Capsule networks for learning structure in language and music

Andrei Furdui, Alexander Lell, Vasileios Charatsidis

University of Amsterdam

Introduction

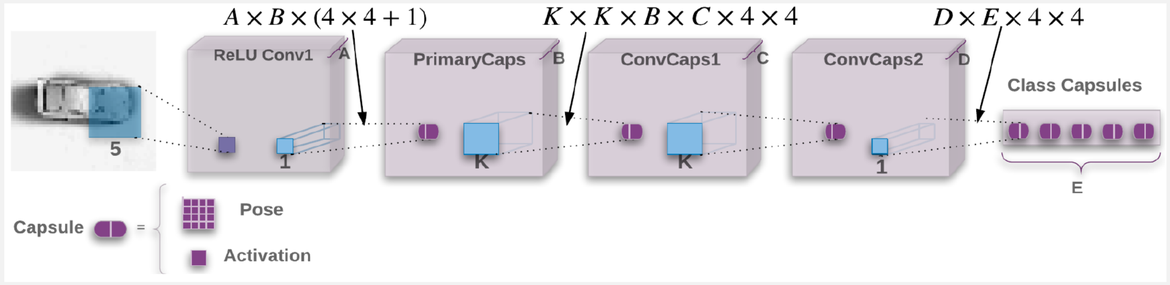
Capsule networks [1] [2] are a neural network architecture potentially superior to CNNs. Developed by Geoffrey Hinton, they are insensitive to rotations and other changes in the image, and excel at segmentation tasks while requiring much less data to train. Capsules represent their state not just by a scalar value, but an activation probability and a pose matrix. The activation probability represents the likelihood of the capsule to convey its information to the next layer. The pose matrix represent the relation between the 'view' of the capsule (its input) and the pattern that the capsule itself represents.

Objectives

The aim of the project is to investigate the behaviour and performance of the **Capsule Network** architecture and apply it to tasks it was not originally designed for: classification of natural language and music. For this purpose, we considered two subtasks:

- **Sentence classification**: We compared two very different architectures in the task of sentence classification in the TREC question dataset consisting of 5400 train data and 500 test data with 6 different classes.
- **Music Mode Classification**: We experiment with mode classification of Gregorian Chants, using the Cantus dataset [3], a massive collection of historic liturgical songs from which we extracted 57000 chants, consisting of their melodic transcripts and their associated mode.

Capsule Network Architecture



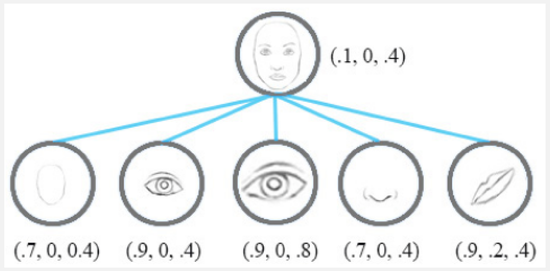


Figure: How a Capsule Network detects features

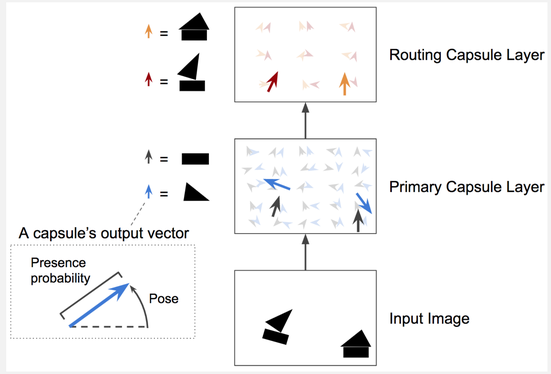
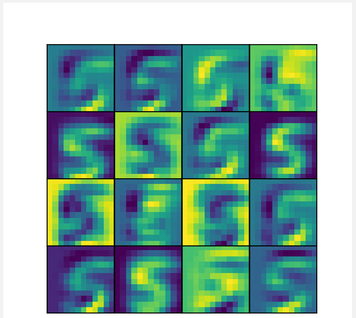


Figure: Visualization of the Routing Mechanism

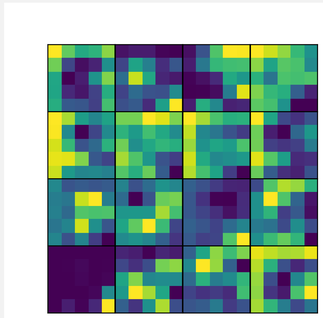
Methods

Visualization In order to visualize capsule networks, we extracted the capsule activations for untrained and trained capsule networks and visualized them with heatmaps.

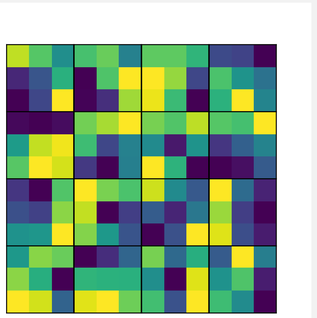
Audio Mode Classification We applied the capsule layer network to the classification of modes, a type of musical scale coupled with a set of characteristic melodic behaviors. The capsule network was adjusted to use 1-dimensional convolutions instead of 2-dimensional convolutions.



(a) first layer activations



(b) second layer activations



(c) third layer activations

Gregorian Chant Mode Classification

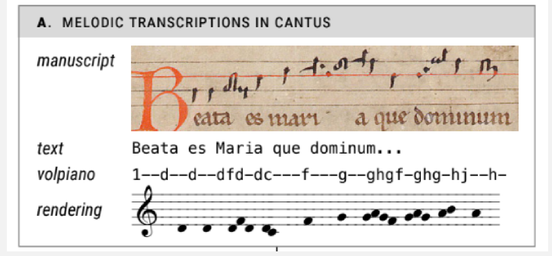


Figure: Gregorian Chant Representations

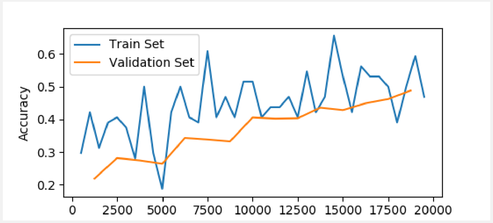


Figure: Training Results

Sentence Classification

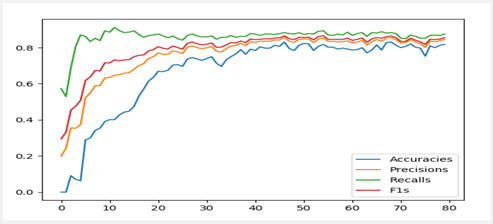


Figure: TREC Question Sentence Classification training results

References

[1] Sara Sabour, Nicholas Frosst, and Geoffrey Hinton. Dynamic routing between capsules. 2017.

[2] Geoffrey Hinton, Sara Sabour, and Nicholas Frosst. Matrix capsules with em routing. 2018.

[3] Cantus: A database for latin ecclesiastical chant – inventories of chant sources. available from <http://cantus.uwaterloo.ca/>.