Proceedings of the 19th Conference on Computer Science and Intelligence Systems (FedCSIS) pp. 141–148
DOI: 10.15439/2024F4228
ISSN 2300-5963 ACSIS, Vol. 39

d'Alembert Convolution for Enhanced Spatio-Temporal Analysis of Forest Ecosystems

Rytis Maskeliūnas
CoE Forest 4.0
Vytautas Magnus University
Kaunas, Lithuania
rytis.maskeliunas@vdu.lt

Robertas Damaševičius
CoE Forest 4.0
Kaunas University of Techology
Kaunas, Lithuania
robertas.damasevicius@ktu.lt

***Abstract*—This paper presents a novel approach to enhance the spatio-temporal analysis of forest ecosystems using the d'Alembert convolution method, which, integrating elements from wave equation theory and convolutional neural networks, enables the comprehensive analysis of remote sensing images by capturing both spatial and temporal variations. This methodology not only improves feature extraction, but also helps address the challenges associated with traditional image processing techniques, which often overlook the temporal dynamics of forests. The results show significant improvements in the analysis of forest ecosystems. Specifically, the higher performance metrics compared to existing methods, including higher accuracy in classifying various forest types and more effective monitoring of changes over time.**

***Index Terms*—Convolutional Neural Networks, Spatio-Temporal Analysis, Remote Sensing, d'Alembert Operator.**

I. Introduction

THE IMPORTANCE of forests in global environmental health, biodiversity conservation, and economic resources is undeniable. Forests play a crucial role in carbon sequestration, climate regulation, and providing habitats for a variety of species [1]. However, they are constantly under threat from various factors, including climate change, deforestation, pests, and diseases [2]. Traditional forest management methods, often relying on periodic and manual surveys, are inadequate in the face of these rapidly evolving challenges [3]. The need for more efficient, accurate, and real-time monitoring and management methods is more pressing than ever [4].

Forestry management is entering a new era of technological innovation, marked by the integration of advanced computational methods and environmental science. The advent of smart forestry, using data-driven approaches, has opened new pathways for sustainable forest management and environmental conservation [5]. Advancements in remote sensing technologies, such as satellite imagery and aerial photography, have propelled the field of forestry into the digital age [6]. Hyperspectral imaging, in particular, has become a valuable tool for monitoring vegetation health, biomass estimation, and detecting changes within forest ecosystems [7]. However, the sheer volume and complexity of the generated data pose significant challenges in terms of processing and analysis [8]. Conventional image processing techniques, while beneficial, often fail in extracting the full spectrum of information hidden within multidimensional spatial and temporal data sets [9]. However, the complexity and dynamic nature of forest ecosystems pose great challenges in terms of data collection, analysis, and interpretation [10], one of the challenges still present in smart forest, as outlined in [11].

Convolutional neural networks (CNNs) have revolutionized the field of image analysis [12]. However, their application in forestry has been somewhat limited, focusing mainly on spatial data without fully exploiting the temporal dimension [13]. By extending the convolution operation to incorporate differential operators that account for both spatial and temporal changes, similar to the components of the d'Alembert operator, we propose a novel method that not only enhances feature extraction from remote sensing images, but also captures the dynamic changes occurring within forest ecosystems over time by adapting the convolution operation to include differential operators that account for spatial and temporal changes, this method allows for a more nuanced extraction of features from remote sensing images, surpassing the capabilities of traditional convolutional methods. The approach allows for an increased level of detail in analyzing both spatial and temporal variations in forest ecosystems, contributing to a deeper understanding of forest dynamics.

II. Related Works

Existing research has demonstrated a robust exploration of remote sensing applications, taking advantage of advanced machine learning and deep learning techniques to address a spectrum of challenges in land use and cover change (LULC), environmental monitoring and resource management. Several researchers utilize machine learning models to analyze and interpret LULC changes and their impacts on ecosystems. For example, Saha et al. (2024) employ geospatial techniques and machine learning to assess the degradation of the Deepor wetland in India, highlighting high losses due to urbanization and agricultural expansion [14]. Similarly, Thien et al. (2023) examined the spatiotemporal dynamics of LULC in Vietnam's Red River delta, attributing changes predominantly to urban development [15]. In a broader scope, Masolele et al. (2021) deploy spatial and temporal deep learning methods to classify land use after tropical deforestation, underscoring the supe-

rior performance of spatio-temporal models over conventional approaches [16]. Maretto et al. (2021) further this discourse by mapping deforestation in the Amazon, demonstrating how spatio-temporal deep learning improves monitoring accuracy [17]. Others focused on the development of sophisticated machine learning algorithms to refine remote sensing data retrieval and analysis. Fonseca et al. (2023) innovate in multi-temporal SAR image analysis through wavelet spatio-temporal change detection, achieving high accuracy with reduced computational demands [18]. On a similar note, Dimyati et al. (2023) and Jing et al. (2023) introduce methods to monitor mangrove changes and combine remote sensing images, respectively, showcasing the potential of these advanced techniques in managing complex environmental datasets [19], [20].

