

## Article

# VF-Mask-Net: A Visual Field Noise Reduction Method Using Neural Networks

Zhenyu Zhang <sup>1</sup> , Haogang Zhu <sup>1,2,3,\*</sup> and Lei Li <sup>4</sup>

<sup>1</sup> State Key Laboratory of Software Development Environment Lab, Beihang University, Beijing 100191, China; zhenyu@buaa.edu.cn

<sup>2</sup> Key Laboratory of Data Science and Intelligent Computing, Zhongfa Aviation Institute, Beihang University, Hangzhou 311115, China

<sup>3</sup> Zhongguancun Laboratory, Beijing 100194, China

<sup>4</sup> School of Biological Science and Medical Engineering, Beihang University, Beijing 100191, China; raylee\_1989@buaa.edu.cn

\* Correspondence: haogangzhu@buaa.edu.cn

**Abstract:** Visual Field (VF) measurements, crucial for diagnosing and treating glaucoma, often contain noise originating from both the instrument and subjects during the response process. This study proposes a neural network-based denoising method for VF data, obviating the need for ground truth labels or paired measurements. Using a mask-imposed VF as an input for the neural network, while the original VF serves as a training label, we evaluated performance metrics such as the accuracy, precision, and sensitivity of denoised VFs. Orthogonal experiments were also employed to assess the impact of mask number, mask structure, and replacement strategy on model accuracy. This study reveals that mask number, replacement strategy, and their interaction significantly affect the accuracy of the denoising model. Under recommended parameters, VF-Mask-Net effectively enhances the accuracy and precision of VF measurements. Furthermore, in deterioration detection tasks, denoised VFs display heightened sensitivity compared to their pre-denoising counterparts.

**Keywords:** glaucoma; visual field; neural network; noise reduction; deterioration detection



**Citation:** Zhang, Z.; Zhu, H.; Li, L. VF-Mask-Net: A Visual Field Noise Reduction Method Using Neural Networks. *Electronics* **2024**, *13*, 646. <https://doi.org/10.3390/electronics13030646>

Academic Editors: Andres Hernandez-Matamoros, Tun-Wen Pai and Hamido Fujita

Received: 28 December 2023

Revised: 30 January 2024

Accepted: 2 February 2024

Published: 4 February 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Glaucoma is one of the most common causes of irreversible vision impairment and blindness [1]. Visual field (VF) measurement using standard automated perimetry (SAP), which assesses differential light sensitivity (DLS) across a subject's field of view, is one of the most important methods in glaucoma diagnosis [2]. As a physical and psychological measure, the measurement process of VF introduces a lot of noise, which makes the measurement results inaccurate [3]. However, a precise measurement of VF is crucial for an accurate diagnosis and corresponding appropriate clinical treatment. In particular, when the VF is used for deterioration detection, the signal change given by the VF measurement needs to be larger than the perturbation of the noise [4].

Various approaches have been adopted to enhance the accuracy of VF measurements. Certain researchers have endeavored to refine the SAP procedure [5–7], while others have sought to augment the application of prior knowledge [8–12]. These endeavors aim to ameliorate the accuracy of VF measurements and mitigate the test–retest variability in the measurement process.

There are also some researchers who use data analysis methods to analyze collected VF data and try to solve problems in the domain of grid-like data noise reduction [13]. Spatial domain methods [14,15] use the correlation between pixels/image patches to denoise the original image. In particular, a spatial filtering method that takes into account the distribution of optic nerve fibers has been shown to reduce VF measurement noise and improve the sensitivity of deterioration detection [16,17].

Transform domain denoising methods [18,19] observe that the distribution of signal and noise is different in transform domain morphology: the signal is usually concentrated in a certain interval of the transform domain, while the noise is usually uniformly distributed in the entire transform domain. Then, the noise reduction of the original data is implemented by designing the filter in the transform domain. On the basis of the above methods, researchers have observed repetition and redundancy in the information in images; thus, they propose methods based on non-local information, non-local means filtering (NLM) [20], and block matching and 3D filtering (BM3D) [21]. By exploiting the similarity between the patches in the image, these methods demonstrate better noise reduction performance.

Now, with the development of neural networks, methods based on machine learning exhibit better performance. Multi-layer perceptron was introduced into the field of image noise reduction in 2012, and the denoising performance can now reach a level similar to that of BM3D, proving the effectiveness of neural network methods in noise reduction tasks [22,23]. By performing normalization for each training mini-batch, batch normalization normalizes the output distribution of each layer of the neural network, facilitating accelerated training and mitigating gradient-related issues [24]. Additionally, the residual neural network incorporates residual connections, enabling the direct transfer of gradients and alleviating the vanishing gradient problem [25]. The combined use of batch normalization and residual neural network techniques makes the training of deep networks feasible. Leveraging these technologies in a deep denoising network results in superior noise reduction performance [26,27].

However, these methods based on machine learning are supervised learning methods, and they need labels to complete the training of the model or require the type and intensity of the noise to be preset. For natural images, ground truth images without noise used as labels are often difficult to obtain. Moreover, in the VF measurement, the noise of DLS at each test point cannot be simply assumed to be a Gaussian distribution [28]. Noise2noise presents a denoising framework that utilizes noisy images as labels for training [29]. It asserts that noise-free images are not essential as labels and provides theoretical evidence that training with noisy images yields results comparable to training with noise-free images. Additionally, noise2void demonstrates that the process of learning itself can also lead to effective noise reduction [30].

When it comes to VF measurement, similar to natural images, the ground truth VFs are also often not available. Thus, supervised approaches cannot simply be used to train a denoising neural network. In test–retest datasets, each subject takes multiple measurements over a short period of time, and their visual function is clinically controlled at a stable level. According to the idea of noise2noise, we can select two VFs in multiple measurements of a subject, one as the input and the other one as the label to train the neural network, that is, VF2VF [31]. However, VF2VF is limited to the test–retest dataset for training; the test–retest dataset is still scarce, and there are still not many measurement data available.

Therefore, in this paper, we propose a training method called VF-Mask-Net using VFs themselves as labels for training. VF-Mask-Net addresses the challenges associated with labels and pairing in denoising networks, significantly broadening the scope of applicable training sets. Building upon this foundation, we systematically analyzed the influence of various mask factors, encompassing mask number, mask structure, replacement strategy, and their interactions, on noise reduction performance. According to the parameters recommended based on our experiments, VF-Mask-Net stands out as the sole method, to the best of our knowledge, capable of concurrently enhancing both accuracy and precision in VF data. Moreover, in comparison to other noise reduction methods, VF-Mask-Net exhibits notable advantages in augmenting sensitivity.

## 2. Materials and Methods

### 2.1. Datasets

All VFs were measured with the HFA (Carl Zeiss Meditec, Jena, Germany) using the 24-2 test pattern and the SITA (Swedish Interactive Thresholding Algorithm) Standard testing algorithm. This test measures DLS at 54 test locations on the retina. For the convenience of processing, grid-like VF data were flattened into structured 54-dimension vectors in an order from top to bottom and from left to right. Specifically, the 26th and 35th dimensions of the vector indicate blind spots.

Two different types of datasets are discussed in this paper: the long-term follow-up dataset and the test–retest dataset. In the test–retest dataset, subjects were clinically held in a stable visual state. Multiple measurements were taken in a short period. Therefore, multiple eye measurements in the test–retest dataset can be considered as multiple observations of a unified visual state. Thus, the variance among VFs in these repeat measurements indicates the inherent measurement variability [32]. Furthermore, the VF series for each eye, and the same series with arbitrary reordering, represent stable series with no underlying change [33]. While in the long-term follow-up dataset, the visual function of subjects may have changed due to aging or deterioration, the measurements in the dataset cannot be considered as multiple observations of the same visual function state.

