Around the clock - capsule networks and image transformations

Maximilian Schmidt\textsuperscript{1,*}
\textsuperscript{1} Center for Industrial Mathematics, University of Bremen, Bibliothekstr. 5, 28359 Bremen, Germany

Capsule networks are a promising new model class for image processing tasks. We investigate and compare their classification performance with convolutional neural networks on a new entry level image data set. In our experiments the capsules are able to learn the effect of rotated input images from data and outperform a comparable convolutional architecture. The results also show that convolutional and capsule networks need structural adjustment to respond to transformations that are not included in the training set.

© 2021 The Authors Proceedings in Applied Mathematics & Mechanics published by Wiley-VCH GmbH

1 Introduction

In the last few years, convolutional neural networks (CNN) have become the state-of-the-art for many image processing tasks. Recently, capsule networks \cite{1} were introduced as a natural extension of CNNs. Capsules are supposed to have several equivalents in the brain and consist of an affine transformation followed by a non-linear function.

Definition 2.1 (Neuron) A neuron \( \mathcal{N} : \mathbb{R}^d \rightarrow \mathbb{R} \) is a parametrized mapping of the input \( x \in \mathbb{R}^d \)

\[
\mathcal{N}(x) = \varphi \left( w^T x + b \right),
\]

where \( w \in \mathbb{R}^d \) is a weight vector, \( b \in \mathbb{R} \) the bias and \( \varphi : \mathbb{R} \rightarrow \mathbb{R} \) is a non-linear activation function. Capsules can be defined as a generalization of neurons.

Definition 2.2 (Capsule) Let \( X := \{ x^{[i]} \in \mathbb{R}^d \mid i = 1, ..., I \} \) be a set of input vectors. A capsule \( C : \mathbb{R}^{d \times I} \rightarrow \mathbb{R}^m \) applies affine transformations followed by a non-linear routing process \( \nu : \mathbb{R}^{m \times I} \rightarrow \mathbb{R}^m \)

\[
C(X) := \nu \left( \left\{ v_i^{[i]} := W^{[i]} v^{[i]} + b^{[i]} \mid i = 1, ..., I \right\} \right).
\]

Here \( W^{[i]} \in \mathbb{R}^{m \times d} \) is a transformation matrix and \( b^{[i]} \in \mathbb{R}^m \) the bias vector for input \( i \). The intermediate results \( v_i^{[i]} \in \mathbb{R}^m \) are called votes. The routing process \( \nu \) combines the votes to generate the capsule vector and can be understood as a clustering.

A layer of a neural network generally consists of multiple neurons or capsules. Definition 2.2 allows for an easy adaption of popular layers, like fully connected and convolution layers, to the capsule setting \cite{1}.

The exact choices of the routing process and the parameter spaces are free. This is beneficial for the representative power of the network. Nevertheless, it does not give any mathematical guarantees how the network will react to transformations of the input. To get an equivariant network, i.e. that the output changes in a predictable manner with the input, one has to limit these choices based on the transformation class. There exist several derivations for equivariance restrictions for CNNs and capsules \cite{2}. However, the transformation class must be known in advance and follow a specific (group) structure.

2 Capsule networks and image transformations

One of the main building blocks of artificial neural networks are the so called neurons. They are inspired by their biological equivalents in the brain and consist of an affine transformation followed by a non-linear function.

3 CloCk data set

The Classify on the Clock (CloCk) \cite{3} data set is composed of 43,200 synthetic images of an analog clock. Each time step between 00:00:00 and 11:59:59 is represented as a single image. The data set contains versions with different levels of detail for each image. In this paper we use the "full" version, which includes markings and numbers (cf. Figure 1).

The movement of the hands is connected by a linear relationship. By introducing an additional rotation \( \omega \), the forward system for the determination of the angles \( \alpha \) of the hands becomes non-linear.

