

CNCA: Toward Customizable and Natural Generation of Adversarial Camouflage for Vehicle Detectors (**NeurIPS 2024**)

Linye Lyu, Jiawei Zhou, Daojing He, Yu Li



(a) Normal

(b) Colorful Graffiti

(c) Zebra Stripes

(d) Colorful Balls

(e) Yellow Black Graffiti

Background

- Deep neural networks are widely used in computer vision tasks.
- Physical adversarial attacks threaten DNN-based systems in the real world.
- Unnatural and conspicuous physical adversarial camouflages.



FCA [1]



DTA [2]



ACTIVE [3]

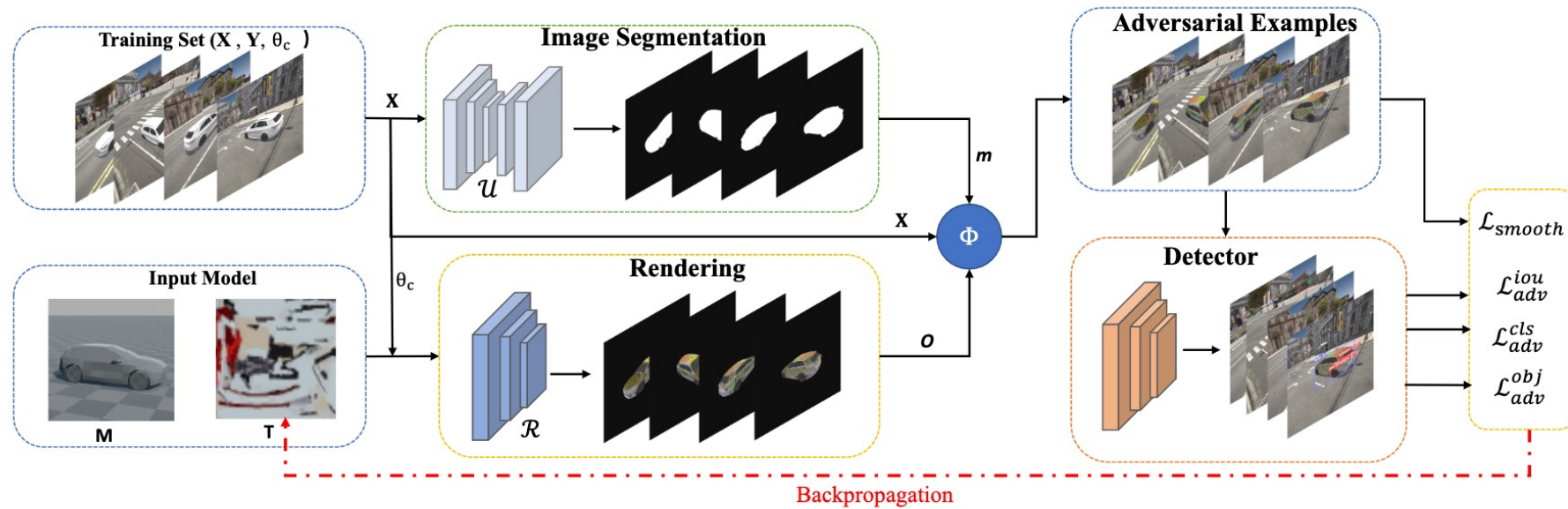
[1] Wang, Donghua, et al. "Fca: Learning a 3d full-coverage vehicle camouflage for multi-view physical adversarial attack." Proceedings of the AAAI conference on artificial intelligence. Vol. 36. No. 2. 2022.

[2] Suryanto, Naufal, et al. "Dta: Physical camouflage attacks using differentiable transformation network." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.

