This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Source-free Adaptive Gaze Estimation by Uncertainty Reduction

Xin Cai[1,2], Jiabei Zeng[1], Shiguang Shan[1,2,3], Xilin Chen [1,2]
[1]Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100090, China
[2]University of Chinese Academy of Sciences, Beijing, 100090, China
[3]Peng Cheng Laboratory, Shenzhen, 518055, China
{caixin20s,jiabei.zeng,sgshan,xlchen}@ict.ac.cn *

Abstract

Gaze estimation across domains has been explored recently because the training data are usually collected under controlled conditions while the trained gaze estimators are used in nature and diverse environments. However, due to privacy and efficiency concerns, simultaneous access to annotated source data and to-be-predicted target data can be challenging. In light of this, we present an unsupervised source-free domain adaptation approach for gaze estimation, which adapts a source-trained gaze estimator to unlabeled target domains without source data. We propose the Uncertainty Reduction Gaze Adaptation (UnReGA) framework, which achieves adaptation by reducing both sample and model uncertainty. Sample uncertainty is mitigated by enhancing image quality and making them gaze-estimation-friendly, whereas model uncertainty is reduced by minimizing prediction variance on the same inputs. Extensive experiments are conducted on six cross-domain tasks, demonstrating the effectiveness of UnReGA and its components. Results show that UnReGA outperforms other state-of-the-art cross-domain gaze estimation methods under both protocols, with and without source data. The code is available at https://github.com/caixin1998/UnReGA.

1. Introduction

Gaze encodes rich information about the attention and psychological factors of an individual. Techniques that use eye tracking to infer human intentions and understand human emotions have found an increasingly wide utilization in fields including human-computer interaction [20,35,36], affective computing [11], and medical diagnosis [21,46]. The most prevalent way to estimate human gaze is using commercial eye trackers, which suffer from high cost or custom invasive hardware. To overcome the limitation on devices and environments, researchers have made great progress on appearance-based gaze estimation methods with the development of deep learning [4,6,12,56,57].

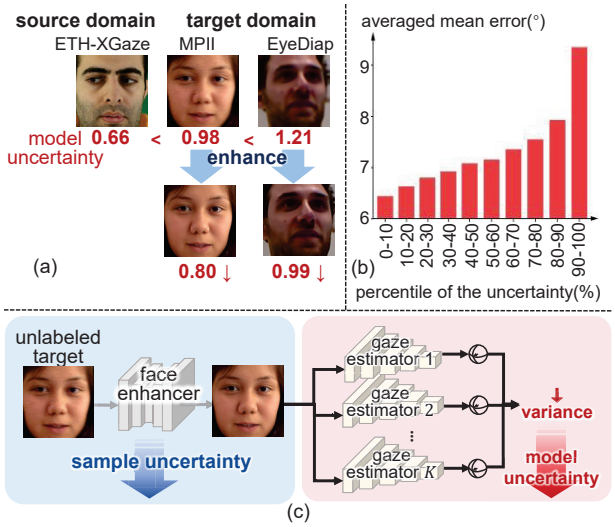

Figure 1. (a) The source-trained model shows high uncertainty on samples from different domains. (b) Statistics of errors and model uncertainty by the same gaze estimator on different samples. The error increases as the uncertainty grow. (c) To accomplish unsupervised source-free domain adaptation, the UnReGA reduces the sample uncertainty by enhancing the input images and reduces the model uncertainty by minimizing the prediction variance.

Notwithstanding the achievements, the appearance-based gaze estimators meet the most challenging problem that their performance drops significantly when they are trained and tested on different domains, e.g., the domains with different subjects, image quality, background environments, or illuminations. Usually, gaze estimators are trained on the data collected under controlled conditions where true gaze is feasible to be measured and recorded by the deployed devices. Then, these gaze estimators would be applied under a much different and uncontrolled environment.

To adapt the source-data-trained model to the target data, researchers propose methods to narrow the gap between the different domains [16,34,42,45]. Most of the methods re-

*This work is partially supported by National Key RD Program of China (No. 2018AAA0102405), National Natural Science Foundation of China (No. 62176248).

quire data from both the source and target domains during the adaptation. However, in the application of gaze estimation, the source data is likely to be neither available nor efficient during the adaptation. First, most gaze models are trained with face images which might be not accessible due to privacy or bandwidth issues. Secondly, processing source data might not be computationally practical in real-time gaze estimation on the target domain. Therefore, we formulate gaze estimation as an unsupervised source-free domain adaptation problem, where we cannot access the source data when fitting the model to the target.

To address the source-free domain adaptation issue, we propose to adapt the source-trained gaze estimators to the target domain by reducing both the *sample uncertainty* and *model uncertainty* on the unlabeled target data. *Sample uncertainty* captures noise inherent in the input images, such as sensor noise and motion blur, which is also referred to as aleatoric uncertainty [24]. *Model uncertainty* is determined by the inconsistency of predication or model perturbations, which is also referred to as epistemic uncertainty [15, 24]. We formulate it as the variance of different estimators' predictions on the same sample. We assume that reducing the two uncertainties helps to reduce the gaze estimator's errors across different domains due to three observations: 1) Estimators show high model uncertainty on samples that are distributed far away from the training data and show low uncertainty on the nearby samples [24, 28]. As shown in Fig. 1(a), the ETH-XGaze-trained estimator has average model uncertainties of 0.66, 0.98, and 1.21 on the samples from ETH-XGaze [53], MPIIGaze [57], EyeDiap [14], respectively. EyeDiap has the most different distribution from ETH-XGaze and shows the highest model uncertainty. 2) Reducing the sample uncertainty pulls together the source and target data, and accordingly reduces the estimator's model uncertainty on target data. In Fig. 1(a), the model uncertainties on MPIIGaze/EyeDiap decrease when we reduce the sample uncertainty by image enhancement, because by doing this, we reduce the image quality discrepancy between MPIIGaze/EyeDiap and ETH-XGaze. 3) Model uncertainty empirically shows a positive correlation with gaze estimation error in cross-domain scenarios. Fig. 1(b) plots how the errors change with model uncertainty. We train 10 gaze estimators from ETH-XGaze and then, for each sample in MPIIGaze, we compute the model uncertainty and the mean error of the estimators' predictions. We sort the samples by the model uncertainty in ascending order and group them by every 10-th percentile. The height of each bar in Fig. 1(b) denotes the averaged mean error over the samples within each group. As can be seen, the top 10 percent of the model uncertainty corresponds to the smallest error.

To this end, we propose an Uncertainty Reduction Gaze Adaption (UnReGA) framework that accomplishes the source-free adaptation by minimizing both the sample and model uncertainty. As illustrated in Fig. 1(c), we first transfer the input images into a gaze-estimation-friendly domain by introducing a face enhancer to enhance input images without changing the gaze. Rather than low-quality images, high-quality images convey more details about the eyes and contribute to less sample uncertainty and better generalization ability of the source-trained gaze estimators. Next, we update an ensemble of source gaze estimators by minimizing the variance of their predictions on the unlabeled target data. Finally, we merge the updated estimators into a single model during inference. Our empirical experiments demonstrate that the updated estimator outperforms the not-adapted source estimator on the target domain.

