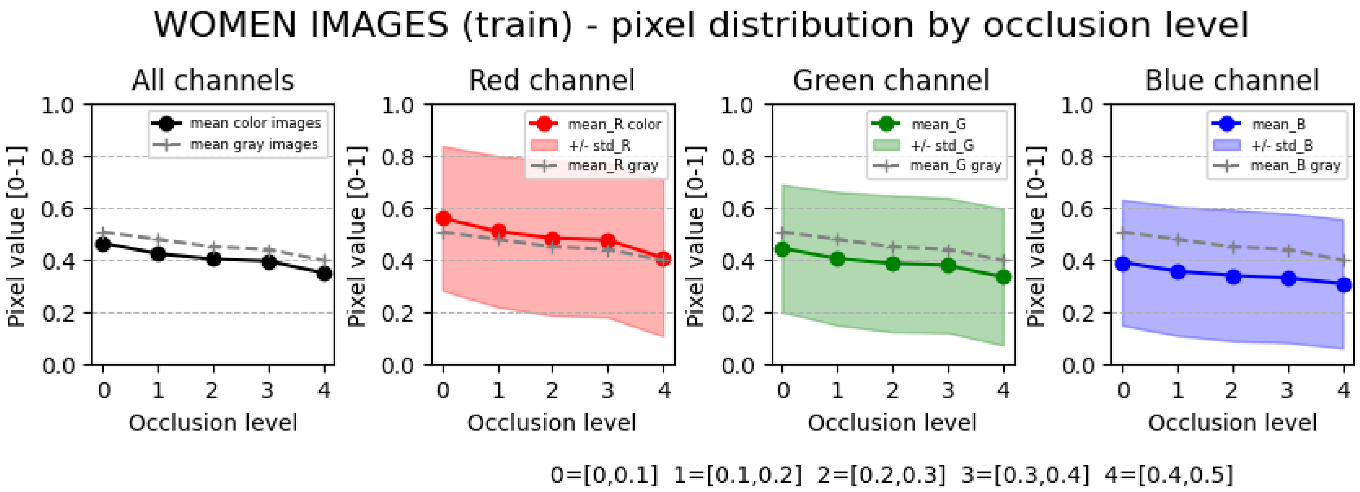
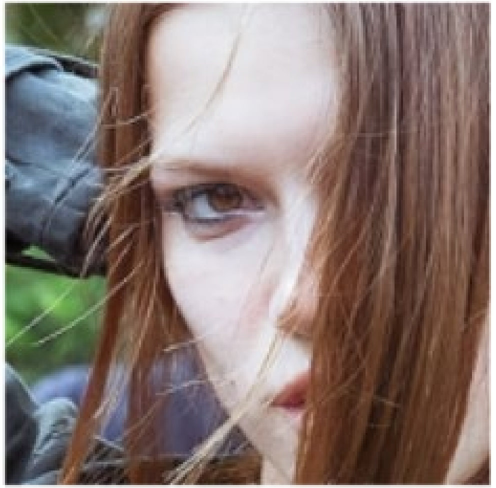
Face occlusion prediction





Data challenge 2

June 2024, MS IA & BGD, Telecom Paris

Anaële BAUDANT-COJAN, Yuchen XIA, Pety Ialimita RAKOTONIAINA

Data challenge description

- The goal of this challenge is to regress the percentage of the face that is occluded.

- Error is weighted by

Occlusion percentage

$$Err = \frac{\sum_i w_i (p_i - GT_i)^2}{\sum_i w_i}, w_i = \frac{1}{30} + GT_i$$

- We also want to have similar performances on female and male, the gender label is given for the train database.

$$Score = \frac{Err_F + Err_M}{2} + |Err_F - Err_M|$$

Our approach

1. Choice of best pretrained CNN model
2. Weighted MSE with occlusion level
3. Weighted MSE with occlusion level & gender
4. Finetuning and drop-out
5. Pixel distribution : analysis by image source (database 1, 2 or 3) and color/grayscale
6. Ensemble learning with 2 models : color and gray scale
7. Face parsing and semantic segmentation

CONCLUSION : 0.00089 final score on test dataset

1. Choice of best pretrained CNN model

- Mobilnet_v3_small – 10 epochs – 0,0031 (test)
- Resnet – 10 epochs 0,0022 (5min/epoch)
- Efficient – 10 epochs 0,00226 (11min/epoch)

⇒ Resnet-18 was best model to continue as much smaller (and thus faster) thant Efficient net.

