Supplementary Materials: Physical Attack for Stereo Matching

Paper #357

1 Summary of Contents

Here is supplementary material for 'Physical Attacks on Stereo matching'. More details on the experimental setup and dataset are presented in section 2. We present more results for PSMNet, AANet, STTR in section 3, including images and analysis. Supplement more experimental results of black-box attacks in section 4, including pictures and tables. More cross-domain generalization attack results are given in section 5. section 6 supplements datasets captured in real-world attacks and more cases.

2 Dataset and experimental setup

We use the Kitti raw dataset for training. The dataset was taken on September 26, 28, 29, 30 and October 3, 2011. We use stereo pairs captured on September 26 and September 28, totaling 20,015 pairs, accounting for about 80% of the total dataset. The Scene Flow dataset is a large scale synthetic dataset and provides dense ground truth disparity maps. So we use Scene Flow dataset to test generalization.

During training, we crop the image size to 384 $\times$512, set the batch size to 1, and set the random number seed to 0.

3 More discussion on White-box attacks

White-box attacks on stereo matching networks are very effective. As shown in **??**. We qualitatively observe the disparity map output before and after the network is attacked, as well as the error map, we can find that the patch's attack on AANet covers the patch area and the surrounding area of the patch. The attack of the patch on PSMNet is in the surrounding area of the patch, and the geometry and content extraction of PSMNet enables the network to better match the inside of the patch. The attack on STTR is designed in the whole map, and the affected area is large but scattered. This is caused by the global attention of STTR.

It is worth noting that when a small patch (here, a patch of 0.3% of the image size is used) is used to attack STTR, under the action of global attention, the disparity map output by STTR, its error is dispersed in the whole image. While the disparity map errors output by PSMNet and AANet are always local. Qualitative results can be seen from figs. 1 to 3.

4 More discussion on Black-box attacks

We present the black-box attack results against PSMNet, AANet, PSMNet in **??**. As can be seen from the table, successful black-box attacks are carried out on all three networks. Among them, the

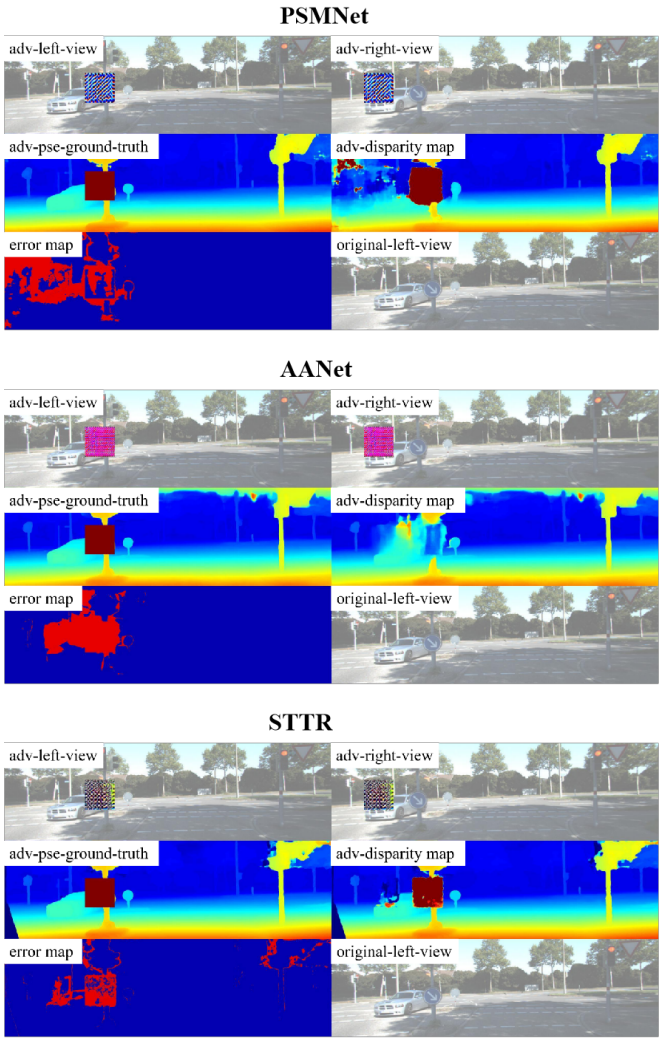

Figure 1. White-box attacks on PSMNet, AANet, STTR. The patch size used is 2.7% of the image size. Using feature correlation as an attack target. If not emphasized, the following pictures are the same as here.

