CentraleSupélec | université PARIS-SACLAY

EUGLOH
European University Alliance for Global Health

Co-funded by the European Union

L2S Laboratoire Signaux & Systèmes

Blur Now, Learn Later: Temporary Resolution Reduction in CNNs

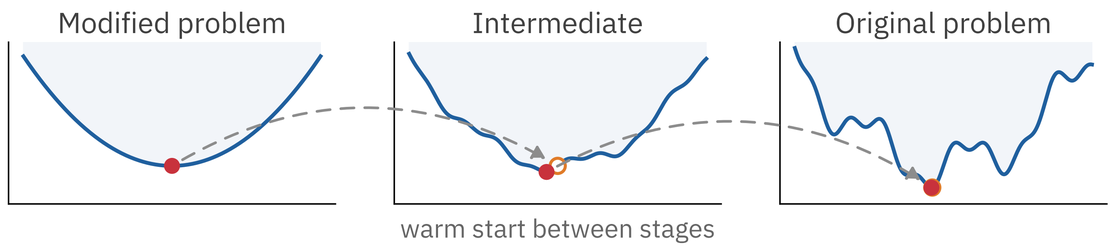
Aubin Charley Alexandre Corrard Idriss El khamlichi Maxime Nicaise

CentraleSupélec, Université Paris-Saclay

Full paper (PDF)

Motivation

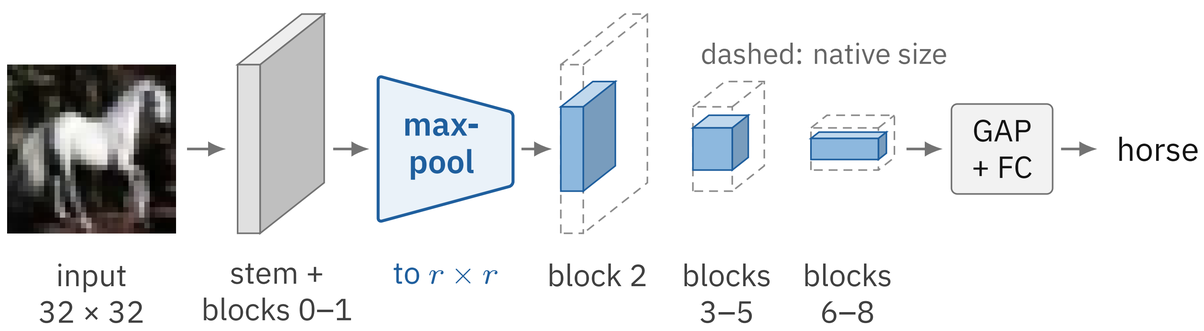
Can a temporary change of internal feature-map resolution improve the final classifier?

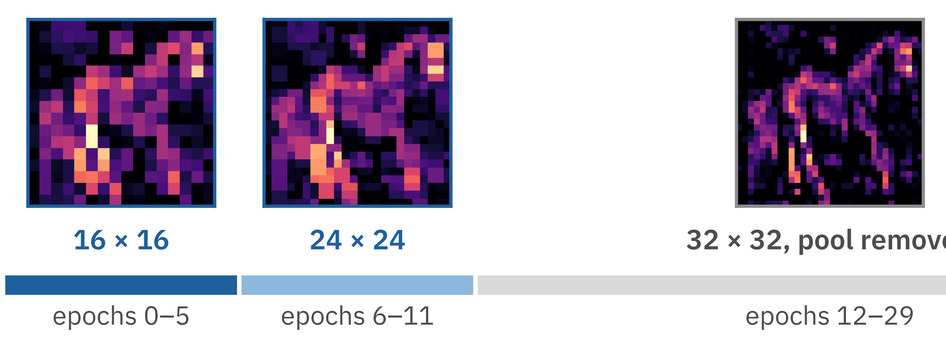


Continuation methods [1,2] train on modified problems that end at the original one.

