Systematic Literature Review on Deep Learning Approach for Detecting Deforestation and Predicting Highly Threatened Areas

A.I.Hewarathna
17APC3081
Department of Computing and Information Systems
Faculty of Computing
Sabaragamuwa University of Sri Lanka

Content

Abstract

This paper provides an overview of the progress of deep learning segmentation techniques and their use in change detection using satellite imagery. The emphasis is on studying segmentation patterns and investigating alternative architectural techniques used by researchers. Furthermore, the study looks into the use of input data in studies. A special emphasis is placed on how domain-specific change detection approaches can be used to improve forest cover change detection. Not only does the research underline the importance of change detection across domains, but it also dives into how ideas and approaches from various areas can be effectively utilized to improve forest cover change detection. Planning, disaster monitoring, and environmental research can all teach us valuable lessons. These domains provide insights and strategies that can aid in the detection and continual monitoring of changes in forest cover. The findings of this review contribute to the advancement of deep learning approaches, particularly in the context of monitoring and managing forest ecosystems. This overview of the literature provides a thorough grasp of the evolution of deep learning segmentation approaches for change detection in satellite data. Furthermore, it investigates the potential advantages of leveraging change detection techniques from other domains to increase the accuracy and efficiency of forest cover change detection.

Keywords:

Forest cover, Change detection, Semantic segmentation, Change detection, Satellite images, Remote sensing, U-net, Semantic segmentation survey

Introduction

While forests once covered more than 50% of the Earth's land area approximately 8,000 years ago, their coverage has significantly decreased, currently accounting for only 30% of the total land area. The primary cause of forest cover loss is human actions, specifically the permanent conversion of forests to other land uses, commonly known as deforestation. Deforestation, alongside forest degradation, forms a mixed duo that contributes to the alarming rates of forest loss. Deforestation involves the irreversible conversion of forests to alternative land uses or the long-term reduction of tree cover in specific forested regions. On the other hand, forest degradation encompasses changes occurring within forests that compromise their ability to provide essential structures and functions, ultimately reducing their capacity to deliver valuable ecosystem services and products. Multiple drivers are responsible for the escalating rates of deforestation and forest degradation, which can be categorized into two main types: direct and indirect drivers. Direct drivers predominantly arise from human activities, such as agricultural expansion, excavation activities, urbanization and infrastructure development, as well as both man-made and natural fires. Conversely, indirect drivers encompass factors like population migrations, changes in agro-technology, production factors, political influences, regulatory frameworks, market demands, and climate change-induced alterations in soil and terrain characteristics. Understanding and addressing these drivers are crucial for effective forest management and conservation efforts. By comprehending the complex interplay between direct and indirect drivers, policymakers, researchers, and stakeholders can develop targeted strategies and interventions to mitigate deforestation and forest degradation. Furthermore, recognizing the multifaceted nature of these drivers enables the formulation of comprehensive and sustainable approaches to safeguard forests, ensuring their invaluable contributions to ecological stability, biodiversity conservation, and the provision of vital ecosystem services.

Deforestation and forest cover change

A forest is a complex environment dominated by trees. It is nature's very efficient system, with a remarkable capacity to perform photosynthesis, and it has a huge influence on both plants and animals in the environment. Forests may flourish in a wide range of conditions, giving rise to a diverse range of soil types, plant species, and animal life. These variances are caused by the various environmental elements that determine forest ecosystems. Trees are the primary structural elements of a forest, contributing to the overall functioning of the ecosystem. They absorb carbon dioxide and release oxygen through photosynthesis, which meets animal oxygen demands while simultaneously controlling the Earth's temperature. The dense canopy of trees offers shade and safety, forming a microclimate that is home to a wide variety of creatures including birds, insects, and mammals. The complexities of a forest's interactions extend beyond trees to include other plant and animal species. The forest is home to a variety of plant species, each of which fills a distinct niche and contributes to overall biodiversity. This diversity of vegetation provides food, shelter, and nesting grounds for a diverse range of animal species. Animals, in turn, contribute to seed dispersal and pollination, boosting the regeneration and reproductive processes of the forest.

Most forests can divide into three types, boreal forests, temperate forests and tropical forests. Boreal woods, often known as taiga, are located in North America, Asia, and Europe between 50 and 60 degrees latitude. Underneath the boreal woods is a territory formed by glaciers, which have left their mark on the geology, hydrology, and soils of the region. The severe cold environment of boreal woods makes living difficult, resulting in limited species variety when compared to temperate and tropical forests. Plants and animals that do exist in boreal woods have evolved to deal with short growth seasons and freezing temperatures. Because of their breadth and isolation, boreal forests are major carbon sinks. Boreal forests have the shortest growth season of the three forest types, lasting roughly 130 days. Conifers are the most common form of a tree, although there are some well-adapted deciduous trees as well, such as willows, poplars, and alders. Black and white fir, jackpine, balsam fir, and tamarack are among the prominent species. Blueberry and cranberry plants in the understory provide high-energy food for animals. Climate change poses a significant danger to boreal forests. Almost 80% of boreal forests are built on top of permafrost, a layer of frozen earth that remains frozen all year. The ground becomes soft and marshy as temperatures rise at excessively rapid rates, and many trees finally lose support and die. Temperate woods are found in the mid-latitudes and have four distinct seasons. The zone is dominated by secondary forests, with only a few areas of old-growth temperate forest remaining. Temperate woods are home to species that have adapted to seasonality. To save energy, deciduous trees such as maples, hickories, oaks, and many more shed their leaves and go dormant in the fall and winter. Bears, bobcats, squirrels, and deer live in temperate woods and can stockpile food, modify their diet, or hibernate to survive the winter's shortage of nourishing foods. Many endangered species live in temperate woods. In the United States, 12 mammal species are categorized as Endangered by the Fish and Wildlife Service. Many endangered species live in temperate woods. In the United States, 12 mammal species are categorized as Endangered by the Fish and Wildlife Service. Tropical woods, located between the Tropics of Cancer and Capricorn at 23 degrees north and south, are among the most biodiverse ecosystems on the planet. These woods span barely a tenth of the planet's area yet are home to half of all species. They are also among the most vulnerable to human activity. Tropical woods provide generally steady environments, allowing life to flourish. They are the world's hottest and rainiest woods, with temperatures ranging from 68 to 77 degrees Fahrenheit and 79 to 394 inches of rain every year. Tropical woods are famous for their incredible biodiversity. The Amazon jungle, for example, is home to 10% of all known species. The variety of tropical forests makes them particularly effective at nutrient processing. Decomposers swiftly break down dead and rotting materials, which are then promptly taken up by another creature. As a result, tropical forest soils are depleted of nutrients. Many tropical trees have created shallow root systems that stretch throughout the forest floor and can more easily absorb nutrients to deal with weak soils. Many fascinating tropical forest species are on the verge of extinction. The African forest elephant, for example, is classified as Critically Endangered by the IUCN owing to habitat degradation and poaching. Human activities such as logging, land clearance for agriculture, and poaching endanger tropical forests' future. Over 12 million hectares of land will be planted in 2020 alone.

Deforestation continues to be an enduring issue displaying varying rates and intensities over time. The impact of forest loss is twofold: it diminishes biodiversity and ecological services while also contributing to climate change. In the past forests covered 50% of the Earth's surface area. Presently wooded land only accounts for about 30%. Although both human and natural factors contribute to the decline, in forest cover it is primarily actions, known as deforestation that serve as the catalyst. Deforestation is intricately linked to the expansion of agriculture prompted by growing consumer demand. Additionally, subsistence agriculture plays a role in sustaining the livelihoods of inhabitants further exacerbating deforestation. Forest fragmentation often precedes deforestation leading to forest degradation and loss of animal species—a recurring trend in forests. Unsustainable logging practices and forest fires also contribute to the deterioration of forests. The phenomena of forest degradation and deforestation are manifestations of socio-political transformations within societies, including factors like urbanization, commodification, globalization, agricultural intensification and increasingly the environmental repercussions of climate change. Several hypotheses have emerged to explain the mechanisms driving changes in forest cover such, as theories regarding land use spillover and transitions.

