2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

Unsupervised Local Editing of Ocular Images via Reverse-Attention Block

Renzhong Wu[1], Shenghui Liao[1]✉, Jianfeng Li[2], Lihong Liu[3], Xiaoyan Kui[1] and Yongrong Ji[4]

[1] School of Computer Science and Engineering, Central South University, Changsha 410083, China
[2] School of Computer Science and Engineering, Jishou University, Jishou 416000, China
[3] Department of Rehabilitation, The Second Xiangya Hospital, Central South University, Changsha 410011, China
[4] Department of Ophthalmology, Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, 200025, China

Abstract. Predicting the postoperative appearance after strabismus surgery is of great significance for improving patients' understanding of the postoperative outcomes, enhancing communication between doctors and patients, and alleviating preoperative anxiety. Although some researchers have used image generation models to predict postoperative appearances for various diseases, existing methods typically rely on paired data for model training. Generative Adversarial Networks (GANs) have demonstrated strong application potential in image generation tasks, and cycle consistency loss has promoted the development of unsupervised image generation techniques. However, traditional cycle consistency loss often results in the retention of unnecessary traces from the source image in the generated images. To address these issues, we propose an unsupervised image generation model based on GANs. By incorporating a reverse-attention block into the generator, the model is guided to focus on key editing regions. Additionally, we employ reverse-attention consistency loss to maintain identity consistency while reducing unnecessary trace residues. Furthermore, we introduce a multi-scale discriminator to ensure that the generated images have more reasonable texture details. Experimental results demonstrate that our model effectively reduces trace residues in the generated postoperative images and produces details that are more consistent with reality.

Keywords: Image-to-image Translation, Reverse-attention Consistency Loss, Unsupervised Learning, Generation Adversarial Networks.

1 Introduction

Strabismus refers to a condition where the eyes cannot simultaneously focus on the same target, resulting in misaligned vision [1,2]. Strabismus may also be accompanied by a certain degree of ptosis (drooping of the upper eyelid), which not only affects the patient's appearance but may also lead to a range of visual and psychological issues.

Therefore, early diagnosis and treatment of strabismus are crucial. By presenting predicted postoperative visual effects, it can enhance the patient's understanding and trust in the treatment process, thereby reducing anxiety and unease [3].

As computer technology advances, image generation models have garnered significant attention [4-6]. Many researchers utilize image generation models to predict the postoperative appearance for various diseases [7-9]. Despite achieving good prediction results, these models typically require paired data for training. However, obtaining paired data often demands more human and material resources. Since unsupervised models do not require paired data, a feasible approach is to design an unsupervised generative model to predict postoperative appearance.

Generative Adversarial Networks (GANs) [10] models have demonstrated outstanding image generation capabilities in various tasks [11-13]. CycleGAN [14] introduced the concept of cycle consistency loss, enabling the learning of image transformations from unpaired datasets. As an unsupervised model, CycleGAN does not require additional conditional information and does not demand data alignment during the training phase. Subsequently, many mainstream models (such as: MUNIT [15], DRIT++ [16], starGAN V2 [17]) adopted the cycle consistency loss. However, the cycle consistency loss, in order to ensure that the generated images can be reconstructed into the original images, leaves traces of the source image in the generated target images (as shown in Fig. 1). To address the trace issue, ACL-GAN [18] proposed adversarial consistency loss, which effectively removes the traces. However, the quality of the generated images is not ideal. Additionally, these methods only consider domain-to-domain transformations and do not focus on whether the details of the generated images are reasonable.

