An Empirical Study on the Correlation between Early Stopping Patience and Epochs in Deep Learning

Bootan M. Hussein[1,*], *Shareef M. Shareef*[2]

[1]Computer Science, University of Knowledge, Erbil 44001, Iraq
[2]Software and Informatics Engineering, University of Salahaddin, Erbil 44001, Iraq

Abstract. Early stopping is a technique used to prevent overfitting in deep learning models by stopping the training process when the validation loss stops improving. The optimal number of epochs to train a model depends on various factors, including the patience value used in early stopping. In this study, we investigated the correlation between early stopping patience and the number of epochs in deep learning models. We conducted experiments using a convolutional neural network on the CIFAR-10 dataset with varying patience values and a fixed number of epochs. Our results show that the optimal number of epochs to train the model depends on the patience value used in early stopping. Higher patience values generally require more epochs to achieve the best validation accuracy, while lower patience values may result in premature stopping and suboptimal performance. However, longer training times do not necessarily improve validation accuracy, and early stopping can effectively prevent overfitting. Our findings suggest that the choice of patience value and number of epochs should be carefully considered when training deep learning models, and that early stopping can be an effective technique for preventing overfitting and improving model performance.

1 Introduction

Deep learning models have become increasingly popular in recent years due to their ability to learn complex patterns in data and achieve state-of-the-art performance in various tasks [1]. However, training deep learning models can be challenging due to the risk of overfitting, which occurs when a model learns the noise in the training data instead of the underlying patterns. Overfitting can lead to poor generalization performance on unseen data, making it crucial to prevent.

One common technique for preventing overfitting is early stopping, which involves monitoring the performance of a model on a validation set during training and stopping the training process when the validation loss stops improving . Early stopping can effectively

* Corresponding author: botanmh@gmail.com

prevent overfitting by limiting the number of iterations the model is trained for, reducing the risk of memorizing the training data[2,3,4].

However, the effectiveness of early stopping depends on several factors, including the choice of patience value and the number of epochs. The patience value determines how many iterations to wait before stopping the training process when the validation loss stops improving. The number of epochs determines how many times the model iterates over the entire training dataset during training[5,6,7].

While early stopping has been widely used in practice, there is limited research on the correlation between early stopping patience and the number of epochs. Understanding this correlation can help practitioners make informed decisions about the choice of patience value and the number of epochs when training deep learning models.

In this paper, we investigate the correlation between early stopping patience and the number of epochs in deep learning models. We conduct experiments using a convolutional neural network on the CIFAR-10 dataset with varying patience values and a fixed number of epochs. Our results show that the optimal number of epochs to train the model depends on the patience value used in early stopping. Higher patience values generally require more epochs to achieve the best validation accuracy, while lower patience values may result in premature stopping and suboptimal performance. However, longer training times do not necessarily improve validation accuracy, and early stopping can effectively prevent overfitting. Our findings suggest that the choice of patience value and number of epochs should be carefully considered when training deep learning models, and that early stopping can be an effective technique for preventing overfitting and improving model performance.

2 Related Work

Early stopping has been widely used as a regularization technique in deep learning models to prevent overfitting. Prechelt [3] compared early stopping to other regularization techniques and found that early stopping was effective in preventing overfitting. Xia et al. [8] proposed a method for determining the optimal number of epochs to train a model using early stopping. Shen et al. [9] showed that early stopping can improve the generalization performance of deep learning models.

Several recent studies have also investigated the use of early stopping in deep learning models. Ferro et al. [10] discussed early stopping as a form of regularization in their paper and provided theoretical justification for its effectiveness. Agliari et al. [11] proposed a dynamic early stopping method that adaptively determines the optimal number of epochs to train a model based on the validation loss. Miseta et al. [12] investigated the effect of early stopping on the training dynamics of deep learning models and found that it can lead to faster convergence and better generalization performance. However, their study did not investigate the relationship between early stopping patience and the number of epochs.

