Improving Methods on Explainable Diabetic Retinopathy Detection and Retinal Image Generation

Jiabao Sean Xiao

January 2024

Abstract

In recent years, deep learning has demonstrated great capability in classifying the label and severity grade of different diseases, a few of which try to give explanations on how to make predictions. Drawing inspiration from Koch's Postulates, which serve as a cornerstone in evidence-based medicine (EBM) for identifying pathogens, our work aims to harness the interpretability of deep learning for medical diagnosis. We endeavor to elucidate the decision-making process of a diabetic retinopathy (DR) detector by identifying and isolating the neuron activation patterns it relies upon, thereby establishing a direct connection between these patterns and the presence of lesions for a pathologically informed explanation. To be specific, based on former researchers introducing innovative pathological descriptors derived from the activated neurons within the DR detector, encapsulating both the spatial and visual characteristics of lesions, we present Patho-GAN2, an innovative network designed to generate medically accurate retinal images, facilitating the visualization of symptoms encoded by these descriptors. The images produced by our method surpass those generated by earlier approaches in terms of quantitative measures.

Index Terms: Explainable deep learning, explainable artificial intelligence, medical image analysis, medical image generation, generative adversarial network.

Contents

1 Introduction **2**

2 Related Works **2**
2.1 Diabetic Retinopathy Detection 2
2.2 Interpretable Deep Learning 2
2.3 Generative Adversarial Networks . . . 3
2.4 Synthesizing Biomedical Images 3

3 Pathological Descriptor **4**
3.1 DR Detection Network 4
3.2 Activation Network 4
3.3 Retinal Pathological Descriptor 4

4 Visualizing Pathological Descriptor **4**
4.1 Patho-GAN Overview 4
4.2 Generator Network 5
4.3 Discriminator Network 5
4.4 Loss Settings 5

5 Experimental Results **5**
5.1 Dataset 5
5.2 Lesion Manipulations 6
5.3 Quantitative Evaluation 6

6 Discussion **6**
6.1 Interpretability gained with Patho-GAN2 6
6.2 Future Works 7

7 Conclusion **7**

8 Acknowledgement **7**

1 Introduction

NOWADAYS, deep learning has emerged as a widely adopted approach in the field of medical imaging analysis, with notable applications in the detection of diabetic retinopathy (DR) [1], [2] and the classification of brain cancer [3]. While having demonstrated a high level of accuracy in identifying specific disease labels and assessing severity stages, a significant challenge remains in the algorithms' lack of interpretability, which is particularly prevalent within the deep learning community and poses a critical barrier to the broader acceptance of these technologies. Hence, it is vital to enhance the ability of interpreting the decision-making process of these algorithms in medical imaging applications, or the adversarial attacks [4] might lead to serious problems. What's more, other techniques such as self-driving automobiles [5] also encounter such limitations.

In this paper, we manage to improve a technique based on Koch's Postulates to in some way interpret how convolutional neural network (CNN) based imaging medical detector makes decisions [6]. Note that the interpretive framework is not confined to this application alone but is adaptable to a broad range of deep learning-based medical imaging models, we take diabetic retinopathy (DR), a prevalent condition leading to vision impairment in individuals with diabetes, as a case study.

Initially, we delineate and procure pathological descriptors that encapsulate the neuron activations intrinsically linked to the diagnostic predictions of the DR detector [7]. The process of distinguishing these pertinent neurons, often numbering in the thousands, from the millions present in a neural network, mirrors the task of isolating a potential pathogen from an organism afflicted by disease. Based on Koch's Postulates, we introduce pathological descriptors into the binary vessel segmentation to create a medically plausible retinal image that manifests the anticipated lesions with the help of Patho-GAN2. Utilizing the pathological descriptors and binary vessel segmentation, the anticipated symptoms with precise quantification and location can be successfully exhibited in the generated image.

Given that our descriptors are lesion-oriented and include spatial coordinates, we possess the capability to systematically adjust their positions and quantities, thereby facilitating the generation of images featuring controlled types and extents of symptoms. Furthermore, enhancements made to Patho-GAN [6], which serves as a viable data augmentation technique for medical image analysis, substantiate the superior performance of our upgraded model, Patho-GAN2.

