GANs training stabilization techniques applied to image translation

Matthieu Toulemont, Jean-Baptiste Sevestre

Ecole des Ponts, ENS Paris-Saclay

matthieu.toulemont@eleves.enpc.fr, jean-baptiste.sevestre@ens-paris-saclay.fr

January 15, 2019

Overview

Cycle-GAN

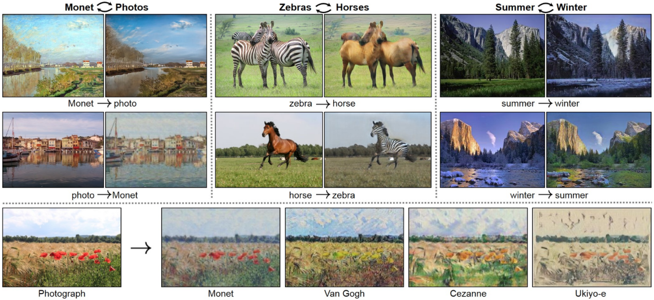


Figure: Picture Taken from the authors' github

W-GAN GP

$$L = \underbrace{\underset{\tilde{\boldsymbol{x}} \sim \mathbb{P}_g}{\mathbb{E}} \left[D(\tilde{\boldsymbol{x}}) \right] - \underset{\boldsymbol{x} \sim \mathbb{P}_r}{\mathbb{E}} \left[D(\boldsymbol{x}) \right]}_{\text{Original critic loss}} + \underbrace{\lambda \underset{\hat{\boldsymbol{x}} \sim \mathbb{P}_{\hat{\boldsymbol{x}}}}{\mathbb{E}} \left[(\|\nabla_{\hat{\boldsymbol{x}}} D(\hat{\boldsymbol{x}})\|_2 - 1)^2 \right]}_{\text{Our gradient penalty}}.$$

Figure: Gradient Penalty in WGAN Loss

In WGAN, critic weight clipping can lead to undesired behavior. Guljurani and al. (2017) propose gradient penalty (WGAN-GP), which does not suffer from the same problems.

SN-GAN

This is a great paper!

Ian Goodfellow

20 Nov 2017 ICLR 2018 Conference Paper45 Public Comment Readers: Everyone

Comment: This is a great paper! I don't think this paper explains the importance of its results nearly enough and I'm concerned that it may not be obvious what a breakthrough it is just from skimming the abstract.

"We tested the efficacy of spectral normalization on CIFAR10, STL-10, and ILSVRC2012 dataset, and we experimentally confirmed that spectrally normalized GANs (SN-GANs) is capable of generating images of better or equal quality relative to the previous training stabilization techniques" is a major understatement. This paper represents an extraordinary advance on the ILSVRC2012 dataset.

Before this paper, there was only one GAN that worked very well at all on ILSVRC2012: AC-GAN. AC-GAN was sort of cheating because it divided ImageNet into 100 smaller datasets that each contained only 10 classes. The new SN-GAN is the first GAN to ever fit all 1000 ImageNet classes in one GAN.

Scaling GANs to a high amount of classes has been a major open challenge and this paper has achieved an amazing 10X leap forward.

Figure: Comment of Ian Goodfellow about SN-GAN in OpenReview

In the Spectral Normalization GAN, the normalize the weights of the discriminator with respect to their spectral norm. They also compared the traditional GAN loss with a hinge loss.

However in our experiments we have kept the losses used in the cycle-gan and used the spectral normalization on both discriminators.

