逢 甲 大 學

資訊工程學系

碩 士 論 文

|  |
| --- |
|  |
| **應用生成對抗網路增強圖像**  **解析度**  **Image Super Resolution using** **Generative Adversarial Networks** |
|  |
| 指導教授：王益文  研 究 生：黃羿衡 |
|  |

中華民國一百零七年十一月

****

# 誌謝

首先感謝家人朋友對我的支持，在學期間感謝指導教授王益文與陳德生老師的指導，適時的給予我解決問題的方向，並用引導的方式激發我獨立思考，也很感謝資訊系上所有的老師與助教的指導與協助。

感謝實驗室的學長陳伯維、林季暉、楊博鈞、管宏成、劉達彥、謝鈞傑、黃政皓、黃博熙；同學陳振偉、楊佰穎；學弟林宗賢、林杰昱、鄭宇哲在學業上的研究、討論，度過這段研究所的日子，感謝所有幫助過我的人。

# 摘要

電腦發明以來，科學家們就寄望藉由電腦的效能創造出人工智慧，然而當時硬體效能低落，也沒有足夠可運用的數據量，但是隨著半導體的發展、數據的收集，人工智慧又慢慢的回到舞台上，類神經網路近年來被許多的領域所應用。

本篇基於對抗生成網路所實現的超分辨率，並引入結構相似性(Structural Similarity)等損失函數方法，並使用Wasserstein 距離取代原本Kullback–Leibler距離，使判別器有更佳的判別能力，使產生出來的高解析度圖像符合人類視覺感知標準。

**關鍵詞：**深度學習、超分辨率、生成對抗網路

# Abstract

Since computer was invented, scientists hope to create artificial intelligence stuff through the efficiency of computers. However, the computer performance is too low to get this target, and it’s not have enough data for train, but with the semiconductor technology improvement and data collection, the artificial intelligence gets attention again, neural networks is widely used in many fields.

We proposed a new architecture base on Generative Adversarial Network, and adding Structural Similarity as an loss function, and using Wasserstein distance replace

Kullback-Leibler distance to improve the capability of discriminator, to generate a better high-resolution image.

Keywords: Deep Learning, Super-Resolution, Generative Adversarial Networks

# 目錄

[誌謝 i](#_Toc536176283)

[摘要 ii](#_Toc536176284)

[Abstract iii](#_Toc536176285)

[目錄 iv](#_Toc536176286)

[圖目錄 vi](#_Toc536176287)

[表目錄 vii](#_Toc536176288)

[第一章 緒論 1](#_Toc536176289)

[1.1 研究動機 1](#_Toc536176290)

[1.2 研究方法 1](#_Toc536176291)

[第二章 研究背景 2](#_Toc536176292)

[2.1 卷積神經網路(Convolutional Neural Networks) 2](#_Toc536176293)

[2.1.1 卷積層(Convolutional layer) 3](#_Toc536176294)

[2.1.2 池化層(Pooling Layer) 4](#_Toc536176295)

[2.1.3 激活函數(Activation Function) 5](#_Toc536176296)

[2.2 生成對抗網路(Generative Adversarial Network) 5](#_Toc536176297)

[2.3 超分辨率對抗生成網路(Super Resolution using Generative Adversarial Network) 7](#_Toc536176298)

[2.4 殘差網路(Residual Network) 8](#_Toc536176299)

[2.5 Wasserstein生成對抗網路(Wasserstein GAN，WGAN) 9](#_Toc536176300)

[2.6 梯度懲罰(Gradient Penalty) 9](#_Toc536176301)

[2.7 結構相似性(Structural Similarity) 10](#_Toc536176302)

[2.8 多尺度結構相似性(Multi-scale Structural Similarity) 11](#_Toc536176303)

[2.9 縮放卷積層(Resize Convolutional Layer) 12](#_Toc536176304)

[第三章 研究方法 13](#_Toc536176305)

[3.1 系統架構 13](#_Toc536176306)

[3.1.1 特徵擷取及表示(Feature extraction and representation) 14](#_Toc536176307)

[3.1.2 非線性映射(Non-linear mapping) 15](#_Toc536176308)

[3.1.3 像素重建(Reconstruction) 15](#_Toc536176309)

[3.1.4 損失函數(Loss Function) 15](#_Toc536176310)

[3.1.5損失逆傳導(Loss Backpropagation) 15](#_Toc536176311)

[3.2 資料集與參數設定(Setting) 16](#_Toc536176312)

[3.2.1 資料集(Dataset) 16](#_Toc536176313)

[3.2.2 參數設定與輸入輸出(Parameters and Input/Output) 17](#_Toc536176314)

[第四章 實驗結果 18](#_Toc536176315)

[4.1 系統環境及實作細節(Experiment of Environment) 18](#_Toc536176316)

[4.2 實驗結果與分析(Results and Analysis) 18](#_Toc536176317)

[4.2.1 比較SRGAN、Wasserstein-SRGAN、與本文方法 18](#_Toc536176318)

[4.2.2 比較重構方法 19](#_Toc536176319)

[第五章 結論 21](#_Toc536176320)

[參考文獻 22](#_Toc536176321)

# 圖目錄

[圖 2.1卷積運算 3](#_Toc536176322)

[圖 2.2最大池化(Max Pooling)與平均池化(Mean Pooling) 4](#_Toc536176323)

[圖 2.3殘差塊(Residual block) 8](#_Toc536176324)

[圖 2.4改善前後權重分佈模型 10](#_Toc536176325)

[圖 2.5多尺度結構相似性(Multi-scale Structural Similarity) 11](#_Toc536176326)

[圖 2.6棋盤效應(Checkerboard artifact) 12](#_Toc536176327)

[圖 3.1生成網路架構 13](#_Toc536176328)

[圖 3.2對抗網路架構 14](#_Toc536176329)

[圖 3.3損失逆傳導流(Loss Backpropagation Flow) 16](#_Toc536176330)

[圖 4.1輸出結果(1) 19](#_Toc536176331)

[圖 4.2輸出結果(2) 20](#_Toc536176332)

[圖 5.1未來改善架構 21](#_Toc536176333)

# 表目錄

[表4.1實驗結果(1) 18](#_Toc536176334)

