

基於生成對抗網路架構使用未批配圖像遷移達成色彩柔和

The Use of Unpaired Image-to-Image Translation for Soft Color Based On Generative Adversarial Network

周佳鋒¹ 劉遠楨²

JOU, JIA FENG¹ LIU, YUAN CHEN²

¹ 國立臺北教育大學 資訊科學研究所 研究生

¹ National Taipei University of Department of Computer Science
College of Science Student

E-mail : lucky48tw@gmail.com

² 國立臺北教育大學 資訊科學研究所 教授

² National Taipei University of Department of Computer Science
College of Science Professor

E-mail : liu@tea.ntue.edu.tw

摘要

以往訓練神經網路，必須透過人工繁複作業收集、整理、標記大量資料，再將資料放入神經網路中訓練。但是對於特定領域資料收集並不容易且通常所費不貲。相對地，使用不成對數據集和未監督式架構進行圖像對圖像的轉換，不但能增加資料集資料，也能讓資料集多樣化。而眾多研究也證明了這種資料預處理的方法。例如，調整顏色和色相分佈，同時保留常規的顏色屬性。這也代表對象的實際顏色很重要，其顏色影響整個模型最後的準確度以及最後訓練結果。因此，圖像轉換相關技術的研究有其必要性。在深度學習上提供資料預處理，擴增資料量這個方法也會提升整體訓練結果。

本研究使用風格轉換技術作為神經網路的基礎架構，搭配跨領域轉換遷移技術的損失函數，將一個領域的特徵轉移到另一個領域的特徵上，可以靈活地學習將已經在不同領域學到的特徵連接在一起。因此，在一個領域內的條件下，就可以強化及控制潛在變數，透過控制跨領域轉換遷移技術，在條件內學習領域內生成特定樣本。研究結果證明輸出端會保留更多輸入端的 RGB 顏色特徵，並且比單純使用風格轉換技術產出的圖像顏色更柔和。

關鍵字：生成對抗網路、圖像轉換、風格轉換

Abstract

In the past, training neural networks was a complex manual work. A large amount of data had to be collected, sorted, and labeled. And then put the data into the neural work for training. But collecting data in a specific field is difficult and expensive. On the contrary, using unpaired datasets and unsupervised architectures for image-to-image translation not only increase the dataset, but also add variety to the dataset. Several studies have proved this data preprocessing method. For example, they adjusted colors and hue distribution but retained the regular attributes as well. This shows that the actual color of the object is crucial, because it affects the final accuracy of the entire model and the final training result. Therefore, the study about image-to-image translation related technology is in demand. For data preprocessing in deep learning, the approach of increasing dataset can enhance accuracy of training.

The study used the image style transfer technology as the basic framework of neural networks, which combined with the loss function of unsupervised cross-domain transfer and migration technology. The feature of one domain was transferred to the feature of another domain. One can flexibly learn from what has been learned from another domain, and then the features were connected together. Therefore, under the condition of one domain, the potential variables could be strengthened and controlled. Specific samples were generated by this transfer technology. The result proved that the output retained more input RGB color features, and the colors became softer than the one produced only by using the image style transfer technology.

Keywords : Generative Adversarial Network 、 image-to-image translation 、 image style transfer

壹、前言

自從 2014 年生成對抗網路架構(Generative Adversarial Network, GAN)[1]發表後，相關的研究開始如雨後春筍般的冒出，GAN 的核心架構就是透過一個生成器與一個鑑別器，如果應用在影像上則可以這樣理解，生成器目的是產出一張圖片而鑑別器則是要鑑定這張圖片有沒有符合標準，如果有則給予通行，真正的產生圖片，如果否則退回再讓生成器繼續生成新的一張圖片，一直重複上述步驟直到鑑別器給予通行，目的就是讓兩者相互對抗後產生出結果。

