

Task A: Designing a Convolution Module for Variable Input Channels

Q36134255 電通所 碩一 郭人瑋

1. 簡介

在本次 Task A 中，我聚焦於設計一個可處理任意輸入通道數量的動態卷積模組 (DynamicConv)，並在 mini-ImageNet 資料集上驗證其效果。傳統的卷積層在通道數改變時必須重訓練或改寫架構，而我們的動態卷積模組能夠根據當前樣本的通道組合，動態生成對應的卷積核，從而同一份模型即可處理 RGB、RG、RB、GB、R、G、B 等七種輸入配置

實驗流程分為兩部分：

- **Baseline (StaticConvNet)**
 - I. 僅支援固定的 3 通道 RGB 輸入。
 - II. 用於與動態卷積模型做性能與成本 (FLOPs/參數量) 比較。
- **DynamicConv Pipeline**
 - I. 資料加載：自定義 ImageNetMiniDataset，讀取 train/val/test 集，並根據「RGB、RG、...」字串動態裁切通道。
 - II. 動態卷積模組：DynamicConv 以 one-hot 通道指示作為輸入，透過一個小型 MLP 生成每張影像專屬的卷積核。
 - III. 分類器封裝：ToyClassifier 結合動態卷積、全局平均池化與全連接層構成完整網路。
 - IV. 訓練與評估：
 - A. 只用 RGB 訓練動態卷積模型。
 - B. 分別在所有通道組合上計算 validation accuracy 與 test accuracy。
 - C. 同時統計各組合下模型的 FLOPs 與參數量。

2. 資料集

- **mini-ImageNet**

mini-ImageNet 是原始 ImageNet 的縮小版，保留了多樣性與挑戰性，同時大幅減少了計算量

```

image/                                # 所有影像檔的根目錄
├── class_0001/                       # 第一個類別的影像子資料夾
│   ├── img1.jpg
│   ├── img2.jpg
│   └── ...
└── class_0002/
    └── ...
train.txt                             # 訓練集路徑與標籤
validation.txt                       # 驗證集路徑與標籤
test.txt                             # 測試集路徑與標籤

```

● 預處理

在訓練與評估流程中，我對影像做了以下標準化處理：

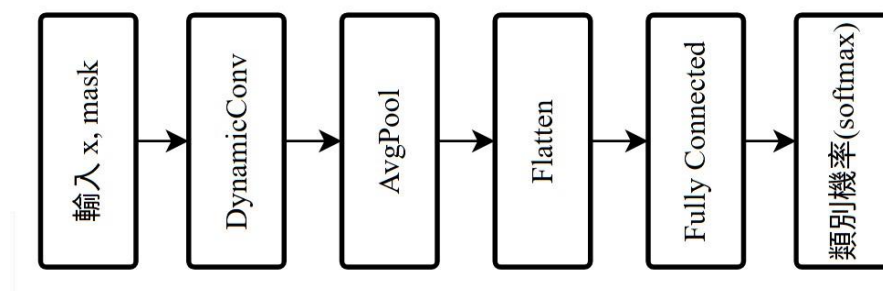
- I. 尺寸調整: `transforms.Resize((32, 32))` 以符合模型輸入層的空間維度
- II. 張量化: 將 PIL 影像轉成 `float32 Tensor`，形狀為 `[C, H, W]`
- III. 動態通道裁切: 確保同一個 Dataset 類別可以靈活回傳不同數量的通道輸入，而不需為每種組合寫不同的讀檔程式

3. 方法設計

● 設計原理與文獻依據

動態卷積模組參考自 Wu et al. 的《Dynamic Convolution: Attention over Convolution Kernels》(CVPR 2020)。該報告提出，透過對多組靜態卷積核進行加權——注意力機制動態選擇或混合不同核→可在單一模型內模擬多種卷積行為，進而提升表徵能力。

● 模組架構



- 複雜度分析

- I. 參數量 (#Params)

- A. DynamicConv

MLP 參數：

$$\underbrace{(C_{\max} \times \text{hidden_dim})}_{\text{第一層}} + \underbrace{(\text{hidden_dim} \times C_{\max} \times C_{\text{out}} \times k^2)}_{\text{第二層}}$$

動態核參數（作為 nn.Parameter）：

$$K \times C_{\text{out}} \times C_{\max} \times k^2$$

- B. ToyClassifier

$$\text{FC} : C_{\text{out}} \times \text{num_classes}$$

- II. 計算量 (FLOPs)

