APTOS 2019:
Blindess Detection

Vancouver Kaggle Meetup

Hello!

I am Alice Roberts

- ★ Recent SFU graduate in Applied Mathematics & Statistics
- ★ Seeking a job in data science / data analytics

Contact me at:

Email: Roberts.alice04@gmail.com

LinkedIn: https://www.linkedin.com/in/alice-victoria-roberts/

Github: https://github.com/aliceroberts10

Our Goal for Today:

- ★ Introduction
- ★ Overview of Data
- ★ Exploratory Data Analysis
- ★ Background Concepts
- ★ Top available solutions
- ★ Discussion/Conclusion

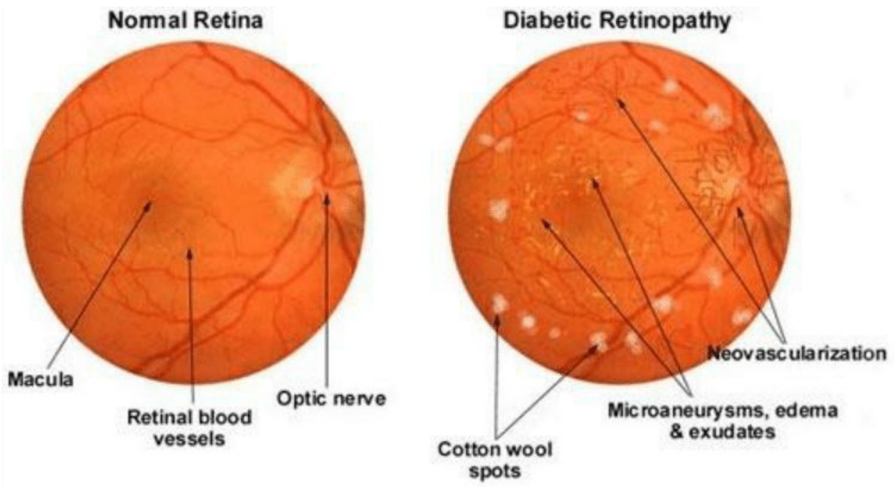
1.

Introduction

What is the problem?

Blindess Detection: Diabetic Retinopathy

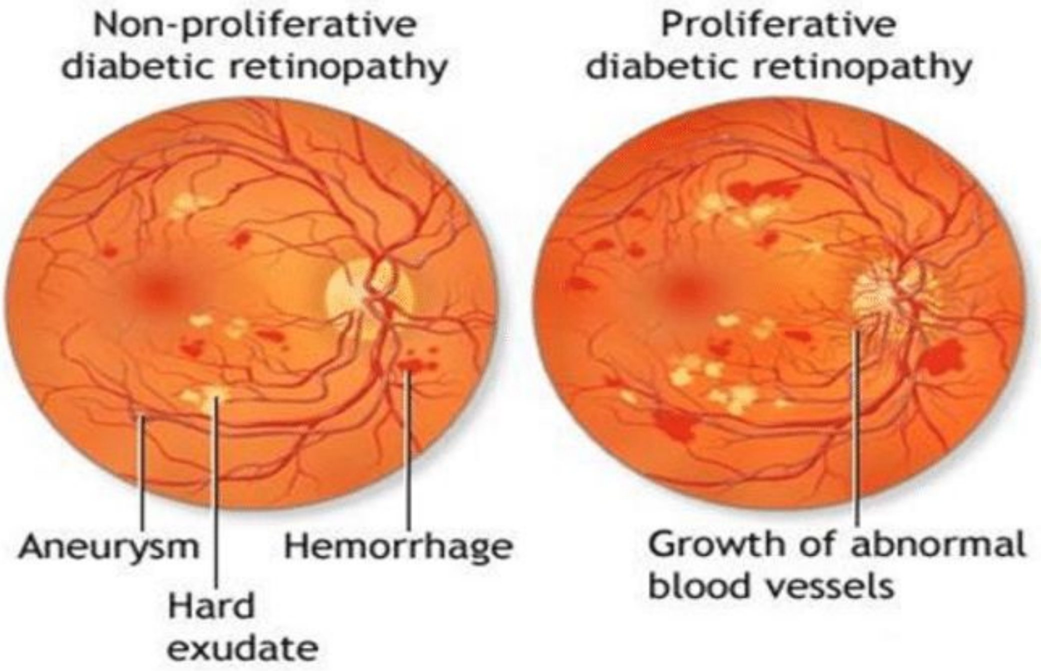
What is Diabetic Retinopathy?



Task:

Given thousands of images of retinas, can we classify whether or not the person has or is at risk of blindess ?

Non-Proliferative vs Proliferative DR



Proliferative Diabetic Retinopathy:
Abnormal blood vessel growth

How Do Humans Classify DR?

Early NPDR – At least one microaneurysm present on retinal exam.

Moderate NPDR – Characterized by multiple microaneurysms, dot-and-blot hemorrhages, venous beading (when the walls of major retinal **veins** lose their normal parallel alignment) and/or cotton wool spots.

Severe NPDR – In the most severe stage of NPDR.
It is diagnosed using the "**4-2-1 rule**." as follows:

- Diffuse intraretinal hemorrhages and microaneurysms in **4 quadrants**
- Venous beading in ≥ **2 quadrants**, or
- Intraretinal microvascular abnormalities (IRMA) in ≥ **1 quadrant**.

Proliferative NPDR - Characterized by all symptoms above as well as abnormal blood vessel growth. At this point, sudden vision loss is highly probable/expected.

Motivation

Current Problem:

Aravind Eye Hospital technicians from India have to travel to rural areas to capture images and then rely on highly trained doctors to review the images and provide diagnosis.

Proposed Solution :

Machine Learning Models can:

- Speed up disease detection
- Identify diabetic retinopathy automatically
- Prevent lifelong blindness
- Can be modified to detect other diseases such as glaucoma

Competition Description

- 2,943 Teams
- 3,522 Competitors
- 60,678 Entries

Winners	Cash Prize	Scores
1st place: Guanshuo Xu	$25,000	0.936129
2nd place: [ods.ai] Eye of Private LB	$12,000	0.934310
3rd place:[ka.kr] Save our eyes	$8,000	0.933720
4th place: Best Over Fitting	$5,000	0.933693

Scoring Method:

Quadratic weighted kappa (0,1)

2.

Data Overview

Testing, Training, Sample Submission

What was given?

