Generalizing Gaze Estimation with Rotation Consistency

Yiwei Bao¹ Yunfei Liu¹ Haofei Wang² Feng Lu¹, ² *
¹ State Key Laboratory of VR Technology and Systems, School of CSE, Beihang University
² Peng Cheng Laboratory, Shenzhen, China
{baoyiwei, lyunfei, lufeng}@buaa.edu.cn wanghf@pcl.ac.cn

Abstract

Recent advances of deep learning-based approaches have achieved remarkable performance on appearance-based gaze estimation. However, due to the shortage of target domain data and absence of target labels, generalizing gaze estimation algorithm to unseen environments is still challenging. In this paper, we discover the rotation-consistency property in gaze estimation and introduce the ‘sub-label’ for unsupervised domain adaptation. Consequently, we propose the Rotation-enhanced Unsupervised Domain Adaptation (RUDA) for gaze estimation. First, we rotate the original images with different angles for training. Then we conduct domain adaptation under the constraint of rotation consistency. The target domain images are assigned with sub-labels, derived from relative rotation angles rather than untouchable real labels. With such sub-labels, we propose a novel distribution loss that facilitates the domain adaptation. We evaluate the RUDA framework on four cross-domain gaze estimation tasks. Experimental results demonstrate that it improves the performance over the baselines with gains ranging from 12.2% to 30.5%. Our framework has the potential to be used in other computer vision tasks with physical constraints.

1. Introduction

Gaze is one of the most important cues for human intention prediction. It has been used in a variety of applications such as virtual/augmented reality [21,30], human-computer interaction [18,35,37], and medical analysis [3,20]. To obtain accurate gaze estimations, various systems have been developed. Appearance-based gaze estimation is one of the most promising approaches, since it has the lowest hardware requirements.

With the advancement of deep learning techniques, Convolutional Neural Networks (CNN) have achieved significant performance improvement in many computer vision tasks. Gaze estimation task is no exception, various CNN-based gaze estimation methods have been proposed over the last decades [8]. These systems usually have different inputs: eye images [9,24,29,39,42], face images [19,22,43] or both face/eye images [1,7,23]. However, existing methods suffer from severe performance degradation when adapting to new domains, which is mainly caused by the difference between the domains, e.g., subject appearance, image quality, shooting angle and illumination.

One of the major challenge of gaze domain adaptation is that we usually do not have access to target domain labels in real world scenarios, and we cannot directly train the gaze estimator in target domain. To address this problem, unsupervised domain adaptation approaches aim to find a gaze-relevant constraint generalizing the model to target domain without label. Kellnhofer et al. propose to supervise gaze estimation model with an domain discriminator by adversarial learning [19]. Similarly, Wang et al. employ an appearance discriminator and a head pose classifier for adaptation [39]. More recently, Liu et al. propose to guide the model with outliers [25]. Although some unsupervised domain adaptation approaches for gaze estimation have been proposed, it is still a challenging task.
To build a gaze-relevant constraint to supervise the model without requiring ground truth labels, we dive into the physical nature of gaze. We find that the human gaze, as a 3D direction vector, is rotation-consistent. Rotating the face image results in the same rotation angle of the gaze direction, we call this rotation-consistency property. And we define the relative rotation angle as the sub-label, meaning that it is not an absolute angle, but the relative difference angle before and after the rotation. This rotation consistency could serve as the desired gaze-relevant constraint without ground truth. Although training with rotated images in source domain does not improve gaze estimation accuracy because user faces are already aligned by normalization [42], we argue that the rotation consistency property provides a gaze-relevant optimization target for adaptation.

In light of this, we present the Rotation-enhanced Unsupervised Domain Adaptation (RUDA) framework for gaze estimation. Our approach creates sub-labels between original and randomly rotated images. The estimator is generalized to target domain via rotation consistency of estimation results with no target domain label required and low computation cost. The contributions of this work are as follow:

- We propose the Rotation-enhanced Unsupervised Domain Adaptation (RUDA) framework for gaze estimation. The RUDA first trains a rotation-augmented model in source domain, then adapts the model to target domain using the synthesized images with physically-constrained gaze directions.
- We found the rotation consistency property, which can be used to generate sub-labels for unsupervised gaze adaptation tasks. To facilitate adaptation, we design a novel distribution loss which supervise the model with rotation consistency and sub-labels.
- Experimental results demonstrate that the RUDA framework achieves consistent improvement over the baseline model on four cross-domain gaze estimation tasks, ranging from 12.2% to 30.5%. It achieves surprisingly good results, even outperforms some state-of-the-art methods trained on target domain with labels.

