NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

School of Computer Science and Engineering
College of Engineering

Attack on Training Effort of Deep Learning

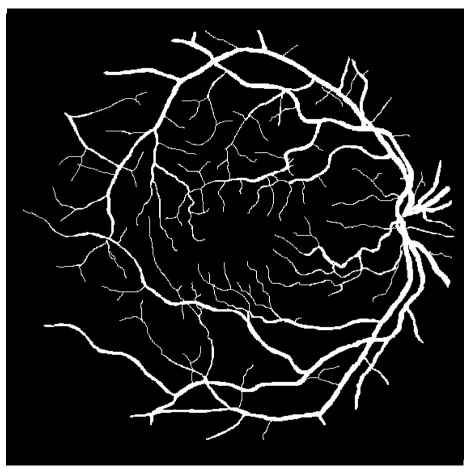
Student: Chan Wen Le
Supervisor: Prof. Liu Yang

Motivation

Retinal vessel segmentation is key to diagnosis of ocular diseases. However, the current approaches have 2 limitations:

Small number of annotated images available	Dataset are all of high quality fundus images
Overfitting + Overtraining. Limited generalisation performance to unseen images	Limited performance on common **low-quality** fundus images and ill-crafted adversarial examples


Retinal images

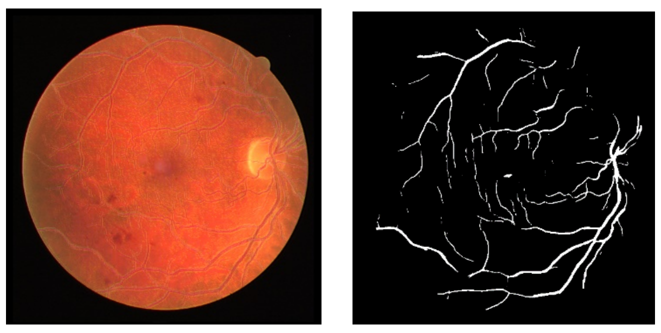

Segmentation label

Contributions

1 Demonstrate the effect of limitations and vulnerability of DNNs through proposing 2 adversarial attack methods

Pixel-wise Attack

Apply a Light-Enhancement curve iteratively on each pixel's illumination



11% decrease in Accuracy, 0.26 decrease in sensitivity

Threshold-based Attack

Create non-uniform illumination through disproportional change in different region's illumination

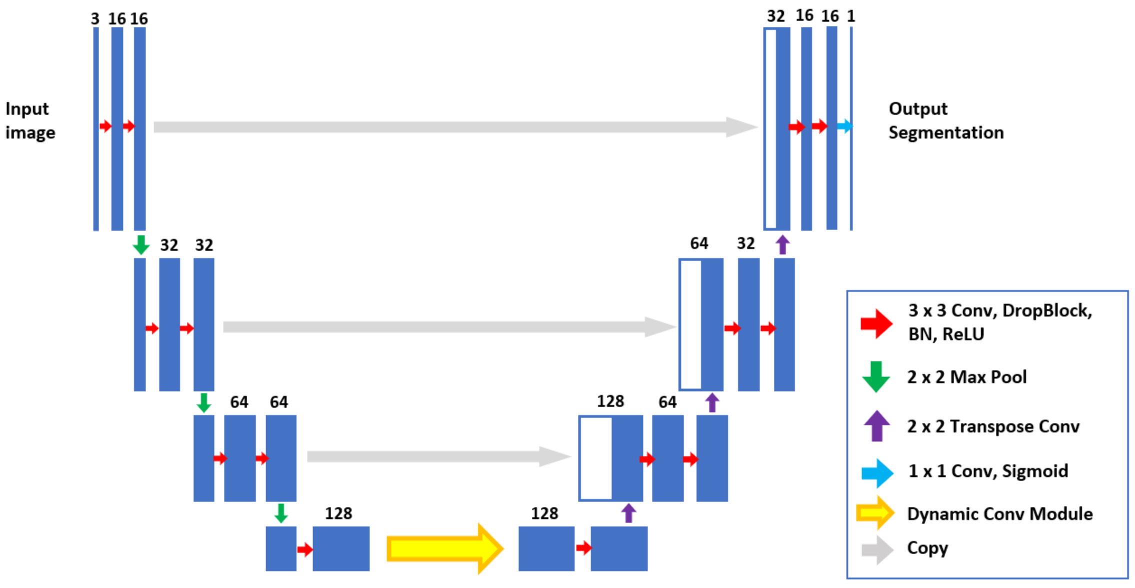
6% decrease in Accuracy, 0.64 decrease in sensitivity

High attack success rate revealed that realistic illumination pose a potential threat to current DNN approaches

2 Increase robustness of automated retinal vessel segmentation

New Network Architecture

Improve current network by adding Dropblock and Dynamic Convolution on U-Net backbone



Adversarial Training

Min-Max formulation: Use inner loop to generate adversarial examples and outer loops to train model using generated examples

Higher segmentation performance on adversarial examples and synthesised low-quality images