Engineering Applications of Artificial Intelligence 50 (2016) 142–158

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Adapting pedestrian detectors to new domains: A comprehensive review

Kyaw Kyaw Htike [a,*], David Hogg [b]

[a] UCSI University, Kuala Lumpur, Malaysia
[b] University of Leeds, Leeds, UK

ARTICLE INFO

Article history:
Received 16 November 2015
Received in revised form
27 December 2015
Accepted 16 January 2016

Keywords:
Domain adaptation
Pedestrian detection
Feature learning
Scene-specific detector
Transfer learning

ABSTRACT

Successful detection and localisation of pedestrians is an important goal in computer vision which is a core area in Artificial Intelligence. State-of-the-art pedestrian detectors proposed in literature have reached impressive performance on certain datasets. However, it has been pointed out that these detectors tend not to perform very well when applied to specific scenes that differ from the training datasets in some ways. Due to this, domain adaptation approaches have recently become popular in order to adapt existing detectors to new domains to improve the performance in those domains. There is a real need to review and analyse critically the state-of-the-art domain adaptation algorithms, especially in the area of object and pedestrian detection. In this paper, we survey the most relevant and important state-of-the-art results for domain adaptation for image and video data, with a particular focus on pedestrian detection. Related areas to domain adaptation are also included in our review and we make observations and draw conclusions from the representative papers and give practical recommendations on which methods should be preferred in different situations that practitioners may encounter in real-life.

1. Introduction

Due to the fact that visual perception is vital to most intelligent life forms, *computer vision* has become one of the most important and active research areas in the field of Artificial Intelligence. Computer vision is about automatic analysis and understanding of visual data (such as images and videos) to extract useful information.

There are many sub-areas within the field of computer vision, one of which is object detection which forms the foundation of many intelligent scene understanding systems. Due to its significance, object detection has received a lot of attention in computer vision (Andreopoulos and Tsotsos, 2013).

Pedestrian detection in particular plays an important role in real world outdoor scenes, especially in urban areas. Although many proposed domain adaptation algorithms in literature could potentially be used for learning detectors for a variety of different object categories (such as pedestrians, cars, buses and bicycles), we focus on the task of domain adaptation for pedestrian detection since pedestrians are of most interest in many applications of computer vision.

2. Motivation

Pedestrian detection in monocular images is a challenging task and a lot of progress has been made in this area (Dollár et al., 2012; Enzweiler and Gavrila, 2009; Geronimo et al., 2010). Most state-of-the-art pedestrian detectors require a supervised training stage based on a labelled dataset that is obtained from manual annotation of pedestrians (*e.g.* delineation of pedestrians by bounding boxes) and a sufficient number of non-pedestrian images (Dalal and Triggs, 2005; Dollár et al., 2012; Girshick et al., 2011).

2.1. Training and generalisation

The objective of the labelled dataset is to provide the classifier (being learnt) with large intra-class variations of pedestrians and non-pedestrians so that the resulting classifier is *generalisable* to never-before-seen test data. This *generalisation* property is a sought-after property for most machine learning classification and regression tasks.

* Corresponding author at: School of Information Technology, UCSI University, Kuala Lumpur, Malaysia.
E-mail addresses: ali.kyaw@gmail.com (K.K. Htike), d.c.hogg@leeds.ac.uk (D. Hogg).

http://dx.doi.org/10.1016/j.engappai.2016.01.029
0952-1976/

When training a pedestrian detector, the goal is often: "Given *any* unseen test image, the detector should locate all the pedestrians in the image". In this paper, we term such a detector as a *generic (pedestrian) detector* and the training data from which the detector was trained as a *generic (pedestrian) dataset*.

2.2. Generic datasets

For a generic dataset, collected positive and negative examples are not (deliberately) limited to a particular scene and viewpoint and the aim of such a dataset is to collect as many variations of pedestrians as possible to produce detectors which should ideally perform well for any unseen test data. Examples of generic pedestrian datasets are INRIA Person Dataset (Dalal and Triggs, 2005), Daimler Mono Pedestrian Detection Benchmark Dataset (Enzweiler and Gavrila, 2009) and Caltech Pedestrian Dataset (Dollár et al., 2009). The INRIA dataset consists of images of upright people taken from a variety of personal image collections. Pedestrian training data of the Daimler and the Caltech datasets are extracted from videos recorded with on-board cameras in vehicles being driven around various places in urban traffic. All these datasets consist of training data from a variety of scenes and places, and as a result, the intra-class variations of pedestrians in such datasets is large. Fig. 1 illustrates this observation.

