Retinal Image Synthesis for CAD development *

Appan K. Pujitha and Jayanthi Sivaswamy

Center for Visual Information Technology, IIIT Hyderabad, India.
pujitha.ak@research.iiit.ac.in and jsivaswamy@iiit.ac.in

Abstract. Automatic disease detection and classification have been attracting much interest. High performance is critical in adoption of such systems, which generally rely on training with a wide variety of annotated data. Availability of such varied annotated data in medical imaging is very scarce. Synthetic data generation is a promising solution to address this problem. We propose a novel method, based on generative adversarial networks (GAN), to generate images with lesions such that the overall severity level can be controlled. We demonstrate the reliability of the generated synthetic images independently as well as by training a computer aided diagnosis (CAD) system with the generated data. We showcase this approach for heamorrhage detection in retinal images with 4 levels of severity. Quantitative assessment results show that the generated synthetic images are very close to the real data. Haemorrhage detection was found to improve with inclusion of synthetic data in the training set with improvements in sensitivity ranging from 20% to 27% over training with just expert marked data.

Keywords: Synthetic Images, Generative Adversarial Networks, Deep Neural Net.

1 Introduction

Generation of synthetic medical data is aimed at addressing a range of needs. Early examples are generating digital brain phantoms [1] and synthesizing a whole retinal image [11] using complex modeling. These were aimed at aiding the development of algorithms for denoising, reconstruction and segmentation. Recently simulation of brain tumors in MR images [13] has also been explored to aid CAD algorithm development. With the advent of deep learning, modeling of complex structures and synthesizing images has become easier with a class of neural networks called generative adversarial networks or GAN [4].

GAN is an architecture composed of two networks, namely, a generator and a discriminator. Functionally, the generator synthesizes images from noise while the discriminator differentiates between real and synthetic images. GAN have recently been explored for a variety of applications: detection of brain lesions [15], predicting CT from MRI images [12], synthesizing normal retinal images from vessel mask [2], segmenting anatomical structures such as vessels [6] and optic disc/cup [17].

* This work was supported by the Dept. of Electronics and Information Technology, Govt. of India under Grant: DeitY/R&D/TDC/13(8)/2013.

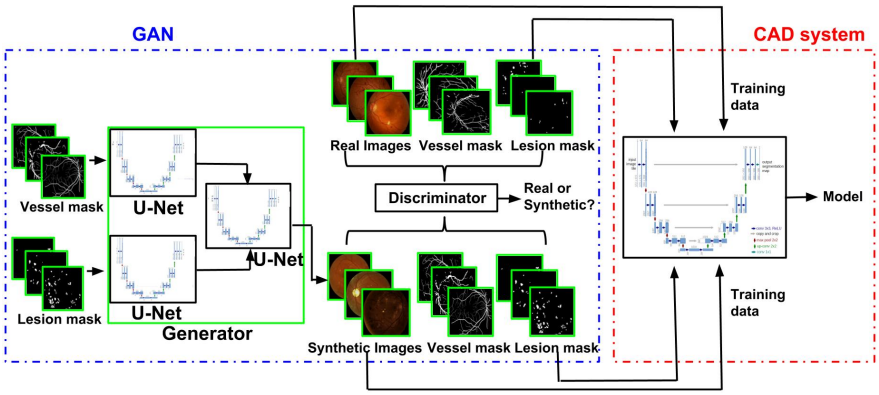

Fig. 1: Proposed end-to-end pipeline for generation of abnormal retinal images and its use in a CAD system for detection of haemorrhages.

We propose a GAN for generating images with pathologies in a *controlled manner* and illustrate how the generated synthetic images can be used to address the data sparsity problem which hampers the development of robust CAD solutions for abnormality detection. We choose *staging* of diabetic retinopathy (DR) from given color retinal images as a case study. The ETDRS standard for staging of DR is based on the number and location of haemorrhages [19]. However, very few images are publicly available with local markings of haemorrhages. Recent deep learning-based methods [5], [20] overcome this problem by sampling a large public dataset (with only image-level annotations) to get local annotations for a much smaller subset of images which are abnormal. These annotations are privately held and hence such measures are not beneficial to a wide community for building a robust CAD solution.

2 Method

The proposed method consists of three modules: (i) pre-processing, (ii) synthetic image generation and (iii) CAD for haemorrhage detection. As a part of the pre-processing step, given retinal images are corrected for non-uniform illumination using luminosity and contrast normalization [7].

