Learning Visual Representation with Homological Labels

ICIAM2023, Waseda, Tokyo, 21 Aug. 2023

Shizuo KAJI (Inst. of Maths-for-Industry, Kyushu Univ.)
Joint with Yohsuke Watanabe (ZOZO inc.)

codes and preprint
https://github.com/shizuo-kaji/PretrainCNNwithNoData

Deep Learning and Topological Data Analysis

Human is good at

- Rough estimation
- Panoramic view
- Discovering rules/**invariance** from a small number of examples
- **Explaining** the reason

Deep Learning(DL) is good at

- Precise observation
- Memorising/imitating examples
- Processing huge data
- Accurate operation

Deep Learning (DL)

Data-driven
local

Topological Data Analysis (TDA)

Maths-based
global

complementary

Background

- DL achieves high performance but has some weakness
- TDA has been proven effective in capturing data features that conventional techniques have missed

Biases in Deep Learning

Algorithmic biases

- image models are locally minded

Convolutional Neural Networks are shortsighted

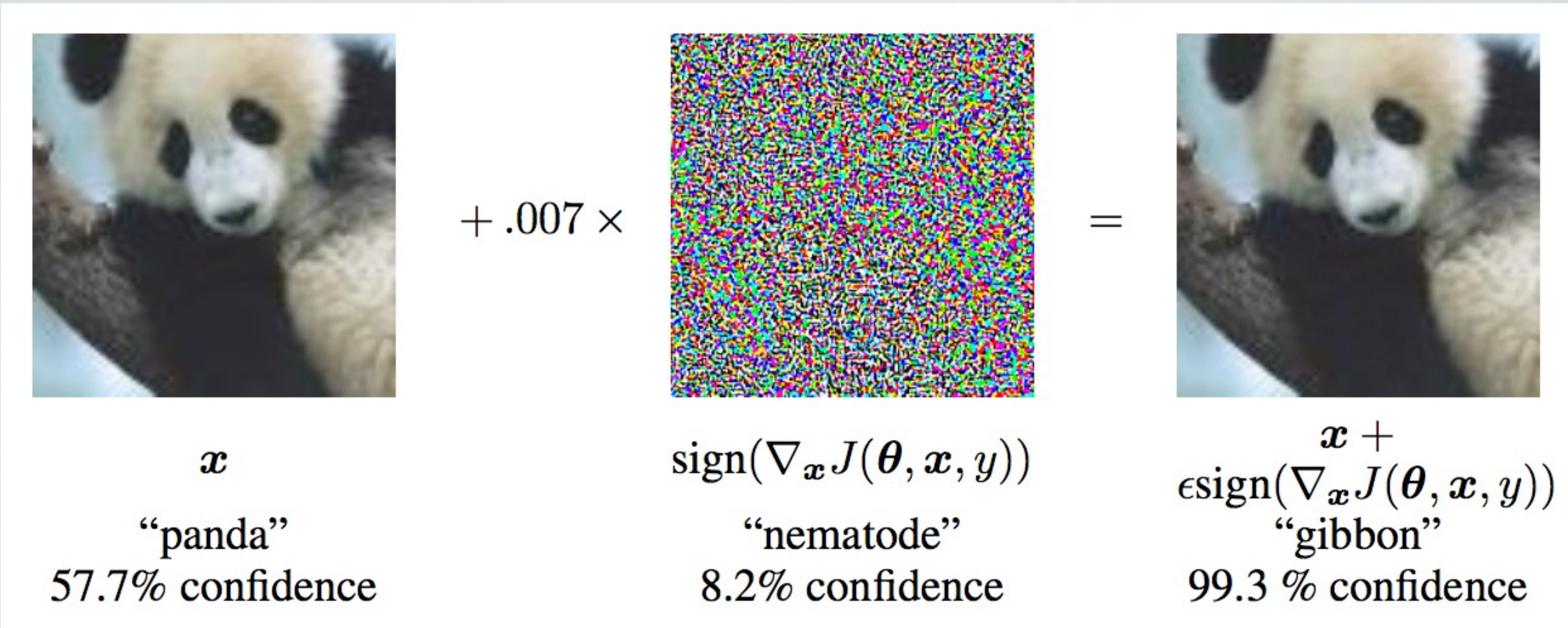
(a) Texture image
81.4% **Indian elephant**
10.3% indri
8.2% black swan

(b) Content image
71.1% **tabby cat**
17.3% grey fox
3.3% Siamese cat

(c) Texture-shape cue conflict
63.9% **Indian elephant**
26.4% indri
9.6% black swan

ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness, Geirhos et al. 2019

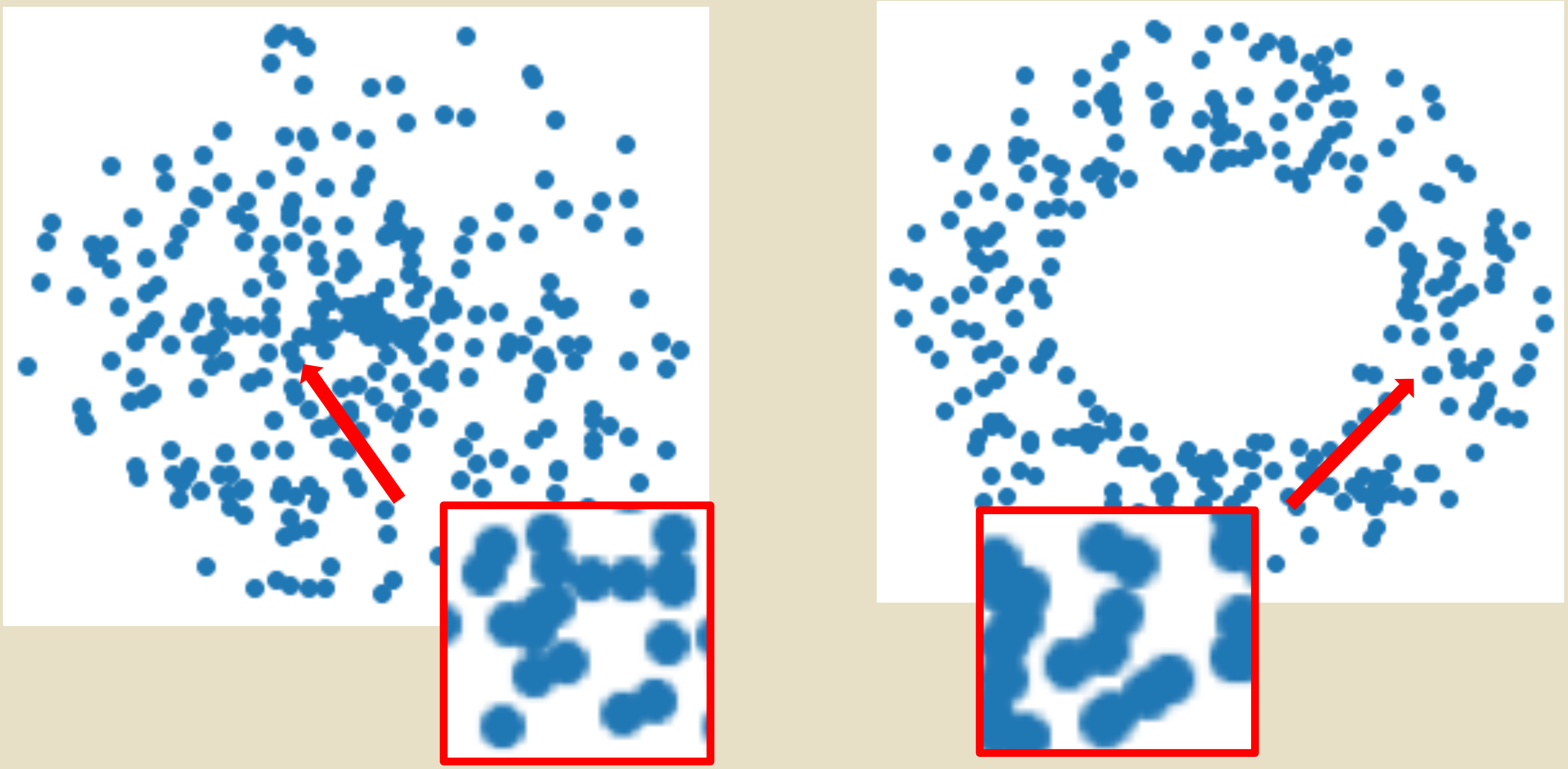
CNNs are easily deceived



Explaining and Harnessing Adversarial Examples
Goodfellow et al. 2014

CNNs are too sensitive to local information

Convolution is a *local* operation



They look similar locally,

but we see a clear difference if we zoom out

c.f. Manifolds are locally all Euclidean and homology distinguishes the global topology of them.

Biases in Deep Learning

Data biases

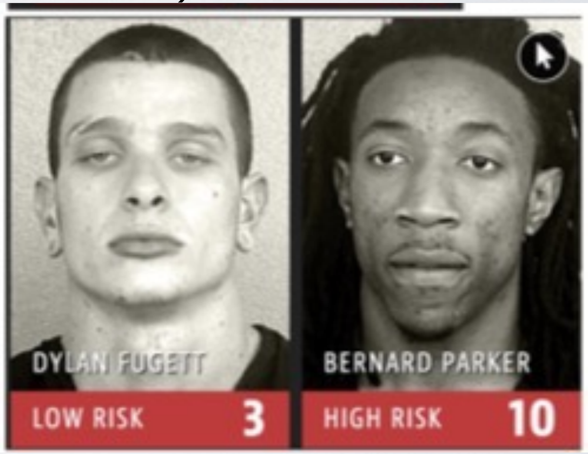
-- not only labels but also images themselves are biased

Concerns with real image

- Huge cost for data collection and annotation
  (ImageNet consists of 14M manually-labelled images)

- Bias in the annotation and images
  (Labels reflect the bias of the labellers.
  The Image distribution itself is also biased.)