III. Theoretical Foundations

A. Convolutional Neural Networks in Image Processing

Convolutional Neural Networks (CNNs) have revolutionized the field of image processing and computer vision. They are a specialized kind of neural network designed for processing data with a grid-like topology, such as images. A CNN learns to recognize patterns and features in images through the process of convolution, pooling, and fully connected layers.

The convolution of a function f with a kernel g is defined as:

$$(F * G)(x, y) = \sum_{i=-a}^{a} \sum_{j=-b}^{b} F(i, j) \cdot G(x - i, y - j) \quad (1)$$

where F is the image, G is the kernel, and x, y are spatial coordinates in the image, and a and b represent the half-width and half-height of the kernel G, respectively. The kernel G slides over the image F, computing the sum of element-wise products at each position.

CNN architecture has several types of layers. Lower layers capture basic features such as edges and textures, while deeper layers identify complex patterns specific to the training data:

- **Convolutional Layer** performs the convolution operation. It applies a set of learnable filters (kernels) to the input image. Each filter extracts different features from the input.

$$C_{out} = \text{ReLU}(C_{in} * K + b) \quad (2)$$

  where C_{in} is the input, K is the convolutional kernel, b is a bias term, and ReLU is the activation function, typically a Rectified Linear Unit.
- **Pooling Layer** reduces the spatial size (height and width) of the input volume, making the network computation more efficient. It also helps to make the network invariant to small translations of the input.
- **Fully Connected Layer** has connections to all activations in the previous layer. These layers are typically used at the end of the network to perform classification based on the features extracted by the convolutional and pooling layers.

B. The d'Alembert Operator in Wave Equations

The d'Alembert operator is a second-order differential operator that is also widely used in the fields of physics and engineering, particularly in the study of wave propagation and vibrations. The d'Alembert operator is defined in the context of a four-dimensional space-time continuum, combining time and space derivatives. In a three-dimensional space with time, the operator is represented as:

$$\Box = \frac{\partial^2}{\partial t^2} - \nabla^2 \quad (3)$$

where $\frac{\partial^2}{\partial t^2}$ is the second derivative with respect to time, and ∇^2 is the Laplacian operator, which is a scalar differential operator defined as the divergence of the gradient of a function, representing the sum of second spatial derivatives:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \quad (4)$$

The wave equation is particularly suitable for analyzing spatiotemporal data in remote sensing images because it inherently captures the propagation of information over both space and time, which aligns with the dynamic nature of environmental phenomena. It combines the second-order temporal derivative with the Laplacian operator, effectively linking temporal changes to spatial variations. This dual capability allows the wave equation to model how disturbances, such as changes in vegetation or land cover, propagate through time and space, providing a comprehensive framework for tracking these dynamics. By incorporating the d'Alembert operator into convolutional analysis, the network can utilize these properties to enhance feature extraction, capturing both the immediate spatial details and their evolution over time. This results in a more robust analysis of remote sensing data, as it allows for the detection and interpretation of complex temporal patterns and spatial structures within the forest ecosystem, essential for accurate monitoring and assessment. The wave equation for a scalar field $\psi(x, y, z, t)$ in a three-dimensional space can be written as:

$$\Box\psi = \frac{\partial^2 \psi}{\partial t^2} - \nabla^2 \psi = 0 \quad (5)$$

This equation describes how a wave propagates in space and time. The term $\frac{\partial^2 \psi}{\partial t^2}$ represents the acceleration of the wave, while $\nabla^2 \psi$ accounts for the spatial spread of the wave.

C. d'Alembert Operator for Convolutional Analysis

CNNs handle spatial data by applying convolutional filters to extract features such as edges, textures, and patterns from static images. These filters are applied across the spatial dimensions (height and width) of the image, capturing local spatial hierarchies and invariances. Temporal processing, on the other hand, works by capturing the changes and dynamics over time, which is critical for understanding phenomena like forest growth or seasonal variations in remote sensing data. To achieve this, the network incorporates layers or mechanisms that can capture temporal dependencies - the d'Alembert operator, which, traditionally used to describe wave propagation,

is adapted to account for both spatial and temporal changes by modifying the convolution operation to include differential components. By using a kernel enhanced with the d'Alembert operator, the network simultaneously processes the spatial features (through the traditional convolution) and the temporal features (through the operator's temporal derivative component). Threfor, this integration allows the network to learn more comprehensive feature representations that encapsulate both static and dynamic aspects of the data, providing a robust and versatile framework for tasks requiring spatio-temporal analysis. The balance is achieved by including both standard convolutional layers for spatial processing and modified layers to handle temporal dynamics effectively.

To integrate the d'Alembert operator into the convolution process, we first redefine the convolution operation to include differential components. Consider a remote sensing image sequence represented as $I(x, y, t)$, where (x, y) are spatial coordinates, and t is the time dimension. The adapted convolution operation, incorporating the d'Alembert operator, can be mathematically represented as:

$$C_{\text{d'Alembert}}(x, y, t) = (K * (\Box I))(x, y, t) \tag{6}$$

where K is the convolution kernel, $*$ denotes the convolution operation, and $\Box$ is the d'Alembert operator. The d'Alembert operator applied to the image sequence I is defined as:

$$\Box I(x, y, t) = \frac{\partial^2 I}{\partial t^2}(x, y, t) - \nabla^2 I(x, y, t) \tag{7}$$

The Laplacian component $\nabla^2 I(x, y, t)$ of the d'Alembert operator enhances the extraction of spatial features by emphasizing areas with high spatial frequency, such as edges and textures in the image.

