Webcam-Based Pupil Diameter Prediction Benefits from Upscaling

Vijul Shah[1][a], Brian B. Moser[1,2][b], Ko Watanabe[1,2][c] and Andreas Dengel[1,2][d]

[1]*RPTU Kaiserslautern-Landau, Germany*

[2]*German Research Center for Artificial Intelligence (DFKI), Germany*

{firstname.lastmame}@dfki.de

Keywords: Pupil Diameter Prediction, Image Super-Resolution.

Abstract: Capturing pupil diameter is essential for assessing psychological and physiological states such as stress levels and cognitive load. However, the low resolution of images in eye datasets often hampers precise measurement. This study evaluates the impact of various upscaling methods, ranging from bicubic interpolation to advanced super-resolution, on pupil diameter predictions. We compare several pre-trained methods, including CodeFormer, GFPGAN, Real-ESRGAN, HAT, and SRResNet. Our findings suggest that pupil diameter prediction models trained on upscaled datasets are highly sensitive to the selected upscaling method and scale. Our results demonstrate that upscaling methods consistently enhance the accuracy of pupil diameter prediction models, highlighting the importance of upscaling in pupilometry. Overall, our work provides valuable insights for selecting upscaling techniques, paving the way for more accurate assessments in psychological and physiological research.

1 INTRODUCTION

The widespread adoption of eye-tracking technology in daily life is accelerating, as highlighted by innovations like Apple's camera-based eye tracking (Apple Inc., 2024), (Greinacher and Voigt-Antons, 2020). As a fortunate side-effect, these technologies enable the analysis of human cognitive states, which are deeply connected to observable features in the eyes (Dembinsky et al., 2024a), (Dembinsky et al., 2024b). While much of the existing research focuses on blink detection (Hong et al., 2024) and gaze estimation (O'Shea and Komeili, 2023), (Yun et al., 2022), (Bhatt et al., 2024), which employ biomarker usage (Liu et al., 2022), infrared reflections (Fathi and Abdali-Mohammadi, 2015), or image analysis techniques (Hisadome et al., 2024), there is comparatively less emphasis on measuring pupil diameters (Sari et al., 2016), (Caya et al., 2022). Yet, accurately capturing pupil size is critical for assessing various physiological and psychological conditions: Recent research shows that the diameter of the pupil can indicate levels of stress (Pedrotti et al., 2014), focus (Lüdtke et al., 1998), (Van Den Brink et al., 2016), or cognitive load (Kahneman and Beatty, 1966), (Pfleging et al., 2016), (Krejtz et al., 2018). Moreover, pupil size is linked to the activity of the *locus coeruleus* (Murphy et al., 2014), (Joshi et al., 2016), a crucial brain region for memory management over both short and long terms (Kahneman and Beatty, 1966), (Kucewicz et al., 2018). It is also vital in other medical contexts, such as evaluating the pupillary responses relating to neurological conditions like Alzheimer's disease (Granholm et al., 2017), (Tales et al., 2001), (Kremen et al., 2019), schizophrenia (Reddy et al., 2018), Parkinson's disease (Micieli et al., 1991), opioid use (Murillo et al., 2004), mild cognitive impairment (Elman et al., 2017), and in patients with brain injuries in intensive care settings (Kotani et al., 2021). Therefore, precise estimation of pupil diameter is essential for advancing the effectiveness of image-based eye-tracking technologies.

The introduction of the EyeDentify (Shah et al., 2024) dataset, which offers webcam-based eye images with corresponding pupil diameters, marks a significant advancement in pupilometry research. Unlike previous datasets (Ni and Sun, 2019), (Khokhlov et al., 2020) that were either not publicly accessible or recorded under highly controlled conditions, EyeDentify provides a diverse array of recordings featuring varying seating positions and distances. Thus

[a] https://orcid.org/0009-0008-5174-0793
[b] https://orcid.org/0000-0002-0290-7904
[c] https://orcid.org/0000-0003-0252-1785
[d] https://orcid.org/0000-0002-6100-8255

Shah, V., Moser, B. B., Watanabe, K. and Dengel, A.
Webcam-Based Pupil Diameter Prediction Benefits from Upscaling.
DOI: 10.5220/0013162800003890
Paper published under CC license (CC BY-NC-ND 4.0)
In *Proceedings of the 17th International Conference on Agents and Artificial Intelligence (ICAART 2025) - Volume 2*, pages 376-385
ISBN: 978-989-758-737-5; ISSN: 2184-433X
Proceedings Copyright © 2025 by SCITEPRESS – Science and Technology Publications, Lda.

potentially advancing the development of consumer-grade pupillometers that are capable of handling diverse eye colors and are easily accessible without any significant efforts, position constraints, and technical expertise. However, the primary challenge with this dataset is the low quality of the images, which can be attributed to the recording camera quality and the small size of the eyes within the images. This necessitates the application of image upscaling techniques to enable the effective use of deep neural networks for pupil diameter prediction.

In this work, we explore the impact of various image Super-Resolution (SR) techniques on the accuracy of webcam-based pupil diameter predictions. Image SR aims to transform low-resolution images into high-resolution counterparts, potentially enhancing the clarity and detail of visual data used in training models for more accurate pupil diameter estimation (Moser et al., 2023). We demonstrate that employing advanced, pre-trained SR models can substantially improve the accuracy of pupil diameter predictions in low-quality, webcam-based images. Yet, we found that different image SR methods affect pupil diameter estimation differently. The effectiveness of SR methods varied, with some enhancing the features necessary for precise pupilometry more effectively than others. Nevertheless, we can conclude that using upscaling methods, in general, improves the performance of pupil diameter prediction models. Overall, our comparative analysis provides clear guidance on selecting appropriate SR techniques for pupilometry.

2 RELATED WORK

In this section, we briefly review the usage of image SR as a pre-processing step for downstream tasks and survey the state-of-the-art of pupil diameter estimation.

