Grenze International Journal of Engineering and Technology, Jul 2019

Computer Aided Diagnosis System for Diabetic Retinopathy using Deep Learning-based CNN Method

M.K.Kaushik[1], D. Mani Teja[2], V.Naga Mahendra Reddy[3] and S. Pravallika[4]
[1-4]Rajeev Gandhi Memorial College of Engineering and Technology , Nandyal
Email: kaushiknicky24@gmail.com, maniteja.dibbamadugu@gmail.com, varakutimahendra@gmail.com, pravallika.sannala98@gmail.com

***Abstract*—Diabetic Retinopathy (DR) is one of the major causes of blindness in the world. Regular screening of diabetic patients for DR has been shown to be a cost-e□ective and important aspect of their care. The accuracy and timing of this care is of significant importance to both the cost and e□ectiveness of treatment. The diagnosis of diabetic retinopathy through colour fundus images requires experienced clinicians to identify the presence and significance of many small features which, along with a complex grading system, makes this a di□cult and time-consuming task. The main objective of this paper is deep learning-based CNN method is introduced for the problem of classifying DR in fundus imagery. This is a medical imaging task with increasing diagnostic relevance .With the help of Convolutional Neural Networks (CNN) approach to diagnosing DR from digital fundus images and accurately classifying its severity. A network is developed with CNN architecture which can identify the intricate features involved in the classification task such as micro-aneurysms, exudate and haemorrhages on the retina and consequently provide a diagnosis automatically and without user input. This network is trained on the publicly available Kaggle dataset and demonstrate impressive results, particularly for a high-level classification task.**

***Index Terms*— Diabetic Retinopathy, Convolutional Neural Networks(CNN), microaneurysms, exudate, haemorrhages,kaggle dataset, Python.**

I. INTRODUCTION

According to International Diabetes Federation survey 2018 a whopping sum of 425 million people are affected worldwide. Out of which 159 million belong to Western Pacific, 82 million belong to South-east Asia, India is considered to be diabetes capital of the world with diabetic population which is predicted to hit 69.9 million by 2025.One study reveal that almost 75% of the people with DR belong to developing countries. In last few years DR has been the prime reason for blindness, which is growing rapidly, particularly among the adults aged 20-74 years. More than 60% of the patients with type-2 diabetes suffer from DR According to studies performed by Wisconsin Epidemiologic Study on DR, in the younger-onset patient group almost 86%of blindness is caused due to DR.40 to 50 percent of americans who is having diabetes suffer with diabetic retinopathy[1]. Diabetic retinopathy is an eye diease that causes due to damage of light sensitive tissue at the back of the eye. These blood vessels can swell and leak. or they can close and stop blood from passing through blood vessels. It is majorly occurred in people who have been suffering from diabetes for a long period of time for atleast 10 years. In the early stages of diabetic retinopathy, you maynot have the symptoms but as the condition progresses, the symptoms[2] include poor night vision,blurred vision. The diabetic retinopathy can be identified by series of photos. A specialist doctor is required to examine the diabetic retinopathy correctly. The DR includes the

Grenze ID: 01.GIJET.5.2.502_1

following characteristics Hard Exudates, Microaneursysms, Soft Exudate, Hemorrhages, Neovascularization, Macular Edema.

A. Microaneurysm

Generally, formation of microaneurysm is the seed for Diabetic Retinopathy,it occurs due to swelling of blood vessels in eye. Sometimes these blood vessels may leak blood into retina and cause cloudy vision.It appears as a red spot on the retinal image with diameter ranging from 34μm to 50μm.

B. Hard Exudates

It forms due to the leakage of lipoproteins and some other proteins from retinal blood vessels. Visually it looks like white or yellowish-white deposits with sharp margins and are found near Microaneurysms.

C. Soft Exudates

White lesions in Retinal Nerve Fibre Layer are called Soft exudates,these occur due to blockage of arterioles,which result in inadequate blood supply to RNFL,that eventually influences the axoplasmic flow and forms as debris across the retinal ganglion cell axons.

