AirSupport

Vital Monitor and ID Detection through Machine Vision for Improving EMS Communication Efficiency

Group 11:
Robert Huang

Clinical Mentors:
Dr. Nick Dalesio
Dr. Laeben Lester

Computer Vision Mentor:
Dr. Mathias Unberath

JOHNS HOPKINS UNIVERSITY

AirSupport

Brief Problem Summary

Smart Glasses in healthcare to **record and stream** live feed [1].

Two Objectives:

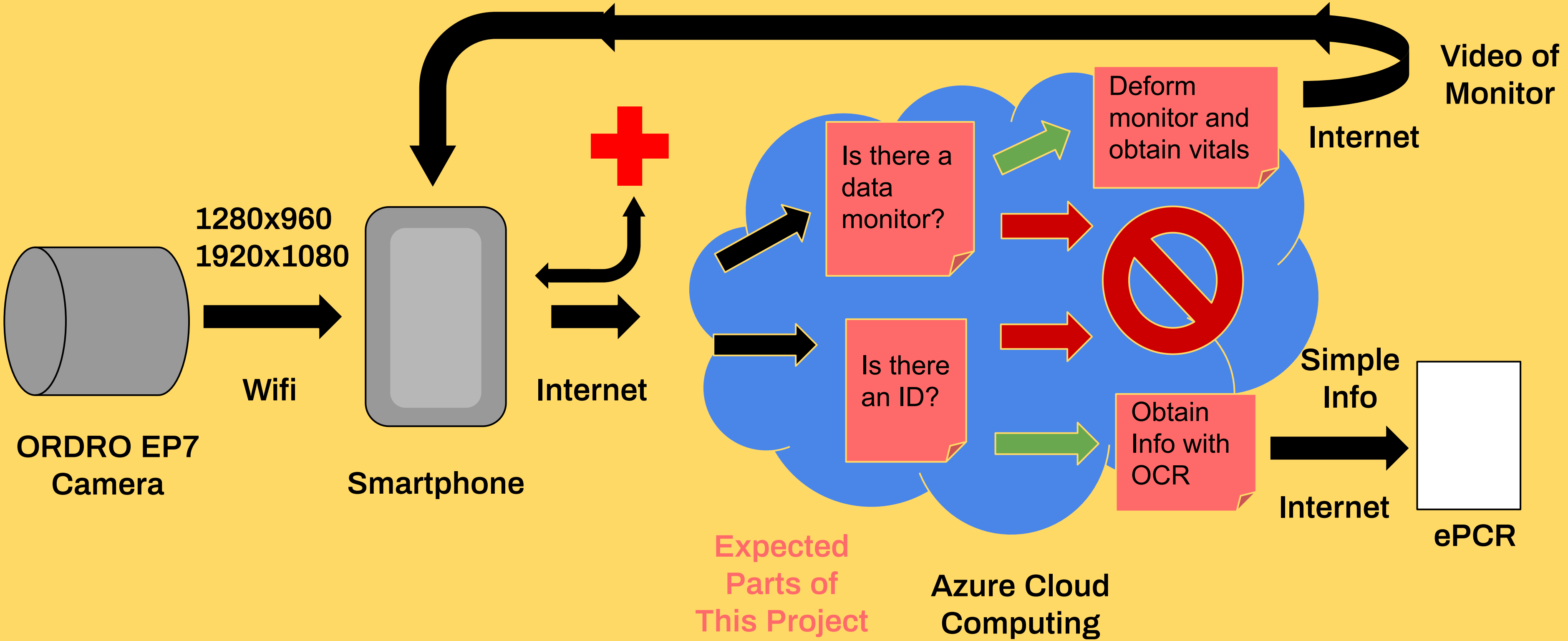
- **Objective 1:** Automatically record and insert information into a digital medical note, from Smart Glasses Feed
  - On the field, streamlining documentation of simple information **can save 1-10min**, which can be spent on healthcare. [2]

