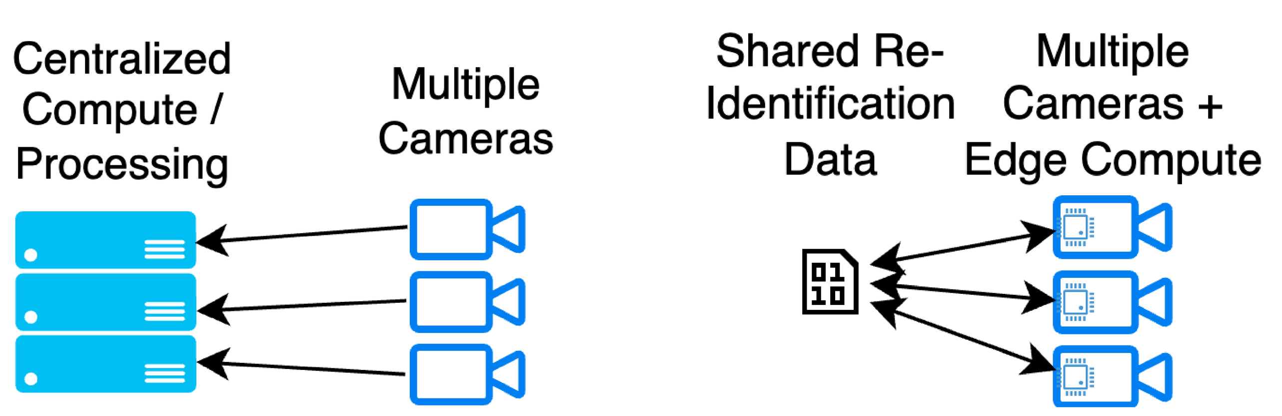
Distributed Low-Power Re-Identification of People

Eric Ye [1] Ryan Rumana [2]

[1]Electrical Engineering, Stanford [2]Computer Science, Colorado School of Mines

Introduction

Person detection and re-identification (ReID) can be used to track individual people across multiple cameras with varying viewpoints. This has applications in marketing and public safety. A centralized solution involves streaming each camera's feed to a central server for processing, which is bandwidth-intensive, costly, and does not scale with the number of cameras. We propose implementing multi-object tracking (MOT) and ReID algorithms on a distributed set of cameras, reducing the need for bandwidth by avoiding raw video streaming. This approach will enable efficient and scalable tracking across multiple cameras with minimal data transfer requirements.



(a) Centralized re-identification (b) Decentralized re-identification

Figure. Comparison of centralized and decentralized re-identification schemes.

Methods & Data

Methodology

We use Raspberry Pis with cameras as our stand-in for a security camera with some level of low-power embedded processing. We limit our models to run on the Raspberry Pi CPU (as opposed to its GPU) for simplicity. Our Raspberry Pis use Yolov8 nano [2] for creating bounding boxes alongside ResNet-50 [3] for ReID, sharing embeddings over HTTP. Embeddings are correlated using cosine similarity without processing video frames, reducing network overhead.

Datasets

We used the pre-trained version of Yolov8 nano which is trained on the Common Objects in Context (COCO). ResNet-50 was trained and evaluated on the Market1501 dataset [1] which is a pre-assembled collection of 1501 people captured from multiple different cameras and camera angles. This dataset is optimized for full body images like those captured by security cameras.

