PhsyicsGen: Can Generative Models Learn from Images to Predict Complex Physical Relations?

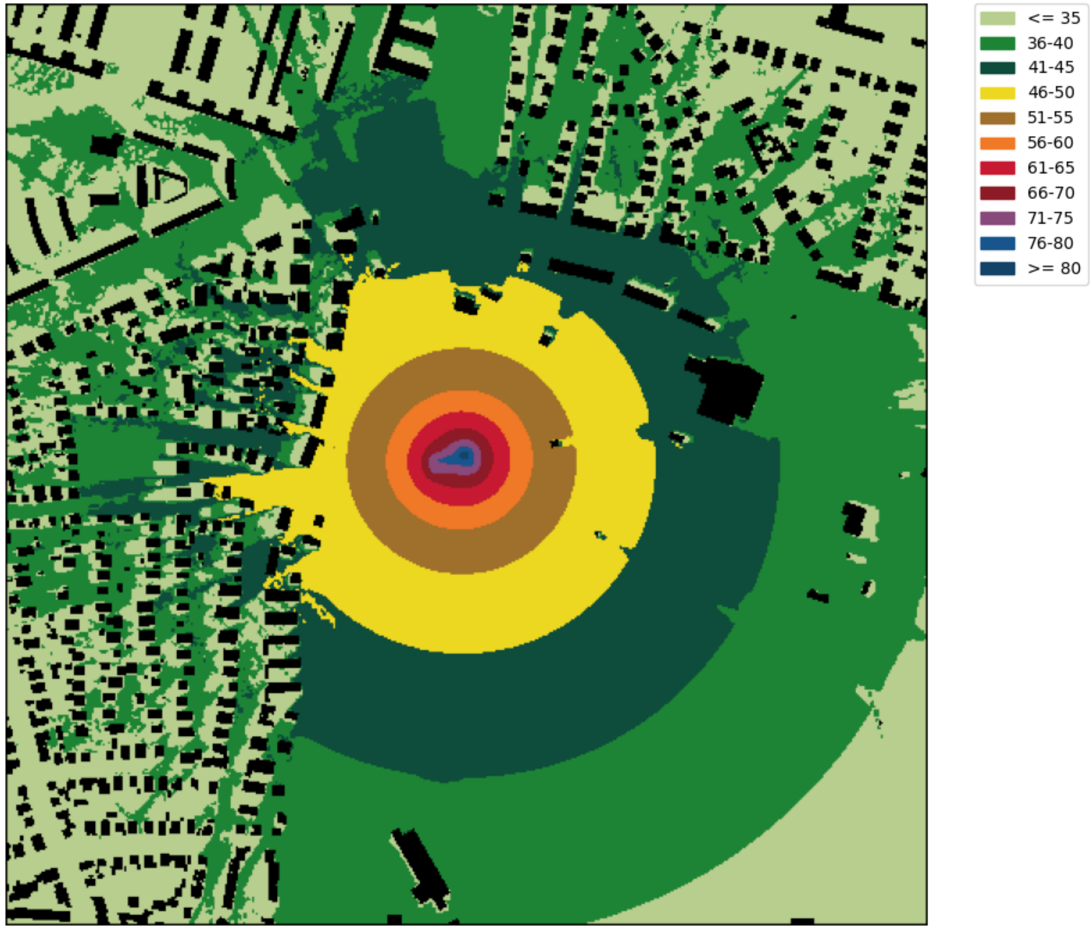
Martin Spitznagel, Jan Vaillant, Janis Keuper

