

Peer Review to “On the Robustness of Multi-view Convolutional Neural Networks with Applications in 2D and 3D Image Classification”

From: Dazhi Li

Summary to the report:

This paper did a research on the robustness of MVCNN in 3D image classification when model structure is modified and 2D image classification accuracy comparison to SIFT+CNN2 (and CNN) when dataset is augmented (rotation/Gaussian noise). They have proven that MVCNN has a strong architecture which will not be influenced by structure changes. Although the result of SIFT+CNN2 is not ideal to them, they came up with potential problems and propose a new comparison to traditional CNN architecture to prove that MVCNN is still useful and even better than state-of-the-art CNN in 2D image classification.

High level discussion:

From my own perspective, this paper makes reader easy to understand. They have talked what is Convolutional Neural Network from the basic idea of convolutional kernel and try to explain everything clearly. However, I think there should be more introduction about MVCNN about what kind of problems it is proposed to solve and how it can solve this kind of problem better than state-of-the-art CNN. I mean this paper could explain more on MVCNN since it is the main topic. Also, I think there should be a motivation part about why this paper wants to prove the robustness of MVCNN. The result is significant, but I just want to understand the motivation. The experiment is well done and interesting to a large audience. For me, I understand that there is a robust CNN model named MVCNN which can do both 3D and 2D image classification job even better than most CNN networks. I guess readers like me would learn a lot from this paper. Finally, the experiment result is reproducible although I have not tried yet.

Low level discussion point:

I particularly like the result part of 2D image classification experiment. When using SIFT+CNN2 does not show us an ideal result they just throw out their guess on potential problems. And they still use experiment to prove that they are correct. And they are so clever to use another kind of CNN to finish their experiment in alternative way. Here I have a question that, I remember SIFT feature extraction usually operates on grayscale image, but CIFAR-10 dataset provides colorful images. Do you just transform the image to grayscale image, or you tried in another way? If you just use grayscale image, have you considered the information loss? Moreover, I wish I could see more structural changes on MVCNN. You have proved that it is robust, but I wish to see how robust it is if you have further plan on it. Anyway, it is a good paper and interests me a lot.

Nitpicks:

1st page: 12th line in Abstract, add “a” before similar performance or use similar performances

2nd page: 2nd line after figure 1,

Please use “the” instead of “then”

3rd page: 10th line “we describe in the following sections”,

It would be better to say “we will describe in the following sections”

4th page: Last 2 line, “Putting this altogether”,

Please use “these” instead of “this”

5th page: 3rd line in Related works, use “directly” instead of “direct”

6th page: 2nd line in Figure 2 annotation, add “the” before “label ‘airplane’”

8th page: 6th line in 3D Image Classification Experiment,

Please use “use” instead of “uses”

10th page: 6th line in SIFT+CNN2,

Please delete the “of” before “the same”

13th page & 3rd page: “We direct the reader to”

I have looked up the dictionary about “direct somebody to do something”, it said we usually use direct somebody to somewhere. If we use direct somebody to do something it means authorized push somebody to do something. I think there is a better expression this place.

My peer review summary:

I think they are doing a great job as their experiment is complete. Their paper is very clear as they even explained what CNN is. At last, I think their paper is interesting for reader to learn something like how to find problem and correct it through unideal experiment.