PhD in Computer and Control Engineering

Research Title: Deep learning techniques for advanced audio-visual media understanding

Funded by	**Rai - Radiotelevisione Italiana (Italy)** *This PhD project will be in collaboration with the SmartData@Polito Interdepartmental Center*

Supervisors	Dr. Luca Cagliero (Politecnico di Torino), Dr. Alberto Messina (Rai Centre for Research and Technological Innovation, Italy)

Contact	Luca Cagliero (luca.cagliero@polito.it), Alberto Messina (alberto.messina@rai.it)

Context of the research activity	Cognitive data analysis consists in the set of algorithms, applications and computing platforms to perform tasks that mimic human intelligence. Image tagging, video highlight creation, automatic subtitling are just some examples of how cognitive data analysis can facilitate and speed up complex tasks usually done by hand in the media domain. Deep Neural Networks (DNN) is a family of technologies implementing such kind of systems. Despite there already exist a number of commercial solutions for cognitive understanding of multimedia data, many of which nowadays based on DNN, they are still far from being effectively usable in industrial practise. First, they are targeted at performing either very specific tasks within restricted domains (e.g. celebrity recognition) or very generic tasks (e.g., image tagging). Second, the success of DNN systems relies on the availability of very large training (i.e. labelled) data sets, called "ground truth". However, manual generation of the ground truth is a long, cost and error-prone process and available data are normally little useful in practical cases. Finally, almost all of state-of-art DNN architectures are mono-modal, i.e. they analyse input media sources such as text, audio, image and video data one at a time, and mono-tier, e.g. they analyse a video as a static sequence of un-related images, little considering temporal or other relations that characterise the content over its timeline.

	This project is performed in collaboration with Rai - Radiotelevisione Italiana, Centre for Research and Technological Innovation, and has the objective to advance the state of the art in deep learning for audio-visual media analysis and understanding by addressing the aforementioned problems. Example research areas of interest includes but is not limited to the following: artificial intelligence, computer vision, audio and speech processing, natural language processing, semantic web. The research activity fits to the aims and scope of research of the SmartData@PoliTo interdepartmental centre.

Objectives	The objective of this PhD project is to develop advanced techniques for audio-visual media understanding applicable in media industry context, addressing the following problems:
	• Designing and implementing efficient semi-automatic ground truth generation methodologies based on background structured and semi-structured knowledge and available multi-modal (i.e. textual, acoustic and visual) data;
	• Devising new DNN architectures for multi-tier (i.e. considering spatial, temporal, causal and other types of relationships) audio-visual content analysis, indexing and representation;
	• Developing novel ontologies to
	   I. represent ground truth information
	   II. map multi-modal features to multi-modal semantic concepts;
	   III. define multi-tier relations between those concepts
	• Analysing the impact of the developed DNN solutions to the media production/archiving workflows in terms of accuracy performance (e.g. Precision, Recall, etc.) and cost-benefit ratio.

Skills and competencies for the development of the activity	Suitable candidates should have the following skills: • Very good programming at least in Matlab, python and C/C++ languages • Excellent reading/writing in English • Good machine learning and multimedia processing background • Good skills in knowledge representation and reasoning (ontologies, RDF, OWL) • Some background in web programming technologies and database management would be a plus