Melanoma Detection Using Regular Convolutional Neural Networks

Important
It isn't important
Very important

Aya Abu Ali and Hasan Al-Marzouqi,
Department of Electrical and Computer Engineering,
Petroleum Institute, Khalifa University of Science and Technology.
{aymabuali, hasalmarzouqi}@pi.ac.ae

***Abstract*—In this paper, we propose a method for classifying melanoma images into benign and malignant using Convolutional Neural Networks (CNNs). Having an automated method for melanoma detection will assist dermatologists in the early diagnosis of this type of skin cancer. A regular convolutional network employing a modest number of parameters is used to detect melanoma images. The architecture is used to classify the dataset of the ISBI 2016 challenge in melanoma classification. The dataset was not segmented or cropped prior to classification. The proposed method was then evaluated for accuracy, sensitivity and specificity. Comparisons with the winning entry in the competition demonstrate that one can achieve a performance level comparable to state-of-the-art using standard convolutional neural network architectures that employ a lower number of parameters.**

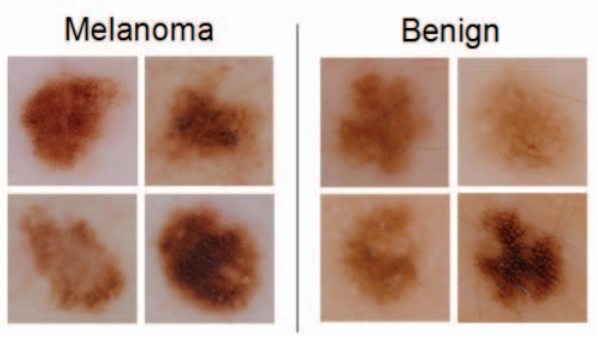

Fig. 1. Melanoma and benign skin images [7]

I. Introduction

Melanoma is a type of cancer that begins in the pigment cells (melanocytes) of the skin. Melanoma disease appears on the skin as pigmented moles or marks. It can also spread to other body organs. Melanoma can be caused by the excess exposure to ultraviolet radiation from the sun. The ultraviolet radiation can damage the cells directly and cause changes in the immune system [1]. Melanoma is considered to be one of the deadliest type of all skin cancers [2]. It accounts for 75 percent of skin cancer deaths [3].

According to the American Cancer Society, if melanoma is detected at an early stage, it can have higher rates of survival [4]. Diagnosing melanoma can be a complex task. An article in the Journal of Investigative Dermatology [5], explains the complexity of the clinical process of diagnosing melanoma, it involves several skin examination components such as anamnestic data (medical history), comparative and differential recognition as well as pattern analysis. The manual detection can be a lengthy process and it requires highly trained dermatologists. The accuracy of dermatologists is estimated to be about 75 to 85% [6].

The fact that melanoma body marks can be confused with normal pigments of the skin makes it hard to classify the skin pigment into benign or malignant. Figure 1 shows melanoma and benign skin images, both pigments seem to be similar. Having an automated algorithm to classify melanoma images will support early diagnosis and help improve melanoma detection performance.

One of the clinical methods used in identifying melanoma is the ABCD rule [8]. The ABCD rule was introduced in 1985 as a method to diagnose melanoma. The ABCD acronym stands for Asymmetry, Border irregularity, Color and Diameter. The letter E was added to the ABCD acronym in 2004 which stands for evolving [9]. Dermatologists use these parameters to identify melanoma. Each criteria has certain features that are recognized to classify melanoma into benign and malignant.

According to [5], the ABCD method, in some cases, wasn't helpful to classify malignant and benign skin moles. In addition, The method couldn't recognize some malignant moles at early stages, for example malignant melanoma with a small diameter. It adds that although clinical methods have improved discrimination, they haven't removed the challenge of differentiating malignant melanoma and nevi (birth marks).

Many learning approaches attempt to extract features from the data images as a step prior to image classification, while CNNs are capable of learning all the details without engineering the features.