2.1 Super-Resolution as Pre-Processing

Image SR is the process of transforming a LR image into a HR one, effectively solving an inverse problem (Moser et al., 2023). More explicitly, a SR model $M_\theta : \mathbb{R}^{\mathrm{H}\times\mathrm{W}\times\mathrm{C}} \to \mathbb{R}^{s\cdot\mathrm{H}\times s\cdot\mathrm{W}\times\mathrm{C}}$ is trained to inverse the degradation relationship between a LR image $\mathbf{x} \in \mathbb{R}^{\mathrm{H}\times\mathrm{W}\times\mathrm{C}}$ and the HR image $\mathbf{y} \in \mathbb{R}^{s\cdot\mathrm{H}\times s\cdot\mathrm{W}\times\mathrm{C}}$, where s denotes the scaling factor and the degradation relationship can be described by

$$\mathbf{x} = ((\mathbf{y} \otimes k) \downarrow_s + n)_{JPEG_q}, \tag{1}$$

where k is a blur kernel, n the additive noise, and q the quality factor of a JPEG compression. In a supervised setting, the training is based on a dataset $\mathbb{D}_{SR} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{N}$ of LR-HR image pairs of cardinality N and on the overall optimization target

$$\theta^* = \arg\min_{\theta} \mathbb{E}_{(\mathbf{x}_i,\mathbf{y}_i)\in\mathbb{D}_{SR}} \|M_\theta(\mathbf{x}_i) - \mathbf{y}_i\|^2 \tag{2}$$

Trained SR models are utilized across a wide array of fields, enhancing everything from medical imaging, where increased image clarity can have critical implications for patient care, to satellite imagery that provides more detailed insights into Earth's geography (Song et al., 2022), (Tang et al., 2021). In consumer electronics, such as smartphones and high-definition televisions, SR technologies significantly improve the visual quality, creating more engaging and realistic digital experiences (Zhan et al., 2021), (Shi et al., 2016). With the rapid advancements driven by deep learning and cutting-edge generative models, the field of image SR has experienced significant progress (Moser et al., 2024b), (Li et al., 2023), (Bashir et al., 2021). This work, however, does not seek to develop new image SR methodologies. Instead, it leverages SR technology as a preprocessing step to enhance the precision of pupil diameter measurements for images in everyday settings.

Similar applications of pre-trained SR models for downstream tasks inspire our goal in related fields, such as image recognition (Kim et al., 2024), (He et al., 2024), remote sensing (Chen et al., 2024), dataset distillation (Moser et al., 2024a), and others (Liu, 2024), (Jiang et al., 2024). For instance, *Chen et al.* utilized image SR to improve the quality of semantic segmentation (Chen et al., 2023a). In a different context, *Mustafa et al.* adopted image SR as a defensive strategy against adversarial attacks on image classification systems (Mustafa et al., 2019). Similarly, *Na et al.* applied image SR to boost the performance of object classification algorithms (Na and Fox, 2020). By integrating image SR into our workflow, we aim to refine the input data quality, thus enabling more accurate and reliable analyses in pupil diameter estimation.

2.2 Pupil Diameter Estimation

Ni et al. introduced a method named BINOMAP for estimating pupil diameter, utilizing dual cameras - referred to as master and slave - as a binocular geometric constraint for analyzing gaze images (Ni and Sun, 2019). This model is built on Zhang's algorithm, which recorded a mean absolute error of 0.022 ± 0.017mm (Zhang, 1999). Similarly, *Caya et al.* used a camera positioned 10cm away from the subject's face to capture facial images. These images were then processed on a Raspberry Pi, which involved converting RGB images to grayscale, adjust-

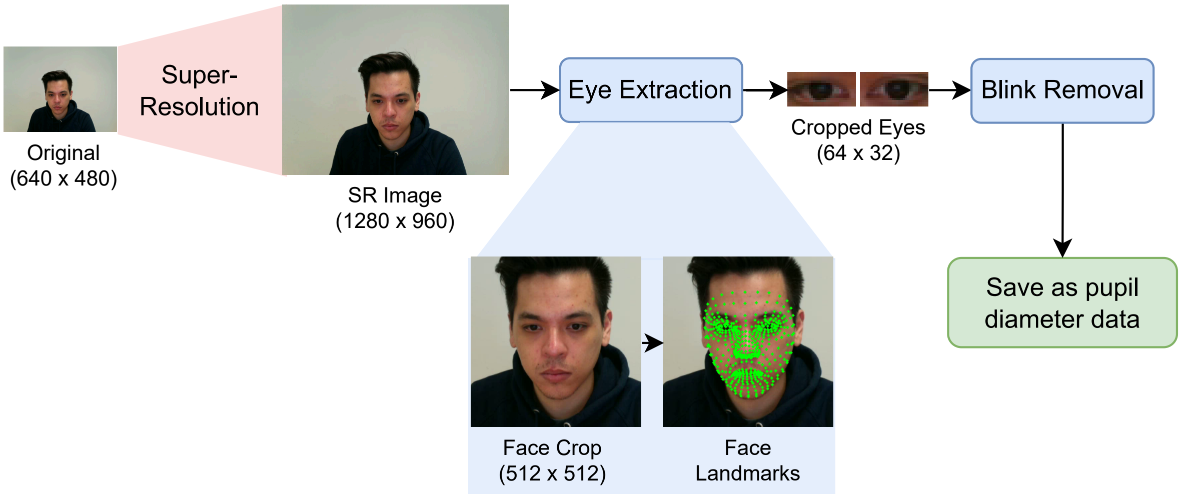

Figure 1: Pipeline of our data preprocessing with image SR. As a first step, we super-resolve the raw data with a pre-defined scaling factor (here 2×). Next, we used Mediapipe to extract the respective cropped eye images (64×32), left and right, for face detection and landmark localization. Subsequently, we applied blink detection on the cropped eyes using the Eye Aspect Ratio (EAR) and a pre-trained vision transformer for blink detection, as described in EyeDentify (Shah et al., 2024). Cropped eye images are then saved based on the EAR threshold and model confidence score.

ing contrast and brightness, reshaping images, and applying the Tiny-YOLO algorithm for pupil diameter estimation (Khokhlov et al., 2020). Their approach resulted in measurement accuracies with a percent difference of 0.58% for the left eye and 0.48% for the right eye. Both works face significant constraints related to specific conditions, including the necessity for dual cameras and maintaining a constant, fixed distance between the face and the camera. Another major limitation of these works is that their datasets are not publicly available, contrary to the EyeDentify dataset (Shah et al., 2024).