INSTITUTE FOR MACHINE LEARNING AND ANALYTICS

HOCH SCHULE OFFEN BURG

GEFÖRDERT VOM Bundesministerium für Bildung und Forschung

Objective: Active noise reduction through AI-based control of drilling rigs

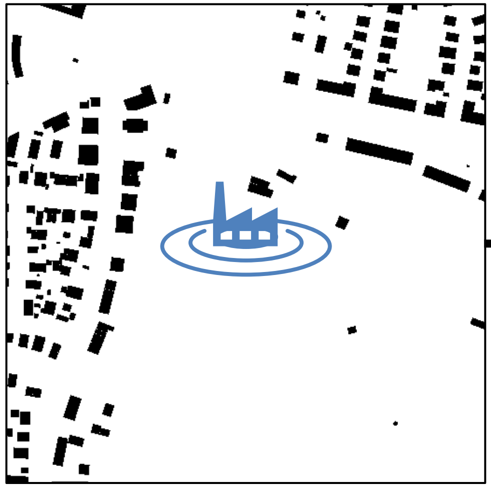


HERRENKNECHT VERTICAL

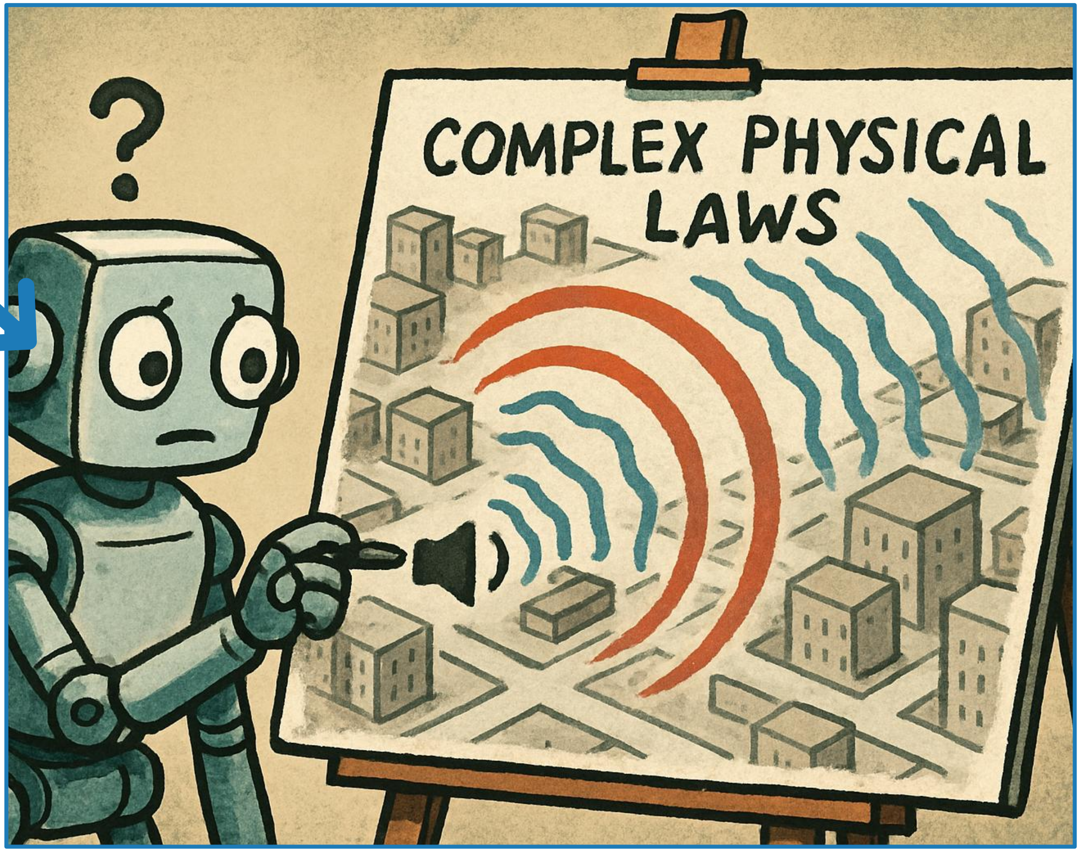
HOCH SCHULE OFFEN BURG

- ~9–10 min simulation time

Introduction & Motivation



DL-model

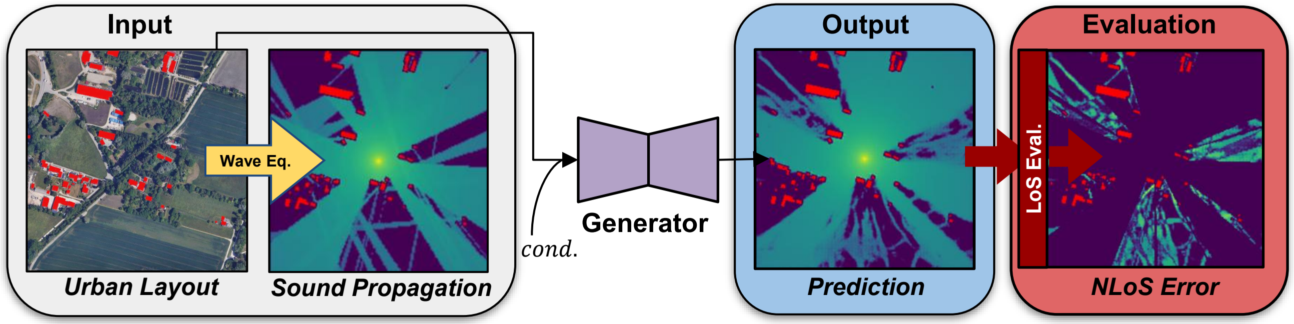


Are generative models able to learn **complex physical** relations from input-output image pairs?

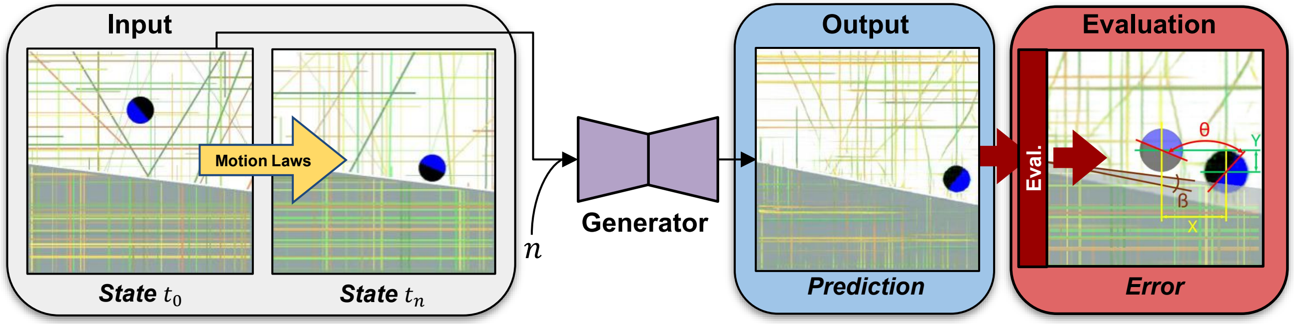
What **speedups** can be achieved by replacing differential equation-based simulations?