2 Related Works

2.1 Diabetic Retinopathy Detection

Diabetic retinopathy (DR) is a prevalent condition that leads to vision impairment or blindness in individuals with diabetes, affecting a substantial population of 347 million people [8]. Ophthalmologists traditionally assess the severity of DR based on the type and quantity of associated lesions, such as the count of microaneurysms. Consequently, there is a significant demand for automated detection systems to alleviate the workload of ophthalmologists and facilitate early diagnosis in diabetic patients, thereby impeding the progression of DR [1], [2].

In 2015, a competition facilitated by Kaggle [9] aimed at the automated classification of retinal images into five categories as per the International Clinical Diabetic Retinopathy Disease Severity Scale [10] was conducted. The leading techniques in this competition utilized deep learning algorithms. Despite their effectiveness in terms of sensitivity and specificity, these deep learning methodologies were not designed to provide straightforward explanations for their predictions. Subsequent techniques (e.g. [11], [12]) redirected efforts towards identifying the positions of lesions through a weakly supervised learning model. Nevertheless, the deployment of these approaches frequently necessitated an extensive compilation of lesion annotations curated by medical professionals. Following Yuhao Niu and his colleagues[6], we choose o_O detector, a CNN-based approach that was among the top three performers in the mentioned contest.

2.2 Interpretable Deep Learning

Interpretable deep learning has become a significant area of study aimed at demystifying the black box of neural networks. Several methodologies have been developed to enhance the interpretability of deep neural networks (DNNs).

Feature visualization [13], for example, aim to represent visually the features and patterns that neural networks learn to recognize, thereby providing insights into the network's focus and biases. Attention mechanisms [14], particularly in sequence models, highlight the elements of the input data deemed most relevant for predictions, offering a direct view into the model's decision-making process.

Gradient-based back-propagation techniques such as Class Activation Mapping (CAM) [15], Gradient-weighted Class Activation Mapping (GradCAM) [16], GradCAM++ [17], and Deep Learning Important FeaTures (DeepLIFT) [18] generate heat maps

through back-propagation to highlight discriminative regions instrumental for class prediction.

To augment the reciprocal comprehension among the artificial intelligence and clinical medicine communities, an innovative framework [6] was proposed that elucidates the neuron-level representation by adhering to the principles of evidence-based medicine (EBM). The framework, although assessed within the context of Diabetic Retinopathy Detection [10], possesses the versatility to be adapted for the explication of various other medical deep learning algorithms. Initially, we employ a method to autonomously discern the neuron representations encoding symptom information pivotal to disease classification. Subsequently, through the generation of synthetic lesions in encoded DR descriptors [7], we facilitate the correlation between pathological lesions and corresponding neuron activation patterns.

2.3 Generative Adversarial Networks

Introduced in 2014, Generative Adversarial Networks (GANs) [19] implemented the concept of a zero-sum game for the creation of lifelike images. Following this, Conditional GANs (CGANs) [20] endeavored to introduce supplementary data to steer the generative process. DC-GAN [21] combined CNN with traditional GAN to achieve a shocking effect. Pix2pix [22] exploited the U-Net [23] converged with adversarial training and accomplished great results. Moreover, CycleGAN [24] employed a dual-GAN framework and introduced a cycle consistency loss, facilitating style transfer across unpaired datasets.

2.4 Synthesizing Biomedical Images

Conventional methodologies for synthesizing biomedical imaging leverage the extensive corpus of medical and biological knowledge accrued by humanity, in conjunction with sophisticated simulation techniques, to yield realistic outcomes. Among the most notable endeavors in this domain are probably GENESIS [25], NEURON [26], L-Neuron [27] and so on.

Amidst the advancements in deep learning, techniques such as Tubs-GAN [28] have emerged, synthesizing realistic retinal and neuronal imagery through data-driven approaches. These methods generate images by incorporating styles from a reference retinal image and vascular patterns from a binary segmentation map. Furthermore, Patho-GAN [6] introduced a pathologically controllable approach capable of generating realistic retinal images featuring medically plausible symptoms. In this manuscript, we present Patho-GAN2, an enhanced methodology derived from the foundational principles of Patho-GAN.

