EE P 5OO D: LLMs and ChatGPT || Course Recap

Dr. Karthik Mohan

Univ. of Washington, Seattle

November 19, 2023

Today

1. Course Recap
2. Course Evaluation
3. Mini-Project Demo and Presentations
4. Office Hours

Course Recap

LLMs

1. **Language Model:** A mathematical and computational model for machines to understand Language

Course Recap

LLMs

1. **Language Model:** A mathematical and computational model for machines to understand Language
2. **Large Language Models (LLM)** - Trained for trillions of tokens - Equivalent of 50 Million Books

Course Recap

LLMs

1. **Language Model:** A mathematical and computational model for machines to understand Language
2. **Large Language Models (LLM)** - Trained for trillions of tokens - Equivalent of 50 Million Books
3. **Foundation Models:** Refers to large AI models that can then be fine-tuned on downstream tasks.

Course Recap

LLMs

1. **Language Model:** A mathematical and computational model for machines to understand Language
2. **Large Language Models (LLM)** - Trained for trillions of tokens - Equivalent of 50 Million Books
3. **Foundation Models:** Refers to large AI models that can then be fine-tuned on downstream tasks.
4. **Pre-Trained Models:** Typically refer to foundation models that have been trained on massive amounts of data. Includes LLMs such as BERT, GPT and also Large Vision Models

Course Recap

Models and APIs

1. **LLM API vs Foundation Model:** Foundation Models are the back-bone of an LLM API such as ChatGPT. LLM or Vision API typically has much more fine-tuning than a Foundation Model

Course Recap

Models and APIs

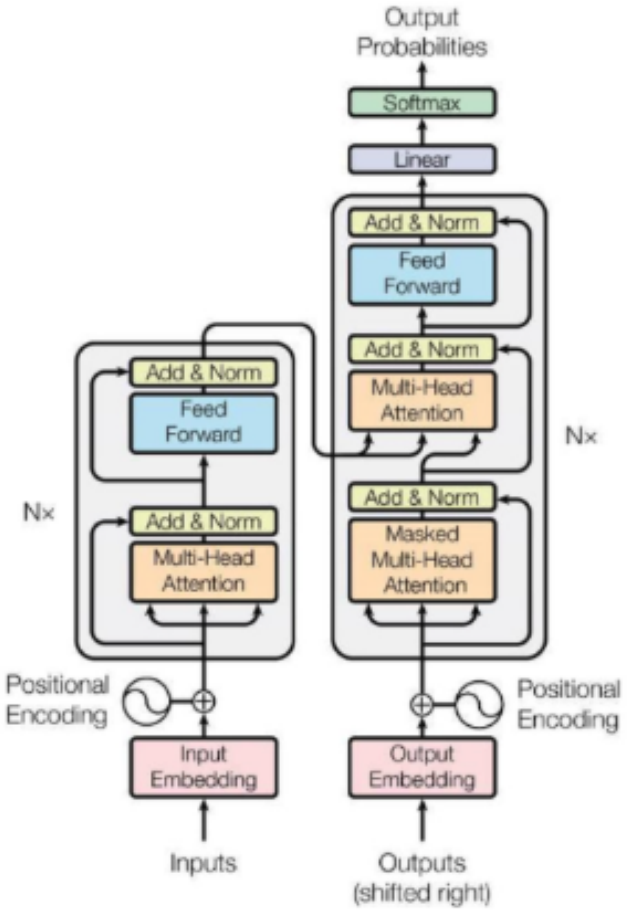
1. **LLM API vs Foundation Model:** Foundation Models are the back-bone of an LLM API such as ChatGPT. LLM or Vision API typically has much more fine-tuning than a Foundation Model
2. **Example:** GPT-2 and GPT-3 are foundation models but GPT3.5 and GPT4 are APIs that include Reinforcement Learning for Human Feedback on top of GPT-3

Course Recap

Models and APIs

1. **LLM API vs Foundation Model:** Foundation Models are the back-bone of an LLM API such as ChatGPT. LLM or Vision API typically has much more fine-tuning than a Foundation Model
2. **Example:** GPT-2 and GPT-3 are foundation models but GPT3.5 and GPT4 are APIs that include Reinforcement Learning for Human Feedback on top of GPT-3
3. **Example:** Stable-Diffusion is a **open-source** Large Vision Model that can generate Images from Text. But **Mid-Journey** and **Dall-e-3** are APIs that take Stable-Diffusion outputs, refine them and produce high-quality images.

