

Individual Studies in Medical Image Analysis

Domain Adaptation on Stroke Outcome Prediction

Advisor: Dr. Wei-Yang Lin

第二組

Name:	ID number:
Jia-Cheng, Syu	611410013
Yu-Ting, Liu	611410083
Wei-Chieh, Tseng	611410146
Cheng-You, Gao	611410166

Outline

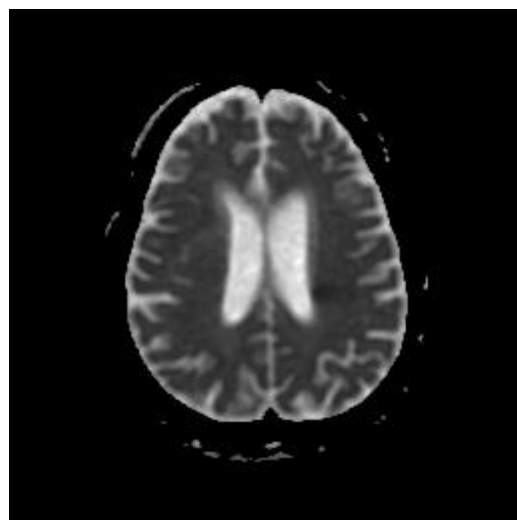
- Introduction
- Proposed Method
- Experiment Result
- Conclusion

Outline

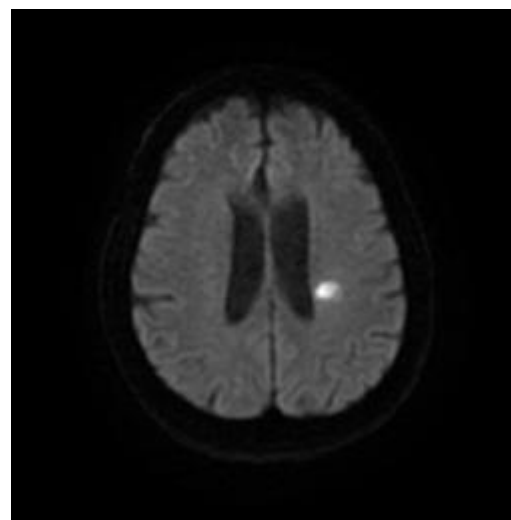
- Introduction
- Proposed Method
- Experiment Result
- Conclusion

Introduction

- 使用 Domain Adaptation 來預測中風患者三個月後的失能程度，Source Domain 是 DWI (Diffusion Weighted Imaging) 影像序列，Target Domain 是 ADC (Apparent Diffusion Coefficient) 影像序列。



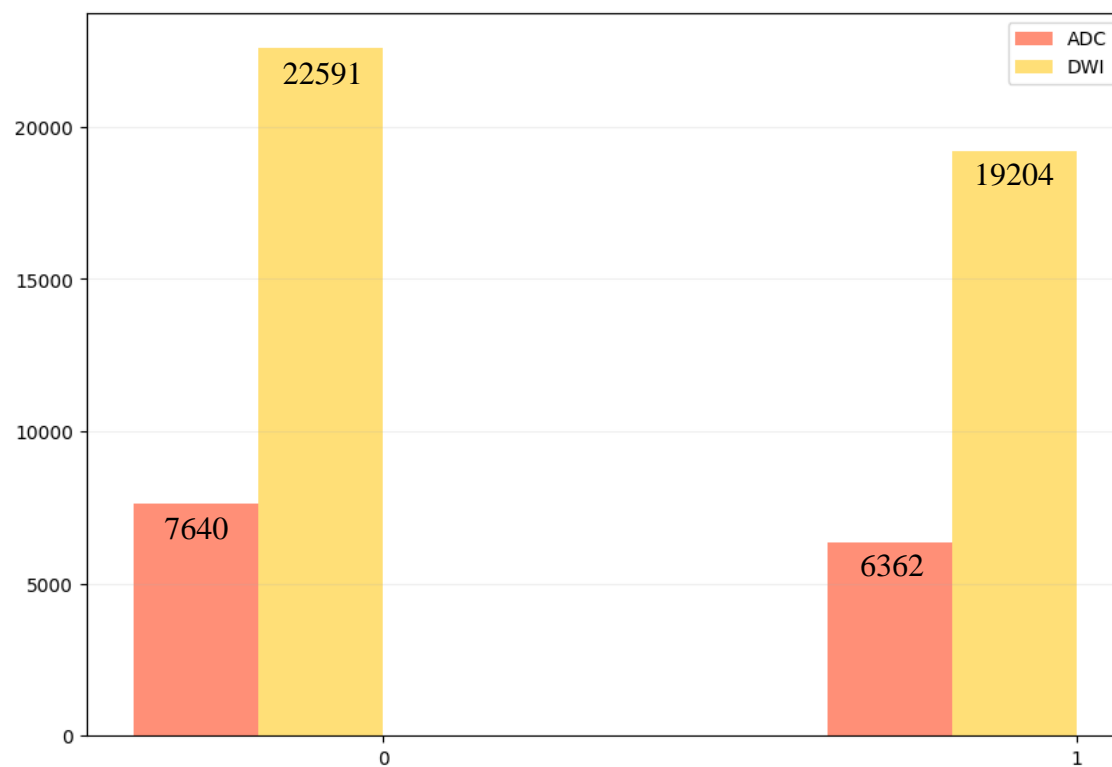
ADC 影像



DWI 影像

Introduction

- 資料集來自戴德森醫療財團法人嘉義基督教醫院
- 影像包含每個病人的 ADC 及 DWI 影像序列



Outline

- Introduction
- Proposed Method
 - DANN
 - HDA
 - GVB
- Experiment Result
- Conclusion

