

國立臺灣師範大學跨域科技產業創新研究學院 AI 跨域  
應用研究所

碩士論文

Graduate Institute of AI Interdisciplinary Applied Technology

College of Interdisciplinary Industry-Academia Innovation

National Taiwan Normal University

Master's Thesis

基於色調映射與模型可解釋性技術的人臉偵測優化

Face Detection Enhancement Using Tone Mapping and  
Model Explainability Techniques

李少榆

Shao-Yu Lee

指導教授: 高文忠 博士

Advisor: Wen-Chung Kao, Ph.D.

中華民國 114 年 8 月

August 2025

# 致謝

歲月匆匆，碩士生活差不多要告一段落。回首研究與生活的交織歷程，若無許多師長、同儕與家人的扶持，便無今日之成果。在此謹以至誠謝意，向所有伸出援手、給予鼓勵的人們致敬。

首先，要感謝我家人無條件的陪伴和支持，他們是我面對未知時最大的定錨，無論潮起潮落，我知道總是有個地方願意接納我。

感謝我的指導教授高文忠院長以及瑞昱半導體股份有限公司的資深主管林鴻志博士，他們給了我很多人夢寐以求的合作機會和舞台。老師會在我研究撞牆時給我時間與建議，那樣的教育既深切又溫柔。很多時候老師給我的不僅僅是學術上的建議，更多時候是教我人生中重要哲學以及待人處事的道理；林博士也總是在百忙中抽出時間確保產學合作的順利進行。與瑞昱務實且開放的討論氛圍，使我受益良多，也體悟學術研究與產業應用之間搭建橋梁的意義。

感謝我的實驗室阿宅夥伴們（仁祥、馬華、13、柏翰、淳嘉、忠謀、璽文、秋貴、翔元、陳灝、鈺珊、宇彤、Ray、高、傳恩、林慶、Liu、知穎），我們在深夜除錯、學術討論及共享咖啡因的日子裡相互扶持；每一次技術切磋與吐槽都是前進的推力。

特別感謝中興資管互助會的成員、我的女友蘊芯還有林政宏教授，你們的出現給了我很多力量，這股力量幫助我度過了人生中一小段痛苦且無聊的歲月。

若有掛一漏萬，尚祈海涵。再次向所有曾給予我指導、協助與鼓勵的人致上最深的謝意。

# 基於色調映射與模型可解釋性技術的人臉偵測優化

學生：李少榆

指導教授：高文忠

國立臺灣師範大學 AI 跨域應用研究所



邊緣攝影機在極端背光與低光環境下，因對比失衡與雜訊升高，常導致人臉偵測表現顯著退化。本研究以全域與區域色調映射為核心，結合輕量化偵測器進行系統性評估與消融，聚焦於「前端影像增益」與「小樣本重新訓練」的相對效益與互補效應。結果顯示，在背光與低光影像集中，最佳組合可將檢測精度由 11.6% 提升至 50.7%，並明顯改善困難區域的人臉可見度與穩定性。基於此結論，我們提出適用於資源受限情境的實作指引，說明前端增益與輕量偵測的搭配原則與取舍，提供可部署方案，並為後續自適應色調映射與輕量偵測器的協同設計奠定基礎。

關鍵字：色調映射、低光／背光人臉偵測、輕量化模型、邊緣運算

# Face Detection Enhancement Using Tone Mapping and Model Explainability Techniques

student : Shao-Yu Lee

Advisors : Dr. Wen-Chung Kao

Graduate Institute of AI Interdisciplinary Applied Technology  
National Taiwan Normal University

## ABSTRACT

Edge cameras operating under extreme backlit and low-light conditions often suffer substantial degradation in face detection due to contrast imbalance and elevated noise. Centered on global and local tone-mapping strategies, this study conducts a systematic evaluation and ablation in combination with lightweight detectors, focusing on the relative benefits and complementary effects of front-end image enhancement and small sample retraining. Results show that, on backlit and low-light image sets, the best configuration increases detection accuracy from 11.6% to 50.7% and markedly improves face visibility and stability in challenging regions. Building on these findings, we provide implementation guidelines for resource-constrained settings, outlining the pairing principles and trade-offs between front-end enhancement and lightweight detection, offering deployable solutions and laying the groundwork for subsequent co design of adaptive tone mapping and lightweight detectors

Keywords: Tone Mapping; Low-light/Backlit Face Detection; Lightweight Models; Edge Computing

# 目錄

	<b>Page</b>
致謝	i
目錄	iv
表目錄	vii
圖目錄	viii
<b>第一章 緒論</b>	<b>1</b>
1.1 研究背景 . . . . .	1
1.2 研究問題 . . . . .	1
1.3 研究目標與具體貢獻 . . . . .	2
1.4 論文架構 . . . . .	3
<b>第二章 相關技術與文獻回顧</b>	<b>4</b>
2.1 色彩空間 . . . . .	4
2.2 影像前處理與色調映射技術 . . . . .	5
2.2.1 全域色調映射 . . . . .	6
2.2.1.1 攝影色調再現 . . . . .	6
2.2.1.2 直方圖均衡化 . . . . .	7
2.2.2 區域色調映射 . . . . .	11
2.2.2.1 影片錄製系統之及時色調重現 . . . . .	11
2.2.2.2 對比受限自適應直方圖均衡 . . . . .	14
2.2.2.3 區域對比增強 . . . . .	15

2.2.2.4	基於梯度範數的色調曲線 . . . . .	17
2.3	臉部偵測模型發展與輕量化趨勢 . . . . .	19
2.3.1	評估指標：mAP 與 mAR . . . . .	19
2.3.1.1	AR 與 mAR . . . . .	20
2.3.2	單階段檢測器 . . . . .	20
2.3.3	無錨框與輕量錨框式偵測架構 . . . . .	22
2.3.4	訓練策略：預訓練權重、資料擴增與再訓練 . . . . .	23
<b>第三章</b>	<b>研究方法</b>	<b>26</b>
3.1	系統概觀 . . . . .	26
3.1.1	輸入資料與前端影像處理模組 . . . . .	26
3.1.1.1	原始影像處理流程 . . . . .	27
3.1.1.2	色調映射之實作與參數設定 . . . . .	28
3.1.2	非極大值抑制 . . . . .	30
3.2	輕量臉部偵測器 . . . . .	31
3.2.1	模型概覽 . . . . .	31
3.2.2	模型設計差異 . . . . .	32
3.3	訓練流程與資料擴增 . . . . .	33
3.3.1	資料集與切分 . . . . .	34
3.3.2	UltraFace 兩階段訓練流程 . . . . .	34
3.3.2.1	推論設定固定化 . . . . .	35
3.3.2.2	訓練流程與資料擴增小結 . . . . .	36

3.3.3 評估指標與實驗配置 . . . . .	36
3.3.3.1 偵測評估指標 . . . . .	36
3.3.3.2 測試資料切分 . . . . .	36
3.3.3.3 小結 . . . . .	37
<b>第四章 實驗結果與分析</b>	<b>38</b>
4.1 色調映射與預訓練模型偵測效能 . . . . .	38
4.1.1 mAP/mAR 矩陣 . . . . .	39
4.1.2 結果分析 . . . . .	40
4.1.3 色調映射與預訓練模型偵測效能 —— 綜合小結 . . . . .	41
4.2 UltraFace 重新訓練實驗 . . . . .	42
4.2.1 UF-1K×7 分割之訓練結果 . . . . .	43
4.2.2 UF-7K 聯合 vs. 原始權重 . . . . .	45
4.2.3 結論與討論 . . . . .	45
4.3 成功與失敗案例討論 . . . . .	46
4.4 結論 . . . . .	50
<b>第五章 結果與未來展望</b>	<b>52</b>
5.1 限制與未解議題 . . . . .	52
5.2 未來展望 . . . . .	52
<b>References</b>	<b>54</b>
<b>自傳</b>	<b>57</b>
<b>學術成就</b>	<b>58</b>

## 表目錄

2.1 輕量 SSD 系列臉部偵測模型比較 . . . . .	21
2.2 訓練策略一覽 . . . . .	25
3.1 Tone Mapping 方法、參數與設定理由（統一符號與單位） . . . . .	29
3.2 輕量偵測器規格 . . . . .	32
3.3 UltraFace-Slim 重新訓練超參數（UF-7K-Joint / UF-1K×7-Split 相同） . . . . .	35
4.1 背光測試集 $mAR_{0.5}$ （七種 Tone Mapping × 五款預訓練模型） . . . . .	39
4.2 測試集 $mAR_{0.5}$ （七種 Tone Mapping × 五款預訓練模型）	40
4.3 UF-7K 聯合：AP 指標 . . . . .	44
4.4 UF-1K×7 分割：重訓前後 mAP . . . . .	44
4.5 UF-1K×7 vs. UF-7K：mAP@0.5 . . . . .	45

## 圖目錄

1.1 背光場景：增益前後對比 . . . . .	2
2.1 色調再現 (Reinhard) 增益比較 . . . . .	7
2.2 HE：低光影像與直方圖 . . . . .	9
2.3 HE：背光影像與直方圖 . . . . .	10
2.4 Tone Reproduction for Video 流程圖 . . . . .	13
2.5 tr_video 細節提取 . . . . .	13
2.6 tr_video 範例圖 . . . . .	14
2.7 local contrast enhancement 流程圖 . . . . .	16
2.8 tr_video 範例圖 . . . . .	17
2.9 GNB 亮度壓縮示意圖 . . . . .	19
2.10 三步訓練流程示意 . . . . .	24
3.1 推論流程圖 . . . . .	26
3.2 前端影像處理流程示意圖 (至色調映射與伽瑪校正) . . . . .	28
3.3 低光：七種色調映射示例 . . . . .	30
3.4 背光：七種色調映射示例 . . . . .	30
3.5 關鍵點與邊界框：質性比較 . . . . .	32
3.6 增益反例：HE 後反而漏檢 . . . . .	33
4.1 色調映射對五款偵測器之 $mAP@0.5$ . . . . .	39
4.2 色調映射對五款偵測器之 $mAR@0.5$ . . . . .	40
4.3 UF-1K $\times$ 7 分割：訓練／驗證損失 . . . . .	43
4.4 UF-7K 聯合：訓練／驗證損失 . . . . .	45
4.5 低光：GNB 成功案例 . . . . .	47
4.6 低光：GNB 成功案例 (二) . . . . .	47

4.7 背光：HE 成功案例 . . . . .	48
4.8 背光：LCE 成功案例 . . . . .	48
4.9 HE 失敗案例：雜訊過度放大 . . . . .	49
4.10 CLAHE 失敗案例：區域對比過度增益 . . . . .	49
4.11 色調映射：成敗示例（預訓練） . . . . .	50
4.12 UF-7K 聯合訓練：整體成功示例 . . . . .	50



# 第一章 緒論

## 1.1 研究背景

臉部偵測技術為計算機視覺領域中之核心任務之一，廣泛應用於身份驗證、人機互動、監控系統與行動裝置等多種場域。隨著深度學習技術的發展，特別是在卷積神經網路（Convolutional Neural Networks, CNNs）架構的推動下，臉部偵測演算法的準確率與即時性皆獲得顯著提升，使其逐漸邁向實務部署與邊緣裝置應用。然而，在實際應用情境中，臉部偵測模型往往面臨極具挑戰性的輸入影像條件，如背光、低亮度、光源不均、臉部遮蔽與低對比等情形。這些光照與成像品質問題將顯著影響模型之穩定性與泛化能力，進而導致誤判、漏判等現象。

為此，業界與學界開始重視影像前處理在臉部偵測任務中的潛在效益，期望透過色調映射、色調再現等手法，提升輸入影像品質，進而優化後續模型之判斷效果。色調映射技術最初多用於高動態範圍影像（High Dynamic Range, HDR）與圖像顯示增強領域，近年來亦被逐漸導入至 AI 視覺任務中，作為調整局部對比與細節的策略。透過適當的色調映射方法，可在不改變影像語意的前提下增強局部特徵表現，對於提升臉部偵測模型在惡劣光源條件下的表現具有潛在幫助。本研究聚焦於探討不同色調映射方法作為影像前處理手段，對輕量化臉部偵測模型效能之影響，期望從模型行為層面理解前處理對推論結果之實際作用。

## 1.2 研究問題

深度學習雖已顯著推進臉部偵測的實用性，但在低光或強背光等高動態範圍場景下，輸入影像的訊雜比（Signal to Noise Ratio, SNR）驟降，色調映射或對比度增強雖能凸顯輪廓，卻也可能同步放大感光雜訊與偽影，進而造成誤判。此一問題在嵌入式或低功耗裝置尤為嚴重：硬體資源限制了複雜影像處理與大規模再訓練的可行性，因此如何在「前端影像處理」與「後端模型訓練」之間取得平衡，成為值得深入探討的課題。

本文聚焦三項互相關聯的研究議題。首先，我們檢驗不同全域／區域色調映射法做為前處理時，對檢測結果的正負效應，並說明在雜訊水平升高的情況下，過度增益可能反而拉低精度。其次，我們比較「僅使用公開預訓權重」與「針對七種增益資料集重新微調」兩種流程，量化再訓練在低光、背光場景下的效益，結果顯示適度微調可將檢測精度平均提升三成以上。最後，本文探討各模型（UltraFace、RetinaFace、SCRFD、YOLO 等）在不同參數組合下對增益演算法的敏感度。



Figure 1.1: 背光場景影像之增益前後對比。(a) 未經前處理之原始影像，在強背光條件下人臉細節嚴重失真；(b) 套用本研究色調映射增益後之影像，顯著恢復人臉輪廓與細節，並維持背景亮度層次。

### 1.3 研究目標與具體貢獻

本研究聚焦於「影像前處理——特別是色調映射——與輕量化臉部偵測模型之互動關係」，旨在釐清適度增益對低光／背光場景偵測效能的實際助益，並建立可供業界於嵌入式環境採用的實作準則。具體目標包括：(1) 量化不同色調映射策略對臉部偵測性能（mAP、mAR）的影響；(2) 比較五款主流輕量模型在七種前處理條件下的敏感度與相容性；(3) 評估「僅用預訓權重」與「針對增益資料集再訓練」兩種工作流程之差異，探討前處理與模型協同設計的必要性。為達成上述目標，本研究完成以下貢獻：首先，自行蒐集並標註 1,000 張低光與強背光影像，據此擴增出七套對應增益資料集，後續再以 7,000 張樣本擴大微調，建

立具代表性的評測基準。

其次，系統性整合區域對比度增益、攝影色調再現、直方圖均化等等六種演算法，並以模組化管線接軌五款輕量偵測器；實驗證實，經過再訓練後，檢測指標平均可在極端光源場景提升逾 30%，同時揭示過度增亮易引入偽影而使性能反跌的潛在風險。最後，透過質性對照（成功／失敗案例）與結果分析，我們驗證前處理對偵測置信度、定位穩定度與誤檢／漏檢的影響；並據此彙整「場景光源特性×增益方法×模型選型」之實務建議，作為未來低功耗人臉偵測系統的設計依據。

## 1.4 論文架構

本論文共分為五章，各章內容與相互邏輯關係說明如下。第二章首先回顧影像前處理領域中與色調映射相關之經典與近年方法，區分全域與區域兩大分支，並整理其在臉部偵測場域的應用現況與研究缺口；同時概述輕量化臉部偵測模型（UltraFace、SCRFD、YOLO-Face 等）的設計脈絡與評估指標。第三章說明整體實驗平台與方法論：先介紹資料集蒐集與標註流程，接續描述七種色調映射策略的實作細節與參數設置，再闡述五款偵測模型之訓練／微調策略，以及 mAP、mAR 評估。第四章呈現系統化實驗結果，分別比較（1）各前處理於預訓模型上的即插即用效益、（2）再訓練後的性能提升幅度與模型敏感度差異。第五章綜合研究發現，歸納前處理與模型協同設計之建議原則，並討論本研究在嵌入式應用、資料擴增與自適應增益方面的潛在延伸方向，作為後續研究與實務部署之參考依據。

## 第二章 相關技術與文獻回顧

本章旨在回顧影像色調映射與臉部偵測模型之發展脈絡，並分析前處理策略在複雜光源場景中的應用現況與研究缺口。內容分為兩大部分：(1) 色調映射技術的分類與代表性方法；(2) 輕量化臉部偵測模型及其演進。

