2019-02043 - Post-Doctoral Research Visit F/M Data Repurposing: compression of large scale image and video databases

Contract type : Public service fixed-term contract
Level of qualifications required : PhD or equivalent
Fonction : Post-Doctoral Research Visit

About the research centre or Inria department

Inria, the French national research institute for the digital sciences, promotes scientific excellence and technology transfer to maximise its impact.
It employs 2,400 people. Its 200 agile project teams, generally with academic partners, involve more than 3,000 scientists in meeting the challenges of computer science and mathematics, often at the interface of other disciplines.
Inria works with many companies and has assisted in the creation of over 160 startups.
It strives to meet the challenges of the digital transformation of science, society and the economy.

Assignment

Context: Free upload of multimedia data on the social networks and cloud storage services makes consumers think that the storage cost is zero. This assertion is wrong twice. The cost is not only pecuniary (funded thanks to the advertisement or data selling) but more importantly also ecological. In 2012, the number of data centers in the world was estimated at about 500.000, consuming roughly the equivalent of 30 nuclear power stations. During the past years, the number of data centers has literally exploded, reaching more than 8 millions nowadays.
All the compression algorithms, including the most powerful ones, are overwhelmed by such an explosion. Their limitation resides in the *rate-fidelity criterion* they are optimizing. Indeed, the goal of a traditional compression algorithm is to describe the data with the minimum bits and with a minimum error compared to the original data. When the data itself is expanding exponentially, even reaching a good compression ratio will only slow down the volume growing instead of stopping it. It now becomes critical to change paradigm and put compression problem in another perspective.

Motivations: This is the promise of Data Repurposing tackled during the postdoctoral project. The core principle is to replace a large input data collection by a data of another format (more compact and ergonomic), giving the opportunity to reach drastic compression ratio. *Fidelity is replaced by subjective exhaustiveness*. Basically, a repurposed data is supposed to replace the input data collection. The success of such an approach depends on its ability to restitute to the user the information contained in the initial data. The information here, is no longer the information defined in Shanon's theory, but corresponds to what a user retains/learns/feels from a data collection. We call it the perceived information. The *subjective exhaustiveness* refers to the fact no perceived information of the input collection is lost in the repurposed data.

Objectives: i) to define this new compression scenario called *Data Repurposing*, suited for massive collections of image and video data ii) to evaluate the theoretical performance of such a paradigm and iii) to propose innovative coding approaches targeting unprecedented compression ratios for both image and video collections.

Main activities

The post-doctoral researcher will focus on the following activities:

- Learn features from the images and videos: this will be based on dictionaries learned from data or auto-encoder derived from CNN [1].
- Define new metrics to evaluate the amount of information spanned by a image/video data collection [2]
- Develop compression algorithms (e.g., sampling) for these metrics [3,4]

References:

[1] Charte, D., Charte, F., García, S., del Jesus, M. J., & Herrera, F. (2018). A practical tutorial on autoencoders for nonlinear feature fusion: Taxonomy, models, software and guidelines. *Information Fusion*, 44, 78-96.

[2] Kuhnel, L., Fletcher, T., Joshi, S., & Sommer, S. (2018). Latent Space Non-Linear Statistics. *arXiv preprint arXiv:1805.07632*.

[3] Belhadji, A., Bardenet, R., & Chainais, P. (2018). A determinantal point process for column subset selection. *arXiv preprint arXiv:1812.09771*.

[4] Puy, G., Tremblay, N., Gribonval, R., & Vandergheynst, P. (2018). Random sampling of bandlimited signals on graphs. *Applied and Computational Harmonic Analysis*, 44(2), 446-475.

Skills

The candidate should have
· strong background in signal/image processing, machine learning
· notions of deep learning and information theory would be appreciated,

Benefits package

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT
- Social, cultural and sports events and activities
- Access to vocational training

Remuneration

About Inria

Inria, the French national research institute for the digital sciences, promotes scientific excellence and technology transfer to maximise its impact. It employs 2,400 people. Its 200 agile project teams, generally with academic partners, involve more than 3,000 scientists in meeting the challenges of computer science and mathematics, often at the interface of other disciplines. Inria works with many companies and has assisted in the creation of over 160 startups. It strives to meet the challenges of the digital transformation of science, society and the economy.

Instruction to apply

Please submit online : your resume, cover letter and letters of recommendation eventually

For more information, please contact thomas.maugey@inria.fr

Defence Security :
This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST).Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

Recruitment Policy :
As part of its diversity policy, all Inria positions are accessible to people with disabilities.

Monthly gross salary amounting to 2653 euros.