

# Improving CNN Generalization with Data Augmentation and Label Smoothing

Aria Ahmadi

June 2025

## Abstract

Convolutional Neural Networks (CNNs) have achieved remarkable success in image classification, yet *overfitting* remains a central challenge—particularly when data are scarce and model capacity is high. This paper investigates two complementary regularization strategies: **(i)** aggressive data augmentation and **(ii)** label smoothing. Experiments on the 10-class CIFAR-10 dataset show that a lightweight CNN attains **84.55%** validation accuracy in the baseline setting. Applying augmentation or label smoothing independently lifts performance to **85.89%** and **85.62%**, respectively, while their *combination* yields **88.44%**. These findings confirm that even simple architectures benefit substantially from carefully selected regularizers.

## 1 Introduction

CNNs underpin a wide range of modern vision systems—from mobile apps to autonomous vehicles. However, their millions of parameters can easily memorize training samples, harming generalization on unseen data. Standard defenses such as dropout and batch normalization help, but in practice must be complemented with additional techniques.

This study evaluates the effectiveness of *data augmentation* and *label smoothing*—two broadly adopted strategies that inject prior knowledge and discourage overconfident predictions. We assess each technique in isolation and in tandem using a purposely compact CNN trained on CIFAR-10.

## 2 Related Work

Data augmentation has long been recognized as a simple yet powerful means of increasing effective dataset size. Automated policies—*AutoAugment* and *RandAugment*—search for optimal transformations to maximize validation performance. Alternative schemes such as *Mixup*[3] and *CutMix*[4] blend images or patches to create harder training examples.

**Label smoothing** was introduced by Szegedy *et al.*[1] to prevent classifiers from becoming overconfident; subsequent work shows it also improves calibration and robustness. Our contribution is a concise ablation showing how these two approaches interact in a high school-level research setting.

## 3 Methodology

### 3.1 Dataset and Preprocessing

We use the standard CIFAR-10 benchmark containing 60,000 color images of size  $32 \times 32$  pixels. Following common practice, we reserve 80% for training and 20% for validation. Input pixels are

normalized to  $[-1, 1]$ . During training we apply random horizontal flips and rotations of up to  $\pm 15^\circ$ .

### 3.2 Model Architecture

Our **SimpleCNN** comprises two convolutional blocks followed by a fully connected (FC) classifier. Each conv block contains two  $3 \times 3$  convolutions, batch normalization, ReLU activations,  $2 \times 2$  max pooling, and 25% dropout. The FC block flattens the feature map, applies a 256-unit dense layer with ReLU and 50% dropout, and finally a 10-way linear classifier.

Layer	Type	Output Shape	Params
Input	–	$32 \times 32 \times 3$	–
Conv1	$3 \times 3$ @64	$32 \times 32 \times 64$	1.8k
Conv2	$3 \times 3$ @64	$32 \times 32 \times 64$	36.9k
MaxPool	$2 \times 2$	$16 \times 16 \times 64$	0
Conv3	$3 \times 3$ @128	$16 \times 16 \times 128$	73.9k
Conv4	$3 \times 3$ @128	$16 \times 16 \times 128$	147.6k
MaxPool	$2 \times 2$	$8 \times 8 \times 128$	0
Flatten	–	8192	0
FC1	256	256	2.1M
FC2	10	10	2.6k

Table 1: Detailed architecture and parameter counts for **SimpleCNN**.

### 3.3 Training Procedure

All variants are trained for 100 epochs with the Adam optimizer (learning rate  $1 \times 10^{-4}$ ). Label smoothing uses  $\epsilon = 0.1$ . We evaluate four settings:

1. **Baseline**: standard cross-entropy, no augmentation.
2. **Data Aug Only**: horizontal flips + rotations, cross-entropy.
3. **Label Smoothing Only**: label-smoothed loss, no augmentation.
4. **Aug + LS**: both techniques combined.

## 4 Results and Analysis

### 4.1 Ablation Study

### 4.2 Training and Validation Curves

### 4.3 t-SNE Feature Visualization

## 5 Conclusion

Aggressive data augmentation and label smoothing each mitigate overfitting in isolation, but their combination yields the highest performance (+3.74 pp) on CIFAR-10 using a compact CNN. The

Experiment	Augment?	Label Smooth?	Val Acc.
Baseline	×	×	84.55%
Data Aug Only	✓	×	85.89%
Label Smoothing Only	×	✓	85.62%
Aug + LS	✓	✓	<b>88.44%</b>

Table 2: Validation accuracy for each training configuration.

qualitative t-SNE analysis supports these gains. Future work will explore more advanced policies (e.g., Mixup, CutMix) and deeper backbones.

## References

- [1] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. *Rethinking the Inception Architecture for Computer Vision*. In CVPR, 2016.
- [2] Alex Krizhevsky and Geoffrey Hinton. *Learning Multiple Layers of Features from Tiny Images*. Technical Report, University of Toronto, 2009.
- [3] Hongyi Zhang, Moustapha Cisse, Yann Dauphin, and David Lopez-Paz. *mixup: Beyond Empirical Risk Minimization*. In ICLR, 2018.
- [4] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo. *CutMix: Regularization Strategy to Train Strong Classifiers With Localizable Features*. In ICCV, 2019.



Figure 1: High-level diagram of the SimpleCNN used in this study.

