Understanding Capsulenets

Yashasvi Baweja
Computer Science Department
Indraprashta Institute of Information Technology, Delhi
Delhi, India
yashasvi15116@iiitd.ac.in

***Abstract*—Recently there has been a new line of research coming up with the advent of the paper by Sarbour et. al [1]. The paper introduces a new concept called Capsulenets which leverages the limitations of present day Convolutional Neural Networks(CNNs). This project aims further to deepen the understanding for Capsulenets and testing their efficacy for larger datasets.**

***Index Terms*—Capsulenets, CNNs, Equivariance**

I. Introduction

Present day CNNs are performing well with the many real world problems and giving state-of-the-art results on many use cases. Working of a CNN can be described as:

“The model takes input an image, the first layer convolves it with some kernel function (which act as the weights of the neural network). This convoluted image is then fed into subsequent layers and thus making the model deeper and more sophisticated.”

Although, CNNs do offer a compelling case for achieving human like vision, but such novel idea still lacks something which is called ”equivariance”. How our brain perceives an image is completely defined by the properties of that object. We can still recognize the object even if it is slightly ocluded, or if the object is at some angle or lacks illumination. Our brain has this innate quality of ”equivariance” which helps us in recognizing the object even if it is presented to us in different shape, size, angle or illumination. Such a quality is absent from CNNs. A simple CNN trained to identify person from a gallery with pose variation also becomes a tedious task(ex. Multipie face dataset trained from scratch). Now if we could somehow encode this equivariance in our present CNN models, then we could achieve similar to how our mind perceives an object.

Most of the work has been done on introducing Capsulenets, and trying them on multiple datasets but there is still some scope for thorough analysis of the time taken for training Capsulenets. As Sabour et. al [1] mentions in their future work, the present model takes a lot of parameters, and also takes a lot of time to train the model that too for smaller datasets like CIFAR10 and MNIST. Thus aim of this project would be to reduce the number of parameters for training and improving the speed of the model by using Dropout [2] on capsules and Dropconnect [3] on the routes.

II. Related Work

Wang et. al [4] formulate the equation given by Sabour et. al [1] as a cluster like loss equation subject to minimization. Also there is a recent paper by Hinton et. al [5] which uses Expectation Maximization for routing between the capsules.

III. Dataset

Since the main focus of this project would be to improve the speed of the model, MNIST and CIFAR10 datasets would be used. This would help in giving an accurate comparison with the proposed approach [1].

IV. Timeline

- September End: Use Dropout on capsules
- October Mid: Use Dropconnect on Routes
- November Mid: Use Capsulenets for modelling faces.

References

[1] S. Sabour, N. Frosst, and G. E. Hinton, “Dynamic routing between capsules,” in *Advances in Neural Information Processing Systems*, 2017, pp. 3856–3866.

[2] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[3] L. Wan, M. Zeiler, S. Zhang, Y. Le Cun, and R. Fergus, “Regularization of neural networks using dropconnect,” in *International Conference on Machine Learning*, 2013, pp. 1058–1066.

[4] D. Wang and Q. Liu, “An optimization view on dynamic routing between capsules,” 2018. [Online]. Available: https://openreview.net/forum?id=HJjtFYJDf

[5] G. E. Hinton, S. Sabour, and N. Frosst, “Matrix capsules with em routing,” 2018.