P- ISSN 1991-8941 E-ISSN 2706-6703
2025,(19), (01):233 – 247

Journal of University of Anbar for Pure Science (JUAPS)

Open Access

Automatic Image Segmentation and Visualization Using CNN and LSTM–GRU Models for Remote Sensing Applications

Noor Eid Ibrahim[1]*, Azmi T. Hussain Alrawi[1], Muthanna M. Abdulhameed[2]

[1]College of Computer Science and IT, University of Anbar, Ramadi, Iraq
[2]College of Engineering, University of technology, Baghdad, Iraq

ARTICLE INFO

Received: 23 / 06 /2024
Accepted: 18/ 08 /2024
Available online: 20/ 06 /2025

DOI: 10.37652/juaps.2024.150377.1265

Keywords:
Remote Sensing Images, Image Segmentation Technique, Clustering, CNN, Automated Image Management.

ABSTRACT

Rapid urbanization driven by population and economic growth reshapes land use patterns and challenges sustainable development. This trend, especially in densely populated areas, requires innovative urban planning and conservation strategies. Remote sensing provides essential imagery for monitoring these changes and capturing the evolving urban landscapes. In this study, we proposed an innovative framework that can automatically manage and visualize the deep learning-based image segmentation output with limited supervision in remote sensing applications. The framework was constructed by integrating recent advances in machine learning models to resolve the problems of scalability and accessibility in processing remote sensing images. We illustrated our methodology with a series of steps, which included preprocessing of remote sensing images, dividing the remote sensing images logically, converting the image data into grayscale images, and developing clustering models, such as K-means and self-organizing maps, to cluster images into logical groups through the same regular pixel intensity. Thereafter, two primary deep learning architectures were implemented, which included a convolutional neural network (CNN) and a hybrid long short-term memory (LSTM)–gated recurrent unit (GRU), specifically designed to efficiently and effectively process image data. The CNN achieved a loss value of 0.015308, mean absolute error (MAE) of 0.083680, mean squared error (MSE) of 0.015308, and root mean squared error (RMSE) of 0.135449, indicating that the model provides accurate image reconstruction. The LSTM–GRU model yielded a slightly higher loss of 0.015364, MAE of 0.076740, MSE of 0.015364, and RMSE of 0.151264, indicating that it can preserve spatial hierarchies and contextual understanding due to slight performance variability. The two models demonstrated excellent processing capabilities that provide evidence on how they can be utilized in urban planning, environmental monitoring, and natural resource management.

Introduction

Urban sprawl/extended urbanization in less-developed areas, such as portions of Asia and Africa, frequently gives rise to unplanned settlements, including slums, shantytowns, and urban villages. These unplanned settlements are prevalent due to the rapid urbanization and the demand for low-cost housing among low-income urban citizens. Unplanned settlements are usually closely spaced, low-rise structures that offer inexpensive housing for low-income urban residents.

However, these unplanned settlements typically feature extreme residential dissimilarities, limited or no access to amenities, unsanitary conditions, and safety concerns.

Given that some cities attempt to manage destitute settlements, meet sustainability goals, or respond to institutional pressures, understanding the challenges and current status of unplanned settlements is essential for urban planners and policymakers. In some territories, policymaking and supporting reconstructions are addressed by unplanned settlement management and enhancement of the schemes. Urban village is the most prevalent form of an unplanned urban settlement in China. Traditionally, unplanned sites in urban regions have been identified by land use departments through

*Corresponding author at : College of Computer Science and IT, University of Anbar, Ramadi, Iraq
ORCID: https:// https://orcid.org/0009-0009-7690-603X ,
Tel: +964 7827475643
Email: noo21c1011@uoanbar.edu.iq

field studies aimed at collecting property dimensions. However, these field studies are costly and time-consuming, particularly when covering urban villages. An additional technique is to use remote sensing to compensate for the high costs associated with their existence, widespread use, and limited updating of basement safety schemes. Low-altitude aerial photos or unmanned aerial vehicle images are utilized to collect and digitalize building footprints, a process that is often time-consuming. Consequently, an efficient and smart system that utilizes graphics is urgently needed [1-9].

Mapping urban buildings using remote sensing has a long-established history and is well documented in the literature [10]. Increasingly available high-resolution optical images and object-based image analysis (OBIA) have become prominent methods for mapping urban buildings [9,11,12,13]. Image target detection methods typically use current spectral and spatial information. Lately, random forest, a machine learning tool, has been broadly used for mapping built-up areas [15,16]. Despite the availability of various classification strategies for land use/land cover mapping in urban areas, pixel-based or object-based methods may struggle to distinguish individual buildings in high-density built-up areas, often aggregating several buildings into a single segment [17]. Urban buildings exhibit high variability in spectral, textural, and shape features, and the traditional remote sensing framework is inadequate for capturing new feature characteristics [7,18,19,20]. Image segmentation is a process that divides an image into sections based on neighboring pixels with similar feature, such as brightness, texture, and color [13], which can be challenging for areas with a high density of built-up areas. OBIA segmentation faces significant challenges in this context, including the selection of scale parameter and rule application. Boundary delineation and shape preservation are difficult due to the noisy and textural information on the edge of the segments [21]. However, a universally applicable and implemented methodology is yet to be published to date. Consequently, how to develop a robust and reliable building segmentation method for mapping informal urban building remains a challenge.

By contrast, several studies have shown that high-density slums can be mapped from remote sensing images based on their physical characteristics, which distinguish them from formal outer settlements [1,2,4,16]. Finally, mapping slum areas should involve not only delineating entire areas but also identifying individual building types. For example, information on the potential of the benefits after urban renewal, from the perspective of public and private decision-makers, would require data on the number of reconstructed buildings, their distance from fire sources, and investments in public services and environmental amenities [22,23]. Accordingly, defining and distinguishing buildings are highly important. The semantic segmentation issue is another well-known challenge in computer vision, which involves masking spatial regions of interest. In recent years, machine learning technologies have gained popularity, and deep learning techniques have attracted increasing interest within the remote sensing society for data processing [14,17,24,25,26,27]. Deep convolution neural networks have been widely used for semantic segmentation [20,27]. The most well-known deep-learning algorithm for image segmentation is the fully convolutional neural network (CNN) proposed by Long et al. [28]. This deep-learning framework enables end-to-end, pixel–pixel semantic classification possibilities of their recognition as long as they have been trained on their own network. Consequently, CNN has gained attention as it can extract relevant and important contextual information for decision making and image identification [28,29]. Thus, CNN-based semantic segmentation has been used for various pixel images, such as road extraction, building extraction, urban land use classification, pool semantic classification, vehicle detection, and damage and aftermath categorizations [1,24,30,31,32].

