

Transformers

Session 1 - Introduction

Sylvain Le Corff - Charles Ollion



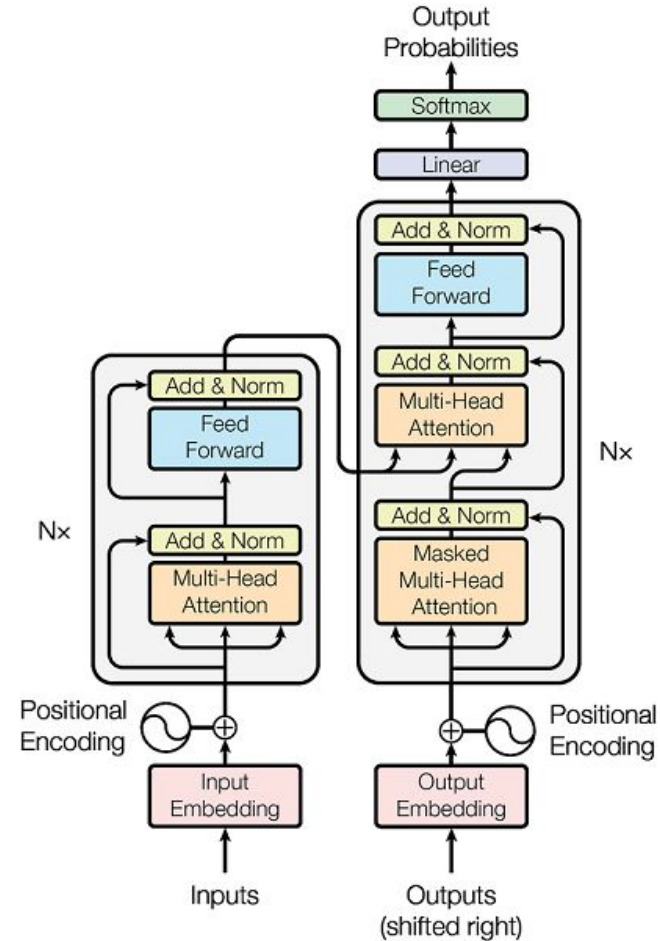
A Deep Learning Building Block

Similar to Recurrent Neural Network (RNN) or Convolutional Neural Network (CNN)

Particularly adapted to process **sets** or **sequences**: in Natural Language Processing (NLP), the input is a sequence of words

Very parallelizable architecture: good for large scale Deep Learning

[Attention is all you need](#)



Where are Transformers today : NLP

The screenshot shows the Hugging Face website interface. At the top left is the Hugging Face logo and name. A search bar is located at the top center with the placeholder text "Search models, datasets, users...". On the top right, there are links for "Models" and "Da".

On the left side, there is a sidebar with a "Back to tag list" button. Below it, the "Tasks" section includes a "Search tags" input. Under "Natural Language Processing", there are several task tags: "Fill-Mask", "Question Answering", "Summarization", "Table Question Answering", "Text Classification", "Text Generation", "Text2Text Generation", "Token Classification", "Translation", "Zero-Shot Classification", "Sentence Similarity", "Conversational", and "Feature Extraction".

