Research Article

Knowledge Graph-Based Image Recognition Transfer Learning Method for On-Orbit Service Manipulation

Ao Chen ¹, Yongchun Xie,^{1,2} Yong Wang,^{1,2} and Linfeng Li¹

¹Beijing Institute of Control Engineering, Beijing, China

²Science and Technology on Space Intelligent Control Laboratory, Beijing, China

Correspondence should be addressed to Ao Chen; chenao526@foxmail.com

Received 30 April 2021; Accepted 24 June 2021; Published 6 August 2021

Copyright © 2021 Ao Chen et al. Exclusive Licensee Beijing Institute of Technology Press. Distributed under a Creative Commons Attribution License (CC BY 4.0).

Visual perception provides state information of current manipulation scene for control system, which plays an important role in on-orbit service manipulation. With the development of deep learning, deep convolutional neural networks (CNNs) have achieved many successful applications in the field of visual perception. Deep CNNs are only effective for the application condition containing a large number of training data with the same distribution as the test data; however, real space images are difficult to obtain during large-scale training. Therefore, deep CNNs can not be directly adopted for image recognition in the task of on-orbit service manipulation. In order to solve the problem of few-shot learning mentioned above, this paper proposes a knowledge graph-based image recognition transfer learning method (KGTL), which learns from training dataset containing dense source domain data and sparse target domain data, and can be transferred to the test dataset containing large number of data collected from target domain. The average recognition precision of the proposed method is 80.5%, and the average recall is 83.5%, which is higher than that of ResNet50-FC; the average precision is 60.2%, and the average recall is 67.5%. The proposed method significantly improves the training efficiency of the network and the generalization performance of the model.

1. Introduction

On-orbit service (OOS) manipulation is the terminal operation that service spacecraft carrying out on target spacecraft after their rendezvous and docking. Typical OOS manipulation uses robotic arm to perform tasks like repairment, module replacement, refueling, assistant deorbiting on end-of-life, or malfunctioning spacecrafts [1]. The visual perception system provides the control system with the status information of the current operating scene, which plays an important role in the OOS manipulation. Currently, the visual perception system used for on-orbit service manipulation mostly adopts the target detection technology based on visible light imaging. In teleoperation, the visual perception system only needs to provide surveillance video and representative task for which is “Robot Refueling Mission” [2]. For partial or fully autonomous manipulation tasks, the visual perception system needs to identify targets with pose measurements, as in tasks like “Orbital Express” [3] and “ETS-VII” [4]. Despite that the visual perception system can measure cooperative markers, for noncooperative targets, only simple geometric features

like points, straight lines, circles, and curves can be identified with certain prior information.

Recent development of deep learning, especially deep CNNs, has endowed many applications in visual perception. In 2012, Krizhevsky et al. proposed AlexNet [5] and won the ILSVRC-2012 challenge with a top-5 error rate of 15.3% adopting AlexNet. Since then, deeper CNNs, such as VGG [6], Inception Net [7], and ResNet [8], have been proposed which continue improving the accuracy of target recognition. In addition, these networks are also used as the basic feature extractor and have been widely used in tasks such as target detection, behavior recognition, and scene understanding. The visual perception method based on CNNs can not only provide the pose information of the target in the current scene for the traditional control method but can also be used as a part of deep reinforcement learning (DRL) to extract more abstract features through the agent-environment interaction. Compared with traditional visual perception method, the application scope of CNNs is wider.

When the training data is distributed all over its marginal distribution function and the test data and the training data

meet the condition of independently identically distribution (iid), the deep CNNs do show great advantages in feature extraction of complex images. However, the scene of on-orbit service manipulation does not satisfy the iid condition and has its particularities: (1) the DRL method used for on-orbit service manipulation is typically first trained in a simulation environment and then transferred into the real environment. For the visual perception system, the targets built in the simulation environment are not perfectly match those correspondence (i.e., of shape and size) in the real environment, leading to the data distribution shift between training and test dataset; (2) insufficient lighting simulation in simulation environment aggravates the aforementioned data distribution shift. As a result, for on-orbit service manipulation, the deep CNNs mentioned above cannot be directly applied on the visual perception system.

