

# 深度學習演算法應用於轉動影像去模糊

張偉治 Q36111281  
June 19, 2023

## Abstract

我們使用半二次分解將解模糊的逆卷積過程分解成迭代的形式，根據該迭代的數學式，我們設計出端對端的解模糊模型，並使用部分 REDS dataset 資料進行訓練，最後我們以訓練集以外的 REDS dataset 對訓練後的模型進行實驗與相關分析。

## 1. Introduction

在工業製造上，常見的影像應用有物件的瑕疵分析、行為的控制或是異常的偵測等，而在這些應用情境中，影像的模糊與否會是相關應用的第一個瓶頸，因為模糊的影像會將低電腦視覺的精確度與可靠性。在物理上，我們可以使用更高幀數的攝像頭來解決影像模糊的問題，但是相對的會增加工業製造的成本，同時高幀數的攝像頭相對於一般的攝像頭較無法承受工廠的極端環境，因此基於成本與設備可靠性的考量，我們不能只從物理條件的角度去發想解決影像模糊的方法。

為此我們希望使用深度學習的演算法，對一般攝像頭捕捉到的模糊影像進行解模糊，使工廠端能用較低廉的成本得到與高幀數攝像頭相同可靠度的影像。為了驗證我們的解模糊模型的品質，我們將在 REDS dataset 上與其他解模糊模型進行比較與探討。最後將解模糊模型應用在轉動影像上，對高速旋轉下所產生的轉動動態模糊影像進行解模糊，希望能透過解模糊模型得到可靠的影像以利後續的應用。

### 1.1 Literature review

在過去深度學習的相關研究尚未蓬勃發展前，影像模糊的解決辦法通常基於 Wiener 濾波器去模糊的設計想法：尋找與模糊影像有最小誤差的圖片作為重建的圖片，此方法用線性化的手法解決了一般非線性最小化的問題。但由於真實雜訊的特性，Richardson 在 1970 年提出遞迴式的去模糊方法 --Lucy-Richardson Deblurring，該方法和 Wiener 濾波器一樣，都是針對反卷積做運算，但是 Lucy-Richardson Deblurring 應用了貝式定理，使得去模糊的過程可以迭代運算，並且該方法會將去模糊處理前後的影像作比較，以消除額外的雜訊。從上述早期的去模糊演算法的發展脈絡可以了解到，去模糊的關鍵在於點擴散函數(Point Spread Function)的設計，模糊的成因是原始像素與鄰近像素平均所導致，因此解模糊的過程必須有足夠的臨近像素資訊。

在深度學習相關研究蓬勃發展後，解模糊的核心思想承襲了 Wiener 濾波器和 Lucy-Richardson Deblurring 的架構，要對點擴散函數或是任意模糊核進行逆卷積，傳統 Model-based 的方式是試圖建立真實的解模糊算法，其優點是具有高度彈性，能根據輸入調整模糊核、縮放以及噪點處理；而近幾年流行的 Learning-based 方法—即深度學習的方式，是透過訓練資料學習解模糊的

參數，而非人工設計解模糊模型，而這樣的解模糊模型效果遠優於傳統 Model-based 的方法。然而 Learning-based 的方法缺點也很明顯，就是應用情境容易受限於訓練資料的內容，一旦模型完成訓練，今天新的解模糊影像大小不同於訓練資料，Learning-based 的模型就會無法與之匹配。

因此在這篇報告中，藉由將傳統的手法與深度學習的工具相結合，透過將逆卷積的過程藉由半二次分解的方式拆解成迭代的數學形式，在迭代的數學式中包含由傳統解模糊手法所計算的超參數，也包含涉及馬可夫鍊計算的先驗因子，後者即可利用合適的深度學習模型學習之。希望能藉由上述手法設計出如傳統解模糊演算法一般的端對端深度學習模型。

## 2. Mathematical model

假設今天有一清晰平面圖片  $x$ ，則線性模糊後的影像被定義為：

$$y = (x \bullet psf)_{\downarrow_s} + \varepsilon(x) \quad (1)$$

其中， $\bullet$  表示卷積； $psf(x)$  表示點擴散函數 PSF(Point Spread Function)； $\varepsilon(x)$  表示雜訊。

