

**Ce rapport
présente les différentes étapes
de l'expérimentation pour la
compétition kaggle siim
covid19.**

Introduction	3
Expérimentation	4
Collecter les données	4
Analyser les données	4
Traiter les données	8
Modéliser avec yolov5s	8
Modéliser avec vgg16	11
Submission	13
Conclusion	13

Introduction

Nous allons saisir l'opportunité de mettre à profit nos compétences acquises au cours de cette formation en participant à une compétition Kaggle active.

Kaggle est une plateforme collaborative proposant de résoudre des problèmes de data science et qui dispose d'un système de rémunération sous forme de compétition.



On souhaite réduire la mortalité du COVID-19 en traitant plus rapidement les patients. Les radiographies peuvent être obtenues en quelques minutes contrairement aux tests moléculaires.

Mais le COVID-19 ressemble beaucoup à d'autres pneumonies sur les radiographies pulmonaires. L'étude d'un patient peut nécessiter plusieurs radiographies pour identifier le type de la pneumonie.

On nous demande de classer la pneumonie du patient en 4 catégories :

1. négative
2. typique
3. indéterminée
4. atypique

Puis d'identifier pour chaque radiographie les zones d'intérêts montrant une opacité responsable de la maladie.

Il faut donc réaliser une classification et une localisation d'objets sur des images.

En ce qui concerne la soumission de nos modèles pour l'évaluation, l'hôte de la compétition demande la réalisation d'un notebook, celui-ci n'a pas d'accès à internet. La sortie de notre notebook doit contenir les prédictions de classification et d'object detection dans un fichier commun spécifique submission.csv

Les différentes étapes de l'expérimentation sont disponibles sur kaggle publiquement :

Your Work Shared With You Bookmarks Recently Run ▾

	[part3] siim-covid19-submission	1	...
	Updated 1h ago		
	Score: 0.058 · 0 comments · SIIM-FISABIO-RSNA COVID-19 Detection +3		
	[part1] siim-covid19-model-image-yolov5	1	...
	Updated 4h ago		
	0 comments · SIIM-FISABIO-RSNA COVID-19 Detection +2		
	[part2] siim-covid19-model-study-vgg16	1	...
	Updated 16h ago		
	0 comments · SIIM-FISABIO-RSNA COVID-19 Detection +2		
	[part0] siim-covid19-first-look-resized-512px	1	...
	Updated 19h ago		
	0 comments · SIIM-FISABIO-RSNA COVID-19 Detection +1		

Expérimentation

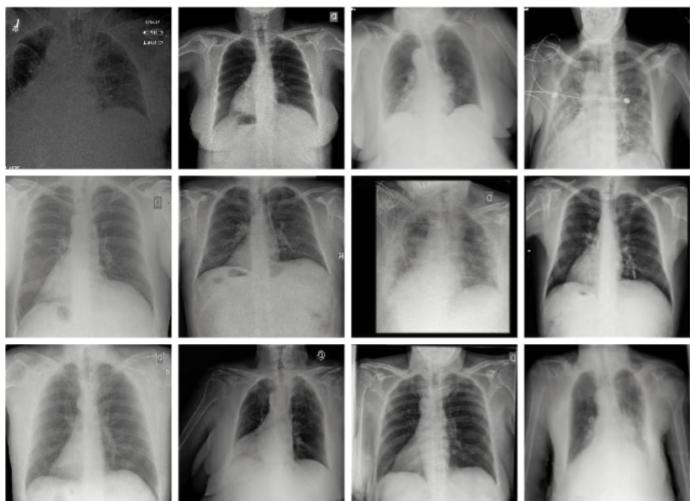
Collecter les données

Kaggle propose un environnement de développement avec un accès direct aux données sur leur serveur.

