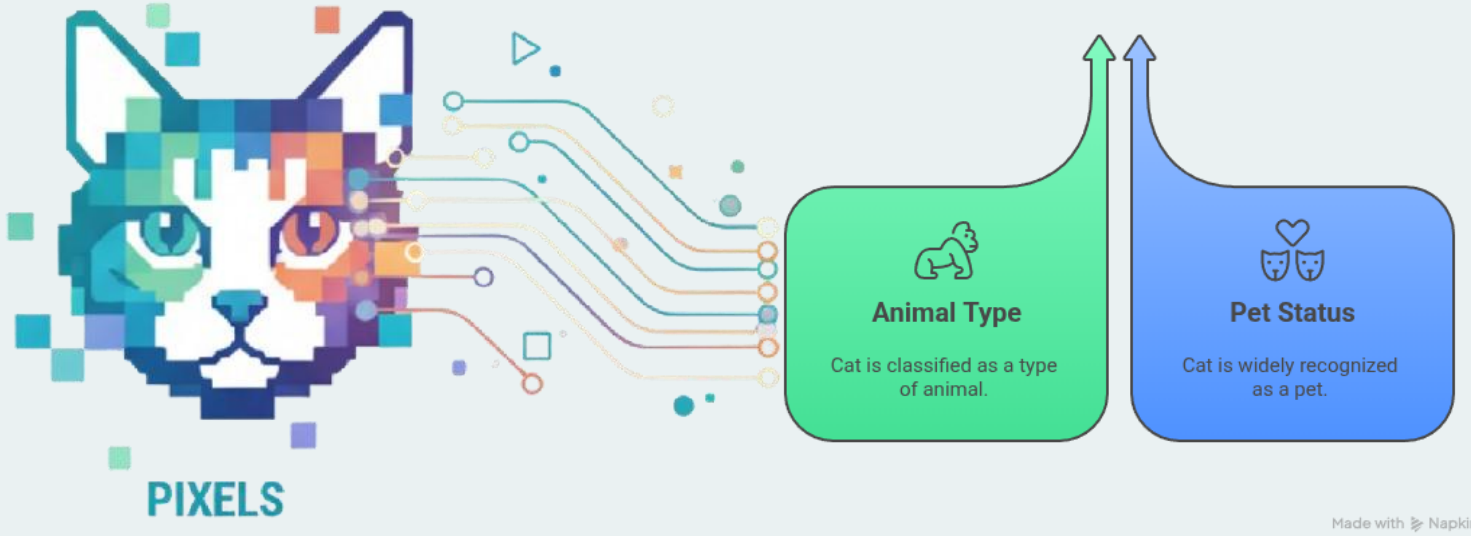
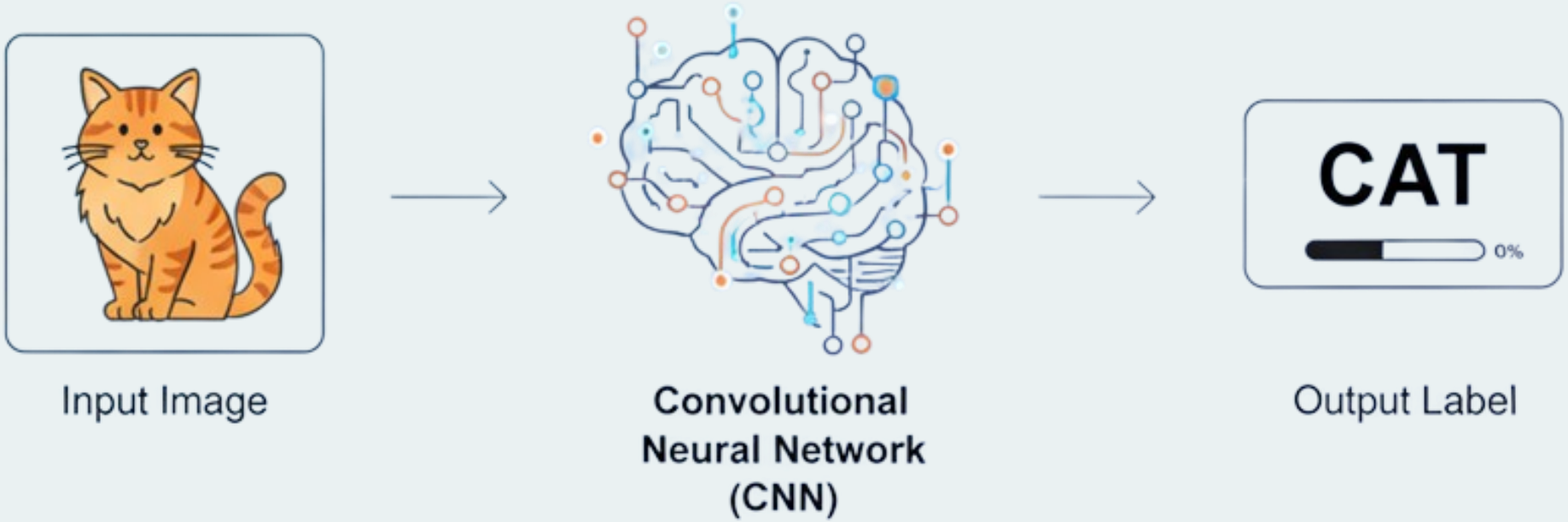
From pixels to prompts:
A beginner's guide to CNN and transformers.



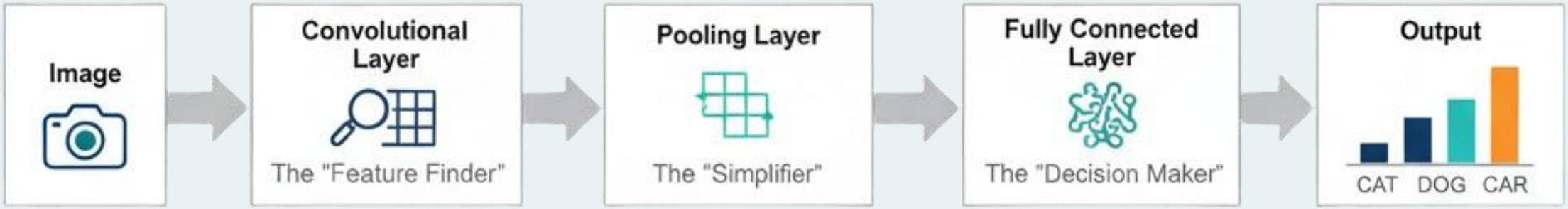
Aashish Thakur

What is Convolution Neural Network (CNN)?



