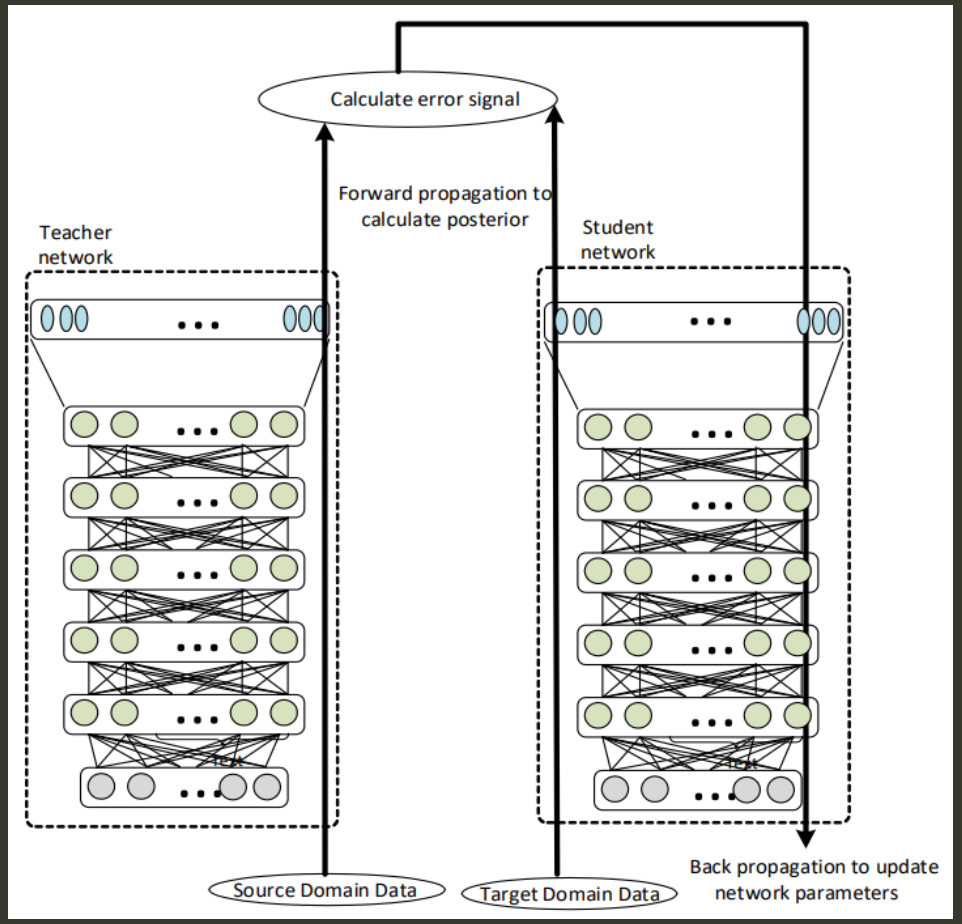
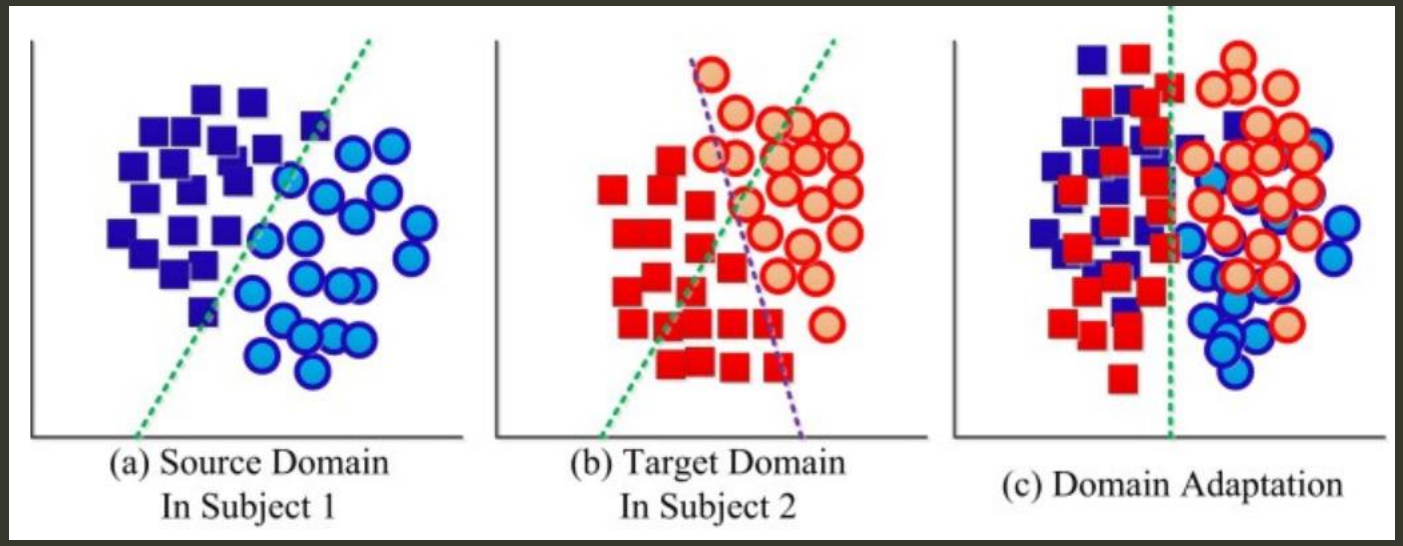
Adversarial Domain Adaptation