Course Recap



Course Recap

Transformers

1. **Game Changer:** Were a game changer when it was introduced in the paper, Attention is all you need in 2017

Course Recap

Transformers

1. **Game Changer:** Were a game changer when it was introduced in the paper, Attention is all you need in 2017
2. **Transformer Architecture:** Consistes of both encoders and decoders

Course Recap

Transformers

1. **Game Changer:** Were a game changer when it was introduced in the paper, Attention is all you need in 2017
2. **Transformer Architecture:** Consistes of both encoders and decoders
3. **Encoder based Foundation Models:** Mainly BERT and variations of BERT - Still used for text classification, sentiment analysis, document summarization, etc

Course Recap

Transformers

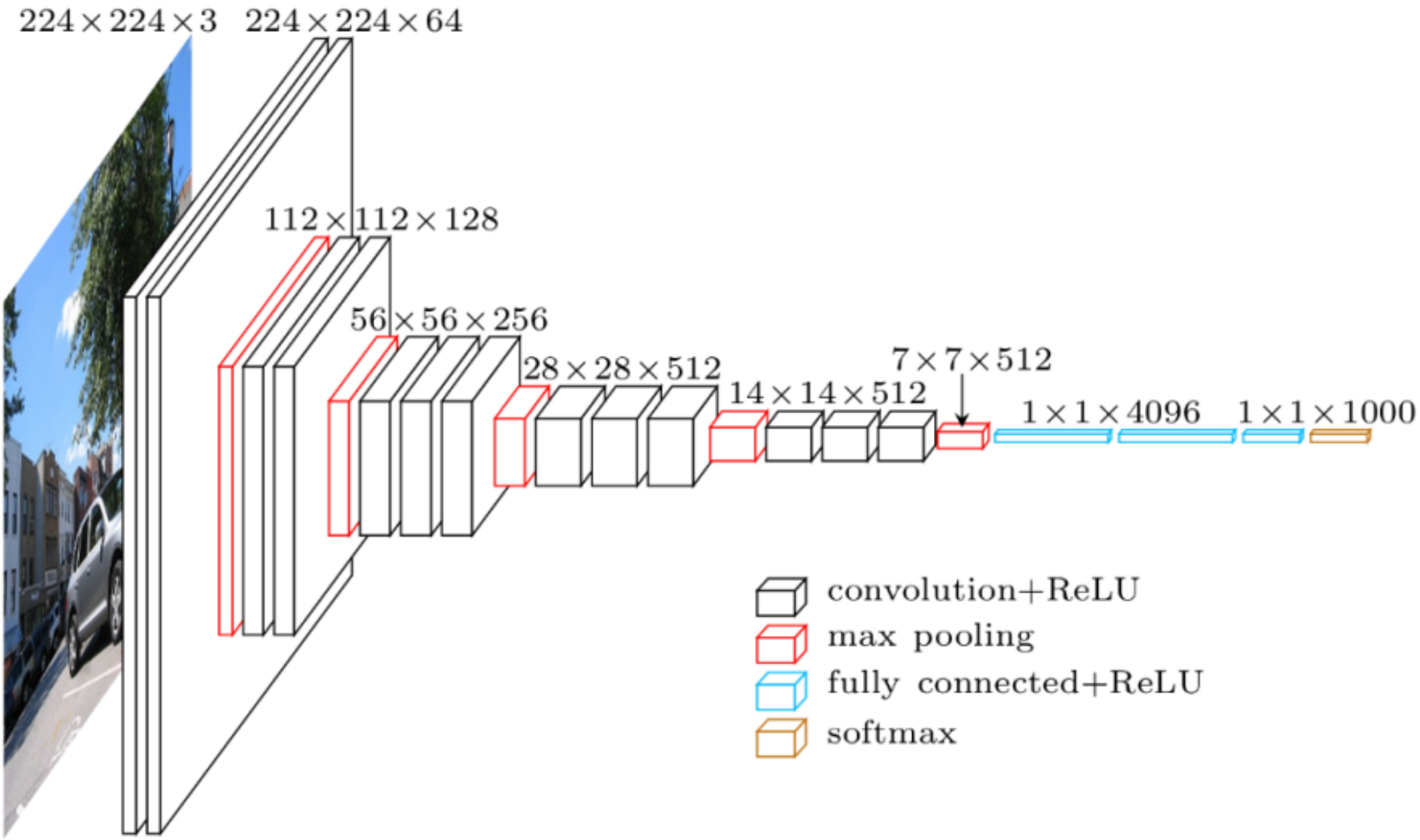
1. **Game Changer:** Were a game changer when it was introduced in the paper, Attention is all you need in 2017
2. **Transformer Architecture:** Consistes of both encoders and decoders
3. **Encoder based Foundation Models:** Mainly BERT and variations of BERT - Still used for text classification, sentiment analysis, document summarization, etc
4. **Decobder based Foundation Models:** Started with GPT (Auto-Regressive models). Then moved to GPT-2, GPT-3 and now GPT-3.5 and GPT-4

Course Recap

Transformers

1. **Game Changer:** Were a game changer when it was introduced in the paper, Attention is all you need in 2017
2. **Transformer Architecture:** Consistes of both encoders and decoders
3. **Encoder based Foundation Models:** Mainly BERT and variations of BERT - Still used for text classification, sentiment analysis, document summarization, etc
4. **Decobder based Foundation Models:** Started with GPT (Auto-Regressive models). Then moved to GPT-2, GPT-3 and now GPT-3.5 and GPT-4
5. **AI revolution:** Was paved by the advent and advance of transformers by many researchers and also OpenAI. Not to mention better compute resources.