點擴散函數的部分也可以替換成任意模糊核  $k$ ，因此公式可以改寫成：

$$y = (x \bullet k)_{\downarrow_s} + \varepsilon(x) \quad (2)$$

要將模糊影像解模糊，我們可以採用 Learning-based model 來得到解模糊模型，因此可以通過最小化解模糊誤差來重建解模糊模型，公式寫成：

$$k_{blur} = \arg \min_k \left\| (x \bullet k)_{\downarrow_s} - y \right\| \quad (3)$$

### 2.1 Single image deblur

在 MAP (maximum a posteriori) 框架下，清晰影像可以藉由最小化以下 energy function 來得到：

$$E(x) = \frac{1}{2\sigma^2} \left\| y - (x \bullet k)_{\downarrow_s} \right\|^2 + \lambda \Phi(x) \quad (4)$$

其中，加法前項是資料項，加法後項的  $\Phi(x)$  項是先驗項、 $\lambda$  是權重參數。從公式中我們可以看出資料項和先驗項高度耦合，若採用 Learning-based model 對其進行求解不僅需要大量學習資料、收斂速度慢，更重要的是會需要大量的參數記憶解耦的過程。為解決上述問題，我們使用變量分裂法，引入輔助變量  $z$  簡化 energy function 的求解。變量分裂法是用於解決目標函數是兩個函數之合的優化問題，常見的變量分裂法有增廣拉格

朗日和半二次分解(half-quadratic splitting)兩種。在接下來的任務中，我們選擇使用半二次分解來求解，因為半二次分解收斂較為快速同時公式較為簡潔，我們因輔助變量  $z$  將上式分解：

$$E(x) = \frac{1}{2\sigma^2} \|y - (z \bullet k)_{\downarrow s}\|^2 + \lambda \Phi(x) + \frac{\mu}{2} \|z - x\|^2 \quad (5)$$

其中， $\mu$  為懲罰參數。因此上述 energy function 的問題可以通過迭代  $x$  和  $z$  求解：

$$\begin{cases} z_k = \arg \min_z \|y - (z \bullet k)_{\downarrow s}\|^2 + \mu \sigma^2 \|z - x_{k-1}\|^2 \\ x_k = \arg \min_x \frac{\mu}{2} \|z_k - x\|^2 + \lambda \Phi(x) \end{cases} \quad (6)$$

根據上述公式，若  $\mu$  足夠大，則  $x$  和  $z$  可以大約被視為固定點，但這樣的假設會導致收斂時間增加，因此比較好的做法是懲罰參數會隨著迭代次數增大。

透過上述變量分裂，我們將數據項和先驗項解耦。對於數據項，我們可以使用快速傅立葉變換進行求解；先驗項則被化簡成降噪問題。

$$z_k = F^{-1} \left( \frac{1}{\alpha_k} (\overline{F(k)} F(y) + \alpha_k F(x_{k-1}) - \overline{F(k)} \bullet \frac{(F(k) \overline{F(k)} F(y) + \alpha_k F(k) F(x_{k-1})))}{(F(k) F(k) + \alpha_k)}) \right) \quad (7)$$

### 3. System framework

該模型是透過式(6)與式(7)進行設計，為了計算出端對端解模糊模型的損失函數，我們將模型拆成三個部分，分別是：1) 計算超參數的 Hyper Parameter Net；2) 進行圖片解模糊核迭代計算的 Data Net；3) 透過解模糊核計算出解模糊影像的 Priori Net。

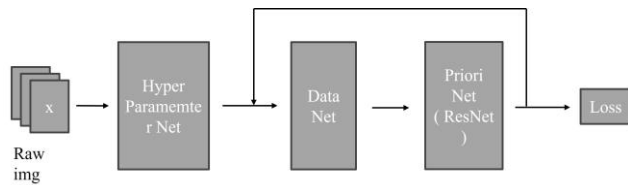


figure 1. 本篇報告中所使用的 Loss 計算架構圖。

在計算解模糊核  $z_k$  之前，我們需要先藉由 Hyper Parameter Net 計算如模糊核或是式(7)中  $\alpha_k$  等超參數，這些參數不參與訓練，但需要於迭代前進行設定。該超參數網路由三個卷積層組成，輸出的激活函數則使用 softplus，會使用 softplus 是因為在影像去模糊的任務中，需要確保迭代之間的連續性，因此選用比 ReLU 更平滑的 softplus 作為激活函數。

