

A. More few-shot transfer results

We observe similar scaling laws on more datasets, including Oxford IIIT Pets [27], CIFAR-100 [23], and Caltech-UCSD Birds [46]. The results are presented in Figure 9.

B. Pre-training details

We pre-train all the ViT models using adafactor optimizer with half-precision momentum. We use the default $\beta_1 = 0.9$ and $\beta_2 = 0.999$ (clipping threshold) for adafactor. We use batch size 4096 for all the models smaller than *ViT-g*. For models *ViT-g* and *ViT-G*, to speed up the training, we scale up the batch size at most to 32 768 and distribute the training to 2048 TPUv3 chips. We set weight decay to 3.0 for the “head” and 0.03 for the “body”. All the models are pre-trained at resolution 224×224 , with inception crop followed by random horizontal flip pre-process. We use reciprocal square-root schedule with a linear learning rate warmup of 10k steps. We cooldown the training at multiple steps as noted in Tables from Section F.

C. Adaptation details

We report both the few-shot linear regression and finetune results on mutiple datasets. For few-shot linear regression, we simply solve the l_2 -regularized linear regression problem, using the frozen embeddings extracted from 224×224 resolution images.

For finetune evaluation, we use SGD optimizer with momentum. We use batch size 512 and gradient clipping at global norm 1. We do not use weight decay for finetuning. Following [22, 42], we use higher resolution for finetuning. More specifically, we use 384×384 resolution for ViT models smaller than *ViT-g*, and 518×518 resolution for both *ViT-g* and *ViT-G*. We use Polyak averaging [29] only for the *ViT-G* model during fine-tuning, similar to [15]. We use a cosine learning rate schedule for 20k steps by default, except a flat learning rate for *ViT-G* with Polyak averaging. We linearly warm up the learning rate for 500 steps. We sweep over two learning rates $\{0.03, 0.01\}$ and choose the better one using a held-out 2% training split. On VTAB tasks, we use a fixed 0.01 learning rate with a cosine learning rate schedule. We train for 2 500 steps in total.

D. Impact of resolution and patch size

In this section, we answer the question of “what happens if we scale up the resolution, while keeping number of tokens fixed?”. We perform experiments on ImageNet-21K to verify this point, by scaling the resolution and patch size linearly together. We observed in Table 3 that the quality difference is pretty subtle if we increase patch and resolution together. What matters for ViT architecture is the total number of

patches, which has already been covered in Table 2 with different patch sizes: 32, 28, 16, and 14.

Table 3. Results of different resolutions and patch sizes.

Model	B/32	B/48	B/64	S/16	S/24	S/32
Res.	224	336	448	224	336	448
INet10	64.43	64.65	64.67	63.42	63.79	63.50

E. Full table of few-shot results

We provide the 5-shot learning and 10-shot learning results, on the four datasets from Figure 9. Both *ViT-g/14* and *ViT-G/14* are summarized in Table 4. All the other smaller ViT models are summarized in Table 5, Table 6, Table 7, Table 8, Table 9, Table 10, Table 11, Table 12, Table 13. We are aware of a few missing rows, which do not affect the trend for the scaling laws plot.

F. Full table of finetune results

We provide the finetune results on ImageNet, as well as the results evaluated on the other two ImageNet V2 and ImageNet ReaL test splits. Results for all the models could be found from Table 14, Table 15, Table 16, Table 17, Table 18, Table 19, Table 20, Table 21, Table 22. We are aware of a few missing rows, which do not affect the trend for the scaling laws plot. We show the total steps and the cooldown steps for each model, as well as the best finetune learning rate selected on ImageNet held-out 2% training split.

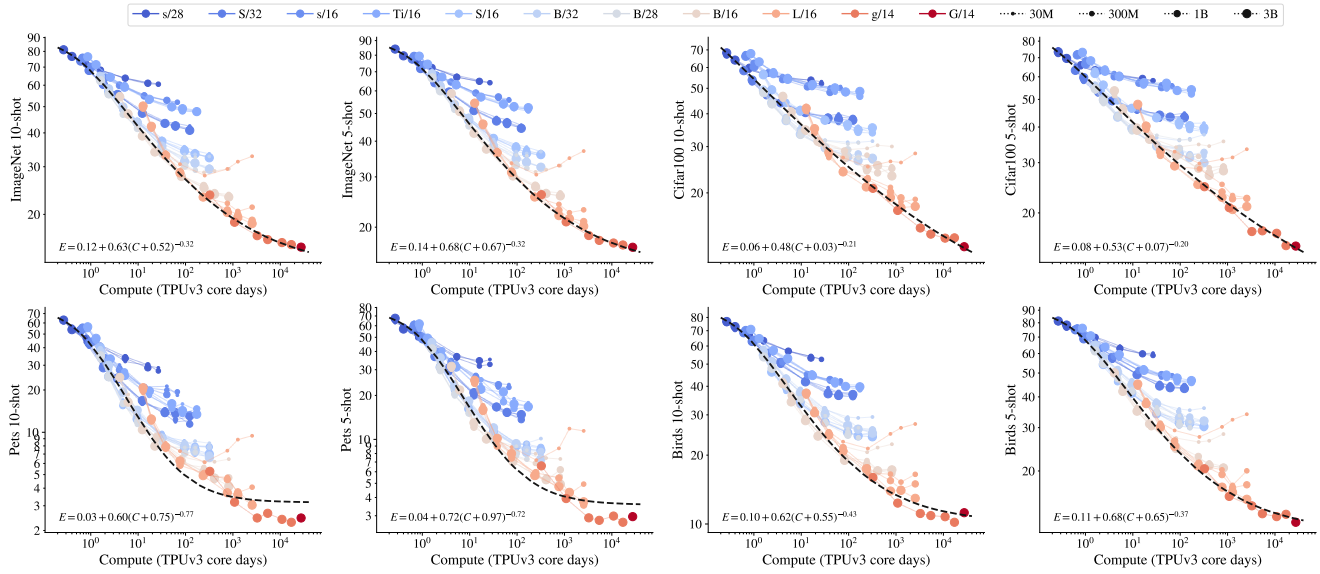


Figure 9. Representation quality as a function of total training compute. Representation quality is measured as few-shot error rate on four datasets. Sometimes, like in pets 5/10shot, the law does not fit the evidence perfectly, maybe the models are not ideal or the law is not universal.

Table 4. Tabular representation of the few-shot results (%) for model ViT-g/14 and ViT-G/14.