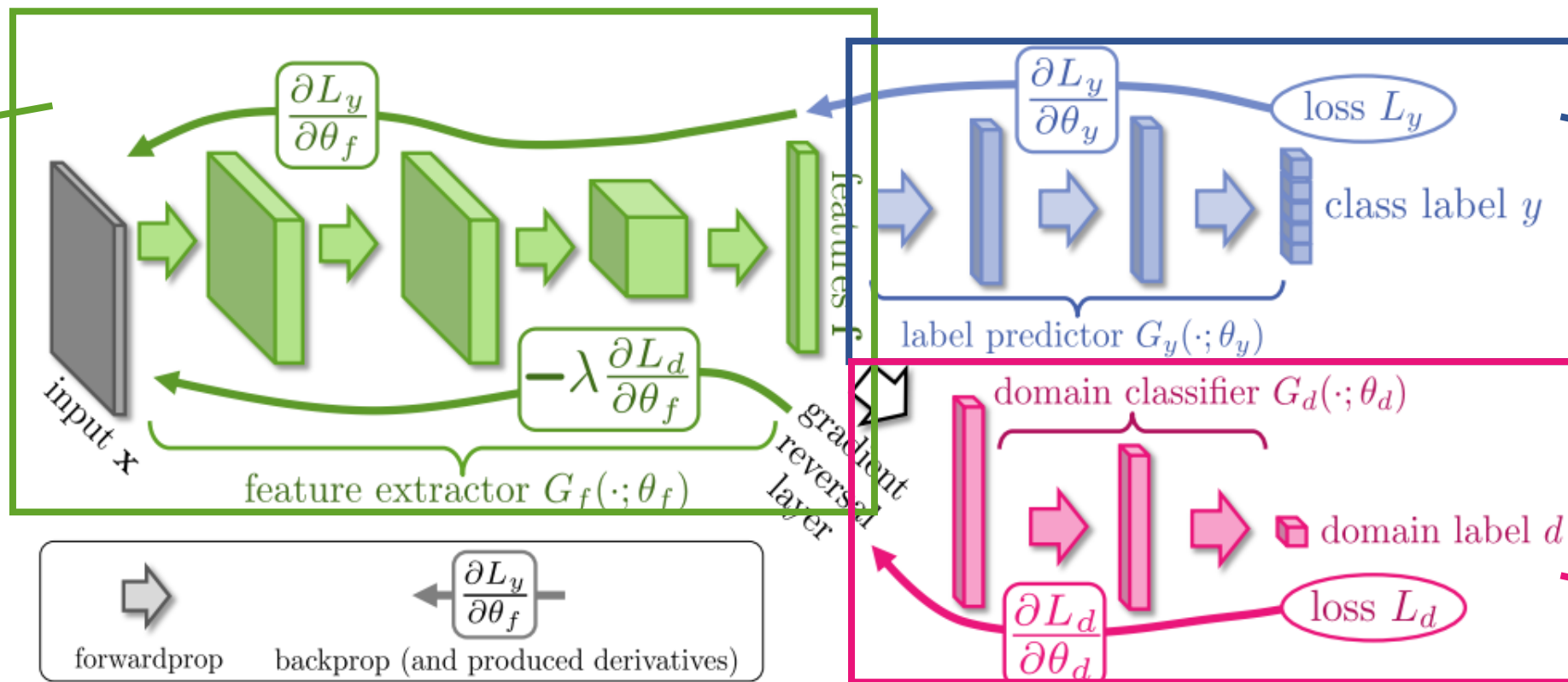
Outline

- Introduction
- Proposed Method
 - DANN
 - HDA
 - GVB
- Experiment Result
- Conclusion

