September 29th 2019

Post-doctorate topic: Semi and Weakly supervised Convolutional Networks for jointly people detection and tracking

In this proposal, we focus upon the problem of joint people detection and tracking in computer vision, which is a foundation stone for several applications such as health monitoring and autonomous vehicles. Most contemporary systems for people detection and tracking make use of fully-supervised techniques. One downside of fully-supervised machine learning techniques is their need for large amounts of annotated data. Data annotation is a time-consuming and costly undertaking which suffers from several technical challenges such as inter and intra-annotator variance. Thus, dependence on fully-supervised techniques is unable to take advantage of large amounts of unannotated data in the form of videos which are available easily in this digital age. In comparison to fully annotated datasets, unannotated datasets are much difficult to work with; begin devoid of any supervisory information in the form of ground-truth. One practical approach towards reducing one's dependence on extensive annotations is to build systems, which can learn effectively from partial or incomplete annotations. There are two paradigms of machine learning, which deal with partial and incomplete annotations – semi-supervised learning and weakly supervised learning. In the context of joint detection and tracking of people, semi-supervised learning models the scenario where the training data is annotated with bounding boxes and tracking information for a subset of people (i.e. all people in an image or video are not annotated). Weakly supervised learning considers the scenario where, no bounding box or tracking information is available for any person; only a label providing information as to whether a person exists in a given frame or not is provided. A potential solution to these two scenarios has major ramifications as they preclude a complete and exhaustive annotation of large amounts of data. This translates into major cost savings, not to mention access to a practical way to harness large amounts of unannotated data available.

An understanding of the feature space topology as well as the interpretation of deep neural networks is essential to the solution of our problem in the aforementioned paradigms. This proposal therefore opens the route of opportunities for understanding and interpreting deep learning models rather than considering them as black box models. Tools from differential geometry and topology can be employed to study the feature space topology and understanding topological patterns of features representing people. We aim to develop a strong knowledge base in these areas and work towards interpreting deep neural networks in weakly and semi-supervised scenarios for the solution of this problem.

We will validate our approach against publicly available annotated datasets such as Caltech, BDD100K and MOT. In particular, we will test the localization performance of our weakly supervised approach vis-a-vis that of fully supervised approaches.
In addition, to demonstrate the generalization performance of our work, we will utilize OpenImages dataset which comes with partial annotations (i.e not all people in an image are labelled) for qualitative evaluation.

We outline the major research topics we aim to address:
1. Evolution of feature space topology in deep neural networks

CENTRE DE RECHERCHE
SOPHIA ANTIPOLIS - MÉDITERRANÉE
2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex
Tél. : +33 (0)4 92 38 77 77
Fax : +33 (0)4 92 38 77 65

www.inria.fr

Training of deep neural networks takes place over large number of iterations; each iteration modifying the network weights and thereby changing the feature space topology. It has been shown using visualization techniques like t-SNE, which as a network converges; the feature space is partitioned into C clusters where C is the number of classes in a given problem. Feature spaces in deep neural networks are Riemannian manifolds and the evolution of a feature space through training iterations is therefore a sequence of manifolds. Thus while the nature of its convergence (assuming convergence of the network) is known, little is known about the nature of this sequence prior to convergence. Uncovering this feature space topology lends us an understanding of how features of various object classes are learnt relative to each other -- an important basis to formulate robust loss functions and training strategies for weakly supervised detection and tracking.

2. Bayesian non-parametrics as a weakly supervised learning strategy
Non-parametric Bayesian systems adjust their complexity in accordance with new data. An assortment of rich tools such as Dirichlet processes and Indian buffet processes form a major cornerstone of non-parametric Bayesian models. However, there exists very sparse literature on the incorporation in deep learning based techniques. By their nature, non-parametric Bayesian methods can be used to impart greater generalization capacity to a learning system. This makes it of great interest to develop and work out methods to incorporate Bayesian non-parametric approaches into mainstream weakly supervised techniques.

These learning methods are grouped together under the umbrella title of weakly supervised learning, a term indicating learning from incomplete or inexact annotations [3].
Manual annotation in the field of pedestrian detection and tracking requires the identity and bounding box annotations localizing every pedestrian in a dataset. This is a time-consuming and costly process, which is prone to many challenges such as inter-annotator variance [1, 2]. Compared to bounding box annotations, collecting information about is much easier and faster.
Research in weakly supervised learning based detection and tracking systems has the capacity to make better and faster utilization of the wealth of image and video data available today. By eliminating the need for exact manually annotated bounding boxes, it removes a basic bottleneck in the machine learning pipeline. Considering the expected widespread applicability of surveillance applications, a weakly supervised learning system makes adjustments to previously trained systems to new environments (a process called fine-tuning in deep learning literature) easier and faster.

The application file should include:

- a detailed CV,
- a letter of motivation from the candidate related to the subject of the post-doc topic,
- the PDF of the thesis manuscript and some articles published as the first author,
- all documents deemed useful for evaluating the candidate's profile (letters of recommendation, etc.).

Contact: François Brémond <francois.bremond@inria.fr>
Sophia-Antipolis,
Research Director
Head of INRIA STARS Team
http://www-sop.inria.fr/members/Francois.Bremond/