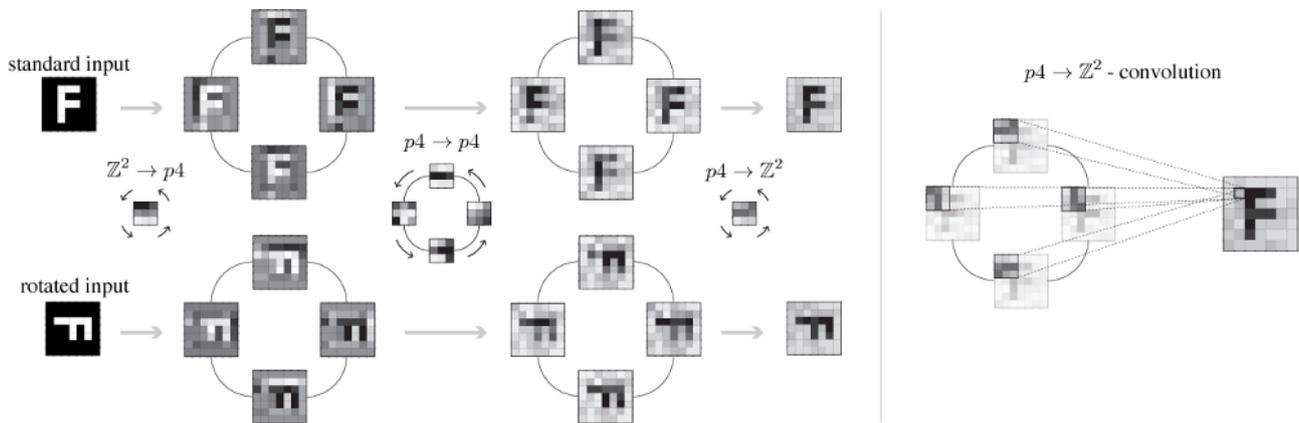


Réseaux convolutifs équivariants pour la segmentation sémantique



Représentation schématique d'un G-CNN deux couches d'après [3]. Le modèle est par construction équivariant à des groupes de transformations, ici les rotations de $\frac{\pi}{2}$. Les activations d'une image tournée sont identiques aux activations tournées de l'image originale.

1 Contexte

Les réseaux de neurones convolutifs (CNN) constituent depuis quelques années le nouvel état de l'art en vision par ordinateur et traitement d'images. Ces réseaux sont construits autour de l'opérateur de convolution qui est par construction équivariant par translation : si une image d'entrée subit une translation, son image convoluée sera translatée de la même manière. Cette propriété est particulièrement intéressante en reconnaissance d'objets dans la mesure où la position d'un objet dans l'image n'influe en général pas sur sa nature sémantique.

Cependant, la convolution ne présente pas d'équivariance à des transformations plus complexes. Ainsi, les CNNs souffrent de performances dégradées lorsque l'on les applique sur des images ayant subi une rotation ou une réflexion, dès lors que cette transformation n'est pas observée dans le jeu d'entraînement soit naturellement, soit par l'utilisation de technique de *data augmentation*.

Plusieurs approches ont été introduites dans la littérature pour tenter de rendre les CNN équivariants à des transformations géométriques complexes, soit en explicitant les contraintes d'équivariance au sein de la fonction de coût [4], soit par construction en travaillant sur des notions de groupes de symétrie [1, 7, 3, 5, 6]. Les modèles basés sur cette dernière approche sont dits *Group Equivariant Convolutional Neural Networks* (G-CNN). L'objectif de ce stage est d'investiguer la mise en œuvre de ces techniques dans des applications concrètes pour évaluer leur capacité à contraindre l'équivariante des réseaux convolutifs à différents types de transformations.

2 Objectifs

Le premier objectif de ce stage est de se familiariser avec la théorie des G-CNNs ainsi que leur implémentation en se basant sur [1, 2]. Les performances de ces réseaux seront d'abord évaluées en classification sur des bases classiques telles que MNIST et CIFAR. Nous nous intéresserons par la suite à des applications en segmentation sémantique d'images, par exemple des images médicales [3], aériennes et satellitaires. En effet, ces images présentent des équivariante naturelles aux rotations et aux réflexions.

Dans un deuxième temps, nous comparerons les représentations apprises par ces modèles équivariants par construction à ceux des CNN classiques ainsi qu'à d'autres modèles qui imposent l'équivariance par contrainte [4].

3 Profil

Nous recherchons pour ce stage un-e candidat-e de niveau M2 ou dernière année d'école d'ingénieur avec une formation en mathématiques appliquées, en informatique ou en traitement du signal. Le ou la candidat-e idéal-e a une appétence pour la recherche scientifique et des bases théoriques en apprentissage automatique. Des notions de traitement d'image ou de mathématiques algébriques (théorie des groupes, algèbre générale) sont un plus pour ce sujet.

Une connaissance de la programmation avec Python est préférable, il est toutefois envisageable pour un-e candidat-e connaissant un autre langage de programmation de se former à Python au cours du stage. Une première expérience avec une bibliothèque d'apprentissage profond telle que TensorFlow ou PyTorch est la bienvenue.

4 Organisation

Cette offre de stage porte sur un stage d'une durée de 5 à 6 mois avec une date de début flexible au printemps 2021. Le stage se déroulera au centre de recherche et d'études en informatique et en communications (CEDRIC) du Conservatoire national des arts et métiers (Cnam) à Paris, 3^e arrondissement.

Le **CEDRIC** est un laboratoire fondé en 1988 rassemblant plus de 80 enseignants-chercheurs regroupés dans 7 équipes thématiques. Ses activités couvrent divers champs de recherche allant de la fouille de données multimédia aux radiocommunications en passant par l'apprentissage statistique, les médias interactifs et l'optimisation combinatoire.

Le stage sera co-encadré par Clément Rambour, Nicolas Audebert et Nicolas Thome de l'équipe **Données complexes, apprentissage et représentations**.

5 Candidater

Envoyer une candidature (CV + brève explication de votre motivation) par email à :

clement.rambour@cnam.fr, nicolas.audebert@cnam.fr et nicolas.thome@cnam.fr.

Références

- [1] T. Cohen and M. Welling. Group equivariant convolutional networks. In *International Conference on Machine Learning*, pages 2990–2999, 2016.
- [2] Jasper Linmans and others. Sample Efficient Semantic Segmentation using Rotation Equivariant Convolutional Networks, 2018.
- [3] B. S. Veeling et al. Rotation Equivariant CNNs for Digital Pathology. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, pages 210–218, 2018.
- [4] Y. Wang et al. Self-supervised scale equivariant network for weakly supervised semantic segmentation. In *Computer Vision and Pattern Recognition*, 2020.
- [5] M. Weiler, F. A. Hamprecht, and M. Storath. Learning Steerable Filters for Rotation Equivariant CNNs, 2018.
- [6] M. Winkels and T. S. Cohen. 3D G-CNNs for Pulmonary Nodule Detection, 2018.
- [7] D. E. Worrall et al. Harmonic Networks : Deep Translation and Rotation Equivariance, 2017.