

LogoGANs：基於生成對抗網路之複數標誌組合與生成

潘品齊

電機工程學系

國立中正大學

嘉義，臺灣

pinchi1609@gmail.com

劉立頌

電機工程學系

國立中正大學

嘉義，臺灣

aliu@ee.ccu.edu.tw

Abstract—對於 Logo 標誌設計師來說，根據使用者所需設計具有特色之標誌是一個漫長且困難的過程，隨著生成對抗網路在圖像的任務中快速發展，越來越多相關的研究投入到圖像生成中，其中對於 Logo 標誌的生成研究也不在少數，Logo 標誌合成具有多模態的圖像領域，至今仍然是生成對抗網路的一大挑戰。隨著電腦科技與深度學習技術的蓬勃發展，現在已有許多不同架構之生成對抗網路可用於 Logo 標誌的生成，本研究重點在於改善組合生成對抗網路的模型表現並結合空間變換網路，為了解決生成器、判別器及空間變換網路三者之訓練間模式崩潰以及輸出圖像層次與顏色遮擋的問題，我們也提出新的考試機制訓練方式，讓模型能夠將複數圖像組合出全新的 Logo 標誌，幫助設計師在創作時提供更多靈感。

Keywords—生成對抗網路、組合生成對抗網路、空間變換網路、圖像組合、圖像生成、圖像上色

I. 緒論

多模態圖像合成一直是生成對抗網路 (Generative Adversarial Network, GAN) [1] 中的一大挑戰，由於 Logo 具有多模態、難以分類且沒有連續潛在空間的特性，Ajkil Mino 等人於 2018 年提出 LoGAN [2] 架構，利用輔助條件生成對抗網路 (Auxiliary Classifier GAN, ACGAN) 並引入 WGAN-GP 的訓練模式，將該網路用於生成幫助設計師進行創作過程的 Logo，該網路能夠生成以十二種不同顏色為條件的 Logo。而 Cedric Oeldorf 等人於 2019 年提出 LoGANv2 [3] 架構，該網路將 ProGAN [4] 以及 StyleGAN [5] 架構應用於 Logo 生成上，使其能夠生成更高質量及多樣化的輸出。

在過去的 Logo 生成研究方向大多是以 Logo 的顏色或是提高 Logo 圖像的解析度為目標進行研究，但使用者所需要的標誌往往具有特定圖案，例如寵物店的標誌會需要具有動物類型的圖案、餐廳的標誌會需要具有餐具或是食物等圖案組成，若是僅使用顏色來生成或分類 Logo，對於 Logo 設計師在設計相關標誌時，能提供的幫助仍然有限。而為了解決特定圖像 Logo 標誌生成的問題，本研究試圖改善現有的組合生成對抗網路 (Compositional GAN) [6] 架構，以解決特定圖像下標誌組合間的問題，使網路能夠在非監督且非成對的數據集下進行訓練，結合多個具有特定圖像之標誌，並使其產生具有 Logo 風格之圖案。

在圖像組合的任務中，若將來自不同圖像的對象疊加在一個圖像中，這些對象之間的外觀和幾何差異是導致合成圖像不夠真實的一大問題。Chen-Hsuan Lin 等人於 2018 提出的空間變換生成對抗網路 (ST-GAN) [7] 使用空間變換器網路作為生成器，通過執行幾何校正來扭曲對象以適應固定的背

景圖像。在本研究中將會探討如何改善現有的組合生成對抗網路模型，並介紹空間變換網路在其中的訓練方式，基於這些研究的改善層面，讓網路能夠在非監督與非成對的數據集上進行訓練，並應用於組合 Logo 標誌當中；現有的圖像風格遷移研究大多是對整張圖片進行風格遷移，在本研究中將介紹如何將組合 Logo 標誌進行著色，使其具有一般品牌標誌之風格。接下來第二章將介紹使用空間變換網路以及組合生成對抗網路的相關研究，第三章會描述我們的實驗架構模型與改善方法，在第四章將介紹實驗的訓練流程、實驗結果及結果探討，並在最後第五章進行總結及本研究的未來展望。

