

Improving Domain Generalization on Gaze Estimation via Branch-out Auxiliary Regularization

Ruijie Zhao, *Student Member, IEEE*, Pinyan Tang, and Sihui Luo*, *Member, IEEE*

Abstract—Despite remarkable advancements, mainstream gaze estimation techniques, particularly appearance-based methods, often suffer from performance degradation in uncontrolled environments due to variations in illumination and individual facial attributes. Existing domain adaptation strategies, limited by their need for target domain samples, may fall short in real-world applications. This letter introduces Branch-out Auxiliary Regularization (BAR), an innovative method designed to boost gaze estimation’s generalization capabilities without requiring direct access to target domain data. Specifically, BAR integrates two auxiliary consistency regularization branches: one that uses augmented samples to counteract environmental variations, and another that aligns gaze directions with positive source domain samples to encourage the learning of consistent gaze features. These auxiliary pathways strengthen the core network and are integrated in a smooth, plug-and-play manner, facilitating easy adaptation to various other models. Comprehensive experimental evaluations on four cross-dataset tasks demonstrate the superiority of our approach.

Index Terms—Gaze estimation, domain generalization, data augmentation, contrastive learning, deep learning.

I. INTRODUCTION

GAZE estimation, which determines the gaze direction of an individual, has widespread applications across various domains such as autonomous driving [1], [2], virtual reality (VR) [3], and psychological research [4]. There are two primary approaches to gaze estimation: model-based and appearance-based methods. Model-based methods [5], [6], [7], [8], [9], [10] rely on expensive eye trackers and require detailed calibration for each user. On the other hand, appearance-based methods [11], [12], [13], [14], [15] leverage cost-effective webcams and deep learning models to infer gaze directions, offering a more accessible solution.

Appearance-based methods demonstrate impressive performance within controlled environments [16]. However, challenges arise in uncontrolled settings, where variations in illumination and diverse individuals pose obstacles to accurate gaze estimation, thereby hindering the generalization of gaze estimation models in cross-domain situations.

Recent studies [17], [18], [19], [20], [21] have explored the use of unsupervised domain adaptation techniques to extend gaze estimation models from source to target domains. However, these methods typically assume access to target domain samples, which poses challenges for practical application.

This work was supported in part by the Natural Science Foundation Projects of Zhejiang Province [Grant NO. LQ22H090011 and LQ22F020020, 2021].

Ruijie Zhao, Pinyan Tang, and Sihui Luo (*corresponding author) are with the faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo, China (e-mail: 216001944@nbu.edu.cn, tangpinyan@nbu.edu.cn, luosihui@nbu.edu.cn).

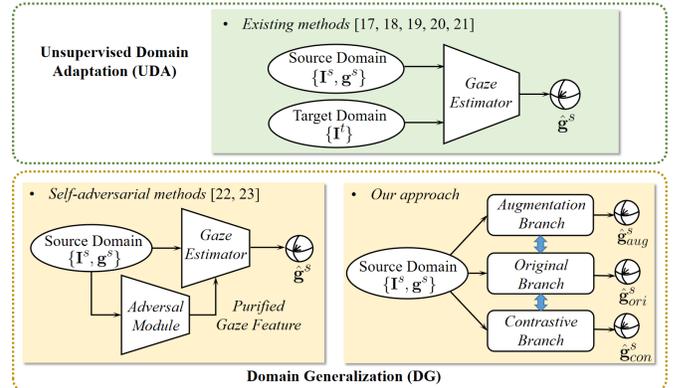


Fig. 1. Comparison of our method with conventional UDA and DG methods for gaze estimation. UDA methods generally rely on target domain data to enable knowledge transfer from the source domain to the target domain. Conventional DG methods leverage uncontrolled adversarial learning techniques that may cause feature elimination problems. Our method, in contrast, requires no access to target domain data and leverages flexible auxiliary consistency regularization branches to enhance model generalization.

