Open Set Domain Adaptation with Zero-shot Learning on Graph

1st Xinyue Zhang
School of Artificial Intelligence,
University of Chinese Academy of Sciences.
State Key Laboratory of Management
and Control for Complex Systems,
Institute of Automation,
Chinese Academy of Sciences.
Beijing, China
zhangxinyue2020@ia.ac.cn

2nd Xu Yang*
State Key Laboratory of Management
and Control for Complex Systems,
Institute of Automation,
Chinese Academy of Sciences.
Beijing, China
xu.yang@ia.ac.cn

3rd Zhiyong Liu
State Key Laboratory of Management
and Control for Complex Systems,
Institute of Automation,
Chinese Academy of Sciences.
Beijing, China
zhiyong.liu@ia.ac.cn

***Abstract*—Open set domain adaptation focuses on transferring the information from a richly labeled domain called *source domain* to a scarcely labeled domain called *target domain*, while classifying the unseen target samples as one *unknown* class in an unsupervised way. Compared with the close set domain adaptation, where the source domain and the target domain share the same class space, the classification of the unknown class makes it easy to adapt to the real environment. Particularly, after the recognition of the unknown samples, the model can either ask for manually labeling or further develop the classification ability of the unknown classes based on pre-stored knowledge. Inspired by this idea, we propose a model for open set domain adaptation with zero-shot learning on the unknown classes in this paper. We utilize adversarial learning to align the two domains while rejecting the unknown classes. Then the knowledge graph is introduced to generate the classifiers for the unknown classes with the employment of the graph convolution network (GCN). Thus the classification ability of the source domain is transferred to the target domain, and the model can distinguish the unknown classes in detail with prior knowledge. We evaluate our model on digits datasets and the result shows superior performance.**

***Index Terms*—open set domain adaptation, zero-shot learning, knowledge graph, graph convolutional network, adversarial learning**

I. Introduction

In the last decades, deep learning models have shown good performance in various tasks, especially in visual perception. The training of the deep learning network relies on plenty of labeled data. However, most of the existing large labeled datasets are collected from the Internet. The images in these datasets are normative and unified, which are different from the images relevant for a specific application. Besides, depending on the application, the images may be obtained by different typed of visual sensors or with different perspectives of sensors. It costs a lot to retrain the classification model in different situations. Thus it is important to deal with the gap among domains. They should be able to utilize the well-labeled samples in the source domain to classify the samples in the unlabeled target domain, which is related to domain adaptation. There are already some researches on domain adaptation, such as [1], [2], [3], and [4]. The alignment of the domain gap makes the robot adapt well to dynamic and unstructured environments.

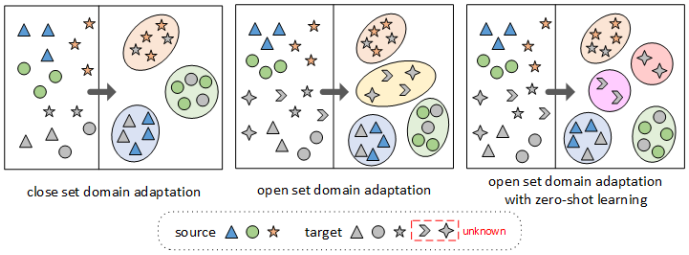

Fig. 1. An overview of the proposed domain adaptation with zero-shot learning. Close set domain adaptation aligns the target domain to the source domain. Open set domain adaptation not only aligns the domain gap but also rejects the unknown classes as one class. Open set domain adaptation with zero-shot learning further gives detailed classification on the unknown classes, which is more complex and valuable.

Except for the domain gap among different datasets, the variation of the classes also makes it hard for the model to adapt to a new dataset. Depending on the application and the scale of different datasets, the model may come across classes that are not contained in the source domains. With the traditional domain adaptation methods, the unknown classes are mistakenly aligned due to the absence of training samples of unknown classes in the source domain. The imbalance of the types of classes brings over-fitting problems and is not suitable for classification in the open world. Thus it is important for the robot to reject the unknown classes and only align the shared classes. This problem is known as open set domain adaptation, which is first proposed by [5] and followed by for instance [6], and [2]. In the setting of the open set domain adaptation, the target domain contains both the classes of the source domain and the additional new classes. The model not only aligns the target domain to the source domain but also

This work was supported by Beijing Science and Technology Plan Project (grant Z201100008320029).

rejects the unknown classes.