- **Objective 2:** Provide View of Data Monitors Remotely with AI and Computer Vision, from Smart Glasses Feed
  - On the field, EMTs obtain remote professional advice. AI can make physicians **2-3x more confident** and **give treatments 2-3x faster**. [3]

[1] Vuzix Corporation, 2020
[2] Lester, 2021
[3] Schaer et al., 2016

AirSupport

Technical Approach - Overall Workflow Review



JOHNS HOPKINS UNIVERSITY

Previous Technical Approach - Objective 1: Detection and Information Extraction of IDs

- Steps:
  1. **Detect the presence of an ID**
     - YOLO Deep Learning Framework
     - Dataset: Images of many state’s Driver's’ License

  2. **Read and categorize information on the ID.**
     - OCR using deep learning technique by Wojna, Zbigniew, et al, 2017. Combine with **Generic Text Parsing** to sort information.

AirSupport

Current Technical Approach - Objective 1: Detection and Information Extraction of IDs

- Steps:
  1. **Detect the presence of an ID**
     - YOLO Deep Learning Framework
     - Dataset: Images of many state's Driver's' License
  2. **Detect the spatial orientation of an ID and deform**
     - Hough Transform Edge Detection
  3. **Read and categorize information on the ID.**
     - Tesseract OCR w/ Python and **Generic Text Parsing** to sort information.

AirSupport

Previous Technical Approach - Objective 2: Detection and Deformation of Data Monitors

- Steps:
  1. **Detect the display** of a vital signs monitor or ultrasound image.
     - YOLO Deep Learning Framework for Detection
     - Dataset: Images of the device/monitor from multiple angles.

  2. Capture video when a display monitor is on camera and **preprocess and deform video** such that the monitor appears head-on.
     - Blob Feature registration and deformation.

  3. **Read vitals signs data.**
     - Optical Character Recognition (OCR) using deep learning technique by Wojna, Zbigniew, et al, 2017.

JOHNS HOPKINS UNIVERSITY

Current Technical Approach - Objective 2: Detection and Deformation of Data Monitors

- Steps:
  1. **Detect the display** of a vital signs monitor or ultrasound image.
     - YOLO Deep Learning Framework for Detection
     - Dataset: Images of the device/monitor from multiple angles.

  2. Capture video when a display monitor is on camera and **preprocess and deform video** such that the monitor appears head-on.
     - Blob Feature registration and deformation.

  3. **Read vitals signs data.**
     - Tesseract OCR w/ Python

AirSupport

Deliverables Progress and Changes

	Activity	Deliverable	Expected Completion
Min	Obtain datasets of IDs, Medical Device with Monitors, ~~and characters~~	Dataset of IDs, Medical Devices, ~~and characters~~	2/19 ✓
	Code the YOLO framework for detection of IDs. Train on ID dataset. Code and assess deformation with Hough. Code the OCR framework. ~~Train with character dataset~~. Test OCR with camera data. Code Generic Text Parsing Code and assess its performance with a constructed ground truth dataset. Combine code to read IDs.	Documentation of overall code and performances of each section.	3/15 ✓
Expected	Duplicate and Train YOLO on Medical Device dataset.	Documentation of code and performance of YOLO on Devices.	3/20
	Code Blob Feature Detection and deformation algorithm. Qualitatively assess how many deformations are acceptable.	Documentation of Blob Feature Algorithm and its performance.	3/30
Max	Incorporate the algorithms into the current overall workflow. Assess their performance in the workflow using a written test.	Documentation of incorporation. Testing procedures for performance assessment. Results of the tests.	5/1

Documentation

Datasets:

- Stored on Google Drive due to size
- Document describing dataset stored in the same folder (see wiki Deliverables or Milestones)

Code Documentation:

- Stored on Google Drive to interact with datasets
- Document describing architecture and results stored in the same folder (see wiki Deliverables or Milestones)

[4] Gould et al, 2009

[5] Shaw, 2020

License Dataset


[4]

[5]

Dataset Characteristics:

- 23834 Data Augmented Images of ID
  - USA IDs transformed and brightness-adjusted and pasted onto empty backgrounds
- 11723 Natural Images of ID
  - 22 of USA IDs
  - 11701 of non-USA IDs
- 7209 Negative Images
  - 700 Images of empty backgrounds
  - 6509 Images of Potential Confusers: Smartphones, Billboards, Books, etc.