Enhancing the generalization capabilities of gaze estimation models without direct access to the target domain presents significant hurdles. Recent research [22], [23] employs adversarial learning to prevent models from learning gaze-irrelevant features, thereby mitigating domain-specific influences like illumination and identity. Nonetheless, the inherent complexity of gaze estimation and the risk of oversimplifying by removing irrelevant features might divert the model from its core aim of precise gaze estimation.

To address these challenges, we introduce Branch-out Auxiliary Regularization (BAR), a straightforward and flexible approach to enhance domain generalization in gaze estimation models. This framework enhances a pre-trained original gaze estimation network with two additional consistency regularization branches: the augmentation and the contrast branch. The augmentation branch inputs sophisticatedly augmented samples to reinforce the model’s generalization capability against environmental variations. Simultaneously, the contrast branch utilizes semantically consistent samples, which share similar gaze directions but differ in identity, to enforce the model to generate similar predictions. These enhancements enable the model to ignore gaze-irrelevant attributes and focus on learning gaze-consistent features, thereby avoiding the pitfalls of adversarial learning methods that may lead to the feature elimination problem. Fig. 1 illustrates the difference between conventional methods and our approach.

To summarize, the contributions of this letter are twofold.

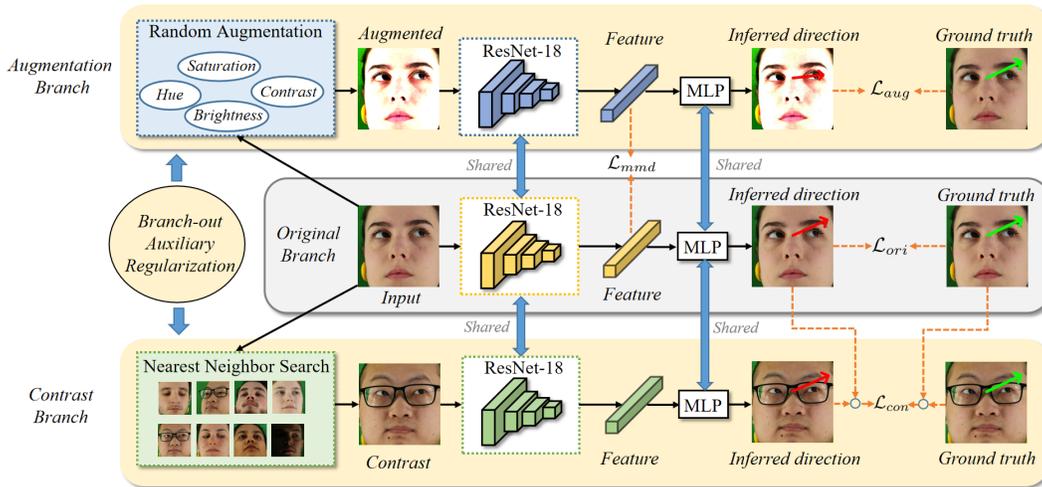


Fig. 2. The overall framework of our approach. We extend the original gaze estimation network by integrating two auxiliary consistency regularization pathways in a plug-and-play manner. Notably, the auxiliary branches are exclusively employed during the training phase and do not influence the test phase.

First, we present Gaze-BAR, a plug-and-play domain generalization method for gaze estimation that eliminates the need for target domain data and can be easily deployed to existing models to improve accuracy. Second, we propose to utilize auxiliary consistency regularization branches with sophisticated augmented samples and semantically consistent cross-identity samples as input to foster the model’s generalization across varying environmental conditions and identities. Experimental results demonstrate that our method significantly surpasses the baseline model in performance and outperforms state-of-the-art domain generalization approaches for gaze estimation.