- ★ train.csv - the training labels
- ★ test.csv - the test set
- ★ sample_submission.csv
- ★ train.zip - the training set images
- ★ test.zip - the public test set images

Training set: (3,662)

	id_code <fctr>	diagnosis <int>
1	000c1434d8d7	2
2	001639a390f0	4
3	0024cdab0c1e	1
4	002c21358ce6	0
5	005b95c28852	0
6	0083ee8054ee	4
7	0097f532ac9f	0
8	00a8624548a9	2
9	00b74780d31d	2
10	00cb6555d108	1

Labels:

0 - No DR

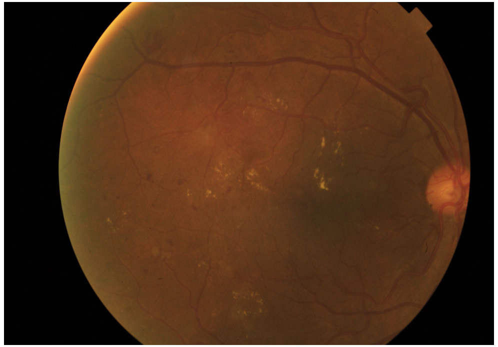
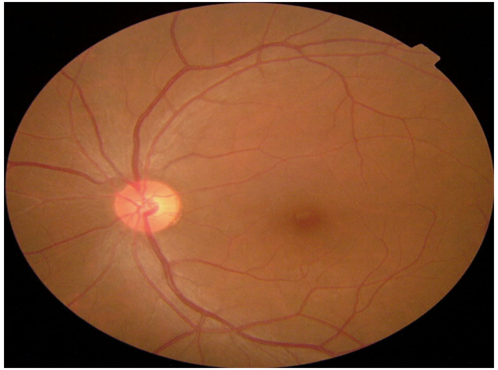
1 - Mild

2 - Moderate

3 - Severe

4 - Proliferative DR

Example:

Id Code:		**Diagnosis:**
000c1434d8d7		2
002c21358ce6		0

Labels:

0 - No DR

1 - Mild

2 - Moderate

3 - Severe

4 - Proliferative DR

Test Set: (1,928)

	id_code <fctr>
1	0005cfc8afb6
2	003f0afdcd15
3	006efc72b638
4	00836aaacf06
5	009245722fa4
6	009c019a7309
7	010d915e229a
8	0111b949947e
9	01499815e469
10	0167076e7089

Sample Submission:

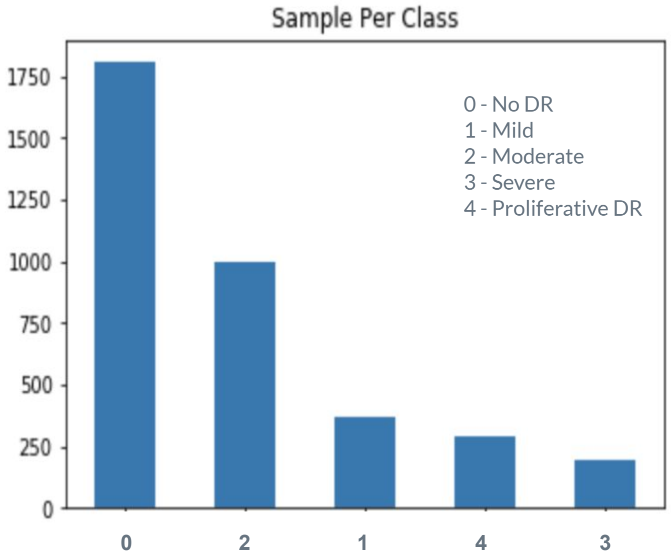
	id_code <fctr>	diagnosis <int>
1	0005cfc8afb6	0
2	003f0afdcd15	0
3	006efc72b638	0
4	00836aaacf06	0
5	009245722fa4	0
6	009c019a7309	0
7	010d915e229a	0
8	0111b949947e	0
9	01499815e469	0
10	0167076e7089	0

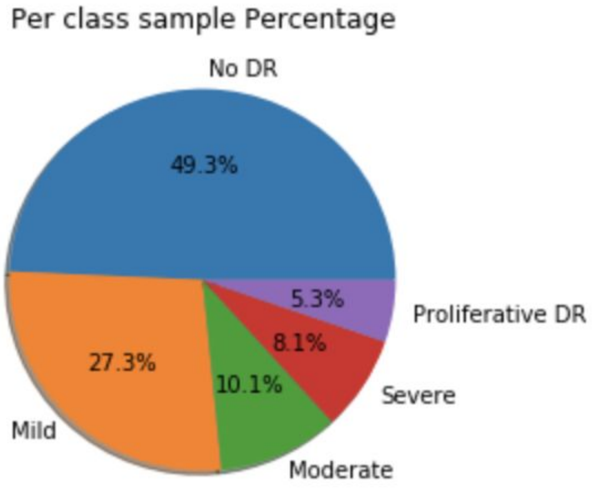
3.

