

Randomness-restricted Diffusion Model for Ocular Surface Structure Segmentation

Xinyu Guo, Han Wen, Huaying Hao, Yifan Zhao, Yanda Meng, Jiang Liu, Yalin Zheng, Wei Chen, Yitian Zhao

Abstract—Ocular surface diseases affect a significant portion of the population worldwide. Accurate segmentation and quantification of different ocular surface structures are crucial for the understanding of these diseases and clinical decision-making. However, the automated segmentation of the ocular surface structure is relatively unexplored and faces several challenges. Ocular surface structure boundaries are often inconspicuous and obscured by glare from reflections. In addition, the segmentation of different ocular structures always requires training of multiple individual models. Thus, developing a one-model-fits-all segmentation approach is desirable. In this paper, we introduce a randomness-restricted diffusion model for multiple ocular surface structure segmentation. First, a time-controlled fusion-attention module (TFM) is proposed to dynamically adjust the information flow within the diffusion model, based on the temporal relationships between the network's input and time. TFM enables the network to effectively utilize image features to constrain the randomness of the generation process. We further propose a low-frequency consistency filter and a new loss to alleviate model uncertainty and error accumulation caused by the multi-step denoising process. Extensive experiments have shown that our approach can segment seven different ocular surface structures. Our method performs better than both dedicated ocular surface segmentation methods and general medical image segmentation methods. We further validated the proposed method over two clinical datasets, and the results demonstrated that it is beneficial to clinical applications, such as the meibomian gland dysfunction grading and aqueous deficient dry eye diagnosis.

Index Terms—Ocular surface; medical image segmentation; diffusion model.

This work was supported by the National Science Foundation Program of China (62422122, 62272444), Zhejiang Provincial Natural Science Foundation of China (LR22F020008), the Key Research and Development Program of Zhejiang Province (2024C03206, 2024C03101), the Youth Innovation Promotion Association CAS (2021298), and the Key Project of Ningbo Public Welfare Science and Technology (2023S012).

X Guo, H Hao and Y Zhao are with the Laboratory of Advanced Theranostic Materials and Technology, Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences, Ningbo, China; H Wen and W Chen are with Eye Hospital, Wenzhou Medical University, Wenzhou, China; Y Zhao is with the School of Aerospace, Transport and Manufacturing, Cranfield University, Bedfordshire, UK; Y Meng and Y Zheng are with the Department of Eye and Vision Sciences, University of Liverpool, Liverpool, UK; J Liu is with the Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, China.

X Guo and H Wen contributed equally to this work. Corresponding authors: Yitian Zhao (e-mail: yitian.zhao@nimte.ac.cn) and Wei Chen (e-mail: chenweimd@hotmail.com).

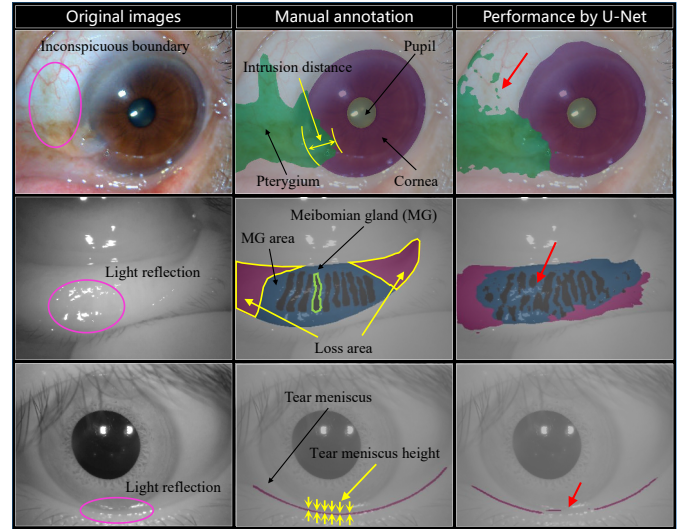


Fig. 1: Illustrations of ocular surface structure segmentation performance. From left to right: original images, manual annotations of different ocular surface structures, and U-Net segmentation performance. Pink circles indicate the challenging regions for the automated segmentation method. Yellow arrows indicate the visualisation of relevant quantitative indicators. Red arrows indicate over- and under-segmentation regions by the existing automated method.

I. INTRODUCTION

PATHOLOGICAL alterations of ocular surface structures such as meibomian gland, tear meniscus and conjunctiva, are associated with meibomian gland dysfunction, dry eye syndrome, and pterygium, which are among the most common conditions in ophthalmology [1] and can cause different adverse consequences. Severe cases of dry eye syndrome, for instance, can result in permanent vision loss [2], while pterygium encroaching upon the pupil area can lead to apparent vision impairment [3]. Therefore, early detection and precise diagnosis of ocular surface diseases are of utmost importance.

In clinical practice, morphological quantification of ocular surface structures or lesions is routinely performed to support the diagnosis and monitor ocular surface diseases. For instance, as shown in Fig. 1, the extent of corneal involvement of the pterygium is commonly used in the diagnosis and classification of pterygium severity [4]; the morphological

characteristics and loss of meibomian glands are often crucial in the diagnosing of meibomian gland dysfunction and blepharitis [5]; tear meniscus height, and rate of meibomian gland loss provide significant indicators for the diagnosis of dry eye syndrome [6]. However, the manual annotation and measurement of these structures in ocular surface images is time- and labour-consuming: thus, the automated detection of these structures is of great importance in clinical practice.

In the last two decades, we have witnessed the a rapid development of automated extraction of ocular surface structures. Previous studies on the segmentation of ocular surface structures and lesions primarily involved traditional algorithms based on morphology [7], filter [8] and threshold [9], [10]. Recently, deep learning has been employed for the ocular surface image segmentation, such as IF-Net, Group-PPM-Net, and so on [11]–[17]. Despite the fact that the deep learning-based techniques have achieved promising results in single structure segmentation, the aforementioned methods only designed a specific method for the detection of a single structure, and there exists no universally applicable ocular surface structure segmentation method. This implies that designing and training a new model will always be required for the segmentation of other structures, leading to lower efficiency in terms of timeliness.

From the perspective of image analysis, there remain challenges regarding ocular surface structures. Most of the aforementioned deep learning-based methods utilize an encoder-decoder architecture for pixel-level classification prediction of ocular surface structures, and often perform poorly in regions with inconspicuous boundaries due to the illumination condition, or acquisition environment [18]. For example, the top row of Fig. 1 illustrates pterygium segmentation performances by a human expert, and by an automated method based on U-Net [19]. It can be seen clearly that the pterygium region with inconspicuous boundaries (the pink oval in the left image) is hard to detect with precision using U-Net. In addition, the specular highlight caused by light reflection, as indicated by the pink oval region in the left two bottom rows of Fig. 1, may further aggravate segmentation errors. In these cases, U-Net-like models may produce poor segmentation results: fragmented structures, and jagged edges. These phenomena are mainly caused by pixel-level-based classification placing primary emphasis on inferring the labels of individual pixels, which may lead to overlooking the intrinsic morphological features of the ocular surface structures [20]–[22].

A. Motivations

Generative models may effectively alleviate the problem of lacking global information in pixel-level segmentation [23], [24], and diffusion models are potential solutions to address the above issues. Diffusion models learn the underlying distribution of data by modeling the denoising process [25], and recently diffusion models have been widely-utilized in medical image segmentation, such as SegDiff [26], MedSegDiff [27], and SCDM [17]. Diffusion models may compensate for the morphological distortions caused by inconspicuous boundaries and specular highlights, by learning the prior distribution of manual annotations [28].

However, diffusion models place more attention on the overall distribution of the generated results, and struggle to constrain the randomness of the generation process [29]. The randomness of the diffusion model helps it achieve excellent performance in the generation field, but causes unsatisfactory performance in segmentation tasks with a unique manual annotation [26]. There are three challenges for the randomness of diffusion models in segmentation tasks. First, it is difficult for a network with a fixed architecture to regulate the utilization of the information at different steps of the generation process. Under-utilization of image information may lead to a lack of structural constraints, making the generation process uncontrollable [17]. Second, the input to the diffusion model includes noise of varying intensities at different times. Due to the random mapping of the morphological distribution into Gaussian noise, the segmentation results are inevitably disturbed [27]. Last, but not least, the diffusion model removes predicted noise iteratively to recover the target. While this indirect prediction can generate diverse results and be beneficial in some use cases, it may also lead to error accumulation.

B. Contributions

Being well acquainted with the above challenges, in this work, we aim to propose a universal segmentation framework for multiple ocular surface structures. By incorporating morphological priors of ocular surface structures seamlessly with a diffusion model, we compensate for the morphological distortions caused by ambiguous structure boundaries and specular highlights, with each of the elements addressing one of the aforementioned challenges. Our contributions can be summarized as follows:

- For the first time, we designed a universal ocular surface structures segmentation method that can detect the pterygium, pupil, cornea, eyelid, meibomian glands, meibomian gland area, and tear meniscus, within a single model. The experimental results have shown the model's superiority in segmentation, and potential to make a strong clinical impact in ocular surface disease grading and diagnosis. The proposed method offers ophthalmologists a standardized tool, alleviating the model selection issue for segmentation of each different structure. The datasets are available on request, and the code is available at <https://github.com/iMED-Lab/Randomness-restricted-Diffusion-Model>.
- We proposed a novel randomness-restricted diffusion model, to compensate for morphological distortions caused by ambiguous structure boundaries and specular lights by learning the morphological priors. To alleviate the randomness of the diffusion model in segmentation tasks, we introduce a time-controlled fusion-attention module (TFM). It dynamically adjusts the information flow within the network depending on the relationships between the network inputs at different times.
- To mitigate the disturbance caused by noise at different times on the segmentation results, we introduced a low-frequency consistency filter (LCF) to selectively filter out high-frequency noise components from the multi-scale features extracted from the encoder at different times. Furthermore, we utilized the diffusion model to directly predict masks, and introduce a new loss function to mitigate the error accumulation

resulting from indirect predictions.

II. RELATED WORK

A. Ocular surface image segmentation

Early work on segmenting ocular surface structures and lesions primarily relied on traditional image processing algorithms [7]–[10]. Yang *et al.* [7] used pixel intensity gradients and morphological operations to extract the lines along the centre of and between the glands. Turgay *et al.* [8] used 2D Gabor filters to estimate the width and stretching ratio of the local structure to segment the meibomian gland. Gao *et al.* [9] developed a method based on the colour difference between pterygium and pupil to detect pterygium within the pupil region. Yang *et al.* [10] implemented automatic tear meniscus segmentation using a multi-threshold algorithm. However, these methods were designed for a specific single task and often have limitations in terms of segmentation performance and robustness.

