Comparison and Analysis of CNN and MLP Neural Networks

Vu Viet Thai – B23DCCE085

1 Compare the training process based on the results in the folder *Draw-the-learning-curve*:

- Neuron MLP:
  Training loss and accuracy:
  - MLP starts with an accuracy of $\tilde{4}2\%$, gradually increasing to $\tilde{7}2\%$ after 10 epochs.
  - Validation accuracy also increased similarly, reaching about 76%.
  - Loss decreased from 1.65 to about 0.78.

  Comment:
  - Slow convergence speed.
  - The level of loss is still high at the end of the training process.
  - The training and validation accuracy have a gap, although not big, but sufficient to show that the model has not generalized well.

- Neuron CNN:

  Training loss and accuracy:
  - Starting from an accuracy of $\tilde{5}1\%$, quickly reaching $\tilde{9}5\%$.
  - The validation accuracy is even higher, reaching $\tilde{9}6\%$ at epoch 10.
  - Loss decreases steadily from 1.35 to about 0.15, showing effective learning.

  Comment:
  - Fast convergence speed.
  - No signs of overfitting: training and validation accuracy are similar to each other.
  - The loss level decreases steadily and is low at the end of training.

2 Comparison Based on Confusion Matrix in the folder ***Confusion-matrix***:

- Neuron MLP:

  MLP has significantly lower performance. The majority of samples were incorrectly classified as the class "frog":

  - *Airplane* → *Frog*: 764 times.
  - *Automobile* → *Frog*: 687 times.
  - *Bird* → *Frog*: 809 times.
  - *Cat* → *Frog*: 780 times.

  This reflects that MLP cannot clearly distinguish image features – because MLP processes images as flat vectors, it does not leverage spatial information.

  Easy-to-distinguish classes like *airplane*, *automobile*, and *ship* are still severely confused, which proves that the model cannot learn a strong representation from images.

- Neuron CNN:

  CNN demonstrates very high classification performance across most classes:

  - *Frog*: 670 correct samples / nearly absolute.
  - *Ship*: 616 correct samples.
  - *Truck*: 670 correct.

  However, there are still some confusions:

  - The classes *cat*, *dog*, *bird*, and *deer* often confuse each other due to similar visual features (color, shape).
  - Class *bird* is misclassified as *dog*: 142 times.
  - *Deer* misclassified as *dog*: 186 times.

  Nevertheless, CNN's confusion remains focused and reasonable, reflecting very good feature recognition capability.

3 The cause of the difference:

The significant difference in performance between CNN and MLP can be explained by their architectural characteristics. CNN's use convolution operations to extract local features, pooling layers to reduce size and create translation invariance, along with the ability to learn hierarchical features from low-level to high-level. This is particularly suitable for image data, where spatial relationships and local patterns play a crucial role.

MLP's, with a fully connected architecture, process each pixel as an independent feature without leveraging the spatial structure of the image. This leads to a loss of important information about relative position and spatial relationships between pixels, significantly reducing classification capability.

4 Conclusion:

Aspect	Neuron MLP	Neuron CNN
Architecture	Simple Dense	Modern Convolutional
Accuracy	72% (Training) – 76% (Validation)	95% (Training) – 96% (Validation)
Loss	0.78 (end)	0.15 (end)
Classification Power	Weak (many mistakes)	Strong (very few mistakes)
Learn image features	Not good	Good (extracted in spatial domain)

Comparison of MLP and CNN

This experimental result clearly affirms the advantage of CNN in processing image data. CNN not only achieves higher accuracy but also shows a more stable and efficient learning process. The 20% difference in accuracy (96% vs 76%) is a significant gap, proving that selecting the appropriate network architecture for the type of data is extremely important. For computer vision tasks, CNN is clearly a better choice than MLP, while MLP may be more suitable for other types of data such as tabular data or time series data that do not have a clear spatial structure.