

Evaluation of Unconditioned Deep Generative Synthesis of Retinal Images

Kaplan Sinan, Lensu Lasse, Laaksonen Lauri, Uusitalo Hannu

This is a Final draft version of a publication
published by Springer, Cham
in International Conference on Advanced Concepts for Intelligent Vision Systems

DOI: 10.1007/978-3-030-40605-9_23

Copyright of the original publication: © Springer Nature Switzerland AG 2020

Please cite the publication as follows:

Kaplan S., Lensu L., Laaksonen L., Uusitalo H. (2020) Evaluation of Unconditioned Deep Generative Synthesis of Retinal Images. In: Blanc-Talon J., Delmas P., Philips W., Popescu D., Scheunders P. (eds) Advanced Concepts for Intelligent Vision Systems. ACIVS 2020. Lecture Notes in Computer Science, vol 12002. Springer, Cham. https://doi.org/10.1007/978-3-030-40605-9_23

**This is a parallel published version of an original publication.
This version can differ from the original published article.**

Evaluation of Unconditioned Deep Generative Synthesis of Retinal Images

Sinan Kaplan¹, Lasse Lensu¹, Lauri Laaksonen¹, and Hannu Uusitalo¹

¹Computer Vision and Pattern Recognition Laboratory, Lappeenranta-Lahti University of Technology LUT, P.O. Box 20, 53850 Lappeenranta, Finland
`lasse.lensu@lut.fi`

²Department of Ophthalmology, Faculty of Health and Biotechnology, Tampere University and Tays Eye Center, Tampere, Finland

Abstract. Retinal images have been increasingly important in clinical diagnostics of several eye and systemic diseases. To help the medical doctors in this work, automatic and semi-automatic diagnosis methods can be used to increase the efficiency of diagnostic and follow-up processes, as well as enable wider disease screening programs. However, the training of advanced machine learning methods for improved retinal image analysis typically requires large and representative retinal image data sets. Even when large data sets of retinal images are available, the occurrence of different medical conditions is unbalanced in them. Hence, there is a need to enrich the existing data sets by data augmentation and introducing noise that is essential to build robust and reliable machine learning models. One way to overcome these shortcomings relies on generative models for synthesizing images. To study the limits of retinal image synthesis, this paper focuses on the deep generative models including a generative adversarial network and a variational autoencoder to synthesize images from noise without conditioning on any information regarding to the retina. The models are trained with the Kaggle EyePACS retinal image set, and for quantifying the image quality in a no-reference manner, the generated images are compared with the retinal images of the DiaRetDB1 database using common similarity metrics.

Keywords: Deep generative model · Generative adversarial network · Variational autoencoder · Retinal image

1 Introduction

The unique structure of the eye makes it possible to noninvasively image the retina, a part of our central nervous system. This opportunity and the known relation of retinal changes to several eye, neuronal and systemic diseases makes it attractive for the researchers to study the images for screening and diagnosing diseases like diabetic retinopathy, age-related macular degeneration glaucoma, cerebrovascular and cardiovascular diseases, multiple sclerosis and Alzheimer's disease [21]. The diagnosis of abnormal medical conditions is mainly based on the

visual evaluation of the signs typical for the disease. The abnormalities are traditionally analyzed by experts giving their statements based on their education and expertise. This work is time consuming and often subjected to variations between the judgments of individual experts. In addition, quantitative changes in these images are usually more difficult to determine and hard to grade accurately. Therefore, automatic image processing methods are a well-motivated option to help an expert’s work and for enabling more accurate quantification of the changes, follow-ups of progressive diseases and high-throughput screening programs [26].

The relevant information about the aforementioned diseases can be gathered from red-green-blue (RGB) images of the retina [21]. While the development in the field of retinal image analysis improves with the help of advances in technology, the demand for retinal image data increases. To study and detect possible abnormalities connected with eye and systemic diseases, the availability of retinal data is crucial to utilize machine learning methods. Although there are publicly available retinal image data sets provided by research institutes and hospitals, these data sets suffer from unbalanced occurrences of abnormalities (related to medical conditions or symptoms visible on the images) and inaccurate annotations.