data explorer

119.68 GB

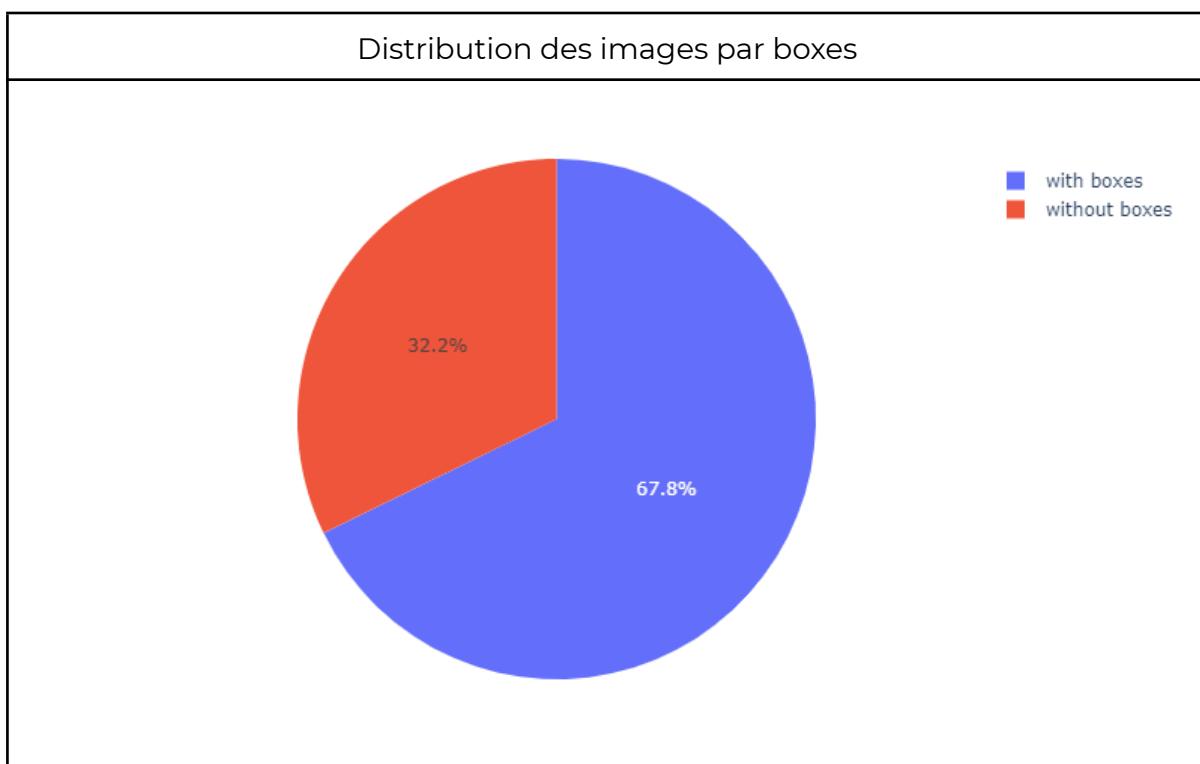
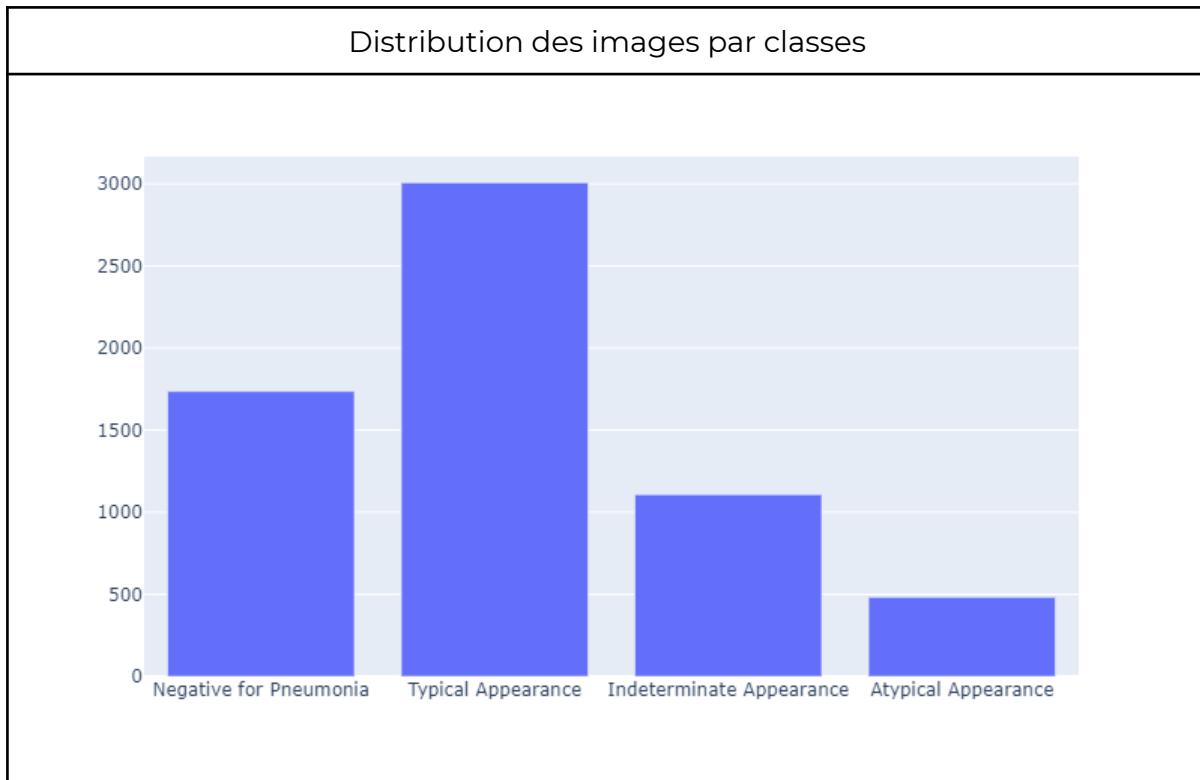
- ▶ test
- ▶ train
- ☰ sample_submission.csv
- ☰ train_image_level.csv
- ☰ train_study_level.csv