D. Hemorrhages

It occurs due to leakage from weak blood vessels .It appears as a red spot with a dimensional range of 125μm.It is clasified into 2 categories: flame and dot-blot HME,where flame emerges at nerve fibers and dot-blot HME emerges at venous end of capillaries and are round in shape and smaller than MA.

E. Neovascularization

It is the formation of new blood vessels that usually appear on the internal surface of retina. It occurs mostly in the proliferative DR.

F. Hemorrhages

It is identified as swollen part of the retina that usually occurs because of retinal capillaries.It affects vision capability.Most people with macular edema will have symptoms range from slightly blurry vision to noticeable vision loss.

G. Diabetic Retinopathy types:

The diabetic retinopathy includes five stages of classification .1)Stage I:No Diabetic Retinopathy,2)Stage II: Mild Non-Proliferative DR,3)Stage III:ModerateNon-Proliferative DR,4)Stage IV: Severe Non-Proliferative DR,5)Stage V:Proliferative DR.

TABLE I : INTERNATIONAL CLINICAL DIABETIC RETINOPATHY &DIABETIC MACULAR EDEMA DISEASE SEVERITY SCALES[3].

Stage	**Observable Findings**	**Severity**
I	No abnormality	No DR
II	One Micro-aneurysms	Mild non-proliferative DR
III	Any of the following: -Micro-aneurysms -Retinal dot-blot haemorrhages -Hard exudates or cotton wools.	Moderate Non-proliferative DR
IV	Any of the following -more than 20 intra-retinal hemorrhages in each of 4 quadrants -definite venous beading in 2 or more quadrants. -Intra-retinal microvascular abnormalituy in 1 or more quadrants	Severe non-proliferative DR
V	One or both of the following: -Neavascularization -Vitreous hemorrhage	Proliferative DR

II. Block Diagram

Convolutional Neural Networks in image classification have a huge importance. Usually every Neural network consists of 3 major layers namely input layer,hidden layer and output layer.

A. Input layer

It is the layer where the inputs are given to the model,The total number of features in the data decides the total number of neurons in input layer.The main job of input layer is to deal with all the inputs only.Each neuron in the layer has to represent a independent variable which has its influence over the output of the neural network.

B. Hidden layer

The hidden layer is the next layer in neural networks,this layer receives the inputs from the input layer. There may be one or more hidden layers based on the model and data size.The job of hidden layer is to process the inputs that are received from previous layer ,so it is responsible for feature extraction from the input data.

C. Output layer

The output layer is the last layer in the network .It takes outputs from the hidden layers and then feed those outputs into a logistic function like sigmoid or softmax .The use of these activation functions is to convert the output from each class into probability score of each class.

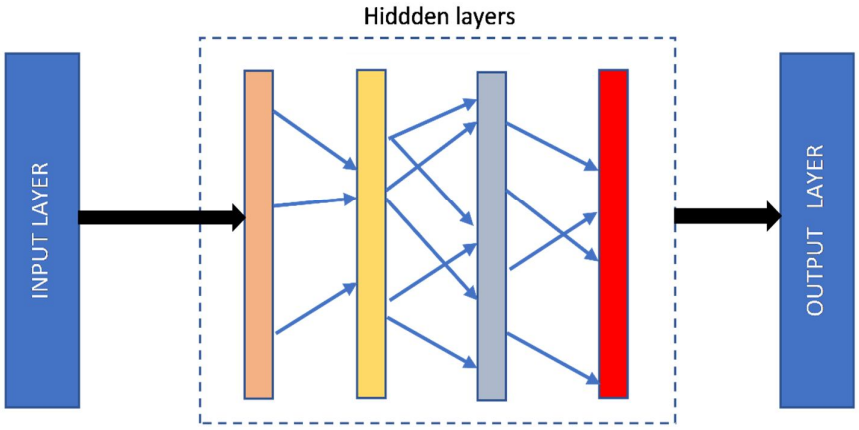

Fig 1 – General representation of a neural network

As already said, there may be any number of hidden layers in the network,The accuracy of the network depends upon the number,type and the combination of Hidden layers.
The block diagram shows different layers in the convolutional neural network,each of the layer performs different tasks on inputs and each layer is better suited for some tasks than others.

