RotInv-ConvNet：一種抗旋轉的卷積類神經網路

楊佳誠
資訊工程學系
國立中央大學
桃園，台灣
roman880523@g.ncu.edu.tw

陳弘軒
資訊工程學系
國立中央大學
桃園，台灣
hhchen@acm.org

***Abstract*—卷積類神經網路 (Convolutional Neural Network, ConvNet) 及其延伸的各類模型常被用來處理各種電腦視覺任務，此類神經網路中的卷積層 (convolution layer) 是卷積類神經網路的骨幹。然而，只要將圖像稍加旋轉，儘管人類看起來是一樣的圖像，但旋轉前和旋轉後的圖片經過卷積層的運算後會產生不同的輸出。為了讓旋轉前和旋轉後的圖片在卷積運算後有相同或類似的輸出，通常會將資料進行資料擴增 (data augmentation) 的前處理 (如：將訓練資料中的圖片做隨機轉動當作額外的訓練資料)。本論文試圖提出一種新的方法，透過限制卷積層參數的值來實驗是否能在不進行資料擴增的情況下，讓模型對旋轉後的圖像給予相同或類似的輸出。實驗結果顯示：在不做資料擴增的前提下，新模型對於旋轉過的圖像的判別率優於一般的卷積類神經網路。然而，若允許採用資料擴增，則一般的卷積類神經網路對於旋轉過的圖像的判別率依然優於我們提出的新模型。我們的成果開源於 GitHub: https://github.com/roman-yang/Rot-CNN。**

1. 導論

卷積類神經網路是現代電腦視覺技術的重要基石，廣泛應用於影像分類 (image classification)、物件辨識 (object detection)、語義分割 (semantic segmentation) 等任務中。然而，卷積類神經網路對於圖片的旋轉非常敏感：一張圖片 I_i 只要稍加旋轉成為圖片 I_i^r ，卷積類神經網路即無法認得旋轉後的圖片 I_i^r 與 I_i 本質上是同一張圖片，即：一個卷積類神經網路不具備旋轉不變性 (rotation-invariant)：當任意的圖片 I_i 經過旋轉成為 I_i^r 後，通常 $f_\theta(I_i) \neq f_\theta(I_i^r)$ (其中 f_θ 是一個卷積類神經網路, θ 為此網路的待學習參數)。這顯示：卷積類神經網路或許仍與人類的視覺感知網路有相當的不同。

實務上，為了讓卷積類神經網路對於圖片的旋轉運算比較不敏感，通常需要以資料擴增的方法 (如圖片隨機旋轉、翻轉、局部的像素擾動等)，讓模型同時把原來的圖片及擴增後的圖片當作訓練資料。這種方式雖然看似在工程上讓卷積類神經網路得以認得轉動後的圖片，但本質上只是設法調整卷積類神經網路的參數 θ 使得 I_i^r 與 I_i 在透過卷積類神經網路的運算後能得到相似的輸出 (即：$f_\theta(I_i) \approx f_\theta(I_i^r)$)，並未從卷積類神經網路的設計上解決此網路對旋轉過於敏感的問題。

本論文試圖從卷積類神經網路的網路架構設計著手，讓模型的架構先天上就具備一定程度的旋轉不變性，意即：給定隨機參數 θ，我們希望此卷積類神經網路的架構能讓 $f_\theta(I_i) \approx f_\theta(I_i^r)$。我們稱這個抗旋轉的卷積類神經網路為 Rotation-Invariant Convolutional Neural Network，簡稱為 RotInv-ConvNet。我們希望即使訓練資料中不包含因原圖片旋轉而產生的擴增資料，RotInv-ConvNet 依然能夠因為自身的抗旋轉特性讓 $f_\theta(I_i) \approx f_\theta(I_i^r)$。

我們在 MNIST、FashionMNIST、及 CIFAR-10 上實測 RotInv-ConvNet。實驗結果顯示：在不使用圖片旋轉做資料前處理的前提下，RotInv-ConvNet 使用較少的參數量便能比其他基準模型更準確地預測經過旋轉的圖片的類別。然而，對於未經過旋轉的測試圖片，RotInv-ConvNet 的效果則不如其他基準模型。

