Unseen Food Creation by Mixing Existing Food Images with Conditional StyleGAN

MADiMa on ACMMM workshop
Oct 21st, 2019
Nice, France

Daichi Horita, Wataru Shimoda and Keiji Yanai
UEC Tokyo, Japan.

Unseen Food Creation by Mixing Existing Food Images with Conditional StyleGAN

MADiMa on ACMMM workshop
Oct 21st, 2019
Nice, France

Daichi Horita, Wataru Shimoda and Keiji Yanai
UEC Tokyo, Japan.

1. Motivation

- **GANs** show remarkable success in various tasks such as **image generation** and **image translation** and have been applied to food image generation.

RamenGAN[Ito+ MADiMa18]

R2GAN[Zhu+ CVPR19]

1. Motivation

- **However**, both GANs fails to generate realistic food images.

→ The quality is still low and it is difficult to generate appetizing and delicious-looking images.

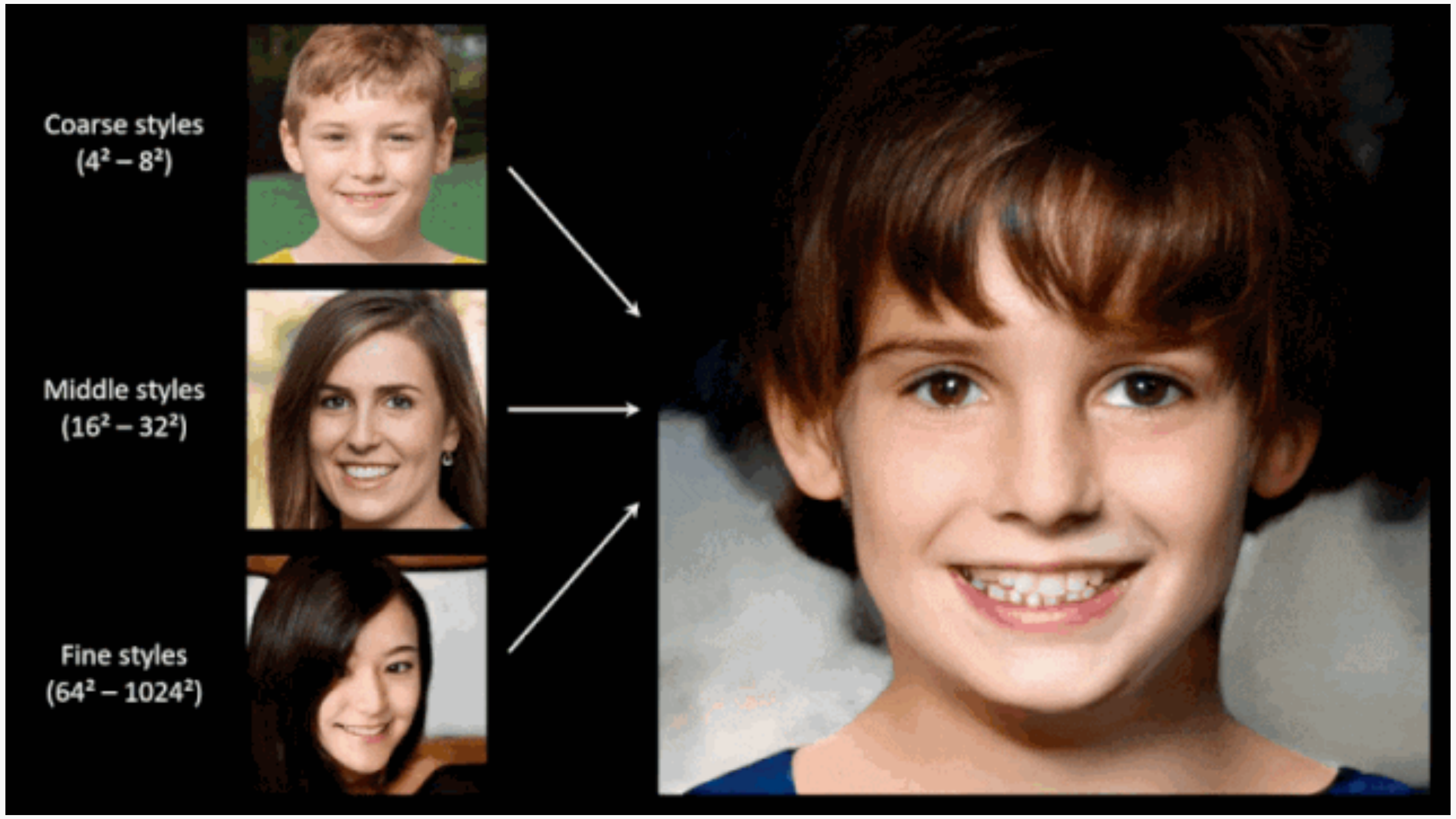
RamenGAN[Ito+ MADiMa18]

R2GAN[Zhu+ CVPR19]

1. Motivation

- **StyleGAN**[Karras+ CVPR19] is one of the state-of-the-art GANs in an unsupervised manner.
- **However**, it is difficult to control the latent space…

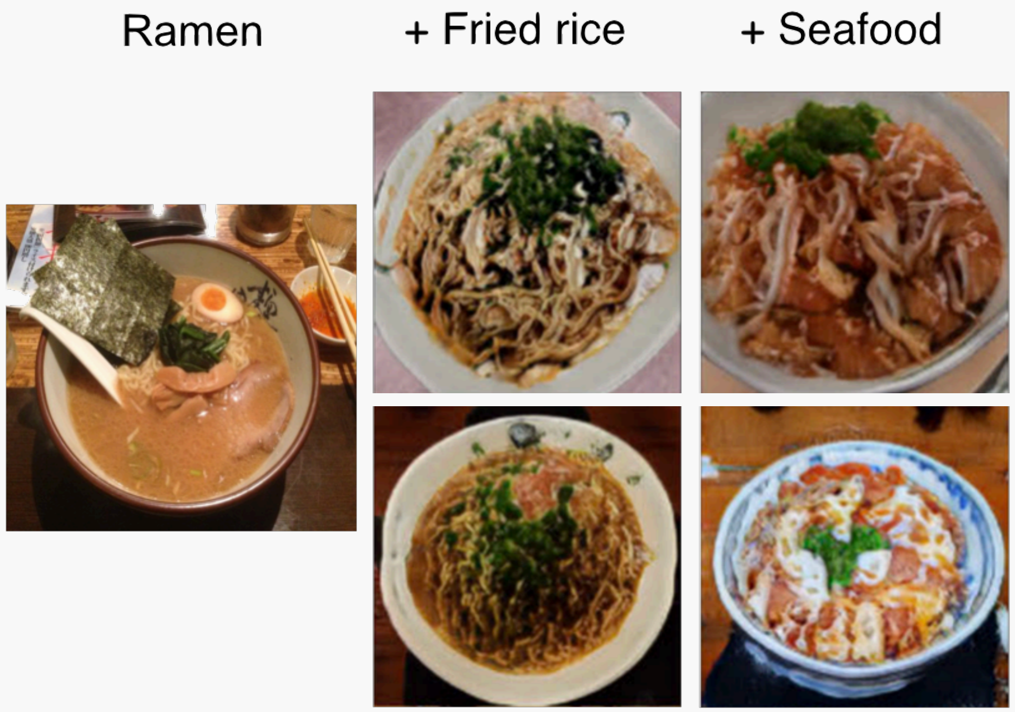
➡ **We extend the generator to manipulate the latent space**



1. Motivation

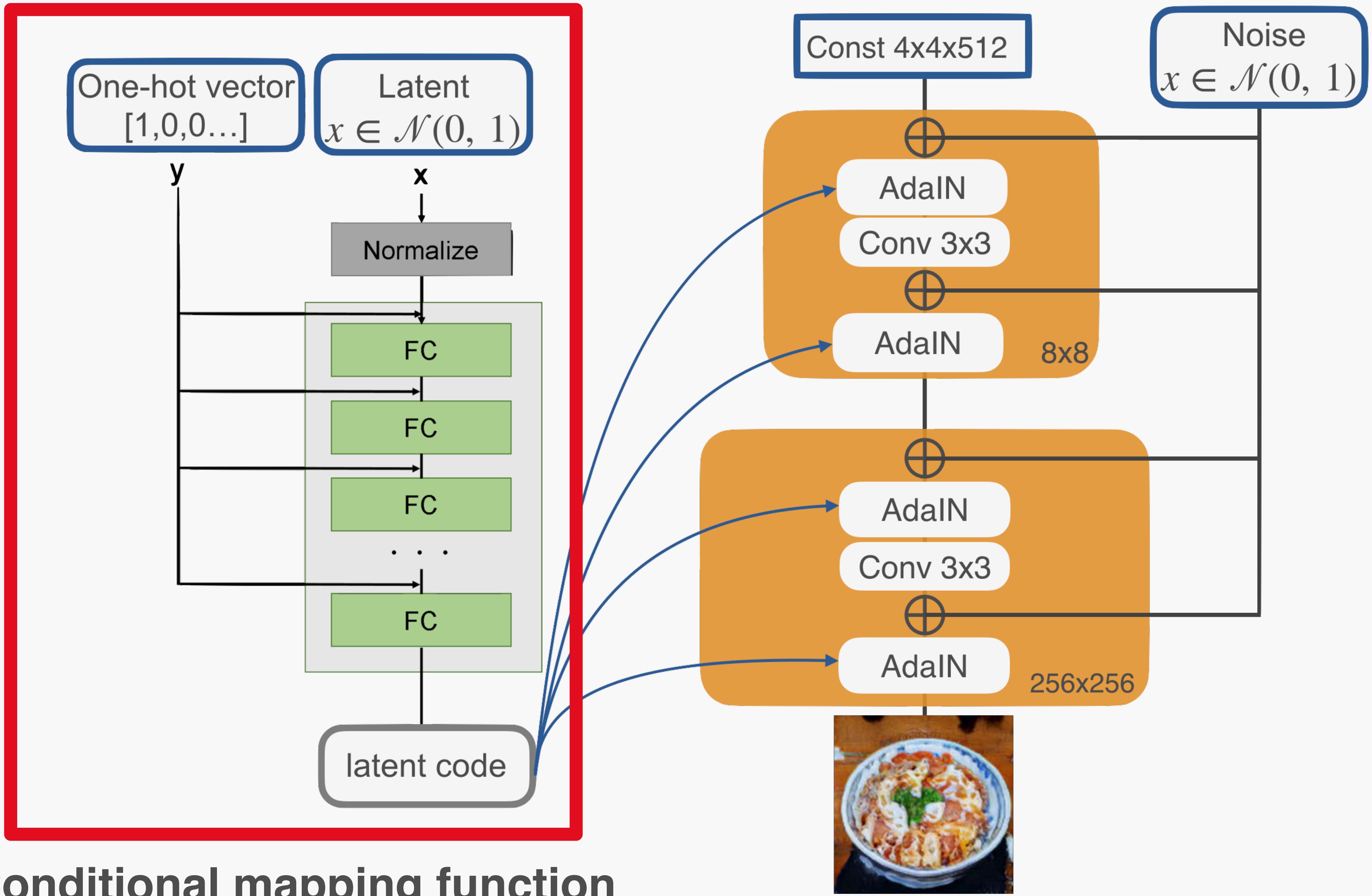
- Thanks to introduce conditional vectors, our proposed model can **generate the specific food** and **create the unseen food images by mixed multiple kinds of foods!**

Images generated by one condition.



