A Reproducibility

Our code will be openly available at [IPLVEs]. Our code will be made publicly available upon publication. We ran our experiments on 2 A100s 40G. Our implementation is based on PyTorch v.1.12.0 (Paszke et al., 2019) and Transformer v4.25.1 (Wolf et al., 2020) for Python 3.9.13 and builds on code from the repositories in Table 6.

Models	Params	Datasets	Links
BERT	4.4M (Tiny) 11.3M (Mini) 29.1M (Small) 41.7M (Medium) 110.1M (Base) 334M (Large)	BooksCorpus; English Wikipedia	https://huggingface.co/google/bert_uncased_L-2_H-128_A-2 https://huggingface.co/google/bert_uncased_L-4_H-256_A-4 https://huggingface.co/google/bert_uncased_L-4_H-512_A-8 https://huggingface.co/google/bert_uncased_L-8_H-512_A-8 https://huggingface.co/bert-base-uncased https://huggingface.co/bert-large-uncased
GPT2	117M (Base) 762M (Large) 1542M (XL)	WebText	https://huggingface.co/gpt2, https://huggingface.co/gpt2-large, https://huggingface.co/gpt2-xl
OPT	125M 6.7B 30B	BooksCorpus; CC-Stories; CCNewsV2; The Pile; Pushshift.io Reddit dataset	https://huggingface.co/facebook/opt-125m, https://huggingface.co/facebook/opt-6.7b, https://huggingface.co/facebook/opt-30b
LLaMA-2	7B 13B	2 trillion tokens	https://huggingface.co/meta-llama/Llama-2-7b, https://huggingface.co/meta-llama/Llama-2-13b
SegFormer-B0	3.4M		https://huggingface.co/nvidia/segformer-b0-finetuned-ade-512-512,
SegFormer-B1	13.1M		https://huggingface.co/nvidia/segformer-b1-finetuned-ade-512-512,
SegFormer-B2	24.2M	Pretrained with ImageNet-1K	https://huggingface.co/nvidia/segformer-b2-finetuned-ade-512-512,
SegFormer-B3	44.0M	Finetuned with ADE20K	https://huggingface.co/nvidia/segformer-b3-finetuned-ade-512-512,
SegFormer-B4	60.8M		https://huggingface.co/nvidia/segformer-b4-finetuned-ade-512-512,
SegFormer-B5	81.4M		https://huggingface.co/nvidia/segformer-b5-finetuned-ade-640-640
MAE_{Base}	86M		https://huggingface.co/facebook/vit-mae-base,
MAE_{Large}	304M	ImageNet-1K	https://huggingface.co/facebook/vit-mae-large,
MAE_{HUGE}	632M		https://huggingface.co/facebook/vit-mae-huge
ResNet18	11.7M		
ResNet34	21.8M		
ResNet50	25.6M	ImageNet-1K	https://pypi.org/project/img2vec-pytorch/
ResNet101	44.6M		
ResNet152	60.2M		

Table 6: Details of all models used in our experiments.

B More Results

LM	VM	P@1	P@10	P@100
BERT_BASE	MAE-Base	2.0	11.9	34.6
BERT_BASE	MAE-Huge	2.1	12.6	35.1
BERT_BASE	MAE-Large	2.2	12.5	35.5
BERT_BASE	ResNet101	3.1	17.6	43.6
BERT_BASE	ResNet152	3.2	17.9	44.0
BERT_BASE	ResNet18	2.5	15.4	40.5
BERT_BASE	ResNet34	2.7	16.1	41.8
BERT_BASE	ResNet50	3.0	17.2	42.8
BERT_BASE	SegFormer-B0	2.2	10.9	32.9
BERT_BASE	SegFormer-B1	2.5	14.2	38.2
BERT_BASE	SegFormer-B2	3.1	15.6	40.3
BERT_BASE	SegFormer-B3	3.0	16.1	41.4
BERT_BASE	SegFormer-B4	2.8	15.5	40.5
BERT_BASE	SegFormer-B5	2.7	15.6	40.6
BERT_LARGE	MAE-Base	2.3	12.2	36.9
BERT_LARGE	MAE-Huge	2.5	13.3	37.6

Table 7 continued from previous page

LM	VM	P@1	P@10	P@100
BERT_LARGE	MAE-Large	2.2	13.2	37.8
BERT_LARGE	ResNet101	3.4	18.6	45.6
BERT_LARGE	ResNet152	3.4	18.5	45.9
BERT_LARGE	ResNet18	2.8	16.2	41.5
BERT_LARGE	ResNet34	3.2	16.9	43.4
BERT_LARGE	ResNet50	3.2	18.0	44.7
BERT_LARGE	SegFormer-B0	2.5	11.5	34.3
BERT_LARGE	SegFormer-B1	2.9	15.4	40.9
BERT_LARGE	SegFormer-B2	3.1	16.4	42.4
BERT_LARGE	SegFormer-B3	3.2	16.7	43.2
BERT_LARGE	SegFormer-B4	3.1	16.5	42.6
BERT_LARGE	SegFormer-B5	2.9	16.2	42.4
BERT_MEDIUM	MAE-Base	1.7	9.4	26.2
BERT_MEDIUM	MAE-Huge	1.6	9.4	25.8
BERT_MEDIUM	MAE-Large	1.6	9.3	25.7
BERT_MEDIUM	ResNet101	2.6	13.6	33.9
BERT_MEDIUM	ResNet152	2.5	13.6	34.0
BERT_MEDIUM	ResNet18	2.2	12.0	31.5
BERT_MEDIUM	ResNet34	2.5	12.9	32.5
BERT_MEDIUM	ResNet50	2.3	13.2	33.1
BERT_MEDIUM	SegFormer-B0	1.5	8.5	25.5
BERT_MEDIUM	SegFormer-B1	2.0	11.2	29.9
BERT_MEDIUM	SegFormer-B2	2.2	13.0	31.5
BERT_MEDIUM	SegFormer-B3	2.1	12.6	31.9
BERT_MEDIUM	SegFormer-B4	2.0	12.5	31.1
BERT_MEDIUM	SegFormer-B5	1.9	12.3	31.2
BERT_MINI	MAE-Base	0.9	6.1	19.6
BERT_MINI	MAE-Huge	0.9	5.8	18.2
BERT_MINI	MAE-Large	0.9	5.6	19.2
BERT_MINI	ResNet101	1.4	9.0	25.1
BERT_MINI	ResNet152	1.5	9.1	25.4
BERT_MINI	ResNet18	1.3	8.2	23.4
BERT_MINI	ResNet34	1.4	8.7	24.2
BERT_MINI	ResNet50	1.4	9.0	24.8
BERT_MINI	SegFormer-B0	1.1	6.5	20.7
BERT_MINI	SegFormer-B1	1.1	7.3	21.9
BERT_MINI	SegFormer-B2	1.4	8.4	23.4
BERT_MINI	SegFormer-B3	1.3	7.9	23.6
BERT_MINI	SegFormer-B4	1.2	7.5	23.1
BERT_MINI	SegFormer-B5	1.3	7.8	22.7
BERT_SMALL	MAE-Base	1.3	8.6	25.2
BERT_SMALL	MAE-Huge	1.1	8.6	24.9
BERT_SMALL	MAE-Large	1.1	8.8	25.0
BERT_SMALL	ResNet101	2.1	13.0	32.2
BERT_SMALL	ResNet152	2.1	12.9	32.5
BERT_SMALL	ResNet18	1.8	11.7	30.2
BERT_SMALL	ResNet34	2.1	12.3	31.7
BERT_SMALL	ResNet50	2.0	12.7	32.0
BERT_SMALL	SegFormer-B0	1.3	8.1	24.7