Source: AI Generated

The Core Building Blocks of a CNN



Source: AI Generated

The Convolutional layer: The “Feature Finder”

Figure : 1

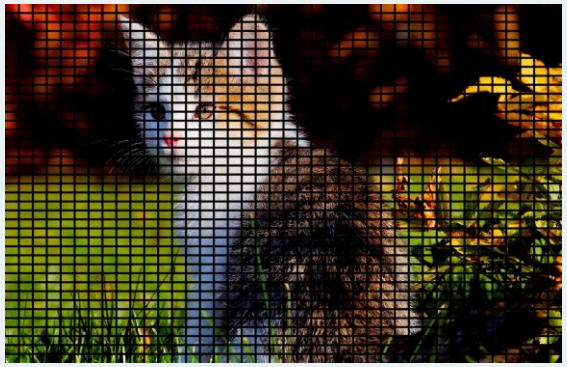
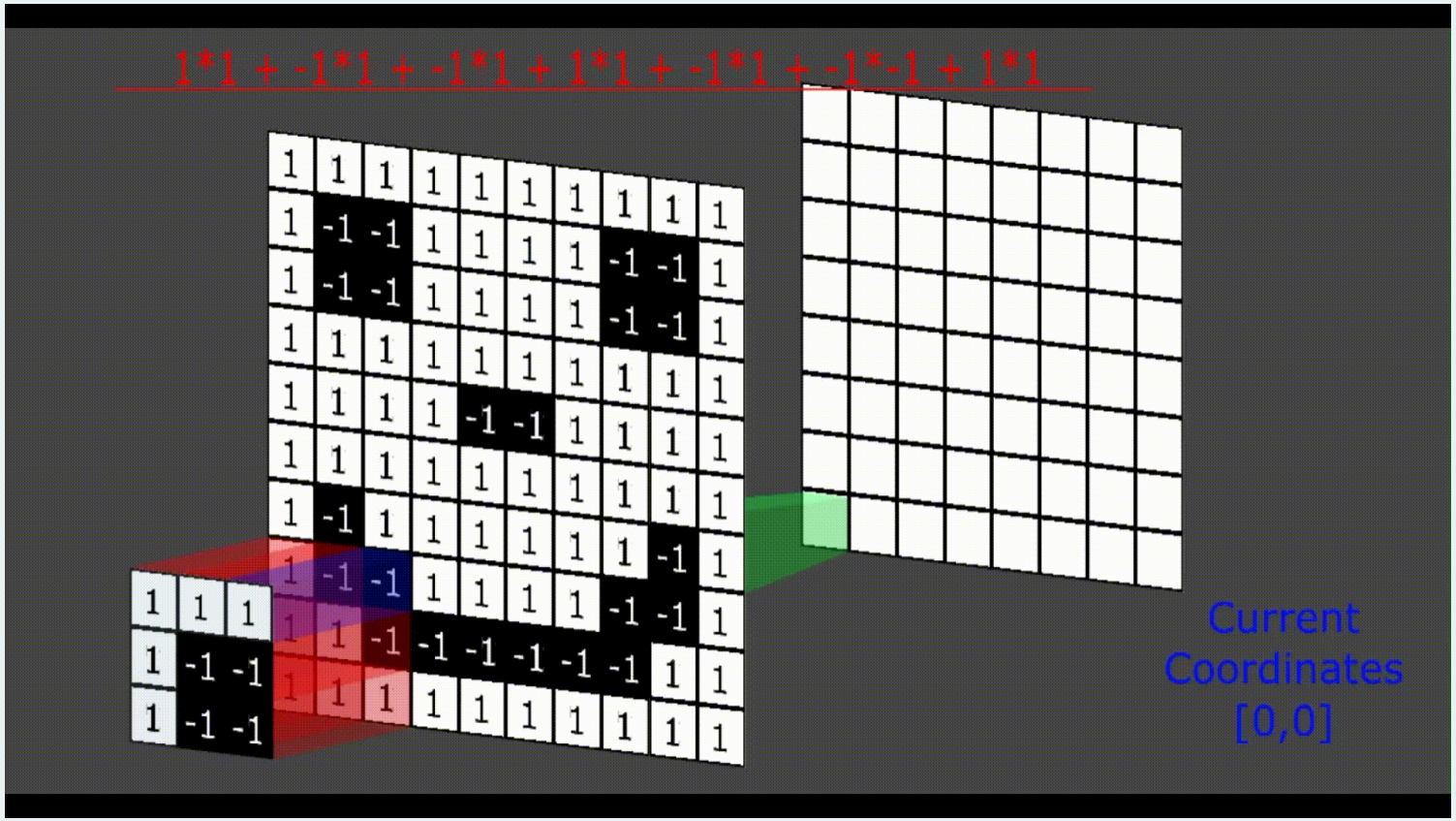
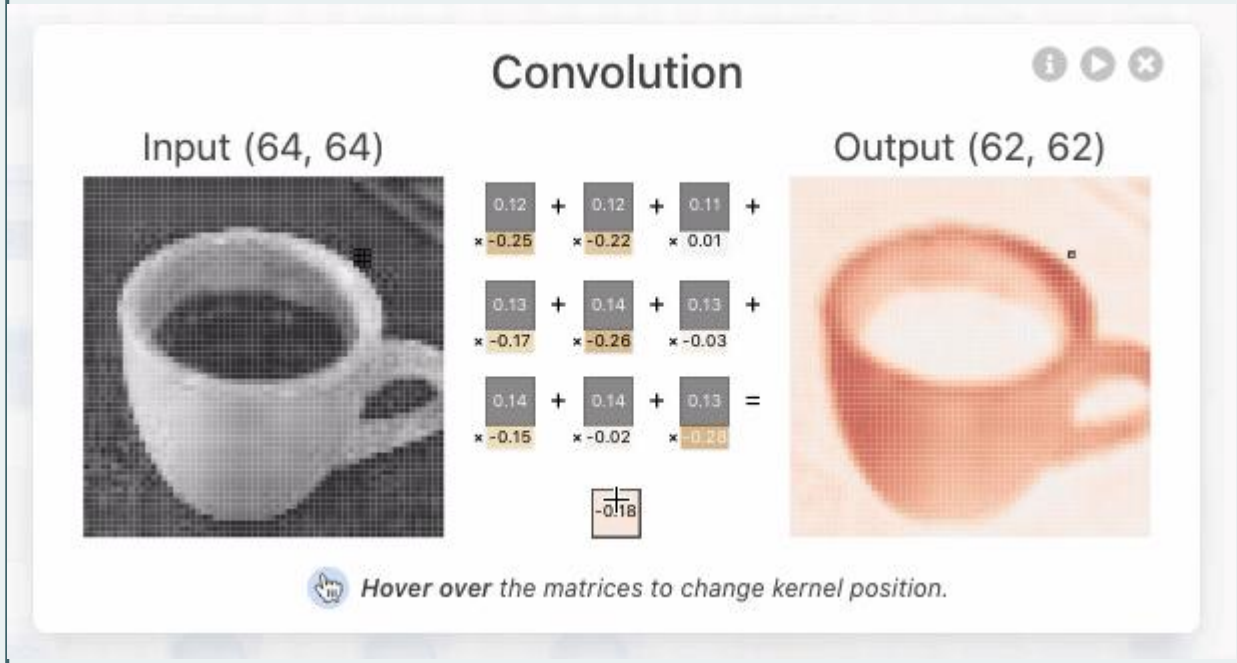