EDA

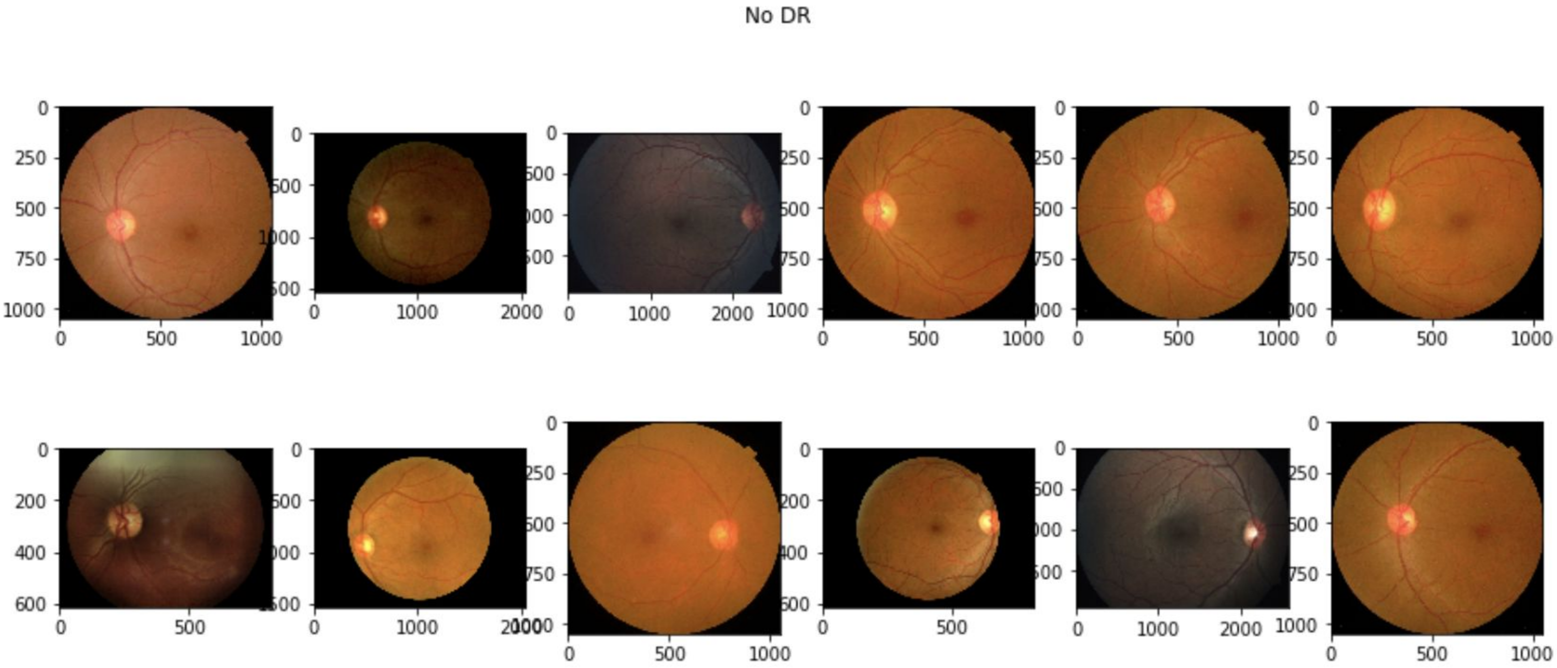
Exploratory Data Analysis

Summary

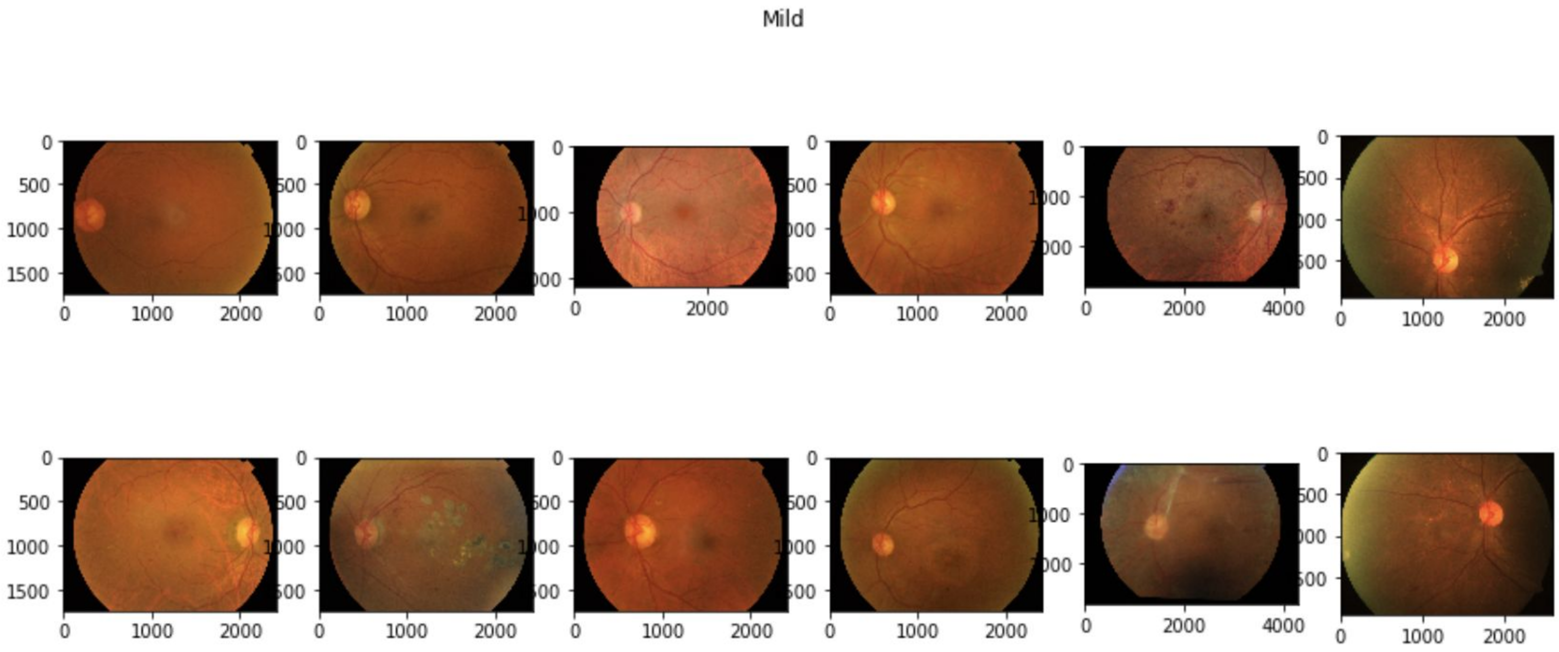

Sample Per Class
1750
1500
1250
1000
750
500
250
0
0 - No DR
1 - Mild
2 - Moderate
3 - Severe
4 - Proliferative DR
0
2
1
4
3


Per class sample Percentage
No DR
49.3%
5.3%
Proliferative DR
8.1%
Severe
27.3%
10.1%
Mild
Moderate

