Scholar Publishing Group
International Journal of Big Data Intelligent Technology
https://doi.org/10.38007/IJBDIT.2022.030401
ISSN 2790-0932 Vol. 3, Issue 4: 1-9

Research on Image Recognition for Small Sample

Mingyuan Xin* and Ang Ling Weay

Malaysia University of Science and Technology (MUST), Kuala Lumpur, Malaysia

xin.mingyuan@phd.must.edu.my

**corresponding author*

Keywords: Convolutional Neural Network, Transfer Learning, Image Classification, Target Detection

Abstract: This paper mainly analyzes the problems existing in image recognition technology, concentrating on the state of the research and current issues with picture identification in tiny sample space. This paper reviews the algorithms and frameworks proposed by scholars in small sample spatial image recognition technology in recent years and analyzes the existing research gaps.

1. Introduction

The typical regulation schemes can be exemplified as follows: A new wave of scientific and technological advancements as well as industrial transformation are being significantly fueled by artificial intelligence. and deep learning technology occupies an absolute dominant position in the field of artificial intelligence. Image recognition is the most basic and most important task in computer vision. It refers to the use of computer processing methods to analyze and understand the categories and attributes of images. The study on picture recognition in various common computer vision tasks, such as target detection, semantic segmentation, and image classification, is summarized in this paper.so as to facilitate more researchers to understand this field. The research idea of this paper is shown in Figure 1.

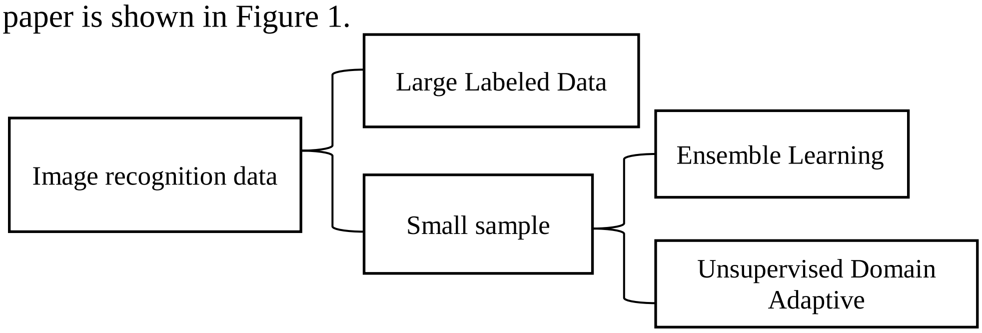

Figure 1. Research route

1.1. Deep Learning Techniques Rely on Large Amounts of Labeled Data

To construct a data set that can cover the complete sample distribution of application scenarios requires huge economic and human costs to collect data and annotate data. For example, ImageNet is the most popular benchmark data set for image classification in the machine learning community, containing over 14 million labeled images, and has become a touchstone for evaluating the ability of computer vision models in a wide range of visual tasks (Yuan et al., 2019). More than 25,000 people were hired to complete annotation for about 300 days (Yuan et al., 2020).

1.2. The Application Range of Small Sample Labeled Data is Larger

Deep neural network has strong fitting and characterization capabilities, and the extracted features are more generalized than those extracted manually (Muhammad Ghifary, W. Bastiaan Kleijn, Mengjie Zhang, David Balduzzi, 2016) However, the complexity of computer vision tasks makes it difficult for the deep learning model trained on specific data sets to be well transferred to new scenes with different data distributions (Chiyuan Zhang et al., 2016; Chongzhen Zhang et al., 2020). Once the model is applied to other data sets with different data distribution, its performance may decrease significantly. Acquiring a huge number of unlabeled data is also simpler than obtaining a large amount of annotated data. The training and learning of an image recognition model can be completed with a tiny quantity of annotation data.Image recognition model oriented to small samples, in order to reduce the dependence of image recognition model learning on labeled samples. Thereby lowering the cost of manual annotation and costs, make full use of existing similar tasks at the same time a large number of labeled data and relatively easy to get a lot of no annotation data mining the data contained in the rich semantic and structural information to help get a better image recognition model, for deep learning for image recognition is more significance and value.

