

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

針對長尾視覺辨識之自適應目標增強策略

Adaptive Targeted Augmentation Strategy for Long-Tail

Visual Recognition



Che-Wei Fan

指導教授：林政宏 博士

Advisor: Cheng-Hung Lin, Ph.D

中華民國114年7月

July 2025

## 致 謝

首先，我由衷感謝我的指導教授林政宏教授，這兩年來給予我極大的支持與悉心指導。林老師為人親切，總是營造輕鬆、尊重自主的學習氛圍，讓我能沒有過多壓力的情況下穩定地投入研究工作。在論文撰寫的過程中，老師不僅提供了精準的方向與建議，也在我迷惘時耐心引導，使我逐步建立起屬於自己的學術思維。更重要的是，老師常以身作則，讓我學會在面對壓力與困難時如何從容應對。

感謝周柏永學長，在我初次接觸人工智慧領域時，學長手把手教我如何撰寫模型、處理資料，讓我在短時間內能夠熟悉整個訓練流程，打下穩固基礎。也感謝郁詠學姊與彥廷學長在研究與實作上的協助，讓我少走了許多彎路；同時感謝侑霖學長毫無保留地分享網路與後端相關知識，使我在跨領域學習上受益良多；在生活方面，感謝郁鈞學姊總是細心照顧我們，經常帶著大家吃吃喝喝、出遊放鬆，使研究生涯中多了許多溫暖與快樂的回憶。感謝博政學長在我第一次擔任講師時提供許多實用的建議，讓我能更有信心地站上講台。

此外，也要感謝宣懿總是與我分享生活中的趣事，在壓力之餘增添不少歡笑與支持，柏序在我專注於論文時，協助分擔產學合作案的工作，讓我能兼顧研究與實務，得以穩步推進。謝謝仲軒學長多次邀請我們到家中聚會，營造溫馨友善的氛圍，也感謝以秀的陪伴，總是在我們疲憊時為團隊注入活力。特別感謝原住民同學易修，總是願意傾聽我天馬行空的想法，並給予我誠懇的回饋，陪伴我度過許多思考與試錯的過程，感謝政軒在論文階段最忙碌的時期主動協助聯繫口試委員，幫助整個流程順利推進，減輕我不少壓力；也感謝學弟妹俐璇、之耀、品妤、品誼與琮善，在課程、活動與計畫上給予我許多實質協助與精神支持，讓整個實驗室團隊氛圍充滿活力與互助。

最後，感謝即將成為菸酒生的柏勳，總是貼心地提醒我那些容易忽略的小細節，讓我在計畫執行上更加周全，也減少許多不必要的失誤。最重要的是，我由衷感謝我的家人一直以來的理解、支持與包容，是他們始終如一地相信我、鼓勵我，才能讓我無後顧之憂地完成學業、順利走到今天這一步。

# 針對長尾視覺辨識之自適應目標增強策略

學生：范哲璋

指導教授：林政宏

國立臺灣師範大學 AI 跨域應用研究所碩士班

## 摘要

監督學習中的長尾問題是由於現實世界資料集中固有的不平衡性所引起的，其中少數幾個類別或樣本佔據了資料分佈的大部分，而大多數類別（「尾部」）則擁有顯著較少的樣本。這個問題對傳統的監督學習算法構成了挑戰，因為這些算法通常優先優化在頻繁（頭部）類別上的表現，而犧牲了在罕見（尾部）類別上的表現。在近期提出的方法中，資料增強技術如 MixUp 和 CutMix 被廣泛應用於解決長尾問題。MixUp 通過對兩張影像進行插值，而 CutMix 則將一張影像的剪切區域貼到另一張影像上，從而合成更多樣化的訓練樣本。然而，據我們所知，目前尚無研究明確探討應該配對或結合哪些影像來達到最佳效果。為了解決這個挑戰，本研究提出了一種名為特徵感知分數選擇 (Feature-Aware Score-Based Selection, FASS) 的新策略。在應用 MixUp 或 CutMix 之前，FASS 根據影像的特徵表現動態選擇並配對影像。與傳統增強方法主要著重於增強少數類別樣本不同，FASS 動態識別與特徵相關的目標類別，以提升模型區分相似特徵的能力。當 FASS 與其他先進方法結合時，在 CIFAR-100 和 ImageNet-LT 等基準資料集上，FASS 展現出卓越的性能，達到了最新的最佳表現。

**關鍵字：**長尾分佈、資料增強

# Adaptive Targeted Augmentation Strategy for Long-Tail Visual Recognition

Student: Che-Wei Fan

Advisors: Dr. Cheng-Hung Lin

Graduate Institute of AI Interdisciplinary Applied Technology  
National Taiwan Normal University

## ABSTRACT

The long-tail problem in supervised learning arises due to the inherent imbalance in real-world datasets, where a few classes or instances dominate the data distribution while a majority of classes (tail) have significantly fewer examples. This issue poses challenges for traditional supervised learning algorithms, which often prioritize optimizing performance on the frequent (head) classes at the expense of rare (tail) classes. Among the recent proposed approaches, data augmentation methods like MixUp, which interpolates two images, and CutMix, which pastes a cutout region from one image onto another, are widely used to address the long-tail problem by synthesizing more diverse training examples. However, to our knowledge, there has been no research explicitly investigating which images should be paired or combined for augmentation to achieve the best results. To address this challenge, we propose Feature-Aware Score-Based Selection (FASS), a novel strategy that dynamically selects and pairs images based on their feature representations before applying Mixup or CutMix. Unlike traditional augmentation methods that primarily focus on enriching minority class samples, FASS dynamically identifies feature-relevant target classes to improve the model's ability to distinguish closely related features. When integrated with other advanced methods, FASS demonstrates superior performance on benchmarks such as CIFAR-100 and ImageNet-LT, achieving state-of-the-art results.

**Keywords:** long-tail distribution, data augmentation

# 目 錄

致 謝 .....	ii
摘 要 .....	iii
ABSTRACT .....	iv
目 錄 .....	v
表 目 錄 .....	vii
圖 目 錄 .....	viii
第一章 緒論 .....	1
1.1 研究背景與動機 .....	1
1.2 研究目的 .....	4
1.3 研究方法概述 .....	5
1.4 研究貢獻 .....	6
1.5 論文架構 .....	8
第二章 文獻探討 .....	9
2.1 權重調整技術 (Re-weighting Techniques) .....	9
2.2 重採樣技術 (Re-sampling Techniques) .....	10
2.3 資料增強策略 (Data Augmentation Strategies) .....	11
2.4 其他應對策略 (Other Methods) .....	12
第三章 研究方法 .....	14
3.1 圖片資料預處理 .....	14
3.2 類別分數設計與目標類別選擇機制 (Feature-Aware Selection) .....	15

3.3 分數導向的資料增強混合策略 (Score-Based MixUp / CutMix) .....	16
3.4 特徵分離損失設計 (Feature Separation Loss) .....	18
3.5 方法總結與演算法.....	20
第四章 實驗結果.....	23
4.1 資料集與任務說明.....	23
4.2 模型與訓練細節.....	24
4.3 CIFAR10-LT、CIFAR100-LT 資料集實驗結果.....	25
4.4 ImageNet-LT 資料集實驗結果.....	27
4.5 消融實驗分析.....	28
第五章 結論與未來展望.....	32
參 考 文 獻.....	34



## 表 目 錄

表 1 :FASS 演算法.....	20
表 2 :資料集資訊.....	24
表 3 :訓練參數表.....	25
表 4 :CIFAR-100-LT 結果.....	26
表 5 :CIFAR-10-LT 結果.....	27
表 6 :ImageNet-LT 結果.....	28
表 7 :不同資料增強方法之消融實驗.....	29
表 8 :FASS 策略之消融實驗.....	29



## 圖 目 錄

圖 1 特徵空間分佈圖 .....	3
圖 2 架構圖 .....	21
圖 3 不同損失函數比例 $\alpha$ 之 Top-1 Accuracy .....	30
圖 4 策略應用頻率 $\rho$ 之 Top-1 Accuracy .....	31



## 1.1 研究背景與動機

隨著深度學習技術的快速發展，特別是在電腦視覺領域的應用中，如圖像分類、目標檢測、語義分割與行為辨識等任務，已取得許多令人矚目的成果。經典模型如 ResNet[1]、EfficientNet[2]以及近年的 Vision Transformer[3]等，這些模型在標準資料集如 ImageNet、CIFAR-10/100、COCO 等資料集上取得了卓越的準確率與泛化能力，成為現今視覺任務的主流解法。

然而，這些模型的成功往往建立在一個隱性前提上——資料的「完美性」。換言之，它們多半是在樣本均衡、數量充足、類別分布合理的資料集上進行訓練與驗證。然而在實務應用上，這樣的條件難以滿足。舉例而言，在自然生態監測中，透過相機陷阱蒐集的野生動物影像，常呈現嚴重的「長尾分布」(long-tailed distribution)，也就是只有少數類別頻繁出現，而其他稀有物種則僅出現極少次數。

此類資料不平衡的情況不僅存在於自然環境領域，也廣泛發生於醫學診斷（例如罕見病辨識）、安全監控（罕見行為或事故的檢測）、電子商務（冷門商品推薦）等場域。由於深度學習模型在訓練過程中傾向於強化主流類別，造成少數類別（tail class）難以被正確分類，這不僅削弱模型的整體效能，也造成結果偏頗，甚至在某些任務中導致嚴重的結果。

