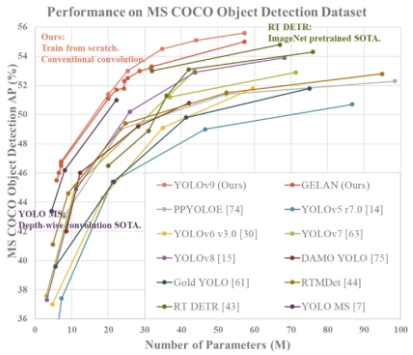
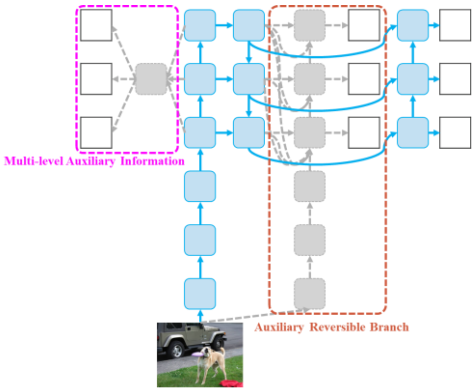
適用於智慧服務的可信賴AI先進技術研究 (陳信希)

說明：本計畫之主軸為開發兼具可信賴性與效能的人工智慧先進技術。就AI技術生命週期各階段，進行可信賴AI技術研究，包含資料收集、標註處理、模型訓練、驗證、到落地使用，探討造成AI技術信賴問題的原因及解決方法。團隊開發的各種AI技術元件，可廣泛應用在不同的智慧服務。以通用型智慧助理作為展示範例，說明各項AI技術如何應在智慧助理系統中，包含語音展頻辨識、未授權資料加密、YOLOv9物件偵測、影像清晰化、場景文字辨識、Taiwan LLM與評測、及提升LLM效能的提示方法。

Descriptions: The focus of this project is to develop advanced AI technologies that can achieve both trustworthy and high efficiency. At each stage of the AI technology lifecycle, including data collection, annotation processing, model training, verification, and inference. We conduct research on trustworthy AI technologies on the root causes of trust issues in AI technology and finding solutions. All different kinds of AI technology components developed by the team can be applied widely in different intelligent services. Using a general-purpose intelligent assistant as a demo example setup, we illustrate how these AI technologies can be applied in an intelligent assistant system. These include speech recognition with bandwidth expansion techniques for noisy environments; perturbation for unauthorized data; YOLOv9 object detection; image debluring techniques; scene text recognition; Taiwan LLM (Large Language Model) and evaluation methods ;and prompt methodology to enhance LLM performance.