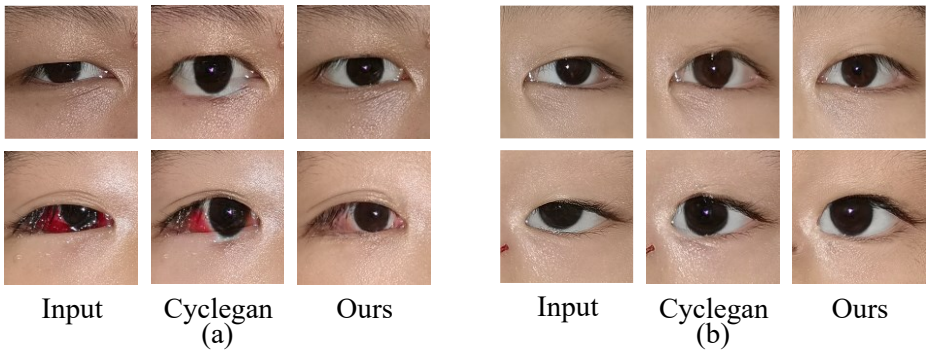

Fig. 1. Examples of existing issues in baseline.

To address these issues, we have designed an unsupervised image generation model to predict the postoperative appearance of strabismus surgery. This model can maintain consistent identity information while reducing unnecessary traces, resulting in more natural predicted images. It enables patients to better understand the postoperative outcomes, enhances communication between doctors and patients, and reduces unnecessary anxiety.

Our main contributions are as follows:

1. We propose a novel unsupervised model for local editing of strabismus images. The core of this method lies in the introduction of the reverse-attention block and the reverse attention consistency loss.

2. The design of the reverse-attention block enables the generation model to precisely locate the areas that require editing. Through the reverse-attention mechanism, the model can dynamically adjust its focus, thereby avoiding interference with irrelevant regions during the editing process. This mechanism not only enhances the accuracy of the local edit but also strengthens the model's robustness, enabling it to handle complex strabismus images effectively.

3. The proposed reverse-attention consistency loss aids the generation model in producing images that retain identity information while minimizing unnecessary traces. This loss function's design allows the model to balance local edit with global consistency during the editing process, resulting in more natural images.

4. The use of a multi-scale discriminator assists the model in evaluating the quality of generated images across multiple scales. This approach facilitates a better capture of both local details and global structures, leading to generated images with more plausible texture details.

2 Method

Our goal is to design an unsupervised model to predict the postoperative outcomes of strabismus. The model architecture is illustrated in Fig. 2.

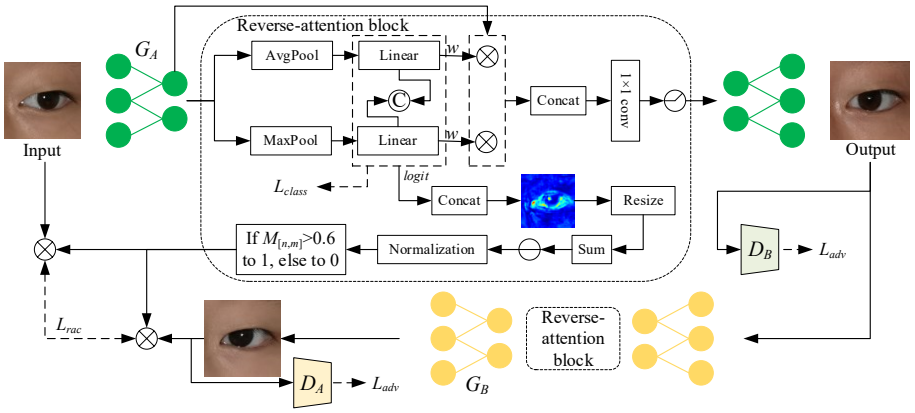

Fig. 2. Overall architecture.

Our model architecture employs an adversarial framework with dual generators and dual discriminators, specifically including Generator A, Generator B, Discriminator A, and Discriminator B. Among these, Generator A is responsible for mapping preoperative strabismus images to postoperative images, while Generator B performs the reverse transformation, reconstructing postoperative images back to preoperative ones. The two discriminators are respectively tasked with assessing the authenticity of images: Discriminator A focuses on distinguishing real preoperative images from the reconstructed images generated by Generator B, while Discriminator B is responsible for discerning real postoperative images from the predicted images generated by Generator A.

