



IP PARIS

BGDIA704

Restitution data challenge

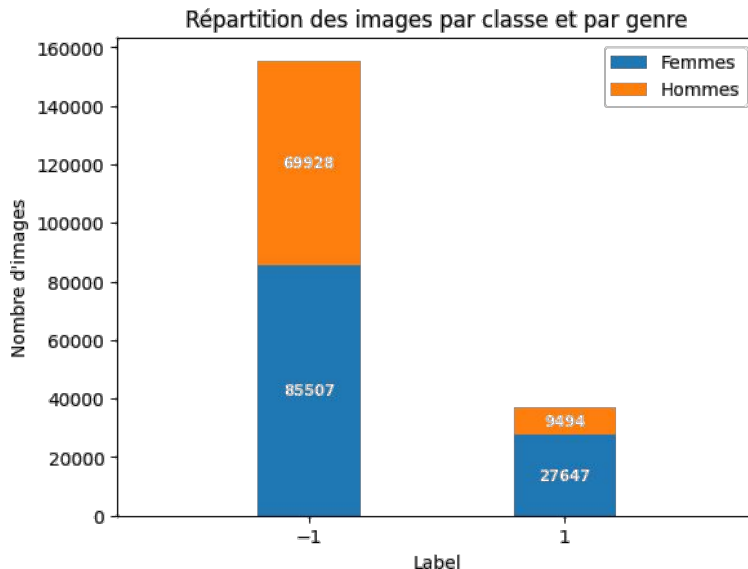
12 juin 2023

Enjeux

- Classification binaire d'images
- Labels non explicites
- Jeu de données déséquilibré
- Réduction de biais (métrique d'évaluation)

Méthodologie

- Exploration de modèle état de l'art
 - ViT, modèles multimodaux (CLIP, Align, etc.)
- Réduction de biais via fonction de loss ou oversampling
- Test de modèles entraînés via Self Supervised Learning



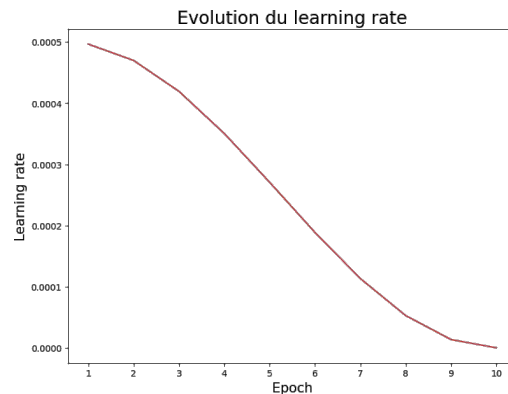
Entraînement

Finetuning d'un modèle DINO

- Entraîné sur ImageNet1k (HuggingFace)
- Architecture basée sur ViT
- Taille de patches 8 x 8
- 6 attention heads
- Version small (21M de paramètres)

Scheduler

- Cosine with warmup (5% du nombre de steps)

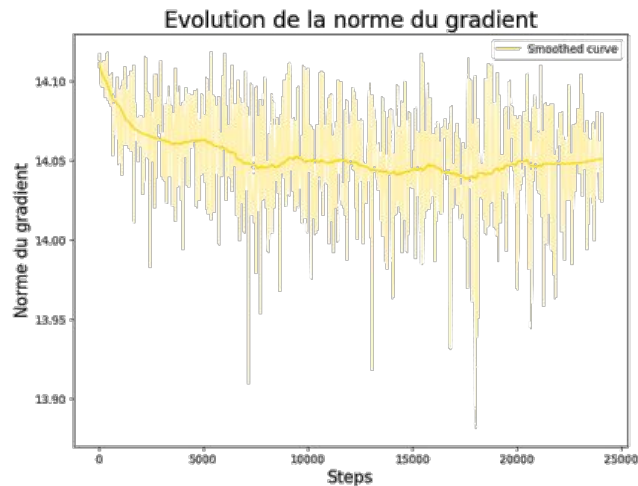


Optimizer SING

- AdamW avec normalisation du gradient (layerwise)

