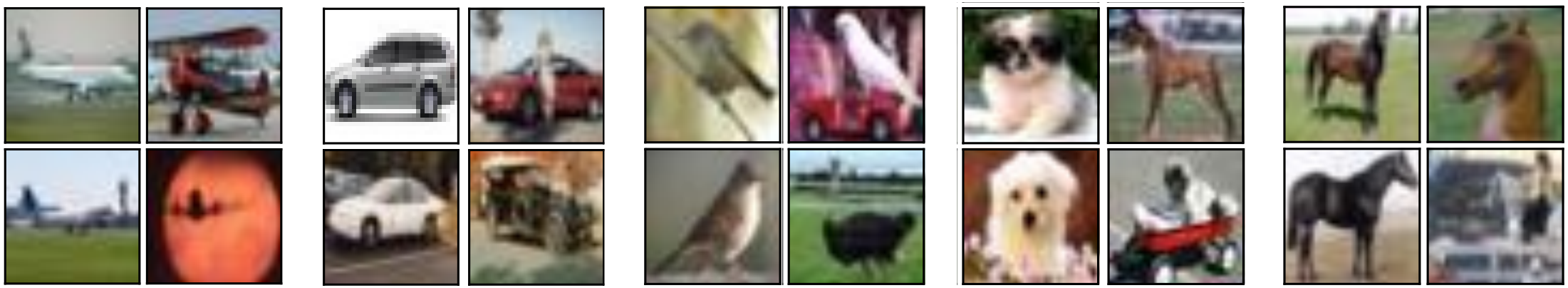
**Easy** / **Hard** samples from different classes of CIFAR10



Airplane | Automobile | Bird | Dog | Horse

# SGD Biased towards Early Important Samples for Efficient Training

International Conference on Data Mining 2023 - 03.12.2023

Alessio Quercia[1,2], Abigail Morrison[1,2], Hanno Scharr[1], Ira Assent[1,3] – [1]Forschungszentrum Jülich, [2]RWTH Aachen, [3]Aarhus University

Member of the Helmholtz Association

RWTH AACHEN UNIVERSITY

AARHUS UNIVERSITY

HDSLEE HELMHOLTZ School for Data Science in Life · Earth · Energy

JÜLICH Forschungszentrum

# Goal & application

Goal: given a model architecture, an optimizer, a training dataset and a test dataset

- make training more data efficient → highest test accuracy with the shortest training time

# Goal & application

Goal: given a model architecture, an optimizer, a training dataset and a test dataset

- make training more data efficient → highest test accuracy with the shortest training time