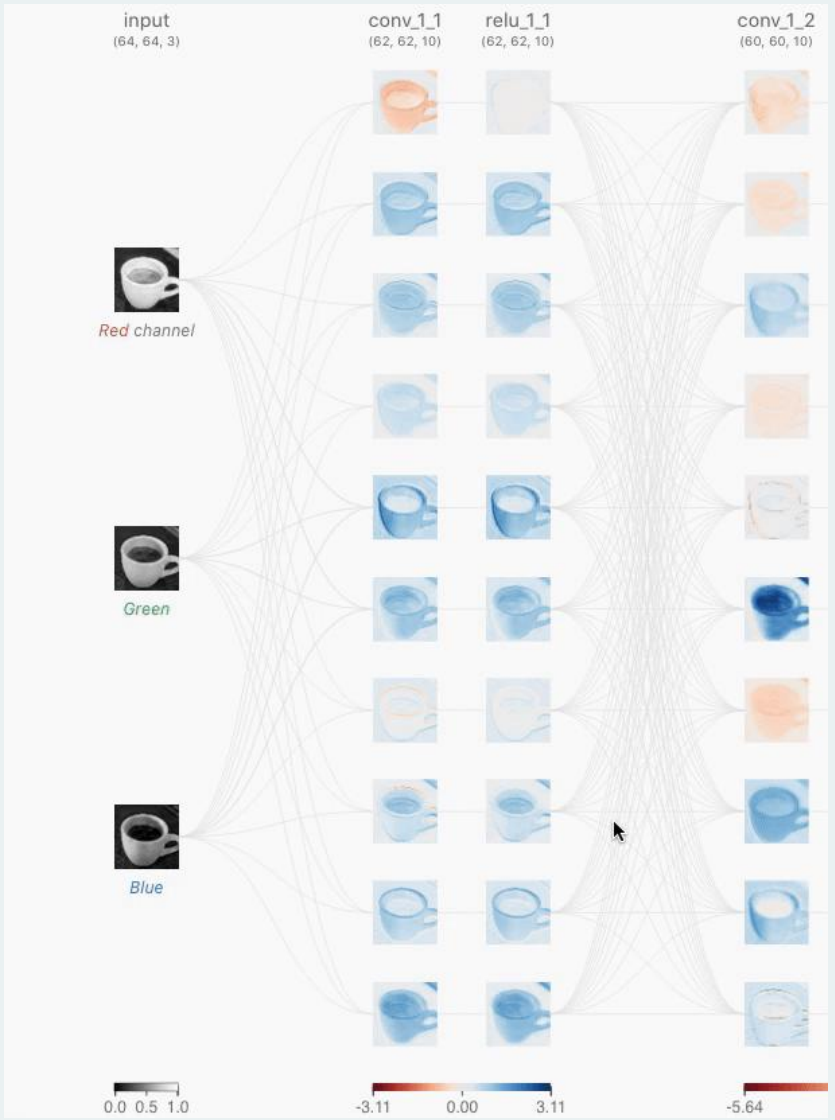
Figure : 2



Source :Max-Pooling Layer (CNN Visualization)

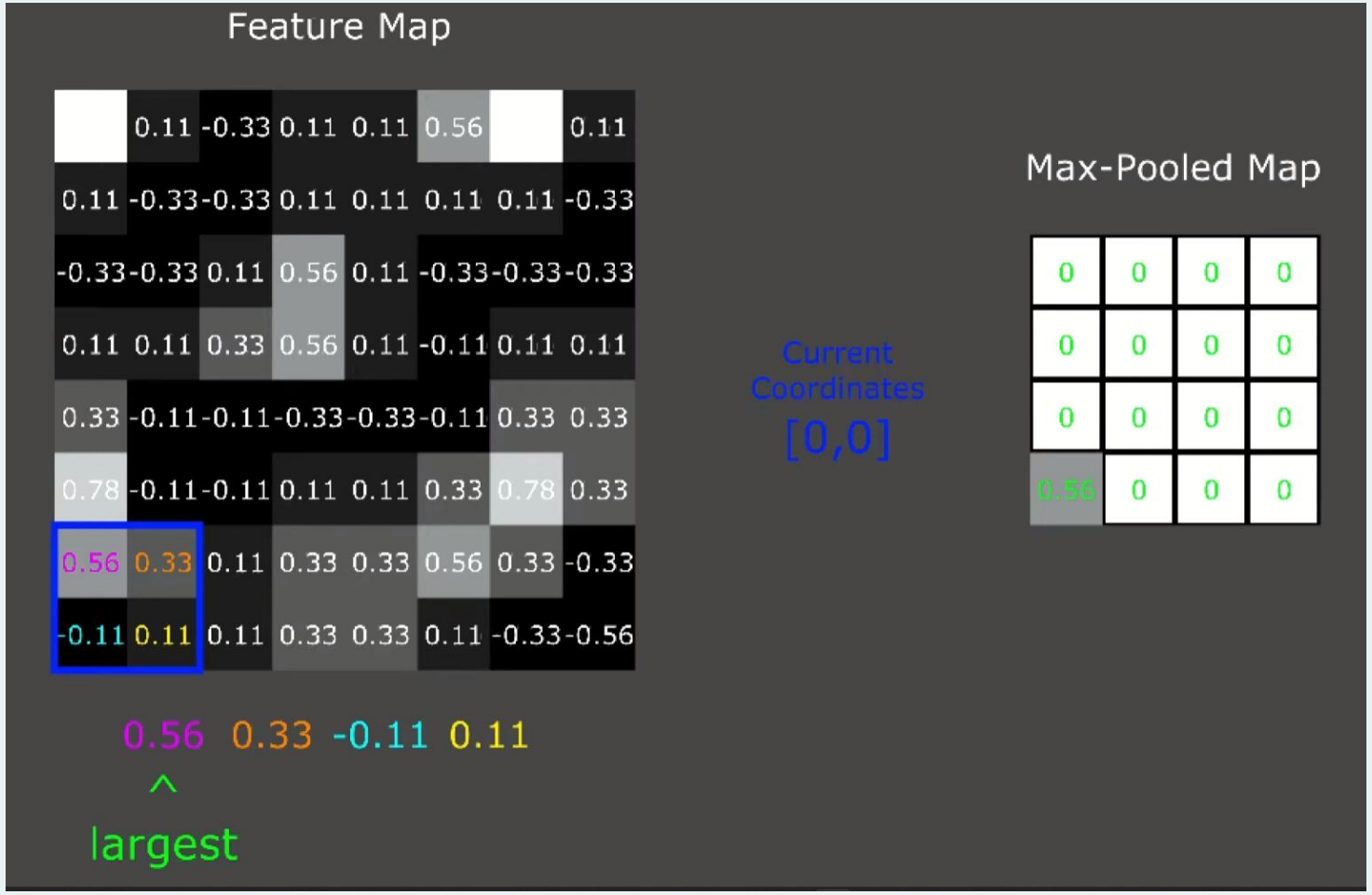
The Convolutional layer: The “Feature Finder”