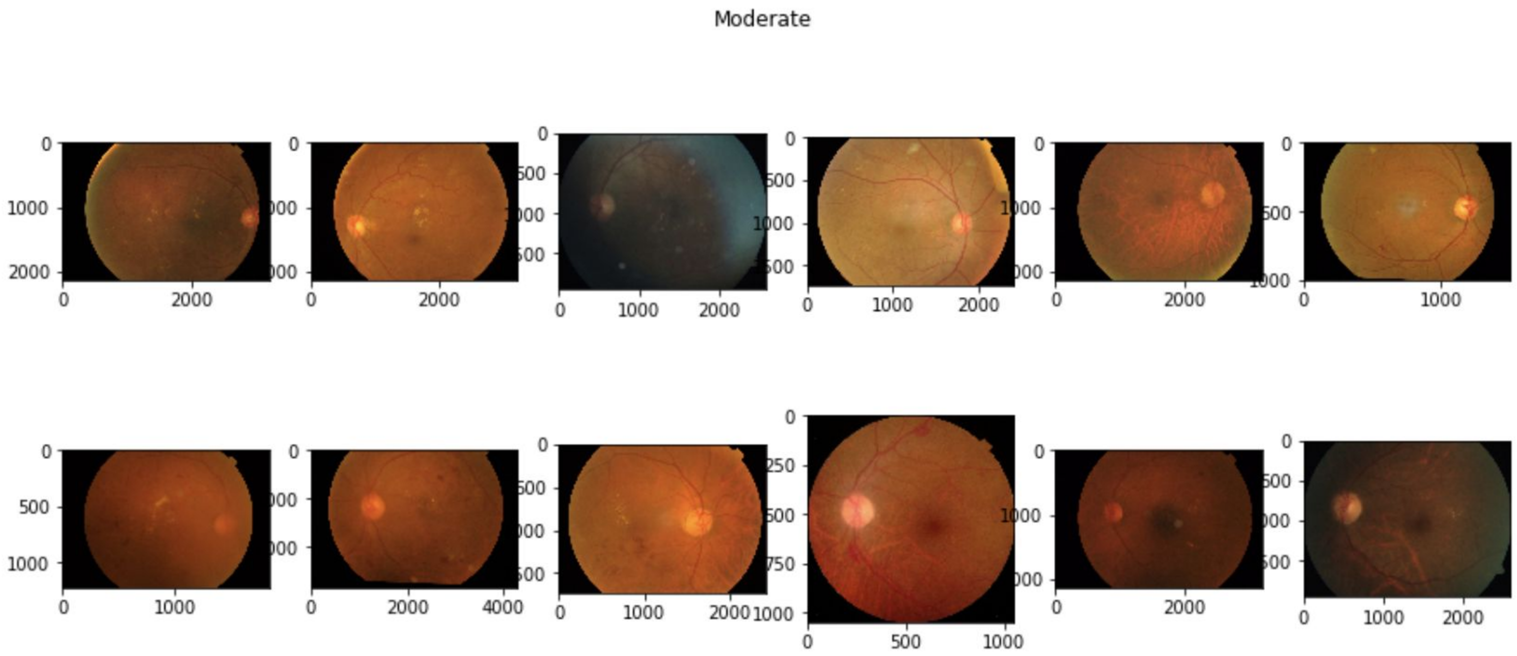
0 - No DR



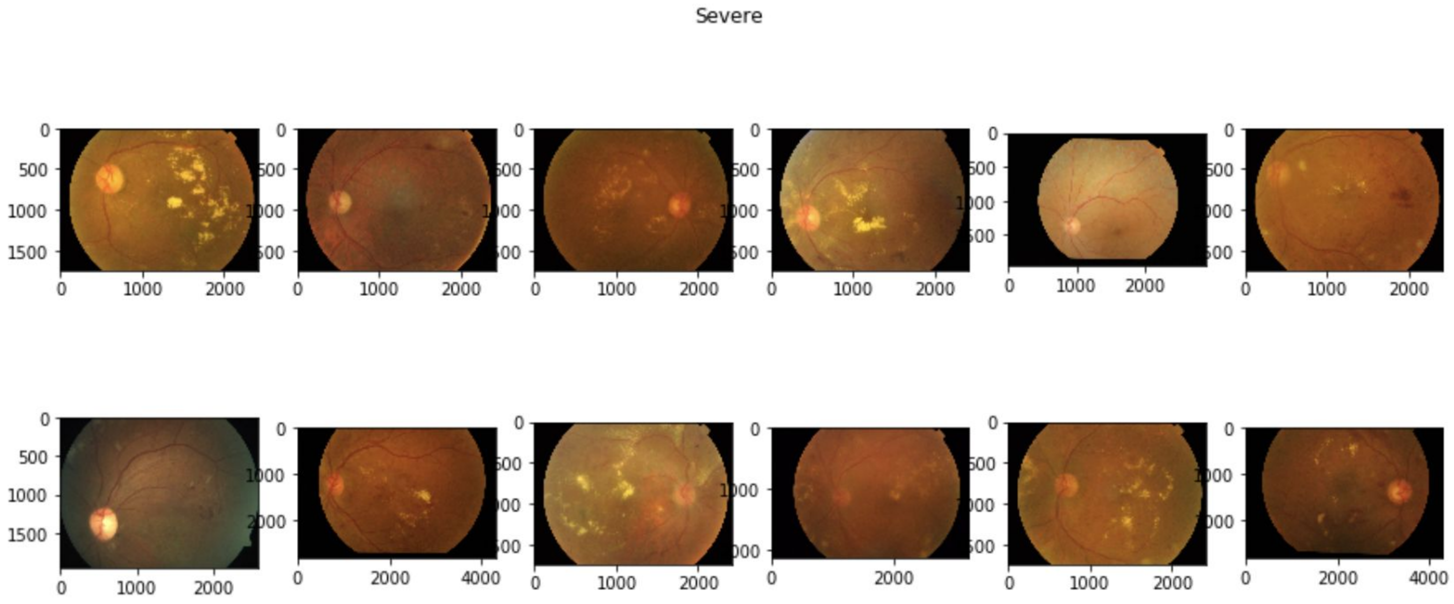
1 - Mild DR



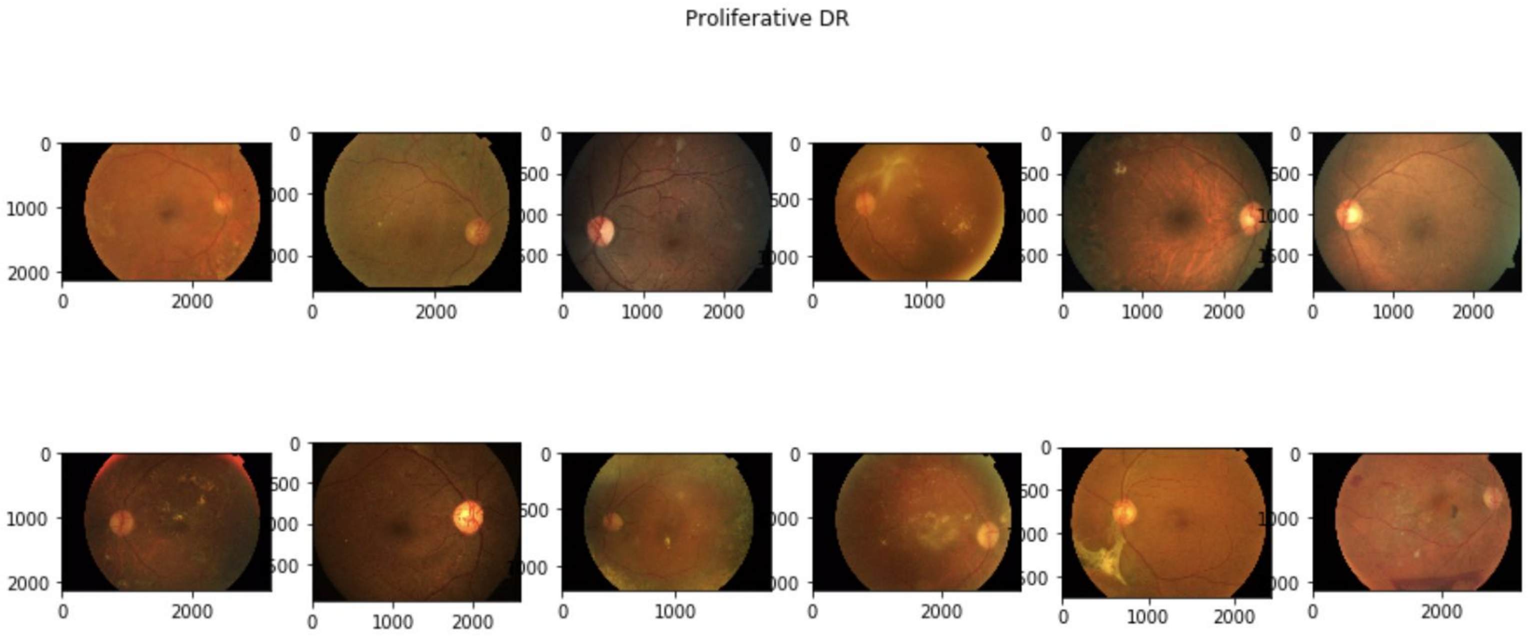
2 - Moderate DR



3 - Severe DR



4 - Proliferative DR



4.

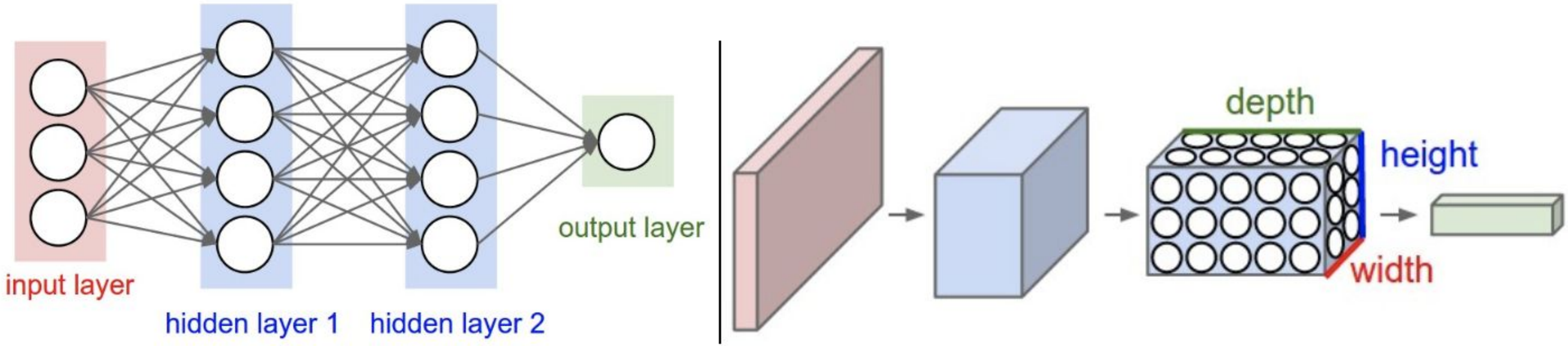
Background Concepts

CNNs, EfficientNets, etc.

Overview

- Convolution Neural Networks
- Vanishing Gradient Problem
- ResNets
- EfficentNets
- DenseNets
- InceptionNets
- Squeeze-and-Excitation Networks

Convolution Neural Networks



Note: Regular Neural Nets don't scale well to full images.

CovNet architectures make the explict assumption that the inputs are images.