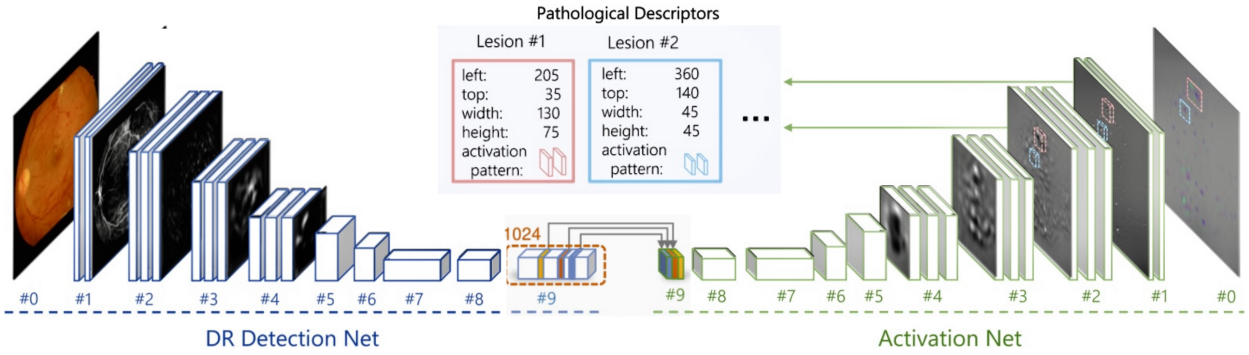

Figure 1: The extraction of pathological descriptors begins with inputting a pathological reference image into a detection network. Features extracted are inversely propagated to the input pixel space using a symmetric activation network, yielding activation projections that highlight lesion locations and features. These projections are then segmented into small patches around the lesions to create pathological descriptors.. Note that the illustration excludes skip connections between the detection and activation networks.

3 Pathological Descriptor

In the forthcoming section, we delineate the methodology for extracting lesion-based pathological descriptors, aimed at encoding the activated neurons within a Diabetic Retinopathy (DR) detection framework [7].

3.1 DR Detection Network

Here, o_O DR detector [7] utilized in this paper will be introduced. The system processes retinal fundus imagery, specifically of dimensions 448×448×3, to classify diabetic retinopathy severity into five categories, ranging from 0 to 4. Illustrated in the left segment of Figure 1, the architecture of the diabetic retinopathy (DR) detection network incorporates multiple blocks. Each block is composed of two to three convolutional layers, accompanied by a pooling layer. With the incremental addition of layers, the architecture coalesces into a bottleneck feature measuring $1 \times 1 \times 1024$. To inject non-linearity into the network and circumvent the phenomenon of neuron inactivation, an Exponential Linear Unit (ELU) [29] is integrated subsequent to each convolutional and dense layer.

The architecture is subjected to training utilizing the Kaggle Diabetic Retinopathy dataset [9], employing Nesterov momentum across a span of 250 epochs. A suite of data augmentation techniques is implemented, encompassing dynamic data re-sampling, random stretching, rotational adjustments, flipping, and modifications in coloration.

3.2 Activation Network

Within the vast network comprising millions of neurons, a mere few thousand play a crucial role in activating bottleneck features and ultimately influencing the model's predictions. To delve into the dynamics of these select neurons, we employ an activation network [30], which involves a methodology analogous to back-propagation, targeting the six-dimensional key bottleneck features to generate activation projections for the detector feature layers.

The depiction in the right segment of Figure 1 illustrates our activation network as an inverse configuration of the DR detector, utilizing duplicated weights. This arrangement ensures that for every layer present in the detector, a corresponding reverse layer exists within the activation network, maintaining identical stride and kernel sizes. Specifically, (1) a convolutional layer in the detector corresponds to a layer performing transposed convolution in the activation network, employing the same weights albeit with the kernel flipped both vertically and horizontally; (2) for every max pooling layer, an unpooling layer performs a quasi-inverse function, identifying the maximum elements through a skip connection (not depicted in the figure) while filling the remaining elements with zeros; (3) a ReLU function within the detector finds its counterpart in the activation network, which negates negative activation projections; (4) the fully connected dense layer is conceptualized as a 1×1 convolution. The process leverages auto-differentiation capabilities offered by Tensorflow [**?**] to back-propagate through each layer.

