# Improved Training of Wasserstein GANs

Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, Aaron Courville

# Outline

- Introduction
- Wasserstein GAN
- Gradient Penalty
- Experiments
- Conclusion

# Introduction

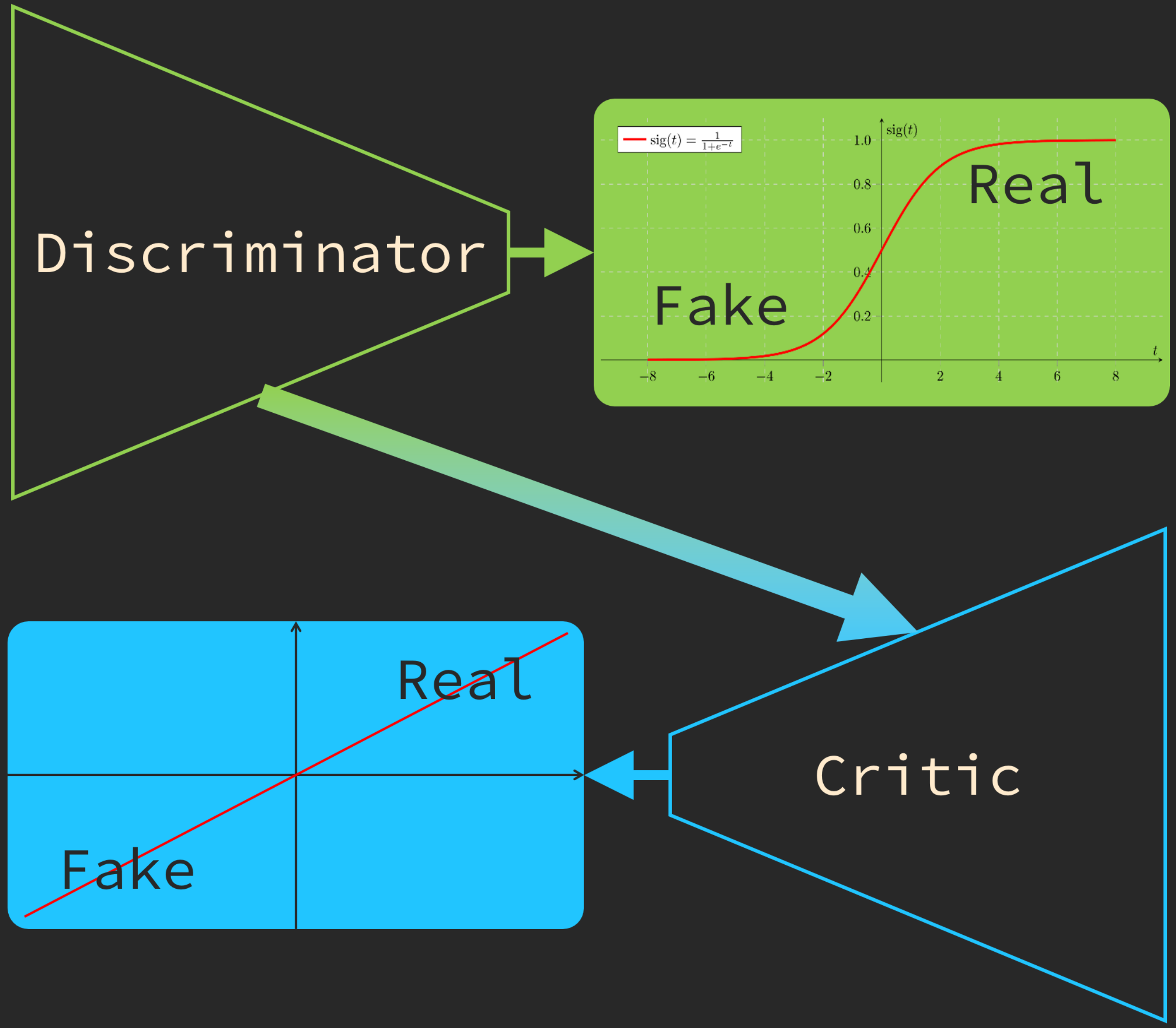
GAN 雖然是種強大的生成模型，但卻容易因為梯度消失導致訓練不穩定。

而 WGAN 利用一些簡單的改動便大幅改善原始 GAN 的問題。

然而在 WGAN 使用的權重剪裁，還是有可能導致梯度發散或消失。

為此，本論文改用「梯度懲罰」的損失函數來約束權重，使得 WGAN 可以穩定收斂。

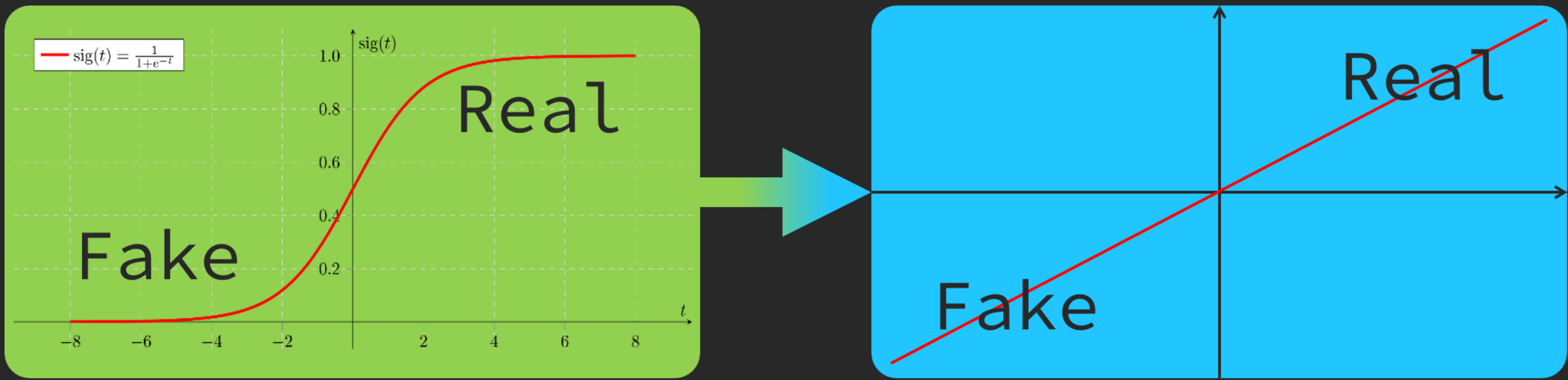