patches optimized by PSMNet and STTR respectively have the most significant effect on the black-box attack of AANet. But we can also see that the attack effect of the patch in the black box attack has decreased. The attack effect of the natural checkerboard patch on PSMNet and STTR and the black-box attack effect of the optimized patch are similar in some indicators. Analyzing the architecture of the network, we found that PSMNet and STTR use 3D convolution

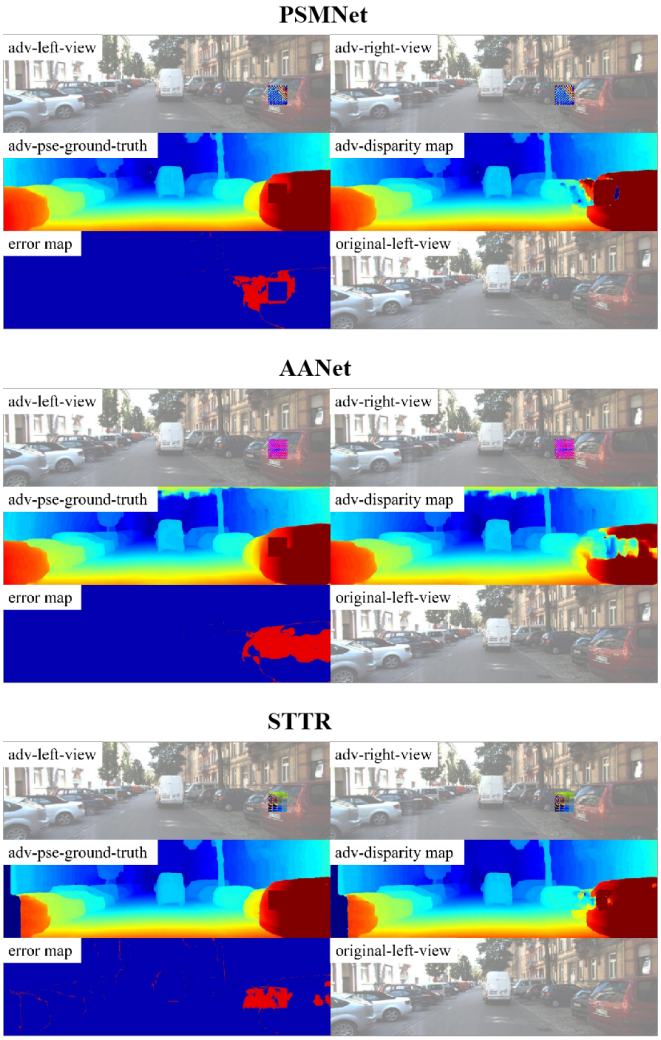

Figure 2. White-box attacks on PSMNet, AANet, STTR. The patch size used is 1.2% of the image size.

