

---

# RecombiNet: A dataset for better ImageNet

---

Tobias Christian Nauen<sup>1,2</sup> Brian Moser<sup>2</sup> Federico Raue<sup>2</sup> Stansilav Frolov<sup>2</sup> Andreas Dengel<sup>1,2</sup>

## Abstract

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi. Lorem ipsum dolor sit amet.

ImageNet (?)

## 1. Experiments

- Train on all variants, then evaluate on all variants.
- Train different models... ViT, ResNet, Swin, ? at multiple sizes.

---

<sup>1</sup>Department of Computer Science, RPTU Kaiserslautern-Landau, Kaiserslautern, Germany <sup>2</sup>German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany. Correspondence to: Tobias Christian Nauen <tobias.christian.nauen@dfki.de>.

*Proceedings of the 41<sup>st</sup> International Conference on Machine Learning*, Vienna, Austria. PMLR 235, 2024. Copyright 2024 by the author(s).

- Finetune for downstream tasks (classification).
- Finetune for dense tasks / semantic segmentation.
- Evaluate Diversity and Affinity from (?).
- Intermediate bg-fg combinations between only *same* and *all*.
- Train with different data augmentation setups on top. 3-Augment, AutoAugment, Real Guidance. With special focus on operations like Mixup & CutMix.
- Look into filtering foregrounds and backgrounds based on segmentation quality + size...

## 2. Training on RecombinationNet

## 3. Evaluating on RecombinationNet

## References

Segment & Recombine

Model	Mode	Foreground Rotation	ImageNet Accuracy [%]	Recombine same Accuracy [%]	Recombine all Accuracy [%]	Backgrounds Accuracy [%]
ViT-S/16	ImageNet		79.73	68.2	50.6	15.7
ViT-S/16	same	$\pm 0$	82.3	82.0	67.9	27.2
ViT-S/16	same	$\pm 10$	82.2	81.9	67.9	27.3
ViT-S/16	same	$\pm 20$	82.2	82.0	67.9	27.0
ViT-S/16	same	$\pm 30$	82.2	82.1	67.8	27.0
ViT-S/16	all	$\pm 0$	76.8	76.4	76.5	03.1
ViT-S/16	all	$\pm 20$				
ViT-S/16	same					
ViT-S/16	same					
ViT-S/16	same					
ViT-S/16	same					

Table 1: Training on RecombinationNet v24-10 (300 ep), evaluating on ImageNet.

Model	Train DS	ImageNet Accuracy [%]	Recombine same Accuracy [%]	Recombine all Accuracy [%]	Backgrounds Accuracy [%]
ViT-Ti/16	IN1k	76.1	64.5/67.3	47.3	12.8
ViT-Ti/16	RN same	64.7	75.4	59.4	19.5
ViT-Ti/16	RN all	53.5	70.6	70.6	03.2
ViT-S/16	IN1k	79.6			
ViT-S/16	RN same	69.7	82.0	67.9	27.0
ViT-S/16	RN all	50.1	76.6	76.4	02.9
ViT-B/16	IN1k	78.0	65.9/68.8	48.2	16.1
ViT-B/16	RN same	70.1	81.5	70.7	30.9
ViT-B/16	RN all	44.9	75.9	76.0	02.5
Swin-Ti	IN1k	77.9	66.5/69.2	48.9	15.6
Swin-Ti	RN same	63.6	81.4	70.1	29.1
Swin-Ti	RN all	09.4	76.9	75.7	00.6
Swin-S	IN1k	79.7	67.9/70.9	50.5	16.8
Swin-S	RN same	65.0	82.5	72.4	34.1
Swin-S	RN all	FAIL			
ResNet-34	IN1k	75.3	65.9/67.8	51.7	11.8
ResNet-34	RN same	63.5	75.7	63.9	17.9
ResNet-34	RN all	37.2	72.0	71.4	02.0
ResNet-50	IN1k	78.4	69.2/71.2	55.9	13.9
ResNet-50	RN same	66.8	80.3	70.4	23.5
ResNet-50	RN all	21.4	75.6	75.6	01.1
ResNet-101	IN1k	79.5	70.2/72.1	57.1	14.8
ResNet-101	RN same	68.9	81.8	72.7	27.3
ResNet-101	RN all	24.0	76.9	76.8	01.5

Table 2: Training on RecombinationNet v24-10? (300 ep), evaluating on ImageNet.

