SUGISAKA MASANORI

Journal of Advances in Artificial Life Robotics
Vol. 2(3); December (2021), pp. 125-130
ON LINE ISSN 2435-8061; ISSN-L 2435-8061
https://alife-robotics.org/jallr.html

Research Article

Cycle-Generative Adversarial Network for Generating a Pseudo Realistic Food Dataset Using RGB and Depth Images

Obada Al aama[1], Yuma Yoshimoto[2], Hakaru Tamukoh[2]

[1]Department of Life Science and Systems Engineering, Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

[2]Department of Human Intelligence System, Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

ARTICLE INFO

Article History

Received 25 November 2020
Accepted 25 February 2022

Keywords

Cycle-GAN
Food dataset
RGB-D images

ABSTRACT

Constructing a food dataset is time and effort consuming due to the requirement for covering the feature variations of food samples. Additionally, a large dataset is needed for training neural networks. Generative adversarial networks (GANs) are a recently developed technique to learn deep representations without extensively annotated training data. They can be used in several applications, including generating food datasets. This paper advocates the use of Cycle-GAN to generate a large pseudo-realistic food dataset based on a large number of simulated images and a small number of real images in comparison to traditional techniques. A single depth camera in three different angles and a turntable are arranged to capture real RGB-D images of food samples. 3D modeling software is used to generate simulated images using the same configuration of captured real images. Results showed that Cycle-GAN realistic style transfer on simulated food objects is achievable, and that it can be an efficient tool to minimize real image capturing efforts.

1. Introduction

Recently, generative adversarial networks (GANs) have attracted particular attention and have been widely studied since 2014, and several algorithms have been proposed [1]. GANs can generate an image that resembles a real one which then used for training as a dataset. Basically, there are various dataset being used to train GAN including human face image dataset such as CelebA and a numeric character image dataset such as MNIST [2].

GANs have produced promising results in many generative tasks, such as photo-realistic image generation. In fact, obtaining a large dataset needed for training neural networks is time and effort consuming. The Cycle-GAN is a technique that involves the automatic training of image-to-image translation models with relatively high resolutions, compared with other networks [3]. Other related work is the semi-automatic RGB dataset generation for object detection and classification [4]. This paper aims to employ the automated Cycle-GAN technique to synthesize a large food dataset by the means of converting the simulated images to more relevant realistic images instead of only using a large real image dataset. Accordingly, it generates the nearest-to-real images through training on a small number of real images simultaneously with a large number of simulated images. The method used in this study synthesizes a large food dataset effortlessly compared with the traditional imaging techniques. This paper is an extended version of the proceeding of ICAROB 2021 [5].

2. Research Concept

The main concept focuses on using a depth camera to capture both RGB and depth images simultaneously, while

Corresponding author's E-mail: *aama.obada-walidal425@mail.kyutech.jp, yoshimoto@brain.kyutech.ac.jp, tamukoh@brain.kyutech.ac.jp*
URL: www.lsse.kyutech.ac.jp

also employing Autodesk 3D-Maya software [6]-to obtain the simulated images from 3D food models. Afterwards, training the Cycle-GAN using the obtained images is done to investigate the capability of Cycle-GAN to produce the pseudo realistic images. The training is based on few real images and several simulated images.

3. Methodology

3.1. Real image capturing

The image capturing setup, as shown in Figure 1, was comprised of the Intel® RealSense™ Depth Camera D435i, a turntable on which the food objects were placed, and a fixture to hold the depth camera in different pose configurations. The capturing process was carried out through adjusting the depth camera to be in front of an actively rotating turntable where the food objects including an apple, a kiwi, a mini-tomato and a mushroom were placed. Both RGB and depth images were captured using this camera. The settled pose configurations of image capturing were as follows; the distance from the object was 50 cm and three different capturing angles (0°, 45°, and 90°) were used. As for the turntable, the rotation rate was 1.22 rpm. As for the image capturing configuration, the frame rate was 30 fps, and the output image resolution was 640×360 pixels.

