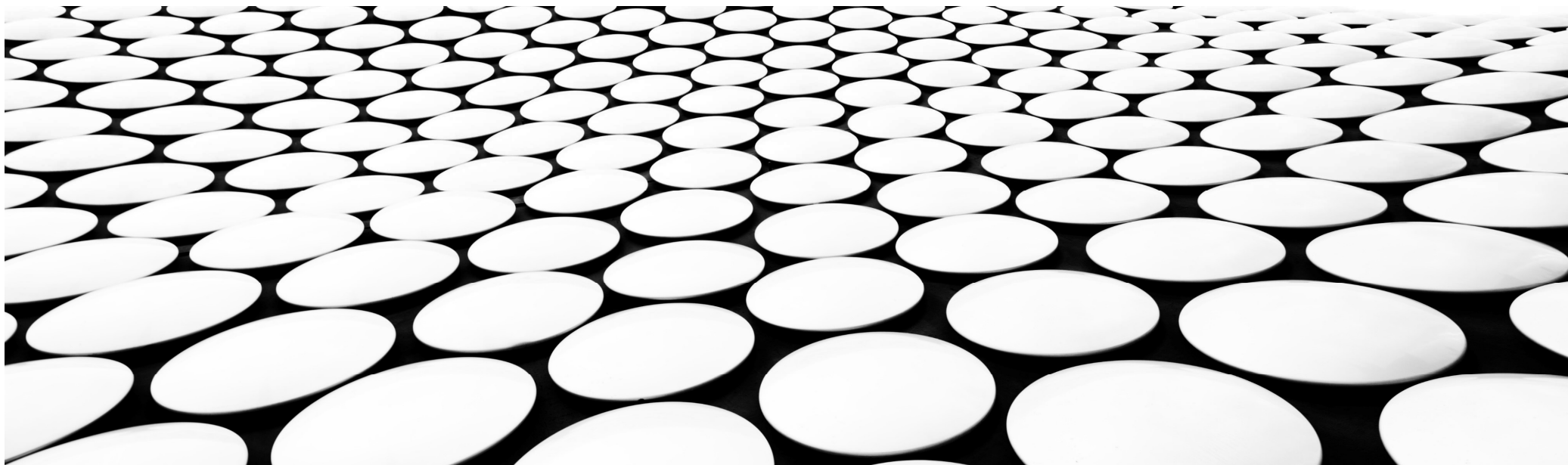

機器學習期末論文研讀報告

常博愛 資工三408410086



研讀列表

	Title	From	Time	Author
1.	Neural machine translation by jointly learning to align and translate	Accepted at ICLR 2015 as oral presentation	<i>[Submitted on 1 Sep 2014 (v1), last revised 19 May 2016 (this version, v7)]</i>	<u>Dzmitry Bahdanau</u> , <u>Kyunghyun Cho</u> , <u>Yoshua Bengio</u>
2.	Visualizing and Understanding Convolutional Networks	arXiv:1311.2901 [cs.CV]	<i>[Submitted on 12 Nov 2013 (v1), last revised 28 Nov 2013 (this version, v3)]</i>	Matthew D. <u>Zeiler</u> and Rob Fergus
3	A neural probabilistic language model	Part of Advances in Neural Information Processing Systems 13 (NIPS 2000)	2000	<u>Yoshua Bengio</u> <u>Réjean Ducharme</u> Pascal Vincent Christian <u>Jauvin</u>

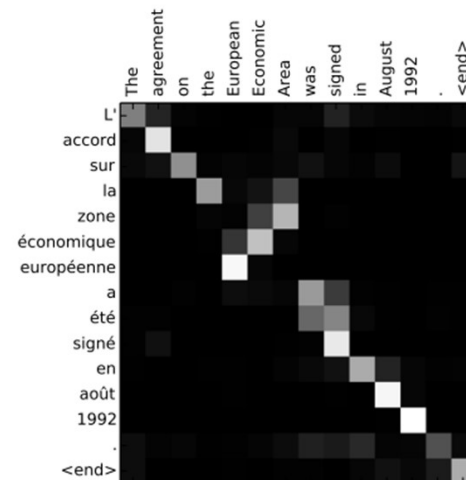
1. NEURAL MACHINE TRANSLATION BY JOINTLY LEARNING TO ALIGN AND TRANSLATE

背景：

- 以往神經網路再翻譯的任務上都是建立在encoder-decoder的架構上，**固定長度的context vector會成為效能的bottleneck**，所以這篇文章提出了dynamic context vector的概念，讓model自己去搜尋input和related predict的相關部分，使得效能增加。
- 這也是**最早提出attention mechanism**的paper。

主要内容：

- 模型會自到去找出source sentence中對當前最富有資訊量的部分，並透過這些部分與context vector結合來進行預測。也就是引入注意力機制，看當前哪個部分是最重要的！
- 比較attention(左邊)機制下和原本encoder-decoder(右邊)的效果