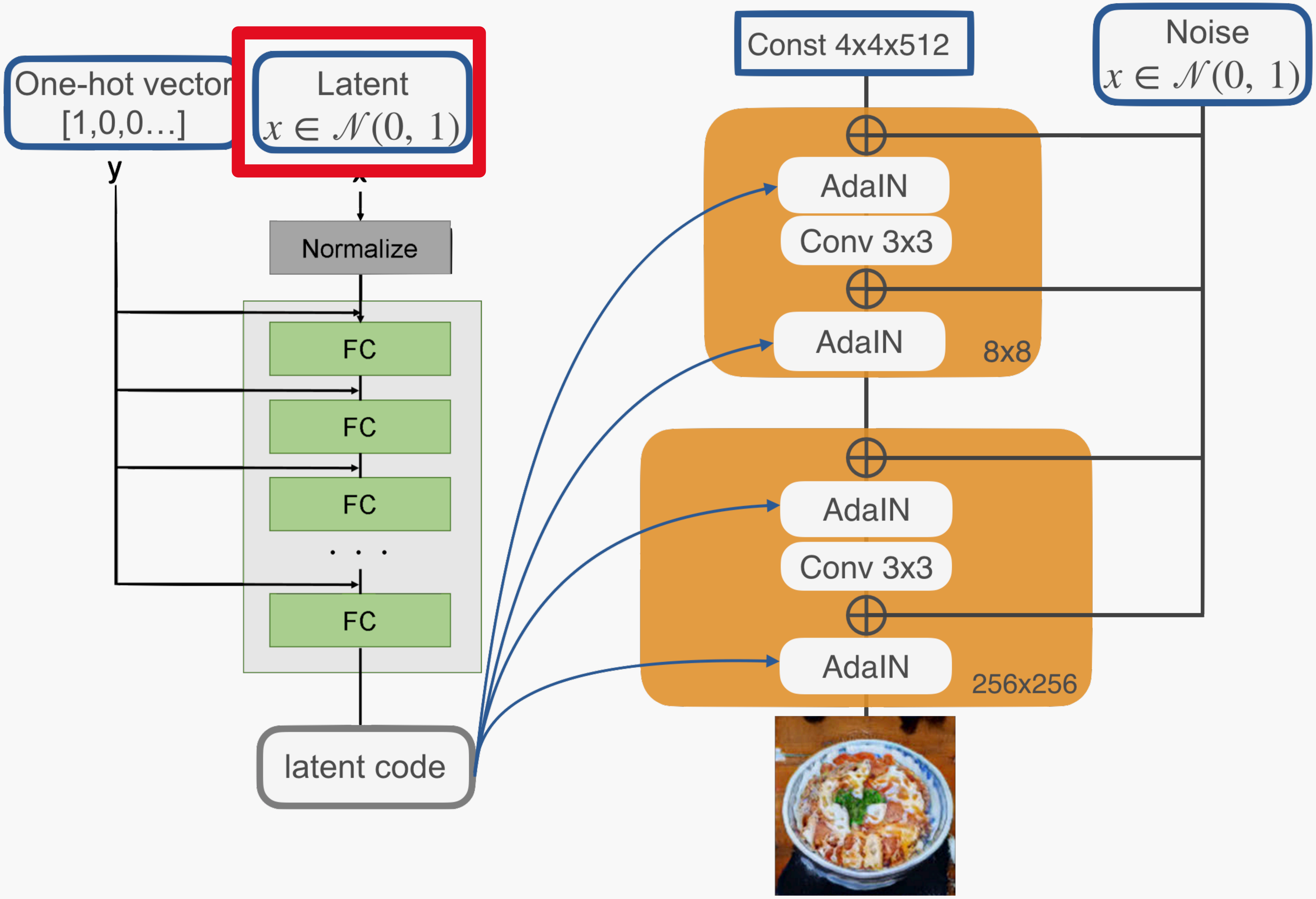
Images generated by mixing conditions.

2. Approach - Conditional style-based generator

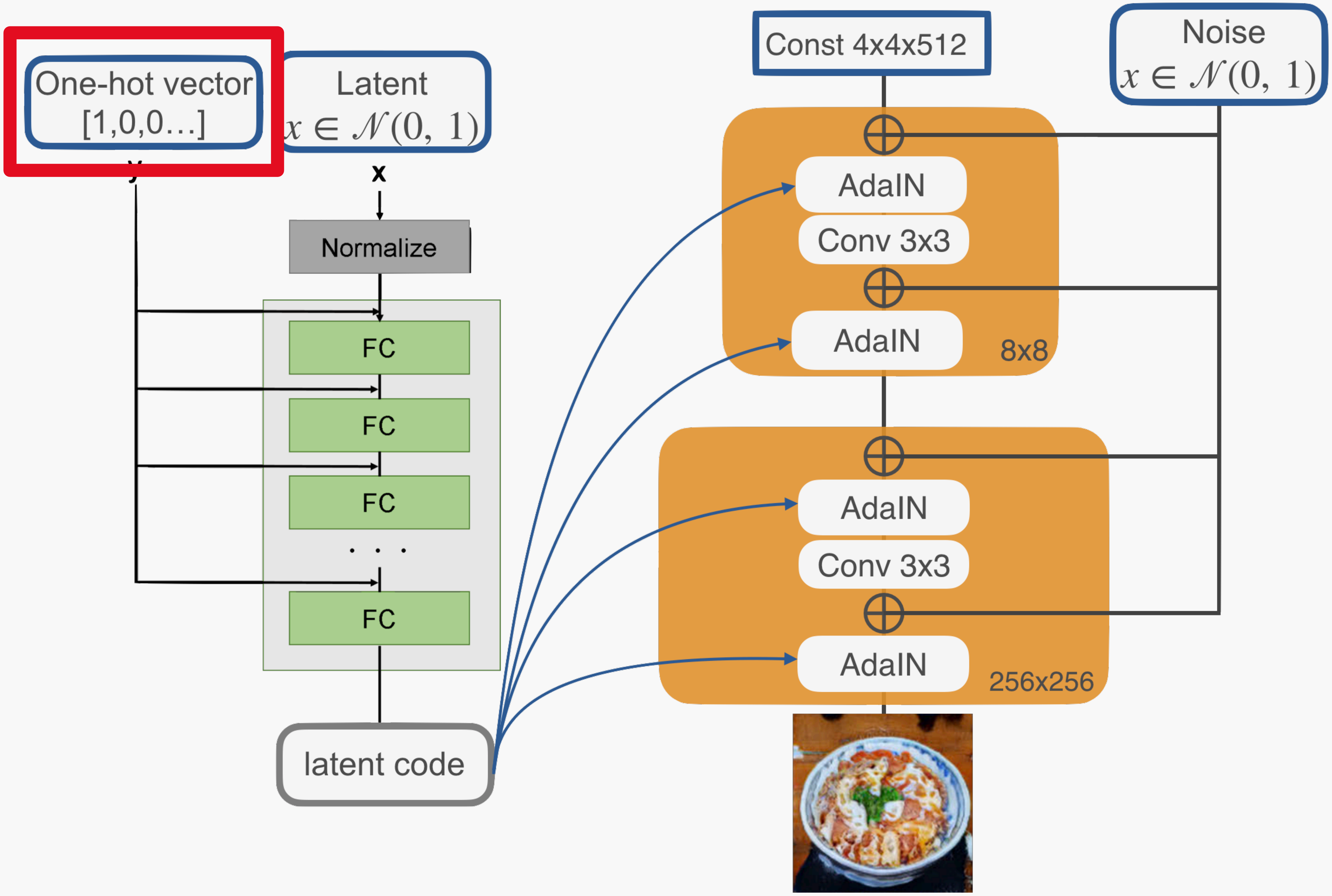


Conditional mapping function

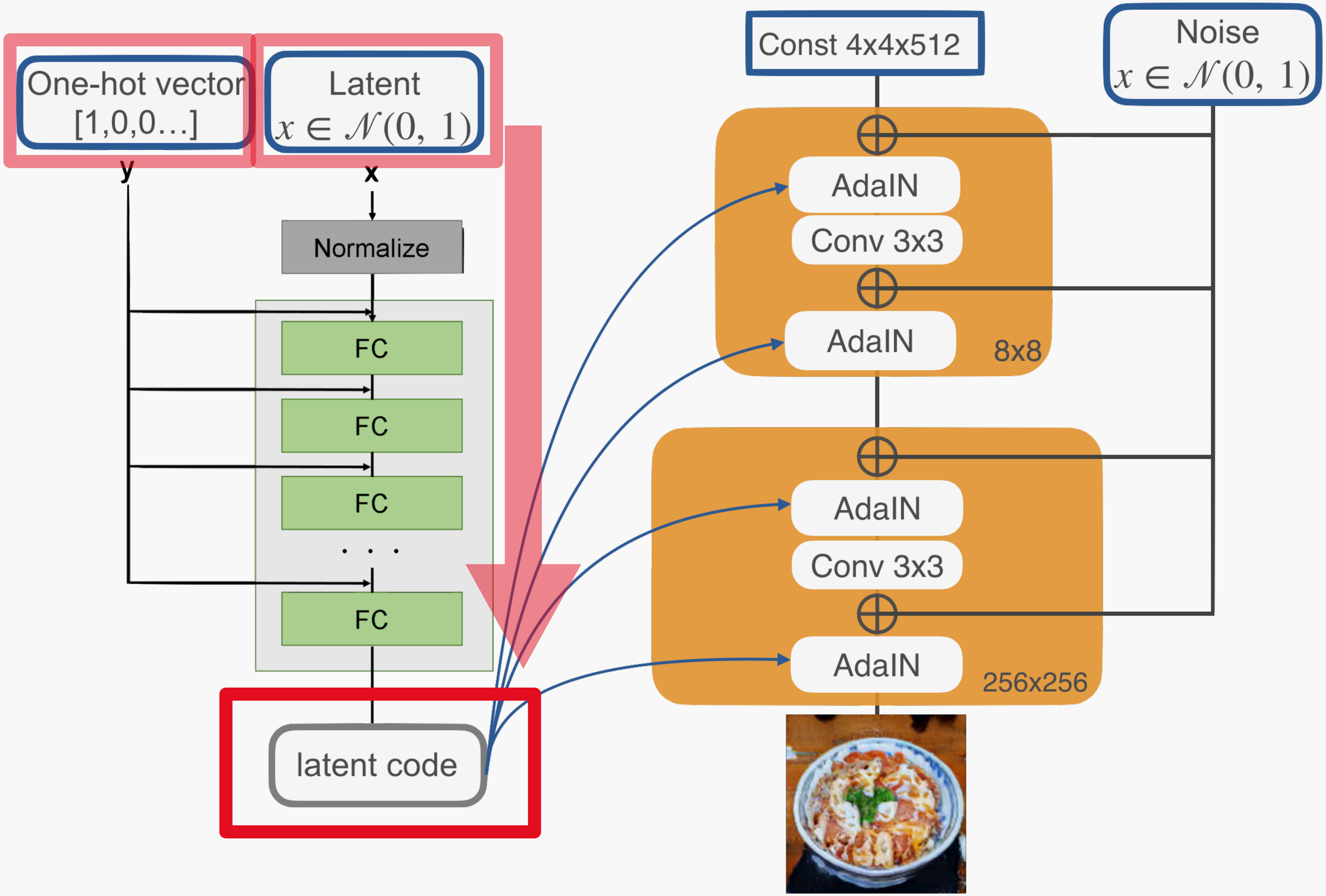
2. Approach - Conditional style-based generator



