Replacing Labeled Real-image Datasets with Auto-generated Contours

CVPR 2022

Hirokatsu Kataoka[*], Ryo Hayamizu[*], Ryosuke Yamada[*], Kodai Nakashima[*], Sora Takashima[*,], Xinyu Zhang[*,**], Edgar Josafat MARTINEZ-NORIEGA[*,**], Nakamasa Inoue[*,**], Rio Yokota[*,**]**

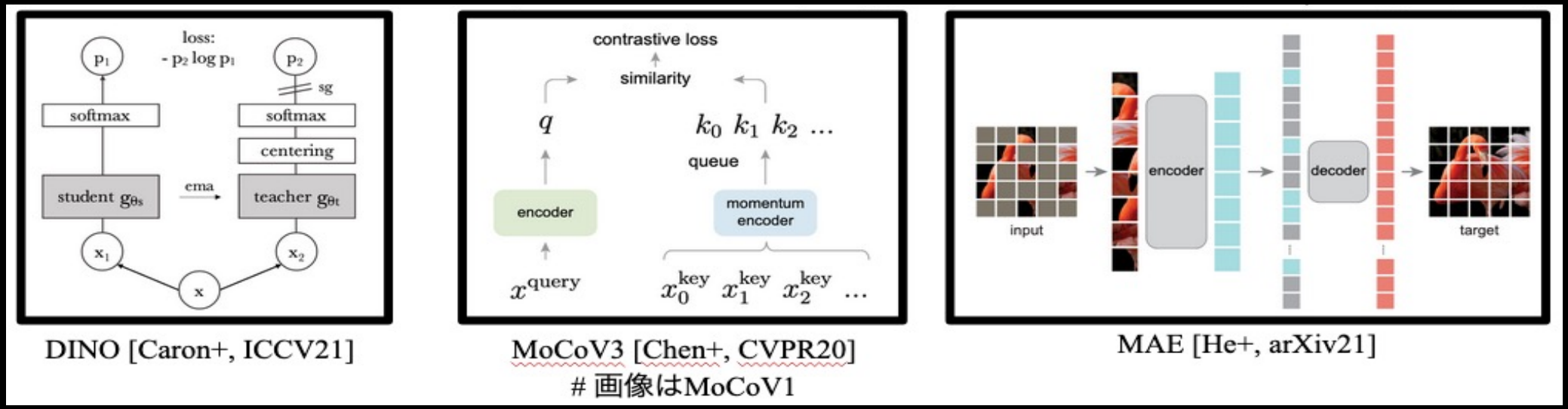
*** National Institute of Advanced Industrial Science and Technology (AIST)**
****Tokyo Institute of Technology**

深刻化する大規模画像データセット関連問題

教師あり学習(Supervised Learning; SL)から自己教師あり学習(Self-Supervised Learning; SSL)へ

- CV分野で最も”アツい”トピック

【Vision TransformerのSSL手法】



DINO [Caron+, ICCV21]

MoCoV3 [Chen+, CVPR20]
画像はMoCoV1

MAE [He+, arXiv21]

78.2　　83.2　　83.6

モデルはViT-Base/16，精度はImageNet-1k Fine-tuning時（Top-1 Acc.を記載）

深刻化する大規模画像データセット関連問題

教師あり学習(Supervised Learning; SL)から自己教師あり学習(Self-Supervised Learning; SSL)へ

- CV分野で最も”アツい”トピック

データ公開停止

攻撃的ラベルを含み，修正困難であるとして80M Tiny Imagesが公開停止に追い込まれた

https://groups.csail.mit.edu/vision/TinyImages/

差別的出力

性別・人種などデータの偏りが不公平・差別的出力につながるとの結果

https://arxiv.org/pdf/1912.07726.pdf

未公開データ

JFT-300M/3B / IG-3.5Bなど未公開データでの学習が最高水準の結果になり続けている

http://openaccess.thecvf.com/content_ICCV_2017/papers/Sun_Revisiting_Unreasonable_Effectiveness_ICCV_2017_paper.pdf
https://arxiv.org/pdf/1805.00932.pdf

商用利用禁止

不特定多数画像を収集しているため，ImageNetをはじめ教育・研究目的に利用制限

https://image-net.org/download.php

不透明性問題

標準画像データセットにおいても約6%が不適切・誤りを含む教師ラベルであると判断

https://www.technologyreview.jp/s/238631/error-riddled-datasets-are-warping-our-sense-of-how-good-ai-really-is/

ラベル付コスト

データセットの規模は大きくなる一方，人間の労力がかかりすぎてしまう

※自己教師あり学習で根本解決できるのは「ラベル付コスト」のみ

実画像による大規模データセットを利用する障壁はあまりにも大きい

実画像を用いずに視覚機能を獲得

膨大な画素空間からImageNetはなぜ良好な視覚機能を獲得できるのか？

Fractal geometry from ImageNet dataset

深層学習では視覚機能をある種の
自然法則から学んでいるのでは？

自然法則から画像に直接投影・学習することを着想

Pre-training without Natural Images (ACCV20 Best Paper H. M. Award / IJCV22)

数式ドリブン教師あり学習

Formula-driven Supervised Learning (FDSL)

