Formula-driven Supervised Learning with Recursive Tiling Patterns

Hirokatsu Kataoka[1], Asato Matsumoto[1], Eisuke Yamagata[2], Ryosuke Yamada[1], Nakamasa Inoue[2], Yutaka Satoh[1]

1: AIST 2: TITech

ICCV Workshop 2021 (Pre-video)

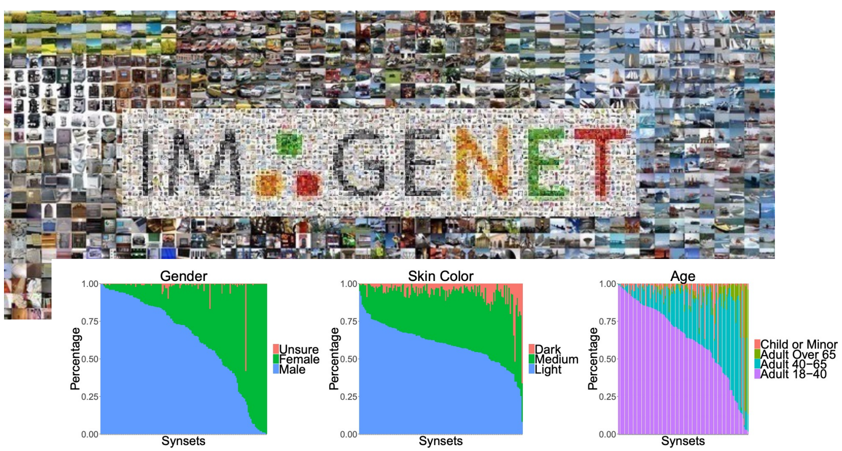
Motivation

ImageNet pre-trained CNNs have become the breakthrough

- ✓ Good feature representation for transfer learning

Datasets have problems

- ✓ Privacy-violating and ethics-related labels
- ✓ ImageNet is limited to academic- and educational-purpose



Annotation

FATE

fairness, accountability, transparency and ethics

Privacy

To overcome these issues, it is better to automatically create image datasets without any natural images

[1] Kaiyu Yang, et al., “Towards Fairer Datasets: Filtering and Balancing the Distribution of the People Subtree in the ImageNet Hierarchy,” FAT* 2020.

Formula-driven Supervised Learning (FDSL)

Fractal DataBase

- Automatically make image patterns and their labels
- With any mathematical formulas and functions

The FractalDB pre-trained model heavily relied on a large amount of parameter tuning

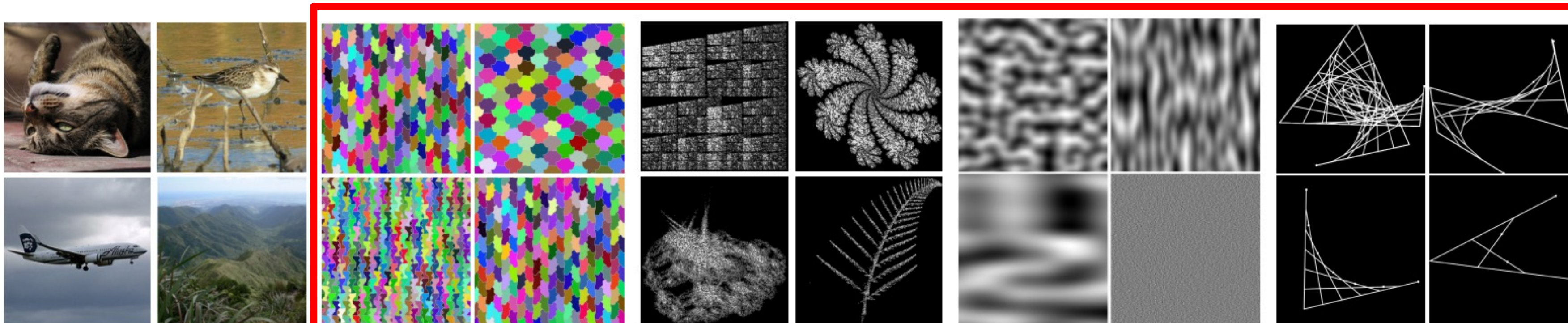
[2] Hirokatsu Kataoka, et al., "Pre-training without Natural Images," ACCV 2020.

Proposed method: TileDB

Key points

1) To make a dataset from fewer hyperparams than FractalDB
2) To acquire good feature representations through pre-training
3) To perform similarly to a human supervision

Formula-driven Supervised Learning



(a) ImageNet. (b) Tiles. (c) Fractals. (d) Perlin noise. (e) Bezier curves.

