Generative Adversarial Learning for Reducing Manual Annotation in Semantic Segmentation on Large Scale Microscopy Images

•••

Kumar Ayush
B. Tech (Hons.) - Computer Science and Engineering

Advisor
Dr. Pabitra Mitra

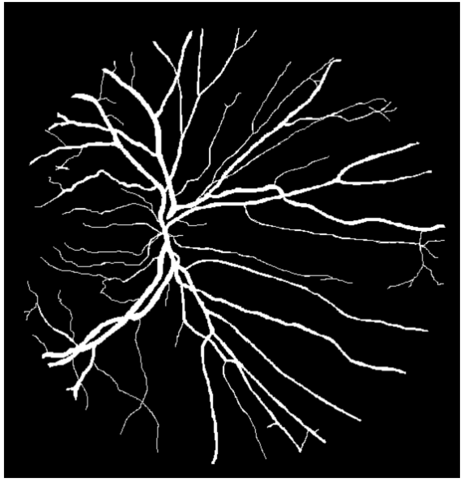
Motivation

Labeling Data

Unlabeled Data

Cheap and Abundant!

Human Expert/Special Equipment/Experiment



Labeled Data

Expensive and Scarce!

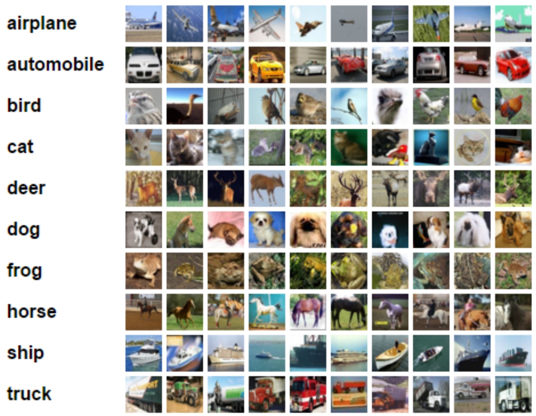
Convolutional Neural Network(CNN) based **semantic segmentation** require **extensive pixel level manual annotation** which is **daunting** for **large microscopic images**.

Success of Generative Adversarial Networks

- Radford et al., have shown convincing evidence that unsupervised training of **a deep convolutional adversarial pair** learns a **hierarchy of representations**.

- They have demonstrated the applicability of these **rich image representations** for **supervised tasks** such as **CIFAR-10 classification**.

Reference : Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. A. Radford, L. Metz, S. Chinatala. ICLR 2016



CIFAR-10

Success of Generative Adversarial Networks

- Augustus Odena, extended GANs to the **semi-supervised context** by forcing the discriminator network to **output class labels**.

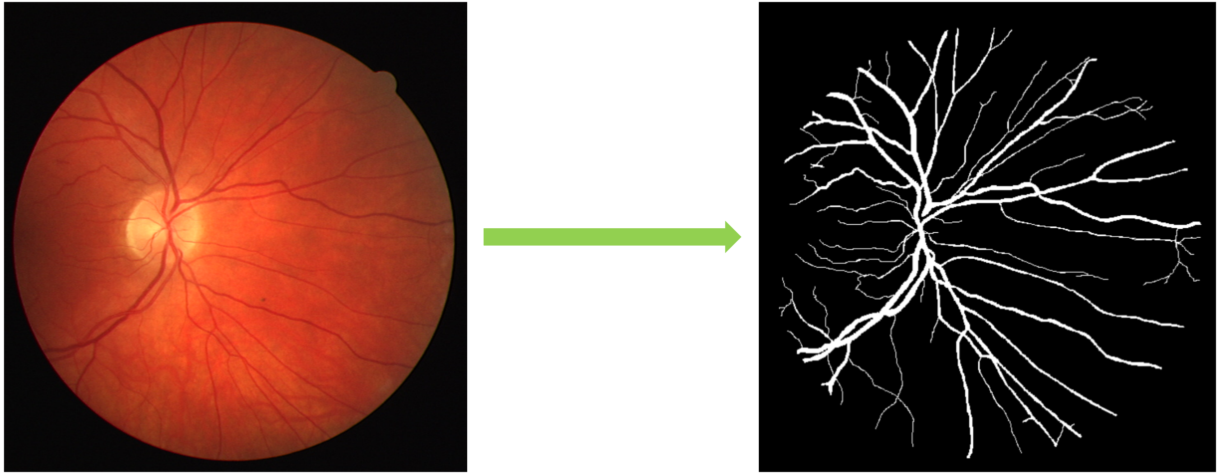
- It was shown that SGAN **improves classification performance** on restricted data sets over a baseline classifier with **no generative component**.

Reference : Semi-Supervised Learning with Generative Adversarial Networks. Augustus Odena. Data Efficient Machine Learning workshop at ICML 2016

MNIST

Objective

Generative Adversarial Learning for Reducing Manual Annotation in Semantic Segmentation on Large Scale Microscopy Images



Automated Vessel
Segmentation in Retinal Fundus Image as Test Case

Generative Adversarial Networks

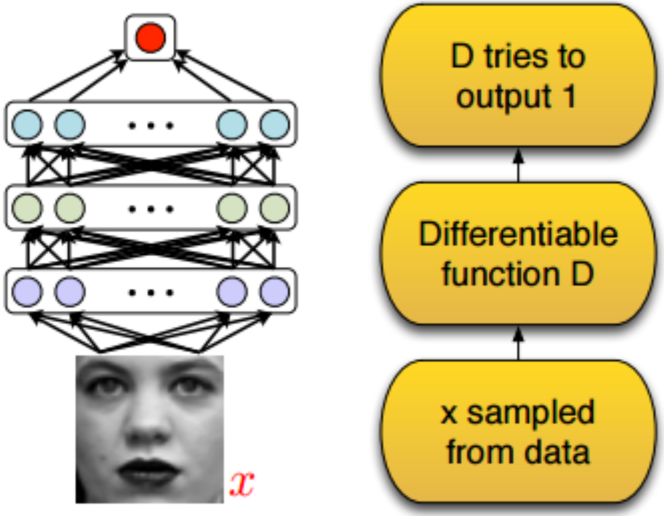
Generative Adversarial Networks

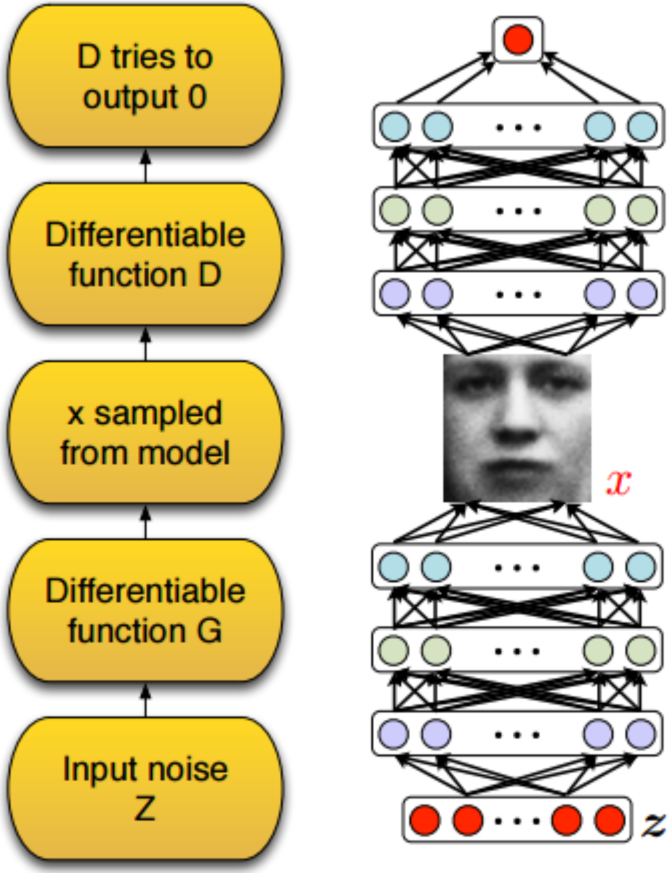
- A game between two players:
    1. Generator G
    2. Discriminator D

- D tries to discriminate between:
    - A sample from the actual data distribution.
    - And a sample from the generator G.