2. Literature Review

2.1. Ensemble Learning

The model variance can be decreased effectively by ensemble learning. can raise either one of the neural network training and testing costs or both of them. Snapshot Ensembling is one of the most successful ensemble methods for neural networks currently available. It uses the cyclic cosine annealing method to train neural networks to obtain multiple local minima of the loss function, and then uses the network entities corresponding to these local minima as the ensemble members. InterBoost, introduced by Xiaoxu Li, reweighted the initial data by two sets of complimentary weights initially [1]. An integrated circuit is obtained which can gradually improve the accuracy of the underlying network while encouraging the differences between the underlying networks. Each base network becomes more accurate as a result of the interaction between them, and the diversity between them is maintained as much as feasible. Two precise and diversified base networks are ultimately obtained.

2.2. An Unsupervised Domain Adaptive Method for Minimizing Divergence

2.2.1. Image Classification Feature Space

Early unsupervised domain adaptation approaches for image classification problems collect features by minimizing the divergence alignment between statistics of the source domain and the target domain. Rozantsev and otherspropósed a dual-flow network, in which the network weights of the source domain and the target domain were not shared, and then the network weights of the two domains were correlated through a loss function(Rozantsev et al., 2019). CORAL is another commonly used method to measure divergence between domains, which was first proposed (Baochen & Kate, 2016). The method uses linear transformation to align the distribution and second-order statistical information (covariance) of eigenvalues in the feature space of source and target domains to achieve unsupervised domain adaptation. CORAL requires feature extraction, linear transformation, and training of a SVM classifier separately, so CORAL is not an end-to-end learning. In order to be able to train end-to-end.

In recent years, other methods to calculate the divergence of feature distribution between different domains have been gradually proposed. However, Damodaran et al. only carried out cross-domain matching and ignored the intra-domain structure (Damodaran et al., 2018), that is, it is simple for the decision boundary of source domain learning to misclassify the target sample when it is dispersed at the cluster's border or far from its associated class center. Due to this, Xu et al. presented a weighted optimum transmission technique that dynamically measures the sample-level inter-domain differences using spatial typical information and intra-domain structure.so as to achieve precise pin-by-pair optimal transmission and reduce negative migration of samples near the decision boundary of the target domain[2]. Considering that the previous methods ignore the information of the class when minimizing the field differences, it may lead to asymmetric alignment between classes, resulting in poor generalization performance. Kang et al. proposed a new measurement method, CDD (Contrastivedomain Discrepancies), which explicitly modeled the intra-class and inter-class discrepancies, and compressed the features of the in-class samples by minimizing the intra-class discrepancies.

2.2.2. Using Generative Adversarial Networks, Unsupervised Domain Adaptation Methods

The unsupervised domain adaptive technique based on the generative adversarial notion executes the distribution alignment by training a domain discriminator, in contrast to the unsupervised domain adaptive method based on the minimal divergence. and generally performs the adversarial learning in the feature space and input space of image classification for domain adaptation. introduced adversarial idea into image classification tasks for domain adaptation earlier[3].

Considering that using only one domain discriminator to determine which domain the extracted features come from may blur the boundary between different classes in the target domain, resulting in poor classification effect, Saito et al. made two different predictions for images in the target domain to improve the classification effect[4]. This is achieved by using the parameter regularization of the classification network to constrain features as far away from the classification boundary as possible [5], or by using two classifiers for collaborative training [6]. Cui et al. believed that only extracting domain invariant features would ignore features unique to different domains, so they proposed a modeling module for such different features for generator and discriminator respectively[7]. For generator, this module can transform both source domain and target domain to intermediate domain, reduce the difficulty of direct migration between different domains, and reduce the influence of residual unique features in the extracted domain invariant features. For discriminators, this module is helpful to improve the discriminating ability of discriminators and balance the training process of confrontation. Some approaches have begun to introduce graph theory and meta-learning methods into domain

Adaptation[8]. Wei et al. introduced meta-learning to alleviate the problem of different optimization objectives between domain alignment tasks and classification tasks, and regarded domain alignment tasks and classification tasks as meta-training and meta-testing of the same group of samples[9]. After that, some methods began to focus on real-time performance. Li et al. quickly realized the domain adaptive task of image classification with relatively high precision under the condition of low computing resources[10], and proposed confidence score learning and inter-class balance self-supervised training strategies to improve the classification accuracy in the target domain.

