Modeling Saliency Dataset Bias

Matthias Kümmerer Harneet Singh Khanuja* Matthias Bethge

Tübingen AI Center
University of Tübingen

firstname.lastname@bethgelab.org

Abstract

Recent advances in image-based saliency prediction are approaching gold standard performance levels on existing benchmarks. Despite this success, we show that predicting fixations across multiple saliency datasets remains challenging due to dataset bias. We find a significant performance drop (around 40%) when models trained on one dataset are applied to another. Surprisingly, increasing dataset diversity does not resolve this inter-dataset gap*, with close to 60% attributed to dataset-specific biases. To address this remaining* generalization gap*, we propose a novel architecture extending a mostly dataset-agnostic encoder-decoder structure with fewer than 20 dataset-specific parameters that govern interpretable mechanisms such as multi-scale structure, center bias, and fixation spread. Adapting only these parameters to new data accounts for more than 75% of the generalization gap, with a large fraction of the improvement achieved with as few as 50 samples. Our model sets a new state-of-the-art on all three datasets of the MIT/Tuebingen Saliency Benchmark (MIT300, CAT2000, and COCO-Freeview), even when purely generalizing from unrelated datasets, but with a substantial boost when adapting to the respective training datasets. The model also provides valuable insights into spatial saliency properties, revealing complex multi-scale effects that combine both absolute and relative sizes.*

1. Introduction

Understanding and predicting where we look is valuable for numerous reasons. Scientifically, it provides insights into visual processing in the retina and the brain, memory, emotions and cognitive processes and task driven behaviour. Practically, it enhances applications such as better compression methods, optimized layouts, robotics, and efficient allocation of computational resources. The field of saliency prediction – the endeavor of predicting where humans look in images using *saliency models* – is highly active, with a plethora of different saliency models being developed.

The standard benchmark in the field is the MIT300 dataset [27] of the MIT/Tuebingen Saliency Benchmark [33]. Recently, the performance on this benchmark has started to plateau [32], raising questions about whether the field has reached its limits in spatial saliency prediction [15].

While there is plenty of room for extensions such as incorporating dynamic eye movements [1, 17, 55, 70], stimulus dynamics [15, 40], or more top-down and task-driven effects [9, 52, 56], spatial saliency remains crucial for many applications. For concluding that spatial saliency is solved to an extent that is useful for real-world applications, models would need to perform well beyond datasets seen during training. To that end, we conduct a systematic study across five different large saliency datasets and test how well models transfer from one or multiple of these datasets to another. We find a large generalization gap even when training on multiple different saliency datasets. This indicates that the advantages of training on vast amounts of data – the key behind the success of recent (foundation) models – are outweighed by dataset-specific differences.

To address this issue, we propose a saliency model that is able to account for dataset biases with less than 20 interpretable parameters. Adapting only the bias parameters to new data closes about 75% of the generalization gap and can be done on as little as 50 samples. This allows us to substantially improve the state-of-the-art on the MIT300 [27], CAT2000 [4] and COCO-Freeview [10] benchmarks in generalization, adaptation and full-training settings. Furthermore, analyzing the learned dataset specific parameters provides valuable insights into the variability of saliency across datasets. Our main contributions are as follows:

- We identify and quantify large performance penalties when transferring predictions from one or many saliency datasets to an unseen saliency dataset (inter-dataset gap and generalization gap).

*Now at Georgia Tech

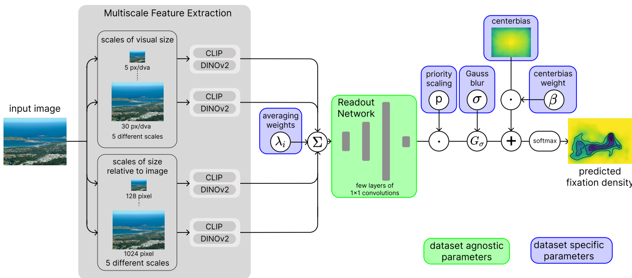

Figure 1. Model Architecture: An input image is rescaled into different resolutions, some defined in total image size in pixels, others in pixels per degree of visual angle. For each image, deep activations from CLIP and DINOv2 encoders are extracted and averaged across scales, from which a priority map is decoded which is then postprocessed with Blur, priority scaling and centerbias. See Appendix Figure 9 for a larger version.

- We attribute the generalization gap to multiple dataset biases that hinder high saliency prediction performance on unseen datasets: center bias, multiscale distribution, priority scaling and fixation scatter
- We propose a new saliency model built on a multiscale backbone architecture which features a simple decoder and incorporates less than 20 interpretable dataset-specific parameters to account for the identified dataset biases.
- We demonstrate that adapting only the dataset biases parameters to unseen data successfully closes about 75% of the generalization gap in a very data-efficient and parameter-efficient way.
- We set a new state-of-the-art on the MIT300, CAT2000, COCO-Freeview benchmarks for generalization, adaptation and full training settings, improving AUC performance by at least 1% compared to previous state-of-the-art on all benchmarks.

2. Related Works

Since Itti and Koch proposed the first image computable saliency model [24], a myriad of different saliency models have been proposed. Early saliency models mostly followed the spirit of the Feature Integration Theory [64] and proposed that saliency depends on low-level image features such as contrast and edges [5, 20, 29, 51, 63, 75]. Later, more and more high-level and semantic image features have been used to predict where people look [7, 27] and nowadays in the computer vision community saliency denotes whatever predicts where people look. The eDN model [65] introduced deep learning to the field and DeepGaze I [34] introduced transfer learning from deep features to improve prediction performance. Since then, all high-performing saliency models are deep learning based models using some kind of feature transfer from other computer vision tasks [22, 30, 31, 36]. The concept of deep feature transfer has since then been augmented with a variety of techniques to improve prediction performance.

