

Received 27 August 2025, accepted 18 September 2025, date of publication 23 September 2025,
date of current version 29 September 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3613398

RESEARCH ARTICLE

Sea Ice Concentration Mapping Based on Sentinel-2 Multispectral Imagery Using Multiscale Deep Learning Approach

MARZURAIKAH MOHD STOFA¹, MOHAMAD AIZUDDIN ISMAIL², JAESUNG LEE³,
AND MOHD ASYRAF ZULKIFLEY¹, (Member, IEEE)

¹Department of Electrical, Electronic and Systems Engineering, Universiti Kebangsaan Malaysia, Bangi, Selangor 43600, Malaysia

²ICT Consulting and System Integration Department, Huawei Technologies (Malaysia) Sdn. Bhd., Kuala Lumpur 55188, Malaysia

³Department of Artificial Intelligence, Chung-Ang University, Seoul 06974, Republic of Korea

Corresponding author: Mohd Asyraf Zulkifley (asyraf.zulkifley@ukm.edu.my)

This work was supported in part by the Universiti Kebangsaan Malaysia through the Dana Padanan Kolaborasi Geran Pembiayaan Sepadan under Grant DPK-GPS-JORDAN-2024-021, and in part by the Ministry of Higher Education Malaysia through the Fundamental Research Grant Scheme under Grant FRGS/1/2022/TK07/UKM/02/4.

ABSTRACT One of the important climate indicators is the condition of sea ice concentration in the Arctic and Antarctic regions. Measuring this concentration using remote sensing and automated artificial intelligence allows continuous monitoring at relatively low cost. The purpose of this study is to develop a robust and scalable deep learning model to accurately estimate sea ice concentration from satellite imagery, supporting climate monitoring efforts. Hence, Ice-SPP-net (ISPP-net) is introduced as an innovative approach to measuring sea ice concentration using Sentinel-2 imagery, leveraging a novel multiscale convolutional neural network (CNN) architecture. This model is specifically designed to address challenges posed by the complex and heterogeneous nature of sea ice, which conventional fixed-scale CNNs struggle to assess accurately. By incorporating Spatial Pyramid Pooling (SPP) modules, the ISPP-net effectively captures and integrates features across multiple scales, enhancing the model's ability to discern both fine details and broader contextual information essential for precise measurements. The model was trained and evaluated using annotated Sentinel-2 datasets and validated through a comprehensive ablation study to determine the optimal architecture configuration. Results show that placing the SPP module at the third encoder layer, using three parallel paths, yields the best performance. This configuration outperforms models like U-Net, SegNet, DABNet, TeraNet, and FC-DenseNet, achieving a mean Intersection over Union of 49.20% and an accuracy of 89.30%. ISPP-net provides critical data to refine climate models and improve predictions of future climate scenarios. Future enhancements will integrate more data sources and refine the architecture, supporting SDG 13: Climate Action.

INDEX TERMS Sea ice concentration mapping, satellite imaging, artificial intelligence, convolutional neural network, multiscale approach.

I. INTRODUCTION

Sea ice plays a crucial role in the Earth's climate system, affecting ocean circulation, weather patterns, and global temperatures [1]. Its presence significantly influences the albedo effect by reflecting solar radiation [2], thereby cooling the

planet. However, climate change is causing rapid sea ice melting in polar regions, leading to significant environmental and economic impacts. Over the past four decades, Arctic Sea ice has significantly diminished, with some CMIP6 project models predicting that a completely ice-free summer could occur before 2050 [3]. Therefore, accurate sea ice concentration is the key factor to these changes and predicting future climate scenarios.

The associate editor coordinating the review of this manuscript and approving it for publication was Monidipa Das.

Beyond its climatic effects, sea ice is vital for polar species, indigenous communities, and global sea levels [4]. The loss of sea ice disrupts ecosystems and livelihoods, highlighting the need for comprehensive monitoring and analysis. This aligns with the United Nations' Sustainable Development Goal 13: Climate Action, emphasizing urgent measures to combat climate change. Reliable sea ice concentration data is crucial for informing policy and effective climate action [5]. Satellite imagery, especially from missions like Sentinel-2, provides high-resolution data essential for monitoring these remote regions. However, the inherent complexity and variability of sea ice patterns makes reliable measurement and analysis quite difficult [6].

Traditional methods for measuring sea ice concentration, such as passive microwave sensors [7], often provide coarse resolution data and miss finer details. Although synthetic aperture radar (SAR) and optical imagery have improved precision, they still face challenges like cloud cover and limited spatial resolution. Recent advancements in remote sensing [8] and machine learning [9] now enable more precise and automated measurement methods, leveraging high-resolution optical imagery to overcome these limitations.

Conventional convolutional neural networks (CNN), which use a uniform scale throughout the network, have been widely used in image analysis tasks [10]. These conventional CNNs apply the same scale of convolutional filters across all layers, which can limit their ability to capture both fine and coarse features simultaneously [11]. While effective for many applications, these networks may struggle with complex datasets like Sentinel-2 imagery, where the features of interest vary significantly in scale and detail [12]. Figure 1 depicts various patterns and concentrations of sea ice captured from Sentinel-2 satellite imagery. Each image showcases the diverse and complex nature of sea ice formations, highlighting differences in concentration, texture, and structure. These images exemplify the challenges in accurately measuring and classifying sea ice concentration due to its heterogeneous and evolving characteristics.

The gap in the research is evident in developing models that will effectively deal with the multiscale spatial variability evident in sea ice imagery. Most of the current CNN-based methods lack the adaptive multiscale technique, which limits their usability in precisely segmenting sea ice characteristics of various textures and scales.

Thus, to address these limitations, the study aims to propose and test a deep learning-based semantic segmentation model, Ice-SPP-net (ISPP-net), which exploits multiscale feature extraction to enhance the pixel-wise segmentation of sea ice concentration using Sentinel-2 imagery. This aim is motivated by the requirement to deal with issues like the necessity to differentiate ice and open water or low-concentration areas, the necessity to consider a variation in sea ice texture and density, and the necessity to consider the difference in spatial resolution of various satellite data sets. The proposed model, through the integration of multiscale

pooling mechanisms, seeks to result in a high degree of sensitivity to both small-scale features, e.g., cracks or thin edges of the ice, and large-scale structures, e.g., ice floes or pack ice areas, which are important when segmenting and measuring the sea ice.

This study introduces a novel multiscale convolutional neural network (CNN) specifically designed for the measurement of sea ice concentration using Sentinel-2 imagery. Unlike standard CNNs, the multiscale approach of multiscale CNNs incorporates filters of various sizes within the network architecture. This enables the model to capture both local and global features of sea ice, enhancing the accuracy and robustness of the measurements. By processing the imagery at multiple scales, the multiscale CNN can effectively integrate detailed textural information with broader spatial patterns, providing a more comprehensive analysis.

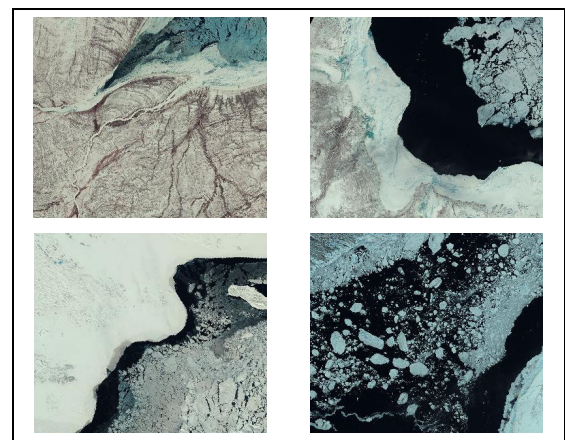


FIGURE 1. Sample of various sea ice concentration from Sentinel-2 imagery.

The multiscale architecture offers several advantages over the conventional single-scale CNN [13], [14]. Firstly, it allows the network to capture small-scale features such as cracks and melt ponds on the ice surface, which are crucial for understanding sea ice dynamics but are often missed by single-scale networks. Secondly, it enhances the network's ability to recognize larger-scale structures like ice floes and leads, which are vital for estimating overall ice concentration. This dual capability significantly improves the model's performance in heterogeneous environments where feature sizes vary widely.

The development of this multiscale CNN is motivated by the need for improved analytical tools that can handle the high-dimensional, complex nature of satellite imagery data [15]. By incorporating multiscale feature extraction, the multiscale CNN aims to address the limitations of existing methods, offering a more detailed and precise measurement of sea ice concentration. This advancement not only enhances our understanding of sea ice dynamics but also contributes to broader climate change studies by providing reliable data for predictive modeling and analysis [16].

In the subsequent sections, we delve into the specifics of the proposed multiscale CNN architecture, its training process, and the evaluation methodology. The results section will present the performance of the multiscale CNN, highlighting its superiority over conventional methods. Finally, the discussion will contextualize these findings within the broader field of remote sensing and environmental monitoring, emphasizing the implications for future research and practical applications.

