Quaternion Capsule Networks

Barış Özcan¹
Furkan Kinli²
Furkan Kirac³

Department of Computer Science, Özyeğin University, İstanbul, Türkiye
¹baris.ozcan.10097@ozu.edu.tr, {2furkan.kinli, ³furkan.kirac}@ozyegin.edu.tr

Abstract
Capsules are grouping of neurons that allow to represent sophisticated information of a visual entity such as pose and features. In the view of this property, Capsule Networks outperform CNNs in challenging tasks like object recognition in unseen viewpoints. In this paper, we present Quaternion Capsules (QCN) where pose information of capsules and their transformations are represented by quaternions. Quaternions are immune to the gimbal lock, have straightforward regularization of the rotation representation for capsules, and require a smaller number of parameters than matrices. The experimental results show that QCNs generalize better to novel viewpoints with fewer parameters, and also achieve on-par or better performances with the state-of-the-art Capsule architectures on well-known benchmarking datasets.

Quaternion Capsules

- \( \hat{v}_{ij} = f_j(u_i) = w_{ij} \ast u_i + w^*_{ij} \)
- \( w_{ij} = \begin{bmatrix} \cos \theta_j, & \sin \theta_j \end{bmatrix} w_i \)
- \( w_i \) is the child capsule pose and \( w_{ij} \) is the learned transformation between \( i^{th} \) and \( j^{th} \) capsule in layers \( k \) and \( k + 1 \) respectively.
- \( w^*_{ij} \) stands for the conjugate and \( \ast \) stands for quaternion product.
- \( \theta \) is the rotation axis and \( \hat{\theta} \) is the rotation angle (Figure 1).
- EM Routing is used without any modification for fair comparison.

Our Architecture

- Experiments on smallNORB with branched and non-branched versions of QCN and Matrix Capsules.
- Branched version of Matrix Capsules is consist of the same capsule extractor layers as in QCN, while non-branched versions have two consecutive residual blocks with 64 and 96 channels.
- The results demonstrate that using separate branches for pose and activation increases the overall performance of QCN, though the number of parameters is reduced.

Effect of Branching

- Table 1. The effect of branching on the error rates and parameters for both Matrix and Quaternion Capsules.

Results on Common Datasets

- Table 2. Results on Common datasets where capsule networks are tested.

Generalization to Novel Viewpoints

- Training and test samples are from different viewpoints of the classes in smallNORB dataset.
- Azimuth experiment: training and test set are divided w.r.t. azimuth angles.
- Elevation experiment: training and test set are divided w.r.t. elevations.
- Training is stopped where training performances are similar for each model to test generalization fairly.

Table 3. Generalization to novel viewpoint results. Methods are trained on Familiar viewpoints and tested on novel viewpoints.

References