RESEARCH ARTICLE

Deep Learning Approach for Accurate Segmentation of Sand Boils in Levee Systems

MANISHA PANTA[1,2], MD. TAMJIDUL HOQUE [1,2], KENDALL N. NILES [3], JOE TOM[3], MAHDI ABDELGUERFI[1,2], AND MAIK FALANAGIN[4]

[1]Canizaro Livingston Gulf States Center for Environmental Informatics, The University of New Orleans, New Orleans, LA 70148, USA
[2]Department of Computer Science, The University of New Orleans, New Orleans, LA 70148, USA
[3]U.S. Army Corps of Engineers, Vicksburg, MS 39183, USA
[4]U.S. Army Corps of Engineers, New Orleans, LA 70118, USA

Corresponding author: Md. Tamjidul Hoque (thoque@uno.edu)

ABSTRACT Sand boils can contribute to the liquefaction of a portion of the levee, leading to levee failure. Accurately detecting and segmenting sand boils is crucial for effectively monitoring and maintaining levee systems. This paper presents SandBoilNet, a fully convolutional neural network with skip connections designed for accurate pixel-level classification or semantic segmentation of sand boils from images in levee systems. In this study, we explore the use of transfer learning for fast training and detecting sand boils through semantic segmentation. By utilizing a pretrained CNN model with ResNet50V2 architecture, our algorithm effectively leverages learned features for precise detection. We hypothesize that controlled feature extraction using a deeper pretrained CNN model can selectively generate the most relevant feature maps adapting to the domain, thereby improving performance. Experimental results demonstrate that SandBoilNet outperforms state-of-the-art semantic segmentation methods in accurately detecting sand boils, achieving a Balanced Accuracy (BA) of 85.52%, Macro F1-score (MaF1) of 73.12%, and an Intersection over Union (IoU) of 57.43% specifically for sand boils. This proposed approach represents a novel and effective solution for accurately detecting and segmenting sand boils from levee images toward automating the monitoring and maintenance of levee infrastructure.

INDEX TERMS Sand boils, levee, segmentation, deep learning, u-net, transfer learning, feature extraction, representation learning.

I. INTRODUCTION

Levees are earthen structures constructed along water bodies that protect numerous commercial and residential properties. They demand regular monitoring and maintenance due to the possibility of being compromised by settling, the risk of overtopping, flood-induced sand boils, seepages, and animal burrowing [1]. Among them, sand boils, also known as sand volcanoes, are one of the most common and dangerous hazards, especially during floods, which occur when the velocity of water flowing from the flood side to the protected side is sufficiently large to erode the soil and allow sand and water to seep through the soil surface. These formations provide evidence for the presence of weak points in a levee [2], [3], indicating the potential for dangerous soil erosion. Hence, it is crucial to detect them precisely to identify new sand boils and monitor their growth for flood-fighting and levee-monitoring purposes during floods. Currently, manual visual inspections of levees are central to monitoring levee systems [2]. which is laborious and costly. This research work introduces the automatic detection of sand boils via advanced image segmentation algorithms using images.

The advancement in deep learning, specifically the use of convolutional neural networks (CNNs), has prompted

The associate editor coordinating the review of this manuscript and approving it for publication was Zhan-Li Sun.

the development of new image segmentation approaches based on Deep learning (DL) [4]. These methods have shown promising results in improving the accuracy and efficiency of semantic segmentation models. However, designing and training deep CNNs for each new segmentation task from scratch can be computationally expensive and time-consuming, especially when the training dataset is limited. Transfer learning [5] has emerged as a promising solution to overcome this challenge by leveraging pretrained models on a related task to improve performance on a new task [6]. Accordingly, this paper introduces a novel, fully convolutional neural network by leveraging the power of transfer learning and advanced DL techniques to locate and segment sand boils from the images of levee systems.

This study aims to detect and segment sand boil regions using an enhanced version of a Fully Convolutional Neural Network [7] (FCN) as an encoder-decoder architecture. We propose a novel transfer learning-based semantic segmentation algorithm for sand boil segmentation in the levee systems. The central hypothesis is that it is possible to use a sizeable pretrained model without possibly increasing the size of the model through controlled transfer learning for semantic segmentation. Our proposed approach leverages a pretrained model as a feature extractor, which enhances efficiency and reduces the need for extensive training data. We integrate advanced deep learning techniques into the encoder-decoder-based architecture to address the over-fitting induced by a pretrained model. These strategies collectively enhance the model's ability to capture spatial information and contextual dependencies across multiple scales, leading to improved segmentation performance. Therefore, the following are the significant contributions of this research paper.

- Introduction of an annotated dataset of sand boil images available for semantic segmentation tasks.
- Comparative analysis of CNN-based state-of-the-art semantic segmentation algorithms on the proposed sand boil image dataset.
- A proposed architectural design that features a pretrained model as an integrated feature extractor for encoder blocks to improve efficiency and reduce extensive training data needs.
- A proposed controlled transfer learning approach that incorporates a pyramidal pooling channel spatial attention model and Principle Component Analysis (PCA) in a parallel manner, followed by a residual connection for facilitating better information flow between layers.
- An approach for fine-tuning the bottleneck layer that makes the model robust and generalizable. Moreover, ablation studies demonstrate that the model with this transfer learning approach adapts to learned representation from the source dataset to align fast with the target task.

The experimental evaluations indicate that the proposed algorithm accurately segments sand boils and outperforms state-of-the-art semantic segmentation algorithms. Overall, the proposed architecture and the building blocks contribute

to the field by demonstrating the potential of deep learning and semantic segmentation for sand boil detection in levee systems and providing a more accurate and efficient approach to this critical task.

II. BACKGROUND AND RESEARCH GAP
A. SEMANTIC SEGMENTATION
Semantic segmentation is a process that involves object detection and the allocation of a semantic label or category to each pixel of objects. Semantic segmentation algorithms provide a detailed understanding of the context through pixel-level analysis of the images. In recent years, deep learning-based semantic segmentation methods have achieved significant breakthroughs due to the progress of large datasets, powerful computing power, and optimization algorithms. Most state-of-the-art architectures for semantic segmentation are based on Convolutional Neural Networks [8] to extract a meaningful representation of objects from the images. The existing deep learning algorithms have shown increased accuracy in various application domains, ranging from biomedical imaging [4], autonomous driving [4], [9], scene understanding [4], and remote sensing operations [10] in comparison to the traditional segmentation methods relying on mathematical and statistical approaches and manual feature engineering [11], [12], [13].

FCNs [7] are the foundation of modern encoder-decoder-based successful deep learning models for semantic segmentation that modify the structure of CNNs and other networks by replacing fully connected layers with convolutional layers to generate a segmentation mask of the same size as the input. Another architecture, SegNet [14], is an encoder-decoder-based architecture that uses pooling indices from the encoder to upsample feature maps in the decoder to improve segmentation, preserving high-resolution information. Likewise, PSPNet [15] uses the pretrained ResNet101 [16] as the feature extraction layer and introduces the pyramid pooling module on top of the encoder to integrate global contextual information by pooling features at different scales. U-Net [17], on the other hand, is a widely popular architecture, especially in medical image segmentation tasks with challenging and small datasets, where overfitting is a common problem. U-Net is an encoder-decoder architecture with connections between corresponding encoder and decoder blocks, facilitating high-resolution features combined with low-resolution contextual information. U-Net++ [18] improves on U-Net through nested and dense skip connections that promote deep supervision without increasing the depth of U-Net architecture. VNet [19] is similar to U-Net, using 3D convolutional layers, and is used for 3D volumetric image segmentation.

Furthermore, to optimize the performance of semantic segmentation architectures, DeepLabv2 and DeepLabv3 [20], [21] introduced the Atrous Spatial Pyramid Pooling (ASPP) model that applies atrous convolution to gather multi-scale information and reduces computation instead of

using fully connected layers. Attention U-Net, proposed by Oktay et al. in [22], is designed to help the model focus on more relevant image regions during the segmentation process. Attention U-Net extends the original architecture by incorporating attention gates to enhance the model's ability to focus on relevant image areas during segmentation. This approach can lead to improved performance, especially in cases where the objects of interest are small or have a complex background. MultiResUNet [23] extends the U-Net architecture to more efficiently capture multi-scale features. It features multi-resolution blocks in both the encoder and decoder paths and employs residual connections to facilitate the flow of gradients during training.

The state-of-the-art encoder-decoder-based architectures have shown improved segmentation results on challenging datasets consisting of small, irregularly shaped objects located against complex backgrounds or with poor quality. They can capture high-level semantic information and fine-grained details of the input images. Typically, an encoder-decoder-based architecture has an encoder that compresses the input image into a lower-dimensional feature representation and a decoder that reconstructs a segmentation map from the compressed representation. In light of sand boil identification and segmenting different sizes against complex backgrounds in images being our problem domain, applying this approach is suitable to achieve successful results.

B. TRANSFER LEARNING

Transfer learning is a technique used to apply the knowledge gained by pretrained models on a large dataset from the source domain to adapt learning in the target domain. This method is beneficial when dealing with limited annotated data or when the target domain significantly differs from the source domain. In recent years, transfer learning has been widely employed in semantic segmentation tasks to address the challenges of insufficient training data and enhance models' generalization capabilities [6].