II. RELATED WORKS

Sea ice images captured from satellites come in a variety of shapes, sizes, textures, and colors, which complicates the process of ice identification, especially for small ice floes. In response to these challenges, researchers have deployed a range of deep learning frameworks aimed at enhancing the detection rate of sea ice features. The application of these approaches has demonstrated versatility across various domains [17], including natural disaster alert systems [18], disease diagnosis [19], stock market prediction [20], and sea ice monitoring [21]. Multiple studies have investigated and affirmed the effectiveness of deep learning techniques in sea ice detection systems proposed by several researchers Zhang et al. [22]; Zhao et al. [23], highlighting the ongoing efforts to overcome the complexities associated with accurate detection in the face of diverse ice characteristics. Despite these gains, the model remained a fixed-scale convolutional model with no adaptive pooling and multiscale modules. Consequently, its performance may be compromised when utilised on heterogeneous imagery, high or low resolution, or textural features.

Before the advent of modern technology, researchers used manual observations and in situ measurements coupled with low-resolution satellite imagery for mapping and measuring sea ice [24]. These traditional methods, although fundamental, were labor-intensive and lacked the spatial and temporal resolution necessary to capture the full complexity of sea ice dynamics. Today, sea ice mapping has a broad spectrum of methods, from traditional observations to advanced remote sensing and modern technology-based approaches. The discipline also primarily utilizes different types of images, each with its advantages and problems.

The study of sea ice using aerial imagery data obtained from drones and satellites is more revolutionary in high-resolution data across large areas. The spatial resolution of drone imagery is high enough to capture the details of individual ice floes and local variations in ice characteristics. The flexible use of drones for targeted data collection in areas of interest makes them of very high value in monitoring specific locations or events. Some limitations on the part of the drones include a small coverage area; weather conditions, such as high winds and low visibility, will affect effective data collection, introducing noise.

Despite difficulties, aerial images captured by drones offer highly useful detail for analyzing ice floes, melt ponds, and

variations in ice thickness. Using drones offers higher resolution than satellites, allowing for more detailed separation and classification of features in sea ice. Zhang et al. [25] and Singh et al. [26] demonstrated that advanced U-Net-based deep learning models can reliably extract detailed features from high-resolution aerial images, despite challenges inherent to aerial data. Zhang et al. proposed a Deep Residual U-Net, where the original U-Net is augmented with a residual connection that modulates the boundary accuracy and network convergence, and thus has proved to be an architectural enhancement that can be transferred to elongated ice floe segments when narrow objects such as roads are to be detected. Meanwhile, Singh et al. used deep CNNs in river ice segmentation, emphasizing different ice types classification based on UAV data with high spatial resolution. Though both methods effectively preserved fine spatial features and minimal data preprocessing efforts, the designs were not multiscale adaptive. Then, the model was not tested for adequacy after scalability to UAV geographically distributed data before the addresses presented in this publication. This constrains their strength in use in more expansive and compositionally diverse sea ice environments, as illustrated by lower-resolution or multi-modal data sources.

The Icenet network proposed by Zhang et al. [27], designed and trained for the segmentation of river ice from UAV imagery, demonstrates superior performance compared to existing state-of-the-art methods in river ice segmentation tasks. Another novel aspect integrated within the Icenet network is using positional and channel-wise attentive mechanisms to harness more segmentation accuracy and boost overall performance. The newer iteration, IcenetV2, builds upon the original Icenet architecture with enhancements for river ice segmentation. These improvements support trend prediction, detailed analysis, and more accurate calculation of drift ice cover density for practical applications [28]. Icenetv2 proved it was more efficient than other models in estimating the drift ice cover's density and assisting the trend forecast with high-resolution information provided by the UAV. Although the presented results show that ICENet and IcenetV2 work well on localized and high-quality UAV datasets, the experiment was conducted only in homogeneous environments. Accordingly, extrapolation of the findings on a large-scale satellite data or mixed-modal sea ice circumstances remains a problematic issue, especially in the event of varied textures, different spatial resolutions, and complicated environmental variability.

Satellite imagery offers broad coverage and frequent observations, making it ideal for widespread monitoring of sea ice on a regional and global scale [29]. However, its lower spatial resolution compared to drones limits its ability to capture fine details, potentially affecting the accuracy of sea ice assessments in areas with complex ice compositions or mixed ice and water conditions. Issues with cloud cover can also impact data quality, necessitating advanced processing techniques to ensure reliable results. Despite these challenges, satellite

images, analyzed with deep-learning methods, are crucial for understanding key sea ice characteristics like extent and concentration.

Research using satellite imagery involves investigations into Arctic Sea ice mapping through Sentinel-1 SAR scenes [30] and estimation of Baltic Sea ice concentration [31], which highlights the effectiveness of the CNN with UNet++ architectures in semantic segmentation tasks for sea ice detection and classification from satellite-derived data. In another study, the spatial resolution of AMSR2 images for the purpose of sea ice concentration estimation was enriched through a Super-Resolution (SR) network in which the spatial resolution was improved four times, making sea ice concentration estimation even more accurate [32]. Merging Sentinel-1 SAR and AMSR2 sources in a multi-modal CNN brings segmentation accuracy improvement to increase the overall reliability of sea ice charts over any kind of previous manual process [33]. This reduces the manual efforts and saves the time spent on interpretation. In both articles, the CNN was proved to be applicable in terms of semantic segmentation on the satellite-derived data, particularly the one extended with network architectures such as UNet++. Nevertheless, the drawbacks are still connected to using uniform convolutional kernels and challenges to represent multiscale ice characteristics in different environmental and spatial circumstances.

On the other hand, ship-mounted camera imagery obtained at ground level generates distinct observations regarding sea ice dynamics and interaction within that environment. Observations from ships provide a close view of sea ice; thus, researchers can study how ice features and processes in this system interact with the environment. High-resolution images from onboard cameras provide valuable data on ice concentration and distribution that enhance understanding of ice-ocean-atmosphere interactions. Ship-based cameras can be free of many of the limitations posed by aerial photography regarding the need for drones to be flying for long times, their limited endurance in flight, their payload capacity, and, at the same time, offer an almost unlimited number of observations over a long period. This provides insight into seasonal variations in sea ice conditions.

The near real-time nature of the observations from cameras mounted on ships makes them very useful for close-range ice analysis and navigational decision support [34]. Furthermore, deep neural network semantic segmentation-based sea ice classification studies and the improved sea ice identification showed that images from ship cameras can be used to train DNN models for precise detection and classification of sea ice types over a wide range of environment conditions [35]. Both research studies substantiate the reliability of deploying ship-based imagery when locating sea ice in various weather conditions and other operational restrictions. However, these studies are optimized for small-scale settings and have low scaling capabilities for large-scale or multi-modal remote sensing problems.

For applications with real-time constraints that run on resource-constrained platforms, such as embedded systems or mobile devices, in ship-mounted camera studies for fast processing of sea ice imagery and providing navigational risk assessments and operational support, the use of DeepLab models optimized for efficient inference is essential [34]. Zhang et al. [22] introduced the Ice-Deeplab model, a refined deep neural network structure based on the original DeepLabv3+ model that constructs attention mechanisms in the convolutional networks and fuses shallow layer features to improve the exact delineation of boundaries in multi-scale sea ice features depending on seasonal variations.

Despite the various deep learning methods that have been used in sea ice segmentation, including U-Net, DeepLab, and IceNet, the majority of previous studies rely on consistent convolutional scales or are oriented towards small-scale, homogeneous investigations. There has been less focus on including multiscale modules such as SPP or ASPP systematically and assessing their evaluation in large and diverse satellite images like Sentinel-2. To fill these gaps, our research suggests the architecture ISPP-net with a spatial pyramid pooling component at an optimal depth to facilitate the segmentation performance of diverse and large-scale sea ice data.

III. METHODOLOGY

This section delineates the primary approach employed in this study, emphasising the formulation of the Ice-SPP-net (ISPP-net) architecture for precise sea ice concentration mapping via Sentinel-2 imagery. The methodology is designed to explicitly delineate the challenge, articulate the network architecture, and elucidate the innovations implemented via multiscale feature extraction.

A. PROBLEM DEFINITION

The key problem addressed in this study is the challenge of accurately measuring sea ice concentration from satellite imagery due to the heterogeneous, multiscale, and complex nature of sea ice formations. Traditional convolutional neural networks (CNNs) tend to have weaknesses in representing fine and coarse spatial information, and hence stand to be less effective in segmenting such imagery characteristics. This is a weakness that encourages better architecture of a more adaptive and unshakable system.

