首創增強式學習 AI 影像壓縮技術

文／翁健棋

自 Open AI、Google 所開發之 ChatGPT、Bard 等 AI 聊天機器人問世以來，人工智慧領域內相關技術逐漸為大眾所熟知，其所能應用領域之廣，執行產出之高效率、高完成度無不令人驚嘆。伴隨 GPU 運算能力大幅飆升、演算法成熟，AI 技術已被廣泛應用於許多智慧應用服務中，然而於多媒體領域內，關鍵之「影像／視訊壓縮技術」卻無法取得重大演進，導致此難題持續困擾產學研各界二十載有餘。

近幾年受新冠疫情影響，串流影音娛樂、遠距互動需求顯著提升，考量到時下流行之娛樂影音平台（如：YouTube、Netflix）與視訊會議服務皆十分仰賴影像／視訊壓縮技術協助，由本院資訊工程學系彭文孝教授領軍開發的「增強式學習 AI 影像壓縮技術」便希望透過人工智慧學習進化之特性，進行影像編碼優化，實現壓縮效能突破之可能。彭文孝教授於受訪時表示，若想在未進行壓縮之情形下傳輸視訊影像，以每秒傳輸之畫面數約為 20 至 60 張來說，遠超一般網路所能負荷之頻寬，如何進行有效的影像／視訊壓縮是各界專家、學者苦思已久，長久以來面臨突破瓶頸的重要課題。

原先影像視訊使用之壓縮技術，係採用人類開發的數學演算法，彭文孝教授團隊選擇跳脫既定思維，欲嘗試透過 AI 技術取代數學演算法，提高壓縮效能，同時顧及影像品質需求，突破現有技術限制。目前全球將 AI 應用於影像／視訊壓縮領域的發展趨勢，共有 AI-based、AI-assisted、Hybrid-based 三大面向。在 AI-assisted 影像／視訊壓縮技術方面，開發團隊首創採用「增強式學習」進行編碼優化，該技術可在不更改既有編解碼器前提下，實現壓縮效能之提升；團隊已將此技術發表於 2021 年的 Data Compression Conference，且分別在台灣及美國完成專利申請。

另一方面，開發團隊也在 AI-based 端對端學習式影像與視訊壓縮技術領域有所突破。利用最新的 Normalizing Flow 生成模型，可使目標影片在相同位元率下，擁有最佳重建影像品質，同時在效能上超越傳統的壓縮標準 HEVC，甚至可接近 2020 年由 ITU（國際電信聯盟）和 ISO（國際標準化組織）最新制定的視訊壓縮標準 VVC。此外，學習式影像與視訊壓縮技術相較傳統 HEVC 之壓縮品質，在主觀視覺效果上顯著提升，可見技術之突破性與發展潛力。未來待 4k 和 8k 等高畫質視訊影像需求普及，能進行高效能、高品質產出之 AI 技術導向影像／視訊壓縮技術將扮演關鍵角色，左右相關產業之發展趨勢。

憶起開發時所遭遇的瓶頸，彭文孝教授特別感謝國家高速網路與計算中心所提供的硬體設備支持。原先團隊在投入開發該技術時，皆需透過平行運算將研究室內 20 多台的電腦串連起來，方能免強滿足專案需求，但仍舊無法避免 AI 參數調整、模型建置時所需付出之時間成本。自從租用國網中心所提供的台灣杉一號服務後，受惠於該平台支援多種深度學習框架容器的特性，原先需要數小時的環境建置時間，縮短到數秒鐘即可完成，還同時解決了運算架構維護曠日費時、電腦設備難以定時更新等問題。

目前，由彭文孝教授所領軍的開發團隊不單已完成多項專利申請，同時與國際大廠簽訂合作專案，更在國際壓縮標準組織 JPEG AI 委員會舉辦的學習式圖像編碼提案徵求中獲得第二名之佳績。期待團隊所開發之增強式學習 AI 影像壓縮技術能於未來普及應用，顛覆並突破現有的技術限制！