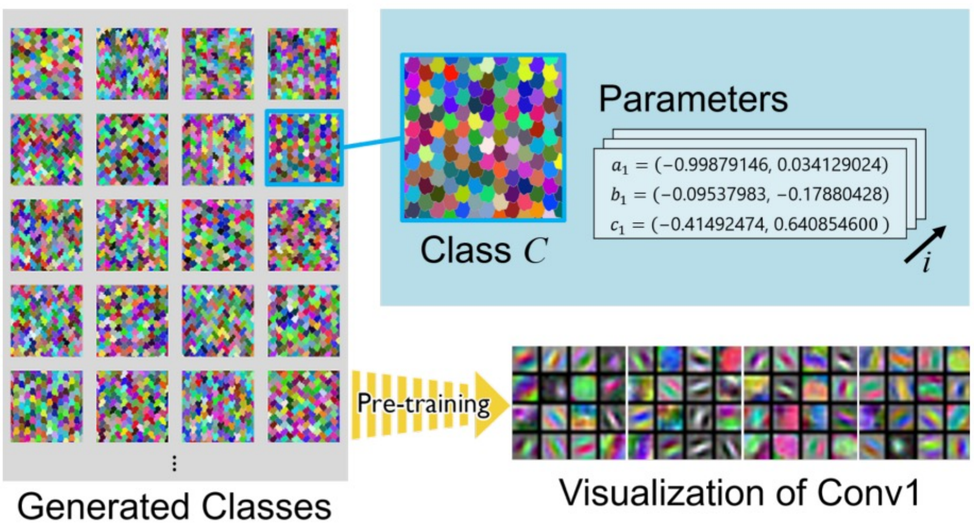
This is worthwhile to propose like a new pretext task in SSL

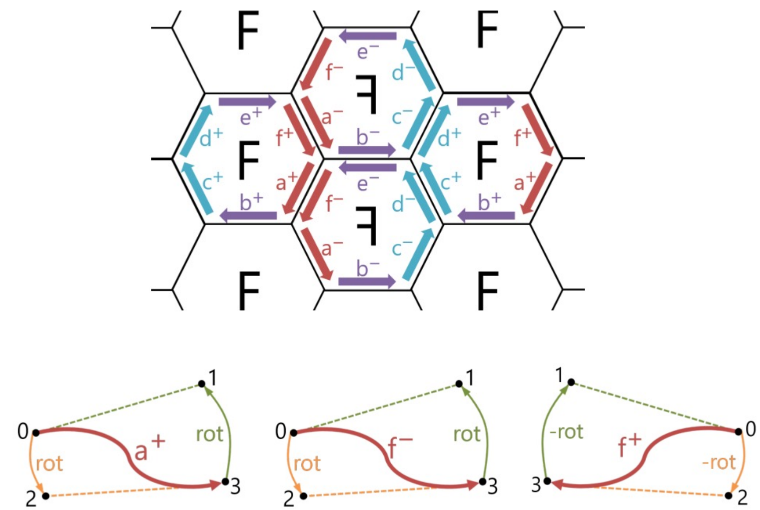
Proposed method: TileDB

TileDB

1) are created by adding three operations to regular hexagonal tiles: moving vertices, deforming edges, and moving symmetrically in a specular direction