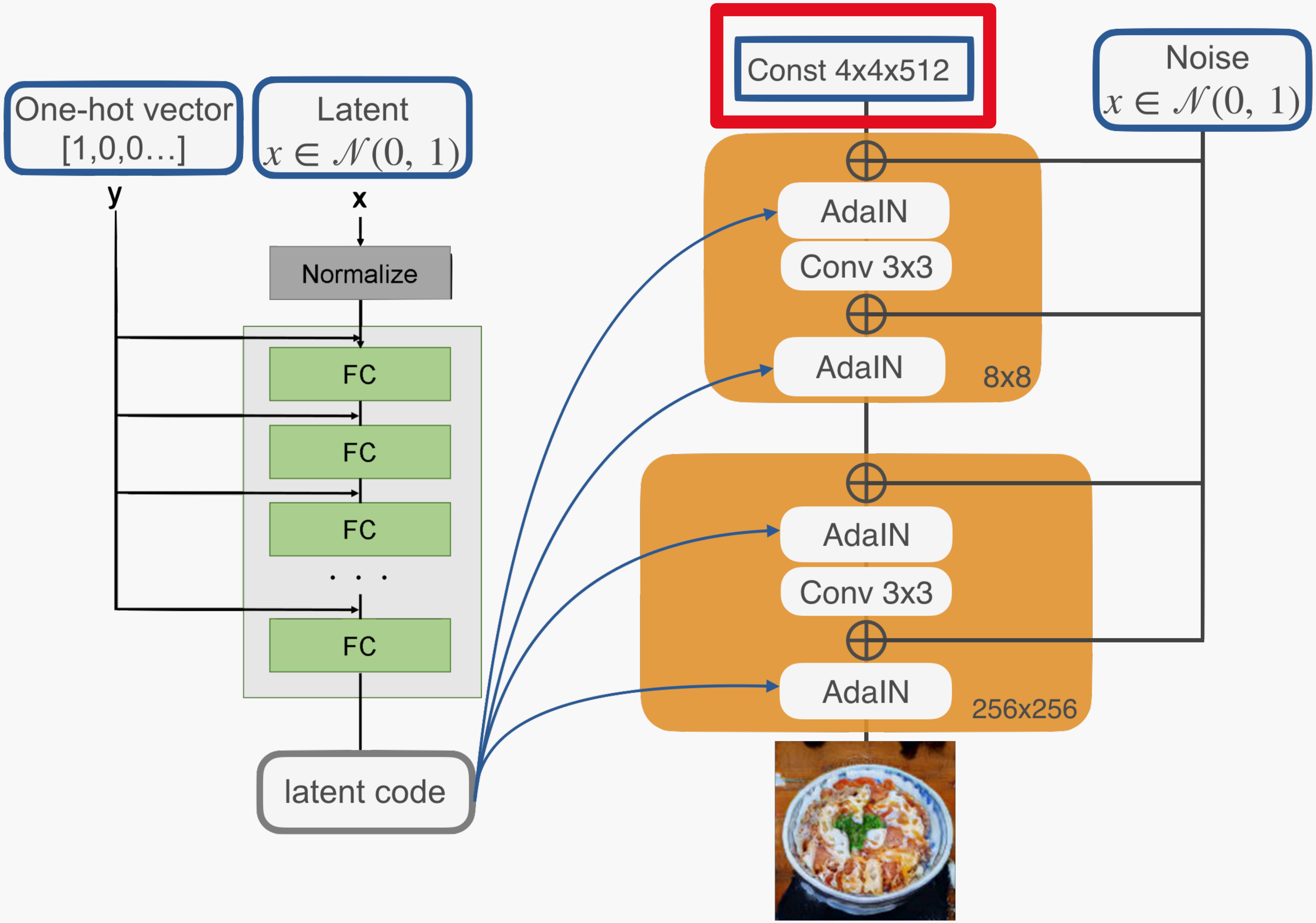
2. Approach - Conditional style-based generator



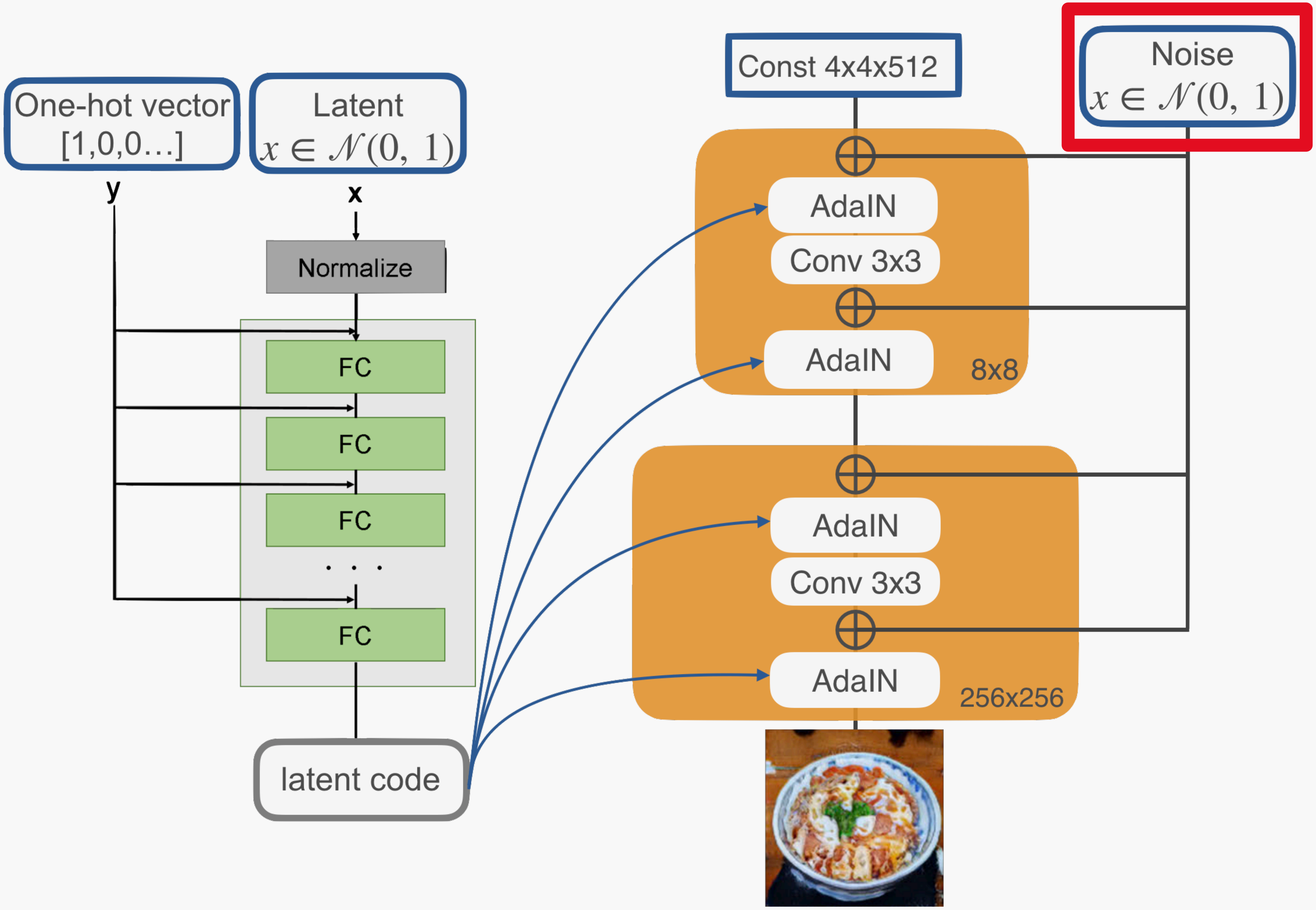
2. Approach - Conditional style-based generator



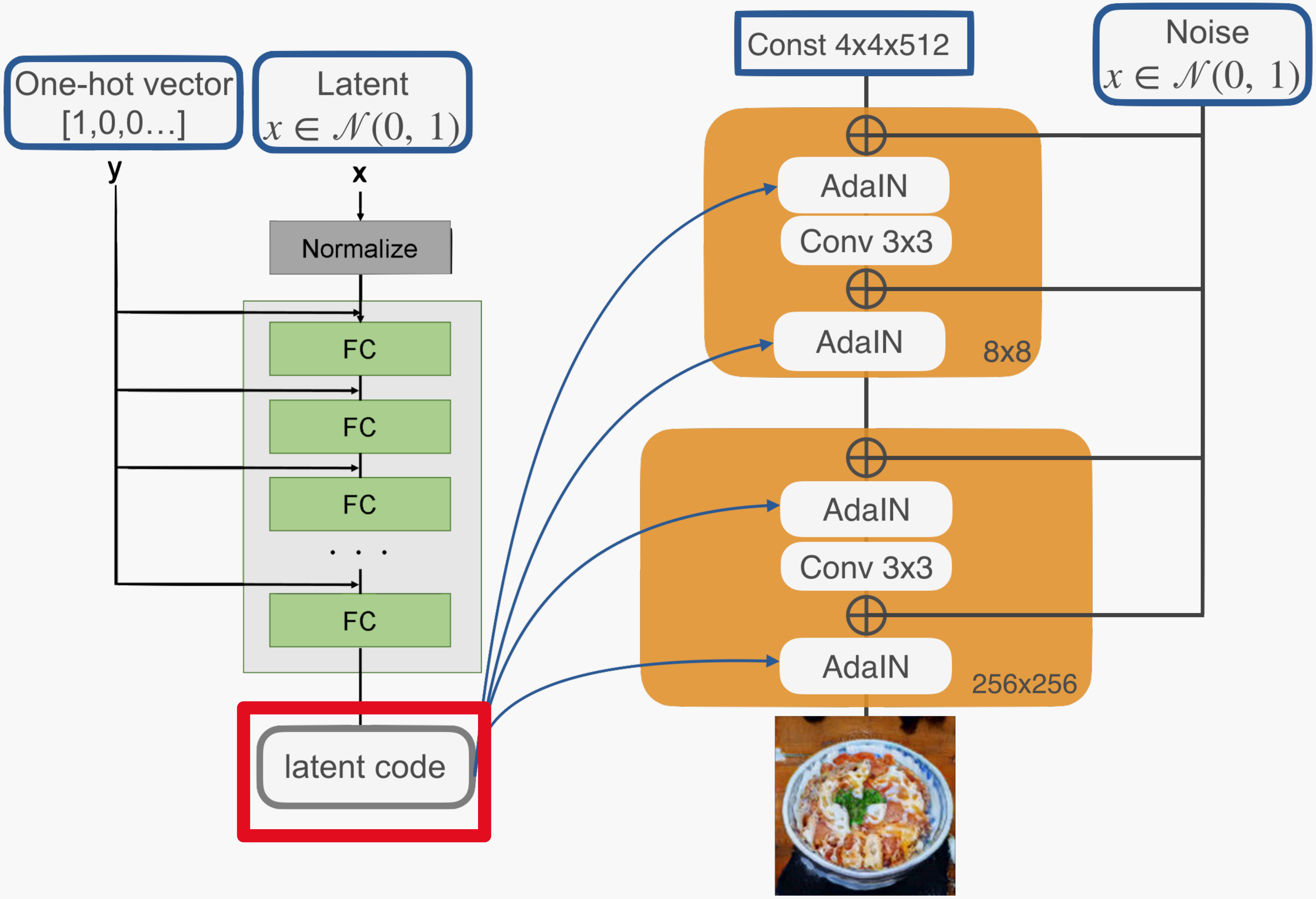
2. Approach - Conditional style-based generator