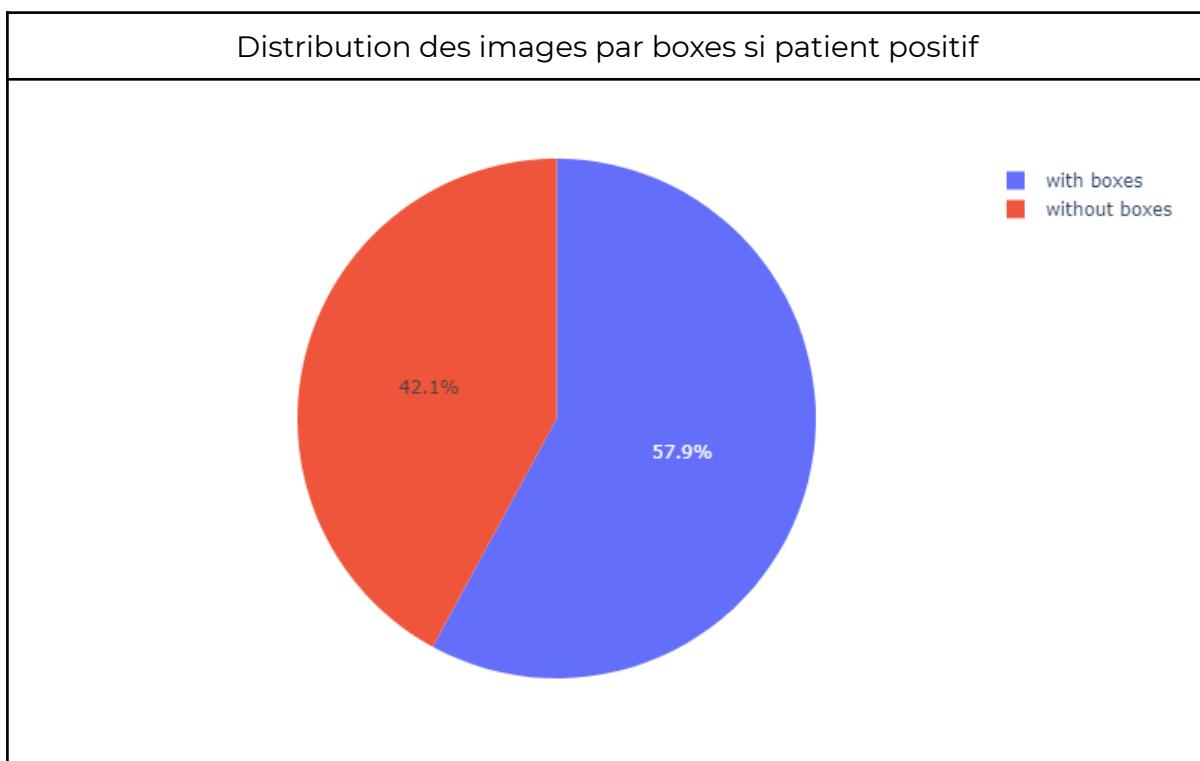
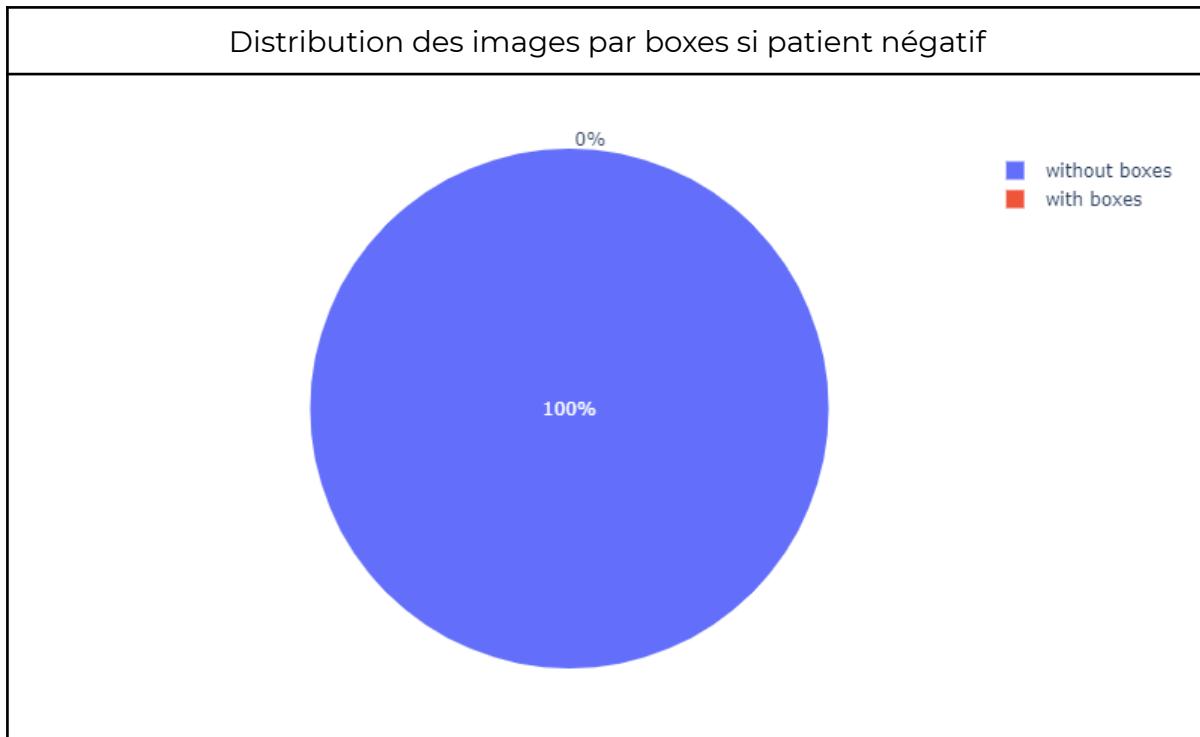