[3] Suryanto, Naufal, et al. "Active: Towards highly transferable 3d physical camouflage for universal and robust vehicle evasion." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

Related Work

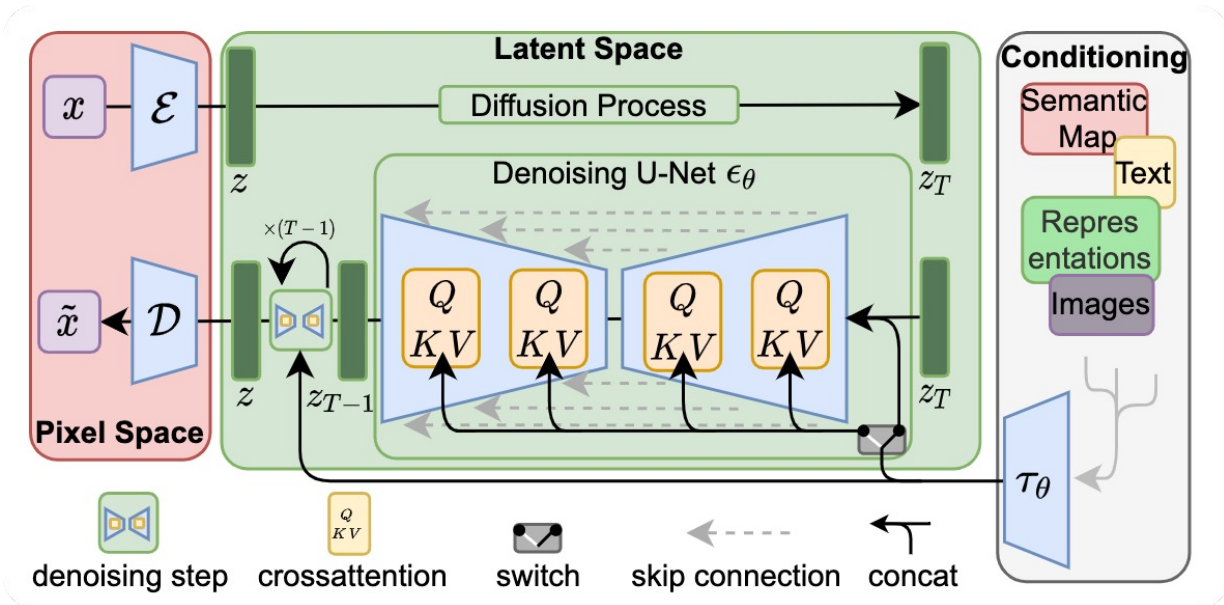
Adversarial Camouflage Generation Framework



FCA [1]

Related Work

Diffusion Models



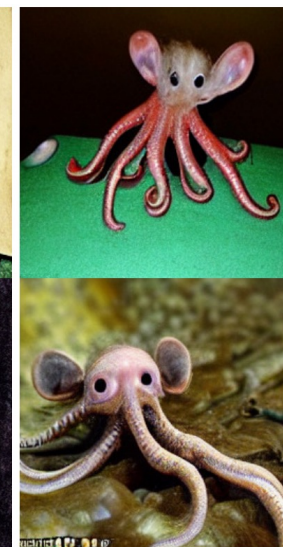
'A street sign that reads "Latent Diffusion"'



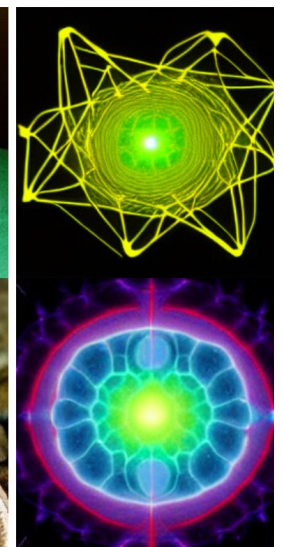
'A zombie in the style of Picasso'



