Date ___/___/___

Topic: No Routing needed between capsules 12/11 June 2023

Abstract:

More capsule network designs rely on traditional matrix multiplication between capsule layers on computationally expensive routing mechanisms to deal with the capsule dimension entanglement that the matrix multiplication introduces. By using Homogeneous vector capsules (HVCs) which use element wise multiplication rather than matrix multiplication, the dimensions of the capsules remain untangled. In this work, we study HVCs are applied to highly structured MNIST dataset in order to produce a direct comparison to the capsule research direction of Geoffrey Hinton, et al. In our study, we show that a simple convolutional neural network using HVCs performs as well as the prior best performing capsule network on MNIST using 5x5 fewer parameters, 4x fewer training epochs, no reconstruction subnetwork and requiring no routing mechanism. The addition of multiple classification branches to the network establishes a new state of the art for MNIST dataset with an accuracy of 99.87% for an ensemble of these models, as well as establishing a new state of the art for single model. (99.83% accurate)

Conclusion:

Page No. ___

In this work, we proposed using a simple convulational neural network and established design principles on a basis for a network architecture. When we presented a design that branched out of the series of stacked convlusions at different points to capture different levels of abstraction and effective receptor fields and from these branches, rather than flattening to individual scalar neurons, used homogenous vector capsules instead.
We also investigated three different methods of merging the output of the branches back into a single set of logits. Each of the three merge strategies generated models that could be ensembled to create new state of the art results.
Beyond the network architecture, we proposed a robust and domain specific ~~area~~ data augmentation strategy aimed at simulating a wider variety of renderings of the digits.
In doing this work, we established new MNIST state of the art accuracies for both a single models and an ensemble.
In addition to the network design and augmentation strategy, the ability to use an adaptive gradient descent method allowed us to achieve this on consumer hardware and was an enabling factor in both initial explorations and the training of all 322 trials of experiments referenced in this work.