棒棒生

失去原有 domain 的能力

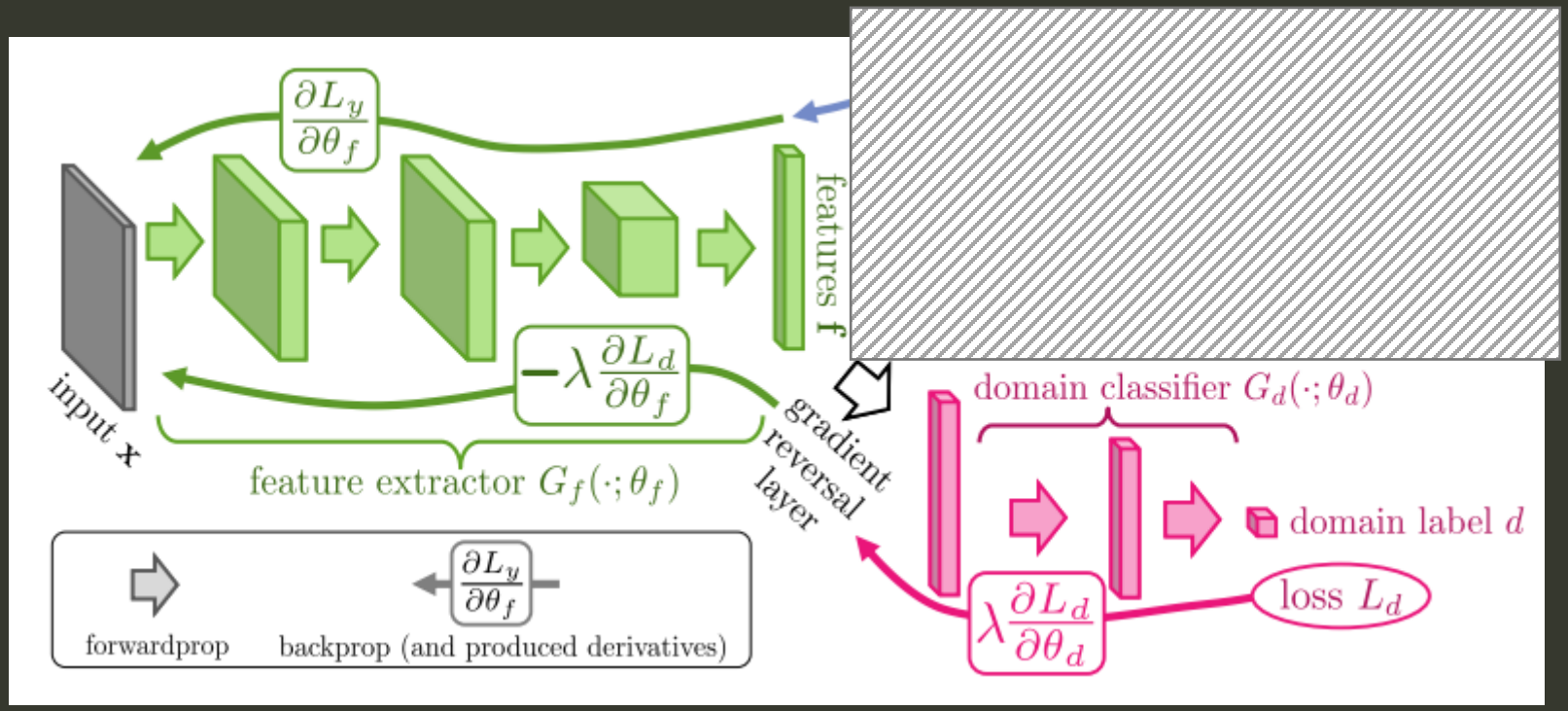
- Q: 有沒有更好的辦法?
- A: Feature-level adaptation