2.1 GAN for Synthesis of Retinal Images with Pathologies

Generating *normal* retinal images from vessel mask has been attempted earlier [2] with a single U-net for the generator and a 5-layer convolutional neural network for the discriminator. Our interest is in generating images with haemmorages (HE) towards synthesis of exemplars for different stages of DR. HE are often indistinguishable from vessel fragments and therefore the input to the generator has to enable distinguishing between these both structures. Further, exemplar generation requires gaining control of the locations, size and density of HE. Hence, we propose a GAN architecture (shown in Fig.1) with a generator consisting of two parallel networks: one with a vessel mask as input and another

with a lesion mask as input. The output of the networks, based on the U-net architectures, are merged and fed to a third U-net architecture which generates the whole retinal image with lesions. The generator thus maps from vessel (v_i) and lesion (l_i) masks to a retinal image (r_i). A 5-layer convolutional neural network as in [2] is used for the discriminator to distinguish between the real and synthetic sets of images, with each set consisting of vessel and lesion masks along with retinal images.

The GAN learns a model as follows: the discriminator iteratively reduces its misclassification error by more accurately classifying the real and synthetic images while the generator aims to deceive the discriminator by producing more realistic images. The overall loss function that is to be optimized is chosen as a weighted combination of 3 loss functions: L_{adv}, L_{SSIM} and L_1 as defined below in eq.1-4 to produce sharp and realistic images (here, G and D, represent generator and discriminator respectively).

(i) The adversarial loss function L_{adv} is defined as

$$\begin{aligned} L_{adv}(G,D) = \mathbb{E}_{(v,l),r\sim p_{data}((v,l),r)}[log(D((v,l),r))] \\ + \mathbb{E}_{v,l\sim p_{data}(v,l)}[log(1-D((v,l),G(v,l)))] \end{aligned} \tag{1}$$

where $\mathbb{E}_{(v,l),r\sim p_{data}}$ represents the expectation of the log-likelihood of the pair $((v,l),r)$ being sampled from the underlying probability distribution of real pairs $p_{data}((v,l),r)$, while $p_{data}(v,l)$ is the distribution of real vessel and lesion masks.

(ii) The Structure Similarity (SSIM) [18] index is useful in quantitatively measuring the structural similarity between two images $(r, G(v,l))$. It also has been shown to perform well for reconstruction and generation of visually pleasing images.

$$SSIM(p) = \frac{2\mu_r\mu_{G(v,l)} + C_1}{\mu_r^2 + \mu_{G(v,l)}^2 + C_1} \cdot \frac{2\sigma_{rG(v,l)} + C_2}{\sigma_r^2 + \sigma_{G(v,l)}^2 + C_2} \tag{2}$$

where (μ_r, $\mu_{G(v,l)}$) and (σ_r,$\sigma_{G(v,l)}$) are the means and standard deviation computed over patch centered on pixel p, C_1 and C_2 are constants. The loss L_{SSIM} can be computed as:

$$L_{SSIM} = 1 - \frac{1}{N}\sum_{p\epsilon P} SSIM(\tilde{p}) \tag{3}$$

where $\tilde{p}$ is the center pixel of a patch P in the image I.

(iii) The loss function L_1 is used mainly to reduce artifacts and blurring and is defined as

$$L_1 = \mathbb{E}_{(v,l),r\sim p_{data}((v,l),r)}(\|r - G(v,l)\|_1) \tag{4}$$

The overall loss function to be minimized is taken to be

$$L(G,D) = L_{adv} + \lambda_1 L_1 + \lambda_2 L_{SSIM} \tag{5}$$

where λ_1 and λ_2 control the contribution of the L_1 and L_{SSIM} loss functions respectively.

2.2 CAD for Haemorrhage Detection

We chose the U-Net [16] to build a CAD solution for detection of HE (referred to as CADH). This is used to demonstrate that the synthetic images (generated by our proposed GAN) are a reliable resource in training the U-net. The U-net architecture consists of a contracting and an expansive path. The contracting path is similar to a typical CNN architecture, whereas in the expanding path, max-pooling is replaced by up-sampling. There are skip connections between contracting and expanding paths to ensure localization. The U-net is modified in terms of the number of filters at each convolutional layer and the loss function. The number of filters at each stage is reduced to half to simplify computations. The loss is modified to account for the misclassification of lesions.

The U-net architecture provides the segmentation of HE. The segmented HE are counted and the image is classified into the respective grade accordingly (as given in section 3.1: training data for CADH).

