InvisibiliTee: Angle-agnostic Cloaking from Person-Tracking Systems with a Tee

Yaxian Li\textsuperscript{1,2}, Bingqing Zhang\textsuperscript{1,2}, Guoping Zhao\textsuperscript{1,2}, Mingyu Zhang\textsuperscript{1,2}, Jiajun Liu\textsuperscript{4 \ast}, Ziwei Wang\textsuperscript{4}, and Jirong Wen\textsuperscript{1,2,3}

\textsuperscript{1} School of Information, Renmin University of China, Beijing, China
\textsuperscript{2} Beijing Key Laboratory of Big Data Management and Analysis Methods, Beijing, China
\textsuperscript{3} Gaoling School of Artificial Intelligence, Renmin University of China, Beijing, China
\textsuperscript{4} Data 61, CSIRO, Pullenvale, Australia

Abstract. After a survey for person-tracking system-induced privacy concerns, we propose a black-box adversarial attack method on state-of-the-art human detection models called InvisibiliTee. The method learns printable adversarial patterns for T-shirts that cloak wearers in the physical world in front of person-tracking systems. We design an angle-agnostic learning scheme which utilizes segmentation of the fashion dataset and a geometric warping process so the adversarial patterns generated are effective in fooling person detectors from all camera angles and for unseen black-box detection models. Empirical results in both digital and physical environments show that with the InvisibiliTee on, person-tracking systems’ ability to detect the wearer drops significantly.

Keywords: Object Detection · Human Tracking · Adversarial Attack.

1 Introduction

Person-tracking systems are widely deployed in metropolitan areas across the world for various purposes. According to the latest Comparitech report\footnote{https://www.comparitech.com/vpn-privacy/the-worlds-most-surveilled-cities/}, approximately 770 million cameras have already been used globally, which include many smart cameras enabled with person-tracking systems. While making contributions to public safety, these network-enabled systems suffer from software/hardware vulnerabilities and are often prone to cyber-attacks\footnote{https://www.scmp.com/news/china/article/1727145/chinese-surveillance-camera-supplier-confirms-hacking-loophole}, raising serious concerns about privacy breaches. Thus, average citizens are now facing even stronger privacy risks. For instance, after hacking into camera-based person-tracking systems, it is not only possible but even unchallenging to precisely recover a person’s daily routine.

\ast Corresponding author. Jiajun.liu@csiro.au
\dagger Code is available at https://github.com/invisibilitee/invisibilitee
\ddagger https://www.comparitech.com/vpn-privacy/the-worlds-most-surveilled-cities/
To learn how this issue is perceived by the general public, we conducted a small-scale questionnaire-based survey. Among the 20 participants, all are concerned with the person-tracking systems’ potential breach of privacy and 90% responded that they feel the need to leverage latest technology to protect personal privacy, as a counter-balance to the evolving person-tracking technology. This interesting study inspires our research for the “InvisibiliTee”, a tee that cloaks wearer in front of the tracking system. As illustrated in Fig. 1 people can successfully fool the person detection system with the “InvisibiliTee” on. Before diving into the technical details of the “InvisibiliTee”, we first show how the survey is conducted and what we could conclude from the results.

1.1 User Study

As shown in Table 1 there are 4 questions in the questionnaire. It involves the urgency of privacy issues, the acceptance of technical means and adversarial patterns. The 20 participants surveyed are college students aged between 22 to 30, with higher education background and living experience in metropolis. They are also the target demographic group that InvisibiliTee tries to help - younger generations with stronger awareness of privacy risks and open mindset towards early technology adoption. Among the 20 people surveyed, all are worried about the camera’s invasion of personal privacy, of which 55% are concerned and 35% are extremely concerned. 90% say they feel the need to leverage technical approaches to protect personal privacy and over a half show the willingness to wear an “InvisibiliTee”. There are 7 people who think the patterns of “InvisibiliTee” not fashionable enough to wear. But others do consider these patterns acceptable, among which pattern C in Fig 2 is most popular, chosen by 10 people.

