

Visual Question Generation for Class Acquisition of Unknown Objects

Kohei Uehara¹[0000–0003–0066–319X], Antonio Tejero-De-Pablos¹,
Yoshitaka Ushiku¹, and Tatsuya Harada^{1,2}

¹ The University of Tokyo
{uehara, antonio-o-t, ushiku, harada}@mi.t.u-tokyo.ac.jp
² RIKEN

Abstract. Traditional image recognition methods only consider objects belonging to already learned classes. However, since training a recognition model with every object class in the world is unfeasible, a way of getting information on unknown objects (i.e., objects whose class has not been learned) is necessary. A way for an image recognition system to learn new classes could be asking a human about objects that are unknown. In this paper, we propose a method for generating questions about unknown objects in an image, as means to get information about classes that have not been learned. Our method consists of a module for proposing objects, a module for identifying unknown objects, and a module for generating questions about unknown objects. The experimental results via human evaluation show that our method can successfully get information about unknown objects in an image dataset. Our code and dataset are available at <https://github.com/mil-tokyo/vqg-unknown>

Keywords: Visual question generation, Unknown object recognition, Unknown object class acquisition, Real world recognition

1 Introduction

In recent years, in large-scale image classification tasks, image classifiers with deep convolutional neural networks (CNN) have achieved accuracies equivalent to humans [22, 7]. The recognition capabilities of these methods are limited by the object classes included in the training data. However, for an image recognition system running in the real world, for example a robot, considering all existing object classes in the world during training is unfeasible. If such a robot was able to ask for information about objects it cannot recognize, the robot would not have to learn all classes in advance. In this paper, we define an unknown object as an object belonging to a class not included in the training data. In order to acquire knowledge about the unknown object class, the most reliable way is to obtain information directly from humans. For example, the robot can present an image to a human and ask them to annotate the class of an object, as in active learning [12]. When the class is unknown, selecting the appropriate object and generating a suitable question about it is a challenging problem, and has not