3 Implementation

3.1 Datasets

Both GAN and CADH were trained on *pathological* images. These are drawn from DRiDB [14] (31 images) and a locally sourced dataset denoted as *LoD* (58 images). Testing of CADH was done at i) lesion level on 40 pathological images from DIARETDB1 [8] and ii) at a stage-level on 308 abnormal images + 892 normal images (without HE) from MESSIDOR [3].

Lesion markings are available for DIARETDB1 from four experts, while for DRIDB and *LoD* it is from one expert. The consensus of 3 experts was considered to derive a binary mask for DIARETDB1. The ground truth of all the three datasets were overlapped with the respective images and thresholded to get a pixel-level lesion mask. The vessel masks, whenever unavailable were derived using method described in [10]. Images from all datasets was cropped and resized 512x512 before feeding them to GAN or CADH.

Training Data for GAN Training of the GAN requires both lesion and vessel masks. the lesion masks for the training data are available, but vessels masks are available only for DRiDB. It is tedious and time consuming to mark the vessels in each of the retinal images. Hence, vessel masks were derived using a method [10] which has proved to perform relatively well for vessel segmentation even in the presence of pathologies.

Training Data for CADH For training the CADH, a heterogeneous mixture of data were combined, namely, expert annotated data and synthetic data and augmented data. The DRiDB and *LoD* datasets were sources of expert annotated data. Augmented data was derived by applying random transformations to the images. This included random rotation between -25^0 to 25^0, random translation in vertical / horizontal directions in the range of 50 pixels, and random horizontal / vertical flips.

Finally, the synthetic retinal images were generated using GAN as follows. The vessel and lesion masks were taken randomly from *LoD* and DRiDB. The

lesion masks were modified using the same random transformations such as flipping the lesions sector wise, flipping horizontally and vertically, rotations and translations. Retinal images containing HE are graded with severity levels as in [3]: grade 0/1 (no HE), grade 2 (1-5 HE) and grade 3 (more than 5 HE). The lesions masks were derived to provide exemplars for each level using these rules. The number of lesions in each category were maintained by masking out few lesions or adding new lesions from another lesion mask randomly. Fig. 2 shows a sample of the vessel, lesion masks and generated synthetic images at grade 2 and 3 severity levels.

3.2 Computing Details

The models were implemented in Python using Keras with Theano as backend and trained on a NVIDIA GTX 970 GPU, 4GB RAM. Training was done with random initialized weights for 2000 epochs by minimizing the loss functions described in Section 2.1 using Adam optimizer. For model parameters, learning rate was initialized to 2×10^{-4} for GAN and 1×10^{-5} for CADH. A batch size of 4 was considered for both cases and other parameters were left at default values. Class weights were outlined as the inverse ratio of the number of positive samples to negative samples and modified empirically.

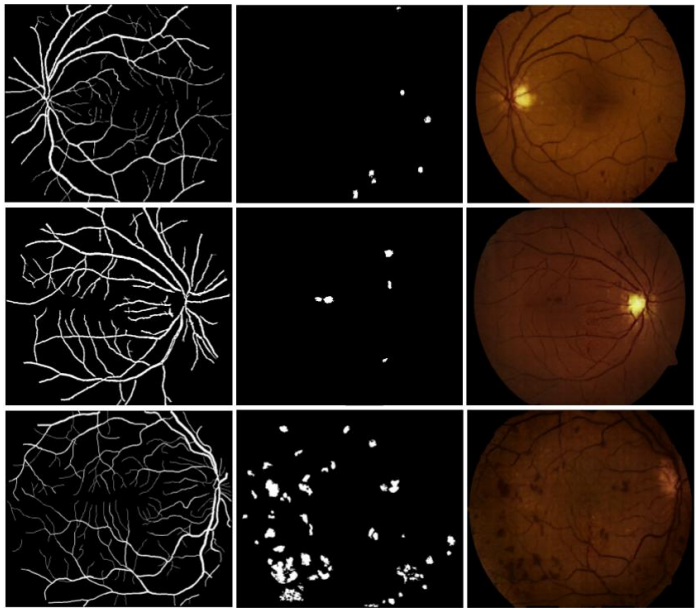

Fig. 2: Vessel, lesion masks and synthetic images generated at grade 2 (first two rows) and grade 3 (last row) levels.

3.3 Evaluation Metrics

The synthetically generated images were evaluated quantitatively and qualitatively. The mean and standard deviation of the Q_v score described in [9] was computed over all images (40 abnormal) in DIARETDB1.

