

Image Classification with Multiple Models and Features

*Deep Learning HW1

黃亮臻 RE6124035

Institute Of Data Science, NCKU

March 16, 2024

Abstract—本次作業的目標為探索不同的影像特徵擷取方法，觀察其在不同模型下的表現。在這裡，我使用了 FAST+BRIEF、ORB、Color Histogram、HOG 四種特徵擷取方法，與 SVM、KNN、Random Forest、AdaBoost 四種機器學習模型，進行影像的分類任務。程式碼提供於：<https://github.com/edogawa-liang/DL-2024spring>

I. INTRODUCTION

隨著機器學習和電腦視覺領域的迅速發展，影像分類已成為重要的研究方向，例如人臉識別、醫學影像分析、自動駕駛等眾多領域皆有廣泛應用。影像分類的關鍵在於如何從原始影像中擷取有效的特徵，並將其用於分類模型中。好的特徵不僅能提高分類的準確率，還能提高模型的泛化能力。

II. THE PROPOSED METHOD

由於本次作業使用的是常見的機器學習模型，因此這裡會著重於介紹特徵擷取方法。以下四種方法中，除了 Color Histogram，其餘方法皆不支持彩色圖，需先將影像轉成灰階。

FAST + BRIEF. 在這裡，我使用 FAST 快速檢測角點作為特徵點，並使用 BRIEF 描述特徵點。

FAST (Features from Accelerated Segment Test) 是一種角點檢測算法，能夠快速識別出角點，即影像中局部特徵變化顯著的點。原理為選擇影像中的一個像素點，設定一個閾值，檢查這個點其圓形鄰域內的像素點是否明顯比中心點亮或暗，若亮或暗的程度超過閾值，則認為該點是一個角點。FAST 的主要優點是速度快，但對噪聲敏感，且當影像縮放時，檢測到的角點可能會改變。

BRIEF (Binary Robust Independent Elementary Features) 是一種快速的特徵描述符算法。對於 FAST 算法檢測到的每一個角點，BRIEF 算法分析該點周圍的小區域，通過隨機或預定模式採樣像素點對，並按

照預設的像素對比較順序，生成一個代表每對比較結果的二進位字符串作為特徵描述符。例如，若第一個像素亮度大於第二個，則記錄為 1，否則為 0。最終，每個角點都會獲得一串二進位的描述符。

ORB. ORB (Oriented FAST and Rotated BRIEF) 是一種結合 FAST 關鍵點檢測和 BRIEF 特徵描述符的算法。透過 FAST 計算特徵點的方向來增加旋轉不變性，並使用影像金字塔實現一定程度的尺度不變性。使用 BRIEF 計算特徵描述符時，也會根據每個關鍵點的方向來旋轉採樣方式。這些改進提升了特徵匹配的準確度和穩健性。

Color Histogram. 由於影像中的每一個像素都是由 R、G、B 三個通道的數值組成，此方法將照片中三個顏色的分布繪製成直方圖，檢視色彩的分布，以此作為影像的特徵。該方法的優點為影像的特徵不會因為旋轉或縮放而改變，以及計算過程簡單快速。然而，這種計算方式只有考量色彩佔比，沒有考慮到特徵的位置與形狀，影像的亮度變化也會導致繪製的 Color Histogram 不同。

HOG. Histogram of Oriented Gradients (HOG) 透過描述梯度方向的分布，捕捉影像中局部目標的形狀。具體而言，影像首先被分割成多個可重疊的小區域，稱為 Cell。接著，選擇數個相鄰的 Cell 組成一個 Block，在每個 Cell 內部，計算像素點梯度的方向，並將這些方向訊息繪製成直方圖，這些直方圖結合在一起形成了 Block。由於直接計算梯度方向容易受到亮度和陰影變化的影響，因此對每個 Block 進行對比度歸一化處理，以增強模型對光照變化的穩健性。

Pooling Strategy. 使用挑選局部特徵 (ex: 角點檢測) 的方法時，由於不同影像的特徵數量可能有差異，導致每張圖的特徵維度不一致，而影響後續的分類模型，因此，在這裡我使用了類似詞袋模型的策略，為所有影像提供了一個固定維度的特徵。主要想法為首先將所有影像的特徵描述符併在一起，接著使用分群

TABLE I
SETTINGS OF FEATURE EXTRACTOR

Feat. Extractor	Setting	Dim.
FAST + BRIEF + pooling	npoints: 256 (default)	32
	Kmeans: k=32	
ORB + pooling	nfeatures: 500 (default)	32
	Kmeans: k=32	
ColorHist	Bins of R * G * B = 8	512
HOG	WinSize: (128, 128)	405
	BlockSize: (64, 64)	
	BlockStride: (32, 32)	
	CellSize: (64, 64)	
	Nbins: 5	

算法 (ex:Kmeans) 將他們分成 K 組，對於每張影像，預測其所有描述符屬於 K 組中的哪一組，並統計每個組別的出現次數，生成一個直方圖，這個直方圖就是該影像維度 K 的特徵向量。以上介紹的特徵擷取方法中，FAST + BRIEF 與 ORB 需搭配使用此策略。

