Magical Rice Bowl: A Real-time Food Category Changer*

Ryosuke Tannno
NTT Communications

Daichi Horita Wataru Shimoda Keiji Yanai
The University of Electro-Communications, Tokyo

ABSTRACT

In this demo, we demonstrate "Real-time Food Category Change" based on a Conditional Cycle GAN (cCycle GAN) with a large-scale food image data collected from the Twitter Stream. Conditional Cycle GAN is an extension of CycleGAN, which enables "Food Category Change" among ten kinds of typical foods served in bowl-type dishes such as beef rice bowl and ramen noodles. The proposed system enables us to change the appearance of a given food photo according to the given category keeping the shape of the given food but exchanging its textures. For training, we used two hundred and thirty thousand food images which achieved very natural food category change among ten kinds of typical Japanese foods: ramen noodle, curry rice, fried rice, beef rice bowl, chilled noodle, spaghetti with meat source, white rice, eel bowl, and fried noodle.

KEYWORDS

Food Category Change, Food Image Transformation, Conditional Cycle GAN, Food Image Generation

ACM Reference Format:
Ryosuke Tannno and Daichi Horita Wataru Shimoda Keiji Yanai. 2018. Magical Rice Bowl: A Real-time Food Category Changer. In *2018 ACM Multimedia Conference (MM '18), October 22–26, 2018, Seoul, Republic of Korea.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3240508.3241391

1 INTRODUCTION

In recent years, CycleGAN [7] is is drawing a lot of attention, which enables us to transform an given images to another domain image such as horse to zebra and edge images to paintings. The advantage of CycleGAN is that it can learn an image transformation model, which is represented with an encoder-decoder network, with unpaired training samples of two domains. However, CycleGAN has disadvantage that it can learn image transformation between only two fixed paired domains. This limitation makes it difficult that CycleGAN becomes more practical beyond fun.

Our objective is to make a system which takes a food image and a food category to be transferred as inputs, then outputs a new food image which corresponds to the given food category. To do that, we propose to extend CycleGAN by adding conditional inputs into conditional CycleGAN (cCycleGAN). In addition, to generate realistic images, the number of training images is the important key. We have gathered 230,000 food images which consist of 10 kind of food categories from Twitter stream for food image transformation. We have been keeping gathering images from the Twitter stream for more than eight years [5], and we mined the images corresponding to any of the ten food categories to create a large-scale food photo dataset for food category change. We show that it enabled high quality mutual transformation on a food domain with conditional CycleGAN (cCycle GAN). In addition, we show the number of the training images is important to get more realistic images.

In the most of the works related to GAN-based methods, a human face image dataset such as CelebA and a numeric character image dataset such as MNIST have been used as main target domains. Recently Jiang et al. applied a GAN-based image transformation to fashion style transfer [1]. On the other hand, there exists no work for a food image generation or transformation using GAN so far. In this demo, we propose food image transformation ,which converts a given food image to another category of a food image, as a new application of GAN-based image transformation. At the conference site, we will show a real-time food image transformation system working on smartphones as well as a note PC with a GPU.

We think food image transformation is promising from practical point of view in addition to being fun itself. For future work, we will combine virtual reality (VR) with this food image transformation, which will enables new eating experience. For example, when we are unable to eat high-calorie foods due to dietary restrictions, we can eat low-calorie foods while seeing high-calorie foods in VR glasses.

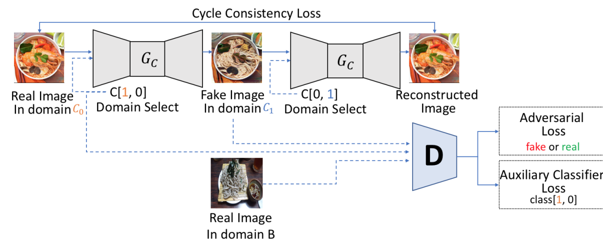

Figure 1: The architecture of the Conditional CycleGAN.

2 METHOD

2.1 Conditional CycleGAN

We show the network of Conditional CycleGAN (cCycleGAN) in Figure 1 which is an conditioned extension of CycleGAN. cCycleGAN can convert a given image to the image

MM '18, October 22–26, 2018, Seoul, Republic of Korea

ACM ISBN 978-1-4503-5665-7/18/10.
https://doi.org/10.1145/3240508.3241391

which belongs to the indicated category by adding a conditional input to an image transformation network of CycleGAN [7]. To use a conditional vector effectively, in cCycleGAN we added Auxiliary Classifier Loss L_{acl} [4] to the discriminator in addition to Adversarial Loss L_{adv}. The discriminator of cCycleGAN classifies not only real or fake but also a category of images. By the discriminator, a multi-class generator can be trained. The major characteristic of CycleGAN is using Cycle Consistency Loss L_{cyc} which forces the doubly-transformed image to be back the same as an given input image when it is transformed twice from class c via the other class c' back to c. Finally the loss of cCycleGAN is represented by a following equation:

