Machine Learning
Feature Pooling

Mostafa S. Ibrahim
Teaching, Training and Coaching for more than a decade!

Artificial Intelligence & Computer Vision Researcher
PhD from Simon Fraser University - Canada
Bachelor / MSc from Cairo University - Egypt
Ex-(Software Engineer / ICPC World Finalist)

Summarizing Numbers

- Do you remember ways to aggregate information of array of numbers?

- Mean (Arithmetic, geometric, harmonic)
- Min / Max / Median
- This kind of information summarizing/aggregating is known as feature pooling in ML context

Feature pooling

- When we apply the convolution layer we end up with
  - Another huge dimensions WxH
    - We can’t just use that with the loss function
    - We can’t keep processing with such dimensions due to computational cost!
  - Extracted features are on very small grid, e.g. 3x3
- Feature pooling can help us tackle these 2 problems
  - Help next layers to reduce the computational cost
  - Help next layers o see bigger view!
- The most common pooling techniques
  - Max Pooling (very common)
  - Average Pooling
- Convolution layers are typically followed by pooling to reduce the size!

Max pooling

- In max pooling, each region of the feature map is divided into **non-overlapping** sections (usually 2x2 or 3x3), and the **maximum value** within each section is selected as the representative value for that region
  - This helps in preserving the most **salient features** within each region.
    - **Salient features** refer to the **most distinctive** features in the region.
      - Such as edges/corners in images

Max pooling

- Assume a max-pooling of 3x3
  - Then divide to 3x3 blocks and take maximum per block
  - New dimensions W/3, H/3
- There are **no learnable parameters**!
- Average pooling: just take the average

15	2	3	4	5	6
7	8	9	10	11	12
1	2	3	16	17	**18**
1	20	21	7	**23**	4
25	2	27	1	2	30
6	**32**	3	1	2	5

15	18
32	23

Pytorch

- pool = nn.MaxPool2d(kernel_size=2, stride=2)
- Typically we use the kernel size = stride to make them non-overlapping

Beyond DNN

- Pooling is a general technique to aggregate information
- Imagine having K feature vectors each of D features representing some input
  - For example, we have K camera watching an object
- We can concatenate all of them as representation
  - We might call it concatenation pooling
  - Pros: all information is available
  - Cons: huge ⇒ computational cost
- We can pool all of them into a single D features
  - Index by index take the maximum
  - A = {1, 5, 7, 2, 6}
  - B = {6, 9, 1, 1, 8}
  - Pooling ⇒ {6, 9, 7, 2, 8}

"Acquire knowledge and impart it to the people."

"Seek knowledge from the Cradle to the Grave."