

Robustifying Correction Filter For Deep Learning Models

深度學習模型使用的更正濾波器之補強方法

組別：B155 指導教授：黃朝宗 組員姓名：曾雋卿、陳泓賓

報告摘要

Correction Filter [1]此篇論文提出一種所謂的 correction filter：非特定降取樣所得的低解析度圖片，在經由超解析模型放大之前，可透過 correction filter 轉換成特定降取樣算法所得到的圖，由此來避免資料不匹配的問題，進而補強超解析度模型。

然而實際上，就算搭配 correction filter，有些情況下超解析度模型的表現並沒有提升，甚至會下降。本專題中，我們將針對這個問題，研究其原因及解決方案。

為了找出問題癥結，我們首先進行各種不同情境下的實驗，以瞭解 correction filter 的性質。基於[1]的原始碼，我們做出針對三種不同降取樣算法的 correction filter。同時我們訓練了分別用同樣的三種降取樣算法得出的圖片作為訓練資料，得到的三個超解析度模型。我們把用不同標準差的高斯濾波器分別降取樣得出的圖片做為測試圖片。我們以 correction filter 對測試圖片進行更正，得到更正後的圖片，並以更正後和未更正的圖片分別作為超解析度模型的輸入。測量兩者的輸出相對於原始高解析度圖片的 PSNR (peak signal-to-noise ratio)。

透過分析測量到的數據，佐以數位訊號處理中關於濾波器的理論，以及機器學習當中 loss function 的相關理論，並觀察論文的原始碼，我們對更正效果產生差異的原因提出猜測：可能與「correction filter 的輸入圖片當初降取樣時使用的降取樣濾波器」(以下簡稱 S)的 passband 範圍有關。本專題中我們用的 S 是高斯濾波器，其 passband 寬度對應到時域上的標準差。因此，我們進一步量測更多以 S 的標準差為操縱變因得出的數據，並驗證了這個假設，由此找出 correction filter 的缺點。明白了它的限制之後，我們基於理論，對一開始的問題提出可行的解決方案：透過設計、調整 correction filter 的轉換目標，使之與 S，兩者的 passband 盡可能吻合，correction filter 就可以有較優秀的表現。

一、關於 Correction filter

原論文提出的 correction filter，其目的是：將「非特定降取樣所得到的低解析度圖片」，轉換成如同「特定降取樣方法所得到的低解析度圖片」。根據輸入圖片的降取樣濾波器為已知或未知，其分成兩種情況：

輸入圖片使用的降取樣濾波器 k 已知(稱為 non-blind 情況)：

此時可直接由原論文中提出的公式(1)計算得出 correction filter \mathbf{h} 。

$$\mathbf{h} = \text{IDFT}\left\{\frac{\text{DFT}\{(k_{bicubic} * \text{flip}(k_{bicubic})) \downarrow_a\}}{\text{DFT}\{(k * \text{flip}(k_{bicubic})) \downarrow_a\}}\right\}, \quad (1)$$

$k_{bicubic}$ 為 bicubic filter，flip 代表空間上的順序顛倒(對應到頻域的 complex conjugate)， \downarrow_a 表示縮小 a 倍，* 是卷積，DFT 與 IDFT 分別是 discrete fourier transform and its inverse。

從頻域來看，整個式子，就是把 bicubic filter 和 k 分別都乘上翻轉的 bicubic filter 並縮小(這是考慮降取樣的效果)，再相除。為了數值穩定性，上述 \mathbf{h} 的算式，可透過加入非零的 ϵ 確保分母不為零(F_{numer}, F_{denom} 分別對應上面式子中的分子與分母項， ϵ 預設為 0)： $\mathbf{h} = \text{IDFT}\left\{\frac{F_{numer} \cdot F_{denom}^*}{|F_{denom}|^2 + \epsilon}\right\}$

輸入圖片使用的降取樣濾波器 k 未知(稱為 blind 情況)：

利用四層的線性卷積神經網路，透過疊代的方法，學習(猜)出「輸入圖片使用的降取樣濾波器 k 」。再將 k 代入 non-blind 式(1)算出 \mathbf{h} 。