The concept of the "forest transition" suggests that forests, within jurisdictions or landscapes may undergo a transition from decline, to growth over a period. However, deforestation poses a difficult and varied dilemma characterized as a "wicked problem." It is complicated by various forces working at different scales and cannot be adequately addressed by singular-oriented remedies. Efforts to stop deforestation in one area may accidentally cause deforestation in another, a process known as leakage. Extensive study has been conducted to better understand the dynamics of forest cover loss and deforestation, as well as the variables that influence them. Furthermore, there is a developing awareness of the reactions and interventions done by both governmental and non-governmental organizations to reduce deforestation. The research ranges from global studies that look at the geographical effect of drivers to examinations that look at the links between consuming and producing countries. Furthermore, a rising number of meta-analyses on deforestation causes have looked at context-specific factors and interactions within specific locations. The methodology and information sources used in this research have gotten increasingly sophisticated, resulting in a better knowledge of the drivers and patterns of deforestation. However, further research is required to measure the efficacy of numerous approaches. Despite advances in understanding the causes of deforestation and the suggested remedies, deforestation remains a persistent problem. It has shifted patterns throughout time and has proven difficult to diminish, halt, and reverse, especially in tropical areas. Maintaining favourable trends in areas where considerable reductions in deforestation have been achieved has been challenging, underlining the necessity for long-term measures. Deforestation reduction success has frequently been ascribed to a mix of measures conducted at multiple levels, necessitating coordination between public and private players.

Deforestation fronts refer to areas that are highly vulnerable to extensive deforestation. A report by WWF in 2015 titled "Saving Forests at Risk" identified 11 deforestation fronts that were projected to account for more than 80% of anticipated deforestation between 2010 and 2030, potentially affecting up to 170 million hectares. In this analysis, a different approach is taken to provide updated insights into the dynamics of these deforestation fronts. The methodology employed for reassessing the deforestation fronts analysis involved a two-step process. Firstly, an Emerging Hotspot Analysis was conducted, which is a commonly used approach to assess forest cover loss and deforestation, particularly at national scales. This analysis utilized hexagons of $10km^2$ within country boundaries and relied on empirical evidence of deforestation derived from Terra-i data spanning from 2004 to 2017. Based on the analysis, 24 active deforestation fronts were identified: nine in Latin America, eight in Africa, and seven in Asia and Oceania. These fronts represent areas with the highest concentration of deforestation. They account for a significant proportion of the total deforestation observed in Latin America, sub-Saharan Africa, Southeast Asia, and Oceania. Comparing the current analysis with the 2015 assessment, several fronts remain in the same locations, such as the Amazon, Central Africa, lower Mekong, and Indonesia. However, many fronts have expanded and encroached further into forested areas. Additionally, new deforestation fronts have emerged in West Africa (e.g., Angola), Madagascar, and Latin America, including the northeastern portion of the Amazon in Guyana and Venezuela, as well as the Maya Forest in Mexico and Guatemala. This highlights the dynamic nature of deforestation and the ongoing challenges in curbing its expansion.

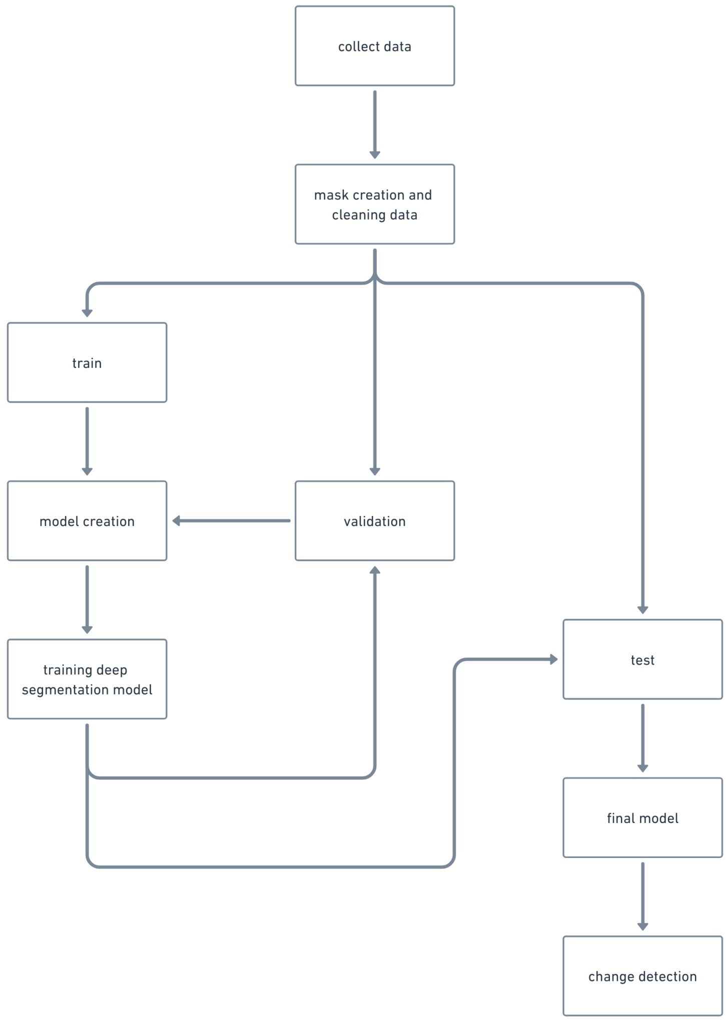
Change detection with deep segmentation

The practice of finding variations in the condition of an item or phenomenon by watching it multiple times in time is known as change detection. It essentially allows us to quantify changes over time using data acquired by Earth-orbiting satellites. Because of the satellites' repeated and regular observation of the Earth's surface, which ensures constant image quality, it is particularly valuable for a variety of applications. Change detection, for example, is critical in analyzing land use changes, monitoring shifting cultivation practices, assessing deforestation, studying changes in vegetation growth patterns, understanding seasonal changes in pasture production, detecting crop stress, monitoring disasters, measuring snow melt, analyzing thermal characteristics during the day and night, and observing other environmental changes. Handling the data manually to identify changes using sequential pictures is an extremely difficult undertaking. However, because most satellite data is digital, it lends itself well to computer-aided analysis, greatly simplifying the process. This capacity enables us to identify and quantify changes throughout time, offering vital insights into a wide range of environmental events and human activities that affect the Earth's surface.

Understanding how cities grow and impact the environment is crucial for policymakers to tackle sustainability challenges and address climate change worldwide. This study focuses on the Yangtze River Delta (YRD) region, in China as an example. It examines the long-term process of urbanization and its effects on the landscape as the carbon emissions and absorption associated with these changes. Additionally, the study explores the factors driving these transformations. Unlike studies, this research takes an approach to analyzing the dynamic process of urbanization and its environmental consequences. It incorporates disciplines, region-specific data and advanced analytical methods. The results reveal that urbanization in the YRD region has undergone three stages: a decline from 1960 to 1978 followed by uneven development from 1978 to 2010. The landscape changes have been rapid, with urban areas expanding into lands and forests. These land use alterations have contributed to increased carbon emissions and absorption. Notably, carbon emissions have substantially increased from 1995 to 2010 in areas leading to higher net carbon intensity. Based on these findings the research proposes policy implications. One way to tackle the issues caused by urbanization and work towards a future is, by making improvements to our social welfare systems. We can also implement policies to conserve our land and control urban sprawl. Additionally promoting reforms that prioritize a low-carbon structure in Jiangsu province can contribute to the overall sustainability of the region. By taking these steps we can address the challenges presented by urbanization protect our environment and move towards a future, for our area.

