Supplementary Materials for

FungiTastic: A multi-modal dataset and benchmark for image categorization

1 *The authors accept full responsibility for any rights violations related to this dataset. The dataset is*
2 *published under the GNU General Public License (GPL3) for non-commercial use only.*

1 Online resources

4 The complexity and size of the dataset require hosting the dataset in multiple spaces. Since the *Kaggle*
5 *dataset* is allowing to store datasets just up to around 140GB we provide images for the vanilla
6 FungiTastic dataset in smaller resolution. The full-size images, as well as the Croissant metadata, are
7 available for download through dataset documentation.

8 Apart from the dataset, we also provide a wide variety of assets. Notably, we provide:

9 • *FGVC*, a Pytorch-based framework designed for training deep learning classification models,
10 hyperparameter fine-tuning, Weights&Biases logging, ONNX model export and HuggingFace
11 deployment.

12 • *Sample notebooks* that allows loading pre-trained models, prediction and making submission on
13 Kaggle benchmarks.

14 • All baseline *pre-trained models* available through the HuggingFace model hub allowing a straight-
15 forward integration and usability.

16 • A comprehensive documentation with tables data samples and training strategies.
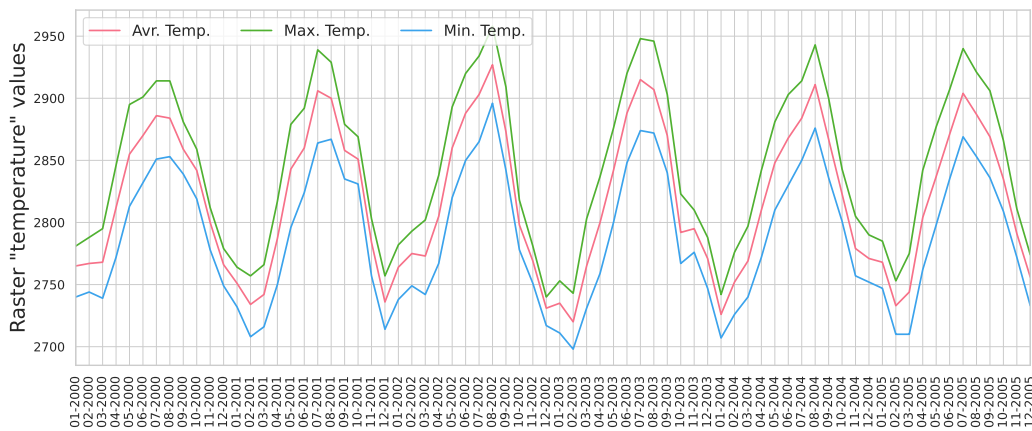
2 Supporting Figures

Figure 1: **Sample of available temperature data**. First 5 years out of 20 available years.

Figure 2: **Additional samples of fruiting body part segmentation**.