Convolutional Neural Networks for CIFAR10

Universidad Carlos III de Madrid

Duarte Moura

Roberto Martínez - Guisasola Guerrero

November 21, 2024

This project studies the effect of regularization and temperature scaling on model calibration by analyzing LeNet5 and DenseNet121 architectures trained on CIFAR-10's bird vs. cat classification task. Key metrics such as Expected Calibration Error (ECE) and Reliability Diagrams are used to evaluate model performance.

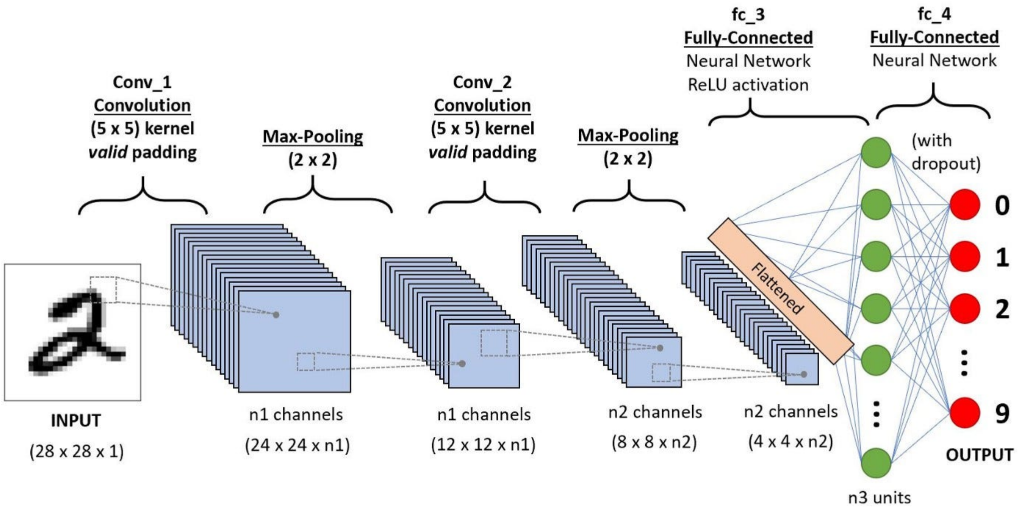


Figure 1: CNN architecture

1 Introduction

Convolutional Neural Networks (CNNs) use convolutional layers to reduce data dimensionality. These layers extract features from different parts of an image, which are then passed to fully connected layers for classification.

2 Methodology

2.1 LeNet architecture

The LeNet architecture was defined in 1989. It consists of two convolutional layers, each followed by a ReLU activation function and a MaxPooling layer. Finally, there are 3 fully connected layers of 120, 84 and 10 output units respectively.

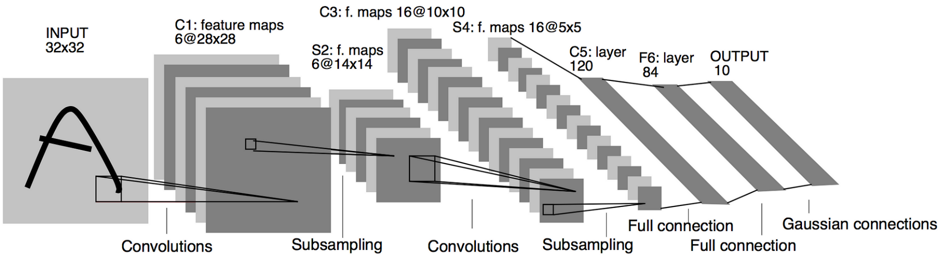


Figure 2: LeNet architecture

2.2 DenseNet121

In order to compare calibration on smaller lex complex and larger more intricate models, we decided to repeat the training on the DenseNet121 model.

2.3 Dataset

We have used CIFAR10, a dataset of 50,000 training images and 10,000 test images, belonging to 10 different categories, so that there are 6,000 images per category. The division of the dataset has been done in 3 subsets:

- **train**: used to perform the training of the CNN and adjust its parameters.
- **val**: used to evaluate the performance of the model during training. 20% of the training images have been extracted to generate this subset.
- **test**: used to evaluate the performance of the model after the training is finished.

