Advanced Multi-Scale Enhanced U-Net for Efficient Land Cover Classification of Remote Sensing Images

Syed Zaheeruddin[1], K. Suganthi[2]
Research Scholar, School of SENSE, Vellore Institute of Technology, Chennai, India[1]
Associate Professor, Center for Cyber-Physical Systems, Vellore Institute of Technology, Chennai, India[2]

***Abstract*—For monitoring the environment, building cities, assessing crops, and studying the climate, it is very important to be able to accurately classify land cover from remote sensing images. Deep learning has made semantic segmentation work much better, especially with encoder-decoder designs like U-Net. Still, ordinary U-Net models have trouble capturing multi-scale contextual relationships, distinguishing narrow borders, and successfully emphasizing traits that are distinctive to an area. This work presents an Advanced Multi-Scale Enhanced U-Net (AMSE-U-Net) to address these difficulties. The AMSE-U-Net combines (i) multi-scale feature extraction, (ii) squeeze-and-excitation channel attention, and (iii) attention-gated skip connections. The model improves learning of both local and global features while getting rid of background noise that isn't useful. Tests done on common remote sensing datasets show big improvements in Intersection over Union (IoU), pixel precision, and boundary delineation when compared to standard U-Net and similar models. The suggested AMSE-U-Net works better for generalization with only a little amount of extra processing power, making it good for monitoring land cover and the environment.**

Keywords*—*Land cover classification; remote sensing; UNet; satellite images; AMSE-U-Net; multi-scale features; semantic segmentation

I. Introduction

Remote sensing is a highly useful way to find out what sort of land cover is in many various regions. It is vitally crucial to classify land cover from satellite photos in order to discover forests area, manage resources, preserve the environment, anticipate the weather, and locate cities, etc. [1]. Land cover classification gives us a lot of information about how the ecosystem on Earth works by carefully sorting the Earth's surface into different categories of land. This creates a complete picture of the many types of land coverings. This makes it easier to comprehend and keep an eye on changes in our natural environment as a whole [3], [4]. This systematic classification makes it easier to comprehend and manage different ways that land is used. Remote sensing technology has changed the way traditional surveys work, like field surveys, aerial photography, map analysis, community engagement, and other early remote sensing techniques. These new methods can effectively collect and analyze land use and cover data, giving us high quality images and important information about vegetation cover, water bodies, and land use, building etc. [5].

There are two main types of traditional land use categorization methods: pixel-based and object-oriented. Deep learning methods have shown to be quite effective in automating mapping through semantic picture segmentation [6]. Recently, in semantic segmentation, particularly in land use and land cover segmentation, traditional machine learning (ML) methods have been declining in popularity. Penatti et al. shown that convolutional neural networks significantly surpass traditional machine learning approaches in land cover categorization [7]. DeepGlobe 2018 Satellite image understanding challenge [8], all the top list were predominantly led by deep neural networks (DNNs) techniques [9]–[11].

Land cover looks different in different areas, therefore data from one location may not be useful for mapping other areas. Also, satellite images may be taken in many different spectral bands [12]. It is very important to make a model that accurately predicts different types of land cover so that we can find aquatic and urban areas on the Earth's surface. This is important for keeping track of changes in the environment and the growth of human settlements. Semantic segmentation is figuring out how to get from broad to specific parts of a scenario. The first step is categorization, which makes a guess about the input. Deep learning, particularly fully convolutional neural networks (FCNs) [13], has transformed semantic segmentation. U-Net has become a very popular model in remote sensing because of its symmetric encoder-decoder construction and skip links [14]. Even though UNet has shown proven to be one of the best approach of semantic segmentation, but it has its own set of drawbacks as listed out below:

- Weak multi-scale representation: Land cover classes vary drastically in scale and texture.
- Limited global context: Standard convolutions capture restricted receptive fields.
- Generalization issues: Satellite imagery differs across countries, sensors, and acquisition conditions.
- Detail loss: Downsampling operations often remove crucial fine boundary features.

To overcome the the above said limitations of UNet model we proposed a Advanced Multi-Scale Enhanced U-Net (AMSE-UNet), which includes the modifications as listed below:

1) Atrous Multi-Scale Enhancement Module (AMEM)
2) Hybrid Channel–Spatial Attention Block (HCSAB)
3) Cross-Level Context Fusion (CLCF)

With the above said modification the proposed model is used for the classification of remote sensing images into the distinct regions showing land cover and land use segments within the image.