II. DOMAIN GENERALIZATION FOR GAZE ESTIMATION

A. Preliminaries

This section outlines the domain generalization task for gaze estimation. We begin by defining the source (training) domain as $S_{train} = \{(\mathbf{I}_i^s, \mathbf{g}_i^s)\}_{i=1}^{N_s}$, where \mathbf{I}_i^s represents the input image from the source domain, \mathbf{g}_i^s denotes the corresponding gaze direction, and N_s indicates the number of samples in the source domain dataset. Similarly, the target (testing) domain is denoted as $S_{test} = \{(\mathbf{I}_i^t, \mathbf{g}_i^t)\}_{i=1}^{N_t}$.

The formulation of the gaze estimation is represented as:

$$\hat{\mathbf{g}} = R(F(\mathbf{I})), \quad (1)$$

where $F(\cdot)$ represents the feature extractor, and $R(\cdot)$ denotes the Multi-Layer Perceptron (MLP) to regress the feature vector into gaze direction.

The predicted angular error $E[\mathbf{g}, \hat{\mathbf{g}}]$ is defined as the angular disparity between the predicted gaze direction $\hat{\mathbf{g}}$ and the actual gaze direction \mathbf{g} :

$$E[\mathbf{g}, \hat{\mathbf{g}}] = \arccos \left(\frac{\mathbf{g} \cdot \hat{\mathbf{g}}}{\|\mathbf{g}\| \|\hat{\mathbf{g}}\|} \right). \quad (2)$$

Therefore, the objective is to optimize the model parameters θ to minimize the inference loss in the target domain:

$$\theta = \arg \min_{\theta} \sum_i E[\mathbf{g}_i^t, \hat{\mathbf{g}}_i^t]. \quad (3)$$

With restricted access to the target domain, our approach aims to guide the model to focus on gaze-specific features while filtering out irrelevant ones. We identify two distinct categories of gaze-irrelevant factors:

- Environmental factors: Variations in image brightness, contrast, and tone lead to disparities.
- Facial attribute factors: Individual appearance variations such as skin complexion, facial morphology, and expressions.

To mitigate the effects of these gaze-irrelevant attributes, we introduce two auxiliary consistency regularization branches to the original model: the augmentation and the contrast branch. The augmentation branch encourages learning the consistent features regardless of varying environmental factors, while the contrast branch aims to disentangle gaze-specific features from irrelevant features concerning facial identity. The overall framework of our approach is shown in Fig. 2.

B. Original Branch

We employ ResNet-18 [24] as the feature extractor backbone of the original gaze estimation model. The extracted feature vector is then passed through a single-layer MLP to regress the gaze direction. Besides, the original branch is trained with the L1 Loss between the predicted gaze direction $\hat{\mathbf{g}}_{ori}$ and the ground truth gaze direction \mathbf{g}_{ori} :

$$\mathcal{L}_{ori} = |\mathbf{g}_{ori} - \hat{\mathbf{g}}_{ori}|. \quad (4)$$

C. Augmentation Branch

To address the challenge of varying illumination, we introduce the augmentation branch. Samples with random adjustments in brightness, contrast, and saturation are fed to this branch, simulating diverse environmental conditions to enhance the model’s resilience to such variations. This branch is trained with the L1 Loss \mathcal{L}_{aug} , which is formulated as:

$$\mathcal{L}_{aug} = |\mathbf{g}_{ori} - \hat{\mathbf{g}}_{aug}|, \quad (5)$$

where $\hat{\mathbf{g}}_{aug}$ represents the predicted gaze direction for the augmented image.

Furthermore, to acquire consistent feature extraction across diverse illumination conditions, we align features from the original and augmented images by minimizing their disparity using Maximum Mean Discrepancy Loss (denoted as \mathcal{L}_{mmd}):

$$\mathcal{L}_{mmd} = \left\| \frac{1}{N} \sum_{i=1}^N \varphi(x_i) - \frac{1}{N} \sum_{j=1}^N \varphi(y_j) \right\|, \quad (6)$$

where x_i and y_j denote the features extracted from the original and augmented images, φ represents the Gaussian kernel function, and N is the size of the feature.