3 METHODOLOGY

The goal of this work is to apply SR models of the form $M_\theta : \mathbb{R}^{\mathrm{H}\times\mathrm{W}\times\mathrm{C}} \rightarrow \mathbb{R}^{s\cdot\mathrm{H}\times s\cdot\mathrm{W}\times\mathrm{C}}$ to improve the quality of eye images derived from face webcam images, denoted as $\mathbb{D}_{eyes} \subset \mathbb{D}_{faces}$, which is crucial for accurate pupil diameter estimation and cognitive state analysis. More formally, we aim at constructing $\mathbb{D}^{M_\theta}_{eyes} = \{(M_\theta(\hat{\mathbf{x}}_i), \mathbf{y}_i)\}_{i=1}^{N}$, where $(\hat{\mathbf{x}}_i, \mathbf{y}_i) \in \mathbb{D}_{eyes} \subset \mathbb{R}^{\mathrm{H}\times\mathrm{W}\times\mathrm{C}} \times \mathbb{R}$, $\hat{\mathbf{x}}_i \in \mathbb{R}^{\mathrm{H}\times\mathrm{W}\times\mathrm{C}}$ denotes the webcam images of eyes and $\mathbf{y}_i \in \mathbb{R}$ their respective pupil diameter size. Due to the sparsity of available training data in this eye-monitoring domain (Shah et al., 2024), we primarily refer to pre-trained SR models with given parameters θ instead of training a model M_θ from scratch. Figure 1 illustrates the overall pipeline, which integrates SR, i.e., M_θ, before any face detection, eye localization, cropping, and blink detection. This revised methodology leverages the strengths of existing SR models while tailoring their application to meet the specific demands of eye feature analysis.

Initially, we planned to apply pre-trained SR techniques directly to isolated images of the left and right eyes, as suggested by the authors of EyeDentify (Shah et al., 2024). However, this approach faces significant limitations, such as the rarity of eye images in image SR training datasets, e.g., DIV2K (Agustsson and Timofte, 2017) or Flicker2K (Timofte et al., 2017). State-of-the-art SR models like HAT (Chen et al., 2023b) or face SR models like GFPGAN (Wang et al., 2021a) are primarily optimized for everyday or full-face images. When these models are applied directly to eye images, their effectiveness diminishes due to a mismatch in the data distribution and latent space, which are tailored for the complexities of everyday or entire face features, as shown in Figure 2. To address this issue, we propose a more general approach: instead of applying SR directly to eye webcam images $\mathbb{D}^{M_\theta}_{eyes} \subset \mathbb{D}^{M_\theta}_{faces}$, we utilize the entire face webcam images $\mathbb{D}^{M_\theta}_{faces}$. Thus, our revised goal is to derive

$$\mathbb{D}^{M_\theta}_{faces} = \{(M_\theta(\mathbf{x}_i), \mathbf{y}_i)\}_{i=1}^{N}, \tag{3}$$

where $\mathbf{x}_i \in \mathbb{D}_{faces} \subset \mathbb{R}^{\mathrm{H}\times\mathrm{W}\times\mathrm{C}}$ denotes the webcam full-face images before any eye-cropping $g : \mathbb{R}^{\mathrm{H}\times\mathrm{W}\times\mathrm{C}} \rightarrow \mathbb{R}^{\mathrm{H}'\times\mathrm{W}'\times\mathrm{C}}$ with $\mathrm{H}' \ll \mathrm{H}$ and $\mathrm{W}' \ll \mathrm{W}$ happened, i.e., $\hat{\mathbf{x}}_i = g(\mathbf{x}_i)$. This allows the SR models trained on classical SR datasets $\mathbb{D}_{SR}$ to operate within

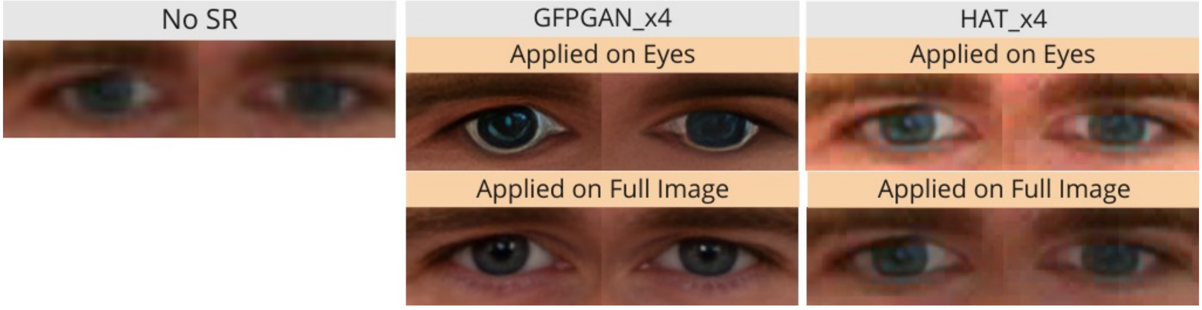

Figure 2: Comparison of applying image SR models on the cropped eye images versus applying them on the entire image. While the SR approximations on the entire image lead to results plausible to the respective input, the SR models applied to the cropped eye images lead to very distinct images. For instance, GFPGAN (left) produces unnatural pupils, whereas HAT (right) emits brightness shifts.

their optimal data distribution context, i.e.,

$$\begin{aligned}
&\|\mu_{\mathbb{D}_{SR}} - \mu_{\mathbb{D}_{faces}}\|^2 \ll \|\mu_{\mathbb{D}_{SR}} - \mu_{\mathbb{D}_{eyes}}\|^2 \text{ and} \\
&Tr\left(\Sigma_{\mathbb{D}_{SR}} + \Sigma_{\mathbb{D}_{faces}} - 2\sqrt{\Sigma_{\mathbb{D}_{SR}}\Sigma_{\mathbb{D}_{faces}}}\right) \\
&\quad \gg Tr\left(\Sigma_{\mathbb{D}_{SR}} + \Sigma_{\mathbb{D}_{eyes}} - 2\sqrt{\Sigma_{\mathbb{D}_{SR}}\Sigma_{\mathbb{D}_{eyes}}}\right),
\end{aligned}$$

where $Tr(\cdot)$ denotes the trace of a matrix, $\mu_{(\cdot)}$ the means and $\Sigma_{(\cdot)}$ the respective covariances. After enhancing the overall facial images, we proceed with localized feature extraction focused on the eyes. This includes precise eye localization, cropping, and subsequent analyses such as blink detection, which we can describe as a function φ_{blink} such that $|\mathbb{D}_{eyes}^{M_\theta}| \gg |\varphi_{blink}\left(\mathbb{D}_{eyes}^{M_\theta}\right)|$.