1. NEURAL MACHINE TRANSLATION BY JOINTLY LEARNING TO ALIGN AND TRANSLATE

心得：由於encoder-decoder具有**context vector**必須要足以包含整個**input sentence**的**information**否則效果就會不好的缺點，當句子長度變長時這個問題就會浮現出來。而引入注意力機制則可以把所有input sequence壓成一個固定的context vector，透過了attention機制選擇一個subset區域，特別針對那一區段來動態產生context vector，使得context vector能夠更有變化。

2. VISUALIZING AND UNDERSTANDING CONVOLUTIONAL NETWORKS

背景：

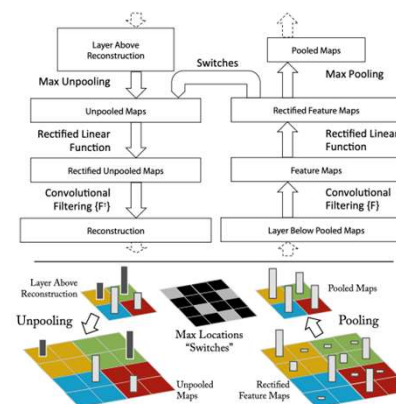
近年來CNN興起的三個因素：

1. 越來越多的有label的圖片訓練資料集
2. 越來越強的GPU
3. 更好的正則化策略，例如dropout

然而即便CNN架構在圖像分類領域上取得了巨大的成功，但是，還沒有研究去說他們為何可以有如此好的表現，以及應該如何針對模型架構去改進。因此這篇文章透過**反卷積(DeConv)**的技術提出了一種**視覺化方法**，使得可以去觀察每一層的Conv layer。此外這篇文章還透過遮擋圖片的一部分來觀察模型輸出，藉此找出圖片中哪些區塊對於模型分類是比較重要的部分。

主要内容：

要去解釋CNN的一種方式就是看中間層的feature map哪些被激活了，作者透過Deconvnet來將特定的feature map映射回到輸入層，這樣就可以看到feature map和input layer之間的對應關係。流程如下：



2. VISUALIZING AND UNDERSTANDING CONVOLUTIONAL NETWORKS

主要内容:

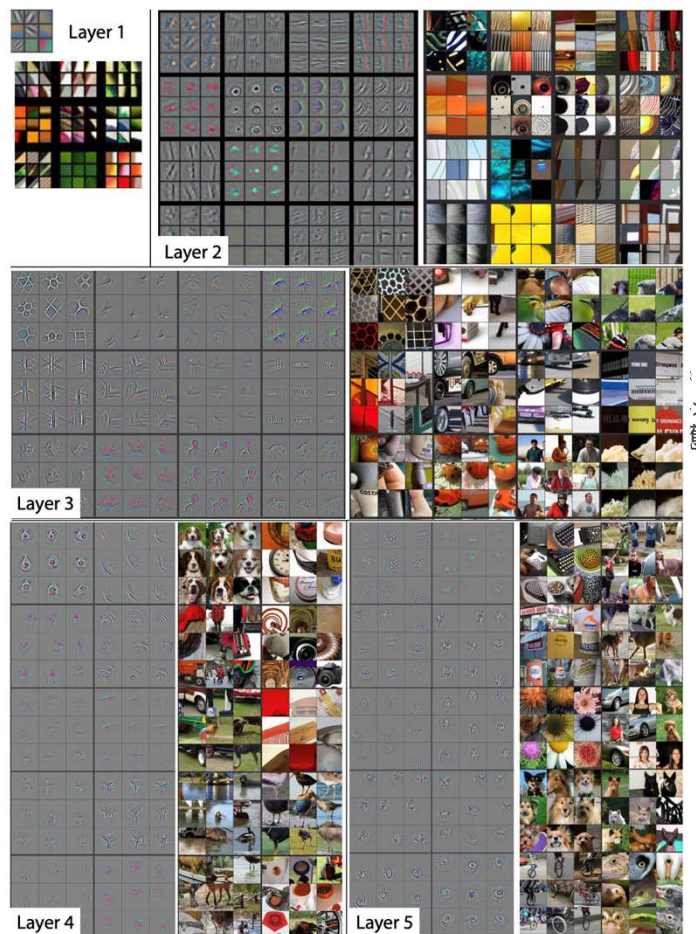
2.Convnet Visualization

图片中左半邊是DeConv的視覺化，右半邊的是該點(activation)對應圖片的感受野區塊(receptive field)，也就是從該點往回推，當推回input layer時被影響到的範圍。比方說input layer \rightarrow (3x3 Conv) \rightarrow output layer，那在output layer的一個點實際上對應到了input layer的3x3區塊

從這張圖中我們觀察到:

show 9個activation的原因是，透過多張圖來觀察，可以發現Conv具備空間不變性: 即使圖片有差異，但學到的特徵都很類似

前幾層學到的特徵都在一些顏色或邊角特徵、越後面則會學到較複雜的紋理、局部、整體特徵



2. VISUALIZING AND UNDERSTANDING CONVOLUTIONAL NETWORKS

心得：

transpose convolution並不能從feature map還原原始的input data，他只是還原出原始的input shape。(就這點來說其實逆矩陣，) 但這篇論文的貢獻在於提出了一個CNN可視化的方法，透過反卷積，使得我們可以去觀察模型不同層對應到input時哪些是比較重要的部分，這對於後續的解釋性AI(Explainable AI, XAI)有非常重要的影響。