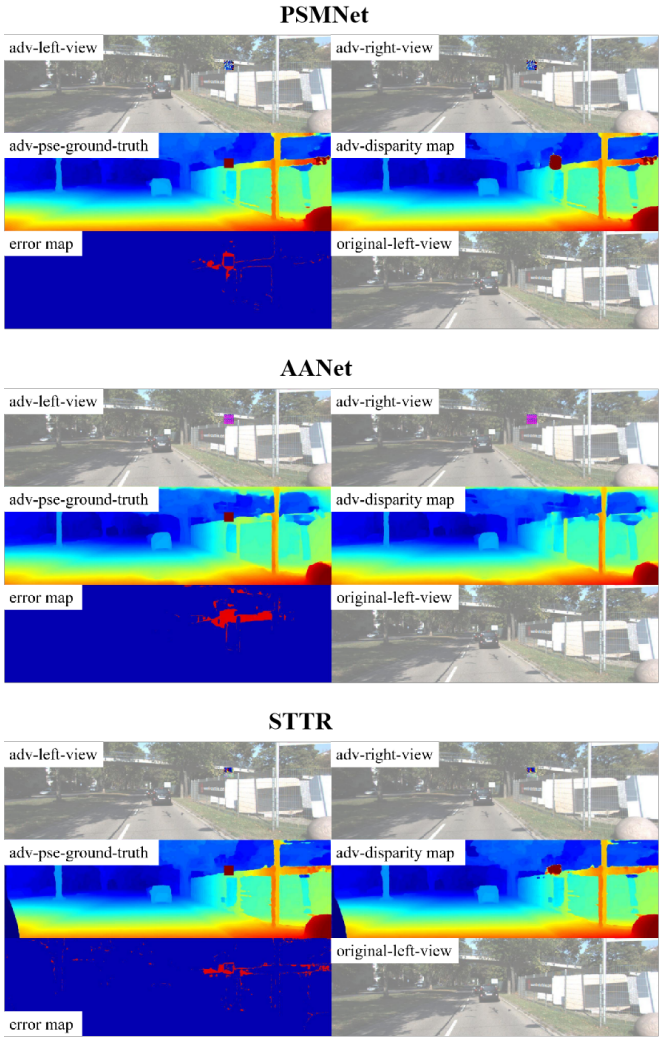

Figure 3. White-box attacks on PSMNet, AANet, STTR. The patch size used is 0.3% of the image size.

and attention mechanisms to optimize the network, respectively, and the larger receptive field and capture of geometric information can resist patch attacks optimized for other types of networks, capturing textures in patches to achieve the correct match. But this brings a lot of computing power, time, and storage consumption that need to be faced. How to develop a lightweight network that resists adversarial attacks, we leave it to later researchers. fig. 4 is the results of black-box attacks on all networks. It can be seen that in black-box attacks, different patches have similar effects on the same network.

5 More discussion on cross-domain generalization attacks

In the main text we discuss the cross-domain generalization attack on KITTI2012. Here we discuss cross-domain generalization attacks on Scene Flow. As in the setting of previous stereo matching methods, we attack 200 stereo pairs in the test set of FlyingThings3D. table 1 and fig. 6 show the qualitative and quantitative results of cross-domain generalization attacks on the scene flow dataset. It can be seen that for the sceneflow dataset, the optimized patches have achieved good attack results. It can be seen that the attack effect of the patch optimized by the feature attack is better than the patch of the direct attack and the natural adversarial examples. It can be seen that the optimization patch carries information such as vulnerable textures, so that a good attack effect can be obtained across datasets.

6 More discussion on Real-world attacks

We use the ZED 2 stereo camera for our experiments. The output resolution is set to 672×376. Since the depth ground truth cannot be obtained, we photographed two sets of stereo pairs as controls, and the two sets of pictures were taken in the same environment. A set of pictures is pasted with the printed patch, and a set of pictures is still the original environment. The location and angle of the print patch sticking are random. Eliminate unqualified photos such as inconsistencies before and after, we organize 2 groups of pictures, a total of 24 pairs of stereo pairs, 48 pictures. fig. 7 is a qualitative result of a real-world attack on all networks, and it can be seen that we only used the PSMNet-optimized patch, but the attack on all networks was very successful. This confirms the real-world threat of patch attacks. But patch attacks also have failed examples. The biggest problem is the reflection, which makes the patch texture disappear, so that the texture features of the patch have no effect on the network. However, the reflection of light leads to the appearance of large areas without