Loss

CycleGAN and SN-CycleGAN (with LS-Loss)

$$
\begin{aligned}
minV(G) = &\ \mathbb{E}[(D_A(G_A(A)) - 1)^2] + \mathbb{E}[(D_B(G_B(B)) - 1)^2] + \\
&+ \lambda_A \mathbb{E}[||G_B(G_A(A)) - A||_1] + \lambda_B \mathbb{E}[||G_A(G_B(B)) - B||_1]
\end{aligned}
$$

$$minV(D_A) = \mathbb{E}[(D_A(B) - 1)^2] + \mathbb{E}[(D_A(G_A(A)))^2]$$

$$minV(D_B) = \mathbb{E}[(D_B(A) - 1)^2] + \mathbb{E}[(D_B(G_B(B)))^2]$$

CycleWGAN-GP

$$
\begin{aligned}
minV(G) = &\ -\mathbb{E}[D_A(G_A(A))] - \mathbb{E}[D_B(G_B(B))] + \\
&+ \lambda_A \mathbb{E}[||G_B(G_A(A)) - A||_1] + \lambda_B \mathbb{E}[||G_A(G_B(B)) - B||_1]
\end{aligned}
$$

$$minV(D_A) = -\mathbb{E}[D_A(B)] + \mathbb{E}[D_A(G_A(A))] + \lambda \mathbb{E}[(||\nabla_t D_A(t)||_2 - 1)^2]$$

$$minV(D_B) = -\mathbb{E}[D_B(A)] + \mathbb{E}[D_B(G_B(B))] + \lambda \mathbb{E}[(||\nabla_t D_B(t)||_2 - 1)^2]$$

Metrics

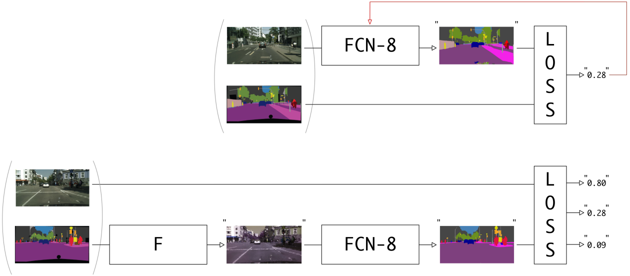


Figure: Cityscapes Labels→Photo ; how FCN score is calculated

FCN score introduced by Isola and al. is similar to the Inception Score. "If the generated images are realistic, classifiers trained on real images will be able to classify the synthesized image correctly as well."

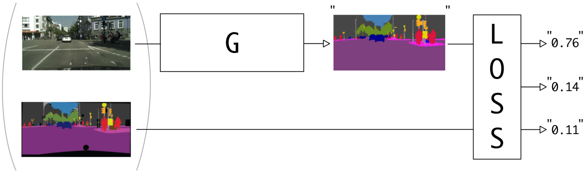


Figure: Cityscapes Photo→Labels : how Semantic Segmentation Metrics is calculated

The advantages of this metrics are that they are "raw" metrics and don't need any external Neural Networks to work, contrary to Inception-Score or FCN-Score.

Experiments

In order to evaluate the impact of the methods presented above, we have worked on three datasets :

- To measure the quality of the photo generated we have used the FCN score on the Cityscapes dataset.
- To measure the quality of the labels generated we have used standard classification metrics (Per Pixel Accuracy, Per Class Accuracy, Class IOU) on all datasets.
- We have used the same hyperparameters as in the original papers.

Cityscapes datasets, results

Model	Per-pixel acc.	Per-class acc.	Class IOU
CycleGAN	0.75	0.33	0.25
WCycleGAN-GP	0.85	0.38	0.27
SN-CycleGAN	0.81	0.34	0.27
Ground Truth	**0.87**	**0.45**	**0.36**
CycleGAN [2]	0.52	0.17	0.11
pix2pix [6]	**0.66**	**0.23**	**0.17**

Table: FCN-scores for different methods, evaluated on Cityscapes labels $\rightarrow$ photo.

The difference between the paper's performance and ours for the CycleGAN lies in the fact that we used a different architecture for the FCN (We used a pretrained VGG as encoder).

Photo Generation

Figure: Left to right : Ground Truth, CycleGAN, W-CycleGAN-GP, SN-CycleGAN

Cityscapes datasets, results

Model	Per-pixel acc.	Per-class acc.	Class IOU
CycleGAN	**0.54**	**0.19**	**0.13**
WCycleGAN-GP	0.49	0.13	0.09
SN-CycleGAN	0.53	0.18	0.12
CycleGAN [2]	0.58	0.22	0.16
pix2pix [6]	**0.83**	**0.36**	**0.29**

Table: Classification performance of photo $\rightarrow$ labels for different methods on Cityscapes dataset.

