

## A Experimental Environment

All of the experiments are conducted on a server with AMD EPYC 9554 64-core Processor and an NVIDIA L40 GPU, running Ubuntu 22.04. During the physical evaluation, the patch is printed out with an EPSON L18050 printer, *i.e.*, a patch of physical size  $(u, v) = (0.891m, 1.26m)$  is printed out by filling A3 size paper.

## B Implementation Details

### B.1 Calculation for Digital Deployment

During the digital experiments, the patches are first digitally deployed into the scene. As our aim is to achieve physical attack, the digital deployment is required to follow the physical constraints. Fortunately, such a physical-constrained digital patch deployment can be realized with the calibration information provided by the dataset. We show the detail calculation for the KITTI dataset below.

For KITTI dataset, we suppose the patch’s physical size and depth are predefined with regard to the camera 0 (the reference camera). Given the patch with physical size of  $(w_p, h_p)$ , we intend to deploy it into a KITTI scene with physical depth  $e$ .

1. Get the corresponding calibration information for the scene from ‘calib\_cam\_to\_cam’ folder provided by KITTI scene flow dataset.
2. Retrieve the three rectified calibration matrix  $P_{rect\_02}$ ,  $P_{rect\_03}$ ,  $R_{rect\_00}$ .
3. Specify the physical shifting  $(x_{shift}, y_{shift})$  (in meters) of the patch center with regard to the camera 0 principal axis.
4. Set the homogeneous coordinates for the corners of the patch as:
  - $top\_left = (-w_p/2 + x_{shift}, -h_p/2 + y_{shift}, e, 1)$ ,
  - $top\_right = (w_p/2 + x_{shift}, -h_p/2 + y_{shift}, e, 1)$ ,
  - $bottom\_left = (-w_p/2 + x_{shift}, h_p/2 + y_{shift}, e, 1)$ ,
  - $bottom\_right = (w_p/2 + x_{shift}, h_p/2 + y_{shift}, e, 1)$ .
5. For the pixel coordinates of the patch’s corner in the right stereo image, get them with  $(P_{rect\_02} \times R_{rect\_00}) \cdot c$  where  $c$  is the corners homogeneous coordinates.
6. Similarly, get the corresponding pixel coordinates for the left stereo image with  $(P_{rect\_03} \times R_{rect\_00}) \cdot c$ .
7. Finally, the patch is deployed into the scene by applying perspective transformation to fit the patch into the calculated pixel region.

For the calculation of DrivingStereo, the process is the same as KITTI except for the reference camera is camera 1 thus  $P_{rect\_101}$ ,  $P_{rect\_103}$  and  $R_{rect\_101}$  from their calibration file is adopted.

### B.2 Physical Stereo Image Capture

During the experiments of physical evaluation of our patch, we have to manually capture the scene due to the lack of RGB stereo capturing device. We adopt the i3DSteroid mobile application which is specifically designed to facilitate the capturing of stereo images with mobile phone of model iPhone 14 pro. In specific, a cutting board of A3 size is utilized to place the mobile phone for a 30cm baseline simulation. Then at each place a picture is taken by the mobile phone as one of the stereo image. However, we are aware of the inaccuracy of this stereo image capturing process, which is why we further test our patch with a commercial RGB-D camera, *i.e.*, the IntelRealSense D435i.

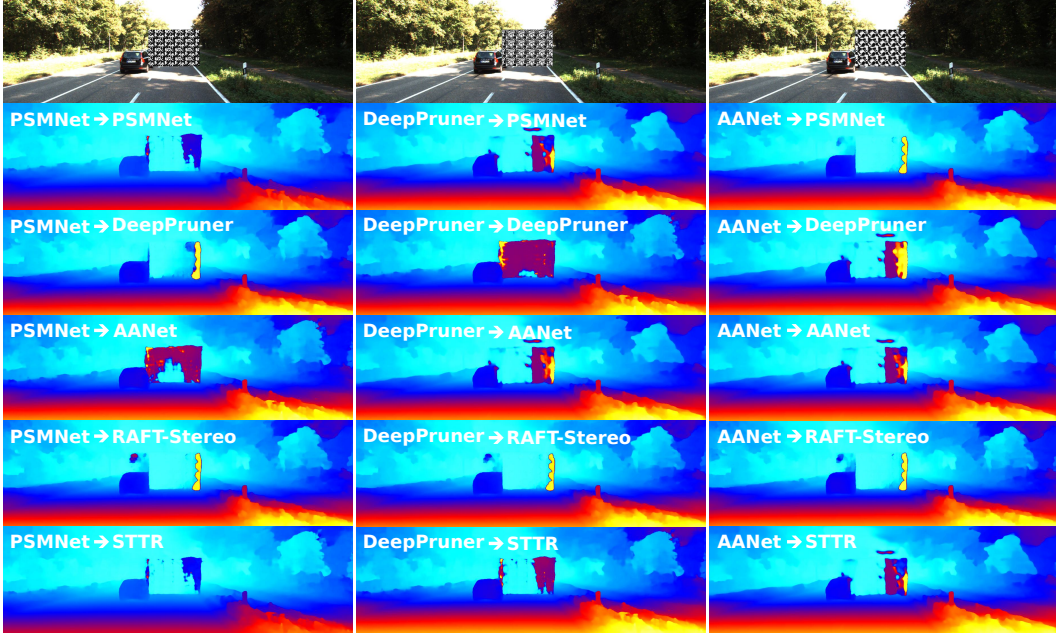


Figure 1: Illustration of i3DSteroid.

