Effective Data Augmentation With Diffusion Models

Brandon Trabucco [1], **Kyle Doherty** [2], **Max Gurinas** [3], **Ruslan Salakhutdinov** [1]
[1] Carnegie Mellon University, [2] MPG Ranch, [3] University Of Chicago, Laboratory Schools
brandon@btrabucco.com, rsalakhu@cs.cmu.edu

Abstract

Data augmentation is one of the most prevalent tools in deep learning, underpinning many recent advances, including those from classification, generative models, and representation learning. The standard approach to data augmentation combines simple transformations like rotations and flips to generate new images from existing ones. However, these new images lack diversity along key semantic axes present in the data. Current augmentations cannot alter the high-level semantic attributes, such as animal species present in a scene, to enhance the diversity of data. We address the lack of diversity in data augmentation with image-to-image transformations parameterized by pre-trained text-to-image diffusion models. Our method edits images to change their semantics using an off-the-shelf diffusion model, and generalizes to novel visual concepts from a few labelled examples. We evaluate our approach on few-shot image classification tasks, and on a real-world weed recognition task, and observe an improvement in accuracy in tested domains.

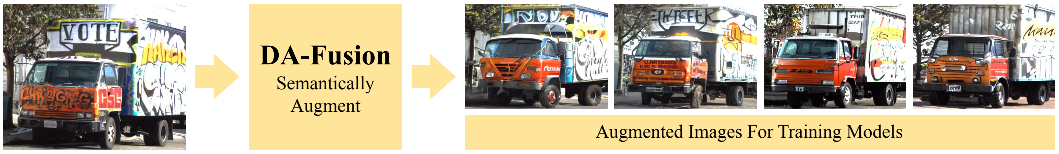

Figure 1: Real images (left) are semantically modified using a publicly available off-the-shelf Stable Diffusion checkpoint. Resulting synthetic images (right) are used for training downstream classification models.

1 Introduction

An omnipresent lesson in deep learning is the importance of internet-scale data, such as ImageNet (Deng et al., 2009), JFT (Sun et al., 2017), OpenImages (Kuznetsova et al., 2018), and LAION-5B (Schuhmann et al., 2022), which are driving advances in Foundation Models (Bommasani et al., 2021) for image generation. These models use large deep neural networks (Rombach et al., 2022) to synthesize photo-realistic images for a rich landscape of prompts. The advent of photo-realism in large generative models is driving interest in using synthetic images to augment visual recognition datasets (Azizi et al., 2023). These generative models promise to unlock diverse and large-scale image datasets from just a handful of real images without the usual labelling cost.

Standard data augmentations aim to diversify images by composing randomly parameterized image transformations (Antoniou et al., 2017; Perez and Wang, 2017; Shorten and Khoshgoftaar, 2019; Zhao et al., 2020). Transformations including flips and rotations are chosen that respect basic invariances present in the data, such as horizontal reflection symmetry for a coffee mug. Basic image transformations are thoroughly explored in the existing data augmentation literature, and produce models that are robust to color and geometry transformations. However, models for recognizing coffee mugs should also be sensitive to subtle details of visual appearance like the brand of mug; yet, basic transformations do not produce novel structural elements, textures, or changes in perspective. On the other hand, large pretrained generative models have become exceptionally sensitive to subtle visual details, able to generate uniquely designed mugs from a single example (Gal et al., 2022).

Our key insight is that large pretrained generative models *complement the weaknesses* of standard data augmentations, while *retaining the strengths*: universality, controllability, and performance. We propose a flexible data augmentation strategy that generates variations of real images using text-to-image diffusion models (DA-Fusion). Our method adapts the diffusion model to new domains by