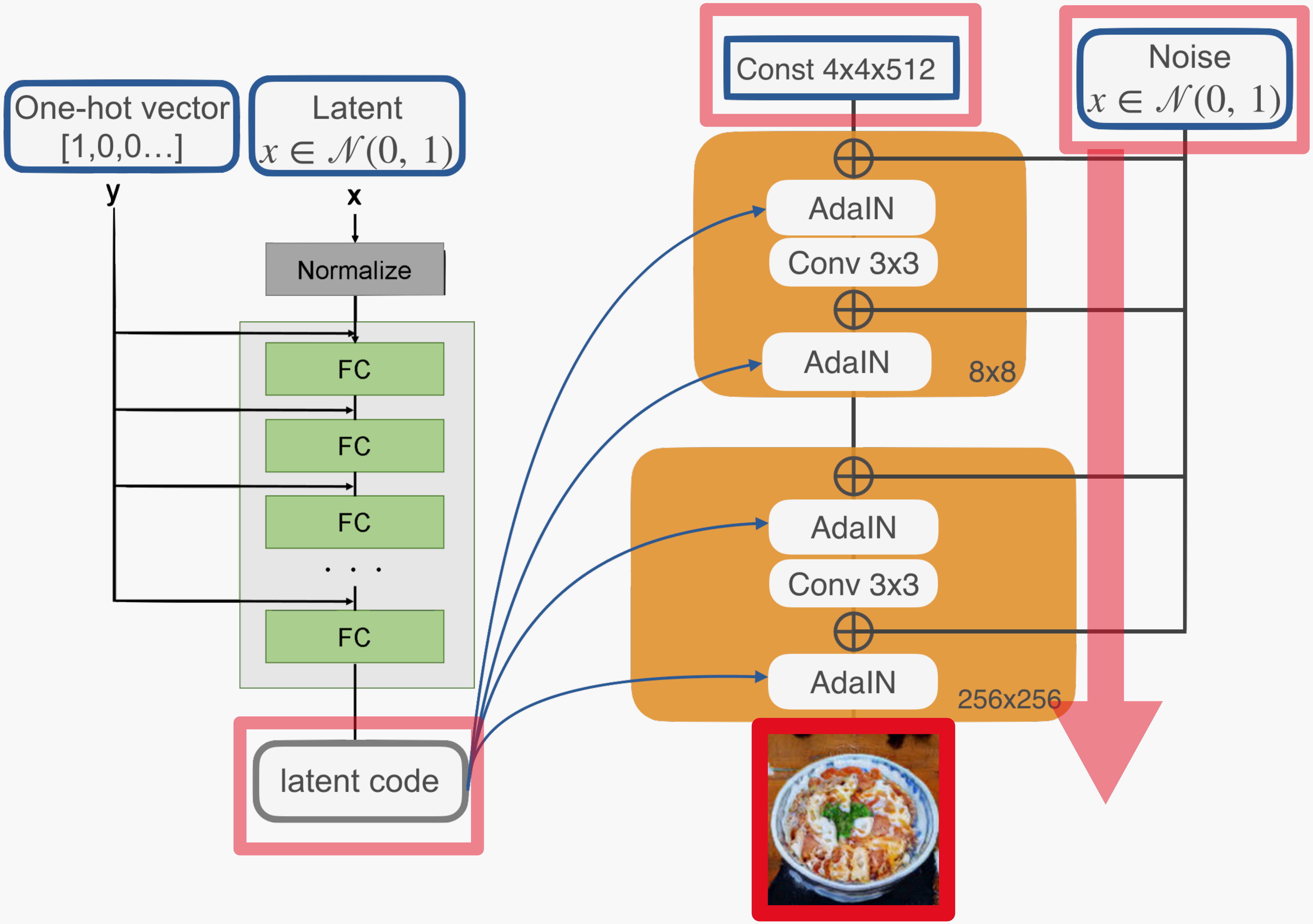
2. Approach - Conditional style-based generator



2. Approach - Conditional style-based generator

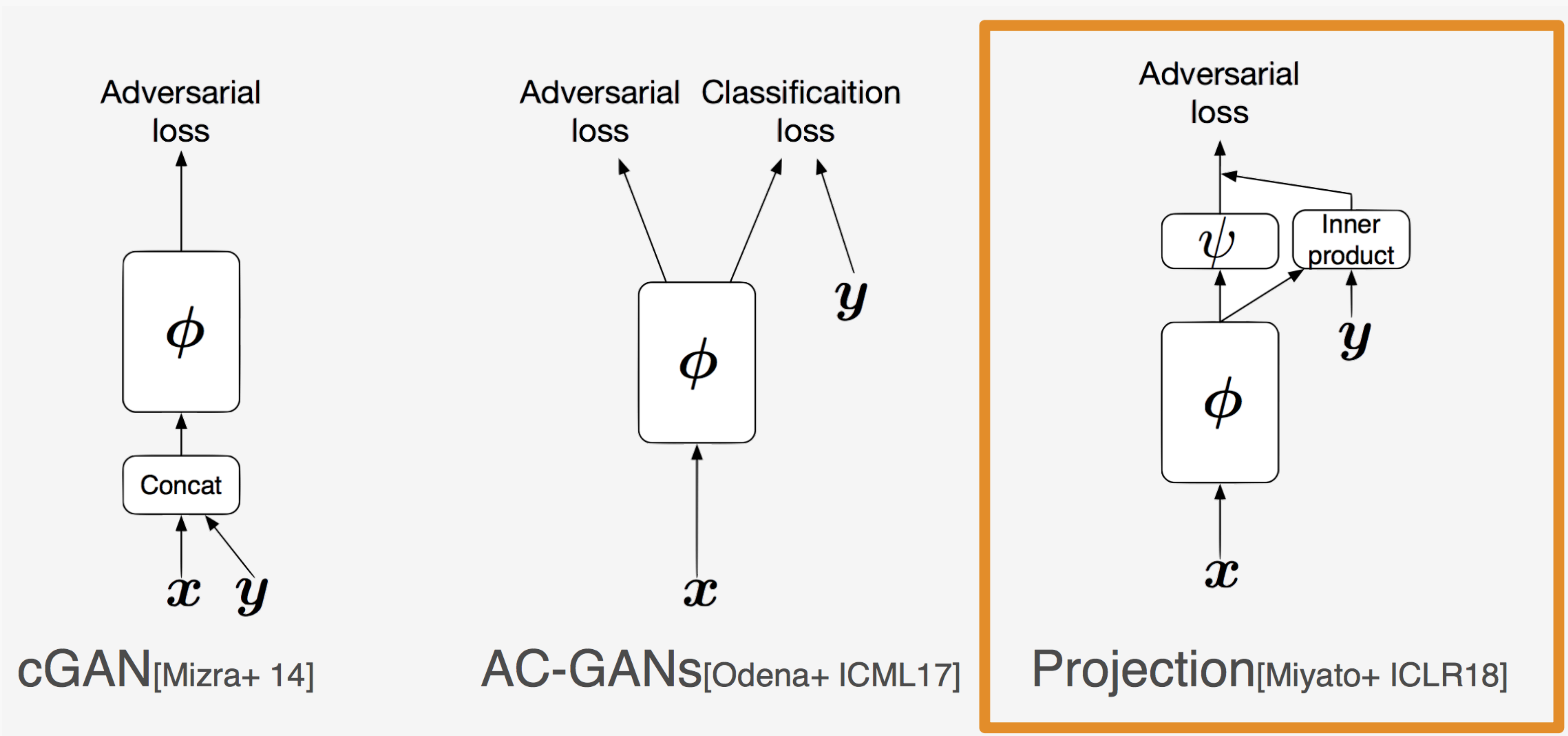


2. Approach - Conditional style-based generator



2. Approach - Conditional discriminator

- We adopt Projection Discriminator[Miyato+ ICLR18]
- cGAN and AC-GANs didn't progress learning.



Discriminator architecture (This figure is borrowed from [Miyato+ ICLR18])

