Computer Vision and Image Understanding
journal homepage: www.elsevier.com

Multi-Perspective Cross-Class Domain Adaptation for Open Logo Detection

Hang Su[a,b,**], Shaogang Gong[b], Xiatian Zhu[c]

[a] *Shenzhen University, Shenzhen, CN*
[b] *Queen Mary University of London, London E1 4NS, UK*
[c] *Vision Semantics Limited, London E1 4NS, UK*

ABSTRACT

Existing logo detection methods mostly rely on supervised learning with a large quantity of labelled training data in limited classes. This restricts their scalability to a large number of logo classes subject to limited labelling budget. In this work, we consider a more scalable *open logo detection* problem where only a fraction of logo classes are fully labelled whilst the remaining classes are only annotated with a clean icon image (e.g. 1-shot icon supervised). To generalise and transfer knowledge of fully supervised logo classes to other 1-shot icon supervised classes, we propose a *Multi-Perspective Cross–Class* (MPCC) domain adaptation method. In a data augmentation principle, MPCC conducts feature distribution alignment in two perspectives. Specifically, we align the feature distribution between synthetic logo images of 1-shot icon supervised classes and genuine logo images of fully supervised classes, and that between logo images and non-logo images, concurrently. This allows for mitigating the domain shift problem between model training and testing on 1-shot icon supervised logo classes, simultaneously reducing the model overfitting towards fully labelled logo classes. Extensive comparative experiments show the advantage of MPCC over existing state-of-the-art competitors on the challenging QMUL-OpenLogo benchmark (Su et al., 2018).

1. Introduction

Logo detection is a long-standing computer vision problem (Doermann et al., 1993) with significant real-world applications ranging from brand trend prediction in smart business (Romberg et al., 2011; Romberg and Lienhart, 2013) to vehicle recognition in intelligent transportation (Pan et al., 2013) and document image logo retrieval (Pham, 2003). It is inherently challenging due to no clear definition of what makes a logo. The difficulty is further amplified by the presence of unconstrained contexts and varying logo instance scales (Fig. 1).

Existing logo detection methods have made progress on recognising a limited number of logo classes in most cases. They often exploit state-of-the-art object detection models such as Fast (Girshick, 2015) and Faster R-CNN (Ren et al., 2015), or YOLO (Redmon and Farhadi, 2017) that require supervised learning from large labelled training data per class. There have

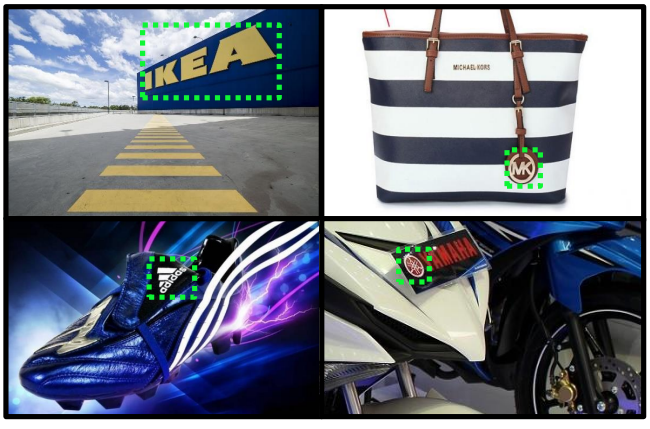

Fig. 1: Illustration of logo detection challenges.

**Corresponding author: Tel.: +86 13604887829;
e-mail: 9175ak@gmail.com (Hang Su)

Table 1: Statistics of logo detection datasets in the literature.

Dataset	Classes	Images	Availability
BelgaLogos (Joly and Buisson, 2009)	37	1,321	✓
FlickrLogos-27 (Kalantidis et al., 2011)	27	810	✓
FlickrLogos-32 (Romberg et al., 2011)	32	2,240	✓
Logo32plus (Bianco et al., 2017)	32	7,830	✓
Logo-In-The-Wild (Tzk. et al., 2018)	1196	9,393	✓
SportsLogo (Liao et al., 2017)	20	1,978	✓
MICC-Logos (Sahbi et al., 2012)	13	720	×
LOGO-NET (Hoi et al., 2015)	160	73,414	×
OpenLogo (Su et al., 2018)	352	27,083	✓

been a number of logo detection datasets developed in the literature (Joly and Buisson, 2009; Kalantidis et al., 2011; Romberg et al., 2011; Bianco et al., 2017; Tzk. et al., 2018; Liao et al., 2017; Sahbi et al., 2012; Hoi et al., 2015; Su et al., 2018) (Table 1). However, their scaling ability is limited in terms of both class and image, due to the high cost of collecting and labelling in-the-wild logo images.

There are a few attempts of tackling the scalability limitation in learning a logo detection model. For example, web logo image collection and annotation are explored in (Su et al., 2017a). The resulting dataset contains a high proportion of noisy labels and negative images with severe class imbalance. More recently, a new open logo detection setting is introduced (Su et al., 2018) where only a fraction of logo classes are associated with fully labelled training data and the objective is to train a detection model generalisable to 1-shot icon supervised logo classes. To this end, Su et al. (2018) develop a data augmentation method to generate context-consistent training images for 1-shot icon supervised logo classes that are weakly supervised by only a clean per-class icon image. This method directly handles the problem of lacking training data. However, it still suffers the domain shift problem between the synthetic training and genuine test images of unlabelled logos, despite improved consistency between logo objects and background context. This leads to model performance degradation.

In this work, we address the aforementioned limitation of open logo detection. Similar as (Su et al., 2018), we retain the use of synthetic training data for 1-shot icon supervised logo classes. Importantly, we further introduce a *Multi-Perspective Cross-Class* (MPCC) domain adaptation method. Specifically, MPCC takes as input the genuine training images of labelled classes, synthetic training images of 1-shot icon supervised classes, and auxiliary non-logo object detection images (e.g. MS COCO) simultaneously in model training. The aim is to transfer supervision information of labelled logo and non-logo instances in genuine scenes to 1-shot icon supervised logo classes by joint domain adaptation, whilst alleviating the underlying risk of model overfitting towards labelled logo classes. Unlike the conventional domain adaptation, this task focuses on cross-class knowledge transfer between genuine labelled images of logos and non-logo objects, as well as the synthetic images of 1-shot icon supervised logos.