II. Literature Review

Land use and land cover mapping was examined by Lilay and Taye through the use of high-resolution Sentinel-2 images and a collection of deep learning models, which included CNN-RF, CNN-SVM, and LinkNet-ResNet34 [15]. The strategy that they took was successful in achieving an accuracy of 88.2%, which demonstrates the advantages of combining comprehensive satellite data with powerful neural networks. In the study, persisting issues such as misunderstanding among visually similar land cover categories and insufficient training samples are highlighted. Additionally, the study suggests that future work should make use of deeper spectral information and extend datasets.

SA-SC U-Net is an improved form of U-Net that Khan and Jung developed [16]. It contains self-attention mechanisms and separable convolutions, and it was called after the enhanced U-Net variant. When the model was evaluated using aerial images from Dubai, it achieved a 91 percent accuracy rate, highlighting its capacity to extract comprehensive spatial context for the purpose of land cover research.

Hybrid approach that integrates a dynamic arithmetic edge-detection technique with a BiLSTM-U-Net architecture to delineate land use and land cover elements from hyperspectral pictures was proposed by Yele. The suggested model demonstrated outstanding performance on the DeepGlobe dataset, with 98% accuracy, 99% recall, and a 99% F1-score, surpassing most current approaches in managing high-dimensional remote sensing data [17].

Tzepkenlis et al. [18] used the U-TAE (U-Net with Temporal Attention Encoder) architecture, initially developed by Sainte Fare Garnot and Landrieu, to facilitate land cover mapping using satellite image time series. Their methodology streamlines preprocessing through the utilization of temporal median composites and enhances performance by using multi-sensor data from Sentinel-1, Sentinel-2, and ALOS, in conjunction with channel attention modules. The enhanced model surpasses methods such as random forest, U-Net, and SegFormer while utilizing less trainable parameters. Additional efforts are scheduled for temporal change modeling and object-based segmentation.

A dual-velocity ensemble classification method was presented by Horry et al., this method combines the advantages of a Vision Transformer and a fast-training CNN [19], [20]. Their staged training pipeline demonstrated its potential for scalable remote sensing applications by achieving a level of accuracy of 65 percent on the So2Sat LCZ42 dataset. A hybrid model of HRNet and PSPNet was utilized by Sun and Zheng in order to enhance the pixel wise segmentation of high resolution aerial and satellite images [21]. Their approach makes use of multiband analysis and adaptive frequency band selection in order to categorize nine different types of terrain, hence achieving significant increases in the quality of segmentation performance. The authors emphasize the importance of conducting further research on multimodal fusion and change-detection-oriented segmentation, as stated in the literature.

According to their findings, the Res UNet system that is fitted with a ResNet 50 encoder provides the optimal balance between the level of accuracy and the level of architectural complexity. The authors also highlight the significance of efficient preprocessing and highlight the limitations of standard U-Net models in terms of their ability to capture a wider range of contextual information. Chueprasert et. al. carried out this investigation into the comparative evaluation of deep learning methods for the extraction of buildings from satellite imagery [22].

The study emphasizes the necessity of merging government datasets with unlabeled satellite imagery to enhance agricultural monitoring. The classification framework proposed by Payani et al. [23] was implemented using Deep Feed Forward Neural Networks (DFNN) and 1D Convolutional Neural Networks. This framework was based on spectral indices and used data from Landsat. The DFNN model they built using the KAPLCU dataset had a test accuracy of 99.57%, which was higher than the accuracy of the 1D CNN. the accuracy achieved by the 1D CNN.

The MST DeepLabv3+ model was presented by Wang et al involved the integration of SENet and MobileNetV2 with transfer learning in order to provide semantic segmentation that was both efficient and accurate [24]. A MIoU of 82.47% and an OA of 92.13% were achieved by the model on the ISPRS dataset. These results demonstrate the model's ability to handle high-resolution remote sensing scenes while still maintaining a lightweight design.

Dataset for Indian research was made by Rengma and Yadav with their investigation into the creation of a medium-scale land use and land cover [25]. More specifically, they examined and evaluated both conventional machine learning techniques and contemporary deep learning methods. Machine learning and deep learning models are shown to have comparable levels of accuracy in their experiments, highlighting the fact that the selection of an algorithm ought to be determined by the computational resources that are available. In their recommendation, they suggest extending evaluations across multiple states in India by utilizing imagery with a higher resolution.

Behera et al. [26] examined the integration of Object-Based Image Analysis (OBIA) with machine learning classifiers for LULC mapping using LISS IV imagery.The study highlights challenges in segment boundary definition and computational demand while demonstrating the value of OBIA for context-aware land cover mapping. The best accuracy, 75.2%, was achieved using the CART classifier.