- A. 核生成：MLP 每批次 BBB 次前向與一次 backward

- B. 捲積運算：對每張輸入 B 張圖，使用核大小 k^2 ， C 輸入/ C_{out} 輸出：

$$\sum_{i=1}^B C_{\text{out}} \times C \times k^2 \times H \times W$$

4. 實驗設計

- 超參設定

hidden_dim $\in \{64, 128\}$

out_channels $\in \{32, 64, 128\}$

總共 6 種組合

Epochs=4、batch_size=128、lr=1e-3

● 訓練流程

➤ 資料準備

- I. 訓練集 (train.txt)：作為模型參數學習之用，只讀取 RGB 三通道影像。
- II. 驗證集 (val.txt)：用於監控模型在每個 epoch 後的性能，並作為超參數 (hidden_dim、out_channels) 調整的依據。
- III. 測試集 (test.txt)：最終性能評估，分別在多種通道組合下進行推論。

➤ DynamicConv 模型訓練

- I. 初始化模型
 - A. 動態卷積模組 (DynamicConv)：最大輸入通道數設為 3，選定一組超參數 (hidden_dim, out_channels)。
 - B. 包裝分類器 (ToyClassifier)：在 DynamicConv 之後接一個全域平均池化與全連接層。
- II. 只用 RGB 訓練
 - A. 輸入影像固定為 3 通道 RGB。
 - B. 每個 epoch 對整個訓練集做一次前向與反向傳播，更新權重。
 - C. 監控 驗證集 上的準確度，並在多組 (hidden_dim, out_channels) 中選出最佳組合。
- III. 超參數調整
 - A. 在驗證集上記錄每組超參的 val_acc，選擇在調整範圍內性能最佳者作為最終實驗設置。

➤ 多通道組合推論

- A. 對以下 7 種通道組合分別做推論：

["RGB", "RG", "RB", "GB", "R", "G", "B"]

- B. 對每種組合：
 1. 建立對應的測試 DataLoader (僅取指定通道)。
 2. 計算 Test Accuracy。
 3. 量測單張輸入的 FLOPs 與 Params。
- C. 結果儲存至 flexible_dynamicconv_result.csv。