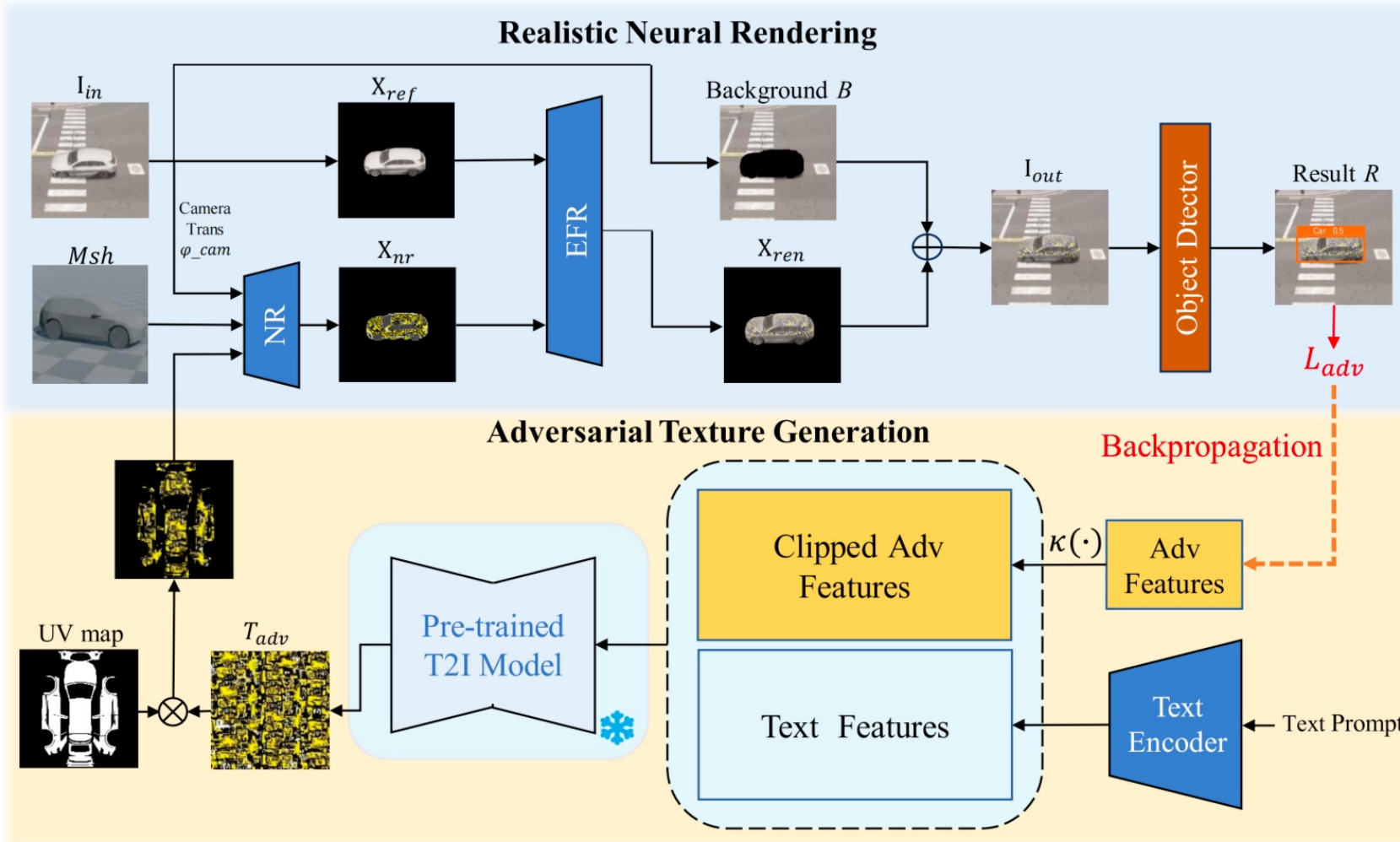
'An image of an animal half mouse half octopus'



'An illustration of a slightly conscious neural network'



CNCA Framework



- Differentiable neural rendering
- Environment feature rendering
- User-specific prompt to control texture content
- Adversarial features for attack signal
- Clipping strategy to balance attack and naturalness