2. 相關研究

卷積類神經網路中的卷積層及池化層 (pooling layer) 讓此類型的神經網路具備一定程度的位置不變性 (shift-invariant) 及尺度不變性 (scale-invariant) [1], [2]。以物件偵測 (object detection) 為例，即使物件在圖片中的位置及物件在圖片中所佔的比例有所不同，卷積類神經網路仍能在一定程度上辨識出這些物件。

然而，卷積類神經網路並不具備旋轉不變性，這使得訓練卷積類神經網路時，通常需要將訓練資料中的圖片隨機旋轉當作額外的訓練資料。過去曾有少數的研究探討如何讓卷積類神經網路本身即具備一定程度的旋轉不變性，就我們的觀察，這些方法可分為兩大類：第一種仍然需要旋轉訓練資料，並在網路中強制讓旋轉前後的圖片在網路中經過特定的轉換後具備一定的特性 [3], [4]，例如：能成為相近的矩陣。然而，由於這一類的方法仍仰賴資料擴增，與實務上主流方法的差異僅只於：主流方法希望端至端地讓旋轉前後的圖片具備相同的預測結果、而上述方法希望旋轉前後的圖片在某些設計過的特性上相似。第二種是將歐式空間 (Euclidean space) 的圖片轉換至其他坐標空間 (如：極坐標空間，polar coordinate system)，使得旋轉不變性在新空間中成立 [5]。

本研究試圖提出一個不同的思路來讓卷積類神經網路具備一定程度的旋轉不變性。與上述的兩類方法不同的地方在於：我們不希望新的技術使用任何資料擴增的前處理；另外，我們希望能夠維持在原有的歐式空間直接做運算，不需要負荷空間轉換的額外成本。

3. RotInv-ConvNet 模型設計

RotInv-ConvNet 與一般卷積類神經網路的關鍵不同之處在於我們加入了一個對旋轉及翻轉不敏感的函數 g_θ

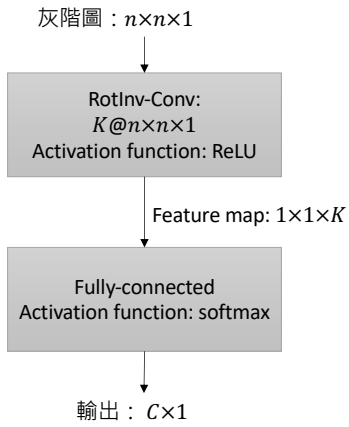

圖 1. 針對尺寸為 $n \times n$ 灰階圖片的 RotInv-ConvNet 設計，K 為卷積的個數，C 為類別的數量

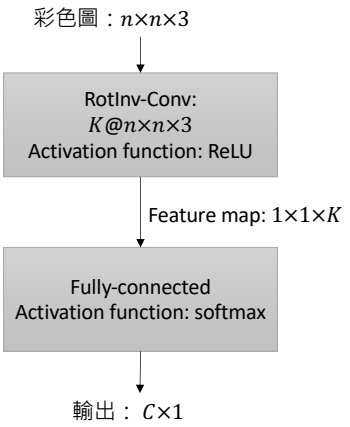

圖 2. 針對尺寸為 $n \times n$ 彩色圖片的 RotInv-ConvNet 設計，K 為卷積的個數，C 為類別的數量

(θ 為待學習的參數)，以下說明此函數的具體作法。以下的說明以灰階圖片為例，故輸入圖片的頻道 (channel) 數為 1，我們將在最後說明如何將 RotInv-ConvNet 延伸至彩色圖片。

