Pacific Northwest
NATIONAL LABORATORY

Bundle Networks

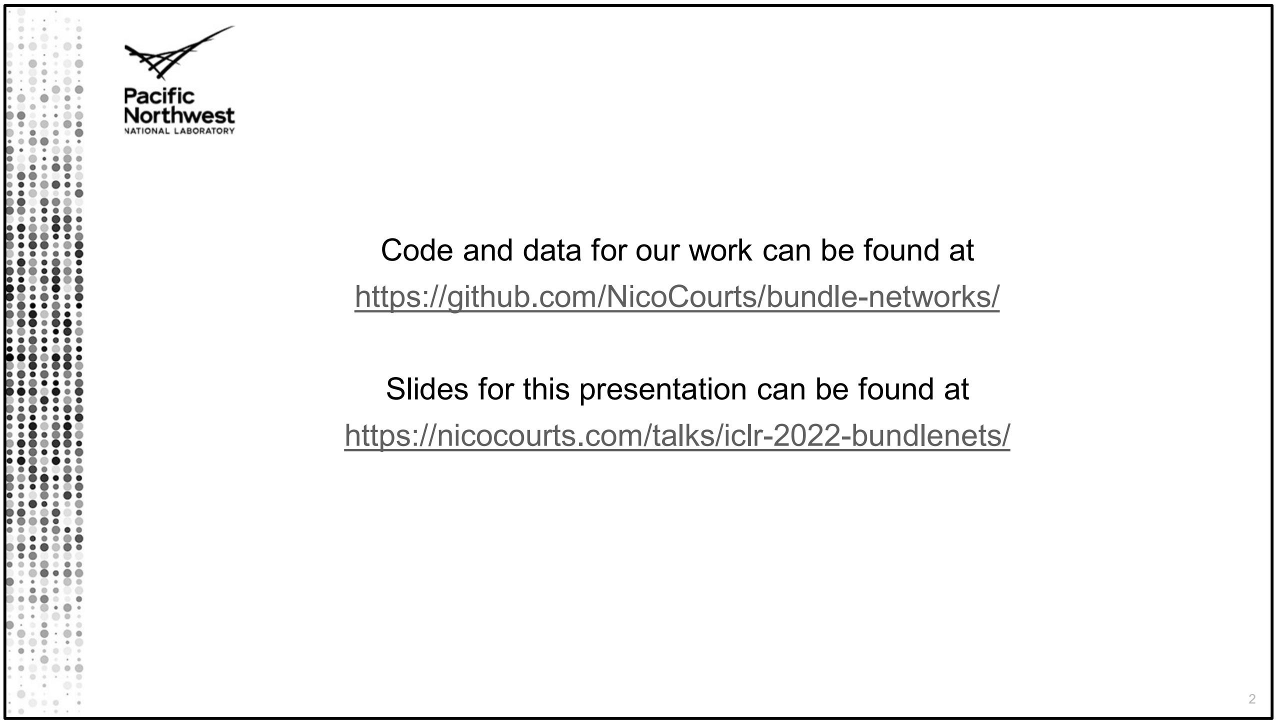
Fiber Bundles, Local Trivializations, and a Generative Approach to Exploring Many-to-one Maps

Nico Courts
University of Washington
Pacific Northwest National Laboratory
nico@nicocourts.com

