Deep Learning
Project Report - Image Classification

Tom Baran & Olivier Blot & Yohann Djafer

Digital Sciences Department - Second Year
2023-2024

Contents

List of Figures

1 Introduction

1.1 Context and Objectives

The aim of this project is to develop an artificial intelligence capable of classifying images of bakery products, specifically baguettes, pain au chocolat, and croissants. Accurate classification of these products is crucial, particularly for applications such as bakery automation, automatic inventory, or personalized recommendations in online stores. The main goal is to achieve high accuracy in recognizing these different types of pastries to minimize classification errors and improve the efficiency of automated systems. Moreover, a similar Japanese AI has proven surprisingly useful in recognizing certain types of cancer cells.

1.2 Problem Statement

One of the main issues encountered is the frequent confusion between croissants and pain au chocolat. This confusion is due to visual similarities between these two pastries, despite their distinct features. Another major constraint is the limitation of computational resources, particularly restricted access to GPUs required to train efficient deep learning models. This limitation directly impacts the ability to experiment with more complex models or conduct extended training.

1.3 Methodology

To address these challenges, several methodological approaches were adopted:

- **Data Augmentation:** Data augmentation was used to enrich the training dataset. Techniques such as rotation, zooming, color transformations, and cropping were applied to generate variations of existing images, thereby increasing data diversity and improving model robustness.
- **Transfer Learning:** Transfer learning involves using pre-trained models on larger and more generic datasets, then adapting them to the specifics of our dataset. This approach allows us to leverage knowledge acquired by complex models without needing to train from scratch.
- **Fine-tuning:** Fine-tuning consists of carefully adjusting the weights of pre-trained models on our specific dataset. This step helps optimize model performance for the pastry classification task while efficiently using limited computational resources.

2 Initial Network

2.1 Initial Model

We started by creating a neural network inspired by the one from lab exercise TP3 (dog and cat image recognition).

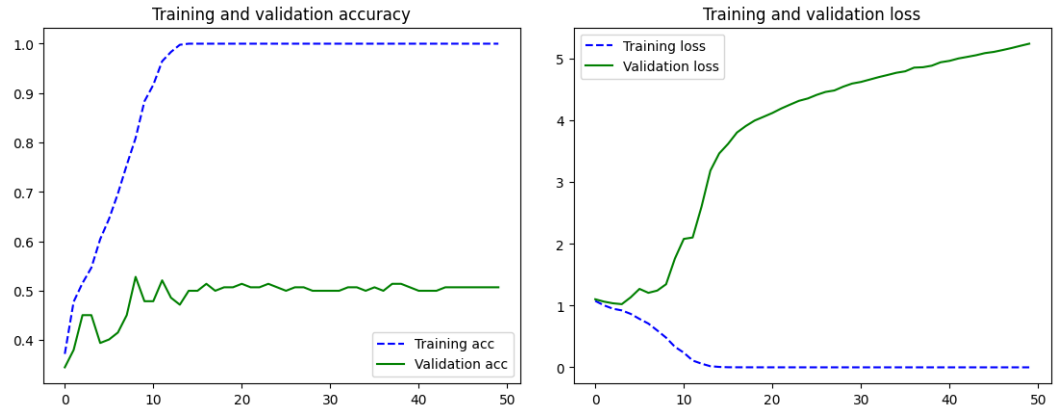
2.1.1 Methodology Used

The model begins with a convolutional layer of 32 filters of size 3x3 with ReLU activation, processing input images of size 128x128 with three color channels (RGB). This is followed by a max pooling layer to reduce spatial dimensions. It continues with a second convolutional layer of 64 filters (3x3), followed by another max pooling. A third layer with 96 filters and a fourth with 128 filters, each followed by max pooling, extract increasingly complex features. The outputs are then flattened into a 1D vector. This vector is passed through a dense layer with 512 neurons and ReLU activation. Finally, a softmax output layer produces probabilities for the three classes: baguette, pain au chocolat, and croissant.

The model was compiled using sparse categorical crossentropy loss, the Adam optimizer with a learning rate of 0.0003, and evaluated using sparse categorical accuracy. It was trained over 50 epochs with a batch size of 10, using validation sets to assess performance during training.