DANN

Unsupervised Domain Adaptation by Backpropagation

最大化
label分類
準確率 &
最小化
domain
分類準確
率

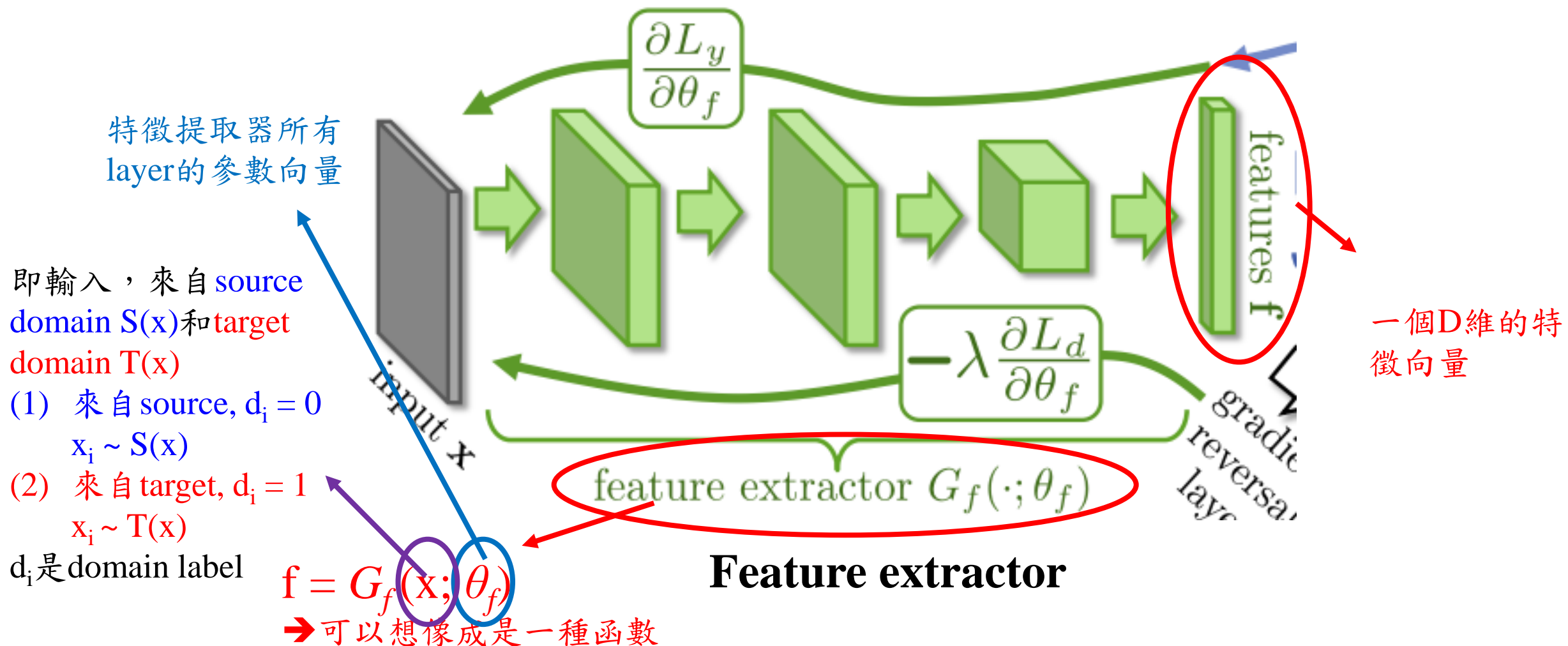


最大化
label分類
準確率

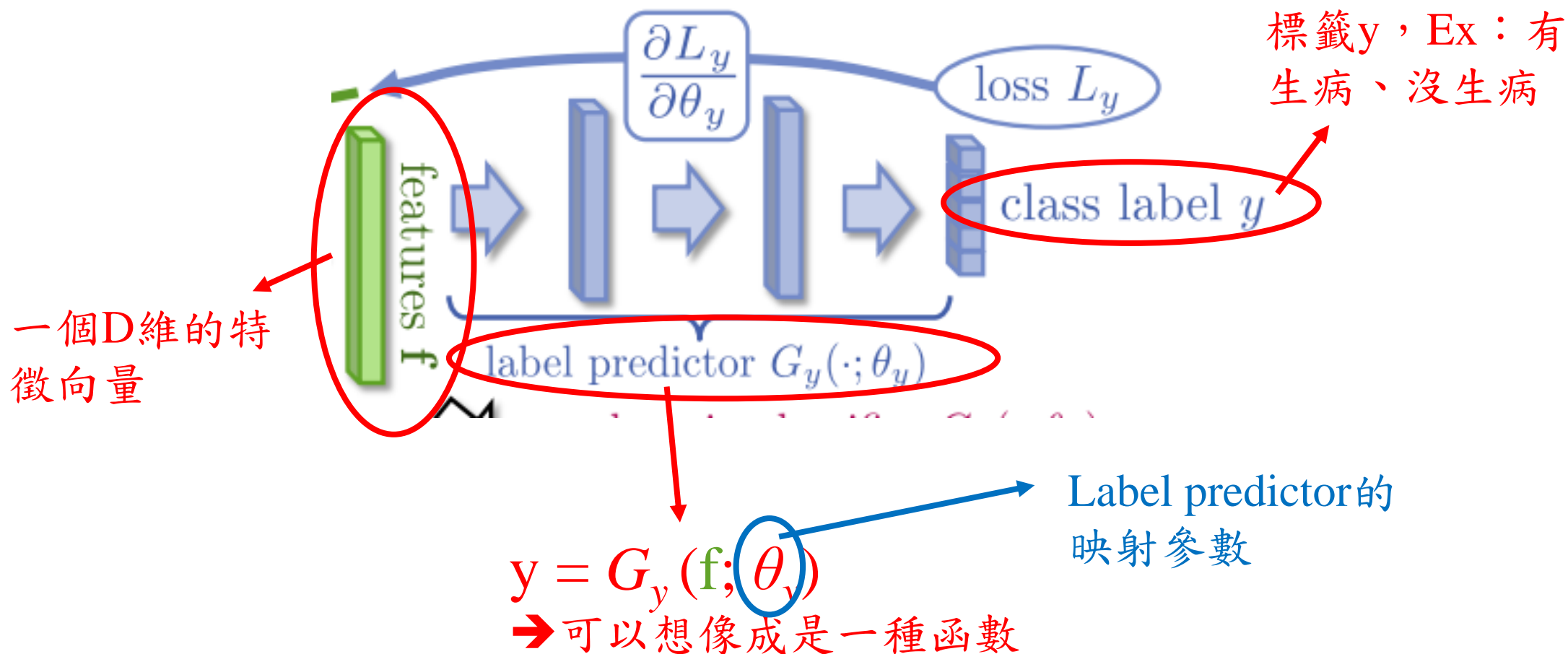
最大化
domain分
類準確率

適用於
Source data有label
Target data沒有label。

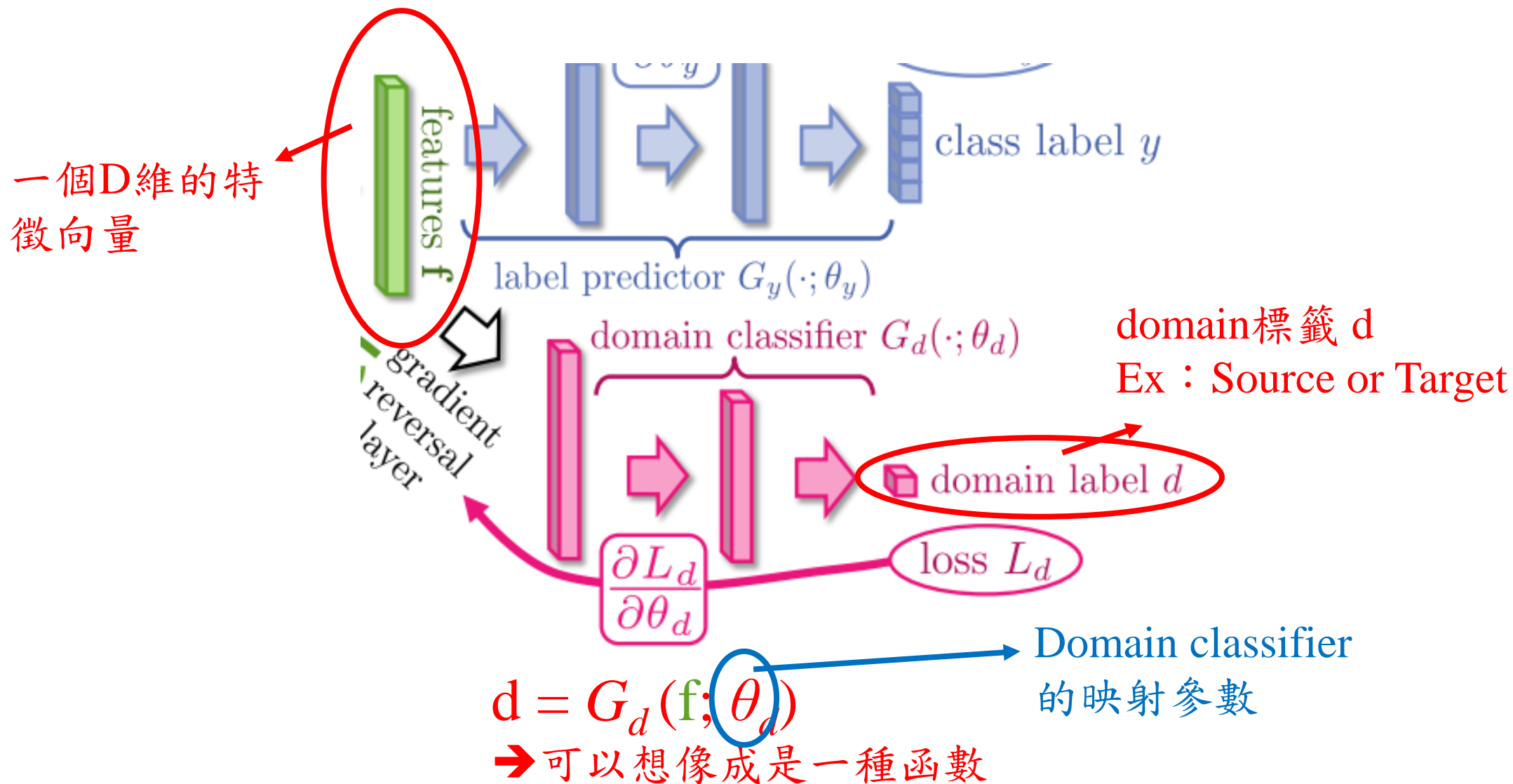
DANN



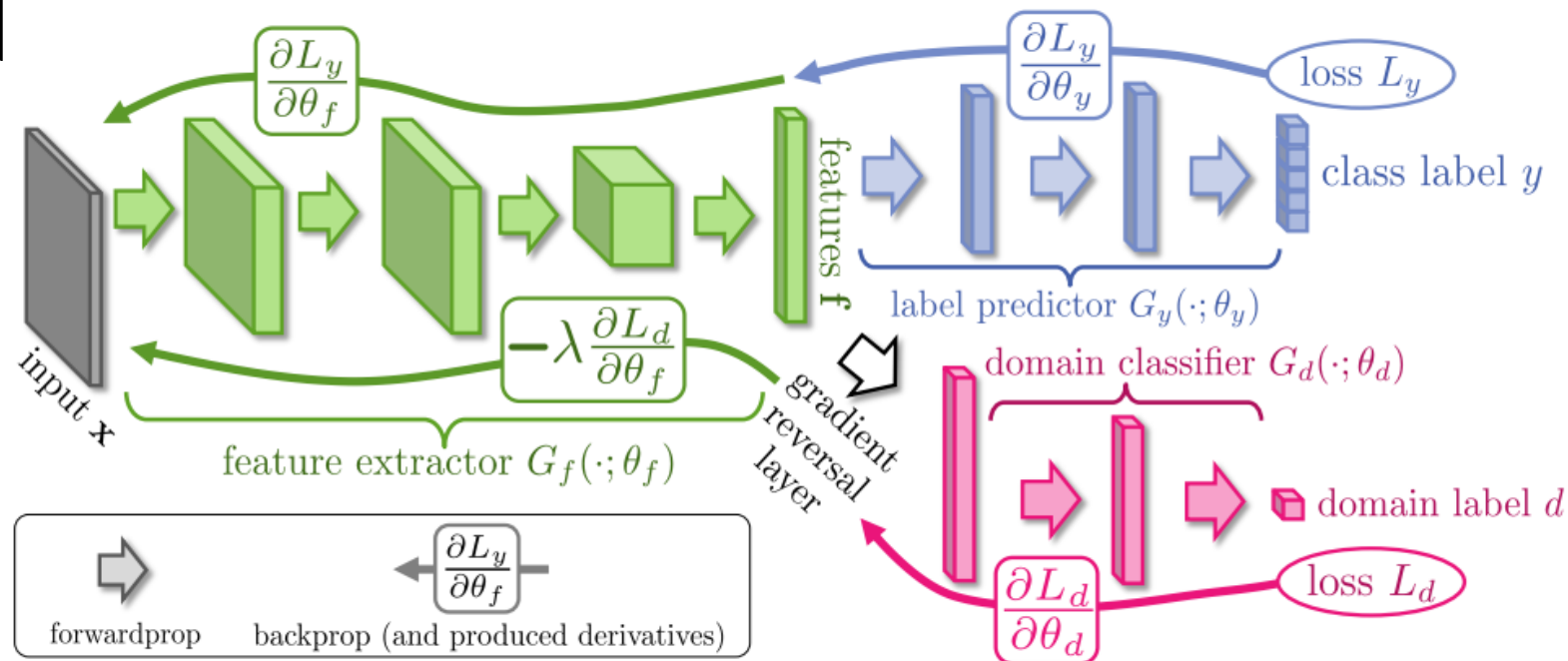
DANN



DANN



DANN



loss L_d 越大越好
 → 代表特徵 f 具有 domain-invariant

即找到的 feature 是無法被 domain classifier 正確的區分

loss L_y 越小越好
 → 特徵 f 具有識別能力

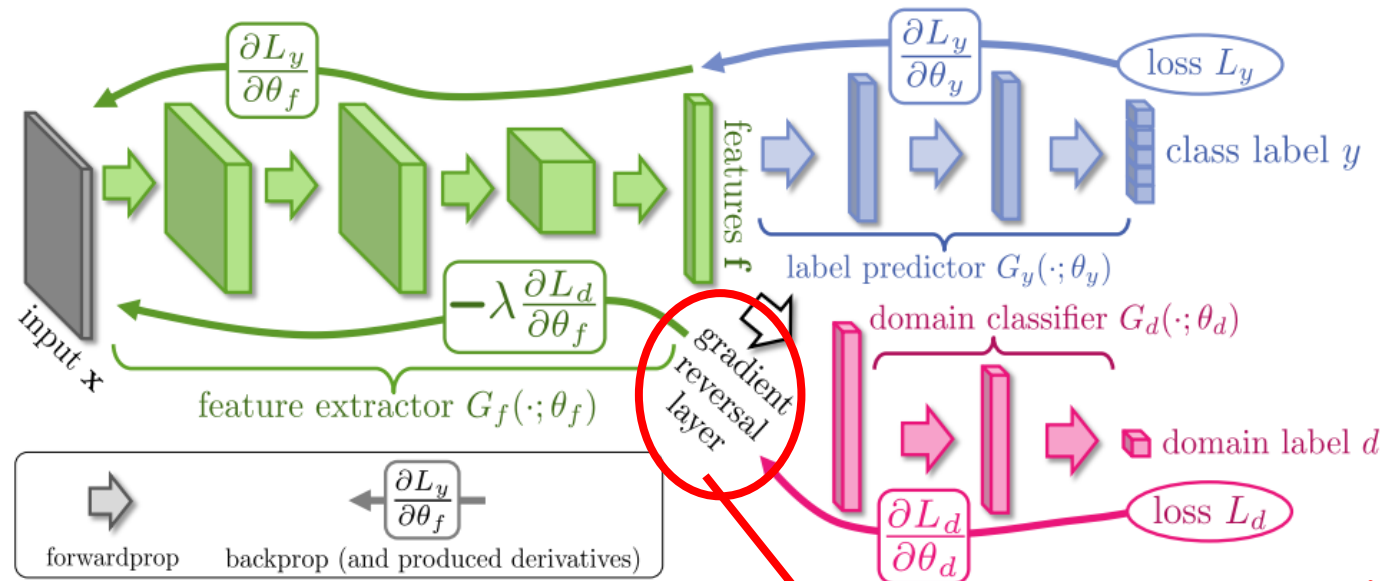
$$E(\theta_f, \theta_y, \theta_d) = \sum_{\substack{i=1..N \\ d_i=0}} L_y(G_y(G_f(\mathbf{x}_i; \theta_f); \theta_y), y_i) -$$

$$\lambda \sum_{i=1..N} L_d(G_d(G_f(\mathbf{x}_i; \theta_f); \theta_d), y_i) =$$

$$= \sum_{\substack{i=1..N \\ d_i=0}} L_y^i(\theta_f, \theta_y) - \lambda \sum_{i=1..N} L_d^i(\theta_f, \theta_d) \quad (1)$$

表示第 i 個訓練樣本對應的 Loss

DANN



Feature extractor 參數 $\theta_f \leftarrow \theta_f - \mu \left(\frac{\partial L_y^i}{\partial \theta_f} - \lambda \frac{\partial L_d^i}{\partial \theta_f} \right)$ (4)

Label predictor 參數 $\theta_y \leftarrow \theta_y - \mu \frac{\partial L_y^i}{\partial \theta_y}$ (5)

Domain classifier 參數 $\theta_d \leftarrow \theta_d - \mu \frac{\partial L_d^i}{\partial \theta_d}$ (6)

μ 為學習率

Back propagation

引入梯度反轉層(GRL)
 → 會接收後面一層的梯度，將其乘以 $-\lambda$ ，傳遞給前面一層
 λ 是一個常數，且在 Back propagation 時不會被更新

※在forward階段時，GRL會做恆等轉換

Outline

- Introduction
- Proposed Method
 - DANN
 - HDA
 - GVB
- Experiment Result
- Conclusion

HDA (Heuristic Domain Adaptation)

- HDA是為了解決在domain-invariant representations中難以分離domain-specific characteristics的問題。
