ENERGY VARIATION IN TRAINING MODERN COMPUTER VISION ARCHITECTURES

[1]DAVID CORTES, [2]CARLOS JUIZ, [3]BELEN BERMEJO

[1,2,3]Computer Science Department, University of the Balearic Islands, Palma, Spain
E-mail: [1]d.cortes@uib.es, [2]cjuiz@uib.es, [3]belen.bermejo@uib.es

Abstract - This study examines the energy consumption variation that occurs when training seven modern computer vision architectures — MobileNetV3-Small, MobileNetV3-Large, EfficientNet-B0, EfficientNet-B1, ViT-B/32, ConvNeXt-Tiny, and ViT-B/16 — for ImageNet-1k classification. All models are trained under identical baseline conditions: 40 epochs, a homogeneous subset of 10,000 images (ImageNet-10k), and two NVIDIA Tesla P100 GPUs at the Center for Bioinformatics and Computational Biology of Colombia (BIOS). The study finds a strong Pearson correlation (r = 0.85) between the number of floating-point operations (GFLOPs) and total energy consumption in kilowatt-hours, confirming that computational complexity serves as a robust predictor of training energy cost. Total consumption ranges from 0.137 kWh (MobileNetV3-Small) to 3.026 kWh (ViT-B/16), representing a 22-fold variation across architectures. EfficientNet-B0 achieves the best precision-to-energy ratio, attaining 97.15% Top-5 validation accuracy at only 0.457 kWh, while Vision Transformer variants exhibit the lowest classification performance under data-limited conditions, consistent with their known dependency on large-scale pretraining datasets. These findings underscore the importance of incorporating energy efficiency as an explicit criterion in architecture selection, particularly within shared high-performance computing environments.

Keywords - energy consumption, deep learning, image classification, ImageNet, efficient architectures, FLOPs, GPU computing, Green AI, Vision Transformer, Efficient Net.