Table 7 continued from previous page

LM	VM	P@1	P@10	P@100
BERT_SMALL	SegFormer-B1	1.6	10.5	28.7
BERT_SMALL	SegFormer-B2	1.8	12.4	30.4
BERT_SMALL	SegFormer-B3	1.8	12.3	30.6
BERT_SMALL	SegFormer-B4	1.7	12.0	30.4
BERT_SMALL	SegFormer-B5	1.5	11.8	29.8
BERT_TINY	MAE-Base	0.4	2.8	12.2
BERT_TINY	MAE-Huge	0.5	2.6	11.4
BERT_TINY	MAE-Large	0.4	2.8	11.9
BERT_TINY	ResNet101	0.9	5.1	17.8
BERT_TINY	ResNet152	1.0	5.2	17.8
BERT_TINY	ResNet18	0.8	4.4	16.3
BERT_TINY	ResNet34	1.0	4.7	17.2
BERT_TINY	ResNet50	1.0	4.7	17.6
BERT_TINY	SegFormer-B0	0.5	3.8	14.1
BERT_TINY	SegFormer-B1	0.6	3.6	14.5
BERT_TINY	SegFormer-B2	0.7	4.1	15.5
BERT_TINY	SegFormer-B3	0.8	4.2	15.4
BERT_TINY	SegFormer-B4	0.6	4.1	14.6
BERT_TINY	SegFormer-B5	0.7	3.9	14.9
GPT2_BASE	MAE-Base	0.3	1.5	5.2
GPT2_BASE	MAE-Huge	0.3	1.3	4.6
GPT2_BASE	MAE-Large	0.3	1.5	5.0
GPT2_BASE	ResNet101	0.6	2.8	9.4
GPT2_BASE	ResNet152	0.6	2.9	9.4
GPT2_BASE	ResNet18	0.4	2.3	8.0
GPT2_BASE	ResNet34	0.4	2.6	8.8
GPT2_BASE	ResNet50	0.5	2.7	9.0
GPT2_BASE	SegFormer-B0	0.3	1.4	5.8
GPT2_BASE	SegFormer-B1	0.4	2.5	8.5
GPT2_BASE	SegFormer-B2	0.6	2.9	8.9
GPT2_BASE	SegFormer-B3	0.6	2.9	9.6
GPT2_BASE	SegFormer-B4	0.6	3.2	9.5
GPT2_BASE	SegFormer-B5	0.5	2.9	8.3
GPT2_LARGE	MAE-Base	2.2	12.1	33.5
GPT2_LARGE	MAE-Huge	2.1	13.4	34.3
GPT2_LARGE	MAE-Large	2.3	13.6	34.7
GPT2_LARGE	ResNet101	3.3	17.4	43.2
GPT2_LARGE	ResNet152	3.2	17.8	43.8
GPT2_LARGE	ResNet18	2.9	14.5	38.6
GPT2_LARGE	ResNet34	3.1	15.4	40.2
GPT2_LARGE	ResNet50	3.3	17.4	42.7
GPT2_LARGE	SegFormer-B0	1.9	9.7	29.8
GPT2_LARGE	SegFormer-B1	2.5	13.9	37.0
GPT2_LARGE	SegFormer-B2	2.8	15.0	38.7
GPT2_LARGE	SegFormer-B3	2.7	15.8	40.1
GPT2_LARGE	SegFormer-B4	2.7	15.1	39.4
GPT2_LARGE	SegFormer-B5	2.5	14.9	38.7
GPT2_XL	MAE-Base	2.2	13.2	36.1
GPT2_XL	MAE-Huge	2.3	14.8	37.4

Table 7 continued from previous page

LM	VM	P@1	P@10	P@100
GPT2_XL	MAE-Large	2.5	14.7	37.4
GPT2_XL	ResNet101	3.2	18.7	46.4
GPT2_XL	ResNet152	3.2	19.1	46.6
GPT2_XL	ResNet18	3.1	15.4	41.2
GPT2_XL	ResNet34	3.3	16.6	42.9
GPT2_XL	ResNet50	3.3	18.4	45.8
GPT2_XL	SegFormer-B0	2.0	10.6	31.3
GPT2_XL	SegFormer-B1	3.0	14.9	39.4
GPT2_XL	SegFormer-B2	2.8	16.0	41.1
GPT2_XL	SegFormer-B3	2.9	16.6	42.7
GPT2_XL	SegFormer-B4	3.1	16.7	41.9
GPT2_XL	SegFormer-B5	2.8	16.0	41.5
Llama-2-13b	MAE-Base	2.9	14.6	38.0
Llama-2-13b	MAE-Huge	3.1	17.0	40.6
Llama-2-13b	MAE-Large	3.1	16.2	40.5
Llama-2-13b	ResNet101	4.3	22.6	53.0
Llama-2-13b	ResNet152	4.3	22.9	53.5
Llama-2-13b	ResNet18	3.3	16.7	42.4
Llama-2-13b	ResNet34	3.5	18.0	44.3
Llama-2-13b	ResNet50	4.0	22.3	52.1
Llama-2-13b	SegFormer-B0	1.9	9.8	29.4
Llama-2-13b	SegFormer-B1	3.1	15.7	42.0
Llama-2-13b	SegFormer-B2	3.2	17.2	43.2
Llama-2-13b	SegFormer-B3	3.7	18.2	45.8
Llama-2-13b	SegFormer-B4	3.4	17.6	45.0
Llama-2-13b	SegFormer-B5	3.3	17.2	44.1
Llama-2-7b	MAE-Base	2.2	10.3	28.9
Llama-2-7b	MAE-Huge	2.5	12.0	29.5
Llama-2-7b	MAE-Large	2.4	11.7	30.4
Llama-2-7b	ResNet101	3.3	18.7	41.4
Llama-2-7b	ResNet152	3.3	18.9	42.3
Llama-2-7b	ResNet18	2.7	13.4	34.0
Llama-2-7b	ResNet34	3.0	14.3	35.4
Llama-2-7b	ResNet50	3.4	18.6	40.9
Llama-2-7b	SegFormer-B0	1.5	7.9	22.9
Llama-2-7b	SegFormer-B1	2.7	12.7	32.6
Llama-2-7b	SegFormer-B2	2.8	13.8	34.7
Llama-2-7b	SegFormer-B3	2.8	14.8	37.0
Llama-2-7b	SegFormer-B4	2.7	14.2	35.0
Llama-2-7b	SegFormer-B5	2.8	13.3	34.8
OPT_125M	MAE-Base	1.5	9.8	28.2
OPT_125M	MAE-Huge	1.4	10.1	27.6
OPT_125M	MAE-Large	1.5	10.4	28.1
OPT_125M	ResNet101	2.6	14.8	37.1
OPT_125M	ResNet152	2.7	14.8	37.2
OPT_125M	ResNet18	2.2	12.6	34.1
OPT_125M	ResNet34	2.4	13.3	35.5
OPT_125M	ResNet50	2.7	14.1	36.4
OPT_125M	SegFormer-B0	1.6	8.9	26.8

Table 7 continued from previous page

LM	VM	P@1	P@10	P@100
OPT_125M	SegFormer-B1	2.2	11.9	32.9
OPT_125M	SegFormer-B2	2.2	13.1	34.3
OPT_125M	SegFormer-B3	2.2	13.5	35.3
OPT_125M	SegFormer-B4	2.2	13.1	34.6
OPT_125M	SegFormer-B5	2.0	12.6	34.6
OPT_30B	MAE-Base	3.8	18.2	45.7
OPT_30B	MAE-Huge	3.6	20.1	47.2
OPT_30B	MAE-Large	3.9	19.9	47.6
OPT_30B	ResNet101	4.9	25.2	57.2
OPT_30B	ResNet152	4.9	25.6	57.6
OPT_30B	ResNet18	4.7	21.2	50.4
OPT_30B	ResNet34	4.8	22.4	52.2
OPT_30B	ResNet50	4.9	24.8	56.5
OPT_30B	SegFormer-B0	3.1	14.0	39.3
OPT_30B	SegFormer-B1	4.1	19.9	48.9
OPT_30B	SegFormer-B2	4.6	21.2	50.9
OPT_30B	SegFormer-B3	4.3	21.9	52.6
OPT_30B	SegFormer-B4	4.4	21.1	52.0
OPT_30B	SegFormer-B5	4.2	21.2	51.3
OPT_6.7B	MAE-Base	3.2	17.0	43.0
OPT_6.7B	MAE-Huge	3.1	18.6	44.1
OPT_6.7B	MAE-Large	3.4	18.4	44.5
OPT_6.7B	ResNet101	4.4	23.7	54.6
OPT_6.7B	ResNet152	4.5	24.2	55.0
OPT_6.7B	ResNet18	4.4	19.5	47.8
OPT_6.7B	ResNet34	4.6	20.9	49.4
OPT_6.7B	ResNet50	4.5	23.4	54.0
OPT_6.7B	SegFormer-B0	2.8	13.0	37.2
OPT_6.7B	SegFormer-B1	3.9	18.7	46.1
OPT_6.7B	SegFormer-B2	4.4	20.3	48.1
OPT_6.7B	SegFormer-B3	4.0	21.1	50.3
OPT_6.7B	SegFormer-B4	4.2	20.4	49.5
OPT_6.7B	SegFormer-B5	4.0	19.8	48.7

Table 7: All results for Exclude-1k set.

