

**“Stable Diffusion - Image to Prompts.”**

**Group 7**

**Aman Patel, Kamal Nileshbhai Panchal,**

**Praveen Kumar Thanniru, Rahul Reddy Parupati**

**Department of Applied Data Science, SJSU**

**DATA 255: Deep Learning Tech.**

**Prof. Simon Shim**

**14th March 23**

### **Abstract**

The project The Stable Diffusion - Image to Prompts : Stable diffusion has emerged as a promising approach for generating high-quality images from textual prompts in the field of deep learning. This approach involves iteratively refining a set of noise vectors using a diffusion process, where noise is gradually added to the image at each iteration, followed by a diffusion step that smoothes out the noise. In the context of image to prompt generation, stable diffusion involves generating an image from a textual prompt by iteratively refining a set of noise vectors that are conditioned on the prompt. The generated image can then be used to further refine the prompt or to generate related prompts. Although stable diffusion requires a large amount of computational resources, it has shown promise in generating high-quality images from textual prompts and is likely to continue to be an active area of research in the field of deep learning. The Kaggle competition aims to find creative ways to create natural language prompts from photos. A critical challenge for applications like image retrieval, content-based picture indexing, and image captioning is the creation of descriptive prompts for images. Yet it's still difficult to produce top-notch natural language descriptions for intricate scenarios. The objective is for them to come up with innovative ways that surpass current challenges regarding formulating top-notch guidelines applicable in complex situations. Those partaking are given an opportunity not only to improve their aptitude but also gain valuable experience on how best they can apply natural language processing techniques alongside machine learning principles towards solving various problems encountered within our present-day existence. Overall, the Stable Diffusion - Picture to Prompts Kaggle competition offers an interesting chance for participants to develop image

identification and natural language processing while simultaneously fighting for recognition and awards.

**Dataset link**

<https://www.kaggle.com/competitions/stable-diffusion-image-to-prompts/overview>

**Description**

The Stable Diffusion - Picture to Prompts dataset is made up of a sizable number of pictures and the natural language questions that go with them. The dataset contains a wide range of photos, from straightforward items and sceneries to intricate scenarios with several elements and connections between them. The pictures in the collection are from many open datasets, such as COCO, OpenImages, and Flickr30k. Each image in the collection is tagged with a corresponding natural language question that defines the content of the image. A training set and a validation set are created from the dataset. The validation set is used to assess how well the models created by competition participants perform. It is smaller than the training set and comprises fewer pictures and prompts.

**Proposed Technologies**

Technologies likely to be used in this project are namely:

Python: A high-level computer program that is frequently employed in projects involving deep learning and machine learning.

TensorFlow: An open-source platform that offers a variety of tools for creating and training deep neural networks and is used for constructing and deploying deep learning models.

Keras: A high-level neural networks API which may be installed on top of TensorFlow that enables quick deep learning model exploration and prototyping.

PyTorch: A dynamic computation graph-supporting open-source machine learning toolkit that offers a versatile and effective framework for developing deep learning models.

NumPy: A Python package that offers assistance for processing data in huge, multi-dimensional matrices and arrays, which is frequently used in deep learning projects.

Matplotlib: A Python package for visual development that may be used to analyze and display model results.

OpenCV: Preprocessing the input photos for the model can be done using an open-source computer vision library that offers a variety of tools for image processing and computer vision applications.

## References

<https://jalammar.github.io/illustrated-stable-diffusion/>

<https://github.com/CompVis/stable-diffusion>

"Open Images 2019 - Object Detection" competition on Kaggle:

<https://www.kaggle.com/c/open-images-2019-object-detection>

"Stable Distributions Through Noise for Improved Training of Deep Neural Networks" by Pratik Chaudhari, Adam Oberman, and Stefano Soatto (<https://arxiv.org/abs/1906.11052>)

"Diffusion Models Beat GANs on Image Synthesis" by Sergey Bartunov, Dmitry Vetrov, and Andrew Sosnovskikh (<https://arxiv.org/abs/2105.05233>)

"Text2Shape: Generating Shapes from Natural Language by Learning Joint Embeddings" by Ankush Gupta, Julia Letchner, Srikumar Ramalingam, and Derek Hoiem (<https://arxiv.org/abs/2011.08751>)

"Generating Images from Natural Language Descriptions with Deep Learning" by Scott Reed, Zeynep Akata, Xincheng Yan, Lajanugen Logeswaran, Bernt Schiele, and Honglak Lee (<https://arxiv.org/abs/1511.02793>)

"ControlGAN: Conditional GAN with Intrinsic Rewards for Image Synthesis with Controlled Semantic Complexity" by Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila (<https://arxiv.org/abs/2008.00951>).