The layers have neurons arranged in 3D: width, height and depth.

Problems with Deeper CNNs

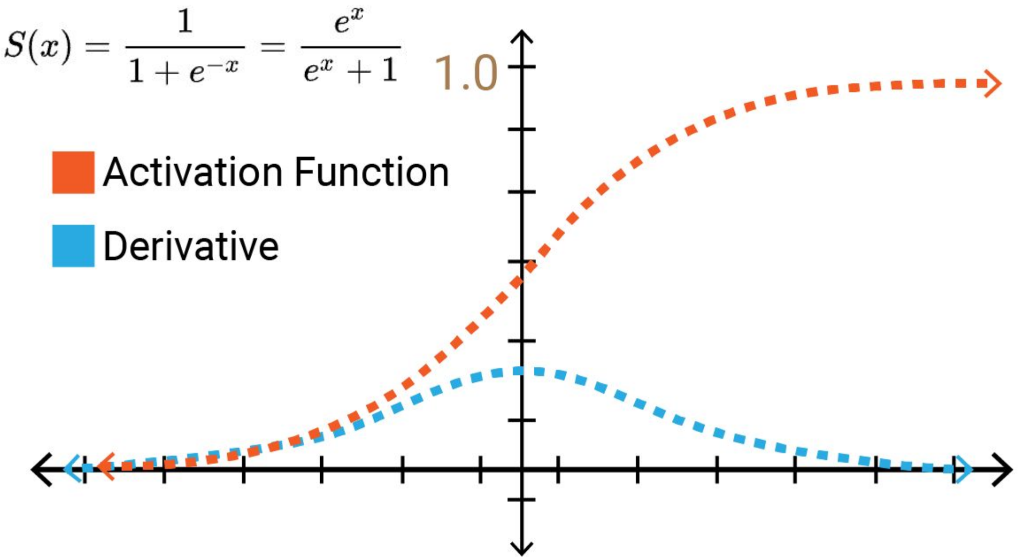
Universal approximation theorem: we know that a feedforward network with a single layer is sufficient to represent any function.

However, the layer might be massive and the network is prone to overfitting the data. Therefore, there is a common trend that our network architecture needs to go deeper.

However, increasing network depth does not work by simply stacking layers together.

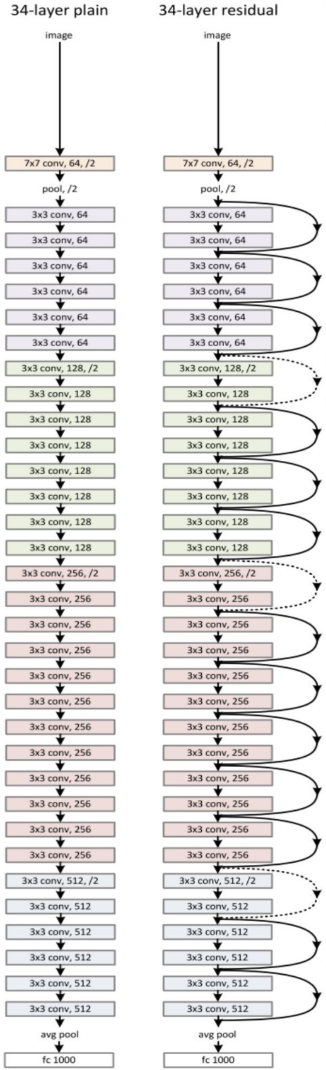
Why?

