

Trustworthy and Reliable Large-Scale Machine Learning Models
Workshop at ICLR 2023
(Highlighted Talk)

Leaving Reality to Imagination: Robust Classification via Generated Datasets

Hritik Bansal








Aditya Grover



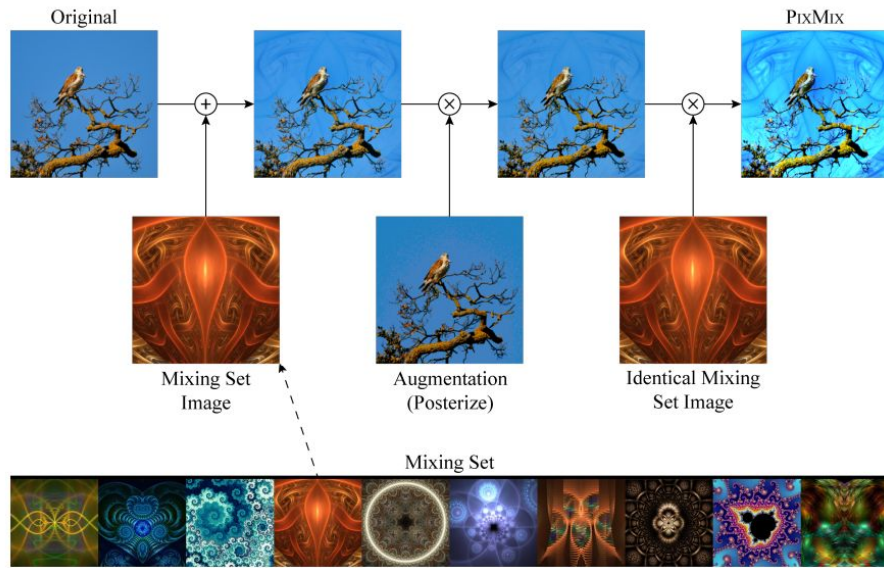
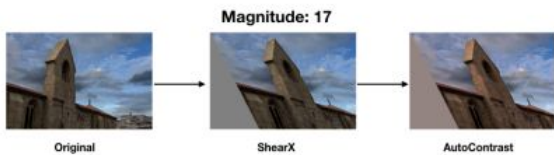
Generalization

- Ultimate aim of machine learning is train models that generalize beyond the training distribution.
- **Example:** ResNet-101 trained on ImageNet-1K training set performs well on its test set but undergoes performance degradation on **natural variations of the same objects** in the dataset i.e., rendition, sketches, and view-points.

	Dataset Examples	ImageNet ResNet101
ImageNet		76.2
ImageNetV2		64.3
ImageNet-R		37.7
ObjectNet		32.6
ImageNet Sketch		25.2

Enlarging Training Data Improves Generalization

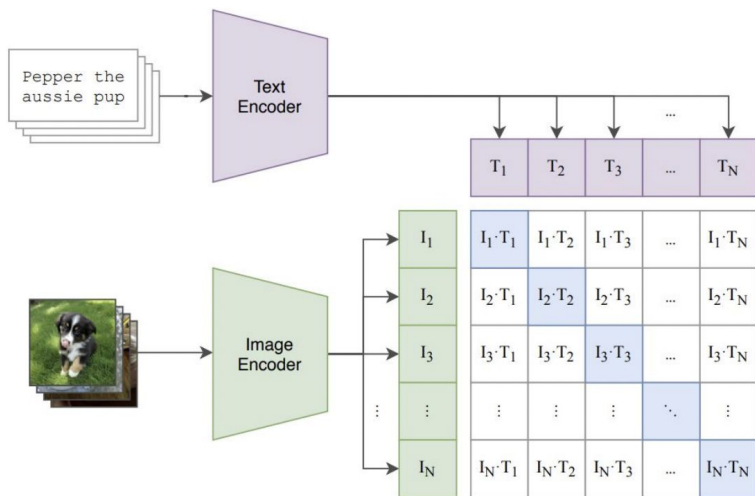
1. Data Augmentation



Enlarging Training Data Improves Generalization

2. Large-scale pretraining on “in-the-wild” diverse data

CLIP: 400M Image-Text, ALIGN: 2B Image-Text



	Dataset Examples	ImageNet ResNet101	Zero-Shot CLIP	Δ Score
ImageNet		76.2	76.2	0%
ImageNetV2		64.3	70.1	+5.8%
ImageNet-R		37.7	88.9	+51.2%
ObjectNet		32.6	72.3	+39.7%
ImageNet Sketch		25.2	60.2	+35.0%

