

# 基於卷積神經網路單影像超解析度優化之研究

A Research of Single Image Super Resolution Optimization

Base on Convolution Neural Network

劉遠楨<sup>1</sup> 謝子玄<sup>2</sup>

LIU, YUAN CHEN<sup>1</sup> HSIEH, TZU HSUAN<sup>2</sup>

<sup>1</sup>國立台北教育大學 資訊科學系研究所 教授

<sup>1</sup> National Taipei University of Education Department of Computer  
Science Professor

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

<sup>2</sup>國立台北教育大學 資訊科學系研究所 研究生

<sup>2</sup> National Taipei University of Education Department of Computer  
Science Student

E-mail: s110616041@stu.ntue.edu.tw

## 摘要

隨著深度學習技術的不斷進步，許多卷積神經網路(CNN)的架構被應用在超解析度成像(Super Resolution, SR)任務中，但許多基於卷積神經網路的模型都有非常多的參數與深度，導致模型訓練也變得越來越困難，每個參數較難收斂到最佳值，而無法表現出完整的模型性能。我們重新審視了EDSR，並且針對不同的激勵函數、損失函數以及優化器做適當的評估，分析出不同超參數對於超解析成像的影響，並且在不大幅度修改模型結構的前提下優化模型的訓練，使得模型擁有較高的準確度。

我們透過實驗證實，以最小的幅度修改EDSR仍可以提升模型預測的準確度，並且在不同的縮放因子上皆得到很不錯的準確度。透過SmoothL1損失函數、PReLU激勵

函數、RMSProp優化器，在縮放因子2、3、4，不論是PSNR或SSIM都有正向的性能提升。

關鍵字：類神經網路、超解析度成像、超參數

## ABSTRACT

With the continuous advancement of deep learning technology, many convolutional neural network architectures have been applied in super-resolution tasks, but many convolutional neural network-based models have many parameters and depth of the model make it more and more difficult to train the model, and it is more difficult for each parameter to converge to the optimal value. Most of the research focuses on the design of the model, while ignoring the influence of hyperparameters on the model, resulting in the model not being able to exert the complete ability of the model and spending more time training. In this paper, through numerous experiments, we find out the suitable hyperparameter combination for the super-resolution imaging model, optimize the training of the model without greatly changing the model architecture, and make the model have a high accuracy. Experiments have confirmed that under different magnification factors and different model architectures, our optimized models and training methods have obtained high prediction accuracy of super-resolution imaging in various validation sets.

Finally, we found a set of hyperparameters suitable for super-resolution imaging, using SmoothL1 loss, PReLU excitation function, RMSProp optimizer, and using post-upscaled sub-pixel convolution such a combination can make the same model architecture have more high accuracy.

Keywords: Artificial neural networks、Super resolution、hyperparameters

# 壹、前言

## 一、研究背景與動機

超解析度成像是將低解析度圖片轉換成高解析度圖片的一種圖片修復方式。尤其是隨著數位影像越來越廣泛使用與傳播，超解析度成像更顯得重要。超解析度成像適用於計算機視覺中的各種情境，例如教育、軍事、監控影像、醫學影像、衛星遙測圖像、影像串流、遊戲渲染、古老的影像重建……，任何需要更多影像細節的情景，都可以使用到超解析度成像技術。

超解析度成像可區分成基於時間序列與靜態圖像。本文針對靜態圖像的單張圖片超解析度成像(SISR)進行研究，已經有許多演算法來解決此問題，其中可以區分成基於插值、基於重建、基於機器學習的演算法。從低解析度重構成高解析度的任務有很高的不適應性，由於高解析度圖片可以映射成許多種的低解析度圖片，且低解析度圖片可以是由不同種高解析度圖片所生成，因此我們想進一步優化，使超解析度成像得到更高的預測值。

## 二、研究目的與結果

Lim、Son、Kim、Nah、Lee(2017)提出EDSR，是超解析度成像的一大里程碑，榮獲2017 NTIRE超解析成像最佳論文獎。我們重新審視了EDSR，並且針對不同的激勵函數、損失函數以及優化器做適當的評估，分析出不同超參數對於超解析成像的影響，並且在不大幅度修改模型結構的前提下優化模型的訓練，使得模型擁有較高的準確度。

