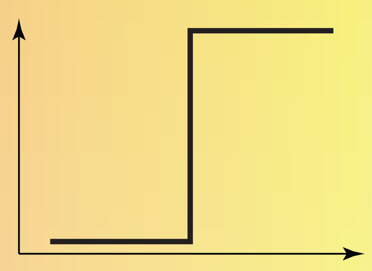
ZeroGrads: Learning Local Surrogate Losses for Non-Differentiable Graphics

MICHAEL FISCHER, TOBIAS RITSCHEL
UNIVERSITY COLLEGE LONDON

SIGGRAPH 2024
DENVER+ 28 JUL — 1 AUG

The Problem

- **Many programs are not differentiable**

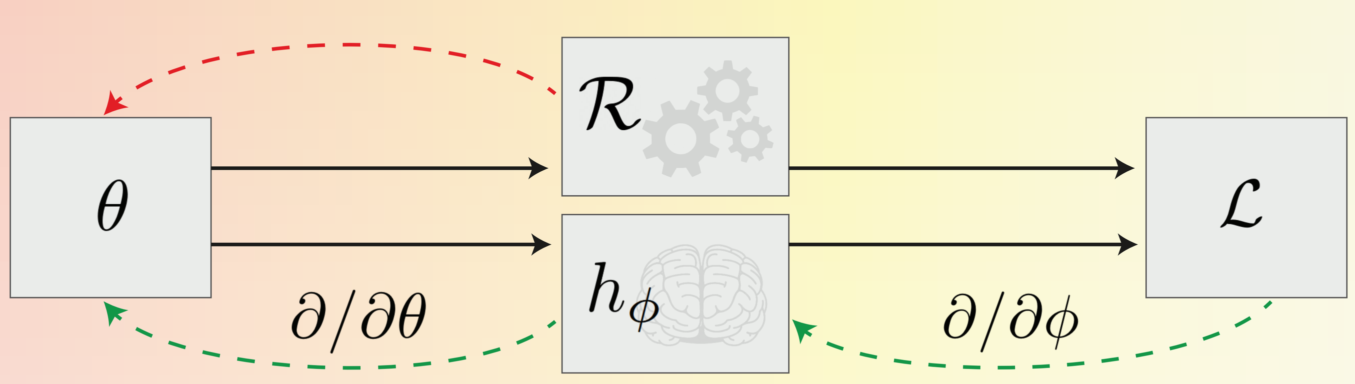


non-differentiable languages | discontinuous integrands | $\frac{\partial f(\theta)}{\partial \theta} = 0$ zero gradients or plateaus

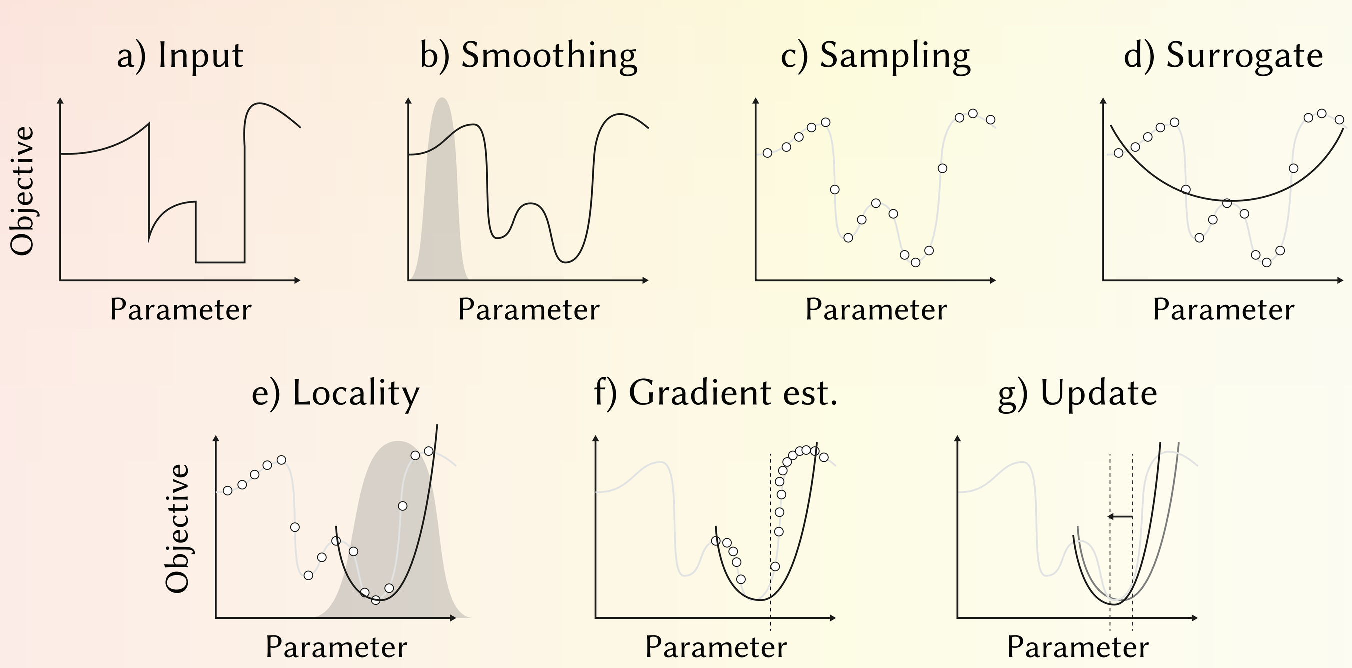
- **Specific solutions:** don't scale across apps
  e.g., what if we want to differentiate through Blender?
- **DFO algorithms:** don't scale w.r.t. dimensionality
  e.g., what if we want to optimize a triangle mesh?