2) basically contains 1,000 categories and 1,071 images per category



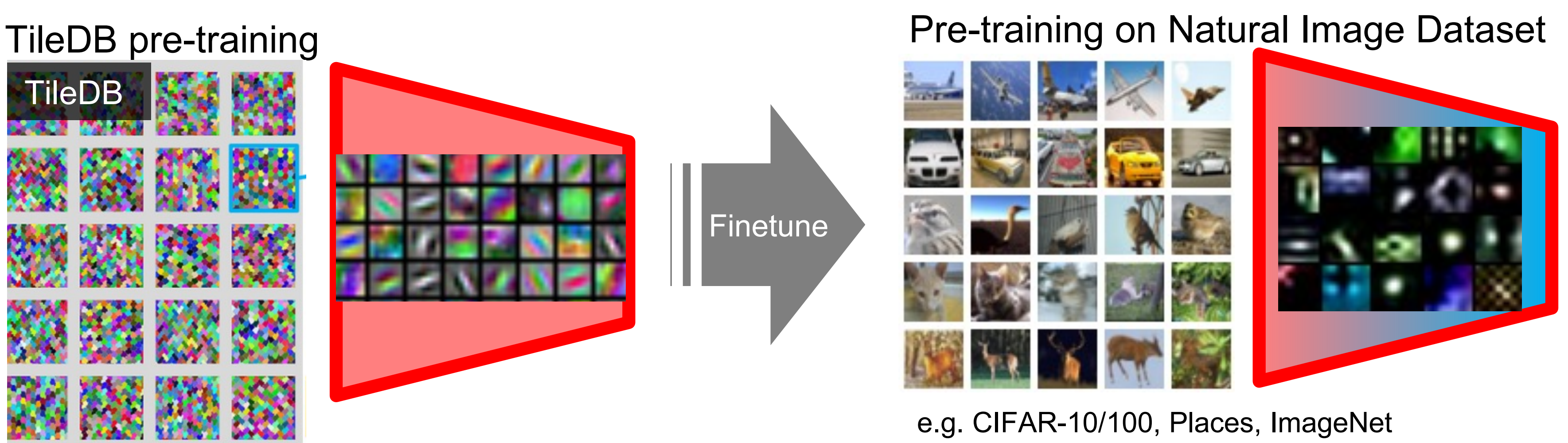


Experimental setting

Pre-training & Fine-tuning

– Pre-training without any natural images

– Fine-tuning in an ordinal way



Results: Exploration study

Compare Formula-driven Supervised Learning with other principles

Pre-training	C10	C100	P30	VOC12
From scratch	87.6	62.7	70.7	58.9
BezierCurveDB [17]	89.7	68.1	73.6	65.4
PerlinNoiseDB [17]	90.9	70.4	74.2	**69.9**
TileDB (proposed)	**92.3**	**73.5**	**75.0**	69.4

Our method partially surpasses the BezierCurveDB / PerlinNoizeDB pre-trained models

Results: Comparisons with conventional work

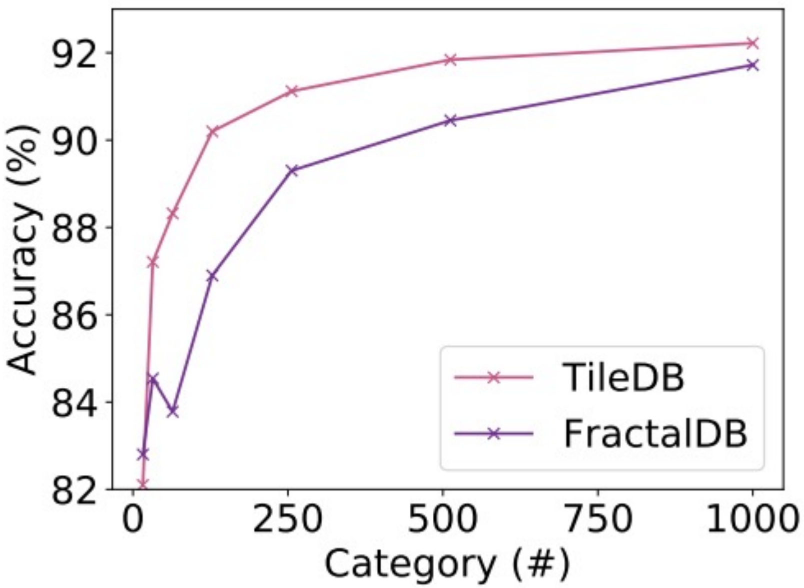
Compared our proposed method with representative pre-trained models in SL and SSL

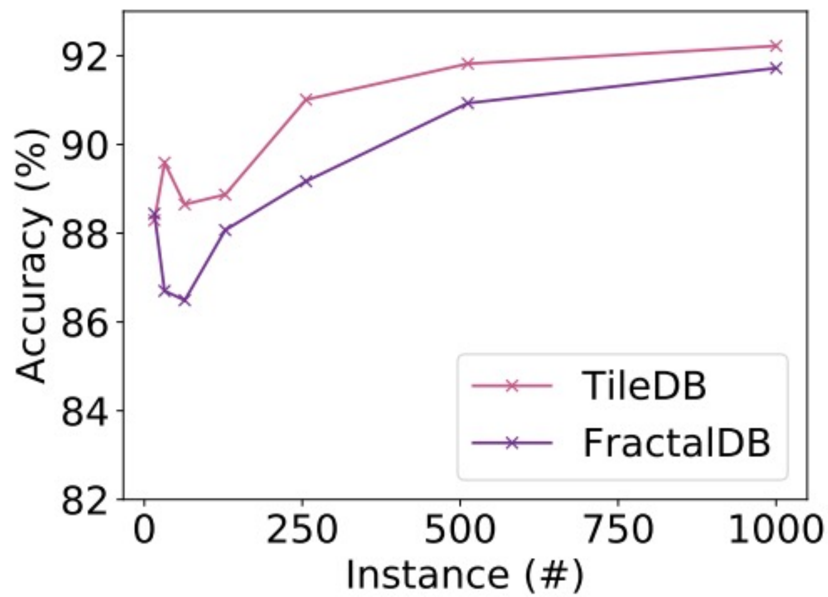
Pre-training dataset	Image	Label	C10	C100	P30	VOC12
From scratch	–	–	87.6	62.7	70.7	58.9
DeepCluster	Natural Image	Self-supervision	89.9	66.9	75.1	67.5
DeepCluster	Formula	Self-supervision	83.1	57.0	72.8	60.4
Places-30	Natural Image	Human-supervision	90.1	67.8	–	69.5
Places-365	Natural Image	Human-supervision	94.2	76.9	–	78.6
ImageNet-100	Natural Image	Human-supervision	91.3	70.6	–	72.0
ImageNet-1k	Natural Image	Human-supervision	**96.8**	**84.6**	**79.5**	**85.8**
TileDB (proposed)	Formula	Formula-supervision	**92.5**	**73.7**	**78.0**	**71.4**