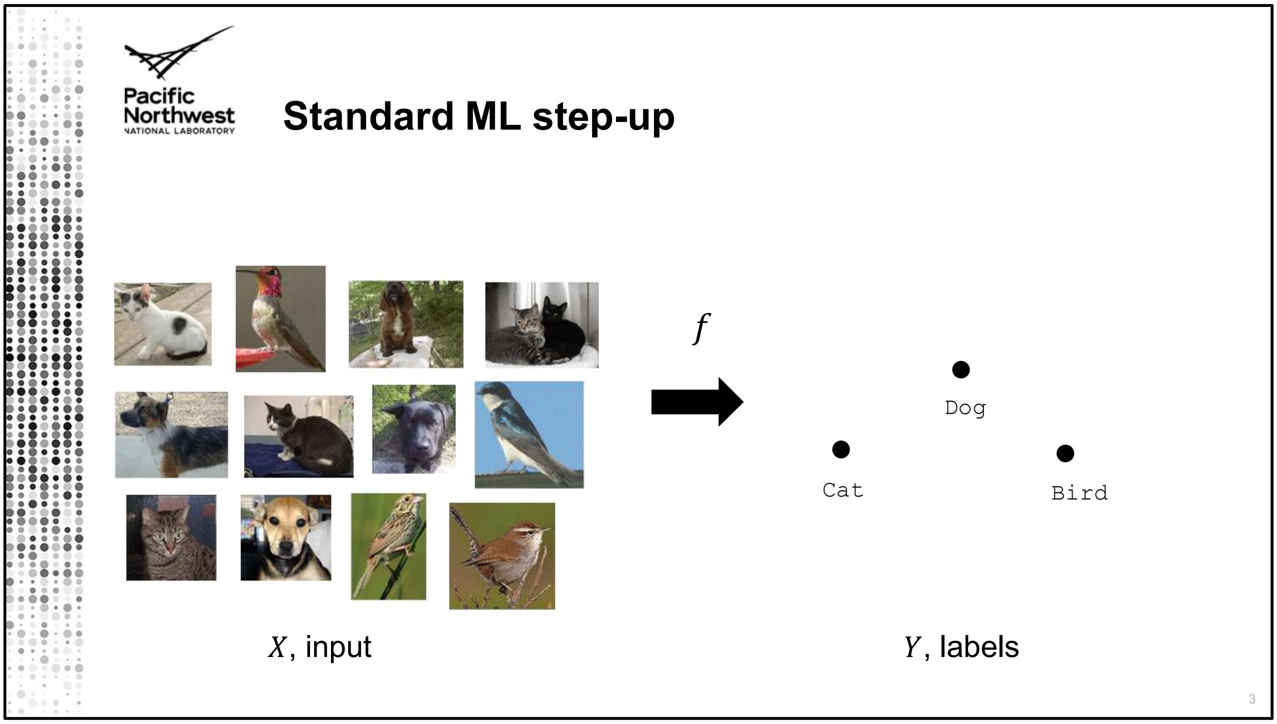
Henry Kvinge
Pacific Northwest National Laboratory
University of Washington
henry.kvinge@pnnl.gov

U.S. DEPARTMENT OF ENERGY BATTELLE
PNNL is operated by Battelle for the U.S. Department of Energy

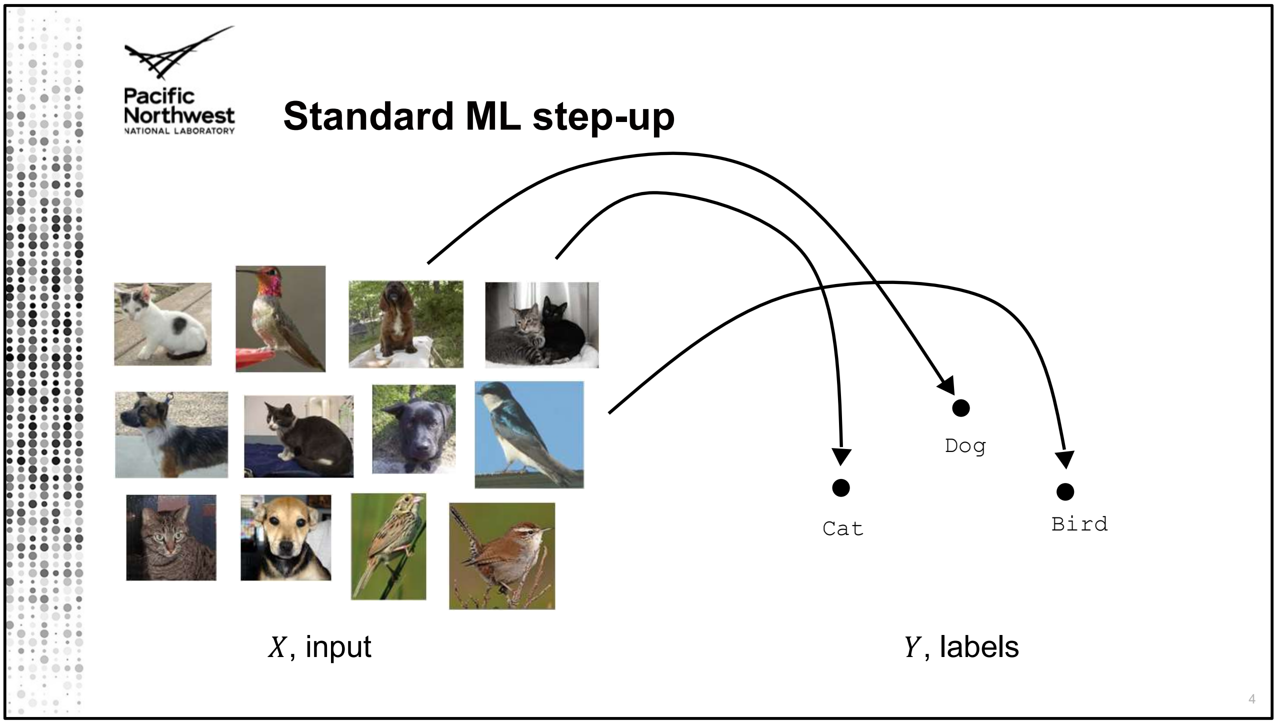
Hello and welcome to our talk today on our paper -- Bundle Networks: Fiber Bundles, Local Trivializations, and a Generative Approach to Exploring Many-to-one Maps. My name is Nico Courts and my co-author on this work was Henry Kvinge.