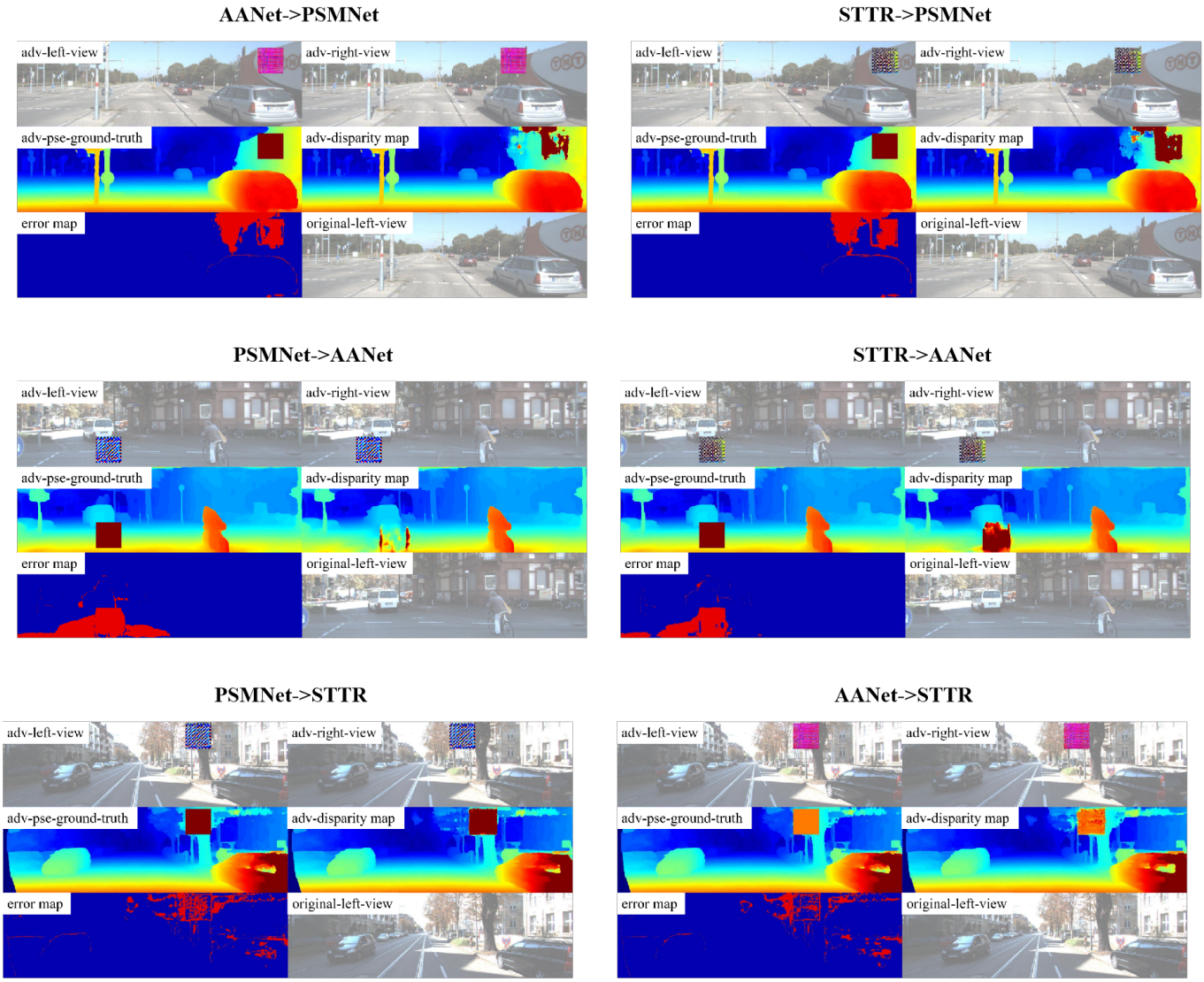

Figure 4. Black-box attacks on PSMNet, AANet and STTR.

Dataset				SCENE FLOW							
method	patch-init	loss	size (%)	D1-all (%)	EPE	adv-PeoD1-all (%)	adv-D1-all (%)	adv-EPE	R-adv-D1-all (%)	R-adv-EPE (%)	R-adv-PeoD1-all (%)
PSMNet	checker	cCV	2.7	6.192	5.412	3.399	6.252	5.835	2.22	15.67	125.89
		DM	2.7			2.931	6.244	5.732	1.93	11.85	108.56
	random		2.7			2.851	6.244	5.719	1.93	11.37	105.59
	checker		2.7			3.031	6.262	5.716	2.59	11.26	112.26
AANet	checker	cCV	2.7	2.981	0.9185	2.515	5.105	1.488	78.67	21.09	93.15
		DM	2.7			1.378	4.018	1.188	38.41	9.98	51.04
	random		2.7			0.931	3.609	1.088	23.26	6.28	34.48
	checker		2.7			2.175	4.843	1.363	68.96	16.46	80.56
STTR	checker	cCV	2.7	9.107	2.15	5.534	11.79	3.472	99.37	48.96	204.96
		DM	2.7			5.136	11.86	3.518	101.96	50.67	190.22
	random		2.7			2.773	10.27	3.168	43.07	37.70	102.70
	checker		2.7			4.923	11.8	3.613	99.74	54.19	182.33

Table 1. Cross-domain generalization attacks on scene flow dataset.

texture, which significantly reduces the accuracy of stereo matching, which is still one of the core problems that plague stereo matching. fig. 5 is an example of a failed attack. Comparing fig. 7 and fig. 5, we can see the impact of reflections on real-world attacks.

