CNN for Hand-Based Binary Gender Classification

Matteo Rocco (2218397)
Alessio Taruffi (1940484)
Patrizio Renelli (1937842)
Nicolò Candita (2216597)

Academic Year 2024–25
Sapienza University of Rome
Fundamentals of Data Science

Abstract

Hand morphology and texture exhibit subtle patterns that may vary across genders. Leveraging these patterns for automated gender classification can aid applications such as biometric authentication and user personalization. Manual classification, however, is labor-intensive and prone to inconsistencies. In this project, we implemented a system that leverages a custom 2-flows AlexNet-based CNN to perform this task [1]. By focusing on AlexNet exclusively, we aim to highlight the impact of image preprocessing and architecture optimization in achieving high classification accuracy. The results showcase AlexNet's effectiveness in extracting discriminative features from hand images for binary gender classification.

1 Related Work

Recent advancements in gender classification using hand images have shown remarkable results. Mahmoud Afifi's (2019) study [2] on the 11K Hands dataset, comprising 11,076 images, employs a two-stream CNN to predict gender and extract deep features, which are then combined with Local Binary Patterns (LBP) to train ensemble SVM classifiers for biometric identification, achieving over 90% accuracy with dorsal images performing best. Gholamreza Amayeh et al. (2008) [3] focused on hand shape geometry, extracting features like Zernike moments and Fourier descriptors from hand silhouettes and using score-level fusion and LDA classifiers to achieve 98% accuracy in gender classification. Mukherjee et al. (2023) [4] introduced the JU-HD dataset for gender classification, fine-tuning pre-trained CNNs to achieve 90.49% accuracy on JU-HD and near-perfect accuracy on 11K Hands using Inception-v3.

2 Introduction

Biometric features, such as hand morphology and texture, have long been studied as reliable indicators for identity verification and gender classification.

Despite the advances in biometric technologies, the specific task of recognizing gender through hand features has been relatively unexplored. This gap presents an opportunity to investigate the potential of Convolutional Neural Networks. By experimenting with the limits of CNNs, we aim to understand their capability to discern subtle patterns in hand morphology and texture that may correlate with gender.

This report details the dataset preparation, preprocessing techniques, and architectural adaptations employed to optimize the chosen CNNs for this novel application. Additionally, we discuss the performance outcomes and the broader implications of using CNNs for gender classification, contributing to a growing field of research at the intersection of biometrics and artificial intelligence.

3 Dataset and Sampling

3.1 Overview and Characteristics

The dataset used for this project is a public dataset called "11k Hands", containing hand images from various subjects, each accompanied by associated metadata. It includes a total of 11,076 hand images with a resolution of 1600x1200 pixels. The images were captured from a base of 190 subjects ranging in age from 18 to 75 years old. The dataset features photos of both sides of the hand (palm and dorsal) and provides multiple images for each subject. The available metadatas for each picture are:

1. subject ID;
2. subject gender;
3. subject age;
4. hand skin color;
5. hand side;

6. right or left hand;
7. presence of accessories or nail polish.

Such metadatas will be essential for image sampling, in the CNN training phase and during the testing phase to evaluate the correctness of the predictions obtained. The following classifications can be made using the available metadata:

Table 1: Division by Hand Side

Hand Side	Left	Right
Dorsal	2786	2895
Palmar	2585	2810

Table 2: Division by Accessories

Presence of Accessories	Number
No	7865
Yes	3211

Table 3: Division by Skin Color

Color	Number
Dark	758
Fair	3493
Medium	6495
Very Fair	330

Table 4: Division by Gender

Gender	Number
Female	7109
Male	3967

From the classifications reported above, the dataset would seem to be correctly balanced with regard to the global subdivision of the data based on the side (palm/back) and the side (right/left).

The main problem of the dataset concerns the subdivision of the images by subject, by gender and, within the individual samples, the distribution relative to the analyzed side (palm/back).

3.1.1 Subdivision by Subject

If we were to divide the images according to the subject belonging to it, we can notice a non-homogeneous distribution by subject, as the weight of some, within the dataset, is much greater than the others.

Subject	Subject's number %	Associated images	Associated images %
Top 19	10,00%	2318	20,93%
Top 30	15,79%	3429	30,96%
Bottom 30	15,79%	733	6,62%

Table 5: Distribution of subjects and associated images .