- G tries to "trick" D by generating samples that are hard for D to distinguish from actual data.

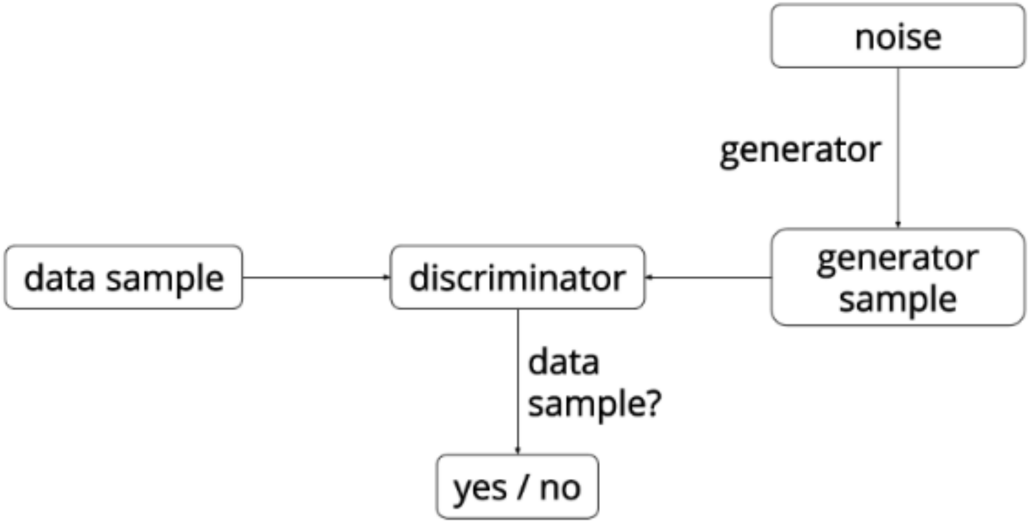
Adversarial nets Framework





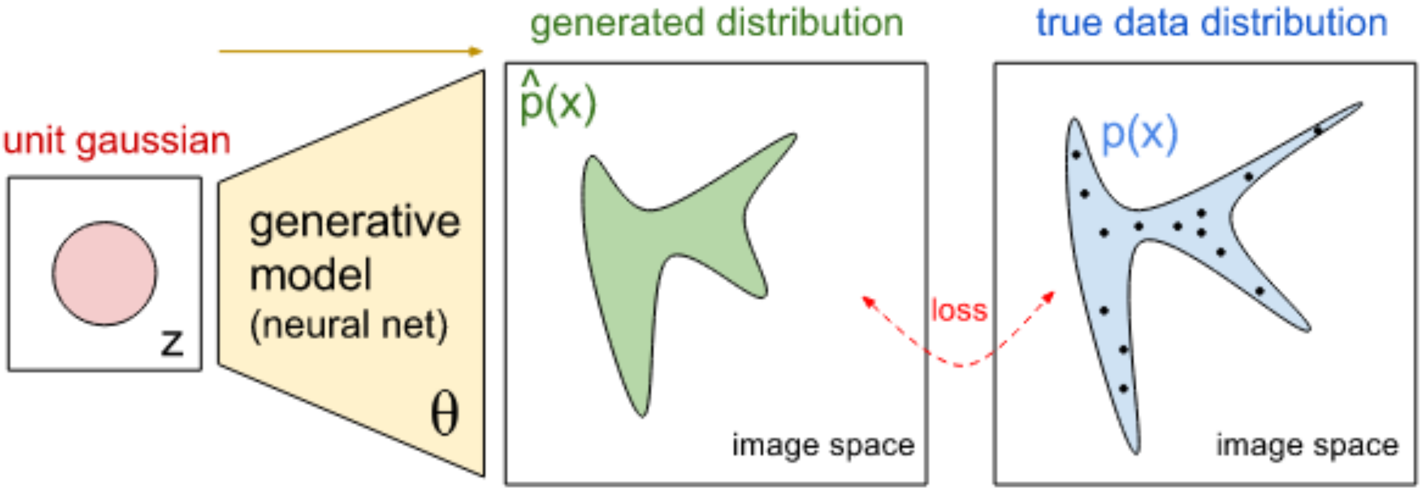
Reference : Generative Adversarial Nets. Ian Goodfellow et al. NIPS 2014

Training GANs



$$\min_G \max_D V(D,G) = \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))]$$

Training GANs

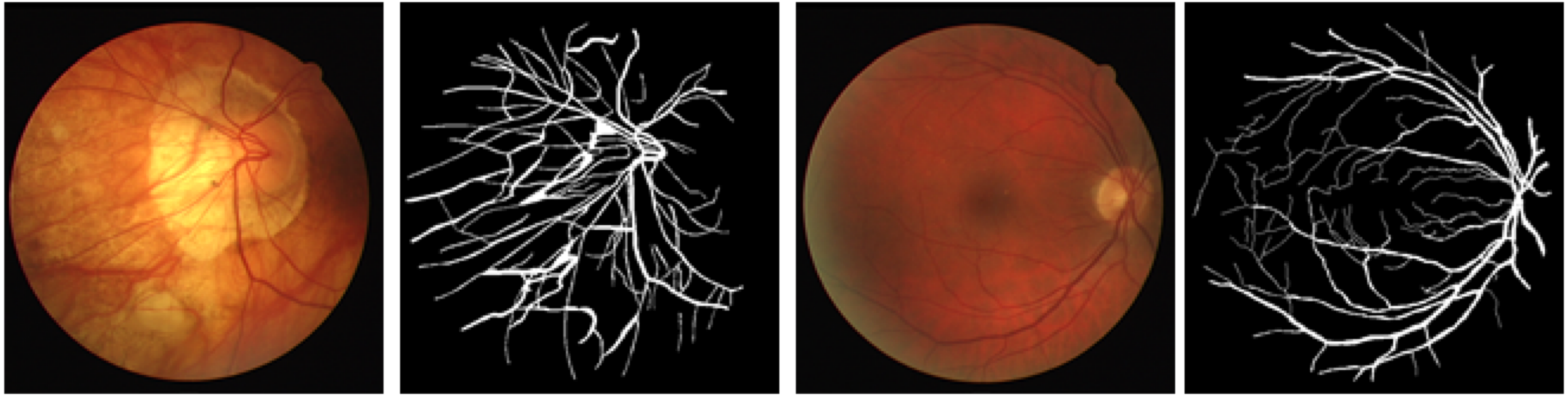


Source: https://openai.com/blog/generative-models/

Green distribution starting out random and then the training process iteratively changes the parameters θ to stretch and squeeze it to better match the blue distribution.