Our contributions are summarized as:

1. We formulate gaze estimation as an unsupervised source-free domain adaptation problem and propose an **Un**certainty **Re**duction **G**aze **A**daption (UnReGA) framework that adapts the trained model to target domain without the source data by reducing both the *sample uncertainty* and *model uncertainty*.
2. We propose the variance minimization and pseudo-label supervision mechanisms in UnReGA to address the adaptation issue without source data for regression tasks, while most existing source-free adaptation methods are designed for classification tasks. We validate the effectiveness of the two mechanisms in source-free adaptive gaze estimation.
3. We evaluate the efficacy of UnReGA and its components on cross-domain gaze estimation tasks. Extensive experiments show UnReGA outperforms other state-of-the-art cross-domain gaze estimation methods under both protocols, with and without source data.

2. Related Work

Cross-domain Gaze Estimation: With the development of deep learning, many efforts are made in appearance-based gaze estimation [1, 7, 9, 12, 37, 38, 55] to reduce prediction errors on public gaze datasets [12, 14, 23, 40], e.g., MPIIGaze [57], ETH-XGaze [53] and GazeCapture [26]. However, training data for these estimators are often collected under controlled conditions, limiting their applicability in diverse real-world scenarios. Therefore, recent studies have explored gaze estimation methods across domains.

According to the availability of the source and unlabeled target data, we review the cross-domain gaze estimation methods under three settings: domain generalization [44], unsupervised domain adaptation with source data [49], unsupervised domain adaptation without source data [32].

For domain generalization, the target domain is unknown so we do not adapt the gaze estimator to a specific domain but improve its generalization ability across different domains during the training. Park *et al.* [37] proposed to learn

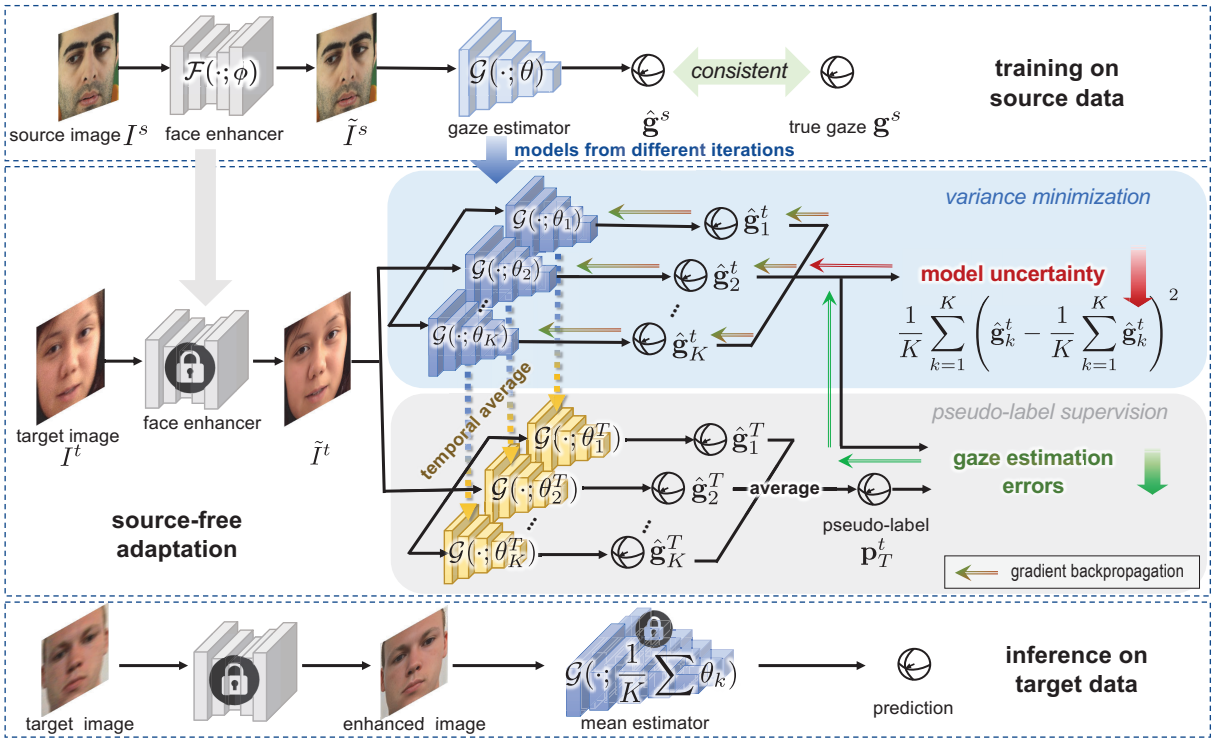

Figure 2. Illustration of the UnReGA framework with three stages. In the training stage on source data (top), we train the face enhancer and the gaze estimator using source data. In the source-free adaptation stage(middle), we update the set of gaze estimators from different training iterations by two mechanisms (variance minimization and pseudo-label supervision) to reduce the model uncertainty and preserve the gaze estimation ability. In the inference stage on target data (bottom), we predict the gaze by the mean estimator.

a rotation-aware latent representation of gaze and Cheng *et al.* [5] proposed to extract domain-agnostic gaze feature to improve the methods' generalization capabilities.

For domain adaptation with source data, existing methods utilize labeled source data and unlabeled target data. These methods simultaneously minimize prediction errors on the source data while adapting the model to the target domain using various techniques, e.g., adversarial learning [45], outlier guidance [34], and contrastive regression [3].

For domain adaptation without source data, optimizing gaze estimators' performance on both source and target domains simultaneously is impractical. Although the strategies in [3, 34] are feasible to adapt the model to the target domain, their performance drops when the supervision from the source domain is absent. Because without the supervision of the true gaze, the models lose their gaze estimation ability. To address this, Bao *et al.* [2] proposed a self-training strategy by keeping rotation consistency on augmented target images for adaptation without source data.

Source-free Domain Adaptation: The domain adaptation problem without source data is also explored in other computer vision tasks, e.g., image classification [29,32], semantic segmentation [13, 27] and object detection [30, 31]. To solve this problem, existing works leverage the knowledge hidden in the source model by pseudo-labeling [13, 32], feature alignment [10, 50, 51], self-supervised learning [3, 18, 33, 41], batch normalization adaptation [39] et al. Most of the methods are designed for classification problems but might fail in regression. Our proposed method addresses the issue in gaze estimation, which is a regression problem. We are inspired by the work [24], which computes uncertainty with an ensemble of models to measure the domain shift, to reduce cross-domain gaze errors by reducing uncertainty. Similarly, regarding entropy as a measure of uncertainty, the source-free adaptation method using Entropy Minimization [13, 32, 43, 58] accomplishes the adaptation by reducing uncertainty in classification tasks.

3. Uncertainty Reduction Gaze Adaptation

We present the **Un**certainty **Re**duction **G**aze **A**daptation (UnReGA) framework to solve the unsupervised source-free domain adaptation problem for gaze estimation.

3.1. Problem Definition

Let $\mathcal{D}_s = \{(I_i^s, \mathbf{g}i^s)\}_{i=1}^{N_s}$ be the source domain data, where I_i^s and $\mathbf{g}_i^s$ represent the i-th image and its true gaze label, respectively. The source domain consists of N_s samples, which are typically obtained under controlled condi-

tions where ground truth labels are available. Let $\mathcal{D}_t = \{I_i^t\}_{i=1}^{N_t}$ denote the target domain images captured under different conditions in real-world scenarios. The goal of unsupervised source-free domain adaptation is to estimate the gaze of the target images when we cannot access to the source and target data simultaneously. Thus, we train the source models on the source data without knowledge of the target data and then adapt these models to the unlabelled target data in absence of the source data.