Change detection (CD) utilizing multitemporal satellite pictures is an important remote sensing application. With technological improvements, we now have access to very high spatial resolution (VHR) photographs with great spatial correlation among pixels. To effectively capture change information, we offer deep change vector analysis (DCVA), a unique unsupervised and context-sensitive framework based on convolutional neural network (CNN) features. DCVA begins with a pre-trained multilayered CNN, which acts as a poor model for extracting deep features capable of modelling spatial connections between surrounding pixels and complicated objects. We use an automated feature selection technique to choose features that emphasize both high and low prior probability change information in order to produce an unsupervised system. These characteristics from many layers are then integrated into a deep feature hypervector, resulting in a multiscale representation of the scene. We may get coherent multitemporal deep feature hypervectors that allow pixel-wise comparisons by utilizing the same pre-trained CNN for semantic segmentation of single pictures. This allows us to build deep change vectors that capture not just changes but also spatial context information. To detect altered pixels, we examine the amplitude of these deep change vectors. Following that, deep change vectors corresponding to the discovered altered pixels are binarized to provide a compressed binary representation that includes information about the direction (type) of change.

Deep learning is a type of machine learning approach that is based on artificial neural networks and representation learning. In comparison to conventional machine learning approaches, deep learning gains significant power and flexibility by describing the world as a layered hierarchy of concepts, with each notion defined by simpler concepts and more abstract representations computed in terms of less abstract ones. We can detect newly built structures in a certain location by analyzing satellite remote sensing picture data, which helps us comprehend land use changes. Various approaches, such as visual object extraction, semantic segmentation, and change detection, are used in this identification process. The difficulty is recognizing changes in remote sensing photos from different time periods and separating newly built structures from existing ones. Traditional mathematical modelling-based approaches are limited in their ability to achieve high recognition accuracy and detection precision. To solve these constraints, researchers used the SegNet neural network, a deep convolutional neural network. SegNet has demonstrated outstanding performance in semantic segmentation tasks for single pictures. However, when directly applied to building change detection, it tends to produce suboptimal results with low accuracy.



The LCZ scheme is a novel categorization system that provides a research foundation for analyzing Urban Heat Islands (UHI) and harmonizing urban temperature measurements throughout the world. Deep learning-based techniques in LCZ mapping have a lot of potential with the emergence of deep learning. Researchers concentrated their investigation on three main cities in China for this study. They created a deep convolutional neural network called Residual combined Squeeze-and-Excitation and Non-local Network (RSNNet) for this project. To identify LCZ using freely accessible Sentinel-1 SAR and Sentinel-2 multispectral images, this architecture contains two critical components: the Squeeze-and-Excitation (SE) block and the non-local block. These two satellite data sets are often utilized in urban research. They ran a series of tests to better understand how Sentinel-1 SAR data influences RSNNet's performance in LCZ mapping. The results showed that integrating SAR data with multispectral data enhances LCZ classification accuracy. This discovery illustrates the potential for applying deep learning and combining various types of satellite data to increase our understanding of urban environments and the accuracy of urban mapping efforts.

Datasets

For the forest cover change segmentation task, the researcher faces a challenge with the existing datasets. The current datasets available for this task comprise only single or bi-temporal images, which are not suitable for volumetric segmentation. Volumetric segmentation involves the analysis of multiple temporal images to detect changes in the forest cover over time accurately. To address this limitation, the researcher plans to create a brand-new dataset tailored specifically for volumetric segmentation. This new dataset will incorporate historical images obtained from Google Earth, NASA World View, and the LandSat database. These sources offer a wealth of satellite imagery captured at various time points, allowing the researcher to obtain multi-temporal data for the segmentation task. The process of creating the new dataset involves gathering a diverse set of satellite images spanning different time periods, ideally covering several years or decades. These images should capture the same geographical regions, enabling a comparative analysis of forest cover changes over time. The historical images from Google Earth and NASA World View, along with the LandSat database, provide a valuable resource for generating such a dataset. Once the new dataset is curated, the researcher can use this data to train a segmentation model capable of analyzing changes in forest cover over time. The volumetric segmentation approach will leverage the multi-temporal information to identify areas of forest growth, deforestation, or any other changes occurring within the forested regions. By creating this specialized dataset, the researcher aims to enhance the accuracy and effectiveness of the forest cover change segmentation task, providing valuable insights into the dynamics of forest landscapes and their transformation over time.

Google Earth Historical Images

When we want to study changes in forest cover over time in specific areas, we need a specialized dataset that captures satellite images from different periods. This allows us to see how forests have evolved over the years. To create this dataset, we use the historical image feature in Google Earth. This feature provides archived satellite images from various time points in the past. So, we start by identifying the regions or areas where we expect to see significant deforestation or forest cover changes.

Next, we use Google Earth to access the images that cover these specific areas at different time periods. These images need to be georeferenced and registered, meaning we align them accurately to a common spatial reference system. This ensures that we can make direct comparisons between the images. Once we have the images, we need to preprocess them by adjusting their size and correcting for any atmospheric effects. This helps to enhance the quality of the images and ensures they are ready for analysis. Now, we organize the preprocessed images into something called "volumetric data cubes." These are like stacks of images, with each cube containing images from different time slots. We include various details about each image, such as the acquisition date and location, for easy reference later. Finally, we use this newly created volumetric dataset to analyze and detect changes in forest cover over time. By doing this, we gain insights into how forests have changed in the identified areas. This information is crucial for understanding deforestation patterns and trends, which can help inform conservation efforts and better forest management practices.