Course Recap



Course Recap

Vision Models

1. **CNN:** Convolutional Neural Nets was one of the first popular models for processing images in the context of machine learning and data science.

Course Recap

Vision Models

1. **CNN:** Convolutional Neural Nets was one of the first popular models for processing images in the context of machine learning and data science.
2. **Convolutions:** Have been known to be a fundamental building block for extracting useful and also refined information (e.g. edge detection, etc) about images such as edges for more than a few decades now

Course Recap

Vision Models

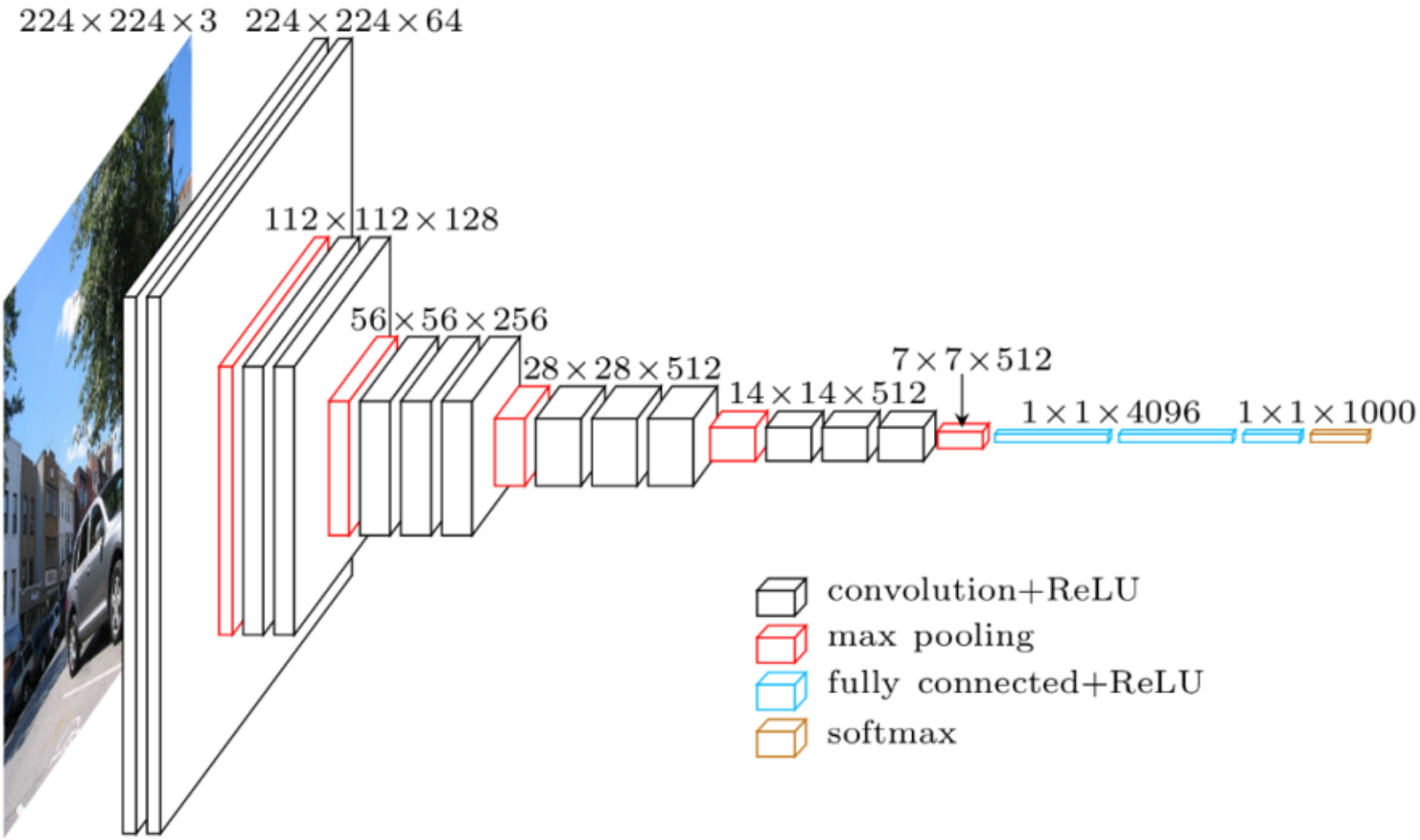
1. **CNN:** Convolutional Neural Nets was one of the first popular models for processing images in the context of machine learning and data science.
2. **Convolutions:** Have been known to be a fundamental building block for extracting useful and also refined information (e.g. edge detection, etc) about images such as edges for more than a few decades now
3. **Large Vision Models:** Typically have a series of convolutions and max-pooling layers to get a refined representation (or embedding) of images.