3.1 SR Techniques

Regarding SR methodologies, we identify two primary factors that fundamentally influence the performance and outcomes of SR models M_θ: the architecture of the models and their training objectives to optimize θ (Moser et al., 2024b). Based on the latter, SR models can be broadly categorized into two groups: regression-based models, which typically employ a regression loss, and generative SR models, which utilize adversarial loss mechanisms. These distinctions are crucial as they result in varying SR approximations, which can subsequently impact the accuracy of pupil diameter estimations. To encompass the breadth of techniques available and ensure a comprehensive evaluation, we have selected at least two distinct approaches from each category:

- **Regression-Based Models.**
  - **SRResNet.** A general SR method that draws architectural inspiration from ResNet (He et al., 2016; Ledig et al., 2017).
  - **HAT.** A state-of-the-art vision transformer designed for image SR (Dosovitskiy et al., 2020; Chen et al., 2023b).
- **Generative Models.**
  - **GFPGAN.** A face-oriented SR GAN model designed specifically to enhance facial features within images (Wang et al., 2021b).
  - **CodeFormer.** A face-oriented VQ-VAE based model (Zhou et al., 2022).
  - **Real-ESRGAN.** A more generalized SR GAN approach, which is considered to offer robust solutions for generating photorealistic textures and details in everyday situations (Wang et al., 2022).

3.2 EyeDentify++

As a result of the examination of GFPGAN, CodeFormer, Real-ESRGAN, HAT, and SRResNet SR models for pupil diameter estimation, we can create five additional datasets containing left and right eye images separately, which we call EyeDentify++ [1]. Due to the different SR approximations, the later stages, where we recognize faces, crop eyes, and detect blinks, result in retaining and discarding different amounts of images. More formally, $|\varphi_{blink}\left(\mathbb{D}_{eyes}^{GFPGAN\times2}\right)| \neq |\varphi_{blink}\left(\mathbb{D}_{eyes}^{HAT\times2}\right)|$. Figure 3 compares the number of images in the original dataset with those in the SR datasets after blink detection. The results indicate that SR enhances the accuracy of blink classification by improving the calculation of the EAR ratio through clearer eye landmark detection on the 2x and 4x up-scaled images and providing higher-quality images for feature extraction in the subsequent blink detection phase (Shah et al., 2024).

[1] https://vijulshah.github.io/webcam-based-pupil-diameter-estimation/

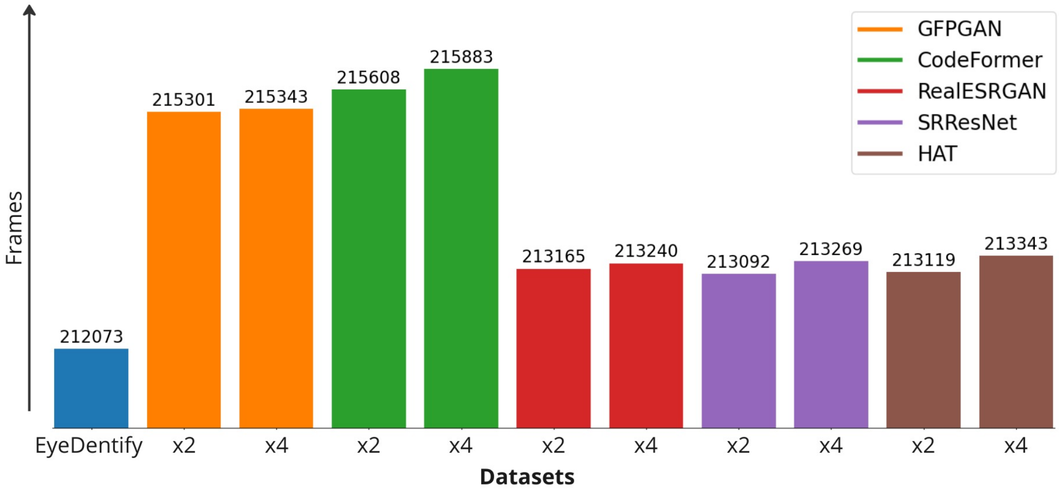

Figure 3: Comparison of applying pre-trained SR models on the EyeDentify Dataset.

4 EXPERIMENTS

In this section, we present our experimental part, which consists of model and training details as well as quantitative and qualitative results.

4.1 Model Details

For pupil diameter prediction, we employed the same regression models as suggested in EyeDentify (Shah et al., 2024): ResNet18, ResNet50, and ResNet152, with the same model configuration and processing steps. The datasets created through SR methods were used to train and evaluate these ResNet models. We upscaled the eye images by 2x and 4x using bi-cubic interpolation to reach 64 x 32 and 128 x 64 dimensions. We then refined the images using SR models (e.g., GFPGAN, CodeFormer, Real-ESRGAN, HAT, and SRResNet).

4.2 Training Details

We followed the training setup from the original work (Shah et al., 2024). Using 5-fold cross-validation, we trained ResNet18, ResNet50, and ResNet152 from scratch on all datasets for 50 epochs, with a batch size of 128, separately for left and right eyes. We used the AdamW optimizer with default settings, a weight decay of 10^{-2}, and an initial learning rate of 10^{-4}, which was reduced by 0.2 every 10 epochs.