D. Contrast Branch

To enhance generalized gaze estimation, we tackle facial attribute variability through the contrast branch. This branch focuses on identifying positive samples with similar gaze directions but differing in facial attributes. These samples are selected from the source domain using the K-Nearest Neighbor (KNN) algorithm. For efficient searching, we employ the Ball-Tree method [25], which expedites nearest neighbor searches by leveraging the query point's location and its proximity to spherical regions, thus significantly cutting down search time.

In gaze estimation, $(pitch, yaw)$ denotes the vertical and horizontal angles of an individual's gaze direction, with pitch and yaw referring to up-down and left-right eye movements, respectively. Thus, we define the distance d between two points in the searching tree as follows:

$$d = \sqrt{(pitch_2 - pitch_1)^2 + (yaw_2 - yaw_1)^2}. \quad (7)$$

Here, $(pitch_1, yaw_1)$ and $(pitch_2, yaw_2)$ represent the coordinates of the query point and a point in the tree respectively.

The contrast branch aligns the gaze directions of positive samples with original images by minimizing contrast loss (\mathcal{L}_{con}), thereby directing the model to learn gaze-relevant features and disentangle gaze-irrelevant features concerning identity attributes. \mathcal{L}_{con} is computed as follows:

$$\mathcal{L}_{con} = |\hat{\mathbf{g}}_{con} - \hat{\mathbf{g}}_{ori}|, \quad (8)$$

where $\hat{\mathbf{g}}_{ori}$ and $\hat{\mathbf{g}}_{con}$ represent the gaze predictions for the original image and the positive samples, respectively. However, due to the precision error stemming from the KNN search, disparities arise between the gaze direction of positive samples and the annotated direction of the original image. Therefore, we modify the \mathcal{L}_{con} as:

$$\mathcal{L}_{con} = |(\hat{\mathbf{g}}_{con} - \hat{\mathbf{g}}_{ori}) - (\mathbf{g}_{con} - \mathbf{g}_{ori})|, \quad (9)$$

the subtraction term $(\mathbf{g}_{ori} - \mathbf{g}_{con})$ helps mitigate precision errors introduced by KNN.

E. Overall Loss Function

To unify the learning objectives of the original, augmentation, and contrast branches, we combine their respective loss functions into a single formulation. This holistic loss function, denoted as \mathcal{L}_{total} , is formulated as:

$$\mathcal{L}_{total} = \mathcal{L}_{ori} + \lambda_a \mathcal{L}_{aug} + \lambda_m \mathcal{L}_{mmd} + \lambda_c \mathcal{L}_{con}. \quad (10)$$

Here, $\lambda_a, \lambda_m, \lambda_c$ are hyperparameters and we set $\lambda_a = \lambda_m = \lambda_c = 1.0$ in our implementation.

III. EXPERIMENT

A. Data Preparation and Evaluation Metric

In this letter, we validate the efficacy of our model across four distinct datasets: ETH-XGaze (\mathcal{D}_E) [28], Gaze360 (\mathcal{D}_G) [13], MPIIGaze (\mathcal{D}_M) [29], and EyeDiap (\mathcal{D}_D) [30]. ETH-XGaze encompasses 110 subjects, contributing a total of 1.1M images, with a training set provided for 80 subjects. We partition 5 subjects from this pool for validation purposes. Gaze360 comprises 238 subjects, yielding 172K images after excluding those devoid of eye-related content. Following the protocol proposed by Sugano et al. [31], we process MPIIGaze to extract 45K images from 15 subjects. EyeDiap offers 94 video clips from 16 subjects, from which we sample one image every 15 frames for analysis.

Following [22], [23], we designate ETH-XGaze and Gaze360 as training sets, while MPIIGaze and EyeDiap are reserved for model evaluation. Consequently, four cross-dataset tasks are investigated in our experiments: $\mathcal{D}_E \rightarrow \mathcal{D}_M$, $\mathcal{D}_E \rightarrow \mathcal{D}_D$, $\mathcal{D}_G \rightarrow \mathcal{D}_M$, and $\mathcal{D}_G \rightarrow \mathcal{D}_D$.