The contributions of this work are: (1) We address the problem of domain shift between synthetic images of 1-shot icon supervised logo classes and genuine images of fully supervised logo classes in model training for open logo detection. This is the first attempt of addressing a *cross-class domain shift* problem for open logo detection. To this end, we provide a theoretical analysis of this cross-class domain shift problem in a probabilistic viewpoint, in order to achieve model learning for generalising to 1-shot icon supervised logo classes given a limited training set of labelled logo classes. (2) We formulate a *Multi-Perspective Cross-Class* (MPCC) domain adaptation method. By exploring unsupervised domain adaptation, MPCC aligns the feature distribution among synthetic logo images, genuine logo images, and non-logo object images in a joint model learning process. Extensive experiments show the performance advantage of the proposed MPCC method for open logo detection over state-of-the-art alternative approaches on the public QMUL-OpenLogo benchmark.

2. Related Work

Logo Detection. Most earlier methods for logo detection exploit hand-crafted visual features in sliding window localisation scheme. For example, SIFT features are often used for logo retrieval (Joly and Buisson, 2009), vehicle brand recognition (Psyllos et al., 2010) and brand logo matching (Sahbi et al., 2012). Common alternative representation options include bag-of-visual-word (Boia et al., 2014; Kalantidis et al., 2011; Revaud et al., 2012; Romberg and Lienhart, 2013), triangulation geometry representation (Kalantidis et al., 2011), and Histograms of Oriented Gradient (HOG) (Li et al., 2014). Due to remarkable success of deep learning (Nanni et al., 2017), recent state-of-the-art logo detection methods leverage generic object detection networks (Girshick, 2015; Ren et al., 2015; Redmon and Farhadi, 2017). For instance, Iandola et al. (2015) and Liao et al. (2017) exploit Fast R-CNN. Later on, Faster R-CNN are often selected (Hoi et al., 2015; Su et al., 2017b) due to the superior efficiency and performance. Universal logo characteristics is explored by considering a binary logo detection problem (Tzk. et al., 2018; Fehérvári and Appalaraju, 2019). To alleviate the training data labelling effort, web logo images with noisy annotation are mined for training detection (Su et al., 2017a). Commonly, these methods focus on *supervised learning* with the need for accurately labelling fine-grained object-

Fig. 2: Examples of (a) fully supervised logo classes, (b) 1-shot icon supervised logo classes, and (c) synthetic logo images from QMUL-OpenLogo.

level bounding box on the training data per logo class. They are therefore not scalable because it is time-consuming for collecting such training data annotation particularly considering the existence of many logo classes in real-world applications.

To scale up the learning algorithms, Su et al. (2017a) propose to leverage the rich web information from the online Internet multimedia data streams that contain weak but noisy label information. Whilst being highly noisy, this method can easily acquire a massive number of in-the-wild images without manual labelling efforts therefore scalable and facilitating the training of deep neural network models. One weakness of using web data is the low quality of supervision with severe class imbalance which dramatically increases the model learning difficulty. Elaborative logo image synthesis pipeline is also proposed (Montserrat et al., 2018) by depth estimation on background image for generating more realistic synthetic data. However, this method is computationally expensive, reducing its usability in large scale learning scenarios. How to use the already labelled logo detection data in a scalable manner seems a promising approach. To this end, an open logo detection setting is introduced (Su et al., 2018) where only a proportion of logo classes are associated with labelled training images. The goal is to learn a detection model that can be deploy to detect 1-shot icon supervised logo classes.

Image synthesising is an effective approach to solving scarce logo training data. Eggert et al. (2015) applied synthetic data to train SVM models for company logo detection. Gupta et al. (2016) and Jaderberg et al. (2016) generated scene-text images for learning text recognition models, a problem very similar to logo detection. Montserrat et al. (2017) employed synthetic images with both brand logo and toy classes. Letessier et al. (2012) created a synthetic dataset FlickrBelgaLogos by pasting logo instances to background web images. Generative Adversarial Networks were also used to generate clean logo images Sage et al. (2017), which however are not suitable for logo detection on in-the-wild images with complex background. The recent work CAL (Su et al., 2018) attempts to address this problem by synthesising context coherent training images for 1-shot icon supervised logo classes. However, it is extremely challenging to achieve this due to the difficulty of simulating the genuine logo instance contexts in real-world scenes.

We tackle this same challenge from a different modelling perspective – *cross-class domain adaptation*. Beyond using the synthetic training images, we further address the feature distribution discrepancy between fully supervised and 1-shot icon supervised classes for better optimising the model detection capability. We additionally leverage less relevant auxiliary object images for reducing the model overfitting risk towards fully labelled logo classes.

Zero-Shot Object Detection. Open logo detection is conceptually similar to the notion of zero-shot object detection where no labelled training data are available for test classes. It is extended from the zero-shot classification problem with the aim of enabling the machines to detect objects visually unseen before (Xian et al., 2017). These methods often rely on the semantic relationships between seen and unseen classes via manually labelling mid-level attributes (Lampert et al., 2009) and/or learning text vector embeddings from large scale corpus (Mikolov et al., 2013). However, such side information is hard and time-consuming to obtain and currently unavailable for open logo detection, which renders all the corresponding methods inapplicable. Besides, existing methods are not designed to work with clean icon images as in this context.

