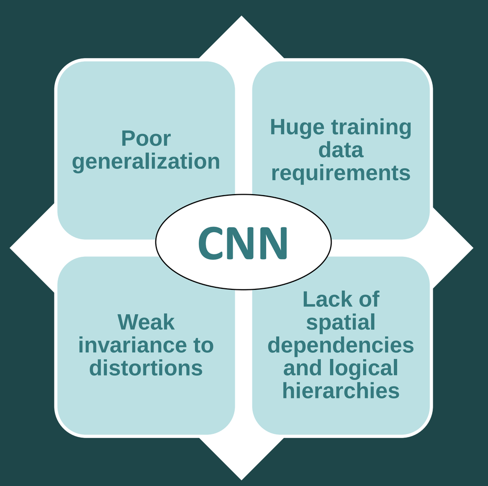
# Lung nodule classification by means of capsule neural networks

Kolezhuk L., Goatman K. A.

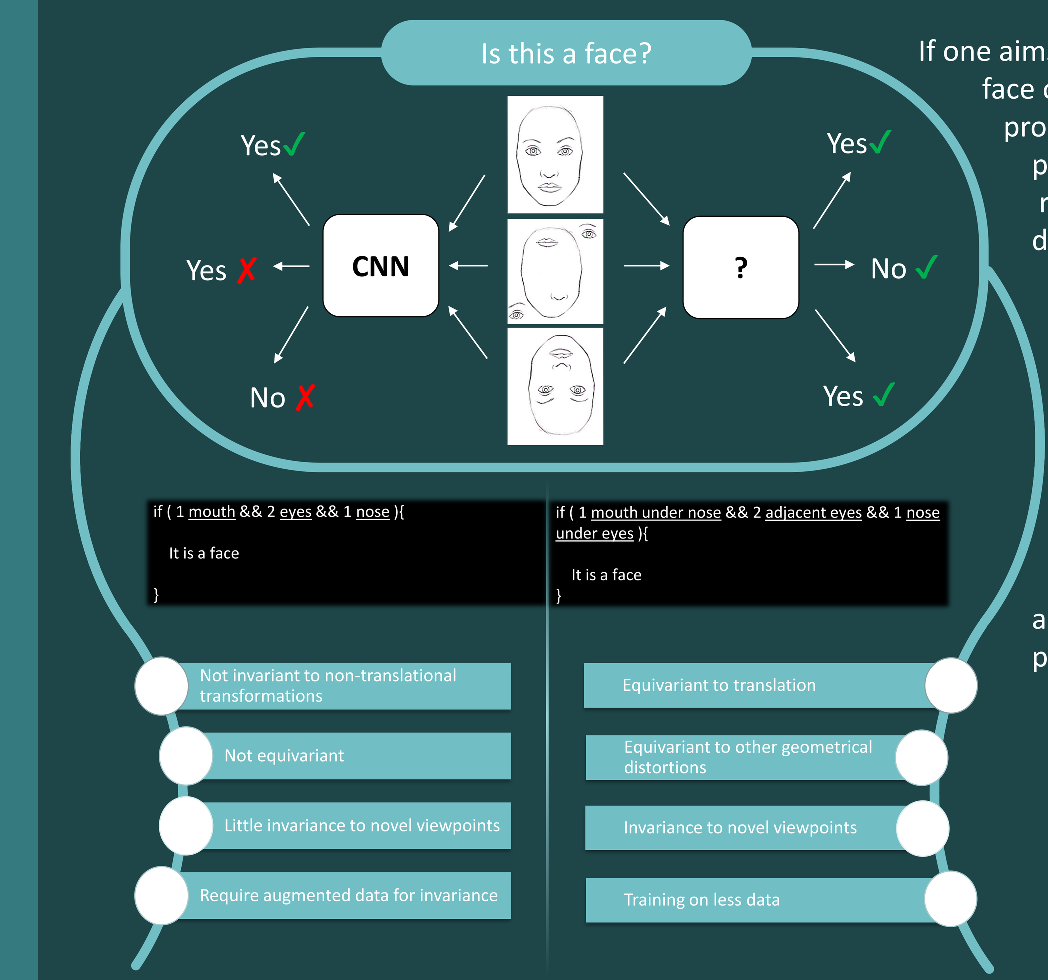
## Background




**CNN**s are the key to obtaining solutions with remarkable results to countless modern computer vision tasks. However, there are some limitations introduced by their concept: absence of invariance to geometrical distortions other than translation; poor generalization; do not consider spatial relationships between features on the image but only their presence; require huge training datasets; low stability to change of viewpoints; no logical structure and neuron hierarchies; generically large networks easily fit random labels to data. Some of this issues are associated with the type of data routing that is used to pass signals to deeper layers of the network. The most common routing algorithm is **max-pooling**.