LM	VM	Bins	P@10	P@100
BERT_LARGE	MAE-Base	high	4.6	17.5
BERT_LARGE	MAE-Base	low	16.0	46.3
BERT_LARGE	MAE-Base	medium	9.3	31.1
BERT_LARGE	MAE-Huge	high	4.1	17.7
BERT_LARGE	MAE-Huge	low	17.7	47.0
BERT_LARGE	MAE-Huge	medium	10.6	32.4
BERT_LARGE	MAE-Large	high	4.1	18.1
BERT_LARGE	MAE-Large	low	17.8	47.1
BERT_LARGE	MAE-Large	medium	10.1	32.5
BERT_LARGE	ResNet101	high	6.6	22.5

Table 8 continued from previous page

LM	VM	Bins	P@10	P@100
BERT_LARGE	ResNet101	low	25.1	56.2
BERT_LARGE	ResNet101	medium	13.5	40.1
BERT_LARGE	ResNet152	high	6.4	22.4
BERT_LARGE	ResNet152	low	25.0	57.0
BERT_LARGE	ResNet152	medium	13.4	39.8
BERT_LARGE	ResNet18	high	6.0	19.8
BERT_LARGE	ResNet18	low	21.6	51.6
BERT_LARGE	ResNet18	medium	12.1	35.9
BERT_LARGE	ResNet34	high	6.7	21.2
BERT_LARGE	ResNet34	low	22.5	53.6
BERT_LARGE	ResNet34	medium	12.1	38.0
BERT_LARGE	ResNet50	high	6.1	22.4
BERT_LARGE	ResNet50	low	24.2	55.0
BERT_LARGE	ResNet50	medium	13.2	39.1
BERT_LARGE	SegFormer-B0	high	4.6	17.2
BERT_LARGE	SegFormer-B0	low	14.7	42.6
BERT_LARGE	SegFormer-B0	medium	9.2	28.6
BERT_LARGE	SegFormer-B1	high	6.1	20.5
BERT_LARGE	SegFormer-B1	low	20.2	50.8
BERT_LARGE	SegFormer-B1	medium	11.5	34.6
BERT_LARGE	SegFormer-B2	high	5.7	20.5
BERT_LARGE	SegFormer-B2	low	22.3	53.4
BERT_LARGE	SegFormer-B2	medium	11.8	35.2
BERT_LARGE	SegFormer-B3	high	5.5	21.5
BERT_LARGE	SegFormer-B3	low	22.9	53.8
BERT_LARGE	SegFormer-B3	medium	12.0	36.9
BERT_LARGE	SegFormer-B4	high	6.2	20.3
BERT_LARGE	SegFormer-B4	low	22.3	53.6
BERT_LARGE	SegFormer-B4	medium	11.8	35.7
BERT_LARGE	SegFormer-B5	high	6.1	20.2
BERT_LARGE	SegFormer-B5	low	21.7	53.2
BERT_LARGE	SegFormer-B5	medium	12.0	35.8
GPT2_XL	MAE-Base	high	11.4	30.0
GPT2_XL	MAE-Base	low	11.8	37.8
GPT2_XL	MAE-Base	medium	14.7	38.0
GPT2_XL	MAE-Huge	high	12.8	31.3
GPT2_XL	MAE-Huge	low	13.4	38.8
GPT2_XL	MAE-Huge	medium	16.5	39.6
GPT2_XL	MAE-Large	high	13.1	31.4
GPT2_XL	MAE-Large	low	13.0	38.8
GPT2_XL	MAE-Large	medium	16.3	39.4
GPT2_XL	ResNet101	high	15.5	39.5
GPT2_XL	ResNet101	low	17.2	46.8
GPT2_XL	ResNet101	medium	21.2	49.8
GPT2_XL	ResNet152	high	15.9	39.6
GPT2_XL	ResNet152	low	17.3	46.9
GPT2_XL	ResNet152	medium	21.8	50.2
GPT2_XL	ResNet18	high	13.0	35.0
GPT2_XL	ResNet18	low	13.4	41.1

Table 8 continued from previous page

LM	VM	Bins	P@10	P@100
GPT2_XL	ResNet18	medium	17.8	44.5
GPT2_XL	ResNet34	high	13.8	36.2
GPT2_XL	ResNet34	low	14.4	42.8
GPT2_XL	ResNet34	medium	19.4	46.5
GPT2_XL	ResNet50	high	15.4	38.8
GPT2_XL	ResNet50	low	16.6	46.7
GPT2_XL	ResNet50	medium	21.1	48.9
GPT2_XL	SegFormer-B0	high	8.7	25.3
GPT2_XL	SegFormer-B0	low	9.6	30.5
GPT2_XL	SegFormer-B0	medium	12.0	35.2
GPT2_XL	SegFormer-B1	high	12.3	33.1
GPT2_XL	SegFormer-B1	low	13.4	39.6
GPT2_XL	SegFormer-B1	medium	17.3	42.6
GPT2_XL	SegFormer-B2	high	13.7	35.2
GPT2_XL	SegFormer-B2	low	13.8	40.8
GPT2_XL	SegFormer-B2	medium	18.5	44.3
GPT2_XL	SegFormer-B3	high	14.2	36.4
GPT2_XL	SegFormer-B3	low	14.8	42.0
GPT2_XL	SegFormer-B3	medium	18.6	46.5
GPT2_XL	SegFormer-B4	high	14.0	35.6
GPT2_XL	SegFormer-B4	low	15.1	41.5
GPT2_XL	SegFormer-B4	medium	18.9	45.6
GPT2_XL	SegFormer-B5	high	13.5	35.6
GPT2_XL	SegFormer-B5	low	14.4	41.3
GPT2_XL	SegFormer-B5	medium	18.1	44.5
Llama-2-13B	MAE-Base	high	7.2	20.3
Llama-2-13B	MAE-Base	low	19.6	49.9
Llama-2-13B	MAE-Base	medium	14.4	39.1
Llama-2-13B	MAE-Huge	high	8.0	22.0
Llama-2-13B	MAE-Huge	low	23.5	52.8
Llama-2-13B	MAE-Huge	medium	16.5	42.2
Llama-2-13B	MAE-Large	high	8.0	22.4
Llama-2-13B	MAE-Large	low	21.7	52.8
Llama-2-13B	MAE-Large	medium	16.2	41.3
Llama-2-13B	ResNet101	high	11.7	32.6
Llama-2-13B	ResNet101	low	29.5	65.6
Llama-2-13B	ResNet101	medium	23.2	55.4
Llama-2-13B	ResNet152	high	12.0	33.2
Llama-2-13B	ResNet152	low	29.7	65.9
Llama-2-13B	ResNet152	medium	23.4	55.7
Llama-2-13B	ResNet18	high	8.4	23.6
Llama-2-13B	ResNet18	low	21.7	53.9
Llama-2-13B	ResNet18	medium	17.3	44.5
Llama-2-13B	ResNet34	high	9.2	25.5
Llama-2-13B	ResNet34	low	23.0	55.8
Llama-2-13B	ResNet34	medium	19.0	46.4
Llama-2-13B	ResNet50	high	11.4	32.0
Llama-2-13B	ResNet50	low	28.8	64.6
Llama-2-13B	ResNet50	medium	23.4	54.2