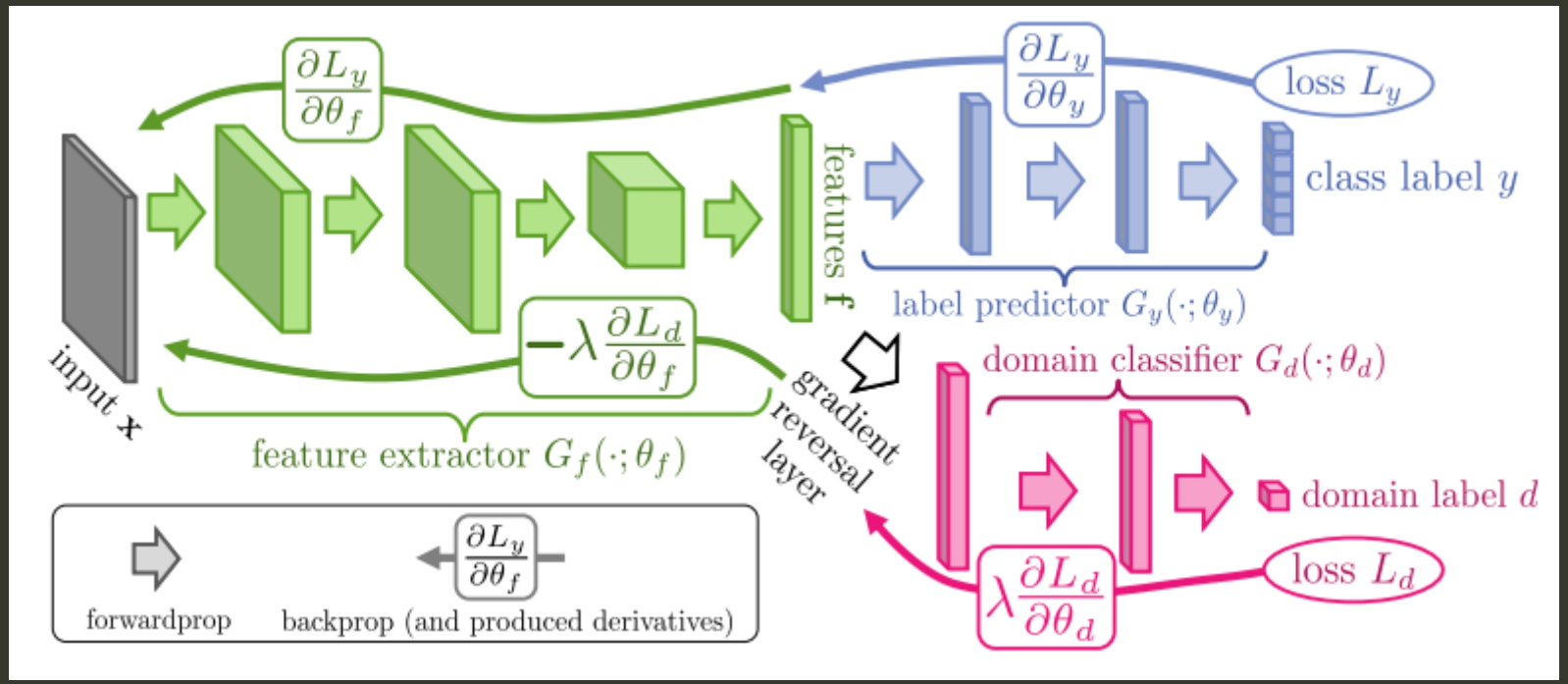
- Conditions:
  - Source domain 有 labels
  - Target domain 無 labels
- 將 $P\big(G(x_{src})\big) = P\left(G\big(x_{tgt}\big)\right)$ 兩個分布弄成一樣, 但怎麼做?
  - 早期使用 MMD (Maximum Mean Discrepancy)
  - 現在我們可以使用 GAN
- 只能保證 source domain 的分類, 那為什能 work?
  - 由於我們消除了 domain 之間的差異, 因此可以期望這時候的 source domain classifier 也能作用在 target domain