PhysicsGen Benchmark – The Three Tasks

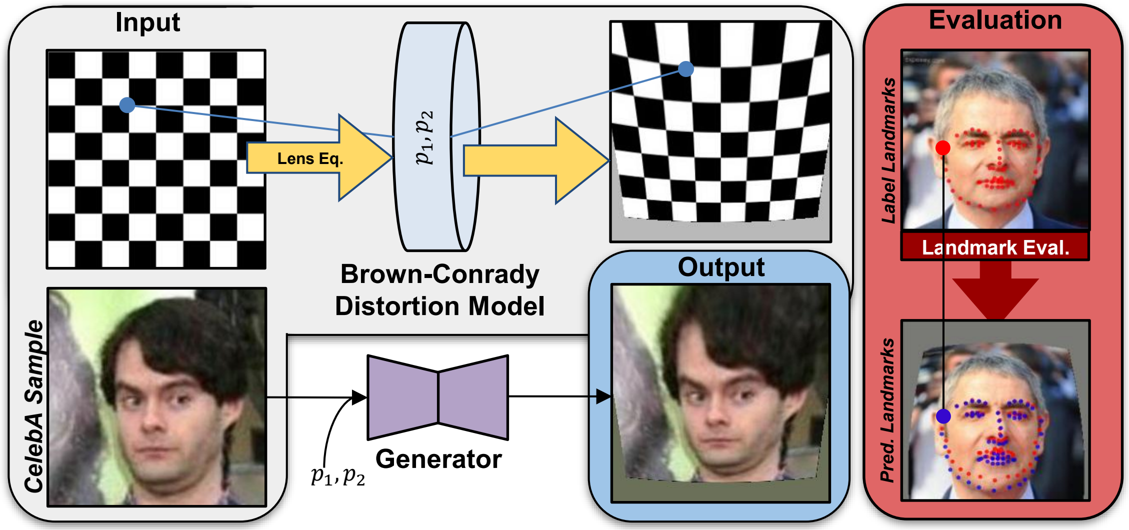
Urban Sound Propagation



Dynamics of rolling and bouncing movements



Lens Distortion

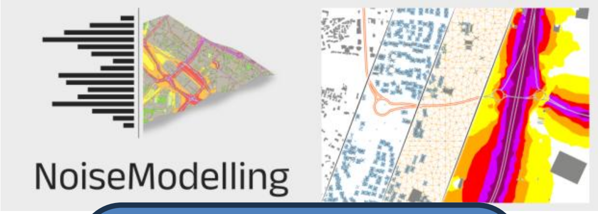


Urban Sound Propagation

PhysicsGen Task 1

100.000 Pairs

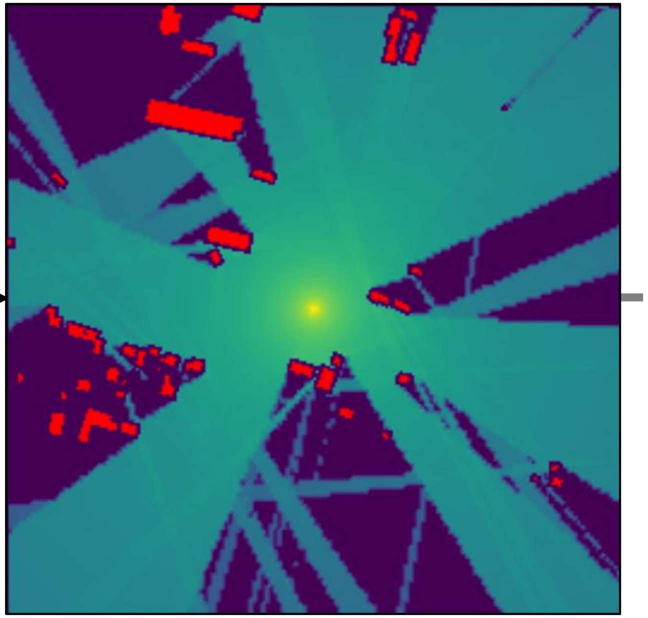
OSM



Sound Propagation Simulation

- Decibels (dB)
- Humidity, Temperature
- Order of Reflections

Prediction



Doc.: https://noisemodelling.readthedocs.io/en/latest/

PhysicsGen Task 2

Goal: Simulate optical lens distortion

- CelebA subset (50k faces, 256×256)

100.000 Pairs

Input

Brown-Conrady Distortion Model

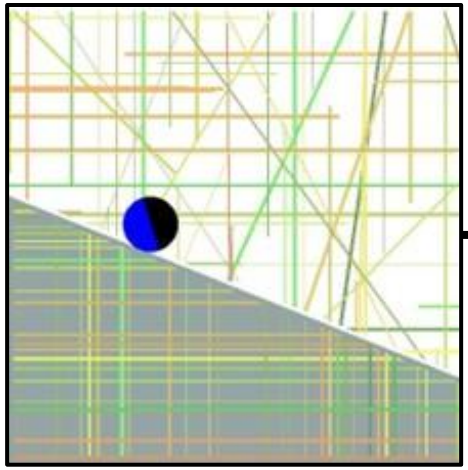
- distortion parameters (p1,p2)