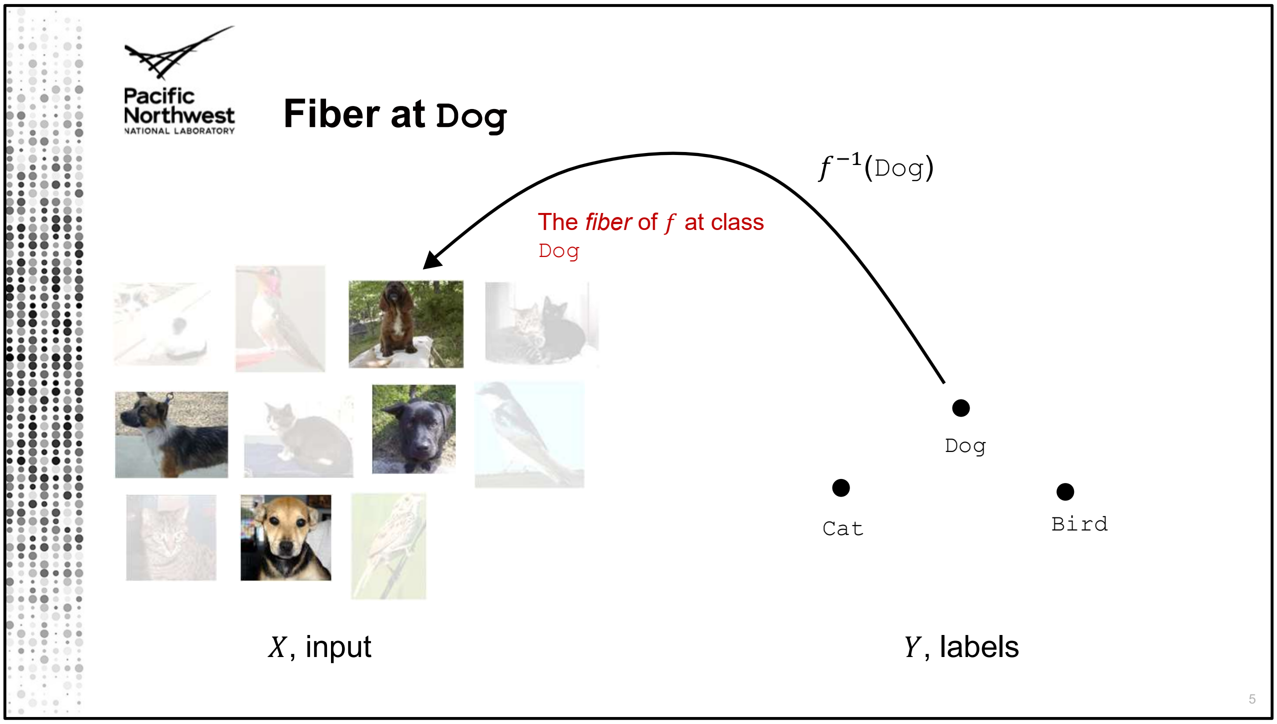
The code and data that we used in our work can be found at the GitHub link on the screen and the slides for the presentation can be found at the URL below.



