

人工智慧 個人專題

四電子三乙 B10602110 呂和軒

論文名稱: A Neural Algorithm of Artistic Style

一、簡介

圖像風格轉移的目標是從一幅圖像中合成另一幅圖像的紋理，但要對紋理合成進行計算，以保留目標圖像的重要內容。大多數用於紋理合成的算法都是採用非參數方法，比如使用特徵對應圖、圖像色度、邊緣提取等方法來實現，但有一個基本問題始終沒有解決。他們都只是將目標圖像的底層圖像特徵使用於風格轉換，並沒有使用更高階的圖像語義訊息，而一個理想的遷移算法應該能從目標圖像中提取語義圖像特徵，並且計算紋理合成，根據風格圖像的紋理來渲染目標圖像的語義內容。

二、相關研究

在神經網路方法前，傳統圖像風格遷移有一個共同的思路，目標在於如何分析某一種風格的圖像，給那一種風格建立一個數學或者統計模型，再改變要做轉換的圖像，讓它能符合所建立的風格模型，根據風格模型來進行圖像重建的任務，風格建模方法可分為基於統計分佈的參數化紋理建模方法及基於 MRF 的非參數化紋理建模方法，圖像重建則分為基於線上優化的慢速重建與離線優化的快速重建，可見參考圖(1)。

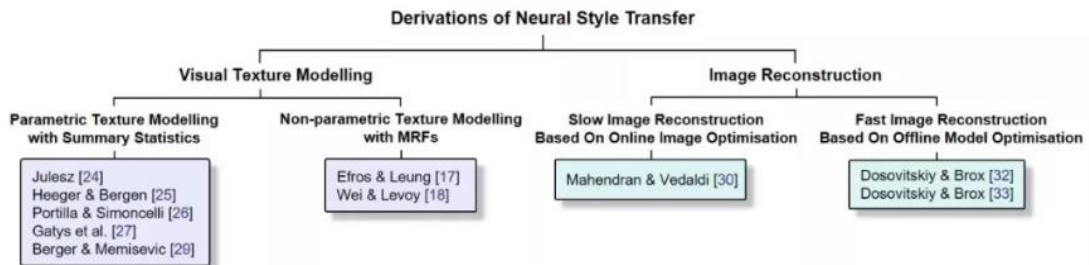


圖 1、不同紋理模型與重建方法

根據不同的方法與發展，目前圖像風格轉換方法種類如圖(2)所示，可分為慢速與快速的轉換方法，慢速轉換的有基於統計分佈參數化及 MRF(Markov Random Field)非參數化的轉換，而快速轉換則主要基於神經網路模型，藉由預先訓練解決計算量大、速度慢的問題，根據可轉換的風格數量分為單模型單風格 (PSPM)、單模型多風格 (MSPM) 和單模型任意風格 (ASPM)，後續許多方法皆是基於此篇論文的概念，引入高階的圖像語意特徵來提高風格轉換的品質，將圖像風格轉換帶領至神經網路的時代。

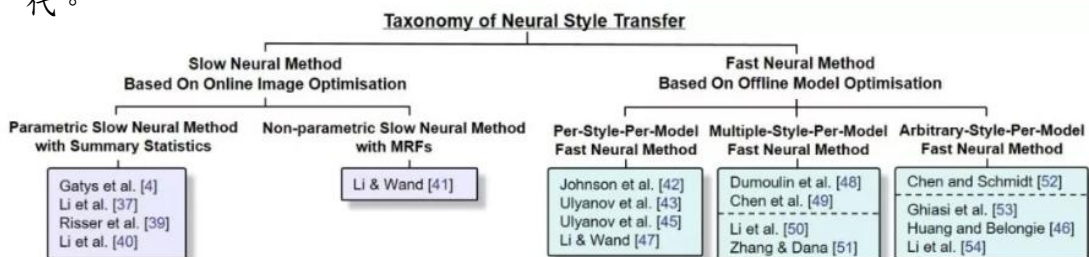


圖 2、目前風格轉換方法種類

三、實現方法

為了有效的提取圖像的語義訊息，文中引入了卷積神經網絡，利用卷積神經網絡來獲得更深層的圖像特徵，文中使用 VGG 網絡進行運算，VGG 常用於圖像特徵抽取，本文使用了由 16 個卷積層和 5 個池化層的 VGG19，不使用任何全連接層，將內容圖和風格圖進行卷積運算，經由高性能 VGG19 計算得到的廣義特徵，透過格拉姆矩陣及平方差評估圖像風格與內容的損失，定義出需要最小化的損失函數，對圖像進行梯度更新，以下小節依序介紹圖像內容表達、圖像風格表達及風格轉換流程。