Paradigm shift in the notion of a dataset



Stable Diffusion



min-DALL.E



ediff-1



Midjourney



DALL.E-2



DALL.E-mini











ImaGen




GLIDE


Generated data is pervading the internet!


image (image)	prompt (string)	seed (uint32)	step (uint16)	cfg (float32)	sampler (string)	width (uint16)	height (uint16)
	"a renaissance portrait of..."	2,480,545,905	50	16	"k_euler_ancestral"	512	768
	"portrait of a dancing eagle..."	2,250,159,284	50	9	"k_lms"	512	640
	"epic 3 d, become legend shiji! gp..."	4,292,948,605	50	7	"k_lms"	512	768
	"an airbrush painting of cybe..."	2,374,713,726	50	12	"k_lms"	512	768
	"concept art of a silent hill..."	2,320,897,141	50	6	"k_lms"	640	512


openaidalle  [Follow](#) [Message](#)  

734 posts 571K followers 0 following



 **midjourney** [Join](#)
r/midjourney



 **StableDiffusion** [Join](#)
r/StableDiffusion

[Posts](#) [Wiki](#) [Stable Diffusion](#) ▼ [Downloads](#) ▼ [Tutorials](#)

Characteristics of the Modern Text to Image Generative Models

1. Pre-trained on a large-scale diverse dataset, e.g., LAION-2B, with **open vocabulary** annotations.
2. Can generate high-fidelity photorealistic images for a wide range of concepts.
3. **Infinite data generators:** Can be queried repeatedly through various conditioning mechanisms.

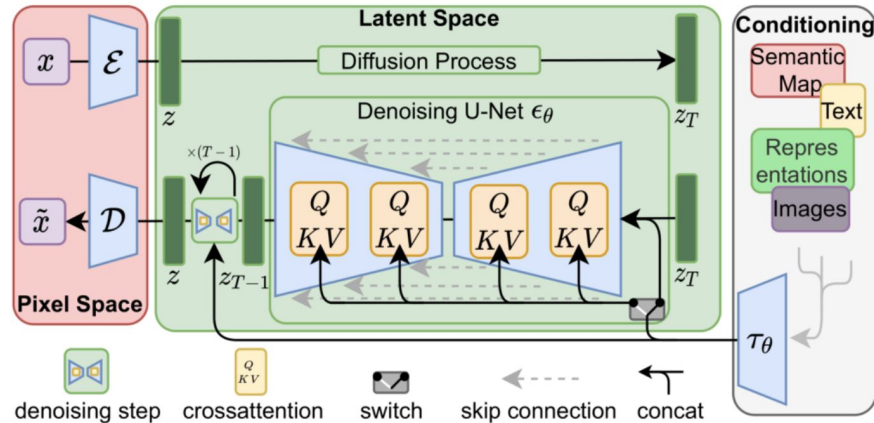
Contribution

- We show that training image classifiers on real data augmented with the generated data improves its **accuracy** and **robustness** on natural distribution shift datasets over standard training, and popular data augmentation techniques.
- We introduce an **evolving** dataset **ImageNet-G-v1** for better evaluation, critique, and design of generated datasets.
 - Why evolving? With improvements in generative modeling, the quality of generated datasets will improve.