A standard machine learning task is to assign labels to a group of images. For instance: if we had the classes cat, dog, and bird, our task could be to assign to each image the correct class.



A correctly trained model will learn the “forward direction,” that is, it will learn to associate to each image the correct label or class.



In any task where multiple images are assigned the same label, however, it is possible to consider the "reverse direction:" the fiber over a particular label. Recall that the fiber of a map is the set of all points which are mapped to a particular label. In this case, we have highlighted the fiber over the label "dog."

Pacific Northwest NATIONAL LABORATORY

A science application

- With the optimal processing parameters (e.g., processing temperature and pressure) advanced manufacturing can produce materials with extraordinary properties.
- However, materials scientists want to know more than a single set of processing parameters that will result in a desired material property, they want to understand the distribution of ALL processing parameters that will result in the same property.
- Then they can choose those parameters that are physically feasible, resource and energy efficient, etc.

6

In a more concrete example, we can consider the process for synthesizing a material in a lab. With the optimal choice of processing parameters, advanced manufacturing can produce materials with extraordinary properties. However, researchers are often interested in understanding the collection of all processing parameters that will produce a product with a particular property. With this knowledge, they will be empowered to choose the parameters that, for instance, are physically feasible or make efficient use of resources and energy.

Underlying question

Both of these ML tasks are **many-to-one**.

Question: Given a mapping $f: X \rightarrow Y$ and a distribution D on X, can we build a model that will learn the distribution on each fiber $f^{-1}(y)$?

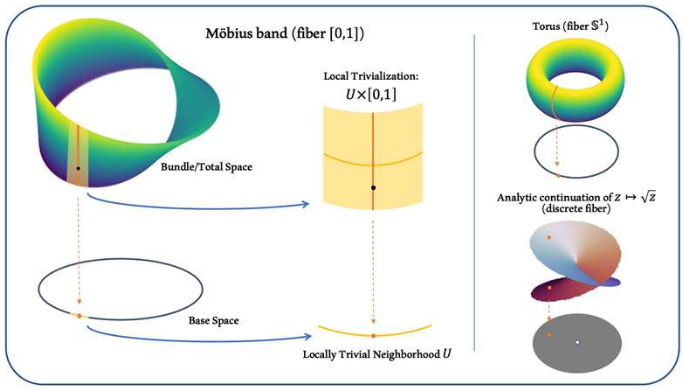
If so, we can succeed at creating a **fiberwise-accurate reconstruction** of this task.

7

Both of the examples discussed above naturally form many-to-one tasks. More generally, we can ask whether, given mapping f from X to Y and a distribution D on X, we can build a model that will learn the distribution on each fiber $f^{-1}(y)$. If so, we can leverage this to create a fiberwise-accurate reconstruction of this data.

Pacific Northwest NATIONAL LABORATORY

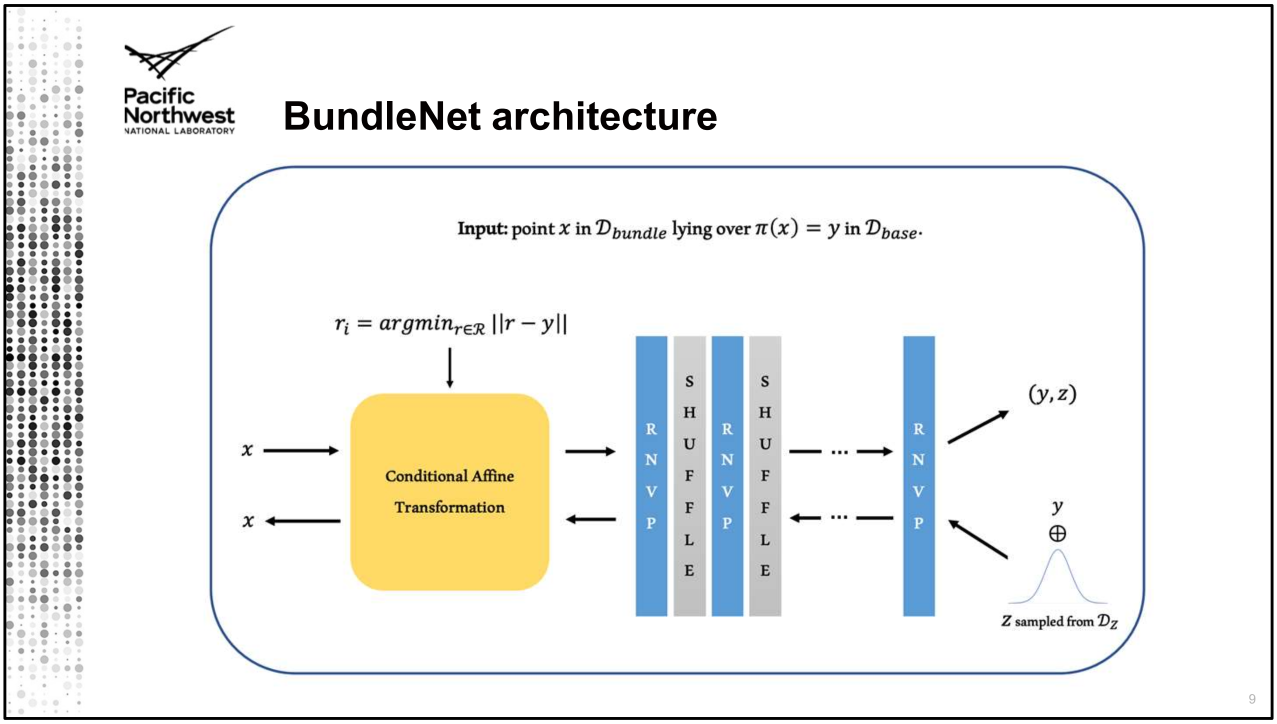
A Mathematical Construction: Fiber Bundles



