ITAI 1378 – Computer Vision Artificial Intelligence

Andrew Badzioch

Kolby Boyd

Florentin Degbo

Natalia Solozano

Jeralynn Sparks

Professor Patricia McManus

Spring 2024 CN: 15259

L01: Exploring Real-World Applications of Computer Vision

Artificial Intelligence has always been a topic of discussion for many decades now, mostly thought of as a tool to ease the workload of humanity on their day-to-day tasks in the workforce. However, the development of technology has made this tool more readily available, resulting in more innovative and easier ways to let the rest of the population interact with A.I (Artificial Intelligence). on a personal or professional level. This progress brings us to an application called Dall-E 2. Dall-E 2 is an A.I. tool that utilizes different methods, which are discussed later in this essay, to generate images from textual descriptions that are unique in their creation. It also could edit existing photographs through a feature called "in-painting."

One vital question would be: how are these images being created by the application? Dall-E 2 makes use of neural networks which register, code, decode and learn from the datasets provided to it. Beyond just the comprehension of what the text says, it can figure out the relationship between objects and distinguish the complexities of shadow perception and depth that only the human eye and brain can capture. This is achieved by the following process:

- Text Encoder

- Text Embedding

- Prior

- Image Embedding

- Decoder

The text encoder determines the key words from the textual sample and the text embedding step runs them through an A.I. neural network called CLIP (Contrastive Language Image Pre-Training). CLIP was trained to learn the relationship between images and text to be able to generate a picture which is then processed in the image embedding step, also using CLIP. It is important to mention that between those two steps Dall-E 2 implements a diffusion prior model consisting of a network that adds noise to an image until it is unrecognizable, then it removes the noise to be able to learn how to generate an image. The decisive step is the most important one in the Decoder section as it uses a modified diffusion model called GLIDE (Guided Language to Image Diffusion for Generation and Editing), that during the training phase adds textual information to the generating process to give the network the ability to learn the difference between what is in the picture. For example, the breed, color, and size of a dog provides a more human-like approach on how our eyes and brain capture images and comprehend the various aspects inside an image. Through this training, the application achieves an editorial approach to image generation, and with the two upscaling steps it can render a better resolution image at the end.

Even though Dall-E 2 was revolutionary in Image Generating A.I., it does contain some limitations. One technical limitation would be that it is not good at generating images from a rather complex coherent text. Another limitation would be an ethical one, since the system is working with information from the internet that might have a biased dataset resulting in the network to learn relationships within words and textual information with a more closed-minded approach.

In conclusion, the application Dall-E 2 alongside other applications, like ChatGPT, have brought the artificial intelligence to the larger masses instead of just the tech-savvy community, allowing the general population to be included in the conversation by providing consumer-based feedback to the industry. One negative aspect could be related to copyright, one should reflect upon who has ownership of A.I generated content: the company that created the app or the individual who provided the sample prompt for the network to run?

References

5 use cases of dalle-3. KDnuggets. (n.d.). https://www.kdnuggets.com/5-use-cases-of-dalle-3

Kerner, S. M. (2023, April 21). *What is dall-e (dall-e 2) and how does it work?*. Enterprise AI. https://www.techtarget.com/searchenterpriseai/definition/Dall-E

Dall·E 2 explained - model architecture, results and comparison. YouTube. (2022, May 10). https://youtu.be/Z8E3LxqE49M?si=VhtOrPMptBlBZVd9