Transfer learning techniques commonly include using a pretrained model as a feature extraction backbone or fine-tuning the pretrained model. Pretrained models are deep neural networks trained on large datasets to extract features from images. They can be considered expert feature extractors because they have already learned to identify compelling features in widely divergent images, which can be helpful for other tasks, such as image segmentation. The pretrained model learns to identify multiple image features and stores them as learned weights in the model's layers [24]. At the same time, fine-tuning involves initializing the pretrained model's weights with the learned representations from the source domain and updating them in the target domain using backpropagation. This process can be achieved by training the entire model or freezing some layers and updating only the remaining ones. The choice between them depends on the availability of labeled data and the specific requirements of the target segmentation task. This

paper introduces both techniques for preparing a U-Net-like architecture, combining feature extraction and fine-tuning approaches with learning to segment sand boil regions in the images.

C. LEVEE SYSTEMS AND SAND BOILS

Levee systems are embankments or walls constructed along rivers, lakes, or other bodies of water to protect surrounding areas from flooding. Levees are essential in reducing the risk of the detrimental effects of flooding on human lives, property, and infrastructure. Therefore, the timely monitoring and maintenance of a levee system to identify and locate several deficiencies, like cracks, sand boils, seepages, and animal burrowing, is crucial to minimize the threats of potential levee failure [1]. One fundamental reason that can lead to levee failure is internal soil erosion, as indicated by the formation of sand boils [25] around a levee system.
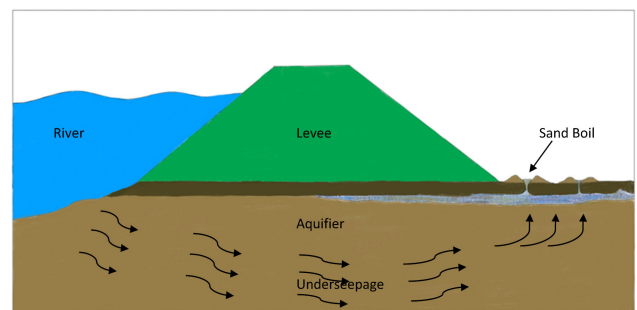

FIGURE 1. Illustration of a cross-sectional view of sand boil formation in a levee system. The high water level in the river forces water to infiltrate through a permeable sand aquifer, resulting in a sand boil on the surface as sand and water emerge from the porous region.

Sand boils are common hazards in a levee system associated with subsurface erosion and under seepage [25]. An illustration of a cross-sectional view of sand boil formation in a levee system is shown in Fig. 1. When hydraulic pressure increases during critical high-water periods, if the seepage velocity is sufficient to initiate and progress piping due to erosion of soil particles, the water seeps through the levee's porous surface. This leads to the accumulation of sand mixed with water into cone-shaped formations within and around the system [25], [26], [27]. A real-world example of a sand boil formed in a levee system is shown in Fig. 2. The emergence of these structures implies the occurrence of internal soil erosion, which in turn highlights potential weak points within a flood control structure or levee. When such erosions cause the piping (or the void) under the levee to expand significantly, it can lead to embankment failure [3]. Hence, accurately detecting sand boils and monitoring their growth becomes crucial for maintaining the integrity of levee systems during floods and minimizing disaster risk.

Traditionally, sand boils are detected through visual inspection, which involves manually investigating the levee for water or sand eruption signs [1] through site visits. This method is time-consuming, labor-intensive, and may

FIGURE 2. Example of sand boil in a levee system collected by the army corps of engineers in the new orleans district.

not detect small, camouflaged sand boils. Another technique involves probes or coring devices inserted into the levee to test for the presence of sand boils [26]. This method is more accurate than just a visual inspection, but it can be expensive and may cause further damage to the levee.

Previous research has explored the application of image processing and computer vision techniques to automate soil identification in order to distinguish it from plants, roots, grains, and other objects [28], [29]. However, traditional methods like manual inspection [1] and aerial surveys [30] have typically been used for sand boil detection Thus far. Although some studies have investigated soil and plant segmentation utilizing deep learning approaches [31], [32], the first attempt to facilitate the levee assessment by automatically locating sand boils in images has been demonstrated in [33]. In [33], the authors generated a synthetic image dataset and introduced machine learning-based object detection techniques that are feasible to automate the sand boil detection process. The object detection techniques presented in the [33] identify and locate sand boils with a bounding box, which may not always be feasible since sand boils can have varying sizes and complexity in terms of uneven texture and color between sand boils and surrounding areas, which can make it difficult to distinguish them based on visual cues alone, requiring more sophisticated segmentation techniques that consider the object's underlying geometry.

D. RESEARCH GAP
Deep learning and semantic segmentation for sand boil detection in levee systems offer several advantages over traditional methods. First, it can achieve superior accuracy and efficiency in identifying sand boils, especially when they are small or difficult to notice. Second, this approach can be applied to extensive datasets of levee images, enabling more comprehensive and frequent monitoring of levee systems. Third, it is more cost-effective than conventional methods,

necessitating minimal human intervention and equipment. It also enables the possibility of identifying issues through remote and automated means, thereby mitigating health and safety hazards for field inspectors. Therefore, this study aims to leverage the advantages of deep learning and semantic segmentation techniques to overcome the challenges associated with manually detecting sand boils in levee systems.

Machine learning and deep learning approaches for applications in levee system monitoring are still in their early stages [33], [34], [35]. Especially in [33], the authors investigated machine learning algorithms and proposed a stacking-based algorithm for detecting sand boils from images using a bounding box approach for object detection. The object detection approach provides a bounding region for sand boils and is particularly useful when the precise boundaries of the object are not required. In contrast, the semantic segmentation approach allows for pixel-level identification and precise localization of the sand boil regions, offering a distinct advantage over the traditional bounding box approach. Therefore, this paper showcases the feasibility of using image segmentation techniques for sand boil detection. Although end-to-end image semantic segmentation architectures, specifically FCNs [7] and U-Net variants, have been successful in medical image segmentation [36], autonomous driving [4], [9], scene understanding [4], and remote sensing operations [10], their application in levee monitoring systems has also gained attention in recent years [35]. Recent studies have shown that U-Net-like architectures have practical applications in levee crack [37] and sinkhole [38] detection.

The U-Net architecture has a unique structure, with an encoding path that captures spatial or contextual information and a symmetric decoding path that enables precise localization. This type of architecture is well-suited to problems that require both context and precision in identifying the location and boundary of the object. U-Net architecture and its variants have shown remarkable performance [39] even with limited training data, accommodating complex and irregular object shapes, textures, and edges–characteristics shared by faults in the levee systems, such as cracks, sand boils, seepages, sinkholes, and animal burrowing. Since obtaining training and evaluation data for these levee system faults is challenging, CNN-based end-to-end image segmentation architectures appear suitable for their detection.

The proposed method in this study addresses the feasibility of using semantic segmentation to detect sand boils in levee systems and presents a lightweight model. Specifically, the proposed method uses an encoder-decoder architecture with skip connections built from ResNet50V2-based encoder blocks to decoder blocks. Furthermore, a lightweight semantic segmentation model efficient in inference time, cost, and compute requirements during training and deployment offers a practical solution for accurately detecting sand boils in resource-constrained environments, including edge devices, low-battery-powered systems, and real-time applications.
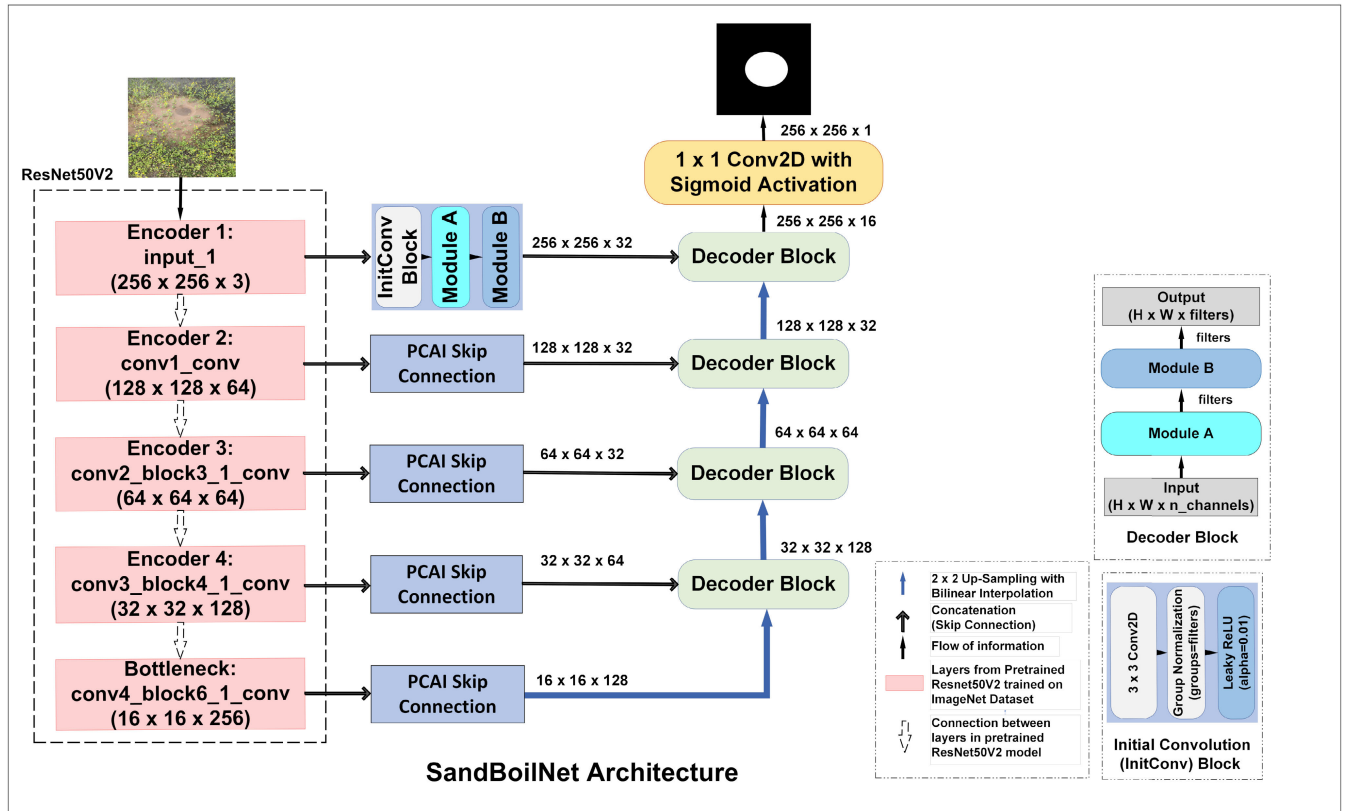
FIGURE 3. Proposed SandBoilNet architecture featuring an encoder-decoder design with skip connection. The architecture comprises three main building blocks: Module A is the proposed inception module shown in Fig. 4. PPCSAttention Block is the proposed attention module displayed in Module B of Fig. 5. PCAI Skip Connection Block, represented in Fig.6, is a skip connection incorporating multi-scale filters and a PCA-based channel-spatial attention module.