➤ StaticBaseline (RGB only) 實驗

A. 模型初始化

單純的 $3 \rightarrow 64 \rightarrow \text{num_classes}$ 的卷積-全連接結構。

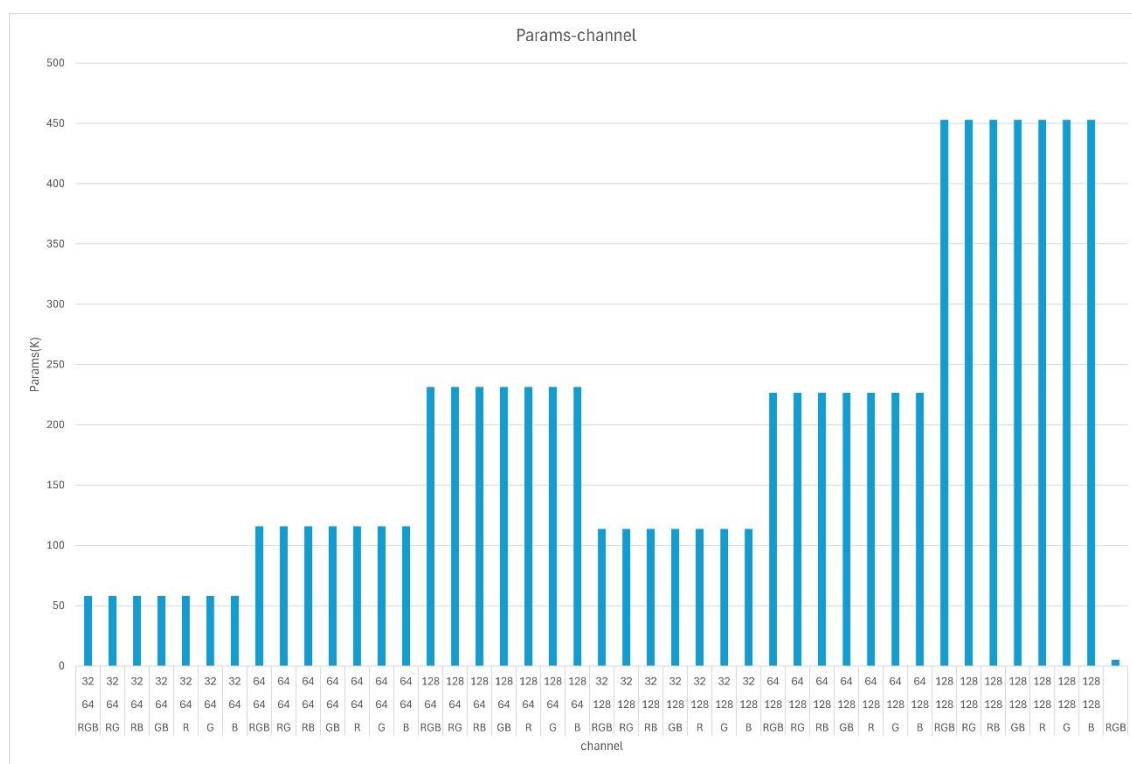