5 RESULTS

Table 1 presents 5-fold cross-validation results for ResNet18, ResNet50, and ResNet152 on SRx2 and SRx4 datasets. Compared to the original EyeDentify dataset, we can observe that upscaling greatly benefits pupil diameter prediction.

Scale Sensitivity. Table reveals a complex relationship between the scale factor and the performance of SR methods. There is no consistent trend of improvement or deterioration as the scale increases from ×2 to ×4 across all methods.

Potential Overfitting. Certain SR methods exhibit exceptional performance in specific configurations but perform poorly in others. For instance, while ResNet152 shows improved results with bicubic interpolation at ×2 scale, it tends to overfit with SR at higher scales. This variability could indicate overfitting to particular network architectures, highlighting a need for robustness in classifier selection rather than focusing solely on image enhancement.

Best Models. Across different setups, bicubic upsampling frequently achieves optimal performance for both left and right eyes, particularly notable in the ResNet18 architecture. However, advanced SR methods like Real-ESRGAN and SRResNet also consistently demonstrate lower error rates, underscoring their potential effectiveness in specific configurations. These findings suggest a balanced approach in selecting SR methods, considering both traditional techniques and advanced models based on specific needs.

Visualizations. Figure 5 shows the Class Activation Maps (CAM) (Zhou et al., 2016) from the final convolution layer for each model, tested on a participant viewing the same display color across all datasets.

Table 1: Quantitative Mean Absolute Error (MAE) ↓ comparison across different pre-trained SR methods and pupil diameter prediction models for both left and right eyes. The lowest errors are highlighted.

Eye	Scale	Method	ResNet18	ResNet50	ResNet152
Left	×1	No SR	0.1329 ± 0.0235	0.1280 ± 0.0164	0.1259 ± 0.0176
	×2	Bi-cubic	0.1340 ± 0.0196	0.1402 ± 0.0327	0.1225 ± 0.0166
		GFPGAN	0.1428 ± 0.0360	0.1486 ± 0.0195	0.1339 ± 0.0122
		CodeFormer	0.1328 ± 0.0245	0.1476 ± 0.0364	0.1442 ± 0.0189
		Real-ESRGAN	0.1265 ± 0.0179	0.1369 ± 0.0153	0.1384 ± 0.0195
		SRResNet	0.1286 ± 0.0139	0.1249 ± 0.0062	0.1391 ± 0.0261
		HAT	0.1251 ± 0.0129	0.1277 ± 0.0241	0.1418 ± 0.0197
	×4	Bi-cubic	0.1375 ± 0.0192	0.1382 ± 0.0287	0.1497 ± 0.0275
		GFPGAN	0.1397 ± 0.0244	0.1230 ± 0.0122	0.1348 ± 0.0183
		CodeFormer	0.1383 ± 0.0170	0.1404 ± 0.0201	0.1413 ± 0.0164
		Real-ESRGAN	0.1338 ± 0.0178	0.1306 ± 0.0160	0.1316 ± 0.0183
		SRResNet	0.1384 ± 0.0234	0.1345 ± 0.0163	0.1509 ± 0.0242
		HAT	0.1330 ± 0.01191	0.1305 ± 0.0115	0.1454 ± 0.0179
Right	×1	No SR	0.1548 ± 0.0273	0.1501 ± 0.0214	0.1452 ± 0.0163
	×2	Bi-cubic	0.1402 ± 0.0327	0.1558 ± 0.0214	0.1500 ± 0.0194
		GFPGAN	0.1470 ± 0.0328	0.1628 ± 0.0286	0.1499 ± 0.0130
		CodeFormer	0.1480 ± 0.0188	0.1519 ± 0.0288	0.1542 ± 0.0423
		Real-ESRGAN	0.1505 ± 0.0235	0.1502 ± 0.0154	0.1526 ± 0.0350
		SRResNet	0.1531 ± 0.0213	0.1490 ± 0.0328	0.1391 ± 0.0261
		HAT	0.1477 ± 0.0321	0.1349 ± 0.0226	0.1413 ± 0.0372
	×4	Bi-cubic	0.1383 ± 0.0287	0.1319 ± 0.0222	0.1424 ± 0.0232
		GFPGAN	0.1595 ± 0.0157	0.1559 ± 0.0204	0.1498 ± 0.0137
		CodeFormer	0.1450 ± 0.0152	0.1454 ± 0.0296	0.1441 ± 0.0211
		Real-ESRGAN	0.1396 ± 0.0164	0.1321 ± 0.0375	0.1520 ± 0.0336
		SRResNet	0.1462 ± 0.0234	0.1345 ± 0.0163	0.1446 ± 0.0220
		HAT	0.1489 ± 0.0136	0.1379 ± 0.0198	0.1369 ± 0.0236

The CAM visualizations show that upscaling affects where prediction models focus their attention, with variations in the same image revealing shifts in attention patterns. The top-performing models usually show high activation corresponding to the shape of the eye (see best-performing, boxed examples). Thus, image upscaling influences both the model's focus and its performance.

6 LIMITATIONS

This study faces several challenges, as shown in Figure 4. Participants were recorded in natural postures with varying distances from the webcam and no strict positioning guidelines, leading to inconsistencies like movement (A), gaze shifts (B), head/body turns (C), and actions like talking or smiling (D). Differences in eye structure, skin tone, and iris color across diverse nationalities and demographics make it difficult to generalize the model. Variations in lighting and screen color changes further affect the perceived eye and pupil colors (E, F, G, H). Additionally, Figure 2 and Figure 4 (A, C, E, G, H) highlight that GAN-based models introduce artifacts like glare, altered eye size, and changes in iris color, complicating model training.