Remarque : hyperparametres pas affiné à ce stade

2. Weighted MSE with occlusion level

Added label on occlusion level to apply weightedMSE

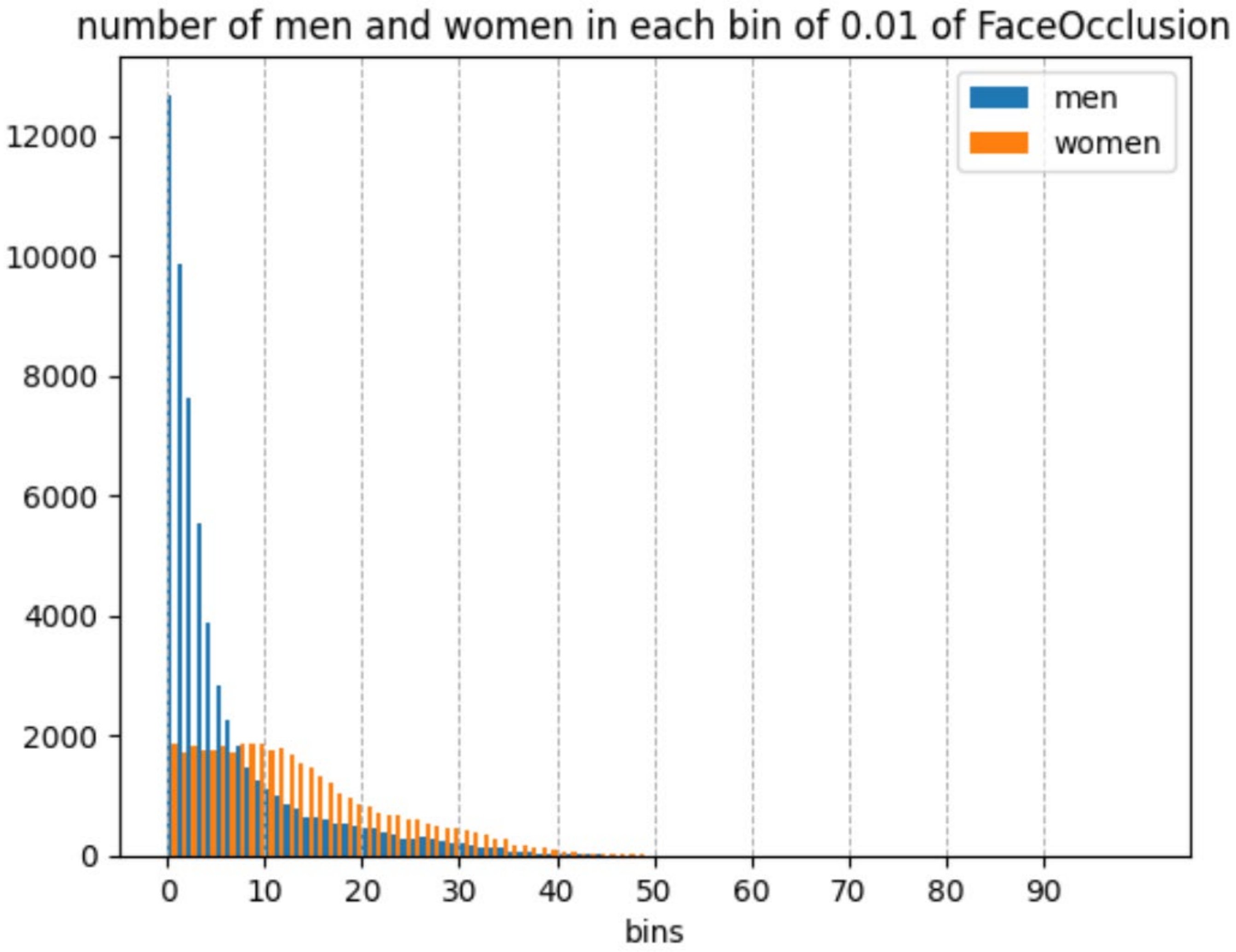
- 1st label strategy (8 classes): step of 0.05 jusqu'à 0.35
- 2nd label strategy (11 classes): step of 0.04 upto 0.4
- **3rd label strategy (10 classes): step of 0.025 up to 0.1,step of 0,05 up to 0.35**

