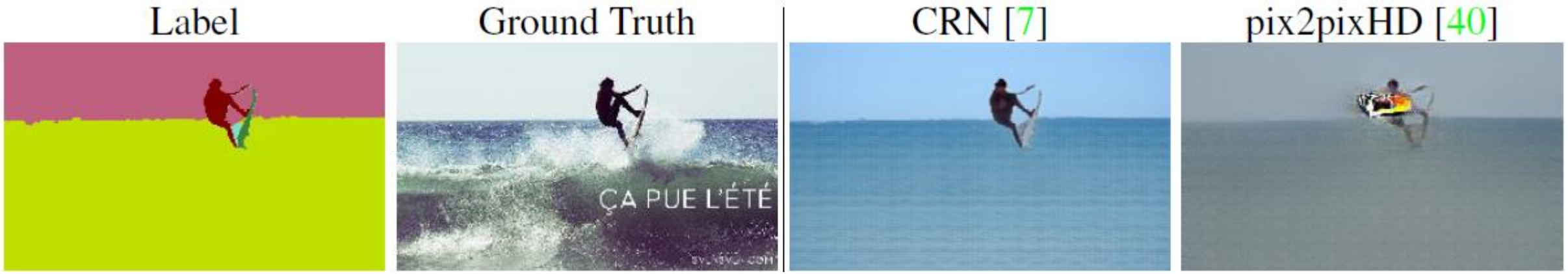
Paper review

SPADE : SPatially-Adaptive DEnormalization

석사과정 김 진용

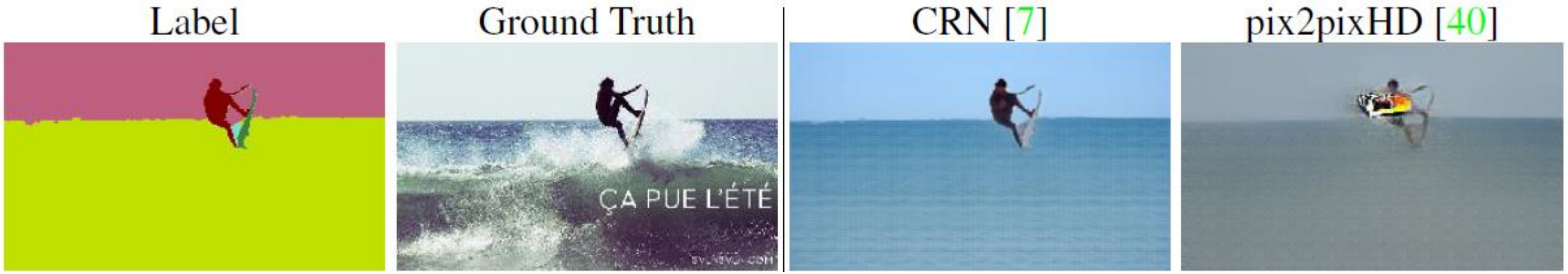
AI Lab. ARTIFICIAL INTELLIGENCE LAB.

본 논문은 Image synthesis시, Sementic segmentation으로 condition을 주는 상황에 대한 연구를 함.



노란 바탕(바다)에 사람과 서핑보드를 얹으니 주변과는 상관없이 정보를 Wash away
-> 부자연스러움

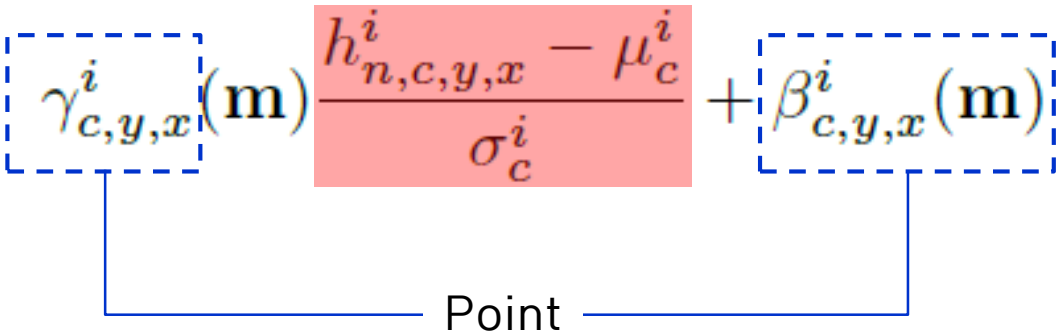
그러나, 기존방법에서는 Sementic segmentation의 정보를 "Wash away(씻어내림)"하는 경우가 많음



노란 바탕(바다)에 사람과 서핑보드를 얹으니 주변과는 상관없이 정보를 Wash away
-> 부자연스러움

따라서 , 본 논문에서는 공간적으로 적응되고(Spatially adaptive) Sementic layout의 informatio을 효과적으로 전달하는 spatially-adaptive normalization을 제안한다.

