scientific reports

Check for updates

OPEN

Versatile cataract fundus image restoration model utilizing unpaired cataract and high-quality images

Zheng Gong[1,3], Zhuo Deng[1,3], Weihao Gao[1], Wenda Zhou[2], Yuhang Yang[2], Hanqing Zhao[2], Lei Shao[2], Wenbin Wei[2] & Lan Ma[1]✉

Cataract is one of the most common blinding eye diseases and can be treated by surgery. However, because cataract patients may also suffer from other blinding eye diseases, ophthalmologists must diagnose them before surgery. The cloudy lens of cataract patients forms a hazy degeneration in the fundus images, making it challenging to observe the patient's fundus vessels, which brings difficulties to the diagnosis process. To address this issue, this paper establishes a new cataract image restoration method named Catintell. It contains a cataract image synthesizing model, Catintell-Syn, and a restoration model, Catintell-Res. Catintell-Syn uses GAN architecture with fully unsupervised data to generate paired cataract-like images with realistic style and texture rather than the conventional Gaussian degradation algorithm. Meanwhile, Catintell-Res is an image restoration network that can improve the quality of real cataract fundus images using the knowledge learned from synthetic cataract images. Extensive experiments show that Catintell-Res outperforms other cataract image restoration methods in PSNR with 39.03 and SSIM with 0.9476. Furthermore, the universal restoration ability that Catintell-Res gained from unpaired cataract images can process cataract images from various datasets. We hope the models can help ophthalmologists identify other blinding eye diseases of cataract patients and inspire more medical image restoration methods in the future.

The cataract is one of the most common causes of blindness. The World Health Organization estimates that cataracts will result in 40 million blindness in 2025[1]. Cataracts are typically caused by the deposition of proteins and form clouding of the lens in the eye. Cataracts usually develop with age but can also be caused by external factors such as trauma, diabetes, prolonged use of certain medications, or exposure to ultraviolet radiation. As cataracts grow, they can cause symptoms such as cloudy or blurred vision, faded colors, glare, poor night vision, and double vision.

Furthermore, cataracts also cause blurry clouding in retinal fundus photographing images and affect the diagnosis of other ophthalmic diseases through this method. Fundus images have been expansively used in the fundus disease clinical diagnosis or computer-aided diagnosis systems. Since cataracts can cause lens opacity, the fundus images of cataract patients will suffer from fogging, blurring, and other degradation. It is challenging to make clinical diagnoses through low-quality cataract fundus images. Therefore, the low-quality fundus images could result in the risk of misdiagnosis and uncertainty in preoperative planning.

Fundus image restoration can effectively solve the fundus image degradation caused by cataracts. Research in fundus image restoration has been carried out for many years. Traditional fundus image restoration methods[3–6] are mainly based on handcrafted priors. However, these methods achieve poor performance in clinical applications due to their limited prior knowledge or poor generalization ability.

Recently, deep Convolutional Neural Networks(CNNs)[7–10] have been used in natural image restoration and achieved impressive results. CNNs have introduced into fundus image restoration due to the success in nature image restoration[11–15]. Meanwhile, the Transformer[16] has been introduced into fundus image restoration to address the limitations in capturing long-range dependencies and achieve remarkable performance. The advantage of the Transformer is capturing long-range dependencies. The effective combination of CNNs and

[1]Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, China. [2]Beijing Tongren Eye Center, Beijing Key Laboratory of Intraocular Tumor Diagnosis and Treatment, Beijing Ophthalmology and Visual Sciences Key Lab, Beijing Tongren Hospital, Capital Medical University, Beijing 100730, China. [3]These authors contributed equally: Zheng Gong and Zhuo Deng. ✉email: malan@sz.tsinghua.edu.cn

Transformers may further improve the restoration performance of deep-learning models in cataract image restoration.

Since deep learning methods are mostly data-driven, existing cataract image restoration methods rely on a large number of cataracts and corresponding clear fundus image pairs. However, practical difficulties appear in cataract fundus image collecting. The degradation of cataract images is pathological, which means that clear images must be collected after surgeries to remove the clouding in the lenses. Nevertheless, collecting fundus images is not necessary after cataract surgery and may cause further damage to patients. Therefore, few cataract-clear image pairs were collected for now. Some cataract patients may have corresponding clear fundus images due to surgery follow-up, but, the long time gap of image collecting reduces the significance of these image pairs. There remains a lack of paired cataract images and clear images.

To get training image pairs, the artificial degradation algorithm[2] was first brought out in 1989 and is used in many works even till now. Other models such as Gaussian filters[9,12,13] are designed to synthesize cataract-like images from high-quality (HQ) fundus images. However, these models can barely achieve good performance due to simple design. As shown in Fig. 4b, these cataract-like images fundamentally differ from real clinical cataract images.

In this paper, we set out to address the cataract image restoration problem. To alleviate the issue of lack of data, we propose a new cataract-like image synthesizing model, Catintell-Syn, which is a GAN model that uses fully unsupervised data to generate paired cataract-like images with realistic style and texture. Based on these simulated images, we develop a novel cataract fundus image restoration method, Catintell-Res, including a CNN-based generator and a Transformer-based discriminator. Specifically, the basic unit in the generator is the Dense Convolution Block(DCB), which can capture local degradation features effectively. Unlike the generator, the basic unit of the discriminator is the Window-based Self-attention Block(WSB). The self-attention mechanism captures the non-local self-similarity and long-range dependencies, which can complement the shortcomings of CNNs. The Transformer-based discriminator can indirectly allow the generator to focus on non-local features through its classification ability with GAN architecture. Furthermore, the visual synthetic degradation comparison results show that the cataract-like images synthesized by our Catintell-Syn are closest to real cataract images in degradation style. Extensive experiments demonstrate that the Catintell-Res achieves remarkable performance in both synthetic cataract-like data and real cataract data.We applied numerical metrics, the AI-based fundus image quality assessment method, and a user study to evaluate our method comprehensively. Finally, Catintell-Res is applied to real cataract images from various external datasets to verify its generalization performance and proved effective.

Our contributions can be summarized as follows:

1. We propose a new image synthesizing method, Catintell-Syn, a deep learning model that only uses unpaired HQ and cataract images to generate realistic cataract images.
2. We develop a novel Transformer & CNN-based method, Catintell-Res, for cataract fundus image Restoration. Considering the significant performance on multiple datasets.
3. Comprehensive quantitative and qualitative experiments demonstrate that our Catintell models outperform other state-of-the-art cataract image restoration algorithms.

Related work

Fundus image restoration

Traditional fundus image restoration and enhancement methods[3–6] are mainly based on hand-crafted priors. For example, Setiawan et al. introduce CLAHE into fundus image enhancement[3]. Mitra et al.[4] combines CLAHE with Fourier transform to enhance cataract images. He et al.[5] filter images as an edge-preserving smoothing operator and remove haze degradation efficiently. Cheng et al.[6] propose a structure-preserving guided retinal image filtering (SGRIF) in fundus image restoration. However, these methods achieve poor performance in clinical applications due to their limited prior knowledge or poor generalization ability.

CNN[7–10] have been used in natural image restoration and achieved impressive results. CNNs have introduced into fundus image restoration due to the success in nature image restoration[11–15]. For instance, Zhao et al.[11] propose an end-to-end deep CNN to remove the lesions on the fundus images of cataract patients. Sourya et al.[12], Shen et al.[13], and Raj et al.[14] customize different synthetic degradation models to simulate the degradation types in actual clinical practice better. Luo et al. report a two-stage dehazing algorithm, which restores cataract fundus images under the supervision of segmentation[15]. Li et al.[9] propose a network to annotation-freely restore cataract fundus images (ArcNet).

Meanwhile, the Transformer[16] has been introduced into fundus image restoration to address the limitations in capturing long-range dependencies and achieve remarkable performance. Deng et al.[16] focus on real fundus image restoration and propose the first Transformer-based method (RFormer) in real fundus image restoration.

Generative adversarial network

Generative Adversarial Network (GAN) is firstly introduced in[17] and has been proven successful in image synthesis[18–20], and translation[19,20]. Subsequently, GAN is applied to image restoration and enhancement[8,11,12,21,22]. For instance, Wang et al.[8] propose the ESRGAN in single image super-resolution. Zhang et al.[21] propose a new method that combines two GAN models, a learning-to-Blur GAN and learning-to-DeBlur GAN. Jiang et al.[22] focuses on low-light image enhancement and develop an unsupervised generative adversarial network(EnlightenGAN). Meanwhile, some works[16,23] are dedicated to improving the underlying framework of GAN, such as replacing the traditional CNN framework with Transformer. Jiang et al.[23] propose the first Transformer-based GAN, TransGAN, for image generation. The introduction of Cycle-GAN further improved

the performance of fundus restoration models by generating its own LQ-HQ image pairs[24,25]. However, since these methods are not trained on cataract images, they can hardly be directly applied in cataract restoration.