Application: supervised image classification

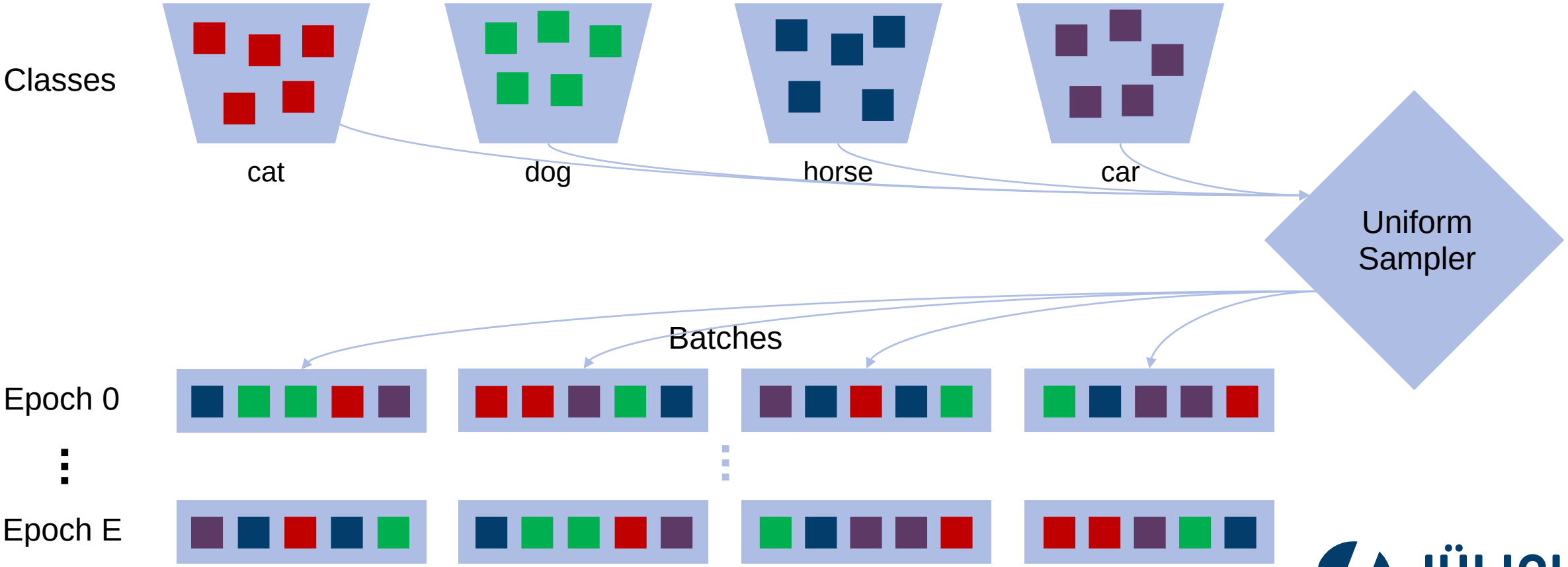
CIFAR10 classes

Easy

Hard

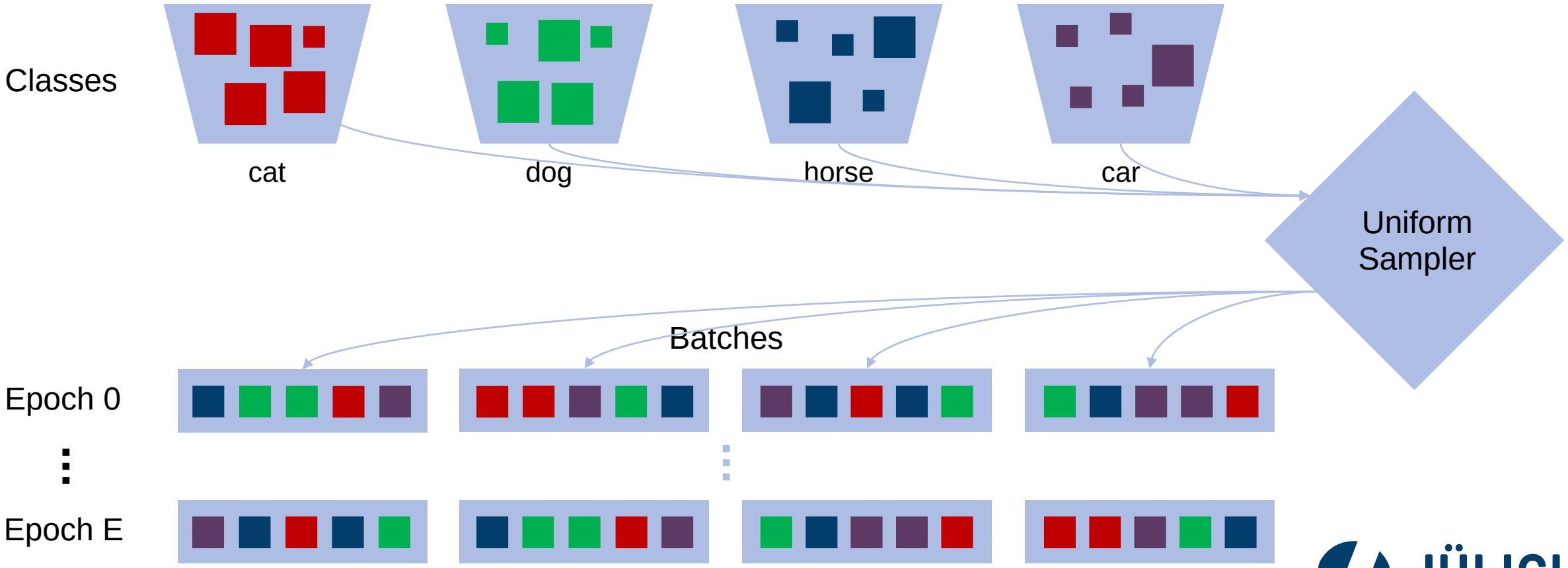
# Pretrain for E epochs with vanilla SGD…

Vanilla SGD uniformly draws all samples in each epoch → Equally important for training



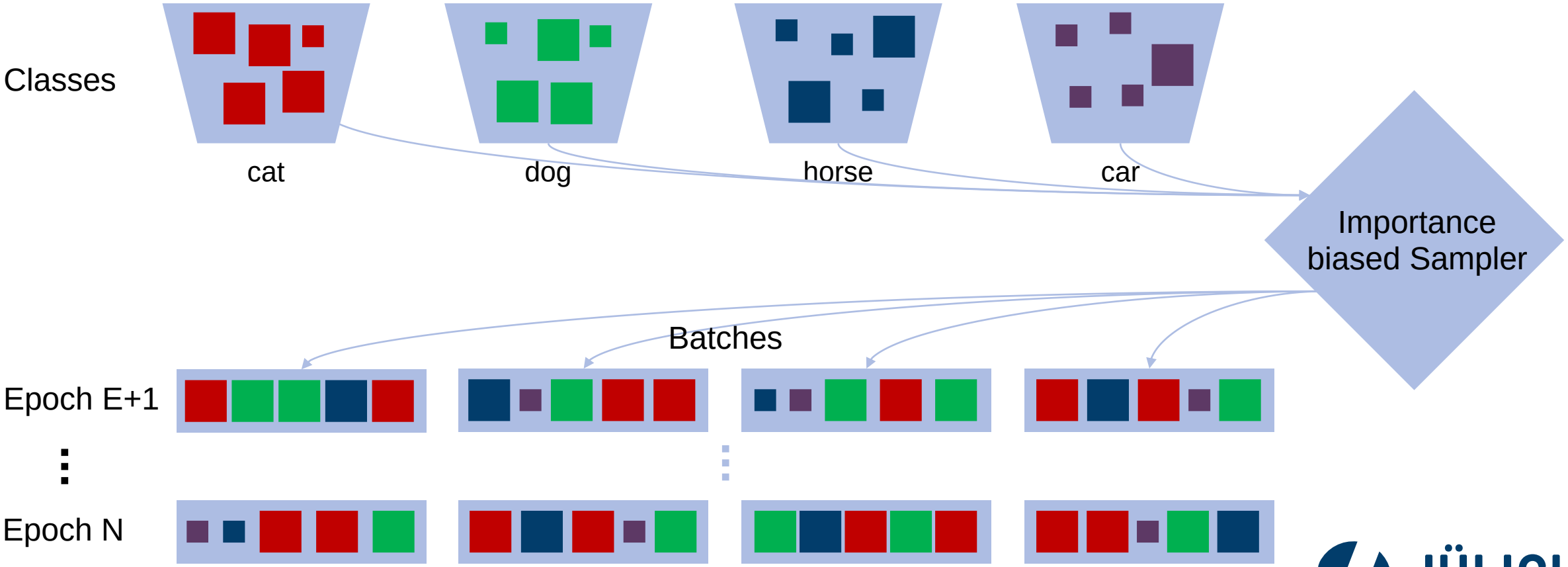
# Pretrain for E epochs with vanilla SGD…