Motivated by a strong need for privacy protection in the current world where person-tracking systems are often abused, an angle-agnostic black-box
adversarial attack method, namely InvisibiliTee, is proposed in this paper. Its advantages are summarized as follows: 1, **Cross-model generalizability.** The learned attack patterns achieve competitive black-box attack results on unseen detection models. 2, **Cross-scene/subject generalizability.** The same learned adversarial pattern can be used by applied on unseen wearers in unseen scenes without re-training. 3, **Angle-agnostic.** In the digital world, adversarial patterns can be applied on images taken from different angles. In the physical world, we use fabric with fully printed adversarial pattern to tailor a T-shirt, achieving angle-agnostic “invisible” effect and qualitative analysis shows effectiveness of attack to some extent.

## 2 Literature Review

In person tracking system, a critical step is to utilize object detection algorithms to generate bounding box to crop the images with persons. Fast R-CNN [8] introduced a RoIPooling operation to accelerate detection speed while maintaining detection performance. He et al. proposed Faster R-CNN [20], utilizing Regional Proposal Network to generate proposals instead of time-consuming selective search [23]. R-FCN [4] comes up with position-sensitive RoIPooling for further improving efficiency. Single-stage frameworks, such as YOLO [18,19,5] and SSD [15,6,14], directly classify and regress bounding boxes. Although detectors are getting better performances, recent studies have proven that the deep object detection models are vulnerable to attacks from adversarial examples [219]. Adversarial attack methods can be divided into two categories, which are **white-box** attack and **black-box** attack. In white-box attack, the attacker can interact with the machine learning system in the process of generating adversarial attack data, obtaining the gradient during the training process. While in black-box attack, attackers do not know the algorithms nor parameters of the target model. But attackers can still observe the output through any input. Several attempts have been made on attacking Re-ID systems via black-box methods [2430]. Although existing works demonstrate the feasibility
of performing black-box attacks on classification and Re-ID vision systems, the research on person detection system has not yet been sufficiently investigated.

Liu et al. [16] propose DPATCH, which is a universal black-box adversarial attack method. It borrows the idea of adversarial patch [1] that simultaneously attacks bounding boxes regression and classification. Later, Lee et al. [13] consider a normalized steepest ascent approach, PGD [17], to update loss function. Thys et al. [22] try to attack the object with high level of intra-class variety such as person. Recently, Wu et al. [26] design an “invisibility cloak” that decreases the scores of a series of object detector. However, to the best of our knowledge, only a few works [26] focused on designing a pattern on a T-shirt which can attack detectors without direction restrictions. We propose “InvisibiliTee”, which is an angle-agnostic adversarial attack method.

3 Method

Fig. 3. An overview of normal object detection networks in persons-tracking systems (a) and the framework of InvisibiliTee (b)(c). (a) is a normal detection pipeline. (b) illustrates the process of adversarial pattern training, including dynamic warping, semantic masking and pattern training. (c) describes the process of adversarial attack in both digital and physical scenario.

3.1 Overview

As illustrated in Fig. 3(b), the InvisibiliTee’s network contains two parts: the attack head and object detection system. Attack head is designed to generate the
shared adversarial pattern which consists of three steps. 1, **Dynamic Warping.** According to the clothing key points annotation, the angle and position of the pattern are adjusted dynamically through warping. 2, **Semantic Masking.** We obtain the clothing polygon mask from annotation and attach the pattern onto the original image to generate the attack picture. 3, **Pattern Training.** Adversarial pattern consists of color pixels, which are directly derived from learnable neural network parameters. During training, only the parameters of adversarial pattern are updated, while the pre-trained detection model remains fixed. A set of adversarial attacking losses and geometric constraint losses are proposed to update pattern parameters.