Data Net 則是計算解模糊核  $z_k$  的網路，Data Net 不包含模型的訓練參數，因此它只是一個用以計算出  $z_k$  的函式。在 Hyper Parameter Net 計算出  $\alpha_k$  等超參數後，Data Net 使用 Hyper Parameter Net 的輸入、目標影

像或是已經經過數次迭代的目標影像作為輸入，同時因為在式(6)中我們一直都是使用同一張圖進行迭代，因此我們可以設計適當的殘差作為  $F(x_{k-1})$  項的輸入。

在 Priori Net 中，我們設使用 U-Net 的思路設計，同時使用含自注意力機制的殘差網路作為模型的架構之一。在目標影像輸入模型後會經過一層卷積層作為目標影像的採樣頭，採樣後的結果會經過三次的下採樣殘差網路將特徵處理成適合 Attention body 計算的形狀，這些特徵在 Attention body 中會被計算成每一個 patch 的特徵向量，也就是每一個 patch 的解模糊核，緊接著的上採樣殘差網路會根據這些解模糊核，逐 patch 地計算出每一個 patch 的解模糊影像。

### 3.1 Experiment process

我們將使用 REDS dataset 做為我們與其他解模糊模型的比較標準。由於 REDS dataset 的圖像 (180\*320\*3) 以及模型參數較大 (因為模型在每次迭代都會對影像進行快速傅立葉轉換)，這都會導致我們所使用的電腦記憶體空間不足，因此用來訓練的資料是從 REDS dataset 的訓練集中選擇前 20 個影片的 2 幀，共 40 張圖片作為訓練，每個 batch 一張圖片，共進行兩個 epoch。

而在原本的實驗規劃中，為了貼近我們愈解模糊的情境，我們會以 REDS dataset 的原始影像做更進一步的幀數刪減以及融合，因為我們的目標是希望能對電磁感應非接觸式脫泡攪拌機的待攪拌物進行解模糊，該設備的正常工作轉速落在 1000 rpm 到 1500 rpm 之間，因此為模擬高速狀態下的線性模糊，模擬前、後幀具有較高差異性的特性，我們將根據 REDS dataset 的模糊設計方式重新產生模糊程度更高的模糊影像進行訓練與驗證。同時因為線性模糊與旋轉模糊不同，旋轉模糊通常以鏡頭中心為旋轉點，旋轉角度越大影像會越模糊，同時模糊的程度會與物體和旋轉軸的距離成相關，相同轉速下距離旋轉軸越遠的物體模糊程度越高，因此我們需要蒐集轉動模糊像的資料集，以此作為訓練集並分析其效果。

### 3.2 Introduce REDS dataset

REDS dataset 的原始影像是幀數為 120 fps、解析度為 720 x 1280 的真實動態影像，其生成的清晰連續圖像是以 24 fps 進行子採樣而得到。該資料集通常用於影像去模糊以及超解析度的任務上。



figure 2. REDS dataset 文件中的解釋。



除了原始影像外，該資料及包含經過模糊處理的模糊連續圖像、低解析度連續圖像以及低解析度模糊連續圖像。其中，模糊的設計是透過原始 120 fps 影像訓練一個 CNN 的非線性插植模型，接著應用該模型將原始 120 fps 影像的幀數提升到虛擬的 1920 fps，再以 24 fps 進行子採樣即得到模糊的影像。



figure 3. 清晰影像



figure 4. CNN 非線性插植後子採樣得到的模糊影像。

#### 4. Result

我們將我們訓練好的模型應用到 REDS dataset 的測試集當中。在影像具有較小動態模糊的例子中，模型的解模糊結果如下圖，直接觀察圖片，我們可以發現原始的模糊影像和清晰影像並無太大的區別，而兩種解模糊的結果都有不錯的表現，影像從具有晃動感的模糊，變成較為銳利的清晰影像。



figure 5. 由左上至右下分別是：1) 模糊影像；2) 模糊影像解模糊結果；3) 清晰影像；4) 清晰影像解模糊後結果。

雖然根據模糊影像和清晰影像解模糊後的結果都有不錯的表現，但若將解模糊後的圖片放大，我們會注意到雖然影像變得銳利，但解模糊後影像變得具有油畫感，同時可以觀察到 patch 的邊界，並且影像中的線條部分有抖動的情況。