Course Recap

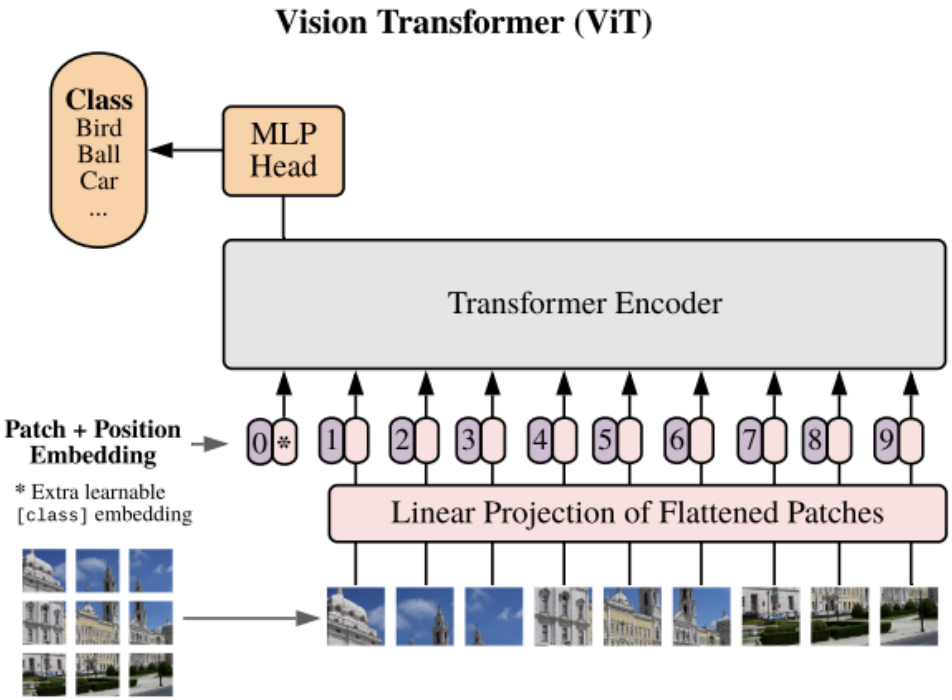
Vision Models

1. **CNN:** Convolutional Neural Nets was one of the first popular models for processing images in the context of machine learning and data science.
2. **Convolutions:** Have been known to be a fundamental building block for extracting useful and also refined information (e.g. edge detection, etc) about images such as edges for more than a few decades now
3. **Large Vision Models:** Typically have a series of convolutions and max-pooling layers to get a refined representation (or embedding) of images.
4. **Visual Transformers:** Are an extension of Transformer Architecture from Language Processing to Vision. The initial conversion of images into tokens for the transformer does involve convolutions