The SALICON model [22] uses a two-scale backbone. Here, we extend and adapt this approach to use a substantially larger number of scales of absolute and relative size, and we don't concatenate the features from multiple scales, but average them. EML-NET [25] uses multiple layers from multiple backbones, which is related to our approach of combining deep features from CLIP and DINOv2 backbones. The UNISAL model [15] trains a saliency model jointly for video and image saliency datasets and uses some domain specific modules to adapt to the different domains. Here we focus only on combining multiple image saliency datasets and analysing the differences within this domain. Furthermore, in UNISAL, most domain specific parameters are not interpretable (batch norm adaptation and linear weights for deep features). Finally, while both UNISAL and we learn domain or dataset specific parameters, our number of parameters is magnitudes smaller, enabling data-efficient adaptation to new datasets. SalFBNet [13] introduced feedback connections from deeper layers of the encoder network to early layers to allow adaptation of low-level features with high-level knowledge. In addition, they pretrained the model on a pseudo saliency dataset obtained by averaging the predictions of top performing saliency models on a large set of images. DeepGaze IIE [39] employed an ensembling strategy (see also [65]) to average predictions from multiple internal saliency models using different encoder backbones and showed that this results in high performance and good confidence calibration on new datasets. DeepGaze IIE currently represents the state of the art in all datatsets and benchmarks presented in this paper. Most saliency models use convolutional architectures which keeps the spatial information in a very direct way, more recently also transformer architectures have been introduced [14, 43, 72]. Here, we combine convolutional and transformer based backbones. Recently, researchers tried to improve saliency prediction by adding time [2], Augmentations [3], explicit global semantic interactions [71] or self-knowledge distillation [60, 61]

The field of eye movement prediction has recognized that different conditions can affect attention behaviour [73] and models have been proposed to capture this. However, these works usually included much more different modalities where differences in behavior are more expected (e.g., images/video [15] or natural images/webpages and different observer tasks [38]. Opposed to this, [8] compare different freeviewing image datasets and analyze semantic differences in the fixation distribution, but do not adapt them to new datasets.

The problem of dataset biases has been recognized in other fields of machine learning before [42, 62] and a variety of adaptation methods tailored more towards large scale datasets have been proposed. Most notable and successful are the test time adaptation methods [53, 67]. Our approach distinguishes itself from typical test-time adaptation methods in that we build on few mechanisms tailored towards the saliency use case, resulting in few interpretable parameters allowing insights into differences between datasets and highly data-efficient adaptation.

3. Model

Overall model architecture Our model architecture is visualized in Figure 1 and can be seen as a variant of the DeepGaze architecture [36, 39] that has been extended with more modern backbones, a multiscale architecture and where some parameters have been made dataset-dependent. Firstly, an input image is rescaled into different resolutions. Secondly, for each resolution, deep activations from pretrained backbones are extracted. Thirdly, the activations from all scales are scaled into a common resolution and combined into a weighted average. Fourthly, the resulting feature maps are decoded into a single spatial priority map with a *readout network* [36], consisting of five layers of 1x1 convolutions. The small number of parameters makes the readout network trainable on datasets with only 1000 images. Lastly, the spatial priority map is multiplied with a priority scaling factor, blurred with a Gaussian and combined with a precomputed center bias log density map which can be down weighted with a weight factor. The result is converted into a probability distribution over pixels using a softmax. In the following we detail our extensions to previous DeepGaze architectures. For a precise mathematical formulation and more details on the model architecture see Appendix H.

Multiscale feature extractor Our multiscale feature extractor rescales the input image into different resolutions before using the backbones to extract features. Importantly, we use two different sets of scales: The *relative scales* rescale the images to different sizes given in pixels. The relative scales find patterns that have a certain size relative to the full image. The *absolute scales* on the other hand rescale images to have a fixed resolution in pixel per degree of visual angle (px/dva) and hence depend on how large the image was seen by observers. We call these the absolute scales, because they are sensitive to the absolute visual size of objects independent of the image size. Using both scales together allows to disentangle saliency effects of relative and absolute visual size and build a model that can be applied to new datasets in a meaningful way without having to reason about the right input resolution of an image (opposed to all other compared models).

Pretrained Backbones For each resolution we extract deep activations from two different pretrained backbones which are then concatenated: CLIP [49] is very good at encoding global information about a scene [28, 41, 74], while DINOv2 [45] is designed to encode very precise local information. Together they should allow to reason about objects within context.

Dataset Biases Nearly all parameters of our model are encompassed by the readout network with a total of 26,460 parameters which are trained jointly across datasets. The remaining parameters are what we call *dataset bias parameters*: they are dataset specific and control interpretable mechanisms. The dataset bias parameters are: the *multiscale averaging weights* which constitute 10 dataset bias parameters that model the relationship between how large the object is both relative to the image but also in terms of how large the object is perceived visually [48]; the *priority scaling* models how much more salient, *e.g.* a face is compared to a house, and we assume it will vary depending on the experimental conditions, *e.g.*, the engagement of the observers; the *blur size* specifies the size of the blur kernel (specified in dva) as we assume that datasets might differ in how much fixations are spread out around objects; the *center bias* constitutes 2-5 dataset bias parameters depending on how exactly the center bias is modeled (see Appendix E for more details). It is the spatial prior of the model and encodes a tendency to fixate more towards the center of the image [58]. Lastly, the *center bias weight* accounts for the fact that a part of the empirical centerbias from the data, which we use in our models as spatial prior, is explained from image content [16] and hence already predicted by the readout network. Hence, we allow our models to downweight the center bias and make this weight dataset dependent as it depends on the image selection in the datasets. In total, this amounts to fewer than 20 dataset bias parameters with interpretable values. We refer to model variants where the bias parameters are identical across datasets as *naive*, as opposed to the full, *bias-aware* model.

Generalization and Adaptation Since our model has dataset-specific parameters, applying it to a new dataset requires specifying them. In the generalization setting, we use the average of the dataset bias parameters across all training datasets (including the center bias, where we average the fixation distributions). In the adaptation setting, we finetune the bias parameters on the new dataset.

4. Experiment Setup

Datasets We use five different datasets in our study: MIT1003 [27], CAT2000 [4], COCO Freeview [10, 72], DAEMONS [57] and FIGRIM [6]. Where included, we discard the initial forced fixation. Unlike all other datasets, FIGRIM technically is not freeviewing data since subjects are doing an image reidentification task where they have to identify repeated image presentations. However, in freeviewing experiments, subjects are commonly told that they will later have to re-identify images to keep engagement up. Hence, we assume that in FIGRIM, the eye movement data for the first presentation of each image to a subject is essentially freeviewing data. We exclude eye movement data from repeated image presentations. We noticed that the CAT2000 dataset contains an artifact, which we filtered out (see Appendix L). For more details on the used datasets including validation splits see Appendix I.

Loss function and training settings Since our model computes a 2d fixation probability distribution, we can compute average log-likelihood for ground truth fixations by taking the log-probability of the pixel for each fixation and averaging values. Log-likelihood has been shown to be a very powerful loss function for saliency models that generalizes very well to all commonly used saliency metrics [32, 35, 37]. All our models are first pretrained on the SALICON dataset [26] and then trained in the actual training setting. Overall, we have 4 main training setups: (1) we train one model individually on each dataset and evaluate it on all datasets; for each dataset we train (2) one bias-naive model on the four other datasets (leave-one-out setting) and evaluate the target dataset and (3) one bias-aware model which is evaluated with averaged or adapted dataset parameters; (4) lastly, we train the full model on all five datasets. For more details on the training see Appendix J.