Transfer learning studies how to transfer model learned from the source domain to the target domain. The algorithm based on identifiability proposed by Thrun and Pratt [9] is considered to be the first transfer learning algorithm. In 1995, Thrun and Pratt carried out discussion and research on “Learning to learn,” wherein they argue that it is very meaningful to research lifelong machine learning algorithms that can store and use previously learned knowledge. According to the knowledge being transferred, the transfer learning algorithm can be divided into instance-based transfer method, feature representation-based transfer method, model parameter-based transfer method, and relationship-based transfer method [10]. Current transfer learning methods are mostly based on the transfer of feature representation, which is to adapt the marginal distribution of the input sample $P(\mathbf{x})$, prediction conditional probability distribution $P(\mathbf{y} | \mathbf{x})$, or the joint distribution $P(\mathbf{x}), P(\mathbf{y} | \mathbf{x})$ of different domains to realize the transfer. The key of applying this adaptation method in deep CNNs is to share the shallow model parameters between source and target domain and use the maximum mean discrepancy (MMD) to adapt deep model parameters in order to achieve feature transfer. The Deep Domain Confusion (DDC) method proposed by Tzeng et al. [11] uses AlexNet as the backbone, adds an adaptation layer in front of the classifier, and designs a loss function containing adaptation errors to train and learn network parameters. Long and Wang proposed Deep Adaption Network (DAN) [12] as an extension of the DDC. In DAN, three layers in front of the classifiers have all been adapted, and the multicore MMD is introduced. Compared with DAN, DDC achieves better results on the classification transfer task. [13] proposes a selective adversarial networks to promote positive partial transfer learning using domain discriminator. Similar to [13, 14] adopt an importance weighted adversarial networks to identify shared labels between source and target domains and strengthen the knowledge transfer. In order to transfer the segmentation network to unseen domains, [15] proposes a domain adaption module to calculate adversarial loss to transfer the feature from target domain, and the adaption module is inserted in two different levels.

One method closely relates to transfer learning is few-shot learning (FSL). FSL studies how to use a small scale of

labeled data to train models. The representative method is metalearning (ML). In ML, the training dataset is randomly divided into a series of metatasks, and each metatask is composed of a support set and a batch set. For metatask learning, the model can learn how to predict the data in the batch set based on the support set [16, 17]. In FSL, there is also a type of metric-based method. This method models the distance distribution between samples, similar to the method of layer adaptation in transfer learning. The representative method is Siamese Neural Networks (SNNs) [18], Matching Networks (MNs) [19], and Prototypical Networks (PNs) [20]. [21] constructed the semantic relationship between various samples based on word embedding, thereby strengthening the learning of few-shot data.

However, current transfer learning methods and few-shot learning methods train models through data grouping and metric optimization and do not explicitly introduce prior knowledge, which limits the efficiency and the generalization performance of those methods.

This paper presents a method called Knowledge Graph-based Image Recognition Transfer Learning (KGTL), which realizes end-to-end target recognition. Firstly, a deep CNN (e.g., ResNet50) is utilized to extract target features from complex collected images. Secondly the knowledge graph is adopted to explicitly encode semantic relationships between the targets. Graph Convolutional Networks (GCNs) are used to train the transferable classifier, and eventually, the transferable classifier is utilized to classify the extracted features. The proposed method can transfer the target recognition model learned in the simulation environment to the real environment. The average target recognition accuracy rate reaches 80.5%, and the average recall rate reaches 83.5%, which is higher than general CNNs based on ResNet50 with a fully connected layer classifier, whose average accuracy rate is 60.2%, and average recall rate is 67.5%. Under the premise of ensuring recognition accuracy, the proposed method significantly reduces the amount of real images needed in the training process, which rather relies on images collected in a simulation environment easier to be obtained with low cost. It improves training efficiency and generalization of the model, and it is of great significance for studying how to transfer the target recognition model learned on the ground to space.