The development of our proposed multiscale CNN is driven by the need to accurately analyze and measure sea ice concentration from satellite imagery, addressing the limitations of conventional uniform scale CNNs. These standard networks often fail to adequately process the varied and intricate features of sea ice, such as ice floes or melt ponds, which require high-resolution local data, and large-scale structures that necessitate a broader contextual perspective.

For complex sea ice imagery, capturing multi-scale contextual information is crucial. The Spatial Pyramid Pooling (SPP) module enhances the model's ability by pooling

features at multiple scales, integrating fine details and broader context, which is essential for segmenting diverse sea ice formations. The Atrous Spatial Pyramid Pooling (ASPP) module uses dilated convolutions to capture wide context while maintaining fine detail. This study aims to determine which module, SPP or ASPP, is best suited for sea ice segmentation by integrating them into our multiscale CNN architecture and evaluating their performance in improving feature extraction and segmentation accuracy. By employing multiscale feature extraction, the network can capture details at multiple levels, enhancing its ability to differentiate between various types of sea ice and water, thereby significantly improving segmentation and classification accuracy in complex image analysis tasks.

The multiscale modules in our proposed network are designed to overcome these limitations by integrating features from various scales, thereby providing a richer and more comprehensive representation of the input data. This approach enhances the network's ability to handle the heterogeneity of sea ice images, resulting in more precise and reliable measurements.

B. BASE ENCODER-DECODER MODEL

This subsection will present the base architecture we used to generate semantic segmentation and the structure upon which our proposed model will be laid. The encoder-decoder architecture was selected because it demonstrated the best results in dense predictive tasks, including pixel-wise classification. We describe how the convolutional layers, skip connections, and upsampling mechanisms are arranged to enable the model to retain the low-level geometric detail and high-level semantics, key components representing the subtle variations in sea ice concentration over heterogeneous satellite images.

The classification of sea ice and open water can be defined as a semantic segmentation problem. In this study, we propose a base encoder-decoder CNN model with addition of multiscale module. This model comprises five layers of encoders and five layers of decoders, with each encoder layer being connected to its corresponding decoder layer via skip connections. This model architecture enhances feature preservation and improves gradient flow during training, facilitating better localization and detailed segmentation. The skip connections also allow for a richer backpropagation of gradients, reducing the problem of vanishing gradients and improving the model's ability to learn from complex images.

The input to this model is an image with dimensions of 256×256 pixels. A convolutional (Conv) layer, a batch normalisation (BatchNorm) layer, a rectified linear unit (ReLU) activation function, and a pooling layer make up each encoder layer. This configuration is systematically repeated across all five encoder layers, enabling the progressive extraction of more complex and higher-level features from the input image. In contrast, the decoder layers each start with an upsampling layer to enhance the spatial resolution, followed by a Conv layer, BatchNorm, and a ReLU activation function. This setup

effectively reconstructs the decoded features back to the size of the original image, ensuring detailed and accurate image segmentation.

The architecture concludes with a final layer featuring a Softmax activation function. This layer outputs a probability distribution across the different segmentation classes, allowing for the classification of each pixel in the image. This comprehensive setup ensures that the model not only captures detailed features but also accurately classifies them, making it highly effective for tasks requiring precise image segmentation.

C. MULTISCALE MODULE ARCHITECTURE

This subsection proposes the multiscale feature extraction modules added to the base network as a solution to the flaws of fixed-scale CNNs. We describe in detail the Spatial Pyramid Pooling (SPP) and Atrous Spatial Pyramid Pooling (ASPP) designs. These modules were intended to improve the capacity of the model to detect information from various receptive fields concurrently, facilitating better partitioning of sea ice characteristics that exhibit huge variation in dimensions, structure, and appearance. The theory and use of each module are explained in terms of remote sensing applications.

The spatial pyramid pooling (SPP) module is designed to pool features at multiple scales and concatenate them to form a robust multi-scale representation of the input data. It operates by applying pooling operations with different window sizes, capturing spatial hierarchies of features that are crucial for understanding complex patterns in sea ice imagery. By pooling features at varying scales, SPP can integrate both fine-grained local details and broader contextual information, which enhances the network's ability to segment and classify sea ice accurately.

SPP can integrate fine-grained local details and broader contextual information by pooling features at varying scales. After pooling, each branch output is bilinearly upsampled (resize step) to the original feature-map resolution so that all branches share the same spatial dimensions before concatenation. The resize operation is non-learnable and uses standard bilinear interpolation. Figure 2 depicted the architecture of SPP integrated within a CNN.

On the other hand, the ASPP module employs atrous convolutions (also known as dilated convolutions) with different dilation rates to capture multi-scale contextual information without reducing the spatial resolution of the feature maps. Atrous convolutions allow the network to have a larger receptive field, enabling it to capture more contextual information while maintaining fine detail resolution. The ASPP module effectively integrates features from different scales, making it well-suited for tasks that require a detailed understanding of spatial hierarchies, such as sea ice segmentation.

Atrous convolutions enable the network to have a larger receptive field without downsampling, allowing it to capture fine details and broader context. Since all atrous branches operate on the same input feature map without altering spatial size, the outputs are spatially aligned and do not

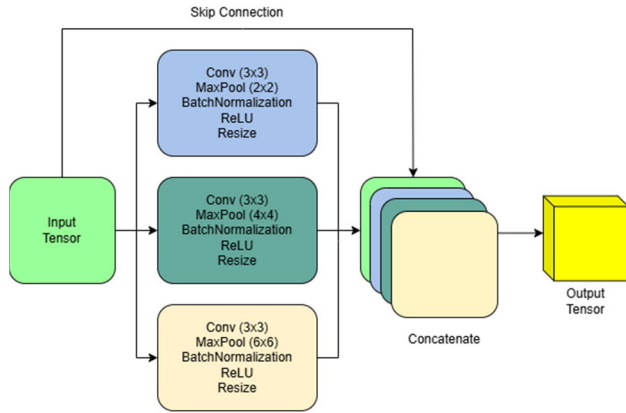


FIGURE 2. SPP module architecture.

require resizing before concatenation. This design makes the ASPP module well-suited for tasks like sea ice segmentation, which demand multi-scale contextual understanding. Figure 3 depicts the architecture of ASPP integrated within a CNN.

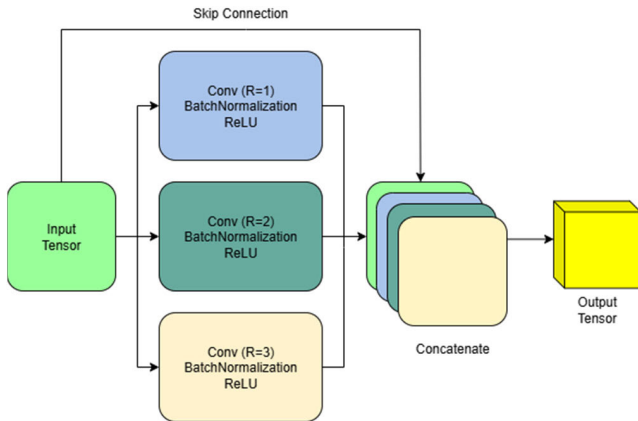


FIGURE 3. ASPP module architecture.

D. MULTISCALE MODULE EMBEDDING

In this subsection, we explain how we strategically combine the multiscale modules, SPP and ASPP, with the base network and measure the outcome on the segmentation task performance. The ablation study was performed to ascertain the best type of module, where in the encoder to implement it, and how many paths of different configurations to parallelize, to find a balance between performance and cost of computation, and we describe the methodology. We will ensure that we can determine the most optimal architecture approach for it by testing various configurations systematically.

To address the limitations of the uniform scale CNN, we modified the base encoder-decoder model by incorporating multiscale modules. These modules are strategically placed within the network to capture and integrate features at different scales, enhancing the model's ability to process

complex sea ice imagery. The optimal network configuration was identified through a comprehensive ablation study that tested various aspects of network design. We evaluated different multiscale architectures, notably Spatial Pyramid Pooling (SPP) and Atrous Spatial Pyramid Pooling (ASPP), to determine the most effective architecture. Additionally, the placement of these multiscale modules was examined within different layers of the encoder, specifically at layers 3, 4, and 5, to ascertain the most beneficial location for feature extraction. We also experimented with the number of parallel paths within the multiscale modules, ranging from 2 to 4, aiming to find the best compromise between performance enhancement and computational efficiency. This systematic approach allowed us to fine-tune our network design to achieve optimal performance for our specific application requirements.

E. ISPP-NET

This subsection summarizes the results of the ablation experiments and provides the final architecture of the model, which was named Ice-SPP-net (ISPP-net). We explain the best configurations that resulted in the best segmentation accuracy. An example is using an SPP module on an encoder layer with three parallel paths, namely the third layer of the encoder. A rationale for this configuration is described, and the following architecture is outlined in detail. Here we have come to the end of our methodological approach and have a good, scalable, and massively efficient sea ice concentration mapping model.