DATASET

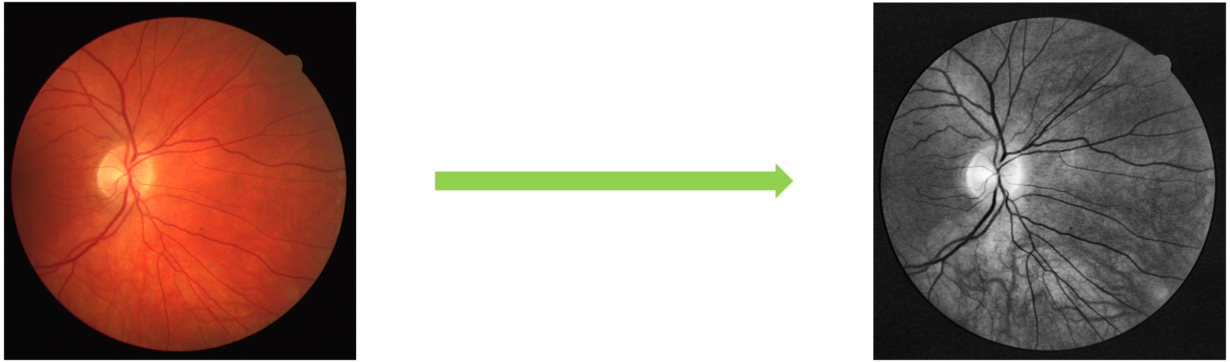
DRIVE: Digital Retinal Images for Vessel Extraction



- The dataset contains 20 images for training and 20 for testing.
- Blood vessel in each image is manually marked by human observers trained by an experienced ophthalmologist.

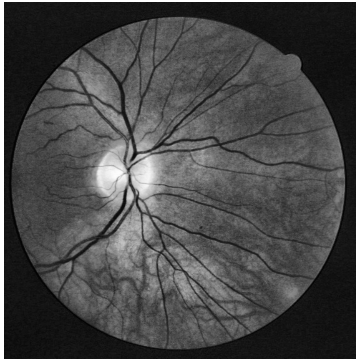
Reference: J.J. Staal, M.D. Abramoff, M. Niemeijer, M.A. Viergever, B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina", IEEE Transactions on Medical Imaging, 2004 Apr.

Preprocessing & Patch Creation



- Vascular structures manifest best contrast in **green channel**.

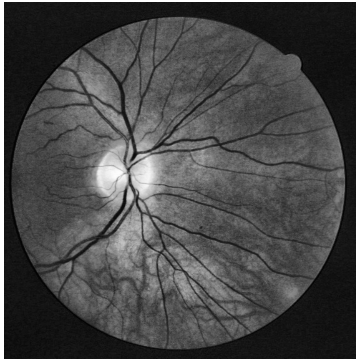
Preprocessing & Patch Creation



- Vascular structures manifest best contrast in **green channel**.

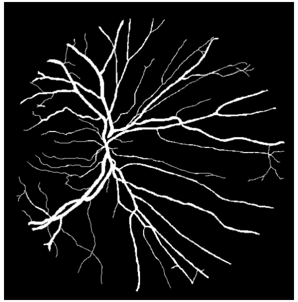
- **Contrast Limited Adaptive Histogram Equalization** (CLAHE) is used for compensating irregular illuminations.

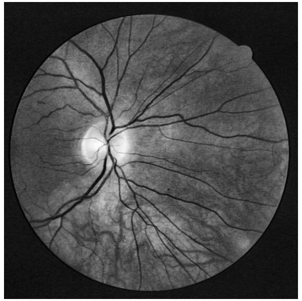
Preprocessing & Patch Creation

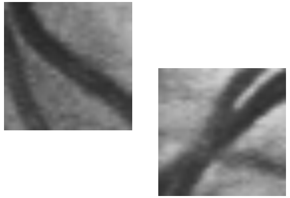


- Vascular structures manifest best contrast in **green channel**.
- **Contrast Limited Adaptive Histogram Equalization** (CLAHE) is used for compensating irregular illuminations.
- **64X64 dimensional patches** were extracted and label of central pixel is assigned as the class label of the entire patch.

Preprocessing & Patch Creation







Vessel Patches

Ground Truth Binary Image

I_{gr}

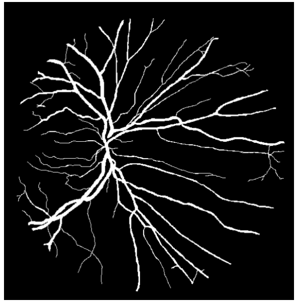
Skeletonization

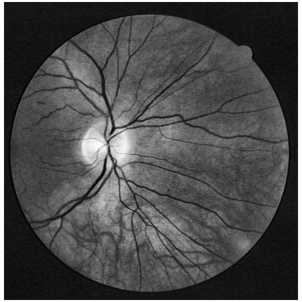
I_{gr}^{s}

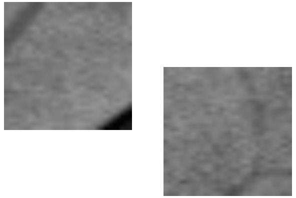
Vessel Patches from green channel image at pixels where

$I_{gr}^{s} = 1$

Preprocessing & Patch Creation







Background Patches

Ground Truth Binary Image

I_{gr}

Morphological Dilation

I_{gr}^{d}

Background Patches from green channel image at pixels where

$I_{gr}^{d} = 0$

Network Architecture

Semi-Supervised Learning with GAN

- The original version of GAN can be implemented with **2-way softmax** output from discriminator network to find a distribution over [REAL, FAKE].

Semi-Supervised Learning with GAN

- The original version of GAN can be implemented with **2-way softmax** output from discriminator network to find a distribution over [REAL, FAKE].

- We incorporate a **K+1-way softmax** layer at output of discriminator, where now the prediction labels will be **[Class 1, Class, 2,.., Class K, FAKE]**.

Semi-Supervised Learning with GAN

- The original version of GAN can be implemented with **2-way softmax** output from discriminator network to find a distribution over [REAL, FAKE].

- We incorporate a **K+1-way softmax** layer at output of discriminator, where now the prediction labels will be **[Class 1, Class, 2,.., Class K, FAKE]**.

- The revised Discriminator can be termed as a **Discriminator-Classifier** network (DC net).

