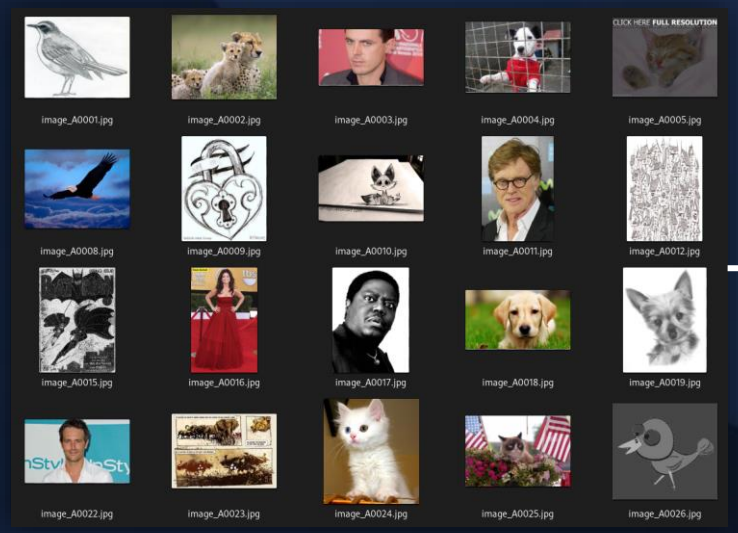
Face and Gesture Analysis

Face Identification Challenge 2024

Àlex Montoya Pérez

TABLE OF CONTENTS

CHALLENGE EXPLANATION



Dataset of 1200 images

Example image

Classify

- Users (80 labels)
- Impostors
- No facial features

FIRST APPROACH

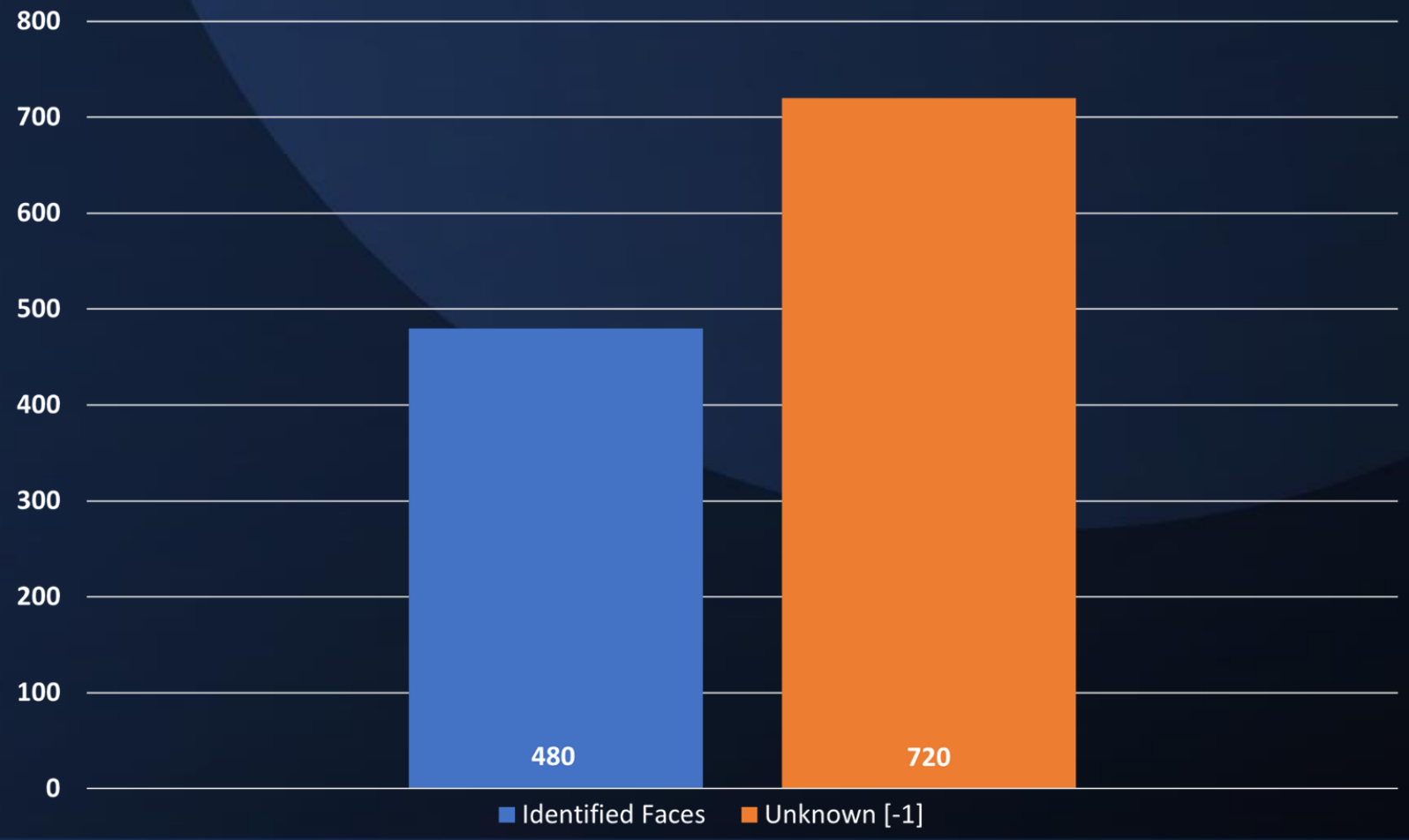
Naive approach:

- Used a **simple CNN** (5 layers).
- Applied **static transformations** to dataset, before splitting.
- Passed every image through Viola-Jones, **kept them irregardless of face or not.**

RESULT

TRAINING score:
92.62%
TESTING score:
18.57%

DATA ANALYSIS



- Dataset was heavily unbalanced.
- Some identities had only 4 images.

DATA AUGMENTATION

Class i

Search

Viola Jones

Class i

Problem

Not taking into account unrecognized faces

Solution

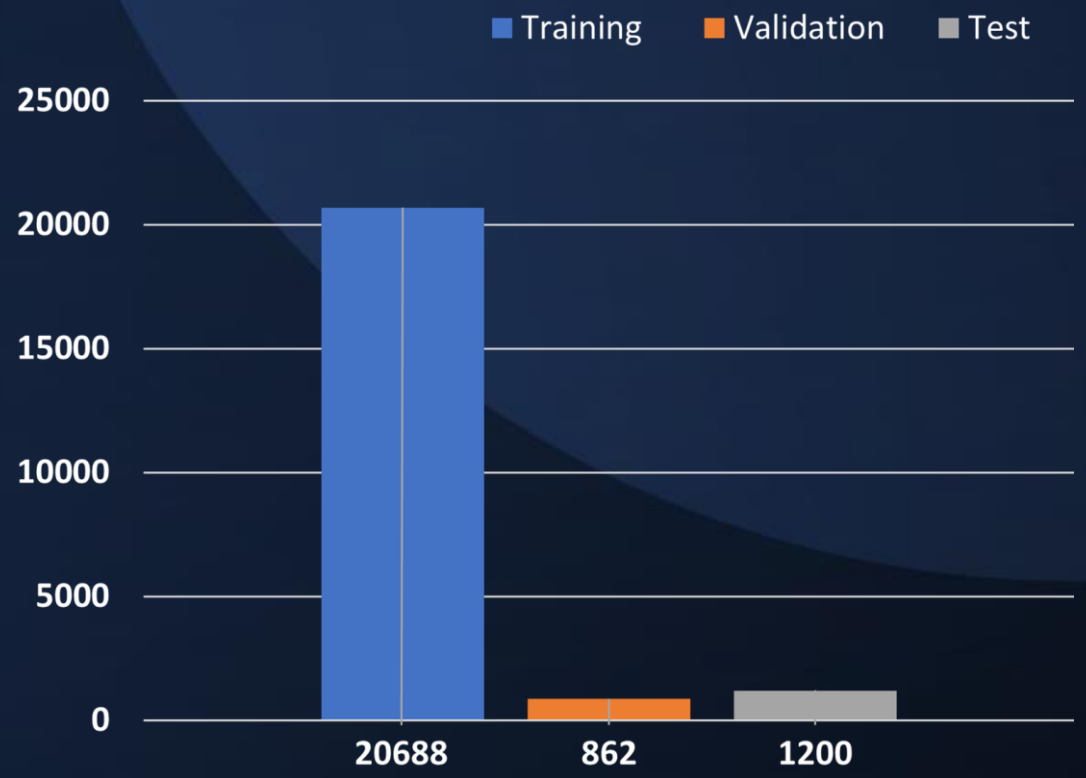
Increase the number of unrecognized faces with https://thispersondoesnotexist.com/

Class -1

DYNAMIC TRANSFORMATIONS

- Random horizontal flip (every time).
- Random rotation in the [-30, 30] degrees range.
- Gaussian blur.
- Resizing the image to a lower resolution (50x50).
- Changing its brightness and contrast.
- Randomly cropping a section.

