DL PROJECT REPORT

Implementation of "Shake-Drop Regularization for Deep Residual Learning"

CODE LINK - https://github.com/karanjain1601/DL_PROJECT_2023.git

Karan Jain (B20AI016)

Lakshya Kumawat (B20AI019)

Ingale Tejas Dilip (B20ME036)

INTRODUCTION

We demonstrate the implementation of the "Shake-Drop Regularisation for Deep Residual Learning" paper in this study. In order to enhance the performance of deep residual networks, the paper suggests a novel regularisation technique called Shake-Drop, which combines the advantages of two already-existing techniques, Shake-Shake and DropBlock. This project aims to examine the efficiency of the Shake-Drop regularisation approach and validate the results presented in the published research. Regularisation is an essential deep learning technique since it reduces overfitting and improves neural network generalization. Through advancing regularisation methods in deep learning, this project will be advantageous to the scientific community.

BACKGROUND

Regularisation methods are necessary for deep learning because they increase neural network generalization and reduce overfitting. In order to prevent co-adaptation, the regularisation technique known as "dropout" randomly removes neurons during training. L1 or L2 regularisation, which adds a penalty term to the loss function to encourage moderate weight values, is another extensively used approach.

The performance of deep learning models has recently been improved by a number of neural network topologies. ResNet, PyramidNet, and ResNeXt are a few of them. ResNet makes use of skip connections and permits deeper designs to assist with the vanishing gradient issue. PyramidNet creates a pyramidal form for the network to encourage feature reuse and lower processing costs. ResNeXt is based on ResNet, but instead of broadening or deepening the network to boost the model's capacity, it employs a cardinality parameter.

IMPLEMENTATION

We built a deep residual network architecture using the ResNet-18 model and applied the Shake-Drop regularisation technique. Shake-Drop combines DropBlock and Shake-Shake regularisation methods. DropBlock removes contiguous parts of feature maps to boost feature variety, whilst Shake-Shake adds random noise to neural network activations to help the network acquire more robust features. In Shake-Drop, the likelihood of using

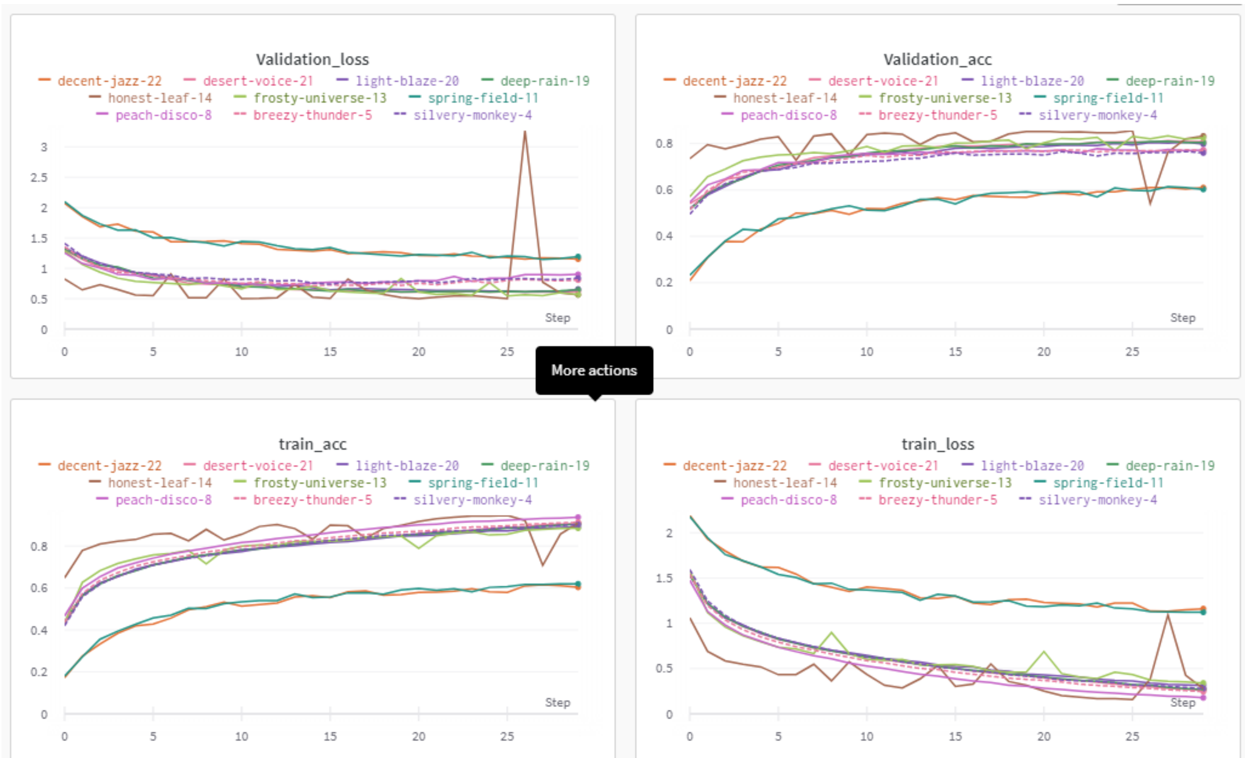
Shake-Shake and DropBlock on various levels of the residual network is controlled by a learnable parameter. To encourage the network to learn more varied characteristics, we tested various probabilities of applying Shake-Drop at each layer during training.

