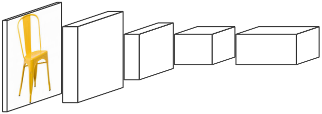
Rethinking deep active learning: Using unlabeled data at model training

Oriane Siméoni
Mateusz Budnik
Yannis Avrithis
Guillaume Gravier

Inria, Irisa, Univ Rennes, CNRS

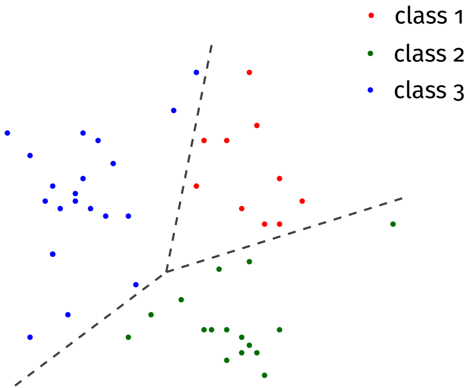
Training DL model for the classification task



- Requires large annotated datasets

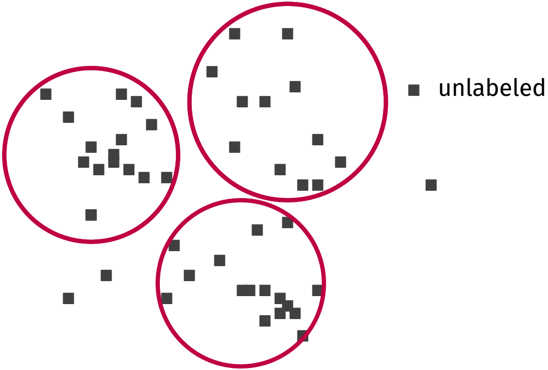
- Annotation done by humans
- Long and fastidious process