Analyser les données

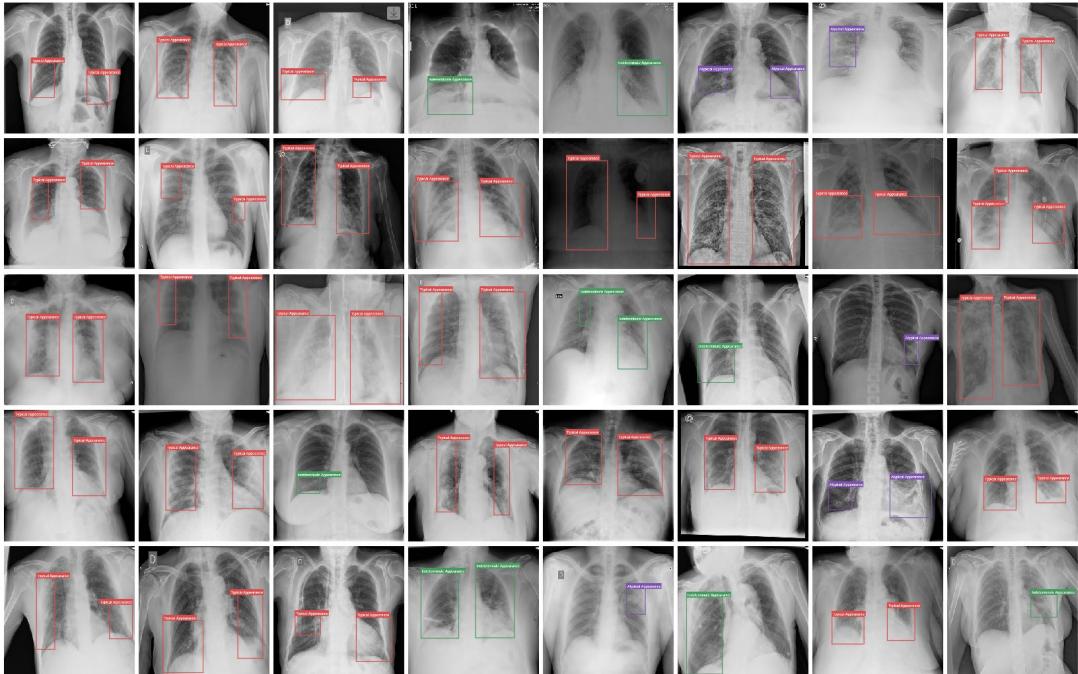
Il est nécessaire de réaliser une analyse des classes et des boxes.

Projet 8 - Participez à une compétition Kaggle !