With the advancement of deep learning, pixel-level classification networks such as U-Net have been widely applied to ocular surface image segmentation [11]–[16]. Setu *et al.* [11] used a pre-trained U-Net model for meibomian gland segmentation. Liu *et al.* [12] proposed an IF-Net for segmentation of meibomian glands and missing areas, which used an information fusion module and a parallel path connection to reduce information loss and compensate for information insufficiency. Abdani *et al.* [13] proposed a group PPM network for pterygium segmentation, which included a spatial pyramid pooling (PPM) module to improve the network’s ability to detect pterygium at different stages. Wan *et al.* [14] extended DeepLabv3 with an ASPP module to improve the network’s multi-scale contextual awareness for tear meniscus segmentation; Wan *et al.* [15] added an attention gate to the U-Net++ architecture to extract pterygium-related information; Deng *et al.* [16] introduced a multiscale perceptual block on top of the U-Net used to segment the meibomian gland region. These methods have achieved better generalisation and accuracy compared to traditional approaches. However, they are designed for individual ocular surface structures. Furthermore, due to the limitations of pixel-level classification in capturing semantic relationships, the challenges of ambiguous structure boundaries and specular highlights remain in current research.

B. Diffusion models

Diffusion models may capture the underlying distribution of the original data through learning the reverse diffusion process, providing a potential approach for learning the morphological priors of ocular surface structures. Inspired by non-equilibrium statistical physics, Jascha *et al.* [30] firstly defined the basic form of diffusion models. Building on this foundation, Ho *et al.* [31] proposed the denoising diffusion probabilistic models (DDPMs), providing a reliable parameterization scheme for training diffusion models. Nichol *et al.* [32] introduced a series of improvements to diffusion models, including noise selection and sampling acceleration, significantly enhancing the generation quality of these models. Prafulla *et al.* [33] used

the gradient of an image classifier to guide the reverse diffusion process, outperforming generative adversarial networks in generation tasks and demonstrating the potential applications of diffusion models.

Currently, diffusion models have been applied to several medical image segmentation tasks. These studies use images as conditional inputs to control the generation of segmentation results. Guo *et al.* [17] proposed skipping the early diffusion process using pre-segmentation results, which provides initial structural constraints. Wu *et al.* [27] proposed to remove high-frequency noise components from the noised mask to obtain the regions of interest in the image. However, due to the variability of noise levels in masks during the diffusion process, the mask cannot provide effective regions of interest in the early stages of the reverse process. Tomer *et al.* [26] proposed to use multiple sampling iterations to mitigate the stochasticity of the sampling process in diffusion models, specifically for nuclei segmentation. Chen *et al.* [34] proposed to employ the stochastic generation of multiple masks using diffusion models to emphasize regions of interest, thereby providing valuable insights for clinical research. However, repeated sampling [26], [34] comes at the cost of increased computational time.

While the introduction of morphological priors can enhance the anatomical plausibility of segmentation results, the inherent randomness of diffusion models may lead to some generated outcomes that do not align with the image.

III. METHODS

This section will present the proposed randomness-restricted diffusion model. It comprises the overall network architecture, the time-controlled fusion-attention module, the low-frequency consistency filter, and the loss function.

A. Overview

Fig. 2 illustrates the overall diagram of the proposed randomness-restricted diffusion model (R²diff), which consists of a forward and a reverse process. The forward process is a fixed-parameter Markov chain with T time steps. As shown by the yellow arrow in Fig. 2, the mask $x_0 \sim q(x_0)$ is gradually degraded with added random noise at each time step t , until it is transformed into a Gaussian noise $N(\mathbf{0}, \mathbf{I})$, defined as:

$$q(x_t | x_{t-1}) = N(x_t; \sqrt{1 - \beta_t}x_{t-1}, \beta_t \mathbf{I}), \quad (1)$$

where β_t is a pre-defined parameter that monotonically increases with t . We define α_t as $1 - \beta_t$ and $\bar{\alpha}_t$ as $\prod_{s=0}^t \alpha_s$: the noised mask x_t at each time step can then be defined as:

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon, \epsilon \sim N(\mathbf{0}, \mathbf{I}). \quad (2)$$

To recover the mask x_0 from noise, we use neural networks to simulate the reverse process of the diffusion model (the green arrow in Fig. 2), which is defined as another Markov chain [17]:

$$p_\theta(x_0) = p(x_T) \prod_{t=1}^T p_\theta(x_{t-1} | x_t, y, t), \quad (3)$$

$$p_\theta(x_{t-1} | x_t, y, t) = N(x_{t-1}; \mu_\theta(x_t, y, t), \sigma_t^2 \mathbf{I}), \quad (4)$$

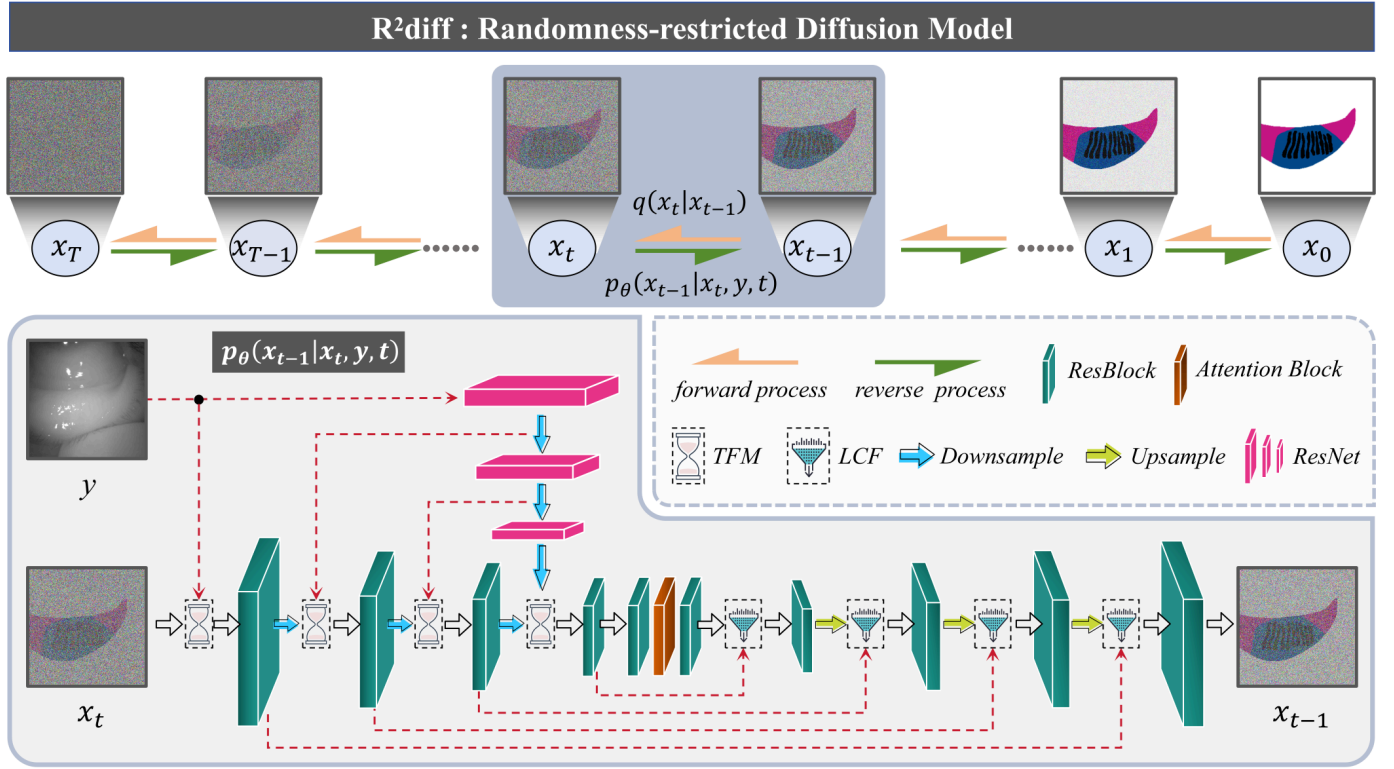


Fig. 2: Architectures of the proposed randomness-restricted diffusion model ($R^2\text{diff}$), and the reverse process. Architectures of the proposed randomness-restricted diffusion model ($R^2\text{diff}$), and the reverse process. The inputs to the network include the noised mask x_t and the original image y . In each layer of the encoder, the noised mask feature and the image feature are simultaneously fed into TFM, which can adjust the weights of these two features based on their relationship at different time steps. In each layer of the decoder, LCF uses the higher-level feature as a reference to filter out the noise components of the skip-connection feature.

where y represents the ocular surface image, $\mu_\theta(x_t, y, t)$ is obtained from the prediction of the network, and σ_t^2 is set to β_t . As shown in Fig. 2, the basic network is a combination of a U-Net and a ResNet [35]. The U-Net has six layers, with each layer having [C, C, C, 2C, 4C, 8C] channels, respectively. In the U-Net backbone, each layer is constructed using ResBlocks, and an attention block is applied to the lowest layer to enhance the network’s capability for contextual modeling. Additionally, the time steps t are introduced as sinusoidal time embeddings (4C-dimensional vectors) and then passed through a linear layer into each ResBlock. The implementation details of U-Net mentioned above are identical to the network in [32]: C is set to 64. The ResNet acts as a conditional network, encoding the image into multi-scale features that are progressively embedded into each layer of the U-Net according to their resolution. The input to the network consists mainly of the noise mask x_t and the original image y , with the output being the mask x_0 . Due to the varying noise intensity of the noise mask, to avoid overfitting, we propose that in each layer of the U-Net’s encoder, the noise mask feature M and the image feature I are simultaneously fed into the time-controlled fusion-attention module (TFM). On the other hand, to prevent the noise components of the mixed multi-scale features, which contain morphological priors, from

generating non-existent structures, we use the Low Frequency Consistency Filter (LCF) to filter out these noise components before each layer of the decoder.

B. Time-controlled Fusion-attention Module

In the proposed segmentation framework, the diffusion model takes the noised mask x_t and the ocular surface image y as the input. As indicated in Eq.(2), $\bar{\alpha}_t$ is a parameter that decreases with time t , which suggests that the noise intensity of the x_t increases with time t and the mask information in x_t continues to decrease. Therefore, x_t gradually transits from x_0 to an approximate Gaussian noise as time t increases. During the training process, we use the network to simulate the denoising process at random time t . As shown in Fig. 2, in the early stages of the diffusion process, x_t has a high similarity to x_0 due to the weaker noise intensity. The network inevitably takes shortcuts, and at this point, the primary information flow is from x_t to x_0 . The fixed network architecture design hinders the network’s ability to automatically adjust the information flow at different times. This can lead to the diffusion model’s inability to effectively utilize image information when the noise intensity in x_t is strong. Thus, the generation process of the diffusion model may be uncontrollable when there is a lack of image information constraints.

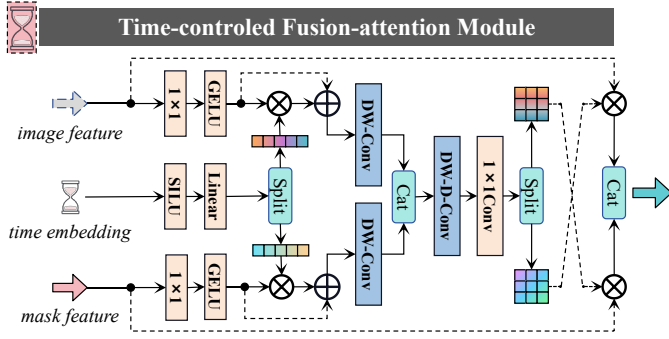


Fig. 3: of the time-controlled fusion-attention module. TFM embeds the time embedding into image and mask features using a linear layer. It then leverages depth-wise and depth-wise dilation convolutions to explore the spatial relationships between these features. By integrating temporal and spatial information, the module generates two attention maps, which are used to adjust the weights between the two types of features.