Ryan Steed and Aylin Caliskan, 2021
"Image representations learned with unsupervised pre-training contain human-like biases"



Angwin, J., Larson, J., Mattu, S. & Kirchner, L. (2016) "*Machine Bias*"

- Security issues
  (model inversion attack)

- Rights and privacy issues
  (ImageNet use "wild" images on Internet)

Topological Image Analysis

Observe locally, understand globally

Persistent Homology of an image

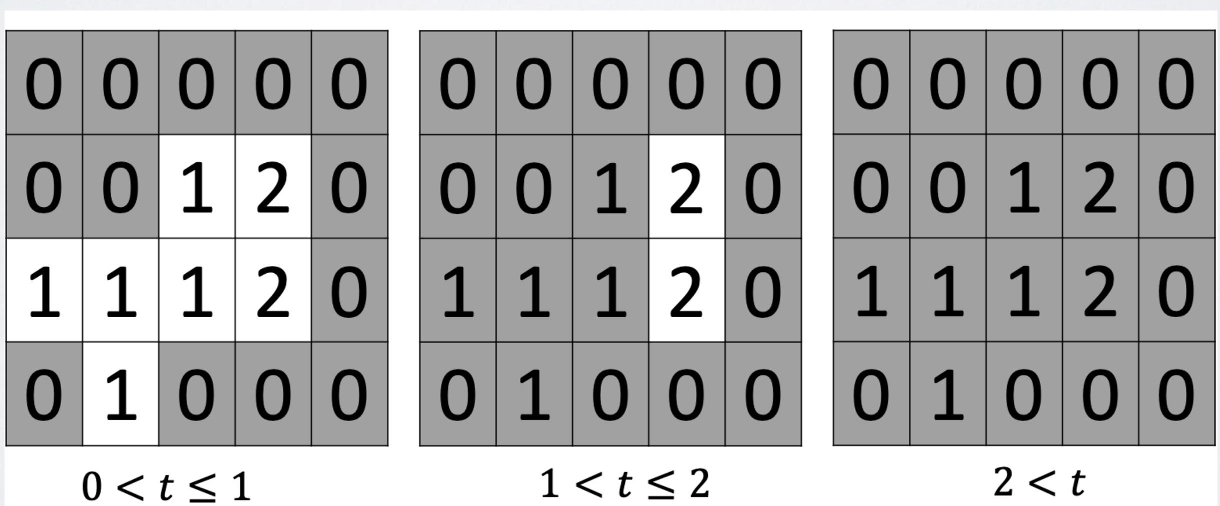
Characteristics of PH

- It captures global topological features
- Proved stability against pixel value change
- Isometry invariance (translation, rotation)

Persistent Homology of an image

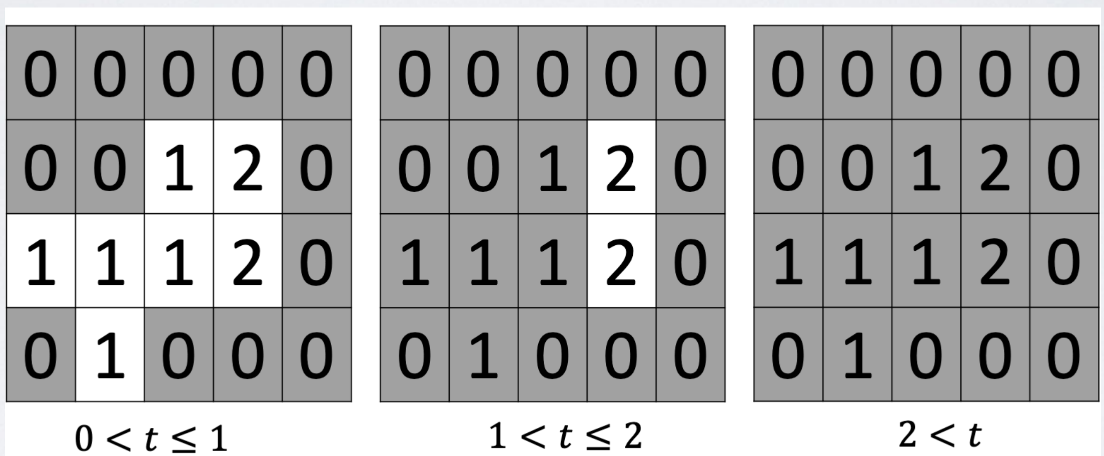
A real-valued function f: X -> R defined over a topological space X (in our case, the square) defines an increase sequence of subspaces

$$X_t := \{x \in X | f(x) < t\}$$



Persistent Homology of an image

PH records the birth and death thresholds of connected components (PH0) and holes (PH1) in the form [birth,death].

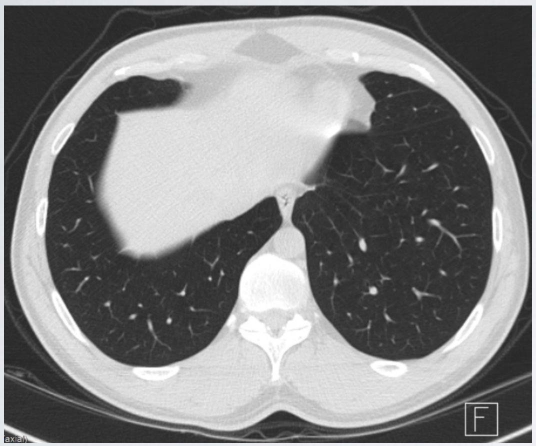


Ex. PH0 = {[0,1],[0,2]}: two connected components born at t=0

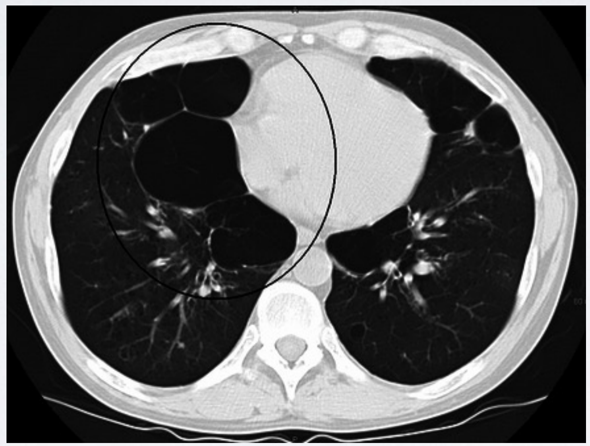
PH1 = {[1,2]}: a hole born at t=1 and filled at t=2.

EX. MEDICAL IMAGE SEGMENTATION

Images from Wikipedia



healthy

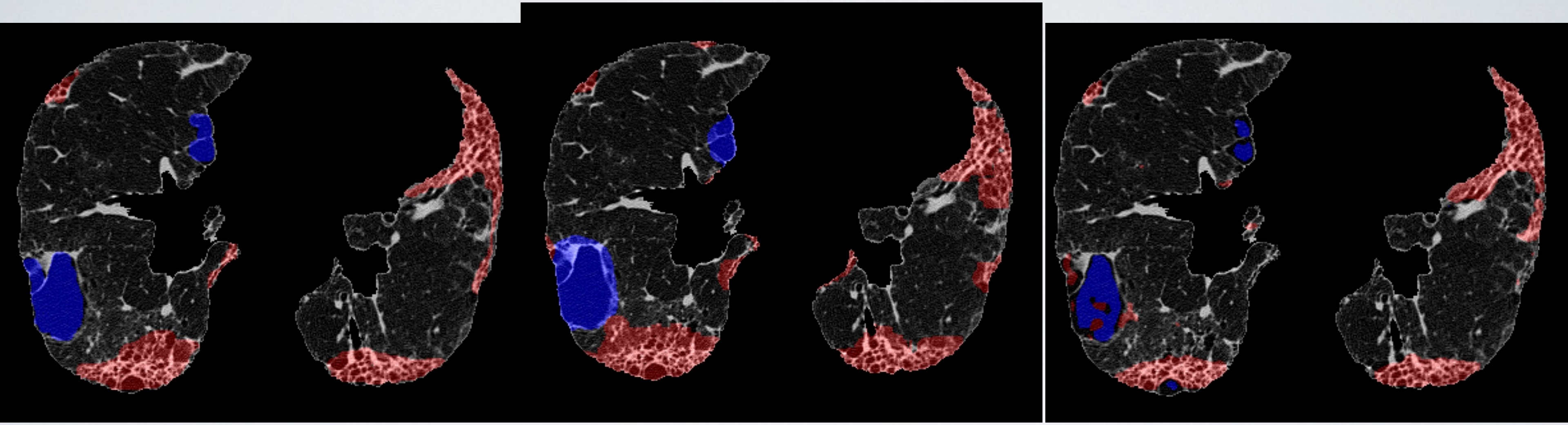


COPD

IPF

COPD: Chronic obstructive pulmonary disease is the third leading cause of death (WHO 2019)
IPF: Idiopathic pulmonary fibrosis is a progressive and irreversible disease