Table 8 continued from previous page

LM	VM	Bins	P@10	P@100
Llama-2-13B	SegFormer-B0	high	4.6	15.3
Llama-2-13B	SegFormer-B0	low	13.4	39.0
Llama-2-13B	SegFormer-B0	medium	9.5	29.4
Llama-2-13B	SegFormer-B1	high	7.7	23.7
Llama-2-13B	SegFormer-B1	low	21.0	53.4
Llama-2-13B	SegFormer-B1	medium	15.6	43.9
Llama-2-13B	SegFormer-B2	high	8.5	24.7
Llama-2-13B	SegFormer-B2	low	22.6	54.0
Llama-2-13B	SegFormer-B2	medium	17.8	45.7
Llama-2-13B	SegFormer-B3	high	9.3	27.6
Llama-2-13B	SegFormer-B3	low	23.8	55.9
Llama-2-13B	SegFormer-B3	medium	18.5	48.6
Llama-2-13B	SegFormer-B4	high	8.6	26.1
Llama-2-13B	SegFormer-B4	low	23.2	55.0
Llama-2-13B	SegFormer-B4	medium	17.8	48.5
Llama-2-13B	SegFormer-B5	high	8.5	25.8
Llama-2-13B	SegFormer-B5	low	22.1	54.5
Llama-2-13B	SegFormer-B5	medium	17.9	46.9
OPT_30B	MAE-Base	high	10.8	31.0
OPT_30B	MAE-Base	low	24.1	56.5
OPT_30B	MAE-Base	medium	19.6	50.4
OPT_30B	MAE-Huge	high	12.5	31.6
OPT_30B	MAE-Huge	low	26.0	59.2
OPT_30B	MAE-Huge	medium	21.9	52.0
OPT_30B	MAE-Large	high	12.4	32.2
OPT_30B	MAE-Large	low	25.4	59.2
OPT_30B	MAE-Large	medium	21.7	52.4
OPT_30B	ResNet101	high	15.6	40.3
OPT_30B	ResNet101	low	31.9	67.2
OPT_30B	ResNet101	medium	28.0	64.4
OPT_30B	ResNet152	high	15.5	40.6
OPT_30B	ResNet152	low	31.8	67.8
OPT_30B	ResNet152	medium	29.1	64.8
OPT_30B	ResNet18	high	13.2	34.9
OPT_30B	ResNet18	low	25.8	59.8
OPT_30B	ResNet18	medium	23.9	56.6
OPT_30B	ResNet34	high	13.5	36.2
OPT_30B	ResNet34	low	27.7	61.9
OPT_30B	ResNet34	medium	25.6	58.8
OPT_30B	ResNet50	high	14.8	39.8
OPT_30B	ResNet50	low	31.5	67.3
OPT_30B	ResNet50	medium	27.9	62.7
OPT_30B	SegFormer-B0	high	7.9	25.2
OPT_30B	SegFormer-B0	low	17.8	48.1
OPT_30B	SegFormer-B0	medium	16.0	45.0
OPT_30B	SegFormer-B1	high	11.8	32.8
OPT_30B	SegFormer-B1	low	25.4	59.6
OPT_30B	SegFormer-B1	medium	22.2	54.9
OPT_30B	SegFormer-B2	high	12.6	35.2

Table 8 continued from previous page

LM	VM	Bins	P@10	P@100
OPT_30B	SegFormer-B2	low	26.0	60.5
OPT_30B	SegFormer-B2	medium	24.5	57.0
OPT_30B	SegFormer-B3	high	13.3	37.5
OPT_30B	SegFormer-B3	low	26.1	60.7
OPT_30B	SegFormer-B3	medium	25.7	59.1
OPT_30B	SegFormer-B4	high	12.5	35.8
OPT_30B	SegFormer-B4	low	25.7	60.7
OPT_30B	SegFormer-B4	medium	24.6	59.4
OPT_30B	SegFormer-B5	high	12.6	35.9
OPT_30B	SegFormer-B5	low	26.5	60.2
OPT_30B	SegFormer-B5	medium	24.1	57.5

Table 8: All language dispersion results (Exclude-1k set).

LM	VM	Bins	P@10	P@100
BERT_BASE	MAE-Huge	high	8.8	28.2
BERT_BASE	MAE-Huge	low	14.8	39.5
BERT_BASE	MAE-Huge	medium	13.5	36.3
BERT_BASE	ResNet152	high	15.4	39.5
BERT_BASE	ResNet152	low	20.4	49.0
BERT_BASE	ResNet152	medium	17.9	43.6
BERT_BASE	SegFormer-B5	high	15.4	39.3
BERT_BASE	SegFormer-B5	low	15.6	40.7
BERT_BASE	SegFormer-B5	medium	15.7	41.6
BERT_LARGE	MAE-Huge	high	10.2	32.3
BERT_LARGE	MAE-Huge	low	15.3	41.3
BERT_LARGE	MAE-Huge	medium	13.8	38.1
BERT_LARGE	ResNet152	high	16.1	41.2
BERT_LARGE	ResNet152	low	19.6	50.4
BERT_LARGE	ResNet152	medium	19.8	46.1
BERT_LARGE	SegFormer-B5	high	16.8	42.3
BERT_LARGE	SegFormer-B5	low	14.7	42.0
BERT_LARGE	SegFormer-B5	medium	17.1	42.7
BERT_MEDIUM	MAE-Huge	high	5.9	19.2
BERT_MEDIUM	MAE-Huge	low	10.9	29.6
BERT_MEDIUM	MAE-Huge	medium	10.7	27.3
BERT_MEDIUM	ResNet152	high	11.1	29.3
BERT_MEDIUM	ResNet152	low	15.1	38.4
BERT_MEDIUM	ResNet152	medium	14.6	34.3
BERT_MEDIUM	SegFormer-B5	high	12.1	29.9
BERT_MEDIUM	SegFormer-B5	low	12.4	32.7
BERT_MEDIUM	SegFormer-B5	medium	12.3	31.0
BERT_MINI	MAE-Huge	high	3.7	12.1
BERT_MINI	MAE-Huge	low	7.0	21.4
BERT_MINI	MAE-Huge	medium	6.3	20.0
BERT_MINI	ResNet152	high	6.4	21.3
BERT_MINI	ResNet152	low	10.7	29.4

Table 9 continued from previous page

LM	VM	Bins	P@10	P@100
BERT_MINI	ResNet152	medium	10.2	25.8
BERT_MINI	SegFormer-B5	high	7.3	21.0
BERT_MINI	SegFormer-B5	low	7.9	24.3
BERT_MINI	SegFormer-B5	medium	8.1	22.8
BERT_SMALL	MAE-Huge	high	5.3	18.4
BERT_SMALL	MAE-Huge	low	9.7	29.0
BERT_SMALL	MAE-Huge	medium	10.2	26.1
BERT_SMALL	ResNet152	high	10.2	27.3
BERT_SMALL	ResNet152	low	14.3	37.8
BERT_SMALL	ResNet152	medium	14.1	32.4
BERT_SMALL	SegFormer-B5	high	11.0	28.0
BERT_SMALL	SegFormer-B5	low	12.0	31.8
BERT_SMALL	SegFormer-B5	medium	12.4	29.7
BERT_TINY	MAE-Huge	high	1.2	6.7
BERT_TINY	MAE-Huge	low	3.6	14.8
BERT_TINY	MAE-Huge	medium	2.5	11.6
BERT_TINY	ResNet152	high	3.5	14.6
BERT_TINY	ResNet152	low	6.2	20.8
BERT_TINY	ResNet152	medium	6.0	18.1
BERT_TINY	SegFormer-B5	high	3.3	13.9
BERT_TINY	SegFormer-B5	low	4.5	16.2
BERT_TINY	SegFormer-B5	medium	3.9	14.8
GPT2_BASE	MAE-Huge	high	0.9	3.4
GPT2_BASE	MAE-Huge	low	1.4	5.1
GPT2_BASE	MAE-Huge	medium	1.6	5.3
GPT2_BASE	ResNet152	high	2.0	6.9
GPT2_BASE	ResNet152	low	3.8	11.8
GPT2_BASE	ResNet152	medium	2.9	9.7
GPT2_BASE	SegFormer-B5	high	2.5	7.5
GPT2_BASE	SegFormer-B5	low	3.5	9.5
GPT2_BASE	SegFormer-B5	medium	2.8	8.1
GPT2_LARGE	MAE-Huge	high	12.0	32.1
GPT2_LARGE	MAE-Huge	low	14.0	35.3
GPT2_LARGE	MAE-Huge	medium	14.2	35.2
GPT2_LARGE	ResNet152	high	15.4	39.5
GPT2_LARGE	ResNet152	low	19.4	47.8
GPT2_LARGE	ResNet152	medium	18.5	44.1
GPT2_LARGE	SegFormer-B5	high	15.4	37.6
GPT2_LARGE	SegFormer-B5	low	13.8	38.8
GPT2_LARGE	SegFormer-B5	medium	15.4	39.6
GPT2_XL	MAE-Huge	high	13.5	35.0
GPT2_XL	MAE-Huge	low	14.5	38.8
GPT2_XL	MAE-Huge	medium	16.6	38.1
GPT2_XL	ResNet152	high	16.7	42.0
GPT2_XL	ResNet152	low	20.3	51.3
GPT2_XL	ResNet152	medium	20.2	46.7
GPT2_XL	SegFormer-B5	high	16.8	39.5
GPT2_XL	SegFormer-B5	low	14.7	42.6
GPT2_XL	SegFormer-B5	medium	16.6	42.3

