

# Dataset Pruning Using Early Exit Networks

ALPEREN GÖRMEZ AND ERDEM KOYUNCU

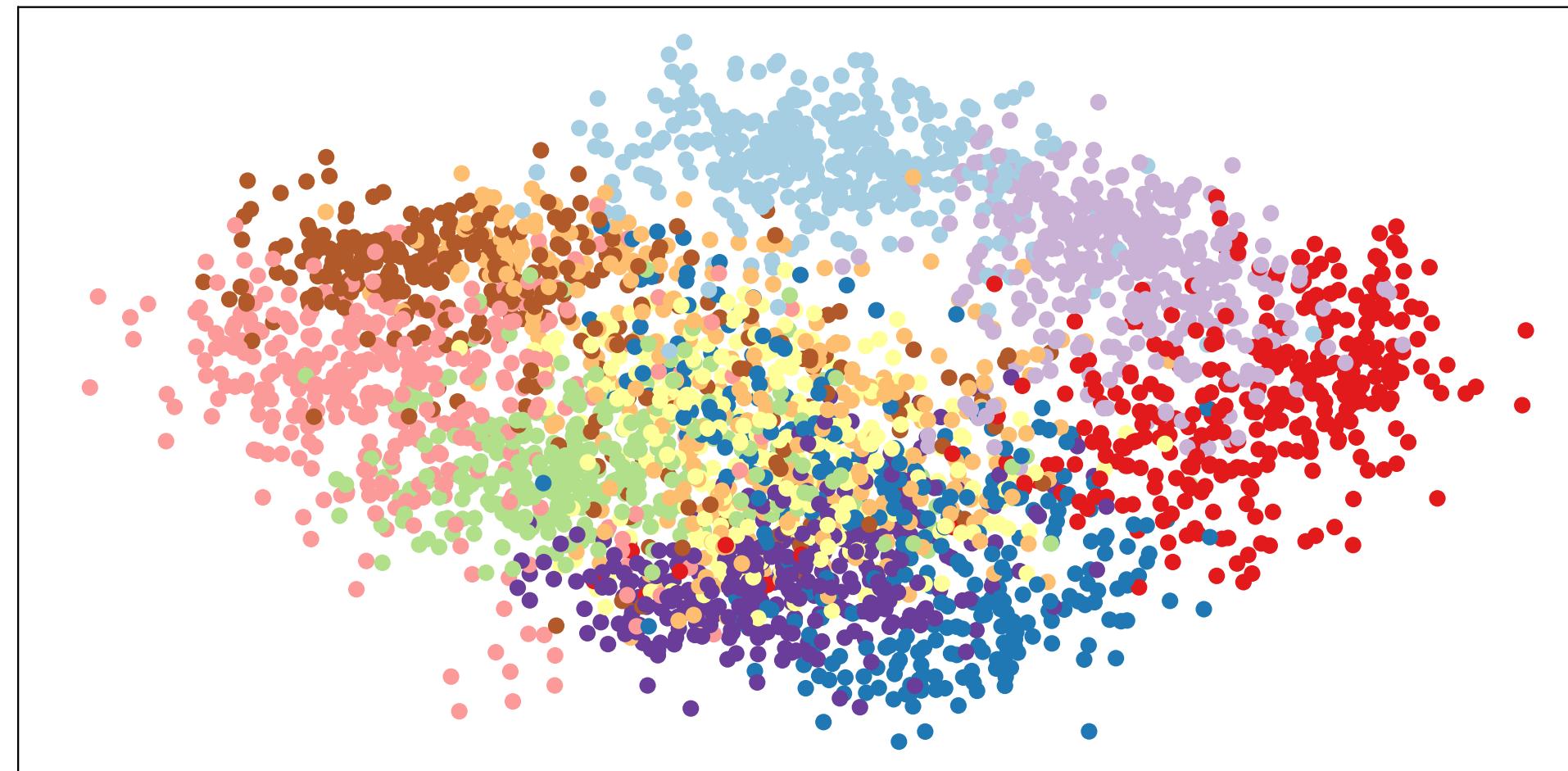
Department of Electrical and Computer Engineering, University of Illinois Chicago

