REPLACING LABELED REAL-IMAGE DATASETS WITH AUTO-GENERATED CONTOURS

Hirokatsu Kataoka^{*}, Ryo Hayamizu^{*}, Ryosuke Yamada^{*}, Kodai Nakashima^{*,**}, Xinyu Zhang^{*,**}, Edgar Josafat Martinez-Noriega^{*,**}, Nakamasa Inoue^{*,**}, Rio Yokota^{*,} National Institute of Advanced Industrial Science and Technology (AIST) *, Tokyo Institute of Technology**

We show that the performance of FDSL can match that of ImageNet-21k without the use of real images, human-/self-supervision during the pre-training of ViTs.

FORMULA-DRIVEN SUPERVISED LEARNING (FDSL)

- Real images come with privacy/copyright issues and societal biases
- Do we actually need real images to pre-train vision transformers?

To enhance the performance of FDSL, we test the following hypotheses 1 & 2

HYPOTHESIS 1: OBJECT CONTOURS ARE WHAT MATTER IN FDSL DATASETS

- In our preliminary study we found that attention was focused on the outer contours of the fractals
- We created a new dataset that consists only of contours Radial Contour DataBase (RCDB)
- Despite the lack of any texture, RCDB performed close to FractalDB and outperformed ImageNet-21k

HYPOTHESIS 2: INCREASED NUMBER OF PARAMETERS IN FDSL PRE-TRAINING

- We tested various synthetic datasets with varying complexity of
- For RCDB, we changed the number of polygons, radius, line widtl
- Complex images increases the difficulty of the pre-training task and I

Attention Image

(Pre-training)

COMPARISON: IMAGENET-1K / MSCOCO

Pre-training Image

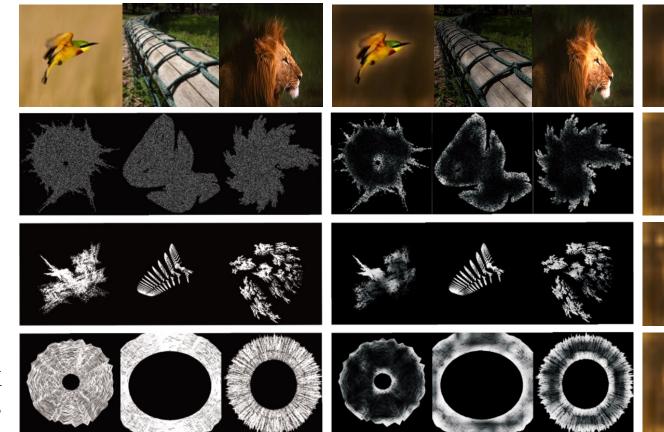
Pre-training

IN-21k ImageNet

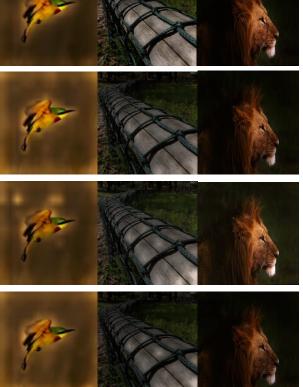
FDB-21k FractalDB

EFDB-211 ExFractalDB

RCDB-21 RadialCor



Attention Image (Fine-tuning on ImageNet-1k)



METHOD

- Classes are defined by parameters used to generate the images
- We generate 1k instances per class through transformations
- We use the same parameters as the DeiT paper
- Each of the large-scale runs were performed only once

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			
Perlin Noise Dead Leaves	Bezier Curr	ves RCDB		FractalDB			

of images	Pre-training	C10	C100	Cars	Flowers
th, resizing factor, and Perlin noise	BC RCDB		81.4 (1.1) 82.2 (0.6)		
d leads to better downstream performance		97.2 (0.4)			98.9 (0.6)

Fine-tuning	Pre-training	COCO Det	COCO
@ ImageNet-1k Top-1 w/ ViT-Base		AP_{50} / AP / AP_{75}	AP ₅₀ /
81.8	Scratch	63.7 / 42.2 / 46.1	60.7 / 3
	ImageNet-1k	69.2 / 48.2 / 53.0	66.6 / 4
81.8	ImageNet-21k	70.7 / 48.8 / 53.2	67.7 / 4
	ExFractalDB-1k	69.1 / 48.0 / 52.8	66.3 / 4
82.7(Ours1)	ExFractalDB-21k	69.2 / 48.0 / 52.6	66.4 / 4
$O \angle . / (Ours1)$	RCDB-1k	68.3 / 47.4 / 51.9	65.7/4
82.4 _(Ours2)	RCDB-21k	67.7 / 46.6 / 51.2	64.8 / 4
O <i>∠</i> .´★(0urs2)	Swin Transformer head	chana Maak D CNN haa	d 60 anas