Semi-Supervised Learning with GAN

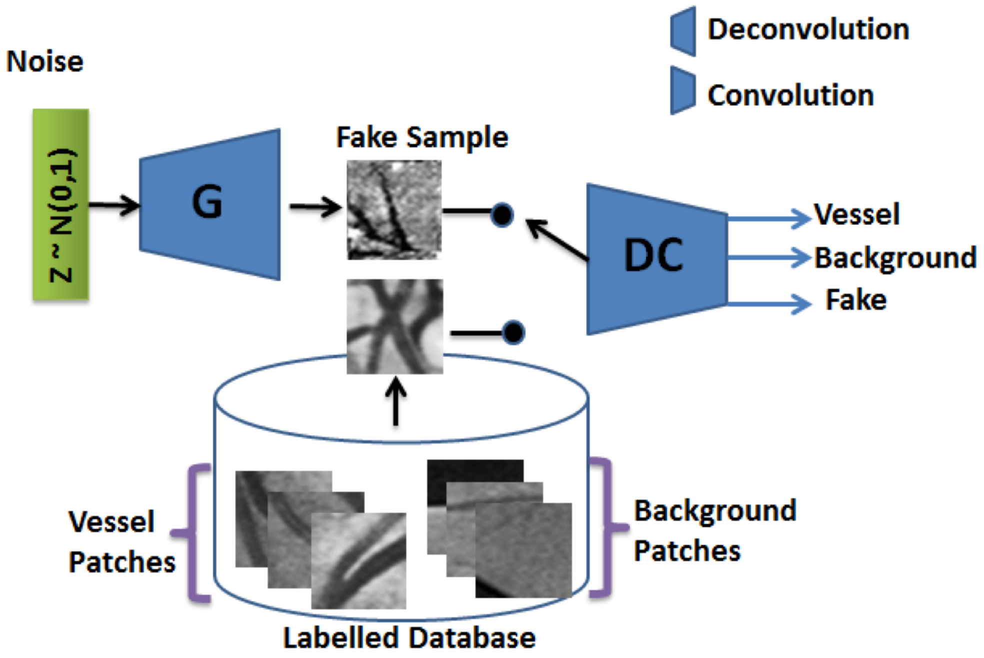
- The original version of GAN can be implemented with **2-way softmax** output from discriminator network to find a distribution over [REAL, FAKE].

- We incorporate a **K+1-way softmax** layer at output of discriminator, where now the prediction labels will be **[Class 1, Class, 2,.., Class K, FAKE]**.

- The revised Discriminator can be termed as a **Discriminator-Classifier** network (DC net).

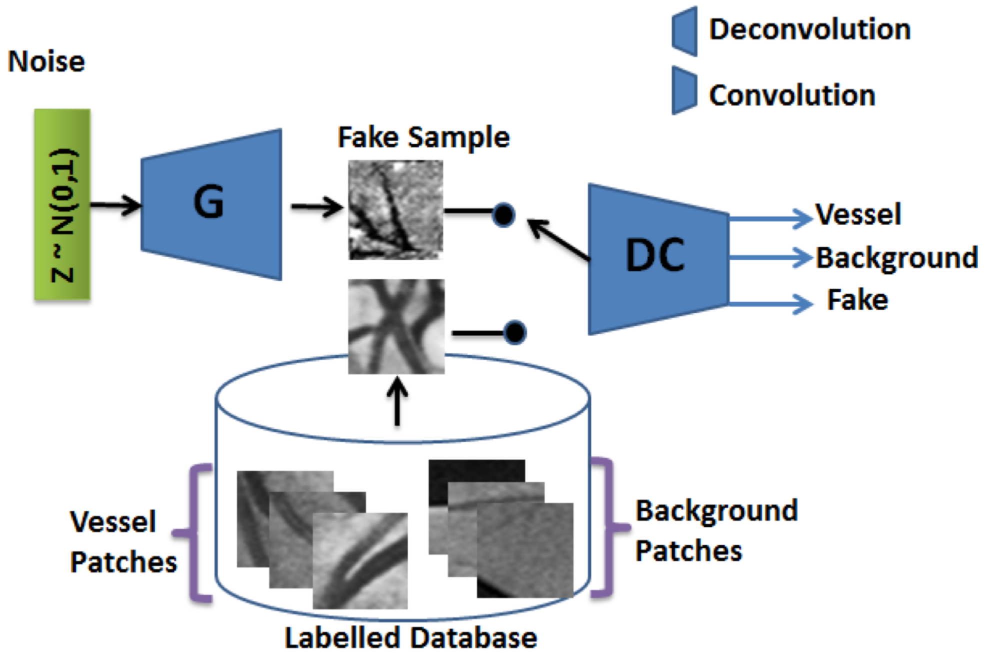
- The DC net now has to minimize two types of losses, viz.
  a) **classification loss ($\boldsymbol{L_c}$)**
  b) **adversarial loss ($\boldsymbol{L_{adv}}$)**

Proposed Model of GAN



- G takes in a **300-D standard normal noise** vector to **create a fake example**, G(z) via a series of deconvolution operations.

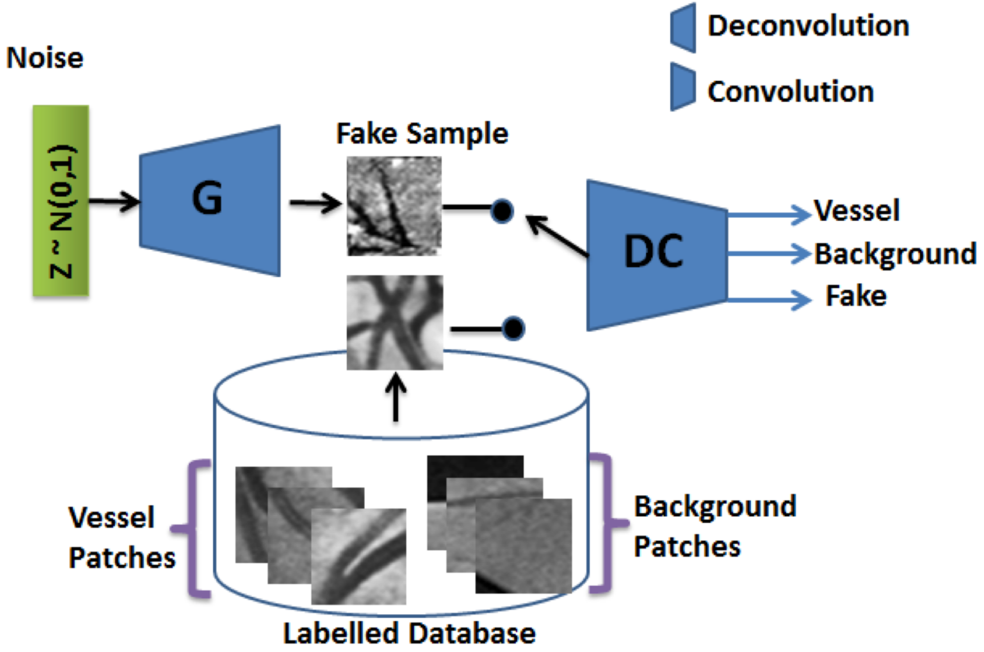
Proposed Model of GAN



- G takes in a **300-D standard normal noise** vector to **create a fake example**, G(z) via a series of deconvolution operations.