The temporal derivative component $\frac{\partial^2 I}{\partial t^2}(x, y, t)$ captures changes in the image sequence over time, highlighting the dynamic changes in the forest environment.

The integration of the d'Alembert convolution into a CNN framework necessitates a modification of the standard convolutional layers. This modification involves applying a kernel that is enhanced with the d'Alembert operator, enabling the network to capture both spatial and temporal variations more effectively. The d'Alembert-enhanced kernel is designed to incorporate both the spatial features, captured by the traditional convolution kernel, and the temporal features, introduced by the d'Alembert operator. Consider a standard convolution kernel K and its adaptation with the d'Alembert operator.

$$K_{\text{d'Alembert}}(x, y, t) = K(x, y) + \lambda \cdot (\Box I)(x, y, t) \tag{8}$$

Here, $K(x, y)$ is the standard convolution kernel, $\Box I$ represents the application of the d'Alembert operator on the image sequence I, and λ is a weighting factor that balances the spatial and temporal components.

The convolution operation in a CNN is modified to use this d'Alembert-enhanced kernel. The modified convolution operation for an input image sequence I at layer l is:

$$C^l_{\text{d'Alembert}}(x, y, t) = (K^l_{\text{d'Alembert}} * I^l)(x, y, t) \tag{9}$$

where $*$ denotes the convolution operation, and l indicates the layer in the CNN.

The d'Alembert convolution allows the CNN to extract features that encapsulate both spatial variations (such as edges, textures) and temporal changes (such as growth patterns, environmental dynamics). This dual capability is required for the analysis of remote sensing images of forests, where both spatial and temporal indicators are required to understand forest health and dynamics.

$$F^l_{\text{d'Alembert}} = \text{Activation}(C^l_{\text{d'Alembert}}(x, y, t)) \tag{10}$$

Here, $F^l_{\text{d'Alembert}}$ represents the feature maps obtained after applying the d'Alembert convolution at layer l, and Activation denotes the activation function used in the CNN (e.g., ReLU).

D. Architecture of d'Alembert Network

The architecture of the d'Alembert network is presented in Table I and is discussed in detail below.

TABLE I
ARCHITECTURE OF THE D'ALEMBERT NETWORK

Layer Type	**Output Size**	**Kernel Size**	**Other Parameters**
Input Layer	480x480x3	-	-
Convolutional Layer	470x470x32	11x11	Stride=1, Padding=Valid
Activation Layer	470x470x32	-	ReLU
Pooling Layer	235x235x32	2x2	Stride=2, Type=Max
d'Alembert Conv Layer	225x225x64	11x11	Stride=1, Padding=Valid, $\lambda = 0.1$
Activation Layer	225x225x64	-	ReLU
Pooling Layer	112x112x64	2x2	Stride=2, Type=Max
Fully Connected Layer	1024	-	-
Activation Layer	1024	-	ReLU
Fully Connected Layer	512	-	-
Activation Layer	512	-	ReLU
Output Layer	7	-	Softmax

The basis of the d'Alembert network is of convolutional layers for extracting spatial features from imagery. Each convolutional layer applies a set of learnable filters to the input image, detecting features such as edges, textures, and shapes. These features are needed for the structural components of the forest, such as the canopy density and tree boundaries. The convolution operation combines image data with a kernel (filter) through a dot product that aggregates local pixel values to produce a feature map, highlighting areas of interest in the image. Following each convolutional layer, an activation function is used (the Rectified Linear Unit (ReLU)). ReLU introduces non-linearity into the model, allowing it to learn more complex patterns. It works by replacing all negative pixel values in the feature map with zero, maintaining only positive values that correspond to detected features, and it helps to overcome the problem of vanishing gradients, ensuring that the network continues to learn effectively throughout its depth. The inclusion of pooling layers (max pooling), reduces the spatial size of the representation, making the computation more manageable, and the network more robust to variations

in the image. By downsampling the feature maps, the pooling layers help reduce the amount of data that needs to be processed while preserving the most essential information, such as the dominant features within a local patch of the image. Which is particularly useful in forest imagery, where specific features, such as tree clusters and clearings, need to be emphasized over large, uniform areas.

The main novelty in our d'Alembert network is the adaptation of the d'Alembert operator into the convolution process. This operator is applied here to account for both spatial and temporal changes in forest imagery by modifying the convolution operation to include differential operators (Laplacian for spatial and second-order time derivative for temporal features). In this way, the network can capture dynamic changes in the forest, such as growth, deforestation, or seasonal variations. Toward the end of the network, fully connected layers are used to interpret the features extracted and learned by the convolutional and pooling layers. These layers consolidate the learned features into a format suitable for classification or regression tasks, such as identifying different types of forests or assessing forest health. Each neuron in these layers connects to all activations in the previous layer, allowing the network to learn non-linear combinations of the high-level features. The final layer of the network (softmax layer) outputs a probability distribution over the target classes. For forest imagery analysis, these classes include different types of land cover, such as dense forest, degraded forest, water bodies, etc. The softmax function converts the logits from the fully connected layer into probabilities by exponentiating and normalizing each output, providing a clear, interpretable classification result.

