Synthetic gaze data augmentation for improved user calibration

Gonzalo Garde[0000−0002−3635−6530],
Andoni Larumber-Bergera[0000−0002−8075−2426],
Sonia Porta[0000−0002−3687−4477],
Rafael Cabeza[0000−0001−7999−1182], and
Arantxa Villanueva[0000−0001−9822−2530]

Public University of Navarra (UPNA), Calle Arrosadía, 31006 Pamplona, Spain
http://www.unavarra.es/gi4e
{gonzalo.garde,andoni.larumbe,sporta,rcabeza,avilla}@unavarra.es

Abstract. In this paper, we focus on the calibration possibilities of a deep learning based gaze estimation process applying transfer learning, comparing its performance when using a general dataset versus when using a gaze specific dataset in the pretrained model. Subject calibration has demonstrated to improve gaze accuracy in high performance eye trackers. Hence, we wonder about the potential of a deep learning gaze estimation model for subject calibration employing fine-tuning procedures. A pretrained Resnet-18 network, which has great performance in many computer vision tasks, is fine-tuned using user's specific data in a few shot adaptive gaze estimation approach. We study the impact of pretraining a model with a synthetic dataset, U2Eyes, before addressing the gaze estimation calibration in a real dataset, I2Head. The results of the work show that the success of the individual calibration largely depends on the balance between fine-tuning and the standard supervised learning procedures and that using a gaze specific dataset to pretrain the model improves the accuracy when few images are available for calibration. This paper shows that calibration is feasible in low resolution scenarios providing outstanding accuracies below 1.5° of error.

Keywords: Gaze Estimation · Calibration · Transfer Learning.

1 Introduction

In the 70's one of the seminal papers about eye tracking was published [15]. This paper described a system based on the well-known pupil center-corneal reflection vector (PR-CR) and assumed a highly focused image of the eye area, i.e. a high resolution image of the eye region as is shown in figure 1 (left). Since then, outstanding works have been published in the field most of them persisting first, in the use of PC-CR technique and its variations and second, employing high resolution eye area images [18] [10] [16]. As a result of these researches, commercial solutions were developed.

As far as it is known, all of them require a subject calibration procedure consisting in asking the subject to gaze specific points in the screen. During the calibration, the system is adapted to the subject position and eye physiology. Depending of the type of system this *adaptation* process can be more or less explicitly carried out. On the one hand, the calibration of system geometry-based gaze estimation trackers permits to infer the value of different variables related to the user such as the corneal radius among other subject specific characteristics [6] [10]. On the other hand, the calibration of polynomial-based gaze estimation trackers is used in order to estimate the unknown coefficients of a polynomial, normally a second degree polynomial [2]. Regardless of the type of system, there is an agreement about the fact that a calibrated system provides better subject accuracies than the ones obtained by an average model. Outstanding accuracies below $0.5°$ can be found in the literature and provided by systems manufacturers.

The hardware of those systems presents some requirements. First, the use of an infrared light which provides better image quality and is responsible for the corneal glint which is key for an accurate gaze estimation. Second, in order to get a highly focused eye image, long focal lengths have to be used $\sim 35mm$. The necessity of this special optics, infrared lighting and filters are some of the reasons that prevent high performance eye tracking from being a plug-and-play technology.

In the last decade, a sustained effort has been made by the community in order to provide a more versatile eye tracking technology that can be implemented using webcams or the mobile phone cameras. The removal of the infrared light sources and the lower focal lengths produces a drastically different subject's images as system input (see figure 1, right). The techniques employed for high resolution images are not longer valid in the new scenario. Thus, a new exploration field is opened for deep learning in the gaze estimation field tackled as a computer vision problem. To train a model from scratch using deep learning approach, it is necessary to employ large datasets. Sadly, to create a high quality dataset with enough number of images it's a costly process. In order to address this limitation, it is common to use a model pretrained over a large database as an initial point, which improves training time and results. Int the case of gaze estimation, works can be found in the literature showing the potential of deep learning techniques [8] [12] [7]. Unfortunately, the reported accuracies are far from being comparable to the ones obtained by high resolution systems, e.g. $3°$ to $5°$ [4] [3]. One of the potentials assumed for deep learning it its generalization ability, that can be defined as the possibility of gaining knowledge from training data in a learning process and apply the gained knowledge on new data. It is agreed that model's generalization capability is a pursued property of the network also for gaze estimation. However, from the experience obtained for high resolution systems it is a known fact that the adaptation of the system to the specific user, i.e. calibration, results in better accuracies, which could enable the use of gaze estimation for other applications were accuracy would be critical. Hence, it is a relevant open issue how to introduce the calibration in the field of deep learning gaze estimation. The few examples found in the literature ap-

