Apprentissage semi-supervisé et faiblement supervisé pour la segmentation sémantique

Sujet de stage de M2 – printemps 2020

Contexte

La compréhension de scènes est un enjeu majeur de la recherche en perception artificielle. Il s'agit non seulement d'identifier les objets imagés mais aussi de comprendre les relations qui les lient : la fracture est-elle horizontale ou verticale, les livres sont-ils rangés, le piéton traverse-t-il ? Depuis quelques années, les approches les plus efficaces de l'état de l'art reposent sur des réseaux de neurones convolutifs profonds (CNN) permettant la détection ou la segmentation d'objets d'intérêt dans les images. Toutefois, le paradigme d'apprentissage supervisé demeure le plus populaire et les modèles profonds sont ainsi gourmands en annotations. Or, l'étiquetage exhaustif des objets, voire des pixels, d'une image est un procédé d'annotation coûteux et qui nécessite souvent l'intervention d'experts (par exemple, des médecins) dont le temps est précieux. Récemment de nouvelles approches d'apprentissage dites semi-supervisée ou faiblement supervisées [Dur+17] se sont intéressées à la réduction de la quantité et de la qualité des annotations nécessaires à l'obtension des performances à l'état de l'art en reconnaissance de formes, jusqu'à se passer entièrement d'annotations [Buc+19]. L'objet de ce stage est ainsi d'étudier les approches parcimonieuses en supervision pour la compréhension d'images.

Enjeux et objectifs

Il existe dans la littérature scientifique plusieurs approches de classification d'images semi-supervisées, par préentraînement non-supervisé [Car+18 ; NF16], propagation d'étiquettes [Rad+18 ; Kho+17] ou contraintes géométriques [Xie+19].

Le premier objectif de ce stage consiste à adapter les techniques de classification semi-supervisées et non-supervisées au problème de la segmentation sémantique. En effet, ces approches exploitent généralement des notions d'invariance ou d'équivariance à des objets à des transformations géométriques qu'il est possible de retrouver ou de modéliser dans le cadre de la compréhension de scènes. Par exemple, faire tourner un objet de 90° doit produire une segmentation où le masque de l'objet correspond à une même rotation du masque initial.

Le second objectif du stage est de réduire le niveau de supervision des exemples d'entraînement, c'est-à-dire d'apprendre à partir d'annotations moins fines ou incomplètes. En pratique, il est rare de pouvoir travailler sur des bases de données où les images ont été complètement annotées

au niveau pixellique et il est bien souvent nécessaire de se contenter d'annotations partielles ou grossières, plus rapides à obtenir.

La mise en application des méthodes développées pourra se faire sur différentes applications déjà étudiées au sein du laboratoire : conduite de véhicules autonomes, cartographie d'images satellitaires, analyse d'images médicales ou segmentation d'images naturelles génériques.

Profil recherché

Nous recherchons un ou une candidate de niveau master 2 ou école d'ingénieur avec une spécialité en mathématiques, en informatique ou en traitement du signal. Le ou la candidate doit démontrer un certain goût pour la recherche et des bases théoriques adéquates en apprentissage automatique, apprentissage profond et traitement d'image. Une aptitude à la programmation, de préférence avec Python, est indispensable. Une première expérience avec une bibliothèque d'apprentissage profond telle que TensorFlow ou PyTorch est un plus.

Les candidatures (incluant un CV, une lettre de motivation et un relevé de notes) sont à envoyer à Nicolas Audebert (nicolas.audebert@cnam.fr) et Nicolas Thome (nicolas.thome@cnam.fr).

Organisation

Le stage est prévu pour une durée de 5 à 6 mois avec un début modulable au printemps 2020. Il se déroulera au centre de recherche et d'études en informatique et en communications (CEDRIC, `https://cedric.cnam.fr`) du CNAM (`https://www.cnam.fr`) à Paris (3ème arrondissement). Le CEDRIC est un laboratoire fondé en 1988 rassemblant 75 enseignants-chercheurs regroupés dans 7 équipes thématiques. Ses activités couvrent divers champs de recherche allant de la fouille de données multimédia aux radiocommunications en passant par l'apprentissage statistique, les médias interactifs et l'optimisation combinatoire.

Le stage sera co-encadré par Dr. Nicolas Audebert (équipe Vertigo) et Prof. Nicolas Thome (équipe MSDMA).

Références

[Buc+19] M. Bucher, T.-H. VU, M. Cord et P. Pérez, "Zero-Shot Semantic Segmentation", in *Advances in Neural Information Processing Systems 32*, 2019, p. 466-477.

[Car+18] M. Caron, P. Bojanowski, A. Joulin et M. Douze, "Deep Clustering for Unsupervised Learning of Visual Features", in *The European Conference on Computer Vision (ECCV)*, 2018.

[Dur+17] T. Durand, T. Mordan, N. Thome et M. Cord, "WILDCAT : Weakly Supervised Learning of Deep ConvNets for Image Classification, Pointwise Localization and Segmentation", in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[Kho+17] A. Khoreva, R. Benenson, J. Hosang, M. Hein et B. Schiele, "Simple Does It : Weakly Supervised Instance and Semantic Segmentation", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[NF16] M. Noroozi et P. Favaro, "Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles", in *Computer Vision – ECCV 2016*, 2016.