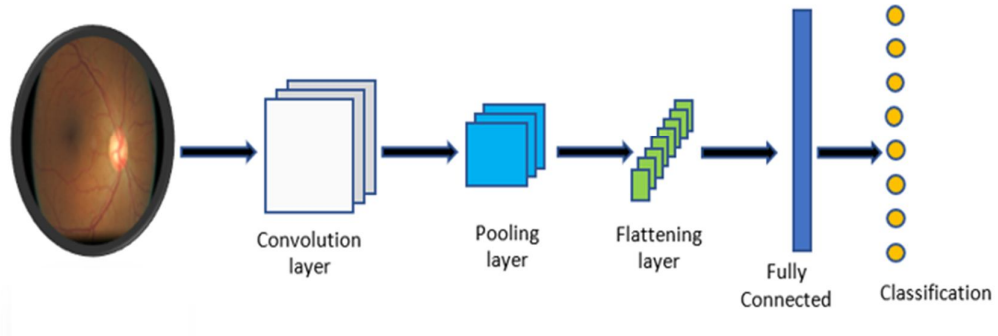

Fig 2 - Schematic representation of Convolutional Neural Network

D. Convolution layer

A Convolutional layer is the most important layer in CNN.Convlution[10] means process of adding each element of the image to its local neighbours weighted by kernal that is, this Convolution layer creates a convolution kernal which is convolved with the input layer and produces an output..The convolution operation can be performed as an element wise matrix multiplication ,where one of the matrices is an image and the other is a filter or kernal that changes the original image into other form ,but it is to be noted that this multiplication is not similar to that of normal matrix multiplication.Convolution layers are usually used to detect patterns in images and are also used to map the identical features between different images.The main aim of convlution layer is to learn feature representation of inputs and to transform the original function into another form to get more information.This layer has a vital role in image processing to blur,sharpen and enhance images.The size of the convolution layer varies in accordance with the input image size and the type of feature.Usually 3X3,5X5,1X1 sized convolution layers are used. In hidden layers only one convolutional layer may be used, or a number of convolutional layers can also be used,it is to be noted that the sizes of all convoutional layers need not be same.

E. Podding layer

This layer is normally implemented to reduce the dimentionality of the image having higher dimensions,thereby reducing the computational complexity.Convolutional layers often followed by these pooling layers.This layer performs some calculations depending upon the type of operation,for instance in Maxpooling layer, it operates on every depth slice of the input matrix and resizes it .

The highest values from a 2X2 filter are taken to reduce the output future map.Similarly when Global average pooling(GAP) layer is used a more extreme type of dimensionality reduction takes place,where a tensor with dimensions h x w x d is reduced in size to have dimensions 1 x 1 x d. GAP layers reduce each h x w feature map to a single number by simply taking the average of all h,w values.

F. Padding

In order to keep the size of the images constant even after performing convolution operation ,padding is used.For this we add zeros at the bordersof the input image before performing any operations.Information loss can be avoided by using padding.there are two types of padding namely Same padding,No padding.In same padding zeros are added to previous image where as for no padding no zeros are added .

G. Dropout layer

Overfitting is the major problem during training ,in order to reduce this effect Dropout concept is implemented.Adding or removing of neurons during the training process is called dropout rate[8].The advantage of Dropsample method is without any extra decision rule dropped samples are selected.

H. Over fitting

Over fitting can be identified based on the metrics that are given for our training data and validation data during training. Validation metrics are worse than the training metrics then our model is over fitting. reducing the over fitting by adding more data to the training set and use data augmentation technique.

I. Under fitting

A model is said to be under fitting when it's not able to classify the data it was trained on. We can tell that a model is under fitting when the metrics given for the training data are poor, meaning that the training accuracy of the model is low and/or the training loss is high. Under fitting can reduce the under fitting by increase the complexity of the model and add more features to the input samples.