2018 年，Google AI 研究團隊發表了 AutoAugment 技術[2]，AutoAugment 主要用來強化預訓練的資料。計算機視覺領域近年來在深度學習上的一些重大突破，可以將部分功勞歸因於大量帶標籤的訓練資料的可用性，神經網路模型的性能通常會隨著增加訓練資料的質量，也就是資料的多樣性和資料數量而提高。但是，收集足夠數量又必須多樣性的數據來訓練模型通常非常困難。解決此問題的一種方法是，或者由專家手動設計數據增強方法，例如旋轉和翻轉，這些方法通常用於訓練性能良好的視覺模型。AutoAugment 是一種為計算機視覺數據集設計自定義數據增強策略的自動方式，例如，指導基本圖像轉換操作的選擇，例如水平、垂直翻轉圖像、更改圖像顏色、顏色標準化等。不僅可以預測要組合的圖像變換，而且還可以預測每個圖像使用的變換的概率和幅度，因此並非總是以相同的方式操縱圖像。AutoAugment 專注於稍微調整顏色和色相分佈，同時保留常規的顏色屬性。這表明對象的實際顏色很重要，因為對整個模型最後的準確度相當有影響。

本研究架構使用 CycleGAN[3]作為神經網路的基礎架構，搭配跨領域轉換遷移技術(Domain Transfer Network, DTN)[4]，將一個領域的特徵轉移到另一個領域，可以靈活地學習將已經在不同領域學到的特徵連接在一起，這樣在一個領域內的條件下，就可以強化並控制潛在變數並在條件內學習領域內生成特定樣本。能夠讓神經網絡更加地保留輸入特徵，並可以在已定義的領域有條件地生成樣本，而獲得很好的效果並且是在無監督的跨領域圖像生成中訓練端到端模型，假設這兩個域相關性，再循環訓練。

透過控制 DTN，我們證明輸出端會保留更多輸入端的 RGB 顏色特徵，並且比 CycleGAN 產出的圖像顏色更自然。

貳、文獻探討

一、生成對抗網路

GAN (Generative Adversarial Network) 全名叫做對抗生成網絡或者生成對抗網絡。GAN 由兩組神經網路構成，分別為生成器 G(Generative) 和鑑別器 D(Discriminative)，生成器的目的就是將隨機輸入的高斯噪聲映射成圖像但是此圖像為生成出的假圖，判別器則是判斷輸入圖像是否來自生成器的概率，即判斷輸入圖像是否為假圖的概率。GAN 的訓練也與 CNN 大不相同，CNN 是定義好特定的損失函數，然後使用梯度下降及其改進算法進行優化參數，盡可能用局部最優解去逼近全局最優解。但是 GAN 的訓練是個動態的過程，是生成器 G 與鑑別器 D 兩者之間的相互博弈過程。可以理解為 GAN 的目的就產生出一堆假的東西，並且又要有接近偽造目標的品質。就是要讓生成器 G 生成的假圖可以順利的騙過鑑別器 D，那麼最優狀態就是生成器 G 生成鎖生成的假圖可以在鑑別器 D 的判別結果為 0.5，這樣代表鑑別器 D 已經不知道到底是真圖還是假圖。整個框架是生成器 G 與鑑別器 D 兩者之間的相互博弈的動態過程。



參、研究實施與設計

一、研究方法

基於 CycleGAN 架構下，會有兩組 GAN 架構神經網路為 $G(x-y)$ 、 $D(X)$ 與 $G(y-x)$ 、 $D(Y)$ ，每組 GAN 的生成器使用 ResNet[5]，鑑別器使用 PatchGAN[6]，這樣搭配可以大幅度簡化參數，提升整體網路架構的訓練速度。整體架構流程如圖 1。