2. Related work

Gaze Estimation. Early studies estimate gaze by reconstructing a 3D eyeball model and calculate gaze from the anatomical eye structure. These methods usually offer accurate gaze estimates, while they require personal calibration and dedicated devices such as depth camera [34, 38, 40], infrared camera [28] and infrared lights [15].

Calibration-free appearance-based gaze estimation with single web camera received favor of researchers in the last decades. In 2015, Zhang et al. first propose to estimate gaze from eye images using CNN [42]. Following this work, a number of gaze estimation dataset have been released [10, 19, 23, 31, 33, 41, 43]. Based on them, various deep learning-based approaches using different inputs have been proposed: using eye images [9, 24, 29, 39, 42], using face images [19, 22, 43] or using both [1, 7, 23].

More recently, cross domain gaze estimation task attracted more and more attention. Park et al. proposed to learn a person-specific gaze estimation network with few samples by meta-learning [29]. Guo et al. eliminated the inter-personal diversity by ensuring prediction consistency [16]. Cheng et al. proposed to improve cross dataset accuracy without target domain data by eliminating gaze-irrelevant feature [6]. Liu et al. [25] proposed a plug-and-play cross-domain gaze estimation framework with the guidance of outliers. Although it significantly outperforms the existing methods, their method requires as many as 20 models for collaborative learning. Zheng et al. [45] propose to redirect head and gaze in a self-supervised manner by embedding transformation including rotation, which helps down stream tasks like gaze estimation. In other tasks like 3D hand pose estimation, rotation has also been used as a constrain for self-supervised learning [32].

Unsupervised Domain Adaptation. Unsupervised domain adaptation (UDA) is one of the common tasks in computer vision, which has been extensively studied for a long time. Early UDA methods use geodesic distance as the subspace distance to learn domain-invariant representations [12, 14]. Inspired by this, some researchers proposed to reduce domain gap by matching the statistics of source and target domain [2, 26]. Chen et al. propose a representation subspace distance (RSD) that aligns features from two domains specifically for regression tasks [4].

Inspired by the generative adversarial net [13], adversarial learning have been adopted for UDA tasks. For example, a min-max game between feature extractor and domain discriminator is built to close the domain gap [27, 36, 44].

Although the above-mentioned methods achieve considerable improvement, most of them are designed for classification tasks, instead of regression task. The RSD proposed by Chen et al. [4] is specifically designed for regression tasks, however, we found that their approach dose not perform well on gaze estimation task. Therefore, the UDA for gaze estimation still remains to be explored.

3. Rotation Consistency in Gaze Estimation

There are two main challenges in unsupervised gaze adaptation tasks: 1) the shortage of target domain samples for adaptation, and 2) the absence of ground-truth labels in target domain. Various data augmentation approaches have been proposed to generate training data in source domain, e.g., color jittering, introducing noise, flipping, translation and rotation. However, existing data augmentation
Why rotation consistency? Rotation is a commonly-used data augmentation approach in computer vision. However, in gaze estimation tasks, training with rotated images brings little performance gain in both within- and cross-dataset tasks. In fact, it is more often used for data normalization: by rotating and scaling the virtual camera, the user’s face is changed to the same size and location, while the x-axis of the camera coordinate system and the user head coordinate system are aligned and the z-axis of the camera coordinate system is perpendicular to the image plane. As a result, rotation operations help the camera look at different faces in a unified way (top-left in Fig. 2).

On the other hand, our proposed rotation consistency-based strategy plays a different role. It aims at solving the shortage of target domain data and absence of target label problem in cross-domain gaze estimation, and it indeed boosts the performance. Fig. 2 illustrates the idea of rotation consistency. It bridges the relative rotation angles between the image and the 3D gaze. In this way, for unsupervised domain adaptation, although the real gaze directions are unknown, the relative rotation angles can serve as the sub-labels to train the network. In addition, we can generate as many target images as we want with different sub-labels if we rotate the image with different angles.