Domain-Adversarial Training of Neural Networks



正常的 GAN 架構

Domain-Adversarial Training of Neural Networks



- 需加上 Classifier 否則會有 trivial solution

GAN 架構

$$Div(P_d \parallel P_G) = \max_D \{\mathbb{E}_{x\sim P_d} D(x) - \mathbb{E}_{x\sim P_G} f\big(D(x)\big)\}$$
$$G^* = arg\min_G \{Div(P_d \parallel P_G)\}$$

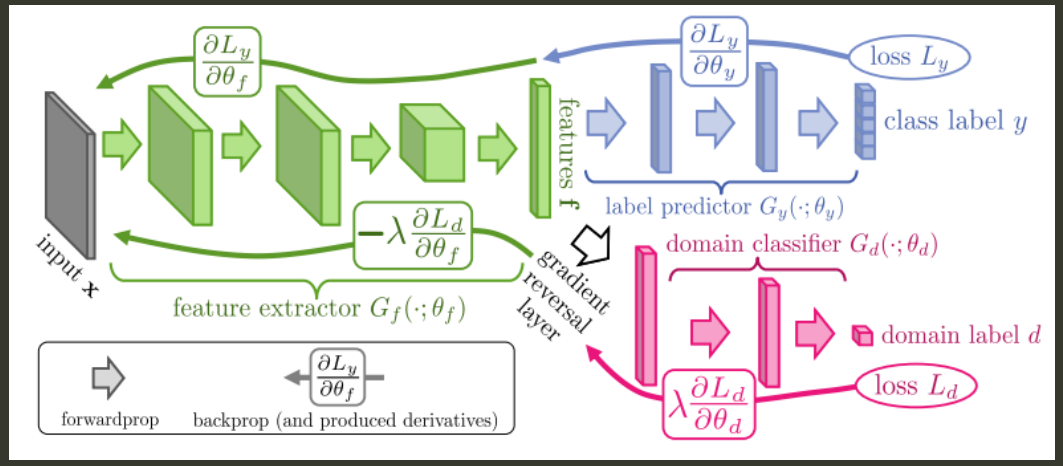
- 解最佳化 D 就是在量測兩個 distribution 的 divergence
  - 帶入不同的 f 會表示不同的 divergence 量測
  - 實務上我們指量測一個 batch 的 distribution, 因此也不需要解得太精確
- 因此 G 就是 divergence 最小的那個 (G^*)
  - 當 divergence = 0, 表示 $P_d = P_G$, 成功把兩個 distribution 重合
- 以上可以看出為何 GAN 都先 optimize D, 再 optimize G

GAN 架構

$$Div(P_d \parallel P_G) = \max_D \{\mathbb{E}_{x\sim P_d} D(x) - \mathbb{E}_{x\sim P_G} f\big(D(x)\big)\}$$

