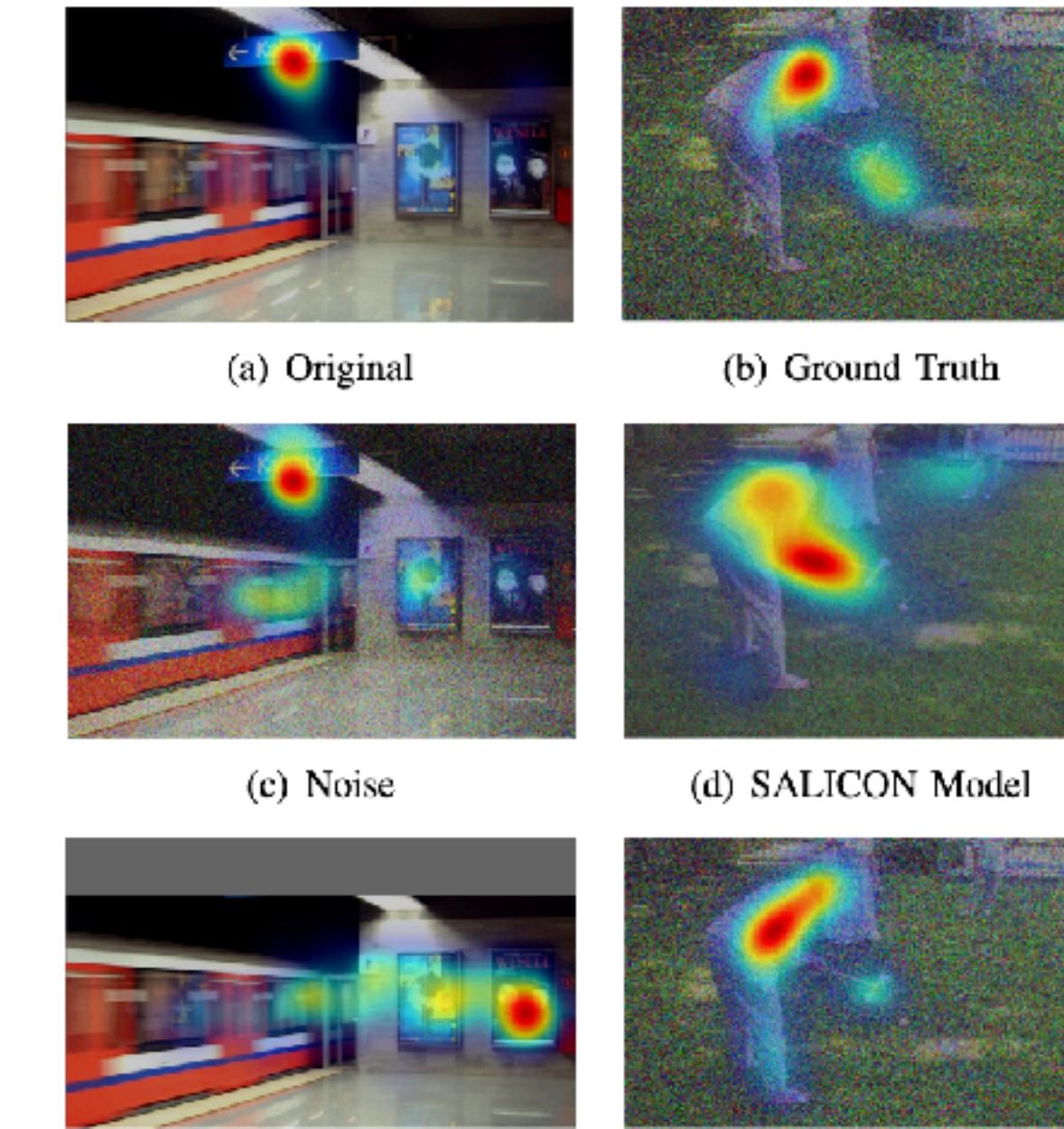


Related Work

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*

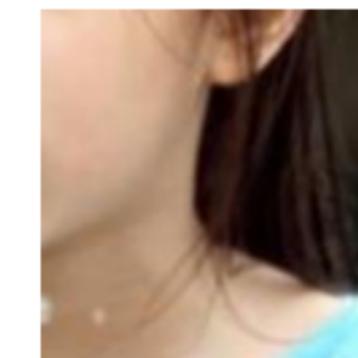
Abstract—Data size is the bottleneck for developing deep saliency models, because collecting eye-movement data is very time-consuming and expensive. Most of current studies on human attention and saliency modeling have used high-quality stereotype stimuli. In real world, however, captured images undergo 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-