Comparison Models We include two baseline models. The *centerbias model* is a KDE which, for each image, uses the fixation locations from all other images in the dataset. The centerbias quantifies how well fixations can be predicted without taking image content into account.The *gold standard model* estimates inter-observer consistency and is implemented as suggested by [32], as a regularized KDE which is crossevaluated in a leave-one-subject-out-setting. In some figures, we specify the gold standard as a range with the upper limit being the mixture of all observer's gold standard models. For more details on the baseline models, see Appendix K. For comparing to previous state-of-the-art we include DeepGaze IIE [39], UNISAL [15], SalFBNet [13] and EML-Net [25]. Each of the models is applied in the resolution which resulted in the highest performance on each the dataset.

Metrics For our internal analyses and evaluations, we use information gain [32, 35]. Information gain (IG) measures difference in log-likelihood between a candidate model and a baseline model: $IG(\hat{p}|p_{\text{baseline}}) = \frac{1}{N}\sum_{i=0}^{N}\left(\log\hat{p}(x_i \mid I_i) - \log p_{\text{baseline}}(x_i \mid I_i)\right)$, where x_i are the fixation locations and I_i denotes the image they occured on. Unlike other metrics, information gain is a ratio scale where differences and ratios of thereof are meaningful, which is needed to answer questions like "how much of the generalization gap has been closed" [32]. We report information gain relative to the center bias model, quantified in bit per fixation. For comparing to other models, we

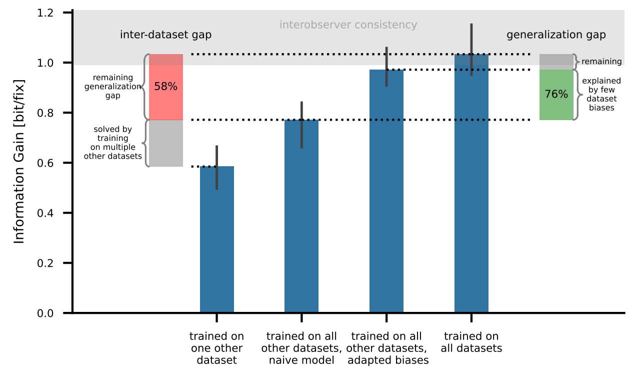

Figure 2. Performing well on unseen datasets is hard due to dataset biases: we show model performances averaged across all five datasets under different training conditions. Generalizing from one dataset to another incurs a substantial performance penalty ("inter-dataset gap"), which largely cannot be fixed by simply training on more other datasets (remaining "generalization gap"). However, accounting for a few dataset biases parameters and adapting them can mostly close the generalization gap. Performances are mean dataset performances averaged across the five datasets, error bars for paired comparisons are according to [11] and [44].

furthermore evaluate the commonly used AUC [59] metric. For probabilistic models (DeepGaze IIE, UNISAL and our models) we evaluate AUC on the log densities [37].

5. Results

Dataset bias parameter adaptation is crucial for improved performance on new datasets In Fig. 2 we show our main results: Compared to full training on all five datasets, we find that naively generalizing from one dataset to another one results in a substantial performance drop (*inter-dataset gap*) of more than 40% information gain. Opposed to what one might expect, this gap cannot be closed by simply training on more diverse data: When training naively on four other datasets, 58% of the inter-dataset-gap still remains: there is a substantial *generalization gap* to unseen datasets. By adapting just the dataset-specific parameters of our bias-aware model to the target dataset, we close 76% of the generalization gap and reach close to full performance. For all training setups we report averages over mean scores from all five datasets. In Tab. 1 we show performances per dataset and also evaluate AUC. In App. B we also compare with older saliency models, which on CAT2000 outperform some DNN based models.

Adaptation bias parameters is data efficient We test how well we can adapt the dataset bias parameters to unseen datasets when using limited data in our leave-one-dataset-out setting. To this end, we adapt the dataset parameters on small random subsets of the target dataset and evaluate on

Table 1. Performance of our model and previous state-of-the-art models. Best performance in is indicated in bold, second best is underlined. "generalization" refers to training on the respective four other datasets and evaluation with average dataset biases, "adaptation" refers to training on the respective four other datasets and evaluation after adapting the dataset bias parameters to the target dataset. Models are sorted by average AUC. See App. B for an extended version with more comparison models.

Model	MIT1003		CAT2000		COCO Freeview		DAEMONS		FIGRIM		average	
	IG	AUC	IG	AUC	IG	AUC	IG	AUC	IG	AUC	IG	AUC
EML-NET	-	0.842	-	0.766	-	0.817	-	0.766	-	0.832	-	0.805
SalFBNet	-	0.883	-	0.858	-	0.868	-	0.774	-	0.886	-	0.854
UNISAL	1.006	0.887	0.099	0.865	0.712	0.873	0.712	0.809	0.771	0.892	0.660	0.865
our model, generalization	1.172	0.902	0.249	0.878	0.889	0.886	0.538	0.800	0.883	0.905	0.746	0.874
DeepGaze IIE	1.113	0.894	0.315	0.878	0.846	0.881	1.006	0.822	0.877	0.899	0.831	0.875
our model w/o biases, generalization	1.123	0.898	0.259	0.879	0.897	0.887	0.625	0.808	0.954	0.907	0.772	0.876
our model, adaptation	1.217	0.904	0.469	0.887	0.965	0.890	1.149	0.840	1.059	0.911	0.972	0.886
our model, trained on all	**1.240**	**0.905**	0.522	0.891	1.031	0.895	1.258	0.848	**1.117**	**0.915**	1.034	0.891
our model, trained per dataset	1.217	0.903	**0.535**	**0.891**	**1.040**	**0.895**	**1.272**	**0.850**	1.105	0.914	**1.034**	**0.891**
Gold Standard (subject-LOO)	1.213	0.901	0.494	0.885	0.869	0.880	1.347	0.850	1.054	0.907	0.995	0.885
Gold Standard (upper bound)	1.829	0.945	0.873	0.920	1.511	0.935	1.722	0.899	1.642	0.947	1.515	0.929

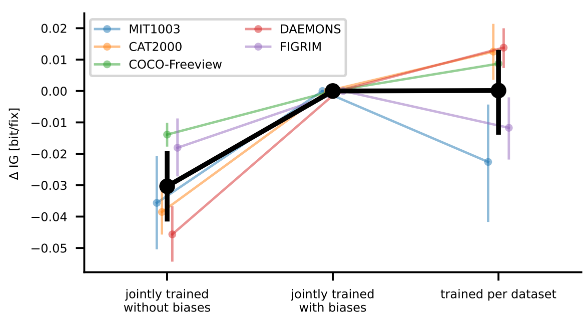

Figure 3. Performing well on multiple datasets is hard without taking dataset biases into account: We compare the performance of our full jointly trained model with dataset biases to the performance of a naively trained model and see that the naive model performs worse.

the full validation split. Due to the small sample sizes, for the model centerbias, we use a mixture of a KDE, a centered Gaussian and a uniform component (for more details see Appendix E).