*The pooling operation used in convolutional neural networks is a big mistake and the fact that it works so well is a disaster. If the pools do not overlap, pooling looses valuable information about where things are. We need this information to detect precise relation-ships between the parts of an object.*

*G. Hinton*




If one aims to determine if the object present on the image is a face or not, a typical CNN with max-pooling would quite probably misclassify completely distorted images as presented in the figure due to not considering the spatial relationships between features. Moreover, a rotated or distorted image would also generate irrelevant output unless the network has been trained on data that already contained such distortions as result of aug-mentation or extensive sets of training samples.

In **medical imaging**, a classification network that could base its predictions also on the spatial relationships between objects of the input would be very beneficial. Such network would allow to learn more complex class feature distribution patterns, and hence increase the accuracy of its predictions.

## Goal

In this work we investigate a recently proposed approach that claims to tackle these issues by introducing advanced grouping of neurons into capsules, that are designed to encode not only the presence of certain features, but also their instantiation parameters. The advantages and limitations of this architecture are studied. Our work expands the use of capsule networks to the task of false positive reduction for pulmonary nodule detection in lung CT scans while achieving improved classification accuracy when compared to a conventional CNN with max-pooling.

## Experiments and results

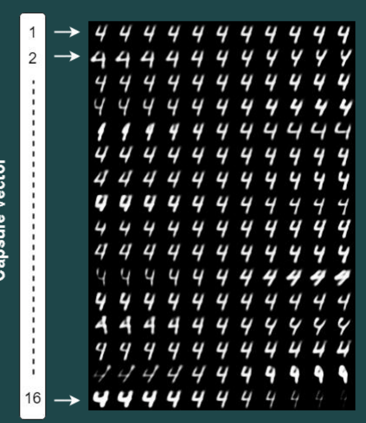
### Handwritten digit classification

In order to verify the correctness of the implementation we have tested CapsNet for the MNIST digit classification task. The following experiments were performed:

- MNIST classification performance
- Robustness to random affine transformations of the input
- Encoded representation quality investigation

| Dataset | Test error rate % | | |
|---|---|---|---|
| | CapsNet (Sabour et al. 2017) | CNN (state-of-the-art) | CapsNet (our implementation) |
| MNIST | 0.25 | 0.39 | 0.29 |
| affNIST | 21 | 34 | 25 |
| multiMNIST | 5.20(80% overlap) | 5.2(4% overlap) | --- |

Performance sufficiently similar to the one claimed by Sabour et al. (2017) was achieved by our implementation. Significant robustness to affine transformations of input data has also been verified indicating the conceptual and functional correctness of the implementation.

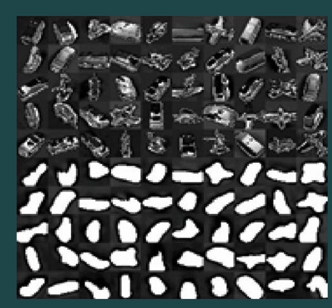



Taking into account that this data is rather simplistic, while having no background and small details in the object bodies, it is of great interest to study how the architecture is able to cope with more complex data.

