

# Attention (BKUP)

## Transfomer

- All the sentence processed at the same time
- What is an Attention Map?
  - a grid that tells how much **focus** a model places on different parts of the input text when processing a specific token
- Example: "flag" → "American"
  - The brightest square (highlighted) shows it's paying extremely high attention to the token "**American**"
  - **Attention allows a model to dynamically create relationships between tokens**

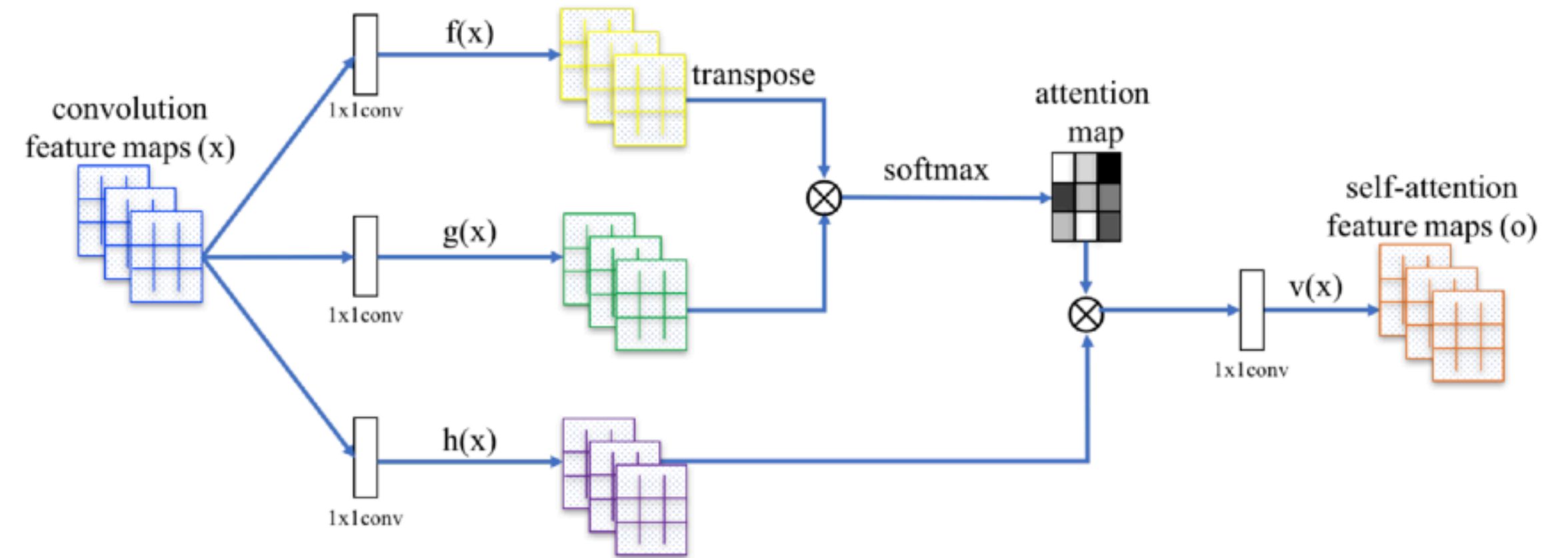
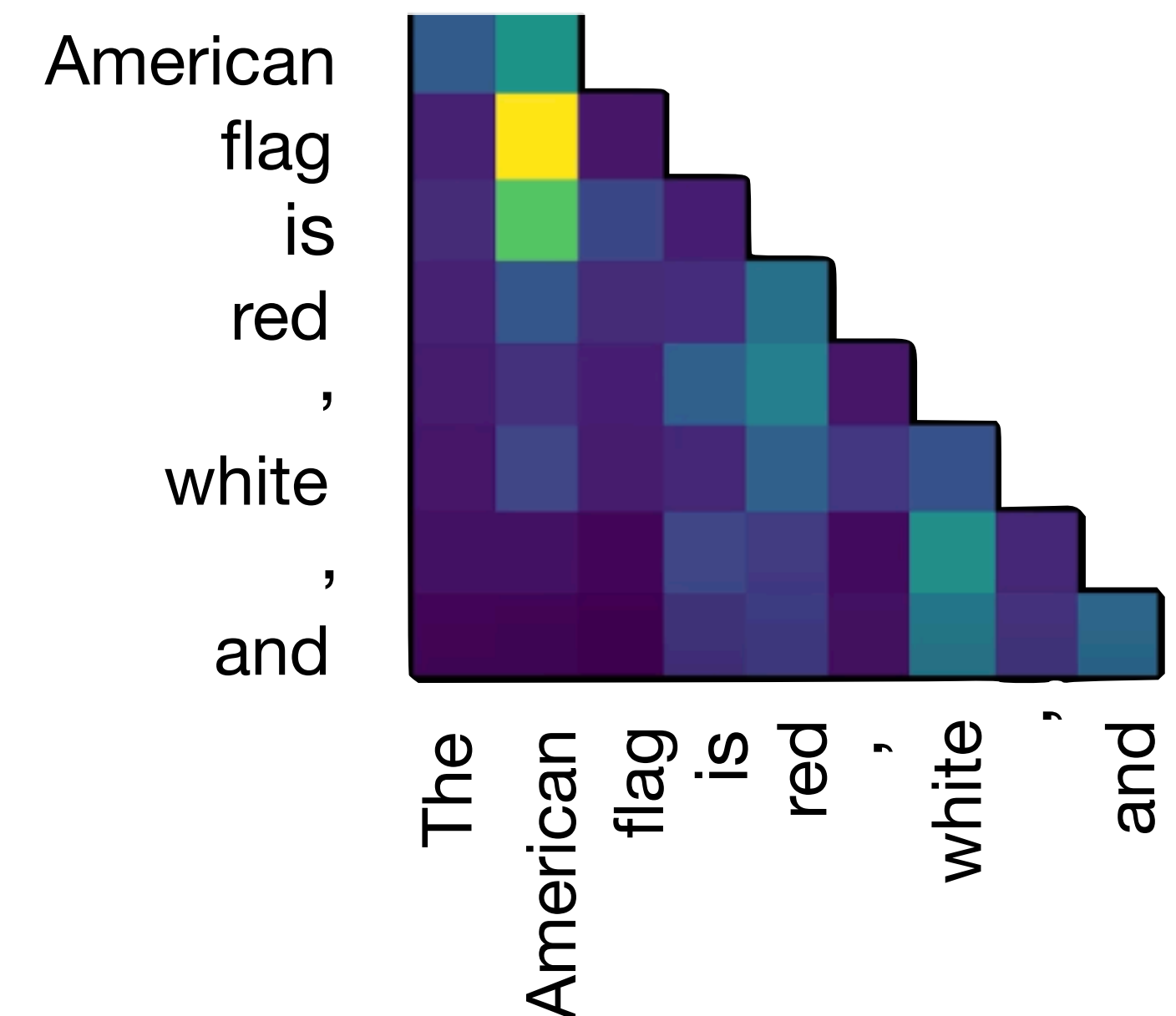


Figure 20: The self-attention mechanism in SAGAN. (Image source: Fig. 2 in Zhang et al., 2018)



# Attention (BKUP)

## Transfomer