### 2.1 色彩空間

CIE (Commission Internationale de l'Éclairage) 是制定色度學標準的國際組織，包含標準觀察者、色度匹配函數與多個標準色彩空間。本文使用的 CIE 1931 XYZ、CIE L\*a\*b\* 與 Yxy 等表示，皆出自或衍生自 CIE 的定義。

在討論色調映射之前，首先必須將感知亮度與色度分量分離。依照文獻 [1]，常見作法可分為兩類：(i) 線性 RGB  $\rightarrow$  CIE XYZ  $\rightarrow$  Yxy；(ii) 線性 RGB  $\rightarrow$  CIE L\*a\*b\*。兩者皆能將亮度抽離為單一標量，但其在色調映射應用上仍存在差異，說明如下。實務上皆會先轉換至 XYZ 色彩空間，再依需求轉換至 CIE L\*a\*b\* 或 Yxy。

$$(i) \text{ 正規化並反 } \gamma: C_s = \frac{C_{\text{bit}}}{255}, C' = \begin{cases} \frac{C_s}{12.92}, & C_s \leq 0.04045, \\ \left(\frac{C_s + 0.055}{1.055}\right)^{2.4}, & C_s > 0.04045, \end{cases} \quad (2.1)$$

$$(ii) \text{ sRGB } \rightarrow \text{ XYZ (D65)}: \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4124564 & 0.3575761 & 0.1804375 \\ 0.2126729 & 0.7151522 & 0.0721750 \\ 0.0193339 & 0.1191920 & 0.9503041 \end{bmatrix} \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} \quad (2.2)$$

**(1) CIE L\*a\*b\*** Lab 空間以人眼等色彩差異為初衷，L 分量在中亮度區具近似線性之明度感知。然而實務上，若僅對 L 做強烈的非線性增益，而不同步依

公式調整 a、b，常會出現「彩度漂移」——在低光區過度提升 L 造成色彩去飽和、畫面偏灰。

$$f(t) = \begin{cases} t^{1/3}, & t > \delta^3, \\ \frac{t}{3\delta^2} + \frac{4}{29}, & t \leq \delta^3. \end{cases} \quad (2.3)$$

$$\begin{cases} L^* = 116 f\left(\frac{Y}{Y_n}\right) - 16, \\ a^* = 500 \left[ f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right) \right], \\ b^* = 200 \left[ f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right) \right]. \end{cases} \quad (2.4)$$

其中  $\delta = \frac{6}{29}$ ， $(X_n, Y_n, Z_n)$  為參照白點（本研究採用 D65： $X_n = 95.047, Y_n = 100.000, Z_n = 108.883$ ）。

**(2) CIE Yxy** CIE  $Yxy$  空間將物理亮度  $Y$  與色度  $(x, y)$  完全分離；其中  $Y$  為線性光度， $(x, y)$  為色相／飽和度之 2D 座標，轉換自 CIE XYZ 如下 [2]：

取亮度  $Y$  並計算色度座標  $(x, y)$ ，即可獲得  $Yxy$  表示：

$$x = \frac{X}{X+Y+Z}, \quad y = \frac{Y}{X+Y+Z}, \quad Y = Y. \quad (2.5)$$

## 2.2 影像前處理與色調映射技術

本節旨在回顧影像訊號處理（Image Signal Processing, ISP）領域中與色調映射（Tone Mapping, TM）相關之代表性方法，並釐清「全域（global）」與「區域（local）」兩大分支的差異與應用情境。

在人眼視覺系統裡，虹膜會因環境光強弱自動收縮或放大，使我們在室內暗

處與戶外強光之間仍能辨識景物的細節與層次；然而，大多數影像感測器與 8 位元儲存格式（如 JPEG、BMP、PNG）並無法在單張影像中覆蓋這樣寬廣的亮度範圍。當明暗比超過感測器容許值時，高亮區會產生過曝，陰影區則陷入雜訊或全黑，細節隨之流失。

傳統相機為捕捉高動態範圍場景，往往需拍攝多張不同曝光度的影像，再以融合（fusion）合成 14 或更高位元的 HDR 影像。然而此作法不僅增加儲存量與拍攝不便，也難以滿足即時應用或嵌入式裝置的資源限制。因此，「如何僅憑單張輸入影像，就能在有限位深度內重現接近人眼感知的亮度階」形成了色調映射研究的核心課題。

全域色調映射透過一條統一的亮度曲線快速壓縮動態範圍，計算量低、硬體友善；唯其對高亮與深暗區只能折衷，難兼顧局部對比。區域色調映射（Local/Multiscale TM）改以空間適應或多尺度濾波調整亮度，能保留細節並抑制過曝，卻付出了更高的計算與記憶體成本。本研究將於後續章節實作各種區域以及全域色調映射，並探討它們在臉部偵測任務中的實際效益與限制。

## 2.2.1 全域色調映射

全域色調映射以單一亮度轉換函數  $f(L)$  同步作用於整張影像，流程通常為「亮度統計 → 曲線推導 → 點對點映射」，計算與記憶體開銷最小，嵌入式友善。

### 2.2.1.1 攝影色調再現

全域色調映射代表性方法有 Reinhard 等人提出之「Photographic Tone Reproduction」[3] 以攝影膠卷之「鍵值」（key value）概念為基礎，先取對數平均亮度

$a$  縮放輸入亮度，再施以壓縮函式：

$$L_{\text{avg}} = \exp\left(\frac{1}{N} \sum_{x,y} \ln[L_w(x,y) + \delta]\right), \quad (2.6)$$

$$L(x,y) = \frac{a}{L_{\text{avg}}} L_w(x,y), \quad (2.7)$$

$$L_d(x,y) = \frac{L(x,y)}{1 + L(x,y)}, \quad (2.8)$$

$L_{\text{avg}}$  為對數平均亮度， $\delta$  防止  $\log 0$ ；式 (2.8) 可再引入  $L_{\text{white}}$  控制高光 roll-off 以避免過曝。



Figure 2.1: 色調再現 (Reinhard) 增益比較。(a) 原始影像；(b) Reinhard 全域映射， $\alpha = 0.18$ ；(c) 同上， $\alpha = 0.27$ ；(d) 同上， $\alpha = 0.50$ 。

### 2.2.1.2 直方圖均衡化

直方圖均衡化 (Histogram Equalization, HE) 是一種經典的全域對比增強技術，其核心思想是利用影像灰階機率分佈 (Probability Density Function, PDF) 與累積分佈函數 (Cumulative Distribution Function, CDF)，將輸入影像之灰階值重新映射，使輸出影像之直方圖趨於均勻分佈，從而增強整體對比度 [4]。

首先，令輸入影像大小為  $M \times N$ ，離散灰階集合為  $r_k$  ( $k = 0, 1, \dots, L-1$ )，其

PDF 定義為

$$p_r(r_k) = \frac{n_k}{MN}, \quad \sum_{k=0}^{L-1} n_k = MN, \quad (2.9)$$

其中  $n_k$  為灰階  $r_k$  的像素數量。接著計算 CDF

$$\text{CDF}(r_k) = \sum_{j=0}^k p_r(r_j), \quad (2.10)$$

並將每一灰階值映射為

$$s_k = (L - 1) \text{CDF}(r_k), \quad 0 \leq s_k \leq L - 1. \quad (2.11)$$

式 (2.11) 代表將累積機率線性縮放至  $[0, L - 1]$ ，理論上可使輸出灰階分佈近似均勻 [5]。

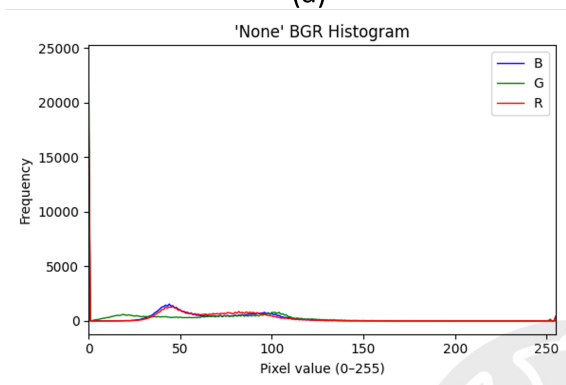
然而，HE 屬全域操作，對於光照分布極度偏斜或高 ISO 影像，易出現過度對比或雜訊放大等問題。為了緩和此缺陷，後續發展出自適應直方圖均衡化 (AHE) 及對比受限 AHE (CLAHE) [6]，以提升區域一致性與抑制雜訊。



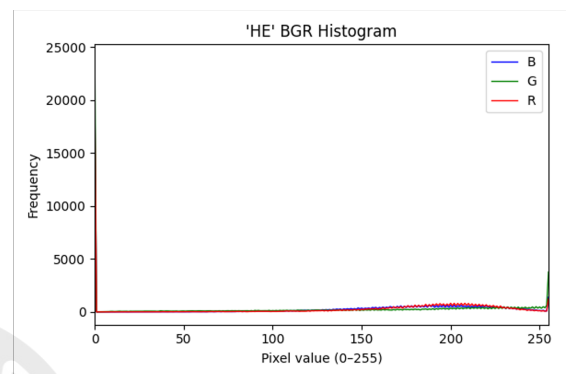
(a)



(b)



(c)



(d)

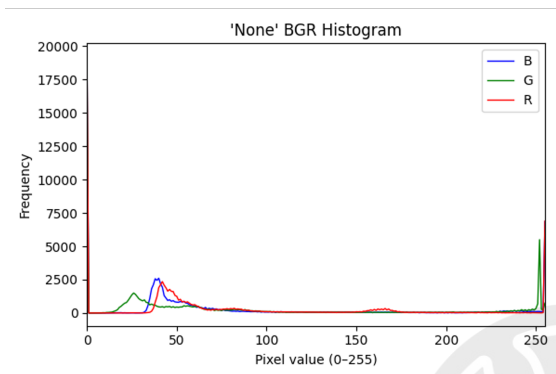
Figure 2.2: 低光影像：直方圖均衡化前後之影像與 BGR 直方圖。(a) 未套用色調映射的原始影像，整體亮度不足且細節沈於暗部；(b) 套用 HE 後之影像，灰階分佈擴展至高亮區，但感光雜訊同步放大；(c) 原始影像之 BGR 直方圖，能量主要集中於 0-80；(d) HE 後之 BGR 直方圖，灰階分佈被拉伸至 0-255，三通道高值端顯著抬升，顯示全域對比提升但雜訊與過曝風險增加。



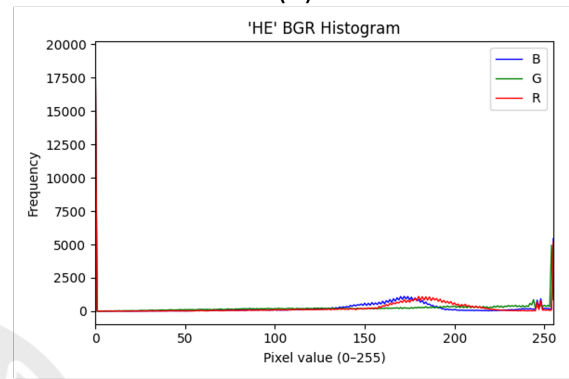
(a)



(b)



(c)



(d)

Figure 2.3: 背光影像：直方圖均衡化前後之影像與 BGR 直方圖。(a) 未套用色調映射的原始影像，受強背光影響臉部細節嚴重壓暗；(b) 套用 HE 後之影像，灰階分佈向高亮擴展，但背景高光區域更易過曝；(c) 原始影像之 BGR 直方圖，能量集中於低灰階；(d) HE 後之 BGR 直方圖，分佈拉伸至 0-255，三通道高值端抬升，對比提升同時帶來雜訊與過曝風險。

由圖 2.2 以及圖 2.3 可觀察到，原始影像受低光或背光影響，灰階分佈集中於暗端，經 HE 處理後，灰階分佈被拉伸至整個動態範圍，影像明暗對比及細節均獲得顯著提升。然而在高亮區域（如天空）仍可能產生輕微過曝，顯示 HE 雖能快速改善全域對比，對於局部光照不均的情況仍有其侷限。

綜合而言，HE 具有實作簡單、計算量低之優點，在醫學影像、監控影像與嵌入式裝置等領域仍被廣泛採用。然而在本研究聚焦的低光源人臉偵測應用中，為避免過度放大感測雜訊及保留區域細節，仍需進一步探討局部自適應之增強方法，並與 HE 作系統化比較。

全域色調映射具備計算量低、向量化容易、硬體友善等優勢，惟因單曲線難因應局部強背光／陰影，後續章節將比較本節方法與區域映射演算法，探討其於臉部偵測任務之效益與限制。

## 2.2.2 區域色調映射

全域色調重現 (Global Tone Mapping, GTM) 僅依整張影像的統計量建立單色調曲線 (Tone Curve)，在背光、局部照度極端不均或 HDR 場景時，往往造成前景壓暗或高光壓縮過度，無法同時兼顧暗部細節與邊緣對比 [7]。為解決此一侷限，局部色調重現 (Local Tone Mapping, LTM) 於空間域引入鄰域資訊或先分離基底／細節層，再對基底亮度施加自適應壓縮，使 tone-curve 隨畫素位置動態變化。此策略能在提亮暗區的同時保留高梯度邊緣，並透過運算量控制（例如改用簡化濾波或區塊化處理）在嵌入式 SoC 上維持及時性，特別適合本研究聚焦的低光源人臉偵測應用。

### 2.2.2.1 影片錄製系統之及時色調重現

Reinhard 的攝影色調再現透過全域對數平均亮度  $L_{avg}$  來決定曝光，但忽略區域差異；在背光或高反差場景中，前景細節常被壓暗，而高光又可能過度擠壓，造成影像失真。為兼顧即時性與區域對比，Kao 等人 [8] 提出「頻率分割色調再現」演算法，其主要流程如下：

1. **高低頻分離**：對輸入亮度  $Y$  施以  $3 \times 3$  均值濾波取得低頻平滑影像  $L_Y$ ，並以  $H_Y = Y - L_Y$  取出高頻細節。此法較傳統雙邊濾波計算量低，可於嵌入式 SoC 以硬體加速器即時運作。
2. **低頻多次 Tone Mapping**：先計算低頻之對數平均

$$L_{avg} = \exp\left(\frac{1}{N} \sum_{x,y} \ln[L_Y(x,y) + \delta]\right), \quad (2.12)$$

其中  $\delta \ll 1$  避免  $\ln 0$  的奇異值。再依場景自適應的 *key value*  $k$  (預設 0.18) 與白點  $L_{\text{white}}$  進行兩段式壓縮：

$$L'(x, y) = \frac{k}{L_{\text{avg}}} L_Y(x, y), \quad (2.13)$$

$$L_d(x, y) = \frac{L'(x, y)}{1 + L'(x, y)}, \quad (2.14)$$

$$L_{d2}(x, y) = \frac{L_d(x, y)(1 + L_d(x, y)/L_{\text{white}}^2)}{1 + L_d(x, y)}, \quad (2.15)$$

其中式 (2.15) 的白點抑制項可避免高光被壓成灰階 [9]。Kao 等人實驗發現，對  $L_Y$  連續套用兩輪 (2.12)-(2.15) 可在不增加可見雜訊的前提下進一步提亮暗部。

### 3. 重組與正規化：將未經壓縮的高頻加回

$$Y_{\text{out}}(x, y) = L_{d2}(x, y) + H_Y(x, y). \quad (2.16)$$

再重新正規化至 8-bit 範圍輸出，以維持紋理對比與色彩忠實度。

### 4. 跨幀曲線共享：影片處理時，每 30 張影格僅首張完整更新色調曲線，其餘影格沿用同一曲線，可抑制亮度閃爍並將 VGA 解析度效能提高至 30 fps 以上 [9].

相較於單次全域 Reinhard，頻率分割式 TR 能在不顯著增加運算量下，於背光主體與高光區保留更多細節；對於本研究關注的人臉亦能提供更平衡的亮度映射，後續實驗將量化其對臉部偵測精確程度的提升。