Table 1. Networks for small-sample image classification

Classification		Year	Methods	Generate	The data set
Ensemble learning		2018	Fangyi Zhu [2]	-	LabelMe, UIUC-Sports
		2020	Inter Boost [1]	-	UIUC-Sports LabelMe 15Scenes,Caltech101
Minimized divergence		2018	Rozantsev [3]	-	Of, MN, USPS, SVHN
		2019	CAN [4]	-	Of, VisDA
		2020	RWOT [5]	-	MN, USPS, SVHN, Of I-DA, O-H, VisDA
Generate against	Spatial character-istics	2018	Sohn [6]	-	MN, MN-M
		2020	Cui [7]	-	Of, O-H, VisDA
		2021	Meta Align [8]	-	Of, O-H
		2021	DDA [9]	-	Of, VisDA
	Input space	2018	SBADA-GAN [10]	S1T	MN, MN-M, USPS, SVHN, SYNS, GT
		2020	Ye [11]	T àT,S àS	MN, USPS, SVHN

2.2.3. Input Space

The minimum-divergence based approach and the antagonistic approach in the feature space are both domain adaptive in the feature space. With the continuous development of GAN, some work uses GAN for image generation and image style transfer to achieve data enhancement and generate more training data. During data enhancement, To determine if an image is real or artificially made, the discriminator is used. which wants the generated image to be as close to the real image as possible to fool the discriminator. Used to train image classification networks for source and target domains together with source domain images. the two classification networks were used to simultaneously predict image categories from target domain. The domain adaptive methods above aim to learn the domain invariant features. These techniques frequently require the source domain classifier to be updated in order to adapt to the target domain, hence they can fail to produce good results in both the source and target domains. In view of this, Ye et al. no longer train the classifier to adapt to the target domain[11], however, keep the classification performance of the source domain constant while improving the classification performance of the target domain.

3. Methodology

Compared with the image classification task, the research on the target detection task of small sample space is still in its infancy, so there are few relevant literatures. This paper mainly studies the method shown in Figure 2.

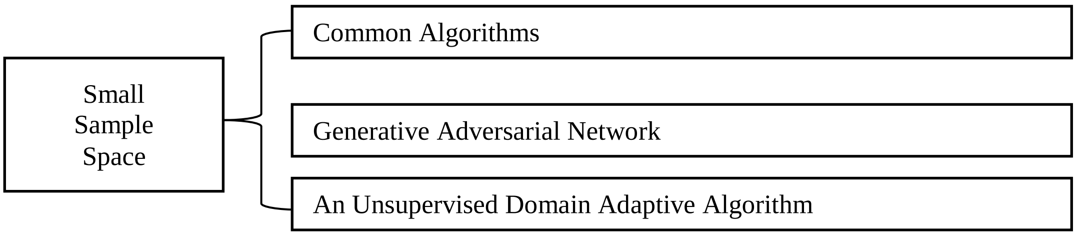

Figure 2. Research on small sample space target detection task

3.1. Common Target Detection Algorithms

The two types of common deep learning target identification algorithms are the single-stage and two-stage methods, respectively. R-CNN series based on candidate region serves as a representation of the two-stage detection procedure. Based on regression analysis, YOLO and SSD represent the single-stage detection algorithm. Scholars have now conducted extensive study on two-stage target detection algorithms and single-stage target detection algorithms, providing them with a theoretical foundation. The two-stage target detection method is dominating in terms of detection accuracy, but it has to be enhanced in order to increase detection speed. The single-stage target detection technique provides advantages in terms of detection speed, but the model must be regularly refined in order to enhance detection accuracy. As a result, some researchers mix the two algorithm models to achieve a balance between accuracy and speed of detection.

