EXPLAINABILITY OF CAPSULE NETWORKS AND THEIR APPLICATION FOR BRAIN TUMOR CLASSIFICATION

Atefeh Shahroudnejad, Parnian Afshar, and Arash Mohammadi
Concordia Institute for Information System Engineering, Montreal, QC, Canada

MOTIVATIONS/CONTRIBUTIONS

Motivation:

- When deep learning architectures are utilized for making critical decisions such as the ones that involve human lives (e.g., in medical applications), it is of paramount importance to understand, trust, and in one word “explain” the rational behind deep models decisions.
- Commonly used Convolutional Neural Networks (CNNs), typically, require large amount of training data and fail to handle some kinds of transformations.
- CapsNets are novel deep structures recently proposed as an alternative counterpart to the CNNs.
- CapsNets are robust to rotation and affine transformation, and require far less training data.

Contribution:

- We investigate and analyze structures and behavior of the CapsNets and illustrate potential explainability properties of such networks.
- We adopt and incorporate CapsNets for the problem of brain tumor classification.

CAPSULE NETWORKS

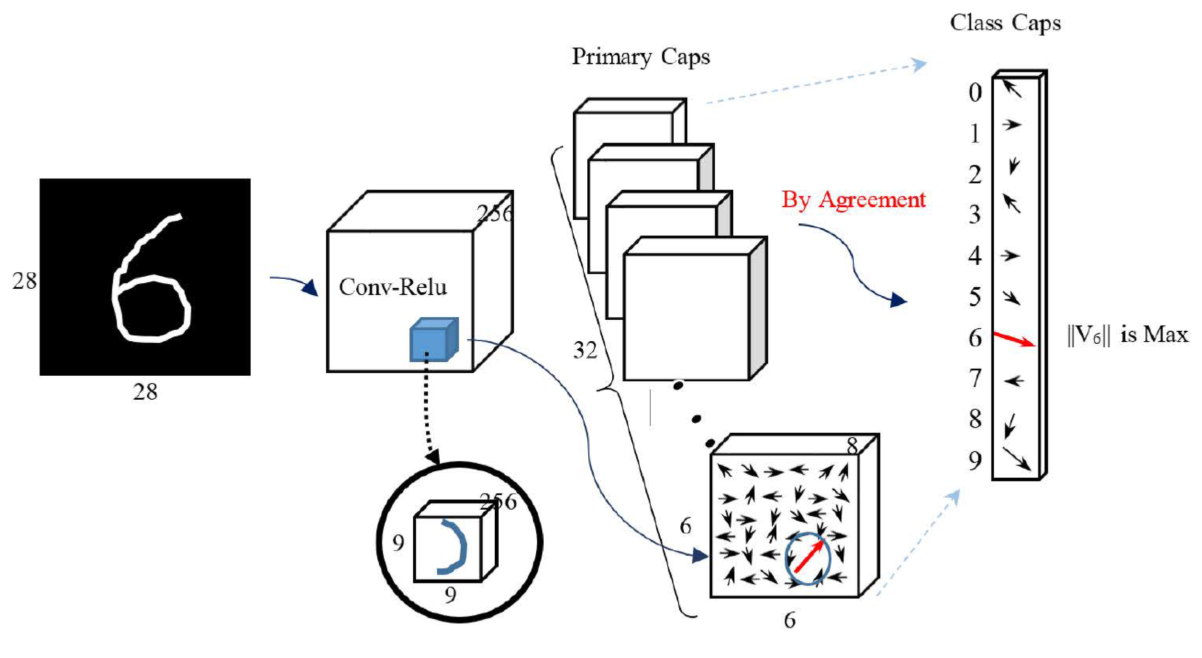

Figure 1: CapsNet architecture.

- Capsules are groups of neurons modeled with activity vectors representing various pose parameters. Length of an activity vector show the probability that a specific entity exists.
- Each primary capsule tries to predict the output of the parent capsules as follows

$$\hat{u}_{j|i} = W_{ij} u_i.$$

- Based on ***Routing by Agreement***, Capsules in the primary layer send their outputs to all parent Capsules, and the parent Capsule’s output is computed as

$$s_j = \sum_i c_{ij} \hat{u}_{j|i}.$$

- The c_{ij} is updated in the routing process based on the agreement between the squashed of s_j and $\hat{u}_{ij}$ using the fact that if the two vectors agree, they will have a large inner product.
- CapsNet also has three layers of fully connected neurons which try to reconstruct the input using the instantiation parameters from the Capsule associated with the true label.