[表4.2實驗結果(2) 19](#_Toc536176335)

# 第一章 緒論

## 1.1 研究動機

照片解析度在硬體方面取決於感光元件以及鏡頭的特性，但是隨著演算法的發展，軟體的方法超分辨率(Super Resolution)，可將低解析度影像增強成一張較佳解析度的圖像。通常會分成：1、參考自己內部像素來生成較高解析度圖像細節的部分(Internal Sample)，例如大家所熟知的內插法(Interpolation)，亦或是2、參考多張外部圖像(External Sample)，再產生對應原圖紋理輪廓的細節，例如用卷積層類神經網路，來訓練多張原圖與高解析度圖像之間的關聯性，訓練完成之後，即可以模擬對應高解析度時的細節，例如早期的超分辨率卷積層神經網路(Super Resolution Convolutional Neural Network，SRCNN)[1]，以及近年來的超分辨率對抗深層網路(Super Resolution Using Generative Adversarial Network，SRGAN) [2]。

## 1.2 研究方法

本文先將圖像資料之低解析度與高解析度給予網絡訓練，學習其中關聯性。網路架構基於SRGAN架構實作，對抗生成網路主要分成生成器(Generator)，與判別器(Discriminator)。生成器的目標在於由低解析度圖像來產生一張以假亂真的高解析度圖像，判別器則是要判別出圖的真偽，擔任一個評分員，來協助生成器提升產生高解析度圖像的效果。

# 第二章 研究背景

在本章節將會介紹卷積神經網路(Convolutional Neural Network，CNN)，以及以由卷積神經網路衍生的對抗生成網絡(Super Resolution Using Generative Adversarial Network，SRGAN)。

## 2.1 卷積神經網路(Convolutional Neural Networks)

卷積層神經網絡(Convolutional Neural Network，CNN)，因為其卷積層的特殊性，對於找出圖像空間裡的關聯性十分有幫助，而像素之間的關聯性通常只會分佈在其周圍，距離較遠的像素，其關聯性相對較低，故不需要以全連接層的方式來處理圖像問題，可以大大的降低演算法的時間複雜度(Time Complexity)與空間複雜度(Space Complexity)，所以普遍被用於影像的相關問題上。

CNN基本上可以分為下面兩個部分:(1)特徵擷取器、(2)分類器。特徵擷取器主要由多層卷積層(Convolutional Layers)，穿插著池化層(Pooling Layers)以及激活函數(Activation Function)所構成，主要用來對輸入圖像特徵的擷取。分類器大部分為全連接層神經網絡(Fully Connected Neural Network，FNN)，用於對前面提取的特徵做關聯性或是分類。也能夠單純當成特徵其取器，用特定的特徵來做為一種資料表示型態，類似於線性代數中的基底向量，稀疏編碼(Sparse Coding)的概念。

### 2.1.1 卷積層(Convolutional layer)

卷積層(Convolutional Layer)，影像處理中使用核(Kernel)以滑動視窗(Sliding Windows)的方式來對圖像做處理的原理是一樣的，但其大多在滑動視窗時會重疊(Overlap)部分區域的方式。目的在降低資訊流失。但實際上滑動視窗的步伐(Stride)大小，仍要依照實際用途來做較佳的調整。圖2.2為卷積層的動作圖解，輸入為7 x 7圖像(I)，運算的核(K)為3 x 3，以滑動視窗步伐為1的方式，對於被核涵蓋到的圖像像素範圍，作矩陣元素積(Matrix Element Wise Multiplication)，輸出為5 x 5的特徵圖(I\*K)。

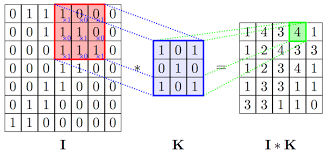


圖 2.1卷積運算

在卷積層中核內的每個值，為神經網絡必須去學習的權重，每個學習出來的核都會成為一個濾波器(Filter)，將輸入與輸出之間的映射關係，所需要的特徵擷取出來，輸出也就是特徵圖(Feature Map)。

### 2.1.2 池化層(Pooling Layer)

池化層的目的主要用為減少資料量(Down-sampling)，保留大部分比較重要的資訊，有效的計算量減少。但是此運算的壞處在於可能會失去某些訊息，所以必須對於前面卷積層的參數及輸出目標做適當的搭配。亦可以用於各層之間輸出與輸入維度耦合的調整。池化運算一般可以分為兩種：(1)最大池化(Max Pooling)、(2)平均池化(Mean Pooling)，運算的方式也是利用池化核與圖像交集的區域，做選出區域內最大值做為代表值或求區域內各個元素平均值的運算，詳細運算方法如圖2.3所示輸入為4 x 4矩陣，池化核為2 x 2，步伐為2。

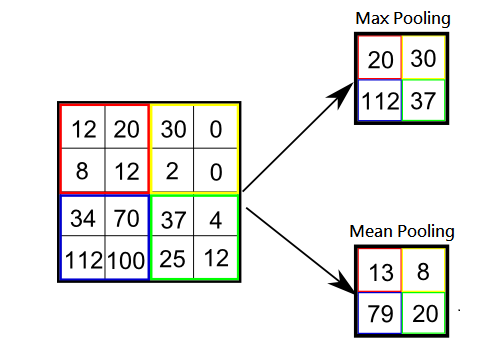


圖 2.2最大池化(Max Pooling)與平均池化(Mean Pooling)

### 2.1.3 激活函數(Activation Function)

神經網絡像前傳遞時,將輸入透過與權重相乘在與偏值相加,此時必須透過激活函數來做正規化輸出值，因為前面所做的運算屬於線性函數轉換當我們所需要求得的問題函數十分的複雜，線性函數的轉換已經無法滿足時，激活函數就扮演很重要的角色，它能使網路對於複雜問題有更好的擬合能力，通常常用的激活函數有下列幾個S函數(Sigmoid function)，其輸出範圍介於[0，1]、雙曲正切函數(Hyperbolic tangent function)，其輸出範圍介於[-1，1]，以及線性整流函數(Rectified Linear Unit，ReLU)則為輸入值小於0輸出為0；輸入值大於0則輸出為輸入值。以及ReLU的變形，帶泄露線性整流(Leaky Rectified Linear Unit，Leaky ReLU)，不會完全濾掉負值域的輸入，是一種具有正負極(bipolar)輸出值的函數。每種激活函數都有其較佳的使用情境，例如ReLU對於較深層的神經網絡有較好的收斂速度，比較不會有梯度消失的問題，傳統的S函數與雙正切函數的連續性是早期因為類比IC建構神經網路條件下的產物。輸出值域具有正負極(bipolar)的激活函數，因為有正有負，存在負數值的懲罰機制，相對於只有正值輸出的激活函數使網路有較快的收斂能力，使用得宜能夠減少訓練的時間。