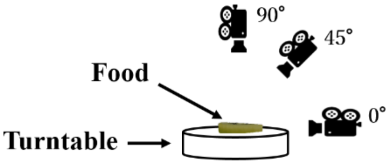

Fig. 1. Image capturing setup showing three different pose configurations.

The obtained images were further processed as shown in Figure 2. All images for each food class were organized into separate folders respectively. Each class has 13140 images; 6570 for RGB and the same number for depth in Jet-color mapping [7], which are compised of 2190 images from each selected angle (2190 * 3 capture poses * 2 modalities = 13140 images). The Robot Operating System (ROS) was used to facilitate the image capturing process, specificaly the Rosbag record utility to store the images in .Bag files. During the capture process, region of interest (ROI) cropping at 150×150 pixels and normalization were applied simultaneously to the images that were stored in the resulting .Bag file format. Extraction of images was carried out using the CV_bridge library [8], [9] in order to organize the captured images into respective directories. Moreover, the saved images were further processed by removing the background after the manual detection of the background color range for RGB via Chroma-key masking [10], and applying the same mask on the depth images.

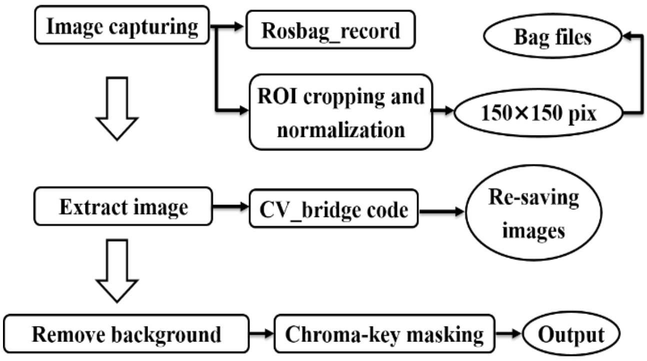

Fig. 2. Process flow illustrating the different procedures of real image capturing beginning from image capturing and ending with processed outputs.

3.2. Simulated image generation

The simulated 3D model for each object class (an apple, a kiwi, a mini-tomato and a mushroom) was downloaded freely from web sources as an .obj format. Thereafter, these models were imported to the Autodesk 3D Maya software to obtain the simulated images using the same cofigurations applied for capturing the real images (capturing angles, object capturing distance, turntable rotation rate, capturing rate and resolution), resulting in the same number of captured images as those obtained from real capturing. Finally, the background was also algorithmically removed for both RGB and depth images during the rendering step. However, the simulated depth images were obtained in grey-scale color model. The whole setup of the process is presented in Figure 3.

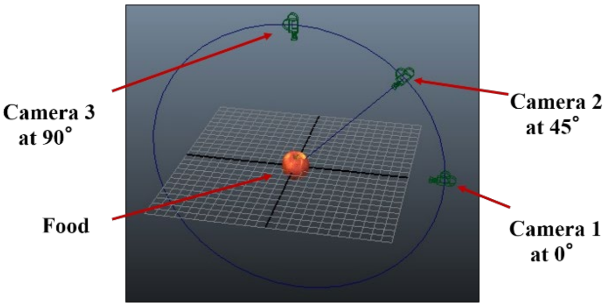

Fig. 3. Simulated images generation

3.3. Cycle-GAN image generation

As an initial experiment, for each food class 500 real images and 500 simulated images were used to train the Cycle-GAN neural network. The epoch was setup to be 200 epochs, yielding a training time of about 11 hours. Thereafter 6070 images were tested to validate the Cycle-GAN neural network. The main concept and various processes applied through Cycle-GAN is illustrated in Figure 4.