進一步來說，資料不平衡問題（data imbalanced problem）對深度學習模型的泛化能力構成極大的挑戰。現實中的長尾分布資料不僅使模型在訓練過程中偏重於數量多的類別，更重要的是，這種偏重導致模型對稀有類別樣本產生的信心通常較低，造成在實際應用中的錯誤分類機率顯著提高。

為了解決這些問題，近年來研究者提出多種方法，例如重取樣（re-sampling）[4][5]、重加權（re-weighting）[6][7][8]以及資料增強（data augmentation）[9][10][11][12]等。重加權策略透過調整各樣本類別的權重，以增加少數類別在訓練過程中的重要性。此類方法主要分為類別層級（class-level）和實例層級（instance-level）的權重調整，儘管這些方法簡單易用，但在實踐中通常對超參數敏感，且易於產生過擬合現象。此外，重取樣策略通過調整類別的樣本數量來平衡類別間的數量差距，例如透過過度採樣少數類別或欠採樣多數類別，然而這種方法可能導致模型學習到不具代表性的資訊或忽略部分重要特徵。

另一方面，以 MixUp[9] 和 CutMix[10] 為代表的資料增強策略，因其簡單有效，已被廣泛應用於處理長尾分布的影像資料。MixUp 通過線性插值的方式，將兩個不同類別的影像與標籤混合，而 CutMix 則將一個影像中的部分區域替換為另一影像的區域，這些方法能有效增加資料多樣性，進而降低少數類別過擬合的風險。然而，現有的大多數研究僅停留在「內容層級」的增強操作，未深入探討資料在「特徵空間」中的相關性與區別性。

為了進一步提升資料增強策略的效果，一些研究開始探索特徵空間中的類別相關性。這些研究試圖在特徵空間中辨識類別之間的關聯性，並基於這種關聯性動態地選擇增強樣本，從而提高資料增強操作的有效性。例如，近期提出的漸進式資料增強方法 CUDA[13]與 DODA[14]，透過動態調整資料增強的強度與頻率，有效降低模型過擬合的風險。此外，也有研究提出利用類別特徵分數動態進行樣本配對的特徵感知資料增強策略。

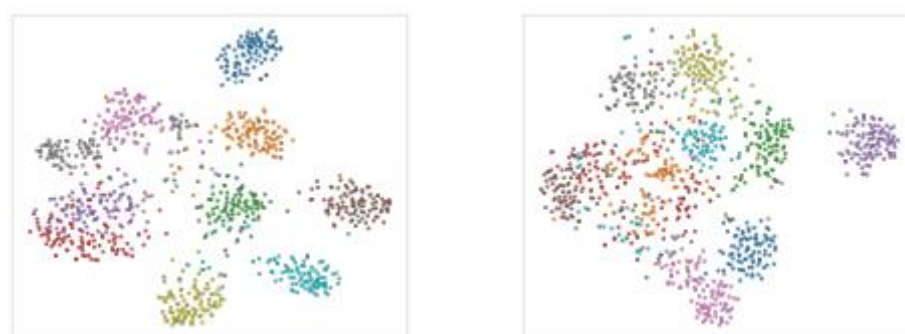
在此基礎上，本研究提出的 FASS（Feature-Aware Score-Based Selection）策略是透過資料特徵動態選取資料的一種策略，其核心思想是透過在訓練過程中動態更新類別的特徵分數，進行針對性資料混合。FASS 策略首先對每一個類別進行特徵分數分析，透過模型訓練過程中各類別的平均信心分數更新此分數；

接著，根據這些分數動態選擇特徵相關度最高的目標類別，以進行資料混合操作。資料混合階段，採用 MixUp 與 CutMix 方法，並動態調整混合比例，使增強後的資料更加貼合特徵空間中的真實分布。

除此之外，FASS 策略也特別在資料混合後的訓練過程中僅使用單一方的真實標籤進行監督，從而確保模型能清晰地學習關鍵特徵而不被混合標籤所干擾。例如，若將來自貓 (cat) 與狗 (dog) 兩類的影像進行 MixUp 操作，傳統方法可能會將該混合樣本標記為 50% 貓與 50% 狗。而在 FASS 中，若選定的監督類別為貓，則即使輸入影像混合了部分狗的區塊，其標籤仍為 100% 貓。這種設計能迫使模型在相近類別干擾的情況下，依然準確識別貓的特徵，進一步強化對目標類別的辨識能力。



(a) 未使用 FASS 之資料分布



(b) 使用 FASS 之資料分布

圖 1 特徵空間分佈圖

綜上所述，本研究旨在彌補現有資料增強策略忽略特徵空間層級關聯性的缺

陷，提出一種自適應的、動態的目標化增強策略 (FASS)，透過動態分析特徵空間中的類別分數，選擇最具特徵相關性的樣本進行資料增強，以顯著提升模型在長尾視覺辨識問題中的表現，進一步推動視覺辨識技術在真實世界應用中的穩健性與有效性。

## 1.2 研究目的

本研究的主要目的，在於解決傳統資料增強策略在處理長尾分布資料時的兩大問題：

1. **忽略特徵空間重疊現象**：傳統的長尾資料處理方法大多著重於提升少數類別的樣本數量，然而這種做法卻未能有效處理另一個重要問題，即特徵空間的重疊性。事實上在長尾分布的情境下，尾部類別不僅因資料稀少而難以學習，更常出現其特徵表示被模型誤判為少數幾個與其特徵相似的頭部類別的現象。這種誤判並非隨機分布於所有頭部，而是集中於與該尾部類別語意或外觀接近的主流類別，導致模型對這些尾部類別產生系統性的辨識錯誤。換言之，長尾分布問題下的分類困難，不僅是樣本不平衡，更包含了特定類別對之間在特徵空間中的混淆。若能識別這些高度相似的類別對，並有針對性地進行訓練與增強，將有助於顯著提升尾部類別的辨識能力。
2. **增強策略選擇隨機性高**：當前主流的資料增強技術，如 MixUp[9]、CutMix[10]，雖然能夠藉由生成混合樣本來提升資料多樣性，緩解類別不平衡問題，然而在進行混合時，增強對象多為隨機挑選，並未考慮其與當前樣本在特徵空間中的相似程度或潛在的混淆關係。這種缺乏選擇性與針對性的增強方式，可能導致混合樣本在特徵層次上不具意義，甚至會干擾模型對關鍵區別性特徵的學習，使得增強效果受限。

舉例來說，當一個尾端類別的樣本與一個在特徵空間中完全無關的主流類別樣本進行 MixUp[9] 或 CutMix[10] 操作時，生成的樣本雖然增加了訓練資料量，但卻無法強化模型對易混淆類別的判別能力。相反，若能策略地選擇在

特徵空間中相近的類別進行混合，將有助於區分這些類別的表示，提升分類邊界的清晰度。因此，本研究提出一種新的特徵導向資料增強策略 FASS

(Feature-Aware Score-Based Selection)，目的是透過動態分析樣本在訓練過程中所學得的類別信心分數，辨識與之特徵相近的類別，再以 MixUp[9] 或 CutMix[10] 進行有意義的混合操作，並僅以一方的標籤作為監督基準，強化模型在特徵層級上的可分性。

### 1.3 研究方法概述

本研究於 CIFAR-100-LT[8] 與 ImageNet-LT[15] 等標準長尾資料集上進行大量實驗，並與現有多種方法（如 LDAM[8]、BCL[16]、RIDE[17]、DODA[14]、CUDA[13] 等）進行組合與比較，以驗證其增強效能與通用性，本研究所提出之 FASS (Feature-Aware Score-Based Selection) 策略，有效解決長尾分布下常見的特徵混淆與尾部類別學習不足問題。其核心機制可分為以下四個主要步驟，彼此相互連結，構成一套以解決語意特徵相似為核心的訓練流程：

1. **動態特徵分數分析**：系統於每一個訓練 epoch 結束後，會將所有訓練樣本於每一次模型 forward 時所產生的信心分數依照其實際所屬類別進行加總與平均，形成一個類別對類別的分數矩陣 (class-to-class softmax score matrix)。該矩陣可視為一種近似語意空間中各類別間的相對相似程度，並被儲存在 Data Selector 中供後續資料選擇階段使用。這種方式不同於傳統以樣本數或距離為依據的相似性評估，能更貼近模型實際的視角，動態反映當前學習狀態下類別間的混淆情形。
2. **目標類別選擇**：對當前 epoch 中所抽取的每一筆樣本，根據其所屬類別與其他類別在上一步所建構之分數矩陣中的信心分數，找出與之語意最相近的一個目標作為混合對象。此選擇機制仰賴類別間的特徵相似性而非隨機挑

選，使得後續資料混合能更具挑戰性與訓練價值，有助於針對混淆區域進行重點學習。

3. **進行資料混合 (MixUp/CutMix)**：本研究設計了一個以類別信心度比值為基礎的動態混合權重( $\lambda$ )計算公式，針對每一組 source-target 樣本對，依據其特徵差異程度調整 MixUp 或 CutMix 的混合比例。此方式避免了過去方法中固定之  $\lambda$  導致的語意模糊問題，使得生成之混合樣本具備更強辨識性。不同於傳統資料增強僅為提升樣本數，本方法更強調「語意邊界強化」，透過混合相似的資料點來迫使模型學習更穩定的分類。
4. **單側監督訓練**：在本研究中，本研究採用單側監督式的訓練策略，即僅使用原始類別樣本的標籤進行監督。與傳統 MixUp 或 CutMix 同時對兩個類別施加監督不同，此策略的核心目的在於引導模型學習在存在相似類別干擾下仍能準確辨識當前類別的能力。換言之，本研究想讓混合樣本中的頭部類別特徵不再被視為目標，而是作為背景或干擾因素，透過單側監督使模型強化對頭部類別本質特徵的聚焦與區辨。此設計有助於提升模型在特徵空間中對頭部類別的魯棒性，尤其是在長尾分布下尾部類別容易被相似頭部類別所掩蓋的情況中，展現更穩定的學習效果。