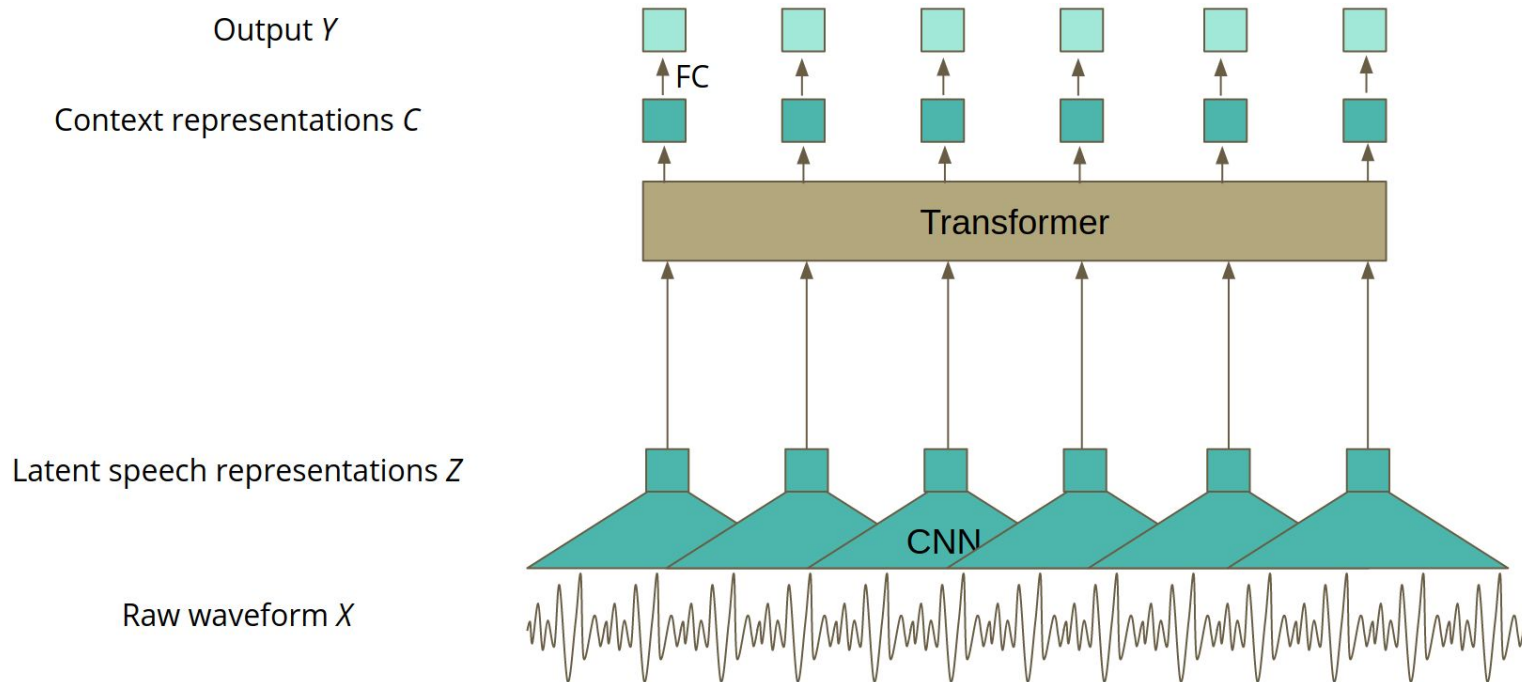
The main content area displays a list of models under the heading "Models 17,937". A "Search Models" input is present. The first three models shown are:

- bert-base-uncased**: Fill-Mask • Updated May 18 • 26.7M • 42
- roberta-large**: Fill-Mask • Updated May 21 • 5.34M • 15
- gpt2**: Text Generation • Updated May 19 • 4.57M • 15

The fourth model shown is **distilbert-base-uncased-finetuned-sst-2-english**: Text Classification • Updated Feb 9 • 2.97M • 11.