- 生成規則（フラクタル幾何）から画像と教師ラベルを同時生成

生成規則(式)

ラベルの生成 (ランダムサンプリング)

$$\Theta = \{(\theta_i, p_i)\}_{i=1}^{N}$$

データの生成 (IFS)

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

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

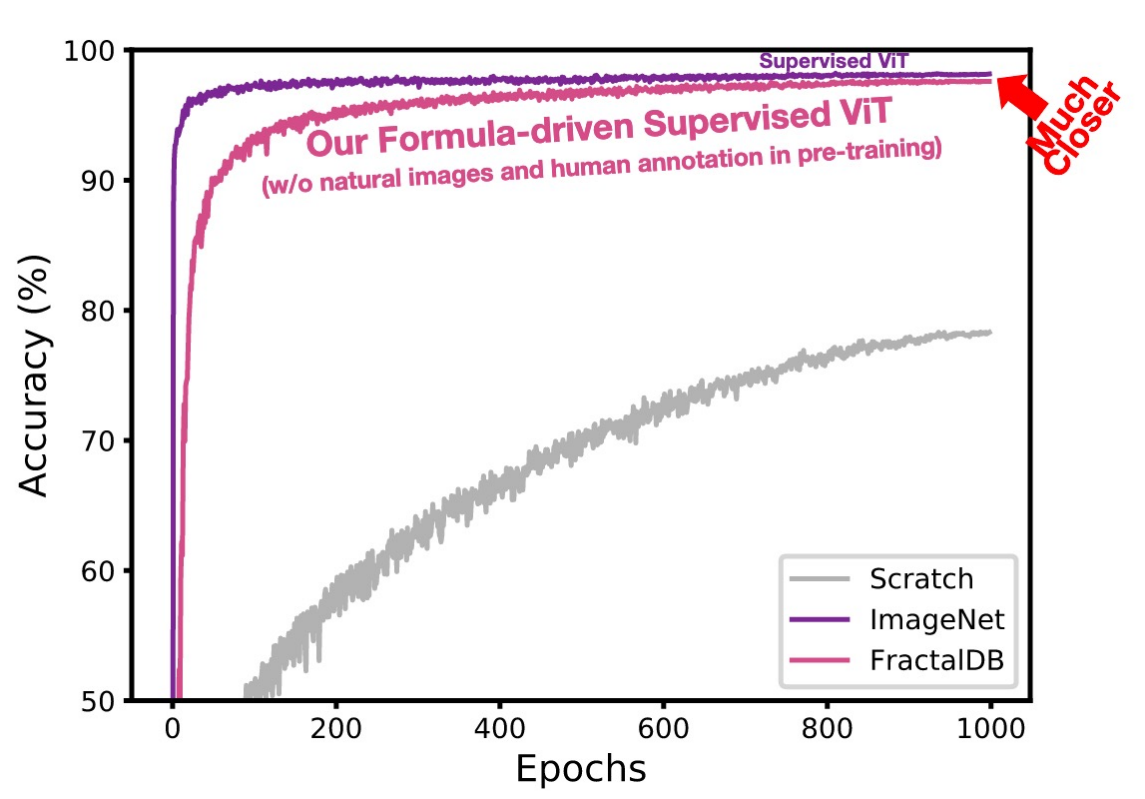
$$p_i = p(w^* = w_i) \qquad \boldsymbol{x}_{t+1} = w^*(\boldsymbol{x}_t)$$

画像・教師ラベル：生成規則より自動生成

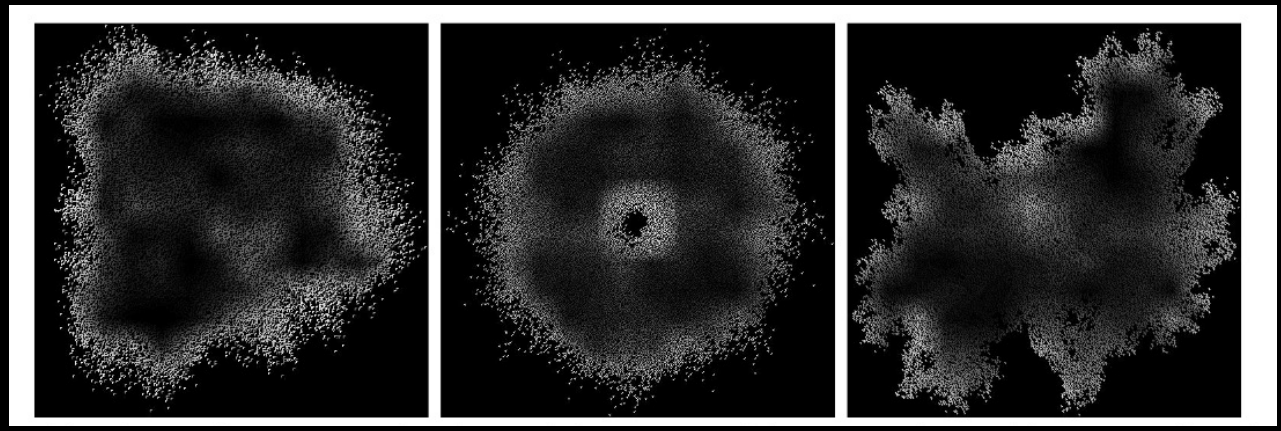
Can Vision Transformers Learn without Natural Images? (AAAI22)

