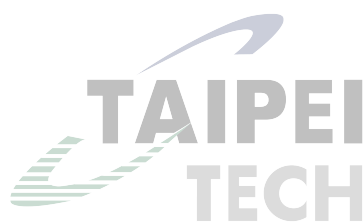


國立臺北科技大學  
資訊工程研究所

碩士學位論文  
編號：106598056

解析卷積神經網路於物件偵測

研究生：謝柏鋒





國立臺北科技大學

資訊工程研究所

碩士學位論文

解析卷積神經網路於物件偵測  
Analysis convolutional neural network of  
object detection



研究生：謝柏鋒

指導教授：謝東儒 博士

中華民國 一百零七 年 六 月

國立臺北科技大學  
研究所博士學位論文口試委員會審定書

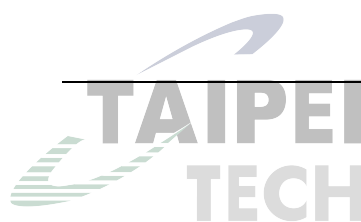
本校\_\_\_\_\_研究所\_\_\_\_\_君

所提論文，經本委員會審定通過，合於博士資格，特此證明。

學位考試委員會

委 員：\_\_\_\_\_

\_\_\_\_\_



\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

指導教授：\_\_\_\_\_

所 長：\_\_\_\_\_

中 華 民 國 一 百 零 四 年 ○ 月 ○ 日

國立臺北科技大學  
研究所碩士學位論文口試委員會審定書

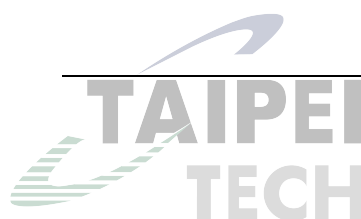
本校\_\_\_\_\_研究所\_\_\_\_\_君

所提論文，經本委員會審定通過，合於碩士資格，特此證明。

學位考試委員會

委員：\_\_\_\_\_

\_\_\_\_\_



\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

指導教授：\_\_\_\_\_

所長：\_\_\_\_\_

中華民國 一百零四 年 ○ 月 ○ 日

# 中文摘要

論文名稱：解析卷積神經網路於物件偵測

頁數：五十頁

校所別：國立臺北科技大學資訊工程研究所

畢業時間：一百零七學年度第二學期

學位：碩士

研究生：謝柏鋒

指導教授：謝東儒博士



關鍵詞：YOLO、物件偵測、機器學習

近年來，卷積神經網路有許多令人突破性的發展。我們的目的是在於解析最近對於物件偵測技術的性能分類非常良好的 YOLO(You Only Look Once)。對於解析深度學習已經有許多相關的論文，但對於解析卷積神經網路的相關論文卻較少，本論文是用簡易的方式說明卷積神經網路的運作方式，並且將過程呈現，可以讓一般大眾更加認識機器學習的運作方式，同時也讓專家方便於解析卷積神經網路的架構，並且能夠迅速改善原本的架構，使其加速。

## 英文摘要

Title: DAnalysis convolutional neural network of object detection

Pages: 50

School: National Taipei University of Technology

Department: Electrical Engineering

Time: June, 2019

Degree: Master

Researcher: PO-FONG HSIEH

Advisor: TUNG-JU HSIEH, Ph.D.



Keywords:

Start writing abstract from here. Start writing abstract from here. Start writing abstract from here. Start writing abstract from here. Start writing abstract from here. Start writing abstract from here. Start writing abstract from here. Start writing abstract from here.

# 致謝

所有對於研究提供協助之人或機構，作者都可在誌謝中表達感謝之意。  
標題使用 20pt 粗標楷體，並於上、下方各空一行 (1.5 倍行高，字型 12pt 空行) 後鍵入內容。致謝頁須編頁碼 (小寫羅馬數字表示頁碼)。

## I. 此範本參考下列網站的資料：

- [台大碩博士論文 LaTeX 範本](#)
- [陳念波老師的元智大學論文樣板](#)
- [台灣科技大學同學編寫的碩博士論文 Latex 模板](#)

## II. 原作者參考並修改自下列網站的資料：

- [如何用 LaTeX 排版臺灣大學碩士論文](#)  
—台灣大學論文 L<sup>A</sup>T<sub>E</sub>X 樣版原創者 [黃子桓](#) 的教學網頁
- [LaTeX 常用語法及論文範本](#)  
—[Hitripod](#) 所修改的範本，這裡參考了許多他所寫的格式和內容
- [使用 LaTeX 做出精美的論文](#)
- [XeTeX：解決 LaTeX 惱人的中文字型問題](#)
- [台灣大學碩士、博士論文的 Latex 模板](#)

# 目 錄

口試委員會審定書	i
中文摘要	iii
英文摘要	iv
致謝	v
目 錄	vi
圖 目 錄	viii
表 目 錄	ix
1 導論	1
1.1 導論	1
1.2 Table	3
2 相關文獻討論	4
2.1 Manga Vectorization and Manipulation with Procedural Simple Screentone. . .	4
2.1.1 回答下列問題	6
2.2 YOLO9000:Better, Faster, Stronger. . . . .	7
2.2.1 回答下列問題	8
2.3 Deep Residual Learning for Image Recognition. 卷積影像深度學習 . . . . .	9
2.3.1 回答下列問題	11
2.4 Only Look Once, Mining Distinctive Landmarks from ConvNet for Visual Place Recognition. 只看一次，在 ConvNet 找到特殊的地標，用於地點識別 . . .	12
2.4.1 回答下列問題	13
2.5 Visualizing and Understanding Convolutional Networks . . . . .	14
2.5.1 回答下列問題	15
2.6 Visualizing Convolutional Neural Networks for Image Classification . . . . .	15
2.6.1 回答下列問題	16
2.7 Visualization of Neural Network Predictions for Weather Forecasting . . . . .	17
2.7.1 回答下列問題	19
2.8 Deep learning for computational biology . . . . .	19
2.8.1 回答下列問題	20



