Medical Imaging Using Capsule Network

Khatri Bimal Mukesh Kumar
bimalgangawal@gmail.com
SIT TUMKUR

Madhupriya K Suttakoti
madhupriyaks05@gmail.com
SIT TUMKUR

ABSTRACT

Health care is the high priority sector where individuals expect high level of services and care regardless of cost. In the beginning, interpretation of medical data was being done by medical experts. After the success of deep learning for real world application, it has started providing exciting results with good accuracy for medical imaging and it will be seen as a primary methodology for future applications in health sector. However, Convolution Neural Networks(CNN) cannot handle input transformation and it requires large amount of training data and good coaching knowledge. Capsule Network is a recently announced machine learning architecture to overcome the shortcomings of CNN. A Capsule is a nested set of neural layers where an activity vector represents the instantiation parameter of specific kind of an entity such as an object or a part of it. Capsule networks are unit that are different from the neural networks in the way that network vertices are capsules instead of neurons, which are of high dimensional and the output of capsule is a vector which represents a group of parameters related to the input. Length of the activity vector is used to represent the probability that an entity exists and its orientation is used to represent the instantiation parameter. Each capsule is trained to trace a specific feature of an image. Combining them allows an Artificial Intelligence system to identify and spot different views of the same image. This technique is much better than the existing system in a challenge to recognise objects from different angles.

***Index Terms*— Capsule networks, Medical Imaging, Tumor Detection.**

1.INTRODUCTION

Cancer is the leading cause of morbidity, mortality around the world and this thought because of the common kind of illness found in youngsters and in adults. Due to the terrific advancement in the image acquisition devices, the data available from different health sector is giant which is making image analysis a challenging, difficult and fascinating one. Tumor type classification by individuals is extremely time consuming and error prone task. The rapid growth in medical images and intervention requires comprehensive tedious and verbose efforts by medical experts which is subjective, prone to error and they may have large variations from different expert. One solution was to use machine leaning techniques to automate diagnosis process however, but the traditional machine meaning techniques were not sufficient to deal with complex learning and capacity to subsume with large medical image for accurate and efficient diagnosis. Deep leaning had helped to pick out and extract features and also to construct new one from it. It additionally measures predictive target and provides affordable prediction models to assist physicians efficiently.

Machine learning(ML) and Artificial Intelligence(AI) have progressed quickly in recent years. Techniques of ML and AI have played a key role in medical field like image interpretation, image registration, medical image processing, computer- aided diagnosis, image fusion and image segmentation, image guided therapy, image retrieval and analysis. Techniques are used to extract information from the images and to effectively and efficiently represent the information. These techniques facilitate and aid doctors to diagnose and predict veracious and faster the risk of diseases and prevent them in time. They also enhance the ability of doctor and researchers to understand that how to analyse the generic variations which leads to disease. Techniques includes Conventional algorithms without learning such as Support Vector Machine(SVM), Neural Network(NN), K-Nearest Neighbour(KNN) etc. and Deep learning algorithms such as, Recurrent Neural Network(RNN), Long Short Term Memory(LSTM), Convolutional Neural Networks(CNN), Generative adversarial Networks(GANs), Extreme Learning Model(ELM), etc.

Convolutional Neural networks performs exceptionally great while classifying images which are very similar to data set and they have extensive and depth learning capacity. They can infer the nature of an associated input image without any prior knowledge. But, they are not robust to affine transformation and if the images have rotation, tilt or any other different orientation then CNNs have poor performance. They do not take the spatial relationships within the image into considerations. To improve their generalisation, CNNs need to be provided with training data consisting of all kinds of rotation and transformation. CNNs performs poor confronting small data sets, which is the case for most of the medical image data sets.