J. Data set

The data set was provided by the kaggle data set,this is huge set composed of digital fundus images of retina, assigned to five classes indicating severity of the disease(0- no disease, 5- most severe case).The dataset contains about 35000 high resolution images, most of them are higher than 1800 x 1800 px .In that 900 images are used for five stage classification.The images were taken in different conditions: they vary due to light condition, quality of images.The whole dataset is divided into 3 sets namely train set, validation set, and test sets.Under sampling and over sampling are performed to adjust high imbalance of number of images within the classes.After obtaining even class distribution the images are scaled to have a constant resolution of 512x512px to fit these images at the neural network input.

K. Data Augmentation

Data Augumentation means creating a new data based on modifications to existing data.New data can be created by flipping the images either horizontally or vertically, zoom in or zoom out,or even vary the colour of images.This method can be used when there is limited data set available and also to obtain different variations of same data.

L. Softmax Activation function

This activation function is used for multi-classification,this function calculates the weights of each target class over all the possible target classes.Hence after obtaining the weights of the image,it determines the target class for the given inputs.
The probability range is 0 to 1.

$$F(z_i) = \frac{\exp(z_i)}{\sum_j \exp(z_j)}$$

M. Inception net v3

VGGNet[4] has the simple architecture,this is high computational network, on the other hand inception architecture of google net[5] is designed to perform well even under strict constraints on memory and computations.To achieve that google net utilizes the channel cancatenation.to reduce the mathematical calculations ,use bottle neck(1 x 1) layers reduces the mathematical operations of convolution by a factor of 10.Google Net employed only 5 million parameters,where as in Alexnet thre will be 60 million parameters are employed,VGG net uses 3 x more parameters than Alex Net. Inception Net v3 [6] utilizes factorizing convolution strategies to further to increase the computational efficiency.

III. Proposed Method

Diabetic Retinopathy classification by a computer aided device is going to be a crucial step for the next decades as there is a chance of increase of DR exponentially, An accurate and early detection would be favourable to provide better treatment. Hence a sophisticated and peculiar classifier is needed. The root of obtaining a better classifier lies in the selection of appropriate dataset.Kaggle provides an unbiased dataset,in this paper the retinal fundus images which are used for training and testing are from kaggle dataset[2].This data set consists of a huge collection of around 35,000 of DR affected fundus images with different severity levels that are classifoed by trained clinicians. From the total database 900 images are used for training and validating the neural network.The given dataset has a major problem of class imbalance,that is majority of the images in the data belong to one class and that too negative class (No DR).Hence the neural network will train more for negative class and classifies majority of the testing images to No DR,which is not a satisfactory output.So there is a requirement of managing class-imbalance in the dataset.In order to make the classes equal we use a concept called Data augumentation. Initially before performing Data augumentation the images have to be preprocessed, In prepocessing step all the images in the dataset are cropped and resized to squares of 512 X 512.This is because it is difficult for a network to be trained with different image sizes.As mentioned in section 2 Data augumentation is used to obtain many images from the given limited set of images.The images are obtained by cropping,rotating,flipping,zooming and shrinking the original images.In this paper 900 images are selected from the dataset and class imbalancing problem is overcomed by performing data augumentation.By considering these equally divides images of different classes the model is trained ,with incepetion net v3 architecture.From the balanced dataset[9] of 900 fundus images,the images are distributed between training and validation images with 736 and 164 respectively.Training set consists of 146 images of each class and the network is trained with these training images where as the remaining validation images are used to validate the training process.
At the first layer input images are taken and are given to hidden layers,A well structured architecture is required to obtain better accuracy.As already mentioned,Inception net v3 architecture is implemented for effective training and classification of images.This inception net v3 architecture consists of inception modules which are cascaded with each other and the last layers are removed and implemented as per the requirement.This removing and adding of new layers to the already trained network is called transfer learning.Tranfer learning is said to be the efficient process as it dramatically reduces the computation steps and saves time in training the network.Each layer connects with the adjacent layer with some weights,in transfer learning we use image net[7] weights for training.Training process mainly relies on the type of dataset and epochs.An epoch can be explained as an iteration where each and every image in the training dataset flow through all the network once.The more the

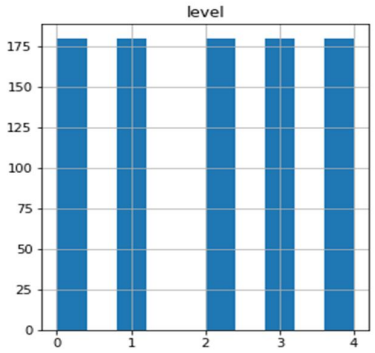

Fig 3 - Bar graph showing balanced data set

number of epochs,better the training process,but there is a stage at which further increase of epochs doesn't alter the training accuracy.