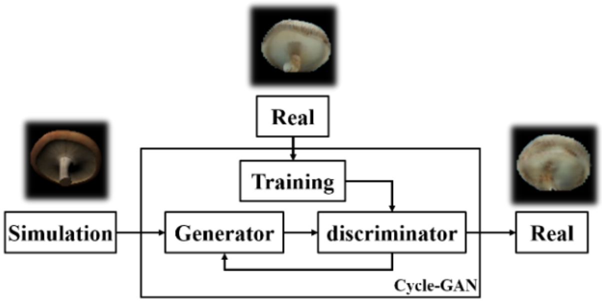

Fig. 4. The main concept of applying Cycle-GAN to generate pseudo real images.

4. Results and discussion

4.1. Real imaging

Four classes were selected as food objects which were mushroom, mini-tomato, apple, and kiwi. For each object class both RGB and depth were captured in three different capturing angles 0°, 45°, and 90°, resulting in 13140 total images as 2190 images for each category at each capturing angle. In Figures 5, 6, 7, and 8 the real images of different food objects are shown, along with the applied mask, and the obtained output. Moreover, at 0° the turntable surface appeared in all captured images and this required an extra cropping step to remove the interfered surface part before applying the Chroma-key masking. However, at 45° and 90° this process was not necessary and the Chroma-key masking was applied directly.

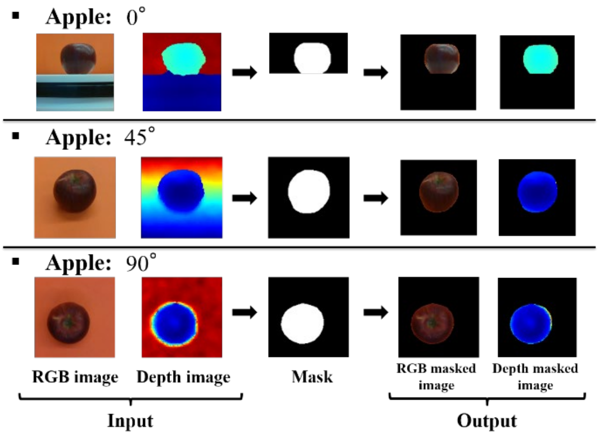

Fig. 5. Real captured RGB and depth images for apple, and the output after masking.

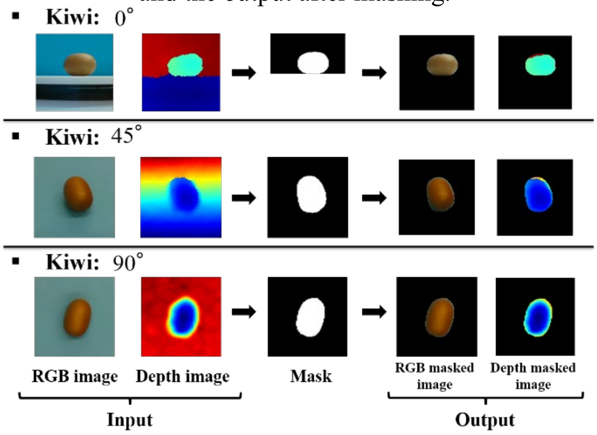

Fig. 6. Real captured RGB and depth images for kiwi, and the output after masking.

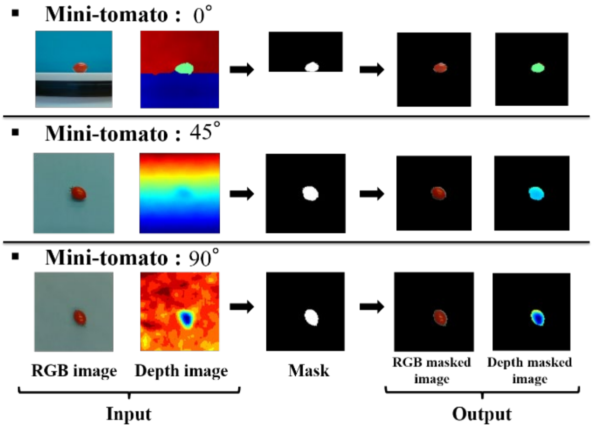

Fig. 7. Real captu0red RGB and depth images for mini-tomato, and the output after masking.