To further enhance the model's performance, we introduced a reverse attention consistency loss. This loss function not only effectively preserves the identity information of the input images but also significantly reduces artifacts and residual traces that may arise during the image transformation process. Through this constraint mechanism, the model maintains the authenticity of the generated images while ensuring the smoothness and naturalness of the transformation process.

2.1 Generators and discriminators

In terms of generator design, we propose an innovative reverse attention mechanism module (reverse-attention blocks). This module enhances the model's ability to extract features from key areas, particularly demonstrating significant advantages in handling the fine structures of the eye region. Additionally, it effectively reduces artifacts in the generated images.

The discriminator employs a multi-scale architecture, consisting of two parallel sub-discriminators: the Global Discriminator evaluates the overall consistency of the image, while the Local Discriminator focuses on the authenticity of specific regions, such as the eye area. This design draws inspiration from the core idea of the pix2pix framework, helping the generator produce higher quality images, ensuring both the overall structural rationality and enhancing the realism of local details.

2.2 Reverse-attention block

When real preoperative and postoperative images pass through the generator, the generator is typically able to capture the differences between the two domains. However, these regions of focus are often relatively scattered and lack concentration on critical areas. Additionally, the cycle consistency loss requires that the generated postoperative images be able to reconstruct the source images, which to some extent results in unnecessary source image traces in the generated postoperative images.

To address the above issues, we propose the reverse-attention blocks. As shown in Algorithm 1, this module has two outputs. The first output guides the generator to focus on the key regions where preoperative and postoperative images differ. The second output obtains the reverse-attention regions, restricting the model to only consider consistency in the non-critical regions where preoperative and postoperative images differ.

Algorithm 1：Reverse-attention block

Input：Real images A and B, feature map *x* obtained from generator downsampling

Output:

1：Enhanced feature map x''

2：Reverse attention matrix *M*

Procedure:

1. Build a classification network:

Process input through two parallel branches:

✧ Branch 1: Global Average Pooling (GAP) → Fully Connected Layer (FC)

✧ Branch 2: Global Max Pooling (GMP) → Fully Connected Layer (FC)

Concatenate features from both branches

Perform classification using concatenated features

2. Feature Enhancement:

Obtain weights *w* from GAP-FC and GMP-FC layers

Concatenate the weights *w*

Apply concatenated weights to original feature map *x*

Generate enhanced feature map x' incorporating both average and max pooling information

3. Feature Transformation (Output 1):

Apply 1×1 convolution to x'

Process through ReLU activation function

Generate optimized feature map x''

4. Generation Reverse Attention Matrix M (Output 2):

Obtain logit from from GAP-FC and GMP-FC layers

Concatenate logit to get L

Apply resize, summation, negation, and normalization operations to L to get *N*

Apply thresholding (predefined threshold = 0.6):

$$M_{[n,m]} = \begin{cases} 1, N_{[n,m]} > \text{ threshold} \\ 0, N_{[n,m]} \leq \text{ threshold} \end{cases}$$

Obtain final reverse attention matrix *M*.

As shown in Algorithm 1, after the true preoperative and postoperative images pass through the downsampling stage of the generator, features x are obtained. These features x are then processed by an additional classification network to capture the importance of each feature, thereby identifying the key regions that distinguish preoperative and postoperative images. Subsequently, convolution and ReLU operations are applied to enhance the features, resulting in Output 1. Output 2 utilizes a reverse-attention mechanism to exclude key regions with strong attention while recording the coordinates of all non-key regions and assigning a value of 1 to these coordinates, with the remaining regions assigned a value of 0. In the Reverse-attention Consistency Loss described in Section 2.2, these binary matrices are multiplied element-wise with the true preoperative image and the reconstructed preoperative image generated from the synthetic postoperative image derived from the true preoperative image. Finally, the Manhattan distance between these two results is calculated. The advantage of this design lies in the fact that Output 1 strengthens the features of key regions, while Output 2 preserves

identity information effectively while reducing artifacts in the generated images. The reverse-attention block enhances the realism and detail quality of the generated images.

