Review of the Application of Ontology in the Field of Image Object Recognition

Zheyuan Ding
College of System Engineering
National University of Defense Technology
Changsha 410073, China
+86 18274959955
750539738@qq.com

Li Yao
College of System Engineering
National University of Defense Technology
Changsha 410073, China
+86 18274959955
750539738@qq.com

Bin Liu
College of System Engineering
National University of Defense Technology
Changsha 410073, China
+86 18274959955
750539738@qq.com

Junfeng Wu
College of System Engineering
National University of Defense Technology
Changsha 410073, China
+86 18274959955
750539738@qq.com

ABSTRACT

Image object recognition is an important research field in computer vision. It has a wide application prospect and practical significance in the information age. Although the current image recognition technology has achieved high accuracy in some tasks, the computer has many deficiencies in the automatic recognition of images such as fine-grained recognition, recognition of complex scenes. In these tasks, some issues exist like insufficient precision, complex high-level semantics which is difficult to identify and so on. This paper reviews the application of ontology in image object recognition. It is found that combining ontology knowledge model and traditional image recognition technology can improve recognition accuracy, enhance high-level semantic recognition ability, reduce the demand of the large number of training samples, and improve the scalability of the image recognition system. Otherwise, this paper also summarizes the frontier research of ontology applied in the field of image object recognition and the difficulties of deep integration of different technologies and ontology.

Keywords

image object recognition; ontology.

CCS Concept

Computing methodologies → Object recognition

1. INTRODUCTION

Image object recognition is to realize the recognition of the image by comparing the stored information with the current information [1]. Specifically, it uses image processing and recognition technology to identify the category of objects in the image. With the increasing demand for automatic recognition of computers, the recognition of deep semantics in complex images has become a frontier research direction. The premise of image object recognition is the description of image. The description is to obtain the abstract representation of the image features of the objects or the relationship between them using numbers or symbols.

Image object recognition technology has a wide range of applications, such as biomedicine, satellite remote sensing, robot vision, object tracking, unmanned vehicle driving, transportation, military, etc. [2].

From the perspective of image type, image object recognition includes professional domain image recognition (medical image recognition [3], remote sensing image recognition [4], etc.), general image recognition (mainly for the recognition of general object, including animal, car, face, etc.).

From the perspective of image recognition technology, image object recognition includes object recognition based on traditional methods, object recognition based on deep learning, etc. [5]. The image object recognition of the traditional method includes image preprocessing, image segmentation, image feature extraction, constructing a feature-based classifier, and identifying a object. Usually the image features necessary in the process require specialized knowledge to design the feature extractor to convert the original image into a suitable internal feature representation or feature vector. The image recognition method based on deep learning is a feature learning method, which can input original data to the machine, and then automatically find out the methods that need to be detected and classified, so that the machine can automatically learn to extract the image features [5].

At present, the recognition method based on deep learning has many outstanding achievements in the field of artificial intelligence, such as speech recognition and image recognition. However, deep learning also has obvious shortcomings: poorly interpretable, easy to fit [6], increased training difficulty and training time due to higher network complexity, and large number of labeled training samples [3]. At the same time, there are still some problems in the field of image recognition that cannot be solved by the current deep learning methods, such as zero-sample

ICCMS 2019, January 16–19, 2019, Melbourne, VIC, Australia

ACM ISBN 978-1-4503-6619-9/19/01…$15.00
DOI: https://doi.org/10.1145/3307363.3307387

image classification and multi-label image recognition in the case of missing tags.

