LLMs and ChatGPT || Text2Image, Image2Text, Segmentation and more!!

Dr. Karthik Mohan

Univ. of Washington, Seattle

February 13, 2024

Today

Focus

We will focus a lot on the interactions between text and images in the lecture today and the associated Foundation models and APIs You will also get to play around with these in the In-class coding exercise today!

Today

1. **Image2Text** API for generation Text from Image

Today

Image Captioning

1. **Image2Text** API for generation Text from Image
2. **Text2Image** API for generating an Image from Text

Today

1. **Image2Text** API for generation Text from Image
2. **Text2Image** API for generating an Image from Text
3. **Stable Diffusion** Pre-Trained Model for **Text2Image**

Today

1. **Image2Text** API for generation Text from Image
2. **Text2Image** API for generating an Image from Text
3. **Stable Diffusion** Pre-Trained Model for **Text2Image**
4. **Segmenting Images** and tagging them

Today

1. **Image2Text** API for generation Text from Image
2. **Text2Image** API for generating an Image from Text
3. **Stable Diffusion** Pre-Trained Model for **Text2Image**
4. **Segmenting Images** and tagging them
5. Image Embeddings and Image-Image search

Today

1. **Image2Text** API for generation Text from Image
2. **Text2Image** API for generating an Image from Text
3. **Stable Diffusion** Pre-Trained Model for **Text2Image**
4. **Segmenting Images** and tagging them
5. Image Embeddings and Image-Image search
6. Foundation Models for Images - **CNNs and ViTransformers**

Foundation Models for Images

Types

CNNs (e.g. Inception, AlexNet, etc) and Visual Transformers or ViTransformer

Foundation Models for Images

Types

CNNs (e.g. Inception, AlexNet, etc) and Visual Transformers or ViTransformer

Building Blocks

Like legos can be used to build a whole factory - Foundation models can be put together across modes (multi-modal) to create interesting and beautiful applications. Text2Image is one such example that combines multiple foundation models - Transformers, ViTransformers, CNNs and also AutoEncoders.