7 FUTURE WORK

Future work should explore additional SR methods and incorporate more diverse data conditions to ensure the robustness of pupil diameter estimation in real-world settings. Fine-tuning SR models on eye-cropped datasets like FFHQ (Karras et al., 2018) or CelebA-HQ (Huang et al., 2018) could help SR models adapt to varying lighting conditions, skin tones, and eye structures, improving dataset quality. Although SR methods cannot fully resolve these challenges, they can enhance features, making them more discretely detectable by deep learning models. Com-

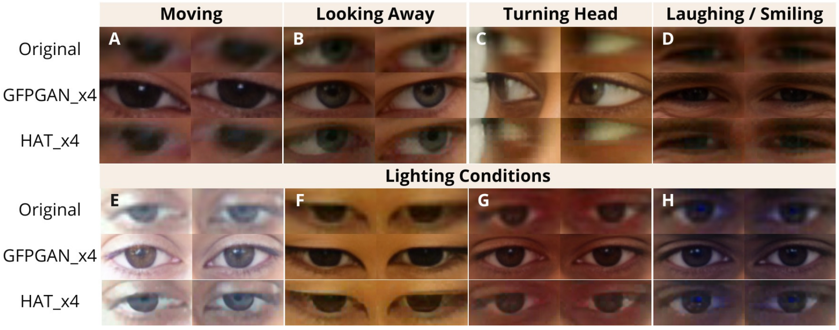

Figure 4: Challenges in estimating pupil diameter without and with SR: Participants A, B, C show head movements and gaze shifts; Participant D shows eye size variation while smiling; Participants E, F, G, H experience different lighting effects—E in bright light, F with a yellow tint, G's face appearing red, and H's face appearing blue.

bining SR with image-to-image translation models like Pix2Net (Jin et al., 2024), which converts RGB images to near-infrared (NIR), could improve feature extraction, particularly in low-contrast scenarios where darker irises make pupil features difficult to detect. Additionally, real-time SR techniques, such as those introduced by (Zhan et al., 2021) and (Shi et al., 2016), could enable mobile and web-based applications for real-time pupilometry without specialized equipment. These advancements will not only enhance the accuracy of eye-tracking technologies but also make them more accessible, laying a strong foundation for future innovations in both pupilometry and eye-tracking technology.

8 CONCLUSION

In this work, we investigated the role of SR techniques in enhancing the accuracy of pupil diameter prediction from webcam-based images, which is crucial for assessing psychological and physiological states. Our experiments, across multiple upscaling methods and neural network architectures, demonstrate that SR can significantly refine the feature details necessary for more precise pupil measurements. Key findings indicate that while the benefits of SR are clear, they are not uniformly distributed across different scales and methods. For instance, although traditional bicubic upscaling often performs well, advanced SR techniques like Real-ESRGAN and SRResNet generally provide superior error rates under specific conditions. In conclusion, while SR presents a promising avenue for enhancing low-quality, webcam-derived images for pupilometry, it requires nuanced application and thorough validation to fully realize its benefits.

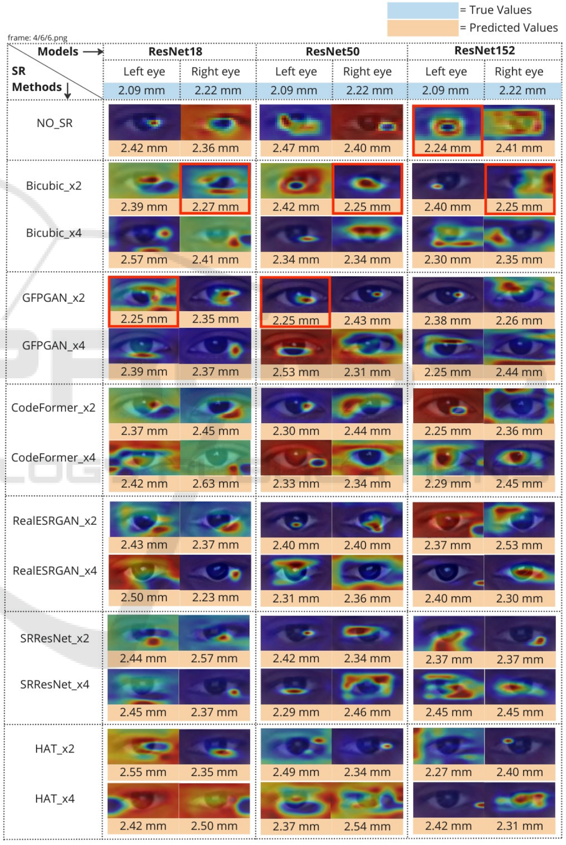

Figure 5: Class Activation Map (Zhou et al., 2016) visualizations for the final convolutional layer of ResNet18, ResNet50, and ResNet152 are shown for a test participant viewing the same display color with No-SR, SRx2, and SRx4 eye images. The true and predicted values represent the original and estimated pupil diameters.

ACKNOWLEDGEMENTS

This work was supported by the DFG International Call on Artificial Intelligence "Learning Cyclotron" (442581111) and the BMBF project SustainML (Grant 101070408).

REFERENCES

Agustsson, E. and Timofte, R. (2017). Ntire 2017 challenge on single image super-resolution: Dataset and study. In *CVPRW*, pages 126–135.

Apple Inc. (2024). Apple announces new accessibility features, including eye tracking, music haptics, and vocal shortcuts. Accessed: 2024-06-06.

Bashir, S. M. A., Wang, Y., Khan, M., and Niu, Y. (2021). A comprehensive review of deep learning-based single image super-resolution. *PeerJ Computer Science*, 7:e621.

Bhatt, A., Watanabe, K., Dengel, A., and Ishimaru, S. (2024). Appearance-based gaze estimation with deep neural networks: From data collection to evaluation. *International Journal of Activity and Behavior Computing*, 2024(1):1–15.

Caya, M. V. C., Jorel P. Rapisura, C., and Despabiladeras, R. R. B. (2022). Development of pupil diameter determination using tiny-yolo algorithm. In *2022 IEEE 14th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*, pages 1–6.

Chen, C.-C., Chen, W.-H., Chiang, J.-S., Chien, C.-T., and Chang, T. (2023a). Semantic segmentation using super resolution technique as pre-processing. In *IET International Conference on Engineering Technologies and Applications (ICETA 2023)*, volume 2023, pages 109–110. IET.

Chen, J., Jia, L., Zhang, J., Feng, Y., Zhao, X., and Tao, R. (2024). Super-resolution for land surface temperature retrieval images via cross-scale diffusion model using reference images. *Remote Sensing*, 16(8):1356.