The ablation study revealed that the optimal configuration for our multiscale CNN involved using the SPP module placed at layer 3 of the encoder with 3 parallel paths. This configuration consistently yielded the highest performance across various evaluation metrics. This optimized architecture leverages multiscale feature extraction to effectively handle the diverse and complex nature of sea ice images, resulting in superior segmentation and classification performance. Figure 4 shows the final architecture of our proposed Ice-SPP-net (ISPP-net) with SPP module embedded at the 3rd layer of the model encoder. Table 1 provides a detailed summary of ISPP-Net architecture. “Number of Filters” indicates the number of output channels in the convolutional layer. “Kernel Size” and “Stride” refer to the spatial dimensions (height $\times$ width) of the convolution or pooling operations. Numbers (e.g., 64, 128, 256, 512) indicate the number of filters at each corresponding stage.

IV. EXPERIMENT AND DISCUSSION

A. DATASET

This process will begin with the collection of sea ice data, which are satellite images of Hudson Bay Sea ice in the Canadian Arctic Archipelago. These images are taken from the Sentinel-2 satellite database that has been uploaded to the Kaggle platform website. In addition, the Canadian Regional Ice Chart was also used in this study to generate a mask for

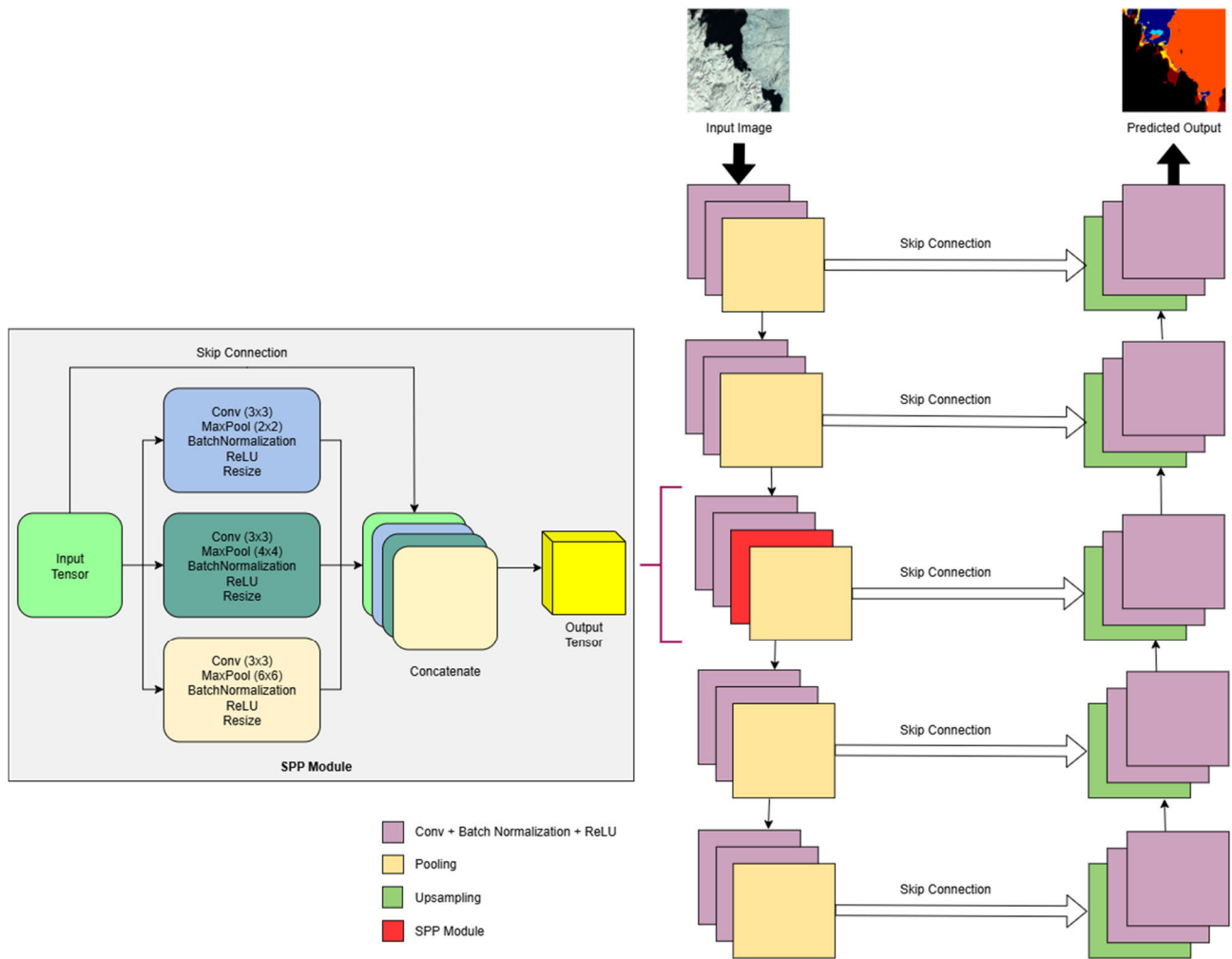


FIGURE 4. Proposed architecture of ISPP-net.

each of the satellite image data to classify the sea ice concentration into eight classes. These Sentinel-2 satellite images offer detailed depictions of icy areas with an accuracy of 10-meters per pixel. The satellite images obtained are 3392 and dated from 01.01.2016 to 31.07.2018. Figure 5 illustrates example of a set of input image and its annotated sea ice map and Table 2 shows the division of ice concentration into eight classes.

The satellite image dataset is then set to a size of 256×256 for each small patch. The selected spot size is based on previous related studies that have been conducted by researchers, and the majority of those studies have shown satisfactory segmentation performance and use a spot size resolution of approximately 256×256 . We divided the dataset into a 70% training set and a 30% testing set, following the approach of Kang et al. [36], who also employed a two-way partitioning strategy in sea ice segmentation. The training set (2374 images) was used to fit the model, while the testing set (1018 images) remained strictly untouched during model development and was used only for final evaluation.

The training set was intentionally larger to ensure the model could effectively learn the underlying patterns and generalize well. The splitting process was randomized to maintain representativeness and minimize bias among the dataset. This ensured that both training and testing subsets reflected the diversity of the overall dataset. To mitigate potential overfitting and support generalization, we applied several data handling techniques, including class balancing through over-sampling and random shuffling of the dataset in each run. In future work, we intend to adopt a three-way split with a dedicated validation set to enhance model tuning and ensure unbiased performance evaluation.

Furthermore, images with fewer pixels for some classes, were oversampled. This will help in dealing with class imbalances in the dataset which can bias the model towards more pixel classes. In addition, the data set also goes through a random image imputation process. This process helps the model learn from diverse scenarios, as the training environment was enriched through additional data variations. Flip, rotate images and carve out smaller patches to explore every corner

of the icy landscape. In this way, it further diversifies the training experience, and prepares our model for real-world uncertainties.

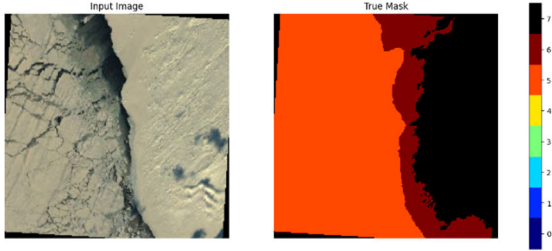


FIGURE 5. Example of a set of input image and its annotated sea ice map.

B. EVALUATION METRICS

The performance of the suggested ISPP-net and its benchmarked models are assessed using two evaluation metrics. The chosen metrics are mean intersection over union (*mIoU*) and accuracy (*ACCU*). True positive (*Tp*), true negative (*Tn*), false positive (*Fp*), and false negative (*Fn*) are the fundamental units on which these metrics are based. False positives (*Fp*) and false negatives (*Fn*) occur when the predicted class does not correspond with the actual ground truth, indicating erroneous positive and negative predictions, respectively. Conversely, true positives (*Tp*) and true negatives (*Tn*) occur when the prediction accurately matches the ground truth label for both positive and negative cases. Based on these indices (*Tp*, *Tn*, *Fp*, and *Fn*), the *mIoU*, and *ACCU* can be calculated by the following formula:

$$mIoU = \frac{1}{j} \sum_{i=1}^j \frac{T_p}{T_p + F_p + F_n} \quad (1)$$

$$ACCU = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \quad (2)$$

in which the j stands for the j^{th} classes. While *mIoU* measures the overlap between the predicted segmentation and the actual ground truth as the ratio of their intersection to their union, *ACCU* focuses primarily on the true detection rate for both positive and negative cases.