**Conversion between the image and gaze rotation angle.**

Given a normalized image \( I \), we use the center of image as rotation center \( O \), and rotate the image with \( \theta \) (clockwise), the rotation matrix \( R \) can be defined as follows:

\[
R = \begin{bmatrix}
\cos \theta & -\sin \theta \\
\sin \theta & \cos \theta
\end{bmatrix}.
\]  

(2)

For each pixel location \( I_i \in I \), the rotated pixel location is \( RI_i \). Gaze is a 3D direction vector \( g \) defined in camera coordinate system. Therefore, the corresponding rotation matrix for gaze direction is

\[
R^g = \begin{bmatrix}
R & 0 \\
0 & 1
\end{bmatrix}.
\]  

(3)

As a result, the rotated gaze direction is \( R^g g^T \). In actual training, the gaze direction is denoted as a 2D Euler angles \( g = [y, p] \), where \( y \) is the yaw angle and \( p \) is the pitch angle. Thus, conversion between 2D Euler angles and 3D direction vector is needed before and after rotation.

**4. Method**

**4.1. Task Definition**

For UDA tasks, we are given a fully labeled source domain and a small amount of unlabeled samples from target domain. Let \( D_s = \{I_s^i, g_s^i\}_{i=1}^{N_s} \) represents \( N_s \) images with gaze label \( g_s \) in source domain, and \( D_t = \{I_t^i\}_{i=1}^{N_t} \) represents \( N_t \) images without gaze labels in target domain. Our goal is to generalize a gaze estimation network \( F_\theta \) with parameter \( \theta \) that performs well in \( D_t \). Only a small subset of unlabeled target domain samples \( D_t' \) is used for adaptation. Before that, \( F_\theta \) is pretrained on \( D_s \). In the following, we will introduce the details of our proposed method.
4.2. Rotation-Enhanced Unsupervised Domain Adaptation for Gaze Estimation

Fig. 3 shows the overview of the proposed RUDA framework, which consists of two phases: 1) the Rotation-Augmented Training (RAT) phase and 2) the Consistency-Guided Domain Adaptation (CGDA) phase. We first train a rotation-augmented model to predict gaze on rotated images. Then, based on rotation consistency, sub-labels generated by image rotation are used to guide the SRM module and calculate the proposed Distribution Loss for unsupervised domain adaptation.

4.2.1 Rotation-Augmented Training

In order to adapt the model to target domain guided by rotation consistency property, the model should be able to estimate the gaze on rotated images. Thus, in RAT, we train a rotation augmented model in source domain.

Gaze estimators are usually trained with labeled source domain data \{I^s, g^s\} with \(L_1\) loss function:

\[
\arg \min_{\theta} \mathcal{L}_1(\hat{g}^s_t, g^s),
\]

where \(\hat{g}^s_t = F_\theta(I^s_t)\) is the estimation result. To predict gaze from rotated images, we also train the model \(F_\theta\) with rotation augmented source domain samples. For each image \(I^s\) in the training set of source domain, we randomly rotate it \(K\) times to obtain a new set \(I^s:\)

\[
I^s = \{RI^s|k = 1, 2, \ldots, K\}.
\]

Here we record the rotation matrix \(R\) as the sub-label for the rotated image set \(I^s\). A group of estimation results of rotated images \(I^s\) is denoted as \(\{\hat{g}^s\} = F_\theta(I^s)\).

To maintain stable estimation across different rotation angles, we train the model with our proposed distribution loss function \(L_D\) and \(L_1\) loss. In \(L_D\), the mean value of estimation results is supervised by gaze label \(g\) and the STD of \(\{\hat{g}\}\) is supervised by sub-label. We explain the details of \(L_D\) in Sec. 4.2.3.

In a nutshell, the RAT phase can be formalized as:

\[
\arg \min_{\theta} (\mathcal{L}_1(\hat{g}^s, g^s) + \mathcal{L}_D(\{R\}, \{\hat{g}^s\}, g^s)).
\]

4.2.2 Consistency-Guided Domain Adaptation

In RAT phase, we generalize the rotation-augmented model \(F_\theta\) to the target domain with the guidance of sub-label based on the rotation consistency property. We also introduce a temporal average model \(\bar{F}_\theta\), which produces pseudo labels to prevent the estimation collapse.

First, we obtain the sub-label by randomly rotating the unlabeled sample \(I^t \in D_t\) by \(K\) times:

\[
I^t = \{RI^t|k = 1, 2, \ldots, K\}.
\]

Ideally, the rotation angle between the estimations of \(I^t\) and the estimation of original image \(I^t\) should be equal to sub-labels \(\{R\}\) according to Eq. (1). \(F_\theta\) is supervised by rotation consistency with sub-label \(\{R\}\) instead of true label.
If rotation consistency is the only constraint applied to $F_\theta$, the estimation results collapse to the $z$-axis of camera coordinate system since it is the rotation axis. Inspired by [11], we introduce a temporal average model $\bar{F}_\theta$ that produces stable pseudo labels to avoid collapse.

