PhD Proposal 2017

School: Ecole Centrale de Lyon	
Laboratory: LIRIS CNRS UMR 5205	**Web site: http://liris.cnrs.fr**
Team: Imagine / Liris ECL	**Head of the team: Pr. Liming Chen**
Supervisor: Pr. Liming Chen, Dr. Emmanuel Dellandréa	**Email: liming.chen@ec-lyon.fr emmanuel.dellandrea@ec-lyon.fr**
Collaboration with other partner during this PhD:	**In China:**

Title: Bio-inspired deep computation models and their application to 2D+t visual data analysis
Scientific field: Computer Vision, Machine Learning
Key words: deep learning, computational model, psychological models, concept recognition

Details for the subject:

Background, Context:

Representation learning aims at learning transformations of raw input data into discriminative and compact representations that can be effectively exploited in machine learning tasks, e.g., classification, detection. This is motivated by the fact that real-world data, *e.g.*, images, video and sensor measurements, is usually complex, noisy, redundant and highly variable. Deep learning allows computational model that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. In recent years, these methods have dramatically improved the state of the art in a number of fields, including object detection and recognition. However, the structure of these models is arbitrary and based on a simplistic modelling of the human brain. Thus, using a bio-inspired approach to deep networks could allow a better understanding of the model behavior and higher efficiency for high level semantic concept recognition from raw data. Therefore, we propose to exploit results from the neural basis of emotion and perceptual mechanisms in humans to enable principled design of deep models relying on appropriate features, best neural architectures, multi-modal fusion models between visual and audio features, for a more accurate and realistic concept prediction from raw data.

Research subject, work plan:

Among the psychological factors that can influence data perception and recognition, memory and affect are among the most important ones. However, computational models estimating the emotions elicited by visual data are a special case where the stimulus is a video scene. Thus, only specific sections of the psychological models could be first investigated in order to maximize the performance of computational models. First, psychologists suggest that the evaluation of an emotion is an iterative process. For example, Russell defines the Core Affect [1] which is a process of recursive and continuous evaluations, also at the core of the appraisal evaluation postulated by Scherer [14]. Thus, this key aspect of affect has to be considered in affective and memory sensitive computational models. However, most computational models do not take into account the recursivity of the emotional episode. This is not the case for HMM-based [2] and RNN-based [3] affective frameworks. Indeed, HMMs are statistical models of sequential data, inherently able to take into consideration consecutive emotional changes through hidden state transitions. However, they are composed of a specific number of discrete states and thus cannot be used to directly infer dimensional scores. Malandrakis et al. converted discrete affective curves obtained with HMMs into continuous curves using a Savitzky-Golay filter [4], but the continuous curves are thus approximations and cannot recover the precision lost by discretizing the affective space. RNNs-based affective frameworks are also able to take into account the temporal transitions between consecutive emotions. Combined with LSTM cells, they can learn efficiently long-term dependencies, as shown by previous work for video emotion recognition [3,8,9].

Beyond the recursive aspect of the models, their structures themselves can be inspired from neuroscience and psychological work. Indeed, psychological theories describe affect as generated through the interaction of bottom-up and top-down processes (but with emphasis on bottom-up process) [5], which is confirmed by neuroscience work demonstrating that both types of responses activate the amygdala (bottom-up responses activating the amygdala more strongly) [6]. Such descriptions of bottom-up and top-down responses can help future movie content analysis work to create computational models with appropriate structures as in other fields of studies [7].

Therefore, the aim of this thesis proposal is to leverage recent advances in computer vision and machine learning, and investigate novel machine learning algorithms in combination with neuroscience and psychologist models for the analysis of 2D+t data. A favored application of this investigation will be video analysis.

References:

[1] J. A. Russell, "Core affect and the psychological construction of emotion." Psychological Review, vol. 110, no. 1, pp. 145–172, 2003.

[2] M. Xu, J. S. Jin, S. Luo, and L. Duan, "Hierarchical movie affective content analysis based on arousal and valence features," in Proceedings of the 16th ACM international conference on Multimedia, ser. MM '08, 2008, pp. 677–680.

[3] M. Nicolaou, H. Gunes, and M. Pantic, "Continuous prediction of spontaneous affect from multiple cues and modalities in valence- arousal space," IEEE Transactions on Affective Computing, vol. 2, no. 2, pp. 92–105, April 2011.

[4] N. Malandrakis, A. Potamianos, G. Evangelopoulos, and A. Zlat- intsi, "A supervised approach to movie emotion tracking," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2011, pp. 2376–2379.

[5] H. Leder, B. Belke, A. Oeberst, and D. Augustin, "A model of aesthetic appreciation and aesthetic judgments," British Journal of Psychology, vol. 95, no. 4, pp. 489–508, Nov. 2004.

[6] K. N. Ochsner, R. R. Ray, B. Hughes, K. McRae, J. C. Cooper, J. Weber, J. D. Gabrieli, and J. J. Gross, "Bottom-up and top- down processes in emotion generation common and distinct neural mechanisms," Psychological science, vol. 20, no. 11, pp. 1322–1331, 2009.

[7] L. Itti, "Models of bottom-up and top-down visual attention," Ph.D. dissertation, California Institute of Technology, 2000.

[8] Y. Baveye, E. Dellandrea, C. Chamaret, and L. Chen, "LIRIS- ACCEDE: A video database for affective content analysis," IEEE Transactions on Affective Computing, vol. 6, no. 1, pp. 43–55, Jan 2015.

[9] Y. Baveye, E. Dellandrea, C. Chamaret, and L. Chen, "Deep learning vs. kernel methods: Performance for emotion prediction in videos," in Affective Computing and Intelligent Interaction (ACII), 2015.