A *fiber bundle* is a base space X, a total space E, and a surjection $\pi: E \to X$, satisfying *local triviality*. This means that E is a space parametrized by X and locally every neighborhood of E looks like a product with a fiber Z.

8

A natural candidate among mathematical concepts is that of a fiber bundle, the intuition for which is the following: It is a pair of spaces consisting of a base space X and a total space E along with a surjection from E to X satisfying a property called local triviality. That means that E is a space that is parameterized by X and, furthermore, that E admits a covering by neighborhoods where each looks like a product with some fiber space Z.



In our model, we fix an open cover for our base space to begin with and choose a representative point for each neighborhood. We then use these representative points to condition the model on the neighborhood to which a data point belongs. In the overview for our model architecture currently shown on the screen, the conditional affine component is the only part that's conditioned on our actual neighborhood representative, while the rest of the layers are all left unaffected by the conditioning. The idea being that the conditional segment learns to map each neighborhood to a common space that is used globally by the remaining layers.

Results

Table 1: The Wasserstein-1 metric for global data generation across all four datasets and models. Each metric is applied to a trained model as detailed in section 5.1 with 95% confidence intervals.

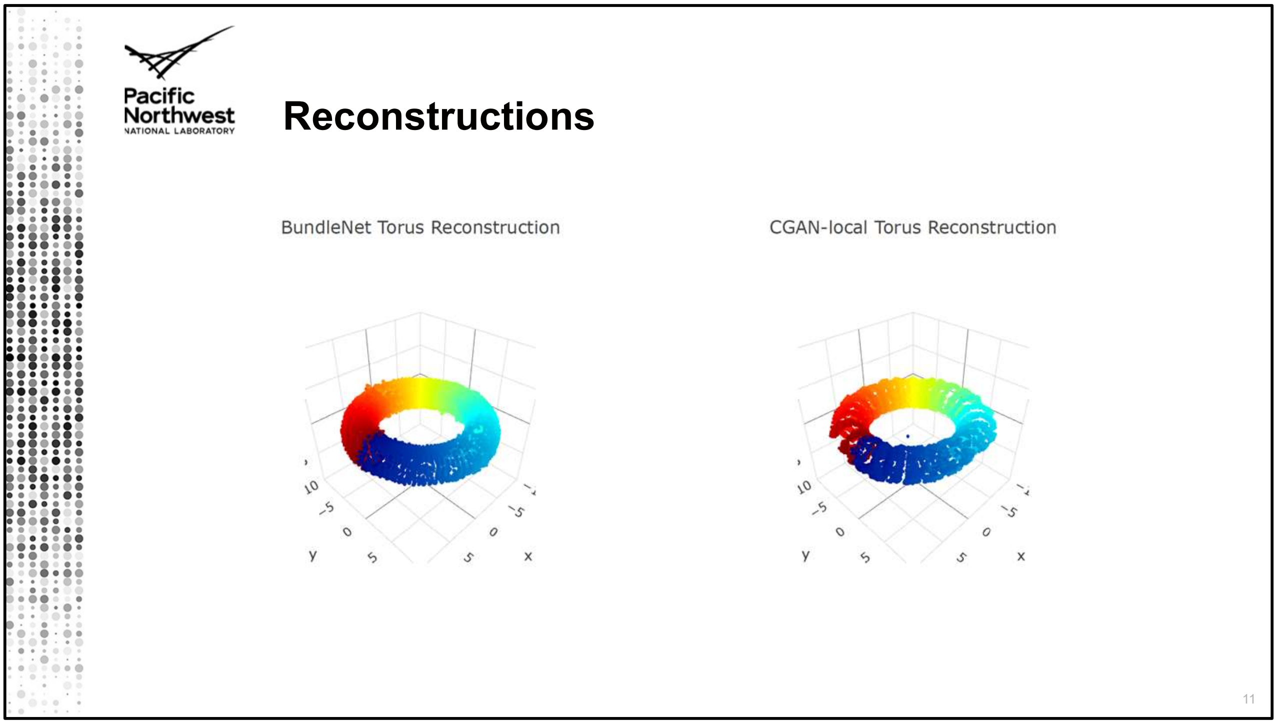
	BundleNet (ours)	WGAN	CGAN	CGAN-local
Torus	$\mathbf{0.461 \pm 0.020}$	2.191 ± 0.011	7.114 ± 0.015	1.610 ± 0.031
