Journal of Information Systems Engineering and Management
2025, 10(42s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Hybrid Deep Learning Models in Image Classification: Integrating CNNs with Attention, Capsule Networks, and Graph Neural Networks

[1]Amitabha Mandal, [*2]Biswajit Mondal, [3]Prabal Kumar Sahu, [4]Chandan Das, [5]Nilkamal Bhunia, [6]Biswajit Saha,
[1]Asst. Professor, Dept. Computer Science & Engineering, Dr. B. C. Roy Engineering College, Durgapur, amitabha.mandal@bcrec.ac.in
[2]Asst. Professor, Dept. Computer Science & Engineering, Dr. B. C. Roy Engineering College, Durgapur, biswajit.mondal@bcrec.ac.in(Corresponding Author)
[3]Asst. Professor, Dept. Information Technology, Dr. B. C. Roy Engineering College, Durgapur, prabal.sahu@bcrec.ac.in
[4]Asst. Professor, Dept. Computer Science & Engineering, Dr. B. C. Roy Engineering College, Durgapur, Chandan.das@bcrec.ac.in
[5]Asst. Professor, Dept. Electronics & Communication Engineering Dr. B. C. Roy Engineering College, Durgapur, nilkamal.bhunia@bcrec.ac.in
[6]Asst. Professor, Dept. Computer Science & Engineering (AIML) Dr. B. C. Roy Engineering College, Durgapur, biswajit.saha@bcrec.ac.in

ARTICLE INFO

Received: 26 Dec 2024

Revised: 10 Feb 2025

Accepted: 18 Feb 2025

ABSTRACT

Image classification has been transformed by convolutional neural networks (CNNs), yet single-architecture solutions are increasingly reaching performance plateaus on complex, fine-grained, and cross-domain tasks. A new research frontier therefore explores hybrid deep learning models that fuse complementary architectural paradigms—attention mechanisms, Capsule Networks, recurrent/transformer layers, and graph neural networks (GNNs)—with CNN backbones to capture richer spatial hierarchies, relational cues, and long-range dependencies.

This review synthesizes 2020-2025 literature on such hybrids, with a focus on models that (i) insert channel- or self-attention modules into CNN feature pipelines; (ii) replace late fully connected layers with Capsule Networks to exploit part–whole relationships; (iii) append GNN layers to reason over pixel-region graphs; and (iv) orchestrate multi-branch designs combining several of the above. We analyse 60+ primary studies, benchmarking gains on ImageNet, CIFAR, hyperspectral, medical, and remote-sensing datasets. Hybrid schemes commonly deliver 2-8 % accuracy improvements and enhanced robustness to occlusion and viewpoint change.

Nevertheless, they incur higher FLOPs, memory footprints, and hyper-parameter complexity. A critical contribution of this review is a taxonomy (Figure 2) and a consolidated performance table (Table 1) that links architectural choices to empirical gains. We discuss optimisation strategies (knowledge distillation, sparse attention, lightweight graph convolutions) and examine open challenges: cross-domain generalisation, explainability, and sustainable energy budgets. Finally, we outline future directions—neuro-symbolic fusion, federated hybrid learning, and automated architecture search—to guide the next wave of research.

Keywords: Hybrid models; CNN; attention mechanisms; Capsule Networks; image classification; graph neural networks

1 INTRODUCTION

Convolutional neural networks have underpinned virtually every breakthrough in image understanding during the past decade. Hierarchical convolutional filters excel at local pattern extraction, yet they struggle with two recurring problems: (i) loss of global context—distant pixels seldom influence each other through limited receptive fields—and (ii) equivariance rather than *equivalence* to part–whole relationships, leading to brittle predictions under rotation, occlusion, or object re-arrangement. Attention mechanisms, Capsule Networks, recurrent transformers, and graph neural networks each address one dimension of this shortfall. Attention explicitly re-weights spatial or channel information, Capsule layers encode pose-aware vectors, transformers model long-range token