Mosaic: Generating The New Yorker Style Cartoons using Text-to-Image Diffusion Models

Mehul Sudrik, Rajesh Nagula, Utsav Oza

ECE-GY 7123 Deep Learning

Problem Statement

> "Sometimes I wonder if I'm too old to be a cartoonist, but then I remember that I'm just not funny enough."

This self-deprecating joke highlights the challenges of creating The New Yorker style cartoons, that uniquely combines a whimsical art style, witty humor, and a subtle commentary on modern life. Cartoon enthusiasts and creative professionals alike know that creating such cartoons can be a daunting task that requires both artistic talent and a knack for satire. In this project, we aim to explore techniques to simplify the cartoon creation process by using Text-to-Image Diffusion models to specifically generate high-quality The New Yorker style cartoons from natural language captions.

Literature Survey

Text-to-image diffusion models (Saharia et al. 2022) are a type of generative model that can produce high-quality images from textual descriptions. Diffusion (Ho, Jain, and Abbeel 2020) is a probabilistic process that involves gradually adding noise to an image until it becomes completely random, and then gradually removing the noise until it becomes the target image. The goal of diffusion models is to learn the latent structure of a dataset by modeling the way in which data points diffuse through the latent space.

The challenge, however, is that it is unclear how the diffusion process can be directly exercised to generate images of specific unique concepts, or compose them in new roles and novel scenes. Textual Inversion (Gal et al. 2022) allows us to teach text-to-image diffusion models new concepts - it takes a very small number of images of a user-provided concept, like an object or a style, and learns to represent it through new "words" in the embedding space of a frozen text-to-image model. These "words" can be composed into natural language sentences, guiding personalized creation in an intuitive way.

Project Considerations

The latest version of the AAAI style file is available on AAAI's website. Download this file and place it in the TEX search path. Placing it in the same directory as the paper should also work. You must download the latest version of the complete AAAI Author Kit so that you will have the latest instruction set and style file.

Dataset

In the LATEX source for your paper, you **must** place the following lines as shown in the example in this subsection. This command set-up is for three authors. Add or subtract author and address lines as necessary, and uncomment the portions that apply to you. In most instances, this is all you need to do to format your paper in the Times font. The helvet package will cause Helvetica to be used for sans serif. These files are part of the PSNFSS2e package, which is freely available from many Internet sites (and is often part of a standard installation).

Model

In the past, AAAI has corrected improperly formatted files submitted by the authors. Unfortunately, this has become an increasingly burdensome expense that we can no longer absorb). Consequently, if your file is improperly formatted, it will be returned to you for correction.

In the past, AAAI has corrected improperly formatted files submitted by the authors. Unfortunately, this has become an increasingly burdensome expense that we can no longer absorb). Consequently, if your file is improperly formatted, it will be returned to you for correction.

Goals and Deliverables

Please check all the pages of your PDF file. The most commonly forgotten element is the acknowledgements — especially the correct grant number. Authors also commonly forget to add the metadata to the source, use the wrong reference style file, or don't follow the capitalization rules or comma placement for their author-title information properly. A final common problem is text (expecially equations) that runs into the margin. You will need to fix these common errors before submitting your file.

References

Gal, R.; Alaluf, Y.; Atzmon, Y.; Patashnik, O.; Bermano, A. H.; Chechik, G.; and Cohen-Or, D. 2022. An Image is

Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion. arXiv:2208.01618.

Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising Diffusion Probabilistic Models. arXiv:2006.11239.

Saharia, C.; Chan, W.; Saxena, S.; Li, L.; Whang, J.; Denton, E.; Ghasemipour, S. K. S.; Ayan, B. K.; Mahdavi, S. S.; Lopes, R. G.; Salimans, T.; Ho, J.; Fleet, D. J.; and Norouzi, M. 2022. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. arXiv:2208.01618.