Version	Version ID	ViT-Ti			ResNet34		
		TIN Acc. [%]	TRN same max Acc. [%]	TRN same max pt=.6 Acc. [%]	TIN Acc. [%]	TRN same max Acc. [%]	TRN same max pt=.6 Acc. [%]
24-01-10	TINS	66.7	60.1	60.7	77.9	71.4	73.2
25-01-13	TINS_v2	66.7	61.0	62.4	77.9	72.2	74.2
25-01-17 man	TINS_v3_f1	66.7	61.8	61.9	77.9	73.5	74.2
25-01-17 auto	TINS_v3_f2	66.7	61.4	61.7	77.9	73.2	74.1
25-01-24 first	TINS_v4_f1	66.7	62.7	62.7	77.9	73.6	74.9
25-01-24 all man	TINS_v5_f1	66.7	61.9	62.3	77.9	73.7	74.5
25-01-24 all auto	TINS_v5_f2	66.7	61.5	62.2	77.9	73.8	74.4
25-02-03 all man	TINS_v6_f1	66.7	61.3	62.1	77.9	73.1	75.0
25-02-04 all man	TINS_v7_f1	66.7			77.9		

Table 3: Training on TinyImageNet and evaluating on different TinyRecombNet versions.

Segment & Recombine

Version	Train DS	ViT-Ti/16	ViT-S/16	ResNet34	ResNet50	ResNet101
		TIN Acc. [%]	TIN Acc. [%]	TIN Acc. [%]	TRN same max Acc. [%]	TRN same max pt=.6 Acc. [%]
v25-01-24 all-man TINS_v5_f1	TIN	66.7/65.7/65.9	68.9	77.9/77.9/78.1	79.1	79.8
	TRN same/range	<b>66.7</b> /67.2/66.9	<b>72.3</b>	75.4		
	TRN same/range pt=1.0	<b>67.5</b> /66.6	<b>71.9</b>	76.1		
	TRN same/range pt=0.8	65.8(?) /66.2	<b>70.8</b>	75.9		
	TRN same/range p→t	<b>66.7</b> /66.6	<b>72.1</b> /71.6/71.8	75.6		
	TRN same/range pt=1.0 p→t	<b>66.9</b> /66.5/66.5	<b>72.7</b> /73.1	75.8		
	TRN same/range pt=0.9 p→t	<b>66.9</b>	<b>73.2</b>			
	TRN same/range pt=0.8 p→t	<b>67.8</b> /64.7/67.1	<b>73.5</b> /70.1/72.2	75.9		
	TRN same/range pt=0.7 p→t	66.6/66.3	71.7			
	TRN same/range pt=0.6 p→t	<b>66.9</b>	<b>71.1</b>	76.0		
	TRN same/range pt=0.5 p→t	66.4	<b>72.0</b>	76.3		
	TRN same/range pt=0.5 p→t	66.6	<b>69.3</b>	75.3		
	TRN all/range	56.6	62.2	63.3		
	v25-02-03 all-man TINS_v6_f1	TIN	66.7	68.9	77.9	
TRN same/range		<b>67.3</b>	<b>72.9</b>	76.0		
TRN same/range pt=1.0		<b>67.2</b>	<b>73.6</b>	76.2		
TRN same/range pt=0.8		66.6	<b>73.1</b>	75.5		
TRN same/range p→t			<b>73.6</b> / <b>74.2</b>	75.8		
TRN same/range pt=1.0 p→t		<b>67.4</b>	<b>74.2</b>	76.0		
TRN same/range pt=0.9 p→t		<b>67.4</b>	<b>74.1</b>	76.0		
TRN same/range pt=0.8 p→t		<b>68.4</b>	<b>72.4</b>	76.0		
TRN same/range pt=0.7 p→t			<b>72.2</b>	75.7		
TRN same/range p→t ed blr 2		66.3	<b>72.3</b>	72.5		
TRN same/range p→t ed blr 4	65.3	<b>69.8</b>	74.5			
v25-02-04 all-man RETEST TINS_v7_f1	TIN	66.7	68.9	77.9		
	TRN same/range	68.7/66.4/67.5	72.1/73.0/72.2			
	TRN same/range	67.5 ± 1.2	72.4 ± 0.5			
	TRN same/range pt=1.0	67.1/66.5/66.6	72.8/72.4/72.5			
	TRN same/range pt=1.0	66.7 ± 0.2	72.6 ± 0.1			
	TRN same/range pt=0.8	66.1/66.1/66.3	72.6/71.4/71.5			
	TRN same/range pt=0.8	66.2 ± 0.1	71.8 ± 0.9			
	TRN same/range p→t	68.5/66.1/66.8	73.0/72.3/73.3			
	TRN same/range p→t	67.1 ± 1.2	72.9 ± 0.5			
	TRN same/range pt=1.0 p→t	67.6/66.1/67.2	72.5/73.1/73.3			
	TRN same/range pt=1.0 p→t	67.0 ± 1.2	73.0 ± 0.3			
	TRN same/range pt=0.8 p→t	67.5/67.5/66.5	72.2/73.3/73.3			
	TRN same/range pt=0.8 p→t	67.2 ± 0.7	72.9 ± 0.8			
	TRN s/r p→t pt=0.8 edgb=2	67.6				
	TRN s/r p→t pt=0.8 edgb=4	65.8				
	TRN s/r p→t pt=0.8 orig p=0.1	69.4				
	TRN s/r p→t pt=0.8 orig p=0.2	70.4				
	TRN s/r p→t pt=0.8 orig p=0.33	70.0	74.1			
	TRN s/r p→t pt=0.8 orig p=0.5	70.0				
	TRN s/r p→t pt=0.8 orig p=lin	69.5				
TRN s/r p→t pt=0.8 orig p=invlin	67.5					
TRN s/r p→t pt=0.8 orig p=cos	71.3					
TRN s/r p→t pt=0.8 orig p=0.5 edgb=2	69.5					
TRN s/r p→t pt=0.8 orig p=0.5 edgb=4	70.6					
TRN s/r p→t pt=0.8 orig p=cos edgb=2						
TRN s/r p→t pt=0.8 orig p=cos edgb=4						