Fundus image quality assessment

To evaluate the performance of cataract image restoration algorithms, it is essential to employ fundus image quality assessment (FIQA) methods. Although numerous natural image quality assessment (NIQA) techniques, such as BRISQUE, BPRI, and RichIQA[26–30], have been developed to assess image quality from various sources, fundus images, as a type of medical image, differ significantly from natural images and thus require specialized quality assessment methods. Traditional FIQA approaches[31–33] primarily rely on hand-crafted models, which have demonstrated inadequate performance for high-precision assessments. Recently, CNNs have been applied to FIQA[34,35], yielding superior results. Consequently, we utilize CNN-based methods to evaluate the images restored by our proposed algorithm and other methods.

Methodology

Overview

The Catintell model can be divided into two parts with similar structures: Catintell-Syn for image generation and Catintell-Res for cataract image restoration. Both of the Catintell models have the conditional GAN structure. The overall structure can be depicted in Fig. 1.

Catintell-Syn receives HQ fundus images and generates synthetic cataract-like images of the same size. Catintell-Syn is trained with unaligned data from the Catintell dataset. Because cataract fundus images from the Catintell Image dataset have different sizes and height-width ratios, the HQ images are cropped to the same size and ratio to accelerate the convergence of Catintell-Syn. Meanwhile, this model receives low-quality cataract fundus images as input and outputs corresponding restored images. It can accept inputs of various sizes and

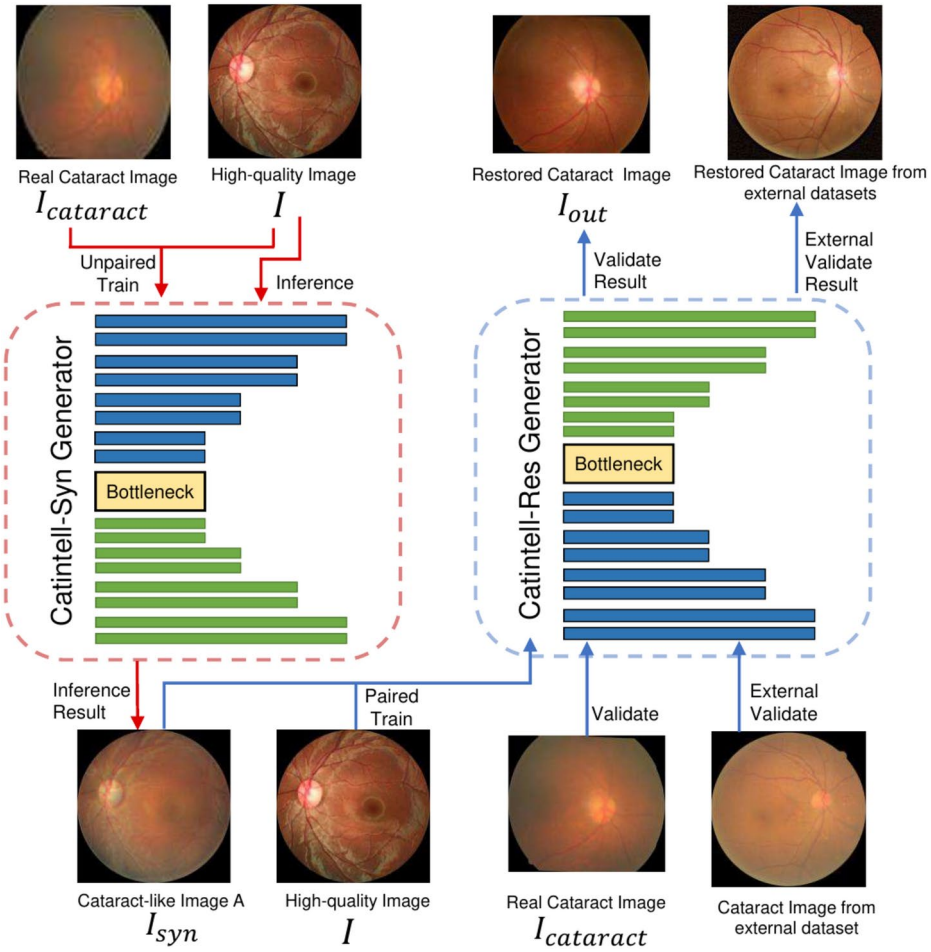

Fig. 1. Catintell Model Workflow. We use two GAN models to generate synthetic cataract images and restore cataract images separately. The idea is to collect the information contained in real cataract images and let Catintell-Syn learn from it. Then Catintell-Res learns from synthetic data generated by Catintell-Syn and works on real cataract images from various datasets. Existing methods focus on learning from synthetic data generated by an old method[2], which may not contain the features of real cataract images. But Catintell extracts features directly from real cataract images and applies them to real cataract image restoration.

height-width ratios and restore real cataract images. We use group convolution, internal small-range dense structures, and residual structures to improve performance.

After training with unpaired cataract data, we use Catintell-Syn to synthesize images highly similar to real cataract images. Then, these paired synthesized images are utilized to train Catintell-Res. This model follows a "Pixel to Pixel" principle to restore fundus images with the same spatial size. Finally, the trained Catintell-Res can restore real cataract images from various sources.

Ethical statement

This research utilized the images of human subjects (retinal fundus images), and identifying images are not included. All usage of data and experiments involving human subjects are approved by the ethical committee of the Beijing Tongren Hospital. The data utilized in this research are collected by the Beijing Tongren Hospital with the informed consent of patients. All research was performed in accordance with relevant guidelines and regulations.

Catintell model

The structures of Catintell models are similar GAN architectures, therefore, here, we take the model used in the cataract image restoration stage, the Catintell-Res as an example, which is shown in Fig. 2a. Catintell-Res takes a cataract image $\mathrm{I}_{in} \in \mathbb{R}^{H \times W \times 3}$ as input. First, the input is processed by an input projection layer (5×5 convolutional layer) to get the initial feature $\mathrm{I}_0 \in \mathbb{R}^{H \times W \times C}$, where C is the feature dimension, and set to 32 in Catintell-Res. Then, the feature is encoded by three Dense Conv Blocks with a skip connection and downsampled with a convolutional layer. In the encoding stage, this operation is performed four times, and the spatial size of the feature can be denoted as $\mathrm{X}_i \in \mathbb{R}^{\frac{H}{2^{i+1}} \times \frac{W}{2^{i+1}} \times 2^{i+1}C}$. Here, $i = 0, 1, 2, 3$ indicates the four stages. Afterward, the feature is processed by the bottleneck layers, another three Dense Conv Blocks, while its height, width, and channel are kept the same. Then, the feature is upsampled with four upsampling layers, each followed by one Dense Conv Block, and its spatial size is transferred to $\mathrm{X}_i \in \mathbb{R}^{\frac{H}{2^{8-i}} \times \frac{W}{2^{8-i}} \times 2^{8-i}C}$. Here, $i = 5, 6, 7, 8$ indicates the four upsampling stages. There are also skip connections between encoding and decoding stages of the same spatial size. Finally, the feature is processed by an output projection layer (5×5 convolutional layer) to provide the output image $\mathrm{I}_{out} \in \mathbb{R}^{H \times W \times 3}$.

The discriminator of Catintell-Res is a lightweight SWIN-Transformer[36]. The structure of the discriminator is shown in Fig. 2b. We use BCE loss as GAN loss in Catintell-Res.

The structure of Catintell-Syn follows the same workflow, but its depth and width are lower. We shrink its size to reduce its encoding level and reduce its generation ability because cataracts only affect the lenses of the eyes and seldom cause vessel lesions in fundus images. If the generation ability of Catintell-Syn is too strong, we can observe some artifact lesions on the generated images. Therefore, the depth and width are optimized to 3 stages and 16 feature dimensions to degrade fundus images but not generate lesions.

