Invisible Reflections: Leveraging Infrared Laser Reflections to Target Traffic Sign Perception

Takami Sato[*1], **Sri Hrushikesh Varma Bhupathiraju**[*2], Michael Clifford[3], Takeshi Sugawara[4], Qi Alfred Chen[1], and Sara Rampazzi[2]

1 UCI 2 UF 3 TOYOTA InfoTech Envisioning Mobility 4 UEC

* denotes co-first authors.

Infrared (IR) laser is not visible to humans

To human eye
(normal camera with IR filter)

Infrared (IR) laser is not visible to humans

To human eye
(normal camera with IR filter)

To camera
without IR filter

Infrared (IR) laser is not visible to humans

To human eye
(normal camera with IR filter)

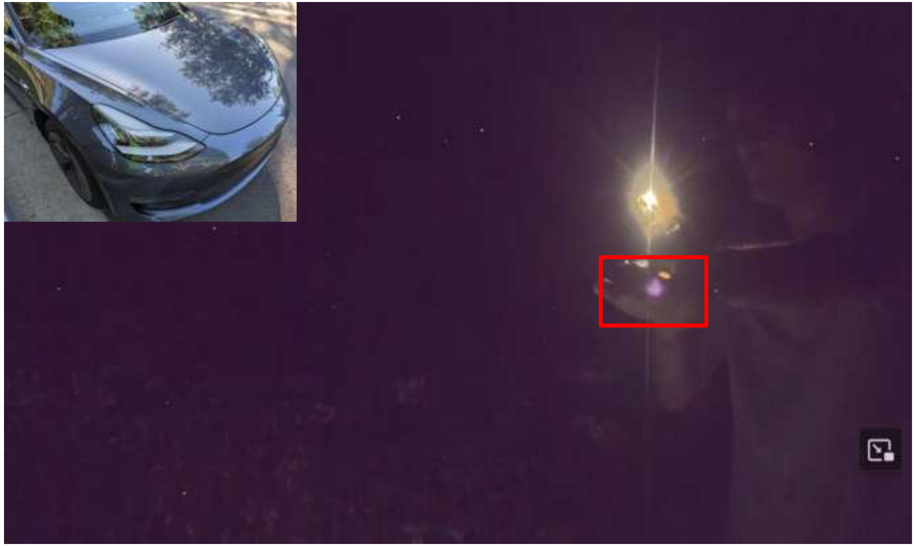
To camera
without IR filter

Can Infrared Laser Reflection (ILR) be a new attack vector?

Autonomous Vehicle Cameras without IR filters

ICSL (I Can See the Light) Attack [Wang et al., CCS'21]

- IR light is detected as red light

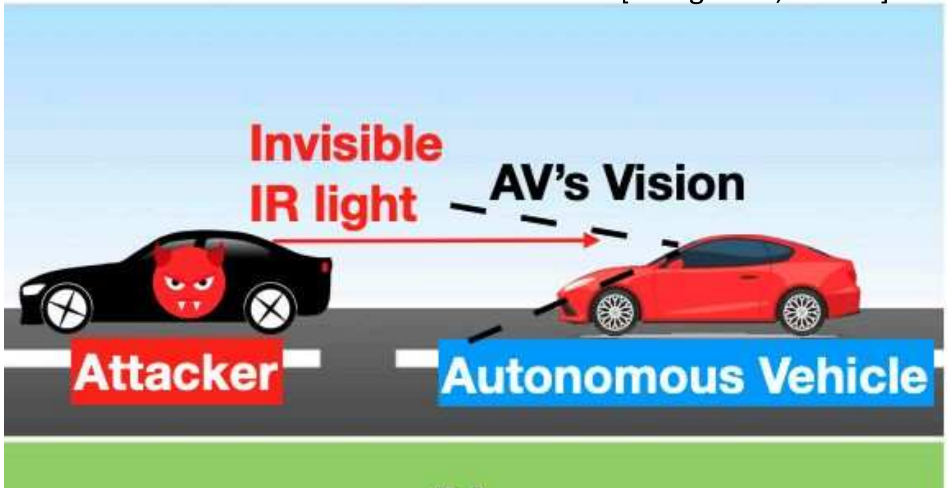


We also confirmed that a commodity car with AV does not have IR filter.

Limitations of Existing Attacks: Visibility for Human

ICSL (I Can See the Light) Attack

[Wang et al., CCS'21]



Patch Attacks

[Eykholt et al., 2018]

[Jia et al., 2022]

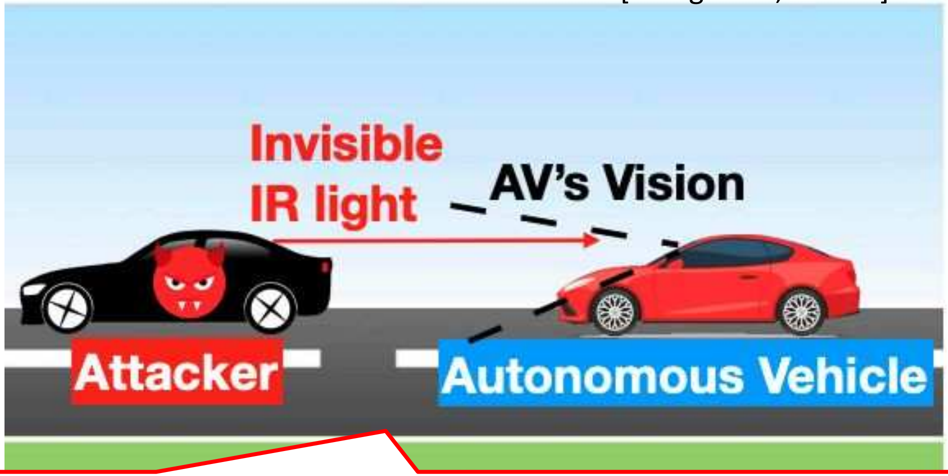
[Chen et al., 2019]

[Zhao et al., 2019]

Limitations of Existing Attacks: Visibility for Human

ICSL (I Can See the Light) Attack

[Wang et al., CCS’21]



Limitation ①

- **Need accurate aiming at driving target**
- **Not designed for attacking traffic sign**

Patch Attacks

[Eykholt et al., 2018]

[Jia et al., 2022]

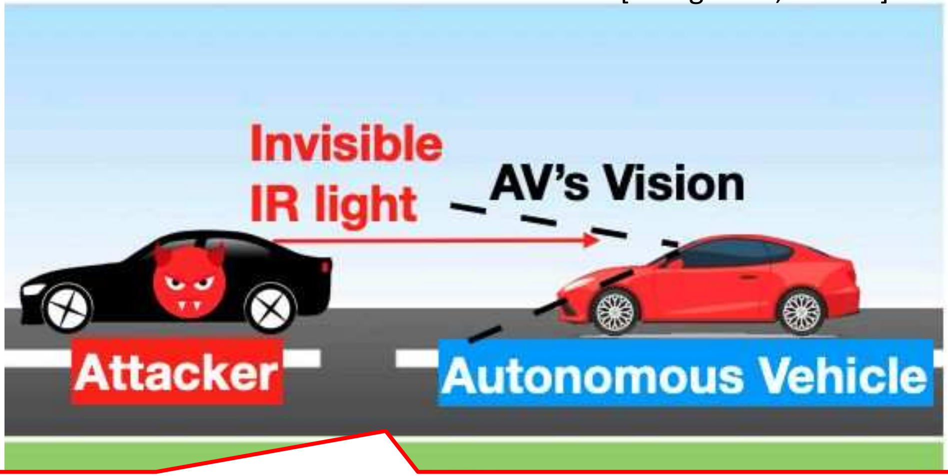
[Chen et al., 2019]

