CS 4644/7643: Deep Learning
Spring 2024
HW4 Solutions

ARTHUR SCAQUETTI DO NASCIMENTO

Due: 11:59pm, April 4, 2024

4.1.1) (Denoising Diffusion Probabilistic Models)

- **Key contributions:** This work proposes a novel technique for generative modeling. The process is essentially training a denoising autoencoder to perform Markov Chain Monte Carlo sampling, which is completely different from other generative models like GANs and VAEs. To the best of my knowledge this is the paper that introduced diffusion models as an option for generative AI.

- **Strengths:**

  - This paper introduced one of the most powerful techniques for denoising diffudion models. The mere fact that it had beaten all benchmarks for a relatively simple dataset (CIFAR10) is already a strength.
  - This is one of the few works in ML/AI where the authors are very explicit about why they came up with that frameworks, and how they did it backed up by a lot of math. As opposed to papers that achieve a phenomenal performance in whatever task they do, but at the end you don't get why it works they way it does; works where they either over rely on an assumption or over engineer some blackbox.
  - If my thought process is correct, they introduced diffusion models to genAI. This is arguably the highest strenght, given the results we see from Midjourney, DALL-E, etc.

- **Weaknesses:**

  - This work only displays image generation, while it could test and showcase the generation capabilities on different domains. Specifically, I expected/wanted to see policy or trajectory generation, especially given that the most senior author is Pieter Abbeel.
  - A light weakness is the heavy language. Even for a NeurIPS paper, this is a very dense work to read. Even though there is a lot to explain in a limited amount of pages, I think that the authors could have done a better job in moving content around, resizing and rephrasing or even reducing the scope of this work for the NeurIPS submission, and publish the complete, longer version in a journal, with more room for writing.

4.2.1) (Denoising Diffusion Probabilistic Models) Personal takeaways:
The first thing I takeaway from this paper is that I should actually read a paper that has some sort of "hype" regardless of the heavy language and that I take over 15min to understand even the abstract + conclusions. Not a very pleasant takeaway, but that is something I am incorporating in my research style.
Another is that diffusion models are not just something to generate images from text. That said, I have a better understanding of one of my labmate's work, and wonder how this could be incorporated to model-based RL to generate policies for a POMDP. That is something I might be interested on doing in the upcoming months.

4.1.2) (High-Resolution Image Synthesis with Latent Diffusion Models)

- **Key contributions:** In contrast to DDPM, where VAEs where completely set aside, this paper puts together a smart architecture for combining diffusion models with VAEs (taking advantage of the dimensionality reduction in the latent space), which, combined with transformers, was a huge leap to achieve those amazing results from text-to-image (now even text-to-video) generation.

- **Strengths:**

  - Just like the DDPM paper, this is work is a hallmark just by the fact that the authors achieve state-of-the-art for some set of scores in different tasks. While some might not see this as an academic strength, I do and my point for that is that novel research is nothing but small increments of knowledge on major building blocks.
  - The authors introduce a very clever architecture (arguably overengineered, but I don't think this is less of a merit) that drastically decreases computational resource consumption for tasks as complex as generating high quality images.
  - This work introduces the possibility of many modalities of conditioning, including text and semantics. I am not sure if this is the first work that did this.

- **Weaknesses:**

  - Although they did a good job with the LDMs, they did not achieve high quality for details, such as specific edges and texture.
  - This is yet another paper that "just" puts two things together, achieve good results and just show the math that is already in the literature. They don't back their approach with math, as the DDPM paper did. They are basically working with blackboxes that happen to work well for those tasks with enough tunning.

4.2.2) (High-Resolution Image Synthesis with Latent Diffusion Models) Personal takeaways: Even though I listed putting two things and making them work as a weakness of this paper, a personal takeaway is that maybe it is ok to just do that. Maybe the math can be revisited later and more conclusions could be drawn? I am saying this because they did pave the road (at least to some extent) to SORA, so maybe showing two black-boxes together and show a proof of concept is enough for the progress of science as we know it. That is a more philosophical takeaway, but epistemology has been something that has been haunting me.