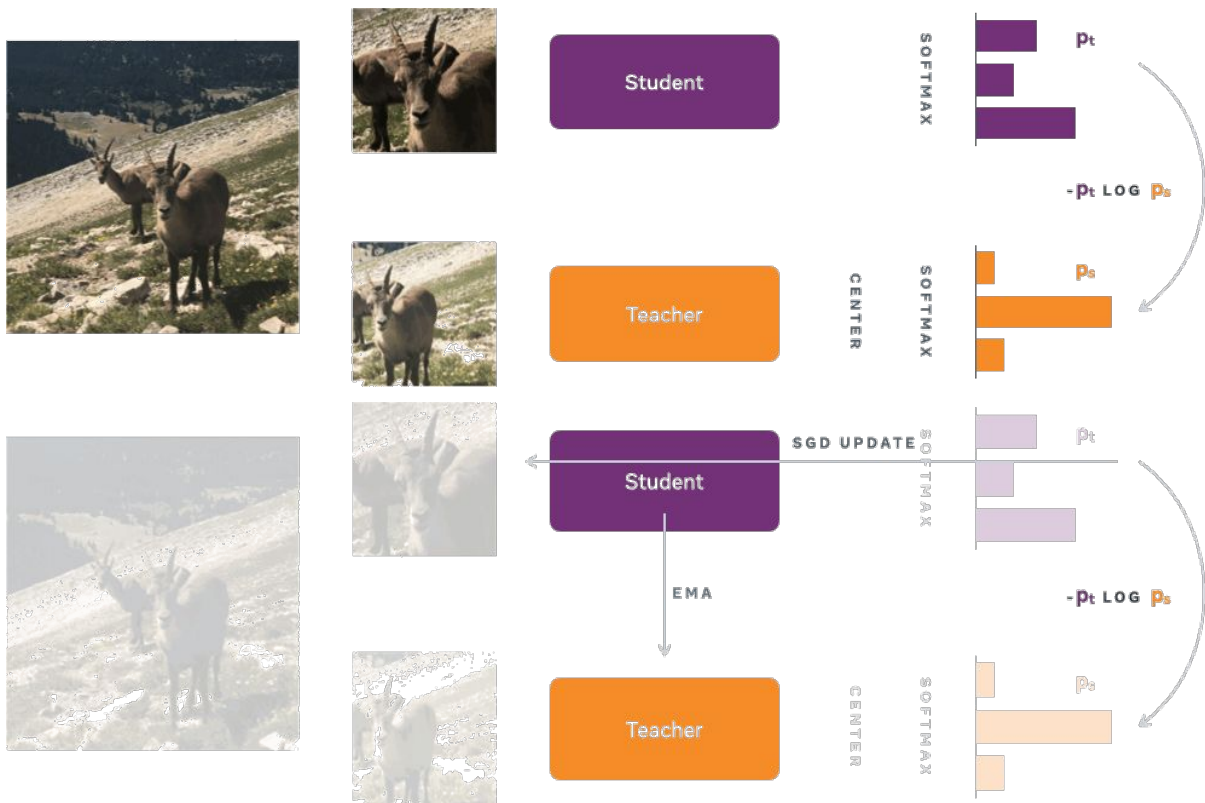
Hyperparamètres

- Batch size 64
- Learning rate 5×10^{-4} / Weight decay 5×10^{-3}
- Label smoothing 0.1
- Split train / test : 0.8 / 0.2 avec stratification sur classe ET genre



Focus sur le modèle

DINO (self-Distillation with NO labels)



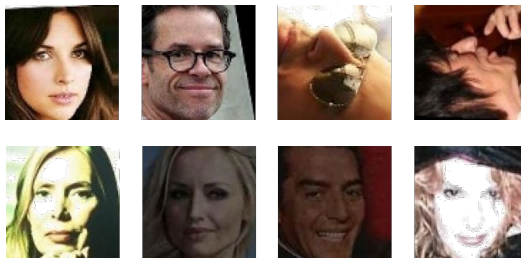
Self-Supervised Learning

- Same architecture for teacher and student
- Loss : difference of outputs between master and student
- Stop gradient for teacher : update through EMA