Label Generation

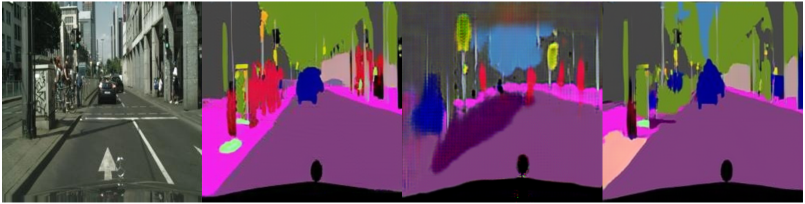


Figure: Left to right : Ground Truth, CycleGAN, W-CycleGAN-GP, SN-CycleGAN

Photo to Maps datasets, results

Model	Per-pixel acc.	Per-class acc.	Class IOU
CycleGAN	**0.66**	**0.37**	**0.27**
WCycleGAN-GP	0.48	0.20	0.12
SN-CycleGAN	0.64	0.34	0.24

Table: Classification performance of aerial photo → maps for different methods on data scraped from Google Maps.

Maps : Images

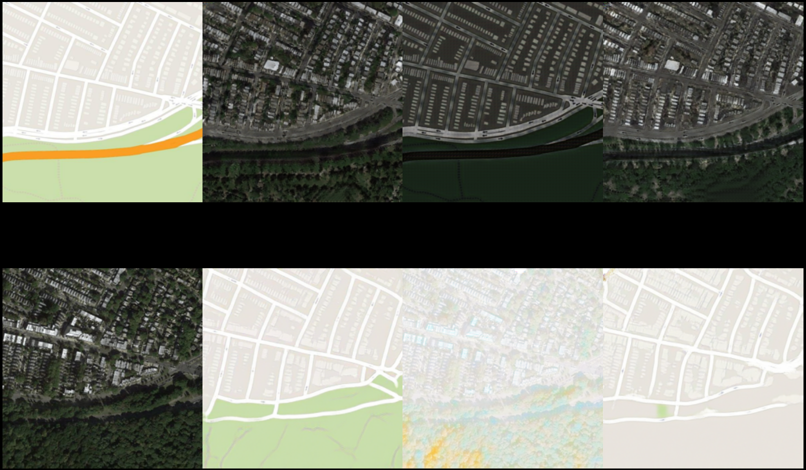


Figure:

CMP Facades dataset results

Model	Per-pixel acc.	Per-class acc.	Class IOU
CycleGAN	**0.29**	**0.15**	**0.09**
WCycleGAN-GP	0.24	0.14	0.07
SN-CycleGAN	0.21	0.11	0.06

Table: Classification performance of photo $\rightarrow$ labels for different methods on CMP Facades [].

CMP Facades : Images

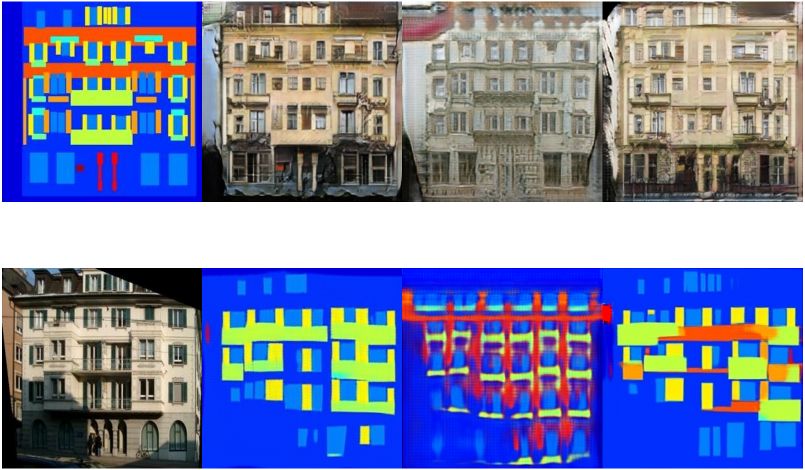


Figure: Left to right : Ground Truth, CycleGAN, W-CycleGAN-GP, SN-CycleGAN

Conclusion and further work

From these experiments we can outline a few conclusions :