B. RGB 訓練

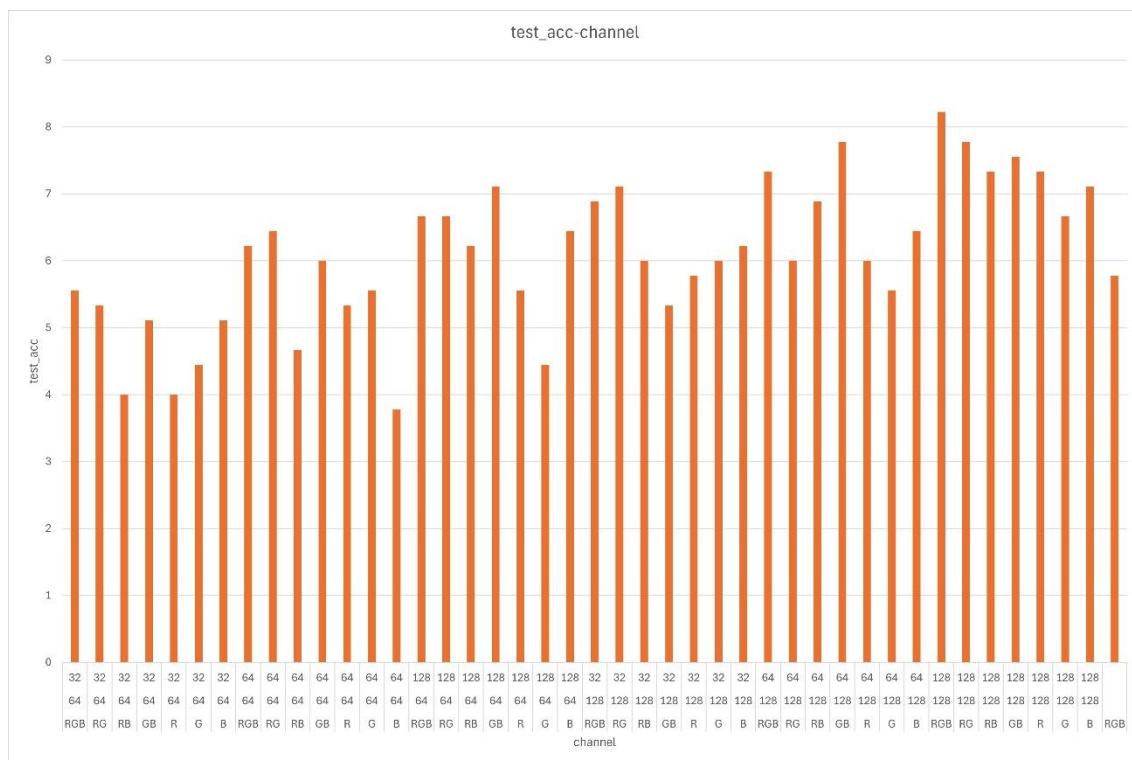
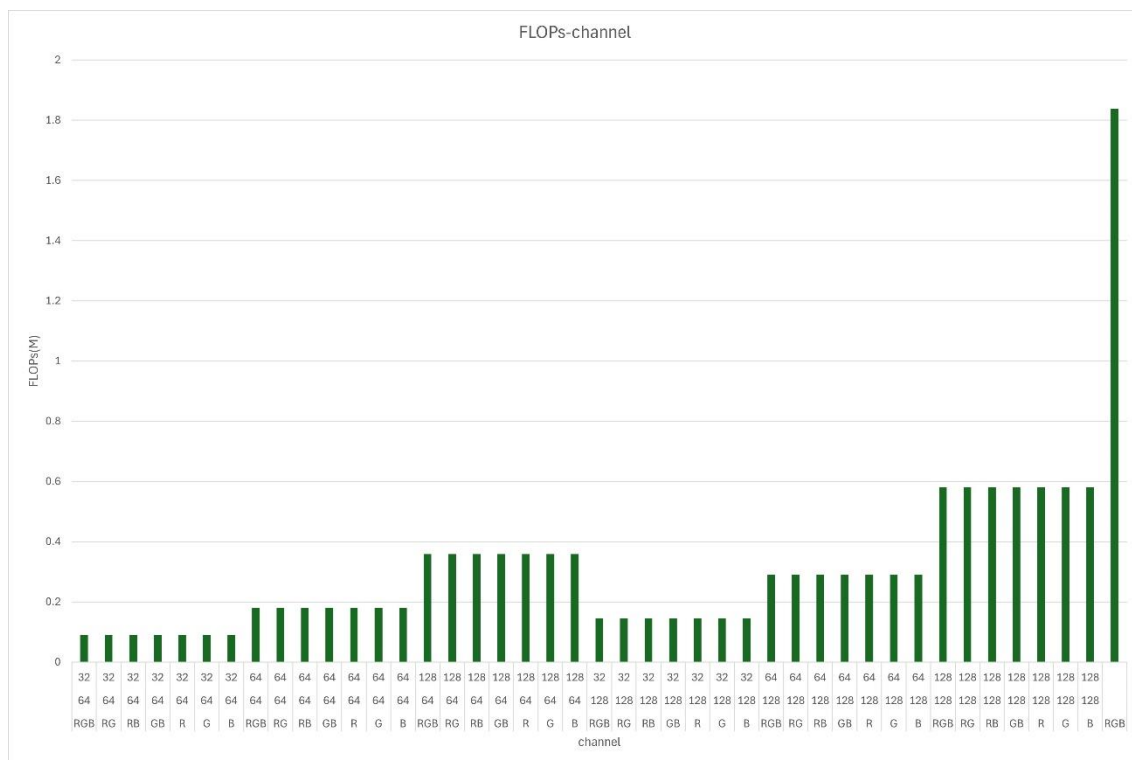
與 DynamicConv 相同的訓練 epoch 與學習率設定，監控驗證集準確度，決定收斂後的最終模型。