## 2.2 生成對抗網路(Generative Adversarial Network)

對抗生成網路(Generative Adversarial Network，GAN)[3]是近年來受到熱烈討論及研究的網絡架構，其特殊的地方在於，它是由兩個神經網絡所組成，分別為生成網絡(Generator)，另一個為判別網絡(Discriminator)，就傳統的類神經網路而言，通常是要學習一對一，或是多對一的函數映射，生成對抗網路則是學習目標的機率分佈模型，固其有一般函數沒有的一對多映射的特性，此種特質更接近智能生物機制。

起始生成網路隨機產生一些輸入(Random Input)，接者透過這個隨機輸入z，由生成網路函數G，產生一些輸出G(z)，再將此組輸出與一些真實資料(Ground Truth)，與其對應的真實(Truth，T)數據和生成網絡所產生的虛假(Fake，F)數據做標籤，輸入給判別網絡，做判別真偽的訓練，訓練過程中生成網路的目標就是努力的生成出能夠騙過判別網絡的虛假資料，反之判別網路則是努力的判別出生成網絡產生的假資料，兩個網路形成了一個動態的博弈過程，在比較理想的狀態下是G和D形成D(G(Z))=0.5，也就是生成網絡產生的資料，判別網路認為它是T和F的機率各半，此時生成網路有較好的效果。x表示真實數據，z表示輸入生成網絡的隨機參數，E表示事件期望值，Loss的計算如下式，log(1-D(x))表明對於判別網路必須極大化此式的兩部分：(1)判別網路對真實數據判別為真、(2)log(1-D(G(z))則是將生成網絡生成的數據判別為假的期望值。反之對於生成網路必須最小化(2)。訓練完成之後，生成網絡就是一個可以隨機產出我們所希望接近真實資料分布的虛假資料。最佳化函數如下式：

( 1 )

生成對抗網路訓練時期實際的損失函數(Loss function)計算分為兩個部分組成:

1. 生成器損失(Generator loss):
   1. 生成器所產生的資料與真實資料的誤差值。
   2. 來自判別器對生成器的矯正。
2. 判別器損失(Discriminator loss)：訓練判別器判斷真偽，用來矯正判別器的能力。

## 2.3 超分辨率對抗生成網路(Super Resolution using Generative Adversarial Network)

超分辨率對抗生成網路(Super Resolution Using Generative Adversarial Network，SRGAN)基本原理與GAN類似，分為生成網絡(Generator，G)與判別網路(Discriminator，D)。生成網路是我們的目標，其用於學習input與output圖像低解析度與高解析的映射關系，也可以說是學習圖像增強方法的分佈模型，判別器則是用於判斷生成器產生圖像的真偽，藉由增強判別器的能力，協助生成器調整，使生成器產生更佳的高解析度圖像。

對於生成網絡部分，損失(Loss)的計算由兩部分所組成：(1)生成器的內容損失(Generator Content Loss)、(2)由判別器給予矯正生成網路的損失(Generator Discriminator Loss)。其中(1)又有兩個計算損失的機制，其中一個為G所產生的高解析度圖像(Super Resolution，SR)與真實圖像(Ground Truth，GT)的均方差(Mean Squared Error，MSE)；另一部分為分別將SR與GT送入預先訓練好，且去除後面的全連接層的CNN架構VGG19[4]模型，由最後幾層的卷積層所產生的SR特徵圖與GT的特徵圖再利用MSE來計算出Loss。

判別網絡部分的損失，則為單純的計算SR與GT，與其標籤所產生的誤差，計算二進制誤差(Binary Code Error)。

## 2.4 殘差網路(Residual Network)

殘差網路[5]是將統計學殘差(Residual)概念的運用，套用於神經網絡架構上。隨著問題越來越複雜，網路架構也逐漸往更深、更多層的方向發展，在某些情境下，對於問題所需求的網路能力不確定時，預先網路建構得很深，讓網路在訓練的時候透過學習決定所需要使用的層數，是一種相對通用型的架構。

殘差網路是由多個殘差塊(Residual block)所組成(如圖)，主要概念是源自統計學，這裡是意圖使網路學習輸入與輸出的殘差(residual)，減少一些網路的學習負荷，縮減網路的寬度。塊內包含多個神經網絡層，通常需要兩層以上並包含非線性的激活函數，如此單一個塊才會有較好的非線性映射能力，否則單單輸入與權重(weight)之乘積，再加上偏值(bias)，如果只依靠一個線性函數，對於非線性問題之輸入與輸出關係的映射能力也較差。

殘差塊表示如圖。x為輸入，F(x)為代表此區塊的映射函數之輸出，x identity為捷徑(shortcut)，可以想成如果訓練映射函數時不需要用到某些層數時，可以繞過(bypass)掉某些層，或是可以調整需要多少上層輸出資訊，與本層輸出做混合，保留一些先前的資訊，將更多資訊像後傳遞，利於特徵的提取或提升網路的能力。

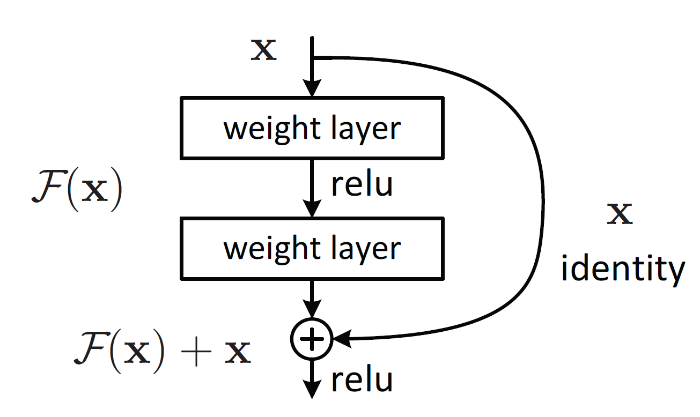


圖 2.3殘差塊(Residual block)