3.3 Retinal Pathological Descriptor

For a given fundus image denoted as x, a set of descriptors, represented as $D(x)$, can be extracted. These descriptors encode crucial information including the coordinates, dimensions, and activation patterns of lesions within x. Furthermore, these descriptors, D, are instrumental in reconstructing activation projections. This reconstruction is accomplished by incorporating the encoded lesion-associated neuron activity into the activation maps. We referred to [31] for details.

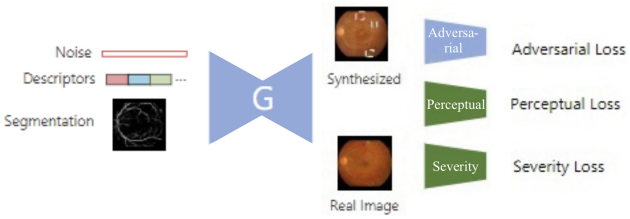

Figure 2: Summary of the Patho-GAN2 training process: The Generator undergoes training with limitations imposed by three sub-networks and associated losses. The trainable components are depicted in blue, while the pre-trained and fixed elements are illustrated in green.

4 Visualizing Pathological Descriptor

4.1 Patho-GAN Overview

As depicted in Figure 2, the generator network is engineered to fabricate a retinal image, incorporating vascular structures within a segmentation map alongside specified symptoms within descriptors. In the process of training this network, the inclusion of three auxiliary networks is necessitated: the discriminator network, the perceptual network, and the Diabetic

Retinopathy (DR) detection network, accompanied by three pertinent loss functions. The discriminator network, in tandem with the generator, is tasked with differentiating between synthesized and authentic images, thereby fostering an adversarial improvement in the generator's performance. To augment the fidelity of pathological and physiological detail, a pre-trained perceptual network is employed to impose constraints on detail reconstruction. Additionally, the integration of a pre-trained DR detector aims to guarantee that the synthesized image convincingly displays the appropriate symptoms. In contrast to the dynamic training of the generator and discriminator, both the perceptual network and the DR detector remain static, with no updates during the training phase. Upon completion of the training, the generator is capable of producing synthesized fundus images in a singular forward pass, utilizing vessel segmentation, descriptors, and a noise vector.

4.2 Generator Network

The generator network, symbolized as G_θ, processes inputs consisting of a vessel segmentation image y within the binary space $0,1^{W\times H}$, pathological descriptors $\mathcal{D}(x)$, and a noise vector $z \in \mathbb{R}^Z$. Here, y furnishes physiological details, $\mathcal{D}(x)$ imparts pathological insights, and z introduces variability to emulate the conditional distribution of images. This procedure culminates in the generation of a diabetic retinopathy fundus image $\hat{x}$, represented in the continuous interval $[0,1]^{W\times H\times 3}$. The synthesis operation is formally described by the function $\hat{x} = G_\theta(y, \mathcal{D}(x), z)$.

For the architecture of the generator, an SA-UNet [32] is employed. The initial vessel segmentation y undergoes dimensional reduction through six stages of Convolution-BatchNorm-ELU, characterized by a kernel size of 3, a stride of 2, and the absence of pooling layers. Pathological descriptors $\mathcal{D}(x)$ are leveraged to form activation projections, which are subsequently integrated during the down-sampling phase. The noise vector z undergoes a transformation through a fully connected layer and convolution before being merged into the generator's bottleneck. The expansion phase incorporates six blocks of Resize-Convolution-BatchNorm.

4.3 Discriminator Network

The discriminator network, denoted as D_γ, is tasked with differentiating between generated and authentic images. This network can also be conceptualized as a discriminant function, expressed as $p = D_\gamma(X, y)$, where p ranges between 0 and 1. For an authentic image $X = x$, the value of p is expected to approach 1, whereas for a synthesized image $X = \hat{x}$, p is anticipated to approach 0. The architecture of the discriminator incorporates five sequential blocks, each consisting of Convolution-BatchNorm-ELU configurations, utilizing a kernel size of 3, a stride of 2, and omitting any pooling layers.