3.1.2 Subdivision by Subject and Side

By analyzing the images for each subject in detail, we find a non-homogeneous subdivision with regards to the number of images for each side.

Subject	Number of subjects	Percentage of subjects
With minimum threshold at 40%	33	34.78%
With minimum threshold at 50%	178	93.68%

Table 6: Table summarizing subject thresholds and percentages

Taking into account all the above analyses, a sampling process was created that allows the model to be trained correctly.

3.2 Sampling

3.2.1 Implementation

In addition to the problems indicated above in the Problem section, in the sampling phase, to avoid the creation of bias, it was decided to exclude from the training phase the images containing accessories. Furthermore, by using a single dataset for both the training and test phases, it was decided to set a policy to avoid using the images already used in the training phase also for the test phase.

The implemented sampling takes some of its characteristics from two different sampling techniques:

1. **Random Sampling:** Each element of the dataset has the same probability of being selected, ensuring an unbiased sample;

2. **Oversampling and Undersampling:** Used specifically to balance unbalanced datasets, by adding or removing elements of some classes.

We can have a more in depth look at the pipeline implemented for sampling.

After having scanned and loaded into memory the file containing the metadata of the images and subjects contained within the dataset, we went on to build a custom data structure (python dictionary) that allows us to manipulate and select the data.

As we have structured our model, the sampling phase is of cardinal importance since each selection has a particular criterion to respect. First, the identifier of the person is extracted, of which the images of the two sides necessary for the analysis pipeline will be extracted. Subsequently, the images are extracted and will be marked with a 'check' flag to indicate that it has already been selected. An image indicated with the 'check' flag equal to true cannot be selected until the beginning of another epoch. It is clear that an image chosen in the training cannot be selected in the testing phase, to avoid falsifying the test by using images on which the model was trained.

In particular, during the sampling phase for training, all images containing accessories were excluded. This was possible by checking the value of the 'accessories' field, present in the data frame containing the global metadata, which indicates the presence of accessories (e.g. rings). These images can be misleading for the training of our model, since they present occlusions. Such occlusions can lead to the learning of further bias in the model which will necessarily lead to a drop in accuracy during the testing phase. Furthermore, to allow a better balance of the extracted data, we opted to randomly choose a pair of images (palm/back) for 2 subjects of different gender for each iteration.

4 Preprocessing

4.1 First Implementation

In the first implementation of the model, 2 CNNs were used, each for one side of the hand, which required different input image formats.

4.1.1 LeNet Preprocessing

The preprocessing implemented for LeNet consists of the following:

- Image normalization to scale pixel values to the range [0, 1];
- Conversion to gray scale;
- Contrast enhancement;
- Resizing to 32×32 pixels for compatibility with LeNet;
- Restoration of pixel values to the range [0, 255].

4.1.2 AlexNet Preprocessing

The preprocessing implemented for AlexNet consists of the following:

- Image normalization;
- Application of Gaussian blur for low-level feature extraction;
- Resizing to 227×227 pixels for compatibility with AlexNet;
- Restoration of pixel values to the range [0, 255].

4.2 Second Implementation

In the second implementation of the model, it was decided to use two instances of AlexNet as CNNs for both sides of the hand, in order to avoid a significant performance loss in the final fusion phase of the results of the two CNNs.

5 Pipeline Overview

During the development of the project, 2 different pipelines were developed, modifying over time the specific pre-processing phases and the CNNs used.

5.1 First Pipeline Model

The first implementation of the pipeline was composed of two different streams, which include specific pre-processing phases and different CNNs and a final fusion evaluation phase to merge the results obtained from the two streams. The streams were divided as follows:

- back: use of LeNet
- palm: use of AlexNet

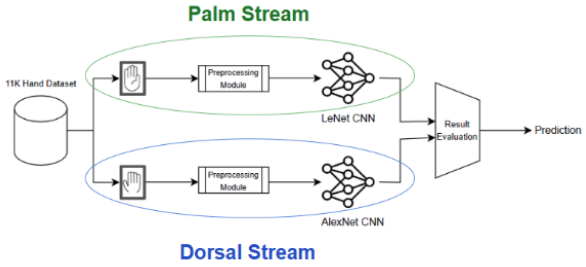


Figure 1: First pipeline model implemented

5.1.1 Palm Stream

- **Input:** randomly select, following the sampling rules illustrated above, a palm image for each iteration
- **Preprocessing Module:** perform the specific preliminary operations to prepare the data, as illustrated in the section 4.1.1
- **LeNet CNN:** Uses a custom convolutional neural network used to process pre-processed palm data and generate useful features for prediction