3.1 Experiments - FOOD13 Dataset

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

Bibimbap

Fried rice

Beef bowl

Pork cutlet bowl

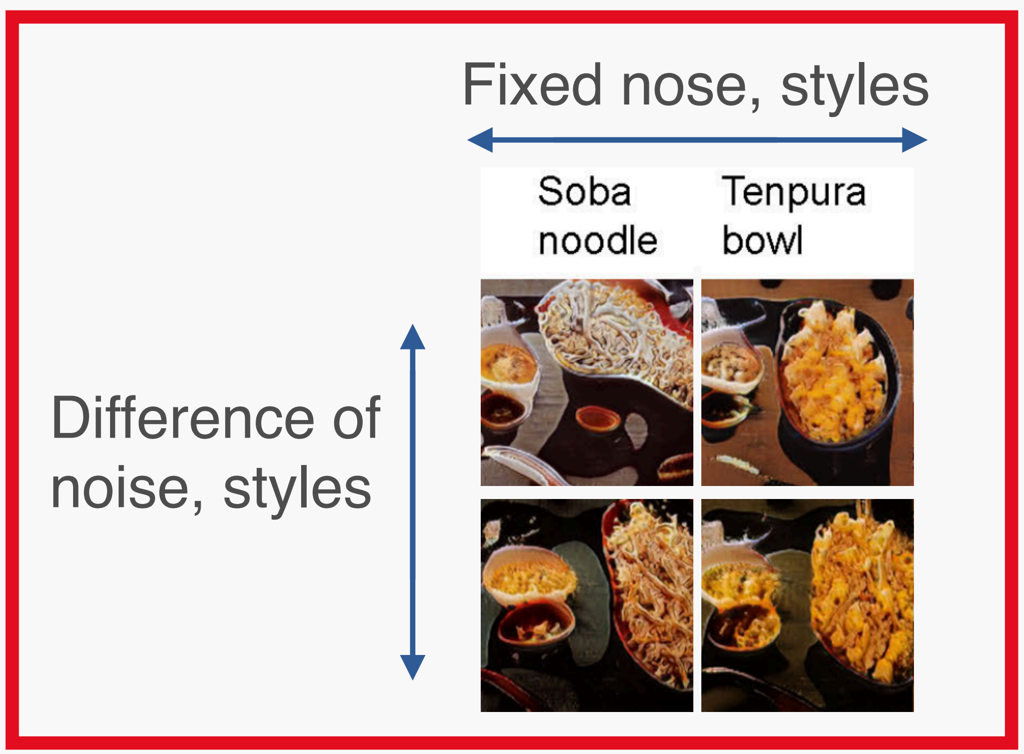
Ramen

Eel bowl

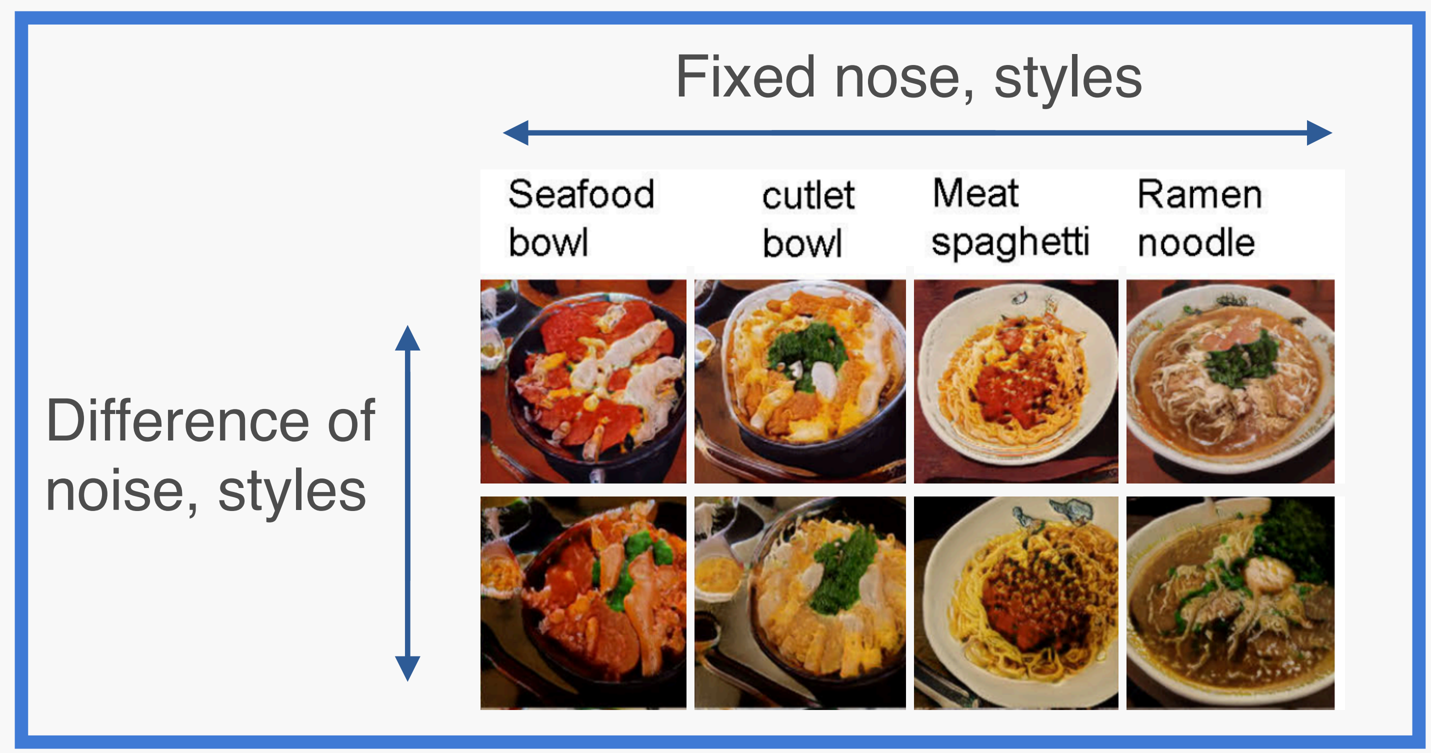
Category	# number of images
bibimbap	9433
fried rice	28406
beef bowl	9720
steamed rice	6387
ramen noodle	80000
eel bowl	5100
fried noodle	25000
pork cutlet bowl	10000
chilled noodle	13600
seafood bowl	10000
tempura bowl	10000
meat spaghetti	7000
soba noodle	3300
total	227946

3.2 Experiments - Manipulation of latent space

- Each row of images are generated from **the fixed input noise, styles** and **one condition vectors.**
- Our model can generate **an arbitrary class of images thanks to the condition vectors.**



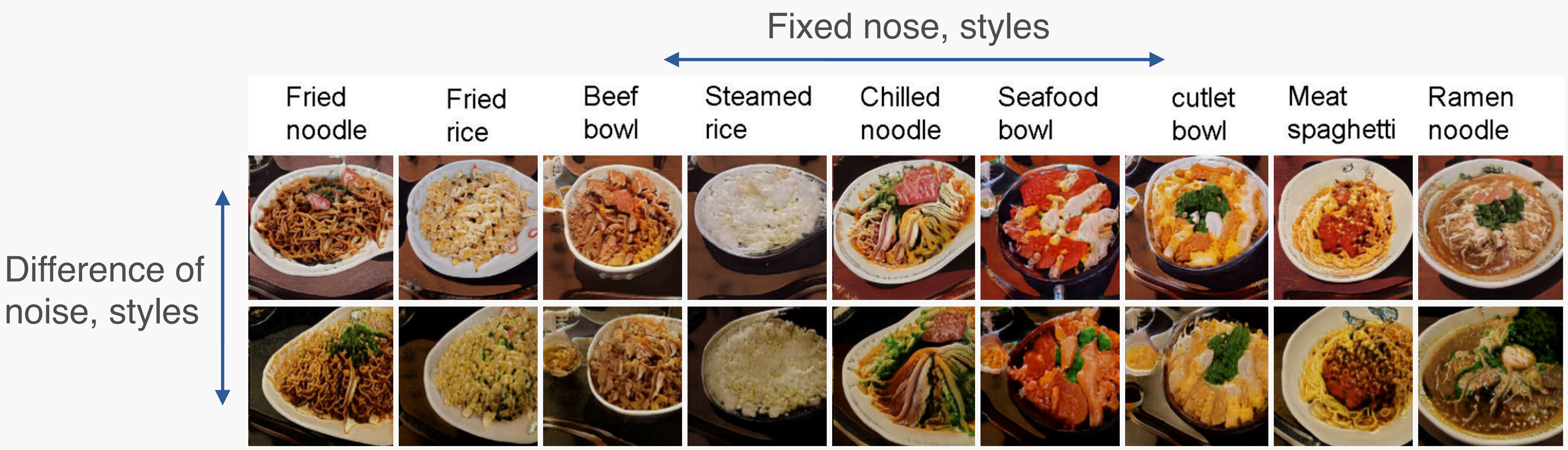
Unsuccessfully generated samples



Successfully generated samples

3.2 Experiments - Manipulation of latent space

- Thanks to the same input noises and styles, our model outputs **a plate with the same shape**.



Successfully generated samples

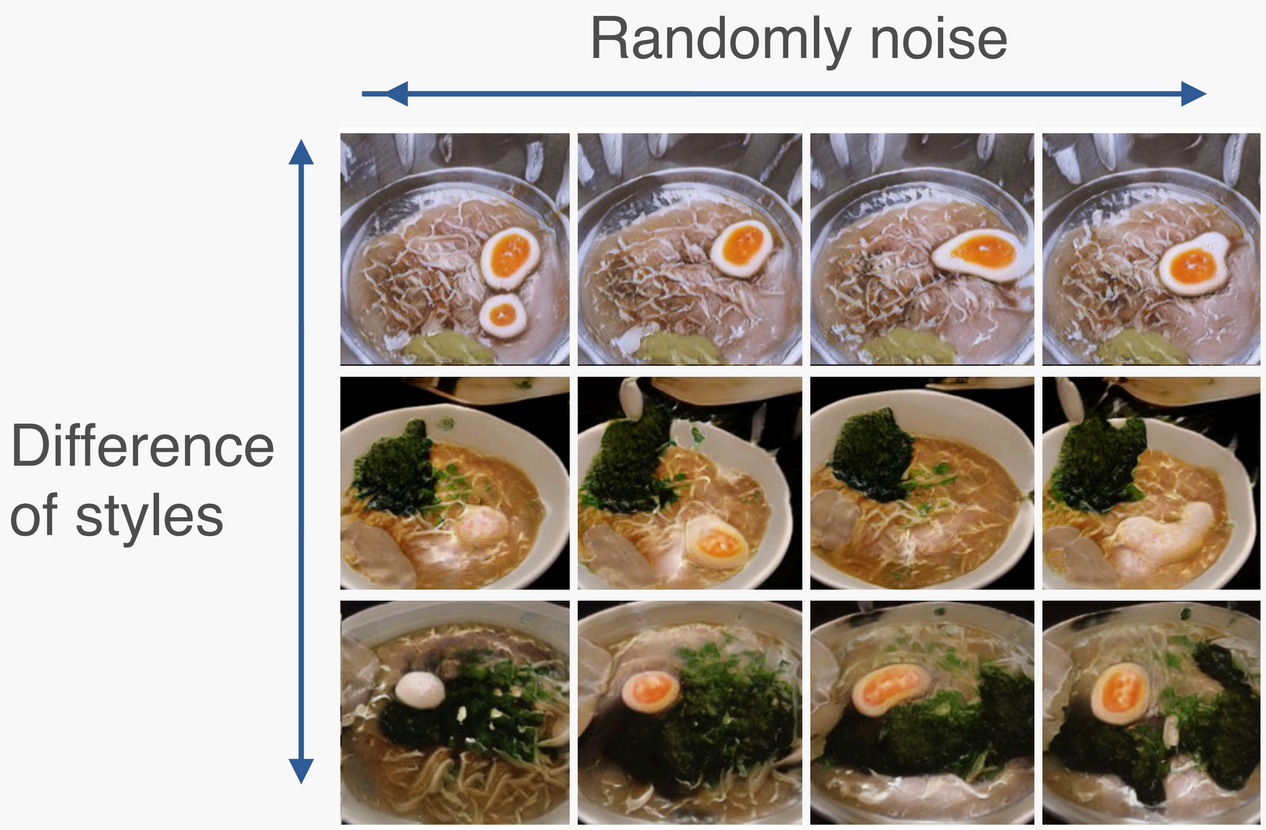
3.2 Experiments - Manipulation of latent space