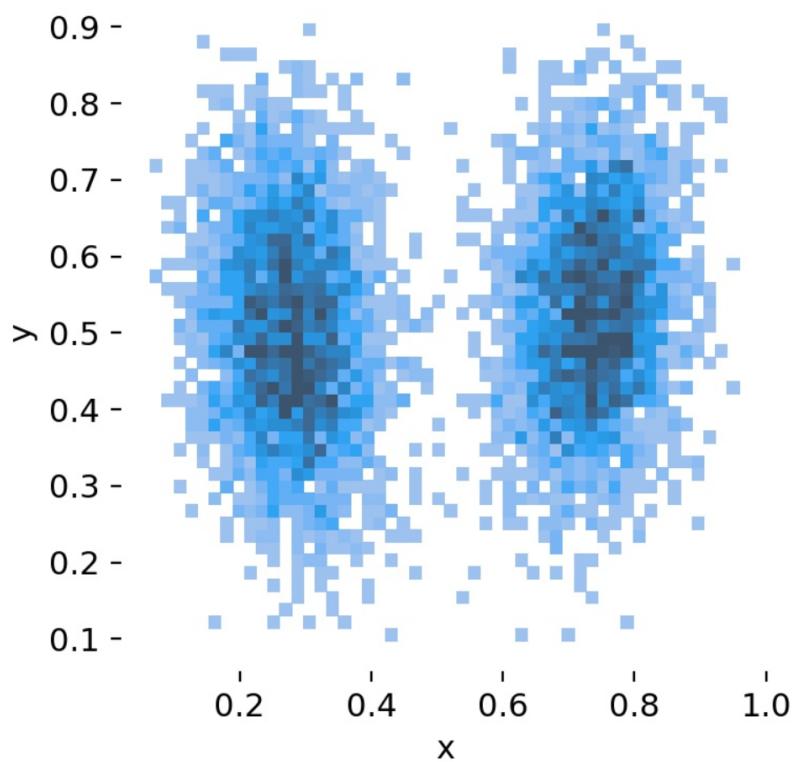




Visualisation des radiographies avec zones d'intérêts



Analyse de la position des boxes



Traiter les données

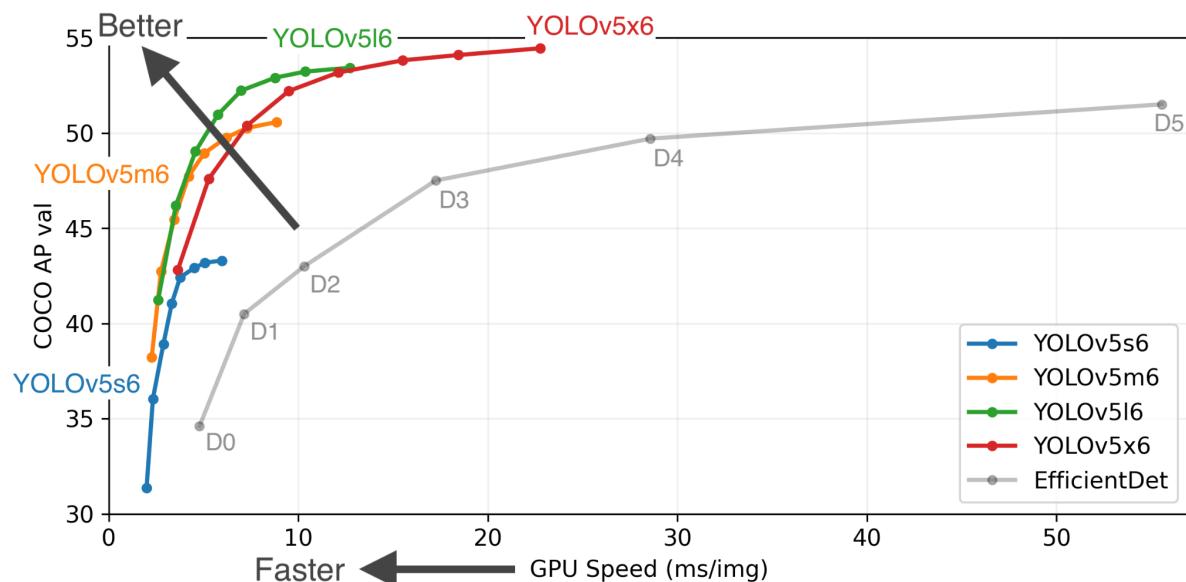
- extraction des images à l'aide de pydicom
- inversion de couleur suivant profil monochrome
- redimensionnement des images avec cv2 ~2500x2500 to 512x512px

Modéliser avec yolov5s

[@github.com/ultralytics/yolov5](https://github.com/ultralytics/yolov5)

YOLOv5 🚀 est une famille d'architectures et de modèles de détection d'objets pré-entraînés sur l'ensemble de données COCO, et représente la recherche open source d'Ultralytics sur les futures méthodes d'IA de vision, intégrant les leçons apprises et les meilleures pratiques développées au cours de milliers d'heures de recherche et développement