II. 背景知識與文獻探討

生成對抗網路已被廣泛用於各種圖像生成應用，其網路中的損失函數已被證明在優化基於可用信息的高質量圖像方面是有效的，本研究重點為複數標誌組合之研究，所以本章節將依序介紹空間變換網路及生成對抗網路在先前研究中的成果，詳細敘述如下：

A. 空間變換網路 (Spatial Transformer Networks)

卷積神經網路已經是一個能力強大的分類模型，但是它仍然受限於計算數據在空間上多樣性的影響。Max Jaderberg 等人於 2015 年採用一種新的可學習模型—空間變換網路 (Spatial Transformer Networks, STNs) 來解決此問題，如圖 1 所示，該網路模型由三個部分組成：localisation net、grid generator 及 sampler。空間變換網路可用於輸入層，也可插入到卷積層或者其他層的後面，不需要改變原來卷積神經網路模型的內部結構。空間變換網路能夠自適應地對數據進行空間變換和對齊，使得卷積神經網路模型對平移、縮放、旋轉或者其他變換等保持不變性。此外，空間變換網路的計算速度很快，幾乎不會影響原有卷積神經網路模型的訓練速度。

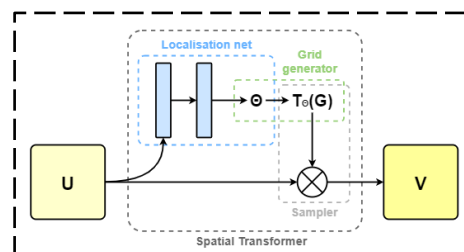


圖 1、空間變換器網路架構

B. 組合生成對抗網路 (Compositional GAN)

生成對抗網路可以生成非常複雜和逼真的圖像，但通常只能從單張圖像中採樣，而忽略了場景中可能存在多個

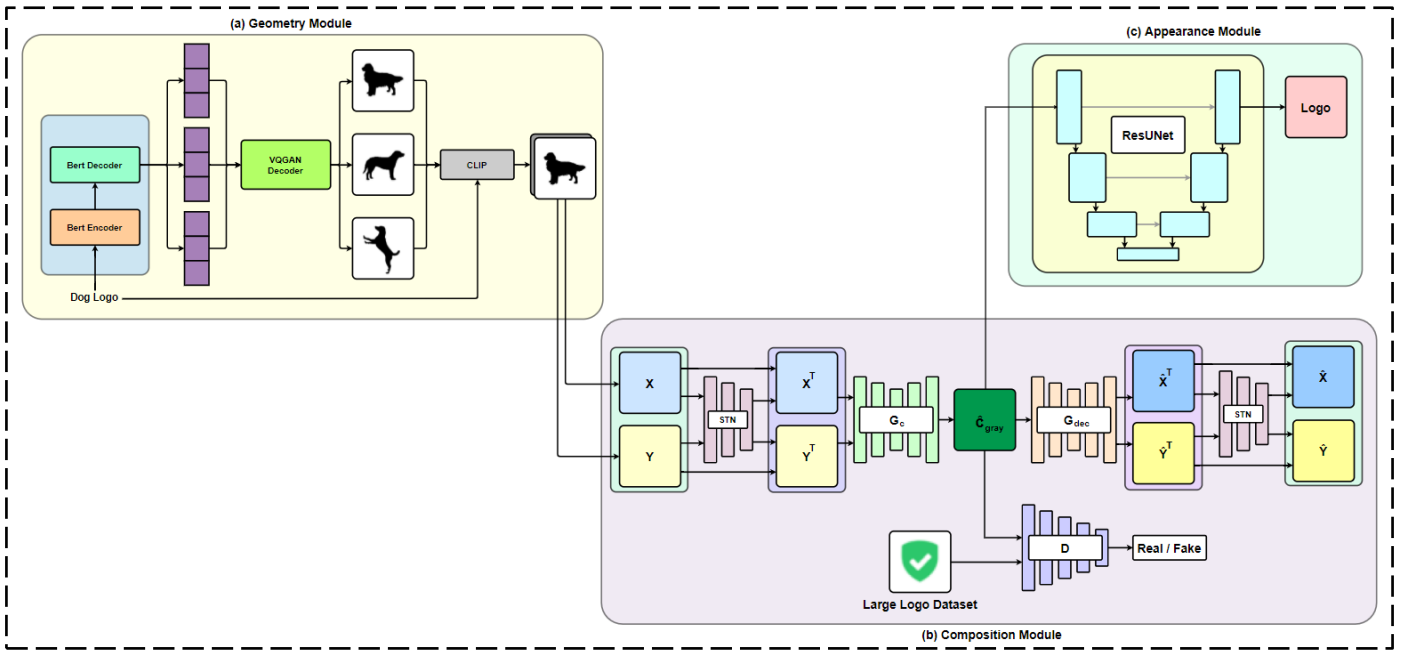


圖2、LogoGANs 模型架構

物件之間的空間交互狀態。若是想要獲得圖片中不同對象之間複雜的空間交互狀態，包括它們的相對縮放、空間佈局、遮擋或視點變換，是一個具有挑戰性的問題。為了解決圖片合成上空間交互關係的問題，組合生成對抗網路 (Compositional GAN) [6] 提出了一種組合分解架構網路 Composition-by-Decomposition (CoDe) 來組合圖片。給定來自兩個不同分佈的對象圖像，Compositional GAN 模型可以根據輸入對象的紋理和形狀從它們的聯合分佈中生成逼真的合成圖像。