#### 1.4 研究貢獻

本研究針對長尾分布問題中常見的「特徵混淆」與「尾部類別辨識困難」等挑戰，提出一套具備特徵導向能力之資料增強策略 FASS (Feature-Aware Score-Based Selection)，以補足既有方法中在語意選樣與辨識穩定性上的不足。整體貢獻為以下五點，並具體展現在理論設計與實驗成效層面：

- **提出特徵導向增強策略 FASS**：引入類別特徵信心分數的概念，於每一訓練週期後動態分析模型對各類別的平均信心分數的分布，進而作為樣本混

合對象的選擇依據。不同於傳統以隨機選樣或靜態類別結構為主的混合方式，FASS 更重視特徵相似性與語意脈絡的一致性，所產生之混合樣本更具辨識挑戰性。

- **改善類別間特徵分離能力**：為強化模型對混淆類別的辨識能力，本研究在資料混合訓練過程中採用單側監督策略，僅對其中一方樣本（通常為尾部類別）提供監督訊號。此設計刻意透過不對稱的監督機制，將另一方（頭部類別）的特徵視為背景干擾，促使模型在挑戰情境下學習更具辨識性的尾部特徵。此策略顯著提升了模型對於特徵相近類別（例如 CIFAR-10 中的卡車與汽車）的分離能力，進一步推動模型學習出更清晰且具語意可分性的特徵結構。
- **具高度可擴展性與兼容性**：FASS 策略並非取代現有處理長尾資料分布的方法，而是以「可併用強化模組」的方式設計，能無縫整合至 LDAM[8]、BCL[16]、CMO[12]、CUDA[13]、RIDE[17]等方法中，無論是在 re-weighting、re-sampling 或 expert-based learning 等架構下，皆能提供進一步的資料選擇與特徵強化效果，實驗亦證明此種融合後的系統能產生加乘性效益。
- **在多個資料集上表現優異**：於多個主流資料集上驗證其泛化能力與穩健性。本研究在 CIFAR-100-LT 與 ImageNet-LT 等實際長尾分布資料集上進行全面實驗，並於多種不同比例的 Imbalance Factor（如  $IF = 10$ 、 $50$ 、 $100$ ）條件下均展現卓越表現。此外，在 CIFAR10-LT 資料集上，FASS 亦能穩定提升 baseline 表現，進一步證實其適用於不同難度層級與資料規模之場景。
- **建立特徵視覺化證據**：透過 t-SNE 降維視覺化方法(圖 1)，直觀呈現應用 FASS 前後的特徵空間變化。結果顯示，在導入 FASS 策略後，模型所學得之特徵空間具有更佳類別可分性與語意結構邊界，直觀呈現出語意導向混合所帶來的實質改善。

## 1.5 論文架構

本論文共分為六章，內容架構如下：

- **第一章 緒論**：說明研究背景、動機與目的，並簡述方法概念、研究貢獻與論文結構。
- **第二章 文獻探討**：回顧目前主流的長尾分類方法，包含重加權、重取樣、資料增強與其他近年提出之方法，並進行歸納與比較。
- **第三章 研究方法**：詳述所提出之 FASS 策略，包括資料預處理、類別分數計算與目標類別選擇機制、資料增強策略、特徵分離損失設計、FASS 方法總結與其演算法。
- **第四章 實驗結果**：介紹實驗所使用之資料集、評估指標與訓練設定，並比較本方法與現有技術之表現差異，最後比較消融實驗。
- **第五章 結論與未來展望**：分析研究結果，探討其意涵與可能應用場景，並提出改進方向與後續研究建議。

## 第二章 文獻探討

本章在探討與本研究主題長尾視覺識別（Long-Tailed Visual Recognition）密切相關之研究背景與技術發展。由於在現實世界中，資料往往呈現高度不平衡的分佈，即大部分類別樣本極少（tail classes）而少部分類別樣本充足（head classes），使得傳統的深度學習模型容易偏向於學習資料量較多的類別，忽略了少數類別的辨識準確性。為此，研究者提出多種應對策略，包含重加權（re-weighting）[6][7][8]、重採樣（re-sampling）[4][5]、資料增強[9][10][11][12]（data augmentation）、學習策略[18][19] 與專家系統架構[20][21]等方法。

### 2.1 權重調整技術（Re-weighting Techniques）

在長尾學習的研究中，類別不平衡是一項關鍵挑戰。為了減少這類不平衡所造成的分類錯誤，學者們提出了「權重調整」方法（Re-weighting），透過改變不同樣本對損失函數的貢獻程度，來讓模型在訓練過程中更加重視樣本數較少的少數類別。這些方法可大致區分為兩類：類別層級（class-level）與實例層級（instance-level）。

類別層級的權重調整[8][14][22][23]根據每一類別的樣本數量或其有效樣本數，調整其在訓練中的重要性。實例層級方法[24]則會根據樣本的難易度或罕見程度進行個別調整，以實現更細緻的平衡。

然而，這些方法常面臨幾項挑戰，例如對超參數設定極為敏感、訓練過程中可能產生過擬合問題，以及計算資源消耗高等。為了解決這些問題，一些研究者提出混合式方法[25][26]，將類別層級與實例層級策略整合使用，但仍需妥善調整參數以獲得穩定表現。

在前述類別層級與實例層級方法之外，近期研究更著重於動態調整與可學習權重的設計，以降低人工超參數依賴。[22]提出的 CB-Focal 於訓練早期採用類別平衡權重，後期再漸進式過渡至 Focal Loss[27]，藉此兼顧尾端樣本的充足梯度與模型穩定性。[23] 的 LDS (Label Distribution Smoothing) 則將標籤機率分散至相鄰類別，等效於在 logits 空間施加 soft re-weighting，能有效減輕極端權重導致的過擬合。

## 2.2 重採樣技術 (Re-sampling Techniques)

重採樣技術 (Re-sampling) 也是應對類別不平衡的常見手段，透過過度採樣 (oversampling) [4] 少數類別或欠採樣 (undersampling) [28] 多數類別來調整訓練資料的分佈，實作簡單卻分別易導致過擬合與資訊流失。為了改善這些限制，近年研究者提出多種進階重採樣策略，嘗試兼顧資料平衡與語意保留效果。例如結合 Grad-CAM 的引導機制[29] [30]，能保留圖像的關鍵區域特徵，避免模型學習到無效資訊，進而提升重採樣過程的語意一致性與辨識效果。

另一類方法則聚焦於特徵空間重採樣 (feature-space re-sampling) [31][32]，在不複製原始樣本的前提下，直接在深層特徵表示上進行補強，以增加少數類別在潛在空間的表徵多樣性與判別能力，這類方法雖然能有效提升尾端樣本的學習效果，卻通常需要額外的訓練資源與更高的計算成本，實作上相對較為複雜。綜合上述，無論是權重調整還是重採樣方法，在實務應用上皆可能產生過擬合風險與參數調整的挑戰。

因此，無論是權重調整還是重採樣方法，在實務應用上皆可能產生過擬合風險與參數調整的挑戰。因此，這些方法常與資料增強策略如 MixUp、CutMix 結合使用，以提升模型在長尾分布下的魯棒性與泛化能力，並有效緩解尾端類

別樣本不足所造成的學習瓶頸。

### 2.3 資料增強策略 (Data Augmentation Strategies)

資料增強在處理長尾分佈的視覺識別任務中扮演著極為關鍵的角色。透過對原始圖像進行轉換，生成額外且多樣的訓練樣本，有助於提升模型的泛化能力與對少數類別的辨識能力。

其中，MixUp[9] 與 CutMix[10] 是目前應用最廣泛的技術之一。MixUp 透過線性插值將兩張不同類別的圖像與其標籤混合；CutMix 則將一張圖像的一部分裁剪貼到另一張圖像上，以產生新的複合樣本。這些技術能增加資料多樣性，減少少數類別過擬合的風險，進而改善類別不平衡下的表現。

近年來，更進階的作法引入 CAM[33]、Grad-CAM[29] 引導資料增強的區域，讓增強更集中於具有判別力的特徵區域，進一步強化少數類別的表徵學習 [12][30]。例如 [30] 策略會計算 Grad-CAM 熱圖，將增強操作集中在語意相關的區塊，避免 CutMix 或 MixUp 混合無效區域造成類別特徵模糊。

除了靜態的資料增強，一些研究也發展出具自適應能力的動態資料增強機制。例如 CUDA[13] 與 DODA[14] 分別利用模型預測的信心分數或 loss gradient 等訊號，動態調整不同資料的增強強度與抽樣機率。這些方法的共通特點是採用漸進式策略，初期讓模型聚焦學習穩定特徵，待具備基本辨識能力後，再引入高變異度的混合資料，以防過早引入噪訊，特別對少數類別效果顯著。儘管上述方法多半著重於創造更多樣的輸入樣本以支援尾端學習，但資料增強對特徵空間的影響機制也逐漸受到重視。傳統的 MixUp 與 CutMix 並不考慮混合樣本在潛在特徵空間中的分布與可分性，可能導致某些混合樣本落入

