

DEEP LEARNING

FOR

- ASSIGNMENT 03

COMPUTER VISION

Write about Transformer models.

Aneetta Sara Shany

MTech CSE

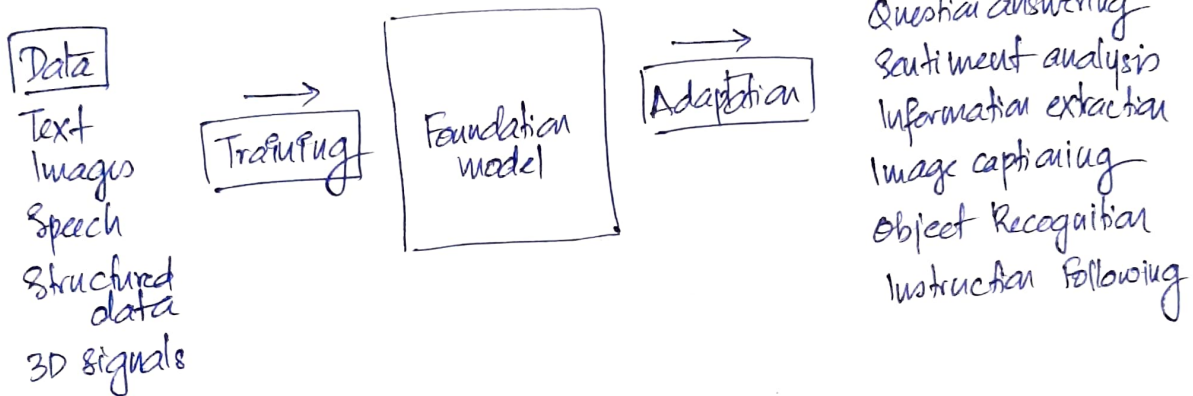
Software Engineering

April 2, 2023.

Transformer models

Transformers are a class of deep learning models that are defined by some architectural traits. Transformer model is a neural network that learns context and thus, meaning by tracking relationships in sequential data like the words in this sentence.

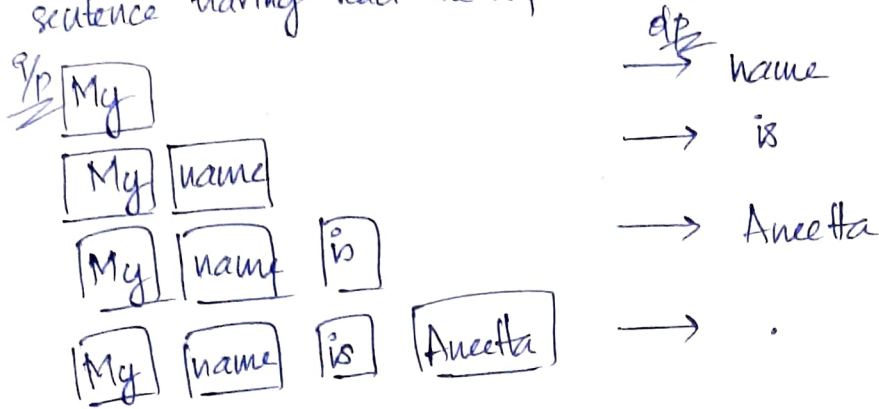
Transformers are translating text and speech in near real-time, opening meetings & classrooms to diverse and hearing-impaired attendees. Stanford researchers called transformers 'foundation models' as they drive a paradigm shift in AI.



Like most neural networks, transformer models are basically large encoder/decoder blocks that process data. Transformers use positional encoders to tag data elements coming in and out of network. Attention units follow these tags, calculating a kind of algebraic map of how each element relates to the others.

Transformer model which is a deep learning architecture originally proposed for natural language processing tasks, but has since been extended to other domains such as computer vision. The Vision Transformer (ViT) is a variant of the Transformer architecture designed specifically for image recognition tasks.

Language modelling example of task predicting the next word in a sentence having read the n previous words.

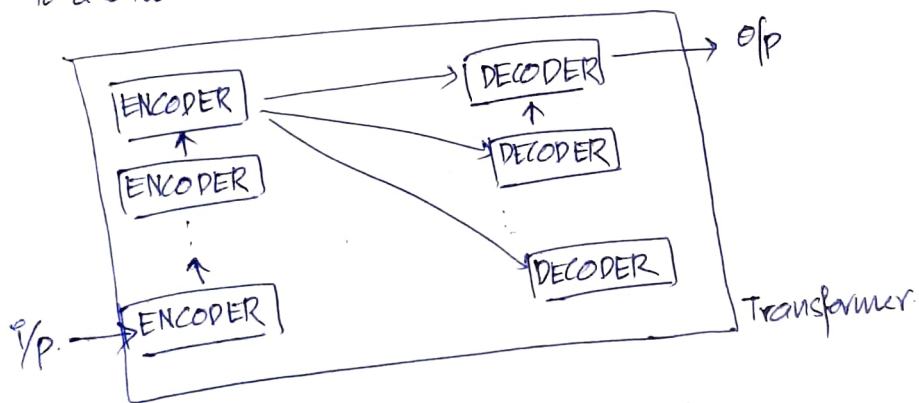


Most common transformer models created recently are:

BERT, DistilBERT, T5, GPT-2.

We can install transformers package by \Rightarrow `pip install transformers`.

Transformers are considered a combination of both CNNs with attention. Attention boosts of how fast the model can translate from one sequence to another.



Feed Forward Neural Network
self attention
Encoder

Feed Forward
Encoder-Decoder Attention
self attention
Decoder.

\Rightarrow Transformers models provides advantages over other neural network architectures with their encoder-decoder models.