3.2. UnReGA Framework

To solve the unsupervised source-free domain adaptation in gaze estimation, we propose an **Un**certainty **Re**duction **G**aze **A**daptation (UnReGA) framework, which makes the pre-trained gaze estimators suitable for the target data by reducing their uncertainties on the target. Fig.2 illustrates the UnReGA framework, which comprises three stages: source model training, source-free adaptation and inference on target data. In the training on source data, we train the face enhancer and the gaze estimator with the enhanced images as input. The face enhancer reduces the sample uncertainty by improving the input images' quality and makes them more suitable for gaze estimation across domains. We keep a set of trained gaze estimators at different iterations during the training process for the next adaptation stage. In source-free adaptation, the set of gaze estimators is updated by the variance minimization mechanism and pseudo-label mechanism. The two mechanisms reduce the model uncertainty on target data and preserve the models' ability in accurate gaze estimation. In inference, by taking the mean parameters of the updated estimators, the set of models is merged into a single one, which is used to predict the gaze for target images. Below, we present details of the three stages.

3.3. Training on Source Data

We collaboratively train a gaze-estimation-friendly keeping-gaze face enhancer and gaze estimator during the training stage. The collaboration improves the generalization ability of the gaze estimator on different domains, although they are trained on the source data **without** knowledge of the target domain. Fig. 3 shows how we train the face enhancer and gaze estimator. We first pretrain the gaze estimator and the face enhancer respectively, and then finetune the face enhancer and the gaze estimator sequentially.

Pretrain the gaze estimator: We employ a ResNet18 [17] as the gaze estimator $\mathcal{G}(\cdot;\theta)$. It is pretrained on the annotated source data $\mathcal{D}_s = \{(I_i^s, \mathbf{g}_i^s)\}_{i=1}^{N_s}$ by minimizing the discrepancy between the prediction and the true gaze:

$$\min_{\theta} \mathbb{E}_{(I_i^s,\mathbf{g}_i^s)\in\mathcal{D}_s} \|\mathcal{G}(I_i^s;\theta), \mathbf{g}_i^s\|_1, \tag{1}$$

where $\mathcal{G}(I_i^s;\theta)$ and $\mathbf{g}_i^s$ is the prediction and the true label of the original source image I_i^s, respectively.

Pretrain the face enchancer: We employ a general image super-resolution model Real-ESRGAN [47] as the face enhancer $\mathcal{F}(\cdot;\phi)$ by removing the last up-sample module of Real-ESRGAN and ensuring the input and output to be of the same resolution. The green part in Fig. 3 illustrates how we pretrain $\mathcal{F}(\cdot;\phi)$ on a high-quality image dataset. Given the high quality image I^h, we degrade its quality through degradation methods in [47]. Then, we fed the degraded image I^l into the face enhancer and obtain the enhanced on $\tilde{I}^l$. Similar to [47], to train the face enhancer, we force the generated image $\tilde{I}^l$ and the original I^h to be consistent and un-distinguishable by minimizing a reconstruction loss and adopting an adversarial mechanism.

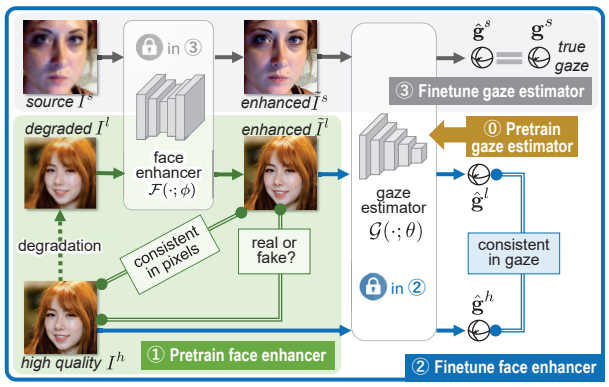

Figure 3. Training strategy on source data. First, we pretrain a face enhancer with a set of high-quality images(in green part). Second, we finetune the face enhancer by adding the gaze consistent constraints for both the labeled source images and unlabelled images (in the blue rectangle). Third, we finetune the gaze estimator with an enhanced source (in the grey part).

Finetune the face enchancer: Although recovering details of the face, the pretrained face enhancer is likely to change the gaze by changing the eyes' appearance. To keep the gaze unchanged, we finetune the face enhancer by forcing the enhanced images to have the same gaze as the label or as the one from the original image. The blue rectangle in Fig. 3 shows how we finetune the face enhancer. For the image I^s in the source domain, we enhance it into $\tilde{I}^s$ by the face enhancer and then predict its gaze using the gaze estimator $\mathcal{G}(\cdot;\theta)$. We require the predictions $\mathcal{G}(\tilde{I}^s;\theta)$ to be consistent with the true gaze by minimizing the gaze estimation loss ℓ_g over all the source data:

$$\min_{\phi} \ell_g = \min_{\phi} \mathbb{E}_{(I^s,\mathbf{g}^s)\in\mathcal{D}_s} \|\mathcal{G}(\tilde{I}^s;\theta), \mathbf{g}^s\|_1. \tag{2}$$

For high-quality images, we optimize the reconstruction loss and adversarial loss as we do in the pretraining stage. Additionally, we force the original high quality I^h and the enhanced low quality $\tilde{I}^l$ to have the same predictions by minimizing the gaze consistent loss ℓ_{gc}:

$$\min_{\phi} \ell_{gc} = \min_{\phi} \mathbb{E}_{I^h\in\mathcal{D}_h} \|\mathcal{G}(\tilde{I}^l;\theta), \mathcal{G}(I^h;\theta)\|_1, \tag{3}$$

where $\mathcal{D}_h$ is the set of high quality images. $\mathcal{G}(\tilde{I}^l;\theta)$ and $\mathcal{G}(I^h)$ are gaze predictions of $\tilde{I}^l$ and I^h, respectively.

In summary, we finetune the parameters ϕ of the face enhancer by freezing the gaze estimator and minimizing the sum of ℓ_g, ℓ_{gc}, and the losses ℓ_{pre} used in the pretraining the face enhancer [47] as: $\min_\phi\ \ell_g + \ell_{gc} + \ell_{pre}$.

Finetune the gaze estimator: The grey part in Fig. 3 shows the finetuning procedure. Freezing the parameters of the face enhancer, we update the parameter θ of the gaze estimator to boost its performance on the enhanced images. The objective is:

$$\min_{\theta} \mathbb{E}_{(I^s,\mathbf{g}^s)\in\mathcal{D}_s} \|\mathcal{G}(\tilde{I}^s;\theta), \mathbf{g}^s\|_1, \tag{4}$$

where $\tilde{I}^s$ is the enhanced source image.

After training on source data, we obtain a keeping-gaze face enhancer and a gaze estimator that predicts gaze for the enhanced images. It is noted that we keep a set of gaze estimators from different iterations during the finetuning of the gaze estimator, which will be used in the next stage of source-free adaptation.