<b>3</b>	<b>方法</b>	<b>22</b>
3.1	解析 YOLO 黑盒子 . . . . .	22
3.1.1	Darknet-19 表格 . . . . .	22
<b>4</b>	<b>結果與討論</b>	<b>24</b>
4.1	第一階層子標題 . . . . .	24
4.1.1	第二階層子標題 . . . . .	24
<b>5</b>	<b>結論</b>	<b>25</b>
5.1	結論 . . . . .	25
5.2	未來展望 . . . . .	25
	<b>附錄 A：第一個附錄名稱</b>	<b>26</b>
	<b>附錄 B：第二個附錄名稱</b>	<b>27</b>
	<b>符號彙編</b>	<b>28</b>



## 圖 目 錄

1.1	Yolo Predivtions. . . . .	1
1.2	Yolo v3 與其他演算法示意圖. . . . .	2
1.3	試誤法 . . . . .	2
1.4	系統示意圖. . . . .	3
3.1	原圖 . . . . .	22
3.2	偵測結果 . . . . .	22



# 表 目 錄

1.1 各演算法派別比較 . . . . .	3
------------------------	---



# 第一章 導論

## 1.1 導論

卷積神經網路為近年來在物件偵測及辨識度上表現最為突出的深度學習架構，本論文中，我們選定現在在物件偵測上效能與辨識度最高的 YOLO(You Only Look Once) 作為解析的範本，介紹卷積神經網路的運作過程，比較其他種物件偵測的演算法，拆解 YOLO 內部的運作過程，探討 YOLO 能夠領先於其他演算法的關鍵原因，並改善其只能在 C++ 或 python 中使用之限制，改為在 HTML 上使用。

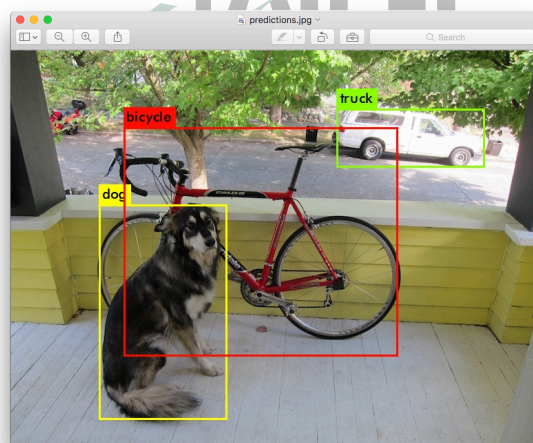


圖 1.1: Yolo Predivtions.

在卷積神經網路中，目前我們所知若是將神經網路的深度加高，可以提升準確率，但過多的深度卻會造成其效率降低。

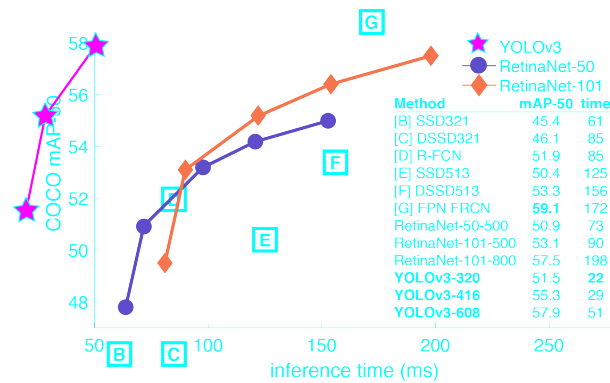


圖 1.2: Yolo v3 與其他演算法示意圖.

所以現今大部分的演算法只能使用試誤法，以低效率的方式去測驗比較此演算法，粗略估計此演算法的深度與效能達到平衡，常常會有過高的深度或準確率未達預期的情形發生，為了解決此問題，本論文著重於解析卷積神經網路的架構及其運作方法，以簡單明瞭的方式闡述其過程，並且分析原本架構的問題，改善原本的架構。