2.4 Model calibration

Model calibration refers to the ability of the neural network to produce probabilities that correctly reflect its confidence in the predictions. In a well-calibrated model, the probability assigned to a prediction matches the actual frequency of hits. However, deep neural networks are often poorly calibrated, typically exhibiting overconfidence in their predictions. To improve the calibration, we use temperature scaling, and we have measure the model calibration with the Expected Calibration Error (ECE).

2.5 Temperature scaling

Temperature is a tuning parameter in the temperature scaling method, used to calibrate the outputs of a neural network to ensure that its estimated probabilities accurately reflect the model's confidence in its predictions. By applying a temperature factor to the output probabilities before passing them through the softmax function, it is possible to 'smooth' these probabilities, thereby reducing model overconfidence. In our project, we use temperature scaling to adjust the predictions of a LeNet network, evaluating how this adjustment improves the match between the predicted probability and the actual accuracy of the network.

2.6 Training Setup

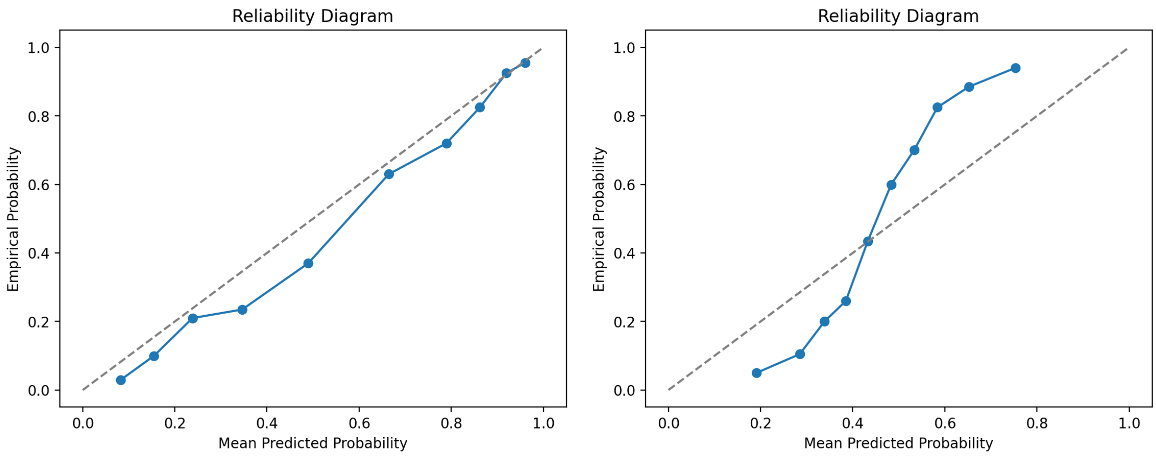
- **Hyperparameters**:
  - Epochs: 20
  - Learning Rate: 0.001
  - Batch Size: 64
- **Optimizer**: Adam
- **Loss Function**: Negative log likelihood loss (NLLLoss) for models without temperature and Cross Entropy Loss for the temperature calibration.

Regularization techniques, including early stopping (patience of 5 epochs), dropout (probability of 0.5), and batch normalization, were employed during training.

3 Results

3.1 Calibration

LeNet shows a better calibration compared to DenseNet121, reflected in a closer alignment to the ideal diagonal and a lower ECE. This means that LeNet produces prediction probabilities that more accurately represent the actual probability of success, which is beneficial for applications that require a reliable measure of confidence in model predictions. Figure 3 shows the reliability diagrams for both cases.



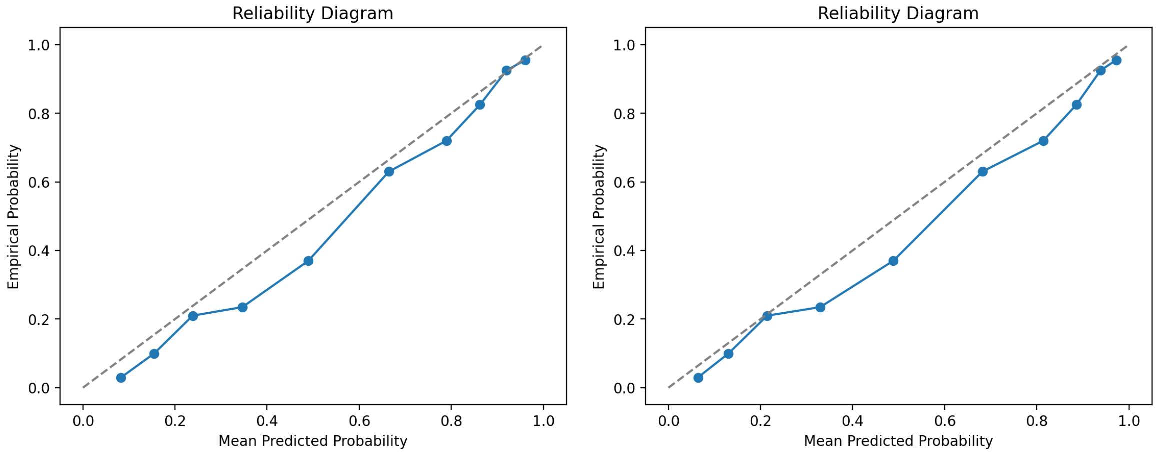
(a) LeNet calibration reliability diagram (b) DenseNet121 calibration reliability diagram