Correction filter 與超解析度模型在使用時的關係如下圖所示。

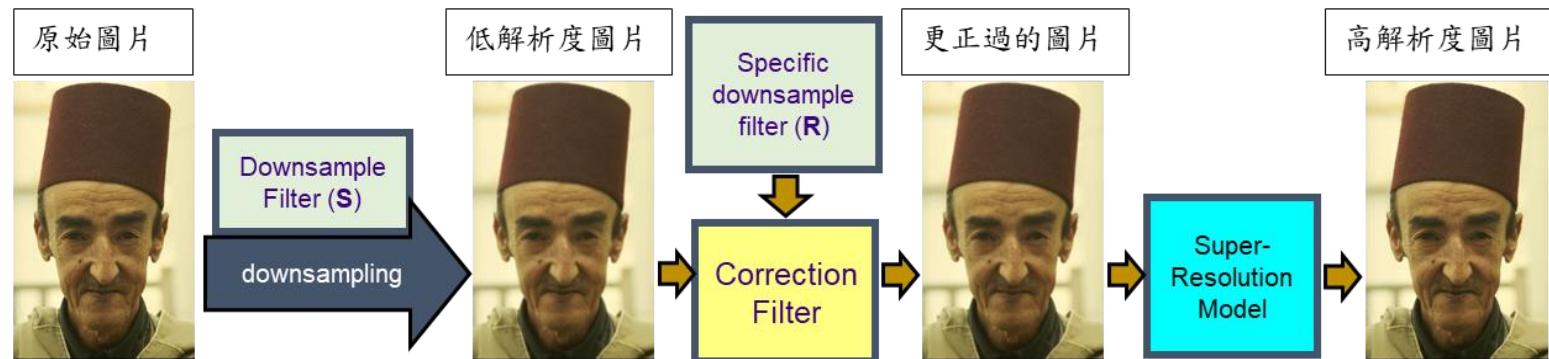


圖 1. 系統示意圖

二、研究方法與結果

(一) 名詞定義與說明

1. R：欲更正成的降取樣濾波器。即超解析度模型訓練資料集使用的降取樣濾波器。
2. S：測試圖片使用的降取樣濾波器。於本專題的實作中，S 皆為高斯濾波器。而內文所述的「S 的標準差」，皆是指 S 在時域上的標準差。
3. BIC、BIL：分別為 Bicubic、Bilinear 兩字的簡稱。
4. Non-blind：S 為已知的設定。透過論文提出的算法，用 S、R 計算得出 \mathbf{h} (the desired correction filter)。
5. Blind：S 為未知的設定。此設定下會透過卷積神經網路，猜出可能的降取樣濾波器 S，再帶入 non-blind 的方法計算出 \mathbf{h} 。

(二) 研究方法

我們使用[1]作者的原始碼，並加以修改，作出分別針對 bicubic、bilinear、以及 box (width = 2)三種降取樣方法的 correction filter。

超解析度模型部份，我們使用 EDSR baseline [2] 作為超解析度模型的架構。我們將 DIV2K 這個含有 800 張圖片的資料集，分別以上述三種降取樣方法搭配兩種縮小倍率(兩倍、四倍)，得到六組資料。再以此作為訓練資料，訓練出六個不同的超解析度模型。

這六個模型，在下文中分別稱作 BIC_x2、BIC_x4、BIL_x2、BIL_x4、box_x2、box_x4；前面的字串代表訓練資料使用的降取樣方法，後面的數字代表放大倍率。

測試資料部分，我們將 Set14 與 BSD100 這兩個分別含有 14、100 張圖片的資料集，以數種不同標準差的高斯濾波器搭配兩種縮小倍率，得到低解析度圖片。我們將這些圖片及其超解析後得到的圖片，稱為未更正組。而 Non-blind 設定下經 correction filter 更正過的圖片，及其超解析後得到的圖片，稱為 NB 或 Non-blind 組。依此類推，Blind 設定下的便稱為 Blind 組。

我們比較未更正組、NB 組、blind 組更正後的超解析圖片，在不同情境下的 PSNR，作為更正效果標準。情境的變數包括但不限於：

1. 降取樣的縮小倍率(同時也是超解析的放大倍率)：兩倍、四倍
 2. 降取樣時使用的高斯濾波器的標準差：主要有 $\frac{1.5}{\sqrt{2}}$ 、 $\frac{2.5}{\sqrt{2}}$ 、 $\frac{3.5}{\sqrt{2}}$ 、 $\frac{4.5}{\sqrt{2}}$
 3. 超解析度模型訓練資料所使用的降取樣方法：bicubic, bilinear, box
- 更正過的圖片，我們只會根據 R，輸進對應的模型進行超解析(例如：R = BIC 的更正圖片，只會由 BIC_x2、BIC_x4 進行超解析。)