* Corresponding author: e-mail maximilian.schmidt@uni-bremen.de

Proc. Appl. Math. Mech. 2020;20(1):e202000179. © 2021 The Authors Proceedings in Applied Mathematics & Mechanics published by Wiley-VCH GmbH
Definition 3.1 (Rotated clock system) Let $n_{sec}, n_{min} \in \{0,\ldots,59\}, n_{hour} \in \{0,\ldots,11\}$ be the current time point and $\omega \in [0^\circ, 360^\circ]$ a rotation angle. The angles $\alpha$ for the hands of the clockwise rotated system are then given by

$$\mod \left( \begin{bmatrix} 6^\circ & 0 & 0 \\ 0.1^\circ & 6^\circ & 0 \\ 0 & 0.5^\circ & 30^\circ \end{bmatrix} \begin{bmatrix} n_{sec} \\ n_{min} \\ n_{hour} \end{bmatrix} + \omega, 360^\circ \right) = \begin{bmatrix} \alpha_{sec} \\ \alpha_{min} \\ \alpha_{hour} \end{bmatrix}.$$

4 Setup and results

We create four different test scenarios from the CloCk test set to investigate the effect of transformed inputs on the model performance (cf. Table 1). For each image and scenario the rotation angle $\omega$ is uniformly drawn from the respective range. The ranges are discretized in steps of $1^\circ$ to decrease the difficulty for the models. The task is to classify the displayed hours, minutes, and seconds.

We compare two similar convolutional and capsule architectures with around 3.5 million parameters each. The capsule architecture uses Dynamic Routing [1] as routing process. There are no restrictions on the weights of the models that would accommodate for the rotated inputs. Therefore, the networks have to learn the effect of the transformation from data. For each architecture we train three different models to study the need for transformed data in the training set. CONV and CAPS are trained on unrotated images. Additional "12" and "360" indicate rotation ranges $[0^\circ, 360^\circ]$ of the training data.

Table 1: Accuracy of the convolutional and capsule models on the CloCk test set for four different rotation ranges.

| Model   | $\omega = 0^\circ$ | $\omega \in [-12^\circ, 12^\circ]$ | $\omega \in [-30^\circ, 30^\circ]$ | $\omega \in [0^\circ, 360^\circ]$ |
|---------|-------------------|-----------------|-----------------|-----------------|
| CONV    | 0.92 0.84 1.00    | 0.56 0.20 0.29  | 0.37 0.09 0.14  | 0.08 0.02 0.02  |
| CONV12  | 0.92 0.74 0.98    | 0.91 0.79 0.98  | 0.48 0.37 0.48  | 0.10 0.06 0.09  |
| CONV360 | 0.09 0.00 0.02    | 0.09 0.00 0.02  | 0.08 0.00 0.02  | 0.08 0.00 0.02  |
| CAPS    | 0.95 0.99 1.00    | 0.48 0.24 0.22  | 0.32 0.10 0.10  | 0.10 0.02 0.03  |
| CAPS12  | 0.95 0.83 1.00    | 0.94 0.84 1.00  | 0.56 0.39 0.51  | 0.11 0.08 0.09  |
| CAPS360 | 0.82 0.50 0.96    | 0.84 0.52 0.96  | \textbf{0.82 0.51 0.96} | \textbf{0.83 0.51 0.96} |

The results in Table 1 indicate, that both architectures benefit from examples of transformed data in the training set. They are unable to generalize the concept of rotations from smaller angle ranges in the training set to full rotations in the test case. This underlines the importance of equivariance restrictions if the transformation class is known (cf. Section 2).

The capsule architecture outperforms the CNN in all test scenarios. The model CAPS360 is able to learn the effect of the rotations and shows a constant performance in all test cases.

5 Conclusion

We compared the classification performance of two similar convolutional and capsule architectures on rotated images of a clock. Both architectures required examples of the transformation in the training set to work best in the respective setting. The capsule models outperformed the CNNs in all test cases. This indicates an improved ability to adapt to transformations. Together with the concept of equivariance, capsule networks are a promising model class to face transformations that commonly arise in image processing tasks.

Acknowledgements  Maximilian Schmidt acknowledges the support by the Deutsche Forschungsgemeinschaft (DFG) within the framework of GRK 2224/1 “$\pi^3$: Parameter Identification – Analysis, Algorithms, Applications”.

Open access funding enabled and organized by Projekt DEAL.

References

[1] S. Sabour, N. Frosst, and G. E. Hinton, Dynamic routing between capsules, Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, California, USA, (Curran Associates Inc., Red Hook, NY, USA, 2017), pp. 3859–3869.
[2] T. S. Cohen, and M. Welling, Steerable CNNs, 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings, (OpenReview.net, 2017)
[3] M. Schmidt, Classify on the clock (CloCk) - an entry level image data set for neural networks, Version 1.0.0, Data set, Zenodo, http://doi.org/10.5281/zenodo.3939033