Benchmarking Robustness to Text-Guided Corruptions

Mohammadreza Mofayezi and Yasamin Medghalchi

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Introduction

Towards evaluating the robustness of image classifiers to **text-guided corruptions**:

1. *Diffusion models* used to edit images to different domains
2. Others use synthetic or hand-picked data for benchmarking
3. No *manual labeling* needed with our method, allowing creation of large-scale benchmarks with less effort.
4. Observed that *convolutional models* are more robust than *transformer architectures*.
5. *Data augmentation* techniques can improve the performance on both the original data and the edited images.

Text-guided Robustness Benchmark

Our Method

1. *Null-text inversion* used to edit ImageNet with prompts.
2. Benchmark has edited images in *domains* like Drawing, Weather, Color, Texture, and Context.

ImageNet Hierarchy

3. ImageNet images divided into *9 subclasses* with specific prompts for each to generate meaningful images: Animal, Plant, Person, Vehicle, Furniture, Tool, Food, Structure, Landscape.

Super Class	Sub-class	Index
Organism	Animal	1
	Plant	2
	Person	3
Artifact	Vehicle	4
	Furniture	5
	Tool	6
	Food	7
Geological Formation	Structure	8
	Landscape	9

Prompt Hierarchy

4. Recent text-guided models struggle to apply all prompts to whole images.
5. To solve this, we introduce hand-engineered prompts for each subclass to make good edits.
6. Images won't convert to damaged or different image if prompts can't be performed. This is crucial for using the process without prompt engineering.

Experiments

Edited images fed to multiple image classifiers to determine their *sensitivity to prompts* and *robustness*.

1. **Experimental Setup:** Classes selected from each super class of ImageNet, 10 images/class, random prompt assigned based on engineering.
2. **Architecture affects robustness:** Swin-Transformer better on *original data*, but ConvNeXt, ResNeXt, and deep ResNet better on *edited images*.
3. **Data augmentation improves robustness:** *Style Transfer* and *AugMix* tested on ResNet-50 and improve accuracy on corrupted images.
4. **Domains affect robustness:** All domains reduce classifier accuracy, with *drawing* being the most difficult.