

Selection via Proxy*:

Increasing the Computational Efficiency of Deep Active Learning

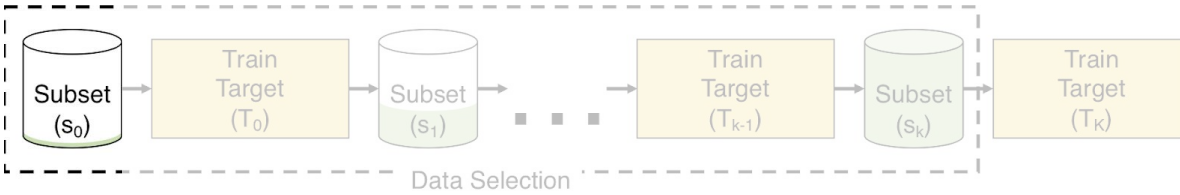
Cody Coleman, Christopher Yeh, Stephen Mussmann, Baharan Mirzasoleiman,
Peter Bailis, Percy Liang, Jure Leskovec, Matei Zaharia

*This is a subset of work from our [ICLR 2020 Paper](#). [Code is also available!](#)

- **Active learning** is a powerful data selection technique to reduce labeling costs

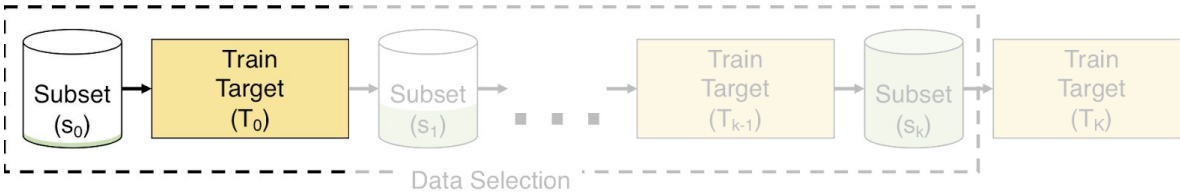


Traditional Approach



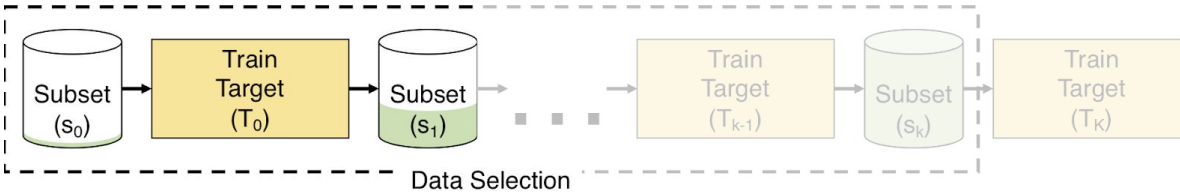
- **Active learning** is a powerful data selection technique to reduce labeling costs

Traditional Approach



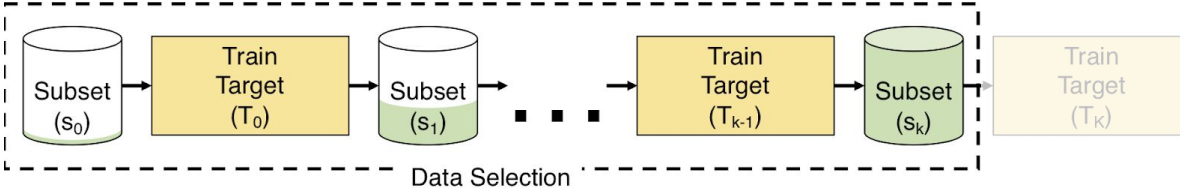
- **Active learning** is a powerful data selection technique to reduce labeling costs