# Wasserstein GAN



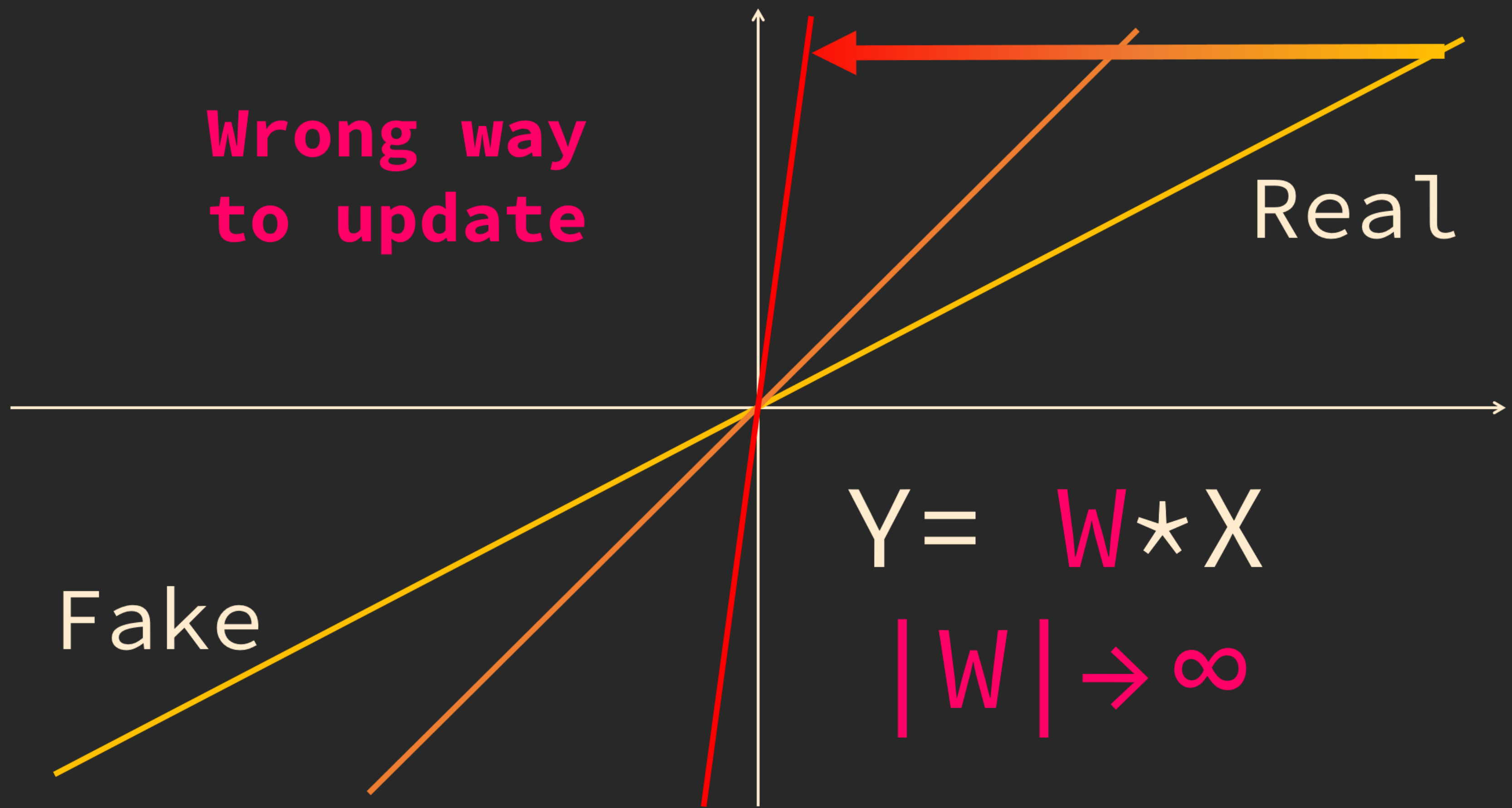
# 判別器 => 評估器

解決梯度消失的問題

1. 移除輸出的 Sigmoid
2. 移除 loss 的 log

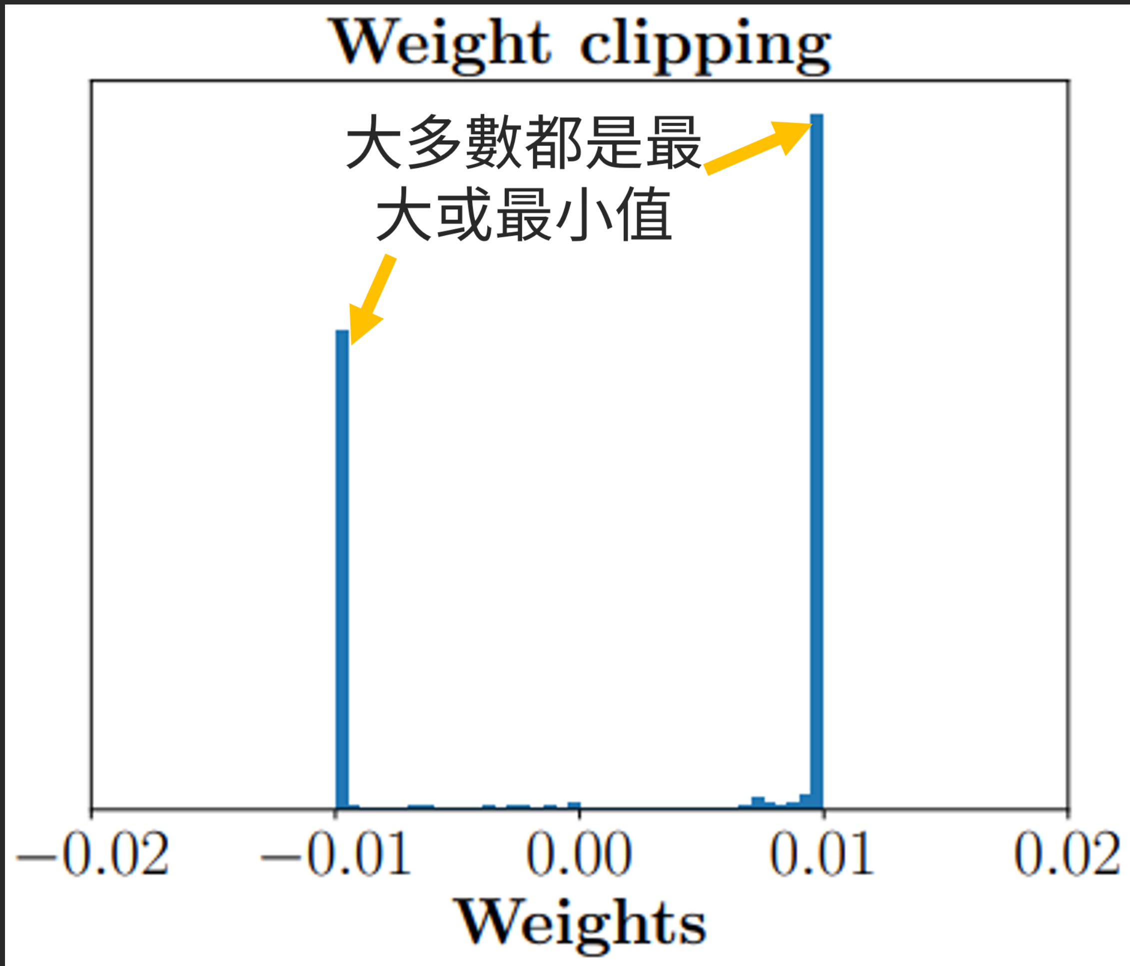


# 判別器 => 評估器




不做限制的話可能會導致權重無限制增長

