Report – ID Document Line Extraction System

Task Code: CVCY000

Position: Computer Vision Engineer

Objective: Develop a pipeline to extract individual lines of text from scanned Egyptian ID cards (any orientation) as cropped images for downstream OCR.

1. Dataset Choice

Selected two of the Egyptian ID datasets to ensure domain relevance and robustness:

- Card Finder (5,119 images) – used for ID detection. It includes diverse real-world conditions such as lighting variations and occlusions, making the detector robust under practical settings.
- RE6 (7,449 images) – used for angle detection and line segmentation. Its larger scale and consistency provide broader coverage, supporting reliable orientation correction and fine-grained line extraction.
- Both datasets are annotated in YOLO format, enabling smooth integration into detection pipelines and reproducibility.

Additional Preprocessing:

- During dataset analysis, some images were found to be unlabeled yet contained ID content. These images were filtered out to maintain annotation consistency.
- The dataset included both polygon and bounding box annotations. All polygon annotations were converted to bounding boxes to standardize the training format for detection models.

2. System Pipeline & Model Architecture

2.1 ID Card Detection

YOLOv8n was chosen for ID card boundary detection due to its strong performance on small and thin objects like ID text lines. Unlike older YOLO versions (v3–v7), YOLOv8 is anchor-free, reducing the need for manual anchor tuning and improving recall for small objects. It also offers faster training, built-in augmentations, and lightweight inference suitable for deployment.

Light augmentations were applied to improve generalization without distorting ID structure. Hyperparameter tuning refined model performance.

2.2 Rotation Correction

Rotation correction was trained using the RE6 dataset. Since the dataset predominantly contained images with 0° orientation, synthetic rotations and extensive augmentations were applied to simulate real-world variations. The model used MobileNetV3-Small as the backbone, with a regression head predicting the 2D unit vector [sin(θ), cos(θ)] representing rotation. This approach allows precise angular regression and stable training using Smooth L1 Loss, without discretizing angles into predefined classes.

2.3 Line Segmentation

Line segmentation was performed using YOLOv8n in detection mode. Hyperparameters were optimized through grid search, and early stopping was applied to avoid overfitting. This approach was preferred over full segmentation networks, as object detection was sufficient to extract text lines accurately.

3. Loss Functions

- YOLOv8 uses CIoU loss for bounding box regression, which accounts for overlap, center distance, and aspect ratio—important for thin ID text lines.
- Objectness loss: Binary cross-entropy ensures the model distinguishes text lines from background.
- Classification loss: Minimal in this case, since the task involves a single class (text line)

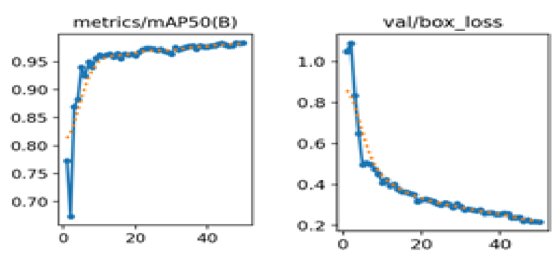
Angle Rotation:

- The model predicts image rotation as a 2D unit vector [sin(θ), cos(θ)]. **Smooth L1 Loss** is used to regress this vector:
    - **Reason for choice:** It is robust to outliers, provides stable gradients, and effectively penalizes deviations in both vector components, ensuring accurate angular predictions even with noisy or ambiguous rotations.

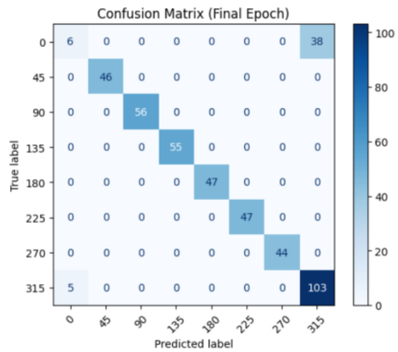
4. Performance Analysis and Evaluation

For ID Detection

The training and validation loss curves show steady convergence, indicating stable learning and good generalization. The YOLO model achieves high results, reflecting nearly perfect detection accuracy for id bounds. Together, the low loss and high mAP demonstrate that the model effectively detects ID bounds across varied orientations and augmented conditions.



For Rotation Correction



Metrics were selected per subtask:

- ID Detection and Line Segmentation → Precision, Recall, mAP@0.5, mAP@0.5–0.95
- Rotation Correction → classification accuracy, weighted F1-score and angular error.

Key results:

- ID Detection (Card Finder) → Best mAP@0.5 = 0.9827 (valid), mAP@0.5: 0.9772 (test)
- Rotation Correction (RE6) → Train/Val Loss = 0.0073/0.0002, Acc = 88.4%, F1 = 0.855, Train Mean Angular Error = 1.36° ± 1.14°.
- Line Segmentation (RE6) → Validation mAP50: 0.9950, Validation Precision: 1.0000, Validation Recall: 1.0000

5. Limitations

Rotation Correction

- Misclassifies some correctly oriented (0°) images, showing weak handling of near-zero cases.
- A large number of images are incorrectly corrected to **180° rotations on test set**, indicating a systemic bias in orientation classification.
- Mean Angular Error is relatively high (**7.92° ± 4.54°**).
- Struggles with subtle, non-zero rotations that need finer adjustment.

ID Detection

- Model lacks diverse negative samples; needs more varied background data to handle real-world non-ID regions.