1) Explainable feature (vs blackbox DL)
2) Robust and easily transferable (vs DL needs re-training)
3) 3D analysis (vs conventional 2D slice-based analysis)



Doctor

Persistent Homology

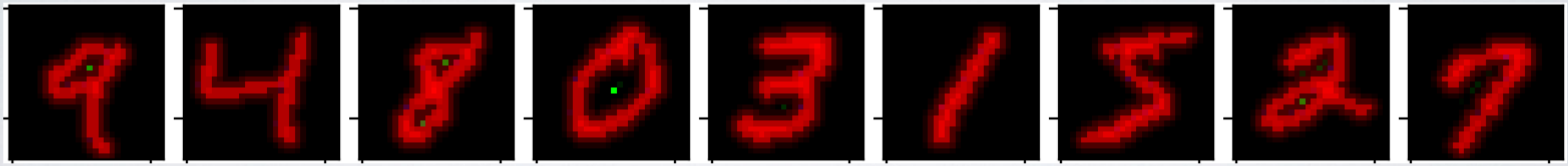
Deep Learning (Unet)

Number of parameters: 5 (PH) vs over 5 million (DL)

Moreover, the 5 parameters have physiological interpretations

With N. Tanabe et al. (Kyoto University Hospital), 2021

EX. IMAGE CLASSIFICATION WITH CNN+PH



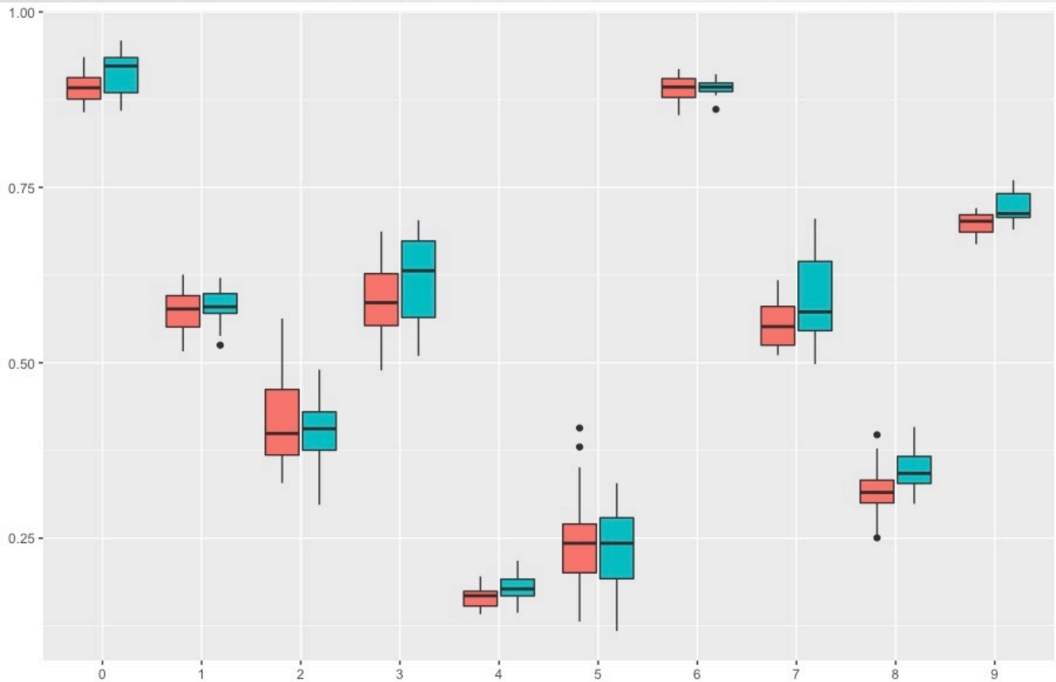
The MNIST Dataset
60k(train)+10k(test) images
10 classes (0,1,…,9)
28x28 black-and-white images
Accuracy of SoTA is over 99.8%

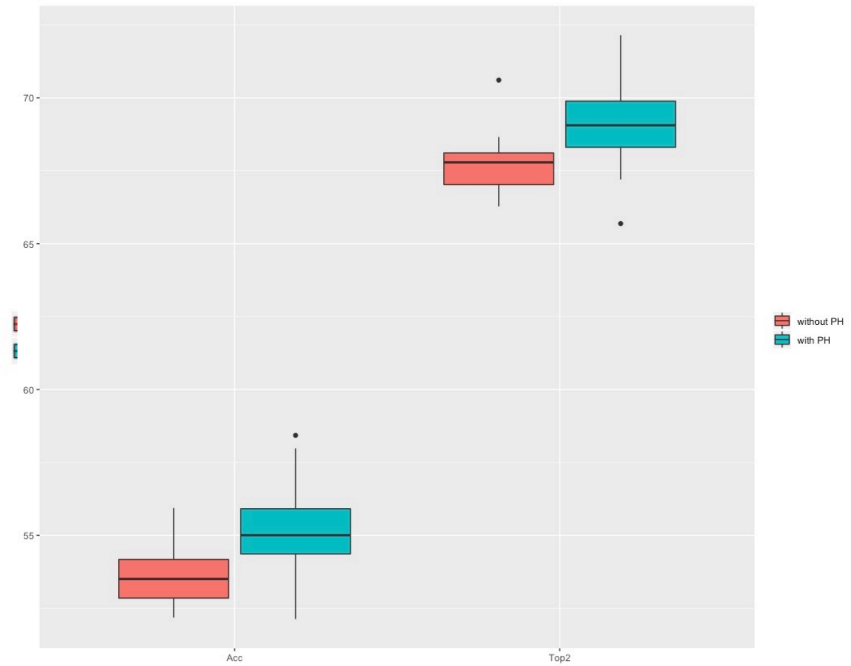
Too easy as a benchmark

Reduced MNIST
Only 10 training images
(one image per class;
one-shot learning)

Difficult!

REDUCED MNIST CLASSIFICATION RESULTS (RED: ORIGINAL BLUE: +HOMOLOGY)





per class accuracy

total accuracy

Adding homological information improves the performance

With Sudo and Ahara, 2020

Teaching Topology to Neural Networks with Persistent Homology

(1) Synthetic image generation

(2) Label generation

Goal

Transfer learning based on pretrained CNNs has some problems

- Huge labelled images are necessary for pretraining (e.g., ImageNet)
- Privacy and bias issues in the training dataset
- The learned model is biased towards texture

Solution: Pretraining with synthetic images with a mathematical task

Training Convolutional Neural Networks without using natural images

- No need for data collection
- No need for manual labelling
- Acquires robust image features based on topology

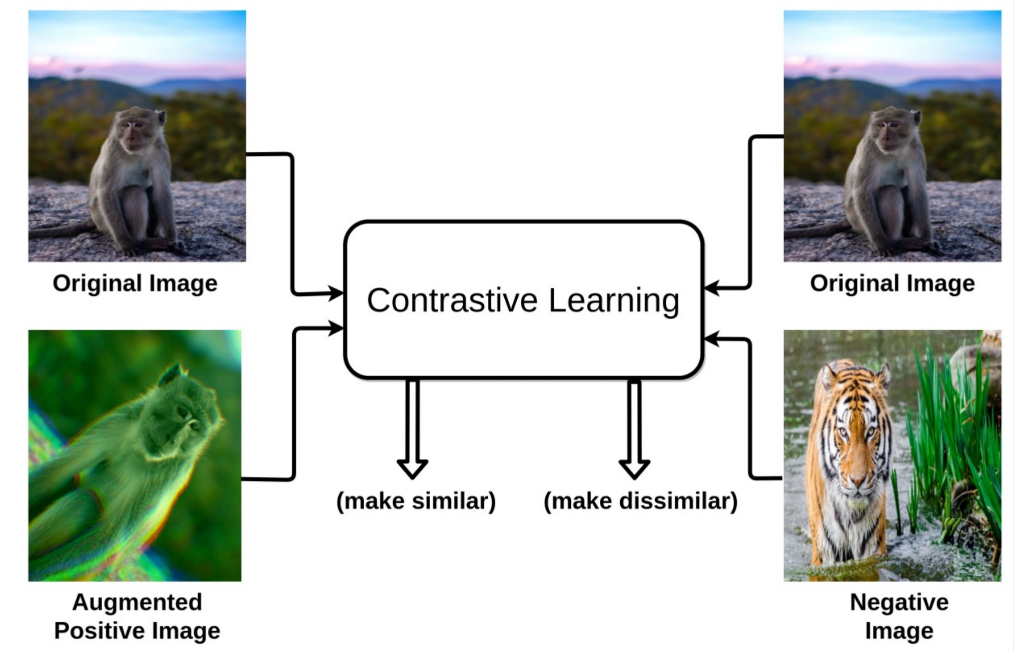
Topology helps to eliminate manual labour and fairness concerns in data preparation!