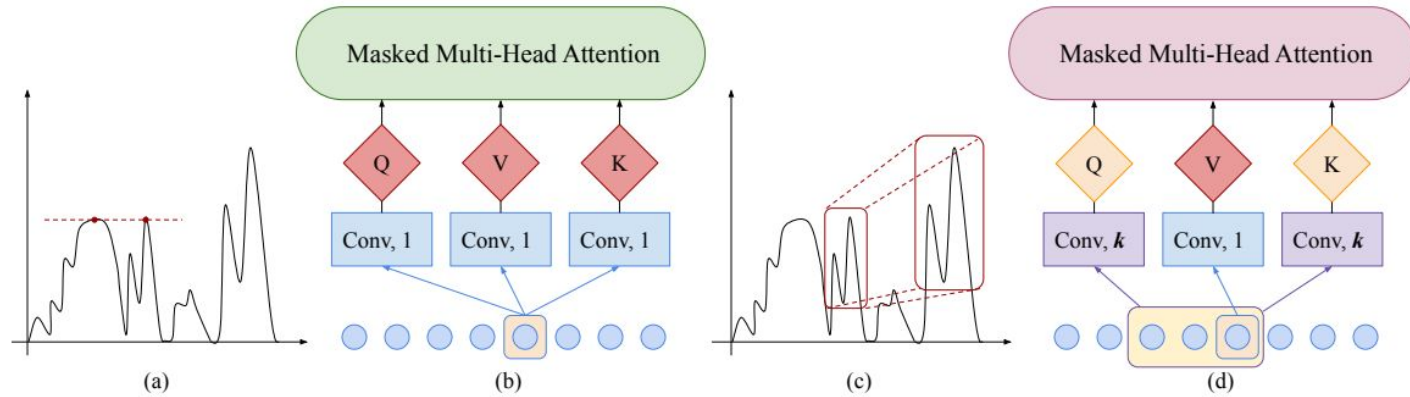
<https://huggingface.co/models>

Audio analysis



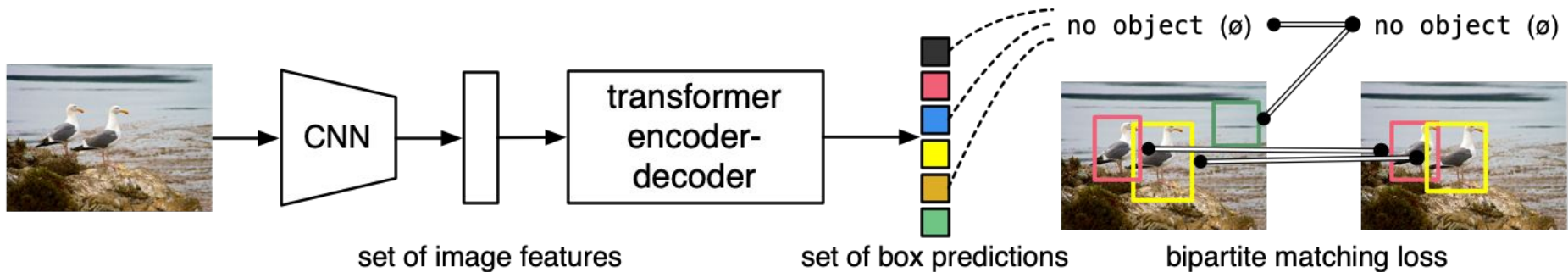
[wav2vec2.0 learning the structure of speech from raw audio](#)

Time-series forecasting



[Enhancing the Locality and Breaking the Memory Bottleneck of Transformer on Time Series Forecasting](#)

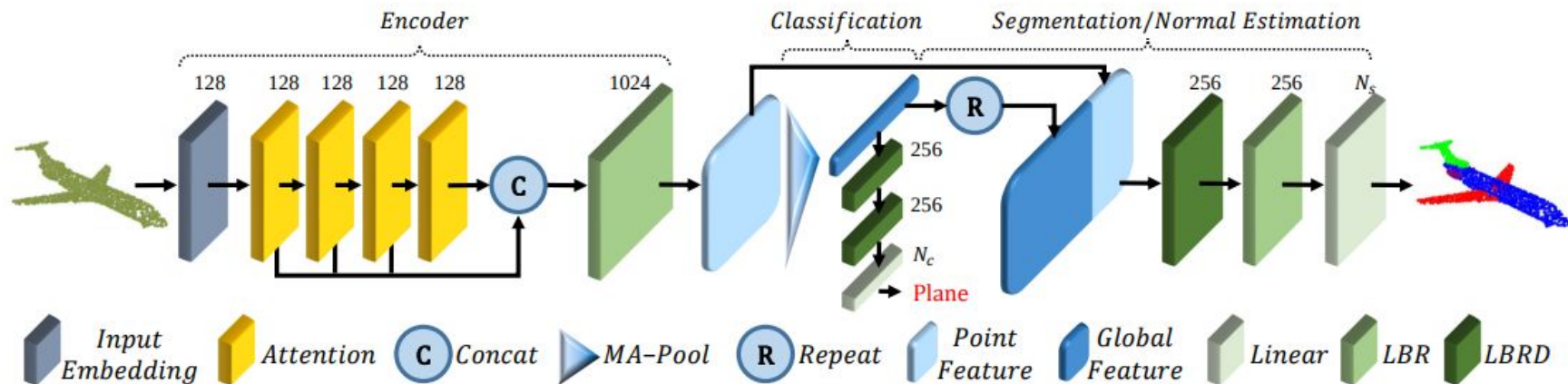
Computer Vision



[End to End Object Detection with Transformers](#)