Method

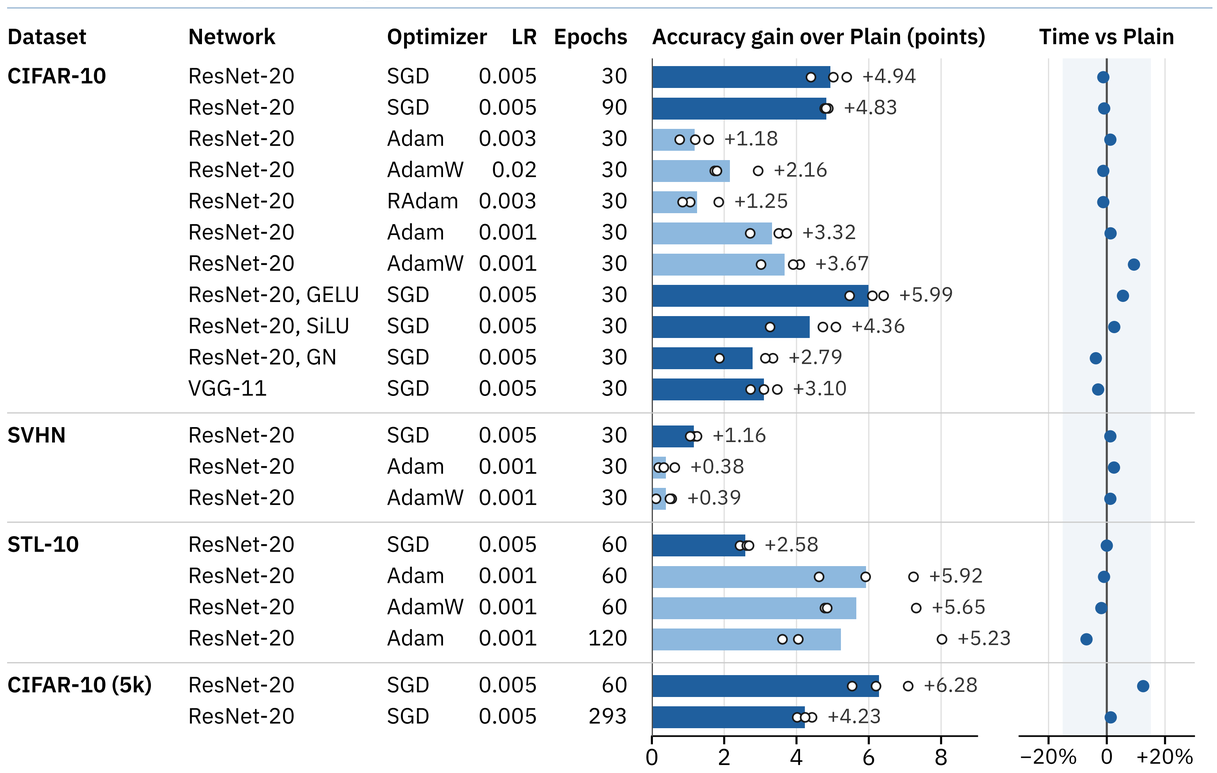
R (*Resolution*): during the first 12 of 30 epochs, an adaptive max-pool after block 1 of ResNet-20 [3] shrinks the feature maps (the "blur") to $r \times r$. Each position keeps the strongest response in its window. **Plain**: the same network trained normally.





Real block-1 feature map of a trained ResNet-20 on CIFAR-10 [4] (one channel), max-pooled as during training. Later layers process the smaller maps directly; the pool is then removed, with **no added parameters or operations at inference**.

Works in every setting, at the same cost



R beats Plain in every setting, at a similar run time. One row per setting (CIFAR-10 [4], SVHN [5], STL-10 [6]), three seeds; bars: means, dots: seeds, dark: SGD, light: Adam, AdamW, RAdam; right: run time relative to Plain at equal updates.

Matched baselines

We also compare with published methods acting on feature-map spatial information: CBS [7] blurs the maps with a decaying Gaussian; SDPoint [8] downsamples them randomly.

	30 epochs		160 epochs
Method	SGD	AdamW	AdamW
Plain	75.46 ± 0.85	84.28 ± 0.20	84.95 ± 0.26
R	**80.46** ± 0.51	86.65 ± 0.14	85.83 ± 0.65
CBS-P	78.01 ± 0.61	80.56 ± 0.24	83.33 ± 0.47
CBS-B	78.71 ± 0.90	83.23 ± 0.38	83.71 ± 0.31
SDPoint	79.83 ± 0.34	**87.57** ± 0.59	**88.94** ± 0.12

R is best with SGD; SDPoint is more accurate with AdamW. Test accuracy (%), mean ± SD; CIFAR-10, ResNet-20, no augmentation; best in bold. CBS-P: published decay; CBS-B: adapted to the budget.

Declarations of interest: none.

Main result

Systematic
Higher mean test accuracy in all 20 settings without augmentation: **+0.38 to +6.28 points**, across datasets, architectures and optimizers.

Consistent
60 of 60 seed-level comparisons with Plain are positive.

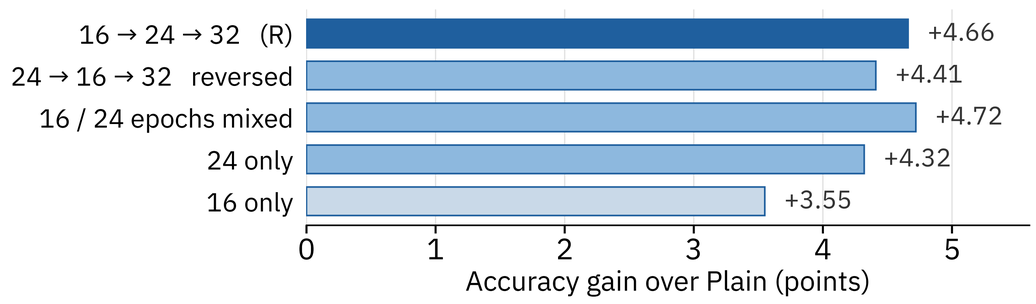
Free
Nothing added at inference; training time close to Plain's, and lower with long SGD training.