⇒Best result 0.0017 with 20 epoches

- Weights : 1 / frequency of class

- **Using weighted MSE with occlusion levels (10 classes) to train model improved results significantly**

3. Weighted MSE with occlusion level & gender



Strategy to rebalance by gender and occlusion level

- 10 occlusion levels
- 2 gender

= 20 classes

3.1 Occlusion level + gender weights :

Added label on occlusion level & gender to apply weightedMSE for 3 epochs

- 0.0028 {'male': 1, 'female': 1.5} corresponding to imabalance of dataset: 60% male / 40% female
- 0.0025 {'male': 1, 'female': 1.1}
- 0.0031 {'male': 1, 'female': 1.2}
- 0.00188 {'male': 1, 'female': 1.05}
- 0.0035 {'male': 1, 'female': 1.01}
- 0.0036 {'male': 1, 'female': 1.04}
- 0.00236 {'male': 1, 'female': 1.06}
- 0.003 {'male': 1, 'female': 1.055}
- 0.0037 {'male': 1, 'female': 1.08}

Increasing epochs didn't improve score : 0.003 with 10 epochs, 0.0024 with 20 epochs

Combined occlusion and gender weights improved to 0.0018 (but instable)

3.2 Ground-truth ponderation on top of occlusion level and gender weights

- Added weights as in « final score » (1/30 + GTi) to **occlusion level** weights (used with mean ground truth of each class)
  - 0.00525 for 3 epochs

- Added GT weights to **occlusion level and gender** wieghts :
  - 0.0033 with {male : 1 female 1.5} (for memery 0.0028 without GT)
  - 0.00317 with {male : 1 female 1.1} (for memery 0.0025 without GT)
  - 0.00409 with {male : 1 female 1.06} (for memeroy 0.0023 without GT)
  - 0.0036 with {male : 1 female 1.05} (for memery 0.0018 without GT)

=> Adding ground truth ponderation to weights dit not improve

4. Finetuning and drop-out

- Finetuned Resnet 18 with weightedMSE on occlusion level :
  - 0.017 with 3 epochs
  - 0.0114 with 50 epochs

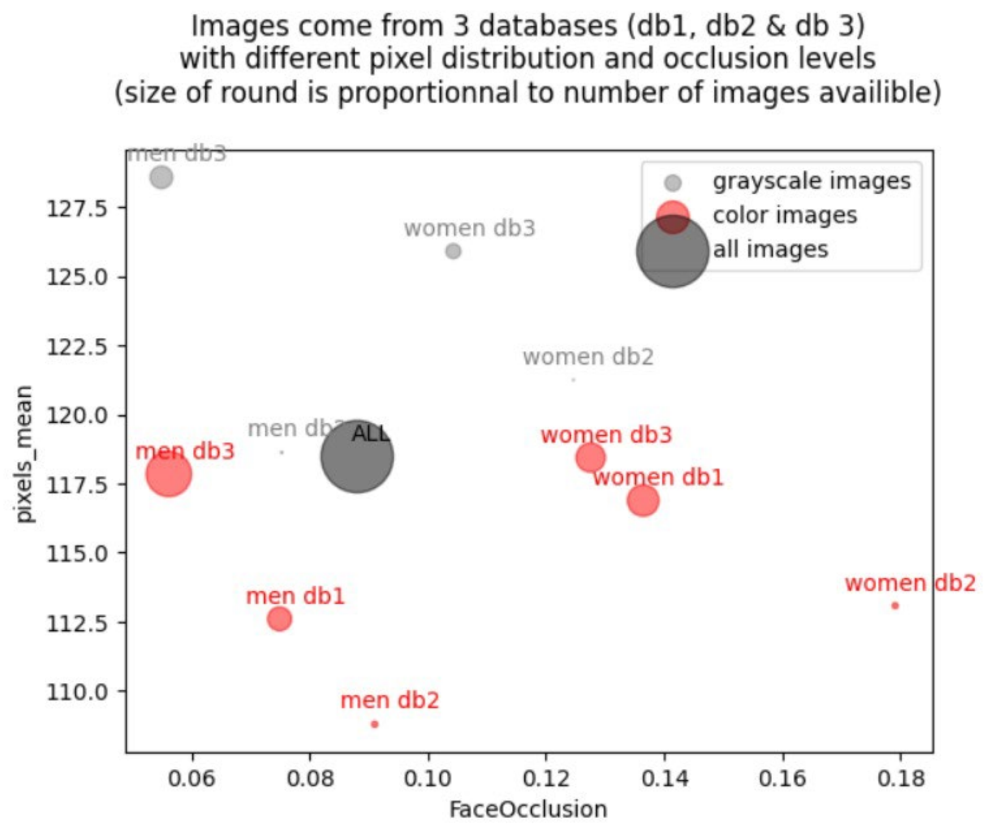
- Dropout with same model
  - 0.00455 with 3 epochs
  - 0.00457 with 10 epochs
  - 0.00256 with 50 epochs