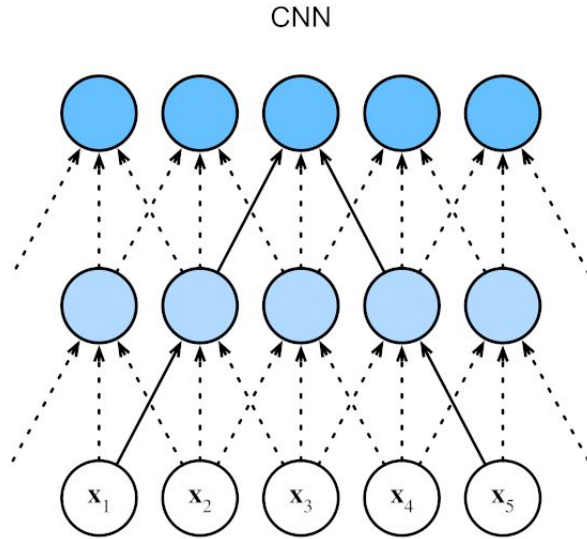
Any set-structured problem

Input is a set of vectors, with an explicit structure (sequence, image) or implicit (list of objects, graph, point cloud)

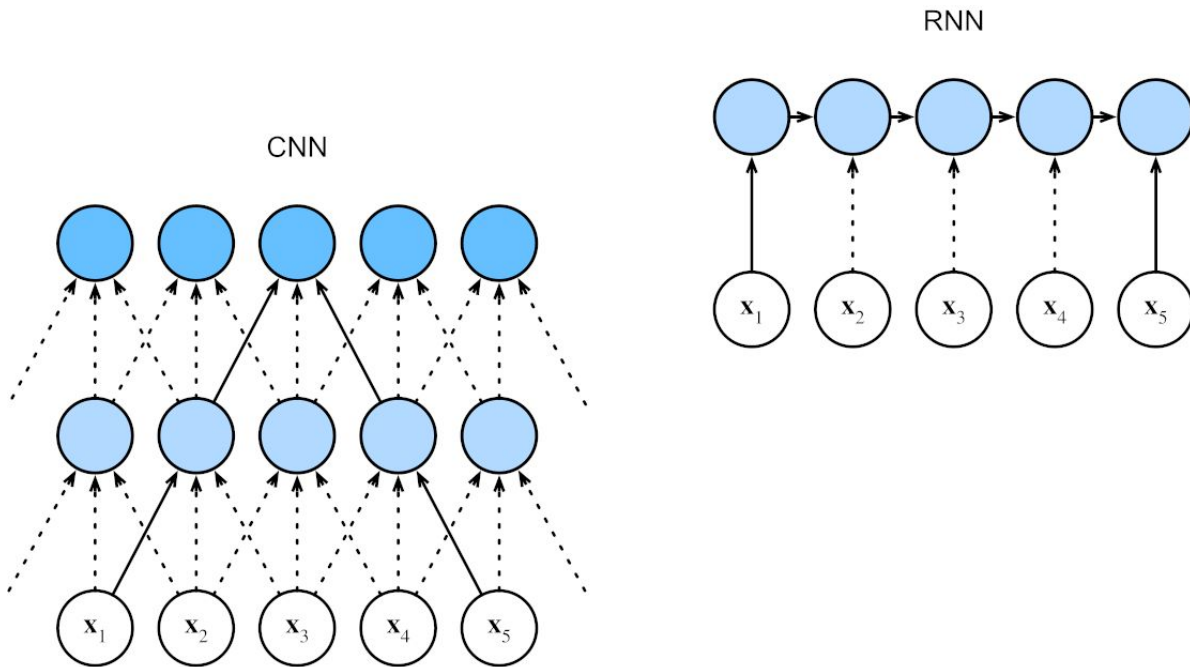


[PCT: Point cloud transformer](#)

Main concept: Self-Attention



Main concept: Self-Attention



Main concept: Self-Attention

