EXTRACTING INFORMATION FROM ID CARD IMAGES USING OCR COMBINED WITH YOLO AND ATTENTION

Nguyen Phuong Lan

University of Information Technology, Vietnam National University Ho Chi Minh City

What ?

This study focuses on developing an efficient system for extracting information from ID cards using YOLO for detection, OCR for text recognition, and an attention mechanism to enhance accuracy. Vietnamese text processing poses unique challenges due to diacritical marks and complex characters. By integrating VietOCR with YOLO and effective preprocessing techniques, the system ensures high precision in extracting both handwritten and printed text, making it suitable for real-world applications in digital identity verification and document automation.

Why ?

Recognizing text from ID cards, including CMND and CCCD, is a critical task in digital transformation, where AI and computer vision play a key role. Automating this process enhances efficiency, accuracy, and security in various sectors such as government administration, finance, and public services. A reliable system reduces manual effort, minimizes errors, and speeds up identity verification, paving the way for seamless integration into smart city initiatives and digital ecosystems.

Overview

The front side of a 9-digit identity card and the front side of a non-chip citizen identification card.

Figure 1: Non-chip citizen identification card

Figure 2: A 9-digit identity card

Training

Figure 3: Information Extraction Pipeline

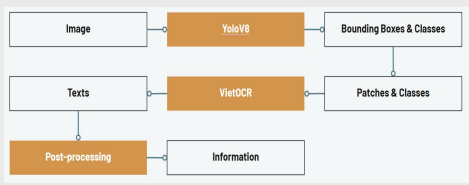
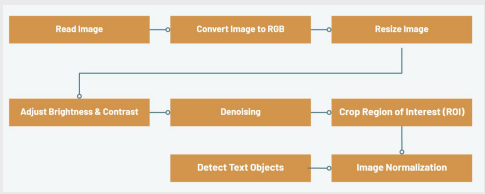


Figure 4: Processing Workflow



Output

A JSON object contains extracted personal information fields.

```
{
  "typeID": "Card type (ID card, Citizen ID, Passport, etc.)",
  "id": "ID card or Citizen ID number",
  "name": "Full name",
  "gender": "Gender",
  "birth_day": "Date of birth",
  "expire_date": "Expiration date",
  "issue_date": "Issue date",
  "issue_place": "Place of issue",
  "nationality": "Nationality",
  "origin_location": "Place of origin",
  "recent_location": "Place of residence"
}
```

Description

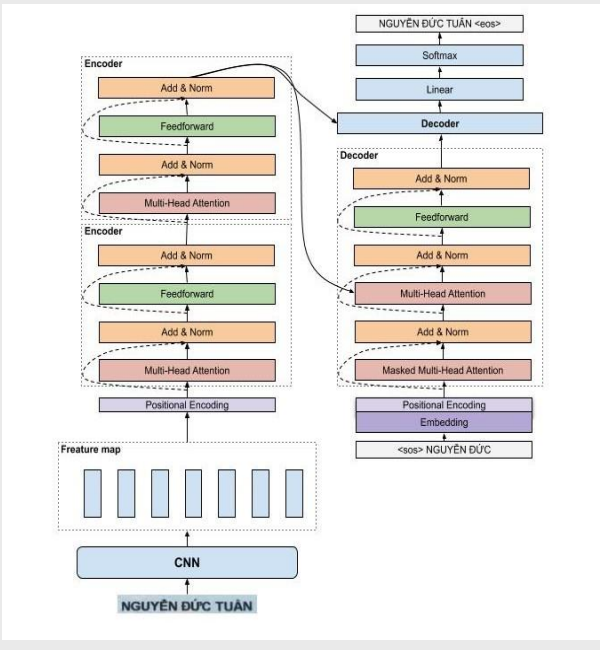
Extracting information from ID cards is essential for **automated identity verification**, streamlining processes in **finance, administration, and public services**. This study integrates **YOLO** for detecting key regions, **OCR** for text extraction, and an **attention mechanism** to improve accuracy, particularly for Vietnamese text with complex diacritical marks. The combination of these techniques ensures **high precision, efficiency, and adaptability** in real-world applications.

1. Yolo

YOLOv8 enhances ID card detection by accurately identifying key regions (photo, ID number, name, etc.), ensuring high-speed and precise localization for OCR.

2.Viet OCR

Vietnamese text. Its ability to process diacritical marks and complex characters ensures accurate recognition. Combined with preprocessing techniques like image binarization and noise reduction, VietOCR improves OCR accuracy, making text extraction more reliable for real-world applications.



3.Evaluate

$$Bleu\ (N) = Brevity\ Penalty \cdot Geometric\ Average\ Precision\ Scores\ (N)$$

BLEU (Bilingual Evaluation Understudy) is used to assess the accuracy of **VietOCR** in extracting text from ID cards. By comparing the recognized text with ground truth references, **BLEU** quantifies how closely the extracted information matches the actual content. A higher **BLEU** score indicates better **OCR** performance, reflecting precise character recognition and minimal errors in handling Vietnamese diacritics and complex text structures.

NII

Nguyen Phuong Lan – University of Information Technology, Vietnam National University Ho Chi Minh City

TEL : 0903975939 Email : lannp.19@grad.uit.edu.vn