Related Work



Reference



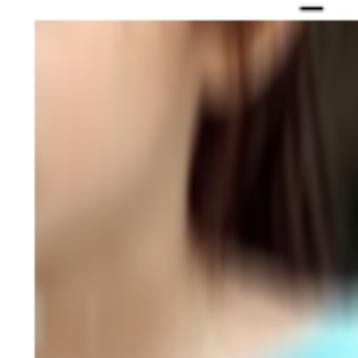
Boundary



Compression_1



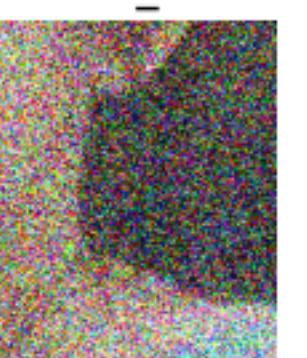
MotionBlur_2



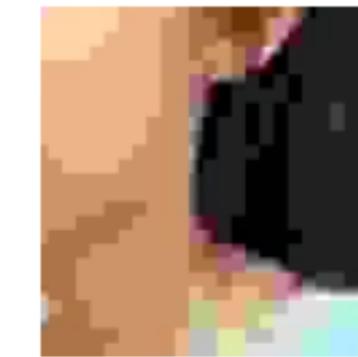
Noise_1



Noise_2



Compression_2



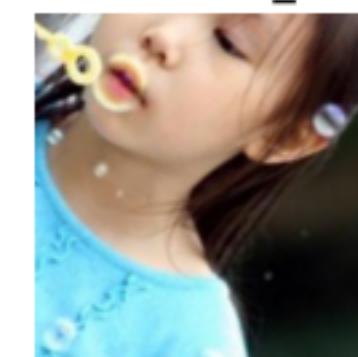
ContrastChange_1



ContrastChange_2



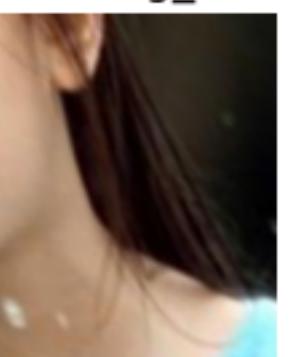
Rotation_1



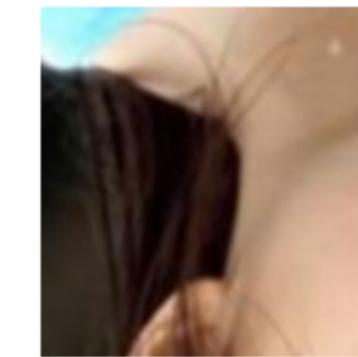
Rotation_2



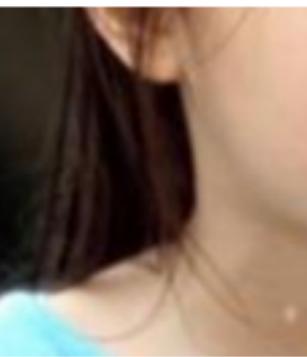
Shearing_1



Inversion



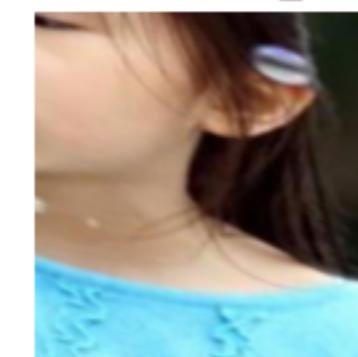
Mirroring



MotionBlur_1



Shearing_2



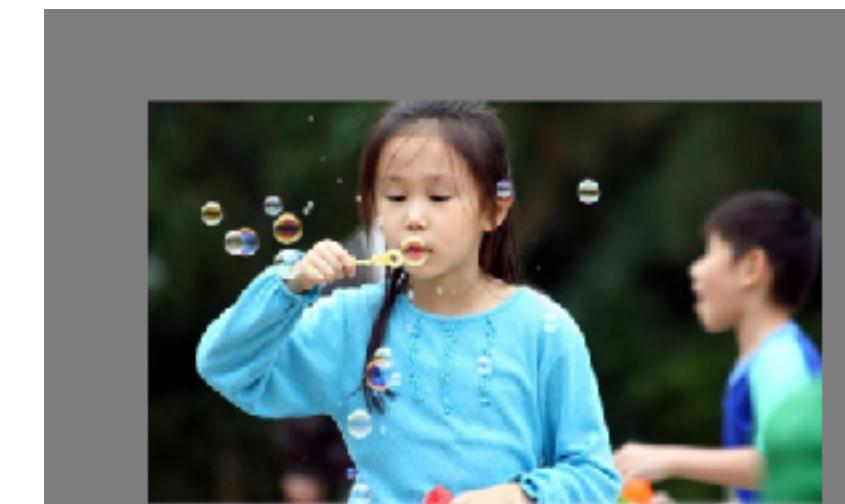
Shearing_3



Cropping_1

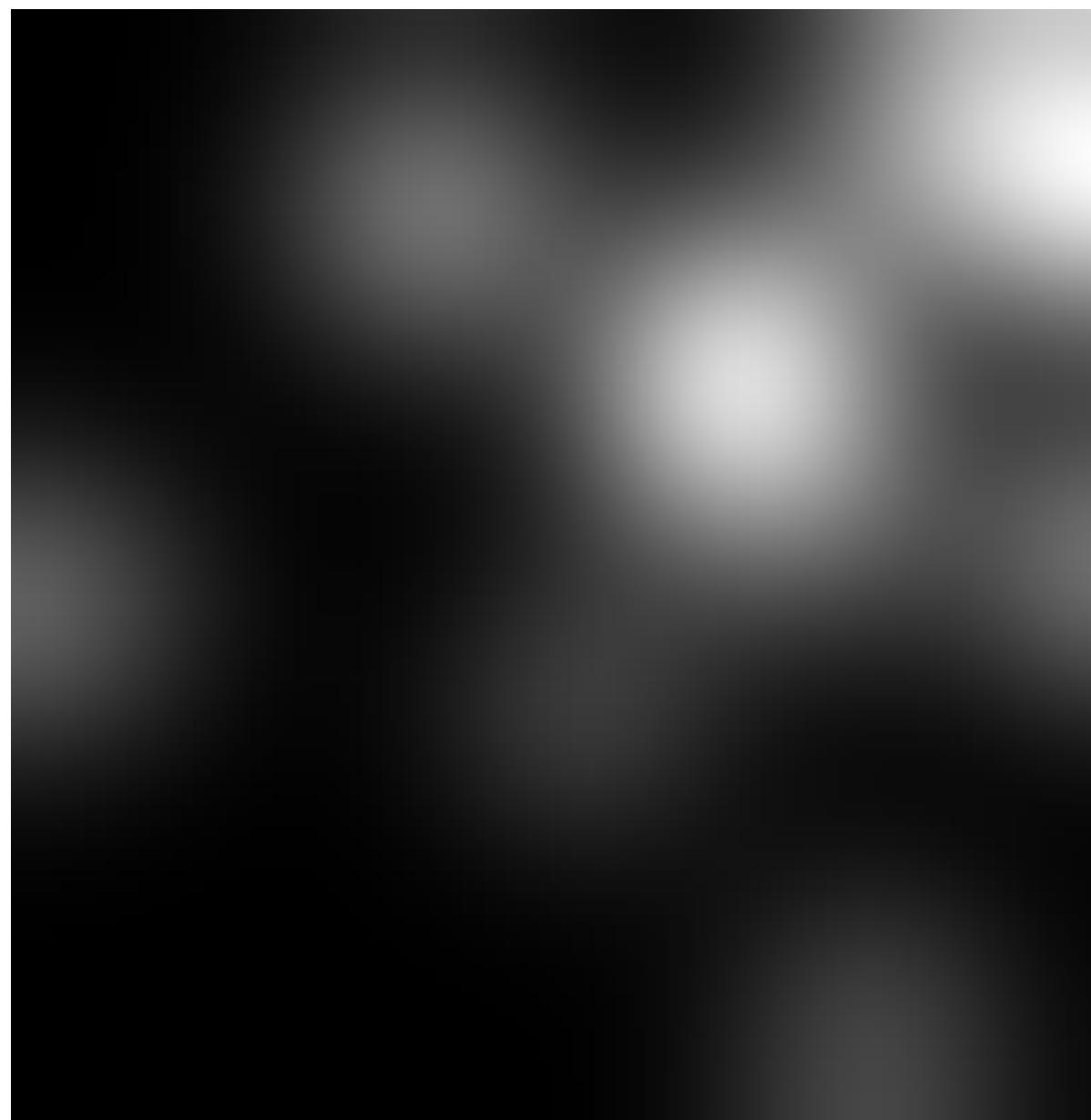
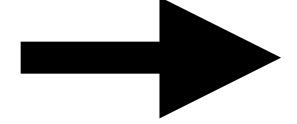