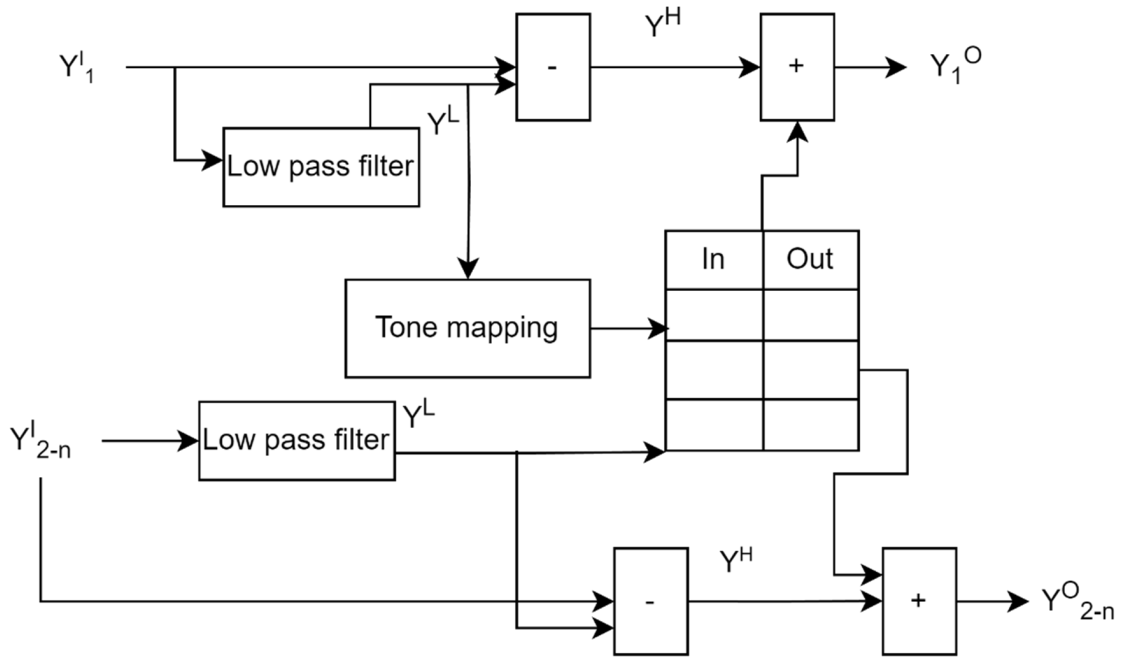


Figure 2.4: Tone Reproduction for Video 流程圖



Figure 2.5: `tr_video` 的細節提取：左為輸入之亮度/灰階影像，右為 `tr_video` 流程產生的細節圖 (detail map)。此圖僅示意細節保留與邊緣加強的效果。

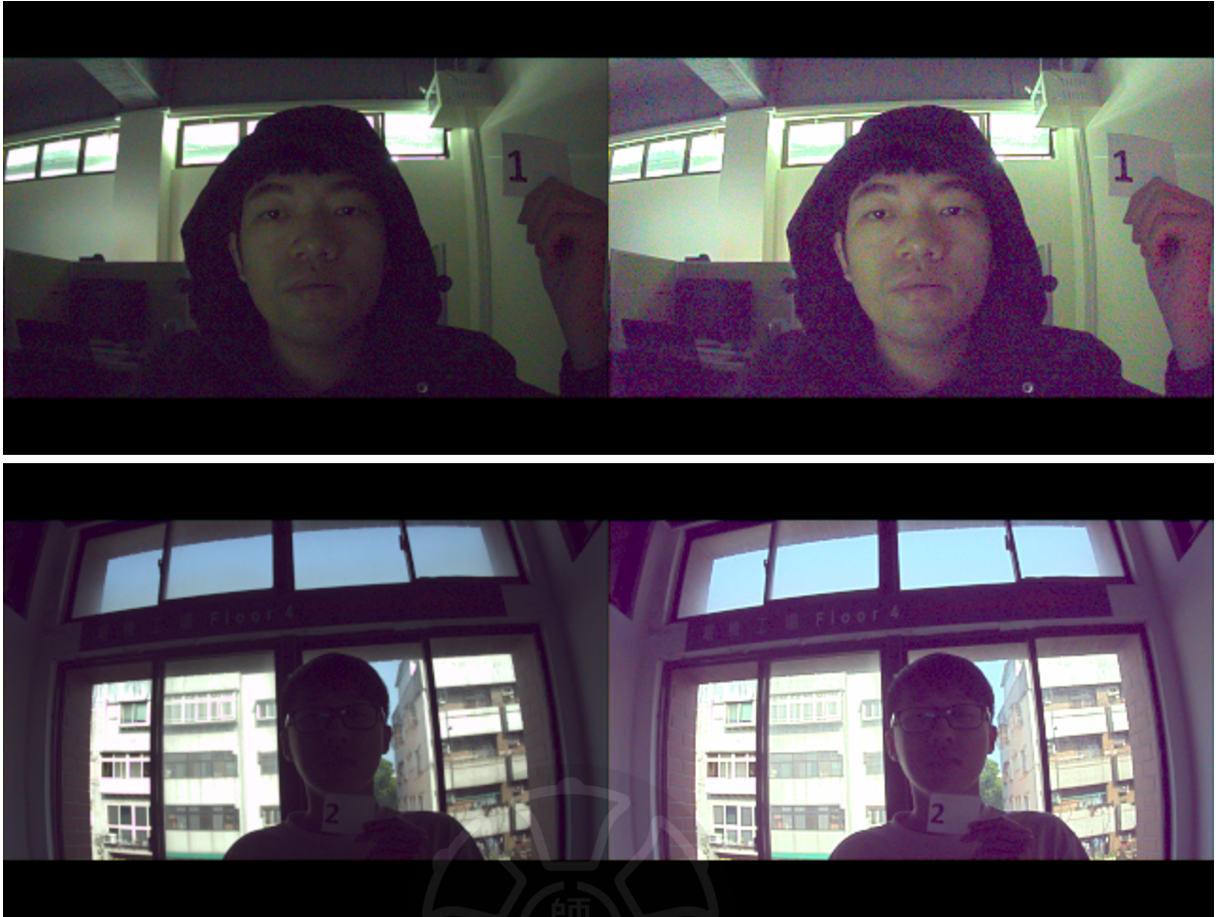


Figure 2.6: `tr_video` 範例圖：每一列為一個場景，左為原始影像，右為 `tr_video` 處理結果；上列為室內低光、下列為逆光場景。

### 2.2.2.2 對比受限自適應直方圖均衡

CLAHE[6] 係由 Zuiderveld 等人發展而來，用以解決傳統自適應直方圖均衡 (AHE) 在低亮度影像中容易放大雜訊、產生過度對比的問題。方法首先將輸入影像劃分為  $m \times n$  個不重疊的小區塊 (tiles)，每個區塊獨立計算 256-bin 灰階直方圖  $h_{ij}[r]$ 。為抑制單一灰階的像素過度集中，演算法對每個區塊設定一個限幅值

$$T = \alpha \frac{|\mathcal{T}|}{N_{\text{bins}}}, \quad (2.17)$$

其中  $\alpha$  為對比限制因子 (常取 2.0)。若  $h_{ij}[r] > T$ ，則將超出部分累加為溢出量，再平均分配回所有 bins，從而壓抑雜訊尖峰。完成限幅後，以累積分佈函數

(CDF) 產生區塊的查找表：

$$\text{LUT}_{i,j}(r) = \left\lfloor 255 \frac{\sum_{k=0}^r h_{i,j}[k]}{\sum_{k=0}^{255} h_{i,j}[k]} \right\rfloor. \quad (2.18)$$

並將區塊內像素灰階  $r$  映射為  $s = \text{LUT}_{ij}(r)$ 。

為避免區塊邊緣產生可見接縫，CLAHE 於映射後利用四鄰區塊之輸出值執行雙線性插值，計算最終像素值

$$S(x, y) = \sum_{(p,q) \in \{\text{LL,LR,UL,UR}\}} w_{pq}(x, y) s_{pq}(x, y). \quad (2.19)$$

其中  $w_{pq}$  為依據像素到四鄰區塊中心距離所定義的權重，滿足  $\sum w_{pq} = 1$ 。由於直方圖計算僅限於區塊內，整體時間複雜度約為  $\mathcal{O}(HW + mn N_{\text{bins}})$ ，易於在 GPU 或 FPGA 上平行實作。

在低光源與背光情境，CLAHE 能有效拉伸暗區灰階，同時透過限幅機制避免高頻雜訊被過度放大，且所需運算量仍可接受於嵌入式 SoC 平台。

### 2.2.2.3 區域對比增強

黃志祥等人 [10] 提出之區域對比增強 (Local Contrast Enhancement, LCE) 流程，先以雙邊濾波器移除感測雜訊，再計算對數域的區域對比圖，最後透過限幅直方圖均衡與相似濾波器平滑化，在不產生光暈的前提下同時提亮暗部與保留邊緣細節，特別適合低曝光單張影像。以下分述其主要步驟。

首先，將線性亮度  $Y(x, y)$  以雙邊濾波器  $\mathcal{B}_{\sigma_d, \sigma_r}\{\cdot\}$  產生平滑版本  $Y_B = \mathcal{B}\{Y\}$ ，並取鄰域平均  $V = \text{mean}_{\mathcal{N}(x,y)}(Y_B)$  作為局部亮度基準。區域對比定義為

$$L_C(x, y) = \log \left( \frac{Y_B(x, y) + 1}{V(x, y) + 1} \right) \quad (2.20)$$

能有效補償照明差異並壓縮整體動態範圍。由於高頻雜訊可能主導  $L_C$  的直方圖

分佈，作者先以限幅值  $T = \alpha|\text{tile}|/256$  ( $\alpha \approx 2$ ) 對 256-bin 直方圖實施剪頂，再以累積頻率函數做均衡化以拉伸對比圖動態範圍。

均衡後的  $L_H$  雖突顯暗部紋理，仍可能伴隨隨機雜訊。為抑制顆粒感，方法利用「相似濾波器」 $\mathcal{S}_{\alpha_1, \alpha_2}\{\cdot\}$ ——只在鄰域像素  $P$  與中心像素  $C$  之亮度滿足  $\alpha_1 Y_H(C) \leq Y_H(P) \leq \alpha_2 Y_H(C)$  (預設  $\alpha_1 = 0.9, \alpha_2 = 1.1$ ) 時才參與加權平均——得到平滑對比圖  $L_A$ 。

最終輸出亮度以原亮度與平滑對比值之幾何平均重組：

$$Y_{\text{out}}(x, y) = \sqrt{Y(x, y) L_A(x, y)} \quad (2.21)$$

作者並在後續套用攝影式 Tone Reproduction 進一步正規化，使影像對數平均亮度逼近 18 並於  $3888 \times 2592$  解析度的實驗中較全域 Reinhard、Retinex 與 Photoreceptor TM 在室內、戶外與夜間場景均呈現更高的暗部可見度且無明顯光暈。

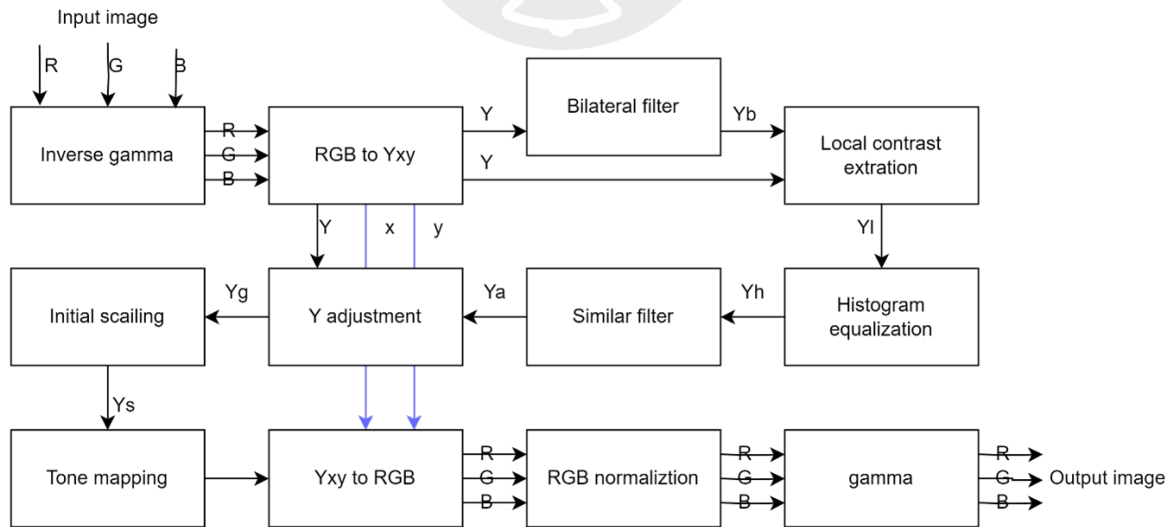


Figure 2.7: local contrast enhancement 流程圖

由於所有運算僅牽涉單尺度雙邊濾波、區域均值、一次直方圖均衡與低複雜度相似濾波，整體時間複雜度約  $\mathcal{O}(HW)$ ，易於在嵌入式 SoC 或 GPU 上行內並

行化；黃在 Core2 Duo E8400 上可於 2.6 Mpixel 影像達每張 170 ms 之處理時間，其中雙邊濾波為主要瓶頸。



Figure 2.8: LCE 範例圖：每一列為一個場景，左為原始影像，右為 LCE 處理結果；上列為室內一般光、下列為逆光場景。

#### 2.2.2.4 基於梯度範數的色調曲線

Kojima 與 Suetake 發現，單張影像若同時包含極亮與極暗區域 (wide-dynamic-range scene)，傳統 HE / CLAHE 或 Retinex 系列常帶來雜訊放大與黑白飽和等問題，且運算量偏高。為兼顧即時性與視覺品質，作者於 2024 年提出 *Gradient-Norm-Based Tone Curve* (GNB) [7]。

GNB 先對灰階亮度  $I$  架構「梯度加權直方圖」

$$h_w(l) = w_E(l) \sum_p \|\nabla I(p)\| \delta(I(p), l). \quad (2.22)$$

$$w_E(l) = \exp\left(-\frac{el}{l_{\max}}\right). \quad (2.23)$$

其中參數  $e$  控制平坦區抑制強度。接著以 Otsu 方法在高亮半區求得分界  $a$ ，據此計算累積分佈  $\alpha(l)$  並產生 tone curve

$$q(l) = \alpha(l) s(l) + (1 - \alpha(l))l, \quad s(l) \in \{0, a, l_{\max}\}. \quad (2.24)$$

考慮 tone curve 在  $\alpha$  斜率趨近平坦處可能削弱紋理，作者進一步以雙邊濾波取得局部平均  $\bar{I}$ ，再依  $\alpha(\bar{I}(p))$  做局部加權混和

$$I_{\text{out}}(p) = \alpha I(p) + (1 - \alpha) q(I(p)). \quad (2.25)$$

同時維持亮度與細節。

**實驗成效** 在 254 張 WDR 測試集中，GNB 於 LOE、NIQE、BIQME 三項指標皆優於 HE / CLAHE、SRIE、ZERO-DCE 等方法，黑白飽和比例亦最低。作者以 MATLAB 實作在 Ryzen 7 桌機處理  $900 \times 573$  影像僅 0.047s，較 Retinex 與曝光融合類演算法快一至兩個數量級，驗證其「低複雜度且即時」的特點。

**評註** GNB 透過「梯度作權」和「局域對比補償」避免過度增益與雜訊放大，同時保留 LUT 式 tone curve，便於硬體或 ISP 前端部署。本研究第二章已將 GNB 納入七種 Tone Mapping 基準，第四章亦證實其在背光場景對 mAP / mAR 的提升幅度與原文一致。

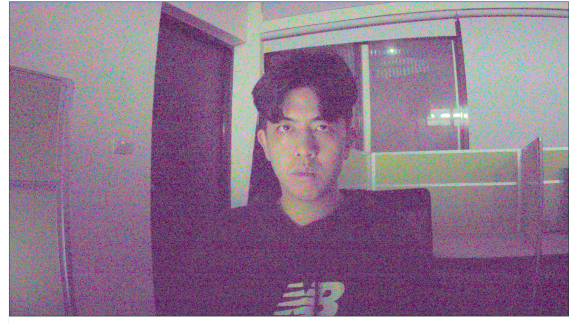
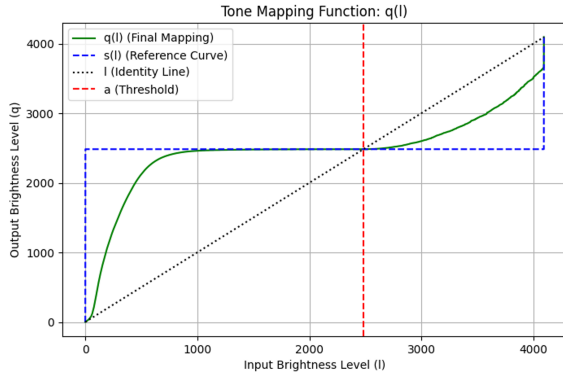