我們透過實驗證實，以最小的幅度修改EDSR仍可以提升模型預測的準確度，並且在不同的縮放因子上皆得到很不錯的準確度。透過SmoothL1損失函數、He、Zhang、Ren、Sun (2015)提出的PReLU激勵函數、Hinton(2018)提出的RMSProp優化器，在縮放因子2、3、4，不論是PSNR或SSIM都有正向的性能提升。

## 貳、文獻探討

### 一、過去的超解析演算法

在深度學習廣泛應用之前，大多採用插值的方式進行超解析度成像，例如Keys(1981)提出的雙三次插值，其優點是運算速度快，不需要訓練繁重的參數，而缺點是得到的結果通常不如人意。

在深度學習廣泛運用後，Dong、Loy、He、Tang(2014)提出SRCNN，首次性的透過深度學習與卷積神經網路應用在SR上，將低解低度圖片先用插值法放大，再利用極少量的卷積層進行運算，得到的結果已優於基於插值的演算法，也因此開啟了深度學習在超解析度成像領域中的更多研究。Dong、Loy、Tang (2016)又提出FSRCNN，可直接輸入低解析度圖片，再透過反卷積輸出成高解析度圖片。

Kim、Lee、Lee(2016, 2016)提出VDSR與DRCN。前者想法是認為低解析度圖片與高解析度圖片在低頻成分上極大的相似，其模型透過差值先得到目標解析度大小的圖片，再透過二十層的卷積層來學習輸入圖像與輸出圖像之間的差值，最後將輸入圖像與輸出圖像相加得到復原結果。VDSR透過3x3的卷積層與更深層的網路架構並且取得更廣泛的視野，並且結合殘差學習，降低梯度消失帶來的問題；後者透過遞迴的方式重複在相同的捲積層運算，最後再輸出結果。

無論是SRCNN、VDSR或DRCN都是透過差值放大圖像後再進行操作，這樣在進行卷積運算時都會產生較大的運算量。Shi等人(2016)提出ESPCN，輸入低解析度圖片進入模型，輸出前透過子像素卷積將特徵圖的通道重新排列至長度與寬度中，得到一個通道變小而長度寬度變大的輸出成高解析度圖片，並且在卷積的過程中模型可學會如何透過更少的訊號來進行放大影像，將插值法的目的透過卷積來訓練，並且得到更好的成績。

Huang、Liu、Weinberger、Maaten(2017)提出DenseNet，此論文提出稠密連接，透過將不同層級的特徵圖連接起來，達到特徵重複使用的效果，這讓其可以使用更少參數的情況下得到相同的性能。而Tong、Li、Liu、Gao(2017)基於DenseNet並應用於超解析度成像議題上，提出SRDenseNet，透過稠密區塊來重複使用區域特徵，並且最後使用子像素卷積進行上採樣，得到了不錯的成果。

Ledig等人(2017)提出SRGAN，它不同於以往的基於卷積神經網路的模型，是使用Creswell等人(2018)提出的生成對抗網路Generative Adversarial Network(GAN)的架構處理超解析度成像的問題，使用感知損失(perceptual loss)與對抗損失(adversarial loss)來恢復圖片的真實感，不把重點著重在PSNR，也在高放大倍率中得到了一定的細節還原。

Lim、Son、Kim、Nah、Lee(2017)提出了EDSR與MDSR，其認為在中許多層的設計是給高等級的計算機視覺問題所使用，例如分類或偵測的問題，因此他們移除了原始He、Zhang、Ren、Sun(2016)提出的ResNet中的批量標準化層(Batch normalization layer)並加以改良，最後將模型的深度與參數都提升，得到了相當不錯的結果。

## 參、研究方法

### 一、簡化版模型

在本節中，將會描述我們提出的模型修改方式。我們分別使用了Kim等人(2016)的VDSR、Kim等人(2016)的EDSR與Zhang等人(2018)提出的RCAN當作我們實驗基準模型，為了加速實驗流程，所有的模型都降低了通道數量與模型深度，以下將會詳細說明我們實驗中使用的模型參數。