Möbius band	$\mathbf{0.264 \pm 0.004}$	2.048 ± 0.013	1.844 ± 0.015	6.497 ± 0.101
Wine Quality	$\mathbf{1.733 \pm 0.011}$	99.22 ± 0.392	4.054 ± 0.014	2.114 ± 0.011
Airfoil Noise	$\mathbf{1.448 \pm 0.022}$	4.741 ± 0.034	2.508 ± 0.025	1.735 ± 0.013

Table 2: The Wasserstein-1 metric for fiberwise data generation across all four datasets and the three models that have a conditional component. Intervals represent 95% CIs over 10 trials.

	BundleNet (ours)	CGAN	CGAN-local
Torus	$\mathbf{0.251 \pm 0.013}$	7.996 ± 0.015	0.450 ± 0.018
Möbius band	$\mathbf{0.279 \pm 0.011}$	9.860 ± 0.038	8.228 ± 0.904
Wine Quality	$\mathbf{1.917 \pm 0.172}$	3.666 ± 0.233	2.926 ± 0.169
Airfoil Noise	$\mathbf{3.124 \pm 0.089}$	3.563 ± 0.158	$\mathbf{3.076 \pm 0.063}$

- **Models**: WGAN – Wasserstein GAN, CGAN – Conditional GAN, CGAN local – CGAN with distinct conditioning data for distinct neighborhoods.
- **Datasets**: Torus – Torus bundle (synthetic), Mobius band – Mobius bundle (synthetic), Wine Quality, Airfoil Noise – UCI database

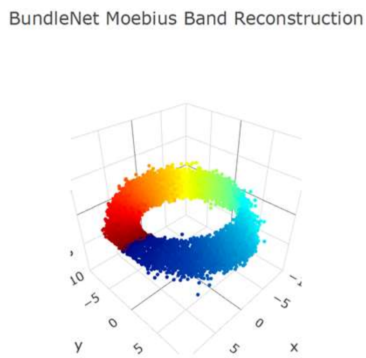
We tested several models against our model, all of which were based on the GAN architecture. One was a Wasserstein GAN, another a conditional GAN, and the third was an iteration on conditional GAN that used finitely-many neighborhood representatives and used a very similar training schema to the one that we use in BundleNets. Across the board, on our two synthetic data sets and two real-world data sets, we found that our model outperformed all the other models that we considered. Further metrics were examined and bore similar results and can be found in our paper.

