Meeting Report of RACV 2016

Jinna Cui

2016.9.18-2016.9.20

RACV2016(Research and Application in Computer Vision 2016) is held during 9.18-9.20, it is hosted by CCF(China Computer Federation), and undertaken by Shanghai Tech University. This meeting is a good chance to exchange experience and ideas with other scholars. Also it is a good place to get answer of questions and clarify confusion we met in our daily study.

The part which attracts me most is the competition. There are three competition topics: Video Object Detection and Retrieval, Deblurring Blurred Vehicle Plate, Video Semantic Annotation. Different teams from different schools got great grades, most of them are from Tsinghua University, Peking University, Beijing University of Posts and Telecommunications University, Zhejiang University and Chinese Academy of Sciences. At the part of Video Object Detection and Retrieval, all three teams use the Faster R-CNN net. In their reports, they mentioned the difficulties and confusions during the preparation of the competition. Such as difficulties in detecting small objects and the problem of different aspect ratio, and they also share with us how they overcome these difficulties: extract features from many different convolution layers maybe very useful, in this way, some small objects detection can be better. And as they say, the mixed feature can express more image information. In our research, when we use CNN to extract features, especially some small features, this method maybe useful.

图 1: Challenge 1

图 2: Challenge 2

Besides difficulties and solutions, they also show how they optimize the detection such as iteration, add weight to boundary and invert the images to get the detection results and combine with the detection results of origin images. After these optimizition, they got higher accuracy rate.

图 3: Final result

Students from Peking University emphatically introduced their methods to extract features.

They combined low-level visual features and deep learning features. So that they can get both general features and detail features. This is really a good idea, both low-level visual features and deep learning features have their own superiority and weakness, a good combination can work effectively, they also get higher accuracy with this method. So when we do features extraction, we can try to use this method.
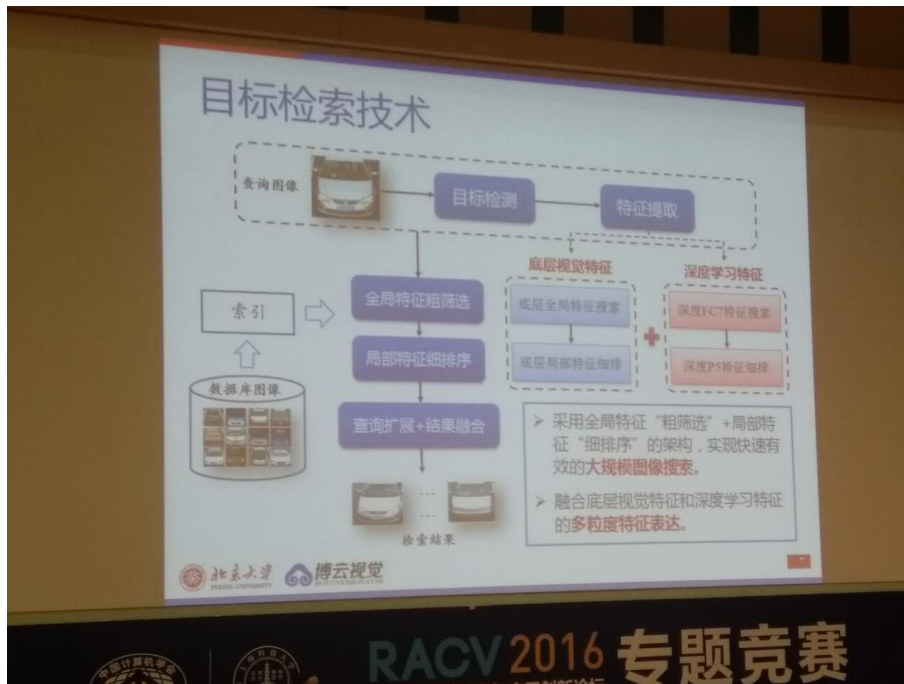

图 4: Combination of low-level visual features and deep learning features

And Zhejiang University students add anchor to different convolution layers and they divided the training into two parts: training motor vehicles and non-motor vehicles together, training non-motor vehicles and pedestrians together. Because non-motor vehicles both have many same features with motor vehicles and pedestrians. So the separation of different objects detection works well.

Conclusion: all of them used Faster-R-CNN to detect objects in the video, Faster-R-CNN still works well in video object detection, also we can see from it, in the video processing, the common method is CNN, they transform video processing into image processing, RNN is not commonly used in video processing. Although we all know RNN is really useful in sequence-signal processing, but it doesn't work well in video processing, they all mentioned that, they also want to apply RNN into video processing, but it seems a little difficult, maybe not a little. But at least, we know, there may be another way which may get better result.

In the part of deblurring blurred vehicle plate, different teams proposed different ideas, one thing exceeds my expectation is the effect, some very blurred samples even some samples angle makes the plate cannot be seen clearly . Team from NUPT use Interpolation algorithm. They used bicubic interpolation, more complex, but gets better result. Team from Fudan University firstly

divided blurred plate into four types, after that, they solved these four problems by four different methods. And they also use deconvolution and upsampling, their inspiration comes from CNN.