Figure 2.9: GNB 亮度壓縮示意圖

## 2.3 臉部偵測模型發展與輕量化趨勢

自 Viola-Jones 引入 Adaboost 級聯分類器以來 [11]，臉部偵測已歷經由傳統特徵（Haar、HOG、LBP）到深度卷積網路的世代交替。近年行動裝置、物聯網與邊緣攝影機需求激增，促使模型朝「即時、低耗能、可量化」方向演進 [12]。本節按偵測框架與輕量化手段，回顧代表性文獻，並說明其對後續研究的啟示。

### 2.3.1 評估指標：mAP 與 mAR

物件偵測常以 IoU（Intersection over Union）衡量預測框與真值框的重疊：

$$\text{IoU}(B_p, B_g) = \frac{|B_p \cap B_g|}{|B_p \cup B_g|}. \quad (2.26)$$

在 IoU 門檻  $t$  下，配對成功視為真陽性（True Positive, TP），否則為假陽性（False Positive, FP）；未被配對之真值為假陰性（False Negative, FN）。

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (2.27)$$

將所有偵測依信心分數由高到低排序，逐步累計 TP/FP 可得到 Precision-Recall (PR) 曲線；AP（Average Precision）為該曲線下的面積，記為  $\text{AP}_t^c$ （類別  $c$ 、IoU 門檻  $t$ ）。mAP（mean AP）則對 AP 做平均：

- **VOC 標準 (mAP@0.5)**: 僅在  $t=0.50$  計算一次 AP，對所有類別平均

$$\text{mAP}@0.5 = \frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} \text{AP}_{0.50}^c. \quad [13]$$

- **COCO 標準 (mAP@[.5:.95])**: 在  $t \in \{0.50, 0.55, \dots, 0.95\}$  (10 個門檻) 皆計算 AP，先對門檻平均、再對類別平均

$$\text{mAP} = \frac{1}{|\mathcal{C}|} \cdot \frac{1}{10} \sum_{c \in \mathcal{C}} \sum_{k=1}^{10} \text{AP}_{t_k}^c. \quad [14]$$

COCO 亦回報  $\text{AP}_{50}, \text{AP}_{75}$  與依物件面積分組的 APS, APM, APL (閾值為  $\text{small} < 32^2, 32^2 \leq \text{medium} < 96^2, \text{large} \geq 96^2$  畫素)。

### 2.3.1.1 AR 與 mAR

在固定 IoU 門檻  $t$  與每張圖允許的最大偵測數  $K$  (也記為  $\text{maxDets}$ , COCO 常用  $K \in \{1, 10, 100\}$ ) 下, **AR (Average Recall)** 定義為能達到的最高召回率, 記為  $\text{AR}_{t,K}^c$ ; **mAR** 為跨門檻 (與類別) 的平均:

$$\text{mAR}@K = \frac{1}{|\mathcal{C}|} \cdot \frac{1}{10} \sum_{c \in \mathcal{C}} \sum_{k=1}^{10} \text{AR}_{t_k, K}^c. \quad (2.28)$$

mAP 著重「在不同召回與 IoU 要求下維持高精確率」; mAR 反映「在偵測數上限  $K$  下最多能找回多少真值」。

### 2.3.2 單階段檢測器

單階 (one-stage) 臉部偵測多以 Singe Shot MultiBox Detector (SSD) [15] 為雛形: 利用卷積特徵金字塔在多個解析度同時回歸邊界框與分類分數, 省去 RPN (Region Proposal Network; 在共享特徵圖上以小型滑動網路對多個錨框預測「物件性」與邊界框回歸, 用以產生類別無關的候選框) 及二階篩選流程, 天

然兼顧速度與泛化性。針對臉部特有的尺度差異與密集分布，研究者沿三條主軸推進：特徵金字塔強化、輕量骨幹設計、以及量化與剪枝。

早期改進側重於提高小尺度召回率。S<sup>3</sup>FD[16] 在每層特徵圖均加入密集 anchor 與 *max-out background label*，可於 WIDER\_Face Hard 子集將 mAP 提升九個百分點。RetinaFace[17] 則進一步在 detection head 引入五點人臉關鍵點監督與 dense regression，並配合自適應 anchor 分配，於 640×640 解析度即達 96% Val\_Hard mAP，而推論速度仍保持每張影像 23 fps。

隨著邊緣裝置需求興起，輕量骨幹成為另一焦點。UltraFace[18] 採用 MobileNet0.25 替換 VGG，並縮減輸入至 320×240，在 Snapdragon 820 行動平台仍能維持 27 fps 與 80% mAP。後續多家 SoC 業者以此為基底，導入 depthwise separable、ShuffleNet v2 乃至 Ghost 模組，將參數量壓縮至 1-2 MB 卻僅損失 1-2 個 mAP，足見 SSD 架構之可插拔性。

第三條主軸是 量化與剪枝。以 RetinaFace 為例，SmoothQuant[19] 在不調整 權重分布的情況下即可將 FP32 模型直轉 INT8，mAP 損失低於 0.3；結構性通道 剪枝亦被證實可在 Arm Cortex-A55 將 UltraFace 延遲降低 35%。這些結果顯示， SSD 系列憑藉單階特徵金字塔與 anchor 機制，能在精度、延遲與模型大小之間 取得平衡，仍是目前行動端與邊緣端臉部偵測的主流選擇。

本研究後續實驗即以 RetinaFace-MobileNet0.25 與 UltraFace 為基線，結合第 二章所述多款色調映射演算法，比較其在低光源與背光場景的人臉檢測效能；細 節將於第四章呈現。

Table 2.1: 輕量 SSD 系列臉部偵測模型比較

模型	backbone	輸入	Params	FLOPs	mAP <sub>Hard</sub>	FPS(裝置)
RetinaFace-M0.25[17]	MobileNet0.25	640 <sup>2</sup>	1.7 M	1.09 G	80	23@(TITAN Xp)
UltraFace[18]	MobileNet0.25	320×240	1.1 M	0.09 G	39.5	27@(SD 820)
S <sup>3</sup> FD-Lite[16]	VGG-Slim	640 <sup>2</sup>	24 M	1.92 G	85.2	18@(TX2)

### 2.3.3 無錨框與輕量錨框式偵測架構

以中心點或邊界框四頂點偏移量直接回歸，省去錨框 (anchor) 設計、匹配與冗餘 NMS 開銷。然而在臉部偵測領域，許多模型為兼顧精度與部署彈性，仍保留 *anchor* 設計，改以 (i) 重參數化骨幹、(ii) 高效標籤分配、(iii) 量化／剪枝，來達到邊緣即時需求。*SCRFD* 與 *YOLO-face* 即屬此「輕量 anchor-based」代表。

Guo 等人提出的 SCRFD (可伸縮且緊湊的人臉偵測器；Scalable and Compact Face Detector) 聚焦兩個問題：其一，在邊緣裝置上兼顧小算力與對小臉／密集場景的召回；其二，傳統以交並比 (IoU) 門檻進行樣本匹配在尺度差異大時不穩。其技術要點如下：

- 重參數化主幹網路 (re-parameterizable backbone)：訓練期採多分支卷積提高表達力；推論期將多分支等效折疊為單一路徑卷積，降低分支帶來的計算與記憶體訪問成本，減少延遲。
- 最佳傳輸分配 (Optimal Transport Assignment, OTA)：把「正負樣本指派」改寫為最小成本最佳傳輸問題，同時考量分類與定位代價，自動挑選正樣本，避免僅以 IoU 門檻造成的小臉／擁擠場景分配失真。
- 多尺度錨框與輕量特徵金字塔：提供多個算力檔位，便於依硬體需求選型。

在 WIDER FACE 的 Hard 子集上，SCRFD 相較 RetinaFace 有小幅領先，且計算量可低至約 0.5 GFLOPs (GFLOPs：十億次浮點運算的量級)。官方報告顯示 SCRFD-2.5G 在 NVIDIA Jetson TX2 約 60 影格／秒；其 8 位元整數量化版本 (INT8) 在 Rockchip RK3588 仍可達約 50 影格／秒，具備良好邊緣部署潛力。Liang 等人基於 YOLOv5 針對人臉偵測進行輕量化改寫，提出 YOLO-face，欲在低功耗裝置上同時降低參數與 FLOPs，並提升對小臉、側臉與遮擋的定位穩定

性。其技術亮點包括：

- 錨框式解耦輸出頭 (decoupled head)：分類與框回歸分支分離，減少梯度互相干擾，使訓練更穩定、收斂更快。
- 部分洗牌模組 (Partial Shuffle Blocks)：以分組／通道洗牌與深度可分離卷積取代傳統瓶頸 (bottleneck)，顯著降低參數量與乘加次數，同時維持感受野與特徵融合。
- 座標注意力 (Coordinate Attention)：在通道注意力中引入座標編碼，保留位置信息，對受遮擋或細長結構的臉部區域更敏感，改善定位。

小型版本 YOLO-face-S 僅約 1.3 百萬參數、計算量約 1 GFLOPs；在 WIDER FACE 驗證集 Hard 切分約 89.6% 的檢測精度 (mAP@0.5)。於 Jetson Nano (功耗低於 10 W) 實測約 45 影格／秒，較同級 Mobile-RetinaFace 約快三成。

雖然無錨框架構 (anchor-free) 在設計上較為簡潔，但透過重參數化與最佳傳輸分配等策略，輕量的錨框式模型亦能在精度與速度之間取得良好平衡；再配合 8 位元整數量化 (INT8) 可進一步降低記憶體占用並提升吞吐，符合邊緣裝置的部署需求。

綜上，anchor-free 架構雖簡化設計流程，但經由重參數化與高效分配策略，輕量 anchor-based 模型亦能在精度與速度上取得平衡，並透過 INT8 量化進一步壓縮記憶體占用。

### 2.3.4 訓練策略：預訓練權重、資料擴增與再訓練

輕量臉部偵測器之所以能兼顧速度與精度，除了網路結構設計，訓練流程亦是關鍵。廣泛驗證有效的三步訓練流程：

- (1) 大規模預訓練 先使用 ImageNet 或 COCO[20, 14] 等通用資料集進行 su-

pervised 預訓練，再將權重遷移到 WIDER Face[21]，可讓輕量骨幹（MobileNet、ShuffleNet 等）獲得穩定的低層特徵，縮短收斂時間並提高收斂上限。

(2) 任務導向資料擴增 基礎隨機翻轉、尺度縮放之外，我們採用 Howard[22] 提出的色彩抖動（亮度／對比／色溫漂移），並混合 Cutout[23] 與 GridMask[24] 遮擋策略，以模擬暗光與遮擋場景中常見的曝光不均、局部缺失。此類擴增已被多份實務報告證實能顯著提升 Hard 子集召回率。

(3) 目標域再訓練與知識蒸餾 完成初步微調後，固定骨幹、僅微調 detection head，隨後在較小學習率下解凍全網路以細調參數。若需進一步壓縮模型，參考 Hinton 等人提出的知識蒸餾 [25]，以高精度教師（如 ResNet50 + RetinaFace）指導輕量學生（MobileNet0.25 + RetinaFace），往往可在不增加推論成本的前提下再取回約 1 個 mAP。

以上三階段流程在行動端與 IoT 裝置已成為實務標配，亦為本研究評估色調重現管線奠定穩固的訓練基線。

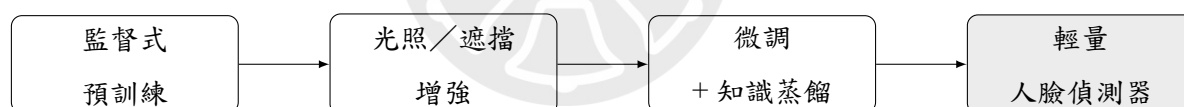


Figure 2.10: 三步訓練流程示意。依序為監督式預訓練、光照／遮擋增強與目標域微調（含知識蒸餾），最終得到可部署的輕量人臉偵測器。

Table 2.2: 訓練策略與代表性工作（以大型通用資料集預訓練、照明增強與知識蒸餾為例）[14, 21]

策略	代表文獻	重點做法／效果
預訓練	COCO → WIDER FACE	建立通用特徵；縮短收斂時間
照明增強	照明增強 (Illumination Augmentation)	曝光隨機化 + 遮擋剪裁；Hard 子集召回率 ↑ 3.7%
知識蒸餾	知識蒸餾 (Knowledge Distillation)	ResNet-50 → MobileNet-0.25；Hard 子集 mAP +1.0

如表 2.2 所示，預訓練能提供穩定低層特徵，照明導向擴增可顯著強化在暗光與遮擋場景的召回率，而知識蒸餾則進一步收復因輕量化損失的精度。



## 第三章 研究方法

第二章已詳述七種影像增益策略演算法之原理。本章聚焦於系統整合、實作細節與訓練流程，說明如何將色調映射與輕量臉部偵測器結合，以因應低光／背光場景。<sup>1</sup>

### 3.1 系統概觀

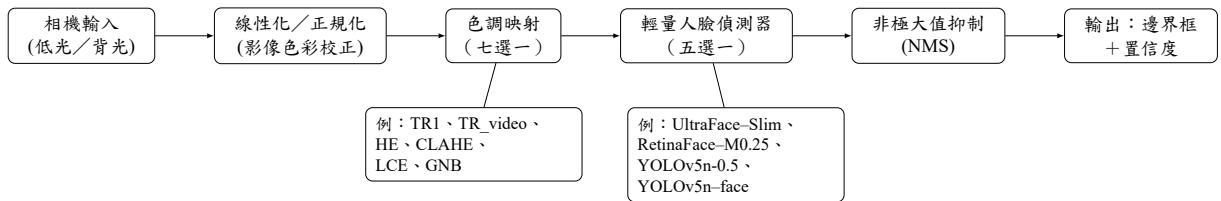


Figure 3.1: 邊緣推論流程：線性化與正規化後，選擇一種色調映射，再由輕量偵測器推論並經非極大值抑制 (Non-Maximum Suppression, NMS) 輸出結果。

圖 3.1 顯示本研究的雙模組架構——本研究的系統採前後兩段式：前段以色調映射負責增亮與對比拉伸，後段以輕量人臉偵測器產生邊界框、關鍵點與置信度。兩段之間以統一的張量介面對接（ $H \times W \times 3$ 、0-1 正規化的浮點影像），因此前處理與偵測器可以在不改動權重的情況下互換，具備可插拔與易於部署的特性。

以下兩小節會針對原始影像的處理流程以及非極大抑制做介紹。

#### 3.1.1 輸入資料與前端影像處理模組

本節說明我們在邊緣情境下所採用的前端影像處理（ISP）模組之目的與設計原則。研究目標是：在不改變影像語意（物件類別、數量與幾何對應不變）的前提下，將輸入影像正規化到一致的亮度域，並提供色調映射等增益策略的可調接口，以量化不同前處理對輕量化人臉偵測器之實際助益。

<sup>1</sup>文獻中亦有 Tone Reproduction 之稱呼，本文統一使用 Tone Mapping。

### 3.1.1.1 原始影像處理流程

本研究所有實驗影像均由 RTK rts 感光元件擷取，解析度  $1920 \times 1080$ ，位元深度 12 bit (Bayer RGGB Raw)。