[Zhao et al., 2019]

Limitations of Existing Attacks: Visibility for Human

ICSL (I Can See the Light) Attack

[Wang et al., CCS’21]



Limitation ①
- **Need accurate aiming at driving target**
- **Not designed for attacking traffic sign**

Patch Attacks

[Eykholt et al., 2018]

[Jia et al., 2022]

[Chen et al., 2019]

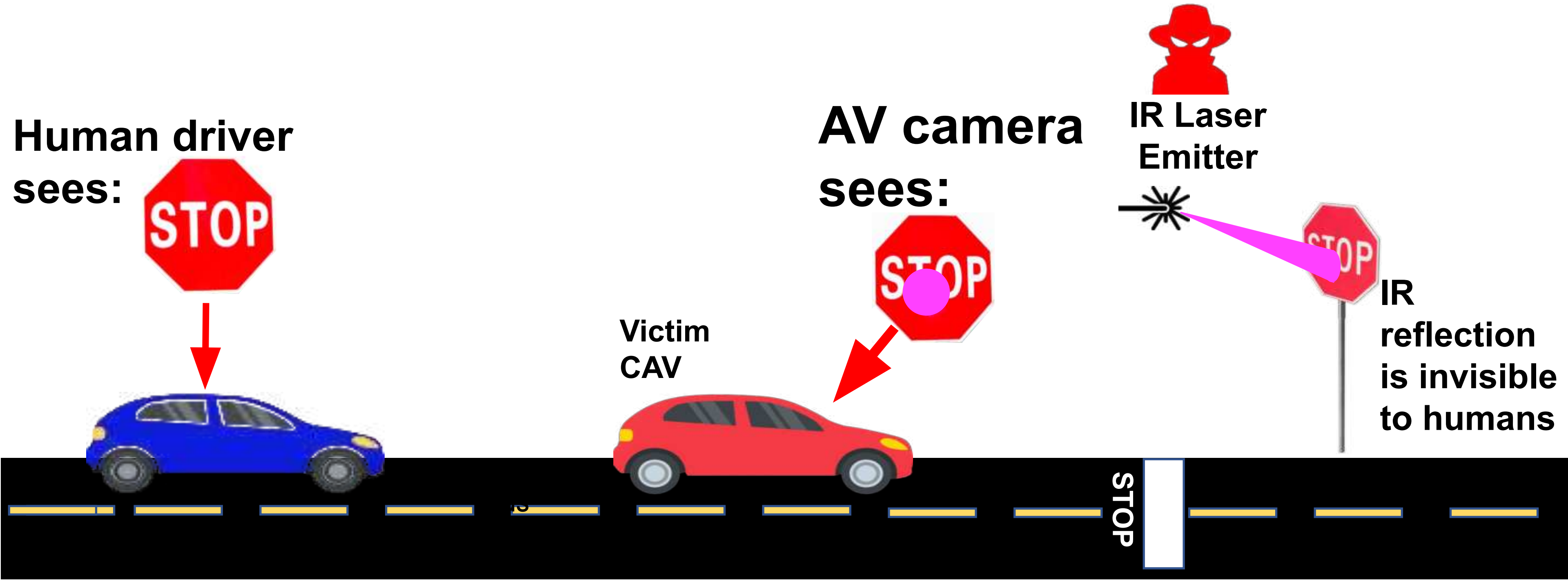
[Zhao et al., 2019]

Limitation ②

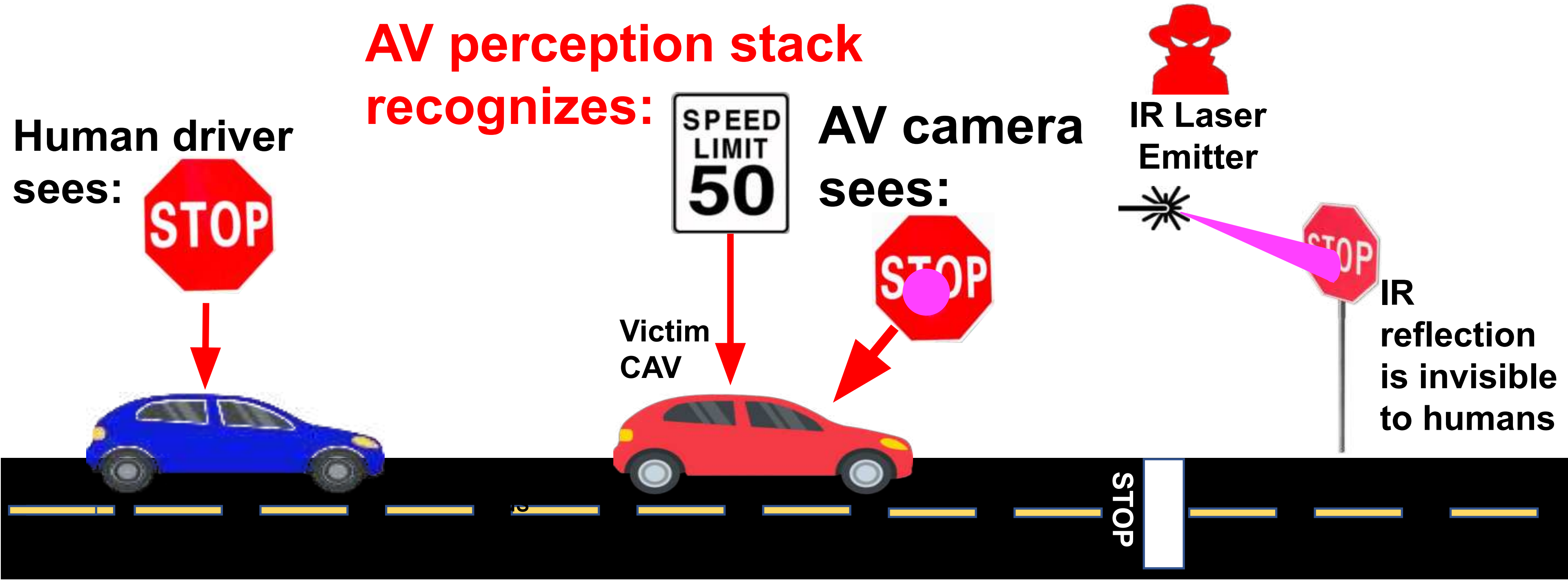
Attack attempt visible to human

Our Attack Vector: Infrared Laser Reflection (ILR)