Cropping_2



Background

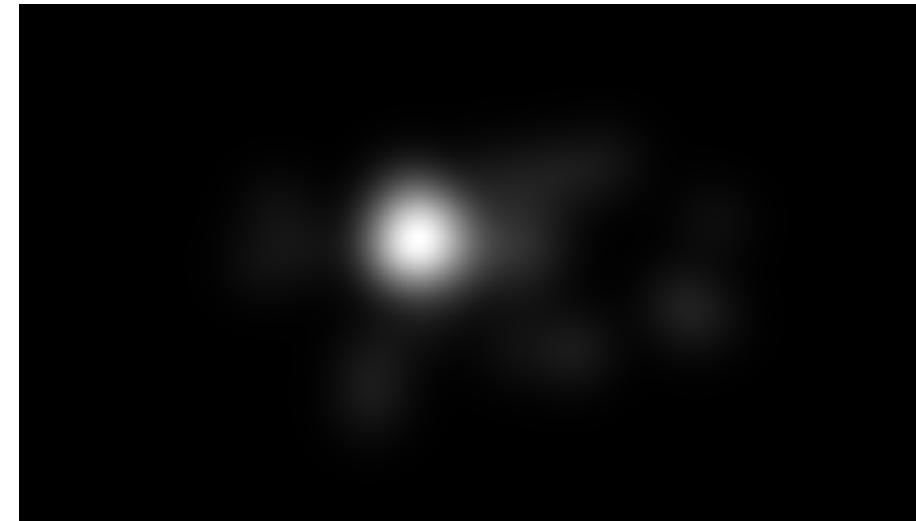
Saliency maps



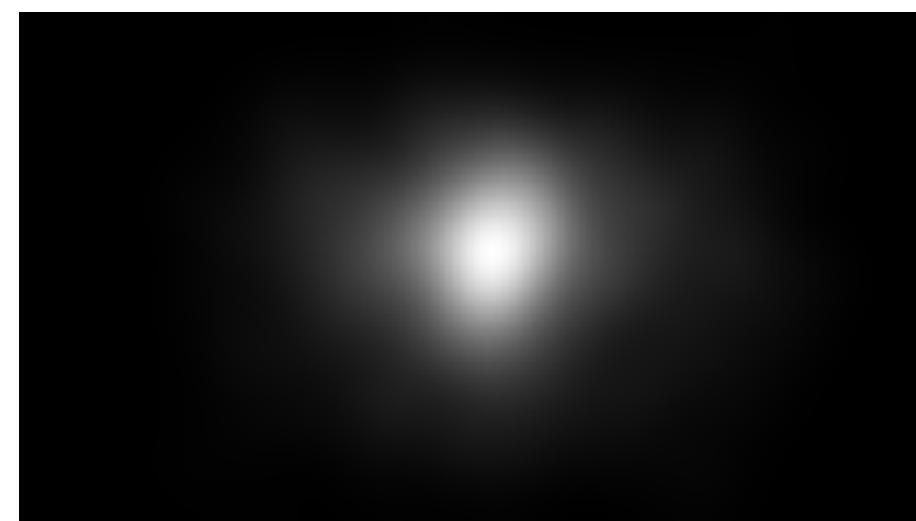
Background



Real



Center bias



Background

MIT/Tuebingen Saliency Benchmark

Leaderboard CAT2000

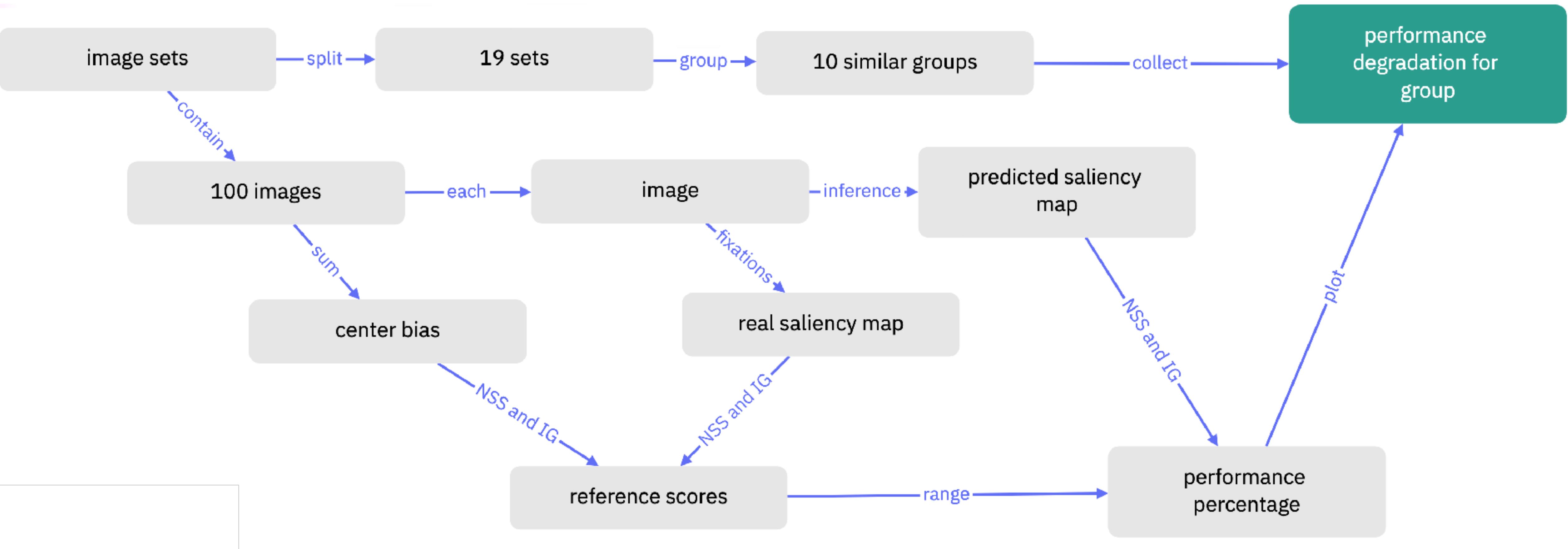
Click on a metric name to sort that metric. Models evaluated as probabilistic models are shown with green background. The performance under the metric which a model has been trained on is shown in italic. The code for evaluation models can be found [here](#).

Name	Published	Code	IG	AUC	sAUC	NSS	CC	KLDiv	SIM	Date tested
Gold Standard	<i>Gaussian kernel density estimate using all fixations of an image with uniform mixture component. Crossvalidated over subjects. Joint performance.</i>		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)	<i>Gaussian kernel density estimate using all fixations of an image with uniform mixture component. Crossvalidated over subjects. Leave-one-subject-out performance</i>		0.4730	0.8840	0.6930	2.4878				First tested 2023-07-05 Last tested 2023-07-05
DeepGaze II E	A. Linardos, M. Kümmerer, O. Press, M. Bethge: DeepGaze II: 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