圖3顯示 Compositional GAN 中 Composition-by-Decomposition 的網路架構，其輸入為兩個不同的圖像，該方法的中心思想是通過 self-consistency loss function 監督兩個圖像 (X, Y) 的合成，確保生成的合成圖像可以進一步分解回各自的單獨對象圖像。合成是使用 Conditional GAN (G_c, D_c) 執行，該 GAN 將兩個通道的 RGB 圖像 (X, Y) 作為輸入，以生成相應的合成輸出，並適當地合成兩個輸入圖像。然後將生成的圖像輸入另一個 Conditional GAN (G_{dec}, D_{dec})，使用 L1 損失函數將其分解回其組成對象 (\hat{X}, \hat{Y}) 。

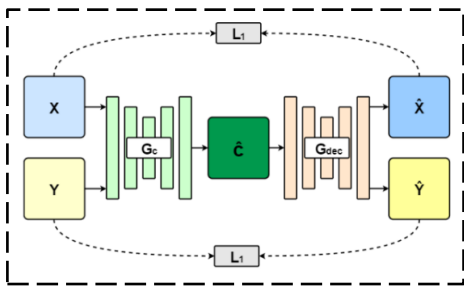


圖3、Composition-by-Decomposition 網路架構

由於 Conditional GAN 不足以在空間上變換物體，該網路需要基於相對空間變壓器網路 (Spatial Transformer Networks, STN) [8] 通過縮放與平移物體圖像將 (X, Y) 轉換為 (X^T, Y^T) 。空間變壓器網路給定 X, Y, C 圖像中物體的遮罩，將所有輸入對象裁剪並縮放到訓練圖像的中心。為了適當的空間佈局，該網路所訓練的是空間變換器網路的變

形體。該空間變換器網路同時採用 multi RGB 圖像按通道連接，並將它們相對於彼此轉換為 (X^T, Y^T) 基於它們在訓練合成圖像中編碼的空間關係，圖4為 Compositional GAN 所使用的空間變換器網路架構。

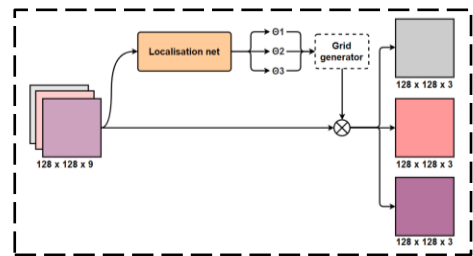


圖4、空間變換器網路架構

III. 研究方法

在第二章所介紹的模型中，以 Compositional GAN 問題進行改善，獲得不同改善模型的切入點，例如：修改模型、加入 STN、訓練方法的改善等，基於這些改善方法，因此本研究以 Compositional GAN 作為模型基礎，我們提出 LogoGANs：改善組合生成對抗神經網路生成 Logo 標誌，該模型主要分成三大部分，分別是幾何模型、組合模型以及著色模型，如圖2所示。

