

SAGAN 別人的簡介

2018年7月7日 下午 03:36

<https://towardsdatascience.com/not-just-another-gan-paper-sagan-96e649f01a6b>

問題點

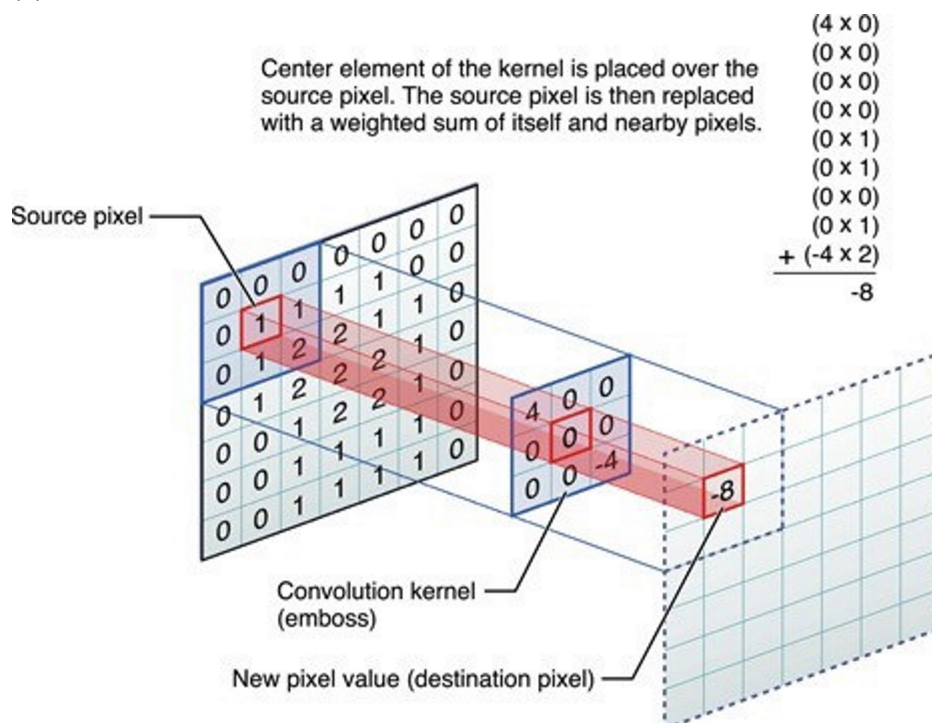
Convolutional GANs (以後都用 CGANs 簡稱)的問題是在學習 差異性大的多類別 時通常學得不好，特別是在某些類別上更難學習像是通常可以學好海、天空等簡單幾何，但狗、馬等特定幾何卻學得不好或者是能夠產生狗的皮毛紋理，但卻生不出清楚的腳部形狀

為甚麼會出現這問題

因為 卷積 神經網路的特性，導致它只能關注某個特定區域

例如圖片左上對圖片右下並不會有任何影響

又或者是如下面這張圖，除了 卷積的範圍外，不會對 -8 的產生造成任何影響



然而你可能會問說，為何不能用超大的 filter 捕捉整個網路呢？

當然可以，但計算效率會變慢

又或者可能會問，為何不用很深的網路，讓感知範圍夠大呢？

當然可以，但很深就代表要更多參數，參數數量可能會過多(可能會過耦合?)