Source: convlayer_detailedview_demo.gif (614×327)

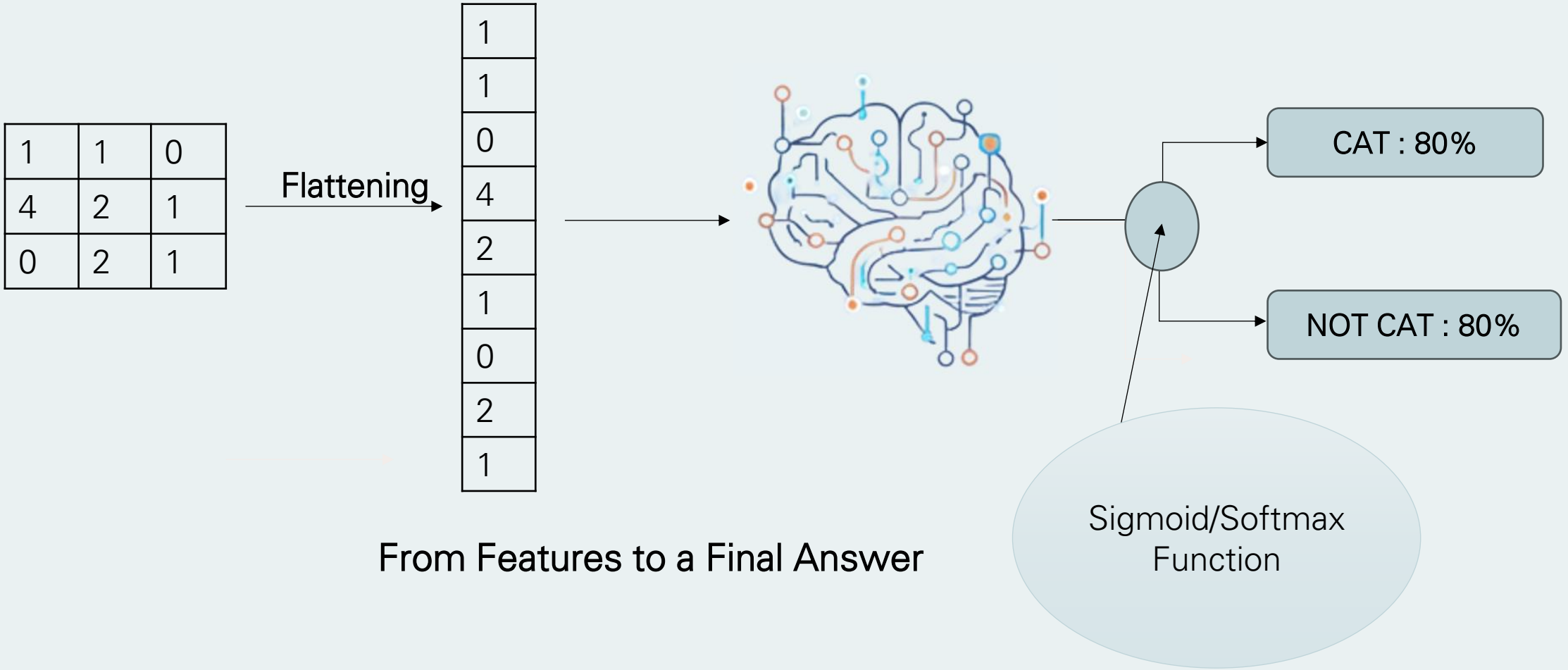
The Pooling Layer: Making Sense of Features



Max Pooling

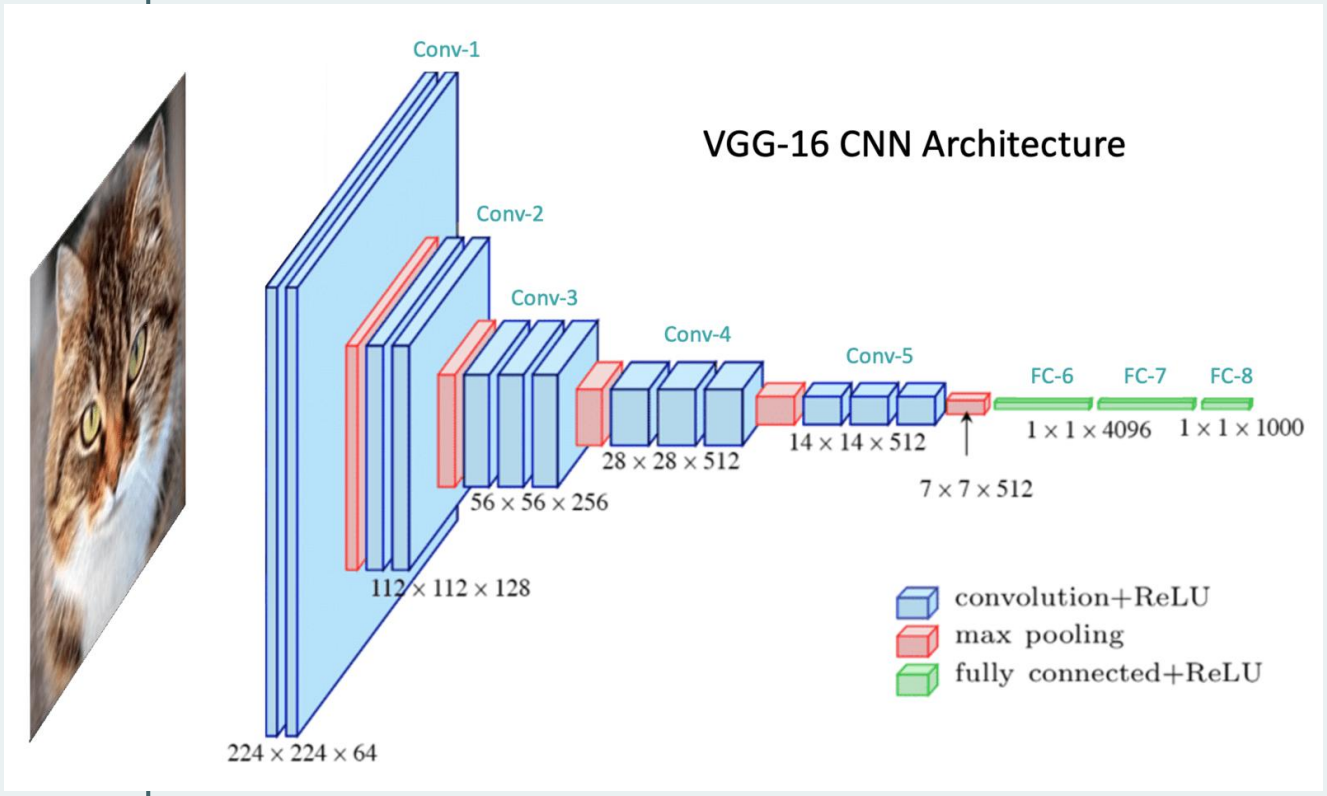
Source :Max-Pooling Layer (CNN Visualization)