III. PROPOSED ARCHITECTURE

The proposed architecture for the sand boil dataset utilizes an encoder-decoder design with a skip connection, incorporating multi-scale filters and a PCA-based channel-spatial attention module to achieve effective feature extraction, as shown in Fig. 3. The key objective is to segment sand boil regions by reducing the number of training parameters and employing controlled transfer learning through partial fine-tuning and feature compression techniques. We use ResNet50V2 [40] model trained on ImageNet [41] dataset. To enhance the model's efficacy in selecting relevant features for the target domain, the architecture integrates domain adaptation via PCA feature representation in addition to integrating improved pyramidal pooling channel-spatial attention module. The architecture comprises three primary building blocks, including an inception-like module, that collectively improve the model's performance in segmenting sand boil regions. Fig. 6 shows the two types of PCA-based skip connection modules used in the architecture. Module D implements a residual connection between two feature recalibration modules applied in parallel to the feature map from selected layers of the ResNet50V2 model, followed by a proposed multi-scale filters inception module for high-dimensional feature representation. In contrast,

Module C represents low-dimensional feature representation from the PCA layer as a residual connection to the output of the parallel branch with attention block and inception module. These blocks represent two different approaches to skip connection from encoder blocks to the corresponding decoder blocks that assist the decoder layers in recovering the fine-grained spatial details. The decoder block in the proposed architecture is simple, with the serialized application of the proposed inception module and attention module to the concatenated feature map of the upsampling path in the decoder block. We discuss details on the remaining building blocks in the subsections below.

A. LEAKYRELU INCEPTION MODULE

Inspired by the Inception family [42], we propose a LeakyReLU inception module that improves the segmentation of sand boil appearing at different scales and orientations by capturing features at multiple scales using different filter sizes in parallel. As represented in Fig. 4, the LeakyReLU inception module takes an input tensor of size [batch_size, height, width, n_channels] along the number of channels for the output tensor, i.e., filters. The module comprises convolution operations with filters of 1×1, 3×3, and 5×5 to obtain multiple-scale features. Each convolution operation is
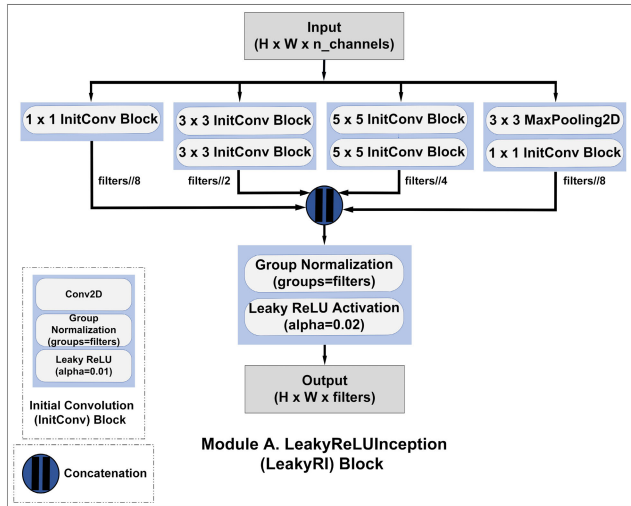
FIGURE 4. Module A is a multi-scale filters-based proposed inception module using GroupNormalization for standardizing the feature map and LeakyReLu activation to add non-linearity.

followed by group normalization and LeakyReLU activation, forming an Initial Convolution (InitConv) Block. To capture additional global features, max-pooling with a filter size of 3×3 and stride of 1 is applied on the input tensor, followed by an InitConv block of filter size 1×1. All the outputs are concatenated along the channel axis to form an output tensor [batch_size, height, width, filters], followed by group normalization and leaky rectified linear unit (LeakyReLU) activation with an alpha of 0.02.

The proposed architecture uses Group Normalization (GN) [43] with a group size equal to the total number of channels employed in the inception module. The intuition behind using GN is mainly because of the assumption that features within a channel are related and meaningful enough to normalize together and the intention of using a small batch size during training. In this module, feature maps from different convolutions are concatenated along the channel dimension to form a larger feature map containing diverse features. When applied, GN standardizes the concatenated feature map for each channel independently but across all spatial locations in that channel. This reduces dependency on the batch size and internal covariate shift and helps ensure that all channels have the same scale.

Additionally, the LeakyReLU activation function is used in the architecture instead of ReLU activation. The ReLU activation function can sometimes lead to dead neurons, especially in the deeper architecture when the input is consistently less than zero; this is also called the "dying ReLU" problem. Using LeakyReLU prevents the "dying ReLU" problem by allowing small negative activation values and improving the model's ability to learn more complex and nuanced features [44]. The combination of group normalization and LeakyReLU allows the model to converge faster and achieve better performance, stability, and robustness, especially when the batch size is small for the sand boil dataset.

B. PYRAMIDAL POOLING CHANNEL-SPATIAL ATTENTION MODULE

Pyramid Pooling Module (PPM) was introduced in [20] to capture global contextual information using custom kernel sizes of 1×1, 2×2, 3×3, and 6×6 strides. The pooling layers of varying scales applied to the input feature map capture relationships and dependencies among different regions of an input feature map. Likewise, Squeeze and Excitation (SE) Block [45] and Convolutional Block Attention Module (CBAM) [46] are two simple yet powerful attention mechanisms that each focus on using pooling and sigmoid layers along with elementwise addition and multiplication operations to emphasize essential features and suppress trivial ones in the network architecture. Inspired by these light-weighted attention mechanisms for CNNs, we propose a novel Pyramidal Pooling Channel-Spatial (PPCS) Attention module in this research, shown in Fig. 5.

In this module, we developed a novel approach combining pyramidal pooling channel attention block and spatial attention for improved performance, as depicted in Fig. 5. The proposed method incorporates varying kernel sizes of 2×2, 4×4, and 8×8 on max-pooling layers to focus on multi-scale spatial details. Additionally, global max-pooling layers are utilized after each layer to acquire channel-wise descriptors, followed by a dense layer with ReLU activation that learn the relevant channel-wise attention weights. Eventually, weighted features from all scales are aggregated through element-wise addition, while sigmoid activation is employed to represent channel-specific information accurately. Lastly, when this refined feature map is multiplied element-wise with the input tensor using an attention map, highlighting the most critical channels would be achieved, leading to enhanced overall results.