I、Content representation (圖像內容表達)

VGG19 可以分為五個卷積區塊，每個卷積區塊的輸出都是輸入圖像的一組特徵，每層所包含的圖像訊息深度不同，透過 VGG19 輸入圖像 \vec{X} ，抽取對應的特徵，內容特徵為卷積的輸出，由 N_l 個 filter 產生 N_l 個大小為 $M_l[\text{height} * \text{width}]$ 的特徵圖，存儲在矩陣 $F^l \in \mathbb{R}^{N_l \times M_l}$ 中， l 代表第 l 層的卷積輸出， N_l 代表具有 N 個卷積核， M_l 表示一維特徵圖的大小，以長乘寬計算，以這樣的形式來表達圖像內容特徵。

令 \vec{p} 和 \vec{x} 分別表示原始內容圖像和生成的圖像， P^l 和 F^l 分別為它們對應的第 l 層的特徵表示；將這兩個特徵表示之間的平方差損失定義為式(1)，損失函數的偏導數對應 l 層的激活函數為式(2)，利用內容損失函數使內容圖像和生成圖像的特徵越相近越好，使得生成圖像近似於輸入的內容圖像。

$$\mathcal{L}_{\text{content}}(\vec{p}, \vec{x}, l) = \frac{1}{2} \sum_{i,j} (F_{ij}^l - P_{ij}^l)^2. \quad (1) \quad \frac{\partial \mathcal{L}_{\text{content}}}{\partial F_{ij}^l} = \begin{cases} (F^l - P^l)_{ij} & \text{if } F_{ij}^l > 0 \\ 0 & \text{if } F_{ij}^l < 0, \end{cases} \quad (2)$$

II、Style representation (圖像風格表達)

為了獲得輸入圖像的風格特徵，用 VGG19 獲得紋理信息，由不同濾波器響應結果的相關性組成，其中特徵關係使用格拉姆矩陣 G^l 表示， G^l 的計算方式如式(3)所示，用來衡量第 l 層中不同濾波器輸出之間的相關性，表達圖像的紋理特徵，和像素位置無關。

$$G_{ij}^l = \sum_k F_{ik}^l F_{jk}^l. \quad (3)$$

令 \vec{a} 和 \vec{x} 分別表示欲轉換的風格圖像和生成的圖像， A^l 和 G^l 分別為它們對應的第 l 層的格拉姆矩陣；將這兩個格拉姆矩陣之間的平方差定義為式(4)，並對於所有卷積層的損失進行權重加總，得到風格損失函數式(5)，風格損失函數的偏導數對應 l 層的激活函數為式(6)，利用風格損失函數使風格圖像和生成圖像的紋理越相近越好，使得生成圖像的風格近似於欲轉換的風格圖像。

$$E_l = \frac{1}{4N_l^2 M_l^2} \sum_{i,j} (G_{ij}^l - A_{ij}^l)^2 \quad (4) \quad \mathcal{L}_{\text{style}}(\vec{a}, \vec{x}) = \sum_{l=0}^L w_l E_l, \quad (5)$$

$$\frac{\partial E_l}{\partial F_{ij}^l} = \begin{cases} \frac{1}{N_l^2 M_l^2} ((F^l)^T (G^l - A^l))_{ji} & \text{if } F_{ij}^l > 0 \\ 0 & \text{if } F_{ij}^l < 0 \end{cases} \quad (6)$$

III、Style transfer (風格轉換)

