Axe 1 : Journée scientifique de la FR Math-STIC

Apprentissage appliqué aux données non-structurées

Jeudi 21 novembre 2019
Amphithéâtre Darwin

9h00
Accueil et présentation de la journée
Par Anissa MOKRAOUI et Roberto WOLFLER

9h30
Evaluation de la fiabilité des algorithmes de "Machine Learning"
Par Pierre DUHAMEL
Laboratoire des signaux et Signaux (L2S), Supelec

10h15
Deep Convolutional Auto-Encoders as Multiscale Inverse Problems
Tomás ANGLES
ENS de l'université PSL (ENS Ulm)

11h-11h15 : Pause café

11h15
Machine Learning driven Variable Frame-Rate in Video Broadcast Applications
Wassim HAMIDOUCHE
Institut d'Électronique et de Télécommunications de Rennes (IETR, UMR CNRS 6164)

12h00
Learning transforms for image/video compression
Aline ROUMY
INRIA Rennes

12h45-14h15 : Pause déjeuner

14h15
Enhancing HEVC spatial prediction by context-based learning
Li WANG
Telecom Paris Tech (Groupe Multimédia)

15h00
Convex Functions and Neural Networks: a Case Study in Natural Language Processing
Joseph LE ROUX
LIPN, Université Paris 13

15h45
(To be announced)
Par Nistor GRAZAVU
LIPN, Université Paris 13

16h30
Clôture de la journée

Résumés des exposés

Pierre DUHAMEL

Evaluation de la fiabilité des algorithmes de "Machine Learning"

Les algorithmes d'apprentissage (et spécialement les algorithmes d'apprentissage profond (Deep Learning) nécessitent la disponibilité de grandes bases de données qui soient représentatives des situations dans lesquelles ils seront utilisés. Cette représentativité peut être mise en cause pas divers phénomènes, tels que la suradaptation (overfitting), l'évolution des statistiques des signaux d'intérêt, etc… il est donc très important d'évaluer si les sorties de tels algorithmes sont fiables.

Un tel objectif pourrait avoir de nombreuses applications, et permettrait de répondre à des questions comme
1- le réseau a été entrainé et est utilisé sur de nouvelles entrées. Si les statistiques de ces entrées évoluent, quand est-il nécessaire de recommencer l'entrainement ?
2- peut-on obtenir une mesure de précision des sorties de l'algorithme, de telle sorte à "pondérer" les décisions correspondantes ?
3- peut-on détecter des "outliers" ?

Même si nous en pouvons répondre à toutes ces questions, l'exposé expliquera tout d'abord comment poser le problème, en faisant un rapide état de l'art. Puis l'exposé se concentrera sur un approche "boite noire", dont nous expliquerons la large utilité, et pour laquelle nous proposerons une première série de méthodes avec les simulations correspondantes.

Tomás ANGLES

Deep Convolutional Auto-Encoders as Multiscale Inverse Problems

Constructing generative models of high-dimensional random processes is a central problem in Machine Learning, Image Processing, and Mathematics. Auto-Encoders based on Deep Convolutional Networks seem to solve this problem by taking advantage of strong regularities in natural signals, but the specific mathematical mechanisms to exploit these regularities are not well-understood. We present a model of Deep Convolutional Auto-Encoders for which we can explicitly state the mechanisms used to obtain a generative model. We propose a multiscale encoder based on a scattering transform that Gaussianizes the random process using a central limit theorem and exploits spatial regularities at different scales as well as the dependencies across scales found in natural signals. Furthermore, the encoder is structured such that the decoder inverts a code by solving a sequence of linear inverse problems at different scales. These inverse problems are stabilized using sparsity by learning a sequence of convolutional dictionaries, one per scale. Our model provides insights and potential explanations for the properties of generated signals obtained with Deep Convolutional Auto-Encoders.

