

中大「深度學習」AI醒「目」 準確率超越人類肉眼

AI
新時代之
視聽無雙

視覺及聽覺是人類最重要的認知渠道，而對於電腦來說，單純接收影像及聲音並不困難，但要進而辨別及判斷視聽訊息，做到真正「觀看」及「聆聽」，卻是相當大的挑戰。

本港大學及科技產業研究人員，正致力突破人工智能（AI）的視聽層面，並成功開發出相關的監察及教學應用。其中中文大學研究團隊以先進的「深度學習」技術，讓電腦系統「帶眼識人」，其準確率已超越人類，一度登上全球之冠；團隊更讓系統學會通過鏡頭辨認人群動向，監察活動安全性，支援防止罪案以至反恐，成為保障公共安全的得力助手。 ■香港文匯報記者 鄭伊莎



■王曉剛與團隊透過「深度學習」方式，輔以大型數據庫，研發能自動識別人面的AI系統。 劉國權 攝

電腦帶「眼」識人 監察人流助反恐

中大電子工程學系副教授王曉剛與團隊，透過模擬大腦行為建構人工神經網絡的「深度學習（deep learning）」方式，研發能自動識別人面的AI系統，讓電腦「帶眼識人」。他表示，系統集合了逾五百個人工神經元，分別代表着人面不同的屬性，可描述出具有不同特徵的人面，例如有沒有戴眼鏡。

團隊又把神經元格式化，讓部分神經元能判斷人的性別或年齡，當神經元的反應由強變弱時，即其所見的人面亦由男變女，或由老年人變成年輕人。

為讓系統變得「聰明」，研究團隊收集了逾萬張名人相片，不斷訓練它識別容貌。王曉剛補充指，每次給一種人面特徵讓神經網絡去分辨，當它算錯了，就給反饋信號，告訴它算錯了，再糾正網絡中的參數，「讓它學會識別這些人面特徵的表達，所以之後也能辨識出沒有見過的人。」

若神經元增加將有何優勢？「它的識別能力會更強，當你要識別上千萬的人，就需要更複雜的網絡了。」

雜的網絡了。」

準確率達99.53% 一萬視頻訓練

在約兩年前，中大的人面識別系統「認人」準確率突破99%，首次超越人類肉眼準確度並登上世界第一，現時準確率更進一步提高至99.53%。研究團隊去年再成功提升系統的智能，讓其單靠「觀看」影片畫面便推算出人群密度，同時監察人群活動安全性。因其應用層面廣泛，在反恐、執法等領域均可發揮極大作用。

王曉剛解釋指，由於人與人之間會互相遮擋，要令系統學會分析人的行為頗有難度，於是他們建立了全球首個有一萬條視頻的數據庫，去訓練系統的智能，「裡面有逾八千個場景，人群可以分為94個屬性，例如發生在什麼地方、有什麼人、在做什麼事，這些屬性可完整地描述人群行為。」

「通過監控鏡頭，系統可以辨認人群的活動、檢測其密度，以及跟蹤和監測相關運動的

方向，若發現有異常情況，即可能涉及危險或恐怖活動，系統可以立即作出警報或報警。」

他續說，系統已應用在北京、成都、深圳等地的安全部門及交通部門、地鐵及廣場等，協助監察公共安全。

續研系統手機化 料數年內完成

在技術層面而言，中大AI視覺系統經過極龐大的數據庫訓練，其智能大為提升，能分析逾千萬甚至近億的數據。那麼研究是否可止步了？王曉剛搖頭說：「現在我們在探討如何讓人面識別系統應用在手機上，所以要把神經網絡及有關參數變小，令它變得很快；而在參數變小時對數據的描述及分辨能力卻要夠高，這是很大的技術難題。」不過，現時AI技術發展迅速，他有信心數年內會完成。

除了人面外，王曉剛與團隊亦正以深度學習訓練AI系統檢測及辨認死物，現時準確率已由3年前的20%進步至60%，能檢測逾二百物體，包括車輛的去向等。

「神經元」增至百層利信息流

話你知

什麼是深度學習（deep learning）？王曉剛解釋指，深度學習是以一套建構神經網絡的方法，來模擬人類大腦的運作，「人的大腦有很多神經元，當中有很多連結，通過這些連結可以傳遞信號。」

他們建立的人工神經網絡，便可發揮此作用。研究團隊利用高效能圖形處理器，以海量的數據訓練電腦學習，分辨人面及識別物件。

人類大腦的視覺皮層有約6層神經

組織，幫助眼睛識別物件。王曉剛指，1990年代電腦的人工神經元僅得3層至5層，隨著科技發展，現已增至數百層至千層，「為何深（deep）就重要？因為當神經網絡的層數加深後，表達能力會有幾何增長」，從而獲得更高智能，最終超越了人類認人的水平。

研究團隊並致力建構更複雜的神經元模型。王曉剛說，以往「深度學習」只是一層層之間互相傳遞信號，現時已發展至可由上而下的層數之間傳遞信號，令信息流更為複雜。 ■記者 鄭伊莎