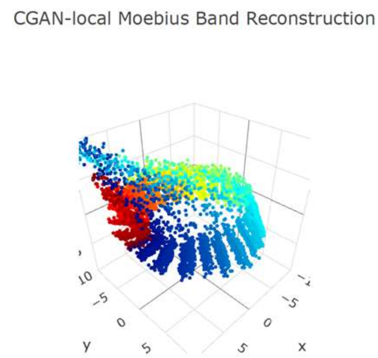


To get a holistic view of the performance differences between our model and the other models tested, we look at how each does at reconstructing the true distribution of points on our synthetic datasets after training. We find that our model creates a very convincing global distribution as well as highly accurate fibers on the torus example while the next-best model, CGAN-local, suffers collapse within each neighborhood as well as having incomplete fibers.

Pacific Northwest
NATIONAL LABORATORY

Reconstructions





12

We see similar results in reconstructions of the Moebius band, where BundleNets creates a convincing reconstruction, but where CGAN-local again suffers from intra-neighborhood collapse as well as an additional failure mode.

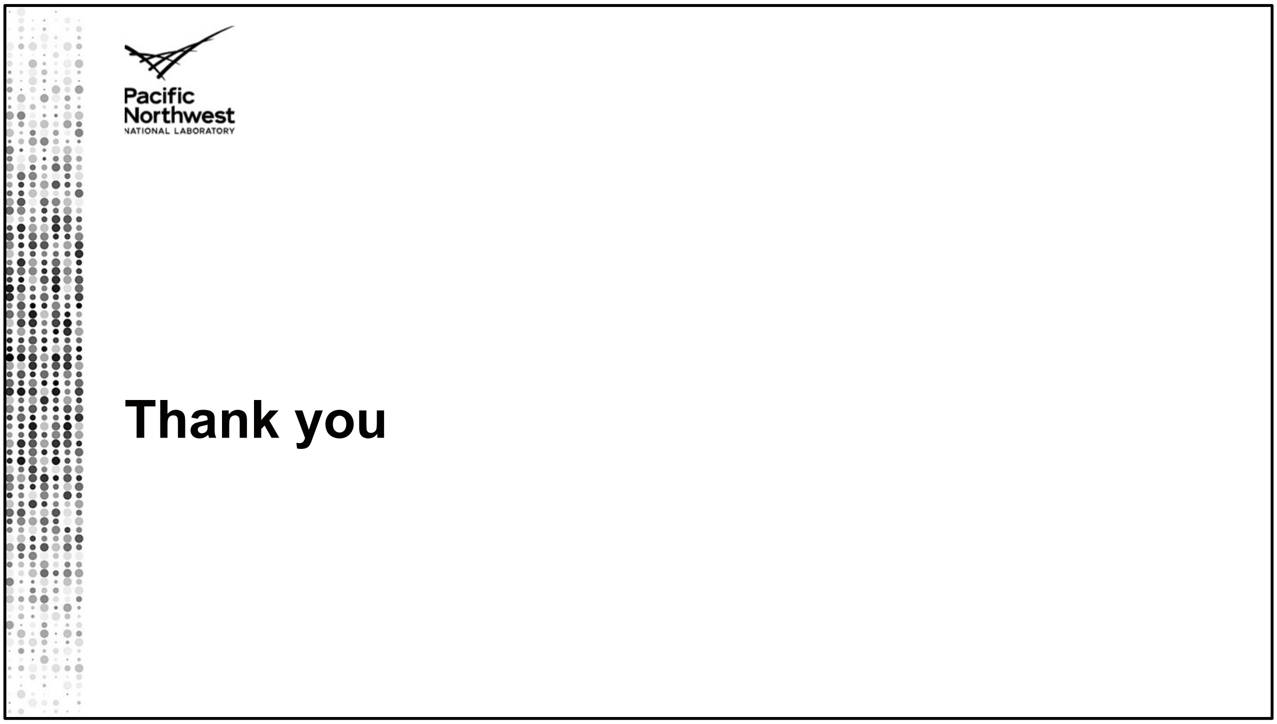
Pacific Northwest NATIONAL LABORATORY

Further questions studied

- What should be our prior for the distribution over a fiber?
- How many neighborhoods are necessary for good performance?
- Is BundleNet effective when different fibers $f^{-1}(y_1)$ and $f^{-1}(y_2)$ have different topological structure across a dataset?

13

In our paper, we consider several other questions including
•What should be our prior for the distribution over a fiber?
•How many neighborhoods are necessary for good performance?
•Is BundleNet effective when different fibers have different topological structures across a dataset?



Thank you for your attention and please refer to our paper for further discussion.