Open-Set Domain Adaptation through Self-Supervision

Daniele Rege Cambrin, Kylie Bedwell, Tommaso Natta, Ehsan Ansari Nejad
Politecnico di Torino
Corso Duca degli Abruzzi, 24
10129 Torino, ITALY
s290144@studenti.polito.it, s287581@studenti.polito.it
s282478@studenti.polito.it, s288903@studenti.polito.it

Abstract

- sentence describing the problem - sentence describing our proposed method - sentence summarising the results - sentence about the variations - sentence about the variation results

1. Introduction

In the computer vision research area large amounts of unlabeled data are available, however the cost of labeling this data is high [4, 16]. Domain adaptation is one technique that can be used to exploit the unlabeled data by first training a model on labeled data from a different but similar domain (the *source* domain), and then applying this model to the unlabeled data (the *target* domain). This technique assumes the distribution of both source and target domains are similar and describe the same class labels, also known as the *closed-set* scenario [1]. When applied to real-word scenarios however it is possible that the target domain includes previously unseen classes, known as the *open-set* scenario. These extra class labels in the target domain will cause performance degradation of the classification model and should be identified and isolated. The problem thus consists of two steps: first separating the target domain into known and unknown samples; then conducting domain alignment between the source domain and the known samples of the target domain.

Self-supervised learning can be used to separate the known class samples in the target domain from the unknown samples. Self-supervised learning involves the transformation of data using a known transform (for example by using image rotation), then training a model to predict the transformation [14]. When used in an object classification task and considering the image rotation transformation, the correct orientation of an object is domain-invariant. In this way the model can be trained to predict the correct orientation of the image using data from the source domain, then applied to the images of the target domain. If the orientation of a sample in the target domain is predicted correctly then it is considered to be of a *known* class. Contrarily if the orientation is not prediction correctly it is labelled as *unknown*.

Domain adaptation can then be performed between the samples in the source domain and the samples recognized as known in the target domain. When applying domain adaptation to an open-set scenario (Open-Set Domain Adapatation or OSDA) the samples classified as unknown can be treated as a separate class and incorporated into the Closed-Set Domain Adaptation (CSDA) task [2]. The self-supervised rotation task can also be used to reduce the domain shift during this step, using the Rotation-based Open Set (ROS) method developed by Bucci *et al.* [1].

This study investigates the use of a simplified ROS method for object classification on the *Office-Home* dataset [12]. Alternative self-supervised tasks as well as the inclusion of center-loss are also considered and their performance on the object classification task is evaluated.

2. Related Work

Anomaly detection, or outlier/novelty detection, in an open-set scenario can be used to detect samples belonging to the unknown or unseen class. Various different approaches for anomaly detection have been used in the literature as applied to the open-set scenario. Golan and El-Yanic [7] present a method for using geometric transformations to create a self-labeled dataset. In this way the neural classifier learns features that are effective for the detection of anomalies. Sakurada and Yairi [11] on the other hand make use of autoencoders with dimensionality reduction. This method assumes the data have correlations that can clearly separate normal and anomalous samples when reduced to a lower dimensional subspace. After the test data is projected into the subspace it is then reconstructed and the corresponding reconstruction error is used to identify the anomalous samples.

Self-supervised learning has been a key concept aimed at reducing the need for human labeling of data. It has also created opportunities for the use of data in problems where supervision is not possible [15]. Self-supervised learning consists of choosing a self-supervised task (or pretext task) to train alongside the main classification task. One possible self-supervised task is image rotation prediction, which is reported to perform best for visual representation learning [6, 14]. However many options are available, including image-patch based methods [8, 9], horizontal flipping [7], or by solving jigsaw puzzles [3, 8].