defensive measures				median blur							
patch-init	loss	size	blur kernel size	D1-all (%)	EPE	adv-PeoD1-all (%)	adv-D1-all (%)	adv-EPE	R-adv-D1-all (%)	R-adv-epe (%)	R-adv-PeoD1-all (%)
checker	cCV	2.7	5	3.04	1.063	15.89	11.22	3.859	302.96	103.56	588.52
		2.7	9			19.21	11.42	3.071	310.37	74.37	711.48
		2.7	13			24.44	15.42	3.49	458.52	89.89	905.19
		2.7	17			29.49	20.24	3.881	637.04	104.37	1092.22
		2.7	21			33.91	24.94	4.244	811.11	117.81	1255.93

defensive measures				JPEG compression							
patch-init	loss	size	JPEG_QUALITY	D1-all (%)	EPE	adv-PeoD1-all (%)	adv-D1-all (%)	adv-EPE	R-adv-D1-all (%)	R-adv-epe (%)	R-adv-PeoD1-all (%)
checker	cCV	2.7	80	3.04	1.063	19.01	13.62	3.292	391.85	82.56	704.07
		2.7	60			21.04	15.46	3.417	460.00	87.19	779.26
		2.7	40			23.52	17.47	3.552	534.44	92.19	871.11
		2.7	20			29.66	22.72	3.931	728.89	106.22	1098.52

defensive measures			adversarial training							
patch-init	loss	size	D1-all (%)	EPE	adv-PeoD1-all (%)	adv-D1-all (%)	adv-EPE	R-adv-D1-all (%)	R-adv-epe (%)	R-adv-PeoD1-all (%)
checker	cCV	2.7	21.87	3.298	12.93	30.22	6.686	309.26	125.48	478.89

Initial attack results										
patch-init	loss	size	D1-all (%)	EPE	adv-PeoD1-all (%)	adv-D1-all (%)	adv-EPE	R-adv-D1-all (%)	R-adv-epe (%)	R-adv-PeoD1-all (%)
checker	cCV	2.7	3.04	1.063	8.839	9.947	3.752	255.81	99.59	327.37

Table 2. Defenses against adversarial attacks against AANet.

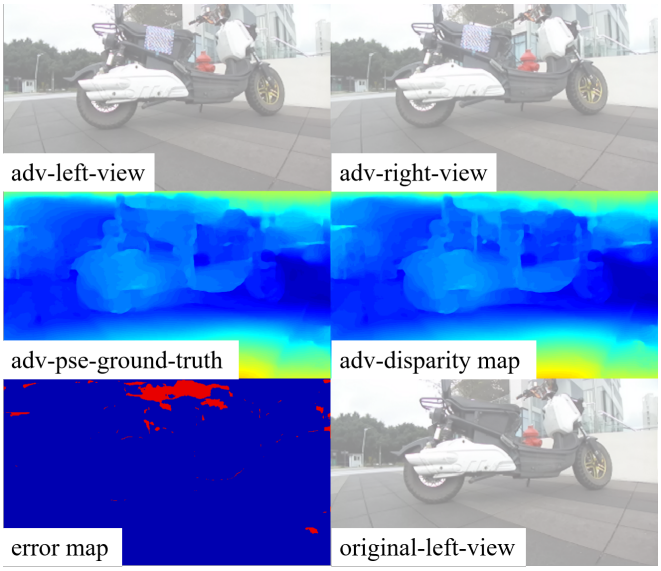

Figure 5. A Failed Case of Real-world attacks on AANet.