Consequently, we propose a time-controlled fusion-attention module (TFM) to dynamically adjust the flow of information within the network based on the relationship between the noised mask x_t and the image y at different times. We use the TFM instead of each ResBlock in each encoder layer of the U-Net. Notably, the noise intensity of x_t at different time steps, and the information in the image y remains unchanged over time. Therefore, as shown in Fig. 3, we use two 1×1 convolutions and GeLU activation to adjust the channels of the image feature I and mask feature M to c channels respectively, where c is the number of channels in the current layer of the U-Net. The time embeddings are passed through a SILU activation and a linear layer to generate two channel-attention maps (c -dimensional vector), used to adjust channel weights of the features: and thus we can obtain two weighted features $\bar{I}$ and $\bar{M}$:

$$atten_I, atten_M = \text{split}(\text{SILU}(\text{Conv}_{1 \times 1}(emb))), \quad (5)$$

$$\bar{I} = (atten_I \oplus 1) \otimes \text{GELU}(\text{Conv}_{1 \times 1}(I)), \quad (6)$$

$$\bar{M} = (atten_M \oplus 1) \otimes \text{GELU}(\text{Conv}_{1 \times 1}(M)), \quad (7)$$

where ‘split’ represents separating channels in the vector, ‘emb’ denotes the time embedding. $\oplus$ and $\otimes$ represent addition and dot multiplication with broadcast mechanism, respectively.

To further regulate the complementarity between these two types of information, we introduce large kernel attention to generate more refined spatial-attention maps, based on the relationship between the image and the noised mask at different time steps. The spatial-attention maps are then employed interchangeably on both the image and the mask feature, enhancing the network’s complementary utilization of these two features. The mechanism can be represented as follows:

$$F = \text{cat}(\text{DW-Conv}(\bar{I}), \text{DW-Conv}(\bar{M})), \quad (8)$$

$$map_i, map_m = \text{split}(\text{Conv}_{1 \times 1}(\text{D-W-D-Conv}(F))), \quad (9)$$

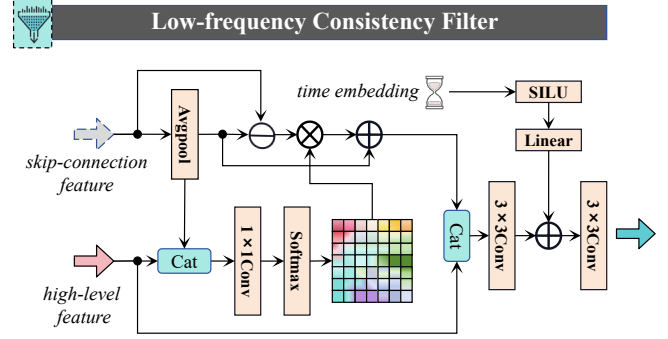


Fig. 4: Illustrations of the low-frequency consistency filter. LCF extracts the high-frequency and low-frequency components of the skip connection feature separately using average pooling. Subsequently, an attention map is generated based on the consistency between the low-frequency components and high-level features, which is used to eliminate the noise in the high-frequency components. Finally, convolutional layers are applied to repair the defects of the filtered features according to different time steps.

$$out = \text{cat}(map_m \times I, map_i \times M). \quad (10)$$

where ‘cat’ and ‘split’ respectively refer to concatenating and separating channels in features. DW-Conv is the 5×5 depth-wise convolution, and D-W-D-Conv is the 7×7 depth-wise dilation convolution [36].

C. Low-frequency Consistency Filter

In the diffusion model, the distribution of the mask is randomly mapped to the Gaussian distribution. While this mapping enables the diffusion model to learn the morphology distribution of the mask, it can also result in the generation of some non-existing structures. To address the noise disturbances of varying intensities at different times, we propose a low-frequency consistency filter (LCF) to selectively remove high-frequency noise components of the skip-connection feature S , which is outside of the region of interest. LCF is used instead of the first ResBlock in each layer of the decoder.

The higher-level feature H obtained through more layers in the encoder is often cleaner. While high-level features may lack local details, they can provide more accurate localization. The skip-connection feature is often used to restore details, which might otherwise preserve much irrelevant information. By leveraging the consistency between higher-level features and the low-frequency component in the skip-connection feature, we can acquire the approximate region of interest. Specifically, taking into account that the average filter is sensitive to Gaussian noise, we use an average filter for extracting low-frequency components. As shown in Fig. 4, we implement this functionality in the network using a 5×5 average pooling layer, denoted as:

$$S_L = \text{avgpool}(S), S_H = S - S_L, \quad (11)$$

where S_L and S_H represent the low- and high-frequency components. Then, we concatenate the low-frequency components with the high-level features and use a 1×1 convolution

followed by a softmax activation to obtain the attention map of the region of interest. The map is subsequently multiplied by S_H to coarsely filter out the noise components. The operation can be described as:

$$\text{map} = \text{softmax}(\text{Conv}_{1 \times 1}(\text{cat}(S_L, H))), \quad (12)$$

$$\bar{S} = \text{cat}(S_L + S_H \times \text{map}, H), \quad (13)$$

where $\bar{S}$ is the skip-connection feature after coarse filtering.

Due to varying noise intensities at different times, the feature $\bar{S}$ exhibits varying degrees of defects after coarse filtering. Therefore, we use convolutional modules to adaptively repair defects at different times. As shown in Fig. 4, we employ two 3×3 convolutions for fine repairing. Taking into account that the noise intensities vary at different time steps, we embed the time between these two convolutions to enhance the network's adaptive capability of repair. Specifically, we use a SILU activation and a linear layer to adjust the dimension of the time embedding and match the channels of the feature after the first convolution. Then we add the time embedding and the feature as input for the second convolution.

D. Loss Function

Effective training of a diffusion model can be performed by optimizing the variational lower bound [31]:

$$\mathcal{L}_{t-1} = \mathbb{E}_q \left[\frac{1}{2\sigma_t^2} \|\tilde{\mu}_t(x_t, x_0) - \mu_\theta(x_t, y, t)\|^2 \right] + C, \quad (14)$$

where C represents the constant term that is independent of the gradient calculation of the network parameters. And $\tilde{\mu}_t(x_t, x_0)$ denotes the mean of the posteriors $q(x_{t-1} | x_t, x_0)$, which can be inferred using Bayesian theory [31].

There are many ways to obtain $\mu_\theta(x_t, y, t)$, and the most commonly used one is to make the network predict the noise ϵ . In this case, $\tilde{\mu}_t$ can be expressed as:

$$\tilde{\mu}_t(x_t, x_0) = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{\beta_t}{\sqrt{1 - \alpha_t}} \epsilon \right). \quad (15)$$

The loss function can be denoted as:

$$\mathcal{L}_{\text{noise}} = \mathbb{E}_{t, x_0, y, \epsilon} \left[\|\epsilon - \epsilon_\theta(x_t, y, t)\|^2 \right], \quad (16)$$

and $\mu_\theta(x_t, y, t)$ can be represented as:

$$\mu_\theta(x_t, y, t) = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{\beta_t}{\sqrt{1 - \alpha_t}} \epsilon_\theta(x_t, y, t) \right). \quad (17)$$

Prediction of the noise can enhance the diversity and authenticity of the diffusion model's generation [32]. However, in segmentation tasks with a unique ground truth, predicting noise implies an indirect prediction of x_0 according to Eq.(2). As a result, errors at each time step of the backward process would accumulate, ultimately leading to generating results that do not meet expectations.

To mitigate the accumulation of errors arising from the indirect prediction of x_0 , we opt to train the network for the direct prediction of x_0 . In this case, we can obtain an alternative parameterization method through Bayesian theory or substituting Eq. (2) into Eq. (15):

$$\tilde{\mu}_t(x_t, x_0) = \frac{\sqrt{\alpha_t - 1}\beta_t}{1 - \alpha_t} x_0 + \frac{\sqrt{\alpha_t}(1 - \alpha_{t-1})}{1 - \alpha_t} x_t, \quad (18)$$

TABLE I: Demographics of MOSS. Here, DE stands for dry eye, MGD for meibomian gland dysfunction, RE for refractive errors. P-I, P-II, and P-III represent Grade 1, Grade 2, and Grade 3 pterygium, respectively.

	MOSS-M	MOSS-T		MOSS-P
Age(yo)	41.86±11.90	37.76±11.99	Age(yo)	43.21±7.64
Sex(F/M)	93/62	85/65	Sex(F/M)	53/92
DE(n)	113	118	P-I(n)	53
MGD(n)	40	42	P-II(n)	61
Conjunctivitis(n)	23	23	P-III(n)	31
RE(n)	24	24	-	-
Cataracts(n)	22	22	-	-

to this end, $\mu_\theta(x_t, y, t)$ can be represented as:

$$\mu_\theta(x_t, y, t) = \frac{\sqrt{\alpha_t - 1}\beta_t}{1 - \alpha_t} x_\theta(x_t, y, t) + \frac{\sqrt{\alpha_t}(1 - \alpha_{t-1})}{1 - \alpha_t} x_t. \quad (19)$$

The resultant loss function can be simplified as follows:

$$\mathcal{L}_{\text{simple}} = \mathbb{E}_{t, x_0, y, \epsilon} \left[\|x_0 - x_\theta(x_t, y, t)\|^2 \right], \quad (20)$$

Importantly, it is often the case that small ocular surface structures, such as the tear meniscus, are disproportionately affected by class imbalance. To this end, we propose a label-frequency weight scheme that assigns greater importance to smaller targets. The final loss function is defined as:

$$\mathcal{L}_{\text{weight}} = \mathbb{E}_{t, x_0, y, \epsilon} \left[\|\sqrt{w} \times (x_0 - x_\theta(x_t, y, t))\|^2 \right], \quad (21)$$

where the weight matrix w is computed based on the label frequencies of each pixel in x_0 , which has the same shape as x_0 :

$$w_{i,j} = \frac{N - \Phi_{x_0}(c_{i,j})}{N}, \quad (22)$$

where $w_{i,j}$ is the element of w at the position (i, j) , $c_{i,j}$ represents the label category at position (i, j) of x_0 , $\Phi_{x_0}(c_{i,j})$ denotes the number of pixels of category $c_{i,j}$ in x_0 , and N is the number of all the pixels in x_0 .

IV. DATASET

To the best of our knowledge, automated ocular surface image processing is rarely explored by contrast with retinal image analysis, and there are few publicly available ocular surface image datasets with manual annotations. This hinders the validation of automatic models, and has limited the progress of ocular surface image segmentation in the medical image analysis community. To this end, we have carefully designed a **Multiple Ocular Surface Structure** dataset (**MOSS**) in order to train, test and validate the proposed ocular surface structure segmentation approach for seven different structures detection purposes. All the images in MOSS were acquired using digital cameras equipped with infrared imaging capabilities, and with an image resolution of 1200×1600 pixels.

