傳統計算機視覺技術落伍了嗎？不，它們是深度學習的「新動能」

OpenCV與AI深度學習 2022-09-12 09:02 發表於重慶

收錄於合集
#深度學習 114 #計算機視覺 143

> 深度學習崛起後，傳統計算機視覺方法被淘汰了嗎？

Deep Learning vs. Traditional Computer Vision

Niall O' Mahony, Sean Campbell, Anderson Carvalho, Suman Harapanahalli,
Gustavo Velasco Hernandez, Lenka Krpalkova, Daniel Riordan, Joseph Walsh

IMaR Technology Gateway, Institute of Technology Tralee, Tralee, Ireland
niall.omahony@research.ittralee.ie

- 論文鏈接：https://arxiv.org/ftp/arxiv/papers/1910/1910.13796.pdf

深度學習擴展了數字圖像處理的邊界。然而，這並不代表在深度學習崛起之前不斷發展進步的傳統計算機視覺技術被淘汰。近期，來自愛爾蘭垂利理工學院的研究者發表論文，分析了這兩種方法的優缺點。

該論文旨在促進人們對是否保留經典計算機視覺技術知識進行討論。此外，這篇論文還探討了**如何結合傳統計算機視覺與深度學習**。文中提及了多個近期混合方法，這些方法既提升了計算機視覺性能，又解決了不適合深度學習的問題。例如，將傳統計算機視覺技術與深度學習結合已經在很多新興領域流行起來，如深度學習模型尚未得到充分優化的全視野、3D 視覺領域。

深度學習VS 傳統計算機視覺

深度學習的優勢

深度學習的快速發展和設備能力的改善（如算力、內存容量、能耗、圖像傳感器分辨率和光學器件）提升了視覺應用的性能和成本效益，並進一步加快了此類應用的擴展。與傳統CV 技術相比，深度學習可以幫助CV 工程師在圖像分類、語義分割、目標檢測和同步定位與地圖構建（SLAM）等任務上獲得更高的準確率。由於深度學習所用的神經網絡是訓練得到而非編程得到，因此使用該方法的應用所需的專家分析和微調較少，且能夠處理目前系統中的海量可用視頻數據。深度學習還具備絕佳的靈活性，因為對於任意用例，CNN 模型和框架均可使用自定義數據集重新訓練，這與CV 算法不同，後者俱備更強的領域特定性。

以**移動機器人的目標檢測問題**為例，對比這兩類計算機視覺算法：

傳統計算機視覺方法使用成熟的CV 技術處理目標檢測問題，如特徵描述子（SIFT、SUR、BRIEF 等）。在深度學習興起前，圖像分類等任務需要用到特徵提取步驟，特徵即圖像中「有趣」、描述性或信息性的小圖像塊。這一步可能涉及多種CV 算法，如邊緣檢測、角點檢測或閾值分割算法。從圖像中提取出足夠多的特徵後，這些特徵可形成每個目標類別的定義（即「詞袋」）。部署階段中，在其他圖像中搜索這些定義。如果在一張圖像中找到了另一張圖像詞袋中的絕大多數特徵，則該圖像也包含同樣的目標（如椅子、馬等）。

傳統CV 方法的缺陷是：從每張圖像中選擇重要特徵是必要步驟。而隨著類別數量的增加，特徵提取變得越來越麻煩。要確定哪些特徵最能描述不同的目標類別，取決於CV 工程師的判斷和長期試錯。此外，每個特徵定義還需要處理大量參數，所有參數必須由CV 工程師進行調整。

深度學習引入了端到端學習的概念，即向機器提供的圖像數據集中的每張圖像均已標註目標類別。因而深度學習模型基於給定數據「訓練」得到，其中神經網絡發現圖像類別中的底層模式，並自動提取出對於目標類別最具描述性和最顯著的特徵。人們普遍認為DNN 的性能大大超過傳統算法，雖然前者在計算要求和訓練時間方面有所取捨。隨著CV 領域中最優秀的方法紛紛使用深度學習，CV 工程師的工作流程出現巨大改變，手動提取特徵所需的知識和專業技能被使用深度學習架構進行迭代所需的知識和專業技能取代（見圖1）。