Training w/ Generated Data

- **Training Datasets (Real and Generated):** ImageNet-1K (1.3M examples), ImageNet-100 (130K examples)
- **Generative Model:** Stable Diffusion V1.5
- **Classifiers:** ResNet-18, ResNeXt-50, ResNeXt-101
- **Eval Datasets:** Im-test set, Im-V2, Im-Rendition, Im-Sketch, ObjectNet.

Generating Data



Stable Diffusion Architecture (image from [paper](#))

- List of class labels in ImageNet dataset e.g., *goldfish*, *stingray*, *labrador* etc.
- List of 80 diverse templates to create prompts for the class labels e.g., *a photo of a [X]*, *a sculpture of a [X]* etc.
- Example prompts: *a photo of a **goldfish***, *a sculpture of a **stingray*** etc.

Evaluation

- **Accuracy** on the natural distribution shift datasets
- **Effective Robustness** on the natural distribution shift datasets
 - Given real test dataset \mathbf{D} (ImageNet-test set), and shifted dataset \mathbf{D}' (ImageNet-Sketch)
 - We perform a linear fit on the accuracies of the standard classifiers \mathbf{F} on the shifted datasets $\mathbf{A}(\mathbf{D}')$, and test dataset $\mathbf{A}(\mathbf{D})$ as $\mathbf{A}(\mathbf{D}') = m\mathbf{A}(\mathbf{D}) + c$
 - For a robust classifier \mathbf{f}' with natural distribution accuracy $\mathbf{B}(\mathbf{D}')$ and test accuracy $\mathbf{B}(\mathbf{D})$ – we can calculate effective robustness as $\mathbf{ER}(\mathbf{f}', \mathbf{D}', \mathbf{D}) = \mathbf{B}(\mathbf{D}') - (m\mathbf{B}(\mathbf{D}) + c)$
 - $\mathbf{ER} > 0$ indicates that the robust classifier is better than the standard training on the shifted datasets when standard training achieves the same accuracy as the robust classifier on the real test dataset.

Results - Accuracy

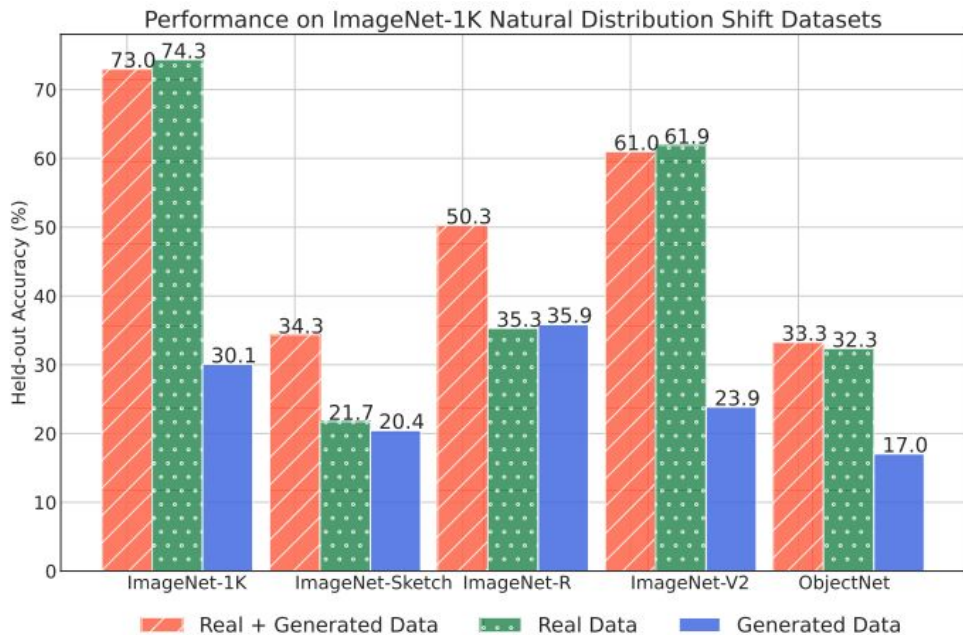
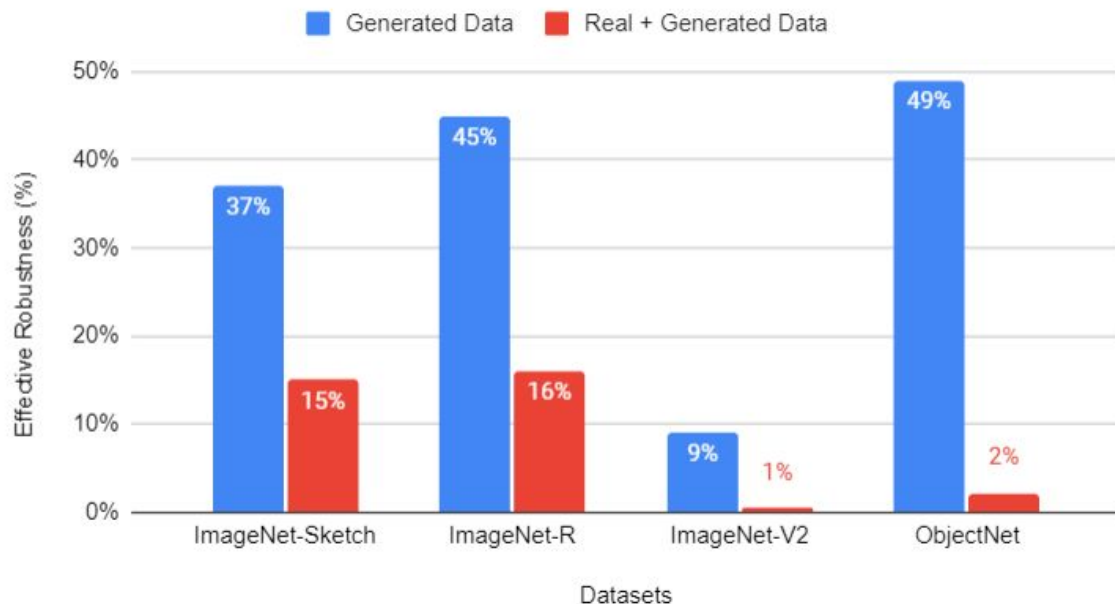


