

Visual Exploration of Latent Spaces in Deep Neural Network Models

Shusen Liu

Postdoc Researcher, Data Analysis Group, CASC



08/08/2018



Latent Space in Deep Neural Network

- Latent space captures the innate variations of the data

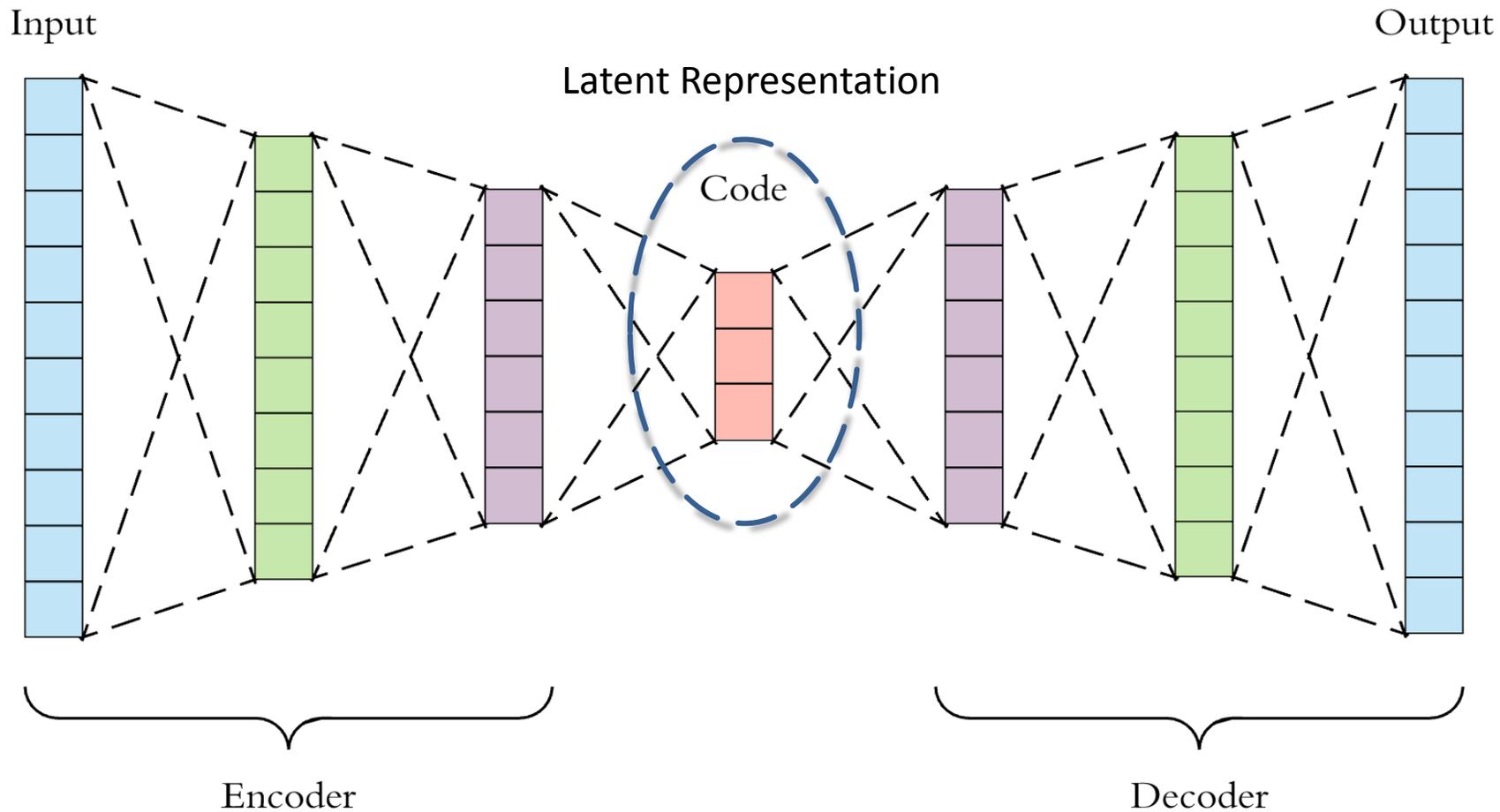
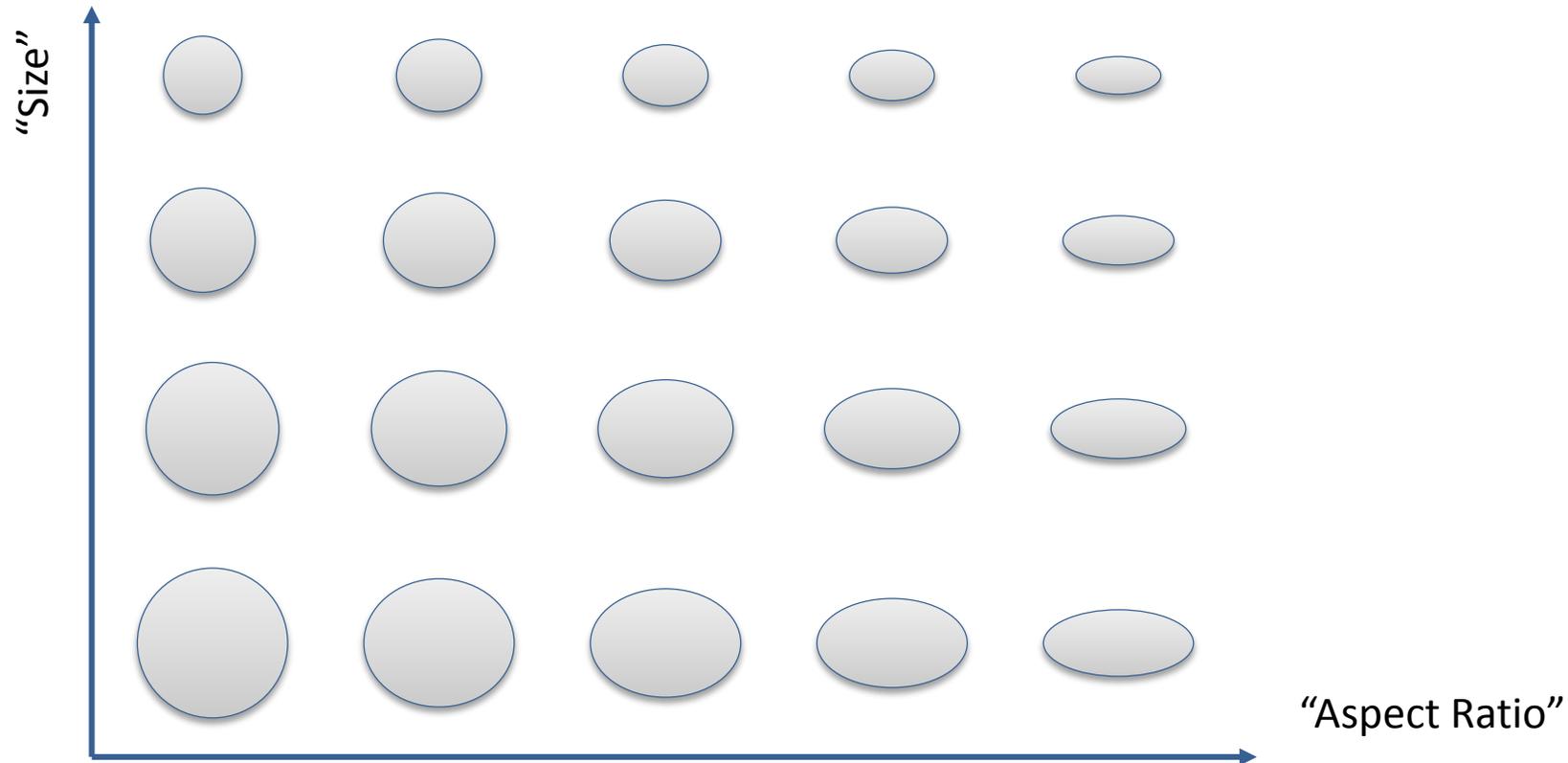


