REPORT

Model Architecture and Training Setup:

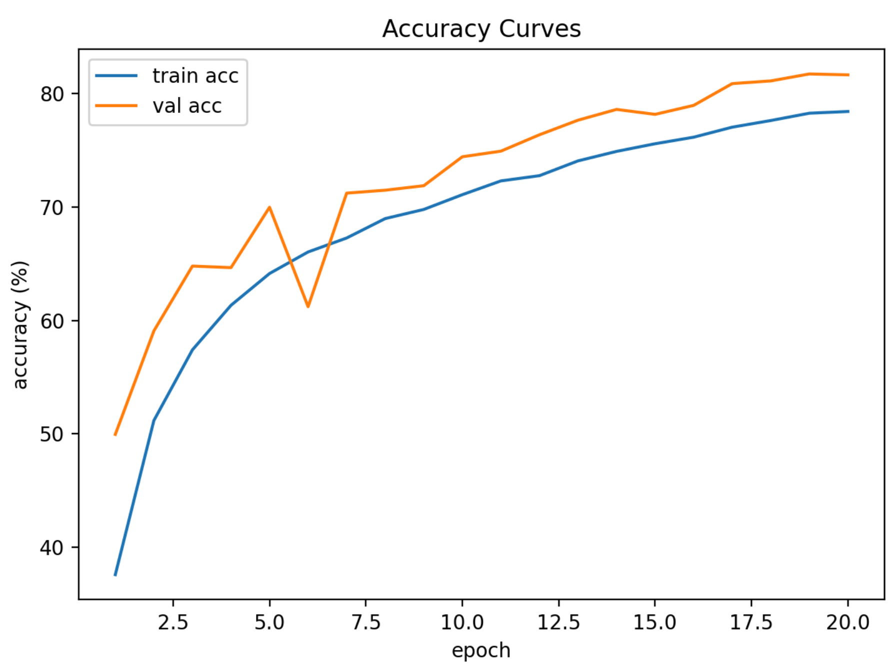
We implemented a custom convolutional neural network (CNN) for CIFAR-10 image classification. The model consists of four convolutional layers organized into two blocks. The first block includes two convolutional layers (3→64 and 64→64), followed by batch normalization, ReLU activations, and a 2×2 max pooling layer. The second block increases the channel depth with two additional convolutional layers (64→128 and 128→128), again followed by batch normalization, ReLU activations, and max pooling. Dropout is applied after pooling to improve generalization and reduce overfitting. The convolutional feature extractor is followed by a fully connected classifier consisting of a linear layer (128×8×8 → 256), a ReLU activation, dropout, and a final linear layer (256 → 10) that outputs class scores for the 10 CIFAR-10 categories. This architecture satisfies the assignment requirements, as it includes more than three convolutional layers, pooling operations, nonlinear activations, and a fully connected classification head.

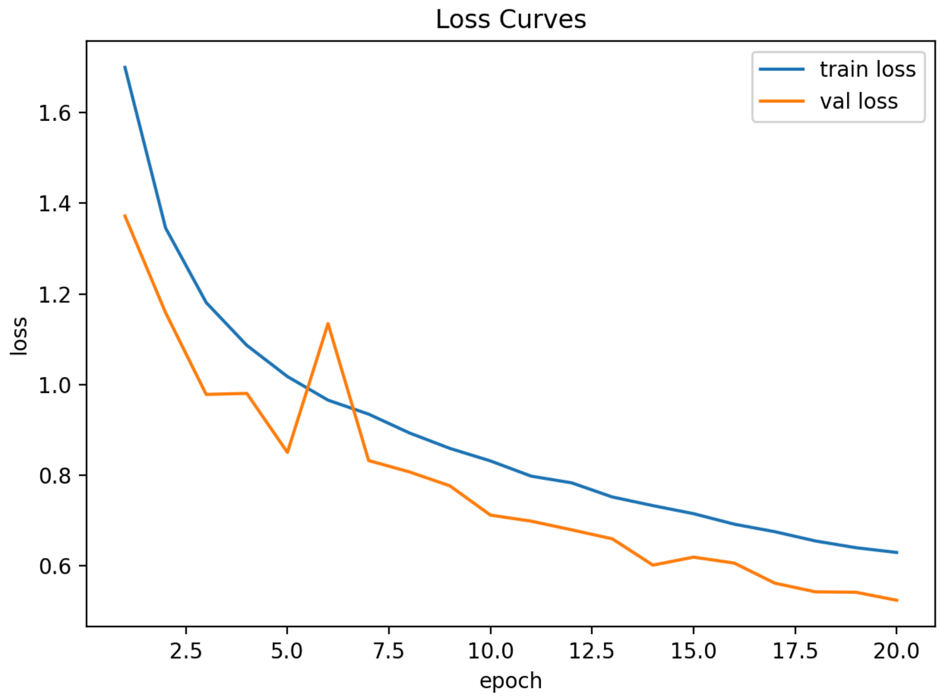
Training Results:

The model was trained for 20 epochs. Training accuracy steadily increased from 37.6% to 78.4%. Validation accuracy improved from 49.9% to 81.7%.