However, the classical image classification approach faces substantial problems in high-density built-up areas. Given that CNNs typically outperform classical image classification methods, deep learning-based semantic segmentation is considered suitable for isolating separate houses in SPAs. Semantic segmentation is a type of deep learning that uses CNNs to grasp the preeminent representations of features and extract the object from an image. The CNN-based approach to learn feature representations from samples has multiple benefits for SPAs. First, this approach is fast because it eliminates the need for human supervision during the learning process. Network training takes several hours on a desktop

computer. Afterward, the classification process takes only a few seconds. Second, the application is automated because deep learning methods can extract feature representations from a vast pool of data, whereas classical visual recognition methods require human effort to extract a few features sufficient for the classification algorithm. Given the significant diversity of current urban land cover, manual feature extraction is impossible. Furthermore, deep learning methods can reapply the training to the tested samples, thus minimizing manual labor. However, training a network on traditional data takes approximately four working days. Moreover, deep learning outperforms classical methods in remote sensing, particularly in areas where boundary line quarrying is most challenging [19,26,27].

This study introduces an innovative framework designed to automate the management and visualization of deep learning-based image segmentation outputs. Our approach addresses the dual challenges of scalability and accessibility in processing remote sensing imagery by leveraging state-of-the-art machine learning models and advanced visualization techniques. We detail the development and implementation of this framework, emphasizing its potential to transform the analysis of remote sensing data by enabling efficient, accurate, and intuitive exploration of segmented images.

In the following sections, we will explore the background and significance of image segmentation in remote sensing, review the current landscape of deep learning techniques tailored for this purpose, and highlight the importance of automated systems in managing and visualizing complex datasets. Our contribution aims to set a new standard for handling segmented remote sensing imagery, paving the way for enhanced environmental monitoring, urban planning, and natural resource management.

Literature Review

After reviewing the literature on the management and visualization of deep learning-based image segmentation in remote sensing aerial photographs, numerous methods and techniques have been proposed. These efforts aim to build on past technological advances to improve the accuracy of target identification and the general processing characteristics of remote sensing images. However, the presence of numerous subsidiary contributors remains a possibility. Key articles include those in which building footprint polygons are indirectly obtained: first, buildings are identified in remote sensing images, and their format is converted from raster to vector. Although buildings typically have well-defined edges, pixel-wise semantic segmentation algorithms fail to accurately delineate the lines between pixels, resulting in highly unpredictable maps. Building footprints can reflect such variation: one involves improving how buildings change shape with boundary regulation during the conversion from raster and vector data.

An optimization method involves improving the accuracy of building boundary classification in remote sensing images. Wu et al. [33] proposed a boundary-regulated network consisting of a modified U-Net and multitasking framework, wherein the segmentation maps are the building outlines. When considering the boundary regulation effect, the method proposed by Wi et al. demonstrates improved accuracy in delineating building boundaries in remote sensing images. This approach is an improvement over traditional U-Net by addressing the complexity of boundaries, although it does not effectively exploit their traits. To this end, the researchers have used a specialized neural network architecture to optimize the deployment of algorithms as descriptively as possible. Marmanis et al. [34] proposed DCNN models for semantic segmentation of high-resolution aerial images to address the problem of indistinct object boundaries. Their model incorporates class boundaries into the segmentation process, significantly improving the distinction between result classes on a pixel-by-pixel basis. Liao et al. [35] proposed a boundary-preserving building extraction approach based on artificial intelligence. Their method strengthened the building boundaries by embedding contour information in the labels, thereby improving performance on the boundary lines of adjacent buildings.

Another optimization method is to improve the building boundary vertices during the data format conversion. Classical vector data processing algorithms include Douglas-Peucker [36], Wang-Müller [37], and Zhou-Jones [38]. Maggiori et al. [39] introduced a novel algorithm that exploits a labeled triangle grid to approximate the classification maps. Despite various

measures to improve the generated vector data of buildings, the final polygons often fail to accurately represent the route footprints. At times, these measures have a significant number of vertex points or exhibit artificial bends. In other instances, these measures are distorted due to the challenges faced by machine recognition software in distinguishing whether multiple buildings are adjacent and determining the optimal way to handle them based on their proximity. Numerous common methods for extracting building footprints were originally designed to generate polygons for individual buildings with regular shapes, such as rectangles.

Polygon extraction differs from pixel classification or segmentation as it involves transforming raster data into vector data, presenting a challenging task for conventional deep learning models. Pixel classification or segmentation, which involves transforming raster data into raster data, is the opposite of polygon extraction. Although polygon extraction is a strict transformation of raster data into vector data, in a deep learning model for pixel-wise classification, the task is simply to determine whether pixels belong to objects or to the background. However, the task changes with polygon extraction. Here, the model must locate and identify the key vertices while perceiving the way they interrelate. In the field of geography, the use of deep learning methods to identify key points for geographical features from remote sensing images is rare. Using deep learning to directly detect key points presents two main problems. The first one is the imbalance between positive samples — of which there are few when detecting a key point — and negative samples; the other problem is the inability to determine how many key points are present in an image. When treated as a classification problem, training and debugging the deep learning network can become particularly challenging. Meanwhile, CNNs face a significant challenge when dealing with outputs of non-fixed lengths.

Song et al. [40] proposed an FCN-based method to detect building corners in aerial images, based on the predicted building footprints. They extracted corners according to the contours of these footprints. However, the performance of this method is severely affected by the accuracy of semantic segmentation. To date, no deep learning-based model has been established to treat the geographical position of building corners as a direct optimization objective.

Methodology

The methodology adopted for the automated management and visualization of deep learning-based image segmentation for remote sensing data in Ramadi Municipality involved a streamlined sequence of steps starting from image acquisition to detailed analysis. First, remote sensing images were gathered and preprocessed to ensure uniformity in size and quality. The subsequent phases involved the application of advanced filtering techniques to improve the dataset, eliminating poor images according to the pixel-magnitude constraints. Thereafter, these images were transformed into gray scales to reduce the computational weight, and clusters were used to arrange the images in a meaningful manner. Afterward, the outcomes were further processed with deep learning models to perform precise image segmentation, followed by the last evaluation and overall data visualization. All particular stages were designed to improve the accuracy and speed of the image analysis and, in future sections, act as a rigorous basis for in-depth exploration. An illustrative graph is presented in Figure 1.

