# Synthesis of realistic food dataset using generative adversarial network based on RGB and depth images

**Obada Al Aama**

*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*


**Hakaru Tamukoh**

*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*
*E-mail:aama.obada-walidal425@mail.kyutech.jp, tamukoh@brain.kyutech.ac.jp*
*www.lsse.kyutech.ac.jp*

## Abstract

Constructing a large food dataset is time and effort consuming due to the need to cover the feature variations of food items. Hence, a huge data is needed for training neural networks. This paper aims to advocate the Cycle-GAN to build up large food dataset based on large number of simulated images and relatively few real captured images thus obtaining more realistic images effortlessly compared with traditional capturing. Real RGB and depth images of variant food samples allocated over turntable were captured in three different angles using real-sense depth camera with different backgrounds. Furthermore, for simulated images, the Autodesk 3D_Maya software was employed using the same parameters of captured real images. Results showed that generally, realistic style transfer on simulated food objects was obtained as a result of employing Cycle-GAN. GAN proved to be an efficient tool that could minimize imaging efforts resulting in realistic images.

*Keywords: Cycle-GAN, Food dataset, RGB-D images.*

## 1. Introduction

Recently, generative adversarial networks (GANs) have attracted particular attention. A GAN can generate an image that resembles a real image. A human face image dataset such as CelebA and a numeric character image dataset such as MNIST have been used for training the GAN as the target domains [1]. GANs have produced promising results in many generative tasks, such as photo-realistic image generation. One of the important obstacles of generating a large dataset is the need of a huge number of real images which is time and effort consuming. Meanwhile, few have reported food image generation or transformation using the GAN thus far.

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 large number of real images, hence, the Cycle-GAN produces the nearest real images based on processing a few numbers of real images along with the huge number of simulated images compared with the traditional imaging techniques. Results obtained from this study represent preliminary results that could be further used to train neural network for object detection which could be applied in various robotics applications.

*Obada Al Aama, Hakaru Tamukoh*

## 2. Research Concept

The main concept focuses over using the depth camera to capture both RGB and depth images simultaneously, on the other hand, employing Autodesk 3D-Maya software to obtain the simulated images from 3D food models. Thereafter, training and testing the Cycle-GAN using those above mentioned images to investigate the capability of Cycle-GAN to produce the realistic images based training and testing few real images and huge simulated images.

## 3. Methodology

### 3.1. Real image capturing

This experiment was carried out through adjusting the Intel® RealSense™ Depth Camera D435i to be in front of turntable where the food objects including Apple, Kiwi, Mini-tomato and Mushroom were allocated. Both RGB and depth images were captured using this camera. The settled parameters of image capturing were as follows; distance from the object was 50 Cm, different capturing angles (0°, 45°, and 90°), turntable rotation rate was 1.22 rpm, capturing rate was 30 fps and the image resolution was 640×360 pixels. Images obtained were further extracted and processed as shown in Figure 1. All images for each food class were categorized in specific folders. Each class has 6570 images for RGB and the same number for depth in Jet-color mapping through obtaining 2190 images from each selected angle. In image capturing step the Rosbag_record was employed and accordingly the Region of Interest ROI cropping at 150×150 pixels and normalization were applied simultaneously resulting in Bag file format [2]. Extraction of images was carried out using the CV_bridge code [3, 4] which role is to re-save the images from RVIZ simulator to the proposed directory. Moreover, the saved images were processed through removing the background after the manual detection of the background color range for RGB via Chroma-key masking [5], and applying the same mask on the depth images.

