Privacy Preserving Object Detection in COSMOS Smart Intersection

Alex Angus Zhuoxu Duan Jingyuan Liu Joseph Yang Xingxing Geng Zoran Kostic

[1]Electrical Engineering Department, Columbia University in NYC

Overview

- COSMOS is an advanced wireless research testbed, equipped with networking, computing and sensing equipment in support of low-latency high-bandwidth experiments.
- COSMOS pilot site at Columbia University facilitates camera-based experiments with smart city applications at the intersection of 120th St. and Amsterdam Ave. in NYC.
- Large video datasets are needed to train deep learning based object detection models in city traffic intersections.
- Faces and license plates in intersection videos compromise personal privacy.
- We create a pipeline to systematically blur faces and license plates using the YOLOv4 object detection model.
- The pipeline blurs 99% of visible faces and license plates recorded by the 1st floor camera.

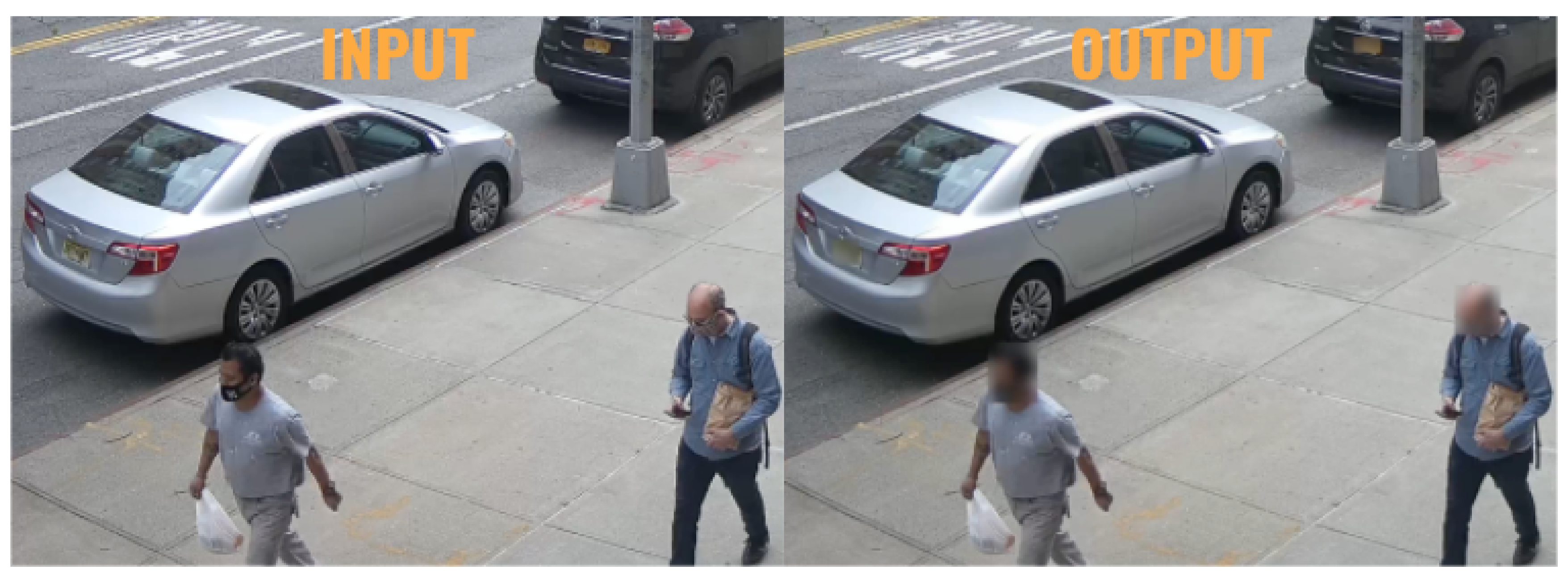

Figure 1:Example input (raw unblurred frame) and output (frame with faces and licenses blurred).

Solution

- Use deep learning models to detect faces and licenses that will then be blurred
- **Data Acquisition:** Collection of 4K resolution traffic intersection videos using first floor cameras.
- **Video Annotation:** To create ground truth labels, we use the browser based annotation tool CVAT (Computer Vision Annotation Tool) to identify, frame by frame, faces and licenses in intersection videos.
- **Object Detection:** Various YOLOv4 object detection models are trained on annotated videos using the open source neural network framework Darknet.
- **Pipeline Integration** Darknet YOLOv4 models are converted to PyTorch for integration into blurring pipeline.

Evaluation Methods

1. **Video Segmentation Validation:** All annotated videos used for training with random video segments selected for validation.
2. **Test Video Validation:** Small subset of annotated videos are withheld from training for testing.
3. **Programmatic Evaluation:** Quantitative evaluation (mAP, precision, recall, etc.) of model inference compared to ground truth labels and measures of performance on small and occluded objects.
4. **Manual Evaluation:** Qualitative evaluation of pipeline performance on visibly discernible objects.
5. **Latency Evaluation:** Analysis of inference speed of face and license detection model running on GPU.