Revolutionizing Image and Video Compression with Artificial Intelligence

Ever since the emergence of AI chatbots like ChatGPT and Bard, developed by OpenAI and Google, the public is getting acquainted with related technologies in the field of artificial intelligence. The wide range of applications, the remarkable efficiency, and quality of their outputs are all astounding to everyone. Thanks to substantial improvements in GPU computing power and algorithm development, AI technology has been widely applied in diverse intelligent application services. However, the crucial domain of "image/video compression technology" in multimedia has remained a long-standing challenge for academia, industry, and research, with no significant breakthrough achieved over the past two decades.

Due to the impact of the COVID-19 pandemic in recent years, there has been a significant increase in demand for streaming video entertainment and remote interaction. Recognizing the heavy reliance on video/image compression technology by popular video streaming services like YouTube and Netflix, as well as video conferences, Professor Wen-Hsiao Peng and his team from the Department of Computer Science have successfully developed an innovative technology called "AI-assisted Image and Video Compression with Reinforcement Learning." Their objective is to enhance image or video coding through the evolutionary characteristics of machine learning, leading to significant advancements in compression efficiency. During an interview, Professor Peng emphasized that the bandwidth requirement to transmit an uncompressed video sequence with a frame rate ranging from approximately 20 to 60 frames per second exceeds the capacity of typical networks to handle. As a result, how to effectively compress video/image data has been a longstanding crucial challenge faced by experts and scholars in various fields.

Previously, video and image compression relied on mathematical algorithms designed by humans. However, Professor Peng's team decided to break away from conventional methods and explore the use of AI technology as a substitute for these algorithms in order to improve compression efficiency while considering visual quality requirements and overcoming existing technological limitations. Currently, there are three major trends in the global development of AI applications in image and video compression: AI-based, AI-assisted, and Hybrid-based approaches. Regarding AI-assisted image/video compression, the team made significant strides by introducing "reinforcement learning" for encoding process. This innovative technique levels up compression efficiency without modifying the existing codec. The team successfully presented this method at the Data Compression Conference in 2021 and subsequently filed patents in both Taiwan and the United States.

On the other hand, the team has made breakthroughs in the field of AI-based end-to-end learning-based image and video compression. By utilizing the latest Normalizing Flow generative model, the new method can attain the optimal visual quality of the reconstructed target video while maintaining the same bitrate. Not only does this surpass the performance of the traditional compression standard, HEVC, but it also approaches the most recent video compression standard, VVC, established by the ITU (International Telecommunication Union) and ISO (International Organization for Standardization) in 2020. Moreover, compared to the conventional HEVC, the learning-based image and video compression technology substantially enhances the subjective visual quality, which indicates the groundbreaking nature and potential development of this technology. In the future, as there is an increasing demand for high-definition video sequences such as 4K and 8K, AI-based image/video compression technology with high efficient and high-quality reconstruction will play a pivotal role in influencing the development trends across relevant industries.

Recalling the bottlenecks encountered during development, Professor Peng deeply expresses appreciation to the National Center for High-Performance Computing (NCHC) for their hardware support. Initially, the team had to connect over 20 computers in the laboratory using parallel processing to barely meet the project requirements. However, it is still inevitable to incur time costs in AI parameter adjustment and model construction. The adoption of NCHC's Taiwania 1 HPC service, which provides extensive support for various deep learning framework containers, has brought significant benefits to the team. Notably, the previous time-consuming process of environment setup, which used to take hours, has been remarkably reduced to a matter of seconds. This solution has simultaneously addressed issues such as the time-consuming maintenance of computing architecture and the timely upgrades of computer equipment.

At present, Professor Peng and the team have accomplished numerous patent applications and established partnerships with major international companies. They have achieved second place in the Call for Evidence on learning-based image coding organized by the JPEG AI Committee, an international compression standard organization. We hope that the team's innovative technology, AI-assisted image compression with reinforcement learning, will revolutionize and surpass current technological constraints to become widespread in the future.