Learning with less supervision



Supervised

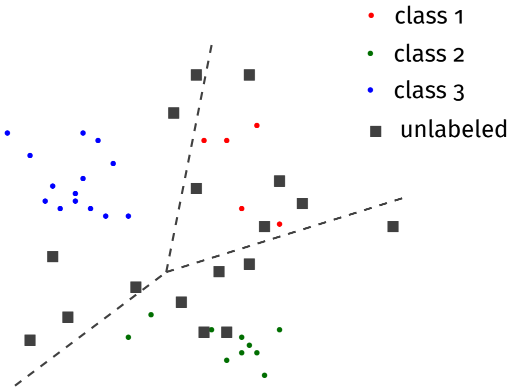
Learning with less supervision



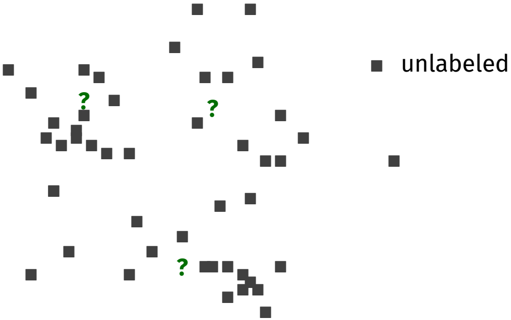
Supervised

Unsupervised

Learning with less supervision

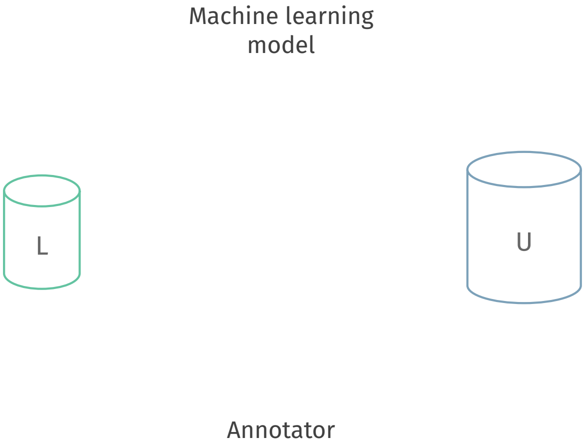


Learning with less supervision

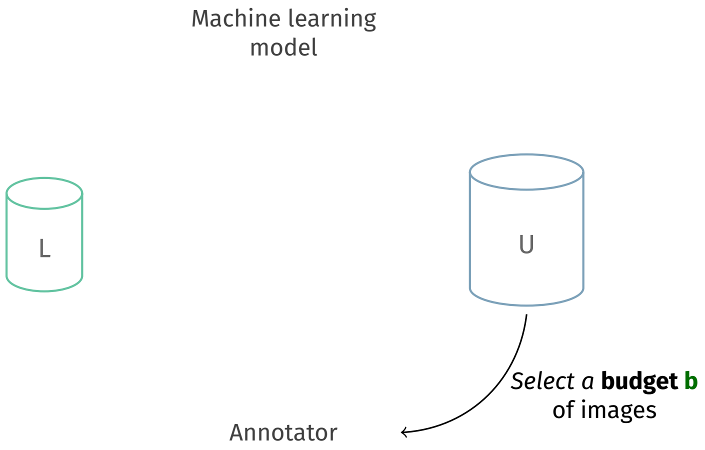


Supervised | Semi-supervised | Active | Unsupervised

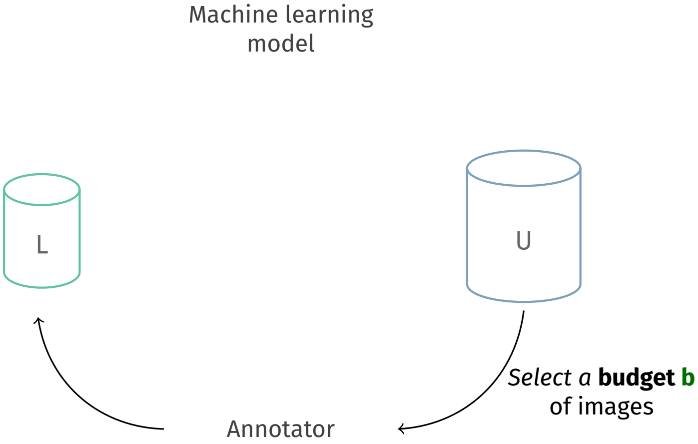
Active Learning



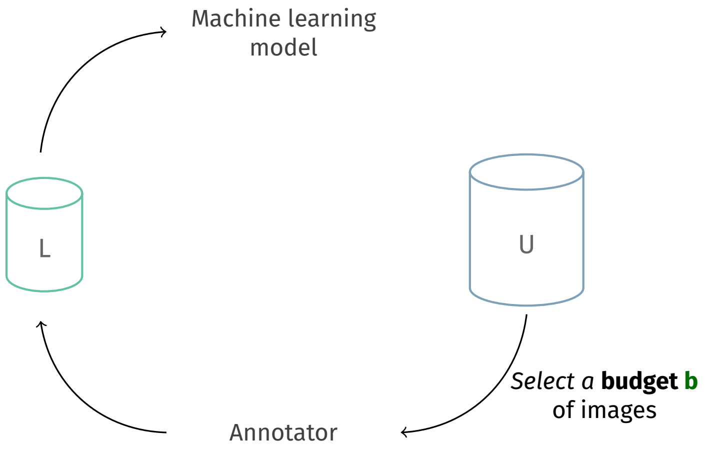
Active Learning



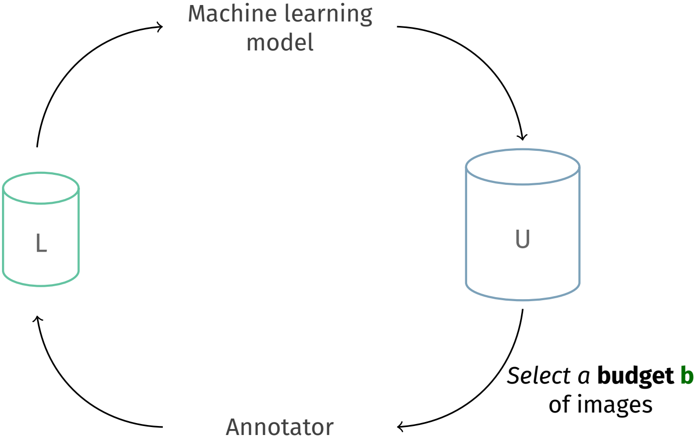
Active Learning



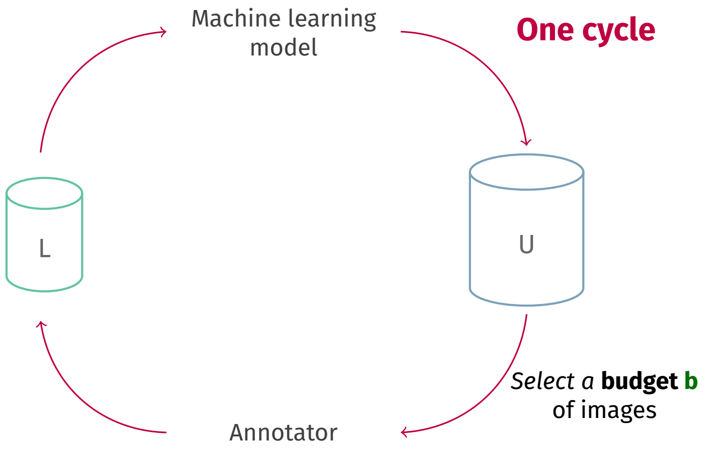
Active Learning



Active Learning



Active Learning



- Relevant before Deep Learning
- Not studied much in the context of Deep Learning

What is the best solution?

Baselines

- Random
  Selects uniformly random images.

[2] Y. Geifman and R. El-Yaniv. “Deep Active Learning over the Long Tail”. In: *arXiv preprint arXiv:1711.00941* (2017)

[2] O. Sener and S. Savarese. “Active learning for convolutional neural networks: A core-set approach”. In: *arXiv* (2018)

[2] K. Wang et al. “Cost-effective active learning for deep image classification”. In: *IEEE Trans. CSVT* 27.12 (2017), pp. 2591–2600

What is the best solution?

Baselines

- Random
  Selects uniformly random images.
- Geometry[1]
  Selects most distant image to its nearest labeled or previously acquired examples.

[2] Y. Geifman and R. El-Yaniv. “Deep Active Learning over the Long Tail”. In: *arXiv preprint arXiv:1711.00941* (2017)

[2] O. Sener and S. Savarese. “Active learning for convolutional neural networks: A core-set approach”. In: *arXiv* (2018)

[2] K. Wang et al. “Cost-effective active learning for deep image classification”. In: *IEEE Trans. CSVT* 27.12 (2017), pp. 2591–2600

