

# Effect of Face Super-Resolution with Degradation Network

## 退化網路對臉孔超解析的影響

組別: A37 組員: 徐浩宇、曾泓硯 指導教授: 孫民

### Abstract

超解析 super resolution (SR)，是一種能夠將低解析度影像擴增到高解析度影像的方法。在這次的專題研究中，我們將此方法應用到人臉的影像上，將低解析度人臉圖片轉成高解析度圖片，並以臉孔特徵復原為主要目標。

我們參照《FSRNet: End-to-End Learning Face Super-Resolution with Facial Priors》這篇論文的架構去建立模型，並嘗試重現論文的結果。同時，我們會提出多種改善模型的方法，並探討不同實踐方法的優劣。

### Introduction

FSRNET 架構:

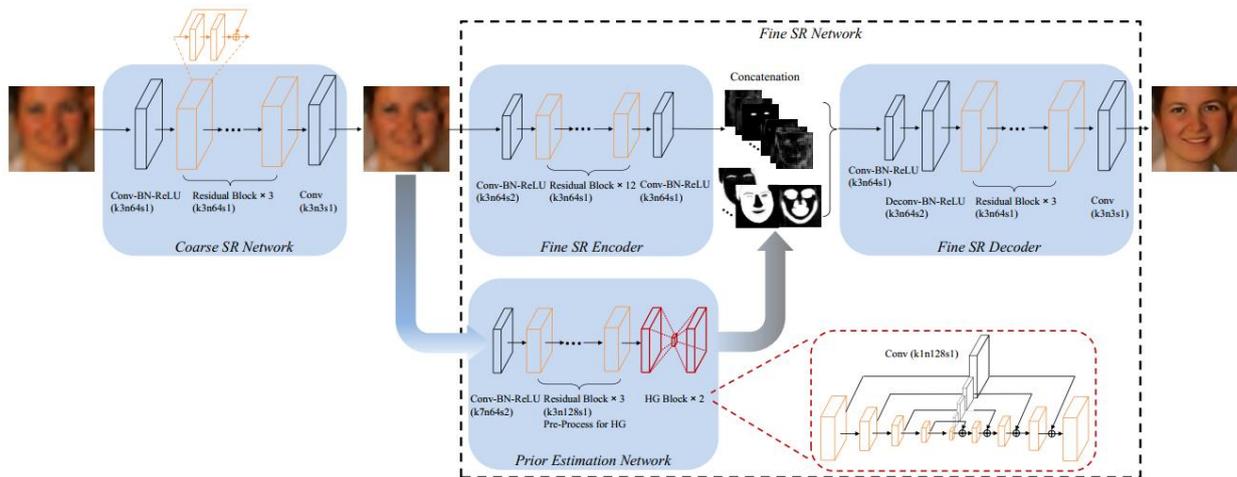


圖 1 FSRNET 的整體架構圖

#### A. Coarse SR Network:

將將低解析度照片(LR)放入到此網絡中，先初步進行特徵的提取，以利後面 Prior Estimation Network 進行臉官的切割 (parsing maps)。

#### B. Fine SR Encoder:

利用 Residual Block 的架構去對 LR 做特徵提取，得到 64x64 的臉部特徵( face embedding )。

#### C. Prior Estimation Network:

除了使用 Residual Block 的架構去保留每一層臉部資訊以外，最後面加上了 Hour-Glass (HG) structure 去產生 parsing maps。

#### D. Fine SR Decoder:

將 face embedding 以及 facial parsing maps 合在一起經過 Fine SR Decoder 做上取樣得到高解析度圖像(HR image)。

FSRGAN 架構:

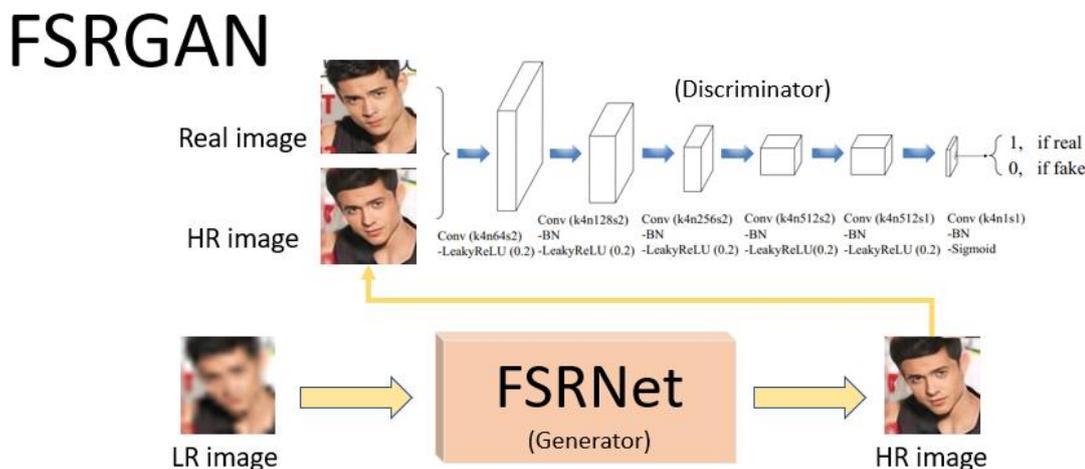


圖 2 FSRGAN 架構圖

將原本 FSRNET 視為一個 Generator  $G$ ，另外增加一個 Discriminator  $C$ 。

我們知道 GAN 在處理 super-resolution 時有很好的效果，GAN 通常能比原本的 MSE-based deep models 產生有更好視覺效果的真實圖片。這邊主要的流程就是使用  $C$  去分辨  $G$  產生的高解析度圖片及真實的高解析度圖片，而  $G$  的目標則是生成越接近真實的高解析度圖片去騙過  $C$ 。

FSRNET with Degrading Network:

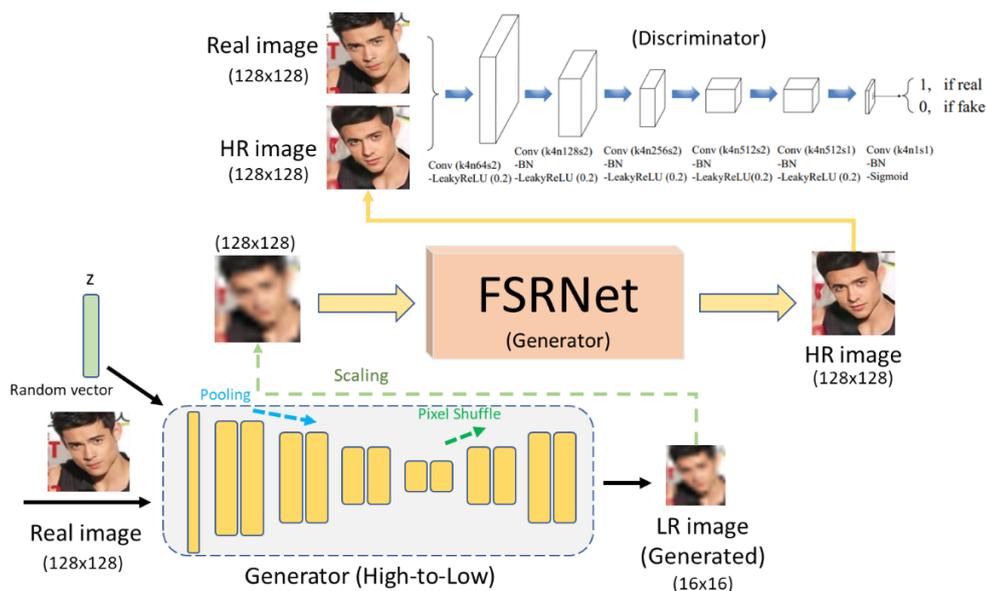


圖 3 加入 Degrading Network 的 FSRGAN 架構圖

主要概念就是讓送進去訓練的照片越接近真實模糊的照片越好，為了達成這個目的，我們額外訓練了一個 Degrading Network [11] 去對高解析度圖片做壓縮。