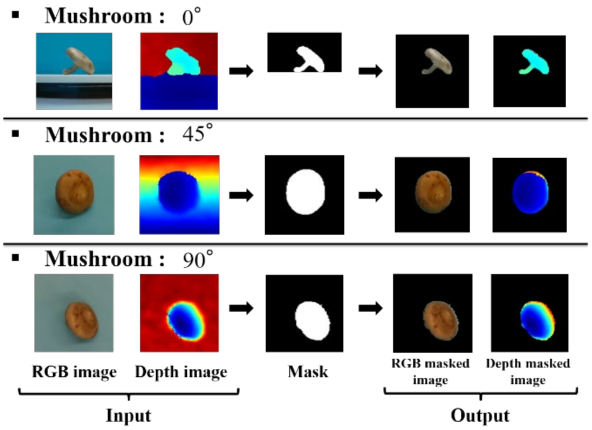

Fig. 8. Real captured RGB and depth images for mushroom, and the output after masking.

4.2. Simulated imaging

As per the process in section 3.2, the same number of simulated images were obtained at each angle for both RGB and depth as those obtained with the real imaging process. However, the obtained depth images were grey-scale. Figure 9 shows the obtained RGB and depth simulated images for each investigated food class.

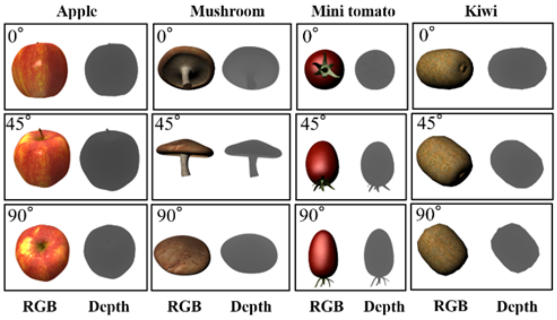

Fig. 9. Simulated RGB and depth images for apple, mushroom, mini-tomato, and kiwi.

4.3. Cycle-GAN

As previously described, 500 RGB Real images in addition to 500 RGB simulated for each object class were used to train the Cycle-GAN neural network. Afterwards, the rest of the total obtained simulated images (6070) were used in the testing step as shown in Figure 10. It was observed from the obtained RGB output that Cycle-GAN is considerably an effective tool for the generation of near to real images in various angles. On the other hand, for the depth images, 500 images were also selected for each object class to train the Cycle-GAN. In the same manner, the rest of the total obtained images were used to test the Cycle-GAN. Figure 11 shows the obtained depth Cycle-GAN output for the trained food classes as example.

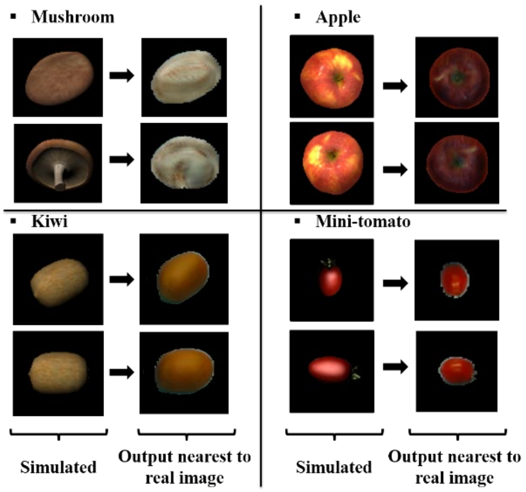

Fig. 10. Samples for Cycle-GAN output for different food classes RGB images at variant angles.