Table 4: Training on TinyRecombinationNet (300 ep), evaluating on ImageNet. TIN=TinyImageNet, TRN=TinyRecombNet, TBN=TinyBackgroundNet. Versions are for train and test sets. v25-02-03 is comparable to v25-01-17, v25-02-04 is comparable to v25-01-24.

Model	Test Dataset → Train Dataset ↓	TinyRecombNet-5-1/same/range				TinyRecombNet-5-1/all/range			
		IG	GradCAM	GradCAM++	Attn	IG	GradCAM	GradCAM++	Attn
ViT-Ti/16	TIN	1.07	1.56	1.83	1.72	1.12	2.20	2.21	1.78
	TRN-5-1/same	1.16	2.00	2.17	1.61	1.31	2.76	2.37	1.62
	TRN-5-1/all	1.51	2.90	2.69	2.37	1.56	3.00	2.74	2.33
ViT-S/16	TIN	1.10	1.59	1.77	1.47	1.15	2.25	2.26	1.52
	TRN-5-1/same	1.18	1.67	1.69	1.38	1.40	2.68	2.46	1.49
	TRN-5-1/all	1.40	2.56	2.70	1.81	1.40	2.65	2.76	1.68
ResNet34	TIN	1.40	1.89	1.71		1.49	2.25	1.83	
	TRN-5-1/same	1.41	1.97	1.86		1.70	2.52	2.07	
	TRN-5-1/all	2.19	2.67	2.64		1.98	2.64	2.60	

Table 5: Relative foreground importance ratio for different models and training datasets. We calculate the per-pixel importance for the foreground label class. Then we aggregate the results for the foreground and background regions. The relative foreground importance is the ratio of the foreground importance divided by the ratio of the foreground size: Rel fg importance :=  $\frac{\text{importance in fg region}}{\text{total importance}} / \frac{\text{fg size}}{\text{total image size}}$ . We average the results on the whole validation set. Training and evaluation were done without background pruning with range fg insertion mode. Using TinyRecombNet v25-01-24.

**TL;DR:** Our dataset/data augmentation improves the focus on the foreground class. Maybe it does not really work with ResNet, because that already has a good focus on the foreground object.

