



# OPTIMISATION DE RÉSEAUX DE NEURONES POUR INTÉGRATION SUR CIBLE EMBARQUÉE

Référence STAGE : BT – Stage optimisation réseau neurones pour l'embarqué – 2024.

## Stage Ingénieur

Spécialités : intelligence artificielle, systèmes d'imagerie, informatique embarquée, imagerie infrarouge

Contexte : Optronique / Défense

Intitulé du stage : Optimisation de réseaux de neurones pour l'intégration « edge computing » sur cible embarquée

Langue de travail : Anglais et Français

## Niveau requis :

Compétences techniques recherchées :

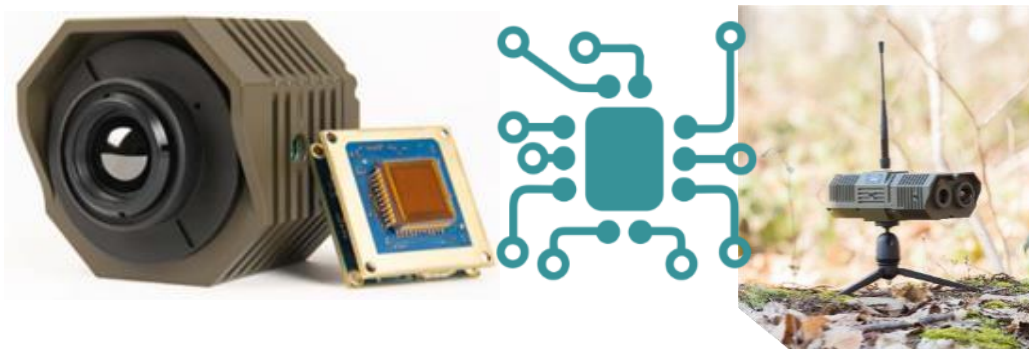
- Traitement Signal / Images / Vidéos
- Deep Learning
- Langage Python, bibliothèques Tensorflow+Keras / PyTorch
- Architectures matérielles d'apprentissage et inférence pour le Deep Learning
- Notions de FPGA / VHDL

Compétences interpersonnelles recherchées :

- Bonne capacité rédactionnelle
- Esprit critique
- Capacité à travailler en équipe et à s'inscrire dans une logique de développement à coût et délais contraints

## Contexte du stage :

Cette proposition de stage s'inscrit dans les processus d'amélioration continue des produits conçus et fabriqués par Bertin Technologies. Plus spécifiquement, il s'agit de poursuivre l'évolution des systèmes d'observation par imagerie avec l'intégration de traitements d'intelligence artificielle proche des capteurs.





Bertin Technologies développe et fabrique une gamme de caméras thermiques, appelée CamSight, ultra-compacte et présentant une très faible consommation électrique. Ces caméras peuvent être intégrées dans de nombreux systèmes d'observation, en particulier des réseaux de capteurs déposés dont le rôle est la surveillance de zones sensibles (figures ci-dessus).

La présente proposition de stage s'intéresse donc aux techniques de Deep Learning pour la classification automatique de menaces (objets, individus, véhicules) pour des applications militaires et de surveillance, reposant sur l'exploitation de données de capteurs d'imagerie thermique principalement (éventuellement visible).

Par l'intermédiaire de ses travaux internes, ou au travers d'échanges avec des organismes partenaires, Bertin Technologies dispose actuellement de plusieurs modèles de détection / classification d'objets basés sur le Deep Learning, fonctionnant dans le domaine infrarouge (IR) thermique.

Ces modèles ont jusqu'à présent été intégrés et utilisés sur des plateformes matérielles classiques et/ou dédiées au Deep Learning et aux réseaux de neurones (architectures basées sur des CPU ou des GPU).

Ces modèles sont à l'heure actuelle en cours de portage dans la plateforme matérielle cible, compatible des contraintes de compacité et de consommation électrique attendues dans le système final (architecture basée sur un FPGA).

Bertin souhaiterait :

- D'une part, finaliser l'intégration complète de ces modèles dans la plateforme matérielle cible, au plus près du capteur et de la caméra (« edge computing »), de manière à disposer d'une chaîne de traitement complète et fonctionnelle
- D'autre part, mettre en œuvre des techniques avancées d'élagage (*pruning*) et de quantification (*quantization*) de ces réseaux afin d'en réaliser une optimisation poussée, pour en limiter l'occupation mémoire et la consommation électrique. Cette optimisation est bien sûr réalisée en compromis avec d'autres paramètres comme la fréquence de traitement des images, les performances de détection du réseau, ...

### **Mission et Objectifs du stage :**

Le but du stage proposé est :

- Au sein des équipes en charge du développement des architectures embarquées pour les produits optroniques, de finaliser l'intégration d'au moins un des modèles existants dans la plateforme matérielle finale (de type FPGA), pour réaliser le portage temps-réel de ce modèle, et procéder à l'évaluation v1 de ses performances
- Au sein des équipes en charge du développement algorithmique, de reprendre la main sur les travaux déjà engagés chez Bertin Technologies, d'élagage et de quantification de réseaux de neurones
- De mettre en œuvre des techniques avancées d'élagage (*pruning*) et de quantification (*quantization*) de ces réseaux, et :
  - Identifier et se doter des outils adéquats, éventuellement manquant actuellement



- Réaliser le monitoring de « l'intérêt » de chacune des couches du réseau, de chacun des poids du réseau, de chacun des filtres du réseau, en vue de supprimer ceux jugés inutiles (recherche bibliographique des méthodes pertinentes, définition méthodologie d'inspection de l'ensemble du réseau, recherche de métriques d'intérêt, développement d'outils graphiques d'analyse...)
- Les travaux seront réalisés en langage Python sous Pytorch ou Tensorflow
- Réaliser les élagages / quantification de différents réseaux disponibles chez Bertin Technologies ou partenaires, avec les outils développés et selon les méthodes définies précédemment par le candidat
- Comparer les performances finales obtenues avec les différents modèles avant / après élagage et quantification avancés, et procéder à l'évaluation v2 des performances

#### **Résultats attendus :**

- Développement et documentation d'une bibliothèque de méthodes de portage hardware de réseaux de neurones, via des techniques d'élagage et/ou de quantification
- Adaptation d'un (ou plusieurs) modèle(s), par ces techniques d'élagage ou de quantification, afin de faciliter son portage sur une architecture temps-réel à base de FPGA
- Portage du (ou des) modèle(s) sur l'architecture temps-réel finale à base de FPGA (optimisations matérielles, conversion via SDK du FPGA sélectionné, intégration software et hardware du réseau élagué)
- Evaluation comparatives de performances de ce modèle avant / après portage, avec celles des modèles déjà existants chez Bertin (taux de reconnaissance, temps de calcul, etc.), non optimisés pour la cible hardware

**Durée :** 6 mois

**Lieu :** Aix en Provence

**Contact :** Aymeric ALAZARINE

Ingénieur Traitement des images / IA chez Bertin Technologies

04 42 60 45 79

[aymeric.alazarine@bertin.group](mailto:aymeric.alazarine@bertin.group)