At the beginning of CGDA phase, $\bar{F}_\theta$ is initialized as a copy of $F_\theta$. After training with $T$ iterations, parameters of temporal average model $\bar{\theta}$ are updated from $\theta$ by Exponential Moving Average (EMA) algorithm:

$$\bar{\theta}^T = \alpha \bar{\theta}^{T-1} + (1 - \alpha) \theta^T,$$

where $\alpha$ is a momentum coefficient. $\bar{F}_\theta$ first estimates the gazes from a group of rotated images. Then, we design a Sub-label-guided Rotate-back Module (SRM) to recover the estimations that corresponds to the original image. According to the rotation consistency property, we rotate the estimation results of rotated images with the inverse matrix of sub-label. The pseudo label $g'$ is defined as the mean direction of recovered gaze estimations:

$$g' = \text{Mean} (\{(R^g)^{-1} \bar{F}_\theta(I^t)\}).$$

The estimation from $\bar{F}_\theta$, i.e., pseudo label is much more stable than $F_\theta$ during adaptation [11], while still being capable of fine adjustment. $F_\theta$ is punished if the estimation deviates far from pseudo label because of collapse.

In CGDA phase, the model is also supervised by $L_D$, while the gaze label is replaced by pseudo label. The adaptation process is summarized as:

$$\begin{align*}
\arg \min_\theta L_D(\{R\}, \{g\}, g'), \\
\bar{\theta}^T = \alpha \bar{\theta}^{T-1} + (1 - \alpha) \theta.
\end{align*}$$

### 4.2.3 Distribution Loss Function

To supervise the model with rotation consistency and sub-label, we propose the Distribution Loss $L_D$, which consists of two terms $L_{\text{mean}}$ and $L_{\text{std}}$. $L_{\text{mean}}$ constrains the estimation to be accurate by gaze label in RAT phase and prevent the estimation from collapse in CGDA phase. $L_{\text{std}}$ constrains the estimations to be consistent with each other by the sub-label $R$. $L_D$ is defined as follow:

$$
L_D(\{R\}, \{g\}, g') = L_{\text{mean}} + L_{\text{std}},
$$

$$
L_{\text{mean}}(\{R\}, \{g\}, g) = \frac{1}{K} \sum_{k=1}^{K} L_1(g', g),
$$

$$
L_{\text{std}}(\{R\}, \{g\}) = \sqrt{\frac{\sum_{k=1}^{K} (g' - \{g\})^2}{K}},
$$

where $\{g'\}$ stands for a group of recovered estimation results by SRM module based on rotation consistency, $\{g\}$ stands for the average direction of $\{g\}$. $L_D$ treats a group of recovered gaze estimation as a distribution. $L_{\text{mean}}$ supervise the model by requiring every sample of the distribution to be equal to the desired mean value $g$. $L_{\text{std}}$ requires the standard deviation of the distribution to be 0, which is the proposed rotation consistency in Eq. (1). The whole procedure of RUDA framework is summarized in Algorithm 1.

### 4.3. Implementation Details

Our method is implemented using PyTorch framework. ResNet18 is used as backbone. $K$ is set to 5 in RAT phase and is set to 20 in CGDA phase. Momentum coefficient $\alpha$ in EMA algorithm is set to 0.99. Batch size is set to 80 and 10 during source domain training and domain adaptation phase respectively. We randomly chose 100 unlabeled images from target domain for adaptation. The model is trained for 10 epochs in source domain and for adaptation. We employed the Adam optimizer with a learning rate of $10^{-4}$ and $\beta = (0.5, 0.95)$.

### 5. Experiments

#### 5.1. Data Preparation

To verify the effectiveness of the RUDA framework, we conducted experiments on four commonly used gaze estimation datasets: ETH-XGaze ($D_E$) [41], Gaze360 ($D_G$) [19], MPIIFaceGaze ($D_M$) [43] and EyeDiap ($D_D$) [10].

- **ETH-XGaze**: ETH-XGaze dataset is collected under laboratory environment with high-resolution cameras. We
follow the original paper and take 750,000 face crops from 80 participants as training set.

- **Gaze360**: Gaze360 dataset is collected in arbitrary environment by a 360° camera. It has a wide distribution over the horizontal axis of gaze. We only use 84900 images with frontal faces.

- **MPIIFaceGaze**: MPIIFaceGaze is collected during daily usage of laptops. We chose 3000 images for 15 subjects respectively as the standard protocol suggests.

