Title: MessIm4: A Vision Transformer-Based Framework for Intelligent Classification of Unwanted Images in Personal Galleries

Author names: Mst. Taposi Rabeya, Shohanur Rahman, Susmoy Biswas, Shantanu Kundu, Abu Kowshir Bitto & Md. Hassan Imam Bijoy

Abstract: In the digital era, image galleries are overwhelmed with a plethora of irrelevant, low-quality, or unintentional captures, frequently neglected and rarely curated. This paper presents MessIm4, a novel and purpose-built dataset that categorizes such messy images into four distinct classes: Blurred Images, No Object Images, Normal Images, and Scanned Messy Documents. A robust deep learning-based classification pipeline is proposed to automatically detect and label these unwanted images, thereby alleviating the manual burden of gallery curation. The framework benchmarks four state-of-the-art models, baseline CNN, MobileNetV2, ResNet50, and Vision Transformer ViT B/16, across rigorous evaluation metrics, including accuracy, precision, recall, and F1-score. The ViT B/16 model exhibits superior performance, achieving an average accuracy of 97.31%, surpassing traditional CNNs and transfer learning counterparts such as MobileNetV2 at 93.60% and ResNet50 at 92.35%. Experimental results underscore the model's proficiency in discerning subtle inter-class differences, particularly in ambiguous scenarios such as no-object versus normal images. This work establishes a pioneering dataset and benchmarking protocol, laying a foundational framework for intelligent photo management systems capable of interpreting content quality and aligning with user intent.

https://link.springer.com/chapter/10.1007/978-3-032-15346-3_42