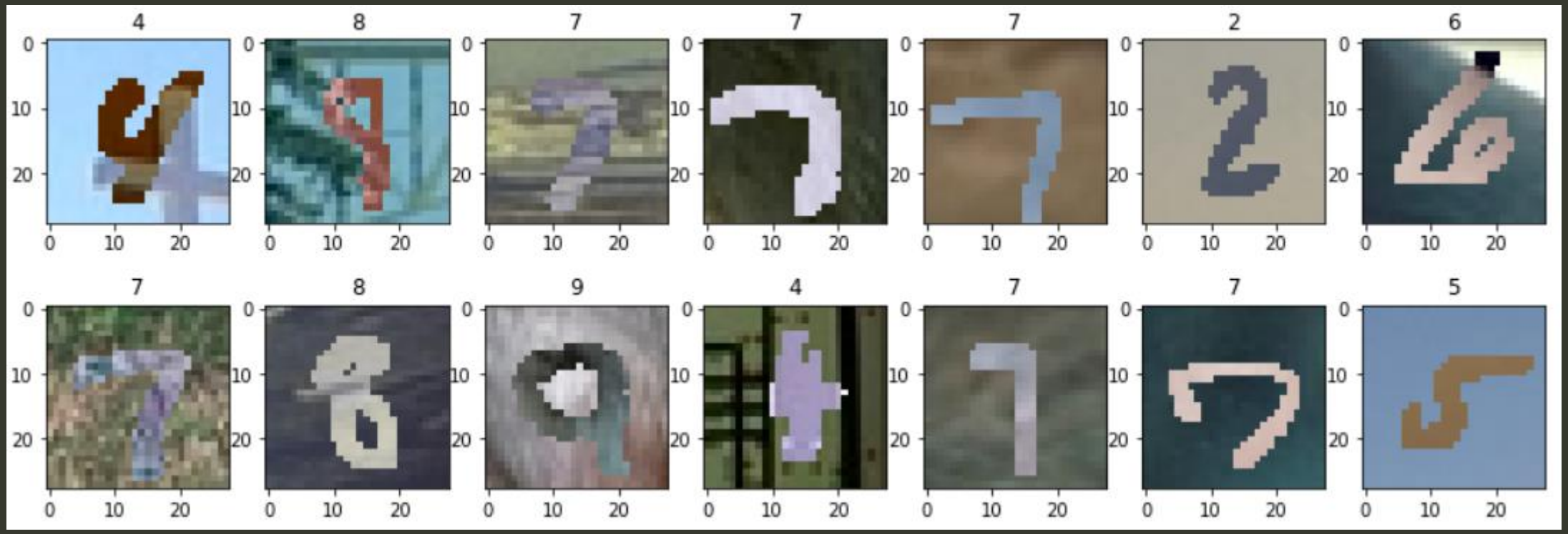
$$G^* = arg\min_G \{Div(P_d \parallel P_G) + pred\_loss(G)\}$$

- 別忘了還有 Classifier
- 訓練 G 除了要能欺騙 D, 同時也要能降低 prediction error
- 相當於 G 的目標函式多了一個 regularization term