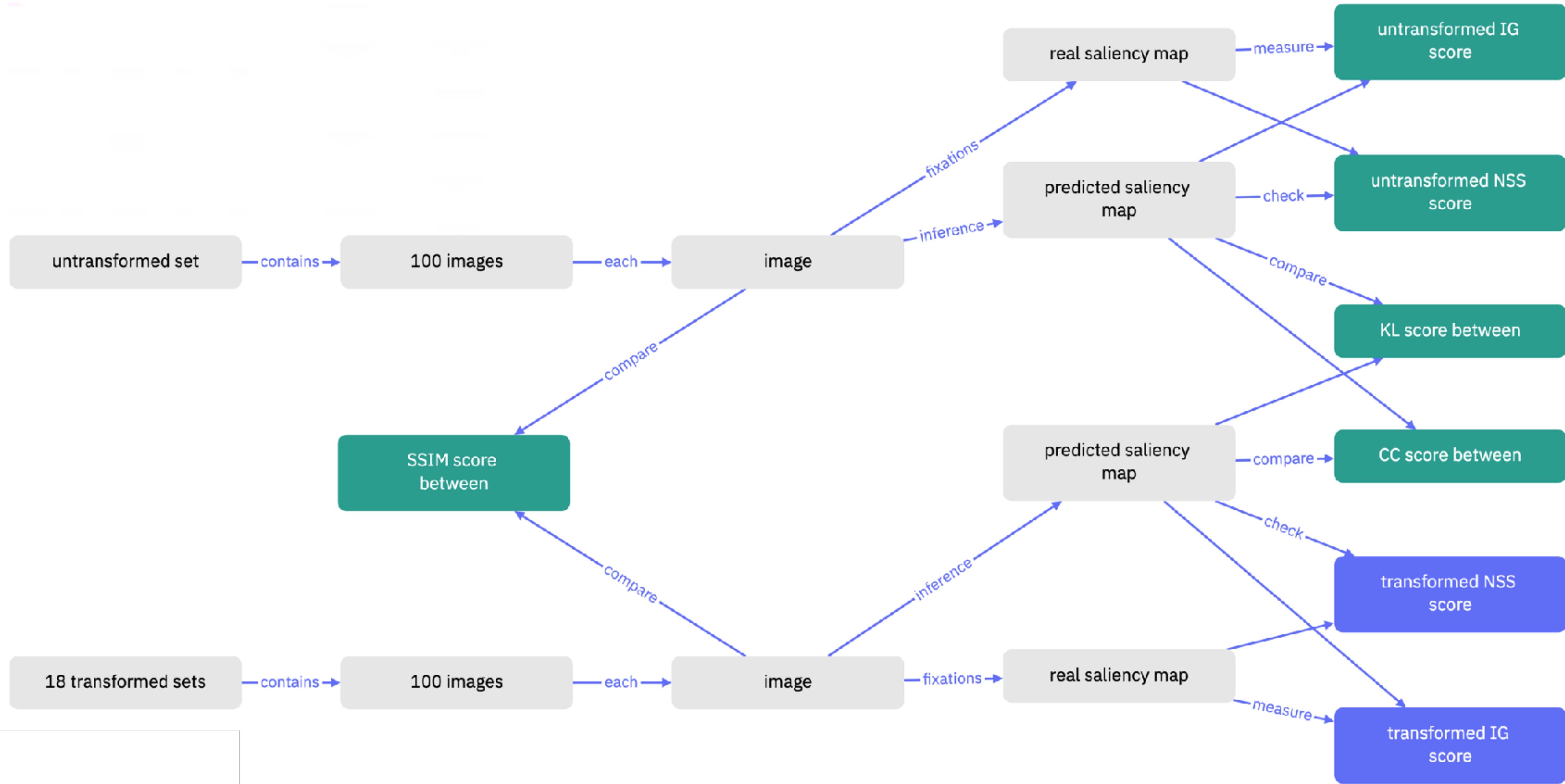
Background

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.