In this study, two long-term follow-up datasets and one test–retest dataset were used. The Moorfields dataset is a long-term follow-up dataset sampled from electronic health records of glaucoma clinics at Moorfields Eye Hospital in London [34]. In this study, 85,988 VF tests from 9169 eyes were sampled. The UKGTS dataset is another long-term follow-up dataset including VF series from the United Kingdom Glaucoma Treatment Study (UKGTS) [35,36]. This study was a randomized, double-masked placebo-controlled clinical trial testing the hypothesis that treatment with a topical prostaglandin analogue, compared with placebo, reduces the frequency of VF deterioration events. The dataset consists of 9316 VF tests from 735 eyes. The Halifax dataset is a test–retest dataset from a study conducted at Dalhousie University, Halifax Canada [37]. In this study, 30 eyes were tested 12 times over in 3 months. More detailed information about each dataset can be found in Table 1.

**Table 1.** Digital features of each dataset. The main characteristic information of the three datasets used in this study is listed in the table for reference.

Dataset	Number of Eyes	Number of VFs	Mean Times of VF Tests	Mean Time between VF Tests	Mean/Longest Follow-Up Duration	Mean of Pointwise DLS (dB)	Std of Pointwise DLS (dB)	Mean MD (dB)
Moorfields dataset	9169	85,988	9.4	1.19 years	8.61/10.24 years	23.7	9.0	3.0
UKGTS dataset	335	4007	12.0	70.0 days	1.96/2.48 years	25.7	7.0	2.4
Halifax dataset	30	360	12.0	1 week	3/3 months	25.9	7.0	2.0

In our study, the Moorfields dataset was used as the training set to train the neural network. The Halifax dataset was used to evaluate the accuracy and precision of different models. In addition, the Halifax dataset and UKGTS dataset were used to evaluate the sensitivity of deterioration detection before and after denoising.

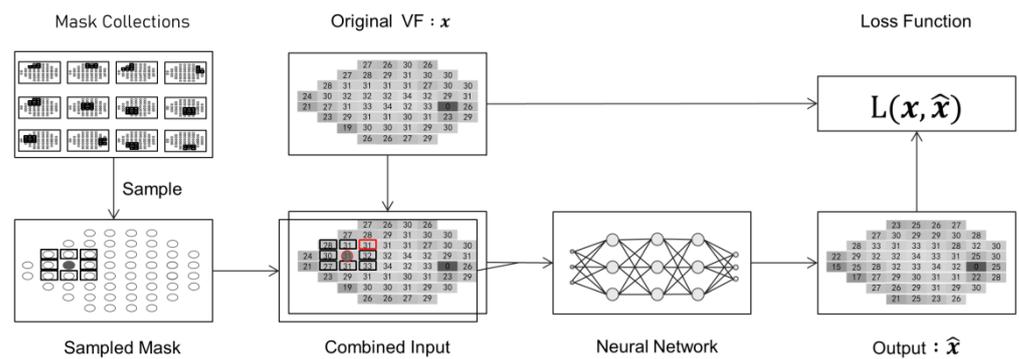
This study adheres to the tenets of the Declaration of Helsinki. Patients' data were anonymized prior to investigation, and this paper does not contain any personal or sensitive information.

### 2.2. VF-Mask-Net

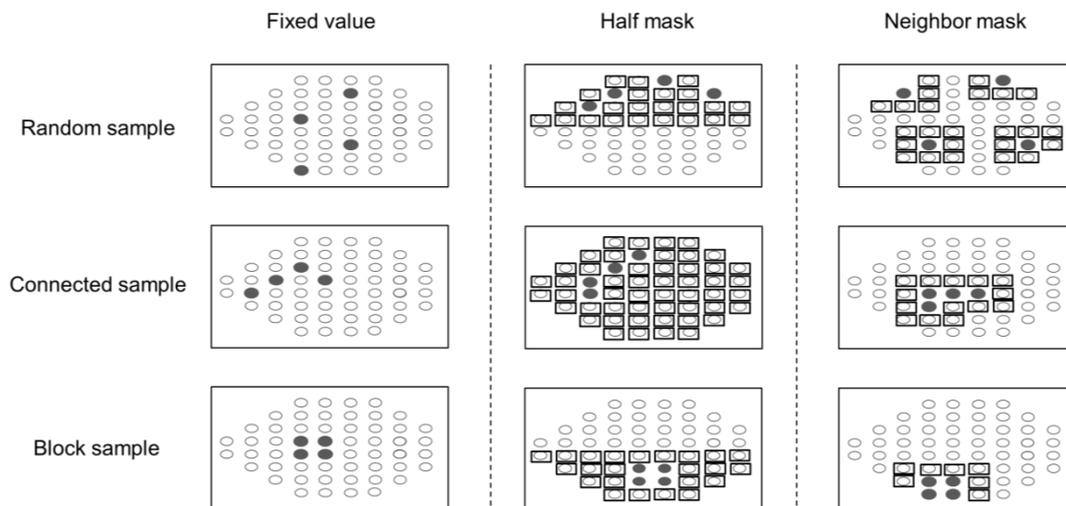
Based on the idea of noise2void [30], VF-Mask-Net is a training method that does not require ground truth as labels and does not require pairing two measurements of a single

eye. Given a single VF without denoising, superimpose a mask on the VF as the input to the neural network, and use the original VF without the mask as the label to train the network.

As depicted in Figure 1, a mask database encompasses masks with various factors. The training set's VF is amalgamated with randomly selected masks from the database to construct the neural network input. Specifically, the DLS value of the pixel at the hollow circle position remains unchanged, while the DLS value of the pixel at the solid circle position (mask pixel) is replaced. The replaced value is sampled from the pixel at the hollow rectangular position. The primary component of the loss function is constituted by the mean squared error between the neural network output,  $\hat{x}$ , and the original VF without denoising,  $x$ .



**Figure 1.** Framework of VF-Mask-Net. Mask collections are the mask set we constructed, which contain masks of different types and different parameters. For more information, refer to Figure 2. During the neural network training process, a mask is randomly selected from the mask collections to superimpose the original VF,  $x$ , and to form the combined input of the neural network. In the example in this figure, the mask pixel of the circular gray shade is overwritten with the DLS value of the surrounding red rectangle. The mean square error between the output of the neural network,  $\hat{x}$ , and of the original VF is used as the loss function.



**Figure 2.** Different mask samples when the mask number takes the value of 4. The black solid circle is the mask pixel that needs to be covered, and the black hollow rectangular is a pool of options available for overriding the mask pixel. When VF is merged with the mask, the value of the solid pixel is overwritten and replaced with a pixel in the pool of options, and the other pixels remain unchanged.

The neural network architecture incorporates a 5-layer fully connected network, encompassing the input layer, three hidden layers, and the output layer. Both the input and output layers consist of 54 neurons, while each hidden layer is composed of 512 neurons.

The activation function for the hidden layers is Rectified Linear Units (ReLUs), and for the output layer, the Hyperbolic Tangent function (Tanh) is employed.