The Fully Connected Layer: Making the Final Call

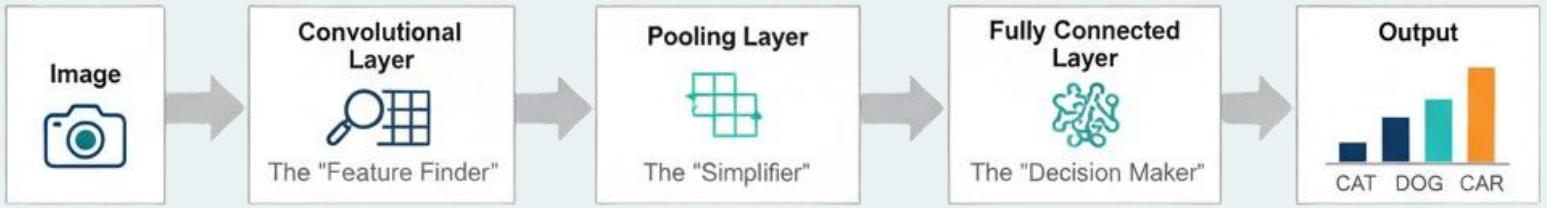


From Features to a Final Answer

CNNs in a Nutshell

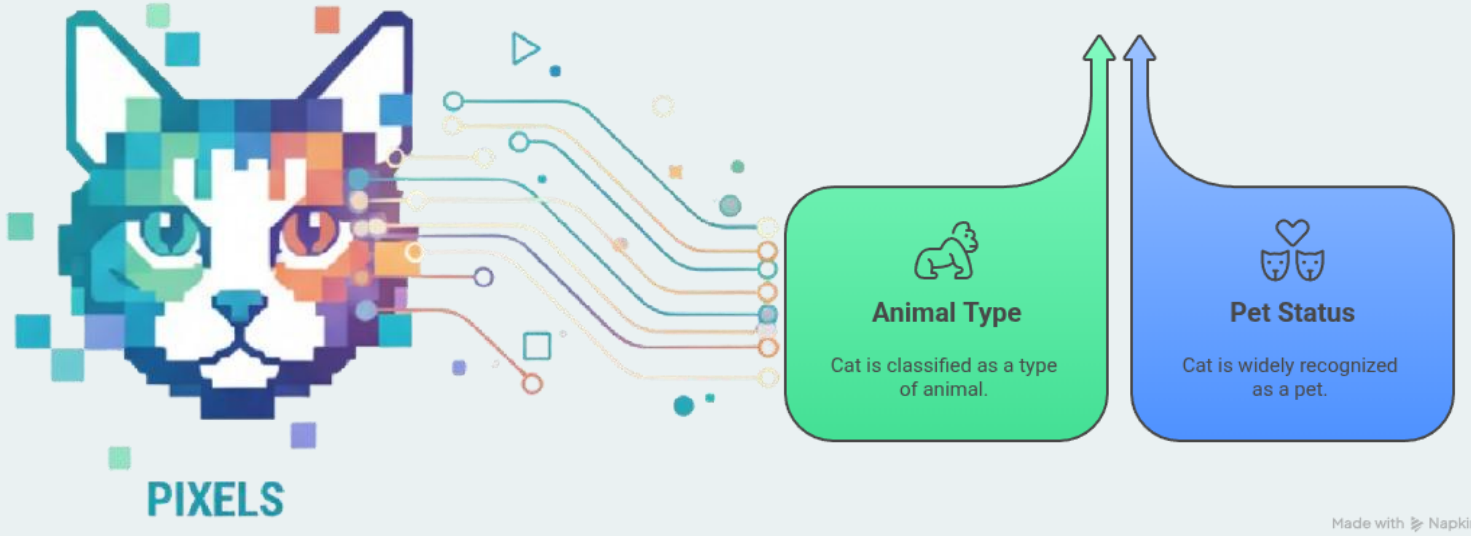



```python
inputs = layers.Input(shape=input_shape)

# --- Block 1 ---
x = layers.Conv2D(32, kernel_size=3, padding='same', activation='relu')(inputs)
x = layers.BatchNormalization()(x)
x = layers.MaxPooling2D(pool_size=2)(x)          # reduces H,W by 2

# --- Block 2 ---
x = layers.Conv2D(64, kernel_size=3, padding='same', activation='relu')(x)
x = layers.BatchNormalization()(x)
x = layers.MaxPooling2D(pool_size=2)(x)          # reduces H,W by 2 again

# --- Block 3 ---
x = layers.Conv2D(128, kernel_size=3, padding='same', activation='relu')(x)
x = layers.BatchNormalization()(x)
x = layers.MaxPooling2D(pool_size=2)(x)          # reduces H,W by 2 again

# --- Classification head ---
x = layers.Flatten()(x)
x = layers.Dense(128, activation='relu')(x)
x = layers.Dropout(dropout_rate)(x)
outputs = layers.Dense(num_classes, activation='softmax')(x)
```



Source:https://learnopencv.com/wp-content/uploads/2023/01/tensorflow-keras-cnn-vgg-architecture.png