Our Solution: ZeroGrads

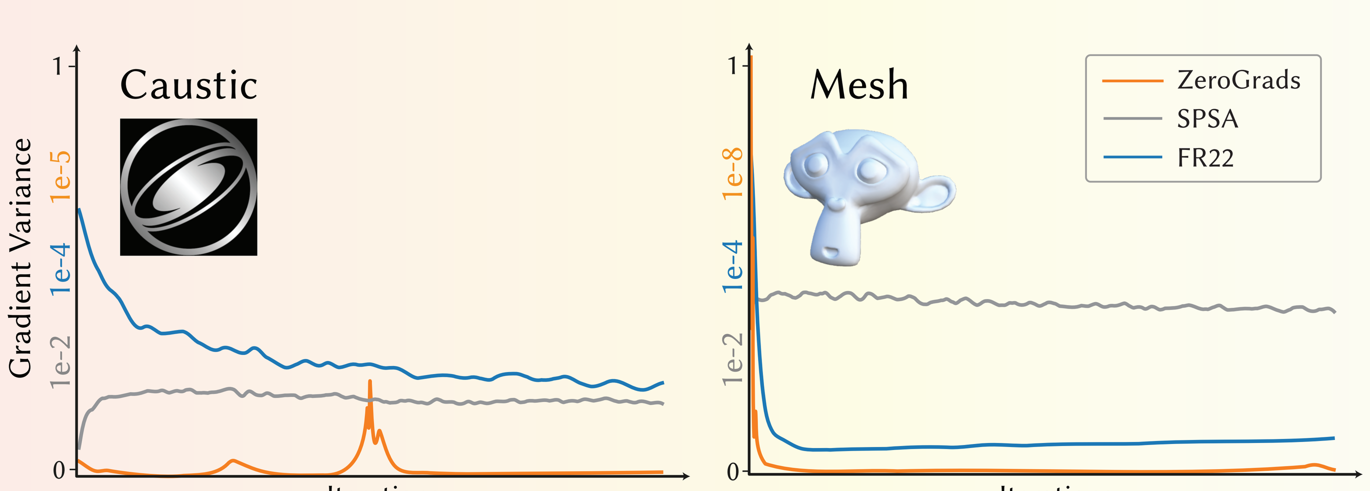
- We cannot **compute** the loss, but we can **sample** it!
- **Fit a function** to the samples: surrogate loss
- Surrogate loss: **analytical gradients**



The original loss might be...
- ... discontinuous: **smoothing**
- ... in large parts irrelevant: **locality**



- Surrogate learning: **self-supervised, on-the-fly**
- Surrogate function: **neural network**
- Network hysteresis: **smooth gradients** over time