3. A NEURAL PROBABILISTIC LANGUAGE MODEL

主要内容:

所謂的 *Neural Probabilistic Language Model*，即是用類神經網路，從語料庫中計算出各個詞彙的語意向量值。

其主要设计类型为一个语言模型，通过对词汇进行embedding预测或验证接下来的语言会表达什么。在輸入層中將詞映射為一個m列的向量，即該詞的向量表示。

語言模型NPLM: (如右圖):

輸入映射層

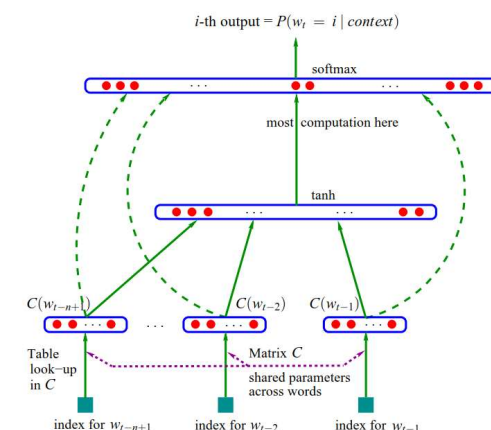
輸入是當前詞 w_t 的前 n 個詞($w_{t-1} \sim w_{t-n+1}$)經過C matrix映射後到了映射層，模型訓練的開始C matrix可以隨機初始化。

將各個詞向量串聯起來作為隱藏層的輸入

隱藏層

由映射層到隱含層函數， y_i 表示詞 w_t 出現的概率

- y 為未經過softmax歸一後的輸出
- h 是隱層節點個數， m 為詞向量的維度， $|V|$ 為詞彙表大小， d 為輸入-隱層的bias， b 為隱層-輸出的bias輸出層
- b ($|V|$ 維): 隱層-輸出層的bias
- d (h 維): 輸入-隱層的bias
- U ($|V| * h$ matrix): 隱層-輸出層的權重矩陣
- H ($h * (n-1)m$ matrix): 輸入-隱層的權重矩陣
- C ($|V| * m$): 詞向量



輸出層

- 由隱藏層到輸出層歸一化指數函數softmax
- 因為是語言模型，需要根據前 n 個單詞預測下一個單詞，所以是一個多分類器，用softmax。整個模型最大的計算量集中在最後一層上，因為一般來說詞彙表都很大，需要計算每個單詞的條件概率

$$\hat{P}(w_t | w_{t-1}, \dots, w_{t-n+1}) = \frac{e^{y_{w_t}}}{\sum_i e^{y_i}}$$

Stochastic gradient ascent

- ϵ 是“學習率”
- 所有單詞 j 的單詞屬性 $C(j)$ 沒有在輸入層出現過。

$$\theta \leftarrow \theta + \epsilon \frac{\partial \log \hat{P}(w_t | w_{t-1}, \dots, w_{t-n+1})}{\partial \theta}$$

3. A NEURAL PROBABILISTIC LANGUAGE MODEL

心得：

Neural Probabilistic Language Model 是2003年期間所提出的語言模型，建構在 *Neural Probabilistic Language Model* 的基礎之上，發展出了各種類神經網路相關的語言模型，一再突破了以往自然語言處理的瓶頸，包括近兩年來很熱門的 *word2vec*，也是從它發展而來。

論文綜合比較與心得總結

- 1.首先：從相同點來說，三篇論文的研究方向都是類神經網路深度學習（CNN），並且都是影響深遠的開山之作。
- 2.另外本人在研讀過三篇論文後才發現正確的閱讀順序建議為2-3-1。其中第一篇和第三篇分別採用CNN加一定的演算法來解決自然語言處理問題，而第二篇則是側重通過介紹反捲積來visualization每層feature map以試圖證明CNN方法可行性的依據，為反駁機器學習不是黑盒子做出了傑出貢獻。
- 3.至於不同點，第三篇論文試圖使用幾率機制將文章的單詞進行embedding,但得到的結果會是固定向量，在文章語句過長的情況下變回產生問題。而引入注意力機制的第一篇論文則在繼承前輩embedding的基礎上拋棄了固定in-decoder的方法，轉而設計動態向量。而第二篇則是研究影像問題，除了上述目的，還通過遮掩圖片的一部分來凸顯CNN學習的主要特征為何。
- 4.通過三篇論文的學習，除了論文本身的知識以外，讓我更受啟發的是在做研究的過程中需要保證方法的可解釋性，並且要能夠推陳出新，在現代GPU與平行計算等強大算力的加持下，機器學習的應用得到了空前發展，因此更應該多多創新，不能僅局限於做調參俠，而是多思考是否可以通過結合最新型的多領域演算法以解決更多問題的同時提高效率。