CT Yolo-AR: Real-time Medical Imaging Process Analytics

Andrew Brown, MD, MBA
St. Michael's Hospital,
Unity Health Toronto
Andrew.Brown@unityhealh.to

Tenzin Migmar
McMaster University,
DSI, University of Toronto
tenzin.migmarr@gmail.com

Wael Sulais
KAUST Academy,
DSI, University of Toronto
waelsulais@mailbox.org

Abstract

We introduce CT YOLO-AR, a two-pronged system combining YOLOv8 segmentation and ResNet18 for Action Recognition to capture process analytics about CT scans. The Action Recognition Net identifies and classifies 6 key events during CT scans, with each video frame labeled accordingly. To ensure patient anonymity, video streams are not saved; instead YOLOv8 generates masks for reference and verification of the Action Recognition model outputs.

1. Introduction

Diagnostic medical imaging services require significant resources, including equipment, materials and labor. Proper coordination of these resources positively impacts the productivity of the radiology department, which, in turn, influences access to care, costs, and quality [1]–[4]. Thus, the availability of near-real-time productivity measurements may improve department operations by providing insight into system bottlenecks.

1.1. Motivation

Medical imaging wait times in Ontario are severely backlogged, especially for low-priority patients. In June 2024, only 58% of adult patients received a scan within their target time [5]. St Michael's Hospital, our pilot project site, shared the same dire statistic, making it an ideal location for a pilot.

1.2. Prior Art

Previous work on the matter has included a variety of methods. Of which, two come into mind.

The first method is to rely on technologist reporting which is cost-effective. However, due to human error, it is likely the logs would be incomplete or contain incorrect information.

The other, more reliable method is to rely on external reporting which typically takes the form of hiring an individual to record the times. For example, they perform reporting was done via a video surveillance system after the fact in [3].

As such, our approach should be something that bridges the gap between being cost-effective while also being accurate.

2. Methodology

In order to accomplish this goal, we will utilize Computer Vision to automate the data collection process. In particular, we use a mixture of Image Segmentation and Action Recognition,

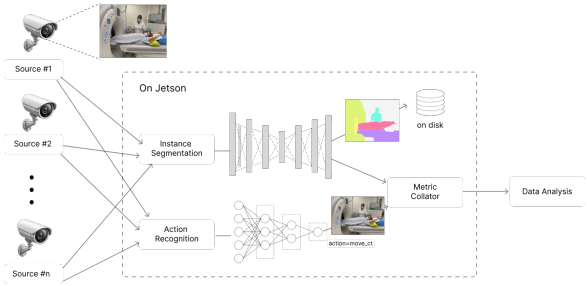


Figure 1: Pipeline architecture

2.1. Dataset Preperation

For image segmentation, we had to collect over 1400 images of CT scan machines from the internet, including some from St. Michael's, as well as manually label them ourselves inside of Roboflow [6].

In particular, we labeled 7 classes of interest: `ct-scanner`, `ct-ring`, `patient`, `scanner-bed`, `stretcher`, and `wheelchair`. The reason we picked these classes in particular was to facilitate automated capturing of certain data that could provide insights,

As for action recognition, the YOLO-AR team recorded around 12 videos of simulated CT scan each around 2 minutes. These were then labelled manually by one team member.

2.2. Image Segmentation

For image segmentation, we chose YOLOv8-m [7] as our model due to it striking a good balance between inference quality and runtime performance. In particular, we fine-tuned a pre-trained MS-COCO [8] checkpoint on our Roboflow [6] dataset with following steps.

From Roboflow, we applied preprocessing to resize and orient the images. Afterwards, we applied augmentations: Flip; Crop; $(\pm 15)°$ Shear, Rotation, and Exposure; $(\pm 25)°$ Hue.

2.3. Action Recognition

We trained a 2D convolution based ResNet18 model on Action Recognition following [9] since this architecture proves to be effective on edge-computing devices.



Figure 2: Output labels of the ResNet18 model

The model was trained for 30 epochs using Nvidia's TAO toolkit on the same rig descibed in Section 2.2 on one GTX Titan X for approximately 1 hour.

3. Results

Instance Class	Box Accuracy	mAP@50 (%)
ct-scanner	98.0	99.7
ct-ring	90.9	94.4
patient	74.3	88.8
scanner-bed	89.9	96.2
stretcher	70.3	75.0
tech	85.1	92.2
wheelchair	79.7	99.5
all	86.9	89.4

Table 1: Image Segmentation Metrics

As Table 1 shows, the image segmentation model performs well on the boudning box task as evident by the "Box Accuracy" metric. It is important to note that both the `wheelchair` and `stretcher` classes are severely underrepresented in the our dataset.

Interestingly, the model does not perform well on patients. We hypothesis this is due to the fact standing patients are underrepresented in the dataest. This manifests as the model picking up standing patients as techs till the lie on the scanner bed.