A. 函數參數之初始化

給定一個二維且大小為 $n \times n$ 的矩陣 $M = [m_{p,q}]$ $(p, q \in \{0, 1, \ldots, n-1\})$。我們根據式 1 計算 M 水平翻轉 (horizontal flip，簡寫為 HF) 的結果。

$$HF\left([m_{p,q}]\right) = [m_{p,n-1-q}]\,. \tag{1}$$

接下來定義操作 R 為將 $n \times n$ 的矩陣 M 進行 90 度逆時針旋轉，其輸入和輸出如式 2。

$$R([m_{p,q}]) = [m_{n-1-q,p}] \tag{2}$$

我們初始化卷積層的方式如下。首先，給定一個任意的 $n \times n$ 的矩陣 M，再將 M 經過式 3的運算來初始化抗旋轉的卷積層 g_θ 的參數。

$$\begin{aligned} g_\theta = \frac{1}{8}\Big(&M + R(M) + R\circ R(M) + R\circ R\circ R(M) + \\ &HF(M) + R\circ HF(M) + R\circ R\circ HF(M) + \\ &R\circ R\circ R\circ HF(M)\Big), \end{aligned} \tag{3}$$

其中，$\circ$ 代表複合函數 (function composition)，即：$f_1 \circ f_2(x) = f_1(f_2(x))$。

經過式 3 初始化後的卷積層 g_θ 是一個水平翻轉、垂直翻轉、或順逆時針旋轉 90 度、180 度、270 度均不會改變的矩陣。

若抗旋轉卷積層需要 K 個卷積，我們只需要先隨機產生 K 個大小為 $n \times n$ 的矩陣 M^k $(k \in \{1, \ldots, K\})$, 再將每個 M^k 按式 3 計算得到初始化的卷積 g_θ^k。

B. 網路結構與學習機制

我們使用未經資料擴增的訓練資料集訓練 RotInv-ConvNet，此網路僅包括一個抗旋轉的卷積層及一個全連接網路將結果轉為長度為 C 的陣列 (C 是分類問題的類別數量)，其網路結構如圖 1所示。假設某一卷積 g_θ^k 經過式 3 初始化後，其參數的值為 $g_\theta^k = \left[\theta_{p,q}^k\right]$ $(p, q \in \{0, \ldots, n-1\}, k \in \{1, \ldots, K\}$，$K$為卷積的個數)，在做反向傳播 (backpropagation) 時，我們先計算出損失 ℓ 對每個 $\theta_{p,q}^k$ 的偏微分值 $\delta_{p,q}^k := \partial\ell/\partial\theta_{p,q}^k$，若定義 $\Delta^k = \left[\delta_{p,q}^k\right]$，為了確保 g_θ^k 在更新參數後仍保持抗旋轉的特性，我們以式 4計算一個對旋轉不敏感的梯度矩陣 $\widehat{\Delta^k}$。

$$\begin{aligned} \widehat{\Delta^k} = \left[\widehat{\delta^k}_{p,q}\right] = \frac{1}{8}\Big(&\Delta^k + R(\Delta^k) + R\circ R(\Delta^k) \\ &+ R\circ R\circ R(\Delta^k) + HF(\Delta^k) + R\circ HF(\Delta^k) + \\ &R\circ R\circ HF(\Delta^k) + R\circ R\circ R\circ HF(\Delta^k)\Big). \end{aligned} \tag{4}$$

最後，我們可以利用梯度下降法 (gradient descent) (如：式 5) 或其他以梯度為導向的數值最佳化方法來更新 g_θ^k 的參數。

$$\theta_{p,q}^k := \theta_{p,q}^k - \eta\widehat{\delta^k}_{p,q}, \tag{5}$$

其中 η 是學習速度 (learning rate) 超參數。

C. 為何 RotInv-ConvNet 具備抗旋轉性質

我們使用式 3來初始化一個特別的矩陣，此初始矩陣經過水平翻轉、垂直翻轉、旋轉 90 度、180 度、270 度都仍與原矩陣相同。我們運用相同的策略，確保梯度矩陣也是一個經過水平翻轉、垂直翻轉、旋轉 90 度、180 度、270 度都不會變動的矩陣，因此，利用梯度下降法訓練時，得到的 K 個卷積 $g_\theta^1, \ldots, g_\theta^K$ 也都是經過水平翻轉、垂直翻轉、旋轉 90 度、180 度、270 度不會變動的矩陣。因此，這種特別的卷積對圖片的旋轉和翻轉比較不敏感。