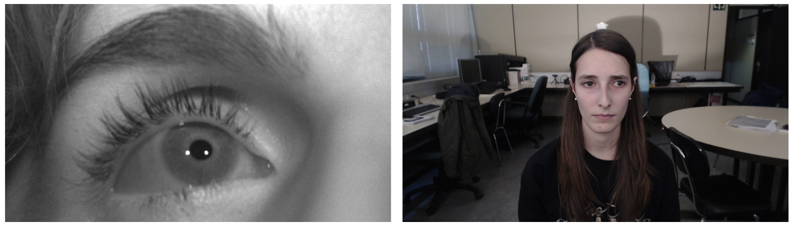

Fig. 1. Image obtained by a high resolution eye tracker using two corneal glints (left). Image obtained by a standard webcam (right) [14].

proach the problem from a fine-tuning perspective employing a few-shot training of an existing model [20] [13].

In this paper, we carry out a thorough study about user calibration for deep learning gaze estimation raised as a question between the balance required among generalization and personalization abilities of the models. This balance is measured according to its impact in the accuracy obtained for the gaze estimation system. We will study the impact of transfer learning over a model pretrained with a computer vision general database, Imagenet [5], versus the results obtained while pretraining over a gaze specialized synthetic database, U2Eyes [17], analyzing the results over a real gaze database, I2Head [14].

This paper is organized as follows, in the next section a summary of related works is performed. In section 3, the working framework is presented, i.e. datasets and the main methodological specifications are provided. Then, subject calibration details are introduced in section 4. The experiments configuration is explained in section 5. In section 6 the accuracies obtained by the different experiments are shown. Finally, the conclusions of our work are addressed.

2 Related Works

Calibration is a well-known procedure in high resolution systems. During the calibration process, the subject is asked to gaze specific points in the screen. The images acquired during calibration are automatically labelled with gazed point information and this permits to fit the gaze estimation function to the subject playing with the eye tracking system. Deep learning gaze estimation is a relatively new topic and the accuracies obtained are not comparable with the ones achieved by high resolution system. Most of the works found are related to compare the architectures to be used, labelled datasets to be used or the evaluation of the alternative training strategies of the models and less attention has been paid to personalization strategies for eye tracking systems. To follow selected relevant works addressing gaze models personalization are presented.

In one of the first relevant works, Krafka *et al.* [11] point out the relevance of using subject images during network training in order to improve the gaze estimation accuracy and its variation according to the calibration points used achieving accuracy values about 1.34 cm. In the work [20],Yu *et al.* try to construct the person-specific gaze estimation models by using few calibration points (few-shot approach). They use a VGG16 achitecture trained with ColumbiaGaze and MPIIGaze datasets. They use a more complex model using synthetic images for gaze prediction consistency. They perform the person adaptation using a few-shot scheme in which 1-5-9 samples are employed to personalize a previously pretrained model. They also confirm that the accuracy improves as the number of calibration images are increased achieving accuracies in the range [2.68°, 5°]. In this work they point out one of the main facts of deep learning gaze estimation which is the lack of labelled binocular data compared to other computer vision problems in which wide datasets area available.