As shown in Fig. 4 we outperform the generalization case with as little as 5 to 10 images for finetuning. With 50 images we already perform close to the full adaptation score. This demonstrates that the dataset bias parameters can successfully be adapted using minimal data, which is an important consideration for real world applications. Since all our dataset bias parameters are interpretable, it may even be possible in some use cases to infer them without any data, yielding performance that surpasses merely averaging the biases of all training datasets.

Dataset bias parameters are critical for achieving strong performance across multiple datasets Opposed to the leave-one-out-setting before, in Fig. 3, we compare for each dataset the performance of our jointly trained model with dataset biases with two other training setups: the model is trained jointly on all datasets but without the dataset-specific bias parameters (left) and the model trained separately for each dataset (right). We see that the joint model without dataset-specific bias parameters performs worse on all datasets compared to models trained separately. This finding supports our hypothesis that datasets differ in ways that overshadow the benefits of better learned patterns due to more diverse training data. Adding dataset specific bias parameters to the jointly trained model compensates for this problem and results in performance comparable to individually trained models. The fact that the joint bias-aware model does not yet outperform individually trained models suggests that additional dataset biases, *e.g.* semantic biases, might be at play which our model does not capture yet.

Interestingly, the performance drop of the naive jointly trained model is not as large as in the generalization case above. This suggests that the naive model implicitly tries to model some biases by detecting the dataset from the input image and using this information in the saliency decoder, by, *e.g.* computing a center bias from border artifacts in the backbones, which will work only as long as we stay within the training domains.

A new state-of-the-art for free-viewing fixation prediction We test our model on the three datasets of the MIT/Tuebingen Saliency Benchmark [33]: MIT300 (the test set for MIT1003), CAT2000 and COCO-Freeview in three different setups: generalization and adaptation of the bias-aware model trained on the respective four other training datasets, and full joint training. For all models and datasets, we set a new state-of-the-art. In particular, with full joint training increasing the main ranking metric AUC by at least 1.1%–1.5% in all datasets, with adaptation a close second, emphasizing the power of our modeling approach (Tab. 2, also Appendix A for all metrics).

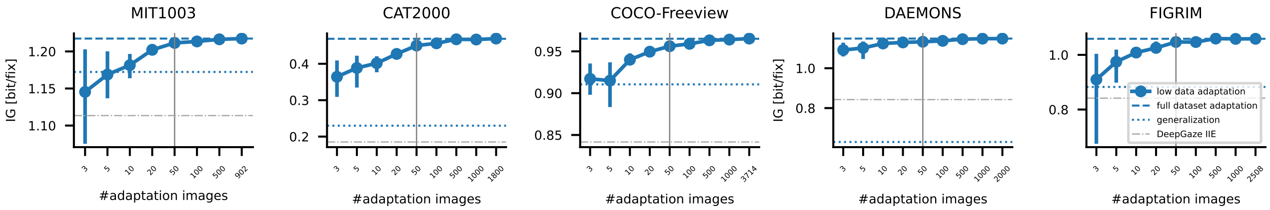

Figure 4. Low data apdatation: adaptation performance depending on the number of images used for adaptation. We outperform generalization with average dataset biases with as little as 5-10 images and reach close to full adaptation performance with around 50 images (vertical lines). Errorbars indicate variance over multiple runs with different random subsets.

Table 2. MIT/Tuebingen Saliency Benchmark: We set a new state-of-the-art on all datasets with generalization, adaptation and full training.

	MIT300			CAT2000			COCO-Freeview		
Model	IG	AUC	sAUC	IG	AUC	sAUC	IG	AUC	sAUC
TempSAL	-	0.863	0.748	-	0.844	0.638	-	0.857	0.708
DeepGaze II	0.951	0.876	0.784	0.084	0.864	0.650	0.664	0.870	0.740
EML-NET	-	0.876	0.747	-	0.831	0.585	-	0.845	0.707
SalFBNet	0.819	0.877	0.786	-	0.855	0.633	-	0.872	0.710
UNISAL	0.951	0.877	0.784	0.032	0.860	0.668	0.749	0.877	0.758
Clueify	-	0.881	0.765	-	-	-	-	-	-
DeepGaze IIE	1.071	0.883	0.794	0.189	0.869	0.668	0.860	0.882	0.767
Ours (generalized)	1.198	0.893	0.814	0.203	0.871	0.689	0.947	0.890	0.786
Ours (adapted)	1.236	**0.894**	0.815	0.433	0.881	0.690	1.011	0.893	0.788
Ours (full joint training)	**1.246**	**0.894**	**0.816**	**0.493**	**0.885**	**0.700**	**1.073**	**0.897**	**0.795**
Interobserver consistency	1.3239	0.8982	-	0.4730	0.8840	0.6930	0.8673	0.8829	-

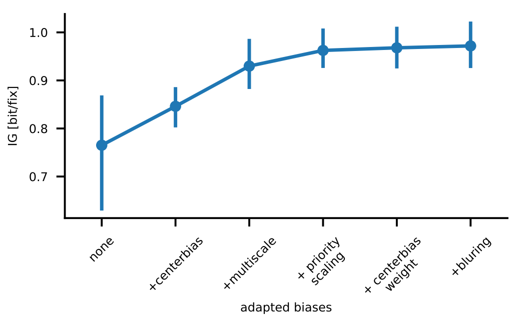

Figure 5. Contribution of different biases to closing the generalization gap. Performances are averages across the five datasets. Error bars [11, 44] are quite large because contributions differ across different datasets (Appendix, Fig. 14)

6. Analyses and Ablations

Importance of different dataset biases: In order to understand which of the different dataset biases implemented in our model are most relevant for closing the generalization gap, we conduct an ablation study in the leave-one-dataset-out setting: We start evaluating each target-dataset in the generalization setting with averaged bias parameters. We then adapt more and more bias parameters on the target dataset to see how performance increases. Results averaged across all five datasets are shown in Fig. 5. We see that centerbias and multiscale weights contribute equally and account for a large part of the performance gain. Priority scaling adds a bit more, the other effects are barely noticable in the dataset average. However, if we compare the performances separate for each dataset (Appendix Fig. 14), we see that which bias matters how much varies from dataset to dataset. MIT1003 and CAT2000 gain from adapting the centerbias weight, and DAEMONS and FIGRIM gain from adapting the blur size, showing that each bias effect is useful for some datasets.

