Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks

Alec Radford, Luke Metz, Soumith Chintala

2021.07.29

세종대학교 무인이동체공학전공

곽수지

Contribution

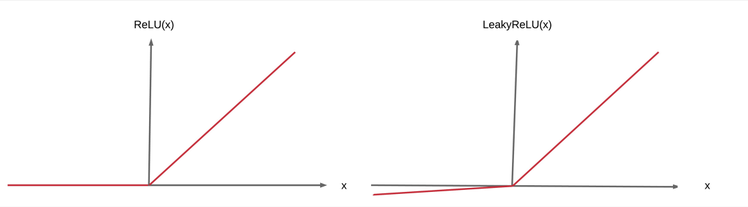
- 안정적 학습이 가능한 convolutional GAN 구조
- Discriminator 성능 평가: 학습된 Discriminator 사용하여 Image Classification 수행
  → 학습된 비지도 학습 기반 표현 모델이 지도 학습(분류)에 적용되었을 때 성능 확인
- Discriminator Filter Visualization (no longer black-box)
- Generator의 vector arithmetic 성질

안정된 DCGAN을 위한 구조

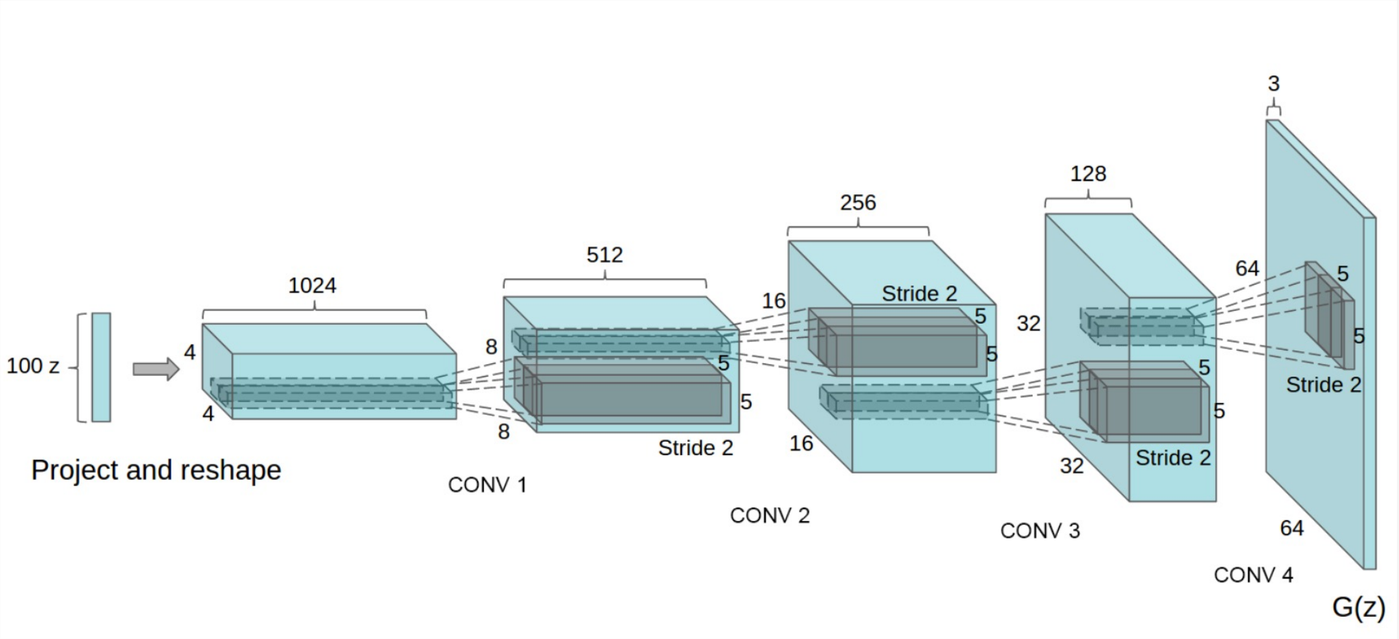
Architecture guidelines for stable Deep Convolutional GANs

- Replace any pooling layers with strided convolutions (discriminator) and fractional-strided convolutions (generator).
- Use batchnorm in both the generator and the discriminator.
- Remove fully connected hidden layers for deeper architectures.
- Use ReLU activation in generator for all layers except for the output, which uses Tanh.
- Use LeakyReLU activation in the discriminator for all layers.