In all experiments of this paper, we leverage the angular error in Eq. (2) as the evaluation metric. Smaller angular error means better inference accuracy.

B. Implementation Details

The experiments were conducted using the PyTorch framework and on a single RTX 4080 GPU. Our model was trained under the batch size of 64 and the Adam optimizer with a learning rate of 1×10^{-4} . We resize and normalize the input images to 224×224 and $[0, 1]$. The model was trained for 10 epochs on ETH-XGaze and 100 epochs on Gaze360.

C. Comparison Methods

1) *Baseline*: The baseline model is the original single-branch gaze estimation model consisting of a ResNet-18 pre-trained on ImageNet as the backbone and an MLP module.

2) *State-of-the-art methods*:

- Gaze estimation methods of Full-Face [16], CA-Net [26], RT-Genie [12], Dilated-Net [27], and ADL* [13].
- Domain generalization methods for gaze estimation including PureGaze [22] and Gaze-Consistent [23].

D. Ablation Studies

This section presents ablation studies aimed at dissecting the contributions of each component within our approach. We assess the impact of removing \mathcal{L}_{aug} , \mathcal{L}_{mmd} , or \mathcal{L}_{con} , comparing these variations with both the baseline model and our complete methodology.

The results, summarized in Table I, indicate that combining the original loss (\mathcal{L}_{ori}) with the additional loss terms leads to significant improvements in gaze estimation accuracy. Notably, the most substantial improvement is observed when all loss components are integrated, yielding a reduction of 16.06% in average angular error compared to the baseline. These findings underscore the efficacy of the proposed multi-loss framework in enhancing gaze estimation across diverse datasets.

TABLE I
ABLATION STUDIES OF DIFFERENT COMPONENTS WITHIN OUR MODEL

Method	$\mathcal{D}_E \rightarrow \mathcal{D}_M(\downarrow)$	$\mathcal{D}_E \rightarrow \mathcal{D}_D(\downarrow)$	$\mathcal{D}_G \rightarrow \mathcal{D}_M(\downarrow)$	$\mathcal{D}_G \rightarrow \mathcal{D}_D(\downarrow)$	Average(\downarrow)
\mathcal{L}_{ori}	$8.02^\circ \pm 0.09^\circ$	$9.11^\circ \pm 0.31^\circ$	$8.04^\circ \pm 0.06^\circ$	$9.20^\circ \pm 0.08^\circ$	8.59°
$\mathcal{L}_{ori} + \mathcal{L}_{aug}$	$7.08^\circ \pm 0.12^\circ$	$7.29^\circ \pm 0.22^\circ$	$7.35^\circ \pm 0.09^\circ$	$8.88^\circ \pm 0.06^\circ$	7.65° ▼10.94%
$\mathcal{L}_{ori} + \mathcal{L}_{con}$	$7.34^\circ \pm 0.10^\circ$	$8.34^\circ \pm 0.26^\circ$	$7.76^\circ \pm 0.05^\circ$	$8.48^\circ \pm 0.12^\circ$	7.98° ▼7.10%
$\mathcal{L}_{ori} + \mathcal{L}_{mmd}$	$7.06^\circ \pm 0.11^\circ$	$7.19^\circ \pm 0.15^\circ$	$7.54^\circ \pm 0.07^\circ$	$9.53^\circ \pm 0.30^\circ$	7.83° ▼8.85%
$\mathcal{L}_{ori} + \mathcal{L}_{aug} + \mathcal{L}_{con}$	$6.81^\circ \pm 0.13^\circ$	$6.85^\circ \pm 0.26^\circ$	$7.07^\circ \pm 0.05^\circ$	$8.34^\circ \pm 0.15^\circ$	7.27° ▼15.37%
$\mathcal{L}_{ori} + \mathcal{L}_{aug} + \mathcal{L}_{mmd}$	$6.64^\circ \pm 0.14^\circ$	$6.76^\circ \pm 0.11^\circ$	$7.39^\circ \pm 0.06^\circ$	$8.66^\circ \pm 0.34^\circ$	7.36° ▼14.32%
$\mathcal{L}_{ori} + \mathcal{L}_{con} + \mathcal{L}_{mmd}$	$7.04^\circ \pm 0.11^\circ$	$7.09^\circ \pm 0.10^\circ$	$7.37^\circ \pm 0.06^\circ$	$9.41^\circ \pm 0.26^\circ$	7.73° ▼10.01%
$\mathcal{L}_{ori} + \mathcal{L}_{aug} + \mathcal{L}_{con} + \mathcal{L}_{mmd}$	$6.35^\circ \pm 0.10^\circ$	$6.72^\circ \pm 0.12^\circ$	$6.96^\circ \pm 0.07^\circ$	$8.79^\circ \pm 0.25^\circ$	7.21° ▼16.06%