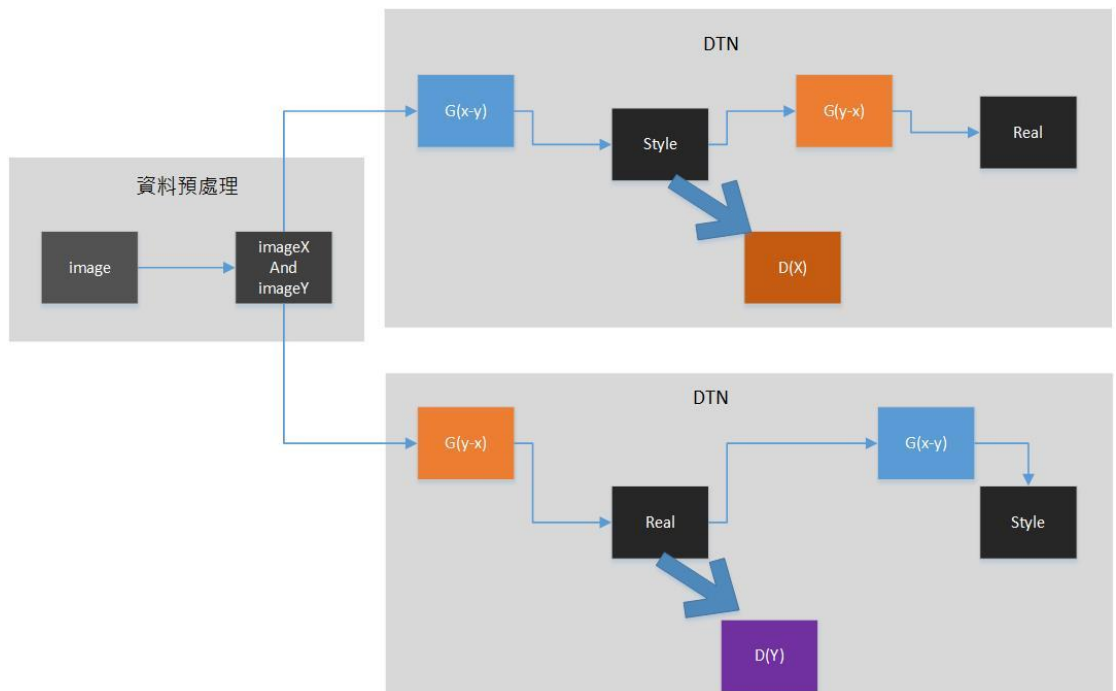


圖 1 整體神經網路架構

二、損失函數設計

原架構整體損失函數如公式(1)。

$$\begin{aligned} L(G, F, D_x, D_y) = & L_{GAN}(G, D_y, X, Y) \\ & + L_{GAN}(F, D_x, Y, X) \\ & + \lambda L_{cyc}(G, F) \end{aligned} \quad (1)$$

使用 DTN 的損失函數可以保留輸入特徵，並可以在已定義的領域有條件地生成樣本，因此在原始損失函數中加上 DTN 損失函數，改良後損失函數如公式(2)。

$$\begin{aligned}
L(G, F, Dx, Dy) = & L_{GAN}(G, Dy, X, Y) \\
& + L_{GAN}(F, Dx, Y, X) \\
& + \lambda L_{cyc}(G, F) \\
& + L_{color}(G, F)
\end{aligned} \tag{2}$$

$L_{color}(G, F)$ 拆解後，等同於公式(3)。

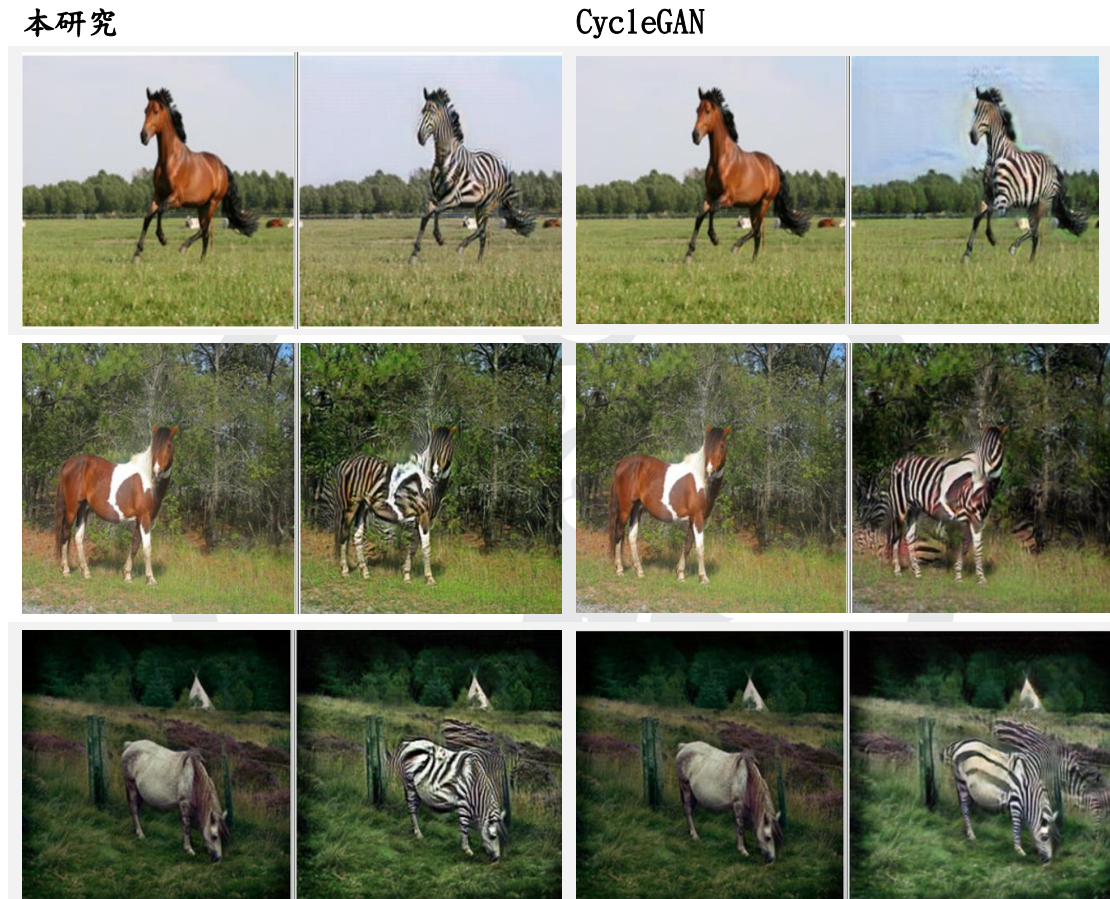
$$L_{color}(G, F) = E_{y \sim P_{data}(y)} [\|G(y) - y\|_1] + E_{x \sim P_{data}(x)} [\|F(x) - x\|_1] \tag{3}$$