## Scope

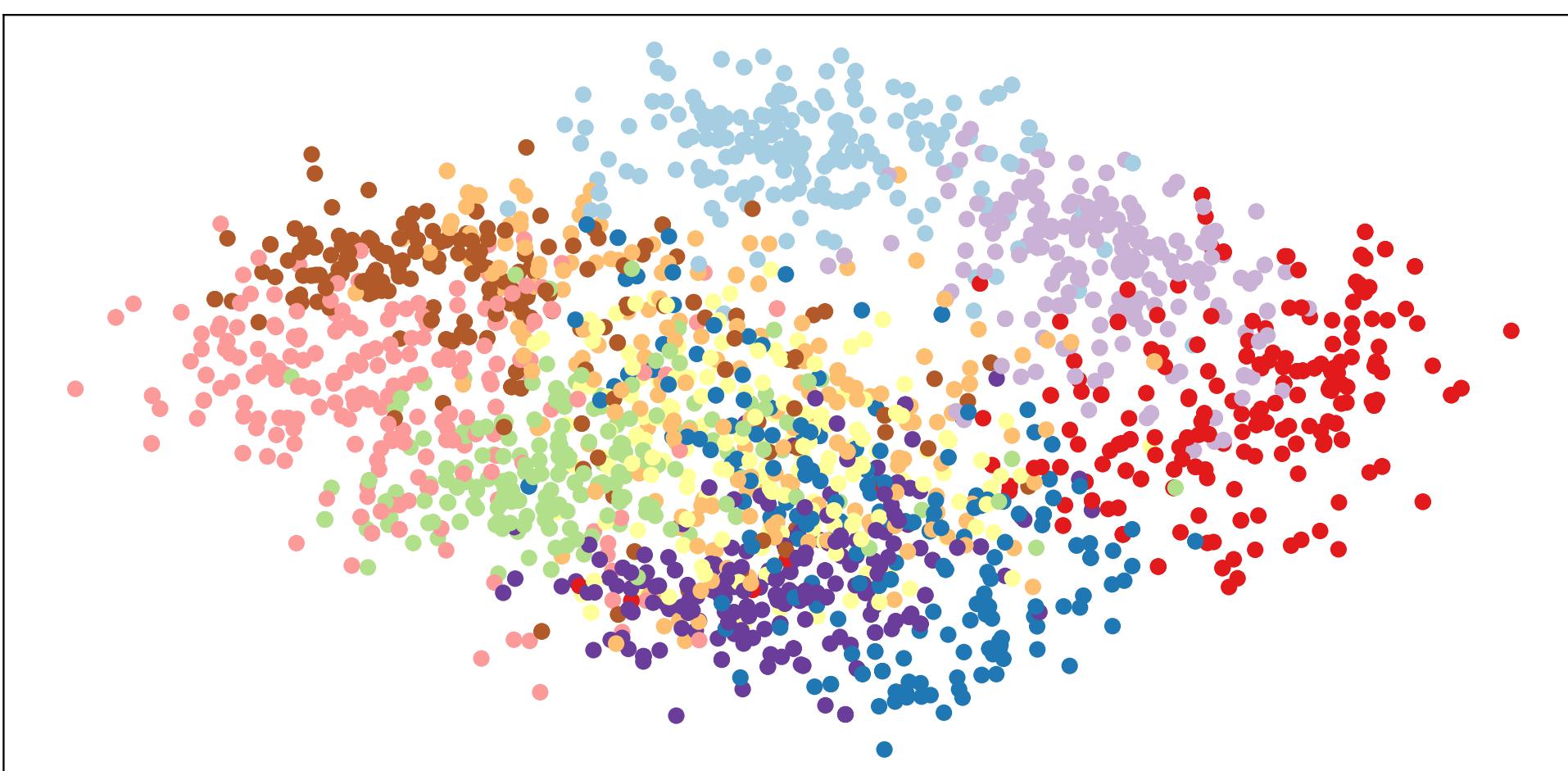
- The Bitter Lesson: An increased amount of computation has been the key enabler for better performance.
- The Trend: Simply train a larger model on a larger dataset.
- This is not sustainable in the long run.
- Our focus** is reducing the training cost by **reducing the training set size**.

## Dataset Pruning

- Goal:** Prune **easy samples** and keep the test set performance intact.
- The key idea is assigning a score of **easiness** to each training sample.



(a) Full training set.



