Proceeding of the IEEE
International Conference on Robotics and Biomimetics
Dali, China, December 2019

Robotic Scene Understanding by Using a Dictionary

Fujian Yan, Saideep Nannapaneni, and Hongsheng He

***Abstract*—Scene understanding is a fundamental task for intelligent robots, especially in human-robot interaction. It is challenging due to the complexity of the human environment. In this paper, the proposed method integrates semantic analysis with object detection that enables robots to perceive scenes and deeply understands working environments. The model can extract deterministic entities of objects by analyzing their dictionary definitions. Therefore, robots can understand a scene at the object-level. These deterministic entities include the category, function, property, and composition of objects, and they can be used to generate feedback on how much robots can understand a scene by describing it in natural language. The feedback on how well robots understand the working space is an essential aspect to eliminate confusion during human-robot interactions. The experiment part of this paper discussed the applicability of the proposed method on robots.**

***Index Terms*—autonomous robots, robot reasoning, semantic scene understanding, cognitive human-robot interaction**

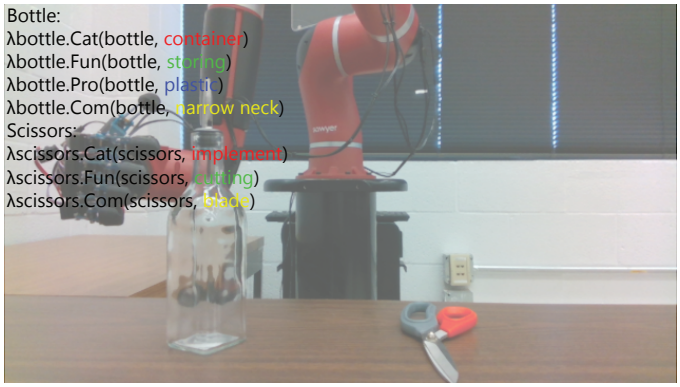

Figure 1. Object semantic comprehension applied to a bottle and the scissors in a scene. The semantics comprehension of both a bottle and the scissors are extracted. Logic expressions, which are formed based on those extracted entities, are displayed.

I. Introduction

The environment of human beings is more dynamic, complicated, and disordered comparing with the well-engineered environment of robots [1]. Previous literatures [2]–[6] discussed the importance of perception systems. There are still challenges for intelligent robots to perform effectively and efficiently in a more natural environment, even though the technique of computer vision is developed. The different methods of scene semantic understanding of robots are discussed in [5]–[7]. The semantic perception method of per-pixel polarization information, which is extracted from monocular RGB images, is proposed in [5]. These papers [8], [9] proposed general frameworks for scene understanding, which focus on binocular vergence, localization, and simultaneously identify multiple objects. Earlier research [10] integrated object recognition with an exhaustive knowledge base, which associated with each object, to understand an environment. By applying this knowledge base, the robot can understand what these objects are and how to manipulate these objects. Besides perceiving objects in a scene, robots need to understand the scene to interact with human beings.

Fujian Yan and Honegsheng He are with Department of Electrical Engineer and Computer Science, Wichita State University, Wichita, KS, 67260, USA fxyan@ wichita.edu,hongsheng.he@wichita.edu.

Saideep Nannapaneni is with Department of Industrial, Systems and Manufacturing Engineering, Wichita State University, Wichita, KS, 67260 USA Saideep.Nannapaneni@wichita.edu

*Correspondence should be addressed to Hongsheng He, hongsheng.he@wichita.edu.

Robots should have the ability to adapt to the uncertainties and complexities of dynamic environments. These abilities of robots can ensure the efficiency and safety of robotic applications. Therefore, a dominant perception, which has an autonomous cognition system, is urgent to fulfill these demands. The problem of scene understanding can be considered as semantic comprehension on every object in a scene. It is not enough for scene understanding in dynamic environments with only one perception system. In a dynamic environment, the complexity of the environment increases exponentially, so robot operators cannot hard-code every function that controls robots. The possibility of robots working in unfaced situations is dramatically high in real-world robotic applications. Whether a robot can perform effectively and efficiently is depending on the ability to solve problems in a dynamic environment.

