Prompt Engineering for Vision Models

Duration: 01 days (08 hours)

In this course, you'll learn to prompt different vision models like Meta's Segment Anything Model (SAM), a universal image segmentation model, OWL-ViT, a zero-shot object detection model, and Stable Diffusion 2.0, a widely used diffusion model. You'll also use a fine-tuning technique called DreamBooth to tune a diffusion model to associate a text label with an object of your preference.

In detail, you'll explore:

Image Generation: Prompt with text and by adjusting hyperparameters like strength, guidance scale, and number of inference steps.

Image Segmentation: Prompt with positive or negative coordinates, and with bounding box coordinates.

Object detection: Prompt with natural language to produce a bounding box to isolate specific objects within images.

In-painting: Combine the above techniques to replace objects within an image with generated content.

Personalization with Fine-tuning: Generate custom images based on pictures of people or places that you provide, using a fine-tuning technique called DreamBooth.

Iterating and Experiment Tracking: Prompting and hyperparameter tuning are iterative processes, and therefore experiment tracking can help to identify the most effective combinations. This course will use Comet, a library to track experiments and optimize visual prompt engineering workflows.