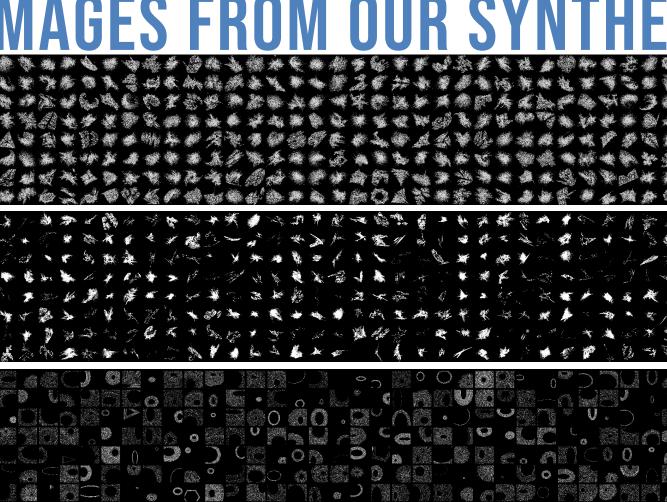
Swin Transformer backbone, Mask R-CNN head, 60 epochs fine-tuning

- CO Inst Seg / AP / AP₇₅ 38.5/41.3 43.1 / 46.5 43.6 / 47.0 **42.8** / 45.9
- 42.8 / 46.1
- 42.2 / 45.5
- 41.6 / 44.7

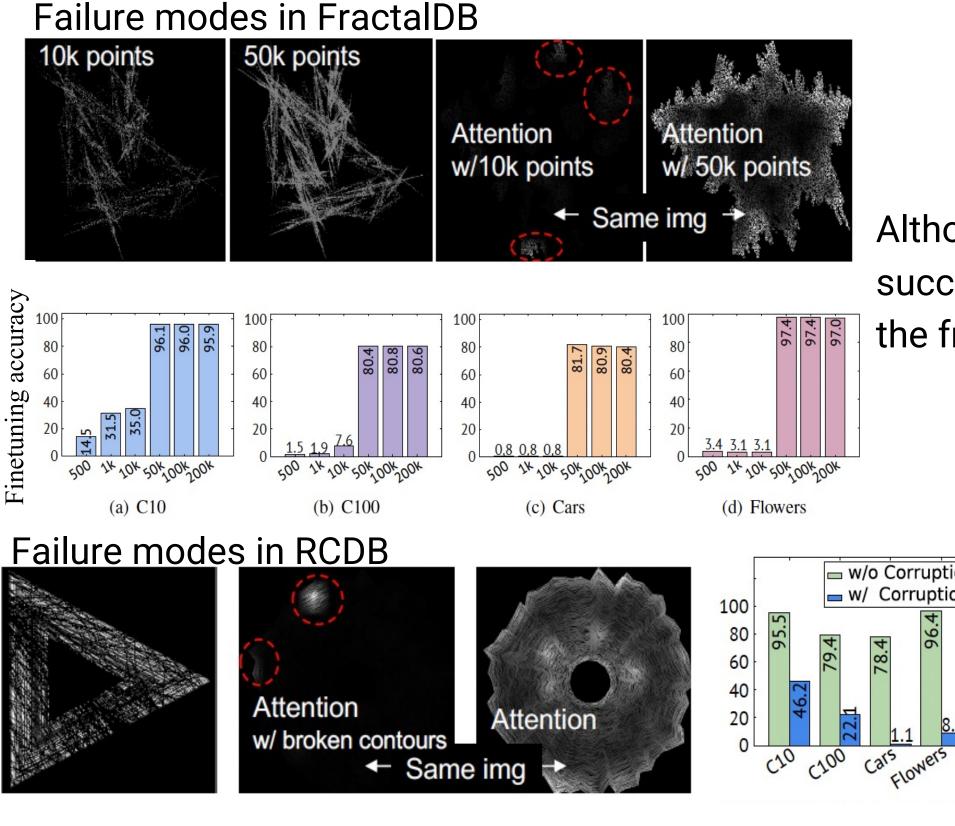
FDB-21 FractalDE

EFDB-21k ExFractalDB

RCDB-21k RadialContourD



FAILURE MODES OF FDSL



CONCLUSION

- We provided empirical evidence that support our two hypotheses Our proposed method can surpass the accuracy of a ViT pre-trained on ImageNet-21k • We performed ablation studies to identify failure modes of FDSL



