Deep learning for rare class recognition

Context

Automatic image and video recognition have strongly progressed these last years, after the introduction of deep learning techniques. It is currently possible to obtain good performance results for the classification of thousands of different objects. However, one limitation of existing deep learning methods is related to their strong dependence on the available volume of data per class. A a consequence, the recognition of poorly represented classes (i.e. only few training images), which make for a majority of surrounding objects, is currently suboptimal. Some early works have approached this research topic by exploiting data belonging to classes that are statistically close to target classes. This PhD topic will capitalize on recent advances in the field and will notably focus on : (1) adaptation of deep learning approaches to rare classes and (2) exploiting semantic and visual similarities between concepts to transfer knowledge from well represented classes to rare ones. Given the prevalence of rare classes in different domains, the results of the work can be exploited in multiple applications, including : (1) e-commerce - with an improved recognition of products and brands ; (2) e-tourism - via a fine recognition of tourist points of interest and of artworks and (3) privacy protection on online social networks - with the creation personalized recognition models for each user. One or several of these application domains will be selected in order to illustrate the results experimentally.

State of the art

Automatic visual recognition witnessed significant progress with the development of deep learning techniques. This progress is mainly explained by theoretical advances that have allowed the efficient implementation of backpropagation for deep network architectures (Hinton et al., 2006). In image recognition (Ciresan et al., 2012), image representation are built in a bottom-up fashion, starting from raw data (i.e. pixels), going through low-level features in the early network layers and outputting semantic representations (i.e. classes). Two practical aspects have also contributed to recent advances in computer vision: (1) the development of powerful GPGPU[1] architectures and (2) the availability of large datasets such as ImageNet[2].
Existing deep network architectures require for a large number of images to be available for each learned class in order to obtain good quality recognition models. For common classes, example images are either already included in ImageNet or can be obtained from search engines, like Google or Bing Images, and the labeled. For the orger classes, which represent the majority of the objects that surround us, the number of associated images is insufficient to efficiently train deep learning models. Among the solutions that were explored to alleviate the lack of data, we can cite :

- Different methods that tackle the imbalance in the number of images per class in a dataset. (He et Garcia, 2009) indicate that the main balancing technique exploits a data sampling that modifies the data distribution in order to create a more balanced dataset. This sampling is problematic for deep networks because it reduces the much needed training examples.
- Knowledge transfer after training a deep architecture on an application domain. Our team contributed to this approach with a promising method that mixes generic and specific classes in order to improve image classification (Tamaazousti et al., 2017). We have also

1 GPGPU *: general purpose processing on graphics processing units* ; effectuer un calcul générique au moyen d'une carte graphique (centrée autour d'un GPU) pour bénéficier de ses capacité de traitement parallèle.

2 http://www.image-net.org/

used transfer learning to automatically geolocate images by exploiting a model learned with tourist points of interest (G. Kordopatis-Zilos et al., 2015). In spite of good performance, this method is limited insofar it requires a sufficiently large seed dataset for the target domain.

- Data augmentation that is classically implemented using geometric transformations or image cropping for the training images (Simonyan et Zisserman, 2015). More recently, domain related augmentations were successfully included in face recognition pipelines (Masi et al., 2016). These transformations include automatic face alignment and facial expression modifications. Data augmentation is an interesting way to follow, especially via the creation of synthetic images, as proposed by (Masi et al., 2016). However, the proposed methods should be generic enough in order to be applicable to different application domains.
- Few shot learning, in its instantiation that exploits siamese networks, was introduced by (Koch, 2015). These networks try to minimize intra-class distances while also maximizing inter-class distances. (Ravi et Larochelle, 2017) have very recently introduced a method that is based on a meta-learner that learns the optimization algorithm which will be exploited to train a learner when few examples are available. Their results are very encouraging and could constitute a very good starting point for future work.

Challenges

The main challenges that need to be addressed during the proposed PhD are:

1. The efficient visual representation of domains for the automatic creation of training datasets that are most efficient in application. Our team has strong experience in this field (Ginsca et al., 2015 ; Vo et al., 2017) that can be leveraged by the PhD candidate.
2. The automatic generation of training examples through the use of generative networks (Goodfellow et al., 2014) is a promising way to tackle data scarcity. The main challenge ahead is to propose domain independent methods and to integrate credible visual contexts for the objects that are inserted in synthetic images.
3. The integration of semantic relations between classes in order reduce imbalance in training datasets. It is interesting to study the possibility to transfer knowledge from richly represented classes to rare classes that are semantically linked to them.
4. Scalability of the proposed method in order to cope with the vast number of existing rare classes, especially when different domains need to be integrated in a unique classification application.

PhD outline

This PhD will start with a state of the art period of 3 – 4 months that will allow the candidate to understand relevant research topics. Technical domains that are most important for the topic are: deep learning, data augmentation and imbalanced learning. The rest of the first year will be dedicated to work focused on tackling one of the main challenges described above, whose application will seem the most promising. The year will be ideally concluded with one or two publications in international conferences.

The second year will be focused on tackling the other challenges. It should result in publications in top tier venues in the domains such as CVPR, ICCV, ACM Multimedia, ECCV or ICLR.

The final year will be dedicated to the integration of the work related to individual challenges et to the proposal of a working classification prototype that highlights rare classes. The second part of this year include the PhD thesis writing and the preparation for the PhD aftermath.

Admin and contact

The PhD is fully funded by CEA and is the successful candidate will work in the Multimedia Team of the Vision and Content Engineering Lab of CEA LIST (http://www.kalisteo.fr/en/).
The successful candidate will be enrolled with the graduate school of Ecole Centrale Paris.

CEA LIST is located on the Saclay Campus, 20 km south of Paris, France.

If interested, please send a CV and cover letter to adrian.popescu@cea.fr

References

Hinton, G.E., Osindero, S., The, T.-W. (2006) A fast learning algorithm for deep belief nets. Neural computation, Vol. 18, No. 7, Pages 1527-1554, July 2006.
Ciresan, D., Meier, U., Schmidhuber, J. (2012) Multi-column Deep Neural Networks for Image Classification, CVPR 2012.
Ginsca, A. et al. (2015) Large-scale Image Mining with Flickr Groups, Multimedia Modelling 2015 (best paper) Sydney, Australia
Goodfellow, Ian J.; Pouget-Abadie, Jean; Mirza, Mehdi; Xu, Bing; Warde-Farley, David; Ozair, Sherjil; Courville, Aaron; Bengio, Yoshua (2014). "Generative Adversarial Networks". arXiv:1406.2661
He, H. et Garcia, E., A. (2009) Learning from Imbalanced Data. IEEE TKDE, 21(9), 2009.
Koch, G. (2015) Siamese neural networks for one-shot image recognition. PhD thesis, University of Toronto, 2015.
Kordopatis-Zilos et al. (2015) CERTH/CEA LIST at MediaEval Placing Task 2015. MediaEVal 2015 Working Notes.
Masi, I. et al. (2016) Do We Really Need to Collect Millions of Faces for Effective Face Recognition? ECCV 2016.
Ravi, S. et Larochelle, H. (2017) Optimization as a model for few-shot learning.
Simonyan, K. and Zisserman, A. (2015) Very deep convolutional networks for large-scale image recognition. In International Conference on Learning Representations, 2015.
Tamaazoustim Y., Le Borgne, H., Hudelot, C. (2017) MuCaLe-Net: Multi Categorical-Level Networks to Generate More Discriminating Features. CVPR 2017.
Vo, P. V., (2017) Harnessing noisy Web images for deep representation, Computer Vision and Image Understanding (to appear).