- The new methods did not improve the classification performances on all datasets on the photo to label task.
- The new methods have improved the FCN score on the Cityscapes dataset.
- However the best FCN score was obtained for the W-CycleGAN-GP which, from a human point of view, did not produce plausible images. The SN-CycleGAN achieved a better quality of image (we think) but got a lower FCN score. Ultimately this makes us question the ability of the FCN score to evaluate how natural images are.

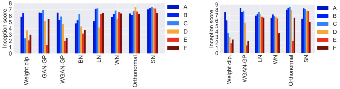
Further work

In order to properly conclude on the effectiveness of these methods we could have made the hyperparameters vary for each method. For instance we could have studied the importance of giving more weight to the wasserstein loss compare to the cyclic loss, etc...

Further work

Table 1: Hyper-parameter settings we tested in our experiments. †, ‡ and $\star$ are the hyperparameter settings following Gulrajani et al. (2017), Warde-Farley & Bengio (2017) and Radford et al. (2016), respectively.

Setting	α	β_1	β_2	n_{dis}
A†	0.0001	0.5	0.9	5
B‡	0.0001	0.5	0.999	1
C*	0.0002	0.5	0.999	1
D	0.001	0.5	0.9	5
E	0.001	0.5	0.999	5
F	0.001	0.9	0.999	5



(a) CIFAR-10 (b) STL-10

Figure 1: Inception scores on CIFAR-10 and STL-10 with different methods and hyperparameters (higher is better).

Figure: Table and Figure from SN-GAN paper

References

- Generative Adversarial Networks: I. Goodfellow et. al, NIPS 2014
- Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, JY Zhu et al., ICCV 2017.
- Wasserstein GAN, 2017
- Improved training of Wasserstein GAN, I. Gulrajani and al.
- Spectral Normalization for Generative Adversarial Networks, T. Miyato and al.
- Image-to-Image Translation with Conditional Adversarial Nets. Isola et al. CVPR 2017
- Improved Techniques for Training GANs, T. Salimans and al., 2016
- Are GANs Created Equal? A Large-Scale Study, Lucic, Kurach and al. 2018
- Cycle GAN Blog, Hardik Bansal
- This is a great paper!, Ian Goodfellow, 2017

Cityscapes : More images

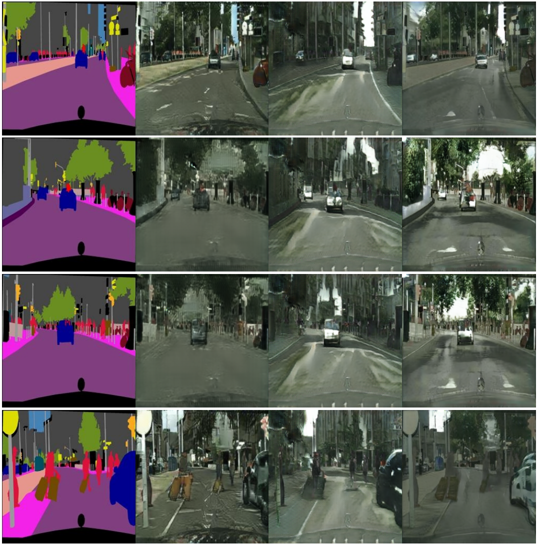
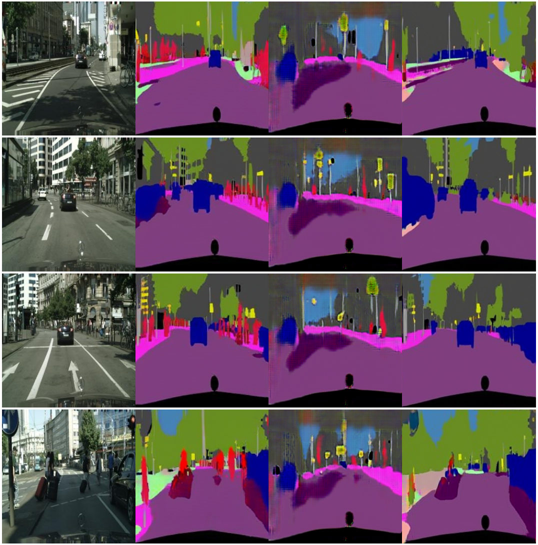


