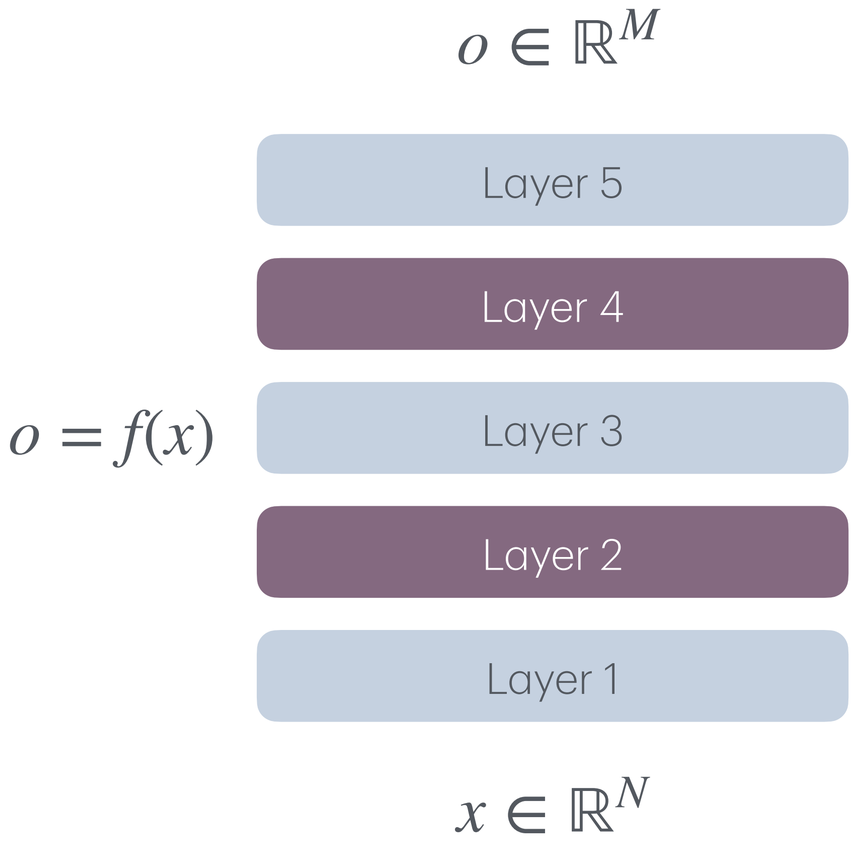
Background: Structures of Deep Networks

Philipp Krähenbühl, UT Austin

What is a Deep Network?

- A really BIG differentiable function
  - Stacks layers of “simple” functions
    - Computation Graph
  - Trained with gradient descent and automated differentiation (backpropagation)