在VDSR中我們使用32個通道數量，並且在殘差部分使用了8層的卷積層進行串聯。雙三次插值使用Python OpenCV的resize函式搭配INTER\_CUBIC參數取得上採樣的圖片當作模型的輸入。優化部分我們修改激勵函數為PReLU、損失函數為SmoothL1、優化器為RMSProp、並且採用後子像素卷積做上採樣。

在EDSR中我們使用32個通道數量，並且使用8個殘差區塊(共16個卷積層)進行串聯。優化部分我們修改激勵函數為PReLU、損失函數為SmoothL1、優化器為RMSProp。

在RCAN中我們使用32個通道數量，並使用3個Residual Group，每個Residual Group包含6個Residual Channel Attention Block。優化部分我們修改激勵函數為PReLU、損失函數為SmoothL1、優化器為RMSProp。

### 二、激勵函數、損失函數、優化器

我們使用了簡化版的EDSR縮放因子為4的模型來測試八個不同的激勵函數，分別是Sigmoid、Tanh、ReLU、LeakyReLU、PReLU、ELU、Swish；也測試了三個

不同的損失函數，分別是 L1損失、L2損失與 SmoothL1損失；另外也測試了六個不同的優化器，分別是 SGD、Momentum、Adagrad、RMSProp、Adam、AdamW。

### 三、虛擬碼

首先我們將圖片分割成高解析度圖片(HR)192\*192大小，與相對應大小的低解析度圖片(LR)，並且設定訓練次數 EPOCHS 設定為300次開始迭代。每次取出圖片時會隨機旋轉與隨機翻轉 HR 與 LR，接著將 LR 減掉 DIV2K 資料集的平均 RGB，並且將 LR 輸入至模型，產生模型預測出的超解析度圖片(SR)並加回 DIV2K 資料集的平均 RGB，再透過 Smooth L1損失計算誤差，最後將產生出的誤差經由 RMSProp 優化器更新製模型參數後不斷重複直到完成迭代。整體的演算法如表1。

表 1 演算法虛擬碼

演算法虛擬碼
<pre> divide the images into HR 192*192 and LR of corresponding size. for epoch from 1 to EPOCHS do   foreach (LR, HR) (images) do     randomly rotate and flip both LR and HR.     Subtract the RGB channels of the LR by the DIV2K RGB average     color channel.     input LR to the model to generate predicted pictures (SR).     add the RGB channels of the SR by the DIV2K RGB average color     channel.     calculate the error between SR and HR through Smooth L1 loss:     <math display="block">\begin{cases} 0.5(HR - SR)^2, &amp;  HR - SR  \leq 1 \\  HR - SR  - 0.5, &amp; \text{otherwise} \end{cases}</math>     update the model parameters through RMSProp optimizer:           </pre>

```


$$v_t = 0.99 * v_{t-1} + 0.01 * \left(\frac{\partial L}{\partial W}\right)^2$$



$$W \leftarrow W - \frac{\eta}{\sqrt{V_t + \epsilon}} * \frac{\partial L}{\partial W}$$


end foreach
end for