In recent years, some studies have combined knowledge-based methods with data mining technology, and achieved good results by integrating human prior knowledge with traditional data mining technology [7, 8]. The ontology-based method is an important method in the field of knowledge engineering. Ontology is a consistent formal description of the concept of domain sharing, and the description of ontology is formal. Traditional ontology representation languages are mostly based on first-order logic or a combination of framework and first-order logic. The ontology not only describes the concept of abstraction, but also clearly defines and describes the relationship between concepts. Through the definition of formal axioms and SWRL rules, it has strong knowledge representation ability and logical reasoning ability. The review of this paper not only includes the improvement of image object recognition technology based on ontology, but also summarizes the application of ontology in different fields of image object recognition. Among them, the former focuses on how to better use the prior knowledge of the ontology to improve the recognition accuracy, while the latter focuses on how to use the logical reasoning ability of the ontology to deal with the semantic reasoning of multi-objective images. Through these researches, the advantages of combining qualitative and quantitative methods are explored in this paper. The advantages and disadvantages of ontology-based image object recognition are also summarized. Besides, the difficulties of combining different technology with ontology are analyzed. At last, we has made the prospect forecast of the application of ontology in the field of image object recognition.

According to the content of image, two parts are summarized as follows.

1 The application of ontology method in single object image recognition.

2 The application of ontology method in multi-objective image recognition.

2. Ontology in Image Object Recognition

There are two main ways to apply ontology. One is to use the conceptual hierarchy of ontology to combine with existing technologies, such as using the concept of ontology for indexing and labeling, and using concept hierarchy to construct different deep neural networks. Another application method is to use the inference ability of the ontology to acquire the attributes and features of multiple objects in the image through the prior art, and use ontology reasoning to obtain high-level semantics. In general single-objective image recognition more is the first application method. And in the multi-objective image recognition problem, there are more methods to apply ontology reasoning ability.

2.1 Ontology in Single Object Image Recognition

The single target for this section refers to the objects which belonging to the same category, or those objects which are similar between each other.

Antonio.M Rinaldi built a multimedia ontology for automatic image classification and annotation [9]. Experiments were performed on car image data.

The ontology constructed by the author not only includes the traditional concept description, but also includes the visual features of the image. The data attributes are saved to the ontology, and the association from the underlying visual features to the high-level semantics is realized.

For the fine-grained visual classification/recognition problem, Wang used the hierarchical concept system of the target to be identified to segment the original image into three different granularity images, respectively, using CNN for training, and obtaining different granularity features through three CNNs [7]. Multi-granularity feature descriptors are combined to train multi-granularity feature classifiers to finally identify the birds in the image. Experiments were performed on the CUB-200 dataset which contains only one bird and some natural background per image.

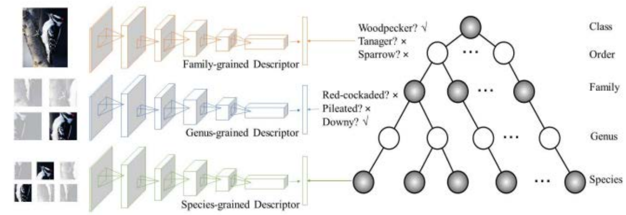

Figure 1. Overview of multiple granularity descriptors framework.

at different granularities. The labels used in each CNN are derived from the classification system concept on the right side of the figure. The top layer CNN corresponds to the image feature of the Family level while the middle layer CNN corresponds to the image feature of the Genus level. And the bottom layer CNN corresponds to the image feature of the Species level. Then, the three different granularity feature vectors are combined and used as an overall feature vector to train the SVM for classification. The experimental results have reached an average accuracy of more than 85%, and the performance is at the forefront of the current fine-grained identification tasks. At the same time, the author proposes that the next research is how to use the background information of the image to further improve the recognition accuracy.

Umang Gupta proposed an ontology-based deep transfer learning method for image classification [8]. Using the domain knowledge in the ontology, an algorithm combining CNN and ontology prior knowledge is proposed to infer the semantics of the abstract pattern. The pre-trained CNN extracts the image features, takes the image features as input, and trains the hierarchical concept of the ontology according to the defined objective function, so as to obtain the weight of each leaf node in the hierarchical concept system, and then according to the defined weight between the parent and child nodes. The weight of the top-level concept is obtained layer by layer, as the score of the judgment category. Experiments were carried out in three types of images including Fort, Tomb and Mosques, which proved that the method significantly exceeded the logic classifier and other transfer learning methods.