Elements of a deep network

Linear Layers

- Linear
- Convolution

- Have weights (trainable parameters)
- Very simple computation

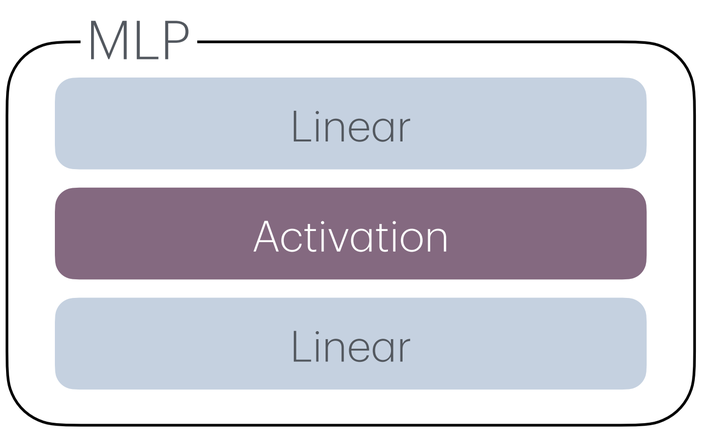
Nonlinear Layers

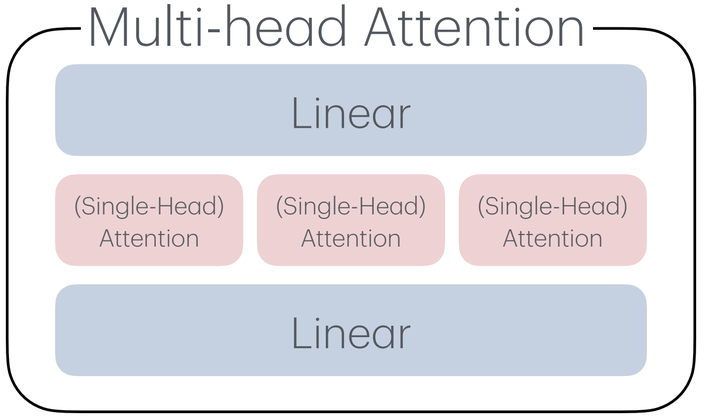
- Normalization
- Activation
- (Single-Head) Attention
- Pooling

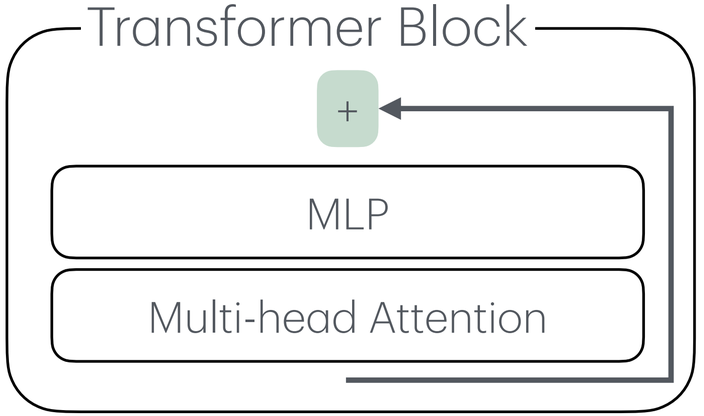
- Very few trainable parameters
- More complex computation