The performance sensitivity to hyperparameters such as learning rate, batch size, and the number of convolutional layers is unfortunatly quite significant in our approach. The learning rate strongly affects the model's ability to effectively integrate both spatial and temporal features; a high learning rate lead to suboptimal convergence in the complex landscape of spatio-temporal data, while a low rate cause excessively slow training, missing critical temporal patterns. Batch size influences the network's capacity to generalize from dynamic forest data; larger batch sizes provide more stable gradient estimates, improving convergence and capturing broader temporal changes, but at the cost of higher memory usage. Conversely, smaller batches enhance generalization but introduce noisy gradients, potentially destabilizing training. The number of convolutional layers directly impacts the depth of spatial feature extraction; insufficient layers fail to capture the intricate textures and edges within forest images, while too many layers could overfit the spatial details and neglect temporal dynamics. Hyperparameter tuning is therefor required to accurately detect and analyze both spatial and temporal variations essential for monitoring forest ecosystems.

IV. Case Study and Experimental Results

A. Datasets

The DeepGlobe Land Cover 2018 dataset [21] is a collection of high-resolution satellite images used for land cover classification challenges, focusing on categorizing land cover into multiple classes. It encompasses geographical landscapes from different parts of the world, offering a robust platform to advance land cover analysis technologies. The dataset consists of 1146 images with 3042 labeled objects belonging to 7 different classes including agriculture_land, urban_land, rangeland, water, barren_land, forest_land, and unknown (see sample images in Figure 1). In this study, we used a subset of the DeepGlobe dataset, DeepGlobe-Forest, which includes only 191 images labeled as forest_land.

The LoveDA [22], [23] dataset is a remote sensing dataset adapted for the study of natural landscapes and their dynamic changes. It consists of multispectral imagery collected from various satellite platforms. The images in the dataset have high spatial resolution, which aids in detailed analysis and facilitates accurate monitoring of small-scale changes. It includes a mix of urban and rural landscapes for a diverse range of scenes. The dataset consists of 5987 images with 20658 labeled objects belonging to seven different classes, including background, road, building, forest, water, agriculture, and barren (see sample images in Figure 2). We used a subset of the LoveDA dataset, called LoveDA-Forest, which has 3043 images labeled as forest. Both the DeepGlobe-Forest and LoveDA-Forest datasets have been used previously in [24]. Preprocessing involved resizing images to a consistent dimension, normalizing pixel values to a standard scale, and augmenting the data with transformations such as rotations and flips to enhance model generalization. Additionally, temporal alignment of sequential images was ensured for the LoveDA dataset to capture temporal dynamics accurately.

TABLE II
Comparison of DeepGlobe-Forest and Loveda-Forest datasets

Characteristic	DeepGlobe-Forest	Loveda-Forest
Sensor	DigitalGlobe	Sentinel-2
Image Size	2448x 2448	1024 x 1024
Spectral Range (μm)	0.4 - 2.3	0.45 - 2.4
Number of Bands	11	13
Spatial Resolution (m)	30	10
Number of Classes	5	7

B. Experimental setting

For this study, the analysis was performed using a custom implementation developed in TensorFlow. The model training used the DeepGlobe-Forest and Loveda-Forest datasets in 200,000 epochs, starting with an initial learning rate of 0.00005. This rate was progressively reduced starting from the 2000th epoch using a linear decay strategy. The input images were cropped to a uniform size of 480×480 for consistency in both datasets. The computational resources included a single NVIDIA RTX 2060 graphics card, an AMD Ryzen 9 5950X CPU and 32 GB of RAM.

C. Performance evaluation

We used performance metrics such as overall accuracy (OA), average accuracy (AA) per class, and the Kappa coefficient (Kappa). The OA metric reflects the proportion of

Fig. 1. Sample images from DeepGlobe dataset: water, barren_land, forest_land, urban_land, and agriculture_land

Fig. 2. Sample images from LoveDA dataset: forest, barren, water, agriculture, background, road, and building

correctly classified images in the test dataset relative to the total sample count. The AA metric represents the average accuracy in each image class, while the Kappa metric provides a measure of accuracy adjusted for the probability of random chance. In addition, precision, recall, and F-1 score were also utilized as performance indicators. The results are summarized in table III.

TABLE III
PERFORMANCE METRICS FOR DEEPGLOBE-FOREST AND LOVEDA-FOREST DATASETS

OA (%)	AA (%)	Kappa	Precision(%)	Recall(%)	F1 Score(%)
DeepGlobe-Forest					
91.19	83.89	0.88	92.13	90.24	91.17
LoveDA-Forest					
86.89	74.34	0.84	87.31	85.45	88.37

For the DeepGlobe-Forest dataset, achieving an OA of 91.19% indicates high accuracy in classifying forest cover types, supported by an AA of 83.89% per class and a robust Kappa coefficient of 0.88, demonstrating strong statistical result reliability. Precision scores around 92.13% and balanced recall rates of 90.24% with a high F1 score of 91.17%, show that the model is able to accurately identify and differentiate forest categories. Similarly, the LoveDA-Forest dataset shows an OA of 86.89%, with precise classification reflected in precision and recall scores of 87.31% and 85.45%, respectively, and an F1 score of 88.37%.