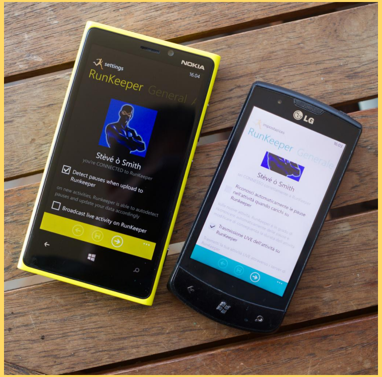
ID transformed and 'pasted' on background.

[4] Gould et al, 2009

[5] Shaw, 2020

License Dataset

Natural Images of Non-US IDs [6]



Negative Images of potential confusers [7]

[6] Arlazarov et al., 2019

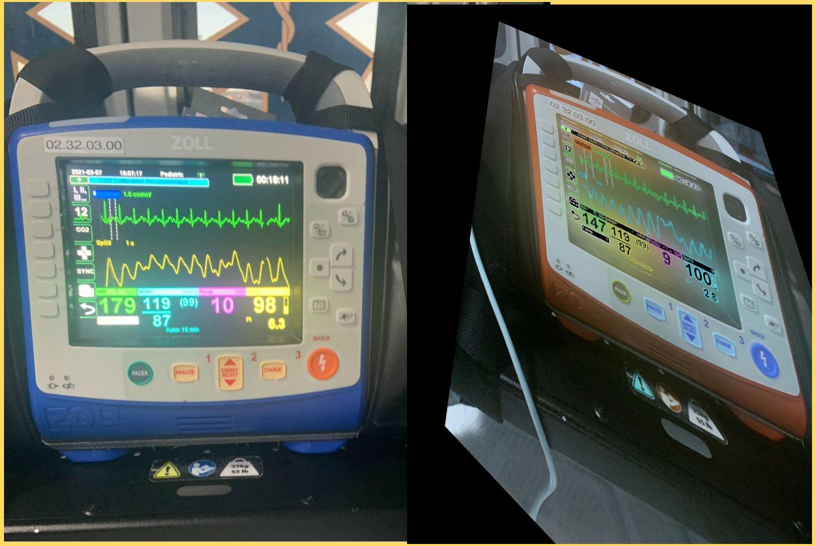
[7] Kuznetsova, 2020

AirSupport

Devices Dataset

Dataset Characteristics:

- 3240 Positive Zoll Images
  - Homography, Color, and Brightness Augmented
- 7209 Negative Images
  - 700 Images of empty backgrounds
  - 6509 Images of Potential Confusers: Smartphones, Billboards, Books, etc.



JOHNS HOPKINS UNIVERSITY

Overall Process

→ YOLOv3

→ Deform

→ OCR

→ Text Parser

[Name: 'Specimen',
ID#: XX518518,
Sex: 'F']

AirSupport

Design Specifications and Performance Objective 1: Detection and Information Extraction of IDs

Step	Target Accuracy (or Loss)	Recorded Accuracy (or Loss)	Target Speed	Recorded Speed
YOLOv3	96%	99%	8ms	30ms
Deformation	<0.01	0.009	2ms	1ms
OCR	96%	96%	3ms	6ms
Text Parsing	96%	99%	3ms	<1ms
Overall	88.8%	94%	16ms	37ms