2.3 Loss function

We use the following loss functions to train our model. Here, x_A represents the real preoperative image, and x_B represents the real postoperative image.

Adversarial loss. We employ the Least Squares loss [19] to train the generator and discriminator. The equations of Adversarial loss ($\mathcal{L}_{adv}$) are shown in equation (1) and equation (2).

$$\mathcal{L}_{adv}^{\mathcal{A}\to\mathcal{B}} = E_{x_B}\left[\left(D_B(x_B)\right)^2\right] + E_{x_A}\left[\left(1 - D_B\left(G_{A\to B}(x_A)\right)\right)^2\right] \tag{1}$$

$$\mathcal{L}_{adv}^{\mathcal{B}\to\mathcal{A}} = E_{x_A}\left[\left(D_A(x_A)\right)^2\right] + E_{x_B}\left[\left(1 - D_A\left(G_{B\to A}(x_B)\right)\right)^2\right] \tag{2}$$

Class loss. The loss of the classifier ($\mathcal{L}_{class}$) in the Reverseattention block are shown in equation (3) and equation (4).

$$\mathcal{L}_{class}^{\mathcal{A}\to\mathcal{B}} = -\left(E_{x_A}\left[\log\left(f_A(x_A)\right)\right] + E_{x_B}\left[\log\left(1 - f_A(x_B)\right)\right]\right) \tag{3}$$

$$\mathcal{L}_{class}^{\mathcal{B}\to\mathcal{A}} = -\left(E_{x_B}\left[\log\left(f_B(x_B)\right)\right] + E_{x_A}\left[\log\left(1 - f_B(x_A)\right)\right]\right) \tag{4}$$

Reverse-attention consistency loss. To reduce unnecessary traces of the source image in the generated image, we designed the Reverse-attention Consistency Loss ($\mathcal{L}_{rac}$). First, we obtain the reverse-attention matrix *M* (a binary matrix) through Algorithm 1 in Section 2.2 Reverse-attention block. Second, we perform element-wise multiplication of these binary matrices with the real preoperative image and the reconstructed preoperative image. Finally, we compute the Manhattan distance between the two results. The advantage of this design is that it effectively preserves identity information while significantly reducing unnecessary consistency constraints, thereby minimizing the residual traces of the original image in the generated image and enhancing the realism and detail quality of the generated image. $\mathcal{L}_{rac}$ is defined as follows:

$$\mathcal{L}_{rac}^{\mathcal{A}\to\mathcal{B}} = E_{x_A}\left[\left|x_A \odot M - G_{B\to A}\left(G_{A\to B}(x_A)\right) \odot M\right|_1\right] \tag{5}$$

$$\mathcal{L}_{rac}^{\mathcal{B}\to\mathcal{A}} = E_{x_B}\left[\left|x_B \odot M - G_{A\to B}\left(G_{B\to A}(x_B)\right) \odot M\right|_1\right] \tag{6}$$

Where *M* is the reverse-attention matrix, $\odot$ denotes element-wise multiplication, and $|\cdot|_1$ denotes the Manhattan distance.

Identity loss. To preserve identity information and prevent mode collapse during training, we employ Identity loss ($\mathcal{L}_{id}$), as shown in the following formula:

$$\mathcal{L}_{id}^{A\to B} = E_{x_B}[|x_B - G_{A\to B}(x_B)|_1] \tag{7}$$

$$\mathcal{L}_{id}^{B\to A} = E_{x_A}[|x_A - G_{B\to A}(x_A)|_1] \tag{8}$$

Total loss. Summary, our total loss is as follows:

$$\mathcal{L}_{total} = \min_{G,f}\max_D(\mathcal{L}_{adv} + 10\mathcal{L}_{rac} + 1000\mathcal{L}_{class} + 10\mathcal{L}_{id}) \tag{9}$$

3 Experiments

3.1 Dataset

The data we used was provided by Shanghai Ninth People's Hospital. Following our preprocessing, the dataset comprises 1345 images, including 757 abnormal eye images and 588 normal eye images. An example of the dataset is illustrated in Fig. 3. The abnormal images are categorized into three types: only strabismus, only ptosis, and those exhibiting both strabismus and ptosis. Additionally, 100 images were randomly selected from the dataset to serve as test images.