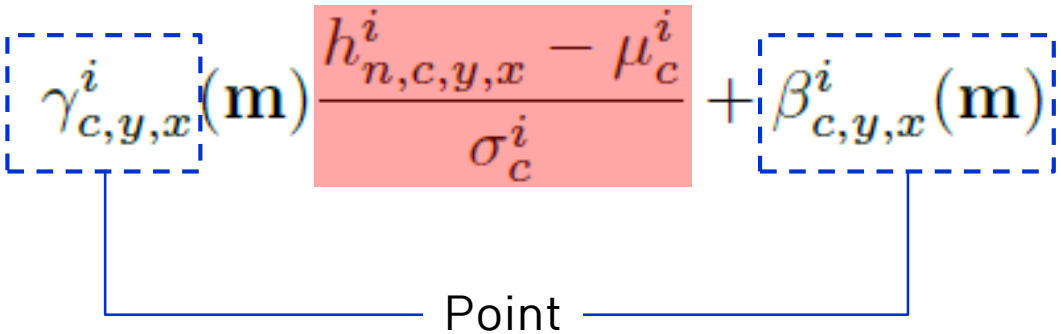
SPADE



Conv에 segmentation map을 통과시켜 만든 감마와 베타값을 기존 batch norm에 대체함

batch norm과는 다르게 Activation map의 크기와 같음

SPADE



Conv에 segmentation map을 통과시켜 만든 감마와 베타값을 기존 batch norm에 대체함

batch norm과는 다르게 Activation map의 크기와 같음

이렇게 될 경우, Segmentation mask를 하나의 값(scalar,vector)로 보지않고 전체적으로 분포된 map형태로 나타나기 때문에 입혀진 segmentation mask의 값이 다 제각기 달라진다.

```
self.mlp_shared = nn.Sequential(
    nn.Conv2d(label_nc, nhidden, kernel_size=ks, padding=pw),
    nn.ReLU()
)
self.mlp_gamma = nn.Conv2d(nhidden, norm_nc, kernel_size=ks, padding=pw)
self.mlp_beta = nn.Conv2d(nhidden, norm_nc, kernel_size=ks, padding=pw)
```

코드 단 몇줄로 극강의 가성비를 보여줌.