Wang et al. [27] suggested DeepG-DLE, a lightweight semantic segmentation model that improves DeeplabV3+ by using GhostNet and depthwise separable convolutions. The model performs well in a wide range of situations, such as when there are shadows or when the class distributions are not balanced. The Fuyang Remote Sensing Image Dataset (FRSID) was created to help with benchmarking. Azedou et al. said that combining spectral indices with Deep Neural Networks (DNNs) can classify land cover from Sentinel-2 data with 94.5% accuracy [28]. Random Search, Hyperband, and Bayesian optimization are some of the optimization methods that made the model work better. Their results show that deep learning in the cloud can be useful for keeping an eye on natural resource.

Albarakati et al. [29] proposed a fused self-attention-based deep learning model for LULC classification using IBNR-

65 and DenseNet-64 feature extractors. The incorporation of Quantum Hyperparameter Optimization significantly reduced computation time while improving classification accuracy to 98.2% on the SIRI-WHU dataset. Amer et al. reviewed state of art methods and concluded that deep learning—particularly CNN-based and hybrid multi-scale models—remains the most effective strategy for remote sensing image classification [30]. Boonpook et al. [31] introduced LoopNet, a novel CNN architecture tailored for semantic segmentation of LULC classes in Landsat-8 imagery. LoopNet effectively captures complex land cover patterns and outperforms traditional models, though it faces difficulties in differentiating urban and miscellaneous classes. The study reaffirms the growing importance of deep learning for accurate LULC mapping.

III. Related Work

Over the past 20 years, the way we classify land cover has changed a lot. It has evolved from basic pixel-based techniques to sophisticated deep learning architectures capable of modeling both spatial and contextual information. This section gives a general idea of these changes and points out the problems that led to the design of the proposed AMSE-U-Net architecture [32], [33]. The introduction of U-Net was a big step forward for remote sensing's ability to make dense predictions. You can mix fine spatial details from the early layers with deep semantic information from the deeper layers thanks to the encoder-decoder structure and skip connections. This hierarchical feature fusion has worked well to separate things with complicated edges, such as roads, rivers, and built-up areas [34].

A number of variants have tried to make the original U-Net better. UNet++ added nested, densely connected skip pathways to make the semantic gap between the encoder and decoder feature maps smaller. Pay attention U-Net used spatial attention gates to get rid of irrelevant activations and bring out important features. Residual U-Net and Dense U-Net made it even easier to reuse features and keep gradients stable by using residual and dense connectivity [35].

Even with these improvements, classical U-Net architectures still have problems when used for remote sensing. Images from remote sensing often show objects that are very different in size and texture, which means that fixed-size convolutional kernels can't capture all of the spatial patterns [35]. U-Net also doesn't have clear ways to model global context, which is important for telling apart land cover types that have similar spectral characteristics. These constraints have spurred growing research interest in hybrid architectures that integrate U-Net's structural advantages with multi-scale and attention-based improvements [36].

A. *Motivation for the Proposed Architecture*

The ideas behind multi-scale processing, attention modeling, and hierarchical feature fusion have led to the creation of hybrid architectures designed for remote sensing. But a lot of the models that are already out there only look at one thing, like multi-scale extraction, channel attention, or spatial refinement, and they don't put them all together into one framework. Remote sensing segmentation, which is hard because of the wide range of scales, textures, and spectral complexities, works best with an architecture that includes multi-scale features, channel recalibration, and attention-guided skip fusion all at once. The proposed AMSE-U-Net combines these ideas by adding Multi-Scale Blocks (MSBs), Squeeze-and-Excitation (SE) modules, and Attention Gates (AGs) to a U-Net backbone. This unified design captures fine spatial details, enhances global context, and improves feature relevance, resulting in a more robust and generalizable segmentation framework for land cover classification.

B. *AMSE Unet*

The proposed Advanced Multi-Scale Enhanced U-Net (AMSE-U-Net) aims to address the issues encountered by standard U-Net and its variants when applied to intricate remote sensing images. Satellite images with high resolution show objects in a wide range of sizes, shapes, and textures. Long-range dependencies and semantic inconsistencies between deep encoder features and shallow decoder features also make it harder to get the right segmentation. To fix these problems, AMSE-U-Net adds three modules: the Atrous Multi-Scale Enhancement Module (AMEM), the Hybrid Channel–Spatial Attention Block (HCSAB), and the Cross-Level Context Fusion (CLCF) unit. All of these parts work together to improve contextual modeling, refine features, and make sure that everything in the network is the same across all scales.