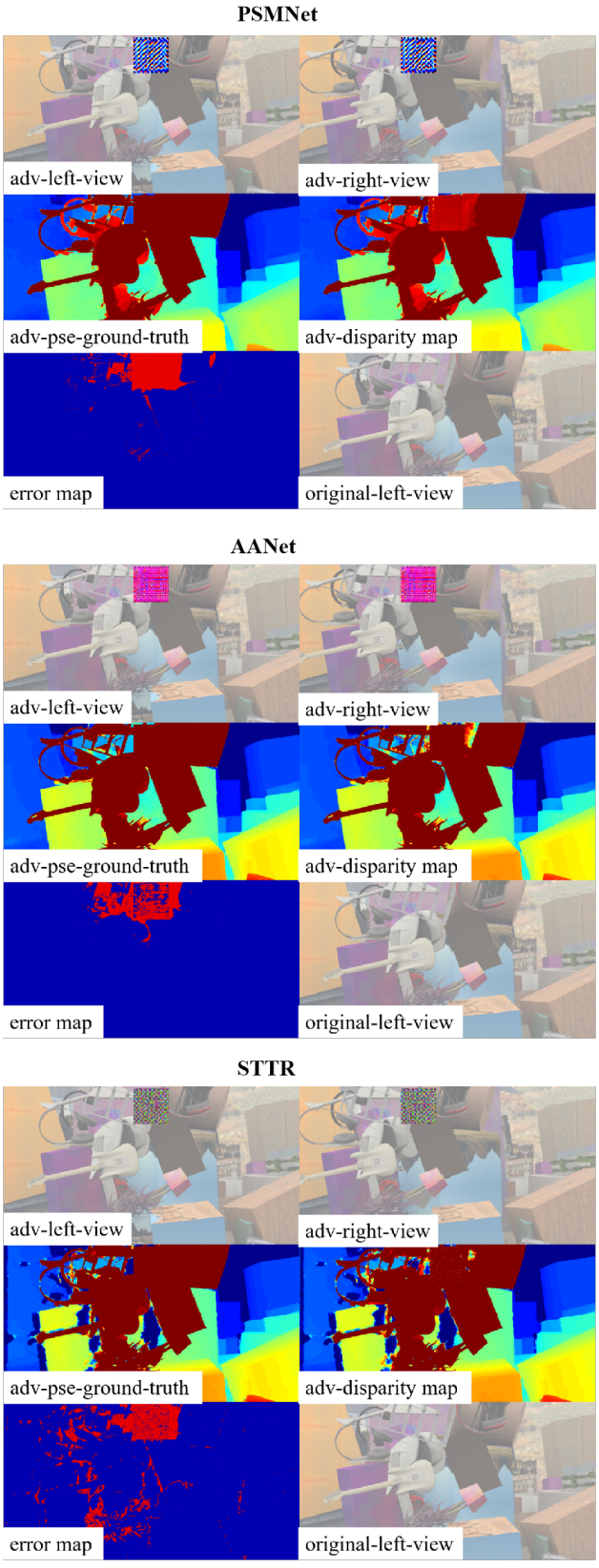

Figure 6. Cross-domain generalization attacks on scene flow dataset.

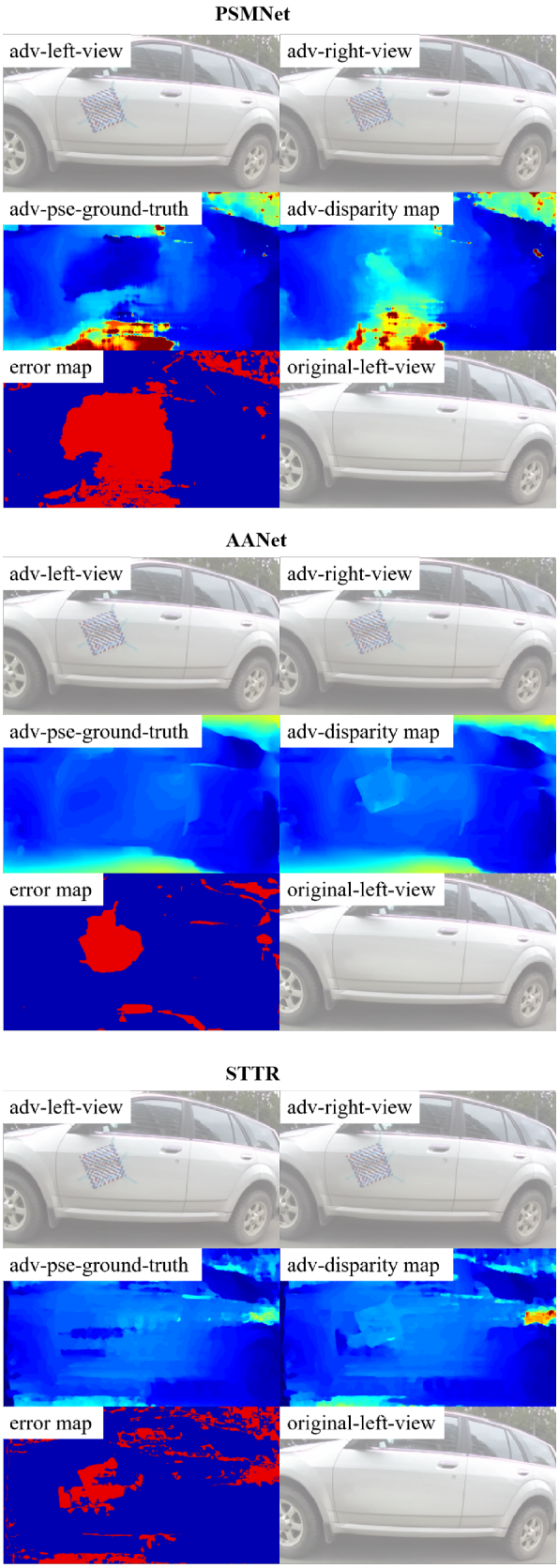

Figure 7. Real-world attacks on PSMNet, AANet, STTR.