See Through the Noise: Improving Domain Generalization in Gaze Estimation

Supplementary Material

In the supplementary material, we provide detailed description of our proposed SeeTN and more experiments results. We first present the training algorithm in Algorithm 1. Then we show additional quantitative experiments, including detailed results on synthetic noisy dataset, fine-tuned results on target domain and more hyperparameter analysis. Finally, we give more visualized results, including feature distribution of unseen domain and noisy sample detection. These results clearly verify the effectiveness of our proposed SeeTN approach.

1. SeeTN Algorithm

The full algorithm for implementing SeeTN is shown in Algorithm 1.

Algorithm 1 SeeTN

Input: Network $F(\cdot, \theta_f)$, Regressor $G(\cdot, \theta_g)$, Dataset $\mathcal{D}_\mathcal{S}$, Prototypes μ.
$\theta_f, \theta_g = \mathrm{WarmUp}(\mathrm{G}(\mathrm{F}(\cdot, \theta_\mathrm{f}), \theta_\mathrm{g}))$
Get initial $\mathcal{D}_\mathcal{S}^\mathcal{C}$ and $\mathcal{D}_\mathcal{S}^\mathcal{N}$ by η
while $t <$ MaxEpoch **do**
  for item $= 1$ **to** num_iters **do**
    From $\mathcal{D}_\mathcal{S}^\mathcal{C}$, draw a mini-batch$\{(x_i^\mathcal{C}, y_i^\mathcal{C}), i = 1...B_C\}$
    From $\mathcal{D}_\mathcal{S}^\mathcal{N}$, draw a mini-batch$\{(x_i^\mathcal{N}, y_i^\mathcal{N}), i = 1...B_N\}$
    Get $f_i^\mathcal{C}, f_i^\mathcal{N}, g_i^\mathcal{C}, g_i^\mathcal{N}, z_i^\mathcal{C}, z_i^\mathcal{N}$
    Update μ by Eqs. (2) and (3) leveraging $z_i^\mathcal{C}$
    Get p_i by Eq. (4) leveraging $z_i^\mathcal{C}$ and $z_i^\mathcal{N}$
    Get $A_{i,j}^{(m)}$ by Eq. (5) leveraging p_i
    Get $A_{i,j}^{(g)}$ by Eq. (6) leveraging $y_i^\mathcal{C}$ and $y_i^\mathcal{N}$
    $\mathcal{L}_\mathrm{gaze} = \frac{1}{B_C} \sum |g_i^\mathcal{C} - y_i^\mathcal{C}|$
    $\mathcal{L}_\mathrm{align}^\mathcal{C} = \frac{1}{B_C(B_C-1)} \sum_{x_i^\mathcal{C}} \sum_{x_j^\mathcal{C}, i \neq j} |A_{i,j}^{(g)} - A_{i,j}^{(m)}|$
    Get $A_{i,j}^{(f)}$ by Eq. (10) leveraging $f_i^\mathcal{C}$ and $f_i^\mathcal{N}$
    $\mathcal{L}_\mathrm{align}^\mathcal{N} = \frac{1}{B_N} \sum_{x_i^\mathcal{N}} \frac{A_{i,:}^{(f)} \cdot A_{i,:}^{(m)}}{\|A_{i,:}^{(f)}\| \|A_{i,:}^{(m)}\|}$
    $\mathcal{L}_\mathrm{all} = \mathcal{L}_\mathrm{gaze} + \mathcal{L}_\mathrm{align}^\mathcal{C} + \lambda \mathcal{L}_\mathrm{align}^\mathcal{N}$
    $\theta_f, \theta_g = \mathrm{Adam}(\mathcal{L}_\mathrm{all}, \theta_\mathrm{f}, \theta_\mathrm{g})$
  end for
  Calculate η by Eq. (7) leveraging x_i and y_i
  Repartition $\mathcal{D}_\mathcal{S}^\mathcal{C}$ and $\mathcal{D}_\mathcal{S}^\mathcal{N}$ by η
end while

2. Additional Quantitative Experiments

2.1. Experiments on the Synthetic Noisy Dataset

First, we provide an explanation of the rationale underlying the synthetic noisy dataset design. In tasks involving noisy labels, simultaneously obtaining both noisy labels and truly accurate annotations from real-world datasets is often challenging, thereby limiting the comprehensive evaluation of methods. As a result, researchers typically conduct controlled experiments on artificially constructed noisy datasets. By examining the noisy labels in the Gaze360, as shown in Fig. 1 of the main paper, we found that their deviations from the true gaze directions are relatively large, and some of the directions are even completely opposite. Thus, we added Gaussian noise with a standard deviation of 60° to randomly selected samples. According to the 3σ principle of the Gaussian distribution, this setting can approximately cover noisy labels with different error levels. Moreover, based on the observations from previous work [3, 5–7, 10] that real-world data typically contain about 8%–38.5% label noise, we introduced noise rates of 10%, 20%, and 30% into the Gaze360 dataset to simulate similar conditions.

