Pre-training without Natural Images

Hirokatsu Kataoka [1], Kazushige Okayasu[1], Asato Matsumoto[1], Eisuke Yamagata[2], Ryosuke Yamada[3], Nakamasa Inoue[2], Akio Nakamura[3], Yutaka Satoh[1]

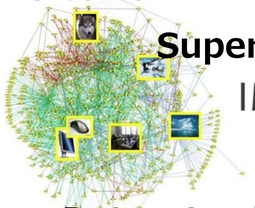
[1] AIST, [2] TITech, [3]Tokyo Denki Univ.

cvpaper.challenge

National Institute of Advanced Industrial Science and Technology
AIST

Can we pre-train CNN w/o any natural images?

- Though the most promising frameworks are supervised/self-supervised learning (SL/SSL), there still exists the dataset-related problems.
- Natural images should be replaced as well as human annotations in pre-training



Supervised Learning

IMAGENET
http://www.image-net.org/

Dataset-related problems
Annotation, FATE, Privacy…

Could we automatically make image patterns and their labels with any functions?

Formula-driven Supervised Learning (FDSL):

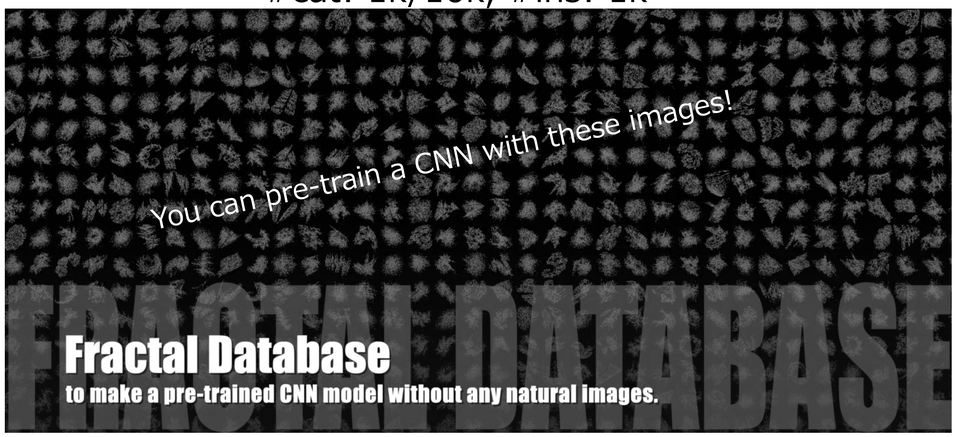
Automatically makes img patterns and the labels with any functions

Fractal geometry from ImageNet

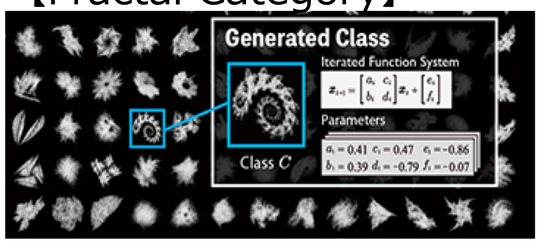
CNN trains a natural principle from ImageNet dataset

Directly render and train Fractals

FractalDB #cat: 1k/10k, #ins: 1k



【Fractal Category】

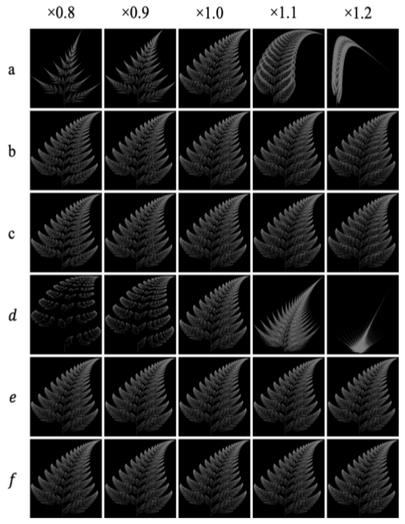


1. Rendering with randomized a〜f, w
2. Add category if filling rate (> r) in the image
   - Iterate 1&2 up to defined #category (C)
   - Param separation makes a different category

$$\mathrm{IFS} = \{\mathcal{X}; w_1, w_2, \cdots, w_N; p_1, p_2, \cdots, p_N\}$$

$$w_i(\boldsymbol{x}; \theta_i) = \begin{bmatrix} a_i & b_i \\ c_i & d_i \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} e_i \\ f_i \end{bmatrix}$$