錯誤類別區域，增加模型混淆風險。為此，一些方法進一步整合資料增強與特徵分析，例如透過 t-SNE 或 UMAP 檢視增強樣本的分布特性，或設計額外損失項引導特徵分離，從而緩解特徵重疊的問題。

近期亦有研究強調將資料增強與權重調整或重採樣機制融合設計，形成多策略協同學習架構。例如[34]結合 Grad-CAM 指引的 CutMix 與 score-based sampling，於每一訓練階段針對尾端樣本進行最具效益的增強與挑選。這種整合策略強調語意一致性、梯度導向與動態回饋，提升模型對尾端類別的識別力，同時保有穩定訓練效果。

因此，資料增強策略已不再僅止於擴增樣本數量，而是朝向「特徵空間的結構重塑與區別性學習」發展。特別是在長尾學習場景中，如何同時兼顧語意表徵強化與特徵分布可分性，將是未來資料增強技術設計的重要方向。

## 2.4 其他應對策略 (Other Methods)

除了上述的權重調整、重採樣與資料增強外，近年來學界亦提出多種方法應對長尾分布所造成的挑戰，涵蓋訓練流程、網路架構與損失設計等多層面策略。

首先，兩階段學習策略 (two-stage approaches) 被廣泛應用於長尾任務中，其核心理念為將「特徵抽取」與「分類器訓練」分離進行。[18]與[19] 即屬此類方法，前者在特徵學習階段使用均勻權重避免過擬合，後者則於特徵穩定後針對分類層進行強化學習。此類方法能降低 representation collapse 的問題，使尾端類別擁有較強的特徵可分性，並有效減少主幹網路過早收斂於多數類別的風險。

其次，專家系統架構（expert-based architectures）提供了具模組化的解法，如 RIDE[17] 設計多個專家子網路，每個子模型專精於不同樣本子分布，並利用路由器（router）模組學習將輸入樣本動態分派至最適合的子模型，提升模型在頭部與尾部類別上的兼容性，可顯著緩解單一模型難以同時學好所有類別的問題。

元學習方法（Meta-learning）為另一種備受關注的方向，代表性研究如[25]與[22]所提出的策略，會根據模型當前訓練狀況，動態調整樣本的損失權重或抽樣機率，使訓練過程能根據模型的需求自適應調整梯度來源。這種策略強調「學習如何學習（learning to learn）」的概念，尤其適用於少數類別樣本有限、分布難以預測的情境，能有效提高泛化能力與穩定性。

此外，對比學習技術（Contrastive Learning）亦被廣泛應用於長尾場景，作為補強特徵表徵的一種手段。傳統的對比學習方法如[18]提出的平衡對比學習，但在長尾資料中，尾端類別難以取得足夠正樣本配對，導致訓練不穩。為此，一些方法如[16]引入類別中心（class prototypes）或調整 sampling 策略，確保少樣本類別在學習過程中有機會被充分對比與強化。[22]更進一步設計了與重加權結合的對比損失函數，讓模型在學習過程中同時兼顧表徵區別性與類別平衡性，顯著提升在長尾分類與檢索任務中的效能。

綜觀上述方法，可發現除了單一策略外，近期研究更傾向將不同策略進行融合，如結合兩階段學習與對比學習、或將專家系統架構與動態 re-weighting 結合，以形成更具彈性與適應性的訓練機制。這些方法皆致力於從多個面向對長尾分布造成的學習偏差加以調節，提升模型在整體類別空間的辨識穩定性與泛化能力，並為後續實務應用提供更具可行性的解法。

### 第三章 研究方法

本研究所提出之研究方法旨在有效解決長尾視覺辨識問題，特別是在資料不均衡情境下的模型泛化能力。具體方法稱之為 FASS (Feature-Aware Score-Based Selection)，結合了資料重採樣、動態特徵選取以及混合資料增強技術，以有效增進模型對長尾類別的辨識效果。

#### 3.1 圖片資料預處理

本研究所採用之資料集呈現出典型的長尾分布特性，即主流類別樣本數量遠多於尾端類別，使得傳統訓練方法容易偏向頭部類別學習，忽略稀有類別之判別特徵。假設訓練集定義為  $\{x_i, y_i\}_{i=1}^N$ ，其中  $x_i$  是輸入樣本， $y_i \in \{1 \dots, C\}$  為其對應的類別標籤，整體樣本數為  $N = \sum_{c=1}^C n_c$ ，而類別樣本數  $n_c$  呈現遞減關係，即  $n_1 > n_2 > \dots > n_c$ ，IF(imbalanced factor) 為最頭部類別樣本數量除以最尾部類別樣本數量  $\frac{n_1}{n_c}$ ，因此實際資料取樣公式如式 (1)：

$$n_i = n_1 \cdot \left(\frac{1}{IF}\right)^{\frac{i-1}{C-1}} \quad (1)$$

為了在資料分布不均的情況下仍能使尾端類別獲得足夠的訓練關注，本研究採用 [14] 提出的「Effective Number of Samples」重採樣方法。不同於傳統以樣本數反比做為類別取樣權重可能導致過度補償，此方法考慮了類別內部資訊的遞減性，具體定義如式 (2)：

$$E(k) = \frac{1 - \beta^{n_k}}{1 - \beta}, \text{ where } \beta = \frac{N - 1}{N} \quad (2)$$

其中  $\beta = \frac{N-1}{N}$ ， $n_k$  為第  $k$  類別的樣本數。此設計的直觀含義是：若某類別樣本大量重複（如資料相似度高），其所能提供的有效資訊增幅會趨於飽和，因此在計算其權重時應予以折減。舉例而言，若一個多數類別的  $n_k = 1000$ ，而某少數類別的  $n_k = 10$ ，則其對應的  $E(k)$  僅會隨著樣本數微幅提升，反映出其邊際資訊遞減的本質，避免過度依賴類別數量而忽略類別間的代表性差異。

透過這樣的權重分配，訓練過程將會以較高的機率取用尾端類別資料，提升其參與梯度計算的機會，並平衡損失函數中不同類別的貢獻程度。相較於直接複製尾端樣本的過度採樣（ROS）可能導致過擬合，本研究方法透過此設計，讓尾端類別能在後續訓練過程中能被有效應用。

重採樣策略的引入，以  $\frac{1}{E(k)}$  當作從採的權重，亦為後續特徵選擇與資料增強（MixUp、CutMix）建立更穩定的基礎，使少數類別在後續步驟中的目標類別選擇與混合比例調控更加具有代表性與區辨力。

### 3.2 類別分數設計與目標類別選擇機制（Feature-Aware Selection）

在長尾資料情境中，即便透過重採樣改善樣本比例，少樣本類別仍可能因其特徵與多樣本類別相似而難以區分，導致分類器產生混淆。為此，本研究設計了 FASS（Feature-Aware Score-Based Selection）策略，其核心概念在於利用訓練過程中模型對各類別的信心水準，動態選出與當前樣本最相似的類別作為資料增強的目標對象，以引導更具針對性的特徵學習，降低特徵空間中的類別重疊現象，具體定義如式 (3)：

$$s_i = \frac{1}{N_i} \sum_{j=1}^{N_i} p_j, \text{ where } p_j = \frac{\exp(z_{j,k})}{\sum_{k=1}^K \exp(z_{j,k})} \quad (3)$$

其中， $z_j$  是模型對於樣本  $j$  的輸出 logits，透過 *Softmax* 函數轉換成機率分佈。針對每個樣本，FASS 選擇最高分數的類別作為目標類別，這項設計具有兩層意義：1. 反映模型對該類別整體的「分類信心」。2. 間接衡量該類別在特徵空間中是否具備明確分離性。

在進行資料增強 (MixUp / CutMix) 時，本研究需為每個經過 balanced re-sample 的樣本  $x_b$  選擇一個「混合對象樣本」 $x_t$ ，本研究採用根據類別分數挑選最具參考性的目標類別機制，選擇方式如式 (4)：

$$y_t = \arg \max s_{k_i} \quad (4)$$

亦即，對於樣本  $x_b$  所屬的類別  $y_b$ ，本研究從其他所有類別中挑選當前信心最高者  $y_t$  作為混合對象，此選擇邏輯有助於使資料增強發生在「容易混淆」的類別之間，強化模型對類別邊界的辨識力。舉例來說，假設某一訓練批次中包含三個類別：類別 A (頭部類別，樣本數多，模型分類信心高)、類別 B (中間類別，信心中等)、類別 C (尾端類別，樣本稀少，信心偏低) 若目前輸入樣本  $x_b$  ground truth 屬於類別 C，且模型對類別分數計算如  $S_A = 0.55, S_B = 0.1, S_C = 0.35$  依據公式 (2)(3)，選定的目標類別即為信心最高的 A。這代表在增強過程中，C 類樣本會與 A 類樣本進行混合，迫使模型更清晰地學習兩者之間的區別。