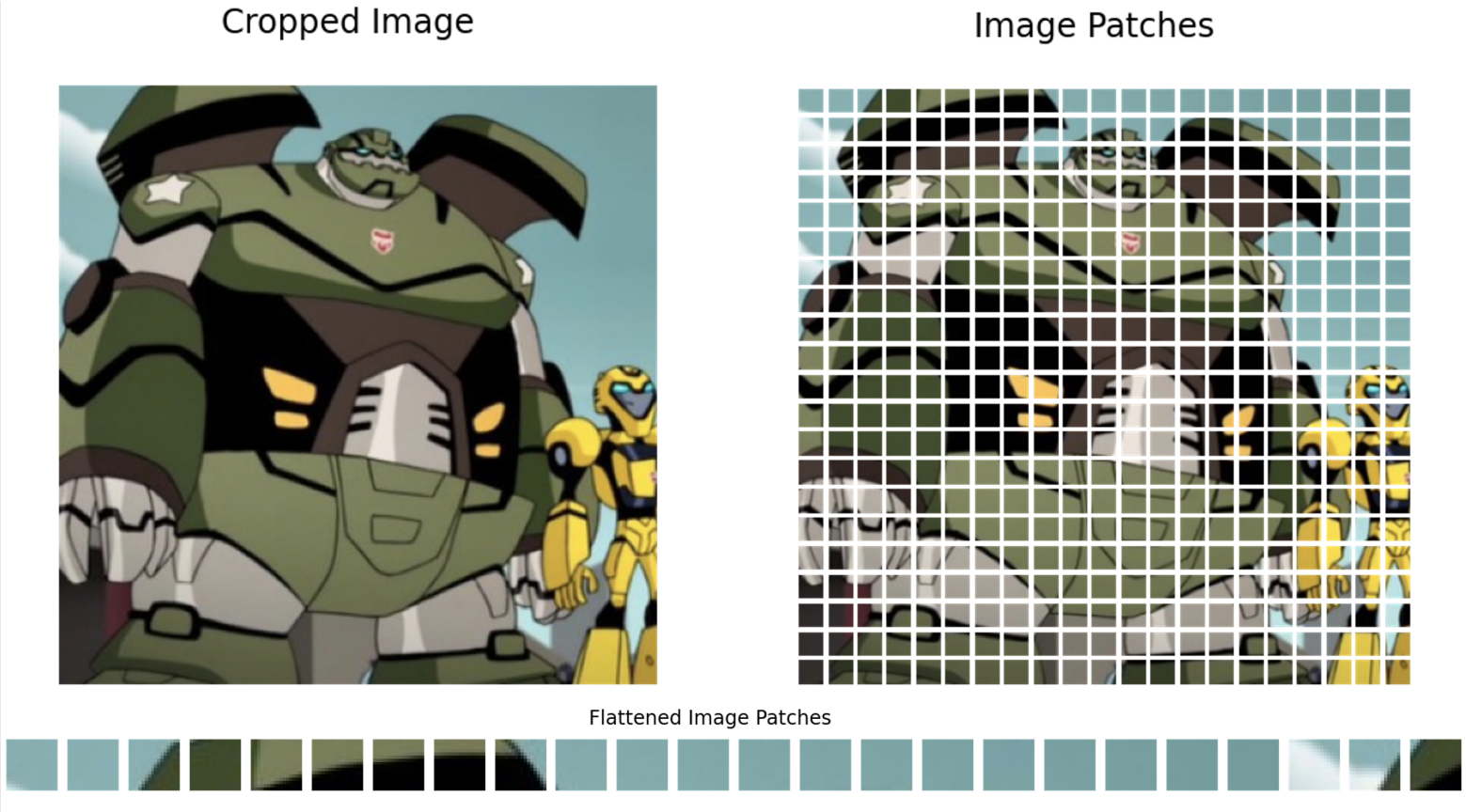
Course Recap



Course Recap



Course Recap



Course Recap

Stable Diffusion

1. **Unique Architecture:** That blends Transformers, Auto-Encoders and Convolutions

Course Recap

Stable Diffusion

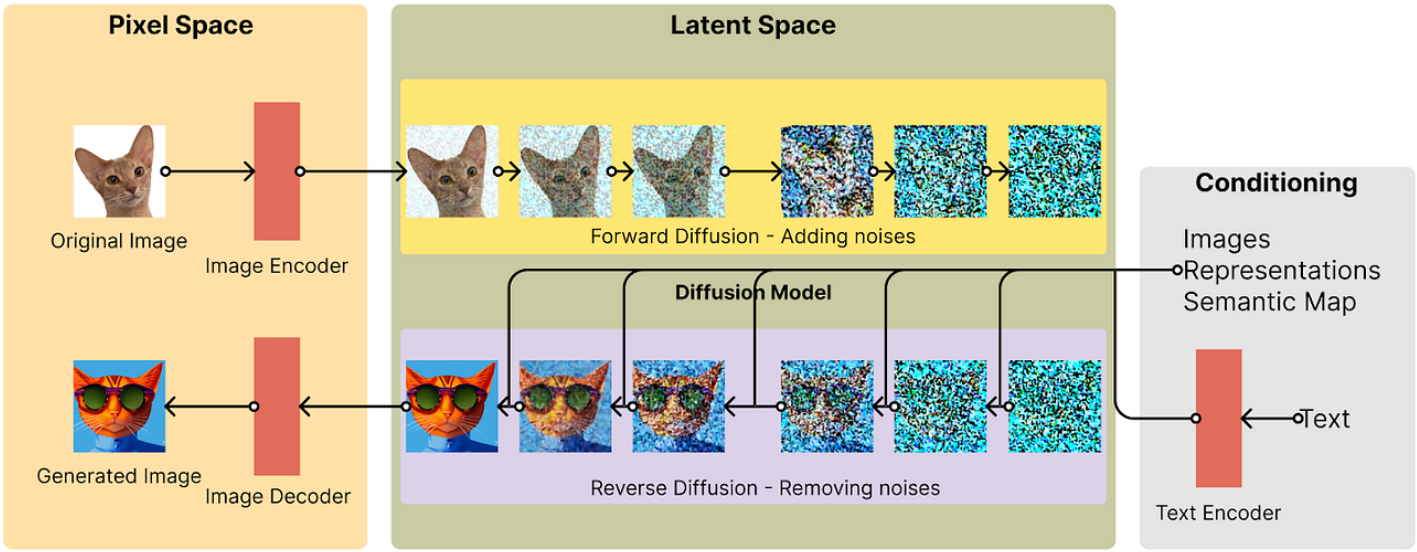
1. **Unique Architecture:** That blends Transformers, Auto-Encoders and Convolutions
2. **Decoding noise:** The inference takes in as input noise + a sentence as input and decodes successively images that progressively have a higher quality to them

