

PhD thesis announcement

Title: Deep learning based image compression

I. Thesis context

Recent developments in acquisition and display technologies tend to add depth (3D) information to allow better immersion in virtual worlds and augmented reality. For instance, one of the potential 3D imaging techniques relies on the use of stereoscopic (or multiview) systems that generate two (or more) images of the same perceived scene. Such data have been widely used in various application fields like 3DTV, computer vision and medicine. As a result, the involved data amounts are prohibitive and constitute a major problem for its practical use. Therefore, it becomes mandatory to design efficient single and stereo images coding schemes in order to improve their reconstruction quality while reducing their storage capacity.

For this purpose, transform coding schemes have been widely used in the literature. However, traditional fixed transforms, which work well for natural images, may not be as well-suited for images presenting specific structures and should be adapted to the inherent characteristics of the input images. Thus, in order to improve the image coding performance, it would be interesting to build adaptive transforms well adapted to the image contents. In this context, and more specifically with wavelet-based compression techniques, many research works have been devoted to lifting schemes (LS) which are composed of prediction filters and update filter to generate the wavelet coefficients. Therefore, in order to build an image content-adaptive transform, a particular attention should be paid to the design of the prediction and update filters by optimizing a rate-distortion criterion [1].

II. Thesis objectives

A key step in image and video compression algorithms corresponds to the prediction stage. Indeed, in a typical LS-based still image coding method, the prediction stage allows to compute the detail wavelet coefficients. This step is often optimized by minimizing an objective function [2]. Moreover, in stereo image coding (resp. video coding), the prediction step aims at generating one view (resp. one future frame) from a reference view (resp. previous frame) based on the estimated disparity (resp. motion) field. This step is also optimized by minimizing the reconstruction error [3].

Recent advances in machine learning algorithms, and in particular in deep learning (DL) techniques, have shown the good performance achieved by such tools in different image processing applications. For instance, most of the first developed research works in DL have been devoted to computer vision tasks like image classification and recognition [4]. Recently, few research works related to inter-frame prediction and image compression have been published [5, 6, 7]. Indeed, the convolution neural networks (CNN) can be seen as feature extractors that transform the image and video into feature space with compact representation, which could be beneficial for image and video compression.

Motivated by the great success of deep learning techniques, the aim of this thesis is to further explore these techniques for improving the image coding performance. More precisely, the objectives of this thesis can be summarized as follows:

- (1) First, an overview of the existing deep convolution neural networks will be performed to identify the appropriate models for image prediction and compression. For instance, particular interest will be given to Fully Connected Networks (FCN), Recurrent Neural Networks (RNN), Auto-Encoder (AE) and Generative Adversarial Networks (GAN) architectures. A comparison study and complete analysis of these different architectures will also be carried out to evaluate their performances and better understand their advantages and limitations.
- (2) Second, and in order to design adaptive image coding methods, we will propose to resort to the appropriate neural networks models to optimize the prediction filters in a conventional lifting scheme-based image coding. This task can be achieved using global as well as local (i.e. patch)-based approaches. We should note that these approaches will be firstly validated in the context of single image coding and then applied to stereo images coding.
- (3) Third, as a continuity of the recent inter-frame prediction research works, the best architecture retained in the previous task (2) will be exploited to improve the prediction of one target view from a reference one in the context of stereo image coding. Due to the similarity of this problem with the video compression one, the developed method may also be validated in the context of video compression.
- (4) Finally, since the main problem with deep learning techniques is the burdens in computation and memory, we will also focus on the efficiency and impact of the different network parameters to obtain efficient compression algorithms appropriate for practical applications.

III. References

- [1] M. Kaaniche, A. Benazza-Benyahia, B. Pesquet-Popescu and J.-C. Pesquet, Non separable lifting scheme with adaptive update step for still and stereo image coding, *Elsevier Signal Processing: Special issue on Advances in Multirate Filter Bank Structures and Multiscale Representations*, vol. 91, no. 12, pp. 2767-2782, 2011.
- [2] M. Kaaniche, B. Pesquet-Popescu, A. Benazza-Benyahia and J.-C. Pesquet, Adaptive lifting scheme with sparse criteria for image coding, *EURASIP Journal on Advances in Signal Processing: Special Issue on New Image and Video Representations Based on Sparsity*, vol. 2012, 22 pages, 2012.
- [3] G. Dauphin, M. Kaaniche and A. Mokraoui, Block dependent dictionary disparity compensation for stereo image coding, *IEEE International Conference on Image Processing*, pp. 4868-4872, Québec, Canada, 2015.
- [4] A. Krizhevsky, I. Sutskever and G. E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, *Advances in neural information processing systems*, vol. 25, no. 2, pp. 109-1105, 2012.
- [5] J. Li, B. Li, J. Xu, R. Xiong, and W. Gao, Fully Connected Network-Based Intra Prediction for Image Coding, *IEEE Trans. on Image Processing*, 2018.
- [6] L. Theis, W. Shi, A. Cunningham, and F. Huszar, Lossy image compression with compressive autoencoders, *International Conference on Learning Representations*, 2017.
- [7] G. Toderici, D. Vincent, N. Johnston, S. J Hwang, D. Minnen, J. Shor and M. Covell, Full resolution image compression with recurrent neural networks, *Computer Vision and Pattern Recognition*, 2017.

Supervision:

Main supervisor: Mounir Kaaniche, Associate professor (HDR) at L2TI, Université Paris 13.
Email: mounir.kaaniche@univ-paris13.fr

Co-supervisor: Gabriel Dauphin, Associate professor at L2TI, Université Paris 13.
Email: gabriel.dauphin@univ-paris13.fr

Salary: 1450 euros/month (Net salary)

Qualifications for applicant

- Master of science in relevant fields (machine learning, image processing, informatics, mathematics) and strong academic record
- Good knowledge in deep learning, mathematics, image and video processing
- Good programming skills (Python, Matlab, C++)
- Good oral and written communication skills

Application procedure

Applications should be sent by email to: mounir.kaaniche@univ-paris13.fr

The application should include:

- A detailed Curriculum Vitae
- A motivation letter explaining why the applicant believes to be suitable for the position offered
- University transcripts
- One or two reference letters

Application deadline is May 3rd, 2019 and the starting date is September 1st, 2019.