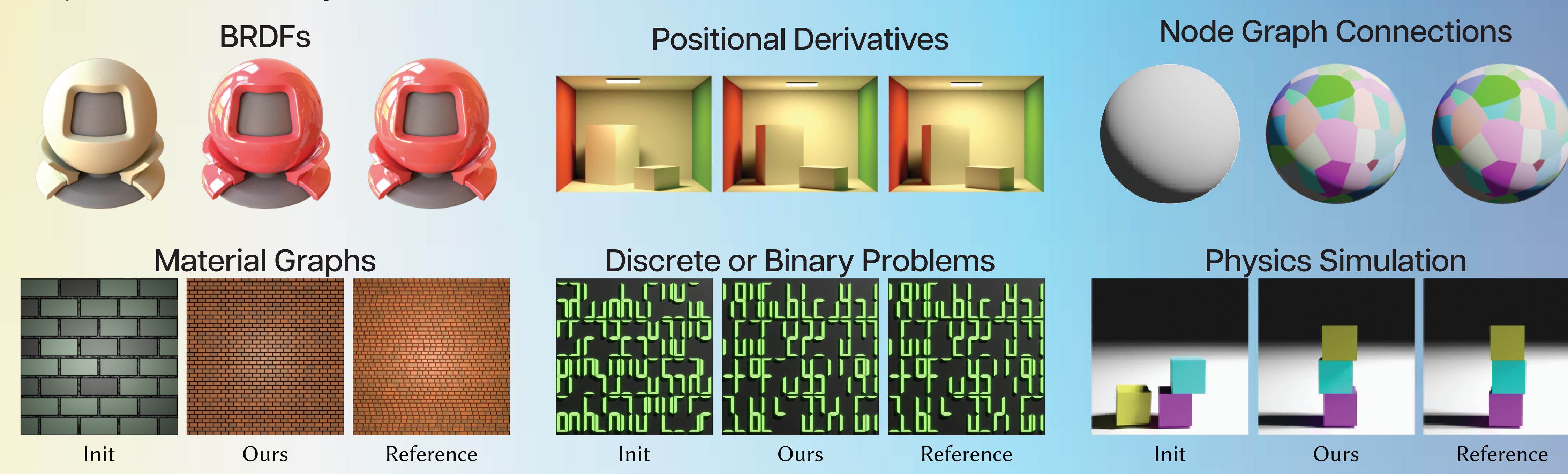


Acknowledgements

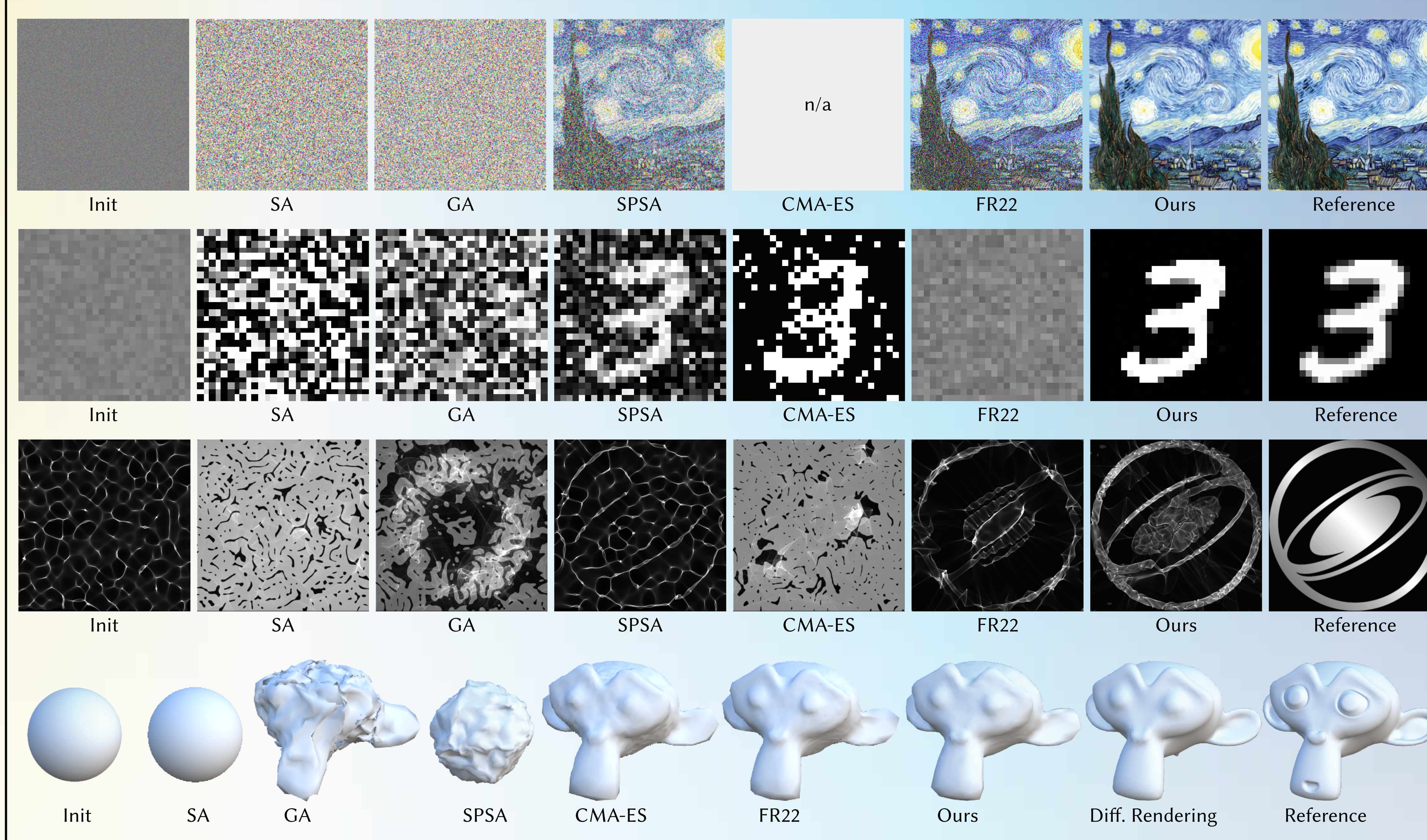
We thank the anonymous reviewers for their helpful feedback. We further thank Meta Reality Labs for their generous support over the years. MF is a recipient of the Rabin Ezra Scholarship.

