Food Image Generation using A Large Amount of Food Images with Conditional GAN: RamenGAN and RecipeGAN

Yoshifumi Ito Wataru Shimoda Keiji Yanai
The University of Electro-Communications, Tokyo

ABSTRACT

Recently, image generation by Deep Convolutional Neural Network has been studied widely by many researchers. In this paper, we describe CNN-based image generation on food images. Especially, we focus on image generation using conditional Generative Adversarial Network (cGAN) with a large-scale dataset. In the experiments, we trained cGAN with a "ramen" image dataset and a recipe image dataset. For "ramen"GAN, we added a dish plate discriminator to make the shape of dishes rounder in generated images. For "recipe"GAN, we generated dish images from cooking ingredients, and tried image-based recipe search with generated images for the recipe database.

KEYWORDS

food image generation, DCGAN, conditional GAN, CNN

1 INTRODUCTION

In recent years, food images are strongly connected with our life. Many food images are posted to blogs or SNS on the Web, while we can search cooking recipe with various kinds of food images. The food images are very important to find target recipes on the Web. In addition, the quality of food images is also critical for increasing view count of recipe. On the other hand, recently Generative Adversarial Network (GAN) [1] which is a method to generate an image with Convolutional Neural Network is drawing a lot of attention.

In this paper, we describe experiments on generating food images with conditional GAN [2] with two kinds of the datasets. The first dataset is a "ramen noodle" image dataset which consists of six kinds of fine-grained ramen categories of images. For the first dataset, we propose a novel approach to use a discriminator of dish plate to make dishes more natural in generate images. For the second dataset, we propose to generate food images from food ingredients.

2 RELATED WORK

2.1 Generative Adversarial Network

Generative Adversarial Network (GAN) [1] is a generative model of deep neural network using two types of networks: a generator and a discriminator. While a generator generates an image, a discriminator classifies if generated images are real or fake. GAN trains two networks alternately, and the networks encourage each other.

The following equation shows the function to be minimized for training of a generator G and to be maximized for training of a discriminator D:

$$\min_G \max_D V(D,G) = E_x \sim p_{data(x)}\Big[logD(x)\Big] + E_x \sim p_{z(z)}\Big[log(1-D(G(z)))\Big]. \quad (1)$$

where x and z represent an input image and a random vector sampled from Gaussian, respectively. D and G are typically defined by deep convolutional neural networks.

2.2 Extension of GAN

There exist many works of extension of GAN. For example, LAPGAN [3] generates a high resolution image from a low resolution image by iterating of image generation. DCGAN [4] extended LAPGAN [3] and generated high resolution images with one iteration.

There are also several researches for stabilization of training of GAN. f-GAN [5] attempted to stabilize the training GAN by change of objective function. They replace JS divergence to f-divergence and achieved better performance. LSGAN [6] stabilized the training of GAN by using Euclidean loss instead of Sigmoid function. Wasserstein GAN (WGAN) [7] used Wasserstein distance, and generated a higher quality image. WGAN-GP [8] improved WGAN by taking a gradient penalty and stabilized the training of GAN. Progressive GAN [9] generated 1024×1024 images by starting from generation of low resolution image and enlarged resolution by stacking convolution layers gradually.

As we mentioned, there are various extensions of GAN. In this paper, we use Conditional GAN (cGAN) for conditioned food image generation, and use WGAN-GP for the experiments with the second dataset.

3 METHOD

First, we collect training images from the Twitter and CookPad dataset, and remove noise from the collected images by the existing food image classifier. Second, we label the collected images with food category labels based on textual information associated with the images, and we train Conditional GAN [2] with the images and labels. Finally, we generate images by the trained network.

The procedure of our work is as follows:.

(1) Collect labeled images.
(2) Train a Conditional Generative Adversarial Network (cGAN).
(3) Train a cGAN with a dish plate discriminator and a WGAN-GP.
(4) Generate images with the trained models.

