

# 專 題 演 講

講 者：邱維辰 教授 (陽明交通大學資工系)

題 目：Multimodal Prompting with Missing Modalities for Visual Recognition

大 綱： Our observation perceived in daily life is typically mulitmodal, such as visual, linguistic, and acoustic signals, thus modeling and coordinating multimodal information is of great interest and has broad application potentials. Recently, multimodal transformers emerge as the pre-trained backbone models in several multimodal downstream tasks, including genre classification, multimodal sentiment analysis, and cross-modal retrieval, etc. Though providing promising performance and generalization ability on various tasks, there are still challenges for multimodal transformers being applied in practical scenarios: 1) how to efficiently adapt the multimodal transformers without using heavy computation resource to finetune the entire model? 2) how to ensure the robustness when there are missing modalities, e.g., incomplete training data or observations in testing? In this talk, I will introduce our simple but efficient approach to utilize prompt learning and mitigate the above two challenges together. If time allows, I will also briefly introduce other recent works from my research group.