



IMG & PDF OCR WEBUI

技術手冊

中央大學 MIAT 實驗室

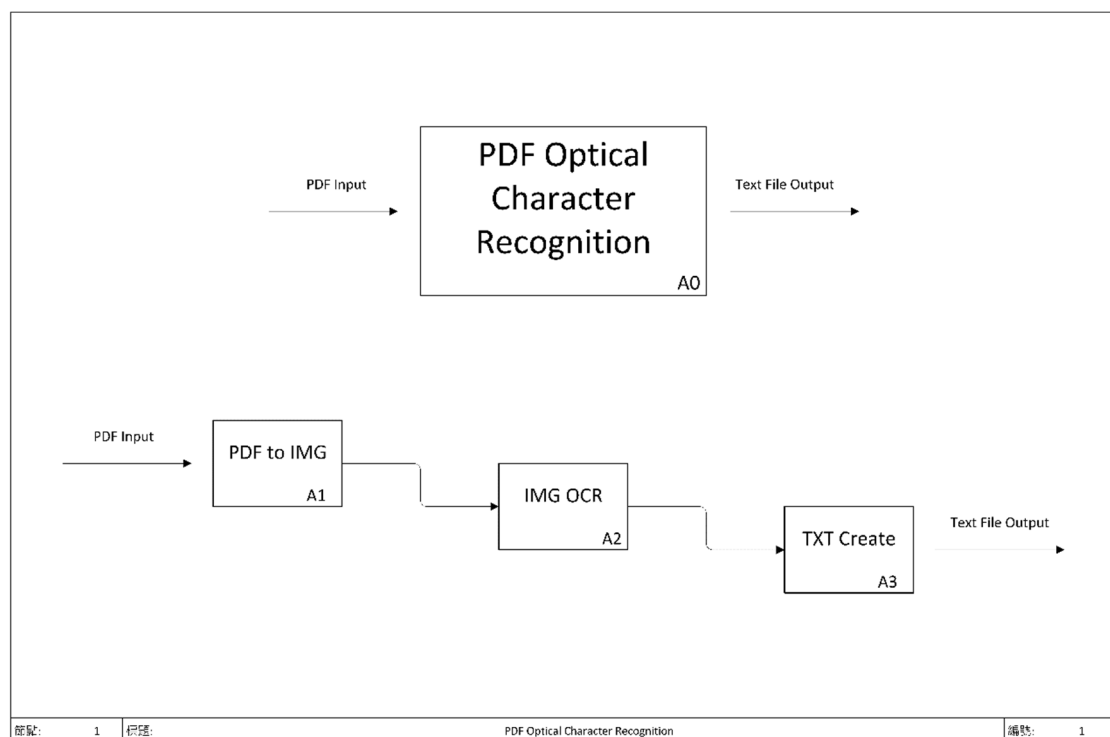
鄧祺文

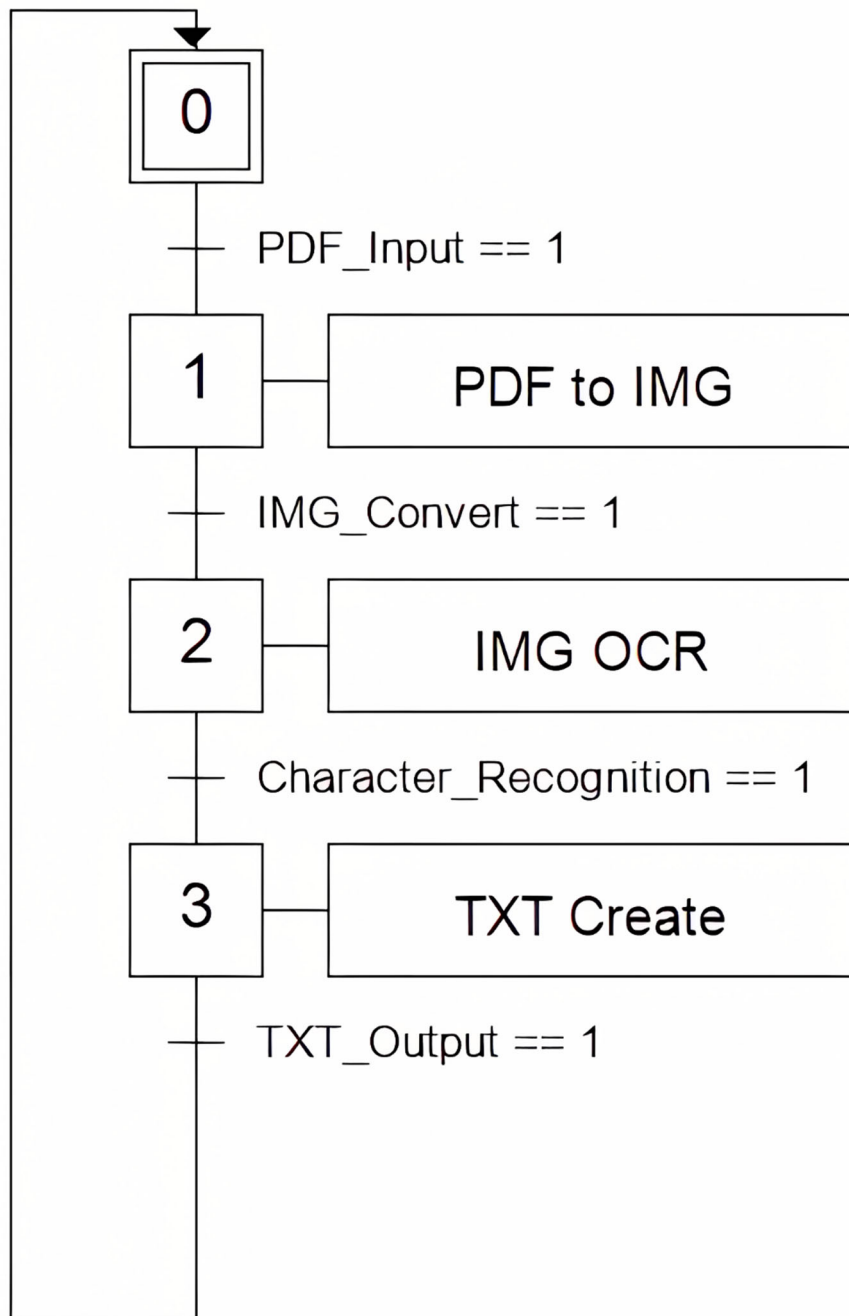
目錄

1. 系統架構設計
2. 辨識模型
3. 網頁功能入口
4. 網頁結果輸出

系統設計架構

- ✧ 功能需求：讀取 IMG 和 PDF 文件，將其中的內容透過光學字元辨識（OCR, Optical Character Recognition）進行擷取，最後輸出成可運用的文字檔案格式。
- ✧ 開發環境：功能 Python, 後端 Flask, 前端 HTML 及 Bootstrap
- ✧ 部屬環境：Docker Container with Ubuntu 20.04, CUDA v11.7.1 and cuDNN v8
- ✧ 專案代碼開源地址：
<https://github.com/toby0622/IMG-Optical-Character-Recognition-Tool>
- ✧ Docker 專案自動化部屬：
<https://github.com/toby0622/IMG-Optical-Character-Recognition-Tool-Docker>
- ✧ IDEF0 和 Grafcet





