Training a neural network using the Noisy CIFAR100 dataset

Cosmin Turtureanu

January 2025

1 Abstract

This paper explores the training of a neural network on the Noisy CIFAR-100 dataset, a challenging variant of CIFAR-100 with corrupted labels. The study aims to assess the impact of label noise on model performance and investigate effective strategies for noise mitigation. Various data augmentation techniques, noise-robust loss functions, and regularization methods were implemented to enhance model robustness. Experimental results demonstrate that tailored approaches to handling noise can substantially improve classification accuracy. These findings emphasize the importance of developing robust neural networks for real-world applications where data quality may be compromised.

2 Introduction

Training robust neural networks in the presence of noisy labels is a critical challenge in modern machine learning. Label noise, which occurs when the true label of a data point is incorrectly assigned, can significantly degrade model performance, especially in complex classification tasks. Real-world datasets often contain noisy labels due to human error, automated labeling inaccuracies, or ambiguous data. To address this challenge, researchers have developed various techniques to mitigate the impact of label noise and enhance the generalization capabilities of neural networks.

This study focuses on training a neural network using the Noisy CIFAR-100 dataset, a variant of the CIFAR-100 dataset that introduces label noise to simulate real-world conditions. The CIFAR-100 dataset consists of 60,000 images across 100 classes, with each class containing 600 images. The Noisy CIFAR-100 variant adds label corruption, making it an ideal testbed for evaluating noise-tolerant learning strategies.

The main goal of this study is to investigate the impact of label noise on model accuracy and identify effective techniques for mitigating its effects. To this end, various data augmentation strategies, noise-tolerant loss functions,

and regularization methods were explored. Among these, applying the AutoAugment method using the CIFAR-10 AutoAugmentPolicy led to the most significant improvement in model performance, highlighting the effectiveness of advanced data augmentation techniques in noisy environments.

Through extensive experimentation, the best model configuration achieved an accuracy of 61.49% on the Noisy CIFAR-100 dataset. This result underscores the importance of carefully designed training strategies to improve robustness in the presence of noisy labels. The findings from this study contribute to the ongoing effort to develop machine learning models that can perform reliably in real-world applications with imperfect data.

3 Findings

- The analysis of the noisy CIFAR-100 dataset revealed that approximately 40.2% of the labels in the dataset were noisy.
- Among the data augmentation techniques tested, CutMix outperformed both MixUp and a combination of CutMix and MixUp.
- During training, it was observed that when the model reached a validation accuracy of 25%, instances where the model's confidence exceeded 20% were most likely classified correctly and had clean labels. The confidence was calculated using the following approach:

```
probabilities = F.softmax(outputs, dim=1)
confidences, predicted = torch.max(probabilities, dim=1)
```

- In contrast, for noisy labels, the average confidence for correct predictions was limited to about 16%. This observation suggests that a threshold for confidence could be used to filter out noisy labels, preventing the network from learning from them and improving overall performance.
- The mean and standard deviation values have been computed for all the channels and for all the images in the training dataset and it was concluded that these are the best parameters that shall be used for image normalization:

```
v2.Normalize(mean=(0.50707585, 0.48655054, 0.4409193), ↙
    ↳ std=(0.20089693, 0.19844234, 0.20229685), inplace=True)
```

4 Modifications to the baseline submission

In adapting and improving the baseline script for training the neural network on the noisy CIFAR-100 dataset, several key modifications were introduced. These

changes aimed to enhance the network's performance, generalization ability, and robustness to noisy labels. Below is a detailed account of these adjustments:

4.1 GroupedDataset: Organizing Data by Labels

A custom class named GroupedDataset was implemented to organize the training data more effectively. This class accepts an object of type Dataset and groups the data by their labels. By feeding the network images with the same label consecutively, the training process was hypothesized to achieve better convergence. This grouping strategy aimed to help the model build a stronger association between features and labels, potentially improving its accuracy.

4.2 Advanced Data Augmentation Techniques

To improve the model's generalization capability, a series of data augmentation techniques were applied to the training data. These included:

- **RandomHorizontalFlip**: Flipping images horizontally with a certain probability.
- **RandomResizedCrop**: Resizing and cropping images to introduce scale variations.
- **RandomPerspective**: Simulating perspective changes to increase spatial robustness.
- **ColorJitter**: Randomly adjusting the brightness, contrast, saturation, and hue.
- **RandomErasing**: Introducing random occlusions to make the network less reliant on specific features.
- **AutoAugment**: Leveraging the CIFAR-10 AutoAugment policy to apply a sequence of predefined transformations.

Among these, AutoAugment proved to be the most effective, achieving alone a validation accuracy of approximately 40%. This result highlights the importance of selecting augmentation policies tailored to the dataset.