Target Distortion

PhysicsGen Task 3

100.000 Pairs

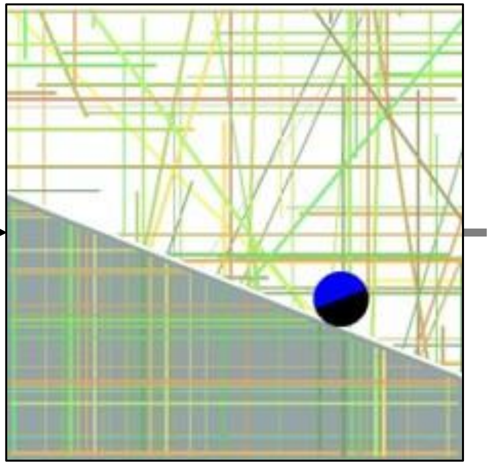
Input t_0



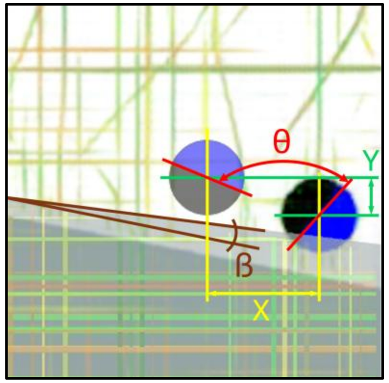
Pymunk Simulation

- Inclination
- Start height
- Time interval

Target t_n

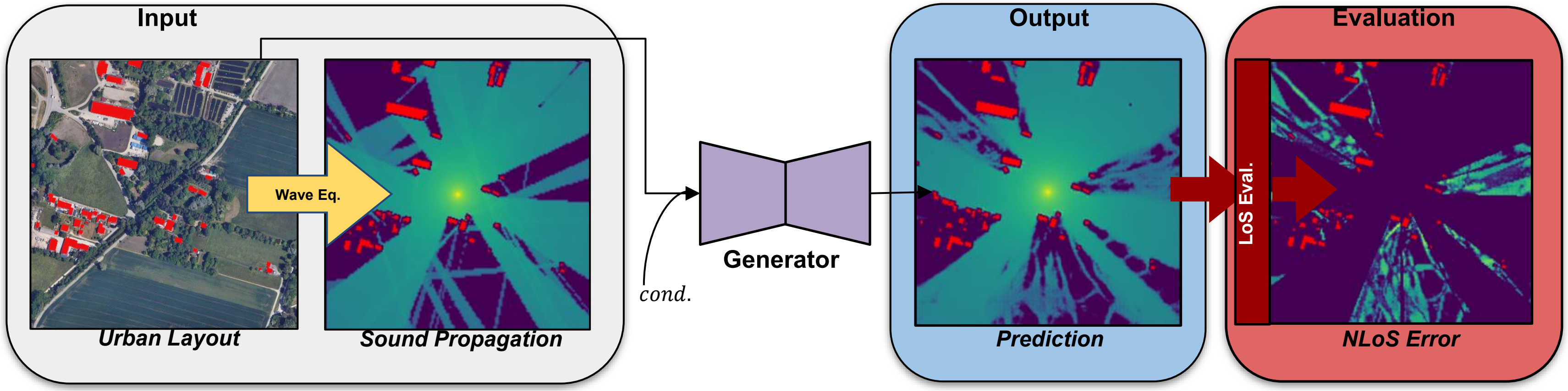


Evaluation:



- Ball position X, Y
- Rotation θ
- Ground inclation β

PhysicsGen Benchmark – Generative Models



Encoder-Decoder Architectures

- U-Net
- Convolutional Autoencoder (ConvAE)
- Variational Autoencoder (VAE)

Generative Adversarial Network

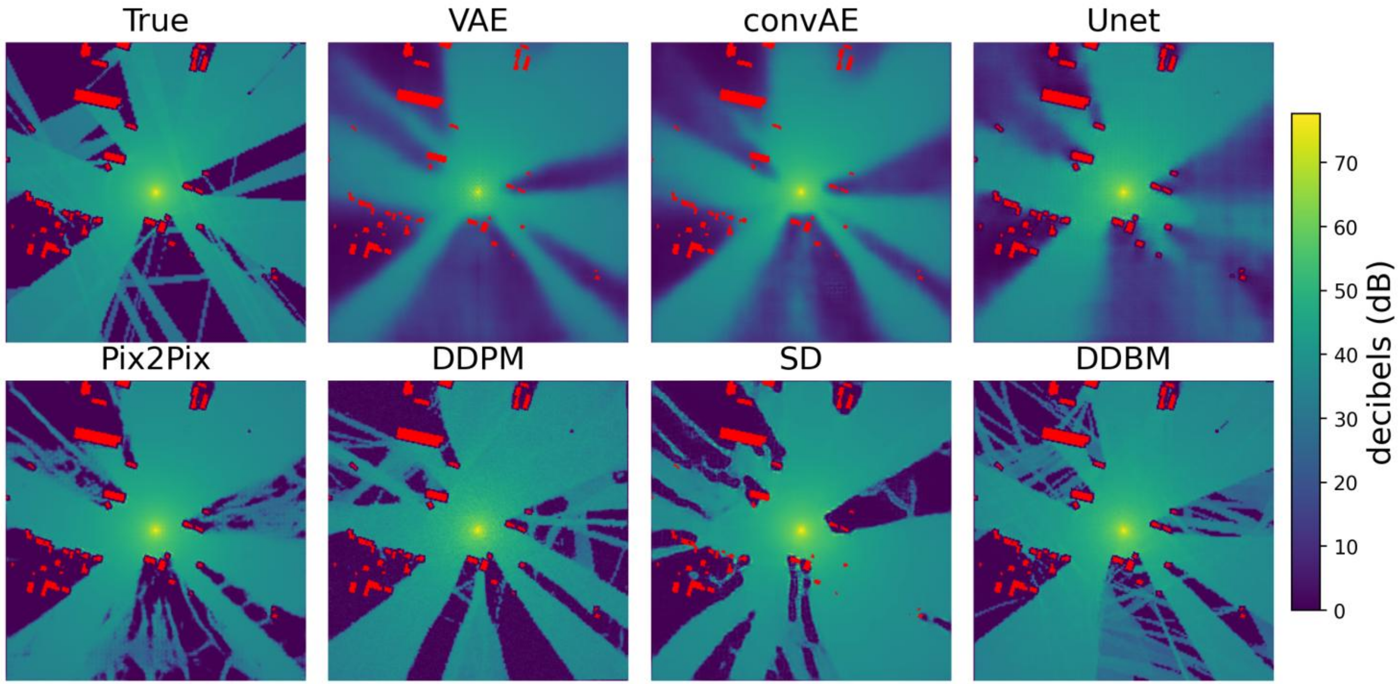
- Pix2Pix

Diffusion Models

- Denoising Diffusion Propabilistic Model (DDPM)
- Stable Diffusion
    - conditioning with and without cross-attention
- Denoising Diffusion Bridge Models (DDBM)