- HDA想法源於Heuristic search $f(n) = g(n) + h(n)$ ，其中 $h(n) \leq h^*(n)$ 則Heuristic domain adaptation為 $G(x) = F(x) - H(x)$ 。
- 其中 $|H(x)| \leq |H^*(x)| = |F(x) - F^*(x)|$ ， $G(x)$ 為domain-invariant characteristics， $F(x)$ 為整體的domain characteristics， $H(x)$ 為domain-specific characteristics。



HDA (Heuristic Domain Adaptation)

Assumption 3.1 *In a certain domain, the difficulty of constructing domain-invariant representations is more than that of constructing domain-specific representations.*

- 在這假設下 $H(x)$ 建模更易於 $G(x)$ ，所以 $H(x)$ 可為 $G(x)$ 的建模作指導。HDA的內部機制是基於下列公式：

$$\epsilon_T(F) \leq \epsilon_S(F) + [\epsilon_S(F^*) + \epsilon_T(F^*)] + |\epsilon_S(F, F^*) - \epsilon_T(F, F^*)|$$

$$\epsilon_S(F) = \mathbb{E}_{(x,y) \sim S} |F(x) - y|$$

$$\epsilon_S(F_1, F_2) = \mathbb{E}_{(x,y) \sim S} |F_1(x) - F_2(x)|$$

$$F^* = \arg \min_F \epsilon_S(F) + \epsilon_T(F)$$

HDA (Heuristic Domain Adaptation)

- 在Heuristic search $F = G + H$ ，而G跟F的目標相同，因此 $F^* = G^*$ ，所以在建模H時是針對S，T上的domain-specific進行建模，因此也可通過F的domain-specific部分，但不通過F*的部分進行建模，就得到以下公式：

$$H = k(F - F^*) \quad k \in (0, 1]$$

- k是positive correlation coefficient，由上述公式能推得以下公式：

$$(1 - k)(F - F^*) = (G - G^*)$$

HDA (Heuristic Domain Adaptation)

$$\epsilon_T(F) \leq \epsilon_S(F) + [\epsilon_S(F^*) + \epsilon_T(F^*)] + |\epsilon_S(F, F^*) - \epsilon_T(F, F^*)|$$

$$(1 - k)(F - F^*) = (G - G^*) \quad H = k(F - F^*) \quad k \in (0, 1]$$

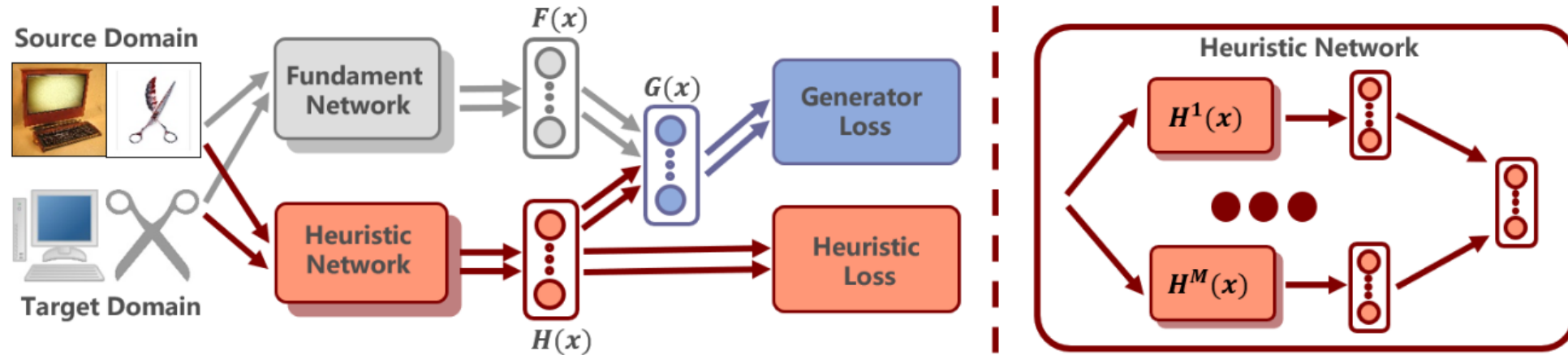
- 此方程式可由S、T實現，則G跟F的risk就是

$\epsilon(G, G^*) = (1 - k)\epsilon(F^* - F)$ 其中G跟F都能正確分類source sample，
所以他們的risk一樣 $\epsilon_S(G) = \epsilon_S(F)$ ，因此就能計算target risk

