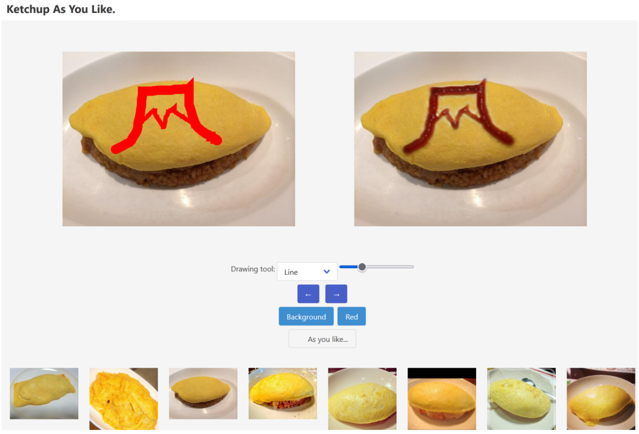
Ketchup As You Like: Drawing Editor for Foods

Shu Naritomi Gibran Benitez-Garcia Keiji Yanai
The University of Electro-Communications, Tokyo
1-5-1 Chofugaoka, Chofu-shi, Tokyo 182-8585 Japan
{naritomi-s,gibran,yanai}@mm.inf.uec.ac.jp

***Abstract*—Omelet rice is a popular Japanese dish made from an egg omelet filled with rice. There is a culture of freely drawing patterns or messages on it with tomato ketchup. The advantage of this practice is that it is easy to create originality and express creativity. We believe that utilizing this advantage would be possible to generate original food models in VR space easily. Therefore, we introduce a web-based real-time application called "Ketchup As You Like", which allows users to freely draw with ketchup on an omelet image. This application is able to generate omelet images with realistic ketchup patterns. Therefore, we introduce a web-based real-time application called "Ketchup As You Like", which allows users to freely draw with ketchup on an omelet image. This application is able to generate omelet images with realistic ketchup patterns. The image generation used in "Ketchup As You Like" is created by combining CNN-based segmentation and generative adversarial networks (GAN). Demo video is available at https://youtu.be/m6AjwEY6Jp8**

***Index Terms*—food image, GAN, image translation, photo editing**


Ketchup As You Like.

Drawing tool: Line

Fig. 1. "Ketchup As You Like" user interface.

I. Introduction

In recent years, due partly to the COVID-19 virus, more and more people communicate on VR SNS (social networking sites) to drown out their loneliness. In VR, environments of social gatherings such as bars or restaurants are often selected. These environments commonly involve food displayed with 3D models of existing dishes. Meanwhile, it is natural that users want to freely create 3D models of foods, not just existing 3D models. Of course, because it's in a VR space, it is not possible to make food that users can eat. Generally, users can only display a pre-created 3D model or assemble a simple model in VR space. Therefore, it is difficult to create something with originality or something that requires realistic texture in real-time in VR space.

This study focused on omelet rice as a food that users can quickly bring out their originality. Omelet rice is a popular Japanese dish made from an egg omelet filled with rice. Because omelet rice has a culture of drawing patterns or messages on eggs with ketchup, it is straightforward to create originality. For this reason, omelet rice is the fourth most popular meal uploaded to Twitter per day in Japan [1]. Thus, we wanted to enable the experience of easily creating this original omelet rice even in VR space. However, there are issues such as how to generate realistic textures easily. For instance, if the omelet rice is decorated with a plain red color as ketchup, there is no reality at all. In addition to the hue, the light condition is also crucial for a realistic look. Therefore, in this research, we will introduce a method for creating realistic textures and a web-based editor for making it easy to use. This method is based on our recent study, Ketchup GAN [2].

II. Proposed System

Figure 1 shows the user interface of the system we propose in this paper. Users can select an omelet image from a pre-selected list shown at the bottom of the UI or upload an omelet image from their own, and the image will be displayed on the left side. The user is free to draw on the displayed image with a red pen. An image is generated on the server-side based on the omelet rice image selected by the user and the pattern drawn with the red pen. Finally, the generated image is displayed on the right side in real-time. This section gives an overview of the system; subsequently, we introduce the networks used internally and the datasets used for learning.