Transfer learning

Self-supervised learning

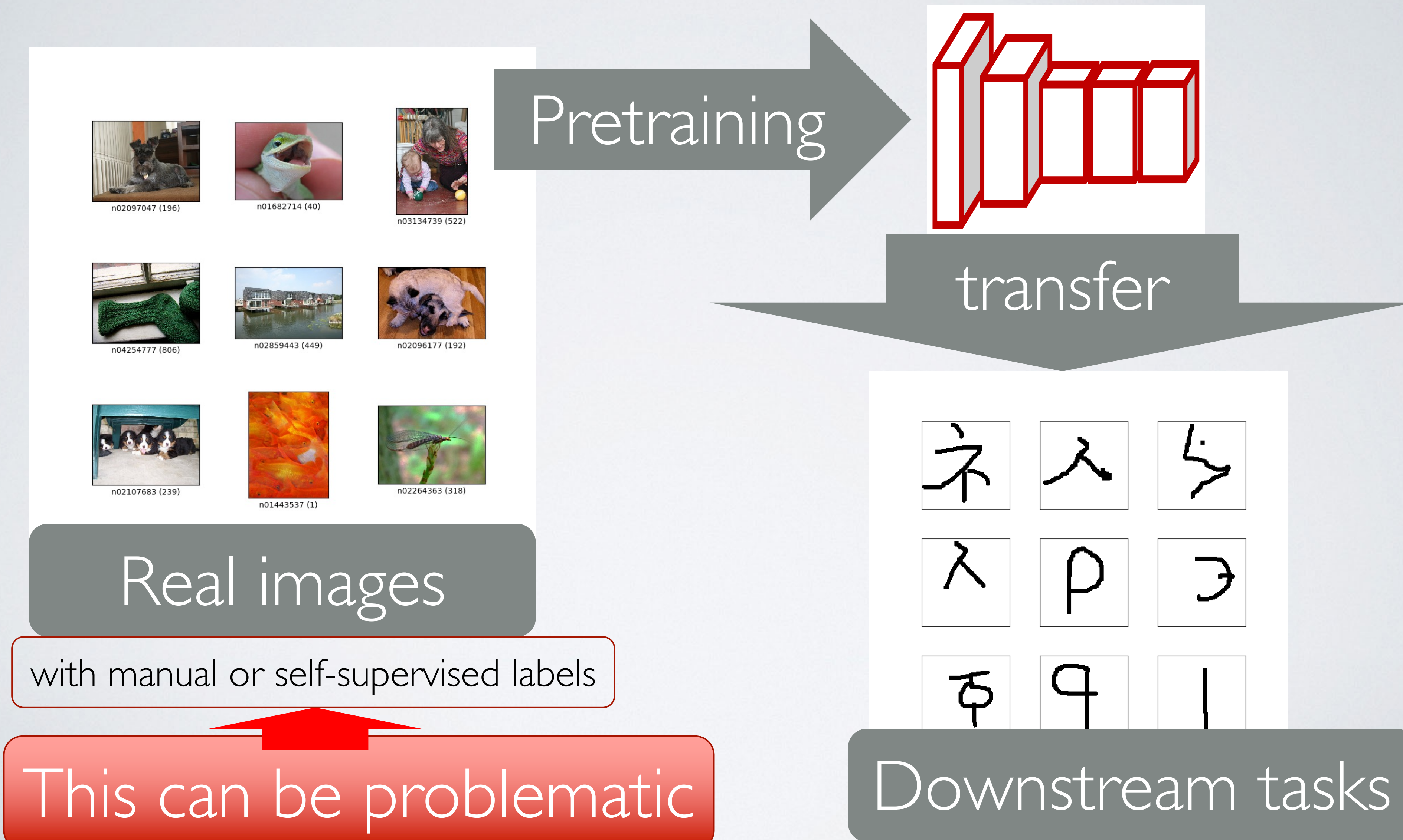
SSL is a method to train models
without manual labels.
SSL has been very successful in NLP.

A popular scheme, contrastive learning, uses the metric in the representation space

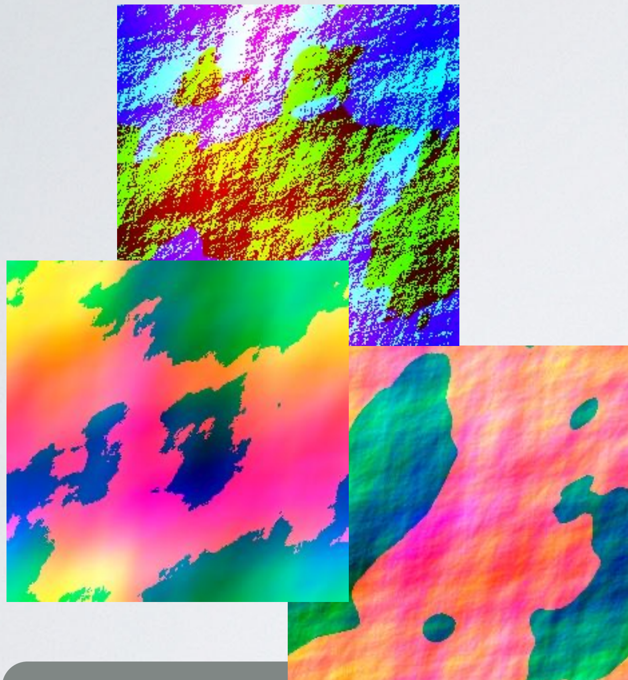


A. Jaiswal et al. 2021

Pretraining with real data



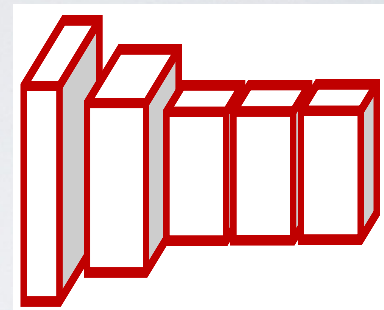
Pretraining with artificial data



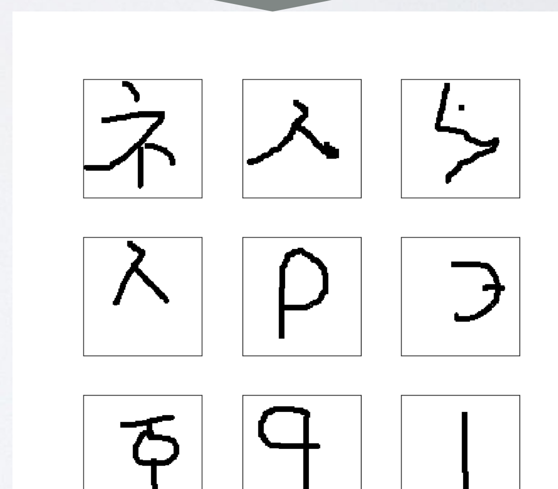
Synthetic images

with mathematically computed labels

pretraining by SSL



transfer



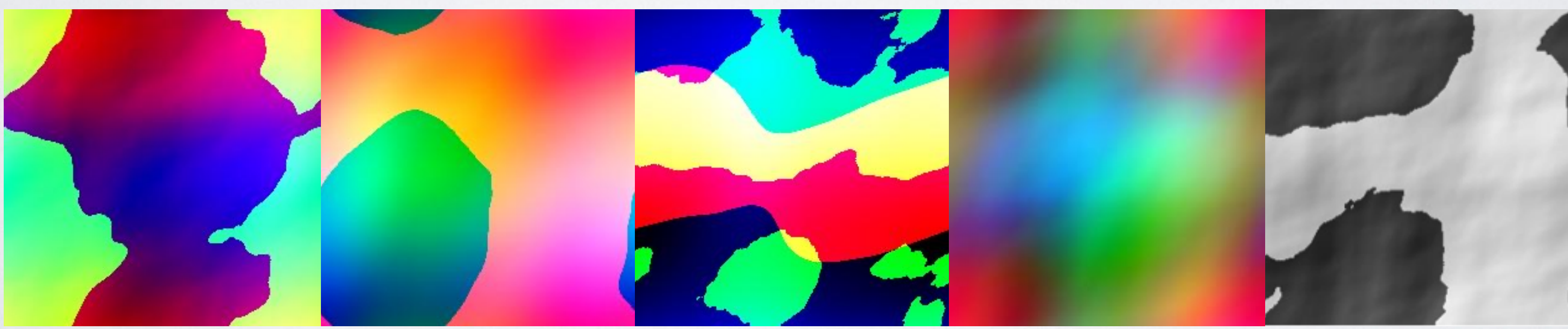
Downstream tasks

Synthetic image generation

Each channel of the image is generated by the following formula where f is an image with uniform random pixel values, and β is uniformly drawn from [1,2].

$$g(x,y) = \mathrm{Re}\left(iFFT \left(\frac{FFT(f)(x,y)}{((x+1)^2 + (y+1)^2)^{\beta}} \right) \right), \tag{1}$$

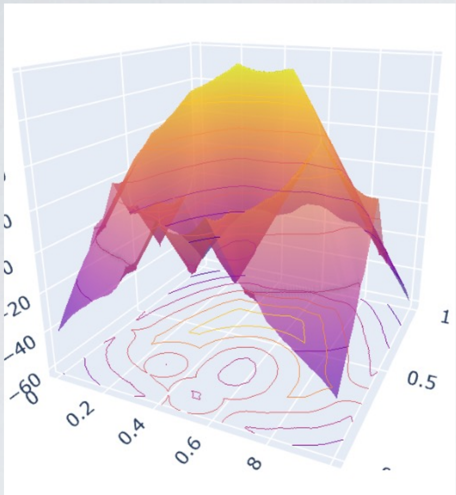
β controls the decay of high-frequency components



Each channel was binarized with a probability of 0.5.

