Thesis topic : multimodal data analysis for scene understanding through curriculum learning

Multimodal learning, that is learning from different modalities, has proven to be an effective way of leveraging multiple information for better decision making [1]. In particular, in the context of computer vision, multimodal visual data provide complementary information about the same scene, thus enhancing the accuracy and robustness of complex scenes analysis. Several approaches have been proposed in literature which mainly focus on the best fusion strategy of the different modalities [2].

However, as the sensing capabilities and properties of different visual sensors differ a lot, it is not obvious in which order to use the available data for training. In general, the data is assumed to be perfectly aligned, i.e. different sensors capture the same scene at the same time and from the same point of view. If this help the fusion step, it also a limitation for practical applications where the different cameras capture data in an asynchronous way.

Moreover, in traditional machine learning algorithms, all training examples are randomly presented to the model, ignoring the various complexities of data samples and the current learning status of the model. Whereas, it has been show that carefully selecting the order in which to present training data for learning improves the generalization capacity and convergence rate of the model [3]. This learning strategy is known as *curriculum learning* and consist in training a model from easier to harder examples. The basic idea is to train a machine learning model with easier data subsets (or easier subtasks) and gradually increase the difficulty level of the data (or subtasks) until the whole training dataset is used [4].

The goal of this project is to explore curriculum learning methods for multimodal visual data analysis. In particular, we aim to leverage multimodal information to answer the following questions:

- how to measure difficulty? That is how to decide the relative “easiness” of each training example?
- how to select examples? That is how to decide the order in which data are used in the training process?

Each of these questions will require using the available information from different modalities, because a data sample considered “easy” in one modality can be “difficult” in another and vice-versa.

Therefore, this research topic opens news questions never addressed before in the literature and can provides new ideas and methods for different tasks such as segmentation, detection or recognition.

References

[1] Baltruaitis et al. “Multilmodal machine learning : a survey and taxonomy”, IEEE PAMI, 2017

[2] Zhang et al. “Deep multimodal data fusion for semantic image segmentation: a survey”, Image and Vision Computing, 2021

[3] Bengio et al. “Curriculum learning”, ICML, 2009

[4] Wang et al. “A survey on curriculum learning”, IEEE PAMI, 2021

Working conditions

The thesis work will be carried out at the IBISC lab under the supervision of Prof. Désiré Sidibé Adequate and necessary materials will be made available for the phd student to carry the project in the best conditions.

Contact: drodesire.sidibe@univ-evry.fr