Conv encoder

In the encoding and decoding stages, the spatial size of feature maps does not change after processing by the Dense Conv Blocks or Conv Encoders. The structure of the Dense Conv Block is shown in Fig. 2c. It comprises two 5×5 convolutional layers and two 1×1 convolutional layers. There is layer normalization between 5×5 and 1×1 convolutional layers and GELU activation between 1×1 convolutional layers. The second 5×5

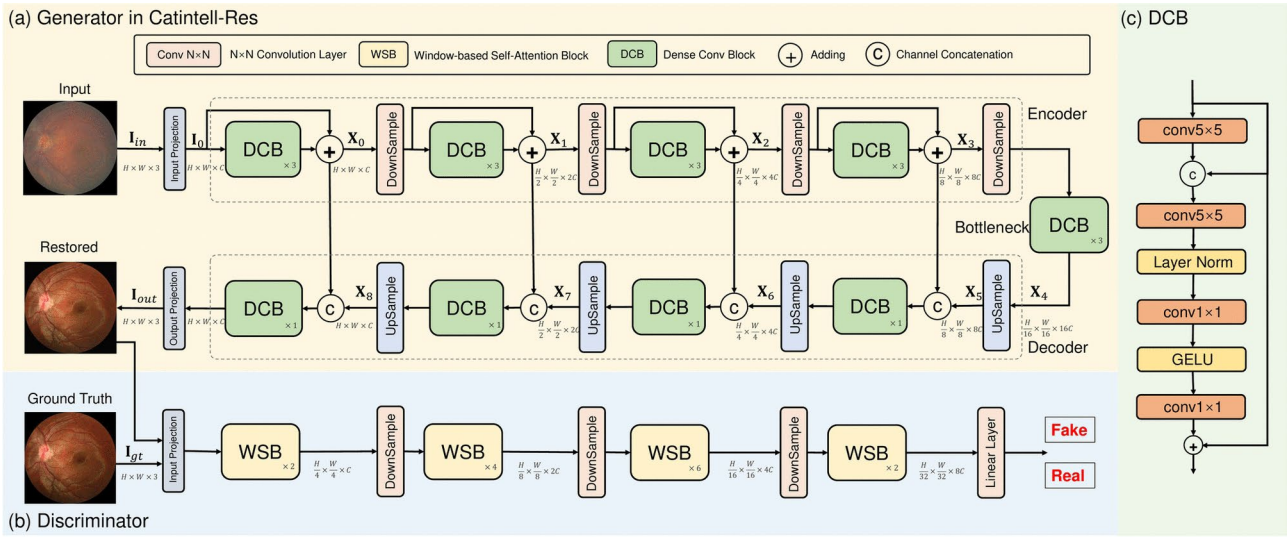

Fig. 2. The structure of the Catintell model. (**a**) The example model has a four-stage convolutional generator with downsampling and upsampling multiplier 2. (**b**) The discriminator of Catintell is a Transformer-based classifier and has four stages. (**c**) Detailed structure of the Dense Conv Block.

convolutional layer not only receives output from the layer ahead but also receives input with a skip connection to form a dense structure. A Conv Encoder contains three Dense Conv Blocks, whose structure is shown in Fig. 2c. There is a skip connection in its structure, which can accelerate its convergence and raise its performance.

Catintell loss functions

To formulate the loss functions, we denote the target HQ image A with I, the input cataract-like image A with I_{syn}, the real cataract image B with $\mathrm{I}_{cataract}$, the output restored image A with I_{out}, the process of degradation generator with $Gen(\cdot)$, and the process of degradation discriminator with $Dis(\cdot)$.

Pixel loss The pixel loss is a fundamental loss function in Catintell models, and we chose to apply it using the SmoothL1 loss function, $\mathscr{L}_{smoothL1}$, which is shown in the Eq. 1.

$$\mathscr{L}_{smoothL1} = \begin{cases} 0.5 \times (\mathrm{I} - \mathrm{I}_{out})^2, -1 < \mathrm{I} - \mathrm{I}_{out} < 1 \\ |\mathrm{I} - \mathrm{I}_{out}| - 0.5, otherwise \end{cases} \tag{1}$$

Fundus perceptual loss Due to the massive difference between the fundus and common images, the perceptual loss shall be modified to suit fundus images. We retrained a VGG-19[37] network to formulate a perceptual loss specifically for fundus images, which is named Fundus Perceptual Loss (FPLoss). The VGG-19 is trained with the EyeQ[35] dataset images with quality labels. The FPLoss works similarly to normal perceptual loss, and it can also give style loss.

Using $\phi(\cdot)$ to denote the feature extractor of VGG-19 and $Gram(\cdot)$ to denote the Gram matrix calculation, if we assume the height and width of extracted feature maps are *H* and *W*, the FPLoss, $\mathscr{L}_{fp}$, can be denoted as following Eq. 2.

$$\begin{aligned} \mathscr{L}_{fp} &= \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}(\phi(\mathrm{I}_{out})(i,j) - \phi(\mathrm{I}_{cataract})(i,j))^2; \\ \mathscr{L}_{fp-style} &= \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}(\phi(Gram(\mathrm{I}_{out}))(i,j) \\ &\quad - \phi(Gram(\mathrm{I}_{cataract}))(i,j))^2 \end{aligned} \tag{2}$$

Identity loss The identity loss $\mathscr{L}_{ide}$ can ensure that the restoration model can keep fundus images unchanged when the input images are HQ images. (Contrary to the cataract image synthesis model Catintell-Syn, which can keep the style of input cataract images) The style and details of a real HQ image shall be kept the same after the process of this restoration model. With input I, the processed image of the degradation branch is $Gen(\mathrm{I})$. The identity loss will calculate the pixel loss of I and $Gen(\mathrm{I})$. To be more specific, the pixel loss applied in the identity loss is SmoothL1 loss, therefore, the loss can be formulated as Eq. 3.

$$\mathscr{L}_{Identity} = \mathscr{L}_{smoothL1}(\mathrm{I}, Gen(\mathrm{I})) \tag{3}$$

GAN loss The discriminators in the Catintell models give predictions of possibility. Therefore, we use BCE loss as GAN loss of Catintell-Res. The calculating of $\mathscr{L}_{GAN}$ is shown in Eq. 4

$$\begin{aligned} \mathscr{L}_{GAN} &= -(\mathrm{P}_Y log(\mathrm{P}_{out}) + (1 - \mathrm{P}_Y) log(1 - \mathrm{P}_{out})), \\ where\, \mathrm{P}_Y &= Dis(\mathrm{I}_{cataract}), \mathrm{P}_{out} = Dis(\mathrm{I}_{out}) \end{aligned} \tag{4}$$

The overall losses of Catintell models can be formulated as follows Eqs. 5 and 6, and each loss is adjusted by loss weight before its loss symbol. The loss weight of each loss is adjusted according to experiments and for better performance. The pixel loss weight is low in the Catintell-Syn model, which has unpaired input images, but significantly higher in the Catintell-Res model. Meanwhile, the ratio of perceptual loss, GAN loss, and identity loss is kept the same for both models. However, as the loss weight of pixel loss in the Catintell-Syn model is too low for a fast convergence, we adjust the loss weight 10 times higher and raise the loss weight of perceptual loss in this model to 1.

$$\mathscr{L}_{Syn} = 0.01\mathscr{L}_{smoothL1} + \mathscr{L}_{fp} + 0.1\mathscr{L}_{ide} + 0.1\mathscr{L}_{GAN} \tag{5}$$

$$\mathscr{L}_{Res} = \mathscr{L}_{smoothL1} + 0.1\mathscr{L}_{fp} + 0.01\mathscr{L}_{ide} + 0.01\mathscr{L}_{GAN} \tag{6}$$

Dataset and experiments

Dataset

To train and test Catintell models, we collected a dataset, named Catintell Image, containing 1144 HQ fundus images and 2436 cataract images from Beijing Tongren Hospital. Meanwhile, the 10-fold validation is also applied. Before training, collected images are randomly sampled 10% as the validation set including 244 cataract images and 114 HQ images 10 times (we intend to create datasets without replication and absence, thus the last set contains 240 and 118 images), while the rest of these images are training set. Meanwhile, as mentioned above, the Catintell is a two-stage model, and the restoration of cataract images happens in the second stage which needs no clear or HQ images in the inference process. Therefore, we collected another 102 cataract images to

examine the performance of Catintell in real cataract image restoration. There are some image samples shown in Fig. 3.

The images and datasets used in this research are available for justified usage and research upon request. For further inquiries, please contact the corresponding authors.

Besides the Catintell dataset, we also use two external datasets to validate the generality of the model. The ODIR[38] and an open-source Kaggle cataract dataset[39] are experimented with to test the model's ability to enhance the quality of real cataract images.

Deployment details

During training, Catintell is applied with PyTorch version 1.10 and trained with CUDA version 11.7. We train each model for 80,000 iterations (equivalently 300 epochs) with the batch size 8 and learning rate 10^{-5} with cosine decay for all sub-models at first and apply a fine-tuning process with the same batch size and learning rate 10^{-6} with linear decay only for Catintell-Res models. The Adam[40] optimizer is applied with 1000 iterations warm up. All experiments are trained using a single NVIDIA Geforce RTX3090 GPU running for 10 hours to complete the training process.