### 3.2. Simulated image generation

The 3D simulated model for each candidate was downloaded from open source as an obj format. Thereafter these models were imported to the Autodesk
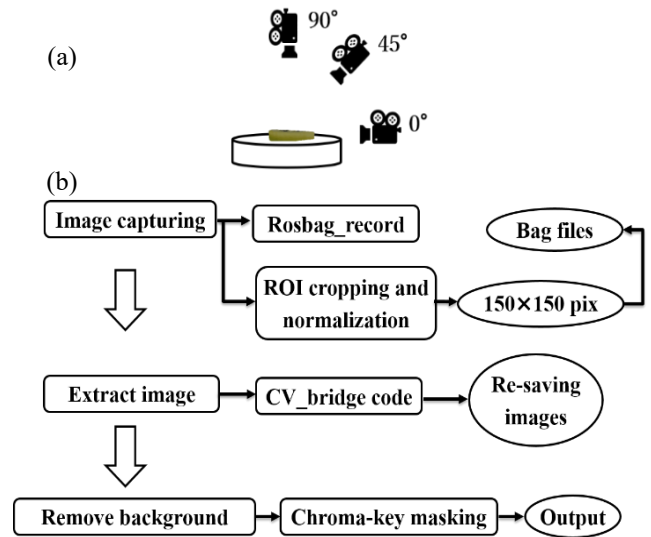


Fig. 1. Schematic diagram illustrating the different procedures of real image capturing beginning from the (a) image capturing process, and (b) different steps of extracting and processing of the obtained images.

3D_Maya software [6] to obtain the simulated images using the same parameters 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 automatically removed for both RGB and depth during the render step. However, the simulated depth images were obtained in grey-scale color model. The whole setup of the experiment was illustrated in Figure 2.
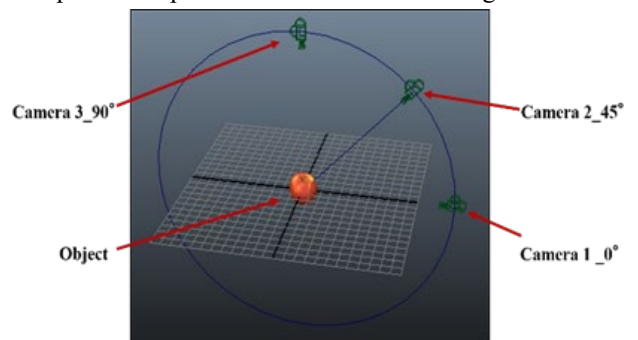


Fig. 2. Simulator experiment setup.

### 3.3. Cycle-GAN image generation

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, and the training time was about 11 hours. Thereafter 6070 images were tested to validate Cycle-GAN neural network. Through Cycle-GAN training the equation used were obtained from 1. The main concept and various processes applied through Cycle-GAN was illustrated in Figure 3.
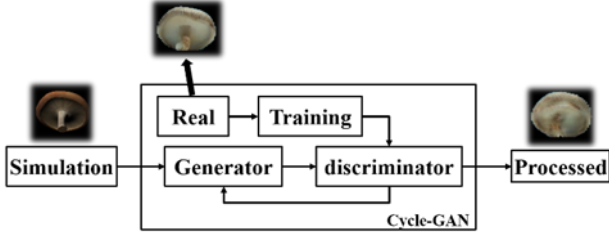


Fig. 3. Cycle-GAN main concept.

## 4. Results and discussion

### 4.1. Real imaging

Four classes were selected as food objects which were mushroom, mini-tomato, apples, and kiwi. For each object both RGB and depth were captured in three different capturing angles 0°, 45°, and 90°, resulting in 26280 total images as 2190 images for each category at each capturing angle. In Figures 4, 5, 6, and 7 the real images of different food objects were shown, along with the applied mask, and the obtained output. Moreover, at 0° the turn table surface appeared in all captured images and this require an extra cropping step to remove the interfered part before applying the Chroma-key masking. However, at 45° and 90° this process was not required and the Chroma-key masking was applied directly.
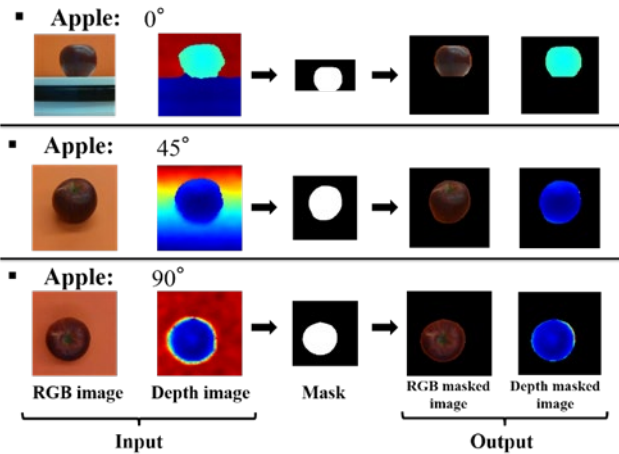