Course Recap

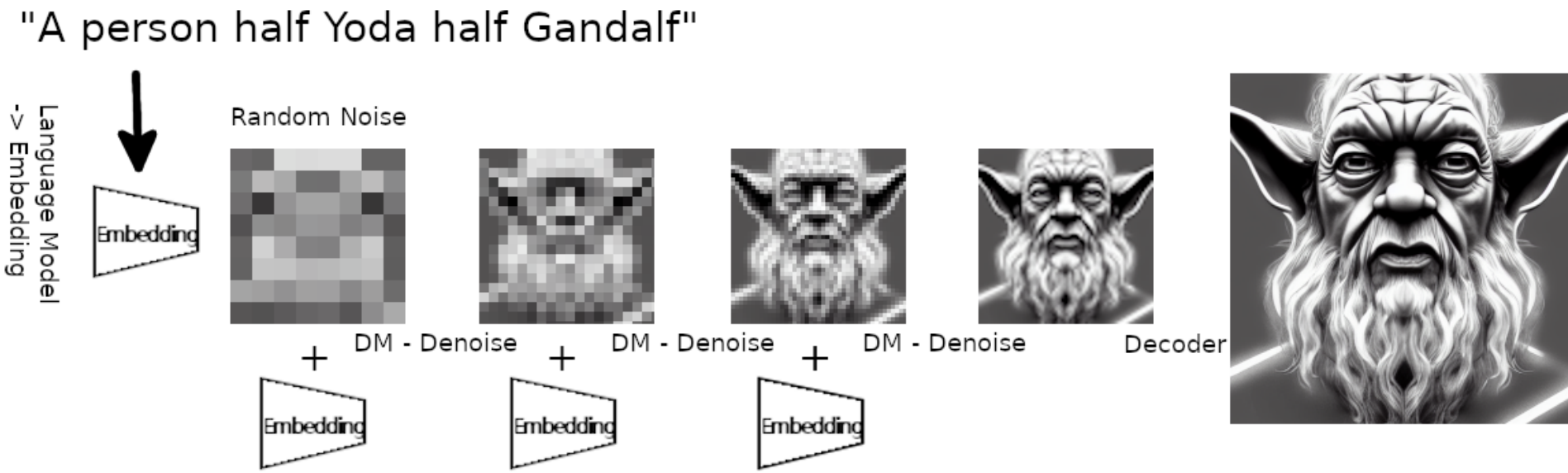
Stable Diffusion

1. **Unique Architecture:** That blends Transformers, Auto-Encoders and Convolutions
2. **Decoding noise:** The inference takes in as input noise + a sentence as input and decodes successively images that progressively have a higher quality to them
3. **Text2Image:** Advances in Stable Diffusion and its refinements with DALL-E-3 and MidJourney have also played a role in making AI very popular

Foundation Model - Stable Diffusion (Text2Image)



Foundation Model - Stable Diffusion (Text2Image)