整體的風格轉換流程如圖(3)所示， \vec{a} 為目標風格圖像， \vec{p} 為目標內容圖像，為了將 \vec{a} 的風格轉移到 \vec{p} 上，必須合成一個新的圖像 \vec{x} ，同步匹配 \vec{p} 的內容和 \vec{a} 的風格表示，需要最小化的損失函數如式(7)所示，為內容損失與風格損失的權重加總， α 和 β 分別是內容和風格重建的權重參數。

$$\mathcal{L}_{\text{total}}(\vec{p}, \vec{a}, \vec{x}) = \alpha \mathcal{L}_{\text{content}}(\vec{p}, \vec{x}) + \beta \mathcal{L}_{\text{style}}(\vec{a}, \vec{x}) \quad (7)$$

轉換流程如下，首先是內容特徵 P^l 和風格特徵 A^l 的提取並與存儲，風格圖片 \vec{a} 經過整個 VGG19 網絡，在所有卷積層上風格特徵 A^l 被計算和存儲(圖 1 左半部)。內容圖片 \vec{p} 同樣經過 VGG19，運算得到內容特徵 P^l 並存儲(圖 1 右半部)。將隨機生成的白雜訊圖像 \vec{x} 輸入至 VGG19 網絡中，取得風格特徵 G^l 及內容特徵 F^l ，並計算內容損失及風格損失，求得整體損失函數 $\mathcal{L}_{\text{total}}$ ，透過梯度反向傳播來反復更新圖片 \vec{x} ，使得合成的風格特徵 G^l 與真實的風格特徵 A^l 相近，合成的內容特徵 F^l 則與真實的內容特徵 P^l 相近，在內容與風格間達到平衡。

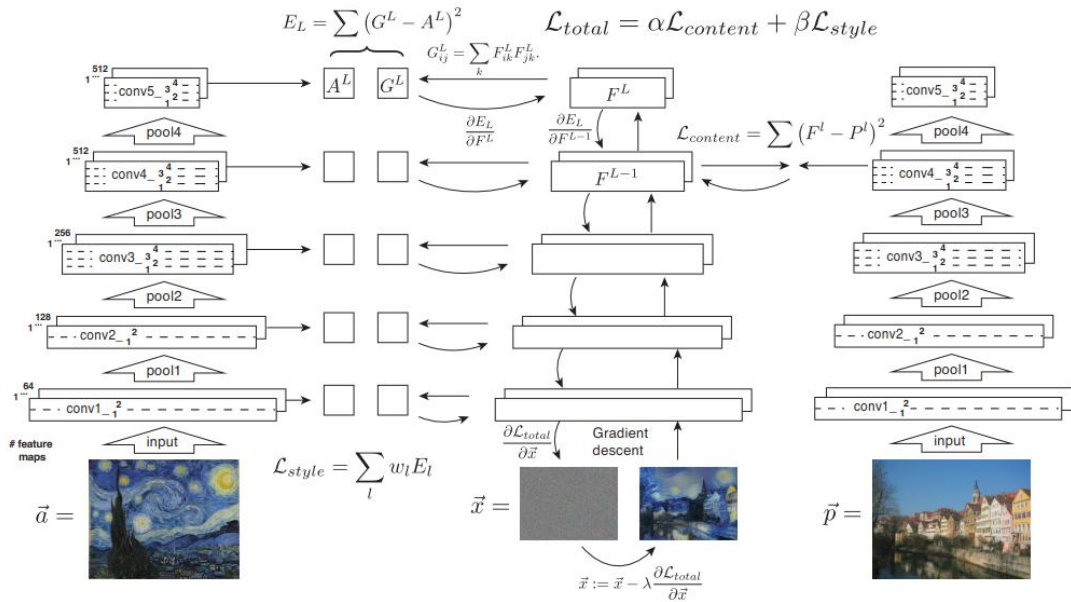


圖 3、風格轉換流程

四、實驗測試與結果

論文提出的風格轉換方法舉有幾種參數可以進行調整，論文中實驗調整了損失函數的加權比重(α 、 β)、VGG19 抽取特徵的層級選擇、輸入圖像 \bar{x} 的初始化方法，不同的參數會造成風格轉換的效果差異，將在以下小節作介紹與結果展示。

I、損失函數的加權比重影響

α 與 β 大小會影響生成圖像的風格強度與內容的保留度，論文中實驗了不同的 α/β 的比重進行風格轉換，成果如圖(4)所示，可以看到當 α/β 越小時，生成的圖像風格會較為強烈，而原始內容越不明顯，反之則風格強度減少，內容保留更加完整。