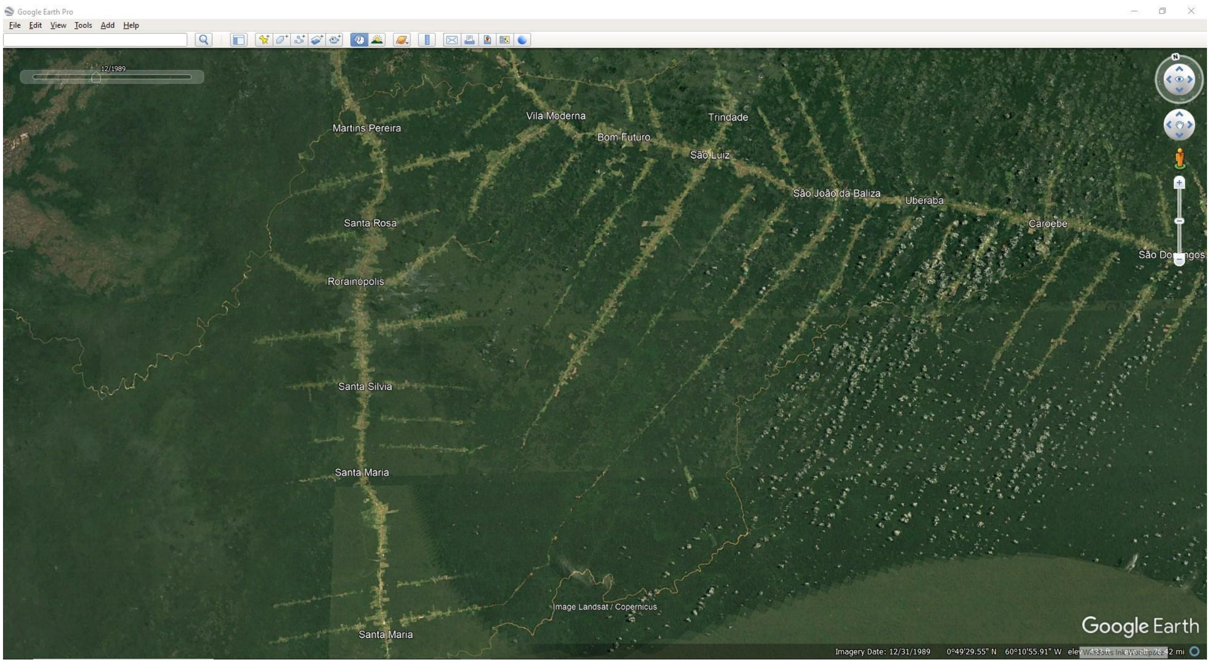


Figure 2: google earth's historical image feature

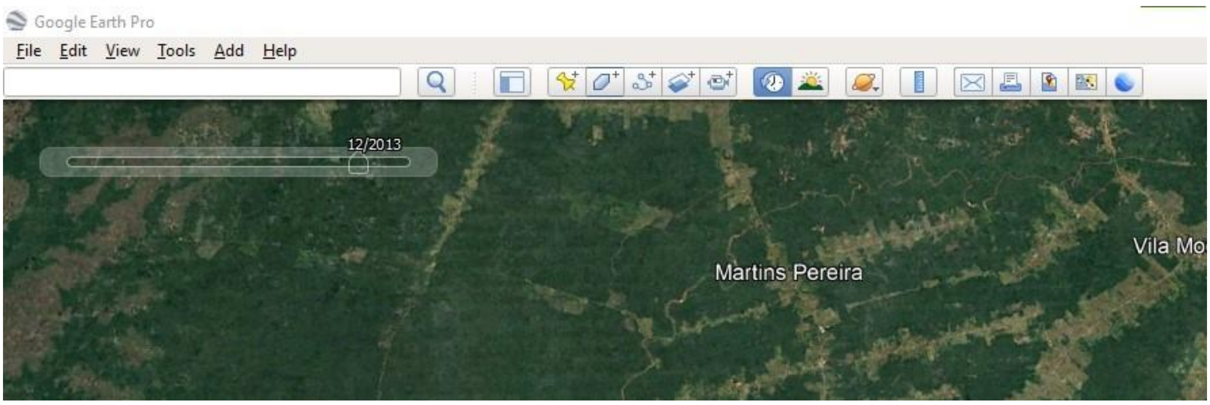


Figure 3: controller to change time slot

NASA EOSDIS World View

In this research project, I'm studying changes in forest cover over time, and I'm using Google Earth as my main source to collect satellite images. Google Earth's historical imagery feature allows me to access archived images taken at different points in the past, which is great for tracking changes in forests over the years. But to ensure that our dataset is complete and accurate, I'm also using two additional sources of data, which we call auxiliary data sources. One of them is NASA's EOSDIS Worldview portal. This portal provides more satellite imagery that complements what we get from Google Earth. You might wonder, why use multiple sources? Well, the main reason is to keep the third dimension of our volumetric data cube unique and consistent across the entire dataset. Think of the data cube as a stack of images, with each layer representing a different time slot. We want to make sure there are no gaps or overlaps in this timeline, so we get a continuous and reliable view of how the forest cover changes over time. By combining data from Google Earth and NASA's EOSDIS Worldview portal, we get a complete picture of the changes happening in the forests we're studying. It helps us see the bigger picture and understand the patterns and trends in forest cover changes more accurately. To know more about NASA's EOSDIS Worldview portal and how it works, you can check out this link: [https://www.earthdata.nasa.gov/worldview]. It's an excellent resource that shows how researchers use satellite imagery to study changes in our environment. using multiple data sources helps us create a comprehensive and detailed dataset, giving us valuable insights into how our forests are changing and how we can better protect and manage them for the future.

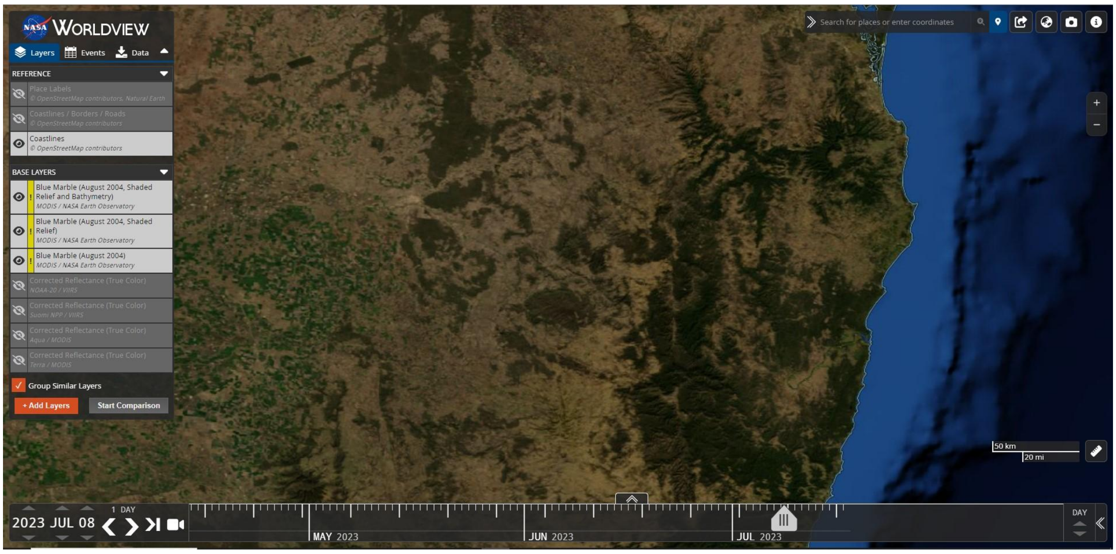


Figure 4: NASA'S EOSDIS web portal

Figure 5: capturing images from the world view portal

Methodology

Inspiration for the research came from the UN's sustainable development goals to protect ecosystems and improve human land usage without disturbing valuable ecosystems. The selection of the research domain and research topic is based on the literature review conducted based on the IEEE Xplore, ACM, Science Direct and also Springer Link scientific databases. The search was directed to deep learning-based methods to detect changes in forest covers and basically change detection on satellite images. Search keywords were semantic segmentation, satellite images, remote sensing, u-net, and change detection. Here are the research query texts I used,

- semantic segmentation AND u-net architecture AND forest
- Keyword:(image segmentation OR semantic segmentation) AND AllField:(u-net) AND AllField:(forests) AND AllField:("remote sensing" OR "satellite images")

First search results gathered and collect all of the paper titles, keywords and abstract. Then the filtering process was started. The first filtering step was done using the paper title. It was assigned to three levels of relevance high, medium and low. The high level represents the most relevant paper titles and the low means, not much relevance or not relevant titles. The second step was to filter high and medium-level relevance papers using their abstracts. Also, assign each paper filtered through the first step into three levels of high, medium and low levels based on abstract relevance.