MOSS contains three subsets: 1) **MOSS-M** contains 155 images from 155 participants. The meibomian gland (MG), meibomian gland area (MGA), and eyelid were manually annotated. 2) **MOSS-T** comprises 150 ocular surface images from 150 subjects. The tear meniscus was manually annotated. 3) **MOSS-P** includes 145 ocular surface images from 145 patients with pterygium. The pterygium, cornea, and pupil areas were manually annotated. It is noted that MOSS-P was

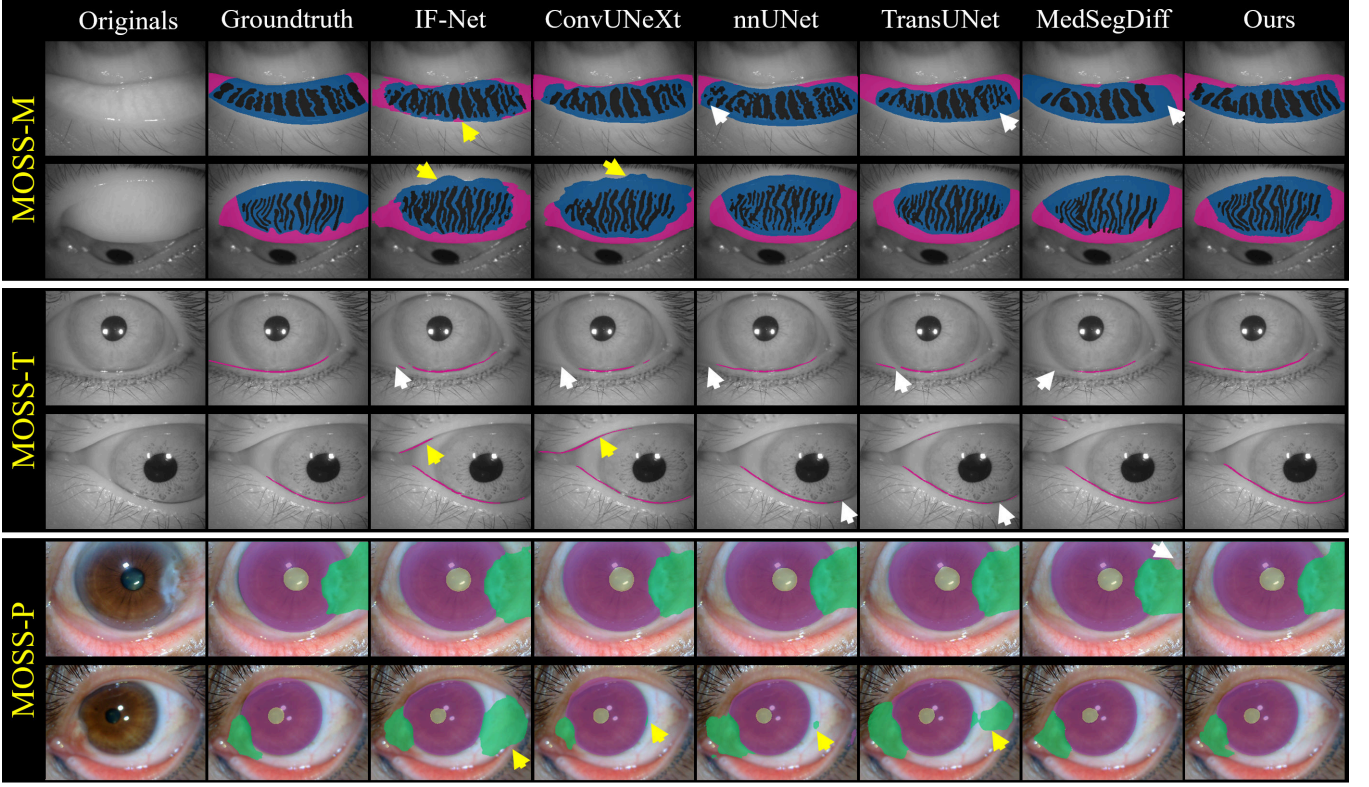


Fig. 5: The detection results of different methods on two randomly selected cases from MOSS-M, MOSS-T, and MOSS-P respectively. Top: MG (black), MGA (blue) and eyelid (pink) detection results; Middle: tear meniscus segmentation results; Bottom: cornea (pink), the pupil (gray), and pterygium (green) segmentation results. Over- and under-segmentation are marked with yellow and white arrows, respectively.

captured using a digital camera, while MOSS-M and MOSS-T were obtained with an infrared camera, due to the insufficient clarity of the meibomian glands and the tear meniscus using visible light cameras. We invited three senior ophthalmologists to annotate seven different ocular surface structures, and their consensus was used as the ground truth. All annotations were performed using the ITK-SNAP [37] software. Datasets are available on request.

To facilitate model training, all images were resized to a resolution of 256×256 . Each dataset was split into training, validation, and test sets in a 3:1:1 ratio. All data were collected at the Eye Hospital of Wenzhou Medical University, Wenzhou, China. The introduction to the participant demographic is provided in Table I. All the data used in this study were approved by the institutional ethics committees and written informed consent from each participant is acquired in accordance with the Declaration of Helsinki.

V. EXPERIMENTAL RESULTS

In this section, the implementation details, evaluation metrics, and experimental results are provided.

A. Implementation details

All experiments were conducted using the publicly available Pytorch library, utilizing two NVIDIA GeForce RTX 3090 GPUs. During the training phase, we employed the AdamW optimizer to optimize the model, with a learning rate set to

0.0001; the batch size was set to 8. For integrity, we applied the same preprocessing steps to all images in this paper, including random rotation, random horizontal flipping, random vertical flipping and random adjustments of brightness, contrast and saturation. The number of diffusion steps T in this paper is set to 100, and β_t is defined as a linearly increasing variable from $\beta_1 = 10^{-4}$ to $\beta_T = 0.02$. For other diffusion model methods, we uniformly use the hyperparameters settings in their papers or codes. For other pixel-level classification methods, we trained them using both the cross-entropy and the Dice loss functions simultaneously.

B. Results

Ten state-of-the-art segmentation methods were compared with the proposed method to validate the effectiveness of the latter. These methods include three specifically designed ocular surface image segmentation methods: IF-Net [12], Improved U-Net++ [15], and Group-PPM-Net [13]; four general medical image segmentation methods: nnU-Net [38], TransUNet [39], Swin-Unet [40], and ConvUNeXt [41]; and three diffusion-based segmentation methods: SegDiff [26], MedSegDiff [27], and SCDM [17]. Dice coefficient score (Dice), G-mean score (G-mean), Kappa score, and 95% Hausdorff Distance (HD95) were calculated for evaluation.

1) *Subjective comparisons:* Fig. 5 illustrates the visualization results using different approaches over two randomly selected images from MOSS-M, MOSS-T, and MOSS-P. It

TABLE II: Quantitative comparison with those obtained by state-of-the-art segmentation methods. The best and second-best methods are represented in red and blue, respectively.

Metrics	Method	MOSS-M			MOSS-T		MOSS-P	
		MG	MG Area	Eyelid	Tear meniscus	Pterygium	Pupil	Cornea
Dice%	Improved U-Net++	71.30±10.27	83.57±11.62	93.11±6.36	73.40±8.73	83.62±11.56	95.90±2.69	96.05±2.37
	IF-Net	70.91±6.34	84.29±7.01	91.55±5.25	72.10±10.27	84.15±20.15	96.71±1.54	96.57±1.45
	Group-PPM-Net	71.53±8.53	82.99±8.28	90.10±7.71	70.95±10.43	79.24±18.38	96.18±1.64	96.30±1.74
	nnU-Net	72.04±8.50	86.92±7.54	94.57±3.95	71.37±14.16	87.98±10.36	97.04±1.05	96.98±1.68
	Swin-Unet	72.68±6.66	87.34±6.38	95.02±3.29	73.25±6.77	87.84±7.66	96.73±1.12	96.69±1.36
	TransUNet	72.22±9.60	87.27±7.41	94.53±3.16	73.30±8.48	88.12±8.95	96.48±1.82	96.68±1.84
	ConvUNeXt	70.94±8.25	87.38±4.88	93.90±3.78	70.22±13.49	88.64±6.85	96.94±1.69	96.83±1.74
	SegDiff	65.44±12.29	79.12±9.06	91.43±6.74	60.93±20.19	75.72±21.86	85.70±16.11	92.33±6.95
	MedSegDiff	67.08±9.06	86.15±4.92	93.90±4.90	61.46±17.90	89.46±6.54	96.31±1.36	96.90±1.55
	SCDM	70.53±8.29	86.86±5.39	93.67±5.86	67.50±9.69	90.24±5.35	97.28±1.01	97.04±1.50
	Ours	74.14±6.49	88.72±5.63	95.23±3.57	75.53±6.20	92.01±4.13	97.38±1.06	97.18±1.56
G-mean%	Improved U-Net++	79.71±9.76	87.08±9.63	95.11±4.37	86.51±7.38	91.13±5.43	96.35±2.68	97.03±1.65
	IF-Net	84.57±5.95	89.15±4.24	93.79±3.42	89.09±9.80	89.10±18.08	97.84±1.75	97.26±1.29
	Group-PPM-Net	82.79±9.53	89.71±5.26	92.90±4.65	90.01±9.32	87.74±17.98	96.67±1.83	97.02±1.36
	nnU-Net	82.50±8.13	90.29±7.12	96.48±1.60	84.04±14.82	92.94±7.83	97.77±1.39	97.72±1.25
	Swin-Unet	84.18±6.43	90.02±5.99	96.51±1.48	89.33±5.26	92.39±6.51	97.74±1.30	97.37±1.12
	TransUNet	82.19±9.48	90.72±6.82	96.04±1.61	87.03±6.86	93.45±3.24	97.60±2.16	97.31±1.67
	ConvUNeXt	83.12±8.73	91.98±4.11	95.60±2.03	86.01±12.17	92.20±5.48	98.03±1.80	97.83±1.12
	SegDiff	79.98±9.08	84.96±9.05	94.82±3.73	85.66±15.71	87.06±18.99	97.27±1.56	94.31±4.86
	MedSegDiff	80.59±9.19	91.09±4.17	96.28±2.53	78.35±21.92	94.07±5.12	97.74±1.60	97.79±0.95
	SCDM	80.64±8.44	91.08±4.56	96.23±2.43	78.82±10.73	94.08±3.77	98.21±1.11	97.72±1.16
	Ours	85.74±5.73	92.19±4.95	97.24±0.97	92.05±4.50	94.47±3.00	98.26±1.25	97.89±1.20
Kappa%	Improved U-Net++	69.04±10.40	79.40±13.57	89.93±9.84	73.15±8.79	81.17±12.13	95.82±2.71	94.20±3.08
	IF-Net	68.08±6.65	79.69±8.28	87.91±7.04	71.82±10.32	82.32±20.35	96.64±1.56	94.88±2.08
	Group-PPM-Net	69.04±8.58	77.68±9.29	85.92±9.21	70.65±10.50	76.30±18.66	96.10±1.67	94.45±2.59
	nnU-Net	69.60±8.69	83.37±8.52	92.19±4.94	71.09±14.32	86.29±10.93	96.98±1.06	95.52±2.31
	Swin-Unet	70.14±6.84	83.89±7.17	92.95±3.87	72.97±6.83	85.08±10.48	96.31±1.39	95.03±1.99
	TransUNet	69.85±9.72	83.71±8.45	92.17±3.75	73.04±8.54	86.35±9.40	96.41±1.83	95.03±2.68
	ConvUNeXt	68.27±8.49	83.45±5.72	91.28±4.53	69.95±13.56	86.90±7.40	96.88±1.70	95.24±2.47
	SegDiff	62.16±13.43	73.26±11.18	87.23±9.86	60.40±20.68	72.43±22.73	85.29±16.64	88.26±11.02
	MedSegDiff	64.05±9.34	81.75±6.34	91.19±6.58	61.10±18.01	87.78±7.24	96.23±1.37	95.33±2.24
	SCDM	68.03±8.38	82.98±6.20	90.99±7.57	67.23±9.75	88.76±5.64	97.25±1.08	95.53±2.14
	Ours	71.68±6.81	85.41±6.61	93.19±4.29	75.27±6.25	90.72±4.62	97.32±1.09	95.81±2.19
HD95	Improved U-Net++	12.70±16.06	26.77±19.01	15.83±17.52	8.02±15.15	61.79±58.09	1.86±1.06	11.51±14.27
	IF-Net	11.50±7.05	27.44±15.40	20.13±17.26	12.69±28.77	41.53±63.16	3.18±7.00	9.30±6.92
	Group-PPM-Net	12.70±13.94	27.45±9.66	20.35±9.39	12.87±24.51	68.42±65.78	1.65±0.66	8.01±4.55
	nnU-Net	11.01±9.45	23.47±16.11	14.83±18.64	10.90±37.14	21.01±20.42	1.24±0.32	10.62±15.48
	TransUNet	10.94±10.49	24.17±15.64	12.44±7.60	13.20±29.56	22.24±29.52	1.63±0.62	7.62±5.45
	ConvUNeXt	11.54±9.58	26.45±17.76	16.39±16.47	12.85±22.92	26.03±35.04	1.41±0.43	6.75±5.13
	Swin-Unet	9.27±7.41	23.10±14.69	13.43±17.60	7.35±19.00	14.74±7.27	1.40±0.48	6.44±3.77
	SegDiff	28.94±33.77	38.49±18.35	28.26±26.11	25.44±50.50	80.66±66.61	47.08±44.12	37.17±33.33
	MedSegDiff	20.39±22.04	28.51±14.14	14.25±12.39	18.22±34.29	15.74±8.31	4.21±8.03	6.32±4.10
	SCDM	12.26±10.38	28.63±19.60	19.20±27.40	6.39±6.53	14.94±6.47	1.25±0.39	6.70±4.22
	Ours	7.89±3.66	22.22±12.36	10.38±6.48	5.48±9.48	10.67±14.85	1.21±0.36	5.81±3.53