It is worth noting that previous open set domain adaptation methods typically classify all the additional new classes into one *unknown* class. However, the unknown class may contain classes that are worth learning. It may be more valuable to detect the unknown classes in detail and develop the ability to classify them with the former information. Since the unknown classes are not included in the source domain, the model lacks the labeled information for the new classes. Current open set domain adaptation methods can not give detailed classification on the unknown part with no labeled images. This problem is related to zero-shot learning. In the zero-shot learning problem, complementary information is collected to transfer the knowledge from the base classes to classify the unknown ones. Inspired by this, with the knowledge stored in the knowledge graph, the classifiers of the unknown classes can be obtained in the target domain with no labeled samples.

Towards this end, we propose a generic model to align the gap between the labeled source domain and the unlabeled target domain while classifying the unknown classes in the target domain, which we call open set domain adaptation with zero-shot learning. The contributions of this paper mainly lie in tackling the following two difficulties.

First, since the unknown classes are not contained in the source domain, we have no labeled samples for supervised training. The lack of labeled data may cause the over-fitting problem of the model, which means the model only classifies the samples as the known classes and can not classify the unknown ones. It is necessary to utilize complementary information to support the inference. Thus we employ the knowledge graph to store some prior knowledge of the known classes and the unknown classes, which contains the structural relations between different classes, beyond the individual attribute representation of each class. The structural information offers a bridge for the inference from the known classes to the unknown ones. With the employment of the graph convolution network, the information propagates among the graph, and the unknown classes gather the information from their neighbor to generate their classifiers. These inference classifiers work as the initial classifiers of the classification model.

The second difficulty is how to adapt the inference classifiers to the target domain. The inference classifiers are suitable to the source domain. It is not able to classify the unknown samples in the target domain because of the domain gap. Thus we introduce adversarial learning to align the domain gap. The classification model consists of two modules, the feature generator and the classifier. Since the generator works to extract the features of the samples and the classifier works to output the class probability, we train them simultaneously in an adversarial way. The classifier is trained to find a boundary for unknown classes, while the generator is trained to make the samples far from the boundary. With adversarial learning, the generator can deceive the classifier into generating aligned features in both domains and reject the unknown classes according to the unknown boundary. Thus the feature of shared classes is aligned in both domains, and the unknown classes are rejected as one class. With the adaptation in both domain gap and class gap, our model is able to classify objects in the dynamic and complex open world. We utilize the knowledge graph and the adversarial learning in a jointly trained framework. We further evaluate our method on digits datasets and demonstrate its effectiveness.

II. Related Works

A. Open Set Domain Adaptation

Open set domain adaptation goes beyond traditional close set domain adaptation. It considers a more realistic classification task in which the target domain contains unknown samples that are not present in the source domain. Open set domain adaptation is first proposed by [5]. They measure the distance between the target sample and the center of the source class to decide whether a target sample belongs to one of the source classes or the unknown class. However, they require the source domain to have unknown samples as well. Later on, [6] propose open set back-propagation (OSBP) for target domain with no unknown samples. They utilize adversarial learning to align the domain gap. The learnable information in the unknown space deserves deep exploitation. We have found few papers that consider the fine-grained classification of the unknown classes in open set domain adaptation, and we aim to fill in the blanks.

B. Zero-shot Learning

Zero-shot learning aims at generating classifiers for unknown classes with no labeled samples. Several pieces of research have been done on this area, such as [7] [8]. Due to the limitation of the available samples, some researchers extract complementary information from the related known classes to support the inference of the unknown ones. Among these methods, building the relationship between classes in the form of a graph seems more reasonable. The special geometry of graphs well shows the complicated relationship, and the unknown classes can gather adequate information from the known ones. [9] built an unweighted knowledge graph combined with word embedding upon the graph convolutional network. With information propagation, novel nodes generate predictive classifiers with common sense. [10] improve upon this model and propose a dense graph propagation to prevent dilution of knowledge from distant nodes.