D. Results of the segmentation performance

The segmentation performance of the models was evaluated using two metrics: Intersection over Union (IoU) and Accuracy (Acc):

$$\text{IoU} = \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}},$$

where p_{ij} denotes the prediction of the category i into category j, and $k+1$ is the total number of categories. The mean Intersection over Union (mIoU) is calculated by:

$$mIoU = \frac{1}{k+1} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}},$$

The formulas for the accuracy and mean accuracy (mAcc) are given by:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN},$$

$$mAcc = \frac{1}{k+1} \sum_{i=0}^{k} \frac{TP_i + TN_i}{TP_i + FP_i + FN_i + TN_i},$$

where TP represents the true positives, TN the true negatives, FP the false positives, FN the false negatives, and $k+1$ indicates the total number of categories.

For comparison we include the values of deeplabv3+ [25], pidnet [26], pspnet [27], knet [28], segformer [29], mask2former [30] and segnext [31] as was determined in the research of Wang et al. [24] (we have not replicated these methods in our work).

TABLE IV
COMPARISON OF MODEL PERFORMANCES ON FOREST AND BACKGROUND CLASSES FROM DEEPGLOBE-FOREST DATASET [24].

Model	IoU (%)		mIoU (%)		Accuracy (%)	
	Forest	Back-ground	Forest	Back-ground	Forest	Back-ground
Deeplabv3+	77.69	79.67	78.68	87.42	88.70	88.06
Segformer	79.98	81.71	80.85	89.02	89.80	89.41
Pidnet-s	78.78	80.96	79.87	87.35	90.21	88.78
Mask2former	80.52	81.61	81.06	91.09	88.17	89.63
Pspnet	79.86	80.79	80.33	91.17	87.22	89.20
Segnext	80.60	81.84	81.22	90.69	88.71	89.70
Knet-s3-r50	80.23	81.22	80.73	91.24	87.63	89.44
SegForest	82.80	83.99	83.39	91.79	90.20	91.00
d'Alembert Network	83.19	84.37	83.89	92.23	90.86	91.52

TABLE V
PERFORMANCE OF MODELS ON LOVEDA-FOREST DATASET [24].

Model	IoU (%)		mIoU (%)		Accuracy (%)	
	Forest	Back-ground	Forest	Back-ground	Forest	Back-ground
Deeplabv3+	64.22	75.88	70.05	80.37	84.85	82.61
Segformer	64.63	76.31	70.47	80.36	85.35	82.86
Pidnet-s	64.36	74.08	69.22	84.82	80.85	82.84
Mask2former	65.67	76.83	71.25	81.69	85.30	83.50
Pspnet	62.68	77.08	69.88	73.93	89.18	81.56
Segnext	64.42	76.96	70.69	78.31	87.01	82.66
Knet-s3-r50	65.99	76.16	71.08	84.11	83.45	83.78
SegForest	68.38	79.04	73.71	82.98	87.14	85.06
d'Alembert Network	69.12	80.05	74.34	83.28	87.93	85.87

For the DeepGlobe-Forest dataset, the d'Alembert Network exhibited the highest metrics in all categories: achieving Intersection over Union (IoU) scores of 83.19% for forest and 84.37% for background, mean IoU (mIoU) of 83.89% and 92.23% respectively, and accuracy scores of 90. 86% and 91. 52%, respectively. These results not only improve upon other advanced models such as SegForest, Pspnet, and Mask2former but also emphasize the network's ability to finely discriminate between forest and non-forest regions, capturing both varying textures of forest landscapes and clear delimitations of background areas. Similarly, on the LoveDA-Forest dataset, the d'Alembert Network again outperformed competing models, recording the highest IoU for the forest at 69. 12% and for the background at 80. 05%. It also achieved the highest mIoU scores of 74.34% for forest and 83.28% for background, along with accuracy figures of 87. 93% and 85. 87%, respectively, showing the model's robustness and its enhanced capability in processing and analyzing remote sensing imagery with high precision, particularly in diverse and dynamic environmental settings.

E. Ablation study

The overly small spatial size of the input image patch leads to a significant loss of important information due to an inadequate receptive field. Conversely, an excessively large spatial size of the input image patch introduces many noisy pixels and suffers from inter-class contamination. Therefore, we have established the spatial size of the input image patch within the range of 5×5, 7×7, 9×9, 11×11, 13×13, 15×15 to evaluate the classification performance across various spatial dimensions. The classification results for two data sets are shown in Figure 3.