FractalDBでVision Transformer（ViT）の事前学習に成功

- 実画像枚数を従来の14,000,000から実質0に



ViTの自己注視可視化結果



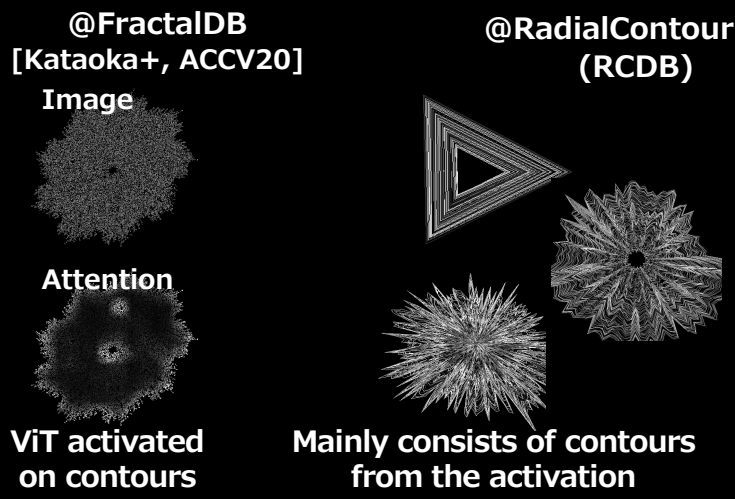
→実はフラクタルが重要なのではなく，
輪郭が複雑であれば良いのでは？

Replacing Labeled Real-image Datasets with Auto-generated Contours (CVPR22)

自動生成輪郭でラベル付実画像データセットを置き換える

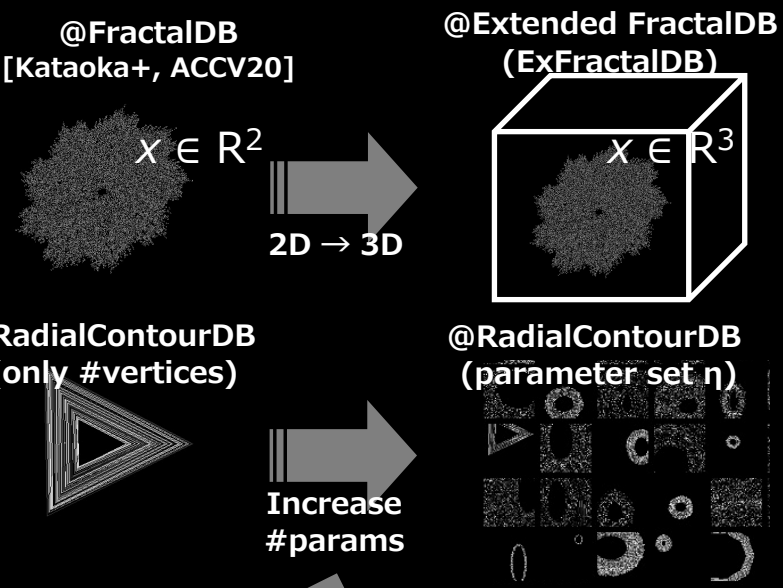
- ふたつの仮説を検証

Hypothesis 1:
Object contours are what matter in FDSL datasets
FDSLの画像表現では物体輪郭こそが重要



輪郭形状の「極端な例」として，画像の主要成分が輪郭である放射輪郭（Radial Contour）を実装

Hypothesis 2:
Task difficulty matters in FDSL pre-training
FDSL事前学習のタスク難易度が精度向上に寄与

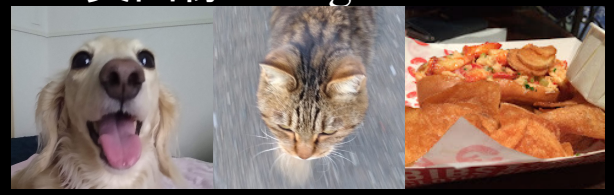


事前学習タスクの難易度と直結する成分は「数式のパラメータ数」

一般物体認識・物体検出・領域分割タスクにおける評価

ImageNet-1k / MS COCOデータセット

実画像: ImageNet-21k



一般物体認識の性能
ImageNet における精度

81.8%

3Dフラクタル幾何画像:
ExFractalDB-21k

82.7%

放射輪郭画像: RCDB-21k

82.4%

Pre-training	COCO Det AP_{50} / AP / AP_{75}	COCO Inst Seg AP_{50} / AP / AP_{75}
Scratch	63.7 / 42.2 / 46.1	60.7 / 38.5 / 41.3
ImageNet-1k	69.2 / 48.2 / 53.0	66.6 / 43.1 / 46.5
ImageNet-21k	**70.7 / 48.8 / 53.2**	**67.7 / 43.6 / 47.0**
ExFractalDB-1k	69.1 / **48.0 / 52.8**	66.3 / **42.8** / 45.9
ExFractalDB-21k	**69.2 / 48.0** / 52.6	**66.4 / 42.8 / 46.1**
RCDB-1k	68.3 / 47.4 / 51.9	65.7 / 42.2 / 45.5
RCDB-21k	67.7 / 46.6 / 51.2	64.8 / 41.6 / 44.7