III. Approach

A. Problem Definition

In open set domain adaptation with zero-shot learning, we have a source domain $D_s = \{(x_i^s, y_i^s)\}_{i=1}^{n_s}$, which contains n_s labeled samples, and x_i^s refers to the ith source images and y_i refers to its label. Target domain is denoted as $D_t = \{x_j^t\}_{j=1}^{n_t}$, which contains n_t unlabeled samples. The class space in the source domain is C_s which we call known classes and contains M classes. The known classes are shared by the class space of the target domain C_t, which contains N classes. It is worth noting that C_t further contains $N - M$ unknown classes C_u, that is $C_t = C_s \vee C_u$. The distribution of the source domain

and target domain is different. Note that the samples in the target domain are all unlabeled and the samples in the source domain are all labeled.

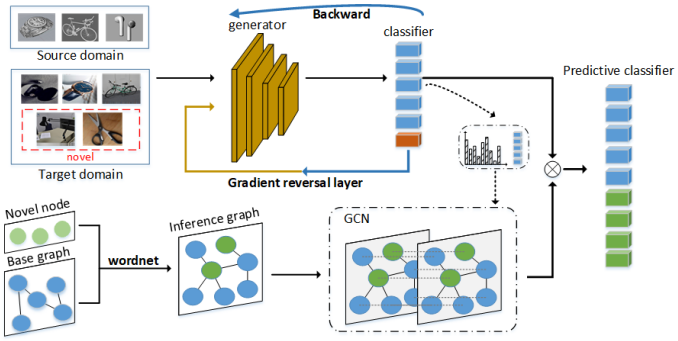

Fig. 2. An overview of our model for domain adaptation with zero-shot learning.

B. classifier inference module

With few labeled samples, humans can make good inferences on unfamiliar things with the related information obtained from books. Our model also extracts the task-based knowledge from a prestored knowledge graph. The knowledge graph is denoted as $G = (V, E)$, where $V = \{v_1, v_2, ..., v_N\}$ is a node-set of all classes including known classes and unknown classes. The nodes features in the knowledge graph are word embedding attributes v_i of different classes. Edge set $E = \{e_{i,j} = (v_i, v_j)\}$ refers to the relationship among classes. The edges in the knowledge graph are decided by the similarity of the attributes between different classes.

Since the labeled samples in the source domain are available. The original recognition model is first trained on the source samples D_s, which is denoted as $C(F(\cdot|\theta)|W^s)$. The recognition model consists of two parts, feature extractor $F(\cdot|\theta)$ and classifier $C(\cdot|W^s)$, where θ and W^s indicate the parameters of the model trained on D_s. Feature extractor $F(x_i|\theta)$ takes an image as input and figures out the feature vector as z_i. The classifier $C(z_i|W^s)$ works to compute the classification score which is denoted as

$$[s_1, s_2, ..., s_M] = [z^T w_1, z^T w_2, ..., z^T w_M] \tag{1}$$

Thus the inference of the classifiers on unknown classes turns to inference of the classification weights w_s on the unknown classes.

With the framework of the graph convolutional network (GCN), our model propagates information among nodes by exploring the class relationship. For one layer in GCN, a node aggregates information from its neighbors. GCN can also be extended to multiple layers to perform a deeper spread. Therefore, the unknown classes can utilize the information from the related known classes and predict the classification weights of their own. The mechanism of GCN is described as

$$H^{(l+1)} = ReLu(\hat{D}^{-\frac{1}{2}}\hat{E}\hat{D}^{-\frac{1}{2}}H^{(l)}U^{(l)}) \tag{2}$$

where $H^{(l)}$ denotes the output of the l^{th} layer, while for the first layer $H^0 = V$. It uses Leaky ReLu as the nonlinear activation function. To reserve the self-information of the nodes, self-loops are added among the propagation, $\hat{E} = E+I$, where $E \in R^{N\times N}$ is the symmetric adjacency matrix and $I \in R^{N\times N}$ represents identity matrix. $D_{ii} = \sum_j E_{ij}$ normalizes rows in E to prevent the scale of input modified by E. The matrix U^l is the weight matrix of the l^{th} layer, which GCN regulates constantly to achieve better performance.

