A propos du centre ou de la direction fonctionnelle

Grenoble Rhône-Alpes Research Center groups together a few less than 800 people in 35 research teams and 9 support research departments.

Staff is localized on 5 campuses in Grenoble and Lyon, in close collaboration with labs, research and higher education institutions in Grenoble and Lyon, but also with the economic players in these areas.

Present in the fields of software, high-performance computing, Internet of things, image and data, but also simulation in oceanography and biology, it participates at the best level of international scientific achievements and collaborations in both Europe and the rest of the world.

Contexte et atouts du poste

General Information: The PhD student will be part of the Inria Thoth team and Jean Kurtzmann Laboratory. The team is specialized in computer vision and machine learning, and the student will be jointly supervised by Kartheek Alahari and Valerie Perrier (Grenoble INP / LJK). Thoth is one of the most recognized European research teams in computer vision, participates in several European projects, and has many international academic (CMU, Berkeley, MPI), and industrial (Facebook, Google, Microsoft, Naver Lab) collaborations.

Location: The student will work at the Inria research center near Grenoble. The unit includes more than 600 people, within 26 research teams and 10 support services. Grenoble is a lively city which hosts many foreign students and researchers. Located in the heart of the French Alps, its direct surroundings offer great outdoor recreation including skiing, cycling, and hiking. Grenoble is well-connected to Lyon and Geneva airports, and Paris can be reached from Grenoble in 3h by train.

Objectives: The goal of this PhD is to develop deep learning algorithms, which lie at the intersection of neural networks and wavelet transforms. These algorithms will be evaluated on computer vision datasets for problems such as image classification, object detection and segmentation, thus making contributions on the experimental front, in addition to developing novel algorithms and analyzing their theoretical properties. Further details are provided below.

Mission confiée

Automatic analysis of digital data, and in particular visual data, which is at the forefront of major industrial, societal and scientific applications, requires dealing with the exponential growth of diverse data. In 2022, it is expected that nearly 80% of Internet traffic will be due to videos, and that it would take an individual over 5 million years to watch the amount of video that will cross global IP networks each day. Thus, there is a pressing and in fact increasing demand to analyze, annotate and index, this visual content automatically for home and professional users alike. In light of this, the main challenge is not only to develop efficient algorithms but also to understand their limitations through theoretical analysis.

Machine learning approaches, and in particular deep learning frameworks, have been popular in our quest for building data analysis algorithms, resulting in significant improvements for speech, text and object recognition, as well as medical image analysis. Despite these impressive successes, several important questions remain unaddressed in this context. In particular, the computational cost required to learn the billions of parameters that these models are composed of, the large amounts of annotated data required during the training phase, and the lack of strong theoretical guarantees for the reliability of the results obtained, even in cases where the training set is sufficiently representative of unseen test data distributions.

A majority of the state-of-the-art methods focus on improving the precision of results by designing new network architectures, exploiting multi-modal data. Some methods have also attempted to reduce the need for fully-annotated training data, with self-supervision [2,3], and to provide an understanding of deep learning models [4]. While these attempts constitute a good starting point to address the challenges in deep learning, the overall analysis continues to remain largely experimental.

In contrast to this largely computational or experimental view, a mathematical approach for data analysis consists of models (e.g., statistical, physical) that exploit the inherent structure of data, and also provide the conditions under which it can be reproduced or generalized with certain error guarantees. Examples of such approaches include compressed sensing, wavelet transforms.

This thesis is at the interface of these two philosophies—it aims to develop novel neural network models, whose parameters control the sparsity of the represented data, thereby exploiting their inherent structure. Specifically, it lies at the intersection of neural networks, where all the convolutional filters are learned [5], and wavelet scattering transform, where information is transferred from one layer to another through fixed wavelet transformations [6]. This goes beyond methods that are appearing recently in this context [7, 8, 9], but are very limited in the number of layers, or do not exploit the properties of compression of wavelets, or do not learn the wavelet transforms.

References

The methods explored during the PhD will include theoretically-grounded initialization strategies for neural networks, the use of wavelet filters in combination with neural networks, invariance in neural networks. They will be evaluated on computer vision datasets.

The results obtained as part of this work will be submitted to top-tier conferences and journals in computer vision (CVPR, ECCV, ICCV, IJCV, PAMI) and machine learning (ICML, JMLR, NeurIPS). The student is expected to attend the conferences and present the accepted paper(s). All the publications will be made available on the Hal repository. Software packages associated with the methods developed will be another research output from this work.

**Compétences**

We are looking for a creative and highly-motivated student (preferably with a Masters degree in Computer Science or Applied Mathematics) with an interest in computer vision and deep learning. We strongly encourage applicants with diverse backgrounds.

Fluent English (written and spoken) is mandatory for effective scientific communication.

This project requires strong mathematics knowledge in linear algebra and statistics, and excellent programming skills: Python or C++. Prior courses or knowledge in the areas of computer vision, signal processing, machine learning is a plus.

**Avantages**

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Possibility of teleworking (after 6 months of employment) and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage

**Rémunération**

Salary (before taxes) : 1982€ gross/month for 1st and 2nd year. 2085€ gross/month for 3rd year.