# Weight Clipping




在 WGAN 原文中使用 Weight Clipping 來限制權重大小。

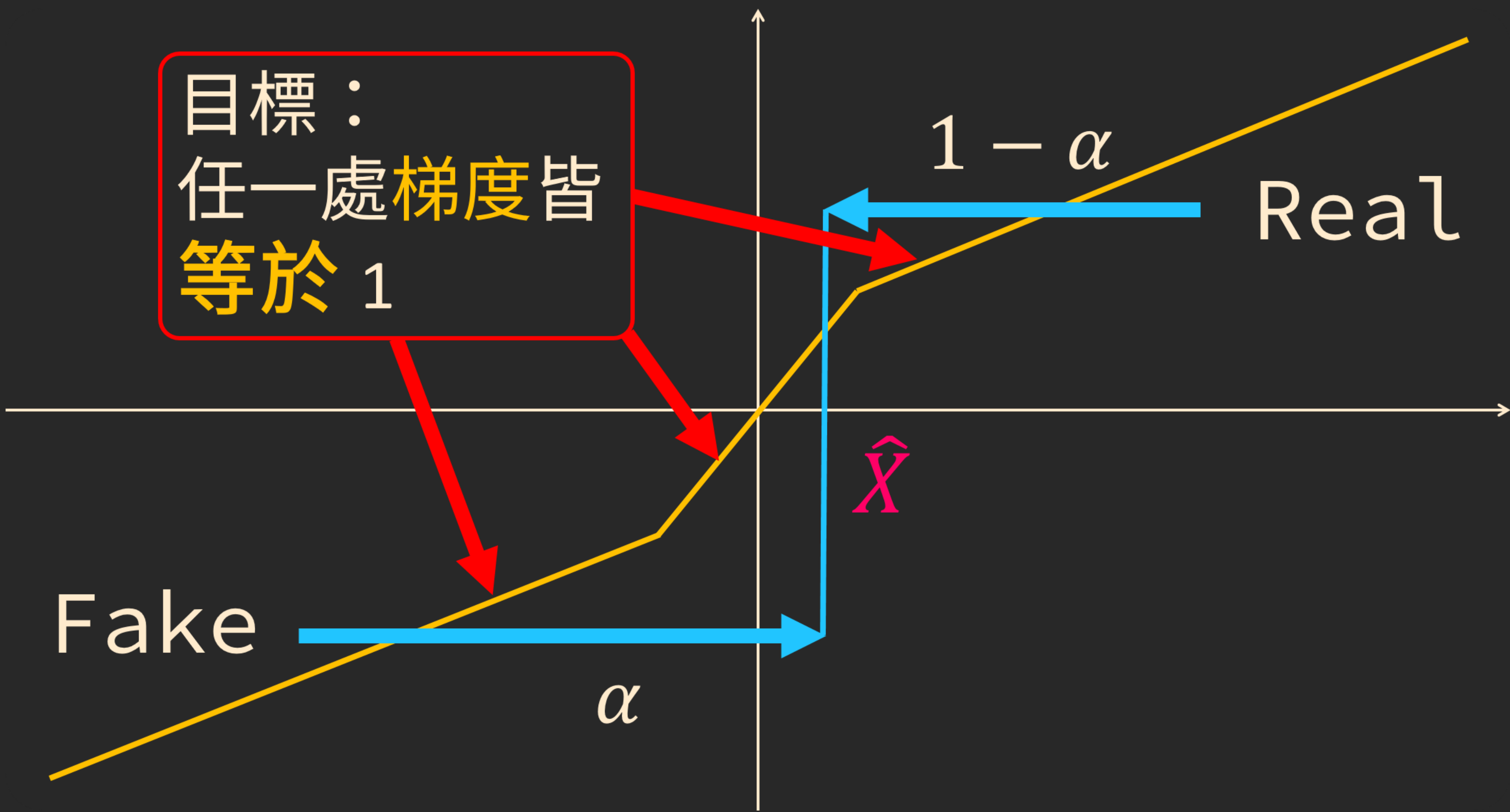
但卻會造成梯度消失或爆炸。

# Gradient Penalty

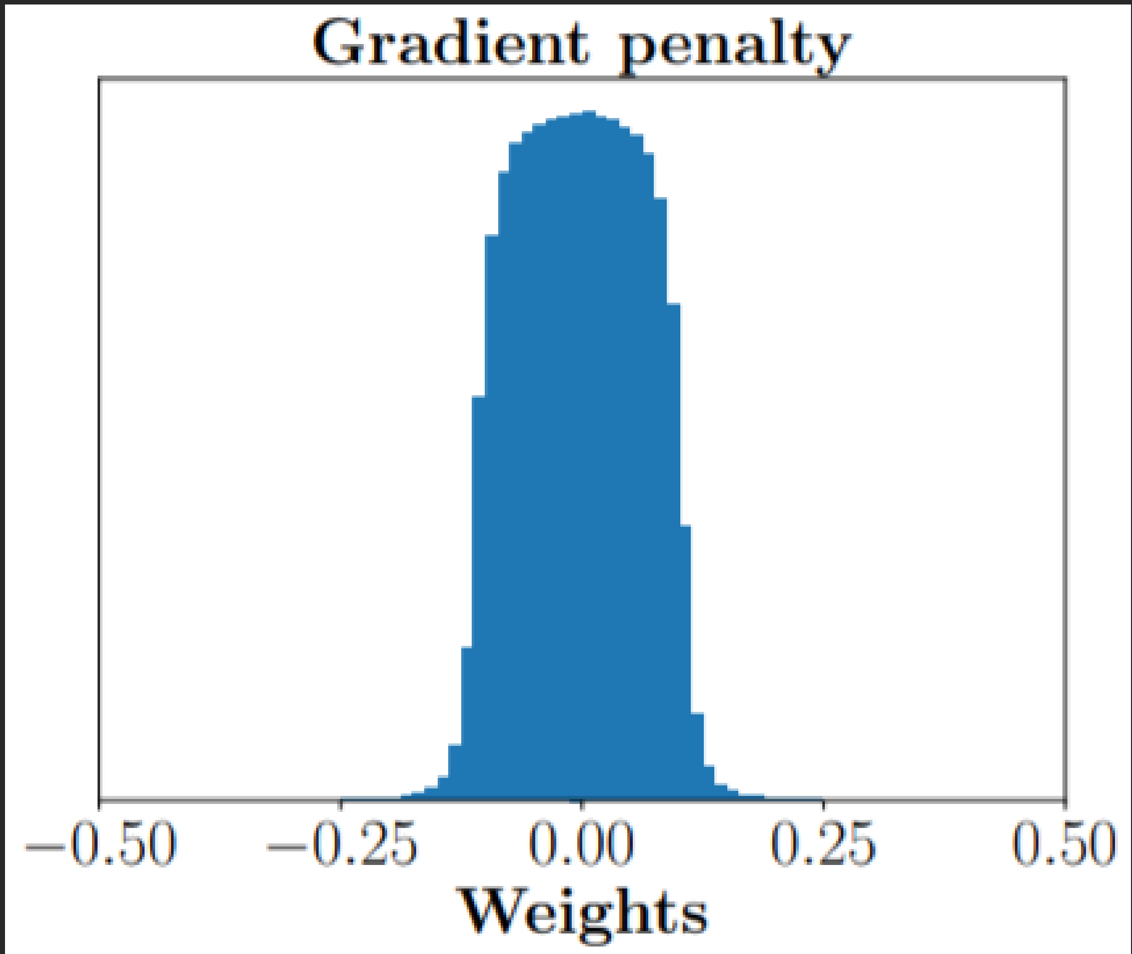
$$\hat{X} = (1 - \alpha)X + \alpha\tilde{X}$$