## Result

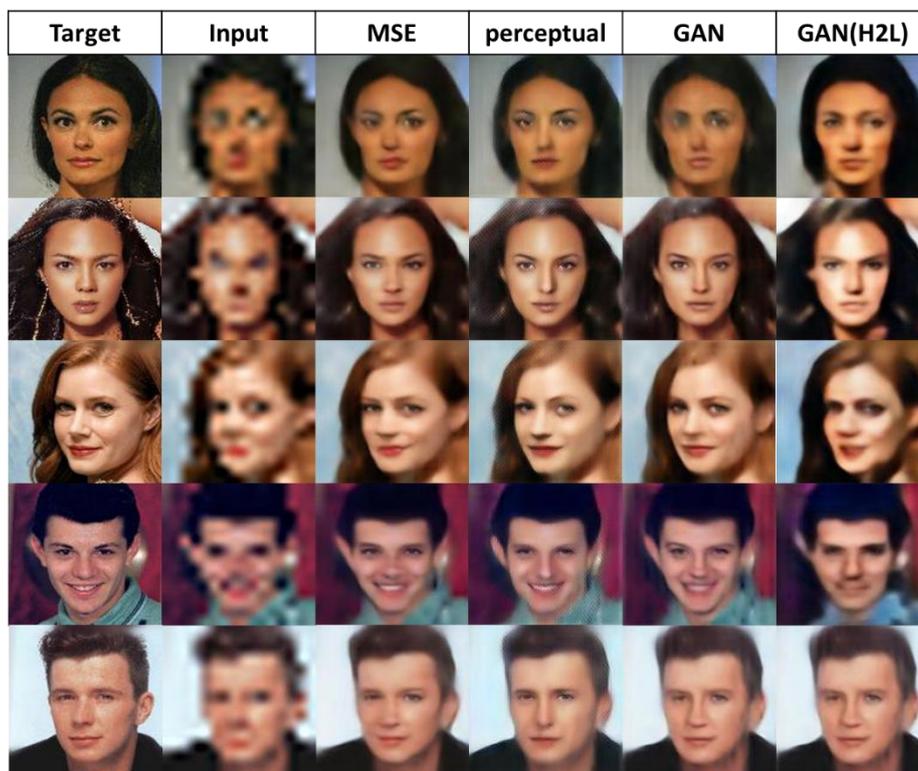


圖 4 各類型 model 復原高解析度圖片比較圖

	FSRNET loss	Perceptual loss ( $\gamma_p = 10$ )	Perceptual loss ( $\gamma_p = 1000$ )	FSRGAN	FSRGAN H2L
PSNR	23.32	23.12	20.64	23.22	19.67
SSIM	0.66	0.625	0.451	0.592	0.49
RMSE	1.4	1.26	1.45	1.30	1.99

表 1 各個方法在訓練集上面的表現

	FSRNET loss	Perceptual loss ( $\gamma_p = 10$ )	Perceptual loss ( $\gamma_p = 1000$ )	FSRGAN	FSRGAN H2L
PSNR	20.58	20.33	17.48	20.4	17.84
SSIM	0.599	0.566	0.441	0.585	0.512
RMSE	1.27	1.22	1.29	1.24	1.85

表 2 各個方法在測試集上面的表現

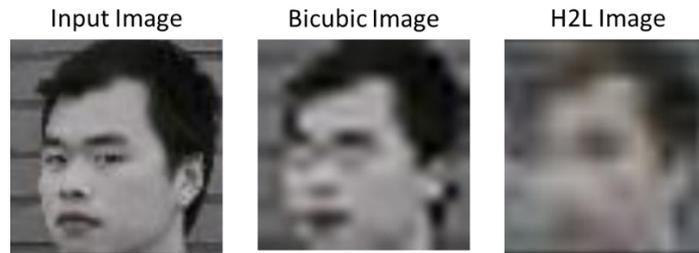


圖 5 Degradation Network 對於圖片壓縮的效果

## Summary

我們的 FSRNET 在訓練集上所做出來的 PSNR 大約在 23 左右，與論文上的 26 相比，還是有一點差距。但是，我們有觀察到其他論文在使用 FSRNET 這個方法時，跑出來的 PSNR 數值跟我們做出來的差不多，所以可以推論我們訓練的 FSRNET 表現還不錯。