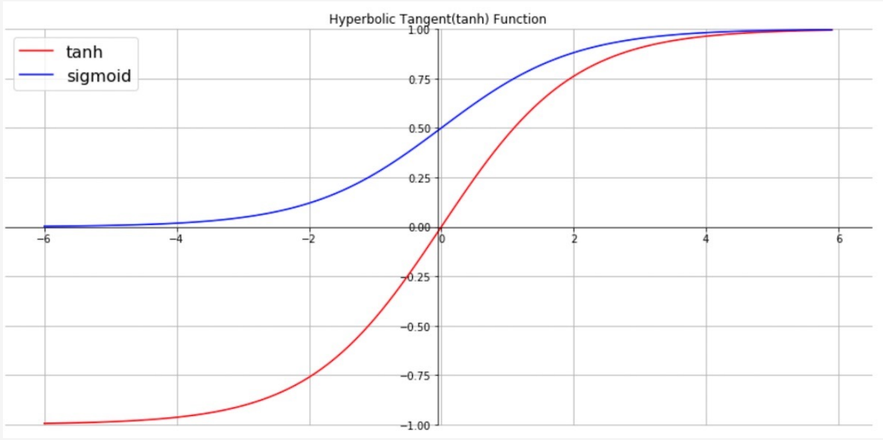
Activation function	Discriminator	Generator
Input layer	LeakyReLU	ReLU
else	LeakyReLU	ReLU
Output layer	LeakyReLU	Tanh