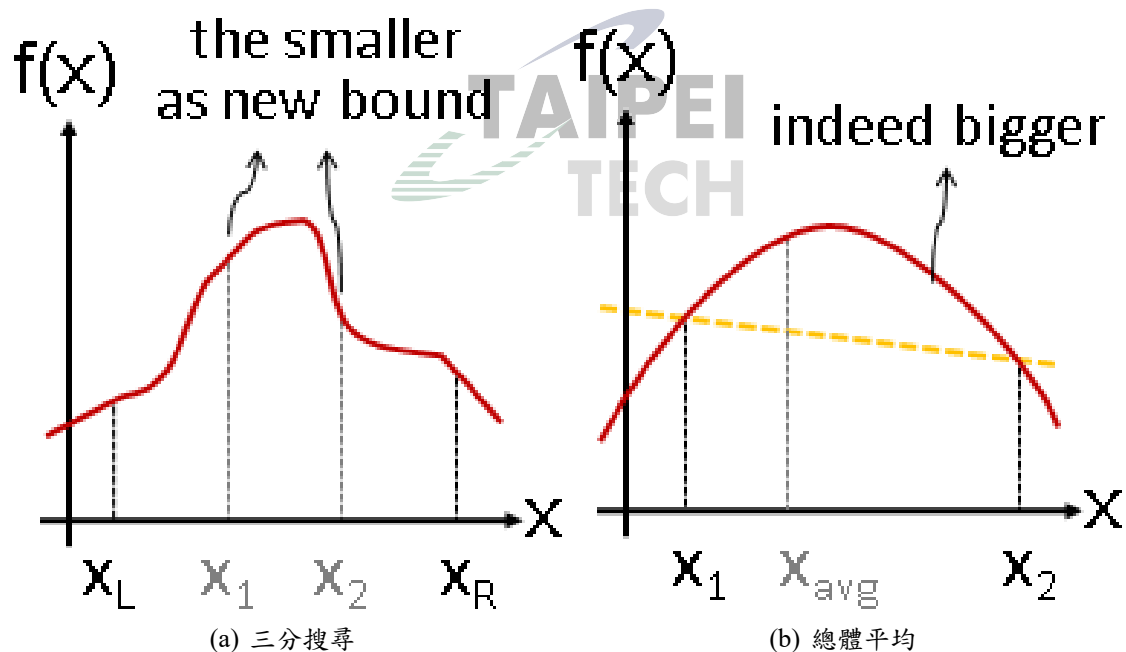


圖 1.3: 試誤法

以往的論文幾乎沒有以視覺化的方式解析卷積神經網路的運作原理，所以在本論文中，希望能以簡單清晰的方式讓人們了解卷積神經網路，並且能在網頁上呈現，其操作性較為直覺性，能夠馬上知道目前選取的位置的功能與效果。

原本使用 Yolo 演算法時，必須先完成其環境安裝，並使用程式的指令才能完成其物件

Layer	Type	Filters	Size/Stride	Output size
0	Convolutional	32	3x3/1	416x416
1	Maxpool		2x2/2	208x208
2	Convolutional	64	3x3/1	208x208
3	Maxpool		2x2/2	104x104
4	Convolutional	128	3x3/1	104x104
5	Convolutional	64	3x3/1	104x104
6	Convolutional	128	3x3/1	104x104
7	Maxpool		2x2/2	52x52
8	Convolutional	256	3x3/1	52x52
9	Convolutional	128	1x1/1	52x52
10	Convolutional	256	3x3/1	52x52
11	Maxpool		2x2/2	26x26
12	Convolutional	512	3x3/1	26x26
13	Convolutional	256	3x3/1	26x26
14	Convolutional	512	3x3/1	26x26
15	Convolutional	256	1x1/1	26x26
16	Convolutional	512	3x3/1	26x26
17	Maxpool		2x2/2	13x13
18	Convolutional	1024	3x3/1	13x13
19	Convolutional	512	1x1/1	13x13
20	Convolutional	1024	3x3/1	13x13
21	Convolutional	512	1x1/1	13x13
22	Convolutional	1024	3x3/1	13x13
23	Convolutional	1024	3x3/1	13x13
24	Convolutional	1024	3x3/1	13x13
25	Route	16		
26	Convolutional	64	1x1/1	13x13
27	Reorg		/2	13x13
28	Route	27*24		
29	Convolutional	1024	3x3/1	13x13
30	Convolutional	425	1x1/1	13x13
31	Detection			



圖 1.4: 系統示意圖.

偵測。本論文提出一個線上的使用環境，只要將圖片拖曳到網頁中，就可以偵測到其內容出現的物件，大幅改善原本使用的限制。

## 1.2 Table



表 1.1: 各演算法派別比較

名稱	應用範圍
符號論學派	用逆向演算法, 用哲學, 心理學, 邏輯思路.
貝氏定理學派	用機率推理, 用統計學
類比推理學派	用支持向量機 (support vector machine), 相似度判斷學學習, 受新理學, 數學最佳化影響
類神經網路學派	用大腦逆向工程 (reverse engineering), 用倒傳遞理論演算法 (back propagation), 受神經科學, 物理學啟發
演化論學派	運用遺傳學 (genetics), 演化生物學 (evolutionary biology), 遺傳程式規劃 (genetic programming)

其中 YOLO 分類於類神經網路學派，使用卷積神經網路實現物件偵測的網路架構。

## 第二章 相關文獻討論

### 2.1 Manga Vectorization and Manipulation with Procedural Simple Screentone.

資料來源: [?]

影片: [影片](#)

1. Abstract—Manga are a popular artistic form around the world, and artists use simple line drawing and screentone to create all kinds of interesting productions.

抽象漫畫是全世界流行的藝術形式，藝術家們用簡單的線條畫和網點來創作各種有趣的作品。

2. Vectorization is helpful to digitally reproduce these elements for proper content and intention delivery on electronic devices.

矢量化有助於數字複製這些元素，以便在電子設備上正確傳送內容和意圖。

3. Therefore, this study aims at transforming scanned Manga to a vector representation for interactive manipulation and real-time rendering with arbitrary resolution.

因此，本研究旨在將掃描的漫畫轉換為交互式操作的矢量表示，並以任意的分辨率進行實時渲染。

4. Our system first decomposes the patch into rough Manga elements including possible borders and shading regions using adaptive binarization and screentone detector.