AMSE UNet is based on the classic UNet design, but instead of using standard convolutional blocks, it uses modules that are more flexible and aware of the situation. The encoder uses AMEM blocks to mix multi-scale context with atrous convolutions that have different dilation rates. This gives it features that are organized in a hierarchy. HCSAB modules make the feature maps better at both the encoder and decoder stages by showing off useful channels and areas of space. Lastly, the CLCF module makes sure that the encoder and decoder's features work well together. This makes skip connections better and helps the decoder make more accurate land cover boundaries.

The three core models of the architecture are shown below:

- Atrous Multiscale Enhancement Module (AMEM): captures multiscale contextual features and long-range spatial dependencies through dilated convolutions.
- Hybrid Channel–Spatial Attention Block (HCSAB): enhances discriminative feature learning by combining channel attention and spatial attention.
- Cross-Level Context Fusion (CLCF): reduces the semantic gap between encoder and decoder layers through adaptive cross-scale feature alignment.

This modular design ensures that both local and global contextual cues are efficiently preserved, enabling robust segmentation of heterogeneous land cover types.

C. *Atrous Multi-Scale Enhancement Module (AMEM)*

The AMEM module is made easier to collect contextual information at different scales without making the calculations harder. Standard convolutions have small receptive fields, which makes it hard for them to capture big structures like fields, bodies of water, or areas of vegetation. Atrous convolutions help with this problem by adding dilation factors that make the receptive field bigger.

The input feature map X, AMEM applies four parallel atrous convolutions with dilation rates {1, 2, 4, 8}:

$$F_d = \text{Conv}_d(X), \quad d \in \{1, 2, 4, 8\}. \tag{1}$$

These multi-scale features are concatenated as shown below:

$$F_{\text{concat}} = [F_1, F_2, F_4, F_8], \tag{2}$$

and fused through a 1×1 convolution to reduce dimensionality and enhance feature compactness:

$$F_{\text{AMEM}} = \text{Conv}_{1\times1}(F_{\text{concat}}). \tag{3}$$

This module provides richer context at each layer of the encoder, enabling better recognition of objects with varying scales and shapes.

D. Hybrid Channel–Spatial Attention Block (HCSAB)

Not all of the features that AMEM extracts are equally useful for segmentation, even though it adds depth to the context. AMSE-U-Net uses HCSAB, which combines channel attention and spatial attention to fully recalibrate features, to selectively improve relevant information.

Channel attention: Channel attention identifies the relative importance of each channel by applying a squeeze-and-excitation operation:

$$M_c = \sigma(\text{MLP}(\text{GAP}(X))), \tag{4}$$

where, GAP denotes Global Average Pooling and σ is the sigmoid activation. This mechanism captures channel-wise dependencies, allowing the network to strengthen informative channels while suppressing irrelevant ones.

Spatial attention: Spatial attention emphasizes important spatial regions by applying average pooling and max pooling across the channel dimension:

$$M_s = \sigma(\text{Conv}_{7\times7}([X_{\text{avg}}, X_{\text{max}}])). \tag{5}$$

Here, X_{avg} and X_{max} represent average-pooled and max-pooled feature maps, respectively. The 7×7 convolution captures spatial relationships over a wide receptive field.

Combined attention output: The recalibrated feature map is obtained by sequentially applying channel and spatial attention:

$$X' = X \odot M_c \odot M_s, \tag{6}$$

where, $\odot$ denotes element-wise multiplication. HCSAB thus enhances feature representations and improves the segmentation of boundaries, small objects, and spectrally similar land cover classes.

E. Cross-Level Context Fusion (CLCF)

Skip connections in classical U-Net simply concatenate encoder and decoder features, despite their semantic disparity. Encoder features tend to capture high-level semantic information at low resolution, whereas decoder features contain fine-grained spatial details. This mismatch often hampers segmentation performance.

To address this, the CLCF module aligns encoder and decoder features before fusion. Given encoder feature map F_{enc} and decoder feature map F_{dec}, CLCF applies transformation functions $\phi(\cdot)$ and $\psi(\cdot)$ to match channel dimensions and semantics:

$$F_{\text{CLCF}} = \phi(F_{\text{enc}}) + \psi(F_{\text{dec}}). \tag{7}$$

This operation ensures consistent representation across network levels, allowing the decoder to more accurately reconstruct object boundaries and spatial details. By mitigating semantic discrepancies, CLCF significantly enhances the effectiveness of skip connections.

IV. Proposed Method: AMSE UNet

A. Architecture Overview of the AMSE UNet

The Advanced Multi-Scale Enhanced UNet (AMSE UNet) is a new type of encoder-decoder architecture that was made to do very precise semantic segmentation of remote sensing images. It improves the classic U-Net by adding three major new features: 1) modules for extracting features at multiple scales, 2) recalibrating squeeze and excitation (SE) channels, and 3) skip connections guided by attention. These parts let the network pick up on small spatial structures while keeping track of global context, which makes it great for mapping complex land cover. Fig. 1 shows the block diagram of the proposed architecture.