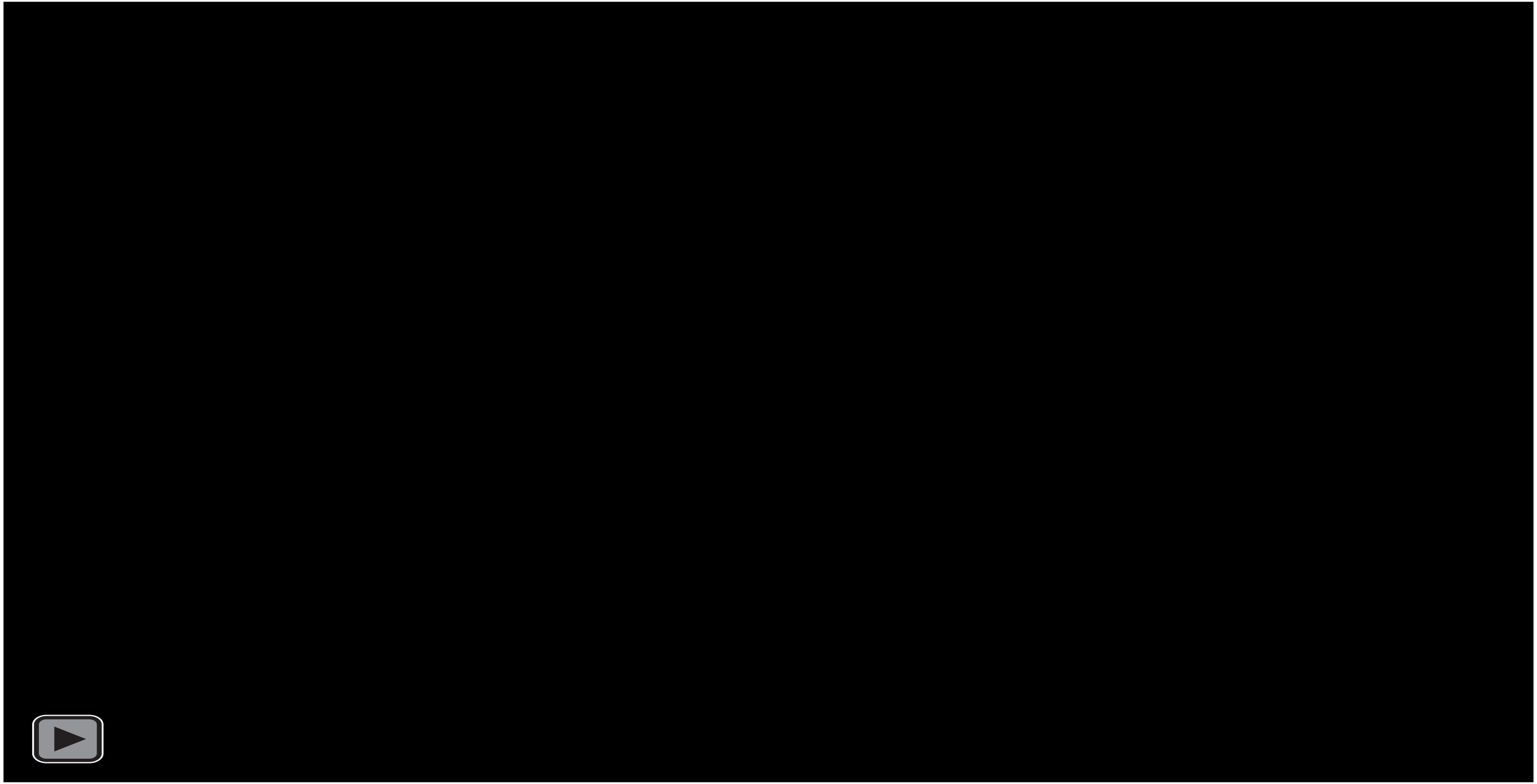
Our Attack Vector: Infrared Laser Reflection (ILR)



Our Attack Vector: Infrared Laser Reflection (ILR)