Data Acquisition

Study Area

The study area in this work is the Ramadi Municipality, Iraq. Ramadi holds a strategically important geographic and strategic location, encompassing a diverse range of landscapes. These landscapes cover high-density urban regions and fertile agricultural lands to expansive deserts. Accordingly, Ramadi presents an excellent research area for the diverse geographic and environmental analyses required. The boundaries that define the study area in this work are those of the administrative boundaries of Ramadi Municipality. This geographical area approximately covers 8340 km^2. This region is positioned in the geographical coordinates 33.430866 to 43.295059. The above-mentioned geographical information indicates that the area is strategically situated along several major roads and encompasses a diverse range of landscapes. This geography is varied and influences various environmental

aspects, such as climate and agricultural activities. The study area was chosen due to its diverse and well-defined landscapes, which provide a solid foundation for the application and analysis of various remote sensing and image segmentation techniques. Consequently, the use of such an area as the area of interest equips the researcher with experience in data collection and management under varying conditions. Hence, the outcome of this study will provide information on how the remote sensing equipment can be used to manage and analyze a topographical diverse area, such as Ramadi.

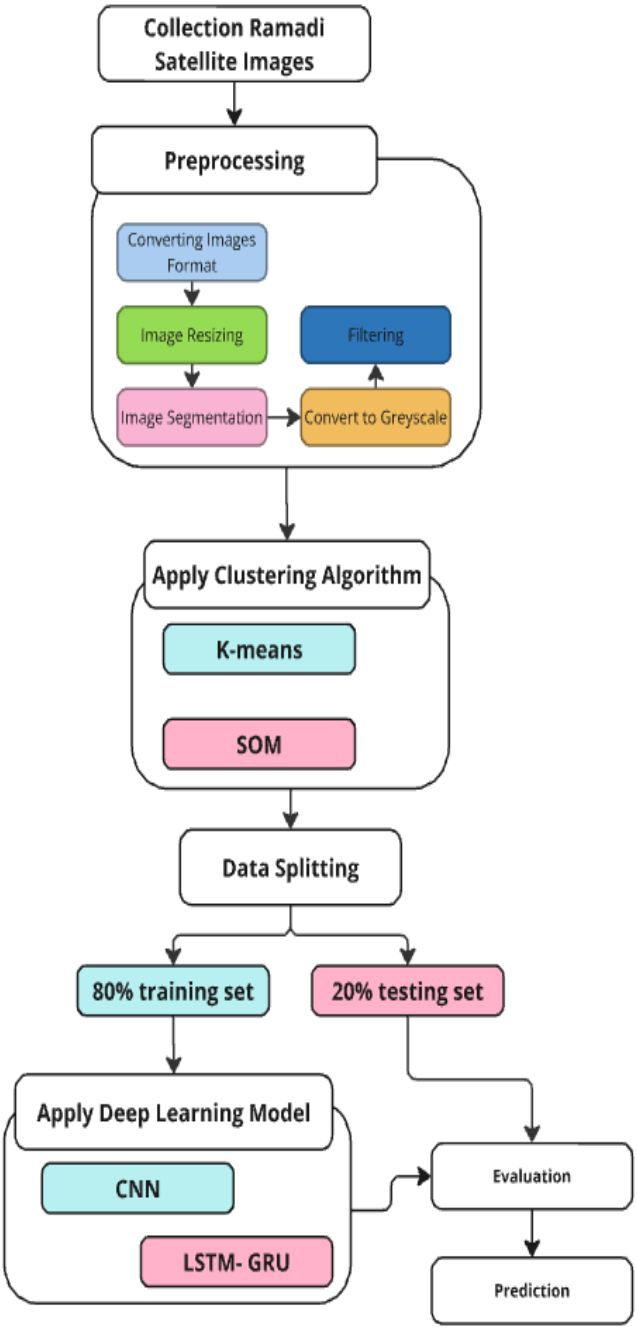

Figure 1: Methodology flowchart.

Data Collection Period

The data collection period was intentionally set between 2004 and 2017. Figure 2 presents an example of an image from the dataset. This time span was deemed most appropriate to ensure the use of only recently captured and high-quality images and facilitate a comprehensive examination of the various seasonal and annual alterations affecting urban and natural

development in Ramadi Municipality. This approach also makes it possible to analyze considerable transformations experienced by the region over time due to human interventions and natural causes. This knowledge can offer an in-depth understanding of the interdynamic nature of urban sprawl, agriculture, and desertification processes.

Figure 2: Example of a sample from the dataset.

The selection of satellite images was based on several criteria to ensure optimal data quality and applicability, implemented through a combination of manual review and programming tools. Cloud cover was a primary criterion, with only images containing less than 10% overcast being selected to maximize clarity and minimize atmospheric interference. Lighting conditions were also considered, favoring images captured during peak light to avoid shadows and excessive glare that could impair analysis. Additionally, images from previously studied areas or those with available topographical maps and plans were prioritized to support data analysis and validation. These criteria were carefully applied, using manual evaluation and automated algorithms, to ensure that the selected remote sensing data would be highly relevant for detailed studies of geographic and environmental conditions. This approach guarantees that decision-makers receive comprehensive, spatially explicit information to aid in regional planning and management.

Preprocessing

During the preprocessing phase of our study, we performed a number of important steps to prepare the remote sensing images for the subsequent steps of the analysis. First, all the images in TIFF format are converted to JPEG files to reduce the file sizes and standardize the format for easy manipulation and analysis. After the conversion, the images are resized to ensure that the dimensions across the dataset are standardized because we need accurate image measurements for analysis. The lowest common image dimension is identified from the existing dataset, and all images are resized to that dimension to ensure that the aspect ratio and image integrity are preserved. The following steps of the methodology, including image segmentation and feature extraction, are efficiently and effectively conducted. Given that the accuracy of the clustering and deep learning models used in the subsequent phase of the study depends on the quality of the preprocessing, this phase must be properly performed.

After the preprocessing, the dataset is divided into training and testing subsets to facilitate model development and evaluation. Specifically, 80% of the

images, totaling 100 images, are allocated to the training set, while the remaining 20%, consisting of 20 images, are designated as the testing set. This division ensures that the models are trained on a comprehensive dataset while being evaluated on a separate subset to accurately assess their performance.

Image Segmentation

Under this phrase, the resized images undergo image segmentation, which divides each image into smaller, more manageable components. This task is achieved by programmatically dividing each image into a 5×5 grid, resulting in 25 segments per image. The 5×5 grid is chosen to provide a balanced level of detail, allowing for a comprehensive coverage and detailed analysis. This approach helps focus the analysis on distinct identified features within the area of interest (Figure 3). Thereafter, the segmented images obtained are saved in a folder specified for processing. Some images are visualized to understand the quality of the segmentation through a visual assessment and ensure a random representation from the land cover. The visualization provides an indication of segmentation accuracy and offers a quick comparison of the natural and artificial land cover areas captured in the segmentation (Figure 2). This mechanism enables the subsequent deep learning models by allowing for the precise application of the models to the clearly identified image areas, significantly improving the analysis time and quality efficiency at the phase level.