$$(\|\nabla_{\hat{x}} Critic(\hat{x})\|_2 - 1)^2$$

# Gradient Penalty

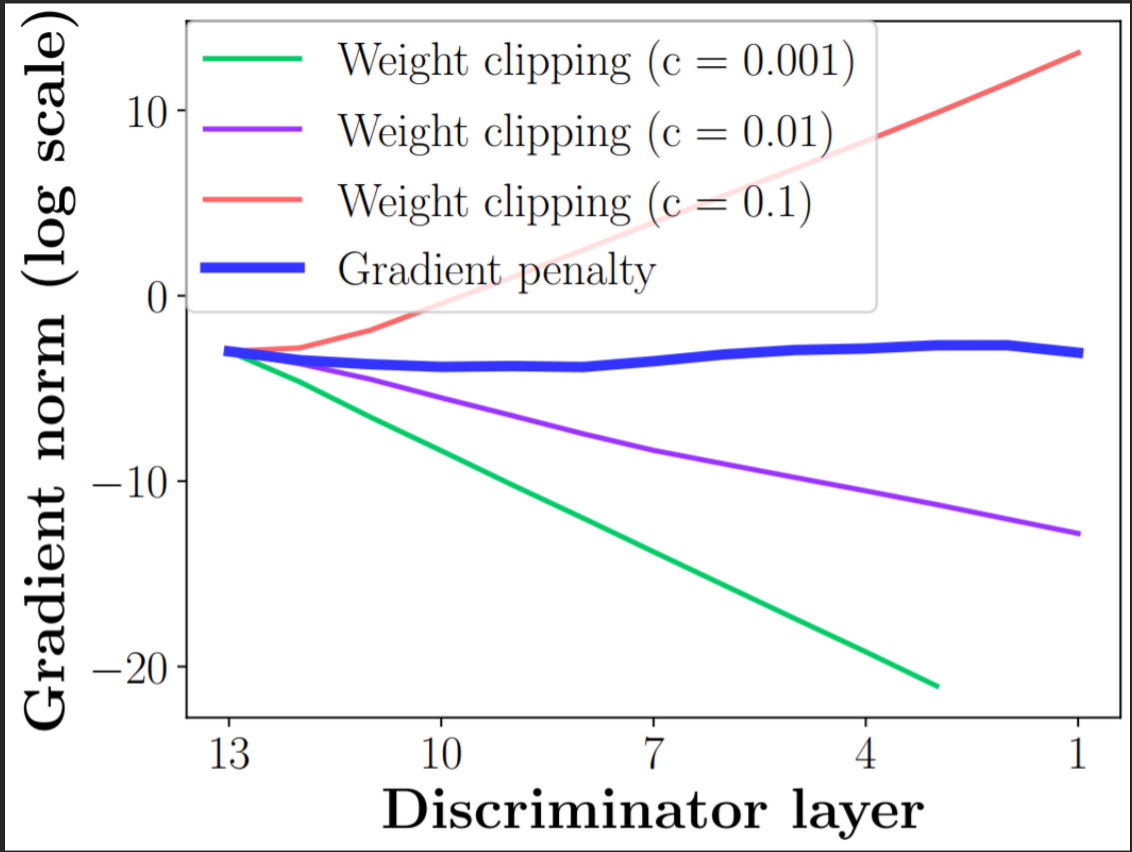


# Gradient Penalty




權重均勻的分佈在 0 的周圍，使模型具有更高的強健性。

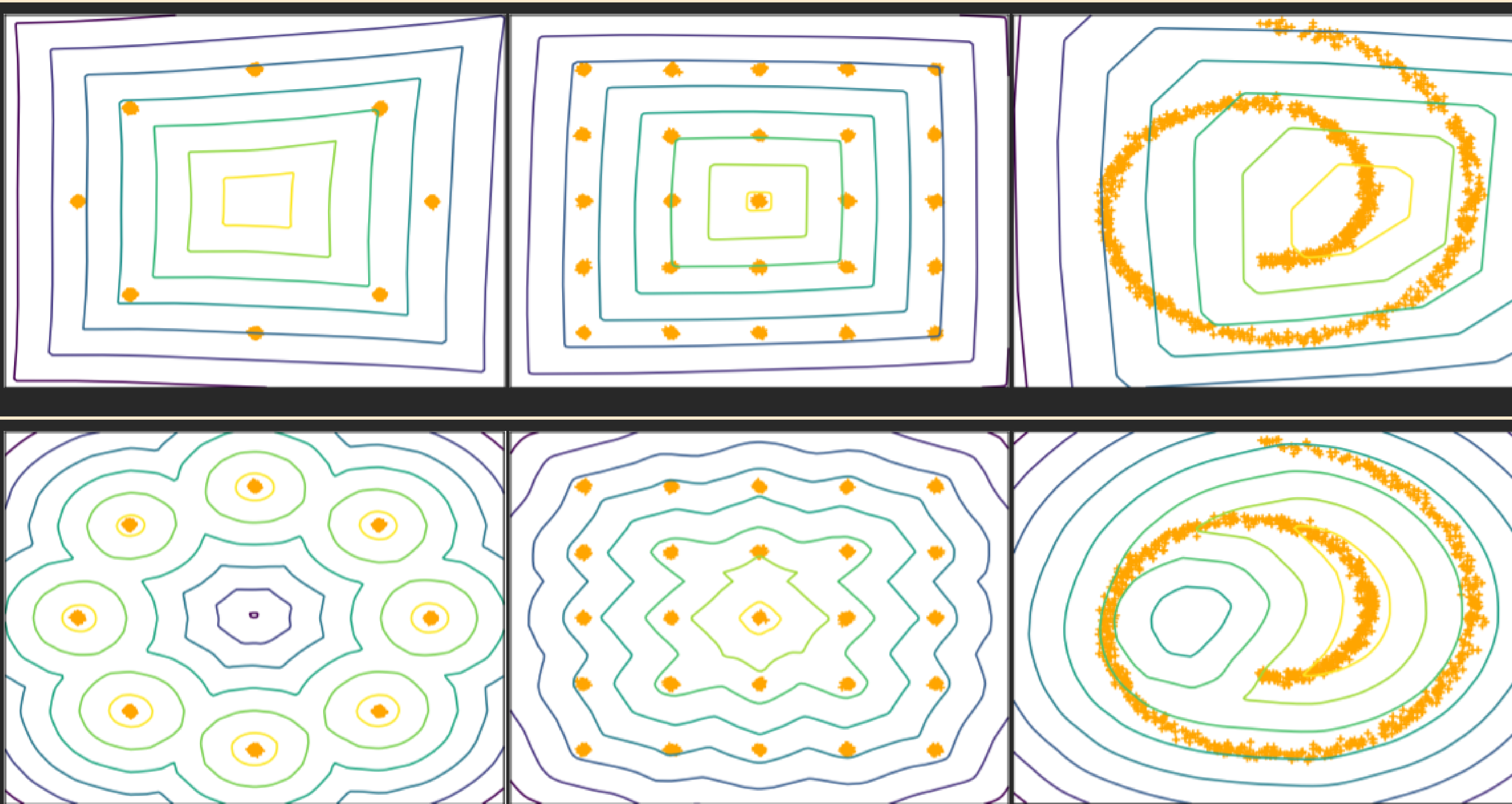
# Gradient Penalty




即便是在深層的模型中，Gradient Penalty 也可以使梯度保持穩定，而不會像 Weight Clipping 一樣發散或消失。

# Gradient Penalty

| | 8 Gaussians | 25 Gaussians | Swiss Roll |
| --- | --- | --- | --- |
| WC | | | |
| GP | | | |



