MATHÉMATIQUES ET INFORMATIQUE
Sciences
Université Paris Cité

Université de Paris Cité
UFR Mathématiques et Informatique

Image Processing: TP3

Master 2 Vision et Machine Intelligente

Noureddine Bertrand

Année universitaire 2024-2025

Question 1 : Open an xml file, and try to understand the information it contains and describe it. From this xml, can you guess what bounding box format is used in this dataset ?

The format used is (x,y) of top left and bottom right corners.

Question 2 : We are going to learn from this data. By looking at it, can you find possible biases in it ?

Yes, there are multiple images of the same person (e.g., 655, 609, 561, 757). By themselves, these repetitions may not introduce significant biases, but given the size of the dataset and the distribution of classes, they could be problematic (see Questions 3 and 4).

Additionally, we can observe that most faces are facing directly at the camera, indicating low variance in picture angles.

There are also images of poor quality (high blur), aggregations of images, or very small faces within a dense crowd.

Question 3 : What is the total number of samples that we have ? Does that seem sufficient to train a model ? What strategy could we use for training ?

We have 854 samples. If the task is to train a model from scratch, this may be problematic ; however, if we are fine-tuning an existing model, it could be sufficient.

If we determine that we need more data, we can use data augmentation techniques such as scaling, cropping, rotation, flipping, noise injection, or color alterations. However, we must be cautious, as some augmentations may not be relevant given the task. Additionally, we could implement cross-validation to optimize the training process.

Question 4 : Propose and implement a methodology to check the labels' distribution. Discuss your findings.

From this distribution, we can observe that the *with_mask* (3232) class is significantly more represented than the other two classes. This imbalance is problematic for the model, as it will have substantially fewer samples for the *without_mask* (717) and *mask_worn_incorrectly* (123) classes.

While data augmentation could be applied to increase the number of samples in the underrepresented classes, the disparity between the class sizes is too great to be entirely compensated. This imbalance may lead to the model being biased toward the with_mask class, potentially diminishing its performance on the other classes.

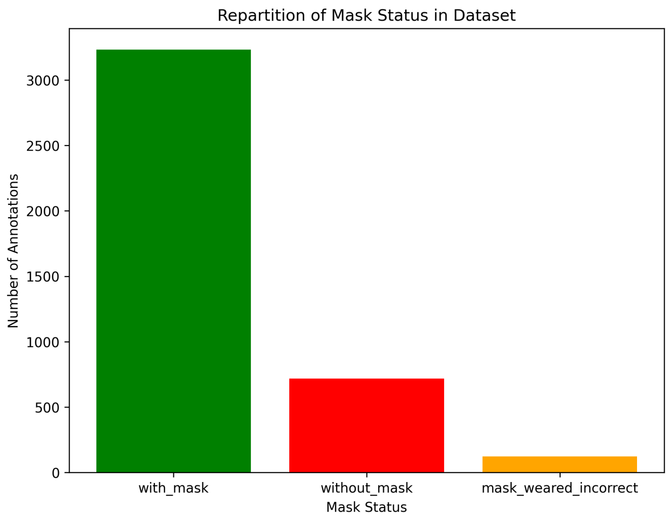

FIGURE 1 – Class Distribution

Question 5 : How did you choose your hyper-parameters ?

I performed a manual search. Given the time and the task, grid search, cross validation or more sophisticated approach like Bayesian Optimization were not considered.

I first tried to over-fit the model than slightly decrease learning rate or increase momentum to see in which direction goes the model loss.

I had the best result with a learning rate of 0.02 and a momentum of 0.1, for the number of epoch and due to the limitation of GPU usage on Colab, I tested up to 10 epoch.

The validation loss is lower than the training loss. This might be caused by the fact that the validation set is "simplier" than the training.