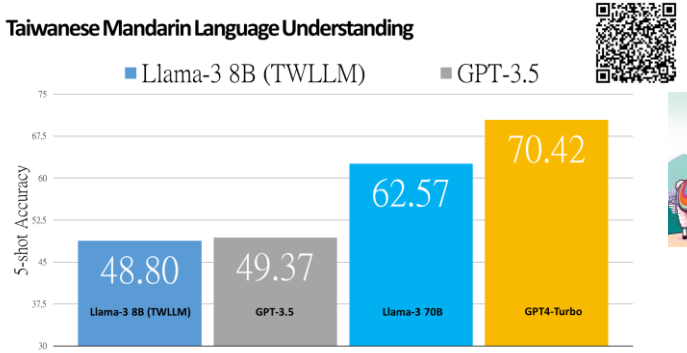
YOLOv9 Object Detection



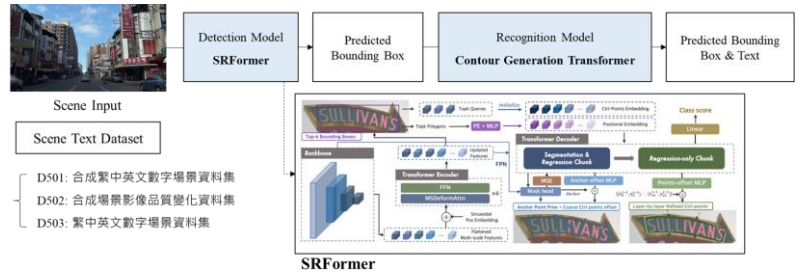


Programmable Gradient Information

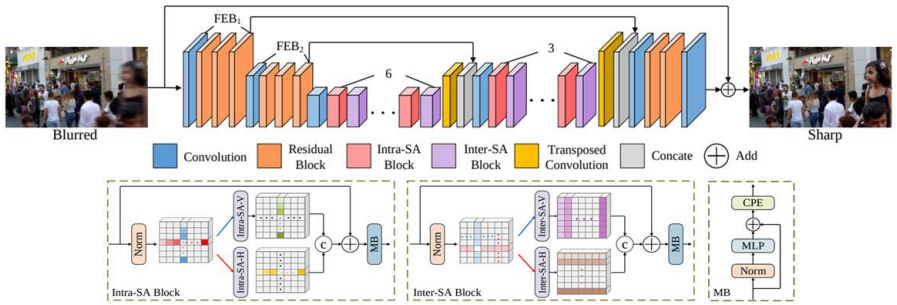
TAIWAN LLM and TMLU



Scene Text Detect and Recognition

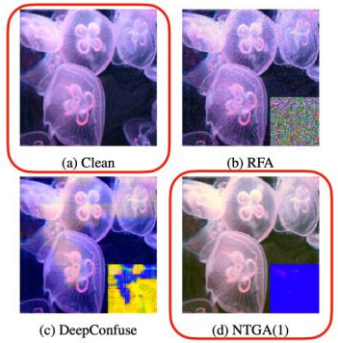


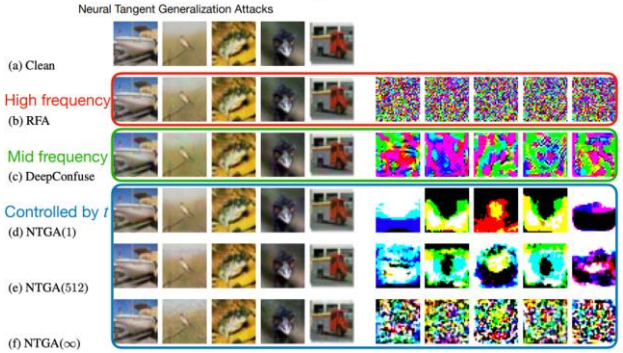
Stripformer - Image Deblur



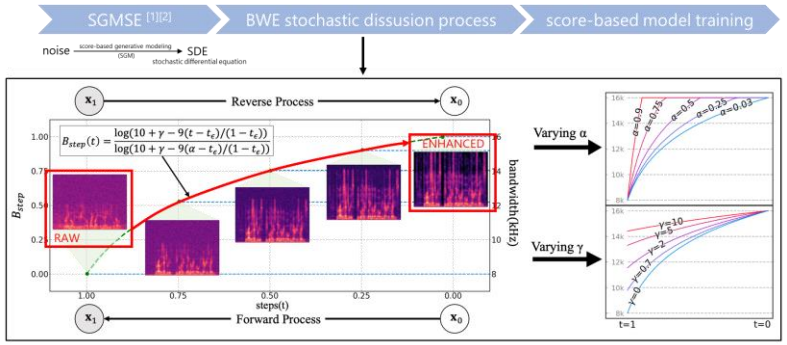
- Efficiently capture long-range information by Intra-SA and Inter-SA
- Successfully deal with blurry patterns with different orientations and magnitudes

Unlearnable NTGA Perturbation





SWiBE - Speech Bandwidth Expand

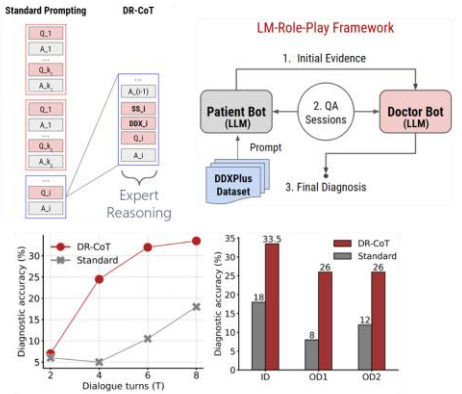


[1] J.Richter,S.Welker,J.-M.Lemercier,B.Lay,andT.Gerkmann, "Speech-enhancement and dereverberation with diffusion-based generative models," IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2023.
[2] S. Welker, J. Richter, and T. Gerkmann, "Speech enhancement with score-based generative models in the complex STFT do- main," in Proc. Interspeech 2022, 2022, pp. 2928–2932.

From Using Prompting Methodology to Proposing New Prompting Methodology

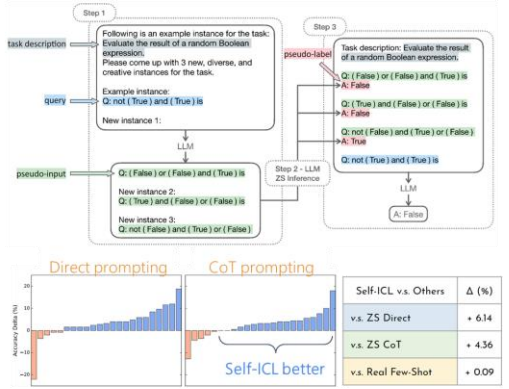
Large Language Models Perform Diagnostic Reasoning @ICLR 2023

Key idea: Demonstrating domain experts' reasoning processes to LLMs significantly boosts their performance



Self-ICL: Zero-Shot In-Context Learning with Self-Generated Demonstrations @EMNLP 2023

Key idea: LLMs are able to enhance their performance by themselves through self-generated demonstrations



Self-ICL v.s. Others	Δ (%)
v.s. ZS Direct	+ 6.14
v.s. ZS CoT	+ 4.36
v.s. Real Few-Shot	+ 0.09