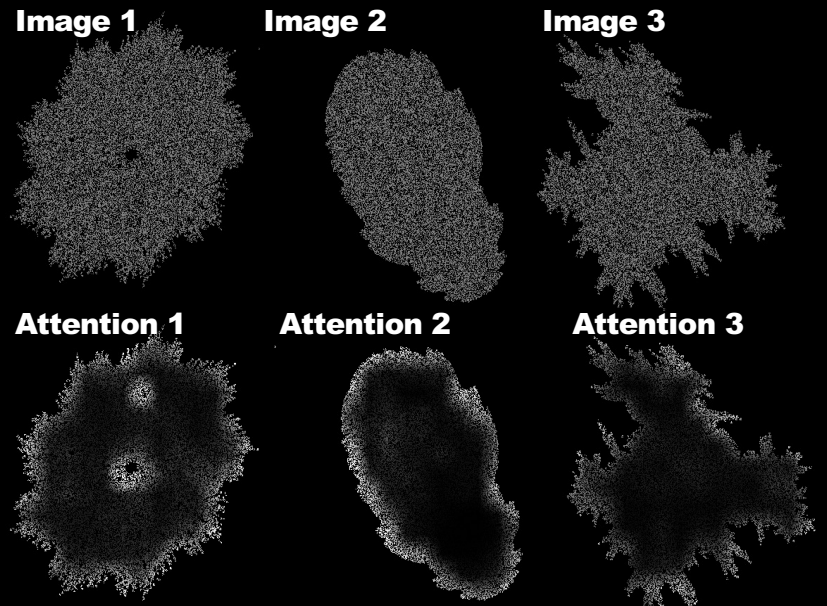
ImageNet-21kを超える事前学習効果！
Fractalでラベル付実画像データセットを超えるのは凄いが，放射輪郭画像のみでも置き換え可能！

輪郭識別のみの事前学習ながら，物体検出・インスタンスセグメンテーションはImageNet-1kに近い事前学習効果を発揮

仮説１：FDSLの画像表現では物体輪郭こそが重要

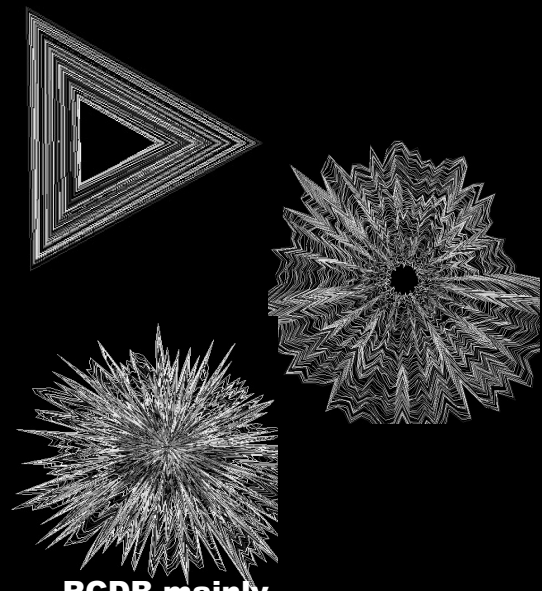
Hypothesis 1:
Object contours are what matter in FDSL datasets

@FractalDB [Kataoka+, ACCV20]



ViT activated on contours of fractal images

@RadialContourDB (RCDB)



RCDB mainly consists of contours

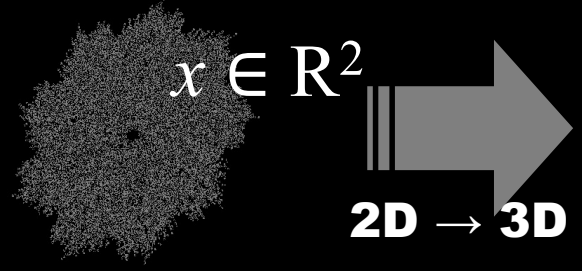
Pre-training	C10	C100	Cars	Flowers
Scratch	78.3	57.7	11.6	77.1
Perlin Noise [21]	95.0	78.4	70.6	96.1
Dead Leaves [3]	95.9	79.6	72.8	96.9
Bezier Curves [21]	96.7	80.3	82.8	98.5
RCDB	**96.8**	**81.6**	84.2	**98.7**
FractalDB [27]	**96.8**	**81.6**	**86.0**	98.3

ハイパラ探索含め，高度なことをせずとも
RCDBはFractalDBに近い精度に到達

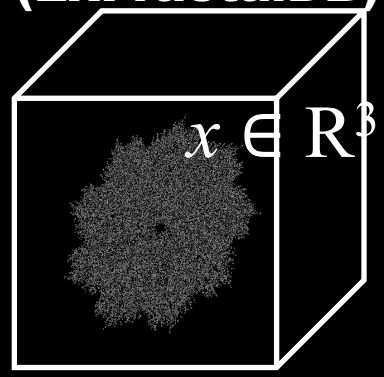
仮説２：FDSL事前学習のタスク難易度が精度向上に寄与

Hypothesis 2:
Task difficulty matters in FDSL pre-training

@FractalDB
[Kataoka+, ACCV20]