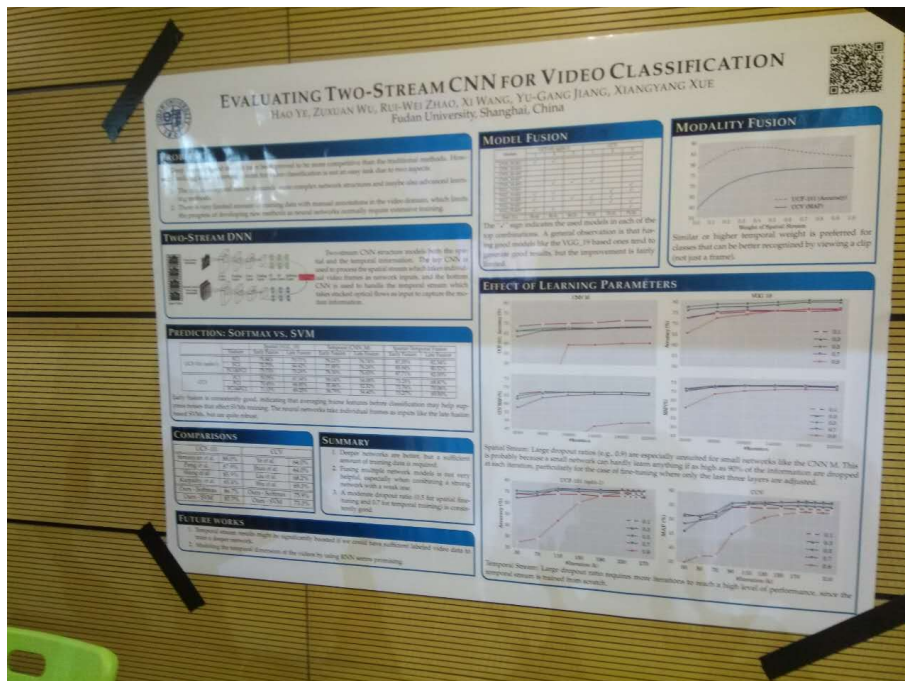图 5: Different kind of problems and different solutions

图 6: Deconvolution and upsampling

Third competition is video database. The same as first competition, all teams use ResNet. They still use CNN net to extract features from images, besides, they also mix up different features, like

images, sounds. They also metioned that, single features database is easier than multi-feature database, but they still get high accuracy in multi-feature database.

Beside the wonderful competition, there are also many amazing wall newspapers. One of them is from Xiamen University, their word is compressing to Convolutional Neural Networks, in fact they only do compression on fully connected layer, after their compression, the net can works as before. They said that they did this because fully connected layer need too many parameters, they want to simplify the net. In my oppoint, if the method is not very complex and very difficult to realize, we can try this method when we use CNN or MLP, everyone wants less parameters.
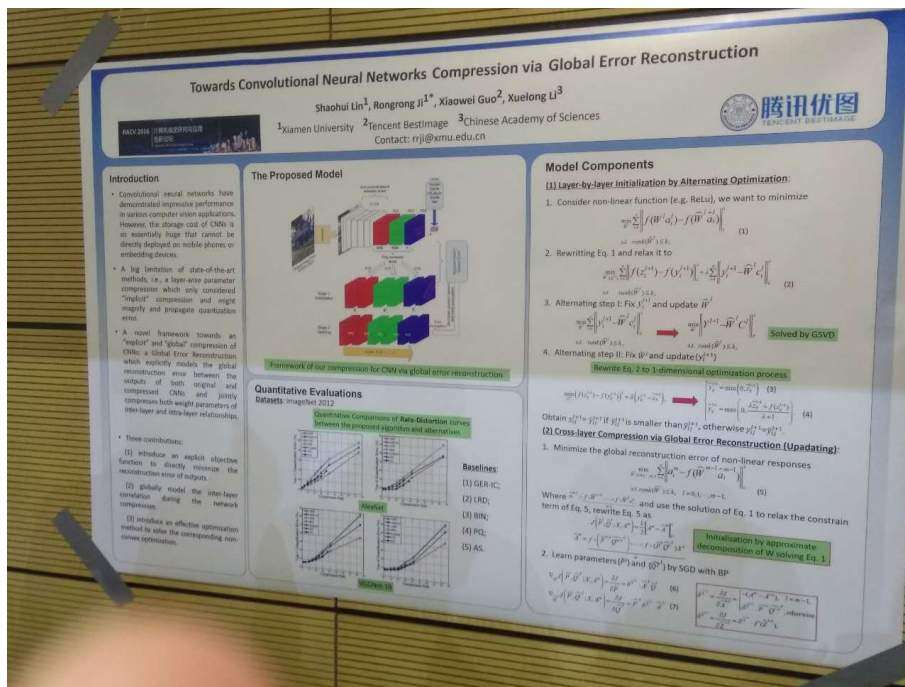

图 7: CNN Compression

We also see the DehazeNet: an end-to-end system for single image haze removal. This is a new application of CNN, this application also makes excited. CNN are usually used in features extraction, but they attempt to do haze removal by CNN, and they succeed they seems have got better rusults than traditional methods. This must be a new application fields of CNN in the future, I start to study CNN a few days ago, I also take great interest in CNN, then I want to read this paper meticulously.

Conclusion: about video processing, they still use CNN to extract features, someone also try to combine CNN and RNN together in video processing, but all people still not process videos only by RNN, although they all believe that RNN may works better than CNN in video processing, maybe this is a breakthrough point in the near future, and many scholars will work on it. Besides video processing, haze removal by CNN also makes me excited. As we all know, CNN are used to be applied to feature extraction, classification, object detection and so on. They are all based on feature extraction, but now, CNN also can be used on haze removal, CNN shows us its strength,

its power and its status that can not be replaced in computer vision filed.