



DEPARTMENT OF INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Robotics, Cognition, Intelligence

# **Evaluation and Generalization of Capsule Networks in Neurorobotics**

Jean A. Elsner







DEPARTMENT OF INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Robotics, Cognition, Intelligence

# **Evaluation and Generalization of Capsule Networks in Neurorobotics**

## **Evaluation und Generalisierung von Kapsel Netzwerken in der Neurorobotik**

Author:	Jean A. Elsner
Supervisor:	Prof. Dr. Alois C. Knoll
Advisor:	Alexander Kuhn
Submission Date:	TBD



I confirm that this master's thesis in robotics, cognition, intelligence is my own work and I have documented all sources and material used.

Munich, TBD

Jean A. Elsner

## Acknowledgments



# Abstract





# Contents

<b>Acknowledgments</b>	<b>iii</b>
<b>Abstract</b>	<b>v</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 State of the Art</b>	<b>3</b>
2.1 Deep Learning for Object Recognition . . . . .	3
2.1.1 Convolutional Neural Networks . . . . .	3
2.1.2 Spiking Neural Networks . . . . .	5
2.2 Limits of Deep Learning Approaches . . . . .	5
<b>3 Capsule Network Architectures</b>	<b>7</b>
<b>4 Experimental Setup</b>	<b>9</b>
<b>5 Results</b>	<b>11</b>
<b>6 Discussion</b>	<b>13</b>
<b>7 Conclusion</b>	<b>15</b>
<b>List of Figures</b>	<b>17</b>
<b>List of Tables</b>	<b>19</b>
<b>Bibliography</b>	<b>21</b>



# 1 Introduction



## 2 State of the Art

This chapter presents an overview of state of the art approaches to object recognition, while focusing on two families of architectures, which are motivated quite differently. Object recognition techniques based on convolutional neural networks (CNNs) currently dominate the field, achieving state of the art performance on many datasets [1, 3]. CNNs however, are only loosely based on biological neurons. Spiking neural networks (SNNs) on the other hand, try to mimic the physical properties of neurons more closely and therefore constitute biologically more plausible models [4]. Generally speaking, CNNs may be regarded as a more engineering-based approach (or top-down), while SNNs are motivated by results from neuroscience and biology (bottom-up approach).

### 2.1 Deep Learning for Object Recognition

Recent years have seen a surge of interest in deep learning methods, especially in the field of computer vision. While the theory behind many deep learning methods has been around for many years, their recent success is mainly due to the availability of large labelled data sets and highly parallel computing powered by GPUs. One of the specific tasks, deep learning based methods excel at, is object recognition: the identification of objects in images or videos (cf. figure 2.1). The significantly better performance of deep neural networks over traditional machine learning methods can be explained by: (i) their hierarchical topology of parameterized non-linear processing units is a fundamentally better probabilistic model and prior for real world data leading to better generalization and (ii) they automatically find good features to extract based on the training data. The potential applications for a robust image classification system are myriad and range from automated driving and image-based diagnosis to robot vision and many more. As deep learning is currently the best candidate for such a system, it is well worth exploring.

#### 2.1.1 Convolutional Neural Networks

CNN architectures are generally distinguished by their use of specific types of neuron-layers, namely, convolutional, pooling and fully connected layers. While wildly different network topologies may be found in literature, characterized by their use of skip connections, number of layers, number of paths etc., CNNs can always be reduced to these three basic layer types.