## 2.5 Wasserstein生成對抗網路(Wasserstein GAN，WGAN)

為了改善對抗生成網路，判別器訓練良好的情況下所產生梯度消失(gradient vanishing)的問題，使判別器對生成器的矯正力減弱，甚至消失，造成生成器效果減弱。傳統的GAN是利用最小化生成器所生成的資料與實際資料的KL距離(Kullback-Leibler Divergence，KL Divergence)，由於必須最大化判別器的判別能力，又必須最小化生成器與真實資料的距離，造成類似雙人博弈的現象，總體的輸贏是固定的，故使網路較難收斂。Wasserstein GAN[6]在此做了一些調整，導入Wasserstein距離(Wasserstein Distance)，並移除判別器末端的Sigmoid 函數，避免損失反向傳遞時落在S函數為分為0的區域，造成梯度消失。並不再讓判別器判斷圖像的真偽，進而讓判別器變成用來對輸入的圖像做評分。利用權重修剪(weight clipping)來限制判別器的能力，並取消對判別器初始化的訓練。使判別器不會因為能力太強導致梯度消失，損失無法傳遞到生成器。

## 2.6 梯度懲罰(Gradient Penalty)

權重修剪的問題在於，權重限制的值域範圍十分狹小，範圍太大容易造成梯度爆炸(gradient explosion)，太小又會導致梯度消失(gradient vanishing)，而對權重的限制會導致權重值大多分佈在限制值的兩個邊界，使判別器退化成一個二分函數(binary function)，大大降低判別器的能力，輔助生成器的訓練效果較差。故由Ishaan Gulrajan等人。提出了Improved Training of Wasserstein GANs(Wasserstein GAN GP)[7]。文中的梯度懲罰(gradient penalty)，透過對梯度的限制改善權重修剪所產生的問題。主要概念是由利普希茨連續(Lipschitz continuity)而來， ，將梯度限制在不超過K，使權重值有較常態的分佈，如圖2.4。有效改善弱化判別器的問題。

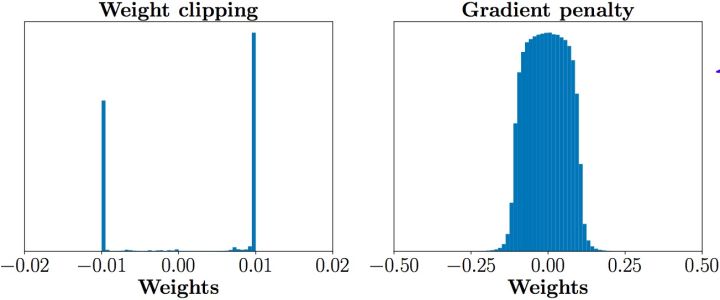


圖 2.4改善前後權重分佈模型

## 2.7 結構相似性(Structural Similarity)

用來衡量兩張數位影像相似度的指標，當兩張圖像，分別為失真圖像與無失真圖像，透過結構相似性可以衡量出失真圖像的品質。

影像資料是存在高度的結構化，在像素之間存有強烈的關聯性，這樣的關聯包含了物件的結構資訊，人類的視覺已經習慣自然的擷取出結構性的資訊，所以結構相似性(Structural Similarity)[8]相較於峰值信噪比(Peak Signal to Noise Ratio， PSNR)更能夠符合人眼對影像品質的直覺。給定圖像x與y分別代表兩張圖像的像素值矩陣，兩者的結構相似性定義為下列式， l(x,y)代表x和y像素的亮度值(Luminance)，c(x,y)代表兩者對比度(Contrast)，s(x,y)代表兩者的結構(Structure)，α、β、γ用於調整l、c、s的重要參數，其值必須大於0；。公式如下列式：

( 1 )

( 2 )

( 3 )

( 4 )

( 5 )

## 2.8 多尺度結構相似性(Multi-scale Structural Similarity)

多尺度結構相似性(Multi-scale Structural Similarity)[9]，通常用於圖像評估方法，是結構相似性(Structural Similarity)的延伸，用多個不同尺度的相同一張失真訊號，多個不同尺度的真實訊號，做結構相似性演算法，在賦予各個尺度不同權重，依據多個尺度共同得出結構的相似程度。演算法如圖2.5。