Some earlier papers attempted to solve robotic scene understanding and object comprehension, and these methods depended on the pre-constructed knowledge base. These methods worked well when the environment is known or not changing. Take robotic arms working in logistics or an assembly line as an example; the shape of packages or dimensions of assembled parts are pre-defined. Robot operators can design path plans or manipulation strategies in advance. Robots can finish tasks in a fast, outstanding way due to their characters in less error rate, and needless for a rest. These systems are vulnerable when facing the undefined shape of packages or parts with unclear dimensions. It is also

978-1-7281-6321-5/19//$31.00 ©2019 IEEE

essential to ensure safety, effective communication between robots and humans when both work in a shared environment. Natural language is the most efficient way to communicate, but it is also ambiguous. Furthermore, as the development of robots, there is an increasing trend of robots that work in daily lives to interact with regular people who have less experience in operating a robot. As a result of that, an effective way that can show how much robots can understand a scene is necessary as well.

In this paper, we proposed a method that can enhance the adaptivity of robots in a dynamic environment by semantically analyzing dictionary definitions of objects in a scene. Our approach enables robots to cognize objects by learning dictionary definitions of objects like a child to learn new knowledge. We propose a method that can not only percept a working environment but also understand the semantics of these objects in the environment based on their dictionary definitions. Objects in the working environment are recognized with the Faster-R-CNN [11]. Descriptions of learned knowledge by robots are formatted based on predefined context-free grammar templates according to these extracted components of dictionary definitions. Explanations with details are provided in the following paragraphs. The main contribution of this work is to assist robots in understanding unfamiliar objects in natural environments by understanding a scene at object-level by referring dictionary definitions, and feedback on how much robots have been learned by descriptions. This paper is structured as follows: The description of object semantics and the details of the semantic model are discussed in section 2. Experiment results and metric evaluations are shown in Section 3, followed by a conclusion in Section 4.

II. Object-level Scene Semantics

This paper proposes an approach that enables robots to understand adaptively their working environments by referring to a dictionary. It has two major components, which can lead to robots to understand objects. These two components are object detection and semantic comprehension of objects. The following parts will discuss details about these two components. Logic references, which are formed by entities that are extracted from dictionary definitions, are also discussed. Fig. 2 illustrates an example of the working flow of the proposed method. In the illustration, according to these detected objects in the scene, the system will consult dictionary definitions for each detected object. By analyzing these definitions of detected objects, general information of each object is extracted. Logic expressions and descriptions are constructed based on these extracted entities.

A. Object Detection and Recognition

In this section, we discuss the method to detect objects in the working space. Before robots can understand the semantics of objects, these objects in the working space are needed to be recognized. We use the Faster-R-CNN [11] to achieve object detection because the Faster-R-CNN has a great balance between computation speed and accuracy [12]. The object detection model, which is used in this paper, contains two parts, and the first one is the regional proposal network. The second one is object detection. The first part takes input of an uncertain size image, the output of that is a group of region proposals with scores that measure the differences between objects and background. The object detection part is processed with a convolution neural network (CNN). After the region of interest (ROI) is pooled, feature maps are extracted with CNN. Two fully connected layers are used to classify the category of object and the regression of the bounding box.

B. Part of Speech and Entity Recognition of Dictionary Items

Both the part-of-speech and the word dependency are other ways to explain the meaning of a sentence besides its semantics. Although there is a difference between entity recognition and part-of-speech taggers, it is often useful to add a sequence of part-of-speech tags with the word IDs for entity recognition.

