Affinity and Diversity: Quantifying Mechanisms of Data Augmentation

Raphael Gontijo-Lopes* Sylvia J. Smullin* Ekin D. Cubuk Ethan Dyer

Google Brain Google Google Brain Google

Data aug is a crucial component of modern ML

data augmentation deep learning

About 112,000 results (**0.10** sec)

The effectiveness of data augmentation in image classification using deep learning

L Perez, J Wang - arXiv preprint arXiv:1712.04621, 2017 - arxiv.org

In this paper, we explore and compare multiple solutions to the problem of **data augmentation** in image classification. Previous work has demonstrated the effectiveness of **data augmentation** through simple techniques, such as cropping, rotating, and flipping input …

Cited by 477 Related articles All 3 versions

[PDF] **The effectiveness of data augmentation in image classification using deep learning**

J Wang, L Perez - Convolutional Neural Networks Vis. Recognit, 2017 - academia.edu

In this paper, we explore and compare multiple solutions to the problem of **data augmentation** in image classification. Previous work has demonstrated the effectiveness of

Many heuristics for why data augmentation works

1) Simulates realistic samples from the true data distribution: *"[augmentation strategies are] reasonable since the transformed reference data is now extremely close to the original data. In this way, the amount of training data is effectively increased"* (Bellegarda et al., 1992).
2) Adds diversity of data
3) Increases the effective dataset size

Each does not seem to tell the full story.

Without Data Augmentation Science, these remain unverified heuristics.

We seek to quantify and verify them.

Affinity and Diversity

Metrics inspired by these heuristics

Affinity: clean-trained model's test accuracy on augmented data (inspired by distribution shift)

Diversity: augmented-trained model's final training loss (how hard augmented data is to fit)

Affinity (aka “robustness”)

Model trained on distribution A approximately captures that distribution.

Low performance on distribution B is indicative of how out-of-distribution B is wrt A, <u>for this model</u>.

If you had a perfect model (that generalized to every unseen distribution) then it would be perfectly robust, AND every distribution would be within the distribution captured by the model.
(Also you probably don’t need to care about extra data augmentations ;))

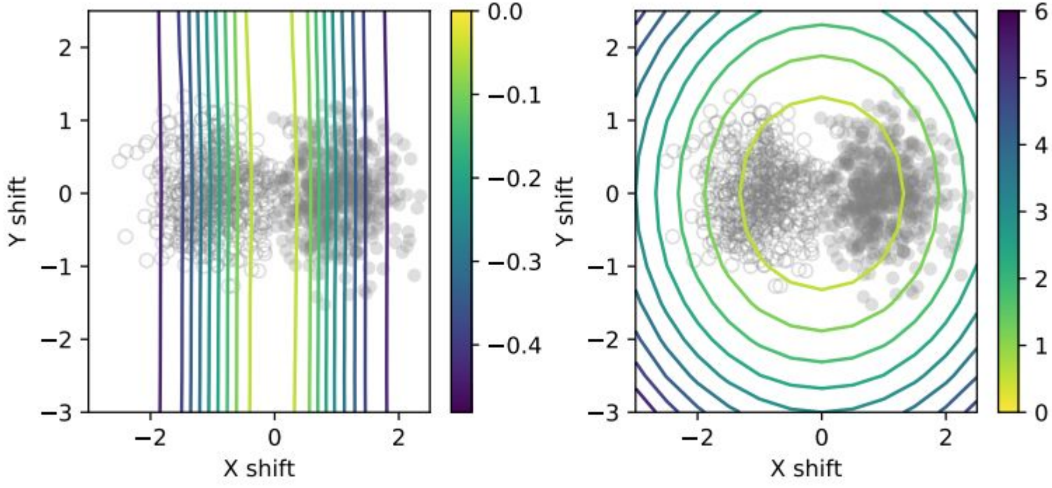
Diversity

Final training likelihood == encoding size of data (Shannon, 1948)

Thus, Diversity is “how hard it is to fit the data.”