In the main paper, we have reported the results of the baseline and SeeTN on the synthetic noisy Gaze360 dataset in the form of line charts. In the supplementary material, we provide the detailed numerical results of our experiments and report the additional results of the DivideMix [4] and SUGE [8] methods.

As shown in Tab. 1, the performance of the baseline decreases as the noise ratio increases. Besides, while DivideMix underperforms on the raw Gaze360 dataset, it remains competitive against the baseline under higher noise ratios. These results suggest that although DivideMix is inherently incompatible with regression tasks, it still performs effectively under relatively high noise ratios. SUGE is specifically designed to address noisy samples in gaze estimation tasks, and it demonstrates good performance improvements across different noise ratio. However, SUGE shows a considerable performance drop when the noise ratio reaches 30%. Our proposed SeeTN consistently achieves superior performance across all settings, and unlike SUGE, it maintains strong robustness even under a 30% noise ratio.

In summary, the aforementioned results emphasize the significance of noise mitigation in enhancing domain generalization, and further validate the effectiveness and robustness of SeeTN.

Table 1. Comparison of domain generalization performance between SeeTN and representative noisy label learning methods in synthetic noisy Gaze360 dataset.

Method	Noise Ratio=0%		Noise Ratio=10%		Noise Ratio=20%		Noise Ratio=30%	
	$\mathcal{D}_G \to \mathcal{D}_M$	$\mathcal{D}_G \to \mathcal{D}_D$	$\mathcal{D}_G \to \mathcal{D}_M$	$\mathcal{D}_G \to \mathcal{D}_D$	$\mathcal{D}_G \to \mathcal{D}_M$	$\mathcal{D}_G \to \mathcal{D}_D$	$\mathcal{D}_G \to \mathcal{D}_M$	$\mathcal{D}_G \to \mathcal{D}_D$
Baseline	7.94	8.73	8.69	10.93	12.17	14.19	14.38	18.99
DivideMix	10.46	12.3	10.32	11.27	10.05	10.13	9.83	11.32
SUGE	7.04	8.32	7.43	8.61	7.40	9.12	9.23	11.37
SeeTN	**6.57**	**7.57**	**7.39**	**8.32**	**7.31**	**7.82**	**7.76**	**8.97**

Table 2. Comparison of domain generalization performance between SeeTN and AGG in a validation set with synthetic noise.

Method	Noise Ratio=0%		Noise Ratio=10%		Noise Ratio=20%		Noise Ratio=30%	
	$\mathcal{D}_E \to \mathcal{D}_M$	$\mathcal{D}_E \to \mathcal{D}_D$	$\mathcal{D}_E \to \mathcal{D}_M$	$\mathcal{D}_E \to \mathcal{D}_D$	$\mathcal{D}_E \to \mathcal{D}_M$	$\mathcal{D}_E \to \mathcal{D}_D$	$\mathcal{D}_E \to \mathcal{D}_M$	$\mathcal{D}_E \to \mathcal{D}_D$
Baseline	10.25	11.04	13.02	13.53	14.23	15.01	15.33	16.72
AGG	9.59	**8.84**	10.52	10.41	12.19	13.21	13.96	18.06
SeeTN	**8.57**	9.37	**9.10**	**9.82**	**8.65**	**9.68**	**8.91**	**10.05**

Table 3. Comparison of fine-tuned results between SeeTN and state-of-the-art domain generalization methods.

Method	$\mathcal{D}_E \to \mathcal{D}_M$	$\mathcal{D}_E \to \mathcal{D}_D$	$\mathcal{D}_G \to \mathcal{D}_M$	$\mathcal{D}_G \to \mathcal{D}_D$
PureGaze	5.30	6.42	5.20	7.36
GazeCF	4.76	5.43	5.34	5.66
CGaG	**4.59**	**5.07**	5.13	5.71
SeeTN	5.04	5.23	**4.66**	**5.52**

Table 4. The ablation of the key components in SeeTN.

Ablation	$\mathcal{D}_E \to \mathcal{D}_M$	$\mathcal{D}_E \to \mathcal{D}_D$	$\mathcal{D}_G \to \mathcal{D}_M$	$\mathcal{D}_G \to \mathcal{D}_D$
SeeTN	**6.58**	**7.18**	**6.57**	**7.57**
w/o prototypes	7.88	8.14	7.64	8.37
w/ L2 distance	7.01	8.15	7.58	8.17

Table 5. The ablation of hyperparameter λ.