IV. RESULTS

Case study : 1

The following accuracy and loss curves are obtained for a model trained for 10 epochs with 900 images.

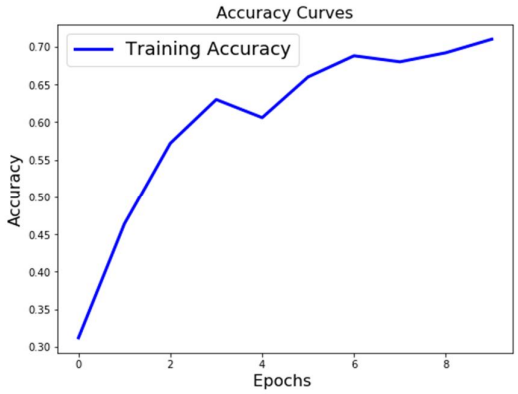

Fig 4 - Training accuracy

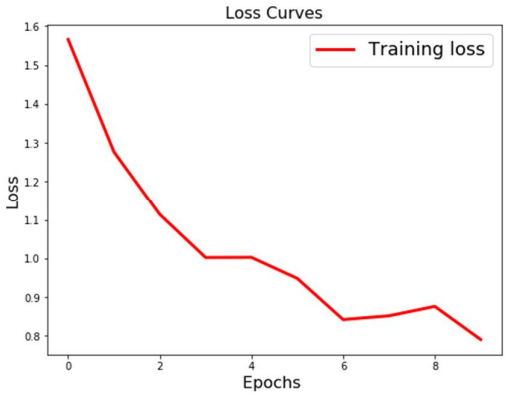

Fig 5 - Training Loss Curve

Case study : 2

The loss and accuracy curves are obtained for a model trained for 5 epochs with 900 images

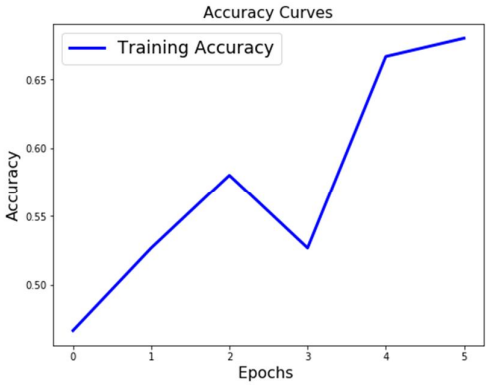

Fig 6 - Training accuracy

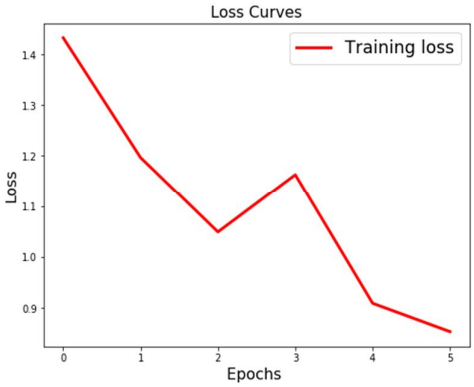

Fig 7 - Training Loss Curve

Classification of images and their probabilities :

