Research Statement

Chi Keung Tang

During the first decade of my academic career (2000–2009) the focus was *vision and graphics* when SIGGRAPH published some of the most impactful papers to date. Relevant research momentum was fueled by high grossing Pixar/Disney movies and computer games in the coming years, where computer vision found its instrumental role in enabling high-quality special efforts which were used to be costly to produce. My SIGGRAPH papers *Lazy Snapping* and *Poisson Matting* published during this time continues to inspire researchers and practitioners, with the patented technologies applied in TV/movie post-production. Notably, they were HKUST's *first SIGGRAPH papers* as well as the first time a Hong Kong-based researcher publishing more than one in the same year. After these impactful papers, we witnessed a sea change in the landscape of graphics research in Hong Kong not limited to HKUST: faculty were invited to serve on SIGGRAPH committees, students and faculty (many are HKUST PhD graduates) have been making consistent presence on the SIGGRAPH map and recognized as leaders in 2D and 3D computer graphics. These first SIGGRAPH papers have gathered over 2500 citations combined and become "must-read" papers in graduate courses in computational photography around the world. The papers are centered around the main theme of *interactive computer vision*, where one focus is to "bring the user into the loop" to achieve computer vision tasks that have been traditionally very difficult to automate. By accepting a few simple and loose hints, such as strokes or scribbles the user casually marks on an image, the automatic system can output segmentation results of pixel or even subpixel accuracy.

Following the revolution brought by deep convolutional neural network in machine learning, *deep learning for computer vision* was the focus during the second decade of my academic career (2010–2019). Over the past decade, I have focused on publishing high-impact work in top computer vision and machine learning venues *CVPR, ICCV, ECCV, NeurIPS*, where in particular *CVPR* is ranked #4 of all publications in h-5 index, only after *Nature, The New England Journal of Medicine*, and *Science*. During this time, *Network Trimming*, an arXiv report which was rejected by NeurIPS'16, have become one of the seminal works for network pruning to achieve efficient deep architectures. Our recent work on *Few Shot Learning*, a series of CVPR and ECCV papers published since 2020, contribute to image and video object detection and semantic segmentation. We proposed novel deep architectures for video object segmentation, video instance segmentation, multiple object tracking and segmentation, and occlusion-aware video object inpainting.

Stepping into the third decade (2020–) of my academic career, I believe *3D computer vision* holds the key to success in the next-generation high-impact applications. My recent computer vision works focus on *Neural Rendering Fields* (NeRFs). NeRFs are one of the hottest emerging topics in computer vision, computer graphics and machine learning due to their impressive 3D results and high potential in augmented reality, storytelling or even "storyliving" technologies leading to high-quality metaverse applications for immersive user experience. To enable such relevant important applications, our fundamental NeRF contributions are essential on unsupervised multi-view object segmentation (NeurIPS'22), region proposal network for general object detection (*NeRF-RPN*), and direct 3D landmark detection on NeRFs (FLNeRF). In *NeRF-RPN*. Given a pre-trained NeRF model, NeRF-RPN detects all bounding boxes of objects in a scene. By exploiting a novel voxel representation that incorporates multi-scale 3D neural volumetric features, we demonstrate it is possible to regress the 3D bounding boxes of objects in NeRF directly without rendering the NeRF at any viewpoint. Our 3D coarse-to-fine FLNeRF model efficiently samples from the NeRF on the whole face with individual facial features for accurate landmarks. To mitigate the limited number of facial expressions in the available data, nonlinear augmentations are applied at facial features in fine scale to simulate large emotions range, including exaggerated facial expressions (e.g., blowing cheek, wide opening mouth, blinking eyes), for training FLNeRF. Consequently, an animator can easily edit, control and even transfer emotion from another face NeRF. With precise landmarks on facial features, exaggerated facial expressions with wide grinning, mouth wide open, cheek blowing can be readily achieved. NeRF-RPN and FLNeRF are argably the first significant work on NeRFs in their respective areas, which are fundamental and essential in enabling future important 3D NeRF-based applications. Demo links: NeRF-RPN, FLNeRF.