CEA/MADiMa'18, July 15, 2018, Mässvägen, Stockholm, Sweden

ACM ISBN 978-1-4503-6537-6/18/07.
https://doi.org/10.1145/3230519.3230598

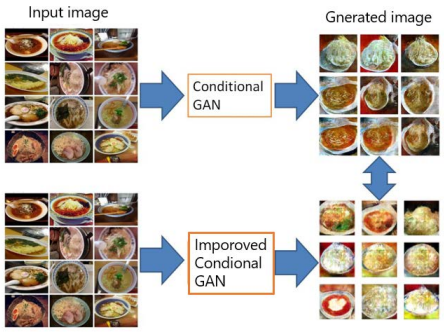

Figure 1: Overview of proposed method.

4 DETAIL OF METHOD

4.1 Collection of food image

In this section, we collect food image for training. In general,a GAN needs a large amount of training data for robust training. Therefore, we use Web existing large-scale food image datasets as data source.

4.1.1 Ramen image dataset. We have been collecting food images from the Twitter Stream for more than eight years. From the Twitter food image database we created, we extract "ramen noodle" images. removed noise images by recognition using existing food classifier trained with UEC-FOOD100 [10].

In this work, we adopt conditioned image generation with cGAN. To train cGAN, conditional labels are needed. We extract high-frequent keywords from the Twitter messages associated with the Twitter ramen images, and pick up the six names of representative fine-grained categories of ramen noodles which include "Plane ramen", "Jiro ramen", "Iekei ramen", "Spicy ramen", "Taiwan ramen", and "Onomichi ramen" as shown in Table 1. Some images are shown in Fig.2.

Table 1: The six fine-grained category of ramen noodles and the number of their images.

Category	Num
Plane ramen	790
Jiro ramen	5,901
Iekei ramen	2,836
Spicy ramen	3,578
Taiwan ramen	1,567
Onomichi ramen	1,228
TOTAL	15,900

4.1.2 CookPad dataset. As the second dataset, we use the CookPad dataset which is available only for academic research purpose [1]. CoodPad is a commercial service for cooking recipe search. This open dataset has more than one million recipe data with food images, and textual recipe information including names of recipes, ingredients, and procedures of cooking.

In the experiment, we use the CookPad dataset as training data for image generation with ingredient information. The dataset has 500,000 kind of recipes, and we use 127,690 kind of recipes which contain ingredient information except for seasoning. To use ingredients as conditions, we selected the following ten different representative ingredients as shown in Table 2.

[1] https://cookpad.com/

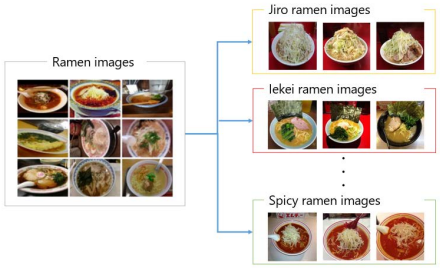

Figure 2: Images of the fine-grained ramen categories.

Table 2: The six fine-grained category of ramen noodles and the number of their images.

Category	Num	Category	Num
Onion	29,610	bacon	7,978
Carrot	22,450	red pepper	5986
Tomato	18,229	tofu	9,540
green pepper	8,143	chicken	7,759
mushroom	7,568	pork	10,427
		TOTAL	127,690

4.2 Conditional GAN

We cannot control images generated by standard GAN using class labels, since the standard GAN generate images from only noise seed vectors. To control food categories of generated images by using Conditional GAN [2]. Conditional GAN learns conditional probability by adding a conditional input which represents a category or some kinds of an attribute of images. This method provides conditional information with both a generator G and a discriminator D. The loss function of conditional GAN is represented as the following equation:

$$\min_G \max_D V(D,G) = E_x \sim p_{data(x)}\Big[logD(x|c)\Big] + E_x \sim p_{z(z)}\Big[log(1 - D(G(z|c)))\Big]. \quad (2)$$

where x, y and z represent an input image, a conditional vector which is commonly represented in a one-hot vector, and a random vector sampled from Gaussian, respectively.