TABLE II
RESULT OF COMPARISON WITH STATE-OF-THE-ART GAZE ESTIMATION AND DOMAIN GENERALIZATION METHODS

Method	$\mathcal{D}_E \rightarrow \mathcal{D}_M$	$\mathcal{D}_E \rightarrow \mathcal{D}_D$	$\mathcal{D}_G \rightarrow \mathcal{D}_M$	$\mathcal{D}_G \rightarrow \mathcal{D}_D$
Full-Face [16]	12.35°	30.15°	11.13°	14.42°
CA-Net [26]	-	-	27.13°	31.41°
RT-Genie [12]	-	-	21.81°	38.60°
Dilated-Net [27]	-	-	18.45°	23.88°
ADL* [13]	7.23°	8.02°	11.36°	11.86°
PureGaze [22]	7.08°	7.48°	9.28°	9.32°
Gaze-Consistent [23]	6.50°	7.44°	7.55°	9.03°
Baseline	8.02°	9.11°	8.04°	9.20°
Gaze-BAR	6.35°	6.72°	6.96°	8.79°

E. State-of-the-Art Comparison

In this section, we compare the performance of our proposed domain generalization method against SOTA gaze estimation methods. Table II presents the angular error results for various methods over four cross-dataset tasks.

Our method consistently outperforms existing SOTA methods in all cross-dataset tasks, achieving the lowest angular error for every cross-domain task. This underscores the advanced generalization ability of our approaches to handle variations in environmental factors and facial attributes.

F. Plug Into Existing Gaze Estimation Models

To substantiate the portability of our approach, we complement our analysis by integrating existing gaze estimation models into our comparative evaluation framework. Table III showcases the performance enhancements achieved by plugging our proposed method, BAR (Branch-out Auxiliary Regularization), into existing gaze estimation models over four cross-dataset tasks.

TABLE III
PERFORMANCE OF OUR BRANCH-OUT AUXILIARY REGULARIZATION PLUGGED TO EXISTING MODELS

Method	$\mathcal{D}_E \rightarrow \mathcal{D}_M$	$\mathcal{D}_E \rightarrow \mathcal{D}_D$	$\mathcal{D}_G \rightarrow \mathcal{D}_M$	$\mathcal{D}_G \rightarrow \mathcal{D}_D$
Full-Face [16]	12.35°	30.15°	11.13°	14.42°
Full-Face [16] + BAR	9.04°	9.45°	8.70°	11.33°
Dilated-Net [27]	-	-	18.45°	23.88°
Dilated-Net [27] + BAR	-	-	14.22°	16.55°
Baseline	8.02°	9.11°	8.04°	9.20°
Baseline + BAR	6.35°	6.72°	6.96°	8.79°

Our results indicate consistent improvements across all evaluated methods when integrated with the BAR framework. Notably, the Full-Face and Dilated-Net models witness a significant reduction in angular error, showcasing the efficacy of our approach in enhancing the generalization performance of existing gaze estimation architectures.