Vanishing Gradient Problem



When the network is too deep, the gradients from where the loss function is calculated easily shrink to zero after several applications of the chain rule.

This result on the weights never updating its values and therefore, no learning is being performed.



What are ResNets?

"The deeper the better"? Well not when it gets too deep!

Problem? Vanishing Gradient Problem

ResNets solves this problem!

With ResNets, the **gradients can flow directly through the skip connections backwards from later layers to initial filters**.

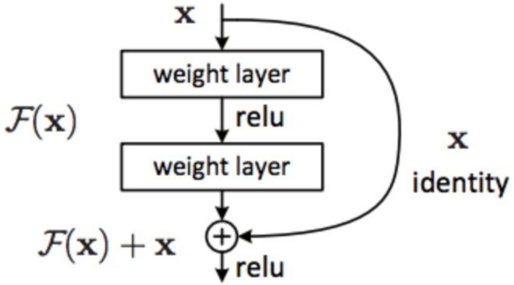
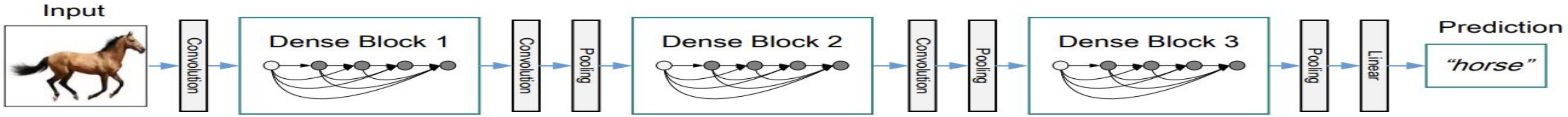


Figure 2. Residual learning: a building block.

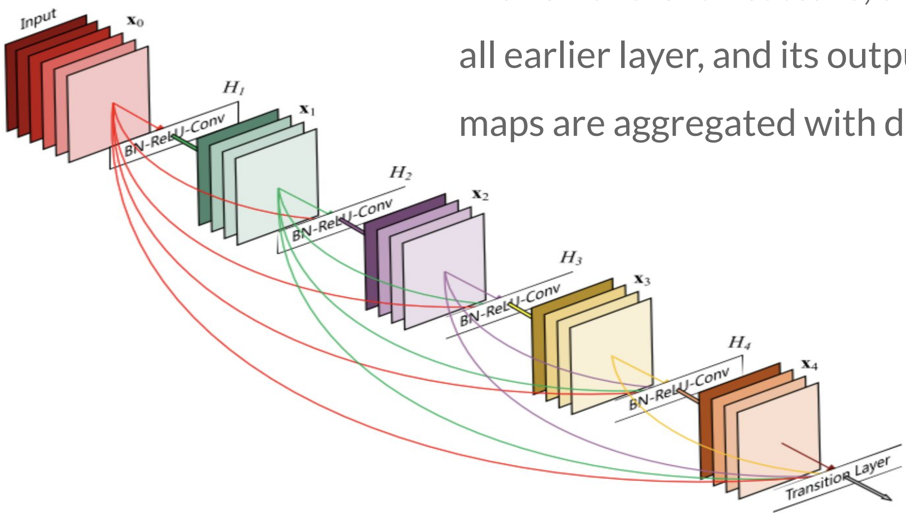
the residual connection directly adds the value at the beginning of the block, **x**, to the end of the block (F(x)+x). This residual connection doesn't go through activation functions that "squashes" the derivatives, resulting in a higher overall derivative of the block.



What are DenseNets?

DenseNet further exploits the effects of shortcut connections — it connects all layers directly with each other.

In this novel architecture, the input of each layer consists of the feature maps of all earlier layer, and its output is passed to each subsequent layer. The feature maps are aggregated with depth-concatenation.

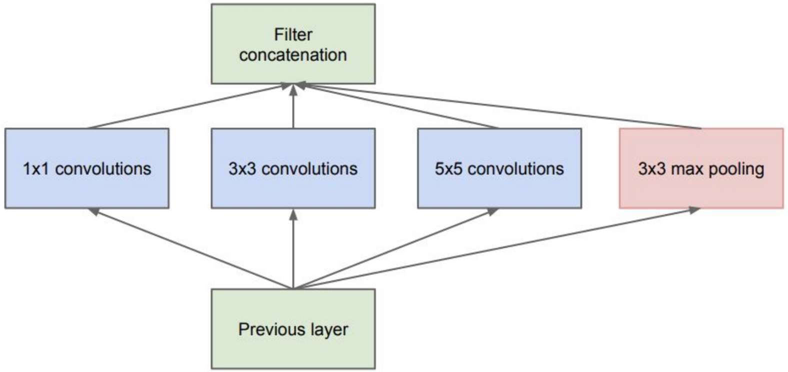


<u>Benefits of DenseNets:</u>

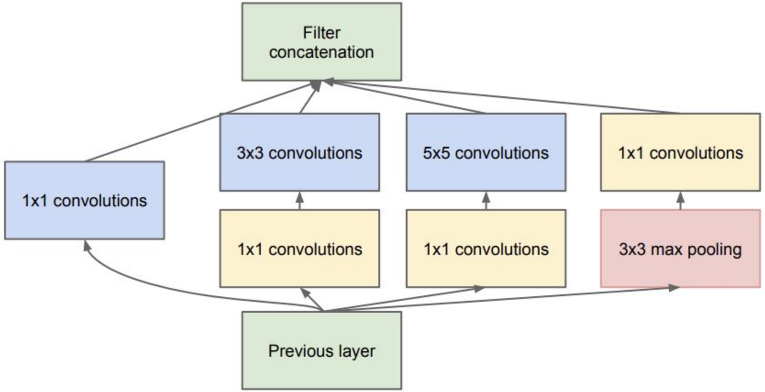
- ★ Vanishing gradients problem
- ★ Encourages feature reuse
- ★ Making network highly parameter-efficient

What are Inception Networks?

design that allows for increasing the depth and width of the network while keeping the computational budget constant



