ID2223 – Scalable Machine Learning and Deep Learning

Project proposal - Capsule networks

Course Coordinators:
Amir H. Payberah
Kamal Hakimzadeh
Jim Dowling

Group *ScalableGroup*

Mohamed Gabr
Óttar Guðmundsson

2018-11-20

Problem description: In the past years, remarkable advancements in image recognition have been achieved using Convolutional Neural Networks. These networks are fed parts of images, through multiple layers, where higher correlational features are extracted and analyzed. By training it so in a supervised way, a set of convolutional and fully connected layers can achieve up to 95% accuracy on popular image-based datasets like ImageNet.
However, CNN's are not as good as they seem. As they have been trained to find features in images, they only capture the feature itself, but not it's relativity to other features. Take for example a face. A face has two eyes, a mouth, and a nose. If you would swap out the position of the eyes for the nose and mouth, a CNN would still classify it as a face even though it clearly isn't. Also, if the image would be flipped upside down the CNN would also have problems classifying it. In short, CNN cant' capture relativity of extracted features.

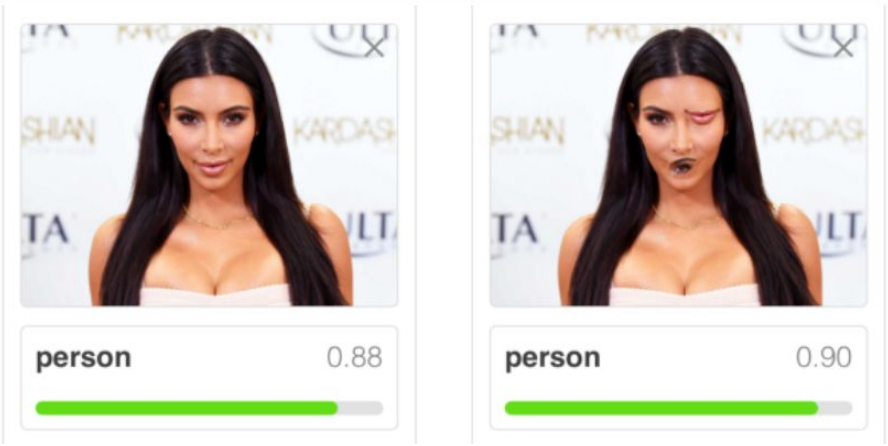

Figure 1: A pretty horrific "face" classified by a CNN

Tools: We plan on using Tensorflow for this problem. We haven't decided if we'll deliver our solution in a Jupyter notebook or as a folder of python classes, but we'll use Jupyter for early sketches of our solution. Furthermore, we might use external libraries such as Numpy for array manipulations/preprocessing and Keras if needed.

Data: The data has not been decided, but our first idea is to manage to train the Capsule networks on the MNIST dataset (the hello world of ML/NN). If we manage to get it working, we will extend the idea on to Cifar10, Cifar100 and then maybe if we have the time (and the computational power) on ImageNet. It would also be interesting to search for a dataset that has images of 3D models and test our method on it, and then compare our results to Alexnet's results that we wrote in lab2. Note that capsule networks tend to need way fewer data points to train on the CNN, we might reflect on this in our results.

Figure 2: 3D models from a different perspective, classified correctly by Capsule networks.

Methodology and algorithm: We plan on implementing Capsule networks, by following the paper mentioned in Figure 3. This will require us to read and analyze this new paper and implement dynamic routing between capsules, primary caps and feature caps for the datasets that we plan on using.

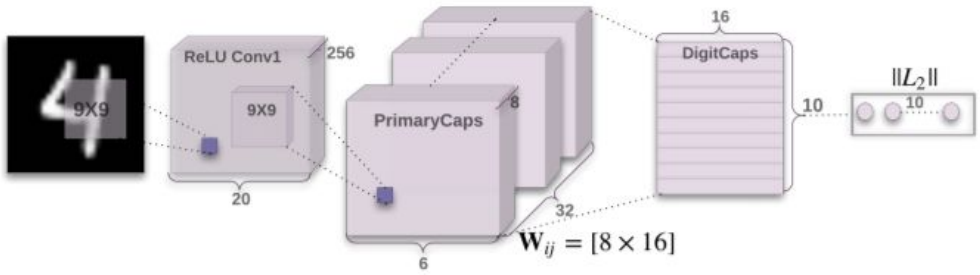

Figure 3: Architectural structure of a capsule network as proposed in the original paper by Hinton et al. 2017 (https://arxiv.org/pdf/1710.09829.pdf)

We think that this project is really valid for the course. Not only are we reaching out for a new topic that is beyond the scope of the material taught, but this also demonstrates our passion and enthusiasm for new concepts in the field. There is a lot of hype about this new architecture and if the hype tends to be real, we want to be knee deep into the concept as soon as possible.