為確保各色調映射演算法在統一亮度域內運作，於套用 TM 之前先進行線性化與灰階正規化，流程如下：

1. **解馬賽克 (Demosaic)** 使用 Edge-Aware 插值將 Bayer 單通道  $\mathcal{R} \in [0, 4095]^{H \times W}$  轉為線性 RGB 影像  $I_{\text{lin}} \in [0, 1]^{H \times W \times 3}$ ，
2. **等比縮小至  $320 \times 240$**  為降低 FLOPs，以 Lanczos 內插將  $I_{\text{lin}}$  等比縮至  $320 \times 180$ ，再於上下補黑框  $2 \times 30$  以維持 16:9 輸入尺寸  $320 \times 240$ ，實測能將 UltraFace 前向計算量由 0.38 G 降至 0.09 G。
3. **白平衡與色彩校正** 依感測器增益計算  $(g_R, g_G, g_B)$ ，隨後乘以  $3 \times 3$  色彩校正矩陣  $\mathbf{C}$  (D65→sRGB)：

$$I_{\text{cc}} = \mathbf{C} \begin{bmatrix} g_R & 0 & 0 \\ 0 & g_G & 0 \\ 0 & 0 & g_B \end{bmatrix} I_{\text{lin}}.$$

得到線性 sRGB 空間影像  $I_{\text{cc}} \in [0, 1]^{320 \times 240 \times 3}$ 。

4. **Tone Mapping** 七種 Tone Mapping 演算法  $f_{\theta_i}$  逐像素作用於  $I_{\text{cc}}$ ，產生最終輸出  $\tilde{I} = f_{\theta_i}(I_{\text{cc}})$ ，隨後餵入輕量偵測器  $g_{\phi_j}$ 。

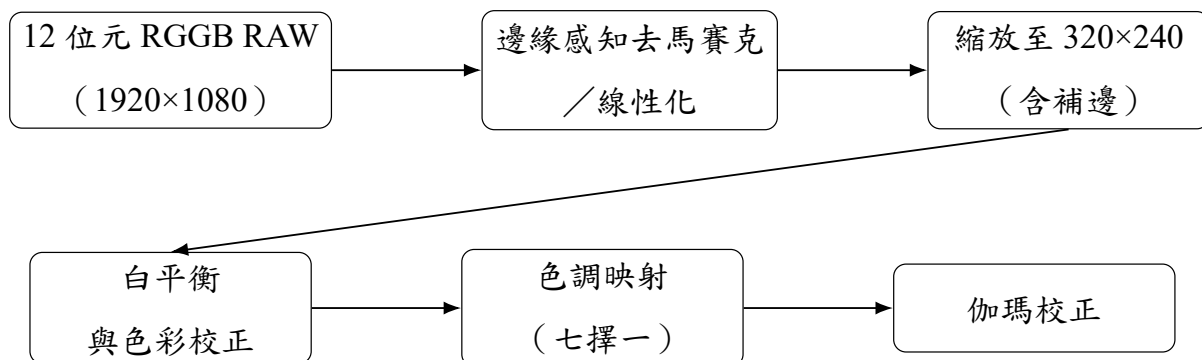


Figure 3.2: 前端影像處理流程示意圖（至色調映射與伽瑪校正）

此四步驟確保所有色調映射在同一線性 sRGB 空間內運作，避免早期 ISP Gamma 壓縮導致的動態範圍損失，並將解析度與計算量降至邊緣裝置可承載範圍，為後續實驗提供一致輸入。

### 3.1.1.2 色調映射之實作與參數設定

本小節僅交代七種色調映射策略在 12-bit 輸入上的落地細節與參數選擇，理論機制已於第二章說明，在此不再重述。所有程式以 Python + NumPy 撰寫，輸入為圖 ?? 所產生之  $I_{cc} \in [0, 255]^{320 \times 240 \times 3}$ 。由於感測器為 12-bit，所有需計算直方圖者（HE / CLAHE / LCE / GNB）皆將 bin 數擴充至 4096，避免量化誤差。

Table 3.1: Tone Mapping 方法、參數與設定理由（統一符號與單位）

方法	目的 / 機制（一句話）	主要參數（域 / 單位）	本研究設定與理由
TR1 (Photo-graphic)	將曝光映射至中灰，壓縮高亮並保留暗部（全域單調曲線）。	$\alpha$ ：中灰係數 ( $0 < \alpha < 1$ )；LUT 長度（階數）。	$\alpha = 0.18$ （兼顧暗部可見度與亮部不飽和）；LUT=4096（對應 12-bit）。
Global HE	依亮度 CDF 做全域均衡，提升動態範圍。	—	無需超參數；結果裁切至 $[0, 1]$ 避免過度伸展。
CLAHE	分塊均衡並限制塊內對比，抑制噪點 / halo（局部）。	tileGrid：分塊大小 ( $m \times n$ ，像素)；clipLimit > 0。	tileGrid=8×8（每塊約 40×30 px，能覆蓋最小人臉 ROI）；clipLimit=2.0。
TR_video	場景平均與局部細節的雙段曲線，暗部提升、亮部保護。	$w_{\text{mean}}$ ：平滑視窗（像素）； $\alpha$ ：混合權重（自適應）。	$w_{\text{mean}} = 3$ （減少閃爍）； $\alpha$ 自適應（依直方圖調整）。
LCE	對數域提升局部對比，並做全域均衡。	$k$ ：視窗大小（像素）	$k = 5 \times 5$ ,
GNB	以梯度強度導引的亮度重映射，搭配邊緣保留濾波。	$a$ ：門檻（式 (18) 自適應）；雙邊濾波 $d, \sigma_s, \sigma_r$ 。	$a$ 自適應； $d = 5, \sigma_s = 10, \sigma_r = 20$ （保邊平滑）。
Gamma	幕次曲線做感知均衡或（反）線性化。	$\gamma > 0$ ；LUT 長度（階數）。	$\gamma = 0.45$ （輸出端編碼；對應相機 $\gamma \approx 2.2$ 的反運算）；LUT=4096。

註：所有操作在線性 sRGB 域；12-bit 影像採 4096 階 LUT。參數域以  $[0, 1]$  或像素為主。

執行時間（Jetson Nano 1.4 GHz CPU） TR1 1.6 ms，HE 1.3 ms，CLAHE 4.9 ms，

TR\_video 4.2 ms，LCE 6.1 ms，GNB 5.4 ms。相較 UltraFace 前向 23 ms，前端開銷 <20%，不影響即時性。



Figure 3.3: 低光情境：七種色調映射之增益示例。排列順序（由左至右、由上而下）：左上起：原圖、TR1、HE、TR\_video、CLAHE、LCE、GNB。



Figure 3.4: 背光情境：七種色調映射之增益示例。排列順序（由左至右、由上而下）：左上起：原圖、TR1、HE、TR\_video、CLAHE、LCE、GNB。

如圖 3.3 及圖 3.4 所示，TR\_video 與 GNB 在不顯著增加雜訊的情況下，最能平衡暗部能見度與全域對比。第四章將以 mAP / mAR 定量驗證各法對五款偵測器之影響。

### 3.1.2 非極大值抑制

**用途與位置**——NMS (Non-Maximum Suppression) 位於偵測器輸出之後，用來去除對同一實體的重複框。若不處理，密集錨框／網格會對同一張臉產生多個高度相近、彼此重疊的預測，這些重複預測在評估時會成為 FP，拉低精確率與

mAP。

**基本流程 (Greedy NMS, 類別內):**

1. 依每個候選框的信心分數由高到低排序。
2. 取當前最高分的框加入輸出集合。
3. 對其餘候選框，凡與該框的重疊度超過門檻  $\tau$  者一律移除。
4. 重複步驟 2-3，直到沒有候選框為止。

此法簡單高效 (向量化後常見  $O(n^2)$  以內的實作， $n$  代表該張影像相同類別框的數量)，實務上通常先套用信心分數門檻再做 NMS，以降低計算量。

去除重複框能夠降低 FFP，進而提升精確率；同時以單一框表徵同一張臉，有助於後續視覺化、追蹤與匯出；輸出更精簡亦可減少 I/O 與後處理成本。NMS 的門檻  $\tau$  反映精確率／召回率的權衡： $\tau$  較低 (抑制較積極) 通常使精確率上升、召回率下降； $\tau$  較高 (抑制較寬鬆) 則相反。

本文設定除特別說明外，採類別內 NMS，先以信心分數 (如 0.25) 過濾，再以  $\tau = 0.45$  執行 NMS；每張圖最多保留 100 個輸出。擁擠場景亦評估 Soft-NMS (以衰減分數代替刪除) 對平均召回率 (mAR, mean Average Recall) 的影響。

## 3.2 輕量臉部偵測器

本節介紹本研究評估的五款輕量模型，說明其架構差異、推論設定以及 UltraFace 重新訓練的動機與配置。

### 3.2.1 模型概覽

選擇依據為：開源可獲、行動／邊緣部署友善、近五年代表作。表 3.2 匯整各模型之骨幹、是否輸出五點關鍵點、輸入解析度與檔案大小。

Table 3.2: 輕量人臉偵測模型規格與資源需求。檔案大小以 fp32 之 PyTorch 檔 (.pth) 計；運算量以乘加次數 (MACs) 表示，單位為百萬 (M)。

模型	主幹網路	關鍵點	輸入大小	年份	檔案大小 (MB)	運算量 (M MACs)
BlazeFace	MobileNet	✓	320×240	2019	0.7	228
UltraFace-Slim	MobileNetV2	—	320×240	2019	1.1	90
YOLOv5n-0.5[? ]	CSPDarkNet	✓	320×320	2022	1.1	571
RetinaFace-M0.25	MobileNet0.25	✓	320×320	2020	1.7	193
YOLOv5n-face	CSPDarkNet	✓	320×320	2022	13.9	2111

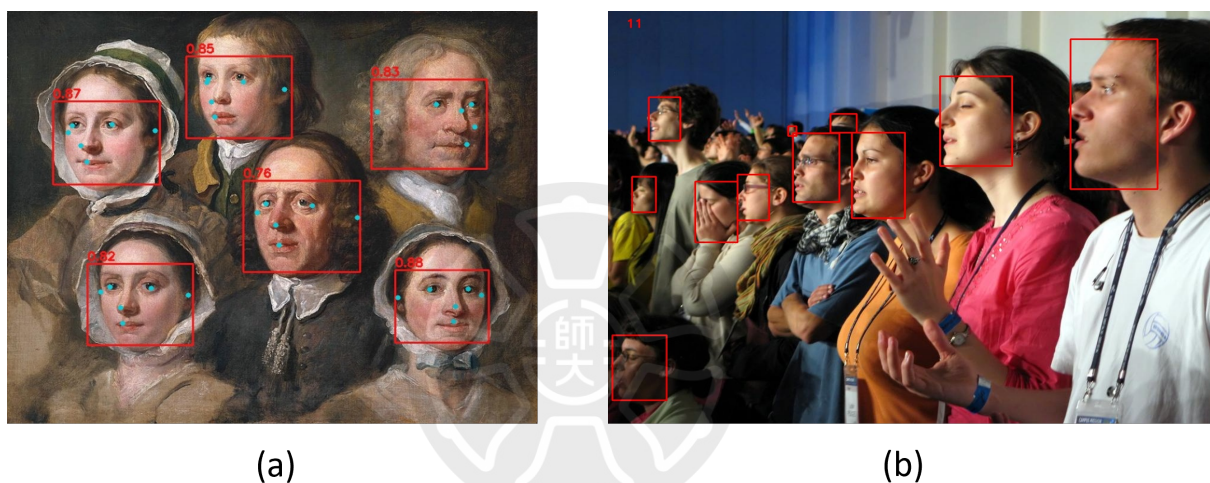


Figure 3.5: 質性比較。(a) 同時輸出五點關鍵點與邊界框的模型；(b) 僅輸出邊界框的模型。兩組範例顯示在「多臉且姿態各異」(左) 與「擁擠側臉」(右) 情境下仍能正確定位人臉。

### 3.2.2 模型設計差異

**BlazeFace** 採極簡 *Anchor-based SSD*，backbone 僅 10 層 depth-wise MobileNet-Lite，偵測頭同時回歸邊界框與五點關鍵點。**UltraFace** 為 MobileNetV2-SSD，本組中唯一不輸出關鍵點的模型，因此檔案最小 (1.1 MB)。**RetinaFace-M0.25** 延用 MobileNet0.25，保留 landmark 分支並以多層特徵金字塔強化小尺度召回，故參數量略高 (1.7 MB)。兩款 **YOLOv5n** 皆為 *Anchor-based YOLO* 且使用 decoupled head；**YOLOv5n-0.5** 在精簡通道 (0.5×) 後仍保留五點 landmark regression，模型 1.1 MB；**YOLOv5n-face** 進一步加入注意力與更寬 head，尺寸因

此增至 13.9 MB。

### 3.3 訓練流程與資料擴增

本節交代影像增益資料集的製作、UltraFace 的重新訓練設定，以及各偵測器權重的使用策略。之所以在本研究階段重新訓練偵測器，源於我們在極端情境（低照度、強背光）下的觀察：部分色調映射增益會非預期地放大雜訊與邊緣暈圈/條帶（banding/halo），形成與所選增益策略耦合的偽特徵，進而導致偵測失準或漏檢。這反映兩層落差：(i) 以 WIDER Face 預訓練的現成模型，其訓練分佈多為一般曝光的網路影像，與我們「經 ISP 與 TM 處理後的影像」存在分佈落差（domain shift）；(ii) 前處理管線與偵測器之間存在演算法耦合，當前處理改變影像統計時，僅靠凍結的既有權重難以適應。



Figure 3.6: 增益反例：未增益可偵測，套用 HE 後反而漏檢。左圖為未套用增益的輸入，偵測器能正確框出人臉；右圖為套用全域直方圖均衡（HE, Histogram Equalization）後的結果，因暗部雜訊被放大且局部對比分佈改變，偵測器未輸出有效框。此例說明前端增益與偵測器權重存在耦合，過度增益可能降低精確率／召回率。

基於此，我們提出假設：若讓偵測器在「增益域」上學習（即在與推論時一致的 ISP / 色調映射後的影像上訓練），則其對雜訊放大與增益耦合的敏感度將降低，於低光/背光子集可獲得較穩定的  $mAP$  /  $mAR$ 。為驗證此假設，我們建立「影像增益資料集」，對每張原始場景產生多種 TM 版本，標註保持不變；接著以 UltraFace-Slim（參數最少、與  $320 \times 240$  相容且程式碼易更換資料集）進行重訓，並與使用原始 WIDER Face 權重的模型在相同評估協定下比較其前端增益的效益

與穩健性。另四款偵測器主要維持原權重以評估前端對不同架構之泛化影響；其重訓版本與完整數據置於附錄以供參考。

本節說明影像增益資料集之製作流程、UltraFace 重新訓練設定，以及五款偵測器在本研究中的權重配置。

### 3.3.1 資料集與切分

**公開資料集與預訓練權重：**本文不以 *WIDER Face* 重新訓練模型，除 UltraFace 之外，其餘四款偵測器皆採用各作者／官方 gitHub 釋出的 *WIDER Face* 預訓練權重作為初始化並在本研究中保持固定。*WIDER Face* 資料集約含 12 萬張影像、39 萬個人臉標註（訓練集 *train*、驗證集 *val*），上述預訓練權重係於 *train* 上訓練並以 *val* 選模而得。預訓練階段的超參數（輸入尺寸、錨框配置、學習率、批次大小、訓練輪數等）均沿用各原始論文／官方實作；其摘要列於表 3.3。

**低光 / 背光資料** 本研究另行拍攝 1000 張 12-bit RGGBRAW（解析度 1920×1080，距離 0.3–1.5 m，臉部 600–115 px），並依照圖 3.1 流程線性化後，套用七種 Tone Mapping（TR）增益  $\{f_{\theta_i}\}_{i=1}^7 = \text{None, TR1, HE, CLAHE, TR\_video, LCE, GNB}$ ，生成 7000 張合成影像；其中 20%（隨機抽）作為 *LowLight-Test*，餘下屬 *LowLight-Train*。

### 3.3.2 UltraFace 兩階段訓練流程

**UF-7K-Joint** 將 1000 張 RAW 之七倍增益影像合併（共 7000），一次性訓練一個模型。

**UF-1K×7-Split** 以七種 TR 逐一對 1000 張 RAW 增益，分別訓練七個模型，檢驗「訓練-測試使用同一增益」的 upper-bound。

**為何僅重訓 UltraFace** UltraFace-Slim 參數最少（1.1 MB），與 320×240 輸入解析度相容度最高，且官方程式碼易於置換資料集。其餘四款保持原始 *WIDER*

Table 3.3: UltraFace-Slim 重新訓練超參數 (UF-7K-Joint / UF-1K×7-Split 相同)