Convolutional neural network based automated algorithms have shown effective results in the detection, classification and segmentation of medical imaging. High accuracy results were achieved by [10] in lung pattern classification for interstitial lung diseases and by [11] in gastric carcinoma (stomach cancer) classification. The paper in [12] shows that state-of-the-art results can be achieved using a pre-trained convolutional neural network without the need to use a specific feature extractor to extract the image's features for the classification of melanoma and benign images. The paper uses 170 clinical

978-1-5386-0872-2/17/$31.00 ©2017 IEEE

images resized to 188×188 pixels as their data set. The paper adopts the method of image augmentation to overcome the problem of the small data set by cropping and rotating of the images which increased the available data set to 6120 images. The proposed method starts by pre-processing the images to remove illumination and noise effects by applying illumination correction. The images are further pre-processed by applying a segmentation mask to separate the lesion's region and a Gaussian filter to smooth the area outside the lesion. The pre-trained CNN in this work has two convolutional layers that has a 5×5 filter. Each convolutional layer is followed by a pooling layer and the final layer is a fully connected layer.

An automated model is proposed using both Convolutional Neural Networks and Support Vector Machines by [13]. The paper compares the melanoma data classification performance of Convolutional Neural Networks and Support Vector Machines based on evaluation metrics such as accuracy and sensitivity. Before classification the proposed method applies texture analysis techniques such as Local Binary Patterns (LBP) [14]. The number of images used in total is 206, 150 images used for training and 56 image for testing for both CNN and SVM models.

Lequan Yu *et. al.* [15] developed a model for classifying ISBI 2016 challenge dataset that can be used with and without a segmentation module. The segmentation process was done by using a fully convolutional residual network (FCRN) to provide accurate lesion segmentation. The FCRN used in this paper has 16 residual blocks, each contains two 1×1 convolutional layers and one 3×3 convolutional layer. The classifier used in this work is a combination of two classifiers; softmax classifier and support vector machine (SVM) classifier. The results found from the two classifiers are averaged. Data augmentation such as rotations, shifts were applied to input images. The paper showed that the results of classification with and without segmentation are close. Classification accuracy with segmentation was found to be 85.5% and without segmentation is 82.8%.

The main contribution of this paper is developing a convolutional neural network model for melanoma classification and comparing its performance with other recently proposed models. The developed architecture is relatively simple and requires a modest number of parameters. In this research work we show that using standard convolutional neural network with a few parameters can achieve comparable results in terms of accuracy and specificity.

II. Proposed Solution

The solution proposed in this research work is to adopt a recently developed deep learning framework to classify melanoma images into benign and malignant. The chosen framework for this research is LightNet [16]. In this work, classification of melanoma images to benign or malignant is done by modifying the convolutional neural network architecture in LightNet. Melanoma images are classified without applying lesion segmentation or complex image pre-processing.

A. Image pre-processing

Before the classification process, simple image pre-processing techniques were applied automatically to all the images in the dataset. Melanoma labels are encoded as a either 1 or 0. All the images are resized to a size of 256×256. The data is normalized by subtracting the mean so that the data is zero-centered. Developed architecture uses mini-batch gradient descent [17]. The labels of the training and testing data are rearranged so that their order is balanced to avoid having too many consecutive images of the same label.

B. Convolutional Neural Network Architecture

The Convolutional Neural Network for this work has 17 layers including the fully connected layer and dropout layers. The layers are distributed in 5 blocks. Each block has the following layers: convolutional, ReLU, pooling and dropout. Except for the last block which contains the fully connected layer and the softmax layer for classification.

Overall the network has 5 convolutional layers, 4 ReLU layers , 3 max pooling layers and 4 dropout layers.
The Convolutional Neural Network Architecture is shown in figure 2.

The filter (kernel) sizes for the convolutional layers are as follows: The first three convolutional layers are 5×5 with padding of 2, the fourth convolutional layer is a 4×4 and the last convolutional layer is 1×1. All the convolutional layers have a stride of 1. The first convolutional layer has 32 kernels and the rest of the convolutional layers have 64 kernels except for the last layer which has 2 kernels. The weights for the filters are initialized using normal random numbers with zero mean and standard deviation of one. The weights are multiplied by one over the square root of the size of the input image and the number of convolutional kernels in each convolutional layer. The initialization method used can be described by the following equation:

$$W_{ij} = U[\frac{-1}{\sqrt{n}}, \frac{1}{\sqrt{n}}], \quad (1)$$

Where W is the wieghts at each layer, U is a uniform distribution on a specific interval and n is the previous layer's size [18]. For example in the first layer the input image has the size of 256×256 and the convolutional layer has 32 kernels, so the number multiplied by the random initialized numbers is $1/\sqrt{256 \times 256 \times 32} = 1/1448$.