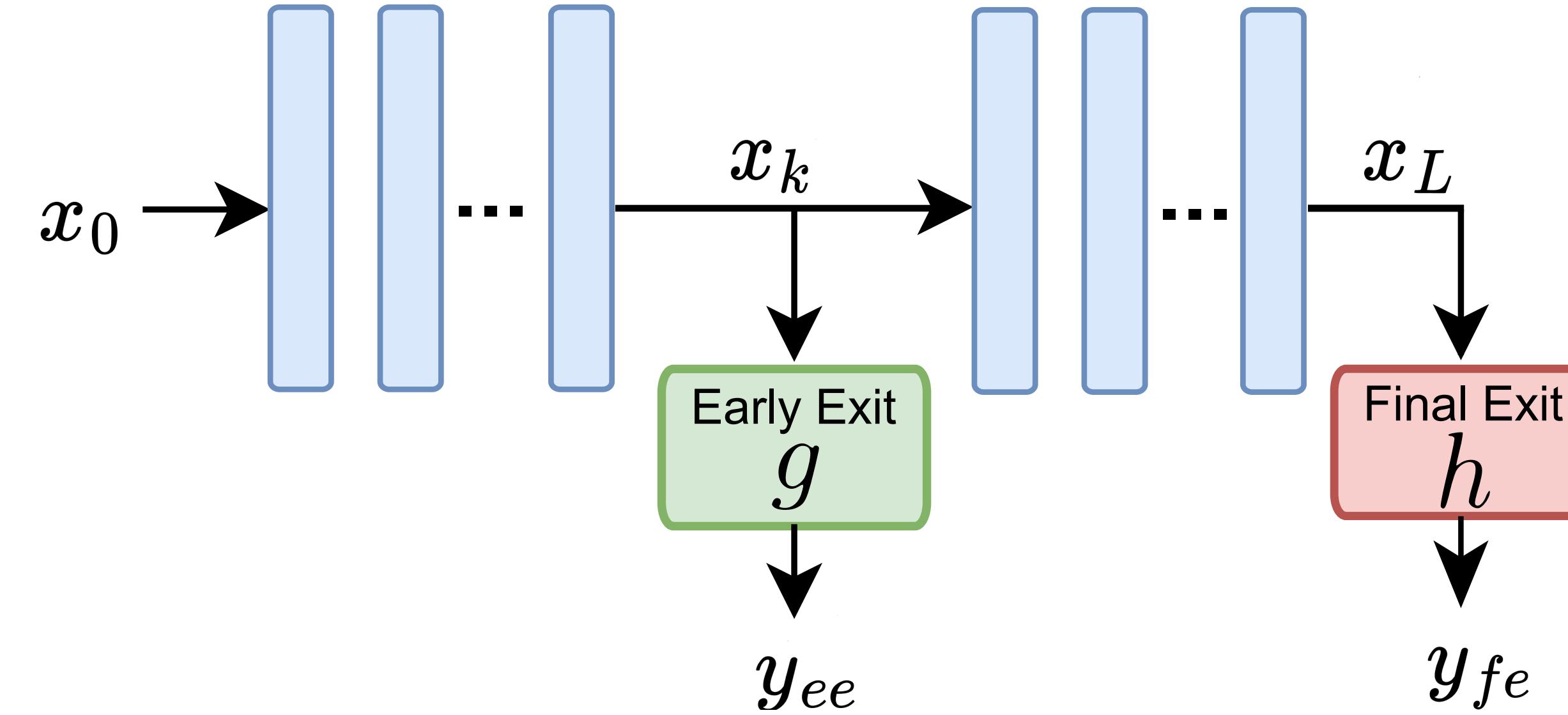
(b) Pruned training set.

## Motivation

- Existing methods are costly. They train an ensemble of models, they train the model fully on the entire dataset.
- They cannot beat random pruning convincingly.
- This work:** Use the model's **innate** ability to detect easy samples.