VFs are min–max normalized when inputted into the neural network, limiting the input to the interval  $[-1, 1]$ , as shown in Equation (1).  $x$  indicates the original VF, and  $\mathbf{x}$  indicates the VF after normalization. Then, the normalized VF is superimposed with a random mask and inputted into the neural network. After the forward propagation of the neural network, the model outputs  $\hat{\mathbf{x}}$ . Then, the output is restored to 0 db to 40 db by inverting the normalization as shown in Equation (2).

$$\mathbf{x} = \frac{x - 20}{20} \quad (1)$$

$$\hat{x} = \hat{\mathbf{x}} \times 20 + 20 \quad (2)$$

The loss function of the model is shown in Equation (3). The first term of the loss function is the mean squared error between the output of the neural network and the original VF. The second term is the L2 norm of all multiplicative parameters of the neural network,  $W$ , which is designed to prevent the model from overfitting the training set,  $X$ . By minimizing the loss function, the parameters of the neural network are optimized iteratively to obtain a denoising model. Root mean square propagation (RMSProp) [38] is used to optimize the loss function. The learning rate is suggested to take the value of 0.00005, and the regularization coefficient  $\lambda$  is suggested to take the value of 0.0002.

$$L(X) = E(\|x - \hat{x}\|_2) + \lambda \|W\|_2 \quad (3)$$

### 2.3. Factors of the Mask

The construction of masks is affected by different factors, and the noise reduction effect is also affected. In this study, we focused on factors such as the number of mask pixels (mask number), the shape of the mask (mask structure), the replacement strategy of the DLS value in the mask, and their interactions. To systematically assess and discern the factors exerting the most substantial influence on noise reduction, we developed an orthogonal table of  $L_{27}(3^{13})$  for in-depth analysis, as shown in Table S1 in the Supplementary Materials.

The factor “mask number” denotes the quantity of mask pixels subject to overwriting. As the number of mask pixels increases, the neural network faces a greater challenge in reconstructing more pixels, thereby intensifying the complexity of the noise reduction task. Conversely, if the mask number is excessively small, the neural network is susceptible to gravitating toward an identity map, potentially compromising the effectiveness of noise reduction. To systematically investigate this phenomenon, we categorized the mask number into three levels—specifically, with values of 4, 9, and 16—and conducted experiments accordingly.

The factor “mask structure” refers to the arrangement of selected mask pixels. In VF measurements, the correlation between each measurement pixel and its neighbors is notably higher than the correlation with non-neighbors. Elevated connectivity between overwritten pixels can erase potentially valuable neighbor information during the DLS overwriting process, thereby amplifying the challenge of VF reconstruction by the neural network. To systematically explore this impact, we categorized the mask structure into three levels—random, connected, and block—and conducted experiments accordingly. In this study, “random” implies that mask pixels are randomly sampled from the 54 dimensions of VF without repetition; “connected” signifies that any two mask pixels are connected; and “block” indicates that mask pixels exhibit higher connectivity. The three rows in Figure 2 illustrate several samples under the three levels of mask structure.

The factor “replacement strategy” denotes the approach employed to replace the original value when overwriting the mask pixels. In this study, we investigated three distinct strategies. The first strategy involves a straightforward replacement of the original value with a fixed value (20 dB). The second strategy takes into account the limited correlation between pixels in the upper and lower halves of the VF. Here, we extract the

DLS value from the corresponding half of the mask pixel to execute the replacement. The third strategy entails sampling DLS values from adjacent pixels of the mask pixel. The three columns in Figure 2 showcase various samples illustrating the three conditions of replacement strategies.

#### 2.4. Assessment of VF Noise Reduction Performance

We evaluated the noise reduction effect of neural network from three aspects, accuracy, precision and sensitivity. The test–retest dataset Halifax dataset, which is independent of the training set, was used for accuracy and precision evaluation. Due to the stability of VF measurements in the short term, the mean value of each eye in the Halifax dataset is taken as an estimate of the ground truth measurement of such eye.

Accuracy reflects the unbiasedness of the denoising model, which is the most critical aspect of the denoising model. In this paper, the mean deviation (MD) is used as the evaluation index for accuracy, as shown in Equation (4).  $E(x)$  represents the ground truth VF vector of the eye, and  $\|\bullet\|_2$  represents the L2 norm of the vector. MD captures the mean deviation between the denoised VF and the ground truth VF. A smaller deviation signifies that the denoised VF is in closer proximity to the ground truth, indicating higher accuracy and superior denoising performance.

$$MD = \|x - E(x)\|_2 \quad (4)$$

Precision gauges the stability of the model’s output. It assesses how the model output responds to small perturbations in the input—essentially measuring the magnitude of change volatility. In this study, precision is evaluated by considering the variance among multiple measurements of one eye in the test–retest dataset, which reflects the volatility of the input VFs. The corresponding output’s variance is then indicative of the model’s precision. Consequently, the standard deviation among multiple measurements of one eye after noise reduction is defined as the precision index, denoted as PD, as outlined in Equation (5), where  $\bar{x}$  represents the sample mean of multiple measurements of the eye.

$$PD = \|x - \bar{x}\|_2 \quad (5)$$

Sensitivity constitutes another crucial aspect that mirrors the performance of VF noise reduction. In the course of noise reduction, subtle yet objectively present signals may be diminished alongside the noise, potentially compromising the sensitivity of VF in tasks related to deterioration detection. To assess sensitivity before and after noise reduction, we employed both a straightforward distance-based deterioration detection method and a widely used approach, Permutation Analyses of Pointwise Linear Regression (PoPLR) [39].

Specifically, in the distance-based methods, we assess the Euclidean distance between the initial VF measurement and subsequent measurements. If the distance surpasses a predefined threshold at a given specificity, it is inferred that deterioration has occurred; otherwise, no deterioration is presumed. As for the PoPLR method, we rely on the deterioration detection index of PoPLR, denoted as  $S_{PoPLR}$  [39]. A VF series is classified as deteriorated when its  $S_{PoPLR}$  exceeds the threshold established in the “no change” dataset.

We conducted a sensitivity comparison before and after denoising by constructing ROC-like curves for the distance-based methods and PoPLR on both the Halifax dataset and the UKGTS dataset. In the Halifax dataset, which served as a “no change” dataset, the deterioration detection index establishes thresholds for determining whether the VF series have deteriorated. The “hit rate” calculated on the UKGTS dataset based on these thresholds has been shown to be positively correlated with the true positive rate [28]. Consequently, the ROC-like curve evaluates sensitivity by plotting the hit rate against the false positive rate. Furthermore, in clinical practice, deterioration detection is only meaningful at lower false positive rate (FPR). Therefore, we specifically analyzed deterioration detection results within the FPR range of 0% to 15%, utilizing the area under the corresponding curve as a quantitative sensitivity index—referred to as the partial area under the curve (partial AUC).

Furthermore, we introduced the concept of warning time for a series, defined as the series number corresponding to a given FPR. In detail, we established the threshold for deterioration detection under the specified FPR and subsequently assessed whether the series has deteriorated based on this threshold. The series number, denoting when deterioration is first detected, was designated as the warning time. This metric offers an alternative perspective on the sensitivity difference before and after noise reduction. A smaller warning time under a given FPR indicates heightened sensitivity in the VF data.

### 3. Results

All experiments were performed on Dell C4130. The operating system was Ubuntu 22.04 LTS. The neural network framework was built using Python 3.10.9 and PyTorch 1.12.0.