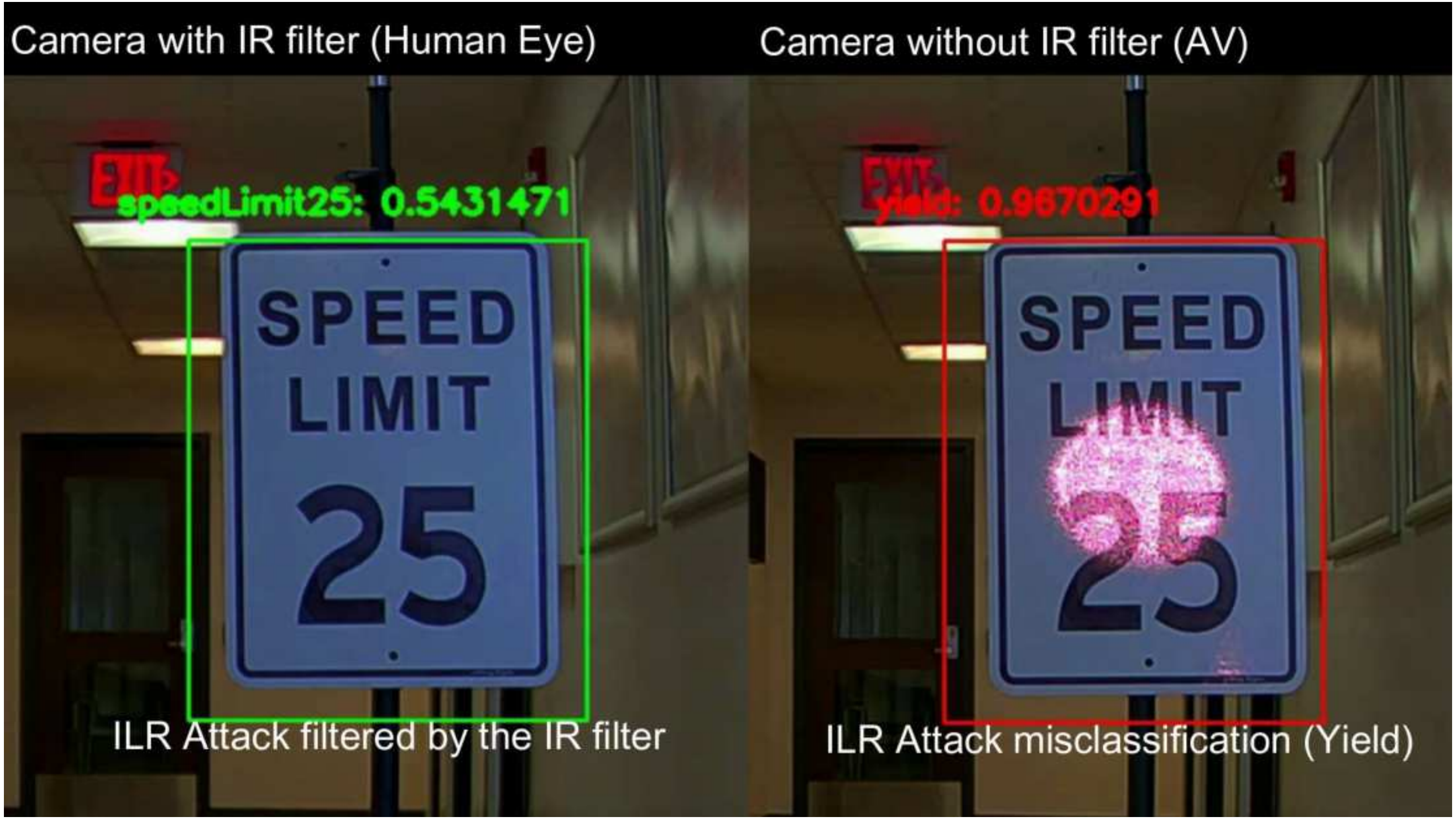


Attack Demo: Indoor Experiment

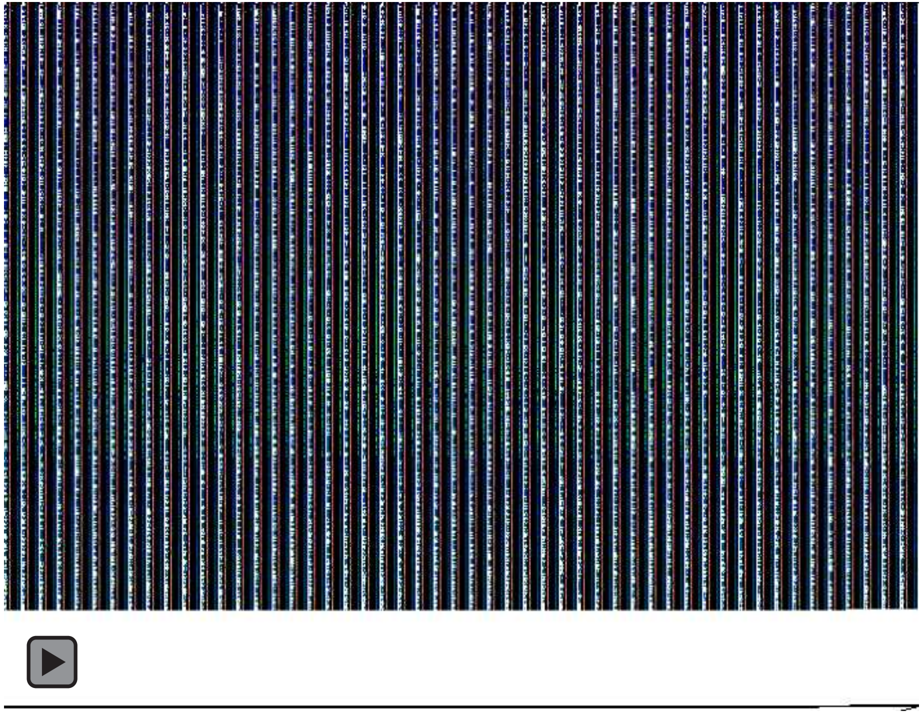


Attack Demo: Indoor Experiment

Attack Demo: Indoor Experiment



Attack Demo: Outdoor Experiment



Attack Demo: Outdoor Experiment

Research Challenges

- **<u>Physical attack capability understanding & modeling</u>**
  - Complicated physical process behind the speckle pattern
    - Pattern is generated from multiple, randomly phased, coherent waves
  - Non-trivial to effectively interpolate unseen ILR attack trace
    - Naive averaging cancels out the speckle pattern
- **<u>Automatic generation of effective attacks on traffic sign recognition model side</u>**
  - Attack effectiveness highly depends on the position, size, and intensity of the speckled pattern
  - Need to be robust to different distances and view angles

Naive trace modeling does not work

Ground Truth

Prediction:
Yield

Attack Modeling

Alpha Blending

Prediction:
Speed Limit (70 km/h)

Ray Tracing

Prediction:
Stop Sign

Naive trace modeling does not work

Attack Modeling

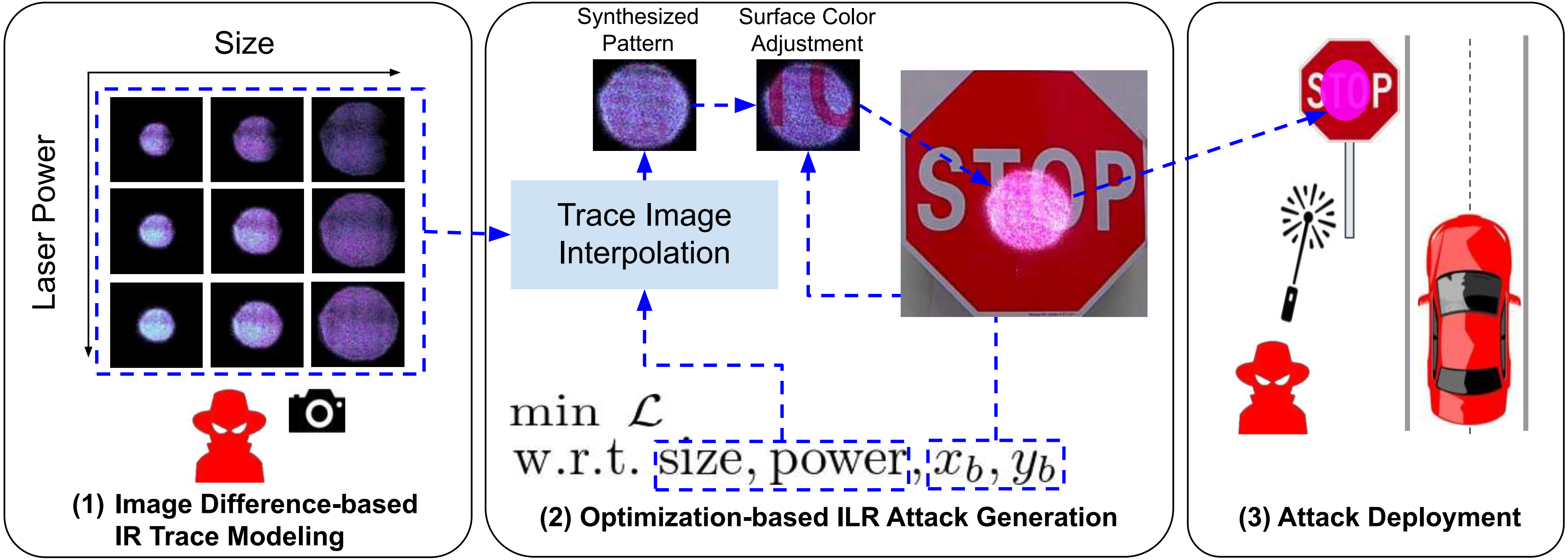
Ground Truth	**Alpha Blending**	**Ray Tracing**

Prediction:
Yield

Prediction:
Speed Limit (70 km/h)