2.2 Ontology in Multi-Object Image Recognition

Multi-target image recognition here refers to the inclusion of multiple types of objects or images containing complex scenes and objects in each image. For images with multiple types of targets, the ontology can better describe and analyze the relationship between different targets, thereby improving the accuracy of target recognition and obtaining higher-level semantic recognition results.

Casey Breen and others used neural networks to identify objects in images [10]. They constructed semantic-based index structures according to the conceptual model of ontology, and obtain more detailed semantic interpretation and meaning in images by checking the relationship between objects. Taking sport as an example, they introduced the implementation process of the method.

The ontology constructed in this paper is obtained through sports terminology and domain experts. Each concept built into ontology contains a tag name and a feature vector. A feature vector is represented by a set of features and weights, each of which represents an object in the image. For example, according to football, goal, and green area, the image is classified as soccer. For images that contain only one football, they are classified as images containing football, rather than a more specific classification concept. In addition, the weight of each concept may not be equal. For example, in the concept of football, the weight of the goal is greater than the weight of the green area.

Another advantage of the system implemented by this method is that it has certain scalability. When a new concept is added to the ontology, the new neural network can be trained to acquire the objects existing in the new concept, and then according to the relationship between the objects. , classify images into different levels of concepts.

Hafidha Bouyerbou used an ontology-based semantic classification method to identify large-scale disasters occurring in satellite images [4]. They proposed an ontology-based semantic hierarchical classification method to carry out experiments with the 2010 Haitian earthquake-related fast bird satellite image data.

Tsai S F. proposed an ontology-based visual understanding method [11]. For a single image (image level) annotation task and multi-image (album level) event recognition, a conceptual architecture consisting of objects, scenes, events, etc. is constructed. It is verified that using the ontology method has better effects in visual recognition tasks than purely using low-level image features, and can better perform high-level semantic recognition.

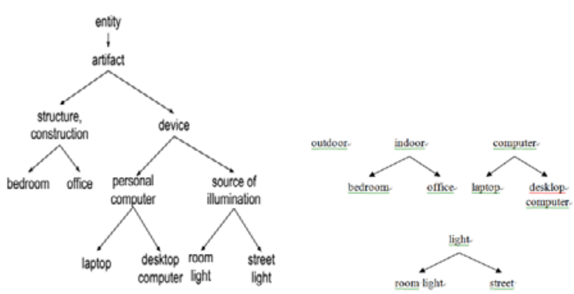

Figure 2. Concepts and relationship architecture.

The authors selected ten validation experiments involving the general concepts of objects and scenes, including indoor, outdoor, bedroom, office, light, room light, street light, computer, laptop, and desktop computer. The left side of Figure 2 shows the conceptual system that the author found in WordNet. Figure 2 shows the subclass relationship of related concepts in WordNet. Experiments were carried out in the collected images, and experiments were carried out with 20%, 35%, and 50% training images respectively. The results show that the ontology-based image semantic recognition, when the training data is small, still has a relatively stable recognition. result. And other common methods such as SVM, as the number of training data decreases, the recognition performance will plummet.

Ly N Q used a behavior ontology to identify behavior in video [12]. First, traditional image recognition technology are used to discriminate objects that exist in a multi-frame image, such as people and backpacks. These objects are defined in the ontology by four attributes. Any object has four attributes. These attributes are position, appearance, component, and name. The position includes two properties of location and size of an object. The appearance includes visual features such as shape, color, and texture. The component includes parts and belonging objects. For example, the legs are part of a human, and the luggage belongs to a human. All the connections of the object are defined in the constructed ontology, including three relationships and two conditions. Three relationships include *space relationships*, *time-space relationships*, and *action relationships*. The *space relationship* includes three types which is *near*, *far*, and *tangential*. The conditions include *position condition* and *action condition*. The position condition contains *in frame* and *out frame*.

