

Statement of Purpose

Shubham Agrawal 22110249

shubham.agrawal@iitgn.ac.in | IIT Gandhinagar

Preliminary Approach: OMR Sheet Bubble Detection

Tools and Technologies

1. **Programming Language:** Python
2. **Libraries and Frameworks:**
 - AI & Machine Learning: TensorFlow, PyTorch, or Scikit-learn
 - Image Processing: OpenCV, PIL
 - Data Handling: Pandas, NumPy
3. **Visualization Tools:** Matplotlib, Seaborn, Dash, Streamlit
4. **Mobile Integration:** TensorFlow Lite, Flutter/Dart for app development

AI Models and Methodologies

1. **Data Preprocessing:**
 - Analyze datasets (e.g., OMR sheet images from the provided GitHub repository).
 - Preprocess images (grayscale conversion, binarization, noise reduction).
 - Identify and extract key regions like bubble sections and corner markers using image processing.
2. **Bubble Detection:**
 - Develop a supervised learning model (e.g., CNN) to classify bubble presence.
 - Use bounding box detection for precise bubble localization.
 - Augment data by applying transformations like rotations or scaling to improve robustness.

3. Sheet Alignment and Marker Detection:

- Implement algorithms for corner marker detection using OpenCV.
- Develop alignment correction methods to handle misaligned or rotated sheets.

4. Error Handling and Feedback Loop:

- Introduce error correction mechanisms for incomplete or misaligned scans.
- Integrate a human-in-the-loop system for reviewing and correcting errors.

Integration and Customization

- Convert the trained model to TensorFlow Lite for efficient deployment.
- Build an app interface that:
 - Accepts scanned OMR sheet PDFs.
 - Detects bubbles and outputs responses in CSV format.

Visualization and Analysis

- Develop tools for error visualization:
 - Heatmaps for misdetections
 - Overlay bounding boxes on images to show bubble detections
- Create dashboards to display accuracy, error rates, and processing statistics.

Reporting and Metrics

- Evaluate and track:
 - Bubble detection accuracy
 - Alignment precision
 - Error rate
- Monitor system efficiency and robustness across diverse templates.