Importance distribution can be identified after a few epochs



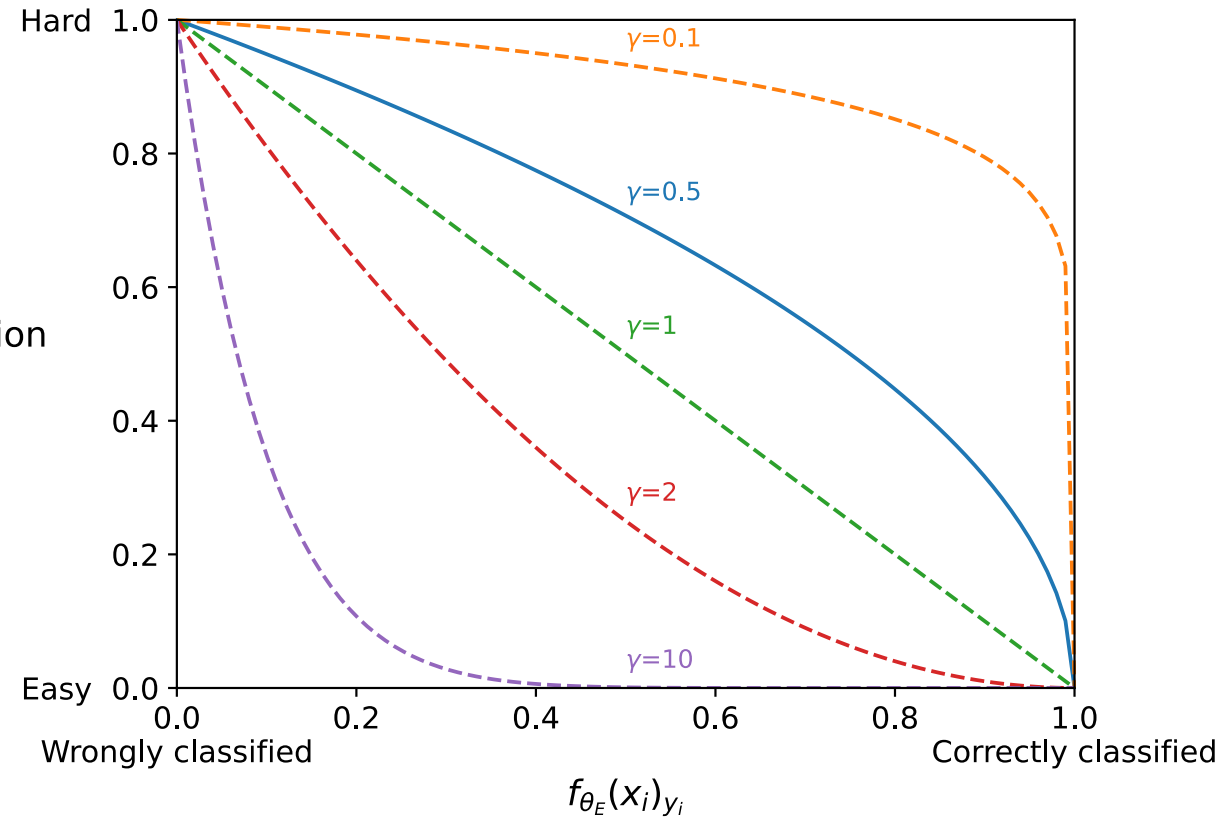
# … continue training until epoch N with importance-biased SGD