Prediction:
Stop Sign

Overview of Attack Generation Pipeline



Our attack generation consists of 3 steps

Overview of Attack Generation Pipeline

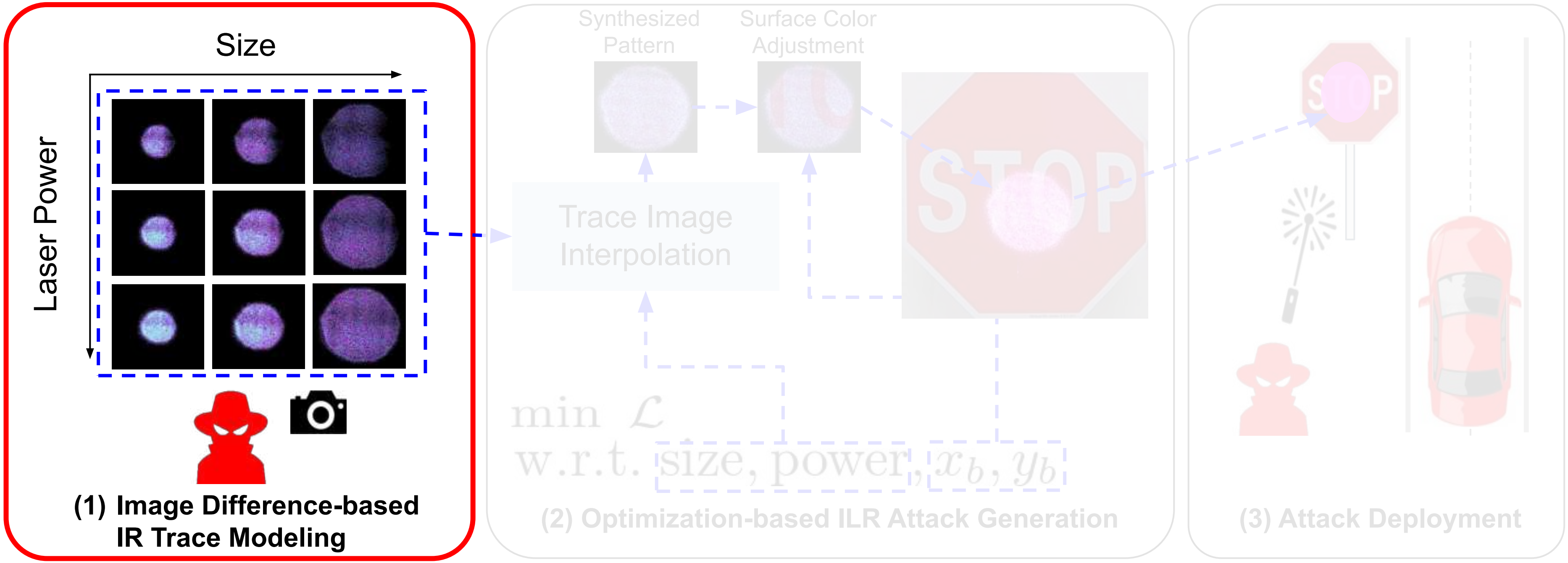
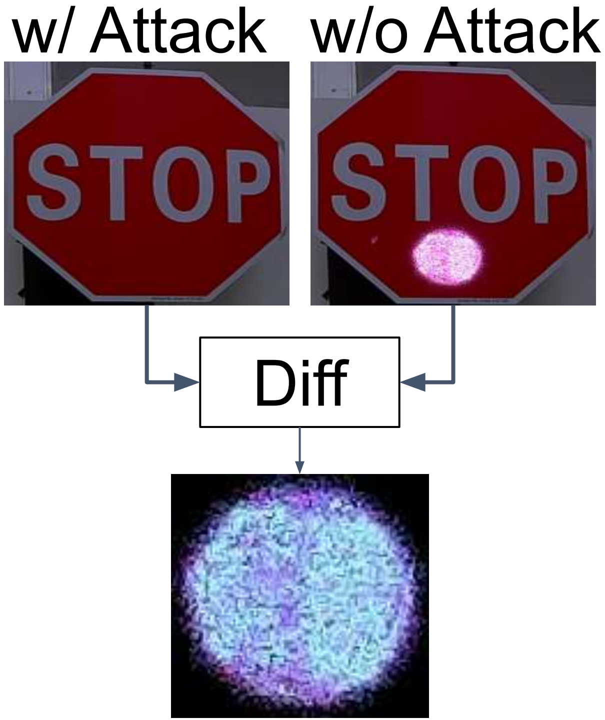
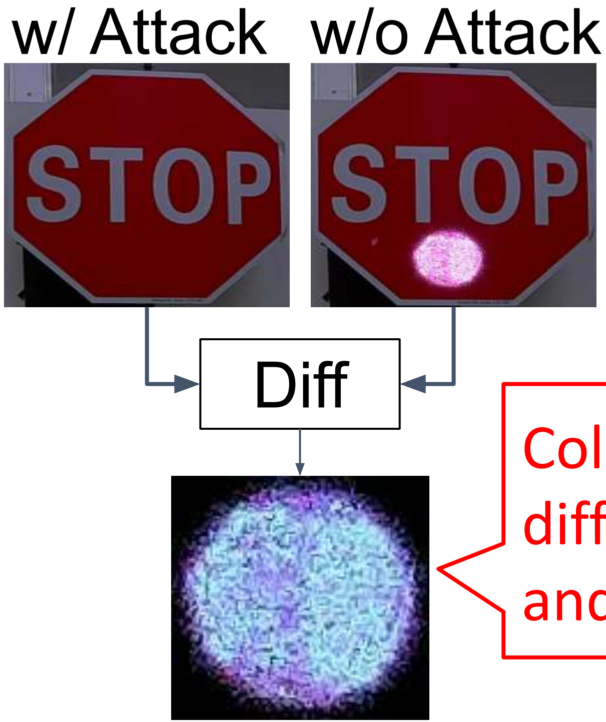


Image Difference-based IR Trace Modeling



Simple but no need to simulate complex speckle patterns

Image Difference-based IR Trace Modeling



Collect trace diffs with different **laser powers** and **diameters.**

Simple but no need to simulate complex speckle patterns

Trace Image Interpolation

How to simulate *non-collected* traces?

- **Impossible to collect them physically**
- **Naive averaging dismisses the pattern**

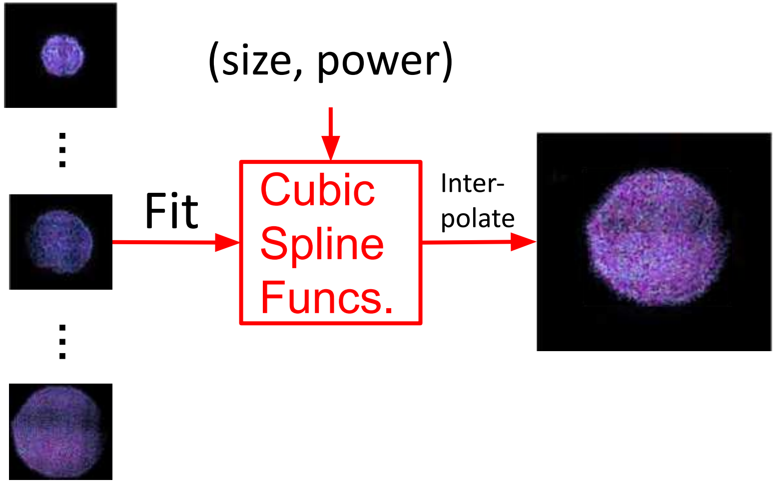
Trace Image Interpolation

How to simulate *non-collected* traces?

- **Impossible to collect them physically**
- **Naive averaging dismisses the pattern**

Pixel-wise Spline Interpolation

- Apply **cubic spline** for each pixel



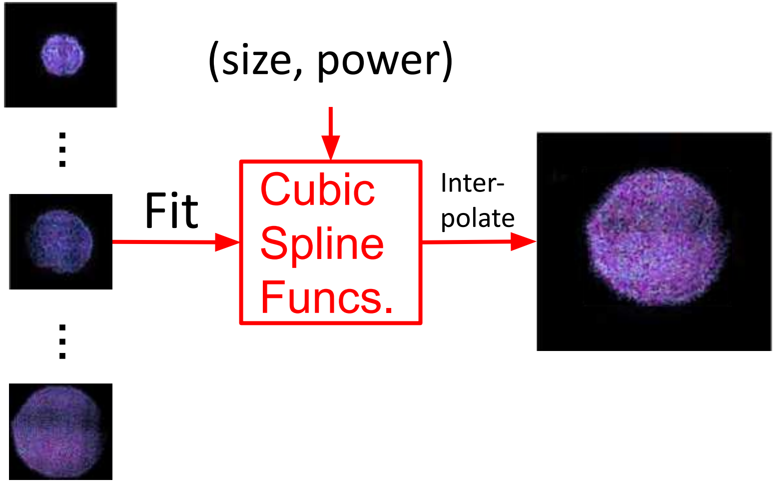
Trace Image Interpolation

How to simulate *non-collected* traces?