[Rad+18] I. Radosavovic, P. Dollár, R. Girshick, G. Gkioxari et K. He, "Data Distillation : Towards Omni-Supervised Learning", in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[Xie+19] Q. Xie, Z. Dai, E. Hovy, M.-T. Luong et Q. V. Le, *Unsupervised Data Augmentation for Consistency Training*, 2019. arXiv : `1904.12848 [cs.LG]`.

Weakly and semi-supervised learning for semantic segmentation

Master internship – spring 2020

Context

Scene understanding is one of the major challenges in computer vision. Not only should the machine be able to identify objects in images, it should also be able to understand relationships between entities : is the bone fractured horizontally or vertically ? are the books on the shelf ? is the pedestrian crossing the road ? Semantic interpretation of images is at the core of the decision-making process for many applications : autonomous driving, satellite imaging, medical diagnosis…For a few years, state of the art techniques have relied on deep convolutional neural networks (CNN) that outperform traditional image processing pipelines for object detection and segmentation. Models such as Faster-RCNN [Ren+15] and U-Net [RFB15] allowed researchers and engineers to make a leap forward on many reference benchmarks. However the supervised learning paradigm remains the most popular one and deep networks have been especially data-hungry compared to shallow methods. Yet, exhaustive labeling of all the objects – or pixels – in images is very costly, especially when the annotation requires the knowledge of experts (for example doctors) whose time is precious. Recent works have investigated so called semi-supervised or weakly supervised [Dur+17] learning strategies to reduce the quantity and the quality of the annotations that are required to reach state of the art accuracies for object classification and segmentation. This internship aims at studying and designing new label-efficient approaches for image understanding.

Intership topic & goals

There are already several semi-supervised image classification techniques in the scientific literature, based on unsupervised (or self-supervised) pretraining [Car+18 ; NF16], label propagation [Rad+18] or geometric constraints [Xie+19].

The first goal of this internship is to transpose semi-supervised and unsupervised image classification strategies to the problem of semantic segmentation. These methods indeed leverage invariance and equivariance of objets to geometrical transformations that can be found in many scene comprehension tasks.

The second goal of the internship is to reduce the level of supervision required to train deep networks, i.e. to learn from coarser or incomplete labels. It is uncommon to work with databases

where images are fully annotated pixelwise and more than not we need to deal with partial labels that are faster to obtain.

Applicant profile

We are looking for an intern preparing a Master or an Engineering Diploma with a strong knowledge of machine learning and deep neural networks. The applicant must demonstrate a will to do research and master the fundamentals of machine learning and deep learning. A strong proficiency with the Python programming language is a must. A first experience with a deep learning software library such as Pytorch or Tensorflow is welcome.

Applications (including a resume, a motivation letter and grades for the current year) are to be sent at Nicolas Audebert (nicolas.audebert@cnam.fr) and Nicolas Thome (nicolas.thome@cnam.fr).

Location

This internship will last for 5–6 months with a start in Spring, 2020. It will take place in the Centre de recherche et d'études en informatique et en communications (CEDRIC, `https://cedric.cnam.fr`), the computer science research department of CNAM (`https://www.cnam.fr`) in Paris. The CEDRIC is a computer science laboratory created in 1988. It is comprised of 75 permanent researchers divided in 7 teams. Its research topics go from multimedia data mining to telecoms, machine learning and interactive medias.

The internship will be co-directed by Dr. Nicolas Audebert (Vertigo team) and Pr. Nicolas Thome (MSDMA team).

Références

[Car+18] M. Caron, P. Bojanowski, A. Joulin et M. Douze, "Deep Clustering for Unsupervised Learning of Visual Features", in *The European Conference on Computer Vision (ECCV)*, 2018.

[Dur+17] T. Durand, T. Mordan, N. Thome et M. Cord, "WILDCAT : Weakly Supervised Learning of Deep ConvNets for Image Classification, Pointwise Localization and Segmentation", in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[NF16] M. Noroozi et P. Favaro, "Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles", in *Computer Vision – ECCV 2016*, 2016.

[Rad+18] I. Radosavovic, P. Dollár, R. Girshick, G. Gkioxari et K. He, "Data Distillation : Towards Omni-Supervised Learning", in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[Ren+15] S. Ren, K. He, R. Girshick et J. Sun, "Faster R-CNN : Towards Real-Time Object Detection with Region Proposal Networks", in *Advances in Neural Information Processing Systems 28*, 2015.

[RFB15] O. Ronneberger, P. Fischer et T. Brox, "U-Net : Convolutional Networks for Biomedical Image Segmentation", in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 2015.

[Xie+19] Q. Xie, Z. Dai, E. Hovy, M.-T. Luong et Q. V. Le, *Unsupervised Data Augmentation for Consistency Training*, 2019. arXiv : `1904.12848 [cs.LG]`.