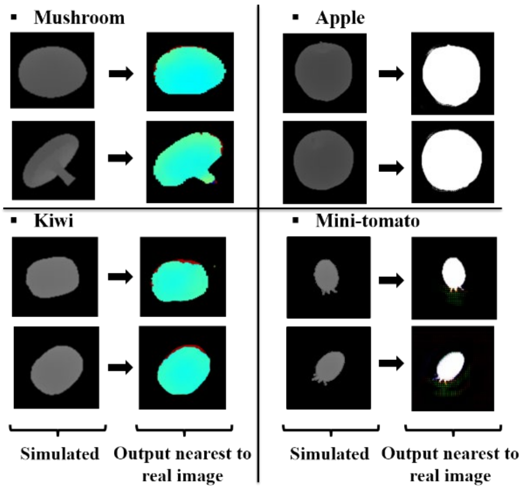

Fig. 11. Samples for Cycle-GAN output for different food classes depth images at variant angles.

4.4. Generated dataset properties

n this section we describe the number of images for each class (RGB or depth) generated by the trained GAN. In our trial, we trained the Cycle-GAN on 500 real captured images and another 500 simulated images. After training, we used 6070 simulated images as input to the trained GAN model, and accordingly obtaind 6070 of near to real images as shown in Tabel 1. As a result, we obtained a large number of generated images using a small number of trained ones. In this case, we managed to obtain about 12 times of the number of real images in the form of near-realistic images. As a matter of fact, we can obtain more output images generated by the GAN based on the number of input simulated images (which are easier to obtain in comparison to real captured images) shown in Table 1.

Table 1. Number of GAN images pre and post training for each class, in case of RGB or depth images.

GAN training input images		**Post training input/output images**	
Real	**Simulated**	**Input simulated**	**Nearest to real**
500	500	6070	6070

5. Conclusion and future work

This paper proposed to synthesize a large food dataset based on the use of a small number of real images along with a large number of simulated images utilizing the Cycle-GAN to produce pseudo realistic images. We used only 500 real and another 500 simulated input images for both RGB and depth for training the Cycle-GAN for each food class. The training set size was considerably smaller when compared to the testing set which was 6070 simulated images. Our proposed method proved to be efficient in synthesizing a large food dataset. A generated dataset could be used to train neural networks for various tasks and applications, such as robotic pick-and-place in kitchens, hospitals and convenience stores.

However, our data represents a preliminary result, so accordingly a future study focusing on the effect of training set size on the realistic quality of the Cycle-GAN output should be conducted. In addition, we aim to focus on widening the investigated food dataset to include additional classes representing Japanese food. We also aim to investigate the quality of using alternative capturing methods, such as 3D scanners that output computerized 3D models of objects [11]. 3D scanners can prove to be more efficient time and effort-wise when compared to single camera-based capturing [12]. Dual stream neural network DS-NN model can be trained using our output dataset for object recognition [13], [14]. A similar model using RGB and Depth images, for increasing operational accuracy in service robots, can be implemented based on our dataset [15].

• References

[1] J. Gui, Z. Sun, Y. Wen, D. Tao, and J. Ye, "A review on generative adversarial networks: Algorithms, theory, and applications," arXiv preprint arXiv:2001.06937, 2020.

[2] D. Horita, R. Tanno, W. Shimoda, and K. Yanai, "Food category transfer with conditional cyclegan and a large-scale food image dataset," in Proceedings of the Joint Workshop on Multimedia for Cooking and Eating Activities and Multimedia Assisted Dietary Management, 2018, pp. 67-70.

[3] D.-h. Kwak and S.-h. Lee, "A novel method for estimating monocular depth using cycle gan and segmentation," Sensors, vol. 20, 2020 , p. 2567.

[4] Y. Ishida and H. Tamukoh, "Semi-automatic dataset generation for object detection and recognition and its evaluation on domestic service robots," Journal of Robotics and Mechatronics, vol. 32, 2020 , pp. 245-253.