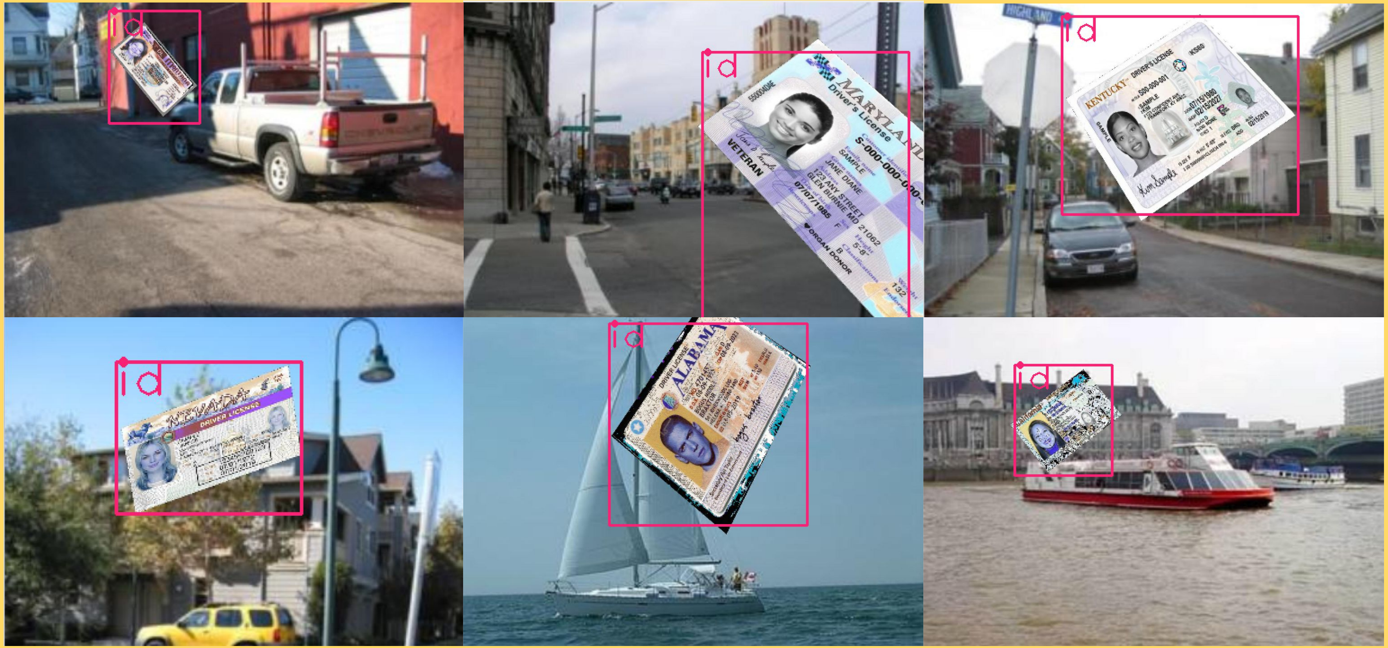
JOHNS HOPKINS UNIVERSITY

AirSupport

YOLOv3 Post-Training Performance

- Average Loss: ~0.03
- Average IOU: ~0.85-0.95
- Average Accuracy: ~0.99 ✓
- Decision Speed: ~30ms ✗
  - Slow for video camera use, but can be run with separate threads for smoothness.

YOLOv3 Post-Training Performance



[4,5]