A. 幾何模型架構(Geometry Module)

幾何模型架構是以 OpenAI 提出的 DALL-E [9] 為基本模型，DALL-E 是一個 Transformer 語言模型，由 Aditya Ramesh 等人於 2021 年 2 月基於 GPT-3 開發的一種新型神經網路，該模型可接收文字和圖像作為輸入，了解物件之間的關係，並在其理解的基礎上建立自動回歸生成高保真圖像。DALL-E 使用離散變分自編碼器 (Discrete Variational Autoencoders, DVAE) 的組合來獲得離散的潛在表示，同時利用基於 Autoregressive Transformer 的解碼器來生成高質量的圖像。由於組合模型架構需要接收二值化圖像當作輸入，我們整理出一個僅有二值化圖像的數據集來微調 DALL-E，

使其能夠生成符合組合模型架構所需之圖像，該數據集包括動植物、日常用品、行為動作等圖像和類型的文字描述。如圖2 (a) 所示。

B. 組合模型架構(Composition Module)

本研究所提出的組合模型架構是以 Compositional GAN 為基本模型，該模型是由 Samaneh Azadi 等人於 2018 年提出，該網路能夠透過 Composition-by-Decomposition (CoDe) 網路來組合物件，但該網路在訓練時仍舊使用成對的數據集進行訓練，為了使網路能夠在非成對的數據上進行訓練，我們保留了原先的 CoDe 網路，並在其中加入循環生成對抗網路 [10] 的 Cycle Consistency Loss，通過 CoDe 網路分解後的 (\hat{X}, \hat{Y}) 必須再次輸入 CoDe 網路進行組合，並且使重組合後的圖像也必須能夠欺騙判別器。我們的目標是確保組合後的合成圖像必須包含對象(X, Y)，使其具有與 Logo 標誌相同的顏色與層次，該網路不是學習所有可能的組合，而是學習數據的分佈模式。

由於 CoDe 網路僅能將圖像進行組合與分解，仍不足以在空間上變換物體，我們在將圖像輸入 CoDe 網路前，需要先通過空間變換網路將圖像進行平移與縮放，如圖2 (b)所示。由於空間變換網路在進行非監督複數圖像空間變換時，容易造成部分圖像過度放大或縮小，我們將計算圖像的像素值作為圖像大小，並計算圖像大小彼此之間的歐式距離納入空間變換網路損失函數當中，以避免某一圖像被網路過度放大或縮小的問題。此外，該網路不需以特定圖像作為背景圖像進行空間變換，大幅增加網路在組合上的更多可能性。

在圖像組合任務上，除了圖像的空間變換與組合之外，圖像間的交互關係也至關重要，圖像間的交互關係包括彼此之間的層次結構及顏色對比。我們將通過空間變換後的圖像作為輸入，分別通過 Densenet121 與 Densenet161 網路，判斷圖像間彼此的層次關係與亮度色彩關係，如圖5所示。圖像物體之間的遮擋關係明確影響著 Logo 標誌的組合與視覺真實感，因此，我們引入層次結構損失來判斷圖像間彼此的層次關係，層次結構損失函數可以表示為：

$$\begin{aligned} L_{\text{comp}}(G) = & X_{\text{weight}} * \lambda_{\text{hierarchy}} \| \hat{C}_{\text{mask}_X} - X_{\text{mask}_X} \|_1 \\ & + Y_{\text{weight}} * \lambda_{\text{hierarchy}} \| \hat{C}_{\text{mask}_Y} - Y_{\text{mask}_Y} \|_1 \\ & + Z_{\text{weight}} * \lambda_{\text{hierarchy}} \| \hat{C}_{\text{mask}_Z} - Z_{\text{mask}_Z} \|_1 \end{aligned} \quad (1)$$