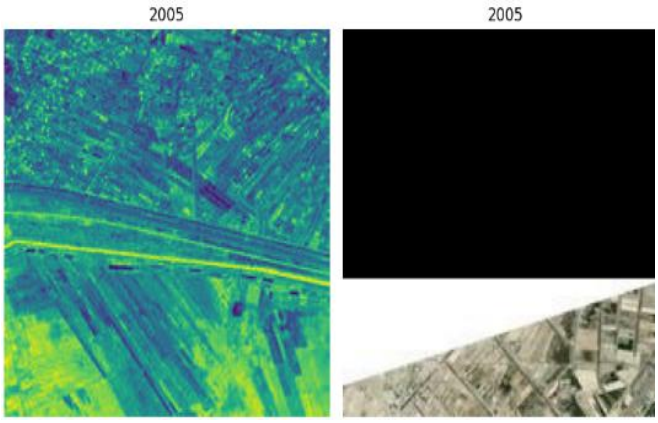

Figure 3: Example of a sample from the image segmentation result “satellite image from 2005”.

Convert to Grayscale

The conversion of the images to grayscale remains a preprocessing step in our workflow but is critical prior to highly sophisticated image analysis tasks. In this step, the segmented color images are transformed into grayscale. This mechanism simplifies the input data by reducing it to a single channel that represents intensity. This simplification of the processed data is desirable because it reduces the workload during processing and simplifies later processes of pattern recognition and feature extraction. The removal of color information ensures that the algorithms will generate patterns and features based on textural and structural cues in the images. This approach is generally sufficient for most image processing tasks, including image feature, edge detection, and as input prior to particularly advanced machine learning tasks. The grayscale images are saved in a systematic way in a designated directory, ensuring easy retrieval to facilitate the next steps in the image processing pipeline. This level of systematic processing facilitates a higher system performance in the efficiency of operation of the image analysis pipeline.

Filtering

Filtering is a crucial step in the image processing pipeline, aimed at enhancing data quality by excluding images that do not meet specific conditions for reliable analysis. This step involves examining each pixel in grayscale images to identify black pixels, which typically indicate non-informative regions or underexposed areas.

In quantifying the quality of each image, the percentage of black pixels within an image is calculated using a predefined intensity threshold, set at ($I_t = 10$) on a scale of 255 (where 0 is pure black, and 255 is pure white). The formula used to determine the percentage of black pixels, (P_b), in an image is expressed as follows:

$$P_b = \left(\frac{\text{Number of pixels with intensity} < I_t}{\text{Total number of pixels}}\right) \times 100.$$

Images containing more than 50% of the black pixels are considered unsuitable for father processing and are automatically eliminated from the dataset. This threshold enables only those images with an adequate proportion of visible features to proceed further, which is vital for the subsequent feature extraction and machine-learning models used in this study. Consequently, only the most relevant and informative images are retained for further

analysis, thereby increasing the quality and robustness of the final results.

Clustering

In accomplish this task, our image processing workflow uses high-level clustering algorithms that consider the image pixels' intensities in the grayscale and classify similarly intense pixels to the same group. This method is necessary for grouping, as it arranges the images in such a way that cameras that are similar are in the same cluster. This mechanism is important as the arrangement is necessary for the next level of analysis, which involves searching for signs that define the images or discrepancies within the images. In this stage, two clustering approaches were utilized, namely, "SOM" and K-means clustering. The former groups the images in a 3×1 grid called a topographical representation in a bid to realize a 1D pattern while maintaining correlations in the input space. The strength of SOM lies in transforming high-dimensional data into lower dimensions, thereby highlighting the interrelationships and dependencies within the data. The latter groups the images into three in each cluster/pol after measuring the distance between the pixels, which lie on the Euclidean assumption, unless proven otherwise.

The effectiveness of these clustering methods is assessed using several statistical measures:

Silhouette score, which computes the closeness or separation of the cluster by measuring the distance within the clusters to that between the clusters [41]. The silhouette score of each image is represented as follows:

$$s = \frac{b - a}{\max(a, b)},$$

where a is the mean distance to the other instances in the same cluster (intra-cluster distance), and b is the mean distance to the instances of the next closest cluster (inter-cluster distance).

Davies–Bouldin Index (DBI), wherein the metric is calculated as the mean ratio of distances within clusters to the distances between clusters. The smaller the DBI, the better the clustering stick together [42]. This factor is calculated as follows:

$$DBI = \frac{1}{k}\sum_{i=1}^{k} \max_{j \neq i}\left(\frac{\sigma_i + \sigma_j}{d(c_i, c_j)}\right),$$

where σ_i is the average distance of all points in cluster i to the centroid c_i, and $d(c_i, c_j)$ is the distance between centroids c_i and c_j.

Calinski–Harabasz Index, which is the ratio of the sum of between-clusters dispersion and of intra-cluster dispersion for all clusters [43]. The greater the value of this function, the better the clusters are defined. This factor is calculated as follows:

$$s = \frac{\mathrm{Tr}(B_k)/(k-1)}{\mathrm{Tr}(W_k)/(n-k)},$$

where B_k is the between-group dispersion matrix, W_k is the within-cluster dispersion matrix, k is the number of clusters, and n is the number of points.

The results of those measurements are crucial for selecting the optimal clustering approach and adjusting its parameters. This mechanism allows for the effective grouping of images, ensuring that they are informative in the subsequent steps of image analysis, particularly when distinguishing between areas and surfaces.

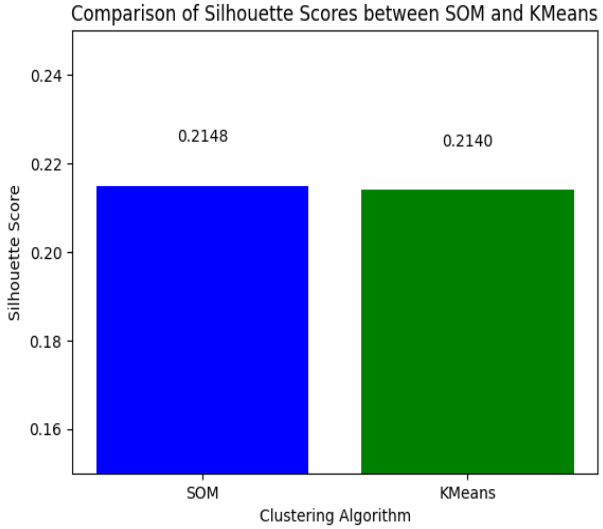

Figure 4: Silhouette score comparison: shows nearly equal performance, with K-means slightly outperforming SOM.