2. Description and Definition of Image Recognition Transfer Learning

We focus on how to transfer the model trained in the source domain (SD) to the target domain (TD). So firstly, concepts and the specific description of the source domain and the target domain in the problem studied in this paper are introduced. Then, integral problem description and definition of image recognition transfer learning are given.

2.1. The Concept of Source Domain and Target Domain. The source domain consists of data $\mathcal{D}_S$ and tasks $\mathcal{T}_S$, i.e., $SD = \{\mathcal{D}_S, \mathcal{T}_S\}$. Data $\mathcal{D}_S$ is sampled from the feature space $\mathcal{X}_S$, whose marginal probability distribution is $P_S(\mathbf{x})$. That is, data $\mathcal{D}_S$ can be described as $\mathcal{D}_S = \{\mathcal{X}_S, P_S(\mathbf{x})\}$. Task $\mathcal{T}_S$

corresponds to the label space $\mathcal{Y}_S$, and its conditional probability distribution is $P_S(\mathbf{y} | \mathbf{x})$. That is, the task $\mathcal{T}_S$ can be described as $\mathcal{T}_S = \{\mathcal{Y}_S, P_S(\mathbf{y} | \mathbf{x})\}$. In this paper, the data in the source domain is collected from a simulation environment (generated by SOLIDWORKS), whose targets are built according to the shape, size, and color of the target in the ground physical environment for on-orbital refueling. The task in the source domain is to recognize the targets in the simulation environment.

The target domain consists of data $\mathcal{D}_T$ and tasks $\mathcal{T}_T$, i.e., $\text{TD} = \{\mathcal{D}_T, \mathcal{T}_T\}$. Data $\mathcal{D}_T$ is sampled from the feature space $\mathcal{X}_T$, whose marginal probability distribution is $P_T(\mathbf{x})$. That is, data $\mathcal{D}_T$ can be described as $\mathcal{D}_T = \{\mathcal{X}_T, P_T(\mathbf{x})\}$. Task $\mathcal{T}_T$ corresponds to the label space $\mathcal{Y}_T$, and its conditional probability distribution is $P_T(\mathbf{y} | \mathbf{x})$. That is, the task $\mathcal{T}_T$ can be described as $\mathcal{T}_T = \{\mathcal{Y}_T, P_T(\mathbf{y} | \mathbf{x})\}$. In this paper, the data in the source domain is collected from the ground physical environment, whose targets are built according to the requirements of the on-orbit refueling mission. The task in the target domain is to recognize the targets in the ground physical environment.

Samples collected from the source domain and the target domain are the RGB images with same resolution, i.e., $\mathcal{X}_S = \mathcal{X}_T$. However, due to modeling error, the marginal distribution of data is different, i.e., $P_S(\mathbf{x}) \neq P_T(\mathbf{x})$. The targets needed to be identified in the source domain and the target domain are the same; that is, $\mathcal{Y}_S = \mathcal{Y}_T$. The distribution of all types of samples is even and consistent in both the source domain and the target domain; that is, the conditional probability distribution of the task is the same, $P_S(\mathbf{y} | \mathbf{x}) = P_T(\mathbf{y} | \mathbf{x})$. After clarifying the description and characteristics of the source domain and target domain involved in this paper, problem of image recognition transfer learning is defined as follows.

2.2. Problem Description. In the image recognition task, we consider the case wherein marginal distribution of the data collected from SD and TD is inconsistent while the feature space is the same, that is

$$\begin{aligned} \mathcal{D}_S &\neq \mathcal{D}_T (\mathcal{X}_S = \mathcal{X}_T, P_S(\mathbf{x}) \neq P_T(\mathbf{x})), \\ \mathcal{T}_S &= \mathcal{T}_T. \end{aligned} \quad (1)$$

During training, a large amount of labeled data in the source domain and a small amount of labeled data in the target domain can be obtained. The problem is to design an image recognition transfer algorithm such that the trained network is also effective for the large amount of data collected in the target domain.