Figure 3: Comparison of reliability diagrams of small and big models

3.2 Temperature

3.2.1 LeNet Model

We have analysed the effect of temperature scaling on the LeNet model, with batch normalisation and dropout since it was the model that generaized better. The ECE value before applying temperature scaling was 0.042, while after applying temperature scaling, the ECE value has decreased to 0.0395, which means that the model is better calibrated. Figure 4 shows the reliability diagrams for both cases. As can be seen, the distance between the model calibration line and the dashed line has been reduced.



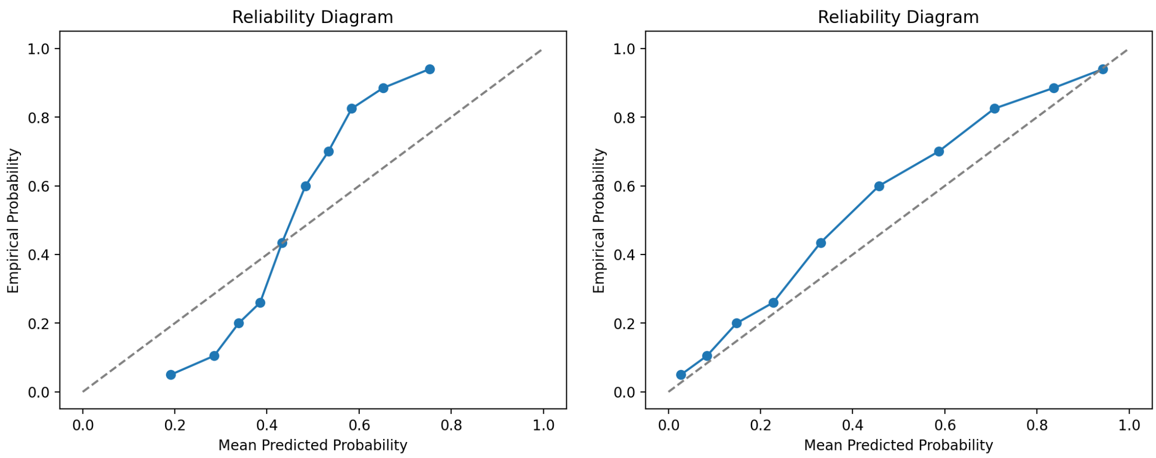
(a) LeNet model without temperature

(b) LeNet model with temperature

Figure 4: Comparison of reliability diagrams with or without temperature in LeNet model

3.2.2 DenseNet121

As expected, the effect of temperature scaling is the same on the DenseNet121 model as on the base model, since in both models the calibration of the models is improved. However, the improvement is greater in DenseNet121 as the model was less well calibrated. This may be due to a higher number of layers, which makes it difficult to adjust the weights correctly. Figure 5 shows the reliability diagrams for both cases. As can be seen, the distance between the model calibration line and the dashed line has been reduced to a greater extent than in the LeNet model. Additionally, this is consistent with the higher initial ECE value of DenseNet121 relative to LeNet.