Sec.	ALC: NO	1.0	uddie in a	1	and the	Mar No.2	Sec	100	1. A.		Bagen	Sec.		Stight.	10	100	100	1.1.1.1		3.000	1000	No.	and the second	and the second	6.65		Station .	10.05	2.4. 5	
			di la	1	1			X			E.F								18		×.					à.				
	No.				A		day.			H			ġ,	Sight	10	Wes-		N. Contraction				翑	÷.	1	\$			st i	C.	
		13 20			11. 1						á.		-					ALC: N	100		-				Â.	1. A.			÷.,	
		Mar	100	ük.		-		and the	-		<u>.</u>			14 14			÷.				il.	din.	14. -	alle.		3	No		1	
		and and a second se				and the second	sing. Shik	ine Alter	tale.	and the second		and and	in and a second	and and a second se	1980	anne. Ase	38	(Ball)					1			24 - 24 -	148 1990 -		and a second sec	
			and the			and the second	1996. -	1999 1999	ni dala National	itan.	1997). 1913	and the second	1	arte a	Sec.		4	No.	and and a second	din D		al anna	ange State	100 M		7X 1. 15	~	-	di s	1996 1916
1				100	7								artise 3		44				and and a second	S.		a the	and the second	Sec.			1990			
	24				1				£	See.		Ø								1	- M	X	No.					39 - 1		Ċ.
1 1 1 1 1	÷.	<u> </u>	de la	-		Sec. Sec.	3PA	de la constante da la constant	\$	X	sk.	Ser as	13	÷.	N.	E.	*		\$	¥	Sec.	τ¢ ²		×	- alter-	ng de la		.	,3% <	Sec.
11	\$	ý		¢.	×.	the c	1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1	đ		*	+	1		1 State		-states	in i	*		ALL A	der.	1	*	X	th.	N.		and the second second	¥	y
	in.		X	*	1	¥						ALC: NO.		1		N. CA	4		ity.		-			â.	7	N.	*		e. Sector	<i>#</i>
	×.	. • • •,																										ý.		
- The		£1	1	*	а. У.		- All	-	1	Ś.	1	No.	i, jan	1	×	X	<i>.</i>			-	"Magar	A.					а Х а	*	1	
		rester		ця́.	17 201	1. 1.	2 mil	4	17 1 - 2		*	- Sec	An		i.	il.	at and	چە ت			3 2		3 4	1	1		-	r en	/` 減	
-	2000 815		1	*		*	-39 K -57	1	4					1	÷. +	Yest.	19 e		~	V	29))) 	v	- 1 <u>6</u>		100 M	and a	1		
12			- 1 980-	.see	ر. مغن	4			*								· · · · · · · · · · · · · · · · · · ·	e e fe		19 June -	1	A.C.	1	-	4	100	-205		15 A 16	
4	₹-	Ū.	A	2				S.	ų.				1	and a f	\$r	F	2.5					1		A.	88	A.C.	5	Х	<i>™</i> ,	a fir
	0	\bigcirc			$\sum_{i=1}^{n}$	С	0		\frown	annage (ann a	and the second	Heter			0	5923 ₀	($ \zeta_{im}^{\mu\nu\nu}\rangle$	*	Ś		\frown					n i	3
								ÍN.	Land				6 6			Ĵ		Ô	<i>C</i>					, 1924			delenatorio della		Û,	ñ
and the		6.285 8)	enan Vissi	No.				V						त्यामाळा ुर्द्धाः			105	Con Star					ner Here		diseased		0		
i AGREN				C.	8				0		Series	al and			11	an a					aand 19 Beers	, M J ^{alia} i	Seede A			0		Ĉ.		
Sec.20		1010000400	1362	193				888° 1		100	5%	0	42			28	10000	10.00	1					Constant of	100			100 1	participation of	
					Ш., "	0	67							38. 1978:		25M 97					0) 德 麗			$\frac{1}{2}$	A.	denselle Marine de la constance	11 J 11 J
			5 .7.95		- 		C .	and Social			1993 1993				S		L.		and the second s	330 (399)		e stal	(B)44		0,0	200 - 200 200 - 200		0		
																														en and de la constante Secondo
	E. Carlos	E.		186.2	- Section of the sect			See As	00000000	Cash		and a second		19 A	100	ALE FORMER				New York		Q		8			See.			

Results, image examples, and attention maps in point-rendered FractalDB and RCDB with corruption

Although the fractal images with 50k⁺ points successfully trained the visual representations, the fractal images with 10k⁻ points failed

- Example of RCDB with broken contours
 - Attention maps with the pre-trained models on RCDB w/ and w/o broken contours, respectively