As a remedy for the lacking retinal image data sets, there is considerable potential in image generation for further development of retinal image analysis methods. While natural medical data is subject to regulations related to the acquisition, storing and anonymization of the medical data, synthetic data avoids the ethical concerns and can be shared and published without privacy issues. In addition, the generated data can be used to improve the performance of baseline supervised methods for lesion detection and image segmentation. This can be achieved by generating synthetic images for a particular context by introducing noise to the machine learning methods in order to build robust models. This way it is possible to increase the balanced heterogeneity and size of the data sets.

An important approach in terms of generating retinal images is to apply generative models. In this paper, deep generative models are studied and employed on the Kaggle EyePACS retinal image data set to generate synthetic retinal images. The aim is 1) to apply deep generative models including a generative adversarial network (GAN) and a variational autoencoder (VAE) for generating synthetic retinal images without conditioning on extra information and 2) to utilize a similarity-based statistical image quality assessment method for quantitative evaluation of the generated retinal images. The work is based on the master’s thesis project of the first author [19].

2 Retinal image generation and deep generative models

Related studies for retinal image generation include can be divided into traditional and deep learning based approaches. In [12], the aim was to reconstruct “only” the vessel tree. In [14, 3], the focus was on reconstructing the retina by preserving the major features of the retina: the optic disk, vessel tree, and fovea.

The former approach produced unrealistic vascular network structures, thus, further development in the latter succeeded to reconstruct more realistic retinal images while preserving the vascular structure.

One of the recent and most interesting developments in deep learning research [23, 10] is related to deep generative models. GANs [16] and VAEs [22] are such generative models. Related to retinal image generation, [2] presents a two-stage GAN for first generating label images for the retinal features and then the actual images. In [5], the focus is on AMD and generation of images as a proxy between humans and the automatic methods. The previous studies in this context focused on generating retinal image by conditioning on vessel trees [2, 5, 8]. In this paper, however, GANs and VAEs are compared to study the limits of deep generative models for synthesizing retinal images without conditioning on any extra information.

2.1 Generative adversarial networks

In 2014, Goodfellow et al. [16] proposed an adversarial network framework as an alternative generative model estimation process for deep generative networks, called GAN. The main principle of a GAN is that there are two neural networks called the generator (G) and the discriminator (D) that compete to maximize their gains. The main goal for both networks is to improve their capabilities to generate and discriminate the data. In this context, G draws samples from a random noise distribution and D discriminates whether the samples are drawn from G or from real data (the training data).

2.2 Variational autoencoders

Different from the GANs, there is a probabilistic graphical deep generative model proposed by [22] or Rezende et al. [29] (combination of deep neural networks and Bayesian inference models) to generate data samples from a latent space representation. This generative directed graphical model is called a VAE.

In a VAE, the latent space is used to draw samples from a probability density function (PDF) and these samples are used to generate new data. The latent space primarily provides the information about the underlying hidden structure inside the data [18].

A VAE consists of two primary blocks that are the *Encoder* and the *Decoder*. These building blocks are formed from a multilayer perceptron (MLP) to learn the parameters of the latent space.

2.3 Retinal image quality assessment

There is ambiguity regarding how to quantitatively evaluate the quality of data generated by deep generative models [30]. This is mostly because the reference for each image does not exist, which makes the evaluation correspond with reduced-reference or no-reference image quality evaluation. The studies conducted with

GANs and VAEs have suggested that the quality of the generated images can be evaluated by performing classification, segmentation and feature extraction methods within the problem domain [16, 6, 32, 22, 27, 4, 31]. Particularly, Inception score, Fréchet inception distance (FID) and sliced Wasserstein distance (SWD) have been commonly used as quantitative metrics of the generated images. In this paper, we applied FID and SWD to measure the quality of the generated retinal images. FID computes the Wasserstein-2 distance between the distributions of values in the embedding layer for the real and fake samples by using a trained classifier. In this paper, we used Inception net [4] trained with the samples from the Kaggle EyePACS set. SWD is a measure of the cost of transforming one distribution to another one using a given cost function [4].

In addition to FID and SWD, we propose a statistical parameters based image quality assessment method. The method is originated from the study in [13] by taking into account the structure of the generated retinal images. The parameters are used to measure the similarity between the generated images and a high-quality retinal image set. For this purpose, DiaRetDB1 [20] is considered as a benchmark. In the scope of this paper, by considering preceding research on retinal images [9], *mean*, *variance*, *entropy*, *skewness* and *kurtosis* are used to extract the statistical features from the retinal images for the quality assessment.

