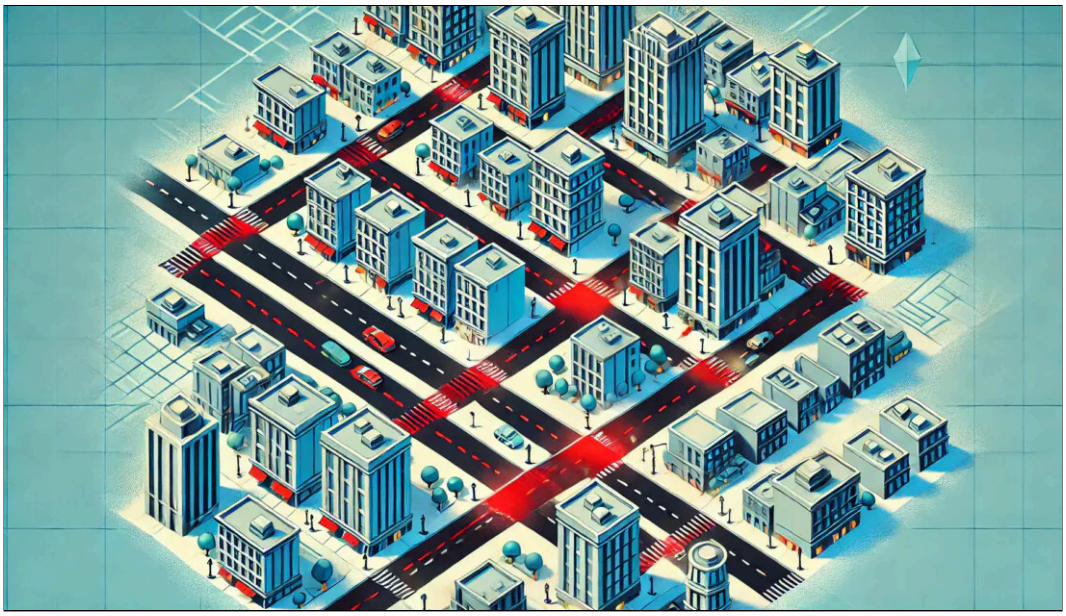
The ‘Secret Routes’ That Can Foil Pedestrian Recognition Systems

Originally published at https://www.unite.ai/the-secret-routes-that-can-foil-pedestrian-recognition-systems/

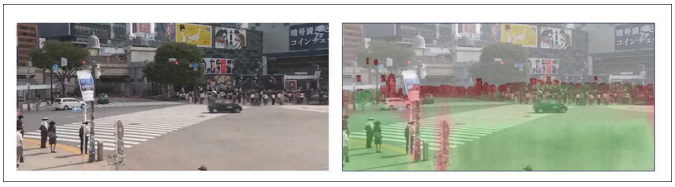


A new research collaboration between Israel and Japan contends that pedestrian detection systems possess inherent weaknesses, allowing well-informed individuals to evade facial recognition systems by navigating carefully planned routes through areas where surveillance networks are least effective.

With the help of publicly available footage from Tokyo, New York and San Francisco, the researchers developed an automated method of calculating such paths, based on the most popular object recognition systems likely to be in use in public networks.

The three crossings used in the study: Shibuya Crossing in Tokyo, Japan; Broadway, New York; and Castro District, San Francisco. Source: https://arxiv.org/pdf/2501.15653

By this method, it’s possible to generate *confidence heatmaps* that demarcate areas within the camera feed where pedestrians are least likely to provide a positive facial recognition hit:



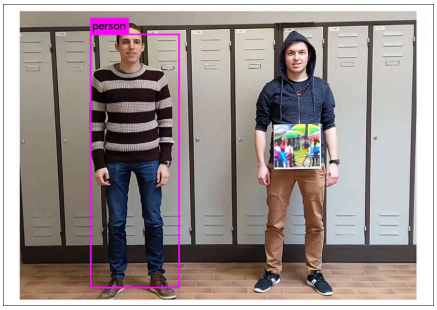
On the right, we see the confidence heatmap generated by the researchers’ method. The red areas indicate low confidence, and a configuration of stance, camera pose and other factor that are likely to impede facial recognition.

In theory such a method could be instrumentalized into a location-aware app, or some other kind of platform to disseminate the least ‘recognition-friendly’ paths from A to B in any calculated location.

The new paper proposes such a methodology, titled *Location-based Privacy Enhancing Technique* (L-PET); it also proposes a countermeasure titled *Location-Based Adaptive Threshold* (L-BAT), which essentially runs exactly the same routines, but then uses the information to reinforce and improve the surveillance measures, instead of devising ways to avoid being recognized; and in many cases, such improvements would not be possible without further investment in the surveillance infrastructure.