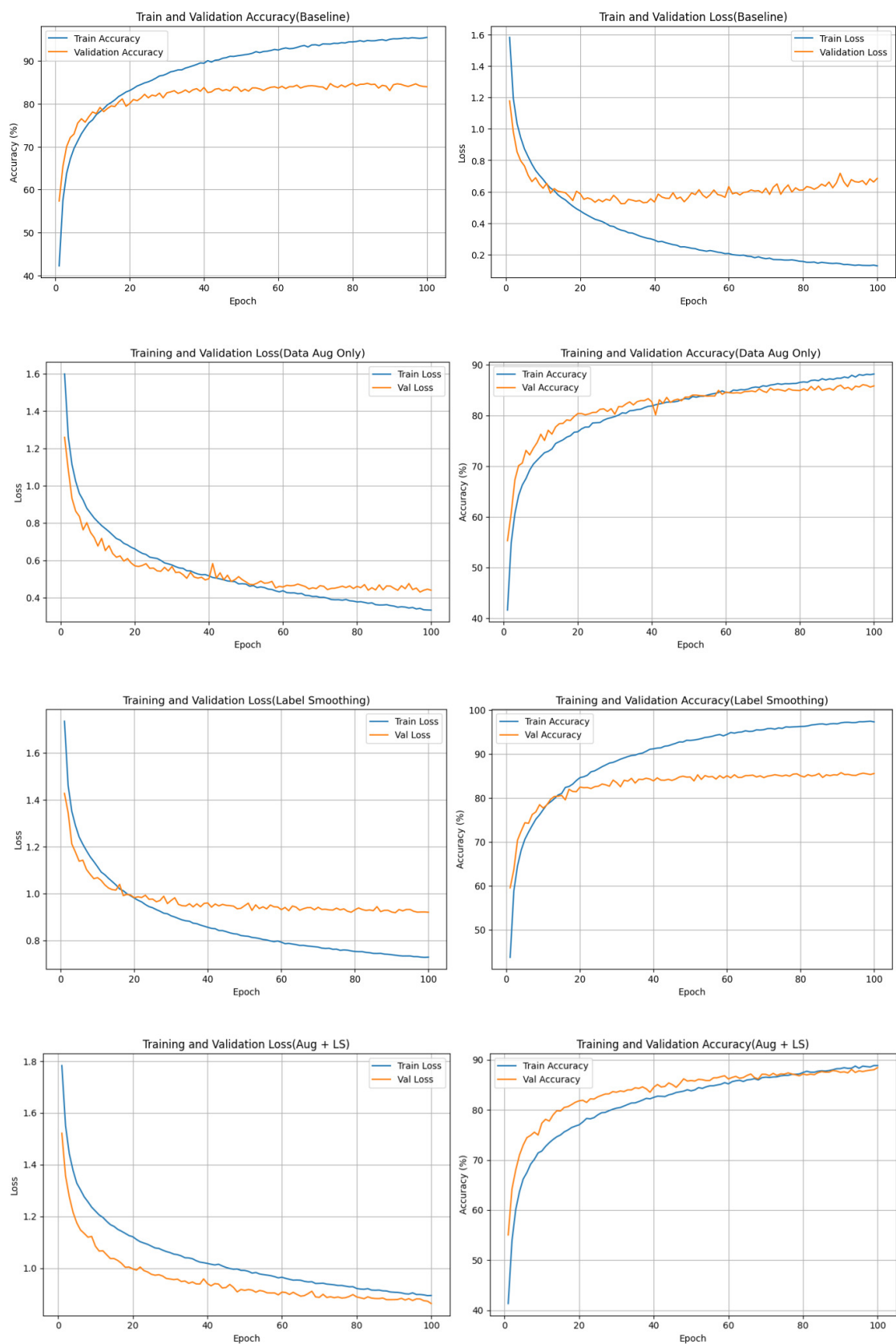


Figure 2: Training (solid) and validation (dashed) curves. Combined Aug+LS exhibits the smallest generalization gap.

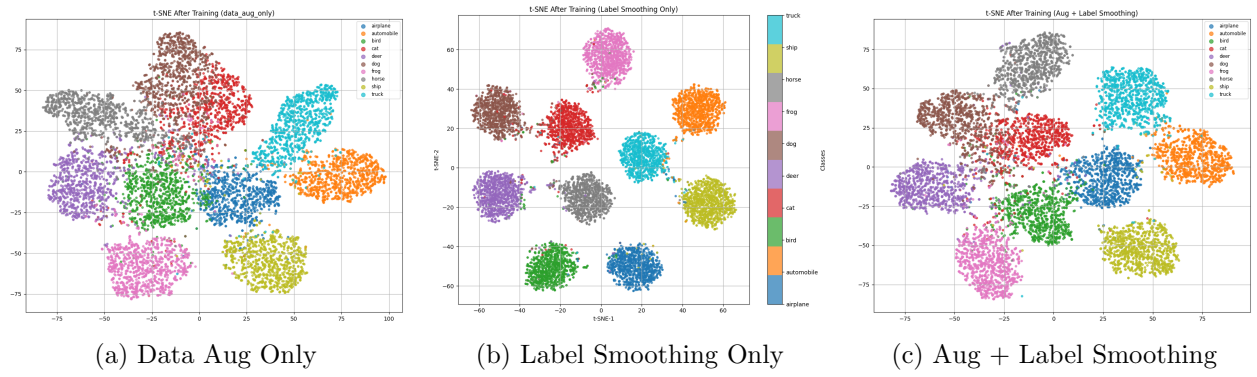


Figure 3: t-SNE embeddings of extracted features.