(三) 量測的數據與觀察

我們首先量測了 2 種放大倍率、4 種不同 S 標準差(每一放大倍率搭配 2 種標準差)、3 種模型(bicubic、bilinear、box)及 3 種設定(未更正、Non-blind、Blind)的組合搭配，共 36 組資料進行觀察。

以下的圖表中，“x2_std. $1.5/\sqrt{2}$ ” 代表放大倍率為兩倍，S 標準差為 $\frac{1.5}{\sqrt{2}}$ ，以此類推。因篇幅限制，圖表僅展示訓練資料集降取樣濾波器 R 為 Bicubic filter 的更正效果(圖 2)，其餘數據則使用表格呈現(表 1)。

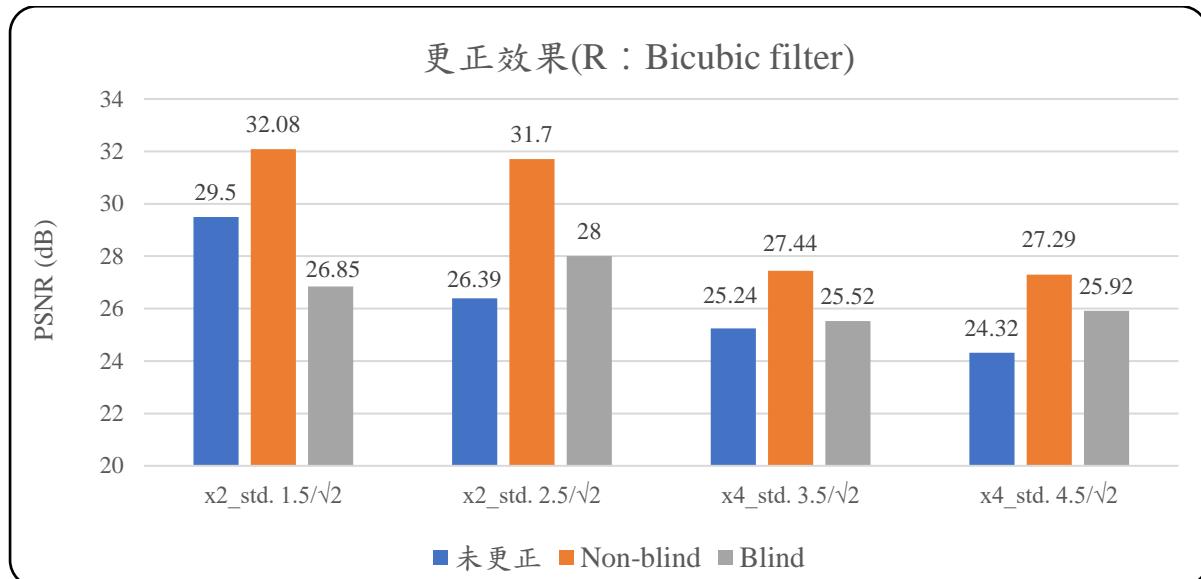


圖 2. 訓練資料集降取樣濾波器 R 為 Bicubic filter 更正效果比較圖

R:	設定:	x2_std. $1.5/\sqrt{2}$	x2_std. $2.5/\sqrt{2}$	x4_std. $3.5/\sqrt{2}$	x4_std. $4.5/\sqrt{2}$
BIC	未更正	29.5	26.39	25.24	24.32
	Non-Blind	32.08	31.7	27.44	27.29
	Blind	26.85	28	25.52	25.92
BIL	未更正	31.07	26.83	25.76	24.64
	Non-Blind	32.05	31.68	27.37	27.19
	Blind	24.58	27.31	24.67	25.47
box	未更正	29.1	26.4	25.22	24.36
	Non-Blind	30.96	29.84	26.07	25.63
	Blind	27.77	27.34	25.64	25.36

表 1. 更正效果表(表格內數值為 PSNR，單位: dB)

從數據中我們發現了一些趨勢與現象：Non-blind 設定下，更正後圖片的 PSNR， $R = \text{BIC}$ 及 $R = \text{BIL}$ 的相近， $R = \text{box}$ 的效果則較不突出。而相同倍率下，S 標準差較大者，更正效果較差一些。Blind 設定下，相同倍率下，S 標準差較小，更正效果較差；甚至有更正後圖片的 PSNR 不如未更正圖片的情況。