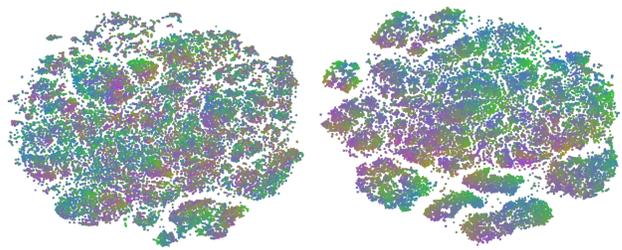


Fig. 3. t-SNE visualization of the extracted features. Similar colors represent similar gaze directions. Each dot denotes a sample from the test set.

G. Visualization of Gaze Feature

To compare and analyze the features acquired by the baseline model and our model, we employ t-SNE [32] to visualize the feature distribution of task $\mathcal{D}_E \rightarrow \mathcal{D}_M$. The results are depicted in Fig. 3, where similar colors represent similar gaze directions. The feature distribution of the baseline model demonstrates that there is no discernible correlation between gaze features and gaze direction. However, in our methodology, a noticeable transition in color from green to purple is observed from the upper right to the lower left corner, validating our model’s capability to extract gaze representations consistent with the gaze direction.

IV. CONCLUSION

In this letter, we introduce branch-out auxiliary regularization, a straightforward and flexible approach to enhance domain generalization in gaze estimation models without requiring direct access to target domain data. Our method leverages data augmentation to address environmental variations and incorporates a contrast branch for aligning gaze directions with positive samples from the source domain, thereby learning consistent gaze features. Comprehensive experimental evaluations on four cross-dataset tasks demonstrate the superiority of our approach.