# Loss

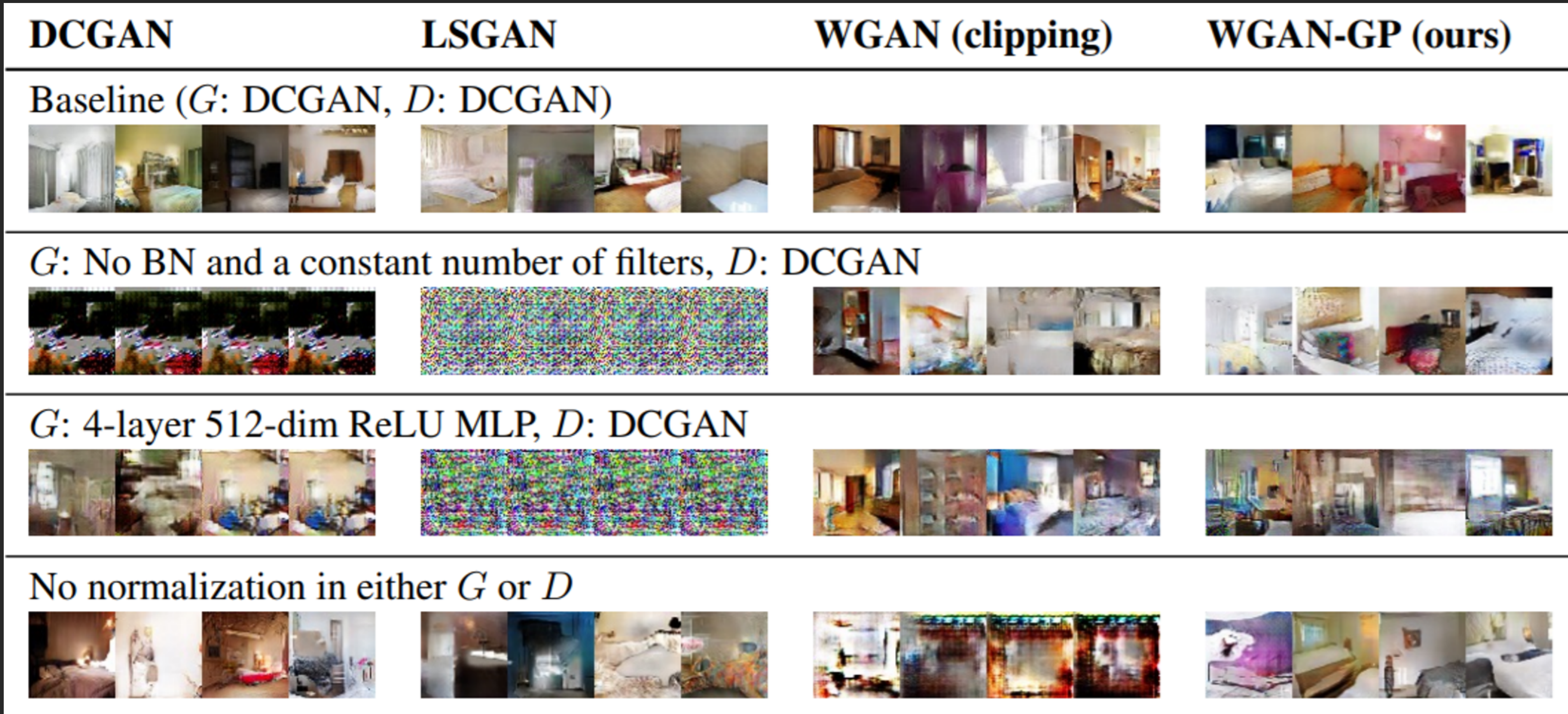
Generate Loss

$$-\underbrace{\boxed{Critic(G(z))}}_{\text{越大越好}}$$

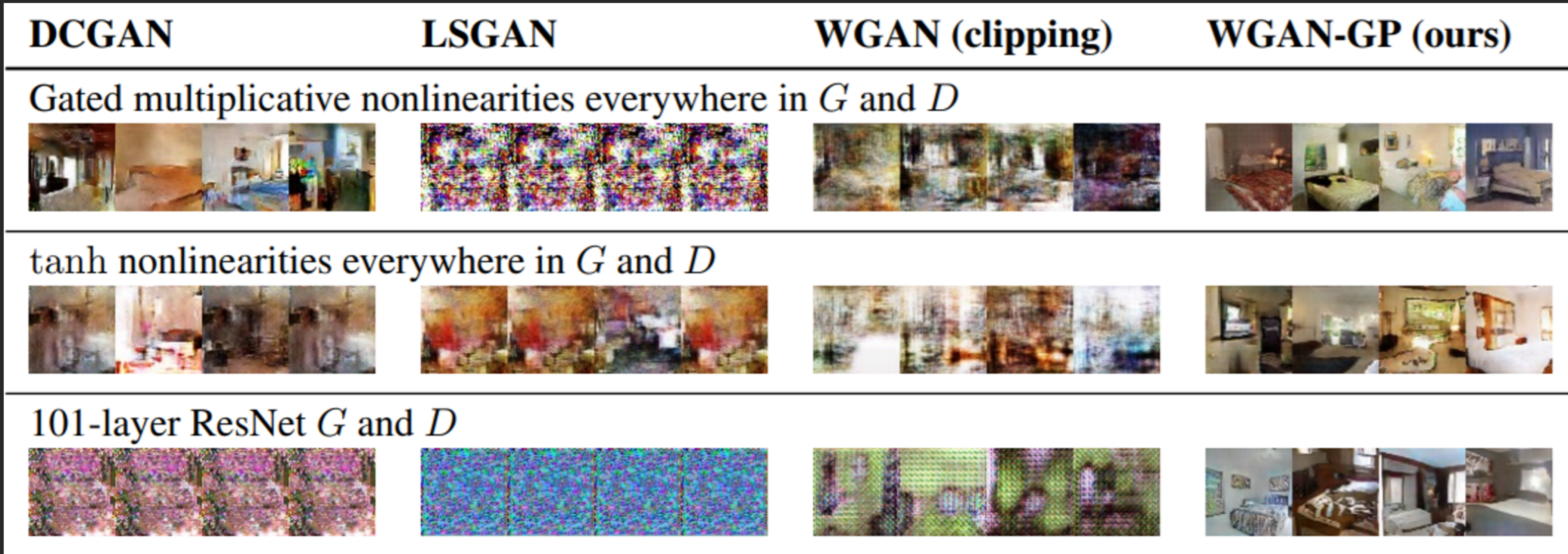
Critic Loss

$$\overbrace{\boxed{Critic(G(z))}}^{\text{越小越好}} - \overbrace{\boxed{Critic(x)}}^{\text{越大越好}} + \lambda(\|\nabla_{\hat{x}} Critic(\hat{x})\|_2 - 1)^2$$