$$\begin{aligned} \epsilon_T(G) &\leq \epsilon_S(G) + [\epsilon_S(G^*) + \epsilon_T(G^*)] + |\epsilon_S(G, G^*) - \epsilon_T(G, G^*)| \\ &\leq \epsilon_S(F) + [\epsilon_S(F^*) + \epsilon_T(F^*)] + (1 - k) |\epsilon_S(F, F^*) - \epsilon_T(F, F^*)| \end{aligned}$$

這樣就能將G的risk限制在一個較低的值上。

HDAN (Heuristic Domain Adaptation Network)



Heuristic Constraint

1. Similarity
$$\cos(\theta) = \frac{G(x) \cdot H(x)}{|G(x)| |H(x)|}$$

2. Independence
$$\text{kurt}(y) = \mathbb{E}_{y \in \mathcal{D}} [N(y)^4] - 3 \{ \mathbb{E}_{y \in \mathcal{D}} [N(y)^2] \}^2$$

3. Termination Condition
$$|H(x)|_1 \approx 0$$

Outline

- Introduction
- Proposed Method
 - DANN
 - HDA
 - GVB
- Experiment Result
- Conclusion

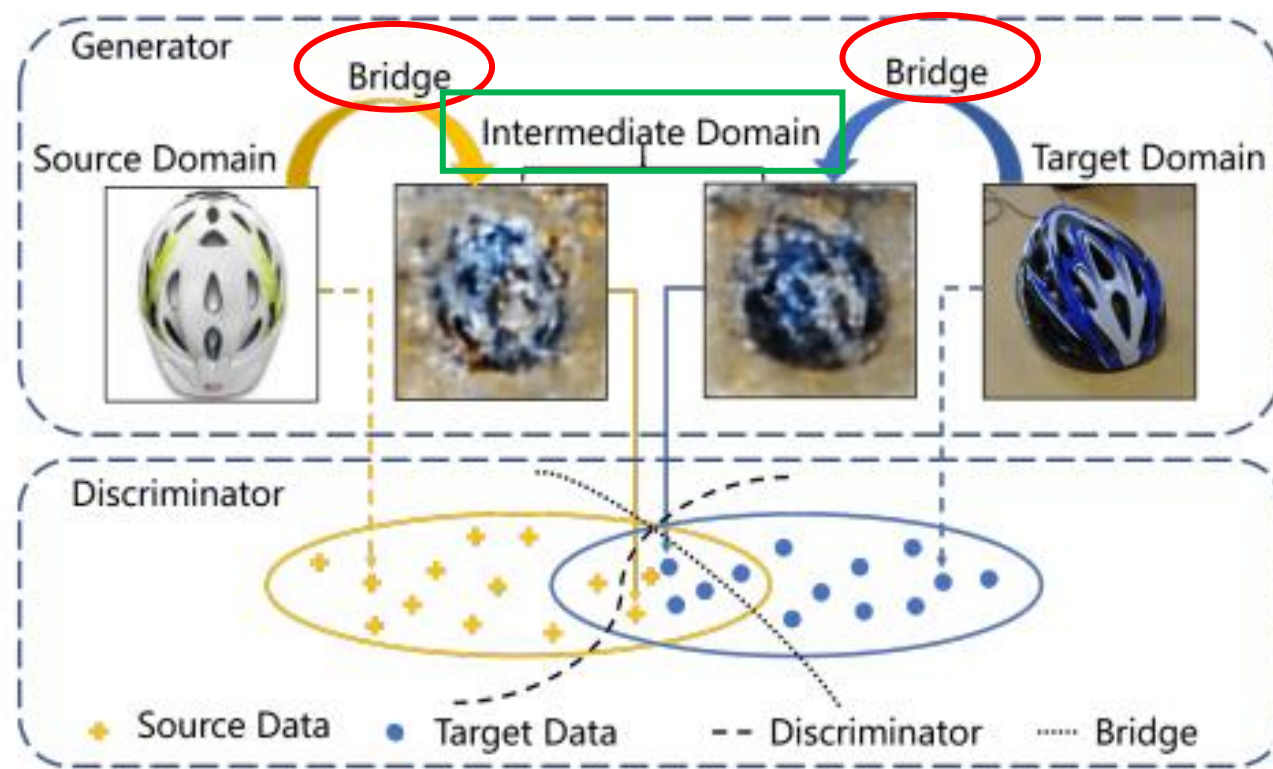
GVB (Gradually Vanishing Bridge)

解決問題：

無監督領域適應中的轉移困難和領域特定特徵殘留的問題，Source Domain和Target Domain之間的差異很大，直接進行轉移非常困難。

目的：

將Source Domain和Target Domain的樣本特徵盡量接近，使模型更具泛化(減少特定特徵的影響)，學習讓兩個領域樣本修正到中間領域裡，使大部分更加對齊，**減少轉移困難**。