The input fundus images are first resized to 768×768 and then randomly cropped to 256×256 patches. The spacial size of 768×768 can ensure details of original images are retained, and 256×256 is set for less GPU RAM usage and data augmentation. Since training GANs using images with varying black areas can complicate the learning process, the HQ-LQ image pairs shall have the same black frame size on the same location. Therefore, the random cropping process is paired in the same location on two images, which can ensure the stability of the training process of both GAN models. Meanwhile, all images are augmented using horizontal/

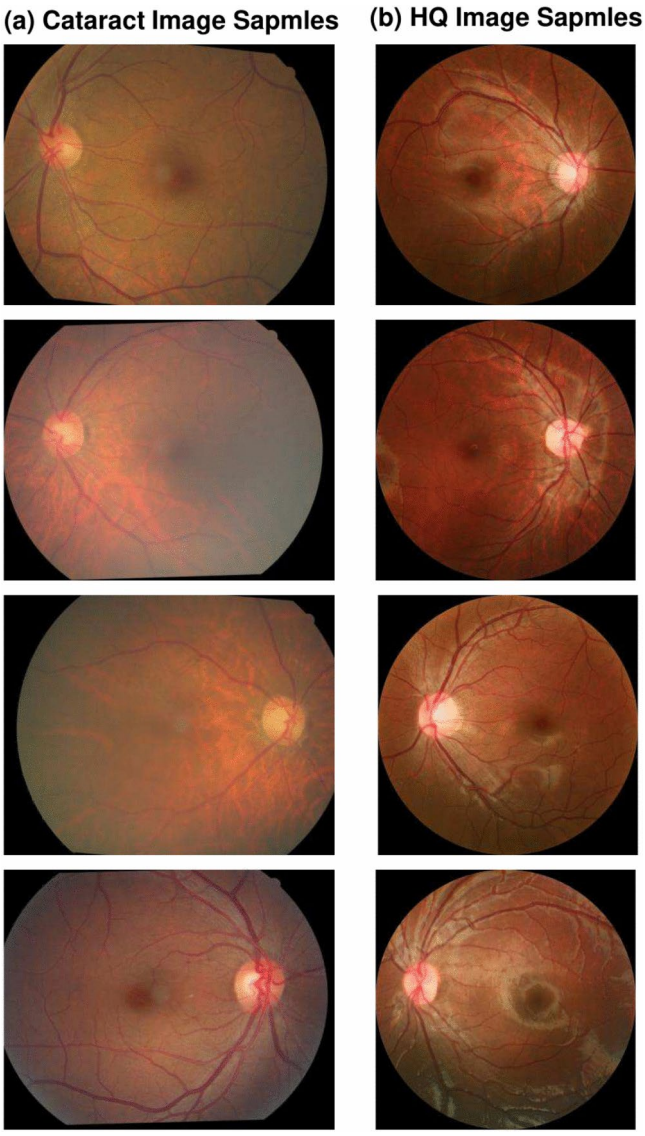

Fig. 3. Sample of our Catintell Image dataset. (**a**) 2436 cataract images were collected in this dataset. (**b**) 1144 high-quality images were collected.

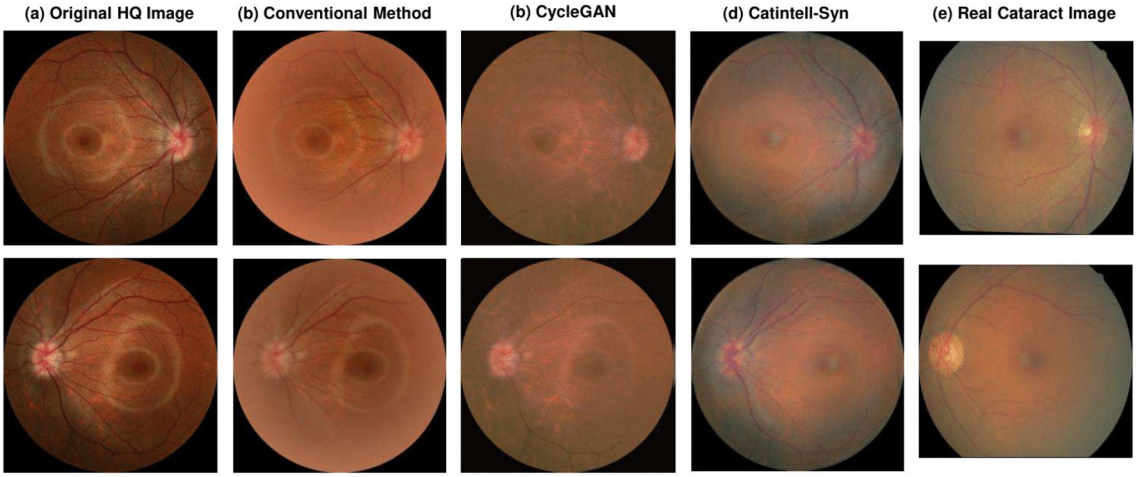

Fig. 4. Result of degraded images from Catintell-Syn and traditional modeling method. (**a**) Source HQ fundus images. (**b**) Synthetic cataract fundus images using traditional method. (**c**) using CycleGAN. (**d**) using Catintell-Syn. (**e**) Real cataract fundus image samples. The images generated by Catintell-Syn are more similar to real cataract fundus images.

Image	HQ	Conventional	CycleGAN	Ours	Real
Rank score	20	53	62	79	86

Table 1. User study of Catintell-Syn.

vertical flipping. These data augmentation methods are not applied to the validation and test stages to ensure consistent output and completeness of cataract images.

The Catintell-Res model can enhance fundus images with different height-width ratios. Therefore, input image shapes in the validating and testing stages are flexible.

The proposed Catintell model is an image restoration model, so we select the PSNR and SSIM as evaluation metrics. The optimization process of hyperparameters in Catintell models is demonstrated in the later part of the experiment section.

Catintell-Syn experiments

We provide qualitative comparisons between Catintell-Syn, CycleGAN[41], and the traditional degradation method[2]. The so-called 'traditional method' was first introduced in 1989[2], and is utilized in many cataract restoration works mentioned above. Though this method can give promising output for various fundus images, it has trouble dealing with images with a height-width ratio other than 1:1. Moreover, this method follows a fixed algorithm workflow regardless of the difference between input fundus images and has outputs almost the same style. The CycleGAN is widely utilized in style transfer research, we also carry out experiments on this method. However, it did not achieve fair results on cataract images.

The results are shown in Fig.4. It can be observed that the degradation style of Catintell-Syn is essentially consistent with real cataract images. Specifically, synthetic degradation closely matches real degradation in both location and severity. Severe degradation is observed in the blood vessels and macula area, while the optic disc region shows mild degradation.

Catintell-Syn user study

To get real feedback from ophthalmologists, we conducted a user study to collect their opinions and rank cataract images synthesized by our Catintell-Syn model. In the study, we provide them with five images: real cataract images, HQ images, images from the conventional method, CycleGAN, and images from our Catintell-Syn model. There are ten sets of these image groups, and the images are given 10,8,6,4,2 scores corresponding to their ranks respectively. (This score setting means to get scores with a maximum of 100. Higher similarity to real cataract images results in higher scores.) The average results of three experienced ophthalmologists and three young ophthalmologists are summarized in Table 1.

The score of images synthesized by Catintell-Syn is slightly lower than the real cataract images and obviously higher than images generated by the conventional method or CycleGAN. Therefore, we conclude that Catintell-Syn succeeds in synthesizing cataract images highly similar to real ones.

Catintell-Res experiments

The calculation of quantitive metrics requires paired images. However, as addressed in the introduction section, the difficulty of acquiring cataract-clear image pairs within a short interval hinders data collection. Therefore,

Method	PSNR	SSIM	Parameters(M)
GLCAE[42]	16.22	0.5627	–
Dark channel prior[43]	15.90	0.7482	–
GCANET[7]	23.61	0.8145	–
ESRGAN[8]	29.47	0.7907	16.72
ARCNet[9]	19.81	0.8709	54.42
GFENET[10]	16.77	0.8521	89.30
pixDA Sobel[44]	18.52	0.8399	54.42
SCRNET[9]	15.76	0.8568	89.28
RFormer[16]	22.96	0.6808	21.66
I-SECRET[45]	19.19	0.7844	10.85
Catintell-Res (ours)	**39.03**	**0.9476**	12.72

Table 2. Compared with SOTA methods. Significant values are in bold.