Figure: Left to right : Ground Truth, CycleGAN, W-CycleGAN-GP, SN-CycleGAN

Cityscapes : More images



CMP Facades : More images

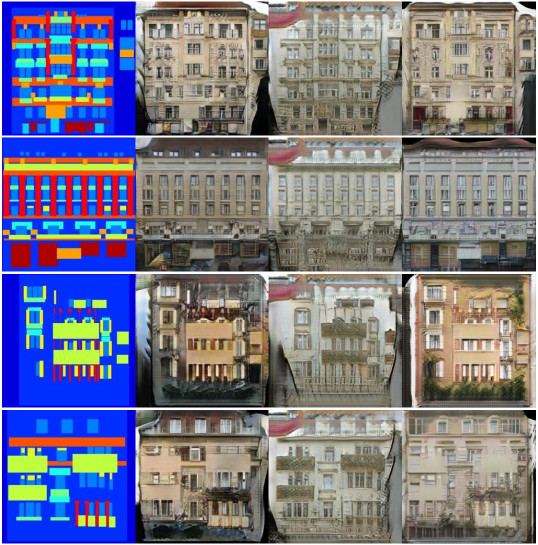


Figure: Left to right : Ground Truth, CycleGAN, W-CycleGAN-GP, SN-CycleGAN

CMP Facades : More images

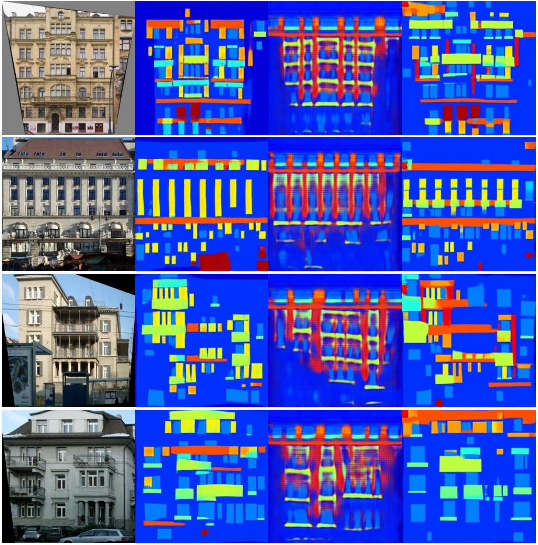


Figure: Left to right : Ground Truth, CycleGAN, W-CycleGAN-GP, SN-CycleGAN

Maps : More images

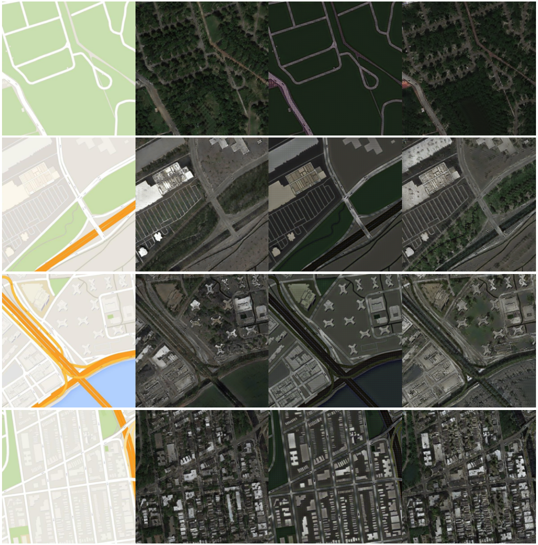
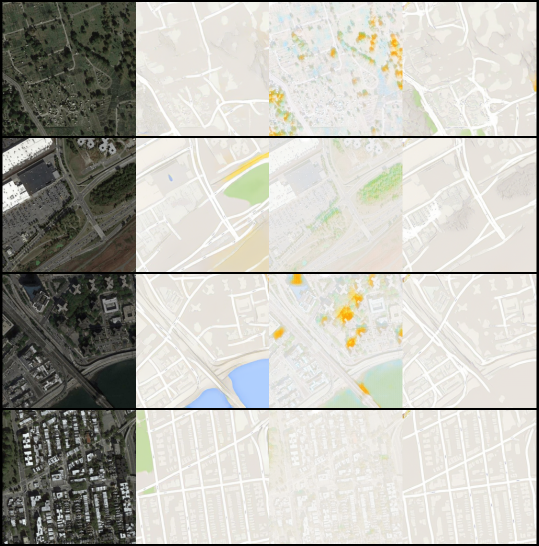


Figure: Left to right : Ground Truth, CycleGAN, W-CycleGAN-GP, SN-CycleGAN

Maps : More images



Supplementary Material - Loss

GAN

$$minV(D) = -\mathbb{E}[lg(D(x))] - \mathbb{E}[lg(1 - D(G(z))]$$
$$minV(G) = \mathbb{E}[lg(1 - D(G(z))]$$

WGAN-GP

$$minV(D) = -\mathbb{E}[D(x)] + \mathbb{E}[D(G(z))] + \lambda\mathbb{E}[(||\nabla_t D(t)||_2 - 1)^2]$$
$$minV(G) = -\mathbb{E}[D(G(z))]$$

SN-GAN (with Hinge Loss)

$$minV(D) = -\mathbb{E}[min(0, -1 + D(x))] - \mathbb{E}[min(0, -1 - D(G(z)))]$$