Data Size	Steps	INet5	INet10	Cifar5	Cifar10	Pets5	Pets10	Birds5	Birds10
<i>ViT-g/14</i>									
3B	120K	74.0	76.4	75.3	79.2	93.4	94.7	79.6	84.0
3B	400K	79.1	81.3	79.1	82.9	96.1	96.8	84.2	87.7
3B	1.2M	81.3	83.3	82.8	85.3	97.1	97.5	85.6	88.9
3B	2M	82.0	83.9	82.7	86.1	97.2	97.3	86.6	89.1
3B	4M	82.4	84.3	83.0	86.5	97.0	97.6	87.0	89.2
3B	6.3M	82.7	84.5	84.6	86.5	97.3	97.7	86.7	89.8
<i>ViT-G/14</i>									
3B	5M	83.0	84.9	84.7	87.5	97.0	97.5	87.6	88.8

Table 5. Tabular representation of the few-shot results (%) for model *L/16*.

Data Size	Steps	INet5	INet10	Cifar5	Cifar10	Pets5	Pets10	Birds5	Birds10
30M	20K	45.5	49.6	52.3	58.0	74.4	81.1	55.1	63.0
30M	30K	53.3	57.4	58.8	64.5	82.8	87.0	61.1	68.3
30M	60K	62.6	66.0	66.2	71.3	90.3	91.6	68.6	74.6
30M	120K	66.5	69.3	67.2	72.8	91.2	92.7	70.8	77.1
30M	400K	69.5	72.2	70.2	74.8	92.8	93.9	72.5	78.6
30M	1.2M	67.1	70.6	69.2	73.6	91.3	92.6	69.8	75.2
30M	2M	65.1	68.7	68.7	73.9	88.2	91.1	67.7	73.7
30M	4M	63.1	67.2	66.6	71.7	88.6	90.6	66.0	72.6
300M	20K	45.0	49.7	52.1	57.3	74.4	79.1	55.5	63.2
300M	30K	54.0	57.4	60.5	65.5	83.1	87.5	61.3	68.3
300M	60K	63.0	66.5	68.0	72.8	90.2	92.2	68.7	74.7
300M	120K	68.5	71.4	70.9	75.4	92.1	93.7	74.1	79.5
300M	400K	74.6	77.0	71.5	77.1	94.2	95.1	78.8	83.3
300M	1.2M	76.0	78.0	74.8	79.0	95.4	95.2	79.8	83.5
300M	2M	77.5	79.5	77.0	81.5	95.8	96.3	82.5	84.8
300M	4M	77.0	78.7	77.4	81.1	95.1	95.9	80.0	83.5
1B	20K	45.9	50.6	52.6	58.9	75.8	80.2	55.8	63.4
1B	30K	54.7	58.4	60.8	66.0	84.5	88.0	62.4	69.5
1B	60K	63.4	66.9	68.2	72.1	91.0	92.3	69.3	75.0
1B	120K	68.5	71.3	70.7	75.7	92.5	94.2	73.5	78.4
1B	400K	74.6	76.9	74.7	77.4	94.1	94.9	78.5	82.4
1B	1.2M	77.1	79.2	76.3	79.8	94.6	95.4	80.7	84.5
1B	2M	78.5	80.0	77.6	80.8	95.7	96.2	82.9	85.8
1B	4M	79.1	81.0	77.5	82.3	96.5	96.9	82.4	85.0
3B	20K	45.7	49.7	51.9	58.1	74.6	79.2	54.9	62.6
3B	30K	54.1	57.7	59.7	64.9	84.0	87.6	62.4	69.4
3B	60K	63.6	66.9	67.2	71.5	89.8	92.1	69.9	75.5
3B	120K	68.9	71.5	70.8	76.0	92.0	93.8	74.8	79.2
3B	400K	74.5	76.8	74.8	78.9	94.4	95.1	79.0	82.6
3B	1.2M	78.0	79.7	77.3	80.8	95.3	96.3	82.8	86.1
3B	2M	78.6	80.5	79.4	82.4	95.7	96.4	83.6	86.0
3B	4M	79.8	81.5	78.9	82.2	96.3	97.0	84.7	87.1

Table 6. Tabular representation of the few-shot results (%) for model *B/16*.

Data Size	Steps	INet5	INet10	Cifar5	Cifar10	Pets5	Pets10	Birds5	Birds10
30M	20K	40.8	45.4	47.2	53.5	68.6	74.8	49.3	57.8
30M	30K	48.0	52.2	53.1	58.7	79.1	81.4	56.7	64.5
30M	60K	56.6	60.5	60.1	64.8	87.0	88.6	64.1	71.3
30M	120K	61.6	65.4	63.4	68.4	89.1	90.8	69.0	75.1
30M	400K	67.1	70.4	66.0	70.8	91.5	93.3	72.8	78.2
30M	1.2M	68.8	71.4	65.2	70.5	92.0	93.8	73.8	79.4
30M	2M	68.2	71.1	65.7	69.7	92.5	93.8	73.5	78.4
30M	4M	67.2	70.5	64.4	70.0	92.7	93.6	71.8	77.7
300M	20K	41.0	45.6	48.4	54.6	69.7	76.1	49.9	57.6
300M	30K	48.6	52.6	53.5	58.7	78.5	83.1	56.5	64.2
300M	60K	57.3	60.6	60.0	65.9	86.8	89.8	63.7	70.8
300M	120K	62.3	65.5	63.8	69.4	89.3	91.1	68.9	75.3
300M	400K	69.3	72.1	68.8	73.1	92.3	94.1	75.5	80.5
300M	1.2M	72.3	74.9	68.4	72.2	93.1	94.4	77.7	81.1
300M	2M	73.3	76.0	71.2	74.7	94.5	95.5	78.4	82.7
300M	4M	73.6	76.3	70.8	74.9	94.4	95.4	79.2	82.7
1B	20K	40.9	45.0	47.0	53.0	69.3	75.8	49.4	58.1
1B	30K	48.5	52.6	53.2	59.1	78.3	83.4	56.6	63.6
1B	60K	57.3	61.0	60.0	65.2	87.9	89.6	65.0	71.0
1B	120K	62.3	65.9	63.4	68.8	89.5	90.6	69.8	75.2
1B	400K	69.2	72.2	67.9	71.1	93.0	93.6	74.6	80.5
1B	1.2M	71.9	74.4	71.5	75.7	94.3	94.8	77.1	81.6
1B	2M	73.8	75.9	72.2	76.2	94.8	95.2	79.0	82.8
1B	4M	74.3	76.7	71.1	75.2	93.8	95.3	79.3	83.4
3B	20K	41.3	45.7	46.4	53.5	68.3	75.4	50.0	58.9
3B	30K	49.2	53.1	54.0	59.5	78.2	83.6	57.9	65.8
3B	60K	57.4	61.0	61.0	65.3	87.0	89.6	65.0	71.7
3B	120K	62.5	66.0	63.8	68.6	90.0	92.0	68.9	75.8
3B	400K	69.8	72.4	68.2	72.9	93.0	93.8	75.3	81.0
3B	1.2M	72.3	74.9	71.3	75.3	94.3	94.7	78.9	83.2
3B	2M	73.8	76.3	72.9	74.6	94.7	95.3	79.0	82.9
3B	4M	74.3	76.8	71.9	76.8	95.1	95.9	79.4	82.8

Table 7. Tabular representation of the few-shot results (%) for model *B/28*.