2.1.2 Results Obtained

- **Validation Accuracy:** 50%
- **Validation Loss:** 1 (after 10 epochs)



(a)
b

(b)
b

Figure 1: Results of the initial model

Validation accuracy for this first model was quite low (50%), compared to 70% at the same stage in TP3. Overfitting also occurred quickly (around 10 epochs), likely due to our relatively small training set of about 1200 images, whereas TP3 had 2000 images and showed less overfitting.

2.2 Transfer Learning

2.2.1 Methodology Used

- **Pre-trained Model:** We used the VGG16 model pre-trained on ImageNet, a large dataset with millions of labeled images.
- **Freezing Initial Layers:** The convolutional layers of the pre-trained model were frozen (weights not updated during training) to retain the learned features.
- **Adding Specific Layers:** We added fully connected layers on top of the VGG16 network to adapt it to our task.

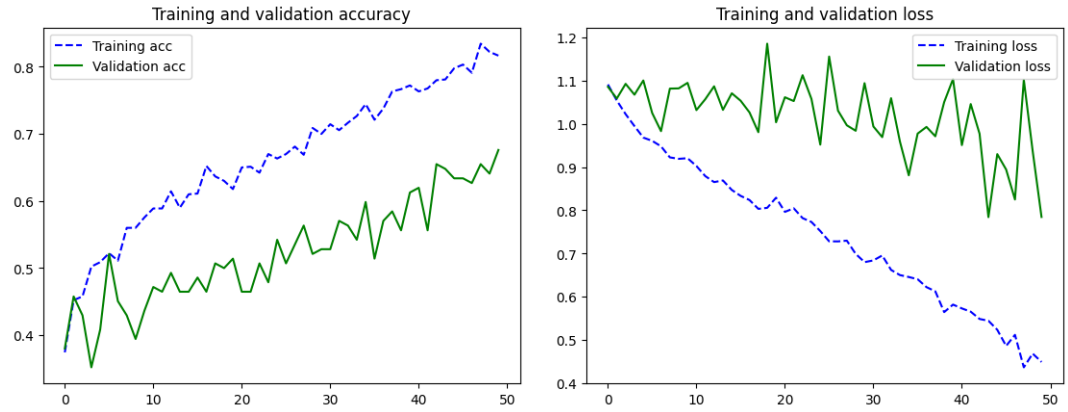
A flattening layer was first added to convert the 3D output of VGG16 into a 1D vector.

Then, a dense layer with 256 neurons and ReLU activation was added to capture complex high-level features.

Finally, a dense output layer with 3 neurons and softmax activation was added to produce class probabilities for our task.

2.2.2 Results Obtained

- **Validation Accuracy:** 65%
- **Validation Loss:** 0.8



(a)
b

(b)
b

Figure 2: Results of the transfer learning model

Transfer learning allowed us to leverage a model trained on a large dataset, yielding higher accuracy than our initial model with our small dataset. This approach optimizes resources and reduces training time while achieving strong performance—demonstrating its effectiveness for image classification tasks with limited data.