從上方的圖表，我們發現 S 的標準差(對應到頻域上 passband 的寬度)在不同倍率及模型下，對更正效果皆有相同趨勢的影響。我們因此著重在不同的 S 標準差，並以其作為操縱變因，放大倍率及模型作為控制變因，量測更多數據，如下方的圖 3 所示。

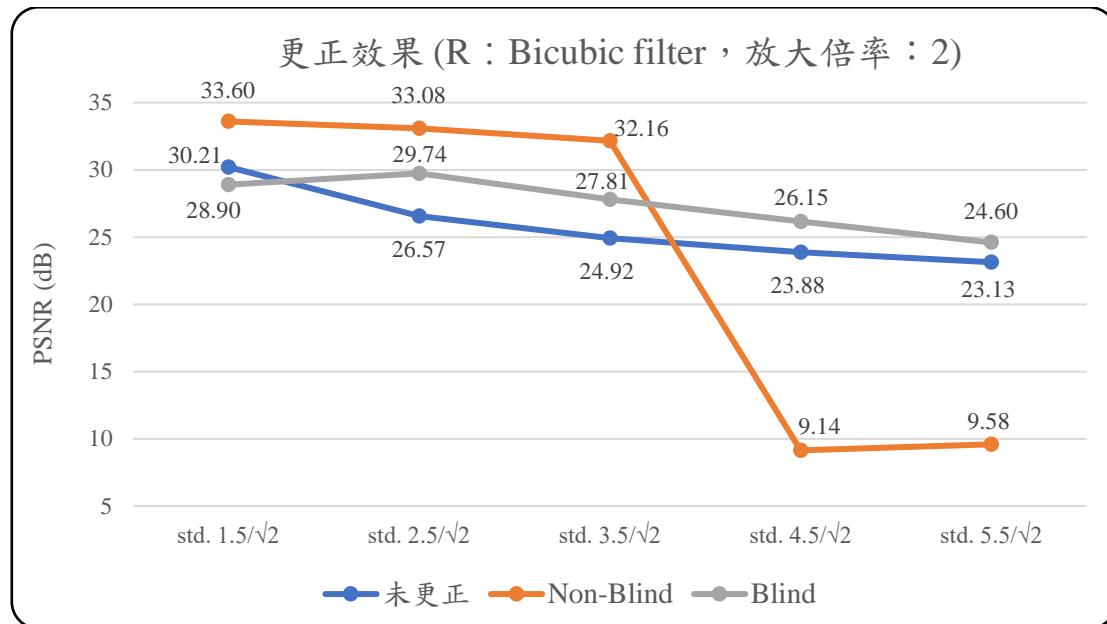


圖 3. 放大倍率 2 倍，訓練資料集降取樣濾波器 R 為 Bicubic filter 更正效果比較圖 (橫軸為 S 標準差)

從圖 3 當中，除了前述的現象，我們更發現：Non-blind 情況下，S 的標準差若大於一定大小，correction filter 會使圖片 PSNR 大幅下降，不如未更正圖片(圖 4)。

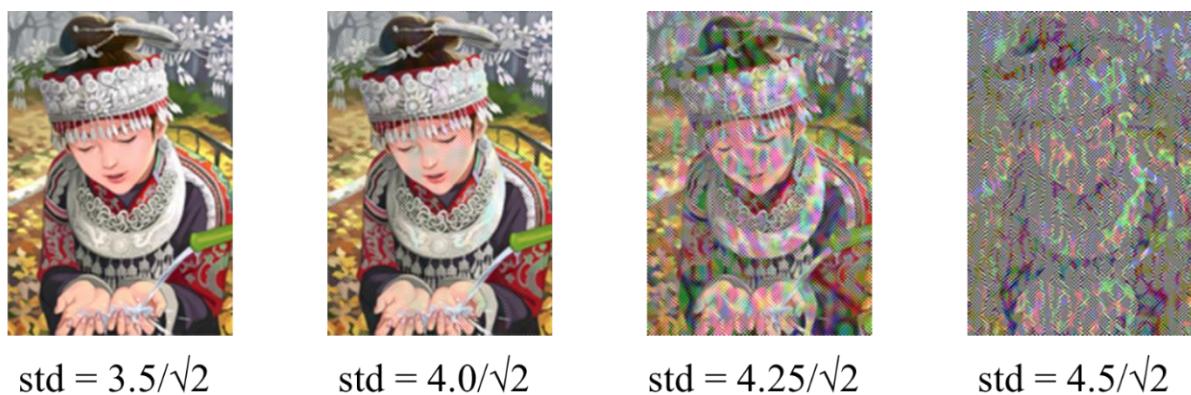


圖 4. 放大倍率 2 倍，訓練資料集降取樣濾波器 R 為 Bicubic filter 更正後低解析度圖片 (橫軸為 S 標準差)