【Instance aug.】



Vary a〜f and rotate images

Experiments

Classification on ResNet-50, Standard Training, Replacing Pre-training
CIFAR-10/100 (C10/C100), ImageNet (IN1k), Places 365 (P365), VOC12, Omniglot (OG)

Method	Pre-train Img	Type	C10	C100	IN1k	P365	VOC12	OG
Scratch	–	–	87.6	62.7	**76.1**	49.9	58.9	1.1
DC-10k	Natural	Self-supervision	89.9	66.9	66.2	**51.5**	67.5	15.2
Places-30	Natural	Supervision	90.1	67.8	69.1	–	69.5	6.4
Places-365	Natural	Supervision	**94.2**	76.9	71.4	–	**78.6**	10.5
ImageNet-100	Natural	Supervision	91.3	70.6	–	49.7	72.0	12.3
ImageNet-1k	Natural	Supervision	**96.8**	**84.6**	–	50.3	**85.8**	17.5
FractalDB-1k	Formula	Formula-supervision	93.4	75.7	70.3	49.5	58.9	**20.9**
FractalDB-10k	Formula	Formula-supervision	94.1	**77.3**	**71.5**	**50.8**	73.6	**29.2**

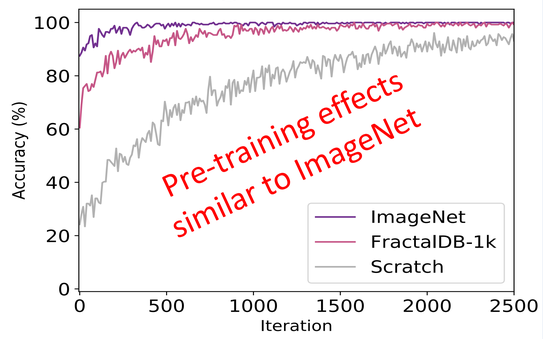
【Baseline】
Scratch: Random param.
ImageNet/Places: SL
DeepCluster: SSL

【Proposed method】
FractalDB-1k/10k: FDSL

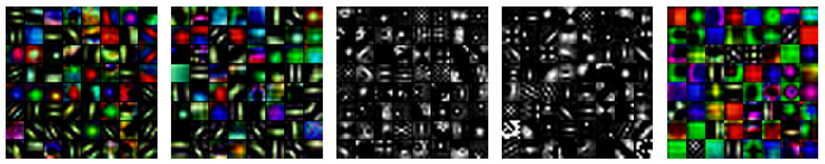
Results

[vs. Scratch] FDSL achieved much higher rates
[vs. SSL] FDSL is better than the DeepCluster with 10k categories (DC-10k)
[vs. SL] FDSL is still better than 100k-order supervised datasets (Places30/ImageNet100)
[vs. SL] FDSL partially surpasses the ImageNet (on OG) /Places (on C100, IN1k) pre-trained models

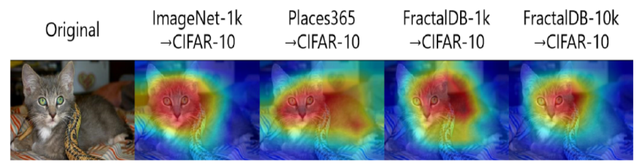
Pre-training effect



Visualization



(a) ImageNet (b) Places365 (c) Fractal-1K (d) Fractal-10K (e) DC-10k



If we could improve the FDSL, ImageNet pre-trained model may be replaced so as to protect FATE, preserve privacy, and decrease annotation labor.

Paper, code and datasets are available:
https://hirokatsukataoka16.github.io/Pretraining-without-Natural-Images/