Figure 1: Accuracy of ImageNet-1K classifiers on ImageNet-1K validation set and its natural distribution shift datasets.

- In majority of the cases, the accuracy on natural distribution shifted datasets is **higher** with **real + generated** data as compared to training solely on the **real** or **generated** data.
- Large gains on ImageNet-Sketch and ImageNet-R by training on real + generated datasets! Solely training on the generated datasets yields lower performance.
- Behaviour on ImageNet-test and ImageNet-V2 is similar.

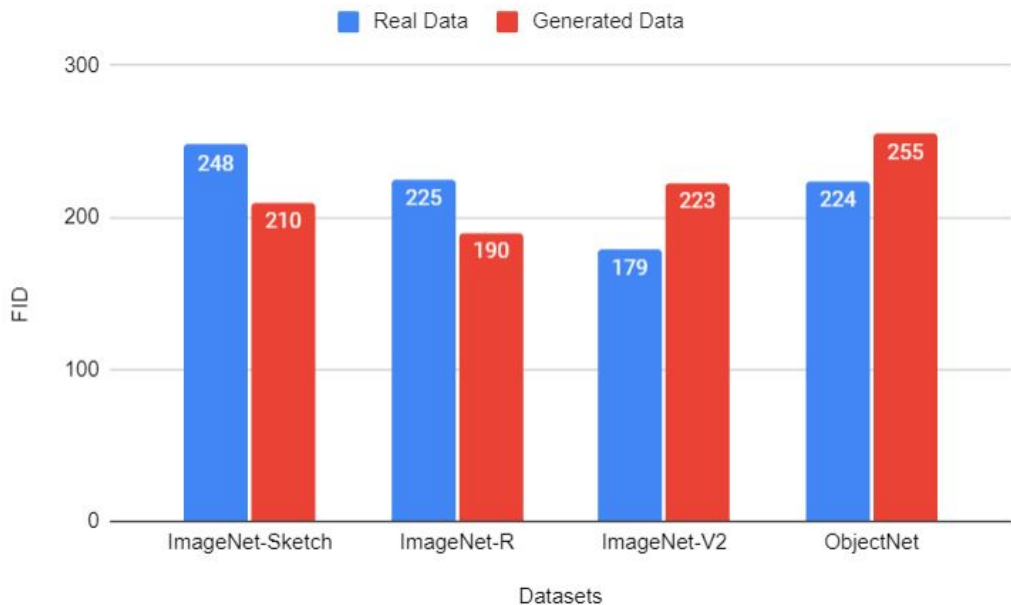
Results - Effective Robustness

Effective Robustness



- Classifiers trained on **generated** data achieve **high** effective robustness.
- Training on real and generated data strikes a **good balance** between accuracy and effective robustness.

FID among datasets



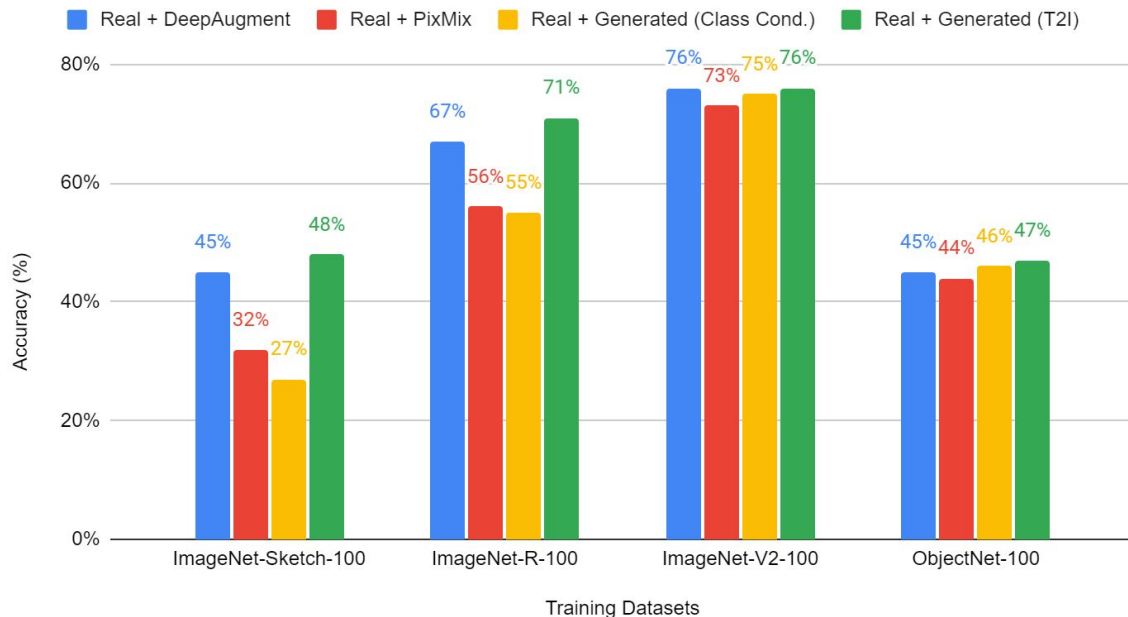
- We report average FID scores per class label between the real/generated data with the NDS datasets.

- **Larger improvements on ImageNet-Sketch/R datasets can be explained by the observed distribution gap.**

- ImageNet-V2 and ImageNet share the same data source - Flickr30K.

