

DÉVELOPPEMENT D'UN LOGICIEL D'AIDE AU TRAITEMENT DE DONNÉES EXPÉRIMENTALES EN UTILISANT L'INTELLIGENCE ARTIFICIELLE

AU COMMISSARIAT À L'ÉNERGIE ATOMIQUE ET AUX ÉNERGIES ALTERNATIVES

PRÉSENTATION DU RAPPORT DU STAGE

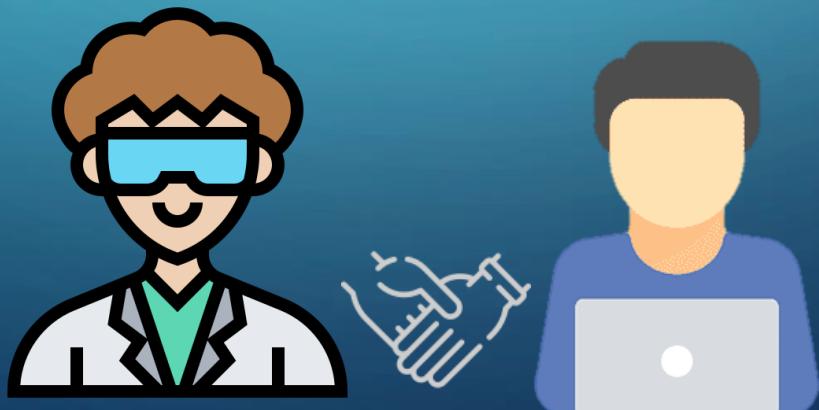
MATHIEU LACROIX

SUMMARY :

- Company presentation
- My mission
 - Part 1 : Application Web
 - Part 2 : Machine Learning
 - Conclusion
- Review

MY WORK IN COLLABORATION WITH RESEARCHERS

- Part 1 : Facilitate their analysis of the results using the Web
- Part 2 : Using machine learning image recognition to recognize the structure of matter



COMPANY PRESENTATION



Laboratoire Léon Brillouin

- Mixed Research Unit
- About 100 employees
- Work on the use of neutrons for the analysis of matter.



Commissariat à l'énergie atomique et aux énergies

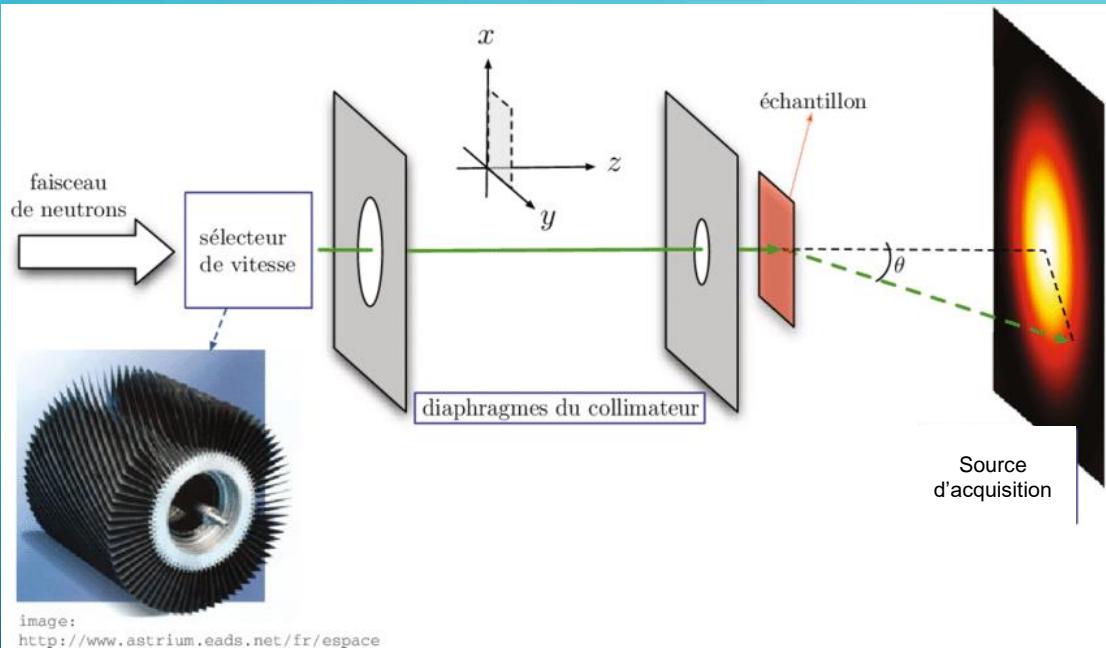
- Public research organization
- Over 21,000 employees
- 1st public research organization in the world ranking of patent filers



