Supplementary Material

1. Parameter Setup And Transformation Details

In this section, we provide more details about parameter setup. We use ADAM optimizer of Tensorflow platform in attack framework with default setup: $learning\ rate = 0.001$, $beta1 = 0.9$, , $beta2 = 0.999$ and $epsilon = 10^{-8}$. We set $\lambda_1 = 10.0$ and $\lambda_2 = 0.1$ in Eq. 6 as initialization. In Algorithm 1, we set fooling rate threshold $r_s = 0.95$, $iter_s = 100$ and the maximum iteration $iter_{max} = 2000$.

Total four methods include UPC**(ours)**, $ShapeShifter$, ERP^2 and $AdvPat$. FR-RES101-0712/FR-VGG16-0712 are used to generate the patterns for these four methods under same setting (*e.g.*, transformation parameters, training set and training epoch, etc).

Table 1. Distribution of transformations and parameters for UPC.

UPC**(ours)**					
	T_r			T_c	
Transform	Parameters	Remark	Transform	Parameters	Remark
Affine	$\mu = 0, \sigma = 0.1$	Perspective Transforms	Affine	$\mu = 0, \sigma = 0.03$	Deformed Simulation
Rotation	$-15^\circ \sim 15^\circ$	Camera Simulation	Cropping	$0.7 \sim 1.0$	Occlude Simulation
Contrast	$0.5 \sim 1.5$	Camera Parameters	Translation	$-0.04 \sim 0.04$	Pattern Location
Scale	$0.25 \sim 1.25$	Distance	Scale	$0.95 \sim 1.05$	Pattern Size
Brightness	$-0.25 \sim 0.25$	Illumination			

Table 2. Distribution of transformations and parameters for $ShapeShifter$.

$ShapeShifter$		
Transform	Parameters	Remark
Translation	$-0.2 \sim 0.2$	Perspective Transforms
Rotation	$-15^\circ \sim 15^\circ$	Camera Simulation
Contrast	$0.5 \sim 1.5$	Camera Parameters
Scale	$0.25 \sim 1.25$	Distance
Brightness	$-0.25 \sim 0.25$	Illumination

Table 3. Distribution of transformations and parameters for ERP^2.

ERP^2		
Transform	Parameters	Remark
Affine	$\mu = 0, \sigma = 0.1$	Perspective Transforms
Cropping	$0.9 \sim 1.2$	Photograph Simulation
Contrast	$0.5 \sim 1.5$	Camera Parameters

Table 4. Distribution of transformations and parameters for $AdvPat$.

$AdvPat$		
Transform	Parameters	Remark
Random Noise	$-0.15 \sim 0.15$	Noise
Rotation	$-15^\circ \sim 15^\circ$	Camera Simulation
Contrast	$0.5 \sim 1.5$	Camera Parameters
Scale	$0.8 \sim 1.2$	Resize
Brightness	$-0.25 \sim 0.25$	Illumination

2. Experiments in Physical World

2.1. Result of Stationary Testing

To evaluate the robustness of our method under different deformations, the person is required to switch from 6 different poses (*i.e.*, standing, sitting, leg lifting, waving hands, fork waist, shaking head) during photographing. We record the average precision $p_{0.5}$ and drop rates of FR-VGG16-0712 and FR-RES101-0712 under three brightness conditions in Table 5.

Table 5. Average precision $p_{0.5}$ in stationary testing after attacking faster r-cnn. We test on a total of 6 different poses (*i.e.*, standing, sitting, leg lifting, waving hands, fork waist, shaking head).