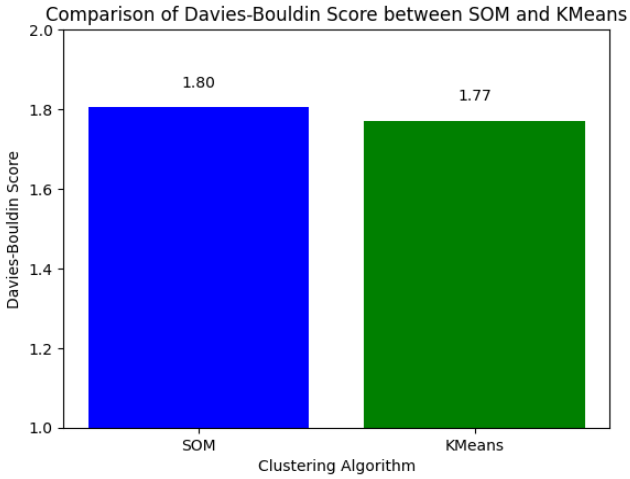

Figure 5: DBI comparison: lower scores indicate better performance, with K-means slightly outperforming SOM.

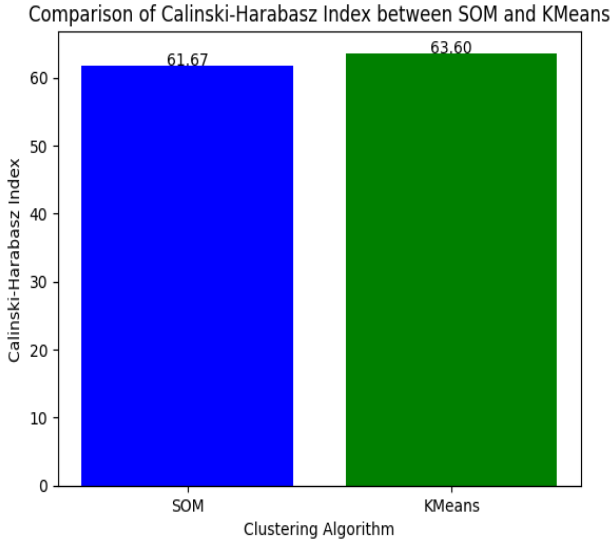

Figure 6: Calinski–Harabasz Index comparison: measures cluster definition quality, showing K-means as superior to SOM.

The visual comparison of the Figures 4, 5, and 6 reveals the efficacy of the clustering algorithms SOM and K-means in clustering grayscale images based on their pixel intensity patterns. In Figure 3, the silhouette scores of SOM and K-means are particularly close to each other, and even narrowly. Meanwhile, K-means outperforms SOM, which does not bring about significant improvement in the cluster's cohesion and separateness. The results indicate that the K-means clusters are slightly more compact and better separated than those of SOM. Figure 4 exhibits the DBI, where both methods display similar scores, with K-means outperforming SOM. K-means forms clusters that are more compact than those formed by SOM, as the Davies–Bouldin score shows a clear relationship with that of Figure 3, the silhouette score. In Figure 5, the Calinski–Harabasz Index shows K-means outperforming SOM again, although the improvement is minimal. The Calinski–Harabasz Index calculates the ratio of between cluster to within cluster sum of squares, with higher values indicating clearer clusters. The greater K-means score suggests that it has a clearer line to separate the clusters than SOM. Thus, K-means would be a better choice of clustering.

CNN Architecture

In our study, we adapted CNN architecture that is specially designed for efficiently working with images. This model consists of multiple convolutional layers and down-sampling layers, also known as max-pooling layers, with batch normalization layers in between. The inclusion of batch normalization helps in stabilizing and accelerating the training process, making it highly efficient. The model starts with an input layer expecting images that are resized to 256×256 pixels in three color channels. The first three sequential 3×3 convolutions form the early part of the CNN, known as the encoder. These layers include 32 filters at first and become increasingly refined, finding edges of objects or textures, up to 128 filters. After convolution, max pooling is applied with a 2×2 window for each output. Max pooling involves reducing the spatial dimensions to focus on the critical components of the spatial data. Next in the CNN is the bottleneck layer, a dense convolutional layer with 256 filters. This section is responsible for heavily distilling the information extracted from input images.

The central component of the CNN on its path back to recreating the full image based on the extracted data is the decoder. This sector restores the feature information to the original pixel-wise size. This task is performed by deconvolutional aka transposed convolution layers. During this process, the dimensions are expanded multiple times, and the output is refined back into an image from a feature map. Each deconvolutional layer is concatenated with the encoder output of the same size to loop the data back into the process to preserve high-resolution features from the input data in a ground-up manner. This task is carried out using skip flow, which is passed through each layer of the encoder. The ultimate output of the CNN is made up of two 3×3 convolutions

and a single 3×3 component, triggering a re-mapping of nodes into a functional output in grayscale form. The last function from the last two layers is a sigmoid, which, with modified x values, forces the output limit between zero and one, similar to the initial model. This terminator function ensures that the convolution will attempt to create clear, nuanced reconstructions of input images for detailed geophysical objects.

Table 1: CNN parameters

Parameter	Value
Input image size	**256×256 pixels**
Number of layers	**12 layers (3 Conv, 3 MaxPool, 3 ConvTranspose, and 3 batch normalization)**
Filter size	**3×3**
Filters per layer	**32, 64, 128, and 256**
Activation function	**LeakyReLU and sigmoid (output)**
Pooling type	**MaxPooling**
Pooling size	**2×2**
Strides in transpose convolution	**2**
Batch normalization	**Yes, after each layer**
Dropout	**50% at the last convolutional layer**
Loss function	**MSE**
Optimizer	**Adam**
Learning metrics	**MAE, MSE, and RMSE**
Validation split	**10%**
Epochs	**2000**

Long Short-Term Memory (LSTM)–Gated Recurrent Unit (GRU)

In our advanced image processing setup, we implemented a hybrid model with LSTM and GRU networks, tailored due to the specific nature of sequence data among the rows of the image. This type of architecture is especially suited for problems where a spatial hierarchy must be maintained among the image segments to obtain a complete contextual understanding. The model begins with the input layer set to take colored images that have been resized to the dimension of 256×256 pixels. Subsequently, a series of TimeDistributed layers applies the same operation to the GRU at each frame, essentially treating every row in the images as a sequence, to find dependencies.

The image data are first taken in by a GRU layer with 32 units, which helps in finding the first level of spatial dependencies. Thereafter, the sequenced output of this level is provided to another GRU layer with 64 units, which helps the model in gaining an in-depth understanding of the spatial connections. At this point, the architecture also includes two layers with LSTM, each with 64 units, which play a critical role in remembering long connections and further refining the feature extraction already performed by GRUs.