$$minV(G) = -\mathbb{E}[D(G(z))]$$

LS-GAN

$$minV(D) = \frac{1}{2}\mathbb{E}[(D(x) - 1)^2] + \frac{1}{2}\mathbb{E}[(D(G(z)))^2]$$
$$minV(G) = \frac{1}{2}\mathbb{E}[(D(G(z)) - 1)^2]$$

Supplementary Material - Loss

CycleGAN and SN-CycleGAN (with LS-Loss)

$$minV(G) = \mathbb{E}[(D_A(G_A(A)) - 1)^2] + \mathbb{E}[(D_B(G_B(B)) - 1)^2] + \\ + \lambda_A \mathbb{E}[||G_B(G_A(A)) - A||_1] + \lambda_B \mathbb{E}[||G_A(G_B(B)) - B||_1]$$

$$minV(D_A) = \mathbb{E}[(D_A(B) - 1)^2] + \mathbb{E}[(D_A(G_A(A)))^2]$$

$$minV(D_B) = \mathbb{E}[(D_B(A) - 1)^2] + \mathbb{E}[(D_B(G_B(B)))^2]$$

CycleWGAN-GP

$$minV(G) = -\mathbb{E}[D_A(G_A(A))] - \mathbb{E}[D_B(G_B(B))] + \\ + \lambda_A \mathbb{E}[||G_B(G_A(A)) - A||_1] + \lambda_B \mathbb{E}[||G_A(G_B(B)) - B||_1]$$

$$minV(D_A) = -\mathbb{E}[D_A(B)] + \mathbb{E}[D_A(G_A(A))] + \lambda \mathbb{E}[(||\nabla_t D_A(t)||_2 - 1)^2]$$

$$minV(D_B) = -\mathbb{E}[D_B(A)] + \mathbb{E}[D_B(G_B(B))] + \lambda \mathbb{E}[(||\nabla_t D_B(t)||_2 - 1)^2]$$

