Convolutional Neural Networks (CNN) Based Bagging Learning for Image Classification Background

Ensemble learning techniques, such as bagging (Bootstrap Aggregating), are widely used in machine learning to improve the performance and robustness of predictive models. Bagging involves training multiple models on random subsets of the dataset and aggregating their predictions to reduce variance and enhance generalization. In this project, we applied bagging with Convolutional Neural Networks (CNNs) to classify the CIFAR-100 dataset. CIFAR-100 is a well-known benchmark dataset containing 60,000 color images divided into 100 classes, with 50,000 training samples and 10,000 test samples. Each image is 32x32 pixels, posing a significant challenge due to its high intra-class variability and low resolution.

Motivation/Challenge

Deep learning models, particularly convolutional neural networks (CNNs), excel at image classification tasks. However, these models often suffer from overfitting when trained on complex datasets like CIFAR-100, especially in scenarios with limited data or insufficient model diversity. Single-model architectures, despite achieving high accuracy, can show significant variability in performance due to biases in data or initialization.

To address these challenges, we aimed to design a robust approach that enhances classification accuracy while reducing overfitting. By employing bagging techniques, we leveraged the collective strength of multiple CNNs trained on diverse subsets of the dataset. This approach allowed us to compare the effectiveness of ensemble methods against traditional single CNN architectures without ensemble.

Methodology

We implemented a bagging ensemble of CNN models trained on CIFAR-100 using the following methodology:

1. Dataset Preparation:

     o CIFAR-100 data was loaded, normalized (scaling pixel values to the range [0, 1]), and one-hot encoded for multiclass classification.

     o Random subsets of the training dataset were created using bootstrap sampling (sampling with replacement), each subset matching the size of the original dataset.

2. Model Architecture:

     o Each CNN model in the ensemble used a standard architecture:

          ▪ Two convolutional layers (32 and 64 filters) with ReLU activation and max-pooling.

- A fully connected dense layer with 64 neurons and a final output layer with 100 neurons for SoftMax-based classification.
  - The models were trained independently on bootstrapped subsets for 10 epochs with a batch size of 64 (full architecture on the last page)

3. Ensemble Prediction:

   - After training, the ensemble combined predictions by averaging the softmax outputs from all models. The result was compared to training and testing CNN without ensembled..

   - Performance was evaluated using accuracy and a confusion matrix on the test dataset.

4. Performance Visualization:

   - Loss and accuracy curves were plotted using the plotnine library to track training and validation performance over epochs.

   - A confusion matrix heatmap visualized the classification performance across all 100 classes, highlighting strengths and weaknesses in specific categories.

Discussion

In our experiments, we observed that the average accuracy of the ensemble dropped significantly to 10.8%, compared to the 29% test accuracy achieved by a single CNN. This result highlights key challenges associated with ensemble methods like bagging when applied to deep learning models on complex datasets. Each model in the ensemble was trained on a subset of the dataset, often through bootstrap sampling. If these subsets were too small or lacked sufficient diversity, individual models tended to overfit to their specific subsets, resulting in poor generalization across the entire dataset. Consequently, the collective performance of the ensemble was hampered, as the individual models failed to effectively capture the broader data distribution. This issue was further compounded when the models in the ensemble were weak, either due to insufficient training epochs or inadequate model complexity, which led to subpar individual performance. When combined, these factors significantly reduced the ensemble's effectiveness compared to the single CNN trained on the full dataset.

Another critical factor contributing to this outcome was the improper weighting of predictions during ensembling. The ensemble utilized simple uniform averaging, which gave equal importance to both strong and weak models. This approach amplified the negative impact of weaker models, introducing noise and incorrect predictions that brought down the

overall accuracy. Additionally, the dataset being split into smaller subsets for each model reduced the amount of training data available, limiting the ability of individual models to learn robust features. In contrast, the single CNN, trained on the full dataset, could leverage the complete data distribution and benefit from techniques like data augmentation, resulting in a much higher accuracy of 29%. This comparison underscores the importance of addressing limitations in data diversity, model training, and ensemble weighting strategies to achieve meaningful improvements in ensemble-based approaches for deep learning.

Conclusion

This study explored the effectiveness of employing bagging ensemble methods with Convolutional Neural Networks (CNNs) for image classification tasks on complex datasets. Our findings revealed significant challenges in leveraging bagging for CNN-based learning, particularly when compared to a single CNN trained on the full dataset. While single CNN models achieved a test accuracy of 29%, the ensemble of CNNs trained on diverse subsets using bagging achieved an average accuracy of only 10.8%. This stark contrast highlights the limitations of the bagging approach in this context.

The observed performance gap can be attributed to several factors. Overfitting of individual models to small, less diverse subsets, combined with insufficient training and improper weighting of predictions, significantly impacted the generalization capability of the ensemble. Additionally, the splitting of the dataset for individual models reduced the data available for training, hindering their ability to learn robust features. In contrast, a single CNN trained on the full dataset benefited from data augmentation and access to the complete data distribution, leading to superior performance. These findings emphasize the importance of addressing issues such as subset diversity, model robustness, and ensemble weighting to improve the effectiveness of CNN-based bagging for image classification. Future research should focus on optimizing subset selection, employing dynamic weighting strategies, and integrating advanced data augmentation techniques to fully harness the potential of ensemble learning in deep learning contexts.