An interesting approach is also described by Linden *et al.* [13]. In this work a complex model consisting in three ResNet-18 are used for three input images (both eyes and face). The network outputs for each eye are concatenated with "calibration parameters" as inputs to a fully connected module. The calibration parameters are the ones to be adjusted during the calibration procedure. Again, the improvement according to calibration points is achieved showing values about 2.76°.

As shown in this summary more attention is paid to the number of points required for calibration than to the specific computer vision domain. If calibration is required, the number of calibration points, although methodologically important, does not involve practical problems for most of the applications since high number of points can be acquired, e.g. by asking the subject to track a point in the screen.

3 Working Framework

3.1 Image Databases

The proposed synthetic framework is based on U2Eyes dataset [17]. The choice of a synthetic dataset to the detriment of a dataset with real images is motivated by the ability of synthetic datasets to provide with a large amount of images while assuring that, in all cases, the labeling is consistent and correct. As we are looking for high accuracies, we believe that this certainty is important. However, one of the main drawbacks of synthetic dataset is that, by their own nature, they have a more limited variability than the one that can be found in real datasets. Twenty different simulated subjects are provided in the public version of the dataset. The images have a resolution of 3840x2160 pixels (4K) and were created using Unity. The provided images represent the eyes area of a subject gazing at different points on a screen simulating a standard remote eye tracking session. The images are annotated with head pose and observed points information. Additionally, 2D and 3D landmarks information of both eyes is provided. For each subject 125 head positions are simulated from which two gazing grids containing 15 and 32 points are observed. Consequently, 5,875 images per user are provided resulting in a total of approximately 120K images. U2Eyes represents a rich appearance variation environment. The authors claim

Fig. 2. Samples extracted from U2Eyes dataset.

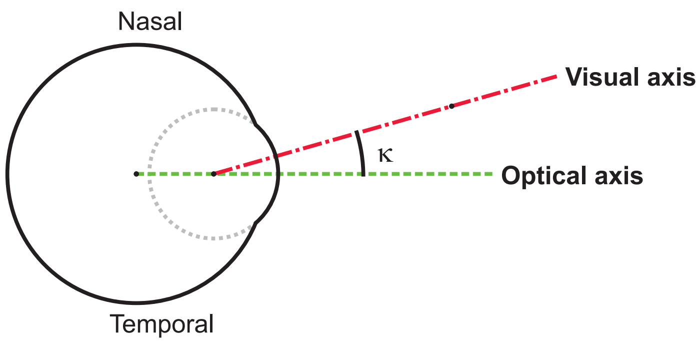

Fig. 3. Simplified model of the eye.

that essential eyeball physiology elements and binocular vision dynamics have been modelled. In figure 2 samples of the dataset are shown.

The 3D eye model employed to generate eye images resembles the simplified eye model for gaze estimation in the literature [1]. The Line of Sight (LoS) is approximated by the visual axis that presents an angular offset, κ, with respect to the optical axis, i.e. eyeball symmetry axis of the eye (see figure 3). The eyeball presents individual's specific characteristics, some of them such as the angle κ cannot be inferred from the image and need to be calibrated.

I2Head dataset is employed as a real benchmark in order to validate the proposed framework. I2Head is a public dataset providing images annotated with gaze and head pose data. This dataset was constructed using a magnetic sensor for pose detection and a careful setup calibration procedure. More details about the design of this dataset can be found in [14]. The dataset contains information about twelve individuals gazing two grids containing 17 and 65 points from eight different head positions, in constrained and free head movement scenarios. For each user, information about 232 ($65 \times 2 + 17 \times 6$) gazing points is provided representing a total of 2,784 points data. The range of gaze angles is approximately ± 20 °.

