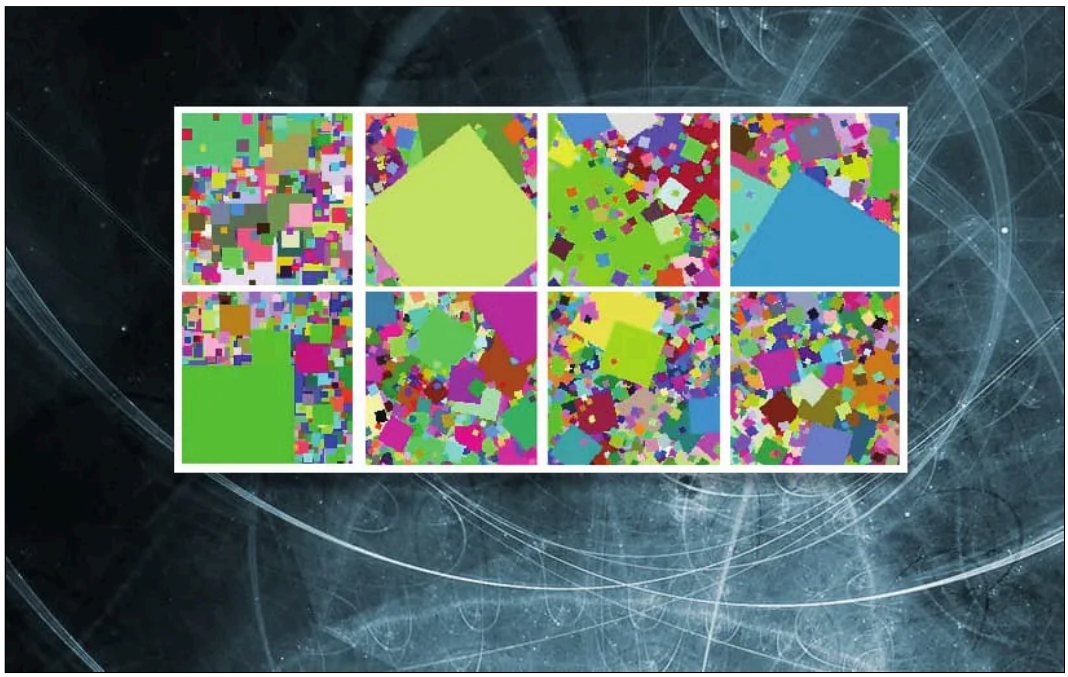
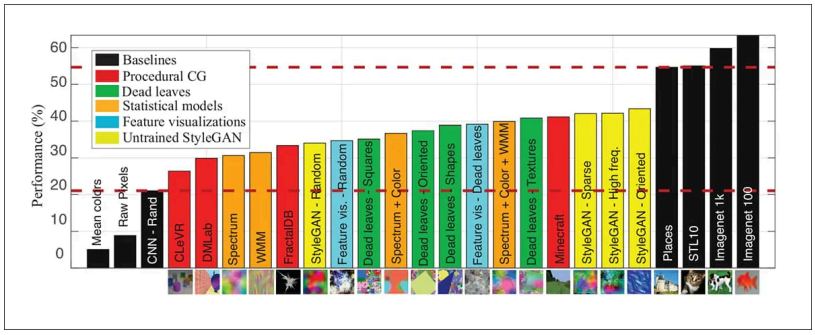
Training Computer Vision Models on Random Noise Instead of Real Images

Originally published at https://www.unite.ai/training-computer-vision-models-on-random-noise-instead-of-real-images/



Researchers from MIT Computer Science & Artificial Intelligence Laboratory (CSAIL) have experimented with using random noise images in computer vision datasets to train computer vision models, and have found that instead of producing garbage, the method is surprisingly effective:



Generative models from the experiment, sorted by performance. Source: https://openreview.net/pdf?id=RQUl8gZnN7O

Feeding apparent 'visual trash' into popular computer vision architectures should not result in this kind of performance. On the far right of the image above, the black columns represent accuracy scores (on Imagenet-100) for four 'real' datasets. While the 'random noise' datasets preceding it (pictured in various colors, see index top-left) can't match that, they are nearly all within respectable upper and lower bounds (red dashed lines) for accuracy.

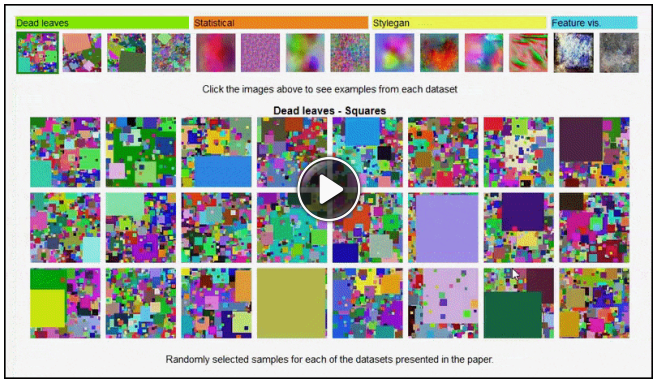
In this sense 'accuracy' does not mean that a result necessarily looks like a face, a church, a pizza, or any other particular domain for which you might be interested in creating an image synthesis system, such as a Generative Adversarial Network, or an encoder/decoder framework.