Network	FR-VGG16-0712							
Schemes	Standing				Sitting			
	L1	L2	L3	Avg (Drop)	L1	L2	L3	Avg (Drop)
Original	1.0	1.0	1.0	1.0 (-)	1.0	1.0	1.0	1.0 (-)
Natural	1.0	0.94	1.0	0.98 (0.02)	1.0	1.0	1.0	1.0 (0.0)
3-Patterns	0.72	0.61	0.67	0.67 (**0.33**)	0.83	0.78	0.67	0.76 (**0.24**)
7-Patterns	0.67	0.56	0.56	0.59 (**0.41**)	0.61	0.50	0.50	0.54 (**0.46**)
8-Patterns	0.22	0.11	0.17	0.17 (**0.83**)	0.28	0.17	0.22	0.22 (**0.78**)
Schemes	Fork Waist				Leg Lifting			
	L1	L2	L3	Avg (Drop)	L1	L2	L3	Avg (Drop)
Original	1.0	1.0	1.0	1.0 (-)	1.0	1.0	1.0	1.0 (-)
Natural	1.0	1.0	1.0	1.0 (0.0)	1.0	1.0	1.0	1.0 (0.0)
3-Patterns	0.78	0.72	0.67	0.72 (**0.28**)	0.72	0.78	0.72	0.74 (**0.26**)
7-Patterns	0.61	0.50	0.56	0.56 (**0.44**)	0.56	0.56	0.50	0.54 (**0.46**)
8-Patterns	0.28	0.17	0.17	0.20 (**0.80**)	0.28	0.28	0.22	0.26 (**0.74**)
Schemes	Rasing Hands				Shaking Head			
	L1	L2	L3	Avg (Drop)	L1	L2	L3	Avg (Drop)
Original	1.0	1.0	1.0	1.0 (-)	1.0	1.0	1.0	1.0 (-)
Natural	0.94	1.0	1.0	0.98 (0.02)	1.0	1.0	1.0	1.0 (0.0)
3-Patterns	0.89	0.78	0.83	0.83 (**0.17**)	0.78	0.78	0.67	0.74 (**0.26**)
7-Patterns	0.72	0.61	0.61	0.65 (**0.35**)	0.61	0.61	0.56	0.59 (**0.41**))
8-Patterns	0.39	0.39	0.28	0.35 (**0.65**)	0.22	0.28	0.11	0.20 (**0.80**)
Network	FR-RES101-0712							
Schemes	Standing				Sitting			
	L1	L2	L3	Avg (Drop)	L1	L2	L3	Avg (Drop)
Original	1.0	1.0	1.0	1.0 (-)	1.0	1.0	1.0	1.0 (-)
Natural	0.94	1.0	1.0	0.98 (0.02)	1.0	1.0	1.0	1.0 (0.0)
3-Patterns	0.83	0.67	0.67	0.72 (**0.28**)	0.72	0.78	0.72	0.74 (**0.26**)
7-Patterns	0.61	0.56	0.61	0.59 (**0.41**)	0.61	0.67	0.50	0.59 (**0.41**)
8-Patterns	0.22	0.22	0.11	0.19 (**0.81**)	0.28	0.22	0.22	0.26 (**0.74**)
Schemes	Fork Waist				Leg Lifting			
	L1	L2	L3	Avg (Drop)	L1	L2	L3	Avg (Drop)
Original	1.0	1.0	1.0	1.0 (-)	1.0	1.0	1.0	1.0 (-)
Natural	1.0	1.0	1.0	1.0 (0.0)	1.0	1.0	1.0	1.0 (0.0)
3-Patterns	0.83	0.72	0.72	0.76 (**0.24**)	0.67	0.78	0.67	0.71 (**0.29**)
7-Patterns	0.61	0.56	0.56	0.57 (**0.43**)	0.67	0.50	0.56	0.57 (**0.43**)
8-Patterns	0.28	0.22	0.22	0.24 (**0.76**)	0.33	0.33	0.22	0.30 (**0.70**)
Schemes	Rasing Hands				Shaking Head			
	L1	L2	L3	Avg (Drop)	L1	L2	L3	Avg (Drop)
Original	1.0	1.0	1.0	1.0 (-)	1.0	1.0	1.0	1.0 (-)
Natural	1.0	1.0	1.0	1.0 (0.0)	1.0	1.0	1.0	1.0 (0.0)
3-Patterns	0.83	0.89	0.83	0.85 (**0.15**)	0.72	0.78	0.78	0.76 (**0.24**)
7-Patterns	0.89	0.61	0.56	0.69 (**0.31**)	0.56	0.61	0.56	0.57 (**0.43**)
8-Patterns	0.39	0.33	0.33	0.35 (**0.65**)	0.22	0.22	0.17	0.20 (**0.80**)

2.2. Qualitative Samples of Physical Experiments

In this section, we provide more qualitative results of FR-VGG16-0712 and FR-RES101-0712 in physical environment in Figure 1. The detection result further shows our attack is invariant to different viewing conditions (*e.g.*, viewpoints, brightness).

Figure 1. **More qualitative results of FR-VGG16-0712 and FR-RES101-0712 on in physical environment.** These universal camouflage patterns are generated using FR-VGG16-0712 and FR-RES101-0712, respectively. Each row applies different pattern schemes (*i.e.*, 8/7/3-Pattern schemes), and captured in different viewpoints and background environments.

