Beyond Pixel Norm-Balls: Parametric Adversaries using an Analytically Differentiable Renderer

Hsueh-Ti Derek Liu[1], Michael Tao[1], Chun-Liang Li[2], Derek Nowrouzezahrai[3], Alec Jacobson[1]

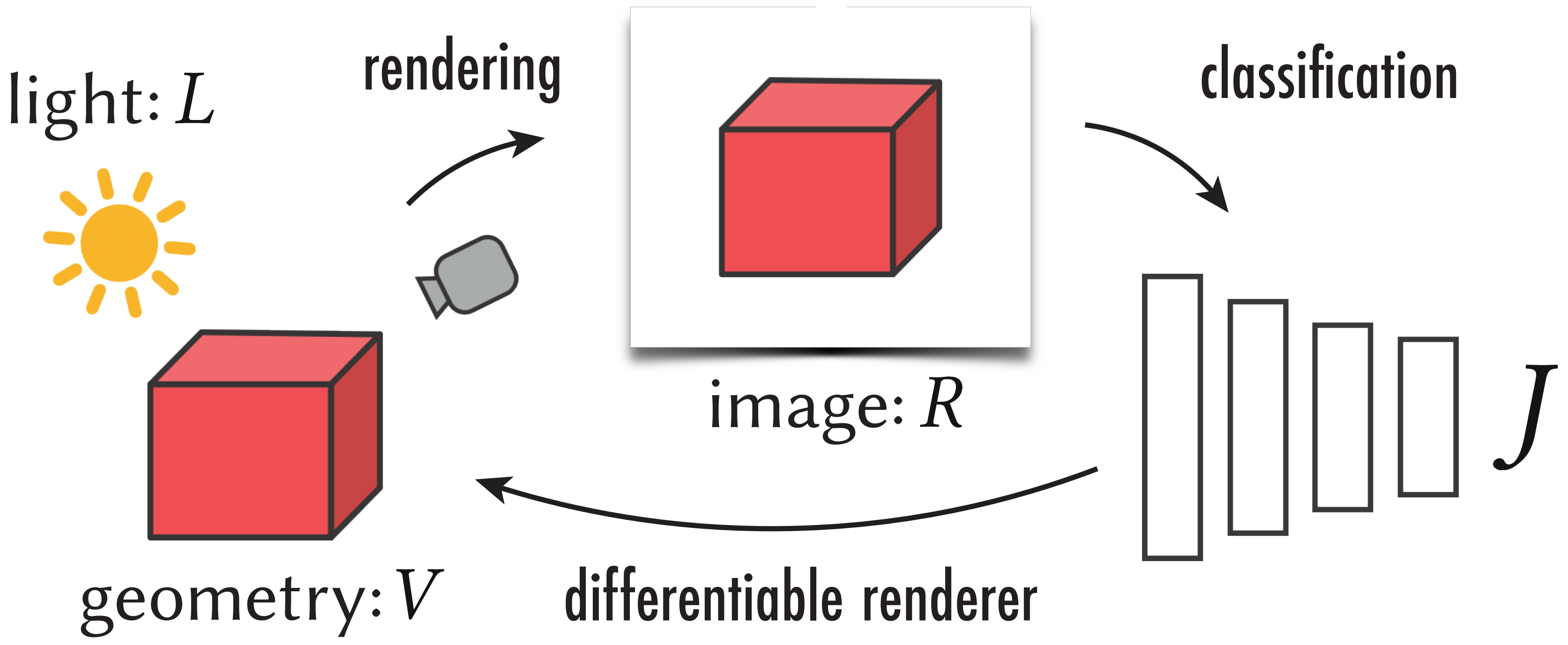
hsuehtil@cs.toronto.edu
http://www.dgp.toronto.edu/~hsuehtil/

Abstract

We compute adversarial examples by perturbing physical parameters instead of pixel colors. We present (1) *adversarial geometry* by 3D shape perturbations, and (2) *adversarial lighting* by scene lighting perturbations.

Method

We use a gradient-based optimization, where the gradients of the cost function J w.r.t 3D scene parameters V, L are cmputed via a physically-based *analytically* differentiable renderer.