B. Encoder: Multi-Scale Feature Extraction

The encoder is composed of a sequence of Multi-Scale Blocks (MSBs) designed to extract hierarchical and spatially rich feature representations. Each MSB processes the input using multiple convolutional branches:

- a standard 3×3 convolution,
- a dilated 5×5 convolution to enlarge the receptive field,
- a 1×1 convolution for channel mixing and dimensionality control.

The outputs of these parallel convolutions are concatenated to form a multi-scale feature map, followed by batch normalization and ReLU activation. Max-pooling operations at each stage downsample spatial dimensions while preserving feature richness. This design enables the encoder to capture both high-frequency details (e.g., edges, boundaries) and low-frequency patterns (e.g., vegetation texture, water bodies).

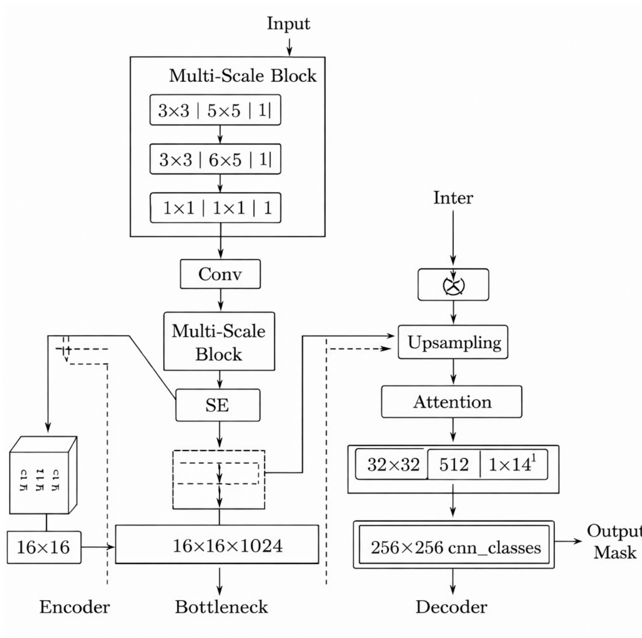

Fig. 1. Block diagram of the proposed AMSE-U-Net architecture.

C. Bottleneck: Contextual Enhancement with Squeeze-and-Excitation

At the network's deepest layer, the bottleneck consists of a Multi-Scale Block followed by a Squeeze-and-Excitation (SE) module. The SE block performs channel-wise attention:

1) Squeeze: Global average pooling aggregates spatial information into a channel descriptor.
2) Excitation: A gating mechanism with sigmoid activation learns per-channel importance.
3) Recalibration: Channels are reweighted to emphasize informative features and suppress noise.

This mechanism strengthens discriminative features, improves robustness against illumination and texture variations, and optimizes the information flow into the decoder.

D. Decoder: Attention-Guided Feature Reconstruction

The decoder reconstructs pixel-level predictions through upsampling and feature fusion. At each scale, feature maps are passed through:

- upsampling via transposed convolution,
- an Attention Gate (AG) applied to the corresponding encoder skip connection,
- a Multi-Scale Block to refine fused information.

Attention Gates suppress irrelevant encoder activations and retain only spatially consistent, task-relevant regions. This significantly improves segmentation quality along complex boundaries such as rivers, roads, and urban edges.

E. Output Layer

The final segmentation map is generated using a 1×1 convolution that maps decoder features to pixel-wise class probabilities. A sigmoid activation is used for binary segmentation, whereas softmax is applied for multi-class land-cover mapping. The output resolution matches the input resolution, maintaining precise localization.

F. Key Advantages of AMSE-U-Net

- Multi-scale feature extraction enhances the model's ability to detect objects of varying sizes.
- SE channel attention selectively amplifies informative channels.
- Attention-based skip connections reduce irrelevant spatial noise.
- Dilated convolutions extend receptive fields without increasing the parameter count.
- Improved boundary precision through deep decoder refinement.

Overall, AMSE-U-Net provides a strong and generalizable framework for remote sensing semantic segmentation, outperforming traditional U-Net variants in capturing fine details while maintaining contextual coherence.