EXPLAINABILITY OF CAPSULE NETWORKS

- The main goal of explainability is to find answers to questions such as: *What is happening inside a neural network? What does each layer of a deep architecture do? What features a deep network is looking for?*
- CapsNets automatically form the verification framework and inherently create the relevance path which eliminates the need for a backward process to construct it, and as such improve trustworthy and explainability of deep networks.

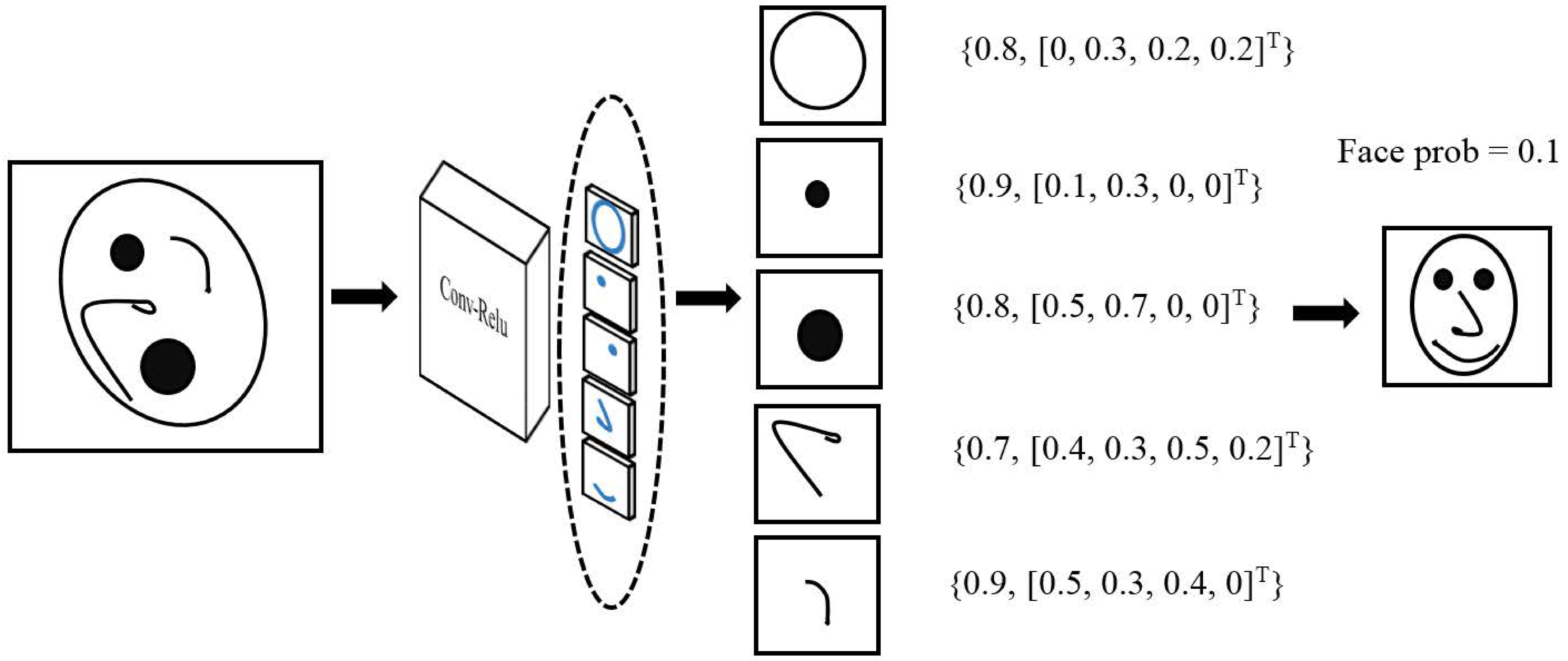

Figure 2: Mismatch among instantiation parameter vectors for the face capsule.

- The vector representation provided by CapsNets is highly informative and can model possible instantiation parameters for components or fragments of an object.
- We can assign to each capsule a set consisting of two segments for explanation purposes: (i) ***Likelihood Values***, which can be used to explain existence probability of the feature that a Capsule detects, and; (ii) ***Instantiation Parameter Vector Values***, which can be used to explain consistency among the layers.
- CapsNet applies non-linear squashing function on output vectors resulting in the unrelated Capsules to become smaller.