3.2. Target Detection Algorithm Based on Generative Adversarial Network

3.2.1. A-Fast-RCNN

In 2017, Wang et al. presented the A-Fast-RCNN algorithm and established the concept of an adversarial network to generate difficult positive samples using adversarial network [12]. A-fast-rcnn combines admissiveness learning with Fast R-CNN, and generates Hard examples through GANs to increase the number of targets with occlusion and attitude change. This approach uses some changes on feature photos as opposed to the conventional method, which directly generates sample images. A Mask layer is added to the Adversarial Spatial Dropout Net-work (ASDN) to realize feature partial occlusion. Mask is selected according to Loss, and then the adversarial network of deformation is processed shown in Figure 3.

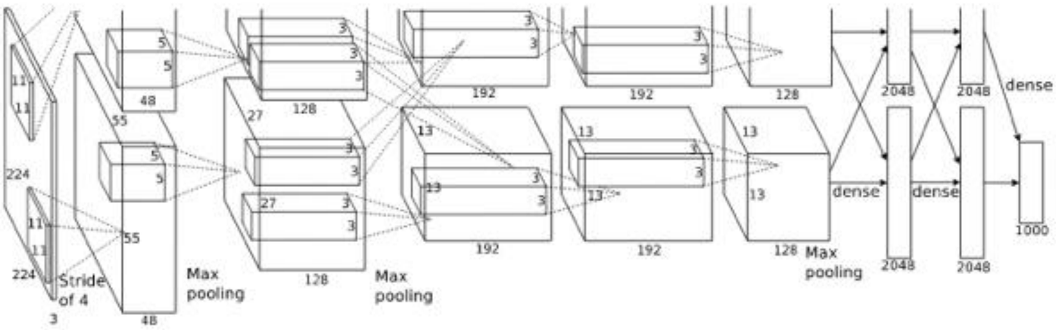

Figure 3. R-CNN convolution model

3.2.2. SOD-MTGAN

Bai et algoal.'s was to increase the Small target Detection's accuracysuggested a multi-task generative adversarial network for end-to-end small object detection in 2018. Algorithm Sod-mtgan [13]. Numerous trials on the COCO datasets show that the detection performance, ability to recover clear super-resolution images from small fuzzy images, and efficiency of this method. particularly for small objects, is superior to the most recent techniques. In order to produce high-resolution details, conventional convolutional generative adversarial networks (CGANs) only produce functions of spatial local points on low-resolution feature maps. According to Zhang et alproposed .'s self-attention Generative Adversarial Network (SA-gan), modeling picture production tasks can be dependent on attention for a long time[14]. It can create information from all feature locations and employs spectral normalization to improve training dynamics, with astonishing results. As shown in Figure 4, the original GAN network structure has problems such as disappearing gradient and difficult convergence, so many improved GAN networks appear.

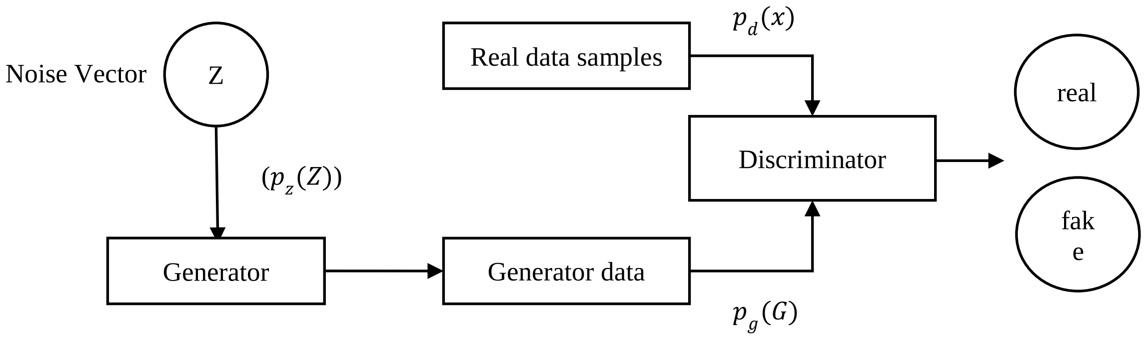

Figure 4. Generate adversarial network model structure

As shown in Figure 5, restrictions and constraints are added to the training process of GAN network to guide the process of data generation, which is the conditional generation adversarial network.