Simple
One max-pool after one block during the first 12 of 30 epochs, then removed.

Schedule ablations

Is the coarse-to-fine order needed?

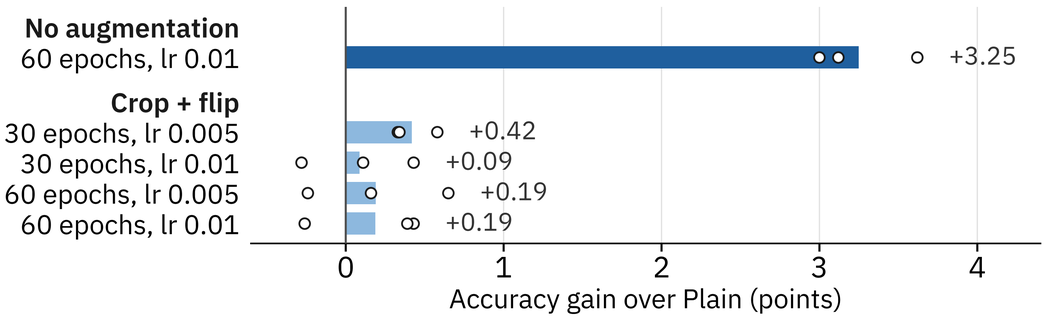
Same reduced-resolution epochs, in another order or at a single side.



Reversing or mixing keeps the gain; using only 24 × 24 keeps most of it, so the temporary reduction matters, not the progression. CIFAR-10, SGD, 30 epochs, five seeds.

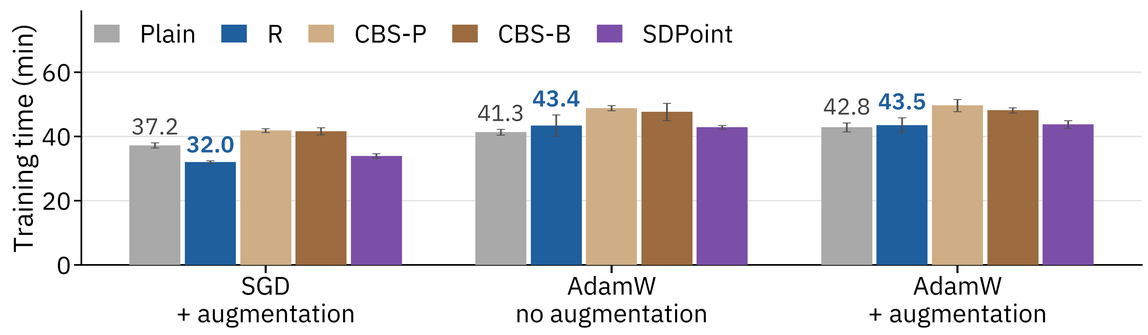
Data augmentation

Does the gain survive crop-and-flip augmentation?



With augmentation the mean gain stays positive but small (+0.09 to +0.42 points), with some seeds below Plain; in longer 160-epoch recipes (see paper) it can vanish. CIFAR-10, SGD (bars: means, dots: seeds).

Training cost



R is the fastest method with SGD and costs about as much as Plain with AdamW; in the 30-epoch reference setting it takes about as long as Plain. Training-loop time at 160 epochs, mean ± SD.

⚠ Limitations

- **Exploratory.** Pooling site, schedule and some learning rates were chosen on CIFAR-10 test results; a validation-selected control still gives **+2.49 ± 0.24 points** (three new seeds, largest tested learning rate).
- **Weaker with strong recipes.** Augmentation and long training shrink the advantage, sometimes to nothing.
- **Mechanism open.** Lower curvature does not explain the gain: SDPoint is flatter yet less accurate.
- **Small effects.** Three seeds per setting do not resolve small differences.

References

[1] Allgöwer & Georg, *Numerical Continuation Methods: An Introduction*, Springer, 1990.
[2] Mobahi & Fisher, *A Theoretical Analysis of Optimization by Gaussian Continuation*, AAAI 2015.
[3] He et al., *Deep Residual Learning for Image Recognition*, CVPR 2016.
[4] Krizhevsky, *Learning Multiple Layers of Features from Tiny Images*, Tech. report, 2009.
[5] Netzer et al., *Reading Digits in Natural Images with Unsupervised Feature Learning*, NIPS Workshop, 2011.
[6] Coates et al., *An Analysis of Single-Layer Networks in Unsupervised Feature Learning*, AISTATS 2011.
[7] Sinha et al., *Curriculum by Smoothing*, NeurIPS 2020.
[8] Kuen et al., *Stochastic Downsampling for Cost-Adjustable Inference and Improved Regularization in CNNs*, CVPR 2018.