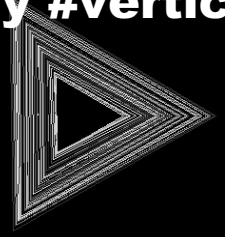


@Extended FractalDB (ExFractalDB)



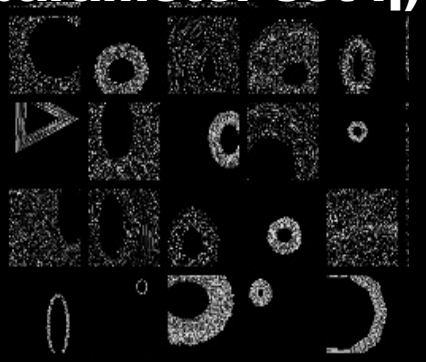
3D Fractalを描画，ランダム視点から2D画像に投影

@RadialContourDB (only #vertices)



Increase #params

@RadialContourDB (parameter set η)



頂点数がメインだが，半径長・輪郭数・滑らかさなど
パラメータセットを調整しつつカテゴリ定義

Pre-training	C10	C100	Cars	Flowers
BC	96.9 (0.2)	81.4 (1.1)	85.9 (3.1)	97.9 (-0.6)
RCDB	97.0 (0.2)	**82.2** (0.6)	86.5 (2.4)	**98.9** (0.2)
ExFractalDB	**97.2** (0.4)	81.8 (0.2)	**87.0** (1.0)	**98.9** (0.6)

事前学習データセットを生成する数式のパラメータ数増加が追加学習の精度に貢献

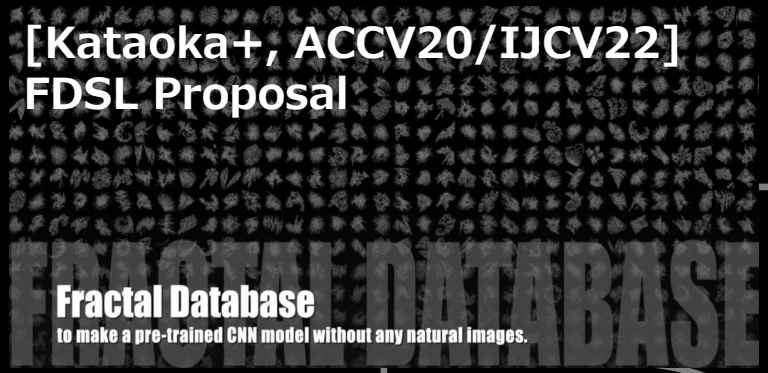
Point Cloud Pre-training with Natural 3D Structures

CVPR 2022

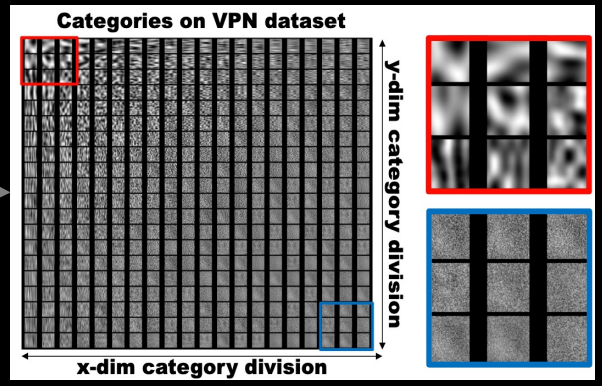
Ryosuke Yamada*, Hirokatsu Kataoka*, Naoya Chiba, Yukiyasu Domae*, Testuya Ogata*, ****

*** National Institute of Advanced Industrial Science and Technology (AIST)**
****Waseda University**

【FDSL Family】



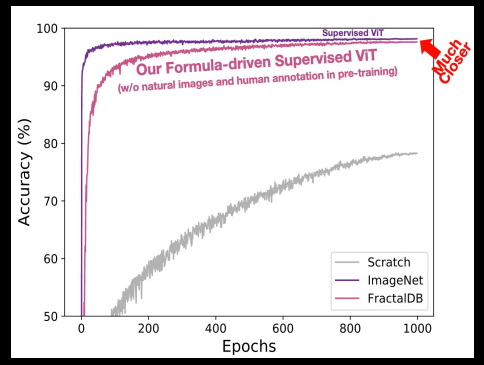
Spatiotemporal Domain



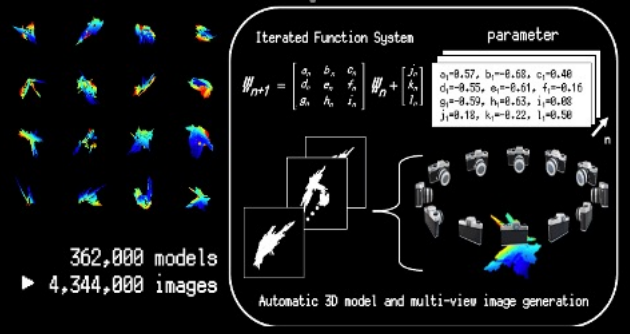
Video Perlin Noise [Kataoka+, WACV22]