```

## 肆、結果與討論

### 一、實驗結果比較

#### (一)不同激勵函數對超解析度成像的影響

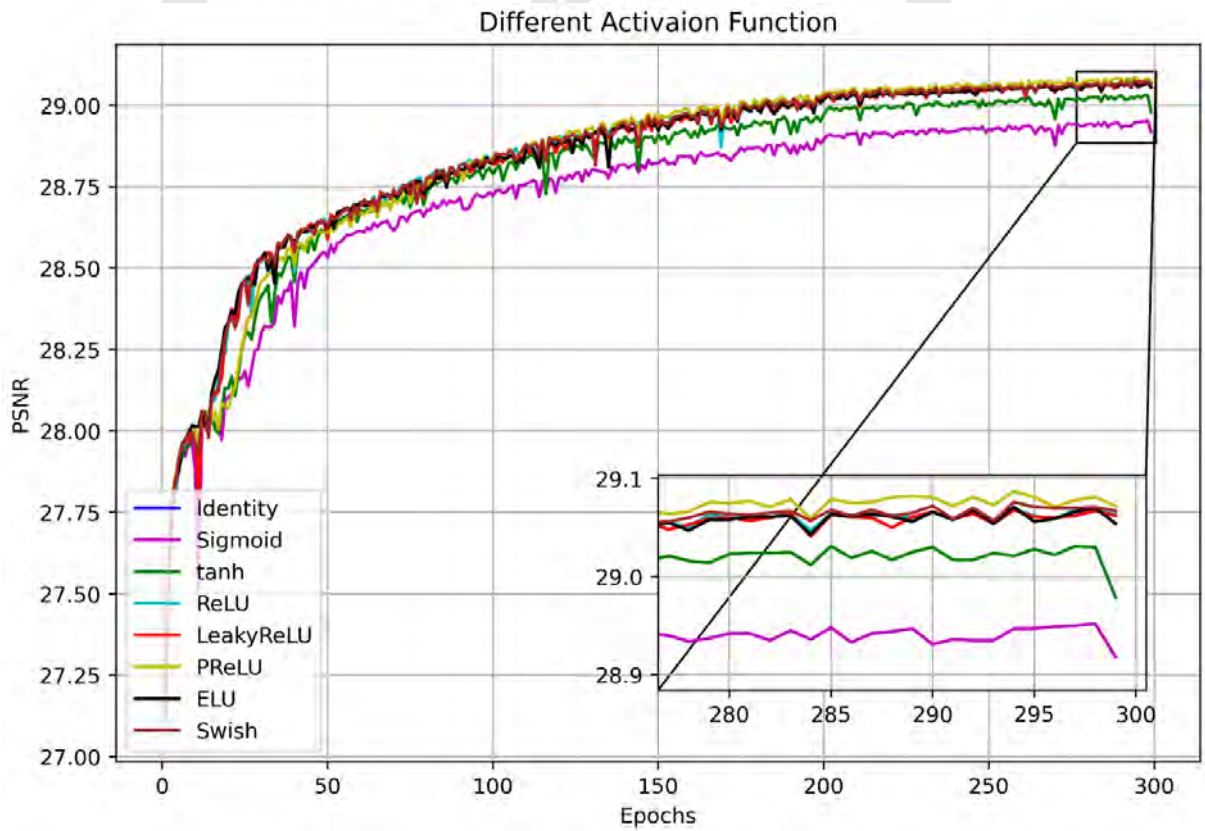


圖 1 不同的激勵函數在簡化版 EDSRX4倍率下的訓練過程 PSNR 值。(右下小圖為大圖的局部放

大

表二中的最佳結果我們用粗體紅色將其標記，次佳結果我們用粗體底線將其標記。其中我們得到PReLU的成效最好，推估是因為其含有參數可訓練的關係，並且較不會發生負數皆為0導致神經網路部分節點無法更新的問題。再者PReLU為ReLU與LeakyReLU之超集，並且保有更高的彈性，可依照問題特性進行訓練，得到較好的準確度。

表 2 不同的激勵函數在簡化版EDSRX4 使DIV2K 10張驗證集的最佳PSNR

激勵函數	簡化版EDSR X4 PSNR
Identity	29.073
Sigmoid	28.952
Tanh	29.031
ReLU	29.069
LeakReLU	29.068
PReLU	<b>29.087</b>
ELU	29.071
Swish	<u>29.076</u>