其中 X_{weight} 、 Y_{weight} 、 Z_{weight} 表示為層次結構模型的輸出權重， \hat{C}_{mask_X} 、 \hat{C}_{mask_Y} 、 \hat{C}_{mask_Z} 分別表示為組合圖像中的 X、Y、Z Mask， X_{mask_X} 、 Y_{mask_Y} 、 Z_{mask_Z} 分別表示每張圖像中的 X、Y、Z Mask。

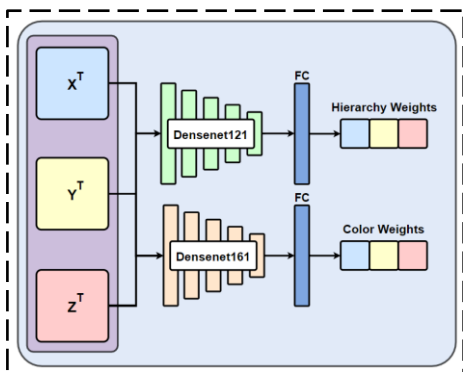


圖5、層次結構模型與亮度色彩模型架構

除了圖像彼此之間的遮擋關係外，組合圖像間的顏色也會影響判別器的判斷。當我們在訓練 LogoGANs 的組合模型時，使用公開資料集 LLD-icon sample [11] 當作真實數據進行訓練，組合模型架構的輸入為複數張二值化圖像，輸出為單張灰度組合圖像，因此，我們使用亮度色彩模型來給予每張圖像灰度顏色，使其能夠輸入著色模型進行上色。

為了避免層次結構與亮度色彩模型的輸出導致單張或多張圖像完全消失，我們也提出了新的考試機制訓練方法，關於考試機制訓練方法的詳細介紹會在第四章提供。

C. 著色模型架構(Appearance Module)

由於數據集較小，本研究所提出的著色模型使用預訓練 ResNet18 作為 Unet 的主要架構進行遷移學習並進一步對其進行微調，該模型是由 Shreyas Kalvankar 等人於 2021 年在探索兩種不同顏色空間 RGB 和 L^*a^*b 時，所修改的模型架構 [12]。我們在模型的輸入及輸出上使用 YCbCr 彩色空間，亦即將灰度圖像當作 Y 通道輸入 ResUnet 網路中，該網路僅需預測並輸出 CbCr 通道，將三者合併後轉換為 RGB 圖像，便能預測組合圖像之顏色，如圖2 (c)所示。

D. 損失函數(Loss Function)

我們的組合網路架構如圖2 (b)所示，該網路的損失函數組成包含：

- 組合分解網路的條件生成對抗網路損失函數：

$$\begin{aligned} L_{cGAN}(G_c, D_c) = & E_{(X, Y, C)} [\log D_c(X^T, Y^T, C)] \\ & + E_{(X, Y)} [1 - \log D_c(X^T, Y^T, \hat{C})] \end{aligned} \quad (2)$$

$$\begin{aligned} L_{cGAN}(G_{dec}, D_{dec}) = & E_{(X, Y)} [\log D_{dec}(\hat{C}, X) + \log D_{dec}(\hat{C}, Y)] \\ & + E_{(X, Y)} [(1 - \log D_{dec}(\hat{C}, X^T)) + (1 - \log D_{dec}(\hat{C}, Y^T))] \end{aligned} \quad (3)$$

- 空間變換網路的對抗損失函數：

$$L_{GAN}(G_c) = E_{r \sim p_r} [\log D(C)] + E_{f \sim p_f} [1 - \log D(\hat{C})] \quad (4)$$

- 分解生成器和空間變換網路的輸出 L1 損失函數：

$$L_{L_1}(G_{dec}) = E_{(X, Y)} [\| (X^T, Y^T) - G_{dec}(\hat{C}) \|_1] \quad (5)$$

$$\begin{aligned} L_{L_1}(STN) = & E_{(X, Y)} [\| (X, Y) - (X^T, Y^T) \|_1] \\ & + E_{(X, Y)} [\| X_{\text{size}} - Y_{\text{size}} \|_1] \end{aligned} \quad (6)$$