3. Experiments for Defense Methods

In order to test whether our proposed UPC can evade from defense methods, we introduce four state-of-the-art defense methods (*i.e.*, HGD [2], Randomization [4], Transformations [1] and Deflection [3]) as attacked target. In our experiment, FR-VGG16-0712 and FR-RES101-0712 are used to compute camouflage patterns. We record the experiment results of 8-Pattern schemes in Table 6. The original rendered images are used to calculate the baseline precision of each method (denoted as "w/o defense" in Table 6). The qualitative results of this experiment are displayed in Figure 2.

We observe that the precisions $\hat{p}_{0.5}$ of all defenses stays at a low level, which means the our proposed UPC successful breaks these state-of-the-art defend methods. An interesting finding is that some methods (*i.e.*, WAVE, TVM) improve the fooling ability of UPC instead of defending against the attacks (see result of "Sitting" column).

HGD [2]. We utilize trained model (*i.e.*, HGD-inception-v3, HGD-resnet-152) to denoise the rendered images.

Randomization [4]. In our experiment, we resize the rendered images to $299 \times 299 \times 3$, and then transform the images (*i.e.*, random resizing, zero-padding) as the input feed into the detectors, denote as RAND.

Transformation [1]. Here we test two defense method (*i.e.*, bit depth reduction (denoted as BIT) and JPEG compression (denoted as JPEG)). For BIT, we set reduce bits number $n = 5$ (*i.e.*, for each pixel value, we reduces last 5 bits). For JPEG, we set compression at quality level $q = 0.75$.

Deflection [3]. We consider four different settings: 1) we use the pixel deflection mechanism alone to defend UPC, denoted as DEF ; 2) we combine pixel deflection with wavelet denoising, denoted as WAVE ($\sigma = 0.04$); 3) pixel deflection and total variance minimization are combined, denoted as TVM; 4) pixel deflection and bilateral filters are composited, denoted as BIL ($size = 5, bins = 1000$). We set pixel deflection number $n = 2000$ during the experiments.

Evaluation Metrics. During the evading experiment, we use a the evaluation metric $\hat{p}_{0.5}$ to evaluate the attack performance:

$$\hat{p}_{0.5} = \frac{1}{|\mathcal{X}|} \sum_{v\sim\mathbb{V}, b\sim\mathbb{B}, s\sim\mathbb{S}} \left\{ \underset{x\in\mathcal{X}}{C(x)} = y, \underset{\hat{x}\in\hat{\mathcal{X}}}{C(\mathcal{D}(\hat{x}))} = y \right\}, \tag{1}$$

where x is normal person and $\hat{x}$ denotes person with camouflage pattern, $\mathcal{D}$ is the defense method we mentioned before, $\mathbb{V}, \mathbb{L}, \mathbb{S}$ denote camera viewpoints, brightness and scenes, respectively; C is the prediction of detector and y is the true label.

Table 6. Average precision $p_{0.5}$ of defense methods for attacking faster r-cnn. We test on a total of 4 different method (*i.e.*, HGD [2], Randomization [4], Transformation [1] and Deflection [3]).