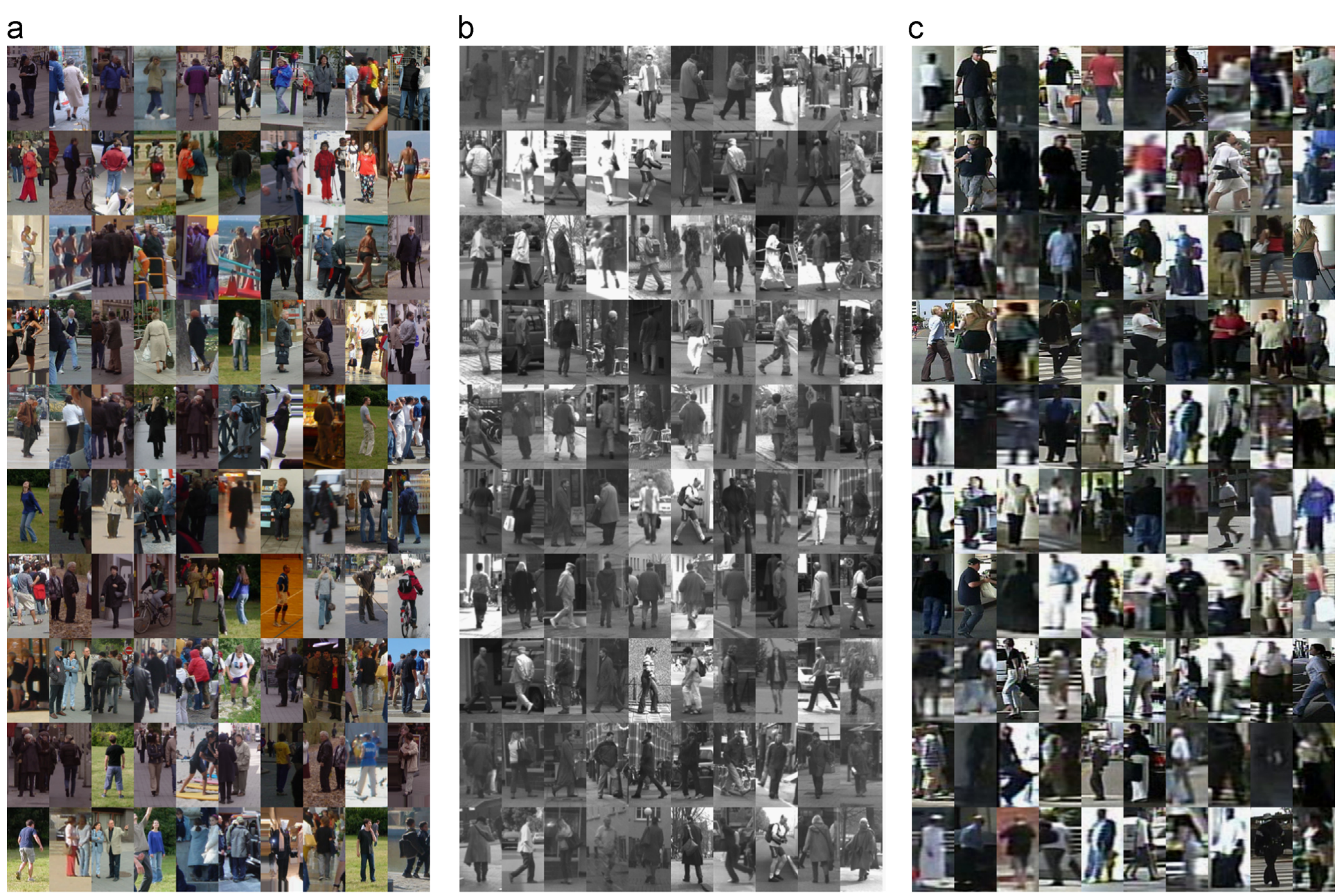
2.3. Problems with generic datasets

Despite the large intra-class variations present in such generic datasets, each of these datasets still has its own inherent bias. For example, since the INRIA dataset is taken from mostly personal digital image collections, many of the people in the training dataset are likely to be intentionally posing for cameras. This may be different from natural pedestrian poses and activities in real-life situations. For the Daimler and Caltech datasets, the pedestrians in the training set are biased to view-points and angles that cameras on-board vehicles could capture. Moreover, pedestrians from these datasets are taken from static images that have been captured using cameras fixed near the same ground plane as the captured pedestrians. This may be considerably different from situations where images of pedestrians are captured by video cameras looking down on a scene (*e.g.* surveillance videos).

2.4. Dataset bias

This dataset bias has been recently studied by Torralba and Efros (2011). No dataset can possibly cover a *representative* set of all the possible variations of pedestrians and non-pedestrians the detector is likely to face at test time. As shown by Dollár et al. (2009) and Roth et al. (2009), detectors may fail to perform satisfactorily when applied to scenes that differ from the original training data in many aspects such as:

- Pedestrian pose.
- Image or video resolution.
- View-point.
- Lighting condition.
- Image or video compression effects.
- Presence of motion blur.



INRIA dataset (Dalal and Triggs, 2005) Daimler dataset (Enzweiler and Gavrila, 2009) Caltech dataset (Dollar et al., 2009)

Fig. 1. Random samples from some generic pedestrian datasets (only pedestrians, *i.e.* positive examples, are shown).