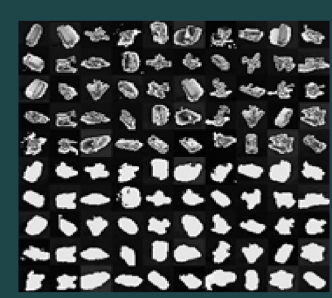
### Classification of images with 3D objects

| Dataset | Test error rate % | |
|---|---|---|
| | CapsNet | CNN |
| smallNORB | 2.7 – Sabour et al.(2017) | 2.0 |
| | 1.8 – EM Caps | |
| | 10.9 – our approach | |
| CIFAR10 | 10.6 – Sabour et al.(2017) | 4.5 |
| | 11.9 – EM Caps | |
| | 28.5 – Xi E. (2017) | |

Once data of greater complexity is processed it becomes troublesome for the network to properly extract and encode sufficient amount of important features from the inputs. The quality of the of this extraction may be judged by the detalization of the reconstructed images. Logically this issue may be tackled by increasing the size of the network by adding more layers, capsule types, extending the dimensions of capsule vectors etc. However, the experiments have proven that any further expansions after the presented ones do not cause any notable raises neither in the reconstruction quality nor the overall classification accuracy.

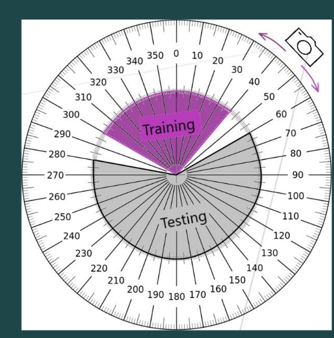


64 convolutional filters, 8 8D capsules in the PrimaryCaps layer



256 convolutional filters, 32 8D capsules in the PrimaryCaps layer

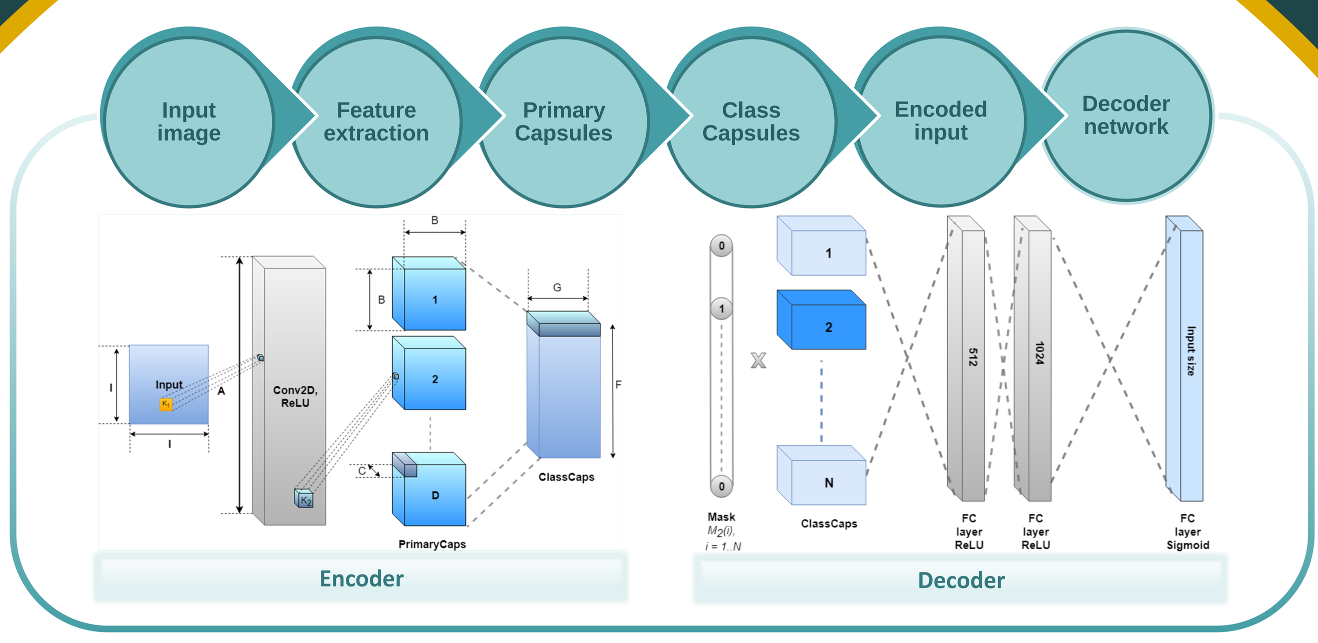
CapsNet is more robust to the change of viewpoints in the data, however it was proven that data augmentation is necessary in order to enforce correct estimation of the objects instantiation parameters. The last allows CapsNet to better deal with previously unseen data than a conventional CNN.