我們的系統首先使用自適應二值化和網點檢測器將補丁分解成粗糙的漫畫元素，包括可能的邊界和陰影區域。

5. We classify detected screentone into simple and complex patterns: our system extracts simple screentone properties for refining screentone borders, estimating lighting, compensating missing strokes inside screentone regions, and later resolution independently rendering with our procedural shaders.

我們將檢測到的網點分類為簡單和複雜的模式：我們的系統提取簡單的網點屬性，用於細化網點邊界，估計光照，補償網點區域內的缺失筆劃，以及後來使用程序著色器獨立渲染的分辨率。

6. Our system treats the others as complex screentone areas and vectorizes them with our proposed line tracer which aims at locating boundaries of all shading regions and polishing all shading borders with the curve-based Gaussian refiner.

我們的系統將其他視為複雜的屏幕區域，並使用我們提出的線追蹤器對其進行矢量化處理，該追蹤器旨在定位所有陰影區域的邊界，並使用基於曲線的高斯細化器來拋光所有陰影邊界。

7. A user can lay down simple scribbles to cluster Manga elements intuitively for the formation of semantic components, and our system vectorizes these components into shading meshes along with embedded Be'zier curves as a unified foundation for consistent manipulation including pattern manipulation, deformation, and lighting addition.

用戶可以放置簡單的塗鴉，以形成語義組件，直觀地對漫畫元素進行聚類，我們的系統將這些組件與矢量嵌入的 Be'zier 曲線一起矢量化為著色網格，作為包括圖案操作，變形和照明在內的一致操作的統一基礎加成。

8. Our system can real-time and resolution independently render the shading regions with our procedural shaders and drawing borders with the curve-based shader.

我們的系統可以實時和分辨率獨立地渲染陰影區域與我們的程序著色器和基於曲線的著色器繪製邊界。

9. For Manga manipulation, the proposed vector representation can be not only magnified without artifacts but also deformed easily to generate interesting results. 對於漫畫操作，



所提出的矢量表示不僅可以被放大而沒有偽影，而且易於變形以產生有趣的結果。

### 2.1.1 回答下列問題

1. What are motivations for this work? 該論文解決的問題

可以讓漫畫放大且沒有偽影，而且能夠易於變形。

2. What is the proposed solution? 該論文解決的方法

線追蹤器對其進行矢量化處理，該追蹤器旨在定位所有陰影區域的邊界，並使用基於曲線的高斯細化器來拋光所有陰影邊界。

將這些組件與矢量嵌入的貝茲曲線一起矢量化為著色網格，作為包括圖案操作，變形和照明在內的一致操作的統一基礎加成。

3. What is the evaluation of the proposed solution? 該論文作者如何評估其結果? (問卷, 量測時間數據, 實驗...)

Show the result.



4. What are the contributions? 該論文的貢獻為和?

有效的漫畫矢量化和元素分解的操縱管道，語義構件的構建和矢量化，以程式簡單的分佈著色。

5. What are future directions for this research? 未來研究方向可以為何?

希望開發一個多層次的網點探測器或其他類型的結構探測器來進行更強大的探測。

希望開發一個識別程序和一個程序著色器來產生這些美學陰影效果。

希望開發一個基於 Gabor 的標識符，用於避免屬性提取的複雜的網點模式。

上下文組件的完成仍然需要大量的手動工作。更多的自動形狀分解與用戶交互較少，類似於抓斗切割是需要的。

## 2.2 YOLO9000: Better, Faster, Stronger.

資料來源: [?]

影片: [影片](#)

1. We introduce YOLO9000, a state-of-the-art, real-time object detection system that can detect over 9000 object categories.

我們介紹 YOLO9000，這是一種先進的實時物體檢測系統，可以檢測超過 9000 個對象類別。

2. First we propose various improvements to the YOLO detection method, both novel and drawn from prior work.

首先，我們提出了對 YOLO 檢測方法的各種改進，既有新穎的，也有先前的工作。

3. The improved model, YOLOv2, is state-of-the-art on standard detection tasks like PASCAL VOC and COCO.

改進型號 YOLOv2 是標準檢測任務（如 PASCAL VOC 和 COCO）的最新技術。

4. Using a novel, multi-scale training method the same YOLOv2 model can run at varying sizes, offering an easy tradeoff between speed and accuracy. At 67 FPS, YOLOv2 gets 76.8 mAP on VOC 2007.

使用新穎的多尺度訓練方法，相同的 YOLOv2 模型可以以不同的尺寸運行，在速度和準確度之間提供簡單的權衡。在 67 FPS，YOLOv2 在 VOC 2007 上獲得 76.8 mAP。

5. At 40 FPS, YOLOv2 gets 78.6 mAP, outperforming state-of-the-art methods like Faster RCNN with ResNet and SSD while still running significantly faster.

在 40 FPS 時，YOLOv2 獲得 78.6 mAP，優於最先進的方法，如使用 ResNet 和 SSD 的 Faster RCNN，同時運行速度明顯更快。

6. Finally we propose a method to jointly train on object detection and classification. Using this method we train YOLO9000 simultaneously on the COCO detection dataset and the ImageNet classification dataset.

最後，我們提出了一種聯合訓練物體檢測和分類的方法。使用此方法，我們在 COCO 檢測數據集和 ImageNet 分類數據集上同時訓練 YOLO9000。

7. Our joint training allows YOLO9000 to predict detections for object classes that don't have labelled detection data. We validate our approach on the ImageNet detection task.