Params.	Value	$\mathcal{D}_E \to \mathcal{D}_M$	$\mathcal{D}_E \to \mathcal{D}_D$	$\mathcal{D}_G \to \mathcal{D}_M$	$\mathcal{D}_G \to \mathcal{D}_D$
λ	0.01	6.88	7.41	6.93	7.74
	0.1	**6.58**	**7.18**	**6.57**	**7.57**
	1	7.02	7.66	6.85	7.70

2.2. Effective of Noise Handling

To further verify that the performance improvement of our method mainly stems from its explicit modeling of noise, we train SeeTN and AGG [1] on a clean validation set and inject 10–30% synthetic label noise. The results are shown in Tab. 2. While both SeeTN and AGG perform similarly under clean labels, SeeTN degrades much more slowly as noise increases, confirming that its gains arise from explicit noise modeling rather than domain generalization alone.

2.3. Fine-tuned Results on the Target Domain

In Tab. 3, we compare SeeTN with several previous works [2, 9, 11] by fine-tuning on 100 randomly selected samples from the target domain. SeeTN achieves better performance on $\mathcal{D}_G \to \mathcal{D}_M$ and $\mathcal{D}_G \to \mathcal{D}_D$.

2.4. More Ablation

To validate the effectiveness of the key components in SeeTN, we construct two simple yet direct ablation studies. The first investigates the effectiveness of the prototype mechanism. Specifically, we remove the prototype module from SeeTN, and all prototype-related computations are directly replaced by the MLP-transformed features. The second examines the choice of similarity metric, where cosine similarity is replaced with the L2 distance. The results are shown in Tab. 4. We observe a substantial performance drop under all cross-domain settings, demonstrating the effectiveness of the proposed SeeTN design.

2.5. Hyperparameter Analysis

In the Tab. 5 of the main paper, we report the ablation experiments of key parameters: K and t. In the supplementary materials, we provide ablation experiments about the parameter λ which controls the weight of $\mathcal{L}_{\text{align}}^{\mathcal{N}}$, as shown in Tab. 5. The performance remains stable with different values of λ. Thus, we choose $\lambda = 0.1$ in the SeeTN.

(a) Visualization of Baseline **(b) Visualization of SeeTN**

Figure 1. The qualitative results of visualizing the features learned by the backbone and SeeTN on the MPIIGaze dataset via t-SNE. In (a), samples with different gaze directions cluster together in the purple-shaded region, while in (b), samples with similar gaze directions form compact clusters. For instance, the four labeled samples, which share similar gaze directions, are dispersed in (a) but closely grouped in (b).

3. Additional Visualization Results

3.1. Visualization of Unseen Domain

In tha main paper, we have shown the feature visualization of our SeeTN in source domain. To further demonstrate the generalizable abilities, we provide the t-SNE visualization results of baseline and SeeTN on the unseen domain MPIIGaze, as illustrated in Fig. 1. It can be observed that the baseline, shown in Fig. 1(a), sometimes clusters samples with different gaze directions together, whereas SeeTN, shown in Fig. 1(b), is more likely to group samples with similar gaze directions. For instance, in the purple-shaded region of (a), samples with different gaze directions cluster together using baseline model, while in (b), our method can enforce the samples with similar gaze directions form compact clusters. Moreover, the four labeled samples, which share similar gaze directions, are dispersed in (a) but closely grouped in (b).

3.2. Visualization of Noisy Sample Detection

We present additional visualizations of noisy samples selected by our proposed indicator η, as shown in Fig. 2(a) and Fig. 2(b). We can see that the ground-truth labels of our selected samples are largely incorrect.

Besides, we also visualize the unseen-domain samples with large test errors for SeeTN in Fig. 3. In fact, a considerable number of noisy labels are also present in the unseen domain, while SeeTN can correctly predict their truly gaze directions, which demonstrates the widespread presence of noise in gaze estimation tasks. Meanwhile, the generalization ability of gaze estimation models across domains also requires further investigation.

