

# Transformers - Summary

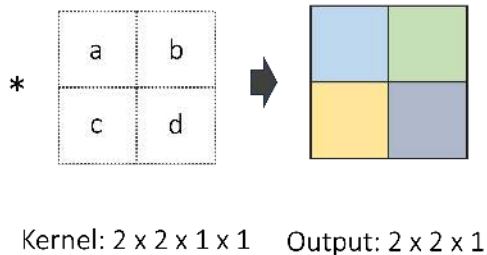
# Fixed vs Flexible Structure

## Images / ConvNets

- Fixed input (resolution)
- Fixed structure



Input:  $3 \times 3 \times 1$



## Language / Transformers

- Variable length
- Diverse structure
- Transformers as powerful as ConvNets
  - In practice, just a bit more data-hungry