Wassim HAMIDOUCHE

Machine Learning driven Variable Frame-Rate in Video Broadcast Applications

The Digital Video Broadcasting (DVB) has proposed to introduce the Ultra-High Definition services in three phases: UHD-1 phase 1, UHD-1 phase 2 and UHD-2. The UHD-1 phase 2 specification includes several new features such as High Dynamic Range (HDR) and High Frame-Rate (HFR). It has been shown in several studies that HFR (+100 fps) enhances the perceptual quality and that this quality enhancement is content-dependent. On the other hand, HFR brings several challenges to the transmission chain including codec complexity increase and bit-rate overhead, which may delay or even prevent its deployment in the broadcast echo-system.

In this talk, I will present our proposed Variable Frame Rate (VFR) solution to determine the minimum (critical) frame-rate that preserves the perceived video quality of the HFR video. The frame-rate determination is modeled as a 3-class classification problem which consists in dynamically and locally selecting one frame-rate among three: 30, 60 and 120 frames per second. Two random forests classifiers are trained with a ground truth carefully built by experts for this purpose. The subjective results conducted on ten HFR video contents (not included in the training set) clearly show the efficiency of the proposed solution enabling to locally determine the lowest

possible frame-rate while preserving the quality of the HFR content. Moreover, our VFR solution enables significant bit-rate savings and complexity reductions at both the encoder and decoder sides.

Aline ROUMY

Learning transforms for image/video compression

Learning transforms is a key ingredient in image/video compression, which allows to adapt to the statistics of the image and more precisely to its local dependencies. In this talk, two types of adaptation will be presented.
First, deep learning techniques for compression of 2D images is considered. We have in particular addressed the problem of learning transforms that would be optimal in terms of energy compaction. This unsupervised learning problem has been addressed by proposing auto-encoders with a rate-distortion cost function, and not only a distortion criterion, as is classically done in machine learning. The proposed neural network can work efficiently at any coding rate using an adaptation to the quantization noise during training.
In the second part of the talk, a transform is learned on the sphere to compress 360-degree images. Omni-directional images are characterized by their high resolution (usually 8K) and therefore require high compression efficiency. Existing methods project the spherical content onto one or multiple planes and process the mapped content with classical 2D video coding algorithms. However, this projection induces sub-optimality. Indeed, after projection, the statistical properties of the pixels are modified, the connectivity between neighboring pixels on the sphere might be lost, and finally, the sampling is not uniform. Therefore, we propose to process uniformly distributed pixels directly on the sphere to achieve high compression efficiency. In particular, a scanning order and a prediction scheme are proposed to exploit, directly on the sphere, the statistical dependencies between the pixels. A transform is also learned to exploit local dependencies while taking into account the 3D geometry.

Li WANG

Enhancing HEVC spatial prediction by context-based learning

Deep generative models have been recently employed to compress images, image residuals or to predict image regions. Based on the observation that state-of-the-art spatial prediction is highly optimized from a rate-distortion point of view, we study how learning-based approaches might be used to further enhance this prediction. To this end, we propose an encoder-decoder convolutional network able to reduce the energy of the residuals of HEVC intra prediction, by leveraging the available context of previously decoded neighboring blocks. The proposed context-based prediction enhancement (CBPE) scheme enables to reduce the mean square error of HEVC prediction by 25% on average, without any additional signaling cost in the bitstream.

Joseph LE ROUX

Convex Functions and Neural Networks: a Case Study in Natural Language Processing

In recent years, neural networks (NN) have become ubiquitous in Natural Language Processing (NLP), with tremendous progress made in Machine Translation, Summarization etc… They offer a rich parameterization of score functions and almost unbounded context.
On the other hand, for tasks enforcing a strong a priori on the output structure, such as parsing which requires that output are arborescences, NNs are mainly used as better replacements for old technology such as feature extractors (recurrent networks) and classifiers (multi-layered perceptrons), whose scores are then passed to combinatorial algorithms that generate discrete structures. One of the problems with this approach is that the combination of the parts of the structures must be fixed in advance by hand (think for instance of a maximum spanning arborescence problem where the objective function is fixed as the sum of the arcs the arborescence).
In this talk I will present some preliminary work which try to depart from this approach and learn the combination function itself, while remaining efficient. We use Structured Prediction Energy Networks (SPENs) and enforce convexity, in some of the input, of the function computed by the neural networks so we can rely on efficient optimization methods to perform learning and decoding.

Nistor GROZAVU

(To be announced)