Geometry derivatives

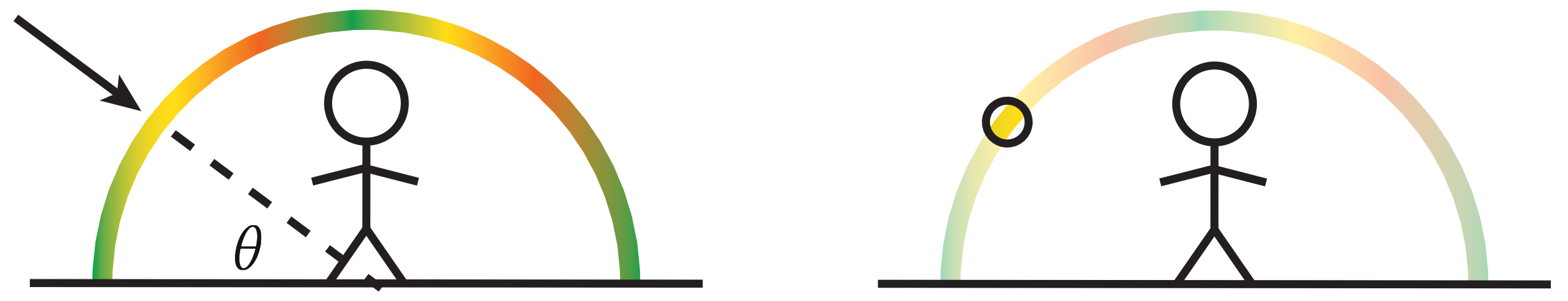
$$\frac{\partial J}{\partial V} = \frac{\partial J}{\partial R}\frac{\partial R}{\partial N}\frac{\partial N}{\partial V}$$

Lighting derivatives

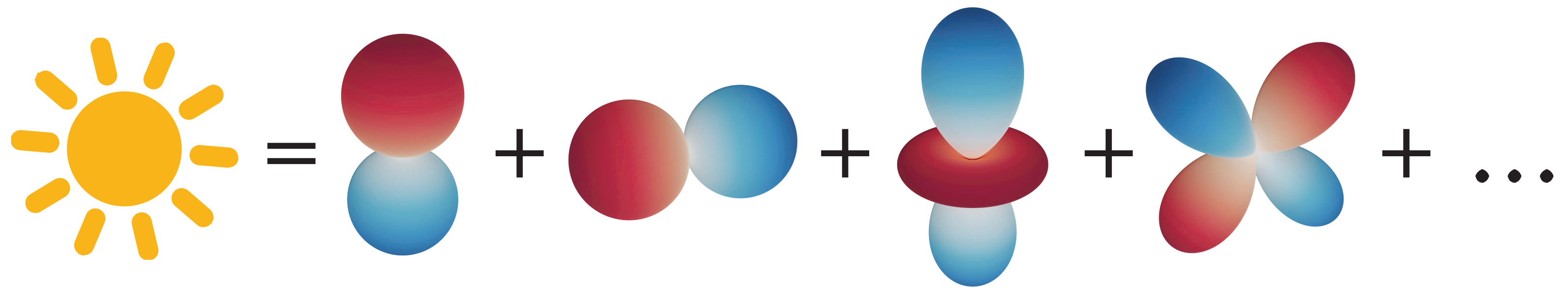
$$\frac{\partial J}{\partial L} = \frac{\partial J}{\partial R}\frac{\partial R}{\partial L}$$