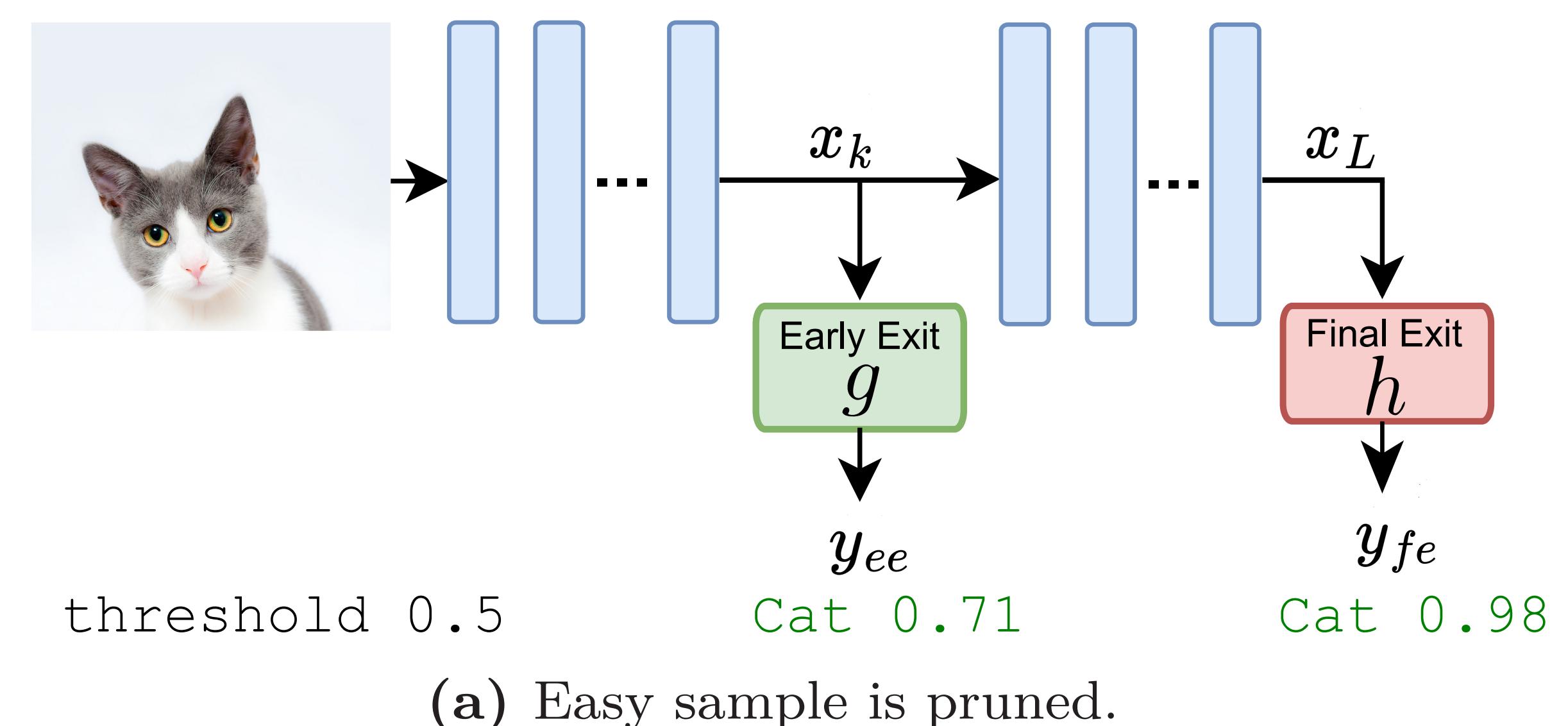
## Early Exit Networks

- Base model + intermediate classifiers.
- Exploit heterogeneous nature of the real world data.
- Easy** data exit **early** from the network at inference time.