Rather, it means that the CSAIL models have derived broadly applicable central 'truths' from image data so apparently unstructured that it should not be capable of supplying it.

Diversity Vs. Naturalism

Neither can these results be attributed to over-fitting: a lively discussion between the authors and reviewers at Open Review reveals that mixing different content from visually diverse datasets (such as 'dead leaves', 'fractals' and 'procedural noise' – see image below) into a training dataset actually *improves* accuracy in these experiments.

This suggests (and it's a bit of a revolutionary notion) a new type of 'under-fitting', where 'diversity' trumps 'naturalism'.



CLICK TO VIEW ANIMATED GIF

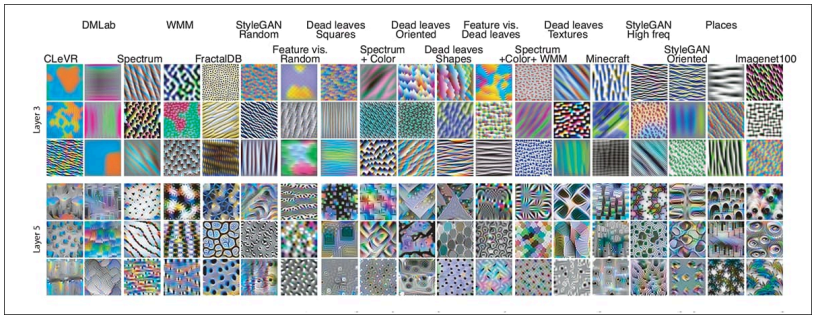
The project page for the initiative lets you interactively view the different types of random image datasets used in the experiment.
Source: https://mbaradad.github.io/learning_with_noise/

The results obtained by the researchers call into question the fundamental relationship between image-based neural networks and the 'real world' images that are thrown at them in alarmingly greater volumes each year, and imply that the need to obtain, curate and otherwise wrangle hyperscale image datasets may eventually become redundant. The authors state:

> *'Current vision systems are trained on huge datasets, and these datasets come with costs: curation is expensive, they inherit human biases, and there are concerns over privacy and usage rights. To counter these costs, interest has surged in learning from cheaper data sources, such as unlabeled images.*
>
> *'In this paper, we go a step further and ask if we can do away with real image datasets entirely, by learning from procedural noise processes.'*

The researchers suggest that the current crop of machine learning architectures may be inferring something far more fundamental (or, at least, unexpected) from images than was previously thought, and that 'nonsense' images can potentially impart a great deal of this knowledge far more cheaply, even with the possible use of ad hoc synthetic data, via dataset-generation architectures that generate random images at training time:

> '*We identify two key properties that make for good synthetic data for training vision systems: 1)naturalism, 2) diversity. Interestingly, the most naturalistic data is not always the best, since naturalism can come at the cost of diversity.*
>
> *'The fact that naturalistic data help may not be surprising, and it suggests that indeed, large-scale real data has value. However, we find that what is crucial is not that the data be* real *but that it be* naturalistic*, i.e. it must capture certain structural properties of real data.*
>
> *'Many of these properties can be captured in simple noise models.'*



Feature visualizations resulting from an AlexNet-derived encoder on some of the various 'random image' datasets used by the authors, covering the 3rd and 5th (f methodology used here follows that set out in Google AI research from 2017.

The paper, presented at the 35th Conference on Neural Information Processing Systems (NeurIPS 2021) in Sydney, is titled *Learning to See by Looking at Noise*, and comes from six researchers at CSAIL, with equal contribution.

The work was recommended by consensus for a spotlight selection at NeurIPS 2021, with peer commenters characterizing the paper as 'a scientific breakthrough' that opens up a 'great area of study', even if it raises as many questions as it answers.

In the paper, the authors conclude:

> *'We have shown that, when designed using results from past research on natural image statistics, these datasets can successfully train visual representations. We hope that this paper will motivate the study of new generative models capable of producing structured noise achieving even higher performance when used in a diverse set of visual tasks.*
>
> *'Would it be possible to match the performance obtained with ImageNet pretraining? Maybe in the absence of a large training set specific to a particular task, the best pre-training might not be using a standard real dataset such as ImageNet.'*

Writer on machine learning, domain specialist in human image synthesis. Former head of research content at Metaphysic.ai, until its dissolution into DNEG's Brahma.ai.
Portfolio site: martinanderson.ai
Contact: martin@martinanderson.ai

Discover More