I2Head does not provides landmark information. Hence, a manual labelling of the eye region bounding box has been performed for a reduced set of images

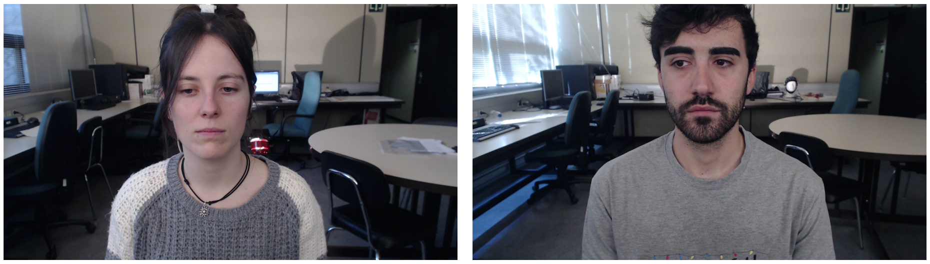

Fig. 4. Samples extracted from I2Head dataset.

Fig. 5. Image preprocessing. The images (a) are rotated (b) so eyes are in the same line, which makes the bounding box (c) contains as meaningful information as possible. Then, a padding process (d) assures that all images have the same size before feeding them to the network.

of the dataset. More specifically, the four central sessions have been selected for each user containing 164 gazing points per user. In order to perform a realistic comparison between the datasets, eye area is extracted from I2Head images. In figure 4 samples of the dataset are shown.

U2Eyes resembles to a large extent I2Head dataset. This is a desired quality as the importance of pretraining in a close domain for gaze estimation is studied. However, although both databases represent standard eye tracking sessions in a remote setup, some differences are found. While the camera is positioned on the top of the screen in I2Head dataset a centered position with respect to the gazing grid is selected in U2Eyes. Consequently, in U2Eyes subjects gaze points above and below the camera while all the points are placed below the camera in the I2Head dataset. In the same manner, the range of head poses is different in both datasets when referring to vertical positions. All these aspects should be taken into account when the alternative training strategies are proposed.

3.2 Image Conditioning

The preprocessing applied to U2Eyes and I2Head data is described: both synthetical and real image preprocessing is equivalent, and it is shown in figure 5. First, the original image, (a), is rotated (b) in order to normalize the roll component. The rotation is done using the angle between the horizontal and the straight line defined by the two outer eye corners. Then, a bounding box is created using the

Table 1. Configurations of the different experiments according to the training mode, *U2Eyes* or *ImageNet*, and the number of users and images in the training and testing phase. The K parameter varies from 0 to 11, i.e. 0 indicates that none of the subjects of I2Head dataset has been included to train the model, except for the subject to be calibrated, and 11 indicates that all the additional subjects are present in the training phase.

Model	**Train # Users/Images**	**Calibration # Users/Images**	**Total Train # Images**	**Test # Users/Images**
ImageNet	K/K*130	1/34	K*130+34	1/130
U2Eyes	K/K*130	1/34	K*130+34	1/130

distance between the two outer eye corners and the image is cropped (c). The rotation before cropping is necessary to prevent the gray pixels that would be present in the U2Eyes images if we would just obtain the bounding box from the original source. Finally, a black edge is added until the image size is 390x85 pixels (d). As the images from both datasets cover a range of well-known distances, the black edge was added to make all images the same size. Thus, the farther to the camera, the bigger the black edge, providing additional depth information to the network.

3.3 Network Architecture

The architecture of our network consists on the Resnet-18 [9] as backbone, followed by a Global Average Pooling layer and three Fully Connected Layers. As the focus of this paper is not to provide a novel architecture but to study the calibration and transfer learning process for gaze estimation, the Resnet-18 was chosen because of its performance over Imagenet [5] classification task while being simple enough to make retraining steps feasible in both, time and hardware requirements. The results obtained over Imagenet ensure that the network is able to extract meaningful features from images. The three fully connected layers at the top of the network make use of these features to compute the x and y coordinates of the look-at-points.