# Robustness

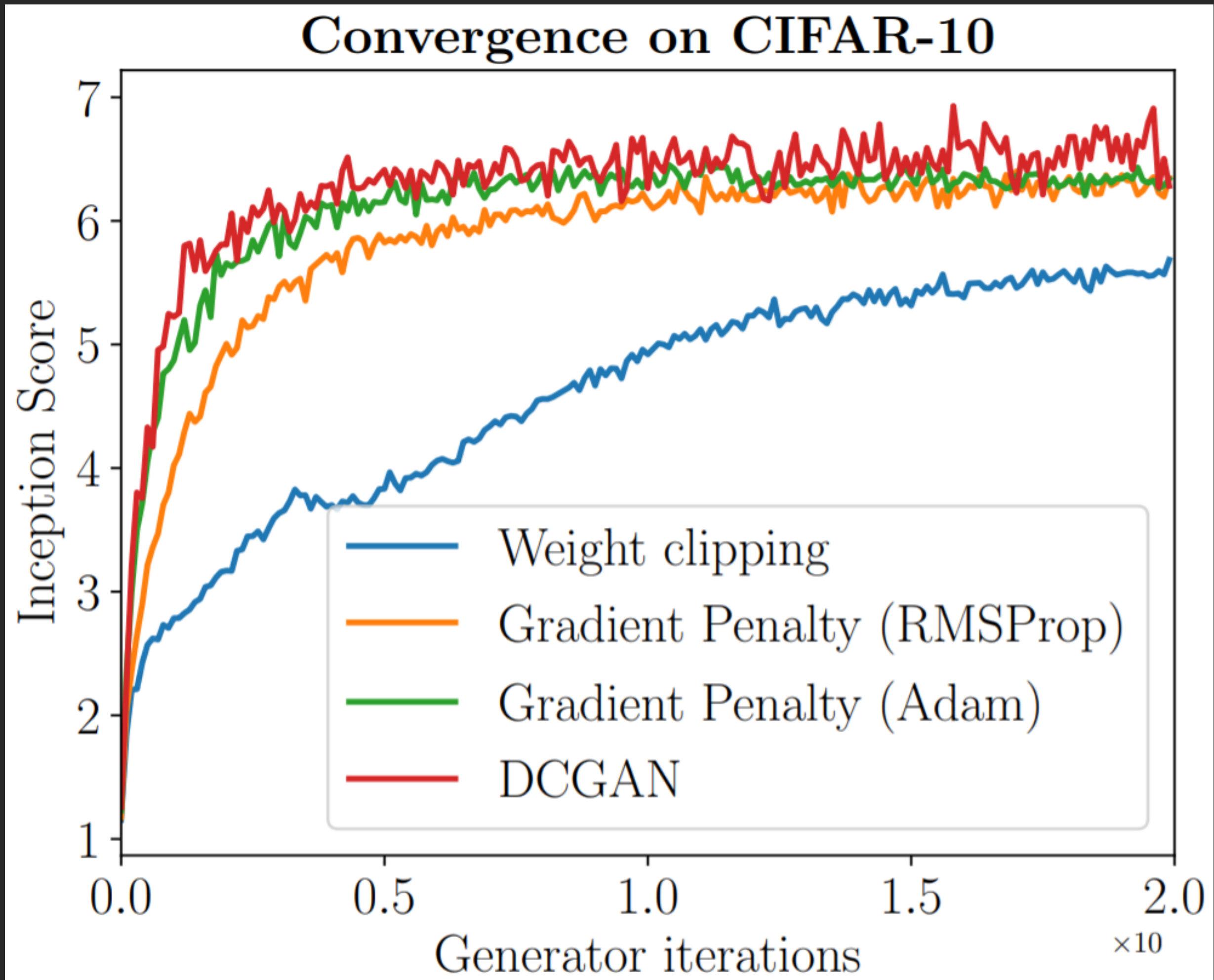

DCGAN
LSGAN
WGAN (clipping)
WGAN-GP (ours)
Baseline ($G$: DCGAN, $D$: DCGAN)
$G$: No BN and a constant number of filters, $D$: DCGAN
$G$: 4-layer 512-dim ReLU MLP, $D$: DCGAN
No normalization in either $G$ or $D$