Generator Model

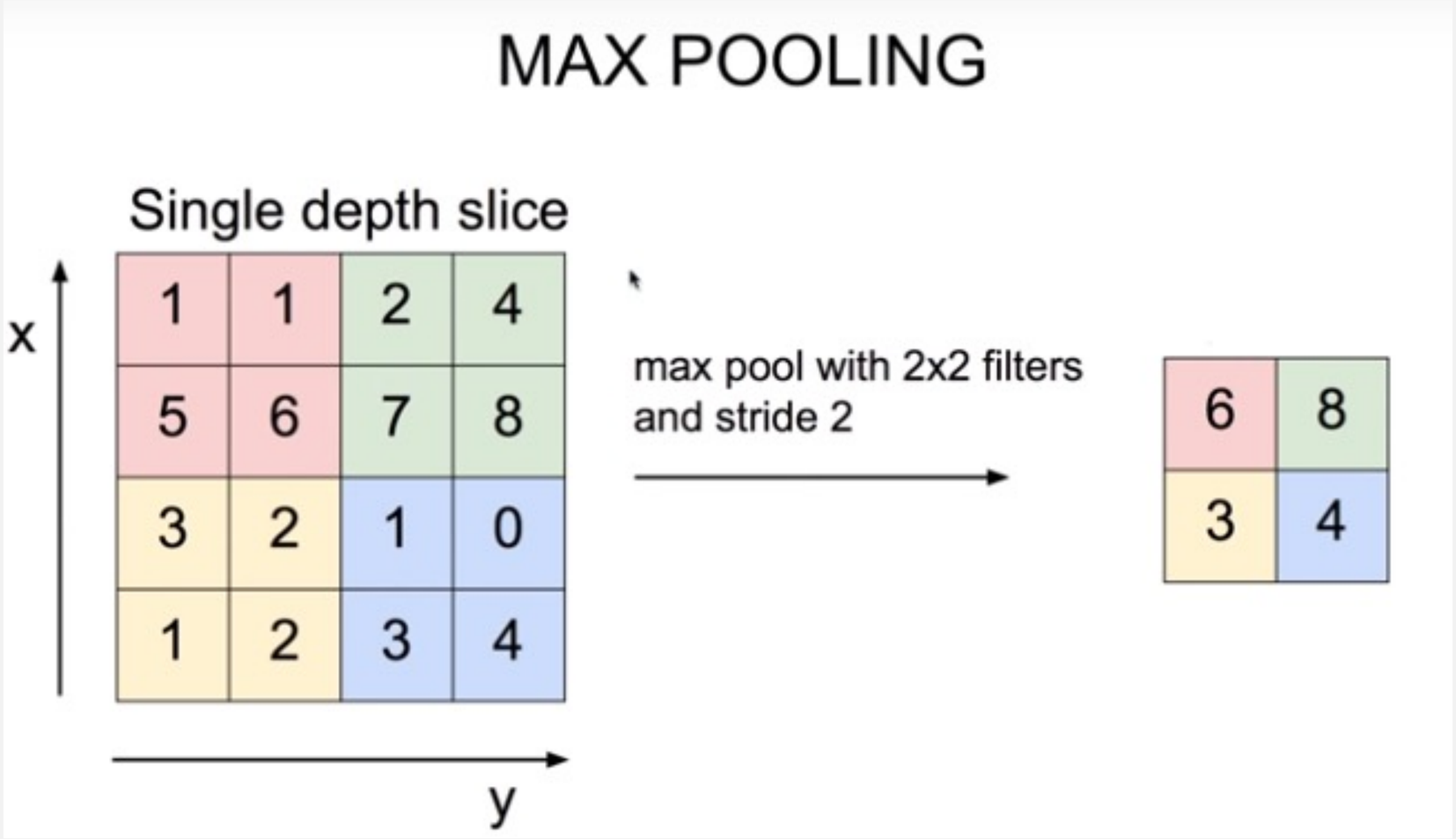


* Image Pixel 값 [-1, 1] scaling → (Output layer Activation Fuction: Tanh)
* Image Augmentation 미적용
* Optimizer: Adam Optimizer



Replace any pooling layers with strided convolutions

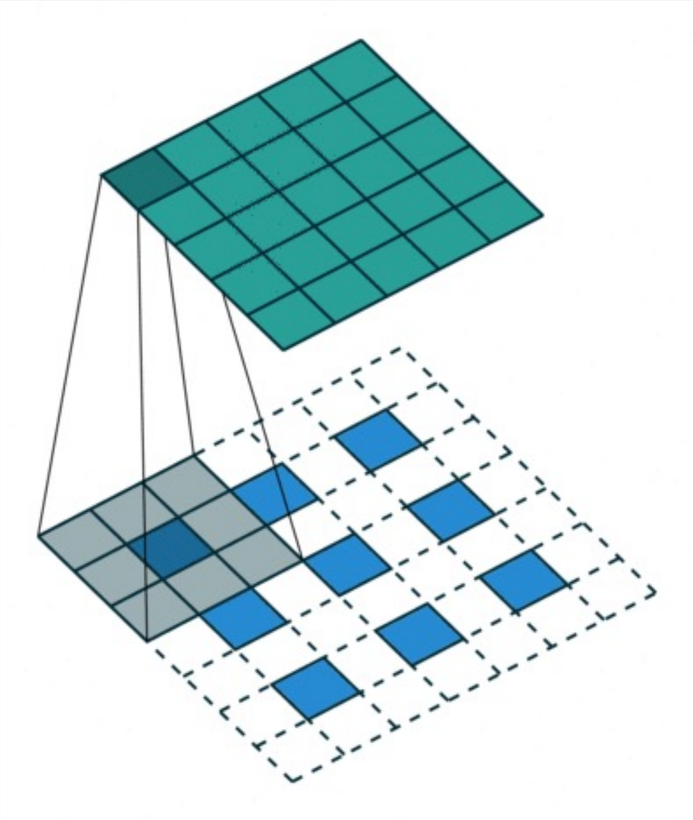
Pooling:



Replace any pooling layers with strided convolutions

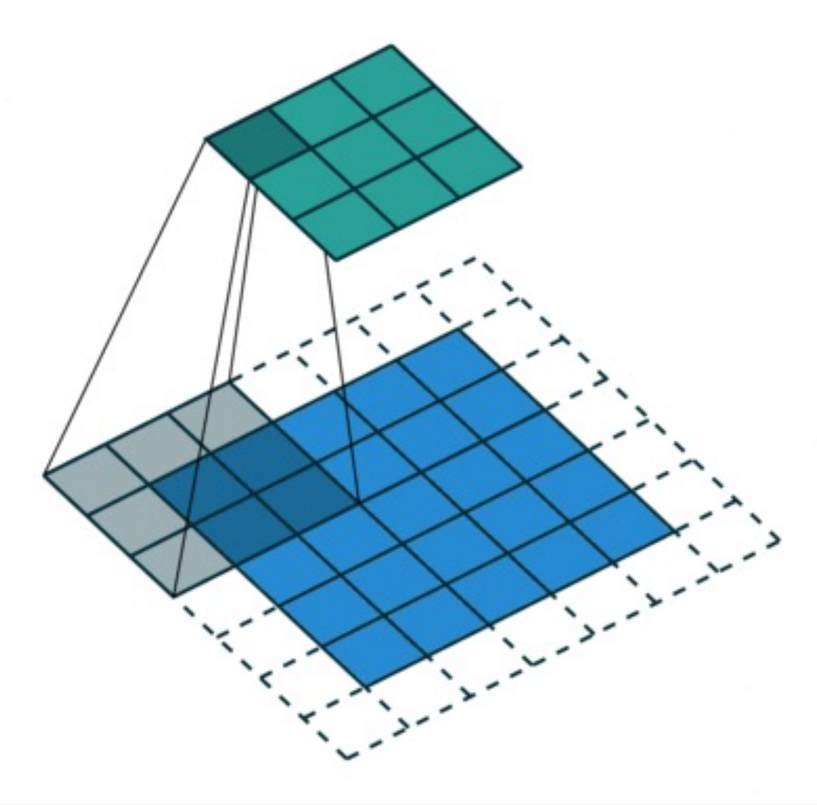
Stride:

Generator



Fractionally-Strided Convolution

Discriminator



Strided Convolution

Generated bedrooms (1 epoch)