Our method oversamples hard samples and undersamples easy samples



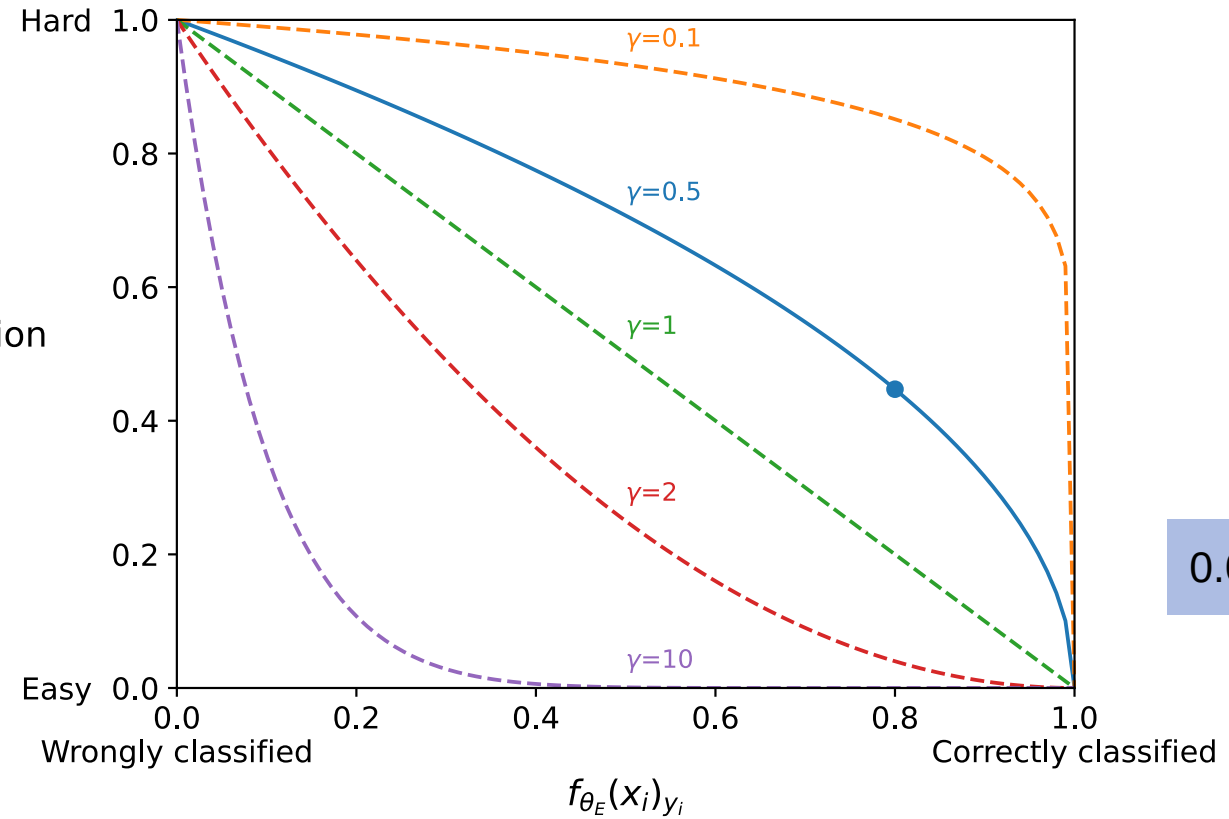
# Importance weights metric

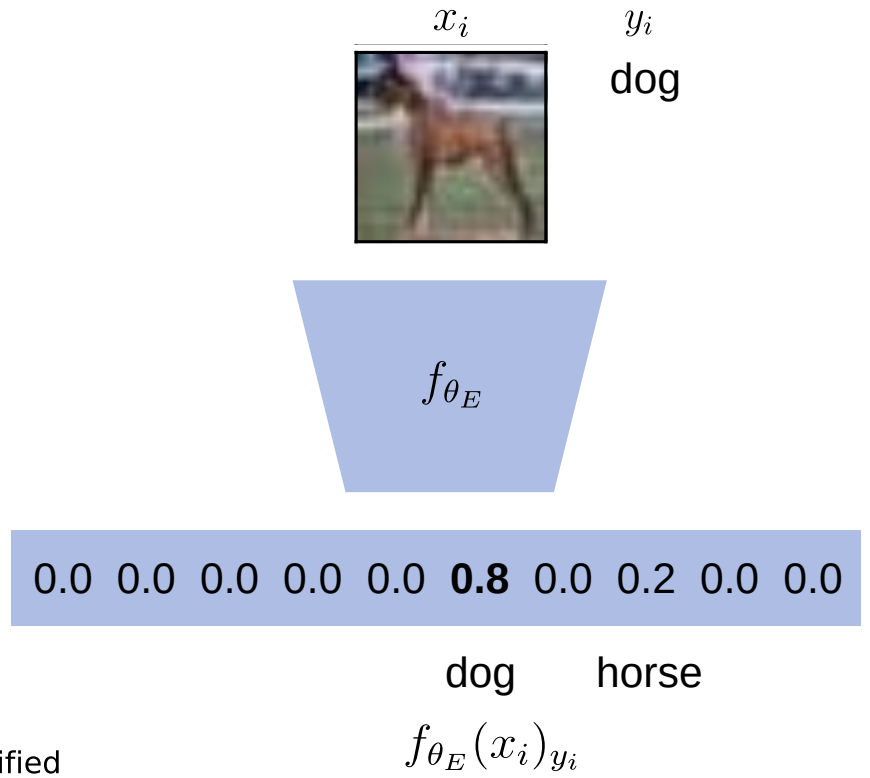
$(1 - f_{\theta_E}(x_i)_{y_i})^{\gamma}$
Importance function



# Importance weights metric

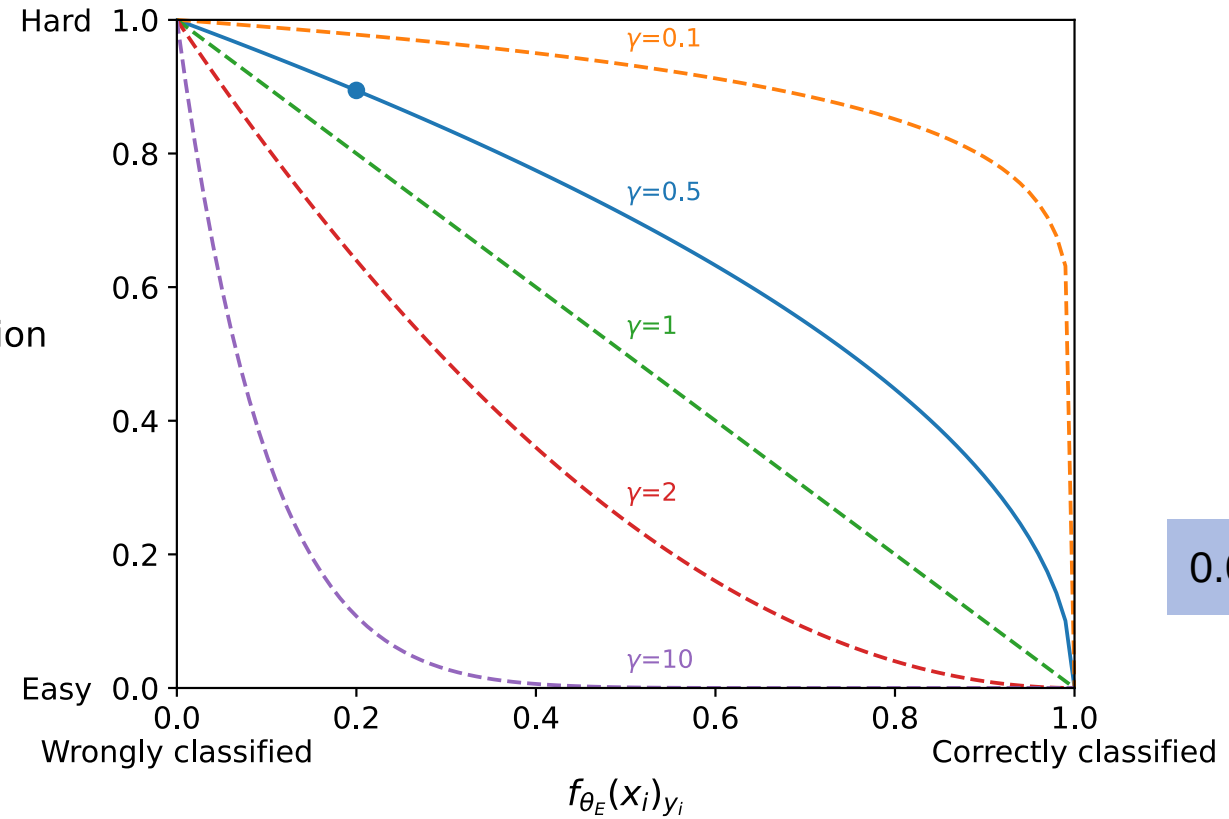
$(1 - f_{\theta_E}(x_i)_{y_i})^{\gamma}$
Importance function