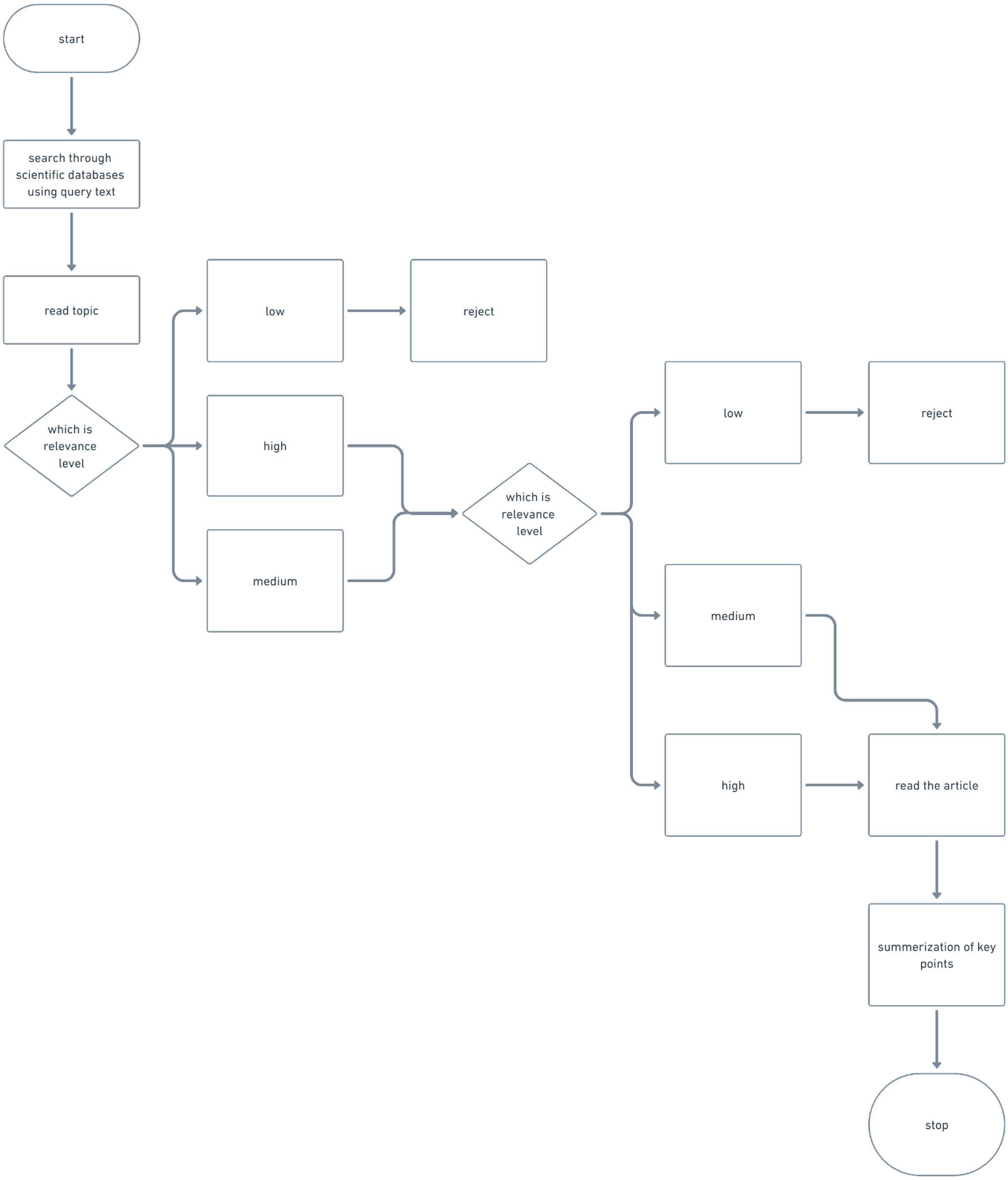


Figure 6: literature review methodology

After both steps only papers have two high-tagged papers choose to proceed. Those selected papers were downloaded and categorised by

- Deep segmentation architecture authors used?
- Type of training image data?
- Focus area of the study?
- Purpose of the study?
- Change detection technique used?

- Dose transfer-learning used?

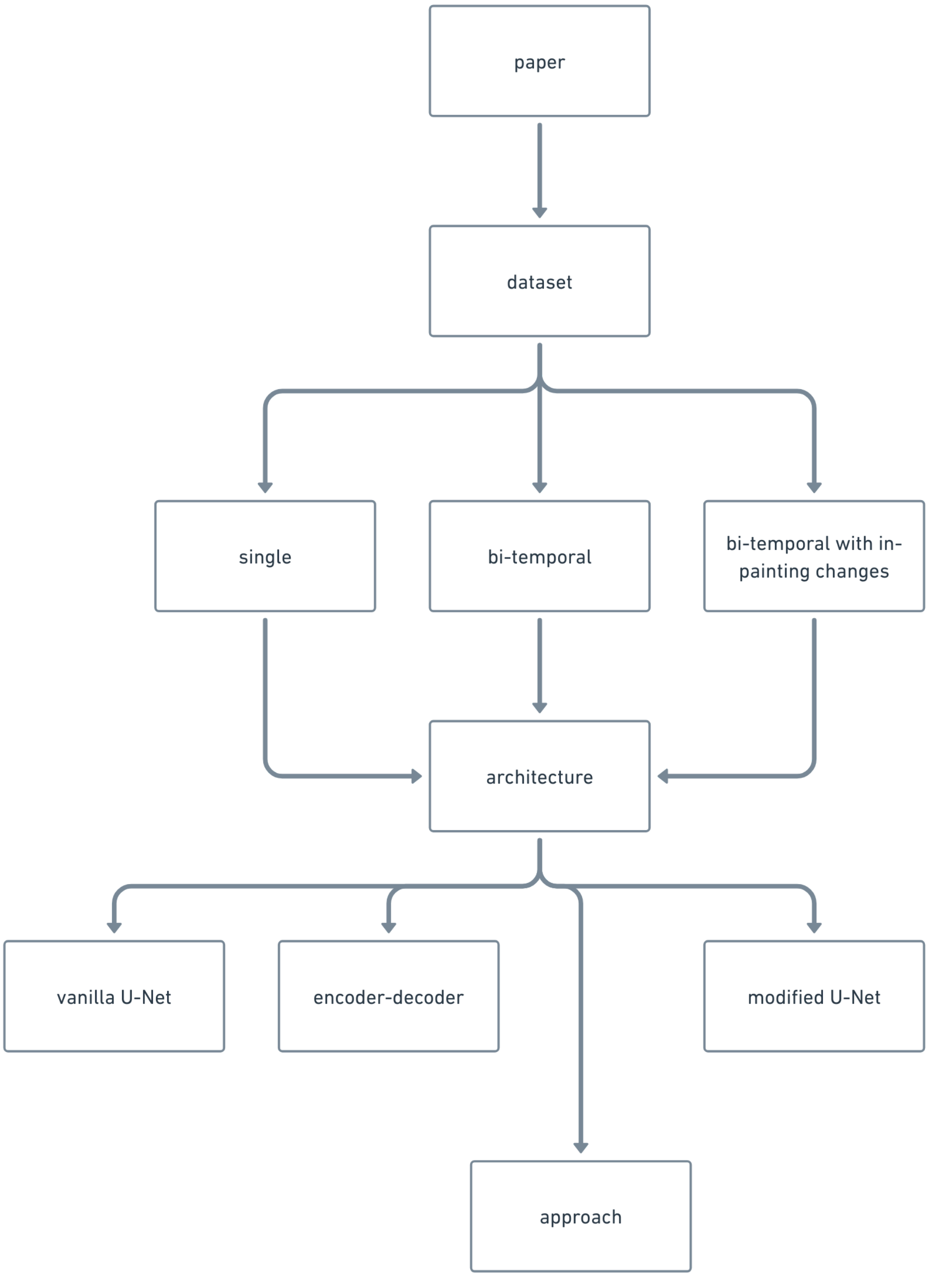


Figure 7: Category of the papers

After all these filtering total of 364 papers were down to 27 most relevant papers.

Survey of forest cover change detection

Sang and Nguyen Duc Minh (2018)[1] developed an approach to areal picture segmentation employing fully convolutional network architecture with ResNet101 as their backbone. In their redesigned network, they additionally used an additional upsampling skip connection. Rather than using traditional RGB (three-channel colour) photographs as input, the proposed models utilised DSM (digital surface model) and nDSM (normalized digital surface model) data. One of their primary goals is to discover a solution to the disparate appearance of artificial and natural items. Hengshuang Zhao et al. (2017) use their pyramid pooling module to investigate the capability of global context information through different-region-based context aggregation. They presented a PSPNet (pyramid scene pooling network). Their goal is to produce high-quality results while analyzing scenes.

Satyam Mohla et al. (2020)[2], AmazonNET is a neural network that extracts burn patterns from multimodal remote sensing photos. The network is made up of UNet, which is a well-known encoder decoder design with skip links that are extensively used in biomedical segmentation. To segment burn scars, the proposed framework employs stacked RGB-NIR channels. Tengfei Bao et al. (2020), The authors presented a feature pyramid module (FPM) and a global attention mechanism module (GAMM) for detecting changes in high-resolution pictures in their research article. The FPM is intended to improve the extraction of semantic information during the feature extraction process, whereas the GAMM is focused on learning image differences. The authors created a dual pyramid attention network (DPANet) specifically for supervised change detection in bi-temporal high-resolution pictures to use these modules. They trained the network with pairs of fixed-size images and obtained a binary pixel-level detection result. The authors proved that their method beats existing deep learning-based approaches and various unsupervised algorithms in terms of change detection accuracy in extended experiments utilizing diverse datasets.

Ying Chen et al. (2019)[3-15], The primary purpose of this work is to detect and localize pixel-level changes in photographs collected at the same location across time. The authors suggest an end-to-end solution based on conditional Generative Adversarial Networks (GANs). They hope to increase the accuracy of the change detection procedure by lowering the difference between the expected and predicted label distributions by using GANs. In particular, the authors employ a conditional GAN network in their research. This network has been trained to generate synthetic images that are similar to the target images. They can discover and emphasize regions where major changes have happened by comparing the generated photos to the actual images. The scientists hope to improve classification results for dense change detection in satellite pictures by using this approach. They concentrate on closing the gap between the expected and predicted distributions of labels, hence enhancing the accuracy and reliability of their change detection model.