D. 彩圖的抗旋轉卷積層設計

以上的敘述雖然針對灰階圖片做說明，但只需經簡單的調整即可適用至彩色圖片上。具體而言，對於彩色圖片的紅色 (R)、綠色 (G)、藍色 (B) 三個頻道，我們各自按照式 3 初始化二維矩陣，並分別使用式 4 計算出對稱的二維梯度矩陣，再以式 5 迭代計算每個參數的值。最後，將各個頻道的二維矩陣疊成三維的張量，即可做為針對彩圖的卷積。彩色圖片的 RotInv-ConvNet 網路結構如圖 2 所示。

4. 實驗

A. 實驗資料

本實驗使用 MNIST、FashionMNIST、CIFAR-10 三個經典的資料集進行實驗。其中，MNIST 為灰階的手寫數字圖片 $(0, 1, \ldots, 9)$，每張圖片的尺寸是 28×28，共有 $60,000$ 張訓練及 $10,000$ 張測試圖片。FashionMNIST 則是灰階的衣飾圖片，包括：襯衫、長褲、毛衣等十類衣飾，每張圖片的尺寸及訓練資料和測試資料的筆數都和 MNIST 相同。CIFAR-10 則是由 $50,000$ 張訓練圖片及 $10,000$ 張測試圖片構成，每張均為 32×32 的彩色圖片，共有飛機、汽車、鳥、狗等十種生活中常見的物件。

我們將 MNIST、FashionMNIST、CIFAR-10 的每個測試資料集又分為以下三種。第一種為原始的測試資料集，以下標示為 Test；第二種是將測試資料集中的圖片隨機旋轉 90 度的整數倍 (即：90 度、180 度、270 度)，以下標示為 Rot90X；第三種則是將測試資料集中的圖片隨機旋轉任意角度，以下標示為 RotRand。

訓練資料方面包括兩種：第一種直接使用原始訓練資料，不做任何資料擴增前處理；第二種則在訓練資料中加上旋轉資料擴增。RotInv-ConvNet 只使用未經資料擴增的訓練資料。兩個要比較的基礎模型 (ConvNet1 及 ConvNet2，模型架構在下節說明) 若使用經過擴增的資料做訓練時，將標示為 AUG，若只使用原始資料做訓練，則不做特別標註。

B. 比較對象

我們將 RotInv-ConvNet 與兩種卷積類神經網路比較。

第一種卷積類神經網路包括兩層卷積層、一個池化層 (pooling layer)、及兩層全連接層，網路結構如圖 3 所示。以下標示為 ConvNet1。這種架構是非常經典的卷積類神經網路架構，與其他卷積類神經網路的主要差別只在於超參數的選擇，如：卷積層層數、全連接層目、池化層數目、卷積層或池化層的大小等。

第二種卷積類神經網路 (以下標示為 CovNet2) 的架構與 RotInv-ConvNet 大致架構相同，第一層的卷積層大小取決於輸入圖片尺寸，最後加上一層全連接層。與 RotInv-ConvNet 的卷積不同之處在於，RotInv-ConvNet 的卷積在初始化以及計算梯度後，會根據式 3 及式 4 調整之，而 ConvNet2 則完全隨機初始化卷積層，也不會對梯度矩陣做式 4 的調整。