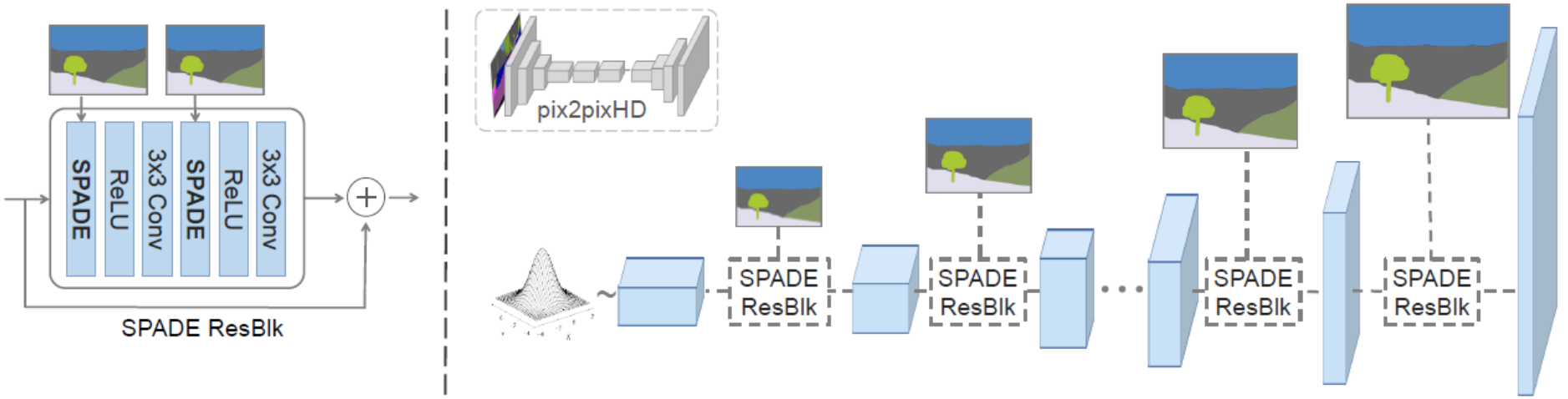


Figure 4: In the SPADE generator, each normalization layer uses the segmentation mask to modulate the layer activations. *(left)* Structure of one residual block with SPADE. *(right)* The generator contains a series of SPADE residual blocks with upsampling layers. Our architecture achieves better performance with a smaller number of parameters by removing the downsampling layers of leading image-to-image translation networks (pix2pixHD [40]).

SPADE를 사용하게되면 SPADE(normalization part)에 image를 주기 때문에
network에 input으로 image를 줄 필요가 없게된다.

-> 그 말은 즉, Random vector(noise)로 input이 들어가게 되는 것이고, **이는 저절로 Multimodal을 실현해줌**

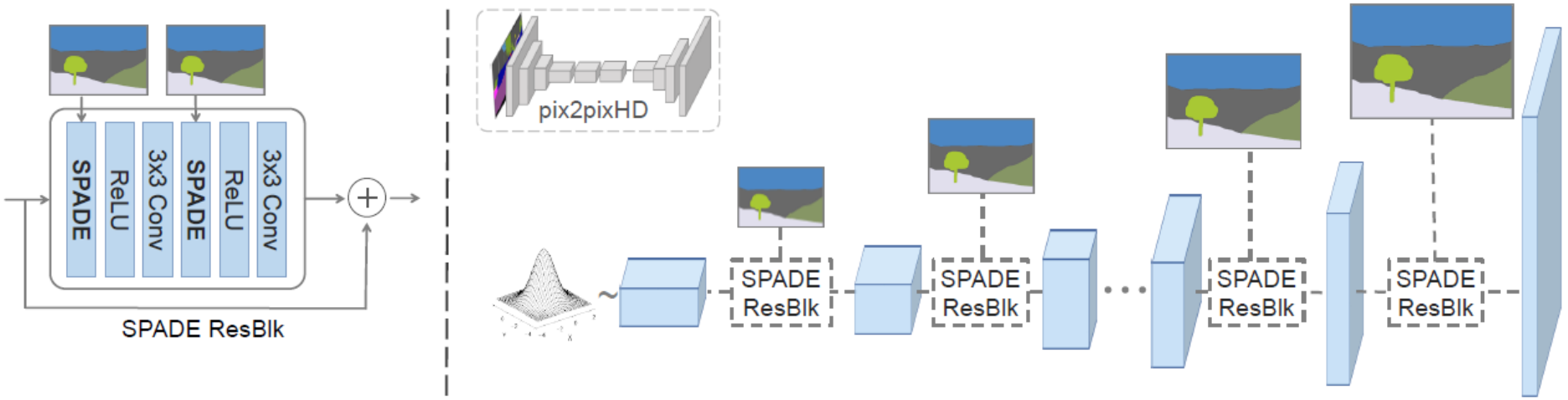


Figure 4: In the SPADE generator, each normalization layer uses the segmentation mask to modulate the layer activations. *(left)* Structure of one residual block with SPADE. *(right)* The generator contains a series of SPADE residual blocks with upsampling layers. Our architecture achieves better performance with a smaller number of parameters by removing the downsampling layers of leading image-to-image translation networks (pix2pixHD [40]).

기존 Pix2PixHD 모델(Local - Global 나눈 generator와 discriminator) 과 유사하며, SPADE block만 차용함.

