Real-Time Recognition of Signboards with Mobile Device using Deep Learning for Information Identification Support System

Shigeo Kitamura
Kansai University
Osaka, Japan
k403662@kansai-u.ac.jp

Kota Kita
Kansai University
Osaka, Japan
k097619@kansai-u.ac.jp

Mitsunori Matsushita
Kansai University
Osaka, Japan
mat@res.kutc.kansai-u.ac.jp

ABSTRACT

In this paper, a framework that uses deep learning on a server to recognize signboards in streets with mobile devices is proposed. The proposed framework enables a user to determine the type of shops in his/her location. Our experimental results revealed that the proposed framework recognized signboards with an 86% accuracy within 1 second.

CCS CONCEPTS

• **Human-centered computing** → **User interface programming**;

KEYWORDS

mobile device, augmented reality, image recognition, deep learning

ACM Reference Format:
Shigeo Kitamura, Kota Kita, and Mitsunori Matsushita. 2018. Real-Time Recognition of Signboards with Mobile Device using Deep Learning for Information Identification Support System. In *Symposium on Spatial User Interaction (SUI '18), October 13–14, 2018, Berlin, Germany.* ACM, New York, NY, USA, 1 page. https://doi.org/10.1145/3267782.3274674

1 INTRODUCTION

When people search for stores in cities, signboards are one of the indicators used in discriminating between stores. However, if the signboard is written in a language that the user does not understand, it is difficult for users to determine the type of the store. Previously, we proposed methods to help people find the right signboard by using augmented reality in an artificial reality environment [1]. He et al. recognized the letters written on the signboard by using CNNs and RNNs [2]. However, even if the name of store is translated, it is not always possible to determine the type of store. In this paper, we propose a framework to acquire store information from images in real-time.

2 DEVELOPMENT

We developed an API on a server equipped with a GPU that recognizes signboards from the images and returns the results in JSON format. The API recognizes signboards in two stages. First, the API detects the bounding box of signboards from the image using YOLO [3]. Next, the signboards are cropped out by YOLO and classified using VGG16 [4]. This method enables recognition using a small amount of training data. In order to confirm that signboards can be recognized instantly, we collected 650 images from around 15 stores in one area of the shopping street of Takatsuki city, Osaka prefecture, and annotated the signboard area in those images. Then, we used 585 pictures as training data and 65 pictures as test data using YOLO. In addition, we collected 100 signboard pictures from 15 stores in the area. For each store, we selected 50 pictures for training data, 25 pictures for validation data, and 25 pictures for test data. Training was performed using VGG16. The accuracy of signboard recognition using test data was 95%. Subsequently, verification was performed using field experiments.

Table 1: Results of the experiment

Signboard	(a)	(b)	Signboard	(a)	(b)
	9	9		10	10
	9	9		10	10
	10	10		9	9
	10	10		8	3
	10	7		10	10
	10	10		10	10
	8	8		9	9
	10	10			

In the experiment, we sent images of the front of the store to the API ten times and counted the number of times that (a) the signboard area was detected correctly; and (b) the signboard images were properly classified. The results of the experiment are shown in Table 1. The overall recognition accuracy was 86%, and recognition was performed within 1 second. Recognition in real-time was achieved by calling the API every 500 milliseconds. In the event that the same store had a number of similar signboards, then the recognition accuracy appeared to decrease. This framework allows users to intuitively access store information from signboards.

REFERENCES

[1] S. Kitamura and M. Matsushita. 2017. Information Identification Support Method for Areas with Densely Located Signboards. In *Adjunct Publication of the 30th Annual ACM Symposium on User Interface Software and Technology (UIST '17).* ACM, New York, NY, USA, 159–160. https://doi.org/10.1145/3131785.3131815

[2] P. He, W. Huang, Y. Qiao, C. C. Loy, and X. Tang. 2016. Reading Scene Text in Deep Convolutional Sequences. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI'16).* AAAI Press, 3501–3508. arXiv:1506.04395v2

[3] J. Redmon and A. Farhadi. 2016. YOLO9000: Better, Faster, Stronger. *Computer Vision and Pattern Recognition. arXiv preprint* (2016). arXiv:1612.08242

[4] K. Simonyan and A. Zisserman. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. *Computer Vision and Pattern Recognition. arXiv e-prints* (2015). arXiv:cs.CV/1409.1556v6

SUI'18, October 13–14, 2018, Berlin, Germany

ACM ISBN 978-1-4503-5708-1/18/10.
https://doi.org/10.1145/3267782.3274674