Fig. 4. Real captured RGB and depth images for apple, and the output after masking.
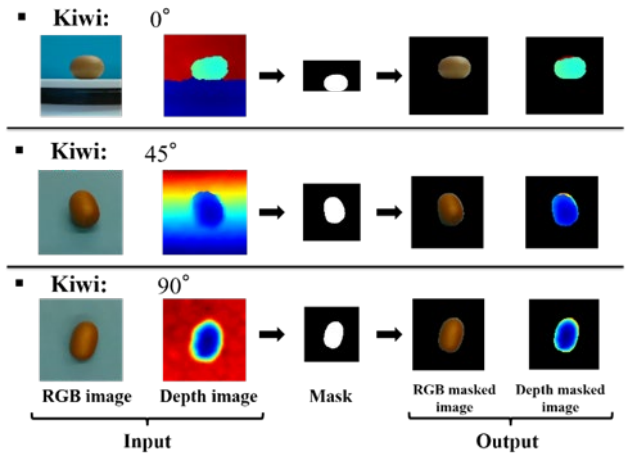


Fig. 5. Real captured RGB and depth images for kiwi, and the output after masking.
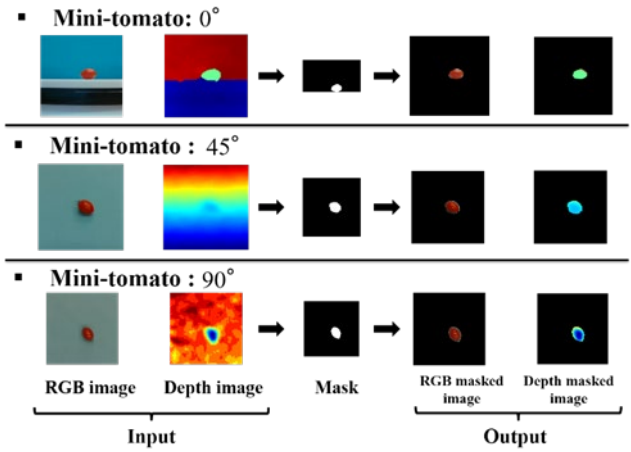


Fig. 6. Real captured RGB and depth images for mini-tomato, and the output after masking.
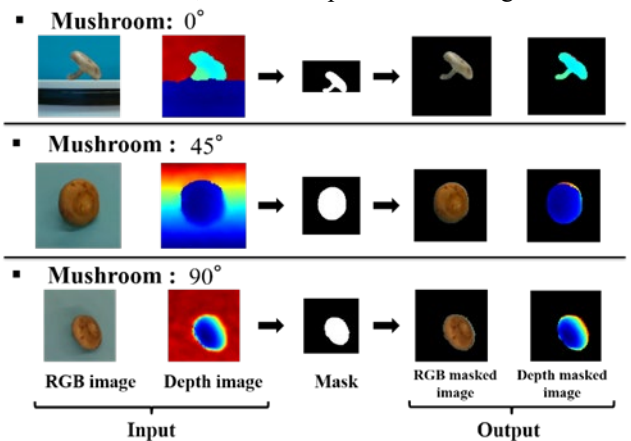


Fig. 7. Real captured RGB and depth images for mushroom, and the output after masking.

### 4.2. Simulated imaging

The same number of simulated images were obtained at each angle for both RGB and depth as those obtained in real imaging. However, the obtained depth images were grey-scale. Figure 8 shows the obtained RGB and depth simulated images for each investigated food class.
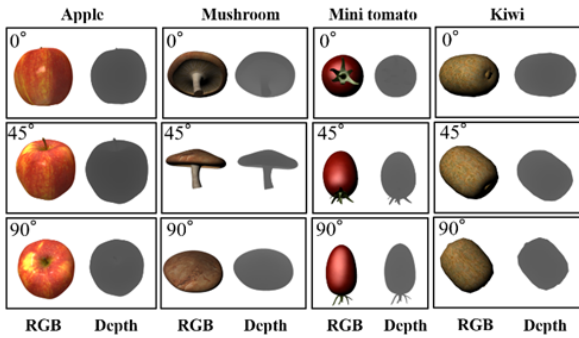


Fig. 8. Simulated RGB and depth images for apple, mushroom, mini-tomato, and kiwi.