REFERENCES

- [1] C. Liu, Y. Chen, L. Tai, H. Ye, M. Liu, and B. E. Shi, "A gaze model improves autonomous driving," in *Proceedings of the 11th ACM symposium on eye tracking research & applications*, 2019, pp. 1–5.
- [2] S. M. Shah, Z. Sun, K. Zaman, A. Hussain, M. Shoaib, and L. Pei, "A driver gaze estimation method based on deep learning," *Sensors*, vol. 22, no. 10, p. 3959, 2022.
- [3] A. Patney, J. Kim, M. Salvi, A. Kaplanyan, C. Wyman, N. Benty, A. Lefohn, and D. Luebke, "Perceptually-based foveated virtual reality," in *ACM SIGGRAPH 2016 emerging technologies*, 2016, pp. 1–2.
- [4] K. Rayner, "Eye movements in reading and information processing: 20 years of research," *Psychological bulletin*, vol. 124, no. 3, p. 372, 1998.
- [5] D. H. Yoo and M. J. Chung, "A novel non-intrusive eye gaze estimation using cross-ratio under large head motion," *Computer Vision and Image Understanding*, vol. 98, no. 1, pp. 25–51, 2005.
- [6] Z. Zhu and Q. Ji, "Eye gaze tracking under natural head movements," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1. IEEE, 2005, pp. 918–923.
- [7] Z. Zhu, Q. Ji, and K. P. Bennett, "Nonlinear eye gaze mapping function estimation via support vector regression," in *18th International Conference on Pattern Recognition (ICPR'06)*, vol. 1. IEEE, 2006, pp. 1132–1135.
- [8] C. Hennessey, B. Noureddin, and P. Lawrence, "A single camera eye-gaze tracking system with free head motion," in *Proceedings of the 2006 symposium on Eye tracking research & applications*, 2006, pp. 87–94.
- [9] J. Chen and Q. Ji, "3d gaze estimation with a single camera without ir illumination," in *2008 19th international conference on pattern recognition*. IEEE, 2008, pp. 1–4.
- [10] R. Valenti, N. Sebe, and T. Gevers, "Combining head pose and eye location information for gaze estimation," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 802–815, 2011.
- [11] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, "Appearance-based gaze estimation in the wild," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 4511–4520.
- [12] T. Fischer, H. J. Chang, and Y. Demiris, "Rt-gene: Real-time eye gaze estimation in natural environments," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 334–352.
- [13] P. Kellnhofer, A. Recasens, S. Stent, W. Matusik, and A. Torralba, "Gaze360: Physically unconstrained gaze estimation in the wild," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 6912–6921.
- [14] Y. Guan, Z. Chen, W. Zeng, Z. Cao, and Y. Xiao, "End-to-end video gaze estimation via capturing head-face-eye spatial-temporal interaction context," *IEEE Signal Processing Letters*, vol. 30, pp. 1687–1691, 2023.
- [15] Y. Cheng, X. Zhang, F. Lu, and Y. Sato, "Gaze estimation by exploring two-eye asymmetry," *IEEE Transactions on Image Processing*, vol. 29, pp. 5259–5272, 2020.
- [16] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, "It's written all over your face: Full-face appearance-based gaze estimation," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 51–60.
- [17] K. Wang, R. Zhao, H. Su, and Q. Ji, "Generalizing eye tracking with bayesian adversarial learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 11907–11916.
- [18] X. Cai, J. Zeng, S. Shan, and X. Chen, "Source-free adaptive gaze estimation by uncertainty reduction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 22035–22045.
- [19] H. Zhang, X. Wang, W. Ren, R. Lin, and H. Liu, "Outlier constrained unsupervised domain adaptation algorithm for gaze estimation," in *International Conference on Intelligent Robotics and Applications*. Springer, 2022, pp. 349–359.
- [20] Y. Wang, Y. Jiang, J. Li, B. Ni, W. Dai, C. Li, H. Xiong, and T. Li, "Contrastive regression for domain adaptation on gaze estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 19376–19385.
- [21] R. Liu, Y. Liu, H. Wang, and F. Lu, "Pnp-ga+: Plug-and-play domain adaptation for gaze estimation using model variants," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.
- [22] Y. Cheng, Y. Bao, and F. Lu, "Puregaze: Purifying gaze feature for generalizable gaze estimation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 1, 2022, pp. 436–443.
- [23] M. Xu, H. Wang, and F. Lu, "Learning a generalized gaze estimator from gaze-consistent feature," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 37, no. 3, 2023, pp. 3027–3035.
- [24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [25] M. Dolatshah, A. Hadian, and B. Minaei-Bidgoli, "Ball*-tree: Efficient spatial indexing for constrained nearest-neighbor search in metric spaces," *arXiv preprint arXiv:1511.00628*, 2015.
- [26] Y. Cheng, S. Huang, F. Wang, C. Qian, and F. Lu, "A coarse-to-fine adaptive network for appearance-based gaze estimation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, 2020, pp. 10623–10630.
- [27] Z. Chen and B. E. Shi, "Appearance-based gaze estimation using dilated-convolutions," in *Asian Conference on Computer Vision*. Springer, 2018, pp. 309–324.
- [28] X. Zhang, S. Park, T. Beeler, D. Bradley, S. Tang, and O. Hilliges, "Eth-xgaze: A large scale dataset for gaze estimation under extreme head pose and gaze variation," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16*. Springer, 2020, pp. 365–381.
- [29] 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*, vol. 41, no. 1, pp. 162–175, 2017.
- [30] K. A. Funes Mora, F. Monay, and J.-M. 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*, 2014, pp. 255–258.
- [31] Y. Sugano, Y. Matsushita, and Y. Sato, "Learning-by-synthesis for appearance-based 3d gaze estimation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1821–1828.
- [32] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne." *Journal of machine learning research*, vol. 9, no. 11, 2008.