Data Size	Steps	INet5	INet10	Cifar5	Cifar10	Pets5	Pets10	Birds5	Birds10
30M	20K	33.9	37.8	46.9	52.3	60.5	66.2	38.7	46.7
30M	30K	40.1	44.3	51.3	57.1	70.7	74.2	45.1	51.9
30M	60K	48.6	52.9	56.8	61.9	80.0	84.4	52.3	60.2
30M	120K	54.2	58.3	59.6	66.0	84.1	87.2	56.7	64.6
30M	400K	61.4	64.8	62.6	68.4	90.2	92.1	63.6	70.3
30M	1.2M	64.2	67.4	63.9	69.3	91.0	92.2	66.4	73.9
30M	2M	64.4	67.5	64.0	69.0	91.6	92.2	68.4	73.7
300M	20K	33.6	37.5	44.7	51.1	58.3	67.2	38.5	46.1
300M	30K	40.0	44.6	51.7	57.0	70.4	75.1	44.2	52.0
300M	60K	48.2	52.8	55.9	61.3	80.2	83.8	52.0	59.7
300M	120K	54.4	58.3	60.9	65.9	84.2	88.6	57.6	65.5
300M	400K	63.1	66.1	65.8	70.8	90.2	91.4	64.5	71.9
300M	1.2M	66.5	69.6	68.2	72.1	92.3	92.9	68.4	74.7
300M	2M	67.9	70.9	68.1	72.7	92.7	92.8	70.0	76.2
1B	20K	33.6	37.9	45.6	51.8	58.9	64.5	38.7	46.0
1B	30K	39.8	44.6	50.9	56.6	70.5	75.5	45.0	51.4
1B	60K	48.3	53.1	56.6	62.4	79.6	84.0	52.6	60.1
1B	120K	54.8	58.5	61.2	66.9	84.9	88.1	57.9	65.0
1B	400K	63.1	66.6	65.3	70.2	89.9	91.5	65.5	72.4
1B	1.2M	67.1	69.8	66.3	70.8	92.0	92.9	69.2	75.6
3B	20K	33.3	37.6	45.6	51.9	58.3	65.1	38.5	46.3
3B	30K	40.0	44.2	50.5	56.4	69.0	75.1	45.4	52.1

Table 8. Tabular representation of the few-shot results (%) for model *B/32*.

Data Size	Steps	INet5	INet10	Cifar5	Cifar10	Pets5	Pets10	Birds5	Birds10
30M	20K	30.6	34.8	44.2	50.2	54.7	62.6	34.7	42.0
30M	30K	37.5	41.4	49.5	55.0	66.9	72.0	41.9	49.1
30M	60K	46.1	50.0	55.9	60.7	77.0	81.0	48.8	57.1
30M	120K	51.7	55.8	59.7	64.2	82.2	85.2	54.1	62.1
30M	400K	59.5	63.4	63.3	68.3	88.5	90.6	62.3	68.3
30M	1.2M	63.2	66.7	64.1	68.8	90.8	91.8	64.3	70.7
30M	2M	63.8	66.7	63.6	68.8	90.6	91.3	64.7	70.9
30M	4M	63.2	67.3	62.4	68.5	89.9	91.6	64.0	70.6
300M	20K	30.5	35.2	45.1	50.0	54.7	60.6	35.8	42.7
300M	30K	37.3	41.0	50.2	55.6	64.4	70.2	40.8	48.0
300M	60K	45.7	49.9	56.6	62.2	75.6	80.8	48.8	56.1
300M	120K	51.9	55.8	61.2	66.0	82.3	86.2	53.5	61.2
300M	400K	60.1	64.1	65.4	70.6	88.9	90.7	61.5	68.1
300M	1.2M	64.0	67.7	66.3	71.4	90.9	91.9	65.1	71.2
300M	2M	66.4	69.3	67.7	73.1	91.6	92.1	67.5	73.7
300M	4M	67.5	70.1	68.4	73.1	91.7	92.2	68.3	74.1
1B	20K	30.6	35.2	43.9	50.4	57.1	61.9	35.1	41.9
1B	30K	37.1	41.9	49.5	55.4	65.3	71.8	41.6	48.5
1B	60K	46.3	50.2	56.0	61.3	76.2	80.3	48.3	56.7
1B	120K	51.9	55.9	60.8	65.2	81.5	85.7	54.6	62.3
1B	400K	60.8	64.5	65.7	70.3	88.6	90.7	62.5	69.1
1B	1.2M	65.1	68.1	66.6	72.3	90.7	91.7	66.6	72.9
1B	2M	66.1	69.4	68.1	71.8	91.4	92.7	66.9	74.1
1B	4M	67.5	70.7	67.4	73.3	91.7	93.1	68.2	74.2
3B	20K	31.5	35.2	44.9	51.0	53.9	62.7	35.1	43.6
3B	30K	37.6	41.8	50.1	55.9	63.7	70.0	42.0	48.8
3B	60K	46.2	50.1	56.7	62.3	77.0	80.7	49.6	57.1
3B	120K	52.0	56.4	60.5	66.6	81.7	85.6	55.2	62.5
3B	400K	61.6	64.3	65.7	70.3	88.0	90.9	62.5	69.8
3B	1.2M	65.3	68.7	67.7	72.6	90.6	92.0	67.3	73.0
3B	2M	66.2	69.1	68.7	73.5	91.9	92.7	68.3	74.0
3B	4M	67.6	70.6	70.0	72.9	92.7	93.4	68.1	75.1

Table 9. Tabular representation of the few-shot results (%) for model *S/16*.

