[BIOINFO 2024]

LadybirdMNIST: ladybird pattern dataset construction via PDE-based pattern generation and pattern classification

Seongmi Woo[1] and Daewon Lee[1,2,*]

[1] *Department of Applied Art and Technology, Chung-Ang University*
[2] *School of Art and Technology, Chung-Ang University*
**Corresponding author: dwlee@cau.ac.kr*

The diverse species of ladybirds and their distinctive patterns make them valuable models for ecological studies and bioinformatics applications. This study aims to generate a dataset of ladybirds using the LPF framework and classify their patterns to construct a high-quality dataset, **LadybirdMNIST**. The LPF framework was used to generate over 400,000 images of the ladybird patterns. To categorize the generated pattern images into distinct classes, we have defined similarity metrics, including mean squared error (MSE), color proportion, and learned perceptual image patch similarity (LPIPS). Moreover, a vector quantized variational autoencoder (VQ-VAE) was trained to extract latent vectors from the ladybird pattern images. These latent vectors were then visualized and clustered using uniform manifold approximation and projection (UMAP), and pattern analysis was conducted based on the similarity values. Then we define 10 species, including both real existing species and synthetically defined species. The current selection process involves choosing 1,000 images for each of the 10 species, for a total of 10,000 images. The **LadybirdMNIST** dataset includes not only pattern images but also PDE parameters. This dataset can be utilized similarly to the MNIST dataset and shows potential for application in physics-informed neural networks (PINNs).