Table 9 continued from previous page

LM	VM	Bins	P@10	P@100
Llama-2-13B	MAE-Huge	high	13.9	33.7
Llama-2-13B	MAE-Huge	low	17.6	43.4
Llama-2-13B	MAE-Huge	medium	19.3	43.7
Llama-2-13B	ResNet152	high	21.3	46.6
Llama-2-13B	ResNet152	low	24.8	62.9
Llama-2-13B	ResNet152	medium	22.6	51.2
Llama-2-13B	SegFormer-B5	high	17.4	40.8
Llama-2-13B	SegFormer-B5	low	16.6	47.2
Llama-2-13B	SegFormer-B5	medium	17.5	44.4
Llama-2-70B	MAE-Huge	high	6.9	18.6
Llama-2-70B	MAE-Huge	low	10.4	25.9
Llama-2-70B	MAE-Huge	medium	10.3	26.7
Llama-2-70B	ResNet152	high	13.8	31.2
Llama-2-70B	ResNet152	low	18.9	46.6
Llama-2-70B	ResNet152	medium	15.8	35.0
Llama-2-70B	SegFormer-B5	high	9.2	25.2
Llama-2-70B	SegFormer-B5	low	10.6	32.4
Llama-2-70B	SegFormer-B5	medium	11.0	28.6
Llama-2-7B	MAE-Huge	high	9.7	23.8
Llama-2-7B	MAE-Huge	low	12.9	32.6
Llama-2-7B	MAE-Huge	medium	12.9	31.2
Llama-2-7B	ResNet152	high	16.7	35.5
Llama-2-7B	ResNet152	low	22.1	51.7
Llama-2-7B	ResNet152	medium	17.9	39.9
Llama-2-7B	SegFormer-B5	high	12.4	30.7
Llama-2-7B	SegFormer-B5	low	13.8	38.5
Llama-2-7B	SegFormer-B5	medium	13.8	35.3
OPT_125M	MAE-Huge	high	8.8	24.5
OPT_125M	MAE-Huge	low	10.9	29.5
OPT_125M	MAE-Huge	medium	10.5	28.1
OPT_125M	ResNet152	high	12.4	33.4
OPT_125M	ResNet152	low	15.7	40.3
OPT_125M	ResNet152	medium	16.2	38.0
OPT_125M	SegFormer-B5	high	12.6	34.2
OPT_125M	SegFormer-B5	low	12.2	33.9
OPT_125M	SegFormer-B5	medium	13.1	35.7
OPT_30B	MAE-Huge	high	18.2	43.5
OPT_30B	MAE-Huge	low	20.8	49.1
OPT_30B	MAE-Huge	medium	21.2	48.4
OPT_30B	ResNet152	high	23.6	51.9
OPT_30B	ResNet152	low	26.7	64.4
OPT_30B	ResNet152	medium	26.5	56.7
OPT_30B	SegFormer-B5	high	22.3	48.3
OPT_30B	SegFormer-B5	low	19.0	52.5
OPT_30B	SegFormer-B5	medium	22.2	53.0
OPT_6.7B	MAE-Huge	high	16.6	40.8
OPT_6.7B	MAE-Huge	low	19.2	46.0
OPT_6.7B	MAE-Huge	medium	19.7	45.0
OPT_6.7B	ResNet152	high	22.2	49.8

Table 9 continued from previous page

LM	VM	Bins	P@10	P@100
OPT_6.7B	ResNet152	low	25.2	61.0
OPT_6.7B	ResNet152	medium	25.1	54.3
OPT_6.7B	SegFormer-B5	high	21.1	46.1
OPT_6.7B	SegFormer-B5	low	17.9	49.6
OPT_6.7B	SegFormer-B5	medium	20.6	50.3
OPT_66B	MAE-Huge	high	16.2	39.2
OPT_66B	MAE-Huge	low	19.2	45.3
OPT_66B	MAE-Huge	medium	19.4	44.7
OPT_66B	ResNet152	high	21.8	49.7
OPT_66B	ResNet152	low	25.2	60.9
OPT_66B	ResNet152	medium	25.1	54.9
OPT_66B	SegFormer-B5	high	20.8	45.1
OPT_66B	SegFormer-B5	low	17.6	49.5
OPT_66B	SegFormer-B5	medium	20.3	49.7

Table 9: All image dispersion results (Exclude-1k set).

LM	VM	Bins	P@10	P@100
BERT_LARGE	MAE-Base	1	11.7	37.2
BERT_LARGE	MAE-Base	2-3	6.6	22.9
BERT_LARGE	MAE-Base	4+	4.6	17.7
BERT_LARGE	MAE-Huge	1	14.1	41.6
BERT_LARGE	MAE-Huge	2-3	5.4	25.2
BERT_LARGE	MAE-Huge	4+	4.8	18.2
BERT_LARGE	MAE-Large	1	13.3	39.9
BERT_LARGE	MAE-Large	2-3	5.8	23.1
BERT_LARGE	MAE-Large	4+	5.0	19.0
BERT_LARGE	ResNet101	1	19.3	45.8
BERT_LARGE	ResNet101	2-3	11.7	30.2
BERT_LARGE	ResNet101	4+	7.4	22.2
BERT_LARGE	ResNet152	1	16.3	46.1
BERT_LARGE	ResNet152	2-3	11.7	32.1
BERT_LARGE	ResNet152	4+	7.1	22.1
BERT_LARGE	ResNet18	1	14.4	42.6
BERT_LARGE	ResNet18	2-3	10.8	27.2
BERT_LARGE	ResNet18	4+	5.7	19.9
BERT_LARGE	ResNet34	1	14.7	44.1
BERT_LARGE	ResNet34	2-3	12.0	30.0
BERT_LARGE	ResNet34	4+	5.9	21.3
BERT_LARGE	ResNet50	1	18.5	46.1
BERT_LARGE	ResNet50	2-3	11.2	29.8
BERT_LARGE	ResNet50	4+	7.3	22.1
BERT_LARGE	SegFormer-B0	1	11.2	34.5
BERT_LARGE	SegFormer-B0	2-3	5.3	24.9
BERT_LARGE	SegFormer-B0	4+	4.2	15.1
BERT_LARGE	SegFormer-B1	1	15.1	43.6
BERT_LARGE	SegFormer-B1	2-3	8.8	27.3

Table 10 continued from previous page

LM	VM	Bins	P@10	P@100
BERT_LARGE	SegFormer-B1	4+	5.3	20.8
BERT_LARGE	SegFormer-B2	1	15.9	41.9
BERT_LARGE	SegFormer-B2	2-3	9.2	29.1
BERT_LARGE	SegFormer-B2	4+	5.3	21.6
BERT_LARGE	SegFormer-B3	1	15.0	47.3
BERT_LARGE	SegFormer-B3	2-3	10.6	31.1
BERT_LARGE	SegFormer-B3	4+	6.4	22.1
BERT_LARGE	SegFormer-B4	1	15.2	45.8
BERT_LARGE	SegFormer-B4	2-3	10.9	30.2
BERT_LARGE	SegFormer-B4	4+	6.1	21.2
BERT_LARGE	SegFormer-B5	1	14.8	45.8
BERT_LARGE	SegFormer-B5	2-3	9.7	30.2
BERT_LARGE	SegFormer-B5	4+	5.8	21.0
GPT2_XL	MAE-Base	1	12.1	35.3
GPT2_XL	MAE-Base	2-3	8.0	21.4
GPT2_XL	MAE-Base	4+	8.0	18.9
GPT2_XL	MAE-Huge	1	13.4	37.0
GPT2_XL	MAE-Huge	2-3	9.0	25.2
GPT2_XL	MAE-Huge	4+	11.0	18.8
GPT2_XL	MAE-Large	1	13.8	36.7
GPT2_XL	MAE-Large	2-3	8.7	22.8
GPT2_XL	MAE-Large	4+	10.2	20.4
GPT2_XL	ResNet101	1	17.4	44.7
GPT2_XL	ResNet101	2-3	16.0	33.4
GPT2_XL	ResNet101	4+	10.6	26.2
GPT2_XL	ResNet152	1	18.6	44.2
GPT2_XL	ResNet152	2-3	15.3	34.9
GPT2_XL	ResNet152	4+	9.9	26.5
GPT2_XL	ResNet18	1	15.1	40.4
GPT2_XL	ResNet18	2-3	9.3	29.5
GPT2_XL	ResNet18	4+	8.3	23.4
GPT2_XL	ResNet34	1	15.9	42.5
GPT2_XL	ResNet34	2-3	12.1	30.0
GPT2_XL	ResNet34	4+	8.6	25.4
GPT2_XL	ResNet50	1	16.5	44.2
GPT2_XL	ResNet50	2-3	14.4	32.2
GPT2_XL	ResNet50	4+	10.0	26.0
GPT2_XL	SegFormer-B0	1	9.7	31.1
GPT2_XL	SegFormer-B0	2-3	6.6	20.9
GPT2_XL	SegFormer-B0	4+	6.4	16.2
GPT2_XL	SegFormer-B1	1	14.3	37.8
GPT2_XL	SegFormer-B1	2-3	10.0	27.3
GPT2_XL	SegFormer-B1	4+	6.4	21.3
GPT2_XL	SegFormer-B2	1	15.9	40.5
GPT2_XL	SegFormer-B2	2-3	10.6	29.2
GPT2_XL	SegFormer-B2	4+	7.2	22.2
GPT2_XL	SegFormer-B3	1	16.5	42.5
GPT2_XL	SegFormer-B3	2-3	12.3	31.1
GPT2_XL	SegFormer-B3	4+	7.5	23.1