A. System Overview

The system overview of "Ketchup As You Like" is shown in Figure 2. As can be seen, the input comes from the web-based application, while all CNN-based networks run on two different servers. One server is running a segmentation network to recognize the egg region from the original omelet rice image. The second server focuses on generating images with the input patterns drawn with ketchup from a segmentation mask built with the egg region.

First, the browser sends two images, an omelet rice image, and a mask image, to the server after each user input stroke. ① The first server (segmentation server) immediately performs a real-time semantic segmentation on the omelet rice image sent

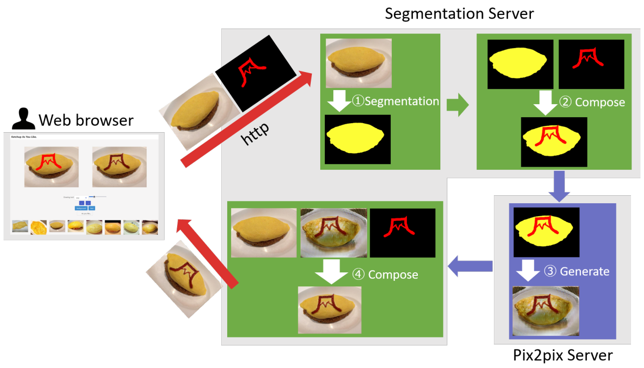


Fig. 2. The overview of "Ketchup As You Like" system.

by the browser. ② Then, a segmentation mask is composed of the omelet rice area obtained in the previous step and the ketchup mask sent from the browser. The segmentation mask is then sent to the second server (Pix2Pix server). Note that in the figure, the egg region of the omelet is shown as yellow and the ketchup region as red for visualization purposes only. In reality, it is a single-channel image with three values packed in 0 (background), 1 (egg), and 2 (ketchup). ③ The Pix2Pix server generates a new omelet rice image based on the entered semantic mask. Note that this process does not use the omelet rice image selected by the user for a conditional image-to-image generation. Therefore we need an extra step to obtain the final output. ④ Finally, the input omelet rice image, the original ketchup mask, and the omelet rice image generated by the Pix2Pix server are combined as follows.

$$mask_b = blur(mask_{src})$$

$$img_{syn} = mask_b * img_{gen} + (1 - mask_b) * img_{src}$$

Where $mask_{src}$ is a mask image with a pixel value between 0 and 1 indicating the user-entered ketchup area. img_{syn} refers to the final synthesized image, img_{gen} is the image generated by Pix2Pix server, and img_{src} is the input omelet rice picture. In brief, the ketchup area of the generated omelet rice image is pasted on the input omelet image. Note that by applying blur to the mask image, the outline of Ketchup can be synthesized more smoothly. Figure 3 shows some results of "Ketchup As You Like" with different drawing patterns.

A framework called Vue.js [3] is used to implement the front end, and the Flask [4] is used for the back end. Vue.js is a reactive JavaScript framework for creating UI. On the other hand, Flask is a micro web framework for Python. You can get a response in about 2 seconds after the browser sends the image and mask to the server.

B. Image Generation

This demo paper network is based on the proposal of Ketchup GAN [2]. We will only give a brief introduction here, so please read the paper [2] for details. FASSD-Net is used for segmentation [5], and SPADE is used for image generation [6]. FASSD-Net [5] is a U-shape encoder-decoder network designed to keep a low computational complexity and exploit high-level features' contextual information by using asymmetric convolutions. SPADE [6] is a GAN-based method for generating realistic images by inputting segmentation masks. SPADE based its architecture on the decoder of Pix2Pix [7], which was the first encoder-decoder-based image translation network that introduces the adversarial loss.