The second part of the network is object detection system. According to the different processing stages, single-stage framework YOLO [18,19,5] and two-stage framework Faster R-CNN [20] are implemented. Note that following the practice of previous research [16], the adversarial pattern here is a group of network parameters. So it can be optimized during model training. It is randomly initialized and updated iteratively. The learned adversarial pattern is tested on other unseen models following black-box attack protocols to evaluate the attacking performance, as shown in Fig. 3(c).

### 3.2 Attack and Geometric Constraint Loss Functions

**Attack Loss** Attack Losses are optimized to fool the detectors into making wrong decisions. When training with YOLO architecture, \( \text{Loss}_{\text{attk}} \) consists of three training losses.

\[
\text{Loss}_{\text{attk}} = \theta_{11} \text{Loss}_{\text{cla}} + \theta_{12} \text{Loss}_{\text{coord}} + \theta_{13} \text{Loss}_{\text{wh}}.
\] (1)

When training with Faster R-CNN architecture, \( \text{Loss}_{\text{attk}} \) consists of two training losses.

\[
\text{Loss}_{\text{attk}} = \theta_{21} \text{Loss}_{\text{cla}} + \theta_{22} \text{Loss}_{\text{bbox}}.
\] (2)

Classification cross-entropy attack loss is expressed as

\[
\text{Loss}_{\text{cla}} = \sum_{i=0}^{C-1} y_i \log (\hat{y}_i)
\] (3)

where \( y_i \) stands for the ground truth category label, \( \hat{y}_i \) indicates the probability that current sample belongs to the \( i \)-th category and \( C \) is the number of categories. Bounding box coordinate cross-entropy attack loss is defined as

\[
\text{Loss}_{\text{coord}} = \sum_{i=1}^{2} x_i \log (\hat{x}_i)
\] (4)

where \( x_i \) stands for the ground truth bounding box coordinate \( x_i = (x_i^1, x_i^2) \), \( \hat{x}_i \) indicates the predicted one. Bounding box width-height mean squared error attack loss is presented as

\[
\text{Loss}_{\text{wh}} = e^{-\frac{1}{2}((w-\bar{w})^2+(h-\bar{h})^2)}
\] (5)
where \( w \) and \( h \) stand for the width and height of the bounding box respectively. Bounding box L1 attack loss is presented as

\[
\text{Loss}_{\text{bbox}} = e^{-|\hat{t}_i - t_i|}
\]  

(6)

\( \hat{t}_i \) is a vector representing the 4 parameterized coordinates of the predicted bounding box, and \( t_i \) is that of the ground-truth box. **Geometric Constraint Loss** To learn a more continuous as well as printable adversarial pattern, we introduce total variation loss and non-printability score to constrain the pattern in the training process.

Total Variation (TV) loss encourages spatial smoothness in the generated image. The TV loss is defined as follows,

\[
L_{\text{tv}}(P) = \sum_{i,j} \sqrt{\left((a_{i,j} - a_{i+1,j})^2 + (a_{i,j} - a_{i,j+1})^2\right)}
\]  

(7)

where \( P \) is the adversarial pattern, \( a_{i,j} \) stands for the pixel value in the pattern of position \((i, j)\). \( L_{\text{tv}}(P) \) minimizes the distance between neighboring pixels. Thus it can make the pattern more natural and smooth.

Non-printability score [22] reflects how difficult it is to print the pattern, and the lower the value, the less distorted it will be printed. It is defined as:

\[
L_{\text{print}}(P) = \sum_{a_{(i,j)} \in P} \min_{c_{\text{print}} \in C} |a_{(i,j)} - c_{\text{print}}|.
\]  

(8)

where \( c_{\text{print}} \) stands for the pre-defined printable colors. \( L_{\text{print}}(P) \) minimizes the distance between pattern pixels and printable pixels.

The Joint Constraint Loss \( \text{Loss}_{\text{cons}} \) can be represented by the equation below:

\[
\text{Loss}_{\text{cons}} = \alpha_1 L_{\text{tv}}(P) + \alpha_2 L_{\text{print}}(P)
\]  

(9)

where \( \alpha_1 \) and \( \alpha_2 \) are hyper-parameters.