Network	FR-VGG16-0712											
Defend Method	Standing				Walking				Sitting			
	L1	L2	L3	Avg (Drop)	L1	L2	L3	Avg (Drop)	L1	L2	L3	Avg (Drop)
w/o defense	0.15	0.03	0.02	0.07 (**0.91**)	0.06	0.05	0.01	0.04 (**0.91**)	0.60	0.47	0.32	0.46 (**0.52**)
HGD-INC [2]	0.38	0.20	0.02	0.20 (**0.78**)	0.11	0.05	0.03	0.06 (**0.89**)	0.70	0.59	0.51	0.60 (**0.38**)
HGD-RES [2]	0.40	0.22	0.02	0.21 (**0.77**)	0.11	0.05	0.02	0.06 (**0.89**)	0.74	0.60	0.54	0.63 (**0.35**)
RAND [4]	0.28	0.07	0.01	0.12 (**0.86**)	0.09	0.05	0.01	0.05 (**0.90**)	0.63	0.51	0.40	0.51 (**0.47**)
BIT [1]	0.56	0.28	0.0	0.28 (**0.70**)	0.26	0.04	0.03	0.11 (**0.84**)	0.50	0.50	0.32	0.44 (**0.54**)
JPEG [1]	0.40	0.14	0.02	0.17 (**0.81**)	0.10	0.09	0.01	0.07 (**0.88**)	0.67	0.56	0.38	0.54 (**0.44**)
DEF [3]	0.22	0.05	0.02	0.10 (**0.88**)	0.06	0.07	0.02	0.05 (**0.90**)	0.60	0.47	0.32	0.46 (**0.52**)
WAVE [3]	0.18	0.13	0.0	0.11 (**0.87**)	0.14	0.07	0.02	0.08 (**0.87**)	0.13	0.19	0.21	0.18 (**0.80**)
TVM [3]	0.60	0.40	0.01	0.34 (**0.64**)	0.23	0.13	0.02	0.13 (**0.82**)	0.57	0.57	0.45	0.53 (**0.45**)
BIL [3]	0.30	0.18	0.01	0.16 (**0.82**)	0.12	0.11	0.05	0.09 (**0.86**)	0.70	0.55	0.45	0.56 (**0.42**)
Network	FR-RES101-0712											
Defend Method	Standing				Walking				Sitting			
	L1	L2	L3	Avg (Drop)	L1	L2	L3	Avg (Drop)	L1	L2	L3	Avg (Drop)
w/o defense	0.10	0.09	0.13	0.11 (**0.88**)	0.05	0.06	0.06	0.06 (**0.93**)	0.49	0.57	0.62	0.56 (**0.43**)
HGD-INC [2]	0.33	0.24	0.30	0.29 (**0.70**)	0.03	0.03	0.04	0.03 (**0.96**)	0.50	0.61	0.55	0.55 (**0.44**)
HGD-RES [2]	0.29	0.23	0.24	0.25 (**0.74**)	0.04	0.04	0.03	0.04 (**0.95**)	0.50	0.65	0.56	0.57 (**0.42**)
RAND [4]	0.27	0.27	0.26	0.27 (**0.72**)	0.10	0.08	0.05	0.08 (**0.91**)	0.50	0.61	0.64	0.58 (**0.41**)
BIT [1]	0.50	0.12	0.08	0.23 (**0.76**)	0.19	0.04	0.03	0.9 (**0.90**)	0.40	0.48	0.45	0.44 (**0.55**)
JPEG [1]	0.25	0.13	0.19	0.19 (**0.80**)	0.05	0.04	0.01	0.03 (**0.96**)	0.44	0.51	0.50	0.48 (**0.51**)
DEF [3]	0.13	0.05	0.15	0.11 (**0.87**)	0.04	0.03	0.04	0.04 (**0.95**)	0.45	0.57	0.58	0.54 (**0.45**)
WAVE [3]	0.12	0.16	0.20	0.16 (**0.83**)	0.11	0.01	0.04	0.05 (**0.94**)	0.03	0.12	0.21	0.12 (**0.87**)
TVM [3]	0.50	0.25	0.18	0.31 (**0.68**)	0.20	0.08	0.02	0.10 (**0.89**)	0.17	0.38	0.33	0.29 (**0.70**)
BIL [3]	0.49	0.27	0.35	0.37 (**0.62**)	0.18	0.15	0.07	0.13 (**0.86**)	0.64	0.75	0.74	0.71 (**0.28**)