我們的聯合培訓允許 YOLO9000 預測沒有標記檢測數據的對象類的檢測。我們驗證了 ImageNet 檢測任務的方法。

8. YOLO9000 gets 19.7 mAP on the ImageNet detection validation set despite only having detection data for 44 of the 200 classes.

YOLO9000 在 ImageNet 檢測驗證集上獲得 19.7 mAP，儘管只有 200 個類中的 44 個具有檢測數據。

9. On the 156 classes not in COCO, YOLO9000 gets 16.0 mAP.

不在 COCO 的 156 個類別中，YOLO9000 獲得 16.0 mAP。

10. But YOLO can detect more than just 200 classes; it predicts detections for more than 9000 different object categories. And it still runs in real-time.

但是 YOLO 可以檢測到超過 200 個類別；它預測超過 9000 種不同對象類別的檢測。它仍然可以實時運行。

### 2.2.1 回答下列問題

1. What are motivations for this work? 該論文解決的問題

易於尋找要偵測的物體，且數量高達 9000 種，在時間上能夠在實現 real-time 的偵測。

2. What is the proposed solution? 該論文解決的方法

提出了一種聯合訓練物體檢測和分類的方法，此新穎的多尺度訓練方法，可以在速度和準確度之間提供簡單的權衡。

3. What is the evaluation of the proposed solution? 該論文作者如何評估其結果? (問卷, 量測時間數據, 實驗...)

YOLO9000 在 ImageNet 檢測驗證集上獲得 19.7 mAP 而不在 COCO 的 156 個類別中，YOLO9000 獲得 16.0 mAP。

4. What are the contributions? 該論文的貢獻為和?

提出一種新穎的訓練方法，在物體的偵測上實現 real-time。

5. What are future directions for this research? 未來研究方向可以為何?

希望使用類似的技術進行 weakly 的監督圖像分割。

希望使用更好的演算法以改進偵測的結果。

## 2.3 Deep Residual Learning for Image Recognition. 卷積影像深度學習

資料來源:Microsoft Research.2016 IEEE [?] 影片:[影片](#)

1. Deeper neural networks are more difficult to train.

更深入的神經網絡更難訓練。

2. We present a residual learning framework to ease the training of networks that are substantially deeper than those used previously.

我們提出了一個殘留的學習框架，以便於對比以前使用的網絡深度更深的網絡進行培訓。

3. We explicitly reformulate the layers as learning residual functions with reference to the layer inputs, instead of learning unreferenced functions.

我們明確地將層重新組合為參考層輸入的學習殘差函數，而不是學習未引用的函數。

4. We provide comprehensive empirical evidence showing that these residual networks are easier to optimize, and can gain accuracy from considerably increased depth.

我們提供全面的經驗證據表明這些殘留網絡更容易優化，並且可以從顯著增加的深度獲得準確性。

5. On the ImageNet dataset we evaluate residual nets with a depth of up to 152 layers—8× deeper than VGG nets [40] but still having lower complexity.

在 ImageNet 數據集上，我們評估了深度高達 152 層的殘留網絡 - 比 VGG 網絡深 8 倍 [40]，但仍然具有較低的複雜度。

6. An ensemble of these residual nets achieves 3.57% error on the ImageNet test set. This result won the 1st place on the ILSVRC 2015 classification task.

這些殘留網絡的集合在 ImageNet 測試集上實現了 3.57% 的誤差。該結果在 ILSVRC 2015 分類任務中獲得第一名。

7. We also present analysis on CIFAR-10 with 100 and 1000 layers. The depth of representations is of central importance for many visual recognition tasks.

我們還提供了 100 和 1000 層的 CIFAR-10 分析。表示的深度對於許多視覺識別任務而言至關重要。

8. Solely due to our extremely deep representations, we obtain a 28% relative improvement on the COCO object detection dataset.

僅僅由於我們極其深刻的表示，我們在 COCO 對象檢測數據集上獲得了 28% 的相對改進。

9. Deep residual nets are foundations of our submissions to ILSVRC & COCO 2015 competitions<sup>1</sup>, where we also won the 1st places on the tasks of ImageNet detection, ImageNet localization, COCO detection, and COCO segmentation.

深度殘留網是我們向 ILSVRC 和 COCO 2015 競賽 1 提交的基礎，我們還在 ImageNet 檢測，ImageNet 定位，COCO 檢測和 COCO 分割任務中獲得了第一名。

### 2.3.1 回答下列問題

1. What are motivations for this work? 該論文解決的問題

讓殘留網絡更容易優化，並且可以增加的深度網路獲得準確性。

2. What is the proposed solution? 該論文解決的方法

提出一個新的深度殘留的學習框架，培訓深度能比以往的深度網路更深。

重新組合參考層的函數，而不是一直新增新的函數。

3. What is the evaluation of the proposed solution? 該論文作者如何評估其結果? (問卷, 量測時間數據, 實驗...)

在 ImageNet 測試集上實作，而誤差只有 3.57%

4. What are the contributions? 該論文的貢獻為和?

為 ILSVRC(ImageNet Large Scale Visual Recognition Challenge) 和 COCO 競賽的基礎

ImageNet 大規模視覺識別挑戰 (ILSVRC) 評估大規模物體檢測和圖像分類的算法。一個高級別的動機是讓研究人員比較各種物體的檢測進度 - 利用相當昂貴的標籤工作。另一個動機是測量用於檢索和註釋的大規模圖像索引的計算機視覺的進展。

5. What are future directions for this research? 未來研究方向可以為何?

目前只是通過設計將 deep 層與 thin 層進行正則化，未來會改善其結果，結合更強的正規化方式。

## 2.4 Only Look Once, Mining Distinctive Landmarks from ConvNet for Visual Place Recognition. 只看一次，在 ConvNet 找到特殊的地標，用於地點識別

資料來源:2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) [?]

