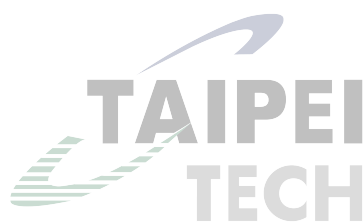


國立臺北科技大學  
資訊工程研究所

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

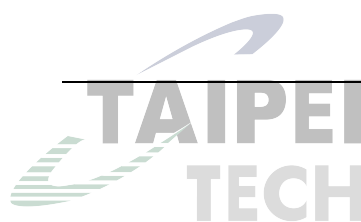
本校\_\_\_\_\_研究所\_\_\_\_\_君

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

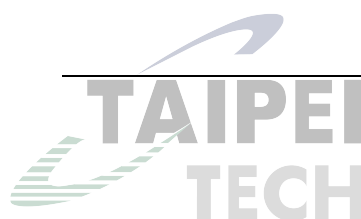
本校\_\_\_\_\_研究所\_\_\_\_\_君

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               | vii  |
| 表 目 錄               | viii |
| 1 導論                | 1    |
| 1.1 導論              | 1    |
| 2 相關文獻討論            | 4    |
| 3 方法                | 6    |
| 3.1 解析 YOLO 黑盒子     | 6    |
| 3.1.1 Darknet-19 表格 | 6    |
| 4 結果與討論             | 8    |
| 4.1 第一階層子標題         | 8    |
| 4.1.1 第二階層子標題       | 8    |
| 5 結論                | 9    |
| 5.1 結論              | 9    |
| 5.2 未來展望            | 9    |
| 附錄 A：第一個附錄名稱        | 10   |
| 附錄 B：第二個附錄名稱        | 11   |
| 符號彙編                | 12   |



## 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 . . . . .               | 6 |
| 3.2 | 偵測結果 . . . . .             | 6 |



# 表 目 錄



# 第一章 導論

## 1.1 導論

You Only Look Once 的架構與一般的類神經網路的架構不同，有特殊的架構，讓人十分難懂，解析物件偵測 (You Only Look Once) 並將其以視覺化的方式呈現，讓不了解運作的人可以直覺性的知道其運作的過程為本論文主要的目的。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。

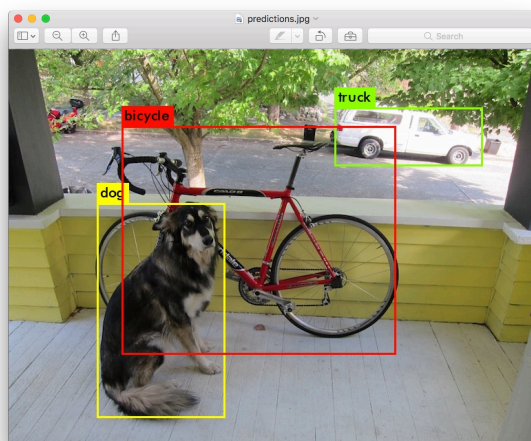


圖 1.1: Yolo Predivtions.