5.1.2 Dorsal Stream

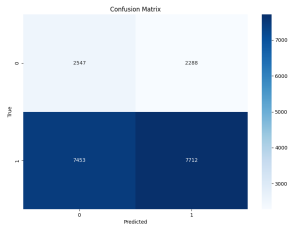
- **Input:** randomly select, following the sampling rules, illustrated above, one image of the back for each iteration
- **Preprocessing Module:** perform the specific preliminary operations to prepare the data, described in the section 4.1.2 in order to prepare the data in the correct input format for the
- **AlexNet CNN:** Represents a more advanced CNN architecture than LeNet, which processes the pre-processed data of the backbone

5.1.3 Result Evaluation

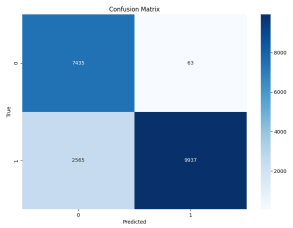
The results obtained from both CNNs (LeNet for the Palm Stream and AlexNet for the Dorsal Stream) are combined or evaluated in an analysis module to generate the final prediction. In the "Result Evaluation" module, the values related to the predictions obtained from the two CNNs are taken, weighted and combined. To do this, it was decided to use the score-level fusion as a method of fusion of the results. This method aims to combine the probabilities, of each class (male and female), returned by the softmax layers of the two CNNs. During the combination phase, it was decided to give different weights to the two CNNs, based on their accuracy, in order to differentiate the contribution they give to the calculation of the overall prediction of the model. In particular, it was decided to give greater importance to the stream related to the back, assigning it a contribution of 60% of the overall evaluation, and penalize that of the palm, assigning it only 40% of the overall evaluation. Finally in the classification layer, the class with the highest probability is taken, which will then be used as the output for the final prediction of the model.

Stream	Accuracy	Precision	Recall	F1-Score
Palmar LeNet	52%	77%	51%	61%
Dorsal AlexNet	86%	99%	80%	88%
Fusion	86%	99%	78%	86%

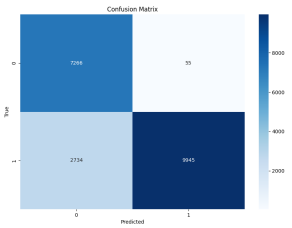
Table 7: Table summarizing the results of the combined AlexNet - LeNet pipeline



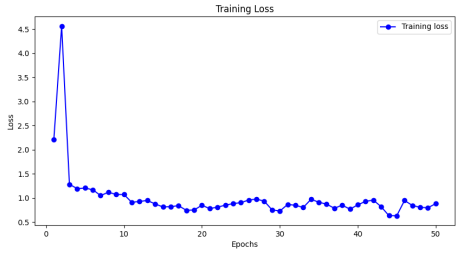
Palmar Confusion Matrix



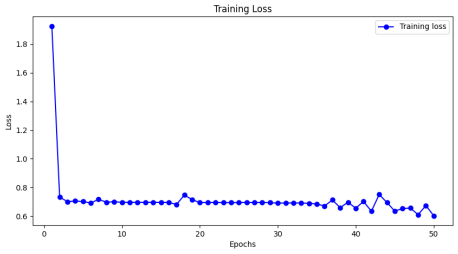
Dorsal Confusion Matrix



Combined Confusion Matrix



Loss for the Dorsal CNN



Loss for the Palmar CNN

5.2 Second Pipeline

After performing various tests on the implemented model, it was found that the use of LeNet CNN worsens the overall performance of the model. For this reason, it was decided to use CNN AlexNet for both streams. This allows us to improve the performance of the model since AlexNet has more hidden layers and we use the same CNN on both streams. The different number of hidden layers has also obstructed the use of different fusion methods, such as feature-level fusion.