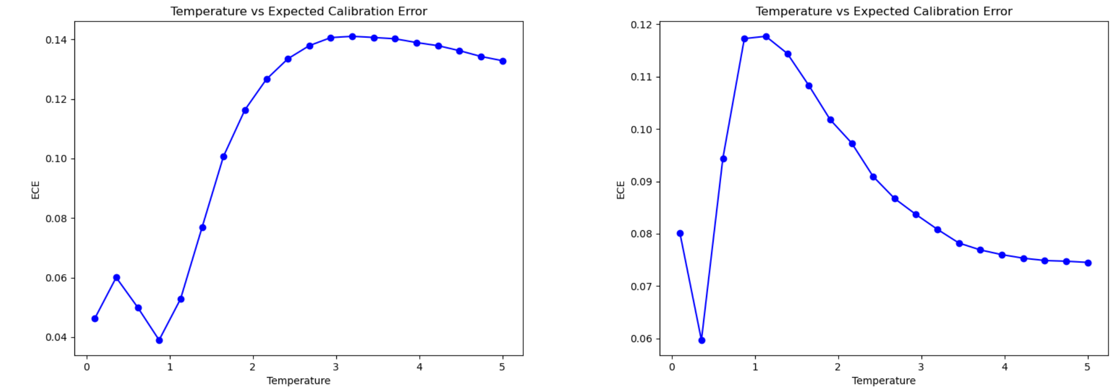
(a) DenseNet121 model without temperature

(b) DenseNet121 model with temperature

Figure 5: Comparison of reliability diagrams with or without temperature in DenseNet121 model

3.2.3 Comparing models

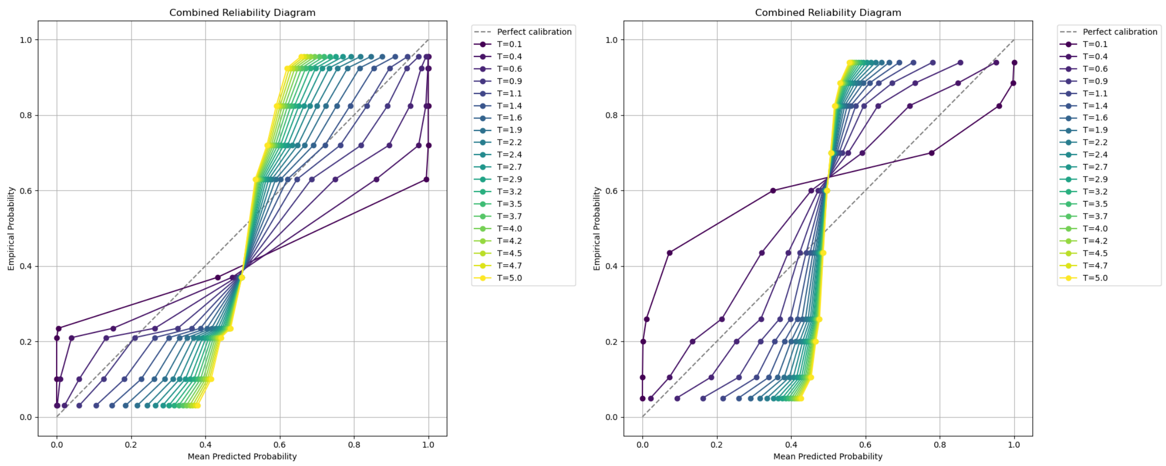
We tested the effect of different temperature values and compared the ECE values obtained. The lower the ECE value, the better the calibration of the model, so the higher the accuracy of the predictions. Figure 6 shows the reliability diagrams for both cases. As can be seen by comparing figures 6a and 6b, the effect of temperature is different in the large models, with lower temperature values giving better results. However, the ECE value is lower in the small models, although as the temperature increases, the ECE value increases.



(a) LeNet model (b) DenseNet121 model

Figure 6: Comparison of ECE and temperatures in both models

Below, we can also see how different temperatures affect the ECE of the different models, which were collected during our temperature study.



(a) LeNet model (b) DenseNet121 model

Figure 7: Comparison of ECE and temperatures in both models

Table 1: Temperature Study in both models

Model	ECE (Pre-temp)	ECE (Post-temp)	Optimal Temp	ECE (Optimal Temp)
LeNet	0.0420	0.0395	0.8740	0.0390
DenseNet121	0.1187	0.0571	0.3580	0.05970

4 Conclusion

This project highlights the importance of calibration in CNNs. Despite LeNet5 showing better initial calibration, DenseNet121 benefited more from temperature scaling due to its complexity. Post-scaling, both models achieved significantly improved ECE, demonstrating the utility of temperature scaling for real-world applications requiring reliable probability estimates.