

Tutorial 3
CSC 420
Vision Transformers

Andrew Guo

Outline

- Brief overview
- Hints for HW Q8
- Interactive tutorial on Jupyter notebook

Outline

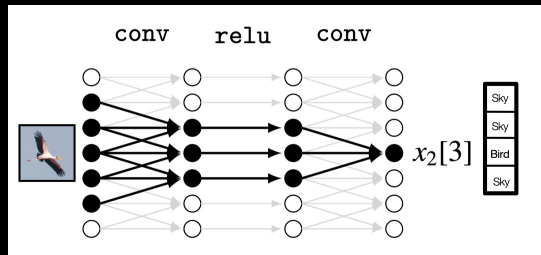
- Brief overview
- Hints for HW Q8
- Interactive tutorial on Jupyter notebook

Limitations of MLPs

- Architectures = inductive biases. We want models that are expressive and encode useful priors.
- MLPs: universal but weak priors → data-hungry; dense layers are compute-heavy.
- CNN strengths: locality, translation equivariance, parameter sharing.

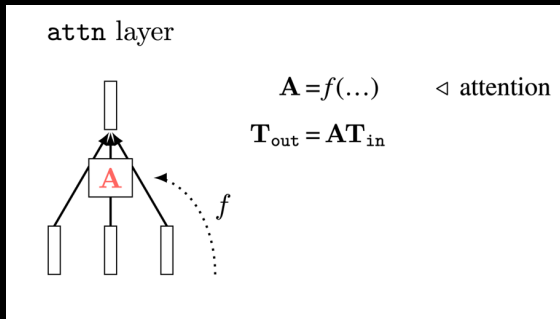
Limitations of CNNs

- Long-range interactions are indirect (need many layers to grow receptive field).
- Harder to model relationships between far-apart regions (e.g., counting or matching distant objects).
- **Motivation:** We want data-dependent, global interactions without giving up parallelism.



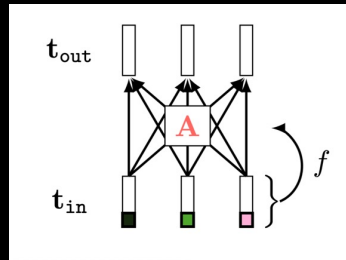
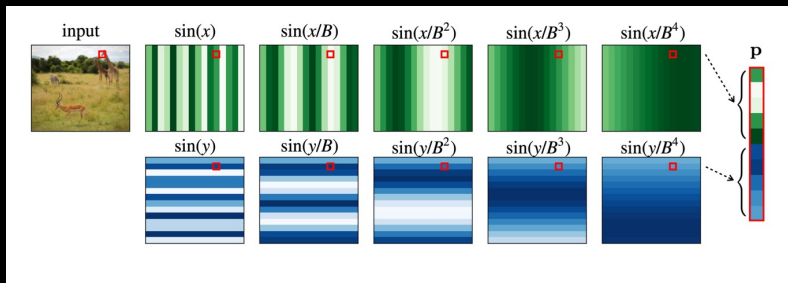
Attention

- Intuition: Let each *token* decide which other tokens to read from, and by how much.
- Data-dependent connectivity (dynamic receptive field).
- Fully parallel (matrix multiplies).



Positional embeddings

- Attention is permutation invariant, order of input tokens does not affect outcome
- For images, this is a problem because then we do not know which token is where in the original image
- Solution: add positional encoding



Outline

- Brief overview
- Hints for HW Q8
- Interactive tutorial on Jupyter notebook

Q8 part 1

- Focus on self attention from lectures, starting from slide 57
- Solution should be relatively straightforward and concise
- Look at the class definitions, try to use the attributes and functions already implemented

Q8 part 2

- See Piazza post: <https://piazza.com/class/me8t5l4u2tw4fm/post/16>
- Please be sure to answer the other parts of the question

HW1 Q8 clarification

Updated 4 days ago by David Lindell

Hi all,

Looks like we accidentally included some code in `a1_q8_helper.py` that implements a solution for the attention visualization problem for Q8.

Feel free to use this code, but make sure you answer the other parts of the question about the role of CLS attention and your analysis of the attention maps for a set of example images.

David

Outline

- Brief overview
- Hints for HW Q8
- Interactive tutorial on Jupyter notebook