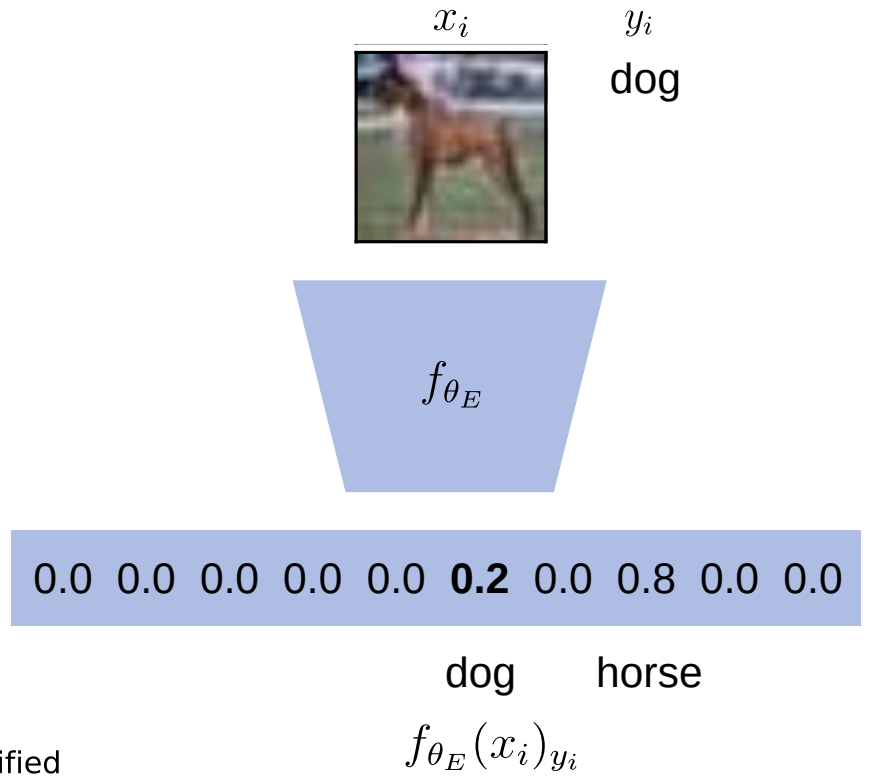


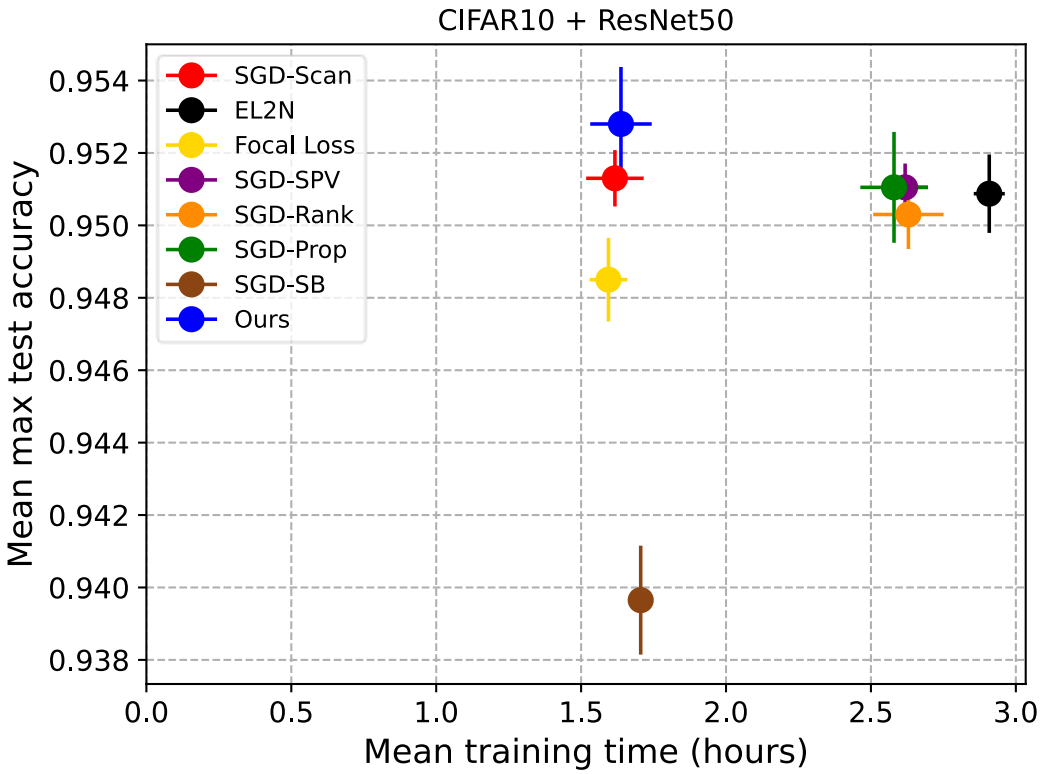
# Importance weights metric

$(1 - f_{\theta_E}(x_i)_{y_i})^{\gamma}$
Importance function

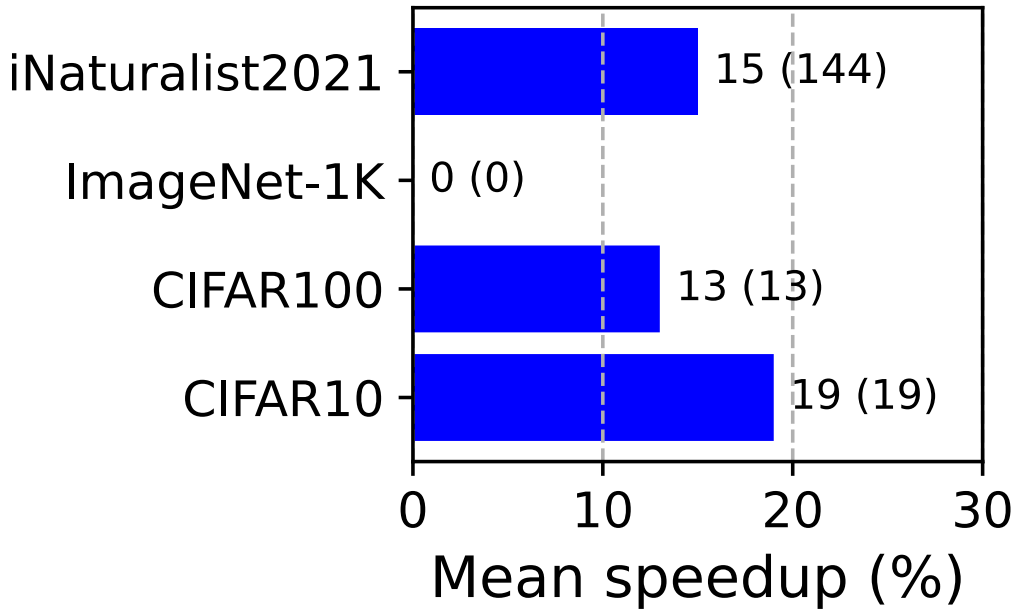


