Joint Feature Selection and Reconstruction

Klas Wijk, supervised by Hossein Azizpour and Ricardo Vinuesa

Optimized measurements of complex, nonlinear systems.

We use machine learning to reconstruct data from sparse measurements while tailoring these measurements for the best possible results.

Keywords

- Machine learning
- Feature selection
- Fluid mechanics

Our impact case is understanding and reducing air pollution in urban environments.

1 Theory: Feature Selection

Feature selection is a fundamental problem in statistics and machine learning: What subset of input values give the best prediction?

The answer is task-dependent. Recent theoretical developments have made tailored (jointly optimized) selections computationally feasible.

Techniques
- Differentiable modeling
- Joint optimization

Metrics
- Task performance
- Accuracy on synthetic datasets

2 Application: Turbulent Flows

Turbulent flows are highly complex physical phenomena governed by the Navier-Stokes equations.

Using data from accurate simulations, we train machine learning models that reconstruct the flow from sparse measurements.

Techniques
- Deep generative models
- Physics constrained learning

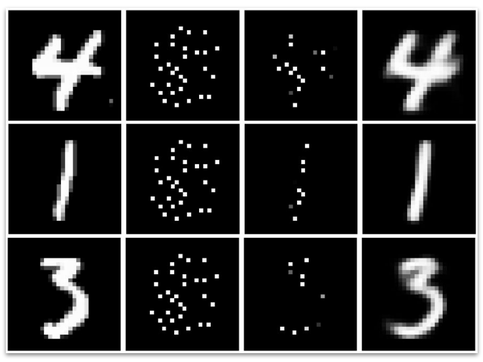
Metrics
- Reconstruction quality
- Physical accuracy

3 Impact: Urban Flows

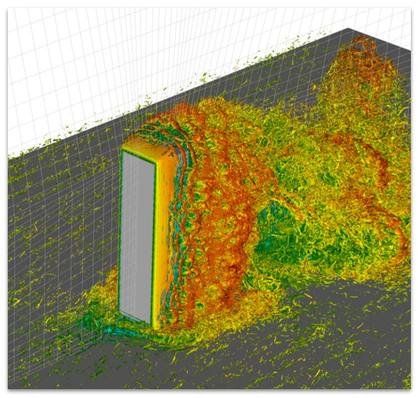
Air pollution poses a health risk to much of the world's population, especially those living in urban areas.

Improved analysis of urban flows can help mitigate these issues.

Measurable Impact
- Particles
- Gases
- Health complications



MNIST, LeCun et al 1998



Turbulent flow, Martínez Sánchez et al. 2023

Photo by Kristen Morith on Unsplash

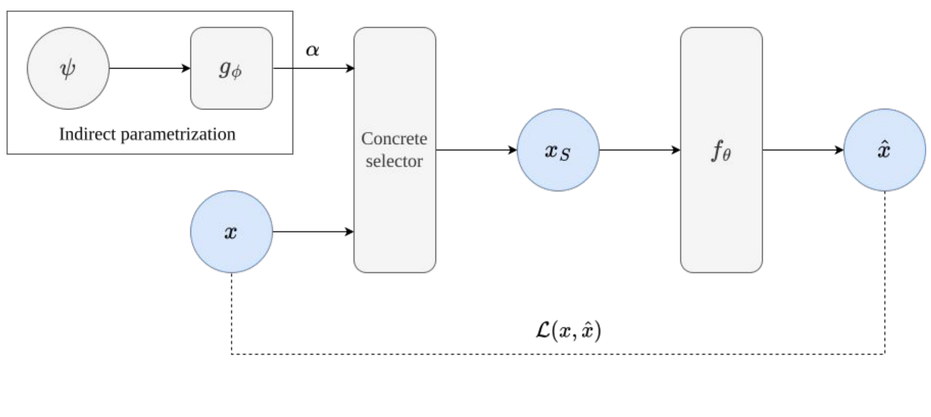
Indirectly Parameterized Concrete Autoencoders

Alfred Nilsson, Klas Wijk, Sai bharath chandra Gutha, Erik Englesson, Alexandra Hotti, Carlo Saccardi, Oskar Kviman, Jens Lagergren, Ricardo Vinuesa, Hossein Azizpour

We improve concrete autoencoders (Balin et al. 2019) by introducing an indirect parametrization of the concrete selector layer, resulting in significantly faster training and fewer redundant selections.

ICML International Conference On Machine Learning

arXiv



Broader Impact

The fundamental methods developed can be used in many other areas:

- Medical Imaging
- Biological sequences
- General machine learning

Funding

- Allows for development of machine learning methods whose relevance go beyond our application.
- Helped securing additional funding which is now contributing to this project.
- Modeling air pollution is important for sustainable urban life and aligned with UN SDG 11.