Gender classification and landmark detection in celebrity face images: FaceNet, Keypointrcnn_resnet50_fpn and Keypointrcnn + FaceNet

Bruno Soares Castro
Cristiano A. Schütz

Faculty of Engineering at the University of Porto - FEUP
International Master in Computer Vision - IMCV

June 2024

1 Introduction

Deep learning has revolutionized the field of computer vision, with Convolutional Neural Networks (CNNs) being the primary driving force behind this success. Over the years, numerous CNN architectures have been proposed, each with unique strengths and weaknesses. In this report, we conduct a comparative analysis of three deep learning architectures: FaceNet, keypointrcnn_resnet50_fpn, and FaceNet + keypointrcnn_resnet50_fpn.

The focus of this case study is to evaluate those models on datasets compatible with the CelebA Mini dataset, consisting of large-scale face attributes dataset with more than 200K celebrity images, each with 40 attribute annotations.

2 Dataset

CelebFaces Attributes Dataset (CelebA Mini) is a large-scale face attributes dataset with more than 200K celebrity images, each with 40 attribute annotations. The images in this dataset cover large pose variations and background clutter. CelebA Mini has large diversities, large quantities, and rich annotations. However, for this project we are using 500 images with its annotations, including:

- **Images**
- **Gender**
- **Keypoints**: Eyes left, eyes right, corner of the mouth left, corner of the mouth right, and nose.

3 Implementation

The implementation of this project involves several key steps, including data loading and preprocessing, model definition, model training, and model evaluation. All of these steps were performed using Python, with the help of popular libraries such as **PyTorch** and **facenet-pytorch** for deep learning, **torchvision** for data loading and transformations, and **scikit-learn** for data splitting and metrics calculation.

3.1 Data Loading and Preprocessing

The CelebA Mini dataset was loaded using the **torchvision** library. A custom dataset was defined to allow for training/validation/test splits using PyTorch's **random_split** function. The data was then split into **training (70%)**, **validation (20%)**, and **test (10%)**.

3.2 Model Definition

FaceNet, keypointrcnn_resnet50_fpn, and FaceNet + keypointrcnn_resnet50_fpn. The FaceNet model is used to predict gender. It has been modified to output two classes corresponding to the woman and man categories.

For the keypointrcnn_resnet50_fpn and FaceNet + keypointrcnn_resnet50_fpn models, we defined three classes to include the background and Gender (woman, man). Additionally, we changed the number of keypoints considered to five keypoints.

3.3 Model Training

The training of the models was performed using the lightning module. This function takes as input the model to be trained, the loss function, the optimizer, the learning rate scheduler, the number of epochs, the batch size, and the device (either CPU or GPU).

For the FaceNet model, the loss was calculated using the **BCEWithLogitsLoss** function from **PyTorch**. The **AdamW optimizer** was used for optimization, with a learning rate of 0.0001 and a weight decay of 0.005. The learning rate was scheduled to decrease after each step size using the **StepLR** scheduler, with a step size of 3 and a gamma value of 0.3.

For keypointrcnn_resnet50_fpn model, the loss is defined by own model, it is not need to define for during training process. The **AdamW optimizer** was used for optimization, with a learning rate of 0.0001 and a weight decay of 0.001. The learning rate was scheduled to decrease after each epoch using the **StepLR** scheduler, with a step size of 3 and a gamma value of 0.3.

For FaceNet + keypointrcnn_resnet50_fpn model, the loss is defined by own model, it is not need to define for during training process. The **AdamW optimizer** was used for optimization, with a learning rate of 0.00006 and a weight decay of 0.01. The learning rate was scheduled to decrease after each epoch using the **StepLR** scheduler, with a step size of 4 and a gamma value of 0.9.

All the models were trained for the maximun 50 epochs, with a batch size of 4. During training, the loss and accuracy on the training and validation sets were logged. The best model weights, based on the validation accuracy, were saved after each epoch. Additionally, different optimization learning rates, learning rate decay factors, and number of epochs were experimented with, but those mentioned above yielded the best results.