Centre National de la Recherche Scientifique

- Public research institution
- Advancing knowledge through research

INTRODUCTION PARTIE 1



Manipulation : bombardement neutronique d'un échantillon

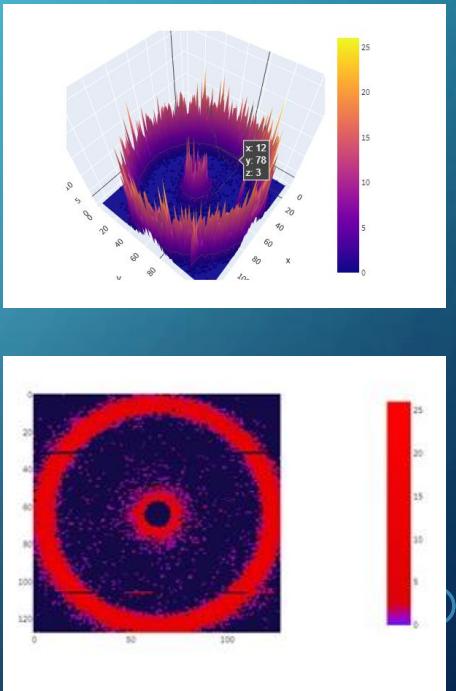


Objectif : développer une application Web pour aider les chercheurs à interpréter les résultats des manipulations neutroniques

