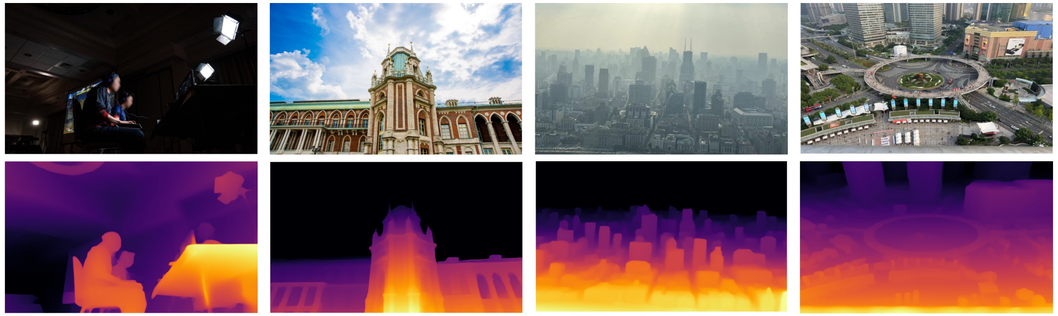
Beware of Road Markings: A New Adversarial Patch Attack to Monocular Depth Estimation

Hangcheng Liu, Zhenhu Wu, Hao Wang, Xingshuo Han, Shangwei Guo, Tao Xiang, and Tianwei Zhang

11/12/2024

Background

- Monocular Depth Estimation (MDE) enables the prediction of scene depths from a single RGB image, having been widely integrated into real-world applications
  - ➢ Autonomous driving
  - ➢ Augmented Reality
  - ➢ 3D modeling
  - ➢ Robotics

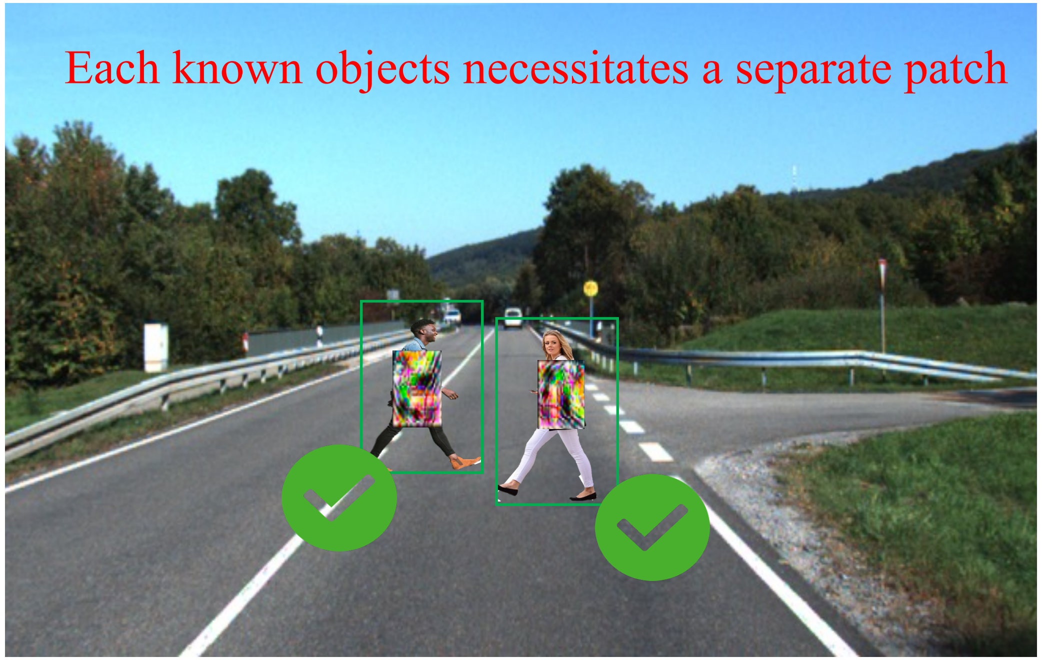


- Existing MDE models suffers safety threats from adversarial attacks
  - ➢ Adversarial patches can effectively change the predicted depth, even in the real world

