Diffusion-based Dataset Augmentation for Downstream Crop Segmentation

Alex Senden[1], Masoomeh Gomroki[2], Robert H. Gulden[2], Christopher J. Henry[1]

[1]Department of Computer Science, University of Manitoba, Winnipeg, Manitoba, Canada

[2]Department of Plant Science, University of Manitoba, Winnipeg, Manitoba, Canada

Corresponding Author: sendena@myumanitoba.ca

Abstract: Semantic segmentation, the task of classifying each pixel in an image according to the object it represents, has become a common approach for automated weed detection. Implemented as a deep neural network, advances in semantic segmentation have led to substantial growth in model size and, consequently, in the required training data. Although unlabelled crop field images are plentiful, creating segmentation masks requires domain knowledge and a significant amount of human labour. We propose a generative approach to alleviate this data labelling bottleneck by training a diffusion model to act as a synthetic data generator, which can then be used to augment the training dataset. This approach requires a small amount of labelled data to bootstrap the process. A Vision-Language Model paired with weak contextual information (such as the imaged crop) creates short captions for the raw training images. The image structure is automatically extracted from annotated images in the form of scribble annotations. This 4-tuple of image, caption, optional scribble annotation, and optional segmentation is used to train a modified version of the FLUX.1-dev image diffusion model. This model learns to incorporate a text caption and scribble-structure to generate realistic crop field imagery complete with segmentation masks.