Results: Digital World Attack

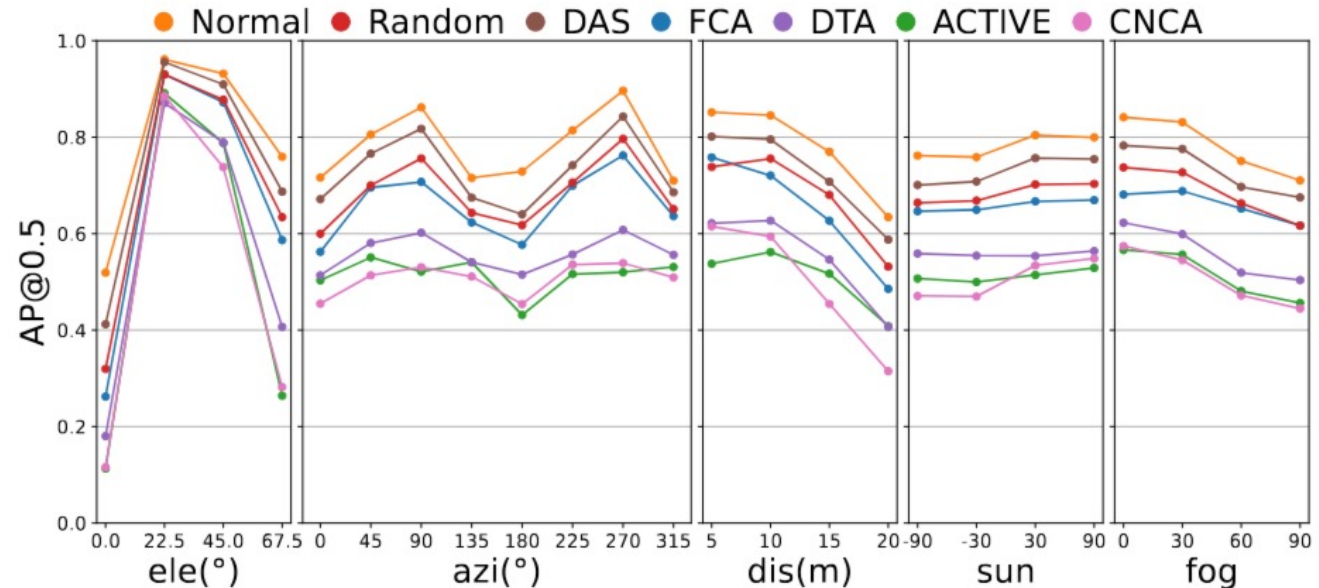
- Different detectors

- single stage
- two stage

METHODS	SINGLE-STAGE			TWO-STAGE			TOTAL
	YOLOv3	YOLOF	DDTR	DRCN	SRCN	FRRCN	
NORMAL	0.712	0.824	0.803	0.778	0.786	0.771	0.779
RANDOM	0.642	0.753	0.625	0.694	0.681	0.672	0.678
DAS	0.671	0.769	0.738	0.715	0.724	0.719	0.723
FCA	0.581	0.725	0.603	0.678	0.642	0.668	0.650
DTA	0.521	0.657	0.402	0.614	0.488	0.562	0.541
ACTIVE	0.473	0.577	0.436	0.534	0.484	0.520	0.504
CNCA	0.485	0.538	0.436	0.536	0.470	0.504	0.495







- Different physical settings

- view angles
- weathers









Results: Subjective Naturalness Evaluation

Table 4: Subjective tests for the naturalness evaluation of our adversarial camouflage with other baselines. The naturalness score is scaled from 1(not natural at all) to 5(very natural). As shown in the results, our score is significantly higher than the other four advanced adversarial camouflages.

IMAGES						
SCORE	4.68 ± 0.67	2.05 ± 1.03	2.11 ± 1.17	1.86 ± 0.98	1.84 ± 0.99	2.84 ± 1.09
SOURCE	NORMAL	DAS	FCA	DTA	ACTIVE	CNCA(OURS)

Results: Customizable Adversarial Camouflage

Table 3: Customizable camouflages with different text prompts. The AP@0.5 of each camouflage is shown below. The normal car texture baseline is 0.712.