The performance of CADH was evaluated using Sensitivity (SN) and Positive Predictive Value (PPV) which are defined as follows: $SN = \frac{TP}{TP+FN}$ and $PPV = \frac{TP}{TP+FP}$. To evaluate against the given local annotations by experts, the pixel wise classification was converted to region wise detection by applying connected component analysis and requiring at least 50% (but not exceeding more than 150%) overlap with manually marked regions to identify true positive detections (TP); else it is false positive (FP). If a region is marked by the expert but was not detected by the model then it is a False negative (FN). The area under the SN vs PPV curve (AUC) is also taken as a measure of performance.

4 Experiments and Results

4.1 Synthetic Image Generation (GAN)

Fig.3 shows two sample synthetic retinal images (containing HE) generated by the proposed GAN model. The first two columns show the vessel and lesion masks given as input to the GAN. Third and fourth columns show the synthetic and the corresponding real images, respectively. The synthetic images appear realistic yet differ from the real images in terms of background color, texture and illumination. Lesion locations are roughly similar but sizes are different as lesion masks are not results of exact segmentations of lesions.

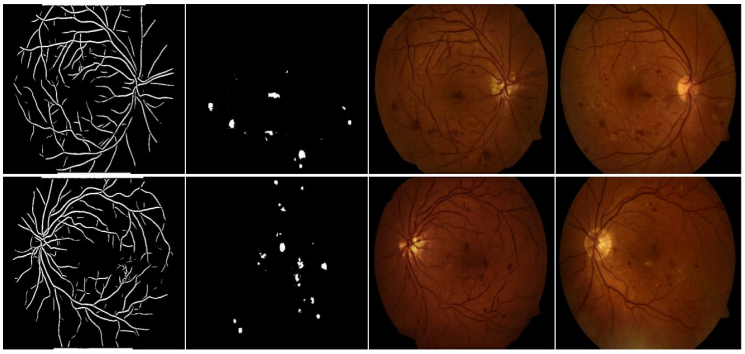

Fig. 3: Results of GAN-based image synthesis. From left to right: vessel mask, lesion mask, synthetic image and corresponding real image.

The mean/ standard deviation of Q_v computed over all images with pathologies in DIARETDB1 is 0.0516/ 0.0144 and over all the synthetic images generated from vessel and lesion mask from DIARETDB1 is 0.0675/0.0239. The Q_v score is higher for images of greater quality, this indicates synthetic images are considered better as they contain less noise.

4.2 CAD for Haemorrhage Detection (CADH)

The utility of the synthetic data for CAD development was tested by training 4 different CADH models by varying the training set content. Denoting the set of real images with expert annotations as E and the set of synthetic images

generated by GAN with the corresponding lesion masks as S, the variants of the training set considered are: (i) only E, (ii) E with data augmentation (E+A), (iii) E and S, (iv) E, S with data augmentation (E+S+A). The computed SN at a fixed PPV and AUC values for these variants are reported in Table. 1. The SN vs PPV curve is shown in Fig. 4.

The tabulated results indicate that addition of synthetic data (E+S) boosts SN by 20.4% and 10% over E and E+A, respectively. The full set of E+S+A yields the best performance with an improvement (over E) in SN by 27.4% and AUC by 19%. This establishes the effectiveness of synthetic data in general and in CAD development. Increasing the number of synthetic images serves to improve the performance (row 5). In order to assess if synthetically derived data has artifacts, the E+S+A variant was tested on an exclusive set of separately generated synthetic images using vessel and lesion masks of DIARETDB1. The obtained results (row 6) shows a minor degradation over that for real images (row 4), implying the generated data is free of artifacts. A recent fast CNN

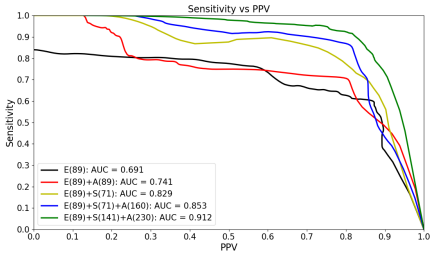

Fig. 4: SN vs PPV curve for CADH.

Table 1: CADH performance on DIARETDB1.