圖 4、不同 α/β 比重的轉換結果

II、特徵層級選擇的影響

在實現方法中有提到，內容圖像輸入至 VGG19 抽取內容特徵時，不同卷積區塊都有一組內容特徵，計算內容損失時僅抽取其中一層來運算，不同深度的圖像特徵會影響生成圖像的結果，論文實現分別測試抽取 Conv2_2 及 Conv4_2 的圖像特徵來比較，結果如圖(5)所示，取用較低層級的內容特徵時，生成出的圖像則保有較細緻的像素資訊，內容圖像的紋理保留更多，而使用高層級的圖像特徵時，捨棄較多的內容像素的細節，保留整體語意訊息，並可以與風格圖像進行較好的融合。

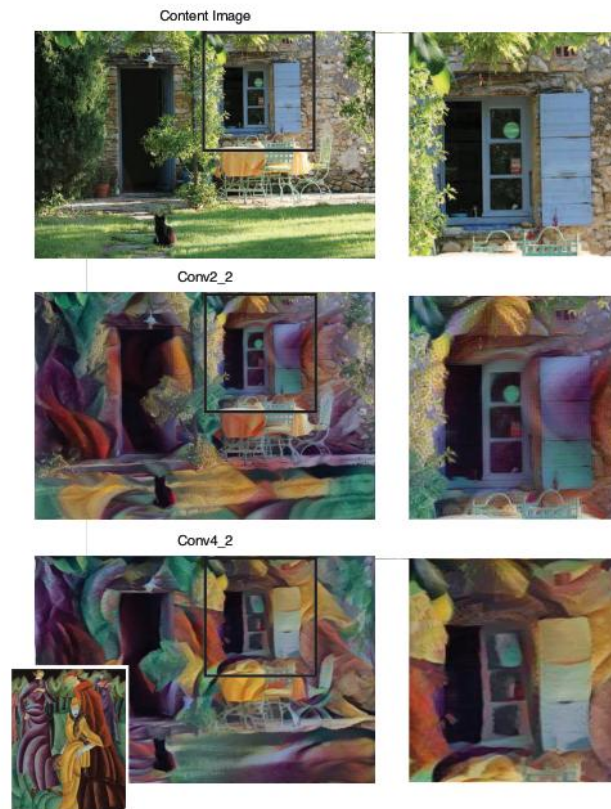


圖 5、不同層級特徵的影響

III、輸入圖像初始化的影響

在實現方法中有提到，生成圖像 \hat{x} 是使用隨機白雜訊進行初始化，再利用損失函數對其做梯度更新，使得 \hat{x} 能最小化損失函數，論文實驗測試使用其他初始化方式，會對最終生成圖像造成哪些影響，論文中分別嘗試直接使用內容圖像、風格圖像作為 \hat{x} ，進行風格轉換，成果如圖(6)所示，可以看到整體轉換結果並沒有太多的差異，但使用固定圖像進行風格轉換僅能得到同樣生成圖像。

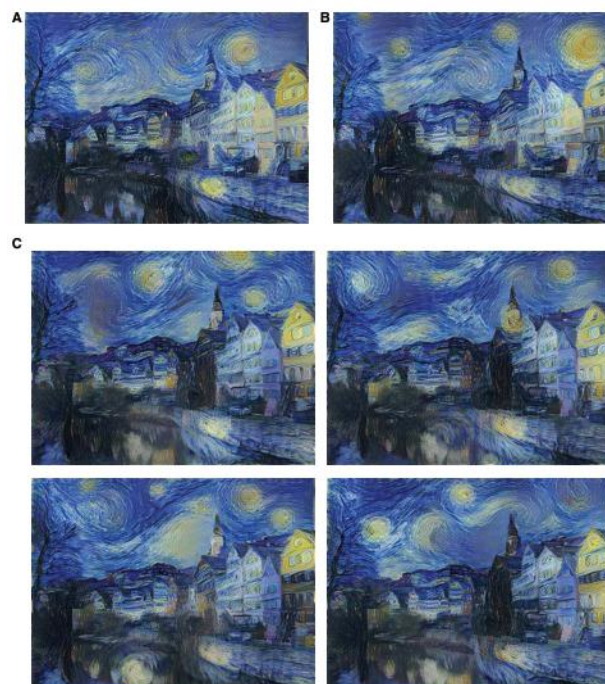


圖 6、不同初始化方式的轉換結果，圖 A 為使用內容圖像作為 \hat{x} 的生成結果，圖 B 為使用風格圖像作為 \hat{x} 的生成結果，圖 C 為四張使用隨機白雜訊初始化的生成結果。