Unsupervised Domain Adaptation. In the literature, most unsupervised domain adaptation methods are focused on the classification problem (Wang and Deng, 2018; Sun and Saenko, 2016; Tzeng et al., 2017). More recently, this has been studied for object detection in various settings by several works (Hattori et al., 2015; Xu et al., 2014; Chen et al., 2018). In particular, Chen et al. (2018) modify the state-of-the-art Faster R-CNN model with domain adaptation layers for transferring knowledge between synthetic and real image domains. However, this study is limited to the closed-class setting where both domain share the same classes. All other existing methods make the same closed-set class assumption. In contrast, we investigate a

more challenging and practical problem of cross-class logo detection adaptation. Specifically, we do not assume the availability of real training data for the target logo classes. This avoids the expense of collecting labelled training data which is often costly or even unavailable in many cases. As such, a feasible solution is to leverage synthetic training data. This leads to a further need for domain adaptation from synthetic data to real data for the unsupervised target logo classes. Besides, we consider multi-perspective domain adaptation by concurrently exploiting both synthetic and less-relevant auxiliary imagery data for further improving the cross-class model generalisation capability in end-to-end model optimisation.

3. Method

Problem definition. In open logo detection, we have access to a training set $\mathcal{L}$ of fully supervised logo classes and a training set $\mathcal{U}$ of 1-shot icon supervised logo classes. For fully supervised logo classes, the training data contain both category and bounding box labels; Therefore, state-of-the-art object detection methods (Ren et al., 2015; Redmon and Farhadi, 2017; Lin et al., 2017) can be applied to train their detector models. For each 1-shot icon supervised logo class, however, only *a single exemplar icon* image (Fig 2) is available. An exemplar icon image is necessary to specify how a target logo class appears visually. The underlying reason is due to the man-made nature – logo class names cannot reflect the corresponding visual appearance in most cases.

In the standard object detection perspective, such exemplar icon images are not appropriate training data. They come without any scene context and bounding box annotations. The ***objective*** of open logo detection is to learn a logo model discriminative for 1-shot icon supervised logo classes, using both fully labelled training data $\mathcal{L}$ and 1-shot icon data $\mathcal{U}$.

Limitation of existing approach. An intuitive and effective approach for open logo detection is to leverage synthetic training data of 1-shot icon supervised logo classes (Su et al., 2018). Synthetic samples are usually generated by placing a clean logo icon at random positions in background scene images, and importantly the logo bounding box and class label supervision can be freely obtained. An attractive merit is that, a potentially infinite number of synthetic images can be produced. This previous method, however, suffers from a domain shift problem – synthetic training images differ from realistic testing imagery in distribution. It leads to significant degradation in model performance (Pan and Yang, 2010). Solving this training-testing domain shift problem is critical, particularly for improving the model performance on 1-shot icon supervised logo classes.

3.1. Multi-Perspective Cross-Class Alignment

To address the aforementioned problem, we introduce a ***Multi-Perspective Cross-Class*** (MPCC) alignment method. The high-level idea is to transfer the knowledge of fully supervised logo classes to 1-shot icon supervised logo classes. Designed as a generic plug module, it can be integrated into existing object detection networks.

Training data. Three types of training data are considered in MPCC: (1) Genuine training data of fully supervised logo classes including both scene images and bounding box annotations. They can be used to train a conventional logo object detection model. (2) Synthetic logo training images for 1-shot icon supervised logo classes, due to the lacking of standard training data. The labels of logo instances can be obtained during synthesis and used for model supervised training. (3) Non-logo object detection training images to augment the appearance distribution of genuine object instances.

It is non-trivial to train an effective model using such heterogeneous training data with different distributions. MPCC solves this problem from a domain adaption perspective, with an overview depicted in Fig 3.

3.1.1. Model Architecture

Overall, we adopt the two-stage model design for logo detection (Ren et al., 2015). Taking as input a specific training image, we compute feature maps, predict region proposals, extract feature vectors, and perform classification and box regression. This method assumes that all the training data are drawn from the same distribution as the test data, which however is not the case in open logo detection. In particular, this assumption does not hold for 1-shot icon supervised logo classes. We introduce two feature alignment components to mitigate this problem.

3.1.2. Alignment between Genuine and Synthetic Images

We propose cross-class distribution alignment between fully supervised and 1-shot icon supervised logo classes. This aims to address the conventional model learning bias towards the distribution of synthetic training data of 1-shot icon supervised logo classes. We consider this problem as a domain adaption problem. Concretely, we regard genuine fully supervised classes and synthetic 1-shot icon supervised classes as two distinctive domains.

In design, we exploit the idea of adversarial gradient learning (Ganin and Lempitsky, 2015) due to its simplicity and good efficacy. We introduce a domain classifier that aligns the feature distributions of genuine and synthetic logo object instances. It can be implemented with a fully connected layer.

We create a genuine-synthetic domain alignment problem as follows. We start by assigning the genuine logo instances of fully supervised classes with domain label "1", and the synthetic ones of 1-shot supervised logo classes with label "0". We then want to train such the model that yields a feature representation space in which an optimal domain classifier cannot distinguish between genuine and synthetic instances. This process is conducted in every mini-batch training. In doing so, aligning the distributions of the two types of logo objects can be well achieved. Consequently, the trained detection model is supposed to have minimal bias towards synthetic logo objects and become more generalisable when applied to the genuine images of 1-shot icon supervised classes.

Loss design. We adopt the softmax based cross-entropy loss function in training. Formally, given an object instance $\boldsymbol{x}_i$, we start by predicting the domain class posterior probability p_i^{genu}