Image source:
<https://towardsdatascience.com/>

Latent Space in Deep Neural Network

- Essential for understanding the underlying pattern of the data



Latent Space of Generative Models

- Generative Models
 - Variational Autoencoder (VAE)
 - Generative Adversarial Network (GAN)
 - Generate new instances by sampling / interpolating in the learned latent space

Latent Space of Generative Models

- Generative Models

- Variational Autoencoder (VAE)
- Generative Adversarial Network (GAN)
- Generate new instances by sampling / interpolating in the learned latent space

Interpolation in Latent Space



[Wu et al. 2017]



[Karras et al. 2017]

The Challenges in Making Sense of the Latent Spaces

- The high-dimensional nature of the representation
 - It is hard for human to understand a space with a dimension higher than 3
- There is no explicit meaning for the dimension in the latent space
 - The meaning of distances and locations within such a space are not clearly defined
- Require both local and global interpretation
 - What are the nearby samples of a given point in the latent space
 - Are there distinct regions in the latent space
- Existing interpretability works mostly focus on supervised models

The State-of-the-Art of Model Interpretation in Deep Learning

- **Attribution:**

Label: bow tie



Which part of image leads to the prediction?



- **Feature Visualization:**

Label: bow tie



What feature does a part of the network seek?

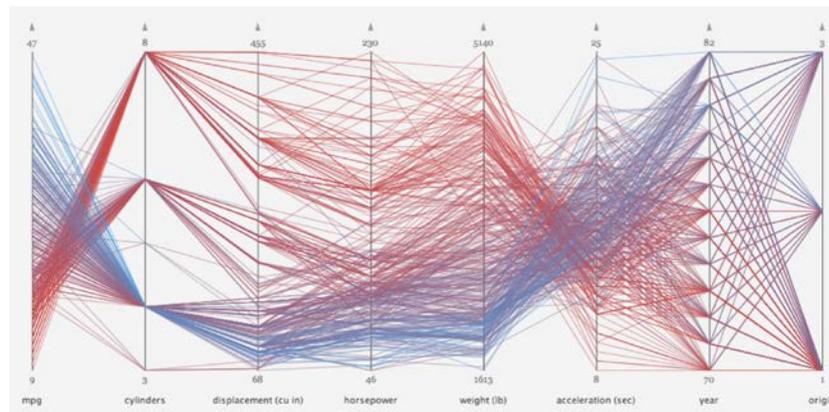


Image source:
<https://distill.pub/2018/building-blocks/>

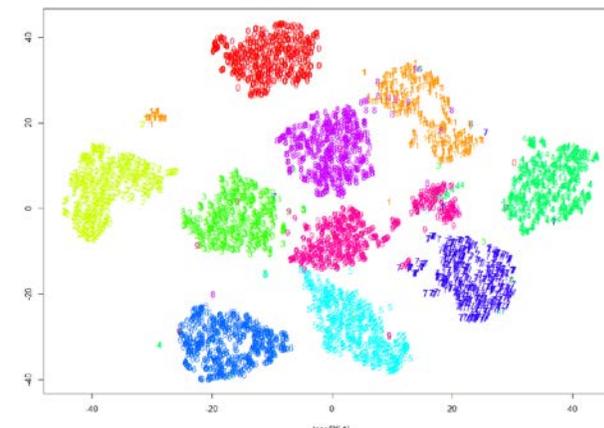
- Few works focused on the interpretability of the latent space

Interactive Data Visualization as a Tool for Latent Space Exploration

- Intuitive visual encodings
 - Parallel Coordinate
 - Dimensionality Reduction (t-SNE)
 - Scatterplot Matrix



Parallel Coordinates

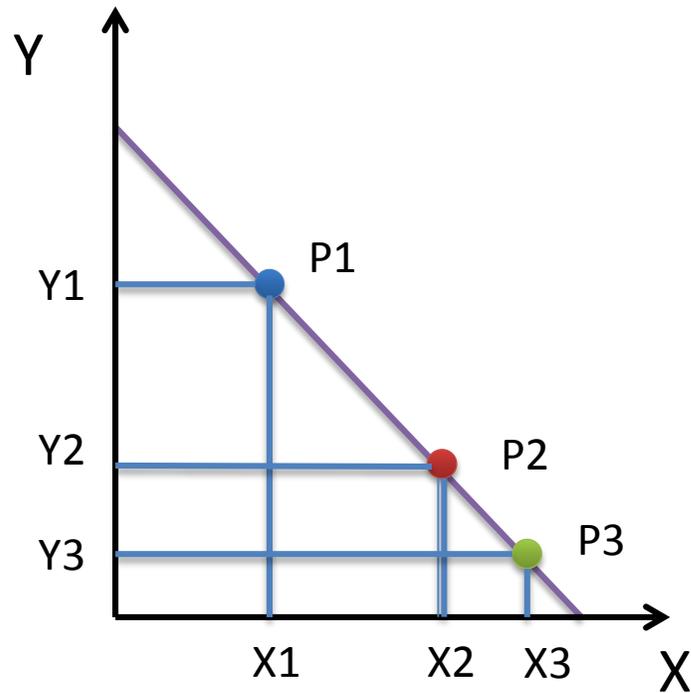


t-distributed stochastic neighborhood embedding (t-SNE)

