Integrating Deep and Handcrafted Features for Enhanced Remote Sensing Image Classification

Vian Abdulmajeed Ahmed
MARS Research Lab, LR17ES05
ISITCom, University of Sousse
Sousse, Tunisia
orcid.org/0009-0002-5924-6139

Khaled Jouini
MARS Research Lab, LR17ES05
ISITCom, University of Sousse
Sousse, Tunisia
orcid.org/0000-0001-5049-4238

Amel Tuama
Computer Engineering Techniques Department
Northern Technical University
Mosul, Iraq
orcid.org/0000-0002-3802-9074

Ouajdi Korbaa
MARS Research Lab, LR17ES05
ISITCom, University of Sousse
Sousse, Tunisia
orcid.org/0000-0003-4462-1805

***Abstract*—Satellite imagery supports critical applications such as land cover mapping, environmental monitoring, disaster assessment, and urban planning. Despite significant advancements, challenges in analyzing satellite imagery persist, primarily due to data variability, atmospheric conditions, and complex land cover patterns. Traditional handcrafted descriptors like Scale-Invariant Feature Transform (SIFT) and encoding techniques such as Bag-of-Visual-Words (BoVW) are effective but often fall short in capturing global context and spatial relationships due to their inherent local nature. The advent of deep learning (DL), propelled by ample data and computational resources, has markedly improved satellite image analysis. However, the reliance on extensive annotated data constrains the wider applicability of DL methods.**

This study harnesses the strengths of both deep and handcrafted features to enhance the classification accuracy of remote sensing images. Specifically, we synergize SIFT descriptors with pretrained MobileNetV2 and VGG16 deep features. While SIFT excels in capturing local features essential for identifying specific image characteristics, pretrained DL models provide enriched representations with global context, spatial relationships, and hierarchical features. This integration aims to overcome the individual limitations of each method, enabling the model to effectively handle perturbations, scale variations, and diverse landscapes. Extensive evaluations on the EuroSAT dataset demonstrate that our approach outperforms, not only SIFT, VGG16, and MobileNetV2 when used separately, but also surpasses state-of-the-art remote sensing image classification approaches. Another salient advantage of our approach is its robust applicability in scenarios with limited labeled data — a prevalent challenge in remote sensing image classification.

***Index Terms*—Remote Sensing, Land Cover Mapping, Features Fusion, Transfer Learning, Scale-Invariant Feature Transform (SIFT), Image Classification.**

I. INTRODUCTION

Satellite imagery offers a unique and comprehensive perspective of the Earth's surface, playing a crucial role in various applications such as land cover mapping, environmental monitoring, disaster response and urban planning [2]. The increasing volume of remote sensing images, fuelled by advancements in Earth observation, presents a critical yet challenging task: extracting valuable information from these extensive and intricate datasets [13]. *Scene and image classification*, which involves assigning predefined semantic classes to remote sensing images, lies at the heart of this challenge. Effective remote sensing image classification requires the capability to discern complex spatial patterns, while maintaining robustness against variations in scale, atmospheric conditions, and noise [23].

Early methods in remote sensing image classification heavily relied on manually designed descriptors, represented by the widely used *Scale-Invariant Feature Transform* (SIFT) [14]. SIFT and similar approaches excel in capturing distinctive local features, such as corners, edges, and textured regions. However, due to their inherent local nature and inability to directly represent the entirety of scenes, SIFT and similar approaches often struggle to adequately capture the complex spatial and contextual information present in remote sensing images [5].

The advent of deep learning, coupled with the increased availability of data, has brought about a paradigm shift in remote image classification. Deep learning models, trained on vast datasets containing millions of labeled images, enable feature extraction capabilities and levels of accuracy beyond the reach of traditional methods [5]. Nevertheless, the data-hungry nature of deep learning limits its scope of application. To mitigate the requirement for vast amounts of annotated training data, *transfer learning* has emerged as a strategic solution. Transfer learning involves harnessing knowledge gained from one task and applying it to a related but different task [20]. In the context of remote sensing image classification, transfer learning entails using a deep learning model pretrained on a large and general-purpose dataset like ImageNet, and fine-tuning it on a smaller task-specific remote sensing dataset. VGG16 [18] and MobileNetV2 [17] are two notable pretrained models frequently employed in this context.

This study leverages the complementary strengths of handcrafted and deep features to enhance the model's performance