The pooling layers have a stride of 4 and the dropout layers have a hyper parameter P, called the dropout rate which determines the amount of dropped out units. The dropout rate used in the first four blocks is 0.1, 0.2, 0.3, and 0.5 respectively.

The output layer used for classification is the softmax classifier. The predicted output is then compared with the actual output and the difference between the predicted output and the actual output is calculated as an error. This error is reduced using the gradient descent algorithm during the back propagation process. In the gradient descent based learning the

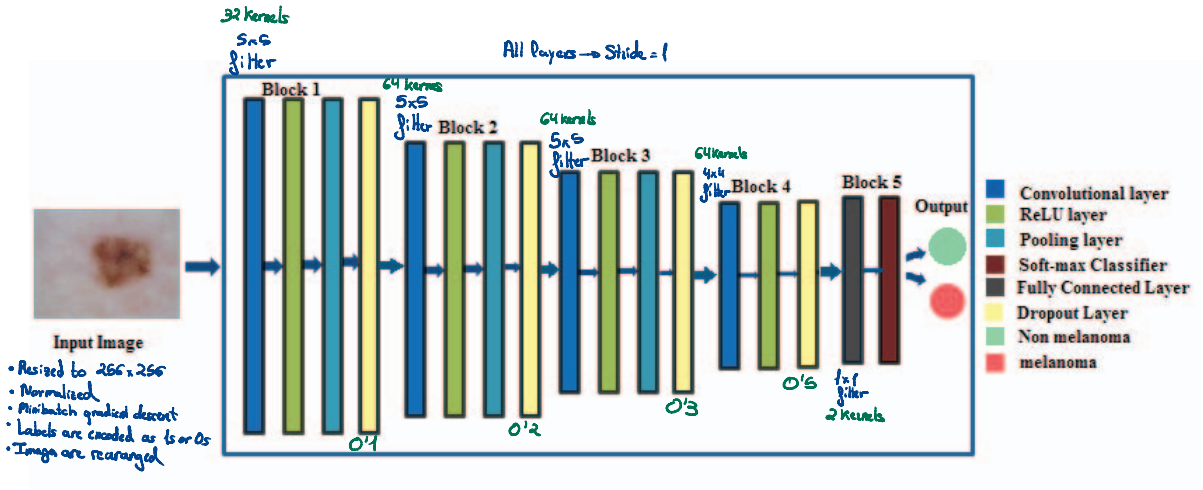

Fig. 2. Proposed convolutional neural network architecture

output function that consists of the weight, input and the bias is computed. A loss function measures the difference between the correct output and the produced output by the algorithm. The aim of the learning process is to obtain a value of the weights that minimizes the loss function.

$$W_{K} = W_{K-1} - \epsilon \frac{dE(W)}{dW} \quad (2)$$

Where W_{K} is the weight to be updated, W_{K-1} is the previous weight, ϵ is the learning rate and $\frac{dE(W)}{dW}$ is the derivative of the loss or error function with respect to the weight parameter. This repeatedly updates the value of W until the algorithm converges. In the back propagation process, the gradients are efficiently computed from the output to the input [19].

The paper in [20] was able to achieve state-of-the-art results in classification of 1.2 million images in a contest of the ImageNet LSVRC-2010 with Deep Convolutional Neural Networks using back-propagation. The deep convolutional neural network was trained to classify the dataset the into 1000 different classes. The network consists of five convolutional layers and three fully-connected layers. After the last fully-connected layer a 1000-way softmax is placed that outputs a distribution over the 1000 class label. The test error results of this paper achieved top-1 and top-5 error rates of 37.5% and 17.0% surpassing previous state-of-the-art results.

The softmax classifier is used for classification. The softmax classifier comes from the concept of the softmax function which outputs prediction probabilities ranging from 0 to 1 [21]. The equation of the softmax function is:

$$y_i = \frac{e^{x_i}}{\sum_{j=1}^{C} e^{x_j}} \quad (3)$$

Where x is the input to be mapped to the output probability y. The denominator is the sum of all the possibilities of the input x and C is 2 which is the number of classes.