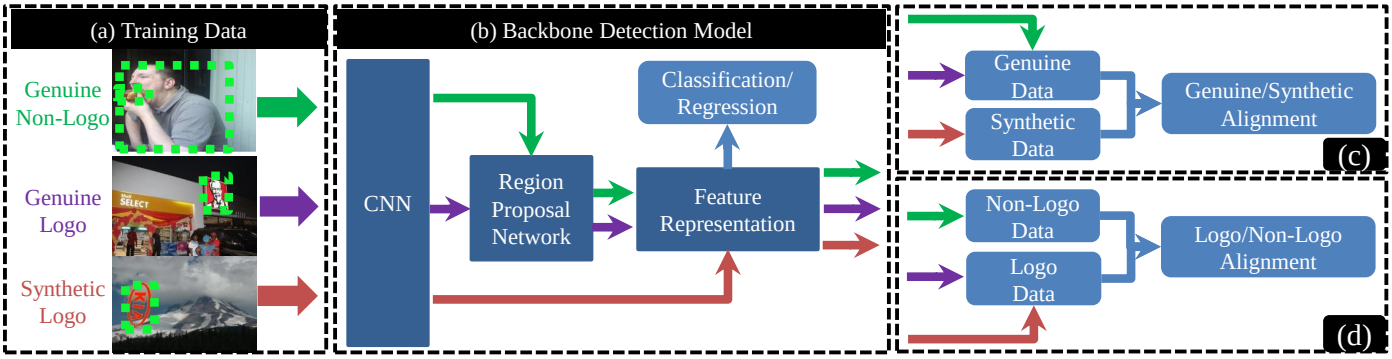

Fig. 3: Overview of the proposed *Multi-Perspective Cross-Class* (MPCC) domain alignment method. MPCC takes as input (a) genuine logo and non-logo object scene images, as well as synthetic logo images of 1-shot icon supervised classes. (b) The baseline detection model (e.g. Faster R-CNN) then extracts feature maps, detects region proposals, and compute feature representations for each proposal. Along with the conventional classification and bounding box regression loss, two domain alignment loss functions are further introduced: (c) one for aligning genuine and synthetic object instances, (d) and the other for aligning logo and non-logo object instances.

on the ground-truth domain class label $y_i \in \{0, 1\}$ using the softmax function as:

$$p_i^{\text{genu}} = \frac{\exp(\boldsymbol{w}_{y_i}^\top \boldsymbol{x}_i)}{\sum_{k\in\{0,1\}} \exp(\boldsymbol{w}_k^\top \boldsymbol{x}_i)} \tag{1}$$

where $\boldsymbol{w}_k$ specifies the classifier parameters of domain class $k \in \{0, 1\}$. The cross-entropy loss for a mini-batch of n_b training objects is then defined as:

$$\mathcal{L}_{\text{ad}}^{\text{genu}} = -\sum_{i=1}^{n_b} \log(p_i^{\text{genu}}) \tag{2}$$

Interestingly, we still minimise the $\mathcal{L}_{\text{ad}}^{\text{genu}}$ loss as in standard training. To achieve the effect that the detection model cannot distinguish genuine objects from synthetic ones, we insert a gradient reversal layer before the genuine-synthetic domain classifier.

3.1.3. Alignment between Logo and Non-Logo Images

As the size of fully supervised logo images is limited due to high labelling cost, we propose to leverage auxiliary non-logo object imagery to further enrich the distribution of genuine instances and improve the effectiveness of feature alignment. However, this may introduce some negative distracting effect due to the intrinsic difference between logo and non-logo objects in appearance. To achieve a consistent solution, we consider again this problem from domain adaptation perspective.

Together with genuine-synthetic domain setup, we further introduce logo-nonlogo domain alignment. Same as genuine-synthetic domain alignment, we employ the adversarial gradient concept. Differently, in this alignment we design the domain label based on if an object instance belongs to logo or not. Specifically, we assign logo instances by domain label "1" and non-logo instance by "0". We then leverage these domain labels to align the feature distribution across logo and non-logo instances. Conceptually, this scheme can be understood as a soft regularisation constraint that encourages the model to selectively learn information particularly useful for logo object detection in an implicit manner.

Loss design. The same softmax based cross-entropy loss function is used as above. We first estimate the domain class probability p_i^{logo} of an object instance $\boldsymbol{x}_i$ on the ground-truth domain label $y_i' \in \{0, 1\}$ as:

$$p_i^{\text{logo}} = \frac{\exp(\bar{\boldsymbol{w}}_{y_i'}^\top \boldsymbol{x}_i)}{\sum_{k\in\{0,1\}} \exp(\bar{\boldsymbol{w}}_k^\top \boldsymbol{x}_i)} \tag{3}$$

where $\bar{\boldsymbol{w}}_k$ denotes the classifier parameters of domain class $k \in \{0, 1\}$. The cross-entropy loss is then computed as:

$$\mathcal{L}_{\text{ad}}^{\text{logo}} = -\sum_{i=1}^{n_b} \log(p_i^{\text{logo}}) \tag{4}$$

where n_b is the batch-size. To realise adversarial gradient learning, we similarly deploy a gradient reversal layer before classification.

3.1.4. Objective Loss Function

Combining the two alignment constraints with the conventional object detection loss $\mathcal{L}_{\text{det}}$, we obtain the MPCC objective loss function as:

$$\mathcal{L}_{\text{mpcc}} = \mathcal{L}_{\text{det}} + \lambda_1 \mathcal{L}_{\text{ad}}^{\text{logo}} + \lambda_2 \mathcal{L}_{\text{ad}}^{\text{genu}} \tag{5}$$

where the hyper-parameters λ_1 and λ_2 control the relative importance ratio of the two adaptation loss terms. Note that, $\mathcal{L}_{\text{det}}$ typically consists of a classification loss and a regression loss. As a model-agnostic design, Eq. (5) can be integrated in existing detection models to boost open logo detection performance.

3.2. Model Implementation

In implementing our MPCC method, we adopt a ResNet-101 based Faster R-CNN (Ren et al., 2015) as the base detection model. The final objective function is an additive aggregation of Faster R-CNN detection loss and our MPCC loss (Eq. (5)). The model can be trained end-to-end by stochastic gradient descent. Other alternative models (Redmon and Farhadi, 2017; Lin et al., 2017) can be similarly considered.

We used the same method as (Su et al., 2018) to synthesise training images for 1-shot icon supervised logo classes (see examples in Fig. 2). To better generalise the detection model to 1-shot icon supervised logo classes, we formulate the objective of RPN as binary (logo and non-logo) classification. This is

Table 2: Open logo detection setting and data statistics.