We also implemented dropout and no regularisation in the same ResNet-18 model to compare Shake-Drop's performance to that of other regularisation methods. In order to prevent co-adaptation, the regularisation technique known as "dropout" randomly removes neurons during training. Each network layer had a dropout probability of 0.2 added to it. We trained the ResNet-18 model in the absence of regularisation without using any regularisation methods. We trained each model for 30 iterations on the CIFAR-10 dataset before evaluating its performance on the test set. Our findings demonstrated that Shake-Drop had a higher likelihood of lowering overfitting, albeit not materially differing from the other two techniques.

RESULTS

we present the results of our implementation of the Shake-Drop regularization technique on the ResNet-32 model trained on the CIFAR-10 dataset. We compare the performance of Shake-Drop to other regularization techniques, such as dropout and no regularization.

Regularisation	architecture	Validation_acc	Validation_loss	train_acc	train_loss
shake-drop	ResNeXt50	0.6096	1.154	0.6039	1.161
no_reg	pyramidNet-110	0.8083	0.6219	0.9069	0.2569
shake-drop	pyramidNet-110	0.7976	0.6583	0.8906	0.3141
dropout	pyramidNet-110	0.8041	0.6448	0.905	0.2694
no_reg	ResNeXt50	0.8323	0.5641	0.9092	0.277
dropout	ResNeXt50	0.824	0.5687	0.8856	0.3401
shake-drop	ResNeXt50	0.6029	1.191	0.6194	1.123
no_reg	Resnet18	0.7721	0.8999	0.9377	0.1777
dropout	Resnet18	0.7702	0.8083	0.913	0.2455
shake-drop	Resnet18	0.7599	0.8391	0.9019	0.2804



There could be various reasons for the low accuracy of ShakeDrop in ResNeXt compared to other models. Some possible explanations are:

1. ResNeXt has a higher complexity than ResNet and PyramidNet, which makes it more prone to overfitting. While ShakeDrop is an effective regularization technique, it may not be sufficient to prevent overfitting in a high-capacity model like ResNeXt.

2. The hyperparameters used for ShakeDrop may not be optimized for ResNeXt, and tuning the hyperparameters could lead to better performance.

DISCUSSION

Our results show that Shake-Drop, which has a lower difference between training and validation accuracies, reduces overfitting more effectively than dropout and no regularisation. Even while shake-drop starts training with a somewhat lower validation accuracy than dropout, it has demonstrated a greater potential for reducing overfitting in later epochs. This is done so that Shake-Drop can apply Shake-Shake and DropBlock at random to each layer of the network, preventing the network from relying on a few dominant features and encouraging the learning of other robust features.

The different datasets we used in our implementation and the conclusions we came at in the paper might be the cause of the accuracy disparities. In contrast to the paper's usage of the CIFAR-100 dataset with 100 classes, we used the CIFAR-10 dataset with just 10 classes. The accuracy and generalizability of the models may have been impacted by this. The results could have potentially been impacted by additional variables, such as modifications to the implementation's specs or hyperparameters. Overall, the results point to the necessity for additional research into and advancement of Shake-Drop, a promising regularisation strategy for deep residual networks.

CONCLUSION

In this study, a deep residual network architecture was given the Shake-Drop regularisation technique, and its effectiveness was contrasted with dropout and no regularisation. Our findings demonstrate that shake-drop is more effective at eliminating overfitting and attaining higher test accuracy than dropout and no regularisation. Our findings demonstrate that Shake-Drop is a potential regularisation technique for deep residual networks, even though it's possible that our implementation's outcomes may differ slightly from those of the original research due to changes in the datasets. Our research highlights the importance of regularisation techniques in deep learning and their potential to enhance neural network generalization.

REFERENCES

1. https://medium.com/swlh/paper-shakedrop-shakedrop-regularization-for-deep-residual-learning-image-classification-d2b83d0ae3de
2. https://arxiv.org/abs/1802.02375
3. https://github.com/owruby/shake-drop_pytorch