[5] O. Al aama and H. Tamukoh, "Synthesis of realistic food dataset using generative adversarial network based on RGB and depth images," The 2021 International Conference on Artificial Life and Robotics (ICAROB 2021), OS19-4, 2021.

[6] Autodesk, INC. Website, 2019. Available online:https:/autodesk.com/maya (accessed on 31 August 2021).

[7] M. M. Rahman, Y. Tan, J. Xue, and K. Lu, "RGB-D object recognition with multimodal deep convolutional neural networks," in 2017 IEEE International Conference on Multimedia and Expo (ICME), 2017, pp. 991-996.

[8] "Converting Between ROS Images And OpenCV Images Python", ROS Wiki Website, 2021. Available online: http://wiki.ros.org/cv_bridge/Tutorials/ConvertingBetweenROS ImagesAndOpenCVImagesPython (accessed on 15 September 2021).

[9] E. Fernandez, L. S. Crespo, A. Mahtani, and A. Martinez, Learning ROS for robotics programming: Packt Publishing Ltd, 2015.

[10] C.-H. Teng, Y.-H. Liao, Y.-C. Chou, and S.-Y. Lin, "Removing blue screen background under non-uniform illumination," in 2017 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), 2017, pp. 307-308.

[11] J. Olatunji, G. Redding, C. Rowe, and A. East, "Reconstruction of kiwifruit fruit geometry using a CGAN trained on a synthetic dataset," Computers and Electronics in Agriculture, vol. 177, 2020, p. 105699.

[12] Y. Abe, Y. Ishida, T. Ono, and H. Tamukoh, "Acceleration of training dataset generation by 3D scanning of objects," ICAROB 2020.

[13] Y. Yoshimoto and H. Tamukoh, "FPGA Implementation of a Binarized Dual Stream Convolutional Neural Network for Service Robots," Journal of Robotics and Mechatronics, vol. 33, 2021, pp. 386-399.

[14] Y. Yoshimoto and H. Tamukoh, "Hardware-Oriented Dual Stream Object Recognition System using Binarized Neural Networks," in 2020 IEEE International Symposium on Circuits and Systems (ISCAS), 2020, pp. 1-5.

[15] Y. Yoshimoto and H. Tamukoh, "Live Demonstration: Hardware-Oriented Dual Stream Object Recognition System using Binarized Neural Networks," in 2020 IEEE International Symposium on Circuits and Systems (ISCAS), 2020, pp. 1-1.

Authors Introduction

Mr. Obada Al aama

He graduated from Al-Baath University Department of Communication and Electronics Engineering, Syria, in 2013. He received Master of Engineering degree from Kyushu Institute of Technology, Japan, in 2019. He is currently a doctoral candidate at the Kyushu Institute of Technology, Japan. His research interests include image processing, deep learning, and neural networks.

Dr. Yuma Yoshimoto

He received the B.E. and Ph.D. degrees from the Kyushu Institute of Technology, in 2018 and 2021.
From 2019 to 2021, he had also been a Research Fellow of the Japan Society for the Promotion of Science (JSPS). Since 2021, he has also been a research fellow at the Kyushu Institute of Technology. His main research interests include image processing using neural networks for robots.

Prof. Hakaru Tamukoh

He received the B.Eng. degree from Miyazaki University, Japan, in 2001. He received the M.Eng and the Ph.D. degree from Kyushu Institute of Technology, Japan, in 2003 and 2006, respectively. He was a postdoctoral research fellow of 21st century center of excellent program at Kyushu Institute of Technology, from April 2006 to September 2007. He was an assistant professor of Tokyo University of Agriculture and Technology, from October 2007 to January 2013. He is currently an associate professor in the graduate school of Life Science and System Engineering, Kyushu Institute of Technology, Japan. His research interest includes hardware/software complex system, digital hardware design,neural networks, soft-computing and home service robots. He is a member of IEICE, SOFT,JNNS, IEEE, JSAI and RSJ.