Expecting the Unexpected: Training Detectors for Unusual Pedestrians with Adversarial Imposters

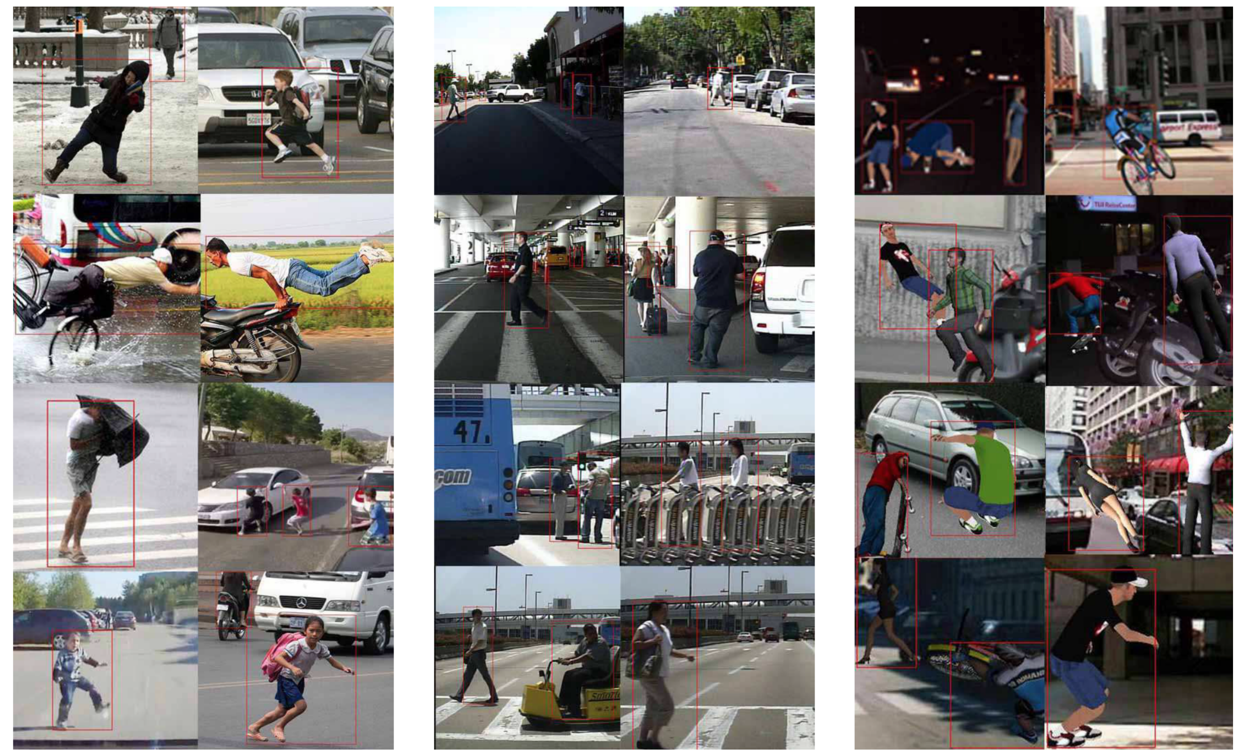
Shiyu Huang, Deva Ramanan

Department of Computer Science and Technology, Tsinghua University, Beijing, China
Robotics Institute, Carnegie Mellon University, Pittsburgh, USA

IEEE 2017 Conference on Computer Vision and Pattern Recognition
CVPR 2017 July 21-26 Honolulu

Introduction

As autonomous vehicles become an every-day reality, high-accuracy pedestrian detection is of paramount practical importance. Pedestrian detection is a highly researched topic with mature methods, but most datasets focus on common scenes of people engaged in typical walking poses on sidewalks. But performance is most crucial for dangerous scenarios, such as children playing in the street or people using bicycles/skateboards in unexpected ways. Such "in-the-tail" data is notoriously hard to observe, making both training and testing difficult.