The final test accuracy was:82.82%

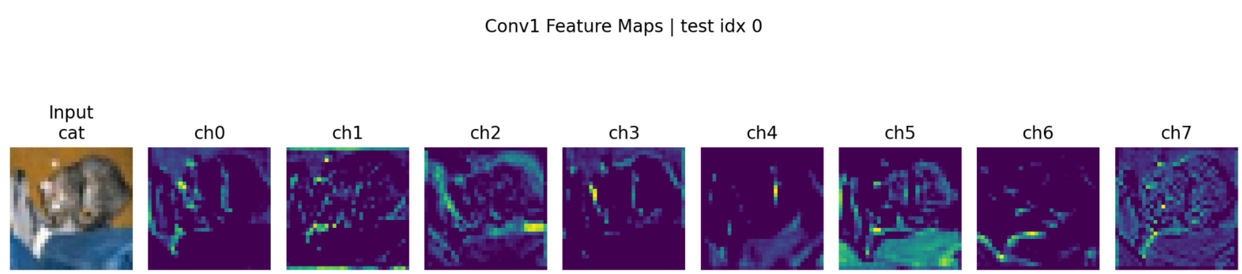
Loss curves show smooth convergence without strong overfitting. Validation accuracy slightly exceeds training accuracy due to dropout and data augmentation during training.



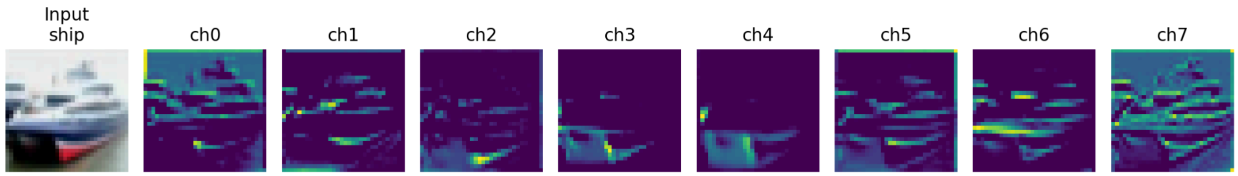


Feature Map Visualization (Early Layer):

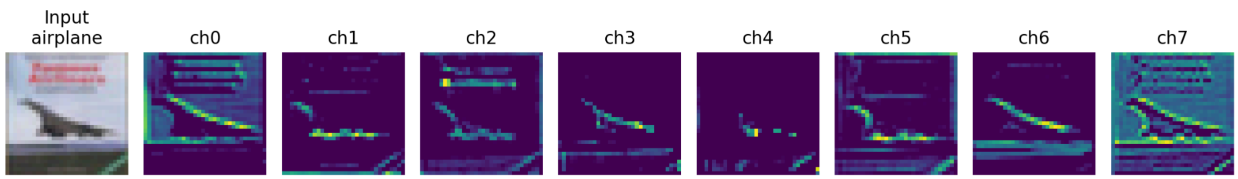
We selected three test images from different classes (ship, cat, and airplane) and passed them through the trained CNN, extracting the feature maps from the first convolutional layer. For each image, we visualized eight channels from this layer to understand what the network learns at an early stage. The feature maps show that different filters respond to different low-level patterns in the input. Some filters clearly detect edges and contours, others respond strongly to color contrasts, and several highlight object boundaries. For example, in the ship image, certain filters activate strongly along horizontal lines corresponding to the water horizon, while in the airplane image, diagonal edge patterns become prominent. In the cat image, curved boundary-like activations appear around the object's outline. These observations indicate that the first convolutional layer learns general low-level features such as edges, textures, and color transitions, which serve as foundational building blocks for deeper, more abstract representations in later layers.