Foundation Models for Images

Types

CNNs (e.g. Inception, AlexNet, etc) and Visual Transformers or ViTransformer

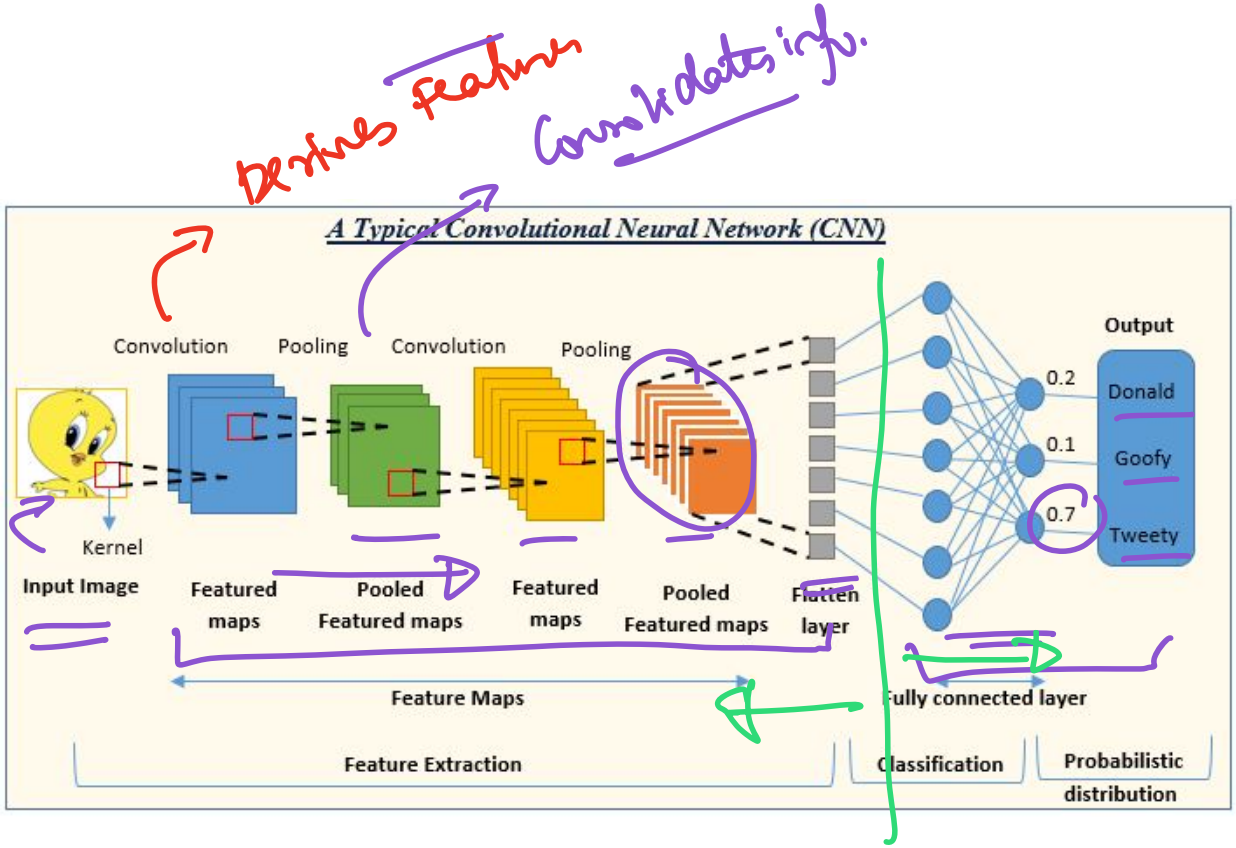
Building Blocks

Like legos can be used to build a whole factory - Foundation models can be put together across modes (multi-modal) to create interesting and beautiful applications. Text2Image is one such example that combines multiple foundation models - Transformers, ViTransformers, CNNs and also AutoEncoders.

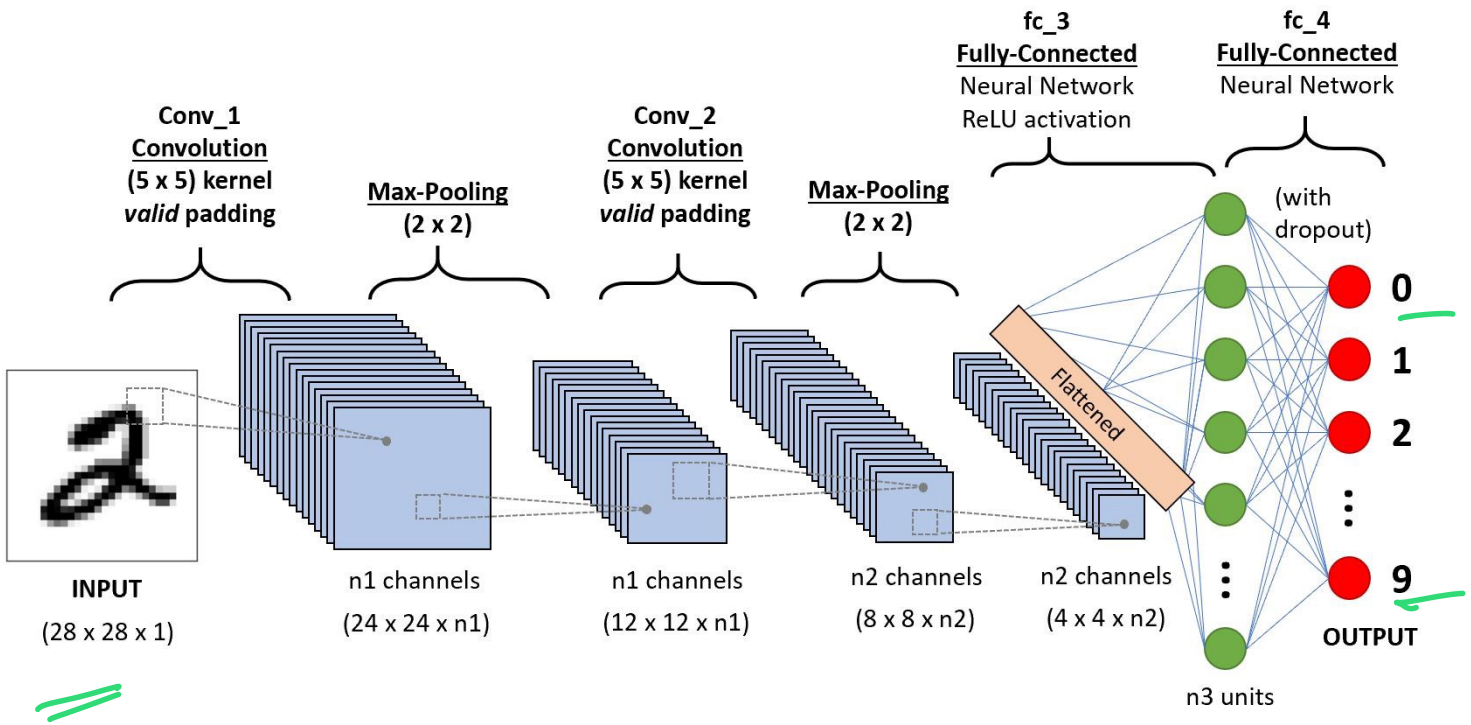
Applications

Classification (cat or dog?), Image2Text, Text2Image, Image Embeddings, Object Detection, Image Segmentation, etc

