Titre : Incremental learning with deep convolutional networks.

Sujet proposé dans : M2R Informatique / Sciences Cognitives, Projet

Responsable(s):
Georges Quénot (Georges.Quenot@imag.fr) LIG-MRIM
Denis Pellerin (Denis.Pellerin@gipsa-lab.grenoble-inp.fr) GIPSA-lab

Mots-clés : Deep learning, multimedia indexing.
Durée du projet : 5 mois, possibilité de continuer en thèse.
Nombre maximal d'étudiants : 1
Places disponibles : 1

Description

Incremental learning with deep convolutional networks.

Considerable amounts of multimedia data are continuously made available online (e.g. Flickr or YouTube) or broadcasted. Access to them or the large-scale analysis of their content is currently still limited either because initial textual annotation is non-existent, incomplete or incorrect or because an interpretation of the "raw" content (pixels or audio samples) is very difficult due to the "semantic gap" between them and the concepts that have meaning and relevance to humans.

Deep Convolutional Neural Networks (DCNN) have recently made a significant breakthrough in image classification [1]. This has been made possible by a conjunction of factors including: findings about how to have deep networks effectively and efficiently converge [2], the use of convolutional layers [3][4], the availability of very powerful parallel architectures (GPUs), findings about how exactly a network should be organized for the task [1], and the availability of huge quantity of cleanly annotated data [5].

Even with dedicated architectures and clever designs, these networks still takes time to train (from days to weeks) and are then restricted to the type of data and to the list of target categories used for their training. They can partially be re-used for other data types and/or other categories [6][7] but this generally requires a complete re-training of a significant part or of all of the whole network for each new data type or each new set of target categories.

The proposed subject is to study methods for incrementally retrain or adapt a DCNN at a low marginal computing cost each time new training data samples and/or new target concept categories are introduced. The system should be able to smoothly and effectively evolve while new information or requirements are given to it. In particular, it should be able to improve itself each time a user detects that it made an error on a given test sample (active learning). Both the learned weights for a given architecture and the network architecture itself are expected to evolve.

A number of already trained DCNN architectures are currently available and can be used as starting points. A number of annotated image and video collections are also available for conducting experiments. A complete DCNN development package (caffe [8]) is also available.

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in Advances in Neural Information Processing Systems 25, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105.

[2] G. B. Orr and K.-R. Mueller, Eds., Neural Networks: Tricks of the Trade, ser. Lecture Notes in Computer Science, second edition. Springer, 2012, LNCS 7700.

[3] K. Fukushima, “Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position,” Biological Cybernetics, vol. 36, pp. 193–202, 1980.

[4] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” in Proceedings of the IEEE, 1998, pp. 2278–2324.

[5] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A Large-Scale Hierarchical Image Database,” in CVPR09, 2009.

[6] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, “How transferable are features in deep neural networks?” CoRR, vol. abs/1411.1792, 2014.

[7] A. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, “CNN Features Off-the-Shelf: An Astounding Baseline for Recognition,” in IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 2014, pp. 512–519.

[8] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, “Caffe: Convolutional architecture for fast feature embedding,” in Proceedings of the ACM International Conference on Multimedia. ACM, 2014, pp. 675–678.