Unseen Food Creation by Mixing Existing Food Images with Conditional StyleGAN

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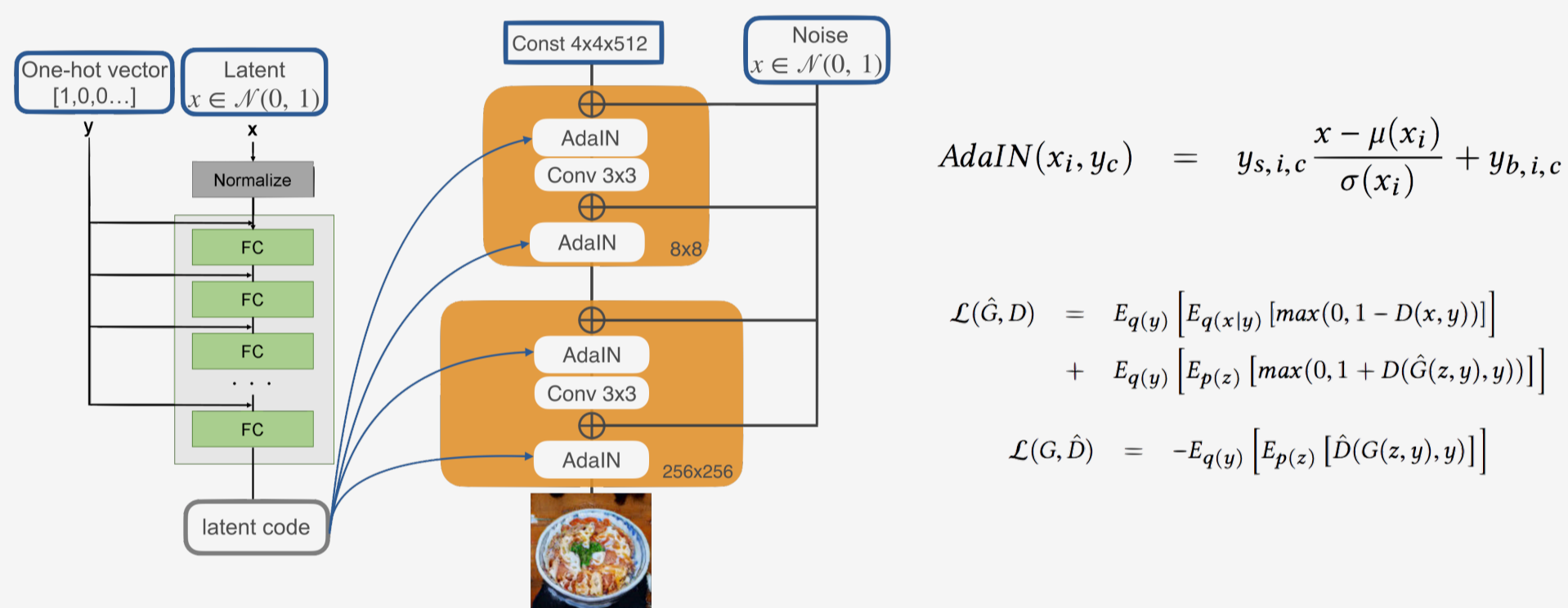
The University of Electro-Communications, Tokyo, Japan

Background

- **StyleGAN**[1] can generate high-quality images in an unsupervised manner.
- However, it is difficult to control latent space…

➡ Conditional version of StyleGAN.

➡ Our model can create the unseen food images by mixing multiple kinds of foods!

One   Two   Three conditions
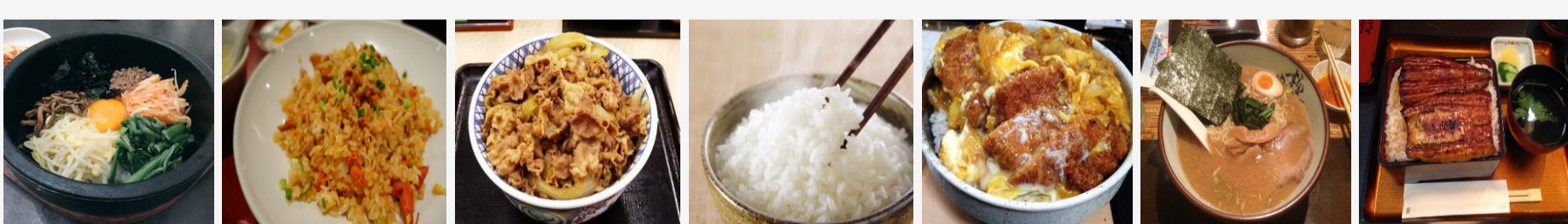
Proposed Approach