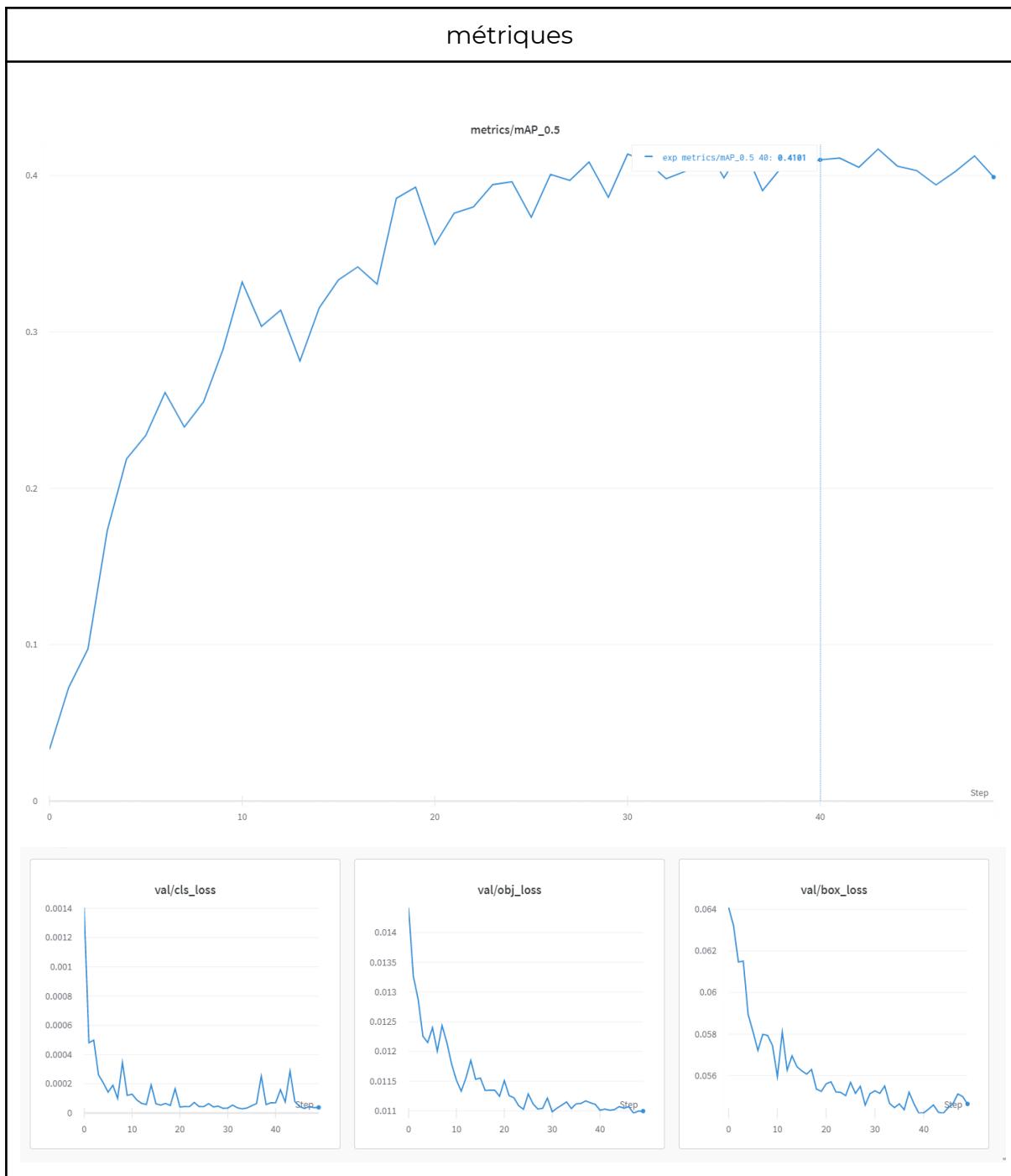
Nous allons utiliser "yolov5s", le plus petit modèle disponible pré-entraîné pour ce projet.