4 Performance Comparison

4.1 FaceNet

Table 1 displays the performance results from test dataset for FaceNet Model. In this model, we are using in order recognize the face and then to classify the gender.

Table 1: Test Results of FaceNet Model Performance

	Loss	**Accuracy**	**Cohen Kappa**
Test Data	0.1827	0.8991	0.7980

Figure 2, We can have a general idea of the performance during training. In this case, the model showed good results, with no signs of overfitting or underfitting during the training.

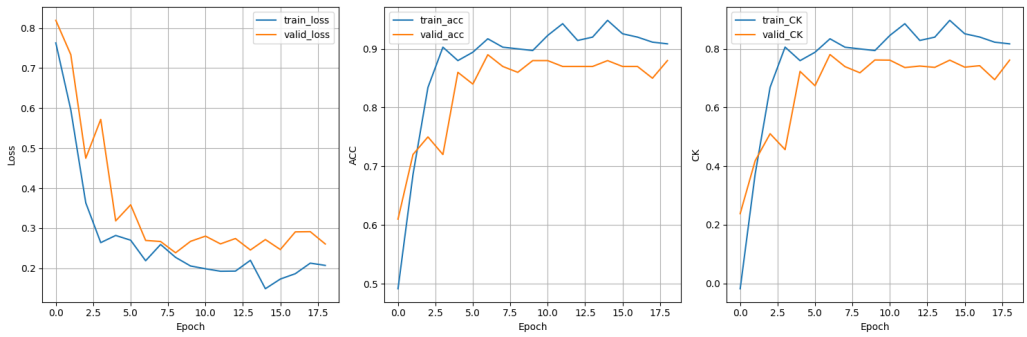


Figure 1: loss function, Accuracy, leaning rate and Cohen Kappa results for CelebA Mini datasets using FaceNet architectures.

The confusion matrix in Table 2 and the classification report in Table 3 provide an overview of the performance metrics for the classification model. The confusion matrix indicates that the model correctly identified 20 women and 25 men, with 2 women and 3 men being misclassified. The classification report details further metrics: the model achieves a precision of 0.93 and a recall of 0.89 for men, and a precision of 0.87 and a recall of 0.91 for women. The overall accuracy of the model is 90%, with both the macro and weighted averages of precision, recall, and F1-score being 0.90, indicating consistent performance across both classes.

Table 2: Confusion Matrix

	Woman	**Man**
Woman	20	2
Man	3	25

Table 3: Classification Report

	precision	recall	f1-score	support
man	0.93	0.89	0.91	28
woman	0.87	0.91	0.89	22
accuracy			0.90	50
macro avg	0.90	0.90	0.90	50
weighted avg	0.90	0.90	0.90	50

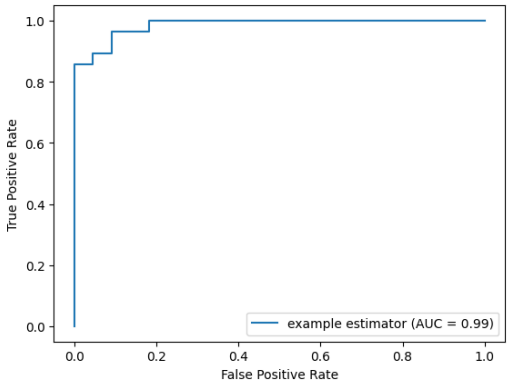
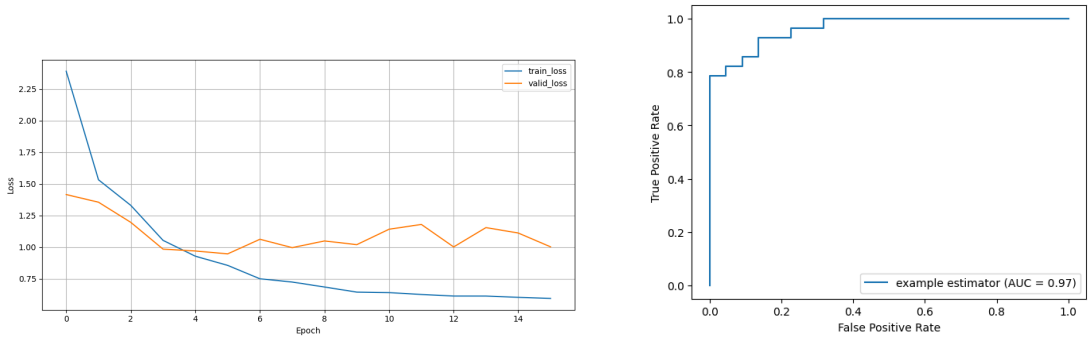