Assuming the dictionary definition of an object S , it is tokenized into a series of words $\mathrm{s_1 \cdots s_n}$, and a series of part-of-speech tags $\mathrm{t_1 \cdots t_n}$. The most reasonable part-of-speech tagging is the series that achieves the highest likelihood of a given word $\mathrm{s_i}$ and a corresponding tagger $\mathrm{t_i}$

$$\hat{\mathrm{t}}_1^{\mathrm{n}} = \underset{\mathrm{t}_1^{\mathrm{n}}}{\mathrm{argmax}} \mathrm{P}(\mathrm{t}_1^{\mathrm{n}} | \mathrm{s}_1^{\mathrm{n}}) \tag{1}$$

where $\hat{t}_1^n$ denotes the most reasonable part-of-speech tag with given n numbers of tokes s_1^n.

Because object's dictionary definitions are more organized compared with random sentence structures, a suitable representation for the entity model is

$$\mathrm{P}(\widetilde{\alpha}) = \max(\mathrm{P}(\mathrm{t}_1^{\mathrm{n}} | \mathrm{s}_1^{\mathrm{n}}) \mathrm{P}(\mathrm{k}_1^{\mathrm{n}} | \mathrm{s}_1^{\mathrm{n}})) \mathrm{P}(\alpha) \tag{2}$$

where $\mathrm{k}_1^{\mathrm{n}}$ is the key indicators appeared in the dictionary definition. $\mathrm{t}_1^{\mathrm{n}}$ is the part-of-speech tags sequence for the given series tokens $\mathrm{s}_1^{\mathrm{n}}$, $\mathrm{P}(\tilde{\alpha})$ is the resolved probability of the entity. $\mathrm{P}(\alpha)$ is SoftMax function [13] result of the semantic analysis model

$$\mathrm{P}(\alpha) = \frac{\mathrm{e}^{\alpha_{\mathrm{i}}}}{\Sigma_{\mathrm{i}} \mathrm{e}^{\alpha_{\mathrm{i}}}} \tag{3}$$

where $\alpha \in \{\mathrm{C}, \mathrm{F}, \mathrm{P}, \mathrm{COM}\}$.

There are phrases in dictionary definitions, which are significant, that can be used as indicators in the sentence. Such as "used for," "used to," or "used as." Verbs, after these phrases are open served as the function of an object. In some special cases, some nouns are used to describe the function of the object. These situations mostly happened when the indicated phrases are "used as." $\exists \mathrm{x}, \mathrm{Labeled(x)} \wedge \mathrm{W(x)} \rightarrow \mathrm{Entity(x)}$, where $\mathrm{W(x)}$ means the word after the indicator phrases. Besides the indicator phrases mentioned above, some phrases such as "consist of" or "belong to" are indicators for compositions of an object.

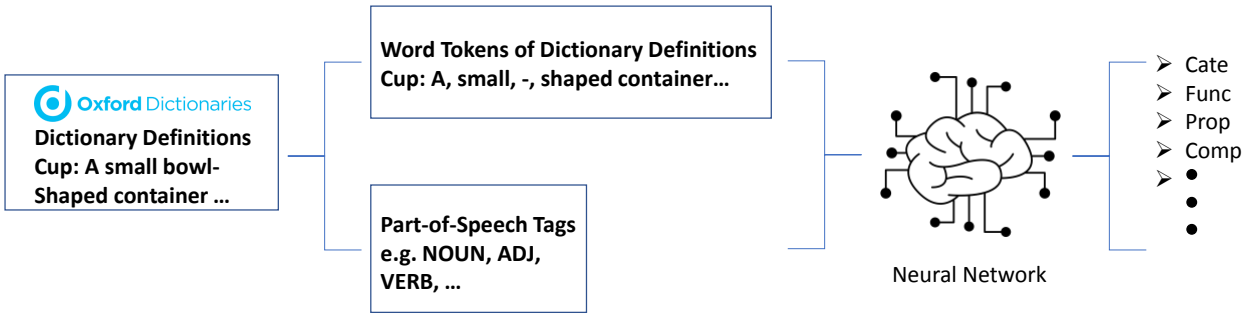

Figure 2. This is a sample, which can show how our approach enables a robot to understand unknown objects in a scene. The first step is to detect objects and recognize them. Then, definitions of these objects can be obtained from the Oxford English Dictionary. After that, important elements, which can describe objects, are extracted.