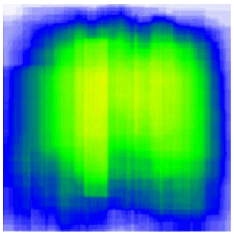


Figure 4: Bounding box heatmap for tech.

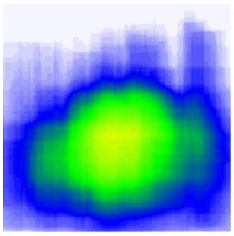


Figure 5: Patient Bounding box heatmap.

The model overfits on the fact patients are always lying down on the scanner bed which is corroborated by distributions in Figure 4 and Figure 5. One way we can counter this is by applying translation ("image shift") augmentations to the training set of patients.

Action Class	Accuracy (%)
open_door	100.0
p_enter_room	91.67
lie_down	75.0
move_ct	91.67
scanning	100.0
remove_blanket	33.33

Table 2: ResNet18 Action Recognition Metrics

In Table 2, we see that the model performs really well on our main actions such as `move_ct` and `scanning`. However, it does not perform on more complex actions such as `remove_blanket`. Overall, the model achieves a mean accuracy of 81.9%.

4. Future Work

The first order of business is increase the sample size in both datasets by incorporating more varied CT scenarios, followed by prescriptive data augmentation techniques that mitigate the observed spatial bias and revisiting model fine-tuning to improve overall performance.

We can extend the system to more than the CT scan room to capture the end-to-end patient journey in similar spirit to [1]. By implement-

ing a multi-object tracking solution that persists in patient identification between cameras interspersed throughout the imaging department, we can obtain a more complete understanding of CT scanning productivity.

Future work could entail merging the instance segmentation and action recognition models into a single, cohesive model in order to reuse the inerent Computer Vision knowledge each model etains rather than running both in parallel to reduce computational overhead and speed-up inferencing.

5. Conclusion

We have created an automated CT scan process data analytics pipeline running on a small-board computer. The pipeline produces results in-line with a minimal viable product (MVP). After applying the future suggestions in Section 4, we hope we can deploy the system in St. Michael's Hospital and get insightful data.

6. Acknowledgements

The authors of the paper would like to give thanks to KAUST Academy for sponsering Wael Sulais' work on YOLO-AR. The project is also supported by the Data Sciences Institute, University of Toronto. The authors gratefully acknowledge the guidance of Dr. Andrew Brown who is the principle investigator of the YOLO-AR project and contributions of the rest of the team.

Bibliography

[1] Y. Y. Cheung, E. M. Goodman, and T. O. Osunkoya, "No More Waits and Delays: Streamlining Workflow to Decrease Patient Time of Stay for Image-guided Musculoskeletal Procedures," *Radiographics: A Review Publication of the Radiological Society of North America, Inc*, vol. 36, no. 3, pp. 856–871, 2016, doi: 10.1148/rg.2016150174.

[2] L. Zhang, A. Hefke, J. Figiel, U. Schwarz, M. Rominger, and K. J. Klose, "Enhancing Same-Day Access to Magnetic Resonance Imaging," *Journal of the American College of Radiology: JACR*, vol. 8, no. 9, pp. 649–656, Sep. 2011, doi: 10.1016/j.jacr.2011.04.001.

[3] C. J. Roth, D. T. Boll, L. K. Wall, and E. M. Merkle, "Evaluation of MRI Acquisition Workflow with Lean Six Sigma Method: Case Study of Liver and Knee Examinations," *AJR. American journal of roentgenology*, vol. 195, no. 2, pp. W150–156, Aug. 2010, doi: 10.2214/AJR.09.3678.

[4] G. Revesz, F. J. Shea, and M. C. Ziskin, "Patient Flow and Utilization of Resources in a Diagnostic Radiology Department," *Radiology*, vol. 104, no. 1, pp. 21–26, Jul. 1972, doi: 10.1148/104.1.21.

[5] Ontario Health, "Wait Times," Jun. 2024. Accessed: Aug. 14, 2024. [Online]. Available: https://www.ontariohealth.ca/public-reporting/wait-times

[6] B. Dwyer, J. Nelson, and T. Hansen, "Roboflow." Accessed: Jul. 18, 2024. [Online]. Available: https://roboflow.com/

[7] R. Varghese and S. M., "YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and Robustness," presented at the 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS), Apr. 2024, pp. 1–6. doi: 10.1109/ADICS58448.2024.10533619.

[8] T.-Y. Lin *et al.*, "Microsoft COCO: Common Objects in Context." Accessed: Aug. 20, 2024. [Online]. Available: http://arxiv.org/abs/1405.0312

[9] "Real-Time Vision AI From Digital Twins to Cloud-Native Deployment with NVIDIA Metropolis Microservices and NVIDIA Isaac Sim." Accessed: Jul. 18, 2024. [Online]. Available: https://developer.nvidia.com/blog/real-time-vision-ai-from-digital-twins-to-cloud-native-deployment-with-nvidia-metropolis-microservices-and-nvidia-isaac-sim/