Figure 2: Curve of ROC for CelebA Mini datasets using FaceNet architectures.

4.2 Keypointrcnn_resnet50_fpn

The loss function value for the test data was 1.4504, which was slightly higher than those presented in Figure 5 (a) for the training and validation data. The ROC AUC was 0.97, indicating that the model has very good discriminatory power in distinguishing between the two classes.



(a) Loss function

(b) Curve of ROC

Figure 3: Results for CelebA Mini datasets using Keypointrcnn_resnet50_fpn architectures.

Table 4 and 7 shows the confusion matrix and classification report, respectively. The confusion matrix indicates that out of 50 instances, the model correctly classified 20 women and 24 men. It misclassified 2 women as men and 4 men as women.

The classification report indicates that the model's precision, recall, and F1-score for identifying men are 0.92, 0.86, and 0.89, respectively. For identifying women, the corresponding scores are 0.83, 0.91, and 0.87. Overall accuracy is 0.88, with balanced performance across both classes.

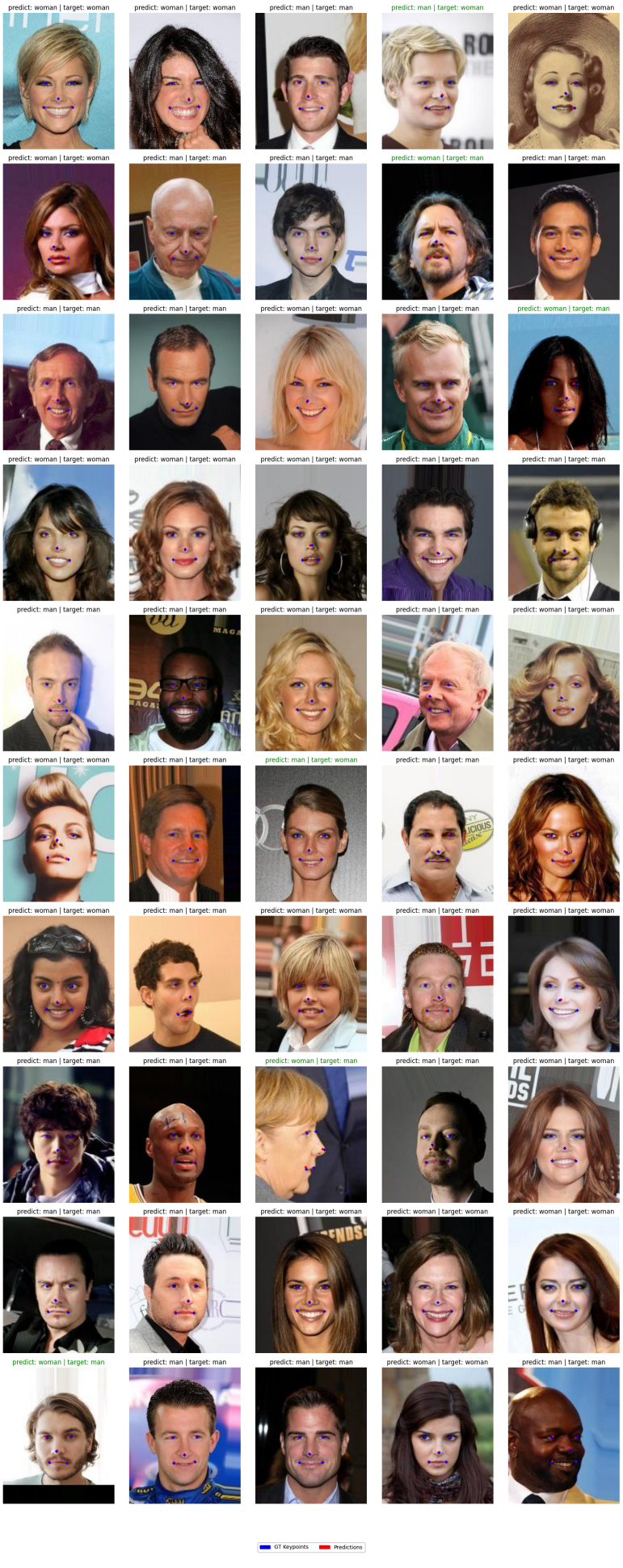


Figure 4: Predict keypoints and gender for Keypointrcnn_resnet50_fpn

Table 4: Confusion Matrix

	Woman	**Man**
Woman	20	2
Man	4	24

Table 5: Classification Report

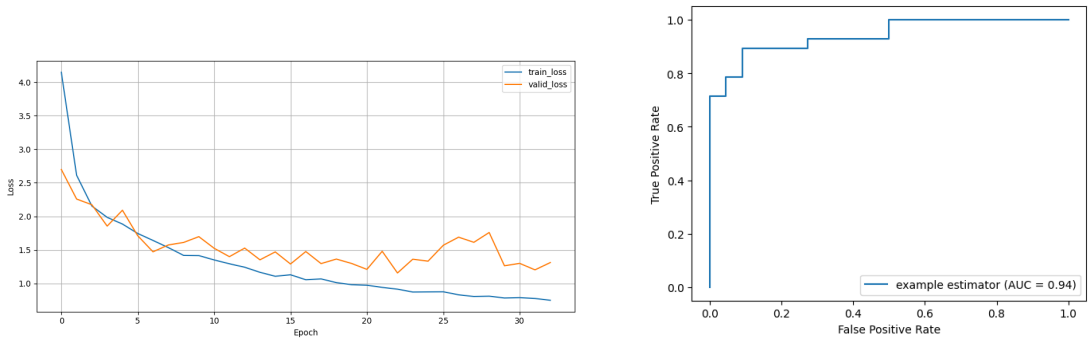
	precision	recall	f1-score	support
man	0.92	0.86	0.89	28
woman	0.83	0.91	0.87	22
accuracy			0.88	50
macro avg	0.88	0.88	0.88	50
weighted avg	0.88	0.88	0.88	50

In Figure 4, the prediction of keypoints and gender of the person can be verified. In the title of each image, the gender prediction is provided, and titles shown in green indicate incorrect predictions. Additionally, for two images, the gender label was reversed; however, the model correctly predicted what was expected. Thus, even though the label was reversed, the prediction was accurate, indicating that the model is effectively predicting gender differences.

In terms of keypoints, the model showed excellent results beyond what we expected, regardless of the head's position.

4.3 Keypointrcnn + FaceNet

In this part, we are replacing resnet_50 with FaceNet. The loss function value for the test data was 1.8442. The ROC AUC was 0.94, indicating that the model has very good discriminatory power in distinguishing between the two classes.



(a) Loss function

(b) Curve of ROC

Figure 5: Results for CelebA Mini datasets using Keypointrcnn + FaceNet architectures.

Although the ROC curve showed different values, the metrics (Table 6) for gender classification were the same for both models (Keypointrcnn_resnet50_fpn and Keypointrcnn + FaceNet). One point to highlight is that, although the number of incorrect labels is the same across the architectures, the images that were misclassified are not the same (Figure 6), except for those images where the label is reversed.

