Design of Convolutional Neural Network for Classifying Depth Prediction Images from Overhead

Shu Sumimoto, Yuichi Miyata, Ryuta Yoshimura,Yoko Uwate and Yoshifumi Nishio
dept. of Electrical and Electronic Engineering
Tokushima University
2-1 Minami-Josanjima, Tokushima 7708506, Japan
E-mail: sumimoto, y.miyata, yoshimura, uwate, nishio@ee.tokushima-u.ac.jp

Abstract— **We predict depth of some objects, such a person, chairs and a soccer ball and so on, in overhead images with Fully Convolutional Residual Networks (FCRN) [1]. This networks can predict depth of RGB images taken by monocular cameras. And we classify images predicted depth. Thus we aim at differentiating person or other objects.**

Keywords; Neural Network; Image classification; Depth prediction

I. INTRODUCTION

Image recognition by deep learning is used in various field. 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 infiltrating various fields, for example delivery, rescue, guard and so on. It is necessary that drones fly safely. Then image recognition by deep learning is becoming important for that drones fly safely. Therefore, I used YOLOv3 for drones to fly safely. It is a popular object detection algorithm. Also it is suited to avoid any dangers, because it can recognize objects quickly. However, standing people in overhead images from the view of drones are not able to be recognized by using YOLOv3. It learns shapes of humans, so it is difficult for YOLOv3 to recognize humans that hide a part of a body.

In this study, we investigate the prediction of the depth of some objects, such humans, chairs and cars in overhead images with FCRN. This system can predict depth of images taken by a monocular camera, so it costs lower than any systems. Depth prediction images from FCRN have 3D data, so Convolutional Neural Network(CNN) can get more data from them than 2D images. We classify the RGB images and the depth prediction images taken by a monocular camera and we aim at differentiating human or other objects by using CNN. Also, the camera has two position, one is lower position and another is higher position. We compare learning and test accuracies of image classification with two camera position.

II. PROPOSED SYSTEM

We propose to classify depth prediction images with FCRN. First, we prepare four types images from overhead, images of a human and a chair taken by a camera which is close to objects, and images of a human and a car taken by a camera which is far from objects.

Second, we predict depth of overhead RGB images of a human, a chair and car taken by monocular cameras in Figs. 1 and 2.

Third, we classify the RGB images and the predicted depth images of a human and a chair or a car with a CNN which has 2 convolutional layers, 2 pooling layers and 2 fully connected layers. When CNN learns, images are compressed. Therefore, training and test images are 28×28 pixels. Learning rate of this CNN is 0.00005.

We compare the learning and test accuracies when a camera is close to an object and when a camera is far from an object. When they are close to a camera, we classify images of a human and a chair. On the other hand, when they are far from a camera, we classify images of a human and a car.

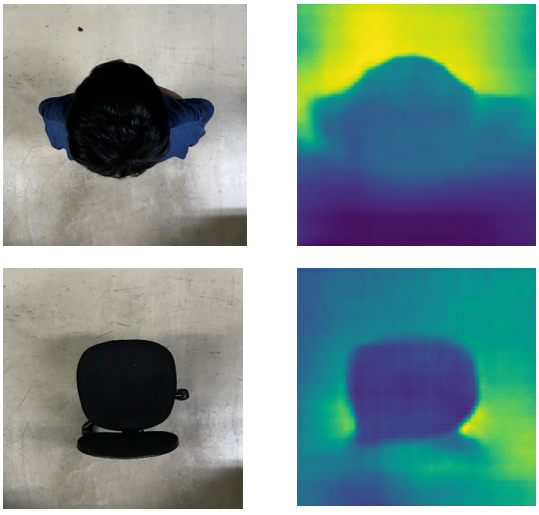

Figure 1. RGB and depth prediction images when a camera is close to objects. (training data)