Traditional Approach



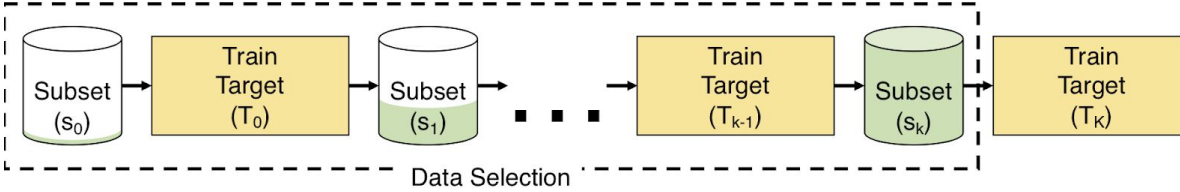
- **Active learning** is a powerful data selection technique to reduce labeling costs

Traditional Approach



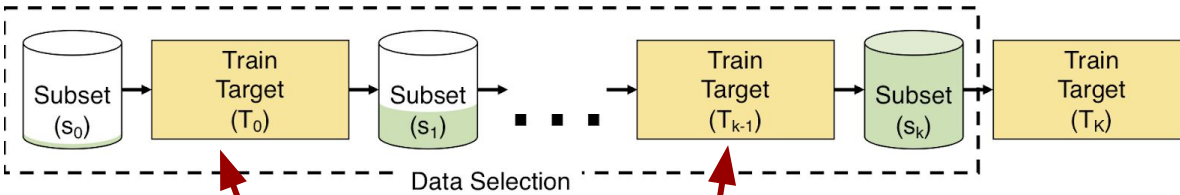
- **Active learning** is a powerful data selection technique to reduce labeling costs

Traditional Approach



- **Active learning** is a powerful data selection technique to reduce labeling costs

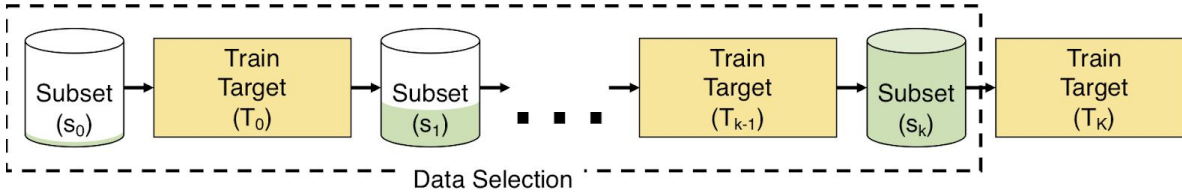
Traditional Approach



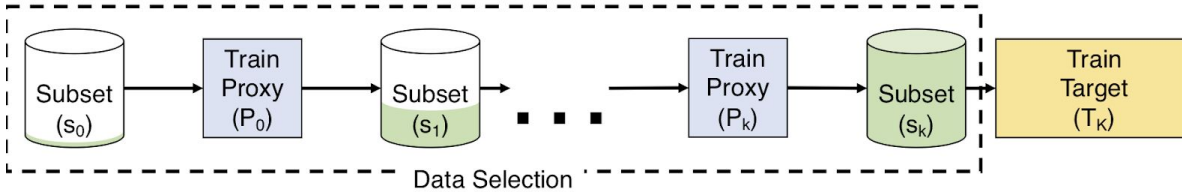
**computationally
expensive**

- **Active learning** is a powerful data selection technique to reduce labeling costs, but **can be computationally expensive**.