Export des données d'acquisition



Création des fichiers .32

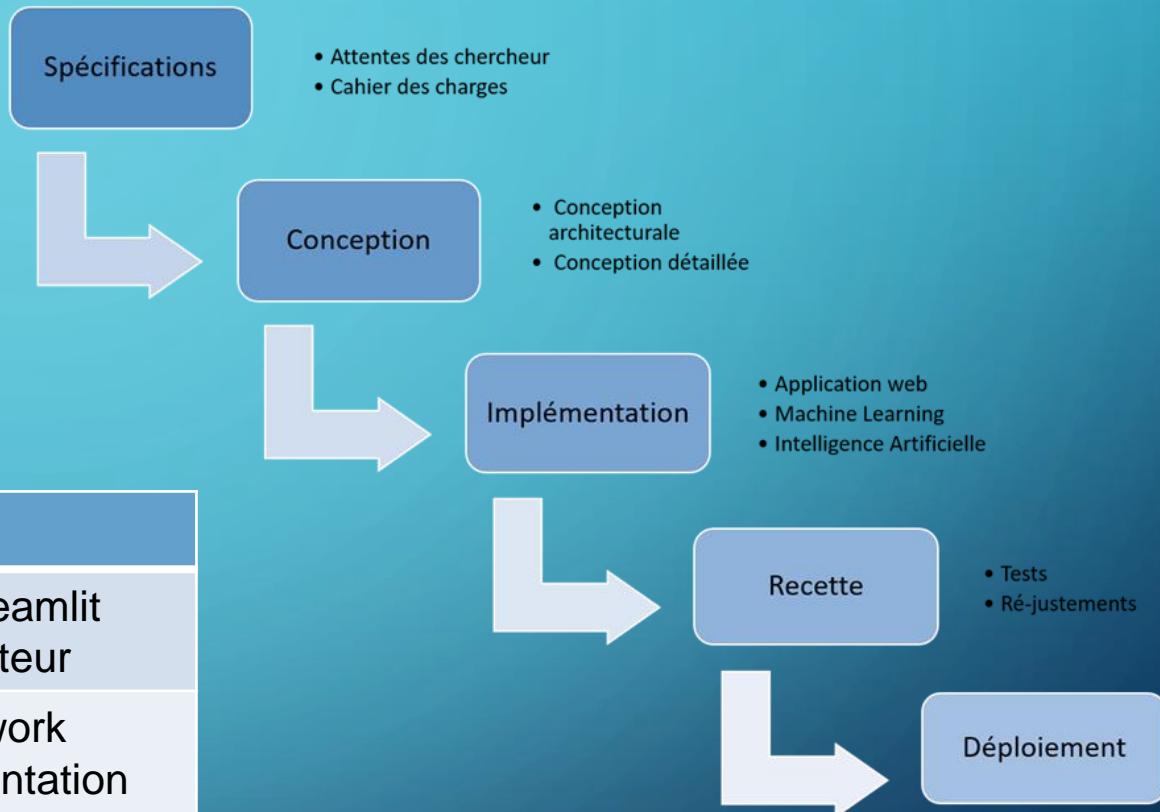


Lecture des résultats et interprétation

PHASE CONCEPTUELLE

- Définition du cahier des charges
- Planification du déroulement du projet en mode agile

Fonctionnalités attendues	Attentes
Création d'une application web interactive	Utilisation de API Streamlit pour l'interface utilisateur
Visualisation des données issues des acquisitions neutroniques	Utilisation de Framework Plotly pour la représentation graphique
Stockage sur une base de données	Développement d'une interface simple d'utilisation pour les chercheurs
Extraction des données depuis les fichiers .32	Utilisation du langage de programmation Python



CONCEPTION DE L'APPLICATION WEB



Un fichier héberge l'ensemble des données d'acquisition et de mesure.

Il est composé de 3 ensembles :



- bloc 1 : taille fixe de 256 octets, contient le détail des paramètres d'acquisition ,
- bloc 2 : les valeurs d'intensité acquises par un détecteur de résolution 128 x 128, en binaire sous le format d'entiers non signés codés sur 4 octets,
- bloc 3 : le détail de paramètres d'acquisition complémentaires, en texte, codés à l'aide de caractères ASCII.

CONCEPTION DE L'APPLICATION WEB



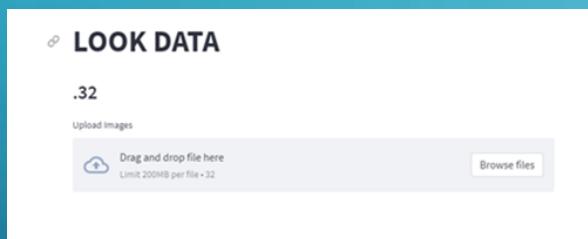
La base de données (cf. image ci-contre) est composée de deux tables:

- La 1^{ère} table « Fichier » stocke les caractéristiques des fichiers,
 - La 2^{ème} table « CheckSum » stocke le CheckSum de chaque fichier.

Pour le stockage des données d'acquisition, nous avons du convertir leur valeur en hexadécimal pour respecter la contrainte du nombre maximal de caractères.