基本設定		優化與排程	
影像尺寸	320×240	Optimizer	SGD
Batch size	24	Momentum	0.9
訓練 epoch	200	Weight decay	$5 \times 10^{-4}$
Checkpoint	每 5 epoch	LR(init)	0.01
排程器		驗證與偵錯	
Scheduler	Multi-step	Validation epoch	5
Milestones	80, 100	Debug steps	100
Loss 與其他		硬體環境	
損失函數	MultiBox (neg/pos=3)	GPU	cuda:0 (可 多 GPU)
IoU overlap thr.	0.35	cudnn.benchmark	True

Face 權重，用以評估「前端 Tone Mapping 對不同架構之泛化影響」。為測試本系統的穩健性，我們也另外用此訓練方法訓練本研究其餘四款模型，將結果放於附錄。

### 3.3.2.1 推論設定固定化

為確保不同模型可比性，推論統一採用下列設定：

- **等比縮放 + Zero-Pad (letterbox)**：先將輸入影像等比縮至最長邊對應之網路邊長（本研究為  $320 \times 180$ ），再以像素值 0 於上下各補邊 30 像素，得到固定尺寸  $320 \times 240$ ；若模型需求  $320 \times 320$ ，則左右補邊成方形。此作法保持幾何比例，標註只需加固定偏移量，黑邊視為背景不參與配對。
- **正規化**：輸入縮放至  $[0, 1]$ 。
- **統一 NMS 門檻**：類別內 NMS 使用

$$\text{prob\_thr} = 0.5, \quad \text{iou\_thr} = 0.5.$$

其中 `prob_thr` 為置信度門檻——分類機率（或物件性）低於 0.5 的候選框

先行丟棄，以抑制低分雜訊框；`iou_thr` 為重疊門檻——當兩框重疊度超過 0.5 僅保留高分者，以移除同一臉的重複預測。兩值取 0.5 可與 AP@0.5 的評估標準對齊並維持穩定的 Precision/Recall 取捨。

### 3.3.2.2 訓練流程與資料擴增小結

本節完成低光／背光資料集製作、七種增益策略與 UltraFace 的訓練設定，並將推論與 NMS 門檻固定（`prob_thr=0.5`，`iou_thr=0.5`）作為後續比較基線。

### 3.3.3 評估指標與實驗配置

為了公平與可重現的比較，本節先說明我們採用的偵測評估指標與其計算方式，接著交代兩套測試資料的切分規則與輸入標準化；所有模型一律使用相同的前處理（等比縮放 + Zero-Pad 至  $320 \times 240$ ）與後處理門檻（`prob_thr=0.5`、`iou_thr=0.5`；見 §3.3.2.1）。

#### 3.3.3.1 偵測評估指標

**AP@0.5**：依官方流程在 IoU 門檻  $\tau = 0.5$  下，以信心分數遞減累積得到 PR 曲線，其面積為 AP；本文以整體 AP@0.5 為主指標。**mAP@0.5**：跨多設定之 AP@0.5 的算術平均（單類別：人臉）。**mAR@0.5**：在  $\tau = 0.5$  計算  $\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$ ，於各測試分組取平均，反映漏檢情形。**推論成本**：報告模型檔案大小與 MACs (FLOPs)。

#### 3.3.3.2 測試資料切分

- **LowLight-Test**：由 300 張 RAW 影像各產生七種 Tone Mapping 版本，隨機抽取 10% 作為測試；條件為照度  $\text{lux} < 4$  或畫面多數區域顯著欠曝。
- **Back-Light-Test**：另取 300 張強背光場景，同樣產生七種增益版本，用於評估逆光人臉之改善。

兩套測試影像在輸入偵測器前皆採等比縮放 + *Zero-Pad* 至  $320 \times 240$ ，並固定  $\text{prob\_thr}=0.5$ 、 $\text{iou\_thr}=0.5$ （見 §3.3.2.1）。

### 3.3.3.3 小結

本章統一了推論前處理與門檻、定義 AP/mAP 與 mAR，並給出兩套測試切分，作為第四章性能比較之共同基線。



## 第四章 實驗結果與分析

本章所有實驗均沿用 CHAPTER 3 既定之評估指標、資料切分與推論門檻設定：偵測表現以兩個指標呈現， $mAP_{0.5}$ ——IoU=0.5 下的平均準確率； $mAR_{0.5}$ ——相同 IoU 下的平均召回率。測試集為自建 LowLight / Back-Light，統一閾值  $prob\_thr = 0.5, iou\_thr = 0.5$ 。

在此基礎上，我們依序比較：

1. 七種 Tone Mapping (TM) 對五款輕量偵測器的 AP 影響 (§4.1)。
2. UltraFace 兩階段再訓練 (UF-7K / UF-1K×7) 對低光與背光場景的精度提升 (§4.2)。
3. 成功與失敗的質性分析，說明 TM 與偵測器之互補 (§??)。

接下來首先呈現 Tone Mapping 與偵測效能之量化結果。

### 4.1 色調映射與預訓練模型偵測效能

本節首先以  $mAP_{0.5}$  與  $mAR_{0.5}$  兩項指標，比較七種 Tone Mapping (TM) 策略對五款預訓練偵測器在背光與低光子集之影響，隨後從平均值與極端案例切入，討論 TM 與網路架構的互補性。

### 4.1.1 mAP/mAR 矩陣

Table 4.1: 背光測試集  $mAR_{0.5}$  (七種 Tone Mapping  $\times$  五款預訓練模型)

Tone Mapping	BlazeFace	RetinaFace-M0.25	UltraFace-Slim	YOLOv5n-0.5	YOLOv5n-face
GNB	0.193	0.362	0.193	0.239	0.232
HE	0.157	0.365	0.210	0.230	0.238
LCE (Local Con.)	0.156	0.334	0.180	0.256	0.235
None (Baseline)	0.116	0.217	0.165	0.137	0.207
TR1 (Reinhard)	0.132	0.257	0.162	0.213	0.224
TR_video	0.141	0.280	0.168	0.225	0.228
CLAHE	0.120	0.244	0.150	0.137	0.176

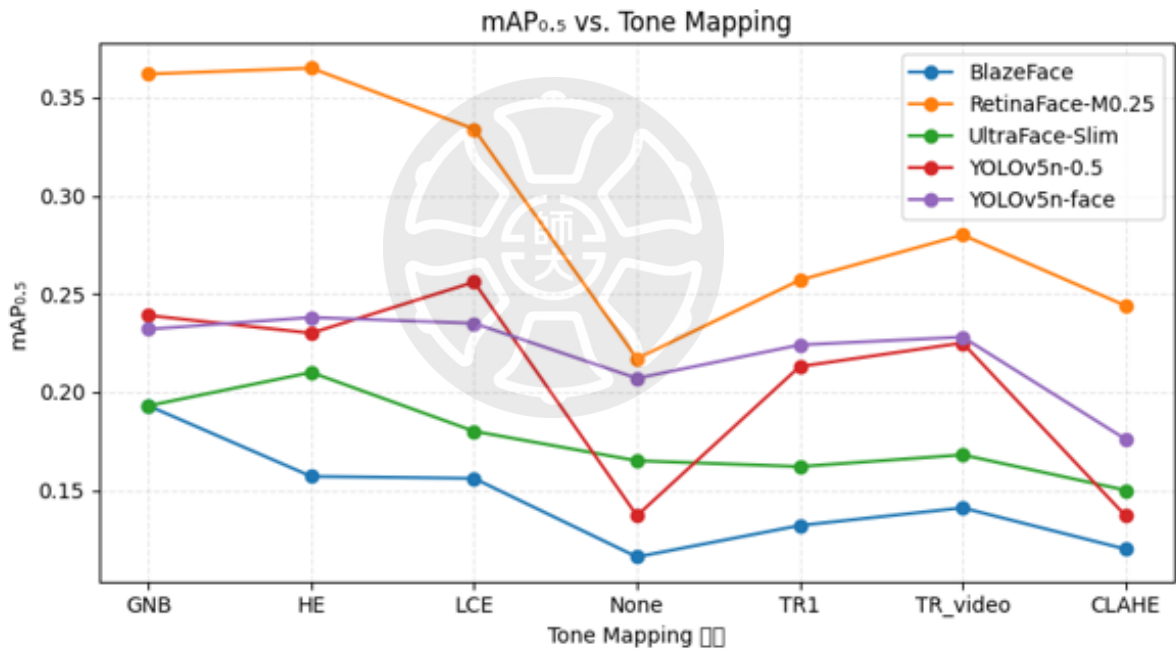


Figure 4.1: 色調映射對輕量人臉偵測器之效能 ( $mAP@0.5$ )。橫軸依序為 GNB、HE、LCE、無增益 (None)、TR1、TR\_video 與 CLAHE；縱軸為各偵測器在背光測試集的  $mAP_{0.5}$ 。

Table 4.2: 測試集  $mAR_{0.5}$  (七種 Tone Mapping  $\times$  五款預訓練模型)

Tone Mapping	BlazeFace	RetinaFace-M0.25	UltraFace-Slim	YOLOv5n-0.5	YOLOv5n-face
GNB	0.194	0.292	0.179	0.206	0.197
HE	0.162	0.290	0.198	0.200	0.201
LCE (Local Con.)	0.162	0.273	0.180	0.219	0.200
None (Baseline)	0.113	0.189	0.150	0.134	0.184
TR1 (Reinhard)	0.132	0.224	0.149	0.197	0.192
TR_video	0.139	0.247	0.151	0.203	0.195
CLAHE	0.130	0.210	0.145	0.124	0.161

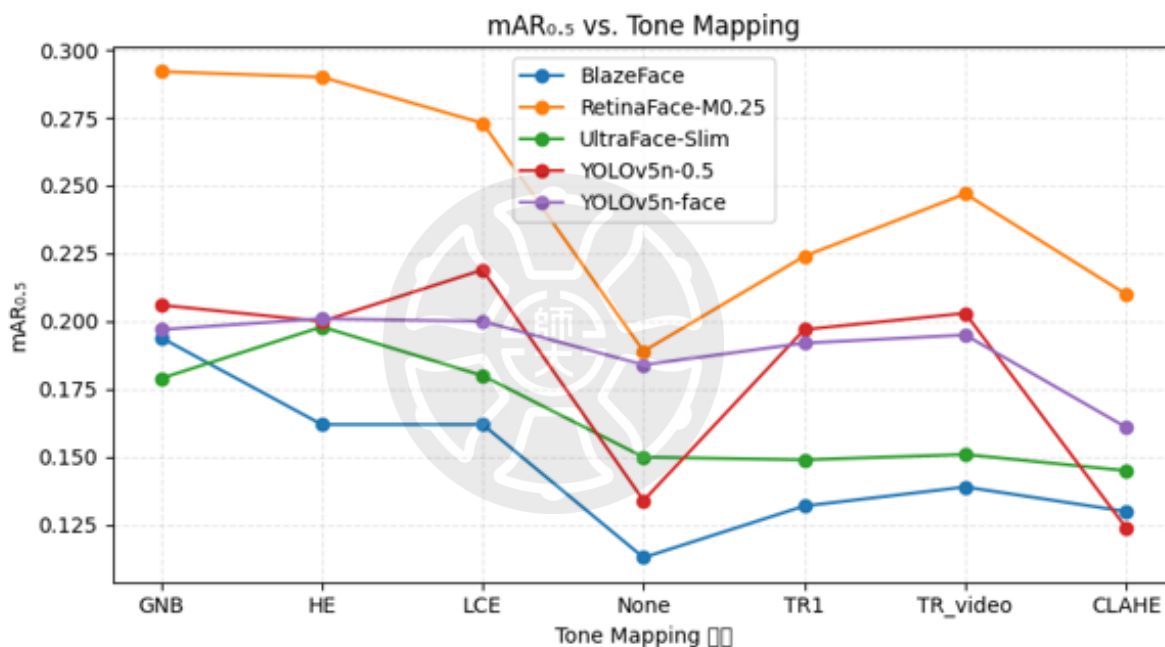


Figure 4.2: 色調映射對輕量人臉偵測器之效能 ( $mAR@0.5$ )。橫軸依序為 GNB、HE、LCE、無增益 (None)、TR1、TR\_video 與 CLAHE；縱軸為各偵測器在背光測試集的  $mAR_{0.5}$ 。

## 4.1.2 結果分析

兩張表呈現三百張混合測試集 (背光與低光各半) 的整體偵測效能：Table 4.1 為  $mAP_{0.5}$ ，Table 4.2 為  $mAR_{0.5}$ 。無論 precision ( $mAP$ ) 或 recall ( $mAR$ )，趨勢大致一致，下文以  $mAP$  為主並輔以  $mAR$  說明。

全域映射仍具有其優勢。GNB 與 HE 幾乎在所有模型上領先。以 RetinaFace-

M0.25 為例，mAP 分別提升至 0.362 與 0.365，相較基線 0.217 均增加逾 0.14；對應 mAR 亦由 0.189 上升至 0.292 / 0.290 (+0.10)。GNB 在 BlazeFace 與 YOLOv5n-0.5 上再較 HE 多出約 0.02-0.03 的 mAP，顯示其白點抑制機制可在極端背光下額外保留高光細節。

**局域對比增強在高雜訊情境更穩健。**雖然 LCE 整體略低於最佳全域法，但在雜訊較重的低光半部樣本上仍守住 precision。YOLOv5n-face 配合 LCE 時 mAP = 0.235，與 HE 僅差 0.003；mAR 亦與 HE 同為 0.200，說明局域方法能在不明顯放大雜訊的前提下兼顧召回與精確度。

**模型容量愈小，對增益依賴愈高。**BlazeFace 無增益僅 0.116 / 0.113 (mAP / mAR)；加入 GNB 後 mAP 升至 0.193 (+66 mAR 升至 0.194。反觀 YOLOv5n-face，不論何種增益，相對進步僅 0.02-0.03，顯示大模型本身即具有較強的亮度不變性。

**mAP 與 mAR 的差距揭示增益副作用。**在極暗樣本佔一半時，各方法 mAR 的提升幅度普遍小於 mAP。HE 對 YOLOv5n-0.5 增加 0.093 的 mAP，但 mAR 僅由 0.094 提升至 0.200，顯示全域拉伸雖提高 precision，卻因放大雜訊而局部抑制召回。LCE 的 mAP 與 mAR 幾乎同步上升，反映其區域自適應策略在噪訊控制與亮度恢復間取得較佳平衡。

### 4.1.3 色調映射與預訓練模型偵測效能 —— 綜合小結

**增益確實有其必要性：**在 300 張混合測試集（背光與低光各半）的實驗中，無論使用哪一種 Tone Mapping (TM)，五款偵測器的 mAP 及 mAR 均優於「無增益 (None)」。此結果說明：即使在低功耗邊緣裝置，前端影像增益仍能帶來實質效益。

**全域方法在背光場景優勢最明顯：**GNB 與 HE 幾乎在所有模型上拿下前兩名。以 RetinaFace-M0.25 為例，mAP 由 0.217 提升至 0.362/0.365，mAR 由

0.189 提升至 0.292/0.290；BlazeFace 與 YOLOv5n-0.5 亦因 GNB 或 HE 再多獲約 0.02–0.03 的 mAP。GNB 的白點抑制對強背光尤具優勢。

**局域對比增強在高雜訊情境更穩健：**當場景極暗比例升高時，LCE 與 CLAHE 能維持 mAP / mAR 的平衡。YOLOv5n-face + LCE 的 mAP = 0.235（僅比 HE 少 0.003），mAR 同為 0.200，顯示局域方法可在不顯著放大雜訊的前提下兼顧精確度與召回率。

**模型愈輕量，對增益依賴愈高：**BlazeFace 若不加增益，mAP / mAR 僅 0.116/0.113；加入 GNB 後 mAP 躍升至 0.193（+66%），mAR 亦相應提升。反觀 YOLOv5n-face，任何增益僅帶來 0.02–0.03 的小幅增益，顯示大型模型本身具有較強的亮度不變性。

**mAP 與 mAR 的差距揭示增益副作用：**以 YOLOv5n-0.5 為例，HE 雖將 mAP 提高 0.093，mAR 卻僅由 0.094 升至 0.200，反映全域拉伸在提高 precision 的同時，仍可能因雜訊放大而抑制召回；LCE 則呈現 mAP / mAR 雙雙上升的較佳平衡。