C. PCA-BASED FEATURE REPRESENTATION

Principal Component Analysis (PCA) is a widely used dimensionality reduction technique in machine learning and data analysis which helps to map higher dimensional feature space into a lower dimensional while preserving the essential features in the dataset [47]. PCA is a widely-used multivariate statistical technique that identifies and extracts valuable insights from complex datasets. In this study, we propose that integrating PCA into our pretrained model's feature map will enhance its ability to adapt to new domains. Since the pretrained layers generate a reduced feature map shape, PCA is used to compute a new set of orthogonal variables or principal components that represent channels adapting to the sand boil dataset's new domain. Therefore, we decreased the number of channels in the output feature map of the pretrained model to half of the input feature map channels by selecting a subset of principal components with the highest variances. This retained the most significant features of the input images while reducing the channel-wise dimensionality, which helped to reveal underlying patterns and relationships within the batch size of data to facilitate
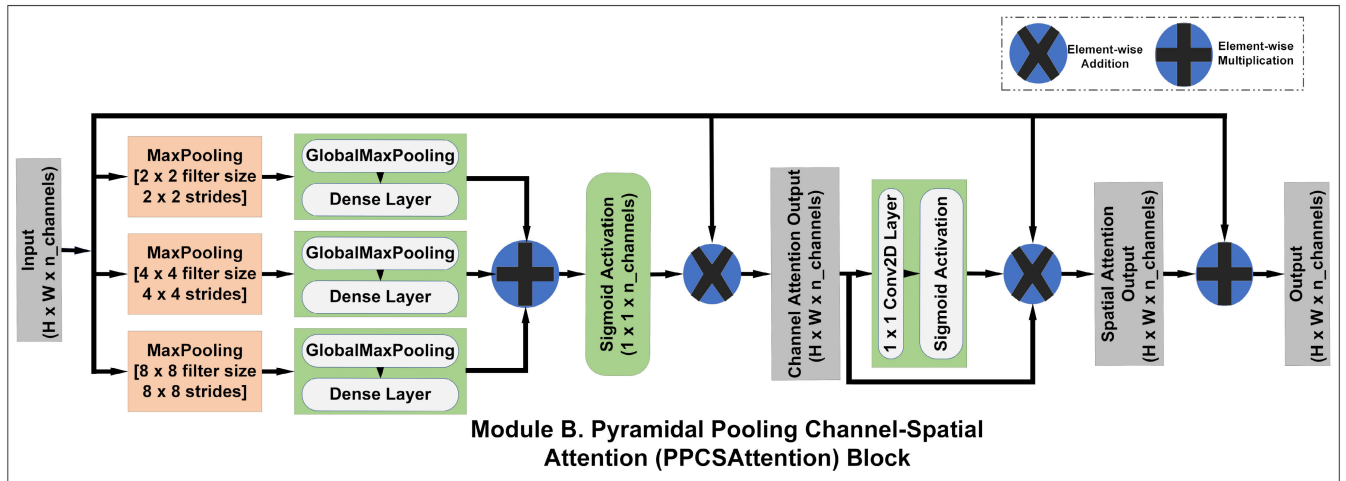
FIGURE 5. Proposed Pyramidal Pooling Channel-Spatial Attention (PPCSAttention) is Module B. The module utilizes max-pooling for multi-scale spatial details, global max-pooling layers to acquire channel-wise descriptors, and a convolution layer to capture important spatial information.

the training. As depicted in Fig. 6, our proposed architecture incorporates a custom-built PCA Layer to facilitate this process effectively.

The PCA layer takes the input tensor from the encoder block with a [batch_size, height, width, n_channels] shape. Initially, we transform this tensor into a flattened array with dimensions of [batch_size, height*width, n_channels]. Each feature vector within this flattened array pertains to a specific location in the original tensor and contains characteristic values for all channels at that particular position. This approach captures correlations between various spatial locations and channels by performing PCA analysis on concatenated feature vectors across all channels in the input tensor. Generating a covariance matrix by centering an input tensor allows the computation of eigenvalues and eigenvectors essential for implementing the PCA technique. PCA on the pretrained model's feature map primarily determines fundamental features while disregarding any extra or redundant ones. Finally, the centered input tensor is projected onto the highest-ranked principal components to achieve a feature map organized by channels, considering their relative contributions towards overall variation as indicated by eigenvalues. The resulting feature vector is subsequently subjected to projection onto half of the new principal components in Module C, whereas in Module D, the feature map is projected to all the principal components.

D. PARTIAL-FINETUNING

This paper introduces partial fine-tuning in the proposed architecture to optimize the learned representation from a pretrained model, ResNet50v2 [40], trained on a large ImageNet dataset [41] (1.4 million images and 1,000 different classes). Both feature extraction and fine-tuning are established transfer learning methodologies frequently employed in image segmentation tasks. We hypothesize that by using the early layers of the pretrained model to extract
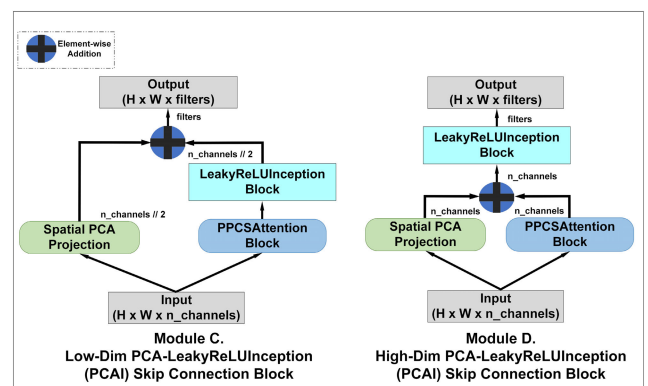

FIGURE 6. Module C and Module D implement a residual connection between spatial PCA projection of pretrained feature map and branch with Pyramidal Pooling Channel Spatial Attention (PPCSAttention) module followed by LeakyReLU inception block.

features and fine-tuning the last few layers, including the bottleneck layer, the overall model will be able to learn on the sand boil dataset while maximizing the features from its existing knowledge base.

In the proposed architecture, all the layers except for the last 48 layers of pretrained ResNet50v2 are fine-tuned, while others are frozen. Lower layers of the pretrained model, which are used as the encoders, are kept frozen for feature extractors, whereas the later layers, which include the bottleneck layer, are fine-tuned. The reason for partial fine-tuning is that the lower layers of the pretrained model are responsible for learning low-level features such as edges, blobs, and corners. The latter layers, closer to the output layer, allow the model to adapt to the task-specific task of sand boil segmentation. Furthermore, partial fine-tuning reduces the potential for overfitting, as adding more fine-tuned layers would increase the number of training parameters in the network [48].

Another essential technique applied during partial fine-tuning is using a low-value learning rate for the optimizer so as not to cause extensive transformations on representations associated with the fine-tuning layer [49]. In addition to this, all the Batch Normalization (BN) layers in the pretrained model are set non-trainable to keep layers in inference mode. When fine-tuning on a sand boil dataset, the data differs significantly from the original ImageNet dataset for training the ResNet50V2 model. This causes batch normalization statistics to be inconsistent, and unfrozen batch normalization will lead to new parameters that do not align with pretrained network optimization. Consequently, freezing batch normalization during fine-tuning prevents parameter conflicts and maintains learned features of the pretrained models without requiring extensive training or facing challenges due to limited data size [50].

IV. DATA AND METHODOLOGY

A. SAND BOIL DATASET

Segmenting sand boils using image segmentation presents two significant challenges, even though sand boils exhibit distinct round/oval shape characteristics. First, the sand boils can be small and intricate, making them difficult to distinguish from the surroundings, as depicted in Fig. 7 (a). Second, various forms of noise can affect images of sand boils, including variations in texture and color between sand boils and their surrounding area, causing difficulties in identifying them solely based on visual cues. The examples shown in Fig. 7 (b), (c), and (d) indicate the second issue. To address these issues, advanced image segmentation algorithms must be developed to handle noisy images and effectively identify sand boils with diverse textures and colors.

The Army Corps of Engineers in the New Orleans District collected a sand boil dataset as part of their levee monitoring efforts. We selected 255 images from this collection for our research, choosing those displaying uniquely identifiable features concerning sand boils against various backgrounds. These images were manually annotated using the VGG Annotator tool [51] using an eclipse shape to identify sand boil in the images. Then binary masks for each image were generated from the exported JSON file with annotations using a Python script. Twenty percent of images with their ground truth were separated as an independent test dataset. Regarding the remaining images, there were not enough for training a deep neural network without encountering overfitting. To address this limitation, we applied 30 augmentation techniques to 204 images using Albumentations package [52]. These techniques included geometric and elastic transformations, channel adjustments, color modifications, and filter operations. The aim was to expand the size of our training and validation datasets in order to enhance their effectiveness during training. These techniques were chosen through an iterative process that involved manual inspection of all the augmented images. We found that
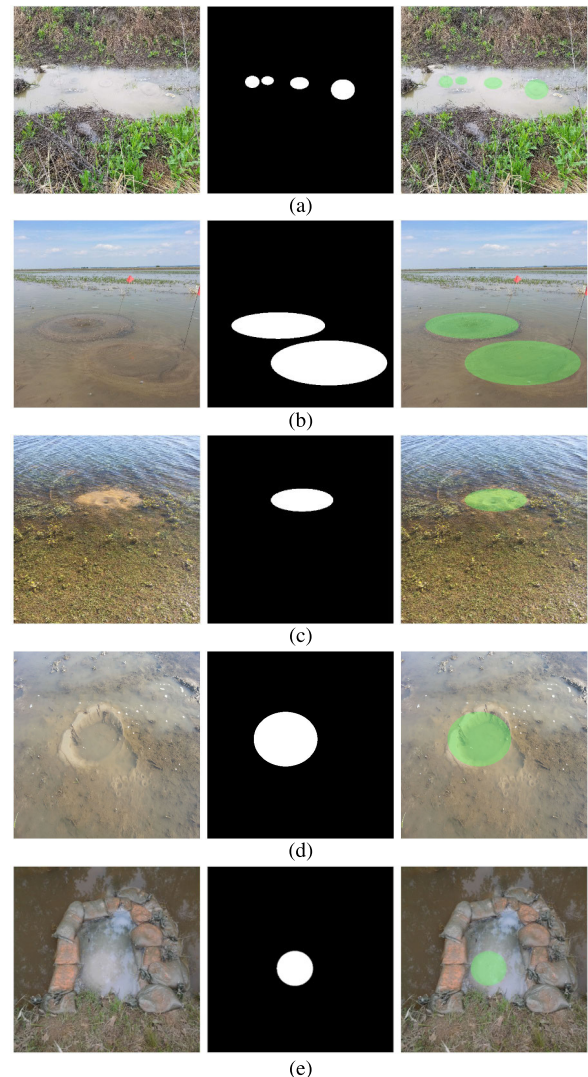

FIGURE 7. Examples from sand boil dataset. Each example (a), (b), (c), (d), and (e) includes three images: the original image, the ground truth, and the ground truth overlaid on the original image. Notably, in example (a), small-sized sand boils are observed in the dried grassland area, while in example (e), a sand boil is seen to be surrounded by sandbags, which serve the purpose of preventing further enlargement.

some augmentation methods, such as random cropping, normalizing, padding, and random snow, did not appear appropriate, so they were discarded, leaving us with 30 augmentations, including geometric transformations, pixel-level modifications, channel transformations, filter operations, and color adjustments, generating 6324 augmented images with their ground truth. While some of these augmentations were applied solely to the images, others were also applied to their corresponding binary masks, provided these techniques did not compromise the integrity of the masks.