3 Experiments and Results

3.1 Data sets

We employed two different data sets for conducting the experiments. The Kaggle EyePACS retinal data set [1] was used for model training, whereas DiaRetDB1 [20] was used for the quality assessment of the generated retinal images. The Kaggle EyePACS set consists of 35,126 images with varying resolutions. DiaRetDB1 set contains 89 high resolution images - WxH (1500x1152). Examples from both sets are shown in Fig. 1.

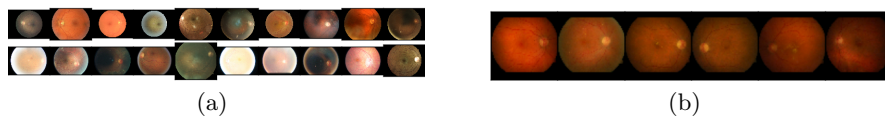


Fig. 1. Example images: (a) the Kaggle EyePACS data set used for training; (b) DiaRetDB1 data set used for quality evaluation.

3.2 Architectural details of deep generative models

This section covers the proposed architectural designs of the GAN and the VAE for the experiments. The implementation of the methods, the associated hyper-parameters for each model and the extra figures are accessible on Github¹.

Architecture of the generative adversarial network By having [28, 11, 15, 24, 28] as a guideline, we propose a GAN architecture seen in Fig. 2 and 3. The detailed description of the architecture is given in [19].

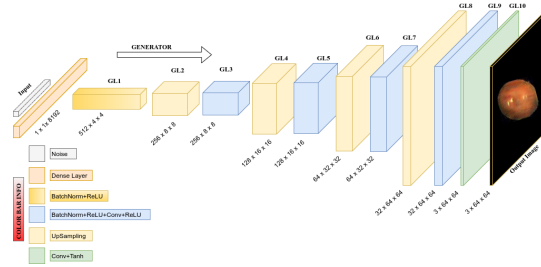


Fig. 2. Architecture of the proposed generator network which is a part of the generative adversarial network, based on convolutional neural network units with their parameters including the number of layers and their dimensions. GL stands for generative layer.

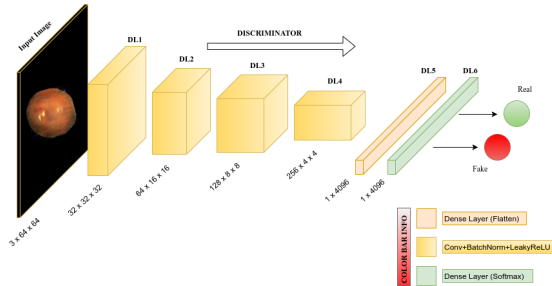


Fig. 3. Architecture of the proposed discriminator network which is a part of the generative adversarial network, based on convolutional neural network units with their parameters including the number of layers and their dimensions. DL stands for discriminator layer.

¹ The source code and the materials: <https://github.com/kaplansinan/MasterThesis>

Architecture of the variational autoencoder The VAE architecture is rather simple and it is based on a MLPs as seen in Fig. 4. By taking [22, 29] as the basis, both the encoder and the decoder blocks are designed to have only one hidden layer with 512 neurons in total. The encoder block takes an image ($3 \times 64 \times 64$) from the training set and encodes it into the latent space by learning the mean μ and variance σ^2 . Similarly, the decoder samples a noise vector z from the normal distribution with the learned mean μ and variance σ^2 and reconstructs it as an image with the dimensions of ($3 \times 64 \times 64$).

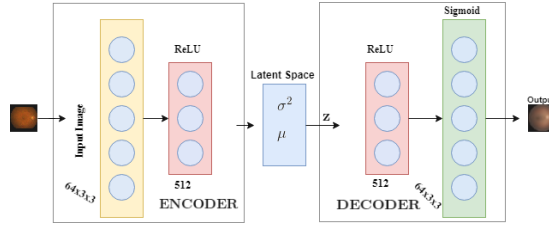


Fig. 4. The proposed variational autoencoder architecture based on multilayer perceptrons which are composed by a single hidden layer both in the encoder and decoder.

3.3 Retinal image generation with the generative adversarial network

Examples of the generated retinal images from the proposed GAN are demonstrated in Fig. 5. While qualitatively analyzing the generated images, one can see that the network is able to capture the global structure of the retina, including the *shape*, *optic disk*, *fovea*, and *macula*. The generated images share same diversity in terms of color as in the Kaggle EyePACS set (refer to Fig. 1). However, detailed structures of the retina such as vessel trees are not generated clearly.