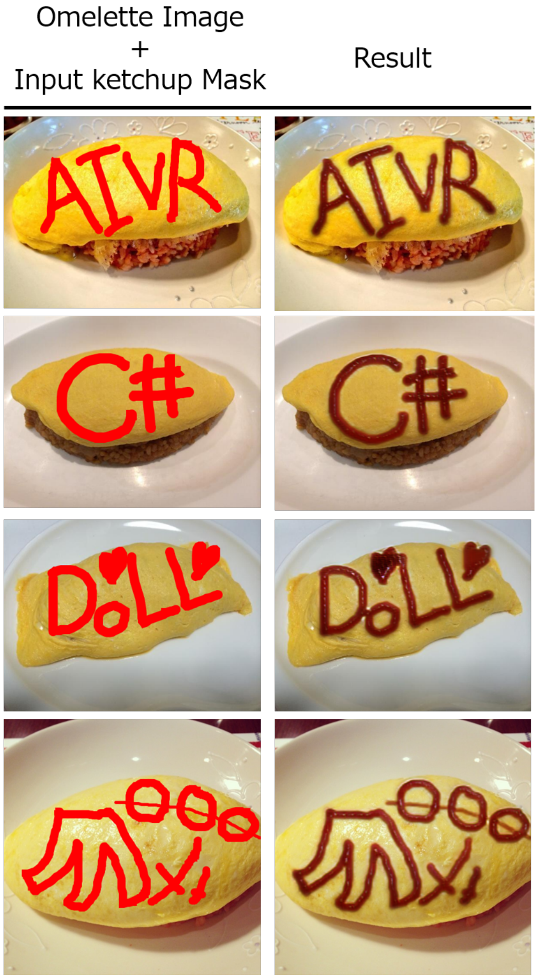


Fig. 3. Results of "Ketchup As You Like" with different drawing patterns.

The images used for training were collected from Twitter. More than 2000 meaningful omelet images with ketchup drawings were selected from more than 100k omelet pictures. The segmentation masks used for training FASSD-Net were generated by weakly supervised learning. For details and access to the database, please refer to [2].

III. Conclusion

In this paper, we proposed a web-based image generation application called "Ketchup As You Like". This application can generate an omelet image with realistic ketchup text by freely writing text and pictures on the input omelet image.

Currently, the demo can generate images only. So it cannot be used in VR space. However, in our latest research [8], we can generate of 3D meshes from a single meal image. We hope to improve creativity in the VR space by combining this technology with the real-time decoration of foods in the future.

References

[1] K. Yanai, K. Okamoto, T. Nagano, and D. Horita, “Large-scale twitter food photo mining and its applications,” in *Proc. of IEEE Fifth International Conference on Multimedia Big Data (BigMM)*, 2019.

[2] G. Benitez-Garcia and K. Yanai, “Ketchup gan: A new dataset for realistic synthesis of letters on food,” in *Proc. of international joint workshop on multimedia artworks analysis and attractiveness computing in multimedia*, 2021.

[3] “Vue.js.” [Online]. Available: https://vuejs.org/index.html

[4] “Flask.” [Online]. Available: https://flask.palletsprojects.com/en/2.0.x/

[5] L. Rosas-Arias, G. Benitez-Garcia, J. Portillo-Portillo, G. Sánchez-Pérez, and K. Yanai, “Fast and accurate real-time semantic segmentation with dilated asymmetric convolutions,” in *Proc. of 2020 25th International Conference on Pattern Recognition*, 2021.

[6] T. Park, M.-Y. L., T.-C. W., and J.-Y. Z., “Semantic image synthesis with spatially-adaptive normalization,” in *Proc. of IEEE Computer Vision and Pattern Recognition*, 2019.

[7] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” in *IEEE Conference on Computer Vision and Pattern Recognition CVPR*, 2017, pp. 1125–1134.

[8] S. Naritomi and K. Yanai, “Hungry networks: 3d mesh reconstruction of a dish and a plate from a single dish image for estimating food volume,” in *Proc. of ACM Multimedia Asia*, 2020.