ECE472 – Quiz 7

Jonathan Lam

October 21, 2020

1. *Which component method do you find most compelling from all of these papers? Justify what's interesting to you.*

   After reading a few papers on ResNets and its variants, the MSG-GAN proposal was one that I thought was most intuitive and compelling. The original ResNet paper (arXiv:1512.03385) proposed some skip connections, and the authors suggested that this helped the gradients "flow" easier through the network, allowing for faster training and deeper networks. Then we read about variants such as restructuring the network so that the "ResNet blocks" were only connected via identity mappings (arXiv:1603.05027), which further un-obstructed the flow of the gradient through the network. Then we read a paper on DenseNets (arXiv:1608.06993), which create many more skip connections in a groups of ResNet blocks, which further improved the performance of (and decreased the depth necessary for) ResNets.

   The overall theme in these (as opposed to "regular" neural networks without skip connections) seems to be that the more skip connections there are, the easier the gradient can flow through the network. Thus, the aim of this paper (arXiv:1903.06048) seemed pretty intuitive to me: add skip connections between the generator and discriminator to promote gradient flow. Like before, the generator and discriminator are all connected as one network, with the generator's output fed into the discriminator. However, previously the output of the generator was the only input to the discriminator; MSG-GAN proposes that all of the intermediate resolutions used by the generator while generating the final image should be fed to the discriminator. This seems very much like skip connections to me.

   The authors of MSG-GAN aim to tackle the instability problem, which they state is due to "insubstantial overlap between the supports of the real and fake distributions," i.e., if the fake image distribution is too far from the real distribution and the generative model does not get informative gradients back. These connections between the generator and discriminator were aimed towards increasing the gradient flow at different resolutions (without adding extra generators and/or discriminators, and without adding extra hyperparameters like in progressive growing). I found it interesting that this intuition indeed helped in this case as well; the authors state that it improved stability, robustness to learning rate, and led to higher-quality images (using the FID metric).

2. *Are there methods that you think are unreasonable?*

   When reading these, I didn't find any unreasonable per se, i.e., such that I don't believe that their methods are sound. I didn't have the time to spend on understanding the math-theoretical argument proposed in "Smoothness and Stability in GANs" (arXiv:2002.04185), so that might be the most unreasonable in a loose definition of "unreasonable." From the abstract, I understand that the authors are attempting to better understanding the inherent instability in GANs. It's very difficult to understand the specifics of their research without any diagrams or simulations, or at least without the math background (e.g., understanding "Jensen-Shannon divergence ... f-divergences ... Wasserstein distance ..., and maximum-mean discrepancy ... Kullback-Leibler divergence"). Without any simple examples, diagrams, or source code, I believe that it is very difficult for anyone without a great knowledge of convergence theory or loss functions to be able to interpret and use these results. Again, this may not be unreasonable because I am not likely the intended audience of this paper (but rather researchers who are hoping to understand GAN stability better), but no matter the reader I think a concrete example would be very helpful.

   (On the other hand, the other three papers all were more reasonable to read and understand, and all provided a suitably-intuitive motivation and concrete examples. The StyleGAN paper (arXiv:1912.04958) aimed to fix GAN image artifacts due to the model forcing itself to fit the normalization method, and the spectral normalization paper (arXiv:1802.05957) attempts to improve stability by doing data-independent spectrum-based (eigenvalue-based) normalization on its weights. The MSG-GAN paper, discussed in the first question, made sense in the context of ResNets. I would say that all of these are reasonable.)