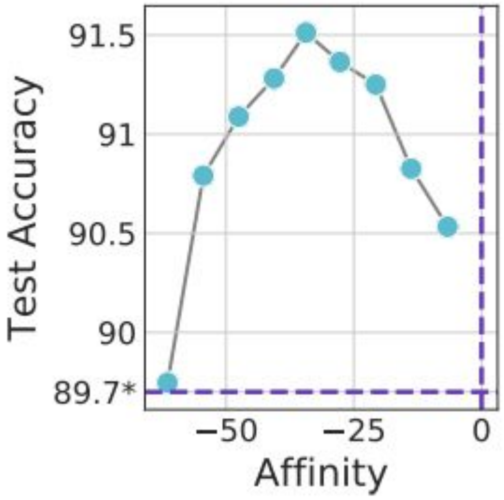
Model-specific metrics

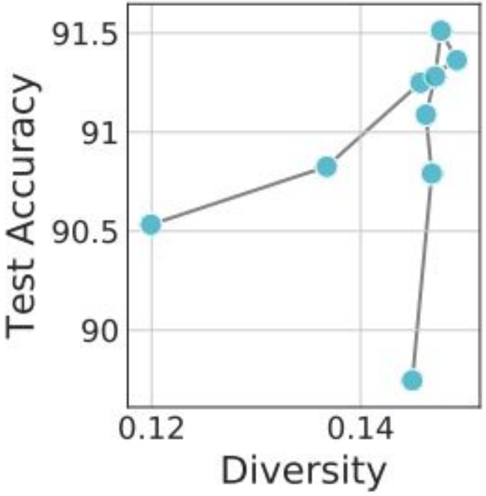
That’s an advantage: when training models, we care about how much the augmented data will shift *with respect to the model’s decision boundary*.

