Foreground segmentation to interpret convolutional neural networks

Environment: QARMA (machine learning) Team at Laboratoire d'Informatique et Systèmes (LIS)
Location: Ecole Centrale Marseille (ECM), Technopôle de Château-Gombert, Marseille
Supervisors: R. Sicre (LIS – ECM), S. Ayache (LIS - Polytech), F. Torres (LIS-ECM)
Salary: legal minimum
Keywords: computer vision, deep learning, interpretability
Contact: ronan.sicre@lis-lab.fr

Computer vision and deep learning received a lot of attention lately due to the great improvements brought by Deep Neural Networks. Over the last decade, these networks are addressing more complex tasks and are reducing their requirement for large amounts of annotated data. However a major limitation remains regarding the interpretability of such highly complex models.

During this internship we will focus on the interpretability of Convolutional Neural Networks (CNN) and more specifically we aim at providing precise saliency maps that highlight the area of an image that is the most important for a given decision.
Numerous methods generating such saliency maps derive from Class Activation Maps (CAM) [1]. GradCAM [2] and GradCAM++ [3] use back-propagated gradient information, while ScoreCAM [4] is based on layer activations. Alternatively several methods address salient object detection or segmentation, which offer a foreground mask of the object that offers interpretability properties. Moreover, weakly supervised semantic or instance segmentation start receiving attention and use similar strategies to produce segmentation masks with only image-level labels.

The candidate will first focus on understanding evaluation protocols of both CAM related methods and salient object segmentation. Our goal is to introduce new evaluation datasets and metrics based on segmentation tasks [5], for interpretability methods.

Then, we want to build an architecture capable of producing or learning improved saliency maps, using only image-level labels. This model should offer improvement on the newly proposed metrics, using erasing methods [6] or affinity learning [7].

[1] Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning deep features for discriminative localization. In CVPR.
[2] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision* .
[3] Chattopadhay, A., Sarkar, A., Howlader, P., & Balasubramanian, V. N. (2018, March). Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In *2018 IEEE winter conference on applications of computer vision (WACV)* IEEE.
[4] Wang, H., Wang, Z., Du, M., Yang, F., Zhang, Z., Ding, S., ... & Hu, X. (2020). Score-CAM: Score-weighted visual explanations for convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*.
[5] Shuhan Chen, Xiuli Tan, Ben Wang, and Xuelong Hu. (2018) Reverse Attention for Salient Object Detection, ECCV.
[6] Kunyang Sun, Haoqing Shi Zhengming Zhang Yongming Huang (2021) ECS-Net: Improving Weakly Supervised Semantic Segmentation by UsingConnections Between Class Activation Maps. ICCV.
[7] Lian Xu, Wanli Ouyang, Mohammed Bennamoun, Farid Boussaid, Ferdous Sohel, Dan Xu (2021) Leveraging Auxiliary Tasks With Affinity Learning for Weakly Supervised Semantic Segmentation. ICCV.