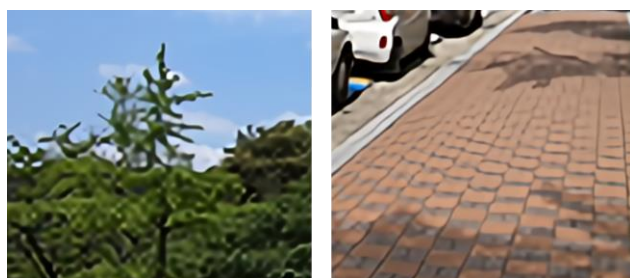


figure 6. 將 figure 5 中解模糊的圖片放大。

接著在影像具有較大動態模糊的例子中，原始的模糊影像和清晰影像有非常大的區別，模糊影像相比清晰影像具有非常大的移動模糊，而清晰影像則十分銳利。在解模糊的結果中，模糊影像解模糊結果解決了拍攝時因為手震所帶來的晃動模糊，但幾乎沒有解決影像中的移動模糊。而清晰影像解模糊後結果相比清晰影像則是在邊緣上變得更加銳利，例如牆面上不同材質磚瓦的各個交接處。

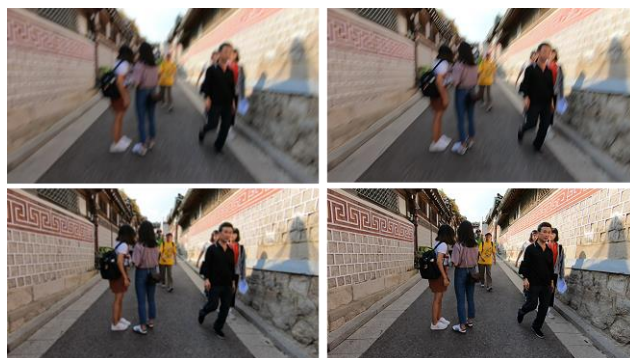


figure 7. 由左上至右下分別是：1) 模糊影像；2) 模糊影像解模糊結果；3) 清晰影像；4) 清晰影像解模糊結果。

將圖片放大觀察人臉細節的部份，解模糊影像都使得影像變得具有油畫的風格，甚至在模糊影像解模糊結果中，人臉已經變成色塊的塗抹，完全丟失人臉的細節。接著對比清晰影像和清晰影像解模糊後結果，由清晰影像得到之解模糊影像效果和使用邊緣強化卷積的效果差不多，都是強化了邊緣但遺失了細節，觀察人臉的眼睛部分，清晰影像雖然具有模糊感，但可以清楚辨識出眉毛、上眼瞼、眼睛三者的特徵，而後者則丟失了眉毛和眼睛之間的細節。

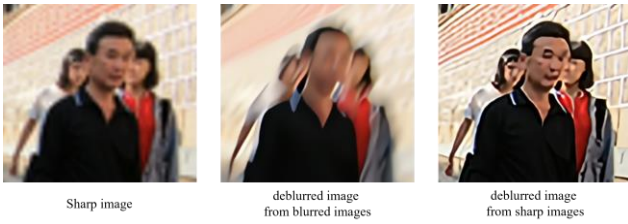


figure 8. 對比 figure7 中清晰影像、模糊影像解模糊結果和清晰影像解模糊後結果在人臉處理的細節。

最後為了實驗解模糊模型應用在轉動影像中的結果，我們使用網路上直升機的影像作為實驗對象。我們的模型完全沒有解決直升機螺旋槳轉動模糊的問題，但對比以下兩張圖，可以發現右圖相比左圖有變得較銳利。