3.4 Implementation Details

In this section, technical details for the sake of reproducibility are described. The experiment is divided into two different phases:

- A first one where a model with the same architecture as the one shown in section 3.3 is trained over the synthetic dataset U2Eyes, saving the trained model, hereafter *U2Eyes-model*, to be used as the initial base for future steps. The goal with this step is to bring closer the domain in which the model is trained to the final environment.
- A second phase where the model is trained over real images obtained from the I2Head dataset in a subject calibration fashion. In this phase, we distinguish

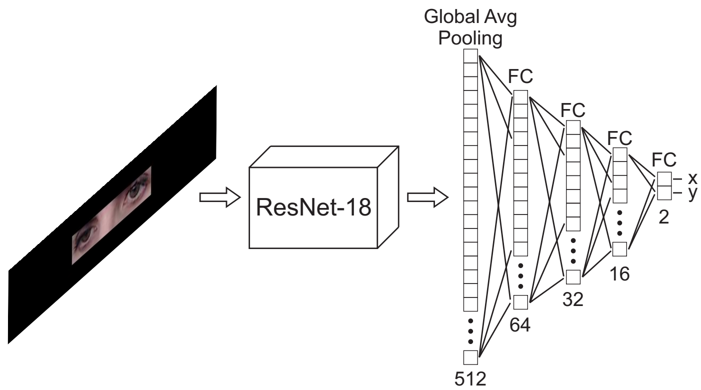

Fig. 6. Architecture proposed. The backbone consist on a Resnet-18 to extract meaningful characteristics from the image. Then, these characteristics are fed into fully connected regressor network to obtain the final gaze components.

among two different situations: using the *U2Eyes-model* as initial point and starting from the Imagenet [5] weights for the Resnet-18 backbone (see section 3.3).

Independently of the training phase, the steps followed to train the model and the training parameters were kept the same for all cases. The models are trained over 240 epochs, using a batch size of 128 images. The loss function employed is the euclidean distance between the estimated look-at-point and the real look-at-point, represented by the following equation:

$$Loss := \frac{1}{N}\sum_{i=1}^{N} \|p - \hat{p}\|_2, \tag{1}$$

where p is the real look-at-point, $\hat{p}$ is the estimated look-at-point, and N is the number of images per batch. Adam optimizer is used to optimize the loss function. The learning rate schedule followed is based on the Cyclic learning rate schedule [19] in a triangular manner for the first 200 epochs, fluctuating among a maximum learning rate value of 0.002 and a minimum learning rate of 0.0002, and then a linear decrease for the remaining 40 epochs where the learning rate goes from 0.0002 to 0.00002. The main specifications of the computer where the experiments were run are: CPU: Intel(R) Xeon(R) CPU E5-1650 v4 @3.60GHz, 128 GB of RAM and a Nvidia Titan X (Pascal) GPU.

For the sake of reproducibility, the weights of the pretrained model over *U2Eyes* and a function for the model that is trained directly over the *ImageNet* weights will be available at Github[1]. It is important to remark that we are not training over U2Eyes neither over Imagenet images in this paper, we are using

[1] https://github.com/GonzaloGardeL/Synthetic-gaze-data-augmentation-for-improved-user-calibration

pretrained models over these datasets and then retraining these models over I2Head images.

4 Subject Calibration

The concept of calibration in gaze estimation refers to the process whereby a model is tuned over a given number of images from an individual to adapt to the specific parameters of that individual. This is especially crucial as there are characteristics that cannot be learned in any other way at the moment and that have a significant impact on the outcome of the regression model. Ideally, when calibrating a system for a user, the fewer images that are necessary for calibration the better, as a large number of images implicates longer calibration times and a more complicated user-machine interaction. On the other side, the improvement obtained is generally directly proportional to the number of images. As a result, a trade-off exists among the number of images, the calibration time and the final accuracy. In any case, the calibration processes work with few images if it is compared with any other approach where deep learning is used. A common strategy these days is to use few-shot training to adapt the parameters of the networks. In this paper, we focus on the advantages of pretraining in a synthetic environment where we are not limited by the number of images or the reliability of the data.

5 Experiments

The final goal of the experiment is to observe the impact of pretraining over a dataset whose domain is closer for gaze estimation before facing a real dataset and the importance of the number of images while calibrating models for gaze estimation. An approximation of the Leave-One-Out strategy is followed in order to study subject calibration. 34 images of the user to be calibrated are included in the training set together with a varying number of additional subjects extracted from I2Head dataset ranging from 0 to 11 users. For each one of these subjects in the training, 130 additional images are used. The condition of no-user-calibration is not contemplated, as the user data is always used while training. A resume of the experiments configuration is shown in table 1.