3.4 Retinal image generation with the variational autoencoder

Examples of the generated retinal images with the VAE can be seen in Fig. 5(b). As in the case of the GAN, the VAE also generates the retinal images by preserving the global structure of the retina that contains an optic disk, fovea, macula and the overall shape of the retina. Yet, the VAE is also unable to generate the vessel trees as a local structure of the retina. Furthermore, the VAE mostly generates the retinal images with the dominant red channel. In addition to that, the generated retinal images appear blurry because of the mean squared error measuring the pixel-to-pixel distance.

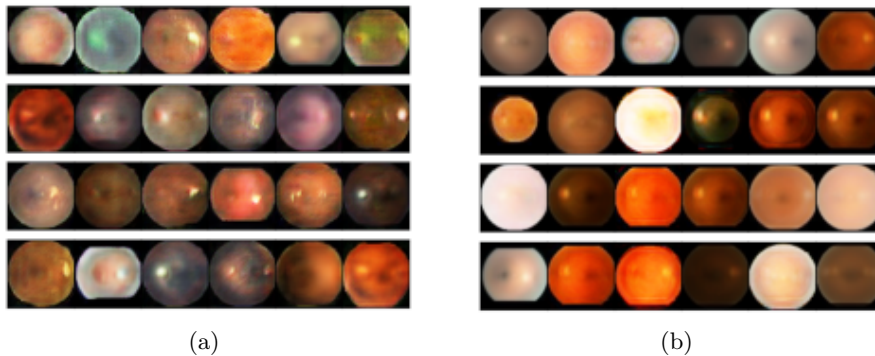


Fig. 5. Randomly chosen examples of the generated retinal images: (a) generative adversarial network, (b) variational autoencoder.

3.5 Quantitative analysis of generated retinal images

In order to assess the quality of the generated retinal images quantitatively, we follow the approach presented in Section 2.3. For the quality assessment, we present both FID and SWD metrics.

Table 1 shows that the visual quality of the generated images by the VAE is higher than the images generated by the GAN. As it is noted in the studies [4, 31], there is a negative correlation between the FID and the visual quality of the generated images.

Table 1. Fréchet inception distance assessment between the Kaggle EyePACS set and the generated retinal images both from the generative adversarial network (GAN) and variational autoencoder (VAE), and DiaRetDB1

	GAN	VAE	DiaRetDB1
EyePACS	161.202	125.00	130.45

Table 2 presented the similarity of distributions represented by the SWD metric between the subsets of The Kaggle EyePACS images and the retinal images generated by the GAN and VAE. To note that SWD is lower when the two distributions are more similar. In this case, the images generated by the GAN have more diversity which is closer to the one of the real data set. Thus, the SWD is lower compared to the images generated by the VAE.

In terms of the evaluation based on the statistical features, the mean, variance, skewness, kurtosis, and entropy were determined for the benchmark DiaRetDB1, the retinal images generated via the GAN and VAE, and five randomly selected subsets of the Kaggle EyePACS set. Each set contains 89 retinal images

Table 2. Sliced Wasserstein distance between the Kaggle EyePACS subsets and the images generated by the generative adversarial network (GAN) and variational autoencoder (VAE).

	Subset1	Subset2	Subset3	Subset4	Subset5
GAN	0.045	0.048	0.043	0.053	0.053
VAE	0.13	0.127	0.17	0.15	0.134

in total as the DiaRetDB1 set consists of 89 images. Once the features were extracted, the next step was to compute the cosine similarity between the features of 1) DiaRetDB1 and the retinal images generated by the GAN, 2) DiaRetDB1 and the retinal images generated by the VAE, and 3) DiaRetDB1 and the subset of retinal images from the Kaggle EyePACS set. The analysis of each statistical feature of the retinal images generated by the GAN and VAE is given in Fig. 6. As can be seen in the figure, the similarity between the retinal images generated by the VAE and DiaRetDB1 is higher than the similarity between the retinal images generated by the GAN and DiaRetDB1 in terms of the mean, variance, skewness and kurtosis. This can be explained by the varying colors of the retinal images generated by the GAN, which is opposite of the retinal images in DiaRetDB1 as it contains mostly retinal images with the dominant red channel. However, in the case of entropy values, the results are quite different from the aforementioned cases as seen in Fig. 6(i-j). It can be explained by the fact that the pixel values in each individual image of each set have low entropy because of the uniformity in the colors. This causes the entropy values between each data set to have similar values, thereby high similarity between entropy values.