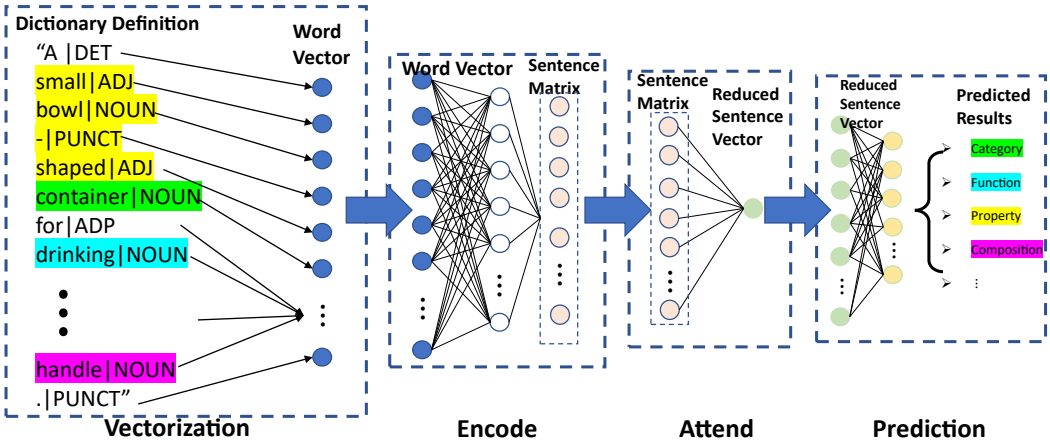

Figure 3. Network architecture. The color codes are used to distinguish different elements. The green, cyan, yellow, and purple is used to illustrate the category, function, property, and composition separately.

The following Table I indicates a list of indicator phrases in dictionary definitions that are based on empirical experience. There are 358 words sampled. Two hundred and seventy-six of the sampled words are used in the transmission area, and 82 of the sampled are words used in the tool area. The apparent frequency in the Table I is the appearance of the indicator phrase.

Table I
INDICATING PHRASES BASED ON EMPIRICAL EXPERIENCE.

Entity Category	Indicator Phrase	Apparent Frequency
Function	Used for/as/to	42
Function	Used as	4
Function	Used to	11
Composition	Consist of	13

Part-of-speech tagging can be evaluated based on rules or stochastic. Rules-based part-of-speech tagging is redundancy. However, the structure for dictionary definitions is comparatively identical. There are rules generated based on part-of-speech and dependency parse to reduce the confusion. $\exists x, \mathrm{Labeled}(x) \wedge \mathrm{POS}(x) \wedge \mathrm{DP}(x) \rightarrow \mathrm{Entity}(x)$, which served as a verify function to ensure the accuracy of the semantic analysis model. $\mathrm{Labeled}(x)$ indicated that the word in the dictionary definition is labeled according to the semantic model. POS(x) is the part-of-speech of each labeled words. In this approach, the part-of-speech for the category and the composition are nouns, the part-of-speech for properties are adjective or adverb, and functions can be noun or verb. DP(X) is used to distinguish the dependency of given words. Dependencies for the category are "root". Dependencies for properties are "amod" and "advcl," which are adjective or adverbial clausal modifier. Dependencies for both functions and compositions are variated, so they are not considered in this rule.

There are also miss-labeled words which can be eliminated by their dependencies. For instance, the verbs with dependency such as "advcl," which is adjective modifier, "acl," which is clausal modifier of noun, or "advcl," which is adverbial clausal modifier should not considered as the function of a object. $\exists x, \mathrm{Labeled}(x) \wedge \mathrm{DP}(x) \rightarrow \neg\mathrm{Entity}(x)$, which is the rule for the above expression.