Transformers - Teaching Computers to Understand Language



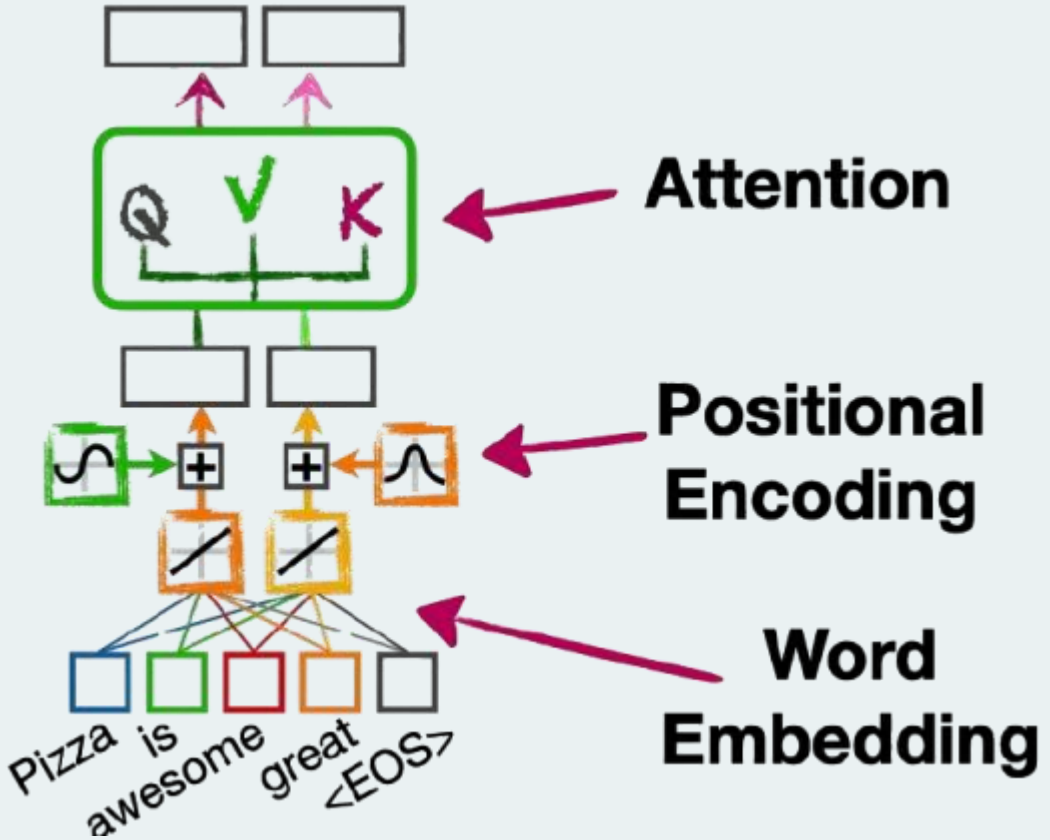
What are Transformers?

The Powerhouse of Modern Language AI

Transformer

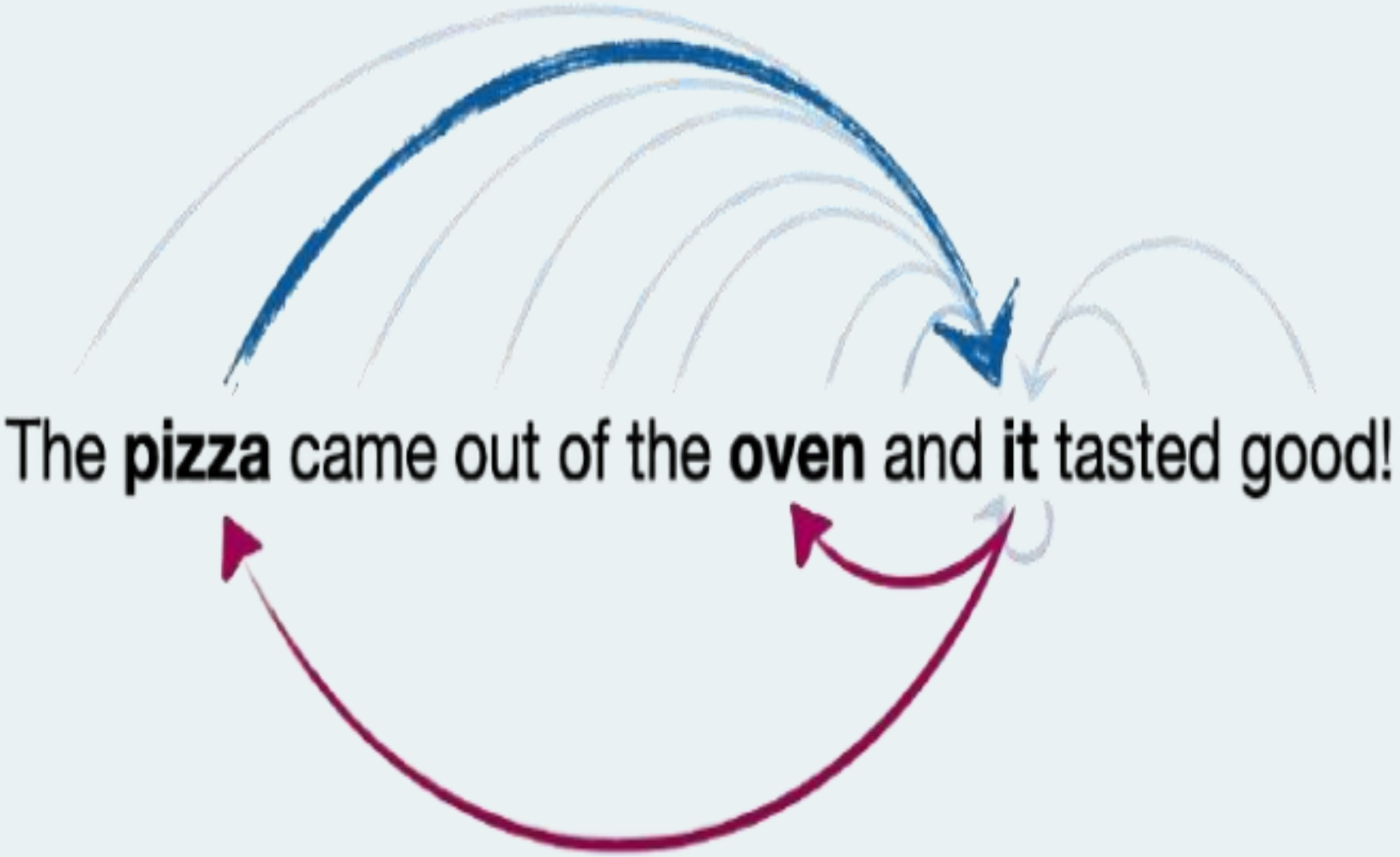
Large Language Model (LLM)