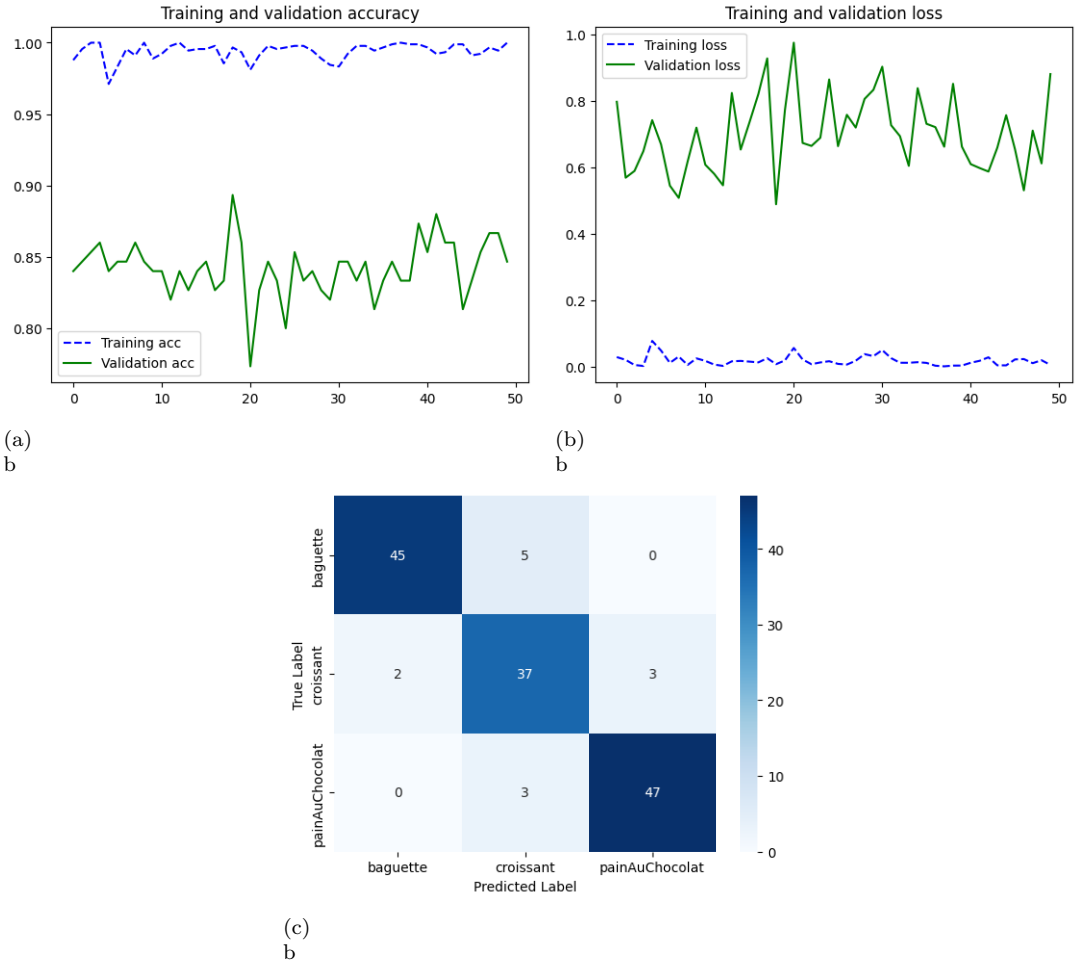
2.3 Fine Tuning

2.3.1 Methodology Used

1. **Initialization with a Pre-trained Model:** We used the VGG16 model as the starting point.
2. **Unfreezing Initial Layers:** Early layers were unfrozen to allow fine adjustments.
3. **Adding Specific Layers:** Fully connected layers (as above) were added to tailor the model to our classification task.
4. **Retraining Final Layers:** The full model (VGG16 + custom layers) was retrained with a lower learning rate.

2.3.2 Results Obtained

- **Validation Accuracy:** 85%
- **Validation Loss:** 0.8
- **Test Accuracy:** 91%
- **Test Loss:** 0.35



(a)
b

(b)
b

(c)
b

Figure 3: Results of the fine-tuned model

The fine-tuned model outperformed the transfer learning model on the validation set. However, results should be interpreted with caution, as the training and validation images are similar, and the transfer learning model was not trained on our specific images.

Still, fine-tuning gave strong test performance with 90% accuracy. The confusion matrix showed good recognition of baguettes, but croissants were harder to distinguish, often confused with pain au chocolat.

In conclusion, fine-tuning proved an effective strategy to boost performance by adapting a pre-trained model to our specific task.

3 Improvement Approach

3.1 Improving the Dataset

The previous results were obtained using training, test, and validation datasets with arbitrarily selected images. We wanted to see if better results could be achieved by modifying the datasets—especially regarding croissants and pain au chocolat. As we observed, the model often confused croissants with pain au chocolat, but not the other way around.

We modified the training images for pain au chocolat to emphasize their distinct features. For example, some images labeled as pain au chocolat actually contained croissants—these were removed.