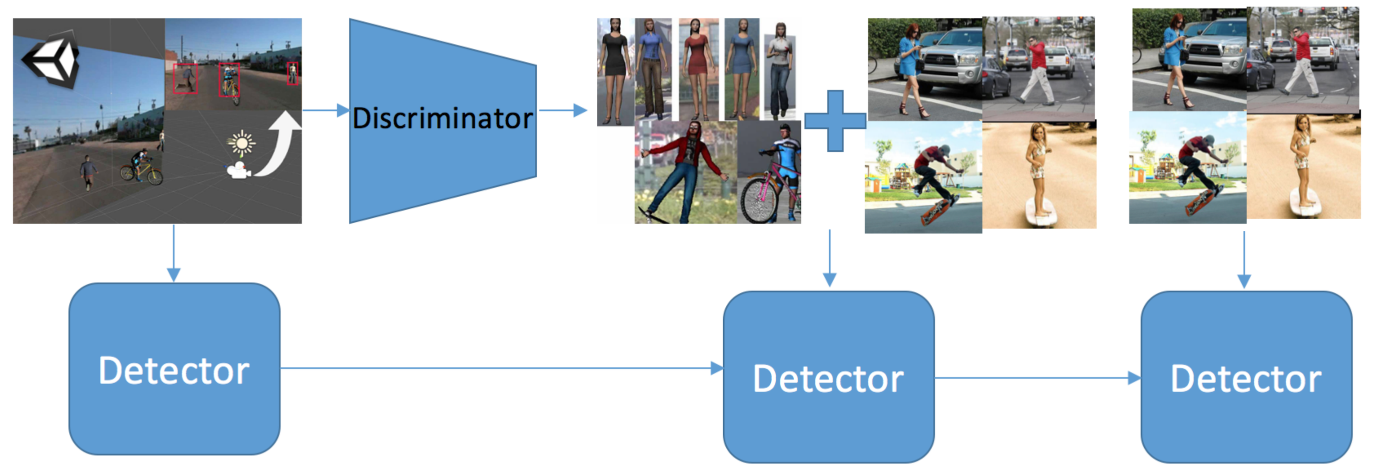
To analyze this problem, we

- Collect a novel annotated dataset of dangerous scenarios called the Precarious Pedestrian dataset;
- Explore the use of synthetic data generated by a game engine;
- Train a discriminative classifier to select a realistic subset, which we deem Synthetic Imposters;
- Demonstrate that such synthetic data can be used to rank algorithms.

Methodology

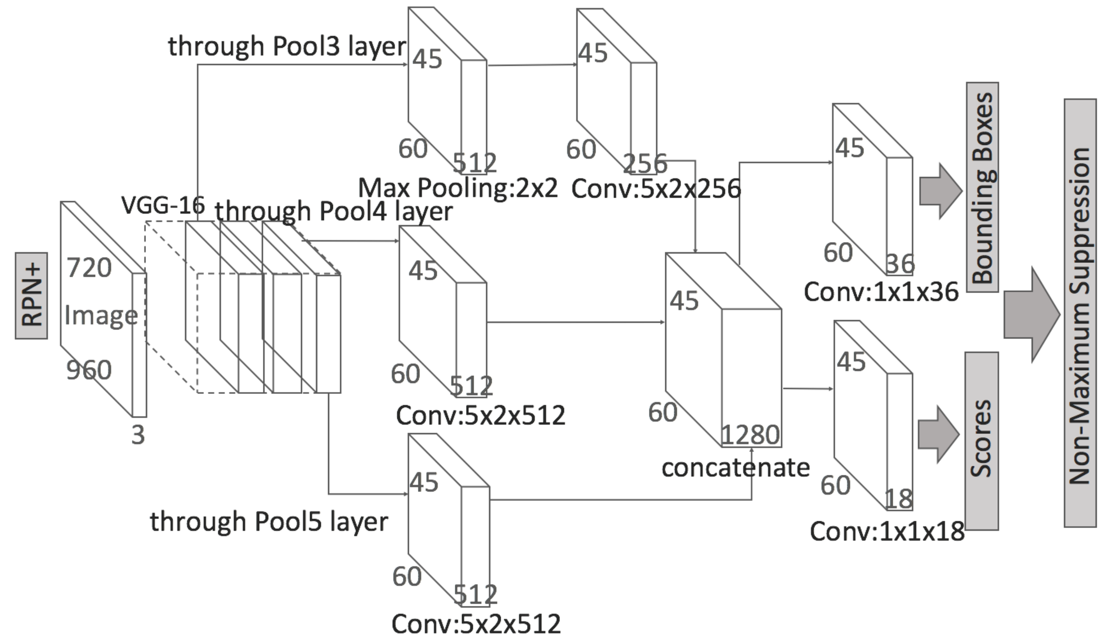
1.Framework

- Generator: Unity3D
- Discriminator: CNN
- Detector: RPN+



2.RPN+

- Fully convolutional network.
- Concatenate several layers on different stages in order to improve the ability of locating people in different resolutions.