GVB

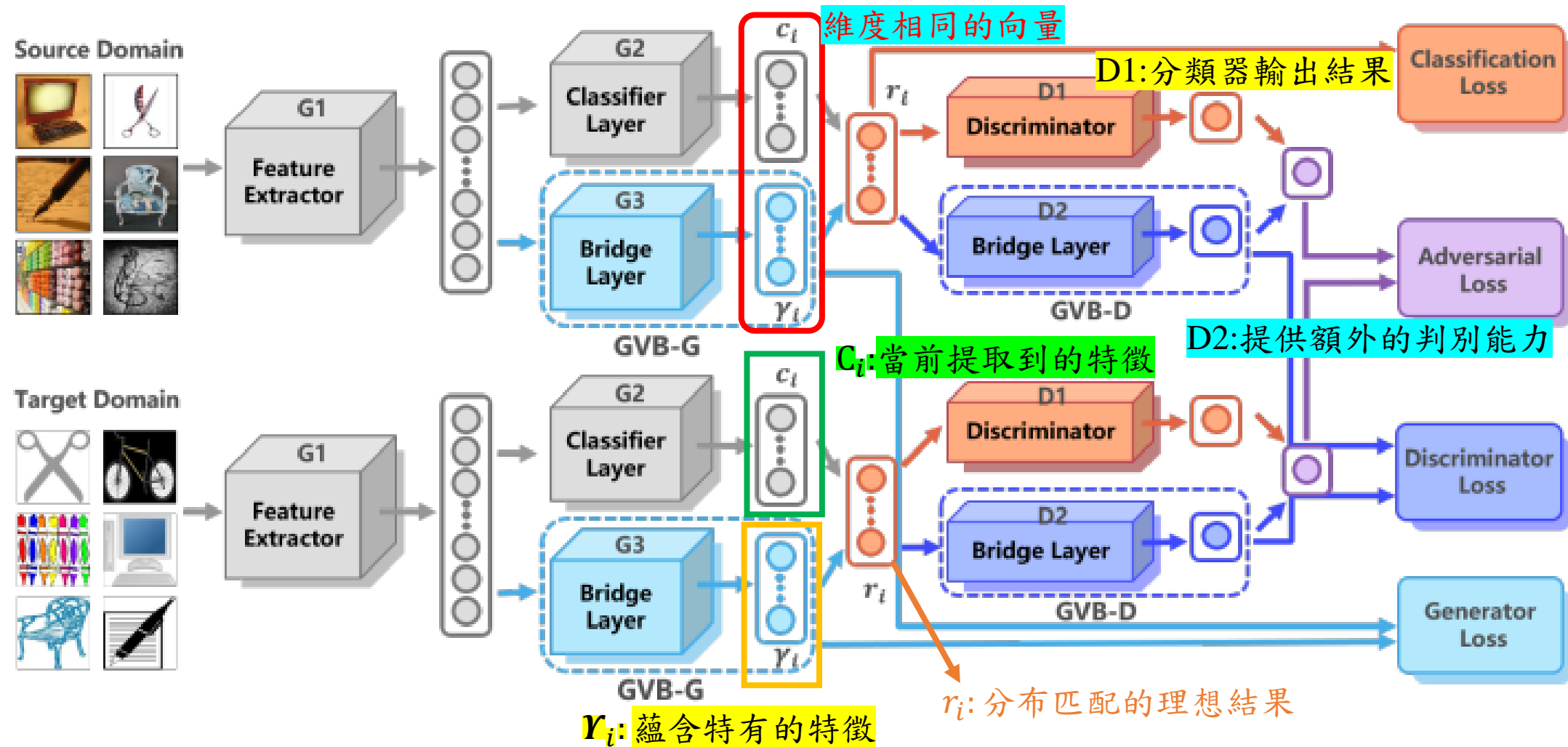
G1: 特徵提取模型

G2: 分類層

D1: 分類器

此架構新增：

G3、D2: 彌補領域差異的橋樑



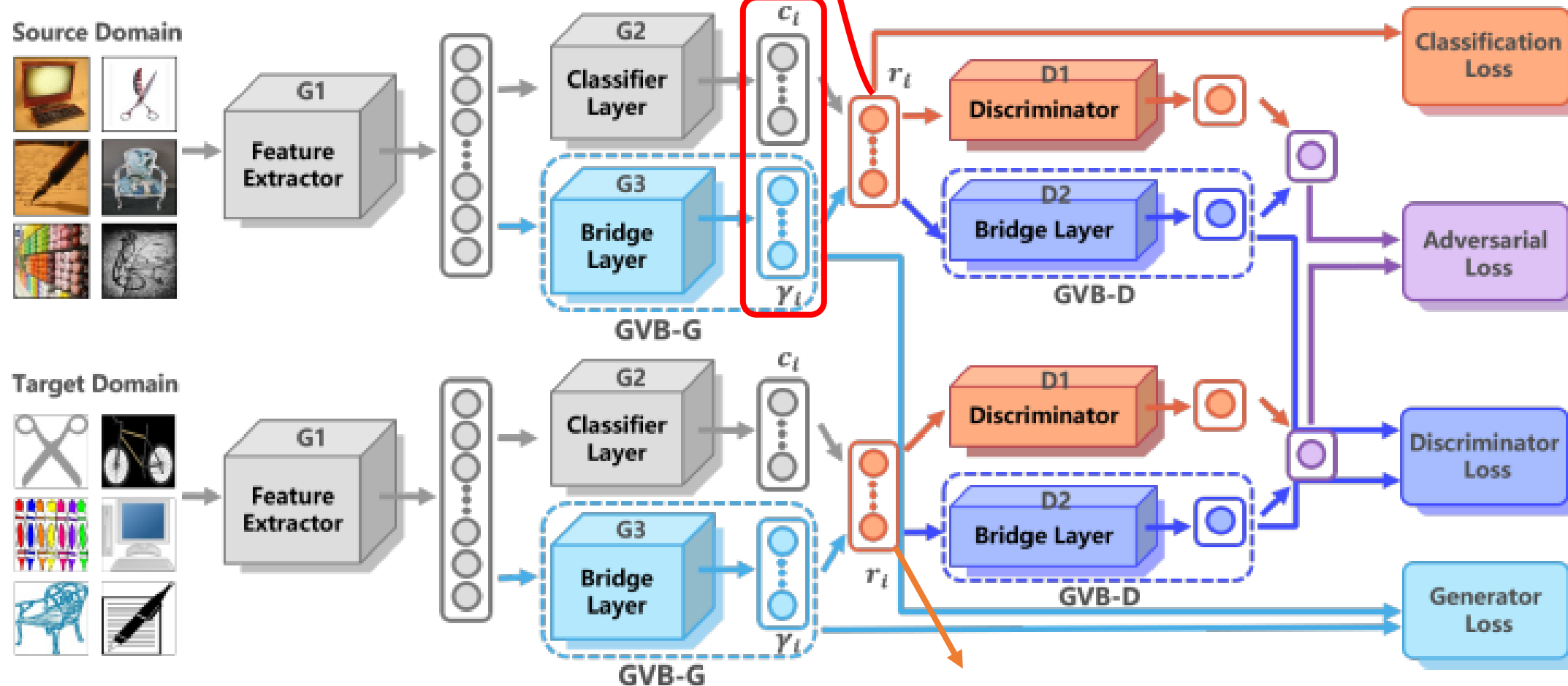
Source Domain & Target Domain 處理過程相同

GVB

domain-invariant representation: $r_i = c_i - \gamma_i$

生成器的輸出 G^* $G_*(x_i) = G_2(G_1(x_i)) - G_3(G_1(x_i))$.

在不同領域之間的分布盡量相似



Source Domain & Target Domain 處理過程相同

GVB

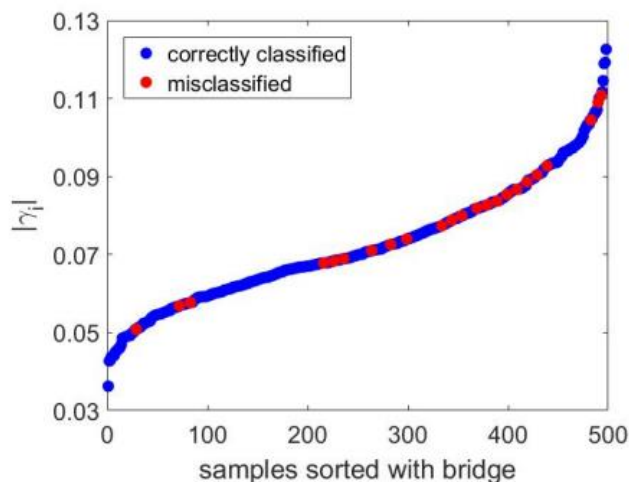
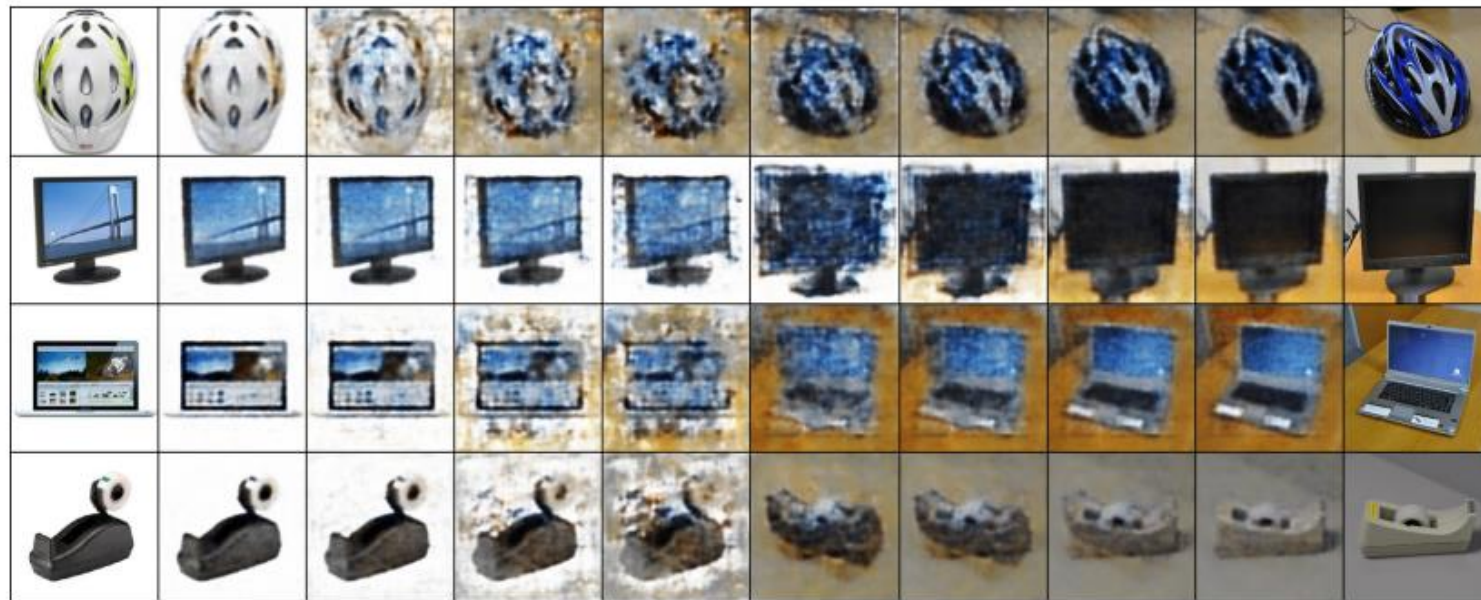
如果 γ_i 的範圍過大，可能有2個原因