Drawbacks of CNNs was resolved by adding different variations of the same image during training. In CNN, each layer understands an image at a more granular level. Invariance or changelessness makes CNN tolerant to tiny changes in the view point. Equivariance makes a CNN understand the proportional and rotational change and adapt itself accordingly such that the spatial positioning inside an image is not lost. Capsules compares to the current state-of-the-art CNNs. Human brain have modules known as "Capsules". These Capsules units are smart at handling different types of visual stimulus and encoding things like reflective power, position, size, hue, orientation, deformation, velocity, texture etc. The brain should have a mechanism for routing low level information to what it believes is the best capsule for handling. Activity vector of each capsule is composed of several cause parameters such as orientation, scaling, and skewness. The length of each activity vector provides the existence probability of the particular object depicted by that capsule. Most key property of CapsNets is called "Routing by Agreement", which means Capsules in lower level predict the outcome of Capsule in higher levels, and also the higher level capsules get activated only if these predictions agree.

The rest of the paper is organised as follows: Section 2 explains needed mathematical background for CNNs. Section 3, considers CapsNets and presents the projected approach followed by the experimental details in Section 4. Finally, Section 5 concludes the paper.

2. PROBLEM FORMULATION

Given a set of N_{train} training MRI(Magnetic Resonance Imaging) images of various tumour organs, the goal is to design a deep learning architecture to identify and classify the type of cancer available in the given test MRI images into completely different categories, i.e., Acute Lymphoblastic Leukemia, Brain Tumor, Esophageal cancer, Bile duct cancer, Bladder Cancer, Eye cancer, Carcinoid Cancer, Adrenocortical Carcinoman, Liver cancer, etc,. In this paper Capsules has been used instead of neurons to build the architecture of deep learning. Before introducing the underlying CapsNets structure and the designed architecture, we briefly analyse basics of CNNs, which are the common form of deep networks for such classification tasks, and highlight their potential drawbacks which have led to the introduction of Capsule networks.

2.1 Convolutional Neural Networks

In Machine learning, a Convent is a feed forward neural network. It makes use of the subsequent properties: (i) Neuron Units in each layer receive inputs from the previous units which are located in a small neighbourhood and performs dot product. This way, elementary features such as edges and corners can be extracted. (ii) These features will be combined in next layers to detect higher order features. Next prominent property is the concept of shared weights, which means similar feature detectors are used for the entire image. (iii) At the end, Convolution Neural Networks usually have several sub-sampling layers. These layers are based on the fact that the precise location of the features can not only enhance, but additionally harmful, because information tends to vary for different instances based on the data. Although CNNs have been proved to be beneficial in many region, they have several drawbacks specially associated with the sub-sampling layers, because these layers provides a small amount of translational unvariedness and they loose the exact location of the foremost active feature detectors. Due to the above mentioned reasons, a new recent architecture called Capsule networks is introduced, which is more robust to translation, orientation and rotation. CapsNets are described in the next section.

3. CapsNet FOR TUMOR CLASSIFICATION

As stated above, the goal is to investigate and design a CapsNet architecture capable of classifying different types of tumour as accurately as possible. In the beginning we present the CapsNet properties.

3.1 Capsule Networks

Capsule Network are groups of neurons with each Capsule within the network consisting of several neurons. The activity vectors of these neurons is composed of various pose parameters(e.g., position, orientation, scaling, and skewness) and the length of these vectors provides the probability that a specific entity exists represented by that capsule. We capitalise on achievability of those benefits and adopt the CapsNet architecture for various tumour type classification problem. Shortcomings and drawbacks of CNNs are mostly related to the pooling layers. As a result, to overcome those in Capsule networks, these layers are replaced with a more appropriate metric known as "Routing By Agreement." Based on this measure, outputs are sent to all parent capsules in the next layer, however, their coupling coefficients are not the same. Each Capsule will try to predict the output of the parent Capsules, and if this prediction conforms to the actual output of the parent Capsule, then the coupling coefficient between these two capsules increases. Considering $\boldsymbol{u_i}$ as the output of capsule $\boldsymbol{i}$, its prediction for parent Capsule $\boldsymbol{j}$ is computed as

$$\hat{\boldsymbol{u}}_{j|i} = \boldsymbol{W}_{ij}\boldsymbol{u}_i, \tag{1}$$

where $\hat{\boldsymbol{u}}_{j|i}$ is the prediction vector for the output of the $\boldsymbol{j^{th}}$ Capsule which is in a exceedingly higher level computed by Capsule $\boldsymbol{i}$ in the below layer, and $\boldsymbol{W_{ij}}$ is the weight matrix that must be learned in the backward pass. Based on the degree of conformation between the capsules within the parent Capsule and also the below layer, coupling coefficients $\boldsymbol{c_{ij}}$ are calculated using the following *softmax* function as follows:

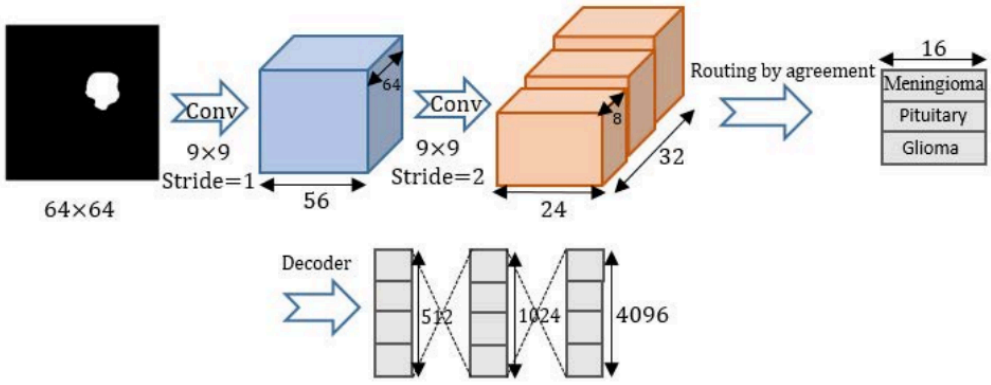

Fig. 1. Proposed model architecture for different tumor classification

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})}, \tag{2}$$

where b_{ij} is the log probability that whether Capsule i should be coupled with Capsule j and it is initially set to 0 at the beginning of the routing by agreement method. Therefore, the input vector to the parent Capsule j is calculated as follows

$$s_j = \sum_i c_{ij} \hat{u}_{j|i}. \tag{3}$$

Finally, the non-linear squashing function is employed to prevent the output vectors of Capsules from surpassing one and forming the final ultimate output of each Capsule based on its initial vector value as defined in Eq. (3)

$$v_j = \frac{\|s_j\|^2}{1+\|s_j\|^2} \frac{s_j}{\|s_j\|}, \tag{4}$$

where s_j is the input vector to Capsule j and v_j is the output. The log probabilities should be updated within the routing method based on the agreement between v_j and $\hat{u}_{j|i}$ using the fact that if the two vectors agree, they will have a large inner product. Therefore, agreement a_{ij} for updating log probabilities and coupling coefficients is calculated as follows

$$a_{ij} = v_j . \hat{u}_{j|i}. \tag{5}$$

Each Capsule k in the last layer is associated with a loss function l_k, which puts high loss value on capsules with long output instantiation parameters when the entity does not actually exists. The loss function l_k is computed as follows

$$l_k = T_k \max(0, m^+ - ||v_k||)^2 + \lambda(1-T_k)\max(0, ||v_k|| - m^-)^2, \tag{6}$$

where T_k is 1 whenever class k is actually present, and is 0 otherwise. Terms $m+$, $m-$, and λ are hyper parameters to be indicated before the learning process. The original Capsule network architecture presented consists of one layer of Convolutional filters and two layers of Capsules. It also has additional three layers of Fully Connected neurons which try to reconstruct the input using the instantiation parameters from the Capsule related with the true label.

3.2.Designing Capsule Network

After exploring several potential architectures of CapsNets, which will be compared with their associated accuracy in section 4, we consider the model which has 64 feature maps in the convolutional layer instead of 256 as is the case in the original architecture.