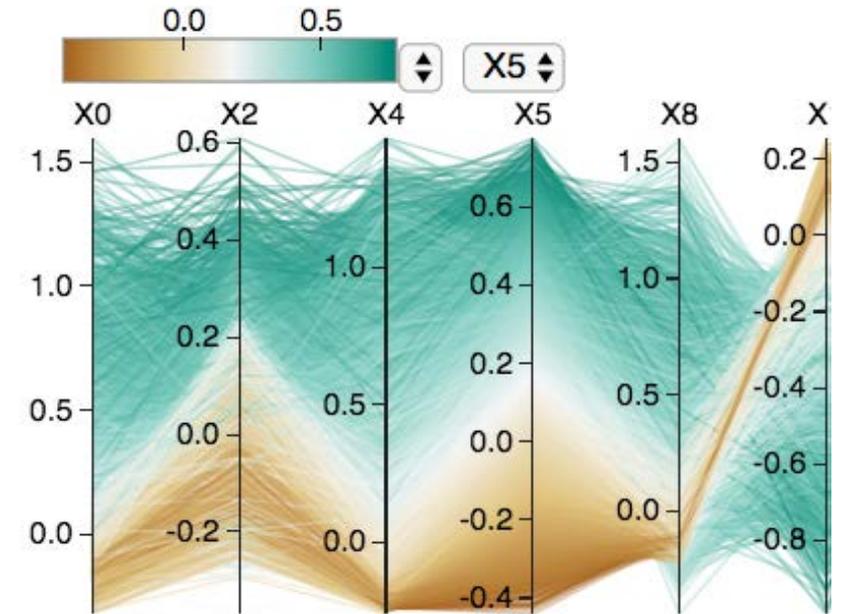
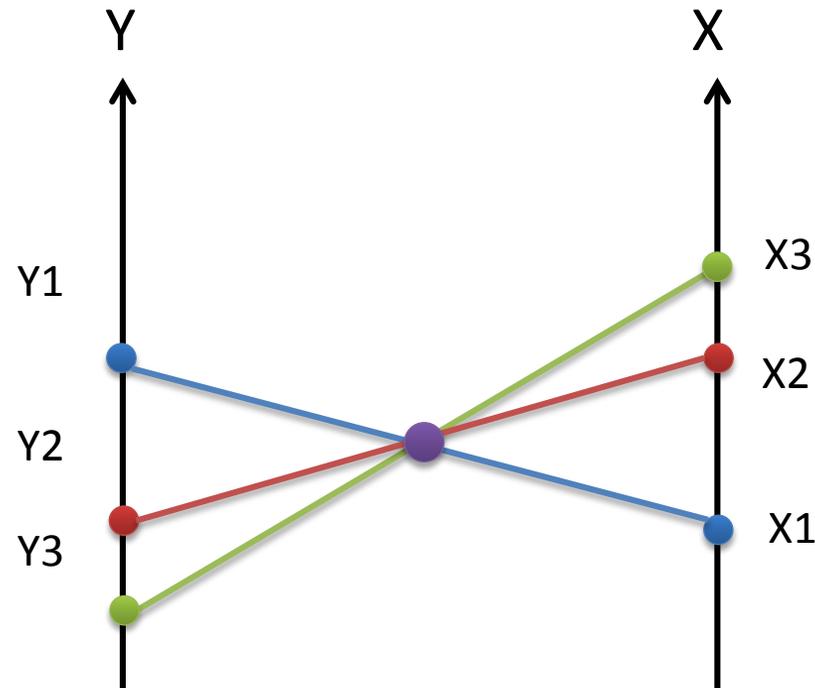
- Exploratory analysis via interactive exploration
 - Linked views
 - User guided exploration