At the time of training, as conditional information, we use fine-grained category labels for the first dataset, and ingredient labels for the second dataset.

4.3 Improvement of Conditional GAN

We introduce some extension of GAN to make generated images more clear.

The outline of dish plates of generated image should be round-shaped. However, they sometimes become distorted. To resolve this, we introduce an additional discriminator which classifies if a dish plate is round or nor. By introducing this additional discriminator to cGAN, the ratio of images with distorted dish plates is expected to reduce.

In addition, we introduce Wasserstein GAN with Gradient Penalty (WGAN-GP) [8] which was known as having ability to generate more clear images than usual GAN.

4.3.1 Discriminator of dish plate outline. We prepared an additional discriminator which is trained with oval figure images to avoid generating food images with distorted dish plates. In particular, we assume that the shape of dish plate is oval such like Fig.3, and we train a single class classifier with complete oval images as positive samples and the other figure images as negative samples. We use this classifier as an additional discriminator to make the shape of generate dish plates rounder. Fig.4 shows the illustration of network using an additional discriminator of dish plate outline.

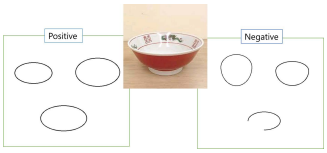

Figure 3: Discriminator of dish plates.

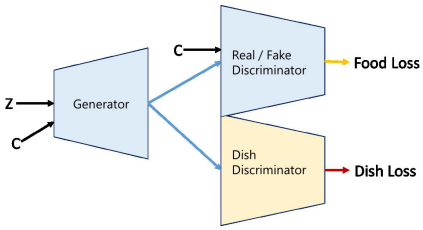

Figure 4: The network of Conditional GAN with the dish plate discriminator.

4.3.2 Use of Wasserstein GAN. We employed Wasserstein GAN (WGAN) [7] to obtain higher quality images. WGAN is an extension of GAN which minimizes Wasserstein distance between training sample distribution and generated sample distribution. The Wasserstein distance is used in transport theory as distance calculation method. On training WGAN, we round the weight of discriminator between $[-c, c]$. This operation is called as Weight Clipping. However, WGAN has several problems such like vanishing of gradients and slow training speed. These problems sometimes lead failure of training. WGAN-GP [8] has been proposed to resolve these problems. WGAN-GP employs a gradient penalty $\lambda E[(|\nabla D(\alpha x - (1-\alpha G(z)))| - 1)^2]$ for the loss. This penalty stabilizes training of GAN, and the loss of WGAN-GP is represented by the following equation:

$$L_D^{\mathrm{WGAN-GP}} = E[D(x|y)] - E[D(G(z|y))] + \lambda E[(|\nabla D(\alpha x - (1-\alpha G(z|y)))| - 1)^2] \quad (3)$$

$$L_G^{\mathrm{WGAN-GP}} = E[D(G(z|y))] \quad (4)$$

5 EXPERIMENTS

In the experiments, we used two kinds of datasets: A ramen image dataset collected from Twitter, and a recipe image dataset imported from the CookPad recipe database. We call two models trained with them "RamenGAN" and "RecipeGAN", respectively.

5.1 RamenGAN

With the first dataset containing "ramen noodle" images with the labels of the fine-grained categories, we trained Conditional GAN (cGAN) model with/without an additional discriminator of dish plates. We used a one-hot vector of a ramen category as a conditional vector for training of cGAN model. Therefore, the conditional vector is six dimension.

We show the results in Fig.5 without the dish constraint, while we show the generated images using discriminator of plate outline in Fig.6. The obtained images with dish loss have round dishes which are similar to ones the training samples have. On the other hand, the generated images without dish loss have relatively distorted dishes. This indicates that the generator without dish loss did not learn concept of "plate" correctly. This is partly because the size and location of ramen bowls are sometimes varies. On the other hand, the generated images using discriminator of plate have clearer outline than the previous ones. The additional discriminator is effective for learning the concept of "plate".

Figure 5: Generated ramen images using a conditional GAN model without dish loss. From the top row to the bottom row, "Jiro ramen" images, "Taiwan ramen" images and "Spicy ramen" images are shown, respectively.