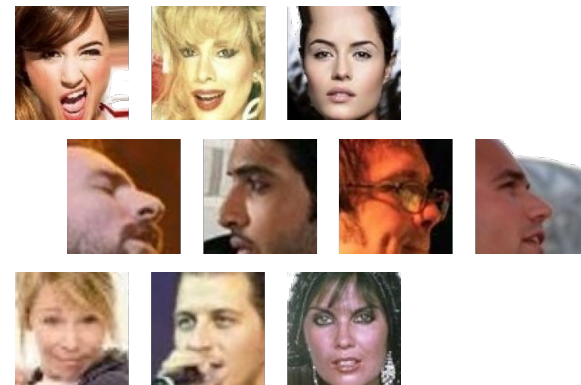
Data Augmentation

Images altérées

- Rotation
- Luminosité



- Déformation
- Crop
- Flou



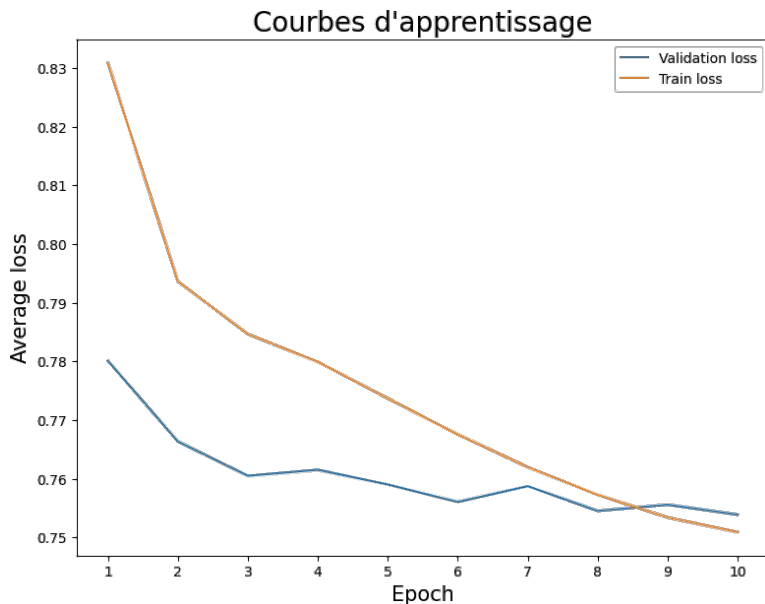
Transformations appliquées (bibliothèque FFCV-SSL)

- RandomResizeCrop
- RandomHorizontalFlip
- Rotate
- RandomErasing
- RandomColorJitter
- GaussianBlur
- Normalize

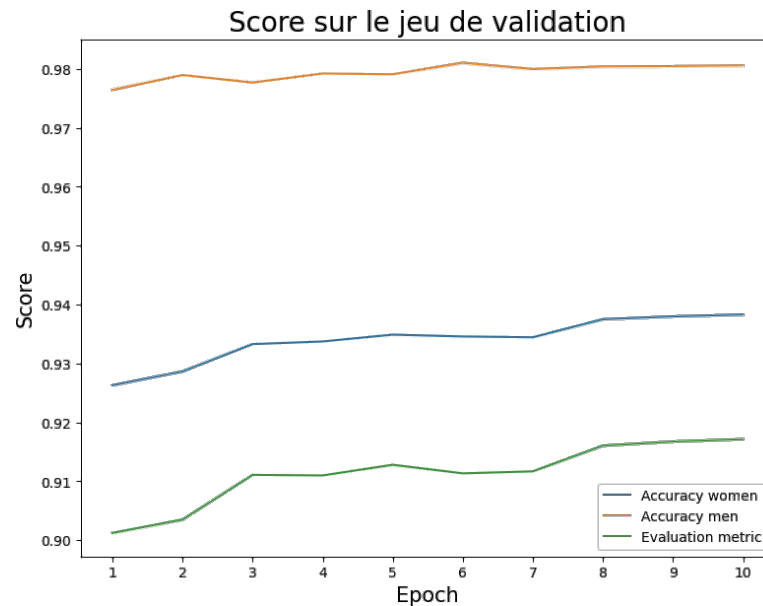


Observations

- Overfit rapide (8 epochs) malgré label smoothing et weight decay
- Gros écart entre distribution de train et de test



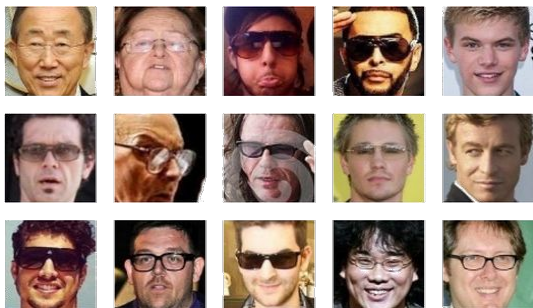
- Tentative de réduction de biais inefficaces