Data Size	Steps	INet5	INet10	Cifar5	Cifar10	Pets5	Pets10	Birds5	Birds10
30M	20K	32.1	36.4	39.2	44.5	56.1	62.3	38.4	45.8
30M	30K	39.1	43.3	46.3	51.4	67.2	74.1	46.8	53.7
30M	60K	47.1	51.2	50.7	55.7	78.4	83.1	54.0	61.6
30M	120K	51.8	55.4	53.8	59.1	82.2	86.1	58.8	66.3
30M	400K	58.4	61.8	57.2	62.6	88.3	90.6	65.0	71.6
30M	1.2M	61.1	64.4	57.1	63.2	90.5	91.3	68.4	73.7
30M	2M	61.9	65.6	58.8	64.1	91.0	92.0	68.0	74.0
30M	4M	63.2	66.5	58.7	65.2	91.3	92.1	69.7	74.7
300M	20K	31.3	36.4	38.9	43.9	57.4	62.2	37.4	45.4
300M	60K	47.4	51.1	50.8	56.6	78.3	83.8	53.4	62.0
300M	120K	52.6	56.4	53.0	58.1	83.1	87.0	58.7	66.2
300M	400K	58.9	62.7	56.1	61.5	88.6	90.2	65.6	72.2
300M	1.2M	62.3	66.0	58.2	63.9	90.9	91.8	69.5	74.9
300M	2M	63.3	66.4	60.1	64.9	91.9	93.0	69.9	75.5
300M	4M	64.5	67.5	61.1	65.9	92.1	93.3	70.3	75.4
1B	20K	31.8	35.8	37.1	42.9	56.1	63.3	38.2	46.0
1B	30K	39.2	43.1	44.2	50.3	68.1	75.1	44.6	52.6
1B	60K	47.2	51.3	50.6	55.8	78.2	84.5	53.9	60.8
1B	120K	51.5	55.8	53.6	59.4	83.3	86.8	59.0	66.8
1B	400K	58.9	62.6	56.7	62.3	88.0	90.6	65.7	72.2
1B	1.2M	61.5	65.2	59.6	64.4	90.7	92.1	67.3	75.0
1B	2M	62.8	66.6	60.8	66.0	90.7	92.0	69.1	75.5
1B	4M	64.0	67.4	61.4	66.2	91.2	92.1	69.5	74.9
3B	20K	32.3	36.5	38.4	43.8	56.2	59.7	37.7	45.4
3B	30K	38.8	43.1	43.8	50.7	68.8	75.1	45.9	53.9
3B	120K	52.6	56.3	53.5	58.8	83.8	87.6	58.5	66.0
3B	400K	59.1	62.7	56.9	62.2	88.7	90.8	65.8	72.1
3B	1.2M	62.1	65.7	58.7	63.6	91.0	92.2	68.6	74.9
3B	2M	63.7	66.5	59.5	65.3	91.4	92.5	68.8	75.7
3B	4M	64.1	67.6	60.3	64.6	91.6	93.0	69.7	75.9

Table 10. Tabular representation of the few-shot results (%) for model *T/16*.

Data Size	Steps	INet5	INet10	Cifar5	Cifar10	Pets5	Pets10	Birds5	Birds10
30M	20K	20.2	23.5	26.9	32.0	37.8	44.6	24.2	29.2
30M	30K	25.7	28.6	31.9	38.0	50.1	54.9	29.7	35.5
30M	60K	32.9	36.0	36.6	41.7	61.9	65.7	36.9	43.1
30M	120K	36.5	40.6	39.1	44.3	67.2	74.2	41.1	48.0
30M	400K	42.3	46.7	41.7	48.2	76.1	82.0	47.4	55.2
30M	1.2M	45.6	49.8	45.4	50.6	80.8	84.9	52.3	58.9
30M	2M	47.4	50.5	45.1	51.0	81.0	84.1	53.6	59.4
30M	4M	48.2	51.6	46.3	52.7	82.2	85.6	53.3	59.5
300M	20K	20.7	23.7	28.0	32.6	43.3	45.3	23.8	30.0
300M	30K	25.8	28.8	32.2	36.8	49.7	55.0	29.6	35.6
300M	60K	33.1	36.4	37.6	42.7	62.2	68.2	37.3	44.0
300M	120K	37.3	41.1	39.9	45.4	67.0	75.1	42.6	48.2
300M	400K	43.2	47.3	42.9	49.3	74.4	81.7	47.7	55.6
300M	1.2M	46.4	50.8	45.6	51.7	81.7	85.4	51.3	59.1
300M	2M	48.0	51.6	46.1	51.8	82.3	85.7	53.3	60.2
300M	4M	49.0	51.9	46.6	52.6	83.1	86.3	54.3	61.4
1B	20K	20.4	23.5	27.7	32.8	40.5	45.3	24.0	29.8
1B	30K	26.0	28.6	31.7	37.5	54.3	54.9	29.5	35.4
1B	60K	32.7	36.0	36.1	42.1	59.5	66.7	36.1	42.8
1B	120K	36.4	40.2	39.2	45.0	68.2	73.1	41.3	47.8
1B	400K	42.9	47.2	43.9	49.3	77.8	80.8	47.9	54.6
1B	1.2M	46.4	49.9	44.9	50.2	81.9	85.1	52.1	59.1
1B	2M	47.5	51.8	46.3	51.9	83.7	86.4	53.9	60.2
1B	4M	48.3	52.2	47.8	53.4	83.5	85.3	54.3	60.2
3B	20K	20.6	23.6	26.9	32.2	38.2	43.6	24.0	29.0
3B	30K	25.6	28.5	31.7	36.9	50.6	53.4	29.2	35.3
3B	60K	32.9	35.7	37.3	43.1	63.1	66.7	36.2	42.5
3B	120K	37.1	41.0	40.0	45.9	68.8	74.6	40.5	47.0
3B	400K	42.9	46.7	42.7	48.3	78.0	80.3	48.7	55.4
3B	1.2M	46.0	50.1	43.0	49.8	78.9	84.0	50.9	58.2
3B	2M	47.1	50.8	46.3	51.6	82.5	85.7	52.3	60.0
3B	4M	47.6	52.1	45.9	51.3	83.2	86.6	53.4	60.2

Table 11. Tabular representation of the few-shot results (%) for model *s/16*.