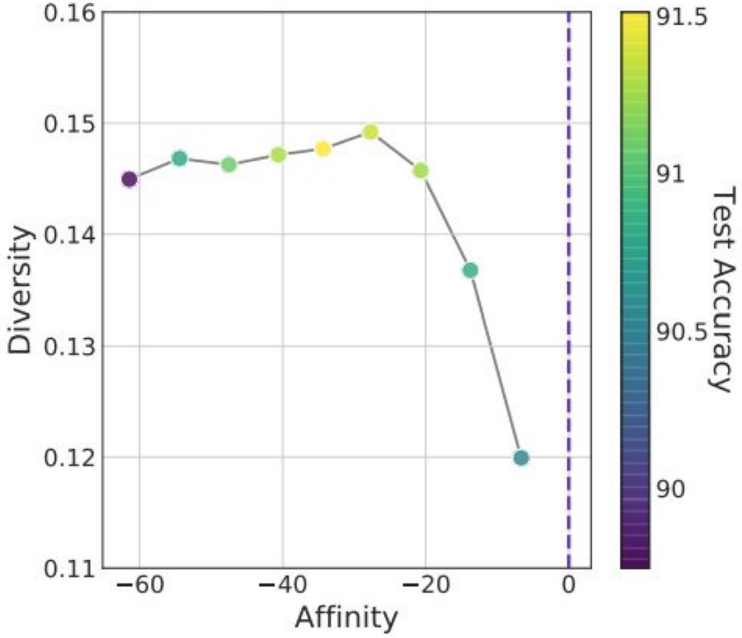


(a) Affinity (b) D_{KL}

Performance is explained by BOTH Affinity and Diversity

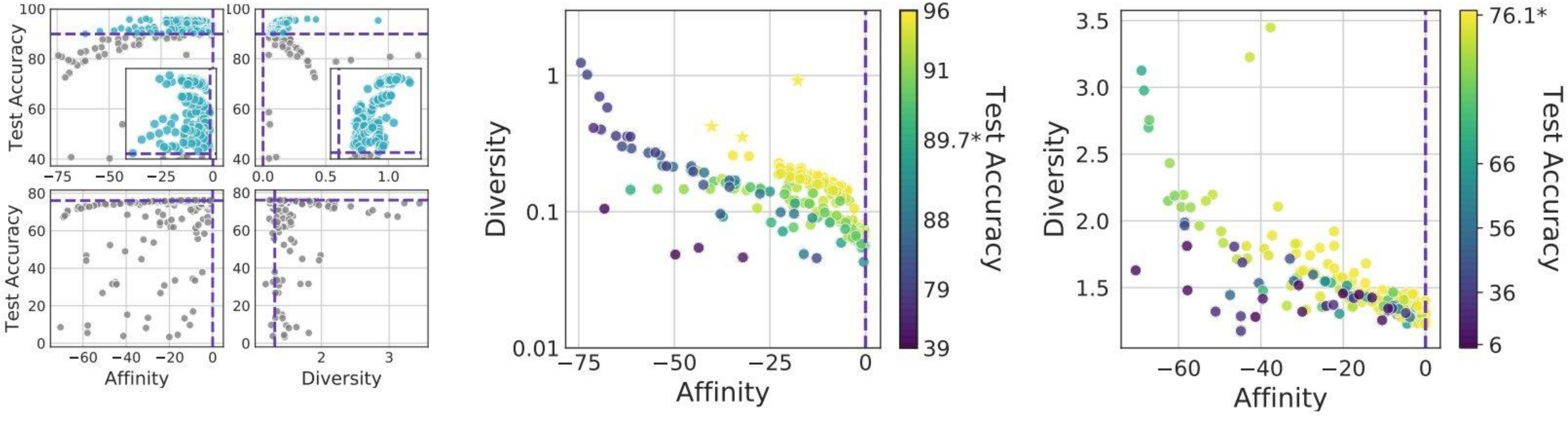






Rotate(fixed, 60deg) on CIFAR-10 is varied from 10% to 90% apply probability.

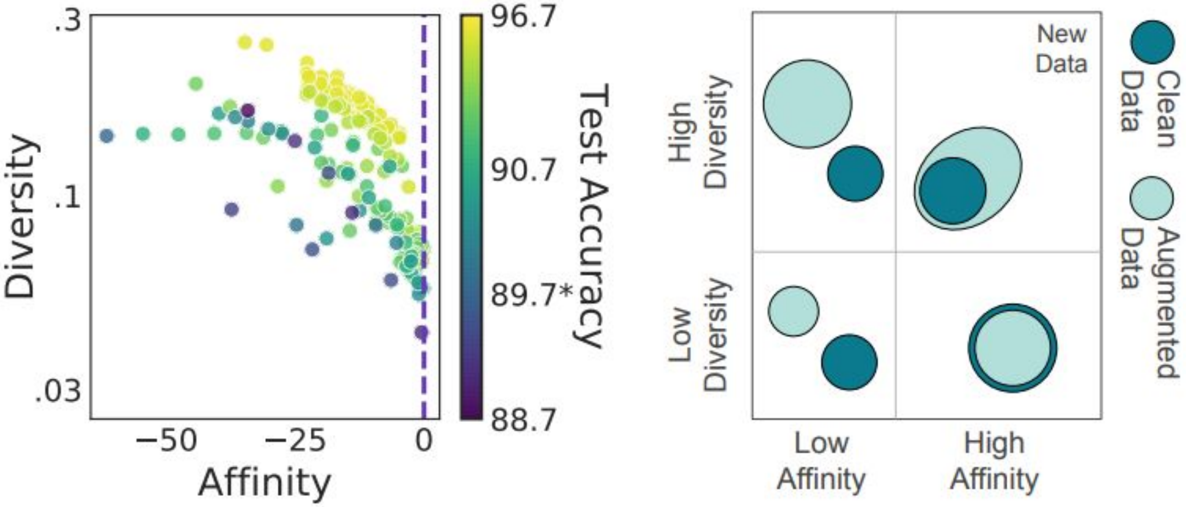
Performance is explained by BOTH Affinity and Diversity