figure 9. 解模糊模型應用於轉動模糊

放大觀察直升機的螺旋槳部分，可以推論模型完全沒有學習到如何解決影像的轉動模糊，這可能是因為訓練資料中並沒有轉動模糊的資料。



figure 10. 轉動模糊處的細節

但觀察該影像轉動模糊以外的細節，像是大樓的部分有變得較為銳利，但一樣有具有油畫感和線條抖動的問題出現。

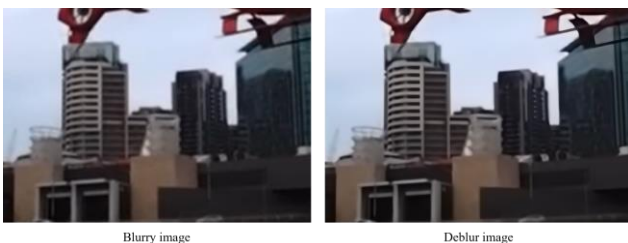


figure 11. 非轉動模糊處的細節

## 5. Conclusion

為了量化分析影像解模糊的效果，我們使用目標的清晰影像做為 ground truth，並計算其對模糊影像、模糊影像解模糊結果和清晰影像解模糊後結果的峰值訊

噪比(Peak signal-to-noise ratio, PSNR)。

圖片 示例	模糊影像	模糊影像解 模糊結果	清晰影像解 模糊後結果
figure 5	35.6	34.53	40.54
figure 7	22.7	22.66	39.95

表 1. 模型應用在 figure 5 和 figure 6 的 PSNR 結果

以模糊影像對清晰影像的 PSNR 作為基準，解模糊後的影像 PSNR 沒有變的較高，反而還降低，這與我們的直接觀察似乎有落差，但這有可能是因為雖然模型解決了拍攝時手震的模糊，但隨之所帶來的油畫風格反而降低了其 PSNR，抑或是影像的 ground truth 本身就帶有手震的模糊，因此模型解決了手震的模糊 PSNR 反而降低。而清晰影像解模糊後結果因為是以 ground truth 作為輸入，因此不適合與其他兩者做比較，但可以從數值中推斷出 ground truth 在經過模型僅有些微的不同，但我們無法僅由 PSNR 的數值判斷這是因為模型解決了清晰影像的手震模糊，還是模型所帶來的油畫風格所導致。

接著我們蒐集其他模型對該資料集的 PSNR 結果，發現其實我的模型在具有較小動態模糊的影像中，結果似乎還算不錯。

模型	Clip 000	Clip 011
DeblurGAN	26.57	22.37
DeepDeblur	29.13	24.28
SRN-Deblur	28.95	25.48
DBN	30.3	24.28
<b>Ours</b>	<b>34.53</b>	<b>22.66</b>
EDVR	36.66	34.33

表 2. 不同模型在 REDS dataset 的 PSNR 比較

根據上述實驗結果與分析，我們得出以下結論：1) 在運動模糊的影像中，沒有對應的 Motion blur kernels 時，僅使用 Gaussian kernels 無法達到好的解模糊效果。因此在未來的任務中，如何從動態模糊的影像中得到對應的 Motion blur kernels 是非常重要的任務；2) 在運動程度低的影像中，我們的模型可以達到還行的效果，但若放大影像，會發現解模糊的影像有明顯的 patch 邊界；3) 從清晰影像得到解模糊影像的結果可以發現，我們設計的模型在解模糊的過程中會大量丟失影像的特徵，猜想可能是因為訓練資料、迭代次數與 epoch 過少，在未來有合適的硬體資源後，可以使用完整的 REDS dataset 來訓練，以驗證該結果是否是因為訓練不足所導致。

## 參考文獻

- [1] 梁智泓 (2019)。基於深度學習於連續影格去模糊網路。〔碩士論文。國立臺灣大學〕臺灣博碩士論文知識加值系統。  
<https://hdl.handle.net/11296/j3333v>。
- [2] Fanous, M.J., Popescu, G. GANscan: continuous scanning microscopy using deep learning deblurring. Light Sci Appl 11, 265 (2022). <https://doi.org/10.1038/s41377-022-00952-z>
- [3] Seungjun Nah, Sungyong Baik, Seokil Hong, Gyeongsik Moon, Sanghyun Son, Radu Timofte, and Kyoung Mu Lee. 2019. NTIRE 2019 challenge on video deblurring and super-resolution: Dataset and study. In Proceedings of the CVPRW.
- [4] Kai Zhang, Luc Van Gool, and Radu Timofte. Deep unfolding

network for image super-resolution. In CVPR, 2020.

- [5] 國立台灣大學電信工程學研究所丁建均教授. 高等數位訊號處理. [2020-07-01].(<http://djj.ee.ntu.edu.tw/ADSP.htm>)
- [6] Wang, X., Chan, K.C., Yu, K., Dong, C., Change Loy, C.: EDVR: video restoration with enhanced deformable convolutional networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (2019)