Reference

Course Recap

Pre-Training vs Fine-Tuning

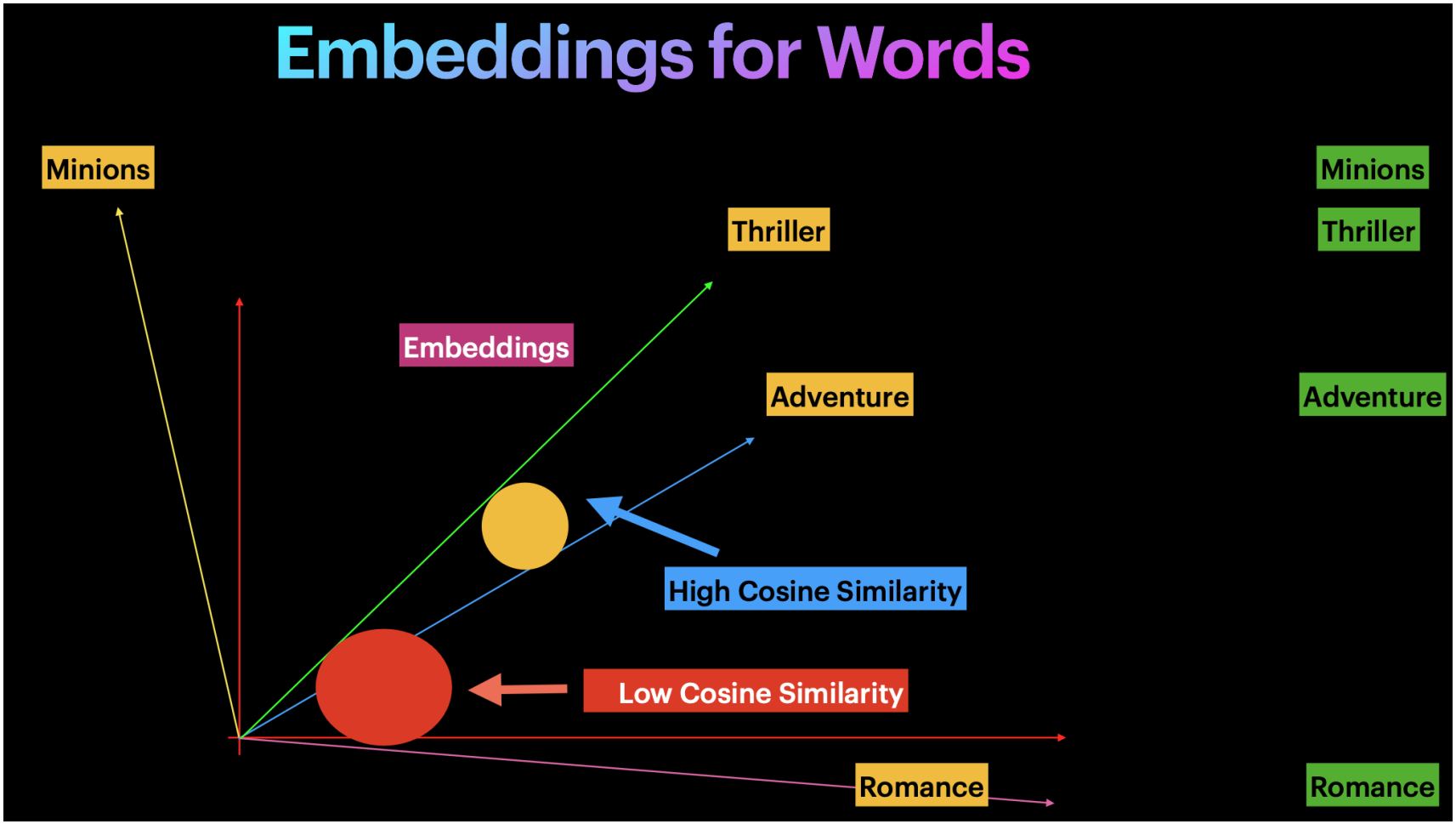
1. **Foundation Models:** For foundation models, sometimes out of the box predictions aren't as good as a fine-tuned model. We saw an example yesterday with "Pokemon" image

Course Recap

Pre-Training vs Fine-Tuning

1. **Foundation Models:** For foundation models, sometimes out of the box predictions aren't as good as a fine-tuned model. We saw an example yesterday with "Pokemon" image
2. **APIs:** LLM and Vision APIs also need better prompting for good responses. There are many prompting tools that can be used to generate better prompts for a desired outcome

Course Recap



Course Recap

Embeddings

1. **Definition:** A concise representation of an object of interest. Object can be a word, sentence, image, video. Usually between 100 to 1000 dimensions

Course Recap

Embeddings

1. **Definition:** A concise representation of an object of interest. Object can be a word, sentence, image, video. Usually between 100 to 1000 dimensions
2. **Text Embeddings:** Glove embeddings are good for understanding a single word but sentence embeddings are better for understanding a whole sentence - Example through Sentence Transformer

Course Recap

Embeddings

1. **Definition:** A concise representation of an object of interest. Object can be a word, sentence, image, video. Usually between 100 to 1000 dimensions
2. **Text Embeddings:** Glove embeddings are good for understanding a single word but sentence embeddings are better for understanding a whole sentence - Example through Sentence Transformer
3. **Image Embeddings:** Can be obtained through CNN or Vi-Transformer Architecture by running it on a input and outputting the last but one layer (i.e. the layer before predictions are made)