The cross-entropy loss between the predictions o or the computed probabilities and the actual output y is used as the loss-function. It is given by [22]

$$f(y, o) = -\sum_{j=1}^{C} y_i \log(o), \quad (4)$$

Learning rate is adapted using the adaptive gradient algorithms (Adagrad) [23].

III. Experimental Setup and Results

A. Dataset

The dataset used for melanoma detection is from the International Symposium on Biomedical Imaging (ISBI) 2016 challenge: Skin Lesion Analysis Towards Melanoma Detection. The images are provided by the International Skin Imaging Collaboration (ISIC) [7]. The ISIC has one of the largest collections of skin dermoscopy images. The dataset for lesion classification consists of 900 training and 379 testing dermoscopic images with labels of benign and malignant. This dataset has 727 benign images and 173 melanoma images in the training subset, as for the testing subset there are 304 benign images and 75 melanoma images. The images' original sizes vary from 1022×767 to 4288×2848 pixels. Each image has a label of either melanoma or non-melanoma (1 for melanoma and 0 for non-melanoma). The program used for classification is Matlab on HP elitebook laptop without using GPU, using a 2.60 GHz CPU and a memory of 8 GB.

B. Implementation Details

As the size of the input images increases, the number of layers of the convolutional neural network needs to increase. The size of input image's was set to 256×256. Since the training dataset is limited the network could overfit. To avoid over-fitting and to make sure that the network converges, the size of the convolutional neural network is adjusted so that there are adequate number of layers in the network. The network is adjusted by changing the filter size of the convolutional layer, the stride and the pooling layers. In addition to manual network's adjustment, dropout layers are added after the pooling layers. The network is constructed

using 17 layers organized in 5 blocks as mentioned in the previous section.

The hyperparameters of the network are set according to the dataset's size and computer's capabilities. The success of the proposed architecture depends on correct setting of network hyperparameters. Examples of our network hyperparameters include: batch size which is the number of training images in one forward pass or backward pass, number of epochs which is one forward and backward pass of all the training examples, and learning rate which is the amount the weights are updated in the network [24].

The batch size is set to 5, as there are only 900 training images. Increasing the batch size requires a large memory. The number of epochs computed is dependent on the training data. Early stopping terminates the algorithm when the training error goes below 15%. The learning rate is set to $1 \times e^{-2} = 0.1353$. The weight decay parameter in the Adagrad algorithm is set to $1 \times e^{-2} = 0.1353$ as well. Hyperparameter values are summarized in table I.

TABLE I
SUMMARY OF HYPERPARAMETER VALUES

Hyper Parameter	Value
Batch Size	5
Learning Rate	$1 \times e^{-2} = 0.1353$
Weight Decay	$1 \times e^{-2} = 0.1353$
Dropout Layer Rate	[0.1,0.2,0.3,0.5]

C. Classification Results

After fixing the hyper parameters and the neural network layers, the network's performance is evaluated and a graph of the testing error verses the number of epochs is plotted. The graph in figure 3 shows the testing error of 0.189. This error rate appears when using the early stopping criteria. This accuracy rate was accompanied with a sensitivity value of 14.86% and a specificity of 98%. Table II summarizes our results. Comparisons are made with two variants of the CUMED algorithm [15]. The CUMED algorithm achieved the best result in ISBI's 2016 skin lesion classification challenge. Our results are comparable to CUMED in terms of accuracy. Including a segmentation part is likely to improve the performance of the proposed method. We also achieve very high specificity values. However, our sensitivity is low.

Proposed model is simpler than CUMED, does not use data augmentation, and uses a significantly lower number of parameters. This reduction in the number of parameters can be crucial in mobile applications where constraints on the size of the network and energy consumption affect the utility of developed tools.