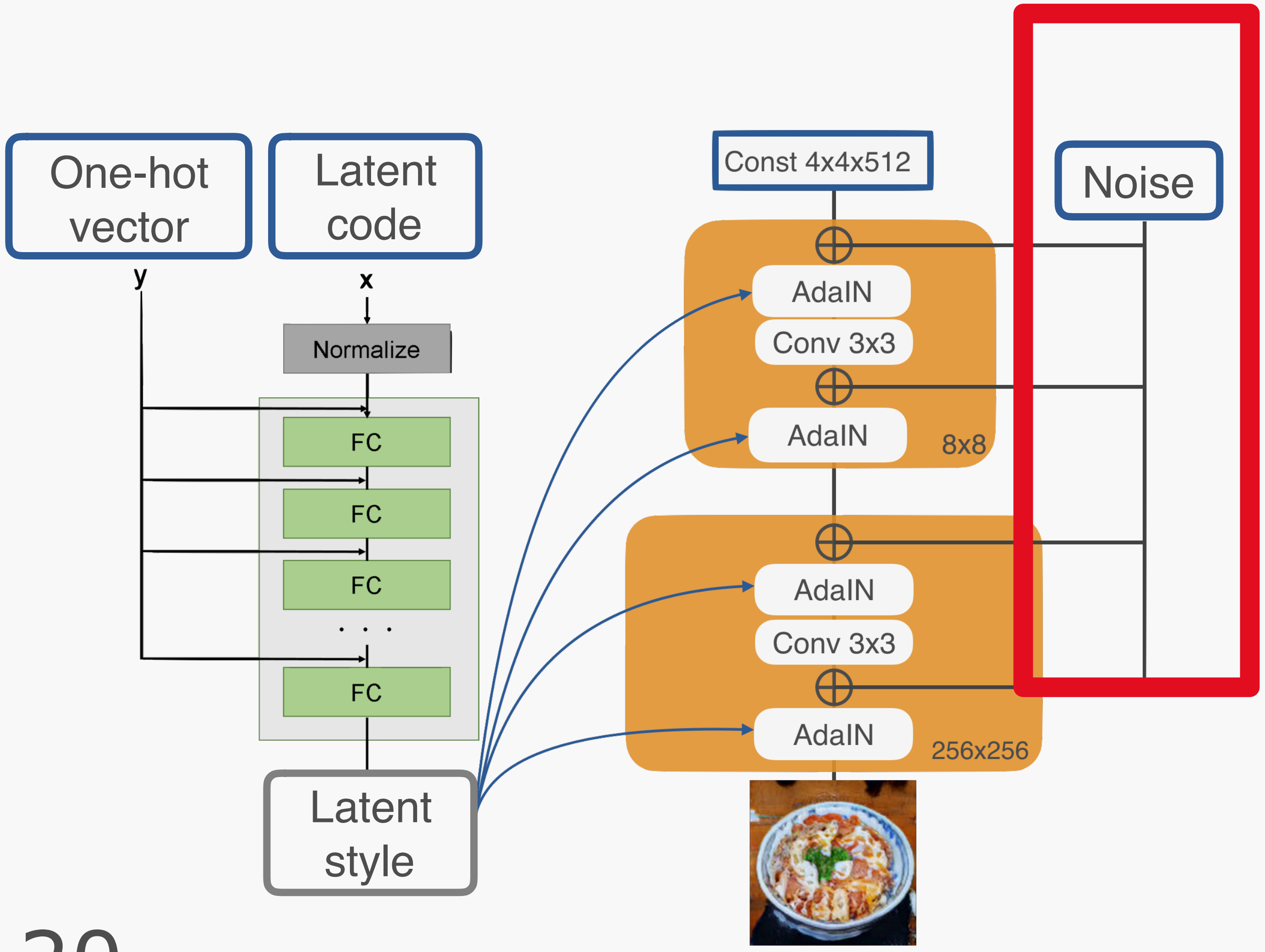
- However, although the noises and styles are fixed, **some images have distorted plate because they have few round samples**.

Unsuccessfully generated samples

3.3 Experiments - Manipulation of random noise

- Ramen images generated from a fixed style, a condition vector, and **randomly sampled noise**.
- Random noise plays a role in expressing differences such as food topping.





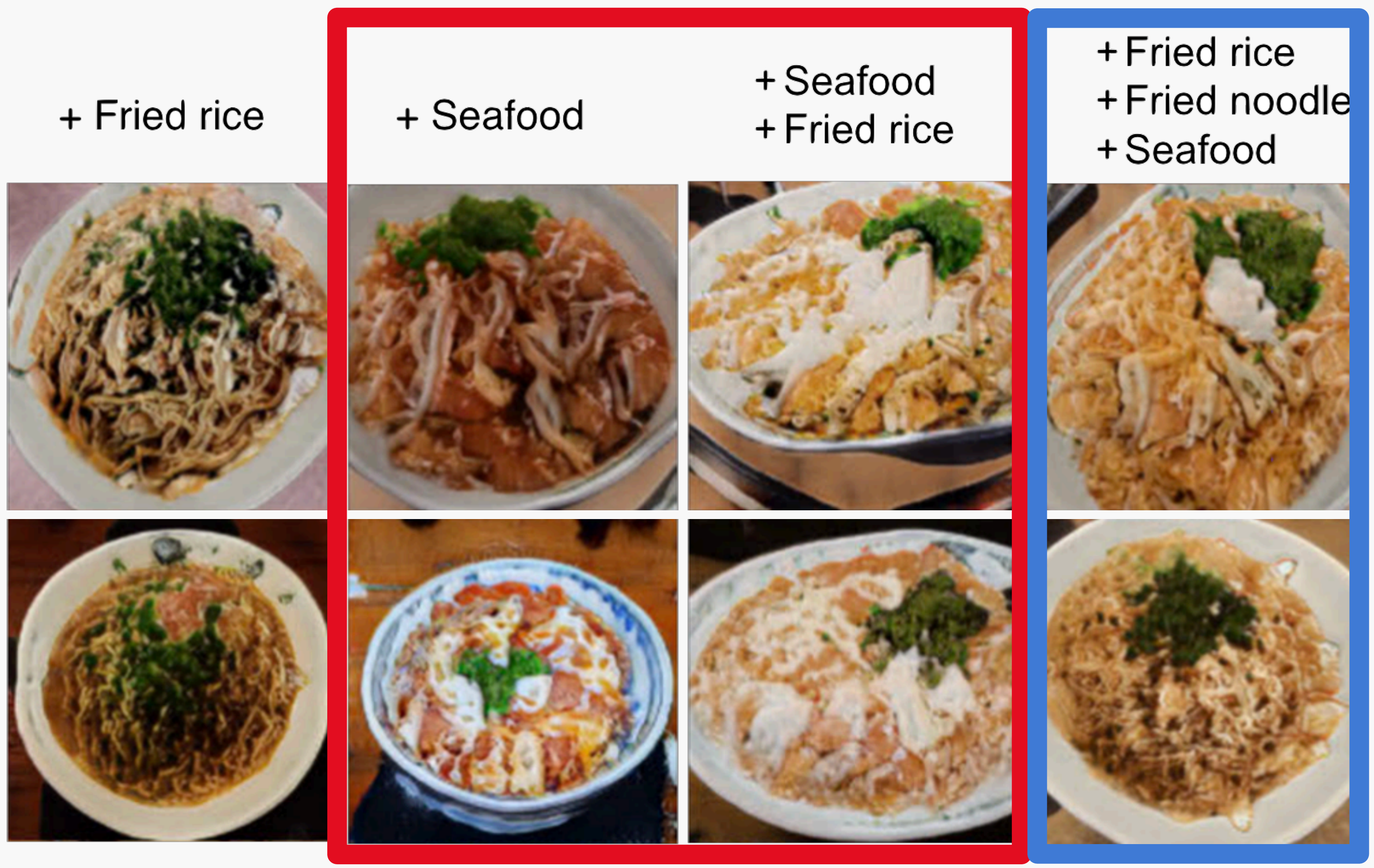
3.4 Experiments - Creation of Unseen Food Image

- Mixing ramen and other foods.

Ramen

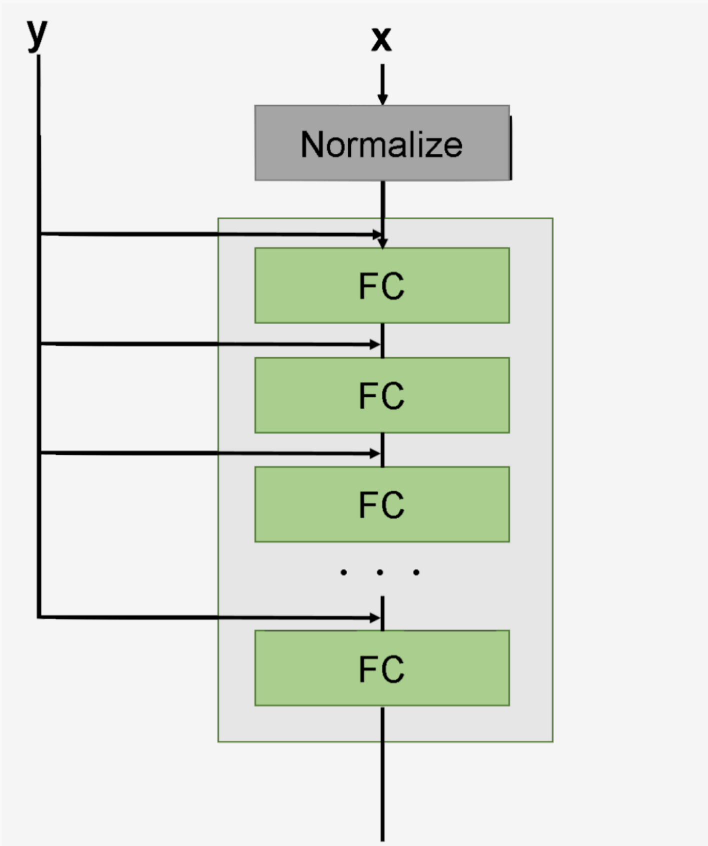
+ conditions

=

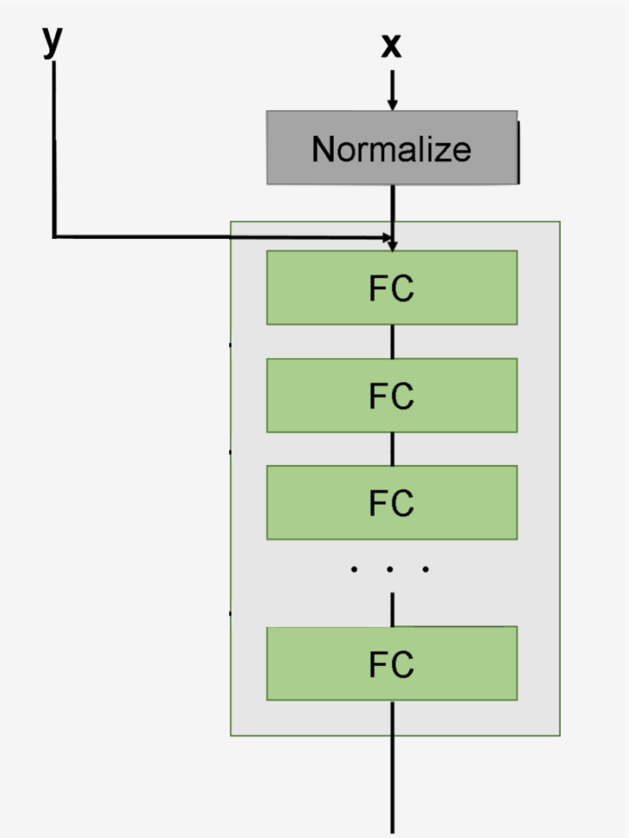


4. Ablation Study - Comparison of generator architecture

- We show the proposed structure(a) is more effective than baseline model(b).
- (b) takes the condition to **only the first layer.**