- DC net is to **assign correct class label (vessel or background) to real examples** coming from stored training database while **assigning G(z) to Fake class.**

Proposed Model of GAN

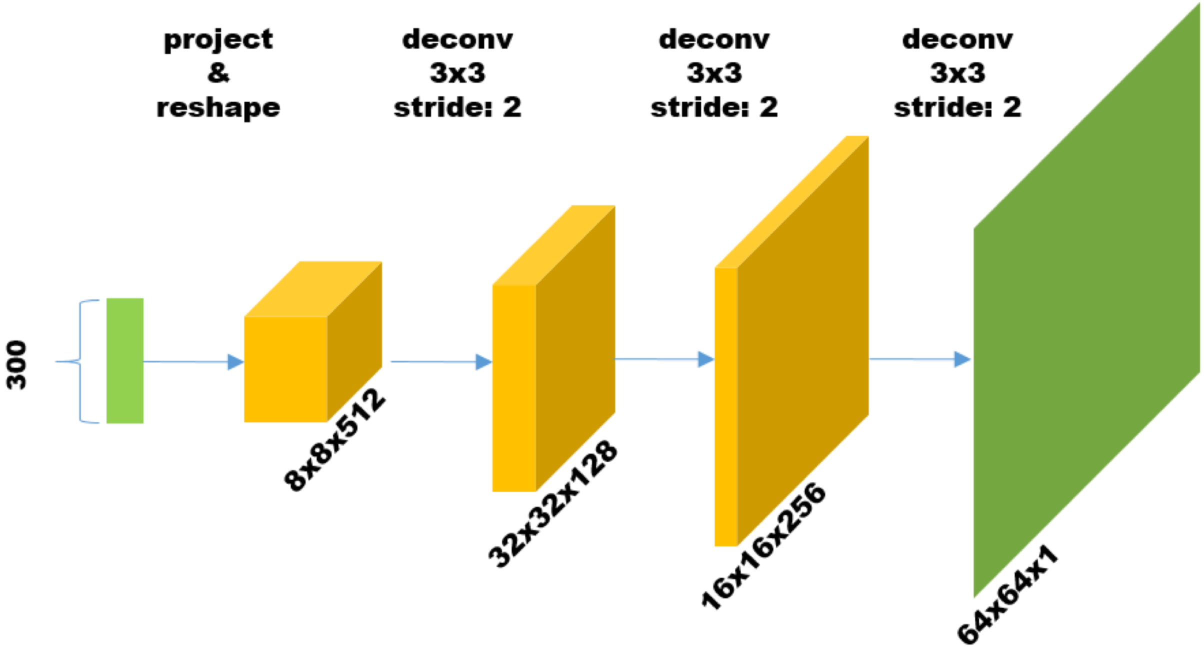


- G takes in a **300-D standard normal noise** vector to **create a fake example**, G(z) via a series of deconvolution operations.

- DC net is to **assign correct class label (vessel or background) to real examples** coming from stored training database while **assigning G(z) to Fake class.**

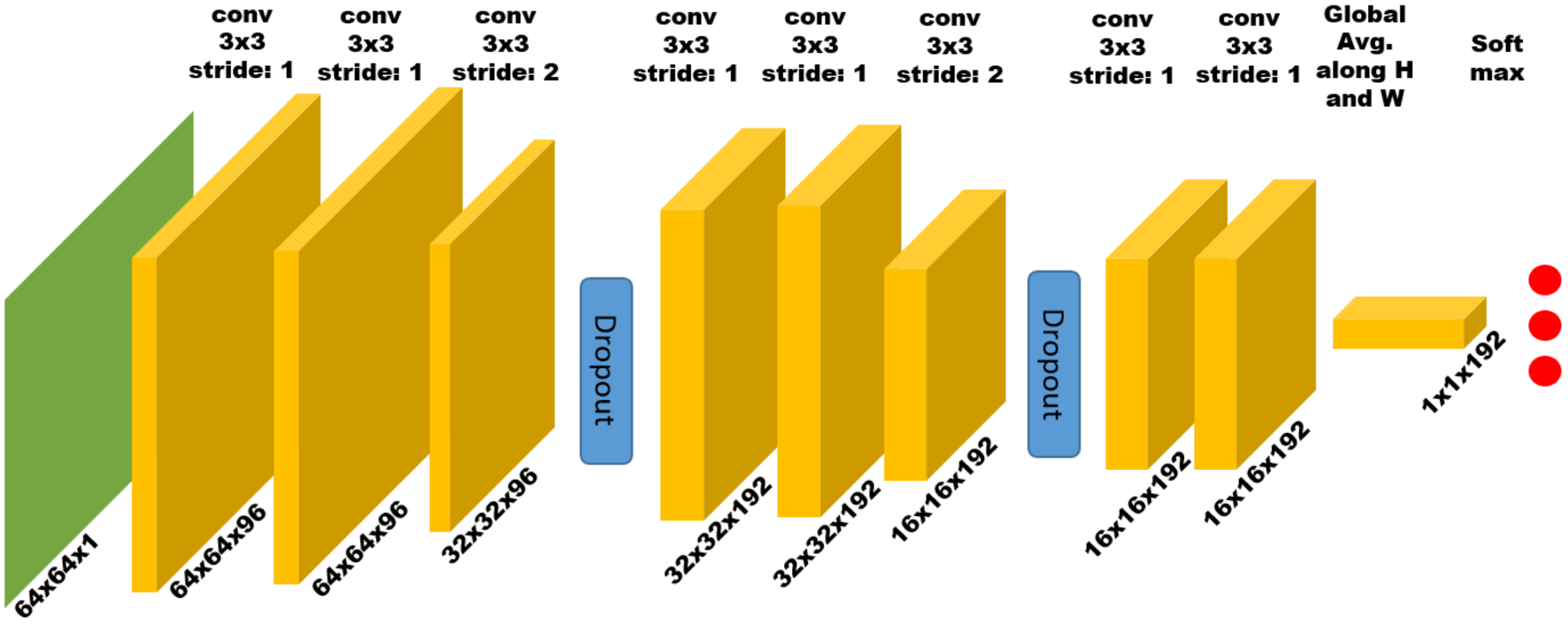
- **Goal of G is to fool DC** in assigning G(z) to any one of the training labels.

Generator Architecture



We apply **instance normalization** after every deconvolution followed by **Rectified linear unit (ReLU)** as non linear activation.

Discriminator/Classifier Architecture



We apply **instance normalization** after every convolution followed by **Leaky Rectified linear unit (LReLU)** as non linear activation.

Semi-Supervised Learning with GAN

The DC net is optimized to minimize both classification loss and adversarial loss.

$$L_c = -\mathbb{E}_{(x,y)\sim p_{data}(x,y)} log\, p_{DC}(y|x; y < K+1)$$

$$L_{adv} = -\mathbb{E}_{x\sim G} log\, p_{DC}(y = K+1|x)$$

Semi-Supervised Learning with GAN

The DC net is optimized to minimize both classification loss and adversarial loss.

$$L_c = -\mathbb{E}_{(x,y)\sim p_{data}(x,y)} log\, p_{DC}(y|x; y < K+1)$$

$$L_{adv} = -\mathbb{E}_{x\sim G} log\, p_{DC}(y = K+1|x)$$

Multi Task Learning

Assign correct class label to a real training sample

Assign high probability to FAKE class when fed with synthetic example

Semi-Supervised Learning with GAN

- The generator is updated in such a way so that the DC net **places minimum probability** over **class k = K+1**.

Semi-Supervised Learning with GAN

- The generator is updated in such a way so that the DC net **places minimum probability** over **class k = K+1**.