The final image was converted to greyscale with a probability of 0.5.

Mathematical labelling of an image

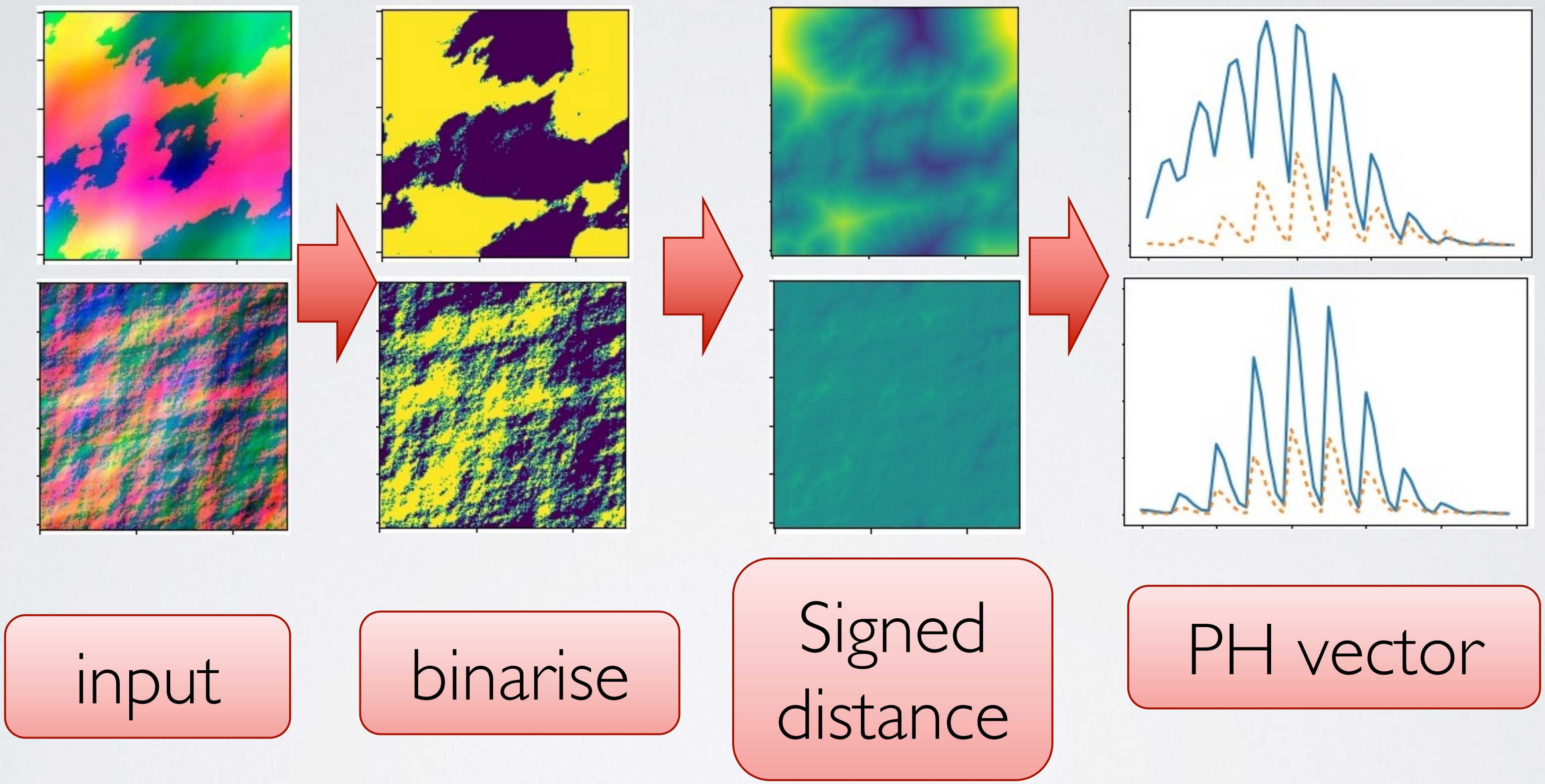


We think of an image as a function defined over the square grid.

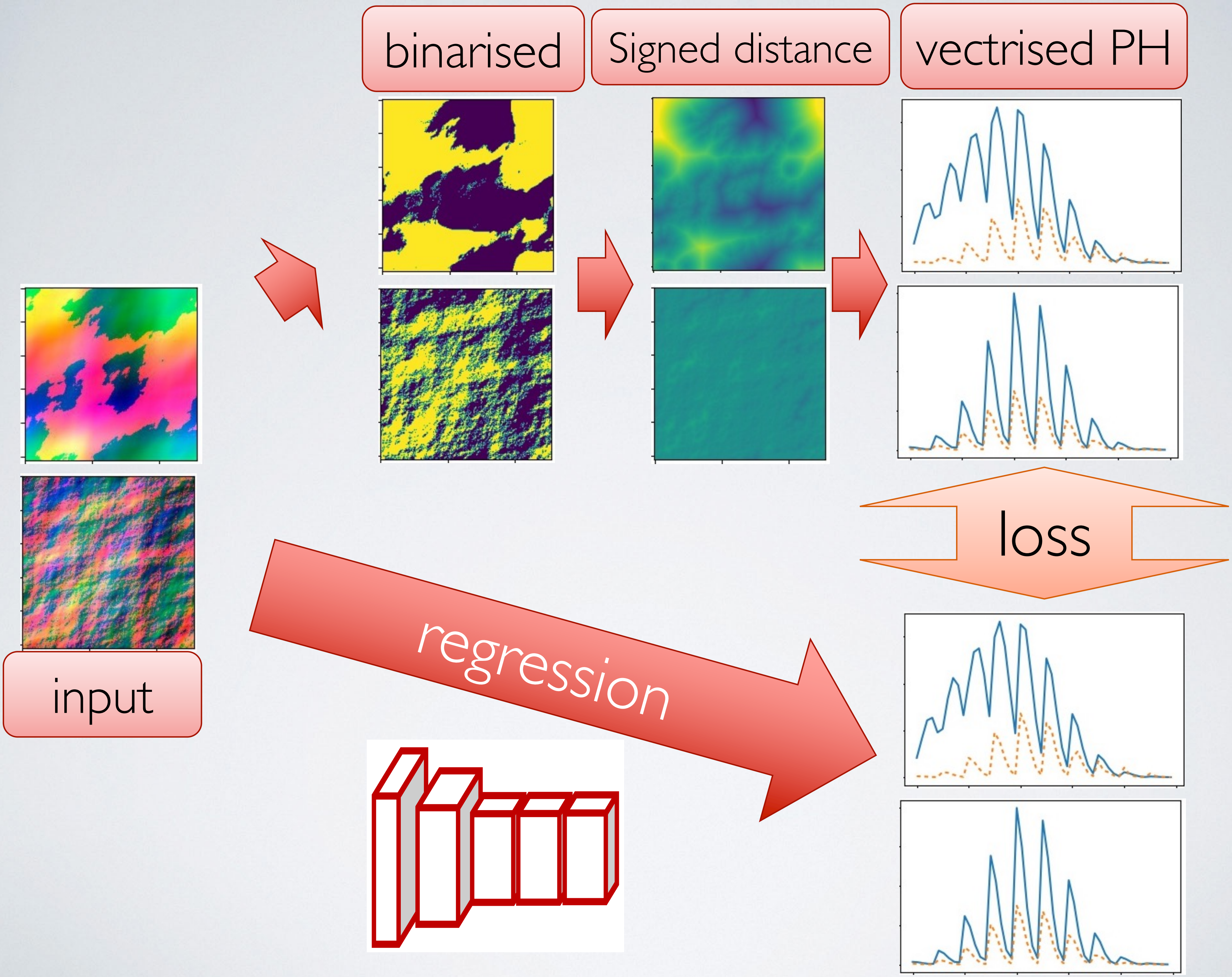
Any mathematical invariant of the function can be used as the label of the image

Through the regression task of the label, the model learns the maths!

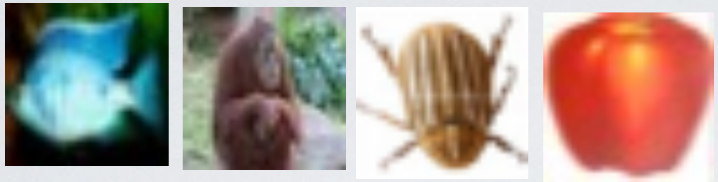
Labelling by Persistent Homology



Remark: although we use synthetic images here, any image dataset can be used.