Figure 6: Generated ramen images using a conditional GAN model with a discriminator of plate. From the top row, "Jiro ramen" images, "Taiwan ramen" images and "Spicy ramen" images are shown.

5.2 RecipeGAN

Differ from the specific ramen image generation, image generation using recipe data is difficult due to the diversity of the dataset. The difficulty of training of standard GAN with diverse images is that generated images tends to get similar to each other which is called as mode collapse. We show the example in Fig.7. On the other hand, WGAN-GP is more robust to mode collapse and can improve the quality of image generation as shown in Fig.8.

We evaluate the results of food image generation by recipe image retrieval using the generated images. We show image-based recipe search results which are retrieved by the generated images. While the result of Conditional GAN is shown in Fig.9, the result of Conditional WGAN-GP is shown in Fig.10. This results indicates that the searched result of WGAN-GP is more reflected to the input recipe than the result of a simple cGAN.

Figure 7: Recipe images generated with cGAN. From the top row, green pepper dishes, tomato dishes, and chicken dishes.

Figure 8: Recipe images generated with conditional WGAN-GP. From the top row, green pepper dishes, tomato dishes, and chicken dishes.

Figure 9: The results of image search using the generated images by Conditional GAN.

6 CONCLUSIONS

In this paper, we described food image generation using conditional GAN. Especially, we attempted image generation for specific category such like ramen dataset and image generation for recipe images trained with ingredient dataset which have large diversity. In addition, we showed that dish discriminator and WGAN-GP are effective for food image domain.

Figure 10: The results of image search using the generated images by Conditional WGAN-GP.

As future works, we consider improvement of further quality of image generation with datasets which has large diversity. To do that, we need to clean up dataset by dividing current category into more detailed category. To change the method of GAN to progressive GAN will be also effective for generating of larger-size images. Though we evaluate proposed method by subjective view in this paper, we need to evaluate the method by objective experiment such like inception score.

Acknowledgements: This work was supported by JSPS KAKENHI Grant Number 15H05915, 17H01745, 17H05972, 17H06026 and 17H06100. In this paper, we used recipe data provided by CookPad and the National Institute of Informatics.

REFERENCES

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, Generative Adversarial Nets, Advances in Neural Information Processing Systems, pp. 2672–2680, 2014.
[2] M. Mirza and S. Osindero, Conditional Generative Adversarial Nets, arXiv:1411.1784, 2014.
[3] E. Denton, S. Chintala, A. Szlam and R. Fergus, Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks, Advances in Neural Information Processing Systems, 1486–1494, 2015.
[4] A. Radford, L. Metz and S. Chintala, Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, Proc. of International Conference on Learning Representations, 2016.
[5] S. Nowozin, B. Cseke and R. Tomioka, f-GAN: Training Generative Neural Samplers using Variational Divergence Minimization, Advances in Neural Information Processing Systems, 271–279, 2016.
[6] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang and S. P. Smolley, Least Squares Generative Adversarial Networks, Proc.of IEEE International Conference on Computer Vision, 2813–2821, 2017.
[7] M. Arjovsky, S. Chintala and L. Bottou, Wassersstein GAN, arXiv:1701.07875, 2017.
[8] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin and A. C. Courville, Improved Training of Wasserstein GANs, Advances in Neural Information Processing Systems, 2769–5779, 2017.
[9] T. Karras, S. Laine and J. Lehtinen, Progressive Growing of GANs for Improved Quality, Stability, and Variation, arXiv:1710.10196, 2017.
[10] Y. Matsuda, H. Hoashi and K. Yanai, Recognition of Multiple-Food Images by Detecting Candidate Regions, Proc. of IEEE International Conference on Multimedia and Expo (ICME), 2012.
[11] T. Salimans, I. Goodfellow, W. Zaremba , V. Cheung, A. Radford and X. Chen, Improved Techniques for Training GANs, Advances in Neural Information Processing Systems, 2234–2242, 2016.