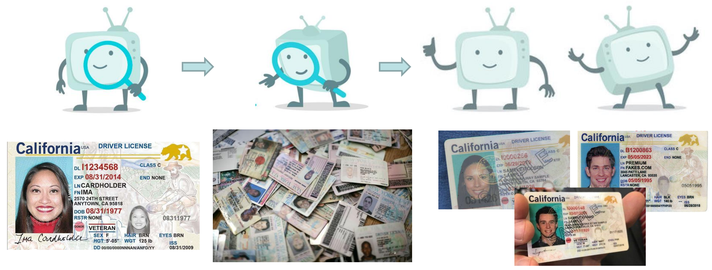
Improved ID-Matching Tool

Inderjot Saggu, 2nd Year Master's in MLDS

Intern-Machine Learning, Document Tooling Team

Primary Project

- Designing an improved ID matching tool for automating document onboarding process and increasing efficiency of modellers.



Query Image **Test Images** **Output Images**

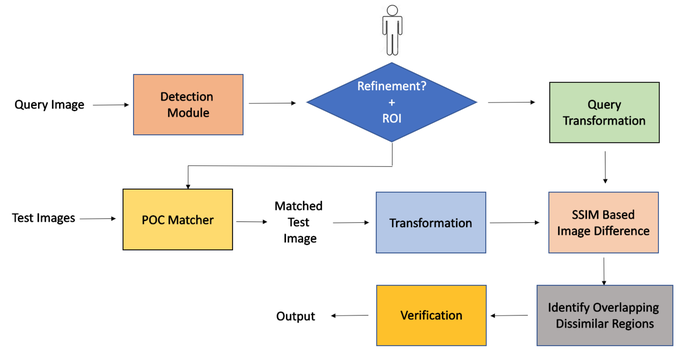
Problem

- Legacy Matcher tool couldn't distinguish between sub categories of ID documents, eg.

- Modellers have to manually separate the above

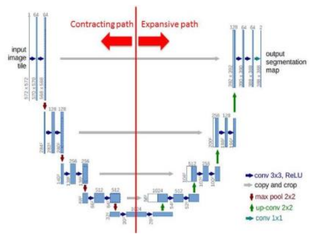
- This requires detection of minute differences across ID-documents in natural setting.
- **Primary Goal**: Significantly reduce FPs while retaining most of the TPs
- **Secondary Goal**: Remove/Reduce Manual Intervention

Improved Matcher Tool



Detection Module

- U-Net based ID segmentation followed by convex-hull, k-means clustering and IoU optimization.
- Achieved a median distance of 7 pixels for four corner detection and IoU of 0.98 for segmentation

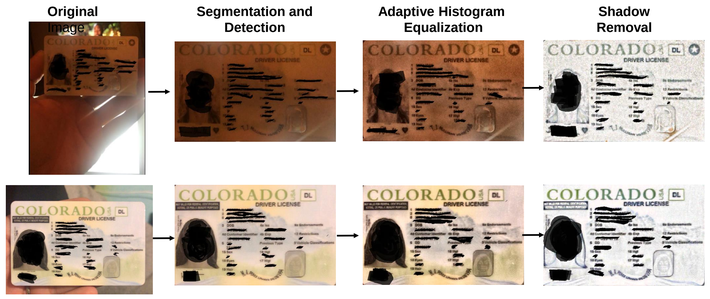


Architecture

Results

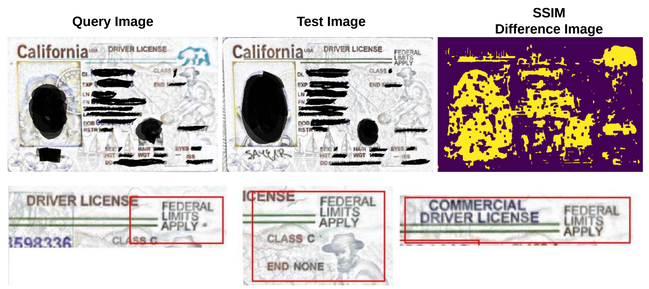
- Refinement module allows user to select four-corners in case of incorrect detection
- RoI requires user to pre-select regions that are important (distinguishing factors)

Transformation Module



- Homography computation, CLACHE, histogram matching, shadow removal

SSIM Image Difference



- Improved Matcher Tool rejects 92% of FPs over legacy tool while retaining over 80% of TPs
- Designed easy-to-use UI, handed finished tool to the Document Modelling Team

Acknowledgments

- Thanks to Tim Hughes, Steve Dyrdahl, Stephen Ritter, scrum team of Emily, Bobby and Joe, and UCSD for a great internship experience