Traditional Approach

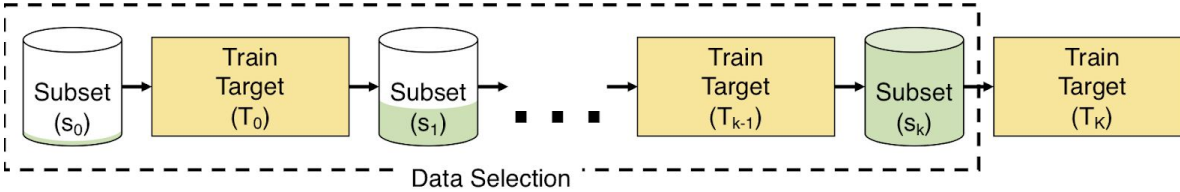


Our Approach: Selection Via Proxy

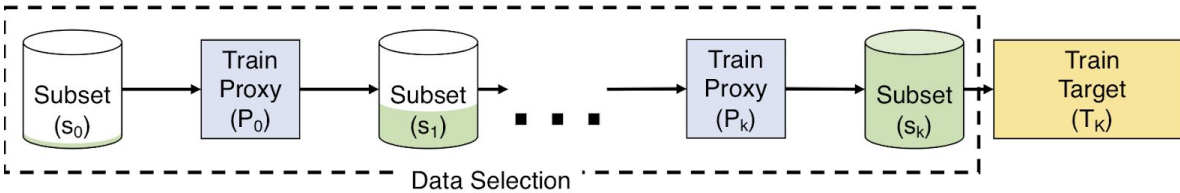


- Active learning is a powerful data selection technique to reduce labeling costs, but can be computationally expensive.
- **Small, less accurate models can serve as inexpensive proxies**

Traditional Approach



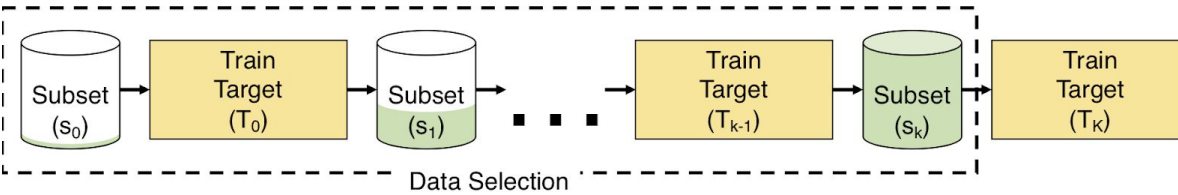
Our Approach: Selection Via Proxy



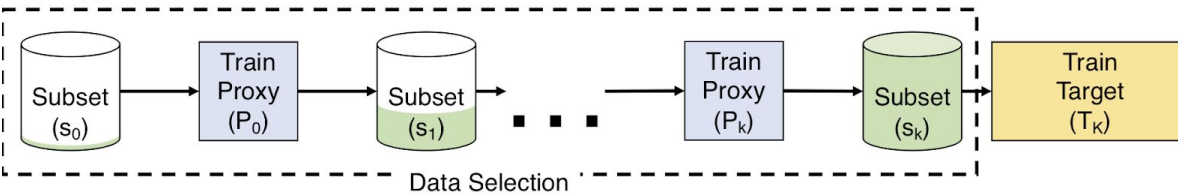
**Reach the
same accuracy!**

- Active learning is a powerful data selection technique to reduce labeling costs, but can be computationally expensive.
- Small, less accurate models can serve as inexpensive proxies