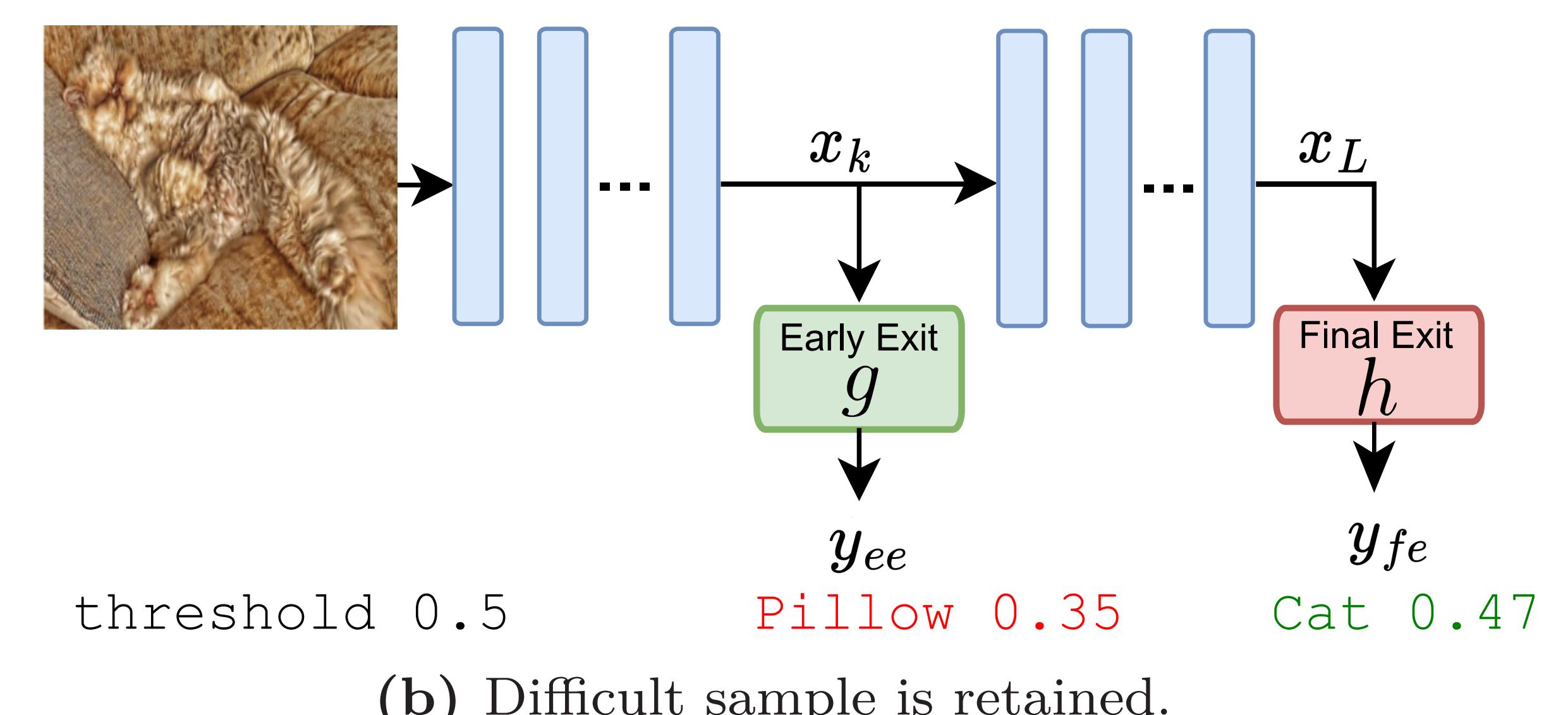


## Our Contribution: EEPrune

- We utilize early exit networks to detect easy samples.
- We prune a training sample if
  - The early exit can predict correctly, **AND**
  - The final exit can predict correctly, **AND**
  - The early exit confidence > threshold.



(a) Easy sample is pruned.



(b) Difficult sample is retained.

