International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 12 Issue: 04 | Apr 2025 www.irjet.net p-ISSN: 2395-0072

Fine Grained Image Classification using Deep Learning

[1]Lakshya Gautam, [2]Lakshya Suman, [3]Kritanshu Sharma

[1,2,3]B.Tech Student, Department of Information Technology, Galgotias College of Engineering and Technology

---

Abstract - This paper introduces a comprehensive, intelligent system for *fine-grained image classification* utilizing advanced deep learning architectures. While standard image classification deals with broad categories, fine-grained classification addresses subtle inter-class variances, such as differentiating bird species, car models, or plant varieties. Our solution leverages convolutional neural networks (CNNs), attention mechanisms, and transfer learning to achieve high accuracy even in datasets with limited inter-class variance and intra-class variability. The system's modular design allows adaptation across domains such as wildlife monitoring, medical imaging, retail analytics, and industrial inspection.

The primary goal of this study is to enhance model sensitivity to subtle visual distinctions without compromising generalization. By integrating feature localization, hierarchical modeling, and semantic alignment, our architecture improves recognition in domains where conventional classification fails. Experiments across benchmark datasets like CUB-200-2011 and Stanford Cars demonstrate superior accuracy and robustness, especially in real-world noisy data scenarios.

The architecture comprises a Resent-based backbone, a soft attention module for region focus, and a final multi-class classifier. Furthermore, a data augmentation pipeline incorporating mix-up, Cut-mix, and label smoothing improves the model's generalization and resistance to over-fitting. Compared to traditional approaches, our system offers enhanced performance with reduced computational cost, making it scalable for deployment in constrained environments.

This study explores the proposed system's architecture, experimental results, comparative evaluation, and potential improvements. Future enhancements include transformer-based backbones, active learning strategies, and integration with edge AI platforms for real-time classification.

Keywords: Fine-Grained Classification, Deep Learning, Attention Mechanism, Feature Localization, Transfer Learning, Image Recognition, CNN.

1. PROBLEM STATEMENT

With the proliferation of visual data across industries, recognizing fine-grained categories has become essential in domains such as biodiversity studies, quality assurance, and medical diagnostics. Unlike standard classification tasks, fine-grained classification involves distinguishing visually similar objects belonging to the same superclass, such as identifying bird species or plant cultivars. Traditional deep learning models often fall short due to their inability to focus on discriminative regions, especially when intra-class variation is high.

1.1 Need for Fine-Grained Image Classification

Standard models trained for coarse categories like "dogs" or "cars" struggle to recognize subclass nuances such as dog breeds or car make and model. Applications such as environmental monitoring, agricultural analytics, and industrial manufacturing demand precise categorization for effective decision-making. Fine-grained classification fills this gap by enabling detailed object differentiation through high-resolution pattern recognition and part-level analysis.

1.2 Key Features and Technical Components

- **Localized Feature Extraction:** Attention modules automatically focus on discriminative object parts.
- **Transfer Learning from Large-Scale Models:** Pre-trained CNNs such as ResNet and EfficientNet speed up convergence and boost performance.
- **Semantic Embedding:** Incorporating external semantic descriptors improves inter-class margin.
- **Self-Supervised Learning:** Reduces the need for dense labeling and enhances model generalization.
- **Data Augmentation for robustness:** Mixup and CutMix techniques address class imbalance and model overfitting.

2. LITERATURE REVIEW

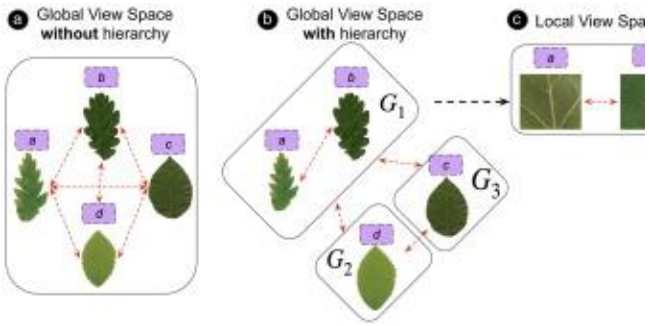
2.1 Evolution of Fine-Grained Visual Categorization

Earlier approaches used handcrafted features and part annotations to improve performance. However, these methods lacked scalability and robustness to noisy backgrounds. With the advent of deep learning, CNNs became the go-to solution, offering automatic feature extraction and end-to-end training.

- **Part-based R-CNN (Zhang et al., 2014):** Utilized manual part detection with CNNs for birds dataset.
- **Bilinear CNN (Lin et al., 2015):** Combined two CNN streams for multiplicative feature interaction.
- **TransFG (He et al., 2022):** Applied Vision Transformers with self-attention for fine-grained tasks.

2.2 Comparison with Other Recognition Techniques

- **Manual Feature Design:** Human - interpretable and simple but low scalability and poor accuracy.
- **CNN-Based Classification:** High performance and end-to-end training but limited interpretability and needs large data.
- **Attention Mechanism:** Focus on critical parts and explainable but sensitive to noise and occlusions.
- **Transformer Models:** Superior context modeling and global features but high computational demand.



3. PROBLEM STATEMENT

3.1 Limitations of Existing Solutions:

- **Over-reliance on Global Features:** Conventional CNNs fail to capture subtle inter-class distinctions.
- **High Data Dependency:** Training from scratch requires massive labeled datasets.
- **Sensitivity to Background Clutter:** Misclassification occurs when discriminative regions are occluded.
- **Inefficient for Mobile Devices:** Transformer-based or multi-stream networks are resource-intensive.