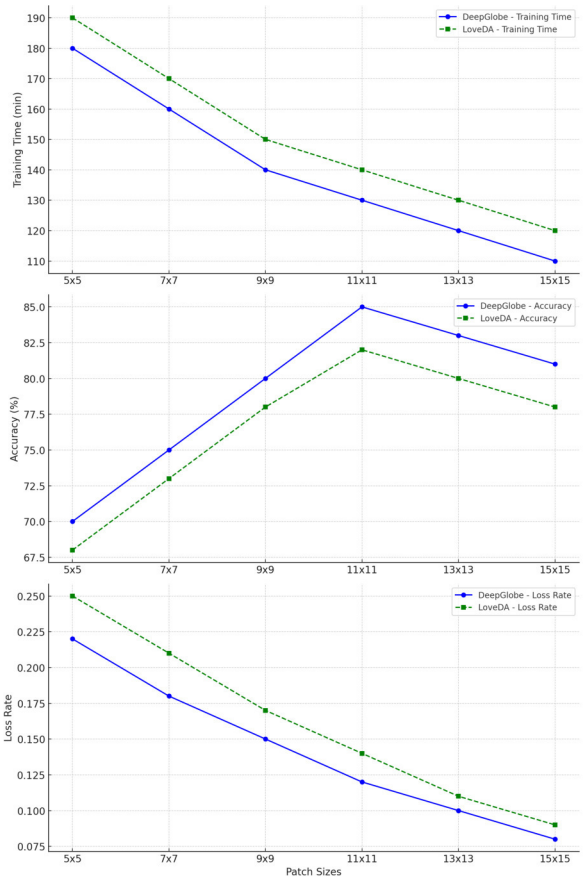

Fig. 3. Validation of the optimal hyperparameters with different patch sizes

Line plots demonstrate the trends in training time, accuracy, and loss rate across various batch sizes for the DeepGlobe-Forest and LoveDA-Forest datasets. For both datasets, as the batch size increases from 16 to 128, the training time consistently decreases, indicating enhanced computational efficiency with larger batches, possibly due to fewer updates needed per epoch. This reduction in training time does not seem to compromise the models' ability to learn effectively, as evidenced by the general increase in accuracy with larger batch sizes. However, the most notable improvement is observed in the loss rates, which decrease significantly as the number of batches increases, suggesting that larger batches help the model to converge more smoothly to a lower loss. This trend reflects the trade-off between computational speed and the stability of the training process, where larger batches provide a more stable but potentially less precise gradient estimation, beneficial for the overall learning process of the model.

We have also varied batch sizes to determine their impact on model performance and training dynamics. Batch sizes of 16, 32, 64, and 128 were systematically tested in multiple training

iterations to observe how they influenced the convergence rate, precision and computational efficiency of the neural network. The experiment aimed to identify the optimal batch size that balances between adequate gradient estimation and efficient resource utilization. The results, including metrics such as training time, model accuracy, and loss convergence rates, were carefully recorded and analyzed to deduce the effects of batch size adjustments on the overall effectiveness of the training process. The results are shown in Figure 4

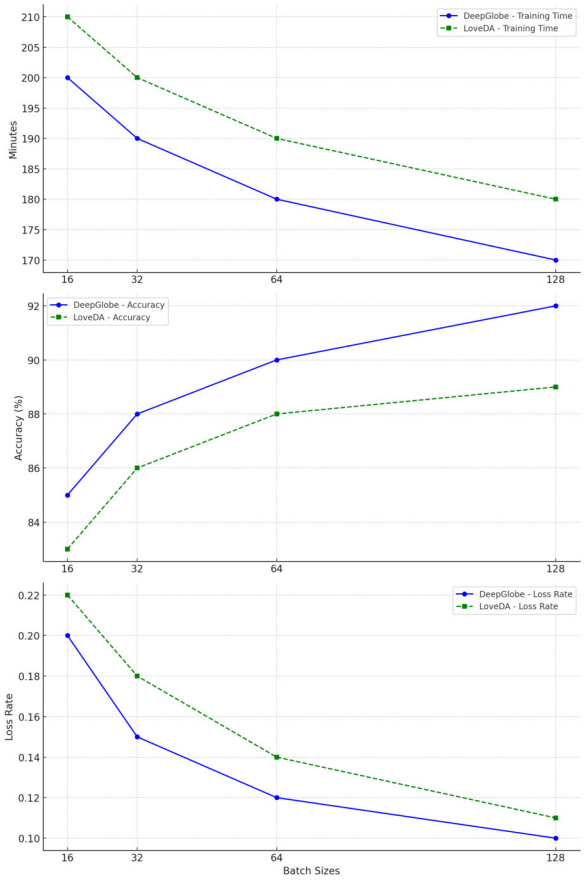

Fig. 4. Validation of the optimal hyperparameters with different batch sizes

Line plots illustrate how different patch sizes affect training time, accuracy, and loss rates for the DeepGlobe-Forest and LoveDA-Forest datasets. As the patch size increases from 5x5 to 15x15, both datasets exhibit a general decrease in training time, suggesting that larger patch sizes enable more efficient training, potentially due to fewer total iterations needed across the dataset. Accuracy trends upward for both datasets as patch size increases, peaking around 11x11 or 13x13 before slightly declining, which indicates an optimal range for capturing relevant features without introducing too much noise or suffering from interclass contamination. Loss rates consistently decrease as the size of the patch increases, reflecting the improved performance of the model with larger patches, which could be attributed to the models' increased ability to capture more comprehensive information about the image, thus potentially enhancing their learning capability. However, the slight decline in accuracy at the largest patch size suggests a trade-off, where too large a patch might start to incorporate irrelevant information or noise, negatively impacting model precision.