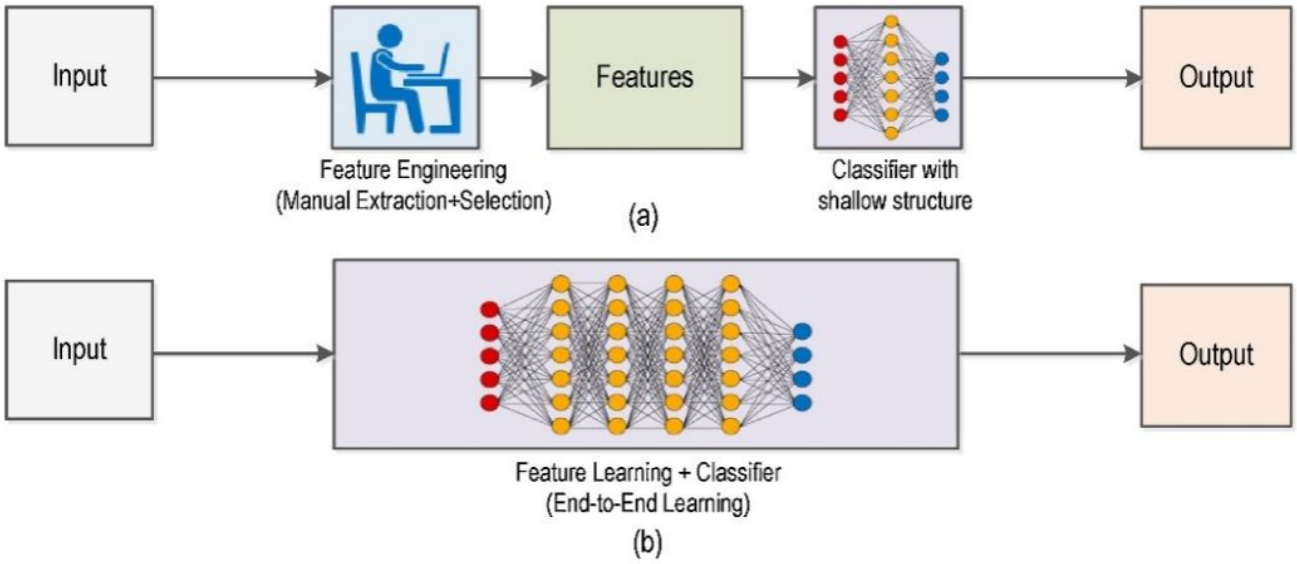


圖1：a）傳統計算機視覺工作流vs b）深度學習工作流。（圖源：[8]）

近年來，**CNN 的發展對CV 領域產生了巨大影響，也使得目標識別能力出現大幅提升**。這種爆發與算力的提升、訓練數據量的增加密不可分。近期CV 領域中深度神經網絡架構出現井噴並得到廣泛應用，這從論文《ImageNet Classification with Deep Convolutional Neural Networks》引用量超3000 次中可見一斑。

CNN 利用卷積核（又稱濾波器）來檢測圖像中的特徵（如邊）。卷積核是權重矩陣，這些權重被訓練用於檢測特定特徵。如名字所示，CNN 的主要思想是在給定輸入圖像上空間性地捲積內核，檢查是否出現檢測所需特徵。為了用數值表示出現某個特徵的置信度，神經網絡執行卷積操作，即計算卷積核與它和輸入圖像重疊區域的點積（卷積核正在查看的原始圖像區域叫做感受野）。

為了促進卷積核權重的學習，研究人員向卷積層的輸出添加偏置項，並饋入非線性激活函數中。激活函數通常是非線性函數，如Sigmoid、TanH 和ReLU。激活函數的選擇取決於數據和分類任務的性質。例如，ReLU 具備更多生物表徵（大腦中的神經元是否處於激活狀態）。因此，在圖像識別任務中，ReLU 會得到更好的結果，因為它對梯度消失問題具備更強的抵抗力，而且它能夠輸出更稀疏、高效的表徵。

為了加速訓練過程，減少網絡消耗的內存量，卷積層後通常跟著一個池化層，用於移除輸入特徵中的冗餘部分。例如，最大池化在輸入上移動窗口，僅輸出窗口中的最大值，從而高效減少圖像中的冗餘部分，留下重要像素。如圖2 所示，深度CNN 可能具備多對卷積和池化層。最後，全連接層將上一層壓縮為特徵向量，然後輸出層利用密集網絡計算輸出類別/特徵的分數（置信度或概率）。將該輸出輸入到回歸函數中，如Softmax 函數，它將所有事物映射為向量且其中所有元素的總和為1。

