Mohammad Sufyaan
ITAI 1378

1. CNN Architecture

In this project, I worked on classifying images into two categories: Chihuahuas and Muffins using a Convolutional Neural Network (CNN). The architecture of CNN consists of three convolutional layers, followed by fully connected layers, which is quite different from the traditional feedforward neural network (NN) used in my previous tasks.

CNNs are designed to handle image data more efficiently by using convolutional layers that apply filters to detect different patterns in images. Each convolutional layer extracts specific features like edges, textures, and shapes from the input images. The CNN layers used in this project are combined with activation functions such as ReLU to introduce non-linearity, and pooling layers like max pooling to reduce the spatial dimensions, which helps with reducing the number of parameters and computational cost.

In contrast, the traditional neural network from the previous workshop worked with fully connected layers from the start. In that case, each neuron was connected to every pixel of the image, leading to a much higher number of parameters and making it harder for the network to learn spatial hierarchies like a CNN does. CNNs, therefore, are better suited for image classification tasks because they effectively capture spatial dependencies.

The model I used for this task consisted of:

- Three convolutional layers: The first layer had 32 filters, the second 64 filters, and the third 128 filters, all with a 3x3 kernel size.
- Max pooling layers: Applied after each convolution to reduce the dimensionality of the feature maps.
- Fully connected layers: After the convolutional layers, the extracted features were flattened, and then passed through two fully connected layers.

This hierarchical structure allowed the model to progressively learn more complex patterns from the images, which improved classification accuracy.

2. Model Performance

During the training process, the model's performance was measured based on loss and accuracy. After training for 10 epochs, the model achieved a training accuracy of around 95%, and a validation accuracy of approximately 93%. These results were promising but not perfect. Upon examining the predictions and comparing them with the ground truth labels, I noticed that the model made some interesting mistakes.

Mohammad Sufyaan
ITAI 1378

The misclassifications often occurred in images where the Chihuahuas or Muffins had similar shapes or colors, such as when the Chihuahua's fur resembled the texture of the muffin. This revealed a limitation in the model's ability to distinguish between classes with visually similar patterns.

Some of the key takeaways from the model's performance include:

- The model learned to classify most images correctly, but struggled with edge cases where the visual similarity between classes was high.
- Data augmentation techniques like random horizontal flips and rotations likely helped improve the generalization of the model, as it made the training set more varied and robust.
- The final validation accuracy was quite good, but could potentially be improved with more sophisticated techniques like fine-tuning or using pre-trained models.

3. Challenges and Limitations

One of the major challenges in this task was dealing with the small dataset size. With only 120 training samples (65 Chihuahuas and 55 Muffins), the model was at risk of overfitting, which could lead to poor performance on unseen data. To mitigate this, I used data augmentation, which helped simulate a larger and more varied dataset by applying transformations like flips and rotations. However, this alone may not be sufficient to fully address the issue of overfitting in real-world applications.

Another challenge was the time it took to train the model. While CNNs are powerful, they can be computationally expensive to train, especially when using a large number of filters and layers. On my hardware, each epoch took several minutes to complete. Utilizing more advanced hardware like GPUs would significantly speed up the process.

4. Comparison to Previous Neural Network

Compared to the traditional neural network used in the previous workshop, the CNN performed significantly better in terms of both training speed and classification accuracy. The previous neural network struggled with high-dimensional data, and its accuracy was much lower, hovering around 75%. The CNN's ability to automatically learn spatial hierarchies and patterns from the data led to improved performance.

One clear advantage of CNNs is that they require less feature engineering. In the previous neural network, we needed to manually define and preprocess the input features, whereas in CNNs, the convolutional layers automatically extract relevant features from the images.

Mohammad Sufyaan
ITAI 1378

5. Reflections on Future Improvements

While the CNN performed reasonably well, there are several ways the model could be improved. For instance, using a pre-trained model such as VGG16 or ResNet with transfer learning could greatly enhance performance. Pre-trained models have already learned general patterns from large datasets like ImageNet, which can be fine-tuned to work with smaller, specialized datasets like ours.

Additionally, exploring more sophisticated data augmentation techniques or using techniques like mixup or Cutout might help the model learn more robust features. Another potential improvement would be to adjust the model architecture, such as adding more layers or experimenting with different types of layers like batch normalization or dropout, to further enhance generalization.

6. Ethical Considerations

When deploying image classification systems, it's important to consider potential ethical issues. In real-world applications, biases in the training data can lead to biased predictions. In our case, this dataset is fairly simple, but in more complex scenarios, such as facial recognition or medical imaging, biases could have serious implications.

Moreover, misclassifications could lead to harm, depending on the application domain. For example, in medical image analysis, incorrect predictions could have life-threatening consequences. Therefore, it's crucial to evaluate and test models thoroughly before deploying them in critical applications.

Conclusion

Overall, this CNN-based image classification task was an insightful experience that reinforced the importance of using convolutional layers for visual data. I gained a deeper understanding of how CNNs work and how they differ from traditional neural networks. While the model achieved good accuracy, there are still challenges and limitations to address, especially in terms of dataset size and model generalization. Going forward, I would like to explore more advanced techniques like transfer learning and work on more complex image datasets to continue improving my skills.