## (二)不同損失函數對超解析度成像的影響

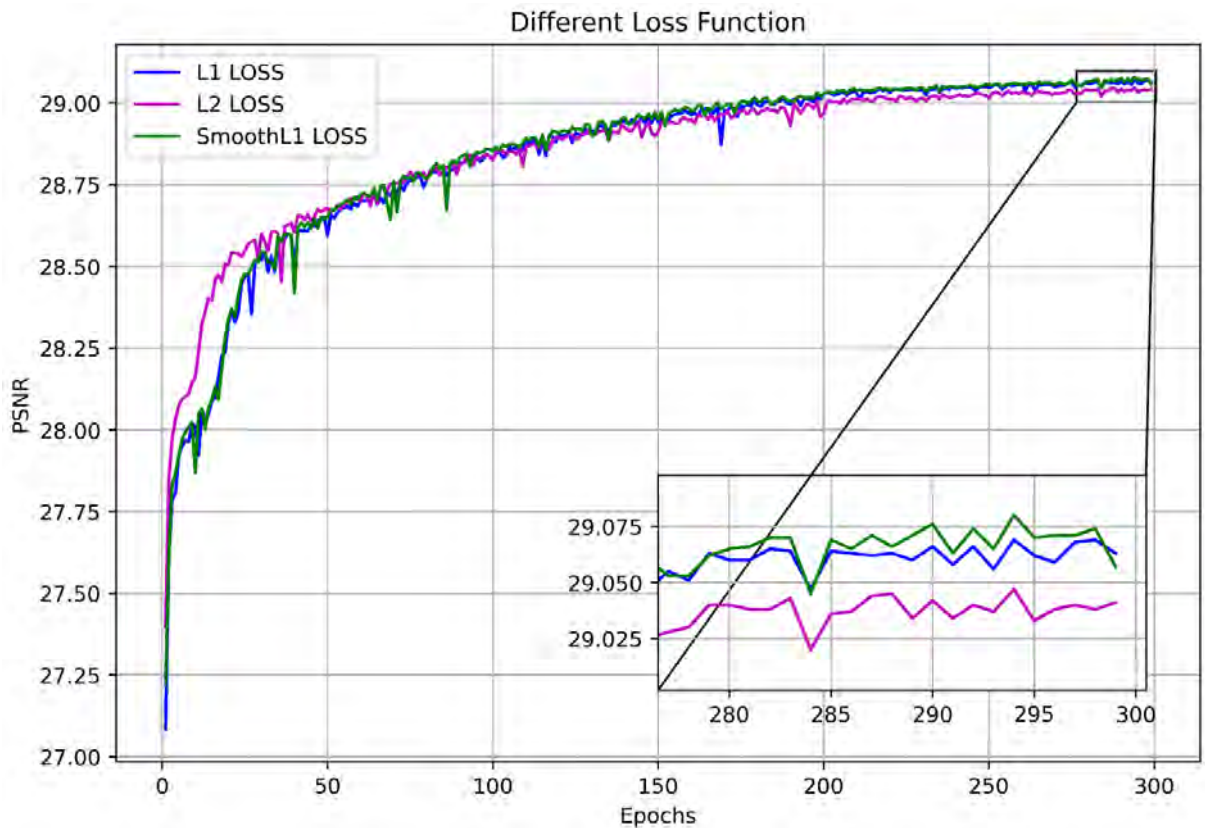


圖 2 不同的損失函數在簡化版 EDSRX4倍率下的訓練過程 PSNR 值。(右下小圖為大圖的局部放大)

表三中的最佳結果我們用粗體紅色將其標記，次佳結果我們用粗體底線將其標記。練前期 L2損失由於其平方的關係，參數能大幅度的變化，進而快速的提升準確度。而訓練中期至末期，不出意料的是使用 SmoothL1損失得到的準確度最好，由於其擁有 L1損失的高穩健性，能避免 L2損失中的異常值，並且在誤差值介於-1至1之間時可以擁有 L2損失的特性，更穩定且緩慢的收斂，並且不會產生無法微分的問題，因而得到較佳的準確度。

表 3 不同的損失函數於 DIV2K 10張驗證集的最佳 PSNR

損失函數	簡化版EDSR X4 PSNR
L1損失(MAE)	<b><u>29.069</u></b>
L2損失(MSE)	29.047
SmoothL1損失	<b>29.080</b>