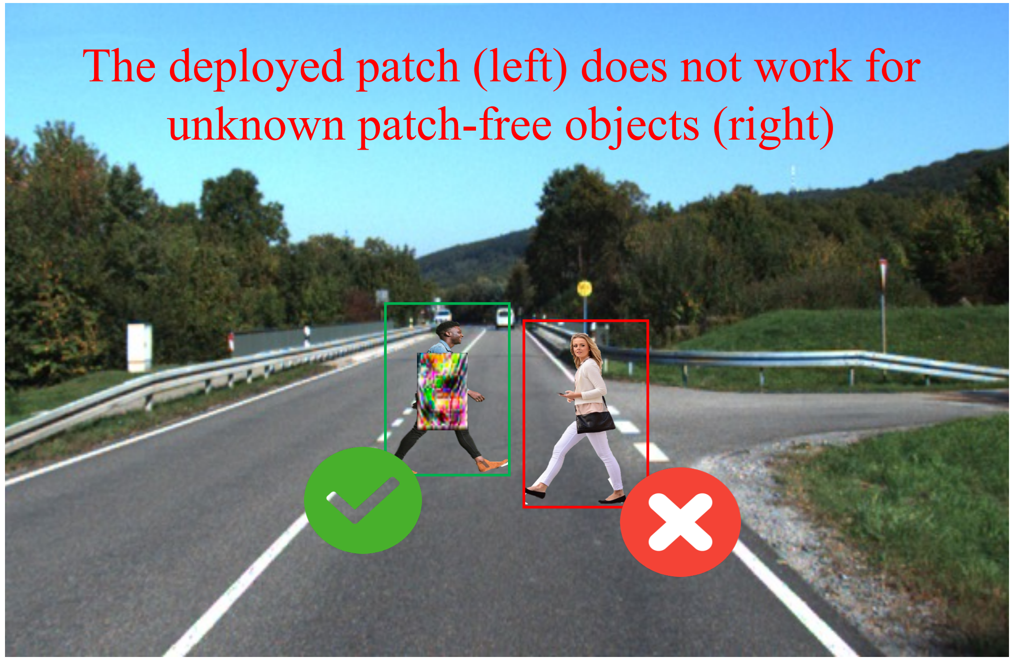
Limitations of Existing Attacks

- Current physical attacks focus on generating object-dependent patches
  - Object-dependency limits the application in complex traffic scenarios, especially in multi-object scenarios