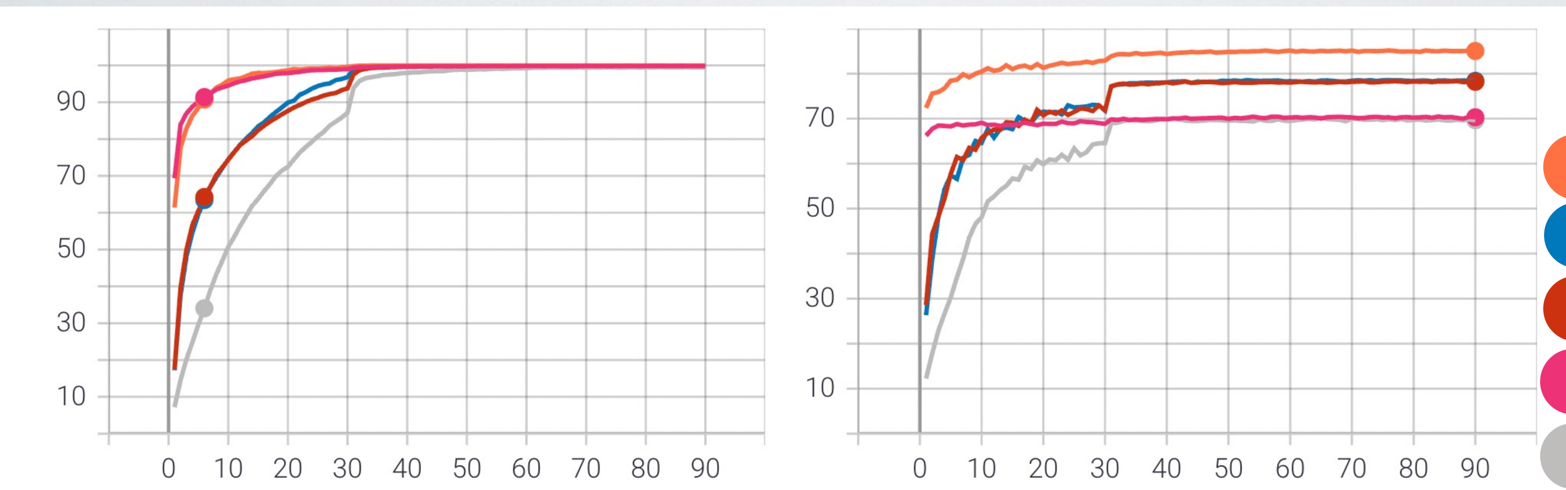

binarised
Signed distance
vectrised PH
input
loss
regression

Benchmark results



CIFAR100

100-category
Natural Image
classification



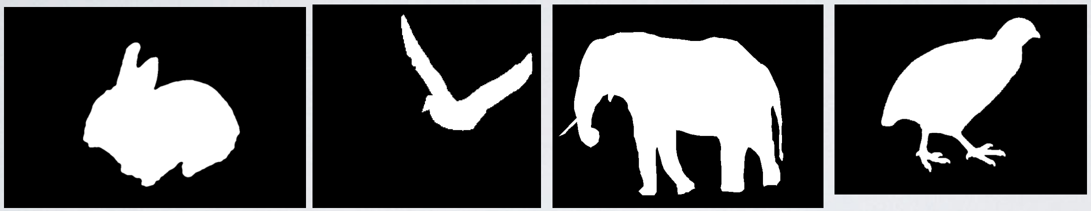
from top to bottom

ImageNet
PH-PI
FractalDB-10k
Label
Scratch

Train accuracy

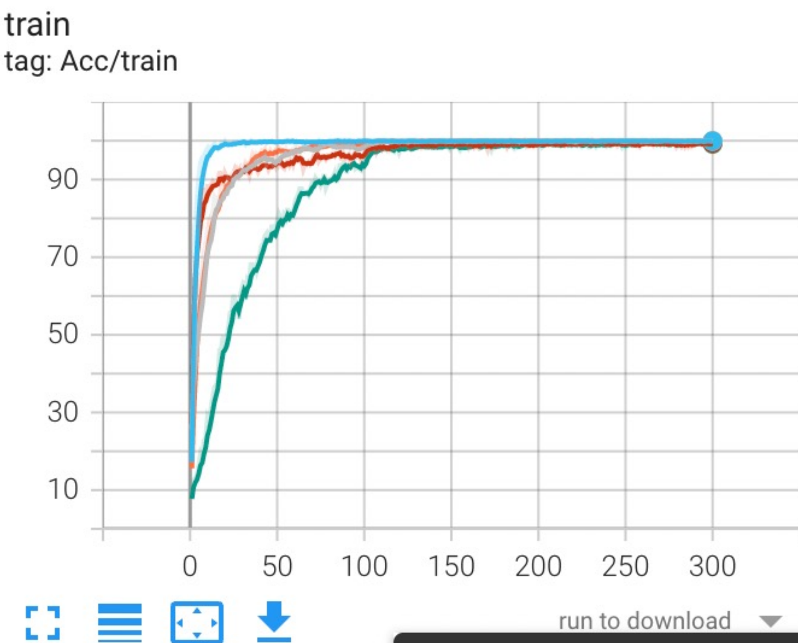
Validation accuracy

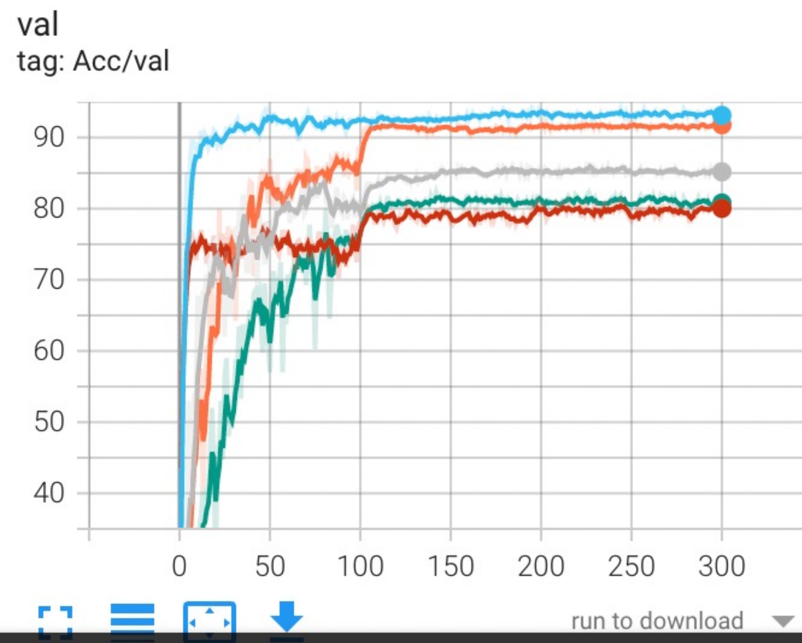
The performance is behind an ImageNet trained model, but better than training from scratch



Animal

20-category
animal silhouette
classification

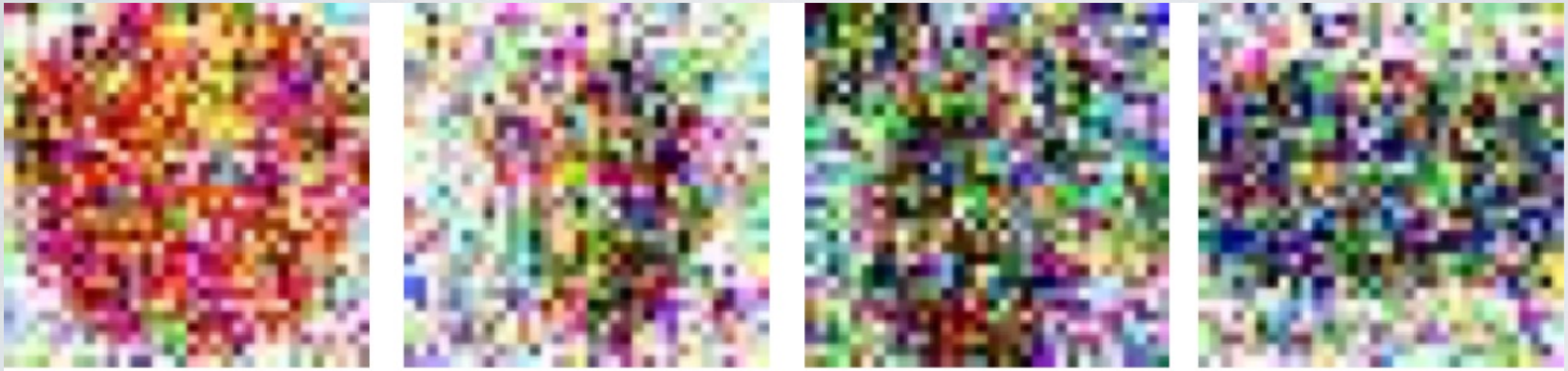




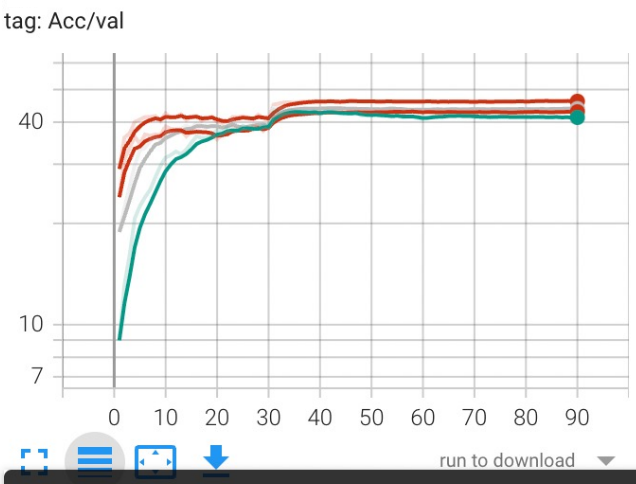
Name	Smoothed	Value
animal0.2/2021_1107_0847_finetuning_IMN_animal_val0.2	93.11	93
animal0.2/2021_1110_1236_finetuning_PH-PI_ft-animal_val0.2	$\boxed{91.79}$	92
animal0.2/2021_1107_1433_finetuning_FDB10k_ft-animal_val0.2	85.17	85
animal0.2/2021_1108_0755_finetuning_Scratch_ft-animal_val0.2	80.8	80.75
animal0.2/2021_1107_2336_class_ft-animal_val0.2	80.04	79.75

pretraining
IMN: ImageNet
FDB: FractalDB
Scratch: no pretraining
class: class label