**Overall Loss Function** The final loss \( L \) of our attack model can be expressed as:

\[
L = \text{Loss}_{\text{attk}} + \text{Loss}_{\text{cons}}
\]  

(10)

3.3 Geometric Warp and Masking

As shown in Fig. 4, we apply geometric warp to the pattern according to the T-shirt key points before attack via a polygonal mask. In Fig. 4, brown lines stand for information supervision while blue lines stand for transformation. Geometric warp operation is also known as perspective transformation, widely used to project the image to a new viewing plane, and its general transformation formula is:

\[
\begin{bmatrix}
    x' \\
    y' \\
    w'
\end{bmatrix} = \begin{bmatrix}
    u & v & w
\end{bmatrix} \cdot T
\]  

(11)

\((u, v)\) are the original image pixel coordinates, \((x = x'/w', y = y'/w')\) are the transformed image pixel coordinates. \( T \) is the perspective transformation matrix. Given corresponding four pairs of pixel coordinates before/after perspective
transformation, the perspective transformation matrix $T$ can be obtained. We warp the pattern image to fit the input image, so the original 4 pairs of coordinates are the 4 corners of the pattern.

After geometric warping, we use the polygonal mask $M$ formed by the T-shirt key points to attach the pattern onto the image, replacing the original T-shirt outward appearance.

$$I_{\text{attk}} = (1 - \text{Mask}) \odot I_{\text{ori}} + M \odot P$$ (12)

4 Attacks in the Digital World

4.1 Dataset and Experiment Setup

The Deep Fashion2 dataset [7] is a comprehensive fashion dataset containing pictures of people wearing various types of clothing with annotations of clothing key points. Each image has around 15-30 key points to outline the clothes, which is shown in Fig.6. We leverage the open sourced object detection project MMDe-
tection [3] and pre-trained models to implement our experiments. YOLOv3 [5] and Faster R-CNN [20] are chosen as the training detection model. The adversarial pattern is trained on the modified Deep Fashion2 dataset with Adam optimizer with initial learning rate $1\times 10^{-3}$. The hyper-parameters $\theta_{11}$, $\theta_{13}$, $\alpha_1$ and $\alpha_2$ are 5, 1, 1, 100, 100 respectively in YOLOv3. And $\theta_{21}$, $\theta_{22}$, $\alpha_1$ and $\alpha_2$ are 500, 10, 18, 100 in Faster R-CNN, balancing the loss terms into the same scale. Fig.5 shows the changing of total loss during training.

| Split | No human Frontal Viewpoint | Size/back Viewpoint | Total |
|-------|----------------------------|---------------------|-------|
| Train | 6,664                      | 56,121              | 8,860 |
| Val   | 1,027                      | 9,855               | 1,674 |

Table 2. Dataset information after reconstruction.
4.2 Experimental Results
We train on a specific object detection model and then transfer the trained pattern to attack other unseen models.

Table 3. The black-box (digital) attack results on SOTA detectors with an adversarial pattern trained on YOLOv3 [5] and Faster-RCNN [20]. The patterns used are shown in Fig. 2 A and B respectively.