The paper therefore sets up a potential technological war of escalation between those seeking to optimize their routes to avoid detection and the ability of surveillance systems to make full use of facial recognition technologies.

Prior methods of foiling detection are less elegant than this, and center on adversarial approaches, such as TnT Attacks, and the use of printed patterns to confuse the detection algorithm.



The 2019 work 'Fooling automated surveillance cameras: adversarial patches to attack person detection' demonstrated an adversarial printed pattern capable of convincing a recognition system that no person is detected, allowing a kind of 'invisibility. Source: https://arxiv.org/pdf/1904.08653

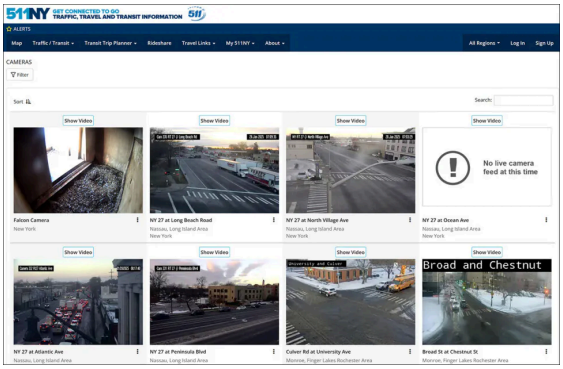
The researchers behind the new paper observe that their approach requires less preparation, with no need to devise adversarial wearable items (see image above).

The paper is titled *A Privacy Enhancing Technique to Evade Detection by Street Video Cameras Without Using Adversarial Accessories*, and comes from five researchers across Ben-Gurion University of the Negev and Fujitsu Limited.

Method and Tests

In accordance with previous works such as Adversarial Mask, AdvHat, adversarial patches, and various other similar outings, the researchers assume that the pedestrian 'attacker' knows which object detection system is being used in the surveillance network. This is actually not an unreasonable assumption, due to the widespread adoption of state-of-the-art open source systems such as YOLO in surveillance systems from the likes of Cisco and Ultralytics (currently the central driving force in YOLO development).

The paper also assumes that the pedestrian has access to a live stream on the internet fixed on the locations to be calculated, which, again, is a reasonable assumption in most of the places likely to have an intensity of coverage.



Sites such as 511ny.org offer access to many surveillance cameras in the NYC area. Source: https://511ny.or

Besides this, the pedestrian needs access to the proposed method, and to the scene itself (i.e., the crossings and routes in which a 'safe' route is to be established).

To develop L-PET, the authors evaluated the effect of the pedestrian angle in relation to the camera; the effect of camera height; the effect of distance; and the effect of the time of day. To obtain ground truth, they photographed a person at the angles 0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315°.

Ground truth observations carried out by the researchers.

They repeated these variations at three different camera heights (0.6m, 1.8m, 2.4m), and with varied lighting conditions (morning, afternoon, night and 'lab' conditions).

Feeding this footage to the Faster R-CNN and YOLOv3 object detectors, they found that the confidence of the object depends on the acuteness of the angle of the pedestrian, the pedestrian's distance, the camera height, and the weather/lighting conditions*.

The authors then tested a broader range of object detectors in the same scenario: Faster R-CNN; YOLOv3; SSD; DiffusionDet; and RTMDet.

The authors state:

> *'We found that all five object detector architectures are affected by the pedestrian position and ambient light. In addition, we found that for three of the five models (YOLOv3, SSD, and RTMDet) the effect persists through all ambient light levels.'*

To extend the scope, the researchers used footage taken from publicly available traffic cameras in three locations: Shibuya Crossing in Tokyo, Broadway in New York, and the Castro District in San Francisco.

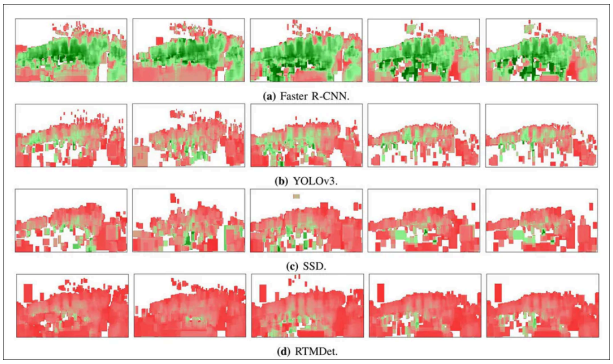
Each location furnished between five and six recordings, with approximately four hours of footage per recording. To analyze detection performance, one frame was extracted every two seconds, and processed using a Faster R-CNN object detector. For each pixel in the obtained frames, the method estimated the average confidence of the 'person' detection bounding boxes being present in that pixel.