### 3.3 分數導向的資料增強混合策略 (Score-Based MixUp / CutMix)

在一般資料增強方法中，如 MixUp[9]和 CutMix[10]，增強比例通常是隨

機決定的，例如從 Beta 分布中隨機抽取參數  $\lambda$ 。然而，在處理長尾分佈問題時，隨機比例可能導致過強或過弱的類別混合，降低增強的針對性與學習效益。為了解決這一問題，本研究引入類別分數 (score) 作為依據，設計一種分數導向的動態混合比例調整機制，根據樣本所屬類別與目標類別的模型信心，自適應調整混合強度，使資料增強過程更具辨識性與穩定性。

$\lambda$  比例調整機制為假設當前樣本  $x_b$  屬於類別  $y_b$ ，其混合對象樣本  $x_t$  屬於目標類別  $y_t$ ，對應分數為  $S_b^b$ 、 $S_b^t$ ，則本研究定義混合比例  $\lambda$  為式 (5)：

$$\lambda = \frac{S_b^b}{S_b^b + S_b^t} \quad (5)$$

這裡  $S_b^b$  表示模型對樣本  $x_b$  分類為自身類別  $y_b$  的信心分數，而  $S_b^t$  表示模型對樣本  $x_b$  分類為目標類別  $y_t$  的信心分數。此公式設計當模型對原始類別  $y_b$  的信心越高（表示該樣本已學得穩固），則保留其內容的比例會相對增加（ $\lambda$  趨近於 1）；相反地，若模型對目標類別  $y_t$  的信心接近或高於原始類別，表示該樣本可能處於模糊邊界區，此時  $\lambda$  會趨近於 0.5，誘導更多的跨類別融合，幫助模型學習區辨性特徵。

MixUp 的增強樣本  $x_m$  公式如式 (6)：

$$x_m = \lambda x_b + (1 - \lambda)x_t \quad (6)$$

CutMix 增強樣本  $x_c$  則公式為式 (7)：

$$x_c = x_b \odot (1 - M) + x_t \odot M \quad (7)$$

其中  $\odot$  為像素級對應相乘， $M$  是二元遮罩，指定被替換的區域。遮罩面積的設計同樣與  $\lambda$  成比例，使圖像的融合程度與模型信心連動。

而為了解決類別誤判特徵多集中於與其相似的主流類別這一問題，本研究進一步將 MixUp 和 CutMix 混合樣本的監督標籤設計進行調整，如式 (8) 所示：

$$y_m = y_c = y_b \quad (8)$$

此設計代表，無論圖像內容為來自原始樣本或與其特徵相近的目標類別，模型皆僅以原始類別  $y_b$  為監督目標。如此一來，模型在訓練過程中將學習在類似頭部類別特徵的干擾下，仍能正確辨識尾端類別，進一步強化尾端類別的特徵表徵能力與區辨魯棒性。

實際訓練中，本研究對每一訓練 batch 中的樣本依序進行上述流程，根據 3.2 節所定義的類別分數選出目標類別  $y_t$ ，再從該類別中隨機挑選一筆樣本  $x_t$ ，與當前樣本進行 MixUp 或 CutMix。此策略不僅避免了隨機挑選導致的語意破壞，也讓每一次增強皆具有根據與學習目的。

### 3.4 特徵分離損失設計 (Feature Separation Loss)

本研究進一步設計了「特徵分離損失 (Feature Separation Loss,  $L_{FS}$ )」，目的在於強化模型對混合樣本中原始類別訊息的保持能力，並提升特徵間的可分性。

MixUp 與 CutMix 雖能提升資料多樣性與魯棒性，但若無進一步約束，容

易導致產生的樣本落入類別邊界模糊區，使模型學習出「模糊中間態」的特徵，進而導致過度混合、分類模糊或信心降低等現象。尤其在尾端類別樣本不足的情況下，這種模糊效果將使少數類別的學習更加困難。

因此，本研究配合 FASS 策略設計一項額外損失項  $L_{FS}$ ，對於每一組經 MixUp 或 CutMix 增強產生的樣本，額外計算其分類損失，使模型在強化原類別學習的同時，也能保持對混合樣本中類別資訊的區辨能力。

給定一筆從原始樣本以重新採樣獲取之資料  $(x_b, y_b)$ ，其經過 MixUp 與 CutMix 所產生的樣本分別為  $x_m$  與  $x_c$ ，本研究定義特徵分離損失為式 (8)：

$$L_{FS} = L(z_m, y_m) + L(z_c, y_c) \quad (8)$$

其中  $z_m$  為 MixUp 樣本通過模型後所輸出的 logits， $z_c$  為 CutMix 樣本的模型輸出， $L(\cdot, y)$  為標準交叉熵損失 (Cross-Entropy Loss)，目標為原始標籤  $y_b$ 。此設計與傳統 MixUp 做法有所不同。傳統方法會將標籤也混合為  $\lambda y_b + (1 - \lambda)y_t$ ，而本研究為避免混合標籤導致混淆，特別採用「單一標籤監督」方式，始終以主樣本的  $y_b$  為監督目標，強化模型對主類別的識別能力。舉例來說假設輸入樣本  $x_\alpha$  屬於尾端類別  $\alpha$ ，經與類別  $\beta$  (多數類別)  $x_\beta$  混合後產生 MixUp 樣本  $x_m = 0.3x_\alpha + 0.7x_\beta$ 、CutMix 樣本  $x_c$  為 C 為主體，部分區域被 A 替代，模型仍會以類別 C ( $y_\alpha$ ) 作為訓練標籤進行監督。雖然這兩筆樣本在視覺上皆含有 A 的成分，但藉由固定標籤  $y_\alpha$  並計算損失，本研究強化模型從混合樣本中「找出對應於 C 的鑑別性特徵」。

最終，本研究的損失函數  $L_{total}$  整合原始資料與增強資料的交叉熵損失，如式 (9)：

$$L_{total} = L(z_i, y_i) + \alpha L_{FS} \quad (9)$$

其中 $\alpha$ 是調控特徵分離損失  $L_{FS}$  的超參數，若  $\alpha$  設定過高，可能會使模型過度依賴混合樣本；設定過低，則難以發揮特徵分離的效果。本研究於實驗中以驗證集調參選定適合的  $\alpha$  值（細節請見第四章實驗結果）。

表 1 :FASS 演算法

---

**Input:** Dataset  $D$ , Model  $f$ , Epoch  $t$ , Score  $S$ , Resampling  $R$ , Augmentation  $A$   
**Output:** Augmented Dataset  $D_{aug}$

- 1: **if**  $t = 1$  **then**
- 2:     Initialize  $S \leftarrow 0$ .
- 3: **end if**
- 4:  $D_{re-sample} \leftarrow R(D), D_{aug} \leftarrow \phi$
- 5: **for each**  $(x_i, y_i) \in D_{re-sample}$  **do**
- 6:     Select target  $y_t \leftarrow \operatorname{argmax} S_i, x_t \text{ from } y_t$
- 7:      $\lambda \leftarrow \frac{S_{y_i}}{S_{y_i} + S_{y_t}}$
- 8:      $\tilde{x} \leftarrow A(x_i, x_t; \lambda)$
- 9:     Add( $\tilde{x}, y_i$ ) to  $D_{aug}$
- 10: **end for**
- 11: **for each class**  $k$  **do**
- 12:      $S[k] \leftarrow \frac{1}{N_k} \sum_{j=1}^{N_k} \frac{\exp(f_k(x_j))}{\sum_i \exp(f_i(x_j))}$      # Update scores
- 13: **end for**
- 14: **return**  $D_{aug}$

---

### 3.5 方法總結與演算法

綜合前述各模組，本研究提出的 FASS (Feature-Aware Score-Based Selection) 策略，藉由結合「類別感知的動態目標選擇」、「分數導向的資料混合