3.4. Source-free Adaptation

During the adaptation stage, we only have access to the unlabelled target data and the source model without the source data. To adapt the trained gaze estimators from different iterations to the unlabelled target data, we unsupervised update the estimators' parameters by minimizing the model uncertainty and preserving the models' ability in gaze estimation via pseudo-labels. The process of unsupervised source-free adaptation is depicted in the middle row of Fig. 2. Specifically, UnReGA enhances the quality of the target image I^t using the trained FaceEnhancer, yielding the enhanced image $\tilde{I}^t$. Then, $\tilde{I}^t$ is fed into two branches: variance minimization and pseudo-label supervision.

Variance Minimization: In the upper branch(variance minimization), UnReGA forces the set of source estimators to have low model uncertainty on the enhanced target image $\tilde{I}^t$. As discussed in the introduction, minimizing the model uncertainty helps to reduce the estimation errors in target data. Inspired by [28], we formulate model uncertainty as the variance of the predictions by the set of models on the same input image. Let $\{\mathcal{G}(\cdot;\theta_k)\}_{k=1}^K$ denote the set of trained estimators, where $\mathcal{G}(\cdot;\theta_k)$ is the k-th model with learned parameters θ_k and K is the number of models. In this work, the K models are saved checkpoints from K different training iterations when we finetune the gaze estimator on the source. They have the same architecture but different parameter values. We update the parameters $\{\theta_k\}_{k=1}^K$ by minimizing the model uncertainty over target data as:

$$\min_{\{\theta_k\}_{k=1}^K} \ell_{vm} = \min_{\{\theta_k\}_{k=1}^K} \frac{1}{K}\sum_{k=1}^{K}\left(\hat{\mathbf{g}}_k^t - \frac{1}{K}\sum_{k=1}^{K}\hat{\mathbf{g}}_k^t\right)^2, \tag{5}$$

where $\hat{\mathbf{g}}_k^t = \mathcal{G}(\tilde{I}^t;\theta_k)$ denotes the prediction of the enhanced target image $\tilde{I}^t$ by the k-th model $\mathcal{G}(\cdot;\theta_k)$. $\frac{1}{K}\sum_{k=1}^K \hat{\mathbf{g}}_k^t$ is the mean prediction by all the K models.

Pseudo-Label Supervision: To preserve the ability in gaze estimation during adaptation, we introduce pseudo-labels to supervise the gaze prediction in target. Since directly using the output of the gaze estimators in the variance minimization branch as the pseudo label may accumulate errors, we generate pseudo labels by employing the temporal average of the models to reduce the accumulated errors.

As can be seen in Fig. 2, we maintain a temporal average version of each estimator $\mathcal{G}(\cdot;\theta_k)$ as $\mathcal{G}(\cdot;\theta_k^T)$ at the T-th iteration during adaptation. The parameters θ_k^T is updated as:

$$\theta_k^T = \frac{T}{1+T}\theta_k^{(T-1)} + \frac{1}{1+T}\theta_k^T. \tag{6}$$

Then, the pseudo-label $\mathbf{p}_T^t$ of the image $\tilde{I}^t$ at the T-th iteration is defined as its mean predictions by the temporal averaged estimators $\{\mathcal{G}(\cdot;\theta_k^T)\}_{k=1}^K$:

$$\mathbf{p}_T^t = \frac{1}{K}\sum_{i=1}^{K}\hat{\mathbf{g}}_k^T = \frac{1}{K}\sum_{i=1}^{K}\mathcal{G}(\tilde{I}^t;\theta_k^T), \tag{7}$$

where $\hat{\mathbf{g}}_k^T = \mathcal{G}(\tilde{I}^t;\theta_k^T)$ is the prediction by the k-th temporal average model at the T-th iteration.

To preserve reliable gaze estimation, we require the predictions of the to-be-learned gaze estimators not to drift away from the pseudo-label by minimizing:

$$\ell_{wpl} = \frac{1}{K}\sum_{i=1}^{K}\omega^t|\hat{\mathbf{g}}_k^t - \mathbf{p}_T^t|, \tag{8}$$

where $\hat{\mathbf{g}}_k^t$ is the prediction on $\tilde{I}^t$ by the k-th estimator. $\omega^t = 1/\sqrt{\ell_{vm}(I^t)}$ weighs the reliability of each I^t's pseudo-label which has a negative correlation with the model uncertainty of I^t. It is noted that we regard ω^t as a coefficient and do not back-propagate gradients through it.

Objective Function for adaptation: During the source-free adaptation stage, only the set of source gaze estimators $\{\mathcal{G}(\cdot;\theta_k)\}_{k=1}^K$ are updated by minimizing the sum of ℓ_{vm} and ℓ_{wpl} over all the target data:

$$\min_{\{\theta_k\}_{k=1}^K} \mathbb{E}_{I^t\in\mathcal{D}_t'}[\ell_{vm}(I^t) + \gamma\ell_{wpl}(I^t)], \tag{9}$$

where $\mathcal{D}_t' \subseteq \mathcal{D}_t$ is a subset of target data, γ is the weight parameter to balance two losses. It is noting that a small set of target data is sufficient for the adaptation.

3.5. Inference on target data

The last row of Fig. 2 shows the pipeline of inference. Given a new image in the target domain, we predict the gaze by sequentially passing it through the face enhancer trained from source data(Sec. 3.3) and the mean estimator $\mathcal{G}(\cdot,\theta^\star)$ of the K gaze estimators $\{\mathcal{G}(\cdot,\theta_k^T)\}_{k=1}^K$ updated on target data(Sec. 3.4). The mean estimator's parameters $\theta^\star = \frac{1}{K}\sum_{k=1}^K\theta_k^T$ is set as the mean value of those in $\{\mathcal{G}(\cdot,\theta_k^T)\}_{k=1}^K$. Using the mean parameters has less computation cost than using the mean predictions and leads to better generalization than a single model [19].

Table 1. Angular gaze errors(°) of the baseline method and the variants of UnReGA on six cross-domain tasks

Method	Average Parameters	Image Enhancement	Source-free Adaptation	$\mathcal{D}_E \rightarrow \mathcal{D}_M$	$\mathcal{D}_E \rightarrow \mathcal{D}_D$	$\mathcal{D}_E \rightarrow \mathcal{D}_C$	$\mathcal{D}_G \rightarrow \mathcal{D}_M$	$\mathcal{D}_G \rightarrow \mathcal{D}_D$	$\mathcal{D}_G \rightarrow \mathcal{D}_C$
Baseline	×	×	×	7.50	7.88	7.81	7.23	8.02	9.49
ModelAvg	✓	×	×	7.18 ▼ 4.2%	7.25 ▼ 8.0%	7.31 ▼ 6.4%	6.90 ▼ 4.6%	7.32 ▼ 8.7%	8.78 ▼ 7.4%
EnhanceFace	✓	✓	×	5.92 ▼ 21.1%	6.31 ▼ 19.9%	6.62 ▼ 15.2%	6.52 ▼ 9.9%	7.05 ▼ 12.1%	7.83 ▼ 17.5%
UnReGA$^-$	✓	×	✓	5.35 ▼ 28.9%	6.06 ▼ 23.1%	5.91 ▼ 24.3%	5.58 ▼ 22.8%	5.84 ▼ 27.2%	6.80▼ 28.3%
UnReGA(w/o avg)	×	✓	✓	5.15 ▼ 31.3%	5.81 ▼ 26.3%	5.84 ▼ 25.2%	5.45 ▼ 24.6%	5.78 ▼ 27.9%	6.58▼ 30.7%
UnReGA	✓	✓	✓	5.11 ▼ 32.3%	5.70 ▼ 27.7%	5.75 ▼ 26.4%	5.42 ▼ 25.0%	5.80 ▼ 27.7%	6.52▼ 31.3%

4. Experiments

Through extensive experiments on cross-domain gaze estimation tasks, we investigate the effectiveness of the UnReGA framework and its components. We also discuss the advantage of uncertainty reduction.