The main evaluation metric used to evaluate and discuss the models' performance in this study is the *mIoU*. Since this study is an 8 classes segmentation task, using *mIoU* is advantageous because it handles class imbalances by equally weighing each class's Intersection over Union, provides a comprehensive measure of accuracy for each class, and assesses the model's ability to manage overlaps and closely located classes. Additionally, *mIoU* offers a balanced performance indicator across all classes, preventing the metric from being skewed towards classes that are easier to predict. It is also optimization-friendly, suitable for gradient-based optimization methods used in deep learning, which allows it to be incorporated directly as a loss function to

optimize model performance across diverse sea ice conditions effectively.

C. EXPERIMENTAL SETTING

The ISPP-net was created using the Keras-Tensorflow framework in Python to enable accurate comparisons, with hyper-parameters finely tuned to maximize classification efficacy. Analytical performance metrics were calculated using the Numpy library. The architecture standardizes the use of one-hot encoding with a SoftMax activation function in the final dense layer. The Adam optimizer, maintaining a steady learning rate, was employed to adjust the model parameters during training sessions, focusing on optimizing the cross-entropy loss function for improved accuracy. Several data augmentation techniques were applied to improve generalization and reduce overfitting. These included random image flipping, rotation, and patch extraction. All satellite images were resized from the original resolution to 256×256 pixels, in order to match the input requirements of ISPP-net and baseline models. Batch sizes were chosen based on the constraints of the model dimensions, with the maximum feasible batch size being used to train the models on a single Nvidia Tesla P100 graphics card available through the Kaggle platform, which supports a batch size of 64 for the ISPP-net and five other baseline models. Table 3 below states the overall hyper-parameter settings for the models' training.

D. BASELINE MODELS

We compare the proposed model of ISPP-net with the following existing models:

- 1) U-net [37]: U-Net is a convolutional neural network designed for precise biomedical image segmentation, characterized by its distinctive U-shaped architecture.
- 2) Seg-net [38]: Seg-net is a deep convolutional encoder-decoder architecture optimized for semantic pixel-wise segmentation
- 3) DAB-net [39]: DAB-net is a compact convolutional neural network featuring depth-wise asymmetric bottleneck blocks for efficient semantic segmentation.
- 4) Ternaues-net [40]: TernauesNet is a U-Net-based convolutional neural network enhanced with pre-trained VGG encoders for improved image segmentation.
- 5) FC-Dense-net [41]: FC-Dense-net is an FCN that integrates the dense connectivity pattern of Dense-net into the architecture for semantic segmentation.

E. ABLATION STUDY

We conducted a comprehensive ablation study to determine the optimal network configuration for our proposed multiscale CNN aimed at sea ice concentration measurement. The objective was to meticulously evaluate different multiscale module designs, their specific placements within the network layers, and the impact of varying the number of parallel paths on overall performance.

TABLE 1. Overall architecture of the ISPP-net.

No.	Layer	Output Size	Number of Filters	Kernel Size	Stride
1	Convolution	256 x 256	64	3 x 3	1 x 1
2	Max Pooling	128 x 128	-	2 x 2	2 x 2
3	Convolution	128 x 128	128	3 x 3	1 x 1
4	Max Pooling	64 x 64	-	2 x 2	2 x 2
5	Convolution	64 x 64	256	3 x 3	1 x 1
6	Max Pooling	32 x 32	-	2 x 2	2 x 2
7	Spatial Pyramid Pooling (SPP)	32 x 32	256-256-256	2 x 2-4 x 4-6 x 6	1 x 1-2 x 2-3 x 3
8	Concatenate	32 x 32	256	3 x 3	1 x 1
8	Convolution	32 x 32	512	3 x 3	1 x 1
9	Max Pooling	16 x 16	-	2 x 2	2 x 2
10	Convolution	16 x 16	512	3 x 3	1 x 1
11	Max Pooling	8 x 8	-	2 x 2	2 x 2
12	Convolution (Upsample)	16 x 16	512	3 x 3	1 x 1
13	Convolution	32 x 32	512	3 x 3	1 x 1
14	Convolution	64 x 64	256	3 x 3	1 x 1
15	Convolution	128 x 128	128	3 x 3	1 x 1
16	Convolution	256 x 256	64	3 x 3	1 x 1
17	SoftMax	-	8	1 x 1	1 x 1

TABLE 2. Division of ice concentration into eight classes.

Class	Value
0	< 10% Ice
1	10% - 30% Ice
2	30%-50%
3	50%-70% Ice
4	70%-90% Ice
5	90%-100% Ice
6	Fast Ice
7	Land

Firstly, we investigated two different multiscale module architectures: Spatial Pyramid Pooling (SPP) and Atrous Spatial Pyramid Pooling (ASPP). These architectures were selected for their demonstrated ability to capture contextual information at multiple scales, a feature that is crucial for accurately analyzing the complex patterns of sea ice as captured in Sentinel-2 imagery. Next, to determine the optimal placement of the multiscale module, we experimented with incorporating the SPP and ASPP modules at three different layers within the encoder of the network: layers 3, 4, and 5. Each of these layers represents a different level of feature abstraction:

TABLE 3. Hyper-parameter settings for the baseline models and ISPP-net.

Item	Hyper-parameter Setting
Batch size	64
Training epoch	100
Input image size	256 x 256 pixels
Backpropagation method	Adam optimizer
Optimizer learning rate	0.001

1) Layer-3: Captures fine-grained, local features essential for detailed ice structure analysis.

2) Layer-4: Represents an intermediate level, balancing detail and broader contextual information.

3) Layer-5: Captures high-level, abstract features which are useful for general context but might miss finer details.

Lastly, we further varied the number of parallel paths within the multiscale modules, testing configurations with 2, 3, and 4 parallel paths. The parallel paths allow the network to process input features at multiple scales simultaneously, which enriches the representation of the input data. This variation aimed to find the configuration that strikes the best balance between computational efficiency and performance:

1) Two Parallel Paths: Provides basic multi-scale representation but may lack depth in capturing complex features.

2) Three Parallel Paths: Enhances multi-scale feature representation while maintaining computational efficiency.

3) Four Parallel Paths: Increases the depth of multi-scale analysis but adds significant computational overhead without proportional performance gains.

Table 4 and Table 5 presented the ablation study results with variation of placement of the module and number of parallel paths within the multiscale module for SPP module and ASPP module, respectively.

TABLE 4. Results of ISPP-Net.

SPP Module	<i>mIoU</i> (%)		
	3 rd Encoder Layer	4 th Encoder Layer	5 th Encoder Layer
Two Parallel Paths	46.00	44.80	43.60
Three Parallel Paths	49.20	45.00	42.00
Four Parallel Paths	47.50	45.20	44.00

TABLE 5. Results of IASPP-net.

ASPP Module	<i>mIoU</i> (%)		
	3 rd Encoder Layer	4 th Encoder Layer	5 th Encoder Layer
Two Parallel Paths	45.40	43.90	43.50
Three Parallel Paths	46.50	43.10	42.00
Four Parallel Paths	45.00	44.20	42.10

The ablation study's findings revealed that the optimal configuration for our multiscale CNN involved using the SPP module placed at layer 3 of the encoder with 3 parallel paths. This specific configuration consistently yielded the highest performance across the evaluation metrics, *mIoU*.

The SPP module demonstrated superior performance compared to the ASPP module in the context of sea ice measurement. The straightforward approach of SPP in pooling features at multiple scales effectively captured the essential multi-scale contextual information needed for this task. This suggests that the complexity introduced by the atrous convolutions in ASPP did not provide additional benefits for this particular application. Placing the multiscale module at layer 3 was the most effective. Layer 3's position in the network allows it to capture detailed local features, which are crucial

for accurately identifying and measuring the characteristics of sea ice. The analysis showed that this layer strikes an optimal balance by providing sufficient detail without losing broader contextual information. The configuration with 3 parallel paths proved to be the most effective. Two parallel paths provided a limited multi-scale representation, insufficient for capturing the complex features of sea ice. On the other hand, four parallel paths introduced unnecessary complexity and computational load without a significant improvement in performance. Three parallel paths, however, offered a balanced approach, enhancing feature extraction while maintaining computational efficiency.

The comprehensive ablation study highlighted that the SPP module, when placed at layer 3 of the encoder with 3 parallel paths, is the most effective configuration for our multiscale CNN. This optimal configuration exhibited a superior ability to accurately measure sea ice concentration from Sentinel-2 imagery, validating its effectiveness for this application. The findings underscore the importance of detailed local feature extraction and balanced multi-scale representation in improving the performance of sea ice measurement models. Thus, developing the best multiscale CNN with our proposed model, ISPP-net.