In addition to the figures, related figures for the other five subsets from the The Kaggle EyePACS can be accessed together with the histogram analysis of each statistical feature of each set on Github². The subsets are chosen to cross validate variations within The Kaggle EyePACS data set.

4 Discussion

Availability of synthetic retinal images is crucial for further development and validation of retinal image analysis methods. To generate retinal images from existing retinal image sources, methods based on deep learning including deep generative models offer potential options.

As an application of the deep generative models, a GAN and a VAE were chosen to generate synthetic retinal images. During the training process of both the GAN and VAE, we explored the following outcomes for the retinal image generation: 1) The overall structure of the retina is successfully generated by applying the GAN and VAE. However, neither one of them is able to model

² The source code and the materials: <https://github.com/kaplansinan/MasterThesis>

the vessel tree structure clearly. 2) The GAN generates the retinal images with distinctive colors as in the training set whereas the VAE captures only the dominant red channel. This is because of the constraints in the VAE in which the data is generated from the normal distribution. 3) In the context of the qualitative analysis of the generated retinal images, the GAN generates sharp retinal images, whereas the VAE generates blurry images because of the mean squared error used to compute the pixel-to-pixel distance between the generated and the actual retinal images.

As the GAN and VAE generate retinal images with a global structure, the image quality evaluation was based on the global statistical features including the mean, variance, kurtosis, skewness, and entropy. The quantitative analysis of the generated images by the proposed similarity based quality assessment method reveals the following outcome: the similarity between the generated retinal images by the VAE and the benchmark data set is higher than the generated retinal images by the GAN and randomly chosen subsets from the Kaggle EyePACS set. This is because of the fact that the VAE often generates retinal images with the dominant red channel.

By considering the quality of the generated retinal images from the application perspective, using these images can be considered as introducing noise to a supervised learning algorithm for lesion detection and image segmentation. Another possible scenario where these images can be applied is that they can be used for generating vessel tree structures. To do so, one can use these generated images as an additional conditioning input information.

The recent studies [17, 7, 25] have shown the potential of combining GANs and VAEs together to synthesize images. Therefore, the retinal image synthesis can be studied to see whether the combination of a GAN and a VAE is capable of generating the vascular tree structure in an unconditioned way. Also, GANs and VAEs can be applied to retinal image generation tasks by conditioning the models on features of retinal images specific for the subsequent use.

5 Conclusion

This paper investigates the utilization of deep generative models for the retinal image generation in an unconditioned manner. In the experimental analysis, we showed that the overall architecture of the retina can be synthesized by applying unconditioned deep generative models. However, the studied deep generative models were unable to generate detailed structures of the retina such as the vascular tree. The proposed quality assessment method revealed that the overall similarity between the retinal images generated by the VAE and the benchmark set (DiaRetDB1) was higher than the similarity between the retinal images generated by the GAN and the benchmark set (DiaRetDB1). The generated images reveal that without conditioning on any additional data, the studied deep generative models are only able to synthesize retinal images with the dominant visible parts in the retina like the fovea and the optic disk. This shows the lack of gen-

eralization capability of these models and their limited learning process in which they tend to learn only the dominant features in the data.