Domain adaptation techniques have advanced significantly in recent years for the closed-set scenario [2], however for real-world applications the closed-set assumption is often not applicable [10]. It has become increasingly important to develop robust techniques for open-set domain adaptation to address this problem. Recent studies in this field include: the development of a generic approach to learn a linear mapping between the features of the source domain and target domain [2]; the use of self-supervision to improve the generalization of models to different domains [3]; partial domain adaptation by using a discrepancy criterion to partially align features whilst avoiding negative transfer [10]. These techniques have been reported to perform well, and increase the applicability of domain adaptation methods to real-world applications [2, 3, 10].

Rotation-based Open Set (ROS) is a specific technique developed by Bucci *et al*. [1]. ROS is a two-stage method for open-set domain adaptation. The first stage separates samples in the target domain into known and unknown categories by training the model on a multi-rotation recognition task. The rotation recognition task includes the use of the center-loss to improve performance by learning a center of the features and minimizing the distances between features and their corresponding centers [13]. The second stage conducts domain alignment, training both semantic and rotation classifiers to classify known target samples. Bucci *et al*. [1] also propose the use of the harmonic mean of the average class accuracies for the known and unknown classes as a more robust and balanced evaluation metric.

3. Method

This study aims to solve the open-set scenario by using self-supervised learning with domain adaptation. Image rotation recognition has been chosen as the self-supervised task due to it's good performance for visual representation learning [6, 14]. The rotation recognition task involves taking the original image sample from the source domain and rotating it clockwise a set amount (for example by 0°, 90°, 180° or 270°). A rotation classifier is then trained to predict the correct orientation of the object. The correct orientation of the object, however, is not an inherent property of an image, for example consider the pens in Figure 1. When analysing a rotated image of a pen it is not possible to infer the original rotation. The original image is therefore included in the rotation classifier and the relative rotation analysed instead. In training the rotation classifier the features of the original sample are concatenated with the features of the rotated sample. By including the original samples the network is also able to learn features that are more discriminative between class labels, focussing more on the shape of the object and less on the texture [1], as shown in Figure 2.

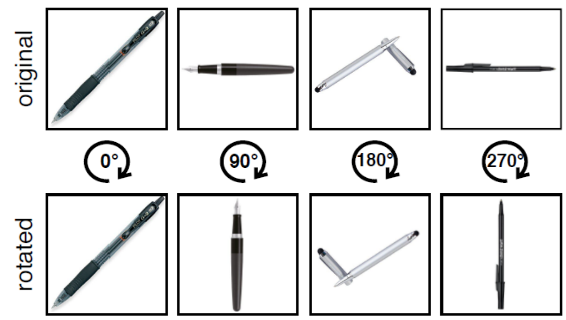

Figure 1. Relative orientations of pens with respect to the original images [1].

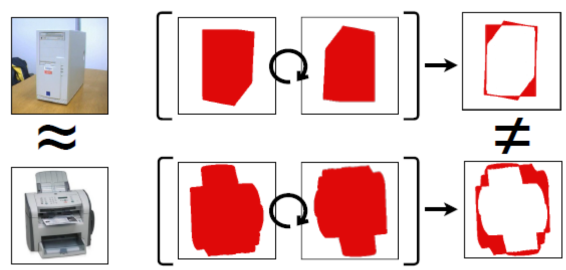

Figure 2. Image rotations help the network learn features that discriminate the shape of objects [1].

To solve the open-set domain adaptation problem a simplified version of the ROS method (Figure 3) is used. This version uses a single-head rotation classifier and does not include the center-loss, however the effect of the center-loss is evaluated subsequently as a variation to this method. Alternative self-supervised tasks are also considered, including horizontal flipping, vertical flipping and through solving jigsaw puzzles. These variations are analyzed in Section 5.