- **Impossible to collect them physically**
- **Naive averaging dismisses the pattern**

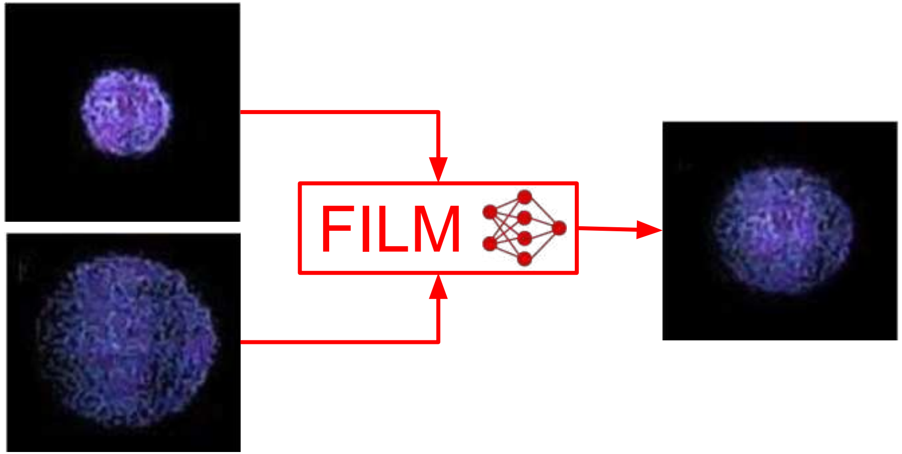
Pixel-wise Spline Interpolation

- Apply **cubic spline** for each pixel

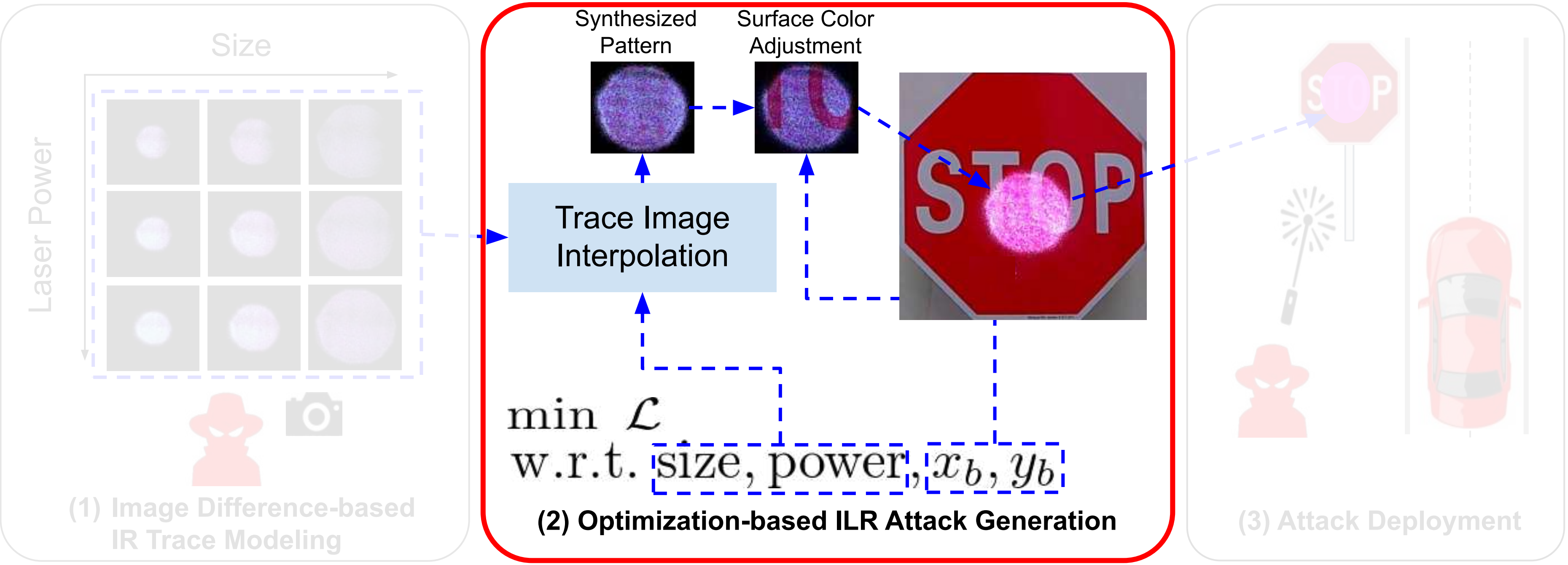


DNN-based Interpolation

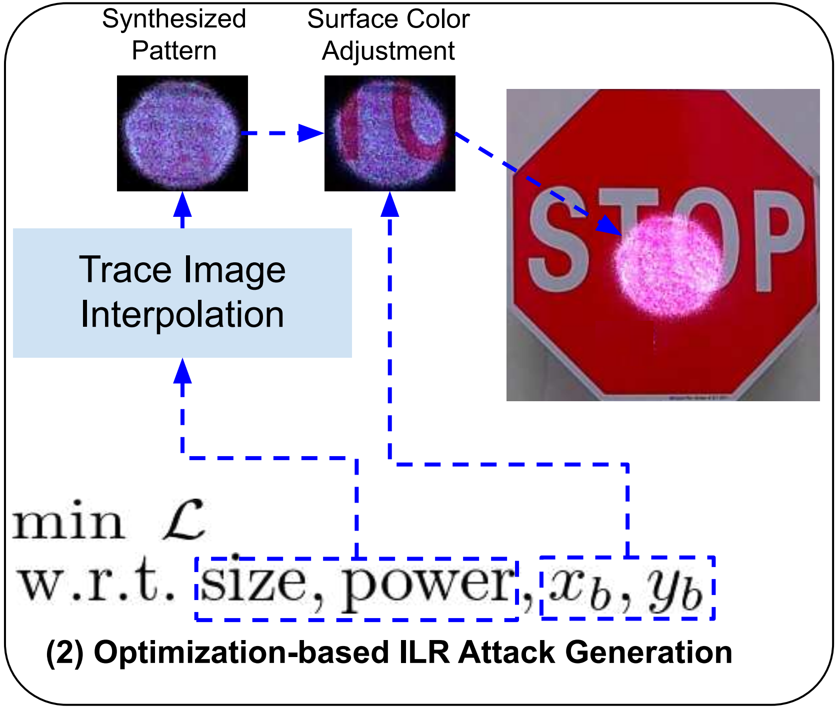
- Apply **FILM** [Reda et al., 2022] model