Course Recap

Embeddings

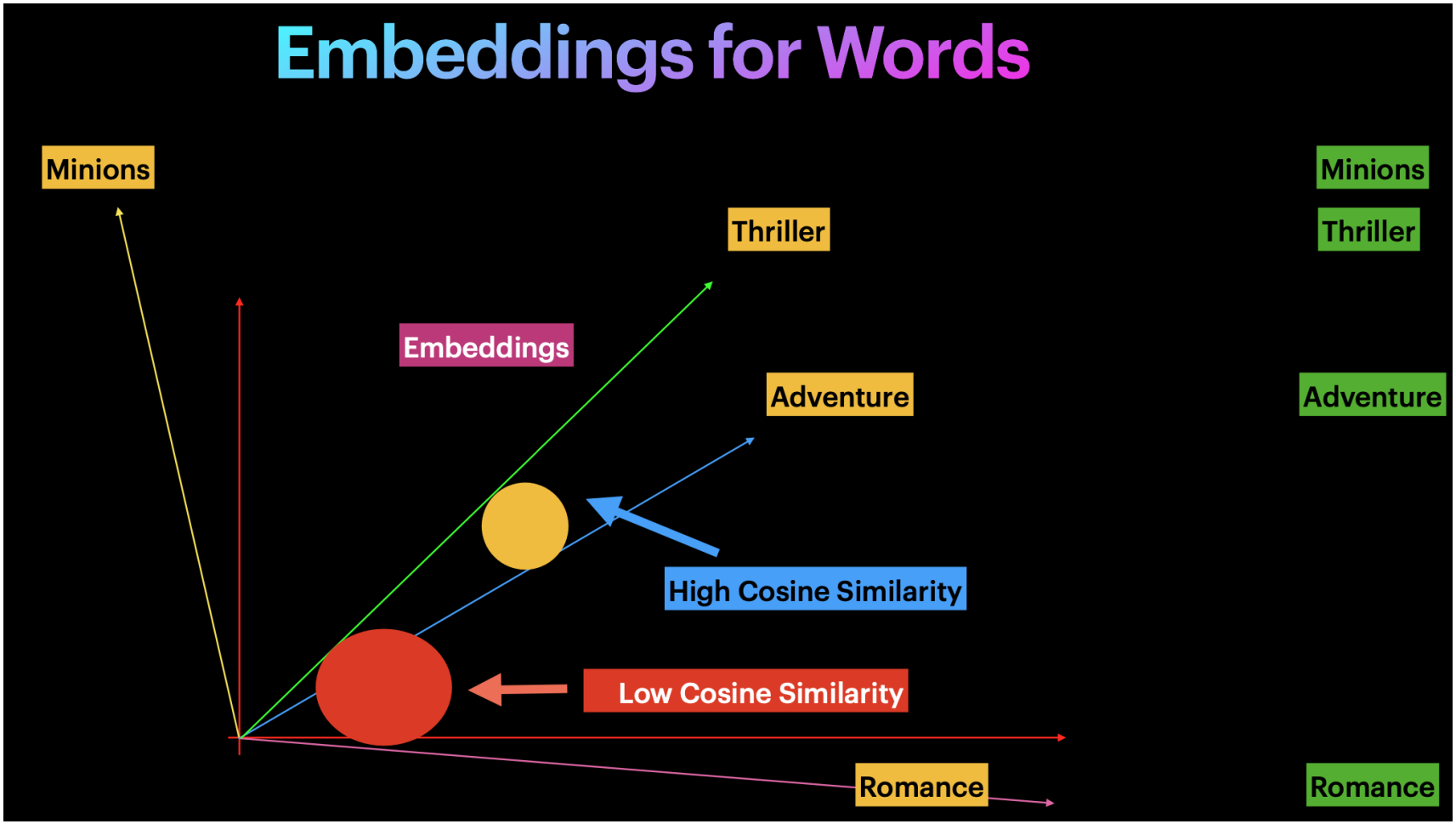
1. **Definition:** A concise representation of an object of interest. Object can be a word, sentence, image, video. Usually between 100 to 1000 dimensions
2. **Text Embeddings:** Glove embeddings are good for understanding a single word but sentence embeddings are better for understanding a whole sentence - Example through Sentence Transformer
3. **Image Embeddings:** Can be obtained through CNN or Vi-Transformer Architecture by running it on a input and outputting the last but one layer (i.e. the layer before predictions are made)
4. **Search:** Embeddings are very useful for search. Example: keyword search, sentence-topic search, image-image search, etc

Course Recap

Embeddings

1. **Definition:** A concise representation of an object of interest. Object can be a word, sentence, image, video. Usually between 100 to 1000 dimensions
2. **Text Embeddings:** Glove embeddings are good for understanding a single word but sentence embeddings are better for understanding a whole sentence - Example through Sentence Transformer
3. **Image Embeddings:** Can be obtained through CNN or Vi-Transformer Architecture by running it on a input and outputting the last but one layer (i.e. the layer before predictions are made)
4. **Search:** Embeddings are very useful for search. Example: keyword search, sentence-topic search, image-image search, etc
5. **Clustering:** Embeddings can also be used to group or categorize objects that are similar. Think of product categories at companies

Course Recap



Course Recap

Applications we looked at

Keyword Search, Image Search, Image Captioning, Text2Image, Paraphrasing Sentences, Image Segmentation

Tools we covered

Google Colab (GPU computations), Streamlit (easy web app development), OpenAI APIs (for text and image), Hugging Face Libraries (for image captioning, stablediffusion, etc)