Traditional Approach



Our Approach: Selection Via Proxy

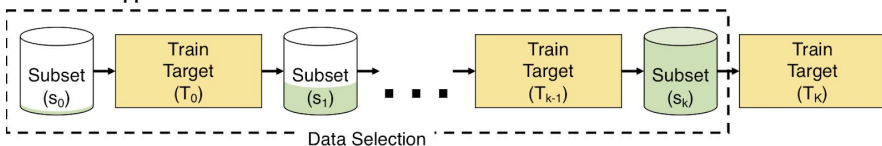


**Reach the
same accuracy!**

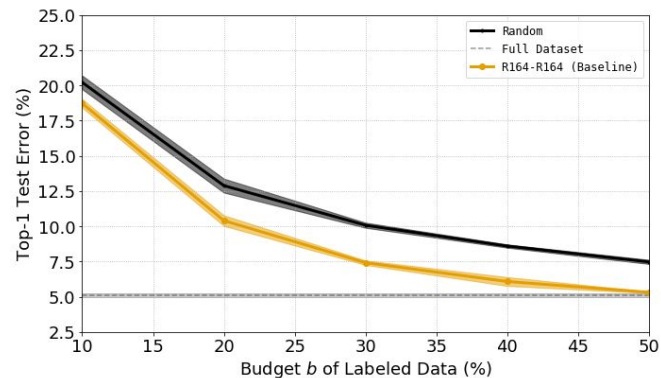
- Active learning is a powerful data selection technique to reduce labeling costs, but can be computationally expensive.
- Small, less accurate models can serve as inexpensive proxies and **accelerate data selection in active learning by up to 41.9x!**

			Active Learning Data Selection Speed-up					
			Budget:	10%	20%	30%	40%	50%
Dataset	Selection Model	Epochs						
CIFAR10	ResNet164 (Baseline)	181	1.0x	1.0x	1.0x	1.0x	1.0x	1.0x

Traditional Approach



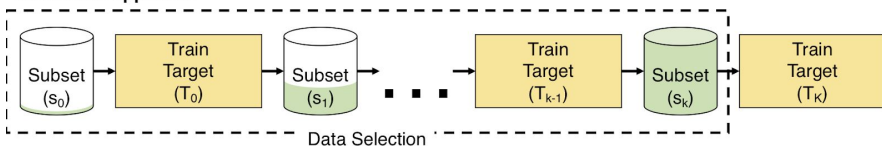
Our Approach: Selection Via Proxy



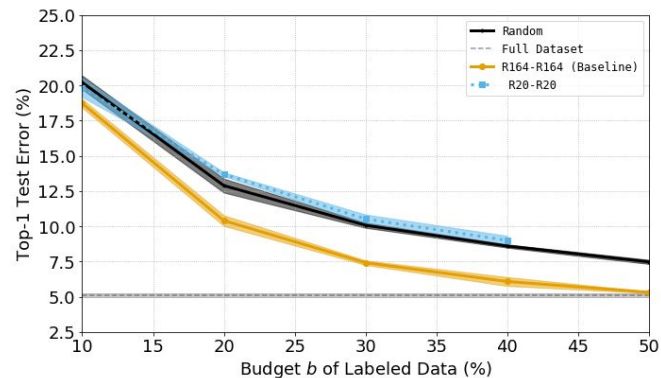
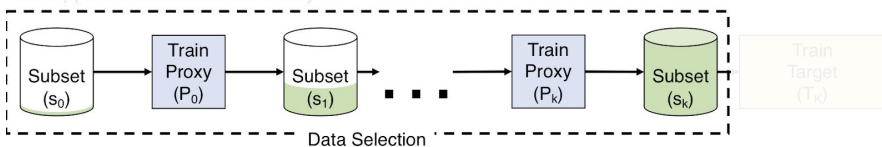
- Traditional active learning with ResNet164 is accurate but slow.

			Active Learning Data Selection Speed-up					
			Budget:	10%	20%	30%	40%	50%
Dataset	Selection Model	Epochs						
CIFAR10	ResNet164 (Baseline)	181	1.0x	1.0x	1.0x	1.0x	1.0x	1.0x