3D Domain

Vision Transformers



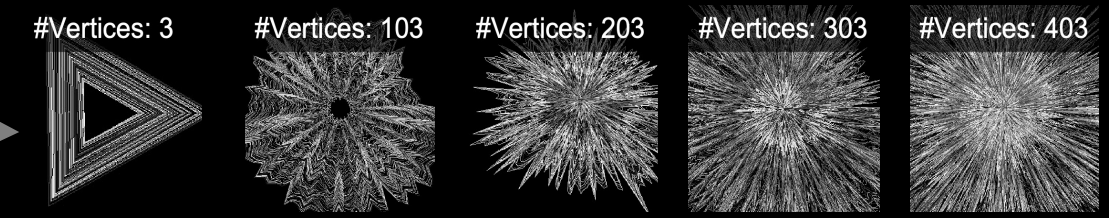
FractalDB Pre-trained ViT [Nakashima+, AAAI22]



3Dドメインにも適用できないか？

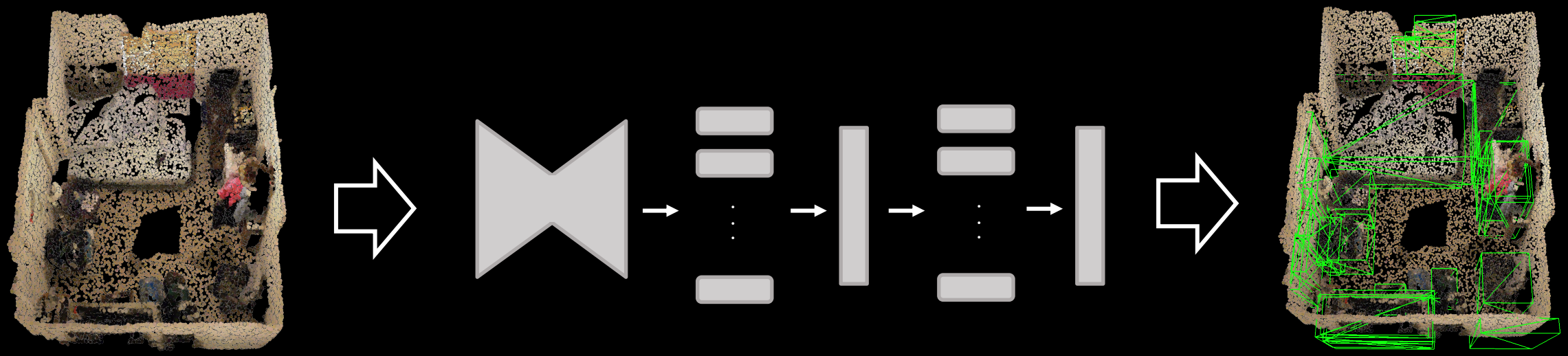
Multi-viewpoint / Point Cloud [Yamada+, IROS22/CVPR22]

Enhanced by Hypotheses



Replacing Labeled Real-image Datasets [Kataoka+, CVPR22]

3Dドメインでは決定版の事前学習データセットが不在
データセット構築のコストが高すぎる



数式ドリブン3D点群事前学習

実世界の生成規則を学習することで
従来の3Dデータセットより
汎用的特徴が獲得できるのでは？

Point Cloud Fractal Database: 3Dフラクタルモデル生成

3Dフラクタルモデルをいかに作るか？ → 変換行列を3Dに拡張するのみ

$3D\ IFS = \{(w_j, p_j)\}_{j=1}^{N}$ w_j: Affine Transformation, p_j: Selection probability

1. 3D-IFS parameters setting

$$w_1 = \begin{bmatrix} 0.57 & -0.68 & 0{,}40 \\ -0.55 & -0.61 & -0.16 \\ -0.59 & 0.63 & 0.08 \end{bmatrix} + \begin{bmatrix} 0.18 \\ -0.22 \\ 0.50 \end{bmatrix}$$