Data Size	Steps	INet5	INet10	Cifar5	Cifar10	Pets5	Pets10	Birds5	Birds10
30M	20K	20.3	23.8	27.2	32.7	36.8	47.1	24.2	30.0
30M	30K	26.3	29.6	32.3	37.9	47.6	54.4	30.2	37.4
30M	60K	32.7	36.0	36.9	43.6	58.9	65.3	36.6	44.0
30M	120K	36.2	39.7	39.0	45.7	65.0	69.9	42.2	48.7
30M	400K	40.9	44.9	41.9	48.1	74.2	79.0	48.7	55.3
30M	1.2M	43.2	47.6	43.7	49.5	75.9	78.7	51.3	57.3
30M	2M	43.7	48.2	43.8	49.2	76.3	81.6	51.2	59.0
300M	20K	20.4	23.7	28.1	32.7	42.3	43.9	24.6	29.7
300M	30K	25.5	29.3	33.5	39.2	49.6	55.0	30.2	36.1
300M	60K	31.4	35.3	37.5	43.4	59.3	66.7	37.0	43.8
300M	120K	35.8	39.4	38.4	44.6	65.3	71.2	41.6	48.9
300M	400K	40.7	44.7	42.5	48.8	72.6	79.2	48.6	55.6
300M	1.2M	43.5	47.3	43.8	50.1	77.2	80.0	51.8	57.8
300M	2M	44.3	48.1	44.4	50.2	77.3	80.6	51.2	57.8
1B	20K	20.6	24.0	27.8	32.8	38.5	44.1	23.7	30.1
1B	30K	26.1	29.9	31.7	37.7	49.3	55.9	30.3	37.3
1B	60K	32.1	36.3	36.8	42.6	60.3	66.1	37.0	43.6
1B	120K	35.5	40.1	40.0	46.1	66.0	72.2	41.6	48.9
1B	400K	41.0	45.2	43.0	49.4	73.3	79.1	48.8	55.0
1B	1.2M	42.8	47.2	45.0	51.8	76.3	81.5	50.7	57.2
3B	20K	20.7	24.5	28.4	33.8	39.7	44.9	24.8	29.8
3B	30K	26.2	30.0	33.5	39.6	51.2	56.2	29.5	36.4

Table 12. Tabular representation of the few-shot results (%) for model $S/32$.

Data Size	Steps	INet5	INet10	Cifar5	Cifar10	Pets5	Pets10	Birds5	Birds10
30M	20K	23.0	26.7	34.1	40.3	43.3	48.1	26.1	32.2
30M	30K	28.8	32.6	40.0	46.3	51.9	59.6	31.2	37.1
30M	60K	36.4	39.9	45.9	51.8	64.5	71.5	39.0	45.2
30M	120K	40.3	44.4	49.8	55.7	71.0	76.9	43.7	50.7
30M	400K	47.9	51.7	54.8	60.1	79.5	83.1	50.5	57.6
30M	1.2M	51.6	55.9	55.2	60.7	84.0	86.6	54.3	60.6
30M	2M	52.5	56.6	56.0	60.1	84.9	88.1	55.7	62.5
30M	4M	54.5	57.6	55.3	61.3	86.0	87.9	56.7	63.2
300M	20K	23.1	27.0	36.1	41.5	42.6	47.0	25.8	32.0
300M	30K	28.4	31.9	42.2	47.1	48.9	58.9	30.3	36.6
300M	60K	35.0	39.3	47.7	52.5	62.9	69.5	36.9	44.8
300M	120K	40.8	44.9	50.4	55.4	71.4	75.6	43.4	50.3
300M	400K	48.4	52.4	54.5	59.9	79.3	83.9	50.6	57.5
300M	1.2M	52.9	56.2	57.3	62.6	83.0	85.5	54.6	61.8
300M	2M	53.4	57.4	57.1	62.9	84.5	87.7	55.2	62.4
300M	4M	55.2	58.5	57.6	62.8	85.4	87.1	55.5	62.6
1B	30K	28.3	32.2	41.4	47.1	50.2	56.2	29.9	36.6
1B	60K	35.7	39.7	47.1	53.1	63.4	70.0	36.9	44.7
1B	120K	40.8	44.7	50.7	56.0	68.6	75.3	43.2	50.1
1B	400K	48.3	52.4	54.0	59.6	80.2	83.4	50.5	57.7
1B	1.2M	52.6	56.7	55.8	61.1	83.2	86.4	55.7	62.4
1B	2M	54.3	58.0	56.7	61.2	84.9	86.6	56.3	63.7
1B	4M	55.4	58.8	56.3	61.6	86.4	88.6	56.5	64.0
3B	20K	22.5	26.3	36.8	41.7	43.4	46.8	25.1	31.5
3B	30K	28.2	32.0	40.4	46.2	49.6	56.9	31.2	37.2
3B	60K	36.0	39.6	46.5	52.0	63.1	71.2	37.5	44.8
3B	120K	40.5	44.5	50.6	56.0	70.0	75.9	42.1	49.1
3B	400K	48.8	52.8	53.5	59.6	79.1	83.2	50.8	58.2
3B	1.2M	52.9	56.6	55.9	61.6	83.3	86.1	55.7	62.9
3B	2M	53.6	57.5	56.6	62.0	84.7	86.2	56.4	63.5
3B	4M	55.6	59.1	56.8	62.2	85.2	86.9	56.8	63.2

Table 13. Tabular representation of the few-shot results (%) for model *s/28*.

Data Size	Steps	INet5	INet10	Cifar5	Cifar10	Pets5	Pets10	Birds5	Birds10
30M	20K	16.0	18.9	24.9	31.9	37.0	36.5	18.6	24.6
30M	30K	20.3	23.4	30.5	35.9	40.4	46.8	23.0	28.2
30M	60K	24.6	28.4	34.7	41.1	48.7	54.4	28.2	34.3
30M	120K	27.7	32.0	37.4	43.3	51.6	58.5	29.7	37.2
30M	400K	32.0	36.3	39.1	45.2	62.0	68.1	35.7	42.9
30M	1.2M	34.8	38.6	40.8	46.3	66.5	70.1	40.1	46.1
30M	2M	35.9	39.3	41.9	47.0	64.7	71.3	39.7	47.4
300M	20K	16.5	19.1	26.8	31.9	32.9	35.8	19.7	23.8
300M	30K	19.9	23.2	29.9	36.2	42.0	44.0	23.7	29.1
300M	60K	24.8	28.4	34.9	41.3	50.3	56.0	28.5	33.6
300M	120K	27.6	31.6	37.0	43.2	54.5	58.4	32.0	37.5
300M	400K	32.9	36.6	39.4	45.4	63.9	65.5	37.3	43.1
300M	1.2M	35.4	39.3	41.2	47.7	68.2	69.8	40.3	45.9
300M	2M	35.9	39.5	41.8	47.9	67.4	72.7	41.1	47.5
1B	20K	16.0	19.0	27.6	33.1	34.3	37.9	19.1	24.1
1B	30K	20.2	23.3	30.1	35.9	41.3	45.8	23.2	27.3
1B	60K	24.5	28.2	33.9	39.9	47.1	53.3	26.6	32.8
1B	120K	27.6	31.8	36.5	43.3	53.6	60.3	30.2	36.6
1B	400K	33.0	36.3	39.8	45.5	63.1	66.4	37.1	43.0
1B	1.2M	35.2	38.9	40.8	46.3	65.4	71.2	40.2	46.9
3B	20K	16.0	18.9	26.6	31.8	32.5	37.1	18.6	23.2
3B	30K	20.4	23.3	30.4	36.0	43.0	46.2	21.7	27.0