(a) CIFAR-10 (top) ImageNet (bottom) (b) CIFAR-10 (c) ImageNet

Each dot is one augmentation with a specific set of hyper-parameters

Performance is explained by BOTH Affinity and Diversity

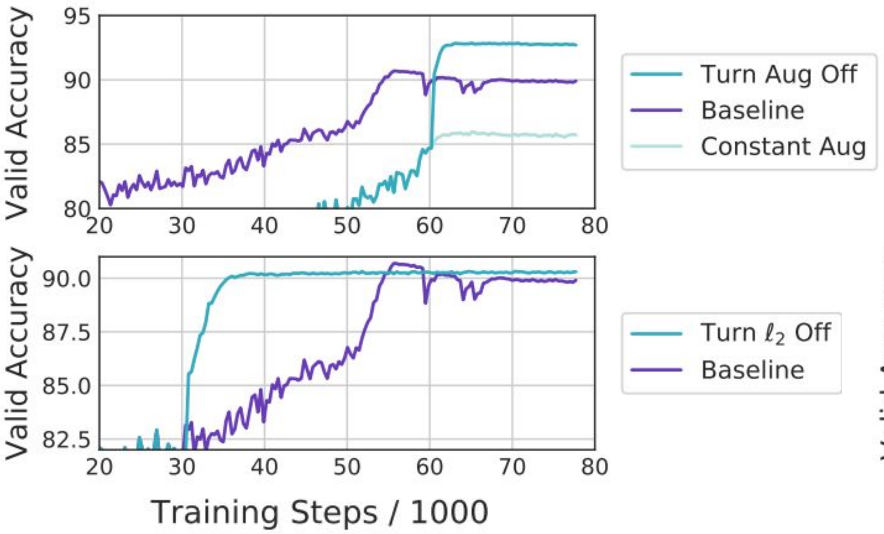


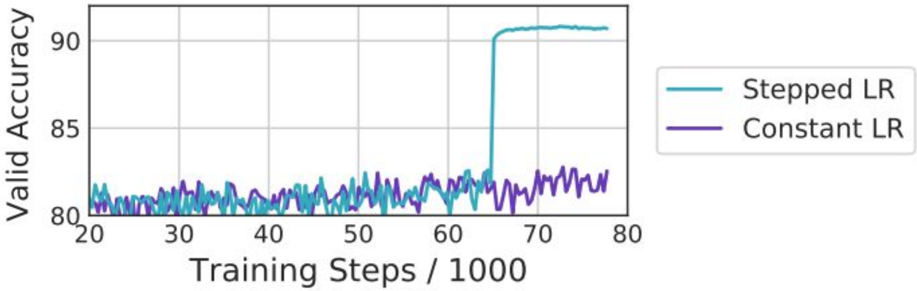
(a) Affinity vs Diversity (b) Model's View of Data

Same as before, but on augs that drop at most 1% in performance

Augmentations behave like regularizers

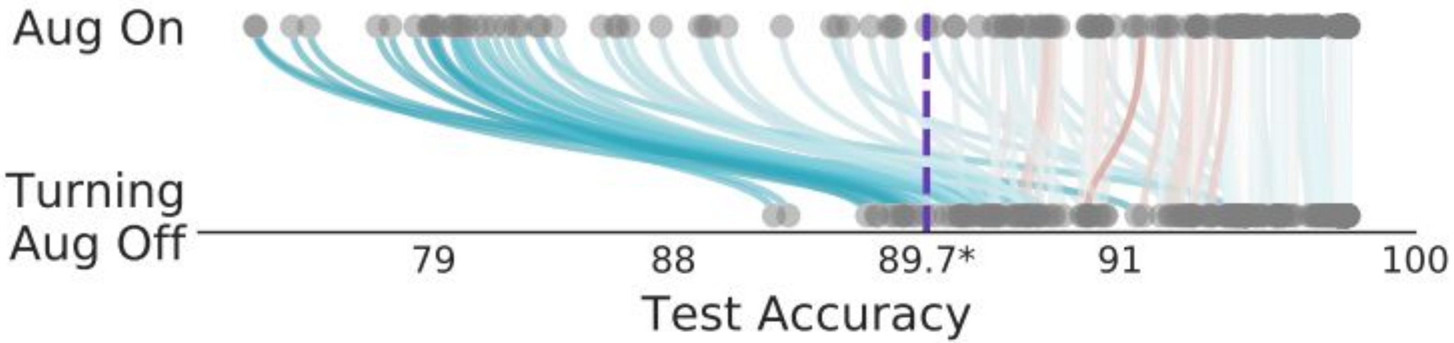
Switching them off part-way through training often helps boost final accuracy.
(phenomena of regularizers observed by various works separately)





