



Apprentissage de l'impact de la radiothérapie sur les cancers ORL grâce à un environnement HPC distribué.

Contexte

L'Institut Curie est né de la volonté d'une femme, Marie Skłodowska Curie, et de l'importance d'une cause, la lutte contre le cancer.

Afin d'accomplir ses missions de recherche et de soins, l'Institut Curie est constitué d'un ensemble hospitalier de pointe en cancérologie et d'un des plus grands centres européens de recherche en cancérologie. Pour le bon fonctionnement de cet ensemble localisé sur la montagne Sainte Geneviève à Paris, l'Institut Curie met aussi à disposition de ses chercheurs et de ses cliniciens des plateformes technologiques. Au sein d'une de ces plateformes, le Dr Didier Meseure, pathologiste et praticien spécialiste des Centres de Lutte Contre le Cancer anime l'unité de Pathologie Expérimentale PATHEX. C'est une plateforme d'expertise tissulaire ayant pour vocation la réalisation du volet histologique des projets de recherche à visée cognitive et pré-clinique des équipes institutionnelles appartenant au Centre de Recherche de l'Institut Curie et des partenaires industriels.

Afin de mieux prévoir l'évolution de certains cancers en réponse au traitement de radiothérapie, une base de données d'images de coupes histologiques de tumeurs est en cours de constitution.

Le projet proposé repose sur l'utilisation de l'intelligence artificielle afin d'élaborer un modèle d'apprentissage visant à prédire le pronostic de certains des cancers traités à l'Institut Curie : cancers de la sphère ORL (voies aérodigestives supérieures) et cancers de l'ovaire.

Objectifs du stage

L'originalité et la difficulté du projet sont liées à la taille des images. En effet, les images associées à chaque biopsie correspondent à des scans de lames obtenus par des microscopes automatisés produisant des images reconstituées dont la définition finale atteint facilement 50 000 x 80 000 pixels. Or les modèles d'apprentissage de réseaux de neurones convolutionnels sont compatibles avec des images de taille inférieure à 1000 x 1000. Il est donc nécessaire d'explorer des stratégies de distribution permettant d'exploiter le potentiel de ces modèles d'apprentissage sans trop perdre les informations globales réparties sur l'ensemble de la coupe histologique.

Après avoir sélectionné un modèle adapté au jeu de données, il s'agira dans un premier temps de définir et de mettre en place une méthodologie d'évaluation des performances (taille des données, taille des patches, empreinte mémoire) en utilisant un unique processeur graphique. Ensuite, le travail portera sur le traitement du même réseau de neurones sur ressources distribuées en suivant la méthodologie préalablement établie, de calibrer les ressources de calcul à mettre en œuvre en adéquation avec la taille des données à traiter, d'identifier

les goulets d'étranglement engendrées et enfin de proposer des solutions d'améliorations au niveau du support d'exécution.

Le stagiaire sera co-encadré par Elisabeth Brunet, enseignante chercheuse de TSP spécialisée dans le parallélisme et Daniel Stockholm, enseignant chercheur biologiste spécialisé dans l'analyse d'images de microscopie et collaborant avec le Dr Meseure sur les applications en imagerie de l'intelligence Artificielle.

Mots clés : Supercomputing, Deep learning, Optimisation de performances, Systèmes parallèles et distribués, GPU, Tensorflow, CNN

Informations pratiques

— Contact : Elisabeth Brunet - elisabeth.brunet@telecom-sudparis.eu