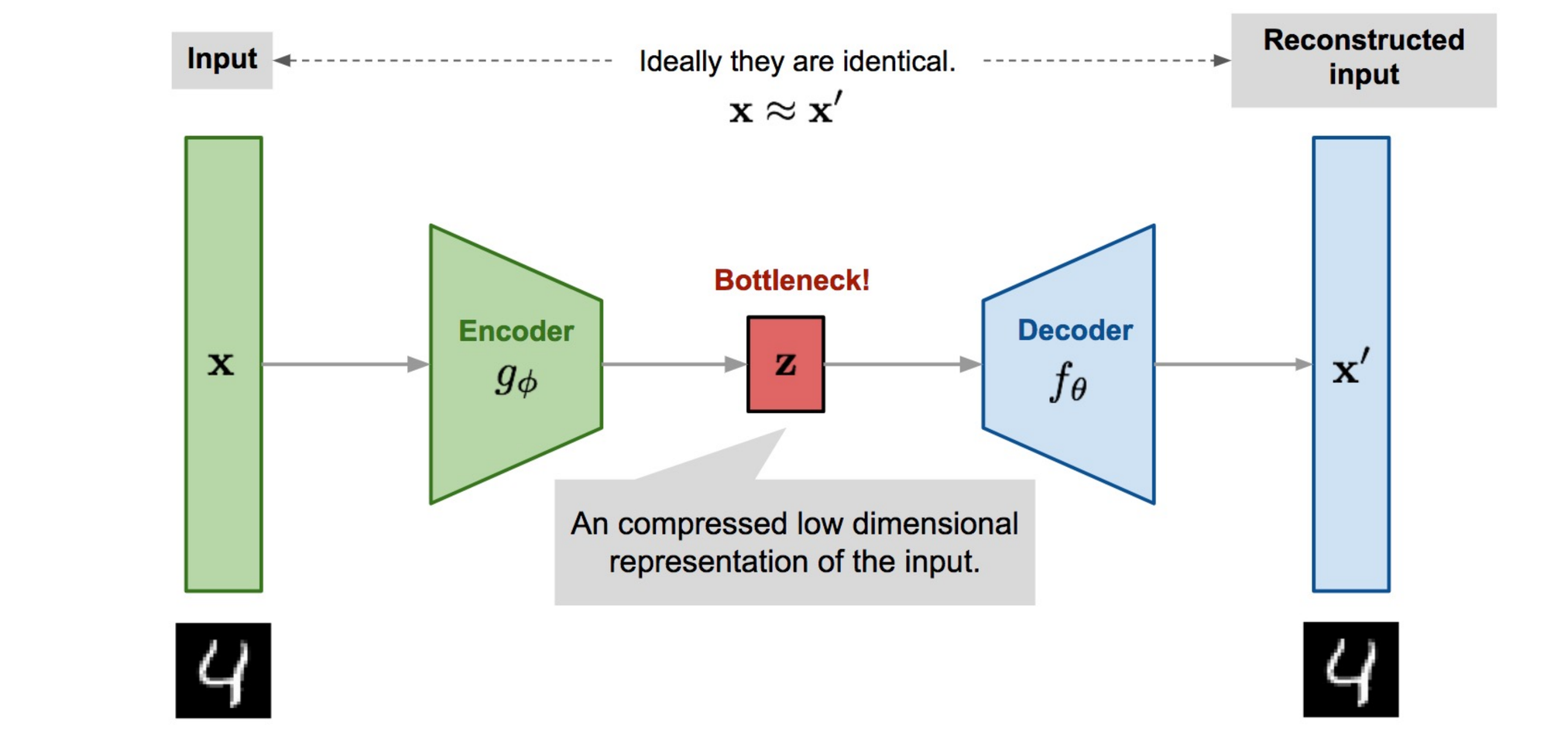
Figure 2: Generated bedrooms after one training pass through the dataset. Theoretically, the model could learn to memorize training examples, but this is experimentally unlikely as we train with a small learning rate and minibatch SGD. We are aware of no prior empirical evidence demonstrating memorization with SGD and a small learning rate.

Generated bedrooms (5 epochs)

Figure 3: Generated bedrooms after five epochs of training. There appears to be evidence of visual under-fitting via repeated noise textures across multiple samples such as the base boards of some of the beds.

유사 중복 이미지 제거: **de-noising autoencoder**

Generator의 학습 이미지 암기 가능성 낮추기 위해 도입
인코딩한 벡터에 임계 값을 도입하여 제거



비지도 표현 학습 알고리즘 성능 평가 **(Discriminator)**

Discriminator,
지도학습 분류 모델의 feature extractor로 사용 + 선형 모델(SVM) 도입하여 분류 모델 생성

Classification Results 1

ImageNet-1k dataset으로 Discriminator 학습 (not pre-trained on CIFAR-10)
→ CIFAR-10 분류 모델

Table 1: CIFAR-10 classification results using our pre-trained model. Our DCGAN is not pre-trained on CIFAR-10, but on Imagenet-1k, and the features are used to classify CIFAR-10 images.

Model	Accuracy	Accuracy (400 per class)	max # of features units
1 Layer K-means	80.6%	63.7% (±0.7%)	4800
3 Layer K-means Learned RF	82.0%	70.7% (±0.7%)	3200
View Invariant K-means	81.9%	72.6% (±0.7%)	6400
Exemplar CNN	84.3%	77.4% (±0.2%)	1024
DCGAN (ours) + L2-SVM	82.8%	73.8% (±0.4%)	512