David John and Ce Zhang (2022)[16], The authors of this paper constructed and analyzed a deep learning network named Attention U-Net. The Sentinel-2 sensor's satellite imagery was used to perform semantic segmentation. The major goal was to detect instances of deforestation in two South American forest biomes: the Amazon Rainforest and the Atlantic Forest. The authors compared the performance of Attention U-Net to that of many other deep learning models, including U-Net, Residual U-Net, ResNet50-SegNet, and FCN32-VGG16. These models were tested on three separate datasets: the three-band Amazon, the four-band Amazon, and the Atlantic Forest. The authors hoped to determine the most effective strategy for detecting deforestation in the selected forest biomes by examining and comparing the outcomes of various models. To determine the model's capacity to reliably detect deforested areas within satellite data, criteria such as segmentation accuracy, precision, and recall were assessed.

Michael Yeung et al. (Michael Yeung)[17, 18], The Focus U-Net is a new deep neural network developed by the academics behind this work. This network includes a unique module known as the Focus Gate, which integrates spatial and channel-based attention mechanisms. The goal of this module is to promote selective learning of polyp-related aspects. Polyps are abnormal growths in the colon. The researchers made many architectural changes to improve the performance of the Focus U-Net. They added short-range skip connections and deep supervision, which improves information flow inside the network and promotes greater learning. In addition to the network architecture, the researchers introduced the Hybrid Focal loss, a new compound loss function. The purpose of this loss function is to address the problem of class imbalance in image segmentation tasks. It combines aspects from two existing loss functions, the Focal loss and the Focal Tversky loss, to more efficiently address the class imbalance. Overall, this work proposes a comprehensive method that incorporates attention processes, architectural changes, and a customised loss function to improve performance in polyp identification and segmentation tasks. The contributions of the researchers aim to improve the performance and accuracy of medical image analysis systems.

LU XU et al. (2019)[19], The authors of this study present a method for detecting changes in high-resolution remote sensing photos that combine pixel-level and object-level analysis. Their research aims to solve the difficulties of salt and pepper noise and false detections that are widespread in pixel-level and object-level change detection approaches, which can complicate image segmentation. To accomplish this, the scientists combine multidimensional features from high-resolution remote sensing photos with random forest classifiers to classify and identify pixel-level changes. The upgraded U-net network is then used to conduct semantic segmentation on the post-phase remote sensing image, yielding object-level segmentation findings. The next step is to combine the results of pixel-level change detection and object-level segmentation to determine the image's changing and unaffected parts. The scientists want to increase the accuracy and reliability of the change detection procedure by integrating these two levels of analysis. In summary, the authors offer a complete approach for detecting changes in high-resolution remote sensing photos that use both pixel-level and object-level information. Their solution combines feature integration, classifier-based pixel-level analysis, semantic segmentation with an enhanced U-net network, and result fusion to identify changing and unchanged areas in an image.

Kostiantyn Isaienkov et al. (2020)[20], The authors focus on using multitemporal data sources, such as Copernicus Sentinel-2, to improve monitoring capabilities of the Earth's surface and environmental dynamics, notably in forest plantations, in this study. Their mission is to detect deforestation in the forest-steppe zone. To do this, the authors present a basic U-Net model tailored exclusively for deforestation identification. They train and test the model with their own dataset, which was derived from Sentinel-2 imagery of the Kharkiv region in Ukraine, encompassing an area of 31,400 square kilometres. In addition to the baseline model, the authors propose many models that can handle time-dependent images effectively. This means that these models can use numerous sequential photos as input for the segmentation model, allowing for a more in-depth investigation of forest changes over time. The fundamental contribution of this study is the development of a baseline model for detecting forest change in Ukraine, with a special emphasis on the forest-steppe zone. Furthermore, the authors enhance the segmentation model's capabilities by including the capacity to use consecutive photos in the baseline model. Overall, the goal of this study is to enhance the field of deforestation detection by integrating multitemporal data sources and offering an improved model capable of analyzing forest changes in Ukraine's forest-steppe zone.

Katherine James and Karen Bradshaw (2019)[21], The authors of this work offer a novel way to improve the performance of picture segmentation models. They present a weight map-based loss function that tackles the issue of low confidence in the annotation at object edges. They hope to limit the detrimental impact of ambiguous annotations by down-weighting the contribution of these edge pixels to the total loss. The authors experiment with numerous combinations to discover the best design for the weight map. They perform their experiment using an aerial imaging collection focusing on vegetation, with the explicit objective of distinguishing one genus of shrub from other land cover types. The authors compare the performance of their weight map-based loss function to that of the inverse class frequency weighted binary cross-entropy loss. They discover that the weight map-based loss outperforms the binary cross-entropy loss through their study. In fact, it improves the F1 score, a typical criterion for evaluating segmentation models, by 4%. Overall, the author's contribution is to provide a weight map-based loss function that effectively solves the issues of ambiguous annotations at object edges in picture segmentation tasks. Their findings illustrate the superiority of this strategy over typical loss functions, demonstrating its potential to improve the performance and accuracy of segmentation models in a variety of applications.

Fahad Lateef and Yassine Ruichek (2019)[22-34], The writers of this work suggest a detailed survey of numerous methodologies employed in a certain subject. Their primary goal is to give a systematic study of these methods by classifying them into ten distinct classes based on their underlying structures. This classification facilitates a better understanding of the common principles used in various strategies. The authors also provide an overview of the publicly accessible datasets that were utilized to evaluate various approaches. This information provides vital insights into the datasets used for performance evaluation and allows researchers to more effectively compare and duplicate results. The research also looks at the assessment matrix that is typically used to assess the accuracy of various methods. The authors offer insight into the common criteria used to analyze the performance of various methodologies by addressing this evaluation matrix. Furthermore, the authors go into individual methodologies, closely scrutinizing their designs to understand how they accomplished the given results. This examination provides insights into the inner workings of various methods and aids in identifying the critical components that contribute to their effectiveness. Finally, the authors wrap up their survey by emphasizing unresolved issues in the sector and suggesting potential remedies. By resolving these issues, researchers can pave the path for future advances in the field. Overall, this study provides a thorough review that categorizes methodologies, provides an overview of datasets, discusses assessment criteria, examines specific architectures, and tackles open issues. It is a great resource for scholars interested in understanding and improving the topic by providing a comprehensive overview of current research and identifying future paths.

Yan He et al. (2023)[35], The authors presented a data augmentation strategy for increasing the training data for forest remote sensing photos in this research. They hoped to improve the diversity and variability of the training dataset by changing the spatial distribution of these photos, which could lead to better performance of machine learning models. In addition, the authors proposed a segmentation network tailored to high-resolution forest remote-sensing photos. Their network concentrated on two major aspects: extracting multi-scale detailed features and exploiting feature information from satellite photos in the NIR (Near-Infrared) band. To address the first point, the network was designed to capture features at various scales, allowing for the extraction of fine-grained and coarse-grained information from images. This multi-scale feature extraction can help in accurately differentiating different forest sections and capturing the photos' rich characteristics. Concerning the second point, the authors stressed the relevance of adding information from satellite photos' NIR band. The NIR band has useful information on vegetation, and by combining it with the other input data, the network may obtain a better understanding of the forested areas. The scientists hoped to create a segmentation network capable of effectively identifying forest sections in high-resolution remote-sensing pictures by combining these strategies. Their contributions include a data augmentation strategy for spatial distribution modification as well as the merging of multi-scale information to boost the network's performance in forest segmentation tasks.