Noised CIFAR100



Apple | Beetle | Chimpanzee | Keyboard

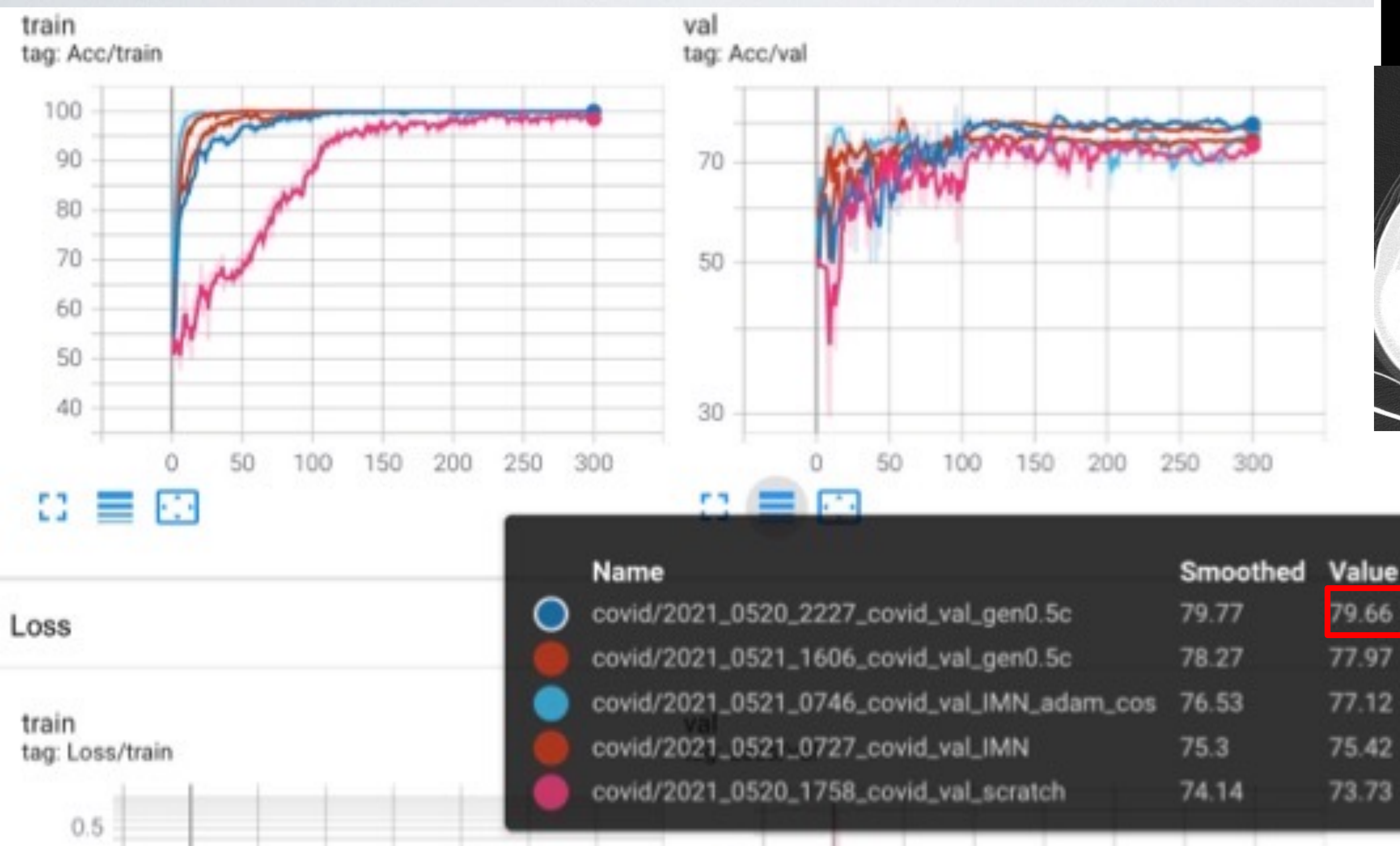


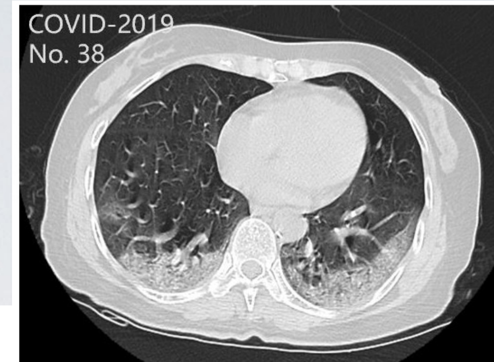
Name	Smoothed	Value	Step
gpu1080x2/2021_1017_2121_finetuning_IMN_noise_CIFAR100s0.3	46.14	46.11	90
gpu1080x2/2021_1017_1632_finetuning_noise_CIFAR100s0.3	43.88	43.95	90
gpu1080x2/2021_1017_1152_finetuning_noise_CIFAR100s0.3_FDB1000	42.94	42.88	90
gpu1080x2/2021_1016_2105_finetuning_noise_CIFAR100s0.3_scratch	41.3	41.35	90

The classification is very hard for human eyes.
How is topology robust against noise?

Covid-19 CT classification

2-class covid vs non-covid classification (COVID-CT dataset)
Various scanning conditions and non-uniform images





Our model shows better performance than the ImageNet pretrained model.
Perhaps because ImageNet does not contain medical images.

PH vectorisation methods

	Scratch	Label	PH-PI	PH-LS	PH-BC	PH-HS
CIFAR100	69.6	70.3	78.4	78.1	76.6	77.9
Animal	80.7	80.1	91.0	90.1	89.1	90.6

Persistence Image (PH-PI)
Persistence Landscape (PH-LS)
Betti curve (PH-BC)
Birth-Life histogram (PH-HS)

Label (PH vectorisation) dimensions

dimension	100	200	400	800
CIFAR100	76.8	77.7	77.4	75.0
Animal	87.4	89.9	90.5	87.5

Number of synthetic images used in pretraining

dataset size	50k	200k	400k	800k
CIFAR100	76.1	77.7	78.4	78.8
Animal	88.6	89.9	91.0	91.6

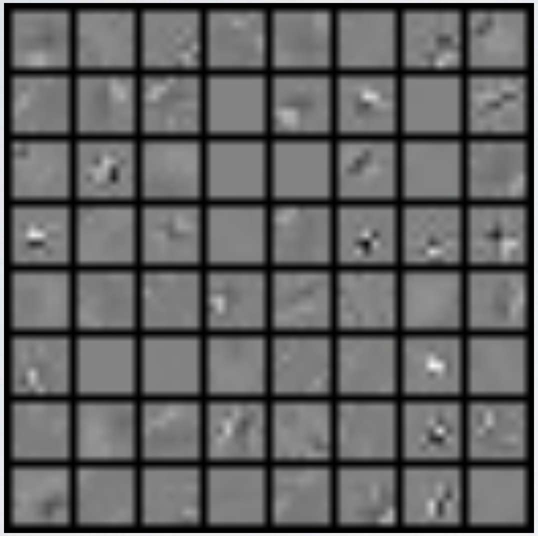
Pretraining with natural images

	Scratch	Label	PH-C	PH-A
CIFAR100	69.6	70.3	75.3	72.4
Animal	80.7	80.1	86.5	83.2

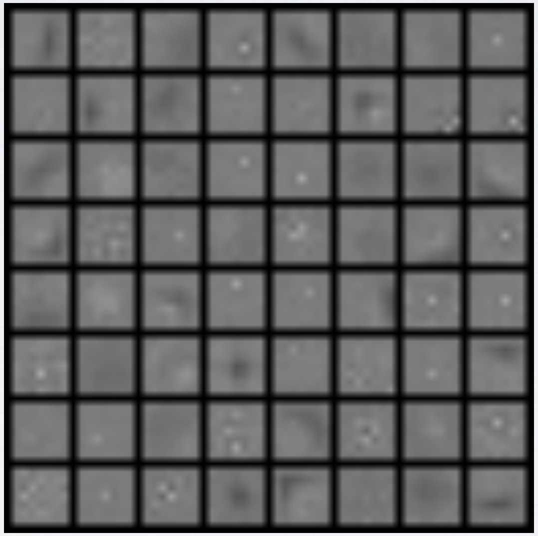
PH-A: animal dataset
PH-C: CIFAR100 dataset
Labels are not used but computed by PH-PI

Interpretation

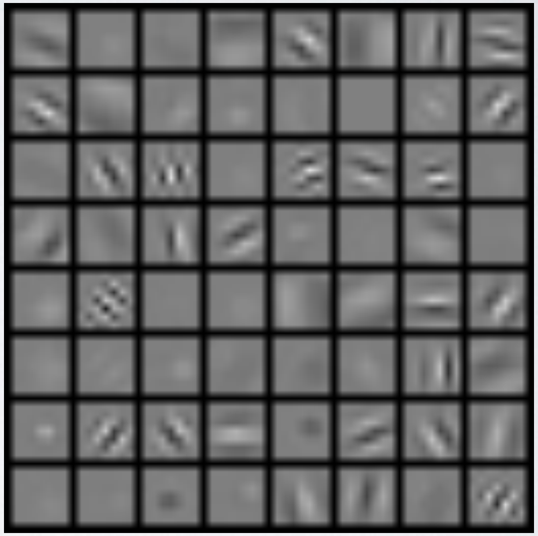
What features are learned?