C. Object Semantic Recognition

There is a hierarchy relation between dictionary definitions and descriptions of objects. Assuming an object's dictionary definition is a root, the object's descriptions are leaves.

This structure can form a hierarchy tree for semantic comprehension. Four necessary elements are used to describe an object for robot applications, which are the category, function, property, and composition of an object. By parsing this hierarchy tree, robots can cognize objects in detail. The entity parsing model is based on the neural network.

The network has four layers. Fig. 3 illustrates the structure of the neural network model. Plain words cannot be directly used as inputs to the neural network model. Because of that, vectorization is needed to embed word tokens of dictionary definitions into vectors. Besides the entities labeling, part-of-speech tags are also useful to improve the accuracy of the prediction model. After a 128-bit vector, an additional 10-bit vector is added as a part-of-speech label. After each word token with a part-of-speech label has been converted to a vector, a dictionary definition can be denoted as a $\mathrm{m}\times\mathrm{n}$ matrix. Each row in this matrix stands for a word token of the dictionary definition. An attend process is needed to feed the matrix results into a feed-forward prediction layer. The trained model aims to carry out the semantics comprehension tasks on dictionary definitions with paraphrasing the category, function, property, composition of an object. The annotation of the data is done manually.

D. Logic Representation of Object Semantic

Robots can use these semantic analysis results to infer and reason surrounding environments. Robots can understand objects in a scene based on their categories, functions, property, and compositions, which are extracted from their dictionary definitions. The general form of the scene understanding is

$$\lambda\Omega.\Gamma(\omega,\gamma_1,\cdots,\gamma_\mathrm{n}) \tag{4}$$

where Ω is the variable that denotes the name of the expression. $\Gamma(\omega,\gamma_1,\cdots,\gamma_\mathrm{n})$ denotes the expression itself. λ is used to denote binding a variable in a function. The following logic tuples are formulated when the definitions of objects in a scene are extracted

$$\begin{aligned}
&\lambda\mathrm{Object.Cat}(\mathrm{object},\mathrm{c}_1,\cdots,\mathrm{c}_\mathrm{n})\\
&\lambda\mathrm{Object.Fun}(\mathrm{object},\mathrm{f}_1,\cdots,\mathrm{f}_\mathrm{n})\\
&\lambda\mathrm{Object.Pro}(\mathrm{object},\mathrm{p}_1,\cdots,\mathrm{p}_\mathrm{n})\\
&\lambda\mathrm{Object.Com}(\mathrm{object},\mathrm{com}_1,\cdots,\mathrm{com}_\mathrm{n})
\end{aligned} \tag{5}$$

where the Object is the name of this expression. The second object denotes the name of each recognized object. Cat, Fun, Pro, and Com stands for category, function, property, and composition of each object. $\mathrm{c},\mathrm{f},\mathrm{p},\mathrm{and}\,\mathrm{com}$ are used to denote category, function, property, and composition in details.

These logic tuples are used to eliminate ambiguous communication between robots and human beings during a shared environment. There is information omitted in communication because the pre-known knowledge of human beings is extraordinarily more extensive than robots. Some commands, which are given by robot operators, are ambiguous. We use these logic tuples, which are constructed by extracted entities, to complete a temporal knowledge base to reason ambiguous commands. Descriptions that are constructed by context-free grammar template to give feedback on robotic comprehension.

III. Experimental Evaluation

A. Experiment Setup

In this experiment, a Sawyer robotic arm was used, which has seven degrees of freedom. An AR-10 humanoid hand was used, and it has ten degrees of freedom. An Intel-RealSence R435 has been used for RGB images acquisition. The definitions of testing objects were from the Oxford Online English Dictionary [14]. The ROS [15] system was used to control the robot. The model was trained on Intel Core i7-5930 processors and a TITAN X GPU.