4.1. Data Preparation

We employ five different gaze estimation datasets as five different domains: ETH-XGaze($\mathcal{D}_E$) [53], Gaze360($\mathcal{D}_G$) [23], GazeCapture($\mathcal{D}_C$) [26], MPIIGaze($\mathcal{D}_M$) [57], and EyeDiap($\mathcal{D}_D$) [14]. ETH-XGaze and Gaze360 are chosen as the source domains and the other three are target domains. We train our models on each source domain and test their adaptation performance on each target domain respectively. In addition, we use FFHQ [22] as our high quality face dataset $\mathcal{D}_h$ to train the face enhancer. **ETH-XGaze** is collected in a laboratory environment with 18 SLR cameras. It contains 756,540 high quality face images of 80 subjects. **Gaze360** is collected in both indoor and outdoor environments with a 360° camera. It contains images from 238 subjects with a wide distribution over gaze. Similar to [3,8], we use 84900 images with frontal faces as the source data. **MPIIFaceGaze** is collected in the daily environment with laptops from 15 subjects. We use 3000 face images from each subject as the target data. **GazeCapture** is collected in the daily environment with mobile phones and tablets. Following [26], we employ 179,496 images from 150 subjects as the target set. **EyeDiap** is collected in laboratory environments with screens and 3D floating balls. Following [2], we use 6,400 images using screen targets as target set and are manually checked by original authors. We process all the face images using the normalization method [54] to eliminate the variability of the camera's degree of freedom.

4.2. Implementation Details

We implement our method using Pytorch. We use Real-ESRGAN model [47] as the face enhancer and Resnet18 as the backbone of gaze estimators. During the training on source data, we pretrain the face enhancer on FFHQ with the same settings as in [47] and finetune it for 20000 iterations with a batch size of 16. We train the gaze estimator using the Adam [25] optimizer with a learning rate of 10^{-4} until 40 epochs. The batch size is 128. We chose $K = 10$ gaze estimators of the last 10 epochs. During the source-free adaptation, we use the Adam optimizer with a learning rate of 2×10^{-5} and set γ in Eq.(9) as 0.01. We randomly choose 100 unlabelled samples from target domain and reported average results of 100 repeated trials. The batch size is 20 and the model is trained for 10 epochs.

4.3. Effectiveness of UnReGA Framework

The UnReGA framework has three key components: face enhancement, source-free adaptation, and mean estimator with averaged parameters. We validate the effectiveness of each component by investigating variants of UnReGA with or without some of its components.

Table 1 reports the angular gaze errors of the baseline method and the variants of UnReGA. For **baseline**, we train a ResNet18 as the gaze estimator with the source data for 40 epochs. For **ModelAvg**, we average the parameters of gaze estimators from the last 10 epochs during the training of baseline. The mean estimator is evaluated on different target domains. As shown in Table 1, compared with the baseline, ModelAvg reduces the error by 4.2%, 8.0%, 6.4% from the source domain $\mathcal{D}_E$ to target $\mathcal{D}_M$, $\mathcal{D}_D$ and $\mathcal{D}_C$, and by 4.6%, 8.7%, 7.4% from source domain $\mathcal{D}_G$ to target $\mathcal{D}_M$, $\mathcal{D}_D$ and $\mathcal{D}_C$. It indicates that averaging the parameters is effective and contributes to better generalization ability.

EnhanceFace omits the source-free adaptation component in UnReGA. It applies the face enhancer on both source and target data, and trains the gaze estimator with the enhanced source data and then employs the mean estimator of the last 10 epochs on the enhanced target images. As shown in Table 1, EnhanceFace further reduces the errors when compared to ModelAvg. The improvements over the baseline are 21.1%, 19.9%, 15.2%, 9.9%, 12.1%, 17.5% on the six cross-domain tasks respectively. It indicates that reducing the sample uncertainty by a face enhancer helps reduce the domain gap and improves the performance on cross-domain tasks considerately. It is worth noting that EnhanceFace does not require any target samples, making it feasible for use in domain generalization scenarios, where target images are unavailable for adaptation.

UnReGA$^-$ omits the component of face enhancement in UnReGA. As shown in Table 1, UnReGA$^-$ significantly improves the performance of baselineand outperforms EnhanceFace. It means that the source-free adaptation mechanism is more effective than face enhancement and is crucial in the proposed UnReGA framework.

UnReGA integrates all three components and is shown

Table 2. Comparison with SOTA cross-domain gaze estimations. Results are reported by angular error (°).

Method	$\mathcal{D}_E \rightarrow \mathcal{D}_M$	$\mathcal{D}_E \rightarrow \mathcal{D}_D$	$\mathcal{D}_G \rightarrow \mathcal{D}_M$	$\mathcal{D}_G \rightarrow \mathcal{D}_D$
Only Source	7.50	7.88	7.23	8.02
w/o source				
PureGaze [5]	7.08	7.48	9.28	9.32
PnP-GA(oma) [34]	5.65	-	6.86	-
CSA† [48]	5.37	6.77	7.30	7.73
RUDA [2]	5.70	6.29	6.20	5.86
w/ source				
Gaze360 [23]	5.97	7.84	7.38	9.61
GazeAdv [45]	6.75	8.10	8.19	12.27
PnP-GA [34]	5.53	5.87	6.18	7.92
CRGA† [48]	5.68	5.72	6.09	6.68
UnReGA$^-$	5.35	6.06	5.58	5.84
UnReGA	**5.11**	**5.70**	**5.42**	**5.80**

† indicates the model employs Resnet50 [17] as the backbone.

to be with the best performance in Table 1.

4.4. Comparison with Cross-Domain Gaze Estimation Methods

To evaluate the superiority of UnReGA, we compare it with state-of-the-art (SOTA) cross-domain gaze estimation methods with or without source data during the adaptation.

The adaptation methods without source data (source-free adaptation) include: **PureGaze** [5] is a SOTA domain generalization method for gaze estimation using gaze feature purification. **CSA** [3] is a SOTA source-free domain adaptation method for gaze estimation using contrastive regression. **PnP-GA (oma)** [34] is a SOTA unsupervised domain adaptation for gaze estimation by outlier-guided model adaptation. we implement it using only outlier loss because other losses proposed by this method need source data. **RUDA** [2] is a SOTA unsupervised gaze adaptation method using rotation consistency.