Parallel Coordinate Plots Basics

Scatterplot

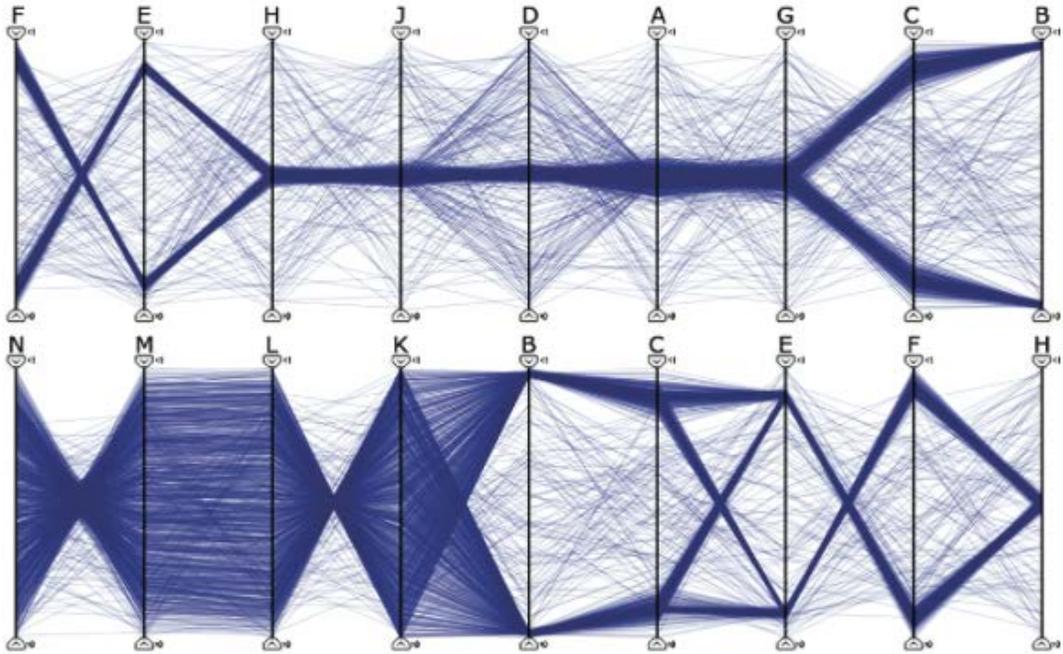


Parallel Coordinates

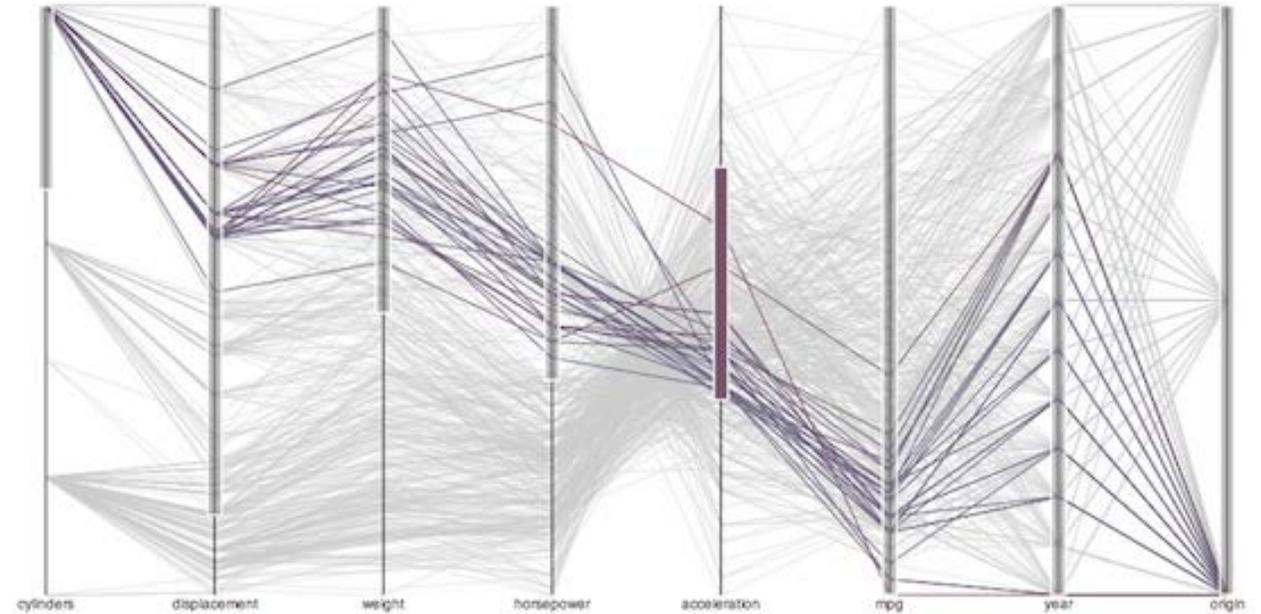


Parallel Coordinate Plots Interactions

Re-arrange Axis

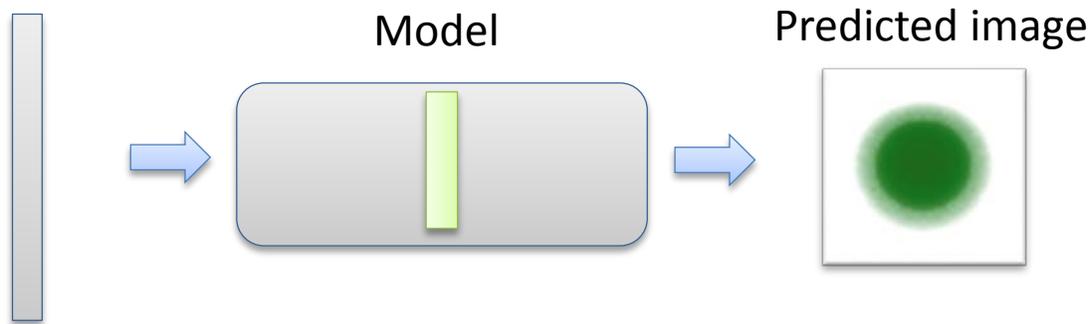


Brushing & Linking



Case Study: Examine the Latent Space of the Physical Simulation

- We have built a deep neural network for predicting physical simulation outputs
 - For an analytical model of the inertia confinement fusion (ICF)
 - Predicting images and scalars (in this case we focus on the images)



- Analysis Goals:
 - Assess the difference choices of latent space dimensions (10D vs 16D)
 - Examine what are the image patterns captured by the latent representation