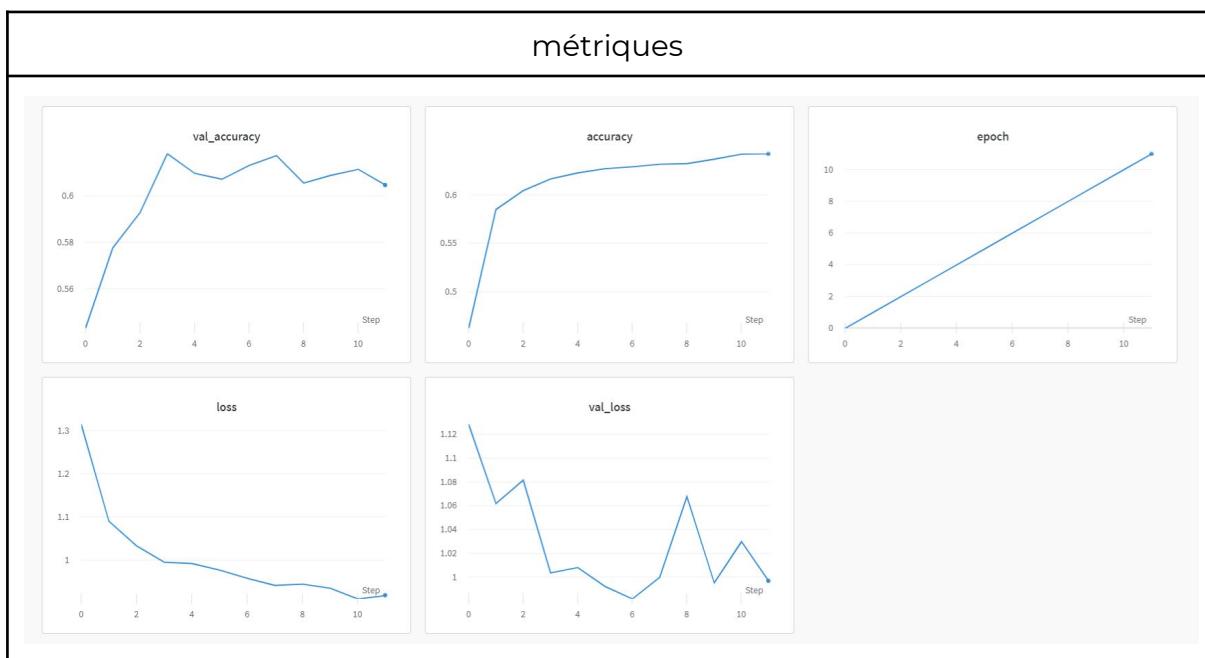
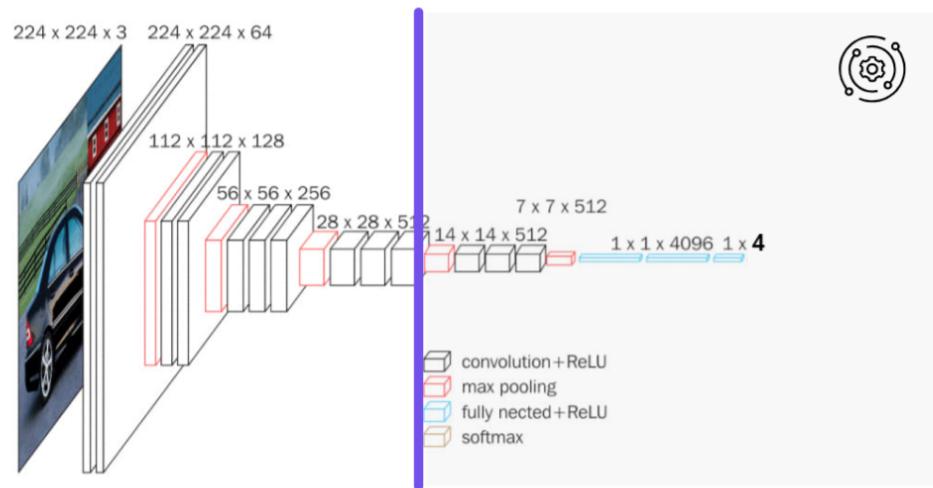
Nous pouvons observer et comparer les différents modèles sur l'interface de wandb après entraînement :