Model	Test Dataset → Train Dataset ↓	TIN	RecombNet-5-1/same/range				RecombNet-5-1/all/range			
			$\eta = 1$	$\eta = 2$	$\eta = 3$	$\eta = 4$	$\eta = 1$	$\eta = 2$	$\eta = 3$	$\eta = 4$
ViT-Ti/16	TIN	66.7	58.7	59.6	60.1	60.2	34.3	34.4	35.3	35.4
	TRN-5-1/same $\eta = -3$	66.0								
	TRN-5-1/same $\eta = -2$	66.8								
	TRN-5-1/same $\eta = 1$	66.7	74.0	74.7	74.8	75.3	49.4	49.1	49.8	49.9
	TRN-5-1/same $\eta = 2$	66.0	74.0	74.9	74.8	74.6	49.6	50.3	50.6	51.4
	TRN-5-1/same $\eta = 3$	64.5	71.8	74.0	74.3	74.3	48.0	48.9	48.7	50.1
ViT-S/16	TIN	68.9	60.8	62.6	63.1	62.5	36.6	37.9	38.7	39.1
	TRN-5-1/same $\eta = -3$	71.3								
	TRN-5-1/same $\eta = -2$	71.5								
	TRN-5-1/same $\eta = 1$	72.3	80.2	79.7	80.4	80.1	56.4	57.5	57.2	57.2
	TRN-5-1/same $\eta = 2$	71.3	79.1	79.6	79.9	80.1	58.2	57.3	58.2	58.2
	TRN-5-1/same $\eta = 3$	71.4	78.6	79.6	79.7	80.1	55.6	57.5	57.9	57.9
ResNet34	TIN	77.9	72.2	72.7	72.7	73.2	54.2	54.6	54.2	54.7
	TRN-5-1/same $\eta = 1$	75.4	83.4	83.7	83.2	83.4	69.2	69.6	68.6	68.81
	TRN-5-1/same $\eta = 2$	76.0	83.1	83.3	83.4	83.7	68.5	69.1	69.2	69.3
	TRN-5-1/same $\eta = 3$	75.8	83.1	83.1	84.0	83.5	67.3	67.8	69.2	69.0

Table 6: Importance of foreground centering via bates distribution.  $\eta$  is the parameter of the bates distribution.  $\eta = 1$  is the uniform distribution. Training and evaluation were done without background pruning with range fg insertion mode.

**TL;DR:** Focussing the foreground object in the center makes the task easier (increasing performance left to right), but this then does not aid training (decreasing performance top to bottom)  $\Rightarrow \eta = 1$  is the optimum (for training).

Augmentation Policy	Dataset	TinyImageNet Accuracy [%]	
		ViT-Ti	ResNet34
minimal w/o cutmix	TinyImageNet	37.7	69.1
	TinyRecombNet-3-2	50.7	69.1
	TinyRecombNet-3-1	51.0	69.2
	TinyDoublecombNet-3-2	44.9	68.6 +
	TinyDoublecombNet-3-1	46.7	68.6
minimal w/ cutmix	TinyImageNet	57.7	<b>73.0</b>
	TinyRecombNet-3-2	<b>60.6</b>	70.3
	TinyRecombNet-3-1	59.8	70.2
3-augment w/o cutmix	TinyImageNet	45.3	71.6
	TinyRecombNet-3-2	<b>59.9</b>	<b>74.4</b>
	TinyRecombNet-3-1	59.8	74.3
3-augment w/ cutmix	TinyImageNet	<b>65.9</b>	<b>78.0</b>
	TinyRecombNet-3-2	65.4	76.0
	TinyRecombNet-3-1	<b>65.9</b>	75.3

Table 7: Training on TinyRecombinationNet (300 ep), evaluating on ImageNet with different data augmentation setups. TinyRecombNet version is v25-01-17-man/auto.

Model	Eval DS	IN1k Baseline	No pruning	$pt = 1.0$	$pt = 0.8$	$pt = 0.6$	$pt = 0.4$	$pt = 0.3$	$pt = 0.2$
ViT-B/16	RN same	78.0	68.8	68.9	69.4	70.2	70.7	71.2	71.2
ViT-B/16	Backgrounds		16.1	16.1	17.9	13.1	15.1		
ResNet-101	RN same	79.5	72.2	72.3	73.0	73.1	73.8	73.8	74.4
ResNet-101	Backgrounds		14.8	14.8	16.4	11.6	13.4		

Table 8: Training on ImageNet and evaluating with different background prune thresholds. **RecombNet v25-01-10.**