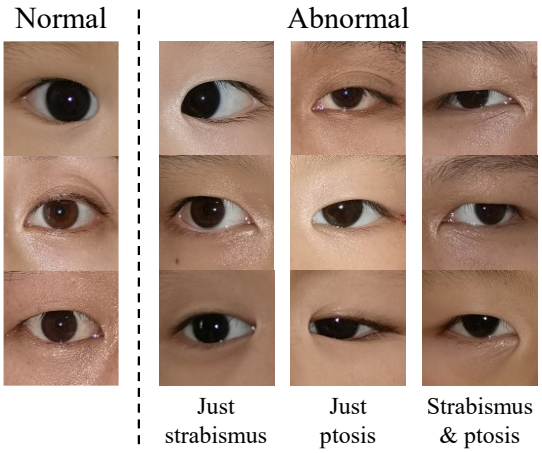

Fig. 3. Dataset samples.

3.2 Qualitative evaluation

To visually evaluate the effectiveness of our proposed method in generating postoperative images, we selected several classical models as our comparison baselines, including g ACL-GAN [18], MUNIT [15], CycleGAN [14], DRIT++ [16], and GP-UNIT [20]. Each model generates postoperative images based on the Input image, and the results are shown in Fig. 4.

Qualitative evaluation results demonstrate that our method outperforms the comparative baselines. From Fig. 4, the following conclusions can be drawn:

1. Image Realism: The images generated by our method most closely resemble real eye images. In contrast, while ACL-GAN-generated images exhibit typical human eye features in shape, they lack realistic skin texture details.