F. Discussion and conclusions

We believe, our approach of using the d'Alembert Convolutional Network in smart forest management, particularly for remote sensing image (RSI) change detection (CD), is a valid alternative compared to existing solutions, for example, the now well-established Spectral-Temporal Transformer (STT) [32]. Both methodologies aim to efficiently capture spectral-temporal features in HSIs, but employ different mechanisms and underlying theories. The STT focus on global spectral-temporal receptive fields with group-wise spectral embedding, linear projection, transformer encoders with an efficient multi-head self-attention mechanism, and a multilayer perceptron head for final change detection. Our approach is different though as it can simultaneously capture spatial features, such as edges and textures, and temporal changes, such as growth patterns or environmental dynamics. The convolution operation in this network is enhanced to include differential operators akin to the d'Alembert operator, creating a more sophisticated mechanism for feature extraction in RSI CD tasks. Furthermore, while STT employs a transformer-based approach with efficient MHSA to reduce computational intensity, the d'Alembert Convolutional Network utilizes the d'Alembert-enhanced kernel in its convolutional layers. This difference in approach leads to a variance in how each model handles the spectral-temporal data, with the d'Alembert network offering a novel perspective by incorporating wave equation principles into image analysis.

To evaluate the efficacy of the proposed d'Alembert convolution approach, we conducted a series of experiments with various parameters on three extensively utilized remote sensing image datasets, designed for change detection. The performance of our method was benchmarked against different established methods. From the analysis of the detection results, it is evident that our d'Alembert approach exhibits good performance in forest change detection, surpassing the comparative methods. For the DeepGlobe-Forest dataset, the d'Alembert Network achieved IoU scores of 83.19% for forest and 84.37% for background, and accuracy scores of 90.86% and 91.52%. These results outperform models like SegForest, Pspnet, and Mask2former benchmarked in the research of Wang et al. [24], showcasing the network's ability to distinguish forest textures and background areas effectively. Similarly, on the LoveDA-Forest dataset, the d'Alembert Network excelled with the highest IoU scores of 69.12% for forest and 80.05% for background, with accuracy rates of 87.93% and 85.87%.

Future work requires further balancing model of complexity and computational resources, considering that while the d'Alembert Convolutional Network enhances spatiotemporal feature extraction, it requires substantial computational power and memory, potentially increasing latency. We're currently working on reducing the number of convolutional layers and

employing techniques like model pruning and quantization to decrease computational load while striving to maintain acceptable performance levels. Naturally, more tests are needed to fully assess the robustness of the model against noise or missing data in the much larger variety of remote sensing images.

FUNDING INFORMATION

This research paper has received funding from Horizon Europe Framework Programme (HORIZON), call Teaming for Excellence (HORIZON-WIDERA-2022-ACCESS-01-two-stage) - Creation of the centre of excellence in smart forestry "Forest 4.0" No. 101059985. This research has been co-funded by the European Union under the project "FOREST 4.0 - Ekscelencijos centras tvariai miško bioekonomikai vystyti" (Nr. 10-042-P-0002)

REFERENCES

[1] S. Trumbore, P. Brando, and H. Hartmann, "Forest health and global change," *Science*, vol. 349, pp. 814–818, 2015.

[2] I. Boyd, P. Freer-Smith, C. Gilligan, and H. C. Godfray, "The consequence of tree pests and diseases for ecosystem services," *Science*, vol. 342, p. 1235773, 2013.

[3] D. Lindenmayer, "Future directions for biodiversity conservation in managed forests: indicator species, impact studies, and monitoring programs," *Forest Ecology and Management*, vol. 115, pp. 277–287, 1999.

[4] M. B. Nuwantha, C. N. Jayalath, M. P. Rathnayaka, D. C. Fernando, L. Rupasinghe, and M. Chethana, "A drone-based approach for deforestation monitoring," in *2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2022.

[5] W. Zou, W. Jing, G. Chen, Y. Lu, and H. Song, "A survey of big data analytics for smart forestry," *IEEE Access*, vol. 7, pp. 46621–46636, 2019.

[6] V. Upadhyay and A. Kumar, "Hyperspectral remote sensing of forests: Technological advancements, opportunities and challenges," *Earth Science Informatics*, vol. 11, pp. 487–524, 2018.

[7] A. Shukla and R. Kot, "An overview of hyperspectral remote sensing and its applications in various disciplines," *IRA-International Journal of Applied Sciences*, vol. 5, pp. 85–90, 2016.

[8] B. Banerjee, S. Raval, and P. Cullen, "Uav-hyperspectral imaging of spectrally complex environments," *International Journal of Remote Sensing*, vol. 41, pp. 4136–4159, 2020.

[9] M. Teke, H. S. Deveci, O. Haliloglu, S. Gurbuz, and U. Sakarya, "A short survey of hyperspectral remote sensing applications in agriculture," in *2013 6th International Conference on Recent Advances in Space Technologies (RAST)*, pp. 171–176, 2013.

[10] R. Prasad and K. Rajan, "Is current forest landscape research approaches providing the right insights? observations from india context," *Ecological Questions*, vol. 20, pp. 85–92, 2015.

