Important terminologies

Vocabulary scene control:

Vocabulary scene control refers to concept or a capability within the computer graphics and virtual environment that allows for flexible and unconstrained control over scene generation or manipulation without pre defined limitations on the vocabulary or scene elements that can be used.

CLIP:

Contransitive language image pre-training enables the training of a single neural network to understand both text and image in a unified way.

Dense segmentation map:

Map refers to a detailed and comprehensive output of segmentation process applied to an image .

Zero shot open model:

Zero shot open model is a naturl language processing model that can perform task without specific training examples, these models are used to generalize across wide range of tasks and topics

Inference:

Refers to the process of making predections decisions and conclusions in the content of artificial intelligence and machine learning

Spatio-textual scene matrix:

The term spatio textual scene matrix appears to refer to a structured representation used in the content of understanding or scene analysis combining visual and textual.

Diffusion model:

Diffusion model holds the responsibility to add noise to the image and learn how to reverse it. In an image generation diffusion model will be adding noise can be consider as adding Gaussian noise. It is noise that has Gaussian distribution or normal distribution

DALLE2:

Distributed and auto regressive language learning with encoding

Prior model:

Trained to translate the tuple(x, y) where x is CLIPimg and y is CLIPtxt.

Decoder model:

Model that converts image into low resolution of the image

Super resolution model:

A super resolution model turn low resolution to higher resolution of the image

Fine-tuning:

Pre-trained model is further trained on a new data-set to increase the performance or to adapt.