影片:[影片](#)

1. Recently, image representations derived from Convolutional Neural Networks (CNNs) have been demonstrated to achieve impressive performance on a wide variety of tasks, including place recognition.

最近，已經證明來自卷積神經網絡（CNN）的圖像表示在包括位置識別在內的各種任務上實現了令人印象深刻的性能。

2. In this paper, we take a step deeper into the internal structure of CNNs and propose novel CNN-based image features for place recognition by identifying salient regions and creating their regional representations directly from the convolutional layer activations.

在本文中，我們更深入地了解 CNN 的內部結構，並通過識別顯著區域並直接從卷積層激活創建區域表示，提出用於位置識別的新穎的基於 CNN 的圖像特徵。

3. A range of experiments is conducted on challenging datasets with varied conditions and viewpoints.

在具有不同條件和觀點的具有挑戰性的數據集上進行了一系列實驗。

4. These reveal superior precision-recall characteristics and robustness against both viewpoint and appearance variations for the proposed approach over the state of the art.

這些揭示了針對現有技術的所提出的方法的優異的精確回憶特性和針對視點和外觀變化的穩固性。

5. By analyzing the feature encoding process of our approach, we provide insights into what makes an image presentation robust against external variations.

通過分析我們方法的特徵編碼過程，我們可以深入了解圖像呈現對外部變化的穩固性。

### 2.4.1 回答下列問題

1. What are motivations for this work? 該論文解決的問題

實作卷積神經網路，改善演算法，應用在識別區域的特徵，並回饋一個相似的圖片。

2. What is the proposed solution? 該論文解決的方法

部分解析卷積神經網路的運作過程與比較，以視覺化的方式呈現。

3. What is the evaluation of the proposed solution? 該論文作者如何評估其結果? (問卷, 量測時間數據, 實驗...)

使用 Caffe 在 TitanX 的設備下輸入單張圖片，先經過 VGG16 網絡約為 59.4□□，在經過他們在 Matlab 提出的方法編譯約 349□□，最後進行兩圖像匹配約 7□□，可以在近乎一秒內找出一個相似的圖片。

4. What are the contributions? 該論文的貢獻為和?

1. 一種新穎的基於 CNN 的特徵編碼方法，用於創建能夠描述若干不同圖像區域的圖像表示，而無需向 CNN 饋送多個輸入。

2. 基於區域的視覺位置識別系統，可同時處理視點和條件的變化。

5. What are future directions for this research? 未來研究方向可以為何?

在本論文中，是使用 CNN 中預訓練過的對象進行辨識，因此未來希望改善 CNN 以提高性能。



## 2.5 Visualizing and Understanding Convolutional Networks

資料來源:Dept. of Computer Science,New York University, USA [?]

影片:[影片](#)

1. Large Convolutional Network models have recently demonstrated impressive classification performance on the ImageNet benchmark Krizhevsky et al. [18].

大型捲積網絡模型最近在 ImageNet 基準測試 Krizhevsky 等人上展示了令人印象深刻的分類性能。

2. However there is no clear understanding of why they perform so well, or how they might be improved.

然而，沒有明確的理解他們為什麼表現如此之好，或者如何改進它們。

3. In this paper we explore both issues.

在本文中，我們探討了這兩個問題。

4. We introduce a novel visualization technique that gives insight into the function of intermediate feature layers and the operation of the classifier.

我們介紹了一種新穎的可視化技術，可以深入了解中間特徵層的功能和分類器的操作。

5. Used in a diagnostic role, these visualizations allow us to find model architectures that outperform Krizhevsky et al.

這些可視化用於診斷角色，使我們能夠找到優於 Krizhevsky 等人的模型架構。

6. on the ImageNet classification benchmark.

在 ImageNet 分類基準上。

7. We also perform an ablation study to discover the performance contribution from different model layers.

我們還進行消融研究，以發現不同模型層的性能貢獻。

8. We show our ImageNet model generalizes well to other datasets: when the softmax classifier is retrained, it convincingly beats the current state-of-the-art results on Caltech-101 and Caltech-256 datasets.

我們展示了我們的 ImageNet 模型很好地概括了其他數據集：當 softmax 分類器被重新訓練時，它令人信服地擊敗了 Caltech-101 和 Caltech-256 數據集的當前最先進的結果。

### 2.5.1 回答下列問題

1. What are motivations for this work? 該論文解決的問題  
介紹卷積神經網路，並將中間的 layer 視覺化出來
2. What is the proposed solution? 該論文解決的方法  
實作並敘述卷積神經網路的原理，將其步驟以圖片展現出來。
3. What is the evaluation of the proposed solution? 該論文作者如何評估其結果? (問卷, 量測時間數據, 實驗...)  
比 Caltech-101 和 Caltech-256 數據集的分類結果更為精準。
4. What are the contributions? 該論文的貢獻為和?  
其目的是詳細介紹卷積神經網路的原理，讓人更容易了解。
5. What are future directions for this research? 未來研究方向可以為何?  
當時的卷積神經網路演算過程還較為簡易，可以挑戰其他卷積神經網路的應用。

## 2.6 Visualizing Convolutional Neural Networks for Image Classification

資料來源: Dept. of Computer Science, New York University, USA [?]