Table 14. Tabular representation of the finetune results (%) for model *ViT-L/16* on ImageNet, ImageNet V2 test set and ImageNet ReaL test set.

Data Size	Steps	Cooldown	LR	ImageNet	ImageNet V2	ImageNet ReaL
30M	20K	10K	0.03	75.4	63.3	82.1
30M	30K	10K	0.03	78.8	67.5	85.0
30M	60K	10K	0.03	82.4	72.5	87.6
30M	120K	50K	0.03	83.8	74.8	88.3
30M	400K	50K	0.03	85.5	76.5	89.0
30M	1.2M	50K	0.03	85.3	76.0	88.7
30M	2M	50K	0.03	85.1	76.2	88.7
30M	4M	50K	0.01	85.6	77.0	89.1
300M	20K	10K	0.03	75.1	63.5	81.9
300M	30K	10K	0.03	79.1	67.7	85.2
300M	60K	10K	0.03	82.7	72.9	87.9
300M	120K	50K	0.03	84.7	75.4	89.1
300M	400K	50K	0.03	86.5	77.5	89.8
300M	1.2M	50K	0.03	87.3	78.8	89.8
300M	2M	50K	0.03	87.7	78.6	89.8
300M	4M	50K	0.01	88.0	79.5	90.3
1B	20K	10K	0.03	75.9	63.9	82.7
1B	30K	10K	0.03	79.5	68.4	85.5
1B	60K	10K	0.03	82.5	72.6	87.8
1B	120K	50K	0.03	84.5	75.4	88.9
1B	400K	50K	0.03	86.7	78.3	89.8
1B	1.2M	50K	0.03	87.2	78.6	89.8
1B	2M	50K	0.03	87.9	78.9	90.0
1B	4M	50K	0.03	88.0	79.5	90.1
3B	20K	10K	0.03	75.5	63.0	82.2
3B	30K	10K	0.03	79.3	68.3	85.4
3B	60K	10K	0.03	82.7	73.5	87.7
3B	120K	50K	0.03	84.7	75.6	89.0
3B	400K	50K	0.03	87.0	78.5	90.1
3B	1.2M	50K	0.03	87.8	79.4	90.0
3B	2M	50K	0.03	87.9	79.6	90.0
3B	4M	50K	0.01	88.5	80.4	90.4

Table 15. Tabular representation of the finetune results (%) for model *ViT-B/16* on ImageNet, ImageNet V2 test set and ImageNet ReaL test set.

Data Size	Steps	Cooldown	LR	ImageNet	ImageNet V2	ImageNet ReaL
30M	20K	10K	0.03	73.0	60.4	80.0
30M	30K	10K	0.03	76.9	64.9	83.4
30M	60K	10K	0.03	80.5	69.5	86.1
30M	120K	50K	0.03	82.2	72.3	87.4
30M	400K	50K	0.03	84.4	74.6	88.5
30M	1.2M	50K	0.03	84.9	75.0	88.7
30M	2M	50K	0.03	84.8	74.8	88.6
30M	4M	50K	0.01	84.9	75.3	88.8
300M	20K	10K	0.03	73.5	61.0	80.5
300M	30K	10K	0.03	77.2	65.2	83.8
300M	60K	10K	0.03	80.6	69.9	86.3
300M	120K	50K	0.03	82.3	72.5	87.5
300M	400K	50K	0.03	84.9	75.5	89.0
300M	1.2M	50K	0.03	86.0	76.7	89.4
300M	2M	50K	0.01	86.2	76.8	89.5
300M	4M	50K	0.01	86.7	77.6	89.7
1B	20K	10K	0.03	73.2	60.7	80.2
1B	30K	10K	0.03	77.0	65.7	83.6
1B	60K	10K	0.03	80.6	70.7	86.4
1B	120K	50K	0.03	82.3	72.0	87.5
1B	400K	50K	0.03	85.1	75.2	89.1
1B	1.2M	50K	0.03	86.0	77.0	89.5
1B	2M	50K	0.03	86.5	77.3	89.6
1B	4M	50K	0.01	86.8	77.5	89.8
3B	20K	10K	0.03	73.4	61.0	80.4
3B	30K	10K	0.03	77.1	65.5	83.7
3B	60K	10K	0.03	80.5	70.0	86.2
3B	120K	50K	0.03	82.5	72.7	87.6
3B	400K	50K	0.03	85.1	75.7	89.1
3B	1.2M	50K	0.03	86.0	77.1	89.4
3B	2M	50K	0.03	86.3	77.0	89.6
3B	4M	50K	0.03	86.6	77.4	89.7

Table 16. Tabular representation of the finetune results (%) for model *ViT-B/28* on ImageNet, ImageNet V2 test set and ImageNet ReaL test set.

Data Size	Steps	Cooldown	LR	ImageNet	ImageNet V2	ImageNet ReaL
30M	20K	10K	0.03	68.8	55.6	76.1
30M	30K	10K	0.03	72.8	59.6	79.8
30M	60K	10K	0.03	76.7	64.5	83.4
30M	120K	50K	0.03	79.1	68.3	85.3
30M	400K	50K	0.03	82.2	72.1	87.4
30M	1.2M	50K	0.03	83.3	73.1	87.8
30M	2M	50K	0.03	83.5	73.4	87.8
300M	20K	10K	0.03	68.9	56.0	76.2
300M	30K	10K	0.03	72.8	60.2	80.0
300M	60K	10K	0.03	77.0	65.0	83.5
300M	120K	50K	0.03	79.4	68.2	85.3
300M	400K	50K	0.03	82.8	72.6	87.7
300M	1.2M	50K	0.03	84.1	74.6	88.5
300M	2M	50K	0.03	84.4	74.6	88.5
1B	20K	10K	0.03	68.6	55.5	75.9
1B	30K	10K	0.03	72.8	60.1	79.9
1B	60K	10K	0.03	76.9	65.1	83.6
1B	120K	50K	0.03	79.4	69.0	85.5
1B	400K	50K	0.03	82.7	73.0	87.6
1B	1.2M	50K	0.03	84.0	74.4	88.3
3B	20K	10K	0.03	68.8	55.3	75.9
3B	30K	10K	0.03	72.6	60.2	79.7

Table 17. Tabular representation of the finetune results (%) for model ViT-B/32 on ImageNet, ImageNet V2 test set and ImageNet ReaL test set.