Foundation Model - CNN (Convolutional Neural Network)

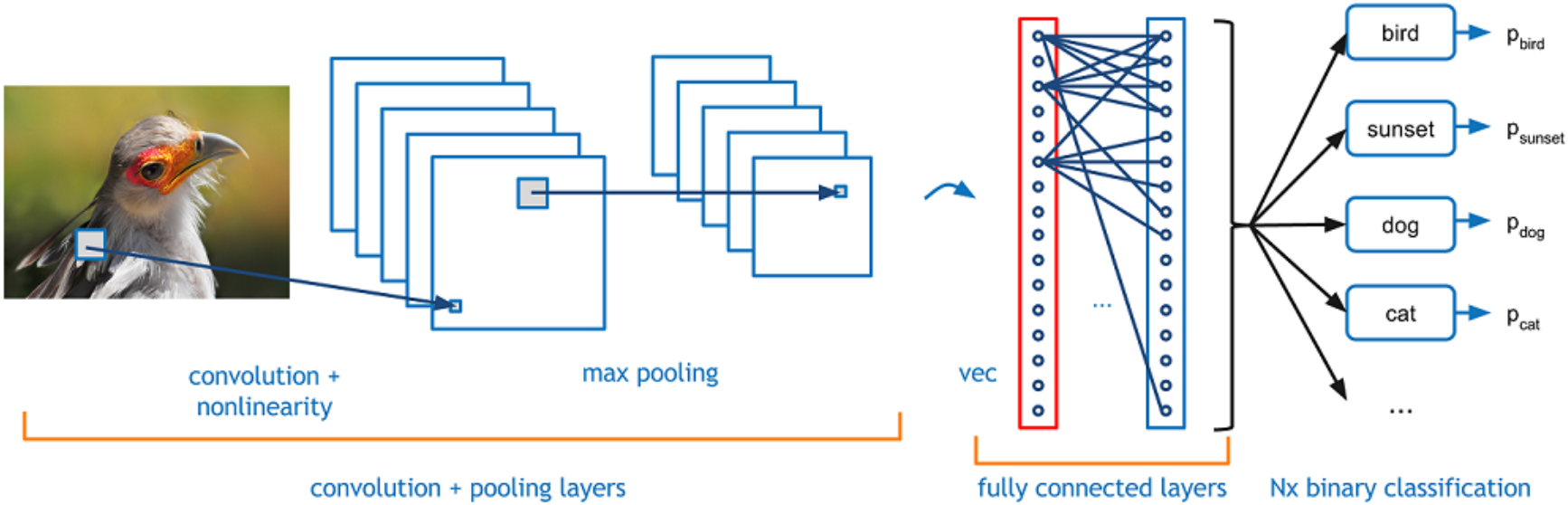


Foundation Model - CNN (Convolutional Neural Network)