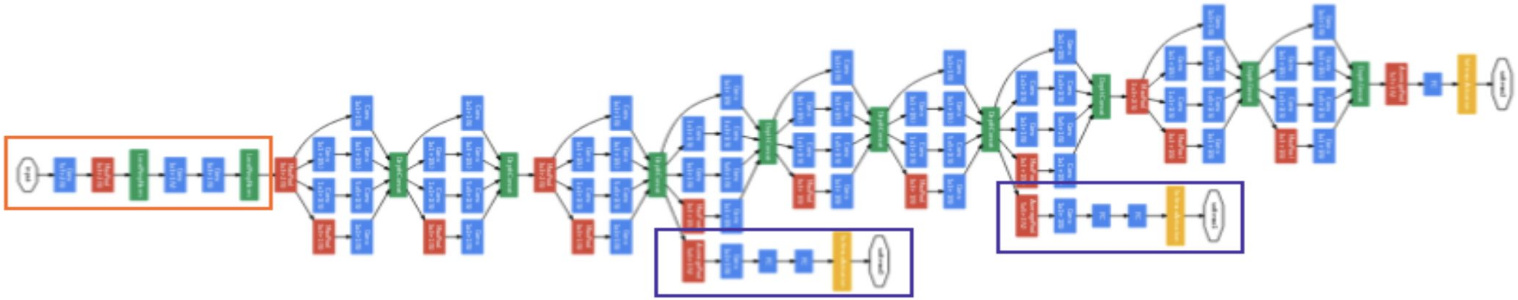
(a) Inception module, naïve version



(b) Inception module with dimension reductions

What are Inception Networks?

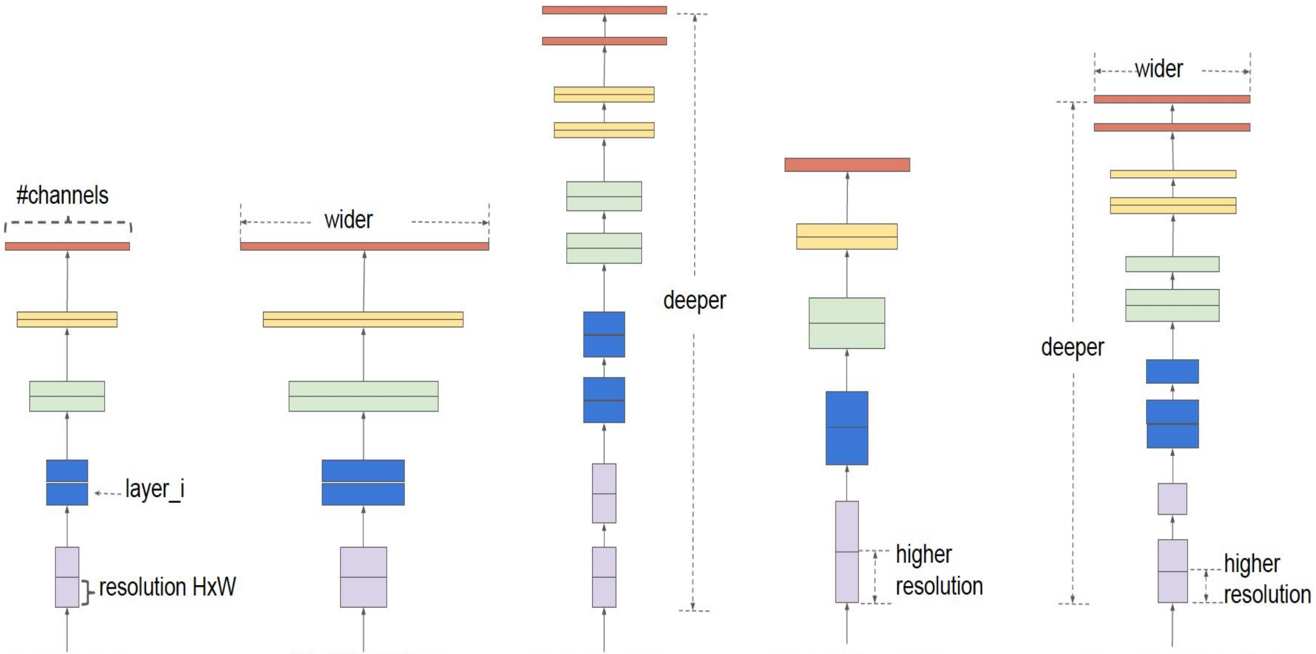
Using the dimension reduced inception module, an NN architecture was built



To prevent the **middle part** of the network from the vanishing gradient problem, two auxiliary classifiers were introduced.

They essentially applied softmax to the outputs of two of the inception modules, and computed an **auxiliary loss** over the same labels.

What are EfficientNets ?

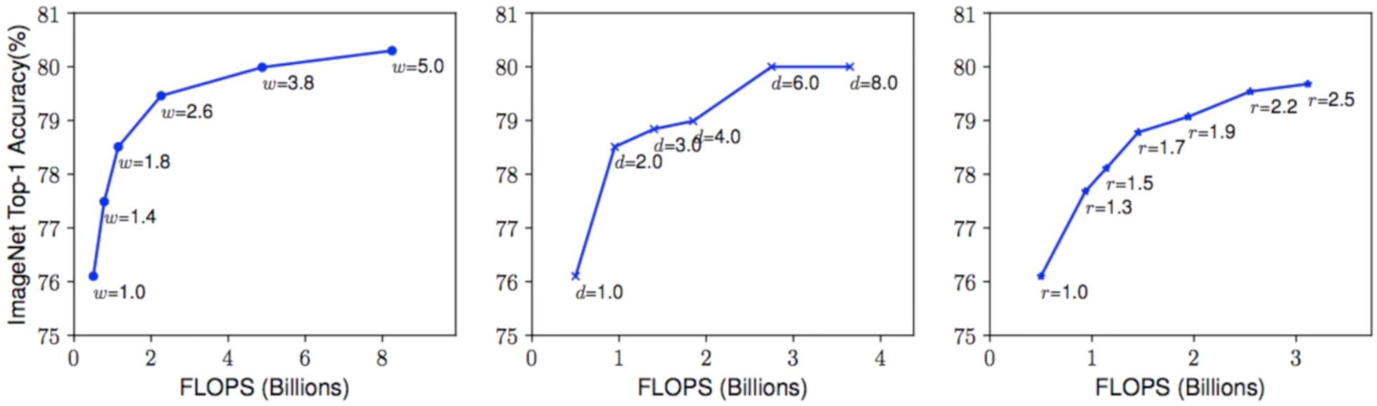


(a) baseline (b) width scaling (c) depth scaling (d) resolution scaling (e) compound scaling

Achieves state of the art accuracy in image classification

Rethinks the way we scale CNN's up

What are EfficientNets ?