[4] Gould et al, 2009

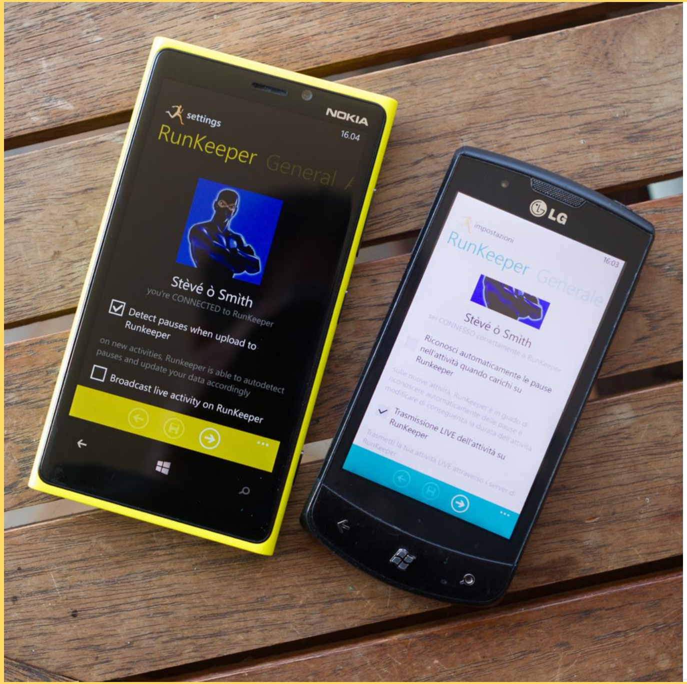
[5] Shaw, 2020

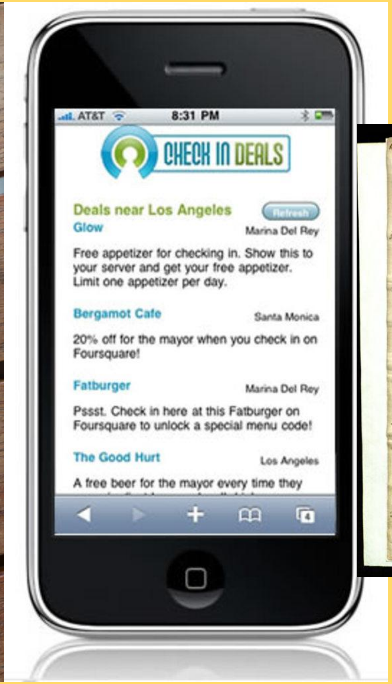
YOLOv3 Post-Training Performance

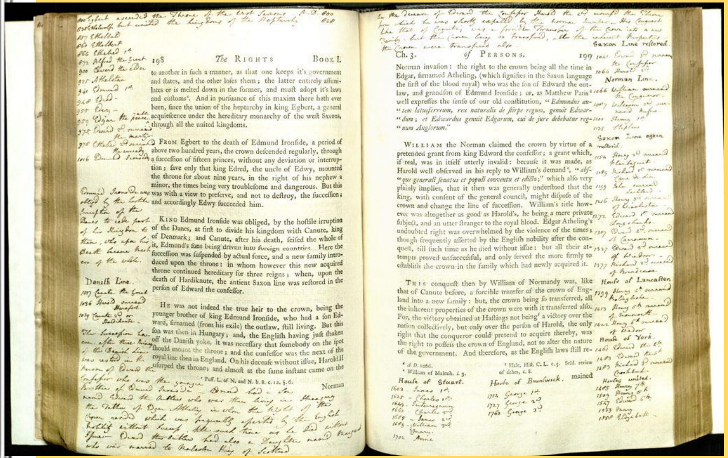
[6]

[6] Arlazarov et al., 2019

YOLOv3 Post-Training Performance







[7]

[7] Kuznetsova, 2020

AirSupport

Deformation Description

- Input: Closely cropped image of ID
- Preprocess to assist Hough Transform
  - Canny Edge Detection
- Edges detected with Hough Transform
- Find contours with certain characteristics indicative of ID:
  - % Area: 75-100% of area
  - # of Corners: 4-5 corners
  - Aspect ratio: 0.5-0.6
- Obtain corners of rectangle and deform.