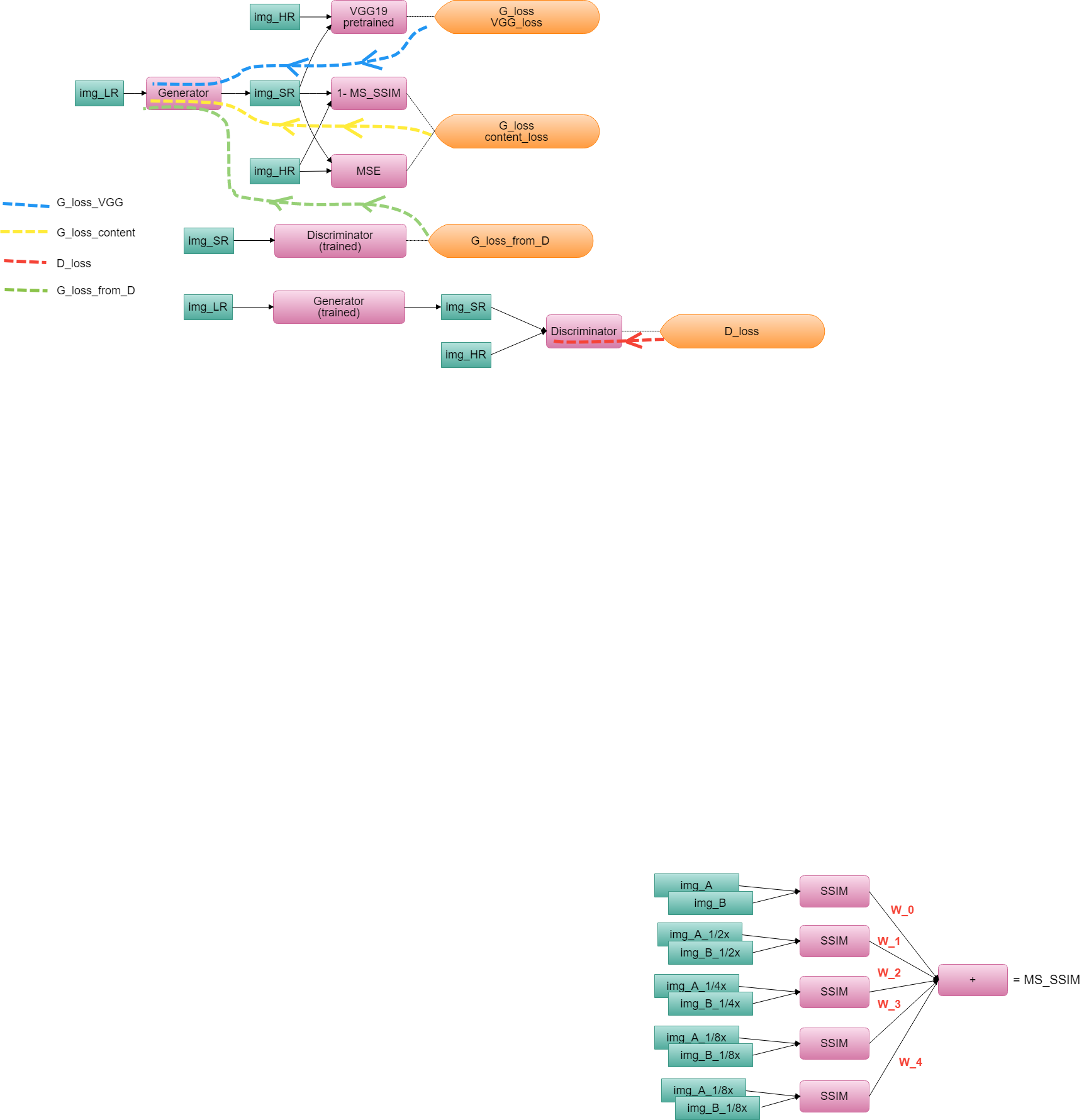


圖 2.5多尺度結構相似性(Multi-scale Structural Similarity)

## 2.9 縮放卷積層(Resize Convolutional Layer)

縮放卷積層常用來取代反卷積層，主要是因為反卷積重建出來的圖像容易有棋盤效應(Checkerboard artifact)如圖2.5，這種現象就是在重建出來的圖像會出現較多且較清楚，類似西洋棋盤狀的紋路。縮放卷積層基本上是用內插法(Interpolation)將輸入圖像或特徵圖放大，取代反卷積層搭配的反池化層(Un-pooling layer)，再用卷積運算來做更進一步的圖像修正。

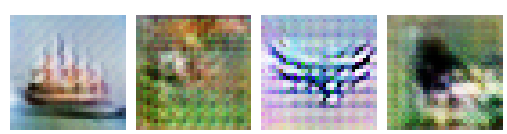


圖 2.6棋盤效應(Checkerboard artifact)

# 第三章 研究方法

超分辨率對抗生成網絡(Super Resolution Using Generator Adversarial Network，SRGAN)為GAN在超分辨率上的應用實作，本篇利用此架構為基礎，在計算內容損失方面，使用結構相似性(Structural Similarity，SSIM)與多尺度結構相似度(Multiscale SSIM)，原本的均方誤差(Mean Square Error，MSE)並用。並導入Wasserstein距離，替代原先的KL距離，改善雙人博弈現象，並加入梯度懲罰，改善權重修剪的缺陷。訓練完成後的SRGAN中生成網絡的部分，便能夠學習到低階析度與高解析度之間的映射關係。