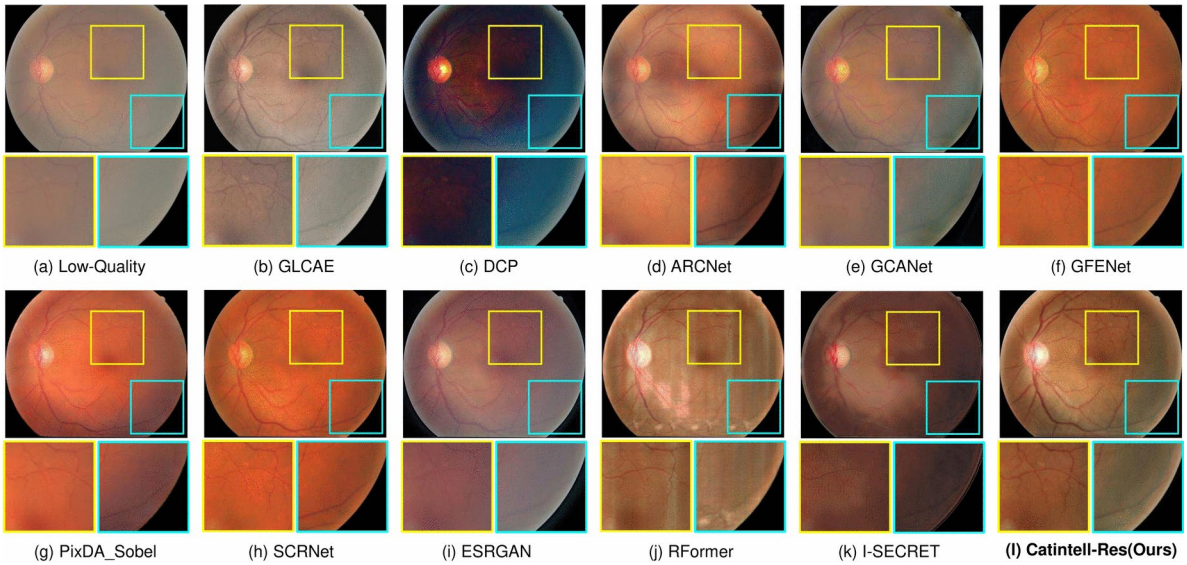

Fig. 5. Restored real cataract image comparisons of Scene 1 on a test image of the Catintell Image dataset. Compared to other methods, the vessels around the macula in the restored image of Catintell-Res are finely enhanced. The overall style of this image is also maintained rather than changed to a dark/orange color.

to meter the performance of Catintell-Res models, we use the simulated cataract-HQ image pairs from the Catintell-Syn model. Moreover, the following models for comparison are also applied with the simulated cataract-HQ images to ensure fair comparisons.

The GLCAE[42] and Dark channel prior(DCP)[43] are modeling methods and need no parameters. They usually follow the same work mode and apply the same modifications to different images. The ESRGAN[8] and GCANET[7] are general image enhancement methods that are yet to be adapted to cataract image restoration. We retrained these models with cataract image pairs to get better results. The ARCNet[9], pixDA Sobel[44], SCRNET[9], RFormer[16], I-SECRET[45], and GFENET[10] are reported fundus image enhancement methods. The ARCNet, pixDA Sobel, and SCRNET use high-frequent information to enhance the restoration process, and RFormer uses Transformers to elevate its performance. These methods need algorithms to degrade the HQ images to get cataract-like images first and then restore the image. Therefore, they actually target a fixed fake cataract image-generating method but not the real style of cataract images. However, the Catintell can learn from the realistic cataract-like images which are proven better in the prior section. Meanwhile, the comparisons with general image restoration can also prove that Catintell is more suitable for the cataract image restoration task.

The results of quantitive metrics are shown in Table 2. We can observe from the results that Catintell-Res has a great ability for image restoration. It extravagantly outperforms other methods both in PSNR and SSIM through learning from the synthesized data. The restoration results of the synthesized images are shown in Fig. 7. The image restored by Catintell exhibits a realistic style and balanced contrast compared to the others.

We also use the test set of the Catintell dataset to examine the restoration ability towards the real cataract image. Two samples of test results are shown in Figs. 5 and 6. As mentioned above, the restoration branch of Catintell can work independently, this test was carried out on the real cataract images which have no corresponding clear

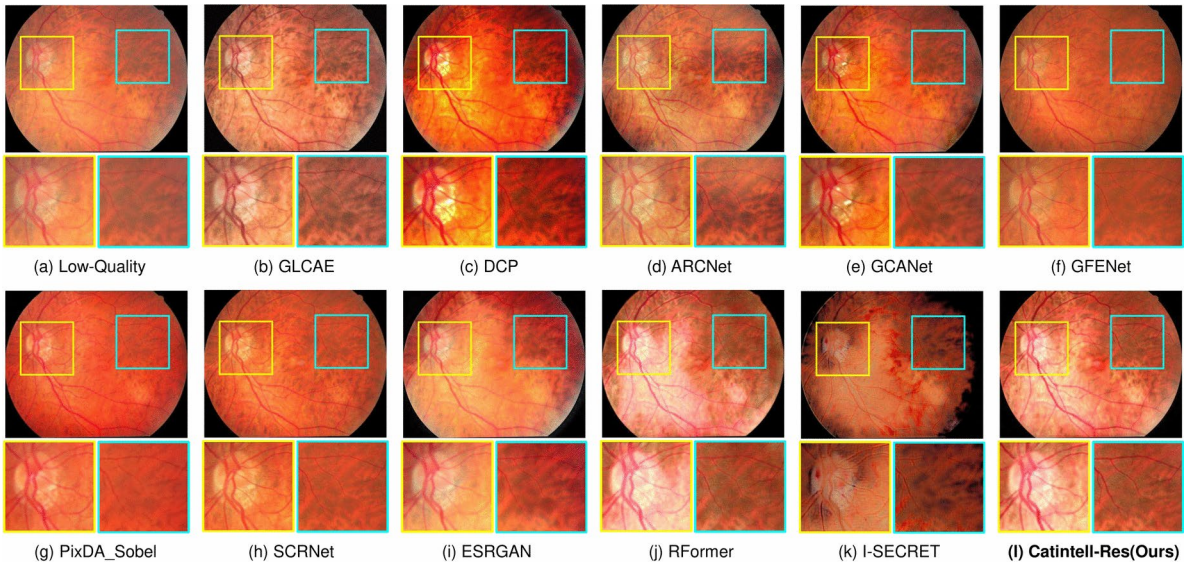

Fig. 6. Restored real cataract image comparisons of Scene 2 on a test image of the Catintell Image dataset. The optic cup/disk area of the fundus image restored by Catintell-Res has clear edges of vessels. In the surrounding area, the vessels are easy to distinguish.

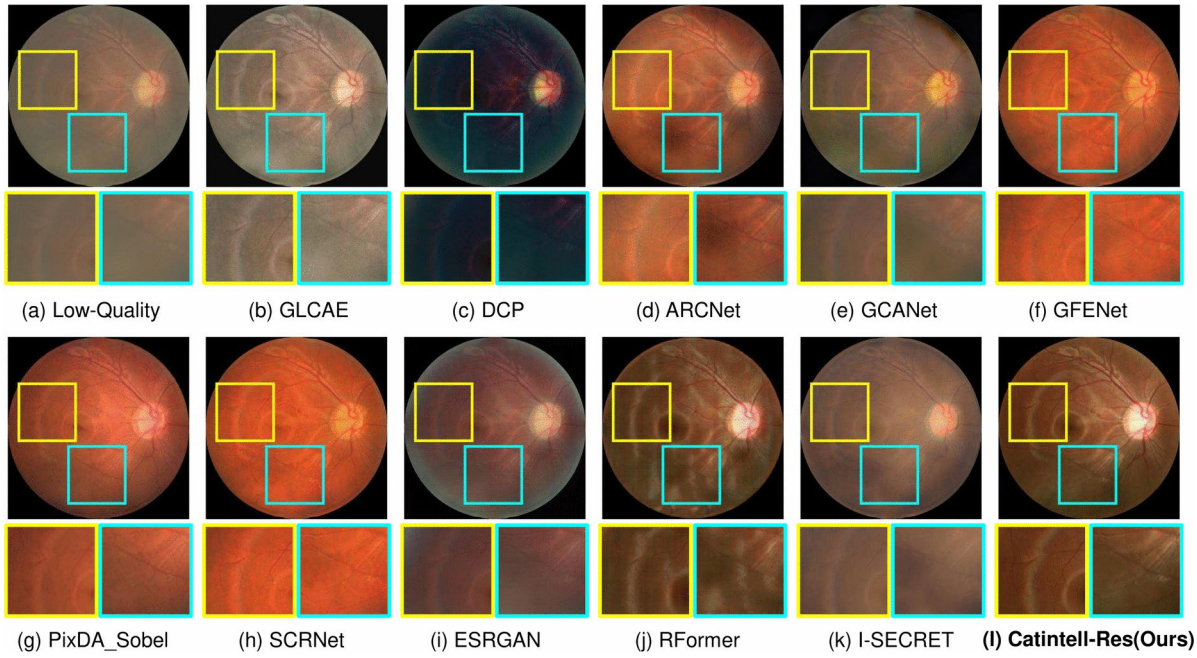

Fig. 7. Restored synthesized cataract image comparisons. Catintell-Res can retain the style of the image and escalate the contrast of the whole image.

images to compare. However, besides this visual exhibition, we also did a user study in the later section to show the results from Catintell getting the highest rating from ophthalmologists from clinical usage.

