Project Report on Multi-Label Image Classification Enhancements

Overview

This report outlines the methodologies implemented to enhance multi-label image classification models. The project commenced with EfficientNet and progressively integrated advanced techniques using Inception v3, including image augmentation, weighted sampling, dynamic thresholding, and selective layer freezing to address the class imbalance and improve model performance.

Initial Experimentation with EfficientNet

The project began with EfficientNet as the foundational model, training all layers to adapt fully to the specific dataset. This approach achieved a modest macro-averaged F1 score of **0.54** on local testing, establishing a baseline for further optimization.

Transition to Inception v3

Seeking a substantial improvement in model performance, I transitioned from EfficientNet to the Inception v3 architecture. This strategic shift proved pivotal, as Inception v3's advanced feature extraction capabilities significantly enhanced classification outcomes. The model was trained over **30 epochs** with a learning rate of **1e-4** using the **AdamW optimizer**, ensuring efficient convergence and robust weight updates. For evaluation purposes, a **static threshold of 0.5** was initially employed to simplify the assessment process.

This transition resulted in a remarkable performance boost. On the leaderboard, the Inception v3 model achieved a score of **0.66 on the professor's testing set**, and a local F1 score of **0.69**—a substantial improvement over the initial EfficientNet baseline of **0.54**. Most impressively, this configuration secured the **highest score in the professor's test set**. This underscores the effectiveness of Inception v3 in leveraging complex feature hierarchies and handling multi-label classification challenges, ultimately demonstrating its superiority in this specific application context.

Advanced Techniques: Augmentation and Weighted Sampling

Building on Inception v3's performance, advanced techniques were implemented to manage class imbalance:

- **Image Augmentation:** Focused on underrepresented classes ("class9", "class15", "class3", "class5", "class12", "class7", "class16"), up to **15** augmented images were generated per original image, capped at **100,000** images per class. Augmentation techniques included:
    - **Random Flips:** Horizontal and vertical.
    - **Random Rotations:** Within ±20 degrees.
    - **Color Jitter:** Adjustments to brightness, contrast, saturation, and hue.

- These strategies increased data diversity and mitigated overfitting by preventing excessive duplication.
- **Weighted Sampling:** To ensure balanced class representation during training, weighted sampling was employed. Minority classes were assigned higher sampling weights based on their inverse frequency, ensuring equitable exposure across all classes and reducing bias toward majority classes.
- **Selective Layer Freezing:** The Inception v3 model was fine-tuned by freezing all layers except the Mixed_6 and Mixed_7 blocks. This allowed higher-level layers to adapt to the dataset's specific features while retaining the pre-trained model's foundational capabilities.
- **Dynamic Threshold Optimization:** Class-specific thresholds were determined by analyzing precision-recall curves to maximize the F1 score. This approach refined binary predictions, leading to an impressive local macro-averaged F1 score of **0.85**.

Despite these enhancements, the model's performance on the professor's external test set was limited, with the F1 score dropping to **0.47**, indicating challenges in generalization.

Simplification by Removing Optimal Thresholds

To improve generalization, the approach was simplified by removing class-specific optimal thresholds and reverting to a universal threshold of **0.5**. This modification enhanced the model's robustness, achieving a local F1 macro score of **0.79** and elevating the test set score on the professor's testing set to **0.56**. This phase highlighted the balance between specialized optimization for known data and the need for broader generalization to unseen data.

Potential and Unexplored Strategies

Lastly, I considered exploring a fully unfrozen Inception v3 model paired with extensive image augmentation, which might outperform all previous configurations. Additionally, integrating Vision Transformers (ViT) was also on my agenda. These strategies, while untested due to time constraints, represent promising future directions that could potentially exceed the 0.66 score achieved on the leaderboard.

Conclusion

Throughout the project, incremental enhancements—from initial model adjustments with EfficientNet to sophisticated techniques with Inception v3—demonstrated significant improvements in handling class imbalance and boosting local performance. However, the challenge of generalizing to external datasets underscored the importance of balancing model complexity with generalization capabilities. The insights gained provide a valuable foundation for future endeavors in image classification and broader machine learning applications.