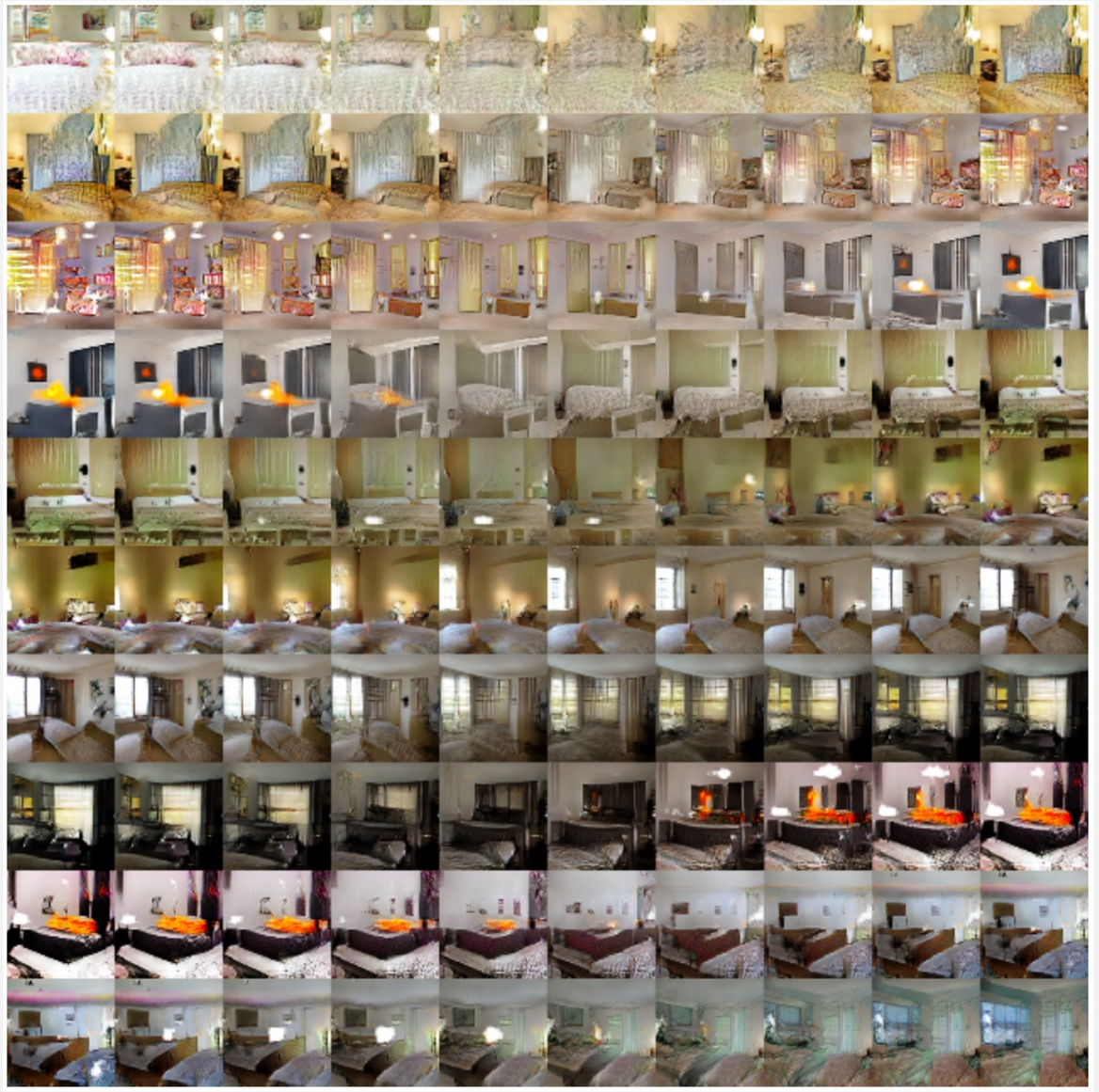
Classification Results 2

DCGAN(Unsupervised), Supervised CNN 비교

Table 2: SVHN classification with 1000 labels

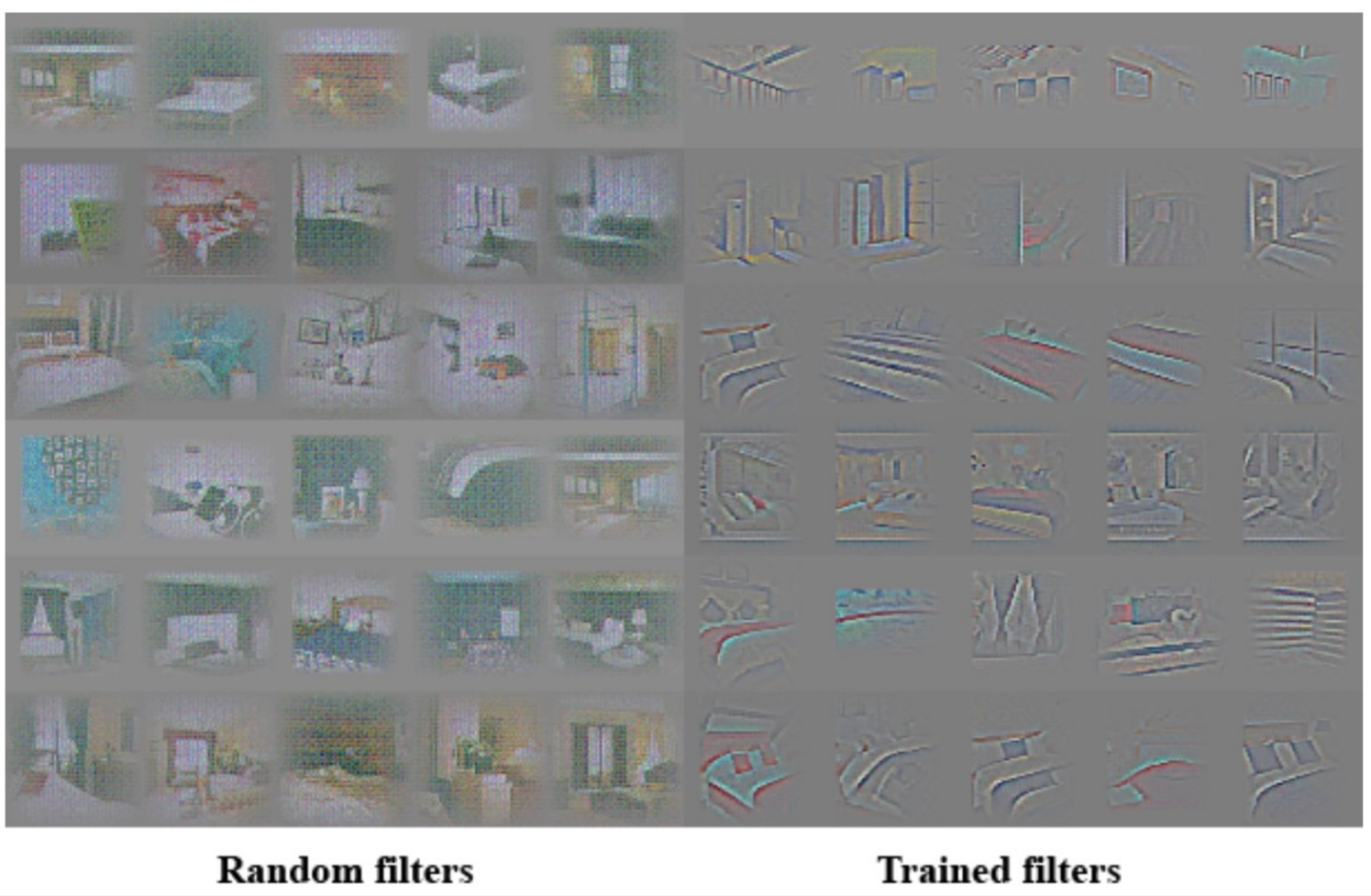
Model	error rate
KNN	77.93%
TSVM	66.55%
M1+KNN	65.63%
M1+TSVM	54.33%
M1+M2	36.02%
SWWAE without dropout	27.83%
SWWAE with dropout	23.56%
DCGAN (ours) + L2-SVM	22.48%
Supervised CNN with the same architecture	28.87% (validation)