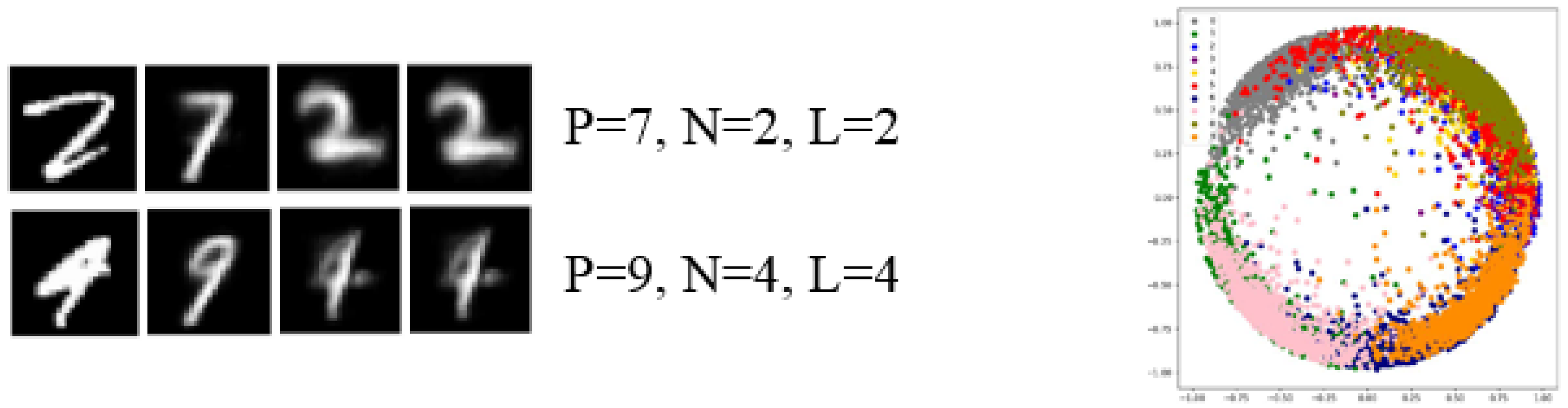

Figure 3: Left: Misclassified samples and CapsNets’ reconstructions based on highest likelihoods. Right: MNIST offline map.

CONCLUSION

- We illustrated potential intrinsic explainability properties of CapsNets.
- W investigated the use of CapsNets for the problem of brain tumor classification, and based on our experiments, these networks can perform better for the segmented tumors than for the whole brain images.

REFERENCES

[1] P. Afshar, A. Mohammadi, and K.N. Plataniotis, “Brain Tumor Type Classification via Capsule Networks,” Submitted to *IEEE International Conference on Image Processing (ICIP)*, available at **arXiv:1802.10200**, 2018.

[2] A. Shahroudnejad, A. Mohammadi, and K.N. Plataniotis, “Improved Explainability of Capsule Networks: Relevance Path by Agreement,” Submitted to *IEEE International Conference on Image Processing (ICIP)*, available at **arXiv:1802.10204**, 2018.

BRAIN TUMOR CLASSIFICATION

- We used a Data-set containing 3064 MRI images of 233 patients diagnosed with different brain tumor types.
- First, different kinds of CapsNet architectures are tested. According to our results, reducing the number of feature maps from 256 to 64 leads to the highest accuracy.

Capsule Network Architecture	Prediction Accuracy
Original architecture	82.30%
Two convolutional layers with 64 feature maps each	81.97%
One convolutional layer with 64 feature maps	**86.56%**
One convolutional layer with 64 feature maps and 16 primary capsules	83.61%
Three fully connected layers with 1024, 2048 and 4096 neurons	83.93%

- Once the best architecture of the CapsNet is selected, we have compared its classification accuracy with a CNN for both brain images and segmented tumors.
- CapsNet outperforms CNN for both types of inputs.