Our model conducts two layers of GCN on the knowledge graph. Unknown classes learn the mechanism of end-to-end learning from known classes through propagation. The output of the GCN is trained by minimizing the loss between the predicted classification weights and the ground-truth weights. The ground-truth weights refer to the classifiers of the known classes, which are extracted from the original recognition model on the source domain.

$$W^{inf} = \text{softmax}(A(\text{relu}(AVU^{(0)}))U^{(1)}), \tag{3}$$

$$A = \hat{D}^{-\frac{1}{2}}\hat{E}\hat{D}^{-\frac{1}{2}}, \tag{4}$$

$$L_{GCN} = \frac{1}{M}\sum_{i=1}^{M}(w_i^{inf} - w_i^s)^2 \tag{5}$$

where $W^{inf} = \{w_1^{inf}, w_2^{inf}, ..., w_N^{inf}\}$ refers to inference classifiers of all the classes, and $W^s = \{w_1^s, w_2^s, ..., w_M^s\}$ denotes the ground truth classifiers of the known classes obtained from the original recognition model. We utilize the M classifiers of known classes from the output of GCN to evaluate the loss. With the supervision of the known classes, the unknown nodes in the inference graph can also generate classifier weights of their own. Finally, with the employment of GCN, the classifier inference module not only generates predictive classifiers of the unknown classes in the target domain but also provides more general classifiers of the known ones.

C. domain adaptation module

With the employment of the classifier inference module, classifiers of the unknown classes are generated. However, these classifiers are only suitable for the source domain since the ground-truth classifiers are extracted from the original model trained on the source domain. Thus the domain adaptation module attempts to align the domain gap between the source domain and target domain, which can transfer the generated classifiers to the target domain.

The inference classifiers W^{inf} are applied to the original recognition model, denoted as $C(F(\cdot|\theta)|W^{inf})$. Note that the number of the classifiers expands from M to N. As mentioned above, the recognition model consists of two parts, the feature generator and the classifiers. To align the domain gap, we employ adversarial learning on the classifiers and the feature generator. The classifiers are trained to set a boundary for the unknown classes in the target domain. With the boundary, unknown classes can be picked out. The proportion of unknown classes in the target domain is denoted

as $p_{un} = \sum_{i=M+1}^{N} p(y = y_i|x^t)$ The classifiers are trained to output $p_{un} = t$, where t is the boundary. The feature generator tries to generate features that can deceive the classifier. That is, the generator tries to generate features far from the boundary. It can choose to decrease or increase p_{un} far from t. Besides, the classification ability on the known classes should be reserved. Thus we also consider the classification accuracy on the source domain during the training process. We use a standard cross-entropy loss for this purpose.

$$L_s(x_s, y_s) = -\log(p(y = y_s|x_s)) \tag{6}$$

$$p(y = y_s|x_s) = (C(F(x_s)))_{y_s} \tag{7}$$

With the cross-entropy loss, the model ensures the classification accuracy on known classes. For the boundary of the unknown classes, we follow the settings in the OSBP and utilize binary cross-entropy loss.

$$L_{adv}(x_t) = -t\log(p_{un}) - (1-t)\log(1-p_{un})) \tag{8}$$

To train the classifier inference module and the domain adaptation module jointly, the overall objective of our model is,

$$\min_C L_s(x_s, y_s) + L_{adv}(x_t) + L_{GCN} \tag{9}$$

$$\min_G L_s(x_s, y_s) - L_{adv}(x_t) + L_{GCN} \tag{10}$$

With the domain adaptation module, the unknown classes in the target domain are separated, and the features of both domains are aligned. We also suggest iterating the classifier inference module and the domain adaptation module for better performance.

IV. Experiments

A. Datasets

We test our model on three digits datasets. Compared to the traditional dataset on domain adaptation. The digits datasets contain a fewer number of classes, which means the number of the known classes is fewer. The task turns to a more difficult zero-shot learning problem. The three digits datasets are MNIST [11], USPS [12], and SVHN [13]. For the class space, we have two settings of unknown classes. In the 3-way setting, the source domain contains seven classes (0-6), while the target domain contains ten classes (0-9). While in the 2-way setting, the source domain contains seven classes (0-7), while the target domain contains ten classes (0-9). In both settings, our goal is to align the known classes in the target domain to the source domain and have the ability to classify the unknown ones.

B. Comparison

To test the performance of our model, we conduct experiments under several settings. However, since there are few models that work on the domain adaptation with zero-shot learning, we compare our model with zero-shot learning models. Besides verifying the value of the inference classifiers, we also compare our model to open set domain adaptation methods with random initialization unknown classifiers. The results are shown in the following table.