4.4 Loss Settings

Our equation follow the GAN's strategy, with:

$$\min_\theta \max_\gamma L\left(G_\theta, D_\gamma\right) = \mathbb{E}_{x,y}\left[L_{\text{adv}}\left(x, \hat{x}, y; \theta, \gamma\right) + w_p \cdot L_{\text{perceptual}}\left(x, \hat{x}; \theta\right) + w_s \cdot L_{\text{severity}}\left(x, \hat{x}; \theta\right)\right]$$

where L_{adv} is the adversarial loss, with $L_{\text{perceptual}}$ and L_{severity} respectively being perceptual loss and severity loss.

Adversarial loss for genuine and generated images is given by $L_{\text{adv}} = \log D_\gamma(x, y) + \log(1 - D_\gamma(G_\theta(y, z), y))$. Optimizing the discriminator's parameters, γ, necessitates maximizing L_{adv}, whereas optimizing the generator's parameters, θ, involves minimizing a composite loss $L_G = \tilde{L}_{\text{adv}} + w_p \cdot L_{\text{percept}} + w_s \cdot L_{\text{severity}}$, where $\tilde{L}_{\text{adv}}$ is the adversarial loss for generated images: $\tilde{L}_{\text{adv}} = -\log D_\gamma(G_\theta(y, z), y)$. Perceptual Loss, quantifying discrepancies between real and synthesized images in a perceptual network's feature space, employs a pre-trained VGG-19 for definition: $L_{\text{percept}} = \|F_V^\lambda(x) - F_V^\lambda(\hat{x})\|$. This approach enhances the accuracy of the generated images in reflecting both pathological details and physiological structures.The introduction of severity loss aims to ensure medical equivalency between the synthesized image $\hat{x}$ and the reference genuine image x. More precisely, this involves enforcing an equivalence in severity levels between the synthesized and real images. The divergence in severity is quantitatively assessed using a trained Diabetic Retinopathy (DR) detector, denoted as o_O. Consequently, the severity loss is articulated as $L_{\text{severity}} = \|\text{DR}(x) - \text{DR}(\hat{x})\|$, facilitating the alignment of pathological severity between the generated and authentic images.

5 Experimental Results

5.1 Dataset

quadMultiple online repositories offer access to fundus image datasets, including DRIVE [33], Kaggle

Dataset Name	DR Lesion Annotation	Vessel Segmentation	Resolution	Number of Images
DRIVE [58]	(No DR)	Yes	584×565	20 train + 20 test
EyePACS Kaggle [10]	Only severity levels	No	1444×1444∼2184×3456	35.1k train + 53.6k test
IDRiD [59]	Pixel-wise lesion segmentation	No	4288×2848	54 train + 27 test
Retinal-Lesions [60]	Lesion annotation in circle & severity levels	No	896×896	337 train + 1256 test
FGADR [61]	Pixel-wise lesion segmentation & severity levels	No	1280×1280	500 train + 1342 test

Figure 3: Retinal Image Datasets and Their Attributes [6]

(EyePACS) [9], IDRiD [34], Retinal-Lesions [35], and FGADR [36]. The DRIVE dataset [33] comprises 40 primary retinal images featuring pixel-level vessel segmentation, albeit with minimal lesion presence. Kaggle (EyePACS) [9] provides a collection of 88.7k retinal images, classified across five Diabetic Retinopathy (DR) severity levels (inclusive of non-DR cases), with each image assigned a specific DR grade. To advance DR research, IDRiD [34] offers pixel-wise segmentation for four distinct lesion types across 81 images. The Retinal-Lesions dataset [35] delineates up to eight lesion varieties using circular regions. The most recent FGADR dataset [36] encompasses 1842 images with pixel-wise segmentation for six lesion types. These datasets' characteristics are detailed in Figure 3 [6].

In our research, images from DRIVE and their associated vessel segmentations are exploited to train a vessel segmentation network, improved SA-UNet [32], [37], [38]. Images and severity classifications from Kaggle are employed in training the DR detector. Subsequently, our PathoGAN model is trained and assessed using data from IDRiD [34], Retinal-Lesions [35], and FGADR [36].