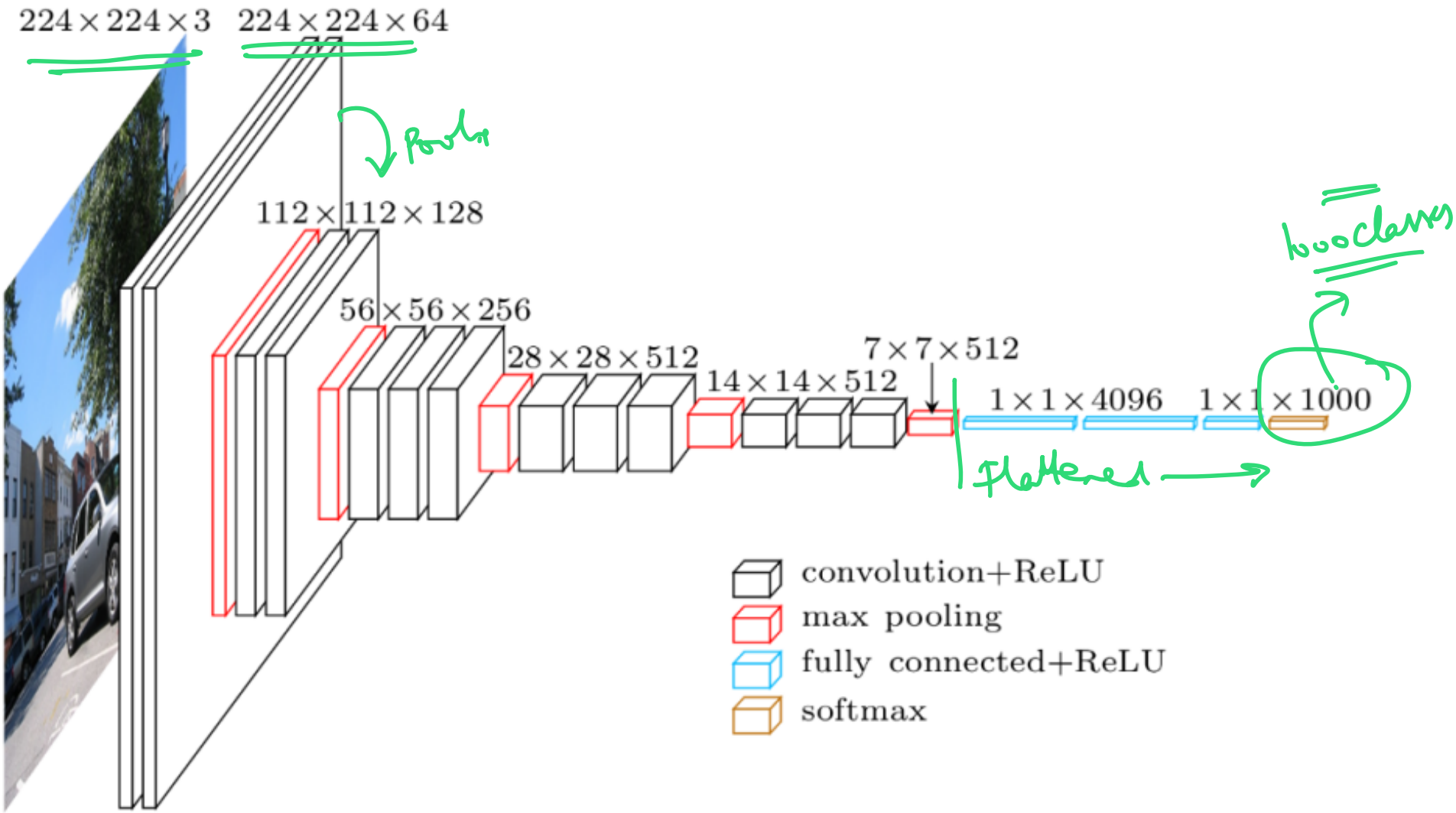


(MNIST digits classification)

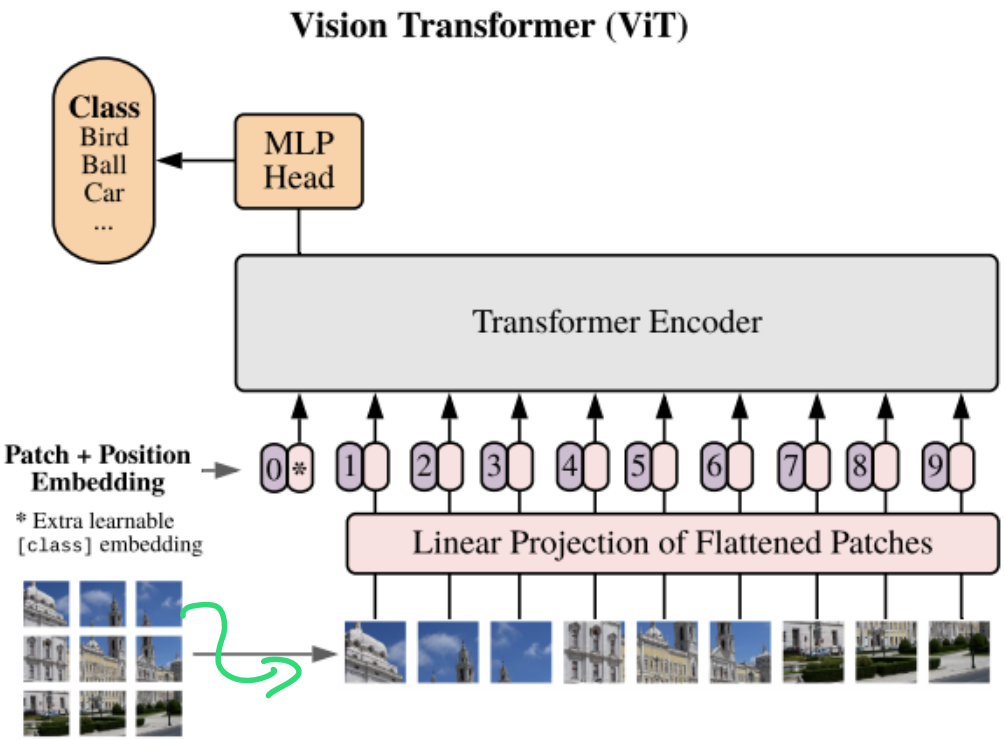
Foundation Model - CNN (Convolutional Neural Network)