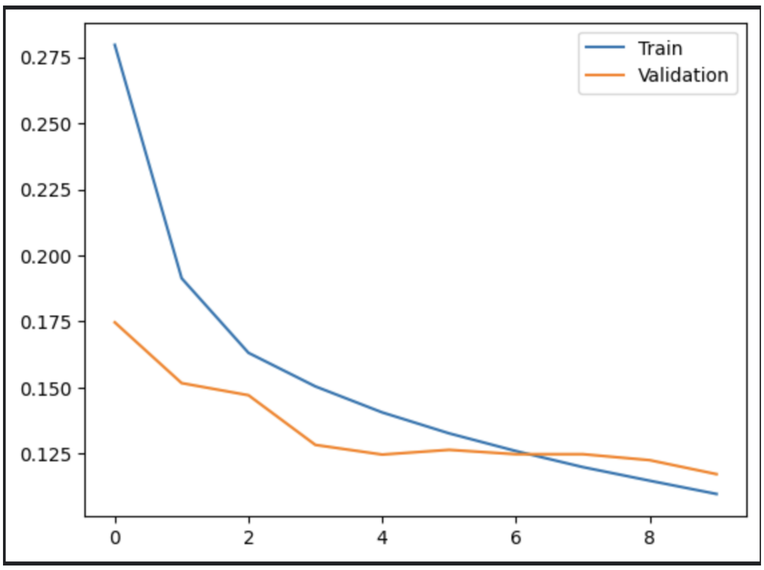

FIGURE 2 – Score

Question 6 : Evaluation

For the evaluation, I used standard metrics : F1 score, precision, recall, avg F1 score, avg accuracy. All were computed based on an IoU threshold of 0.5 and a score threshold of 0.5.

As we have seen on the first lab, the score for the the under represented third class is bad.

The average F1 score is heavily penalize by that.

We can think that our model have issue finding the *mask_ weared_ incorrectly* class, but by looking at the confusion matrix we see that this class is simply not represented in the dataset.

We could also have tried to compare the model with other threshold or have a more balanced dataset.

```
--- Classification Metrics ---
Class           Precision       Recall          F1 Score
------------------------------------------------------------
with_m          0.9356          0.8640          0.8984
without_mask    0.7647          0.8211          0.7919
mask_weared_inc0.0000           0.0000          0.0000

--- Summary ---
Average F1 Score          : 0.5634
Overall Accuracy          : 0.7800
```

Figure 3 – Score

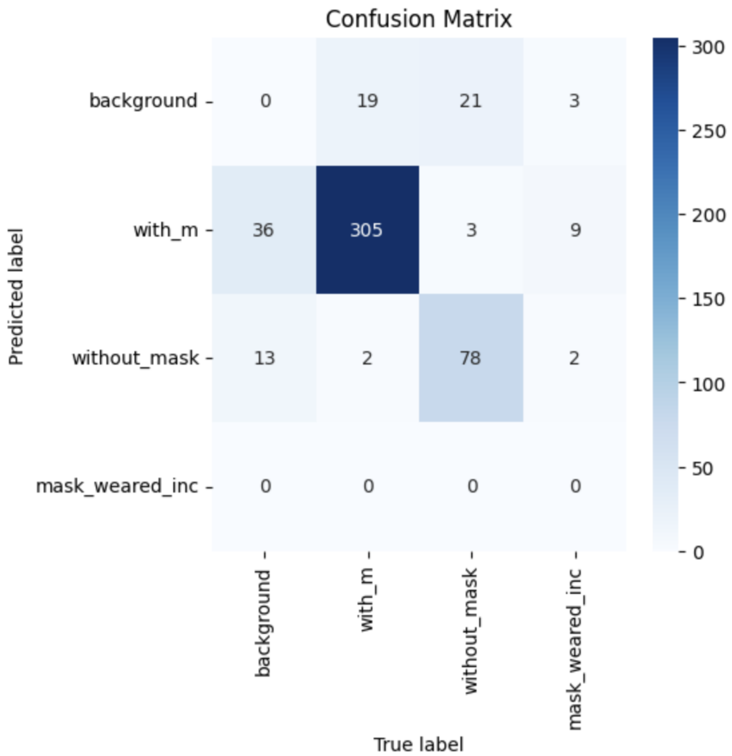

Figure 4 – Confusion Matrix