**ON THE IMPORTANCE OF BACKBONE, PRETRAINING AND HYPERPARAMETER SELECTION FOR HIERARCHICAL FINE-GRAINED IMAGE RECOGNITION**

1*Augusto Christian Surya* (蘇立光)*,*2*Edwin Arkel Rios* (冉恩達), 2*Bo-Cheng Lai* (賴伯承),1*Min-Chun Hu* (胡敏君)

1 Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan

2 Department of Electronics Engineering, National Yang Ming Chiao Tung University, Hsinchu, Taiwan

E-mail: ss110000162@gapp.nthu.edu.tw, edwinarkel.rios@gmail.com, bclai@nycu.edu.tw, anitahu@cs.nthu.edu.tw

*Abstract*

Fine-grained image recognition (FGIR) is a task that requires models to distinguish visually similar subcategories within a broader class. FGIR tasks are naturally organized into hierarchical structures where coarse-level groupings provide valuable semantic context for fine-level classification. Despite this, most prior work in hierarchical FGIR evaluates only a limited set of backbone architectures, leaving the influence of backbone and pretraining largely underexplored. This gap is critical, as the backbone determines the quality of feature representations used across multiple classification heads in hierarchical settings. In this work, we systematically study 21 pretrained models, using convolutional and transformer-based backbones, evaluated under both fine-tuned and frozen backbone settings. We also evaluate 18 transferability metrics to examine their ability to predict performance. Our results show that backbone and hyperparameter selection heavily influence FGIR performance, with Discriminative models such as MoCov3 and SwaV on Resnet being a solid choice due to having desirable qualities, high mean accuracy and low standard deviation. Most of the metrics weakly correlates or inconsistent with transfer accuracy across different settings, highlighting an open challenge for future research.

Keywords - Fine-Grained Object Categorization, Hierarchical Classification, IPPR, CVGIP 2025.