Summary of the layers of our proposed model (as illustrated in Fig. 1) is as follows:

- MRI(Magnetic Resonance Imaging) imagesARE Inputs to the model which are down-sampled to 64 × 64 from 512 × 512, in order to decrease the training time and reduce the number of parameters in the model.
- Second layer will be a Convolutional layer with 64 × 9 × 9 filters and stride of 1 which results to 64 feature maps of size 56×56.
- Second layer is a Primary Capsule layer resulting from 256×9×9 convolutions with stride of 2. This layer consists of 32 Component Capsules which has dimension of 8 each of which has feature maps of size 24×24 (i.e., each Component Capsule contains 24 × 24 individual Capsules which are localised).
- Final Capsule layer includes several Capsules, which is referred to as "Class Capsules," one for each type of tumor. The dimension of these capsules is 16.
- The decoder part is composed of absolutely three fully connected layers having 512, 1024 and 4096 neurons, respectively. We note that, the number of neurons in the last Fully Connected layer is the equal to the number of pixels in the input image, as the goal is to minimise the sum of squared differences between input images and reconstructed ones.

There is one drawback that is observed by exploiting CapsNets for the problem in hand with several parameters to be learned and for the relatively small-scale dataset, which leads to over-fitting. We observed that the performance of the trained CapsNet based on the above mentioned specifications was high for training data, but degraded perceptible on the test data. In other words, care must be taken in the training stage to have a reasonable generalisation capability. We adopted "Early-Stopping Approach" to overcome this problem. According to the approach, at the end of each

epoch in the training process, model is tested on a validation set, and training continues to the point where the validation accuracy starts to decrease. For the goal of tumor type classification, two types of images can be used as the input to the before mentioned Capsule network. We can use either the whole tissue of an organ as the input, or instead only the tumor regions can be segmented first and then use these regions as the input to the classification model. As expressed in the Capsule Network, Capsules tend to model everything in the input image, thus they do not perform as good as potential for images with miscellaneous backgrounds. Due to this fact, we expect our Capsule network to have a better result when fed with segmented tumors rather than the whole image of organ. This is further explored next in Section 4.

4.EXPERIMENTAL SETUP

To test our proposed approach, we have used the MNIST data set . This data set contains 3, 064 MRI images of 233 patients diagnosed with one of the aforementioned tumor types. The most important property of this data set is that it includes both the complete image of the organ and the segmented tumor, which enables us to perform experiments on both types of inputs.

The first part of our experiment is allocated to testing different kinds of Capsule network architectures. We have changed different components of the original framework and calculated the prediction accuracy. According to thes e results, reducing the number of feature maps from 256 (as is the case in the original architecture) to 64 leads to the highest accuracy. However, there are many more architectures that can be explored, which is the focus of our ongoing research work. Next, we evaluate the total loss in a Capsule network, which is composed of two parts: CapsNet loss and Decoder loss. The former calculates the miss-classification error and is determined using Eq. (6). The latter is related to the reconstruction part and is calculated using the square error between the input and the reconstructed image. This loss contributes to the total loss with a smaller weight. We have trained our proposed architecture for 10 epochs and computed the three losses at the end of each epoch . It is observed that training is faster at the beginning and the total loss is mostly dependent on the CapsNet loss.
After selecting the best architecture for the Capsule network, we have compared its classification accuracy with a conventional CNN over the same dataset. The CNN used for comparison is adopted from, which has investigated the problem of different organ tumor

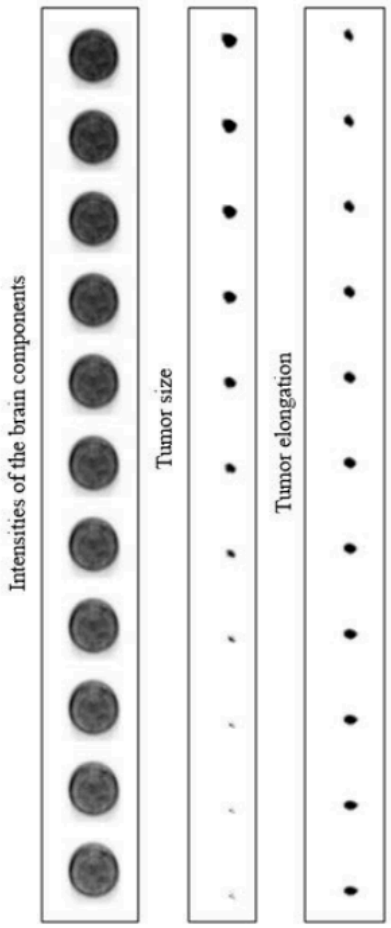

Fig. 2. Features detected by tweaking activation vectors of Class Capsules.