TABLE III: The results of the paired t-test for key structures. A p-value less than 0.05 indicates a significant difference.

Method	MG	Pterygium	Tear meniscus
Improved U-Net++	0.032	<0.001	0.034
IF-Net	<0.001	0.035	0.048
Group-PPM-Net	0.021	<0.001	0.005
nnUNet	0.023	0.023	0.071
TransUNet	0.063	0.017	0.044
ConvUNeXt	<0.001	0.002	0.03
Swin-Unet	0.005	0.002	<0.001
SegDiff	<0.001	<0.001	<0.001
MedSegDiff	<0.001	0.006	<0.001
SCDM	<0.001	0.04	<0.001
Ours	-	-	-

is observed that IF-Net and ConvUNeXt suffer from misclassification issues - they falsely identified specular reflections as eyelid and meibomian gland area, as indicated by the arrows. These two methods also encounter similar issues in tear meniscus and pterygium detection tasks: they falsely detected the upper eyelid boundary as part of the tear meniscus, while

mis-classifying the healthy areas as pterygium. Fig. 5 also reveals that nnUNet, TransUNet and MedSegDiff yield under-segmentation results - it can be clearly seen that the meibomian gland, tear meniscus, and pterygium were incompletely detected in regions with low contrast.

Overall, the proposed method demonstrates superior visual outcomes when compared with its competitors in tasks involving different ocular surface structures. Our method can produce more accurate segmentation results, while preserving the structural integrity of the meibomian glands and tear meniscus segmentation. In pterygium segmentation, the proposed method not only avoids over-segmentation, but also obtains segmentation boundaries that closely match the groundtruth.

2) Quantitative comparisons: To further assess the performance of the proposed method, the quantitative results are reported in Table II. Overall, our method achieved the best performance in the segmentation of all ocular surface structures when compared with either general segmentation methods or specially designed ocular surface segmentation methods.

TABLE IV: Quantitative comparison on the combined dataset, and the DICE score (%) is reported. The best and second-best methods are represented in red and blue, respectively.

Method	MG	MG Area	Eyelid	Tear meniscus	Pterygium	Pupil	Cornea	Average
Imprioved U-Net++	72.16	82.19	92.60	74.12	82.39	96.10	95.91	85.07
IF-Net	71.51	83.90	91.45	71.68	83.96	96.63	96.86	85.14
Group-PPM-Net	71.62	82.38	89.29	62.98	76.97	96.37	96.12	82.25
nnUNet	72.68	85.65	94.15	70.66	87.69	96.89	97.04	86.39
TransUNet	72.91	86.44	94.17	73.54	87.60	96.99	96.46	86.87
ConvUNeXt	71.67	87.11	93.56	69.73	88.27	96.84	96.78	86.28
Swin-Unet	72.26	86.39	94.62	72.01	88.72	96.76	96.60	86.77
SegDiff	64.54	78.12	88.31	49.23	74.12	92.03	85.06	75.92
MedSegDiff	67.21	86.48	93.10	64.58	90.55	96.21	96.81	84.99
SCDM	71.10	86.39	93.21	67.11	90.11	97.31	97.05	86.04
Ours	74.93	87.74	94.81	77.19	92.88	97.34	97.07	88.85

It is noted that, in MOSS-P, the diffusion-based methods demonstrated a significant advantage. For example, SCDM achieved a DICE score of 90.24% in pterygium segmentation, 97.28% in pupil segmentation, and 97.04% in cornea segmentation, ranking second only to our method. However, in datasets like MOSS-M and MOSS-T that contain curvilinear structures, diffusion-based methods are less effective than the other competing methods. For example, compared to Swin-Unet, MedSegDiff yielded 5.6% and 11.79% lower DICE scores in meibomian gland region and tear meniscus segmentation, respectively. This may be due to diffusion-based methods randomly generating structures outside of the target region. This has a substantial negative impact on the segmentation of curvilinear structures.

Compared to other comparative methods, the proposed method effectively learns the morphological priors of manual annotations, and mitigates the issue of random structure generation by the diffusion model. It can be observed that while different methods have their respective advantages, our approach achieved the best performance in all metrics. Specifically, compared to the second method, as shown in Table II, our approach achieved a 1.46% increase in DICE score for meibomian gland segmentation, a 1.34% increase in DICE for meibomian gland area segmentation, a 2.13% increase in DICE for tear meniscus segmentation, and a 1.77% increase in DICE for pterygium segmentation. Furthermore, to ensure the reliability of the experimental results, we focused our statistical analysis on three key structures: the meibomian glands, pterygium, and tear meniscus, excluding simpler structures like the cornea that were segmented well by all methods. Table III shows the paired t-test results in terms of Dice. Most $p < 0.05$ demonstrate that the proposed method performs significantly better than the other compared methods.

3) Combined dataset validation: In addition, for better validation, we used all subsets of MOSS for joint training and tested on a new combined test set. This test set contains 71 ocular surface images under different imaging conditions, including annotations for all seven ocular surface structures provided by experienced ophthalmologists.

For fair comparisons, all experimental settings adhered to those defined by subsection V-A. Table IV presents the quantitative results. Overall, the proposed R²diff achieved the best performance. For example, the average DICE score of R²diff is 1.98% higher than that of the second best method. Specifi-

cally, while other diffusion-based methods perform poorly in segmenting curved structures such as the tear meniscus, R²diff achieves optimal results in the segmentation of all structures. The above experimental results further demonstrate that R²diff is a strong and universal method for ocular surface image segmentation.

C. Ablation Studies

In order to validate the effectiveness of the TFM, LCF, and the new loss function, a series of detailed ablation experiments was conducted.

1) Effectiveness of TFM: As shown in the fourth column of Fig. 6, the addition of TFM to the network significantly improved the segmentation results in terms of fine details. Also, as seen in the second row of Table V, it can be observed that TFM effectively improved the segmentation accuracy for all structures. For example, compared to the backbone network, TFM improved the DICE score by 6.64% in the meibomian gland segmentation, by 5.94% in the tear meniscus segmentation, and by 2.75% in the pterygium segmentation. The reason for this lies in the fact that TFM enhances the utilization of the diffusion model, providing more constraints while fitting the prior shape.

2) Effectiveness of LCF: Another major contribution of this work is the low-frequency consistency filter. Comparing the third and fifth columns of Fig. 6, it can be observed that the LCF effectively improves segmentation results. Particularly, the LCF prevents erroneous generation by our diffusion model, leading to a reduction in over-segmentation regions in the tear meniscus segmentation task. This finding is also supported by the quantitative analysis, as reported in the third row of Table V. The incorporation of the LCF resulted in a 3.55% increase for meibomian gland segmentation, a 3.82% increase for tear meniscus segmentation, and a 1.67% increase for pterygium segmentation. These improvements may be attributed to the fact that the LCF can selectively filter out noise components outside the regions of interest, thus reducing the randomness of the diffusion model.

3) Effectiveness of the weighted loss function: To address the issues of error accumulation and class imbalance, a novel loss function was introduced. As shown in Fig. 7, due to error accumulation caused by prediction noise, $\mathcal{L}_{noise}$ leads to fragmented segmentation results and the generation of noise. $\mathcal{L}_{simple}$ replaces the predicted target ϵ with the mask x_0 . As

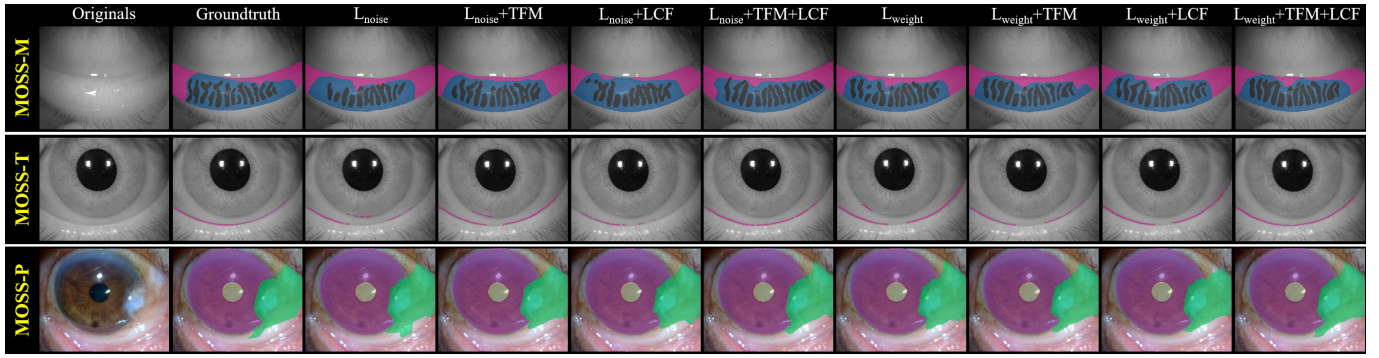


Fig. 6: Illustrative ablation results under varying combinations of different loss functions.