(四) 問題的可能原因

對於前段提及的現象，我們猜測有兩個可能的肇因。以下分為兩點探討：

1. S 的 passband 若小於 R 的 passband，效果會變差。差距越大，效果越差

S 的標準差若大於一定大小，更正後圖片會不如未更正圖片，即圖 5 中 $\text{std. } 4.5/\sqrt{2}$ 、 $\text{std. } 5.5/\sqrt{2}$ PSNR 大幅下降、圖 6 的更正後圖片有大片斑點的情況。

我們把造成這個現象的原因，解釋為：因為在此 S 標準差下，已大幅偏離 correction filter 推導時所用的假設，也因此造成前段所述的幾無更正效果的狀況。

以下將解釋為何會偏離 correction filter 理論假設：

[1]在推導 correction filter 的算法時，其中一個假設為「S 的 null space 和 R 的值域，其交集為 0 向量」，即： $\text{null}(S) \cap \text{range}(R) = \{0\}$ 。其物理意義是：那些在降取樣時丟失的資訊 (null(S))，對升取樣的過程沒有影響。

為了滿足此假設，R 的 passband 需包含在 S 的 passband 之內。隨著 S 標準差變大，其 passband 也逐漸減小。在 R 不變的情形下，便愈來愈偏離假設，就可能使得更正效果變差。

據此，我們得出了可能原因：S 的 passband 若小於 R 的 passband，效果會變差。差距越大，效果越差。

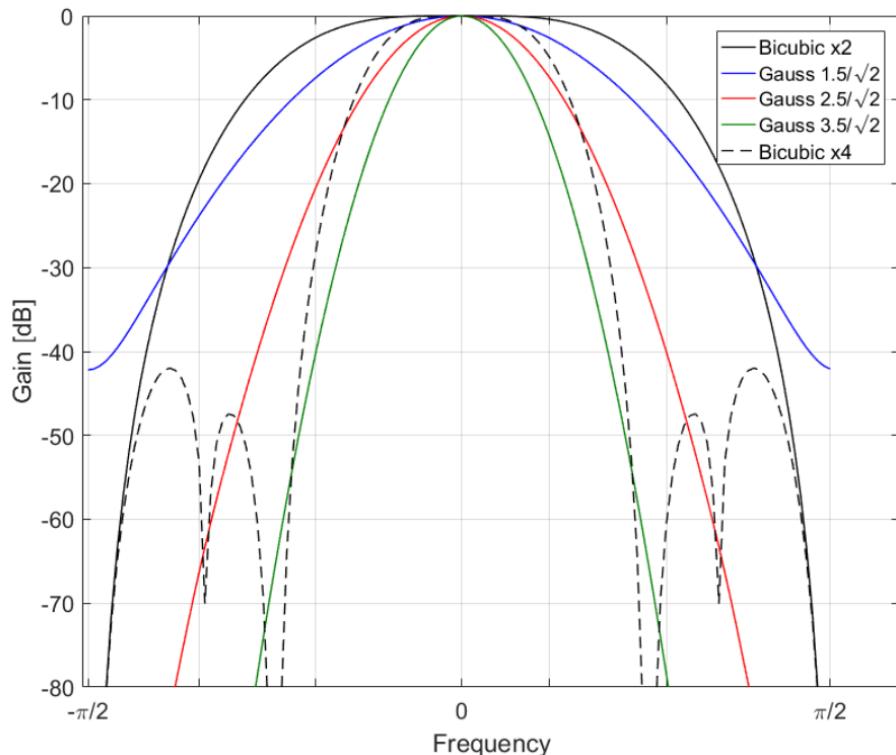


圖 5. 不同 kernel 一維頻譜圖(取自[1])

2. Blind 下，S 的 passband 越寬，其學習計算出來的降取樣濾波器 k 與 S 相差越大

我們認為 correction filter Blind 設定下，用來學習(猜出)低解析度圖片所使用之降取樣濾波器的 CNN 模型，其中的 Loss function，是造成圖 3 的「Blind 設定中 std. $1.5/\sqrt{2}$ 的更正後圖片的 PSNR 不如未更正圖片」，以及圖 1 和表 1 中「相同倍率下，S 標準差較小，更正效果較差」的原因。