V. Results

To evaluate the efficacy of the proposed AMSE-UNet Semantic Segmentation, tests were conducted on DeepGlobe dataset as shown in the Fig. 2. From the dataset 4 sample test images are taken and their obtained masks are shown along with the true masks. With the proposed algorithm the output masks perfectly lock into the different land cover regions of the remote sensing images

A. Experimental Setup

1) Dataset: The dataset used to get the results is called DeepGlobe Land Cover. It has 50 cm satellite images and 7 classes. The DeepGlobe Land Cover Classification dataset is a large-scale benchmark that includes high-resolution satellite images that have been carefully chosen for research on semantic segmentation [2]. The dataset was released as part of the 2018 DeepGlobe Challenge. It has RGB image tiles that are 2448 × 2448 pixels in size and have a ground sampling distance of about 50–60 cm. This makes it possible to do detailed analysis of different landscapes [8]. Each scene is manually labeled with seven land cover types: urban, agriculture, rangeland, forest, water, barren, and unknown. This gives dense pixel-level supervision that is good for training deep learning models. The dataset was gathered from a wide range of locations, so it has a lot of variation in scene structure, object size, and land cover patterns. This makes it a great resource for testing the robustness and generalization of segmentation architectures like U-Net, DeepLabv3+, and transformer-based models. DeepGlobe is one of the most popular benchmarks for remote sensing land cover classification because of its size, the quality of its annotations, and the variety of its classes.

2) Preprocessing: Images were normalized, resized to 512×512, and augmented using flipping, rotation, color jitter, and Gaussian noise.

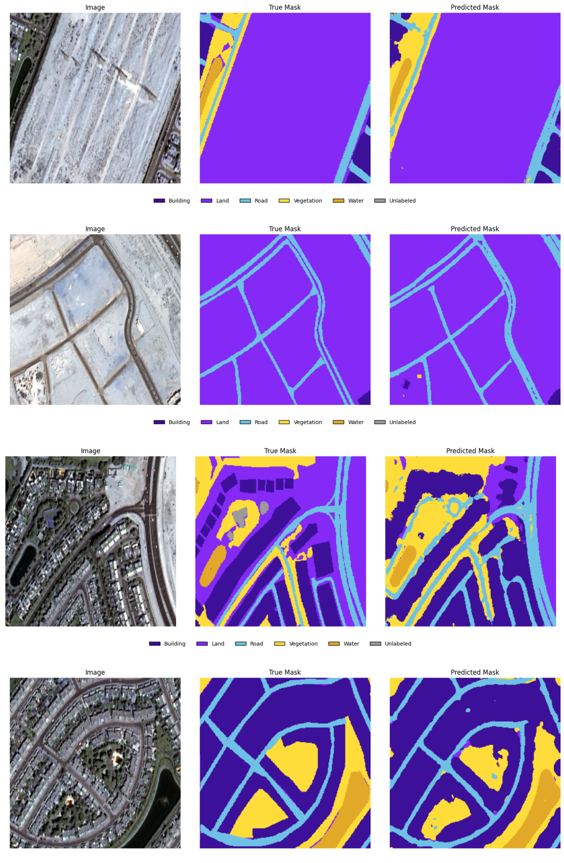

Fig. 2. Columns (i) & (iv) are input images, (ii) & (v) are ground truths, and (iii) & (vi) are predicted masks.

3) Training details: Training of the model done using:

- Optimizer: AdamW
- Learning rate: 10^{-4}
- Batch size: 8
- Loss: BCE + Dice loss
- Epochs: 100

4) System configuration and other settings: The CPU is Huawei-made Intel(R) Core(TM) i3 processor with 2.10GHz speed. The operating system is Windows11 (64-bit operating system), and the development environment is python .For the DeepGlobe dataset, testing images were selected randomly from the dataset. For each of the original images the model predicts the segmentation. The comarative analysis of our approach with other deep learning algorithms is shown in the Table I.

TABLE I. mIoU on DeepGlobe

Model	mIoU (%)
U-Net	61.0
UNet++	63.4
DeepLabv3+	65.8
HRNet-Seg	66.1
AMSE-U-Net (proposed)	**69.3**

VI. Conclusion

In this work, an expanded U-Net-based architecture known as AMSE-U-Net was presented. The purpose of this architecture was to overcome the drawbacks of standard segmentation models in terms of managing the multi-scale complexity of land cover representations. By incorporating the Atrous Multi-Scale Enhancement Module (AMEM), the Hybrid Channel–Spatial Attention Block (HCSAB), and the Cross-Level Context Fusion (CLCF), the proposed model is able to effectively capture a wide variety of contextual information, improve feature discrimination, and enhance semantic alignment across all levels of the network.According to the findings of the experiments, AMSE-U-Net performs significantly better than standard U-Net systems when it comes to segmenting and locating borders. It has the capability of employing multi-scale context and attention-driven refinement in order to produce classifications of various forms of land cover that are more accurate. AMSE-U-Net is a formidable and adaptable framework for the segmentation of remote sensing data that has the potential to be utilized in the fields of environmental monitoring, urban planning, and geospatial analysis.