Table 10 continued from previous page

LM	VM	Bins	P@10	P@100
GPT2_XL	SegFormer-B4	1	16.9	41.2
GPT2_XL	SegFormer-B4	2-3	11.0	31.0
GPT2_XL	SegFormer-B4	4+	8.6	23.8
GPT2_XL	SegFormer-B5	1	15.3	42.5
GPT2_XL	SegFormer-B5	2-3	10.0	32.6
GPT2_XL	SegFormer-B5	4+	8.4	24.9
Llama-2-13B	MAE-Base	1	8.4	25.9
Llama-2-13B	MAE-Base	2-3	6.3	18.8
Llama-2-13B	MAE-Base	4+	4.8	13.3
Llama-2-13B	MAE-Huge	1	10.0	30.7
Llama-2-13B	MAE-Huge	2-3	7.5	21.0
Llama-2-13B	MAE-Huge	4+	5.8	15.1
Llama-2-13B	MAE-Large	1	9.8	29.1
Llama-2-13B	MAE-Large	2-3	6.4	21.3
Llama-2-13B	MAE-Large	4+	4.9	14.7
Llama-2-13B	ResNet101	1	14.8	42.8
Llama-2-13B	ResNet101	2-3	11.8	33.3
Llama-2-13B	ResNet101	4+	6.9	21.7
Llama-2-13B	ResNet152	1	13.0	43.8
Llama-2-13B	ResNet152	2-3	11.5	32.7
Llama-2-13B	ResNet152	4+	7.1	21.1
Llama-2-13B	ResNet18	1	12.1	31.5
Llama-2-13B	ResNet18	2-3	5.7	23.8
Llama-2-13B	ResNet18	4+	4.7	16.4
Llama-2-13B	ResNet34	1	11.5	36.4
Llama-2-13B	ResNet34	2-3	9.7	23.9
Llama-2-13B	ResNet34	4+	4.5	17.6
Llama-2-13B	ResNet50	1	14.3	41.1
Llama-2-13B	ResNet50	2-3	11.7	31.5
Llama-2-13B	ResNet50	4+	6.0	20.5
Llama-2-13B	SegFormer-B0	1	6.9	21.8
Llama-2-13B	SegFormer-B0	2-3	3.3	15.0
Llama-2-13B	SegFormer-B0	4+	2.4	10.0
Llama-2-13B	SegFormer-B1	1	8.7	31.4
Llama-2-13B	SegFormer-B1	2-3	6.5	21.4
Llama-2-13B	SegFormer-B1	4+	4.2	14.8
Llama-2-13B	SegFormer-B2	1	10.6	32.9
Llama-2-13B	SegFormer-B2	2-3	9.4	26.4
Llama-2-13B	SegFormer-B2	4+	4.0	15.3
Llama-2-13B	SegFormer-B3	1	10.7	37.0
Llama-2-13B	SegFormer-B3	2-3	11.0	26.6
Llama-2-13B	SegFormer-B3	4+	4.2	16.4
Llama-2-13B	SegFormer-B4	1	9.1	36.6
Llama-2-13B	SegFormer-B4	2-3	9.4	24.5
Llama-2-13B	SegFormer-B4	4+	3.6	16.7
Llama-2-13B	SegFormer-B5	1	9.5	36.0
Llama-2-13B	SegFormer-B5	2-3	9.3	26.2
Llama-2-13B	SegFormer-B5	4+	4.8	17.2
OPT_30B	MAE-Base	1	17.2	42.5

Table 10 continued from previous page

LM	VM	Bins	P@10	P@100
OPT_30B	MAE-Base	2-3	11.6	26.1
OPT_30B	MAE-Base	4+	9.0	22.8
OPT_30B	MAE-Huge	1	18.8	42.3
OPT_30B	MAE-Huge	2-3	12.6	28.0
OPT_30B	MAE-Huge	4+	9.5	21.9
OPT_30B	MAE-Large	1	18.5	44.9
OPT_30B	MAE-Large	2-3	12.8	28.3
OPT_30B	MAE-Large	4+	9.4	22.7
OPT_30B	ResNet101	1	22.3	53.5
OPT_30B	ResNet101	2-3	19.7	38.6
OPT_30B	ResNet101	4+	12.8	27.9
OPT_30B	ResNet152	1	23.8	53.3
OPT_30B	ResNet152	2-3	18.8	40.6
OPT_30B	ResNet152	4+	11.6	28.6
OPT_30B	ResNet18	1	19.5	44.6
OPT_30B	ResNet18	2-3	16.7	32.9
OPT_30B	ResNet18	4+	11.0	27.2
OPT_30B	ResNet34	1	21.0	47.6
OPT_30B	ResNet34	2-3	18.8	34.9
OPT_30B	ResNet34	4+	10.9	26.0
OPT_30B	ResNet50	1	21.9	53.1
OPT_30B	ResNet50	2-3	19.1	37.6
OPT_30B	ResNet50	4+	11.1	29.1
OPT_30B	SegFormer-B0	1	11.6	37.8
OPT_30B	SegFormer-B0	2-3	9.8	25.9
OPT_30B	SegFormer-B0	4+	6.6	18.8
OPT_30B	SegFormer-B1	1	18.0	44.8
OPT_30B	SegFormer-B1	2-3	16.0	32.7
OPT_30B	SegFormer-B1	4+	8.0	23.8
OPT_30B	SegFormer-B2	1	18.7	48.0
OPT_30B	SegFormer-B2	2-3	17.0	33.9
OPT_30B	SegFormer-B2	4+	9.0	24.4
OPT_30B	SegFormer-B3	1	19.2	51.8
OPT_30B	SegFormer-B3	2-3	18.4	37.0
OPT_30B	SegFormer-B3	4+	9.9	23.8
OPT_30B	SegFormer-B4	1	18.2	51.5
OPT_30B	SegFormer-B4	2-3	18.0	36.1
OPT_30B	SegFormer-B4	4+	10.3	24.7
OPT_30B	SegFormer-B5	1	18.5	49.2
OPT_30B	SegFormer-B5	2-3	16.6	37.4
OPT_30B	SegFormer-B5	4+	11.0	24.9

Table 10: All polysemy results (Exclude-1k set).

LM	VM	Bins	P@10	P@100
BERT_LARGE	MAE-Base	0-10k	6.2	23.4
BERT_LARGE	MAE-Base	10k-100k	13.5	35.9

Table 11 continued from previous page

LM	VM	Bins	P@10	P@100
BERT_LARGE	MAE-Base	100k+	11.2	36.8
BERT_LARGE	MAE-Huge	0-10k	7.3	25.0
BERT_LARGE	MAE-Huge	10k-100k	13.9	37.3
BERT_LARGE	MAE-Huge	100k+	12.6	37.1
BERT_LARGE	MAE-Large	0-10k	7.7	24.6
BERT_LARGE	MAE-Large	10k-100k	14.6	36.7
BERT_LARGE	MAE-Large	100k+	12.2	37.7
BERT_LARGE	ResNet101	0-10k	8.6	27.0
BERT_LARGE	ResNet101	10k-100k	19.7	44.9
BERT_LARGE	ResNet101	100k+	17.7	45.4
BERT_LARGE	ResNet152	0-10k	7.9	26.1
BERT_LARGE	ResNet152	10k-100k	19.5	45.3
BERT_LARGE	ResNet152	100k+	17.6	45.8
BERT_LARGE	ResNet18	0-10k	7.6	25.4
BERT_LARGE	ResNet18	10k-100k	17.5	40.8
BERT_LARGE	ResNet18	100k+	15.2	41.2
BERT_LARGE	ResNet34	0-10k	6.6	26.3
BERT_LARGE	ResNet34	10k-100k	18.4	42.8
BERT_LARGE	ResNet34	100k+	15.8	43.2
BERT_LARGE	ResNet50	0-10k	7.6	26.7
BERT_LARGE	ResNet50	10k-100k	19.6	44.2
BERT_LARGE	ResNet50	100k+	16.9	44.4
BERT_LARGE	SegFormer-B0	0-10k	4.4	20.4
BERT_LARGE	SegFormer-B0	10k-100k	13.5	34.4
BERT_LARGE	SegFormer-B0	100k+	10.3	33.6
BERT_LARGE	SegFormer-B1	0-10k	6.8	24.0
BERT_LARGE	SegFormer-B1	10k-100k	17.2	40.8
BERT_LARGE	SegFormer-B1	100k+	14.2	40.3
BERT_LARGE	SegFormer-B2	0-10k	7.2	24.4
BERT_LARGE	SegFormer-B2	10k-100k	18.0	42.5
BERT_LARGE	SegFormer-B2	100k+	15.4	41.9
BERT_LARGE	SegFormer-B3	0-10k	8.5	26.7
BERT_LARGE	SegFormer-B3	10k-100k	18.6	43.1
BERT_LARGE	SegFormer-B3	100k+	15.5	42.6
BERT_LARGE	SegFormer-B4	0-10k	9.2	26.1
BERT_LARGE	SegFormer-B4	10k-100k	17.4	42.8
BERT_LARGE	SegFormer-B4	100k+	15.8	41.9
BERT_LARGE	SegFormer-B5	0-10k	7.3	26.1
BERT_LARGE	SegFormer-B5	10k-100k	18.2	42.1
BERT_LARGE	SegFormer-B5	100k+	15.2	41.8
GPT2_XL	MAE-Base	0-10k	10.6	24.0
GPT2_XL	MAE-Base	10k-100k	13.0	33.9
GPT2_XL	MAE-Base	100k+	12.6	36.6
GPT2_XL	MAE-Huge	0-10k	13.5	26.4
GPT2_XL	MAE-Huge	10k-100k	13.9	34.6
GPT2_XL	MAE-Huge	100k+	14.5	38.1
GPT2_XL	MAE-Large	0-10k	12.1	26.4
GPT2_XL	MAE-Large	10k-100k	14.5	34.9
GPT2_XL	MAE-Large	100k+	14.1	37.9

