Document Digitization Using YOLOv8 and Conservative Attention-Based TrOCR

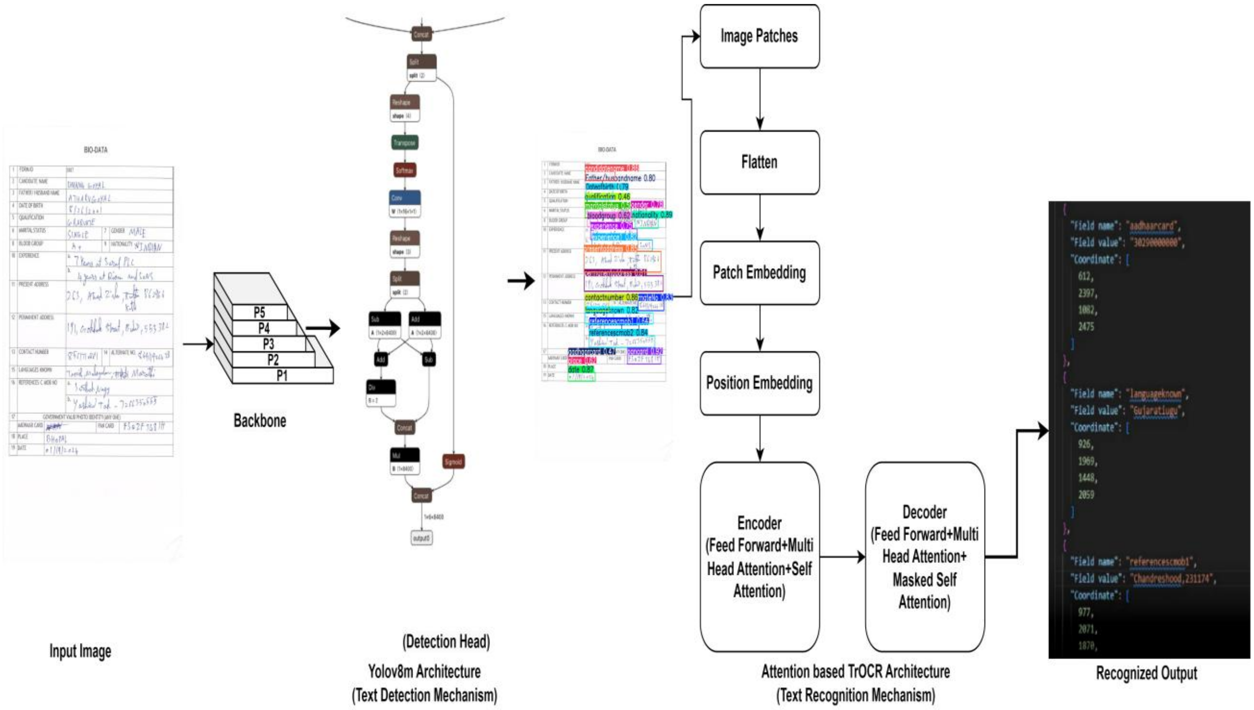
1. Team Details

- **Team Name:** Innovators
- **Team Leader Name:** Dr.M.Senthilarasi
- **Team Leader Institution and Email:** Thiagarajar College of Engineering, Madurai
- **Rest of the Team Members:** Dr.P.Uma Maheswari, Vimal Raj Nagarajan, Iyyappan Ramesh
- **Team Website URL (if any):** Nil
- **Affiliations:** Velammal College of Engineering and Technology, Velammal College of Engineering and Technology, Thiagarajar College of Engineering
- **Link to Codes/Executables:** https://vimal-afk.github.io/dehado/
- **Link to Enhancement Results:** Nil

2. Contribution Details

- **Title of the Contribution:** Document Digitization Using YOLOv8 and Attention-Based TrOCR
- **General Method Description:** This research introduces a two-phase deep learning framework for precise and efficient document digitization. The initial phase utilizes YOLOv8, an advanced object detection model, to identify and extract text areas from scanned or captured documents, even in intricate handwritten documents. During the second phase, the extracted text areas are input into a conservative self-attention driven TrOCR (Transformer-based Optical Character Recognition) model that leverages Vision Transformers and sequence-to-sequence learning for superior text recognition accuracy. The combination of YOLOv8 for accurate handwritten text field detection and TrOCR for strong recognition provides a complete solution that efficiently transforms document images into editable and searchable text. This approach accommodates multilingual and handwritten text, rendering it ideal for uses in archival digitization, mobile scanning, and automated document processing systems.