Interactive Visualization Environment for Latent Space Exploration

Parallel Coordinate

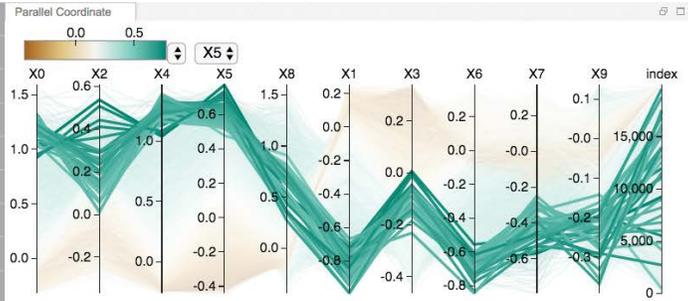
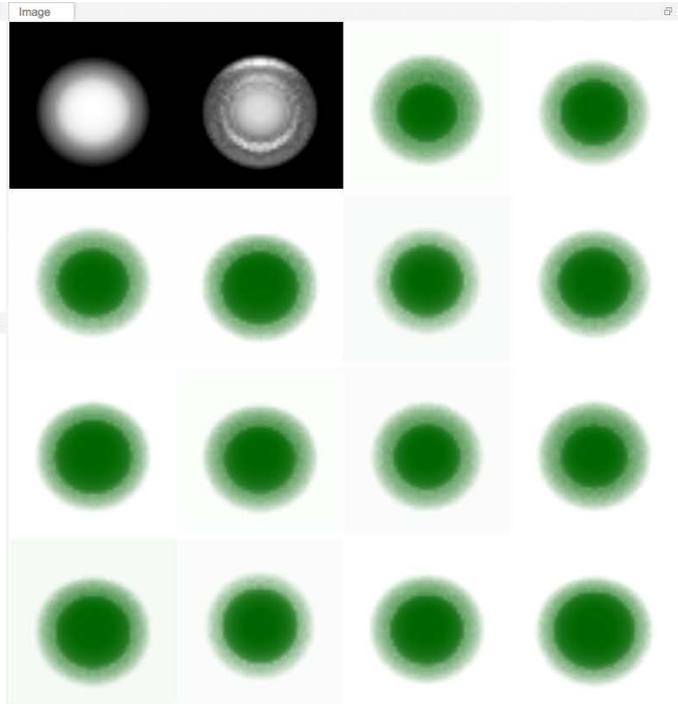


Image Display



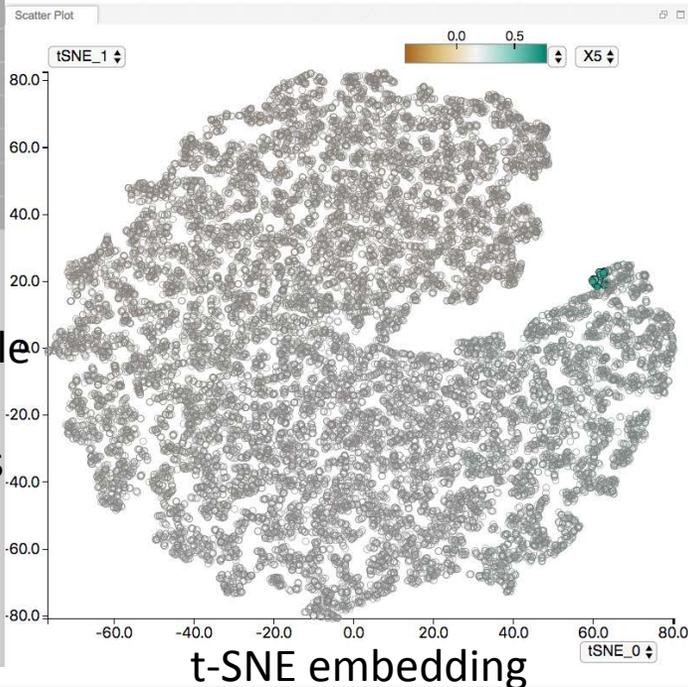
- Part of our high-dimension data visualization tool NDDAV
- Directly interfacing with existing python code

```

from nddavPackage import *
import numpy as np

##### load model sample #####
params = np.load("./data/latent/params.npy").astype("f4")
hidden_h1 = np.load("./data/latent/hidden_h1.npy").astype("f4")
output_imgs = np.load("./data/latent/output_imgs.npy").astype("f4")

##### create vis #####
layout = {
    "row": [
        {"column": ["Parallel Coordinate", 'Scatter Plot']},
        {"column": ["Image", "DimReduction"]}
    ]
}
vis = nddav(layout)
data = vis.addModule(DataHook)
data.setDataFromArray(hidden_h1[0:5000, :])
data.setImages(output_imgs[0:5000, :])
vis.show()
    
```



DimReduction

method:

neighbor:

Dimension reduction control

Filtering

Clustering

DimReduction

Neighborhood

Scatter Plot

Parallel Coordinate

Table

Topological Spine

Topological Multi-Spine

FuncDistance

OptimalAxisAlign

DynamicProjection

PeakShape

PlotPeeling

Image

ViewGraph

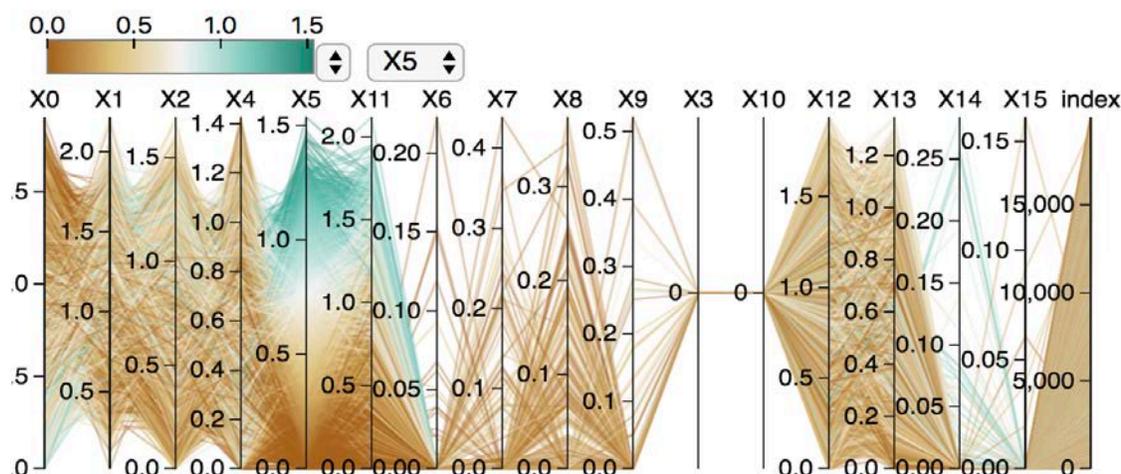
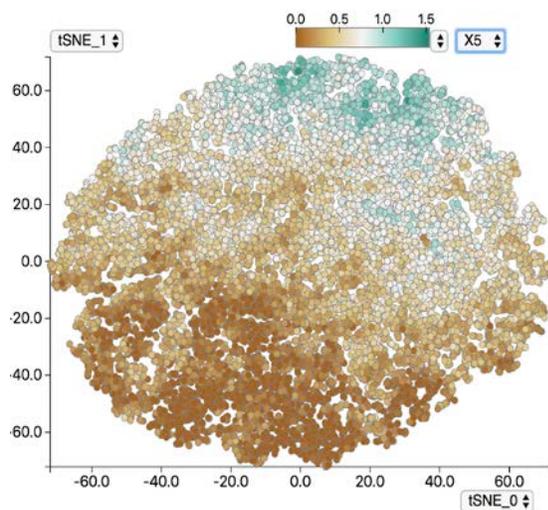
↑

List of Available Visualization Components

Latent Space Comparison

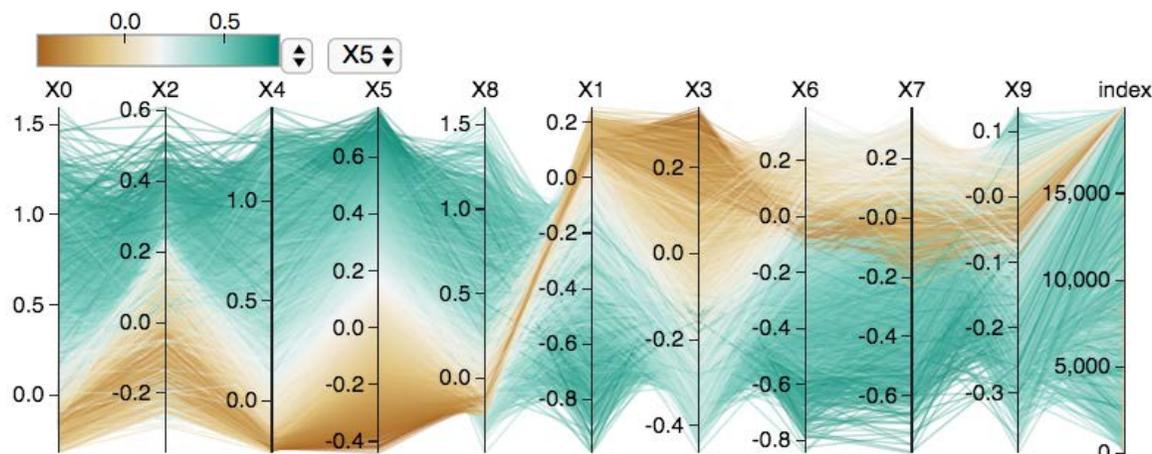
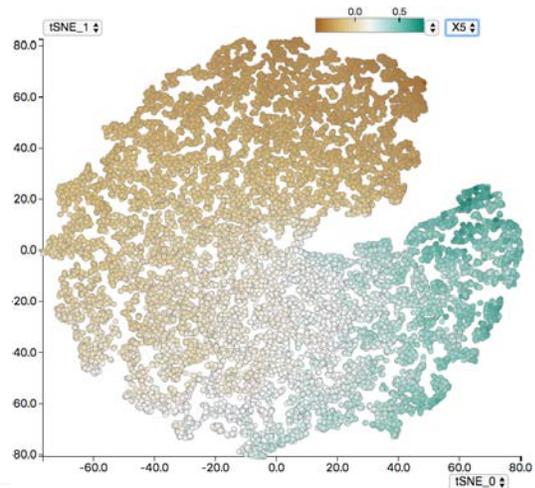
models trained with different bottleneck layer sizes