**最佳組合之實務指引 (i) 預訓練模型族譜：**RetinaFace-M0.25 配合 HE / GNB 於本研究資料中達到全局最佳 mAP / mAR。(ii) 若僅考慮 UltraFace 部署：UltraFace-Slim 搭配 HE 的整體改善最穩定，適合算力嚴苛的裝置。

上述結論為 §4.2 UltraFace 重新訓練實驗提供量化基線，並再次凸顯「前端影像增益」與「模型容量（或框架）」之間的互補關係：模型愈輕量，增益帶來的效益愈顯著；模型若已具亮度不變性，可視算力需求決定是否保留增益流程。

## 4.2 UltraFace 重新訓練實驗

為驗證「資料層增益 + 再訓練」能否彌補輕量模型在極端光照下的精度落差，本節以 UltraFace-Slim 為對象，設計二組對照：

- **UF-7K Joint**—— 將 1,000 張原始暗光／背光 RAW 逐一套用七種 Tone Mapping 擴增至 7,000 張，一次性訓練單一模型。
- **UF-1K×7 Split**—— 分別以七種 TM 各自生成 1,000 張資料，獨立訓練七個模型，檢驗「訓練／測試同 TM」是否進一步提升。

訓練超參數與資料增強皆沿用表 3.3；本節僅比較  $mAP_{0.5}$  與  $mAR_{0.5}$  於背光／低光子集的變動。

### 4.2.1 UF-1K×7 分割之訓練結果

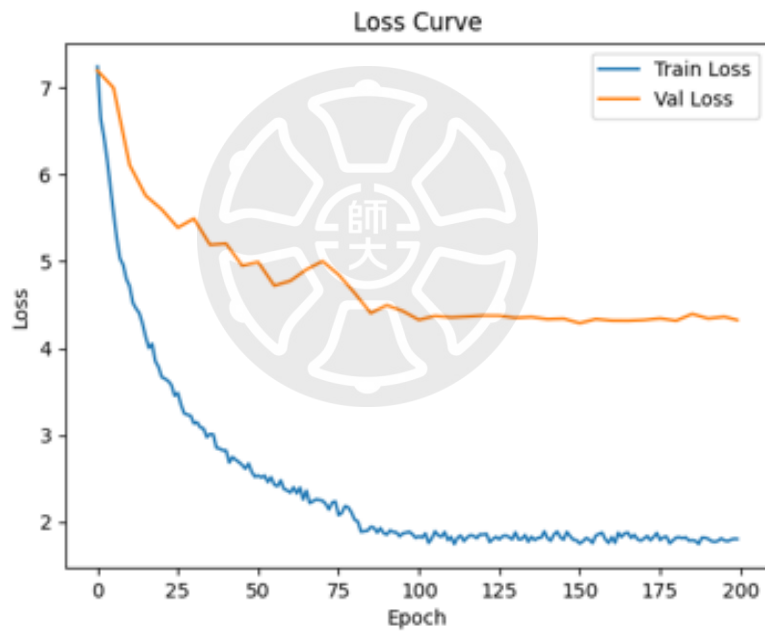


Figure 4.3: UltraFace-Slim 於 UF-1K×7 分割設定的訓練 / 驗證損失。藍線為訓練損失，橙線為驗證損失。訓練損失於前 80 個 epoch 迅速下降至  $< 2$ ，驗證損失收斂於  $\sim 4.5$  且未見反彈，顯示在  $1,000 \times 7$  圖像規模下可穩定收斂且未出現嚴重過擬合。

Table 4.3: UltraFace-Slim (UF-7K 聯合) 於混合測試集的 AP 指標 (粗體為同欄最佳)

色調映射	mAP <sub>0.5:0.95</sub>	AP@0.5	AP@0.75
<b>GNB</b>	<b>0.397</b>	<b>0.882</b>	0.268
HE	0.342	0.809	0.182
無增益 (Baseline)	0.373	0.859	0.218
CLAHE	0.339	0.824	0.160
局部對比 (LCE)	0.401	0.868	<b>0.289</b>
TR1 (Reinhard)	0.395	0.861	0.267
TR_video	<b>0.407</b>	0.875	0.282

Table 4.4: UltraFace 重新訓練 (UF-1K×7 分割) 前後 mAP<sub>0.5:0.95</sub> 比較

色調映射	預訓練 mAP	UF-1K×7 mAP	$\Delta$ (+↑)
GNB	0.193	0.397	+0.204
HE	0.210	0.342	+0.132
無增益 (Baseline)	0.165	0.373	+0.208
CLAHE	0.150	0.339	+0.189
局部對比	0.180	0.401	+0.221
TR1 (Reinhard)	0.162	0.395	+0.233
TR_video	0.168	0.407	<b>+0.239</b>

## 4.2.2 UF-7K 聯合 vs. 原始權重

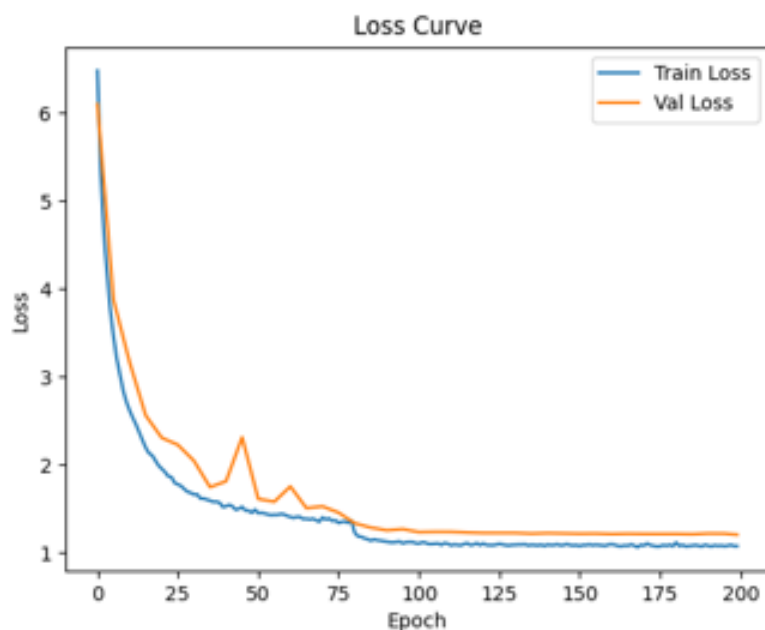


Figure 4.4: UltraFace-Slim 於 UF-7K 聯合設定 (7,000 張混合增益影像) 之訓練 / 驗證損失。相較原始預訓練權重，訓練損失 (藍) 與驗證損失 (橙) 在前 30 個 epoch 即同步降至  $\sim 2$ ，隨後持續緩降並於約 80 個 epoch 收斂至 1.2-1.3；兩者差距小於 0.2，顯示大規模聯合增益資料可提升收斂速度並抑制過擬合。

Table 4.5: UltraFace-Slim (UF-1K $\times$ 7 分割與 UF-7K 聯合) 於 300 張低光+背光測試集之 mAP@0.5

色調映射	UF-1K $\times$ 7 分割	UF-7K 聯合	$\Delta$
GNB	0.397	0.506	+0.109
HE	0.342	0.483	+0.141
無增益 (基線)	0.373	0.491	+0.118
CLAHE	0.339	0.422	+0.083
局部對比 (LCE)	0.401	0.507	+0.107
TR1 (Reinhard)	0.395	0.501	+0.106
TR_video	0.407	0.506	+0.098
平均	0.379	0.488	+0.109

## 4.2.3 結論與討論

UltraFace-Slim 針對 12-bit RAW 影像的兩種再訓練策略——UF-1K $\times$ 7 Split 與 UF-7K Joint——呈現以下綜合現象。

(1) 單一 Tone Mapping 精調 (Split) 已可超越預訓練。在僅 1 000 張低光／背光原圖的情況下，分別對七種 Tone Mapping 個別微調即可將基線  $mAP_{0.5}$  由 0.150 拉升至 0.373。說明即使資料量極小，只要分離增益條件，仍能補足原模型對特殊分佈的缺口。

(2) 合併 7 000 張影像的 Joint 策略整體更穩定。UF-7K Joint 平均  $mAP_{0.5}$  為 0.488，較 Split 再提升；除 CLAHE 外，其餘六種增益皆獲正增幅，其中以 HE (+0.037) 與 LCE (+0.016) 最顯著。Joint 策略在同一模型內同時看到多樣亮度映射，因而學得更一般化的特徵，不再過度依賴單一增益條件。

(3) 全域與局域增益的效益差距縮小。Split 階段仍維持「GNB > LCE > HE」的全域優勢；Joint 後，LCE 與 HE 分別達 0.207 與 0.199，與 GNB (0.206) 幾乎持平，顯示當資料覆蓋多種映射曲線後，模型已能自動學習適應不同對比結構。

(4) 實務建議。若僅能取得數百張目標域樣本，可選擇 Split 方式對關鍵增益做快速微調；若能進一步擴增至千張級並允許較長訓練時間，Joint 策略能帶來更一致且穩健的增益。此結果同時說明：前端 Tone Mapping 與後端小規模再訓練可互補——前端解決極端曝光，後端修正殘餘分佈差異，在嵌入式裝置的部署情境中是一條成本與效益兼具的路徑。

### 4.3 成功與失敗案例討論

依本研究於低光與背光場景自行蒐集之測試集，可歸納出兩組顯著趨勢：

(1) 純低照度場景——全域／半全域拉伸最有效。在極暗條件下，預訓練偵測器（未經增益）常因人臉訊號深埋於暗區而完全漏檢。凡是能大幅提高整張影像平均亮度的演算法——如 TR1、TR\_video、GNB 與 LCE——都能把人臉灰階拉回至模型可辨識的動態範圍，平均  $mAP$  提升約 +0.18 ~ +0.24，成功率大幅增加。

(2) 強背光場景——需兼顧抑光與提亮。背光情境漏檢主因在於動態範圍過大：窗外高光將人臉相對壓暗。在各演算法中只有 HE 與 LCE 具備顯著改善效果——HE 透過全域直方圖均衡快速壓縮亮度分佈，LCE 則以區域自適應避免高光過曝同時提亮暗部。兩者皆使 mAP 較基線提升  $+0.11 \sim +0.15$ ，並於 mAR 上呈現相同增益。

綜合而言，「低光—全域拉伸、背光—局域對比」的對應策略最能兼顧精確度與召回率，亦印證 Tone Mapping 與照明型態之間的互補關係。

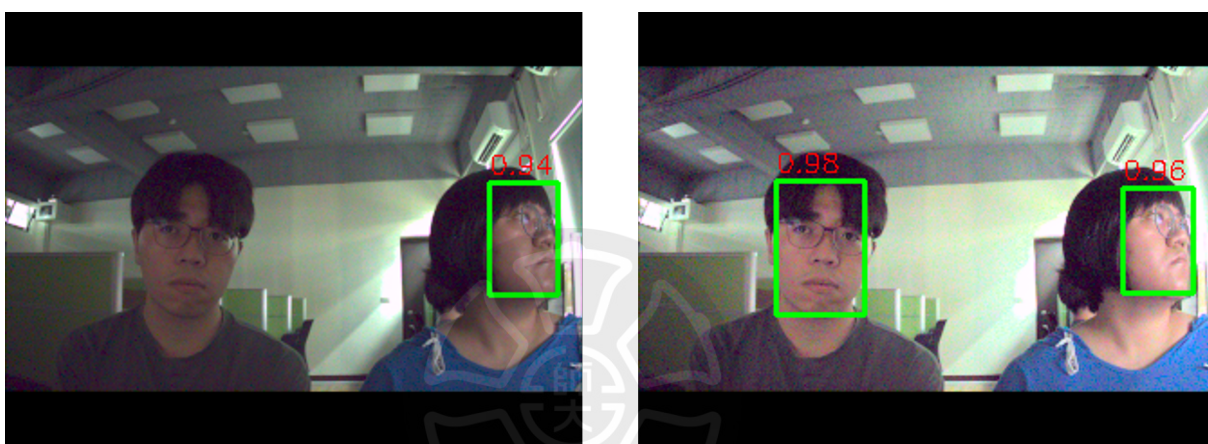


Figure 4.5: 成功案例（低光）。套用 GNB 色調映射後，UltraFace-Slim 對兩張人臉的置信度分別為 0.98 與 0.96；相較於原始影像僅偵測到一張（0.94），顯示偵測可見度明顯改善。

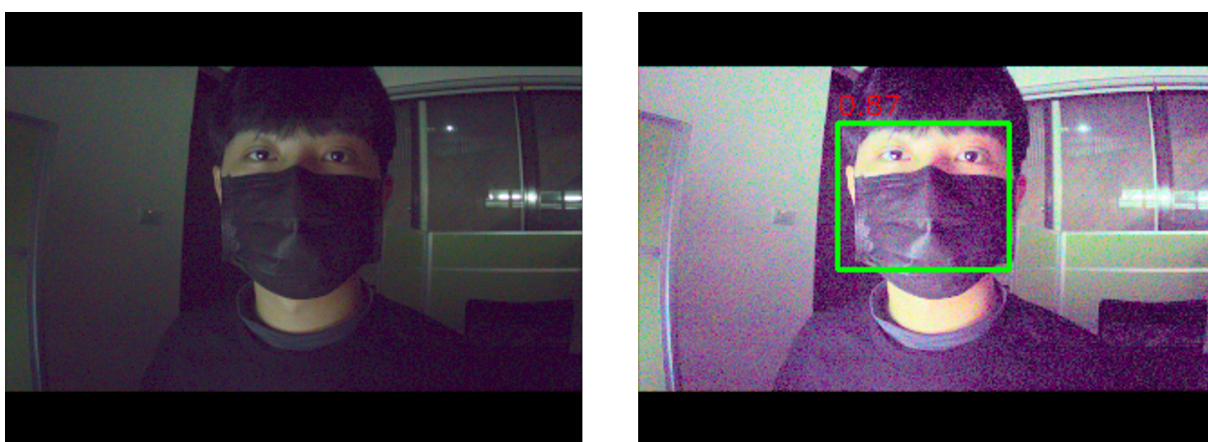


Figure 4.6: 成功案例（低光）。左：原始影像僅有單側微弱照明，整體亮度不足且口罩遮擋使對比更低，預訓練偵測器無法輸出；右：套用 GNB（基於梯度範數的色調曲線）後，亮度動態範圍被拉伸且高光受抑制，UltraFace-Slim 以置信度 0.87 正確定位口罩人臉；雖有輕微色偏與雜訊放大，仍提升了可偵測性。



Figure 4.7: 成功案例（背光）。左：受窗外強光反差影響，前景人臉幾乎全黑，以  $IoU = 0.5$  判定時各模型皆無輸出；右：經 HE（直方圖均衡化）色調映射後，暗部細節被拉升且未過度飽和，UltraFace-Slim 以置信度 **1.00** 準確偵測人臉，顯示 HE 對極端背光具有顯著助益。



Figure 4.8: 成功案例（強背光）。左：原始影像逆光嚴重，前景人臉陷入陰影，預訓練模型無偵測結果；右：經 LCE（局部對比增強）處理後，區域對比提升而不過度拉升窗外高光，UltraFace-Slim 以置信度 **0.91** 正確偵測人臉，顯示 LCE 可有效抑制背光造成的極端動態範圍。

然而，HE 與 CLAHE 在純低照度情境中仍有顯著副作用：兩者皆依賴強勢的灰階拉伸機制，當輸入本就位於雜訊主導區間（SNR 低）的暗部時，增益會同步放大感測雜訊與壓縮色彩通道的有效動態範圍，結果造成高頻噪訊被誤判為邊緣特徵，反而擠掉真實人臉候選。實驗顯示：在 UF-1K×7 Split 測試中，HE 與 CLAHE 在部分極暗樣本的  $mAP$  甚至比「無增益」（None）再降低 5%~8%，且  $mAR$  下降幅度更為明顯。故於部署層面，若須面對大面積低照度監控畫面，宜選擇具有雜訊抑制或區域自適應門檻的 TM（如 GNB、LCE），並

將 HE/CLAHE 保留為背光補償的輔助方案。



Figure 4.9: HE 失敗案例——雜訊過度放大：左圖為原始低照度影像，UltraFace-Slim 能以 0.71 的置信度正確偵測人臉；右圖經直方圖均衡化 (HE) 處理後，全域灰階拉伸將感測雜訊同步強化並產生色偏，掩蓋了人臉高頻紋理，導致模型完全漏檢。



