Image Classification for Training Data Sets Composed of Depth Prediction and Edge Extraction Images from Overhead by Using Convolutional Neural Network.

S. Sumimoto, Y. Miyata, Y. Uwate, Y. Nishio

Tokushima University

{*sumimoto, y.miyata, uwate, nishio*}*@ee.tokushima-u.ac.jp*

Image recognition by deep learning is used in various fields. An automatic drive vehicle uses image recognition by deep learning when it avoids dangers. In this way, it is important for a machine to use image recognition by deep learning. Drones are also recently used in various fields. Then image recognition by deep learning is becoming important for that drones fly safely. Therefore, we used YOLOv3 [1]-[3] for drones to fly safely. It is a popular object detection algorithm. However, standing people in overhead images from the view of drones are not able to be recognized by using YOLOv3. So, Convolutional Neural Network (CNN) needs more images from overhead, but it's difficult to obtain many images from overhead.

In this study, we investigate the prediction of the depth of some objects, such humans and cars in overhead images with Fully Convolutional Residual Networks (FCRN) [4]. This system can predict depth of images taken by a monocular camera. Depth prediction images from FCRN have 3D data, so CNN can obtain more data from them than 2D images. Also, we process RGB images into edge extraction images by using OpenCV. In this way, we try to get high accuracies with few images. We classify the images of a human and a car which are composed of depth prediction images and edge extraction images, blended images by using CNN in Fig. 1. We aim at differentiating human or other objects. We investigate learning and test accuracies of the image classification with some traing data sets. We make four training data sets in Fig. 2, and compare thier test accuracies. First data set has 200 RGB images. Second data set has 100 RGB images and 100 depth prediction images. Third data set has 100 RGB images and 100 edge extraction images. Forth data set has 100 imgages blended by a RGB image and a depth prediction image, and 100 images blended by a RGB image and an edge extraction image.

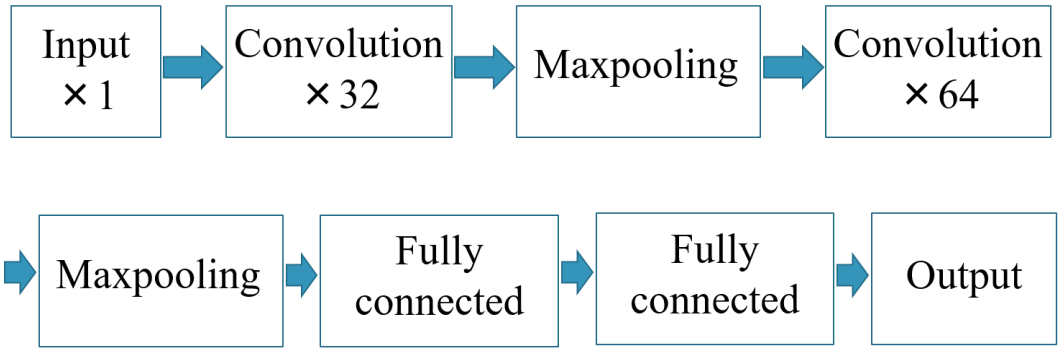
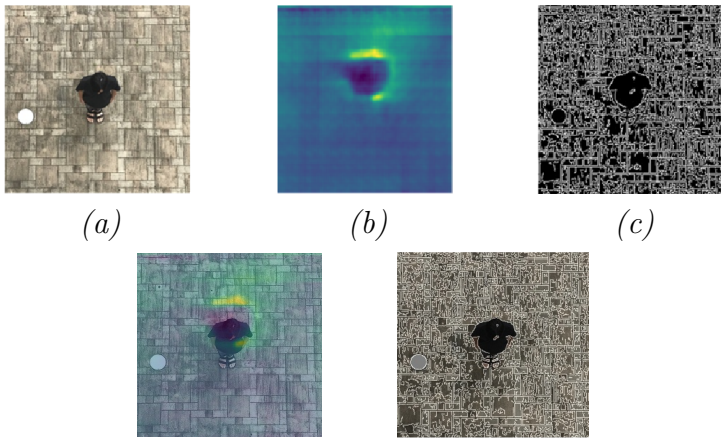


Figure 1: *Structure of CNN.*



(a) *(b)* *(c)*

(d) *(e)*

Figure 2: *Images of a human from overhead.*
(a) A RGB image.
(b) A depth prediction image.
(c) An edge extraction image.
(d) A image blended by a RGB image and a depth prediction image.
(e) A image blended by a RGB image and an edge extraction image.

References

[1] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," arXiv preprint arXiv:1506.02640, 2015.

[2] Joseph Redmon, Ali Farhadi, "YOLO9000: Better, Faster, Stronger," arXiv preprint arXiv:1612.08242, 2016.

[3] Joseph Redmon, Ali Farhadi,"YOLOv3: An Incremental Improvement," arXiv preprint arXiv: arXiv:1804.02767, 2018.

[4] Iro Laina, Christian Rupprecht, Vasileios Belagiannis, Federico Tombari, Nassir Navab, "Deeper Depth Prediction with Fully Convolutional Residual Networks," 2016 Fourth International Conference on 3D Vision (3DV), pp. 239248, 2016.