F. BENCHMARKING RESULTS

In this section, we compared the performance of our proposed model, the ISPP-net, against five baseline models to evaluate its performance. The comparison was based on a series of standardized tests using the same dataset and metrics. Each model was assessed using identical hyper-parameters settings, and the performance was measured through *mIoU* and *ACCU*. Table 6 shows the evaluated metrics for all models in this study. The best result is indicated in bold. It can be observed that ISPP-net achieves the best performance among other baseline models. The results indicate that the combination of a base encoder-decoder architecture and multiscale module of SPP can improve the performance of sea ice concentration segmentation. The proposed model achieves an improvement of 6.2 percentage points in *mIoU* compared to the highest performing baseline model, Seg-net. To further isolate the contribution of the multiscale SPP module, we also evaluated a variant using only the base encoder-decoder architecture without the SPP module. This configuration achieved an *mIoU* of 43.7%, confirming that the SPP module contributes a 5.5 percentage point gain in segmentation performance. Similarly, the performances in terms of *ACCU* is also illustrated in the Table 6.

Figure 6 shows the segmentation result of our proposed ISPP-net. The figure shows the input image of raw satellite data, while the true mask represents the ground truth segmentation. The predicted mask, generated by our ISPP-net, closely aligns with the true mask, demonstrating the model's ability to accurately identify and segment sea ice features.

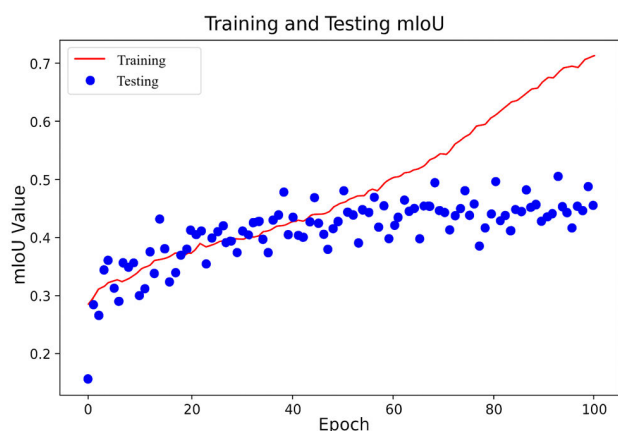
Figure 7 illustrates the training and testing *mIoU* values over 100 epochs. While the training *mIoU* shows a consistent upward trend, the testing *mIoU* fluctuates and diverges

TABLE 6. Comparison results.

Model	<i>mIoU</i> (%)	<i>ACCU</i> (%)
Ternaus-net	34.40	74.20
FC-Dense-net	36.00	76.00
U-net	39.00	78.00
DAB-net	42.40	83.30
Seg-net	43.00	84.20
Base Encoder-Decoder (no SPP)	43.70	84.90
ISPP-net	49.20	89.30

**FIGURE 6.** ISPP-net Segmentation Task Result.

from the training curve after approximately 60 epochs. This widening gap suggests potential overfitting, where the model increasingly adapts to training data at the expense of generalization to unseen data. Although data augmentation techniques such as flipping and rotation were applied to mitigate this effect, further regularization strategies such as early stopping, dropout tuning, or more diverse augmentation may help narrow this gap. These enhancements are proposed for future work to improve model robustness.

**FIGURE 7.** ISPP-net *mIoU* Graph.

In the performance analysis of various models, U-net established a solid baseline with a *mIoU* of 39.00% and *ACCU* of 78.00%, although its capacity to handle complex

multiscale features was limited. Seg-net showed improvement over U-net, achieving *mIoU* of 43.00% and *ACCU* of 84.20%, yet it did not reach the performance levels of our proposed ISPP-net model. DAB-net was competitive with *mIoU* of 42.40% and *ACCU* of 83.30%, closely trailing Seg-net. Ternaus-net, however, recorded the lowest metrics in the group, with *mIoU* of 34.40% and *ACCU* of 74.20%, reflecting its struggles with the detailed features of sea ice imagery. FC-Dense-net performed marginally better than Ternaus-net, posting *mIoU* of 36.00% and *ACCU* of 76.00%, but it still lagged behind the other models.

The superior performance of the ISPP-net can be attributed to its multiscale feature extraction capabilities. Multiscale approaches are particularly advantageous for sea ice imagery, which is characterized by a wide variety of shapes, sizes, textures, and colors. The ability to process features at multiple scales allows the model to capture both fine-grained details and broader contextual information, which is essential for accurate segmentation and classification.

In the context of a sea ice image dataset with 8 classes, the multiscale approach is crucial for distinguishing between different types of ice and water features. The classes include various ice formations and open water that can vary significantly in appearance. By leveraging multiscale feature extraction, the ISPP-net can effectively identify and differentiate these classes, leading to higher accuracy and better segmentation performance.

Overall, the use of multiscale techniques in the ISPP-net has proven to be highly effective for the complex task of sea ice concentration measurement. The enhanced ability to capture detailed and contextual information from the imagery has resulted in a model that significantly improves upon existing methods, making it a valuable tool for both scientific research and operational applications in monitoring and understanding sea ice dynamics.

G. DISCUSSION OF RESULTS AND OBSERVED PERFORMANCE

The experimental outcomes show that the ISPP-net model achieved numerically higher performance compared with the other segmentation models (U-net, Seg-net, DAB-net, Ternaus-net, and FC-Dense-net). In particular, the ISPP-net had the best *mIoU* (49.20%) and *ACCU* (89.30%) standings of all models. This observed performance improvement can be explained by the fact that in our architecture, we have the SPP module, making it possible to extract features across multiple scales, and also by the fact that the network is capable of catching more global context and more local structures of the data. In contrast with the traditional CNNs, which are run at a predetermined scale, ISPP-net is adaptive to the varying patterns of sea ice, further empowering the model to divide up complicated and heterogeneous sea ice gestures.

This performance gap is particularly notable when comparing ISPP-net with Seg-net and DAB-net, which also use encoder-decoder frameworks but lack adaptive

multiscale pooling mechanisms. This difference is particularly noticeable when comparing between ISPP-net and Seg-net and DAB-net, which, however, use encoder-decoder frameworks without an adaptive multiscale pooling layer. ISPP-net shows improvements of more than 6% in mIoU over these models, which suggests that multiscale contextual learning can be beneficial in the segmentation of natural unstructured patterns such as sea ice.

These findings indicate potential benefits. Mapping sea ice concentration with high accuracy is critical to the awareness of the polar regions and climate-related event predictions. The environmental monitoring systems and the efficiency of the safety of the sea navigation in the arctic area can also be positively affected by the improved accuracy of segmentation, which can also be used to support the global climate modeling that refers to sea ice extent and concentration as central variables. In addition, it is possible to use the ISPP-net framework as a guideline in the multiscale modeling of other environmental monitoring applications, e.g., deforestation mapping or flood extent identification.

Although ISPP-net is quite a good performer, it has certain limitations as well. First, Sentinel-2 imagery, which is prone to cloud cover and seasonal change in light, is high-resolution but can be impacted by seasonal changes in the light. Second, applying 256×256 patches provides significant computational advantages but is at the cost of losing larger spatial context between patches. Finally, the currently implemented version is able to process only optical imagery; adding multi-modal information, such as SAR, might be a solution to make it more robust in adverse conditions (e.g., night, cloud cover).

The limitations will be dealt with in future work to include multi-sensor fusion, temporal data, and transfer learning within the Arctic and Antarctic regions. Increasing the capacity of the model to allow its application to a broader geographic and seasonal range will be necessary in the process of deploying it as a practical implementation tool of climate action within the United Nations SDG 13 framework of action.

V. CONCLUSION

In this study, we presented a novel multiscale convolutional neural network approach, specifically the ISPP-net, designed for the accurate measurement of sea ice concentration using Sentinel-2 imagery. The necessity for such advanced techniques stems from the critical role sea ice plays in the Earth's climate system, influencing ocean circulation, weather patterns, and global temperature regulation. Given the ongoing and rapid changes in sea ice extent due to climate change, reliable and precise measurement methods are essential for both scientific research and policymaking.

The ISPP-net utilizes the SPP module to enable multiscale feature extraction, addressing the shortcomings of conventional single-scale CNNs which often fail to effectively handle the intricate variations in sea ice imagery. By processing features at various scales, ISPP-net enhances its ability to discern both detailed and broader contextual

elements, significantly boosting its performance in segmenting and classifying sea ice features. The increased accuracy in measuring sea ice concentration can refine climate models, leading to improved forecasts of climate conditions. Moreover, the valuable insights derived from this research can contribute to worldwide initiatives aimed at climate change mitigation, in line with the United Nations' Sustainable Development Goal 13: Climate Action.

As for future works, the ISPP-net can be extended and elaborated in several meaningful directions. First, it might be possible to increase the robustness of the model against variable weather and lighting conditions by combining multi-sensor satellite data, e.g., by performing an optical (Sentinel-2) and a radar (Sentinel-1) observation in close time proximity together. Second, the use of transfer learning [42] methods would allow the model to generalize to other areas of the polar world with minimum retraining, including the Antarctic. Third, the synthesis of training examples using generative adversarial networks (GANs) [43], [44] may contribute to better performance of segmenting the sea ice, particularly improving the results on classes that are under-represented during training. Lastly, the real-time deployment of ISPP-net on resource-limited systems, i.e., ship-borne or remote sensing systems, would increase its practical applicability in climate monitoring operations. Such future directions are meant to grow the capacity of ISPP-net further to be much more adaptable, scalable, and relevant beyond its environmental application.