5.2 Lesion Manipulations

Given that our descriptors are formulated to be lesion-based, incorporating spatial coordinates, we possess the capability to adjust the positioning of lesions through alterations to their respective coordinates. Furthermore, the quantification of lesions can be modulated by either replicating or eliminating certain descriptors. Take Number Manipulation as an example. Through the replication or deletion of select input descriptors, it becomes feasible to regulate the volume of synthesized lesions, thereby facilitating a reduction or augmentation in their quantity.

5.3 Quantitative Evaluation

We here undertake a quantitative assessment of Patho-GAN2 alongside comparative analyses with other methodologies[6], [28]. This evaluation involves the generation of imagery via these methodologies on datasets IDRiD [34], Retinal-Lesions [35], and FGADR [36]. Subsequently, we appraise the resemblance between the synthetically produced images and the authentic images employing two statistical measures: Frecshet Inception Distance (FID) and Mean Squared Error (MSE). Freshet Inception Distance (FID) is commonly utilized to quantitatively gauge the disparity between two sets of images. Conversely, Mean Squared Error (MSE) quantifies the average pixel-wise variance between pairs of images or correlating sets of images. The outcomes, as presented in Figure 4, delineate that our results surpass those of preceding studies, indicating that Patho-GAN2 is capable of generating retinal images that bear a closer resemblance to real images.

Methods	IDRiD		Retinal-Lesions		FGADR	
	FID ↓	MSE ↓	FID ↓	MSE ↓	FID ↓	MSE ↓
Tub-sGAN [7]	123.22	0.0120	68.98	0.0250	40.57	0.0153
Patho-GAN $_{4\times4}$	81.12	0.0095	**22.28**	0.0 142	20.37	0.0120
Patho-GAN2	**78.89**	**0.0082**	22.32	**0.0139**	**20.03**	**0.0099**

Figure 4: Quantitative Measurement on Different Methods (Lower is better)

6 Discussion

6.1 Interpretability gained with Patho-GAN2

The primary focus of the diabetic retinopathy (DR) detector is directed towards the identification and analysis of lesions. In this research, we endeavor to provide an interpretive framework for comprehending the underlying mechanisms of a deep learning-based DR detector, drawing inspiration from the methodological principles akin to Koch's Postulates. Initially, our approach involves the extraction of pathological descriptors from retinal images, with the aim of characterizing the spatial coordinates, dimensions, and visual attributes of DR lesions. Subsequently, we validate the efficacy of these descriptors by faithfully reconstructing lesions based on them. Through an application of Koch's Postulates, our findings suggest a close association between the activation patterns of neurons and the presence of pathological lesions.

6.2 Future Works

In future works, our interpretation framework is not limited to medical tasks such as DR detection. It also applies on broader critical applications such as medicine, banking and self-driving automobiles. Our method has the potential to enhance decision reliability, algorithm fairness, and ensure network performance of general CNN models.

7 Conclusion

In order to leverage network interpretability within the context of medical imaging, akin to the approach outlined in Koch's Postulates, we enhance an existing technique, Patho-GAN, to facilitate the visualization of pathological descriptors within retinal images representative of pathology derived from previously unseen binary vessel segmentations. The resultant images exhibit medically plausible and controllable symptoms, thereby validating the direct association between specific lesions and the prediction of diabetic retinopathy (DR) grade. This elucidatory endeavor contributes to the broader comprehension within the medical community regarding the predictive mechanisms of deep learning algorithms, thereby fostering opportunities for enhanced collaboration and understanding.

8 Acknowledgement

Consequently, I would like to thank (the list is not ordered):

- my mentor Haoyang He for his guidance on this topic, as the project would not be possible without him.
- my friends such as Hangyu Li, Dongchen Zou for there support and advice.
- the Yuanpei Young Scholars Program for offering such a fantastic opportunity.
- my family for being by my side.
- ChatGPT, an AI-based language model developed by OpenAI, for its help (only) in editing the English language of this manuscript.

References

[1] Sehrish Qummar, Fiaz Gul Khan, Sajid Shah, Ahmad Khan, Shahaboddin Shamshirband, Zia Ur Rehman, Iftikhar Ahmed Khan, and Waqas Jadoon. A deep learning ensemble approach for diabetic retinopathy detection. *Ieee Access*, 7:150530–150539, 2019.