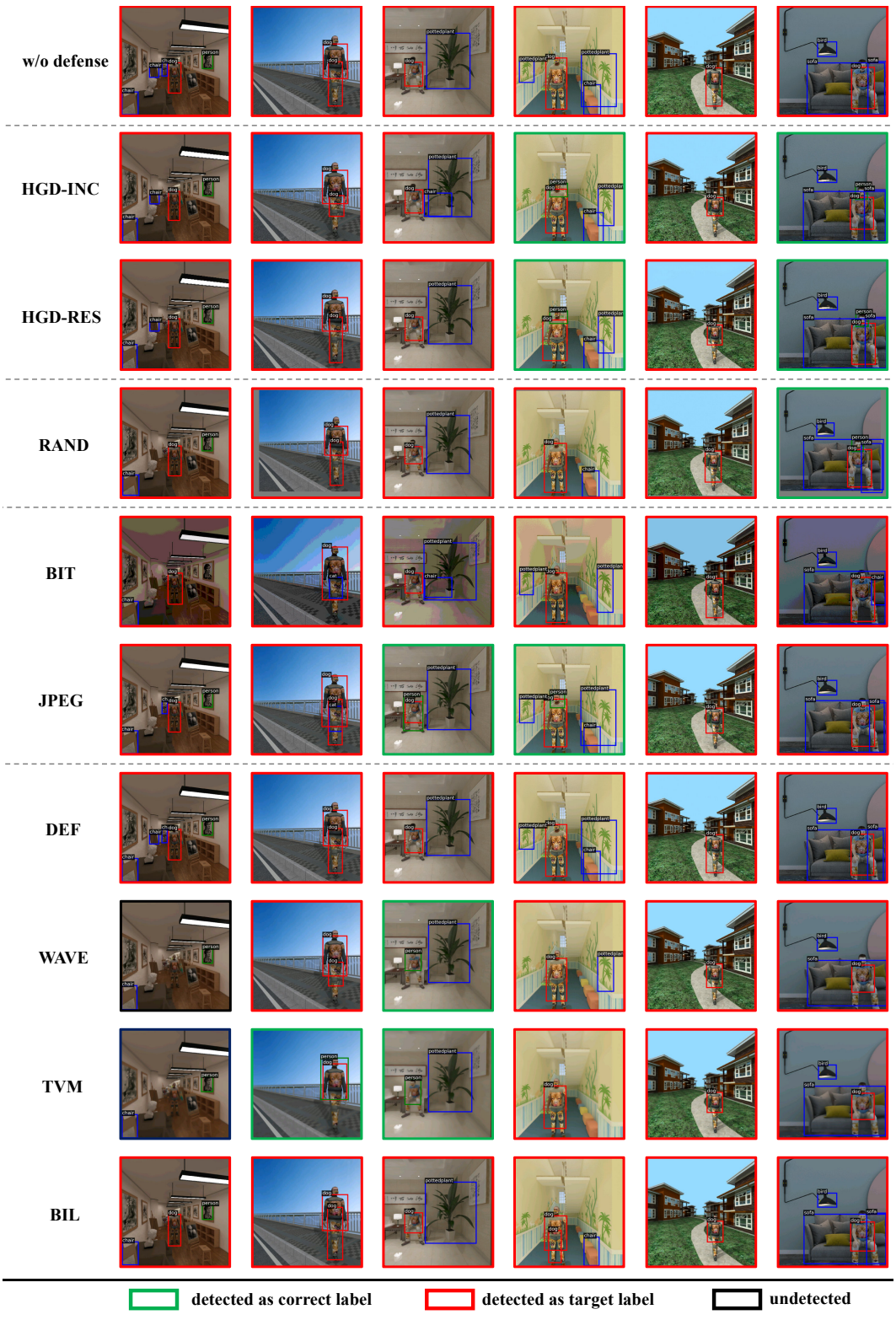

Figure 2. **Sampled qualitative results of experiments for defense methods.** We test four state-of-the-art defense methods (*i.e.*, HGD [2], Randomization [4], Transformations [1] and Deflection [3]). Both quantitative and qualitative result shows these methods can not defend against UPC effectively.

4. Experiments in Virtual Scenes

Here we provide more qualitative results of FR-VGG16-0712 and FR-RES101-0712 in the synthesized virtual environments.

4.1. Qualitative Samples of Existing Methods

In this section, we demonstrate several qualitative results of different methods (*i.e.*, $Shape$, ERP^2, $AdvPat$ and our proposed UPC) in Figure 3.

Figure 3. **More qualitative results under different attack settings in virtual experiments.** Each column uses same physical conditions (*i.e.*, lighting, viewpoints, environment, *etc.*). The camouflage patterns generated from UPC_{rc} achieve the most superior performance and visually similar to natural image, which can be regarded as pattern designs on human accessories.

4.2. Detection Result of Various Physical Conditions

Sampled results captured in different physical conditions (*e.g.*, brightness, background environments) are shown in Figure 4, which further show the power of proposed method.

Figure 4. **More qualitative results of FR-VGG16-0712 and FR-RES101-0712 on virtual environment.** Each row set different virtual environments with the same viewpoint of camera, and each column uses different lighting condition.

4.3. Experiment Results of Other Labels

We show some results of targeting other categories in Figure 5.

Figure 5. **More qualitative results of targeting other categories.** Each row applies different patterns (*i.e.*, boat/car/cat/horse), and captured in different viewpoints and background environments.

5. Visualization Study

In order to study the intrinsic reasons for different performance between existing methods, we visualize the feature maps from last convolutional layer to represent discriminative regions (see Figure 6). We observe that natural scheme can not change the detectors attention though patterns occlude some parts of human body. On the contrary, total four attacks attempt to fool detectors by activating patterns' feature. However, the visualization result demonstrate existing methods (*i.e.*, $Shape$, ERP^2, $AdvPat$) can not depress the activated features of un-occluded parts (*i.e.*, face, hand) effectively, which may lead higher detection accuracy.