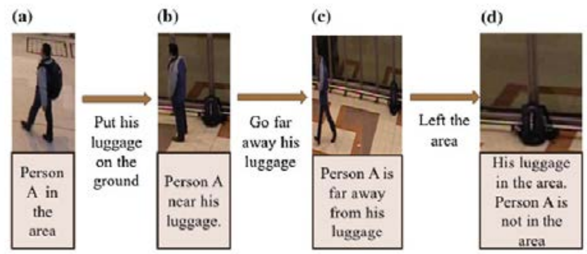

Figure 3. Four scene of behavior identification.

In the figure 3, the identification of left luggage behavior is taken as an example. There are scenes a, b, c, and d recognized by different frame images in the ontology. Each scene contains different space conditions and space relationships, and the objects of the scene have affiliations. There are successive relationships in different scenes. When a certain sequence of scenes is met, and the space conditions, space relationships, and object affiliations of each scene satisfy the conditions, high-level semantics can be inferred—person A left his packages.

Günther M proposed a model-based target recognition method for the construction of indoor semantic maps in robot vision, and the construction of ontology models for different furniture and flat areas in indoor 3D scenes [13]. It is divided into three steps as follows.

1) Original geometrical image detection, extracting the plane area;

2)Classifying the plane area, as an example of hypothesis generation, detecting the furniture object according to the plane area, and calculating the initial pose estimation of the furniture based on the plane area;

3) The final pose adjustment, as a verification example, and place the corresponding model in the scene.

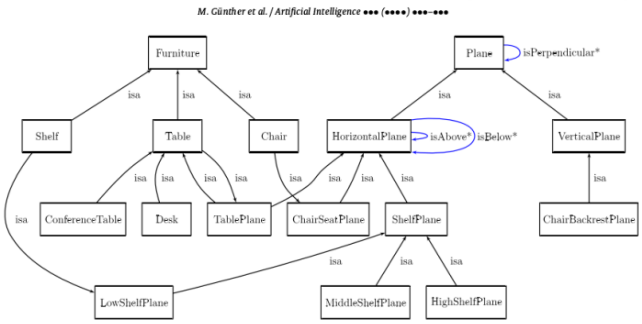

Figure 4. The concepts of indoor scene.

As shown in Fig. 4, the basic concept of the ontology constructed by the method is furniture and a plane area, and the plane part and the different plane area in the furniture are connected by object attributes, such as a seat surface and a table top are horizontal plane areas. At the same time, some SWRL rules are defined to describe how a region is constructed.

Low Shelf Plane(?p) ← Horizontal Plane(?p)∧ has Size(?p, ?s) ∧ swrlb:greater Than(?s, 0.01)∧ swrlb:less Than(?s, 0.5) ∧ has Pos Y(?p, ?h) ∧ swrlb:greater Than(?h, 0.08) ∧ swrlb:less Than(?h, 0.18)

As indicated by the above rule, a Low Shelf Plane is a Horizontal Plane with a size between 0.01 and 0.5 and a height PosY between 0.08 and 0.18.

When acquiring a 3D point cloud (a kind of three-dimensional image data), through surface reconstruction, multiple planar regions and related geometric features are obtained. Each planar region generates an instance in the ontology. And the pellet is used to obtain the corresponding furniture according to the SWRL rules, the definition of concepts, the properties of the different planar regions and the relationship between the regions. Then, after the attitude calculation, an indoor scene map is generated. The effectiveness of the method is verified by comparing the difference between the actual object and the generated graph.

3. CONCLUSION AND PROSPECT

Based on the above papers, we can see that although the ontology method is a traditional knowledge representation method, it still has a vigorous vitality in the latest machine vision tasks. In the different applications of image recognition, the use of ontology method has certain advantages, which can be summarized as the following four points.

