#ret

On the exploratory notebook, I agree with your answers on exercise 0, although I think you could be more clear on your answer to problem 2. Why do repeated values imply optimal values? Are the values returned by image 2 actually optimal? You successfully found two images that produce positive values on exercise 1, and you also made new filters that behaves as you predicted, producing a positive value on the gradient image. I suspect, though, that this is not quite what you were trying to do — it might be interesting to come back to this one and think more about how you would detect a gradient. On the tuning exercise, you successfully increased accuracy to 81%, and I liked that you included all your code for both the models that didn't work that well, as well as the ones that did. You provided a brief summary of your results as to what worked better or worse, but I would challenge you to think more deeply about this: why do you think these are true? Are there experiments you could run that would prove some of your suppositions to be true? For example, you say that more convolutional layers might make things worse because we only have two epochs. How would you test to see if this is in fact the reason?

As for problem 2 with the repeated values, intuitively speaking, given that each group of values is repeating some number as opposed to repeating the same collection of different numbers their must be one optimal group which repeats the highest number. As for the gradients, without either large kernels or some type of memory or hidden cell state, I don't know how to make filters what would check for the gradients consistent between sections. I would love to know the answer to this.

To test my hypothesis for my model, I could always run different types of models for more than 2 epochs and see how they perform. Less layers might also work better because of overfitting, or simply because the features we are trying to pick up aren't that detailed.