CUSTOMIZABLE CAMOUFLAGE GENERATION WITH USER-SPECIFIC TEXT PROMPT.					
<i>colorful graffiti</i>	<i>yellow black graffiti</i>	<i>colorful camouflage</i>	<i>colorful balls</i>	<i>snake texture</i>	<i>zebra strips</i>
					
0.508	0.479	0.485	0.516	0.486	0.561

Results: Physical World Attack

Table 2: AP@0.5 of the different methods in the physical world evaluation.

METHODS	YOLOv3	YOLOX	SSD	CENTERNET	RETINANET	TOTAL
NORMAL	0.778	0.936	0.890	0.916	0.981	0.900
DAS	0.734	0.878	0.847	0.873	0.955	0.857
FCA	0.560	0.770	0.767	0.798	0.921	0.763
DTA	0.566	0.689	0.786	0.854	0.886	0.756
ACTIVE	0.518	0.563	0.574	0.743	0.735	0.627
CNCA	0.439	0.464	0.557	0.698	0.780	0.588

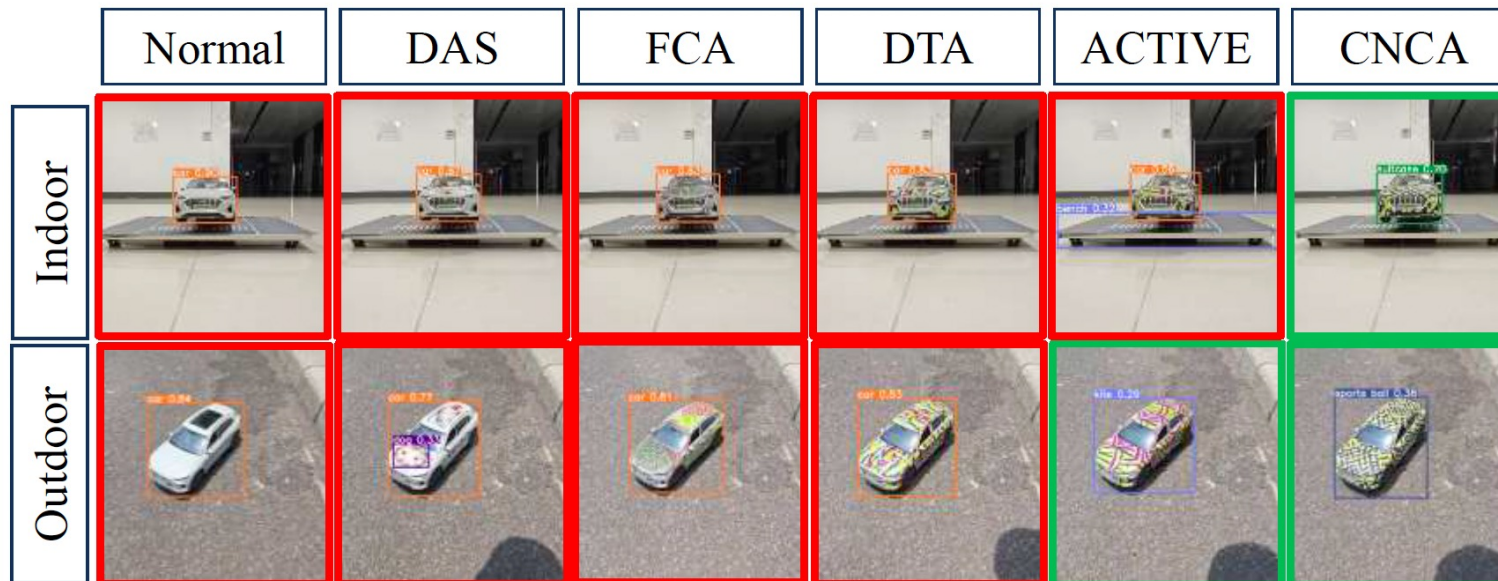


Figure 5: Examples of real-world evaluation for different methods. The evaluation includes both indoor and outdoor environments.

Conclusion

- CNCA: a novel framework leveraging diffusion models to generate customizable and natural adversarial camouflage
- Adversarial feature for gradient-based adversarial camouflage generation.
- Clipping strategy to balance attack performance and naturalness.