JOHNS HOPKINS
UNIVERSITY

Deformation Testing Procedure

- Using the manual data augmentation ->, obtain true transforms of 10000 images.
- Transforms:

n = 1	2	3
4	5	6
7	8	9

- Algorithm calculates a transform.
- Loss = $\sum_n (\text{true-predicted})^2$

[4]

[5]

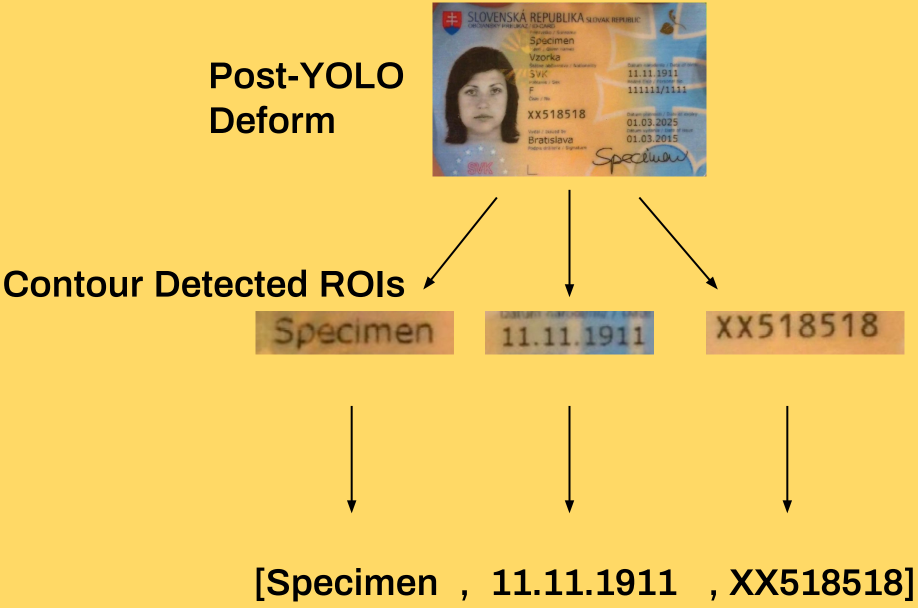
Deformation Performance

- Average Loss: ~0.009 ✓
- Decision Speed: ~1ms ✓

AirSupport

OCR Description

- Preprocess Post-YOLO Deform image for optimal OCR:
  - Binarize, Gaussian Blur, TopHat, BlackHat, Threshold Local
- Find contours that are ‘Regions of Interest’
- Run Tesseract OCR on regions.
- Return list of text detected and their location on the image.



JOHNS HOPKINS UNIVERSITY

AirSupport

OCR Testing Procedure

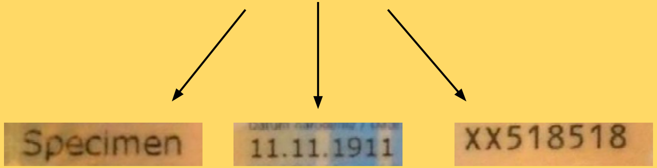
- Images from the smart glasses camera are fed through the YOLO and deformation algorithms.

- 200 ‘Regions of interest’ (AOI) are labeled with their truth in a list.

- The images are fed into the OCR algorithm, and the output of text is compared with truth.
- Accuracy = ROIs correctly detected and read/All

Image

Post-YOLO Deform



ROIs with labels

JOHNS HOPKINS UNIVERSITY

AirSupport

OCR Performance

- Average Accuracy: 96% ✓
- Decision Speed: 6ms ✗
  - One check every # of frames instead of every frame.

Johns Hopkins University

OCR


SLOVENSKA REPUBLIKA
SLOVENSKA REPUBLIKA
SI0VAKRIPIIBUG
SLOVAK REPUBLIC
OBCIANSKY PREUKAZ / ID-CARD
Specimen
Specimen
Vzorka
Vzorka
SVK
SVK
F
F
XX518518
XX518518
Bratislava
Bratislava
11.11.1911
11.11.1911
111111/1111
111111/1111
01.03.2015
01.03.2025
01.03.2015
01.03.2015
5pooww
SVK

AirSupport

Text Parsing Description

- Using heuristics and regular expressions, categorize received list of data.
- Locations of ambiguous items, such as Eye and Hair are looked on an ‘Atlas of state IDs’ to categorize.