Case studies: In Fig. 6, we analyse example predictions from cases where our model outperforms the previous state-of-the-art DeepGaze IIE most (success cases), or where it misses most performance compared to inter-observer consistency (failure cases). We find that our model excels at predicting fixations at hidden faces, objects like instruments and toy animals and gets less distracted by some patterns. This shows that our model not only quantitatively, but also qualitatively improves over DeepGaze IIE. In terms of failure cases, we find that the model underestimates the saliency of low level pattern and abstract drawings, overestimates the salience of background objects and misses fixations on occluded objects. Fixing these issues would both require better high-level understanding of scenes as well as a better understanding of the interplay between high-level

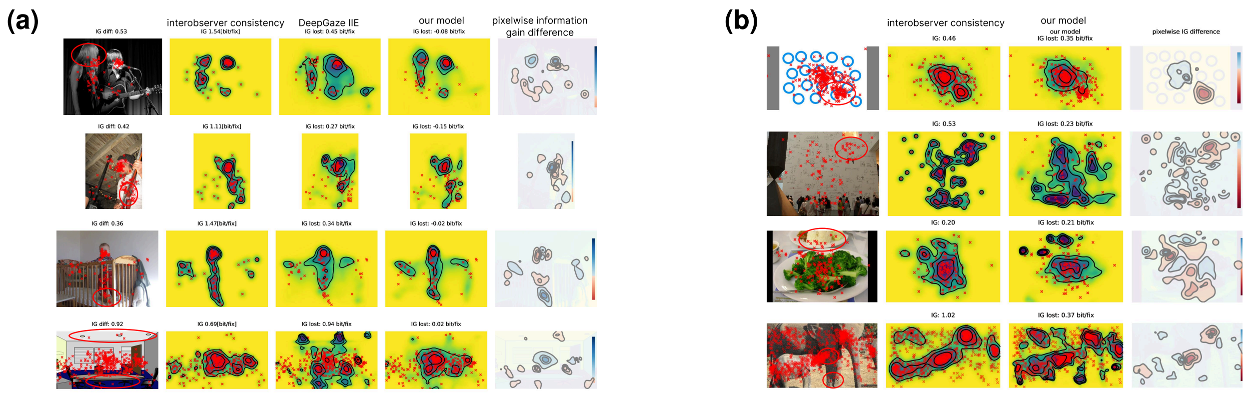

Figure 6. Successes and failure cases. (a) Images where our model improves a lot compared to DeepGaze IIE. Areas with notable structure are highlighted: 1. finds the hidden face, 2. sees the instrument, 3. sees the toy animal under foot, 4. doesn't get distracted by ceiling lamps and chairs. (b) Images where the model misses patterns: 1. misses the low-level pattern in lower right (broken circle), 2. misses the drawing in upper right, 3. overestimates the salience of dish in background, 4. misses that people look at an occluded location and overestimates legs. We use *pixelwise information gain difference* [35] to visualize where model predictions differ.

and low-level visual features and demonstrates that spatial saliency is still not solved.

Ablation relative to previous SOTA: In Fig. 7a we compare the performance of different models which step by step transform the previous SOTA model DeepGaze IIE into our new model to quantify the contribution of the different architectural elements. We see that each of the three main architectural changes (replacing the ensemble over multiple backones with a combined CLIP+DINOv2 backbone, adding a multiscale feature extraction and finally making the bias parameters dataset dependent) contributes to the increased performance of our new model.

Dependency on backbone: To make sure that our claims about inter-dataset gap, generalization gap and closing thereof with few dataset biases do not depend on our specific backbone (CLIP+DINOv2), we run the same experiments with a variety of different backbones. The backbones are choosen to include the backbones from DeepGaze I [34], DeepGaze II [36] and some of the backbones used in DeepGaze IIE [39]. The results are shown in Fig. 7b and confirm that our results hold across backbones.

Different generalization strategies: In Fig. 7c we show that generalization from a naive-model, generalization from the bias-aware model and generalization via ensembling of per-dataset trained models perform roughly on par and far worse than the bias adaptation. Which generalization strategy works best differes from dataset to dataset.

Inspection of dataset bias parameters: In Figure 8 we show how some of the dataset specific parameters differ across datasets for the model jointly trained across all five datasets. In Figure 8a we see that all datasets require both absolute and relative scales but the specific differ substantially across datasets. DAEMONS requires very high-resolution scales, FIGRIM profits from low-resolution relative scales and most other datasets are roughly in the middle. In Figure 8b we show the learned priority scaling parameters across dataset and as a function of presentation time. Again the learned values differ substantially. There is a clear dependency on presentation time, as, e.g., also suggested by Schütt et al. [54], but also variance beyond that (as visible for CAT2000 vs COCO-Freeview). We also find that the Gaussian blur differs substantially across datasets with DAEMONS requiring much smaller blur size, and that the centerbias weights differ across datasets (Appendix Fig. 18).

In Figure 8a we also show the multi scale weights learned for a dataset agnostic model (dashed line). This shows that saliency follows a complicated multiscale distribution, which requires both information about absolute size (left subplot) as well as relative size (right subplot).

Further analyses and ablations In the Appendix we include additional analyses: We conduct a sensitivity analysis for the bias parameters (Appendix C), we test generalization and adaptation on the Toronto, Kienzle and SALICON datasets (Appendix D), we explore differences between models trained across datasets or individually (Appendix G), we assess relevance of different parts of the multiscale architecture in an ablation study (Appendix F) and we compare different strategies for modeling the center bias (Appendix E)

7. Discussion

Our results demonstrate that despite the perceived saturation in performance on the MIT300 benchmark, substantial improvements in image-based saliency prediction are still achievable. While training on larger and more diverse

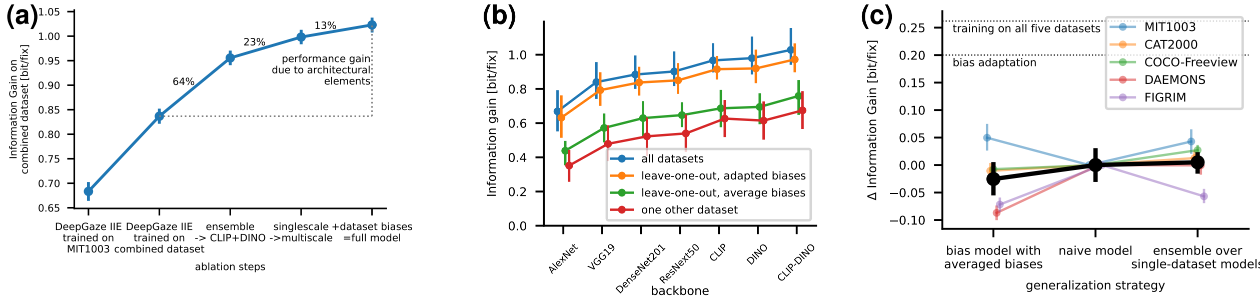

Figure 7. Ablations. (a) Performance gain on the combination of the five dataasets, compared to the previous SOTA model DeepGaze IIE: About half of the performance gain comes from training on the give datasets, the other half is due to architectural changes, comprised of the better backbone, the multiscale architecture and the dataset biases. (b) We run our main experiment with many different backbones to confirm that the results about inter-dataset gap, generalization-gap and adaptation hold across backbones. (c) Generalizing from a naive model, from a bias-aware model with averaging and via ensembling single-dataset trained models performs roughly on par.