3. Knowledge Graph-Based Image Recognition Transfer Learning Method

The traditional image recognition method based on CNNs is to train a fully connected classifier to classify the features extracted by the representation learning network (such as ResNet50). This classifier can only classify the learned image pattern; that is, $\mathcal{D}_S = \mathcal{D}_T$, which cannot perform semantic-

level reasoning and is difficult to be used for image recognition transfer learning.

To this end, we propose an image recognition transfer learning method based on knowledge graph. The method consists of two parts: representation learning based on CNNs and classifier learning based on knowledge graph. In the CNN-based representation learning, we use ResNet50 as an example, which is used to extract image features of complex objects. In the classifier learning part, knowledge graph is used to explicitly encode semantic relationships between the targets, and GCN is adopted to train transferable classifiers. Finally, the learned transferable classifier is applied to classify the extracted image features.

Overviews of the image recognition transfer learning method based on the knowledge graph, the representation learning module based on CNNs, and the classifier learning module based on the knowledge graph are, respectively, introduced as follows.

3.1. Overview of Knowledge Graph-Based Image Recognition Transfer Learning Method. Here is the overview of the proposed knowledge graph-based image recognition transfer learning method, as shown in Figure 1.

The scheme includes a representation learning module based on CNNs (the upper part of Figure 1) and a classifier learning module based on the knowledge graph (the lower part of Figure 1). The representation learning module based on CNNs uses convolution operation to extract image features hierarchically, while the classifier learning module based on knowledge graph adopts GCNs to update the classifier nodes in the knowledge graph integrated with prior semantic relations. Finally, the learned transferable classifier is utilized to classify the extracted image features to obtain the final image recognition result. In Figure 1, metric is defined to calculate the target label based on the extracted features and learned classifiers. The two modules are introduced separately as below.

3.2. Convolutional Neural Network-Based Representation Learning [8]. CNN has good performance on feature extraction of complex images. This paper uses ResNet50 proposed by He et al. as the basic representation learning module. ResNet50 takes the image as input and computes a high-dimensional feature vector $\mathbf{f} \in \mathcal{R}^{N_f \times 1}$ to represent the key feature of the image through hierarchical operations.

ResNet50 first uses one convolution layer and four cascaded convolution modules to extract multilevel features of the image and then adopts a fully connected layer to concatenate the extracted features at each level to generate a multidimensional feature vector, which represents the characteristics of the entire image. Each convolution module mentioned above is composed of multiple residual modules in series. Figure 2 shows the structure of the residual module in the first convolution module. $1 \times 1, 64$, $3 \times 3, 64$, and $1 \times 1, 256$ represent the convolution operation with different channels and different kernel sizes. The residual module sets the results of some convolutional layer close to zero through supervised learning; then, the shot-cut connection in the residual module can ensure that the information

FIGURE 1: Overview of proposed knowledge graph-based image recognition transfer learning method.

FIGURE 2: Structure of residual block.

extracted by the upper layer of the network is maintained and can be transferred to the next layer of network, so as to retain key information in each layer.

Due to the introduction of the residual module, ResNet50 can avoid the problem of performance degradation existing in deep network, so as to bring the advantages of deep CNNs in complex image feature extraction into play.

3.3. Knowledge Graph-Based Classifier Learning. Unlike traditional classifier based on fully connected layers, this paper

proposes a classifier learning module based on knowledge graph. Specifically, the classifier is set as the node of the knowledge graph; knowledge graph is used to explicitly encode semantic relationships between the targets, and then, GCN is applied to propagate defined prior knowledge between the classifier nodes to learn the transferable classifier. Targets from the source domain and corresponding targets from the target domain have known semantic relationships, which is the transferable knowledge. Knowledge graph can explicitly encode prior knowledge to improve the learning efficiency of the network. The reason for the definition of the nodes and edges of the knowledge graph is that the classifier directly corresponds to the target category. Compared with the feature, classifier has high-level semantic information. Therefore, the classifier is chosen as the node, and the semantic relationship between the targets is the edge. In order to learn the transferable classifier, it is necessary to construct the knowledge graph first and then use GCNs to update the classifier parameters on the defined knowledge graph, which are introduced separately as below.