B. Quantitative Experiment Results

Table II
THE PRECISION, RECALL, AND FSCORE SHOWED IN THE TABLE.

	Proposed Method	**SS-CNN[7]**
Precision	82.19%	-
Recall	78.95%	-
Fscore	**80.51%**	41.3%

The Table II showed the evaluation results on the dictionary semantic analysis mode. The model is trained based on the collected dataset of dictionary definitions from the Oxford English Dictionary. There are 243 definitions of daily used objects collected. The fscore of the model evaluated on the dataset of the Oxford English Dictionary is 80.51%. A comparison study has been done between SS-CNN model [7], which is another method used to understand the semantics of a scene by using RGB-D images as input. In contrast to the SS-CNN model is a classifier, which can classify six classes, the proposed method can understand a scene in object-level by referring to a dictionary if objects are recognizable. The proposed method integrates the advantages of both object detection and entities extraction from dictionary definitions to improve semantic understanding.

There is a comparison experiment on the model performance of different groups of objects. There are five different groups of objects evaluated, which are the cutlery, fruit, tool, electronics, and transport. The details of each group is listed in Table IV. The definitions for the tool, the cutlery, and the electronics are focused on the functions of objects. The definition of fruits is focused on property. The definition of transport is focused on composition.

The bar-chart 4 showed the evaluation of the performance of the model on different groups of objects. The performance is better than the whole model evaluation due to the definitions of the first group, the second group and the fourth group are focused on functional description. The frequency of word phrases such as "used for," "used as" and "used to" is

Table III
THERE ARE SOME RESULTS OF THE SEMANTIC ANALYSIS ON DICTIONARY DEFINITIONS OF DAILY USED OBJECTS.

Object	Definition from Oxford [14]	Description
Pool	An artificial pool for swimming in.	A pool is a pool. It is used for swimming. It is artificial.
Screwdriver	A tool with a flattened or cross-shaped tip that fits into the head of a screw to turn it.	A screwdriver is a tool. It is used to turn. It is cross-shaped. It consists of a tip.
Tray	A flat, shallow container with a raised rim, typically used for carrying the food and drink, or for holding small items or loose material.	A tray is a container. It is used for carrying and holding. It is flat, shallow, and raised.
Skillet	A small metal cooking pot with a long handle, typically having legs.	A skillet is a pot. It is used for cooking. It is small and long. It consists of metal.
Washer	A person or device that washes something.	A washer is a device. It is used to wash.
Toothbrush	A small brush with a long handle, used for cleaning the teeth.	A toothbrush is a brush. It is used to clean
Carrot	A tapering orange-colored root eaten as vegetable.	A carrot is a root. It is used as a vegetable. It is tapering and orange-colored.
Basketball	The inflated ball used in basketball.	A basketball is a ball. It is inflated.
Keyboard	A panel of keys that operate a computer or typewriter.	A keyboard is a panel. It is used to operate.
Cup	A small bowl-shaped container for drinking from, typically having a handle.	A cup is a container. It is used for drinking. It is small and bowl-shaped. It consists of a handle.
Grinder	A machine used for grinding something.	A grinder is a machine. It is used for grinding.
Hacksaw	A saw with a narrow fine-toothed blade set in a frame, used especially for cutting metal.	A hacksaw is a saw or a blade. It is used for cutting. It consists of metal. It is narrow and fine-toothed.
Pitchfork	A farm tool with a long handle and two sharp metal prongs, used for lifting hay.	A pitchfork is a tool. It is used for lifting. It consists of a handle, metal, and prongs. It is farm, long, and sharp.