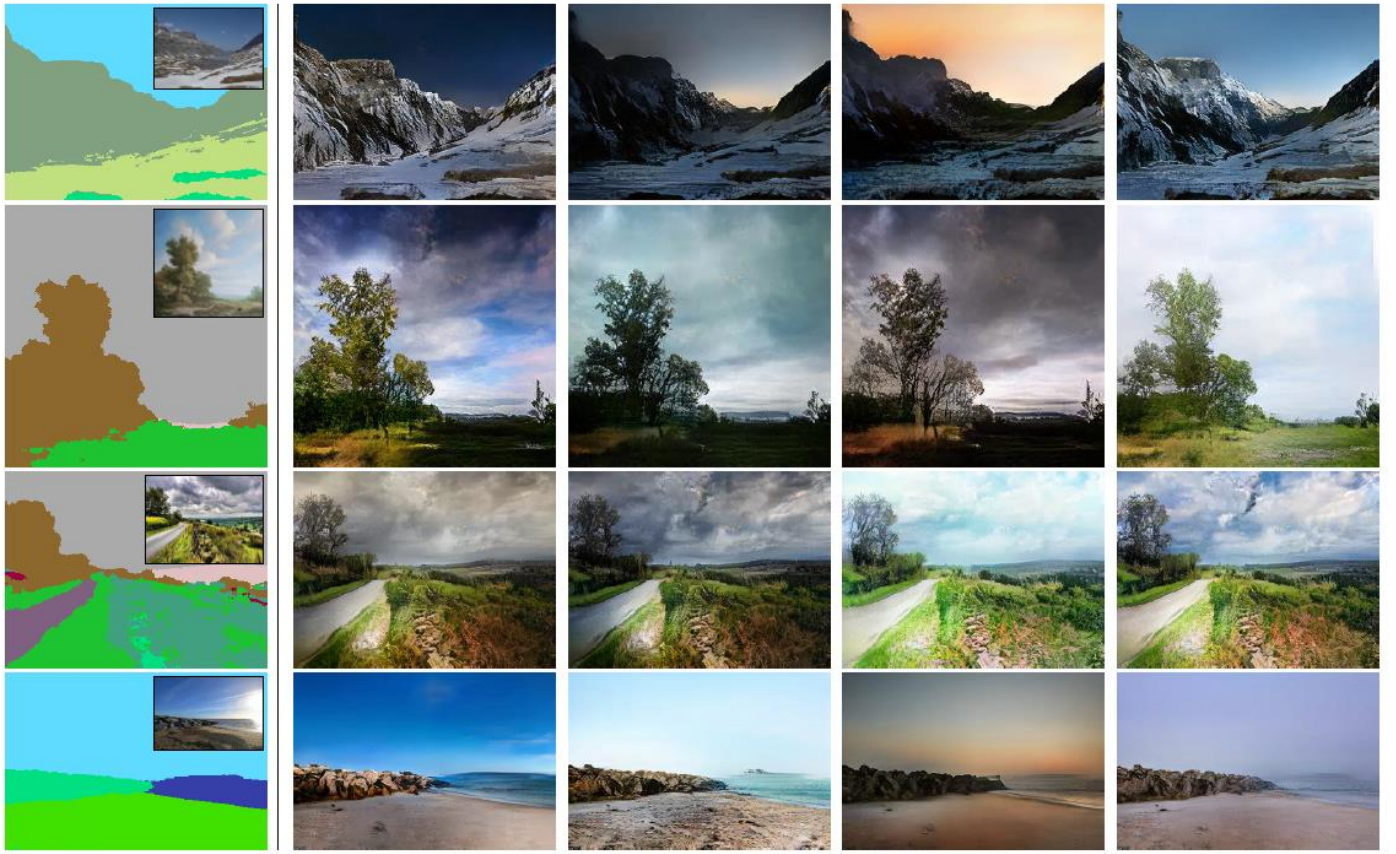


Figure 9: Our model attains multimodal synthesis capability when trained with the image encoder. During deployment, by using different random noise, our model synthesizes outputs with diverse appearances but all having the same semantic layouts depicted in the input mask. For reference, the ground truth image is shown inside the input segmentation mask.

왜 잘될까?

- 일반 nomalize layer보다 더 기존의 sementic 정보들을 보전을 잘하기 때문
- 기존에 SoTA로 쓰이는 InstanceNorm이 경우에는 flat하거나 uniform 한 마스크의 경우에는 정보를 wash away한다.
- 기존에 특정 sementation mask로 모든 픽셀을 도배하고 다른 균일 한 값을 가진 레이블로 올리면
  기존 정보가 wash away하기 때문에 정규화 된 활성화 값이 모두 0이된다. (보전 X)