> *'We found that in all three locations, the confidence of the object detector varied depending on the location of people in the frame. For instance, in the Shibuya Crossing footage, there are large areas of low confidence farther away from the camera, as well as closer to the camera, where a pole partially obscures passing pedestrians.'*

The L-PET method is essentially this procedure, arguably 'weaponized' to obtain a path through an urban area that is least likely to result in the pedestrian being successfully recognized.

By contrast, L-BAT follows the same procedure, with the difference that it updates the scores in the detection system, creating a feedback loop designed to obviate the L-PET approach and make the 'blind areas' of the system more effective.

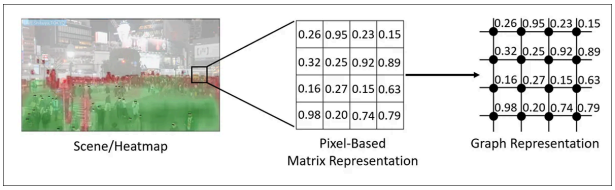
(In practical terms, however, improving coverage based on obtained heatmaps would require more than just an upgrade of the camera sitting in the expected position; based on the testing criteria, including location, it would require the installation of additional cameras to cover the neglected areas – therefore it could be argued that the L-PET method escalates this particular 'cold war' into a very expensive scenario indeed)



The average pedestrian detection confidence for each pixel, across diverse detector frameworks, in the observed area of Castro Street, analyzed across five videos. Each video was recorded under different lighting conditions: sunrise, daytime, sunset, and two distinct nighttime settings. The results are presented separately for each lighting scenario.

Having converted the pixel-based matrix representation into a graph representation suitable for the task, the researchers adapted the Dijkstra algorithm to calculate optimal paths for pedestrians to navigate through areas with reduced surveillance detection.

Instead of finding the shortest path, the algorithm was modified to minimize detection confidence, treating high-confidence regions as areas with higher 'cost'. This adaptation allowed the algorithm to identify routes passing through blind spots or low-detection zones, effectively guiding pedestrians along paths with reduced visibility to surveillance systems.



A visualization depicting the transformation of the scene's heatmap from a pixel-based matrix into a graph-based representation.

The researchers evaluated the impact of the L-BAT system on pedestrian detection with a dataset built from the aforementioned four-hour recordings of public pedestrian traffic. To populate the collection, one frame was processed every two seconds using an SSD object detector.

From each frame, one bounding box was selected containing a detected person as a positive sample, and another random area with no detected people was used as a negative sample. These twin samples formed a dataset for evaluating two Faster R-CNN models – one with L-BAT applied, and one without.

The performance of the models was assessed by checking how accurately they identified positive and negative samples: a bounding box overlapping a positive sample was considered a true positive, while a bounding box overlapping a negative sample was labeled a false positive.

Metrics used to determine the detection reliability of L-BAT were Area Under the Curve (AUC); true positive rate (TPR); false positive rate (FPR); and average true positive confidence. The researchers assert that the use of L-BAT enhanced detection confidence while maintaining a high true positive rate (albeit with a slight increase in false positives).

In closing, the authors note that the approach has some limitations. One is that the heatmaps generated by their method are specific to a particular time of day. Though they do not expound on it, this would indicate that a greater, multi-tiered approach would be needed to account for the time of day in a more flexible deployment.

They also observe that the heatmaps will not transfer to different model architectures, and are tied to a specific object detector model. Since the work proposed is essentially a proof-of-concept, more adroit architectures could, presumably, also be developed to remedy this technical debt.

Conclusion

Any new attack method for which the solution is 'paying for new surveillance cameras' has some advantage, since expanding civic camera networks in highly-surveilled areas can be politically challenging, as well as representing a notable civic expense that will usually need a voter mandate.

Perhaps the biggest question posed by the work is *'Do closed-source surveillance systems leverage open source SOTA frameworks such as YOLO?'*. This is, of course, impossible to know, since the makers of the proprietary systems that power so many state and civic camera networks (at least in the US) would argue that disclosing such usage might open them up to attack.

Nonetheless, the migration of government IT and in-house proprietary code to global and open source code would suggest that anyone testing the authors' contention with (for example) YOLO might well hit the jackpot immediately.

* *I would normally include related table results when they are provided in the paper, but in this case the complexity of the paper's tables makes them unilluminating to the casual reader, and a summary is therefore more useful.*

First published Tuesday, January 28, 2025

Writer on machine learning, domain specialist in human image synthesis. Former head of research content at Metaphysic.ai, until its dissolution into DNEG's Brahma.ai.
Portfolio site: martinanderson.ai
Contact: martin@martinanderson.ai