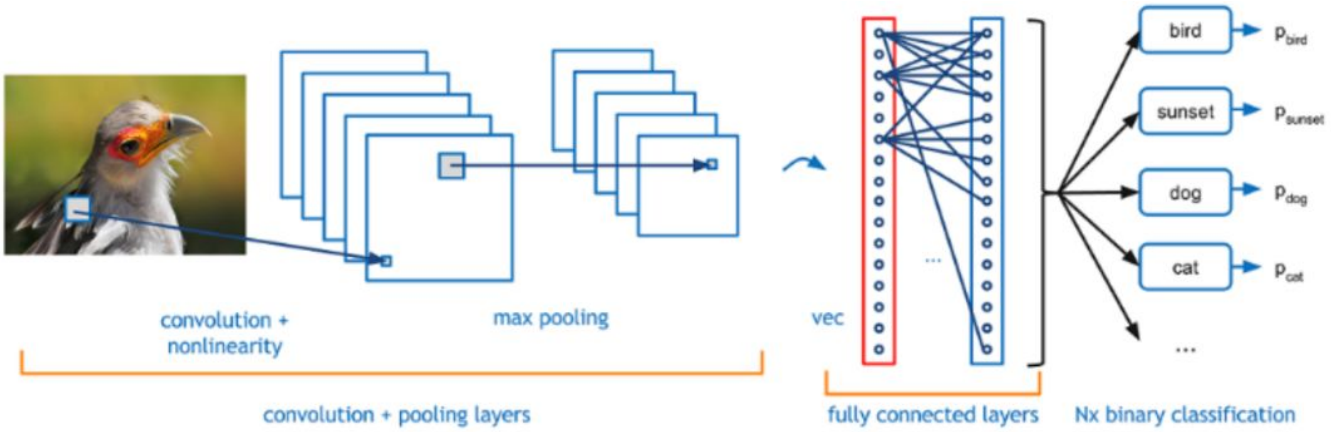


圖2：CNN 構造塊。（圖源：[13]）

但是深度學習仍然只是CV 領域的工具。例如，CV 領域中最常用的神經網絡是CNN。那麼什麼是卷積呢？卷積廣泛應用於圖像處理技術。（深度學習的優點很明確，本文暫不討論當前最優算法。）但深度學習並非解決所有問題的萬靈藥，下文將介紹傳統CV 算法更適合的問題及應用。

傳統CV 技術的優勢

這部分將詳細介紹**基於特徵的傳統方法在CV 任務中能夠有效提升性能的原因**。這些傳統方法包括：

- 尺度不變特徵變換（Scale Invariant Feature Transform，SIFT）[14]
- 加速穩健特徵（Speeded Up Robust Feature，SURF）[15]
- 基於加速分割測試的特徵（Features from Accelerated Segment Test，FAST）[16]
- 霍夫變換（Hough transform）[17]
- 幾何哈希（Geometric hashing）[18]

特徵描述子（如SIFT 和SURF）通常與傳統機器學習分類算法（如支持向量機和K 最近鄰算法）結合使用，來解決CV 問題。

深度學習有時會「過猶不及」，傳統CV 技術通常能夠更高效地解決問題，所用的代碼行數也比深度學習少。SIFT，甚至簡單的色彩閾值和像素計數等算法，都不是特定於某個類別的，它們是通用算法，可對任意圖像執行同樣的操作。與之相反，深度神經網絡學得的特徵是特定於訓練數據的。也就是說，如果訓練數據集的構建出現問題，則網絡對訓練數據集以外的圖像處理效果不好。

因此，**SIFT 等算法通常用於圖像拼接/3D 網格重建等應用，這些應用不需要特定類別知識**。這些任務也可以通過訓練大型數據集來實現，但是這需要巨大的研究努力，為一個封閉應用費這麼大勁並不實際。在面對一個CV 應用時，工程師需要培養選擇哪種解決方案的常識。例如，對流水線傳送

帶上的兩類產品進行分類，一類是紅色一類是藍色。深度神經網絡需要首先收集充足的訓練數據。然而，使用簡單的色彩閾值方法也能達到同樣的效果。一些問題可以使用更簡單、快速的技術來解決。