Split	Classes	Train images	Val images	Test images
Fully supervised logo	176	10,586	1,561	3,121
1-shot icon supervised logo	176	0	0	3,649

inspired by the idea of universal logo detection (Tzk. et al., 2018; Fehérvári and Appalaraju, 2019) – class agnostic localization models can learn the generic characteristics of logo objects more strongly. Note that, we only use the manually labelled bounding boxes to minimise the loss of region proposal net. The intuition is that, synthetic instances are with *unrealistic* background context which may mislead model optimisation. This is verified in our evaluation (see Table 7).

3.3. Computational Complexity Analysis

We analyse the computational complexity of MPCC on top of the baseline method CAL (Su et al., 2018). As a model training strategy, MPCC does not increase the inference cost for logo detection on test images. It maintains the same inference cost as the base detection model. MPCC does increase the cost of model training by 2.3 times, due to more training data used. However, this should not be a big limitation, since training takes place only once.

4. Experiments

Dataset and setting. To evaluate the proposed MPCC model, we utilised the public QMUL-OpenLogo[1] detection dataset (Su et al., 2018). It contains a total of 27,083 images from 352 logo classes, established by combining and refining seven previous logo datasets. To facilitate model training, we adopted the second benchmark setting where 176 logo classes are fully supervised and the remaining 176 are 1-shot icon supervised. Fig. 2 shows example scene images and logo icons. It is noted that, in open logo detection, we focus more on the performance evaluation of 1-shot icon supervised logo classes. The statistics for train/val/test image sets are summarised in Table 2.

Performance metrics. For the performance evaluation of logo detection models, we used the common Average Precision (AP) for individual logo classes, and the mean Average Precision (mAP) over all classes (Everingham et al., 2010). We considered a logo detection as being correct if the Intersection over Union (IoU) between the detected and ground-truth boxes exceeds 50%.

Implementation details. For model optimisation, we adopted the Adam solver (Kingma and Ba, 2014). We set the learning rate of 0.0002, the batch size of 2, the max epoch number of 5. Following (Su et al., 2018), we generated 100 synthetic images for each of 352 logo classes, resulting in 35,200 synthetic logo images. Three types of training data were involved: 10,586 genuine logo images from QMUL-OpenLogo, 35,200 synthetic logo images by synthetic data generation with background images from FlickrLogo-32 (Romberg et al., 2011), and 82,081 non-logo data from COCO 2014 benchmark (Lin et al., 2014). The size ratio corresponds to 1.0:3.3:7.8. We also tested different proportion configurations to verify their effect (Table 8). The model hyper-parameters were setting as $\lambda_1 = 0.1$, and $\lambda_2 = 0.1$ for Eq. (5) by cross-validation on the validation set. Concretely, we first cross-validated λ_1 and λ_2 on the validation set; Once the hyper-parameters were estimated, we merged the validation and training sets to train the final model.

4.1. Comparisons to the State-of-the-Art Methods

Competitors. We compared the MPCC with two synthetic data generation methods SCL (Su et al., 2017b) and CAL (Su et al., 2018) in conjunction with two strong object detection models YOLOv2 (Redmon and Farhadi, 2017) and Faster R-CNN (Ren et al., 2015). Moreover, we further compared to a feature manipulation method (Sage et al., 2017) based on Faster R-CNN. This method quantifies the latent visual attribute discrepancy between genuine and synthetic logo object instances for improving the representation quality of synthetic logo object instances by algebraic addition and subtraction vector operations. The key idea is to represent the logo instances by latent attributes that can be manipulated such that the corresponding instances are accordingly transformed. Specifically, a general logo detector was first trained with both genuine and synthetic logo data to learn logo localisation. Second, a multi-label classifier is trained to classify both the logo classes and genuine/synthetic labels of the logo instances, thus the genuine/synthetic feature boundary was modelled. Third, the genuine and synthetic data of the supervised logo classes were fed into the model to extract their latent features which are used to obtain their mean difference. In the evaluation stage, this genuine-synthetic instance difference was transferred to the unsupervised logo classes to bridge the gap of synthetic training data and genuine test data. This model can be considered as a feature domain alignment strategy in contrast to the adversarial learned MPCC method for imagery pixel alignment. All these competitors were trained on the same training data (if possible by design) for a fair comparison.

Quantitative evaluation. From Table 3, we conclude that:

1. CAL (Su et al., 2018) is superior to SCL (Su et al., 2017b) by generating context more coherent synthetic images. This superiority is consistent over two detectors. Both methods aim to address the cross-class detection challenge by training data synthesis. Despite random context sampling and rendering, the domain mismatch between the genuine and synthetic images still remain at large.
2. YOLOv2 (Redmon and Farhadi, 2017) is shown as a weaker architecture than Faster R-CNN (Ren et al., 2015)

[1]QMUL-OpenLogo: `https://qmul-openlogo.github.io/`

Table 3: Open logo detection performance by MPCC and state-of-the-art methods. Metric: mAP.

Classes	1-shot supervised	Fully supervised	All
YOLOv2 (Redmon and Farhadi, 2017) + SCL (Su et al., 2017b)	12.75%	47.36%	30.06%
YOLOv2 (Redmon and Farhadi, 2017) + CAL (Su et al., 2018)	13.72%	46.60%	30.16%
Faster R-CNN (Ren et al., 2015) + SCL (Su et al., 2017b)	18.63%	49.16%	33.89%
Faster R-CNN (Ren et al., 2015) + CAL (Su et al., 2018)	20.31%	48.19%	34.25%
Faster R-CNN + SCL + Feature Manipulation (Sage et al., 2017)	19.95%	46.90%	33.43%
Faster R-CNN + CAL + Feature Manipulation (Sage et al., 2017)	21.09%	46.72%	33.91%
Faster R-CNN + SCL + **MPCC (Ours)**	23.40%	48.60%	36.00%
Faster R-CNN+ CAL + **MPCC (Ours)**	**24.53%**	**49.41%**	**36.97%**

for open logo detection. The potential reason is that logo object instances vary significantly in size, which makes the region proposal estimation more necessary.

