RACV2016 Reporter

These days I have gone to Shanghai to attend the RACV2016. During the four days I have learned a lot through the reports that were given by some experts on deep learning.

On 18th September,We mostly watched the poster hung on the wall,these poster are mainly about the deep learning methods,object recognition and image segmentation. What especially attracted my attention is the CNN Features off-the-shelf. It tells that we can extract the image features through two ways and improve the classification accuracy by data fusion.

First we know that CNN uses three features to improve the result of extracting image information. The first feature,locality,refers to the connection of a neuron only to the neighboring neurons in the preceding layer,so it has been thought that can reflect the inherent compositional structure of data – the closer pixels are in an image,the more likely they are to be correlated. The second feature of CNN is weight sharing,which means that different neurons in the same layer,connected to different neighborhoods in the preceding layer,share the same weights. It's a good strategy to reduce the parameter of our CNN model. Sharing is motivated by the fact that in the natural images,the semantic meaning of a pattern often does not depend on its location. Finally,the third architectural idea of CNN is pooling,which contains maximum pooling and average pooling. Pooling is essentially an operator that decimates layers. Through the three features,we can get some feature maps that contains the image information.

In the author's poster,he explain that we can use two CNN to get different feature maps. One includes locality information using small convolution kernel,the other can obtain global features using bigger convolution kernel. At last we can fuse the two types of features to get more information. This way is applying to the extraction of an image.

Another attractive poster tells about learning semi-supervised representation towards a unified optimization framework for semi-supervised learning. There are two key points,the first is to construct an affinity matrix. We use the affinity matrix to compute which label the data points belong to. Neither the initially given labels nor the inferred labels are used to update the affinity matrix,so we use the label that we have gotten or computed to update the affinity matrix. The other is performing label propagation,we use the self-taught machinery to optimize the semi-supervised model,which means the label we got can effect how we tag the next image or data points. Here we use self-expressive model to induce the affinity.

As far as I'm concerned,the two poster both tell two key points, reducing the parameters and augmenting the connection between pictures and sequence. The trend to optimize our model is to deep our neuron model,the results prove that the shallow networks are not potential,sometimes they can't extract the correct information of an image. In order to fit the our train samples,the way we need to do is to deep the networks and augment the connection between different objects.

On the 19th we listened to some reports given by some experts,the first speaker spoke in English,I can't understand her completely,I can only understand a little. Here are some my thoughts over her speeching.

She mostly explains that the view is important to understand a picture,we can get different information from different views,take the cylinder as an example,we think the object as a rectangle from the front view while we think the object as a circular from above,so from different view we can get different shapes,some information contribute to our classification while some have a bad effect. The important point is that how we can get a good view. The speaker also explains some main factor can effect the object extraction,they are as follows:

1.Pattern distinctness – a picture can have different patterns.

2.Color - sometimes the color of the object contains some important information.

3.Prior organization – the prior object can effect what the next object express. For example,the pencil places different places means differently.

4.The attribute of the object.

She also mentions that we can mimics the way how we human extract information from a picture,we can design a mechanism in which we can extract object information hierarchically. It's not a bad idea to use the hierarchical networks to extract information.

The second speaker is Jiaya Jia. His report is about image deblurring. Every day we take many photos of fantastic insights,but some of them are blurred. His main research is how to finish image deblurring. In his speeching,he has mentioned that the realization of the bilateral filtering. The code even covers just one line.

Another work he is doing attracts me a lot,he is devoting himself to realizing the human-computer interaction. We can ask the computer some questions over an image. For instance,we can give the computer an image,then we can ask the computer some questions that is about the image,the computer finally give us the answers. For me it's so difficult,because a picture can contain much information,what the computer should do is to understand what an image contains and what the object of the image is doing.

At last,he talks about some skills to optimize the model. He mentioned the ground truth,but I don't know what the meaning of the ground truth is. When I came back,I searched a lot,now I know that the ground truth is used to implement punishment mechanism. In other words,It's the benchmark,the result is close to the ground truth,we increase the weight of the connection,and vice versa.

In afternoon,we attend the RACV2016 match report,during the three reports,I get some skills to improve the results. I summed up some of the main points,here is my summary:

1.Prior information is important for get the correct expression of a picture,on the one hand,it's a part of the whole image,on the other hand,it can also be correlated to the next object. So we can build up a self-taught mechanism to reduce the build-in parameters and improve the accuracy of the classification.

2.It's essential to fuse the local features and global features,we can know that the local features contains subtle information while the global features contains the information about what the object express. Another skill is to choose suitable feature map according to your pictures,if you're using a small set of the pictures to train the model,you'd better to choose the local feature map,and it will perform better if you choose global feature maps when you're training your model using a big set of samples.

3.Using mirror image can decrease the error,for the mirror image can provide more information of a picture. Sometimes when we use a mirror image can avoid the misunderstanding for the image. Besides,mirror image can also increase the number of our samples when we train a model.

4.Model fusion. This is a way that can employ different models,as a result,different models can be applying to different scene classification,but how can we put a good use of these models at the same time? A solution is to fuse these model by weight,for example,we can give a model a suitable weight to stand for how the model contribute to the results. If so,we can realize the optimization by adjusting the weight.

5.If we aim at the classification of the video,it's important to use different types of the information,such as audio,picture frame and sound information. In the speeching of one team,it tells that picture frame contains more information than the others,and the accuracy of using all types of information is better than the accuracy of using each one. So this also shows that data fusion is an important way to improve the accuracy.

6.Some skills about how to raw a picture is important,we can crop the pictures if the pictures contains too much useless information. What's more,we can also add the paddings. Another way to process image is to encode the image and decode it,which means that we can transform the image to a multi-dimension spatial space. The reason that why we encode the image is that we can fit the image by increase the dimensions of the spatial space.

The next day,we attend a panel,the four speakers give us an introduction on their own fields. Then they have a discussion about the trend of the deep learning and give some suggestions about how to find a good job to us students. At the same time,they also tells about the what they and their company is doing. In this part,I get a lot of precious experience,they can be my leader for me to learn deep learning and to be a member of one team.

Finally,the last speaker tells about the gragh cuts for the image. He speak all in English and use a lot of professional vocabulary,so I can understand a little,I write down some words that he emphasizes. He refer to the submodular and the max flow,which I can't understand them at all. So my next step is to trying to understand them and find out the connection of these ideas.

During the four days,I leaned a lot and get some skills to improve the accuracy. Even more precious is that I have the discussion with those experts who have gotten many achivements. I know how to improve myself and what I should do. Their methods of learning can give me a lot of inspiration. They are the people I want to go beyond.

Ziqiang Zheng
25th September