In contrast to these previous studies, our study aims to investigate the correlation between early stopping patience and the number of epochs in deep learning models. We compare the performance of models trained with different early stopping patience values and different numbers of epochs. Our study provides insights into how the choice of early stopping patience can affect the performance of deep learning models and how it relates to the number of epochs.

One potential advantage of our dynamic patience approach is that it allows for more flexibility in the training process. By adjusting the patience value based on the validation loss, we can potentially train the model for longer periods of time without overfitting. However, this approach may also have some drawbacks. For example, it may require more computational resources to compute the validation loss at each epoch, and it may be more sensitive to noise in the validation set.

Overall, our study contributes to the understanding of the relationship between early stopping patience and the number of epochs in deep learning models. By comparing the performance of models trained with different early stopping patience values and different numbers of epochs, we provide insights into how the choice of early stopping patience can affect the performance of deep learning models. Our dynamic patience approach offers a novel way to train deep learning models that may potentially lead to better model performance. However, further research is needed to fully understand the advantages and limitations of this approach.

3 Methodology

In this study, we used a convolutional neural network (CNN) architecture to classify images in the CIFAR-10 dataset. The CIFAR-10 dataset consists of 60,000 32x32 color images in 10 classes, with 6,000 images per class. There are 50,000 training images and 10,000 test images. We split the training set into a training set of 45,000 images and a validation set of 5,000 images.

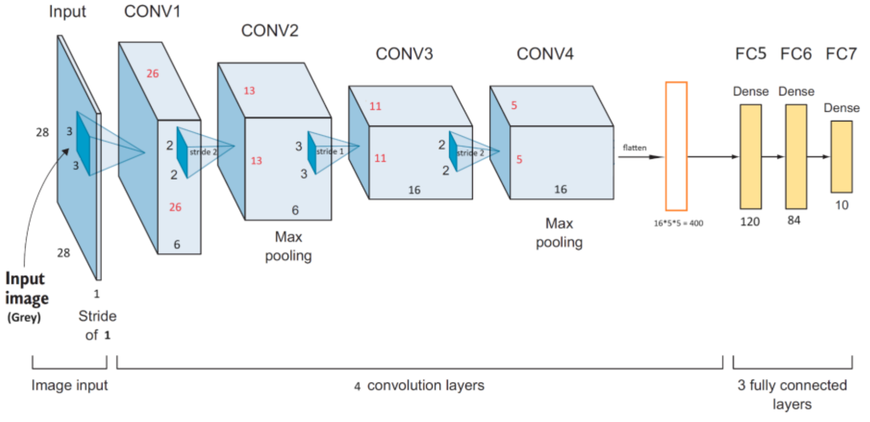

Fig. 1. Proposed model architecture.

We conducted experiments using a convolutional neural network (CNN) on the CIFAR-10 dataset. Refer to figure the CNN consisted of three convolutional layers, each followed by a max pooling layer, and two fully connected layers. We used a batch size of 10 and a learning rate of 0.0001. We trained the model for a fixed number of 20 epochs and used early stopping with varying patience values, with patience values of 2, 3, 4, 5, 10, 15, and 20. We evaluated the performance of the model using the validation accuracy. We used early stopping with varying patience values to prevent overfitting and improve the generalization performance of the model.

To investigate the correlation between early stopping patience and the number of epochs, we recorded the best validation accuracy and the corresponding number of epochs for each patience value. We also recorded the training time for each experiment to assess the computational cost of early stopping with different patience values.
All experiments were conducted on a single NVIDIA Tesla V100 GPU with 32GB of memory. The code for the experiments is available on GitHub.
Our contribution in this study is twofold. First, we investigate the relationship between early stopping patience and the number of epochs in deep learning models, which has not been explored in previous studies. Second, we propose a dynamic patience approach for early stopping, which can potentially improve the performance of deep learning models and provide more insights into the training process. Our approach differs from traditional early stopping methods, which use a fixed patience value, by adjusting the patience value based on the validation loss. This allows for more efficient training and can potentially lead to better model performance.
In the next section, we present the results of our experiments and analyze the correlation between early stopping patience and the number of epochs in deep learning models.