[‘Nevada’,
‘5-02’,
‘211’,
‘BRO’,
‘BLK’]

→

{State: ‘Nevada’,
Height: ‘5-02’,
Weight: ‘211’,
Eye: ‘BLK’,
Hair: ‘BRO’}

→ Nevada atlas check

Heuristics

- Sex: “F” or “M”
- Dates:
  - Expiry: Date in Future
  - Birth: Date furthest in past
  - Issued: Date in Between
- States:
  - ‘Nevada’,’California’,etc.
- Address:
  - Fill up Two - Three lines
  - ‘St’
- Height:
  - 4/5(')?-(0/1)?#(“)?
- Weight:
  - ###
- Eye: BRO, etc.
- Hair: BRO, BLK, etc.

Johns Hopkins University

AirSupport

Text Parsing Testing Procedure

- 1000 Custom Input lists of potential texts obtained, such as dates, names, states, sex, series of numbers, height, weight.
- Accuracy = Correctly categorized items/All items

Text Parsing Performance

- Average Accuracy: ~99% ✓
- Decision Speed: ~1ms ✓

AirSupport

Dependency Progress

Dependency	Need	Contingency	Status	Planned Deadline	Hard Deadline
ID/Medical Device/ Characters Datasets	For training and assessing algorithms	Begin coding without training.	Currently met	2/19	3/20
Computer/Internet	For coding and communication	Use public computers. Use mobile data.	Currently met	Continuous	Continuous
MDAirSupport Sample Product	For testing for incorporation into overall workflow	Write procedures for other MDAirSupport to test.	Will visit robotics lab.	4/1	5/1
Deep Learning Mentor	For Optimizing Algorithm to reduce computation	Spend time researching optimization	Currently met	Continuous	Continuous
Microsoft Azure or other Cloud Computing	For incorporation into overall workflow. May also require for training.	For training: use public JHU computer.	Have requested: Once algorithms are good.	4/1	5/1

Johns Hopkins University

AirSupport

Current Timeline/Milestones

Milestone	Deadline
Obtain datasets.	2/19
Code YOLO	2/22
Train YOLO on IDs and record results	2/25
Code OCR	2/29
Test OCR w/ camera data and record results	3/1
Code and assess deformation with Hough Transform	3/4
Create Ground Truth Set for Generic Text Parsing	3/5
Code and Assess Generic Text Parsing and record results.	3/15
Combine all individual codes to read IDs	3/15

AirSupport

Current Timeline/Milestones

Milestone	Deadline
Checkpoint Presentation	3/18
Duplicate YOLO and train on Devices Dataset and record results	3/20
Generate Test Set for Blob Feature Deform Algorithm Testing	3/25
Code Blob Feature Deform Algorithm	3/28
Assess Blob Feature Deform Algorithm and record results.	3/30

Johns Hopkins University

AirSupport

Current Timeline/Milestones

Milestone	Deadline
Investigate how to write to ePCR.	4/5
Feed video back into smartphone.	4/10
Incorporate all algorithms into workflow.	4/15
Generate tests to assess the functionality of the application.	4/20
Paper Presentation	4/22
Assess overall application functionality.	5/1
Final Paper	5/4
Final Presentations	5/6

AirSupport

Tasks Performed

Team:

Robert Huang

- Milestone tasks not related to obtaining Medical Devices Dataset or OCR Test data.