Self-Attention GANs

使用 NLP 上有名的機制 attention 來維持 效率與大範圍感知 的平衡

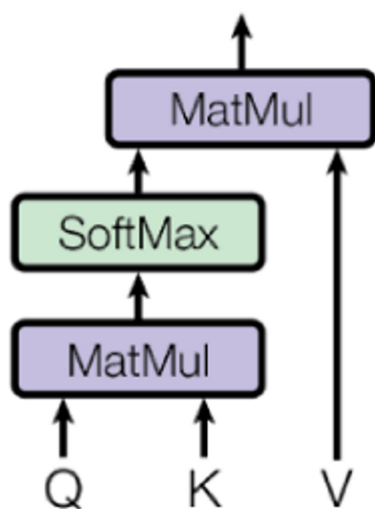
Attention 是甚麼？

其實就是權重

實作上就是讓 query 與 key 決定 value 的影響力

首先讓 query 與 key 相乘，經過 softmax 後得到權重

再與 value 相乘，得到真正的輸出結果



那甚麼又是 Self-Attention 呢？

那就是 Q K V 都是相同的

可有多個

模型

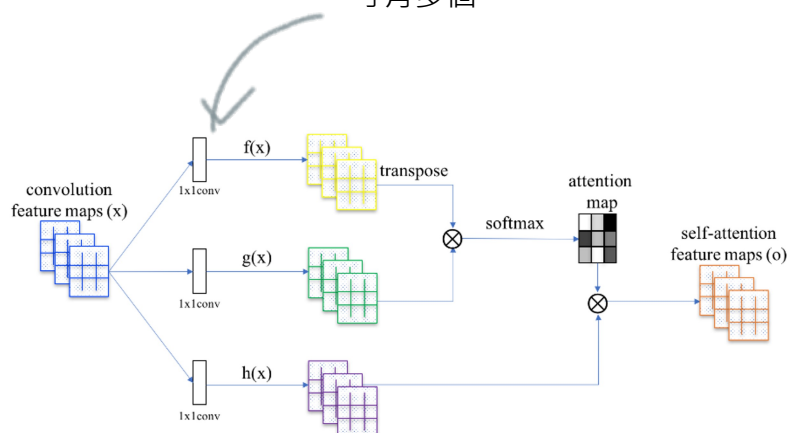


Figure 2: The proposed self-attention mechanism. The \otimes denotes matrix multiplication. The softmax operation is performed on each row.

最後，則是把輸出 o 乘上 可被學習的參數 y，加上 原始圖片 x，才是最後的輸出

$$O = y * o + x$$

論文也建議把 y 初始化為0，他們希望網路能依靠 關注區域附近的線索，因為這樣比較簡單且網路下降時也會學著 給 y 數值並使用 self-attention

這網路有助於捕捉相鄰遙遠的相依細節，並與捲積層互補



Figure 1: The proposed SAGAN generates images by leveraging complementary features in distant portions of the image rather than local regions of fixed shape to generate consistent objects/scenarios. In each row, the first image shows five representative query locations with color coded dots. The other five images are attention maps for those query locations, with corresponding color coded arrows summarizing the most-attended regions.

他們使用 hinge version 的 對抗損失函數，也沒有特別講這個損失函數

?

$$L_D = -\mathbb{E}_{(x,y) \sim p_{data}} [\min(0, -1 + D(x, y))] - \mathbb{E}_{z \sim p_z, y \sim p_{data}} [\min(0, -1 - D(G(z), y))],$$
$$L_G = -\mathbb{E}_{z \sim p_z, y \sim p_{data}} D(G(z), y),$$

z 是生成的圖片， x, y 是真實圖片

一些細節

1. 他們把 self attention 用在 生成與判別網路上
2. 他們把 spectral normalization 用在 生成與判別網路 的 權重上，不像之前的論文只用在其中一個上
並把 spectral norm 設置成 1 去 限制 權重的 Lipschitz 常數
這只是用來控制梯度的
原始論文在此 <https://arxiv.org/abs/1802.05957>
3. 他們使用兩倍更新規則(two-timescale update rule (TTUR))，就只是簡單的對 生成器和判別器用上不同學習率而已
4. 使用 Inception Score (IS, higher is better) 及 Frechet-Inception Distance(FID, lower is better) 作為效能的評估方式

結果

論文解釋了 spectral normalization 及 TTUR 怎麼幫助 GAN 收斂得更好

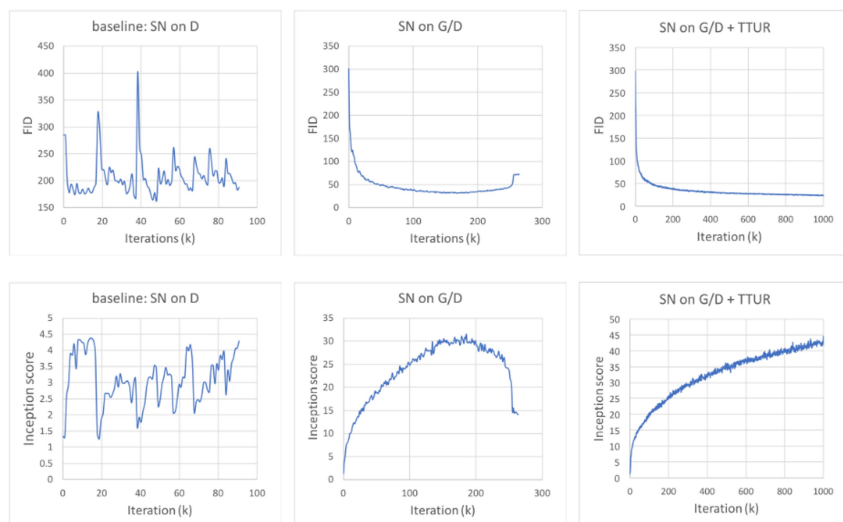


Figure 3: Training curves for the baseline model and our models with the proposed stabilization techniques, “SN on G/D ” and two-timescale learning rates (TTUR). All models are trained with 1:1 balanced updates for G and D .

達到了 state of the art

Model	Inception Score	FID
AC-GAN [31]	28.5	/
SNGAN-projection [17]	36.8	27.62*
SAGAN	52.52	18.65

生成的結果