1. 數據含有特殊屬性的困難樣本
2. 具有較豐富的特徵(細節較多)

透過逐漸減少 γ_i 的整體範圍來減少 γ_i 影響，抑制了特徵，而對 c_i 、 r_i 產生較小的負面影響。

$$\mathcal{L}_G = \frac{1}{(N_s + N_t) * C} \sum_{i=1}^{N_s + N_t} \sum_{j=1}^C |\gamma_{i,j}|,$$

應用於生成器的GVB
稱為GVB-G



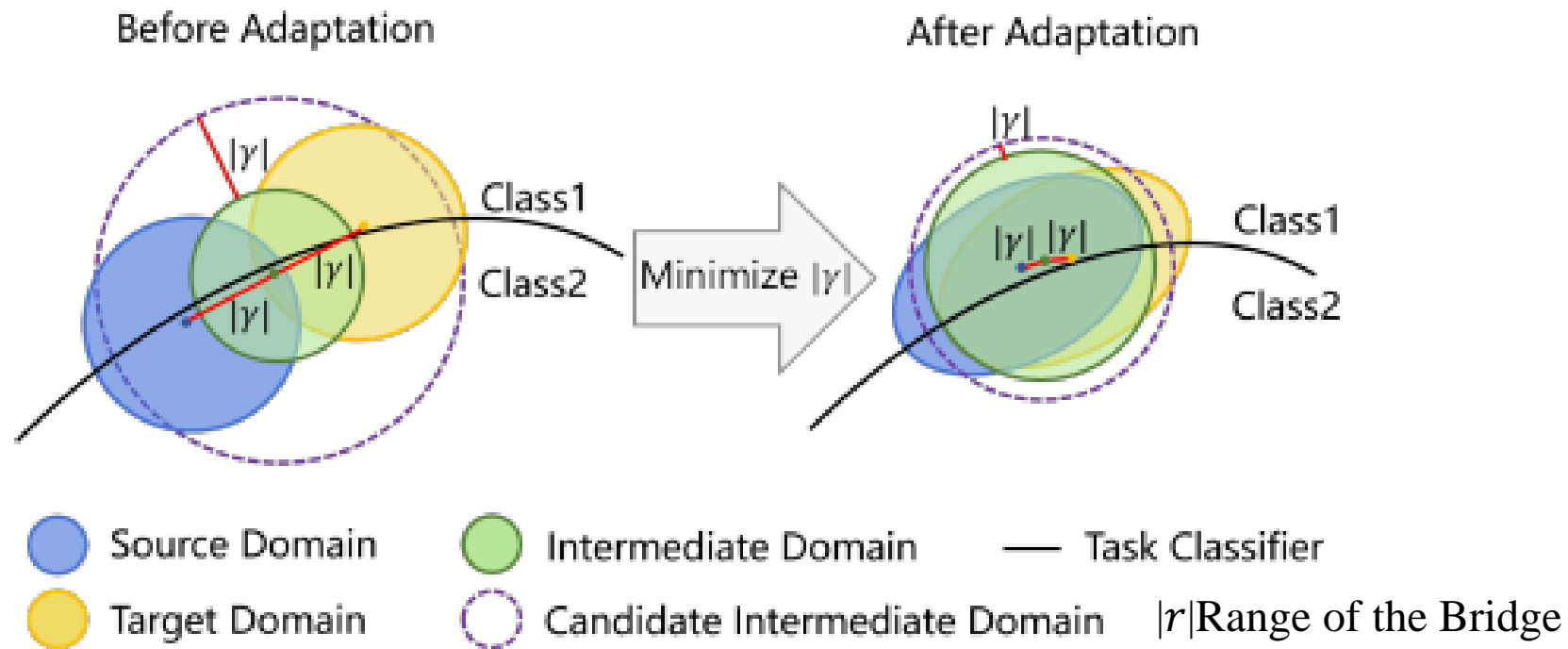
(a) GVB-G

若Bridge的範圍大代表 γ_i 具有較豐富的特徵，因此會影響 c_i 、 r_i ，導致有較高的錯誤分類率。

GVB

Gradually Vanishing Bridge (GVB)