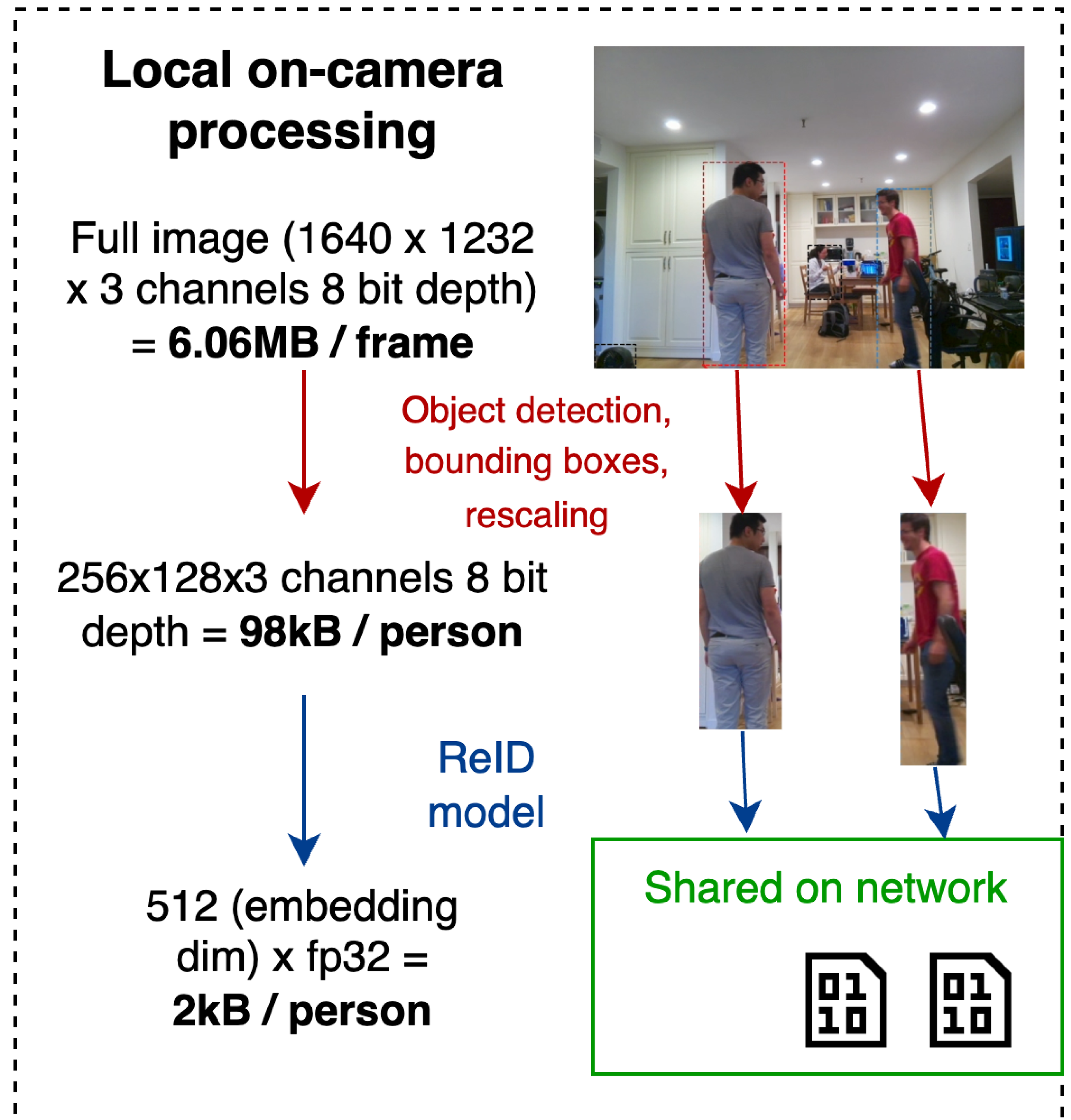


Figure. Decentralized camera image processing flow.

Quantization

To run both models on the Raspberry Pi, we quantize both of them from fp32 to int8. For Yolov8 nano, we use the pretrained pre-quantized model. For ResNet-50, we apply post-training quantization (PTQ) with a calibration sample of 100 random images from the Market1501 dataset.

Results

Model Optimization

Since we used the pre-trained version of Yolov8 we did not evaluate the accuracy after quantization. We calibrated the quantized version of ResNet-50 with 100 training images in the Market-1501 dataset. Surprisingly, the mean average precision (mAP) was **unchanged** after quantization. Quantization improved runtime of the object detection algorithm by approximately 49% and the ReID algorithm by approximately 37%. The overall runtime depends on how many people are detected in-frame, since a ReID embedding is generated for every person in each frame.

	Inference Time		
Model	Before	After	Improvement
Yolov8	395 ms	200 ms	49%
ResNet-50	462 ms	290 ms	37%

Table. Each model's average measured inference time before and after quantization alongside percent improvement.

Challenges & Future Work

Despite the performance improvement from quantization, the inference time may still be too slow to be practical, especially for images of crowds (since processing time scales with the number of people detected). There may be opportunities to improve performance further using different models or different hardware.

We identify potential for improvement in the accuracy of our system. Occasionally someone can step into different lighting or become occluded partially and the system will identify them incorrectly, hurting accuracy. This can be mitigated by more training data, utilizing clustering for a more accurate representation of individuals, capturing movement information through Kalman Filtering, and tweaking the association parameters to better correlate embedding. Additionally, we can gain accuracy by calibrating all cameras to produce more consistent embeddings across a whole deployment.

Conclusion

Our project aimed to address the existing issues with implementing camera ReID infrastructure. We succeeded in reducing or even completely removing the network impact of running ReID depending on camera configuration. While we significantly reduced the computation requirements of the chosen models, further research needs to be done to shrink ReID models enough to attain adequate edge performance. Edge deployment also enables ReID on flexible camera deployments where number and orientation are not necessarily defined. These improvements extend the use cases for ReID into the domain of drones and other moving deployments.

We would like to thank Pete Warden for advising this project.

[1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.

[2] Ultralytics. New - yolov8 in pytorch > onnx > openvino > coreml, 2024.

[3] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1116–1124, 2015.

me@ericye16.com, ryan.rumana@gmail.com