REFERENCES

- [1] S.-K. Kim, H.-J. Lee, J.-S. Kim, S.-H. Kang, E.-J. Yang, K.-H. Cho, Z. Tian, and A. Andrady, "Importance of seasonal sea ice in the western Arctic ocean to the Arctic and global microplastic budgets," *J. Hazardous Mater.*, vol. 418, Sep. 2021, Art. no. 125971, doi: [10.1016/j.jhazmat.2021.125971](https://doi.org/10.1016/j.jhazmat.2021.125971).
- [2] P. Lu, B. Cheng, M. Leppäranta, and Z. Li, "Partitioning of solar radiation in Arctic sea ice during melt season," *Oceanologia*, vol. 60, no. 4, pp. 464–477, Oct. 2018, doi: [10.1016/j.oceano.2018.03.002](https://doi.org/10.1016/j.oceano.2018.03.002).
- [3] A. Keen, E. Blockley, D. A. Bailey, J. B. Debernard, M. Bushuk, S. Delhay, D. Docquier, D. Feltham, F. Massonnet, S. O'Farrell, L. Ponsoni, J. M. Rodriguez, D. Schroeder, N. Swart, T. Toyoda, H. Tsujino, M. Vancoppenolle, and K. Wyser, "An inter-comparison of the mass budget of the Arctic sea ice in CMIP6 models," *Cryosphere*, vol. 15, no. 2, pp. 951–982, Feb. 2021, doi: [10.5194/tc-15-951-2021](https://doi.org/10.5194/tc-15-951-2021).
- [4] A. V. Reyes, A. E. Carlson, J. Clark, L. Guillaume, G. A. Milne, L. Tarasov, E. C. B. Carlson, F. He, M. W. Caffee, K. M. Wilcken, and D. H. Rood, "Timing of cordilleran-laurentide ice-sheet separation: Implications for sea-level rise," *Quaternary Sci. Rev.*, vol. 328, Mar. 2024, Art. no. 108554, doi: [10.1016/j.quascirev.2024.108554](https://doi.org/10.1016/j.quascirev.2024.108554).
- [5] I. Fasoulis, "Governing the oceans: A study into Norway's ocean governance regime in the wake of united nations sustainable development goals," *Regional Stud. Mar. Sci.*, vol. 48, Nov. 2021, Art. no. 101983, doi: [10.1016/j.risma.2021.101983](https://doi.org/10.1016/j.risma.2021.101983).
- [6] L. Niu, X. Tang, S. Yang, Y. Zhang, L. Zheng, and L. Wang, "Detection of Antarctic surface meltwater using Sentinel-2 remote sensing images via U-Net with attention blocks: A case study over the Amery ice shelf," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4301013, doi: [10.1109/TGRS.2023.3275076](https://doi.org/10.1109/TGRS.2023.3275076).
- [7] N. Ivanova, O. M. Johannessen, L. T. Pedersen, and R. T. Tonboe, "Retrieval of Arctic sea ice parameters by satellite passive microwave sensors: A comparison of eleven sea ice concentration algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7233–7246, Nov. 2014, doi: [10.1109/TGRS.2014.2310136](https://doi.org/10.1109/TGRS.2014.2310136).