References

[1] Fayaz, Muhammad, Junyoung Nam, L. Minh Dang, Hyoung-Kyu Song, and Hyeonjoon Moon. 2024. "Land-Cover Classification Using Deep Learning with High-Resolution Remote-Sensing Imagery" Applied Sciences 14, no. 5: 1844. https://doi.org/10.3390/app14051844.

[2] Burcu Amirgan, Arzu Erener,"Semantic segmentation of satellite images with different building types using deep learning methods", Remote Sensing Applications: Society and Environment, Volume 34,2024,101176, ISSN 2352-9385, https://doi.org/10.1016/j.rsase.2024.101176.

[3] K. S. Kalaivani et al., "Land Cover Segmentation in Satellite Images Using Transfer Learning," in Proc. ICoICI, Coimbatore, India, 2024, pp. 714–724.

[4] L. Thomas Ramos and A. D. Sappa, "Multispectral Semantic Segmentation for Land Cover Classification: An Overview," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 17, pp. 14295-14336, 2024.

[5] A. Shreehari, P. R. Chandan, P. V. Vinod and R. Hebbar, "Land Use Land Cover Classification using Deep Learning Techniques: A Comparative Study," 2024 11th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 2024, pp. 7-11, doi: 10.1109/SPIN60856.2024.10511460.

[6] N. Bengana and J. Heikkilä, "Improving Land Cover Segmentation Across Satellites Using Domain Adaptation," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 1399-1410, 2021, doi: 10.1109/JSTARS.2020.3042887.

[7] O. A. B. Penatti, K. Nogueira and J. A. dos Santos, "Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?," 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Boston, MA, USA, 2015, pp. 44-51, doi: 10.1109/CVPRW.2015.7301382.

[8] I. Demir et al., "DeepGlobe 2018: A Challenge to Parse the Earth through Satellite Images," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 2018, pp. 172-17209, doi: 10.1109/CVPRW.2018.00031.

[9] C. Tian, C. Li and J. Shi, "Dense Fusion Classmate Network for Land Cover Classification," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 2018, pp. 262-2624, doi: 10.1109/CVPRW.2018.00049.

[10] G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, "Densely Connected Convolutional Networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 2261-2269, doi: 10.1109/CVPR.2017.243.

[11] T. -S. Kuo, K. -S. Tseng, J. -W. Yan, Y. -C. Liu and Y. -C. F. Wang, "Deep Aggregation Net for Land Cover Classification," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 2018, pp. 247-2474, doi: 10.1109/CVPRW.2018.00046.

[12] A. Yehia, H. Elhifnawy and M. Safy, "Effect of Different Spatial Resolutions of Multi-temporal Satellite Images Change Detection Application," 2019 International Conference on Innovative Trends in Computer Engineering (ITCE), Aswan, Egypt, 2019, pp. 41-46, doi: 10.1109/ITCE.2019.8646510.

[13] Q. Zhu and X. Zu, "Fully Convolutional Neural Network Structure and Its Loss Function for Image Classification," in IEEE Access, vol. 10, pp. 35541-35549, 2022, doi: 10.1109/ACCESS.2022.3163849.

[14] A. Kumar and T. Singh, "U-NET Architecture for Liver Segmentation using Multi Model Scans," 2022 IEEE 19th India Council International Conference (INDICON), Kochi, India, 2022, pp. 1-6, doi: 10.1109/INDICON56171.2022.10039786.

[15] M. Y. Lilay and G. D. Taye, "Semantic Segmentation Model for Land Cover Classification from Satellite Images in Gambella National Park, Ethiopia," SN Applied Sciences, vol. 5, p. 76, 2023.

[16] B. A. Khan and J.-W. Jung, "Semantic Segmentation of Aerial Imagery Using U-Net with Self-Attention and Separable Convolutions," Applied Sciences, vol. 14, no. 9, p. 3712, 2024.

[17] V. P. Yele, R. R. Sedamkar, and S. Alegavi, "Systematic Analysis of Effective Segmentation and Classification for Land Use Land Cover in Hyperspectral Image using Deep Learning Methods," in Proc. 20th CSI Int. Symp. Artificial Intelligence and Signal Processing (AISP), Babol, Iran, 2024, pp. 1–8.

[18] A. Tzepkenlis, K. Marthoglou, and N. Grammalidis, "Efficient Deep Semantic Segmentation for Land Cover Classification Using Sentinel Imagery," Remote Sensing, vol. 15, no. 8, p. 2027, 2023.