[2] Varun Gulshan, Lily Peng, Marc Coram, Martin C Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, Kasumi Widner, Tom Madams, Jorge Cuadros, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *jama*, 316(22):2402–2410, 2016.

[3] Heba Mohsen, El-Sayed A El-Dahshan, El-Sayed M El-Horbaty, and Abdel-Badeeh M Salem. Classification using deep learning neural networks for brain tumors. *Future Computing and Informatics Journal*, 3(1):68–71, 2018.

[4] Samuel G Finlayson, John D Bowers, Joichi Ito, Jonathan L Zittrain, Andrew L Beam, and Isaac S Kohane. Adversarial attacks on medical machine learning. *Science*, 363(6433):1287–1289, 2019.

[5] Yun Tian, Kexin Pei, Suman Jana, and Baishakhi Ray. Deeptest: Automated testing of deep-neural-network-driven autonomous cars. *Cornell University - arXiv,Cornell University - arXiv*, Aug 2017.

[6] Yuhao Niu, Lin Gu, Yitian Zhao, and Feng Lu. Explainable diabetic retinopathy detection and retinal image generation. *arXiv: Image and Video Processing,arXiv: Image and Video Processing*, Jul 2021.

[7] M. Antony. Team o o solution for the kaggle diabetic retinopathy detection challenge. `https://www.kaggle.com/c/diabetic-retinopathy-detection/discussion/15807`, 2016.

[8] Goodarz Danaei, Mariel M Finucane, Yuan Lu, Gitanjali M Singh, Melanie J Cowan, Christopher J Paciorek, John K Lin, Farshad Farzadfar, Young-Ho Khang, Gretchen A Stevens, Mayuree Rao, Mohammed K Ali, Leanne M Riley, Carolyn A Robinson, and Majid Ezzati. National, regional, and global trends in fasting plasma glucose and diabetes prevalence since 1980: systematic analysis of health examination surveys and epidemiological studies with 370 country-years and 2·7 million participants. *The Lancet*, page 31–40, Jul 2011.

[9] Kaggle. Kaggle diabetic retinopathy detection challenge. `https://www.kaggle.com/c/diabetic-retinopathy-detection`, 2015.

[10] Shion Haneda and Hidetoshi Yamashita. [international clinical diabetic retinopathy disease severity scale]. *Nihon rinsho. Japanese journal of clinical medicine,Nihon rinsho. Japanese journal of clinical medicine*, Nov 2010.

[11] Zhe Wang, Yanxin Yin, Jianping Shi, Wei Fang, Hongsheng Li, and Xiaogang Wang. Zoom-in-net: Deep mining lesions for diabetic retinopathy detection. In *Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III 20*, pages 267–275. Springer, 2017.

[12] Yehui Yang, Tao Li, Wensi Li, Haimeng Wu, Fan Wang, and Wensheng Zhang. Lesion detection and grading of diabetic retinopathy via two-stages deep convolutional neural networks. *Cornell University - arXiv,Cornell University - arXiv*, May 2017.

[13] Chris Olah, Alexander Mordvintsev, and Ludwig Schubert. Feature visualization. *Distill*, 2(11):e7, 2017.

[14] Zhaoyang Niu, Guoqiang Zhong, and Hui Yu. A review on the attention mechanism of deep learning. *Neurocomputing*, 452:48–62, 2021.

[15] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning deep features for discriminative localization. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2016.

[16] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[17] Aditya Chattopadhay, Anirban Sarkar, Prantik Howlader, and Vineeth N Balasubramanian. Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Mar 2018.

[18] Avanti Shrikumar, PeytonG. Greenside, and A Kundaje. Learning important features through propagating activation differences. *arXiv: Computer Vision and Pattern Recognition,arXiv: Computer Vision and Pattern Recognition*, Apr 2017.

[19] IanJ Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, and Delhi Delhi. Generative adversarial nets.

[20] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *Cornell University - arXiv,Cornell University - arXiv*, Nov 2014.

[21] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[22] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul 2017.

[23] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science,Lecture Notes in Computer Science*, Jan 2015.

