Wavelet Resampling

Author: Zhizheng Wang

	Resnet50	VGG	GoogleNet	AlexNet
Original	Training Loss: 0.3478 Training Accuracy: 0.8865	Training Loss: 1.4185 Training Accuracy: 0.4525	Training Loss: 0.1304 Training Accuracy: 0.9576	Training Loss: 1.889 Training Accuracy: 0.1816
Upsample	Training Loss: 0.4076 Training Accuracy: 0.8608	Training Loss: 1.4080 Training Accuracy: 0.4694	Training Loss: 0.2000 Training Accuracy: 0.9348	Training Loss: 0.3478 Training Accuracy: 0.1414
Downsample	Training Loss: 0.7420 Training Accuracy: 0.7420	Training Loss: 1.9524 Training Accuracy: 0.1452	Training Loss: 0.4455 Training Accuracy: 0.8548	Training Loss: 1.9472 Training Accuracy: 0.1292

Upsampling:
Upsampling effectively increased the representation of underrepresented classes (e.g., Swimming and Karate). However, the synthetic images generated may have needed to have introduced more diversity or realistic features. Models such as ResNet50 and GoogleNet, which are sensitive to overfitting, may have degraded performance.

Downsampling:
Downsampling reduced image resolution to 128 × 128 and randomly removed images from overrepresented classes. This process may have eliminated critical features and details necessary for classification, particularly for models like ResNet50 and GoogleNet that rely on capturing fine-grained information.
VGG and AlexNet, being older architectures with less robust feature extraction capabilities, suffered significantly when trained on the downsampled dataset.

Wavelet Upsample:
By combining the horizontal (cH), vertical (cV), and diagonal (cD) wavelet coefficients, the wavelet-based method effectively highlights object details, making the synthetic images representative of the original class. This approach emphasizes high-frequency information, such as edges and textures, which are crucial for visual classification tasks. As a result, the synthetic images retain the distinguishing features of each class, ensuring that they contribute meaningfully to model training. However, the models' performance did not improve significantly because the synthetic images, while preserving key features, needed more diversity to introduce new patterns or variability. Additionally, high-performing models like ResNet50 and GoogleNet may already leverage intricate feature extraction capabilities, meaning the added wavelet-based synthetic data contributed redundant information rather than enhancing the overall learning process. This highlights the importance of generating diverse and realistic synthetic samples for more impactful model improvements.

Wavelet Downsample:
The approximation component (cA) from the wavelet transform is utilized to capture low-frequency information, effectively retaining the general structure of images while discarding high-frequency details. This approach is essential for reducing image size while preserving meaningful features. Normalizing cA to a range of 0 – 255 ensures consistency in pixel intensity, making the downsampled images comparable across all classes. The resizing step to 128 x 128 guarantees uniformity in the input size, optimizing the dataset for model training. However, models like ResNet50 and GoogleNet, which rely on high-resolution images to extract intricate details, may experience a decline in performance due to the loss of critical high-frequency information during the downsampling process. While efficient, this trade-off highlights the balance between computational simplicity and feature retention in image classification tasks.

Work Cited:

Yaji, G. S., Sarkar, S., Manikantan, K., & Ramachandran, S. (2012). Dwt feature extraction-based face recognition using intensity mapped unsharp masking and laplacian of gaussian filtering with scalar multiplier. *Procedia Technology*, *6*, 475–484. https://doi.org/10.1016/j.protcy.2012.10.057

Mallat, S., & Hwang, W. L. (1992). Singularity detection and processing with wavelets. *IEEE Transactions on Information Theory*, *38*(2), 617–643. https://doi.org/10.1109/18.119727