Lighting parameterization

Lighting as a spherical function

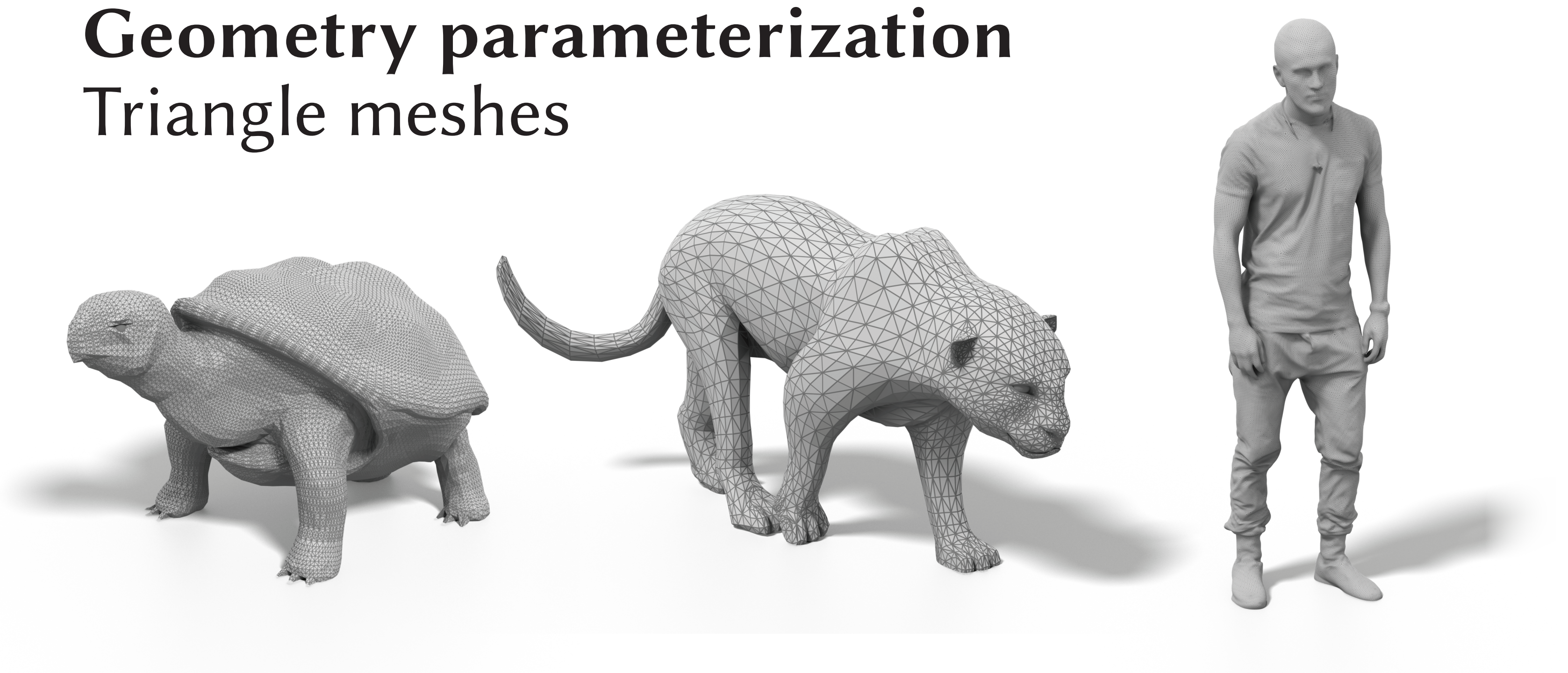


We parameterize lighting with spherical harmonics

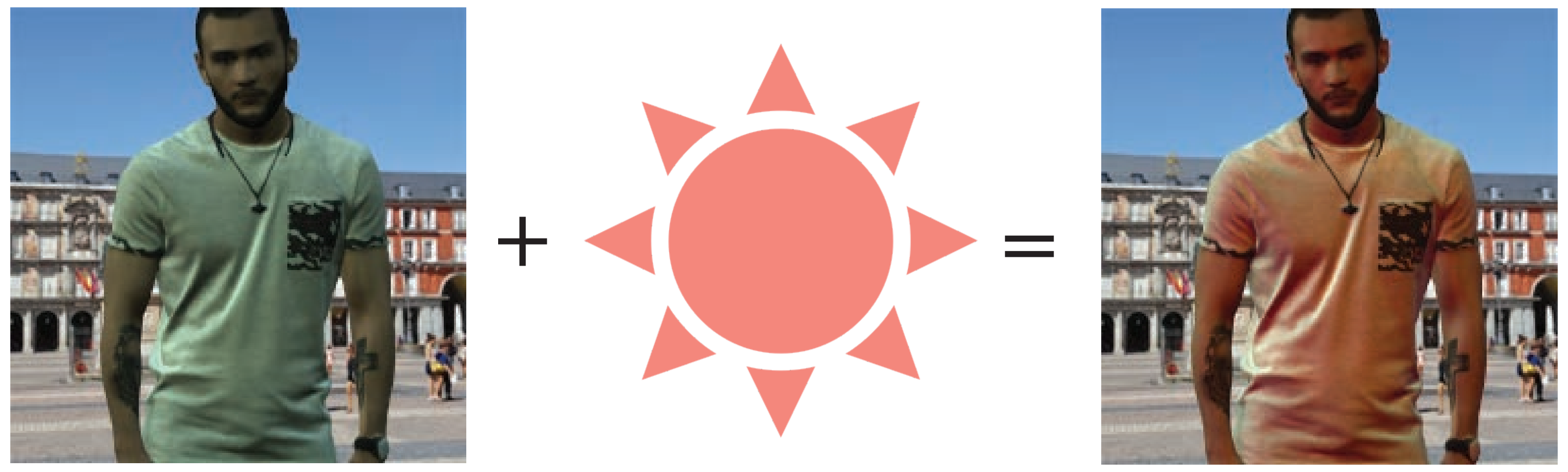