辨識模型

✧ 論文主題：

SVTR: Scene Text Recognition with a Single Visual Model

✧ 論文地址：

<https://arxiv.org/abs/2205.00159>

✧ 模型代碼開源地址：

<https://github.com/PaddlePaddle/PaddleOCR>

✧ 技術簡介：

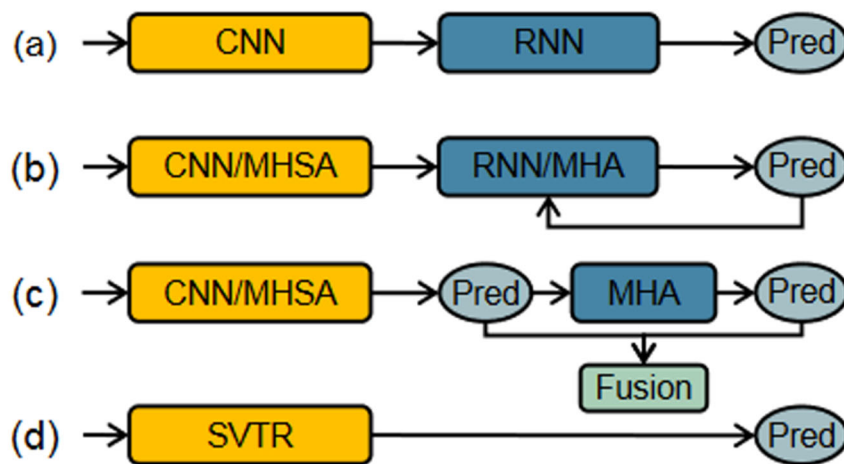


Figure 1: (a) CNN-RNN based models. (b) Encoder-Decoder models. MHSA and MHA denote multi-head self-attention and multi-head attention, respectively. (c) Vision-Language models. (d) Our SVTR, which recognizes scene text with a single visual model and enjoys efficient, accurate and cross-lingual versatile.

場景文字識別可以看作是一個從圖像映射到序列的任務。大多數的識別算法通常由兩個模塊構成，分別包含用於特征提取的視覺模塊，以及用於文本輸出的序列模塊。比如早期基於 CNN-RNN 的 CRNN，和現在一些基於注意力機制，進行自回歸式解碼

的算法，如上圖（圖一）所示。但是這樣設計出之雙階段算法的推理速度往往較慢，難以滿足工業應用的需求。因此該論文從推理速度和模型性能的雙重角度出發，提出了只由 Transformer 構成的純視覺模塊網絡 SVTR，在消費級顯示卡上達到了毫秒級的推理速度，並且參數量僅有 6 M。

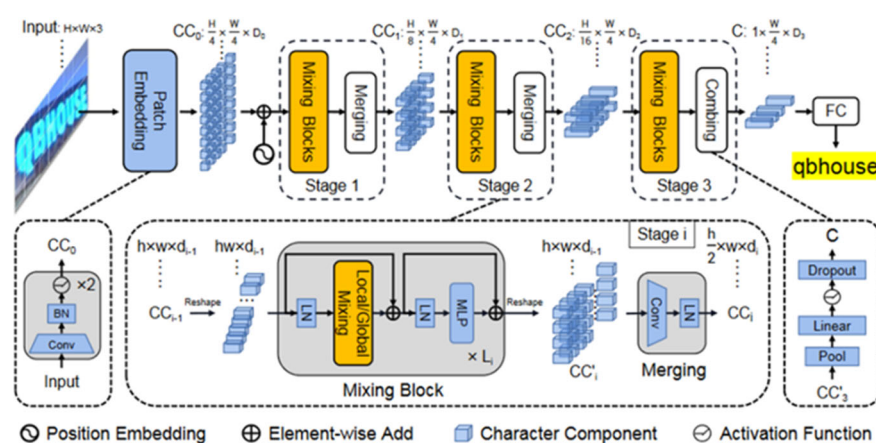


Figure 2: Overall architecture of the proposed SVTR. It is a three-stage height progressively decreased network. In each stage, a series of mixing blocks are carried out and followed by a merging or combining operation. At last, the recognition is conducted by a linear prediction.

上圖（圖二）為該論文所提出模塊網路 SVTR 的整體結構，採用類似於 SwinTransformer 的視覺模型和一個全連接層以及 CTC 解碼器進行文本序列預測。

首先和 ViT 類似，將輸入尺寸為 $H \times W \times 3$ 圖像按照 Patch 進行劃分，得到 $\frac{H}{4} \times \frac{W}{4} \times D_0$ Embeddings。本文采用的 Patch Embedding 操作和 ViT 中的有些許差異，其由兩層步距為 2，卷積核大小為卷積層 3×3 ，以及 BN 層構成。這樣不同的 Patch 之

間是存在著重疊的，如下圖（圖三）所示。經過 Patch Embedding 後的序列將經過一系列的 Stage，每一個 Stage 都由一系列的 Mixing Block 和 Merging Layer 構成。