3.1. Stage I - known/unknown separation

Stage I of the simplified ROS method involves training both an object classifier and a rotation recognition task on data from the source domain, as shown on the left side of Figure 3, where D_s is the source domain dataset, E is the encoder, C_1 is the object classifier and R_1 is the rotation

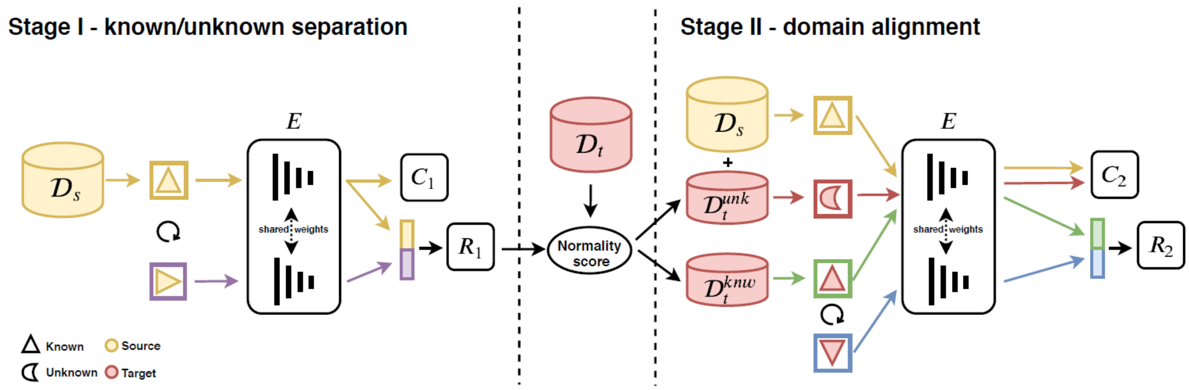

Figure 3. Rotation-based Open Set (ROS) method schematic illustration [1].

classifier. The object prediction is based on the features of the original source samples, whereas the rotation prediction is based on the concatenated features of the original and rotated samples. The object classification and rotation recognition tasks are trained simultaneously to minimize the total loss objective function given by:

$$L_{tot} = L_{cls} + \alpha_1 L_{rot}, \tag{1}$$

where L_{cls} is the loss from the object classifier, L_{rot} is the loss from the rotation classifier, and α_1 is the weight assigned to the rotation recognition task. The value of α_1 is tuned according to the performance of the network, as detailed in Section 4.1.

Target evaluation

Once the rotation classifier has been well trained it can be used to identify the known classes of the target domain, and separate these samples from the unknown classes. The method is illustrated in the center of Figure 3, where D_t is the target domain dataset, D_t^{unk} is the dataset of target samples identified as belonging to an unknown class, and D_t^{knw} is the dataset of target samples identified as belonging to a known class. The rotation classifier is applied to each of the target samples and generates a score for each of the possible rotations. The normality score is then computed as the rotation with the maximum score (highest prediction).

The precision of the normality score is then evaluated using the AUROC (Area Under Receiver Operating Characteristics curve) metric. The AUROC metric reduces analysis of the ROC curve to a single scalar value which can be used to compare the performance of classifiers [5]. The higher the AUROC score the better the classifier overall, with scores > 0.5 showing an improvement over random guessing.

If the AUROC is > 0.5 then the normality score can be used to conduct the known/unknown separation subject to a selected threshold, i.e. if the normality score is above the threshold the sample is considered as known and added to D_t^{knw}, otherwise it is considered as unknown and added to D_t^{unk}. The threshold value is also tuned according to the performance of the network, as detailed in Section 4.1.

3.2. Stage II - domain alignment

In stage II of the simplified ROS method the unknown dataset created from the target domain D_t^{unk} is combined with the samples from the source domain D_s, with the unknown samples representing an additional 'unknown' class. The object classifier C_2 is thus trained to recognise samples belonging to the unknown class. Concurrently, the known target samples D_t^{knw} are used for the source-target adaptation by training the rotation recognition task R_2. The method is shown on the right side of Figure 3. As in stage I, the object classification and rotation recognition tasks are trained simultaneously to minimize the total loss objective function given by:

$$L_{tot} = L_{cls} + \alpha_2 L_{rot}, \tag{2}$$

where α_2 is the weight assigned to the rotation recognition task of stage II. The value of α_2 is also selected to provide the best performance, detailed further in Section 4.1.