IV. FUTURE WORK

In the future, we plan to improve the proposed method by employing parts of the ABCD rule to extract useful input features. A neural network that takes these features as input

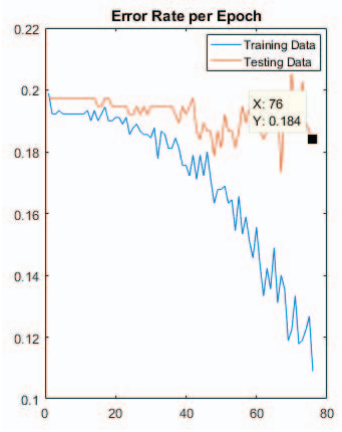

Fig. 3. Training and testing error rates vs. the number of epochs

can be implemented to enhance detection accuracy. The dataset involved in this work has a class imbalance problem because there are more benign class samples than malignant. A simple way to handle this problem is to perform oversampling of the data in which the minority class examples are duplicated. Another way is to use cost-sensitive learning which incorporates the misclassification costs between classes. For example, one of the techniques of this learning method is the application of misclassification costs as a form of data weighting to the imbalanced dataset [25].

Since the data set is limited and there are few training examples, data augmentation can be applied to the data set to increase the number of training samples.

V. CONCLUSION

In this work, a method for melanoma classification based on Convolutional Neural Networks is proposed. The method proposed uses a deep learning framework LightNet to classify melanoma images into benign and malignant. The convolutional neural network architecture in LightNet is modified to work with the ISBI skin challenge data set. Melanoma classification is performed without applying lesion segmentation or complex image pre-processing. Our results are comparable to state-of-the-art results while using a significantly lower number of parameters.

REFERENCES

[1] Deevya L. Narayanan MPH, CPH,Rao N. Saladi MD,Joshua L. Fox MD, FAAD, *"Review: Ultraviolet radiation and skin cancer"*,International Journal of Dermatology,Volume 49, Issue 9 September 2010 Pages 978986.

[2] Aswin.R.B, J. Abdul Jaleel, Sibi Salim.*"Implementation of ANN Classifier using MATLAB for Skin Cancer Detection"*,International Journal of Computer Science and Mobile Computing,December 2013, pg. 87-94.

[3] Jerant, Anthony F.; Johnson, Jennifer T.; Demastes Sheridan, Catherine; Caffrey, Timothy J. *"Early Detection and Treatment of Skin Cancer."*,American Family Physician . 7/15/2000, Vol. 62 Issue 2, p357. 16p. 8.

[4] The American Cancer Society medical and editorial content team *"Survival Rates for Melanoma Skin Cancer, by Stage"*, Available at:[https://www.cancer.org/cancer/melanoma-skin-cancer/detection-diagnosis-staging/survival-rates-for-melanoma-skin-cancer-by-stage.html],May 19, 2016

TABLE II
RETRIEVAL PERFORMANCE IN TERMS OF ACCURACY,SENSITIVITY,AND SPECIFICITY

Method	Accuracy	Sensitivity	Specificity	Number of model parameters
CUMED with segmentation [15]	0.855	0.547	0.931	779264
CUMED without segmentation [15]	0.828	0.427	0.927	389632
Proposed solution	0.816	0.149	0.980	221890

[5] Ashfaq A. Marghoob1,Alon Scope *"The Complexity of Diagnosing Melanoma"*,Journal of Investigative Dermatology (2009), 129, 1113. doi:10.1038/jid.2008.388.

[6] Lin LiEmail author, Qizhi Zhang, Yihua Ding, Huabei Jiang, Bruce H Thiers and James Z Wang, *"Automatic diagnosis of melanoma using machine learning methods on a spectroscopic system"*,BMC Medical Imaging,13 October 2014.

[7] International Symposium on Biomedical Imaging (ISBI) 2016 challenge *"ISBI 2016: Skin Lesion Analysis Towards Melanoma DetectionPart 3: Lesion Classification"*, available at:[https://challenge.kitware.com/phase/5667455bcad3a56fac786791].

[8] Naheed R. Abbasi, Helen M. Shaw,Darrell S. Rigel, Robert J. Friedman, William H. McCarthy, Iman Osman, Alfred W. Kopf, David Polsky , *"Early Diagnosis of Cutaneous Melanoma Revisiting the ABCD Criteria"*,American Medical Association ,Vol. 292, No. 22,December 8, 2004.

[9] Anna Skripnik Lucas, Esther Chung, Michael A. Marchetti, Ashfaq A. Marghoob , *"A guide for dermatology nurses to assist in the early detection of skin cancer"*,Journal of Nursing Education and Practice 2016, Vol. 6, No. 10 June 2, 2016.