classification on the same data set. The layers of this CNN are constructed as follows:

- Convolutional layer with 64 × 5 × 5 filters and strides of 1.
- 2 × 2 Max-Pooling.
- Convolutional layer with 64 × 5 × 5 filters and strides of 1.
- 2 × 2 Max-Pooling.
- Fully connected layer of 800 neurons.
- Fully connected layer of 800 neurons.
- Fully connected layer of 3 neurons.

We compared Capsule network with the CNN for both complete image of organ and segmented tumors. Based on the comparison results, it is observed that CapsNet outperforms CNN for both types of inputs. As as stated earlier, Capsules tend to account for everything in the input image even in the background, and considering the fact that MRI images of organ are taken from different angles such as Sagital and Coronial, backgrounds have lots of variations. Therefore, CapsNet can not handle images as good as segmented tumor images, and this may be one of the reasons for the CapsNet architectures to provide lower accuracy for images. Features detected by tweaking activation vectors of “Class Capsules” in comparison to the case where segmented tumors are used as the input. Nevertheless, CapsNets

result is superior to that of the CNN for different tumor classification, which shows Capsule networks' advantageous over CNNs. One reason behind success of CapsNets in providing better tumor classification results can be attributed to the fact that CapsNets can handle data sets with smaller number of samples better than CNNs. Finally, we investigate the output of the last layer in the CapsNet (referred to as the Class Capsule), which is a vector containing the pose features, however, the CapsNet determines what features to learn on it's own. To provide better explainability, therefore, one option is to tweak them and try to reconstruct the input image using these tweaked vectors (tweaking refers to adding small numbers to the original vectors). When we visualized these tweaked vectors, one can identify/capture the nature of learned features. Fig. 2 illustrates some of the results together with what type of features they seem to be related to. Each column represents one particular reconstructed input using the tweaked features. For instance, the special feature learned in the second column seems to represent the size of tumor as tweaking this feature has changed the size. Similarly, the third column seems to be related to how wide the tumor is.

5. CONCLUSION

In future use of Capsule Networks for the classification of type of tumor can be used in many medical fields. Radiologists do not arrive at a diagnosis of lung cancer from a single CT scan they diagnose a particular type of lung cancer using a sequence of CT scans over a few months. They match the behaviour of the nodules over time with a particular subtype of lung cancer. To match radiologist-level accuracy on the task, CapsNet has been used to develop a time-varying model of cancer that can effectively include a progression of CT scans. The future scope of CapsNet is to automate all medical surgery where it can be guided by machines. Even the non-skilled or the less skilled physician can also perform complex surgeries at low cost.

6.REFERENCES

[1] Sergey Ioffe and Christian Szegedy,
"Batch normalization:Accelerating deep network training by reducing internal covariate shift,"
arXiv preprint arXiv:1502.03167, 2015.

[2] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov,
"Dropout:A simple way to prevent neural networks from overfitting,"
The Journal of Machine Learning Research, vol. 15, no. 1, pp. 1929–1958, 2014.

[3] Experiments on learning by back-propagation" Technical Report CMU-CS-86-126.
Department of Computer Science, Carnegie-Mellon University.

[4] Krizhevsky, A., Sutskever, I. and Hinton, G. E. "ImageNet Classification with Deep Convolutional Neural Networks Advances in Neural Information Processing" 25, MIT Press, Cambridge, MA

[5] "RELU" by Krizhevsky et al., 2012

[6] S. Sabour, N. Frosst, G.E. Hinton, "Dynamic Routing Between Capsules," 31st Conference on Neural Information Processing Systems (NIPS 2017), 2017

[7] Y. Lecun , L. Bottou , Y. Bengio , P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, 1998.

[8]. Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Palc, PierreMarc Jodoin, Hugo Larochelle, "Brain Tumor Segmentation with Deep Neural Networks," Medical Image Analysis, vol. 35, pp. 18-31, Jan. 2017.