1. Explain convolutional neural network, and how does it work?

2. How does refactoring parts of your neural network definition favor you?

3. What does it mean to flatten? Is it necessary to include it in the MNIST CNN? What is the reason for this?

4. What exactly does NCHW stand for?

5. Why are there 7\*7\*(1168-16) multiplications in the MNIST CNN's third layer?

6.Explain definition of receptive field?

7. What is the scale of an activation's receptive field after two stride-2 convolutions? What is the reason for this?

8. What is the tensor representation of a color image?

9. How does a color input interact with a convolution?

Answers

Here are concise answers to your questions:

1. **Explain convolutional neural network, and how does it work?**  
   A CNN is a neural network designed for processing grid-like data (e.g., images). It uses convolutional layers to automatically detect spatial features by applying kernels/filters across the input.
2. **How does refactoring parts of your neural network definition favor you?**  
   It improves code readability, reusability, easier debugging, and helps modularize complex models.
3. **What does it mean to flatten? Is it necessary to include it in the MNIST CNN? What is the reason for this?**  
   Flattening converts a multi-dimensional tensor (e.g., 3D image feature map) into a 1D vector. Yes, it's necessary before feeding the data into fully connected layers which expect 1D input.
4. **What exactly does NCHW stand for?**  
   NCHW stands for: **Batch size (N)**, **Channels (C)**, **Height (H)**, **Width (W)** – the tensor format used in many deep learning frameworks like PyTorch.
5. **Why are there 7×7×(1168−16) multiplications in the MNIST CNN's third layer?**  
   Because each of the 7×7 spatial locations is convolved with a kernel over 1152 (=1168−16) input channels (assuming 16 filters were used previously), leading to that many multiplications per output channel.
6. **Explain definition of receptive field?**  
   The receptive field is the region in the input image that a particular neuron in a deeper layer "sees" or is influenced by, based on kernel size, stride, and depth.
7. **What is the scale of an activation's receptive field after two stride-2 convolutions? What is the reason for this?**  
   It increases exponentially. After two stride-2 convolutions with 3×3 kernels, the receptive field is **7×7**. Each layer doubles the field and adds overlap due to kernel size.
8. **What is the tensor representation of a color image?**  
   A 3D tensor: **(3, Height, Width)** – one channel each for Red, Green, and Blue.
9. **How does a color input interact with a convolution?**  
   Each filter has 3 channels (for RGB), and performs element-wise multiplication across all channels. The results are summed to produce a single output value per spatial position.