TABLE I
Comparation Results

Task	SVHN → MNIST			
Setting	2-way		3-way	
	all	unknown	all	unknown
z-GCN [9]	48.4%	6.2%	39.5%	13.2%
OSBP [6]	58.2%	17.4%	54.0 %	24.3%
our model	67.0%	38.2%	64.3%	46.4%
Task	USPS → MNIST			
Setting	2-way		3-way	
	all	unknown	all	unknown
z-GCN [9]	60.5%	9.5%	54.1%	10.4%
OSBP [6]	61.4%	8.5%	62.3 %	12.4%
our model	67.4%	26.5%	69.2%	24.1%
Task	MNIST → USPS			
Setting	2-way		3-way	
	all	unknown	all	unknown
z-GCN [9]	63.4%	8.6%	62.0%	12.3%
OSBP [6]	51.3%	7.6%	42.2%	13.3%
our model	73.6%	49.3%	68.2%	23.5%

The comparison between our model and other exiting methods is reported in table 1. The 2-way and 3-way settings mean the number of unknown classes is 2 and 3. The class type setting: all and unknown, refer to the classification accuracy on the overall 10 classes and the unknown classes only. The performance is evaluated by the average top-1 accuracy. z-GCN proposed by [9] is a zero-shot learning model with the employment of the knowledge graph. It only considers the class gap while ignoring the domain gap. OSBP proposed by [6] is a model focusing on open set domain adaptation. For comparison, we expand it with randomly initialized classifiers on the unknown classes.

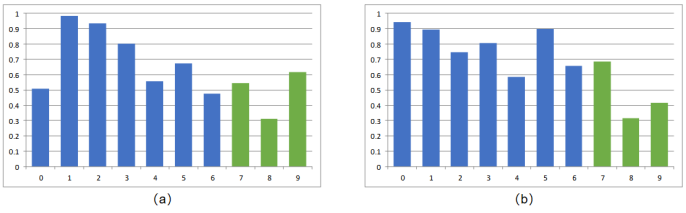

Fig. 3. The classification accuracy in the target domain: (a) the result in 3-way SVHN to MNIST experiment (b) the result in 3-way USPS to MNIST experiment.

From table 1, we notice that the classification accuracy on z-GCN is about ten to twenty percent lower than our model, which demonstrates the importance of domain alignment. Besides, the classification accuracy on the unknown classes of our model is about twenty percent higher than z-GCN. We owe that to the domain gap. Since the unknown classifiers of the z-GCN are generated with the labeled known samples in the source domain, it is not suitable in the target domain, which results in a huge decrease in the accuracy. Compared with zero-shot learning methods, our method transfers the

inference classifiers to the target domain and shows about sixty percent overall accuracy and thirty percent accuracy on the unknown classes. As figure 4 shows, the domain gap between the MNIST and the USPS is large. Overfitting on the source domain and the lack of labeled training images in the target domain affect the results a lot. In the SVHN to MNIST tasks with the 2-way setting, the classification accuracy of z-GCN on all classes is forty-eight, which is twenty percent lower than our model. In the USPS to MNIST and MNIST to USPS tasks, the improvement of our model is about ten percent as well. The result demonstrates that the adversarial learning employed by our model is able to transfer the classification ability from the source domain to the target domain. The domain adaptation is important for the flexibility of the models.

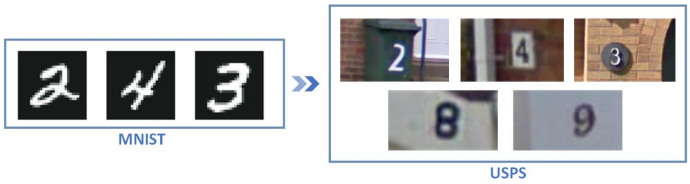

Fig. 4. In the task setting m2u, the model is trained on the MNIST dataset with seven or eight classes and transferred to the USPS dataset with ten classes.