### (三)不同優化器對超解析度成像的影響

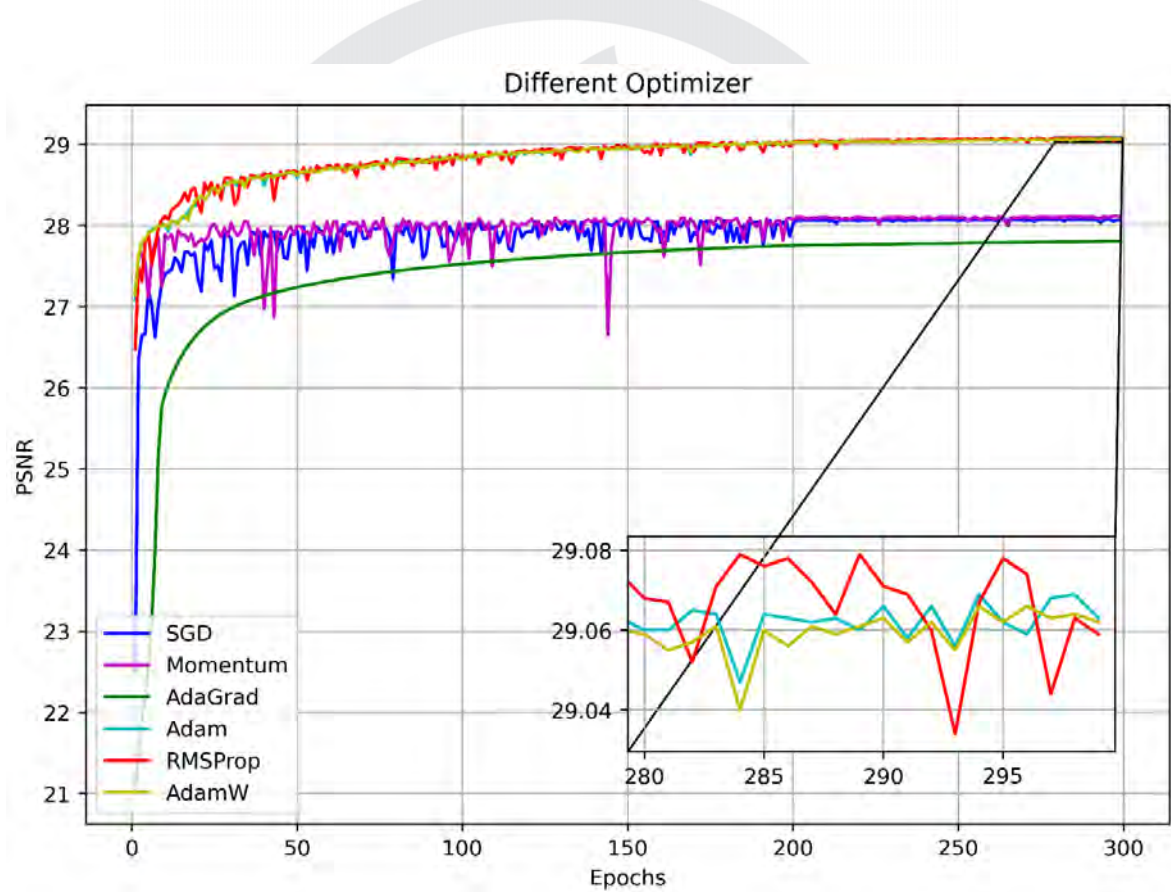


圖 3 不同的優化器在簡化版 EDSRX4倍率下的訓練過程 PSNR 值。(右下小圖為大圖的局部放

大)

表四中的最佳結果我們用粗體紅色將其標記，次佳結果我們用粗體底線將其標記。

表 4 不同的優化器於 DIV2K 10張驗證集的最佳 PSNR

優化器	簡化版EDSR X4 PSNR
SGD	28.071
Momentum	28.116
AdaGrad	27.806
Adam	<b>29.069</b>
RMSProp	<b>29.079</b>
AdamW	29.066