Elements of a deep network

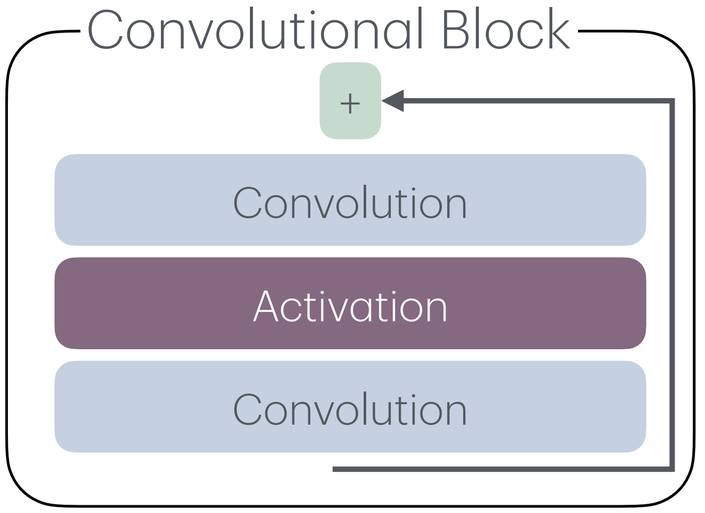
Blocks







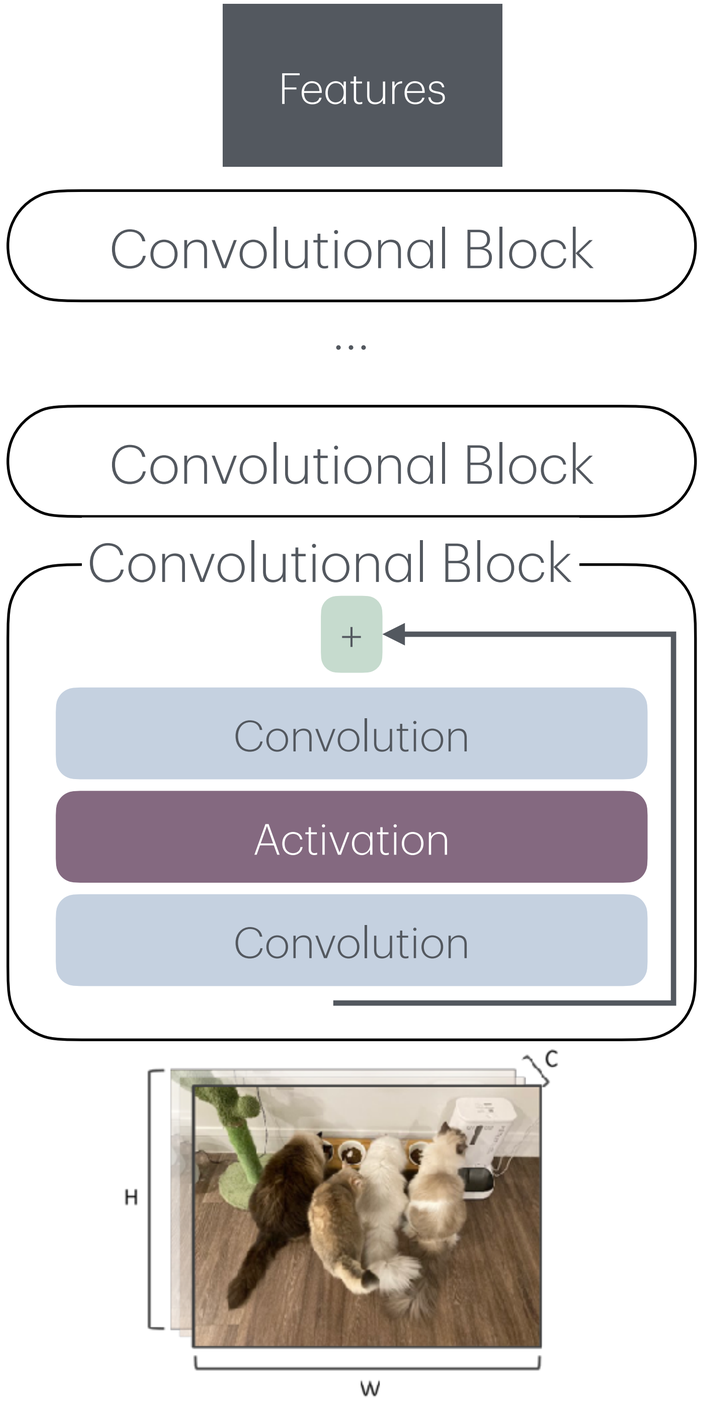
- Combinations of Layers
- Non-Linear
- With Parameters



What is a deep network?

Example: Convolutional Network

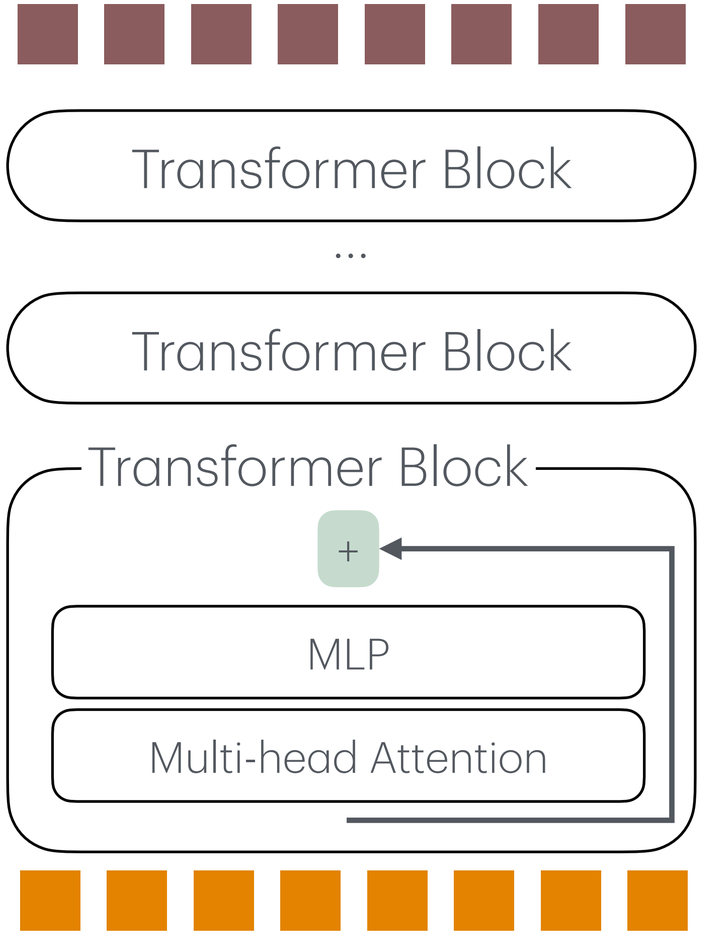
- Convolution: Fast, memory efficient
- Preserve image structure
  All activations are “images”
- Stride
  - Reduce spatial dimension
  - Increase channel dimension
  - Wider receptive field, less computation



What is a deep network?