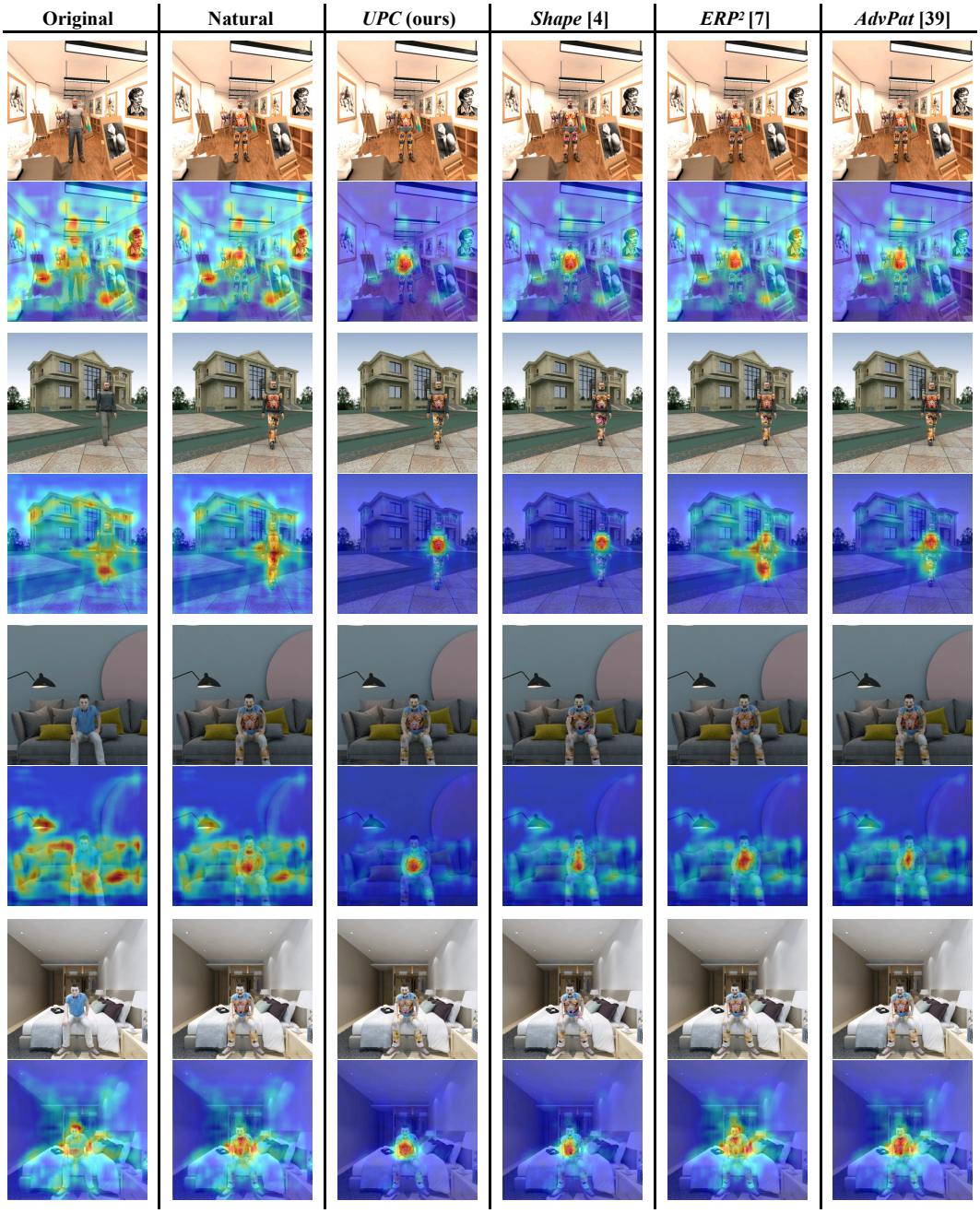

Figure 6. **Visualization of discriminative regions between different methods (*i.e.*, $Shape$, ERP^2, $AdvPat$ and UPC).** The rendered images include Standing/Walking/Sitting poses, and captured in different viewpoints and background environments.

Here we show some visualization results for different pattern schemes and poses both in physical world and virtual scenes (Figure 7). We can observe that the feature of face or hands will be activated when less surface area covered.