如果DNN 對訓練數據以外的數據效果不好，怎麼辦？在訓練數據集有限的情況下，神經網絡可能出現過擬合，無法進行有效泛化。手動調參是非常困難的事情，因為DNN 擁有數百萬參數，且它們之間的關係錯綜複雜。也因此，深度學習模型被批評為黑箱。傳統的CV 技術具備充分的透明性，人們可以判斷解決方案能否在訓練環境外有效運轉。CV 工程師了解其算法可以遷移至的問題，這樣一旦什麼地方出錯，他們可以執行調參，使算法能夠有效處理大量圖像。

現在，**傳統CV 技術常用於解決簡單問題**，這樣它們可在低成本微處理器上部署，或者通過突出數據中的特定特徵、增強數據或者輔助數據集標註，來限定深度學習技術能解決的問題。本文稍後將討論，在神經網絡訓練中可使用多少種圖像變換技術。最後，CV 領域存在很多更具挑戰性的難題，比如機器人學、增強現實、自動全景拼接、虛擬現實、3D 建模、運動估計、視頻穩定、運動捕捉、視頻處理和場景理解，這些問題無法通過深度學習輕鬆實現，但它可以從傳統CV 技術中受益。

傳統CV 技術與深度學習的融合

傳統CV+深度學習=更好的性能

傳統CV 技術和深度學習方法之間存在明確的權衡。經典CV 算法成熟、透明，且為性能和能效進行過優化；深度學習提供更好的準確率和通用性，但消耗的計算資源也更大。

混合方法結合傳統CV 技術和深度學習，兼具這兩種方法的優點。它們尤其適用於需要快速實現的高性能係統。

機器學習度量和深度網絡的混合已經非常流行，因為這可以生成更好的模型。混合視覺處理實現能夠帶來性能優勢，且將乘積累加運算減少到深度學習方法的130-1000 分之一，幀率相比深度學習方法有10 倍提升。此外，混合方法使用的內存帶寬僅為深度學習方法的一半，消耗的CPU 資源也少得多。

充分利用邊緣計算

當算法和神經網絡推斷要在邊緣設備上運行時，其延遲、成本、雲存儲和處理要求比基於雲的實現低。邊緣計算可以避免網絡傳輸敏感或可確認數據，因此具備更強的隱私性和安全性。

結合了傳統CV 和深度學習的混合方法充分利用邊緣設備上可獲取的異質計算能力。異質計算架構包含CPU、微控制器協同處理器、數字信號處理器（DSP）、現場可編程邏輯門陣列（FPGA）和AI 加速設備，通過將不同工作負載分配給最高效的計算引擎來降低能耗。測試實現證明，在DSP 和CPU 上分別執行深度學習推斷時，前者的目標檢測延遲是後者的十分之一。

多種混合方法證明了其在邊緣應用上的優勢。使用混合方法能夠高效地整合來自邊緣節點傳感器的數據。

不適合深度學習的問題

CV 領域中存在一些難題，如機器人學、增強現實、自動全景拼接、虛擬現實、3D 建模、運動估計、視頻穩定、運動捕捉、視頻處理和場景理解，它們很難通過深度學習以可微方式輕鬆實現，而是需要使用其他「傳統」技術。