$$L_{cyc} = \mathbb{E}_{x,c,c'}[||x - G(G(x,c),c')||_1] \quad (1)$$
$$L_{adv} = \mathbb{E}_x[\log D(y)] + \quad (2)$$
$$L_{acl}^{real} = \mathbb{E}_{x,c}[-\log D_{acl}(c|x)] \quad (3)$$
$$L_{acl}^{fake} = \mathbb{E}_{x,c}[-\log D_{acl}(c|G(x,c))] \quad (4)$$
$$L_{Dis} = -L_{adv} + \lambda_{acl} L_{acl}^{real} \quad (5)$$
$$L_{Gen} = L_{adv} + \lambda_{acl} L_{acl}^{fake} + \lambda_{cyc} L_{cyc} \quad (6)$$

where λ_{cyc} and λ_{acl} are weights for Cycle loss and Auxiliary classifier loss.

3 DATASET

By adding Cycle Consistency Loss, we can generate an image which keeps the original image structure. Therefore, in this experiments, we use a constrain to use images which have the same structure "bowl" so that corresponding structure prompts training of Cycle Consistency Loss. Actually we selected ten kind of categories related to "bowl" foods from UECFOOD-100 [3]. We gathered images from the large-scale food image dataset [5] which was created by mining food images from the twitter stream for more than eight years continuously. We sorted the images in the dataset [5] by using confidence scores obtained by a food classifier model which was trained with UECFOOD-100 dataset [3]. We selected the top 20% of images, because the top 20% images within each category are uniform and similar to each other, which are good property for training of food transformation. Finally we prepared 230,000 food images for ten categories in total. We show the ten bowl food categories and the number of selected images from re-ranked images in Table 1. The ratio of the train set is 90% and the ratio of the test set is 10% regarding the total amount of ten kinds of the bowl food images.

Table 1: Training data

food category	# images
ramen	74,007
curry rice	34,216
fried rice	27,854
fried noodles	24,760
white rice	21,324
beef bowl	18,396
chilled noodles	13,499
meat spaghetti	7,138
eel bowl	5,329
buckwheat noodle	3,530
TOTAL	230,053

Figure 2: The leftmost images are input images, and the other ones are generated regarding each of the ten categories.

4 EXPERIMENTS

4.1 Network and training setting

In the original CycleGAN [7], a generator network is the same to the network of Fast Style Transfer [2] which is added several Residual block to a standard Conv-Deconv Network. We proposed the conditional Fast Style Transfer Network [6] before which is the conditional extension of the Fast Style Transfer network. We use the same network of this for cCycleGAN. For training, the input image size is 256×256. As a conditional vector, we use a one-hot vector. After broadcasting the conditional vector to input image size, we concatenate it with an input image in the middle of the encoder part. After updating the discriminator five times, we update the generator one time. NVIDIA Quadro P6000 for training, bath size is 32, optimization method is Adam and iteration epoch is 20. On testing, we generate images with 512×512 resolution.

4.2 Results of food image transformation

We show the results by the proposed method in Fig.2. The left end image is the input image and other 10 images are the transformed images of each of the ten categories, respectively. Our proposed method can transform one certain category of an input to any of the other ten food categories clearly. We transformed given food images to the other food categories of images with keeping shape structure the Cycle Consistency Loss. This means that the generator trained the concept of "bowl". In addition, the generator generated an image which did not only fool the discriminator but also minimized the classification error of discriminator by Auxiliary Classifier Loss. We consider that Auxiliary Classifier Loss is also helpful for generating higher quality image than usual GAN. The images generated by using Auxiliary Classifier loss do not have blur which is frequently appeared if we use a simple GAN model. Note that additional results can be see at https://negi111111.github.io/FoodTransferProjectHP/.

REFERENCES

[1] S. Jiang and Y. Fu. 2017. Fashion Style Generator. In *Proc. of the Twenty-Sixth International Joint Conference on Artificial Intelligence*.

[2] J. Johnson, A. Alahi, and L.F. Fei. 2016. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In *Proc. of European Conference on Computer Vision*.

[3] Y. Matsuda, H. Hoashi, and K. Yanai. 2012. Recognition of Multiple-Food Images by Detecting Candidate Regions. In *Proc. of IEEE International Conference on Multimedia and Expo*.

[4] A. Odena, C. Olah, and J. Shlens. 2017. Conditional Image Synthesis With Auxiliary Classifier GANs. In *Proc. of the 34th International Conference on Machine Learning*.

[5] K. Yanai and Y. Kawano. 2014. Twitter Food Image Mining and Analysis for One Hundred Kinds of Foods. In *Proc. of Pacifit-Rim Conference on Multimedia (PCM)*.

[6] K. Yanai and R. Tanno. 2017. Conditional Fast Style Transfer Network. In *Proc. of ACM International Conference on Multimedia Retrieval*.

[7] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros. 2017. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. In *Proc. of IEEE International Conference on Computer Vision*.