5. **Explain all models that are used in producing performance results (eg RetinaNet) (10 points)**

RetinaNet: Its used to evaluate BiT on object detection. Using the COCO-2017 dataset and train a top-performing object detector, RetinaNet , using pre-trained BiT models as backbones. Due to memory constraints, we use the ResNet-101x3 architecture for all of our BiT models. We fine-tune the detection models on the COCO-2017 train split and report results on the validation split using the standard metric.

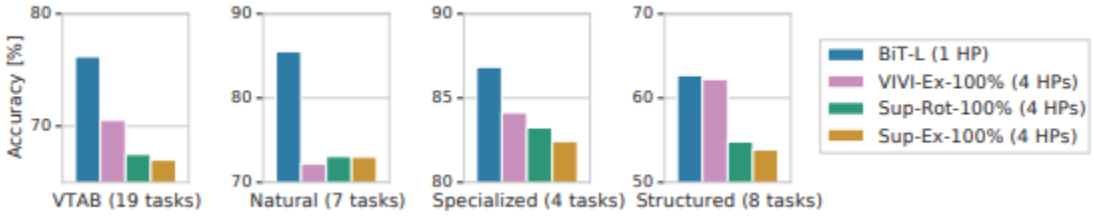
Model	Upstream data	AP
RetinaNet [33]	ILSVRC-2012	40.8
RetinaNet (BiT-S)	ILSVRC-2012	41.7
RetinaNet (BiT-M)	ImageNet-21k	43.2
RetinaNet (BiT-L)	JFT-300M	**43.8**

We can see clear benefits of pre-training on large data beyond ILSVRC-2012: pretraining on ImageNet-21k results in a 1.5 point improvement in Average Precision (AP), while pretraining on JFT-300M further improves performance by 0.6 points.

VTAB: Visual Task Adaptation Benchmark. Consist of 19 tasks and are divided into 7 Natural tasks, 4 Specialized tasks and 8 Structured tasks.

Main improvement is from pictures in the natural task as seen in figure below.

VTAB does well on natural images similar to what it was pre trained on.



BiT-L(JFT-300M): consists of around 300 million images with 1.26 labels per image on average. The labels are organized into a hierarchy of 18 291 classes. Annotation is performed using an automatic pipeline and are therefore imperfect; approximately 20% of the labels are noisy.

BiT-M(IN21K): M is trained on the full ImageNet-21k dataset, a public dataset containing 14.2 million images and 21k classes organized by the WordNet hierarchy. Images may contain multiple labels.

BiT-S(ILSVRC-2012): S is trained on the ILSVRC-2012 variant of ImageNet, which contains 1.28 million images and 1000 classes. Each image has a single label.

Generalist SOTA: including BiT, perform task-independent pre-training

Specialist SOTA: models are those that condition pre-training on each task

	BiT-L	Generalist SOTA	Specialist SOTA
ILSVRC-2012	**87.54 ± 0.02**	86.4 [57]	88.4 [61]*
CIFAR-10	**99.37 ± 0.06**	99.0 [19]	-
CIFAR-100	**93.51 ± 0.08**	91.7 [55]	-
Pets	**96.62 ± 0.23**	95.9 [19]	97.1 [38]
Flowers	**99.63 ± 0.03**	98.8 [55]	97.7 [38]
VTAB (19 tasks)	**76.29 ± 1.70**	70.5 [58]	-

Outperformance on all generalist SOTA in comparison to BiT-L including VTAB.

They don't always outperform specialist models, but they come really close.