Metrics

We use here the following loss : Per-pixel accuracy :

$$\frac{\sum_i n_{ii}}{\sum_i t_i} \tag{1}$$

Per-class accuracy :

$$\frac{1}{n_{cl}} \sum_i \frac{n_{ii}}{t_i} \tag{2}$$

Mean Class Intersection-Over-Union :

$$\frac{1}{n_{cl}} \sum_i \frac{n_{ii}}{t_i + \sum_j n_{ji} - n_{ii}} \tag{3}$$

where n_{ij} is the number of pixels of class i predicted to belong to class j, n_{cl} is the number of different classes, and $t_i = \sum_j n_{ij}$ is the total number of pixels of class i

Generator $G(z)$	Kernel size	Resample	Output shape
z	-	-	128
Linear	-	-	$128 \times 4 \times 4$
Residual block	$[3\times3] \times 2$	Up	$128 \times 8 \times 8$
Residual block	$[3\times3] \times 2$	Up	$128 \times 16 \times 16$
Residual block	$[3\times3] \times 2$	Up	$128 \times 32 \times 32$
Conv, tanh	3×3	-	$3 \times 32 \times 32$

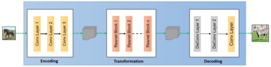
Critic $D(x)$	Kernel size	Resample	Output shape
Residual block	$[3\times3] \times 2$	Down	$128 \times 16 \times 16$
Residual block	$[3\times3] \times 2$	Down	$128 \times 8 \times 8$
Residual block	$[3\times3] \times 2$	-	$128 \times 8 \times 8$
Residual block	$[3\times3] \times 2$	-	$128 \times 8 \times 8$
ReLU, mean pool	-	-	128
Linear	-	-	1

$z \in \mathbb{R}^{128} \sim \mathcal{N}(0, I)$
dense, $6 \times 6 \times 512$
ResBlock up 256
ResBlock up 128
ResBlock up 64
BN, ReLU, 3×3 conv, 3 Tanh

(a) Generator

RGB image $x \in \mathbb{R}^{48\times48\times3}$
ResBlock down 64
ResBlock down 128
ResBlock down 256
ResBlock down 512
ResBlock 1024
ReLU
Global sum pooling
dense $\rightarrow$ 1

(b) Discriminator



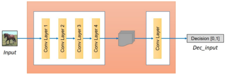


Figure: Generator and Discriminator structure. Left to Right : WGAN-GP, SN-GAN, CycleGAN

SN-GAN : SGD

Algorithm 1 SGD with spectral normalization

- Initialize $\tilde{\boldsymbol{u}}_l \in \mathcal{R}^{d_l}$ for $l = 1, \ldots, L$ with a random vector (sampled from isotropic distribution).
- For each update and each layer l:
  1. Apply power iteration method to a unnormalized weight W^l:
  $$\tilde{\boldsymbol{v}}_l \leftarrow (W^l)^{\mathrm{T}} \tilde{\boldsymbol{u}}_l / \|(W^l)^{\mathrm{T}} \tilde{\boldsymbol{u}}_l\|_2 \tag{20}$$
  $$\tilde{\boldsymbol{u}}_l \leftarrow W^l \tilde{\boldsymbol{v}}_l / \|W^l \tilde{\boldsymbol{v}}_l\|_2 \tag{21}$$
  2. Calculate $\bar{W}_{\mathrm{SN}}$ with the spectral norm:
  $$\bar{W}^l_{\mathrm{SN}}(W^l) = W^l / \sigma(W^l), \text{ where } \sigma(W^l) = \tilde{\boldsymbol{u}}_l^{\mathrm{T}} W^l \tilde{\boldsymbol{v}}_l \tag{22}$$
  3. Update W^l with SGD on mini-batch dataset $\mathcal{D}_M$ with a learning rate α:
  $$W^l \leftarrow W^l - \alpha \nabla_{W^l} \ell(\bar{W}^l_{\mathrm{SN}}(W^l), \mathcal{D}_M) \tag{23}$$