肆、結果與討論

使用 horse2zebra 數據集，在 Colab 環境中訓練後，可以看出兩種方法的差異，如表 1，左邊為本研究結果可以分為輸入原始數據（左）與由生成器最後輸出圖像轉移後的假圖（右），在本研究中圖像顏色看起來更自然更柔和，也比較接近輸入端。

表 1 兩種方法的輸入（左）與輸出（右）



伍、未來展望

以往訓練神經網路，必須透過人工繁複作業收集、整理、標記大量資料，然後在放入神經網路中訓練，但是對於特定領域資料收集是不容易，並且需要付出昂貴代價，不過在 2017 年時有研究團隊發表 CycleGAN，可以使用未匹配數據集使用未監督式架構進行圖像對圖像的轉換，並且對莫內(Oscar-Claude Monet)、梵谷(Vincent Willem van Gogh)、塞尚(Paul Cézanne)等等大師的著名畫作進行圖像轉換，雖然這幾位大師已經不在人世也絕無可能發表新作品，但是我們仍可以利用這樣的技術產生出如果大師們還在的話，看到這樣的景像可以如何用他們本身的繪畫特性來表達眼中所見進一步產生出圖像，也就是利用 GAN 的特性輸入一張可能是風景圖像，對映大師們的特徵，產出帶有大師們特徵的圖像，除了在藝術方面進行轉換外，在季節風景、近似的物件方面也可以達到很好效果的轉換，例如：夏季轉換冬季、黃種馬轉換成斑馬、蘋果轉換成橘子。另外在醫學研究方面的影像資料更加難以蒐集資料，因此在 2021 年醫學研究論文中，有研究團隊發表 HistoGAN[7]，用未匹配圖像轉換技術產生可能病變的細胞來擴增資料集加強資料集圖像的多樣性，並證明了這樣資料預處理的方法是有助於最後訓練結果。

未來不論在任何的機器學習領域上一定都會使用資料來作為訓練集，如果能把原始訓練資料的品質處理好，就能先確保訓練集的品質是沒有問題的，也能有助於最後準確度的提升或是增加資料集樣本的多樣性。

參考文獻

- [1] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial networks. *arXiv preprint arXiv:1406.2661*.
- [2] Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V., & Le, Q. V. (2019). Autoaugment: Learning augmentation strategies from data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 113-123).
- [3] Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision* (pp. 2223-2232).
- [4] Taigman, Y., Polyak, A., & Wolf, L. (2016). Unsupervised cross-domain image generation. *arXiv preprint arXiv:1611.02200*.
- [5] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- [6] Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1125-1134).
- [7] Wei, J., Suriawinata, A., Vaickus, L., Ren, B., Liu, X., Wei, J., & Hassanpour, S. (2019). Generative image translation for data augmentation in colorectal histopathology images. *arXiv preprint arXiv:1910.05827*.