In addition, the augmented images were enhanced by including negative images along with corresponding black masks. The inclusion of contextual negative images contributed to more diverse background variations in the dataset, enabling better learning for the model. The negative dataset

includes images of levees without any faults, images of levees with seepages, pothole samples from pothole600 [53] dataset, images of the levee with cracks [37], and animal burrowing, resulting in a total of 6853 training and validation images. To ensure consistency across experiments, we divided our overall dataset into training (74%) and validation (26%) datasets using a shared seed value. Furthermore, all pixel values were normalized by dividing them by 255 and resized to dimensions of 512×512 prior to being fed into the models. By employing this approach, we expanded our training data set extensively while incorporating greater diversity. As a result, our models exhibited improved performance in detecting sand boils.

B. BASELINE AND EXISTING MODELS

A partial fine-tuning-based approach is utilized for the baseline model, as illustrated in Fig. 8. The early layers of ResNet50v2 [40] serve as feature extractors, with partial fine-tuning applied to the layers at and beyond the bottleneck layer. Establishing a baseline model is crucial for developing the proposed model. It enables establishing performance benchmarks as a reference point for measuring improvements made by proposed models or modifications. By comparing the proposed model with the baseline, we can determine which modifications or enhancements lead to better segmentation results on the sand boil dataset. The baseline model in Fig. 8 has low-level features from the encoders concatenated to the decoder layers via a direct skip connection, denoted as None.

Furthermore, we have a sand boil dataset, eventually informing the development of a robust segmentation model tailored to this specific application. The selection of these U-Net-based models is further supported by their successful application in studies such as levee crack segmentation [37] and sinkhole detection [38], which demonstrate their relevance in detecting levee deficiencies.

C. METRICS AND LOSS FUNCTION

Accurately detecting and segmenting sand boils is crucial in disaster management, and the performance of deep learning models for this task is highly dependent on selecting appropriate loss functions and evaluation metrics. Consequently, the proposed models for sand boil detection were evaluated based on their ability to accurately locate sand boil and compute overlap scores between the predicted and ground truth masks. The sand boil dataset is highly imbalanced, having approximately ten percent of sand boil pixels and the remaining ninety percent of background pixels. In such real-world scenarios, accuracy alone can be misleading as it will not effectively address the performance of the model in the minority class, as we have for the sand boil dataset. So balanced accuracy (BA) serves well as it focuses on the model's true performance as presented in Equation 5. Furthermore, metrics like Intersection over Union (IoU) in Equation 1 and Dice Coefficient (DC) in Equation 2

measure the similarity between ground truth and predicted segmentation mask. Furthermore, the combination of loss functions – Binary Cross-Entropy (BCE) loss and dice loss – provide a more comprehensive evaluation of the performance of the sand boil detection models. Their use is considered more appropriate than pixel accuracy and loss alone [54].

$$IoU = \frac{|Y_{predicted} \cap Y_{gt}|}{|Y_{predicted} \cup Y_{gt}|} \quad (1)$$

$$Dice\ Coefficient\ (DC) = \frac{2 \cdot |Y_{predicted} \cap Y_{gt}|}{|Y_{predicted}| + |Y_{gt}|} \quad (2)$$

$$Sensitivity\ (TPR) = \frac{TP}{TP + FN} \quad (3)$$

$$Specificity\ (TNR) = \frac{TN}{TN + FP} \quad (4)$$

$$Balanced\ Accuracy\ (BA) = \frac{Sensitivity + Specificity}{2} \quad (5)$$

$$Macro\ F1\ Score\ (MaF1) = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (6)$$

$$BCE\ Loss = -(Y_{gt} \cdot \log(Y_p) + (1 - Y_{gt}) \\ \cdot \log(1 - Y_p)) \quad (7)$$

$$BCE\ Dice\ Loss = \alpha \cdot BCE\ Loss + \beta \cdot Dice\ Loss \quad (8)$$

In these equations, $Y_{predicted}$ represents the predicted sets of pixels, while Y_{gt} represents the ground truth sets of pixels. TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative segmentation of sand boil pixels, respectively. Y_p represents the predicted probability of the sand boil class. Equation 1 computes the Intersection over Union (IoU) metric and measures the overlap between the predicted and ground truth segmentation mask. Equation 2 computes the Dice coefficient computing the similarity between predicted and ground truth masks.

Since the goal is to classify each pixel in an image as belonging to either the sand boil (positive) class or background (negative) class, Equation 6 and Equation 7 compute macro F1 score and binary cross-entropy (BCE) loss, respectively, which are commonly used in binary segmentation tasks. The macro F1 score in Equation 6 provides a balanced measure of the average of precision and recall over all test images, providing a harmonic mean of precision and recall. It helps to evaluate the overall performance of the model across all the test images. On the other hand, the BCE loss function measures the difference between the predicted probabilities and the actual labels of the examples in the training data. The BCE loss is used to train the model to accurately predict the presence or absence of the object of interest in the image.

Similarly, Equation 5 provides an average measure of sensitivity and specificity which represents the ability of the model to correctly classify actual sand boil pixels and background pixels, respectively. The loss function used in this research is Equation 8, which computes a weighted combination of the BCE loss and the Dice loss. Here, $\alpha = 0.5$

TABLE 1. Model statistics. TPs, NTPs, and MS (MB) represent trainable parameters, non-trainable parameters, and model size (in megabytes) for each architecture used in the study, respectively. Similarly, TT (Hrs), IT-CPU (Sec), and IT-GPU (Sec) represent training time in hours, inference time in CPU and GPU in seconds, respectively.