[19] V. S. F. Garnot and L. Landrieu, "Panoptic segmentation of satellite image time series with convolutional temporal attention networks," in Proc. IEEE/CVF Int. Conf. Computer Vision, 2021, pp. 4872–4881.

[20] M. J. Horry, S. Chakraborty, B. Pradhan, et al., "Two-Speed Deep-Learning Ensemble for Classification of Incremental Land-Cover Satellite Image Patches," Earth Syst. Environ., vol. 7, pp. 525–540, 2023.

[21] Y. Sun and W. Zheng, "HRNet- and PSPNet-based multiband semantic segmentation of remote sensing images," Neural Comput. Appl., vol. 35, pp. 8667–8675, 2023.

[22] T. Chueprasert, A. Udomchaiporn, and S. Intagosum, "Comparative Analysis of Deep Learning Models for Building Extraction from High-resolution Satellite Imagery," Current Applied Science and Technology, vol. 25, no. 1, 2024.

[23] C. A. Payani, C. Gupta, and A. K. M., "Spectral Indices based Land Cover Classification using Deep Learning," in Proc. ICRAIS, Manipal, India, 2024, pp. 166–171.

[24] Y. Wang, L. Yang, X. Liu, and P. Yan, "An Improved Semantic Segmentation Algorithm for High-Resolution Remote Sensing Images Based on DeepLabv3+," Scientific Reports, vol. 14, p. 9716, 2024.

[25] N. S. Rengma and M. Yadav, "Generation and Classification of Patch-Based Land Use and Land Cover Dataset in Diverse Indian Landscapes," Environmental Monitoring and Assessment, vol. 196, no. 6, p. 568, 2024.

[26] D. K. Behera, G. S. Pujar, R. Kumar, and S. K. Singh, "A comprehensive approach towards enhancing land use land cover classification through machine learning and object-based image analysis," Journal of the Indian Society of Remote Sensing, vol. 53, no. 3, pp. 731–749, 2025.

[27] G. Wang, J. Chen, L. Mo, P. Wu, and X. Yi, "Lightweight Land Cover Classification via Semantic Segmentation of Remote Sensing Imagery and Analysis of Influencing Factors," Frontiers in Environmental Science, vol. 12, p. 1329517, 2024.

[28] A. Azedou et al., "Enhancing Land Cover/Land Use (LCLU) Classification through a Comparative Analysis of Hyperparameters Optimization Approaches for Deep Neural Network (DNN)," Ecological Informatics, vol. 78, p. 102333, 2023.

[29] H. M. Albarakati et al., "A Novel Deep Learning Architecture for Agriculture Land Cover and Land Use Classification from Remote Sensing Images Based on Network-Level Fusion of Self-Attention Architecture," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 17, pp. 6338–6353, 2024.

[30] I. Amer, K. Rahouma, and A. Shahin, "Remote sensing image processing, deep learning classification algorithms survey and review," Journal of Advanced Engineering Trends, vol. 43, no. 1, pp. 239–267, 2024.

[31] W. Boonpook et al., "Deep Learning Semantic Segmentation for Land Use and Land Cover Types Using Landsat 8 Imagery," ISPRS Int. J. Geo-Inf., vol. 12, no. 1, p. 14, 2023.

[32] T. Lu, L. Wan, S. Qi, and M. Gao, "Land Cover Classification of UAV Remote Sensing Based on Transformer–CNN Hybrid Architecture," Sensors, vol. 23, no. 11, p. 5288, 2023.

[33] Irfan, Ayesha, Yu Li, Xinhua E, and Guangmin Sun. 2025. "Land Use and Land Cover Classification with Deep Learning-Based Fusion of SAR and Optical Data" Remote Sensing 17, no. 7: 1298. https://doi.org/10.3390/rs17071298

[34] Li, Zan, Hong Zhang, Zhengzhen Li, and Zuyue Ren. 2022. "Residual Attention UNet++: A Nested Residual-Attention U-Net for Medical Image Segmentation" Applied Sciences 12, no. 14: 7149. https://doi.org/10.3390/app12147149

[35] Zhu, Zhifu, Xiping Yuan, Shu Gan, Raobo Li, Weidong Luo, Rui Bi, and Cheng Chen. 2025. "Texture and Structure Interaction Guided Generative Adversarial Network for Multimodal Remote Sensing Image Change Detection." GIScience & Remote Sensing 62 (1). doi:10.1080/15481603.2025.2565866.

[36] Punn, N.S., Agarwal, S. Modality specific U-Net variants for biomedical image segmentation: a survey. Artif Intell Rev 55, 5845–5889 (2022). https://doi.org/10.1007/s10462-022-10152-1