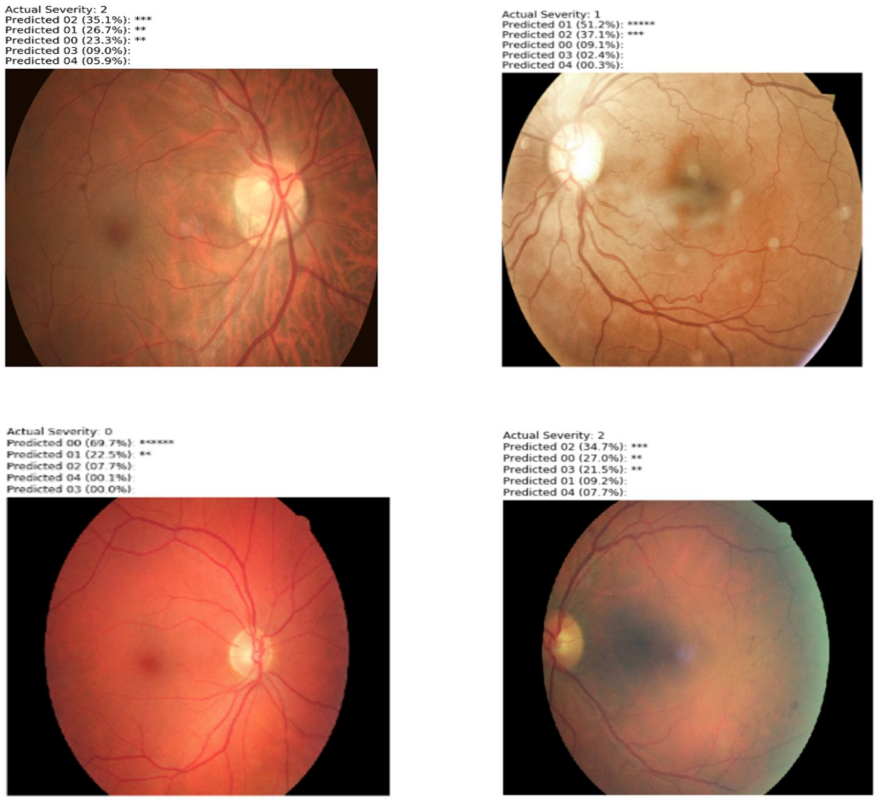

Fig 8 - Result showing Classification probability of an image after training

The accuracy of the training is depended on many factors ,mainly on architecture that is implemented, CPU used, Data set taken for training, Number of epochs etc.From the Case Study1 the maximum accuracy obtained is 66%, this is due to less number of epochs used in training process,where as in Case study 2 the accuracy in training process is 73% ,an increase in the training accuracy can be observed as in this case the number of epochs used is 10.Similarly the loss calculated in training process is different in different case studies .In Case study 1 the loss is more while compared with that of case study 2.But the best trained model should have less prediction loss.The main reason here is also the number of epochs in each case and the similiarity among Data Augumented images produced at the time of training.

V. Conclusion

From the obtained results after training the network with different epochs ,a conclusion can be derived that more the number of epochs more the training accuracy .when trained with 5 epochs which contain 900 fundus images accuracy in determining a class of an image is 66 % where as in case of 10 epochs the accuracy increased to 73 percent. As this model is trained in a regular PC further increase of accuracy is not achieved due to CPU'S limited performance .The accuracy can be furthermore increased by changing the data set, the number of epochs and using high end GPU'S .

References

[1] "Facts About Diabetic Eye Disease | National Eye Institute." [Online]. Available: https://nei.nih.gov/health/diabetic/retinopathy. [Accessed: 09-Feb-2018].

[2] "International Clinical Diabetic Retinopathy Disease Severity Scale." [Online]. Available: http://www.icoph.org/ dynamic/attachments/ resources/diabetic-retinopathy-detail.pdf. [Accessed: 09-Feb-2018].

[3] Xiaoliang Wang1, Yongjin Lu2, “Diabetic Retinopathy Stage Classification using Convolutional Neural Networks” International Conference on Information Reuse and Integration for Data Science 2018 IEEE.
[4] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in Proceedings of International Conference on Learning Representations (ICLR 2015), Sep. 2015
[5] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015), IEEE, Jun. 2015, pp. 19.
[6] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016), IEEE, Jun. 2016, pp. 2818-2826
[7] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009), IEEE, Jun. 2009, pp. 248-255.
[8] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting,” J. Mach. Learn. Res., vol. 15, pp. 1929–1958, 2014.
[9] Kotsiantis S, Kanellopoulos D, Pintelas P. Handling imbalanced datasets: A review. GESTS International Transactions on Computer Science and Engineering, 2006, 30(1): 25-36.
[10] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. pages 1–9, 2014.