

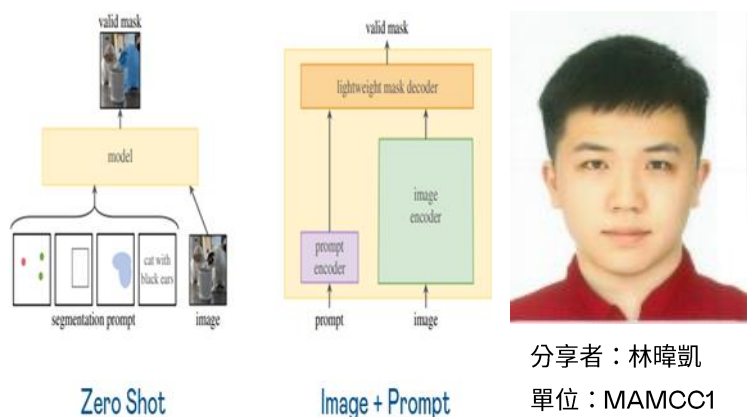
Tap Into The
Possibilities

AI 新技術 - 機器視覺基礎模型

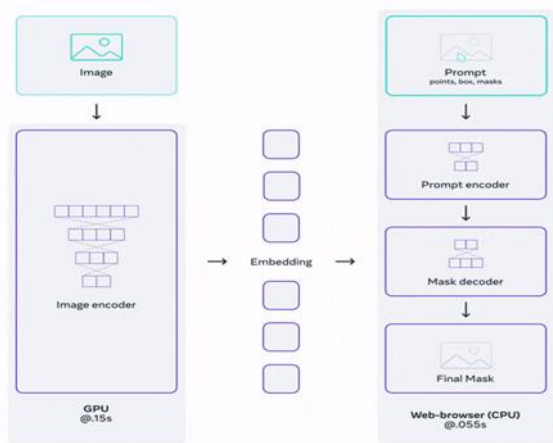
Segment Anything Model

Segment Anything Model：機器視覺的新突破

SAM (Segment Anything Model) 是 Meta 提出的一個全新架構，其初衷是希望做出一個可以認知世界的模型。除了傳統的 Point、Bounding Box 及 Mask 的分割方法，Meta 也將近期很紅的 LLM 概念加入進來，達到僅使用 Prompt 的方式也能直接將想要的物件分割出來，其架構如右圖所示。



SAM 結合實例分割和語義分割



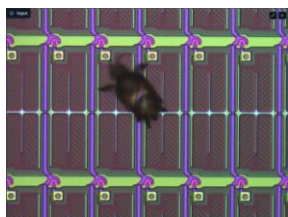
SAM 的模型架構如左圖所示，分別為 Image Encoder、Prompt Encoder 及 Mask Decoder 共三個部分。首先，Image Encoder 使用 ViT，它再特徵擷取上比傳統的 CNN 還有更高的穩定性；其次，Prompt Encoder 使用 CLIP 中已訓練好的 Text Encoder，使文字邊碼能力有著不輸給 DALL·E 2 良好的表現；最後，使用 DETR 和 Mask Classification 相似的架構，可以同時預測出 Semantic 與 Instance-level 的影像分割結果。

Tap Into The Possibilities

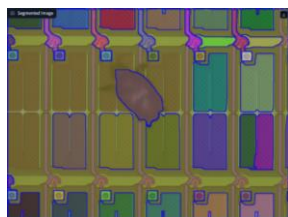
SAM 模型在工業圖像中面臨的挑戰



SA-1B Dataset



原圖



結果

分享者演示了工業圖像在 SAM 的實驗結果，在這個實驗當中並沒有提供模型 Prompt 的資訊，直接讓 SAM 針對圖像上的所有物件進行分割，從結果上來看，SAM 確實已經達到了 Zero-Shot 的水準，但以工業的精度標準來評估，SAM 確實還有一段路要走。分享者提到 SAM 所訓練的資料集是使用 SA-1B，其目的是讓 SAM 認知整個世界，但在認知世界的過程中，也比較難去接觸到工業資料，所以 SAM 對工業資料的物件分割能力會有許多不足之處。

未來展望

SAM 的到來確實對電腦視覺領域帶來巨大的衝擊，這也導致許多人認為 SAM 是在電腦視覺領域的 GPT 3 時刻，它可以更有效地將圖像中的各個物體區分開來，這對自動駕駛、醫學影像等領域都具有重要的應用價值，可促進新圖像分割模型的發展。除了上述應用，SAM 也能作為更大型 AI 系統的一部分，期望結合提示工程 (Prompt engineering) 等，發展出可拆解的系統設計，以用於多模 AI 應用，進而同時處理不同類型的數據，如圖像、文本、語音等，這樣的應用將能更好地認知世界並提供更有有效的解決方案。

分享回饋

MAMCC2 陳翰藝：如果未來關於機台的訊號數據也有大量的資料，應可搭配 SAM 用於判斷訊號的異常情形，但由於訊號資料與原先的圖像資料不同，需考量訓練成本。

MAMCA1 鄭力誠：可以在安全監控、工業自動化和機器人應用中用於對象檢測任務。

MAMCC1 周郁傑：未來可導入人流分析專案，用於解決人群遮擋問題與提升辨識速度。

[更多回饋請點此 \(JamBoard\)](#)

發稿單位：智慧製造發展處

編輯作者：MAMCC1 張晨晏