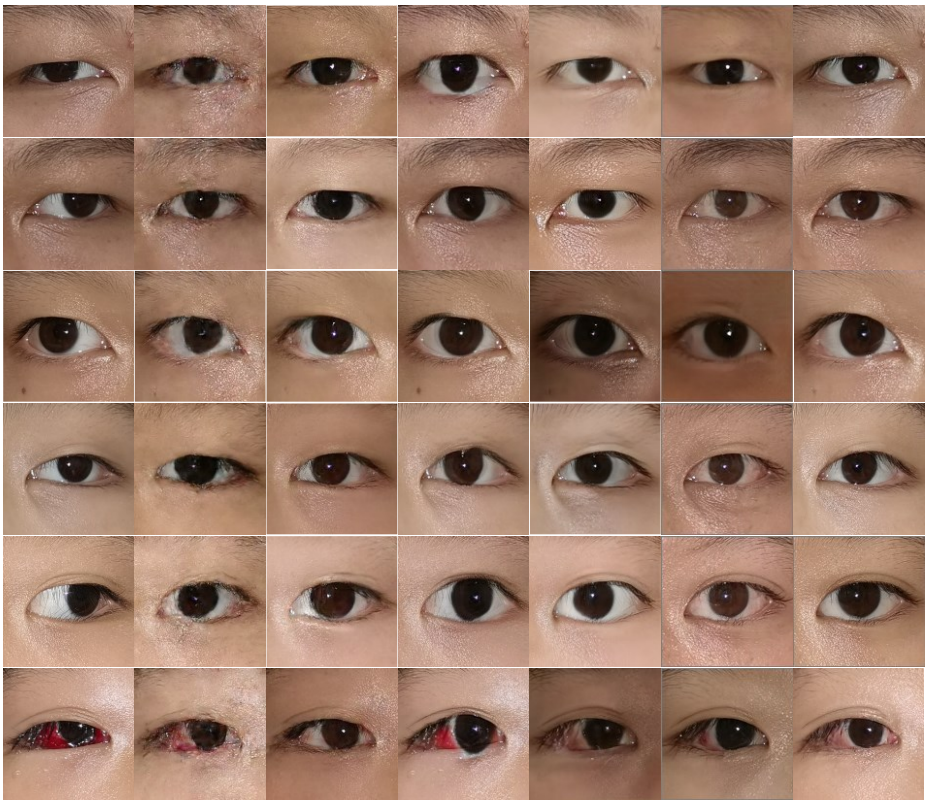

Fig. 4. Visual comparisons.

2. Detail Accuracy: The images generated by our method have details that most closely match real eye images. CycleGAN-generated images show noticeable traces of the input image that are unnecessary (especially in the first row), significantly affecting overall visual quality and detail accuracy. Additionally, MUNIT-generated images exhibit obvious artifacts and insufficient visual clarity. DRIT++-generated postoperative images show a certain deviation in iris shape compared to real iris shapes (particularly evident in the fifth row).

3. Identity Information Preservation: Our method performs outstandingly in preserving patient identity information. Skin color, skin texture, and periorbital features are effectively retained in the generated postoperative images, ensuring image authenticity. While CycleGAN also demonstrates excellent performance in preserving identity information, the generated images still have a certain gap compared to real eye images. The postoperative images generated by DRIT++ and GP-UNIT show significant differences in skin color and texture information compared to the input images.

4. Gaze Conversion Effect: In terms of converting strabismus to orthophoria, our method demonstrates exceptional results. The generated postoperative images show a

natural orthophoric state of the patient's eyes, with eyeball positions consistent with physiological characteristics.

In summary, through comparisons with models such as ACL-GAN, CycleGAN, MUNIT, DRIT++, and GP-UNIT, our method demonstrates significant advantages in image realism, detail accuracy, identity information preservation, and gaze conversion effects. These comprehensive advantages make our method stand out as the optimal solution for postoperative image generation tasks.

3.3 Quantitative evaluation

The Kernel Inception Distance (KID) is a metric used to measure the similarity between generated images and real images, with smaller values indicating that the quality of generated images is closer to that of real images. We evaluated the quality of postoperative images generated by ACL-GAN, CycleGAN, MUNIT, DRIT++, and GP-UNIT models using the KID metric. The results are shown in Table 1.

Table 1. Quantitative results of test datasets.

Model	KID
Cyclegan	0.1667±9.0729e-07
MUNIT	0.5008±1.1145e-06
DRIT++	0.3118±7.4429e-07
ACL-GAN	1.3917±7.4332e-07
GP-UNIT	1.4833±9.6052e-07
Ours	**0.1297±1.1857e-06**

From Table 1, we can see that the KID values of classic models such as CycleGAN, MUNIT, DRIT++, ACL-GAN, and GP-UNIT are 0.1667±9.0729e-07, 0.50081±1.1145e-06, 0.3118±7.4429e-07, 1.3917±7.4332e-07, and 1.4833±9.6052e-07, respectively. In comparison, the KID value of our proposed method (Ours) is 0.1297±1.1857e-06, which is significantly lower than that of the other models. This indicates that, compared to existing classic models, our proposed method has a significant advantage in the quality of generated postoperative images.

Additionally, the KID values of CycleGAN and DRIT++ are 0.1667 and 0.3118, respectively, indicating a certain gap in image generation quality. The KID value of MUNIT is 0.50081, which is much higher than that of other models. This may be related to its multi-modal generation characteristics but also reflects its limitations in generating postoperative images. The KID values of ACL-GAN and GP-UNIT are 1.3917 and 1.4833, respectively, which are significantly higher than those of other models, indicating poor performance in the task of generating postoperative images.