Overview of Attack Generation Pipeline



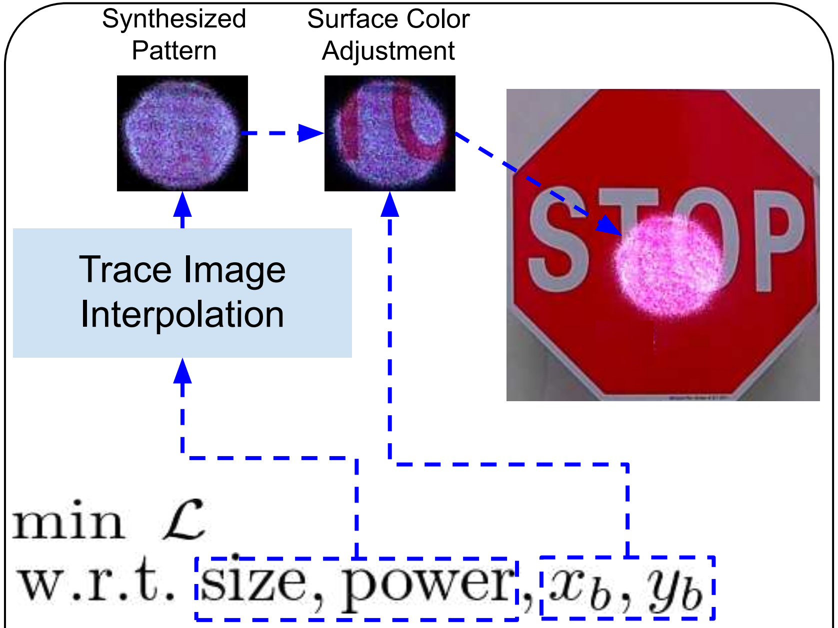
Optimization-based ILR Attack Generation



Black-box attack optimization

- Optimize attack trace w.r.t
  - size
  - power
  - position
- Use a bayesian optimization, Tree-Structured Parzen Estimator

Optimization-based ILR Attack Generation



(2) Optimization-based ILR Attack Generation

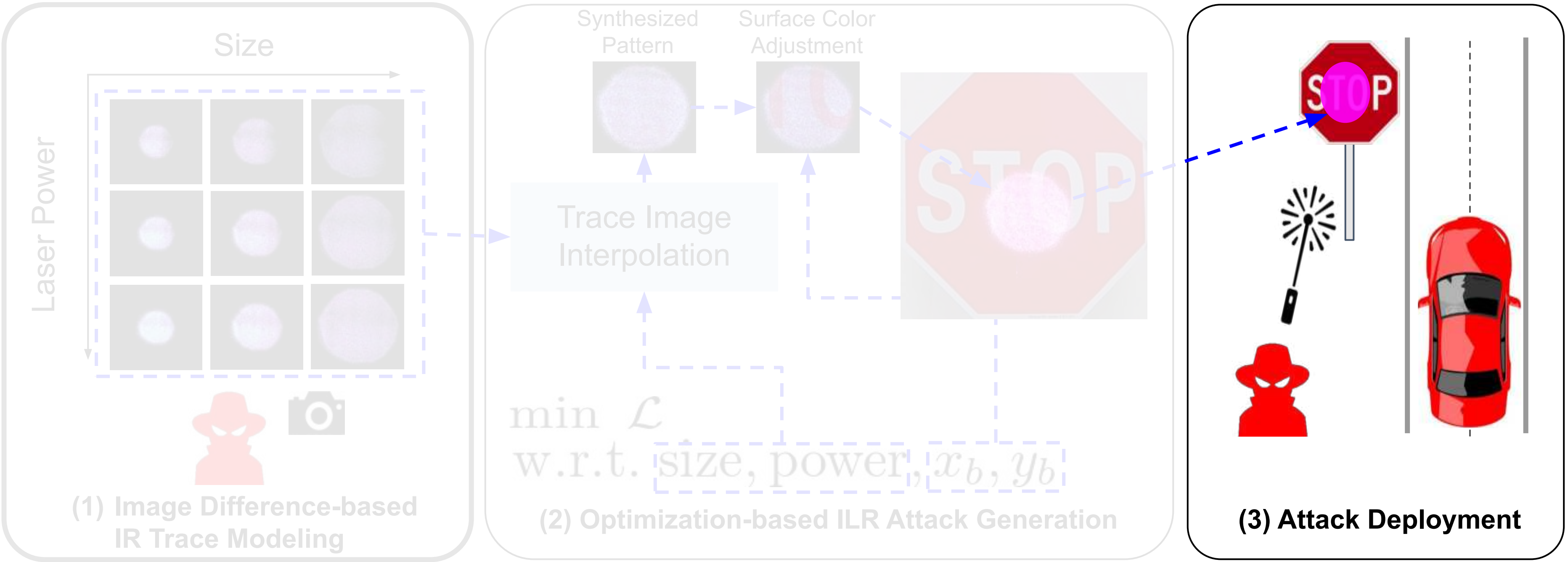
Black-box attack optimization

- Optimize attack trace w.r.t
  - size
  - power
  - position
- Use a bayesian optimization, Tree-Structured Parzen Estimator

Robustness improvement with Expectation over Transformation

- Resizing
- Brightness
- Gaussian Noise
- Rotation
- Shearing etc.

Overview of Attack Generation Pipeline



Attack Evaluation

** All attacks in this paper are physically deployed and evaluated*

Evaluation Criteria

- Effectiveness
- Generality
- Robustness
- Transferability

Evaluation Scenarios

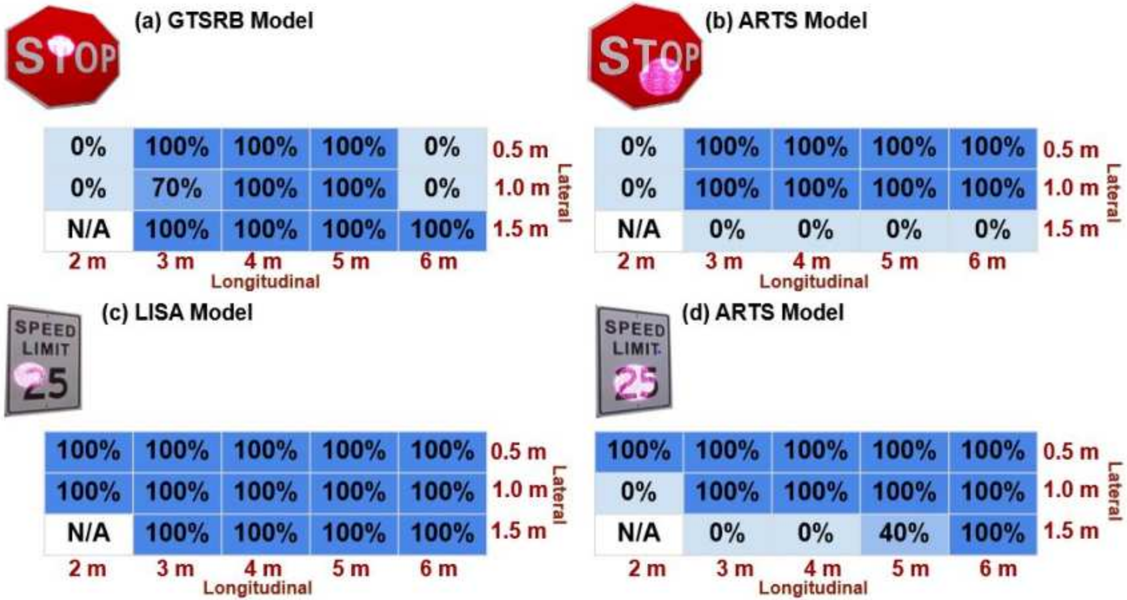
- Indoor
- Outdoor
    - Different lighting conditions
- Dynamic

Attack Factor Effectiveness

Target Traffic Sign Classification Models:
GTSRB, ARTS, LISA

Evaluation Factors:

- Lighting Conditions
- Victim Cameras
- Laser Modules
- Laser Orientations
- Camera Position