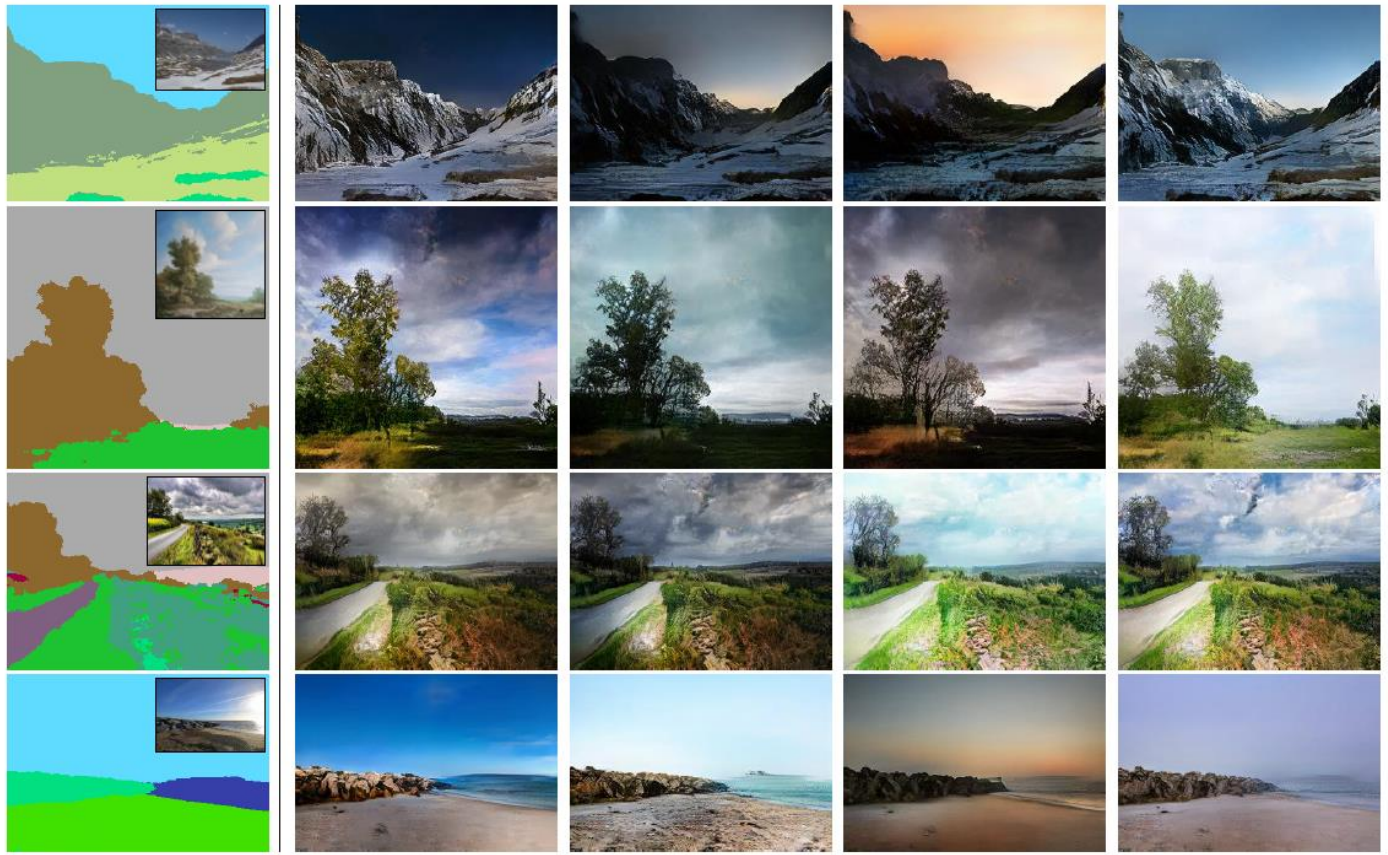


Figure 9: Our model attains multimodal synthesis capability when trained with the image encoder. During deployment, by using different random noise, our model synthesizes outputs with diverse appearances but all having the same semantic layouts depicted in the input mask. For reference, the ground truth image is shown inside the input segmentation mask.

왜 잘될까?

즉, 정리하자면 SPADE를 이용한 방법들은 기존 Normalization을 이용하는 이유인 “적응(Adaptive)”에 키워드를 맞추고
“부분적으로 적응시키며” -> “어색함을 야기하는 Wash Away를 보완”

AI Lab. ARTIFICIAL INTELLIGENCE LAB.