#### 3.1. Improved Accuracy and Precision of Transformed VFs

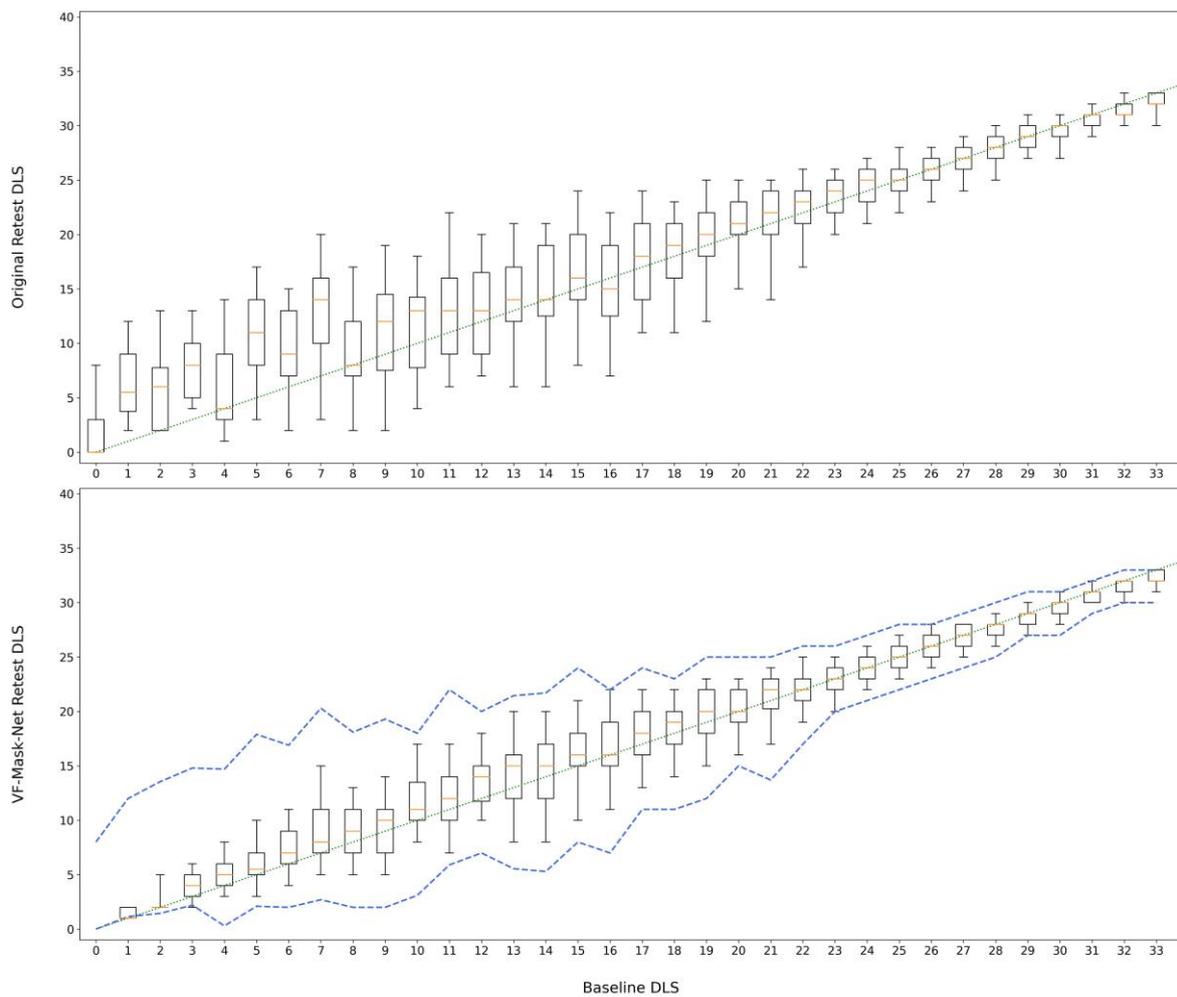
We performed eight experiments under specific conditions: mask number set at 2, mask structure designated as “random sample”, and the replacement strategy specified as “half mask”. The results of these experiments indicate that, on average, the MD of the VF after noise reduction is reduced by 0.34 dB compared to the pre-denoising state. This reduction signifies a closer alignment of the denoised VF with the ground truth compared to the VF before noise reduction. Additionally, the post-denoising precision index PD is, on average, 0.76 dB lower than its pre-denoising counterpart, indicating increased stability in the output of the noise reduction network.

Another indicator of enhanced precision is evident in the boxplots of the retest DLS, as shown in the Figure 3. In VF measurement, when the DLS value of a test point in the VF is denoted as  $d$ , the retest DLS of that test point,  $r$ , conforms to a mixed Weibull distribution, as validated in ANSWERS [28]. The boxplot effectively illustrates the distribution of the retest DLS. Comparing the boxplot after denoising with those before denoising, it is apparent that the denoising model is efficacious in refining precision, as the boxplot after denoising is encapsulated within the boundaries of the boxplots before denoising. This observation substantiates the effectiveness of the denoising model in improving precision.

We compared the accuracy and precision of VF-Mask-Net with those of the traditional noise reduction method spatial filter [16,17], neural network methods such as DnCNN [26] and VF2VF [31], and the reconstruction method variational autoencoder (VAE) [40]. As delineated in Table 2, VF-Mask-Net stands out as the sole method that effectively reduces both MD and PD simultaneously. The spatial filter tends to maintain consistency with the original values, resulting in marginal changes in MD and PD after denoising. DnCNN, VF2VF, and VAE reduce PD but exhibit an increase in MD. Given that MD, representing unbiasedness, holds a higher priority than PD, indicative of output stability, VF-Mask-Net emerges as superior to the other methods when considering the combined factors of accuracy and precision in noise reduction.

**Table 2.** MD and PD of different denoising methods.

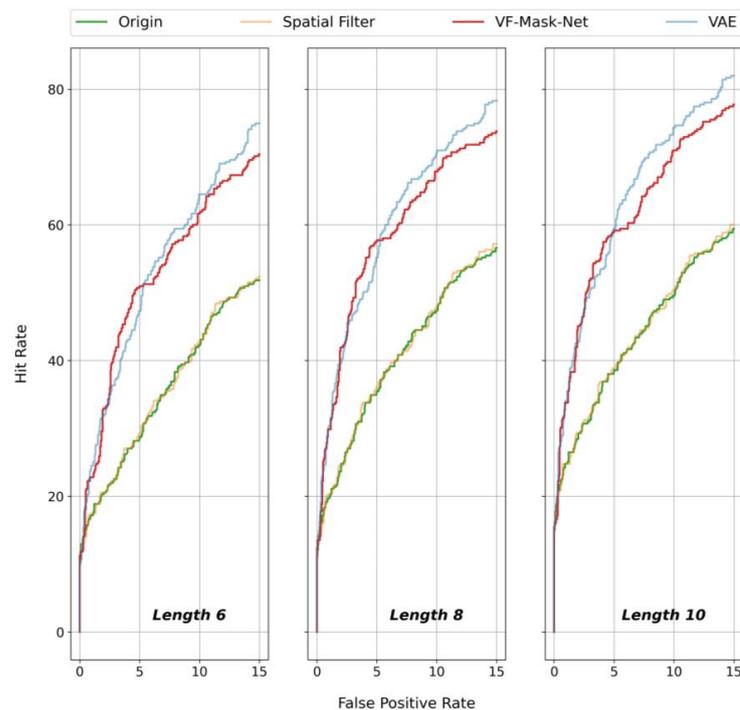
Methods	Original VF	Spatial Filter	DnCNN	VF2VF	VAE	VF-Mask-Net
MD (dB)	2.31	2.36	2.49	2.56	2.54	<b>1.97</b>
PD (dB)	2.12	2.14	0.99	<b>0.79</b>	1.04	1.36



**Figure 3.** Boxplots of retest DLS. The subplot above is the boxplot of retest DLS before denoising, and the subplot below is the one after denoising. The boxes in the boxplot delineate the distribution of the retest DLS from five numerical features, the 5th quantile, the lower quartile, the median, the upper quartile, and the 95th quantile. The green dotted line is the diagonal of the quadrant and serves to facilitate an observation of the deviation between the bin median and the ideal value. The blue dashed line in the subplot below shows the fit to the 5% and 95% quantiles before denoising.