RESULT OF DATA AUGMENTATION

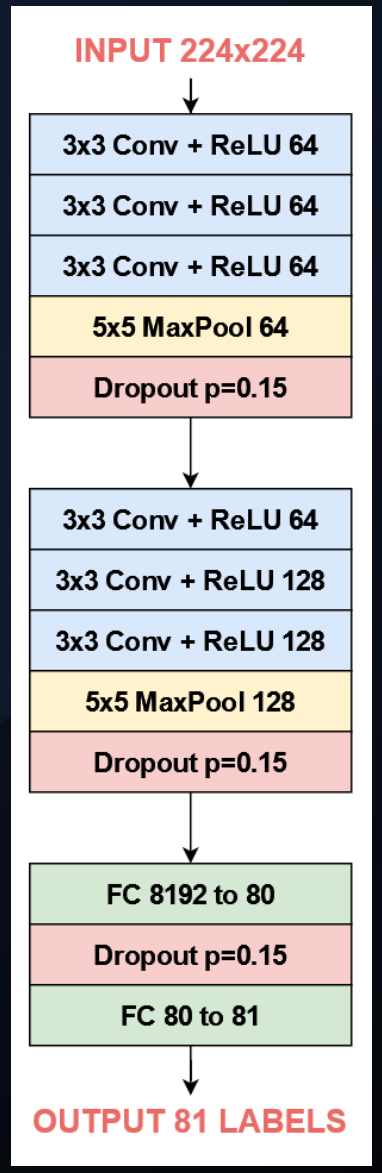


- Transformations are not applied to validation data.
- Removed TRAINING images from original dataset.
- TRAINING dataset is now used for testing with the challenge script.

Example images from our augmented dataset

DEFINING THE CNN MODEL

```
class VGGSimple5(nn.Module):
    def __init__(self, num_classes=10):

        super(VGGSimple5, self).__init__()

        self.conv11 = nn.Conv2d(3, 64, kernel_size=3, stride=1, padding=1)
        self.conv12 = nn.Conv2d(64, 64, kernel_size=3, stride=1, padding=1)
        self.conv13 = nn.Conv2d(64, 64, kernel_size=3, stride=1, padding=1)

        self.conv20 = nn.Conv2d(64, 64, kernel_size=3, stride=1, padding=1)
        self.conv21 = nn.Conv2d(64, 128, kernel_size=3, stride=1, padding=1)
        self.conv22 = nn.Conv2d(128,128, kernel_size=3, stride=1, padding=1)

        self.maxpool = nn.MaxPool2d(kernel_size=5, stride=5)

        self.fc1 = nn.Linear(8*8*128, 80)
        self.fc2 = nn.Linear(80, num_classes)

        self.relu = nn.ReLU()
        self.dropout = nn.Dropout(0.15)
```