CheckSum		Fichier	
<u>Id_CheckSum</u>	INTEGER	<u>Id_Fichier</u>	INTEGER
Hexa_CheckSum	TEXT	Nom_Fichier	TEXT
		Path_Fichier	TEXT
		<u>CheckSum_Fichier</u>	TEXT
		X0	INTEGER
		Y0	INTEGER
		Title	TEXT
		Sample	TEXT
		Counting	INTEGER
		Monitor	INTEGER
		Data	TEXT
		Time	TEXT
		Dist_sample_detector	INTEGER
		Distance_window_detector	INTEGER
		Lambda_read	INTEGER
		Lambda	INTEGER
		Temperature_param	REAL
		Temperature_enceinte	REAL
		Temperature_sample	REAL
		Magnetic_field_value	REAL
		Calibration_file	TEXT
		X_resolution	INTEGER
		Y_resolution	INTEGER

CONCEPTION DE L'APPLICATION WEB



Widget Streamlit

Exemple de requêtes :

- "INSERT INTO Fichier VALUES(?)", (item)
- "SELECT * From Fichier Where CheckSum_Fichier = '%s'%"id

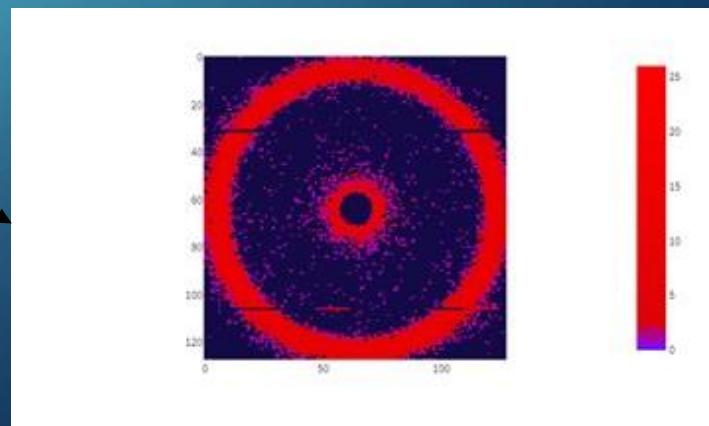


Requête SQL
d'insertion

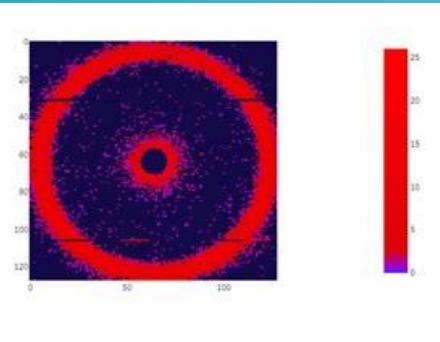
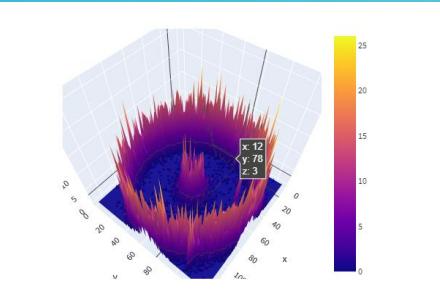


Requête SQL d'extraction

tableau des paramètres		tableau des intensités	
nom	valeur		intensité
0 basename	Xy056	2791	8
1 title	MnGe-	2792	8
2 sample	MnGe-D20	2793	9
3 counting	00030	2794	3
4 monitor	0050866	2795	8
5 date	22/03/201	2796	12
6 time	15:41:13	2797	5
7 max	0005	2798	3
8 selectorSpeed	2202	2799	5
9 lamhrls	0		



INTRODUCTION PARTIE 2



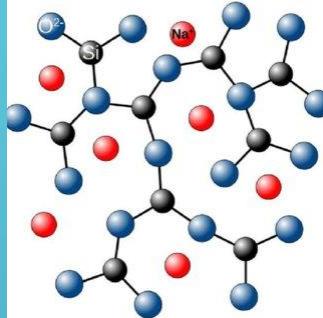
Représentation graphique des données d'acquisition