To test the effectiveness of the inference classifiers, we further conduct experiments on the open set domain adaptation method. From the results in table 1, the classification ability of our model on the unknown classes is about twenty percent higher than random-expanded OSBP. The fine-grained classification accuracy on every class is shown in figure 3. Besides, we notice that OSBP still shows about ten percent accuracy on the unknown classes and sometimes even higher than z-GCN. We owe the classification accuracy on the unknown classes to the rejection mechanism. Since OSBP has the ability to reject the unknown classes as one class, the detailed classification in the one class is much easier. Besides, the inaccurate classifiers on the unknown classes confuse the classifier on the known classes and result in a decrease in the accuracy. To avoid randomness, we perform three different domain adaptation tasks. From the result shown above, we can come to the conclusion that our model shows a good performance on open set domain adaptation with zero-shot learning.

We visualize the output of the model in the target domain with t-SNE. Figure 5 shows the visualization results. The samples from the same class are grouped together, while those belonging to different classes are separated. Besides, the unknown classes are separated with each other in the visualization, like class seven, eight and nine. Although the domain adaptation module reject the unknown class as one class, the inference classifiers still have the fine-grained classification ability with the support of knowledge graph. The visualization demonstrates that the inference classifiers generated from the knowledge graph are discriminative.

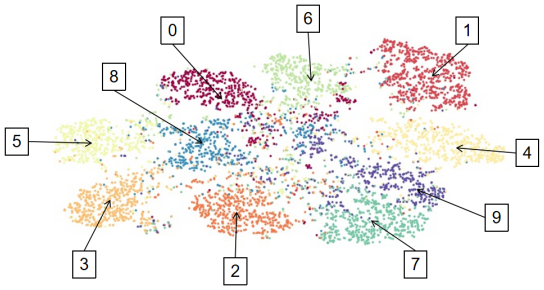

Fig. 5. Visualization of the class samples in the target domain.

V. Conclusion

In this paper, we propose a model on open set domain adaptation with zero-shot learning. Our model not only makes good performance on the alignment of the domain gap but also gives detailed classification on the unknown classes. The ability of the further classification on the unknown classes improves the visual cognitive development ability of the robot, which is important for the robot working in a realistic environment. The experiments show that our model has a good performance on domain adaptation with zero-shot learning.

References

[1] A. Farahani, S. Voghoei, K. Rasheed, and H. R. Arabnia, "A brief review of domain adaptation," *Advances in Data Science and Information Engineering*, pp. 877–894, 2021.

[2] N. Xiao and L. Zhang, "Dynamic weighted learning for unsupervised domain adaptation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15242–15251, 2021.

[3] S. Zhao, X. Yue, S. Zhang, B. Li, H. Zhao, B. Wu, R. Krishna, J. E. Gonzalez, A. L. Sangiovanni-Vincentelli, S. A. Seshia, *et al.*, "A review of single-source deep unsupervised visual domain adaptation," *IEEE Transactions on Neural Networks and Learning Systems*, 2020.

[4] G. Wei, C. Lan, W. Zeng, and Z. Chen, "Metaalign: Coordinating domain alignment and classification for unsupervised domain adaptation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16643–16653, 2021.

[5] P. Panareda Busto and J. Gall, "Open set domain adaptation," in *Proceedings of the IEEE international conference on computer vision*, pp. 754–763, 2017.

[6] K. Saito, S. Yamamoto, Y. Ushiku, and T. Harada, "Open set domain adaptation by backpropagation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.

[7] S. Badirli, Z. Akata, G. Mohler, C. Picard, and M. Dundar, "Fine-grained zero-shot learning with dna as side information," *Advances in Neural Information Processing Systems*, vol. 34, 2021.

[8] W. Xu, Y. Xian, J. Wang, B. Schiele, and Z. Akata, "Attribute prototype network for zero-shot learning," *Advances in Neural Information Processing Systems*, vol. 33, pp. 21969–21980, 2020.

[9] X. Wang, Y. Ye, and A. Gupta, "Zero-shot recognition via semantic embeddings and knowledge graphs," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 6857–6866, 2018.

[10] M. Kampffmeyer, Y. Chen, X. Liang, H. Wang, Y. Zhang, and E. P. Xing, "Rethinking knowledge graph propagation for zero-shot learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11487–11496, 2019.

[11] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[12] J. Friedman, T. Hastie, R. Tibshirani, *et al.*, *The elements of statistical learning*, vol. 1. Springer series in statistics New York, 2001.

[13] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng, "Reading digits in natural images with unsupervised feature learning," 2011.