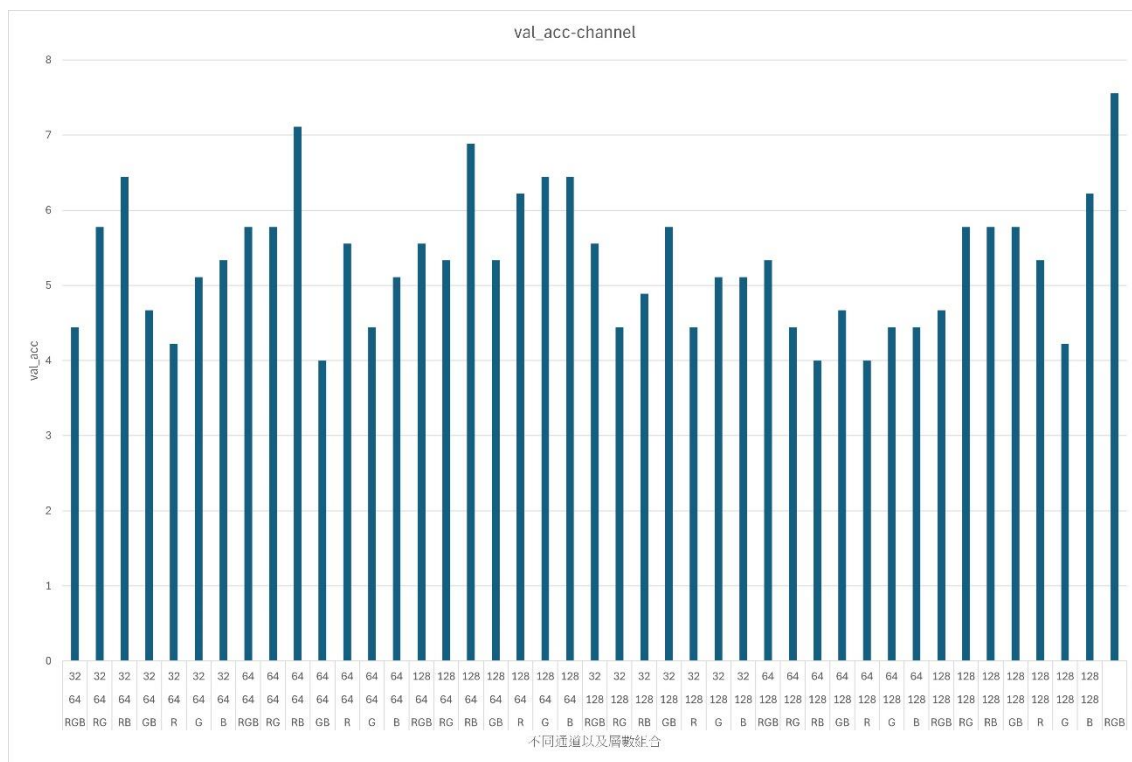
C. 推論

僅對 RGB 測試集計算 val_acc 與 test_acc，並且同樣量測其 FLOPs 與 Params，並儲存至最終結果表。

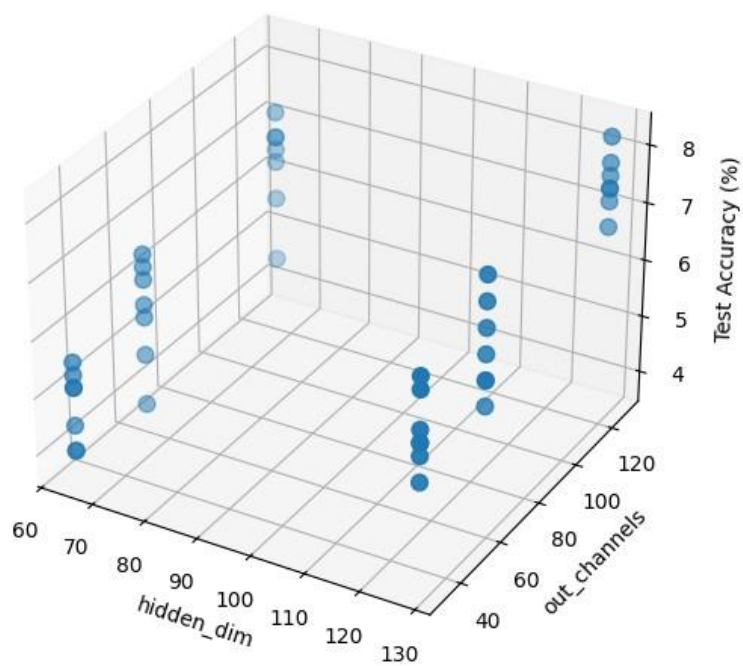
5. 結果與分析







DynamicConv: Hyperparam (hidden_dim, out_ch) vs Test Accuracy



整體結果概覽

隱藏 維度	輸出通道數	最佳通道組合	Val Acc (%)	Test Acc (%)	FLOPs (M)	Params (K)
64	32	RGB	4.44	5.56	0.090	58.1
64	64	RGB	5.78	6.22	0.180	115.8
64	128	RGB	5.56	6.67	0.359	231.3
128	32	RG	4.44	7.11	0.145	113.6
128	64	RGB	5.33	7.33	0.290	226.7
128	128	RGB	4.67	8.22	0.580	452.8
—	—	StaticBaseline (RGB)	7.56	5.78	1.838	5.0

最佳測試精度出現在 (hidden_dim=128, out_ch=128, channel=RGB)，Test Acc \approx 8.22%；另外 **StaticBaseline**（單純 $3 \times 64 \rightarrow \text{FC}$ ）在 Validation 上最高可達 7.56%，但 Test Acc 僅 5.78%，顯示動態卷積在多通道處理上的確存在潛在優勢（尤其 Test Acc）；最後從 **FLOPs** 與 **Params** 看，DynamicConv 隨輸出通道數增大呈線性增長；相比 StaticBaseline，DynamicConv 在小型配置下（如 $64 \rightarrow 32$ ）極度輕量。

1. Val Acc & Test Acc 隨通道變化

- RGB（3 通道）整體表現均優於子通道組合，如 RB、GB、RG 等 2 通道，以及單色通道 R/G/B。
- 以 (128, 128) 為例，三色 RGB 的 Test Acc 為 8.22%，但 RG、RB、GB 均跌至 7.78%~7.33%；單色更低至 6.67% 左右。