- **Representative Image/Diagram of the Method:**



```
{
  "Field name": "aadhaarcard",
  "Field value": "30290000000",
  "Coordinate": [
    612,
    2397,
    1082,
    2475
  ]
},
{
  "Field name": "languageknown",
  "Field value": "Gujaratiugu",
  "Coordinate": [
    926,
    1969,
    1448,
    2059
  ]
},
{
  "Field name": "referencescmob1",
  "Field value": "Chandreshood,231174",
  "Coordinate": [
    977,
    2071,
    1870,
```

- **Loss Function:** CIoU and DFL (Bounding Box), Cross Entropy (Class) (For Detection)
- **Testing of Previously Published Methods:** Yes

  Recognition : pytesseract model, DTrOCRLMHeadModel
- **Use of Extra Data:** No
- **Other Methods and Baselines Tested:** Nil

3. Global Method Description

- **Total Method Complexity (all stages):**

  Time complexity

  Text Detection : Training 7.5ms(avg/image), Testing 5.6ms(avg/image)

  Text Recognition : Training 4.8ms(avg/image), Testing 1.5ms(avg/image)
- **Pre-trained or External Models Used:** No pre-trained models used.
- **Additional Data Used:** No additional data used
- **Training Description:**
  - Framework: Modified Yolov8 for text field detection and Conservative Self Attention based TrOCR model for text recognition

- Hardware: No
- Input: Handwritten application forms (2000)
- Training Data: 1200 (Remaining kept it for validation)
- Augmentation: Nil
- Optimizer: AdamW (Modified Yolov8m)
- Learning Rate: 0.0001(Modified Yolov8m) 1e-5 (Conservative Self Attention TrOCR Model)
- Epochs: 10(Modified Yolov8m) 5 (Conservative Self Attention TrOCR Model)

- **Testing Description:**
  - We can validate the results directly using the proposed hybrid trained model.
- **Quantitative and Qualitative Advantages:**
  - parameter count: 25.4M (Text Detection), 83.5(Text Recognition)
- **Results of Comparison to Other Approaches:**

Model	WER	CER
pytesseract model	0.947	0.721
DTrOCRLMHeadModel	0.888	0.472
Proposed (Validation)	0.319	0.201

- **Results on Other Benchmarks:** Nil
- **Novelty and Prior Publication:** 1. Proposal of modified yolo version infused with transformer mechanism for text detection
  2. Proposal of Conservative Self Attention based TrOCR model for text recognition

4. Technical Table

Dataset	Accuracy	F1 score	WER	CER
Training	75%	73%	0.354	0.185
Validation	72%	69%	0.319	0.201

5. Competition Particularities

- **Any particularities of the solution for this competition compared to other challenges:** Nil

References:

- Robert Turnbull; Evelyn Mannix; "Detecting and Recognizing Characters in Greek Papyri with YOLOv8, DeiT and SimCLR", ARXIV-CS.CV, 2024.
- Mejdl S. Safran; Abdulmalik Alajmi; Sultan Alfarhood; "Efficient Multistage License Plate Detection and Recognition Using YOLOv8 and CNN for Smart Parking Systems", JOURNAL OF SENSORS, 2024.
- Hanif Fakhrurroja; Dita Pramesti; Abdul Rofi Hidayatullah; Ahda Arif Fashihullisan; Harry Bangkit; N. Ismail; "Automated License Plate Detection and Recognition Using YOLOv8 and OCR With Tello Drone Camera", 2023 INTERNATIONAL CONFERENCE ON COMPUTER, CONTROL, 2023.
- Minghao Li; Tengchao Lv; Jingye Chen; Lei Cui; Yijuan Lu; Dinei Florencio; Cha Zhang; Zhoujun Li; Furu Wei; "TrOCR: Transformer-Based Optical Character Recognition with Pre-trained Models", AAAI, 2023.
- Hongkuan Zhang; Edward Whittaker; Ikuo Kitagishi; "Extending TrOCR for Text Localization-Free OCR of Full-Page Scanned Receipt Images", ARXIV-CS.CL, 2022.
- Tongkun Guan; Chaochen Gu; Jingzheng Tu; Xue Yang; Qi Feng; Yudi Zhao; Wei Shen; "Self-Supervised Implicit Glyph Attention for Text Recognition", CVPR, 2023.