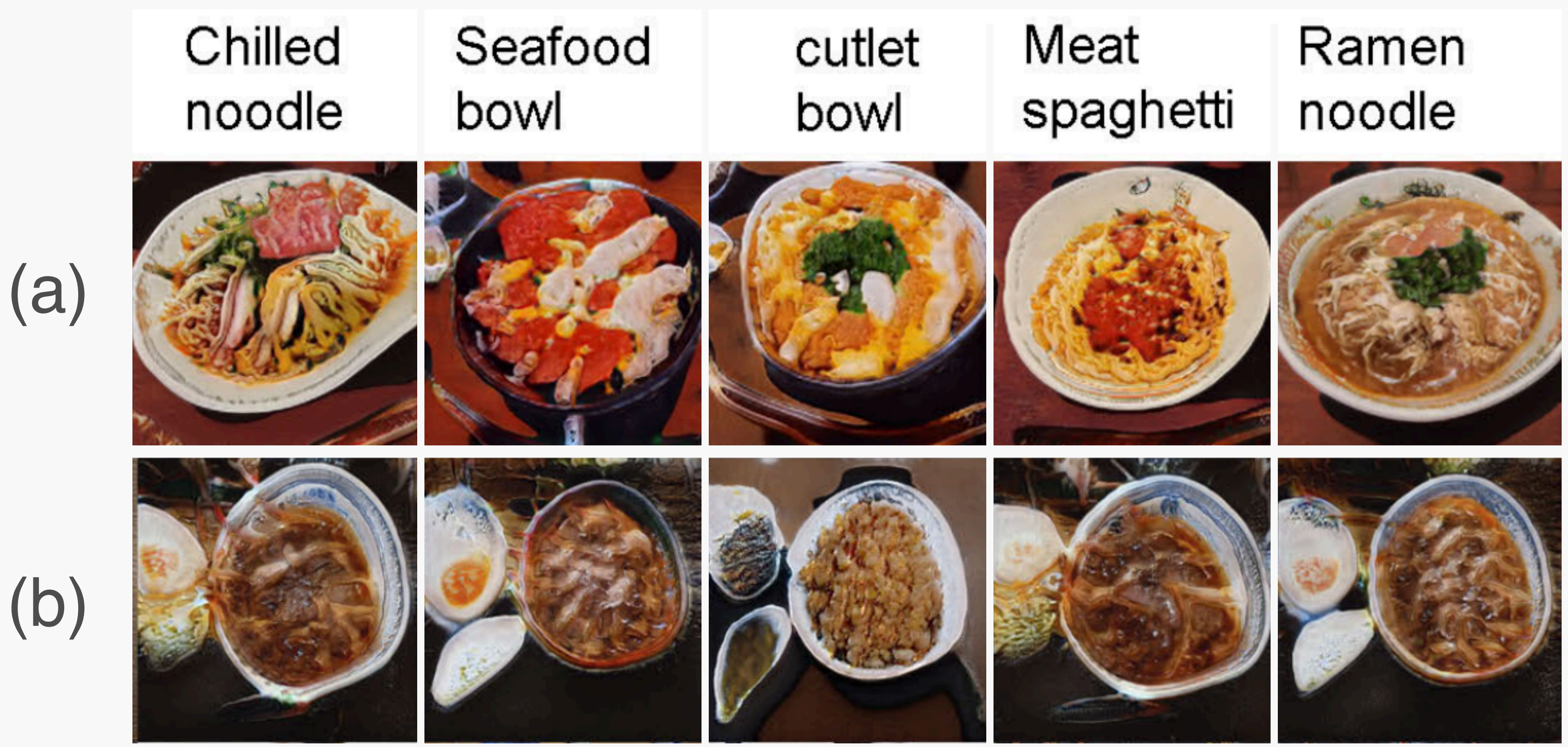
(a)

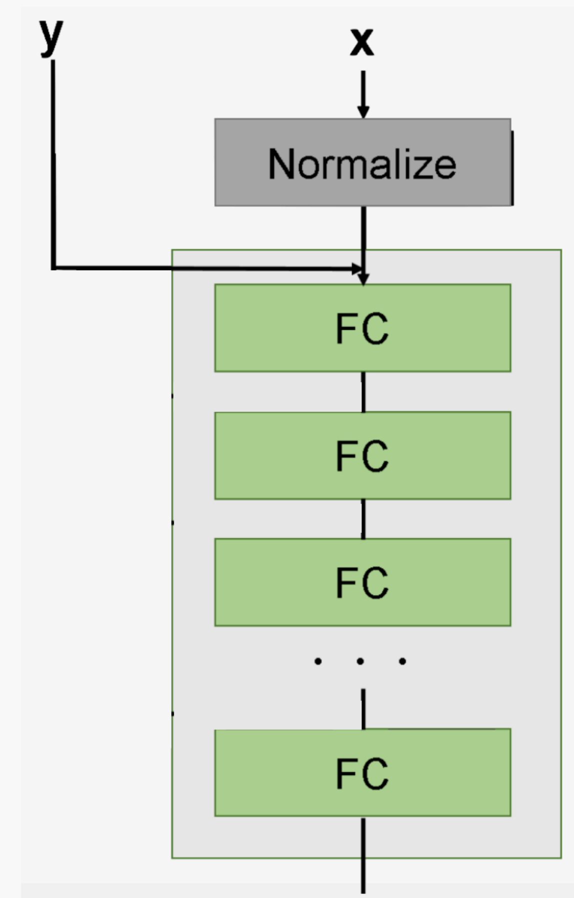


(b)

4. Ablation Study - Comparison of generator architecture

- **Mapping function (b) can not control the class without multiple condition vector inputs.**
- **In other words, (b) is affected only by noise.**

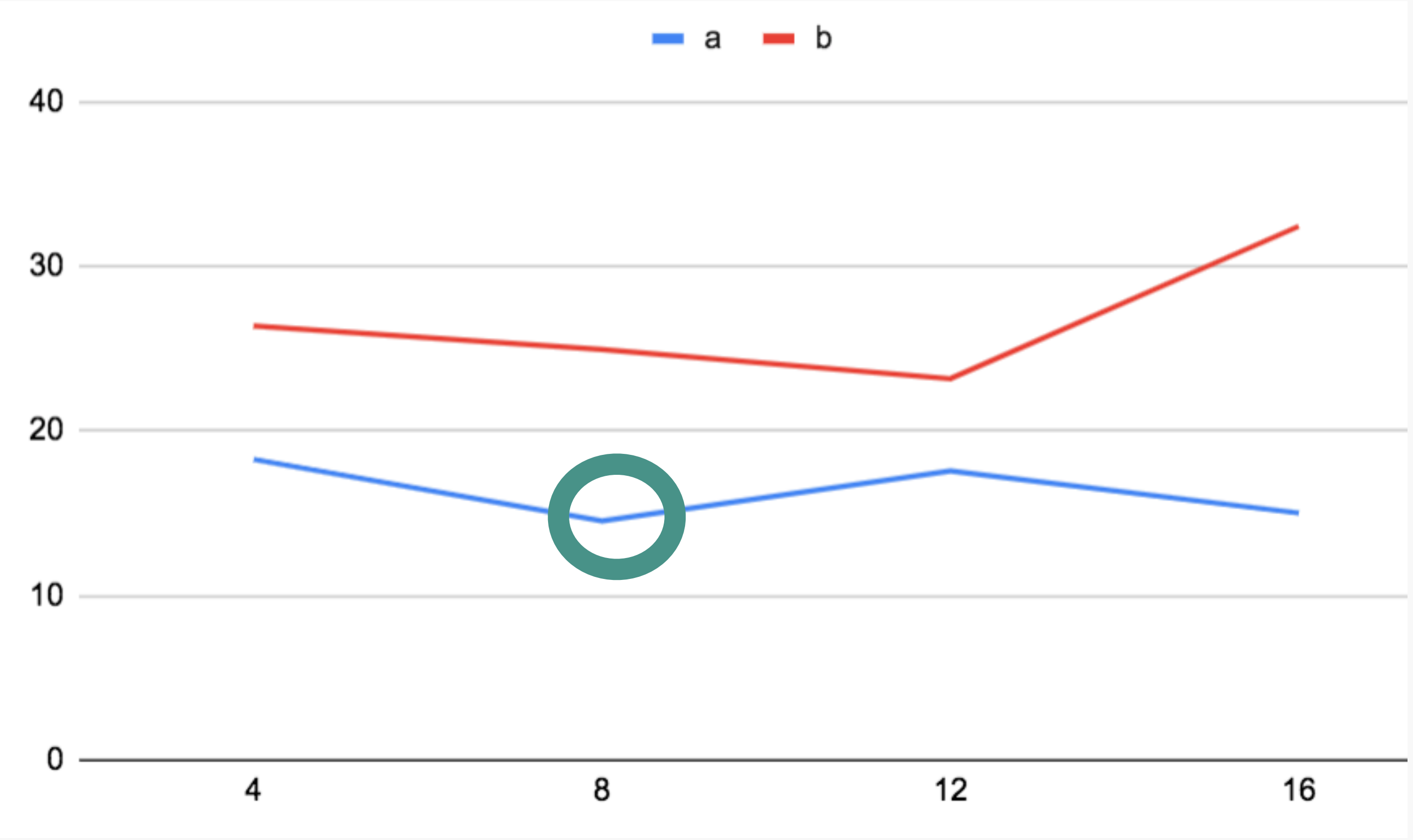




(b)

4. Ablation Study - Comparison of generator architecture

- FIDs is the metric of the distance between test set and generated samples.



- **<u>FIDs is the lowest</u>** when the proposed conditional mapping network consists of **eight FC layers!**

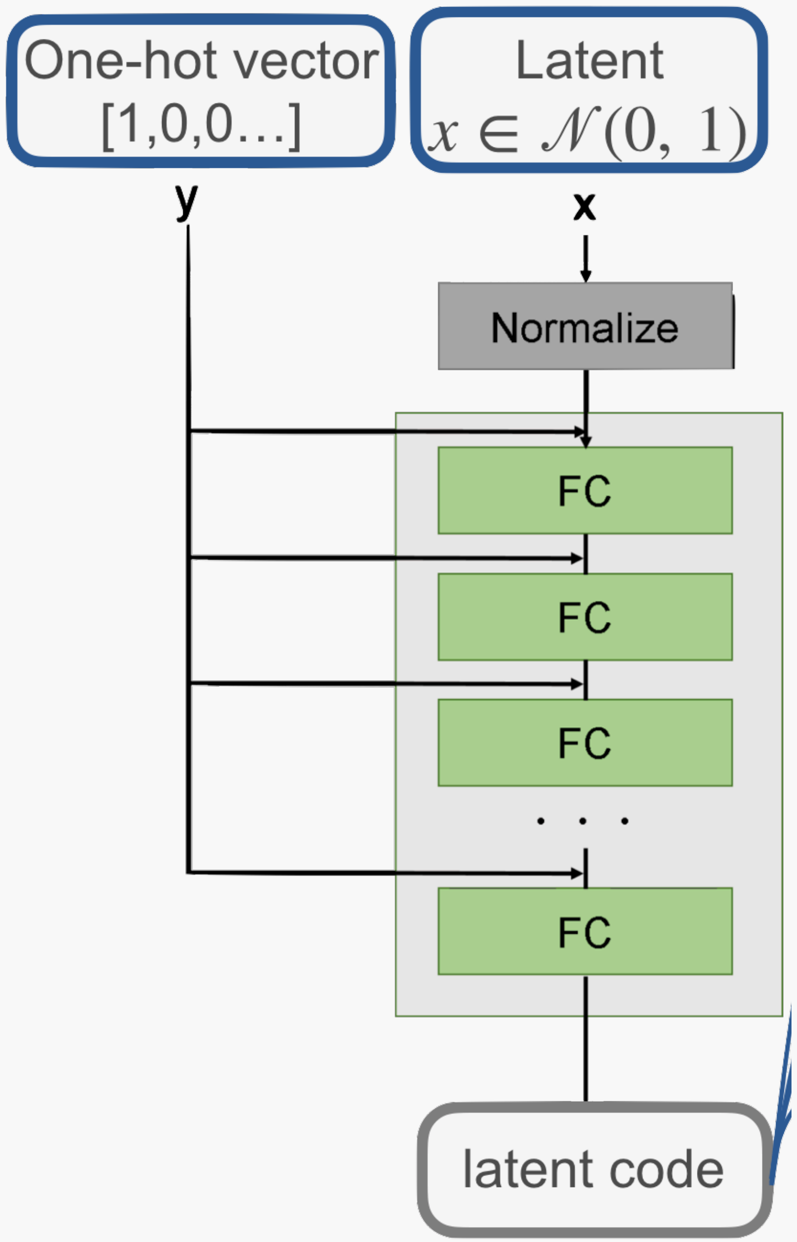
5. Conclusion

- We propose conditional version of StyleGAN to control latent space.
- Our model can create the unseen food images by mixing multiple conditions.
- In the future, we want to produce the healthy and unhealthy food image generation by using the conditions as calories and carbohydrates.