[10] Marios Anthimopoulos, Stergios Christodoulidis, Lukas Ebner, Andreas Christe, and Stavroula Mougiakakou *"Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network"*,IEEE Transactions on Medical Imaging, Vol.35, Issue: 5, May 2016.

[11] Harshita Sharmaa, Norman Zerbe b, Iris Klempert b, Olaf Hellwicha, Peter Hufnagl b *"Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology"*,Computerized Medical Imaging and Graphics, The International Journal on Imaging and Image-Computing in ALL Medical Specialties, June 8, 2017.

[12] E. Nasr-Esfahani, S. Samavi, N. Karimi, S.M.R. Soroushmehr, M.H. Jafari, K. Ward, K. Najarian *"Melanoma Detection by Analysis of Clinical Images Using Convolutional Neural Network"*,38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, Florida, USA, 2016.

[13] Esra Mahsereci Karabulut, and Turgay Ibrikci *"Texture analysis of Melanoma Images for Computer-aided Diagnosis"*,Annual Int'l Conference on Intelligent Computing, Computer Science and Information Systems (ICCSIS-16) April 28-29, 2016 Pattaya (Thailand).

[14] Pietikinen, M., Hadid, A., Zhao, G., Ahonen, T, *"Chapter Local Binary Patterns for Still Images"*,Computer Vision Using Local Binary Patterns Book, 2011.

[15] Lequan Yu, Hao Chen, Qi Dou, Jing Qin, and Pheng-Ann Heng. *"Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks"*, DOI 10.1109/TMI.2016.2642839, IEEE Transactions on Medical Imaging.

[16] Chengxi Ye, Chen Zhao, Yezhou Yang, Cornelia Fermller, Yiannis Aloimonos *"LightNet: A Versatile, Standalone Matlab-based Environment for Deep Learning"*, available at:[https://github.com/yechengxi/LightNet/tree/master/Documentations].

[17] Hinton, Geoffrey, Nitish Srivastava", and Kevin Swersky. *"Overview of mini-batch gradient descent."*, Neural Networks for Machine Learning (2012). APA

[18] Xavier Glorot, Yoshua Bengio *"Understanding the difficulty of training deep feedforward neural networks"*,13th International Conference on Artificial Intelligence and Statistics (AISTATS) 2010, Chia Laguna Resort, Sardinia, Italy. Volume 9 of JMLR.

[19] J. KivinenM. K. Warmuth*"Relative Loss Bounds for Multidimensional Regression Problems"*, Proceedings of the IEEE 86, 11 (1998), 22782324.

[20] Krizhevsky, A., Sutskever, I., Hinton, G.*"ImageNet classification with deep convolutional neural networks"*, In Proc. Advances in Neural Information Processing Systems 25 10901098 (2012).

[21] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P.*"Gradient-based learning applied to document recognition."*, Vol., Issue 3, pp 301329, December 2001.

[22] Li Wan,Matthew Zeiler, Sixin Zhang,Yann LeCun,Rob Fergus*"Regularization of Neural Networks using DropConnect"*, Proceedings of the 30 th International Conference on Machine Learning, Atlanta, Georgia, USA, 2013. JMLR:W and CP volume 28.

[23] Sebastian Ruder *"An overview of gradient descent optimization algorithms"*, available online:[https://arxiv.org/pdf/1609.04747.pdf], 15 June 2017.

[24] Adri'a Romero Lopez *"SKIN LESION DETECTION FROM DERMOSCOPIC IMAGES USING CONVOLUTIONAL NEURAL NETWORKS"*,BioMed 2017 paper available at :[https://imatge.upc.edu/web/sites/default/files/pub/xRomero-Lopez.pdf]

[25] Vaishali Ganganwar *" An overview of classification algorithms for imbalanced datasets"*, International Journal of Emerging Technology and Advanced Engineering,ISSN 2250-2459, Volume 2, Issue 4, April 2012.

Matthieu Devin, Quoc V. Le, Mark Z. Mao, MarcAurelio Ranzato, Andrew Senior, Paul Tucker, Ke Yang, Andrew Y. Ng *"Large scale distributed deep networks"*, available online:[http://papers.nips.cc/paper/4687-large-scale-distributed-deep-networks.pdf]