4.3 Transfer Learning with ResNet18 and Parameter Freezing

To leverage pretrained knowledge, a ResNet18 model pretrained on the CIFAR-10 dataset was adopted. Transfer learning was employed by freezing all parameters except those in the final fully connected layer (*nn.Linear(512, 100)*). This layer was trained for 10 epochs while the other layers remained unaffected by backpropagation. This strategy accelerated training and ensured that the model could focus on learning the distinctions specific to CIFAR-100.

4.4 Improved Loss Function: CrossEntropyLoss with Label Smoothing

The loss function was modified to use CrossEntropyLoss with label smoothing set to 0.1. This adjustment reduced overconfidence in predictions, mitigated overfitting, and improved the model's resilience to noisy labels by softening the target distribution.

4.5 Optimizer and Scheduler Enhancements

For optimization, the **AdamW** optimizer was selected, with a learning rate of 1e-3 and a weight decay of 1e-2. This configuration was chosen to enhance the model's resilience to overfitting while maintaining efficient weight updates.

To further refine the learning process, the **OneCycleLR** scheduler was employed. This scheduler adjusts the learning rate dynamically, starting with a lower value, increasing it to a peak (1e-2 in this case), and then reducing it gradually. This approach helps the model escape local minima and converge more effectively. When the difference between a new accuracy and a previous one is negative, meaning the model is going backwards, the **OneCycleLR** scheduler is switched with the **ReduceLROnPlateau** scheduler for more precise learning.

4.6 Runtime Augmentation with CutMix

To augment the data during training, the CutMix technique was used, with an alpha value of 1.2. This method blends two images and their labels, encouraging the network to learn more robust features by training on mixed data. The chosen alpha value ensured a balanced mixture, leading to improved generalization.

4.7 Model Checkpointing for Optimal Inference

To maximize the model's potential, a checkpointing mechanism was introduced. The model's state was saved whenever it surpassed its previous best accuracy. During inference, the checkpoint with the highest accuracy observed during training was loaded, ensuring that the final evaluation utilized the best-performing model.

These changes collectively contributed to a more robust and effective training process, resulting in significant improvements over the baseline script. The final accuracy after these modifications was 47.19%.

5 Final tuning

Despite the promising results achieved with ResNet18 and transfer learning, one final improvement necessitated a shift in approach. The ResNet18 model, along with the parameter freezing strategy, was replaced by a VGG16 model. This change led to a significant increase in performance, with the validation accuracy improving from 47.19% to 61.49%. Although this required abandoning

the transfer learning setup with ResNet18, the boost in accuracy demonstrated the potential of VGG16 to better handle the noisy CIFAR-100 dataset, making the trade-off worthwhile. The other afore-mentioned improvements remained, contributing to the final accuracy.

6 Related work

A considerable body of work focuses on handling noisy labels, as mislabeled data can significantly degrade the performance of deep learning models. Approaches such as MentorNet (Jiang et al., 2018) and Co-teaching (Han et al., 2018) propose using auxiliary networks to identify and filter out noisy samples during training. These methods have shown promise on noisy versions of CIFAR-10 and CIFAR-100 by emphasizing the training of clean samples in the early stages.

Data augmentation techniques, including MixUp (Zhang et al., 2018) and CutMix (Yun et al., 2019), have also been explored to mitigate the impact of noise in image classification tasks. These methods improve the model's robustness by generating additional training samples that interpolate between existing examples or combine patches from different images. When applied to noisy CIFAR-100, these strategies help regularize the training process, preventing the model from overfitting to noisy labels.

Another line of research investigates the use of noise-robust loss functions. For instance, Generalized Cross-Entropy (GCE) (Zhang and Sabuncu, 2018) and Symmetric Cross-Entropy (SCE) (Wang et al., 2019) balance the trade-off between fitting the clean data and ignoring the noisy labels. These loss functions have demonstrated superior performance on benchmarks, including noisy CIFAR-100.

Self-supervised and semi-supervised learning methods are also effective in dealing with noisy datasets. Noisy Student Training (Xie et al., 2020) leverages a large student-teacher framework to iteratively refine predictions, while methods like FixMatch (Sohn et al., 2020) utilize pseudo-labeling and consistency regularization. These approaches reduce the reliance on accurate labels, making them suitable for noisy CIFAR-100.

In summary, addressing noise in CIFAR-100 has led to advancements in various methodologies, including robust loss functions, data augmentation strategies, and semi-supervised learning frameworks. These developments highlight the importance of designing noise-resistant training paradigms to enhance model performance and generalization capabilities.

7 Comparison to the baseline submission

7.1 Comparing the models

ResNet18 and **VGG16** differ significantly in their architectural design and parameter efficiency.