TABLE V: Quantitative ablation results, and the DICE score (%) is reported.

Method			MOSS-M			MOSS-T		MOSS-P	
L_{weight}	TFM	LCF	MG	MG Area	Eyelid	Tear meniscus	Pterygium	Pupil	Cornea
			64.05±13.10	85.62±8.93	95.00±4.59	66.95±16.90	85.60±17.89	96.78±1.01	97.07±1.70
	✓		70.69±9.94	87.22±8.40	94.95±5.11	72.89±9.17	88.35±16.92	96.79±1.03	97.05±1.80
		✓	67.60±11.10	86.67±7.91	94.53±4.66	70.77±9.38	87.27±17.17	96.95±1.16	97.03±1.53
	✓	✓	71.41±8.82	87.46±5.17	95.12±3.81	72.28±7.31	89.52±10.10	96.41±0.97	97.04±1.76
✓			70.49±9.78	86.43±7.09	95.22±4.33	71.20±7.03	87.13±16.98	97.05±1.29	96.92±2.06
✓	✓		72.39±7.85	87.45±5.42	95.01±3.99	73.99±6.65	91.47±4.93	96.88±1.11	97.00±1.79
✓		✓	72.16±9.40	87.24±4.96	95.13±3.78	72.18±8.60	89.34±7.60	97.02±1.07	97.10±1.53
✓	✓	✓	74.14±6.49	88.72±5.63	95.23±3.57	75.53±6.20	92.01±4.13	97.38±1.06	97.18±1.56

TABLE VI: Impact of different loss functions, and the DICE score (%) is reported.

Loss function	MOSS-M			MOSS-T		MOSS-P	
	MG	MG Area	Eyelid	Tear meniscus	Pterygium	Pupil	Cornea
L_{noise}	71.41±8.82	87.46±5.17	95.12±3.81	72.28±7.31	89.52±10.10	96.41±0.97	97.04±1.76
L_{simple}	73.39±6.73	88.04±4.64	95.02±3.92	73.21±6.78	91.37±4.80	97.24±1.25	97.18±1.55
L_{weight}	74.14±6.49	88.72±5.63	95.23±3.57	75.53±6.20	92.01±4.13	97.38±1.06	97.18±1.56

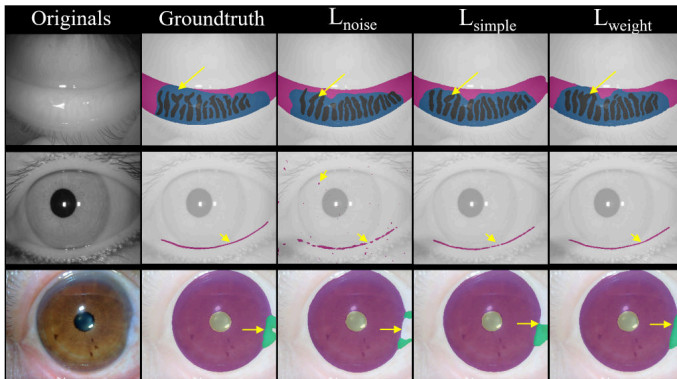


Fig. 7: Illustrative segmentation results with different loss functions.

evident in the fourth columns of Fig. 7, the new segmentation results are more precise. However, class imbalance issues also result in segmentation outcomes with discontinuities and under-segmentation. As shown in the last column of Fig. 7 the weighted loss function L_{weight} effectively further alleviates under-segmentation caused by class imbalance.

As shown in Table VI, training with L_{simple} and L_{weight} both results in better segmentation performances in different

tasks. For example, compared to L_{noise} , training with L_{simple} regularization improved DICE by 1.98% in the meibomian gland segmentation, by 0.97% in the tear meniscus segmentation, and by 1.85% in the pterygium segmentation; compared to L_{simple} , training with L_{weight} regularization improved DICE by 0.75% in the meibomian gland segmentation, by 2.32% in the tear meniscus segmentation, and by 0.64% in the pterygium segmentation. In both subjective and quantitative comparisons, it can be seen that the weighted loss function L_{weight} performs best in tear meniscus segmentation. This is because the tear meniscus is a curvilinear structure, and possesses significant class imbalance.

4) *Effectiveness of components in combination*: To validate the combined performance of the aforementioned components, all the possible combinations of TFM, LCF, and L_{weight} were emulated, as demonstrated in Fig. 6. It may be observed that the segmentation performance of any two-component combination surpasses that of any individual component. The combination of all three components yields the best segmentation results, and the reported scores in Table V further support our finding, with the results of our method achieving the best results of all methods compared. It can be observed that our method attains a DICE score of 74.14% in meibomian gland segmentation, 75.53% in tear meniscus segmentation, and 92.01% in pterygium segmentation.

TABLE VII: Quantitative comparison on the ORIGA dataset. The best and second-best methods are represented in red and blue, respectively.

Method	Optic cup				Optic disc			
	Dice%	G-mean%	Kappa%	HD95	Dice%	G-mean%	Kappa%	HD95
Imprioved U-Net++	86.02±10.60	90.81±9.59	85.26±10.91	8.93±5.78	96.84±2.13	98.11±1.71	96.25±2.52	3.96±3.35
IF-Net	86.26±8.80	91.13±8.08	85.48±9.13	9.22±5.25	96.79±2.42	97.87±1.73	96.19±2.82	3.80±3.78
Group-PPM-Net	81.15±14.85	85.39±14.21	80.25±15.10	13.20±24.33	93.64±6.52	97.52±3.84	92.44±7.46	26.06±27.07
nnUNet	87.37±8.28	92.11±7.38	86.64±8.65	7.94±4.65	96.89±2.22	98.27±1.48	96.31±2.66	3.73±3.60
TransUNet	87.45±7.78	93.40±6.26	86.69±8.15	8.23±4.58	97.06±2.57	98.29±1.68	96.52±2.98	3.37±2.57
ConvUNetXt	86.73±8.66	91.20±7.93	85.98±9.02	8.68±5.01	97.07±2.29	97.70±1.86	96.53±2.66	2.98±2.60
Swin-Unet	85.91±9.73	92.75±8.26	85.09±10.06	10.69±11.25	95.18±5.48	96.42±4.79	94.33±5.91	6.81±8.96
SegDiff	79.93±14.15	86.70±12.84	78.89±14.52	20.30±7.01	93.96±9.70	95.76±8.69	92.64±12.69	13.44±25.49
MedSegDiff	84.73±8.29	93.31±6.54	83.81±8.58	21.50±5.77	96.72±1.86	98.10±1.51	96.12±2.17	4.14±2.87
SCDM	87.56±7.89	92.93±6.68	86.84±8.22	8.12±4.88	96.99±2.20	98.21±1.32	96.42±2.66	3.69±3.57
Ours	88.52±7.11	94.16±5.80	87.82±7.44	7.46±4.29	97.64±1.74	98.81±1.17	97.21±2.01	2.39±1.31

TABLE VIII: Quantitative comparison on the OCTA3M dataset. The best and second-best methods are represented in red and blue, respectively.

Method	Retinal vessel				FAZ			
	Dice%	G-mean%	Kappa%	HD95	Dice%	G-mean%	Kappa%	HD95
Imprioved U-Net++	90.87±2.26	93.64±1.97	90.24±2.35	1.47±2.21	97.31±2.42	98.73±2.00	97.20±2.53	2.92±2.13
IF-Net	90.85±2.09	94.26±2.02	90.21±2.18	1.56±1.98	97.25±1.86	98.42±1.13	97.14±1.95	4.66±5.96
Group-PPM-Net	90.14±2.05	94.82±1.83	89.44±2.13	1.66±1.49	96.70±2.02	98.54±1.10	96.57±2.12	5.15±3.12
nnUNet	91.05±2.04	94.17±1.55	90.42±2.13	1.38±1.30	98.40±1.13	99.07±0.82	98.33±1.18	2.63±2.21
TransUNet	90.42±2.33	93.25±2.09	89.77±2.43	1.63±2.14	97.54±1.30	98.78±0.78	97.44±1.34	2.74±2.30
ConvUNetXt	90.67±2.42	93.84±2.03	90.02±2.52	1.49±2.06	97.61±1.15	98.52±1.03	97.51±1.18	2.24±1.57
Swin-Unet	86.87±2.24	92.37±1.79	85.94±2.32	1.48±1.72	95.62±1.94	97.72±1.43	95.45±1.99	5.45±2.46
SegDiff	90.40±2.10	94.56±1.82	89.72±2.19	1.64±1.95	98.10±1.52	99.05±0.94	98.03±1.57	3.34±4.53
MedSegDiff	90.25±2.19	94.73±1.64	89.56±2.28	1.77±2.30	98.08±1.60	98.66±1.44	98.00±1.67	3.16±3.36
SCDM	90.44±2.06	94.67±1.63	89.77±2.14	1.68±1.85	98.35±1.14	98.97±0.92	98.28±1.19	2.67±2.39
Ours	91.44±2.11	95.08±1.70	90.83±2.20	1.36±1.47	98.73±0.91	99.24±0.84	98.68±0.95	2.07±1.53

D. Generalization validation

To further validate the generalization of the R^2 diff on other ocualr structure segmentation tasks, we conducted experiments on two publicly available datasets, ORIGA [42] and OCTA500 [43].

The ORIGA dataset comprises 650 fundus images from 168 glaucoma patients and 482 normal subjects, with each image annotated at the pixel level for optic cup and optic disc. We validated the performance of different segmentation methods for optic cup and optic disc segmentation on the ORIGA dataset. In our experiments, we followed the dataset splitting strategy outlined by Fu *et al.* [44].

The OCTA500 dataset comprises a total of 500 subjects who underwent OCT/OCTA imaging. To accommodate the input size constraints for various comparison methods, we used the OCTA3M subset in our experiments, evaluating the performance of different methods in FAZ (foveal avascular zone) and retinal vessel segmentation tasks. Furthermore, we adopted the input format and dataset partitioning strategy as detailed by Li *et al.* [45].

For fair comparisons, all experimental settings adhered to those defined by subsection V-A. Table VII and Table VIII present the quantitative results of all the segmentation methods across different segmentation tasks. Overall, the proposed R^2 diff achieved the best performance across all segmentation tasks.

It is noteworthy that the diffusion-based segmentation methods performed well on the ORIGA dataset. For example, SCDM achieved results comparable to some segmentation networks such as nnUNet, with a Dice score 0.19% higher for optic cup segmentation. On the contrary, in the retinal vessel segmentation task, SCDM performed less effectively compared to some excellent pixel-level classification methods. For example, in retinal vessel segmentation, SCDM had a Dice score 0.61% lower than nnUNet. This is because diffusion-based methods are sensitive to curvilinear structures. While large vessels are clearly visible in OCTA3M, diffusion-based methods may show slight deviations due to their inherent randomness.

Compared to other diffusion-based methods, R^2 diff effectively reduces the randomness. Therefore, from the experimental results, it can be observed that compared to SCDM, R^2 diff improves the Dice score by 0.96% and reduces the HD95 by 0.66 in optic cup segmentation, and improves the Dice score by 1% and reduces the HD95 by 0.32 in retinal vessel segmentation. The experimental results demonstrate that the proposed R^2 diff has strong generalizability to other ophthalmic datasets and achieves competitive performance.