References

1. Diabetic retinopathy detection — kaggle (May 2017), <https://www.kaggle.com/c/diabetic-retinopathy-detection>
2. Andreini, P., Bonechi, S., Bianchini, M., Mecocci, A., Scarselli, F., Sodi, A.: A Two Stage GAN for High Resolution Retinal Image Generation and Segmentation. Tech. rep. (Jul 2019), <https://arxiv.org/abs/1907.12296v1>
3. Bonaldi, L., Menti, E., Ballerini, L., Ruggeri, A., Trucco, E.: Automatic generation of synthetic retinal fundus images: Vascular network. *Procedia Computer Science* **90**, 54–60 (2016)
4. Borji, A.: Pros and cons of gan evaluation measures. arXiv preprint arXiv:1802.03446 (2018)
5. Burlina, P.M., Joshi, N., Pacheco, K.D., Liu, T.Y.A., Bressler, N.M.: Assessment of Deep Generative Models for High-Resolution Synthetic Retinal Image Generation of Age-Related Macular Degeneration. *JAMA Ophthalmology* **137**(3), 258–264 (Mar 2019). <https://doi.org/10.1001/jamaophthalmol.2018.6156>, <https://jamanetwork.com/journals/jamaophthalmology/fullarticle/2720489>
6. Burt, P., Adelson, E.: The laplacian pyramid as a compact image code. *IEEE Transactions on Communications* **31**(4), 532–540 (1983)
7. Choi, E., Biswal, S., Malin, B., Duke, J., Stewart, W.F., Sun, J.: Generating multi-label discrete electronic health records using generative adversarial networks. arXiv preprint arXiv:1703.06490 (2017)
8. Costa, P., Galdran, A., Meyer, M.I., Abràmoff, M.D., Niemeijer, M., Mendonça, A.M., Campilho, A.: Towards adversarial retinal image synthesis. arXiv preprint arXiv:1701.08974 (2017)
9. Davis, H., Russell, S., Barriga, E., Abramoff, M., Soliz, P.: Vision-based, real-time retinal image quality assessment. In: *Computer-Based Medical Systems, 2009. CBMS 2009. 22nd IEEE International Symposium on*. pp. 1–6. IEEE (2009)
10. Deng, L., Yu, D., et al.: Deep learning: methods and applications. *Foundations and Trends® in Signal Processing* **7**(3–4), 197–387 (2014)
11. Denton, E.L., Chintala, S., Fergus, R., et al.: Deep generative image models using a laplacian pyramid of adversarial networks. In: *Advances in Neural Information Processing Systems*. pp. 1486–1494 (2015)
12. Fang, L., Li, S., McNabb, R.P., Nie, Q., Kuo, A.N., Toth, C.A., Izatt, J.A., Farsiu, S.: Fast acquisition and reconstruction of optical coherence tomography images via sparse representation. *IEEE Transactions on Medical Imaging* **32**(11), 2034–2049 (2013)
13. Fasih, M.: Retinal Image Quality Assessment Using Supervised Classification. Ph.D. thesis, École Polytechnique de Montréal (2014)
14. Fiorini, S., Ballerini, L., Trucco, E., Ruggeri, A.: Automatic generation of synthetic retinal fundus images. In: *Eurographics Italian Chapter Conference*. pp. 41–44 (2014)
15. Gauthier, J.: Conditional generative adversarial nets for convolutional face generation. Class Project for Stanford CS231N: Convolutional Neural Networks for Visual Recognition, Winter semester **2014**, 5 (2014)

16. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: *Advances in Neural Information Processing Systems*. pp. 2672–2680 (2014)
17. Gorijsala, M., Dukkipati, A.: Image generation and editing with variational info generative adversarial networks. arXiv preprint arXiv:1701.04568 (2017)
18. Hoff, P.D., Raftery, A.E., Handcock, M.S.: Latent space approaches to social network analysis. *Journal of the American Statistical Association* **97**(460), 1090–1098 (2002)
19. Kaplan, S.: Deep generative models for synthetic retinal image generation. Master's thesis, Lappeenranta University of Technology (2017), <http://urn.fi/URN:NBN:fi-fe201708047855>
20. Kauppi, T., Kalesnykiene, V., Kamarainen, J.K., Lensu, L., Sorri, I., Pietila, J., Kalviainen, H., Uusitalo, H.: Diaretdb1-standard diabetic retino-pathology database. IMAGERET Optimal Detection and Decision-Support Diagnosis of Diabetic Retinopathy (2007)
21. Keane, P.A., Sadda, S.R.: Retinal imaging in the twenty-first century: state of the art and future directions. *Ophthalmology* **121**(12), 2489–2500 (2014)
22. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013)
23. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**(7553), 436–444 (2015)
24. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al.: Photo-realistic single image super-resolution using a generative adversarial network. arXiv preprint arXiv:1609.04802 (2016)
25. Makhzani, A., Shlens, J., Jaitly, N., Goodfellow, I., Frey, B.: Adversarial autoencoders. arXiv preprint arXiv:1511.05644 (2015)
26. Patton, N., Aslam, T.M., MacGillivray, T., Deary, I.J., Dhillon, B., Eikelboom, R.H., Yogesan, K., Constable, I.J.: Retinal image analysis: Concepts, applications and potential. *Progress in Retinal and Eye Research* **25**(1), 99–127 (Jan 2006). <https://doi.org/10.1016/j.preteyeres.2005.07.001>, <https://linkinghub.elsevier.com/retrieve/pii/S1350946205000406>, 00587
27. Pu, Y., Gan, Z., Henao, R., Yuan, X., Li, C., Stevens, A., Carin, L.: Variational autoencoder for deep learning of images, labels and captions. In: *Advances in Neural Information Processing Systems*. pp. 2352–2360 (2016)
28. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434 (2015)
29. Rezende, D.J., Mohamed, S., Wierstra, D.: Stochastic backpropagation and approximate inference in deep generative models. arXiv preprint arXiv:1401.4082 (2014)
30. Theis, L., Oord, A.v.d., Bethge, M.: A note on the evaluation of generative models. arXiv preprint arXiv:1511.01844 (2015)
31. Vertolli, M.O., Davies, J.: Image quality assessment techniques show improved training and evaluation of autoencoder generative adversarial networks. arXiv preprint arXiv:1708.02237 (2017)
32. Yin, W., Fu, Y., Sigal, L., Xue, X.: Semi-latent gan: Learning to generate and modify facial images from attributes. arXiv preprint arXiv:1704.02166 (2017)