Interprétation des labels

Visualisation des données

Hommes classe 1



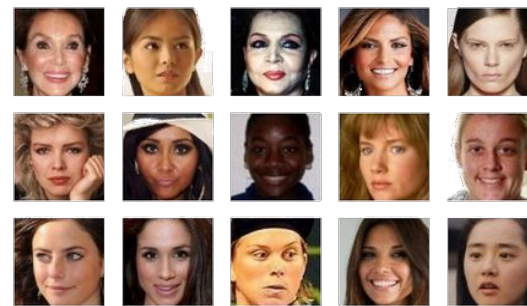
Femmes classe 1



Hommes classe -1

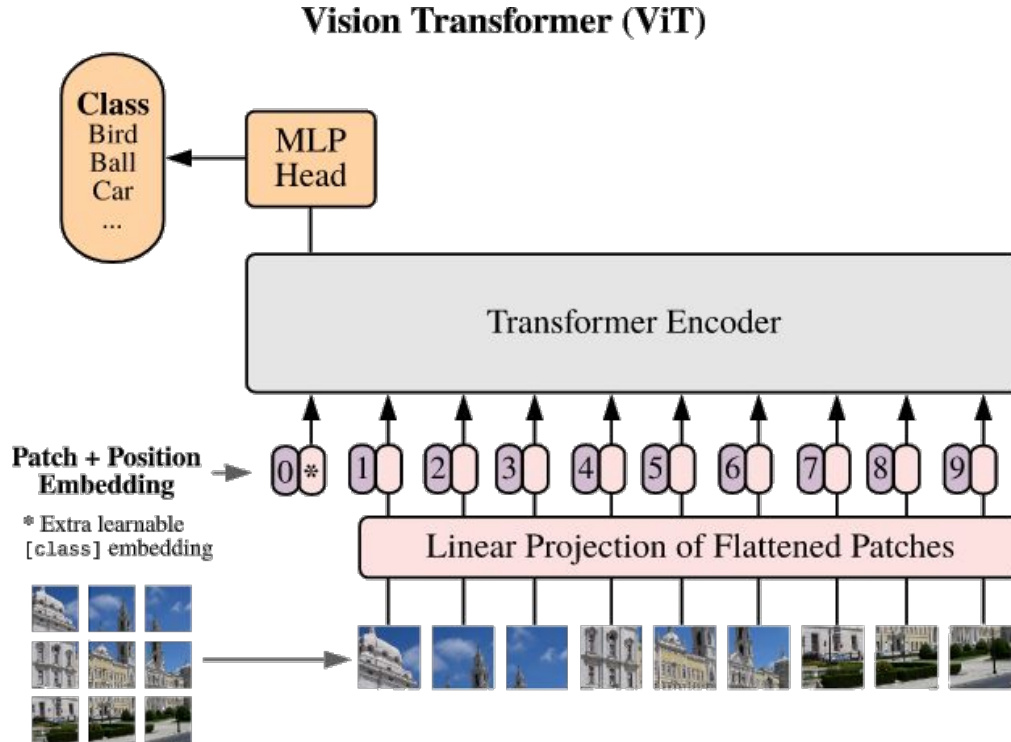


Femmes classe -1

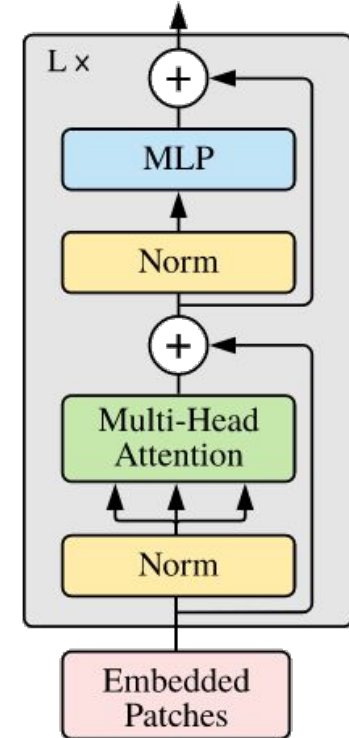


Interprétation des labels

Attention maps (1)