E. Parameters and computational efficiency

We also compared the parameters and computational efficiency of different methods. Table IX shows the parameters (Param), floating point operations (FLOPs), and inference

TABLE IX: Comparison in terms of parameters and computational efficiency.

Method	Input size	Param(M)	FLOPs(G)	Time(S)
Imprived U-Net++	256×256	9.2	34.9	0.012
IF-Net	256×256	38.3	23.1	0.014
Group-PPM-Net	256×256	3.3	13.2	0.029
nnUNet	256×256	41.2	16.3	0.182
TransUNet	256×256	93.2	32.2	0.038
ConvUNeXt	256×256	38.3	21.4	0.016
Swin-Unet	224×224	41.3	34.8	0.034
SegDiff	256×256	126.9	313.1	2.369
MedSegDiff	256×256	132.6	253.1	0.826
SCDM	256×256	149.5	323.0	1.910
Ours	256×256	81.4	67.4	0.785

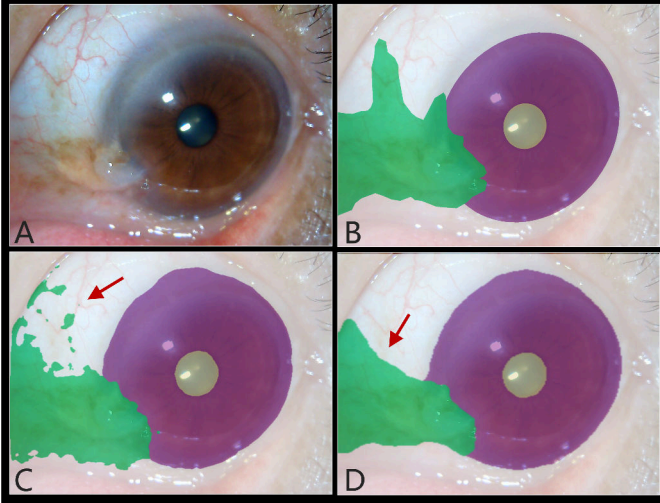


Fig. 8: Visualization of the error analysis. (A) represents an ocular surface image, (B) indicates the ground truth, (C-D) show the segmentation results obtained by U-Net and R^2 diff, respectively.

time of each image of different models at standard input sizes. Although diffusion models can improve the morphological plausibility of segmentation results, diffusion-based method typically have more parameters and FLOPs due to the additional conditional network compared to U-shaped network like TransUNet. Furthermore, diffusion models require multiple inference steps during the backward process, they typically consume more time for segmentation of a single image. Although R^2 diff may have higher computational costs compared to some U-shaped networks like Group-PPM-Net, it shows significant performance improvements. Additionally, compared to other diffusion-based methods, R^2 diff has fewer Param, FLOPs, and inference time.

F. Error analysis

The limitation of this study is the size of the dataset. Due to the limited number of ocular surface structures available for training, the morphological priors learned by the model are limited. Therefore, R^2 diff may have significant segmentation errors when segmenting some rare and complex structures. For example, the pterygium in Fig. 8 has not only a blurred border but also irregular edges, unlike the typical pterygium which has smooth edges. Therefore, while R^2 diff can effectively

ensure the completeness of segmentation results compared to pixel-level classification methods, it cannot guarantee the accurate segmentation of these rare structures. In future research, we will collect more rare samples to extend the applicability of the model.

VI. CLINICAL IMPACT

To validate the clinical impact of our segmentation method, meibomian gland dysfunction (MGD) grading and aqueous deficient dry eye (ADDE) diagnosis was conducted, based on the quantitative segmentation results.

MGD is an independent clinical dataset contains 361 images exhibiting meibomian gland dysfunction. Ophthalmologists commonly assess meibomian gland function based on the loss ratio of meibomian gland area [46]. The loss ratio LR of meibomian gland can be calculated by

$$LR = 1 - \frac{S_{MGA}}{S_{eyelid}}, \quad (23)$$

where S_{MGA} is the area of the meibomian gland region and S_{eyelid} is the area of the eyelid. Depending on the loss ratio from manual measurements, 181 images were scored as grade 1 ($LR < \frac{1}{3}$), 132 images as grade 2 ($\frac{1}{3} < LR < \frac{2}{3}$), and 48 images as grade 3 ($LR > \frac{2}{3}$) by a senior ophthalmologist (with more than 20 years' clinical experience). The higher the loss ratio, the more severe the meibomian gland dysfunction.

ADDE is another clinical dataset, comprising 175 images. All the dry eye subjects were further classified into aqueous deficient dry eye (94 images) and non-aqueous deficient dry eye (81 images) by a senior ophthalmologist (with more than 20 years' clinical experience), based on the tear meniscus height (TMH) measured on keratograph [47]. In clinical practice, a participant may be diagnosed as aqueous deficient dry eye if their TMH < 0.2 mm. In this work, we aimed to obtain the TMH by using digital cameras with infrared light. Due to variable camera positioning, direct computation of TMH based on camera parameters is not feasible. To this end, we followed [48] to obtain the tear meniscus height:

$$TMH = \frac{TMH_{pixel} \times TDC}{D_{pixel}}, \quad (24)$$

where TMH_{pixel} represents the average pixel height of the tear meniscus over a 20-pixel width centered beneath the cornea; D_{pixel} is the pixel length of transverse diameter of the cornea; and TDC is the average transverse diameter of the adult cornea, approximately 11.5mm.

Table X demonstrates the diagnostic performances of all tested methods on MGD and ADDE based on the ocular structures as extracted by different segmentation methods. Overall, the proposed method achieved the best results in MGD grading and ADDE diagnosis. Specifically, compared to the second best method (highlighted in blue in Table X), our method outperformed by relatively large margins - improving the ACC by 6.21% and Kappa by 10.75% in MGD grading, with a 4.79% increase in ACC and 9.54% in Kappa for ADDE diagnosis.

In addition, we report the MGD and ADDE diagnosis performances obtained by a relative junior ophthalmologist

TABLE X: Diagnosis results of the MGD and ADDE based on different segmentation performances.

Method	MGD		ADDE	
	ACC%	Kappa%	ACC%	Kappa%
Imprioved U-Net++	64.97	40.51	68.49	37.12
IF-Net	50.85	1.29	63.61	27.32
Group-PPM-Net	52.54	5.15	60.27	20.41
nnUNet	76.84	59.04	60.96	22.33
TransUNet	70.62	48.44	63.01	26.32
ConvUNeXt	67.80	40.09	60.96	22.50
Swin-Unet	75.14	55.79	67.81	35.66
SegDiff	38.42	10.06	52.74	5.55
MedSegDiff	70.62	47.63	56.16	12.31
SCDM	72.32	53.01	58.90	18.41
Ours	83.05	69.79	73.28	46.66
junior ophthalmologist	78.53	62.95	63.70	27.32
senior ophthalmologist	85.31	74.79	65.07	30.08

(with less than 8 years of clinical experience) and a senior ophthalmologist (with more than 20 years of clinical experience). We asked them to make a diagnostic decision by inspecting the images directly. It may be seen that our method still outperforms the results from the junior ophthalmologist - our method has improved the ACC by 4.52% and Kappa by 6.84% in MGD grading, and shows a 9.58% increase in ACC and 19.34% in Kappa for ADDE diagnosis. It is worth noting that although manual diagnosis by the junior and senior ophthalmologist is well in MGD grading, its performance is poorer in ADDE diagnosis compared with other segmentation methods. The reason is that the tear meniscus is very thin, and its height is difficult to estimate by manual inspection. In clinical practice, the keratograph is the preferred method for measuring the tear meniscus, but keratography is more expensive, and acquisition more time-consuming [49], and so is unsuitable for a large-population screening scenario.

The results of this clinical validation imply that MGD and ADDE diagnosis are two relatively difficult tasks in real-world practice if clinicians must rely solely on image inspection, highlighting that our method will be a potential tool to support the diagnosis and monitoring of diverse ocular surface diseases.

VII. CONCLUSION

In this paper, we conducted for the first time a comprehensive study on the segmentation of various structures and lesions present in different ocular surface images. We constructed three datasets annotated with multiple structures and lesions. To address the issue of morphological degradation caused by inconspicuous structure boundaries and specular highlights, we introduced a randomness-restricted diffusion model to learn the underlying morphological priors from manual annotations. However, due to the inherent randomness of the diffusion model, existing methods may generate some structures that do not actually exist. We have summarized the three reasons for the potential randomness of the diffusion model in the segmentation task with a unique ground truth: which are the lack of structural constraints due to insufficient utilization of image information; disturbance caused by noise in the input; and error accumulation caused by the loss

function. To this end, we proposed a Time-controlled Fusion-attention Module (TFM), a Low-frequency Consistency Filter (LCF), and a new weighted loss to mitigate these three issues. Extensive experiments demonstrate that the proposed method can effectively alleviate the randomness of diffusion models. Compared to state-of-the-art methods, our method can achieve a more complete structure and more accurate edges in ocular surface image segmentation. In addition, the excellent performance of the proposed method in segmenting ocular structures such as the optic cup and vessels further demonstrates its generalization capability. In the end, to verify the performance of the proposed method in clinical practice, we conducted validation in the grading of meibomian gland dysfunction and diagnosis of aqueous deficient dry eye. Experiments have shown that our method surpasses in accuracy both manual diagnosis and other segmentation methods, giving it the potential to be a new tool in clinical applications. Regrettably, due to the limited amount of data, the morphological priors learned by the proposed method are restricted, resulting in poor performance in segmenting some rare and complex ocular surface structures. This may affect the accuracy in diagnosing some rare cases. In the future, we will collect rare samples to expand the model's applicability. In addition, we plan to advance our work by applying it to more types of eye disease on the ocular surface. We intend to leverage these measured indicators, along with ocular surface images to establish a comprehensive multi-modal framework for diagnosing a wide variety of eye surface diseases.