## 3.1 系統架構

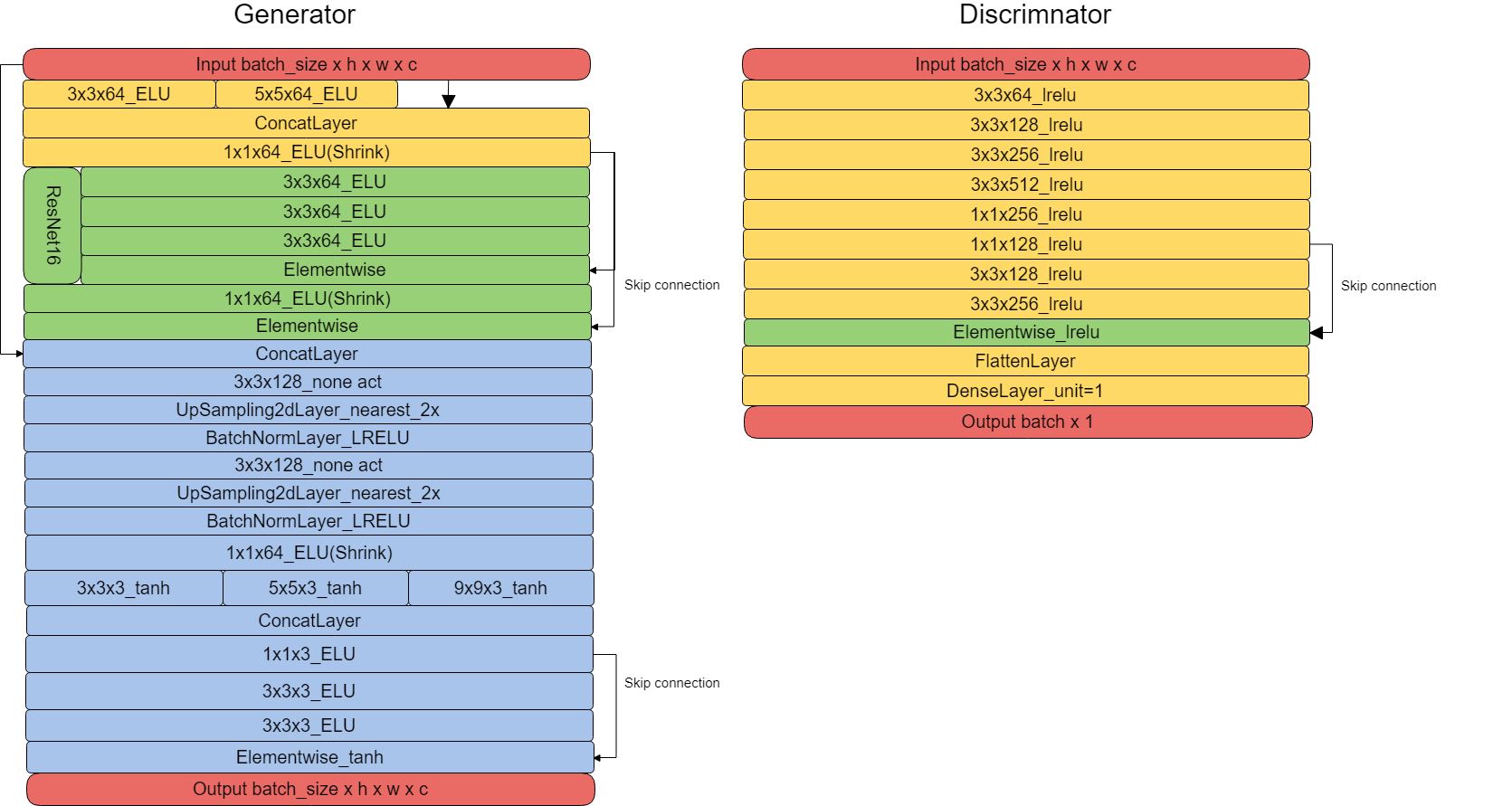


圖 3.1生成網路架構

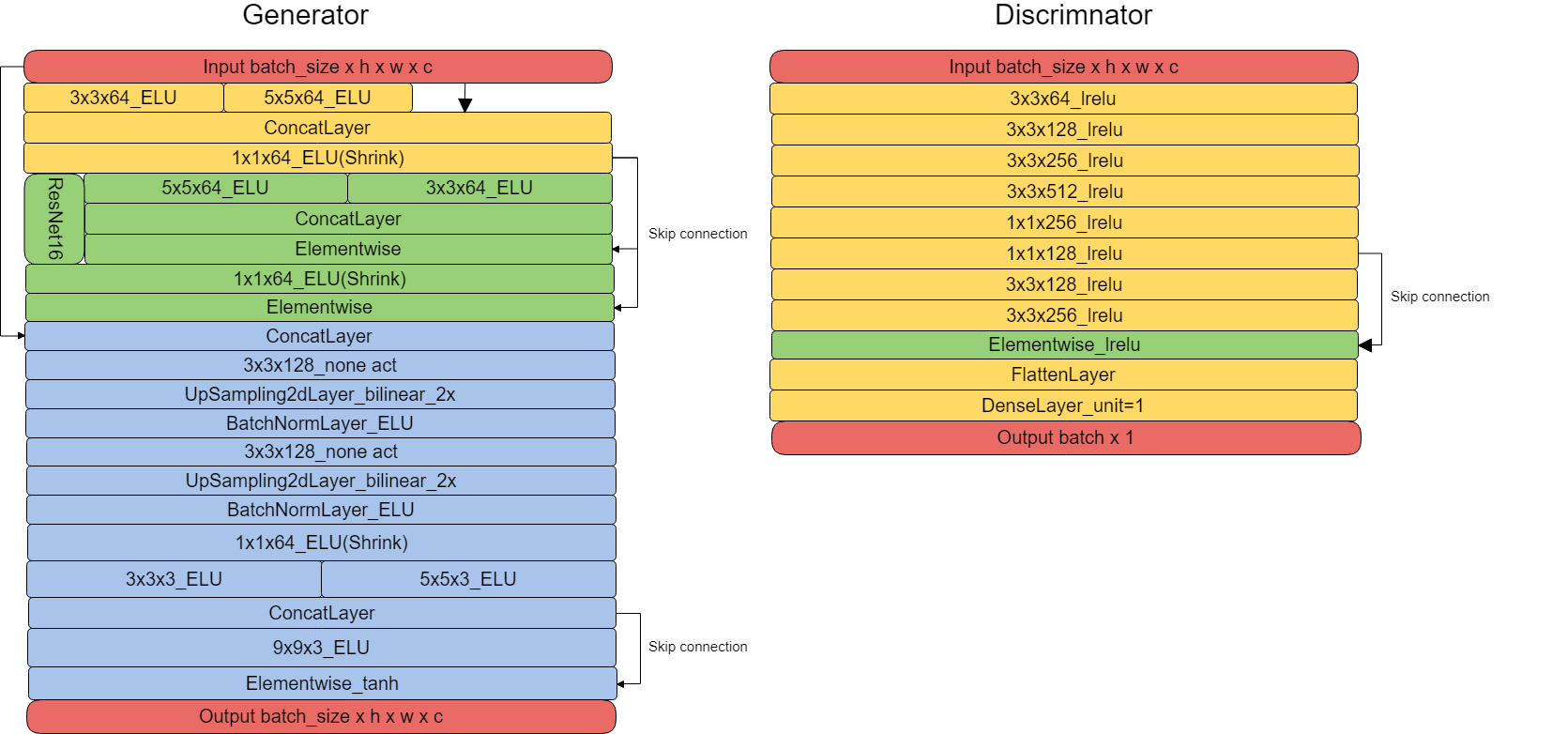


圖 3.2對抗網路架構

本篇作法沿用SRGAN的架構，並對於生成網絡與損失函數部分做些修改與調整，並交叉比對各個組合的效果，由結果分析各種原因。架構主要分為三大部分：(1)特徵擷取及表示(Feature extraction and representation)：此部分用於提取低解析度的特徵圖做為目標函數的輸入。(2)非線性映射(Non-linear mapping)：低解析度的特徵圖與高解析度之間的非線性映射。(3)像素重建(Reconstruction)：將由低解析度提取出的特徵經過非線性映射學習高解析度的像素塊，在此合成高解析度的圖像。

### 3.1.1 特徵擷取及表示(Feature extraction and representation)

本文提取特徵代表，用兩個並聯式卷積層，分別使用3x3與5x5的卷積核，主要原因是希望能夠用不同的視覺感受區域(Receptive field)，獲取更豐富的特徵。將前兩所獲取的特徵堆疊，並且使用一個帶有1x1卷積核的卷積層，用來壓縮這些特徵，一方面提高特徵濃度，另一方面可以節省一些運算量。