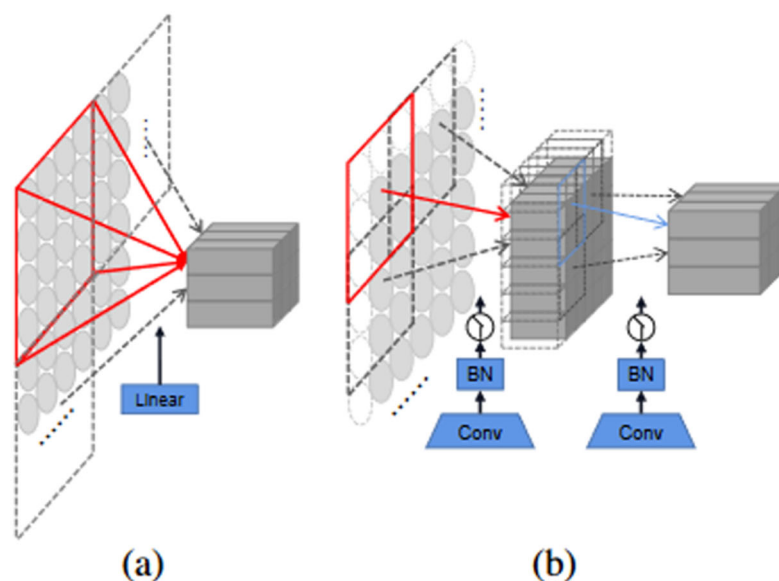


Figure 3: (a) The linear projection in ViT[Dosovitskiy *et al.*, 2021]. (b) Our progressive overlapping patch embedding.

作者認為文本識別需要兩種特征。第一種是局部特征，如筆畫特征。它編碼了字符的不同部分之間的形態特征和相關性。第二種是字符間的依賴性，如不同字符之間或文字與非文字成分之間的相關性。因此，作者設計了兩個混合模塊，即 Global Mixing 和 Local Mixing，通過使用不同大小感受的自注意層來實現。如下圖所示。Global Mixing 層本質上就是一個 Transformer Block，由一個多頭自注意層，一個 Layer Norm 層，以及一個 MLP 層構

成。通過自注意力機制的全局建模特性來進行全局字符建模。

Local Mixing 則是採用了帶窗的自注意層，窗大小設置為了 7×11 。

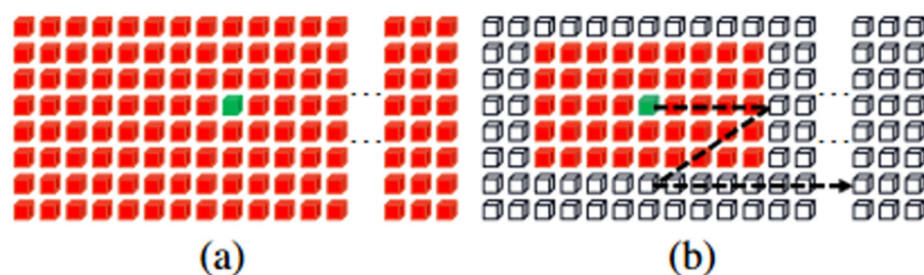
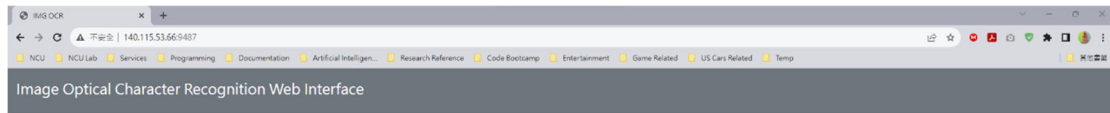


Figure 4: Illustration of (a) global mixing and (b) local mixing.

Merging 層扮演著將輸入序列進行下采樣的角色。其由高度方向步距為 2，寬度方向步距為 1，卷積核大小為 3×3 的卷積層構成。將輸入序列的尺寸由 $h \times w \times d_{i-1}$ 縮小為 $\frac{h}{2} \times w \times d_{i-1}$ 。同時每經過一次 Merging 層，序列的 Channel 維度也會增大，從而彌補在高度上的信息損失。