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。

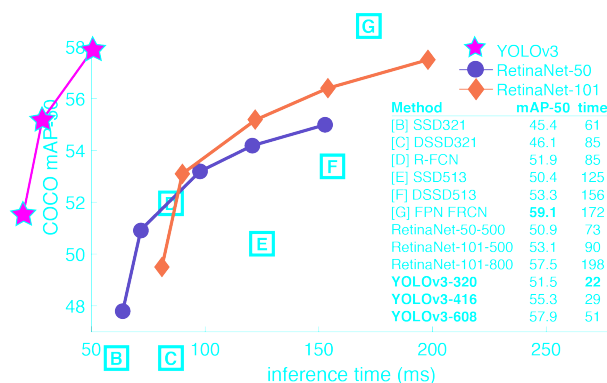


圖 1.2: Yolo v3 與其他演算法示意圖。

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。

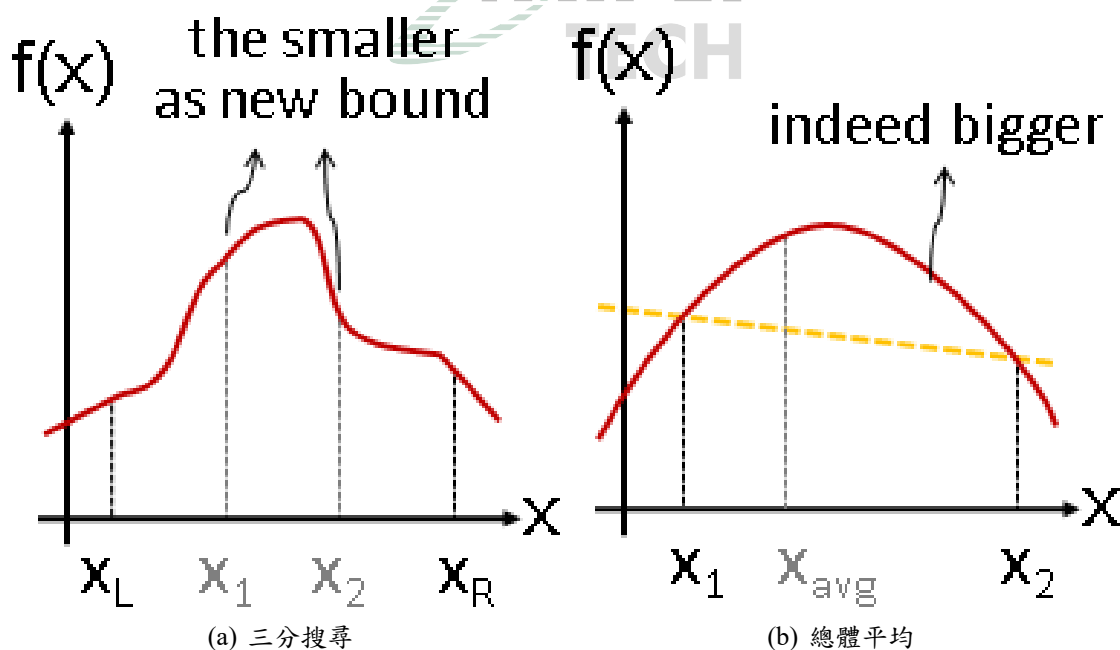


圖 1.3: 試誤法

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。

| 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.

本論文將探討 YOLO 與其他物件偵測的演算法所花費的時間、偵測的準確率與其中的差異，將其結構拆解出來，一方面讓使用者了解其構造與運作模式，另一方面也讓使用者了解 YOLO 快速的原因，並探討參數的內容應該有如何調整才會有更好的結果。



## 第二章 相關文獻討論

視覺化演算法在 CNN 上已經有許多相關的研究，一開始的研究著重於將圖片的特徵，但隨著 CNN 的快速發展，視覺化已經擴展到解釋 CNN 的整體架構與運作方式。主要是解析每個演算法的網路架構和演算法的邏輯，其中有幾個較具代表性的方法：

Erhan 等人提出 Activation Maximization 來對傳統的淺層網路進行解釋。

後來，Simonyan 等人通過將單個 CNN 神經元的最大啟用視覺化合成一個輸入影像模式 (input image pattern)，進一步改進了這種方法。

後續出現了很多工作都是基於這種方法，再利用不同的正則項進行擴充套件，以提高合成影像模式的可解釋性。

Mahendran 等人提出了 Network Inversion 重建基於多個神經元啟用的輸入影像，以此說明每個 CNN 層學習到的綜合特徵圖，揭示了 CNN 網路在網路層層面的內部特徵。

Network Inversion 根據特定層的特徵圖中的原始影像重建輸入影像，這可以揭示該圖層所儲存的影像資訊。

沒有選擇對輸入影像進行重建以實現特徵視覺化，Zeiler 等人提出了基於反摺積神經網路的視覺化方法 (Deconvolutional Neural Network based Visualization, DeconvNet)，該方法利用 DeconvNet 框架將特徵圖直接對映到影像維度，利用反摺積 CNN 結構 (由反摺積層和反摺積層組成) 在特定神經元啟用的原始輸入影像中查詢影像模式。

通過直接對映，DeconvNet 可以突出顯示輸入影像中的哪些模式啟用特定神經元，從而直接連結神經元和輸入資料的含義。

周博磊等人提出了 Network Dissection based Visualization，它從語義層面對 CNN 進行了解釋。

通過引用異構影像資料集——Borden，Network Dissection 可以有效地將輸入影像分割為多個具有各種語義定義的部分，可以匹配六種語義概念 (例如場景，目標，部件，材質，紋理和顏色)。

由於語義直接代表了特徵的含義，神經元的可解釋性可以顯著提高。

以上都是圍繞在以 CNN 為基礎的視覺化，而鮮少有對其他較進階的 CNN 演算法進行視覺化分析。

CNN、R-CNN、Fast R-CNN、Faster R-CNN、Mask R-CNN、SSD、YOLO、YOLOv2、YOLOv3……等，都是屬於使用 CNN 模型，只要輸入一張圖片，並得到該圖片分類的結果，但每個類型所使用的架構並不相同。



## 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

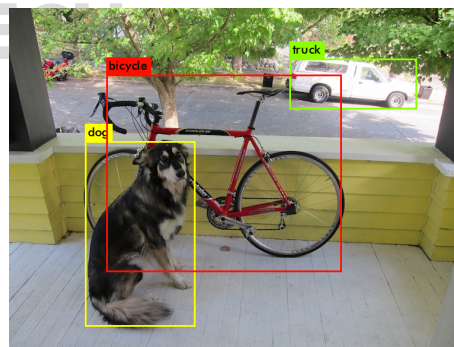


圖 3.2: 偵測結果



| 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