3.3.1. Construction of Knowledge Graph. First, we construct the knowledge graph $\mathcal{G} = \{\mathbf{V}, \mathbf{A}\}$. Nodes $\mathbf{V}$ are the targets to be identified. In order to model the relationship between targets in the source domain and the target domain, the same targets in different domains are represented by different nodes; that is, $\mathbf{V} = \{\mathbf{V}_{SD}, \mathbf{V}_{TD}\}$, $|\mathbf{V}_{SD}| = |\mathbf{V}_{TD}| = N/2$, $|\cdot|$ denotes the number of nodes. The defined nodes in knowledge graph are total N , and the source and target domains contain half and half. Each node is represented by a feature vector, i.e., $\mathbf{x}_V \in \mathcal{R}^{N_{x_V} \times 1}$. $\mathbf{A} \in \mathcal{R}^{N \times N}$ is an adjacency matrix that encodes the relationship between nodes, in which semantic relationship $\epsilon_{\text{semantic}}$ between nodes is considered in this paper. For $\epsilon_{\text{semantic}}$, correlation coefficients are used to connect the same targets from different domains, as shown in Figure 3.

FIGURE 3: Semantic relation-based knowledge graph.

3.3.2. Learning of Classifier Parameters. After the knowledge graph is constructed, GCN is used to transfer the prior semantic knowledge between the nodes to realize the update of the node feature vectors of the knowledge graph. The nodes are initialized using classifier parameters obtained by pretraining of a fully connected classifier connected with ResNet50 in Section 3.2, namely

$$\mathbf{X}^0 = \mathbf{W}_{\text{rep}}^{\text{pre}}. \quad (2)$$

The first-order ChebNet [22] is used to realize the spectral graph convolution, namely

$$\mathbf{x}^{l+1} = \rho \left(\tilde{\mathbf{D}}^{-1/2} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-1/2} \mathbf{x}^l \Theta^l \right). \quad (3)$$

Among them, $\mathbf{x}^{l+1}$ and $\mathbf{x}^l$ are the features of graph convolution layer l and $l+1$, respectively, $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$ is the adjacency matrix added with the node's self-connection, $\tilde{\mathbf{D}}$ is the corresponding metric matrix, and Θ^l is the parameter of the graph convolution of the layer l . After several updates, the output features of the nodes are

$$\mathbf{X}^{\text{out}} \in \mathcal{R}^{N \times N_f}. \quad (4)$$

The output of the image recognition transfer learning method based on the knowledge graph thus is

$$\hat{\mathbf{Y}} = \mathbf{X}^{\text{out}} \mathbf{f} \in \mathcal{R}^{N \times 1}. \quad (5)$$

Among them, $\mathbf{f}$ is the image feature extracted by the representation learning module.

Supervised learning is adopted to learn the parameters of the both the representation learning and classifier learning module. Since target classification task is considered in this paper, cross-entropy loss is used to guide the update of network parameters, namely

$$\text{Loss} = - \sum_{i=1}^N p(c_i) \log q(c_i). \quad (6)$$

Among them, $p(c_i)$ is the probability that the current sample belongs to the class i given by the supervised label, and $q(c_i)$ is the probability that the current sample belongs

FIGURE 4: Ground physical environment.

TABLE 1: Training dataset (unit : piece).

Images from source domain			Images from target domain		
Port	490N	10N	Port	490N	10N
106	185	201	39	36	23

TABLE 2: Test dataset (unit : piece).

Images from source domain			Images from target domain		
Port	490N	10N	Port	490N	10N
10	21	23	517	269	54

to the class i judged by the trained classifier, which is obtained by normalizing the network output (5) by the softmax function, as shown in the following formula

$$q(c_i) = \frac{e^{\hat{Y}_i}}{\sum_{i=1}^N e^{\hat{Y}_i}}. \quad (7)$$

4. Results

Taking image recognition in on-orbit refueling task as an example, we have carried out simulation verification on the proposed algorithm. In order to verify the transfer performance of the proposed algorithm on few-shot data, a simulation environment and a ground physical environment are built, dataset is prepared, and comparative experiment with ResNet50-FC-based general CNN image recognition method is given.