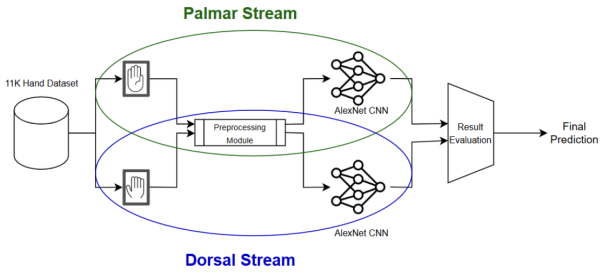
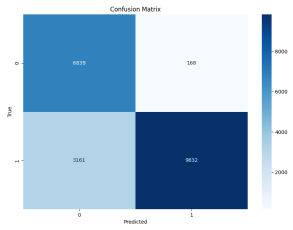


Figure 2: Second pipeline used to enhance the performances on the palmar side

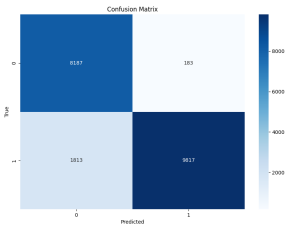
5.2.1 Result Evaluation

Stream	Accuracy	Precision	Recall	F1-Score
Palmar	83%	97%	76%	85%
Dorsal	90%	98%	84%	91%
Fusion	89%	99%	83%	91%

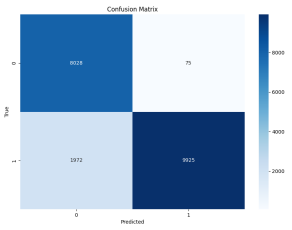
Table 8: Table summarizing the results of the AlexNet-only Pipeline



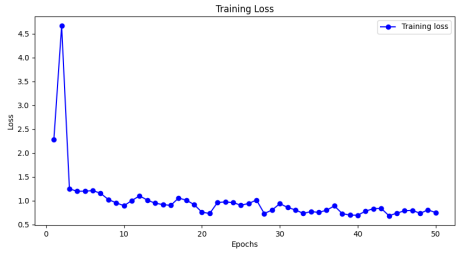
Palmar Confusion Matrix



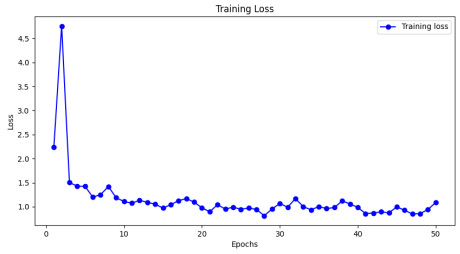
Dorsal Confusion Matrix



Combined Confusion Matrix



Loss for the Dorsal CNN



Loss for the Palmar CNN

6 Conclusions

This study evaluated convolutional neural network (CNN) architectures for gender classification using the "11k Hands" dataset. Multiple pipelines were examined, focusing on the relative performance of simpler models, such as LeNet, compared to more advanced architectures, such as AlexNet. The findings indicate that migrating from LeNet to AlexNet yields a substantial improvement in classification accuracy, underscoring the benefits of employing deeper, more sophisticated networks for this task.

7 Future developments

From what we learned from the experiments, we know that changing the CNNs used could lead to improvements in performance. However, it should be noted that each CNN accepts a particular input format, so if the CNN were changed, the specific pre-processing phase would need to be redesigned. There are several pre-trained models in the field of image analysis, which could be used for various experiments. Among these, we find the implementation of the VGG-Net CNN, which is deeper and therefore could be more accurate than AlexNet. Additionally, we plan to test the models on a different dataset beyond the current 11k Hands dataset to improve generalizability and robustness.

8 Roles

The project's workflow has been approached collaboratively, ensuring that everyone contributed equally without assigning rigid roles. Some particular contributions include:

- Dataset analysis - Patrizio and Matteo
- Preprocessing - Patrizio and Nicolo
- Pipeline implementation - Patrizio, Matteo and Alessio
- Performance evaluation - Alessio and Nicolo
- Training and Testing - Alessio

References

[1] Mahmoud Afifi. Gender recognition and biometric identification using a large dataset of hand images. *CoRR*, abs/1711.04322, 2017.

[2] Mahmoud Afifi. 11k hands: Gender recognition and biometric identification using a large dataset of hand images, 2018.

[3] Gholamreza Amayeh, George Bebis, and Mircea Nicolescu. Gender classification from hand shape. *IEEE Computer Vision and Pattern Recognition Workshop, CVPR*, 06 2008.

[4] Rajesh Mukherjee, Asish Bera, Debotosh Bhattacharjee, and Mita Nasipuri. *Human Gender Classification Based on Hand Images Using Deep Learning*, pages 314–324. 01 2023.