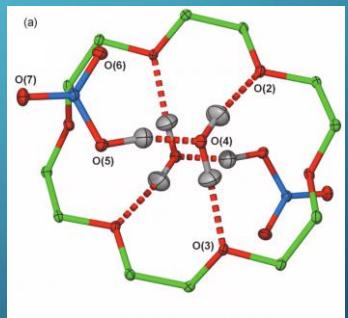
Objectif : Utiliser la reconnaissance d'image par apprentissage automatique à partir de représentation graphique des données d'acquisition pour identifier la structure d'un échantillon



Machine Learning



verre de glass



ellipsoïde

Exemples de structures neutroniques

CONCEPTION DU LOGICIEL DE TRAITEMENT DES DONNÉES

Récupération
des données



Mise en
forme des
images



Augmentation
du nombre
de données

Conception
du modèle

Entraînement
et validation
du modèle

Création d'un jeu de données à l'aide :

- de l'acquisition d'un jeu de données déjà confectionné,
- de l'acquisition d'image unitaire,
- de la prise de photo.

Chaque donnée est caractérisée par :

- son nom
- sa classe



CONCEPTION DU LOGICIEL DE TRAITEMENT DES DONNÉES

Récupération
des données

Mise en
forme des
images

Augmentation
du nombre
de données

Conception
du modèle

Entraînement
et validation
du modèle

128 x128 px



128 x128 px

- Redimensionnement harmonisé de l'ensemble des images pour faciliter le calcul du modèle

CONCEPTION DU LOGICIEL DE TRAITEMENT DES DONNÉES

Récupération
des données

Mise en
forme des
images

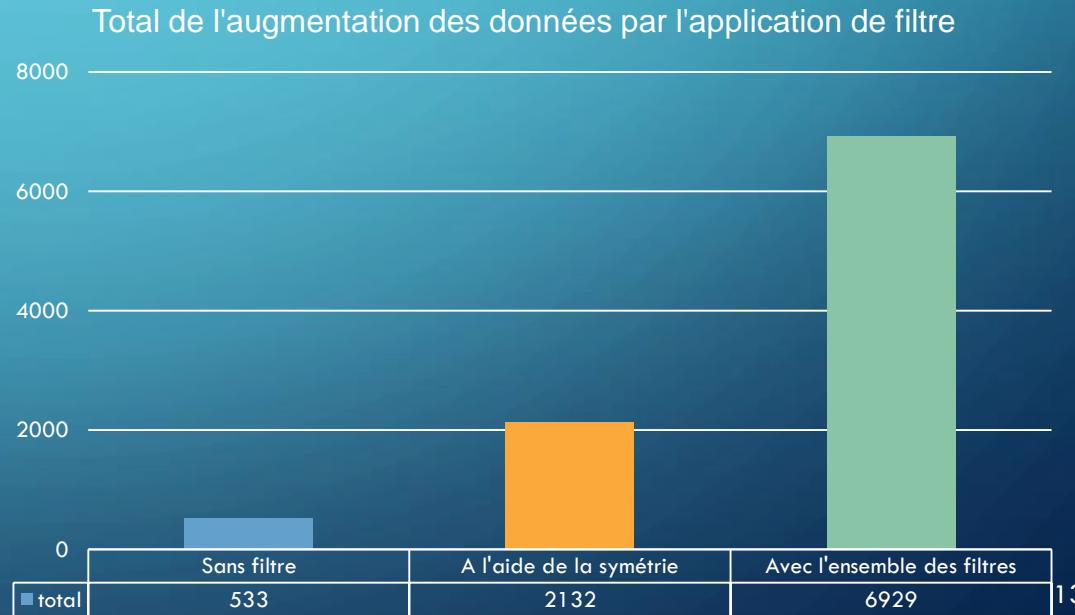
Augmentation
du nombre
de données

Conception
du modèle

Entraînement
et validation
du modèle

Exemples de filtre appliqué aux images :

- Symétrie verticale ou horizontale,
- Floutage,
- Rotation,
- Effet miroir combiné à une rotation,
- un décalage de 10 pixels vers la droite,
- un décalage de 10 pixels vers la gauche...



