

基於GAN語義分割研究之總結報告

(一) 相關工作

1.1 語義分割

基於 Convolution Neural Network (CNN): FCN、DeconvNet、U-Net、SegNet、DeepLab、RefineNet、PSPNet、GSCNN

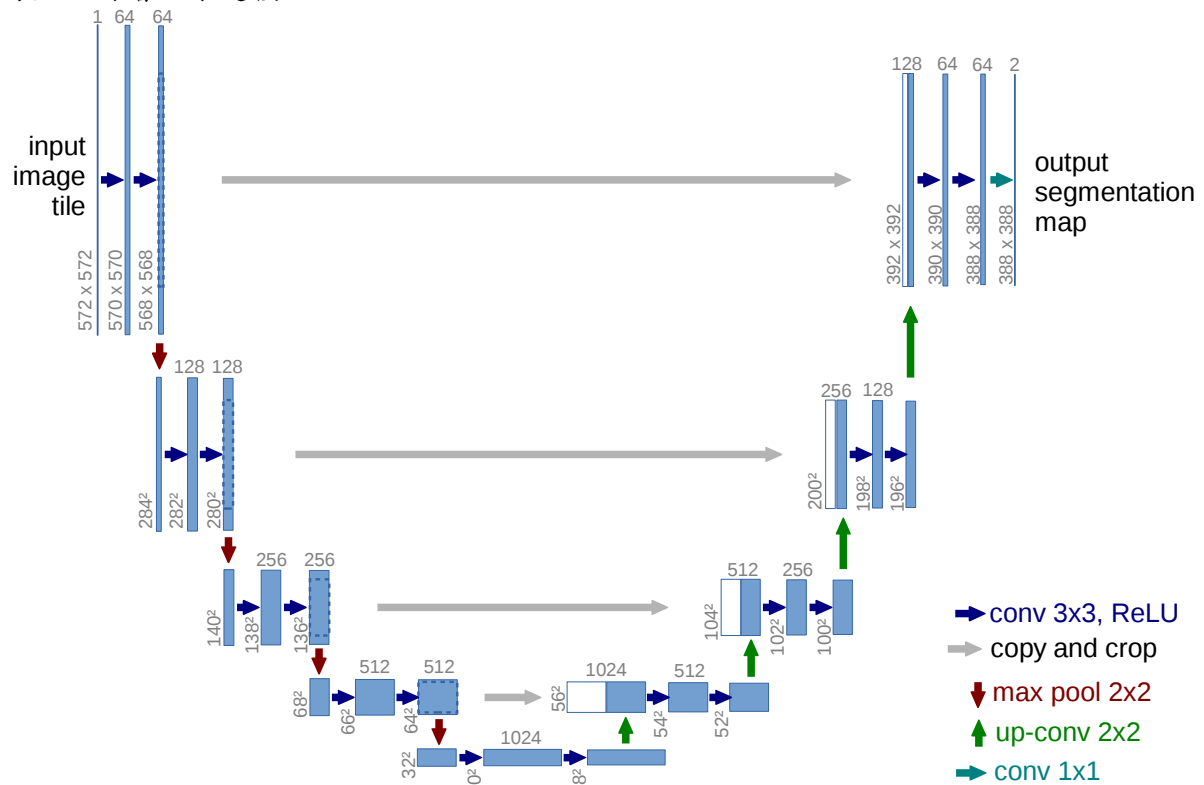
基於 Recurrent Neural Network (RNN): ReNet、ReSeg

基於 Generative adversarial network (GAN): pix2pix、Probalistic Unet

基於 Transformer: HRNet、OCRNet、HRNet-OCR、Point Transformer、SETR

1.2 U-Net

Olaf Ronneberger 在 2015 MICCAI 上發表 [U-Net: Convolutional Networks for Biomedical Image Segmentation](#)，基於編碼器-解碼器(Encoder-Decoder)結構。傳統的 Autoencoder 先使用編碼器來提取特徵，使輸出的尺寸越來越小；再使用解碼器將變小的特徵重建成與原圖一樣大小的新圖像。因為原圖有很多資訊沒能傳遞給解碼器，U-Net 在原有的編碼器和解碼器之間增加了連接。