N

2. Affine transformation

$$\boldsymbol{x}_i = w_j\, \boldsymbol{x}_{i-1}$$

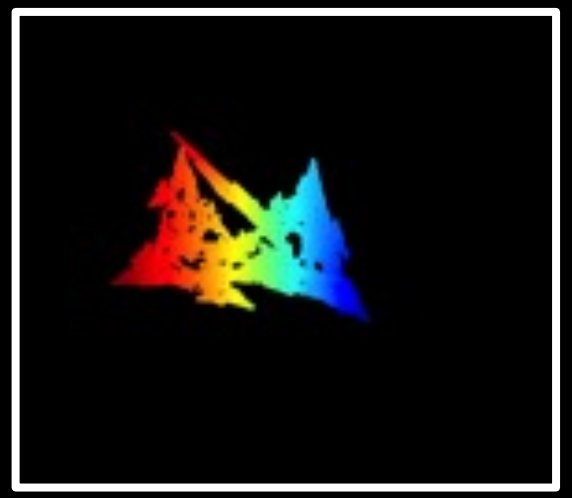
$(i = 1, 2, 3, .., n)$

$\boldsymbol{x} = [x, y, z]^T$

3D fractal model: $P = \{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$

3. Variance check & Fractal category definition

$$min(Var[x], Var[y], Var[z]) = \mathbf{0.17\ \ldots} > 0.15$$



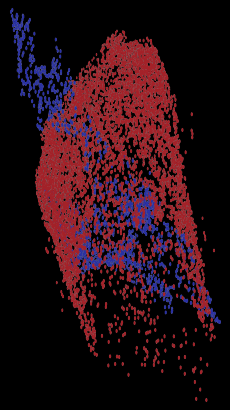
Point Cloud Fractal Database: インスタンス拡張方法

インスタンス拡張 / 3Dシーン構築

インスタンスはMix

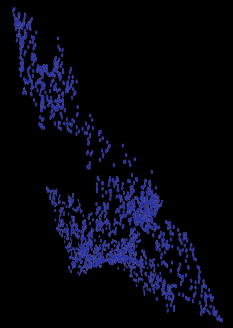
Main Category

Point number: 3,200



FractalNoiseMix

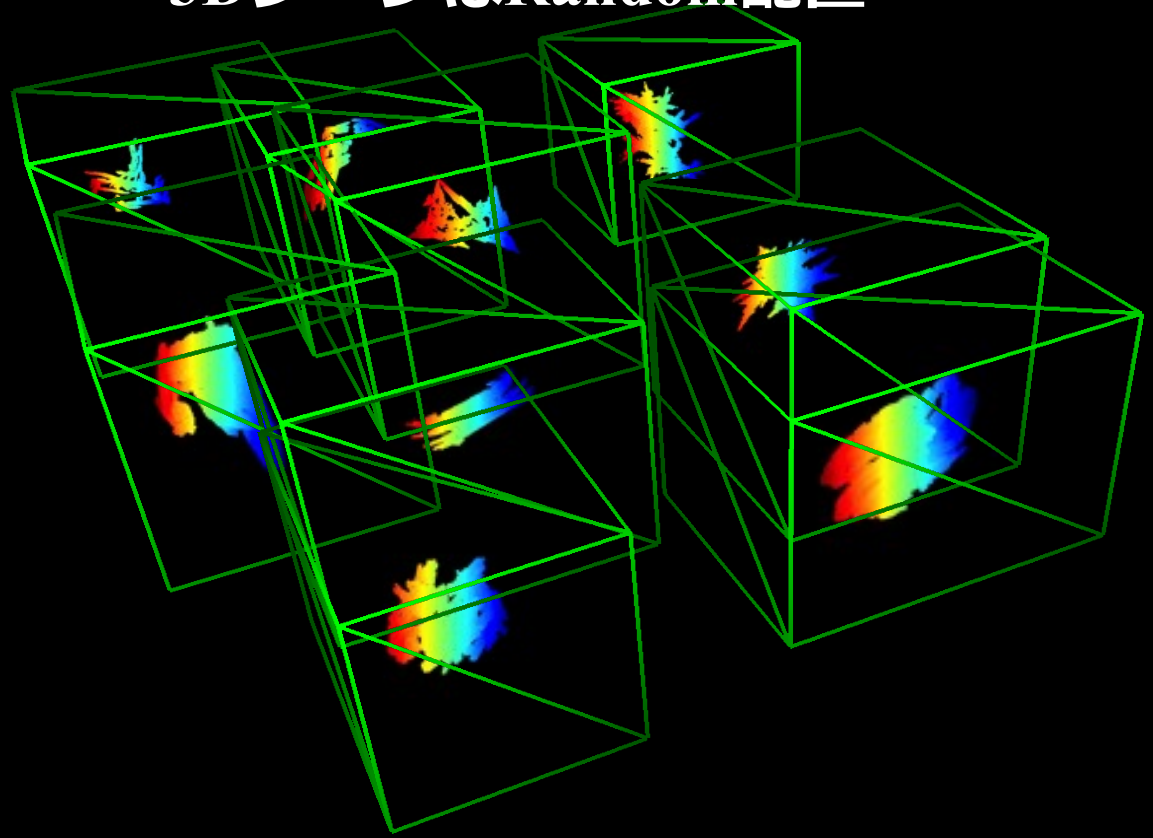
Point number: 4,000



Noise Category

Point number: 800

3DシーンはRandom配置



実験結果: 3D物体検出精度比較

ScanNetV2 / SUN RGB-Dによる比較

Pre-training	Backbone	Parameter	Input	ScanNetV2		SUN RGB-D	
				mAP@0.25	mAP@0.50	mAP@0.25	mAP@0.50
Scratch	PointNet++	0.95M	Geo + Height	57.9	32.1	57.4	32.8
Scratch	SR-UNet	38.2M	Geo	57.0	35.8	56.1	34.2
RandomRooms [51]	PointNet++	0.95M	Geo + Height	61.3	36.2	59.2	35.4
PointContrast [67]	SR-UNet	38.2M	Geo	59.2	38.0	57.5	34.8
CSC [26]	SR-UNet	38.2M	Geo	-	**39.3**	-	**36.4**
PC-FractalDB	PointNet++	0.95M	Geo + Height	**61.9**	38.3	**59.4**	33.9
PC-FractalDB	PointNet++ ×2	38.2M	Geo + Height	**63.4**	**39.9**	**60.2**	35.2
PC-FractalDB	SR-UNet	38.2M	Geo	59.4	37.0	57.1	**35.9**