✅ Affected ❌ Unaffected



Successful example

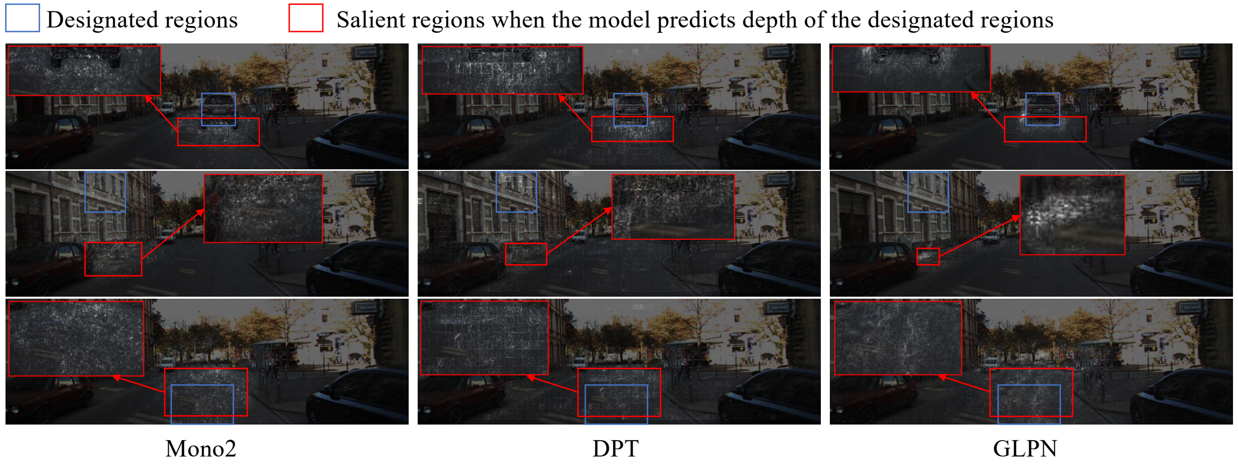


Failed example

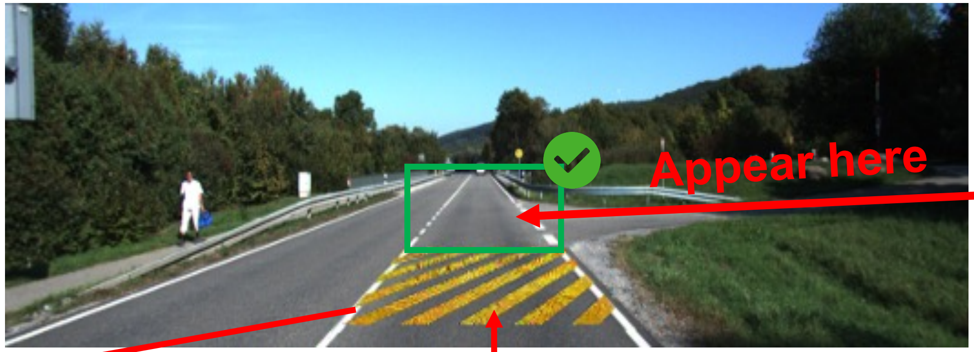
Key Observations & Goals

- **Observations:** mainstream MDE models trained for traffic scenarios exhibit a road-dependency property
  - ➢ Inspiring us to deploy patches on the road

- **Attack goals:** designing a new road patch that meets the following requirements:
  - ➢ **Stealth:** natural appearance
  - ➢ **Effectiveness:** altering the predicted distance of any object that appears in front of it