1.3 Conditional GAN

Mehdi Mirza 在 2014 年發表 [Conditional Generative Adversarial Nets](#)，在生成器 (Generator) 和鑑別器 (Discriminator) 的輸入中加入額外條件信息，生成器生成的圖像只有足夠真實且與條件相符才能順利通過鑑別器，希望可以控制 GAN 生成的圖像而不是隨機生成圖像。

原始GAN：

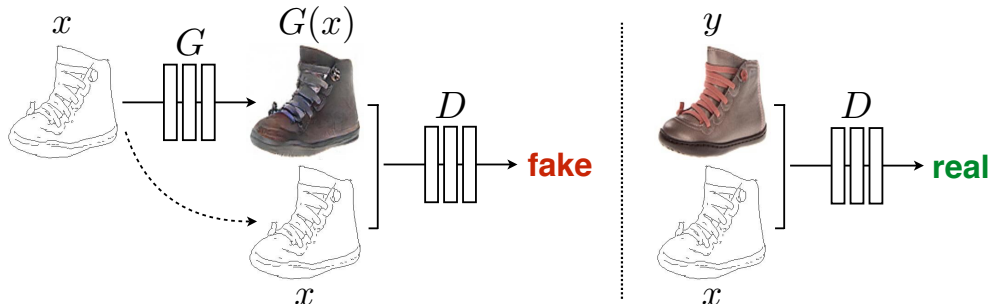
$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))]$$

CGAN：

$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})} [\log D(\mathbf{x}|\mathbf{y})] + \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log(1 - D(G(\mathbf{z}|\mathbf{y})))]$$

1.4 pix2pix

Phillip Isola 在 2017 CVPR 上發表 [Image-to-Image Translation with Conditional Adversarial Nets](#)，將 GAN 應用於有監督的圖像到圖像翻譯(image-to-image translation)。



另外真實圖像 y 和 x 也基於通道維度合併在一起，作為判別器 D 的輸入得到概率預測值。因此判別器 D 的訓練目標就是在輸入不是一對真實圖像 (x 和 $G(x)$) 時輸出小的概率值 (比如最小是 0)，在輸入是一對真實圖像 (x 和 y) 時輸出大的概率值 (比如最大是 1)。生成器 G 的訓練目標就是使得生成的 $G(x)$ 和 x 作為判別器 D 的輸入時，判別器 D 輸出的概率值盡可能大，這樣就相當於成功欺騙了判別器 D 。

$$G^* = \arg \min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G)$$

pix2pix 在訓練時需要成對的圖像 x (輸入圖像的邊緣圖像) 和 y (輸入圖像)， x 作為生成器 G 的輸入，將 $G(x)$ 和 x 基於通道維度合併在一起，作為鑑別器 D 的輸入得到預測概率值，預測概率值表示輸入是否是一對真實圖像，概率值越接近 1 表示鑑別器 D 越肯定輸入是一對真實圖像。 y 和 x 也基於通道維度合併在一起，作為鑑別器 D 的輸入得到預測概率值。因此鑑別器 D 的訓練目標是在輸入不是一對真實圖像 (x 和 $G(x)$) 時輸出小的概率值 (最小值是 0)，在輸入是一對真實圖像 (x 和 y) 時輸出大的概率值 (最大值是 1)。

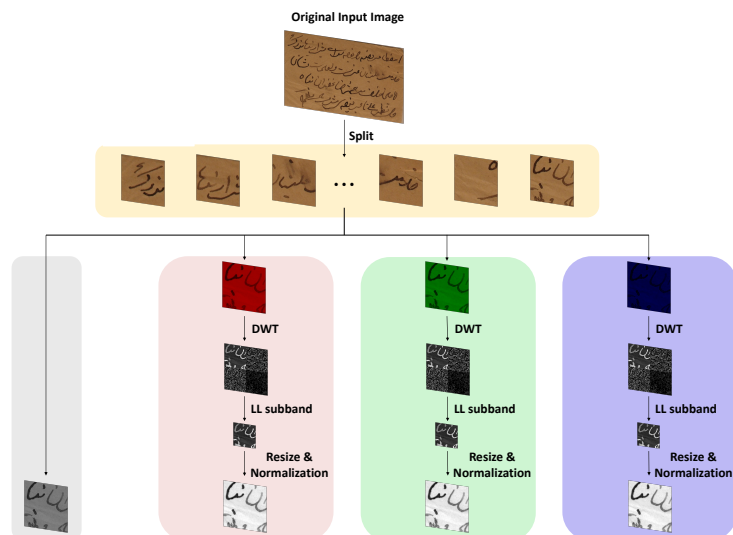
$L1$ 距離用來約束生成圖像 $G(x, z)$ 和真實圖像 y 之間的差異。論文證明了使用 $L2$ 平方差生成的圖像比使用 $L1$ 相差絕對值生成的圖像更模糊。