網頁功能入口

✧ 功能入口



IMG OCR TOOL

☐ Chinese Vertical Right To Left Format

選擇檔案

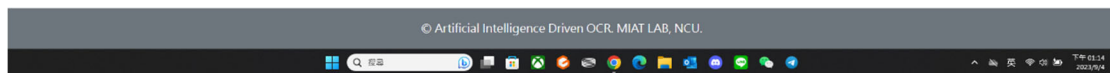
未選擇任何檔案

IMG Upload

選擇檔案

未選擇任何檔案

PDF Upload



✧ 上半部選擇框：支持 JPG、JPEG、PNG

選擇檔案

未選擇任何檔案

IMG Upload

✧ 下半部選擇框：支持 PDF

選擇檔案

未選擇任何檔案

PDF Upload

✧ 圖像特殊排序：中文直式由右至左（預設標準為由上至下）



Chinese Vertical Right To Left Format

✧ 成功讀取資料後，可於下方看見處理進度條

選擇檔案

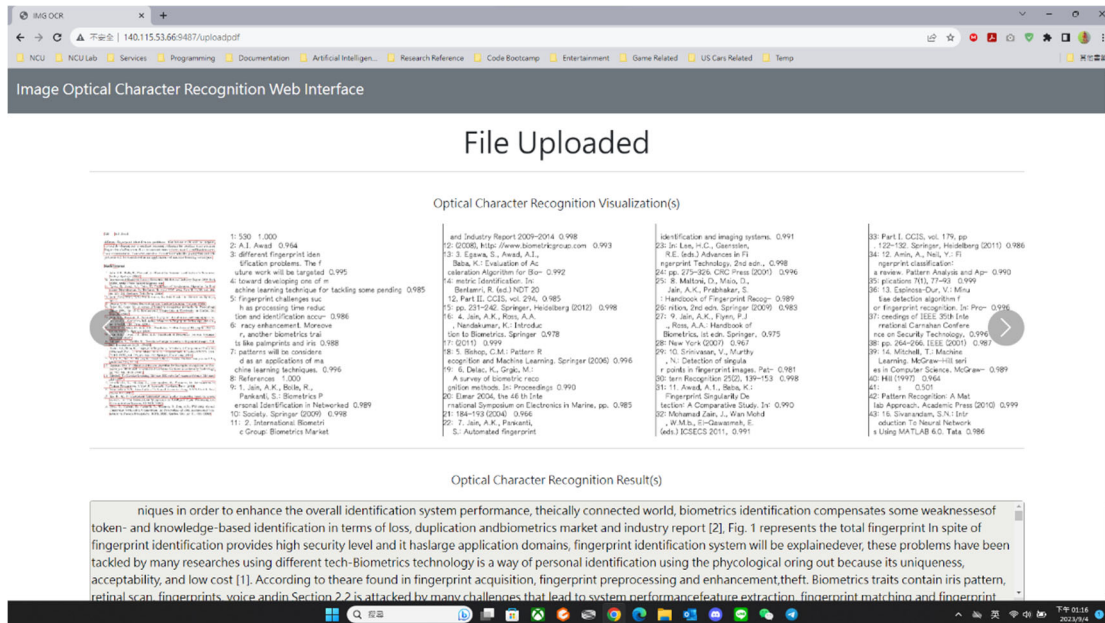
Machine Learning Techniques for Fingerprint Identification A Short Review.pdf