Real road markings in Singapore



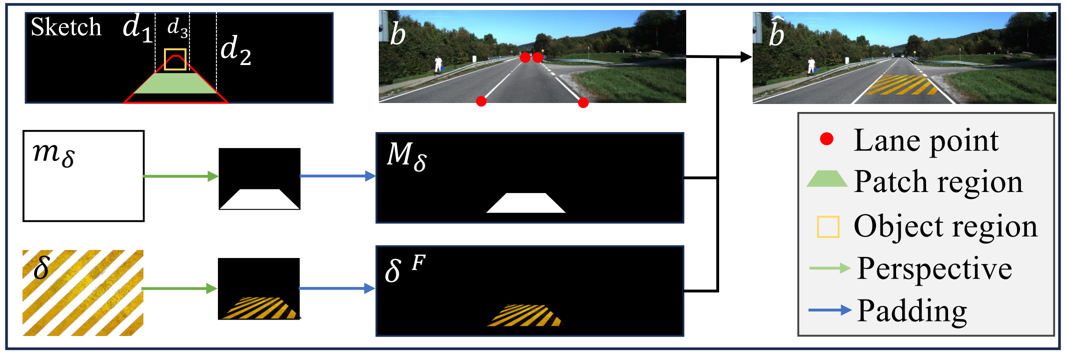
Natural road patch

Possible objects

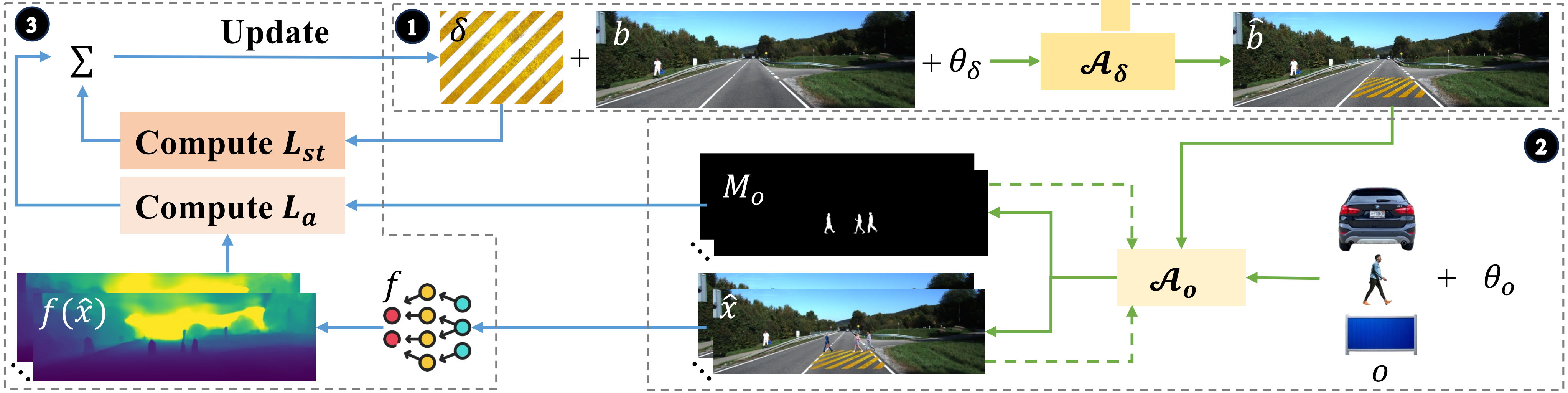
Methods

- **Step 1 and 2** in the overall pipeline are designed for image synthesis
  - ➢ Lane key points based synthesis
  - ➢ Multi-target insertion
- **Step 3** is used to optimize the road patch
  - ➢ Attack loss
  - ➢ Stealth loss
- ➢ **Robustness enhancement:** applying EoT to the patch

Patch insertion



Overall pipeline



Evaluation Results

- **Datasets**: **KITTI**
- **Victim MDE models:**

Backbone	Model
CNN	Depthhints (Dehin) [40]
	Monodepth2 (Mono2) [17]
	Manydepth (Mande) [41]
ViT	Midas [30]
	Adabins (Ada) [7]
	GLPN [22]
	Depth Anything (DeAny) [48]
	DPT-DINOv2 (DPT) [28, 29]

- **Evaluation metrics:**
  1. Mean Relative Shift Ratio (MRSR)

$$\xi_r = \frac{\mathrm{sum}(f_{M_o}(\hat{x}) - f_{M_o}(x))}{\mathrm{sum}(f_{M_o}(x))}$$

  2. Affect Region Ratio (ARR)

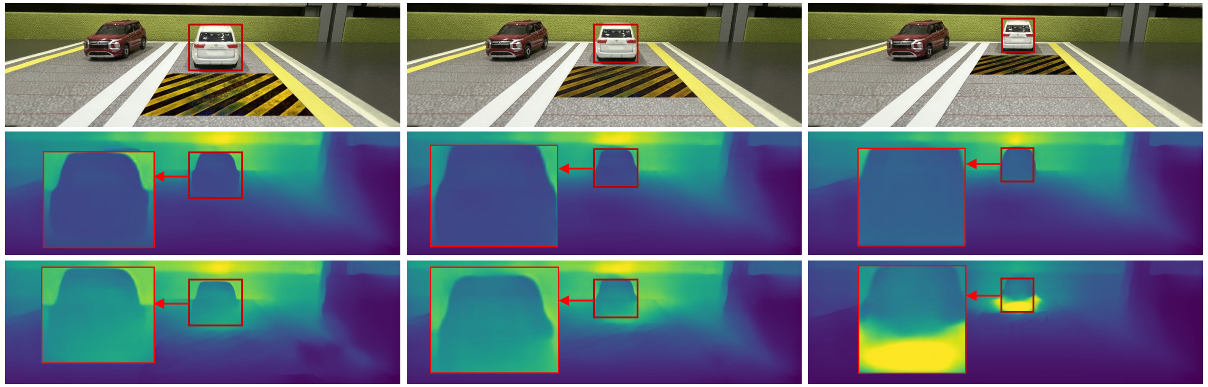
$$\xi_a = \frac{\mathrm{sum}\left(\mathbb{1}\left(f_{M_o}(\hat{x}) > (f_{M_o}(x) \times \eta)\right)\right)}{\mathrm{sum}(M_o)}$$

- **Main results**: the average MRSR over all models is 1.507, i.e., an object located at **12m** will be considered to be at **30 m**. Such an error is enough to delay braking and cause a serious collision.

Metric	Obstacle	CNN			ViT				
		Dehin	Mono2	Mande	Midas	Ada	GLPN	DeAny	DPT
ξ_r	PE	1.319	2.431	0.977	0.329	2.151	0.649	0.518	5.589
	CA	0.941	1.868	0.583	0.240	1.008	0.245	0.469	3.562
	RO	1.108	3.157	1.211	0.370	2.334	0.300	0.505	4.314
	Average	1.123	2.485	0.924	0.313	1.831	0.398	0.497	4.488
ξ_a	PE	0.954	0.960	0.954	0.817	0.999	0.918	0.946	0.999
	CA	0.948	0.969	0.873	0.729	0.998	0.793	0.984	0.998
	RO	0.929	0.999	0.997	0.953	1.000	0.805	0.989	1.000
	Average	0.944	0.976	0.942	0.833	0.999	0.839	0.973	0.999

- **Physical simulation.**

Obstacle regions



The end
Thanks