Traditional Approach



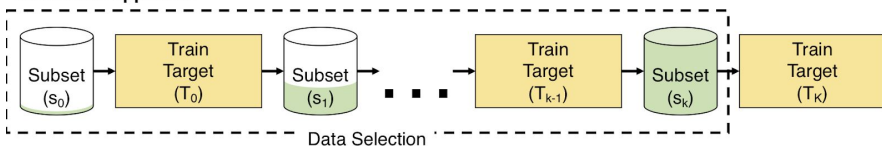
Our Approach: Selection Via Proxy



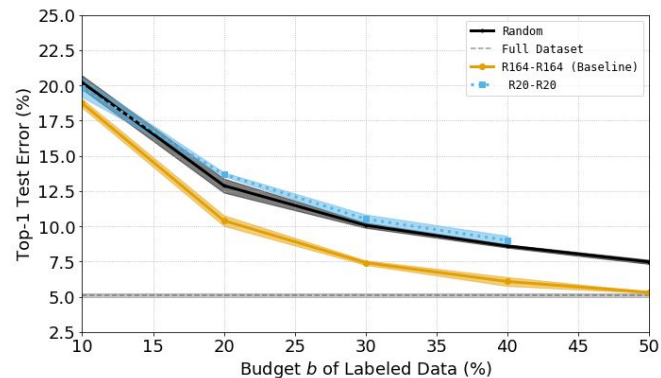
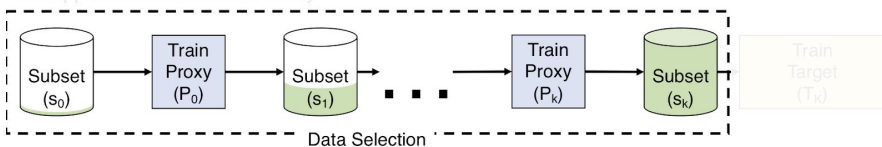
- Traditional active learning with ResNet20 is less accurate than ResNet164 but much faster.

			Active Learning Data Selection Speed-up					
			Budget:	10%	20%	30%	40%	50%
Dataset	Selection Model	Epochs						
CIFAR10	ResNet164 (Baseline)	181	1.0x	1.0x	1.0x	1.0x	1.0x	1.0x

Traditional Approach



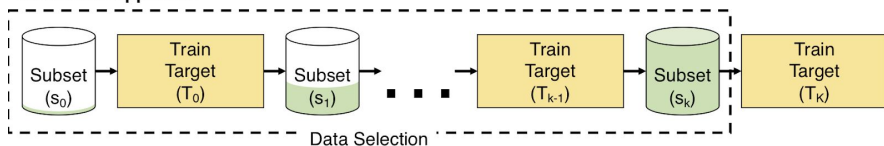
Our Approach: Selection Via Proxy



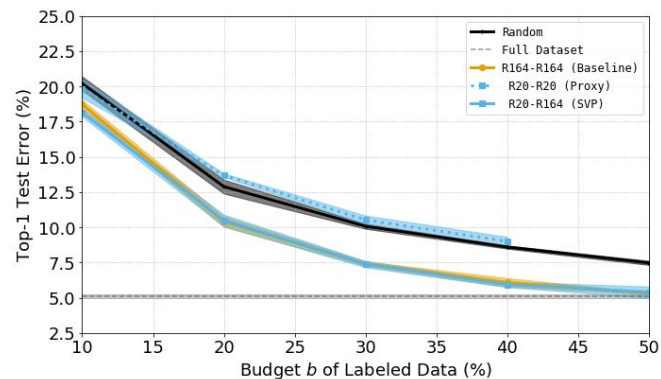
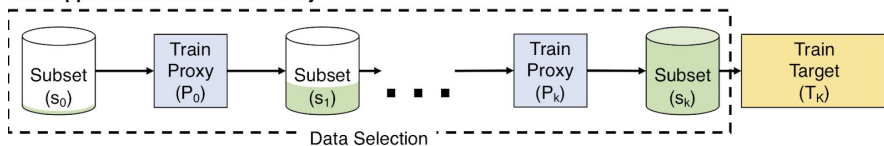
- Can we get the speed of ResNet20 and the final accuracy of ResNet164?