| Target model          | AP@IoU=0.50:0.95 without Faster R-CNN attack | AP@0.50 Faster R-CNN attack | AP@0.75 Faster R-CNN attack |
|-----------------------|---------------------------------------------|------------------------------|-----------------------------|
| YOLOv3 [5]            | 0.254 0.204 0.003                           | 0.601 0.512 0.022           | 0.178 0.131 0.000           |
| SSD [5]               | 0.256 0.184 0.187                           | 0.599 0.532 0.558           | 0.133 0.100 0.102           |
| RetinaNet [18]        | 0.216 0.000 0.000                           | 0.607 0.000 0.000           | 0.120 0.000 0.000           |
| Faster-RCNN [20]      | 0.211 0.081 0.197                           | 0.564 0.087 0.512           | 0.128 0.000 0.000           |
| Mask R-CNN [20]       | 0.210 0.047 0.071                           | 0.573 0.164 0.270           | 0.131 0.000 0.042           |
| Cascade Mask R-CNN [17]| 0.221 0.063 0.089                           | 0.599 0.232 0.291           | 0.124 0.019 0.041           |
| Cascade R-CNN [17]    | 0.211 0.038 0.061                           | 0.607 0.159 0.216           | 0.132 0.010 0.031           |
| Dynamic R-CNN [28]    | 0.227 0.053 0.078                           | 0.600 0.202 0.264           | 0.133 0.015 0.035           |
| CornerNet [12]        | 0.222 0.133 0.084                           | 0.541 0.438 0.249           | 0.157 0.084 0.044           |
| RepPoints [27]        | 0.309 0.046 0.002                           | 0.567 0.190 0.252           | 0.397 0.041 0.023           |
| FreeAnchor [12]       | 0.218 0.000 0.078                           | 0.598 0.000 0.265           | 0.121 0.000 0.038           |
| SADB [23]             | 0.265 0.058 0.076                           | 0.657 0.210 0.249           | 0.175 0.020 0.040           |
| PAA [23]              | 0.243 0.075 0.089                           | 0.607 0.217 0.276           | 0.178 0.036 0.045           |

Black-box Attacks. Table 3 shows the black-box attack performance on several different SOTA detection models, where AP stands for Average Precision, the most commonly used evaluation metric for target detection. Our attack model is able to significantly reduce AP at different IoU ratios. The $AP@IoU = 0.50 : 0.95$ corresponds to average mAP calculated over a range of IoU thresholds, from 0.50 to 0.95 with the uniform step size 0.05. As shown in column 2 and 3, the $AP@IoU = 0.50 : 0.95$ is reduced from about 0.25 to about 0.08. In column 5, it can be seen that even under the most relaxed threshold condition ($IoU > 0.5$ means that it was a hit), our method reduces the evaluation to less than half in most models. When the IoU threshold is set to 0.75, it can be considered that the detector is close to completely invalid after attack.

Effect of Different Pattern Resolution Fig. 7 shows several adversarial patterns with different resolutions. (a) are patterns in size $200 \times 200$ while (b)
Table 4. Black-box experiment results of different pattern resolutions trained on Faster-RCNN. “w/o attack” means no attack is performed. P 200, p 100, p 50 stand for adversarial pattern resolution of 200 × 200, 100 × 100 and 50 × 50 pixels, respectively.

| target model | CornerNet | Cascade R-CNN | Mask R-CNN | Faster R-CNN | Faster-RCNN + softnms | Cascade R-CNN | YOLOv3 |
|--------------|-----------|---------------|------------|--------------|----------------------|---------------|--------|
| w/o attack   | 0.541     | 0.607         | 0.713      | 0.564        | 0.203                | 0.599         | 0.601  |
| p 200        | 0.438     | 0.159         | 0.197      | 0.007        | 0.164                | 0.232         | 0.512  |
| p 100        | 0.437     | 0.150         | 0.165      | 0.126        | 0.136                | 0.208         | 0.529  |
| p 50         | 0.443     | 0.185         | 0.211      | 0.174        | 0.188                | 0.289         | 0.524  |

Fig. 7. Adversarial patterns trained with different resolution in pixels. (a) and (b) are random initialized while (c) is initialized by a texture.

are corresponding patterns in size 100 × 100. The first pattern on left is trained on Faster-RCNN and the last two are trained on YOLOv3. Fig. 8 shows several adversarial patterns initialized by different textures. We use downsampling in training to study the effect of image resolution. These corresponding patterns with different resolutions are firstly random initialized in size 400 × 400 with the same random seed and then they are downsampled to different resolutions before dynamic warping. As Table 4 indicates, high-resolution adversarial pattern performs slightly better in most cases, but the advantage is not obvious. It is believed that a relatively lower resolution can be better considering a limited scale of training set.