The Journey of a Word: Three Core Components



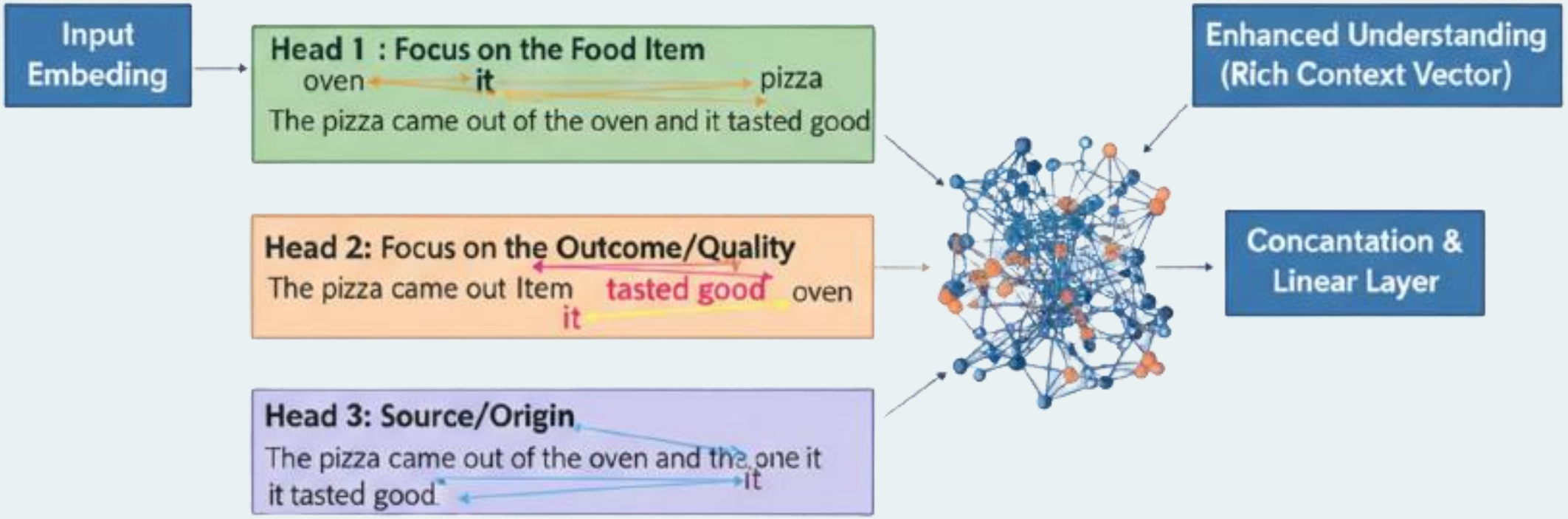
Source: Deeplearning.ai

The Secret Sauce: "Self-Attention“



Source: Deeplearning.ai

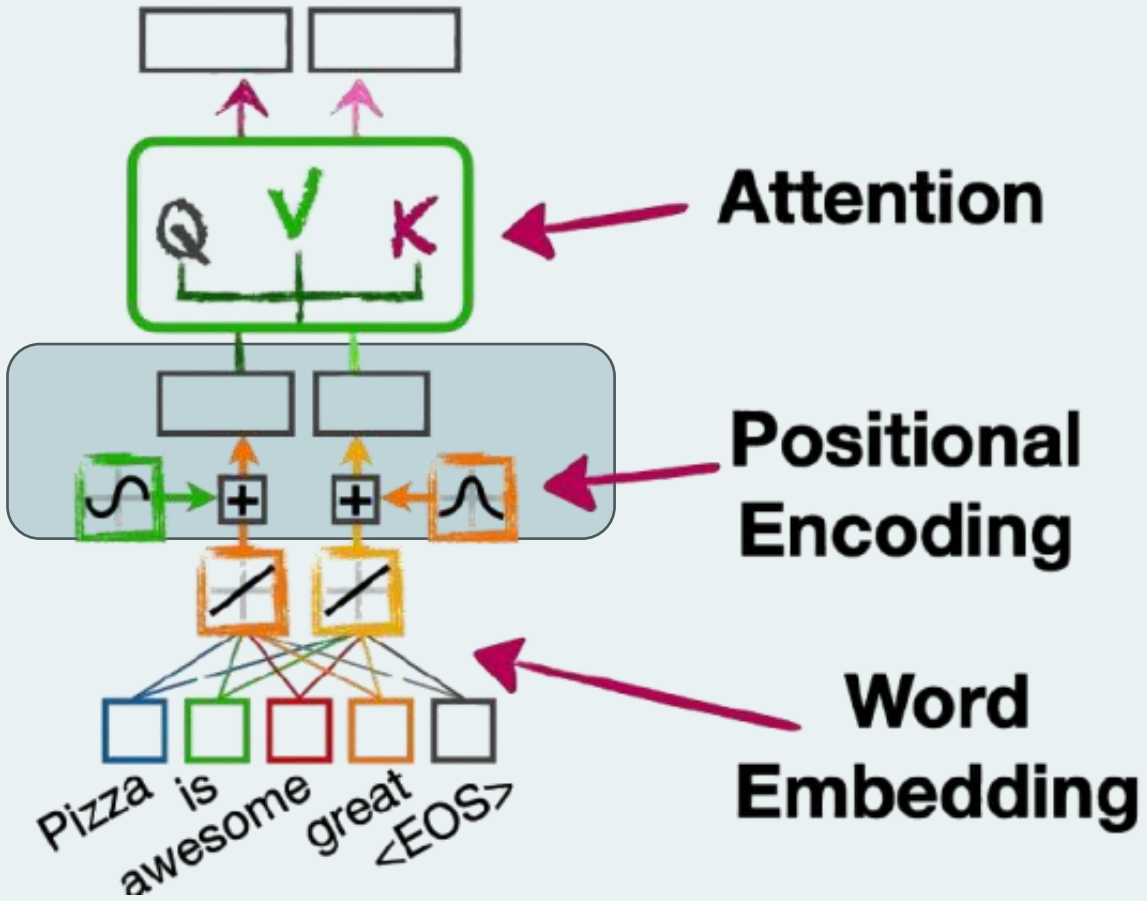
Going Deeper: Multi-Head Attention & Positional Encodings



Positional Encodings

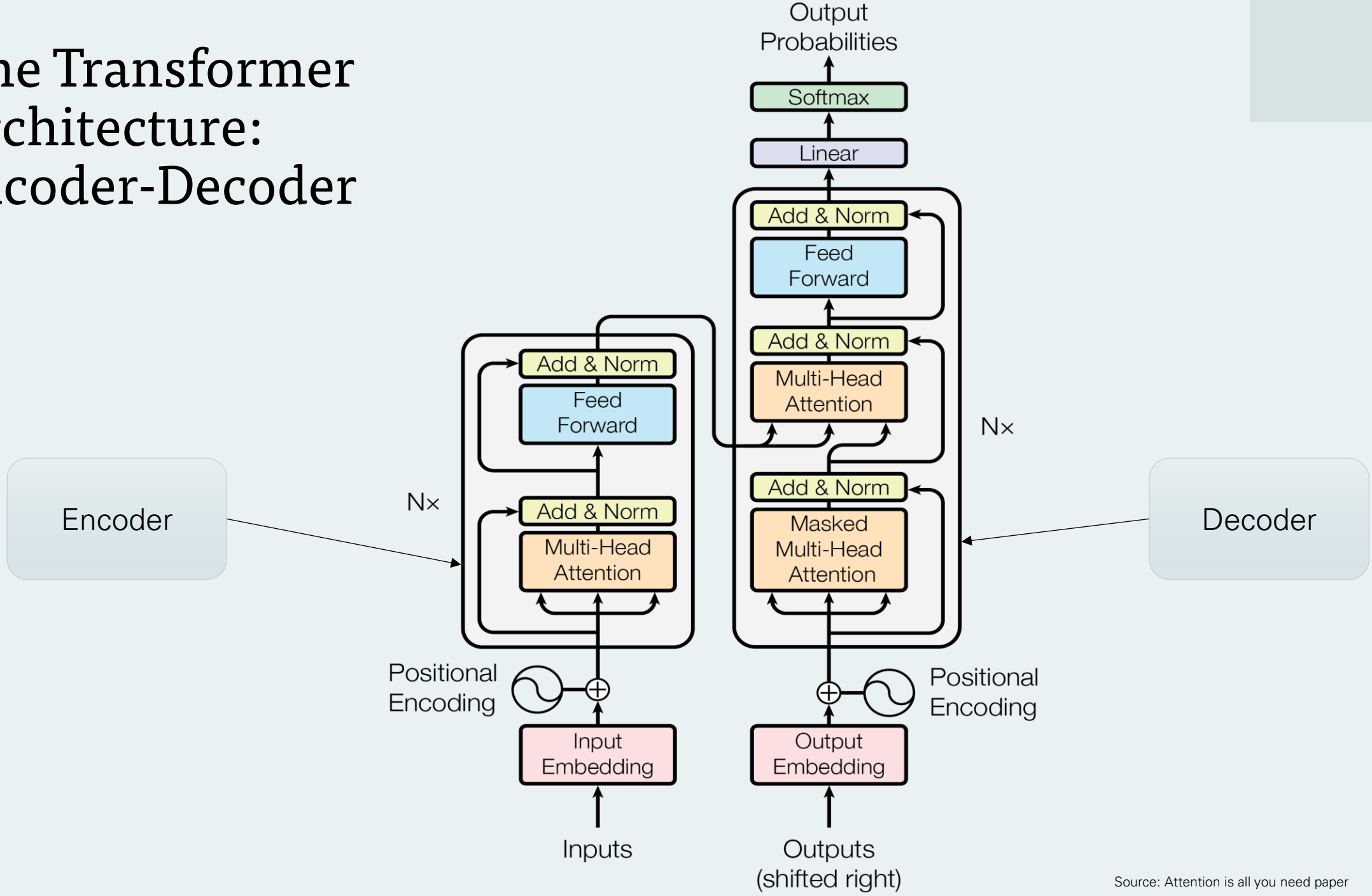
‘The pizza came out of oven and it tasted good’