Thank you!!

Q&A

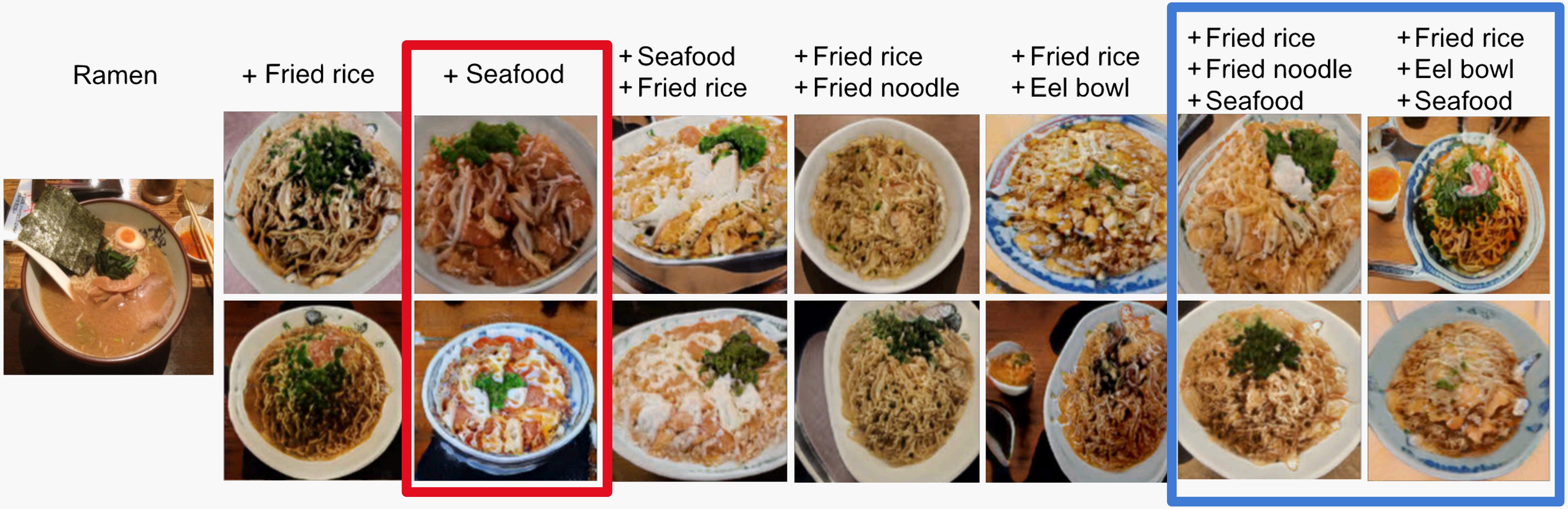
How to input the conditions



- Sample latent noise[N, 512], (N: Batch size).
- Add the one-hot condition to the noise and get the feature [N, 512+C], (C: Num of conditions).
- Input the features [N, 512+C] to FC layer and get the features [N, 512].
- Repeat.

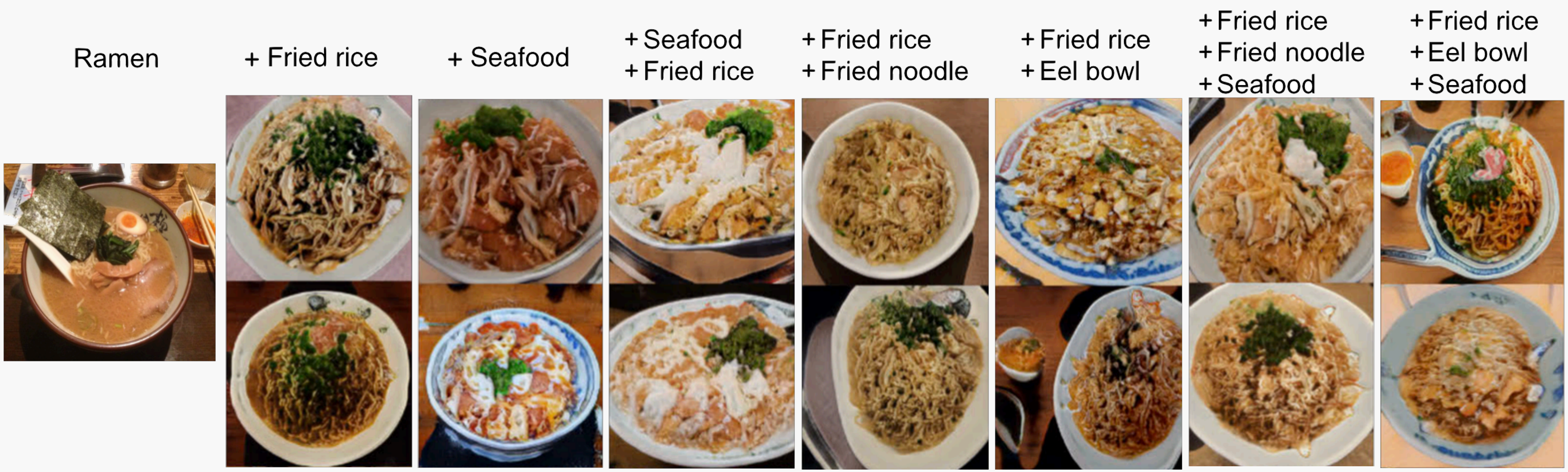
3.4 Experiments - Creation of Unseen Food Image

- There are many red samples of seafood bowl, so the ramen and seafood bowl are red.
- However, when many conditions are put, the effect faded.



1. Motivation

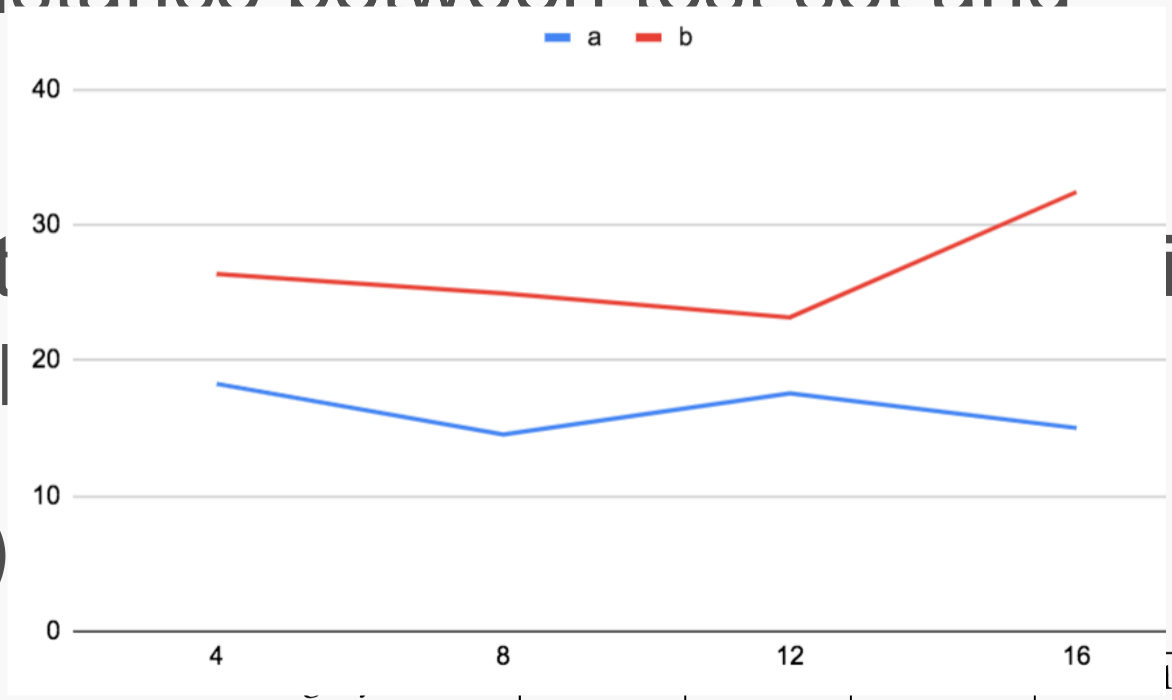
- **In addition,** thanks to introduce conditional vectors, our proposed model can create the unseen food images **by mixing multiple kinds of foods!**



Images generated by our proposed Conditional StyleGAN.

4. Ablation Study - Comparison of generator architecture

- FIDs is the metric of the distance between test set and generated samples.
- **FIDs is the lowest** when t ing network consists of **eight** l
- **FIDs (a) is lower than (b)**



Category	MLP 4	MLP 8	MLP 12	MLP 16
bibimbap	13.61	10.91	20.15	11.89
fried rice	16.30	8.50	14.15	8.65
beef bowl	22.41	11.50	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	40.33	32.53	31.82
fried noodle	10.56	8.51	13.58	16.93
pork cutlet bowl	13.47	10.22	15.91	11.93
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	14.20	20.15	12.32
soba noodle	21.15	13.73	14.29	18.0
average	18.28	14.55	17.58	15.02

FIDs of proposed model(a)

				16
bibimbap	23.11	19.34	21.63	26.13
fried rice	19.35	23.45	19.85	32.23
beef bowl	23.28	21.89	19.26	29.51
steamed rice	29.75	28.87	22.61	43.24
ramen noodle	34.18	30.32	27.15	45.09
eel bowl	51.23	47.13	43.34	45.57
fried noodle	18.41	19.65	20.27	27.16
pork cutlet bowl	22.46	20.77	20.56	22.47
chilled noodle	21.10	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	27.93	23.16	37.65
average	26.39	24.96	23.19	32.44

FIDs of baseline model (b)