Table 6: Confusion Matrix

	Woman	**Man**
Woman	20	2
Man	4	24

Table 7: Classification Report

	precision	recall	f1-score	support
man	0.92	0.86	0.89	28
woman	0.83	0.91	0.87	22
accuracy			0.88	50
macro avg	0.88	0.88	0.88	50
weighted avg	0.88	0.88	0.88	50

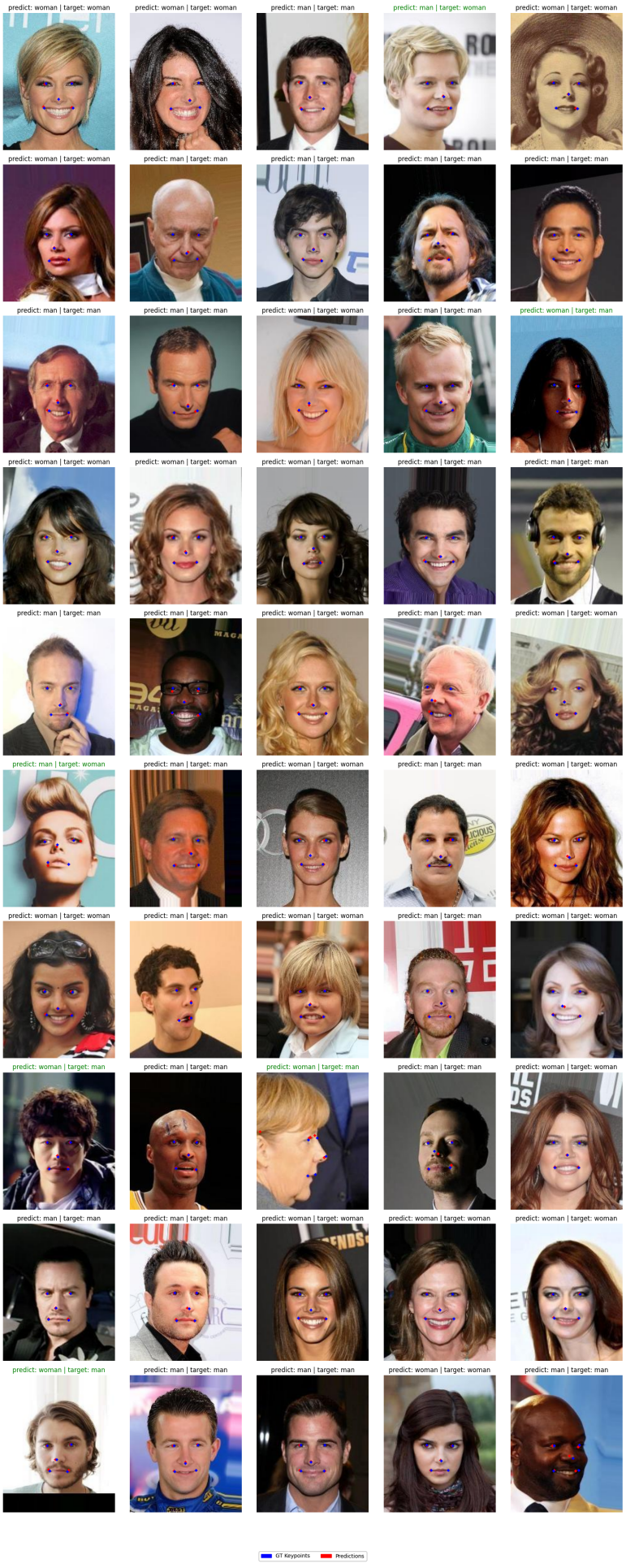


Figure 6: Predict keypoints and gender for Keypointrcnn + FaceNet model

In Figure 6, the prediction of keypoints and gender of the person can be verified. The previous architectures (Keypointrcnn_resnet50_fpn) showed better prediction to keypoints than Keypointrcnn + FaceNet.

5 Conclusion

The architecture **Keypointrcnn_resnet50_fpn** proved to be the best choice for our case, mainly because the keypoint estimation was more accurate with this model.

However, **the Keypointrcnn + FaceNet** seems quite promising. We need to spend more time finding better hyperparameters.