### 3.1.2 非線性映射(Non-linear mapping)

這部分使用一個16個塊的殘差網路，主要想法是企圖讓殘差塊學習產生高頻的訊號所需要的映射關係，並組合高低頻訊號特徵塊，生成高解析度圖像所需要的特徵塊，利於重建高解析度圖像。 使用3x3與5x5兩種大小卷積核，做為不同視覺感受野的特徵映射。

### 3.1.3 像素重建(Reconstruction)

原始SRGAN建構部分是使用像素洗牌卷積(Pixel-shuffle Convolution)，重建出高解析度的圖片，本文將使用縮放卷積(Resize Convolution)來取代原先的重構方法。

### 3.1.4 損失函數(Loss Function)

分別使用均方差(Mean Square Error)、結構相似性(Structural Similarity)、多尺度結構相似性(Multi-scale Structural Similarity)，做為內容損失。

### 3.1.5損失逆傳導(Loss Backpropagation)

損失(Loss)基於反向傳導(Loss Backpropagation)演算法，如圖3.3，向前傳遞到前面的層數，訓練網絡時對各層權重的更新。分成下列幾條路徑：

1. G\_loss\_VGG：表示來自於預先訓練好的VGG19損失函數用於對生成器(Generator)權重的修正。
2. G\_loss\_content：表示來自均方差或是結構相似性損失函數對生成器權重的修正。
3. D\_loss：表示對於判別器(Discriminator)權重的修正。
4. G\_loss\_from\_D：訓練一段時間的判別器對生成器的修正。

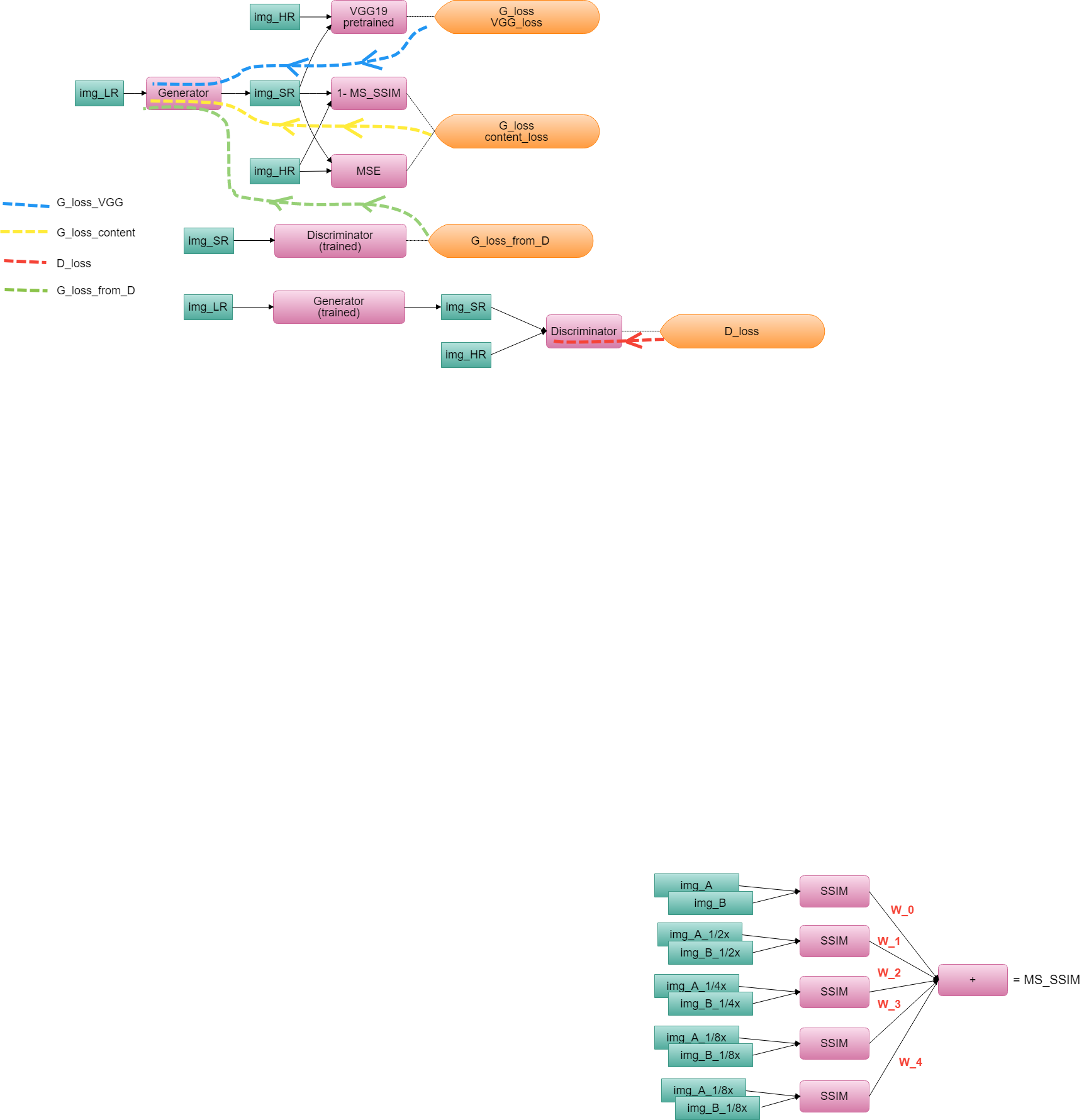


圖 3.3損失逆傳導流(Loss Backpropagation Flow)

## 3.2 資料集與參數設定(Setting)

### 3.2.1 資料集(Dataset)

使用史丹佛大學提供的STL10資料集，選擇卡車與船的部分資料作為訓練及測試資料，訓練集包含896張圖像，評估與測試集共用包含16張圖像。STL10資料集原始大小為96x96，我們在96x96的圖像上，隨機擷取出48x48的高解析度圖像做為真實資料(Ground truth)，接著將48x48的真實資料做雙三次式(Bi-cubic)縮小演算法，至12x12，做為低解析度(Low resolution)的輸入圖像。

