Master Internship: Deep architecture to learn discriminative image parts

Environment: QARMA (machine learning) Team at Laboratoire d'Informatique et Systèmes (LIS)
Location: Centre de Mathématique et d'Informatique (CMI), Technopôle de Château-Gombert, Marseille
Supervisors: R. Sicre (LIS – ECM), T. Artieres (LIS – ECM), S. Ayache (LIS - Polytech)
Salary: legal minimum
Keywords: computer vision, deep learning, image classification, part-based models

The proposed internship is in the field of computer vision, pattern recognition, and machine learning and it more specifically focuses on image classification. Image recognition has been intensively studied in both academia and industry and a large variety of methods exist that aim at building compact, precise, and transferable representation from which classification may be performed accurately.

While Deep Learning has popularized the idea that deep neural architectures may automatically build relevant and informative representations from raw data, alternative works show the interest of representation based on image parts that are learned from training data. In such works, a vocabulary of latent parts is learned per category before building an image representation based on the occurrence of certain parts (without any extra labels). It can be viewed as an Embedding of features, where features are rather high level and more related to the presence or absence of specific parts.

Several work focused on learning part-like representations for image classification [1], [2]. These methods showed superior performance in the cases of complex problems, such as fine-grained classification or few shot learning. Similarly, our model is based on matching image regions to part representations with an iterative soft-assignment algorithm [4], [3], see figure. These works showed the importance of part-based representations to better describe images and discriminate similar classes. Following these approaches, the goal of this internship is to continue the ongoing work on the translation of the part learning process into a new deep CNN architecture that can address the current limitations of part-based methods and allow end-to-end training at larger scale.

Several directions will be studied towards such a translation.

- First, we should adapt the standard convolutional architecture to allow for operations, such as: learning features to represent image regions, learning discriminative parts, representing an image based on the parts (encoding), and finally, learning a task-dependent function (e.g. a classifier).

- Second, a set of constraints should be applied to the architecture to introduce the specificities of parts as in our previous models [4], *e.g.* orthogonality between layers weights to build parts that are different one from another, spatial competition through NMS (non-maximal suppression), and sparsity constraints on activations to satisfy matching constraints between regions and parts.

- Third, we would like to further improve the parts capabilities, by creating hierarchical parts and by learning them in an unsupervised fashion [3].

Contact: ronan.sicre@lis-lab.fr

References

[1] Carl Doersch, Abhinav Gupta, and Alexei A Efros. Mid-level visual element discovery as discriminative mode seeking. In *Advances in Neural Information Processing Systems*, pages 494–502, 2013.

[2] Mayank Juneja, Andrea Vedaldi, C V Jawahar, and Andrew Zisserman. Blocks that shout: Distinctive parts for scene classification. In *International Conference on Computer Vision and Pattern Recognition*, pages 923–930. IEEE, 2013.

[3] Ronan Sicre, Yannis Avrithis, Ewa Kijak, and Frédéric Jurie. Unsupervised part learning for visual recognition. In *Computer Vision and Pattern Recognition*, 2017.

[4] Ronan Sicre and Frédéric Jurie. Discriminative part model for visual recognition. *CVIU*, 2015.