6 Results

The outputs from the experiments were arranged together based on the training user configuration, and the figures and tables that are shown in this section are an average from the experiments of all users. The angular offset (in degrees) between the estimated gaze direction and subject's visual axis is used as comparison metric. This angular offset is calculated by computing the distance between the estimated point and the real point in the grid and the known distance between the real point and the user. For the sake of readability, the results from the

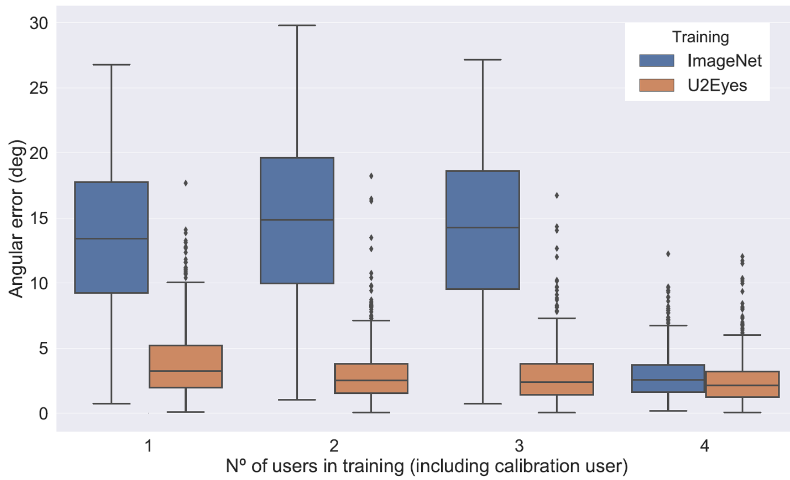

Fig. 7. Experiments when training from 1 user to 4 users. In this figure we observe that the performance of the *ImageNet* models change drastically when the number of users included in the training dataset gets reduced from a specific number. The models trained over the *U2Eyes* strategy are more robust when working with a reduced dataset.

experiments have been plotted in two different figures, figure 7 and figure 8, using both figures the results from training with 4 users as common anchor point. We highlight two cases of interest:

- Correlation between the number of training images and regressor estimation.
- Comparison of *U2Eyes* and *ImageNet* Methods.

6.1 Number of Training Images and Regressor Estimation

The regressor estimations are worse as the number of training images decreased. This is the expected behavior for this problem as by reducing the number of available images for training, it is more difficult for the network to learn the optimal parameters for the problem. In table 2, the mean and median for each case are shown. The results are consistent for both *U2Eyes* and *ImageNet* training modes. When the number of training images is maximum (12 users in training, the calibration user + 11 additional users), outstanding results with degrees of error lower than 1.5º are obtained for the median.

6.2 *U2Eyes* and *ImageNet* Methods

At figure 8, we observe that the results obtained from *U2Eyes* and *ImageNet* are similar but slightly better for *ImageNet* method when the number of additional users goes from 4 to 11. One possible explanation for this behaviour could be