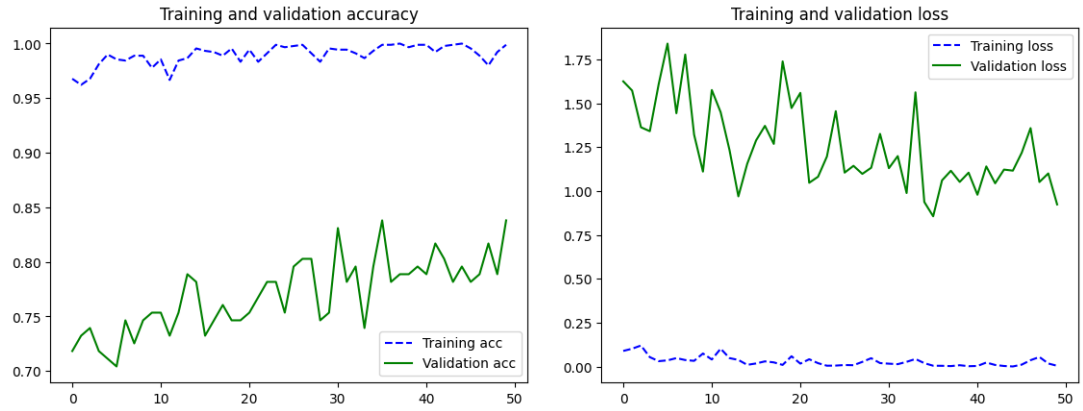
Another important point: the initial dataset had 300 training samples per label, 50 test samples, and 50 validation samples—except for croissants, which had only 42 validation images. Validation data for croissants was increased to 50 for consistency.

Would these changes improve the network? The idea was to provide cleaner, more balanced data, which should theoretically help the model distinguish croissants from pain au chocolat and improve overall classification accuracy.

3.1.1 Improved Results

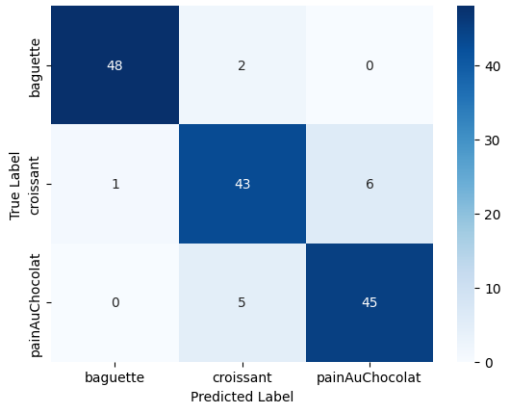
- **Validation Accuracy:** 94
- **Test Accuracy:** 92

In conclusion, modifying the training and validation datasets significantly improved the model's accuracy. Many images in the original dataset were duplicates or too dissimilar from classic pastries, which negatively affected the model.



(a)
b

(b)
b



(c)
b

Figure 4: Results after dataset improvement

3.2 Increasing Image Resolution

After training with the improved dataset, our model reached about 84

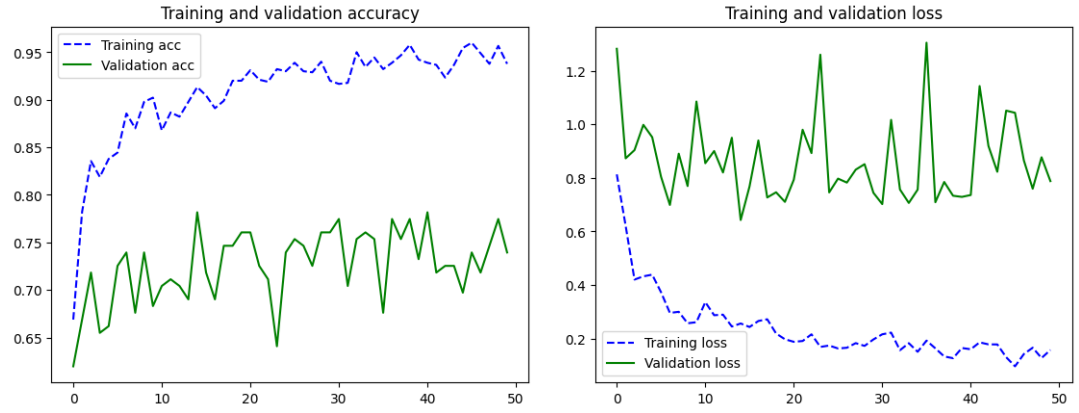
To improve this, we increased image resolution to 256x256 pixels. This higher resolution allows the model to extract finer features, especially helpful for distinguishing visually similar pastries.