影片: [影片](#)

1. Deep convolutional neural networks have recently shown state of the art performance on image classification problems.

最近，深度卷積神經網絡顯示了圖像分類問題的最新性能。

2. However, their inner workings remain a mystery to machine learning experts, particularly when compared to better studied and less complex algorithms such as SVM and Logistic Regression.

然而，他們的工作主要是神秘的機械學習專家，特別是與更好的研究和不太複雜的算法（如 SVM 和 Logistic 回歸）相比。

3. As a result, constructing and debugging effective convolutional neural networks is time-consuming and error-prone, as it often involves a substantial amount of trial and error.

因此，構建和調試有效的捲積神經網絡既費時又容易出錯，因為它經常涉及大量的反復試驗。

4. We introduce deepViz, a system designed to allow experts to understand their models and diagnose issues with the model structure, enabling more rapid iteration during the model construction process and faster convergence to a suitable model for the task at hand.

我們介紹了 deepViz，這是一個系統，旨在讓專家了解他們的模型並診斷模型結構的問題，在模型構建過程中實現更快速的迭代，並且更快收斂到手頭任務的合適模型。

### 2.6.1 回答下列問題

1. What are motivations for this work? 該論文解決的問題

解析卷積神經網路的內容，在以視覺化呈現

2. What is the proposed solution? 該論文解決的方法

該論文開發了 Deep Viz 可以將卷積神經網路解析並呈現

3. What is the evaluation of the proposed solution? 該論文作者如何評估其結果? (問卷, 量測時間數據, 實驗...)

能夠讓專家快速的了解卷積神經網路的架構，以便快速分析

4. What are the contributions? 該論文的貢獻為和?

提出了一個可視化工具 deepViz，旨在幫助專家理解和診斷卷積神經網路視覺分類問題。

5. What are future directions for this research? 未來研究方向可以為何?

未來可改善其可視化的技術，以更好的方式呈現

## 2.7 Visualization of Neural Network Predictions for Weather Forecasting

資料來源:COMPUTER GRAPHICS forumVolume 00 (2018), number 0 pp. 1-12 [?]

影片:[影片](#)

1. Recurrent neural networks are prime candidates for learning evolutions in multi-dimensional time series data.

遞迴神經網路是用於學習多維時間序列數據中的演化的主要者。

2. The performance of such a network is judged by the loss function, which is aggregated into a scalar value that decreases during training.

這種網絡的性能由損失函數判斷，該函數被聚合成在訓練期間減小的標量值。

3. Observing only this number hides the variation that occurs within the typically large training and testing data sets.

僅觀察此數字會隱藏通常較大的訓練和測試數據集中發生的變化。

4. Understanding these variations is of highest importance to adjust network hyper-parameters, such as the number of neurons, number of layers or to adjust the training set to include

more representative examples.

理解這些變化對於調整網絡超參數是最重要的，例如神經元的數量，層數或調整訓練集以包括更具代表性的示例。

5. In this paper, we design a comprehensive and interactive system that allows users to study the output of recurrent neural networks on both the complete training data and testing data.

在本文中，我們設計了一個全面的交互式系統，允許用戶在完整的訓練數據和測試數據上研究遞歸神經網絡的輸出。

6. We follow a coarse-to-fine strategy, providing overviews of annual, monthly and daily patterns in the time series and directly support a comparison of different hyper-parameter settings.

我們遵循粗略到精細的策略，提供時間序列中年度，月度和日常模式的概述，並直接支持不同超參數設置的比較。



7. We applied our method to a recurrent convolutional neural network that was trained and tested on 25 years of climate data to forecast meteorological attributes, such as temperature, pressure and wind velocity.

我們將我們的方法應用於循環卷積神經網絡，該網絡經過 25 年的氣候數據訓練和測試，以預測氣象屬性，如溫度，壓力和風速。

8. We further visualize the quality of the forecasting models, when applied to various locations on the Earth and we examine the combination of several forecasting models.

我們進一步可視化預測模型的質量，當應用於地球上的各個位置時，我們檢查了幾種預測模型的組合。

### 2.7.1 回答下列問題

1. What are motivations for this work? 該論文解決的問題

RNN 的常用方法是使用一系列先前時間步驟來預測下一步驟。會導致無法一次顯示的非常大的數據集。

2. What is the proposed solution? 該論文解決的方法

設計了一個全面的交互式系統，允許用戶在完整的訓練數據和測試數據上研究遞歸神經網絡的輸出。

3. What is the evaluation of the proposed solution? 該論文作者如何評估其結果? (問卷, 量測時間數據, 實驗...)

使用 1990 年至 1999 年的數據來訓練 1990 年至 2016 年的模型和數據。訓練時間為 6 小時。該分辨率在我們的初始測試中表現最佳。

4. What are the contributions? 該論文的貢獻為何?

評估誤差在 2016 年的溫度預測中取平均值，並在最大誤差為華氏 10 度。

5. What are future directions for this research? 未來研究方向可以為何?

改善現在的神經網路系統，加上殘差神經網路，以用來訓練及預測其他屬性。

## 2.8 Deep learning for computational biology

資料來源:COMPUTER GRAPHICS forumVolume 00 (2018), number 0 pp. 1-12 [?]

影片:[影片](#)

1. Technological advances in genomics and imaging have led to an explosion of molecular and cellular profiling data from large numbers of samples.