Geometry parameterization

Triangle meshes

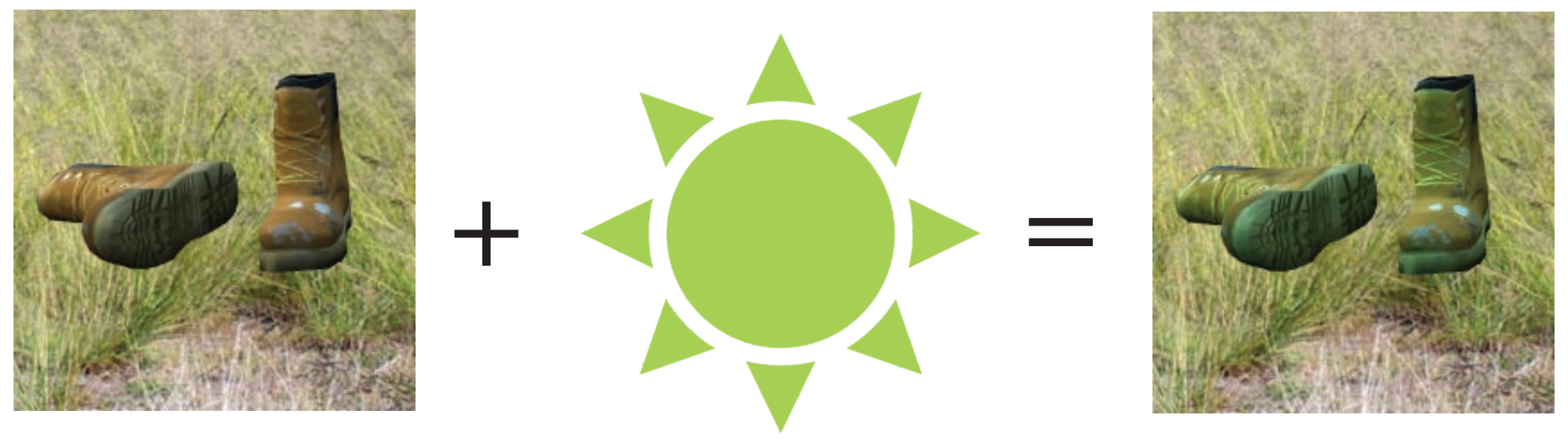


Adversarial lighting



t-shirt 86% miniskirt 28%

boots 98% boots 98% bottle 15% cannon 20%