(a) GTSRB Model

Lateral \ Longitudinal	2 m	3 m	4 m	5 m	6 m
0.5 m	0%	100%	100%	100%	0%
1.0 m	0%	70%	100%	100%	0%
1.5 m	N/A	100%	100%	100%	100%

(b) ARTS Model

Lateral \ Longitudinal	2 m	3 m	4 m	5 m	6 m
0.5 m	0%	100%	100%	100%	100%
1.0 m	0%	100%	100%	100%	100%
1.5 m	N/A	0%	0%	0%	0%

(c) LISA Model

Lateral \ Longitudinal	2 m	3 m	4 m	5 m	6 m
0.5 m	100%	100%	100%	100%	100%
1.0 m	100%	100%	100%	100%	100%
1.5 m	N/A	100%	100%	100%	100%

(d) ARTS Model

Lateral \ Longitudinal	2 m	3 m	4 m	5 m	6 m
0.5 m	100%	100%	100%	100%	100%
1.0 m	0%	100%	100%	100%	100%
1.5 m	N/A	0%	0%	40%	100%

- Attack success rates reach **100%**, when camera is 3-5 m away from traffic sign
- Attack is more successful on speed limits - due to contrast with laser speckles

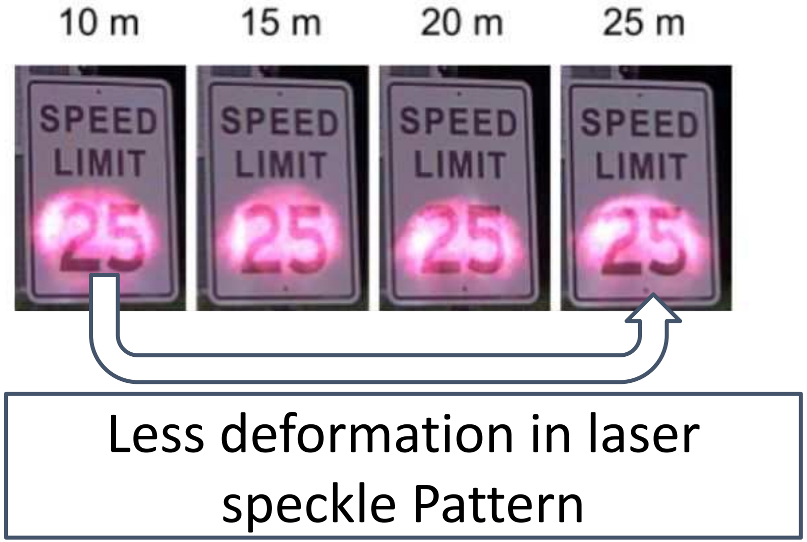
Attack on Object Detectors

- Attack success rates are **100%** at 6 m away from the target
  - YOLOv3 (single-stage object detector) shows higher robustness
- Attack is more robust on Speed Limit.

Target Sign	Detection Model	4 m	5 m	6 m	7 m
Stop Sign	Faster R-CNN (ARTS)	**100%**	**100%**	**100%**	**100%**
	YOLOv3 (COCO)	0%	0%	**100%**	0%
	YOLOv5 (COCO)	10%	90%	**100%**	**100%**
Speed Limit	Faster R-CNN (ARTS)	**100%**	**100%**	**100%**	**100%**
	Faster R-CNN (Mapillary)	**100%**	**100%**	**100%**	**100%**
	YOLOv5 (ARTS)	**100%**	**100%**	**100%**	**100%**

Maximum Attacker Distance

- Attack deployed from 25 meters with low power (26 mW).
- Long range attack due to laser properties
- Longer attack distances deform speckle, require sophisticated optics



Outdoor Attack Evaluation

	Stop Sign				Speed Limit			
	ARTS		GTSRB		ARTS		LISA	
Speed	ASR	SCR	ASR	SCR	ASR	SCR	ASR	SCR
Night Scenario								
5 km/h	**100%**	**100%**	99%	90%	**100%**	0%	99%	31%
8 km/h	**100%**	**100%**	92%	91%	**100%**	0%	**100%**	0%
13 km/h	**100%**	**100%**	85%	85%	**100%**	0%	99%	16%
Day Scenario								
5 km/h	98%	82%	85%	57%	**100%**	18%	**100%**	98%
8 km/h	**100%**	88%	88%	46%	**100%**	50%	**100%**	87%
13 km/h	91%	75%	80%	40%	**100%**	58%	**100%**	98%

Outdoor Attack Evaluation

	Stop Sign				Speed Limit			
	ARTS		GTSRB		ARTS		LISA	
Speed	ASR	SCR	ASR	SCR	ASR	SCR	ASR	SCR
Night Scenario								
5 km/h	**100%**	**100%**	99%	90%	**100%**	0%	99%	31%
8 km/h	**100%**	**100%**	92%	91%	**100%**	0%	**100%**	0%
13 km/h	**100%**	**100%**	85%	85%	**100%**	0%	99%	16%
Day Scenario								
5 km/h	98%	82%	85%	57%	**100%**	18%	**100%**	98%
8 km/h	**100%**	88%	88%	46%	**100%**	50%	**100%**	87%
13 km/h	91%	75%	80%	40%	**100%**	58%	**100%**	98%

- High attack success rates for 2 models trained on popular datasets
  - 🌙 Night time (120 lux) : **≧85%** attack success rate
  - 🌞 Day time (982 lux) : **≧80%** attack success rate

Limitation of State-of-Art Certifiable Defense

- **PatchCleanser** [Xiang et al., 2022] **does not handle ILR attack well**
  - Assume prediction holds without adversarial trace
  - Their intuition doesn’t hold, i.e., small part making can change label
- PatchCleanser’s key idea, 2-round masking, can cause false agreements
- Mis-certifies ≥33.5% of cases of ILR attack traces

Proposed Defenses

Color-Frequency Detection: Physics-based characteristics of laser light reflections

Speckle Color Range based on ambient illumination

- #CF9FFF and #DA70D6 (Low illumination)
- #FFB266 to #CC6600 (High illumination)

Traffic Sign with ILT attack trace | Speckle Color Range | High Spatial Frequency

Evaluated on 300 images during daytime and nighttime scenarios

- 98% True Positive Rate and 2.7% False Positive Rate during daytime conditions
- 92% True Positive Rate and 6.7% False Positive Rate during nighttime conditions

Conclusion

Discovered ILR, **a long-distance and human-invisible attack vector,** that can cause misclassification by **traffic sign recognition systems**.

- **Design a novel methodology to optimize attack**
  - Image difference-based IR trace modeling
  - Trace image interpolation
  - Robust attack generation with black-box optimization
- **Measure the characteristics of ILR with a wide variety of parameters**
- **Perform evaluation in both indoor and outdoor day/night setups**
  - **100%** attack success rate indoor setup
  - **≥80.5%** attack success rate in outdoor driving setup up to 13 Km/h at day and night
- **Demonstrate the limitations in the current state-of-the-art defense**
- **Design a new defense leveraging characteristics of ILR**

Thank you!

Invisible Reflections: Leveraging Infrared Laser Reflections to Target Traffic Sign Perception

Takami Sato, **Sri Hrushikesh Varma Bhupathiraju**, Michael Clifford, Takeshi Sugawara, Qi Alfred Chen, and Sara Rampazzi

takamis@uci.edu and bhupathirajus@ufl.edu

Scan to visit our project website

UCI UF TOYOTA InfoTech Envisioning Mobility UEC