### 4.3. Cycle-GAN

500 RGB Real images in addition to 500 RGB simulated for each object class were used to train the Cycle-GAN. Thereafter rest of the total obtained simulated images (6070) were used in Cycle-GAN testing step rather than those used in the training step as shown in Figure 9. It
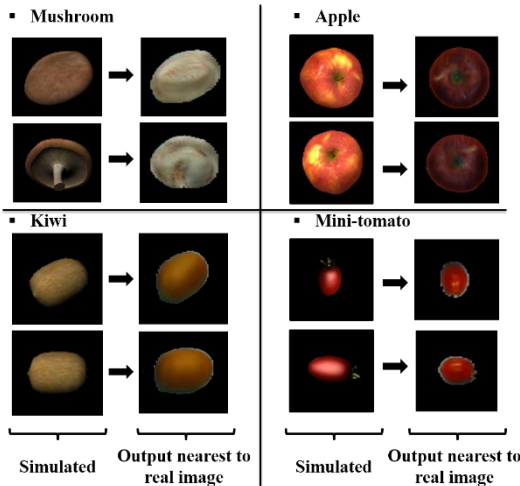


Fig. 9. Samples for Cycle-GAN output for different food classes RGB images at variant angles.

was noticed from the obtained RGB Cycle-GAN output that Cycle-GAN represent an efficient tool in modelling of near to real images in various angles. On the other hand, for the depth images 500 images were selected also for each object class to train the Cycle-GAN, and in the

same manner the rest of the total obtained images were used to test the Cycle-GAN as illustrated in Figure 10 showing the obtained depth Cycle-GAN output for mushroom food class as example.
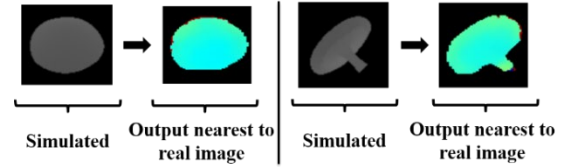


Fig. 10. Samples for Cycle-GAN output for mushroom class depth images.

## 5. Conclusion

This paper proposed to synthesize a large food dataset based on the use of few number of real images along with a huge number of simulated images utilizing the Cycle-GAN to produce realistic images. We used only 500 Real and another 500 simulated for both RGB and depth for training Cycle-GAN for each food class and this number represent a few number compared with that used for testing which was 6070 simulated images. Accordingly, our proposed method proved to be efficient in synthesizing of large food dataset. However, our data represent a preliminary result and a future study focusing on the effect of number of training images on the realistic quality of the output of Cycle-GAN, should be conducted, in addition to focusing on widening of the investigated food dataset including Japanese food.

### References

1. HORITA, Daichi, et al. 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. p. 67-70
2. SHI, Wuzhen, et al. Reference image based method of region of interest enhancement for haze image. In: 2015 IEEE International Conference on Image Processing (ICIP). IEEE, 2015. p. 735-739.
3. Website:"http://wiki.ros.org/cv_bridge/Tutorials/ConvertingBetweenROSImagesAndOpenCVImagesPython", 15/12/2020.
4. MARTINEZ, Aaron; FERNÁNDEZ, Enrique. Learning ROS for robotics programming. Packt Publishing Ltd, 2013.
5. C. Teng, Y. Liao, Y. Chou and S. Lin, "Removing blue screen background under non-uniform illumination," 2017 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-TW), Taipei, 2017, pp. 307-308, doi: 10.1109/ICCE-China.2017.7991118.
6. Autodesk, INC. (2019). Maya. Retrieved from https:/ autodesk.com/maya.