Clinical Mentors:

Dr. Nick Dalesio

- Recorded and Provided Medical Devices Dataset

Dr. Laeben Lester

- Recorded IDs with camera for OCR Tests

Computer Vision Mentor:

Dr. Mathias Unberath:

- Advice for Algorithm Optimization

JOHNS HOPKINS UNIVERSITY

Management Plan Progress

- Meetings:
  - Continued Weekly Meetings with MD AirSupport on Friday 9AM
  - Continued Imaging Meeting with Dr. Mathias Unberath and Dr. Russell Taylor on Friday 2PM
- Programs Used:
  - Communication through Email
  - Code on GitHub
  - Writing Reports and Documentation using OneLeaf and Upload onto CIIS Wiki

Increased Reading List

- Redmon, Joseph, et al. “You Only Look Once: Unified, Real-Time Object Detection.” *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, doi:10.1109/cvpr.2016.91.
- Shaifee, Mohammad Javad, et al. “Fast YOLO: A Fast You Only Look Once System for Real-Time Embedded Object Detection in Video.” *Journal of Computational Vision and Imaging Systems*, vol. 3, no. 1, 2017, doi:10.15353/vsnl.v3i1.171.
- Wojna, Zbigniew, et al. “Attention-Based Extraction of Structured Information from Street View Imagery.” *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, 2017, doi:10.1109/icdar.2017.143.
- Llados, J., et al. “ICAR: Identity Card Automatic Reader.” *Proceedings of Sixth International Conference on Document Analysis and Recognition*, doi:10.1109/icdar.2001.953834.
- Mikolajczyk K, Schmid C. Scale & affine invariant interest point detectors. *International Journal on Computer Vision*. 2004;60:63. doi: 10.1023/B:VISI.0000027790.02288.f2.
- V. V. Arlazarov, K. Bulatov, T. Chernov and V. L. Arlazarov, “A dataset for identity documents analysis and recognition on mobile devices in video stream”, *Comput. Opt.*, vol. 43, no. 5, pp. 818-824, 2019.
- Y. S. Chernyshova, A. V. Sheshkus and V. V. Arlazarov, “Two-Step CNN Framework for Text Line Recognition in Camera-Captured Images,” *IEEE Access*, vol. 8, pp. 32587-32600, 2020, doi: 10.1109/ACCESS.2020.2974051.
- S. Gould, R. Fulton, D. Koller. Decomposing a Scene into Geometric and Semantically Consistent Regions. Proceedings International Conference on Computer Vision (ICCV), 2009.

AirSupport

References

[1] Vuzix Corporation. Vuzix Corporation, 2020, *VUZIX SMART GLASSES AT THE CHI MEI MEDICAL CENTER, TAIWAN*, ss-usa.s3.amazonaws.com/c/308483104/media/21105f5a523ce21ce43889049199725/Vuzix-Chi-Mei-Medical-Case-Study-2020.pdf.
[2] Lester, Laeben. “Inquiry into EMS Documentation Times.” 14 Feb. 2021.
[3] Schaer, et al. “Using Smart Glasses in Medical Emergency Situations, a Qualitative Pilot Study.” *2016 IEEE Wireless Health (WH)*, 2016, doi:10.1109/wh.2016.7764556.
[4] Gould, Stephen, Richard Fulton, and Daphne Koller. 2009. “Decomposing a Scene into Geometric and Semantically Consistent Regions.” *2009 IEEE 12th International Conference on Computer Vision*. doi:10.1109/iccv.2009.5459211.
[5] Shaw, Gabbi, and Frank Olito. 2020. “What a Driver's License Looks like in Every State.” *Insider*. Insider. January 21. https://www.insider.com/what-drivers-license-looks-like-in-every-state.
[6] V. V. Arlazarov, K. Bulatov, T. Chernov and V. L. Arlazarov, “A dataset for identity documents analysis and recognition on mobile devices in video stream”, *Comput. Opt.*, vol. 43, no. 5, pp. 818-824, 2019.
[7] Kuznetsova, Alina, Hassan Rom, Neil Alldrin, Jasper Uijlings, Ivan Krasin, Jordi Pont-Tuset, Shahab Kamali, et al. 2020. “The Open Images Dataset V4.” *International Journal of Computer Vision* 128 (7): 1956–81. doi:10.1007/s11263-020-01316-z.

JOHNS HOPKINS UNIVERSITY

Thank You!
Questions?