The adaptation methods with source data (unsupervised domain adaptation) include: **GazeAdv** [45] is a SOTA unsupervised domain adaptation for gaze estimation by adversarial learning. **Gaze360** [23] is a SOTA unsupervised gaze adaptation method by adversarial learning and pinball loss. **PnP-GA** [34] is a SOTA unsupervised gaze adaptation method by outlier-guided collaborative adaptation. **CRGA** [3] is a SOTA unsupervised gaze adaptation method using contrastive regression. For a more fair comparison, we use 100 target and source samples for adaptation with CRGA.

Table 2 shows the angular errors of UnReGA and other methods on five cross-domain tasks. Both UnReGA$^-$ and UnReGA outperform all the state-of-the-art source-free adaptation methods. Besides, UnReGA outperforms all the unsupervised gaze adaptation methods despite they use of source data for adaptation. Moreover, even without enhancement, UnReGA$^-$ also shows superior performance on these domain adaptation tasks, except for $\mathcal{D}_E \rightarrow \mathcal{D}_D$, slightly inferior compared to CRGA [48], which employs a Resnet50 backbone and use source data during adaptation.

Table 3. Angular gaze errors (°) of methods with pretrained or finetuned face enhancers using different loss functions.

Method	$\mathcal{D}_E \rightarrow \mathcal{D}_M$	$\mathcal{D}_E \rightarrow \mathcal{D}_D$	$\mathcal{D}_G \rightarrow \mathcal{D}_M$	$\mathcal{D}_G \rightarrow \mathcal{D}_D$
pretrained	6.12	6.48	6.74	7.11
finetune w/ ℓ_{gc}	6.03	6.43	6.69	7.08
finetune w/ ℓ_{gc}+ℓ_g	**5.92**	**6.31**	**6.52**	**7.05**

Table 4. Mean angular gaze errors (°) $\pm$ stand deviations for UnReGAs with different loss functions in source-free adaptation.

Method	$\mathcal{D}_E \rightarrow \mathcal{D}_M$	$\mathcal{D}_E \rightarrow \mathcal{D}_D$	$\mathcal{D}_G \rightarrow \mathcal{D}_M$	$\mathcal{D}_G \rightarrow \mathcal{D}_D$
UnReGA$^-$				
w/o adaptation	7.50	7.88	7.23	8.02
ℓ_{vm}	5.48 ± 0.11	6.39 ± 0.17	5.65 ± 0.15	6.50 ± 0.23
ℓ_{wpl}	5.98 ± 0.17	6.10 ± 0.12	5.91 ± 0.14	6.01 ± 0.15
$\ell_{vm} + \ell_{pl}$	5.51 ± 0.17	6.13 ± 0.22	5.70 ± 0.08	5.92 ± 0.21
$\ell_{vm} + \ell_{wpl}$	$\mathbf{5.35} \pm 0.20$	$\mathbf{6.06} \pm 0.17$	$\mathbf{5.58} \pm 0.15$	$\mathbf{5.84} \pm 0.18$
UnReGA				
w/o adaptation	5.92	6.31	6.52	7.05
ℓ_{vm}	5.19 ± 0.11	6.21 ± 0.23	5.56 ± 0.06	6.22 ± 0.11
ℓ_{wpl}	5.26 ± 0.09	5.81 ± 0.06	5.83 ± 0.08	5.92 ± 0.14
$\ell_{vm} + \ell_{pl}$	5.16 ± 0.10	5.75 ± 0.12	5.43 ± 0.06	5.96 ± 0.11
$\ell_{vm} + \ell_{wpl}$	$\mathbf{5.11} \pm 0.09$	$\mathbf{5.70} \pm 0.16$	$\mathbf{5.42} \pm 0.06$	$\mathbf{5.80} \pm 0.12$

4.5. Ablation Study

We investigate the effectiveness of each loss item during the stages of training on source data and source-free adaptation in the UnReGA framework. Due to limited space, the study of other hyperparameters is in *suppl*.

4.5.1 Loss Terms for Training on Source Data

We propose gaze loss ℓ_g and gaze consistency loss ℓ_{gc} to finetune the face enhancer for keeping the gaze unchanged in enhanced images (Sec.3.3). We investigate the effectiveness of ℓ_g and ℓ_{gc} by comparing the methods with and without them. For convenience, we follow the experimental protocol as EnhanceFace in Sec. 4.3. Table 3 reports the results. The results demonstrate that both two losses improve the baseline on four cross-domain tasks.

4.5.2 Loss Terms for Source-free Adaptation

We investigate the mechanisms of variance minimization (with ℓ_{vm}) and pseudo-label supervision (with ℓ_{wpl}) in source-free adaptation under both the settings as UnReGA$^-$ and UnReGA in Sec. 4.3. Table 4 reports the mean gaze errors with different losses. The results demonstrate that adaptation ℓ_{vm} or ℓ_{wpl} individually achieve performance improvement over baseline and adaptation with $\ell_{vm} + \ell_{wpl}$ achieve the best performance. Besides, to verify the effectiveness of weight in ℓ_{wpl}, we substitute ℓ_{wpl} with $\ell_{pl} = \frac{1}{K}\sum_{i=1}^{K} |\hat{\mathbf{g}}_k^t - \mathbf{p}_T^t|$ by removing ω^t in Eq.(8). Results in Table 4 show the advantage of weight in ℓ_{wpl}.

To investigate why adaptation with both ℓ_{vm} and ℓ_{wpl} outperforms adaptation with only one of them, we plot the trend of gaze errors of adaptation with different losses over iterations in Fig.4. The results show that utilizing either ℓ_{vm} or ℓ_{wpl} individually can be beneficial for adaptation. How-

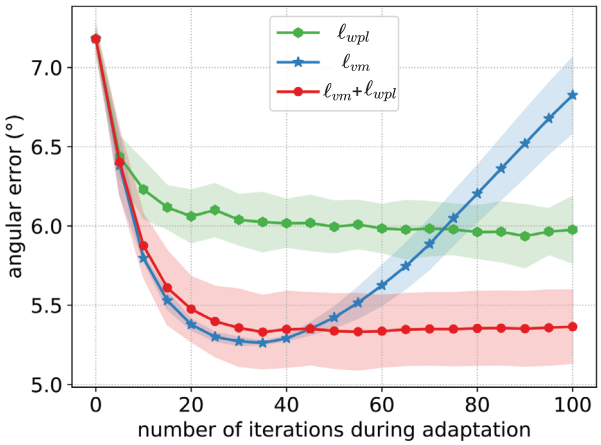

Figure 4. The trend of angular errors (°) over iterations with different loss functions in source-free adaptation stage. The light colors denote the standard deviation of 100 times experiments. The experiments are conducted on $\mathcal{D}_E \rightarrow \mathcal{D}_M$ under UnReGA$^-$ setting.