Chen, X., Wang, X., Zhou, J., Qiao, Y., and Dong, C. (2023b). Activating more pixels in image super-resolution transformer. In *CVPR*, pages 22367–22377.

Dembinsky, D., Watanabe, K., Dengel, A., and Ishimaru, S. (2024a). Eye movement in a controlled dialogue setting. In *Proceedings of the 2024 Symposium on Eye Tracking Research and Applications*, ETRA '24, New York, NY, USA. Association for Computing Machinery.

Dembinsky, D., Watanabe, K., Dengel, A., and Ishimaru, S. (2024b). Gaze generation for avatars using gans. *IEEE Access*, 12:101536–101548.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Elman, J. A., Panizzon, M. S., Hagler Jr, D. J., Eyler, L. T., Granholm, E. L., Fennema-Notestine, C., Lyons, M. J., McEvoy, L. K., Franz, C. E., Dale, A. M., et al. (2017). Task-evoked pupil dilation and bold variance as indicators of locus coeruleus dysfunction. *Cortex*, 97:60–69.

Fathi, A. and Abdali-Mohammadi, F. (2015). Camera-based eye blinks pattern detection for intelligent mouse. *Signal, Image And Video Processing*, 9:1907–1916.

Granholm, E. L., Panizzon, M. S., Elman, J. A., Jak, A. J., Hauger, R. L., Bondi, M. W., Lyons, M. J., Franz, C. E., and Kremen, W. S. (2017). Pupillary responses as a biomarker of early risk for alzheimer's disease. *Journal of Alzheimer's disease*, 56(4):1419–1428.

Greinacher, R. and Voigt-Antons, J.-N. (2020). Accuracy assessment of arkit 2 based gaze estimation. In Kurosu, M., editor, *Human-Computer Interaction. Design and User Experience*, pages 439–449, Cham. Springer International Publishing.

He, C., Xu, Y., Wu, Z., and Wei, Z. (2024). Connecting low-level and high-level visions: A joint optimization for hyperspectral image super-resolution and target detection. *IEEE Transactions on Geoscience and Remote Sensing*.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *CVPR*, pages 770–778.

Hisadome, Y., Wu, T., Qin, J., and Sugano, Y. (2024). Rotation-constrained cross-view feature fusion for multi-view appearance-based gaze estimation. In *WACV*, pages 5985–5994.

Hong, J., Shin, J., Choi, J., and Ko, M. (2024). Robust eye blink detection using dual embedding video vision transformer. In *WACV*, pages 6374–6384.

Huang, H., He, R., Sun, Z., Tan, T., et al. (2018). Introvae: Introspective variational autoencoders for photographic image synthesis. *Advances in neural information processing systems*, 31.

Jiang, T., Yu, Q., Zhong, Y., and Shao, M. (2024). Plantsr: Super-resolution improves object detection in plant images. *Journal of Imaging*, 10(6):137.

Jin, Y., Park, I., Song, H., Ju, H., Nalcakan, Y., and Kim, S. (2024). Pix2next: Leveraging vision foundation models for rgb to nir image translation. *arXiv preprint arXiv:2409.16706*.

Joshi, S., Li, Y., Kalwani, R. M., and Gold, J. I. (2016). Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron*, 89(1):221–234.

Kahneman, D. and Beatty, J. (1966). Pupil diameter and load on memory. *Science*, 154(3756):1583–1585.

Karras, T., Laine, S., and Aila, T. (2018). A style-based generator architecture for generative adversarial networks. arxiv e-prints. *arXiv preprint arXiv:1812.04948*.

Khokhlov, I., Davydenko, E., Osokin, I., Ryakin, I., Babaev, A., Litvinenko, V., and Gorbachev, R. (2020). Tiny-yolo object detection supplemented with geometrical data. In *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, pages 1–5. IEEE.

Kim, J., Oh, J., and Lee, K. M. (2024). Beyond image super-resolution for image recognition with task-driven perceptual loss. In *CVPR*, pages 2651–2661.

Kotani, J., Nakao, H., Yamada, I., Miyawaki, A., Mambo, N., and Ono, Y. (2021). A novel method for measuring the pupil diameter and pupillary light reflex of healthy volunteers and patients with intracranial lesions using a newly developed pupilometer. *Frontiers in Medicine*, 8.

Krejtz, K., Duchowski, A. T., Niedzielska, A., Biele, C., and Krejtz, I. (2018). Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze. *PloS one*, 13(9):e0203629.

Kremen, W. S., Panizzon, M. S., Elman, J. A., Granholm, E. L., Andreassen, O. A., Dale, A. M., Gillespie, N. A., Gustavson, D. E., Logue, M. W., Lyons, M. J., et al. (2019). Pupillary dilation responses as a midlife indicator of risk for alzheimer's disease: association with alzheimer's disease polygenic risk. *Neurobiology of Aging*, 83:114–121.

Kucewicz, M. T., Dolezal, J., Kremen, V., Berry, B. M., Miller, L. R., Magee, A. L., Fabian, V., and Worrell, G. A. (2018). Pupil size reflects successful encoding and recall of memory in humans. *Scientific reports*, 8(1):4949.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In *CVPR*, pages 4681–4690.

Li, X., Ren, Y., Jin, X., Lan, C., Wang, X., Zeng, W., Wang, X., and Chen, Z. (2023). Diffusion models for image restoration and enhancement–a comprehensive survey. *arXiv preprint arXiv:2308.09388*.

Liu, J. (2024). Improving image stitching effect using super-resolution technique. *International Journal of Advanced Computer Science & Applications*, 15(6).

Liu, M., Bian, S., and Lukowicz, P. (2022). Non-contact, real-time eye blink detection with capacitive sensing. In *Proceedings of the 2022 ACM International Symposium on Wearable Computers*, ISWC '22, page 49–53, New York, NY, USA. Association for Computing Machinery.

Lüdtke, H., Wilhelm, B., Adler, M., Schaeffel, F., and Wilhelm, H. (1998). Mathematical procedures in data recording and processing of pupillary fatigue waves. *Vision research*, 38(19):2889–2896.