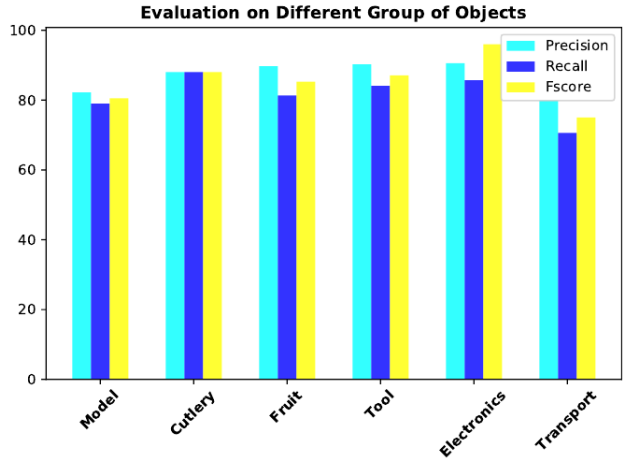

Figure 4. The figure showed the performance of the semantic model on different groups of objects.

Table IV
A LIST OF DAILY-USED OBJECTS THAT ARE USED IN THE COMPARISON EXPERIMENT IN PERFORMANCE. THE DICTIONARY DEFINITIONS OF EACH OBJECTS ARE EXTRACTED FROM OXFORD ENGLISH DICTIONARY[14].

Cutlery	Fruit	Tool	Electronics	Transport
Spoon	Apple	Hammer	Oven	Car
Knife	Pear	Tape	Keyboard	Train
Fork	Banana	Shovel	Mouse	Airplane
Chopstick	Mango	Screwdriver	Computer	Jeep
Tongs	Watermelon	Rake	Television	Truck
Toothpick	Melon	Spanner	Microwave	Boat
Skillet	Peach	Pilers	Lamp	Ship
Pepper Mill	Papaya	Pincer	Heater	Rocket
Ladle	Pineapple	Seal	Monitor	Motorcycle
Slotted Spoon	Grapefruit	Saw	Radio	Bicycle

high. To understand the target object's function is an essential aspect of robotic scene understanding. The performance of the second group is better than the whole model evaluation. The property is also a vital aspect for robots to understand its environment. The performance of the fifth group is lower than the model evaluation. The reason for that is the structure of transport's definition is unorganized, which means that the frequency of word phrases such as "with," "consist of," "belong to" is low. More data is needed to increase prediction accuracy.

Table III is the sample of the object's dictionary definition results. These samples are selected from daily used objects. The first column is the object. The second column is the definition of the object. The third column is the controlled logic language result.

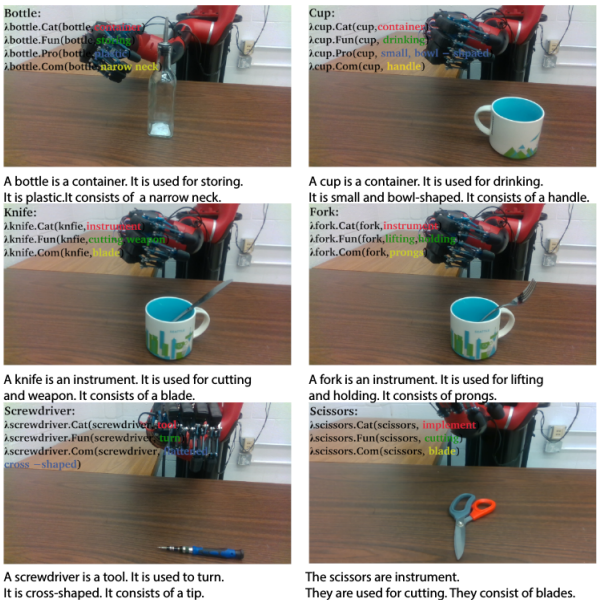

Figure 5. There are some experiment results shown. We applied the semantic comprehension model on scenes that had a bottle, a cup, a fork, a knife, a screwdriver, and the scissors.