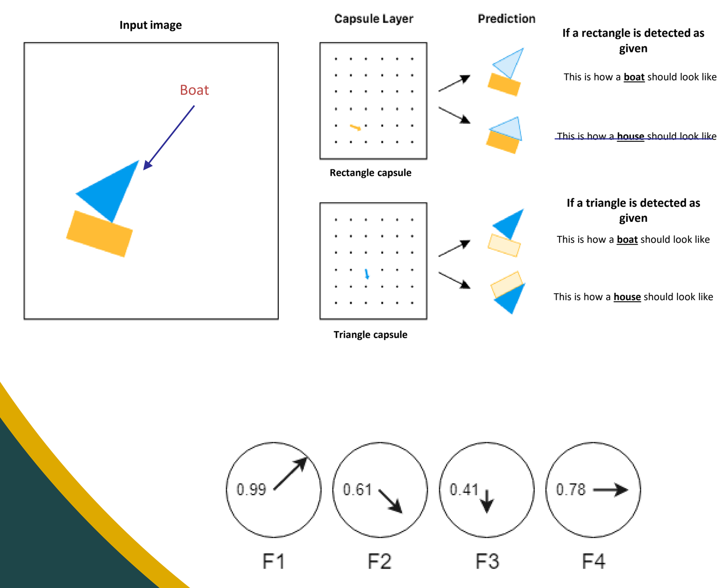

Split of training and testing data for the viewpoint robustness study

| Model | | Test accuracy | | Robustness coefficient |
|---|---|---|---|---|
| | | Familiar viewpoints | Novel viewpoints | |
| 64 conv. Filters, 8x8 P. Caps | CapsNet - 32x32 | 82.19 | 66.78 | 0.81 |
| | CapsNet - 32x32 - augmented | 83.04 | 77.23 | 0.93 |
| | CapsNet - 32x32 (random crops from 48x48) | 84.58 | 68.44 | 0.81 |
| | CapsNet - 32x32 (random crops from 48x48) - augmented | 83.72 | 79.91 | 0.95 |
| **256 conv. Filters, 32x8 P. Caps** | **CapsNet - 32x32 (random crops from 48x48) - augmented** | **81.73** | **80.11** | **0.98** |
| CNN 32x32 | | 80 | 68.97 | 0.86 |
| CNN - 32x32 - augmented | | 83.96 | 77.04 | 0.92 |