Data Size	Steps	Cooldown	LR	ImageNet	ImageNet V2	ImageNet ReaL
30M	20K	10K	0.03	66.6	53.8	73.8
30M	30K	10K	0.03	71.0	57.9	78.0
30M	60K	10K	0.03	75.6	63.5	82.3
30M	120K	50K	0.03	78.0	66.4	84.3
30M	400K	50K	0.03	81.4	70.8	86.8
30M	1.2M	50K	0.03	82.7	72.4	87.5
30M	2M	50K	0.03	83.1	72.7	87.7
30M	4M	50K	0.01	83.0	72.8	87.7
300M	20K	10K	0.03	66.6	53.4	73.9
300M	30K	10K	0.03	70.8	58.0	78.0
300M	60K	10K	0.03	75.5	63.2	82.2
300M	120K	50K	0.03	78.3	66.7	84.5
300M	400K	50K	0.03	81.8	71.4	87.0
300M	1.2M	50K	0.03	83.3	73.4	87.9
300M	2M	50K	0.03	83.7	73.9	88.2
300M	4M	50K	0.01	83.9	74.3	88.3
1B	20K	10K	0.03	66.8	53.7	74.1
1B	30K	10K	0.03	71.1	58.5	78.1
1B	60K	10K	0.03	75.5	63.1	82.2
1B	120K	50K	0.03	78.5	66.9	84.7
1B	400K	50K	0.03	82.0	71.6	87.2
1B	1.2M	50K	0.03	83.4	73.5	87.9
1B	2M	50K	0.03	83.7	73.9	88.1
1B	4M	50K	0.03	84.1	74.4	88.4
3B	20K	10K	0.03	66.7	53.7	73.9
3B	30K	10K	0.03	71.0	58.4	78.1
3B	60K	10K	0.03	75.6	63.4	82.3
3B	120K	50K	0.03	78.3	67.3	84.6
3B	400K	50K	0.03	82.2	71.8	87.1
3B	1.2M	50K	0.03	83.5	73.5	87.9
3B	2M	50K	0.03	83.8	74.0	88.2
3B	4M	50K	0.03	84.1	74.4	88.2

Table 18. Tabular representation of the finetune results (%) for model ViT-S/16 on ImageNet, ImageNet V2 test set and ImageNet ReaL test set.

Data Size	Steps	Cooldown	LR	ImageNet	ImageNet V2	ImageNet ReaL
30M	20K	10K	0.03	67.4	54.5	74.7
30M	30K	10K	0.03	72.5	59.9	79.6
30M	60K	10K	0.03	76.8	65.0	83.2
30M	120K	50K	0.03	78.8	67.8	85.1
30M	400K	50K	0.03	81.5	70.9	87.1
30M	1.2M	50K	0.03	82.5	72.0	87.7
30M	2M	50K	0.03	82.8	72.2	87.8
30M	4M	50K	0.01	83.5	72.8	88.2
300M	20K	10K	0.03	67.8	54.8	75.0
300M	30K	10K	0.03	72.6	60.3	79.7
300M	60K	10K	0.03	76.8	65.3	83.4
300M	120K	50K	0.03	79.0	68.0	85.3
300M	400K	50K	0.03	81.7	71.2	87.3
300M	1.2M	50K	0.03	82.9	72.9	87.9
300M	2M	50K	0.01	83.3	73.4	88.3
300M	4M	50K	0.01	83.9	74.2	88.5
1B	20K	10K	0.03	67.3	54.5	74.6
1B	30K	10K	0.03	72.3	60.0	79.6
1B	60K	10K	0.03	76.6	64.9	83.4
1B	120K	50K	0.03	78.8	67.9	85.2
1B	400K	50K	0.03	81.9	70.6	87.3
1B	1.2M	50K	0.03	82.8	72.4	87.8
1B	2M	50K	0.01	83.2	72.8	88.2
1B	4M	50K	0.03	83.5	72.7	88.3
3B	20K	10K	0.03	67.5	54.5	74.9
3B	30K	10K	0.03	72.3	60.0	79.4
3B	60K	10K	0.03	76.7	64.8	83.3
3B	120K	50K	0.03	79.1	67.9	85.4
3B	400K	50K	0.03	81.9	71.1	87.3
3B	1.2M	50K	0.01	82.8	72.7	87.9
3B	2M	50K	0.03	83.2	73.2	88.1
3B	4M	50K	0.01	84.0	73.8	88.5

Table 19. Tabular representation of the finetune results (%) for model *ViT-Ti/16* on ImageNet, ImageNet V2 test set and ImageNet ReaL test set.

Data Size	Steps	Cooldown	LR	ImageNet	ImageNet V2	ImageNet ReaL
30M	20K	10K	0.03	55.5	43.6	62.5
30M	30K	10K	0.03	61.8	49.2	69.2
30M	60K	10K	0.03	67.8	55.2	75.5
30M	120K	50K	0.03	71.2	58.6	78.5
30M	400K	50K	0.03	74.9	62.8	82.1
30M	1.2M	50K	0.01	76.5	64.5	83.4
30M	2M	50K	0.03	76.7	64.7	83.4
30M	4M	50K	0.01	77.5	65.6	84.2
300M	20K	10K	0.03	55.9	43.7	62.9
300M	30K	10K	0.03	61.7	49.3	69.0
300M	60K	10K	0.03	68.5	55.7	76.0
300M	120K	50K	0.03	71.4	58.8	78.7
300M	400K	50K	0.03	75.2	62.8	82.2
300M	1.2M	50K	0.03	76.7	64.7	83.7
300M	2M	50K	0.01	77.1	65.5	84.1
300M	4M	50K	0.01	77.8	66.1	84.4
1B	20K	10K	0.03	55.8	43.2	62.8
1B	30K	10K	0.03	61.6	49.1	69.0
1B	60K	10K	0.03	67.9	54.8	75.4
1B	120K	50K	0.03	71.1	58.3	78.5
1B	400K	50K	0.03	74.9	63.0	82.1
1B	1.2M	50K	0.03	76.7	64.6	83.6
1B	2M	50K	0.01	77.1	65.4	83.8
1B	4M	50K	0.01	77.7	66.2	84.4
3B	20K	10K	0.03	55.6	43.3	62.5
3B	30K	10K	0.03	61.4	49.2	68.7
3B	60K	10K	0.03	68.1	55.5	75.7
3B	120K	50K	0.03	71.2	58.6	78.7
3B	400K	50K	0.03	75.0	62.8	82.1
3B	1.2M	50K	0.03	76.4	64.7	83.4
3B	2M	50K	0.03	76.9	64.7	83.7
3B	4M	50K	0.01	77.6	65.6	84.3