4.3 Case Studies of Digital Attacks

Fig. 9 shows the digital attack results when we use the adversarial pattern trained on YOLOv3 [5] to attack the target model CornerNet [12]. Only bounding boxes with a confidence score over 0.5 are shown. As Fig. 9 indicates, successful attack results can be roughly divided into following three types, which are namely detection failure, bounding box miss and category error. Most cases belong to the first type, which are (a)1, (a)3-8 and (c)1-5. (a)2 and (c)6 provide a misleading or incomplete bounding box. In (c)7, the target model recognizes it as a cake instead of a person. There are also some failure cases as (c)8. However, wearing a “InvisibiliTee” still manages to make the confidence score drop from 0.86 to 0.64.

5 Attacks in the Physical World

We print the adversarial pattern on fabric and tailor it into an “InvisibiliTee” to evaluate attack effects in physical world.

Fig. 10 shows a typical scene in real life when a pedestrian walked from an outdoor parking lot into a building. Note that the mosaics are added for anonymity only and do not have any impact in the experiments. When he wears
Fig. 8. Adversarial patterns initialized by different textures. In each group, the left one is the texture and the right one is the corresponding adversarial pattern.

Fig. 9. Digital attack results. “w/ attack” and “w/o attack” stand for with attack and without attack respectively.

an “InvisibiliTee”, detectors fail to identify the person. But at the same time it is able to recognize other objects or even people.

5.1 Additional Discussion
There are perceived security concerns for developing technologies to cloak from person-tracking systems, for potential malicious uses. Although this is a valid concern to some degree, we argue that the risk is manageable for the following reasons: 1, The adversarial patterns are generated in a way that makes them easily differentiable from daily clothing, almost making a public statement that the wearer does not intend to be detected by person-tracking systems. They can be extremely noticeable and therefore discouraging for malicious parties to wear. 2, In critical areas where additional security is required, the InvisibiliTee model could be used in adversarial training to improve robustness against such attacks.

6 Conclusion
This paper presented a black-box attack method, InvisibiliTee, to perform digital and physical adversarial attacks on the human detection models for individual privacy preservation. An angle-agnostic scheme trained by the attack
and geometric constraint losses was proposed to generate adversarial patterns. Both digital and physical attacking experiments were conducted on a group of state-of-the-art human detection systems, demonstrating the effectiveness of the learned adversarial patterns. The results have shown that the InvisibiliTee can significantly reduce the average precision of person detection systems especially in the digital attacks. In addition, 3D adversarial attack is an emerging research. Therefore, it is challenging to extent InvisibiliTee to a 3D model.