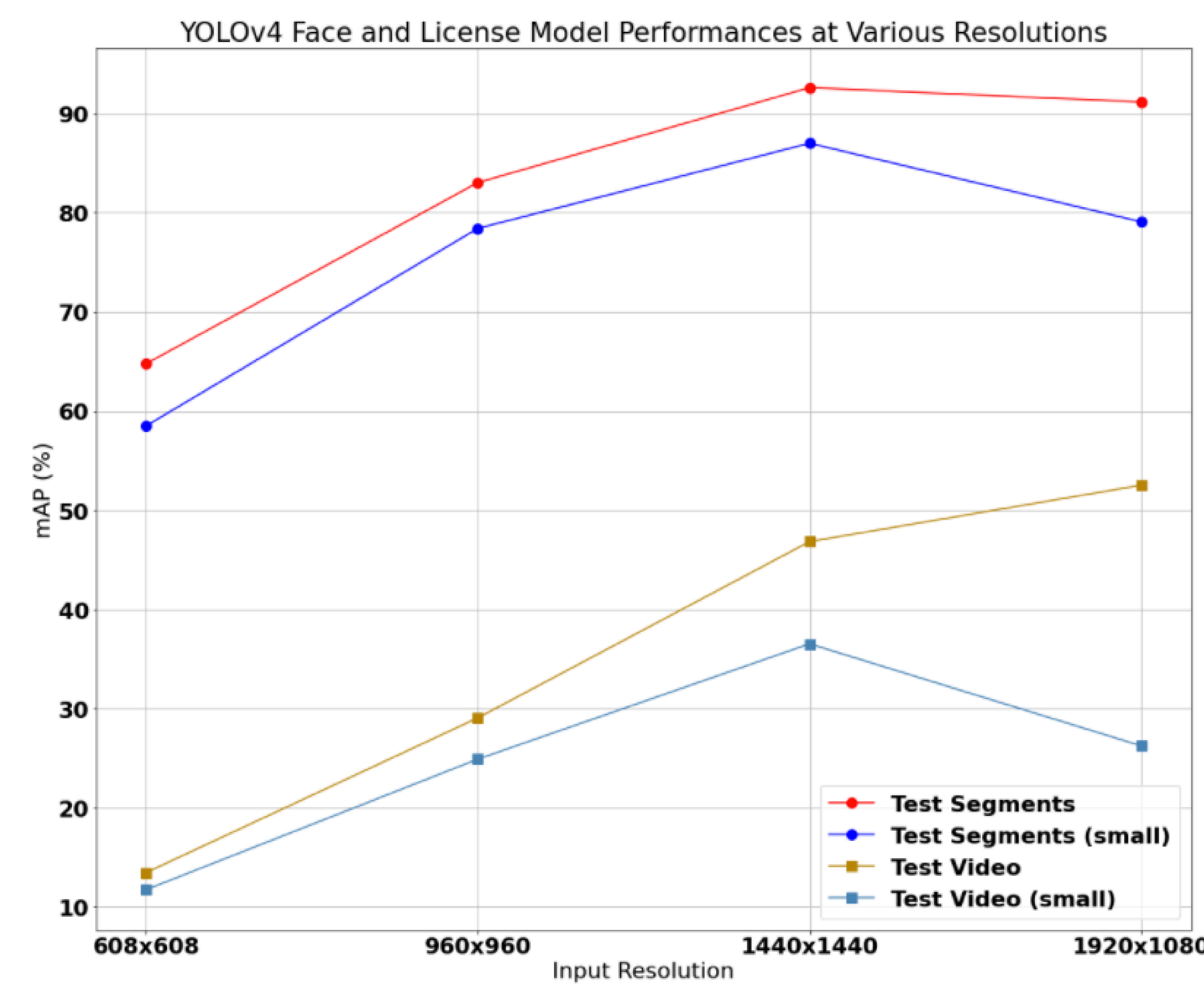

Figure 2:Plot of Face and License Plate mAP as a function of model input resolution for all objects vs only relatively small objects.

Results

Input Resolution	Face AP	License AP	Face Recall*	License Recall*	FPS
960 x 960	84.74%	75.90%	93.93%	99.96%	25.68
1440 x 1440	97.42%	96.29%	98.90%	99.96%	14.08

Figure 3:Table of selected results. * indicates recall value at the selected "visible" threshold; the size at which an object is discernible.

Figure 4:Successful detection and blurring of faces and license plates (top). False negatives highlighting various edge cases (bottom).

Conclusion

- We create a deep learning based pipeline to systematically blur face and license plates in city traffic intersections.
- For a given 1st floor intersection video, we are able to blur faces and license plates with a recall of over 99%.
- **Detection of Edge Cases:** The majority of missed objects involve scenarios not included or sparsely included in the training set.
- **Generalization:** Application of the pipeline on 2nd floor intersection videos shows further data collection and training is needed for generalization to novel intersections.

Acknowledgement

The team is grateful for the support of Columbia School of Engineering, Columbia Data Sciences Institute and Rutgers University Winlab laboratory. This work was supported in part by NSF grants CNS-1827923, OAC--2029295, and CNS-2038984.

COLUMBIA | ENGINEERING
The Fu Foundation School of Engineering and Applied Science