Augmentations behave like regularizers

Importantly: sometimes augmentations that would hurt otherwise, start helping when switched off (you can image: similar to learning rate)

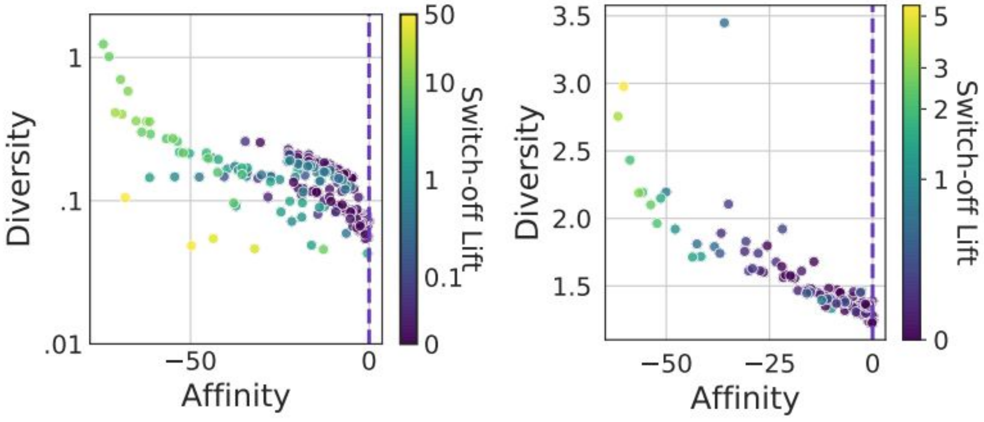


(b) Switch-off Lift on CIFAR-10

Why does switching off help?

We hypothesize that switching augmentation off adjusts the training-time Affinity and Diversity to match that of test-time.

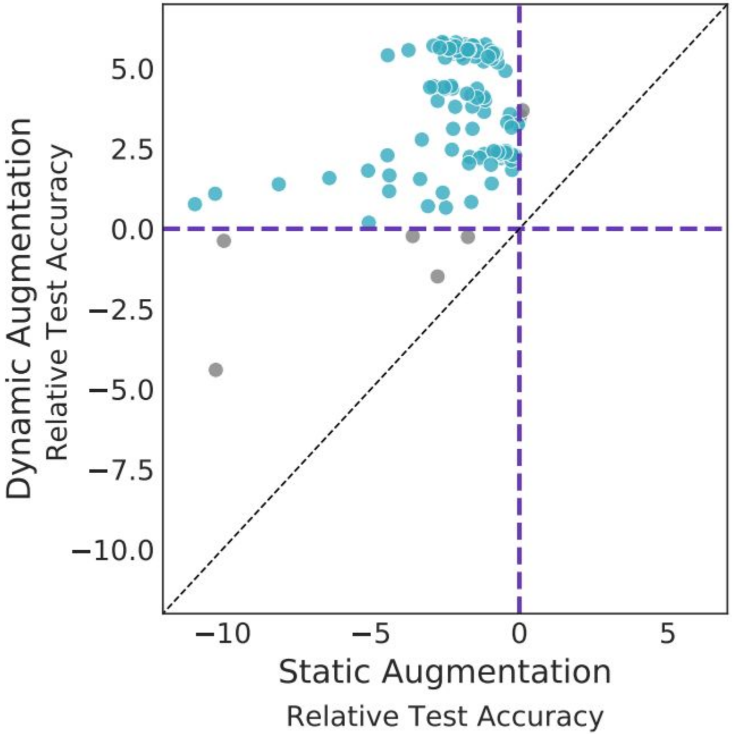
Indeed, Affinity and Diversity *together* can predict the lift from switching augmentation off.



(a) CIFAR-10 (b) ImageNet

Increased effective dataset size is key

When an augmentation is not allowed to increase the effective dataset size ("static augmentation"), it never improves over the baseline.



Key Takeaway

Increased effective training set size is crucial to the performance benefit of data augmentation.

An augmentation's Affinity and Diversity inform how useful the additional training examples are.

Keep in touch!

iraphael@google.com

[@iraphas13](https://twitter.com/iraphas13)

Questions?