### 3.2.2 參數設定與輸入輸出(Parameters and Input/Output)

1. 訓練圖像輸入大小(input size) : 12x12
2. 訓練圖像輸出大小(output size) : 48x48
3. 批次訓練檔大小(batch size) : 16
4. 學習率(learning rate) :
5. 學習率衰減比率(learning rate decay) :
6. 學習率衰減週期(learning rate decay epoch) : 2500
7. 訓練週期(training epoch) : 5000

# 第四章 實驗結果

## 4.1 系統環境及實作細節(Experiment of Environment)

在整個程式系統上，電腦設備使用的CPU為八代Intel Core i7 CPU 4.70-GHz、作業系統為64位元Windows 10、實作平台為Tensorflow版本為0.12.0-rc1。

## 4.2 實驗結果與分析(Results and Analysis)

### 4.2.1 比較SRGAN、Wasserstein-SRGAN、與本文方法

比較SRGAN、Wasserstein-SRGAN、與本文方法，以常用的圖像品質(Image Quality Assessment)，分別為峰值信噪比(Peak Signal to Noise Ratio， PSNR)、結構相似性(Structural Similarity)、多尺度結構相似性(Multi-scale Structural Similarity)、梯度幅度相似度偏差(Gradient Magnitude Similarity Deviation)，來衡量生成圖像品質。結果如表4.1。

表4.1實驗結果(1)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **PSNR** | **SSIM** | **MS-SSIM** | **GMSD** | **Execution time** |
| SRGAN | 19.1321 | 0.5798 | 0.9150 | 0.1481 | 0.8196s |
| W-SRGAN | 19.4785 | 0.6065 | 0.9184 | 0.1428 | 0.8777s |
| My Work | 19.5590 | 0.6352 | 0.9289 | 0.1394 | 0.8795s |

本文主要比較生成之圖片與原始高解析度圖像之間的結構相似性(Structural Similarity)、多尺度結構相似性(Multi-scale Structural Similarity)、主要因為此兩種所衡量影像品質標準，較接近人類視覺系統的自然感知。在這兩個影像評估數值上，本文皆優於之前的工作，但是因為架構稍微大些，使執行速度上略有落後。因為使用結構相似性系列來計算內容損失，與VGG損失的相依關係較大，在相同的訓練週期下，收斂的速度較均方差慢，如有更多時間訓練，效果應該會更佳。均方差方法在約1000個週期數時，網路收斂情況良好，損失已經降到理想範圍，就這點來看本文的架構更具有潛力。結果如圖4.1。



圖 4.1輸出結果(1)

## 4.2.2 比較重構方法

基於本文方法，比較兩種重構方法，分別為像素洗牌卷積(Pixel-shuffle Convolution，PS conv.)與縮放卷積(Resize Convolution，Resize conv.)，結果如表4.2。

表4.2實驗結果(2)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **PSNR** | **SSIM** | **MS-SSIM** | **GMSD** | **Execution time** |
| PS conv. | 19.346 | 0.6286 | 0.9219 | 0.1494 | 0.7620s |
| Resize conv. | 19.559 | 0.6352 | 0.9289 | 0.1394 | 0.8795s |

相較兩種重構方法，各有優點。縮放卷積在重構圖像的效果略勝像素洗牌卷積，這裡猜想可能是前面層數的卷積核數量不足，無法形成足夠的特徵碎塊讓像素卷積重構。縮放卷積則是很直覺地將特徵圖放大，再進行調整。但是執行時間是像素卷積較好，因為特徵碎塊資料重塑形(reshape)，一定比內差法(interpolation)放大特徵圖來的快。



圖 4.2輸出結果(2)

# 第五章 結論

結構相似性(Structural Similarity)較符合人類視覺感知標準，此方法有繼續延伸的價值，但本文的架構仍然有改善的空間，對於使用預先訓練好的VGG19作為計算損失函數的工具是之前論文了不起的智慧，但是我們不確定此方法與均方差損失函數或是結構相似性損失函數之間，是否存在相依關係亦或是互斥性，如此一來在訓練時，對於網絡收斂速度及能力，可能會有不良影響，本文認為可以將預先訓練好VGG19的卷積層第一層加入本文架構中，與特徵擷取及表示(Feature extraction and representation)的部分並聯，強化細部特徵圖效果，使後面的網路有更佳特徵碎塊，用於映射和重構，強化生成網路能力，並去除原本VGG19損失函數，減低損失函數之相依性所產生的影響。結構如圖5.1。

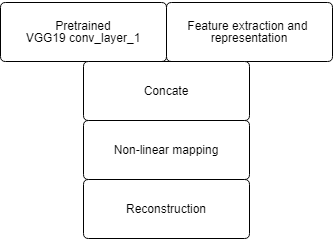


圖 5.1未來改善架構

# 參考文獻

1. Chao Dong, Chen Change Loy, Kaiming He, Xiaoou Tang."Image Super-Resolution Using Deep Convolutional Networks" Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)， 2015.
2. Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, Wenzhe Shi. “Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network” Computer Vision and Pattern Recognition (CVPR), 2017.
3. Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio. “Generative Adversarial Networks” 2014
4. Karen Simonyan, Andrew Zisserman. “Very Deep Convolutional Networks for Large-Scale Image Recognition” Pattern Recognition (ACPR), 2015.
5. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun “Deep Residual Learning for Image Recognition” Computer Vision and Pattern Recognition, 2015.
6. Martin Arjovsky, Soumith Chintala, Léon Bottou. “Wasserstein GAN” arXiv.org Machine Learning (stat.ML); Machine Learning (cs.LG) 2017.
7. Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, Aaron Courville ” Improved Training of Wasserstein GANs ” arXiv.org Machine Learning (cs.LG); Machine Learning (stat.ML) 2017.
8. Zhou Wang, Member, Alan Conrad Bovik, Fellow, Hamid Rahim Sheikh, Student Member, and Eero P. Simoncelli. ”Image Quality Assessment: From Error Visibility to Structural Similarity” IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600−612, Apr. 2004.
9. Zhou Wang, Eero P. Simoncelli and Alan C. Bovik. “Multi-scale Structural Similarity for Image Quality Assessment” The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003.