The data outputted from the last LSTM layer is fully connected and passed through a dense layer with 128 units that use the "ReLu" activation function to introduce nonlinearity and help in finding complex patterns. The final layer of the model is a dense output layer with the reshaping of results into the same format as the input image that uses sigmoid to ensure that the output pixel values are normalized between zero and one, similar to that performed for the input images to preprocess them.

Table 2: LSTM–GRU Parameters

Parameter	Value
Input image size	**256×256 pixels**
Number of GRU layers	**2**
Units in the first GRU layer	**32**
Units in the second GRU layer	**64**
Number of LSTM layers	**2**
Units in LSTM layers	**64 each**
Activation function (LSTM/GRU)	**Tanh (internal) and sigmoid (output)**
Dense layer units	**128**
Activation function (dense layer)	**ReLu**
Output layer activation	**Sigmoid**
Optimizer	**Adam**
Loss function	**MSE**
Metrics	**MAE, MSE, and RMSE**
Training epochs	**2000**
Batch size	**16**
Validation split	**10%**

Experimental Results

In summary, the experimental outcomes between the CNN and the LSTM–GRU hybrid models showed that both models are effective in image processing tasks, as revealed by their metrics presented in Table 4. For instance, the CNN obtained a loss of 0.015308, an mean absolute error (MAE) of 0.083680, a mean squared error (MSE) of 0.015308, and a root mean squared error (RMSE) of 0.135449. The reconstruction's high accuracy provides the evidence suggesting that the CNN's model is good at learning and reconstructing the images' critical features as fed to it.

By contrast, the LSTM–GRU indicates that it had a loss of 0.015364, an MAE of 0.076740, an MSE of 0.015364, and a higher RMSE of 0.151264. The only indicators that the LSTM–GRU model had better performance than the CNN were the lower MAE and the slightly higher RMSE. The lower MAE indicates that the LSTM–GRU was closer to the actual pixel values; however, the final image may still show a slight variation because the RMSE value is relatively higher. Both models have the potential to learn and analyze complex images, although they appear to serve different aspects of the image analysis in terms of detail and the overall reconstruction accuracy.

Table 3: Comparison metrics

Model	Loss	MAE	MSE	RMSE
CNN	**0.01530**	**0.0836**	**0.01533**	**0.1354**
LSTM–GRU	**0.015362**	**0.07673**	**0.015363**	**0.151263**

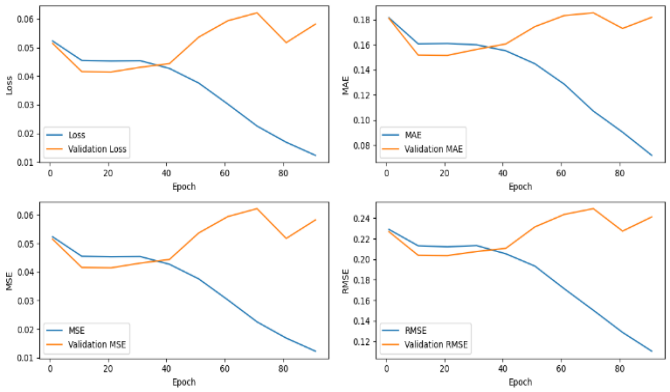

Figure 7: Training metrics for CNN: graphs of training and validation loss

Figure 7 exhibits the progression of training metrics over epochs for a CNN applied to an image processing task. The figure consists of four subplots representing loss, MAE, MSE, and RMSE for the training data (blue line) and validation data (orange line). We observe a notable decrease in training loss and errors across all metrics with the increase in the number of epochs, indicating the model's effective learning and improvement. However, the validation metrics demonstrate fluctuation, suggesting some degree of overfitting or instability in model performance on unseen data as training progresses.

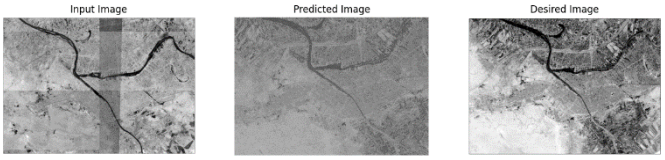

Figure 8: CNN prediction comparison

Figure 8 compares the results of the CNN's predictions with actual imagery. This figure displays three images: the "Input Image", which is the original image fed into the CNN, the "Predicted Image", which is the output from the CNN, and the "Desired Image", which represents the ideal or target outcome for the input image. The "Predicted Image" shows a commendable resemblance to the "Desired Image", indicating that the CNN has effectively learned and replicated significant features from the input. However, discrepancies in clarity and detail are evident between the predicted and the desired images, suggesting areas where the model may still be improved to enhance accuracy and image detail in its predictions. The visual comparison provides a straightforward assessment of the model's current capabilities and limitations in recreating precise image features and textures.

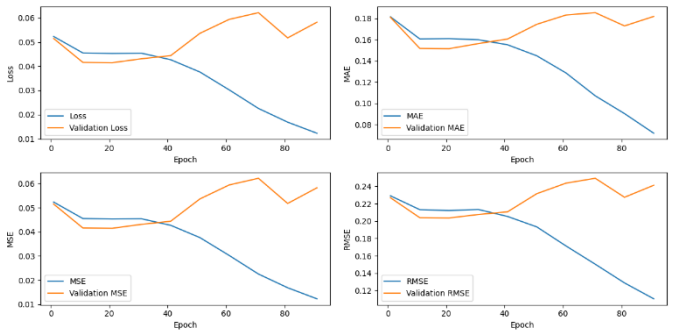

Figure 9: Training metrics for LSTM–GRU: graphs of training and validation loss

Figure 9 exhibits the LSTM–GRU hybrid model for image processing training and validation metrics for different epochs. Data in loss, MAE, MSE, and RMSE. MSE shows a decreasing trend for the training data, implying improvement while learning occurs. However, the validation curve maintains a fluctuating movement that implies changes in the learning patterns for data unseen. MAE and RMSE are similar, with a slight decline in training error balance, followed by a trend of increased error variation on validation, indicating overfitting, or the model is becoming unstable as the training progresses.

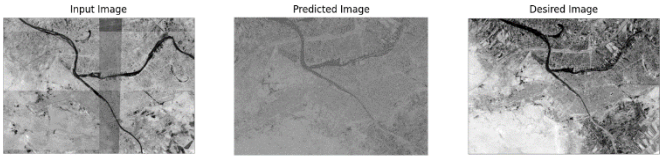

Figure 10: LSTM–GRU prediction comparison

After training the LSTM–GRU model with LSTM units, the best models are used to predict results. Figure 10 presents a visual set that includes an "Input Image", a "Predicted Image", and a "Desired Image", showcasing the model's practical outcomes in predicting 2022 results. The "Input Image" displays the original data provided to the model, the "Predicted Image" shows the model's output, and the "Desired Image" represents the expected outcomes. A comparison of these images indicates that the Predicted Image captures the general physical layout, the main structural elements, and other common features present in the Desired Image. The accuracy of this image varies due to differences in texture, with some aspects remaining unclear. The results indicate the model's accuracy in predicting major topographic features. Areas for improvement have been identified, particularly in capturing finer details and enhancing the quality of the outputs. Strengthening the model in these aspects can lead to highly accurate predictions and improved overall performance.