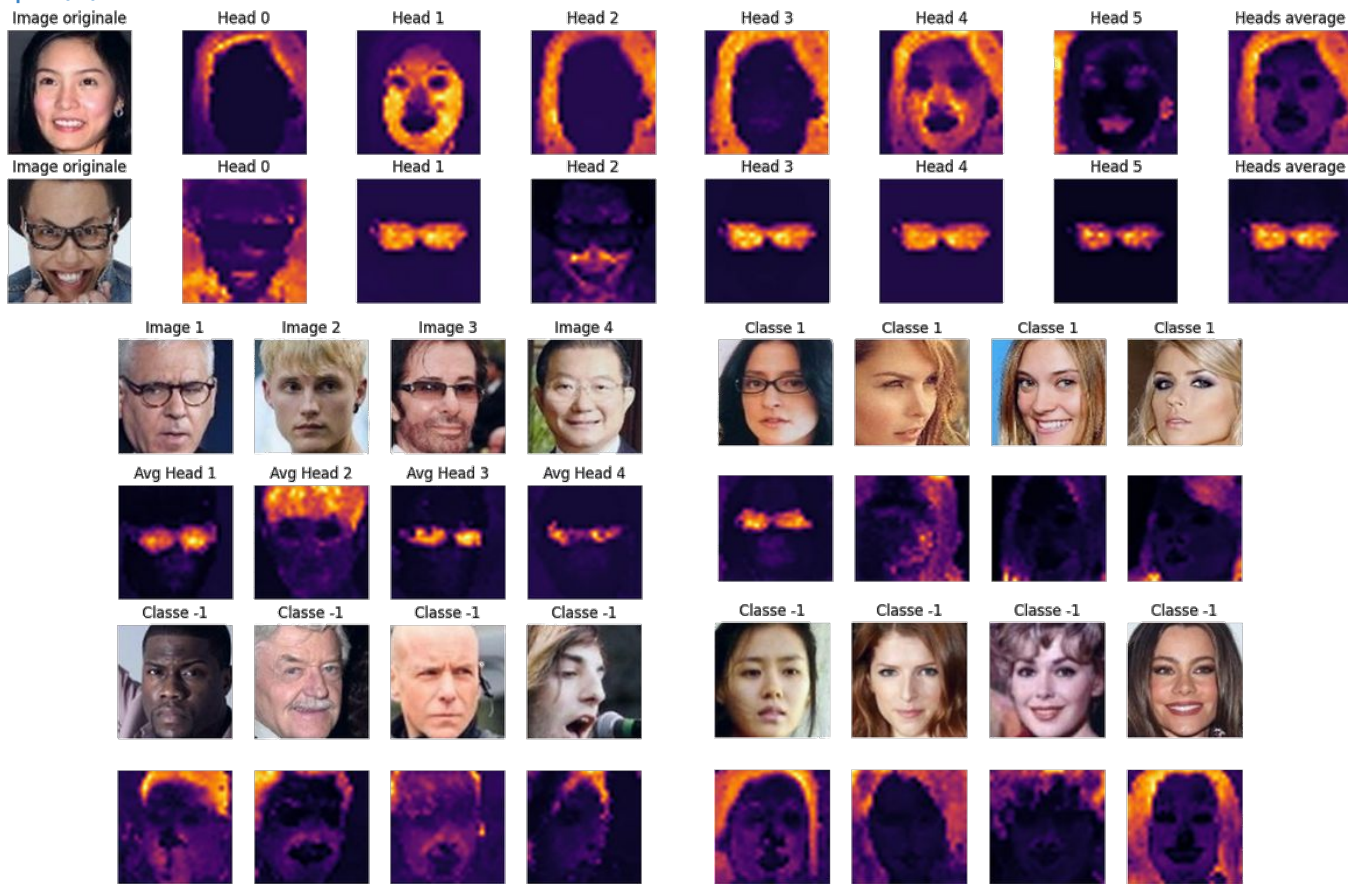


Transformer Encoder



Interprétation des labels

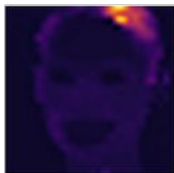
Attention maps (2)



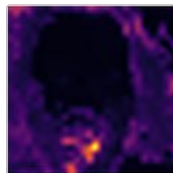
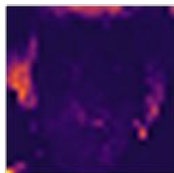
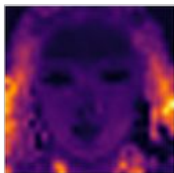
Interprétation des labels

Attention maps : errors

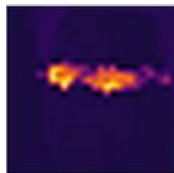
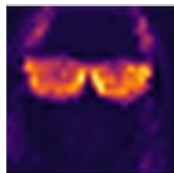
Prédiction Classe 1



Prédiction Classe -1



Prédiction Classe 1



Prédiction Classe -1