			Active Learning Data Selection Speed-up					
			Budget:	10%	20%	30%	40%	50%
Dataset	Selection Model	Epochs						
CIFAR10	ResNet164 (Baseline)	181	1.0x	1.0x	1.0x	1.0x	1.0x	1.0x
	ResNet20	181	3.8x	5.8x	6.7x	7.0x	7.2x	

Traditional Approach



Our Approach: Selection Via Proxy



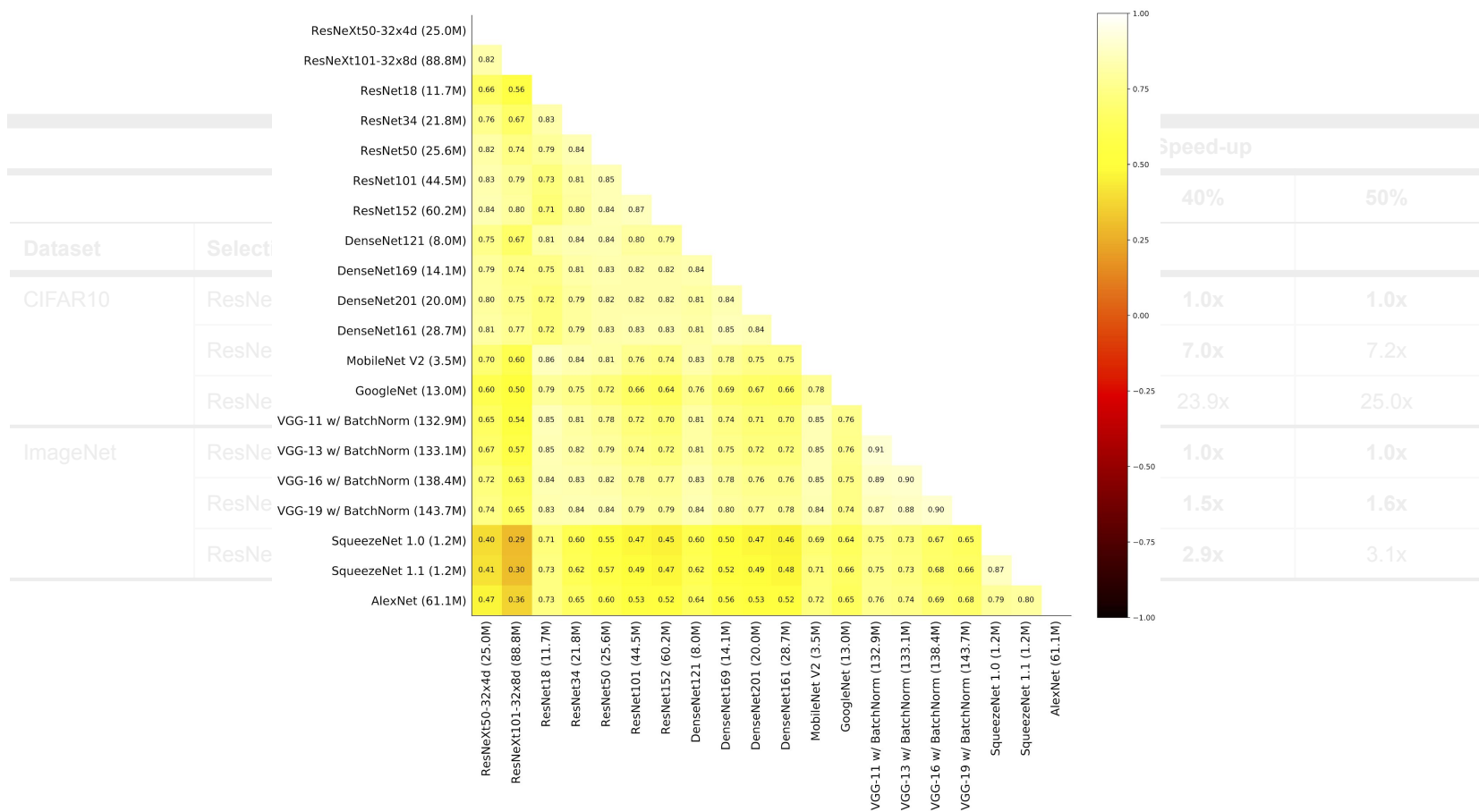
- Can we get the speed of ResNet20 and the final accuracy of ResNet164? **Yes!** Using the data selected by ResNet20 to train ResNet164, yields **up to a 7.0x speed-up without any loss in the final accuracy of ResNet164.**

