Journal & Proceedings of the Royal Society of New South Wales, vol. 154, part 2, 2021, pp. 254–255. ISSN 0035-9173/21/020254-02

Thesis abstract

Zero-shot learning: recognition, tagging, and detection of novel concepts

Shafin Rahman

Abstract of a thesis for a Doctorate of Philosophy submitted to Australian National University, Australia

Recent advancements in deep neural networks have performed favorably well on the supervised object recognition task. Towards an ultimate automated visual recognition system, we identify three key shortcomings of the existing supervised learning approaches. First, the dependency on a significantly large volume of manually annotated examples (e.g., ImageNet dataset with ~10 million images limits the scalability of deep networks. Secondly, once the model learning stage is complete, it is difficult to add new classes continually when new data becomes available. Lastly, such models lack the notion of human-like understanding, i.e., an object can be recognized by humans without having any visual examples and just by a semantic description of its distinctive characteristics.

In this thesis, we investigate the zero-shot learning (ZSL) framework to address these limitations. ZSL aims to perform reasoning about previously unseen objects without observing even a single instance of them. Such a learning paradigm requires no visual examples of novel objects, no re-training to add new classes and does not rely solely on visual information. Considering the relationship among the semantic description of previously seen examples, ZSL incorporates human wisdom to the visual understanding developed by a machine. In this work, we specifically address three critical bottlenecks in ZSL research which give rise to three ZSL problem settings: (a) unified zero-shot recognition (ZSR), (b) zero-shot tagging (ZST), and (c) zero-shot detection (ZSD) of novel concepts. Established ZSR methods are not flexible enough to adapt to one/few-shot learning (O/FSL) scenario where one/few labeled examples of unseen classes become available during the supervised learning stage. To provide a comprehensive and flexible solution, we present a novel 'unified' approach for ZSL and O/FSL based on class adapting principal direction (CAPD) that computes class-specific discriminative information by relating the semantic description of categories. The primary objective is to learn a metric in the semantic embedding space that minimizes intra-class distances and maximizes inter-class distances. In the real-life scenario, instead of a single object per image, a scene may contain multiple seen and unseen concepts together. To adopt this, we present the DeepoTag approach for zero-shot tagging (ZST) to assign multiple labels to an input. This method considers both global and local details to discover seen or unseen concepts from a given scene. We solve this problem by formulating a multiple instance learning (MIL) framework. Unlike traditional MIL solutions, our method runs end-to-end without using offline object proposal generation methods. While most of the ZSL methods provide answers to unseen categories in simple tasks, e.g., single or multi-label classification and retrieval, we also focus on predicting both multi-class category-label and precise location of each instance in a given image. To this end, we introduce a new challenge for ZSL called zero-shot detection (ZSD) that simultaneously recognizes and localizes multiple novel concepts. Similar to traditional object detection, we present zero-shot version of double (Faster R-CNN) and single (RetinaNet) stage end-to-end object detectors. For both of the cases, we design associated loss functions that consider visual-semantic relationships to train the network. In addition to inductive learning approaches, we also propose the first transductive learning method for ZSD to reduce the domain-shift and model-bias against unseen classes convincingly. Our transductive approach follows a self-learning mechanism that uses a novel hybrid pseudo-labeling technique. Finally, we recommend training and testing protocols to evaluate ZSD based on large-scale ILSVRC-2017 and MSCOCO-2014 datasets. In summary, this thesis addresses three main ZSL tasks: recognition, tagging, and detection of novel concepts. It investigates different drawbacks of the current literature and establishes state-of-the-art solutions in each respective sub-task. In particular, the ZSD setting proposed in this thesis is highly challenging, and we hope our initial work will attract further efforts on this important and largely unsolved problem.

Dr Shafin Rahman
ECE
North South University
Bangladesh

E-mail: u5929575@alumni.anu.edu.au

URL: https://openresearch-repository.anu.edu.au/handle/1885/204349