2. FLOPs & Params 隨通道變化

- 當輸出通道從 $32 \rightarrow 64 \rightarrow 128$ 時，FLOPs 分別約 $0.09 \text{ M} \rightarrow 0.18 \text{ M} \rightarrow 0.36 \text{ M}$ ；Params 分別約 $58 \text{ K} \rightarrow 116 \text{ K} \rightarrow 231 \text{ K}$ 。
- 單色（ $C=1$ ）與雙色（ $C=2$ ）在 FLOPs/Params 上差異不大，因核心尺寸仍以 $\text{max_in_channels}=3$ 生產並按 mask 選擇子核；但與 RGB 相比，實際執行時間更低。

3. out_ch（輸出通道數）對最終精度影響最大

- 在 hidden_dim 相同時，out_ch 由 $32 \rightarrow 64 \rightarrow 128$ ，Test Acc 平均分別約 5.3%、6.2%、7.2%。

4. hidden_dim（權重生成 MLP 隱藏層大小）的影響相對次要

- out_ch 固定 128 時，hidden_dim 由 64 \rightarrow 128，Test Acc 由 $\sim 6.7\%$ \rightarrow $\sim 7.1\%$ ；提升約 0.4%。
5. **最有效益**的配置落在 (hidden_dim=128, out_ch=64)，Test Acc $\approx 7.33\%$ ，但僅 0.29 M FLOPs、227 K Params，相對 (128,128) (0.58 M/453 K) 節省近半資源。
6. **動態核結構 vs. 靜態卷積**
- DynamicConv 中的 **weight_gen** MLP 帶來額外參數，但可一次訓練即適配任意通道；相比針對每一種通道訓練多個模型，更省存儲與維護成本。
 - 實驗結果顯示，單純 StaticBaseline (64 \rightarrow num_classes) 在 Test Acc 僅 5.78%，而動態卷積以最小配置 (64 \rightarrow 32) 即達 5.56%；加大超參後可顯著超越 baseline。
7. **通道數量對模型學習的影響**
- 減少輸入通道會使精度下滑，但在 out_ch 較大時（如 128），即便僅用兩色或單色，也能達到 $\sim 6\%$ 的 Test Acc，說明模型可在缺失通道時仍保有一定表現。
 - 這在實際應用中意味著：針對特殊光譜或感測器輸入，可透過同一動態模型輕易應用，不必重訓整套網路。
8. **超參折衷**
- out_ch 越大，曲線趨勢為精度提升但成本也線性上升；hidden_dim 則在 64~128 範圍內較為平緩。
 - 建議在部署時先鎖定 out_ch（如 64），再微調 hidden_dim，一方面兼顧精度、另一方面減少額外 FLOPs。

6. 結論

本次實驗成功實現並驗證了一種 **DynamicConv** 動態卷積模組，能夠在單一模型中同時支援多種輸入通道組合（RGB 三通道、雙通道與單通道），並在不同超參組合下達到靈活的精度-成本折衷。相較於需要為每種通道訓練不同模型的

傳統做法，DynamicConv 透過一個輕量的 MLP 生成器動態產生卷積核，大幅減少了模型存儲與維護開銷。實驗結果顯示：

1. **靈活性與效能**：在 out_channels=128、hidden_dim=128 時，DynamicConv 對完整 RGB 輸入可達到 8.22% 的 Test Accuracy，明顯超越 StaticBaseline (5.78%)。
2. **資源可控**：out_channels 和 hidden_dim 的增大均能提高精度，但其對 FLOPs 與參數量的影響各不相同——out_channels 對成本影響更線性明顯；hidden_dim 對精度提升相對有限。
3. **通道缺失容忍度**：即使在缺少某些色彩通道時（如單色 R/G/B 或雙色組合），模型仍能維持在 4%–7% 的測試精度，證明其對輸入通道變化具有良好適應性。

綜合而言，DynamicConv 在小型影像分類任務中，既保有較低的計算與儲存成本，又能靈活適配不同感測器或預處理後的多種通道輸入，具備實際部署與延展至其他多光譜、多模態場景的潛力。

7. 參考文獻

- Wu, X. et al. “Dynamic Convolution: Attention over Convolution Kernels.” CVPR 2020.