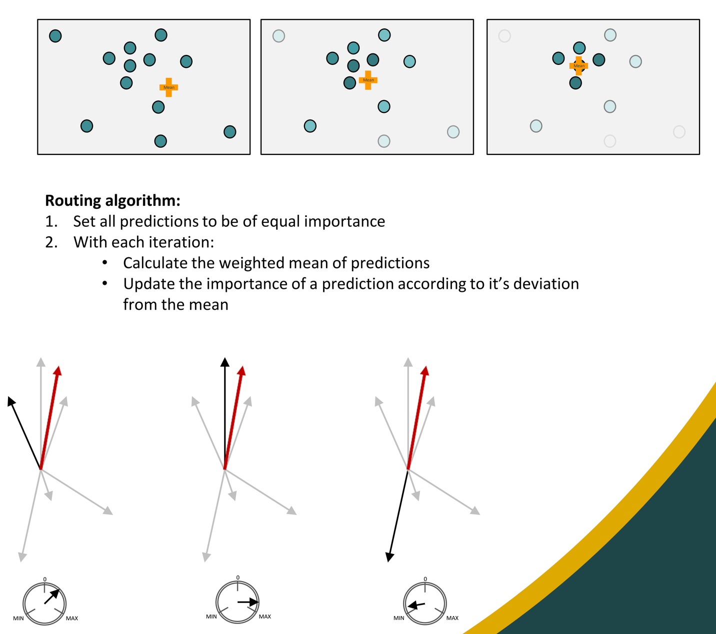
## Capsule network




### Capsule predictions



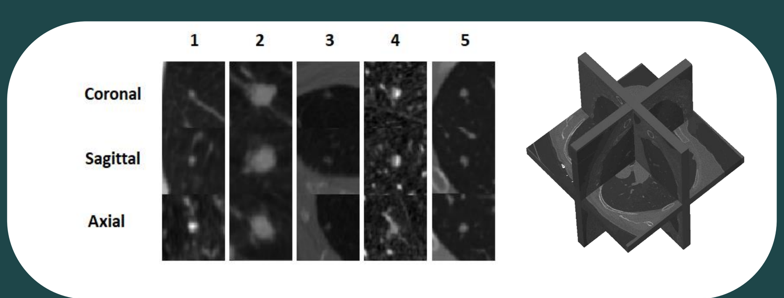

### Dynamic routing




**Routing algorithm:**

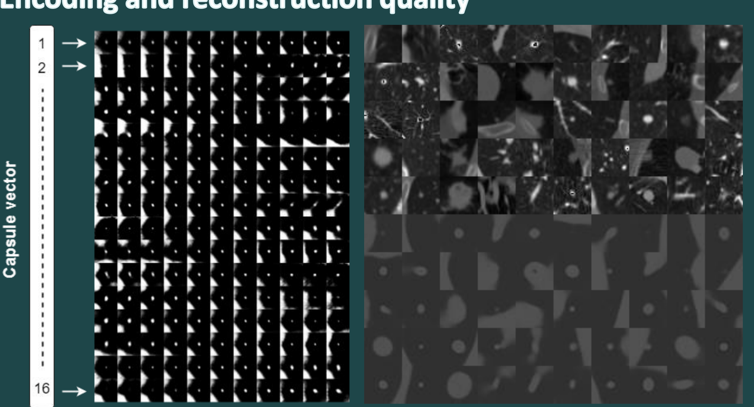
1. Set all predictions to be of equal importance
2. With each iteration:
   - Calculate the weighted mean of predictions
   - Update the importance of a prediction according to it's deviation from the mean

### False positive reduction for pulmonary nodule detection




The LUNA16 dataset is used to target the goal. We extract orthogonal patches for each nodule candidate in CT volumes. The classes are balanced by means of generating 125 samples per view for the positive class, and selective reduction of the number of samples in the negative class.

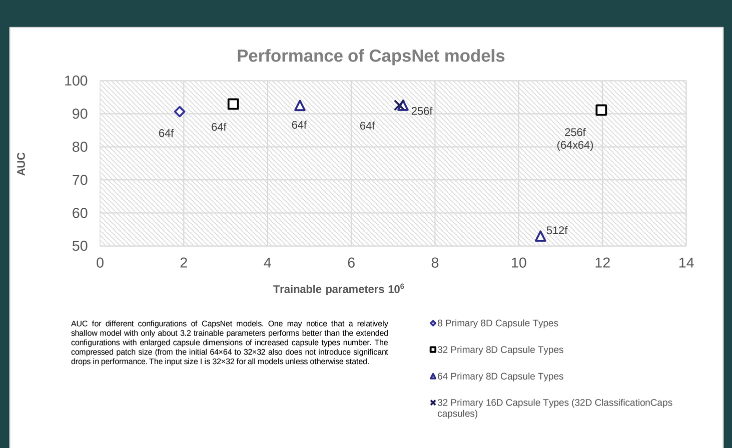
### Encoding and reconstruction quality

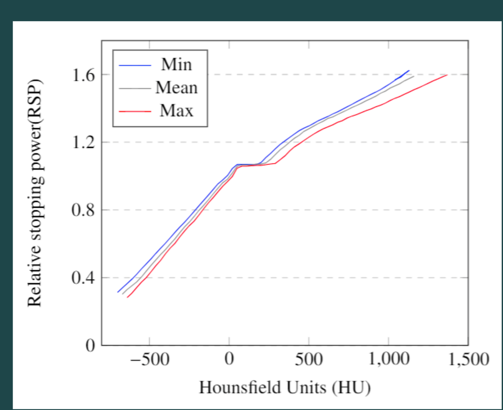



### Performance of CapsNet models

| Configuration | Test AUC | |
|---|---|---|
| | CNN baseline | CapsNet |
| Separate patches | 91.4 | 93.2 |
| Candidate (patch voting) | 93.8 | 96.0 |

Performance of different CapsNet models was investigated. It was determined that the deeper models do not allow to obtain neither better classification accuracy, nor improved feature encoding quality.





| Computation time per training step | |
|---|---|
| CapsNet | CNN |
| 158ms | 19ms |

## Discussion & Conclusion