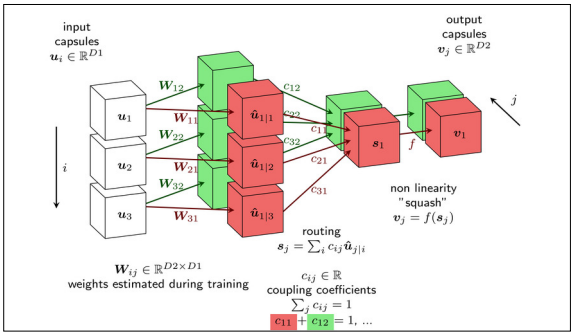
Jonas
Schmid

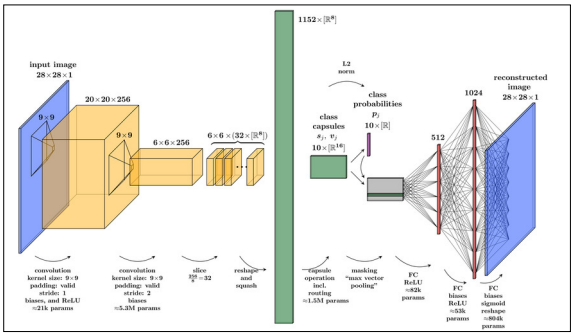
Graduate Candidate	Jonas Schmid
Examiner	Prof. Dr. Guido Schuster
Co-Examiner	Prof. Dr. Aggelos K. Katsaggelos, Northwestern University
Subject Area	Sensor, Actuator and Communication Systems
Project Partner	Image and Video Processing Laboratory (IVPL), Northwestern University, Evanston, IL, USA

Capsule Network

Training and Analysis of a new type of Neural Networks



Computational flow of a hypothetical dense capsule layer with 3 input and 2 output capsules.
Own presentment

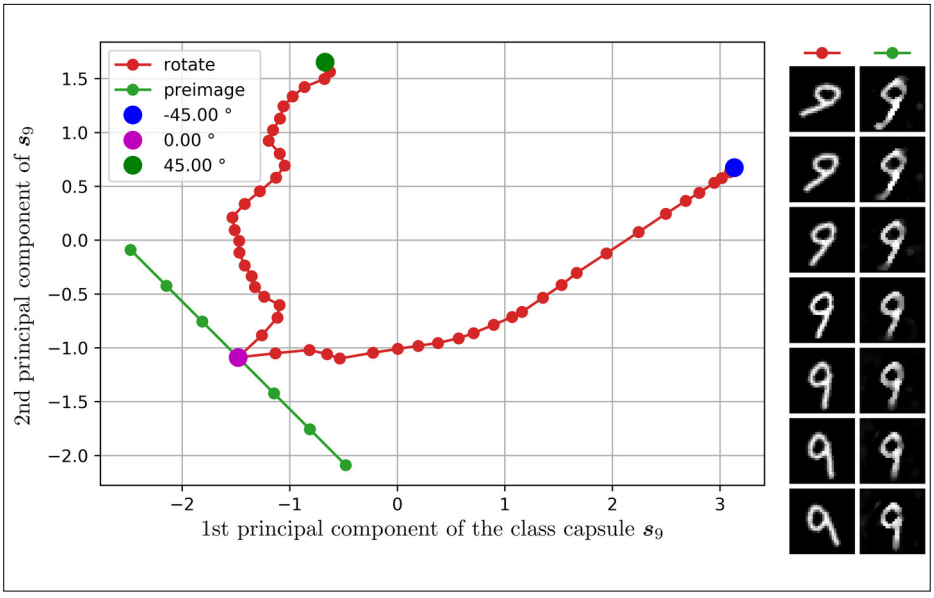


The analyzed capsule network architecture suggested by Sabour, Hinton and Frosst in "Dynamic Routing Between Capsules".
Own presentment

Introduction: In the last years, deep learning, specifically deep convolutional neural networks (CNNs), have outperformed and replaced several classical image processing algorithms. Recently Geoffrey Hinton introduced a new kind of neural networks, the so called capsule networks, suppose to enhance the power of deep learning further. The main idea is, to group similar features, and consider the relationship of different groups of features, in order to alleviate the Picasso problem. A so called capsule describes not only the existence of such a group, like a simple neuron does, it also describes the differences within a group. In other words a capsule indicates the existence of a generic feature and its attributes.

Objective: First of all, the idea and functionality of this new type of neural network will be analyzed on a theoretical basis. Then we are implementing the capsule network architecture with TensorFlow as introduced by Sabour, Hinton and Frosst in ``Dynamic Routing Between Capsules". The major goal is to experimentally analyze the behavior of the capsules, the impact of the routing algorithm and the ``reconstruction as regularization method". Among others, we are using activation maximization and a more general pre-image technique, both tools from the field of explainable artificial intelligence (XAI).

Result: We have successfully implemented a capsule network architecture that achieves a similar accuracy to the original paper. And we could show, that the class capsules at the output of the network, indeed encode the differences within a class, i. e. the output capsules behave as proposed by Hinton. Specifically, simple classic image processing transformations on the input image are encoded in a rather comprehensible way. We were also able to control the pre-image in a similar way as with the classic image processing transformations. But to enforce this behavior ``reconstruction as regularizer" seems to be essential, at least for simple datasets like MNIST.



A rotating '9' leads to the red curve in the capsule vector space. The pre-images of the capsule vectors corresponfing to the green dots show a similar transformation.
Own presentment