Deep Models Under Domain Shift: A Sketch-Based Study

Anonymous Submission
Institution Name Withheld
anonymous@example.com

Abstract

Deep learning models achieve remarkable performance on classification tasks when trained on large, labeled datasets from consistent distributions. However, they often fail to generalize under domain shifts, where training and test data distributions differ. This happens because models assume test data follows the same distribution as training data, which is rarely true in real-world scenarios. Addressing this challenge is crucial for applications demanding reliability and robustness, including forensics, security, design, and medical imaging. Despite advances through data manipulation, learning strategies, and representation learning, critical gaps remain. Sketch recognition exemplifies this challenge: sketches lack the color and texture cues of natural images, causing models trained on RGB images to suffer drastic performance drops, highlighting vulnerabilities under sketch-based distribution shifts.

Model	**ImageNet-1k** (↑)		**ImageNet-Sketch** (↑)	
	Top-1 (%)	**Top-5** (%)	**Top-1** (%)	**Top-5** (%)
AlexNet	56.55	79.09	10.74	22.15
VGG16	71.58	90.40	17.56	32.84
VGG19	72.39	90.88	17.96	33.12
ResNet18	69.76	89.08	20.25	37.31
ResNet50	76.14	92.87	24.12	41.42
ResNet101	77.37	93.56	27.02	45.30
EfficientNet-B0	77.67	93.59	25.13	43.05
EfficientNet-B7	73.92	91.57	29.42	**49.55**
ViT-B16	**81.07**	**95.32**	**29.44**	47.85

Table 1. Showcasing the vulnerabilities of the current state-of-the-art image classification. The networks, since heavily trained on color images, show a drastic reduction in performance when sketch images are utilized for their evaluation. Best values are highlighted.

We benchmark the performance sensitivity of widely used architectures—including ***AlexNet, VGG19, ResNet-50, EfficientNet-B7, and ViT-B16****—on* ***ImageNet-sketch dataset****, and propose cost-effective strategies to improve model robustness. We explore four directions: (1) Colorization of sketches using diffusion-based ControlNet pipelines, (2) Grayscale training and augmentation to reduce reliance on color cues, (3) Frequency-domain training via Discrete Fourier and Wavelet Transforms with low/high-pass filtering, and (4) Image complexity analysis based on edge density and compression ratio.*

Artificial colorization *of sketches for test-time adaptation is achieved using ControlNet with a Canny edge detector, and Stable Diffusion v1.5 as the generative backbone. The final FC layers of ImageNet-pretrained models were* ***fine-tuned on the ImageNet-200 training set using RGB and grayscale inputs****. Advanced models (EfficientNet-B7, ViT-B16) benefited most from RGB fine-tuning, while simpler CNNs (AlexNet, VGG19, and ResNet-50) show negligible gains or performance drops, suggesting they rely on edge- and shape-based features and treat synthetic color cues as noise. In contrast, modern high-capacity models exploit them effectively. Grayscale fine-tuning benefits ViT-B/16 (+9.71% top-1), but degrades EfficientNet-B7 and CNNs' performance, reflecting a mismatch between grayscale inputs and RGB-pretrained backbones, highlighting architecture-dependent effects of colorization.*

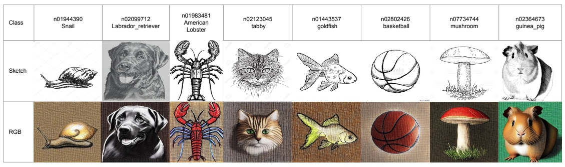

Figure 1. ImageNet sketches and their corresponding colored versions showcase how synthetic color integration changes the visual perception of images.

Training on grayscale images *improves sketch recognition but reduces accuracy on RGB datasets. However, models trained with* ***grayscale augmentation*** *with varying ratios of RGB and grayscale inputs achieve better out-of-distribution robustness and higher accuracy on large-scale validation sets, indicating that color sensitivity can be mitigated through targeted augmentation.* ***Frequency-domain low-pass training*** *was particularly effective: ResNet-50's top-5 accuracy on ImageNet-Sketch rose from 9̃% to over 42% (DFT) and 43% (DWT), highlighting deep networks' reliance on domain-specific cues.* ***Complexity analysis*** *also suggests a slight inverse correlation between image complexity and recognition accuracy, motivating further study into structural biases.*

Overall, our study demonstrates that while no single technique completely resolves domain shifts in vision models, targeted preprocessing and augmentation significantly mitigate performance degradation. These insights highlight promising directions for developing more generalizable and trustworthy AI systems under out-of-distribution scenarios.