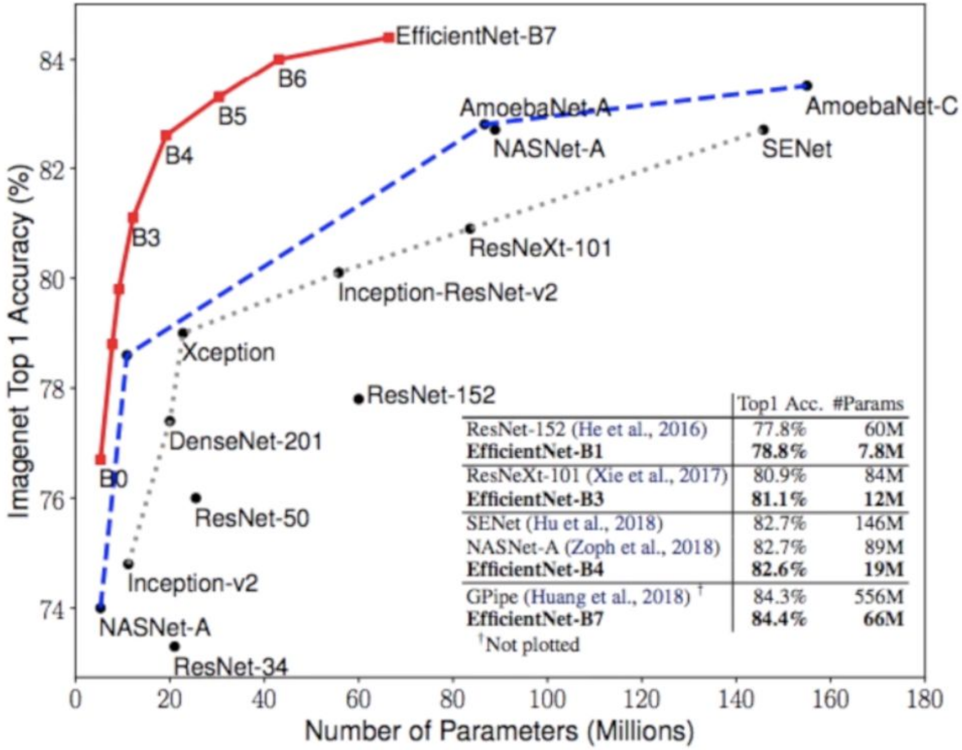
Compound scaling is the way to go!

The idea is to balance the up sampling of width, depth and resolution by scaling them with a constant ratio. Determined by alpha, beta and gamma

$$
\begin{aligned}
\text{depth: } & d = \alpha^{\phi} \\
\text{width: } & w = \beta^{\phi} \\
\text{resolution: } & r = \gamma^{\phi} \\
\text{s.t. } & \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \\
& \alpha \geq 1, \beta \geq 1, \gamma \geq 1
\end{aligned}
$$

What are EfficientNets ?

- ★ Top 5% imagenet accuracy.
- ★ Much more computationally efficient
- ★ Much less parameters



5.

Top Solutions Summary

1st place, 2nd place, 11th place

Best Available Solution - 0.926

Preprocessing: Preprocessing copied from joorarkesteijn's kernel which used ideas from Neuron Engineer's kernel. They used the gaussian blur subtraction method for the DenseNet model only. The images were normalised using appropriate values for the model's pretrained weights.

Models:

- EfficientNet B3 - ord regression - 256px - radius reduction preproc - flip TTA
- EfficinetNet B3 - ord regression - 300px - radius reduction preproc - flip TTA
- DenseNet101 - ord regression - 320px with Ben's preprocessing - no TTA

Loss Function: BCEWithLogitsLoss with modified label smoothing.

Last Pool Layer: generalized mean pooling (better than original average pooling)

Augmentions:

- flip_lr,
- brightness,
- Contrast,
- rotate(360)

Optimiser: Adam

First Place Solution - 0.936129

Preprocessing: Just resizing, nothing special

Models:

```
2 x inception_resnet_v2, input size 512
2 x inception_v4, input size 512
2 x seresnext50, input size 512
2 x seresnext101, input size 384
```

Loss Function: nn.SmoothL1loss()

Last Pool Layer: generalized mean pooling (better than original average pooling)

Augmentions:

```
contrast_range=0.2,
brightness_range=20.,
hue_range=10.,
saturation_range=20.,
blur_and_sharpen=True,
rotate_range=180.,
scale_range=0.2,
shear_range=0.2,
shift_range=0.2,
do_mirror=True,
```

Second Place Solution - 0.933693

Preprocessing: Cropped black background and resized images

Models: EfficientNets :

B3 image size: 300
B4 image size: 460
B5 image size: 456

Loss Function: Mean Square Error (typically used in regression)

Augmentions:

- Blur
- Flip
- RandomBrightnessContrast
- ShiftScaleRotate
- ElasticTransform
- Transpose,
- GridDistortion
- HueSaturationValue
- CLAHE
- CoarseDropout.

Secret Ingredient: pseudo-labeled current test data and then fine-tune networks again using train+pseudolabeled test.

6.

Discussion

& conclusion

Top Methods :

1. EfficientNets
2. InceptionNets
3. DenseNets
4. SeResNext50

Best Loss function ?

- ★ *nn.SmoothL1Loss()*
- ★ Mean Square Error
- ★ WingLoss
- ★ Cauchy loss.
- ★ Huber loss (aka smooth L1 loss)

Related Postings

[Google AI Blog (2016)]

Deep Learning for Detection of Diabetic Eye Disease

[Kaggle Competition (2015)]

Diabetic Retinopathy Detection

[FDA (2018)]

FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems

Thank you!

Any questions? ☻

References

★ https://www.kaggle.com/tanlikesmath/intro-aptos-diabetic-retinopathy-eda-starter
★ https://en.wikipedia.org/wiki/Diabetic_retinopathy
★ https://arxiv.org/abs/1905.11946
★ https://arxiv.org/abs/1512.03385
★ https://arxiv.org/pdf/1608.06993.pdf
★ https://www.kaggle.com/tanlikesmath/intro-aptos-diabetic-retinopathy-eda-starter
★ https://www.kaggle.com/c/aptos2019-blindness-detection/discussion/108065#latest-628246
★ https://www.kaggle.com/c/aptos2019-blindness-detection/discussion/107926#latest-627944
★ https://www.kaggle.com/lextoumbourou/2-x-b3-densenet201-blend-0-926-on-private-lb
★ https://www.kaggle.com/c/aptos2019-blindness-detection/discussion/107947
★ https://arxiv.org/abs/1409.4842
★ https://webeye.ophth.uiowa.edu/eyeforum/tutorials/Diabetic-Retinopathy-Med-Students/Classification.htm
★ https://www.kaggle.com/aroraaman/quadratic-kappa-metric-explained-in-5-simple-steps
★ https://towardsdatascience.com/the-problem-of-vanishing-gradients-68cea05e2625