- 逐漸縮小範圍
- 之間的差異很難最小化到零



GVB

問題：計算成本大，容易陷入局部最小化。

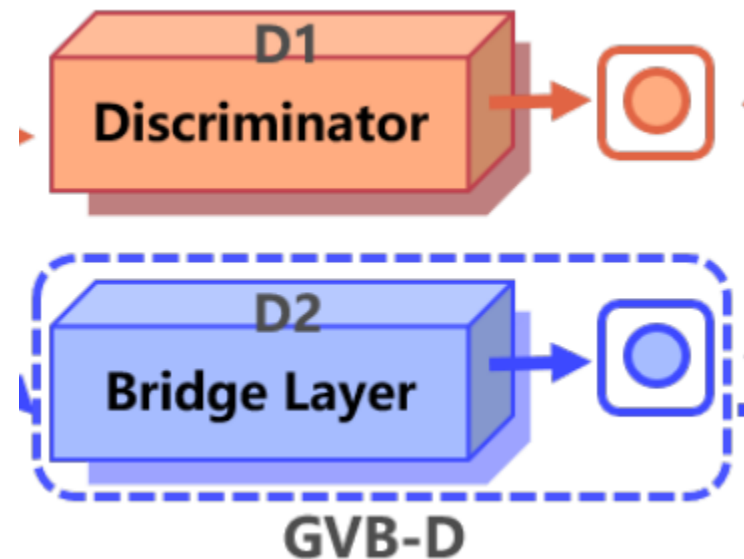
解決方式：新增了D2，主要是模擬D1函數，為D1提供額外的分類能力。

$$D_{\star}(G_{\star}(x_i)) = D_1(G_{\star}(x_i)) + D_2(G_{\star}(x_i)).$$

逐漸減少分類器在Bridge layer的範圍

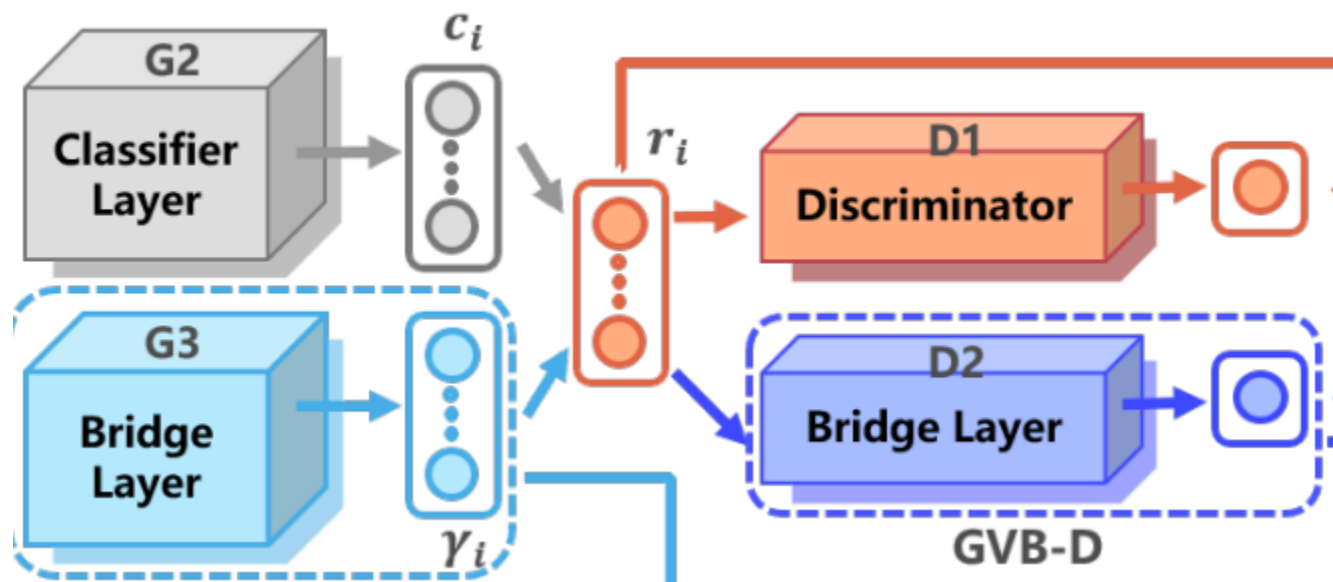
$$\mathcal{L}_D = \frac{1}{(N_s + N_t)} \sum_{i=1}^{N_s + N_t} |\sigma_i|,$$

σ_i : 表示建構的分類器Bridge layer



應用於分類器GVB
稱為 GVB-D

GVB



結合GVB-G和GVB-D
形成生成器和分類器的GVB-GD

$$\begin{cases} G_{\star}(x_i) = G_2(G_1(x_i)) - G_3(G_1(x_i)) \\ D_{\star}(G_{\star}) = D_1(G_{\star}) - D_2(G_{\star}) \\ \mathcal{L}_{ext} = \lambda \mathcal{L}_G + \mu \mathcal{L}_D, \end{cases}$$

使用超參數 λ 和 μ 來調整優化

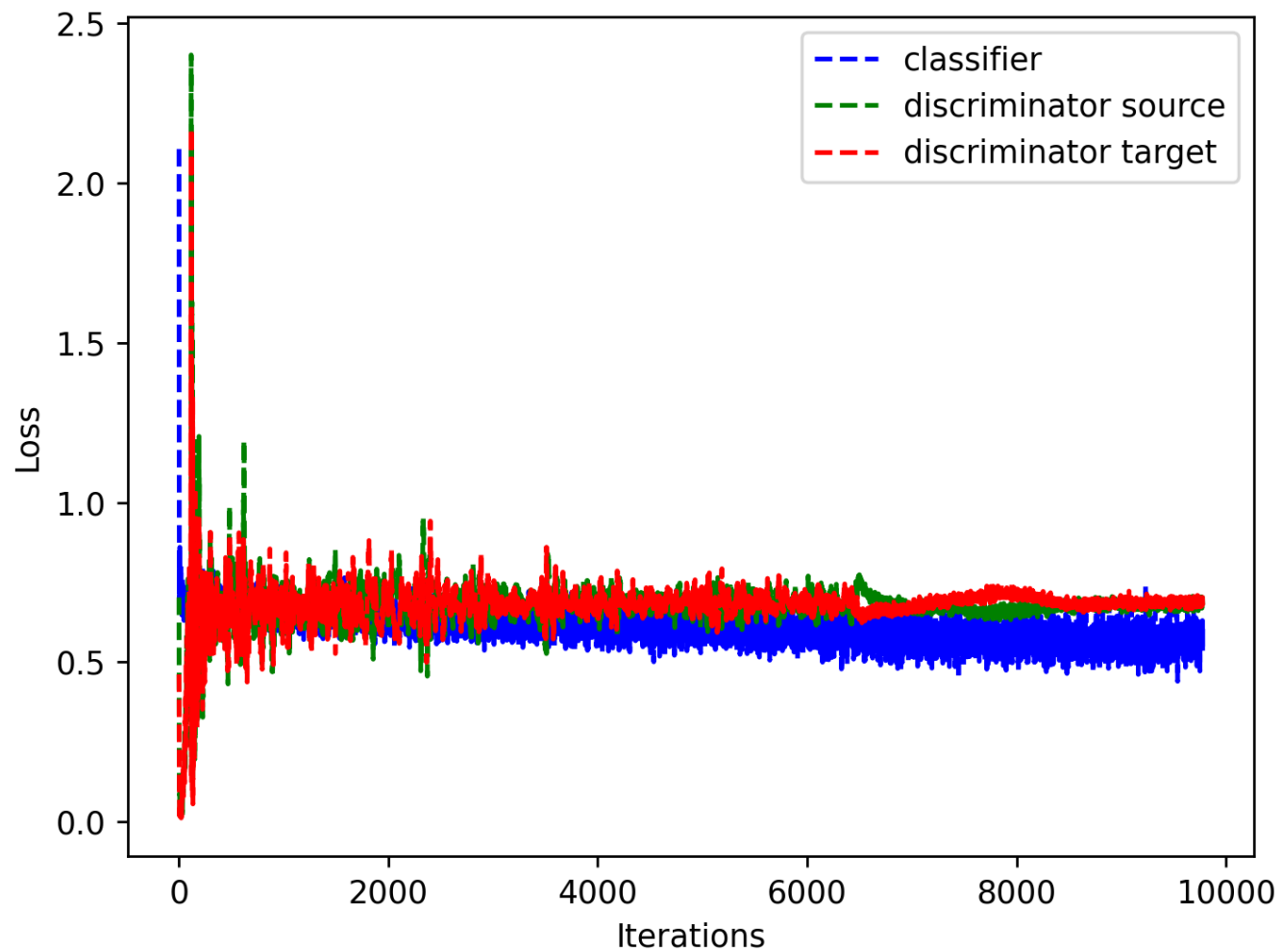
Outline

- Introduction
- Proposed Method
- **Experiment Result**
- Conclusion

Experiment Result

Source Domain	Target Domain	Model	Accuracy	Increment	AUC
ADC	ADC	ResNet50	70%	-	0.68
DWI	ADC	ResNet50	45%	-	0.42
DWI	ADC	DANN	56.2%	+11.2%	0.58
DWI	ADC	HDA	53.5%	+8.5%	0.55
DWI	ADC	GVB	55.1%	+10.1%	0.52

Experiment Result: DANN



Outline

- Introduction
- Proposed Method
- Experiment Result
- Conclusion

Conclusion

透過DANN、HDA和GVB等Transfer Learning的架構實作Domain Adaptation，在準確率的部分，DANN達到56.2%、HDA達到53.5%、GVB達到55.1%，皆優於ResNet50模型的45%。儘管如此，改善效果並不明顯。為了進一步提高準確率，我們將嘗試調整參數或嘗試其他模型架構，以取得更好的結果。