基因學和視覺化技術的進步導致大量分子和細胞樣本的分析數據以爆炸式的增長。

2. This rapid increase in biological data dimension and acquisition rate is challenging conventional analysis strategies.

生物數據維度和採集速率的這種快速增長正在挑戰傳統的分析策略。

3. Modern machine learning methods, such as deep learning, promise to leverage very large data sets for finding hidden structure within them, and for making accurate predictions.

現代機器學習方法，例如深度學習，有望利用大數據來尋找其中的隱藏結構，並進行準確的預測。

4. In this review, we discuss applications of this new breed of analysis approaches in regulatory genomics and cellular imaging.

在本論文中，我們討論了這種新型分析方法在監管基因學和細胞成像中的應用。

5. We provide background of what deep learning is, and the settings in which it can be successfully applied to derive biological insights.

我們提供深度學習的背景，以及可以成功應用以獲得生物學見解的設置。

6. In addition to presenting specific applications and providing tips for practical use, we also highlight possible pitfalls and limitations to guide computational biologists when and how to make the most use of this new technology.

除了提供具體應用並提供實際使用技巧外，我們還強調了可能存在的缺陷和局限性，以指導計算生物學家何時以及如何充分利用這項新技術。

### 2.8.1 回答下列問題

1. What are motivations for this work? 該論文解決的問題

在基因學及細胞成像成熟的現今，能夠快速辨識與預測是其中最重要的事項。

2. What is the proposed solution? 該論文解決的方法

利用機器學習、監督式學習以及卷積神經網路作為物件偵測與分類。

3. What is the evaluation of the proposed solution? 該論文作者如何評估其結果? (問卷, 量測時間數據, 實驗…)

在更深層數階層中，更能找到偵測物體的特徵，會更容易進行分類。

4. What are the contributions? 該論文的貢獻為和?

讓計算生物學家尋找基因學中的隱藏結構，並進行準確的預測。

5. What are future directions for this research? 未來研究方向可以為何?

改善參數與優化。使軟體能夠使用深度學習在更多的生物學問題上。





## 第三章 方法

### 3.1 解析 YOLO 黑盒子

YOLOv2 演算法中，有一個黑盒子 darknet-19，本論文著重於解析並呈現其內容與過程。

#### 3.1.1 Darknet-19 表格

##### YOLO 運作過程

YOLO 輸入圖片與結果



圖 3.1: 原圖

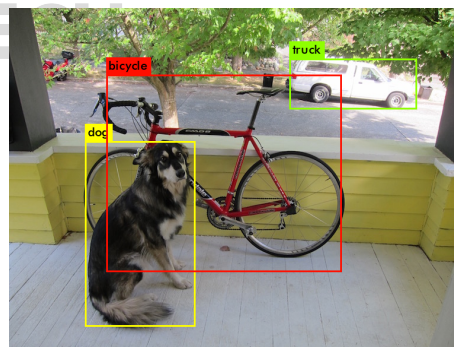


圖 3.2: 偵測結果

Layer	Type	Filters	Size/Stride	Output size
0	Convolutional	32	3x3/1	416x416
1	Maxpool		2x2/2	208x208
2	Convolutional	64	3x3/1	208x208
3	Maxpool		2x2/2	104x104
4	Convolutional	128	3x3/1	104x104
5	Convolutional	64	3x3/1	104x104
6	Convolutional	128	3x3/1	104x104
7	Maxpool		2x2/2	52x52
8	Convolutional	256	3x3/1	52x52
9	Convolutional	128	1x1/1	52x52
10	Convolutional	256	3x3/1	52x52
11	Maxpool		2x2/2	26x26
12	Convolutional	512	3x3/1	26x26
13	Convolutional	256	3x3/1	26x26
14	Convolutional	512	3x3/1	26x26
15	Convolutional	256	1x1/1	26x26
16	Convolutional	512	3x3/1	26x26
17	Maxpool		2x2/2	13x13
18	Convolutional	1024	3x3/1	13x13
19	Convolutional	512	1x1/1	13x13
20	Convolutional	1024	3x3/1	13x13
21	Convolutional	512	1x1/1	13x13
22	Convolutional	1024	3x3/1	13x13
23	Convolutional	1024	3x3/1	13x13
24	Convolutional	1024	3x3/1	13x13
25	Route	16		
26	Convolutional	64	1x1/1	13x13
27	Reorg		/2	13x13
28	Route	27 24		
29	Convolutional	1024	3x3/1	13x13
30	Convolutional	425	1x1/1	13x13
31	Detection			

## 第四章 結果與討論

### 4.1 第一階層子標題

各階層子標題均應置於左側，並於其下方不空行。

#### 4.1.1 第二階層子標題

第二階層子標題之內文。

#### 第三階層子標題

第三階層子標題之內文。



## 第五章 結論

### 5.1 結論

各階層子標題均應置於左側，並於其下方不空行。

### 5.2 未來展望

第二階層子標題之內文。



## 附錄 A：第一個附錄名稱

附錄內容



## 附錄 B：第二個附錄名稱

附錄內容



## 符號彙編

### Symbol Meaning

$\Theta$  Debye's constant or characteristic temperature

$\Omega$  efficiency; number of molecules

$\Psi$  availability of a closed system

$\Delta$  internal energy (change) of reaction

$\Phi$  availability of a closed system

$\iota$  specific irreversibility

$\lambda$  critical state

$\mu$  Joule-Thomson coefficient

$\nu$  stoichiometric coefficient (number of moles in chemical equation)

$\xi$  cutoff ratio