




Jedha


Data Science Bootcamp

 **Featured Code Competition**

Human Protein Atlas - Single Cell Classification

Find individual human cell differences in microscope images

\$25,000
Prize Money

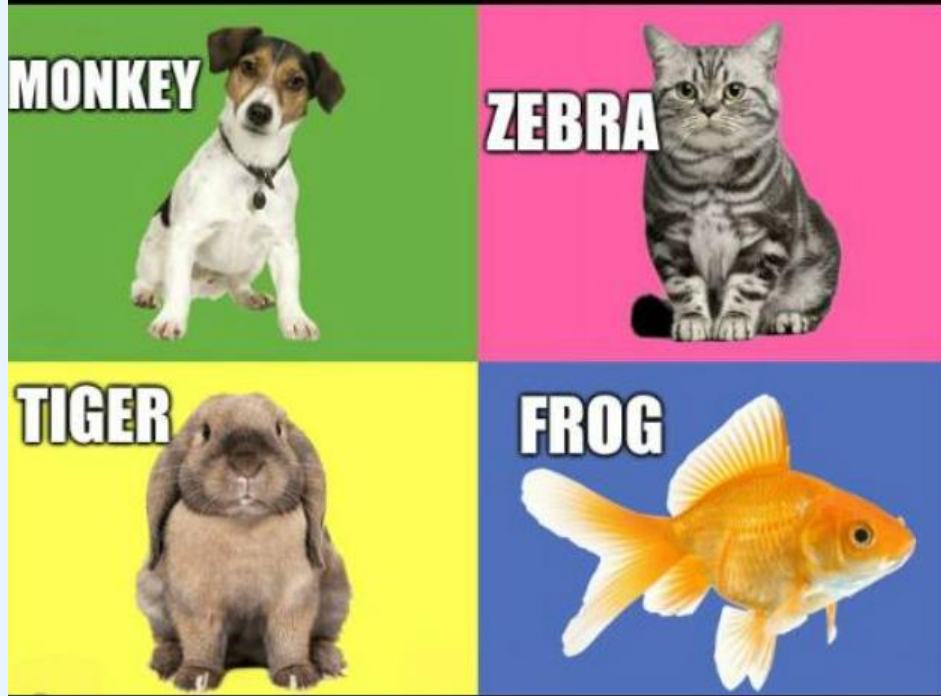
 Human Protein Atlas · 383 teams · 2 months to go (a month to go until merger deadline)

L'équipe: Catherine Duclos
Olga Kurnosova
Hadrien Grimaud

[Submit Predictions](#)



WHEN YOU TEST YOUR FIRST IMAGE CLASSIFIER MODEL





Qui sommes
nous/pourquoi
ce project?

Olga

- Background traitement du signal
- Secteur R&D
- Intérêt pour computer vision

Catherine

- Background Informatique médicale
- Secteur Santé
- Intérêt pour le deep learning

Hadrien

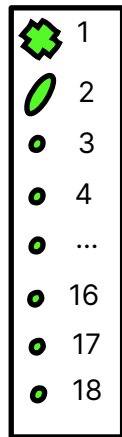
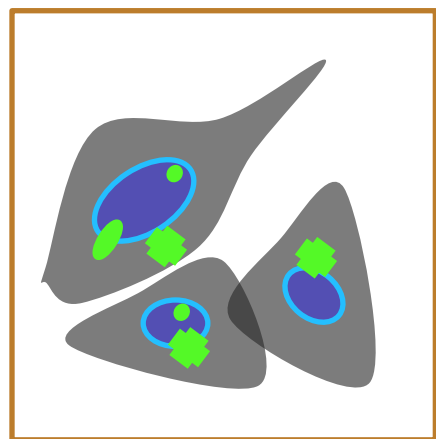
- Background Math/Physique
- Secteur Banque
- Intérêt pour les problèmes complexes



Expérience scientifique, unis par
JEDHA, envie de défi



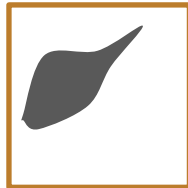
Définition du problème: classification multi-label + segmentation



Label 1 +



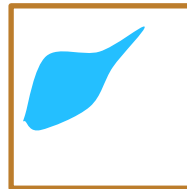
Label 2 +



Label 3 +



kaggle
TEST:



Your score: 100 %



INPUT:

21 806 images

TARGET

18 labels

EXPECTED OUTPUT: label + binary mask



Les étapes de projet:

Data analysis:

Baseline solution

La suite

- Exploration des données
- Identification des problèmes
- Le preprocessing de données
- Le choix du modèle de deep learning
- Les métriques sur le test set
- L'interprétation de résultats
- Les pistes d'amélioration



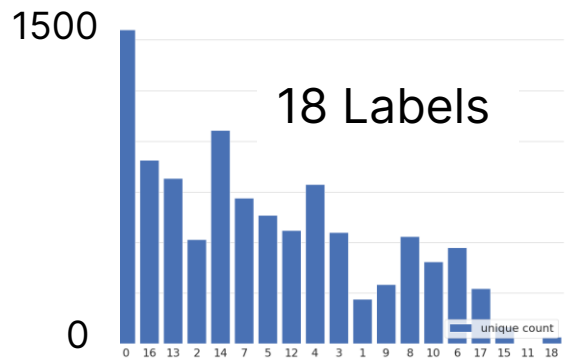
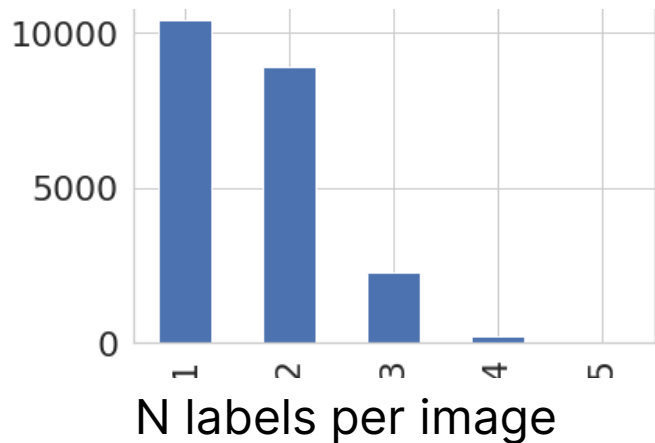
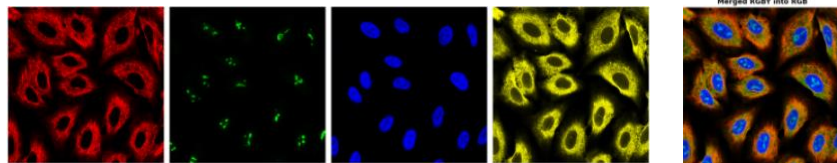
Data analysis:

Baseline solution

La suite

21186 img, 160Go des données en total,

Train vs Test



Unique count labels per images



Data analysis:

Baseline solution

La suite

— 18 modèles de classification binaire. Entraînement sur :

- Un dataset monolabel équilibré
- Augmenté pour atteindre 7000 images
- Diversifié

— Transfer learning appliqué (réseau neurones)

- EfficientNetB0 weights
- Poids de Imagenet
- Dernière couche dense



Data analysis:

Baseline solution

La suite

— Réseau pré-entraîné pour faire segmentation de cellules:

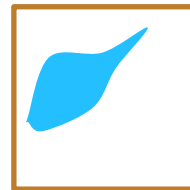
→ HPA Cell Segmentator

- 1 masque binaire par image
- Technique de compression

✱ Label 1 +



◌ Label 2 +



● Label 3 +



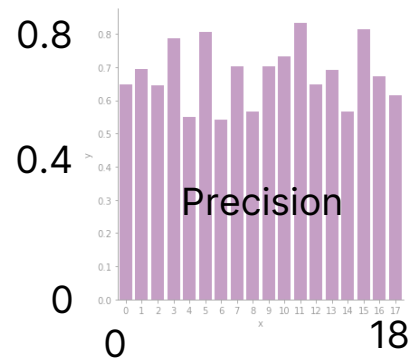
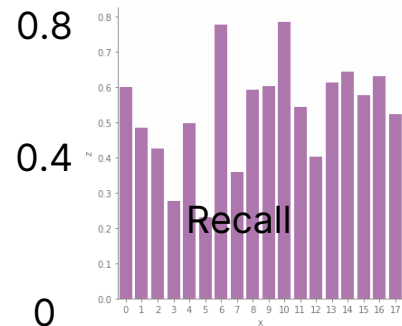
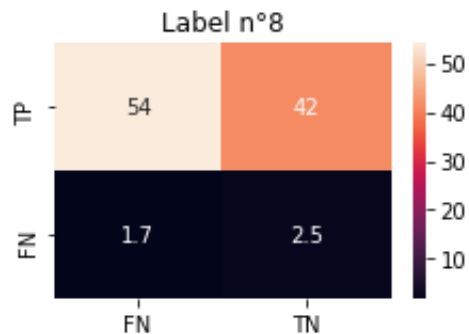
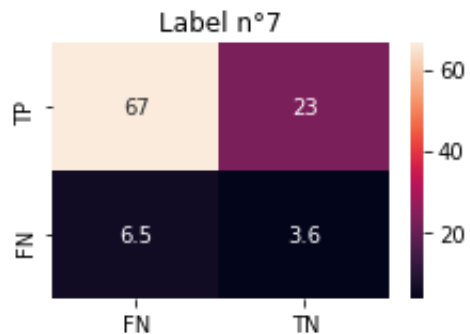
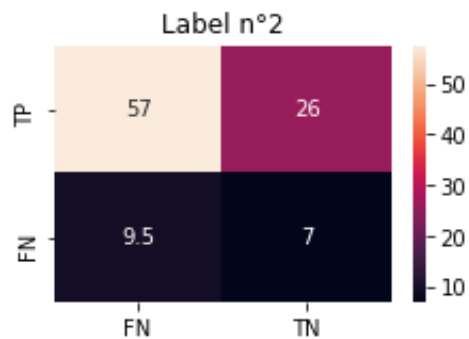
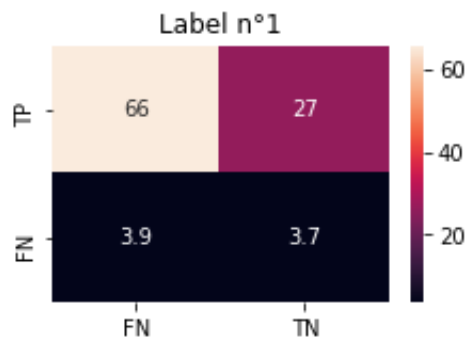
OUTPUT VS EXPECTED OUTPUT



Data analysis:

Baseline solution

La suite





Data analysis:

Baseline solution

La suite

- Optimisation du temps de calcul (DataGenerator)
- Enrichir le dataset
- Essayer des modèles multi-labels.
- Entrainement sur totalité des images
- Affiner la Data Augmentation des images (paramètres)
- Optimiser les (hyper)-paramètres du modèle
- Essayer d'autres modèle pré-entraînés plus complexes



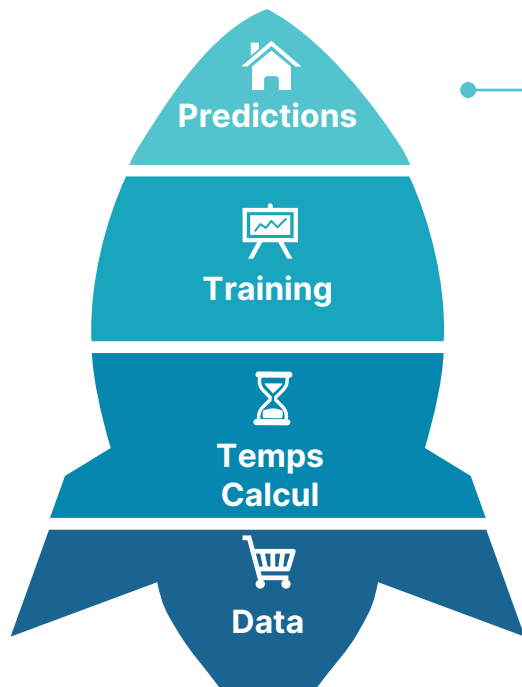
Jedha

Merci,
à bientôt !





Piste d'améliorations



PREDICTIONS

Multi-Masques (niveau cellulaire)



TRAINING

Différents Modèles + Optimisation (Hyper)-Paramètres



TEMPS DE CALCUL

DataGenerator, Numba



DATA

Images ++, Data Augmentation