4.1. Simulation Environment and Ground Physical Environment. Simulation environment is set as the source domain defined in the problem description, and the ground physical environment is set as the target domain in the problem description. The main components in the simulation environment are built based on the satellite board and key components in the ground physical environment. The ground physical environment is mainly composed of the target satellite board equipped with key components including 490N engine, 10N engine, and refueling port for on-orbit refueling. These components are designed according to real

FIGURE 5: Objects from source domain. (a) Port. (b) 490N, (c) 10N.

FIGURE 6: Objects from target domain. (a) Port. (b) 490N, (c) 10N.

on-orbit refueling task, as shown in Figure 4. The images collected from the source domain and target domain meet the data and task requirements in the problem description.

4.2. Algorithm Platform. The hardware system mainly comprises the following: a 64 bit 16 core CPU, a P40 graphics, and an IntelRealSense D435 camera. The software system includes Ubuntu16.04 LTS operating system and Pytorch deep learning platform running in Python 3.6.

4.3. Datasets. According to the limitations of the problem description on the training dataset and the test dataset, we prepared the corresponding datasets. First, the specific image recognition task is given; that is, the network needs to correctly identify three objectives including 490N engine, 10N engine, and the refueling port. Next, the dataset can be prepared according to the constraints of the task and problem description.

The training dataset contains 492 source domain images and 98 target domain images, a total of 590, of which the target domain images account for about 16%, which meets the requirement of obtaining a large amount of source domain labeled data and a small amount of target domain labeled data during training. The scale of each components is shown in Table 1, where port denotes refueling port.

The test dataset contains 54 source domain images and 840 target domain images, a total of 894, which meets the requirements for a large amount of target domain data during the test. The scale of each components is shown in Table 2.

FIGURE 7: Loss function value w.r.t training iteration.

Images of refueling port, 490N engine and 10N engine collected under the source domain are shown in Figures 5(a)–5(c).

Images of refueling port, 490N engine and 10N engine collected under the target domain are shown in Figures 6(a)–6(c).

Similarity can be seen from the objects rendered in source domain and target domain images, except the details in color and shape are variant.

4.4. Training and Test. The labeled training dataset is used to train the knowledge graph-based image recognition transfer

TABLE 3: Precision of training dataset.

Objects	Images from source domain			Images from target domain		
	Port	490N	10N	Port	490N	10N
ResNet50-FC	98.1%	100%	99.5%	17.9%	91.7%	60.9%
KGTL	98.1%	100%	100%	82.1%	86.1%	82.6%

TABLE 4: Recall of training dataset.

Objects	Images from source domain			Images from target domain		
	Port	490N	10N	Port	490N	10N
ResNet50-FC	99.0%	100%	99.0%	46.7%	48.5%	93.3%
KGTL	98.1%	100%	99.0%	82.1%	81.6%	90.5%

TABLE 5: Precision of test dataset.

Objects	Images from source domain			Images from target domain		
	Port	490N	10N	Port	490N	10N
ResNet50-FC	100%	100%	100%	16.8%	95.2%	68.5%
KGTL	100%	100%	100%	70.8%	92.7%	77.8%

TABLE 6: Recall of test dataset.

Objects	Images from source domain			Images from target domain		
	Port	490N	10N	Port	490N	10N
ResNet50-FC	100%	100%	100%	79.1%	37.3%	86.0%
KGTL	100%	100%	100%	92.7%	62.3%	95.5%

learning network, and the Adam method is used to optimize the network parameters. The training curve is shown in Figure 7. Horizontal axis of Figure 7 denotes the iteration epochs with the unit of times, and the vertical axis denotes the corresponding loss with no unit.

In the first 1000 iteration, the variance of loss is large. This is because the hyperparameter batchsize selected in the experiment is 5, and samples are quite different. In the next 1000 iterations, the loss function value gradually converges to stable minimum.