3. Feature manipulation (Sage et al., 2017) further improves the performance, e.g. a mAP gain of 1.32% (19.95-18.63) with SCL and 0.78% (21.09-20.31) with CAL. This suggests the efficacy of such feature level alignment between synthetic and genuine data.
4. MPCC achieves the best performance, indicating the overall result superiority of our method thanks to the principled domain adaptation between classes for supervision knowledge transfer from both labelled logo classes and generic non-logo objects in realistic context. This also reduces the necessity of rendering logo context as implied by the smaller difference between using SCL and CAL in MPCC.

Qualitative evaluation. To visually assess the model performance, we compared MPCC (w/ CAL) with the best alternative model Feature Manipulation (w/ CAL) (Sage et al., 2017) in Fig 4. We observed similar performance comparisons as the numerical evaluation above. Moreover, we also compared the feature distributions of genuine and synthetic logo images from the *MasterCard* class using two models trained with and without the MPCC. Fig 5 shows that MPCC can bring about a more immersed single overlapping region from the distributions of the genuine and synthetic logo data whilst "without MPCC" their distributions are in two more separable regions. This demonstrates that "with MPCC" the synthetic data are more effective for model training.

4.2. Model Component Analysis

Genuine and synthetic domain adaptation. We examined the effect of domain adaptation between genuine and synthetic training images. Table 4 shows that it brings clear mAP gain to the models. This suggests the significance of aligning the synthetic towards the genuine logo training data, which is mainly caused by unrealistic image synthesis in terms of both logo instance appearance and background context.

Table 4: Effect of genuine and synthetic domain adaptation.

$\mathcal{L}_{\text{ad}}^{\text{genu}}$	mAP
×	21.47%(SCL) / 23.24%(CAL)
✓	**23.40%(SCL) / 24.53%(CAL)**

Logo and non-logo domain adaptation. We tested the benefits of using non-logo object detection images (COCO) in a domain adaptation manner. Table 5 shows a positive impact of this component in model performance. This validates our design consideration of transferring generic object instance supervision for enlarging the training data and reducing the model overfit inclination towards the 1-shot icon supervised logo classes.

Table 5: Effect of logo and non-logo domain adaptation.

Non-logo data	$\mathcal{L}_{\text{ad}}^{\text{logo}}$	mAP
×	×	21.62%(SCL) / 21.82%(CAL)
✓	×	21.01%(SCL) / 22.00%(CAL)
✓	✓	**23.40%(SCL) / 24.53%(CAL)**

Non-logo object image source. We evaluated the impact of non-logo object image with two sources: 9,963 PASCAL VOC 2007 images (Everingham et al., 2015) *vs.* 82,081 MS COCO 2014 images (Lin et al., 2014). Table 6 suggests that COCO serves as a better data source as expected. The plausible reason is the availability of more labelled genuine object detection images with richer contexts.

Table 6: Effect of non-logo object image source.

Non-logo image source	mAP
PASCAL VOC 2007	22.02%(SCL) / 23.15%(CAL)
MS COCO 2014	**23.40%(SCL) / 24.53%(CAL)**

Synthetic box supervision. Recall that we deliberately ignore the synthetic bounding box supervision of 1-shot icon supervised logo classes in training the RPN function. We tested this design. Table 7 shows that using synthetic bounding box labels leads to a small model performance decrease. A plausible reason is due to less realistic context within synthetic logo instances.

Table 7: Effect of synthetic box supervision (SBS).

SBS	mAP
✓	23.11%(SCL) / 24.22%(CAL)
×	**23.40%(SCL) / 24.53%(CAL)**

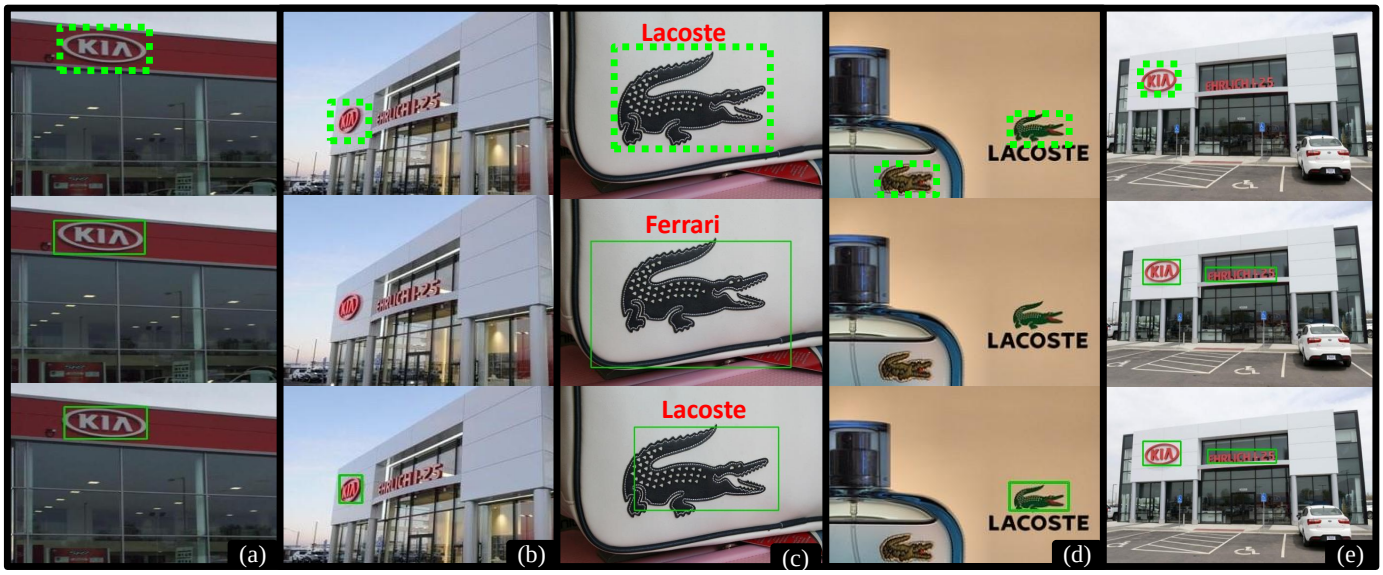

Fig. 4: Five qualitative logo detection examples by Feature Manipulation (FM) (Sage et al., 2017) (2nd row) and MPCC (3rd row) along with the ground-truth (1st row). (a) Both models correctly detect the logo instance; In (b) FM misses the target; In (c) FM produces a miss classification, whist MPCC succeeds; (d) FM fails to identify two "Lacoste" logo instances, while MPCC misses the hard instance with significantly varied appearance on the bottom left; (e) Both models find the correct logo instance whilst making a false positive detection.