Yirong Yuan et al. (2023)[36] The authors of this study suggest a model for ultra-high-precision semantic segmentation. The model is divided into three sections: the local branch, the surrounding branch, and the global branch. Each branch is responsible for extracting different types of information from the supplied data. The model is built with a two-level fusion technique to achieve great precision in segmentation outcomes. The local and neighbouring branches collaborate in the low-level fusion process to acquire high-resolution fine structures in the input data. This enables the model to identify intricate details and subtle boundaries in the segmented regions with accuracy. The model gathers global contextual information from downsampled inputs during the high-level fusion phase. This information gives the model a richer knowledge of the input data, allowing it to incorporate larger-scale patterns and context while segmenting. The model can make more informed judgements regarding the boundaries and interactions between distinct semantic regions by including global context. The suggested model achieves ultra-high precision in semantic segmentation due to the integration of the local, surrounding, and global branches, as well as the two-level fusion process. It captures fine-grained minutiae as well as global context, resulting in accurate and exact segmentation results. The original design of the model architecture and the implementation of the two-level fusion mechanism, which fulfils the demand for high precision in semantic segmentation tasks, are the contributions of this work. The uses local, surrounding, and global information.

Muhammad Talha et al[37], the authors propose an enhanced U-Net architecture called Attention Dense U-Net (ADU-Net) for pixel-wise classification of satellite imagery. They introduce dense decoder connections and an attention mechanism to improve the model's performance. Additionally, the authors investigate the impact of different upsampling strategies in the decoder part of the U-Net architecture. To assess the effectiveness of their proposed models, the authors conduct evaluations on the Gaofen Image Dataset (GID). This dataset focuses on land cover classification and comprises five classes: built-up areas, forests, farmlands, meadows, and water bodies. By incorporating dense decoder connections and an attention mechanism, the ADU-Net architecture aims to enhance the accuracy of pixel-wise classification in satellite imagery. The authors' evaluations on the GID dataset provide insights into the performance of their proposed models and allow for comparisons between different upsampling strategies. Ultimately, their work contributes to advancing the field of satellite image analysis and landcover classification.

Junxiao Wang et al[38], the authors introduce a novel model called the Orientation Attention Network (OANet) for accurate segmentation of ground objects in images. The OANet is designed to learn both orientation features and global semantic features to improve the segmentation performance. The authors first propose an Asymmetrical Convolution (AC) technique to capture the directional anisotropy of objects. This technique helps in understanding the orientation characteristics of objects in the image. By incorporating AC, the model gains insights into the preferred directions or orientations of different objects. Additionally, the authors develop an Orientation Attention Module (OAM) to enhance the intrinsic geometric features of objects. The OAM consists of two branches with stacked asymmetrical convolutions along the coordinate axis. This allows the model to selectively focus on features that are most relevant for accurate segmentation. The OAM dynamically adapts and chooses the features that contribute the most to the segmentation task. By combining the AC and OAM, the OANet model can effectively learn orientation features and global semantic features, which are crucial for the precise segmentation of ground objects. This approach improves the model's ability to capture the geometric properties and distinctive characteristics of objects in the image, leading to more accurate segmentation results. The proposed OANet contributes to the advancement of segmentation techniques and can have potential applications in various fields, such as object recognition, scene understanding, and autonomous systems.

Xingjian Gu et al[39], In this study, the researchers explore the combination of Convolutional Neural Networks (CNNs) and Transformers, two powerful models commonly used in computer vision tasks. They propose a novel model called Adaptive Enhanced Swin Transformer with U-Net (AESwin-UNet) specifically designed for remote sensing segmentation. The motivation behind this work lies in the Transformer's ability to capture long-range dependencies in data, which can be beneficial for computer vision tasks. To leverage the strengths of both CNNs and Transformers, the researchers introduce

AESwin-UNet, which adopts a hybrid Transformer-based U-shaped Encoder-Decoder architecture with skip connections. By incorporating the U-Net structure, AESwin-UNet can effectively extract both local and global semantic features from remote sensing data. This allows the model to capture fine-grained details as well as the broader context, enabling accurate segmentation of objects in the images. The hybrid architecture combines the feature extraction capabilities of CNNs with the long-range dependency modelling of Transformers. This synergistic combination enhances the model's performance in remote sensing segmentation tasks, where precise identification and delineation of objects are crucial. Overall, the proposed AESwin-UNet model represents an innovative approach to remote sensing segmentation by leveraging the advantages of both CNNs and Transformers. It provides a promising direction for advancing the field and improving the accuracy of segmentation in remote sensing applications.

In this research[40], the authors introduce a new model called UNetFormer for real-time urban scene segmentation. The model combines the power of Transformers and the popular UNet architecture to achieve efficient and accurate segmentation results. To enable real-time segmentation, the authors focus on designing an efficient decoder using Transformers. The Transformer-based decoder allows the model to effectively capture complex spatial dependencies in the urban scene data. By incorporating Transformers, which are known for their ability to model long-range dependencies, the model can capture both local and global information necessary for accurate segmentation. For the encoder part of the UNetFormer, the authors select a lightweight architecture called ResNet18. This choice ensures computational efficiency while maintaining a good balance between model complexity and performance. To improve the attention mechanism of the decoder, the authors develop an efficient global-local attention mechanism. This mechanism enables the model to consider both global context and local details, ensuring accurate segmentation of urban scenes. By combining the UNet architecture with Transformers, the UNetFormer model offers real-time segmentation capabilities for urban scenes. It effectively leverages the strengths of both architectures, enabling the model to capture complex spatial dependencies and achieve accurate segmentation results. The proposed UNetFormer model represents a significant contribution to the field of urban scene segmentation, providing a promising approach for real-time segmentation with a focus on efficient global-local attention modelling.

Limitations

The shortage of high-quality and adequate satellite image datasets specifically specialized for forest cover change detection and deforestation identification is one of the key issues in this research. Existing datasets may lack the essential temporal resolution, spatial coverage, or ground truth labels for accurate analysis and model training. It is critical to manage and build comprehensive datasets that represent the numerous aspects of forest cover changes at multiple sizes and time intervals, such as deforestation, reforestation, and forest degradation. The availability of such datasets will allow researchers to more effectively create and test algorithms and models, resulting in more accurate and reliable detection of forest cover changes.

Attention techniques have shown tremendous promise in recent years for increasing the performance of deep learning architectures. Integrating attention-based techniques into U-Net designs can aid in the identification of deforestation and forest cover changes by highlighting relevant geographical and temporal elements in satellite data. However, there is a research gap in investigating the applicability of attention-based U-Net architectures in the context of detecting changes in forest cover. Investigating how attention can be efficiently included into the U-Net architecture and how it improves the identification of changes in forest cover over time would lead to more accurate and reliable detection systems.

Another significant research gap is the scarcity of comparison studies evaluating various U-Net architectures, such as attention and focus-based processes, as well as lightweight models in the domains of forest cover change detection and deforestation identification. In terms of computational complexity, memory needs, and accuracy, many U-Net versions and lightweight models offer trade-offs. Understanding the differences in performance between various architectures and models is critical for picking the best method based on the specific restrictions and objectives of forest cover change detection activities. Comparative studies can shed light on each architecture's strengths and drawbacks, providing insights into their efficacy, computing efficiency, and adaptability.

Addressing these research gaps in forest cover change detection and deforestation identification would help to develop more accurate, efficient, and dependable systems of monitoring and managing forest resources. Researchers can make significant advances in the field by acquiring or curating high-quality satellite image datasets, investigating attention-based U-Net architectures with sequential images, and conducting comparative studies of various U-Net architectures and lightweight models.

Conclusion

To conclude the rapid increase, in deforestation rates worldwide caused by both activities and natural factors requires us to implement detection methods to tackle this critical issue. By using methods, authorities, communities and activists can proactively safeguard forests and maintain sustainable land use practices in urban areas. Over the years significant progress has been made in this field. Notable efforts have been focused on improving detection techniques and experimenting with changes to deep networks. However despite these advancements there are still gaps that need to be addressed in research. It is crucial to address these gaps by developing suitable datasets exploring approaches that consider data as a timeline of changes over time periods (multi temporal approaches) and making further improvements to architectural designs. These measures will play a role in strengthening the fight against deforestation and ensuring the long term preservation of our ecosystems. Continuous research and innovation are essential, for achieving forest conservation and sustainable land management practices.

References

[1] D. V. Sang and N. D. Minh, "Fully Residual Convolutional Neural Networks for Aerial Image Segmentation," presented at the Proceedings of the 9th International Symposium on Information and Communication Technology, Danang City, Viet Nam, 2018. [Online]. Available: https://doi.org/10.1145/3287921.3287970.