PDF Upload

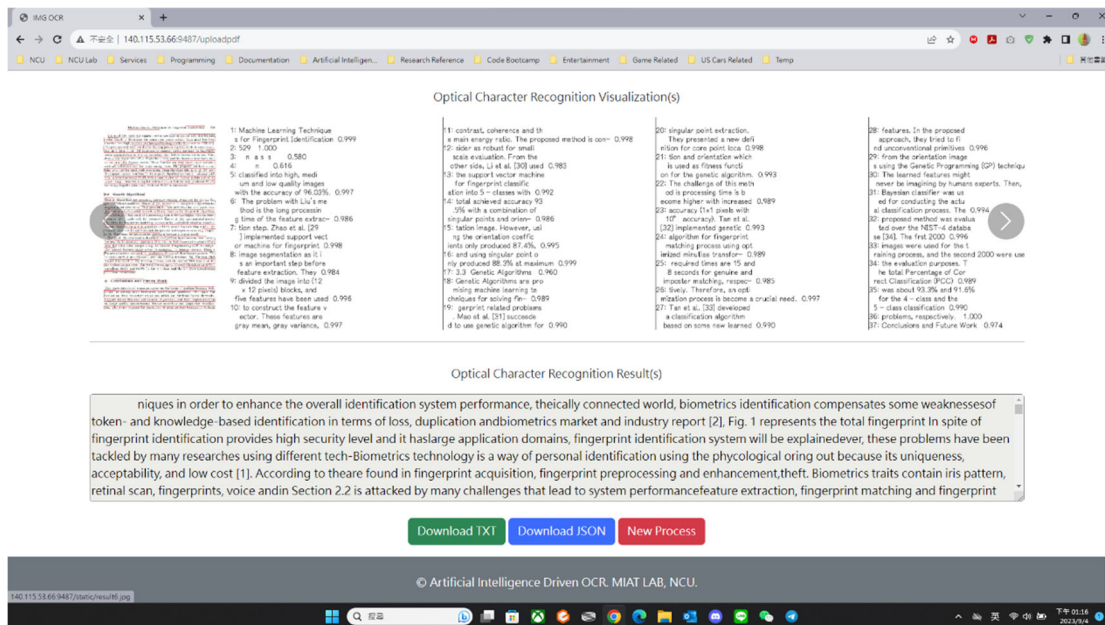
12%

網頁結果輸出

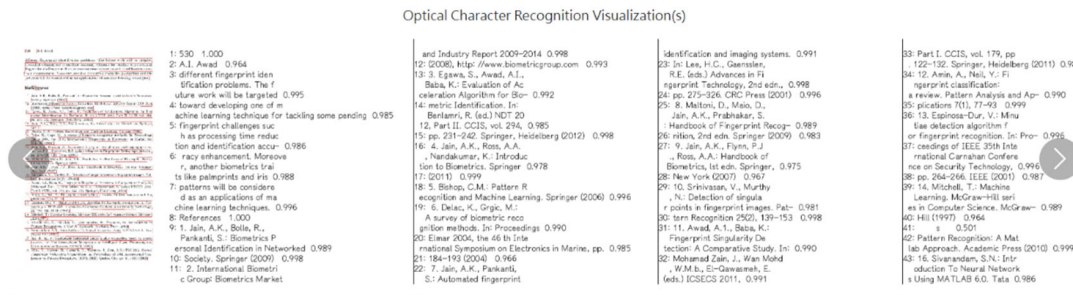
◇ 結果輸出頁面：上半部



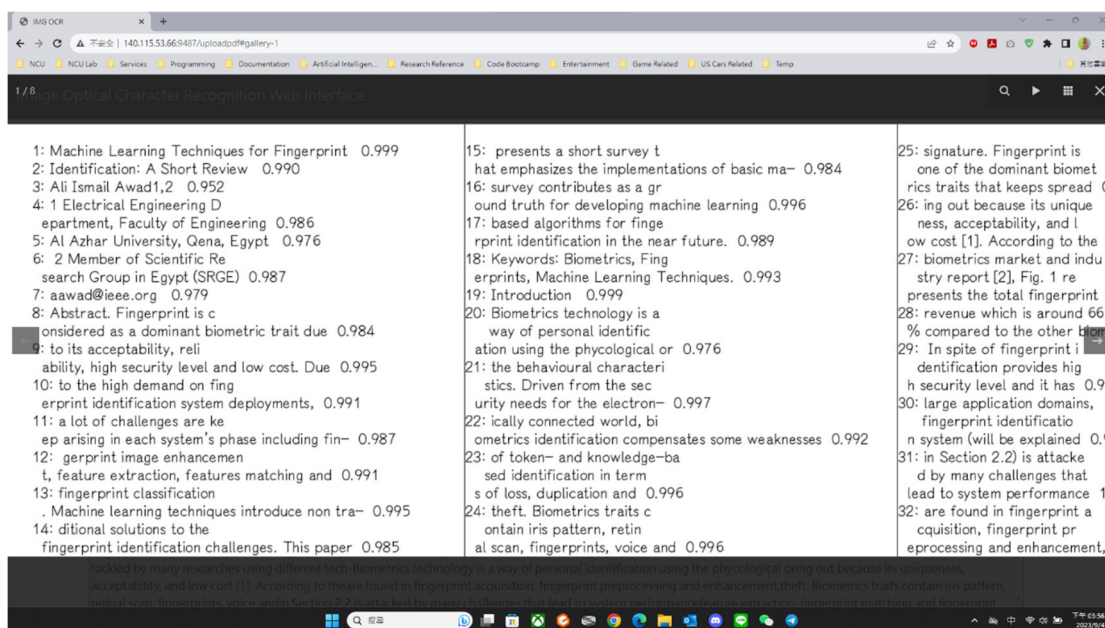
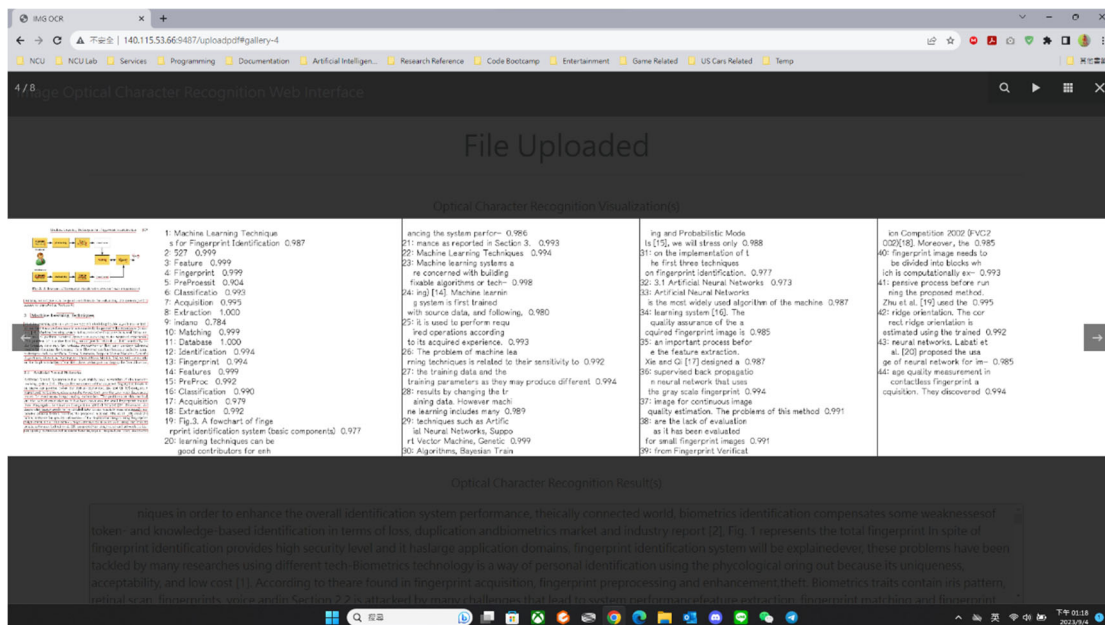
◇ 結果輸出頁面：下半部



- ✧ 輸出視覺化：包含辨識框及輸出結果展示，左右按鍵可變換當前展示頁面



✧ 直接點擊圖片進入相簿模式，可以對圖片進行放大



✧ 完整文字輸出：包含所有頁面辨識出之完整文章

Optical Character Recognition Result(s)

niques in order to enhance the overall identification system performance, theically connected world, biometrics identification compensates some weaknesses of token- and knowledge-based identification in terms of loss, duplication and biometrics market and industry report [2], Fig. 1 represents the total fingerprint. In spite of fingerprint identification provides high security level and it has large application domains, fingerprint identification system will be explained ever, these problems have been tackled by many researches using different tech- Biometrics technology is a way of personal identification using the phycological oring out because its uniqueness, acceptability, and low cost [1]. According to theare found in fingerprint acquisition, fingerprint preprocessing and enhancement, theft. Biometrics traits contain iris pattern, retinal scan, fingerprints, voice and in Section 2.2 is attacked by many challenges that lead to system performance feature extraction, fingerprint matching and fingerprint

✧ 文章打包功能：包含 TXT 文字檔下載、API 串接用 JSON 檔案下載，以及重新開始新一輪識別

Download TXT

Download JSON

New Process