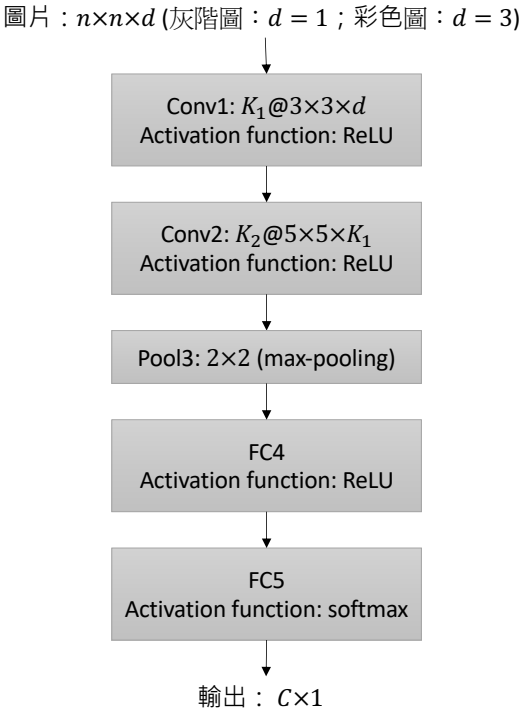

圖 3. ConvNet1 架構

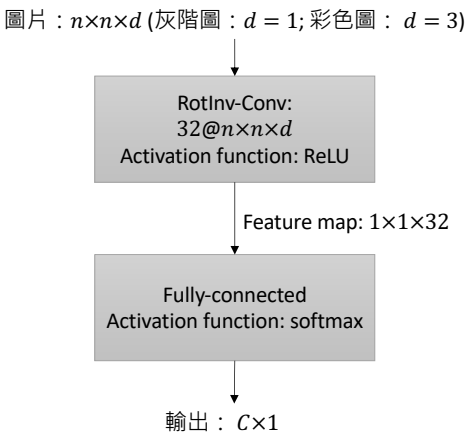

圖 4. ConvNet2 架構

C. 實驗結果

我們比較各種模型在不同的訓練資料集及不同的資料前處理方式後，在各項指標的表現。我們比較的指標包括：測試資料的分類正確率 (accuracy，愈高愈好)、測試資料的交叉熵損失 (cross-entropy loss，愈低愈好)、及模型的參數量 (一般認為在相同的預測效果下參數量愈低愈好)。結果顯示在表 I中。

首先，RotInv-ConvNet 如我們預期的對於 Test 與 Rot90X 的結果始終保持一致，這是因為即使模型未看過 Rot90X 的訓練資料，旋轉 90 度的整數倍後的圖

表 I
各種模型在測試資料上的交叉熵損失及正確率比較。AUG 代表訓練資料中加入隨機翻轉的圖片做資料擴增。對於未使用資料擴增者，我們以粗體標出 ConvNet1、ConvNet2、及 RotInvCovNet 在各種測試資料集下交叉熵損失最低者及正確率最高者。

		MNIST			FashionMNIST			CIFAR-10		
		Test	Rot90X	RotRand	Test	Rot90X	RotRand	Test	Rot90X	RotRand
ConvNet1 (AUG)	loss	0.1266	0.1159	0.1330	0.5251	0.5087	0.5251	1.3254	1.3319	1.3282
	accuracy	0.96	0.96	0.96	0.81	0.82	0.81	0.52	0.53	0.52
ConvNet2 (AUG)	loss	0.5015	0.4993	0.5035	0.8005	0.7920	0.8111	1.8655	1.8632	1.8598
	accuracy	0.85	0.85	0.84	0.71	0.71	0.71	0.33	0.33	0.34
ConvNet1	loss	**0.0223**	6.9584	4.3955	**0.2297**	8.4026	5.3514	**0.8124**	2.9429	3.1170
	accuracy	**0.99**	0.17	0.42	**0.93**	0.07	0.22	**0.72**	0.29	**0.30**
ConvNet2	loss	0.1006	8.3428	5.8147	0.3594	10.0771	6.3615	1.5459	2.3862	2.4464
	accuracy	0.97	0.16	0.33	0.87	0.02	0.19	0.45	0.24	0.23
RotInv-ConvNet	loss	0.8478	**0.8478**	**2.6548**	0.6145	**0.6145**	**3.7220**	1.7946	**1.7946**	**2.0903**
	accuracy	0.72	**0.72**	**0.43**	0.79	**0.79**	**0.33**	0.36	**0.36**	0.26

表 II
模型參數量

	MNIST	FashionMNIST	CIFAR10
ConvNet1	1,044,096	1,044,096	1,437,888
ConvNet2	25,440	25,440	98,656
RotInv-ConvNet	6,624	6,624	24,928