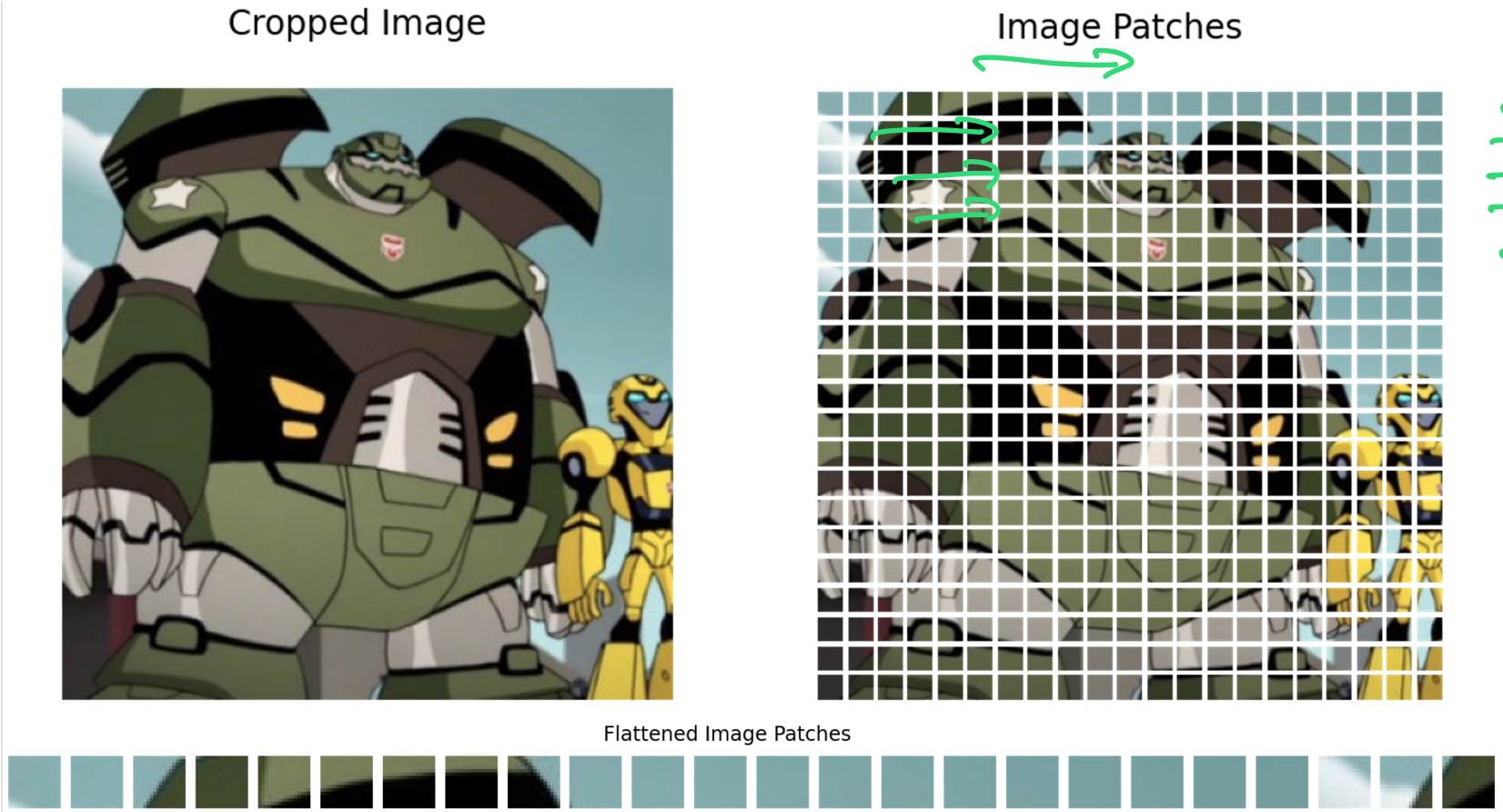
Foundation Model - CNN (Convolutional Neural Network)