Micieli, G., Tassorelli, C., Martignoni, E., Pacchetti, C., Bruggi, P., Magri, M., and Nappi, G. (1991). Disordered pupil reactivity in parkinson's disease. *Clinical Autonomic Research*, 1:55–58.

Moser, B. B., Raue, F., Frolov, S., Palacio, S., Hees, J., and Dengel, A. (2023). Hitchhiker's guide to super-resolution: Introduction and recent advances. *IEEE TPAMI*, 45(8):9862–9882.

Moser, B. B., Raue, F., Palacio, S., Frolov, S., and Dengel, A. (2024a). Latent dataset distillation with diffusion models. *arXiv preprint arXiv:2403.03881*.

Moser, B. B., Shanbhag, A. S., Raue, F., Frolov, S., Palacio, S., and Dengel, A. (2024b). Diffusion models, image super-resolution and everything: A survey. *arXiv preprint arXiv:2401.00736*.

Murillo, R., Crucilla, C., Schmittner, J., Hotchkiss, E., and Pickworth, W. B. (2004). Pupillometry in the detection of concomitant drug use in opioid-maintained patients. *Methods and findings in experimental and clinical pharmacology*, 26(4):271–275.

Murphy, P. R., O'connell, R. G., O'sullivan, M., Robertson, I. H., and Balsters, J. H. (2014). Pupil diameter covaries with bold activity in human locus coeruleus. *Human brain mapping*, 35(8):4140–4154.

Mustafa, A., Khan, S. H., Hayat, M., Shen, J., and Shao, L. (2019). Image super-resolution as a defense against adversarial attacks. *arXiv preprint arXiv:1901.01677*.

Na, B. and Fox, G. C. (2020). Object classifications by image super-resolution preprocessing for convolutional neural networks. *Advances in Science, Technology and Engineering Systems Journal (ASTESJ)*, 5(2):476–483.

Ni, Y. and Sun, B. (2019). A remote free-head pupillometry based on deep learning and binocular system. *IEEE Sensors Journal*, 19(6):2362–2369.

O'Shea, G. and Komeili, M. (2023). Toward super-resolution for appearance-based gaze estimation. *arXiv preprint arXiv:2303.10151*.

Pedrotti, M., Mirzaei, M. A., Tedesco, A., Chardonnet, J.-R., Mérienne, F., Benedetto, S., and Baccino, T. (2014). Automatic stress classification with pupil diameter analysis. *International Journal of Human-Computer Interaction*, 30(3):220–236.

Pfleging, B., Fekety, D. K., Schmidt, A., and Kun, A. L. (2016). A model relating pupil diameter to mental workload and lighting conditions. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, page 5776–5788, New York, NY, USA. Association for Computing Machinery.

Reddy, L. F., Reavis, E. A., Wynn, J. K., and Green, M. F. (2018). Pupillary responses to a cognitive effort task in schizophrenia. *Schizophrenia Research*, 199:53–57.

Sari, J. N., Hanung, A. N., Lukito, E. N., Santosa, P. I., and Ferdiana, R. (2016). A study on algorithms of pupil diameter measurement. In *2016 2nd International Conference on Science and Technology-Computer (ICST)*, pages 188–193.

Shah, V., Watanabe, K., Moser, B. B., and Dengel, A. (2024). Eyedentify: A dataset for pupil diameter estimation based on webcam images. *arXiv preprint arXiv:2407.11204*.

Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., Rueckert, D., and Wang, Z. (2016). Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *CVPR*, pages 1874–1883.

Song, L., Wang, Q., Liu, T., Li, H., Fan, J., Yang, J., and Hu, B. (2022). Deep robust residual network for super-resolution of 2d fetal brain mri. *Scientific reports*, 12(1):406.

Tales, A., Troscianko, T., Lush, D., Haworth, J., Wilcock, G., and Butler, S. (2001). The pupillary light reflex in aging and alzheimer's disease. *Aging (Milan, Italy)*, 13(6):473–478.

Tang, J., Zhang, J., Chen, D., Al-Nabhan, N., and Huang, C. (2021). Single-frame super-resolution for remote sensing images based on improved deep recursive residual network. *EURASIP Journal on Image and Video Processing*, 2021:1–19.

Timofte, R., Agustsson, E., Van Gool, L., Yang, M.-H., and Zhang, L. (2017). Ntire 2017 challenge on single image super-resolution: Methods and results. In *CVPRW*, pages 114–125.

Van Den Brink, R. L., Murphy, P. R., and Nieuwenhuis, S. (2016). Pupil diameter tracks lapses of attention. *PloS one*, 11(10):e0165274.

Wang, X., Li, Y., Zhang, H., and Shan, Y. (2021a). Towards real-world blind face restoration with generative facial prior. In *CVPR*, pages 9168–9178.

Wang, X., Li, Y., Zhang, H., and Shan, Y. (2021b). Towards real-world blind face restoration with generative facial prior. In *CVPR*.

Wang, X., Xie, L., Dong, C., and Shan, Y. (2022). Realesrgan: Training real-world blind super-resolution with pure synthetic data supplementary material. *Computer Vision Foundation open access*, 1(2):2.

Yun, J.-S., Na, Y., Kim, H. H., Kim, H.-I., and Yoo, S. B. (2022). Haze-net: High-frequency attentive super-resolved gaze estimation in low-resolution face images. In *ACCV*, pages 3361–3378.

Zhan, Z., Gong, Y., Zhao, P., Yuan, G., Niu, W., Wu, Y., Zhang, T., Jayaweera, M., Kaeli, D., Ren, B., et al. (2021). Achieving on-mobile real-time super-resolution with neural architecture and pruning search. In *ICCV*, pages 4821–4831.

Zhang, Z. (1999). Flexible camera calibration by viewing a plane from unknown orientations. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*, volume 1, pages 666–673 vol.1.

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., and Torralba, A. (2016). Learning deep features for discriminative localization. In *CVPR*, pages 2921–2929.

Zhou, S., Chan, K., Li, C., and Loy, C. C. (2022). Towards robust blind face restoration with codebook lookup transformer. *NeurIPS*, 35:30599–30611.