III. EXPERIMENTS

本次實驗的資料共有訓練集 99600 張與測試集 200 張影像，共 200 個類別。原先的影像維度為 64*64，先將其投放到 256*256 維後，再使用以上介紹的特徵擷取方法，相關的超參數設定如表 I 所示。接著使用 SVM、KNN、RF、AdaBoost 四種模型進行影像分類的任務，此次實驗未特別挑選超參數，模型參數皆使用預設值，並採用 Accuracy 與 F1 score 作為評估指標。

在此次實驗中，由於 SVM 的計算量較大，因此不論搭配什麼特徵擷取方法，使用 SVM 模型前，都會先經過 PCA 降維降至 32 維再放入模型。實驗結果如表 II。整體來看，模型表現皆不是很好，而使用 ColorHist 特徵擷取的模型在 Accuracy 與 F1 score 普遍優於使用 FAST+BRIEF、ORB 和 HOG 的模型，其中 Random Forest 結合 ColorHist 達到最高的 Accuracy (0.105) 和 F1 score(0.0722)。結果說明，對於此次的任務，基於顏色的特徵擷取方法可能比其他局部二進位 (FAST + BRIEF, ORB) 或是基於梯度的描述符 (HOG) 提供更多的訊息，並且 Random Forest 與 PCA+SVM 普遍表現較好。然而，除了參考測試集的表現以外，由訓練集準確率 (Training Acc) 可以發現，大部分的模型在訓練階段也很難有效學習到特徵，而 Random Forest 則有非常嚴重的過度擬合。這不僅說明了傳統機器學習模型較難處理影像分類任務，也暗示了特徵擷取的步驟中，可能未能充分捕捉到對分類有幫助的資訊。

TABLE II
RESULTS

Feat. Extractor	Model	TrAcc	TestAcc	TestF1
FAST + BRIEF + pooling	PCA+SVM	0.136	0.020	0.010
	KNN	0.238	0.005	0.005
	RF	0.999	0.005	0.005
	AdaBoost	0.030	0.005	0.000
ORB + pooling	PCA+SVM	0.103	0.010	0.004
	KNN	0.222	0.015	0.010
	RF	0.997	0.010	0.004
	AdaBoost	0.024	0.005	0.003
ColorHist	PCA+SVM	0.126	0.080	0.052
	KNN	0.268	0.030	0.019
	RF	1.000	0.105	0.072
	AdaBoost	0.045	0.015	0.008
HOG	PCA+SVM	0.187	0.060	0.044
	KNN	0.213	0.015	0.007
	RF	0.963	0.045	0.026
	AdaBoost	0.026	0.020	0.011

TABLE III
THE TIME REQUIRED FOR FEATURE EXTRACTION AND TRAINING

Feat. Extractor	Model	Dim.	Feat.(s)	Tr.(s)
FAST + BRIEF + pooling	PCA+SVM	32 → 32	560	8245
	KNN	32		215
	RF	32		186
	AdaBoost	32		86
ORB + pooling	PCA+SVM	32 → 32	613	10520
	KNN	32		452
	RF	32		245
	AdaBoost	32		87
ColorHist	PCA+SVM	512 → 32	21	8990
	KNN	512		26
	RF	512		396
	AdaBoost	512		187
HOG	PCA+SVM	405 → 32	210	8454
	KNN	405		14
	RF	405		1333
	AdaBoost	405		504

表 III 呈現了本次實驗在特徵擷取和模型訓練階段所耗費的時間。特徵擷取的方法中，Color Histogram 的速度最快，與第二節提到的相符，即檢測色彩分布作為特徵的計算複雜度較低。而由於處理局部特徵的方法 (FAST+BRIEF 與 ORB) 需要再使用 Kmeans 分群，因此花費較多的時間。模型訓練方面，即使進行了 PCA 降維，SVM 所需的時間仍遠多於其他模型，反映了其較高的運算需求。

IV. CONCLUSIONS

在本次作業中，使用了多種方法分別對影像萃取特徵，並使用常見的機器學習模型學習特徵並做影像分類。實驗結果不盡人意，說明了特徵擷取方法與傳統機器學習皆有其局限性。未來可以嘗試使用深度學習方法，例如卷積神經網路 (CNN)，它能夠透過 filter 自動提取特徵，並將特徵擷取與分類任務接在一個框

架中進行端到端的訓練，有機會提高影像分類的表現。

REFERENCES

- [1] OpenCV Python Tutorials, Available online: <https://opencv-python-tutorials.readthedocs.io/zh/latest/>
- [2] OpenCV-with-Python, Available online: <https://github.com/chewbacca89/OpenCV-with-Python/blob/master/Lecture%205.5%20-%20SIFT%2C%20SURF%2C%20FAST%2C%20BRIEF%20%26%20ORB.ipynb>
- [3] Bag-of-Words Model in NLP, Available online: https://medium.com/@derekliao_62575/nlp%E7%9A%84%E5%9F%BA%E6%9C%AC%E5%9F%B7%E8%A1%8C%E6%AD%A5%E9%A9%9F-ii-bag-of-words-%E8%A9%9E%E8%A2%8B%E8%AA%9E%E8%A8%80%E6%A8%A1%E5%9E%8B-3b670a0c7009
- [4] Udacity Computer Vision, Available online: <https://medium.com/chiukevin0321/%E4%BA%BA%E8%87%89%E8%BE%A8%E8%AD%98-face-detection-face-recognition-7ba98aafla02>