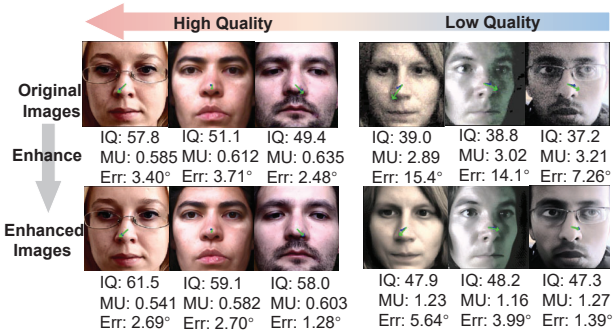

Figure 5. Examples of the high quality and low quality images and their image quality (IQ), model uncertainty (MU) and gaze errors (Err). The blue and green arrows denote the gaze labels and the predictions respectively.

ever, each loss function has its advantages and disadvantages. Specifically, solely employing ℓ_{vm} can significantly reduce gaze errors but errors may increase after a certain iteration, which is challenging to identify without access to labeled validation data. Conversely, utilizing ℓ_{wpl} alone can maintain stable gaze errors after convergence, but its performance is not as excellent as the best iteration of ℓ_{vm} alone. Combining ℓ_{vm} and ℓ_{wpl} during adaptation can leverage the strengths of both loss functions, resulting in satisfactory performance and stable results during optimization.

4.6. Discussion about Uncertainty Reduction

To discuss how image quality influences the model uncertainty and gaze errors, we visualize some high quality and low quality examples and their enhanced pairs in Fig.5. We report their image quality (IQ), model uncertainty (MU) and gaze errors (Err). We measure the image quality (IQ) with a popular blind image quality assessment method [52]

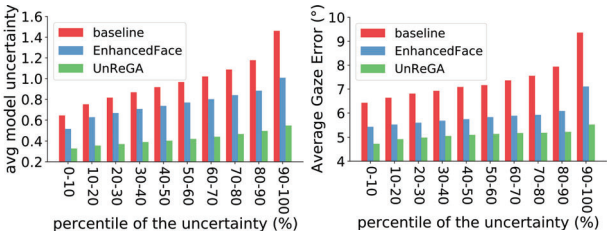

Figure 6. Average model uncertainty and average gaze errors of samples grouped according to the percentile of the baseline's uncertainty.

and output the model uncertainty and gaze errors with a set of gaze estimators. Compared with high-quality images, low-quality samples tend to have higher model uncertainty and higher gaze errors. After face enhancement on samples, the model uncertainty of both high-quality and low-quality images decreases and so do the gaze errors. Moreover, the enhancement of low-quality images brings more performance gain than high-quality images.

To understand the correlation between reducing model uncertainty and reducing gaze errors, we illustrate the model uncertainty and gaze errors of applying EnhanceFace and UnReGA on different samples grouped by model uncertainty in Fig.6. Specifically, we compute the model uncertainty of samples in $\mathcal{D}_M$ with a set of gaze estimators trained with $\mathcal{D}_E$ and sort the samples by the model uncertainty in ascending order and group them by every 10-th percentile. We take the set of gaze estimators as the baseline and apply EnhanceFace and UnReGA on $\mathcal{D}_E \rightarrow \mathcal{D}_M$. Subsequently, we calculate the average gaze errors and average model uncertainty for each group. The results indicate that both EnhanceFace and UnReGA can consistently reduce model uncertainty and gaze errors for groups with different uncertainty over baseline. Moreover, the higher uncertainty of the samples, the more uncertainty and errors can be reduced by EnhanceFace and UnReGA.

5. Conclusion

We present a novel uncertainty reduction gaze adaptation (UnReGA) framework for adapting gaze estimators on the unlabelled target domain without source data. UnReGA improves gaze estimation performance on the target data by reducing uncertainty on the target. Our source-free adaptation method shows significant performance improvements over baseline and also outperforms the SOTA gaze adaptation methods using source during adaptation on adaptation tasks. In the future, the connection between face enhancement and the minimization of sample uncertainty can be discussed by formulating sample uncertainty mathematically and the proposed uncertainty reduction method can be explored on other cross-domain regression problems.

References

[1] Yiwei Bao, Yihua Cheng, Yunfei Liu, and Feng Lu. Adaptive feature fusion network for gaze tracking in mobile tablets. In *The International Conference on Pattern Recognition*, 2020. 2

[2] Yiwei Bao, Yunfei Liu, Haofei Wang, and Feng Lu. Generalizing gaze estimation with rotation consistency. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4207–4216, 2022. 3, 6, 7

[3] Dian Chen, Dequan Wang, Trevor Darrell, and Sayna Ebrahimi. Contrastive test-time adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 295–305, 2022. 3, 6, 7

[4] Zhaokang Chen and Bertram E Shi. Appearance-based gaze estimation using dilated-convolutions. In *Asian Conference on Computer Vision*, pages 309–324. Springer, 2018. 1

[5] Yihua Cheng, Yiwei Bao, and Feng Lu. Puregaze: Purifying gaze feature for generalizable gaze estimation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 436–443, 2022. 3, 7

[6] Yihua Cheng, Shiyao Huang, Fei Wang, Chen Qian, and Feng Lu. A coarse-to-fine adaptive network for appearance-based gaze estimation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020. 1

[7] Yihua Cheng, Feng Lu, and Xucong Zhang. Appearance-based gaze estimation via evaluation-guided asymmetric regression. In *The European Conference on Computer Vision (ECCV)*, September 2018. 2

[8] Yihua Cheng, Haofei Wang, Yiwei Bao, and Feng Lu. Appearance-based gaze estimation with deep learning: A review and benchmark. *arXiv preprint arXiv:2104.12668*, 2021. 6

[9] Haoping Deng and Wangjiang Zhu. Monocular free-head 3d gaze tracking with deep learning and geometry constraints. In *The IEEE International Conference on Computer Vision*, Oct 2017. 2

[10] Ning Ding, Yixing Xu, Yehui Tang, Chao Xu, Yunhe Wang, and Dacheng Tao. Source-free domain adaptation via distribution estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7212–7222, 2022. 3

[11] Sidney D'Mello, Andrew Olney, Claire Williams, and Patrick Hays. Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of human-computer studies*, 70(5):377–398, 2012. 1

[12] Tobias Fischer, Hyung Jin Chang, and Yiannis Demiris. Rt-gene: Real-time eye gaze estimation in natural environments. In *The European Conference on Computer Vision (ECCV)*, September 2018. 1, 2

[13] Francois Fleuret et al. Uncertainty reduction for model adaptation in semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9613–9623, 2021. 3

[14] Kenneth Alberto Funes Mora, Florent Monay, and Jean-Marc Odobez. Eyediap: A database for the development and evaluation of gaze estimation algorithms from rgb and rgb-d cameras. In *Proceedings of the symposium on eye tracking research and applications*, pages 255–258, 2014. 2, 6

[15] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR, 2016. 2

[16] Zidong Guo, Zejian Yuan, Chong Zhang, Wanchao Chi, Yonggen Ling, and Shenghao Zhang. Domain adaptation gaze estimation by embedding with prediction consistency. In *Proceedings of the Asian Conference on Computer Vision*, 2020. 1

[17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 4, 7

[18] Jiaxing Huang, Dayan Guan, Aoran Xiao, and Shijian Lu. Model adaptation: Historical contrastive learning for unsupervised domain adaptation without source data. *Advances in Neural Information Processing Systems*, 34:3635–3649, 2021. 3

[19] P Izmailov, AG Wilson, D Podoprikhin, D Vetrov, and T Garipov. Averaging weights leads to wider optima and better generalization. In *34th Conference on Uncertainty in Artificial Intelligence 2018, UAI 2018*, pages 876–885, 2018. 5

[20] Robert JK Jacob and Keith S Karn. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In *The mind's eye*, pages 573–605. Elsevier, 2003. 1