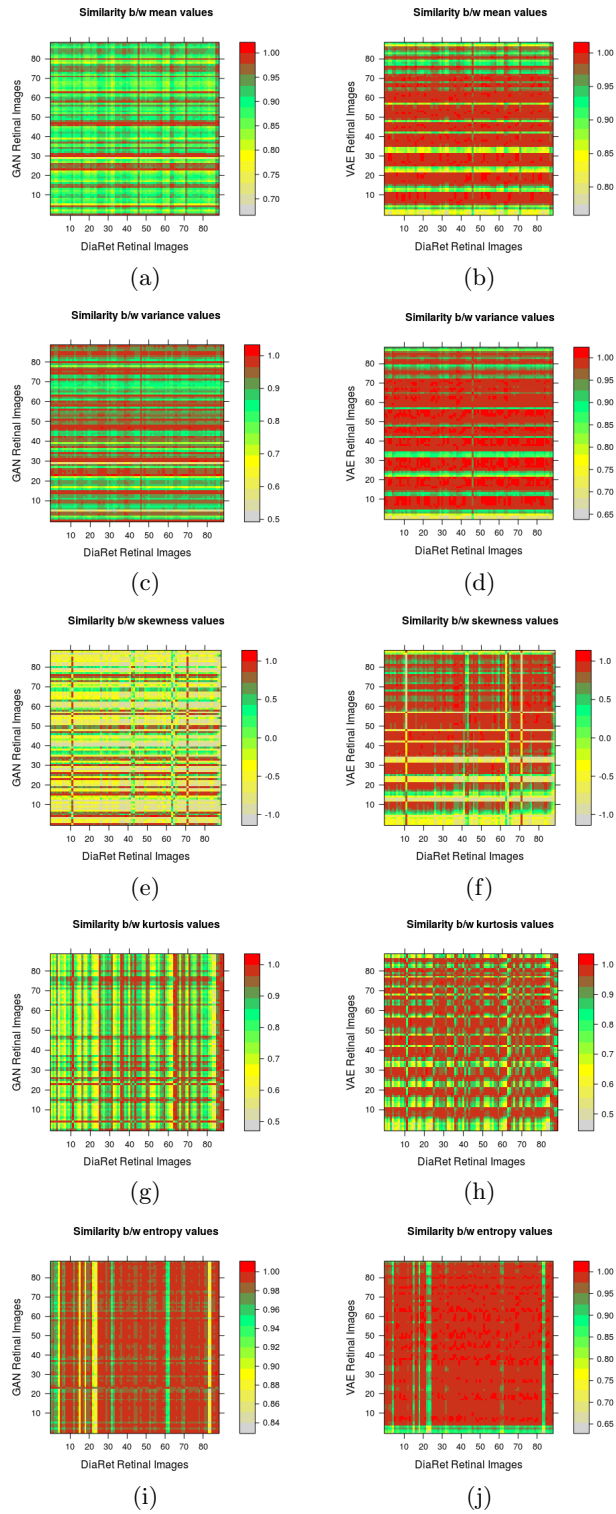


Fig. 6. Comparison of similarity measurements b/w the GAN and DiaRetDB1 and VAE: (a -b) Mean; (c-d) Variance; (e-f) Skewness; (g-h) Kurtosis and (i-j) Entropy.