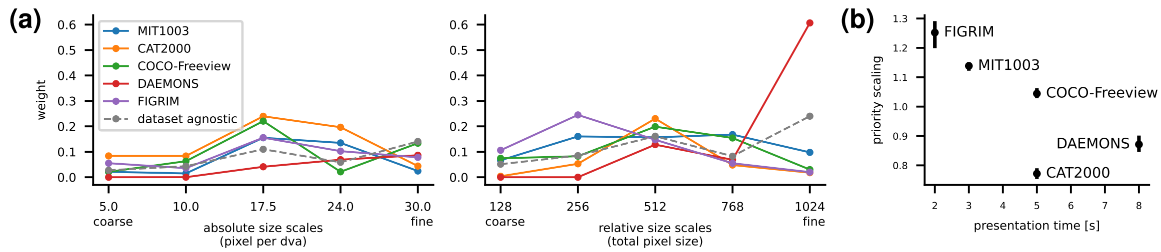

Figure 8. Dataset specific parameters. (a): multiscale weights learned per dataset. For comparison, we also include the weights of the dataset agnostic model. (b) priority scaling learned per datatset as a dependency of presentation time. The reported errors are bootstraped 95% confidence intervals of the mean across four random seeds.

datasets is crucial for achieving high performance, we find that dataset-specific biases hinder good performance on unseen datasets and result in a substantial generalization gap. A large part of this gap can be explained with few interpretable mechanisms than can be estimated very data-efficiently on new data, and which are even important to achieve high performance across many training datasets.

Since to a large degree, the bias parameters are likely to depend on the experimental conditions more than on the images themselves, explicitely modeling them is also important for learning good general saliency representations: it avoids that models need to learn shortcuts to essentially make different predictions depending on which dataset an image seems to belong to. One exception might be the multiscale structure, where we expect also an image dependency, which we hope to explore in the future.

We propose that future saliency research should focus on integrating many available datasets to develop models with robust generalization capabilities. When benchmarking models, it is preferable to evaluate them without retraining on multiple datasets. An interesting approach could be to start reporting aggregated performances over multiple datasets. If models are submitted with their code, benchmarks could evolve towards continual evaluation [19, 47], where new data is regularly added, challenging and potentially reducing the performance of existing models.

This study focused primarily on natural images to remain consistent with typical saliency evaluations. Future work should incorporate a broader range of datasets, including low-level stimuli and other out-of-distribution data. Additionally, the dataset biases considered in this study are limited, and future models could be extended to account for, e.g., varying preferences among different subject cohorts including semantic preferences [12].

An intriguing outcome of our work is that our new model outperforms estimates of inter-observer consistency on many datasets (Tab. 1, Tab. 2). Given that our model still shows some clear failure cases (Fig. 6b), this suggests that the standard methods to estimate per-image inter-observer consistency, usually as KDE [32, 69], may no longer be sufficient. Future efforts should consider new ways to establish a gold standard, such as combining high-performing deep neural networks (DNNs) with models of inter-observer consistency to account for consistent behavior not yet captured by DNNs and guide future model developments.

Acknowledgments This work was supported by the German Research Foundation (DFG): SFB 1233, Robust Vision: Inference Principles and Neural Mechanisms, project number: 276693517. HSK was supported by the Tübingen AI Center.

References

[1] Hossein Adeli, Françoise Vitu, and Gregory J. Zelinsky. A Model of the Superior Colliculus Predicts Fixation Locations during Scene Viewing and Visual Search. *The Journal of Neuroscience*, 37(6):1453–1467, 2017. 1

[2] Bahar Aydemir, Ludo Hoffstetter, Tong Zhang, Mathieu Salzmann, and Sabine Süsstrunk. TempSAL – Uncovering Temporal Information for Deep Saliency Prediction, 2024. 2

[3] Bahar Aydemir, Deblina Bhattacharjee, Tong Zhang, Mathieu Salzmann, and Sabine Süsstrunk. Data Augmentation via Latent Diffusion for Saliency Prediction. In *Computer Vision – ECCV 2024*, pages 360–377. Springer Nature Switzerland, Cham, 2025. 2

[4] Ali Borji and Laurent Itti. CAT2000: A Large Scale Fixation Dataset for Boosting Saliency Research. *CVPR 2015 workshop on "Future of Datasets"*, 2015. 1, 3

[5] Neil D. B. Bruce and John K. Tsotsos. Saliency, attention, and visual search: An information theoretic approach. *Journal of Vision*, 9(3):5–5, 2009. 2, 13

[6] Zoya Bylinskii, Phillip Isola, Constance Bainbridge, Antonio Torralba, and Aude Oliva. Intrinsic and extrinsic effects on image memorability. *Vision Research*, 116:165–178, 2015. 3

[7] Kai-Yueh Chang, Tyng-Luh Liu, Hwann-Tzong Chen, and Shang-Hong Lai. Fusing generic objectness and visual saliency for salient object detection. In *2011 International Conference on Computer Vision*, pages 914–921, 2011. 2

[8] Shi Chen, Ming Jiang, and Qi Zhao. What Do Deep Saliency Models Learn about Visual Attention? *Advances in Neural Information Processing Systems*, 36: 9543–9555, 2023. 2

[9] Yupei Chen, Zhibo Yang, Seoyoung Ahn, Dimitris Samaras, Minh Hoai, and Gregory Zelinsky. COCO-Search18 fixation dataset for predicting goal-directed attention control. *Scientific Reports*, 11(1):8776, 2021. 1

[10] Yupei Chen, Zhibo Yang, Souradeep Chakraborty, Sounak Mondal, Seoyoung Ahn, Dimitris Samaras, Minh Hoai, and Gregory Zelinsky. Characterizing Target-Absent Human Attention. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5031–5040, 2022. 1, 3

[11] Denis Cousineau. Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, 1(1):42–45, 2005. 4, 6

[12] Benjamin de Haas, Alexios L. Iakovidis, D. Samuel Schwarzkopf, and Karl R. Gegenfurtner. Individual differences in visual salience vary along semantic dimensions. *Proceedings of the National Academy of Sciences*, page 201820553, 2019. 8

[13] Guanqun Ding, Nevrez İmamoğlu, Ali Caglayan, Masahiro Murakawa, and Ryosuke Nakamura. SalFBNet: Learning pseudo-saliency distribution via feedback convolutional networks. *Image and Vision Computing*, 120:104395, 2022. 2, 4