What is the best solution?

Baselines

- Random
  Selects uniformly random images.
- Geometry[1]
  Selects most distant image to its nearest labeled or previously acquired examples.
- Uncertainty
  Selects most uncertain images: highest entropy of the classifier output probabilities.

[2] Y. Geifman and R. El-Yaniv. "Deep Active Learning over the Long Tail". In: *arXiv preprint arXiv:1711.00941* (2017)

[2] O. Sener and S. Savarese. "Active learning for convolutional neural networks: A core-set approach". In: *arXiv* (2018)

[2] K. Wang et al. "Cost-effective active learning for deep image classification". In: *IEEE Trans. CSVT* 27.12 (2017), pp. 2591–2600

What is the best solution?

Baselines

- Random
  Selects uniformly random images.
- Geometry[1]
  Selects most distant image to its nearest labeled or previously acquired examples.
- Uncertainty
  Selects most uncertain images: highest entropy of the classifier output probabilities.
- CEAL[2]
  Uses unlabeled data.

[2] Y. Geifman and R. El-Yaniv. "Deep Active Learning over the Long Tail". In: *arXiv preprint arXiv:1711.00941* (2017)

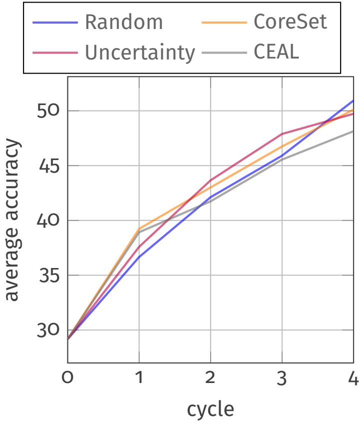
[2] O. Sener and S. Savarese. "Active learning for convolutional neural networks: A core-set approach". In: *arXiv* (2018)

[2] K. Wang et al. "Cost-effective active learning for deep image classification". In: *IEEE Trans. CSVT* 27.12 (2017), pp. 2591–2600

What is the best solution?

Baselines

- Random
  Selects uniformly random images.
- Geometry[1]
  Selects most distant image to its nearest labeled or previously acquired examples.
- Uncertainty
  Selects most uncertain images: highest entropy of the classifier output probabilities.
- CEAL[2]
  Uses unlabeled data.