### 3.2. Higher Sensitivity of Deterioration Detection

As depicted in Figure 4, the distance-based deterioration detection method illustrates a notable enhancement in the hit rate after noise reduction using VF-Mask-Net compared to the pre-denoising state. The hit rate curve of the spatial filter remains largely unchanged before and after noise reduction. The hit rate of VAE closely aligns with that of VF-Mask-Net. Specifically, as outlined in Table 3, the partial AUC of VF-Mask-Net increases by 0.17, 0.18, and 0.17, respectively, compared to the original value for VF series lengths of 6, 8, and 10. Furthermore, at FPRs of 5%, 10%, and 15%, the hit rate after VF-Mask-Net noise reduction is elevated by 21.3%, 20.0%, and 17.3%, respectively, compared to the original values.



**Figure 4.** ROC-like curves of distance-based deterioration detection methods. The *x*-axis represents the false positive rate, and the false positive rate limit is between 0% and 15%. The *y*-axis represents the observed hit rate in the long-term follow-up dataset. The three sub-figures, from left to right, show the relationship between hit rate and false positive rate when the sequence length is 6, 8, and 10, respectively.

**Table 3.** Partial AUCs of different denoising methods.

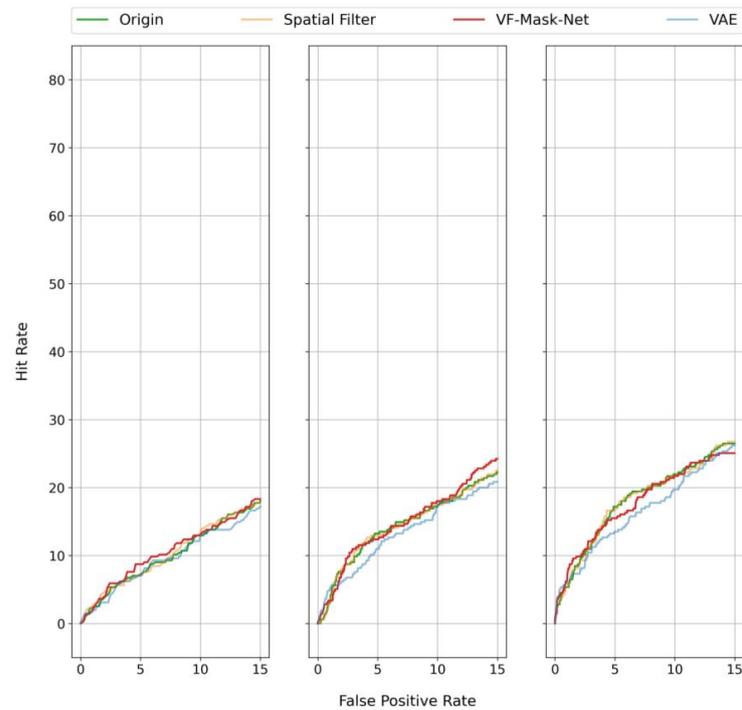
Methods	Distanced-Based Methods			PoPLR		
	6	8	10	6	8	10
Original VF	0.3539	0.4038	0.4336	0.0996	0.1428	<b>0.1800</b>
Spatial filter	0.3551	0.4056	0.4357	0.1018	0.1427	0.1793
DnCNN	0.4718	0.5336	0.5665	0.0887	0.1365	0.1664
VF2VF	0.5131	0.5744	0.6090	0.0920	0.1218	0.1557
VAE	<b>0.5382</b>	<b>0.6059</b>	<b>0.6468</b>	0.0957	0.1316	0.1632
VF-Mask-Net	0.5229	0.5831	0.6107	<b>0.1049</b>	<b>0.1473</b>	0.1771

Furthermore, Table 3 presents the partial AUC results for other noise reduction methods. It is evident that the partial AUC of VF-Mask-Net surpasses that of the conventional spatial filter and the neural network approaches DnCNN and VF2VF. Notably, VF-Mask-Net ranks second, only being bested by the optimal outcome achieved by VAE under the current deterioration detection method.

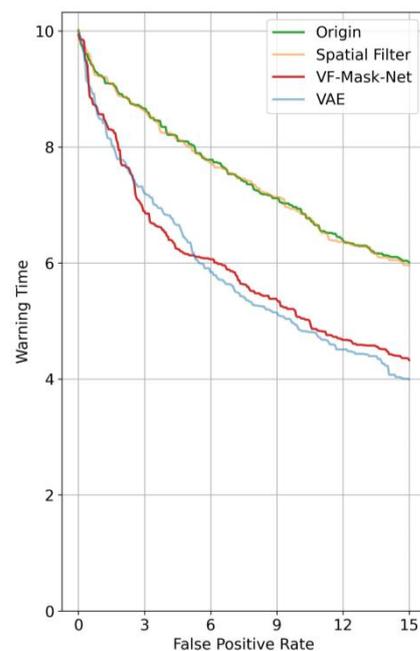
The outcomes of the PoPLR, another method for deterioration detection, are more tightly clustered, as depicted in Figure 5. The curves of the original value, VF-Mask-Net, and spatial filter essentially overlap and slightly exceed the results obtained with VAE. Specifically, as indicated in Table 3, the partial AUC of VF-Mask-Net is optimal for series lengths of 6 and 8 and is comparable to the optimal value for a series length of 10.

Another index indicating sensitivity, warning time, was also subjected to comparisons. As illustrated in Figure 6, when the FPR ranges from 0% to 15%, the warning time curve of VF-Mask-Net remains consistently stable and lower than the warning time curve before noise reduction. This signifies that in clinical practice, at the same error probability, VF data denoised by VF-Mask-Net consistently provide an early warning 1.73 times ahead of the

original data. Comparatively, spatial filter leads by 0.02 times, DnCNN leads by 1.35 times, VF2VF leads by 1.77 times, and VAE leads by 2.03 times. VF-Mask-Net ranks second, only being bested by the optimal outcome, VAE.



**Figure 5.** ROC-like curves of PoPLR. The  $x$ -axis represents the false positive rate, and the false positive rate limit is between 0% and 15%. The  $y$ -axis represents the observed hit rate in the long-term follow-up dataset. The three sub-figures, from left to right, show the relationship between hit rate and false positive rate when the sequence length is 6, 8, and 10, respectively.