五、總結

論文中展示了基於卷積神經網路來實現圖像風格轉換的方法，使用圖像特徵進行風格、內容評估，不同於傳統紋理建模的實現方法，首先使用格拉姆矩陣作為風格紋理的特徵表示，並且能生成效果相當優秀的風格轉換圖像，為 Neural Style Transfer 領域的開山之作，但儘管如此，此方法仍然有許多限制與不足的地方。

第一個不足是圖像生成的運算速度，由於生成過程需要持續的進行損失求導，對生成圖像進行更新，需要迭代多次才能獲得成果較好的生成圖像，因此對於生成高解度的圖像需要花費時間運算，儘管能輸入任意的內容圖象及風格圖像進行轉換，得到的成果也相當不錯，但此方法的運算效率及實用性仍然不高。

第二個問題是生成圖像會受到低階雜訊的影像，雖然對於藝術圖畫的轉換影響較小，但對於內容圖像及風格圖像皆為真實影像時，雜訊的影響會十分明顯，如圖(7)所示，可以看出雜訊嚴重影響生成圖像品質，因此必須雜訊問題進行額外的處理與設計技巧，讓圖像看起來更加自然。

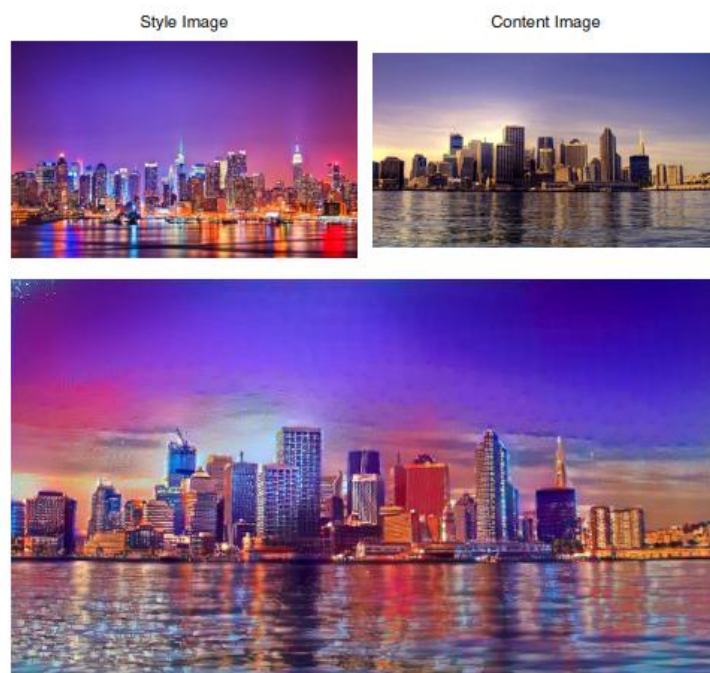


圖 7、真實圖像風格轉換。內容與風格皆使用真實圖像。

儘管此方法具有以上幾點問題，但仍然對 Neural Style Transfer 領域有相當大的貢獻，提供風格轉換一個全新的切入觀點，後續許多研究也都基於此篇論文的啟發，對此方法進行改良與增強，更加快速、強健的演算法及模型相繼提出，帶動了基於深度學習圖像風格轉換的相關研究，此領域的潛力及未來發展非常值得關注與研究。

References

- [1] L. A. Gatys, A. S. Ecker, and M. Bethge, “A neural algorithm of artistic style,” ArXiv e-prints, Aug. 2015.
- [2] L. A. Gatys, A. S. Ecker, and M. Bethge, “Image style transfer using convolutional neural networks,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2414–2423.
- [3] Y. Jing, Y. Yang, Z. Feng, J. Ye, and M. Song. Neural style transfer: A review. T-VCG, 2017.