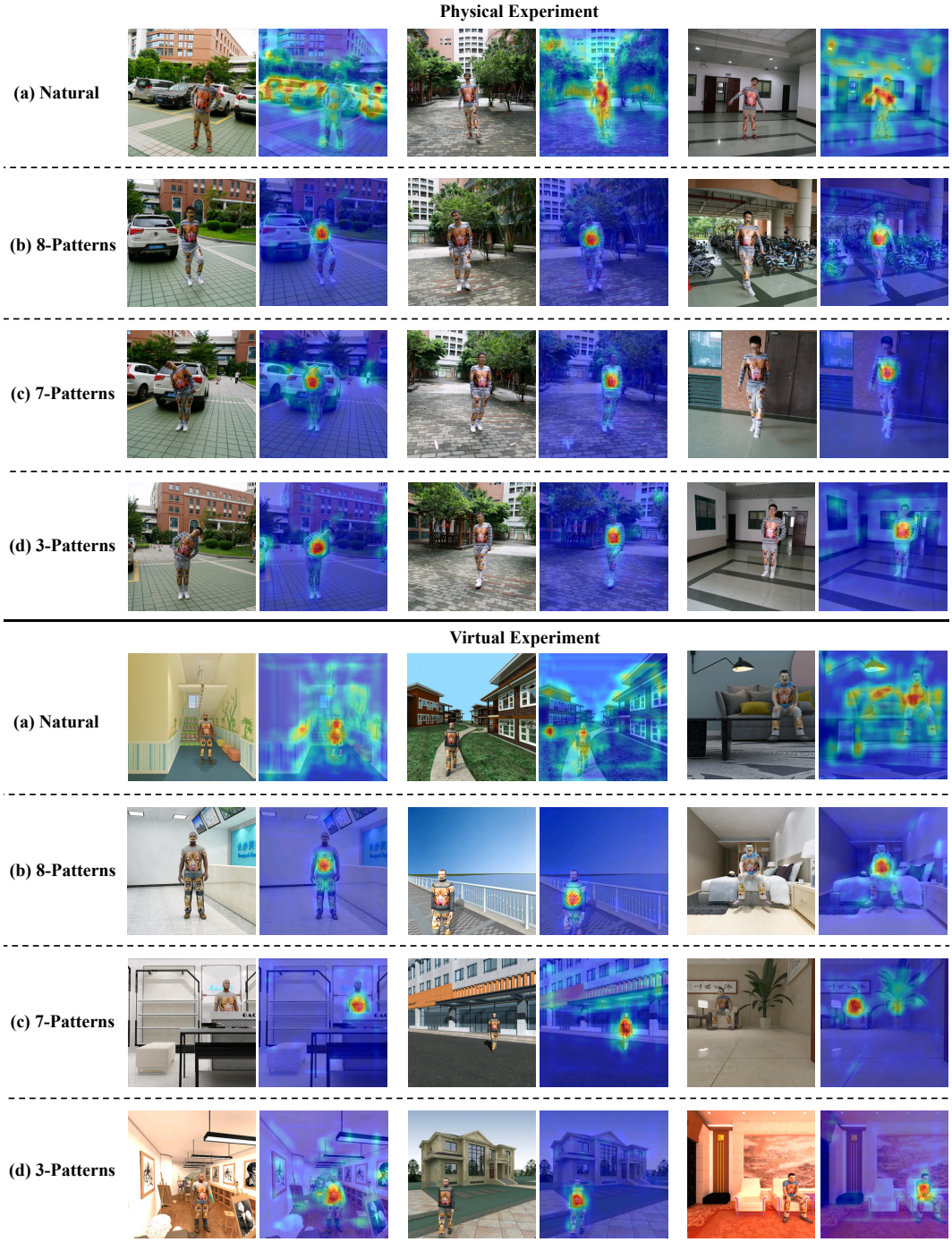

Figure 7. **Visualization of discriminative regions of proposedd UPC between different pattern schemes and poses.** Both physical and virtual results demonstrate similar trends under various physical conditions.

6. Generalization to Other Categories

In this section, we show some qualitative results of UPC on fooling the "car" category in both virtual scenes and physical world. Video of this experiment is available in the supplemental files.