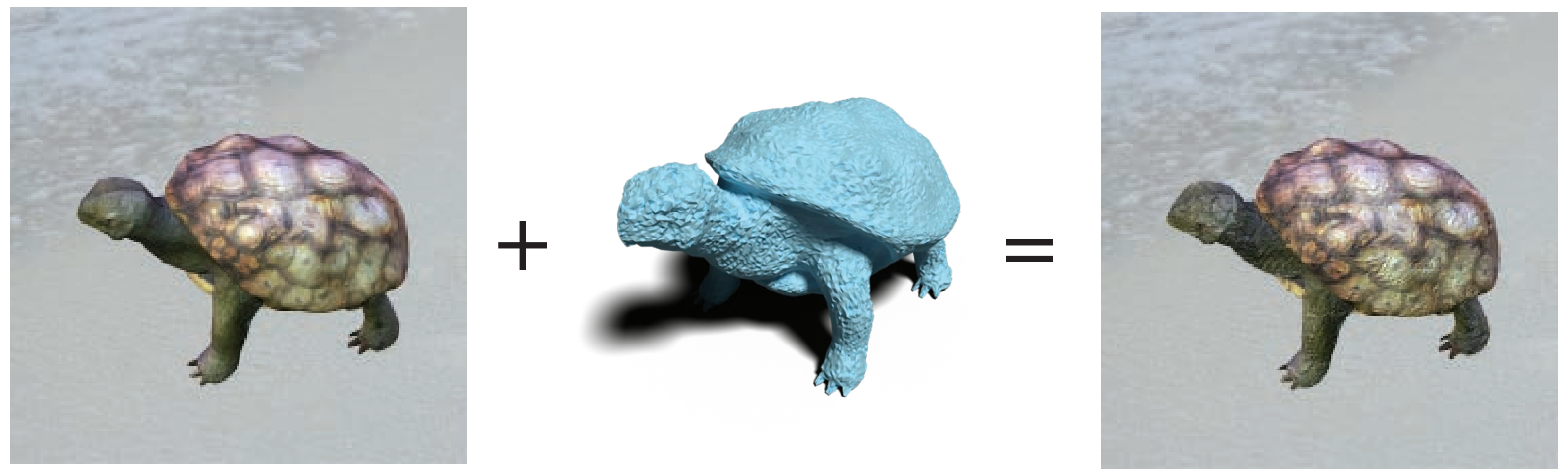


boots 100% sleep bag 98%

Rendered adversarial training

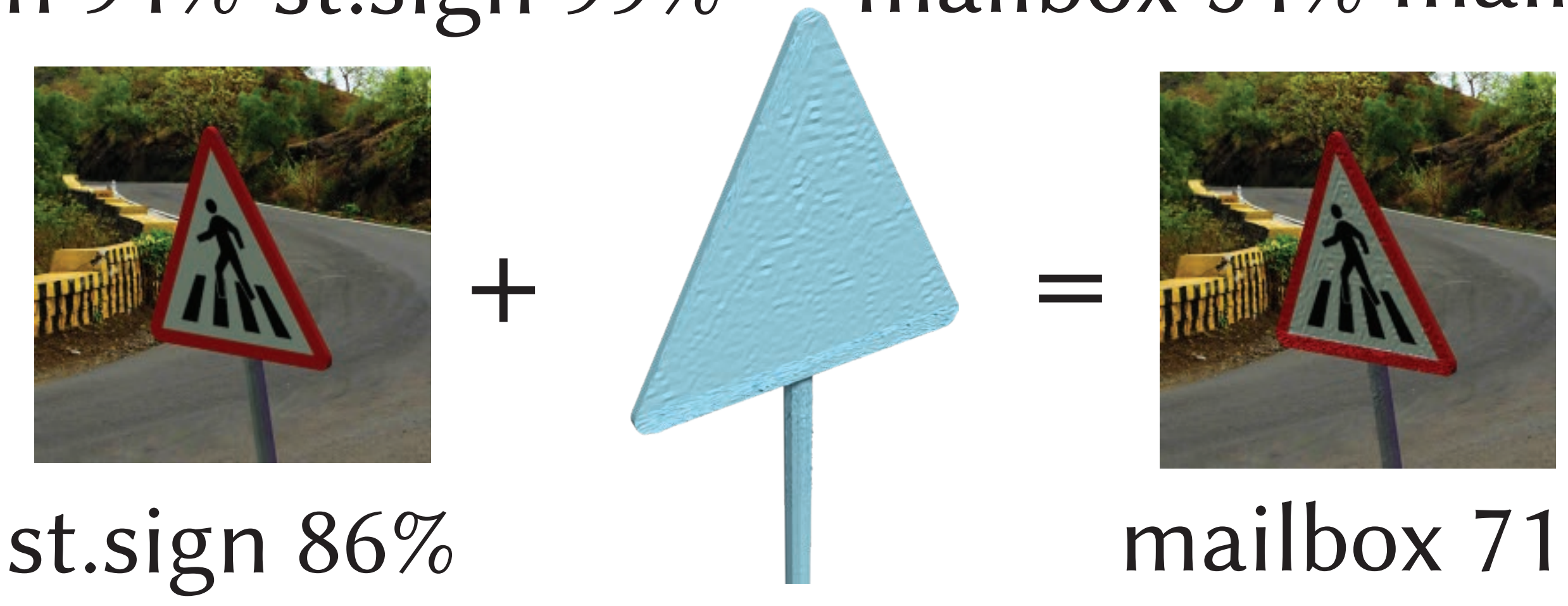
- standard data augmentation: 40.4%
- adversarial data augmentation: **65.8%**