- DC net is therefore **fooled** to believe that the fake example belongs to one of the legitimate K classes of the database. So, for training the generator, we need to maximize, L_G,

$$L_G = -\mathbb{E}_{x \sim G} log\ p_{DC}(y = K + 1|x)$$

Semi-Supervised Learning with GAN

But trying to optimize the generator network with

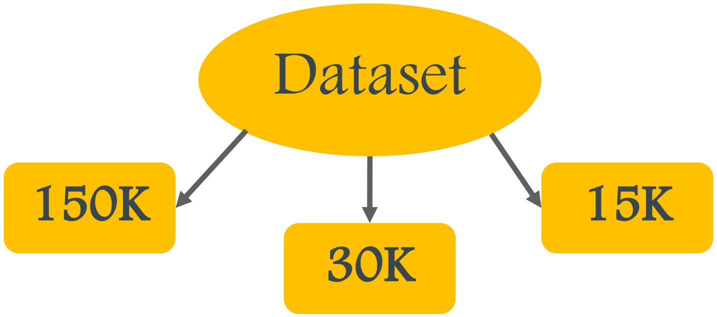
$$L_G = -\mathbb{E}_{x \sim G} log\, p_{DC}(y = K + 1|x)$$

is practically not advisable because in the early phase of training, **magnitudes of gradients propagated to generator are small**. Thus, we instead minimize,

$$L_G = -\mathbb{E}_{x \sim G} log\{1 - p_{DC}(y = K + 1|x)\}$$

Reference : Generative Adversarial Nets. Ian Goodfellow et al. NIPS 2014

Training Details



On each dataset, we train the simple CNN and GAN-CNN from scratch.

ADAM optimizer for both the G and DC net

Slope of leaky ReLU = 0.1

Initial learning rate = 10^{-4} (for both G and **DC**)

Decay factor of **0.8** after every **20 epochs**

Mini-batch size of **64**

Testing Details

20 Test images

At **test time**, a real test examples, x_t is assigned a label, $y^*(x)$, according to,

$$y^*(x) = \underset{y}{\operatorname{argmax}} \; p_{DC}(c = y|x)$$

Results

In retinal vessel segmentation literature, area under the **Receiver Operation Curve**, i.e., **AUC** is taken as a standard metric of comparison. A **larger AUC** signifies a **better segmenter**.

Dataset Size	GAN-CNN	CNN	p-value
150K	0.962	0.960	0.1
30K	0.945	0.921	10^{-3}
15K	0.931	0.916	10^{-5}

Results

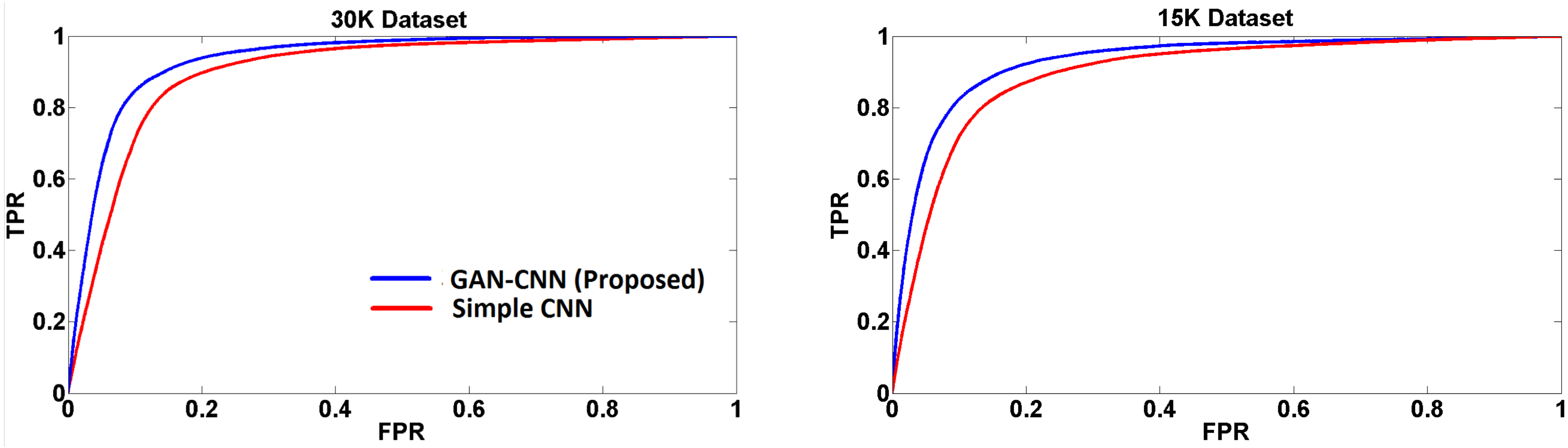
In retinal vessel segmentation literature, area under the **Receiver Operation Curve**, i.e., **AUC** is taken as a standard metric of comparison. A **larger AUC** signifies a **better segmenter**.

Dataset Size	GAN-CNN	CNN	p-value
150K	0.962	0.960	0.1
30K	0.945	0.921	10^{-3}
15K	0.931	0.916	10^{-5}

Conclusion: **p-value (Welch's t-test)** indicates that there is **significant difference** between the **mean AUCs** of the **proposed method** and **simple CNN**, specially when trained on smaller training sets. The **null hypothesis** in this case is that the **mean AUCs** of both paradigms of segmenters are **same**.

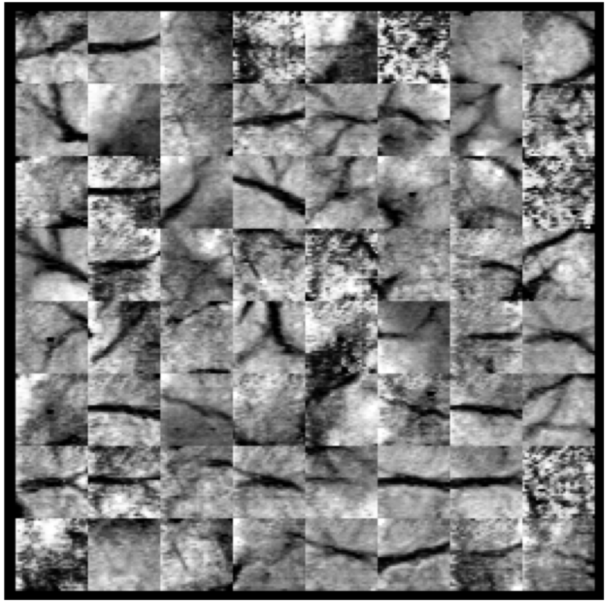
Results

ROC curves of proposed GAN-CNN and simple CNN on the combined 20 test images of DRIVE retina dataset.

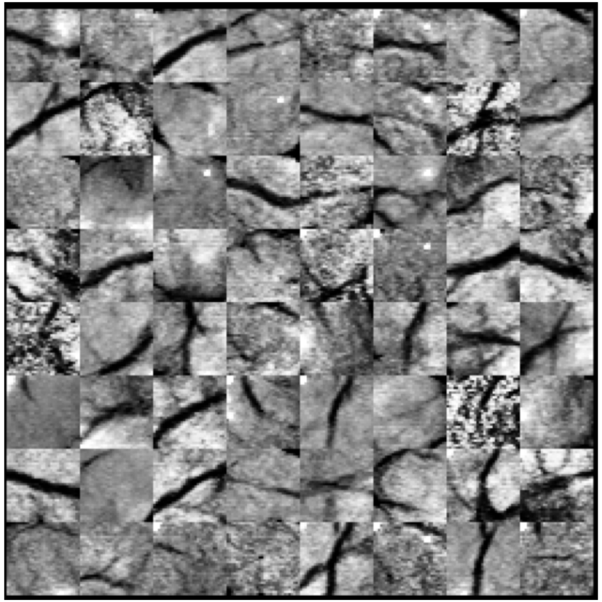


Curves of **GAN-CNN** always tends to be **higher** on the ROC plots compared to **simple CNN** based segmenter. The visualization bolsters our claim that training a **GAN based CNN for semantic segmentation is data efficient**.