- Finetuned Resnet 18 with weightedMSE on occlusion level + gender wieghts:
  - 0.015 with 3 epochs
  - 0.024 with 10 epochs
  - 0.034 with 50 epochs

- **=> Due to poor results and GPU constraints, finetuning & dropout strategies were abandonned**

5. pixel distribution : analysis by image source (database 1, 2 or 3) and color/grayscale

COLOR vs GRAYSCALE IMAGES PIXEL DISTRIBUTION BY SOURCE AND OCCLUSION LEVEL

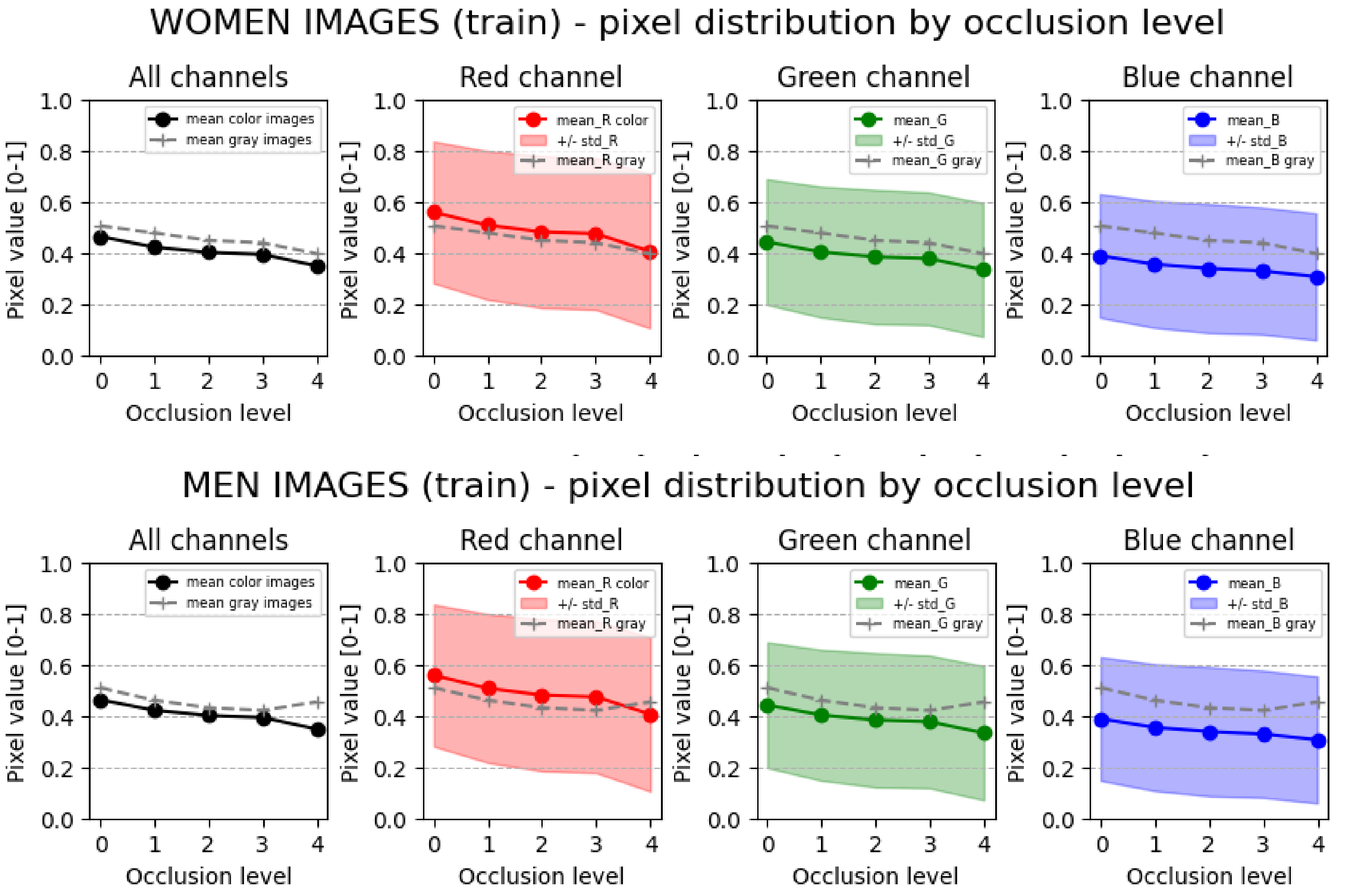


- Distribution is different among different databases
- Distribution is different between grayscale and color image
- Correlation between pixel mean and occlusion level for color images (inverse for grayscale images)
- Global normalisation would maintain distirubtion disparities

Analysis for each R G B channel

- Distribution is extremely different between channels for color and grayscale images
- In particular all channels all equal for grayscale images
- Red channel is much higher than green and blue channels for color images

=> It does not make sense to use the same model for color and grayscale images



0=[0,0.1] 1=[0.1,0.2] 2=[0.2,0.3] 3=[0.3,0.4] 4=[0.4,0.5]

6. Ensemble learning with 2 models : color and gray scale

- Created a function to identifiy grayscale images in test dataset, and predict with color or grayscale model accordingly
- Re-computed weights for each models by level of occlusion for color and grayscale dataset
- Trained color model on 87k images, and grasycale model on 14k images
-