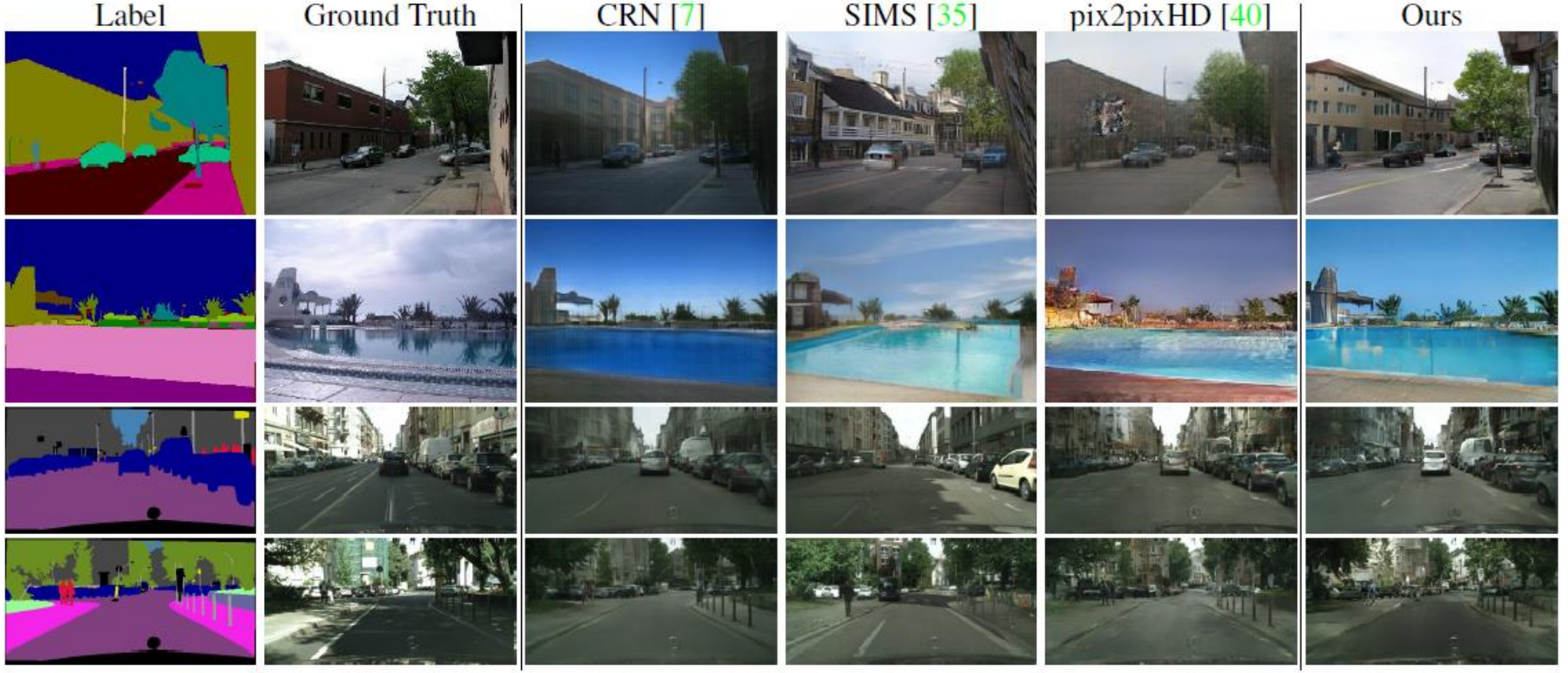


Figure 6: Visual comparison of semantic image synthesis results on the ADE20K outdoor and Cityscapes datasets. Our method produces realistic images while respecting the spatial semantic layout at the same time.

Method	COCO-Stuff			ADE20K			ADE20K-outdoor			Cityscapes		
	mIoU	accu	FID	mIoU	accu	FID	mIoU	accu	FID	mIoU	accu	FID
CRN [7]	23.7	40.4	70.4	22.4	68.8	73.3	16.5	68.6	99.0	52.4	77.1	104.7
SIMS [35]	N/A	N/A	N/A	N/A	N/A	N/A	13.1	74.7	67.7	47.2	75.5	**49.7**
pix2pixHD [40]	14.6	45.8	111.5	20.3	69.2	81.8	17.4	71.6	97.8	58.3	81.4	95.0
Ours	**37.4**	**67.9**	**22.6**	**38.5**	**79.9**	**33.9**	**30.8**	**82.9**	**63.3**	**62.3**	**81.9**	71.8

DEMO

http://nvidia-research-mingyuliu.com/gaugan