我們將程式碼中的 Loss function 整理為下式(2)：(y 為未更正的低解析度圖片，SRmodel(x) 表示使用超解析度模型放大圖片 x，Downsample(x, k, a) 表示將 x 以降取樣濾波器 k 縮小 a 倍，h 為目前為止的疊代所學到的 correction filter)

$$Loss = L1\ Loss(\hat{y}, y) + \text{sum}(\sqrt{|k|}), \quad (2)$$

$$\hat{y} = \text{Downsample}(SRmodel(h * y), k, Scale), \quad (3)$$

其中的 \hat{y} ，是將低解析度圖片 y 和目前的 correction filter 卷積，再做超解析，然後經過 k 這個降取樣濾波器，並縮小成與 y 相同的大小。

Loss function 中 L1 Loss 的使用，會使學習的結果較為平滑，容易丟失高頻部分。因此，S 標準差較小，(即 S 的 passband 較寬、高頻部分較多)，會導致學習得出的 correction filter 與 S 相差較大，不容易有好的更正效果。

(五) 解決方法及討論

知道了造成 correction filter 更正效果產生差異的原因，我們便能根據這些因素來討論解決方式。以下將以測試圖片使用的降取樣濾波器 S 是否已知，即 Non-blind 和 Blind 設定來說明。因為時間因素(如：改寫程式碼、訓練模型所需的時間)，所以我們僅就理論直接討論，並沒有實作出來。

對於 Non-blind 設定：

我們可以透過選擇 passband 小於 S 的 R 進行更正，並搭配訓練資料的降取樣方法為 R 的超解析度模型，進行圖片之超解析。然而，理論上訓練資料用的降取樣方法 passband 越窄，模型的超解析效果會越差。因此，比較穩當的做法，是選用 passband 恰小於 S 的 R。

除此之外，[1] 中也提出了一種解決方式，就是調整 correction filter 算式中的 regularization parameter (ε)。透過提升 ε ，能在 S 的標準差過大時，獲得較穩定的表現，但整體而言更正過的圖片會較為模糊。

對於 Blind 設定：

我們可以訓練、準備多個以不同降取樣方法圖片作為訓練資料的超解析度模型(如同我們訓練了 BIC、BIL、box 模型)，而這些降取樣的方法，對應的濾波器的 passband 各不相同，從窄到寬都有。並將圖片對所有模型皆進行 blind 設定下的更正與隨後的超解析，最後人為挑選出效果最好的輸出圖片。

三、結論

我們量測觀察了不同情況下，correction filter 更正效果的差異，並針對差異探討出其可能原因。透過推論，我們瞭解到 S 及 R 的 passband 大小，對於 correction filter 成效有重要影響。如此一來，我們便可以此解釋各種情境下更正效果差異的原因，並根據個別設定提出了不同應用情況下，可能改善 correction filter 效果的方法。

四、心得感想

在這份實作專題中，需要用到數位訊號處理及機器學習相關的知識。我們雖然都有修習過數位訊號處理概論，但於專題一開始時對機器學習是一知半解。因此在上學期我們做了幾個機器學習相關題目的練習，如聲音辨識，同時也不斷複習數位訊號處理的相關內容。

決定了專題要做的題目之後，前段所做的練習使我們在閱讀、修改程式碼、訓練模型以及最重要的，理解 correction filter 和發現問題的可能原因上，皆能順利進行。

當然，專題不是只靠我們就能完成。我們非常感謝指導教授黃朝宗老師及陳永泰學長於過程中不厭其煩的教導。另外，除了專題題目本身相關知識，老師於每次 meeting 中亦會講解上台報告的技巧，點出 PPT 需要改正的地方……等等不同領域皆受用的專業能力。總總的學習內容，讓我們在這兩學期的實作專題中獲益良多。

五、參考文獻

- [1] Shady Abu Hussein, Tom Tirer, and Raja Giryes. Correction Filter for Single Image Super-Resolution: Robustifying Off-the-Shelf Deep Super-Resolvers. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. arXiv:1912.00157v2, 2020.
- [2] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017.
- [3] Yifan Wang, Federico Perazzi, Brian McWilliams, Alexander Sorkine-Hornung, Olga Sorkine-Hornung, and Christopher Schroers. A fully progressive approach to single-image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 864–873, 2018.
- [4] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. arXiv:1609.04802, 2016.