Old 16D Latent Space



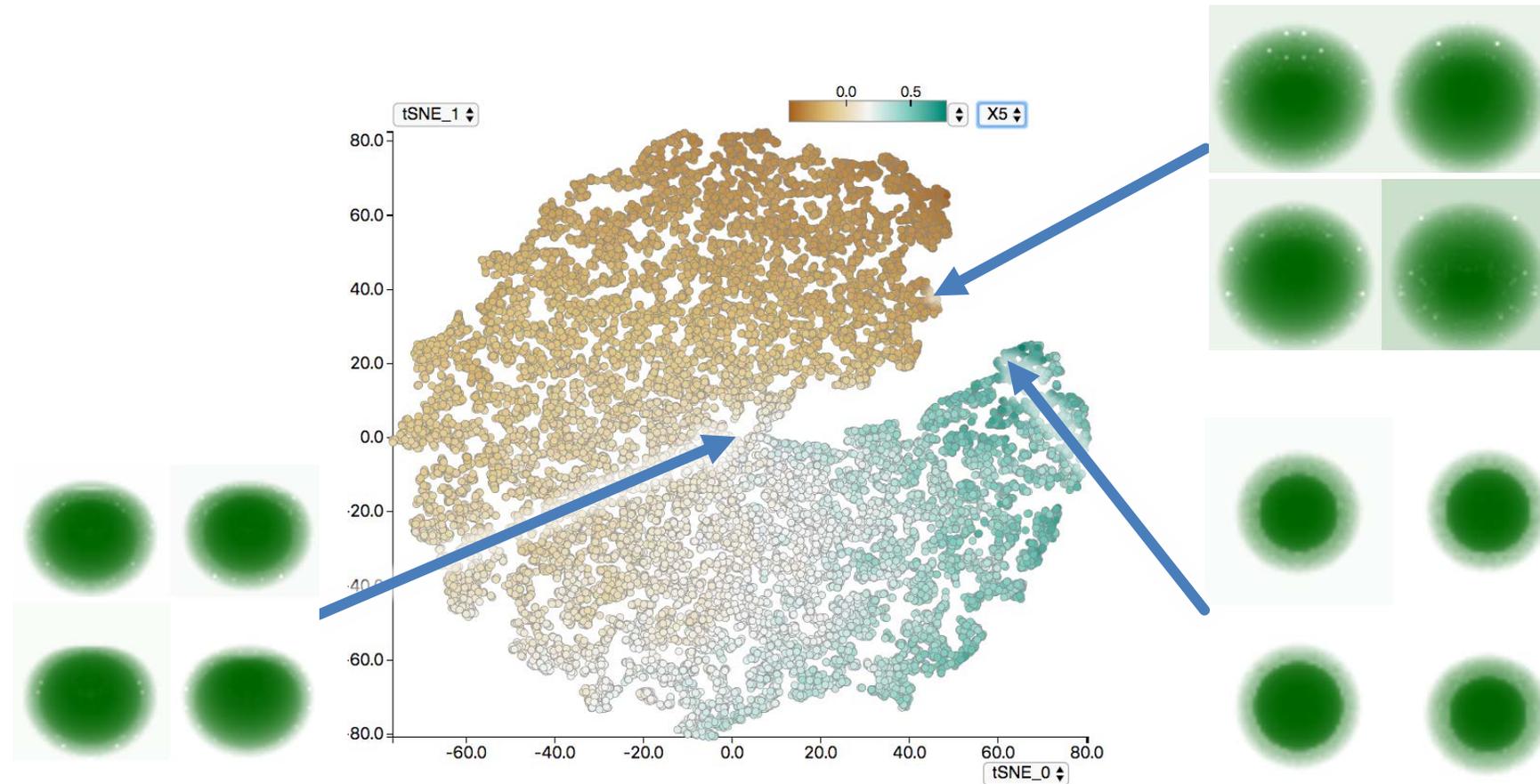
Examine latent dimensions and their relationships

New 10D Latent Space



Explore Distinct Image Patterns in the Latent Space

What image variations are captured in the latent space?





Conclusion & Future Directions

- There is an increasing demand for making sense of latent spaces
- Interactive visualization is a useful tool for understanding latent space
 - Can potentially play an important role in the model design process
- Interpreting the latent space vs. making the latent space more interpretable
 - Visualize how physically meaningful properties distributed in the high-dimensional latent space
 - Adding interpretability constrain in the latent space construction process (i.e., forcing the dimension to be meaningful)



CASC

Center for Applied
Scientific Computing



Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.