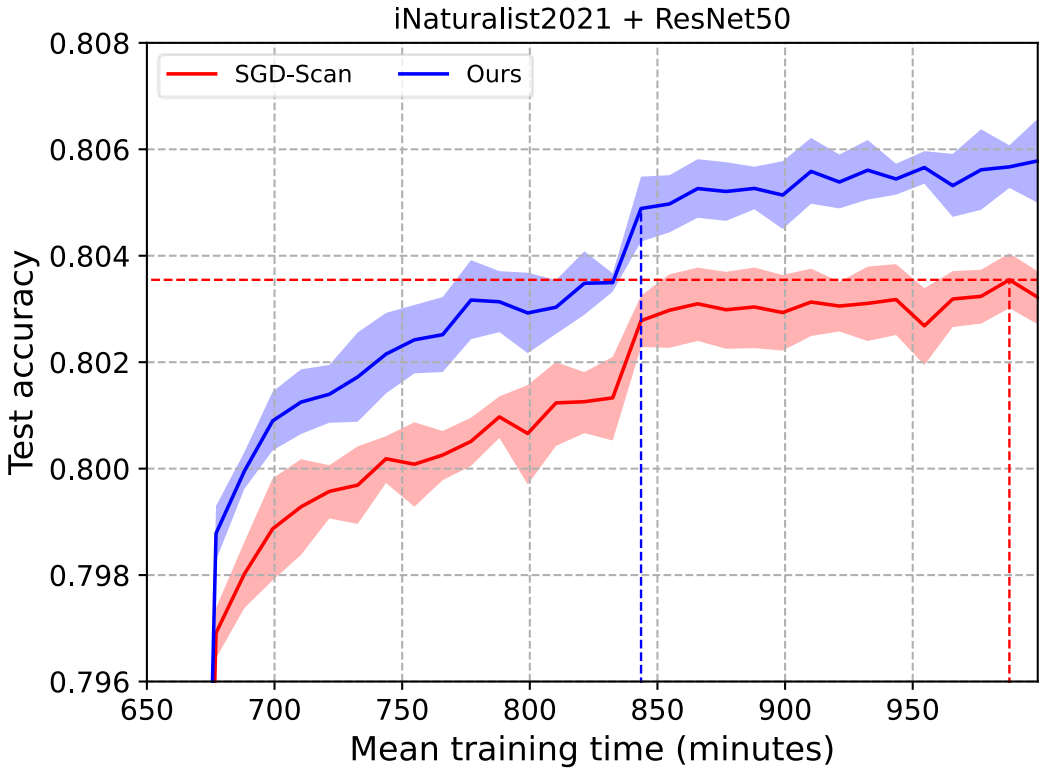
# Test accuracy is improved with minimal time cost



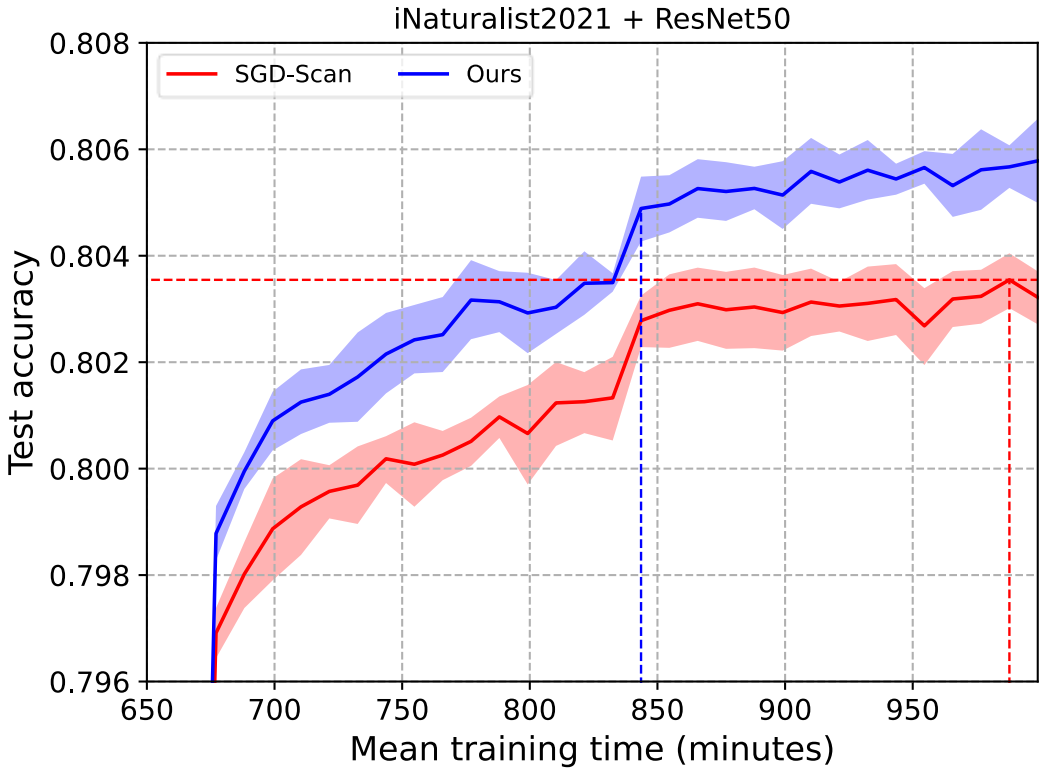
# Speedup is achieved w.r.t. the baseline‘s max accuracy