References

[1] Yiwei Bao and Feng Lu. From feature to gaze: A generalizable replacement of linear layer for gaze estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1409–1418, 2024. 2

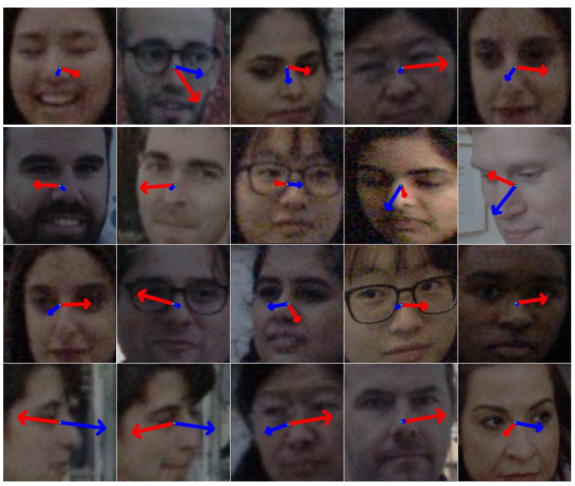
[2] Yihua Cheng, Yiwei Bao, and Feng Lu. Puregaze: Purifying gaze feature for generalizable gaze estimation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 436–443, 2022. 2

[3] Kuang-Huei Lee, Xiaodong He, Lei Zhang, and Linjun Yang. Cleannet: Transfer learning for scalable image classifier training with label noise. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5447–5456, 2018. 1

[4] Junnan Li, Richard Socher, and Steven CH Hoi. Dividemix: Learning with noisy labels as semi-supervised learning. *arXiv preprint arXiv:2002.07394*, 2020. 1

[5] Wen Li, Limin Wang, Wei Li, Eirikur Agustsson, and Luc Van Gool. Webvision database: Visual learning and understanding from web data. *arXiv preprint arXiv:1708.02862*, 2017. 1

[6] Hwanjun Song, Minseok Kim, and Jae-Gil Lee. Selfie: Refurbishing unclean samples for robust deep learning. In *In-*



Gaze360

(a) Noisy Samples Selected from Gaze360

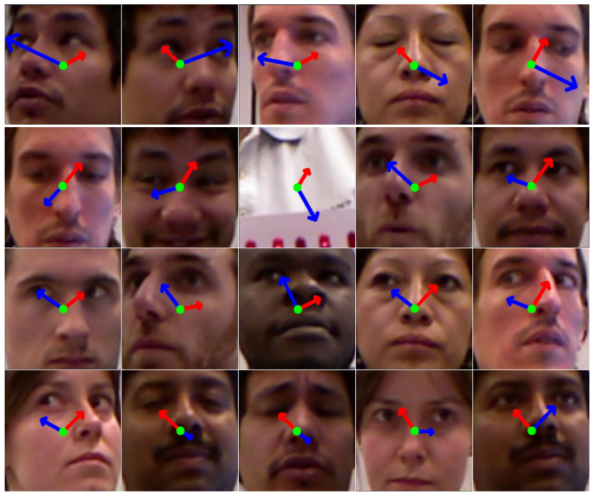
ETH-XGaze

(b) Noisy Samples Selected from ETH-XGaze

Figure 2. Noisy Sample Visualization in Gaze360 and ETH-XGaze. The red and blue arrows indicate the ground-truth directions and the model predictions respectively. The samples selected by the our designed indicator show clear evidence of label noise.

MPIIGaze

(a) Noisy Samples from MPIIGaze



EyeDiap

(b) Noisy Samples from EyeDiap

Figure 3. Noisy Sample Visualization in MPIIGaze and EyeDiap. The red and blue arrows indicate the ground-truth directions and the model predictions respectively. It can be observed that some of samples with incorrect labels are correctly predicted by our SeeTN model in the unseen domain.

ternational conference on machine learning, pages 5907–5915. PMLR, 2019.

[7] Hwanjun Song, Minseok Kim, Dongmin Park, Yooju Shin, and Jae-Gil Lee. Learning from noisy labels with deep neural networks: A survey. *IEEE transactions on neural networks and learning systems*, 34(11):8135–8153, 2022. 1

[8] Shijing Wang and Yaping Huang. Suppressing uncertainty in gaze estimation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 5581–5589, 2024. 1

[9] Lifan Xia, Yong Li, Xin Cai, Zhen Cui, Chunyan Xu, and Antoni B Chan. Collaborative contrastive learning for cross-domain gaze estimation. *Pattern Recognition*, 161:111244, 2025. 2

[10] Tong Xiao, Tian Xia, Yi Yang, Chang Huang, and Xiaogang Wang. Learning from massive noisy labeled data for image classification. In *Proceedings of the IEEE conference on*

computer vision and pattern recognition, pages 2691–2699, 2015. 1

[11] Mingjie Xu, Haofei Wang, and Feng Lu. Learning a generalized gaze estimator from gaze-consistent feature. In *Proceedings of the AAAI conference on artificial intelligence*, pages 3027–3035, 2023. 2