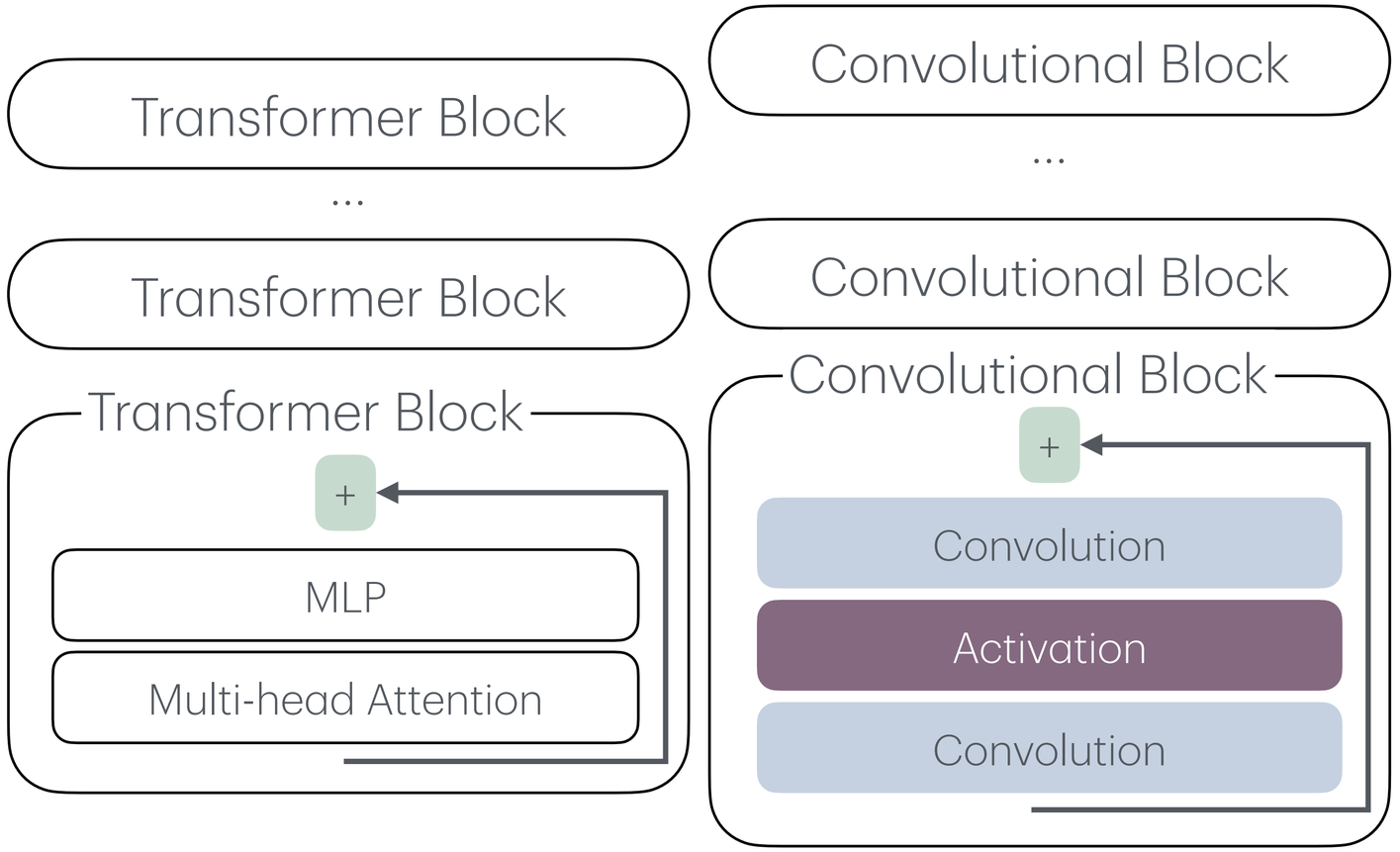
Example: Transformer

- Set-based architecture
  - Input: Set of embeddings / Tokens
  - Output: Set of features
- Positional encoding for ordered data