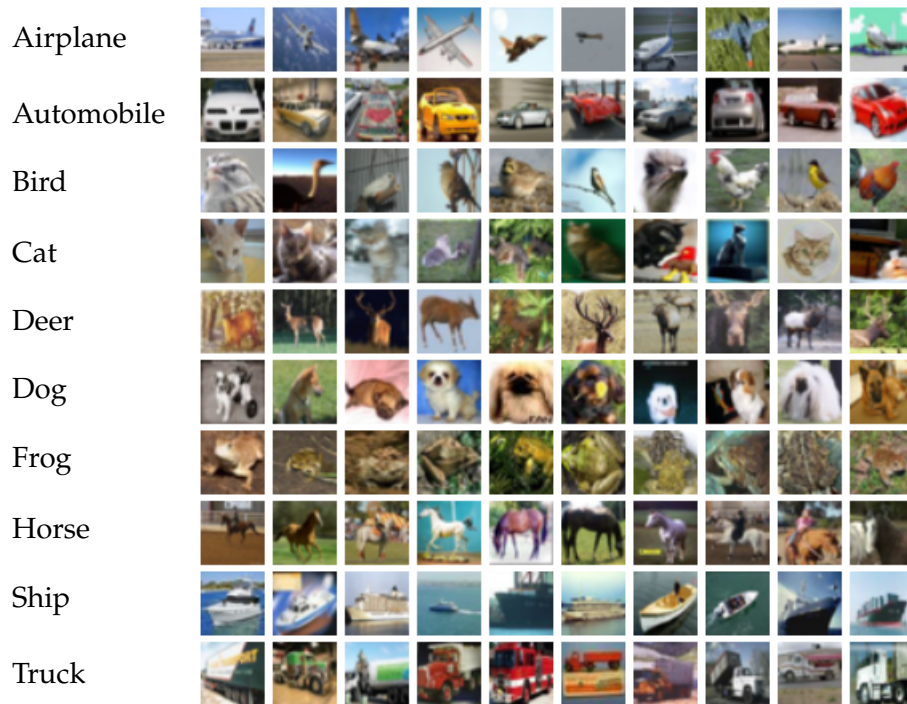


Figure 2.1: Sample images from the CIFAR-10 [2] dataset and their corresponding classes. CIFAR-10 consists of 6000 images at 32 by 32 pixels for each of the 10 classes. Datasets such as this are often used as a benchmark to evaluate the performance of novel deep learning architectures for image recognition.

### Fully Connected Layer

Each neuron in a fully connected layer is connected to all the activations in the previous layer. The activation of a single neuron as the basic non-linear computational unit is calculated by applying a nonlinearity to the weighted sum of its inputs plus a bias.

$$h = g\left(\sum_i w_i x_i + b\right) \quad (2.1)$$

With the nonlinear function  $g$ , the learnable weights  $w_i$ , the input activations  $x_i$  and the learnable bias  $b$ . In the case of a fully connected layer, the activation can be computed using matrix multiplication. In tensor notation this may be written as:

$$\mathbf{h}_l = g_l(\mathbf{W}_l^T \mathbf{h}_{l-1} + \mathbf{b}_l). \quad (2.2)$$

With  $N_l$  denoting the number of neurons in layer  $l$ ,  $\mathbf{W}_l$  is an  $N_{l-1} \times N_l$  dimensional weight matrix,  $\mathbf{b}_l$  an  $N_l$  dimensional vector and  $g_l$  the  $N_l$  dimensional vectorized activation function of layer  $l$

$$g_l(\mathbf{x}) = (g_l(x_1), \dots, g_l(x_{N_l}))^T \quad (2.3)$$

### Convolutional Layer

#### Pooling Layer

#### 2.1.2 Spiking Neural Networks

## 2.2 Limits of Deep Learning Approaches





## 3 Capsule Network Architectures



## 4 Experimental Setup



## 5 Results



## 6 Discussion





## 7 Conclusion



# List of Figures

2.1 CIFAR-10 classes and sample images . . . . . 4



# List of Tables



# Bibliography

- [1] A. Diba, V. Sharma, A. M. Pazandeh, H. Pirsiavash, and L. V. Gool. “Weakly Supervised Cascaded Convolutional Networks.” In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), pp. 5131–5139.
- [2] A. Krizhevsky. *Learning Multiple Layers of Features from Tiny Images*. Tech. rep. 2009.
- [3] W. Ouyang, X. Zeng, X. Wang, S. Qiu, P. Luo, Y. Tian, H. Li, S. Yang, Z. Wang, H. Li, K. Wang, J. Yan, C. C. Loy, and X. Tang. “DeepID-Net: Object Detection with Deformable Part Based Convolutional Neural Networks.” In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39.7 (July 2017), pp. 1320–1334. issn: 0162-8828. doi: 10.1109/TPAMI.2016.2587642.
- [4] A. J. Schofield, I. D. Gilchrist, M. Bloj, A. Leonardis, and N. Bellotto. “Understanding images in biological and computer vision.” In: *Interface Focus* 8.4 (2018). issn: 2042-8898. doi: 10.1098/rsfs.2018.0027. eprint: <http://rsfs.royalsocietypublishing.org/content/8/4/20180027.full.pdf>.