[24] Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei Efros, Summer Winter, Van Gogh, Cezanne Monet, and Ukiyo-E Photos. Unpaired image-to-image translation using cycle-consistent adversarial networks.

[25] James M. Bower, Hugo Cornelis, and David Beeman. *GENESIS, the GEneral NEural SImulation System*, page 1287–1293. Jan 2015.

[26] Nicholas T Carnevale and Michael L Hines. *The NEURON book*. Cambridge University Press, 2006.

[27] Giorgio A Ascoli and Jeffrey L Krichmar. L-neuron: a modeling tool for the efficient generation and parsimonious description of dendritic morphology. *Neurocomputing*, 32:1003–1011, 2000.

[28] He Zhao, Huiqi Li, Sebastian Maurer-Stroh, and Li Cheng. Synthesizing retinal and neuronal images with generative adversarial nets. *Medical Image Analysis*, page 14–26, Oct 2018.

[29] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289*, 2015.

[30] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13*, pages 818–833. Springer, 2014.

[31] Yuhao Niu, Lin Gu, Feng Lu, Feifan Lv, Zongji Wang, Imari Sato, Zijian Zhang, Yangyan Xiao, Xunzhang Dai, and Tingting Cheng. Pathological evidence exploration in deep retinal image diagnosis. *Proceedings of the AAAI Conference on Artificial Intelligence*, page 1093–1101, Sep 2019.

[32] Changlu Guo, Márton Szemenyei, Yugen Yi, Wenle Wang, Buer Chen, and Changqi Fan. Sa-unet: Spatial attention u-net for retinal vessel segmentation. In *2020 25th international conference on pattern recognition (ICPR)*, pages 1236–1242. IEEE, 2021.

[33] J. Staal, M.D. Abramoff, M. Niemeijer, M.A. Viergever, and B. van Ginneken. Ridge-based vessel segmentation in color images of the retina. *IEEE Transactions on Medical Imaging*, page 501–509, Apr 2004.

[34] Prasanna Porwal, Samiksha Pachade, Manesh Kokare, Girish Deshmukh, Jaemin Son, Woong Bae, Lihong Liu, Jianzong Wang, Xinhui Liu, Liangxin Gao, TianBo Wu, Jing Xiao, Fengyan Wang, Baocai Yin, Yunzhi Wang, Gopichandh Danala, Linsheng He, Yoon Ho Choi, Yeong Chan Lee, Sang-Hyuk Jung, Zhongyu Li, Xiaodan Sui, Junyan Wu, Xiaolong Li, Ting Zhou, Janos Toth, Agnes Baran, Avinash Kori, Sai Saketh Chennamsetty, Mohammed Safwan, Varghese Alex, Xingzheng Lyu, Li Cheng, Qinhao Chu, Pengcheng Li, Xin Ji, Sanyuan Zhang, Yaxin Shen, Ling Dai, Oindrila Saha, Rachana Sathish, Tânia Melo, Teresa Araújo, Balazs Harangi, Bin Sheng, Ruogu Fang, Debdoot Sheet, Andras Hajdu, Yuanjie Zheng, Ana Maria Mendonça, Shaoting Zhang, Aurélio Campilho, Bin Zheng, Dinggang Shen, Luca Giancardo, Gwenolé Quellec, and Fabrice Mériaudeau. Idrid: Diabetic retinopathy – segmentation and grading challenge. *Medical Image Analysis*, page 101561, Jan 2020.

[35] Qijie Wei, Xirong Li, Wu Yu, Xiao Zhang, Yongpeng Zhang, Bin-Jie Hu, Bin Mo, Daozhi Gong, Ning Chen, Dayong Ding, and Youxin Chen. Learn to segment retinal lesions and beyond. *Cornell University - arXiv,Cornell University - arXiv*, Dec 2019.

[36] Yi Zhou, Boyang Wang, Lei Huang, Shanshan Cui, and Ling Shao. A benchmark for studying diabetic retinopathy: Segmentation, grading, and transferability. *IEEE Transactions on Medical Imaging*, page 818–828, Mar 2021.

[37] Ariel Rechtman Nadav Potesman. Sa unet improved. *Cornell University - arXiv,Cornell University - arXiv*, Aug 2023.

[38] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8110–8119, 2020.