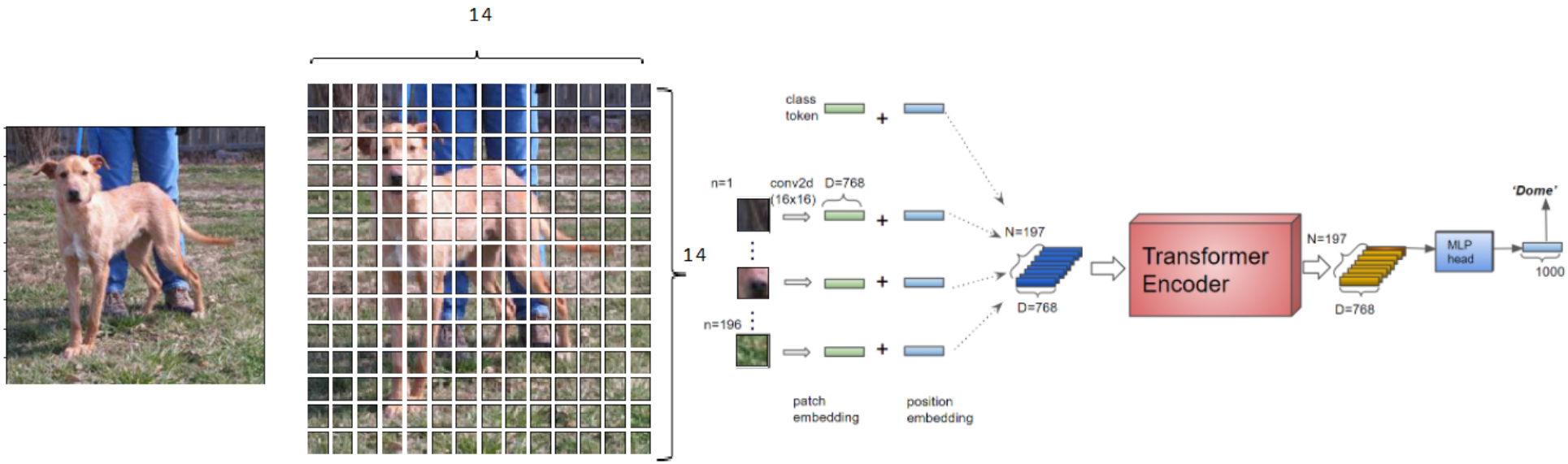
Foundation Model - Visual Transformers (ViT)



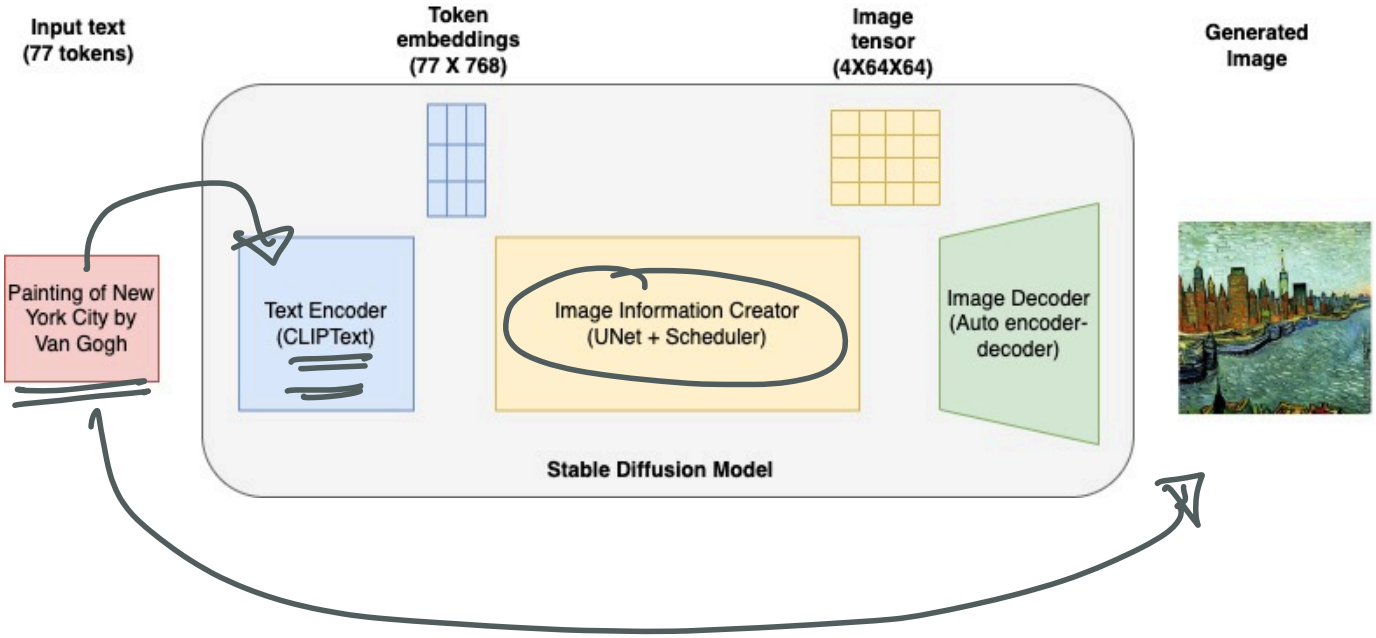
Foundation Model - Visual Transformers (ViT)