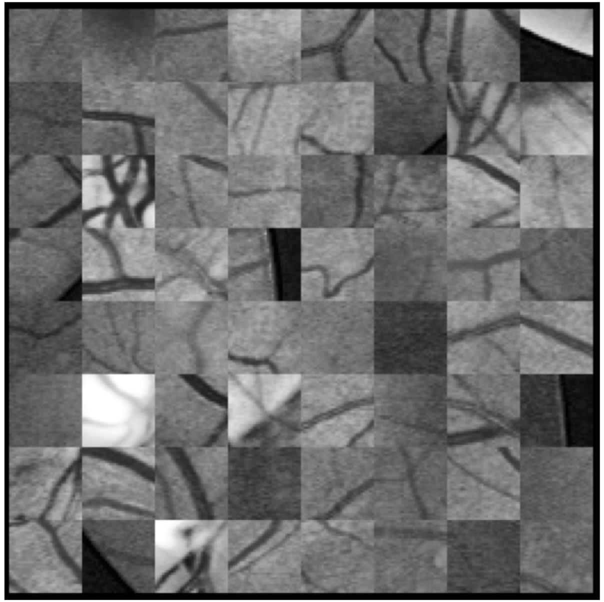
Results



Samples generated during training on 30K dataset



Samples generated during training on 15K dataset



Real samples

Results

Comparison of mean AUC of some of the contemporary deep learning based retinal vessel segmentation algorithm.

Method	Dataset Size	Mean AUC
Maji et al.	60K	0.928
Lahiri et al.	120K	0.950
Fu et al.	330K	0.947
Liskowski et al.	3857K	0.963
GAN-CNN (Proposed)	30K	0.945
GAN-CNN (Proposed)	15K	0.931

Results

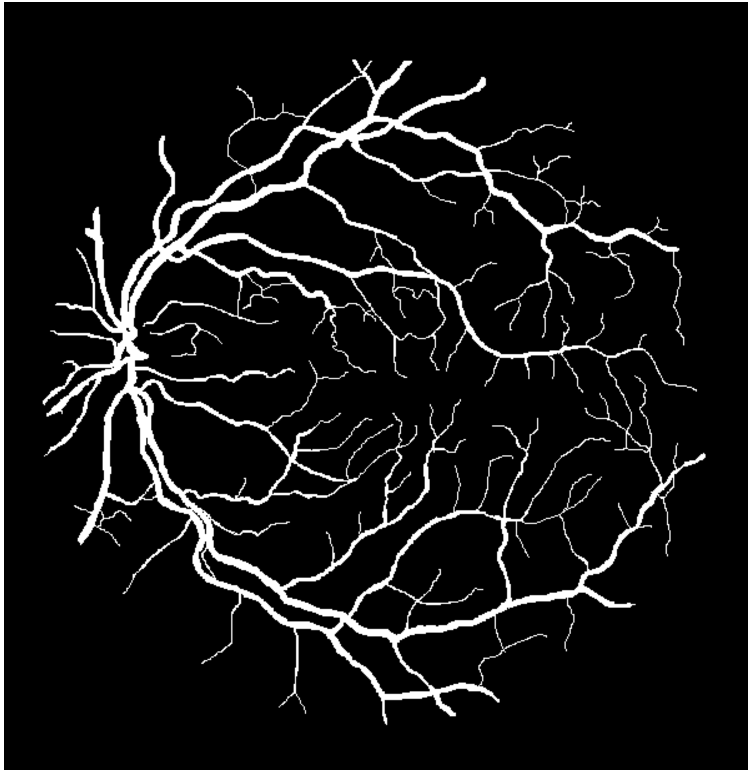
Comparison of mean AUC of some of the contemporary deep learning based retinal vessel segmentation algorithm.

Method	Dataset Size	Mean AUC
Maji et al.	60K	0.928
Lahiri et al.	120K	0.950
Fu et al.	330K	0.947
Liskowski et al.	3857K	0.963
GAN-CNN (**Proposed**)	30K	0.945
GAN-CNN (**Proposed**)	15K	0.931

Conclusion:. Even with **smaller dataset size**, our proposed method performs **comparable (sometimes even better)** than the competing techniques trained with **2X-10X times more training data**.