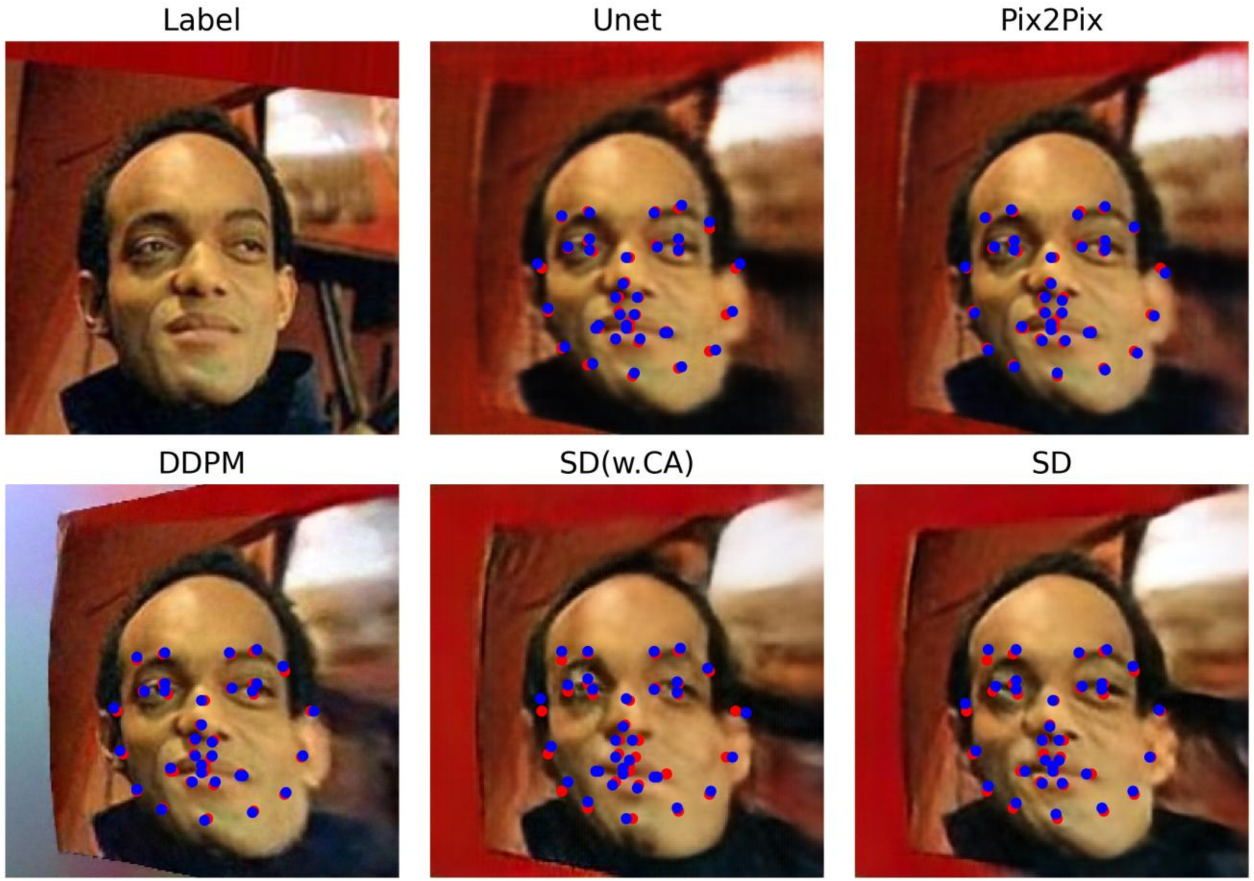
INSTITUTE FOR MACHINE LEARNING AND ANALYTICS

Reflection Task



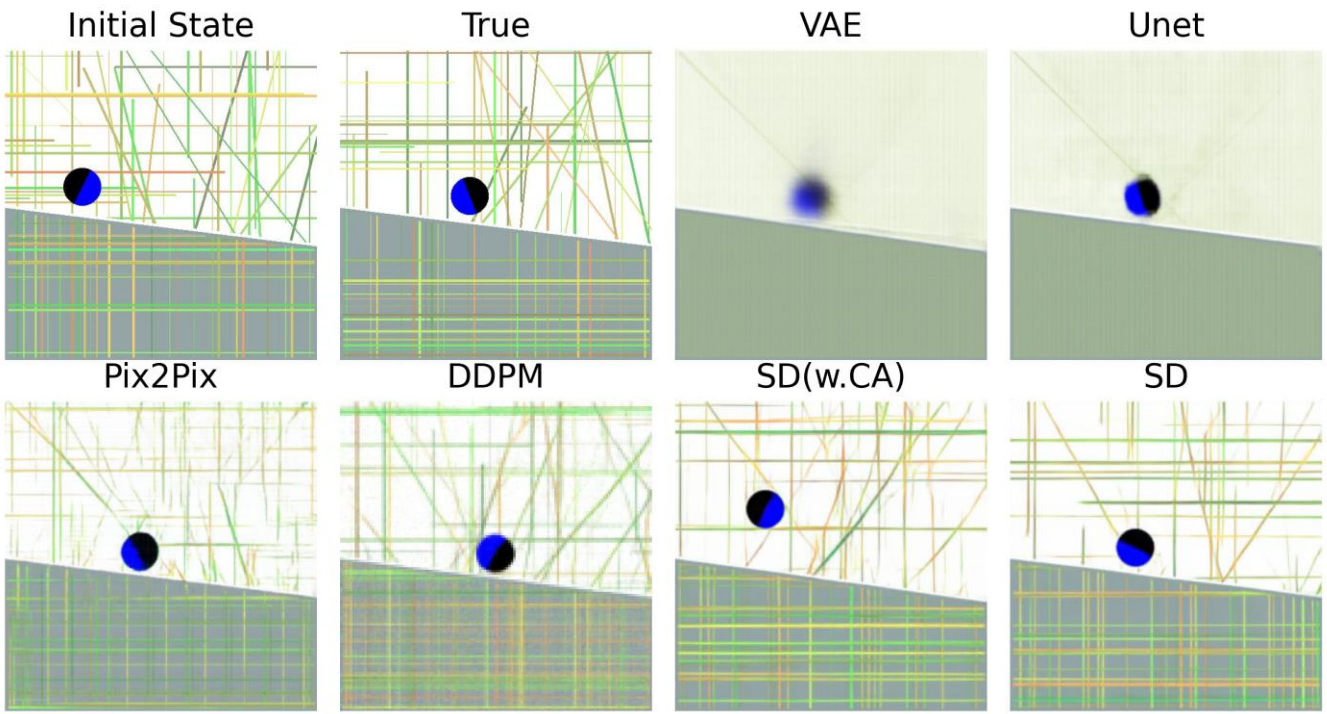
Model	Task	MAE ↓ LoS	MAE ↓ NLoS	wMAPE ↓ LoS	wMAPE ↓ NLoS	Runtime/ ↓ Sample (ms)
Simulation	Base	0.00	0.00	0.00	0.00	204700
convAE	Base	3.67	2.74	20.24	67.13	0.128
VAE [25]	Base	3.92	2.84	21.33	75.58	0.124
UNet [30]	Base	2.29	1.73	12.91	37.57	0.138
Pix2Pix [20]	Base	1.73	1.19	**9.36**	**6.75**	0.138
DDPM [18]	Base	2.42	3.26	15.57	51.08	3986.353
SD(w.CA) [29]	Base	3.76	3.34	17.42	35.18	2961.027
SD	Base	2.12	**1.08**	13.23	32.46	2970.86
DDBM [36]	Base	**1.61**	2.17	17.50	65.24	3732.21
Simulation	Refl.	0.00	0.00	0.00	0.00	251000
convAE	Refl.	3.83	6.56	20.67	93.54	0.128
VAE	Refl.	4.15	6.32	21.57	92.47	0.124
UNet	Refl.	2.29	5.72	12.75	80.46	0.138
Pix2Pix	Refl.	2.14	**4.79**	**11.30**	**30.67**	0.138
DDPM	Refl.	2.74	7.93	17.85	80.38	3986.353
SD(w.CA)	Refl.	3.81	6.82	19.78	81.61	2961.027
SD	Refl.	2.53	5.26	15.04	55.27	2970.86
DDBM	Refl.	**1.93**	6.38	18.34	79.13	3732.21

Lens Distortion



Model	Comb ↓	X ↓ Err	Y ↓ Err	|X-Y| ↓ Shift	Runtime/ Sample (ms)