Conclusion

This study proposed an innovative framework to automate the management and visualization of deep learning based image segmentation outputs on remote sensing data. The framework utilized state-of-the-art machine learning models and enhanced visualization techniques to improve the scalability and accessibility of image processing of remote sensing imagery. Remote sensing technologies have made an affordable manner of gather spatial data and updating base map data without extensive field surveys. Deep learning-based image segmentation is essential for detailed urban planning, environmental monitoring, and natural resource management. In this study, we preprocessed the remote sensing imagery for load reduction. Second, the image was segmented into small manageable parts and merged as grayscale data for easy handling. We utilized K-means and self-organizing map clustering techniques to label the images according to pixel intensity patterns and cluster the segmented parts. Finally, we proposed a deep learning image segmentation model. We proposed two architectures and implemented the primary models: CNN and the hybrid LSTM and GRU model. CNN was implemented to validate the performance of reconstruction accuracy, which revealed high metrices and proved the selection of essential features. The second hybrid LSTM and GRU model depicted performance variations and validated the image in segmentation accuracy and also revealed group 3 data's behavior. However, the models exhibit promising performance in solving complex images and convey insights for further applications.

References

[1]. Wurm, M.; Stark, T.; Zhu, X.; Weigand, M.; Taubenböck, H. Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks. ISPRS J. Photogramm. Remote Sens. 2019, 150, 59–69.

[2]. Mahabir, R.; Croitoru, A.; Crooks, A.T.; Agouris, P.; Stefanidis, A. A critical review of high and very high-resolution remote sensing approaches for detecting and mapping slums: Trends, challenges and emerging opportunities. Urban Sci. 2018, 2, 8.

[3]. Kuffer, M.; Persello, C.; Pfeffer, K.; Sliuzas, R.; Rao, V. Do we underestimate the global slum population? In Proceedings of the Joint Urban Remote Sensing Event 2019, Vannes, France, 22–24 May 2019; pp. 1–4.

[4]. Kuffer, M.; Pfeffer, K.; Sliuzas, R. Slums from space—15 years of slum mapping using remote sensing. Remote Sens. 2016, 8, 455.

[5]. United Nations. Habitat iii issue papers 22—Informal settlements. In United Nations Conference on Housing and Sustainable Urban Development; United Nations: New York, NY, USA, 2015.

[6]. Mahabir, R.; Crooks, A.; Croitoru, A.; Agouris, P. The study of slums as social and physical constructs: Challenges and emerging research opportunities. Reg. Stud. Reg. Sci. 2016, 3, 399–419.

[7]. Li, Y.; Huang, X.; Liu, H. Unsupervised deep feature learning for urban village detection from high resolution remote sensing images. Photogramm. Eng.

Remote Sens. 2017, 83, 567–579.

[8]. Buchanan, T. Photogrammetry and Projective Geometry: An Historical Survey; SPIE: San Francisco, CA, USA, 1993; Volume 1944.

[9]. Bachofer, F.; Braun, A.; Adamietz, F.; Murray, S.; Angelo, P.d.; Kyazze, E.; Mumuhire, A.P.; Bower, J. Building stock and building typology of kigali, rwanda. Data 2019, 4, 105.

[10].Patino, J.E.; Duque, J.C. A review of regional science applications of satellite remote sensing in urban settings. Comput. Environ. Urban Syst. 2013, 37, 1–17.

[11].Blaschke, T. Object based image analysis for remote sensing. ISPRS J. Photogramm. Remote Sens. 2010, 65, 2–16.

[12].Liu, J.; Li, P.; Wang, X. A new segmentation method for very high resolution imagery using spectral and morphological information. ISPRS J. Photogramm. Remote Sens. 2015, 101, 145–162.

[13].Jin, X. Segmentation-Based Image Processing System. U.S. Patent 20,090,123,070, 14 May 2009.

[14].Emmanuel, M.; Yuliya, T.; Guillaume, C.; Pierre, A. Convolutional neural networks for large-scale remote sensing image classifcation. IEEE Trans. Geosci. Remote Sens. 2017, 55, 645–675.

[15].Rodriguez-Galiano, V.F.; Chica-Olmo, M.; Abarca-Hernandez, F.; Atkinson, P.M.; Jeganathan, C. Random forest classification of mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. Remote Sens. Environ. 2012, 121, 93–107.

[16].Duque, J.C.; Patino, J.E.; Betancourt, A. Exploring the potential of machine learning for automatic slum identification from vhr imagery. Remote Sens. 2017, 9, 895.

[17].Huang, B.; Zhao, B.; Song, Y. Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery. Remote Sens. Environ. 2018, 214, 73–86.

[18].Chen, R.; Li, X.; Li, J. Object-based features for house detection from rgb high-resolution images. Remote Sens. 2018, 10, 451.

[19].Li, W.; He, C.; Fang, J.; Zheng, J.; Fu, H.; Yu, L. Semantic segmentation-based building footprint extraction using very high-resolution satellite images and multi-source gis data. Remote Sens. 2019, 11, 403.

[20].Yi, Y.; Zhang, Z.; Zhang, W.; Zhang, C.; Li, W.; Zhao, T. Semantic segmentation of urban buildings from vhr remote sensing imagery using a deep convolutional neural network. Remote Sens. 2019, 11, 1174.

[21].Carleer, A.P.; Debeir, O.; Wolf, E. Assessment of very high spatial resolution satellite image segmentations. Photogramm. Eng. Remote Sens. 2004, 71, 1285–1294.

[22].Hui, E.C.M.; Dong, Z.; Jia, S.H.; Lam, C.H.L. How does sentiment affect returns of urban housing? Habitat Int. 2017, 64, 71–84.

[23].Loures, L.; Vaz, E. Exploring expert perception towards brownfield redevelopment benefits according to their typology. Habitat Int. 2018, 72, 66–76.

[24].Ball, J.E.; Anderson, D.T.; Chan, C.S. Comprehensive survey of deep learning in remote sensing: Theories, tools, and challenges for the community. J. Appl. Remote Sens. 2017, 11, 042609.

[25].Zhang, L.; Zhang, L.; Du, B. Deep learning for remote sensing data: A technical tutorial on the state of the art. IEEE Geosci. Remote Sens. Mag. 2016, 6, 22–40.