下文介紹了CV 領域中的一些新興問題，在這些問題中深度學習面臨新挑戰，而經典CV 技術能夠發揮更大作用。

3D 視覺

3D 輸入的內存大小比傳統的RGB 圖像大得多，卷積核必須在三維輸入空間中執行卷積（見圖3）。

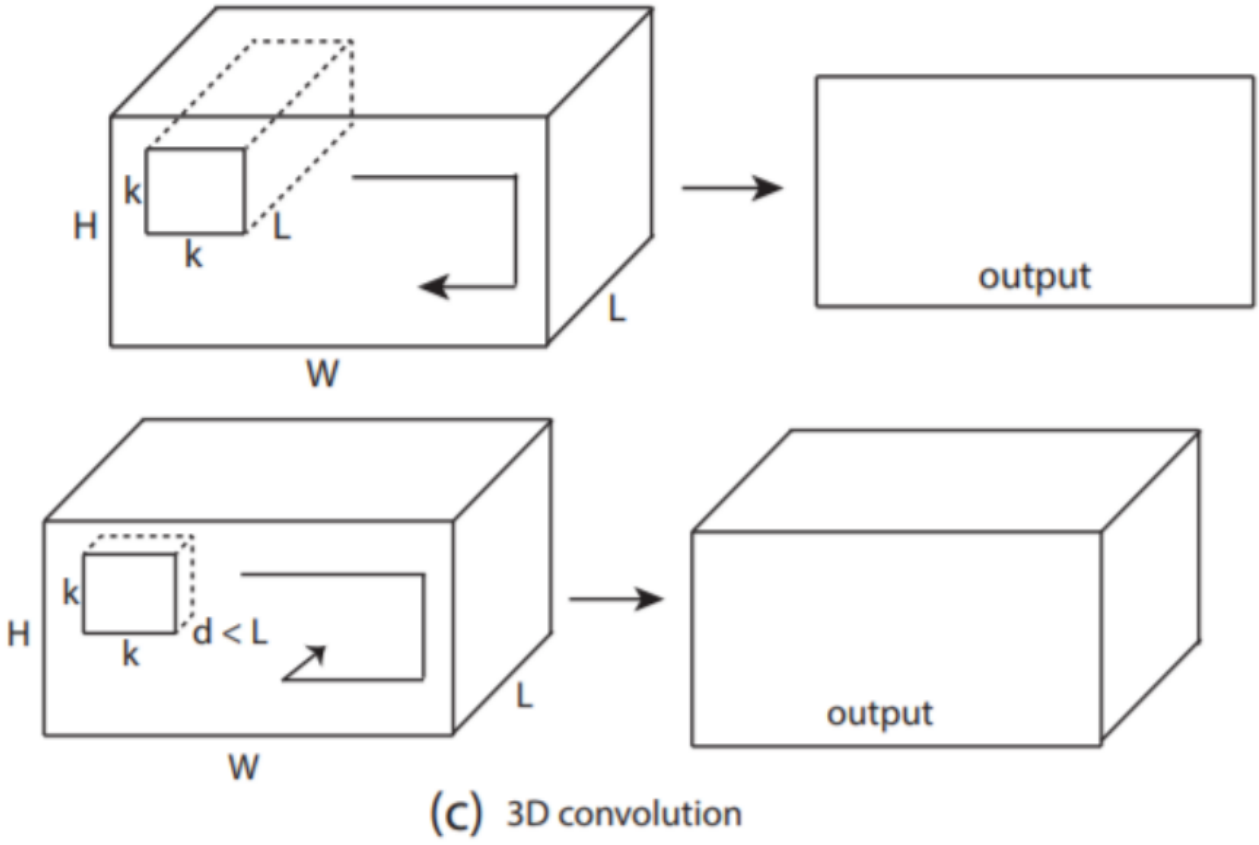
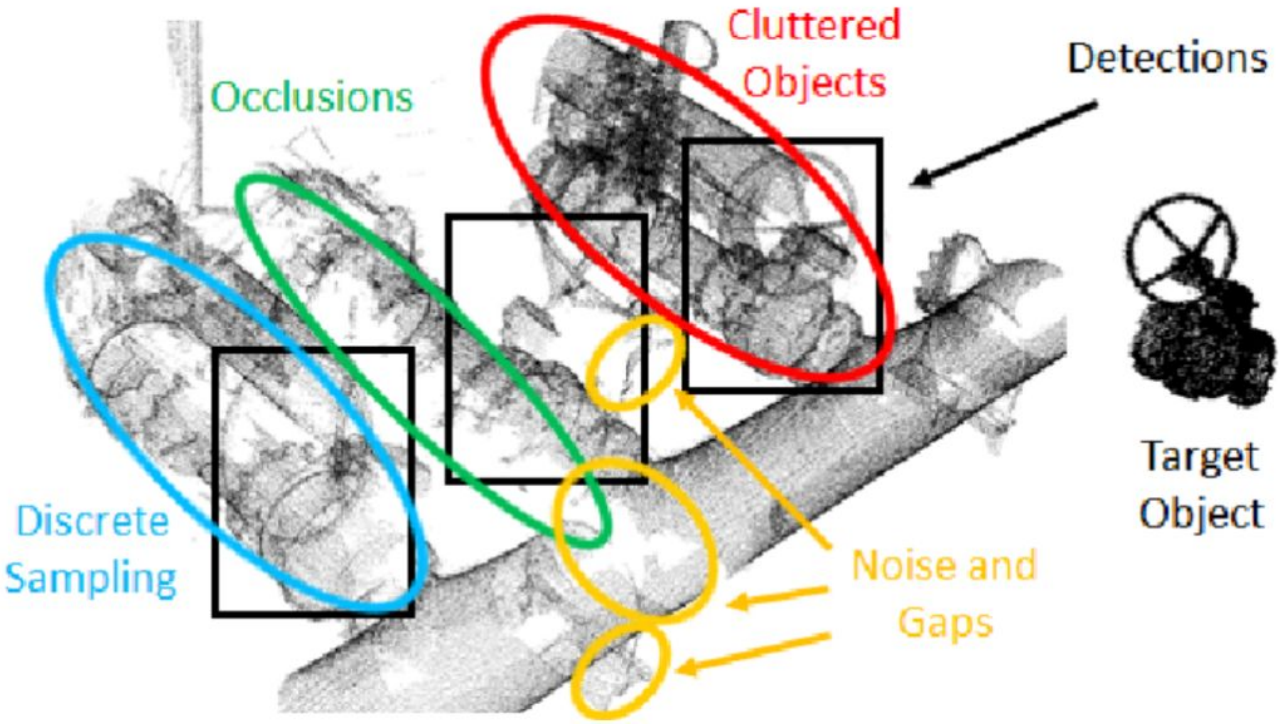