			$p1 \neq 0, p2 = 0$		
Sim.	0.00	0.00	0.00	0.00	153.205
convAE	11.93	6.75	8.13	1.38	0.110
VAE	11.53	6.55	7.83	1.28	0.122
UNet	2.82	1.28	2.15	0.87	0.118
Pix2Pix	2.00	0.99	1.43	**0.44**	0.122
DDPM	**1.93**	**0.94**	**1.39**	0.45	3970.603
SD(w.CA)	3.09	1.59	2.21	0.62	2991.678
SD	2.79	1.41	2.01	0.60	2997.576
			$p1 = 0, p2 \neq 0$		
Sim.	0.00	0.00	0.00	0.00	153.205
convAE	10.56	8.35	4.77	2.21	0.110
VAE	10.40	8.26	4.62	3.64	0.122
UNet	2.36	1.33	1.60	0.27	0.117
Pix2Pix	**1.77**	**1.02**	**1.14**	**0.13**	0.123
DDPM	2.13	1.39	1.23	0.16	3970.603
SD(w.CA)	2.85	1.60	1.94	0.34	2991.678
SD	2.44	1.38	1.64	0.26	2997.576

Dynamics of Movement



	Position X	Position Y	Rotation	Roundness	Error
convAE	4.24 ± 3.9	6.08 ± 5.9	12.2 ± 8.6	1.06 ± 0.0	99%
VAE	4.69 ± 6.1	6.25 ± 6.9	31.0 ± 40	0.90 ± 0.1	95%
UNet	5.53 ± 7.5	10.8 ± 12	15.2 ± 23	0.74 ± 0.2	28%
Pix2Pix	**6.28 ± 8.0**	**11.7 ± 13**	**17.2 ± 21**	**0.56 ± 0.1**	11%
DDPM	7.91 ± 9.0	15.5 ± 14	32.9 ± 34	0.61 ± 0.2	5.7%
SD(w.CA)	40.0 ± 49	24.8 ± 23	61.1 ± 52	**0.53 ± 0.2**	7.3%
SD	8.55 ± 12	16.2 ± 14	34.2 ± 38	**0.47 ± 0.1**	**2%**