Final evaluation

Once the classifiers have been trained the final evaluation can be completed. Three metrics are considered for the evalutation: the accuracy of the object classifier in correctly classifying objects of the known category OS^*; the accuracy of the object classifier in correctly identifying objects of the unknown category UNK; and the harmonic mean HOS between the two accuracys OS^* and UNK, as defined in [1]. The harmonic mean is calculated as:

$$HOS = 2\,\frac{OS^* \times UNK}{OS^* + UNK}, \tag{3}$$

and provides a balanced measure of performance of the classifier at recognizing both known and unknown samples.

4. Experiments

- the network used

4.1. Ablation study

- ablation study: hyperparameter tuning of weights and threshold value, include graphs or tables of values

4.2. Results

5. Variations

- describe each of the variations - present results of the performance of the model and compare with the baseline for each of the variations

6. Conclusions

- summarise the results - add future recommendations

6.1. Acknowledgements

The authors would like to thank Silvia Bucci for her assistance and guidance in completing this study.

References

[1] Silvia Bucci, Mohammad Reza Loghmani, and Tatiana Tommasi. On the effectiveness of image rotation for open set domain adaptation. In *CVPR*, 2020. 1, 2, 3

[2] Pau Panareda Busto, Ahsan Iqbal, and Juergen Gall. Open set domain adaptation for image and action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2):413–429, 2020. 1, 2

[3] Fabio Maria Carlucci, Antonio D'Innocenteand Silvia Bucci, Barbara Caputo, and Tatiana Tommasi. Domain generalization by solving jigsaw puzzles. In *CVPR*, 2019. 2

[4] Gabriela Csurka. Domain adaptation for visual applications: A comprehensive survey. In *CVPR*, 2017. 1

[5] Tom Fawcett. An introduction to roc analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006. 3

[6] Spyros Gidaris, Praveer Singh, and Nikos Komodakis. Unsupervised representation learning by predicting image rotations. In *CVPR*, 2018. 2

[7] Izhak Golan and Ran El-Yaniv. Deep anomaly detection using geometric transformations. In *NeurIPS*, 2018. 1, 2

[8] Dahun Kim, Donghyeon Cho, Donggeun Yoo, and In So Kweon. Learning image representations by completing damaged jigsaw puzzles. In *WACV*, 2018. 2

[9] T. Nathan Mundhenk, Daniel Ho, and Barry Y. Chen. Improvements to context based self-supervised learning. In *CVPR*, 2018. 2

[10] Chuan-Xian Ren, Pengfei Ge, Peiyi Yang, and Shuicheng Yan. Learning target-domain-specific classifier for partial domain adaptation. *IEEE Transactions on Neural Networks and Learning Systems*, 32(5):1989–2001, 2021. 2

[11] Mayu Sakurada and Takehisa Yairi. Anomaly detection using autoencoders with nonlinear dimensionality reduction. In *Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis*, pages 4–11, 2014. 1

[12] Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. Deep hashing network for unsupervised domain adaptation. In *CVPR*, 2017. 1

[13] Yandong Wen, Kaipeng Zhang, Zhifeng Li1, and Yu Qiao. A discriminative feature learning approach for deep face recognition. In *ECCV*, 2016. 2

[14] Jiaolong Xu, Liang Xiao, and Antonio M. Lopez. Self-supervised domain adaptation for computer vision tasks. *IEEE*, 7:156694–156706, 2019. 1, 2

[15] Burhaneddin Yaman, Seyed Amir Hossein Hosseini, Steen Moeller, Jutta Ellermann, Kâmil Uğurbil, and Mehmet Akçakaya. Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data. *Magnetic Resonance in Medicine*, 84(6):3172–3191, 2020. 2

[16] Lei Zhang and David Zhang. Robust visual knowledge transfer via extreme learning machine-based domain adaptation. *IEEE Transactions on Image Processing*, 25(10):4959–4973, 2016. 1