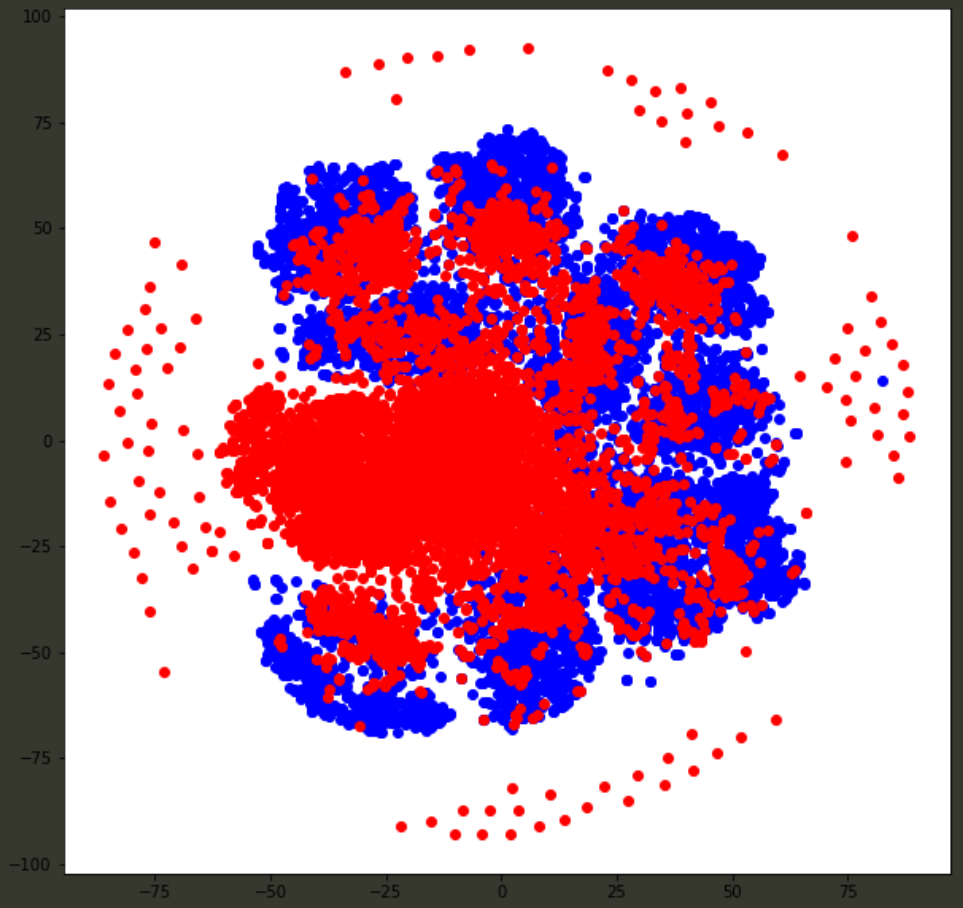


Source and Target Domains

- Source domain: mnist
- Target domain: mnist_m

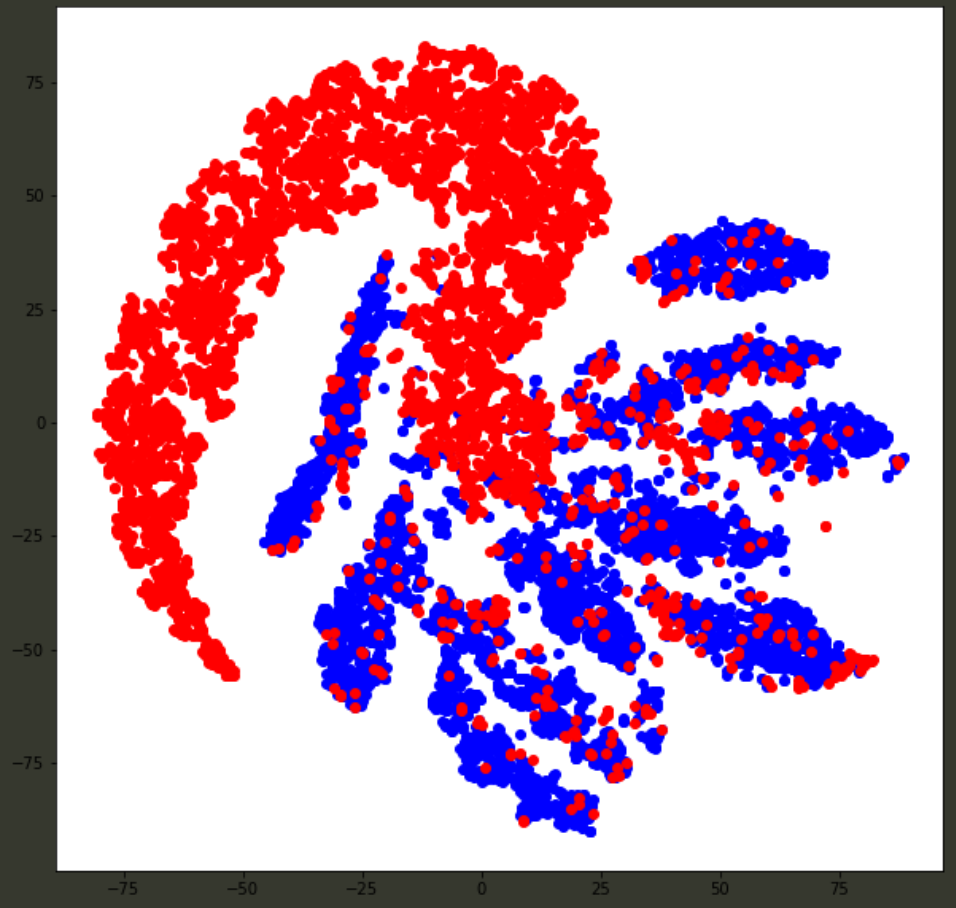


tSNE Before Domain Adaptation

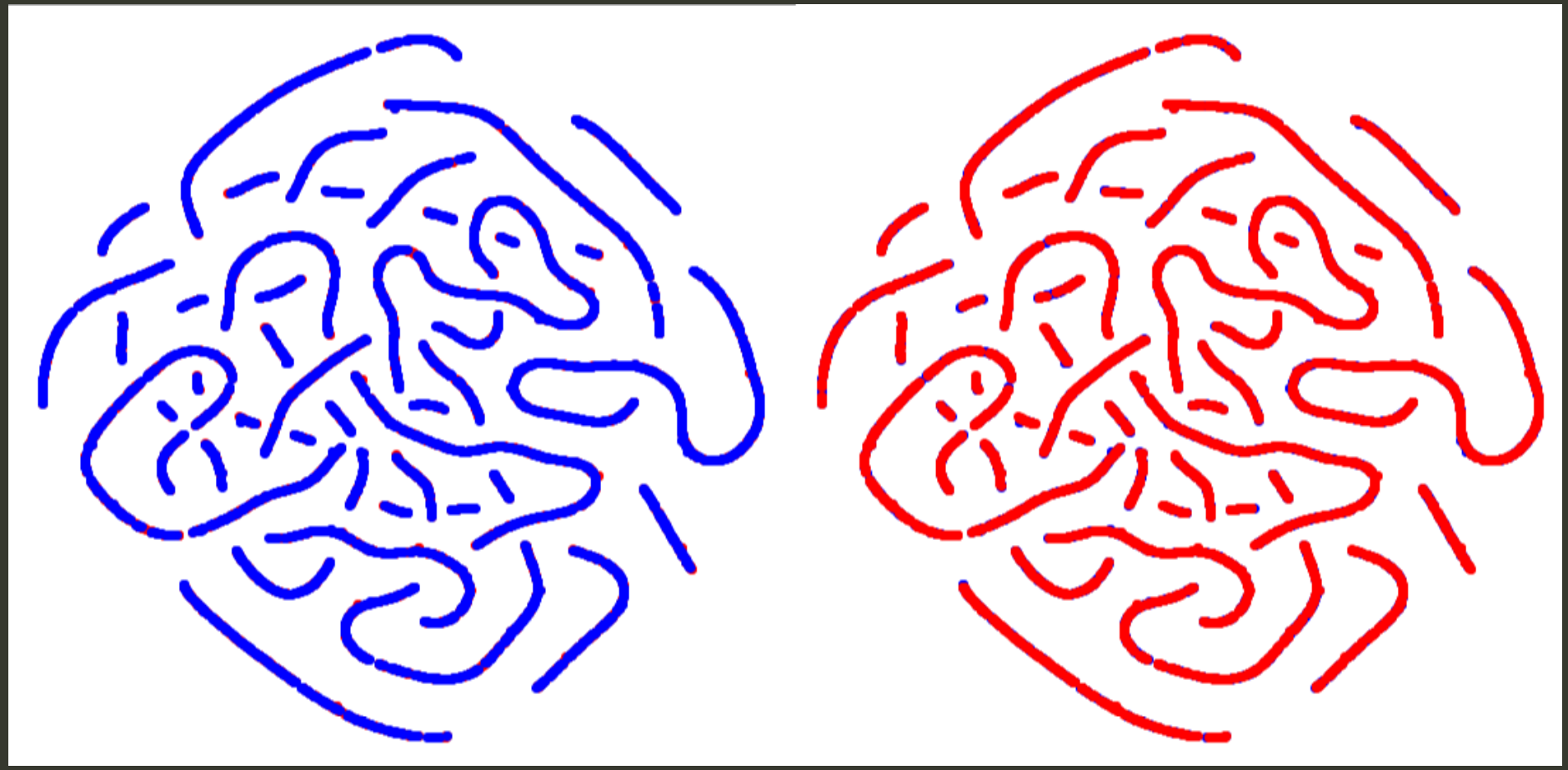


- 藍色: Source
- 紅色: Target

GAN too Weak



GAN too Strong