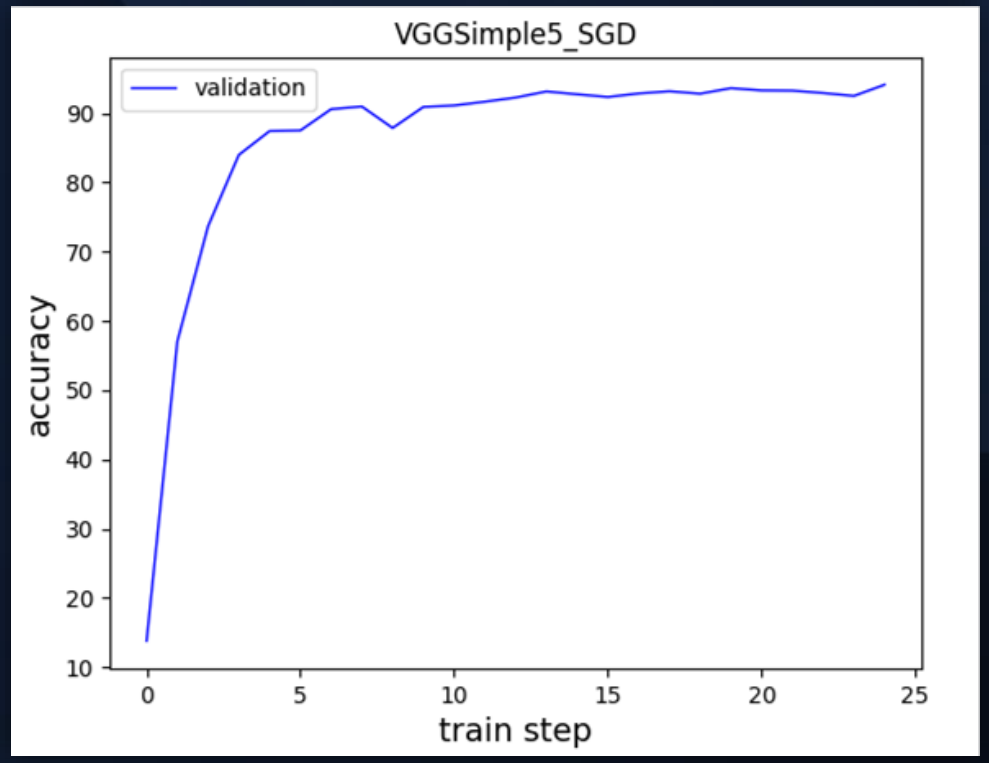


996,017 parameters with 8 layers in total. Based on VGG.

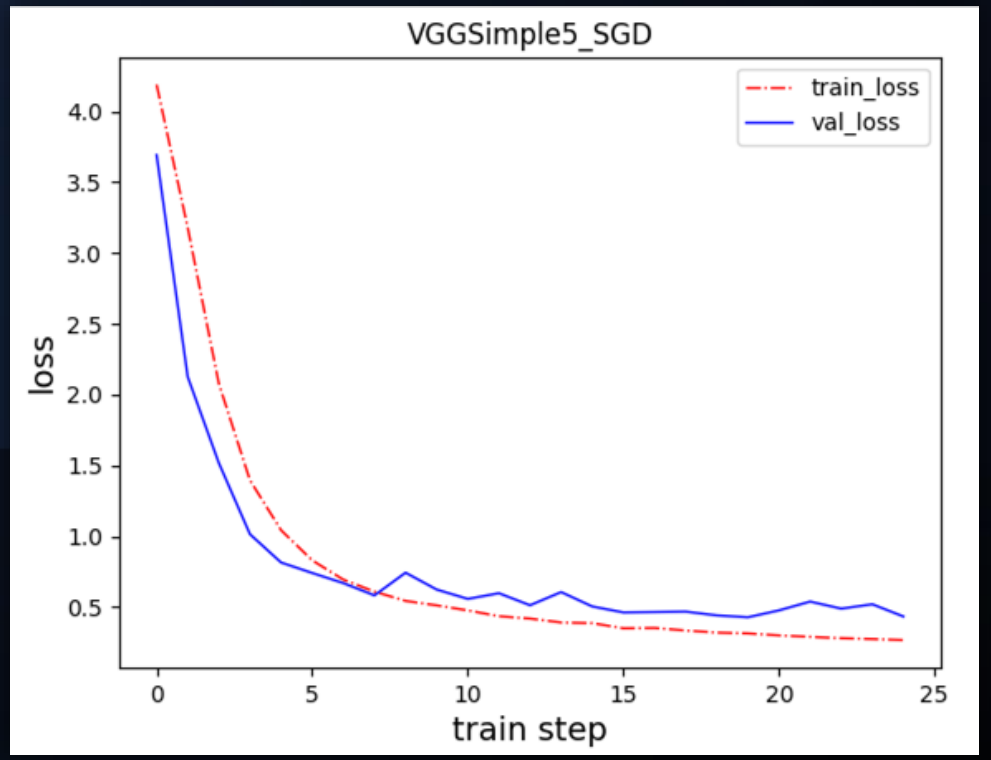
TRAINING THE MODEL

We employed the Stochastic Gradient Descent (SGD) optimizer with a learning rate of 0.01, during 25 training epochs.

Accuracy evolution:

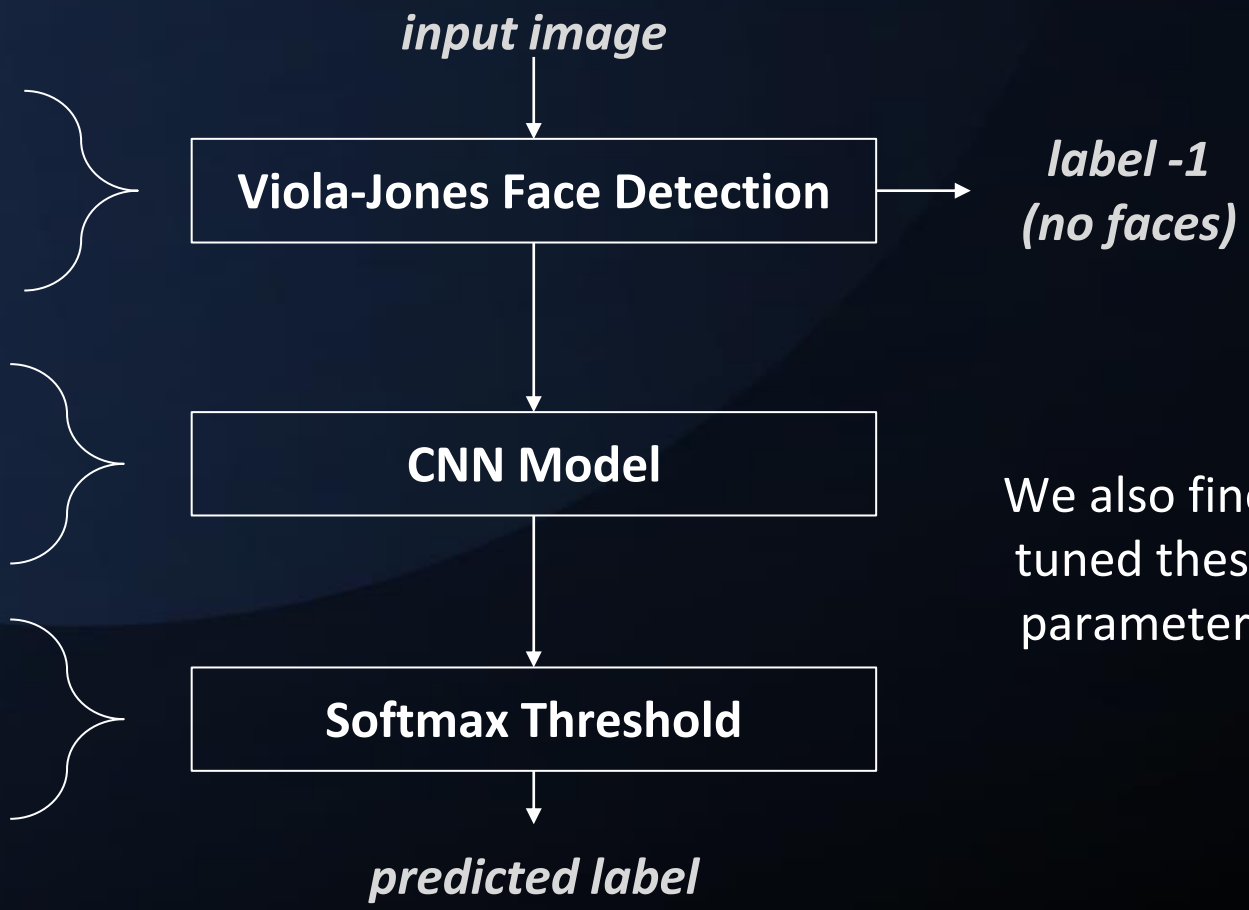


Loss evolution:



BUILDING THE IDENTIFICATION METHOD

```python
SOFTMAX_THRESHOLD = 0.8

def my_face_recognition_function(A):
    image = face_detection(A, scaleFactor=1.1, minNeighbors=6, minSize=[50, 50])

    if image is None:
        return -1

    image = Image.fromarray(image).convert('RGB')

    # Apply the preprocessing
    image = tr(image)
    image = image.view(1, *image.shape)
    image = image.to(device)

    predicted_values = softmax(our_model(image)).tolist()[0]

    max_softmax = max(predicted_values)
    if max_softmax < SOFTMAX_THRESHOLD:
        return -1

    predicted_label = int(number_to_labels[predicted_values.index(max_softmax)])

    return predicted_label
```



We also fine-tuned these parameters

RESULTS ANALYSIS

- Validation accuracy: 95.5%
- Challenge script scores:
  - TRAINING: 83.64% in 48.61 seconds.
  - TESTING: **77.18% in 28.03 seconds**.
- Most mistakes were false negatives (88/144).
- Adjusting Viola-Jones parameters changed the results in the [77, 83]% range.
- Without discarding with Viola-Jones, the TRAINING score was 74%.

```
False positives: 32
False negatives: 88
Wrong class: 24
Total wrong: 144
```

CONCLUSIONS

- A face detection algorithm (Viola-Jones) is essential for the pipeline:
  - Crops input images to the important data (faces) for the model.
  - Easily filters images with no faces (with the cost of missing some real undetected faces).
- Data augmentation was the most crucial step for obtaining a robust model.
- Some possible future improvements:
  - Replace Viola-Jones with a more accurate algorithm (ex.: Deep Learning).
  - Using transfer learning or feature extraction.
  - Even more data augmentation (with real data).

THANK YOU FOR YOUR ATTENTION

Face Identification Challenge 2024

Àlex Montoya Pérez