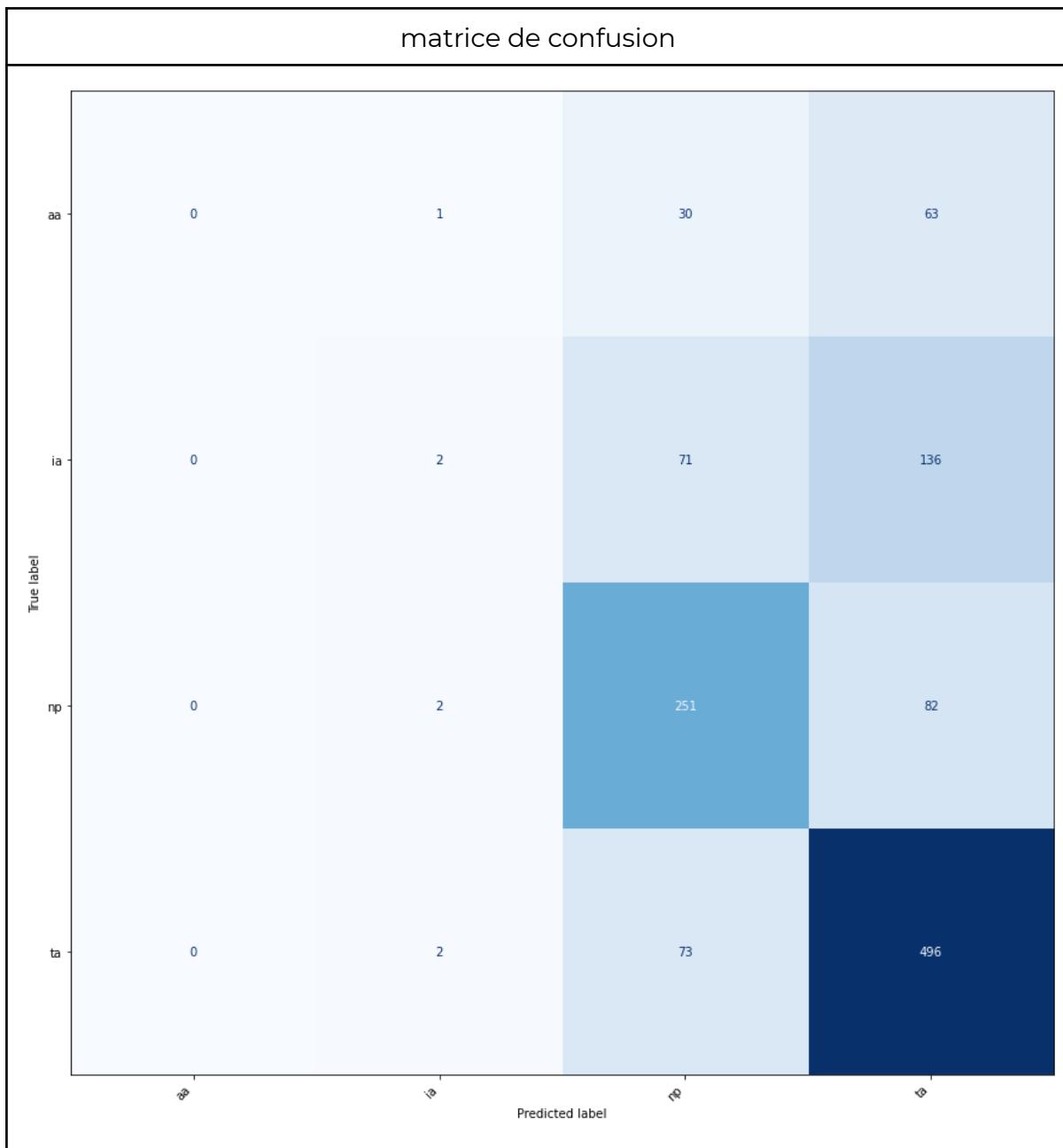
The screenshot shows a table titled "Runs" from the WandB interface. The table has columns for Name, Project, State, and Created. There are 11 rows of data, with rows 11 and 12 partially visible at the bottom. The data is as follows:

Name	Project	State	Created
exp	project8-kaggle-covid19	finished	10 hours ago
earthy-music-95	project8-kaggle-covid19	finished	10 hours ago
amber-sea-94	project8-kaggle-covid19	crashed	10 hours ago
rare-flower-93	project8-kaggle-covid19	finished	22 hours ago
dutiful-water-91	project8-kaggle-covid19	finished	1 day ago
skilled-cloud-90	project8-kaggle-covid19	finished	1 day ago
exp	project8-kaggle-covid19	crashed	2 days ago
exp	project8-kaggle-covid19	failed	2 days ago
exp	project8-kaggle-covid19	finished	2 days ago
model yolov5-512s	project8-kaggle-covid19	finished	1 week ago



Modéliser avec vgg16





On peut voir que la performance de prédictions pour nos 4 classes est faible avec de nombreuses erreurs sur ia et aa.

Submission

							Submit Predictions	...
520	Byungjun Yoon	</> happyhappy-covid1...		0.104	3	4d		
521	Vadim Titko			0.104	1	2d		
522	goofy.gf			0.104	1	1d		
523	Gopal K K 1	</> notebooke591cc8f4b		0.104	10	1d		
524	pablOberhauser			0.104	2	16h		
525	ivanwang			0.104	6	9h		
526	_Magic			0.104	2	8h		
527	the way			0.104	1	2h		
528	G@me Changer			0.104	1	2h		
529	Jessica Lee			0.089	7	2h		
530	omelette du fromage	</> [part3] siim-covid1...		0.063	6	2h		
Your Best Entry ↗								
Your submission scored 0.063, which is an improvement of your previous score of 0.058. Great job!							Tweet this!	
531	Naveen Rajan			0.053	2	1mo		
532	Riad			0.053	1	22d		
533	Carlos Picazo Montoya			0.053	1	21d		
534	Howie Wu			0.053	4	19d		
535	Meenakshi			0.053	2	16d		
536	Bùi Nhật Trường			0.053	1	17d		
537	Anton Chikin			0.050	2	12d		

Conclusion

Nous avons réussi à entraîner un modèle pour la prédiction de boxes sur les radiographies ainsi qu'un modèle de classification pour les études. La performance observée est faible et le score de la submission par conséquence également.

L'amélioration est possible par l'utilisation de méthodes sota plus avancées dans la classification d'images, par des méthodes ensemblistes et l'entraînement de modèles plus complexes sur des serveurs puissants les notebooks kaggle sont disponibles publiquement pour la communauté.