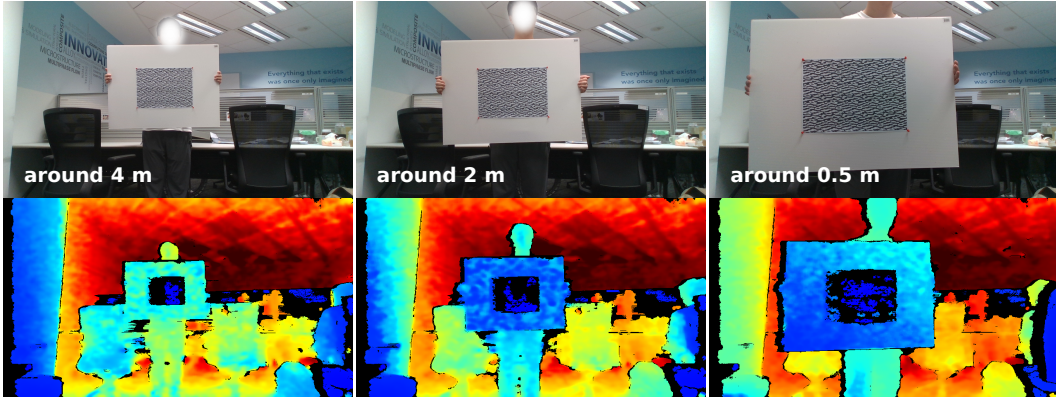
### B.3 Setting for Attack Targets

For the PSMNet, DeepPruner and AANet, we follow the setup of Stereoscopic [1] and directly adopt the API provided at <https://github.com/alexklwong/stereoscopic-universal-perturbations.git>. As for the RAFT-Stereo and STTR, we use the official code release and integrate them following Stereoscopic code. And the checkpoint pretrained on KITTI dataset for both models, *i.e.*, ‘raftstereo-sceneflow.pth’ and ‘kitti\_finetuned\_model.pth.tar’, are adopt for experiments. Note that we set  $k = 3$  for D1-error metric calculation during the evaluation for all the attack target models.





**Figure 4:** Visualization of attack transferability results of our DepthVanish patches against different stereo depth estimators on DrivingStereo scene.



**Figure 5:** Visualization of our DepthVanish patch's robustness across different distance.

### C.3 Distance Robustness

As we have verified the rotation robustness of our patch in the main experiments, here we further show our DepthVanish patch is also robust to distance. As shown in Fig. 5, our DepthVanish patch (optimized with RAFT-Stereo) remains attack effective with the variation of distance.

## D Limitation

Our research has several important limitations regarding the attack methodology. Due to a limited testing scope, we cannot guarantee that our attack methods are robust against the full spectrum of existing DNN architectures for stereo depth estimation. Additionally, we have not evaluated our attacks against existing robustified methods or defense mechanisms that may be implemented in real-world systems, which may overestimate the practical effectiveness of the vulnerabilities we identified. Regarding defensive approaches, our work lacks comprehensive solutions to the vulnerabilities demonstrated. While we discussed several potential defense strategies in the next section, we do not provide thoroughly tested, reliable defense methods with rigorous evaluation of their effectiveness or practical implementability. Furthermore, we have not explored the deeper theoretical principles underlying these vulnerabilities, which limits our ability to provide principled guidance for designing inherently robust depth estimation systems.

## E Broader Impact

**Scientific & Societal Benefit** In conducting our research on digital and physical attacks against stereo depth estimation models, we identify several important benefits to the scientific community and society. ❶ Through our identification of vulnerabilities in current depth estimation algorithms, we highlight critical weaknesses that need addressing before these systems are widely deployed in safety-critical applications. By revealing these issues in a controlled research setting, we enable improvements before real-world failures occur. ❷ Our work advances fundamental understanding of robustness in computer vision systems, particularly for depth perception, which is essential for autonomous vehicles, robotics, augmented reality, and medical imaging systems. This knowledge can lead to more resilient algorithms and implementations. ❸ Our physical attack demonstrations help bridge the gap between theoretical and practical security concerns, providing empirical evidence that can drive industry standards and testing protocols for vision-based systems before deployment.

**Misuse Potential & Security Concern** We acknowledge there are legitimate concerns about how this research could be misused, thus we carefully considered the ethical implications of revealing vulnerabilities in stereo depth estimation systems. Malicious actors might exploit the vulnerabilities we have identified to compromise autonomous navigation systems in vehicles or robots, potentially causing accidents or enabling theft/tampering of autonomous systems. As a result, we implemented several safety controls to minimize misuse risk, including: ❶ limited disclosure of specific technical details that could enable immediate exploitation; ❷ establishment of a reasonable timeline for patches before full disclosure; ❸ creation of a centralized database of proposed attacks accessible only to verified researchers and industry partners. We recommend similar safeguards for related research, including mandatory ethics review for attack demonstrations, the implementation of differential privacy techniques to limit what information is shared, and the development of standardized responsible disclosure protocols specific to vision system vulnerabilities. By maintaining these ethical standards and safety controls, we can continue advancing security research while minimizing potential harm to systems that increasingly underpin critical infrastructure and everyday technologies.

**Possible Defense** Based on our findings, we propose several potential defensive approaches that could help mitigate the vulnerabilities our research identifies. ❶ Ensemble approaches that combine multiple depth estimation techniques (e.g., stereo, monocular, LiDAR fusion) can reduce the effectiveness of attacks targeted at any single method. ❷ Adversarial training with examples similar to our attack vectors could significantly improve model robustness, especially if incorporating both digital and physical attack factors. ❸ Runtime anomaly detection systems that identify sudden or physically implausible changes in depth maps could flag potential attacks for secondary verification. Other possible direction includes physical hardening through careful sensor placement, multi-angle verification, and environmental controls could reduce the effectiveness of physical attacks in critical systems. We suggest regulatory frameworks requiring security testing against known attack vectors like those we have identified could help ensure systems meet minimum safety standards before deployment.