TileDB pre-trained model is still better than 100k-order supervised datasets at all fine-tuned datasets

Results: Comparisons with FractalDB

Relationship between accuracy and #category / #instance on CIFAR-10

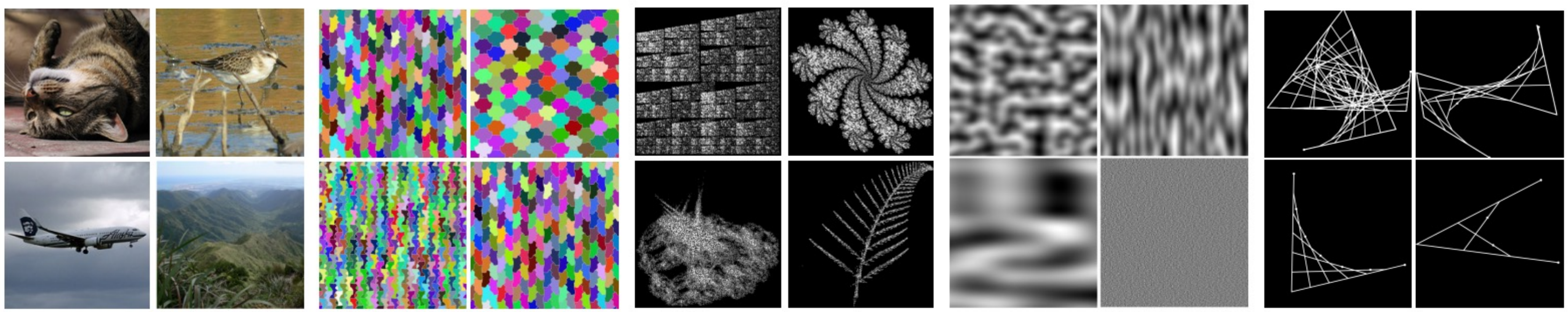




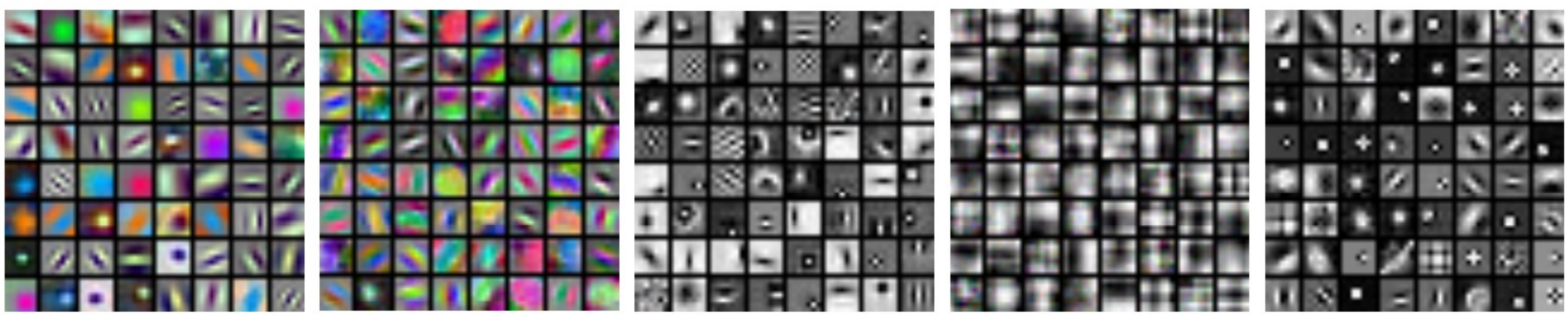
TileDB tends to be a simple tuning in a compact dataset

Results: Comparisons of feature representations

Compared both FDSL and ImageNet in terms of the initial convolutional maps



(a) ImageNet. (b) Tiles. (c) Fractals. (d) Perlin noise. (e) Bezier curves.



(f) PT on ImageNet. (g) PT on tiles. (h) PT on fractals. (i) PT on Perlin noise. (j) PT on Bezier curves.

TileDB pretrained features are similar to ImageNet pre-trained features on ResNet-50

We're also trying to implement how to simply and automatically create a large-scale image dataset for trustworthy computer vision.