The results showed significant improvement. Learning and validation metrics improved, with faster convergence and better final values compared to 128x128 images.

3.2.1 Improved Results

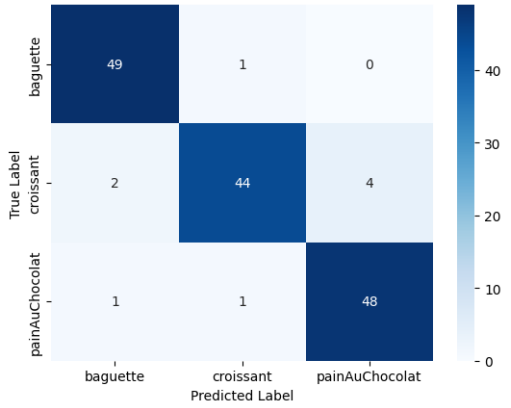
- **Validation Accuracy:** 85
- **Validation Loss:** 1
- **Test Accuracy:** 94
- **Test Loss:** 0.35
- **Training and Validation Accuracy:** Curves showed higher and more stable accuracy with 256x256 images.
- **Training and Validation Loss:** Loss decreased faster and reached lower values.
- **Croissant Recall Rate:** Increased recall for croissants and pain au chocolat, reducing confusion and improving overall accuracy.

Ultimately, increasing image resolution led to better class discrimination—especially for croissants—and was an effective improvement for our image recognition project.



(a)
b

(b)
b



(c)
b

Figure 5: Results after increasing image size

4 Implementation and Testing

4.1 Procedure

To reproduce results and test modifications, we followed these steps:

1. **Cloning the GitHub Repository:** We started by cloning the project repository containing code and data.
2. **Importing Required Modules:** We imported all necessary libraries and modules.
3. **Loading and Analyzing Data:** We executed functions to load and analyze data for initial exploration and preparation.
4. **Running Code Cells:** We then ran code cells starting from section III of the notebook, where the models are defined and trained.

4.2 Training the Network

The neural network was trained using the modified datasets. Model performance was assessed through various metrics and compared to initial results.

1. **Using the Modified Dataset:** The new images—cleaner and better balanced—were used for training, especially to reduce confusion between croissants and pain au chocolat.
2. **Performance Evaluation:** Performance was assessed using metrics like confusion matrix, accuracy, recall, precision, and F1-score.
3. **Results Comparison:** Improved results were compared to the initial ones to assess gains in accuracy and recall—especially for croissants.

These steps allowed us to measure the impact of dataset modifications and confirm the effectiveness of our improvements.

5 Conclusion and Future Work

5.1 Results Summary

Improvements to the dataset and image resolution significantly boosted our image recognition model's performance. The AI now reaches about 95

Cleaning and balancing the dataset—by removing ambiguous images—also contributed to higher performance. These methods reduced inter-class confusion, as reflected in performance metrics and confusion matrices. Overall, our methodology effectively increased both the accuracy and robustness of the model.

5.2 Limitations and Challenges

Despite the improvements, several limitations remain. One key issue is limited access to computational resources, especially GPUs. This restricted our ability to conduct longer training or use more complex models.

Furthermore, data quality and quantity remain essential. Even with efforts to clean and balance the dataset, variations in image quality and class representation may still affect results.

5.3 Future Work

To improve beyond current results, several paths can be explored. Additional data augmentation techniques (e.g., rotation, zoom, color shifts) could boost training diversity and robustness.

More complex models could also be considered. Finally, gaining better GPU access or leveraging cloud resources would enable longer training sessions and testing of advanced architectures—potentially yielding even better results.

A Appendix

GitHub: https://github.com/TomBaran501/BDD_Boulangerie.git