Unsupervised learning of discriminative representation for image recognition

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), T. Artieres (LIS – ECM), S. Ayache (LIS - Polytech)
Salary: legal minimum
Keywords: computer vision, deep learning, image recognition, unsupervised learning

Context

As for machine learning, computer vision has witnessed a core change with the recent repopularization of Deep Neural Networks (DNN) at the end of 2012 [6]. For the first time in several years, DNN largely outperformed previous methods on large scale image classification on the ImageNet dataset [3]. Within a few years, DNN have been applied to various problem, such as image retrieval, detection, instance segmentation, etc. These methods benefit from the pretraining of parts of the network on large annotated corpus to gain superior visual description. DNN now allow the learning of complex non-linear classifiers on large amounts of data and are capable of generalization and adaptation.

However one major limitation persist, which is the need for large amounts of annotated data. The community gained interest on this problem and developed fields such as transfer learning, few-shot learning, unsupervised or self-supervised learning.

We are particularly interested in unsupervised learning in the case of transfer. While the common trend in transfer is to fine-tune a network to a specific recognition task. We observe two directions in the recent works on unsupervised image representation learning.

- Several recent methods focus on learning from large uncurated datasets from scratch. The work of [4] augments an image dataset by taking 4 images orientations and learns to classify each orientation. Jigsaw similarly propose to learn permutations of images regions [7], while [12] learn colorization from grayscale images. Then, deep cluster propose an iterative process, where images are clustered to produce pseudo-labels and classification is learned on the pseudo labels [1]. Moreover, [2] further combines the two ideas and add hierarchical clustering. These processes allow to learn most parameters of a DNN, however these are not adapted to transfer with small amounts of data.

- Other methods focus on adapting or learning image representation in an unsupervised fashion for related tasks. For instance, people study unsupervised detection or use detection module to encode salient parts of images for retrieval [5]. These methods are afficient enough to scale to large datasets.

A few related work focus on unsupervised latent intermediate representations. The work of [10] formulate the problem as unsupervised discriminative clustering on a large dataset of image patches. Similarly, our work [9] learns a set of discriminative part in an unsupervised fashion. Other approaches apply similar methods to fine-grained image recognition [11]. Unfortunately, such methods are rarely end-to-end and are too expensive to scale.

Research project

We are particularly interested in unsupervised learning in the case of transfer. We search method capable of adapting pre-trained representation to new complex tasks without supervision. Problems such as image retrieval, and fine-grained recognition are directly linked to our objectives.

Several works show that using standard large networks on complex problems such as fine-grained recognition or retrieval is not competitive. The best methods often rely on large computation, additive labeled data, complex multi-stage systems.

Our goal is to propose methods that overcome these limitations, by learning unsupervised latent intermediate representations in a light and end-to-end fashion. Specifically, we will first focus on translating latent representations learning such as [9, 8] in an efficient end-to-end learnable architecture. We will further adapt such architecture to unsupervised learning, taking inspiration from [9] and [1]. We believe that a compact architecture with few parameters will enable efficient and unsupervised learning.

Finally, we will study how such architecture can help interpreting the network decision. One idea would be to link part regions to the final classification decision, by adapting the Class Activation Mapping [13]. More specifically the architecture must be adapted to obtain a fully convolutional architecture followed by one unbiased fully connected layer, in the case of classification. One can discover the participation of each part, that can be linked to the higher scoring regions in the image.

Contact: ronan.sicre@lis-lab.fr - thierry.artieres@lis-lab.fr

References

[1] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. Deep clustering for unsupervised learning of visual features. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 132–149, 2018.

[2] Mathilde Caron, Piotr Bojanowski, Julien Mairal, and Armand Joulin. Unsupervised pre-training of image features on non-curated data. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2959–2968, 2019.

[3] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition*, pages 248–255. IEEE, 2009.

[4] Spyros Gidaris, Praveer Singh, and Nikos Komodakis. Unsupervised representation learning by predicting image rotations. *arXiv preprint arXiv:1803.07728*, 2018.

[5] Albert Gordo, Jon Almazán, Jerome Revaud, and Diane Larlus. Deep image retrieval: Learning global representations for image search. In *European Conference on Computer Vision*, pages 241–257. Springer, 2016.

[6] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[7] Mehdi Noroozi and Paolo Favaro. Unsupervised learning of visual representations by solving jigsaw puzzles. In *European Conference on Computer Vision*, pages 69–84. Springer, 2016.

[8] Sobhan Naderi Parizi, Andrea Vedaldi, Andrew Zisserman, and Pedro Felzenszwalb. Automatic discovery and optimization of parts for image classification. In *International Conference on Learning Representations*, 5 2015.

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

[10] Saurabh Singh, Abhinav Gupta, and Alexei A Efros. Unsupervised discovery of mid-level discriminative patches. In *Proceedings of the European Conference on Computer Vision*, pages 73–86. Springer, 2012.

[11] Jian Zhang, Runsheng Zhang, Yaping Huang, and Qi Zou. Unsupervised part mining for fine-grained image classification. *arXiv preprint arXiv:1902.09941*, 2019.

[12] Richard Zhang, Phillip Isola, and Alexei A Efros. Colorful image colorization. In *European conference on computer vision*, pages 649–666. Springer, 2016.

[13] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2921–2929, 2016.