CIFAR-10 ($b = 100$)

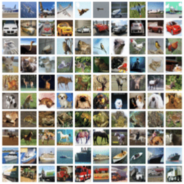
[2] Y. Geifman and R. El-Yaniv. "Deep Active Learning over the Long Tail". In: *arXiv preprint arXiv:1711.00941* (2017)

[2] O. Sener and S. Savarese. "Active learning for convolutional neural networks: A core-set approach". In: *arXiv* (2018)

[2] K. Wang et al. "Cost-effective active learning for deep image classification". In: *IEEE Trans. CSVT* 27.12 (2017), pp. 2591–2600

Experimental details

- Network
  - 13-layer convolutional network[3]
  - model trained from scratch
- Training very dependent on the data $\rightarrow$ 5 repetitions
- Metrics: average accuracy and standard deviation
- Datasets
  - MNIST[4] (10 cls, 60000 imgs)
  - SVHN[5] (10 cls, 73257 imgs)
  - CIFAR-10[6] (10 cls, 50000 imgs)
  - CIFAR-100[7] (100 cls, 50000 imgs)



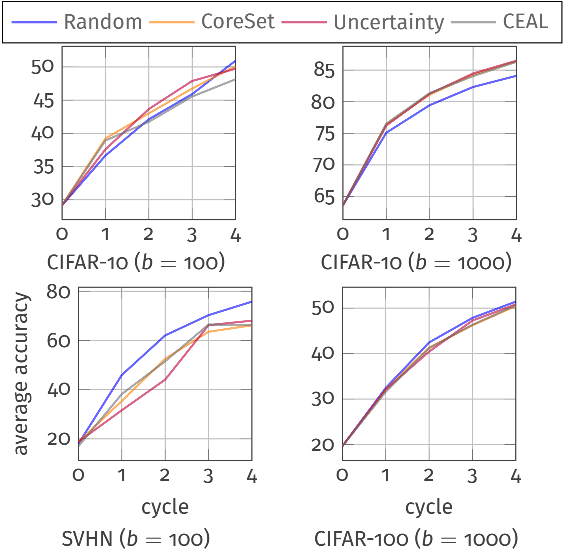
[3] S. Laine and T. Aila. "Temporal ensembling for semi-supervised learning". In: *arXiv preprint arXiv:1610.02242* (2016).

[5] Y. LeCun et al. "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324

[6] Y. Netzer et al. "Reading Digits in Natural Images with Unsupervised Feature Learning". In: *NIPS Workshop on Deep Learning and Unsupervised Feature Learning* (Jan. 2011)

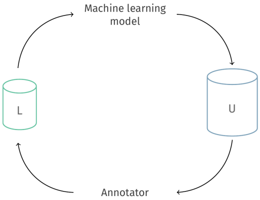
[7] A. Krizhevsky. "Learning Multiple Layers of Features from Tiny Images". In: 2009

What is the best solution?



No clear winner.

