Fusion RAG with Hybrid Search 🔍

Ask your query about the documents where you have provided the data

Enter your Google API Key

••

Enter PDF Folder Path

C:\Z0052PUN - Pranav P R\learnings\resources\personal_proj\fusion_rag_chat\data

Explain what is convolutional neural network?

Based on the provided context, a Convolutional Neural Network (ConvNet) is a type of deep, feedforward network that is easier to train and generalizes better than networks with full connectivity between adjacent layers.

Here are the key characteristics of ConvNets as described in the documents:

- **Purpose:** They are designed to process data that comes in the form of multiple arrays, such as 1D for signals and language, 2D for images or audio spectrograms, or a color image composed of three 2D arrays.
- **Architecture:** They consist of stacked layers, typically two or three stages of convolution, non-linearity, and pooling, followed by more convolutional and fully-connected layers.
- **Hierarchical Feature Detection:** ConvNets exploit the property that many natural signals are compositional hierarchies, where higher-level features (like objects) are formed by composing lower-level ones (like edges and motifs).
- **Pooling Layers:** The role of the pooling layer is to merge semantically similar features into one. It does this by coarse-graining the position of features, for example, by computing the maximum of a local patch of units in a feature map. This helps create invariance to small shifts and distortions.
- **Inspiration:** The convolutional and pooling layers are directly inspired by the classic notions of simple and complex cells in visual neuroscience.

Explain attention mechanisms and transformer

Based on the context provided:

Attention Mechanisms: An attention mechanism is used in sequence modeling to model dependencies between different parts of input or output sequences, regardless of their distance from each other. It allows every position in a sequence to relate to all other positions to compute a representation.

- **Self-attention (or intra-attention):** This is a specific type of attention mechanism that relates different positions within a single sequence to compute a representation of that sequence.
- **Encoder-decoder attention:** In this setup, the queries come from a decoder layer, while the keys and values come from the output of an encoder. This allows every position in the decoder to attend to all positions in the input sequence.

There are different ways to compute attention, such as additive attention and dot-product attention, with the latter being faster and more space-efficient.

Transformer:

The Transformer is a model architecture that, unlike previous models, does not use recurrent or convolutional networks. Instead, it relies entirely on an attention mechanism to establish global dependencies between its input and output. It was the first transduction model to rely completely on self-attention to compute representations of its input and output. This design allows for significantly more parallelization.

The Transformer uses a specific implementation called **Multi-Head Attention**. Instead of a single attention function, it linearly projects the queries, keys, and values multiple times (`h=8` in this work) and runs the attention function on each of these projections in parallel. The outputs are then concatenated. This helps to counteract the reduced effective resolution that can result from averaging attention-weighted positions.

CNN?

Based on the context provided, a CNN, or convolution neural network, is a type of deep neural

network. Its applications include:* **Image Understanding:** Since the early 2000s, CNNs (also referred to as ConvNets) have been successfully used for the detection, segmentation, and recognition of objects and regions in images, such as in traffic sign recognition, segmentation of biological images, and face detection.

- **Image-to-Text Generation:** A deep CNN can extract a representation from an image, which is then used as input for a recurrent neural network (RNN) to generate a caption for that image.
- **Historical Applications:**
  - **Document Reading:** In the late 1990s, a system using a ConvNet was reading over 10% of all cheques in the United States.
  - **Character and Handwriting Recognition:** ConvNet-based systems were deployed by Microsoft for optical character recognition and handwriting recognition.
  - **Object and Face Recognition:** They were experimented with in the early 1990s for detecting objects like faces and hands, as well as for face recognition.
  - **Speech Recognition:** A primitive 1D ConvNet, known as a time-delay neural net, was used for recognizing phonemes and simple words.The architecture of a ConvNet is described as being "reminiscent of the LGN–V1–V2–V4–IT hierarchy in the visual cortex ventral pathway."

What is your question?