| Models	TPs	NTPs	MS (MB)	TT (Hrs)	IT-CPU (Sec)	IT-GPU (Sec)
U-Net	7,760,097	5,888	91.29	1.53	0.46	0.09
MultiResUnet	7,238,228	24,522	80.05	2.64	1.66	0.17
Attention U-Net	8,903,043	9,728	105.02	2.94	0.71	0.09
U-Net++	7,238,228	24,522	107.99	2.85	1.07	0.12
Baseline	1,053,825	7,457,216	42.01	0.07	0.39	0.09
SandBoilNet-CBAM	1,583,287	7,546,368	49.81	1.20	1.19	0.17
SandBoilnet-SE	1,505,573	7,546,368	48.72	1.02	0.94	0.22
SandBoilNet-No-PCA	1,819,182	7,456,256	52.33	0.17	0.88	0.15
SandBoilNet-Low-Dim	1,716,446	7,501,312	51.13	2.01	0.92	0.16
SandBoilNet-High-Dim	1,819,182	7,456,368	52.69	2.78	0.94	0.17
Baseline-Conv	2,234,497	7,458,240	55.87	0.62	0.43	0.09
Baseline-LeakyRI	764,035	7,456,256	47.22	0.96	0.49	0.10
Baseline-CBAM-LeakyRI	866,842	7,456,256	40.6	1.20	0.77	0.13
Baseline-SE-LeakyRI	792,003	7,456,256	39.55	1.02	0.56	0.18
Baseline-PPCSAttention	1,102,707	7,456,256	43.49	1.19	0.62	0.13
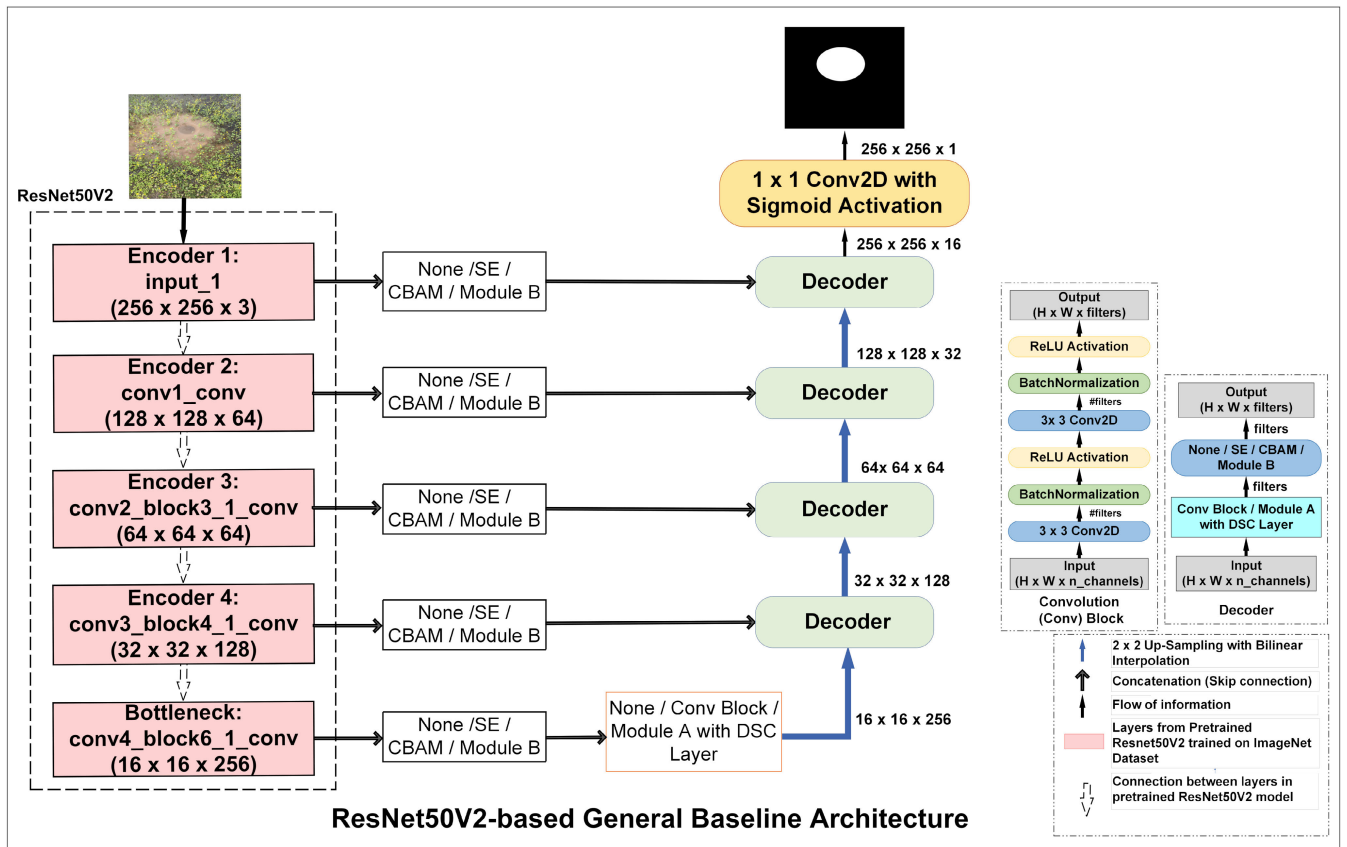
FIGURE 8. Baseline architecture with layers from pretrained ResNet50V2 model as feature extractor with either SE Block or CBAM Block or proposed attention block followed by low parameterized LeakyRI or no attention block followed by either Conv block or LeakyRI block applied as skip connection.

and $\beta = 0.5$ are weights provided to BCE loss and dice loss, respectively, where $\alpha + \beta = 1$. The BCE-Dice loss function encourages the model to segment the object while

maintaining well-defined boundaries accurately. The Dice loss helps ensure that the model's predictions are similar to the ground truth mask. In contrast, the BCE loss helps

minimize false positives and false negatives by assigning a higher penalty to these false predictions. Using the BCE-Dice loss can improve the accuracy of the model's predictions and make them more useful in practical applications.

D. EXPERIMENTAL SETUP

We built the segmentation models using Keras [55] and trained them on a single NVIDIA A100 GPU. For evaluation of use of computational resources, the training and inference time in the same GPU is recorded along with the inference performance of models in the CPU (Intel Xeon Processor E5-2620) as well as observed in Table 1. We initialized the convolutional layers using the He Initialization [56] method with the same seed value to ensure consistency of initial training parameters and training and validation datasets across all models. This approach allowed us to effectively train robust segmentation models for sand boil detection, demonstrating high accuracy and generalization ability on the independent test dataset. The training process of the segmentation models involved minimizing a weighted BCE Dice loss, as given in Equation 8, using an Adam optimizer. The models were trained for a maximum of 200 epochs with a batch size of 8 and an initial learning rate of 4e-4. Early stopping was implemented after eight epochs to prevent overfitting, and the learning rate decayed by 0.06 every six epochs when the validation loss plateaued. Our primary goal was to segment sand boils accurately. Thus, the model with the highest Dice Coefficient (DC) on the validation dataset was saved for evaluation on an independent test dataset.
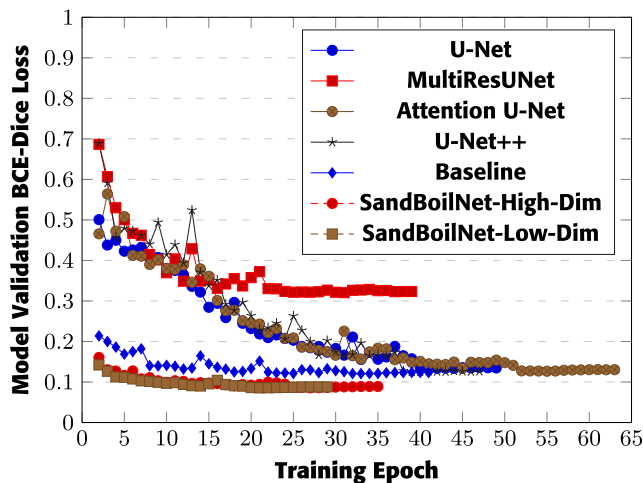

FIGURE 9. BCE-Dice loss on validation data during training epochs for the models. The proposed model, SandBoilNet-Low-Dim, exhibits lower validation loss values over training epochs and smooth learning compared to other models because of controlled transfer learning.

V. RESULTS AND ANALYSIS

This section evaluates the models quantitatively using independent levee sand boil images and qualitatively using the inferences generated by the models. The overall test dataset is used for quantitative and qualitative analysis.

However, in the ablation study, the test dataset is divided into two sections to evaluate the robustness of the models, especially the effectiveness of attention modules. The first dataset component contains test images in which sand boil regions are developed in and near grassland - 30 images with related masks. The second contains 21 test images without grassland. In other words, we are attempting to understand the efficacy of the models in identifying sand boil regions despite the complex and textural backgrounds, such as muddy water, grassland, and mixed ones.

TABLE 2. Metric results of models on levee sand boil dataset. The best metrics results are shown in bold. The model with the highest Intersection over Union (IoU) score is indicated in both bold and underlined.

Models	BA (%)	IoU (%)	DC (%)	MaF1 (%)	TPR (%)	TNR (%)
U-Net	75.00	40.79	53.54	58.47	51.79	98.21
MultiResUNet	73.82	38.14	50.75	56.99	49.85	97.79
Attention U-Net	75.06	40.85	54.30	59.70	51.82	**98.30**
U-Net ++	74.59	37.56	50.73	56.72	51.30	97.89
SandBoilNet-Low-Dim	**85.52**	**57.43**	**70.17**	73.12	**73.28**	97.76
SandBoilnet-High-Dim	85.26	55.59	68.53	**73.25**	73.27	97.25

A. QUANTITATIVE AND QUALITATIVE EVALUATION

The models trained on augmented sand boil images are evaluated using independent test images. Table 1 presents BA, IoU of the sand boil, and Macro F1 score for state-of-the-art and proposed models. Fig. 10 shows the evaluation of the proposed and the state-of-the-art models, U-Net, Attention U-Net, and U-Net++, on independent test images. All the models were evaluated in the same GPU where they were trained. The validation loss graph of each model during training epochs is plotted in Fig. 9. The loss graph indicates that the state-of-the-art models struggle to make accurate predictions on the validation data due to overfitting because of larger training parameters compared to the proposed model. It is evident from the loss chart that baseline and proposed models implementing feature extraction and fine-tuning have more satisfactory learning during training. It also indicates that models that implement some degree of transfer learning converge faster in comparison to those trained from scratch. This suggests a preference towards transfer learning methodologies.