- Condition enables to control the food categories.
- Given latent code z and the condition y, the mapping function f converts to conditional style $(y_{s,i,c}, y_{b,i,c})$ to control AdaIN.
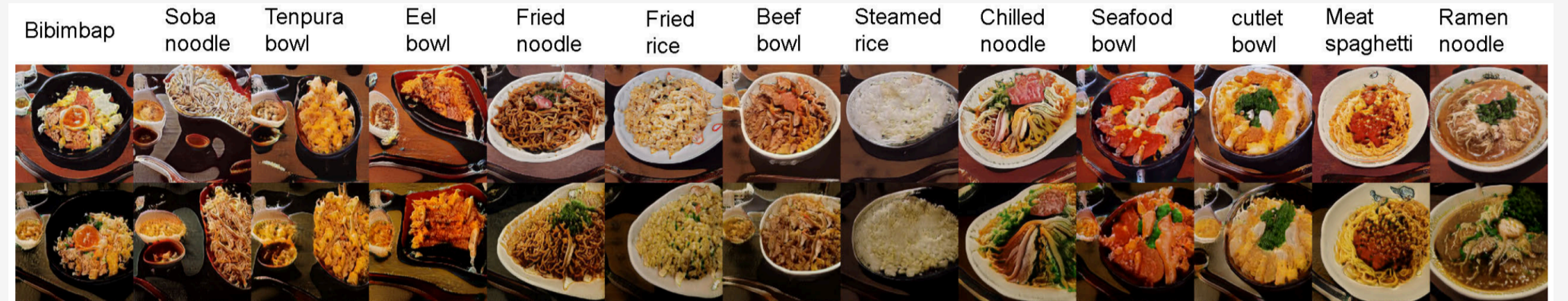
$$AdaIN(x_i, y_c) = y_{s,i,c} \frac{x - \mu(x_i)}{\sigma(x_i)} + y_{b,i,c}$$

$$\mathcal{L}(\hat{G}, D) = E_{q(y)}\left[E_{q(x|y)}[max(0, 1 - D(x, y))]\right]$$
$$+ E_{q(y)}\left[E_{p(z)}[max(0, 1 + D(\hat{G}(z, y), y))]\right]$$

$$\mathcal{L}(G, \hat{D}) = -E_{q(y)}\left[E_{p(z)}[\hat{D}(G(z, y), y)]\right]$$

Dataset

- The dataset consists of 13 categories and 220k images.
- Test set consists of 1000 images of each category.

Category	# number of images
bibimbap	9433
fried rice	28406
beef bowl	9720
steamed rice	6387
ramen noodle	80000
eel bowl	5100
fried noodle	25000
pork cutlet bowl	10000
chilled noodle	13600
seafood bowl	10000
tempura bowl	10000
meat spaghetti	7000
soba noodle	3300
total	227946

Experiments

1. Manipulation of Latent Space

Bibimbap | Soba noodle | Tempura bowl | Eel bowl | Fried noodle | Fried rice | Beef bowl | Steamed rice | Chilled noodle | Seafood bowl | cutlet bowl | Meat spaghetti | Ramen noodle
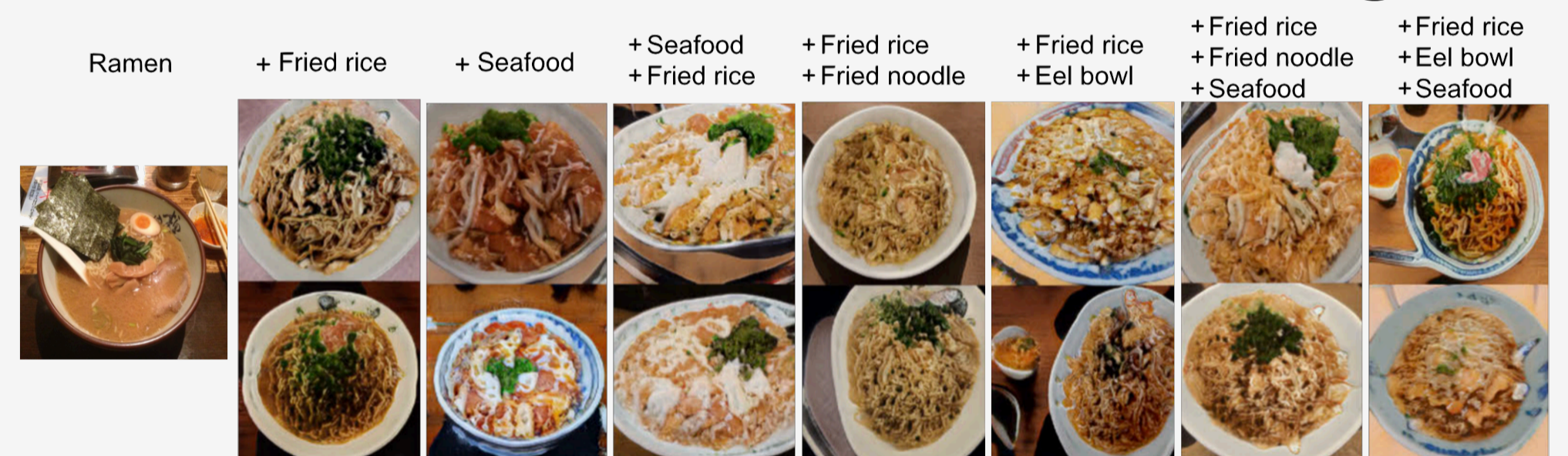
- Images generated from fixed input noises and styles.
- Each row of images has a similar shape and style.
- Soba, Tempura and Eel bowl have distorted plate because they have few round samples…