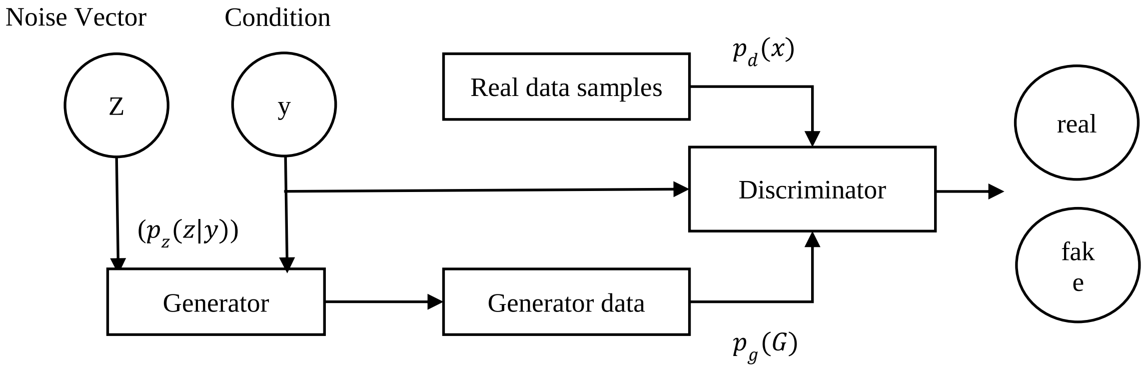

Figure 5. Conditional generation adversarial network structure

3.2.3. SAGAN

In order to produce high-resolution details, conventional convolutional generative adversarial networks (CGANs) only produce functions of spatial local points on low-resolution feature maps. Attention-driven and long-term dependence on modeling picture production tasks are made possible by the self-attention Generative Adversarial Network (SA-gan) introduced by Zhang et al [14]. It can generate data from any feature location and uses spectral normalization to improve training dynamics, producing astounding results.

3.2.4. Your Local GAN

For generative models, Daras et al. suggested Two Dimensional Local Attention Mechanisms. Introduced is a novel local sparse attention layer that preserves two-dimensional geometry and locality [15]. The FID score on ImageNet was optimized from 18.65 to 15.94, replacing the SAGAN (self-attention generating adversarial networks) intensive Attention force layer. A new approach is suggested to reverse the attentional network of antibiosis. The sparse attentional mode of the new layer presented in this method is created utilizing the new information theory criterion of information flow graph.

3.3. An Unsupervised Domain Adaptive Algorithm Based on Generative Adversarial

Unsupervised domain adaptive algorithm in target detection feature space adds domain discriminator to classical target detection framework for adversarial learning. Chen et al. earlier added domain discriminator to the Faster R-CNN framework for target detection domain adaptation [16]. Since target detection needs to recognize and locate one or more objects in the image, domain migration may occur at both image level (such as image style, illumination, etc.) and instance level (such as object appearance, size, etc.). In the target detection task, the target categories contained in the same image are often different and the distribution of target samples of different categories is different. Some previous work did not take such information into account, but treated the distribution of different types of objects in the image as a whole for domain adaptation. Therefore, Zheng et al proposed a two-stage domain adaptation method for feature alignment from coarse-grained to fine-grained [17]. To reduce incorrect category alignment between domains, VS et al. used multiple discriminators to align semantically identical categories in different domains. VS et al proposed the attention mechanism with memory to distinguish different types of targets, so as to send different types of targets to corresponding discriminators [18]. In addition, Xu et al. also enhanced the adaptive effect of such goals by assigning higher weights to categories that appeared less frequently.

Compared with the image classification task, the target detection research method of small sample data is still in its infancy. From the current development situation, the method used in the small sample data target detection task is similar to the basic idea used in the image classification task. In future research, the methods used in other mature computer vision tasks can be used to improve the mobility of the target detection model by combining the problems encountered in the domain adaptation of the target detection task itself.