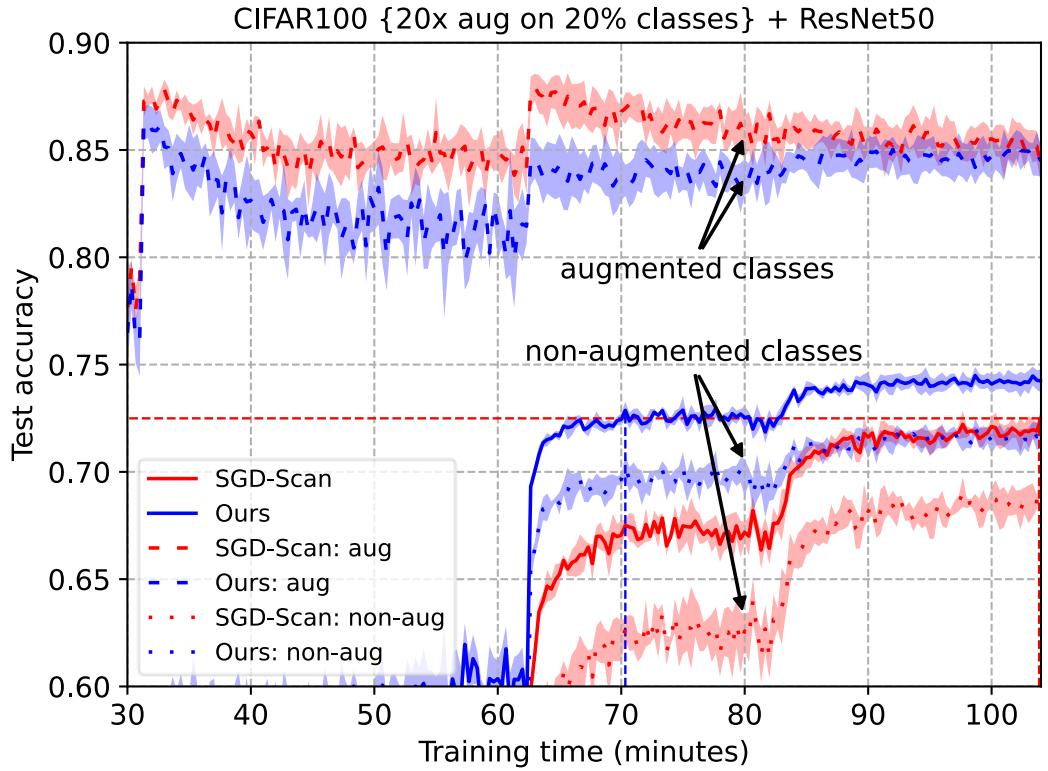


# Classes are automatically balanced

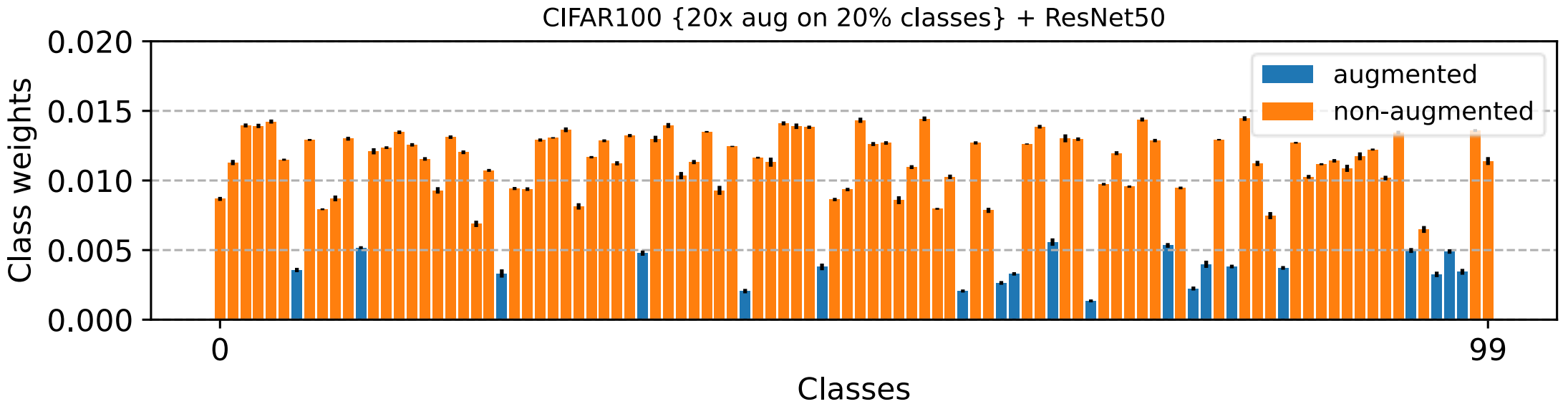


# Classes are automatically balanced



# Classes are automatically balanced



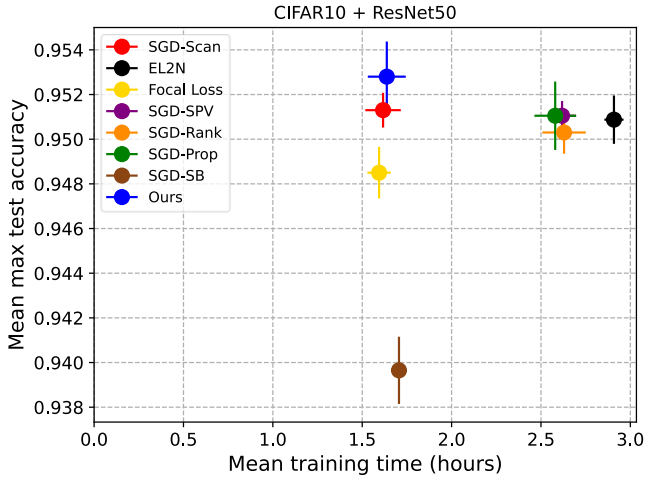
# Conclusion

Using our importance scheme allows to:

# Conclusion

Using our importance scheme allows to:

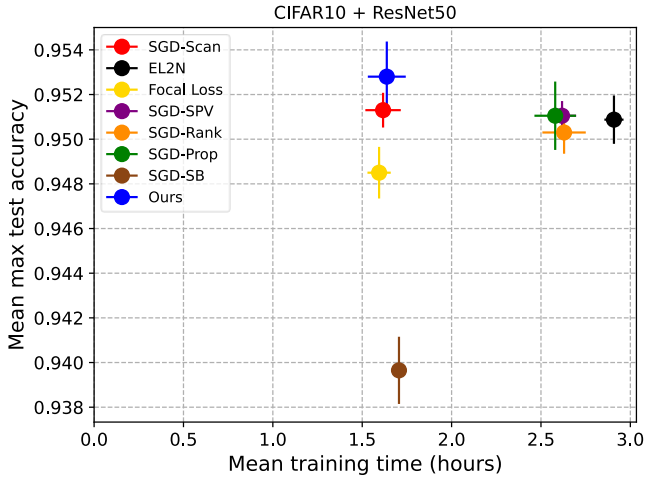
- Improve performance with minimal cost

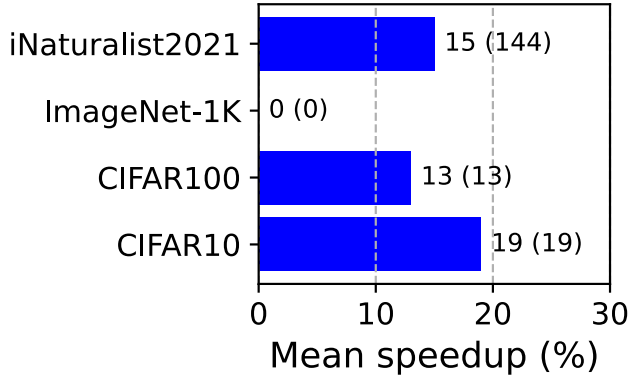


# Conclusion

Using our importance scheme allows to:

- Improve performance with minimal cost
- Speedup learning

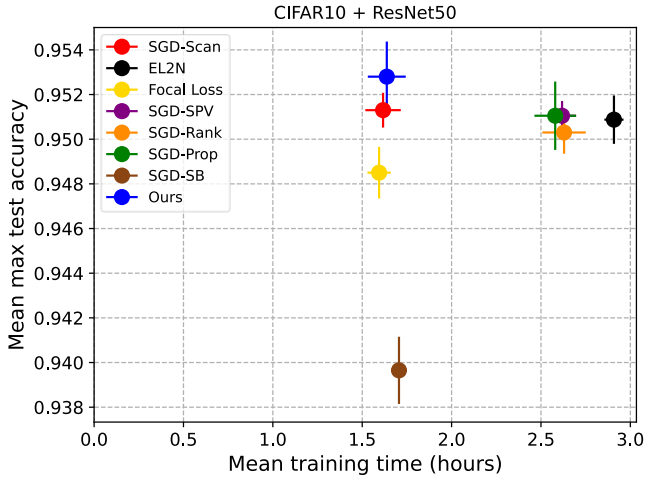


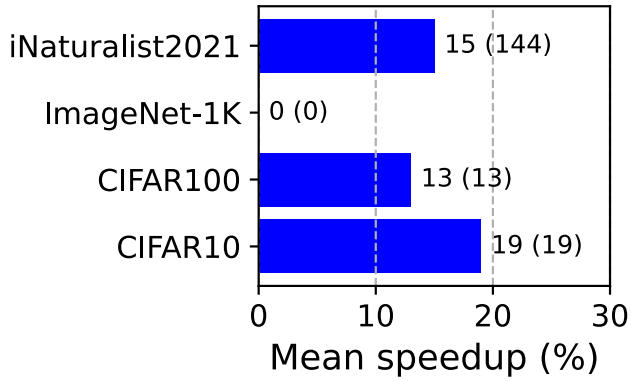



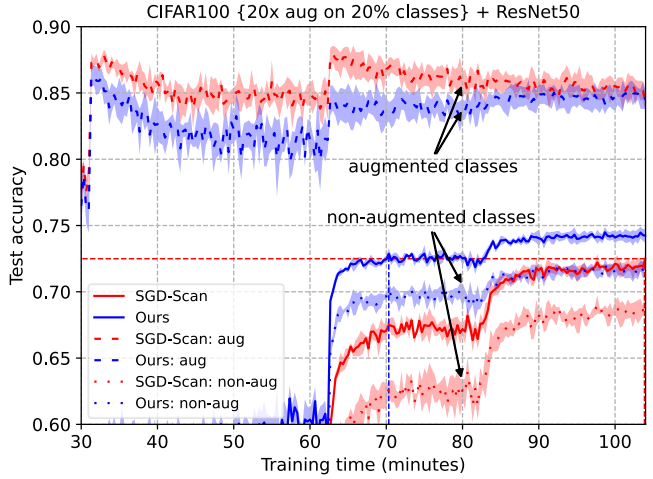
# Conclusion

Using our importance scheme allows to:

- Improve performance with minimal cost
- Speedup learning
- Automatically balance classes



# Conclusion

Using our importance scheme allows to:

- Improve performance with minimal cost
- Speedup learning
- Automatically balance classes

Contact: a.quercia@fz-juelich.de

Acknowledgements:

- This work was funded by the Helmholtz School for Data Science in Life, Earth, and Energy (HDS-LEE).
- The authors gratefully acknowledge the computing time granted by the JARA Vergabegremium and provided on the JARA Partition part of the supercomputer JURECA at Forschungszentrum Jülich.
- The authors kindly thank Mansheej Paul for providing the EL2N scores.