

Robust Image Classification Using Convolutional Neural Networks

Abstract

Convolutional Neural Networks (CNNs) have demonstrated strong performance in image classification tasks; however, their susceptibility to noise, perturbations, and distributional shifts poses challenges for real-world deployment. This project investigates robustness-oriented training strategies for CNN-based image classification using standard vision benchmarks. A baseline CNN model was trained and evaluated under clean data conditions and subsequently assessed under input perturbations such as noise, transformations, and data augmentations. To improve robustness, techniques including data augmentation, regularization, and architectural adjustments were incorporated into the training pipeline. Experimental results show that robustness-aware models maintain more stable performance under perturbed inputs compared to baseline models, with only a marginal reduction in clean-data accuracy. These findings demonstrate that robustness-focused training can significantly enhance model reliability and generalization. The project highlights the importance of evaluating vision models beyond standard accuracy metrics and contributes toward building dependable deep learning systems for practical computer vision applications.