The training time for the proposed model was approximately 1.79 hours on a single NVIDIA A100 GPU. Subsequently, in terms of inference-time computation, generating predicted segmentation masks have an average duration of 0.16 seconds per image with a higher average IoU at around 57.43%. This level of performance is noteworthy, particularly during real-world scenarios where there is a significant disparity between positive and negative pixel
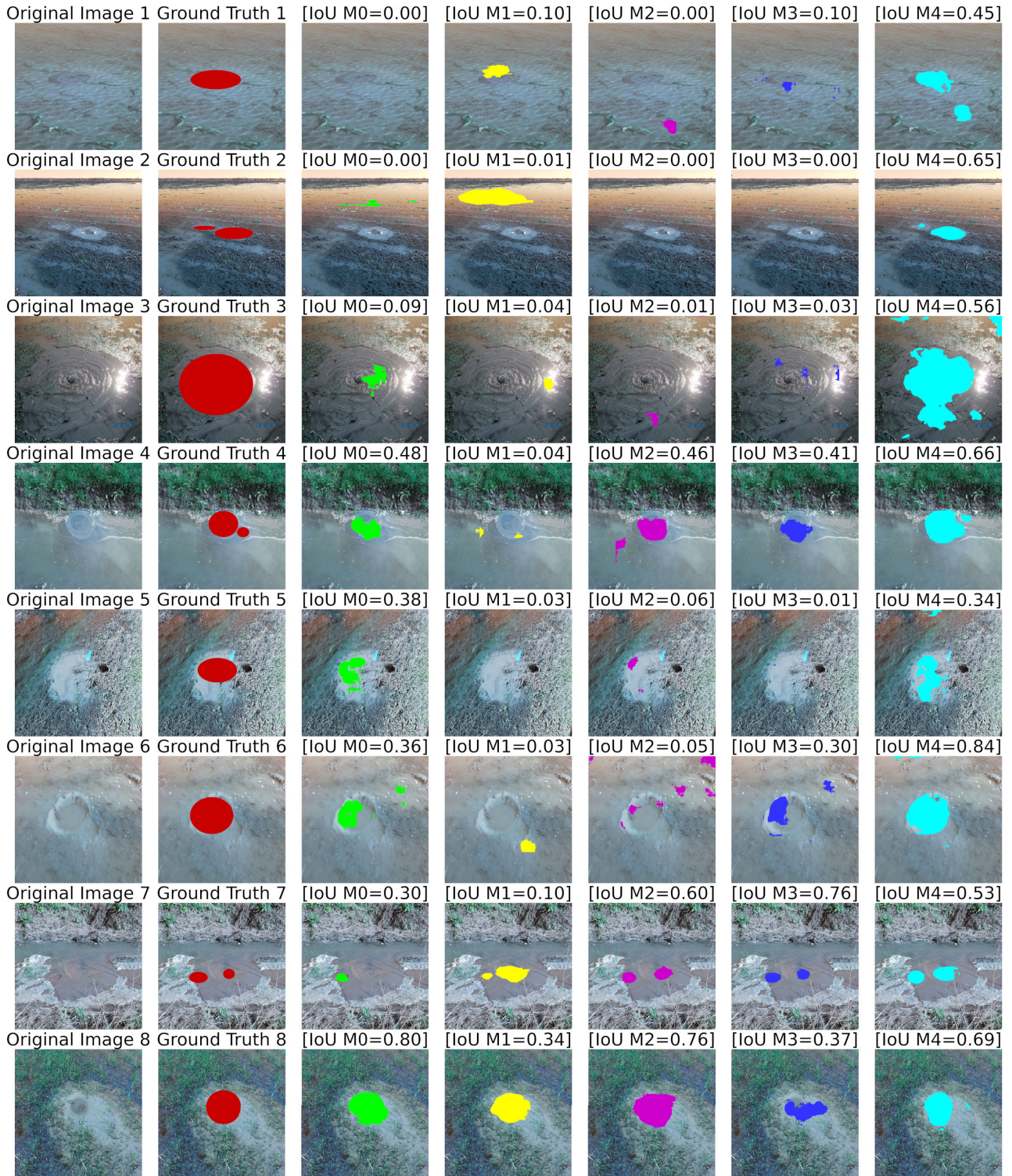
FIGURE 10. Example inferences on test images. Qualitative visual comparison between our proposed model against state-of-the-art models. M1, M2, M3, M4, and M5 represent trained U-Net, MultiResUNet, Attention U-Net, U-Net++, and SandBoilNet-Low-Dim, respectively.

classes in images. Fig. 10 represents the evaluation of the proposed and the state-of-the-art models, U-Net, Attention U-Net, and U-Net++, on independent test images. It shows that models trained from scratch cannot learn the properties

of the sand boil and background with a small dataset and fewer training epochs. The area sand boil regions predicted by existing models are comparatively low compared to the proposed model that uses controlled transfer learning. It can

also be noted that the proposed model identifies the area and boundary of sand boil regions even on a noisy background which is possible through the attention module introduced in the skip connection.

Table 2 shows a quantitative evaluation of models through balanced accuracy (BA), IoU, DC, and Macro F1 scores for the test dataset. The results demonstrate the superior performance of the proposed model. Specifically, the proposed model (SandBoilNet) outperforms U-Net, MultiResUNet, Attention U-Net, and U-Net++ by 28.97%, 33.59%, 28.87%, and 34.60%, respectively, in terms of IoU for sand boil segmentation. This substantial improvement can be attributed to the fine-tuning approach employed in this research paper. By fine-tuning the pretrained weights with the sand boil dataset, the proposed model can adapt and specialize its learned features to the task at hand, resulting in remarkably better performance than the other models.

On average, the models achieve a success rate of fifty-seven percent in locating and segmenting sand boil regions in the levee images. However, due to similarities in color, texture, structure, and porosity between soil and sand boils, the models sometimes mistakenly predict soil pixels as sand boil regions for animal burrow faults within the levee system as shown Fig. 11a. Interestingly, SandBoilNet demonstrated an ability to differentiate between sand boils and seepage regions represented in Fig. 11b, which also consist of soil and water. This distinction may be attributed partially to the circular shape of a typical sand boil, with its soil-like texture resembling that of animal-formed burrows.

Additionally, we evaluated the models based on the inferences on the samples of two separate public test datasets: Pothole600 [53], levee crack test dataset [37] and images of levee with cases of animal burrowing. In these evaluations, we compared the proposed model to state-of-the-art models. The examples for predicting sand boils for out-of-domain tests are illustrated in Fig. 12. Notably, potholes are characterized by smaller, rounded shapes formed in complex backgrounds, whereas cracks are developed in the levee system. The test images with burrows also exhibit similar properties as sand boils can be observed in original image 3 and 8 in Fig. 12. As a result, negative images depicting the formation of potholes on concrete and asphalt surface, burrows and cracks in levees were opted for during model evaluation through inferences.

B. ABLATION STUDY

This section compares advanced deep learning (DL) techniques incorporated within our proposed architecture. Our primary objective is to verify the effectiveness of the combination of building blocks. For this purpose, we evaluated SandBoilNet-Low-Dim against four other configurations. These included a baseline model, SandBoilNet-High-Dim with Squeeze-and-Excitation (SE) block, and a Convolutional Block Attention Module (CBAM) in place of the proposed attention module and SandBoilNet-High-Dim with a feature space same in terms of the number of channels of the

TABLE 3. Metric results of variation of the proposed model with SE block, CBAM block, proposed attention module, without PCA, and low dimensional representation via PCA on the levee sand boil dataset. The best results for each metric are shown in bold. The model with the highest Intersection over Union (IoU) score is indicated in both bold and underlined.

Models	BA (%)	IoU (%)	DC (%)	MaF1 (%)	TPR (%)	TNR (%)
Baseline	73.64	38.89	50.11	56.64	48.95	**98.34**
SandBoilNet-CBAM	82.55	52.36	65.01	69.70	66.95	98.15
SandBoilNet-SE	85.38	55.84	68.67	72.85	73.09	97.67
SandBoilNet-No-PCA	**86.35**	55.50	68.97	72.81	**75.30**	97.39
SandBoilNet-Low-Dim	85.52	**57.43**	**70.17**	73.12	73.28	97.76
SandBoilNet-High-Dim	85.26	55.59	68.53	**73.25**	73.27	97.25

pretrained layer, and SandBoilNet without PCA feature mapping.

Table 3 illustrates the performance of these models using Intersection over Union (IoU), TPR or Recall, and Macro F1 score metrics as key metrics. The results indicate that SandBoilNet-Low-Dim exhibits a superior performance of 57.43% on the independent levee sand boil dataset, thus validating the effectiveness of the proposed architecture. Specifically, when the feature map's channel count is kept the same through PCA, the resulting model, SandBoilNet-PCA-High-Dim, impacts the models' IoU dropping to 55.59%. Likewise, The performance of SandBoilNet-No-PCA is comparatively better in terms of BA and TPR, with 86.35% and 75.30%, respectively; however, IoU is 55.50%. This indicates that the model performs well in terms of overall accuracy and correctly identifying sand boil pixels, but it struggles with precise localization or segmenting them from the background. Overall, the table shows that the model's PCA-based feature presentation helps the model to increase adaptability to the sand boil dataset.