$\Rightarrow$ 0.0016 with 3 epochs,

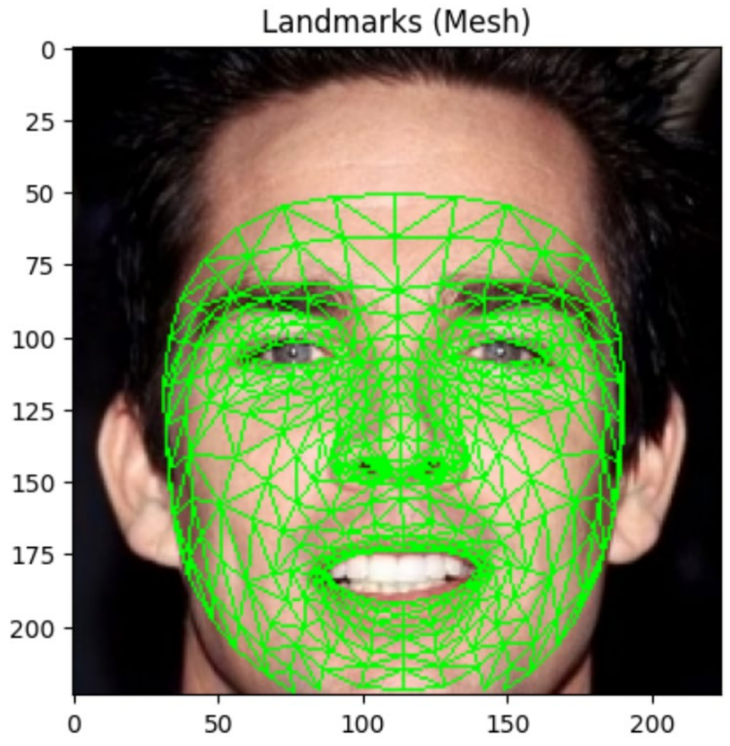
$\Rightarrow$ 0.0013 with 10 epochs

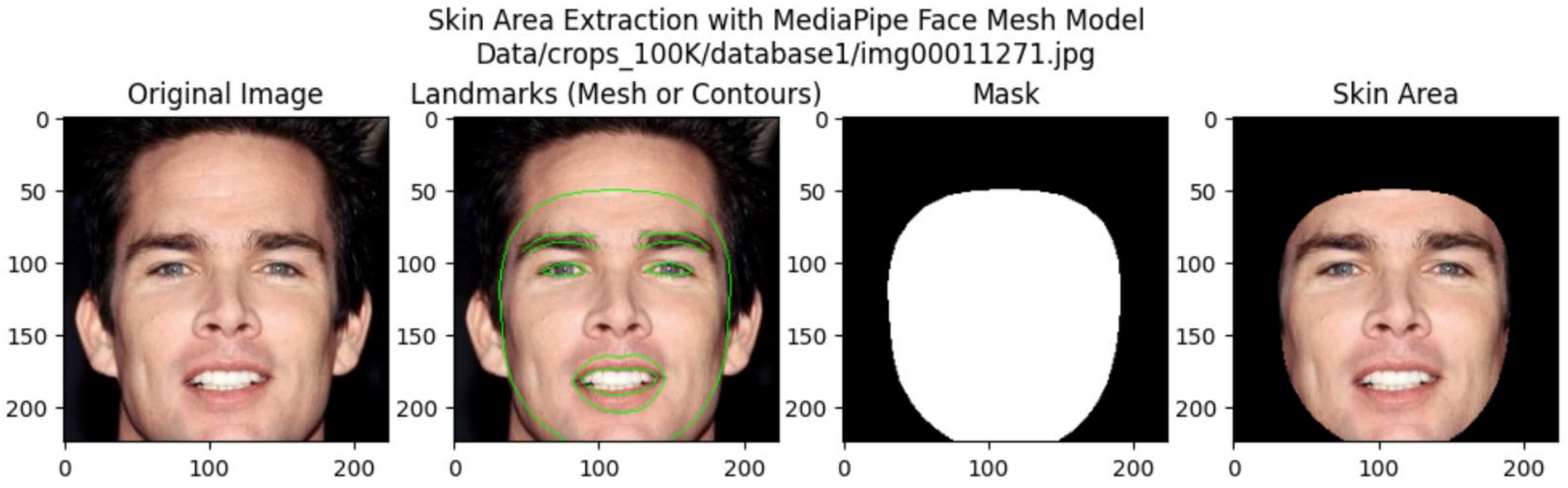
$\Rightarrow$ **0.00089 with 30 epochs**

Ensemble learning with 2 models gave best results, with a weighted MSE loss function (rebalancing only occlusion level)

7. Face parsing and semantic segmentation

- We processed images with mediapipe models to extract 468 landmarks (mesh) coordinates (x, y, z)
- Face detected for 98,5% of images
- **Extermely complicated to train NN on masked images or landmark coordinates (not possibile on Resnet 18)**





Semantic face segmentation (face parsing)

- We explored 2 pretrained models :
  - Easyportrait
  - BiSeNEt
- **Idea :** regroup non-occluded zones (skin, lips, nose, brows, eyes) and calculate occlusion by difference with mediapipe mask
- **Implementation was challenging and computation long,**

=> approach was abandonned

CONCLUSION

- **Best model (0.00089) took into account 2 main disparities:**
  - **Imbalance between occlusion levels in the data set**
  - **Pixel distribution disparity between color images and grayscale images**

- However **gender** is not well accounted for in our best result, the model was instable, and we did not have time (due to long computation times) to fine-tune hyperparameters (learning rate, male-female ratio, batch size…)

- We would have liked to try **ensemble learning with 4 models** :
  - Color male / cole female
  - Grayscale male / grayscale female