(四)不同模型優化前後比較

	放大 倍數	Set5	Set14	BSD100	UrBan100	DIV2K	Manga109
優化前 EDSR	2	37.542/0.9589	33.072/0.9129	31.821/0.8956	30.747/0.9138	33.959/0.9404	37.398/0.9742
優化後 EDSR		37.580/0.9589	33.106/0.9135	31.858/0.8961	30.872/0.9153	34.022/0.9410	37.505/0.9744
優化前 VDSR		37.326/0.9577	31.721/0.8994	27.237/0.8026	28.484/0.8829	33.721/0.9385	29.045/0.9332
優化後 VDSR		37.468/0.9585	33.020/0.9126	31.796/0.8952	30.684/0.9128	33.894/0.9399	37.288/0.9739
優化前 RCAN		37.824/ <b>0.9598</b>	33.332/0.9154	32.046/0.8984	31.625/0.9233	34.359/0.9436	38.259/0.9762
優化後 RCAN		<b>37.841/0.9598</b>	<b>33.389/0.9161</b>	<b>32.047/0.8985</b>	<b>31.643/0.9235</b>	<b>34.373/0.9438</b>	<b>38.312/0.9762</b>
優化前 EDSR	3	33.680/0.9204	29.853/0.8327	28.764/0.7974	27.012/0.8257	30.305/0.8771	32.144/0.9329
優化後 EDSR		33.675/0.9206	29.880/0.8334	28.789/0.7979	27.088/0.8278	30.342/0.8778	32.238/0.9340
優化前 VDSR		29.821/0.8741	25.887/0.7558	27.639/0.7734	23.963/0.7507	30.026/0.8721	24.560/0.8435
優化後 VDSR		33.600/0.9197	29.796/0.8317	28.727/0.7962	26.961/0.8242	30.241/0.8760	31.981/0.9320
優化前 RCAN		<b>34.134/0.9244</b>	<b>30.152/0.8381</b>	<b>28.954/0.8021</b>	27.661/0.8419	<b>30.637/0.8829</b>	32.954/0.9399
優化後 RCAN		31.114/0.9242	30.143/0.8378	28.953/0.8018	<b>27.685/0.8420</b>	<b>30.637/0.8828</b>	<b>32.960/0.9400</b>
優化前 EDSR	4	31.359/0.8809	28.086/0.7685	27.215/0.7249	25.056/0.7497	28.374/0.8232	28.940/0.8848
優化後 EDSR		31.423/0.8821	28.130/0.7697	27.250/0.7262	25.125/0.7525	28.417/0.8243	29.056/0.8868
優化前 VDSR		30.843/0.8728	26.893/0.7353	26.126/0.6978	23.565 0.7022	28.086/0.8166	21.477/0.7345
優化後 VDSR		31.364/0.8812	28.099/0.7684	27.221/0.7247	25.109/0.7509	28.370 0.8230	28.957/0.8851
優化前 RCAN		<b>31.874/0.8886</b>	28.360/0.7754	27.408/0.7310	<b>25.604/0.7693</b>	28.684/0.8305	<b>29.810/0.8984</b>
優化後 RCAN		31.851/0.8883	<b>28.389/0.7758</b>	<b>27.410/0.7315</b>	<b>25.604/0.7705</b>	<b>28.692/0.8312</b>	29.789/ <b>0.8985</b>

## 伍、結論與未來展望

### 一、結論

經過眾多的實驗證明，超解析度成像在超參數上有優化的空間，我們所提出的超參數組合(激勵函數使用PReLU、損失函數使用SmoothL1損失、優化器使用RMSProp)能有效的提升訓練的效率以及模型預測的準確度。不論是使用VDSR、EDSR、RCAN，在縮放因子為2、3、4倍上都在不同驗證集上得到良好的成績。

### 二、未來展望

在我們實驗中每個測試都十分得耗時，尤其是超參數的訓練，往往要等到模型趨於收斂後才能判斷是否有效，若未來當設充足的情況下，可在完整大小的模型上進行訓練與測試，並且嘗試更多超參數以及其組合，以利於未來超解析度成像的發展。

## 陸、參考文獻

- Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A. (2018). Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, 35(1), 53–65.
- Dong, C., Loy, C. C., & Tang, X. (2016, October). Accelerating the super-resolution convolutional neural network. In *European conference on computer vision* (pp. 391–407). Springer, Cham.

- Dong, C., Loy, C. C., He, K., & Tang, X. (2014, September). Learning a deep convolutional network for image super-resolution. In European conference on computer vision (pp. 184-199). Springer, Cham.
- 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).
- Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).
- Keys, R. (1981). Cubic convolution interpolation for digital image processing. *IEEE transactions on acoustics, speech, and signal processing*, 29(6), 1153-1160.
- Kim, J., Lee, J. K., & Lee, K. M. (2016). Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1646-1654).
- Kim, J., Lee, J. K., & Lee, K. M. (2016). Deeply-recursive convolutional network for image super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1637-1645).
- Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., ... & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4681-4690).

- Lim, B., Son, S., Kim, H., Nah, S., & Mu Lee, K. (2017). Enhanced deep residual networks for single image super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 136-144).
- Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., ... & Wang, Z. (2016). Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1874-1883).
- Tong, T., Li, G., Liu, X., & Gao, Q. (2017). Image super-resolution using dense skip connections. In Proceedings of the IEEE international conference on computer vision (pp. 4799-4807).
- Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., & Fu, Y. (2018). Image super-resolution using very deep residual channel attention networks. In Proceedings of the European conference on computer vision (ECCV) (pp. 286-301).