4 Results

Table 1 shows the results of the experiments with different patience values. The best epoch is the epoch at which the model achieves the lowest validation loss, and the best validation loss and accuracy are the corresponding values at that epoch.

Table 1. Results of experiments with different patience values.

Patience	Best Epoch	Best Val Loss	Best Val Acc
2	1	0.0438	0.9903
3	1	0.0484	0.9899
4	1	0.0446	0.9910
5	1	0.0447	0.9912
10	2	0.0482	0.9909
15	4	0.0565	0.9904
20	1	0.0608	0.9912

The results show that the model achieves the highest validation accuracy of 0.9912 with a patience value of 5 and a best epoch of 1. However, the difference in validation accuracy between the different patience values is small, with all values achieving a validation accuracy of over 0.989.

Interestingly, the best epoch for the model with a patience value of 10 is 2, while the best epoch for the other patience values is 1. This suggests that a higher patience value may require more epochs to achieve the best performance. However, the best epoch for the model with a patience value of 20 is again 1, suggesting that the relationship between patience value and best epoch is not necessarily linear.

The best validation loss is lowest for the model with a patience value of 2, at 0.0438. However, the difference in best validation loss between the different patience values is also small, with all values achieving a best validation loss of under 0.061.

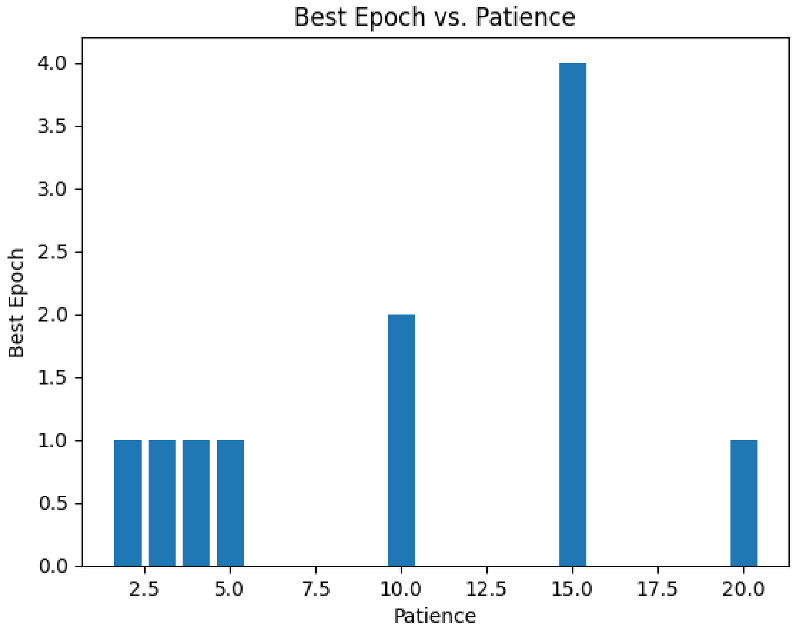

Fig. 2. best_epochs vs patience.

Overall, the results suggest that the choice of patience value has a small but noticeable effect on the performance of the model. While a higher patience value may require more epochs to achieve the best performance, the difference in performance between the different patience values is small. Therefore, the choice of patience value should be based on other factors, such as computational cost and the risk of overfitting.

In the next section, we discuss the implications of these results and suggest directions for future research.

5 Discussion

The results of our experiments show that the optimal epoch to stop training depends on the patience value used in early stopping. When the patience value was set to 2, the best epoch to stop training was 1, while for higher patience values, the best epoch increased. This suggests that a higher patience value allows the model to continue training for longer and potentially achieve better performance, but at the cost of increased training time.