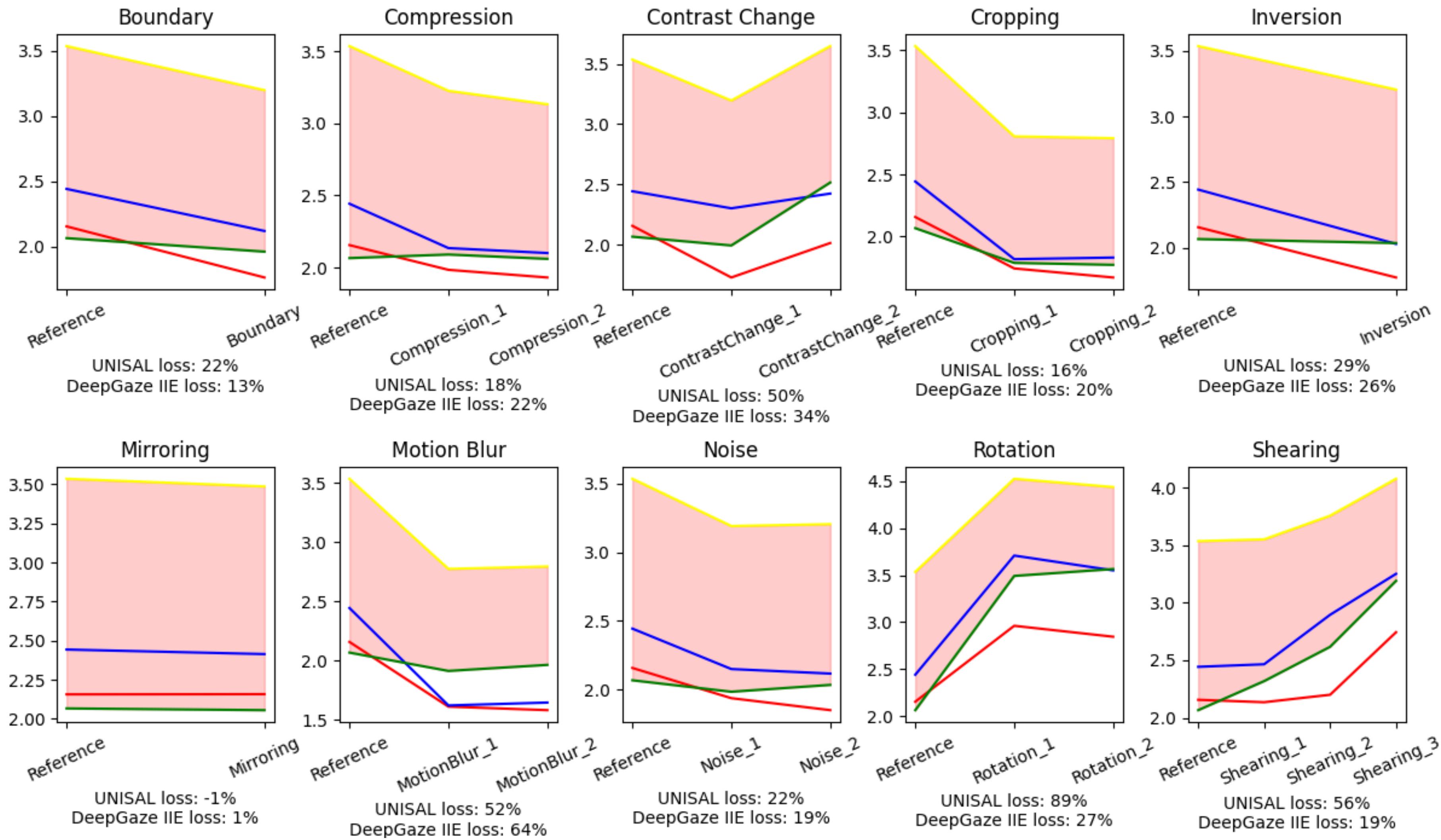
Method



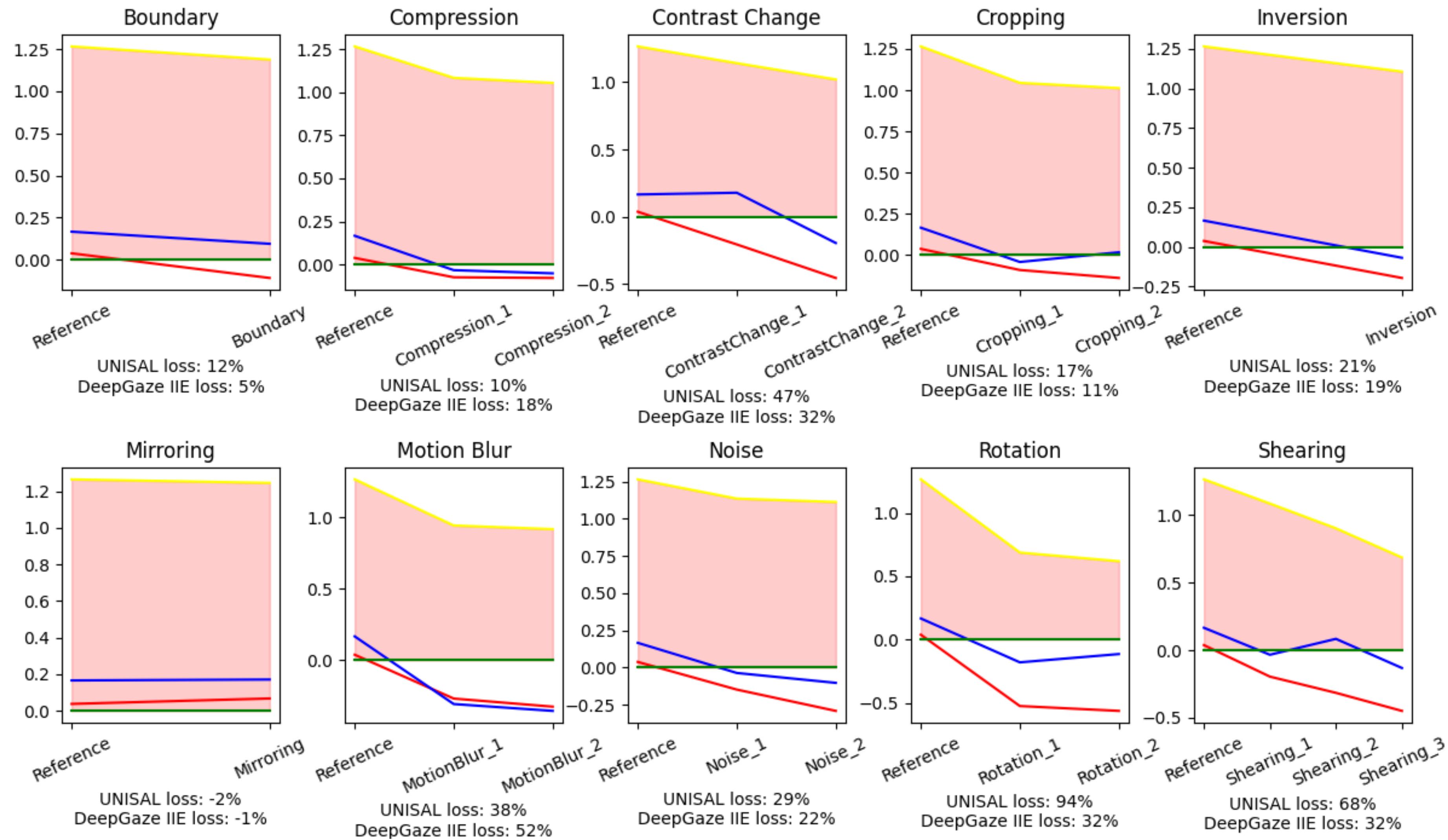
Method



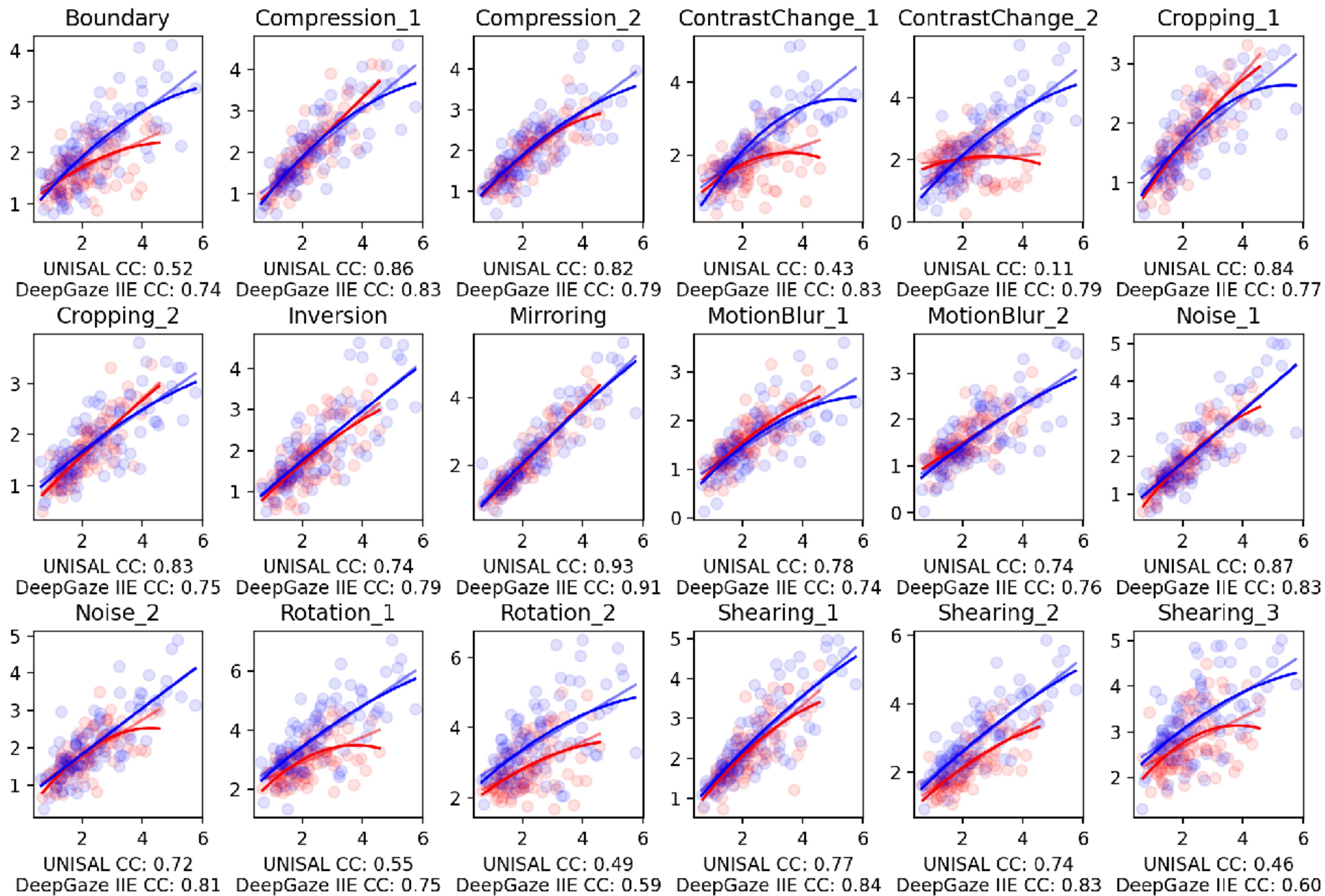
Results



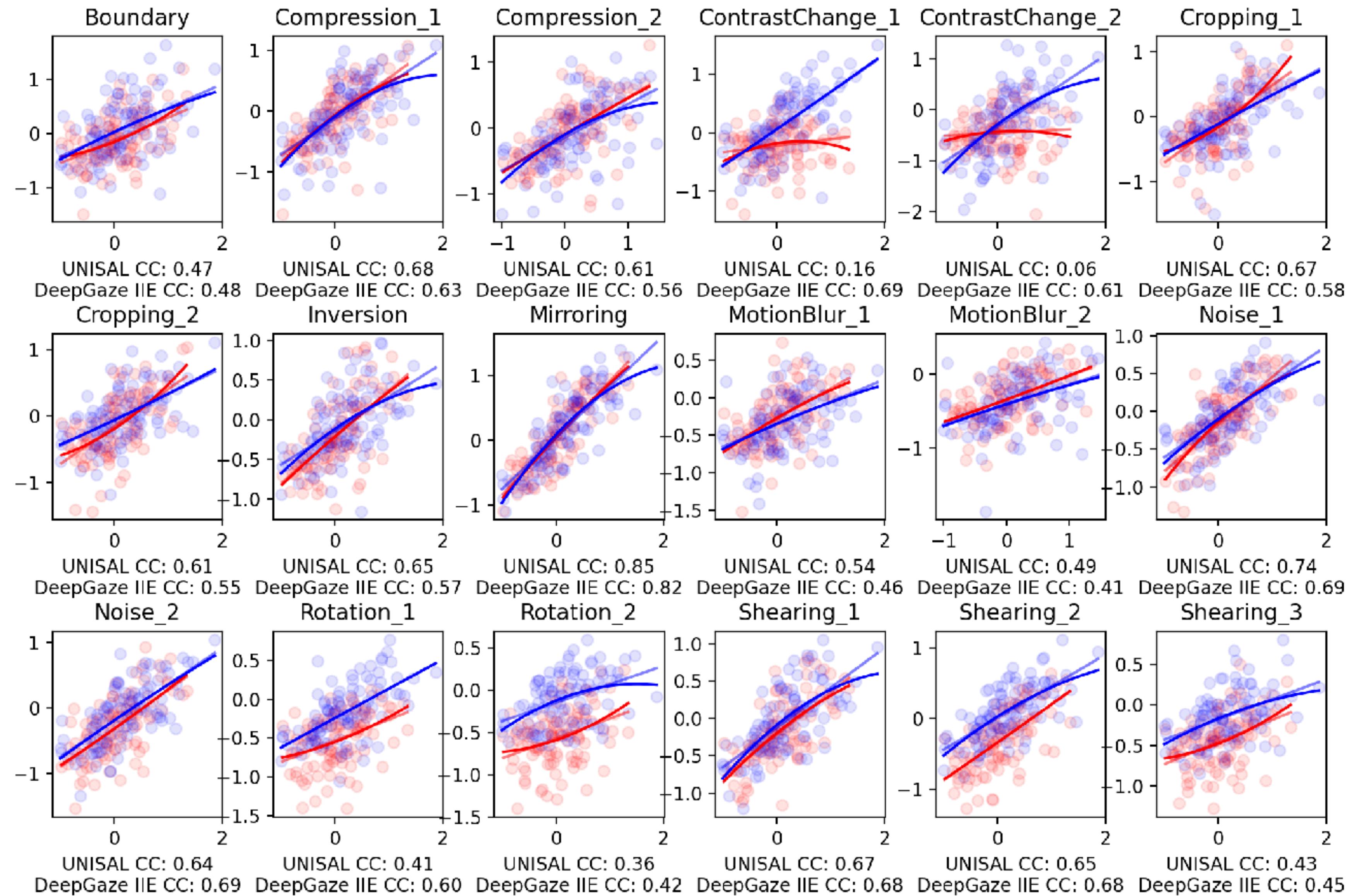
Results



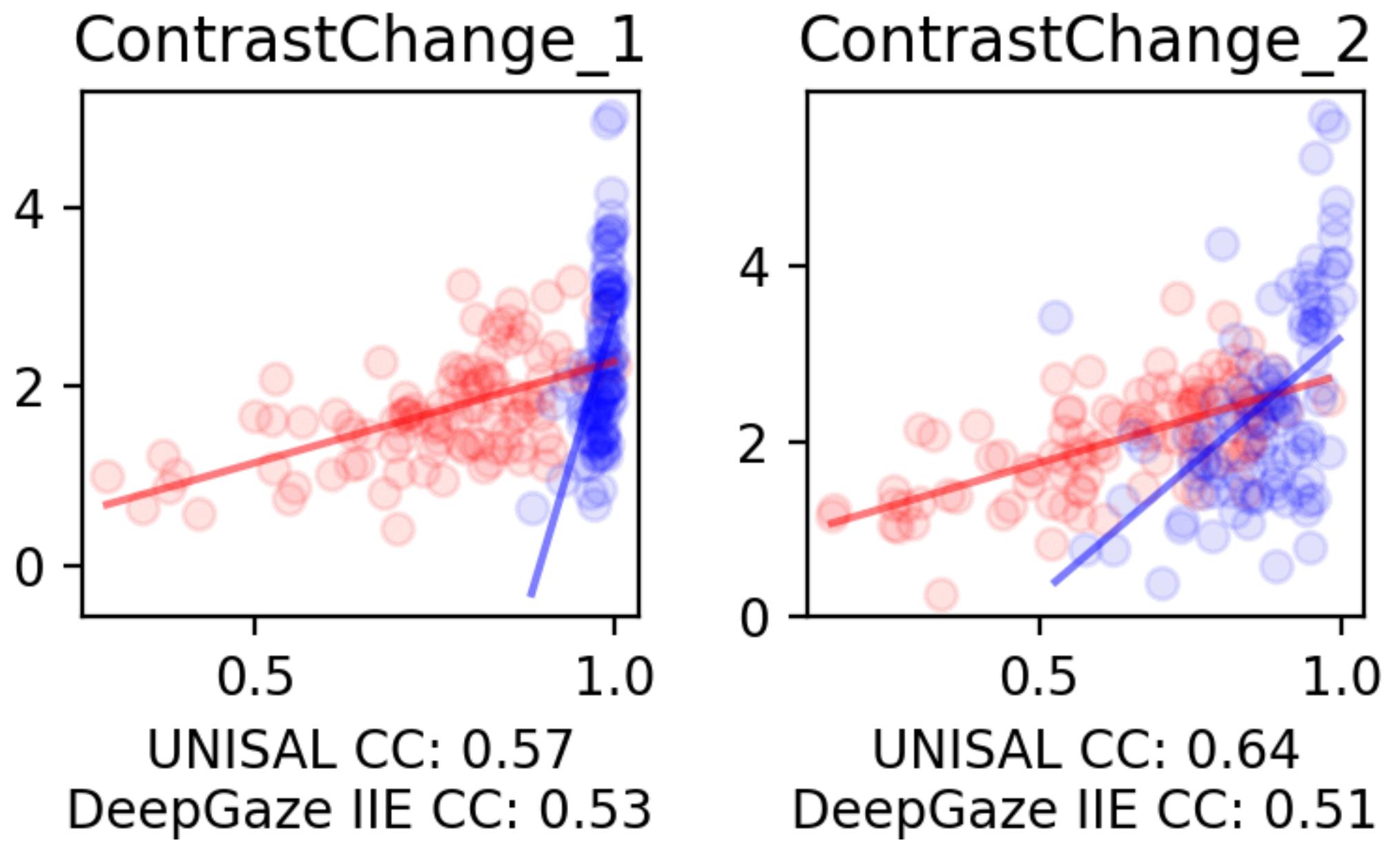
Results



Results



Results



Results

Name	Published	Code	IG [?]	AUC [?]	sAUC [?]	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
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

Our NSS:

2.4429

2.1563

Image Sources

- Che et al. - “How is Gaze Influenced by Image Transformations? Dataset and Model”
Slides 3, 7, 8, 10.
<https://arxiv.org/abs/1905.06803>
- Our work.
Slides 3, 9, 10, 13, 14, 15, 16, 17, 18, 19.
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 6.
https://en.wikipedia.org/wiki/Object_detection
- Screenshot from “Return of the Obra Dinn”
Slide 6.
<https://obradinn.com/>
- MIT/Tuebingen Saliency Benchmark
Slide 11, 20.
https://saliency.tuebingen.ai/results_CAT2000.html
- Bylinskii et al. - “What do different evaluation metrics tell us about saliency models?”
Slide 12.
<https://arxiv.org/abs/1604.03605>

Discussion

Final Examination