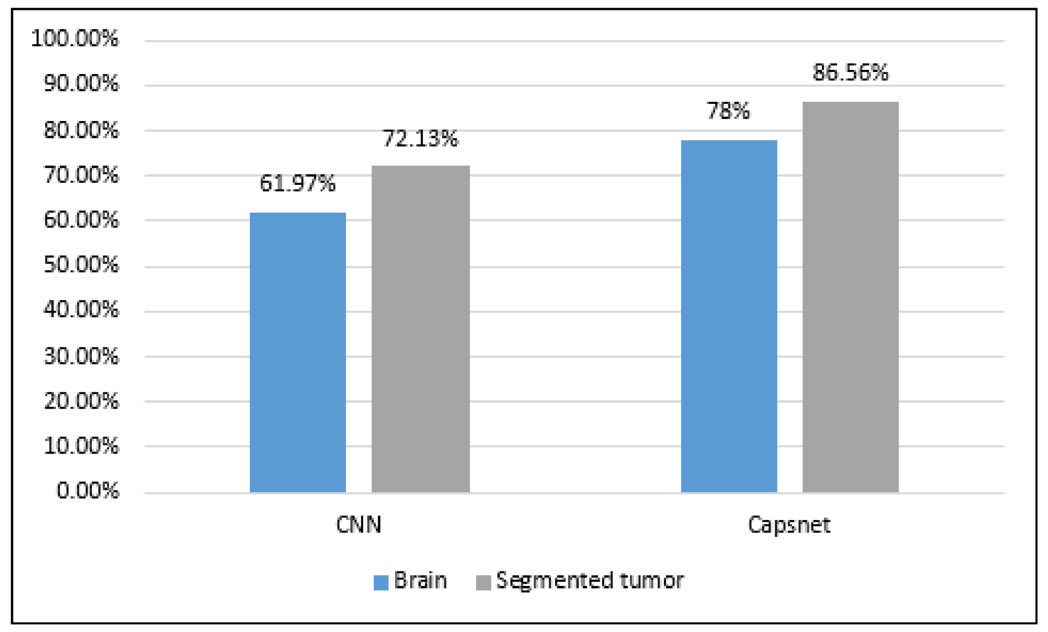

Figure 4: CapsNet vs. CNN.

- We have tweaked the output of CapsNet to visualize what features it has captured.

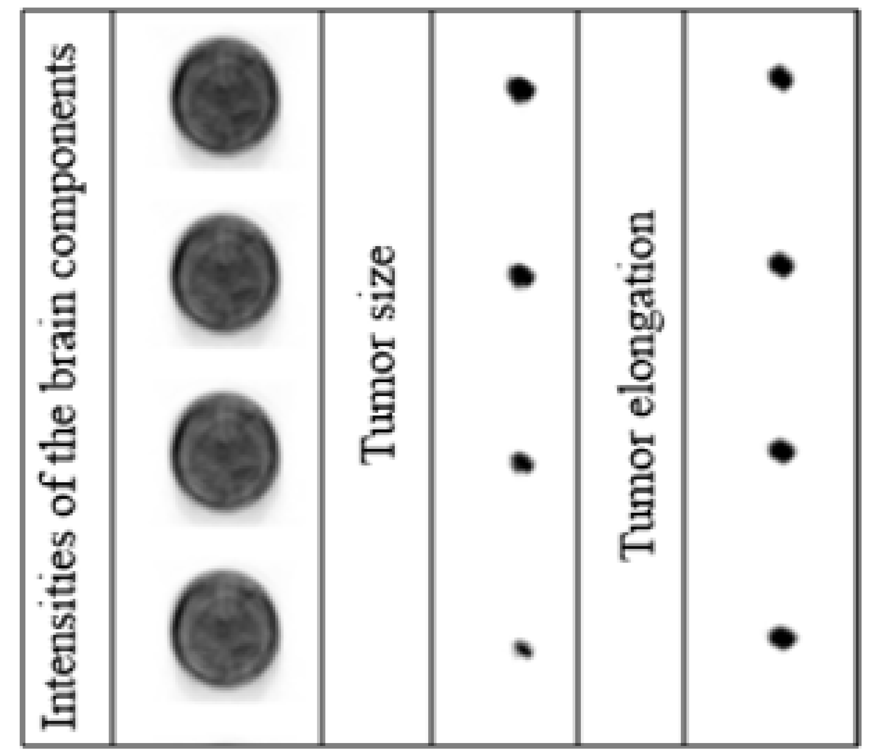

Figure 5: Tweaking the CapsNet output.