- 循環一致性損失函數：

$$L_{cyc}(G_c, G_{dec}) = E_{C \sim p_C} [\| G_c(G_{dec}(\hat{C})) - \hat{C} \|_1] \quad (7)$$

我們還參考了 WGAN-GP [13] 的訓練方式引入的梯度懲罰，以提高生成對抗網路損失函數的收斂性。完整的組合網路模型損失函數表示如下：

$$\begin{aligned} L(G) = & \lambda_1 [L_{cGAN}(G_c, D_c) + L_{cGAN}(G_{dec}, D_{dec})] \\ & + \lambda_2 L_{\text{comp}}(G) \\ & + \lambda_3 L_{L_1}(G_{dec}) \\ & + \lambda_4 L_{cyc}(G_c, G_{dec}) \end{aligned} \quad (8)$$

$$L(STN) = \lambda_5 L_{GAN}(G_c) + \lambda_6 L_{L_1}(STN) \quad (9)$$

在本研究中，訓練的超參數設置為 $\lambda_1 = 1$ 、 $\lambda_2 = 100$ 、 $\lambda_3 = 100$ 、 $\lambda_4 = 10$ 、 $\lambda_5 = 10$ 、 $\lambda_6 = 100$ 。

IV. 實驗結果與討論

A. 實驗一：加入STN的對抗網路訓練

本研究將空間變換網路與組合生成器結合，替代原本的生成器進行訓練，在實驗一中的改善重點是讓空間變換網路與組合生成器合作訓練騙過判別器，但在實際訓練過程中我們發現空間變換網路在訓練前期並不會投入訓練，而是會等到生成器訓練到一定程度後仍然無法騙過判別器，空間變換網路才會實際投入訓練。由於兩個網路的訓練特性，在訓練 LogoGANs 網路時，必須給予組合生成器一定程度的訓

練限制，若在訓練網路時不限制組合生成器的發展，容易造成組合生成器過度訓練及空間變換器無法訓練直到訓練崩潰的問題。

B. 實驗二：考試機制訓練之改善

原始生成對抗網路的訓練方式是透過生成器與判別器互相對抗的概念來做訓練，生成器的目的是學習訓練數據的分佈，生成盡可能真實的圖像，以確保判別器無法區分；判別器需要不斷地學習生成器偽造的圖像，以防止被欺騙。由於當組合模型訓練一定程度後，Compositional GAN 所能控制的圖像分佈差異不大，判別器很容易能區分出偽造的圖像，造成對抗網路的訓練崩潰，為了解決網路訓練崩潰問題，本研究改善原始生成對抗網路的訓練方式，提出新的考試機制訓練，判別器在訓練過程中不會直接學習到生成器偽造的圖像分佈，而是學習判別生成器是否正確組合圖像，如圖6所示。在第三章中提到生成器中的層次結構與亮度色彩模型輸出可能會導致圖像消失問題，若改使用考試機制訓練，當生成器的輸出導致圖像消失時，判別器便能輕易判斷出偽造圖像，促使訓練生成器將消失的圖像層次提升或是增強亮度對比，用以欺騙判別器。

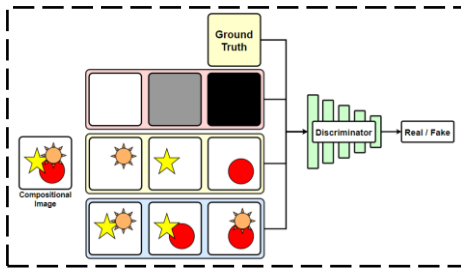


圖6、考試機制訓練

C. 實驗結果

圖7顯示了 LogoGANs 在不同數量下的圖像組合任務之輸出結果，從圖中可以看到該網路能夠將複數圖像透過空間變換網路進行平移與縮放，並與組合生成器順利結合進行訓練，組合出多樣化的 Logo 標誌。我們在相同物件下 LogoGANs 訓練的結果，如圖8所示，在相同物件下 LogoGANs 能夠組合出多種不同的 Logo 標誌，也顯示該網路擁有高隨機性的組合生成能力。