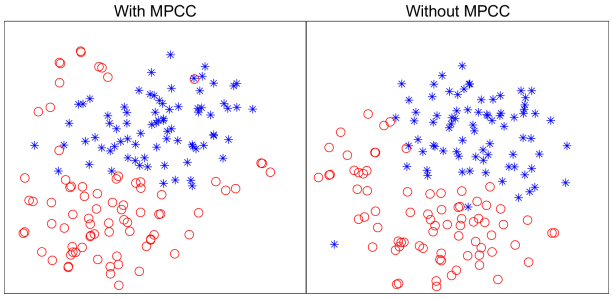

Fig. 5: A visualisation of the t-SNE feature distributions of genuine (red circles) and synthetic (blue stars) logo images from the 'MasterCard' class with (*left*) and without (*right*) the proposed MPCC method. It is evident that MPCC can bring about a more immersed single overlapping region from the distributions of the genuine and synthetic logo data whilst "without MPCC" their distributions are in two more separable regions. This demonstrates that "with MPCC" the synthetic data are more effective for model training.

Training data configuration. Recall that we used three different training sets, including genuine logo images, synthetic logo images and non-logo images, to train our model, with a proportion of 1.0:3.3:7.8. To evaluate the effect of different data combinations, we further tested three more proportional configurations by halving one of the three training sets, individually. Table 8 shows that reducing the amount of any type of training data would negatively affect the model performance. In particular, genuine data and non-logo data are most and least important, respectively.

5. Conclusion

We presented a *Multi-Perspective Cross-Class* (MPCC) domain adaptation method for overcoming the domain shift problem of open logo detection so that synthetic training images of 1-shot icon supervised logo classes can be more discriminatively leveraged. This method scales up existing logo detection models that rely on conventional supervised learning due to no need for large labelled training data per class. Compared to previous alternative methods, it solves the largely ignored domain mismatch problem between synthetic and genuine logo images. MPCC also leverages large auxiliary non-logo object detection images for further improving the model generalisation capability on 1-shot icon supervised logo classes. Empirical evaluations show the performance advantages of our MPCC method over the state-of-the-art competing methods on the standard QMUL-OpenLogo benchmark. We provided component analyses to give insights on the design considerations of our model.

Table 8: Effect of training data configuration.

Data configuration	mAP
Default	**23.40%(SCL) / 24.53%(CAL)**
50% synthetic data	22.34%(SCL)/ 22.50%(CAL)
50% non-logo data	22.40%(SCL)/ 23.04%(CAL)
50% genuine logo data	20.71%(SCL) / 21.36%(CAL)

Acknowledgements

This work was partially supported by the China Scholarship Council, Vision Semantics Limited, the Royal Society Newton Advanced Fellowship Programme (NA150459), Innovate UK Industrial Challenge Project on Developing and Commercialising Intelligent Video Analytics Solutions for Public Safety (98111-571149), and the Alan Turing Institute Fellowship Project on Deep Learning for Large-Scale Video Semantic Search.

References

Bianco, S., Buzzelli, M., Mazzini, D., Schettini, R., 2017. Deep learning for logo recognition. Neurocomputing 245, 23–30.

Boia, R., Bandrabur, A., Florea, C., 2014. Local description using multi-scale complete rank transform for improved logo recognition, in: IEEE International Conference on Communications, pp. 1–4.

Chen, Y., Li, W., Sakaridis, C., Dai, D., Van Gool, L., 2018. Domain adaptive faster r-cnn for object detection in the wild, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3339–3348.

Doermann, D.S., Rivlin, E., Weiss, I., 1993. Logo recognition using geometric invariants, in: Proceedings of 2nd International Conference on Document Analysis and Recognition, IEEE. pp. 894–897.

Eggert, C., Winschel, A., Lienhart, R., 2015. On the benefit of synthetic data for company logo detection, in: Proceedings of the 23rd ACM international conference on Multimedia, pp. 1283–1286.

Everingham, M., Eslami, S.A., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A., 2015. The pascal visual object classes challenge: A retrospective. International Journal of Computer Vision 111, 98–136.

Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A., 2010. The pascal visual object classes (voc) challenge. International journal of computer vision 88, 303–338.

Fehérvári, I., Appalaraju, S., 2019. Scalable logo recognition using proxies, in: IEEE Winter Conference on Applications of Computer Vision, IEEE. pp. 715–725.

Ganin, Y., Lempitsky, V., 2015. Unsupervised domain adaptation by backpropagation, in: International Conference on Machine learning, pp. 1180–1189.

Girshick, R., 2015. Fast r-cnn, in: Proceedings of the IEEE international conference on computer vision, pp. 1440–1448.

Gupta, A., Vedaldi, A., Zisserman, A., 2016. Synthetic data for text localisation in natural images, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2315–2324.

Hattori, H., Naresh Boddeti, V., Kitani, K.M., Kanade, T., 2015. Learning scene-specific pedestrian detectors without real data, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3819–3827.

Hoi, S.C., Wu, X., Liu, H., Wu, Y., Wang, H., Xue, H., Wu, Q., 2015. Logo-net: Large-scale deep logo detection and brand recognition with deep region-based convolutional networks. arXiv preprint arXiv:1511.02462 .