圖3：2D CNN vs. 3D CNN [47]

因此，3D CNN 的計算複雜度隨著分辨率呈現三次方增長。相比於2D 圖像處理，3D CV 更難，因為增加的維度使得不確定性也隨之增加，如遮擋和不同的攝像頭角度（見圖4）。



下一節將涉及處理多種3D 數據表徵的解決方案，這些方法具備新架構和預處理步驟，專用於解決上述挑戰。

幾何深度學習（GDL）將深度學習技術擴展到3D 數據。3D 數據的表徵方式多種多樣，總體上可分為歐幾里得和非歐幾里得。3D 歐幾里得結構化數據具備底層網格結構，允許全局參數化，此外，它還具備和2D 圖像相同的坐標系統。這使得現有的2D 深度學習範式和2D CNN 可應用於3D 數據。3D 歐幾里得數據更適合通過基於體素的方法分析簡單的剛性物體，如椅子、飛機等。另一方面，3D 非歐幾里得數據不具備網格數組結構，即不允許全局參數化。因此，將經典深度學習技術擴展到此類表徵是非常難的任務，近期[52] 提出的Pointnet 解決了這個難題。

對目標識別有用的連續形狀信息常常在轉換為體素表徵的過程中丟失。使用傳統CV 算法，[53] 提出可應用於體素CNN（voxel CNN）的一維特徵。這種基於平均曲率的新型旋轉不變特徵提升了體素CNN 的形狀識別性能。該方法應用到當前最優的體素CNN Octnet 架構時取得了極大成功，它在ModelNet10 數據集上取得了1% 的整體準確率提升。

SLAM

視覺SLAM 是SLAM 的子集，它使用視覺系統（而非激光雷達）登記場景中的路標。視覺SLAM 具備攝影測量的優勢（豐富的視覺數據、低成本、輕量級和低能耗），且沒有後處理通常需要的繁重計算工作負載。視覺SLAM 包含環境感知、數據匹配、運動估計、位置更新和新路標登記等步驟。

對在不同條件（如3D 旋轉、縮放、光照）中出現的視覺對象建模，以及使用強大的遷移學習技術擴展表徵以實現zero/one shot learning，是一道難題。特徵提取和數據表徵方法可以有效地減少機器學習模型所需的訓練樣本數量。

圖像定位中常使用一種兩步方法：位置識別+姿勢估計。前者使用詞袋方法，通過累積局部圖像描述子（如SIFT）來計算每個圖像的全局描述子。每個全局描述子均被存儲在數據庫中，一同存儲的還有生成3D 點雲基準圖的攝像頭姿勢。從query 圖像中提取出類似的全局描述子，數據庫中最接近的全局描述子可以通過高效搜索檢索出來。最接近全局描述子的攝像頭姿勢可以幫助我們對query 圖像進行粗略定位。在姿勢估計中，使用Perspective-n-Point (PnP) [13] 和幾何驗證等算法更準確地計算query 圖像的確切姿勢。

基於圖像的位置識別的成功很大程度上歸功於提取圖像特徵描述子的能力。不幸的是，在對激光雷達掃描圖像執行局部特徵提取時，沒有性能堪比SIFT 的算法。3D 場景由3D 點和數據庫圖像構成。一種方法是將每個3D 點與一組SIFT 描述子結合起來，描述子對應該點被三角化的圖像特徵。然後將這些描述子平均為一個SIFT 描述子，來描述該點的外觀。