REFERENCES

- [1] K.-H. Hung, C.-H. Hsiao, H.-Y. Tan, H.-C. Chen, D. H.-K. Ma, H.-C. Lin, and L.-K. Yeh, "Clinical demographics of pterygium excision and prevalence of conjunctival intraepithelial neoplasia: a 15-year review," *International Ophthalmology*, vol. 40, pp. 1781–1788, 2020.
- [2] J. A. Gomes and R. M. Santo, "The impact of dry eye disease treatment on patient satisfaction and quality of life: A review," *The ocular surface*, vol. 17, no. 1, pp. 9–19, 2019.
- [3] W. Xu, L. Jin, P.-Z. Zhu, K. He, W.-H. Yang, and M.-N. Wu, "Implementation and application of an intelligent pterygium diagnosis system based on deep learning," *Frontiers in Psychology*, vol. 12, p. 759229, 2021.
- [4] T. Shahraki, A. Arabi, and S. Feizi, "Pterygium: an update on pathophysiology, clinical features, and management," *Therapeutic Advances in Ophthalmology*, vol. 13, p. 25158414211020152, 2021.
- [5] M. Y. Adil, J. Xiao, J. Olafsson, X. Chen, N. S. Lagali, S. Ræder, Ø. A. Utheim, D. A. Dartt, and T. P. Utheim, "Meibomian gland morphology is a sensitive early indicator of meibomian gland dysfunction," *American journal of ophthalmology*, vol. 200, pp. 16–25, 2019.
- [6] S. McGinnigle, S. A. Naroo, and F. Eperjesi, "Evaluation of dry eye," *Survey of ophthalmology*, vol. 57, no. 4, pp. 293–316, 2012.
- [7] Y. W. Koh, T. Celik, H. K. Lee, A. Petznick, and L. Tong, "Detection of meibomian glands and classification of meibography images," *Journal of biomedical optics*, vol. 17, no. 8, pp. 086008–086008, 2012.
- [8] T. Celik, H. K. Lee, A. Petznick, and L. Tong, "Bioimage informatics approach to automated meibomian gland analysis in infrared images of meibography," *Journal of optometry*, vol. 6, no. 4, pp. 194–204, 2013.
- [9] X. Gao, D. W. K. Wong, A. W. Aryaputera, Y. Sun, C.-Y. Cheng, C. Cheung, and T. Y. Wong, "Automatic pterygium detection on cornea images to enhance computer-aided cortical cataract grading system," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2012, pp. 4434–4437.
- [10] J. Yang, X. Zhu, Y. Liu, X. Jiang, J. Fu, X. Ren, K. Li, W. Qiu, X. Li, and J. Yao, "Tmis: a new image-based software application for the measurement of tear meniscus height," *Acta Ophthalmologica*, vol. 97, no. 7, pp. e973–e980, 2019.

- [11] M. A. K. Setu, J. Horstmann, S. Schmidt, M. E. Stern, and P. Steven, "Deep learning-based automatic meibomian gland segmentation and morphology assessment in infrared meibography," *Scientific reports*, vol. 11, no. 1, p. 7649, 2021.
- [12] D. Liu, W. Zhu, X. Zhuang, X. Zhang, X. Chen, F. Shi, and D. Xiang, "If-net: information fusion network for meibomian gland area and atrophy area segmentation," in *Medical Imaging 2023: Image Processing*, vol. 12464. SPIE, 2023, pp. 904–911.
- [13] S. R. Abdani, M. A. Zulkifley, and N. H. Zulkifley, "Group and shuffle convolutional neural networks with pyramid pooling module for automated pterygium segmentation," *Diagnostics*, vol. 11, no. 6, p. 1104, 2021.
- [14] C. Wan, R. Hua, P. Guo, P. Lin, J. Wang, W. Yang, and X. Hong, "Measurement method of tear meniscus height based on deep learning," *Frontiers in Medicine*, vol. 10, p. 1126754, 2023.
- [15] C. Wan, Y. Shao, C. Wang, J. Jing, and W. Yang, "A novel system for measuring pterygium's progress using deep learning," *Frontiers in Medicine*, vol. 9, p. 819971, 2022.
- [16] X. Deng, L. Tian, Y. Zhang, A. Li, S. Cai, Y. Zhou, and Y. Jie, "Is histogram manipulation always beneficial when trying to improve model performance across devices? experiments using a meibomian gland segmentation model," *Frontiers in Cell and Developmental Biology*, vol. 10, p. 1067914, 2022.
- [17] X. Guo, Z. Wang, H. Hao, Q. Zheng, J. Zhang, W. Chen, and Y. Zhao, "Structure constrained diffusion models for pterygium segmentation," in *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2023, pp. 1–5.
- [18] Z. Ning, S. Zhong, Q. Feng, W. Chen, and Y. Zhang, "Snu-net: Saliency-guided morphology-aware u-net for breast lesion segmentation in ultrasound image," *IEEE transactions on medical imaging*, vol. 41, no. 2, pp. 476–490, 2021.
- [19] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.
- [20] X. Guo, J. Liu, and Y. Yuan, "Semantic-oriented labeled-to-unlabeled distribution translation for image segmentation," *IEEE transactions on medical imaging*, vol. 41, no. 2, pp. 434–445, 2021.
- [21] Y. Fu, J. Chen, J. Li, D. Pan, X. Yue, and Y. Zhu, "Optic disc segmentation by u-net and probability bubble in abnormal fundus images," *Pattern Recognition*, vol. 117, p. 107971, 2021.
- [22] Y. Zheng, Z. Chen, X. Li, X. Si, L. Dong, and Z. Tian, "Deep level set with confidence map and boundary loss for medical image segmentation," in *2020 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2020, pp. 1–6.
- [23] T. Wang, B. Zhang, T. Zhang, S. Gu, J. Bao, T. Baltrusaitis, J. Shen, D. Chen, F. Wen, Q. Chen *et al.*, "Rodin: A generative model for sculpting 3d digital avatars using diffusion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 4563–4573.
- [24] Z. Huang, K. C. Chan, Y. Jiang, and Z. Liu, "Collaborative diffusion for multi-modal face generation and editing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 6080–6090.
- [25] F.-A. Croitoru, V. Hondru, R. T. Ionescu, and M. Shah, "Diffusion models in vision: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.
- [26] T. Amit, E. Nachmani, T. Shaharabany, and L. Wolf, "Segdiff: Image segmentation with diffusion probabilistic models," *arXiv preprint arXiv:2112.00390*, 2021.
- [27] J. Wu, R. FU, H. Fang, Y. Zhang, Y. Yang, H. Xiong, H. Liu, and Y. Xu, "Medsegdiff: Medical image segmentation with diffusion probabilistic model," in *Medical Imaging with Deep Learning*, 2023.
- [28] A. Graikos, N. Malkin, N. Jojic, and D. Samaras, "Diffusion models as plug-and-play priors," *Advances in Neural Information Processing Systems*, vol. 35, pp. 14 715–14 728, 2022.
- [29] A. Rahman, J. M. J. Valanarasu, I. Hacihaliloglu, and V. M. Patel, "Ambiguous medical image segmentation using diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 11 536–11 546.
- [30] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics," in *International conference on machine learning*. PMLR, 2015, pp. 2256–2265.
- [31] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.
- [32] A. Q. Nichol and P. Dhariwal, "Improved denoising diffusion probabilistic models," in *International Conference on Machine Learning*. PMLR, 2021, pp. 8162–8171.
- [33] P. Dhariwal and A. Nichol, "Diffusion models beat gans on image synthesis," *Advances in neural information processing systems*, vol. 34, pp. 8780–8794, 2021.
- [34] T. Chen, C. Wang, and H. Shan, "Berdiff: Conditional bernoulli diffusion model for medical image segmentation," *arXiv preprint arXiv:2304.04429*, 2023.
- [35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [36] M.-H. Guo, C.-Z. Lu, Z.-N. Liu, M.-M. Cheng, and S.-M. Hu, "Visual attention network," *Computational Visual Media*, pp. 1–20, 2023.
- [37] P. A. Yushkevich, Y. Gao, and G. Gerig, "Itk-snap: An interactive tool for semi-automatic segmentation of multi-modality biomedical images," in *2016 38th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE, 2016, pp. 3342–3345.
- [38] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnu-net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature methods*, vol. 18, no. 2, pp. 203–211, 2021.
- [39] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "Transunet: Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv:2102.04306*, 2021.
- [40] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-unet: Unet-like pure transformer for medical image segmentation," in *European conference on computer vision*. Springer, 2022, pp. 205–218.
- [41] Z. Han, M. Jian, and G.-G. Wang, "Convunet: An efficient convolution neural network for medical image segmentation," *Knowledge-Based Systems*, vol. 253, p. 109512, 2022.
- [42] Z. Zhang, F. S. Yin, J. Liu, W. K. Wong, N. M. Tan, B. H. Lee, J. Cheng, and T. Y. Wong, "Origa-light: An online retinal fundus image database for glaucoma analysis and research," in *2010 Annual international conference of the IEEE engineering in medicine and biology*. IEEE, 2010, pp. 3065–3068.
- [43] M. Li, Y. Chen, Z. Ji, K. Xie, S. Yuan, Q. Chen, and S. Li, "Image projection network: 3d to 2d image segmentation in octa images," *IEEE Transactions on Medical Imaging*, vol. 39, no. 11, pp. 3343–3354, 2020.
- [44] H. Fu, J. Cheng, Y. Xu, D. W. K. Wong, J. Liu, and X. Cao, "Joint optic disc and cup segmentation based on multi-label deep network and polar transformation," *IEEE transactions on medical imaging*, vol. 37, no. 7, pp. 1597–1605, 2018.
- [45] M. Li, K. Huang, Q. Xu, J. Yang, Y. Zhang, Z. Ji, K. Xie, S. Yuan, Q. Liu, and Q. Chen, "Octa-500: a retinal dataset for optical coherence tomography angiography study," *Medical Image Analysis*, vol. 93, p. 103092, 2024.
- [46] A. Tomlinson, A. J. Bron, D. R. Korb, S. Amano, J. R. Paugh, E. I. Pearce, R. Yee, N. Yokoi, R. Arita, and M. Dogru, "The international workshop on meibomian gland dysfunction: report of the diagnosis subcommittee," *Investigative ophthalmology & visual science*, vol. 52, no. 4, pp. 2006–2049, 2011.
- [47] J. S. Wolffsohn, R. Arita, R. Chalmers, A. Djalilian, M. Dogru, K. Dumbleton, P. K. Gupta, P. Karpecki, S. Lazreg, H. Pult *et al.*, "Tfos dews ii diagnostic methodology report," *The ocular surface*, vol. 15, no. 3, pp. 539–574, 2017.
- [48] S. Wang, X. He, J. He, S. Li, Y. Chen, C. Xu, X. Lin, J. Kang, W. Li, Z. Luo *et al.*, "A fully automatic estimation of tear meniscus height using artificial intelligence," *Investigative Ophthalmology & Visual Science*, vol. 64, no. 13, pp. 7–7, 2023.
- [49] P. Arriola-Villalobos, J. I. Fernández-Vigo, D. Díaz-Valle, J. E. Peraza-Nieves, C. Fernández-Pérez, and J. M. Benítez-del Castillo, "Assessment of lower tear meniscus measurements obtained with keratograph and agreement with fourier-domain optical-coherence tomography," *British Journal of Ophthalmology*, vol. 99, no. 8, pp. 1120–1125, 2015.

Randomness-restricted diffusion model for ocular surface structure segmentation

Guo, Xinyu

2024

Attribution-NonCommercial 4.0 International

Guo X, Wen H, Hao H, et al., (2024) Randomness-restricted diffusion model for ocular surface structure segmentation. IEEE Transactions on Medical Imaging, Available online 11 November 2024

<https://doi.org/10.1109/tmi.2024.3494762>

Downloaded from CERES Research Repository, Cranfield University



DSPACE

<https://dspace.org/>

Randomness-restricted diffusion model for ocular surface structure segmentation

Guo, Xinyu; Wen, Han; Hao, Huaying; Zhao, Yifan; Meng, Yanda ...

2025-03

IEEE

<https://doi.org/10.1109/tmi.2024.3494762>

Guo X, Wen H, Hao H, et al., (2025) Randomness-restricted diffusion model for ocular surface structure segmentation. IEEE Transactions on Medical Imaging, Volume 44, Issue 3, March 2025, pp. 1359-1372