In the first real cataract image test scene, the style of the restored cataract image is retained by Catintell-Res, and the vessel details around the macular are restored and become more obvious compared to other methods and the original image. In the second scene, the optic cup/disk and surrounding area of the fundus image restored by Catintell-Res become much clearer, and the overall contrast of this image is raised.

Catintell-Res user study

After validating the restoration ability of Catintell-Res, we carried out another user study to figure out what opinions ophthalmologists hold. In the study, we provide them with eight images, which are original cataract

Method	Catintell-Res (ours)	GLCAE	DCP	ARCNet
Rank score	**99.17**	64.67	44.17	78.17
Method	GCANET	GFENET	I-SECRET	Original Image
Rank score	74.33	71.33	34.83	53.33

Table 3. User study of Catintell-Res. Significant values are in bold.

Method	Label "Good"	"Usable"	"Reject"
GLCAE[42]	48.61	45.14	6.25
Dark channel prior[43]	8.33	24.31	46.53
GCANET[7]	32.64	36.11	9.72
ESRGAN[8]	84.03	11.81	4.17
ARCNet[9]	81.25	11.81	2.78
GFENET[10]	90.28	6.94	2.78
pixDA Sobel[44]	84.72	11.81	3.47
SCRNET[9]	91.67	5.56	2.78
RFormer[16]	84.72	10.42	4.86
I-SECRET[45]	59.03	16.67	8.33
Catintell-Res (ours)	**93.06**	4.86	**2.08**

Table 4. FIQA test results comparison with SOTA methods. Significant values are in bold.

images and images restored by GLCAE[42], Dark channel prior[43], ARCNet[9], GCANET[7], GFENET[10], I-SECRET[45], and our Catintell-Res model. We did not label these images with methods or indicate their source. There are ten sets of these image groups, and the images are given 10,9,8,7...4,3 scores corresponding to their ranks, respectively. (This score setting means to get scores with a maximum of 100.) The average results of three experienced ophthalmologists and three young ophthalmologists are summarized in Table 3.

The images restored by Catintell-Res are the best according to the score among these methods. Therefore, the restoration ability of Catintell-Res has proven effective and powerful, whether in quantitive experiments or user studies.

Catintell-Res FIQA test

To assess fundus image quality, AI-based FIQA methods that provide subjective metrics are also available as mentioned in "Fundus image quality assessment" section. Consequently, we conducted a FIQA test to compare our method with others using an AI-based FIQA approach on real cataract image results. In this study, we selected the widely-used MCF-Net[35], which was proposed with the EyeQ dataset[35]. MCF-Net receives fundus images and assigns quality labels, including "Good," "Usable," and "Reject," indicating high, mediocre, and low image quality, respectively. The restored images from all methods were processed by this FIQA network, and the classes corresponding to the highest output logits were selected as the output labels. The results are presented in Table 4.

From this chart, we observe that Catintell has the highest image ratio in the "Good" category and the lowest image ratio in the "Reject" category. Therefore, the FIQA test demonstrates that Catintell exhibits superior performance in cataract image restoration.

Discussion

Generalized cataract restoration ability

Besides using the synthesized cataract and real cataract images in the Catintell Image dataset, we also test our models on the other open-source cataract dataset. The ODIR dataset is from the ODIR2019 competition[38], which contains several kinds of fundus images of retinal diseases. We use cataract images in the training set of this dataset to validate Catintell-Res. We also collected a dataset from Kaggle named Cataract-Dataset[39]. We use the cataract division of this dataset in this experiment.

Catintell-Res is not further retrained or modified, and the data is directly processed by the trained Catintell-Res model. The results are shown in Fig. 8. We can observe from the figures that Catintell-Res has obtained universal restoration ability through the synthesized data from the Catintell Image dataset, and its ability still functions even on cataract images from other sources. In the Kaggle dataset, the macula of these fundus images is restored to be clear, and the vessels become obvious. This also suits the ODIR-5K dataset, and we can see that Catintell-Res is able to remove most of the blurry area in the real cataract images.

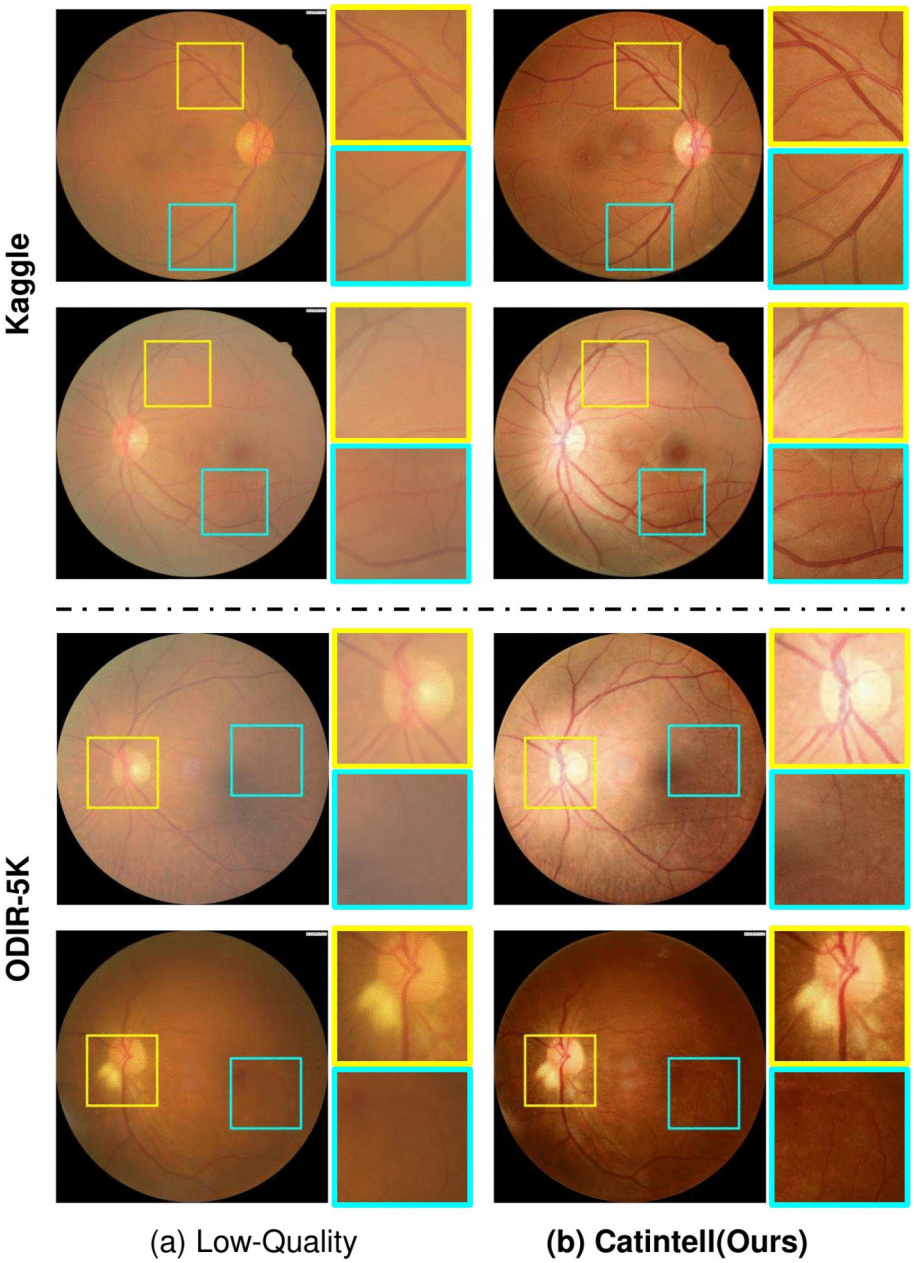

Fig. 8. Restored real cataract image from external datasets. Catintell-Res has the universal ability to restore cataract images collected from other fundus cameras and sources.

Ablation studies

Encoder/decoder

Though we designed a new encoder/decoder structure in Catintell-Res, we also tried other encoder structures to optimize the performance of Catintell-Res. ConvNeXt[46], RRDB(residue in residue dense block) of ESRGAN[8], and W-MSA of SWIN Transformer[36] are applied in the model of same structures to compare their performance.

The results are shown in Table 5. The ConvNeXt encoder/decoder has the best performance except for Catintell, which is why we optimize the encoder/decoder of Catintell with inspiration from ConvNeXt. The encoder/decoder of Catintell is optimized for image restoration and achieves the best performance among those methods.