比重調整」、以及「特徵保持導向的損失優化」三大核心設計，有效改善傳統資料增強方法在長尾場景中所面臨的類別混淆與特徵重疊問題。

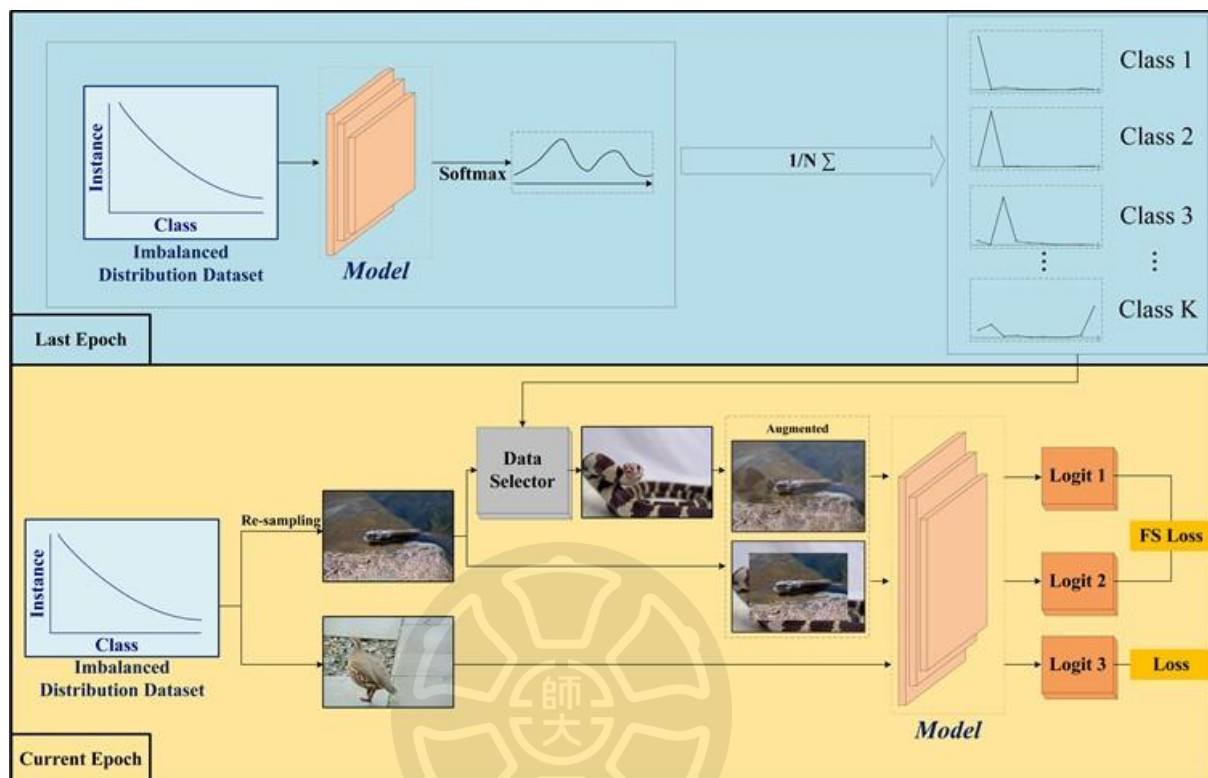


圖 2 架構圖

本方法於訓練階段會先根據上一輪訓練中模型對各類別的平均信心分數，更新類別分數表，作為本輪資料增強參考依據。接著每筆樣本將根據該分數表選擇一個最具區辨性的目標類別進行資料混合，再依據雙方分數差異動態調整混合比例  $\lambda$ ，最終生成具針對性的 MixUp 或 CutMix 增強樣本。整體訓練流程如圖 2 所示。

本研究同時以演算法方式明確定義 FASS 的實作流程，完整列於表 1，FASS 不僅能搭配其他重採樣機制使用，其類別分數的記憶與更新機制(見表 1 之第 11 - 13 行)也可作為一般 score-based 資料增強策略的通用模組進行拓展。FASS 策略的創新之處，在於其並非單純強化資料量或單一方向的權重調整，而

是從類別結構 (score-based)、混合機制 (dynamic  $\lambda$ )、特徵學習 ( $L_{FS}$ ) 三個方面同步優化，並可與其他重採樣或對比學習方法進行模組化整合。透過多輪實驗結果可觀察到，FASS 在保持訓練穩定的同時，能顯著提升尾端類別的辨識率，並於多個 benchmark 長尾資料集上達成競爭性結果。



## 第四章 實驗結果

本章將系統性呈現本研究所提出之 Feature-Aware Score-Based Selection (FASS) 在各種長尾分類資料集上的實驗表現。首先，本研究說明整體實驗設定，包括使用之資料集、模型架構與訓練參數；其次，依序展示 FASS 在 CIFAR100-LT 與 ImageNet-LT 資料集上的分類效果，並與多種現有技術進行比較。接著，本章深入進行消融實驗，評估各組件對模型效能的影響，特別關注 MixUp、CutMix 與特徵分離策略的互補性，以及損失函數比例和策略頻率之超參數分析。最後，本研究綜合所有結果進行討論，針對 FASS 的優勢與應用潛力提出觀察與反思。

### 4.1 資料集與任務說明

為驗證本研究所提出之 Feature-Aware Score-Based Selection (FASS) 策略的通用性與穩定性，本研究選擇三個具代表性的長尾影像分類資料集進行實驗，包括 CIFAR10-LT、CIFAR100-LT 與 ImageNet-LT。這些資料集涵蓋不同的類別數與不平衡程度，能有效模擬實務應用中常見的長尾分布情境。表 2 詳細列出各資料集的訓練樣本數、測試樣本數、類別數與不平衡因子 (Imbalance Factor, IF) 設定。

其中，CIFAR10-LT 與 CIFAR100-LT 分別由 CIFAR-10 與 CIFAR-100 調整而成，透過設定不同的 IF 值模擬從中度到極端不平衡的樣本分布；而 ImageNet-LT 則為真實長尾資料集，含有超過千類，尾部類別樣本甚至少於 20 張，對於模型的辨識能力與泛化能力提出更高挑戰。透過這三組資料集的綜合驗證，可全面觀察 FASS 方法在各種資料量級與不平衡程度下的表現。

表 2: 資料集資訊

Datasets	Training Samples	Testing Samples	Imbalance Factor	Classes
CIFAR10-LT	50,000	10,000	{10,50,100}	10
CIFAR100-LT	50,000	10,000	{10,50,100}	100
ImageNet-LT	115,800	50,000	256	1000

## 4.2 模型與訓練細節

所有實驗皆於 NVIDIA RTX 4090 GPU 平台上執行，採用 PyTorch 框架。為確保與先前研究公平比較，本研究依據 [8] 的 Baseline 參數設定，如表 3。

表 3:訓練參數表

項目	CIFAR10-LT / CIFAR100-LT	ImageNet-LT
模型架構	ResNet-32	ResNet-50
優化器	SGD(momentum = 0.9)	SGD(momentum = 0.9)
Weight Decay	$2 \times 10^{-4}$	$2 \times 10^{-4}$
學習率 初始值	0.1	0.1
學習率 策略	Step decay : 第 160 與 180 epoch 時學習率乘以 0.01	原始 ImageNet-LT benchmark 設定, step decay
批次大小	128	256
訓練週期	200 epochs	100 epochs (依照 benchmark)
實作平台	PyTorch + NVIDIA RTX 4090	PyTorch + NVIDIA RTX 4090

### 4.3 CIFAR10-LT、CIFAR100-LT 資料集實驗結果

本研究於 CIFAR10-LT 與 CIFAR100-LT 兩個經典長尾資料集上進行全面性實驗，評估所提出之 FASS 方法在不同程度資料不平衡下的泛化能力與整合性。CIFAR100-LT 提供 100 類細緻物件分類任務，挑戰性高，常用於驗證長尾方法的實用性與穩定性；而 CIFAR10-LT 則雖類別較少，但仍可有效反映模型在低樣本尾部類別上的表現。

在 CIFAR100-LT 上，實驗涵蓋三種常見 Imbalance Factor (IF=10、50、100)，皆展現出所提方法的強勁表現。即便僅單獨使用 FASS，其在三種失衡程度下的平均 Top-1 accuracy 皆穩定優於 baseline 模型，顯示其內建的語意選樣與特徵分離策略能有效補足尾部類別的語意缺陷。進一步與現有先進策略如 CMO[12]、CUDA[13]、RIDE[17] 等融合時，本方法表現再度顯著提升，於 IF=100 的極端情境中，FASS 搭配 RIDE[17] 所取得的 Top-1 accuracy 高達 51.73%，相較其他 state-of-the-art 技術皆有明顯優勢。詳細結果如表 4 所示，其中灰底區塊為本研究實驗數據，粗體為該欄最佳，紅字則為全表最佳，顯示

本方法在穩定性與準確率表現上皆具備領先地位。

表 4 :CIFAR-100-LT 結果

Method	IF=10	IF=50	IF=100
CE	56.5	43.38	38.1
CMO[12]	60.2	47.0	41.8
CUDA[13]	59.5	46.2	42.0
DODA[14]	59.9	48.0	44.5
CMO+DODA[14]	59.5	49.3	44.9
Ours	60.69	47.70	43.15
CMO+CUDA+Ours	<b>61.92</b>	<b>49.94</b>	<b>45.06</b>
LDAM[8]	57.5	47.2	43.2
LDAM+CUDA[13]	57.9	50.8	47.1
LDAM+DODA[14]	57.9	50.5	48.3
LDAM+Ours	57.85	48.22	43.50
LDAM+CMO+CUDA+Ours	<b>59.77</b>	<b>52.23</b>	<b>48.71</b>
BCL[16]	58.9	49.1	44.2
BCL+CUDA[13]	62.6	52.7	48.8
BCL+DODA[14]	62.7	<b>53.6</b>	<b>51.0</b>
BCL+Ours	<b>63.29</b>	52.28	47.20
RIDE[17]	61.1	52.2	47.5
RIDE+CMO[12]	60.1	53.1	50.0
RIDE+CUDA[13]	59.9	52.1	49.6
RIDE+DODA[14]	60.8	52.6	50.2
RIDE+Ours	62.79	54.37	50.91
RIDE+CMO+CUDA+Ours	<b>63.16</b>	<b>55.78</b>	<b>51.73</b>

而在 CIFAR10-LT 上也可觀察到 FASS 能在 baseline 上帶來明確的準確率

提升。在 IF=100 的設定下，FASS 加入後整體準確率提升約 3.5%，特別是在與如 LDAM[8]、CMO[12] 等重加權方法結合時，顯示其具備良好的相容性與整合性。這也代表 FASS 並非僅適用於複雜且嚴重失衡的任務，在中小型任務中同樣具有實質幫助與泛用性。CIFAR10-LT 的具體比較結果如表 5 所示，紅色字體同樣進一步突顯出全表格中的整體最佳結果。

表 5 :CIFAR-10-LT 結果

Method	IF=10	IF=50	IF=100
CE	86.40	74.80	70.40
Ours	89.65	80.98	76.45
DRW+Ours	<b>90.30</b>	<b>83.62</b>	78.76
LDAM+Ours	90.22	83.33	<b>78.87</b>