2. Manipulation of Random Noise

- Random noise plays a role in differences such as food topping.
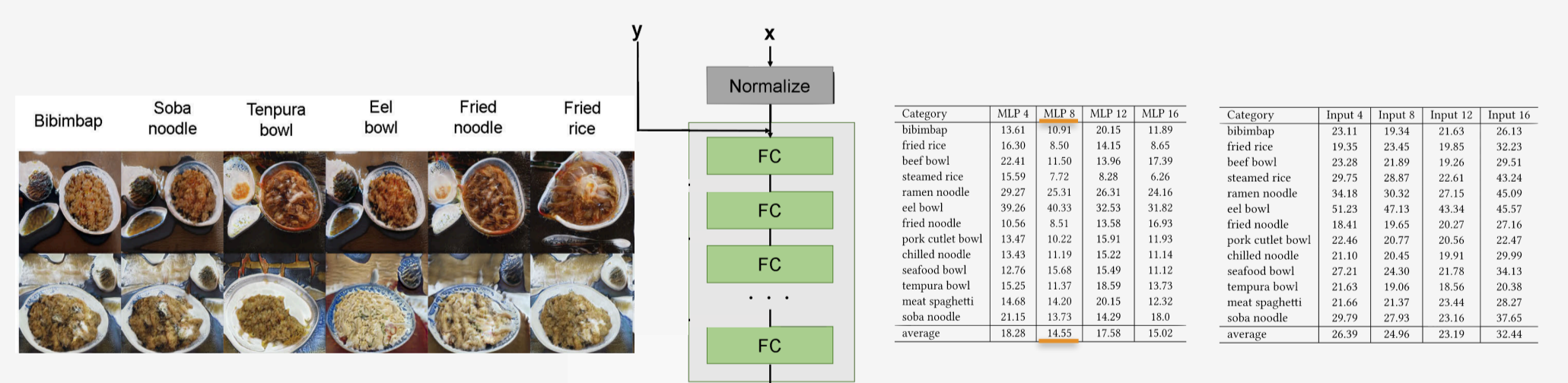- Egg number, appearance, and seaweed are affected.

3. Creation of Unseen Food Image

Ramen | + Fried rice | + Seafood | +Seafood +Fried rice | + Fried rice + Fried noodle | + Fried rice + Eel bowl | + Fried rice + Fried noodle + Seafood | + Fried rice + Eel bowl + Seafood

- Our model can create the unseen food images by mixing multiple conditions.

Ablation Study

Bibimbap | Soba noodle | Tempura bowl | Eel bowl | Fried noodle | Fried rice

Category	MLP 4	MLP 8	MLP 12	MLP 16
bibimbap	13.61	10.91	20.15	11.89
fried rice	16.30	6.58	14.15	8.65
beef bowl	22.41	11.58	13.96	17.39
steamed rice	15.59	7.72	8.28	6.26
ramen noodle	29.27	25.31	26.31	24.16
eel bowl	39.26	49.31	32.53	31.82
pork cutlet bowl	50.56	8.51	13.58	16.93
chilled noodle	13.43	11.19	15.22	11.14
seafood bowl	12.76	15.68	15.49	11.12
tempura bowl	15.25	11.37	18.59	13.73
meat spaghetti	14.68	14.28	20.15	12.52
soba noodle	21.15	13.73	14.29	18.0
average	18.28	14.35	17.58	15.02

Category	Input 4	Input 8	Input 12	Input 16
bibimbap	23.11	19.34	21.63	26.13
fried rice	19.35	23.45	19.85	32.23
beef bowl	23.28	23.89	19.26	29.51
steamed rice	29.75	28.87	22.61	45.24
ramen noodle	34.18	30.32	27.15	45.09
eel bowl	51.25	47.13	43.34	45.57
fried noodle	18.41	19.65	29.27	27.16
pork cutlet bowl	24.46	20.77	20.56	22.47
chilled noodle	21.30	20.45	19.91	29.99
seafood bowl	27.21	24.30	21.78	34.13
tempura bowl	21.63	19.06	18.56	20.38
meat spaghetti	21.66	21.37	23.44	28.27
soba noodle	29.79	27.95	23.19	37.65
average	26.39	24.96	23.19	32.44

- Can not learn with inputting the condition to only the first layer.
- FIDs is the lowest when the mapping network consists of eight FC layers

References

[1] Tero Karras, Samuli Laine, and Timo Aila. A Style-Based Generator Architecture for Generative Adversarial Networks. In Proc.of IEEE Computer Vision and Pattern Recognition(CVPR), 2019.

Conclusion

- Conditional version of StyleGAN to control latent space and to create unseen food images.