Patch size

In the training process of Catintell, we use patches of size 256×256 pixels to avoid heavy computational burden. Since the patch size significantly impacts the model performance, we test different patch sizes in this section.

When the patch size is smaller, the model can use a larger batch size during training to avoid sampling error. However, smaller patches make it difficult for the model to learn the spatial context information of the entire image and prone to overfitting, which in turn leads to a decrease in performance in the validation stage. On the

Method	PSNR	SSIM	Parameters (M)
RRDB[8]	29.61	0.7649	38.56
Swin transformer[36]	24.83	0.4150	15.56
ConvNeXt[46]	34.97	0.8696	21.81
Catintell-Res (ours)	**39.03**	**0.9476**	12.72

Table 5. Ablation study of different encoder/decoder. Significant values are in bold.

Size	PSNR	SSIM	Parameters (M)
128	35.71	0.9357	12.72
192	38.07	0.9279	12.72
256 (Catintell-Res)	**39.03**	**0.9476**	12.72
384	37.51	0.9369	12.72

Table 6. Ablation study of different patch size. Significant values are in bold.

Width	PSNR	SSIM	Parameters (M)
16	36.07	0.9265	3.78
32 (Catintell-Res)	**39.03**	**0.9476**	12.72
48	37.11	0.9365	26.82

Table 7. Ablation study of different width. Significant values are in bold.

Depth	PSNR	SSIM	Parameters(M)
3	36.78	0.9346	3.711
4 (Catintell-Res)	**39.03**	**0.9476**	12.72
5	34.07	0.9005	46.27

Table 8. Ablation study of different depth. Significant values are in bold.

other hand, when the patch size is larger, it consumes more space and forces the batch size to be reduced, and the sampling error increases, making the model hard to converge.

As shown in Table 6, the training results of the model with a patch size of 256 are the best.

Depth and width

Since Catintell-Res uses a U-shaped structure, each encoding/decoding stage is aligned with a downsampling/upsampling, so the number of encoding/decoding stages significantly affects the network depth.

The width of the network is determined by the projection channels of the input projection layer. With the linear increase of the projection channels, the parameters of the model increase quadratically.

To obtain the optimal number of encoding/decoding stages and network width, we conduct the following experiments on the Catintell-Res model, keeping the rest of the structure unchanged and only changing the number of encoding/decoding stages or width to verify its impact on performance. The results are summarized in Table 8 and 7, and the model with four stages and a width of 32 has the best performance. Therefore, this width and depth combination is used in the Catintell-Res model to obtain the best model performance.

Loss weights

Catintell models incorporate four distinct loss functions, which are crucial for their convergence and performance, as detailed in "Catintell loss functions" section, we conducted experiments to evaluate the impact of different loss weights. We maintained the pixel loss weights constant for both models and varied the weights of other losses by a factor of ten, either higher or lower, to observe their effect on performance. The models were trained and fine-tuned using the same strategies and hyperparameters as the original Catintell model, but with different loss weights.

The results are presented in Table 9. From the chart, we observe that the selected ratio for the Catintell models is optimal. Any increase or decrease in the loss weights results in a decline in performance.

Avoid mode collapse

Although many researchers have noted that mode collapse frequently occurs during GAN training[47,48], it is largely avoided in Catintell models. We have implemented several measures to prevent mode collapse.

$\mathscr{L}_{fp}$	$\mathscr{L}_{ide}$	$\mathscr{L}_{GAN}$	PSNR	SSIM	Parameters(M)
10x higher	Same	Same	28.70	0.8925	12.72
10x lower	Same	Same	37.27	0.9453	12.72
Same	10x higher	Same	32.47	0.9392	12.72
Same	10x lower	Same	30.90	0.9256	12.72
Same	Same	10x higher	32.10	0.9283	12.72
Same	Same	10x lower	33.67	0.9376	12.72
	Catintell-Res		**39.03**	**0.9476**	12.72

Table 9. Ablation study of different loss weights. Significant values are in bold.

First, data augmentation has been employed to mitigate mode collapse. Mode collapse often arises when data is insufficient, causing the model to represent a single pattern. To enhance the overall diversity of images, we applied paired random cropping, flipping, and rotating to the training data. This significantly reduced the incidence of mode collapse.

Additionally, the use of identity loss further minimized the likelihood of mode collapse. This loss function ensures that the processed target images retain their style, thereby increasing the consistency of the output style. Identity loss is particularly important for Catintell-Syn due to its relatively low pixel loss weight.

Moreover, the learning rate is another factor influencing mode collapse. Initial experiments with various learning rates revealed that high learning rates could lead to potential mode collapse. Consequently, we adopted "safe" learning rates, which are relatively lower and less prone to causing mode collapse. These learning rates are detailed in "Deployment details" section.

After implementing these measures, we rarely observed mode collapse during the subsequent training and fine-tuning of the Catintell models.

Limitation

Though Catintell-Res has obtained universal restoration ability, it can not process images with severe blur. When fundus images are collected, there is some reason that their quality is not guaranteed. To be more specific, some images suffer from wrong illumination, whether too high or too low. And some may be blocked by eyelids or iris. All of those abnormal images can be named degradation images. For those images with severe degradation, there is no sign of vessels to assist Catintell-Res in escalating image quality. Therefore, Catintell-Res can not handle these images or generate whole images through a little undegraded area. We plan to include more cataract images with severe degradation to improve the synthetic models.

Meanwhile, some of the HQ images we utilized in the Catintell Image Dataset have texture features that usually appear in young healthy eyes (like "sparkling reflections"). This could potentially cause synthetic images to display similar features, leading to ambiguity. However, our evaluations indicate that these features do not appear in synthetic images generated from HQ images that lack these characteristics. Furthermore, tests on real clinical data and external datasets also show no similar features. Therefore, we conclude that the models treat these features as innate rather than generalized, which does not affect the learning and generalizability of the Catintell-Res models. Including more HQ images can further address this issue.

Recently, diffusion models have attracted some interest from researchers in the image-to-image translation field. We also regard diffusion models like works from Rombach et al.[49] and Su et al.[50] as good solutions for both cataract image synthesis and restoration. However, the diffusion models could occupy a massive amount of GPU RAM while training, which is sometimes over 40GB in practice and much higher than the inference process. This RAM burden is too heavy for our GPU to train a diffusion model.

Moreover, suppose the diffusion steps are high or the latent feature size is small. In that case, the generation ability of the diffusion model is too strong to retain enough fidelity for medical usage, and fake focus may be generated due to this. Therefore, we choose not to use diffusion models in our work for now, but, still, diffusion models are of great potential in medical image processing which we plan to exploit in the future.

We plan to:

1 Enlarge the range of images collected in the Catintell Image dataset to elevate the generating ability of Catintell-Syn and Catintell-Res for a more extensive range of LQ and HQ images.
2 Modify the structure of Catintell-Syn to make it able to generate more kinds of degraded images.
3 Transfer the Catintell models to other medical image tasks to extend their application.
4 Apply lightweight diffusion models on fundus image restoration and optimize Catintell models.

Conclusion

In this paper, we address the problems in cataract image restoration through a new synthesizing and restoration method, Catintell. Before our method, there was much difference between conventional simulated and real cataract images; the quality of restored cataract images was not high enough. Our method, Catintell-Syn, uses fully unsupervised data to generate paired cataract-like images with realistic style and texture and successfully alleviates the lack of paired images. Based on the synthetic images, we developed Catintell-Res to restore real cataract images. The structure of these models is optimized for fundus images, and we also added the loss function expertized for ophthalmology in the training stage. Then, we carried out user studies and quantitive

experiments for Catintell models. The results show that Catintell achieves remarkable performance in both synthesizing cataract-like data and restoring real cataract data. The generalization performance of Catintell-Res is verified by real cataract images from various external datasets. We plan to open Catintell models for research and clinic utilization and hope this model can help ophthalmologists with their work in the future.

Data availability

he data and code are available at https://github.com/HudenJear/Catintell for justified usage and research upon request. Please contact the corresponding author if there are any questions for data and code.