## Results

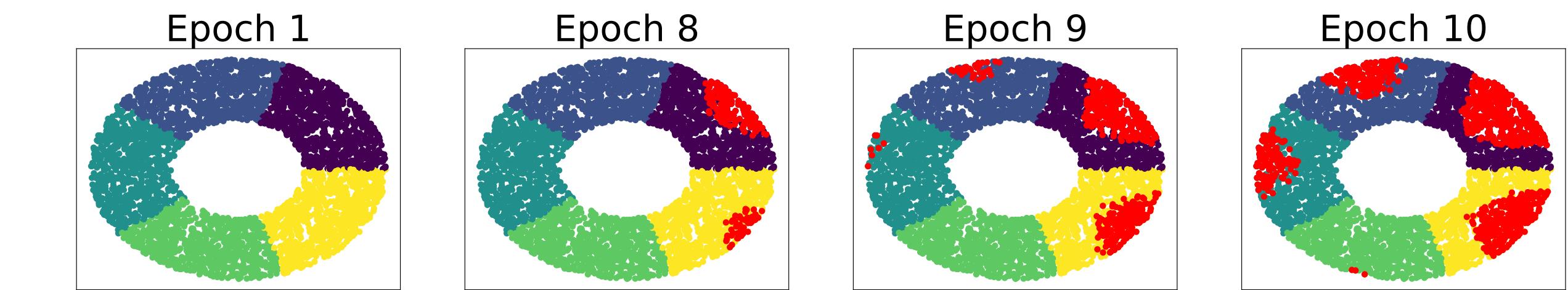
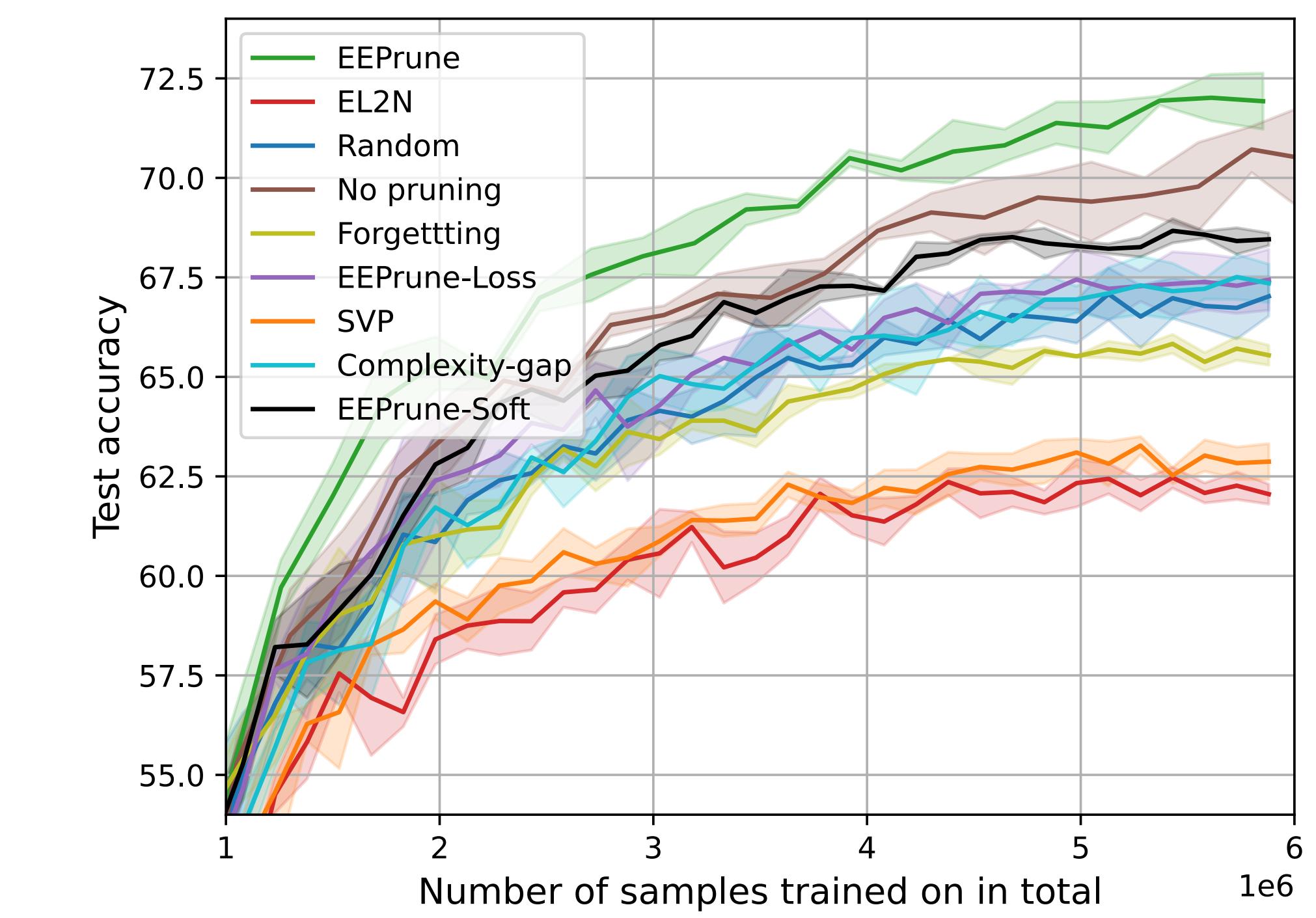
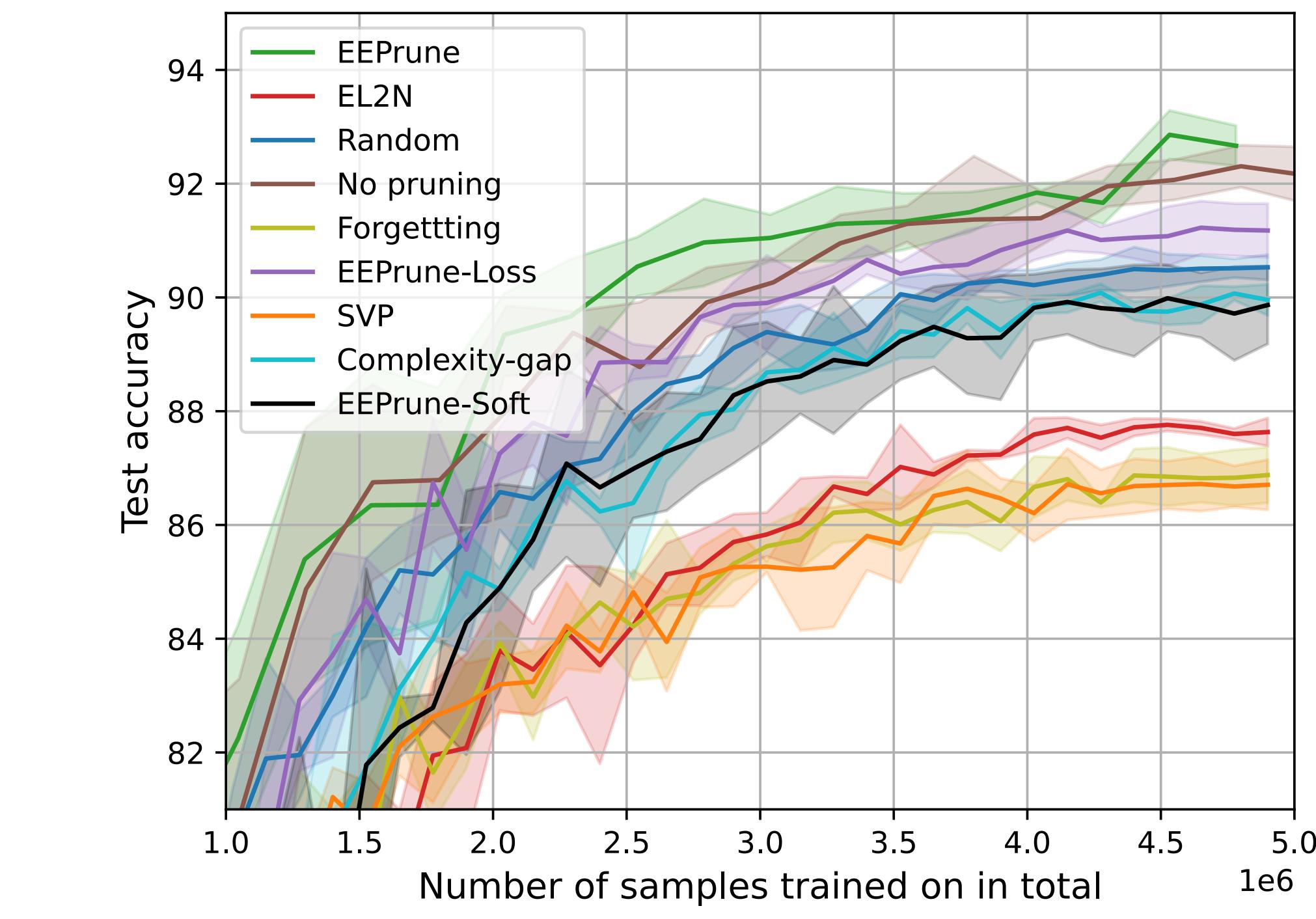


Figure 3: EEPrune discards easy samples (red).



(a) MobileNetV3, CIFAR-100, 40% pruning.



(b) ResNet-50, CIFAR-10, 50% pruning.

Figure 4: Results on CIFAR-10 and CIFAR-100 datasets.

## Conclusion

EEPrune can reduce the training cost more than existing methods, and identify easy samples in a more resource-efficient manner.