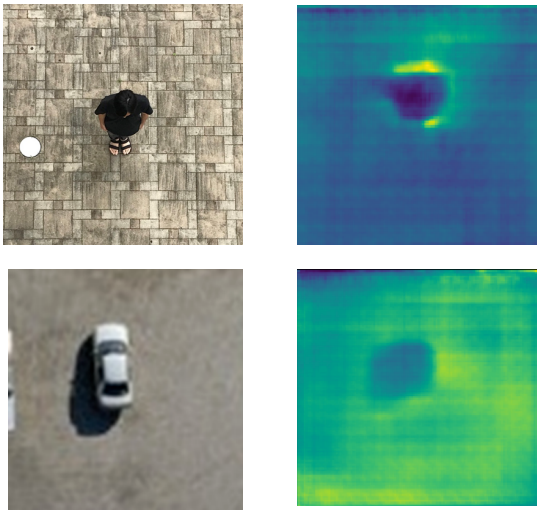

Figure 2. RGB and depth prediction images when a camera is far from objects. (training data)

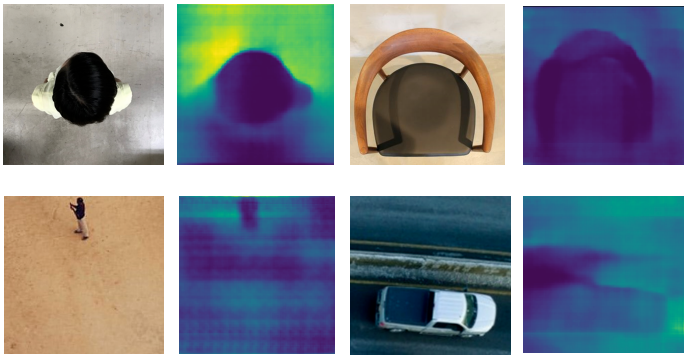

Figure 3. RGB and depth prediction images. (trest data)

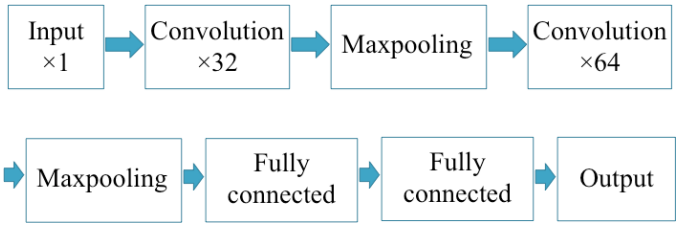

Figure 4. Structure of CNN.

III. Simulation Result

We define as the learning steps 300, the number of the training data sets 80 which are 40 overhead images of a human and 40 overhead images of a chair and the number of the test data sets 10 which are 5 overhead images of a human and 5 overhead images of a chair. They are taken by a camera which is close to an object. Table 1 shows average of learning and test accuracies when we classify only RGB images 10 times and we classify only predicted depth images 10 times.

We define as the learning steps 100, the number of the training data sets 80 which are 40 overhead images of a human and 40 overhead images of a car and the number of the test data sets 10 which are 5 overhead images of a human and 5 overhead images of a car. They are taken by a camera which is far from an object. Table 2 shows average of learning and test accuracies when we classify only RGB images 10 times and we classify only predicted depth images 10 times.

From Tables 1 and 2, when a camera is close to an object, the test accuracy of classifying RGB images is higher than predicted depth images. On the other hand, when a camera is far from an object, the test accuracy of classifying predicted depth images is higher than RGB images.

TABLE I. Average of learning and test accuracies hen a camera is close to an object

	RGB images	***Predicted depth images***
Learning accuracy	0.97	0.90
Test accuracy	0.87	0.72

TABLE II. Average of learning and test accuracies hen a camera is far from an object

	RGB images	***Predicted depth images***
Learning accuracy	0.83	0.76
Test accuracy	0.38	0.58

IV. Conclusion

From these simulation results, we consider it is effective for classifying images taken by a camera which is far from an object to predict the depth of images.

However, the learning accuracies of classifying predicted depth images are lower than RGB images. Then we will try to raise accuracies of classifying predicted depth images.

References

[1] Iro Laina, Christian Rupprecht, Vasileios Belagiannis, Federico Tombari, Nassir Navab, “Deeper Depth Prediction with Fully Convolutional Residual Networks”

[2] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, “You Only Look Once: Unified, Real-Time Object Detection”

[3] Joseph Redmon, Ali Farhadi, “YOLO9000: Better, Faster, Stronger”

[4] Joseph Redmon, Ali Farhadi,“YOLOv3: An Incremental Improvement”