Thanks for your valuable comments. We will explain your concerns point by point.

1 Response to Reviewer #3

Q: Is it reasonable to apply a patch to the target to be identified?

There is no overlap between the real target and the patch. Specicically, the boundary of the patch and the real target is 16 pixels apart in our experiments settings.

Q: How can we know what is the object that the application is tracking?

In academic research, the target position of the first frame is given. In practical applicaiton, the target position of the first frame is often obtained by running an object detector on this frame. Once the target position of the first frame is obtained, we can put the adversarial patch near the target to mislead the tracker. In the next frames, we can specify an arbitrary *fake trajectory* to place the adversarial patch.

2 Response to Reviewer #10

Q: The university shall reside on trackers rather than videos.

There are two similar concepts in the field of adversarial attacks: *university* and *transferability*. The university of adversarial examples is proposed in [Moosavi-Dezfooli *et al.*, 2017], which means an adversarial example is image-agnostic. The transferability means the adversarial example can be used in different networks. We have analyzed the transferability to different backbones/tracking architectures in **Sec. 4.4**.

Q: During testing, the target trajectory is predicted by the tracker, rather than the ground truth. This shall be made clear in the paper.

We will clarify this in the camera-ready paper.

Q: Literature missing in the related works

We will add these missing references in the camera-ready paper.

3 Response to Reviewer #64

Q: CNN attacks are usually expected imperceptible but the proposed method has to add an obviously noticeable fake target patch to tracking frames.

Both imperceptible perturbations and dversarial patches [Brown *et al.*, 2017] are popular adversarial attack methods. In our work, the patch is small, and its area accounts for 4% of the searched image area. Making the patch looks like the background is a very meaningful future work.

4 Response to Reviewer #76

Q: The paper applies FGSM method to the exemplar and Brown et al.' work to the candidate image.

To the best of our knowledge, our work is the first attempt to generate video-agnostic perturbations to attack siamese trackers. Besides FGSM and Brown et al.' work, other adversarial example generation methods such as C&W [Carlini and Wagner, 2017] and PGD [Madry *et al.*, 2017] can also be integrated into our attacking system to further improve the attack effect. In short, we focus on proposing a video-agnostic attacking system for siamese trackers instead of prosing a specific adversarial example generation method.

5 Response to Reviewer #87

Q: How to initialize the imperceptible perturbation δ and the adversarial patch p?

Each element in δ is initialized to 0. p is initialized with a normal distribution, with a mean of 127 and a standard deviation of 1.

Q: Is only one imperceptible perturbation δ and one adversarial patch p attained? How to keep the δ and p video-agnostic and if the unique (δ, p) is really effective to various videos?

Only one δ and one p are obtained after end-to-end training. The unique (δ, p) works because we train it on large datasets including COCO, ILSVRC-VID and the training splits of GOT10k and LaSOT.

Q: I believe that how to use the proposed attack method to improve the object tracking will make this work more valuable.

In this paper, we focus on the attck to the trakcer, and increasing the robustness of trackers against attacks is a very meaningful future work.

References

[Brown *et al.*, 2017] Tom B Brown, Dandelion Mané, Aurko Roy, Martín Abadi, and Justin Gilmer. Adversarial patch. *arXiv preprint arXiv:1712.09665*, 2017.

[Carlini and Wagner, 2017] Nicholas Carlini and David Wagner. Towards evaluating the robustness of neural networks. In *2017 ieee symposium on security and privacy (sp)*, pages 39–57. IEEE, 2017.

[Madry *et al.*, 2017] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards deep learning models resistant to adversarial attacks. *arXiv preprint arXiv:1706.06083*, 2017.

[Moosavi-Dezfooli *et al.*, 2017] Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, Omar Fawzi, and Pascal Frossard. Universal adversarial perturbations. In *CVPR*, pages 1765–1773, 2017.