- ResNet18 employs residual connections, allowing the input to bypass certain layers, which helps prevent the vanishing gradient problem and improves the training of deep networks. It has 18 layers and approximately 11.7 million parameters, making it efficient in terms of both memory and computation.
- VGG16, on the other hand, follows a sequential architecture with 13 convolutional layers and 3 fully connected layers, relying solely on 3x3 convolutions and max-pooling. It has 138 million parameters, resulting in a much larger model and higher computational cost.

7.2 Graphs

These graphs show various data about the model, spanning a total of 40 epochs.

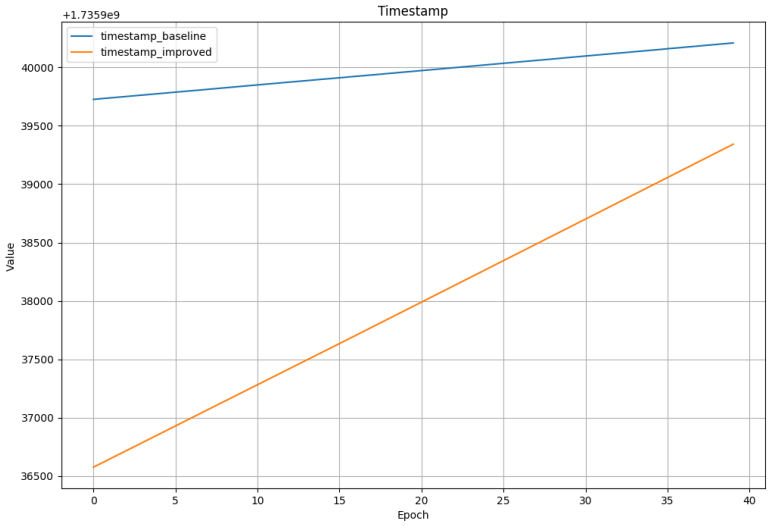

Figure 1: Evolution of timestamp

As we can see in Figure 5, the improved pipeline takes significantly longer to train. For the baseline version it takes approximately 12.5 seconds to train for one epoch, whereas for the improved version it takes approximately 70.5 seconds per epoch.