# Robustness



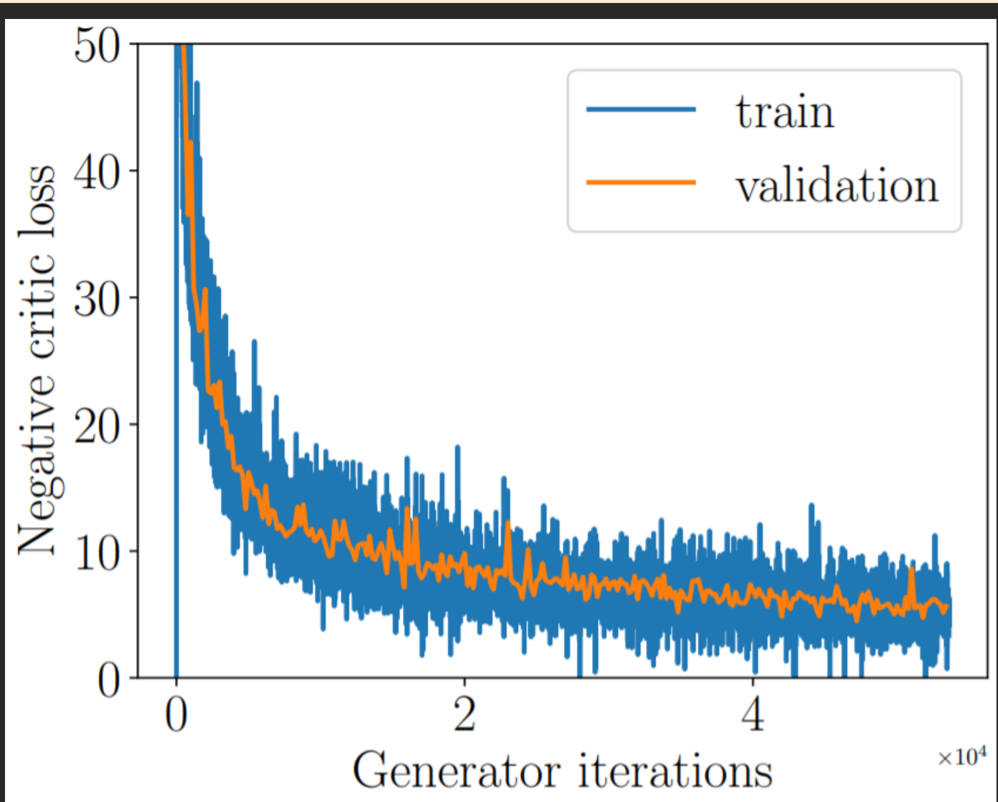
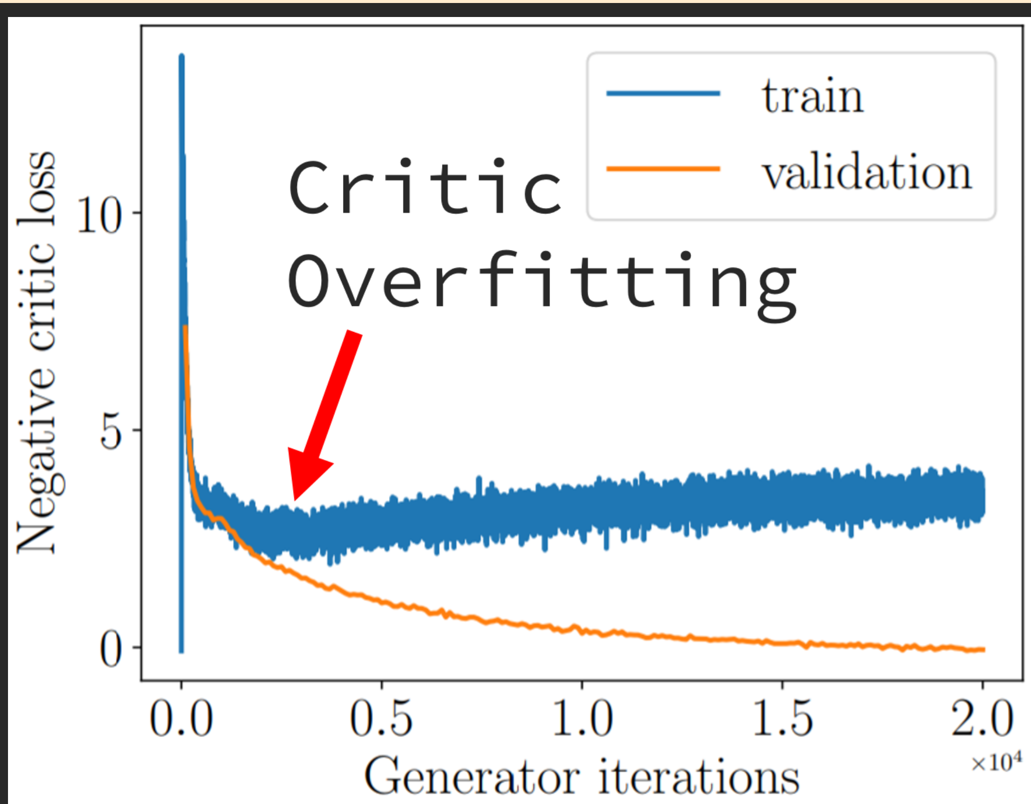

利用 Gradient Penalty 可以使 WGAN 泛化到各模型架構

# Inception Score




可以達到與 DCGCN 相當的水平，且訓練收斂性不會限制於模型架構。

# Overfitting

| Non Overfitting | Overfitting |
|---|---|
|  |  |





當使用 WGAN-GP 時，若真實樣本過少，
Critic 會比 Generator 更快發生 Overfitting

# Conclusion

- WGAN 利用簡單的技巧解決了辨別器造成的梯度消失問題。
- 而 Gradient Penalty 為模型帶來更高的強健性，讓 WGAN 的訓練不受模型架構影響收斂。

# Note

- 在 WGAN-GP 中，所有的 Batch Norm 都替換成 Layer Norm。
- 論文有提到，對雙向 Gradient Penalty 的效果比單向 Gradient Penalty(只對大於 1 的梯度懲罰)更好。
- 非線性非平滑函數(e.g. ELU if α != 1)無法訓練 WGAN-GP