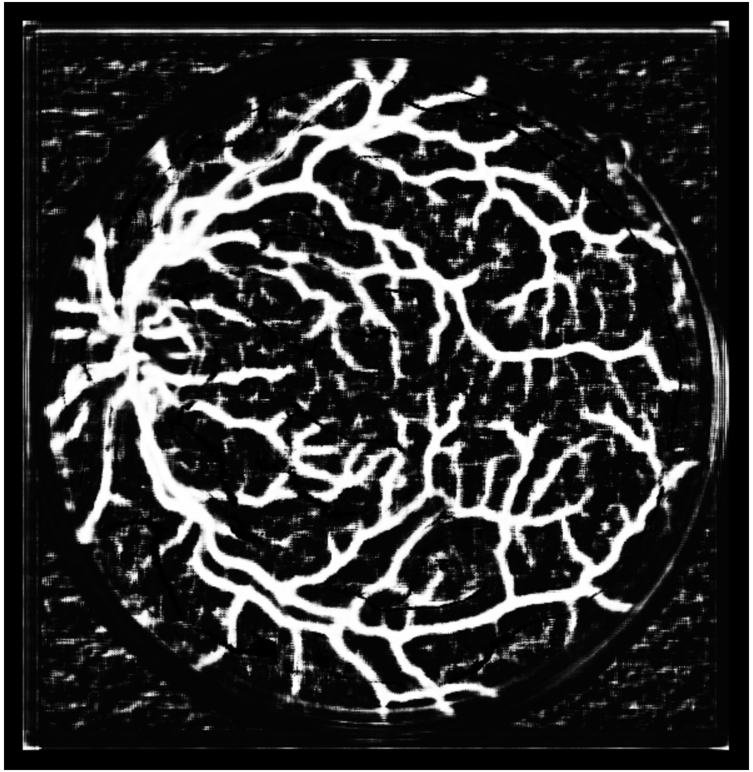
Results

Sample Fundus Image

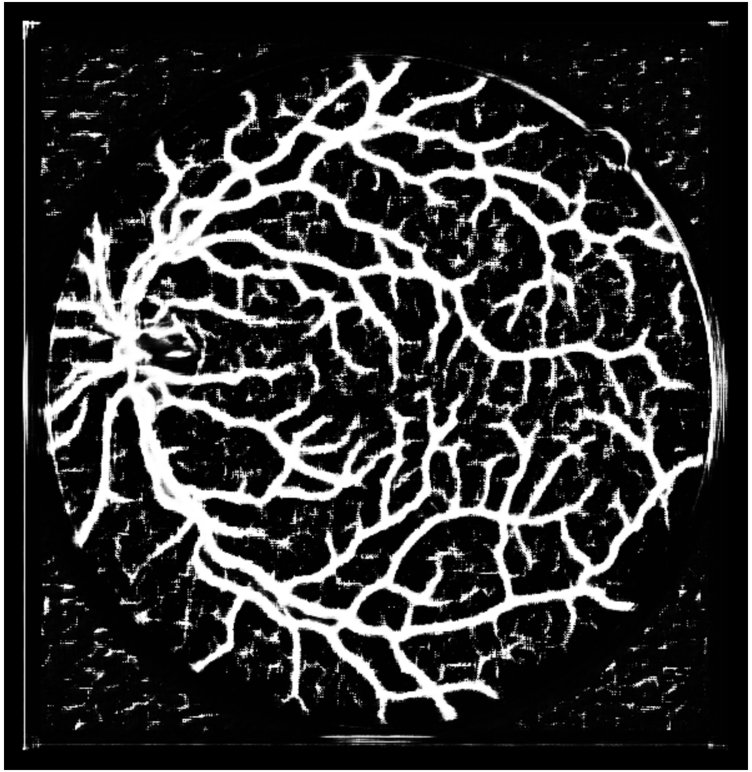


Ground Truth

Results

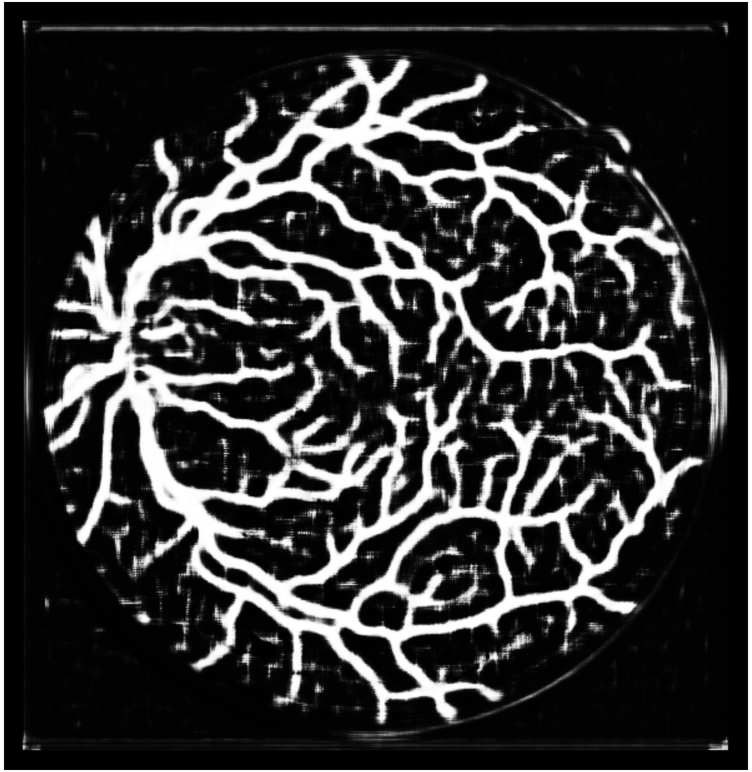


CNN trained on 15K

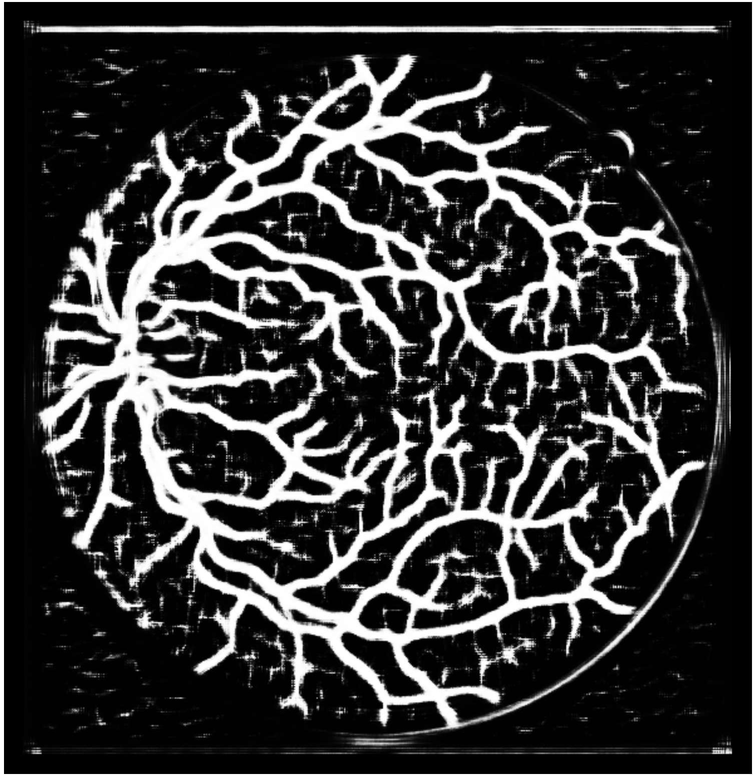


GAN-CNN trained on 15K

Results



CNN trained on 30K



GAN-CNN trained on 30K

Contributions

- To our best knowledge, this is the **first work** which leverages GAN for semi-supervised learning on **large scale fundus imaging** modality for **automated blood vessel segmentation**.

Contributions

- To our best knowledge, this is the **first work** which leverages GAN for semi-supervised learning on **large scale fundus imaging** modality for **automated blood vessel segmentation**.

- We achieve comparable performance (**sometimes even better**) with **recent CNN based segmentation techniques** while **using upto 9X times less training data**.

Contributions

- To our best knowledge, this is the **first work** which leverages GAN for semi-supervised learning on **large scale fundus imaging** modality for **automated blood vessel segmentation**.

- We achieve comparable performance **(sometimes even better)** with **recent CNN based segmentation techniques** while **using upto 9X times less training data**.

- We show that **performance** of simple **CNN based segmenter** starts **deteriorating faster on smaller datasets** compared to GAN-CNN.

Contributions

- To our best knowledge, this is the **first work** which leverages GAN for semi-supervised learning on **large scale fundus imaging** modality for **automated blood vessel segmentation**.
- We achieve comparable performance (**sometimes even better**) with **recent CNN based segmentation techniques** while **using upto 9X times less training data**.
- We show that **performance** of simple **CNN based segmenter** starts **deteriorating faster on smaller datasets** compared to GAN-CNN.
- We show that the **difference of performances** between simple CNN and GAN-CNN is **statistically significant** when trained on **smaller training sets**.

Impact

- We applied the proposed model to the challenging task of vessel segmentation in fundus images, but our concept is generic.

- Fundus Images have:
  - Intricate Branching Pattern
  - Noisy Background
  - Irregular Illumination

Therefore, pixel level manual annotation is much more tedious than image tagging, thus bolstering the importance of our contribution.

Publication

Title: **Generative Adversarial Learning for Reducing Manual Annotation in Semantic Segmentation on Large Scale Microscopy Images: Automated Vessel Segmentation in Retinal Fundus Image as Test Case**

Conference: Computer Vision and Pattern Recognition (CVPR) Workshop on **Computer Vision for Microscopy Image Analysis (CVMI)** 2017

CVPR
July 21-26
HONOLULU
2017

Future Work

Future Work

One possibility is to make use of large amount of unlabeled data by forcing the DC-net to place low likelihood for fake class to these examples

Another possibility is to use class conditional generator network to force it to generate class specific fake examples and forcing the DC-net to classify these fake examples.

Both of these methods are further steps towards improving the performance of the combined DC-net.

Thank You!