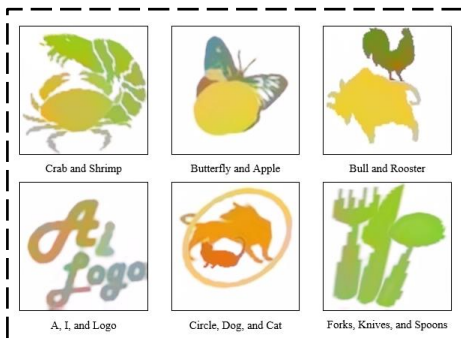


圖7、LogoGANs 輸出結果



圖8、LogoGANs 於相同物件下之輸出結果

V. 結論與未來展望

在本研究中我們提出了一種新的組合生成對抗網路模型，解決了原先組合生成對抗網路需要成對數據進行訓練的問題，並且能在複數圖像間進行組合產生新的 Logo 標誌。透過將生成器結合空間變換網路進行訓練以及利用考試機制訓練避免造成物件消失與網路訓練崩潰的問題，進而提升模型表現。將空間變換網路加入至生成對抗網路的訓練當中，透過一定程度的限制生成器發展，平衡空間變換網路、生成器、判別器三者之間的訓練，避免生成對抗網路訓練崩潰的問題。

在未來展望方面，若使用更好的神經網路模型來替代現有的網路架構，在層次及色彩方面的判斷與表現也能更加優秀，使 Logo 標誌更加美觀。對於組合模型架構，可以嘗試將該模型運用於其他生成對抗網路的任務，或是將其訓練方法加以改進其他網路的訓練當中，提高模型的生成效果。

參考文獻

- [1] I. Goodfellow et al., "Generative Adversarial Nets," in Advances in Neural Information Processing Systems, 2014, vol. 27.
- [2] A. Mino and G. Spanakis, "LoGAN: Generating Logos with a Generative Adversarial Neural Network Conditioned on Color," 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), 2018, pp. 965-970.
- [3] C. Oeldorf and G. Spanakis, "LoGANv2: Conditional Style-Based Logo Generation with Generative Adversarial Networks," 2019 18th IEEE International Conference on Machine Learning And Applications (ICMLA), 2019, pp. 462-468.
- [4] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive Growing of GANs for Improved Quality, Stability, and Variation," in International Conference on Learning Representations, 2018.
- [5] T. Karras, S. Laine, and T. Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 4401-4410.
- [6] S. Azadi, D. Pathak, S. Ebrahimi, and T. Darrell, "Compositional GAN: Learning Conditional Image Composition," arXiv:1807.07560, 2018.
- [7] C.-H. Lin, E. Yumer, O. Wang, E. Shechtman, and S. Lucey, "ST-GAN: Spatial Transformer Generative Adversarial Networks for Image Compositing," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 9455-9464.
- [8] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial Transformer Networks," In NIPS, 2015.
- [9] A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, and I. Sutskever, "Zero-Shot Text-to-Image Generation," Proceedings of the 38th International Conference on Machine Learning, PMLR 139:8821-8831, 2021.
- [10] J. Y. Zhu, T. Park, P. Isola, A. Efros, "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2223-2232.
- [11] A. Sage, E. Agustsson, R. Timofte, and L. Van Gool, "Lld - Large Logo Dataset - Version 0.1," <https://data.vision.ee.ethz.ch/cvl/lld>, 2017.
- [12] S. Kalvankar, H. Pandit, P. Parwate, A. Patil, and S. Kamalapur, "Astronomical Image Colorization and upscaling with Generative Adversarial Networks," arXiv:2112.13865, 2021.
- [13] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved Training of Wasserstein GANs," Part of Advances in Neural Information Processing Systems 30 (NIPS), 2017.

致謝

科技部在研究經費上的補助，計畫編號：109-2221-E-194-022-MY3、111-2813-C-194-019-E