#### 4.4 ImageNet-LT 資料集實驗結果

為驗證本研究方法於更貼近現實場景的大規模長尾分類任務中的適用性與泛化能力，本研究將所提出的 FASS 方法應用至 ImageNet-LT 資料集。該資料集具有極端不平衡的樣本分布，且類別數高達千級，為長尾學習中最具挑戰性的基準之一。為深入觀察模型於不同數量層級下的學習能力，本研究依據每類訓練樣本數將所有類別區分為三類：Many-shot（超過 100 筆訓練樣本）、Medium-shot（介於 20 至 100 筆之間）、以及 Few-shot（少於 20 筆樣本）。此分類方式有助於進一步解析模型對頭部、中段及尾部類別的適應性差異。

整體結果如表 6 所示，其中灰底區域標示本論文實作之結果，而紅色字體則凸顯全表格中整體最優者。從整體 Top-1 準確率來看，FASS 單獨使用時已可達到與 LDAM[8]、BCL[16] 等先進技術相當的水準，顯示其具備穩定且具競

爭力的基本性能。而當進一步結合 RIDE[17]、CMO[12] 或 CUDA[13] 等策略時，整體準確率更進一步推升至 54.8%，展現出優異的整合性與放大效益。相比於僅仰賴頻率補償或特徵後期處理的傳統方法，FASS 能在語意層級上引入具代表性的支持樣本，進一步提高辨識準確度與語意穩定性。

表 6 :ImageNet-LT 結果

Method	Head	Medium	Tail	All
CE	64.0	33.8	5.8	41.6
CUDA	67.1	47.1	13.4	47.2
DODA	67.4	47.5	13.9	48.1
CMO	67.0	42.3	20.5	49.1
LDAM	60.4	46.9	30.7	49.8
RIDE(3experts)	64.9	50.4	34.4	53.6
RIDE(3experts)+CMO	65.6	50.6	34.8	54.0
RIDE(3experts)+CUDA	66.0	51.7	34.7	54.7
Ours	63.8	44.2	21.6	48.5
CMO+Ours	63.9	46.2	24.7	49.9
LDAM+Ours	60.8	46.9	31.6	51.0
RIDE(3experts)+Ours	66.7	51.2	31.8	<b>54.8</b>

#### 4.5 消融實驗分析

為深入探討本研究所提出各模組與策略對整體效能的實質貢獻，本研究以 CIFAR100-LT (Imbalance Factor = 100) 為基礎進行一系列消融實驗。首先針對資料增強策略的組合進行測試，結果如表 7 所示，本研究發現單獨使用 MixUp 或 CutMix 均能顯著改善 baseline 模型之準確率，惟兩者結合使用時，效果最為顯著。此結果反映出兩種方法於語意層級的互補性：MixUp 提供平滑的類別

邊界過渡，而 CutMix 則強化局部語意對齊，兩者共用能更有效促進類別間的特徵分離。

表 7:不同資料增強方法之消融實驗

Method	Accuracy
Baseline	37.06
w/ CutMix	40.52(+3.46)
w/ MixUp	41.51(+4.45)
w/ MixUp + CutMix	44.03(+6.97)

為了確認提出之 FASS 策略是否有效，本研究評估了若移除 FASS 策略（Feature-Aware Score-based Selection）後模型表現之變化。結果顯示如表 8，未使用 FASS 時模型準確率明顯下滑，顯示該模組對於提升尾部類別辨識有實質作用，其藉由引入上個 epoch 所記錄之 softmax 分布資訊，引導資料選擇，得以更準確地補強分類模糊區間。

表 8:FASS 策略之消融實驗

Method	Accuracy
w/o FASS	42.27
w/ FASS	44.03(+2.24)

此外，本研究針對損失函數中  $\alpha$  超參數進行分析，以掌握 Feature Separation Loss 與原始分類損失間的最佳權重配置。實驗結果參照圖 3，當  $\alpha = 0.8$  時表現最佳，若比例過低則特徵分離效果不足，過高則會使模型學習偏離原始分類任務，進而導致過擬合。此結果進一步突顯平衡兩種學習目標的重要性。

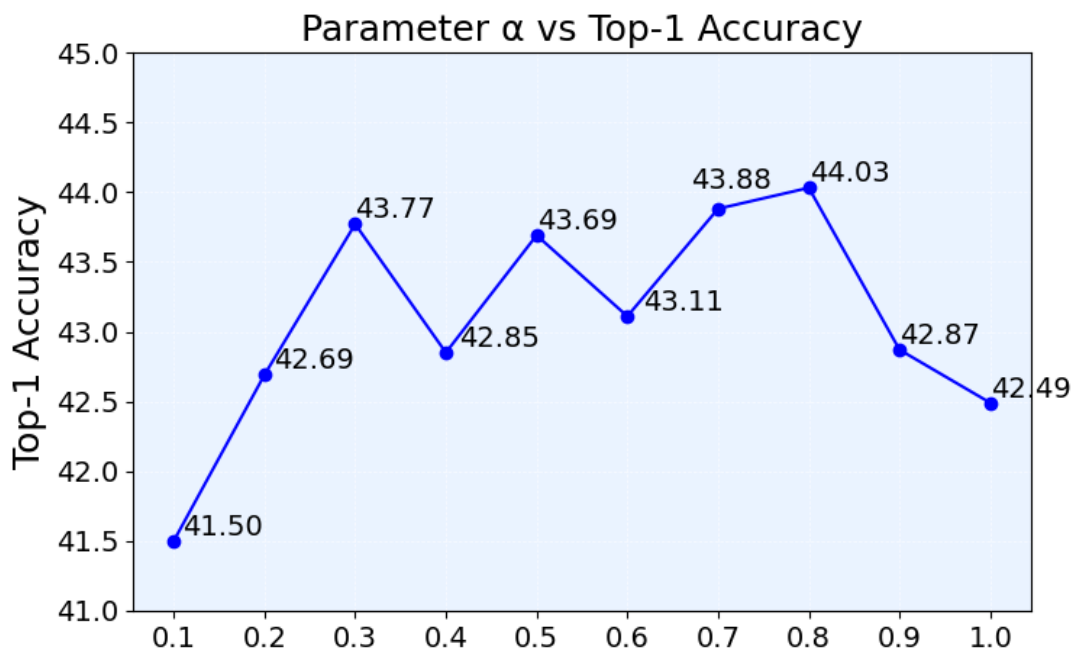


圖 3 不同損失函數比例  $\alpha$  之 Top-1 Accuracy

最後，在 FASS 的應用頻率方面，本研究引入超參數  $\rho$ ，以控制其於每個 mini-batch 中啟用的比例。實驗結果如圖 4，當  $\rho$  設定為 0.3 時整體效能達到高峰，亦即適度啟用 FASS 能有效強化模型之資料選擇能力，但若使用頻率過高，反而可能對模型學習造成干擾，導致收斂不穩。

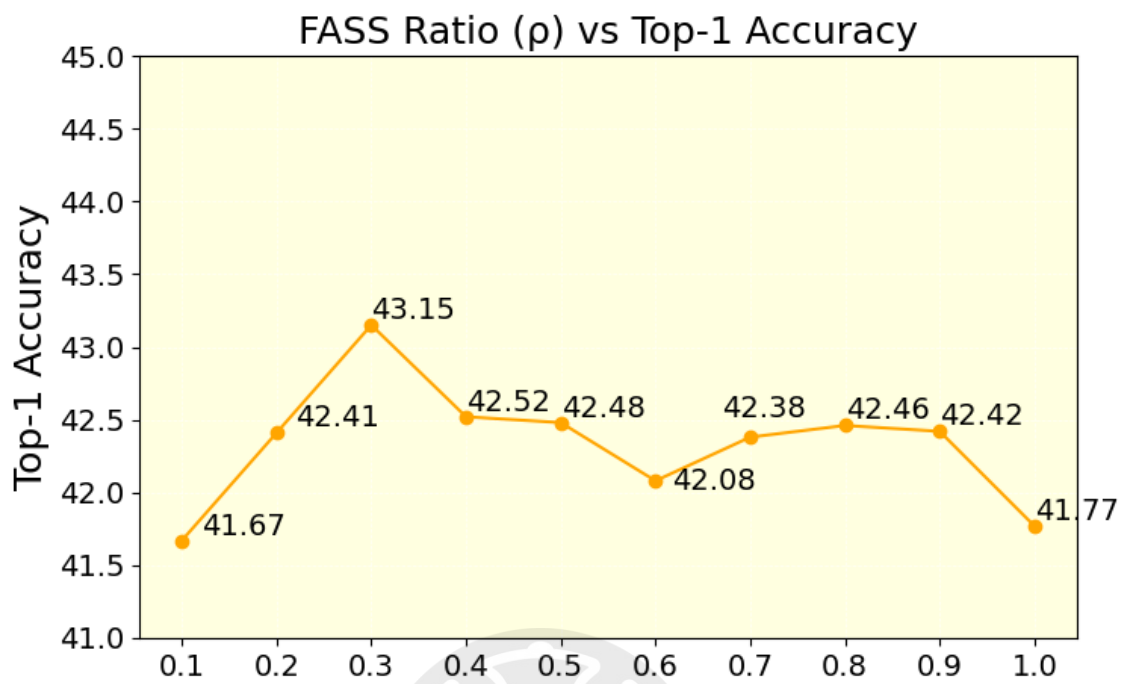


圖 4 策略應用頻率  $\rho$  之 Top-1 Accuracy

## 第五章 結論與未來展望

本研究針對長尾分布 (Long-Tailed Distribution) 中常見的特徵混淆與樣本失衡問題，提出一套具備語意感知能力的訓練機制 Feature-Aware Score-based Selection (FASS)。不同於傳統僅依據樣本數量進行重加權或重採樣的方法，本研究強調「特徵語意」的重要性，透過動態分數記錄、資料選擇與特徵強化的機制，針對尾部樣本進行更具語意辨識力的訓練設計。整體方法兼具穩定性、可擴充性與實務應用潛力，並可與現有方法進行高度整合，展現出強大的兼容性與效能。