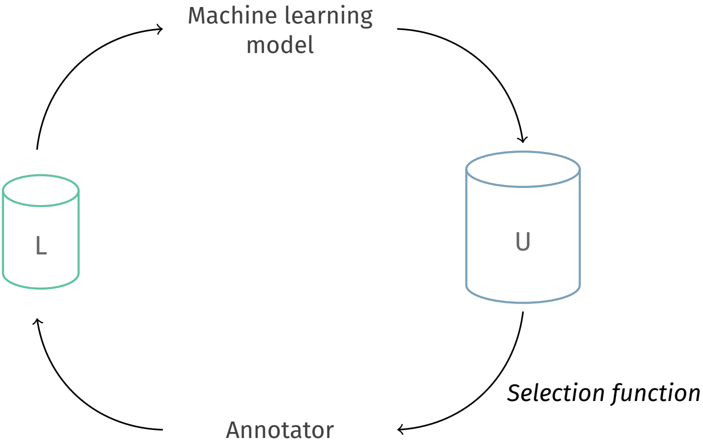
The idea



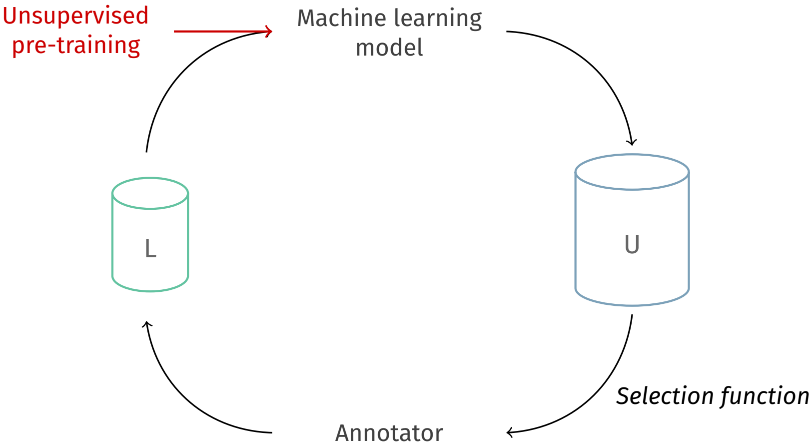
What if we could

- Improve results with no additional supervision
- Use unlabeled data during training

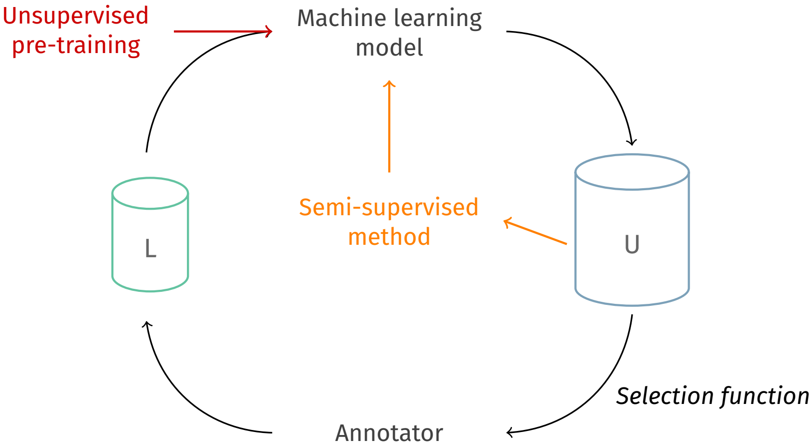
Using more unlabeled data



Using more unlabeled data



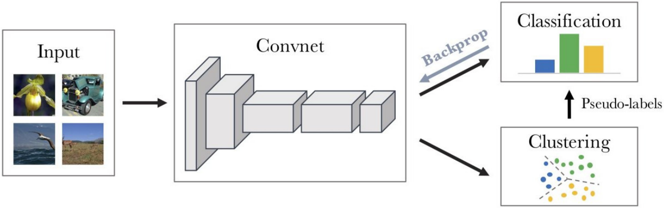
Using more unlabeled data



Integrating information from unlabeled data

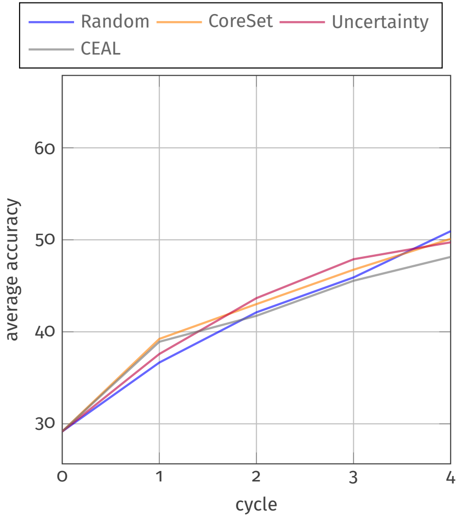
- Improving the model using unlabeled data
- Unsupervised pre-training

- Following Deep Cluster[8] to pre-train CNN
  - Assign classes to data given closest centroids
  - Train the network
  - Re-assign classes



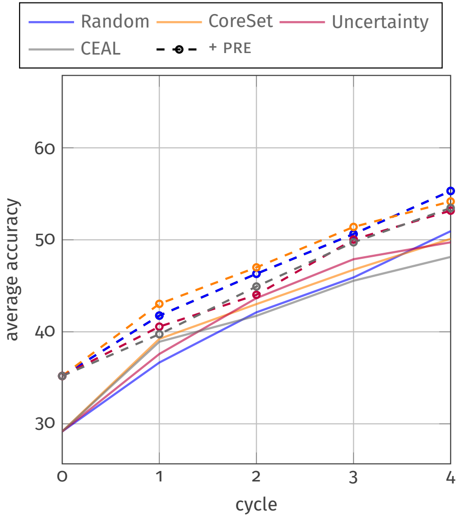
[8] M. Caron et al. "Deep Clustering for Unsupervised Learning of Visual Features". In: *arXiv preprint arXiv:1807.05520* (2018).