PH-PI



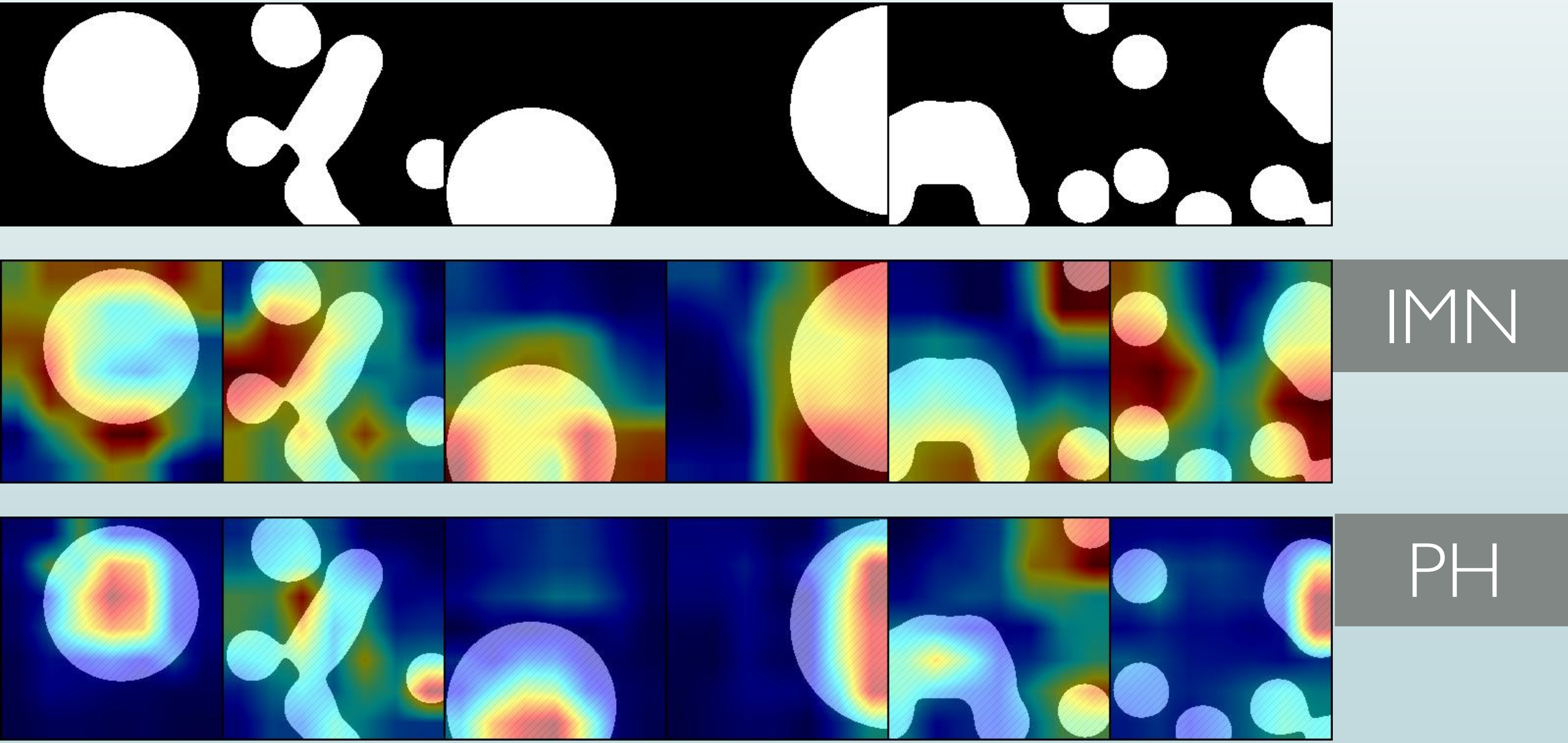
FractalDB-10k



ImageNet

First convolution kernels of pretrained models

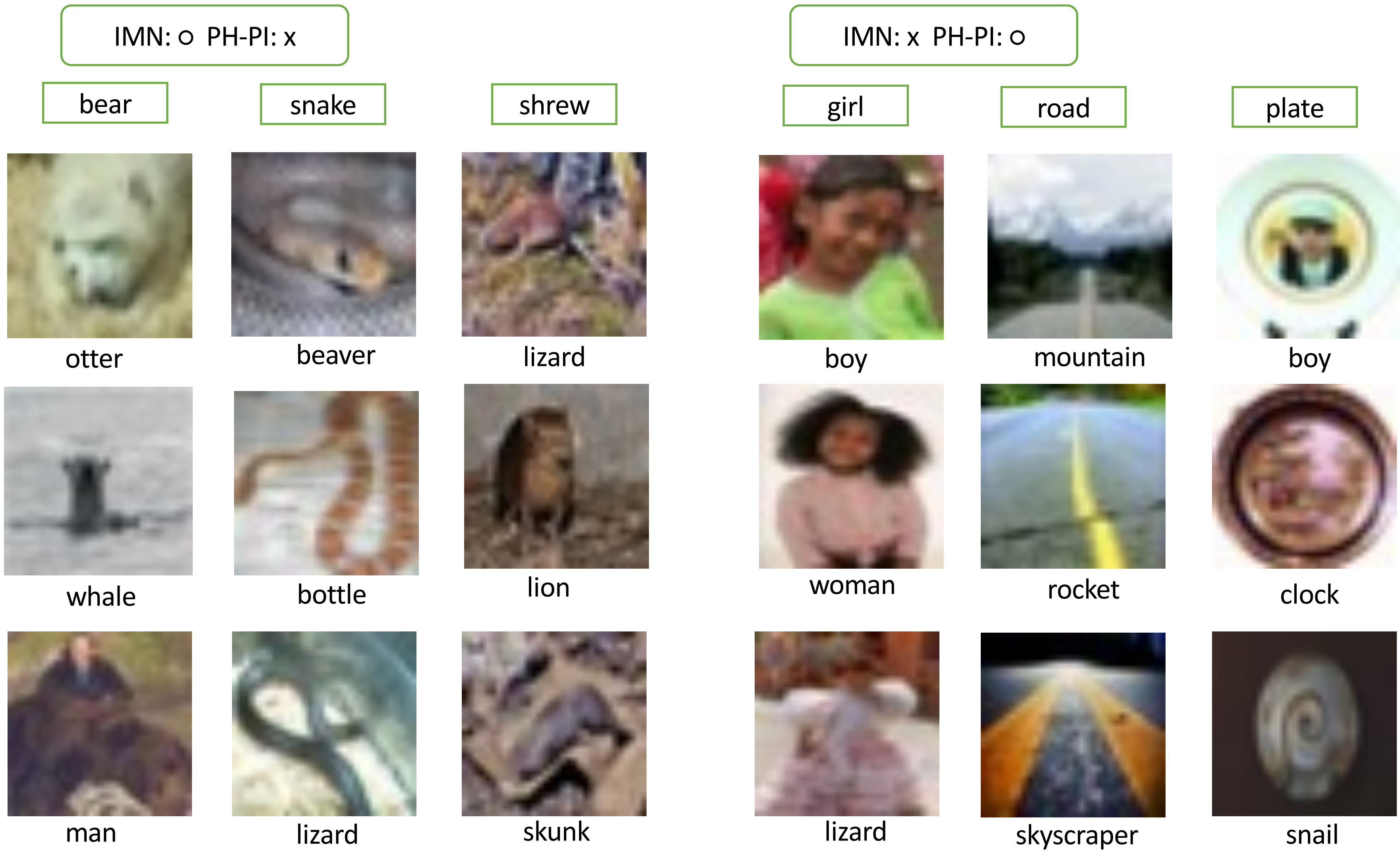
What the model focuses on?



Task: counting the number of connected components.
Visualisation: GradCAM++

IMN focuses more on edges?

What kind of mistakes the model makes?



PH-PI seems to focus more on shape than texture

MATERIALS

- Codes
  https://github.com/shizuo-kaji/PretrainCNNwithNoData

- PH computation
  https://github.com/shizuo-kaji/CubicalRipser_3dim
  S Kaji, T Sudo, K Ahara, Cubical Ripser: Software for computing persistent homology of image and volume data

- TDA Tutorial with Google Colab
  https://github.com/shizuo-kaji/TutorialTopologicalDataAnalysis
  Interactive demo on various techniques of Topological Data Analysis including Cubical Ripser

SUMMARY

- Topology (persistent homology) provides a way to extract image features that are not easy to obtain by conventional method.
- CNNs can be pretrained with synthetic images, requiring no data collection nor manual labelling
- Making CNNs learn global features encoded by topology leads to a performance gain

FUTURE WORK

- Tolerance test against adversarial attacks
- Applicability for other tasks than classification
- Theoretical analysis

Thank you!