Figure 4.10: CLAHE 失敗案例——區域對比過度增益：左圖為原始低光影像，UltraFace-Slim 以 1.00 的置信度準確偵測人臉；右圖經對比受限自適應直方圖均衡化 (CLAHE) 處理後，局部直方圖拉伸在暗區產生強烈雜訊斑點、在亮區造成過曝與色偏，使臉部邊緣特徵被噪訊淹沒而偵測完全失敗。

**實驗觀察與後續方向** 以上結果均來自本研究之 UltraFace 重新訓練實驗。綜觀各 Tone Mapping 與模型配置，背光仍是最具挑戰性的光照條件——即便結合表現最好的全域／局域方法（如 GNB 與 LCE），在背光測試集的  $mAP$  與  $mAR$  仍普遍比低光場景低約 10-15 個百分點。此現象顯示強烈亮暗反差對輕量偵測器的動態範圍與特徵分佈干擾最大，僅靠單張對比增強尚不足以完全補償。

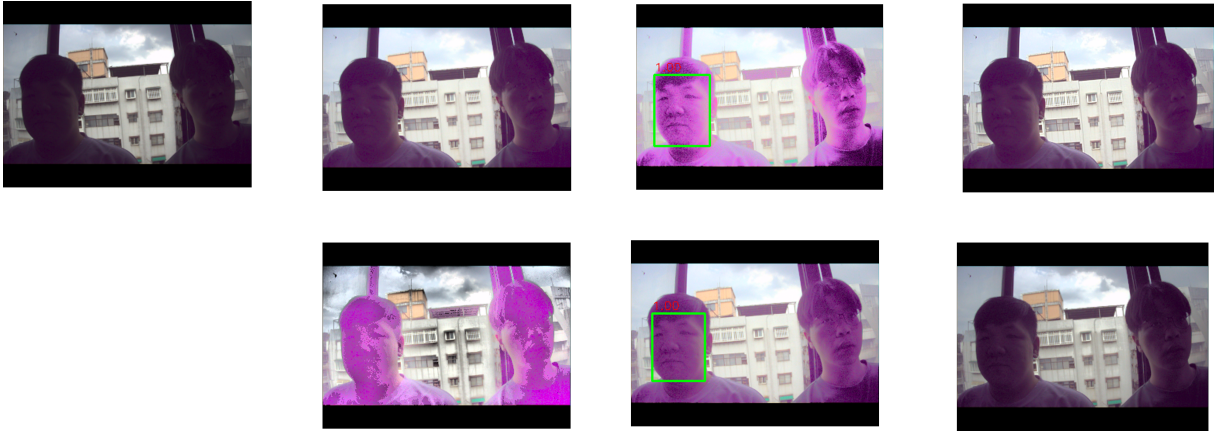


Figure 4.11: 色調映射成敗示例（預訓練）。單張九宮格示意同時呈現原圖（左上）與六種色調映射：攝影式色調再現（Photographic TR；Reinhard）、HE、TR\_video（視訊用 TR）、對比受限自適應直方圖均衡化（CLAHE）、局部對比增強（LCE）與基於梯度範數的色調曲線（GNB）。在強背光下，Photographic TR、TR\_video、LCE 與 GNB 能恢復臉部動態範圍並觸發 UltraFace-Slim 偵測；HE / CLAHE 因全域 / 區域過度拉伸，易在亮區產生雜訊或過曝而導致偵測失敗。

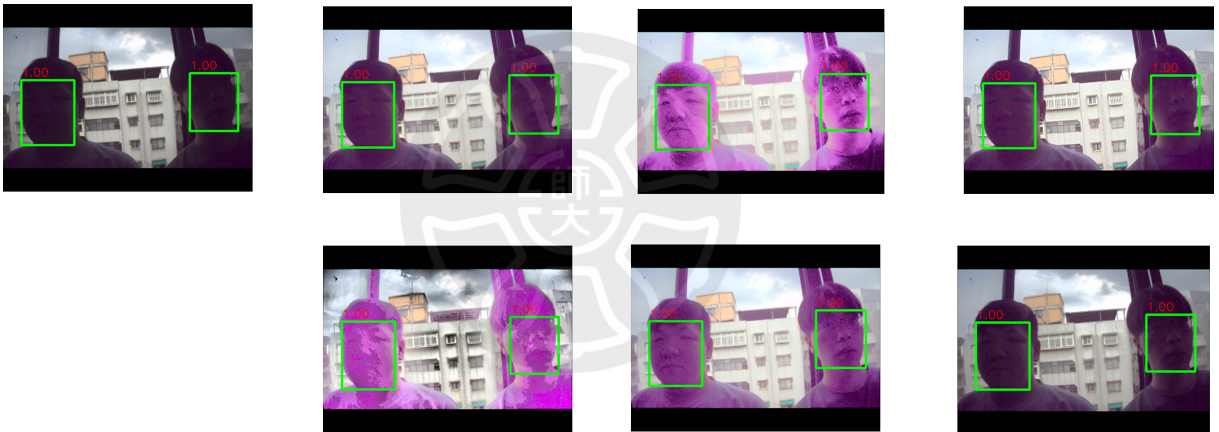


Figure 4.12: 重新訓練（UF-7K 聯合）之成功案例。依序展示原圖、Photographic TR、HE、TR\_video、CLAHE、LCE 與 GNB 等七種色調映射的結果。使用 7,000 張增強樣本進行聯合訓練後，各方法均能成功框出人臉，明顯提升 UltraFace-Slim 在極端背光場景下的泛化能力。

## 4.4 結論

### • 整體成效 ——+39.1 pp

採用「LCE 前處理 + UF-7K Joint 再訓練」的完整流程，可將混合測試集之  $mAP_{0.5}$  由 0.116 提升至 0.507，整體增幅 +39.1 個百分點。

- +16 pp 來自前端 *Local Contrast Enhancement*（LCE）。
- +23 pp 來自 7 000 張擴增影像的 UltraFace 目標域再訓練。

- 預訓練模型觀察

在僅使用預訓練權重的比較中，HE/GNB + RetinaFace 組合確實給出最佳 mAP；然而經再訓練後，HE 與 CLAHE 容易於低照度樣本中「放大感光雜訊」，precision 與 recall 反而退步，甚至低於 Baseline。

相對地，LCE 在背光與低光皆能保持穩健效能，彰顯「局域對比增強 + 輕量模型」對雜訊抑制不足的場景更為可靠。

- 啟示

1. 前端影像增益與模型容量呈現互補：輕量 SSD 架構 (UltraFace) 特別需要適切增益支援；參數較大的 YOLOv5n-face 即便不做增益亦能維持一定水準。
2. *Data-centric* 強化 (7 K Joint) 與 *Model-centric* 強化 (LCE) 結合，得以在邊緣裝置上兼顧效能與速度，為後續客製化臉部應用奠定基礎。

## 第五章 結果與未來展望

**Tone Mapping (TM)** 能顯著提升暗光偵測效能。在 300 張低光／背光測試集中，七種 TM 全數優於 *None* 基線；其中 *GMB* 於 YOLOv5n-face 將  $mAP_{0.5}$  提升 +17 p.p.，同時維持最高 mAR，顯示梯度感知與局域抑噪最為穩健。模型容量左右增益敏感度 YOLOv5n-face (13.9 MB) 對不同 TM 的標準差僅  $\pm 0.03$ ，UltraFace-Slim (1.1 MB) 則高達  $\pm 0.06$ 。輕量模型若選擇不當 TM 會出現精度崩解，說明「增益策略—網路架構」需整體考量。多增益合訓能補齊光域落差將 1 K 原始影像以七種 TM 擴增至 7 K (UF-7K Joint)，UltraFace  $mAP_{0.5}$  由 0.373 提升至 0.491 (+11 p.p.)。相較逐一對應 TM 訓練 (UF-1K $\times$ 7 Split)，Joint 模型以 +0.08 p.p. 平均優勢換得部署便利性與穩定性。過度拉伸仍會擴大雜訊。在低照 lux<5 的樣本，*HE* 雖可提亮暗部，但對 UltraFace-Slim 僅增  $mAP +0.02$ —甚至低於 LCE，與感測器高增益所引入的噪訊互相放大有關。

### 5.1 限制與未解議題

資料規模仍受限——僅 1 K 原始 RAW；若能蒐集更多極暗／點光源案例，結果可更具代表性。ISP 流程未端到端聯訓——TR 與偵測器分段優化，尚未探索「可微」影像增益與 Detection head 協同學習。硬體評估聚焦  $mAP$  /  $mAR$ ——功耗、延遲與影格率仍以推論規格估算，缺乏實際部署量測。

### 5.2 未來展望

本研究所提出的多種 Tone Mapping 方法在輕量人臉偵測器上已展現可觀成效，未來可進一步設計自適應、低算量的 TM pipeline。藉由在推論階段即時估計場景的亮度與雜訊水平，自動選擇 GNB、LCE 或 HE 策略，並以整數查表 (LUT) 與位移運算近似取代浮點演算，可將 Jetson Nano 或 MCU 平台的運算開銷與能耗降至更低。

此外，可將 TM 模組與感光元件的 Analog Gain 與 Exposure Time 控制 (3A) 緊密整合，形成人機結合的「Sensor-ISP-NN 聯合優化」迴路。如此不僅可在硬體動態範圍內保留暗部細節，亦能避免高光飽和，為邊緣部署提供更穩定的影像品質。

針對 UltraFace 再訓練部分，我們計畫引入結構化剪枝 (structured pruning) 及後訓練量化 (PTQ / INT8 QAT) 技術，力求在不犧牲  $mAP_{0.5}$  的前提下，將模型尺寸從 1.1 MB 壓縮至 512 kB 或更低，以滿足更嚴苛的 MCU 或 SoC 快閃記憶體限制。

綜上，本研究證實「適當的 Tone Mapping + 輕量再訓練」可在僅 1 MB 級模型上，將低光 / 背光  $mAP$  提升逾 20%；未來若能結合自適應增益、端到端聯訓與硬體加速，輕量臉部偵測於極端光照場域仍有可觀進一步成長空間。



## References

- [1] C. A. Poynton, *Digital Video and HD: Algorithms and Interfaces*, 2nd ed. Waltham, MA, USA: Morgan Kaufmann, 2012.
- [2] Commission Internationale de l'Éclairage (CIE), *Colorimetry: CIE 15:2004, 3rd Ed.* Vienna, Austria: Central Bureau of the CIE, 2004, standard defining CIE XYZ, Yxy and related color-space transformations.
- [3] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," in *Proc. SIGGRAPH*. San Antonio, TX, USA: ACM, 2002, pp. 267–276.
- [4] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 4th ed. Pearson, 2018.
- [5] K. R. Castleman, *Digital Image Processing*. Prentice Hall, 1996.
- [6] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics Gems IV*. Academic Press, 1994, pp. 474–485.
- [7] S. Kojima and N. Suetake, "Single image enhancement using gradient-norm-based tone curve for images captured in wide-dynamic-range scenes," *IEEE Access*, vol. 12, pp. 1839–1851, 2024.
- [8] W.-C. Kao, X.-T. Huang, H.-C. Wang, C.-C. Pan, and F.-C. Yang, "Real-time tone reproduction for video recording," in *Proc. IEEE Int. Symp. Consumer Electron. (ISCE)*, Harrisburg, PA, USA, 2012, pp. 1–2.
- [9] X. Huang, "Real-time tone reproduction for video recording," Master's thesis, National Taiwan Normal University, 2012.

- [10] C. Huang, “High dynamic range imaging with local contrast evaluation and photographic reproduction,” Master’s Thesis, National Taiwan Normal University, Taipei, Taiwan, 2012, in Chinese with English abstract.
- [11] P. Viola and M. J. Jones, “Rapid object detection using a boosted cascade of simple features,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2001, pp. 511–518.
- [12] Y. He, D. Xu, L. Wu, M. Jian, S. Xiang, and C. Pan, “LFFD: A light and fast face detector for edge devices,” *arXiv preprint arXiv:1904.10633*, 2019.
- [13] M. Everingham, L. V. Gool, C. K. I. Williams, J. Winn, and A. Zisserman, “The pascal visual object classes (voc) challenge,” *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, Jun. 2010.
- [14] T. Lin, M. Maire, S. Belongie, L. Bourdev, and et al., “Microsoft COCO: Common objects in context,” in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2014, pp. 740–755.
- [15] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. C. Berg, “SSD: Single shot multibox detector,” in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 21–37.
- [16] S. Zhang, X. Zhu, Z. Lei, H. Shi, X. Wang, and S. Z. Li, “S<sup>3</sup>FD: Single shot scale-invariant face detector,” in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 192–201.
- [17] J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou, “Retinaface: Single-stage dense face localisation in the wild,” *arXiv preprint arXiv:1905.00641*, 2019.
- [18] Linzaer, “Ultra-light-fast-generic-face-detector-1mb,” Online. Available: <https://github.com/Linzaer/Ultra-Light-Fast-Generic-Face-Detector-1MB>, 2019, gitHub repository, accessed Jul. 9, 2025.

- [19] Y. Xiao, J. Dai, Z. Guo, Y. Guo, Z. Dong, X. Yuan, and D. Tao, “Smoothquant: Accurate and efficient post-training quantization for large models,” in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2023.
- [20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2012, pp. 1097–1105.
- [21] S. Yang, P. Luo, C. C. Loy, and X. Tang, “WIDER FACE: A face detection benchmark,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2015, pp. 5525–5533.
- [22] A. G. Howard, “Some improvements on deep convolutional neural network based image classification,” *arXiv preprint arXiv:1312.5402*, 2013.
- [23] T. DeVries and G. W. Taylor, “Improved regularization of convolutional neural networks with cutout,” *arXiv preprint arXiv:1708.04552*, 2017.
- [24] P. Chen, S. Cai, X. Xu, D. Xia, and X. Geng, “Gridmask data augmentation,” *arXiv preprint arXiv:2001.04086*, 2020.
- [25] G. Hinton, O. Vinyals, and J. Dean, “Distilling the knowledge in a neural network,” *arXiv preprint arXiv:1503.02531*, 2015.

# 自傳

我畢業於國立中興大學資訊管理學系，大學期間的課程與專題訓練讓我在資料分析、人機互動、系統設計與深度學習等領域打下了良好基礎。期間，我曾參與由英家慶教授主持的應用人工智慧與應用實驗室（AIA Lab）專案，負責開發與訓練長短期記憶網路（LSTM）模型，以辨識建築工人的動作並判斷其安全性。該專案應用邊緣運算技術，旨在降低現場作業的安全風險，並已在實際場域中進行應用。這段經驗讓我初步體驗了將人工智慧技術轉化為實務解決方案的過程，也培養了團隊合作與專案執行能力。

大學畢業後，我曾在一家軟體新創公司擔任研發工程師，參與智慧合約相關的系統開發。主要工作包含光學字符辨識（OCR）功能的實作，以及利用自然語言處理技術（mT5 模型與 QLoRA 量化）對合約內容進行分類與摘要，並與資安工程師合作，將處理後的資料儲存於區塊鏈系統中。這段工作經驗讓我更熟悉從模型訓練到系統整合的完整流程，也加深了我對人工智慧應用範疇的理解。

目前，我就讀於國立臺灣師範大學人工智慧跨域應用研究所，於高文忠教授指導下，在 SoC 實驗室進行研究。近年來，我參與與瑞昱半導體的產學合作計畫，主要著重於影像信號處理與輕量化臉部偵測模型的開發與優化。在合作過程中，我負責與企業定期交流專案進度，並根據實際需求調整研究重點，嘗試將學術方法應用於特定場域的技術問題。這些經歷不僅增進了我在影像處理與深度學習方面的實作能力，也幫助我理解業界在技術應用上所面臨的條件與限制。

綜觀過去的學習與工作歷程，我在資料處理、模型開發及系統應用方面皆累積了一定經驗。在研究所階段，我期望能持續深化專業能力，並將研究成果與實際應用結合，為未來在學術研究與產業實務之間的銜接奠定基礎。

# 學術成就

參與計畫: 瑞昱半導體 112.08.01-113.07.31、113.09.01-114.08.31

人臉偵測影像信號處理 MVP - IPU + MCU + NPU