Integrating information from unlabeled data



CIFAR-10 ($b = 100$)

Integrating information from unlabeled data



CIFAR-10 ($b = 100$)

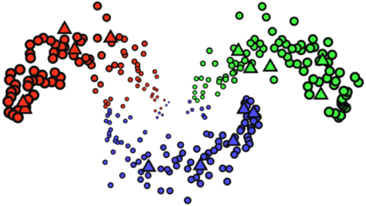
Benefits

- performed only once at the beginning of the process
- can bring up to 6% improvement

Improving active learning cycles

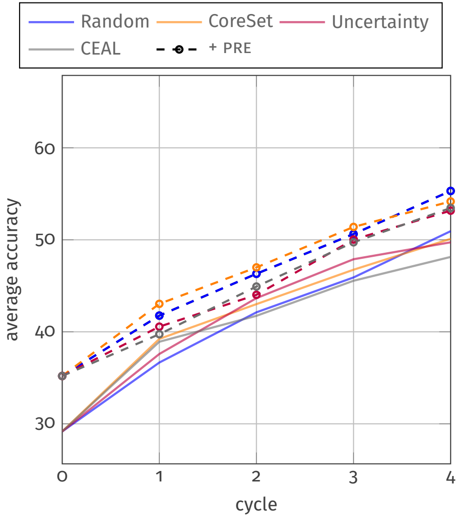
- Use unlabeled data in each cycle
- Adding semi-supervised learning

- Iterative label propagation following Iscen *et al*[9].
  - Construct a reciprocal k-nn graph on data features
  - Label propagation
  - Train classifier using pseudo-labels



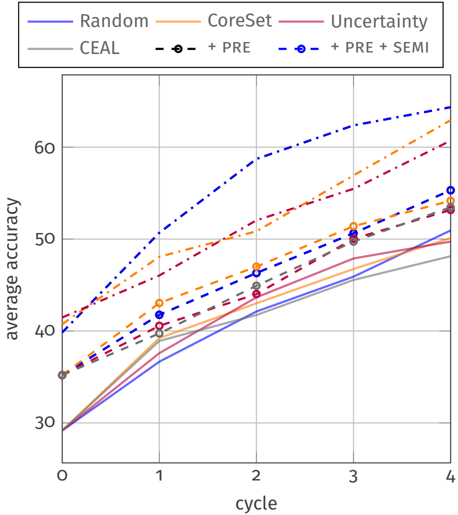
[9] A. Iscen et al. “Label propagation for Deep Semi-supervised Learning”. In: *CVPR*. 2019.

Adding semi-supervision



CIFAR-10 ($b = 100$)

Adding semi-supervision

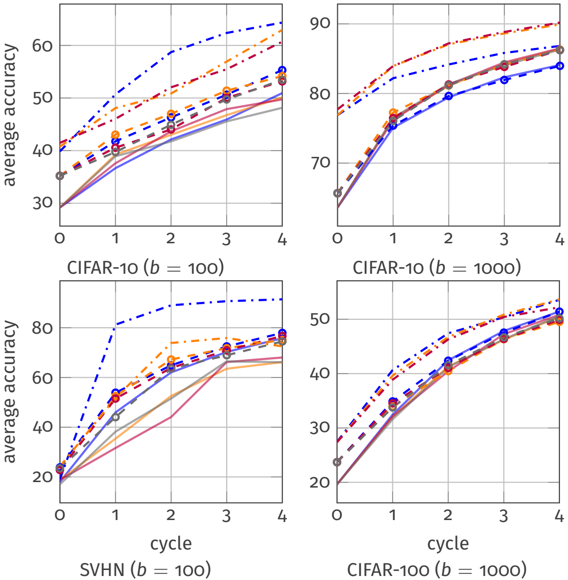


CIFAR-10 ($b = 100$)

Benefits

- Results improved by up to 15% from baselines
- Taking advantage of the whole dataset
- Suits better deep learning models

Adding semi-supervision



Conclusions

Take home message

- Active learning benefits from using unlabeled data
- We obtain better models requiring less labeled data
- Random selection of images is best with small budgets
- The selection method does not appear to matter

Contributions

- First results mixing active learning and unlabeled methods in the context of Deep Learning
- Proposition to rethink Deep Active Learning
  - using a scenario integrating unlabeled data
  - to always compare to Random with small budgets