[14] Yasser Abdelaziz Dahou Djilali, Kevin McGuiness, and Noel O'Connor. Learning Saliency From Fixations, 2023. 2

[15] Richard Droste, Jianbo Jiao, and J. Alison Noble. Unified Image and Video Saliency Modeling. In *Computer Vision – ECCV 2020*, pages 419–435, Cham, 2020. Springer International Publishing. 1, 2, 4, 13

[16] Wolfgang Einhäuser, Merrielle Spain, and Pietro Perona. Objects predict fixations better than early saliency. *Journal of Vision*, 8(14):18–18, 2008. 3

[17] Ralf Engbert, Hans A. Trukenbrod, Simon Barthelmé, and Felix A. Wichmann. Spatial statistics and attentional dynamics in scene viewing. *Journal of Vision*, 15(1):14–14, 2015. 1

[18] Erkut Erdem and Aykut Erdem. Visual saliency estimation by nonlinearly integrating features using region covariances. *Journal of Vision*, 13(4):11–11, 2013. 13

[19] Adhiraj Ghosh, Sebastian Dziadzio, Ameya Prabhu, Vishaal Udandarao, Samuel Albanie, and Matthias Bethge. ONEBench to Test Them All: Sample-Level Benchmarking Over Open-Ended Capabilities, 2024. 8

[20] Jonathan Harel, Christof Koch, and Pietro Perona. Graph-Based Visual Saliency. In *Advances in Neural Information Processing Systems 19*, pages 545–552. MIT Press, 2007. 2, 13

[21] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825): 357–362, 2020. 19

[22] Xun Huang, Chengyao Shen, Xavier Boix, and Qi Zhao. SALICON: Reducing the Semantic Gap in Saliency Prediction by Adapting Deep Neural Networks. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 262–270, Santiago, Chile, 2015. IEEE. 2

[23] John D. Hunter. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, 9(3):90–95, 2007. 19

[24] Laurent Itti, Christof Koch, and Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 20(11):1254–1259, 1998. 2, 13

[25] Sen Jia and Neil D. B. Bruce. EML-NET: An Expandable Multi-Layer NETwork for saliency prediction. *Image and Vision Computing*, 95:103887, 2020. 2, 4

[26] Ming Jiang, Shengsheng Huang, Juanyong Duan, and Qi Zhao. SALICON: Saliency in Context. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1072–1080, 2015. 4, 18

[27] Tilke Judd, Krista Ehinger, Frédo Durand, and Antonio Torralba. Learning to predict where humans look. In *Computer Vision, 2009 IEEE 12th International Conference On*, pages 2106–2113. IEEE, 2009. 1, 2, 3

[28] Amita Kamath, Jack Hessel, and Kai-Wei Chang. What's "up" with vision-language models? Investigating their struggle with spatial reasoning, 2023. 3

[29] Wolf Kienzle, Matthias O. Franz, Bernhard Schölkopf, and Felix A. Wichmann. Center-surround patterns emerge as optimal predictors for human saccade targets. *Journal of Vision*, 9(5):7–7, 2009. 2, 13

[30] Alexander Kroner, Mario Senden, Kurt Driessens, and Rainer Goebel. Contextual encoder–decoder network for visual saliency prediction. *Neural Networks*, 129: 261–270, 2020. 2

[31] S. S. S. Kruthiventi, K. Ayush, and R. V. Babu. DeepFix: A Fully Convolutional Neural Network for Predicting Human Eye Fixations. *IEEE Transactions on Image Processing*, 26(9):4446–4456, 2017. 2

[32] Matthias Kümmerer and Matthias Bethge. Predicting Visual Fixations. *Annual Review of Vision Science*, 9 (1):269–291, 2023. 1, 4, 8, 18

[33] Matthias Kümmerer, Zoya Bylinskii, Tilke Judd, Ali Borji, Laurent Itti, Frédo Durant, Aude Oliva, Antonio Torralba, and Matthias Bethge. MIT/Tuebingen Saliency Benchmark. saliency.tuebingen.ai. 1, 5

[34] Matthias Kümmerer, Lucas Theis, and Matthias Bethge. Deep Gaze I: Boosting Saliency Prediction with Feature Maps Trained on ImageNet. *ICLR Workshop Track*, 2015. 2, 7

[35] Matthias Kümmerer, Thomas S. A. Wallis, and Matthias Bethge. Information-theoretic model comparison unifies saliency metrics. *Proceedings of the National Academy of Sciences*, 112(52):16054–16059, 2015. 4, 7, 16

[36] Matthias Kümmerer, Thomas S. A. Wallis, Leon A. Gatys, and Matthias Bethge. Understanding Low- and High-Level Contributions to Fixation Prediction. In *The IEEE International Conference on Computer Vision (ICCV)*, pages 4789–4798, 2017. 2, 3, 7

[37] Matthias Kümmerer, Thomas S. A. Wallis, and Matthias Bethge. Saliency Benchmarking Made Easy: Separating Models, Maps and Metrics. In *Computer Vision – ECCV 2018*, pages 798–814. Springer International Publishing, 2018. 4

[38] Peizhao Li, Junfeng He, Gang Li, Rachit Bhargava, Shaolei Shen, Nachiappan Valliappan, Youwei Liang, Hongxiang Gu, Venky Ramachandran, Golnaz Farhadi, Yang Li, Kai J. Kohlhoff, and Vidhya Navalpakkam. UniAR: A Unified model for predicting human Attention and Responses on visual content. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. 2

[39] Akis Linardos, Matthias Kümmerer, Ori Press, and Matthias Bethge. DeepGaze IIE: Calibrated Prediction in and Out-of-Domain for State-of-the-Art Saliency Modeling. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12919–12928, 2021. 2, 3, 4, 7

[40] Panagiotis Linardos, Eva Mohedano, Juan Jose Nieto, Noel E. O'Connor, Xavier Giro-i-Nieto, and Kevin McGuinness. Simple vs complex temporal recurrences for video saliency prediction, 2019. 1

[41] Fangyu Liu, Guy Emerson, and Nigel Collier. Visual Spatial Reasoning, 2023. 3

[42] Zhuang Liu and Kaiming He. A Decade's Battle on Dataset Bias: Are We There Yet?, 2024. 3

[43] Jianxun Lou, Hanhe Lin, David Marshall, Dietmar Saupe, and Hantao Liu. TranSalNet: Towards perceptually relevant visual saliency prediction. *Neurocomputing*, 494:455–467, 2022. 2

[44] Richard D. Morey. Confidence Intervals from Normalized Data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, 4(2): 61–64, 2008. 4, 6