此外，我們也比較了加入不同損失函數下的 PSNR 差異，其中原本的 FSRNET loss、perceptual loss 以及 GAN loss 所得出的結果差不多；但是觀察圖片後發現加入 perceptual loss 的圖片高頻訊號較多，放大來看會有雜訊，而 GAN loss 看起來最平順。如果把 perceptual loss 的比例提升的話，圖片看起來銳利度會提高，但雜訊也會變得更明顯。

至於另一個客觀的指標 SSIM 主要是來比較整張圖的結構，而我們的結果大概落在 63%附近，跟論文的 73%相比也是有一段落差，而且可以觀察到 PSNR 跟 SSIM 之間有正相關，SSIM 較高的方法 PSNR 也會比較高。

此外，我們使用了新的評估指標—臉部特徵點的差異(RMSE)，去比較生成圖片對於臉部形狀特徵的復原程度。可看出使用 Perceptual loss 和 GAN loss 的模型表現會略好一些，這個也可以很容易從生成圖片的效果去看出，體現出這個評估標準是比較能反映人眼感官。

我們有嘗試不加入 Prior Estimation Network，在 PSNR 能達到 28 左右的成果，比起輸入圖多了一些高頻的影像，但是五官是模糊的。相較之下，加入了 Prior Estimation Network 對於臉部的重建有明顯的幫助，尤其是臉部五官的細節有明顯增加；儘管如此，背景、頭髮……等其他特徵由於不包含在 Prior Estimation 內，故還原程度低，因此我們認為這個也是降低 PSNR 的原因之一。

最後，討論 FSRNET 加上 Degradation Network 訓練後表現不佳的原因。Degradation Network 除了輸入高解析圖片外，會加上一個 noise vector，理論上能夠產生多樣性高的低解析度影像，也能更貼近真實世界的情形。然而，我們訓練出來的架構雖然能夠產生和 Bicubic 不同的低解析度影像，但是其多樣性過低，因此很難達到我們原本預期的效果。在這樣情況下，新增 Degradation Network 也只是換了生成低解析度輸入圖的架構，並沒有發揮增加訓練集多樣性的優勢，因此是無法應用到真實世界的情況，故我們在 GAN high to low 的表現不盡理想。

## 心得感想

曾泓硯：

超解析一直是一個我想研究的主題，經過這一次專題，我對於加上什麼架構以及損失函數，其對於結果會有什麼影響有了初步的理解；此外，我也透過了這一次的專題更加熟悉了 PyTorch 的架構，也對於如何研究一個 Deep learning model 有了一次完整經驗，但對我來說最大的收穫還是發現了自己研究時的盲點。經過了一年的專題研究，相信自己也算是將這些領域粗略的認識一遍，也希望這次的經驗能夠讓下次研究更加順利，最後謝謝同組的徐浩宇，實驗室的林捷愷學長以及孫民教授的幫忙以及指導。

徐浩宇：

在修這次專題之前，我對於深度學習和機器學習領域的了解不多。但是透過每個禮拜 meeting 和閱讀 paper 的訓練，我慢慢地增加了這個領域基本知識。透過這個專題實作的機會，讓我了解到如何從頭到尾去做模型的訓練，包含：資料前處理、設計架構、評估結果等等，其中最重要的部分就是去培養解決問題的思維，去仔細思考問題成因並解決。雖然這次專題的成果不甚完美，但是認真學到不少東西，也感謝同組組員曾泓硯的合作、實驗室的林捷愷學長及孫民教授的幫忙及指導。