Our proposed method (Ours) achieves the lowest KID value (0.1297), indicating that the postoperative images it generates are closer to real images in terms of quality, with higher reliability and practicality. This result validates the effectiveness of our method in the task of generating postoperative images.

Since the KID value has some deviation from human subjective perception, we further conducted a subjective assessment experiment to validate the quality of the postoperative images generated by the models. Specifically, we invited three volunteers to

vote on the effects of postoperative images generated by different models. In the experiment, volunteers subjectively evaluated different models based on the postoperative image effects generated from the same input image and voted for the model they believed to have the best effect and retain the most identity information from the input image. The experiment included a total of 100 input images, and each volunteer needed to evaluate and vote on the generated results corresponding to all input images. Table 2 shows the voting statistics for each volunteer's evaluation of the postoperative image effects generated by each model.

Table 2. Subjective voting record for test datasets.

Model	1	2	3	Average
Cyclegan	21	16	14	17
MUNIT	8	13	8	9.67
DRIT++	24	25	23	24
ACL-GAN	0	0	0	0
GP-UNIT	20	14	27	20.33
Ours	**27**	**32**	**28**	**29**

From the table 2, it can be seen that Volunteer 1 gave 27 votes to the Ours model, Volunteer 2 gave 32 votes, Volunteer 3 gave 28 votes, with an average of 29 votes, which is significantly higher than the other models. This indicates that the Ours model received the highest satisfaction among the volunteers. The average votes for other models such as CycleGAN, MUNIT, DRIT++, and GP-UNIT were 17, 9.67, 24, and 20.33, respectively, all of which are lower than the Ours model. ACL-GAN received 0 votes, indicating that volunteers considered its generation effect to be the worst.

This result is consistent with our previous analysis based on the KID metric. The KID metric showed that the postoperative images generated by the Ours model were closest to the real images, demonstrating the highest reliability and practicality. The volunteers' votes further validated the superiority of the Ours model in subjective evaluations.

In summary, the Ours model not only performed outstandingly in objective metrics but also received high recognition from volunteers in subjective evaluations. This indicates that the Ours model has high practicality and reliability in generating postoperative images.

4 Limitations and discussion

In this study, we propose a novel framework for local editing of strabismus images. The core of our model consists of the reverse-attention block and the reverse-attention consistency loss. Specifically, the reverse-attention block enhances editing accuracy by guiding the model to effectively focus on key editing areas. Meanwhile, the reverse-attention consistency loss addresses the issue of residual source image traces in generated images, while maintaining identity information, thus solving the problem caused

by traditional cycle consistency loss. Furthermore, our model exhibits superior performance in detail representation, aligning more closely with practical demands and enhancing the naturalness and realism of generated images.

Although our method outperforms state-of-the-art approaches in experimental results, it still has limitations. As shown in Fig. 5, our model generates collapsed images for cases with excessively severe ptosis. The reasons include: on one hand, the lack of data diversity and data imbalance; on the other hand, our model's inability to constrain the generated shapes. In the future, we aim to enhance the quantity and diversity of our dataset and introduce additional masks to constrain the generated shapes.

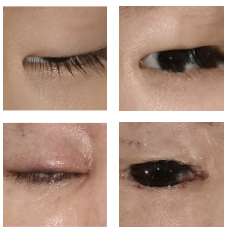

Fig. 5. Case of failure by our model.