Underlined bold: best score　**Baseline**　**Ours**

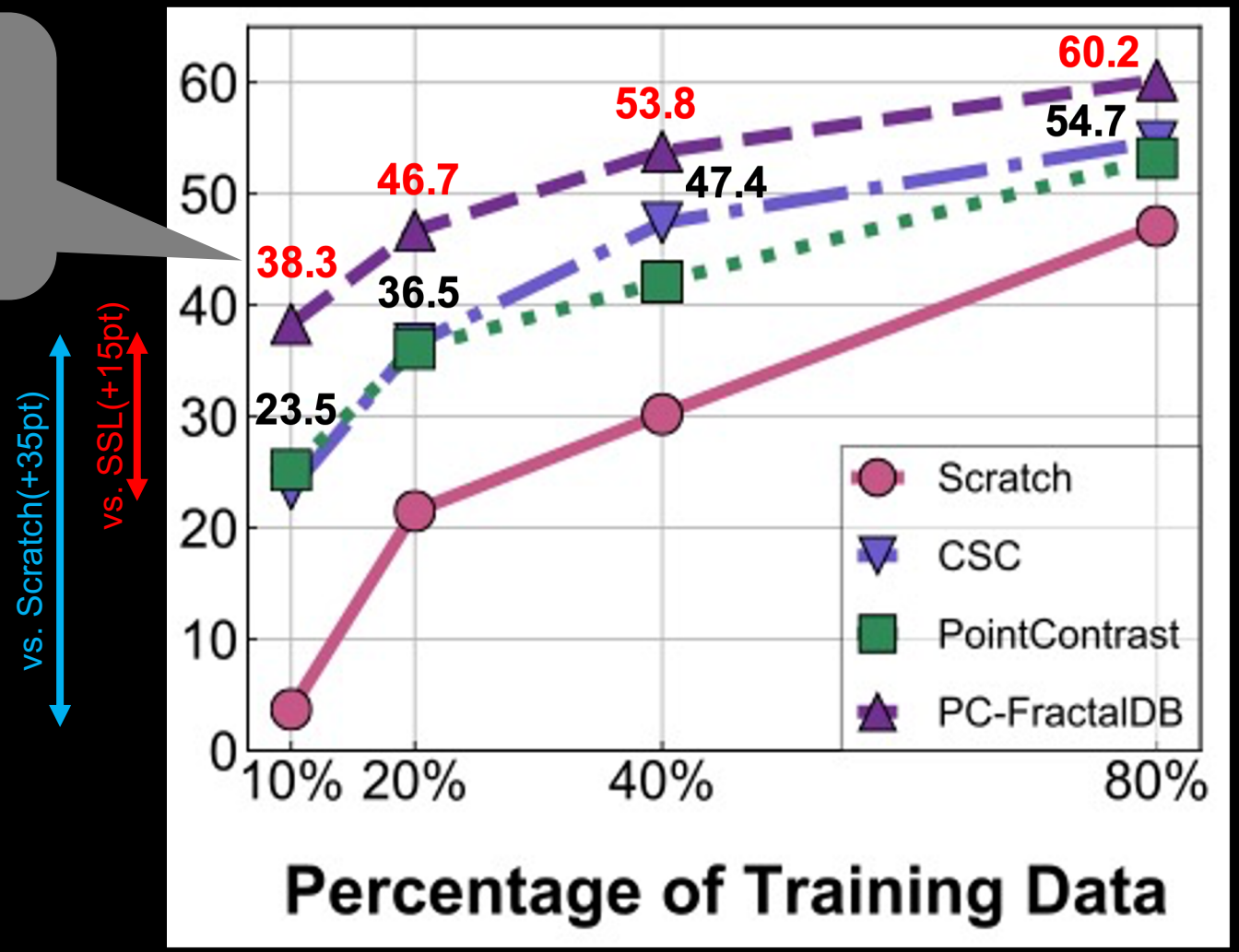
PC-FractalDB 61.9 vs 59.2 (PointContrast; ECCV 2020)
vs 61.3 (RandomRoom; ICCV 2021)

ScanNetV2 / mAP @ 0.25 により計測

実験結果: 少量データ&アノテーションにおける実験

限られたデータの学習においても高精度

データ量10%使用時：
SSL比較で約+15%
Scratch比較約+35%



少量の学習データに対する実験結果（mAP@0.25）

Hirokatsu Kataoka | 片岡裕雄
@HirokatuKataoka

我々の研究により「画像認識AIの事前学習（Pre-training）データを人間が集める時代は終わった」と言えるような世界にしていきたいですね！さらに、AI開発を加速させるべく、今回我々からは商用利用を制限する権利を設けておりません。

https://twitter.com/HirokatuKataoka/status/1536284511696490498