In order to illustrate the effectiveness of the proposed knowledge graph-based image recognition transfer learning method, the general CNN image recognition method without adding the knowledge graph is used as a comparison, namely, ResNet50-FC, which uses a fully connected layer to classify the extracted image features. Two indicators precision and recall are selected to evaluate the image recognition performance of both the methods.

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (8)$$

$$\text{Recall} = \frac{TP}{TP + FN}. \quad (9)$$

Among them, TP means that the classification result is correct, and the result is a positive sample; FP means that the classification result is wrong, and the result is a positive

sample; FN means that the classification result is wrong, and the result is a negative sample. The positive sample means the category considered when calculating the index, and the negative sample means all categories except for the positive sample.

ResNet50-FC and proposed image recognition transfer learning method based on knowledge graph (also written as KGTL) are both used to test the training dataset, and indicators precision and recall are obtained as shown in Tables 3 and 4. ResNet50-FC and proposed image recognition transfer learning method based on knowledge graph are also both used to test the test dataset, and indicators precision and recall are obtained as shown in Tables 5 and 6.

It can be seen that using ResNet50-FC and proposed KGTL methods, precision and recall for image recognition in the source domain are high, both higher than 98.1%. Compared to the ResNet50-FC, using the method proposed in this paper, it has a higher accuracy and recall rate for image recognition in the target domain. The average target recognition accuracy rate reaches 80.5%, and the average recall rate reaches 83.5%, which is higher than the method based on ResNet50-FC, and its average accuracy rate is 60.2%, and the average recall rate is 67.5%.

In order to compare the image recognition effects of the two methods of ResNet50-FC and KGTL on the target domain data, the test results of the target domain data in both the training dataset and the test dataset are given. Precision and recall are shown in Figures 8 and 9, respectively.

FIGURE 8: Precision of the image recognition from the target domain.

FIGURE 9: Recall of the image recognition from the target domain.

It can be seen that for the precision of the 490N engine in the training dataset, the recall of the 10N engine in the training dataset, and the precision index of the 490N engine in the test dataset, the KGTL method is slightly lower than ResNet50-FC, and details are listed in Tables 3–5. For other indicators, KGTL is significantly higher than ResNet50-FC, which shows the superiority of KGTL in target domain image recognition.

Furthermore, ResNet50-FC and the image recognition transfer learning method based on the knowledge graph proposed in this paper have high accuracy and recall rates for source domain data and tasks. This is because the training dataset contains a large number of source domain images (492 images), which proves that the neural network relies on a large number of effective samples, and the effectiveness

is reflected in the test dataset, and the training datasets satisfy the condition of independent and identical distribution.

ResNet50-FC has a low target recognition accuracy and recall rate in the target domain because the training dataset contains sparse target domain data and dense source domain data, and the network is only trained based on the data; therefore, the convolutional neural network suitable for source domain distribution cannot be transferred to the target domain distribution. Compared with ResNet-FC, the proposed knowledge graph-based image recognition transfer learning method performs better (red lines are almost above the blue lines in Figures 8 and 9), and we analyze the reason.

Except for data, the proposed method utilizes the prior semantic knowledge to train the network, which we think is the key to realize transfer learning. The explicit knowledge is another type of information and is more effective than the data for the training procedure of network and can be used as supplement of data. Knowledge graph is the suitable form to encode the knowledge and can be processed by GCNs. Specifically, the image recognition transfer learning method based on the knowledge graph introduces the knowledge graph and explicitly connects the similar categories in the source domain and the target domain according to the semantic relationship, which transfer the knowledge of easily learned classifiers in the source domain and makes it possible to strengthen learning procedure of the classifier in the target domain during training. The relevance of the classifier under different domains makes it easy to transfer the knowledge under the source domain to the target domain. Such a classifier is better than the fully connected layer classifier in ResNet50-FC.