(二) 研究成果

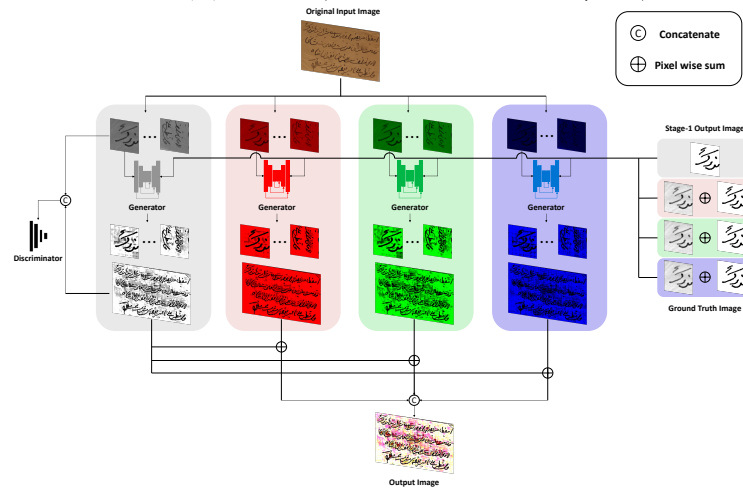
Yu-Shian Lin*, Rui-Yang Ju*, Chih-Chia Chen, Ting-Yu Lin, Jen-Shiun Chiang. [Three-stage binarization of color document images based on discrete wavelet transform and generative adversarial networks](#). *arXiv preprint*.

(三) 研究內容

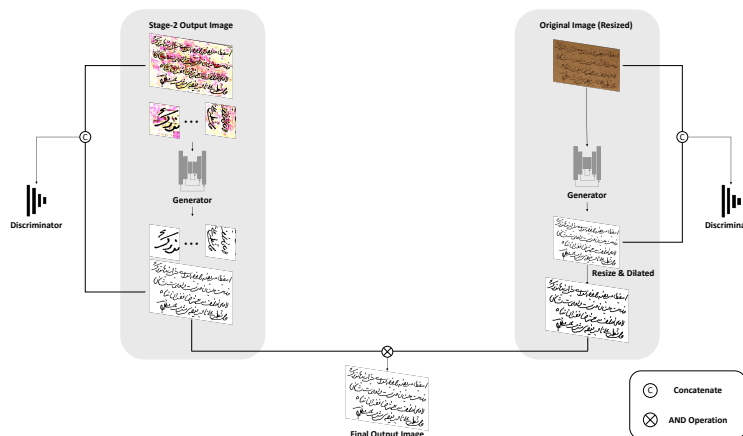
為了從彩色文檔圖像中提取出前景信息，我們提出了一種新的生成對抗網路(GAN)，對彩色文檔進行圖像增強和二值化處理，所提出的方法由三個階段構成。



第一個階段將原始輸入圖像拆分為四種不同顏色的單通道圖像，對 RGB 三個通道的單通道圖像分別使用離散小波轉換(Discrete Wavelet Transform)，保留 LL 頻帶的圖像並進行正規化處理，為第二階段 RGB 三個通道分別訓練 GAN 時的 Ground Truth 圖像做準備。



第二階段利用四個單通道圖像分別訓練四個對抗網路，從局部影像塊(Patch)中去除背景資訊並提取彩色前景資訊。灰階通道圖像訓練 GAN 時直接使用 Ground Truth 圖像，而 RGB 三個通道的單通道圖像訓練 GAN 時使用第一階段的輸出圖像加上 Ground Truth 圖像作為新的 Ground Truth 圖像。



由於第二階段是對局部影像塊(Patch)訓練 GAN，第三階段先對全局圖像訓練 GAN 並輸出結果，再融合兩個輸出(第二階段和第三階段)結果作為最終輸出。