Acknowledgments. This study was funded by National Natural Science Foundation of China (grant numbers No. 62372475, No. U22A2034, No. 62177047).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Hennein, L., Robbins, S.L.: Thyroid-associated orbitopathy: management and treatment. Journal of Binocular Vision and Ocular Motility **72**(1), 32–46 (2022)
2. Wenner, Y., Schneider, J., Drisser, T., Yang, Y.Y., Demeter, T., Fronius, M., Lorenz, B., Kohnen, T., Müller, M., Triesch, J.: Virtual reality-based Harms tangent screen test for strabismus measurement. Graefe's Archive for Clinical and Experimental Ophthalmology, 1–8 (2025)
3. Cai, H., Xu, S.: A randomized trial of psychological intervention on perioperative anxiety and depression of patients with severe blepharoptosis undergoing autologous fascia lata frontal muscle suspension. Annals of Palliative Medicine **10**(3), 3185193–3183193 (2021)
4. Xiao, G., Yin, T., Freeman, W.T., Durand, F., Han, S.: Fastcomposer: Tuning-free multi-subject image generation with localized attention. International Journal of Computer Vision, 1–20 (2024)
5. Peng, J., Qiu, R.L., Wynne, J.F., Chang, C.W., Pan, S., Wang, T., Roper, J., Liu, T., Patel, P.R., Yu, D.S., Yang, X.: CBCT-Based synthetic CT image generation using conditional denoising diffusion probabilistic model. Medical Physics **51**(3), 1847–1859 (2024)
6. Boroujeni, S.P.H., Razi, A.: Ic-gan An improved conditional generative adversarial network for rgb-to-ir image translation with applications to forest fire monitoring. Expert Systems with Applications **238**, 121962 (2024)

7. Sun, Y., Huang, X., Zhang, Q., Lee, S.Y., Wang, Y., Jin, K., Lou, L., Ye, J.: A fully automatic postoperative appearance prediction system for blepharoptosis surgery with image - based deep learning. Ophthalmology Science **2**(3), 100169 (2022)
8. Yoo, T.K., Choi, J.Y., Kim, H.K.: A generative adversarial network approach to predicting postoperative appearance after orbital decompression surgery for thyroid eye disease. Computers in Biology and Medicine **118**, 103628 (2020)
9. Huang, X., Li, Z., Lou, L., Dan, R., Chen, L., Zeng, G., Jia, G., Chen, X., Jin, Q., Ye, J., Wang, Y.: GOMPS: global attention-based ophthalmic image measurement and postoperative appearance prediction system. Expert Systems with Applications **232**, 120812 (2023)
10. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. Advances in Neural Information Processing Systems **27** (2014)
11. Zhou, T., Li, Q., Lu, H., Cheng, Q., Zhang, X.: GAN review: Models and medical image fusion applications. Information Fusion **91**, 134–148 (2023)
12. Wang, W., Zhang, M., Wu, Z., Zhu, P., Li, Y.: Scgan: Semi-centralized generative adversarial network for image generation in distributed scenes. Information Fusion **112**, 102556 (2024)
13. Chen, Y., Lin, H., Zhang, W., Chen, W., Zhou, Z., Heidari, A.A., Chen, H., Xu, G.: ICycleGAN: Improved cycle generative adversarial networks for liver medical image generation. Biomedical Signal Processing and Control **92**, 106100 (2024)
14. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2223–2232 (2017)
15. Huang, X., Liu, M.Y., Belongie, S., Kautz, J.: Multimodal unsupervised image-to-image translation. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 172–189 (2018)
16. Lee, H.Y., Tseng, H.Y., Huang, J.B., Singh, M., Yang, M.H.: Diverse image-to-image translation via disentangled representations. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 35–51 (2018)
17. Choi, Y., Uh, Y., Yoo, J., Ha, J.W.: Stargan v2: Diverse image synthesis for multiple domains. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8188–8197 (2020)
18. Zhao, Y., Wu, R., Dong, H.: Unpaired image-to-image translation using adversarial consistency loss. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX, vol. 16, pp. 800–815. Springer International Publishing (2020)
19. Mao, X., Li, Q., Xie, H., Lau, R.Y., Wang, Z., Paul Smolley: Least squares generative adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2794-2802 (2017)
20. Yang, S., Jiang, L., Liu, Z., Loy, C.C.: Gp-unit: Generative prior for versatile unsupervised image-to-image translation. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(10), 11869–11883 (2023)