[21] Ming Jiang and Qi Zhao. Learning visual attention to identify people with autism spectrum disorder. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3267–3276, 2017. 1

[22] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019. 6

[23] Petr Kellnhofer, Adria Recasens, Simon Stent, Wojciech Matusik, and Antonio Torralba. Gaze360: Physically unconstrained gaze estimation in the wild. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6912–6921, 2019. 2, 6, 7

[24] Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? *Advances in neural information processing systems*, 30, 2017. 2, 3

[25] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6

[26] Kyle Krafka, Aditya Khosla, Petr Kellnhofer, Harini Kannan, Suchendra Bhandarkar, Wojciech Matusik, and Antonio Torralba. Eye tracking for everyone. In *The IEEE Conference on Computer Vision and Pattern Recognition*, June 2016. 2, 6

[27] Jogendra Nath Kundu, Akshay Kulkarni, Amit Singh, Varun Jampani, and R Venkatesh Babu. Generalize then adapt: Source-free domain adaptive semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7046–7056, 2021. 3

[28] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems*, 30, 2017. 2, 5

[29] Rui Li, Qianfen Jiao, Wenming Cao, Hau-San Wong, and Si Wu. Model adaptation: Unsupervised domain adaptation without source data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9641–9650, 2020. 3

[30] Shuaifeng Li, Mao Ye, Xiatian Zhu, Lihua Zhou, and Lin Xiong. Source-free object detection by learning to overlook domain style. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8014–8023, 2022. 3

[31] Xianfeng Li, Weijie Chen, Di Xie, Shicai Yang, Peng Yuan, Shiliang Pu, and Yueting Zhuang. A free lunch for unsupervised domain adaptive object detection without source data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 8474–8481, 2021. 3

[32] Jian Liang, Dapeng Hu, and Jiashi Feng. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In *International Conference on Machine Learning*, pages 6028–6039. PMLR, 2020. 2, 3

[33] Yuejiang Liu, Parth Kothari, Bastien van Delft, Baptiste Bellot-Gurlet, Taylor Mordan, and Alexandre Alahi. Ttt++: When does self-supervised test-time training fail or thrive? *Advances in Neural Information Processing Systems*, 34:21808–21820, 2021. 3

[34] Yunfei Liu, Ruicong Liu, Haofei Wang, and Feng Lu. Generalizing gaze estimation with outlier-guided collaborative adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3835–3844, 2021. 1, 3, 7

[35] Päivi Majaranta and Andreas Bulling. Eye tracking and eye-based human–computer interaction. In *Advances in physiological computing*, pages 39–65. Springer, 2014. 1

[36] Carlos H Morimoto and Marcio RM Mimica. Eye gaze tracking techniques for interactive applications. *Computer vision and image understanding*, 98(1):4–24, 2005. 1

[37] Seonwook Park, Shalini De Mello, Pavlo Molchanov, Umar Iqbal, Otmar Hilliges, and Jan Kautz. Few-shot adaptive gaze estimation. In *The IEEE International Conference on Computer Vision*, October 2019. 2

[38] Seonwook Park, Adrian Spurr, and Otmar Hilliges. Deep pictorial gaze estimation. In *The European Conference on Computer Vision*, September 2018. 2

[39] Steffen Schneider, Evgenia Rusak, Luisa Eck, Oliver Bringmann, Wieland Brendel, and Matthias Bethge. Improving robustness against common corruptions by covariate shift adaptation. *Advances in Neural Information Processing Systems*, 33:11539–11551, 2020. 3

[40] Yusuke Sugano, Yasuyuki Matsushita, and Yoichi Sato. Learning-by-synthesis for appearance-based 3d gaze estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1821–1828, 2014. 2

[41] Yu Sun, Xiaolong Wang, Zhuang Liu, John Miller, Alexei Efros, and Moritz Hardt. Test-time training with self-supervision for generalization under distribution shifts. In *International conference on machine learning*, pages 9229–9248. PMLR, 2020. 3

[42] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7167–7176, 2017. 1

[43] 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

[44] Jindong Wang, Cuiling Lan, Chang Liu, Yidong Ouyang, Tao Qin, Wang Lu, Yiqiang Chen, Wenjun Zeng, and Philip Yu. Generalizing to unseen domains: A survey on domain generalization. *IEEE Transactions on Knowledge and Data Engineering*, 2022. 2

[45] Kang Wang, Rui Zhao, Hui Su, and Qiang Ji. Generalizing eye tracking with bayesian adversarial learning. In *The IEEE Conference on Computer Vision and Pattern Recognition*, June 2019. 1, 3, 7

[46] Shuo Wang, Ming Jiang, Xavier Morin Duchesne, Elizabeth A Laugeson, Daniel P Kennedy, Ralph Adolphs, and Qi Zhao. Atypical visual saliency in autism spectrum disorder quantified through model-based eye tracking. *Neuron*, 88(3):604–616, 2015. 1

[47] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1905–1914, 2021. 4, 5, 6

[48] Yaoming Wang, Yangzhou Jiang, Jin Li, Bingbing Ni, Wenrui Dai, Chenglin Li, Hongkai Xiong, and Teng Li. Contrastive regression for domain adaptation on gaze estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19376–19385, 2022. 7

[49] Garrett Wilson and Diane J Cook. A survey of unsupervised deep domain adaptation. *ACM Transactions on Intelligent Systems and Technology*, 11(5):1–46, 2020. 2

[50] Haifeng Xia, Handong Zhao, and Zhengming Ding. Adaptive adversarial network for source-free domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9010–9019, 2021. 3

[51] Hao-Wei Yeh, Baoyao Yang, Pong C Yuen, and Tatsuya Harada. Sofa: Source-data-free feature alignment for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 474–483, 2021. 3

[52] Weixia Zhang, Kede Ma, Jia Yan, Dexiang Deng, and Zhou Wang. Blind image quality assessment using a deep bilinear convolutional neural network. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(1):36–47, 2018. 8

[53] Xucong Zhang, Seonwook Park, Thabo Beeler, Derek Bradley, Siyu Tang, and Otmar Hilliges. Eth-xgaze: A large scale dataset for gaze estimation under extreme head pose and gaze variation. In *European Conference on Computer Vision*, pages 365–381. Springer, 2020. 2, 6

[54] Xucong Zhang, Yusuke Sugano, and Andreas Bulling. Revisiting data normalization for appearance-based gaze estimation. In *Proceedings of the 2018 ACM symposium on eye tracking research & applications*, pages 1–9, 2018. 6

[55] Xucong Zhang, Yusuke Sugano, Andreas Bulling, and Otmar Hilliges. Learning-based region selection for end-to-end gaze estimation. In *The British Machine Vision Conference*, 2020. 2

[56] Xucong Zhang, Yusuke Sugano, Mario Fritz, and Andreas Bulling. Appearance-based gaze estimation in the wild. In *The IEEE Conference on Computer Vision and Pattern Recognition*, June 2015. 1

[57] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling. Mpiigaze: Real-world dataset and deep appearance-based gaze estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(1):162–175, Jan 2019. 1, 2, 6

[58] Aurick Zhou and Sergey Levine. Bayesian adaptation for covariate shift. *Advances in Neural Information Processing Systems*, 34:914–927, 2021. 3