[26].Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep learning in remote sensing: A comprehensive review and list of resources. IEEE Geosci. Remote Sens. Mag. 2017, 5, 8–36.

[27].Garcia-Garcia, A.; Orts-Escolano, S.; Oprea, S.; Villena-Martinez, V.; Garcia-Rodriguez, J. A review on deep learning techniques applied to semantic segmentation. arXiv 2017, arXiv:1704.06857.

[28].Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015.

[29].Yang, H.L.; Yuan, J.; Lunga, D.; Laverdiere, M.; Rose, A.; Bhaduri, B. Building extraction at scale using convolutional neural network. J. Latex Cl. Files 2015, 14, 1–15.

[30].Bai, Y.; Mas, E.; Koshimura, S. Towards operational

satellite-based damage-mapping using u-net convolutional network: A case study of 2011 tohoku earthquake-tsunami. Remote Sens. 2018, 10, 1626.

[31].Wagner, F.H.; Sanchez, A.; Tarabalka, Y.; Lotte, R.G.; Ferreira, M.P.; Aidar, M.P.M.; Gloor, E.; Phillips, O.L.; Aragão, L.E.O.C. Using the u-net convolutional network to map forest types and disturbance in the atlantic rainforest with very high resolution images. Remote Sens. Ecol. Conserv. 2019..

[32].Garcia-Garcia, A.; Orts-Escolano, S.; Oprea, S.; Villena-Martinez, V.; Garcia-Rodriguez, J. A review on deep learning techniques applied to semantic segmentation. arXiv 2017, arXiv:1704.06857.

[33].Wu, G.; Guo, Z.; Shi, X.; Chen, Q.; Xu, Y.; Shibasaki, R.; Shao, X. A Boundary Regulated Network for Accurate Roof Segmentation and Outline Extraction. Remote Sens. 2018, 10, 1195.

[34].Marmanis, D.; Schindler, K.; Wegner, J.D.; Galliani, S.; Datcu, M.; Stilla, U. Classification with an edge: Improving semantic image segmentation with boundary detection. ISPRS J. Photogramm. Remote Sens. 2018, 135, 158–172.

[35].Liao, C.; Hu, H.; Li, H.; Ge, X.; Chen, M.; Li, C.; Zhu, Q. Joint Learning of Contour and Structure for Boundary-Preserved Building Extraction. Remote Sens. 2021, 13, 1049.

[36].Douglas, D.H.; Peucker, T.K. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. Cartogr. Int. J. Geogr. Inf. Geovis. 1973, 10, 112–122.

[37].Wang, Z.; Müller, J.-C. Line generalization based on analysis of shape characteristics. Cartogr. Geogr. Inf. Syst. 1998, 25, 3–15.

[38].Zhou, S.; Jones, C.B. Shape-aware line generalisation with weighted effective area. In Developments in Spatial Data Handling; Springer: Berlin/Heidelberg, Germany, 2005; pp. 369–380.

[39].Maggiori, E.; Tarabalka, Y.; Charpiat, G.; Alliez, P. Polygonization of remote sensing classification maps by mesh approximation. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 560–564.

[40].Song, W.; Zhong, B.; Sun, X. Building corner detection in aerial images with fully convolutional networks. Sensors 2019, 19, 1915.

[41].SHAHAPURE, Ketan Rajshekhar et NICHOLAS, Charles. Cluster quality analysis using silhouette score. In : *2020 IEEE 7th international conference on data science and advanced analytics (DSAA)*. IEEE, 2020. p. 747-748.

[42].Ros, Frédéric, Rabia Riad, and Serge Guillaume. "PDBI: A partitioning Davies-Bouldin index for clustering evaluation." *Neurocomputing* 528 (2023): 178-199.

[43].Wang, Xu, and Yusheng Xu. "An improved index for clustering validation based on Silhouette index and Calinski-Harabasz index." *IOP Conference Series: Materials Science and Engineering*. Vol. 569. No. 5. IOP Publishing, 2019.

P- ISSN 1991-8941 E-ISSN 2706-6703 *Journal of University of Anbar for Pure Science (JUAPS)* Open Access
2025,(19), (01):233 – 247

تقسيم الصور وتصورها تلقائيًا باستخدام نماذج CNN وLSTM–GRU لتطبيقات الاستشعار عن بعد

نور عيد ابراهيم[1]* ، عزمي توفيق حسين[1] ، مثنى محمدعبدالحميد[2]

[1]كلية علوم الحاسبات وتكنولوجيا المعلومات، جامعة الانبار/ رمادي–العراق
[2]كلية الهندسة، الجامعة التكنولوجية / بغداد– العراق
noo21c1011@uoanbar.edu.iq

الخلاصة:

في هذه الدراسة، اقترحنا إطارًا مبتكرًا يمكنه إدارة وتصور مخرجات تجزئة الصور القائمة على التعلم العميق تلقائيًا مع إشراف محدود في تطبيقات الاستشعار عن بُعد. وقد تم بناء الإطار من خلال دمج التطورات الحديثة في نماذج التعلم الآلي وتقنيات التصور المتطورة لحل مشاكل قابلية التوسع وإمكانية الوصول في معالجة صور الاستشعار عن بُعد. قمنا بتوضيح منهجيتنا من خلال سلسلة من الخطوات، والتي تضمنت المعالجة المسبقة لصور الاستشعار عن بُعد، وتقسيم صور الاستشعار عن بُعد منطقيًا، وتحويل بيانات الصور إلى صور ذات تدرج رمادي، وتطوير نماذج تجميع مثل K–means وخرائط التنظيم الذاتي لتجميع الصور في مجموعات منطقية من خلال نفس كثافة البكسل المنتظمة. ثم تم تنفيذ بنيتين أساسيتين للتعلّم العميق، بما في ذلك الشبكة العصبية التلافيفية والشبكة العصبية التلافيفية الهجينة LSTM–GRU التي تم تصميمها خصيصًا لمعالجة بيانات الصور بطريقة سريعة وفعالة. وحققت الشبكة العصبية التلافيفية قيمة خسارة قدرها 0.015308، ومتوسط متوسط الأرباح والخسائر 0.083680، و0.015308، و0.135449، مما يشير إلى أن النموذج يوفر إعادة بناء دقيق للصور. خلص نموذج LSTM–GRU المطور إلى خسارة أعلى قليلاً بلغت 0.015364، ومتوسط متوسط الأرباح المتوقعة 0.076740، ومتوسط الأضرار المتوسطة 0.015364، وقيمتها 0.015364، وقيمتها 0.151264، مما يشير إلى أنها يمكن أن تحافظ على التسلسلات الهرمية المكانية والفهم السياقي بسبب التباين الطفيف في الأداء.