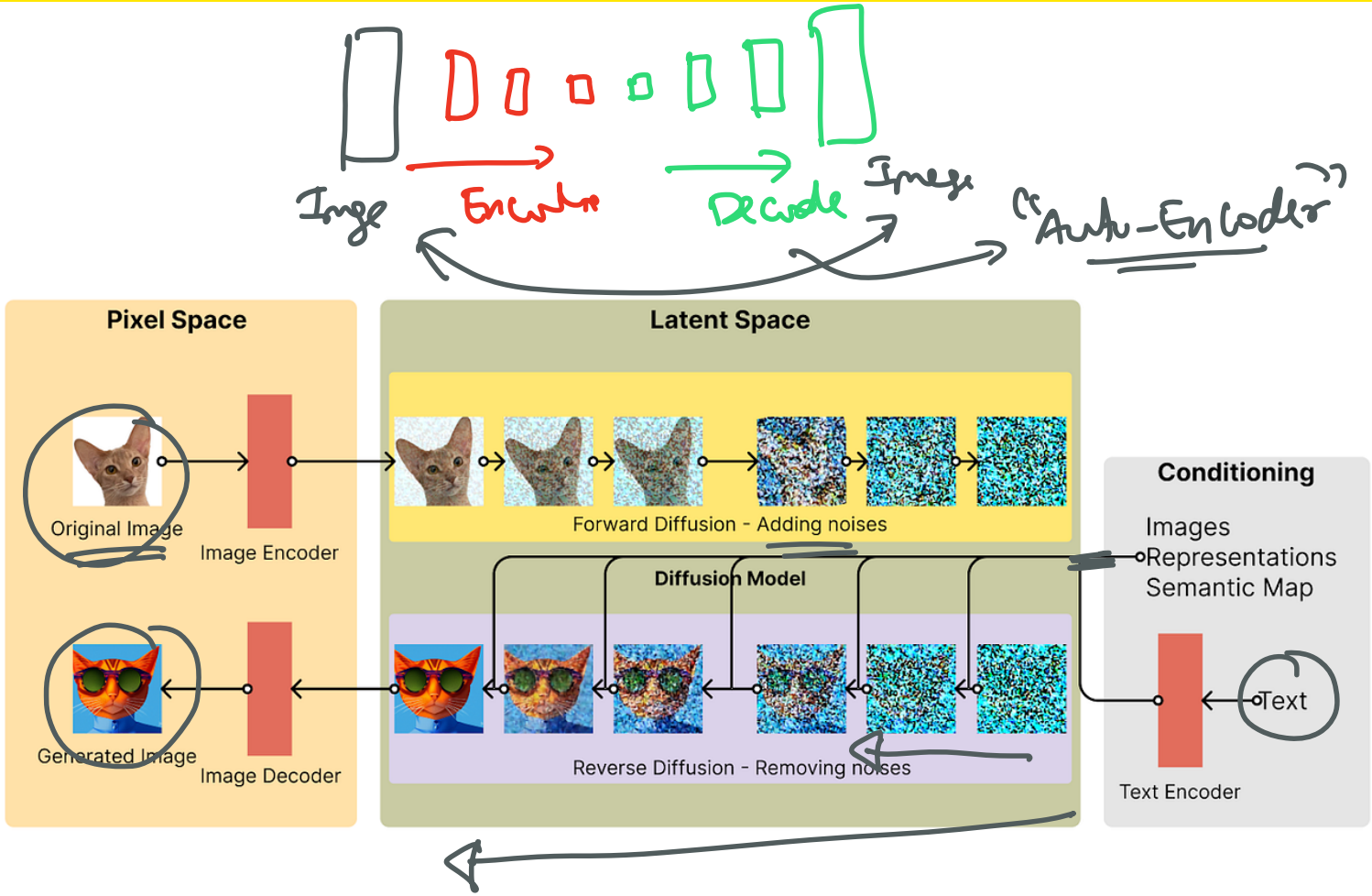
Foundation Model - Visual Transformers (ViT)