Comparison with Data Augmentations

Comparison among data generation strategies



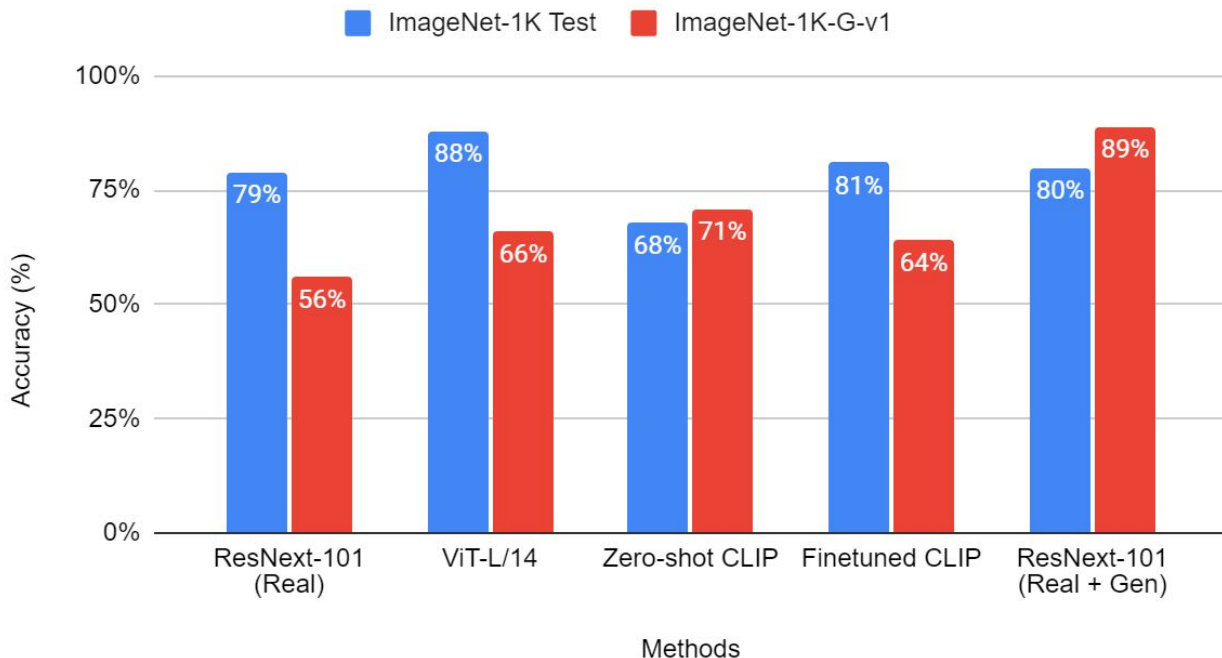
- Data Augmentation with generated data from text to image generative models outperforms popular data augmentation techniques such as **DeepAugment** and **PixMix**.
- It is also better than using a class conditional generative model trained on ImageNet.

ImageNet-G

- Generated Dataset constructed by querying Stable Diffusion.
- We studied its usefulness as a training data in the previous slides.
- A generative model can also be utilized for benchmarking the progress of the classifiers to novel variations in the data synthesized by it.
- Having Generated dataset is very different from a real dataset since real datasets are often stationary, and hard-to-critique is data generation process is hard to replicate.

ImageNet-G Evaluation

Benchmarking on ImageNet-G-v1



- Most of the classifiers undergo a performance reduction when evaluated on the generated data.
- The performance of the classifier trained on real + generated data (**89%**) shows that the current classifiers can do better on the generated data.

Ethics Statement

- **Privacy** concerns regarding the suitable use of the generated data since they are trained on the uncurated datasets, and have potential to memorize them.
- **Bias** amplifications since the generative models might fail to fit certain modes of distribution well.
- Since generative models can create images based on natural language descriptions, they can be prompted to generate **objectionable** content relatively easily.

Read the paper for more experiments on:

- Effect of training data and generated **data size** on accuracy and effective robustness.
- Effect of the choice of **conditioning** (single vs diverse templates), and **generation** strategies (label, image, label and image conditioned).
- Automatic and Human Evaluation of the generated data along the dimensions of **Consistency**, **Quality**, and **Diversity**.

Thank You!

Paper: <https://openreview.net/pdf?id=LjGqAFP6rA>

Code: <https://github.com/Hritikbansal/generative-robustness>

People: Hritik Bansal (@hbXNov), Aditya Grover (@adityagrover_)