1).The ontology method can improve the high-level semantic recognition ability for image recognition.

2).Applying domain prior knowledge to the recognition technology based on ontology can improve the accuracy of image recognition.

3).Based on the ontology method, the demand for a large number of samples in the training phase can be reduced.

4).The ontology method can improve the scalability of the image recognition system.

Generally, the image object recognition technology is based on the information of the actual image data for analysis and recognition, and is mostly obtained based on the data by a quantitative method. The ontology method is a qualitative knowledge representation method. It can be seen from the above referenced literatures that how to combine qualitative and quantitative, make full use of the mature identification technology and the human prior knowledge represented by the ontology, and apply it to various fields of image recognition is a study worthy of further consideration.

Otherwise, from the comparison of ontology-based recognition technology and other object recognition technologies, the ontology-based method relies on image features or other reference information extracted by image recognition technology, and is more suitable for high-level semantic reasoning. The advantage of ontology is that it has strong representation ability of abstract concepts. And it can get more information through the relationship between multiple objects in the image. Therefore, it is suitable for the application area which contains complex relationships between objects in an image such as robot vision, video behavior recognition.

Through the combination of ontology and deep learning methods, combining deep learning with prior knowledge of domain ontology is also a frontier research. But how to apply the logical reasoning ability of ontology to a deeper level of integration with deep learning methods is still difficult.

It is necessary to pay attention to the object recognition using the ontology technology. In other hand, the reasoning based on the ontology method always takes a long time. For the application needed fast response time, the professional inference engine design is generally required. In addition, complex scene modeling is difficult. And too many objects with complex relationships are difficult to represent with concise axioms or constraints, which further leads to longer inference time.

4. ACKNOWLEDGMENTS

The presented work is framed within the National Natural Science Foundation of China (No. 71371184).

5. REFERENCES

[1] Linan Fan, Xiaowei Han. 2007. Image processing and pattern recognition[M].

[2] XU Cai-Yun，Review of Research on Image Recognition Technology[J]，Computer Knowledge & Technology , 2013,10：2446-2447

[3] Esteva A, Kuprel B, Novoa R A, et al. Dermatologist-level classification of skin cancer with deep neural networks [J]. Nature, 2017, 542(7639): 115-118.

[4] Bouyerbou H, Bechkoum K, Benblidia N, et al. Ontology-based semantic classification of satellite images: Case of major disasters [C]. Proceedings of the Geoscience and Remote Sensing Symposium, 2014: 2347-2350.

[5] Yann LeCun , Yoshua Bengio & Geoffrey Hinton. 2015.Deep learning 5Nature volume 521, pages 436–444

[6] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions [J]. 2014, 1-9.

[7] Wang D, Shen Z, Shao J, et al. Multiple Granularity Descriptors for Fine-Grained Categorization [C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 2399-2406.

[8] Gupta U, Chaudhury S. Deep transfer learning with ontology for image classification [C]. Proceedings of the Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics, 2015: 1-4.

[9] Rinaldi A M. Using Multimedia Ontologies for Automatic Image Annotation and Classification [C]. Proceedings of the IEEE International Congress on Big Data, 2014: 242-249.

[10] Breen C, Khan L, Kumar A, et al. Ontology-based image classification using neural networks [C]. Proceedings of the ITCom 2002: The Convergence of Information Technologies and Communications, 2002.

[11] Tsai S F. Toward ontological visual understanding [J]. Dissertations & Theses - Gradworks, 2012.

[12] Ly N Q, Truong A M, Nguyen H V. Specific Behavior Recognition Based on Behavior Ontology [M]. Springer International Publishing, 2016.

[13] Günther M, Wiemann T, Albrecht S, et al. Model-based furniture recognition for building semantic object maps[J]. Artificial Intelligence, 2015, 247:336-351.