另一種方法基於RGB-D 數據構建多模態特徵，而不是深度處理。至於深度處理部分，研究者採用基於表面法線的著色方法，因為它對多種任務有效且具備穩健性。另一種使用傳統CV 技術的替代方法提出基於圖的層級描述子Force Histogram Decomposition (FHD)，它可以定義對象的成對結構化子部分之間的空間關係和形狀信息。該學習步驟的優勢是與傳統詞袋框架兼容，從而出現結合了結構特徵和局部特徵的混合表徵。

360 度攝像頭

由於球面攝像頭的成像特點，每張圖像都能夠捕捉到360 度全景場景，消除了對轉向選擇的限制。球面圖像面臨的一個主要挑戰是超廣角魚眼鏡頭導致的嚴重桶形畸變，這增加了受傳統人類視覺啟發的車道檢測和軌跡追踪等方法的實現複雜度。這通常需要額外的預處理步驟，如先驗校準（prior calibration）和deworming。[60] 提出的一種替代方法將導航看作分類問題，從而繞過了預處理步驟，該方法基於原始未校準球面圖像找出最優潛在路徑方向。

全景拼接是該領域的另一個開放性問題。實時拼接方法[61] 使用一組可變形網格和最終圖像，並結合利用穩健像素著色器的輸入。另一種方法[62] 將幾何推理（線和消失點）提供的準確率和深度學習技術（邊和法線圖）實現的更高級數據提取和模式識別結合起來，為室內場景提取結構化數據，並生成佈局假設。在稀疏結構化場景中，由於缺乏明顯的圖像特徵，基於特徵的圖像配準方法通常會失敗。這時可使用直接的圖像配準方法，如基於相位相關的圖像配準算法。[23] 研究了基於判別相關濾波器（DCF）的圖像配準技術，證明基於DCF 的方法優於基於相位相關的方法。

數據集標註和增強

對於CV 和深度學習的結合存在一些反駁意見，總結為一句話就是：我們需要重新評估方法，不管是基於規則的方法還是數據驅動方法。從信號處理的傳統角度來看，我們了解傳統CV 算法（如SIFT 和SURF）的運算內涵，而深度學習無法展示這些意義，你所需要的只是更多數據。這可以被視為巨大的前進，但也有可能是後退。本論文提到了該爭論的正反方觀點，但是如果未來的方法僅基於數據驅動，那麼研究重點應該放在更智能的數據集創建方法上。

當前研究的基礎問題是：對於特殊應用的高級算法或模型，沒有足夠的數據。未來，結合自定義數據集和深度學習模型將成為很多研究論文的主題。因此研究者的輸出不僅涉及算法或架構，還包括數據集或數據收集方法。數據集標註是深度學習工作流中的主要瓶頸，需要大量的手動標註工作。這在語義分割中尤為明顯，因為該領域需要準確標註每一個像素。[20] 討論了很多有用的半自動流程工具，其中一些利用了ORB 特徵、多邊形變形（polygon morphing）、半自動感興趣區域擬合等算法方法。

克服數據缺乏、減少圖像分類深度學習模型過擬合現象最容易也最常見的方法是，利用標籤不變的圖像變換（label-preserving transformation）人為地擴大數據集。該過程叫做數據集增強，指基於已有數據通過剪裁、縮放或旋轉等方式生成額外的訓練數據。人們希望數據增強步驟需要極少的計算，且可在深度學習訓練流程中實現，這樣變換後的圖像就不必存儲在磁盤中了。數據增強使用的傳統算法方法包括主成分分析（PCA）、噪聲添加、在特徵空間的樣本之間進行內插或外推，以及基於分割標註建模視覺語境周邊物體。

轉載自：機器之心

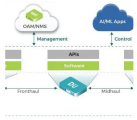
—THE END—

收錄於合集#深度學習 114

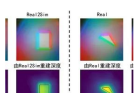
下一篇 · 項目案例：基於YOLO 的鋁型材表面缺陷識別

喜歡此內容的人還喜歡

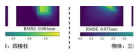
O-RAN的網絡架構
和老康一起學5G



視觸覺傳感器的仿真—現實雙向遷移研究



CAAI認知系統與信息處理專委會



一文盡覽| 計算機視覺中的魚眼相機模型及環視感知任務匯總！

自動駕駛之心