CONCEPTION DU LOGICIEL DE TRAITEMENT DES DONNÉES

Récupération
des données

Mise en
forme des
images

Augmentation
du nombre
de données

Conception
du modèle

Entraînement
et validation
du modèle

Les paramètres du jeu d'entraînement et de validation du modèle appliqués :

- le pourcentage d'images utilisées pour entraîner ou valider (ici, 80%),
- son type « *training* » ou « *validation* »,
- le format des images (ici, 128 x 128 px).

Les paramètres du modèle :

- le type du modèle,
- une méthode de normalisation « *rescaling* » pour d'aider la machine pour les calculs,
- une méthode de convolution qui va appliquer des filtres pour faire ressortir les caractéristiques fortes des images,
- une méthode « *maxpooling2d* » pour réduire la taille des images pour réduire leurs détails.¹⁴

CONCEPTION DU LOGICIEL DE TRAITEMENT DES DONNÉES

Récupération
des données

Mise en
forme des
images

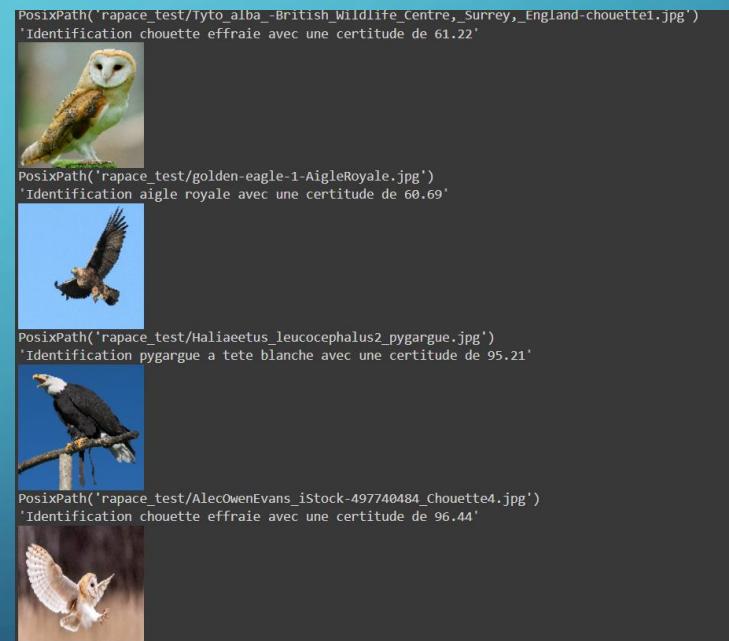
Augmentation
du nombre
de données

Conception
du modèle

Entraînement
et validation
du modèle



Évolution de la précision de l'entraînement
et de la validation du modèle



CONCLUSION

Au terme de mon stage, les objectifs définis ont été atteints :

- les chercheurs disposent maintenant d'une application Web simple, ergonomique et évolutive pour obtenir les représentations graphiques résultant de leurs manipulations ; et ainsi interpréter plus aisément les données acquisition,
- grâce au machine Learning et à la reconnaissance d'image par apprentissage automatique, les chercheurs ont à leur disposition un outil permettant de reconnaître automatiquement et avec une fiabilité élevée la structure d'un échantillon à partir de sa représentation graphique et de la base de données des graphiques.

BILAN ET PERSPECTIVES

Ce stage au sein du CEA à été une expérience enrichissante professionnellement et personnellement.

Les missions confiées m'ont permis de collaborer aux activités du LLB, et ainsi d'apprécier l'apport majeur de l'informatique au secteur de la recherche.

J'ai pu mettre en application les connaissances et compétences acquises à l'IUT, et compléter celles-ci grâce à la gestion de ces 2 projets.

Au final, ces 12 semaines passées au LLB sont venues confirmer mes choix d'orientation professionnelle vers le domaine du développement informatique par la diversité des sujets et des projets.