3. Imposter Selection

- Select realistic synthetic images via discriminator.

$$\min_{I} \max_{D} V(D, I) =$$
$$\mathbb{E}_{\mathbf{x}\sim p_t(\mathbf{x})}[\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z}\sim \mathrm{Unif}(\mathcal{Z}_I)}[\log(1 - D(G(\mathbf{z})))]$$

Results

We follow the evaluation protocol of the Caltech pedestrian dataset, which use ROC curves for 2D bounding box detection at 50% and 70% overlap thresholds.

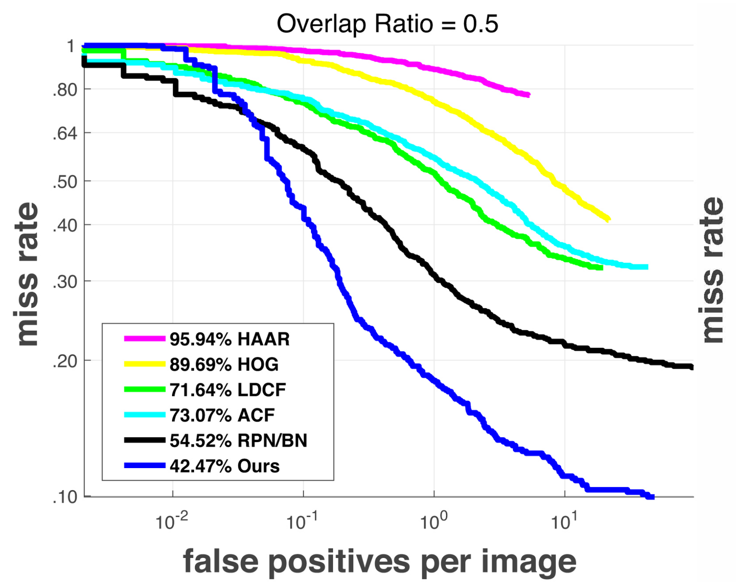
Method	50% overlap	70% overlap
HAAR	95.94%	98.4%
HOG	89.69%	96.28%
ACF	73.07%	88.50%
LDCF	71.64%	87.37%
RPN/BF	54.52%	82.8%
RPN+	58.82%	84.17%
Ours	**42.47%**	**73.70%**

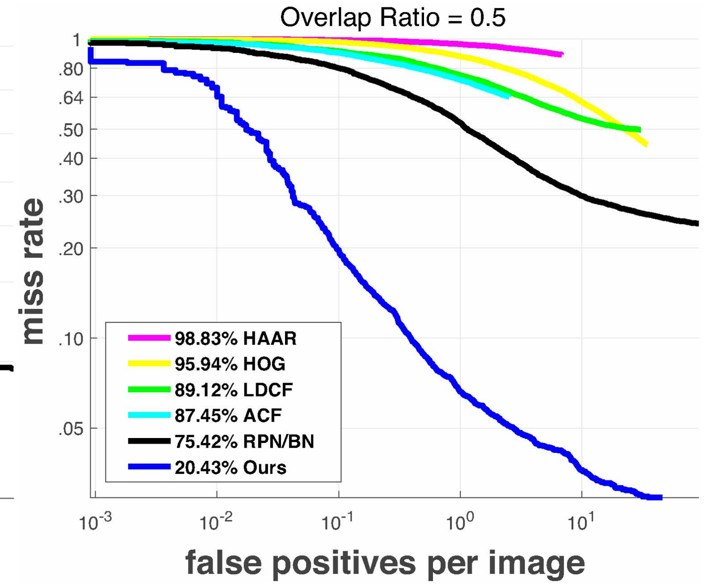
Our proposed method is more robust on precarious dataset.

We study how the training process influences the performance. S, T, and I refer to source datasets(synthetic images), target dataset(real Precarious images), and the Imposter dataset.

Fine-tuning method	50% overlap	70% overlap
S	83.49%	95.18%
T	72.39%	93.70%
$S \Rightarrow T$	48.45%	77.14%
$S \Rightarrow (T \cup I)$	45.97%	74.94%
$S \Rightarrow (T \cup I) \Rightarrow T$	**42.47%**	**73.70%**

We show that synthetic data can also be used for evaluation(left: Precarious Testset, right: Synthetic Testset).





Conclusions

We have explored methods for analyzing "in-the-tail" urban scenes, which represent important modes of operations for autonomous vehicles. We propose the use of synthetic data generated with a game engine. And inspired by generative adversarial networks, we introduce the use of a discriminative classifier to implicitly specify this distribution. We then use the synthesized data that fooled the discriminator to both train and evaluate state-of-the-art, robust pedestrian detection systems.