Table 20. Tabular representation of the finetune results (%) for model *ViT-s/16* on ImageNet, ImageNet V2 test set and ImageNet ReaL test set.

Data Size	Steps	Cooldown	LR	ImageNet	ImageNet V2	ImageNet ReaL
30M	20K	10K	0.03	56.0	43.2	63.0
30M	30K	10K	0.03	62.2	49.4	69.5
30M	60K	10K	0.03	67.8	54.8	75.3
30M	120K	50K	0.03	70.0	57.5	77.7
30M	400K	50K	0.03	73.7	60.9	81.0
30M	1.2M	50K	0.03	75.0	62.4	82.0
30M	2M	50K	0.01	75.2	63.0	82.3
300M	20K	10K	0.03	56.3	43.2	63.3
300M	30K	10K	0.03	62.0	49.5	69.4
300M	60K	10K	0.03	67.4	54.3	75.0
300M	120K	50K	0.03	70.1	57.8	77.6
300M	400K	50K	0.03	73.6	61.2	80.6
300M	1.2M	50K	0.03	74.9	62.8	82.0
300M	2M	50K	0.01	75.4	63.4	82.6
1B	20K	10K	0.03	56.2	44.1	63.2
1B	30K	10K	0.03	62.4	49.7	69.8
1B	60K	10K	0.03	68.0	54.9	75.6
1B	120K	50K	0.03	70.5	57.5	77.8
1B	400K	50K	0.03	73.9	61.6	81.1
1B	1.2M	50K	0.03	75.1	63.2	82.1
3B	20K	10K	0.03	56.4	43.6	63.3
3B	30K	10K	0.03	62.6	49.9	70.1

Table 21. Tabular representation of the finetune results (%) for model ViT-S/32 on ImageNet, ImageNet V2 test set and ImageNet ReaL test set.

Data Size	Steps	Cooldown	LR	ImageNet	ImageNet V2	ImageNet ReaL
30M	20K	10K	0.03	59.3	47.1	66.3
30M	30K	10K	0.03	64.3	51.8	71.7
30M	60K	10K	0.03	70.3	58.1	77.5
30M	120K	50K	0.03	73.4	61.2	80.5
30M	400K	50K	0.03	77.1	65.7	83.6
30M	1.2M	50K	0.03	79.0	67.3	84.9
30M	2M	50K	0.03	79.1	67.9	85.1
30M	4M	50K	0.01	79.7	68.2	85.6
300M	20K	10K	0.03	59.3	47.1	66.2
300M	30K	10K	0.03	64.2	51.0	71.5
300M	60K	10K	0.03	70.1	57.6	77.4
300M	120K	50K	0.03	73.4	60.5	80.6
300M	400K	50K	0.03	77.5	66.3	84.0
300M	1.2M	50K	0.03	79.0	67.9	85.1
300M	2M	50K	0.03	79.6	67.8	85.6
300M	4M	50K	0.03	79.9	68.5	85.8
1B	20K	10K	0.03	59.0	46.2	66.2
1B	30K	10K	0.03	64.0	51.4	71.4
1B	60K	10K	0.03	70.5	57.7	77.7
1B	120K	50K	0.03	73.6	60.8	80.7
1B	400K	50K	0.03	77.6	65.7	84.0
1B	1.2M	50K	0.03	79.5	68.0	85.5
1B	2M	50K	0.03	79.7	68.2	85.5
1B	4M	50K	0.03	80.2	68.1	85.9
3B	20K	10K	0.03	59.3	47.3	66.4
3B	30K	10K	0.03	64.3	51.5	71.6
3B	60K	10K	0.03	70.2	57.2	77.6
3B	120K	50K	0.03	73.5	61.3	80.7
3B	400K	50K	0.03	77.6	65.7	84.0
3B	1.2M	50K	0.03	79.4	67.4	85.4
3B	2M	50K	0.01	79.5	68.5	85.6
3B	4M	50K	0.01	79.9	69.4	86.0

Table 22. Tabular representation of the finetune results (%) for model *ViT-s/28* on ImageNet, ImageNet V2 test set and ImageNet ReaL test set.

Data Size	Steps	Cooldown	LR	ImageNet	ImageNet V2	ImageNet ReaL
30M	20K	10K	0.03	50.3	38.0	56.9
30M	30K	10K	0.03	55.8	43.4	62.8
30M	60K	10K	0.03	61.5	48.5	68.8
30M	120K	50K	0.03	64.1	51.4	71.6
30M	400K	50K	0.03	68.4	55.5	75.7
30M	1.2M	50K	0.03	69.8	57.2	77.4
30M	2M	50K	0.01	70.2	57.5	77.8
300M	20K	10K	0.03	50.3	38.2	56.9
300M	30K	10K	0.03	55.7	43.6	62.7
300M	60K	10K	0.03	61.1	48.8	68.5
300M	120K	50K	0.03	64.0	51.1	71.5
300M	400K	50K	0.03	68.6	55.5	76.0
300M	1.2M	50K	0.03	70.1	57.1	77.6
300M	2M	50K	0.03	70.5	57.1	77.9
1B	20K	10K	0.03	49.9	37.8	56.5
1B	30K	10K	0.03	55.2	42.8	62.3
1B	60K	10K	0.03	61.0	47.9	68.4
1B	120K	50K	0.03	64.0	51.1	71.5
1B	400K	50K	0.03	68.5	55.7	75.9
1B	1.2M	50K	0.03	70.0	57.3	77.3
3B	20K	10K	0.03	49.9	38.0	56.3
3B	30K	10K	0.03	55.4	43.5	62.4