Iandola, F.N., Shen, A., Gao, P., Keutzer, K., 2015. Deeplogo: Hitting logo recognition with the deep neural network hammer. arXiv preprint arXiv:1510.02131 .

Jaderberg, M., Simonyan, K., Vedaldi, A., Zisserman, A., 2016. Reading text in the wild with convolutional neural networks. International Journal of Computer Vision 116, 1–20.

Joly, A., Buisson, O., 2009. Logo retrieval with a contrario visual query expansion, in: ACM International Conference on Multimedia, pp. 581–584.

Kalantidis, Y., Pueyo, L.G., Trevisiol, M., van Zwol, R., Avrithis, Y., 2011. Scalable triangulation-based logo recognition, in: ACM International Conference on Multimedia Retrieval, p. 20.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. International Conference on Learning Representations , 13.

Lampert, C.H., Nickisch, H., Harmeling, S., 2009. Learning to detect unseen object classes by between-class attribute transfer, in: Proceedings of the IEEE conference on computer vision and pattern recognition, IEEE. pp. 951–958.

Letessier, P., Buisson, O., Joly, A., 2012. Scalable mining of small visual objects, in: Proceedings of the 20th ACM international conference on Multimedia, ACM. pp. 599–608.

Li, K.W., Chen, S.Y., Su, S., Duh, D.J., Zhang, H., Li, S., 2014. Logo detection with extendibility and discrimination. Multimedia tools and applications 72, 1285–1310.

Liao, Y., Lu, X., Zhang, C., Wang, Y., Tang, Z., 2017. Mutual enhancement for detection of multiple logos in sports videos, in: IEEE International Conference on Computer Vision, pp. 4856–4865.

Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017. Focal loss for dense object detection, in: Proceedings of the IEEE international conference on computer vision, pp. 2980–2988.

Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft coco: Common objects in context, in: European Conference on Computer Vision, Springer. pp. 740–755.

Mikolov, T., Chen, K., Corrado, G., Dean, J., 2013. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 .

Montserrat, D.M., Lin, Q., Allebach, J., Delp, E.J., 2017. Training object detection and recognition cnn models using data augmentation. Electronic Imaging 2017, 27–36.

Montserrat, D.M., Lin, Q., Allebach, J., Delp, E.J., 2018. Logo detection and recognition with synthetic images. Electronic Imaging 2018, 337–1.

Nanni, L., Ghidoni, S., Brahnam, S., 2017. Handcrafted vs. non-handcrafted features for computer vision classification. Pattern Recognition 71, 158–172.

Pan, C., Yan, Z., Xu, X., Sun, M., Shao, J., Wu, D., 2013. Vehicle logo recognition based on deep learning architecture in video surveillance for intelligent traffic system, in: IET International Conference on Smart and Sustainable City, pp. 123–126.

Pan, S.J., Yang, Q., 2010. A survey on transfer learning. IEEE Transactions on knowledge and data engineering 22, 1345–1359.

Pham, T.D., 2003. Unconstrained logo detection in document images. Pattern recognition 36, 3023–3025.

Psyllos, A.P., Anagnostopoulos, C.N.E., Kayafas, E., 2010. Vehicle logo recognition using a sift-based enhanced matching scheme. IEEE Transactions on Intelligent Transportation Systems 11, 322–328.

Redmon, J., Farhadi, A., 2017. Yolo9000: better, faster, stronger, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7263–7271.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks, in: Advances in Neural Information Processing Systems, pp. 91–99.

Revaud, J., Douze, M., Schmid, C., 2012. Correlation-based burstiness for logo retrieval, in: ACM International Conference on Multimedia, pp. 965–968.

Romberg, S., Lienhart, R., 2013. Bundle min-hashing for logo recognition, in: Proceedings of the 3rd ACM conference on International conference on multimedia retrieval, ACM. pp. 113–120.

Romberg, S., Pueyo, L.G., Lienhart, R., Van Zwol, R., 2011. Scalable logo recognition in real-world images, in: Proceedings of the 1st ACM International Conference on Multimedia Retrieval, ACM. p. 25.

Sage, A., Agustsson, E., Timofte, R., Van Gool, L., 2017. Logo synthesis and manipulation with clustered generative adversarial networks. arXiv preprint arXiv:1712.04407 .

Sahbi, H., Ballan, L., Serra, G., Del Bimbo, A., 2012. Context-dependent logo matching and recognition. IEEE Transactions on Image Processing 22, 1018–1031.

Su, H., Gong, S., Zhu, X., 2017a. Weblogo-2m: Scalable logo detection by deep learning from the web, in: Workshop of the IEEE International Conference on Computer Vision, pp. 270–279.

Su, H., Zhu, X., Gong, S., 2017b. Deep learning logo detection with data expansion by synthesising context, in: IEEE Winter Conference on Applications of Computer Vision, IEEE. pp. 530–539.

Su, H., Zhu, X., Gong, S., 2018. Open logo detection challenge, in: British Machine Vision Conference, p. 16.

Sun, B., Saenko, K., 2016. Deep coral: Correlation alignment for deep domain adaptation, in: European Conference on Computer Vision, Springer. pp. 443–450.

Tzeng, E., Hoffman, J., Saenko, K., Darrell, T., 2017. Adversarial discriminative domain adaptation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7167–7176.

Tzk., A., Herrmann., C., Manger., D., Beyerer., J., 2018. Open set logo detection and retrieval, in: Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, pp. 284–292.

Wang, M., Deng, W., 2018. Deep visual domain adaptation: A survey. Neurocomputing 312, 135–153.

Xian, Y., Lampert, C.H., Schiele, B., Akata, Z., 2017. Zero-shot learning the good, the bad and the ugly. Proceedings of the IEEE conference on computer vision and pattern recognition , 3077–3086.

Xu, J., Ramos, S., Vázquez, D., López, A.M., 2014. Domain adaptation of deformable part-based models. IEEE transactions on pattern analysis and machine intelligence 36, 2367–2380.