Conv1 Feature Maps | test idx 1



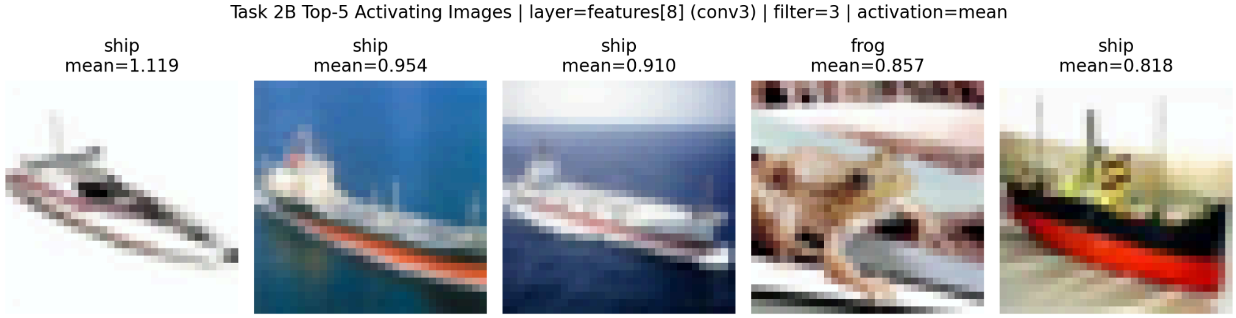
Conv1 Feature Maps | test idx 3

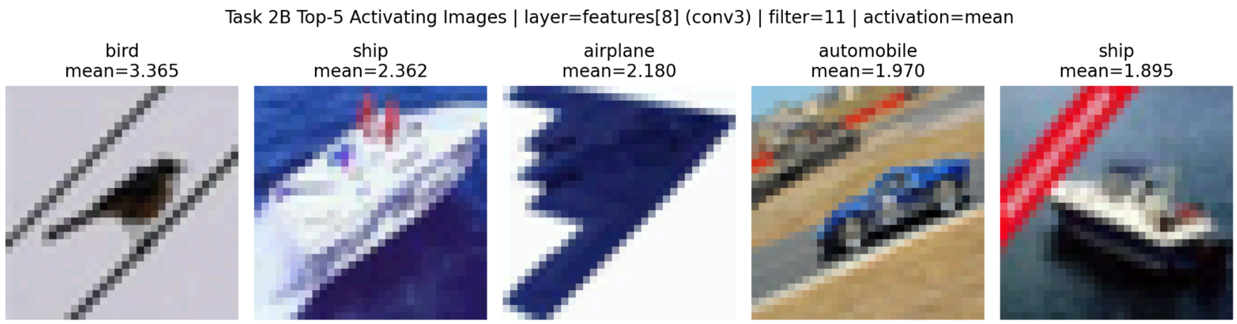


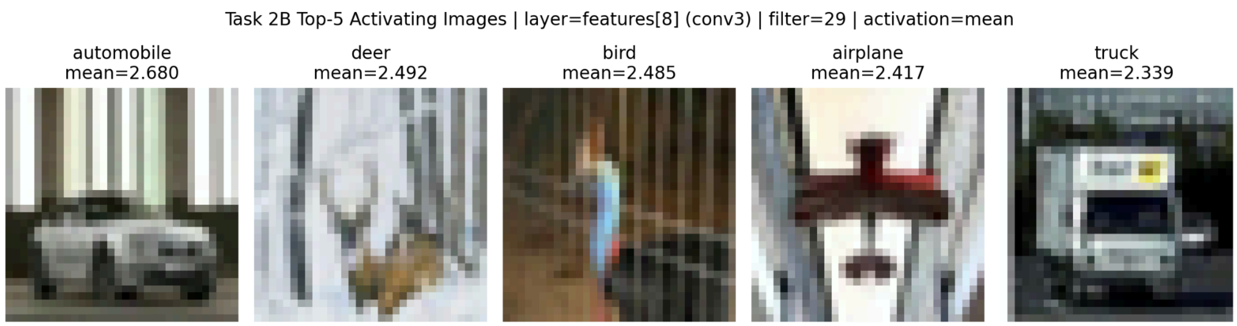
Maximally Activating Images (Task 2B):

To analyze maximally activating images, we selected the convolutional layer `features[8]` (conv3) and examined three filters: 3, 11, and 29. Activation was defined as the mean value of the ReLU feature map for each filter, representing the overall response strength of that filter to a given image. For each selected filter, we identified and visualized the top-5 test images that produced the highest activation values. The results show that Filter 3 responds strongly to ship-like structures and elongated horizontal shapes, suggesting sensitivity to large linear forms. Filter 11 shows high activation for bird and airplane images, indicating that it may be detecting wing-like or diagonal structural patterns. Filter 29 activates frequently on vehicles such as automobiles, trucks, and airplanes, which suggests that it captures rigid object structures or structured edges common in man-made objects. Overall, these filters appear to learn mid-level shape representations and object structure patterns that are more specific than simple edge detectors from early layers, but not yet fully class-specific representations.

Task 2B Top-5 Activating Images | layer=features[8] (conv3) | filter=3 | activation=mean







Discussion and Reflection:

The CNN achieved strong performance, reaching a test accuracy of 82.82%, which shows that even a relatively simple convolutional architecture can effectively classify CIFAR-10 images. From the feature visualizations, it is clear that the early convolutional layers learn basic visual patterns such as edges, textures, and color transitions. As we move deeper into the network, the filters begin capturing more structured and meaningful patterns related to object shape and geometry. The analysis of maximally activating images further confirms that different filters specialize in different types of visual features, such as horizontal structures, diagonal patterns, or rigid object shapes. Overall, this experiment illustrates how convolutional neural networks progressively build hierarchical visual representations, starting from low-level features and moving toward more complex and structured patterns that support final classification decisions.