Received: 31 October 2024; Accepted: 28 January 2025
Published online: 01 April 2025

References

1. Wang, W. et al. Cataract surgical rate and socioeconomics: A global study. *Investig. Ophthalmol. Vis. Sci.* **57**, 5872–5881 (2016).
2. Peli, E. & Peli, T. Restoration of retinal images obtained through cataracts. *IEEE Trans. Med. Imaging* **8**, 401–406. https://doi.org/10.1109/42.41493 (1989).
3. Setiawan, A. W., Mengko, T. R., Santoso, O. S. & Suksmono, A. B. Color retinal image enhancement using CLAHE. In *International Conference on ICT for Smart Society* 1–3 (IEEE, 2013).
4. Mitra, A., Roy, S., Roy, S. & Setua, S. K. Enhancement and restoration of non-uniform illuminated fundus image of retina obtained through thin layer of cataract. *Comput. Methods Programs Biomed.* **156**, 169–178 (2018).
5. He, K., Sun, J. & Tang, X. Guided image filtering. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 1397–1409 (2012).
6. Cheng, J. et al. Structure-preserving guided retinal image filtering and its application for optic disk analysis. *IEEE Trans. Med. Imaging* **37**, 2536–2546 (2018).
7. Chen, D. et al. Gated context aggregation network for image dehazing and deraining. *WACV 2019* (2018).
8. Wang, X. et al. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops* (2018).
9. Li, H. et al. An annotation-free restoration network for cataractous fundus images. *IEEE Trans. Med. Imaging* **41**, 1699–1710 (2022).
10. Li, H. et al. A generic fundus image enhancement network boosted by frequency self-supervised representation learning. Preprint at arXiv:2309.00885 (2023).
11. Zhao, H., Yang, B., Cao, L. & Li, H. Data-driven enhancement of blurry retinal images via generative adversarial networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* 75–83 (Springer, 2019).
12. Sengupta, S., Wong, A., Singh, A., Zelek, J. & Lakshminarayanan, V. DeSupGAN: Multi-scale feature averaging generative adversarial network for simultaneous de-blurring and super-resolution of retinal fundus images. In *International Workshop on Ophthalmic Medical Image Analysis* 32–41 (Springer, 2020).
13. Shen, Z., Fu, H., Shen, J. & Shao, L. Modeling and enhancing low-quality retinal fundus images. *IEEE Trans. Med. Imaging* **40**, 996–1006 (2020).
14. Raj, A., Shah, N. A. & Tiwari, A. K. A novel approach for fundus image enhancement. *Biomed. Signal Process. Control* **71**, 103208 (2022).
15. Luo, Y. et al. Dehaze of cataractous retinal images using an unpaired generative adversarial network. *IEEE J. Biomed. Health Inform.* **24**, 3374–3383 (2020).
16. Deng, Z. et al. Rformer: Transformer-based generative adversarial network for real fundus image restoration on a new clinical benchmark. *IEEE J. Biomed. Health Inform.*[SPACE]https://doi.org/10.1109/JBHI.2022.3187103 (2022).
17. Goodfellow, I. et al. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **27** (2014).
18. Gong, X., Chang, S., Jiang, Y. & Wang, Z. Autogan: Neural architecture search for generative adversarial networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* 3224–3234 (2019).
19. Isola, P., Zhu, J.-Y., Zhou, T. & Efros, A. A. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 1125–1134 (2017).
20. 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* 2223–2232 (2017).
21. Zhang, K. et al. Deblurring by realistic blurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 2737–2746 (2020).
22. Jiang, Y. et al. Enlightengan: Deep light enhancement without paired supervision. *IEEE Trans. Image Process.* **30**, 2340–2349 (2021).
23. Jiang, Y., Chang, S. & Wang, Z. TransGAN: Two pure transformers can make one strong GAN, and that can scale up. *Adv. Neural Inf. Process. Syst.* **34** (2021).
24. Wu, H.-T. et al. Fundus image enhancement via semi-supervised GAN and anatomical structure preservation. *IEEE Trans. Emerg. Top. Comput. Intell.* **8**, 313–326. https://doi.org/10.1109/TETCI.2023.3301337 (2024).
25. Yoo, T. K., Choi, J. Y. & Kim, H. K. CycleGAN-based deep learning technique for artifact reduction in fundus photography. *Graefe's Arch. Clin. Exp. Ophthalmol.* **258**, 1631–1637. https://doi.org/10.1007/s00417-020-04709-5 (2020).
26. Mittal, A., Moorthy, A. K. & Bovik, A. C. No-reference image quality assessment in the spatial domain. *IEEE Trans. Image Process.* **21**, 4695–4708. https://doi.org/10.1109/TIP.2012.2214050 (2012).
27. Min, X., Zhai, G., Gu, K., Liu, Y. & Yang, X. Blind image quality estimation via distortion aggravation. *IEEE Trans. Broadcast.* **64**, 508–517. https://doi.org/10.1109/TBC.2018.2816783 (2018).
28. Min, X. et al. Exploring rich subjective quality information for image quality assessment in the wild. Preprint at arXiv:2409.05540 (2024).
29. Min, X. et al. Blind quality assessment based on pseudo-reference image. *IEEE Trans. Multimed.* **20**, 2049–2062. https://doi.org/10.1109/TMM.2017.2788206 (2018).
30. Zhu, H., Li, L., Wu, J., Dong, W. & Shi, G. MetaIQA: Deep meta-learning for no-reference image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020).
31. Lee, S. C. & Wang, Y. Automatic retinal image quality assessment and enhancement. In *Medical Imaging 1999: Image Processing*, vol. 3661 1581–1590 (SPIE, 1999).
32. Lalonde, M., Gagnon, L., Boucher, M.-C. *et al.* Automatic visual quality assessment in optical fundus images. In *Proceedings of Vision Interface*, vol. 32 259–264 (Ottawa, 2001).
33. Köhler, T. et al. Automatic no-reference quality assessment for retinal fundus images using vessel segmentation. In *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems* 95–100 (IEEE, 2013).
34. Shen, Y. et al. Multi-task fundus image quality assessment via transfer learning and landmarks detection. In *International Workshop on Machine Learning in Medical Imaging* 28–36 (Springer, 2018).

35. Fu, H. et al. Evaluation of retinal image quality assessment networks in different color-spaces. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* 48–56 (Springer, 2019).
36. Liu, Z. et al. Swin transformer: Hierarchical vision transformer using shifted windows. Preprint at arXiv:2103.14030 (2021).
37. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. Preprint at arXiv:1409.1556 (2014).
38. ODIR. Peking University International Competition on Ocular Disease Intelligent Recognition (ODIR-2019) (2019). https://odir2019.grandchallenge.org/.
39. yiweichen. Cataract dataset. https://github.com/yiweichen04/retina_dataset (2019).
40. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. Preprint at arXiv:1412.6980 (2014).
41. 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* 2223–2232 (2017).
42. Tian, Q.-C. & Cohen, L. D. Global and local contrast adaptive enhancement for non-uniform illumination color images. In *Proceedings of the IEEE International Conference on Computer Vision Workshops* 3023–3030 (2017).
43. He, K., Sun, J. & Tang, X. Single image haze removal using dark channel prior. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* 1956–1963. https://doi.org/10.1109/CVPR.2009.5206515 (2009).
44. Li, H. et al. Restoration of cataract fundus images via unsupervised domain adaptation. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)* 516–520 (IEEE, 2021).
45. Cheng, P., Lin, L., Huang, Y., Lyu, J. & Tang, X. I-secret: Importance-guided fundus image enhancement via semi-supervised contrastive constraining. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* 87–96 (Springer, 2021).
46. Liu, Z. et al. A convnet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2022).
47. You, A., Kim, J. K., Ryu, I. H. & Yoo, T. K. Application of generative adversarial networks (GAN) for ophthalmology image domains: A survey. *Eye Vis.* **9**, 6. https://doi.org/10.1186/s40662-022-00277-3 (2022).
48. Srivastava, A., Valkov, L., Russell, C., Gutmann, M. U. & Sutton, C. VEEGAN: Reducing mode collapse in GANs using implicit variational learning. In *Advances in Neural Information Processing Systems* vol. 30 (eds. Guyon, I. et al.) (Curran Associates, Inc., 2017).
49. Rombach, R., Blattmann, A., Lorenz, D., Esser, P. & Ommer, B. High-resolution image synthesis with latent diffusion models. Preprint at arXiv:2112.10752 (2021).
50. Su, X., Song, J., Meng, C. & Ermon, S. Dual diffusion implicit bridges for image-to-image translation. Preprint at arXiv:2203.08382 (2022).

Acknowledgements

This work is supported by the Science and Technology Innovation Committee of Shenzhen-Platform and Carrier (International Science and Technology Information Center) & Shenzhen Bay Lab. This work is funded by Shenzhen Science and Technology Innovation Committee under KCXFZ20211020163813019 and by the National Natural Science Foundation of China under 82000916.

Author contributions

Zheng Gong: Conceptualization, Methodology, Validation, Data Curation, Writing—Original Draft, Visualization. Zhuo Deng: Methodology, Validation, Data Curation, Writing—Original Draft, Visualization. Weihao Gao: Conceptualization, Data Curation, Writing—Review and Editing. Wenda Zhou: Data Curation, Validation, Writing—Review and Editing. Yuhang Yang:} Validation, Writing—Review and Editing. Hanqing Zhao: Data Curation. Lei Shao: Data Curation, Project administration. Wenbin Wei: Funding acquisition. Lan Ma: Funding acquisition.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to L.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.