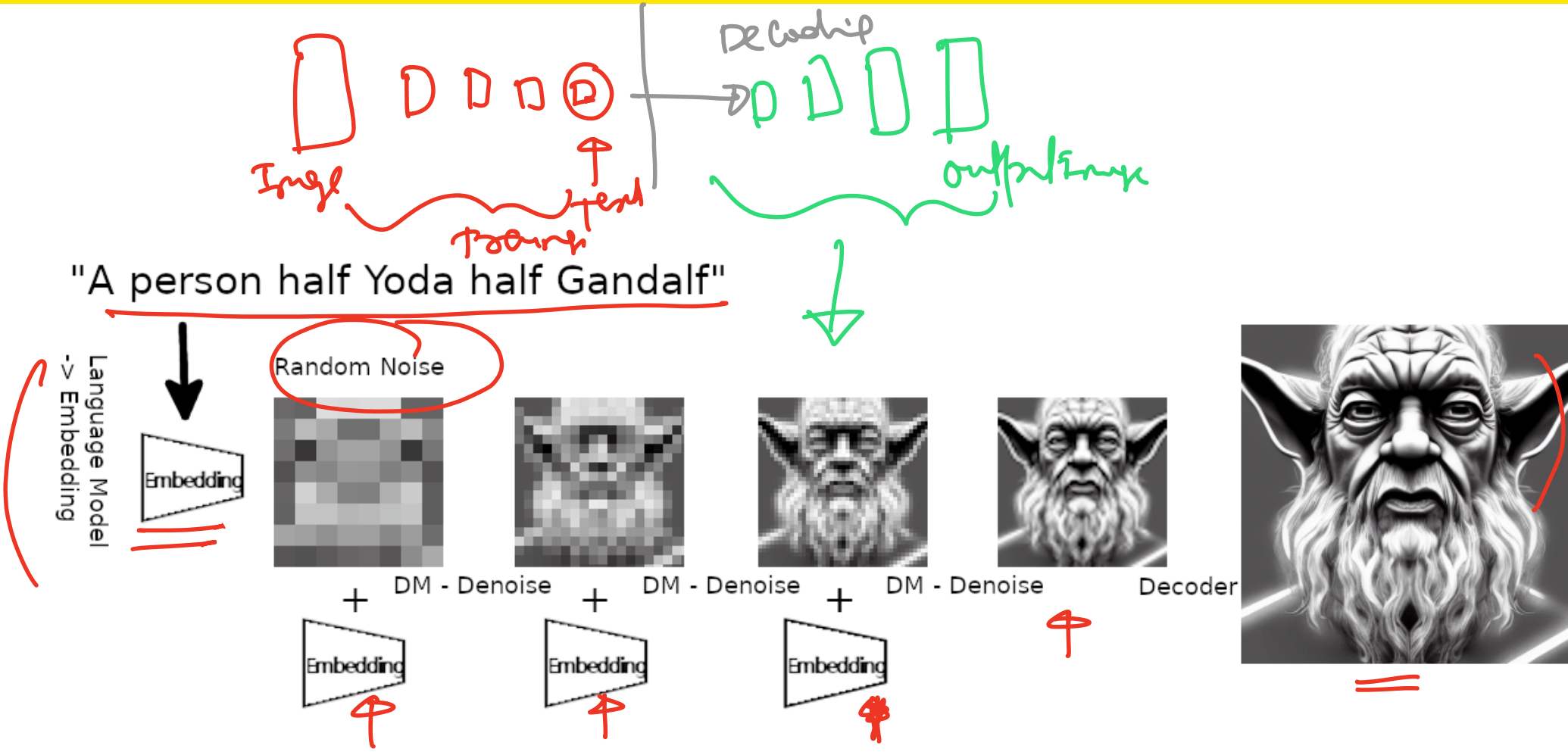
Foundation Model - Stable Diffusion (Text2Image)



Foundation Model - Stable Diffusion (Text2Image)



Foundation Model - Stable Diffusion (Text2Image)



Reference