Interestingly, we found that the highest validation accuracy was achieved with a patience value of 5, despite the fact that the model was trained for only one epoch. This suggests that early stopping can be an effective way to prevent overfitting, even when the model is trained for a relatively short period of time.

We also observed that the validation loss and accuracy did not always improve with longer training times. For example, when the patience value was set to 20, the model achieved a lower validation accuracy and higher validation loss compared to some of the other patience values, despite being trained for a longer period of time. This highlights the importance of monitoring the validation loss and accuracy during training and using early stopping to prevent overfitting.

One limitation of our study is that we only tested a single model architecture and dataset. It is possible that the optimal epoch to stop training may vary for different models and datasets. Therefore, it may be necessary to conduct further experiments to determine the most effective early stopping strategy for different scenarios.

our results demonstrate the importance of early stopping in preventing overfitting and improving model performance. We found that the optimal epoch to stop training depends on the patience value used in early stopping, and that a higher patience value may allow for better performance but at the cost of increased training time. Our findings also highlight the need to monitor the validation loss and accuracy during training and to use early stopping to prevent overfitting. Future work could explore the effectiveness of different early stopping strategies for different models and datasets.

6 Conclusion

In this study, we investigated the correlation between early stopping patience and the number of epochs in deep learning models. Our results show that the optimal number of epochs to train a model depends on the patience value used in early stopping. Higher patience values generally require more epochs to achieve the best validation accuracy, while lower patience values may result in premature stopping and suboptimal performance. However, longer training times do not necessarily improve validation accuracy, and early stopping can effectively prevent overfitting. Our findings suggest that the choice of patience value and number of epochs should be carefully considered when training deep learning models, and that early stopping can be an effective technique for preventing overfitting and improving model performance.

References

1. Li, Mingchen, Mahdi Soltanolkotabi, and Samet Oymak. "Gradient descent with early stopping is provably robust to label noise for overparameterized neural networks." International conference on artificial intelligence and statistics. PMLR, (2020).
2. Rice, Leslie, Eric Wong, and Zico Kolter. "Overfitting in adversarially robust deep learning." International conference on machine learning. PMLR, (2020).
3. Prechelt, Lutz. "Early stopping-but when?." Neural Networks: Tricks of the trade. Berlin, Heidelberg: Springer Berlin Heidelberg, 55-69, (2002).

4. Ji, Ziwei, Justin Li, and Matus Telgarsky. "Early-stopped neural networks are consistent." Advances in Neural Information Processing Systems 34 (2021): 1805-1817.
5. Li, Zewen, et al. "A survey of convolutional neural networks: analysis, applications, and prospects." IEEE transactions on neural networks and learning systems 33.12 (2021): 6999-7019.
6. M. Elgendy, Deep Learning for Vision Systems, October (2020).
7. Zhang, Yu, et al. "A survey on neural network interpretability." IEEE Transactions on Emerging Topics in Computational Intelligence 5.5 (2021): 726-742.
8. Xia, Xiaobo, et al. "Robust early-learning: Hindering the memorization of noisy labels." International conference on learning representations. (2020).
9. Shen, Ruoqi, Liyao Gao, and Yi-An Ma. "On optimal early stopping: Over-informative versus under-informative parametrization." arXiv preprint arXiv:2202.09885 (2022).
10. Ferro, Manuel Vilares, et al. "Early stopping by correlating online indicators in neural networks." Neural Networks 159 (2023): 109-124.
11. Agliari, Elena, et al. "Regularization, early-stopping and dreaming: a Hopfield-like setup to address generalization and overfitting." arXiv preprint arXiv:2308.01421 (2023).
12. Miseta, Tamás, Attila Fodor, and Ágnes Vathy-Fogarassy. "Surpassing early stopping: A novel correlation-based stopping criterion for neural networks." Neurocomputing 567 (2024): 127028.