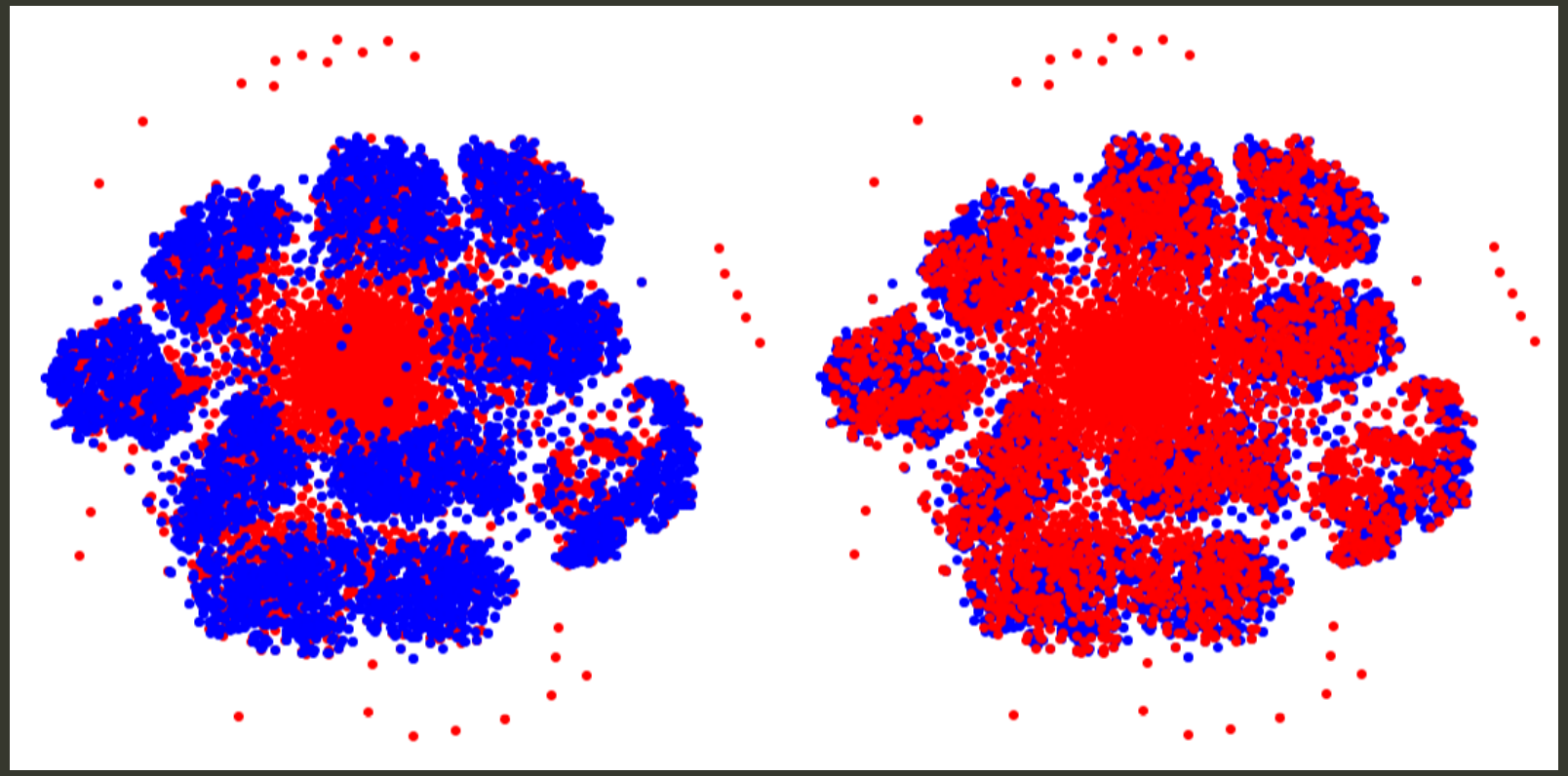


總之我就是訓練不起來啦

- 參考網路上的做法

1. WGAN 改成用 MMGAN (最原始的GAN)
2. RMSProp(1e-4) 改成 Adam(1e-3)
3. 使用網路上的一個更簡單的架構 [github]
4. 改成用 MMGAN 後, 去掉 BN layer 就能訓練起來

tSNE After Domain Adaptation



```
Epoch 100, Train Loss: 0.007638361304998398, Train Acc: 99.94830322265625%
        Test Loss: 0.3424606919288635, Test Acc: 98.93000030517578%
        Target Domain Test Loss: 2.7718491554260254, Test Acc: 83.6300048828125%
```

結論

- 還有另一篇叫 ADDA (Adversarial Discriminative Domain Adaptation), 基本也是上述的 GAN 架構
- 都已經是三、四年前的文章了, 近期有無更好的方法要再看看
- GAN 真心難訓練
- 歡迎給任何建議