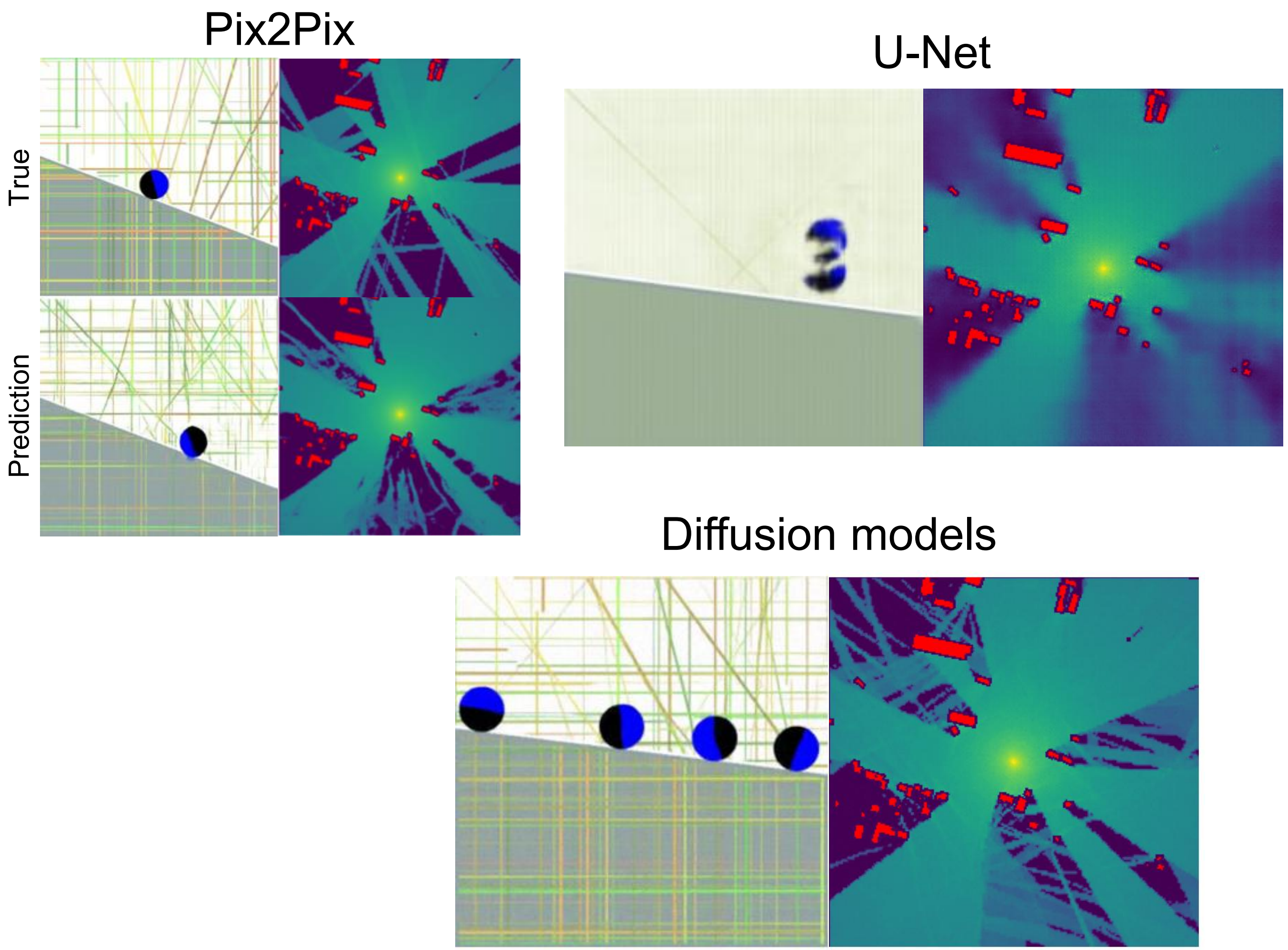
Overall Discussion & Key Insights

Are generative models able to learn **complex physical** relations from input-output image pairs?

- Good for simple 0^{th} and 1^{st} order problems
- Struggle with complex physical relations

What **speedups** can be achieved by replacing differential equation-based simulations?

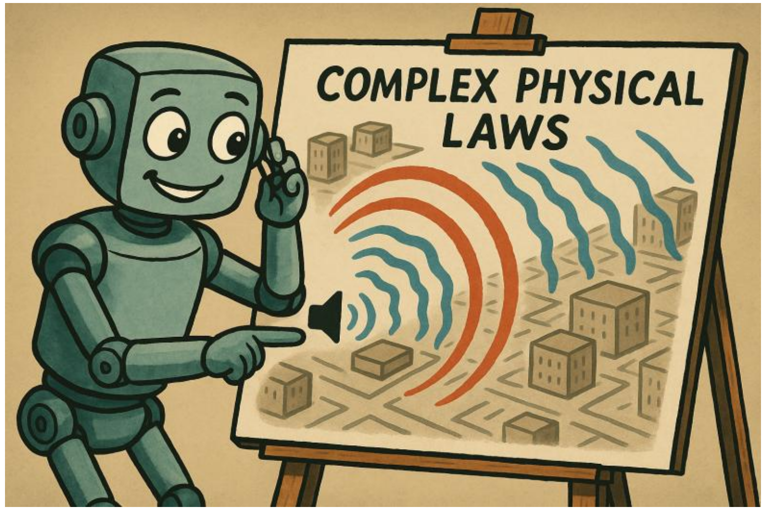
- 2M speedup in sound propagation
- Runtime comparison is practical; CPU vs. GPU



Limitations & Conclusion

Conclusion

- PhysicsGen provides a valuable benchmark for evaluating generative models in physical simulation tasks
- Current models show great promise for speed but require significant advancements to ensure physical correctness



Hugging Face

```python
from datasets import load_dataset

dataset = load_dataset("mspitzna/physicsgen", name="sound_combined", trust_remote_code=True)

# Access a sample from the training split.
sample = dataset["train"][0]

input_img = sample["osm"]
target_img = sample["soundmap_512"]
```