實驗方面，本研究於 CIFAR100-LT、CIFAR10-LT 以及 ImageNet-LT 等經典長尾資料集上進行系統性測試，並採用 ResNet-32 與 ResNet-50 作為主幹架構。結果顯示，在各種不平衡程度下，FASS 均能有效提升分類準確率，其中於 CIFAR100-LT (IF=100) 下甚至創下與 RIDE 結合後的最佳成績，展現其於極端失衡環境中依然具備穩健的表現，即便在較為簡單的 CIFAR10-LT 資料集上，FASS 亦能穩定提升 baseline 表現，並展現出高度模組化的整合潛力。

本研究亦透過消融實驗驗證各項子模組的貢獻性，包含 MixUp 與 CutMix 的互補性、 $\alpha$  參數的損失平衡設計、FASS 模組的選擇效益、以及其應用頻率  $\rho$  之影響。實驗顯示上述設計若任一缺失，均會導致效能明顯下滑，反映本方法在整體架構上具備嚴謹的內部耦合。

本研究所提出之 FASS 框架在設計上具有高度的通用性與延展性，不僅可應用於圖像分類等靜態任務，更有潛力拓展至如物件偵測、語意分割與行為辨識等結構性預測任務。這些任務同樣面臨類別不平衡與特徵混淆的挑戰，因此透過特徵感知的樣本選擇策略與單側監督機制，有望進一步強化模型對尾端類

別或難分樣本的辨識能力。此外，未來亦可將 FASS 應用於更具挑戰性的真實世界資料集，例如醫療影像、自然場景、衛星遙測等領域，進一步驗證其在多樣化資料型態與應用環境中的實用性與泛化能力。



## 参 考 文 献

- [1]. K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770–778.
- [2]. M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," *arXiv preprint arXiv:1905.11946*, 2019.
- [3]. A. Dosovitskiy *et al.*, "An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2021.
- [4]. J. V. Hulse, T. M. Khoshgoftaar and A. Napolitano, "Experimental Perspectives on Learning from Imbalanced Data," in *Proceedings of the International Conference on Machine Learning (ICML)*, 2007.
- [5]. M. Buda, A. Maki and M. A. Mazurowski, "A Systematic Study of the Class Imbalance Problem in Convolutional Neural Networks," *Neural Networks*, vol. 106, pp. 249–259, 2018.
- [6]. C. Huang, Y. Li, C. C. Loy and X. Tang, "Learning Deep Representation for Imbalanced Classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 5375–5384.
- [7]. Y.-X. Wang, D. Ramanan and M. Hebert, "Learning to Model the Tail," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, Curran Associates, Inc., 2017.
- [8]. K. Cao, C. Wei, A. Gaidon, N. Arechiga and T. Ma, "Learning Imbalanced Datasets with Label-Distribution-Aware Margin Loss," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 32, Curran Associates, Inc., 2019.
- [9]. H. Zhang, M. Cisse, Y. N. Dauphin and D. Lopez-Paz, "mixup: Beyond Empirical

- Risk Minimization," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2018.
- [10]. S. Yun, D. Han, S. Chun, S. J. Oh, Y. Yoo and J. Choe, "CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea, 2019, pp. 6022–6031.
- [11]. X. Chen, Y. Zhou, D. Wu, W. Zhang, Y. Zhou, B. Li and W. Wang, "Imagine by Reasoning: A Reasoning-Based Implicit Semantic Data Augmentation for Long-Tailed Classification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, pp. 356–364, 2022.
- [12]. S. Park, Y. Hong, B. Heo, S. Yun and J. Y. Choi, "The Majority Can Help the Minority: Context-Rich Minority Oversampling for Long-Tailed Classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA, 2022.
- [13]. B. Chandran, "[Re] CUDA: Curriculum of Data Augmentation for Long-Tailed Recognition," *Transactions on Machine Learning Research (TMLR)*, 2024.
- [14]. B. Wang, P. Wang, W. Xu, X. Wang, Y. Zhang, K. Wang and Y. Wang, "Kill Two Birds with One Stone: Rethinking Data Augmentation for Deep Long-Tailed Learning," in *Proceedings of the 12th International Conference on Learning Representations (ICLR)*, 2024.
- [15]. Z. Liu, Z. Miao, X. Zhan, J. Wang, B. Gong and S. X. Yu, "Large-Scale Long-Tailed Recognition in an Open World," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019, pp. 2532–2541.
- [16]. J. Zhu, Z. Wang, J. Chen, Y.-P. P. Chen and Y.-G. Jiang, "Balanced Contrastive Learning for Long-Tailed Visual Recognition," in *Proceedings of the IEEE/CVF*

- Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA, 2022, pp. 6908 – 6917.
- [17]. X. Wang, L. Lian, Z. Miao, Z. Liu and S. Yu, "Long-Tailed Recognition by Routing Diverse Distribution-Aware Experts," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2021.
- [18]. J. Cui, Z. Zhong, S. Liu, B. Yu and J. Jia, "Parametric Contrastive Learning," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada, Oct. 2021, pp. 715–724.
- [19]. Z. Zhong, J. Cui, S. Liu and J. Jia, "Improving Calibration for Long-Tailed Recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, 2021, pp. 16484–16493.
- [20]. Y. Jin, M. Li, Y. Lu, Y.-M. Cheung and H. Wang, "Long-Tailed Visual Recognition via Self-Heterogeneous Integration With Knowledge Excavation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, BC, Canada, 2023, pp. 23695–23704, doi: 10.1109/CVPR52729.2023.02269.
- [21]. J. Li, Z. Tan, J. Wan, Z. Lei and G. Guo, "Nested Collaborative Learning for Long-Tailed Visual Recognition," in *\*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)\**, New Orleans, LA, USA, 2022, pp. 6939–6948, doi: 10.1109/CVPR52688.2022.00682.
- [22]. J. Ren, C. Yu, S. Sheng, X. Ma, H. Zhao, S. Yi and H. Li, "Balanced Meta-Softmax for Long-Tailed Visual Recognition," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, pp. 4175–4186, 2020.
- [23]. Y. Hong, S. Han, K. Choi, S. Seo, B. Kim and B. Chang, "Disentangling Label Distribution for Long-Tailed Visual Recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*,

- Nashville, TN, USA, 2021, pp. 6622–6632.
- [24]. S. Park, J. Lim, Y. Jeon and J. Y. Choi, "Influence-Balanced Loss for Imbalanced Visual Classification," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada, 2021, pp. 715–724.
- [25]. J. Shu, Q. Xie, L. Yi, Q. Zhao, S. Zhou, Z. Xu and D. Meng, "Meta-Weight-Net: Learning an Explicit Mapping for Sample Weighting," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 32, Curran Associates, Inc., 2019.
- [26]. C. Wang, S. Gao, P. Wang, C. Gao, W. Pei, L. Pan and Z. Xu, "Label-Aware Distribution Calibration for Long-Tailed Classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 5, pp. 6963–6975, 2024.
- [27]. T.-Y. Lin, P. Goyal, R. Girshick, K. He and P. Dollár, "Focal Loss for Dense Object Detection," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2017, pp. 2999–3007.
- [28]. X.-Y. Liu, J. Wu and Z.-H. Zhou, "Exploratory Undersampling for Class-Imbalance Learning," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, no. 2, pp. 539–550, 2009.
- [29]. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2017, pp. 618–626.
- [30]. J.-X. Shi, T. Wei, Y. Xiang and Y.-F. Li, "How Re-Sampling Helps for Long-Tail Learning?" in *Proceedings of the 37th Conference on Neural Information Processing Systems (NeurIPS)*, 2023.
- [31]. S. Ando and C.-Y. Huang, "Deep Over-Sampling Framework for Classifying Imbalanced Data," *arXiv preprint arXiv:1704.07515*, 2017.
- [32]. S. Li, K. Gong, C. H. Liu, Y. Wang, F. Qiao and X. Cheng, "MetaSAug: Meta

Semantic Augmentation for Long-Tailed Visual Recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, 2021, pp. 5208–5217.

[33]. B. Zhou, A. Khosla, A. Lapedriza, A. Oliva and A. Torralba, "Learning Deep Features for Discriminative Localization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 2921–2929, doi: 10.1109/CVPR.2016.319.

[34]. Y. Shi, Y. Wang, D. Lin and A. Yao, "Contextual Sampling for Long-Tailed Recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, BC, Canada, 2023.



## 自 傳

范哲瑋，於 1999 年出生於桃園市，從小在新竹縣關西鎮長大。

- 新竹縣關西鎮關西國民小學
- 新竹縣立關西國民中學
- 新竹縣立六家高級中學
- 國立嘉義大學資訊工程學系
- 國立臺灣師範大學 AI 跨域應用研究所