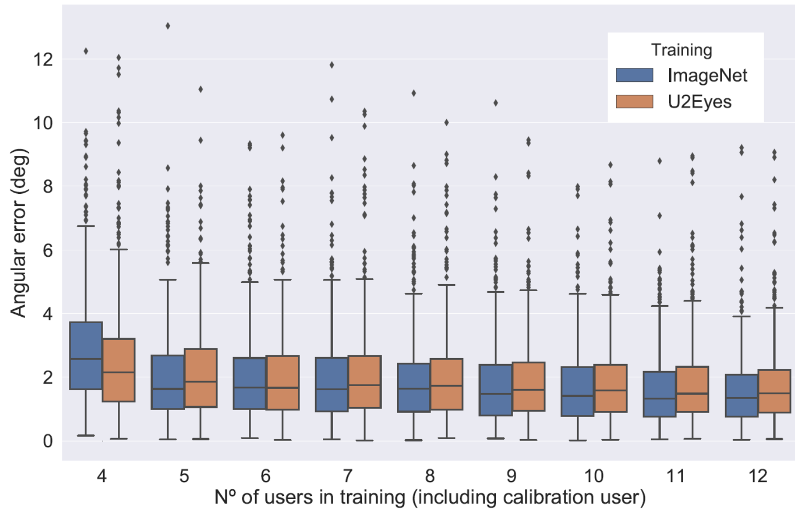

Fig. 8. Experiments when training from 4 users to 12 users. The performance of the models tends to improve as the number of training images increases. Both *ImageNet* and *U2Eyes* models have a similar behavior for each one of the cases.

that, if enough images are used to train the ImageNet model over the specific user, the model benefits from the more variability presented in real images rather than the more limited variability of synthetic ones. Attending to table II, the biggest difference in mean is under 0.17º. As we increase the number of training users, the obtained results from both models are better. However, when adding 3 or fewer additional users to the training, we observe in figure 7 that the behavior is different. The *U2Eyes* models are more robust to the lack of training images than the *ImageNet* models. The hypothesis is that, when the number of training images is low enough, the models that were pretrained in a similar domain (synthetic dataset U2Eyes in this case) are more capable to continue learning than the ones trained in a more general domain. If we compare the results in table 2, we can observe that this difference is up to 10.16º for the median value in the case of training using only the calibration user.

7 Conclusions

In this paper, subject calibration based in transfer learning for low resolution systems is studied. To this end, a deep learning model and two gaze data datasets, i.e. I2Head and U2Eyes, are used containing real and synthetic images. Two experimental setups have been employed in order to validate our hypothesis. The first setup uses Imagenet as start-point while a dataset containing eye synthetic images is employed in the second one. The results presented in this paper show that a calibration strategy is possible for low resolution, achieving first-class

Table 2. Results from the different experiments configurations. The mean and median angular errors in degrees are shown for each. The number of users in training includes the user for which the system has been calibrated. The maximum and minimum values for each one of the columns are emphasized.

Users in training	**Mean (°)**		**Median (°)**	
	Imagenet	**U2Eyes**	**Imagenet**	**U2Eyes**
1	13.615	**3.891**	13.401	**3.243**
2	**14.812**	2.967	**14.880**	2.522
3	14.069	2.861	14.255	2.404
4	2.867	2.488	2.567	2.149
5	2.004	2.149	1.631	1.867
6	2.028	1.965	1.675	1.667
7	1.987	2.039	1.617	1.746
8	1.877	1.968	1.639	1.724
9	1.758	1.860	1.471	1.611
10	1.681	1.818	1.412	1.588
11	1.615	1.777	**1.334**	**1.485**
12	**1.559**	**1.714**	1.344	1.486

performance when adapting a gaze estimation regressor for an specific user, yielding results close to the ones achieved in high resolution, i.e. $\sim 1.5°$ which is one of the main contributions of our work. Furthermore, the importance of providing domain images during the training process has been confirmed and also the benefits of pretraining the regressor in a closer domain instead of in a more general dataset to compensate the lack of useful gaze data images, due to the difficulties in both acquiring and labeling them, that characterize gaze estimation problem.

Acknowledgment

References

1. Abass, A., Vinciguerra, R., Lopes, B.T., Bao, F., Vinciguerra, P., Jr, R.A., Elsheikh, A.: Positions of ocular geometrical and visual axes in brazilian, chinese and italian populations. Current Eye Research **43**(11), 1404–1414 (2018). https://doi.org/10.1080/02713683.2018.1500609, https://doi.org/10.1080/02713683.2018.1500609, pMID: 30009634
2. Cerrolaza, J.J., Villanueva, A., Cabeza, R.: Study of polynomial mapping functions in video-oculography eye trackers. ACM Trans. Computer-Human Interaction **19**(2), 10:1–10:25 (Jul 2012). https://doi.org/10.1145/2240156.2240158, http://doi.acm.org/10.1145/2240156.2240158
3. Chen, Z., Shi, B.E.: Geddnet: A network for gaze estimation with dilation and decomposition (2020), https://arxiv.org/abs/2001.09284
4. Cheng, Y., Huang, S., Wang, F., Qian, C., Lu, F.: A coarse-to-fine adaptive network for appearance-based gaze estimation (2020), https://arxiv.org/abs/2001.00187

5. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: A Large-Scale Hierarchical Image Database. In: CVPR09 (2009)
6. Guestrin, E., M.Eizenman: General theory of remote gaze estimation using pupil center and corneal reflections. IEEE Trans. on Biomedical Engineering **53**(6), 1124–1133 (2006)
7. Guo, T., Liu, Y., Zhang, H., Liu, X., Kwak, Y., In Yoo, B., Han, J.J., Choi, C.: A generalized and robust method towards practical gaze estimation on smart phone. In: 2019 International Conference on Computer Vision (ICCV) Workshops. ICCV'19 (Oct 2019)
8. He, J., Pham, K., Valliappan, N., Xu, P., Roberts, C., Lagun, D., Navalpakkam, V.: On-device few-shot personalization for real-time gaze estimation. In: 2019 IEEE International Conference on Computer Vision (ICCV) Workshops. ICCV'19 (Oct 2019)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015)
10. Hennessey, C., Noureddin, B., Lawrence, P.: A single camera eye-gaze tracking system with free head motion. In: Proceedings of the 2006 Symposium on Eye Tracking Research & Applications. p. 87–94. ETRA '06, Association for Computing Machinery, New York, NY, USA (2006). https://doi.org/10.1145/1117309.1117349, https://doi.org/10.1145/1117309.1117349
11. Krafka, K., Khosla, A., Kellnhofer, P., Kannan, H., Bhandarkar, S., Matusik, W., Torralba, A.: Eye tracking for everyone. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2176–2184 (2016)
12. Linden, E., Sjostrand, J., Proutiere, A.: Learning to personalize in appearance-based gaze tracking. In: 2019 IEEE International Conference on Computer Vision (ICCV) Workshops. ICCV'19 (Oct 2019)
13. Linden, E., Sjostrand, J., Proutiere, A.: Learning to personalize in appearance-based gaze tracking. In: The IEEE International Conference on Computer Vision (ICCV) Workshops (Oct 2019)
14. Martinikorena, I., Cabeza, R., Villanueva, A., Porta, S.: Introducing I2Head database. In: 7th International Workshop on Pervasive Eye Tracking and Mobile Eye based Interaction. PETMEI'07 (2018)
15. Merchant, J., Morrissette, R., Porterfield, J.: Remote measurement of eye direction allowing subject motion over one cubic foot of space. IEEE Transactions on Biomedical Engineering **21**(4), 309–317 (July 1974)
16. Morimoto, C.H., Amir, A., Flickner, M.: Detecting eye position and gaze from a single camera and 2 light sources. In: Proccedings International Conference on Pattern Recognition. pp. 314–317 (2002)
17. Porta, S., Bossavit, B., Cabeza, R., Larumbe-Bergera, A., Garde, G., Villanueva, A.: U2eyes: A binocular dataset for eye tracking and gaze estimation. In: 2019 IEEE International Conference on Computer Vision (ICCV). ICCV'19 (Oct 2019)
18. Shih, S.W., Liu, J.: A novel approach to 3-D gaze tracking using stereo cameras. IEEE Transactions Systems Man and Cybernetics Part-B **34**(1), 234–245 (February 2004)
19. Smith, L.N.: Cyclical learning rates for training neural networks. In: 2017 IEEE Winter Conference on Applications of Computer Vision (WACV) (2017)
20. Yu, Y., Liu, G., Odobez, J.M.: Improving few-shot user-specific gaze adaptation via gaze redirection synthesis. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2019)