[2] S. Mohla, S. Mohla, A. Guha, and B. Banerjee, "Multimodal Noisy Segmentation based fragmented burn scars identification in Amazon Rainforest," 2020.

[3] Y. Chen, X. Ouyang, and G. Agam, "ChangeNet: Learning to Detect Changes in Satellite Images," presented at the Proceedings of the 3rd ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery, Chicago, IL, USA, 2019. [Online]. Available: https://doi.org/10.1145/3356471.3365232.

[4] J. Liu, M. Gong, K. Qin, and P. Zhang, "A Deep Convolutional Coupling Network for Change Detection Based on Heterogeneous Optical and Radar Images," *IEEE Transactions on Neural Networks and Learning Systems,* vol. 29, no. 3, pp. 545-559, 2018, doi: 10.1109/TNNLS.2016.2636227.

[5] M. Gong, X. Niu, P. Zhang, and Z. Li, "Generative Adversarial Networks for Change Detection in Multispectral Imagery," *IEEE Geoscience and Remote Sensing Letters,* vol. 14, no. 12, pp. 2310-2314, 2017, doi: 10.1109/LGRS.2017.2762694.

[6] Y. Chen, X. Ouyang, and G. Agam, "MFCNET: End-to-End Approach for Change Detection in Images," 2018.

[7] L. Bruzzone and F. Bovolo, "A Novel Framework for the Design of Change-Detection Systems for Very-High-Resolution Remote Sensing Images," *Proceedings of the IEEE,* vol. 101, no. 3, pp. 609-630, 2013, doi: 10.1109/JPROC.2012.2197169.

[8] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence,* vol. 40, no. 4, pp. 834-848, 2017.

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, 2015: Springer, pp. 234-241.

[10] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784,* 2014.

[11] P. Luc, C. Couprie, S. Chintala, and J. Verbeek, "Semantic segmentation using adversarial networks," *arXiv preprint arXiv:1611.08408,* 2016.

[12] R. Li *et al.*, "DeepUNet: A deep fully convolutional network for pixel-level sea-land segmentation," *IEEE journal of selected topics in applied earth observations and remote sensing,* vol. 11, no. 11, pp. 3954-3962, 2018.

[13] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*, 2015: pmlr, pp. 448-456.

[14] I. Demir *et al.*, "Deepglobe 2018: A challenge to parse the earth through satellite images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 172-181.
[15] J. S. Deng, K. Wang, Y. H. Deng, and G. J. Qi, "PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data," *Int. J. Remote Sens.,* vol. 29, no. 16, pp. 4823–4838, 2008, doi: 10.1080/01431160801950162.
[16] J. David and Z. Ce, "An attention-based U-Net for detecting deforestation within satellite sensor imagery," *International Journal of Applied Earth Observation and Geoinformation,* vol. 107, p. 102685, 2022, doi: https://doi.org/10.1016/j.jag.2022.102685.
[17] Y. Michael, S. Evis, S. Carola-Bibiane, and R. Leonardo, "Focus U-Net: A novel dual attention-gated CNN for polyp segmentation during colonoscopy," *Computers in Biology and Medicine,* vol. 137, p. 104815, 2021, doi: https://doi.org/10.1016/j.compbiomed.2021.104815.
[18] V. Mnih, "Machine learning for aerial image labeling," University of Toronto, 2013.
[19] L. Xu, W. Jing, H. Song, and G. Chen, "High-Resolution Remote Sensing Image Change Detection Combined With Pixel-Level and Object-Level," *IEEE Access,* vol. 7, pp. 78909-78918, 2019, doi: 10.1109/ACCESS.2019.2922839.
[20] K. Isaienkov, M. Yushchuk, V. Khramtsov, and O. Seliverstov, "Deep Learning for Regular Change Detection in Ukrainian Forest Ecosystem With Sentinel-2," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol. 14, pp. 364-376, 2021, doi: 10.1109/JSTARS.2020.3034186.
[21] K. James and K. Bradshaw, "Segmenting objects with indistinct edges, with application to aerial imagery of vegetation," presented at the Proceedings of the South African Institute of Computer Scientists and Information Technologists 2019, Skukuza, South Africa, 2019. [Online]. Available: https://doi.org/10.1145/3351108.3351124.
[22] L. Fahad and R. Yassine, "Survey on semantic segmentation using deep learning techniques," *Neurocomputing,* vol. 338, pp. 321-348, 2019, doi: https://doi.org/10.1016/j.neucom.2019.02.003.
[23] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," 2015. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2015/file/14bfa6bb14875e45bba028a21ed38046-Paper.pdf.
[24] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8759-8768.
[25] A. Salvador *et al.*, "Recurrent neural networks for semantic instance segmentation," *arXiv preprint arXiv:1712.00617,* 2017.
[26] H. Fan, X. Mei, D. Prokhorov, and H. Ling, "Multi-level contextual rnns with attention model for scene labeling," *IEEE Transactions on Intelligent Transportation Systems,* vol. 19, no. 11, pp. 3475-3485, 2018.
[27] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," 2015.

[28] G. Nanfack, A. Elhassouny, and R. O. H. Thami, "Squeeze-SegNet: a new fast deep convolutional neural network for semantic segmentation," in *Tenth International Conference on Machine Vision (ICMV 2017)*, 2018, vol. 10696: SPIE, pp. 703-710.
[29] P. Bilinski and V. Prisacariu, "Dense Decoder Shortcut Connections for Single-Pass Semantic Segmentation," presented at the The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June, 2018.
[30] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801-818.
[31] D. M. Vo and S.-W. Lee, "Semantic image segmentation using fully convolutional neural networks with multi-scale images and multi-scale dilated convolutions," *Multimedia Tools and Applications,* vol. 77, no. 14, pp. 18689-18707, 2018/07/01 2018, doi: 10.1007/s11042-018-5653-x.
[32] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, "Enet: A deep neural network architecture for real-time semantic segmentation," *arXiv preprint arXiv:1606.02147,* 2016.
[33] T. Pohlen, A. Hermans, M. Mathias, and B. Leibe, "Full-Resolution Residual Networks for Semantic Segmentation in Street Scenes," 2017.
[34] L. Ke *et al.*, "Segment Anything in High Quality," *arXiv preprint arXiv:2306.01567,* 2023.
[35] Y. He, K. Jia, and Z. Wei, "Improvements in Forest Segmentation Accuracy Using a New Deep Learning Architecture and Data Augmentation Technique," *Remote Sensing,* vol. 15, no. 9, p. 2412, 2023. [Online]. Available: https://www.mdpi.com/2072-4292/15/9/2412.
[36] Y. Yuan, J. Cui, Y. Liu, and B. Wu, "A Multi-Step Fusion Network for Semantic Segmentation of High-Resolution Aerial Images," *Sensors,* vol. 23, no. 11, p. 5323, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/11/5323.
[37] T. Muhammad, A. B. Farrukh, G. Sajid, and Z. Hamza, "ADU-Net: Semantic segmentation of satellite imagery for land cover classification," *Advances in Space Research,* 2023, doi: https://doi.org/10.1016/j.asr.2023.05.007.
[38] W. Junxiao, F. Zhixi, J. Yao, Y. Shuyuan, and M. Huixiao, "Orientation Attention Network for semantic segmentation of remote sensing images," *Knowledge-Based Systems,* vol. 267, p. 110415, 2023, doi: https://doi.org/10.1016/j.knosys.2023.110415.
[39] G. Xingjian, L. Sizhe, R. Shougang, Z. Hengbiao, F. Chengcheng, and X. Huanliang, "Adaptive enhanced swin transformer with U-net for remote sensing image segmentation," *Computers and Electrical Engineering,* vol. 102, p. 108223, 2022, doi: https://doi.org/10.1016/j.compeleceng.2022.108223.
[40] W. Libo *et al.*, "UNetFormer: A UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery," *ISPRS Journal of Photogrammetry and Remote Sensing,* vol. 190, pp. 196-214, 2022, doi: https://doi.org/10.1016/j.isprsjprs.2022.06.008.