表 III
訓練每一輪 (epoch) 所需的時間（秒）

	MNIST	FashionMNIST	CIFAR10
ConvNet1	193	184	206
ConvNet2	17	18	21
RotInv-ConvNet	18	19	20

片與未旋轉的圖片在經過對稱卷積層運算後會得到相同的輸出，實驗結果符合我們設計時想要達成的效果。另外，同樣從表 I中顯示：若訓練資料未經過資料擴增的前處理，RotInv-ConvNet 在圖片旋轉過的測試資料 (i.e., Rot90X 與 RotRand) 的正確率及交叉熵損失普遍優於基準模型的 ConvNet1 及 ConvNet2。

然而，RotInv-ConvNet 在未經旋轉測試資料 (i.e., Test) 上效果則遜於 ConvNet1 及 ConvNet2。可能有兩個原因導致這種結果。其一，RotInv-ConvNet 的卷積層中的參數數量小於相同大小的一般卷積層，故其假設空間 (hypothesis space) 較小。其二，RotInv-ConvNet 在對圖片進行卷積運算時，因為卷積的長寬與圖片相同，故圖片的空間資訊消失。

此外，若將訓練資料的圖片以旋轉進行資料擴增再使用 ConvNet1 或 ConvNet2 訓練，產生的模型無論在 Test、Rot90X、及 RotRand 上的效果則優於 RotInv-ConvNet 的效果。就實務上而言，以旋轉進行資料擴增仍是較佳的做法。

在模型的參數量上，ConvNet1 的的參數量是 RotInv-ConvNet 的 50 倍以上至 150 倍以上 (細節如表 II)。這也反映在每一輪 (epoch) 所需訓練的時間上，兩者相差約十倍 (如表 III 所示)。若與 ConvNet2 比較，RotInv-ConvNet 由於上下左右的對稱性質，所需的參數量約為 ConvNet2 的 1/4 (參閱表 II)，訓練時間則大致相同 (參閱表 III)。

5. 結論與未來展望

我們提出的新模型 RotInv-ConvNet 利用對卷積層的限制來讓類神經網路對圖片的旋轉產生抗性，這是因為受限制的卷積層中，對於旋轉 90 度的整數倍及水平、垂直翻轉的圖片會產生相同的輸出。另外，此限制附帶達成減少參數以及提高訓練速度的效果。然而，實證上發現：這種限制需要付出對未經旋轉的圖片辨識度下降的代價。因此，我們的第一個未來目標是繼續延伸 RotInv-ConvNet，使得它能在不影響未旋轉圖片的辨識率的前提下仍能有抗旋轉的能力。

目前 RotInv-ConvNet 的卷積層被限制成與輸入圖片大小一致，導致訊息快速的被濃縮成一個純量值 (scalar)。雖然在同一層能夠以多個卷積運算來讓圖片轉成向量而非純量，但圖片中的空間訊息仍然會因為新的卷積運算而消失。因此，我們的另一目標是嘗試設計長寬尺寸較小的抗旋轉卷積運算單元。當神經網路包含空間資訊後，或許在一定程度上也能提高圖片的辨識率。

References

[1] A. Chaman and I. Dokmanic, "Truly shift-invariant convolutional neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 3773–3783.

[2] Y. Xu, T. Xiao, J. Zhang, K. Yang, and Z. Zhang, "Scale-invariant convolutional neural networks," *arXiv preprint arXiv:1411.6369*, 2014.

[3] G. Cheng, P. Zhou, and J. Han, "Rifd-cnn: Rotation-invariant and fisher discriminative convolutional neural networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2884–2893.

[4] B. Fasel and D. Gatica-Perez, "Rotation-invariant neoperceptron," in *18th International Conference on Pattern Recognition (ICPR'06)*, vol. 3. IEEE, 2006, pp. 336–339.

[5] J. Kim, W. Jung, H. Kim, and J. Lee, "Cycnn: A rotation invariant cnn using polar mapping and cylindrical convolution layers," *arXiv preprint arXiv:2007.10588*, 2020.