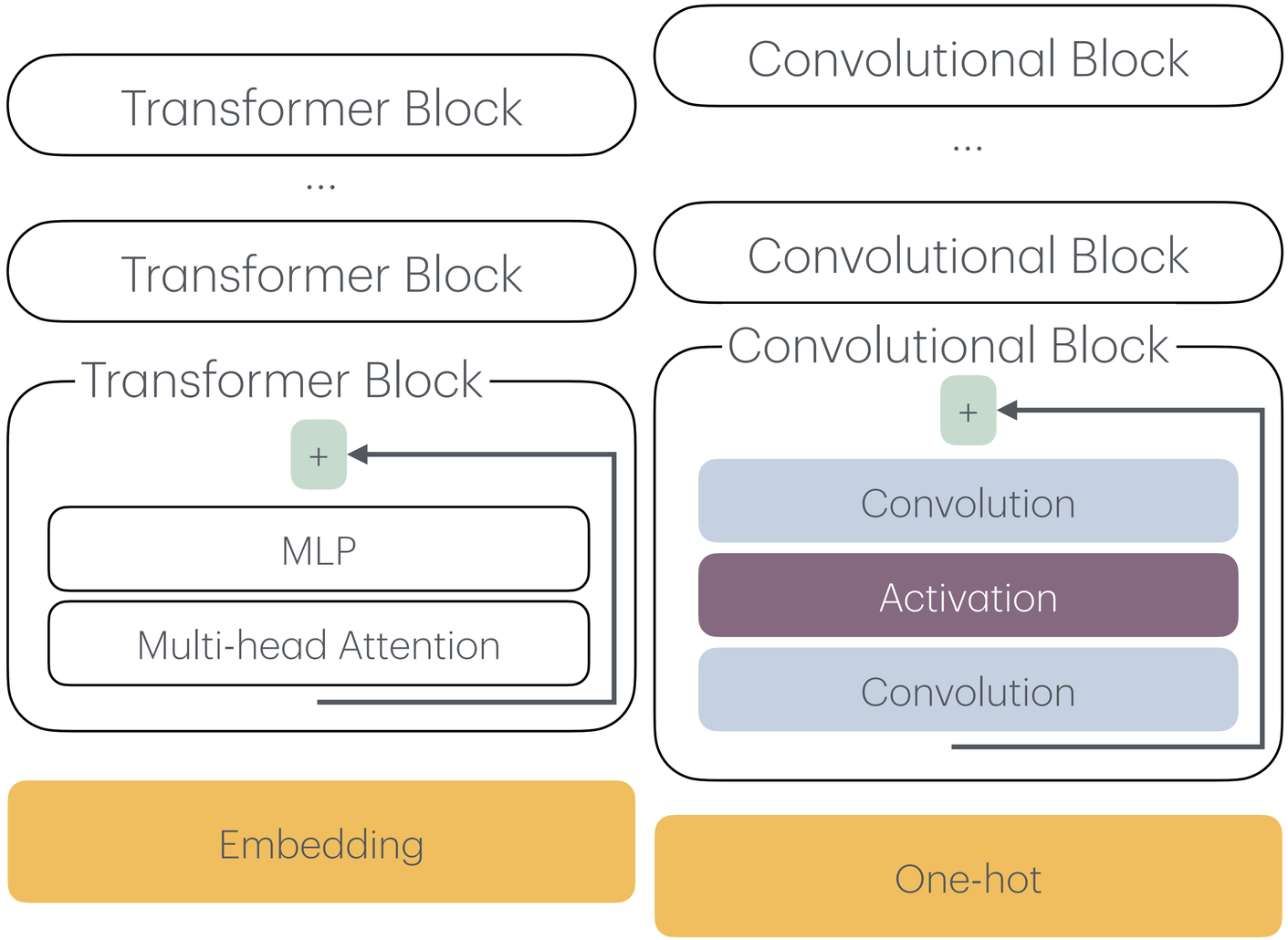
Inputs and Outputs

- Input: Real-valued vector
- Output: Real-valued vector



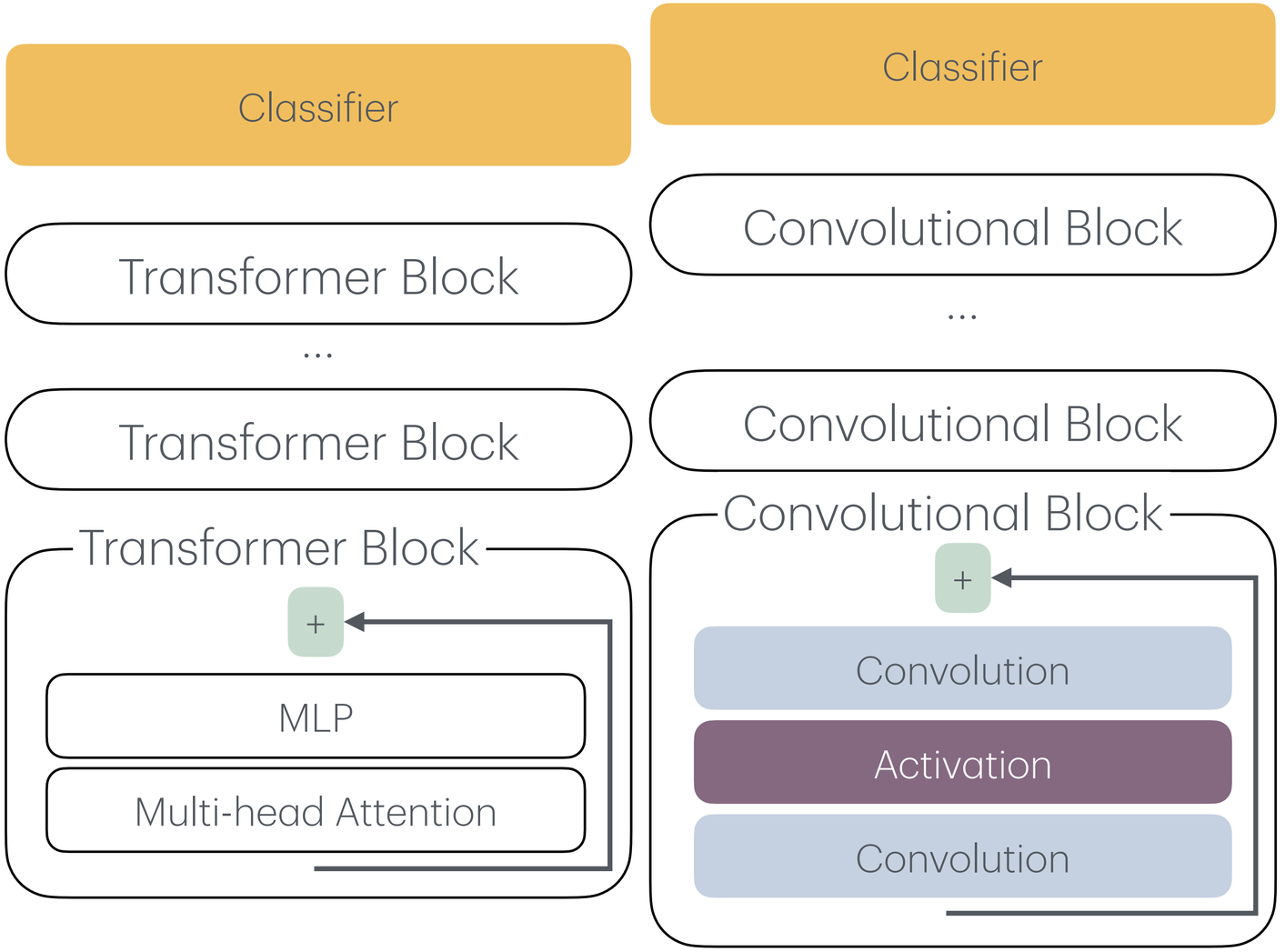
Inputs transformations

- Real-valued inputs
  - Feed in directly
- Categorical inputs
  - Convert to one-hot
  - Convert to embedding



Output transformations

- Real-valued outputs
  - Use directly (**regression**)
- Categorical outputs
  - Predict real-valued probabilities over categories (**classifier**)



What is a Deep Network?

- A really BIG differentiable function
  - Stacks layers of “simple” functions
    - Computation Graph
  - Trained with gradient descent and automated differentiation (backpropagation)