'The pizza tasted good and it came out of the oven’



Source: Deeplearning.ai

The Transformer Architecture: Encoder-Decoder

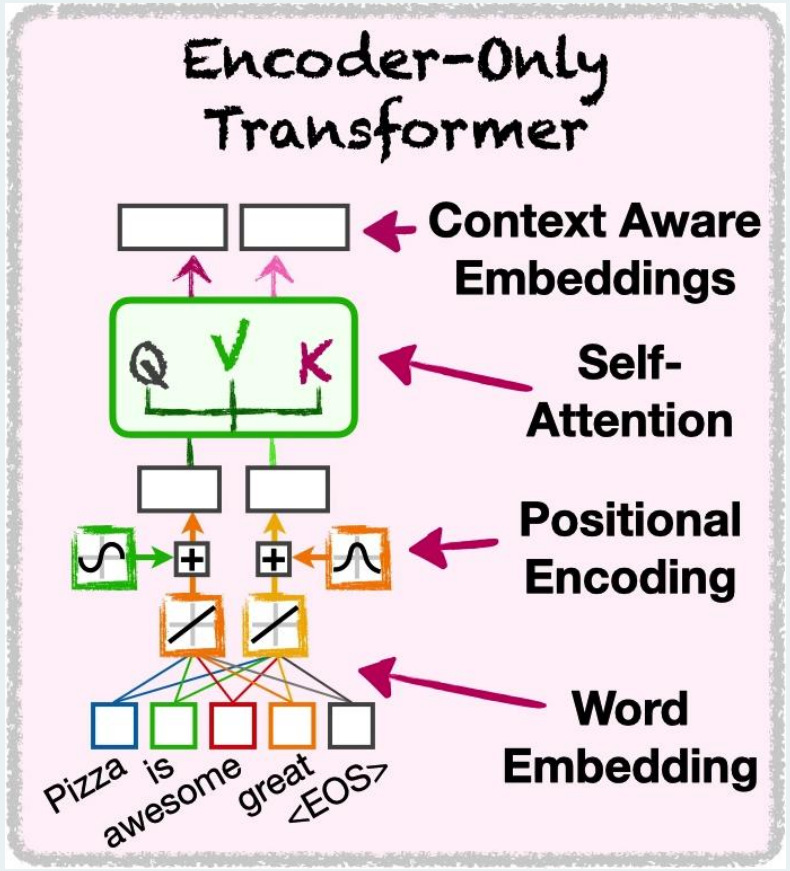


Source: Attention is all you need paper

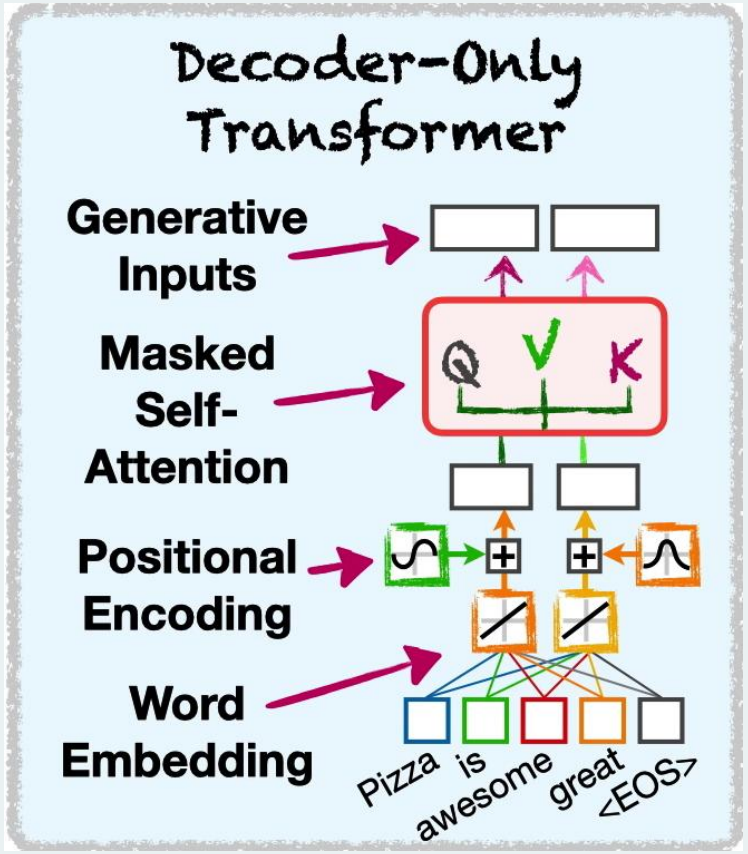
Masked Self Attention

Source: Deeplearning.ai

Transformers and the Rise of LLMs



Eg: Google Bert for sentiment Analysis



Eg: Openai's GPT(Generative Pretrained Transformers)

Source: Deeplearning.ai

Key Takeaways and Next Steps

- **CNNs** are masters of spatial data like images.
- **Transformers** excel at sequential data like text, thanks to the power of self-attention.
- **Next Steps:** We will now dive into the provided Colab notebook to see a CNN in action. I will also provide a link to a basic Transformer notebook for you to explore.

Questions?