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

**Ans:** A convolutional neural network (CNN) is a type of deep learning algorithm used primarily for image recognition and classification tasks. It consists of multiple layers of neurons, including convolutional layers, pooling layers, and fully connected layers. CNNs work by passing input images through a series of convolutional and pooling layers to extract features at different spatial scales. These features are then flattened and fed into one or more fully connected layers for classification or regression.

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

**Ans:** Refactoring parts of a neural network definition can benefit by improving modularity, readability, and maintainability of the code. It allows for easier debugging, testing, and modification of specific components without affecting the entire network structure. Additionally, refactoring can facilitate code reuse and abstraction, leading to more efficient development and better organization of the neural network architecture.

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

**Ans:** Flattening in the context of neural networks refers to converting a multidimensional tensor into a one-dimensional vector. In the case of MNIST CNN, flattening is necessary to convert the output feature maps from the convolutional and pooling layers into a format suitable for input into the fully connected layers. This is necessary because fully connected layers require one-dimensional input vectors.

**4. What exactly does NCHW stand for?**

**Ans:** NCHW stands for "Batch size, Number of Channels, Height, Width," which is a common data format used in deep learning frameworks like PyTorch and TensorFlow to represent multi-dimensional tensors, particularly for convolutional neural networks.

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

**Ans:** There are 77(1168-16) multiplications in the MNIST CNN's third layer because it likely consists of a convolutional layer with a 7x7 kernel, 1168 input channels, and a bias term, and a ReLU activation function applied element-wise. The subtraction of 16 may be due to the bias term being applied to each output channel.

**6.Explain definition of receptive field?**

**Ans:** Receptive field refers to the region of the input space that a particular neuron in a neural network is sensitive to. In convolutional neural networks, the receptive field of a neuron in a particular layer is determined by the size of the convolutional kernels and the strides used in previous layers.

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

**Ans:** After two stride-2 convolutions, the scale of an activation's receptive field is increased by a factor of 4. This is because each stride-2 convolution effectively downsamples the spatial dimensions of the activation by a factor of 2, resulting in an enlarged receptive field.

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

**Ans:** The tensor representation of a color image typically follows the NCHW format, where N is the batch size, C is the number of channels (usually 3 for RGB images), H is the height of the image, and W is the width of the image.

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

**Ans:** A color input interacts with a convolutional layer in the same way as a grayscale input, except that it has multiple channels (usually three for red, green, and blue). Each channel of the input image undergoes separate convolution operations with corresponding kernels in the convolutional layer, resulting in feature maps for each channel. These feature maps are then combined through element-wise addition or concatenation to form the output feature maps of the convolutional layer.