C. Robot Semantic Comprehension on the Real Robot

In Fig. 5, some samples of experiment results on a real robot are shown. These results contained logic tuples and robot semantic comprehension. For logic tuples, the category of objects, the function of objects, the composition of objects, and the property of objects are displayed in red, green, blue, and yellow separately. There are six objects used in this experiment, and the definition of these six objects is from the Oxford English dictionary. The following experiment displayed the results of how the whole system works. All objects are successfully detected, and semantics are analyzed correctly. These descriptions are generated based on predefined context-free grammar templates. Plural nouns such as scissors, hedge trimmers, and grass shears, used pronoun accordingly.

IV. Conclusion

In this paper, we proposed an approach that enables robots to understand a shared environment by searching the dictionary to find definitions of each object in a scene. This approach can increase the adaptivity of robots in a less experienced or new environment, thereby, to improve the performance of a robot. The proposed method integrated the traditional perception method, which is object detection, and semantic analysis on text to improve the performance efficiency of robots. It can also tighten the interaction between robots and human beings by describing a scene that is comprehended by robots. The method proposed in this paper can eliminate ambiguous communications during a shared working space, thereby, to make robots effectively.

V. Discussion and Future Work

In this work, an Intel D435 camera is used to collect visual information of the scene. The method takes RGB images and dictionary definitions of detected objects as inputs. Other devices, which can collect images, are also theoretically suitable for the proposed method. Because of the limitations of devices, images, which are taken by other devices, are not tested. We are currently working on a real-time scene comprehension algorithm that enables robots to give feedback on their understanding of the scene.

References

[1] B. Rosman and S. Ramamoorthy, "Learning spatial relationships between objects," *The International Journal of Robotics Research*, vol. 30, no. 11, pp. 1328–1342, 2011. 1

[2] T. Fong, I. Nourbakhsh, and K. Dautenhahn, "A survey of socially interactive robots," *Robotics and autonomous systems*, vol. 42, no. 3-4, pp. 143–166, 2003. 1

[3] C. Bartneck and J. Forlizzi, "A design-centred framework for social human-robot interaction," in *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No. 04TH8759)*. IEEE, 2004, pp. 591–594. 1

[4] C. L. Breazeal, *Designing sociable robots*. MIT press, 2002. 1

[5] K. Yang, L. M. Bergasa, E. Romera, X. Huang, and K. Wang, "Predicting polarization beyond semantics for wearable robotics," in *2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids)*. IEEE, 2018, pp. 96–103. 1

[6] K. Yang, X. Hu, L. M. Bergasa, E. Romera, and K. Wang, "Pass: Panoramic annular semantic segmentation," *IEEE Transactions on Intelligent Transportation Systems*, 2019. 1

[7] Y. Liao, S. Kodagoda, Y. Wang, L. Shi, and Y. Liu, "Understand scene categories by objects: A semantic regularized scene classifier using convolutional neural networks," in *2016 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2016, pp. 2318–2325. 1, 4

[8] G. A. Camarasa and J. P. Siebert, "A hierarchy of visual behaviours in an active binocular robot head," 2009. 1

[9] G. Aragon-Camarasa, H. Fattah, and J. P. Siebert, "Towards a unified visual framework in a binocular active robot vision system," *Robotics and Autonomous Systems*, vol. 58, no. 3, pp. 276–286, 2010. 1

[10] M. Tenorth, L. Kunze, D. Jain, and M. Beetz, "Knowrob-map-knowledge-linked semantic object maps," in *Humanoid Robots (Humanoids), 2010 10th IEEE-RAS International Conference on*. IEEE, 2010, pp. 430–435. 1

[11] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99. 2

[12] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017. 2

[13] G. E. Hinton and R. R. Salakhutdinov, "Replicated softmax: an undirected topic model," in *Advances in neural information processing systems*, 2009, pp. 1607–1614. 2

[14] O. U. Press, *Oxford Dictionary of English*, O. U. Press, Ed. Oxford University Press, 2010. 4, 5

[15] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "Ros: an open-source robot operating system," in *ICRA workshop on open source software*, vol. 3, no. 3.2. Kobe, Japan, 2009, p. 5. 4