[45] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy V. Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Mido Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Herve Jegou,

Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. DINOv2: Learning Robust Visual Features without Supervision. *Transactions on Machine Learning Research*, 2023. 3

[46] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An Imperative Style, High-Performance Deep Learning Library, 2019. 19

[47] Ameya Prabhu, Vishaal Udandarao, Philip Torr, Matthias Bethge, Adel Bibi, and Samuel Albanie. Lifelong Benchmarks: Efficient Model Evaluation in an Era of Rapid Progress, 2024. 8

[48] Michael J. Proulx and Monique Green. Does apparent size capture attention in visual search? Evidence from the Müller–Lyer illusion. *Journal of Vision*, 11(13): 21, 2011. 3

[49] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning Transferable Visual Models From Natural Language Supervision. *arXiv:2103.00020 [cs]*, 2021. 3

[50] Jeff Reback, jbrockmendel, Wes McKinney, Joris Van den Bossche, Tom Augspurger, Phillip Cloud, Simon Hawkins, Matthew Roeschke, gfyoung, Sinhrks, Adam Klein, Patrick Hoefler, Terji Petersen, Jeff Tratner, Chang She, William Ayd, Shahar Naveh, J. H. M. Darbyshire, Marc Garcia, Richard Shadrach, Jeremy Schendel, Andy Hayden, Daniel Saxton, Marco Edward Gorelli, Fangchen Li, Matthew Zeitlin, Vytautas Jancauskas, Ali McMaster, Pietro Battiston, and Skipper Seabold. Pandas-dev/pandas: Pandas 1.4.1. Zenodo, 2022. 19

[51] Nicolas Riche, Matei Mancas, Matthieu Duvinage, Makiese Mibulumukini, Bernard Gosselin, and Thierry Dutoit. RARE2012: A multi-scale rarity-based saliency detection with its comparative statistical analysis. *Signal Processing: Image Communication*, 28(6):642–658, 2013. 2, 13

[52] Constantin A. Rothkopf, Dana H. Ballard, and Mary M. Hayhoe. Task and context determine where you look. *Journal of Vision*, 7(14):16–16, 2007. 1

[53] Steffen Schneider, Evgenia Rusak, Luisa Eck, Oliver Bringmann, Wieland Brendel, and Matthias Bethge. Improving robustness against common corruptions by covariate shift adaptation. In *Advances in Neural Information Processing Systems*, pages 11539–11551. Curran Associates, Inc., 2020. 3

[54] Heiko H. Schütt, Lars O. M. Rothkegel, Hans A. Trukenbrod, Ralf Engbert, and Felix A. Wichmann. Disentangling bottom-up versus top-down and low-level versus high-level influences on eye movements over time. *Journal of Vision*, 19(3):1–1, 2019. 7

[55] Lisa Schwetlick, Lars Oliver Martin Rothkegel, Hans Arne Trukenbrod, and Ralf Engbert. Modeling the effects of perisaccadic attention on gaze statistics during scene viewing. *Communications Biology*, 3(1): 1–11, 2020. 1

[56] Lisa Schwetlick, Daniel Backhaus, and Ralf Engbert. A dynamical scan-path model for task-dependence during scene viewing. *Psychological Review*, 130(3): 807–840, 2023. 1

[57] Lisa Schwetlick, Matthias Kümmerer, Matthias Bethge, and Ralf Engbert. Potsdam data set of eye movement on natural scenes (DAEMONS). *Frontiers in Psychology*, 15, 2024. 3

[58] Benjamin W. Tatler. The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, 7(14):4–4, 2007. 3

[59] Benjamin W. Tatler, Roland J. Baddeley, and Iain D. Gilchrist. Visual correlates of fixation selection: Effects of scale and time. *Vision Research*, 45(5):643–659, 2005. 4

[60] Chakkrit Termritthikun, Ayaz Umer, Suwichaya Suwanwimolkul, Feng Xia, and Ivan Lee. SalNAS: Efficient Saliency-prediction Neural Architecture Search with self-knowledge distillation. *Engineering Applications of Artificial Intelligence*, 136: 109030, 2024. 2

[61] Chakkrit Termritthikun, Ayaz Umer, Suwichaya Suwanwimolkul, and Ivan Lee. Semi-PKD: Semi-supervised Pseudoknowledge Distillation for saliency prediction. *ICT Express*, 11(2):364–370, 2025. 2

[62] Antonio Torralba and Alexei A. Efros. Unbiased look at dataset bias. In *CVPR 2011*, pages 1521–1528, 2011. 3

[63] Antonio Torralba, Aude Oliva, Monica S. Castelhano, and John M. Henderson. Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, 113(4):766–786, 2006. 2

[64] Anne M Treisman and Garry Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12(1):97–136, 1980. 2

[65] Eleonora Vig, Michael Dorr, and David Cox. Large-Scale Optimization of Hierarchical Features for Saliency Prediction in Natural Images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2798–2805, 2014. 2

[66] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C. J. Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, and Paul van Mulbregt. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, pages 1–12, 2020. 19

[67] Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully Test-Time Adaptation by Entropy Minimization. In *International Conference on Learning Representations*, 2020. 3

[68] Michael Waskom. Seaborn: Statistical data visualization. *Journal of Open Source Software*, 6(60):3021, 2021. 19

[69] Niklas Wilming, Torsten Betz, Tim C. Kietzmann, and Peter König. Measures and Limits of Models of Fixation Selection. *PLOS ONE*, 6(9):e24038, 2011. 8

[70] Calden Wloka, Iuliia Kotseruba, and John K. Tsotsos. Active Fixation Control to Predict Saccade Sequences. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3184–3193, Salt Lake City, UT, USA, 2018. IEEE. 1

[71] Jiawei Xie, Zhi Liu, Gongyang Li, Xiaofeng Lu, and Tao Chen. Global semantic-guided network for saliency prediction. *Knowledge-Based Systems*, 284: 111279, 2024. 2

[72] Zhibo Yang, Sounak Mondal, Seoyoung Ahn, Gregory Zelinsky, Minh Hoai, and Dimitris Samaras. Predicting Human Attention using Computational Attention, 2023. 2, 3

[73] Alfred L. Yarbus. *Eye Movements and Vision*. Plenum Press, 1967. 2

[74] Mert Yuksekgonul, Federico Bianchi, Pratyusha Kalluri, Dan Jurafsky, and James Zou. When and Why Vision-Language Models Behave like Bags-Of-Words, and What to Do About It? In *The Eleventh International Conference on Learning Representations*, 2022. 3

[75] Lingyun Zhang, Matthew H. Tong, Tim K. Marks, Honghao Shan, and Garrison W. Cottrell. SUN: A Bayesian framework for saliency using natural statistics. *Journal of Vision*, 8(7):32–32, 2008. 2