- **EyeDiap**: EyeDiap dataset is collected under laboratory environment with screen and floating targets. Note that due to the misalignment of timeline, some of the labels are not reliable. We selected 6400 sample images that are manually checked by original authors.

We perform gaze normalization proposed by [42] for all datasets except $D_G$, as it does not provide head pose labels. Rotations are performed after gaze normalization. Face images are cropped and resized to 224x224. We further normalize the image pixels to $[0, 1]$ as the final input. More details can be found in [8].

### 5.2. Performance of RUDA Framework

To test the performance of RUDA framework, we implement it based on two state-of-the-art backbone network: ResNet18 and ResNet50 [17]. We train the backbone network on source domain with $L_1$ loss as baseline. As shown in Tab. 1, RUDA framework improves the performance of both backbone network by a big margin. For ResNet18, RUDA framework improves the performance by 30.5%, 12.2%, 19.9% and 23.3% on four cross domain tasks, respectively. For ResNet50, RUDA framework brings 19.2%, 20.7%, 17.6% and 14.4% performance gain. Thanks to the reasonable design of RUDA framework and wide data distribution of ETH-XGaze dataset, the performance of ResNet50+RUDA model on $D_E \rightarrow D_D$ task even outperforms state-of-the-art within dataset gaze estimation methods, e.g., [5, 7]. The results show that RAT strategy alone does not improve cross domain performance, as expected. The ability to estimate gaze from rotated images does not improve estimation accuracy on normalized face images with upright orientation. After rotation consistency guided adaptation, the performance improves significantly. This proves our point in Sec. 3 that rotation consistency contain much more significant relation with physical model of gaze than data augmentation like rotation.

### 5.3. Comparison with SOTA UDA methods

To demonstrate the performance of RUDA framework, we compare it with state-of-the-art unsupervised domain adaptation methods on four cross domain tasks: $D_E \rightarrow D_M$, $D_E \rightarrow D_D$, $D_G \rightarrow D_M$, $D_G \rightarrow D_D$. We choose four typical methods for comparison:

- **ADDA [36]**: Reduce domain gap between source and target domain features by adversarial learning. A discriminator which classifies feature to source or target domain is introduced. 500 target domain images are used in our implementation for better performance.

- **DAGEN [16]**: A SOTA unsupervised domain adaptation method for gaze estimation by embedding representation design. 500 target domain images are used in our implementation for better performance.

- **GazeAdv [39]**: A SOTA unsupervised domain adaptation method for gaze estimation by adversarial learning. Appearance classification and head pose classification are designed as adversarial tasks.

- **Gaze360 [19]**: A SOTA unsupervised domain adaptation method for gaze estimation by adversarial learning.
method for gaze estimation by combination of adversarial learning, image flip and pinball loss.

- **RSD [4]:** A SOTA unsupervised domain adaptation method specially designed for regression tasks. It closes domain gap through orthogonal bases of the representation spaces without changing the feature scale.

For a fair comparison, we replace the backbone of all methods with ResNet18. The result is shown in Tab. 2. Our method outperforms SOTA methods by a big margin. RUDA framework significantly improves the performance on all tasks. Note that general unsupervised domain adaptation methods do not bring any performance improvement, which shows the difficulty of cross domain gaze estimation tasks. Methods designed for gaze estimation may bring performance gain on certain cross domain tasks, while make other tasks worse. Our RUDA framework performs stably and improves the performance in all four tasks.

### 5.4. Ablation Study

To prove the effectiveness of each component in RUDA framework, we conducted ablation study on all four cross domain tasks. In Tab. 3, we show the results for different combination of source domain training strategy, domain adaptation strategy and loss function:

- **R(Σ1, Σ2):** Training with Σ1, Σ2 loss respectively on rotation augmented source domain.
- **RΣD:** The proposed RAT strategy with ΣD loss function.
- **DAΣ1, DAΣ2:** The proposed CGDA strategy in which ΣD loss is replaces by Σ1, Σ2 loss function respectively.
- **DAΣD:** The proposed CGDA strategy with ΣD loss.