			Active Learning Data Selection Speed-up				
Budget:			10%	20%	30%	40%	50%
Dataset	Selection Model	Epochs					
CIFAR10	ResNet164 (Baseline)	181	1.0x	1.0x	1.0x	1.0x	1.0x
	ResNet20	181	3.8x	5.8x	6.7x	7.0x	7.2x
	ResNet20	50	10.7x	18.9x	22.2x	23.9x	25.0x

- **Training ResNet20 for fewer epochs** before selecting points **is within 1% of the accuracy** of the baseline approach but **up to 25x faster**.

			Active Learning Data Selection Speed-up				
Budget:			10%	20%	30%	40%	50%
Dataset	Selection Model	Epochs					
CIFAR10	ResNet164 (Baseline)	181	1.0x	1.0x	1.0x	1.0x	1.0x
	ResNet20	181	3.8x	5.8x	6.7x	7.0x	7.2x
	ResNet20	50	10.7x	18.9x	22.2x	23.9x	25.0x
ImageNet	ResNet50 (Baseline)	90	1.0x	1.0x	1.0x	1.0x	1.0x
	ResNet18	90	1.2x	1.3x	1.4x	1.5x	1.6x
	ResNet18	45	2.1x	2.5x	2.7x	2.9x	3.1x

- Selection via proxy **yields similar results for much larger and more complex datasets** like ImageNet



- Selection via proxy yields similar results for much larger and more complex datasets like ImageNet, **where there is a high ranking correlation across many models.**

			Active Learning Data Selection Speed-up				
Budget:			10%	20%	30%	40%	50%
Dataset	Selection Model	Epochs					
CIFAR10	ResNet164 (Baseline)	181	1.0x	1.0x	1.0x	1.0x	1.0x
	ResNet20	181	3.8x	5.8x	6.7x	7.0x	7.2x
	ResNet20	50	10.7x	18.9x	22.2x	23.9x	25.0x
ImageNet	ResNet50 (Baseline)	90	1.0x	1.0x	1.0x	1.0x	1.0x
	ResNet18	90	1.2x	1.3x	1.4x	1.5x	1.6x
	ResNet18	45	2.1x	2.5x	2.7x	2.9x	3.1x
Amazon Review Polarity	VDCNN29 (Baseline)	15	1.0x	1.0x	1.0x	1.0x	1.0x
	fastText	10	10.6x	20.6x	32.2x	41.9x	51.3x

- By using an extremely fast proxy architecture, we achieve **up to a 41.9x speed-up**.

			Active Learning Data Selection Speed-up				
Budget:			10%	20%	30%	40%	50%
Dataset	Selection Model	Epochs					
CIFAR10	ResNet164 (Baseline)	181	1.0x	1.0x	1.0x	1.0x	1.0x
	ResNet20	181	3.8x	5.8x	6.7x	7.0x	7.2x
	ResNet20	50	10.7x	18.9x	22.2x	23.9x	25.0x
ImageNet	ResNet50 (Baseline)	90	1.0x	1.0x	1.0x	1.0x	1.0x
	ResNet18	90	1.2x	1.3x	1.4x	1.5x	1.6x
	ResNet18	45	2.1x	2.5x	2.7x	2.9x	3.1x
Amazon Review Polarity	VDCNN29 (Baseline)	15	1.0x	1.0x	1.0x	1.0x	1.0x
	fastText	10	10.6x	20.6x	32.2x	41.9x	51.3x

Conclusion: Selection via Proxy (SVP) improves the computational efficiency of active learning by substituting a cheap proxy model for a more accurate, but expensive model.