	Training data (# images)	SN (%)	PPV (%)	AUC
1.	E (89)	63.1	79.4	0.691
2.	E (89) + A (89)	70.6	79.6	0.741
3.	E (89) + S (71)	79.3	79.6	0.829
4.	E (89) + S (71)+ A (160)	86.9	79.8	0.853
5.	E (89) + S (141)+ A (230)	92.7	79.4	0.912
6.	E (89)+ S (71)+ A (160) *	84.2	80	0.834

* detection performance on only synthetic images.

method [5] for binary (HE/no HE) classification used local markings by experts and report a SN/SP on the MESSIDOR dataset of 91.9/ 91.4% which is lower than 94/ 91.7% achieved by CADH trained on E+S(141)+A. Since this dataset provides severity grades for each image, testing of CADH model was done at gradelevel and the obtained values for SN/ SP are 92/ 94.4% for grade 2 and 89.4/ 90.1% for grade 3. This indicates that the model trained with synthetic data (can be generated in abundance) is better than that trained with expert annotated data (which is difficult to obtain).

5 Concluding Remarks

We proposed a novel solution to develop retinal images with HE using generative adversarial networks. The network is trained to generate the retinal image using vessel and lesion masks. Hence, we can develop retinal image with any type of severity, by providing the corresponding lesion mask. The synthetic abnormal images generated are shown to be realistic in the type of lesions produced and also the color, texture using the Q_v metric. These generated images are valuable in developing a CAD system which detects and localizes haemorrhages as addition of synthetic data led to improvement in both SN and AUC by 17%. Our proposed approach can be extended to other lesions (such as hard exudates) and image modalities and thus has a wide potential.

References

1. Collins, D.L., et al.: Design and construction of a realistic digital brain phantom. IEEE Transactions on Medical Imaging 17 3, 463–8 (1998)
2. Costa, P., et al.: End-to-end adversarial retinal image synthesis. IEEE Transactions on Medical Imaging 37 3, 781 – 791 (2017)
3. Decenciere, E., et al.: Feedback on a publicly distributed database: the messidor database. Image Analysis & Stereology 33, 231–234 (Aug 2014)
4. Goodfellow, et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems 27, pp. 2672–2680 (2014)
5. van Grinsven, et al.: Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images. IEEE Trans. Med. Imaging 35, 1273–1284 (2016)
6. Guibas, J.T., et al.: Synthetic Medical Images from Dual Generative Adversarial Networks. ArXiv e-prints (Sep 2017)
7. Joshi, G.D., et al.: Colour retinal image enhancement based on domain knowledge. In: Computer Vision, Graphics Image Processing, ICVGIP. pp. 591–598 (Dec 2008)
8. Kalesnykiene, V., et al.: Diaretdb1 diabetic retinopathy database and evaluation protocol (Jun 2007)
9. Kohler, T., et al.: Automatic no-reference quality assessment for retinal fundus images using vessel segmentation. International Symposium on Computer-Based Medical Systems, CBMS 00, 95–100 (2013)
10. Maninis, K., et al.: Deep retinal image understanding. In: Medical Image Computing and Computer-Assisted Intervention, MICCAI. pp. 140–148 (2016)
11. Menti, et al.: Automatic generation of synthetic retinal fundus images: Vascular network. In: Simulation and Synthesis in Medical Imaging: SASHIMI, Held in Conjunction with MICCAI. pp. 167–176 (Oct 2016)
12. Nie, D., et al.: Medical image synthesis with context-aware generative adversarial networks. In: Medical Image Computing and Computer-Assisted Intervention, MICCAI. pp. 417–425 (Sept 2017)
13. Prastawa, M., et al.: Simulation of brain tumors in mr images for evaluation of segmentation efficacy. Medical Image Analysis 13, 297 – 311 (2009)
14. Prentasic, P., et al.: Diabetic retinopathy image database (dridb): A new database for diabetic retinopathy screening programs research. In: Image and Signal Processing and Analysis, ISPA. pp. 704–709 (2013)
15. Rezaei, M., et al.: Conditional adversarial network for semantic segmentation of brain tumor. CoRR abs/1708.05227 (Aug 2017)
16. Ronneberger, et al.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention, MICCAI. pp. 234–241 (2015)
17. Shankaranarayana, et al.: Joint optic disc and cup segmentation using fully convolutional and adversarial networks. In: Fetal, Infant and Ophthalmic Medical Image Analysis: OMIA Held in Conjunction with MICCAI. pp. 168–176 (2017)
18. Wang, Z., et al.: Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing 13(4), 600–612 (2004)
19. Wilkinson, C.P., et al.: Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. Ophthalmology 110(9), 1677–82 (2003)
20. Yang, Y., et al.: Lesion detection and grading of diabetic retinopathy via two-stages deep convolutional neural networks. CoRR abs/1705.00771 (2017)