Nvidia 首席研究總監 劉洺堉博士：實時肖像生成技術的革命性突破

文／鍾乙君、高嘉豪 多工所碩士生、林廷翰 機器人碩士學位學程碩士生

在當前疫情後時代，線上會議的普及程度前所未有。因此，對於能夠準確代表個人的虛擬形象需求大幅增加。近年來，我們見證了許多驚人的技術突破，其中之一便是影像生成領域的顯著進展。為了探索這一領域的最新成果，我們邀請到劉洺堉博士在陽明交通大學舉辦兩場 演 講「Deep-learning-based Live Portrait」及「Diffusion Models for Image, Video, and 3D generation」，聚焦於他如何以專業來解釋現代生成技術的進展與風險。

劉洺堉博士為 NVIDIA 高級研究總監，並帶領著一個專注於深度生成模型及其應用的研究小組。他的貢獻對於 NVIDIA 產品的開發至關重要，例如 NVIDIA Maxine 和 NVIDIA Picasso 等產品。劉博士在該領域的卓越表現獲得了廣泛認可，因此劉洺堉博士也是影像生成領域中備受矚目的研究者，在深度學習技術的推動方面，成功實現了逼真卻又由人工創造的內容。

劉博士首先介紹了他們在實時肖像生成方面的方法，該方法涉及將影像拆分為不同的特徵，例如外貌、頭部姿態和表情。這種拆分不僅可以在傳輸成本最低的情況下高效重建原始影像，還可以通過操縱不同特徵來生成人工影像。值得一提的是，他們的影像壓縮技術的編碼效率比傳統的 H.264 影像壓縮標準高出十倍，但劉博士也承認，這種方法確實需要更多計算複雜度。

在先前的方法中，遮蔽是一個常見的挑戰，這指的是源圖像中不可見的區域，因此很難進行準確的合成。當僅使用單一源圖像作為輸入時，問題就會出現。為了克服這個限制，團隊使用多個圖像集進行生成，並使用適應不同輸入幀數的變壓器設計。此外，劉博士還展示了他們基於音頻輸入的臉部運動生成工作。通過利用來自 YouTube 的大量可用影像，他們訓練了一個模型，使得能夠將音頻與臉部運動相互關聯，因此能夠僅基於音頻就生成表情。

在肖像生成的過程中，照明也是一個至關重要的方面。在一個設置在陽光明媚的場景中的線上會議中，即使在昏暗的室內環境中，也要讓人物看起來照明逼真。為了解決這個問題，劉博士提出了一項將輸入影像重新照明以適應任意照明環境的方法。這個方法涉及使用各種環境下的多樣人物圖像，可以在更低成本下實現相當水準的效能。

劉博士也強調了在這些技術開發過程中，訓練數據的重要性。正如聊天機器人 (ChatGPT) 的架構在每個版本中保持相對穩定一樣，關鍵的改進在於數據的收集和處理方面。對於肖像生成的每個模塊，高質量的訓練數據都是不可或缺的，以保證其有效運作。

在劉洺堉博士深度影像生成的演講中，他展示了人工智慧在這方面所取得的顯著進展，也為我們提供了有關當前最先進技術的深入見解與寶貴的啟示。他們的工作為增強虛擬互動提供了令人興奮的可能性，同時也有望應用於其他領域。在此，我們要向劉洺堉博士表達最衷心的感謝，感謝他的慷慨分享。他的演講為我們打開了一扇通往人工智慧前沿的大門，也為未來數字科技的發展指明了方向。我們期待著更多的精彩演講和突破性研究，同時也期盼著這些創新成果能為我們的生活帶來更多便利與驚喜。

Nvidia Chief Research Director Dr. Ming-Yu Liu: Revolutionary Breakthrough in Real-time Portrait Generation Technology

In the post-pandemic era, holding online meetings in the workplace has become very common. Therefore, there is a significant increase in demand for accurate representations of individuals through virtual avatars. In recent years, we have witnessed many astonishing technological breakthroughs, one of which is the remarkable progress in the field of image generation. To explore the latest advancements in this field, we invited Dr. Ming-Yu Liu to give two lectures at our school, titled “Deep Learning-Based Live Portrait” and “Diffusion Models for Image, Video, and 3D Generation.” These talks focused on his expertise in explaining the advances and risks of these technologies."

Dr. Ming-Yu Liu is the Chief Research Director at NVIDIA, where he leads a research group dedicated to deep generative models and applications. His contributions are vital to the development of NVIDIA products, such as NVIDIA Maxine and NVIDIA Picasso. His outstanding performance in this field has gained wide recognition in the domain. In the talk, Dr. Liu first introduced their approach to real-time portrait generation, which involves splitting images into different features, such as facial appearance, head pose, and expressions. This segmentation not only allows for efficient reconstruction of the original image at the lowest transmission cost but also enables the generation of artificial images by manipulating different features. In addition, the image compression technology of their encoding efficiency is ten times higher than the traditional H.264 image compression standard. However, Dr. Liu acknowledges that this approach does require more computational complexity.

For example, Occlusion has been a challenge in previous methods. It refers to the invisible areas in the source image that decrease the accuracy of synthesis, especially when using a single source image as input. To overcome this limitation, his team used multiple image sets for generation and designed transformers capable of adapting to different input frame rates. Additionally, Dr. Liu demonstrated their work on facial motion generation based on audio input. By utilizing a vast amount of available video footage from YouTube, a model was trained to associate audio with facial movements. Therefore, expressions can be generated solely based on audio.

Lighting is also another crucial aspect in the portrait generation process. In an online meeting set in a sunny environment, it is necessary for the participants to appear realistically lit even in dim indoor surroundings. To address this issue, Dr. Liu proposed a method that re-illuminates input images to adapt to any lighting environment. This method involves using diverse images of individuals in various lighting conditions and achieves a substantial level of performance at a lower cost. Dr. Liu also emphasized the importance of training data in the development of these technologies. Just as the architecture of chatbots like ChatGPT remains relatively stable in each version, key improvements lie in data collection and processing. High-quality training data is crucial for the effective operation of each module in portrait generation.

In Dr. Liu's lecture on deep image generation, he showcased significant advancements in artificial intelligence in this regard and shared profound insights and valuable experiences about the current state-of-the-art technology. Their work offers exciting possibilities for enhancing virtual interactions and holds promise for applications in other fields. We would like to express our heartfelt gratitude to Dr. Liu for his generous sharing. His lectures demonstrated insight into artificial intelligence and pointed the way to the future development of digital technology. We look forward to more exciting lectures and groundbreaking research and hope that these innovative achievements will bring us greater convenience and surprises in our lives.