Table 11 continued from previous page

LM	VM	Bins	P@10	P@100
GPT2_XL	ResNet101	0-10k	14.4	29.9
GPT2_XL	ResNet101	10k-100k	17.6	45.7
GPT2_XL	ResNet101	100k+	18.4	46.3
GPT2_XL	ResNet152	0-10k	13.5	30.7
GPT2_XL	ResNet152	10k-100k	18.0	45.8
GPT2_XL	ResNet152	100k+	18.8	46.5
GPT2_XL	ResNet18	0-10k	12.8	27.9
GPT2_XL	ResNet18	10k-100k	14.6	40.5
GPT2_XL	ResNet18	100k+	14.9	41.0
GPT2_XL	ResNet34	0-10k	13.0	28.5
GPT2_XL	ResNet34	10k-100k	16.3	42.7
GPT2_XL	ResNet34	100k+	15.9	42.3
GPT2_XL	ResNet50	0-10k	13.4	30.1
GPT2_XL	ResNet50	10k-100k	17.6	44.7
GPT2_XL	ResNet50	100k+	18.1	45.8
GPT2_XL	SegFormer-B0	0-10k	10.7	21.6
GPT2_XL	SegFormer-B0	10k-100k	11.3	30.2
GPT2_XL	SegFormer-B0	100k+	9.5	30.9
GPT2_XL	SegFormer-B1	0-10k	11.2	25.4
GPT2_XL	SegFormer-B1	10k-100k	14.7	38.8
GPT2_XL	SegFormer-B1	100k+	14.3	39.1
GPT2_XL	SegFormer-B2	0-10k	10.3	26.3
GPT2_XL	SegFormer-B2	10k-100k	16.0	40.8
GPT2_XL	SegFormer-B2	100k+	15.4	40.7
GPT2_XL	SegFormer-B3	0-10k	11.0	26.1
GPT2_XL	SegFormer-B3	10k-100k	16.7	42.6
GPT2_XL	SegFormer-B3	100k+	15.9	42.2
GPT2_XL	SegFormer-B4	0-10k	11.7	26.7
GPT2_XL	SegFormer-B4	10k-100k	16.9	42.0
GPT2_XL	SegFormer-B4	100k+	15.9	41.4
GPT2_XL	SegFormer-B5	0-10k	10.0	28.5
GPT2_XL	SegFormer-B5	10k-100k	16.0	40.7
GPT2_XL	SegFormer-B5	100k+	15.4	41.2
Llama-2-13B	MAE-Base	0-10k	8.2	19.3
Llama-2-13B	MAE-Base	10k-100k	13.2	32.1
Llama-2-13B	MAE-Base	100k+	14.8	40.5
Llama-2-13B	MAE-Huge	0-10k	10.1	21.2
Llama-2-13B	MAE-Huge	10k-100k	14.2	33.7
Llama-2-13B	MAE-Huge	100k+	17.7	43.6
Llama-2-13B	MAE-Large	0-10k	9.0	21.6
Llama-2-13B	MAE-Large	10k-100k	14.3	33.8
Llama-2-13B	MAE-Large	100k+	16.5	43.3
Llama-2-13B	ResNet101	0-10k	11.0	26.3
Llama-2-13B	ResNet101	10k-100k	20.2	46.2
Llama-2-13B	ResNet101	100k+	23.3	56.3
Llama-2-13B	ResNet152	0-10k	10.8	27.2
Llama-2-13B	ResNet152	10k-100k	20.3	47.3
Llama-2-13B	ResNet152	100k+	23.6	56.2
Llama-2-13B	ResNet18	0-10k	9.0	22.0

Table 11 continued from previous page

LM	VM	Bins	P@10	P@100
Llama-2-13B	ResNet18	10k-100k	14.8	37.6
Llama-2-13B	ResNet18	100k+	17.0	44.3
Llama-2-13B	ResNet34	0-10k	8.3	22.5
Llama-2-13B	ResNet34	10k-100k	16.7	39.7
Llama-2-13B	ResNet34	100k+	18.2	46.2
Llama-2-13B	ResNet50	0-10k	10.0	25.6
Llama-2-13B	ResNet50	10k-100k	19.9	45.6
Llama-2-13B	ResNet50	100k+	23.0	55.2
Llama-2-13B	SegFormer-B0	0-10k	5.6	15.9
Llama-2-13B	SegFormer-B0	10k-100k	8.9	25.7
Llama-2-13B	SegFormer-B0	100k+	9.7	30.4
Llama-2-13B	SegFormer-B1	0-10k	8.9	19.2
Llama-2-13B	SegFormer-B1	10k-100k	14.5	37.3
Llama-2-13B	SegFormer-B1	100k+	15.7	44.0
Llama-2-13B	SegFormer-B2	0-10k	9.2	21.4
Llama-2-13B	SegFormer-B2	10k-100k	16.1	38.9
Llama-2-13B	SegFormer-B2	100k+	17.3	44.8
Llama-2-13B	SegFormer-B3	0-10k	8.4	20.7
Llama-2-13B	SegFormer-B3	10k-100k	17.5	41.2
Llama-2-13B	SegFormer-B3	100k+	18.0	47.8
Llama-2-13B	SegFormer-B4	0-10k	6.4	21.1
Llama-2-13B	SegFormer-B4	10k-100k	16.0	40.6
Llama-2-13B	SegFormer-B4	100k+	18.0	46.9
Llama-2-13B	SegFormer-B5	0-10k	7.5	22.2
Llama-2-13B	SegFormer-B5	10k-100k	16.1	39.9
Llama-2-13B	SegFormer-B5	100k+	17.2	45.8
OPT_30B	MAE-Base	0-10k	11.7	27.5
OPT_30B	MAE-Base	10k-100k	17.4	42.2
OPT_30B	MAE-Base	100k+	17.9	47.0
OPT_30B	MAE-Huge	0-10k	13.4	27.9
OPT_30B	MAE-Huge	10k-100k	18.1	43.0
OPT_30B	MAE-Huge	100k+	20.3	48.9
OPT_30B	MAE-Large	0-10k	13.4	27.7
OPT_30B	MAE-Large	10k-100k	18.7	44.0
OPT_30B	MAE-Large	100k+	19.6	49.1
OPT_30B	ResNet101	0-10k	15.6	29.0
OPT_30B	ResNet101	10k-100k	24.5	54.5
OPT_30B	ResNet101	100k+	24.9	58.7
OPT_30B	ResNet152	0-10k	15.4	30.6
OPT_30B	ResNet152	10k-100k	24.5	54.6
OPT_30B	ResNet152	100k+	25.4	59.1
OPT_30B	ResNet18	0-10k	13.6	28.0
OPT_30B	ResNet18	10k-100k	20.9	48.4
OPT_30B	ResNet18	100k+	20.6	51.0
OPT_30B	ResNet34	0-10k	13.3	26.9
OPT_30B	ResNet34	10k-100k	22.0	50.2
OPT_30B	ResNet34	100k+	21.9	53.1
OPT_30B	ResNet50	0-10k	14.1	30.2
OPT_30B	ResNet50	10k-100k	23.7	53.8