Results - ZeroGrads can ...

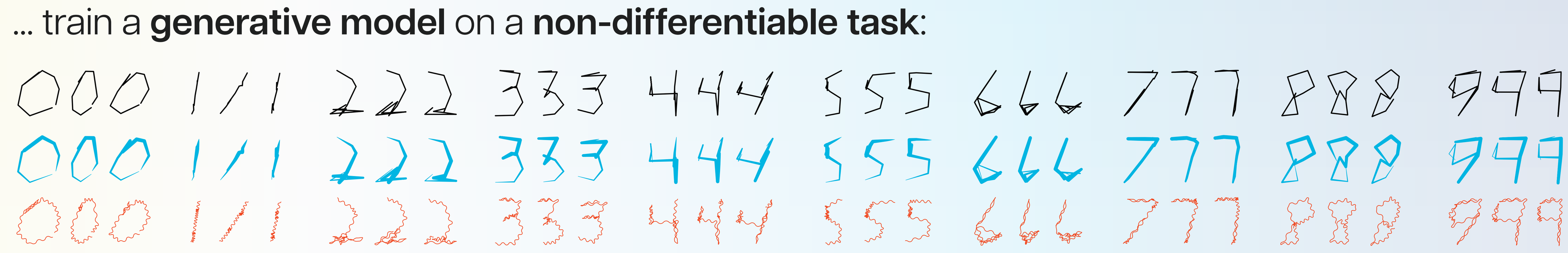
... optimize **arbitrary forward models**:



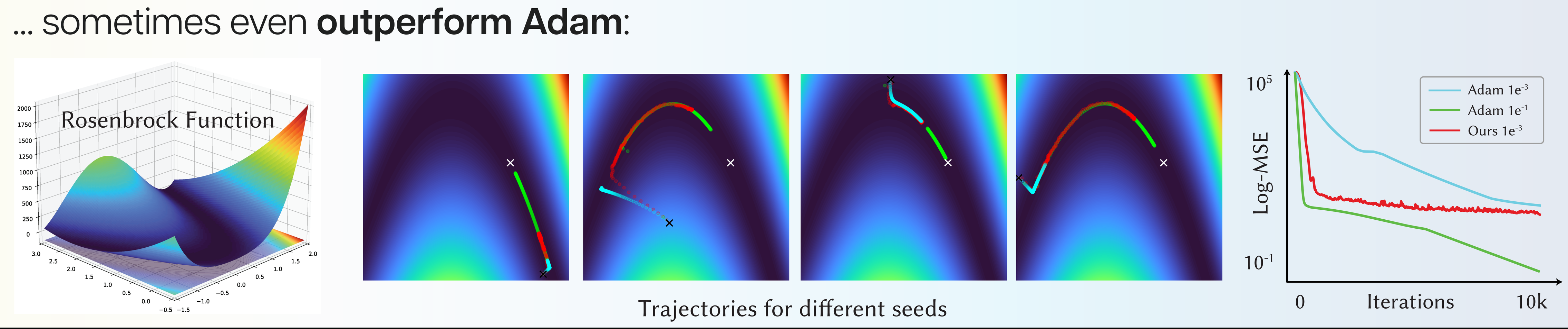
... scale to **higher-dimensional problems**:



... train a **generative model** on a **non-differentiable task**:



... sometimes even **outperform Adam**:



Trajectories for different seeds

Caveats & Limitations

- Higher-dimensional problems:
  require more samples, longer runtime
- No convergence guarantees:
  loss landscape might be too complex
- "No free lunch":
  if GD cannot work, ZeroGrads will not work either