

Responses to Reviewer 2 – round 2 :

The author made a revision to the paper based on my opinion and answered all my questions.

But I still have some comments and hope the author can further explore in this topic. The author demonstrated the GAP (global average pooling) has greater advantages than fully connected layer in the problem issued in their work. In addition to numerical results, maybe the author can also analyze it from other aspects, such as the features within the CNN networks, to give readers more insightful information.

We would like to thank the reviewer for the new feedback on our manuscript. Unfortunately, we cannot provide a sample of the feature maps with our test dataset for confidentiality reasons.

For the networks with and without GAP, the feature maps up to Conv 4 will be very similar. However, with GAP we use 2D feature maps (Conv 5) instead of a fully connected layer. We do not have developed techniques to compare the 2D feature maps to their equivalent in the fully connected layer.

Some intuitive insight on the potential advantages of GAP is available in the original article from Lin et al. We see two benefits to the use of GAP in our network. The first is the direct correspondence between a feature map and an output class directly from the start of the training. They will ultimately represent “categories confidence maps” and preserve the spatial information up until the very last layer of the network. The second benefit is the averaging operation, which sums this spatial information and produce an output that will be robust to small translations of the input.

When using fully connected layers we lose both aspects. Direct correspondence between units from the last layer is not present and by flattening the feature maps from conv 4 we also lose the spatial information.

We added a paragraph in section 4.1 to explain our choice of GAP and introduce the elements explained above in the manuscript.