Walking in the latent space



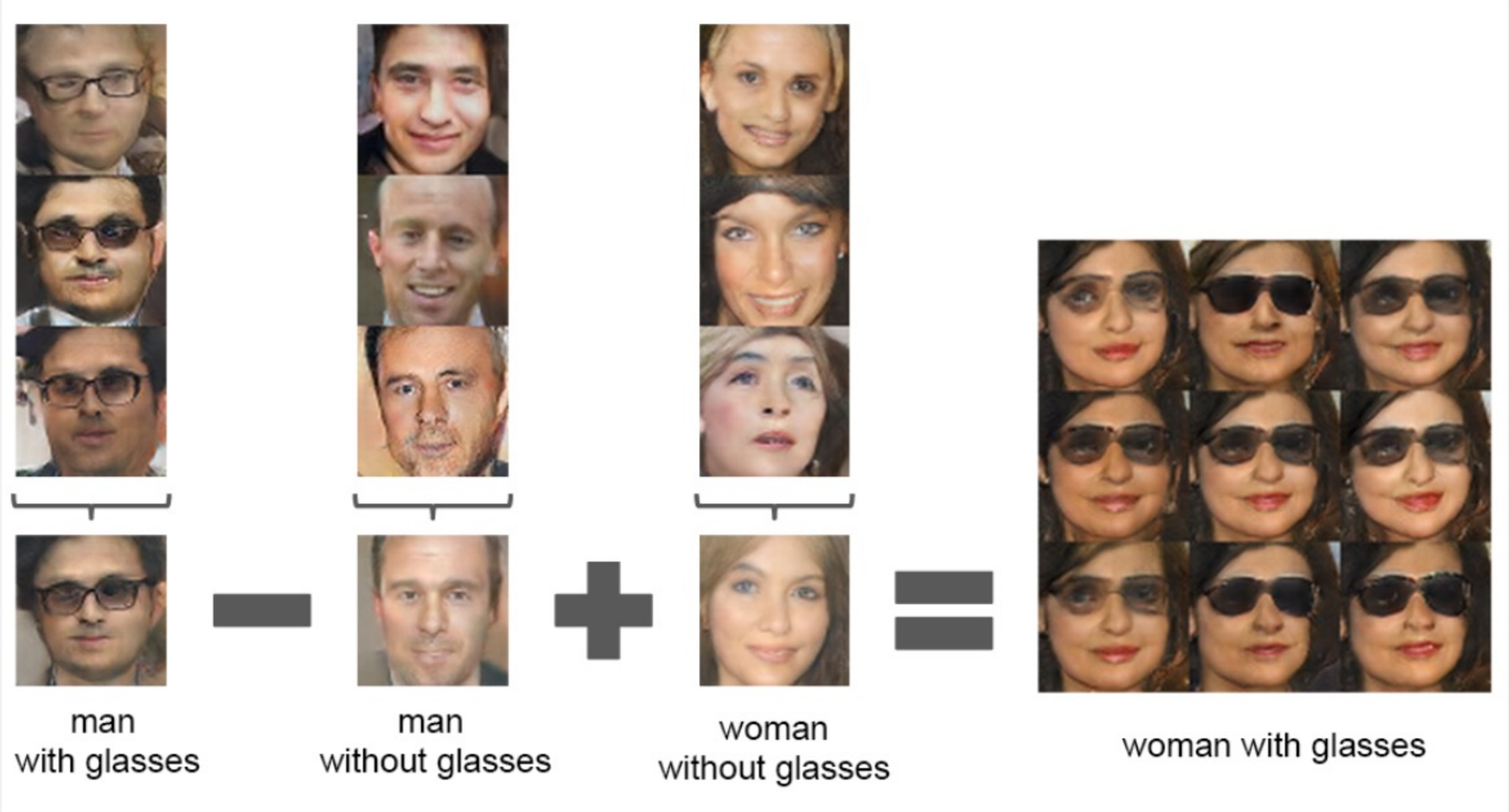
Visualize Discriminator filters (no longer black-box)

Guided Backpropagation 이용



Applying arithmetic in the input space

King – Man + Woman = Queen



Conclusion and Future Work

Conclusion:

안정적인 GAN 학습을 위한 구조 제시
GAN, 생성모델링과 지도 학습을 위한 표현 학습한다는 것 입증

Future Work:

모델의 불안정성 문제 (mode collapse)
Video, Audio 등 다른 도메인 확장 가능성 제시
학습된 latent space의 성질 조사