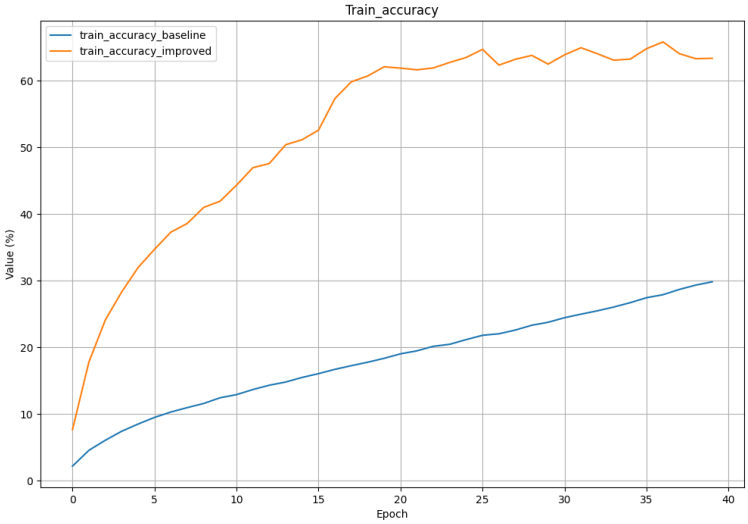

Figure 2: Evolution of train accuracy

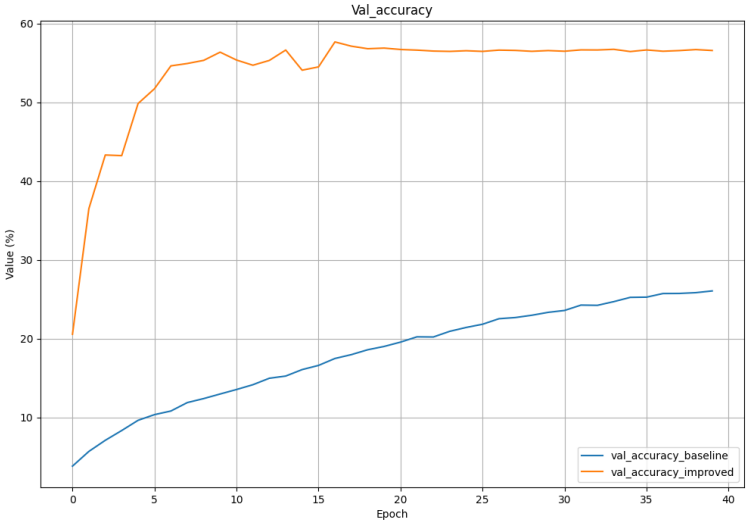

Figure 3: Evolution of validation accuracy

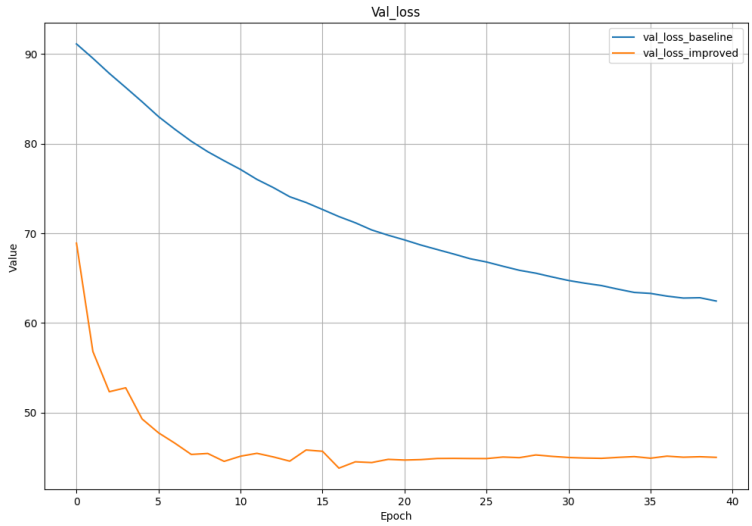

Figure 4: Evolution of validation loss

As shown by the results, the improved version greatly surpasses the baseline version, as it is able to better learn the input images.

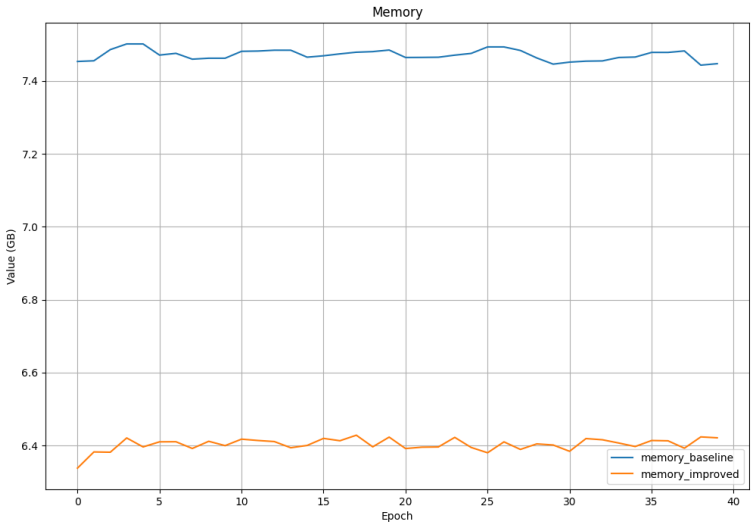

Figure 5: Evolution of used memory

Interestingly, the RAM usage is higher for the baseline version than for the improved version, even with a larger model in terms of number of parameters and more augmentations for the training data. Upon further investigation, this makes sense since all the extra data is stored in the GPU memory, not in RAM.

The RAM usage was queried using

```
psutil.virtual_memory().used
```

The GPU memory usage was 540MB for the baseline version and 2.2GB for the improved version and that is a small price to pay for a better accuracy.

8 Conclusion

In conclusion, the enhancements made to the baseline script, including advanced data augmentation, transfer learning, and the introduction of new optimization strategies, substantially improved the model's performance on the noisy CIFAR-100 dataset. While ResNet18 initially showed promise, the transition to VGG16 marked a pivotal improvement, resulting in a validation accuracy increase from 47.19% to 61.49%. These findings underscore the importance of adapting the architecture and training strategies to the unique challenges posed by the dataset. By filtering noisy labels, applying robust augmentation techniques, and leveraging well-suited models, this work demonstrates a clear pathway to overcoming data quality issues and achieving superior performance.

References

[1] Lu Jiang, Zhengyuan Zhou, Thomas Leung, Li-Jia Li, and Li Fei-Fei. Mentornet: Learning data-driven curriculum for very deep neural networks on corrupted labels. In *International conference on machine learning*, pages 2304–2313. PMLR, 2018.

[2] Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin A Raffel, Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in neural information processing systems*, 33:596–608, 2020.

[3] Yisen Wang, Xingjun Ma, Zaiyi Chen, Yuan Luo, Jinfeng Yi, and James Bailey. Symmetric cross entropy for robust learning with noisy labels. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 322–330, 2019.

[4] Qizhe Xie, Minh-Thang Luong, Eduard Hovy, and Quoc V Le. Self-training with noisy student improves imagenet classification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10687–10698, 2020.

[5] Linjun Zhang, Zhun Deng, Kenji Kawaguchi, Amirata Ghorbani, and James Zou. How does mixup help with robustness and generalization? *arXiv preprint arXiv:2010.04819*, 2020.

[6] Zhilu Zhang and Mert Sabuncu. Generalized cross entropy loss for training deep neural networks with noisy labels. *Advances in neural information processing systems*, 31, 2018.