In Tab. 3, row 1-4 show that training with rotated images on source domain does not improve cross domain accuracy. But training with proposed ΣD appears to be more stable than Σ1, Σ2 thanks to the Σstd term. Results from row 5-7 prove the effectiveness of DAΣD, i.e., the proposed CGDA strategy. CGDA improves accuracy when combined with models from row 1 to 3 on all cross domain tasks. In row 5 to 7, although some of the combination reaches compatible ([row 5, DΣ → DΣ], [row 6, DΣ → DΣ]) or even better ([row 6, DΣ → DΣ]) performance than RUDA in certain tasks, they performs similar or even worse than the baseline ResNet18 in other tasks ([row 5, DΣ → DΣ], [row 6, DΣ → DΣ], [row 7, DΣ → DΣ]). Without ΣD loss in source domain training, combinations from row 5 to row 7 suffer from obvious performance gap for different cross domain tasks. This is a fatal flaw for unsupervised domain adaptation tasks as we have no target domain label to verify whether performance is improved or dropped. Compared with row 8 to row 10, methods with RAT strategy shows apparent stability and improves accuracy in all four tasks.

In row 8 and row 9, we test the combination of proposed CGDA strategy with different loss function in domain adaptation. Compare to row 10, ΣD achieves better overall performance gain than Σ1 and Σ2 loss function.

Above experiments validate the effectiveness of RAT and CGDA strategy. With the help of RAT and CGDA, the proposed RUDA framework achieves the most stable and satisfactory improvement in all four tasks.

### 5.5. Hyper Parameters and Further Analysis

#### 5.5.1 Hyper Parameters

In this section, we carried out experiments to investigate the impact of hyper parameters. Rotation degree is the most important hyper parameter as we create sub-label by rotation. In Tab. 4, we show the results of different rotation range. For a given degree r, we randomly rotate the image in range [-r, r]. In RAT phase, models perform similarly under different rotation degree. After adaptation by CGDA, the accuracy increases with the rotation range. During adaptation, the model is supervised by the rotation consistency of estimation results from rotated images, which corresponds to the Σstd term in distribution loss. Hence, we test models without CGDA on the subset of target domain Dl′ and count STD. As shown in Fig. 4, STD drops as the range of rotation shrinks. In consequence, smaller the range of rotation is, smaller the ΣD is, less uncertainty signal for the model to learn during adaptation.

We also evaluate the impact of the number of rotation for each image during CGDA phase. We set the number of rotation to 10, 15, 20, 25 during adaptation respectively while keeps rotation number in RAT phase at 5. The result jitters when number of rotation changes. But the overall disturbance is relatively subtle compared to range of rotation.
5.5.2 Importance of Rotation Consistency

We design the RUDA framework around the rotation consistency for it connects deeply with the physical nature of gaze. To prove the importance of rotation consistency, we replace rotation consistency with other data augmentation methods in RUDA framework.

Specifically, we choose two commonly-used image augmentations: 1) geometry augmentation, we apply random scaling and random translating to the normalized face images, and 2) noise augmentation, we randomly apply four kinds of different noise to the image including random noise, Gaussian noise and Poisson noise. Examples of three kinds of operation are shown in Fig. 5.

The results are shown in Tab. 6. Although geometry augmentation and noise augmentation are proven to be effective in other computer vision tasks such as classification and object detection, they do not bring any improvement in cross-domain gaze estimation tasks. We argue that it might be because that geometry consistency and noise consistency are easier to achieve as these two augmentation only disturb the appearance of images, do not touch the physical nature of gaze. Rotation brings more uncertainty information, i.e., it changes not only the appearance but also gaze direction.

5.5.3 System Limitations

The proposed RUDA framework have successfully addressed one of the critical problems in unsupervised domain adaptation, i.e., the shortage of training data and the absence of target labels. On the other hand, another common challenge of appearance-based gaze domain adaptation tasks is that the data distribution of source domain and target domain can be different. When the range of source domain is significantly smaller, the adaptation capability decreases. Such a problem has not been well addressed by existing methods. In the future, we can try to handle this problem and combine the resulting technique into our RUDA framework to further improve the system robustness.

6. Conclusions

In this paper, we present the rotation-enhanced unsupervised domain adaptation framework for gaze estimation tasks. Based on the rotation consistency property, the proposed RUDA framework adapts the model to unlabeled target domain. It first trains a rotation-augmented model with RAT strategy in source domain, then generalized to target domain via the guidance of sub-labels, i.e., estimation consistency across different rotation angles in CGDA phase. Experimental results show that the RUDA framework achieves stable and significant improvement in four different cross-domain tasks. The idea of rotation consistency may be applied in other physical related regression tasks such as pose estimation.
References


[34] Li Sun, Zicheng Liu, and Ming-Ting Sun. Real time gaze estimation with a consumer depth camera. Information Sciences, 320:346–360, 2015. 2