■團隊以反饋信號訓練電腦辨別人面特徵。 劉國權 攝

■AI可分析音準、速度等並即時給予分數及評語，化身「音樂助教」。 劉國權 攝

「聽」音階給評語 Scalebook 助習樂應試

音樂助教

科幻電影中，人工智能機器人擁有高智慧，甚至與人類平起平坐；而當下香港的科技發展，亦成功將人工智能結合音樂，讓電腦系統能「聆聽」歌聲及樂曲，再進一步分析有關音準、速度等並即時給予分數及評語，化身「音樂助教」，協助學生學習樂器。

科技公司Playnote創辦人容志偉（Eric）身兼電子工程專家及專業音樂家，更獲英國皇家音樂學院授權，致力開發有助應試的音樂教育程式。早在3年前，該公司針對英國皇家音樂學院考試當中的聆聽測試推出「Auralbook」，以人工智能技術提供模擬考試的練習，即時分析學生



■Eric成功將電腦系統智能提升至「聽」到樂器演奏時的走音。 劉國權 攝

歌聲的音準、打拍子的力度和節奏，有關程式更獲得亞太資訊及通訊科技獎金獎、香港資訊及通訊科技獎金獎等多個獎項，屬本港業界的先驅之一。

升級3年前程式 能判斷走音

至今年Eric成功將電腦系統的智能擴闊及提升，推出能針對樂器演奏的scales（音階）作判斷並即時回饋的「Scalebook」程式，以回應真實考試中，學生以樂器演奏不同大調（major）和小調（minor）的音階（scale）時對音準、速度等不同的要求。

Playnote團隊經歷約3年時間，讓人工智能由「聽」人聲發展到「聽」樂器聲。Eric解釋當中原理：「每種聲音有其特性，人工智能可憑當中的波形（waveform），分析到聲音的類型。」

如何調教「AI耳朵」去分辨走音呢？原來人工智能一秒內已經可以收集上千個數據頻率，Eric指：「如果在一秒的數據裡，部分數據有特別的波幅差別或特別的傾向，就表示出現走音。」

兩年儲數千數據 克3難關

開拓「Scalebook」，團隊需要克服三大困難：準確性、速度，以及計算出不同錯誤的可能性。Eric指，演奏樂器時每個音只相差數個赫茲（Hertz），克服困難的方法就是擴大數據庫（database），「現在有數千個數據，約花兩年時間儲，需要真人現場演奏不同的「錯誤」，讓人工智能『認得』。」

除此之外，「Scalebook」能模擬考官給予評語。Eric透露，要讓人工智能「學會用語言表達」一點也不容易，團體在最後研發階段便主要鑽研該部分，過程需要參考英國皇家音樂學院考官曾經實際所給的評語，為系統字庫提供足夠資料，「『Scalebook』有足夠的字詞去形容樂章，但你問它想吃什麼，就答不上了。」 ■記者 吳希雲

非公開數據難得

摸索前行

「深度學習」一直被業界視為一個黑盒子，要深入研究往往有很多未知之數。過去多年王曉剛致力設計讓「深度學習」應用在人面識別系統的專用模型，期望「打開黑盒子」。過程中自然會有不少問號，「因為前人沒做過，我們實在不知道『深度學習』能做到什麼程度，相當於在黑暗中摸索一樣。」

王曉剛憶述，他的兩組學生在2012年開始利用深度學習，各自以不同的方法去研究人面識別系統，因互相競賽的關係，系統「認人」的準確率亦不斷提升，「比如一組做到了97%，另一組會更努力去解決困難，後來慢慢做到了99.53%，超越了肉眼水平。」

他又提到，在訓練電腦「學習」的過程中，需要龐大的數據量，「當你將人工智能的系統，當成是一個火箭，大數據就是火箭的燃料，網絡的結構就是火箭的發動機，所以你要要有充足的燃料及發動機，火箭才能騰飛。」

不過，大學如何獲得非公開的數據卻是另一困難，「在數據較少的情況下，我們怎樣把通用模型遷移至數據比較少的系統應用，是很困難的。」 ■記者 鄭伊莎



欠感情「不可能練出朗朗」

輔助性質

看着人工智能的發展，由分辨人聲「進化」到分辨更複雜、音域更廣的樂器聲，還能夠提供模擬考試，給予評語，程式基本上已做到音樂老師的工作。那人工智能在音樂教育會否有朝一日取代人類呢？Eric的答案是「絕對不可以」，「缺乏了人類情感，人工智能是不可能訓練出一個朗朗！」

《荀子·樂論》有云：「夫（音）樂者，樂也，人情之所必不免也。」對於Eric來說，任電腦硬件和軟件發展再迅速，人工智能始終缺乏人類的情感，「它能分析到音準和拍子，但人卻可以分析演奏時的感情。正如你會聽CD，但都會購票去聽現場的演奏會，為的是那個你喜歡的音樂家，希望聽到他帶感情的演繹。」

Eric相信，科技對音樂教育來說終究只是輔助性質，協助學生練習，而真正的老師用心觀察學生的進度，指導正確的手勢，並即時矯正。

冀EdTech方面研究更多

觀乎現在香港人工智能技術的發展，Eric指雖然不少資訊科技應用都涉及人工智能，但以音樂為主的確實較少，即使是學術機構中的學者或研究生，可能也只有兩三篇論文會簡單提到，他認為「團隊相信發展教育科技（EdTech）是很自然的事，但香港很多人也覺得『唔正常』，不明白為什麼要這樣做」，期望未來更多人才於人工智能與教育科技中發展。 ■記者 吳希雲