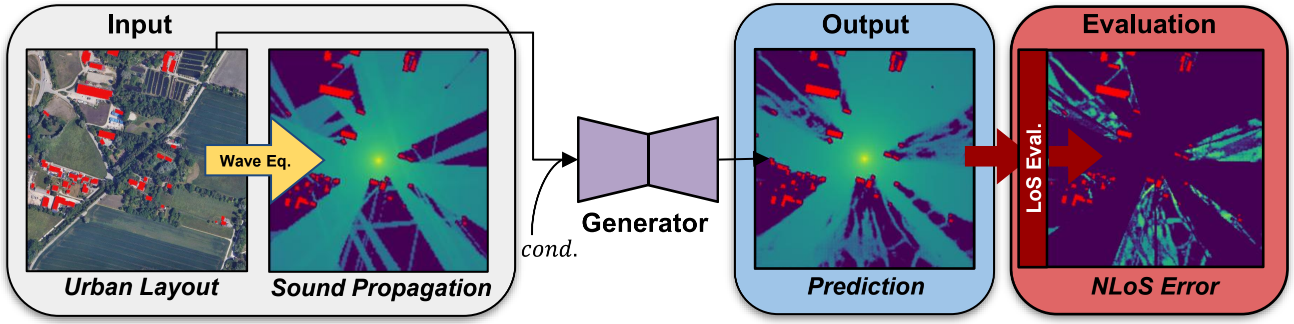
@mspitzna/physicsgen

- Project: https://www.ki-bohrer.de/
- PhysicsGen: https://www.physics-gen.org/

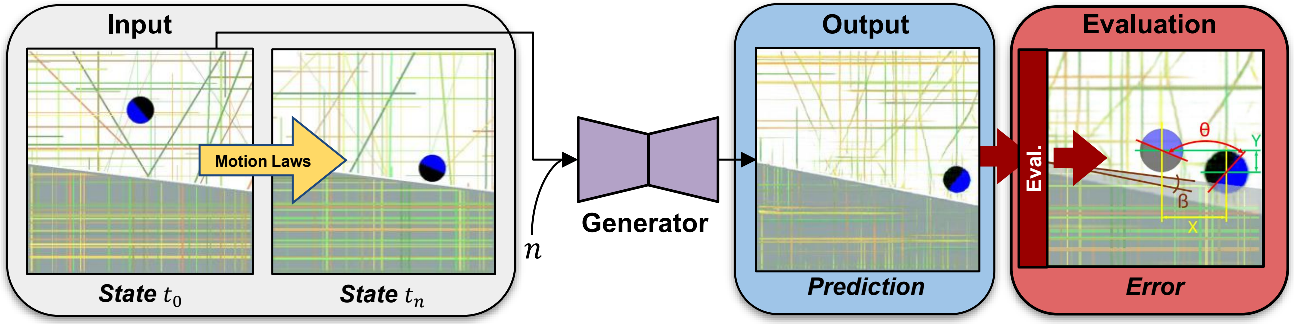
PhysicsGen Benchmark

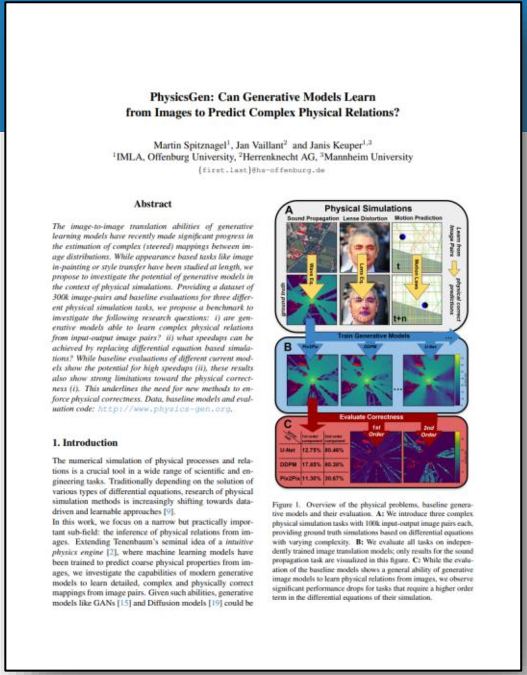
PhysicsGen: Can Generative Models Learn from Images to Predict Complex Physical Relations?

Urban Sound Propagation



Dynamics of rolling and bouncing movements





PhysicsGen: Can Generative Models Learn from Images to Predict Complex Physical Relations?

Martin Spitznagel[1], Jan Vaillant[2] and Janis Keuper[1,3]
[1]IMLA, Offenburg University, [2]Herrenknecht AG, [3]Mannheim University
{first.last}@hs-offenburg.de

Abstract

The image-to-image translation abilities of generative learning models have recently made significant progress in the estimation of complex (steered) mappings between image distributions. While appearance based tasks like image in-painting or style transfer have been studied at length, we propose to investigate the potential of generative models in the context of physical simulations. Providing a dataset of 300k image-pairs and baseline evaluations for three different physical simulation tasks, we propose a benchmark to investigate the following research questions: i) are generative models able to learn complex physical relations from input-output image pairs? ii) what speedups can be achieved by replacing differential equation based simulations? While baseline evaluations of different current models show the potential for high speedups (ii), these results also show strong limitations toward the physical correctness (i). This underlines the need for new methods to enforce physical correctness. Data, baseline models and evaluation code: http://www.physics-gen.org.

1. Introduction

The numerical simulation of physical processes and relations is a crucial tool in a wide range of scientific and engineering tasks. Traditionally depending on the solution of various types of differential equations, research of physical simulation methods is increasingly shifting towards data-driven and learnable approaches [9].
In this work, we focus on a narrow but practically important sub-field: the inference of physical relations from images. Extending Tenenbaum's seminal idea of a *intuitive physics engine* [2], where machine learning models have been trained to predict coarse physical properties from images, we investigate the capabilities of modern generative models to learn detailed, complex and physically correct mappings from image pairs. Given such abilities, generative models like GANs [15] and Diffusion models [19] could be

Figure 1. Overview of the physical problems, baseline generative models and their evaluation. **A:** We introduce three complex physical simulation tasks with 100k input-output image pairs each, providing ground truth simulations based on differential equations with varying complexity. **B:** We evaluate all tasks on independently trained image translation models; only results for the sound propagation task are visualized in this figure. **C:** While the evaluation of the baseline models shows a general ability of generative image models to learn physical relations from images, we observe significant performance drops for tasks that require a higher order term in the differential equations of their simulation.

Accepted at CVPR Nashville JUNE 11-15, 2025

Lens Distortion