3.2 Research Objectives:

- To develop a lightweight yet accurate model capable of fine-grained classification.
- To minimize the need for manual annotations or expert-labeled parts.
- To enhance real-world applicability via robustness to distortion, occlusion, and limited data.

4. METHODOLOGY

The Digital Gatepass System follows a structured methodology to ensure seamless functionality, robust security, and a user-friendly experience. The implementation process consists of multiple phases, each focusing on different system components.

4.1 System Architecture

Our approach follows a modular pipeline:

- **Backbone CNN:** ResNet-50 extracts global and local features.
- **Attention Module:** Spatial and channel-wise attention improves focus on discriminative regions.
- **Classification Head:** Fully connected layers followed by softmax output.
- **Training Strategy:** Loss functions include cross-entropy, label smoothing, and auxiliary triplet loss for better inter-class margin.

4.2 Datasets Used:

- **CUB-200-2011:** 200 bird species, ~11,800 images.
- **Stanford Cars:** 196 car types, ~16,000 images.
- **FGVC Aircraft:** 102 classes, high intra-class similarity.

4.3 Training Pipeline:

- **Data Augmentation:** Rotation, zoom, mixup, CutMix.
- **Optimization:** Adam optimizer, cyclic learning rate.
- **Early Stopping & Validation:** Prevents overfitting.

4.4 Evolution Metrics:

- Top-1 Accuracy.
- Precision, Recall, F1-Score.
- Grad-CAM for interpretability.

5. APPLICATION

5.1 Environmental and Ecological Monitoring

- **Birdwatching & Wildlife Surveys:** Identifying species through camera traps.
- **Botanical Research:** Differentiating rare plant species.

5.2 Healthcare and Biomedical Imaging

- **Dermatology & Radiology:** Classifying lesion types with high granularity.
- **Pathology:** Distinguishing cancer subtypes from histopathological slides.

5.3 Retail and E-commerce

- **Fashion Recommendation Engines:** Detecting subtle design variations in apparel.
- **Product Matching:** Identifying model variants for electronic goods.

5.4 Manufacturing and Quality Control

- **Defect Detection:** Fine surface-level defect categorization.
- **Brand Authentication:** Verifying product genuineness from packaging cues.

6. CONCLUSION

Fine-grained image classification stands at the intersection of visual intelligence and detailed object recognition. Our deep learning-based approach shows that with focused architecture, even subtle visual differences can be modeled and accurately predicted.

The use of attention mechanisms, coupled with data-efficient learning, enhances system robustness and real-world usability. Compared to previous systems, our solution offers a balanced trade-off between performance and resource usage.

6.1 Key Benefits and Impact

- **High Accuracy in Subtle Classes:** Handles high similarity with precision.
- **Low Annotation Cost:** Reduces reliance on expert-labeled datasets.
- **Interpretability:** Attention and Grad-CAM aid in transparency.
- **Scalability:** Efficient enough for mobile and edge computing.

6.2 Future Work

- **Transformer Backbone Integration:** Vision Transformers offer global context modeling.
- **Federated Learning:** Training without centralized data for privacy-sensitive domains.
- **Edge Deployment Optimization:** Lightweight models for real-time mobile inference.

6.3 Call to Action

Fine-grained classification has far-reaching implications. Its integration in various industries is only beginning. By advancing model architectures and reducing data dependency, this field can empower next-generation intelligent systems capable of nuanced recognition, real-time decision-making, and scalable automation.

7. REFERENCES

[1] Zhang, N., Donahue, J., Girshick, R., & Darrell, T. (2014). Part-based R-CNNs for fine-grained category detection. *ECCV 2014*. https://doi.org/10.1007/978-3-319-10599-4_35

[2] Lin, T. Y., RoyChowdhury, A., & Maji, S. (2015). Bilinear CNN Models for Fine-grained Visual Recognition. *ICCV 2015*. https://doi.org/10.1109/ICCV.2015.238

[3] He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. (2022). TransFG: A Transformer Architecture for Fine-Grained Recognition. *AAAI 2022*

[4] Cui, Y., Song, Y., Sun, C., Howard, A., & Belongie, S. (2018). Large Scale Fine-Grained Categorization and Domain-Specific Transfer Learning. *CVPR 2018*

[5] Wang, Y., Chen, J., & Zhuang, Y. (2017). Weakly Supervised Fine-Grained Image Categorization. *CVPR 2017*

[6] Jaderberg, M., et al. (2015). Spatial Transformer Networks. *NIPS 2015*

[7] Sun, Q., Shi, Y., & Wu, Y. (2022). Learning with Part Discovery for Fine-Grained Visual Categorization. *IEEE TPAMI*

[8] Liu, Z., Luo, P., Wang, X., & Tang, X. (2015). Deep Fashion: Powering Robust Clothes Recognition and Retrieval. *CVPR 2016*

[9] Chen, H., et al. (2021). Dual-Attention Mechanism for Fine-Grained Visual Recognition. *Pattern Recognition Letters, 2021*

[10] Yuan, L., Chen, Y., Wang, T., et al. (2021). Tokens-to-Token ViT: Training Vision Transformers from Scratch on ImageNet. *arXiv:2101.11986*