- [8] A. Kumar, J. Yadav, and R. Mohan, "Global warming leading to alarming recession of the Arctic sea-ice cover: Insights from remote sensing observations and model reanalysis," *Heliyon*, vol. 6, no. 7, Jul. 2020, Art. no. e04355, doi: [10.1016/j.heliyon.2020.e04355](https://doi.org/10.1016/j.heliyon.2020.e04355).
- [9] S. Ali, S. A. M. Mostafa, X. Li, S. Khanjani, J. Wang, J. Foulds, and V. Janeja, "Benchmarking probabilistic machine learning models for Arctic sea ice forecasting," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2022, pp. 4654–4657, doi: [10.1109/IGARSS46834.2022.9883505](https://doi.org/10.1109/IGARSS46834.2022.9883505).
- [10] K. S. Min, M. A. Zulkifley, and N. A. Mohamed Kamari, "A review of automated micro-expression analysis," *Jurnal Kejuruteraan*, vol. 34, no. 5, pp. 763–775, Sep. 2022, doi: [10.17576/jkukm-2022-34\(5\)-02](https://doi.org/10.17576/jkukm-2022-34(5)-02).
- [11] S. Khaleghian, H. Ullah, T. Kræmer, N. Hughes, T. Eltoft, and A. Marinoni, "Sea ice classification of SAR imagery based on convolution neural networks," *Remote Sens.*, vol. 13, no. 9, p. 1734, Apr. 2021, doi: [10.3390/rs13091734](https://doi.org/10.3390/rs13091734).
- [12] L. Istomina, H. Niehaus, G. Heygster, and G. Spreen, "Analysis of the Antarctic sea ice optical properties using high resolution satellite imagery," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2023, pp. 79–82, doi: [10.1109/IGARSS52108.2023.10283195](https://doi.org/10.1109/IGARSS52108.2023.10283195).
- [13] M. A. Zulkifley, A. Saleh, A. Akhyar, and A. H. Saputro, "Multiscale semantic segmentation networks for forest fire mapping," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Feb. 2024, pp. 1–6, doi: [10.1109/IMS61812.2024.10512964](https://doi.org/10.1109/IMS61812.2024.10512964).
- [14] D. Wang, C. Zhang, and M. Han, "MLFC-Net: A multi-level feature combination attention model for remote sensing scene classification," *Comput. Geosci.*, vol. 160, Mar. 2022, Art. no. 105042, doi: [10.1016/j.cageo.2022.105042](https://doi.org/10.1016/j.cageo.2022.105042).
- [15] Z.-A. Yang, N.-R. Zheng, X.-Z. Shi, W.-B. Zhou, and F. Wang, "Precise and fast segmentation of sea ice in high-resolution images based on multiscale and knowledge distillation," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2023, pp. 4946–4949, doi: [10.1109/IGARSS52108.2023.10282651](https://doi.org/10.1109/IGARSS52108.2023.10282651).
- [16] A. A. Tiurin, M. I. Vorobiev, O. I. Lisov, A. M. Andrianov, and E. S. Yanakova, "An effective algorithm for analysis and processing of satellite images for semantic segmentation," in *Proc. IEEE Conf. Russian Young Researchers Electr. Electron. Eng. (EIConRus)*, Jan. 2020, pp. 2018–2022, doi: [10.1109/EIConRus49466.2020.9039113](https://doi.org/10.1109/EIConRus49466.2020.9039113).
- [17] E. Elizar, M. A. Zulkifley, R. Muharar, M. H. M. Zaman, and S. M. Mustaza, "A review on multiscale-deep-learning applications," *Sensors*, vol. 22, no. 19, p. 7384, Sep. 2022, doi: [10.3390/s22197384](https://doi.org/10.3390/s22197384).
- [18] T. Liu, T. Chen, R. Niu, and A. Plaza, "Landslide detection mapping employing CNN, ResNet, and DenseNet in the three Gorges reservoir, China," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 11417–11428, 2021, doi: [10.1109/JSTARS.2021.3117975](https://doi.org/10.1109/JSTARS.2021.3117975).
- [19] L. Balkenende, J. Teuwen, and R. M. Mann, "Application of deep learning in breast cancer imaging," *Seminars Nucl. Med.*, vol. 52, no. 5, pp. 584–596, Sep. 2022, doi: [10.1053/j.semnuclmed.2022.02.003](https://doi.org/10.1053/j.semnuclmed.2022.02.003).
- [20] M. A. Zulkifley, A. F. Munir, M. E. A. Sukor, and M. H. M. Shafiai, "A survey on stock market manipulation detectors using artificial intelligence," *Comput., Mater. Continua*, vol. 75, no. 2, pp. 4395–4418, 2023, doi: [10.32604/cmc.2023.036094](https://doi.org/10.32604/cmc.2023.036094).
- [21] B. Dowden, O. De Silva, W. Huang, and D. Oldford, "Sea ice classification via deep neural network semantic segmentation," *IEEE Sensors J.*, vol. 21, no. 10, pp. 11879–11888, May 2021, doi: [10.1109/JSEN.2020.3031475](https://doi.org/10.1109/JSEN.2020.3031475).
- [22] C. Zhang, X. Chen, and S. Ji, "Semantic image segmentation for sea ice parameters recognition using deep convolutional neural networks," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 112, Aug. 2022, Art. no. 102885, doi: [10.1016/j.jag.2022.102885](https://doi.org/10.1016/j.jag.2022.102885).
- [23] J. Zhao, L. Chen, J. Li, and Y. Zhao, "Semantic segmentation of sea ice based on U-Net network modification," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2022, pp. 1151–1156, doi: [10.1109/ROBIO55434.2022.10011899](https://doi.org/10.1109/ROBIO55434.2022.10011899).
- [24] M. M. Stofa, S. R. Abdani, A. M. Moubark, M. A. A. M. Zainuri, A. A. Ibrahim, N. A. M. Kamari, and M. A. Zulkifley, "Recent developments of artificial intelligence methods for sea ice concentration monitoring using high-resolution imaging datasets," *Ecol. Informat.*, vol. 87, Jul. 2025, Art. no. 103132, doi: [10.1016/j.ecoinf.2025.103132](https://doi.org/10.1016/j.ecoinf.2025.103132).
- [25] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual U-Net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, May 2018, doi: [10.1109/LGRS.2018.2802944](https://doi.org/10.1109/LGRS.2018.2802944).
- [26] A. Singh, H. Kalke, M. Loewen, and N. Ray, "River ice segmentation with deep learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 7570–7579, Nov. 2020, doi: [10.1109/TGRS.2020.2981082](https://doi.org/10.1109/TGRS.2020.2981082).
- [27] X. Zhang, J. Jin, Z. Lan, C. Li, M. Fan, Y. Wang, X. Yu, and Y. Zhang, "ICENET: A semantic segmentation deep network for river ice by fusing positional and channel-wise attentive features," *Remote Sens.*, vol. 12, no. 2, p. 221, Jan. 2020, doi: [10.3390/rs12020221](https://doi.org/10.3390/rs12020221).
- [28] X. Zhang, Y. Zhou, J. Jin, Y. Wang, M. Fan, N. Wang, and Y. Zhang, "ICENETv2: A fine-grained river ice semantic segmentation network based on UAV images," *Remote Sens.*, vol. 13, no. 4, p. 633, Feb. 2021, doi: [10.3390/rs13040633](https://doi.org/10.3390/rs13040633).
- [29] S. R. Abdani, M. A. Zulkifley, and N. Hani Zulkifley, "Analysis of spatial pyramid pooling variations in semantic segmentation for satellite image applications," in *Proc. Int. Conf. Decis. Aid Sci. Appl. (DASA)*, Dec. 2021, pp. 397–401, doi: [10.1109/DASA53625.2021.9682339](https://doi.org/10.1109/DASA53625.2021.9682339).
- [30] D. Murashkin and A. Frost, "Arctic sea ICE mapping using Sentinel-1 SAR scenes with a convolutional neural network," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2021, pp. 5660–5663, doi: [10.1109/IGARSS47720.2021.9553206](https://doi.org/10.1109/IGARSS47720.2021.9553206).
- [31] J. Karvonen, "Baltic sea ice concentration estimation from C-band dual-polarized SAR imagery by image segmentation and convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4301411, doi: [10.1109/TGRS.2021.3097885](https://doi.org/10.1109/TGRS.2021.3097885).
- [32] T. Feng, X. Liu, and R. Li, "Super-resolution-aided sea ice concentration estimation from AMSR2 images by encoder-decoder networks with Atrous convolution," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 962–973, 2023, doi: [10.1109/JSTARS.2022.3232533](https://doi.org/10.1109/JSTARS.2022.3232533).
- [33] J. Sheikh, F. Farahnakian, F. Farahnakian, and J. Heikkonen, "Ice-water segmentation using deep convolutional neural network-based fusion approach," in *Proc. 28th Int. Conf. Autom. Comput. (ICAC)*, Aug. 2023, pp. 1–6, doi: [10.1109/icac57885.2023.10275247](https://doi.org/10.1109/icac57885.2023.10275247).
- [34] N. Balasooriya, B. Dowden, J. Chen, O. De Silva, and W. Huang, "In-situ sea ice detection using DeepLabv3 semantic segmentation," in *Proc. OCEANS*, Sep. 2021, pp. 1–7, doi: [10.23919/OCEANS44145.2021.9705801](https://doi.org/10.23919/OCEANS44145.2021.9705801).
- [35] N. M. Alsharay, Y. Chen, O. A. Dobre, and O. De Silva, "Improved sea-ice identification using semantic segmentation with raindrop removal," *IEEE Access*, vol. 10, pp. 21599–21607, 2022, doi: [10.1109/ACCESS.2022.3150969](https://doi.org/10.1109/ACCESS.2022.3150969).
- [36] J. Kang, F. Tong, X. Ding, S. Li, R. Zhu, Y. Huang, Y. Xu, and R. Fernandez-Beltran, "Decoding the partial pretrained networks for sea-ice segmentation of 2021 Gaofen challenge," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4521–4530, 2022, doi: [10.1109/JSTARS.2022.3180558](https://doi.org/10.1109/JSTARS.2022.3180558).
- [37] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, 2015, pp. 234–241, doi: [10.1007/978-3-319-24574-4_28](https://doi.org/10.1007/978-3-319-24574-4_28).
- [38] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017, doi: [10.1109/TPAMI.2016.2644615](https://doi.org/10.1109/TPAMI.2016.2644615).
- [39] G. Li, I. Yun, J. Kim, and J. Kim, "DABNet: Depth-wise asymmetric bottleneck for real-time semantic segmentation," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, Nov. 2019, p. 259.
- [40] C. Li and P. Ji, "TernausNet-based segmentation of cerebral infarction in magnetic resonance images," *J. Radiat. Res. Appl. Sci.*, vol. 16, no. 3, Sep. 2023, Art. no. 100619, doi: [10.1016/j.jrras.2023.100619](https://doi.org/10.1016/j.jrras.2023.100619).
- [41] X. Gao, T. Chen, R. Niu, and A. Plaza, "Recognition and mapping of landslide using a fully convolutional DenseNet and influencing factors," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 7881–7894, 2021, doi: [10.1109/JSTARS.2021.3101203](https://doi.org/10.1109/JSTARS.2021.3101203).
- [42] J. Hao, "Deep learning-based medical image analysis with explainable transfer learning," in *Proc. Int. Conf. Comput. Eng. Distance Learn. (CEDL)*, Jun. 2023, pp. 106–109, doi: [10.1109/CEDL60560.2023.00029](https://doi.org/10.1109/CEDL60560.2023.00029).
- [43] D. K. Vishwakarma, "Comparative analysis of deep convolutional generative adversarial network and conditional generative adversarial network using hand written digits," in *Proc. 4th Int. Conf. Intell. Comput. Control Syst. (ICICCS)*, May 2020, pp. 1072–1075, doi: [10.1109/ICICCS48265.2020.9121178](https://doi.org/10.1109/ICICCS48265.2020.9121178).
- [44] M. A. Zulkifley, S. R. Abdani, and N. H. Zulkifley, "COVID-19 screening using a lightweight convolutional neural network with generative adversarial network data augmentation," *Symmetry*, vol. 12, no. 9, p. 1530, Sep. 2020, doi: [10.3390/sym12091530](https://doi.org/10.3390/sym12091530).



MARZURAIKAH MOHD STOFA received the bachelor's and master's degrees in electrical and electronic engineering from the Universiti Kebangsaan Malaysia (UKM), where she is currently pursuing the Ph.D. degree with the Department of Electrical, Electronic and Systems Engineering. Her research interests include deep learning applications in remote sensing, such as sea ice monitoring, ship detection, and forest fire mapping.



MOHAMAD AIZUDDIN ISMAIL received the bachelor's degree in electrical and electronic engineering from the Universiti Kebangsaan Malaysia. His research interests include deep learning, computer vision, and remote sensing applications, particularly in environmental monitoring, such as sea ice concentration measurement using convolutional neural networks.



JAESUNG LEE received the B.S., M.S., and Ph.D. degrees in computer science from Chung-Ang University, Seoul, South Korea, in 2007, 2009, and 2013, respectively. Currently, he is the Head and an Associate Professor with the Department of Artificial Intelligence, Chung-Ang University, where he is also the Chief of the AI/ML Innovation Research Center. He holds two patents. His research interests include classification, feature selection, multilabel learning with information theory, machine learning, multilabel learning, model selection, and neural architecture search applications. He is an Associate Editor of *Earth, Moon, and Planets*.



MOHD ASYRAF ZULKIFLEY (Member, IEEE) received the bachelor's degree in engineering (mechatronics) from International Islamic University Malaysia and the Ph.D. degree in electrical and electronics from The University of Melbourne. He did his two-year postdoctoral research (deep learning) with the University of Oxford. His main research interest includes deep learning for computer vision applications. He is a member of the Young Scientists Network—Academy of Sciences Malaysia. He has received Turkish Burslari Fellowship, the Top 2% Scientist Award, and the Asian Science Diplomat Award.

...