

Deep Learning
Prof. Mitesh M. Khapra
Department of Computer Science and Engineering
Indian Institute of Technology, Madras

Lecture – 95
Occlusion experiments

Ok, the another thing that you can do is to figure out whether things are working properly or not. So, you can do something known as an Occlusion Experiment. So, these are all your debugging tools sort of to say if you are working in Vision or computer Vision where you are using a convolutional neural network and this is to gain more insight said most of you will get away by just taking an off the shelf convolution neural network training it on your data getting some accuracy and reporting a trade.

But, for those of you who want to really understand what is happening and how can you improve things further. So, this could actually tell you for example, if you want to compare whether a 5 cross 5 filter would have been better than a 7 cross 7 filter. Then you could have observed what these filters are actually learning and in your data does it make sense to have a 5 cross 5 filter versus 7 cross 7 filter. Because maybe the 5 cross 5 filters are not being able to distinguish enough. But, if you had used a smaller or a larger filter things would have been different, right.

So, this is for people who really want to get into the knowhow of how things are working otherwise most of you I do not really expect you to do this is, but this is an important set of tools to have and I would strongly encourage everyone to experiment with them and some of this you will do in the assignment, ok.

(Refer Slide Time: 01:16)

softmax pomeranian wheel ... hound

- Typically we are interested in understanding which portions of the image are responsible for maximizing the probability of a certain class

Mitesh M. Khapra CS7015 (Deep Learning)

So, here is the idea of behind occlusion experiments. So, we are interesting knowing that what patches in the image are actually causing the output to belong to a particular class, right.

So, I have here the figure of a dog and the class being probably predicted is a Pomeranian and I want to know that what patch of the image actually resulted in this output, right. So, have you tried doing this in any other context; if you want to know if you have several features or several things several factors and you want to decide which actually influenced the output, how would you do it? So, what you could do is you could drop one factor, right and see whether your output would have drastically changed if it goes from positive negative then that maybe that was the factor which really mattered, right.

So, if for example, it is a movie review classification, right. So, when you drop certain words from your review. So, you drop the word amazing, great and so on and keep everything else the same. Now, it is quite likely that your probability of the review still being tagged as a positive review will at least drop earlier maybe with these words it was getting tagged as a positive review with 0.9 probability it would come down to 0.6, but now if you drop words like the, and, for and so on then you do not expect the output to change much because these words are not really important indicators of positive or negative, ok.

So, the similar thing that you do here is you occlude certain patches of the image. So, I have shown one occlusion here so, I have replaced that patch by a gray patch and I again feed the image to the convolutional neural network and I see what is the probability of the Pomeranian class right now, and I do it for all such patches in the image. I can do for as many patches as I want.

(Refer Slide Time: 03:03)

softmax pomeranian wheel ... hound

- Typically we are interested in understanding which portions of the image are responsible for maximizing the probability of a certain class
- We could occlude (gray out) different patches in the image and see the effect on the predicted probability of the correct class
- For example this heat map shows that occluding the face of the dog causes a maximum drop in the prediction probability
- Similar observations are made for other images

NPTEL Mitesh M. Khurana CS7015 (Deep Learning) 10/51

And, I create something known as a heat map. So, the red portions here are the ones which do not cause a large drop in the output probability if you occlude them and the blue portions are the ones which cause a large drop in the output probability if you occlude them and it is pretty obvious. Because what is happening is when I cover the face of the dog the probability drops drastically and that is what you would expect, right. So, this is also an indicator that your network has actually learned something meaningful. It is being able to detect this based on the facial features and not just randomly guessing that this is a dog, right and see the similar experiment.

So, for example, if there is a car, sometimes these results are not very at least to me it does not look very intuitive. So, I would have expected that the wheel would have been one of the deciding factors, right. So, if I occlude the wheel the probability should drop drastically, but the other way of looking that it is that it is really learning a lot of redundant features. So, it is not heavily relying on the wheel unlike in the dog case even

if the wheel is occluded it is relying on certain other features which look like cars and hence the probability is not dropping drastically, right.

So, this allows you to interpret what kind of things it is running. So, if its heavily for example, for face detection if its heavily relying on nose to detect the face to say that this is the face; the moment you block the nose it will drop its probability of detecting this as a face, but thats not good right because you want these redundant features. Remember we had discussed this at some other point where it should try to detect the face not only from the nose, but also from the ears, from the hair, from the eyes and so on.

So, if your occlusion is not drastically reducing your probability; that means, it has learned some redundant features which are still allowing it to operate well even though certain portions of the image are not there; that means, it is more robust noise in that case, right. And, here it looks like it is not so robust because it is probably heavy this is the rearview mirror of the car. So, it is probably heavily relying on that feature to detect a car, ok.

Then this is another thing where the true label is an Afghan Hound and for some reason if you occlude the face of the woman its probability decreases, now let us not comment on that, but if you go back and look at the image you might be able to make some observations, right. So, these are things so, this is an indicator that is probably not really learnt it well maybe all the Afghan Hound images that it saw maybe a woman was carrying the dog always, right.

So, its learn this wrong association that when I see a woman with some object it that is the portion which is the dog which is bad, right. So, now, you can see that your network has not learned something interesting and you would want. So, if you look at one network which is predicting a dog based on this kind of a occlusion and another network which is predicting a dog based on this occlusion then you would prefer the other one and so, this is a very interesting experiment to do.