Adversarial geometry



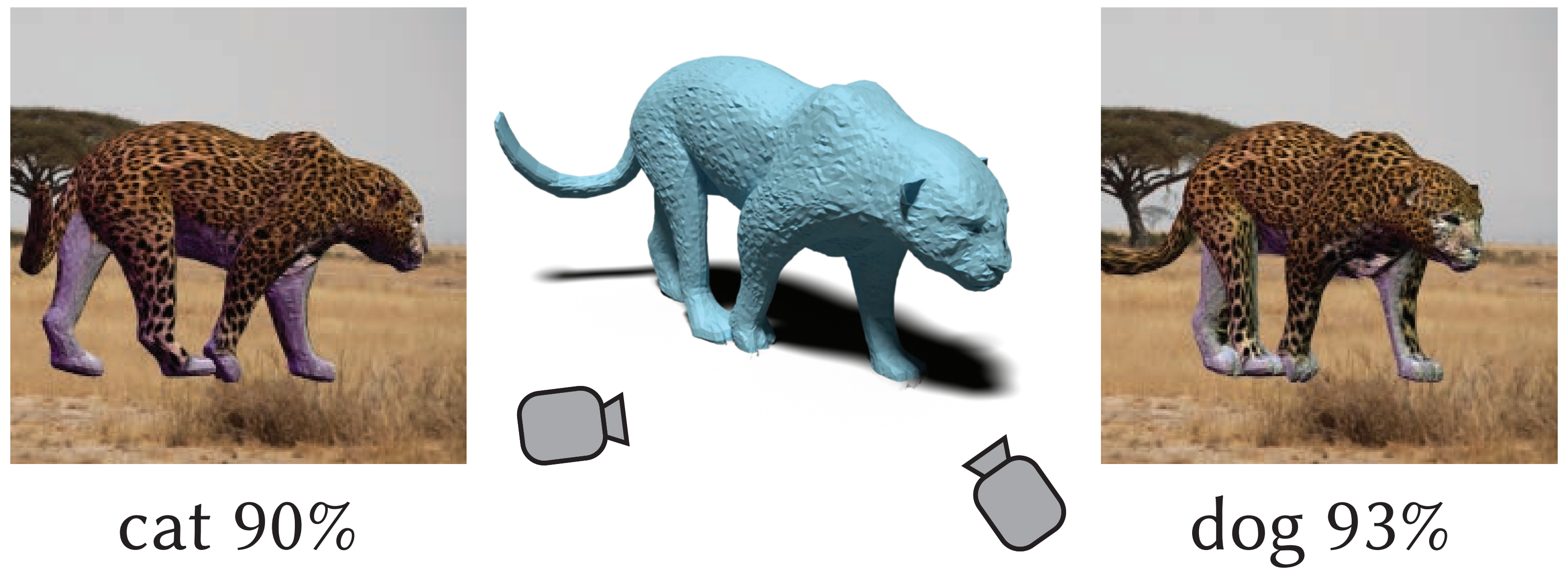
turtle 67% perturbation rifle 87%

st.sign 91% st.sign 99% mailbox 51% mailbox 61%



st.sign 86% mailbox 71%

Deep geometric illusion



cat 90% dog 93%

Future Work

- simulation/rendering adversarial training
- differential renderers for real images
- incorporating real-time rendering techniques
- more parametric perturfbations (e.g., adversarial human poses)

1 Computer Science UNIVERSITY OF TORONTO
2 Carnegie Mellon University
3 McGill