References

1. Brown, B.T., Mané, D., Roy, A., Abadi, M., Gilmer, J.: Adversarial patch. CoRR (2017)
2. Cai, Z., Vasconcelos, N.: Cascade r-cnn: High quality object detection and instance segmentation. TPAMI p. 1–1 (2019). https://doi.org/10.1109/tpami.2019.2956516
3. Chen, K., Wang, J., Pang, J., Cao, Y., Xiong, Y., Li, X., Sun, S., Feng, W., Liu, Z., Xu, J., Zhang, Z., Cheng, D., Zhu, C., Cheng, T., Zhao, Q., Li, B., Lu, X., Zhu, R., Wu, Y., Dai, J., Wang, J., Shi, J., Ouyang, W., Loy, C.C., Lin, D.: MM Detection: Open mmlab detection toolbox and benchmark. arXiv preprint arXiv:1906.07155 (2019)
4. Dai, J., Li, Y., He, K., Sun, J.: R-fcn: object detection via region-based fully convolutional networks. In: NeurIPS. pp. 379–387 (2016)
5. Farhadi, A., Redmon, J.: Yolov3: An incremental improvement. In: CVPR. pp. 1804–2767. Springer Berlin/Heidelberg, Germany (2018)
6. Fu, C.Y., Liu, W., Ranga, A., Tyagi, A., Berg, C.A.: Dssd : Deconvolutional single shot detector. CVPR (2017)
7. Ge, Y., Zhang, R., Wu, L., Wang, X., Tang, X., Luo, P.: A versatile benchmark for detection, pose estimation, segmentation and re-identification of clothing images. CVPR (2019)
8. Girshick, R.: Fast r-cnn. In: CVPR. pp. 1440–1448 (2015)
9. Goodfellow, J.I., Shlens, J., Szegedy, C.: Explaining and harnessing adversarial examples. ICLR (2014)
10. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: CVPR. pp. 2961–2969 (2017)
11. Kim, K., Lee, H.S.: Probabilistic anchor assignment with iou prediction for object detection. In: ECCV (2020)
12. Law, H., Deng, J.: Cornernet: Detecting objects as paired keypoints. In: ECCV 2018. pp. 765–781. Springer Verlag (2018)
13. Lee, M., Kolter, Z.: On physical adversarial patches for object detection. CoRR (2019)
14. Lin, T.Y., Goyal, P., Girshick, B.R., He, K., Dollár, P.: Focal loss for dense object detection. ICCV pp. 318–327 (2017)
15. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S.: Ssd: Single shot multibox detector. ECCV pp. 21–37 (2016)
16. Liu, X., Yang, H., Liu, Z., Song, L., Chen, Y., Li, H.: Dpatch - an adversarial patch attack on object detectors. SafeAI@AAAI (2019)
17. Madry, A., Makelov, A., Schmidt, L., Tsipras, D., Vladu, A.: Towards deep learning models resistant to adversarial attacks. ICLR (2018)
18. Redmon, J., Divvala, K.S., Girshick, B.R., Farhadi, A.: You only look once: Unified, real-time object detection. CVPR (2016)
19. Redmon, J., Farhadi, A.: Yolo9000: Better, faster, stronger. CVPR (2017)
20. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: towards real-time object detection with region proposal networks. TPAMI 39(6), 1137–1149 (2016)
21. Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I.J., Fergus, R.: Intriguing properties of neural networks. ICLR (2014)
22. Thys, S., Van Ranst, W., Goedemé, T.: Fooling automated surveillance cameras: adversarial patches to attack person detection. In: CVPR (2019)
23. Uijlings, J.R., Van De Sande, K.E., Gevers, T., Smeulders, A.W.: Selective search for object recognition. IJCV 104(2), 154–171 (2013)
24. Wang, H., Wang, G., Li, Y., Zhang, D., Lin, L.: Transferable, controllable, and inconspicuous adversarial attacks on person re-identification with deep mis-ranking. In: CVPR (June 2020)
25. Wang, J., Zhang, W., Cao, Y., Chen, K., Pang, J., Gong, T., Shi, J., Loy, C.C., Lin, D.: Side-aware boundary localization for more precise object detection. In: ECCV (2020)
26. Wu, Z., Lim, S.N., Davis, L.S., Goldstein, T.: Making an invisibility cloak: Real world adversarial attacks on object detectors. In: ECCV. pp. 1–17. Springer (2020)
27. Yang, Z., Liu, S., Hu, H., Wang, L., Lin, S.: Reppoints: Point set representation for object detection. In: ICCV (Oct 2019)
28. Zhang, H., Chang, H., Ma, B., Wang, N., Chen, X.: Dynamic r-cnn: Towards high quality object detection via dynamic training. In: ECCV. pp. 260–275. Springer (2020)
29. Zhang, X., Wan, F., Liu, C., Ji, R., Ye, Q.: FreeAnchor: Learning to match anchors for visual object detection. In: NeurIPS (2019)
30. Zheng, Y., Lu, Y., Velipasašar, S.: An effective adversarial attack on person re-identification in video surveillance via dispersion reduction. IEEE Access 8, 183891–183902 (2020)