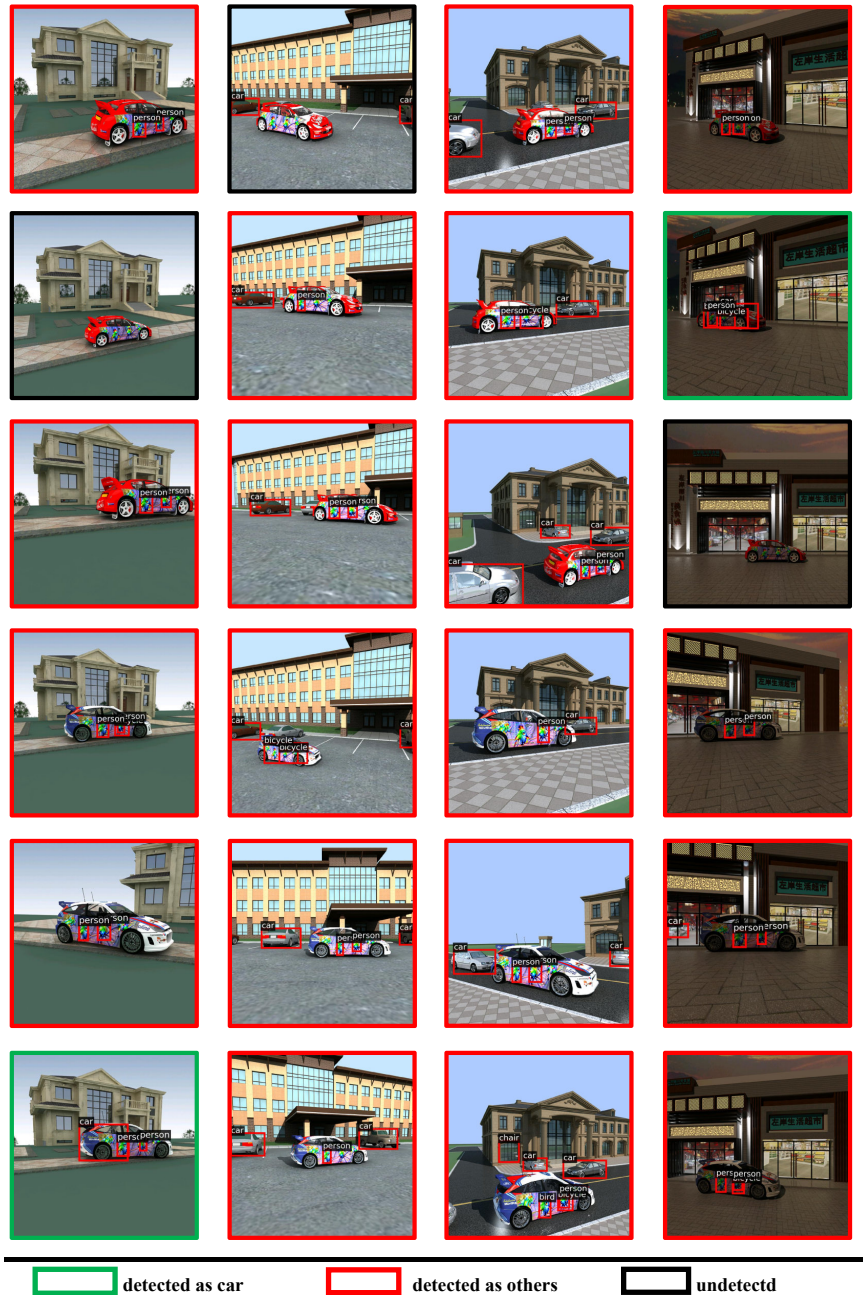

Figure 8. **More qualitative results of attacking the "car" category in virtual scenes.** We use two different car models (red car in top three rows and white car in bottom three rows) to evaluate the generalizablity of UPC.

Figure 9. **More experimental results of fooling the "car" category in physical world.** We attack two different cars, *i.e.*, Volvo XC60 and Volkswagen Tiguan.

References

[1] Chuan Guo, Mayank Rana, Moustapha Cisse, and Laurens Van Der Maaten. Countering adversarial images using input transformations. *arXiv preprint arXiv:1711.00117*, 2017.

[2] Fangzhou Liao, Ming Liang, Yinpeng Dong, Tianyu Pang, Xiaolin Hu, and Jun Zhu. Defense against adversarial attacks using high-level representation guided denoiser. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1778–1787, 2018.

[3] Aaditya Prakash, Nick Moran, Solomon Garber, Antonella DiLillo, and James Storer. Deflecting adversarial attacks with pixel deflection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8571–8580, 2018.

[4] Cihang Xie, Jianyu Wang, Zhishuai Zhang, Zhou Ren, and Alan Yuille. Mitigating adversarial effects through randomization. *arXiv preprint arXiv:1711.01991*, 2017.