5. Conclusion

In order to deal with the problem of image recognition transfer learning caused by inconsistent of the source and target domain data distribution, this paper proposes a knowledge graph-based image recognition transfer learning method by introducing knowledge graph, which connects classifier of the source and target domains by semantic relationship, thus strengthen the transfer and learning of the classifier in the target domain. Compared with the traditional CNN image recognition method based on the fully connected layer classifier, the proposed method has higher accuracy and recall rate for target recognition in the case of few-shot samples and improves the training efficiency and the generalization of the model. The proposed method is effective on reducing the number of real samples needed in the training process, adopting samples in a simulation environment that is easy to obtain with low cost. The preliminary results could be leveraged to the transfer of ground-learned image recognition models to real space manipulation in the future.

Data Availability

The data used to support the findings of this study are available from the author upon reasonable request.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Acknowledgments

This work was supported by the National Science Foundation (U20B2054).

References

- [1] Y. C. Xie, Y. Wang, A. Chen, and L. F. Li, "Leaning based on-orbit servicing manipulation technology of space robot," *Aero-space Contrd and Application*, vol. 45, no. 4, pp. 25–37, 2019.
- [2] Robotic Refueling Mission https://sspd.gsfc.nasa.gov/robotic_refueling_mission.html.
- [3] S. Rotenberger, D. Soohoo, and G. Abraham, "Orbital Express fluid transfer demonstration system," in *Sensors and Systems for Space Applications II*, vol. 6958, article 695808, International Society for Optics and Photonics, 2008.
- [4] H. Yingming, F. Shuangfei, F. Xiaopeng, W. Jingyang, and Z. Feng, "Vision perception technology for space manipulator on-orbit service operations," *Unmanned Systems Technology*, vol. 1, no. 1, pp. 54–65, 2018.
- [5] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *NIPS*, Curran Associates Inc., 2012.
- [6] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, <https://arxiv.org/abs/1409.1556>.
- [7] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, Columbus, USA, 2014.
- [8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, USA, 2016.
- [9] S. Thrun and L. Pratt, *Learning to Learn: Introduction and Overview*, Springer US, Boston, MA, 1998.
- [10] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.
- [11] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, "Deep domain confusion: maximizing for domain invariance," 2014, <https://arxiv.org/abs/1412.3474>.
- [12] M. Long and J. Wang, "Learning transferable features with deep adaptation networks," in *International conference on machine learning*, Lille, France, 2015.
- [13] Z. Cao, M. Long, J. Wang, and M. I. Jordan, "Partial transfer learning with selective adversarial networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, USA, 2018.
- [14] J. Zhang, Z. Ding, W. Li, and P. Ogunbona, "Importance weighted adversarial nets for partial domain adaption," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, USA, 2018.
- [15] Y. H. Tsai, W. C. Hung, S. Schuster, K. Sohn, M. H. Yang, and M. Chandraker, "Learning to adapt structured output space for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Salt Lake City, USA, 2018.
- [16] S. Ravi and H. Larochelle, *Optimization as a model for few-shot learning*, 2016.
- [17] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *International Conference on Machine Learning*, pp. 1126–1135, Sydney, Australia, 2017.
- [18] G. Koch, R. Zemel, and R. Salakhutdinov, "Siamese neural networks for one-shot image recognition," *ICML deep learning workshop*, vol. 2, 2015.
- [19] O. Vinyals, C. Blundell, T. Lillicrap, and D. Wierstra, "Matching networks for one shot learning," *Advances in neural information processing systems*, vol. 29, pp. 3630–3638, 2016, <https://arxiv.org/abs/1606.04080>.
- [20] J. Snell, K. Swersky, and R. S. Zemel, "Prototypical networks for few-shot learning," 2017, <https://arxiv.org/abs/1703.05175>.
- [21] R. Chen, T. Chen, X. Hui, H. Wu, G. Li, and L. Lin, "Knowledge graph transfer network for few-shot recognition," in *Proceedings of the AAAI Conference on Artificial Intelligence*, Hawaii, USA, 2019.
- [22] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, <https://arxiv.org/abs/1609.02907>.