# Noisy CIFAR-100

Andrei Murgulet

January 2026

## 1 Introduction

This raport is meant to present results obtained for the *FII ATNN 2025 Noisy CIFAR-100* competition. CIFAR-100 is a standard benchmark for image classification, but real-world datasets often contain mislabeled samples. The competition provides 50 000 training images with synthetic label flips and clean test labels. The goal is to maximise top-1 accuracy on the hidden test set while using only the noisy training data.

## 2 Method

### 2.1 Architecture

We experiment with two main models:

- **ResNet-18** - light baseline, 11M parameters
- **ResNet-34** - bigger version, 21M parameters

Both models are initialised with Image-Net pre-trained weights

### 2.2 Loss Function

We compare **Cross-Entropy with label-smoothing(set at 0.1)** the standard loss with **Symmetric Cross-Entropy (SCE)**, which additionally penalises the label if the network is very confident about a different class, trying to reduce over-fitting to potentially wrong annotations.

### 2.3 Clean-Data Re-training

In order to minimise the noisy labels I first train ResNet-34 for 10 epochs, collect per-sample cross-entropy losses, and keep the 70 % examples with the *smallest* loss. The network is then re-trained only on this filtered subset. This change only improved as expected the train accuarcy, but the results did not improve from the baseline.

## 2.4  Data Augmentation

- RandomHorizontalFlip

- Random crop

- MixUp & CutMix applied with alpha=0.5

- Standard Normalization

## 2.5  Training Details

- Optimiser: ADAMW , LR $1 \times 10^{-4}$, cosine annealing 40 epochs

- Batch size 16

- Warm-up 5 epochs

- Weight decay $5 \times 10^{-4}$

## 2.6  Training Time

All experiments ran on a single Kaggle GPU T4 x2. Resnet 34 training takes $\approx$ 2 h 30 min for 40 epochs.

# 3  Experiments

## 3.1  Ablation Study

Table 3.1 shows the incremental contribution of each component (40 epochs).

| Configuration | Accuracy (%) | $\Delta$ |
|---|---|---|
| A. ResNet-18 + CE (baseline) | 69.6 | 0.0 |
| B. + Clean-data re-train | 64.7 | -4.9. |
| C. + SCE loss | 70.4 | +0.8 |
| D. + CutMix+ MixUp+ WarmUp | 71.6 | +1.2 |
| E. + Resnet34 | 73.2 | +1.6 |

## 3.2  Benchmark Performance

We compare our final model with previous editions of the course competition mentioned on the overview page in Table 1.

Our submission ranks 6th on the private leaderboard (accessed 12-Jan-2026).

Table 1: Leaderboard comparison.

| Team | Test Acc. (%) |
| --- | --- |
| Baseline (ResNet18) | 71.46 |
| Baseline (ResNet34) | 74.78 |
| Best ATNN 2024 submission | 75.14-75.50 |
| Best ATNN 2025 submission | 84.29 |
| **My Resnet18 score** | **71.6** |
| **My Resnet34 score** | **73.2** |

# 4  Conclusions

Starting from a basic ResNet-18, we gradually added high-loss filtering, strong augmentation, warm-up scheduling and moved to the bigger model ResNet-34. The final 73.25% test score placed my solution on the 6th place. Future work could explore:

- Self-supervised pre-training to further reduce reliance on labels

- Dynamic threshold for small-loss selection