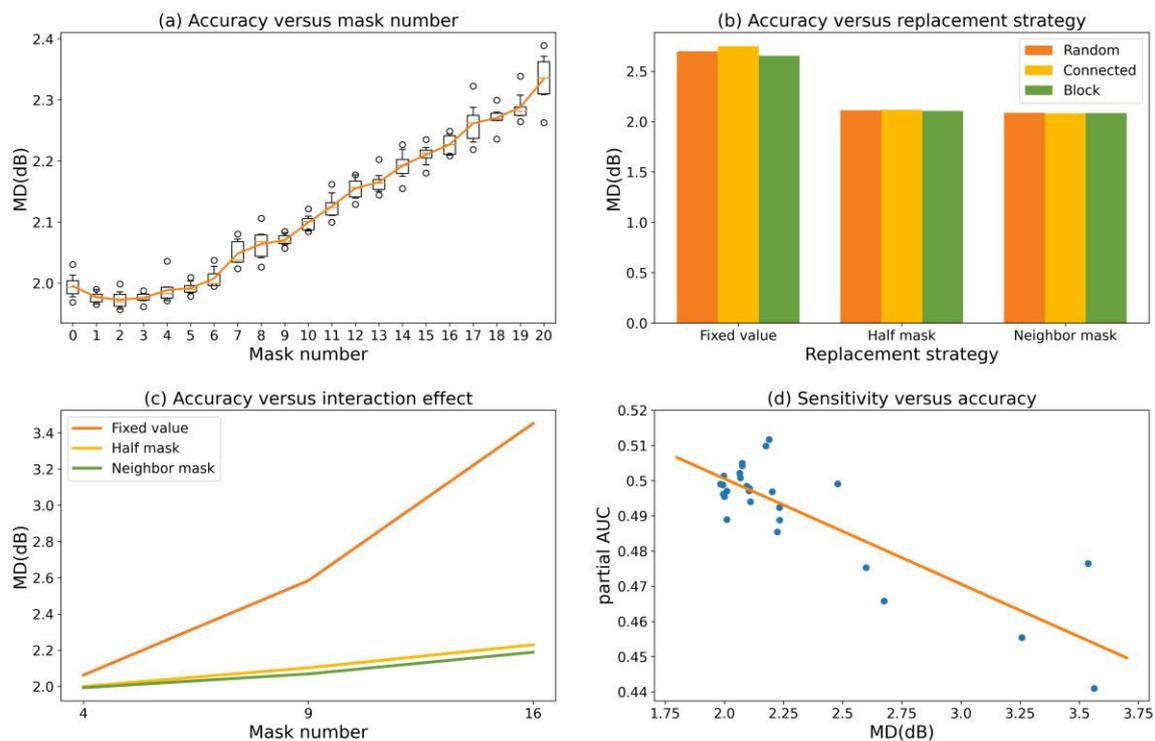


**Figure 6.** Comparison of warning time. The  $x$ -axis represents the false positive rate, and the false positive rate limit is between 0% and 15%. The  $y$ -axis, warning time, represents the follow-up serial number where the abnormality first found is in the series of length 10.

### 3.3. Impact of Different Mask Factors on Accuracy and Sensitivity

We conducted eight repetitions for each experimental point specified in the orthogonal experimental table (refer to Supplementary Materials Table S1) and subsequently carried out variance analysis based on the experimental outcomes. According to the results, among the three primary factors—mask number, mask structure, and replacement strategy—along with the interaction factor formed by pairing each of the three main factors, it was observed that mask number, replacement strategy, and the combined effect of mask number and replacement strategy significantly impacted the accuracy of the model ( $p < 0.01$ ). Conversely, the remaining factors demonstrated no significant impact. A detailed analysis of these findings is provided in this section.

To scrutinize the impact of mask number on accuracy, we conducted 21 sets of experiments with the mask structure designated as “random” and the replacement strategy as “half mask”, systematically varying the mask number from 0 to 20. Each experiment was replicated eight times, and the outcomes are graphically represented in Figure 7a. It is evident that as the mask number ascends from 0 to 2, the MD progressively diminishes, signifying an improvement in the accuracy of noise reduction. Conversely, as the mask number increases from 2 to 20, MD gradually elevates, indicating a diminishing noise reduction accuracy. The denoising network attains optimal accuracy when the mask number is set at 2.



**Figure 7.** The impact of different mask factors on accuracy and the relationship between accuracy and sensitivity. (a) The impact of mask number on accuracy. The boxplot displays the 5% quantile, quarter quantile, median, third-quarter quantile, 95% quantile of the test point. The solid orange line fits the mean value of the results at each test point. (b) The impact of replacement strategy on accuracy. (c) The impact of the interaction effect of mask number and replacement strategy on accuracy. (d) The fitting results of partial AUC and MD based on orthogonal test results. The blue dots describe the corresponding results of the partial AUC and MD of the orthogonal experiments, and the orange line represents the regression curve fitted based on these results.

The choice of replacement strategy also exerts an impact on the accuracy of the denoising network. As per the experimental outcomes derived from the orthogonal test, among the three levels of replacement strategy, there is no significant difference between

“half mask” and “neighbor mask”, while “fixed value” is significantly different from the other two ( $p < 0.01$ ). As illustrated in Figure 7b, the MD under the “fixed value” condition is markedly higher than the MD under the “half mask” and “neighbor mask” conditions, indicating that utilizing “half mask” or “neighbor mask” as the replacement strategy yields higher accuracy. Furthermore, Figure 7b shows that the accuracy differences resulting from the three levels of mask structure are relatively minor.

The interaction between mask number and replacement strategy also plays a role in influencing the accuracy of the denoising network, as depicted in Figure 7c. Firstly, it is apparent that the MD after noise reduction increases with the escalating mask number. Secondly, concerning the replacement strategy, the MD for “fixed value” surpasses that of “half mask” and “neighbor mask”. These observations align with the findings from the individual factor analysis mentioned earlier. The interaction effect between mask number and replacement strategy manifests in the varying change rates of MD for different replacement strategies as the mask number increases. With increasing mask number, the MD for “fixed value” experiences a more rapid ascent compared to the MD for “half mask” and “neighbor mask”, signifying a faster decrease in accuracy.

Furthermore, we conducted an analysis of the correlation between sensitivity and accuracy based on the results of the orthogonal experiments. As illustrated in Figure 7d, as the MD steadily increases, the accuracy of the denoising network progressively diminishes, and concurrently, the sensitivity of the denoising network also experiences a decline. MD, serving as a representation of accuracy, and the partial AUC, indicative of sensitivity, exhibit a significant correlation, with a Pearson correlation coefficient of  $-0.839$  ( $p < 0.01$ ).

In summary, for training the denoising network, our recommendation is to set the mask number at 2 and utilize “half mask” or “neighbor mask” as the replacement strategy. These configurations lead to favorable outcomes in terms of accuracy and sensitivity, as supported by the results of our experiments and analyses.

#### 4. Discussion

The issue of noise reduction in VF data has been thoroughly studied. In this paper, operating without labeled or paired data, we introduce a neural network training method, VF-Mask-Net, designed to enhance the accuracy and precision of VFs. In the subsequent deterioration detection task, our experiments empirically validate that the denoised VFs exhibit superior sensitivity compared to the original ones. Additionally, we conducted an orthogonal experiment to dissect the influence of various mask factors on noise reduction performance. Through this analysis, we identified the pivotal factors affecting noise reduction performance and provided recommended parameters for these factors.

Compared to various denoising methods, VF-Mask-Net exhibits notable advantages. Firstly, as a neural network method, VF-Mask-Net significantly broadens the applicability of noise reduction techniques based on neural networks. Commonly used neural network approaches like DnCNN [26] typically necessitate labeled data for training. VF2VF [31] relies on multiple measurements of a subject for paired training, whereas VF-Mask-Net operates without the need for labels or pairing. This training method significantly lowers the entry threshold for the training set, expanding the range of applicable datasets.