[11] R. E. O. Schultz, T. M. Centeno, G. Selleron, and M. Delgado, "A soft computing-based approach to spatio-temporal prediction," *International Journal of Approximate Reasoning*, vol. 50, pp. 3–20, 2009.

[12] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 10, pp. 6232–6251, 2016.

[13] S. K. Roy, R. Mondal, M. E. Paoletti, J. M. Haut, and A. Plaza, "Morphological convolutional neural networks for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 8689–8702, 2021.

[14] T. K. Saha, H. Sajjad, Roshani, M. H. Rahaman, and Y. Sharma, "Exploring the impact of land use/land cover changes on the dynamics of deepor wetland (a ramsar site) in assam, india using geospatial techniques and machine learning models," *Modeling Earth Systems and Environment*, 2024.

[15] B. B. Thien, V. T. Phuong, and D. T. V. Huong, "Detection and assessment of the spatio-temporal land use/cover change in the thai binh province of vietnam's red river delta using remote sensing and gis," *Modeling Earth Systems and Environment*, vol. 9, no. 2, p. 2711 – 2722, 2023.

[16] R. N. Masolele, V. De Sy, M. Herold, D. Marcos Gonzalez, J. Verbesselt, F. Gieseke, A. G. Mullissa, and C. Martius, "Spatial and temporal deep learning methods for deriving land-use following deforestation: A pan-tropical case study using landsat time series," *Remote Sensing of Environment*, vol. 264, 2021.

[17] R. V. Maretto, L. M. G. Fonseca, N. Jacobs, T. S. Körting, H. N. Bendini, and L. L. Parente, "Spatio-temporal deep learning approach to map deforestation in amazon rainforest," *IEEE Geoscience and Remote Sensing Letters*, vol. 18, no. 5, p. 771 – 775, 2021.

[18] R. V. Fonseca, R. G. Negri, A. Pinheiro, and A. M. Atto, "Wavelet spatio-temporal change detection on multitemporal sar images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, p. 4013 – 4023, 2023.

[19] M. Dimyati, D. A. Umarhadi, I. Jamaluddin, D. Awanda, and W. Widyatmanti, "Mangrove monitoring revealed by mdprepost-net using archived landsat imageries," *Remote Sensing Applications: Society and Environment*, vol. 32, 2023.

[20] W. Jing, T. Lou, Z. Wang, W. Zou, Z. Xu, L. Mohaisen, C. Li, and J. Wang, "A rigorously-incremental spatiotemporal data fusion method for fusing remote sensing images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, p. 6723 – 6738, 2023.

[21] I. Demir, K. Koperski, D. Lindenbaum, G. Pang, J. Huang, S. Basu, F. Hughes, D. Tuia, and R. Raskar, "Deepglobe 2018: A challenge to parse the earth through satellite images," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018.

[22] J. Wang, Z. Zheng, A. Ma, X. Lu, and Y. Zhong, "Loveda: A remote sensing land-cover dataset for domain adaptive semantic segmentation," in *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks* (J. Vanschoren and S. Yeung, eds.), vol. 1, Curran Associates, Inc., 2021.

[23] J. Wang, Z. Zheng, A. Ma, X. Lu, and Y. Zhong, "LoveDA: A remote sensing land-cover dataset for domain adaptive semantic segmentation," Oct. 2021.

[24] H. Wang, C. Hu, R. Zhang, and W. Qian, "Segforest: A segmentation model for remote sensing images," *Forests*, vol. 14, p. 1509, July 2023.

[25] S. C. Yurtkulu, Y. H. Şahin, and G. Unal, "Semantic segmentation with extended deeplabv3 architecture," in *2019 27th Signal Processing and Communications Applications Conference (SIU)*, pp. 1–4, IEEE, 2019.

[26] J. Xu, Z. Xiong, and S. P. Bhattacharyya, "Pidnet: A real-time semantic segmentation network inspired by pid controllers," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 19529–19539, 2023.

[27] X. Zhu, Z. Cheng, S. Wang, X. Chen, and G. Lu, "Coronary angiography image segmentation based on pspnet," *Computer Methods and Programs in Biomedicine*, vol. 200, p. 105897, 2021.

[28] W. Zhang, J. Pang, K. Chen, and C. C. Loy, "K-net: Towards unified image segmentation," *Advances in Neural Information Processing Systems*, vol. 34, pp. 10326–10338, 2021.

[29] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "Segformer: Simple and efficient design for semantic segmentation with transformers," *Advances in neural information processing systems*, vol. 34, pp. 12077–12090, 2021.

[30] H. Zhang, F. Li, H. Xu, S. Huang, S. Liu, L. M. Ni, and L. Zhang, "Mp-former: Mask-piloted transformer for image segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 18074–18083, 2023.

[31] M.-H. Guo, C.-Z. Lu, Q. Hou, Z. Liu, M.-M. Cheng, and S.-M. Hu, "Segnext: Rethinking convolutional attention design for semantic segmentation," *Advances in Neural Information Processing Systems*, vol. 35, pp. 1140–1156, 2022.

[32] X. Li and J. Ding, "Spectral-temporal transformer for hyperspectral image change detection," *Remote Sensing*, vol. 15, no. 14, 2023.