4. Conclusion

Small sample oriented image recognition has achieved outstanding results in image classification, object detection and other computer vision tasks, which provides the possibility to achieve relevant computer vision tasks in the field lacking truth tags. This paper analyzes the problems faced by computer vision tasks in real scenes, proposes the research objectives of small sample image recognition, and reviews the research progress of supervised tasks and unsupervised domain adaptive methods based on deep learning in typical computer vision tasks in recent years. The research of unsupervised domain adaptive provides the possibility for the practical implementation and application of theoretical algorithms, and its integration with deep learning has injected new vitality into the field, which is a worthy direction of future computer vision and deep learning.

Funding

This work was supported by Science and Technology research project of Heihe University: KJY202001.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] X. Li, D. Chang, Z. Ma, Z. H. Tan, J. H. Xue, J. Cao, and J. Guo, Deep InterBoost Networks for Small-Sample Image Classification, Neurocomputing 456, 492 (2021).

[2] F. Zhu, Z. Ma, X. Li, G. Chen, J. T. Chien, J. H. Xue, and J. Guo, Image-Text Dual Neural Network with Decision Strategy for Small-Sample Image Classification, Neurocomputing 328, 182 (2019).

[3] A. Rozantsev, M. Salzmann, and P. Fua, Beyond Sharing Weights for Deep Domain Adaptation, IEEE Trans. Pattern Anal. Mach. Intell. 41, 801 (2019).

[4] G. Kang, L. Jiang, Y. Yang, and A. G. Hauptmann, Contrastive Adaptation Network for Unsupervised Domain Adaptation, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2019-June, 4888 (2019).

[5] R. Xu, P. Liu, L. Wang, C. Chen, and J. Wang, Reliable Weighted Optimal Transport for Unsupervised Domain Adaptation, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 4393 (2020).

[6] K. Sohn, W. Shang, X. Yu, and M. Chandraker, Unsupervised Domain Adaptation for Distance Metric Learning, 7th Int. Conf. Learn. Represent. ICLR 2019 (2019).

[7] S. Cui, S. Wang, J. Zhuo, C. Su, Q. Huang, and Q. Tian, Gradually Vanishing Bridge for Adversarial Domain Adaptation, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 12452 (2020).

[8] G. Wei, C. Lan, W. Zeng, and Z. Chen, MetaAlign: Coordinating Domain Alignment and Classification for Unsupervised Domain Adaptation, 1, 16643 (2021).

[9] S. Li, J. Zhang, W. Ma, C. H. Liu, and W. Li, Dynamic Domain Adaptation for Efficient Inference, 7832 (2021).

[10] P. Russo, F. M. Carlucci, T. Tommasi, and B. Caputo, From Source to Target and Back: Symmetric Bi-Directional Adaptive GAN, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 8099 (2018).

[11] S. Ye, K. Wu, M. Zhou, Y. Yang, S. H. Tan, K. Xu, J. Song, C. Bao, and K. Ma, Light-Weight Calibrator: A Separable Component for Unsupervised Domain Adaptation, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 13733 (2020).

[12] X. Wang, A. Shrivastava, and A. Gupta, A-Fast-RCNN: Hard Positive Generation via Adversary for Object Detection, Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017 2017-Janua, 3039 (2017).

[13] Y. Bai, Y. Zhang, M. Ding, and B. Ghanem, SOD-MTGAN: Small Object Detection via Multi-Task Generative Adversarial Network, Lect. Notes Comput. Sci. (Including Subser. Lect.

Notes Artif. Intell. Lect. Notes Bioinformatics) 11217 LNCS, 210 (2018).
[14] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, Self-Attention Generative Adversarial Networks, Electron. Lett. 55, 459 (2019).
[15] G. Daras, A. Odena, H. Zhang, and A. G. Dimakis, Your Local GAN: Designing Two Dimensional Local Attention Mechanisms for Generative Models, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 14519 (2020).
[16] S. Ren, K. He, R. Girshick, and J. Sun, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, Total Perform. Scorec. 159 (2020).
[17] Y. Zheng, D. Huang, S. Liu, and Y. Wang, Cross-Domain Object Detection through Coarse-to-Fine Feature Adaptation, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 13763 (2020).
[18] V. VS, V. Gupta, P. Oza, V. A. Sindagi, and V. M. Patel, MeGA-CDA: Memory Guided Attention for Category-Aware Unsupervised Domain Adaptive Object Detection, 4516 (2021).