Accuracy stands as the paramount index reflecting the efficacy of noise reduction. According to our experiments, VF-Mask-Net is the only method that effectively reduces accuracy. Though traditional noise reduction methods like spatial filters do not require labels and pairing, they do not manifest a prominent noise reduction effect on VFs. Generating models such as VAE also does not demand labels and pairing, yet the accuracy of the reconstructed VF is reduced, and ensuring the unbiasedness of the VF becomes challenging. The stringent conditions within the training set significantly constrain the volume of available training data, preventing DnCNN and VF2VF from achieving higher accuracy.

The sensitivity of deterioration detection serves as another crucial metric for evaluating noise reduction performance. In this study, we assessed the sensitivity of VF data in detecting deterioration through two commonly employed methods: distance-based methods

and the prevalent trend-based method, PoPLR. In the indicators defined by distance-based methods, VF-Mask-Net exhibits a significant improvement in sensitivity compared to the original data. Relative to other methods, VF-Mask-Net secures the second position and closely approaches the optimal performance of VAE. Regarding the indicators defined by PoPLR, the differences between various denoising methods diminish, yet VF-Mask-Net still maintains a slight advantage over other denoising methods.

Ideally, high accuracy signifies a superior signal-to-noise ratio in the data, implying that the data can more precisely capture changes in the visual situation, thereby enhancing detection sensitivity. The relationship between partial AUC and MD, depicted in Figure 7d, corroborates this hypothesis. Nevertheless, it is important to note that the sensitivity of deterioration detection is also contingent on the specific method employed for such detection. For instance, while the MD of VAE may be higher than that of VF-Mask-Net, it exhibits the most sensitive deterioration detection performance among distance-based methods. Conversely, in the PoPLR method, the sensitivity of VAE is lower compared to VF-Mask-Net.

The selection of the mask also plays a crucial role in the performance of noise reduction. When the mask overwrites an excessive amount of information, it becomes more challenging for the model to reconstruct the overwritten data, potentially leading to a decrease in noise reduction performance. For instance, in the experiment examining the correlation between accuracy and mask number, as the mask number exceeded 2, the MD gradually increased. Further evidence was found in the experiment analyzing the correlation between accuracy and replacement strategy, where the MD of “fixed value”—which overwrites more information—was higher compared to that of “half mask” or “neighbor mask”, which overwrite less information. However, it is important to note that the relationship is not strictly linear, and the “more information overwritten, the better” is not a universal rule. If the mask overwrites too little information, the disparity between the model’s input and output becomes too minimal, potentially causing the model to collapse into an identity map. For example, in the experiment investigating the correlation between accuracy and mask number, the accuracy was higher when the mask number was 2 compared to when it was 1 or 0. This underscores the need for a balanced approach in determining the extent of information overwritten by the mask for optimal noise reduction performance.

When faced with VF measurement data, VF-Mask-Net provides a solution to simultaneously improve accuracy and sensitivity for VFs. VF data, characterized as a grid-like dataset, possess unique attributes. In contrast to natural images, VF data have lower dimensions, and each pixel retains fixed position information, imposing more significant constraints than natural images. This fixed position information facilitates pixel-wise quantitative analysis. Medical images share similarities with VF data, exhibiting relatively fixed structures and simple, clear semantics. These shared features create favorable conditions for the utilization of VF-Mask-Net, and we foresee its meaningful application and advancement in related fields.

In summary, VF-Mask-Net presents an unsupervised denoising method leveraging neural networks. This model enhances the accuracy and precision of VFs, resulting in denoised VFs that exhibit increased sensitivity in deterioration detection.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/electronics13030646/s1>, Table S1: Orthogonal experimental table of mask factors.

**Author Contributions:** Conceptualization, Z.Z. and H.Z.; methodology, Z.Z. and H.Z.; software, Z.Z. and L.L.; validation, Z.Z. and L.L.; formal analysis, Z.Z. and H.Z.; investigation, Z.Z. and H.Z.; resources, H.Z.; data curation, Z.Z. and H.Z.; writing—original draft preparation, Z.Z.; writing—review and editing, H.Z. and L.L.; visualization, Z.Z.; supervision, H.Z.; project administration, H.Z.; funding acquisition, H.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Major Project of Science and Technology Innovation 2030—New Generation Artificial Intelligence (No. 2021ZD0140407), the National Natural Science Foundation of China (No. U21A20523).

**Data Availability Statement:** The data that support the findings of this study are available upon request from the corresponding author.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Flaxman, S.R.; Bourne, R.R.; Resnikoff, S.; Ackland, P.; Braithwaite, T.; Cicinelli, M.V.; Das, A.; Jonas, J.B.; Keeffe, J.; Kempen, J.H. Global causes of blindness and distance vision impairment 1990–2020: A systematic review and meta-analysis. *Lancet Glob. Health* **2017**, *5*, e1221–e1234. [[CrossRef](#)] [[PubMed](#)]
2. Henson, D.B. *Visual Fields*; Butterworth-Heinemann Medical: Oxford, UK, 2000.
3. Sharma, V.; Shen, L.Q.; Pasquale, L.; Elze, T.; Boland, M.V.; Wellik, S.R.; De Moraes, G.; Myers, J.S.; Yousefi, S.; Wang, M. A Deep Autoencoder Model to Denoise Visual Fields in Glaucoma. In Proceedings of the Investigative Ophthalmology & Visual Science, Denver, Colorado, 1–4 May 2022.
4. Spry, P.G.; Johnson, C.A. Identification of progressive glaucomatous visual field loss. *Surv. Ophthalmol.* **2002**, *47*, 158–173. [[CrossRef](#)] [[PubMed](#)]
5. Schiefer, U.; Pascual, J.P.; Edmunds, B.; Feudner, E.; Hoffmann, E.M.; Johnson, C.A.; Lagreze, W.A.; Pfeiffer, N.; Sample, P.A.; Staubach, F. Comparison of the new perimetric GATE strategy with conventional full-threshold and SITA standard strategies. *Investig. Ophthalmol. Vis. Sci.* **2009**, *50*, 488–494. [[CrossRef](#)] [[PubMed](#)]
6. Bengtsson, B.; Heijl, A. Evaluation of a new perimetric threshold strategy, SITA, in patients with manifest and suspect glaucoma. *Acta Ophthalmol. Scand.* **1998**, *76*, 268–272. [[CrossRef](#)] [[PubMed](#)]
7. Bengtsson, B.; Olsson, J.; Heijl, A.; Rootzén, H. A new generation of algorithms for computerized threshold perimetry, SITA. *Acta Ophthalmol. Scand.* **1997**, *75*, 368–375. [[CrossRef](#)] [[PubMed](#)]
8. Zhu, H.; Yang, H.; Crabb, D.; Miranda, M.; Garway-Heath, D.F. Improved precision and accuracy with trail traced threshold test (T4). *Investig. Ophthalmol. Vis. Sci.* **2018**, *59*, 5114.
9. Turpin, A.; Jankovic, D.; McKendrick, A.M. Retesting visual fields: Utilizing prior information to decrease test–retest variability in glaucoma. *Investig. Ophthalmol. Vis. Sci.* **2007**, *48*, 1627–1634. [[CrossRef](#)] [[PubMed](#)]
10. Rubinstein, N.J.; McKendrick, A.M.; Turpin, A. Incorporating spatial models in visual field test procedures. *Transl. Vis. Sci. Technol.* **2016**, *5*, 7. [[CrossRef](#)]
11. McKendrick, A.M.; Turpin, A. Combining perimetric suprathreshold and threshold procedures to reduce measurement variability in areas of visual field loss. *Optom. Vis. Sci.* **2005**, *82*, 43–51.
12. Montesano, G.; Rossetti, L.M.; Allegrini, D.; Romano, M.R.; Crabb, D.P. Improving visual field examination of the macula using structural information. *Transl. Vis. Sci. Technol.* **2018**, *7*, 36. [[CrossRef](#)]
13. Fan, L.; Zhang, F.; Fan, H.; Zhang, C. Brief review of image denoising techniques. *Vis. Comput. Ind. Biomed. Art* **2019**, *2*, 1–12. [[CrossRef](#)]
14. Yang, R.; Yin, L.; Gabbouj, M.; Astola, J.; Neuvo, Y. Optimal weighted median filtering under structural constraints. *IEEE Trans. Signal Process.* **1995**, *43*, 591–604. [[CrossRef](#)]
15. Benesty, J.; Chen, J.; Huang, Y. Study of the widely linear Wiener filter for noise reduction. In Proceedings of the 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, Dallas, TX, USA, 14–19 March 2010; pp. 205–208.
16. Strouthidis, N.G.; Scott, A.; Viswanathan, A.C.; Crabb, D.P.; Garway-Heath, D.F. Monitoring glaucomatous visual field progression: The effect of a novel spatial filter. *Investig. Ophthalmol. Vis. Sci.* **2007**, *48*, 251–257. [[CrossRef](#)]
17. Gardiner, S.K.; Crabb, D.P.; Fitzke, F.W.; Hitchings, R.A. Reducing noise in suspected glaucomatous visual fields by using a new spatial filter. *Vis. Res.* **2004**, *44*, 839–848. [[CrossRef](#)]
18. Jain, P.; Tyagi, V. Spatial and frequency domain filters for restoration of noisy images. *IETE J. Educ.* **2013**, *54*, 108–116. [[CrossRef](#)]
19. Cai, J.-F.; Candès, E.J.; Shen, Z. A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.* **2010**, *20*, 1956–1982. [[CrossRef](#)]
20. Buades, A.; Coll, B.; Morel, J.-M. A non-local algorithm for image denoising. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05), San Diego, CA, USA, 20–26 June 2005; pp. 60–65.
21. Dabov, K.; Foi, A.; Katkovnik, V.; Egiazarian, K. Image restoration by sparse 3D transform-domain collaborative filtering. In Proceedings of the Image Processing: Algorithms and Systems VI, San Jose, CA, USA, 28–29 January 2008; pp. 62–73.
22. Xie, J.; Xu, L.; Chen, E. Image denoising and inpainting with deep neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 341–349.
23. Burger, H.C.; Schuler, C.J.; Harmeling, S. Image denoising: Can plain neural networks compete with BM3D? In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Rhode Island, CA, USA, 16–21 June 2012; pp. 2392–2399.
24. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 448–456.

25. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
26. Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; Zhang, L. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Trans. Image Process.* **2017**, *26*, 3142–3155. [[CrossRef](#)]
27. Zhang, K.; Zuo, W.; Zhang, L. FFDNet: Toward a fast and flexible solution for CNN-based image denoising. *IEEE Trans. Image Process.* **2018**, *27*, 4608–4622. [[CrossRef](#)]
28. Zhu, H.; Crabb, D.P.; Ho, T.; Garway-Heath, D.F. More accurate modeling of visual field progression in glaucoma: ANSWERS. *Investig. Ophthalmol. Vis. Sci.* **2015**, *56*, 6077–6083. [[CrossRef](#)]
29. Lehtinen, J.; Munkberg, J.; Hasselgren, J.; Laine, S.; Karras, T.; Aittala, M.; Aila, T. Noise2Noise: Learning image restoration without clean data. *arXiv* **2018**, arXiv:1803.04189.
30. Krull, A.; Buchholz, T.-O.; Jug, F. Noise2void-learning denoising from single noisy images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 2129–2137.
31. Zhang, Z.; Chen, X.; Zhu, H. VF2VF: Improving Precision while Maintaining Accuracy. In Proceedings of the 2022 IEEE International Conference on Big Data (Big Data), Osaka, Japan, 17–20 December 2022; pp. 5388–5394.
32. Artes, P.H.; Iwase, A.; Ohno, Y.; Kitazawa, Y.; Chauhan, B.C. Properties of perimetric threshold estimates from Full Threshold, SITA Standard, and SITA Fast strategies. *Investig. Ophthalmol. Vis. Sci.* **2002**, *43*, 2654–2659.
33. Patterson, A.J.; Garway-Heath, D.F.; Strouthidis, N.G.; Crabb, D.P. A new statistical approach for quantifying change in series of retinal and optic nerve head topography images. *Investig. Ophthalmol. Vis. Sci.* **2005**, *46*, 1659–1667. [[CrossRef](#)]
34. Zhu, H.; Russell, R.A.; Saunders, L.J.; Ceccon, S.; Garway-Heath, D.F.; Crabb, D.P. Detecting changes in retinal function: Analysis with non-stationary Weibull error regression and spatial enhancement (ANSWERS). *PLoS ONE* **2014**, *9*, e85654. [[CrossRef](#)]
35. Garway-Heath, D.F.; Crabb, D.P.; Bunce, C.; Lascaratos, G.; Amalfitano, F.; Anand, N.; Azuara-Blanco, A.; Bourne, R.R.; Broadway, D.C.; Cunliffe, I.A. Latanoprost for open-angle glaucoma (UKGTS): A randomised, multicentre, placebo-controlled trial. *Lancet* **2015**, *385*, 1295–1304. [[CrossRef](#)]
36. Garway-Heath, D.F.; Lascaratos, G.; Bunce, C.; Crabb, D.P.; Russell, R.A.; Shah, A.; United Kingdom Glaucoma Treatment Study Investigators. The United Kingdom Glaucoma Treatment Study: A multicenter, randomized, placebo-controlled clinical trial: Design and methodology. *Ophthalmology* **2013**, *120*, 68–76. [[CrossRef](#)]
37. Garway-Heath, D.F.; Zhu, H.; Cheng, Q.; Morgan, K.; Frost, C.; Crabb, D.P.; Ho, T.-A.; Agiomyrgiannakis, Y. Combining optical coherence tomography with visual field data to rapidly detect disease progression in glaucoma: A diagnostic accuracy study. *Health Technol. Assess.* **2018**, *22*, 1. [[CrossRef](#)]
38. Hinton, G.; Srivastava, N.; Swersky, K. Neural networks for machine learning lecture 6a overview of mini-batch gradient descent. *Neural Netw. Mach. Learn.* **2012**, *14*, 2.
39. O’Leary, N.; Chauhan, B.C.; Artes, P.H. Visual field progression in glaucoma: Estimating the overall significance of deterioration with permutation analyses of pointwise linear regression (PoPLR). *Investig. Ophthalmol. Vis. Sci.* **2012**, *53*, 6776–6784. [[CrossRef](#)]
40. Asaoka, R.; Murata, H.; Asano, S.; Matsuura, M.; Fujino, Y.; Miki, A.; Tanito, M.; Mizoue, S.; Mori, K.; Suzuki, K. The usefulness of the Deep Learning method of variational autoencoder to reduce measurement noise in glaucomatous visual fields. *Sci. Rep.* **2020**, *10*, 7893. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.