Table 11 continued from previous page

LM	VM	Bins	P@10	P@100
OPT_30B	ResNet50	100k+	24.6	57.7
OPT_30B	SegFormer-B0	0-10k	10.6	24.1
OPT_30B	SegFormer-B0	10k-100k	14.6	37.7
OPT_30B	SegFormer-B0	100k+	13.0	39.4
OPT_30B	SegFormer-B1	0-10k	12.6	25.5
OPT_30B	SegFormer-B1	10k-100k	19.8	47.4
OPT_30B	SegFormer-B1	100k+	19.2	49.4
OPT_30B	SegFormer-B2	0-10k	12.7	26.8
OPT_30B	SegFormer-B2	10k-100k	21.4	49.4
OPT_30B	SegFormer-B2	100k+	20.5	51.3
OPT_30B	SegFormer-B3	0-10k	13.0	25.7
OPT_30B	SegFormer-B3	10k-100k	22.4	51.8
OPT_30B	SegFormer-B3	100k+	21.1	52.9
OPT_30B	SegFormer-B4	0-10k	13.1	26.0
OPT_30B	SegFormer-B4	10k-100k	20.9	50.7
OPT_30B	SegFormer-B4	100k+	20.5	52.6
OPT_30B	SegFormer-B5	0-10k	12.7	26.8
OPT_30B	SegFormer-B5	10k-100k	20.6	49.5
OPT_30B	SegFormer-B5	100k+	20.8	51.9

Table 11: All frequency results (Exclude-1k set).

Model	Explained Variance Ratio (Sum)						
	256	512	768	1024	1280	2048	Max
MAE_{Huge}	0.9735	0.9922	0.9975	0.9994	1.0000	-	1.0000
ResNet152	0.9795	0.9942	0.9974	0.9987	0.9993	1.0000	1.0000
SegFormer-B5	0.9685	1.0000	-	-	-	-	1.0000
$LLaMA\text{-}2_{13B}$	0.5708	0.6662	0.7277	0.7725	0.8077	0.8814	0.8814
OPT_{30B}	0.4926	0.6002	0.6664	0.7164	0.7554	0.8360	0.8360

Table 12: Sum of explained variance ratios for different models and sizes.

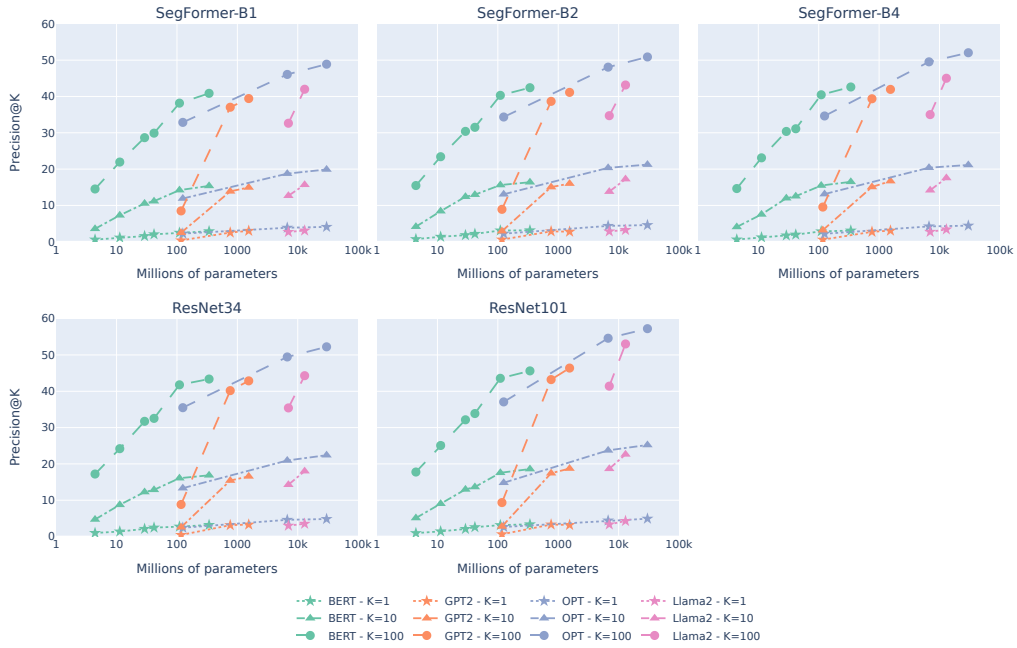

Figure 5: LMs converge toward the geometry of remaining visual models as they grow larger (Exclude-1K set).

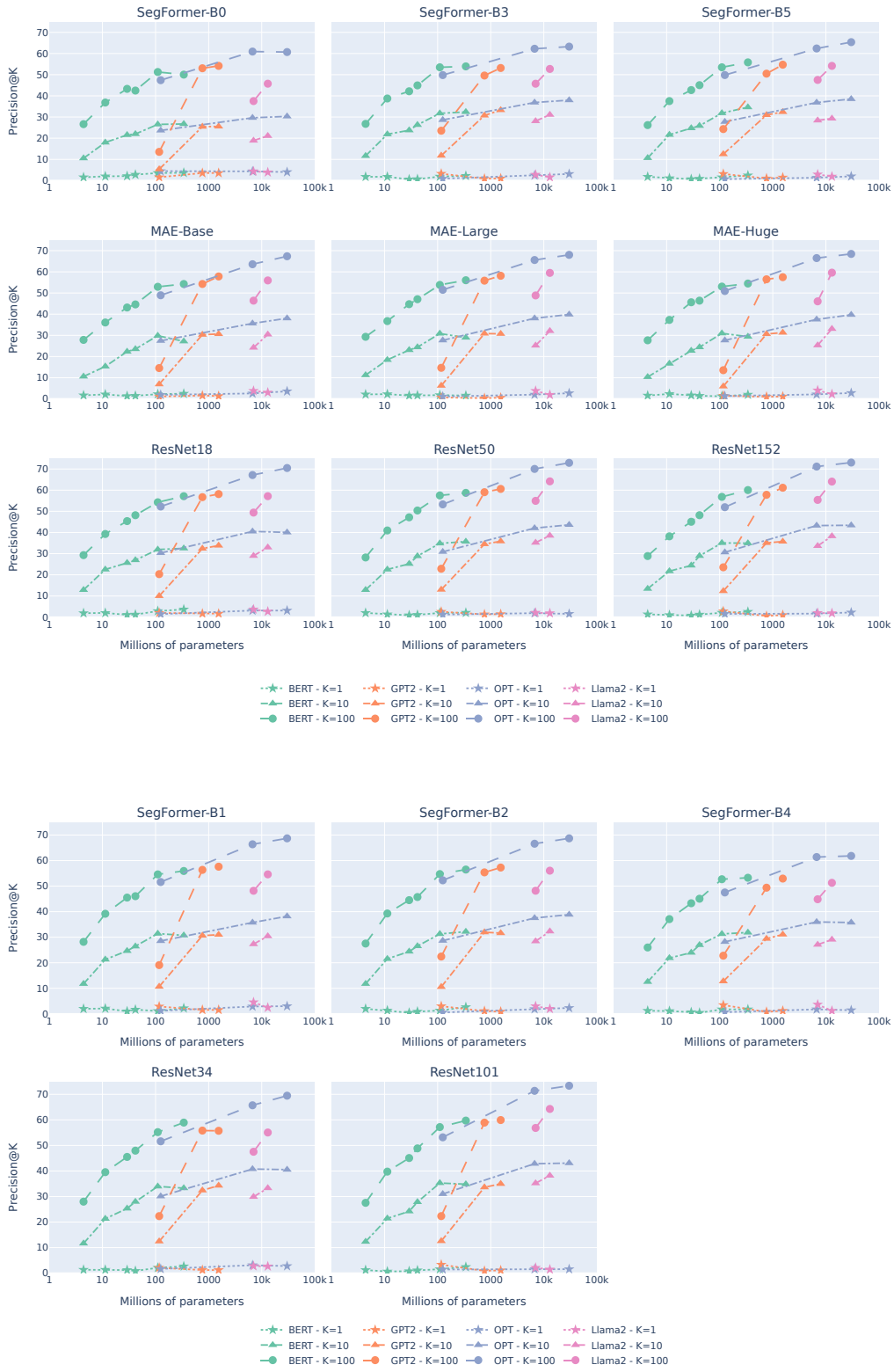

Figure 6: LMs converge toward the geometry of remaining visual models as they grow larger (Only-1K set).