Fig. 13 and Fig. 14 display model inferences on selected independent test images with and without grassland in the background, respectively. It is clear from these figures that SandBoilNet, with the proposed PPCSAttention module, demonstrates outstanding performance in locating and segmenting sand boils, mainly when the background is complex. This suggests that our proposed attention block can effectively harness global information, thereby enabling it to focus on sand boil regions even within complicated environments. Contrarily, the inferences drawn from the SandBoilNet incorporating the SE block and the SandBoilNet without PCA (as presented by B3 and B4, respectively, in Fig. 14) display a certain level of competitiveness with the SandBoilNet featuring our proposed attention module; however, the models encounter difficulties. These challenges become especially prominent when handling complex
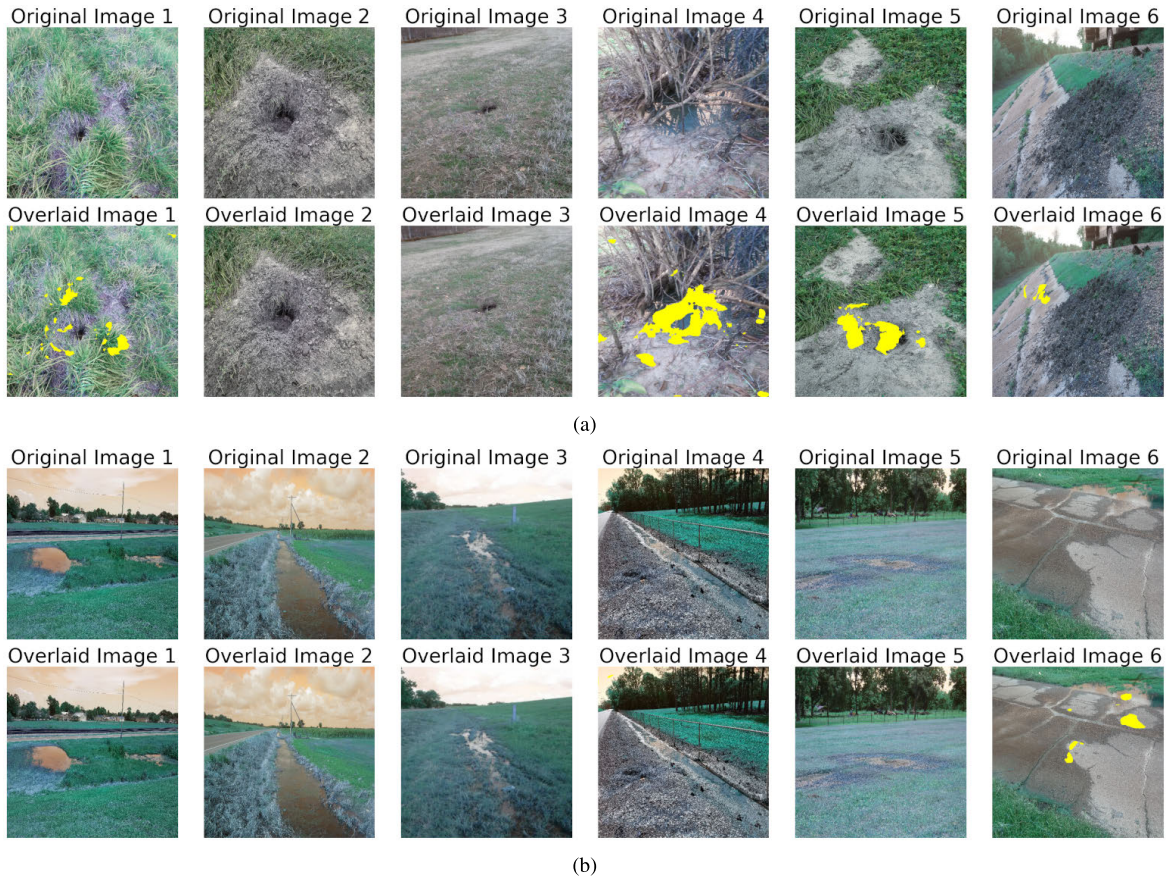
(a)

(b)

FIGURE 11. Examples from the levee system domain. SandBoilNet-Low-Dim model failing to detect other levee-related deficiencies such animal burrowings in (a) and seepages in (b). Each example includes six images: the original image in the first row and predicted segmentation mask overlaid on the original image in the second row.

backgrounds and scenarios where the global context needs to be considered.

TABLE 4. Metric results on the adopted version of SandBoilNet, where the LeakyRI Module is included in the bottleneck layer represented by the model and various attention modules for comparison. The best results for each metric are highlighted in bold. The model with the highest Intersection over Union (IoU) score is indicated in both bold and underlined.

Models	BA (%)	IoU (%)	DC (%)	MaF1 (%)	TPR (%)	TNR (%)
Baseline-Conv [IM0]	80.92	50.54	63.39	68.56	63.98	97.86
Baseline-LeakyRI [IM1]	**84.49**	54.20	67.24	**72.02**	**71.66**	97.33
Baseline-CBAM [IM2]	82.27	50.79	63.15	68.15	66.34	**98.20**
Baseline-SE [IM3]	83.57	53.04	65.77	71.22	69.45	97.68
Baseline-PPCSAttention [IM4]	84.20	**54.70**	**67.83**	71.57	70.57	97.83

In line with our hypothesis, we suggest a fine-tuning approach where only encoder layers adapted from a pre-trained model are set false to trainable, i.e., frozen. The focus of the fine-tuning process is solely on the bottleneck layer. By isolating and prioritizing the fine-tuning of the bottleneck layer, we aim to maximize the learning potential of our model.

To assess the implications of the fine-tuning procedure adopted in this research paper, we integrated a low-parameterized version of the LeakyRI Block into the bottleneck layer, thereby enhancing its adaptability to datasets from the target domain during upsampling. The first adaptation of the SandBoilNet model features no attention modules but includes two convolutional layers followed by Batch Normalization (BN) and ReLU activation after each operation. Meanwhile, all four remaining models include attention modules alongside the low-parameterized version of the LeakyRI module. To obtain a low-parameterized version, the normal convolution of the first block of 3×3 and 5×5 convolutions in the inception module represented in Fig. 4 is replaced by the corresponding filter size of the depthwise separable convolution [57] layer used as DSC layer. These modifications help analyze the significance of parallel convolution branches with multiple-sized filters to capture
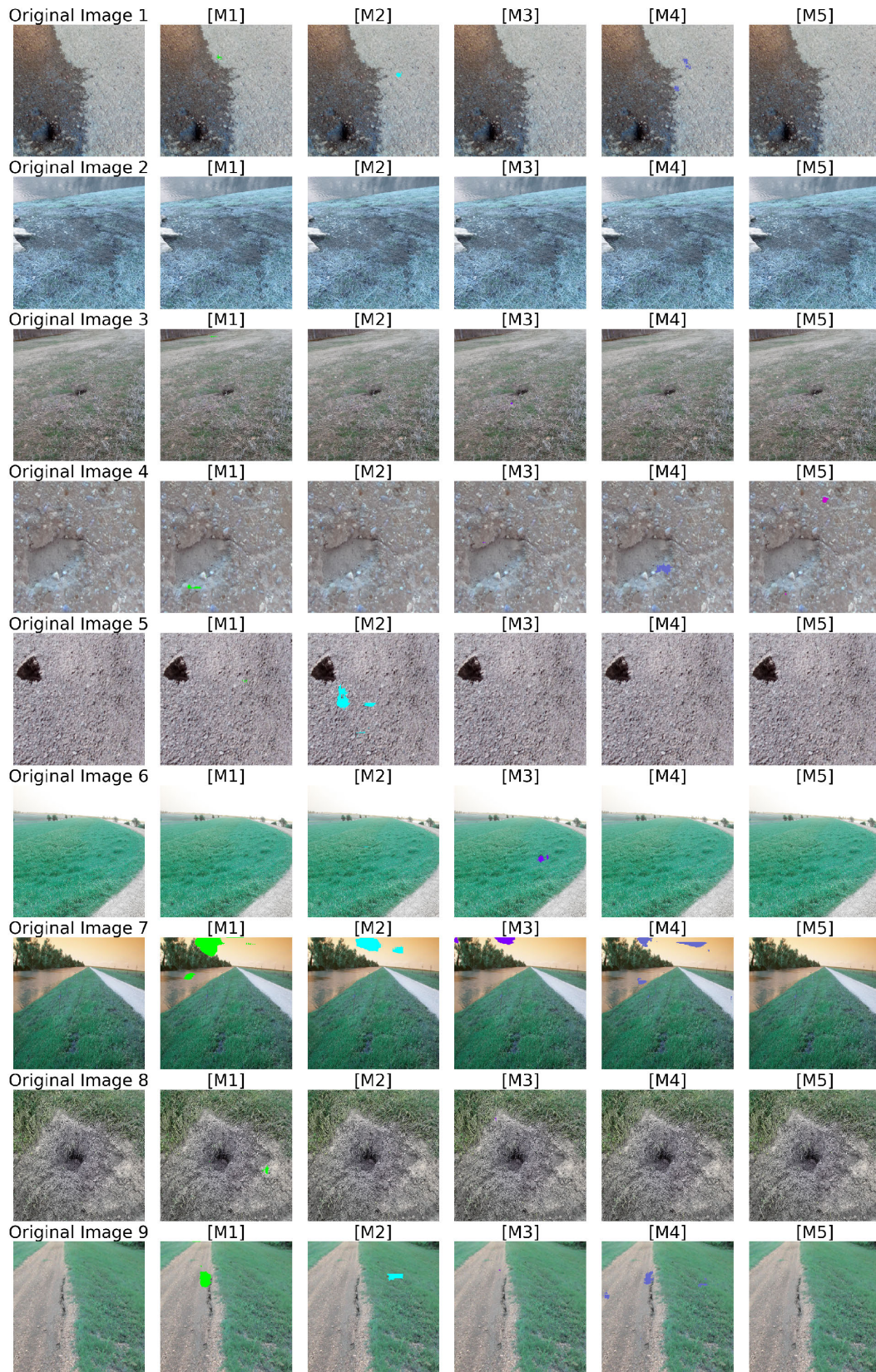
FIGURE 12. Examples from independent out-of-domain negative test dataset. Each example includes six images: the original image and the predicted mask for the original image from M1-U-Net, M2-MultiResUNet, M3-Attention U-Net, M4-U-Net++, and M5-SandBoilNet-Low-Dim, respectively overlaid on the original image.
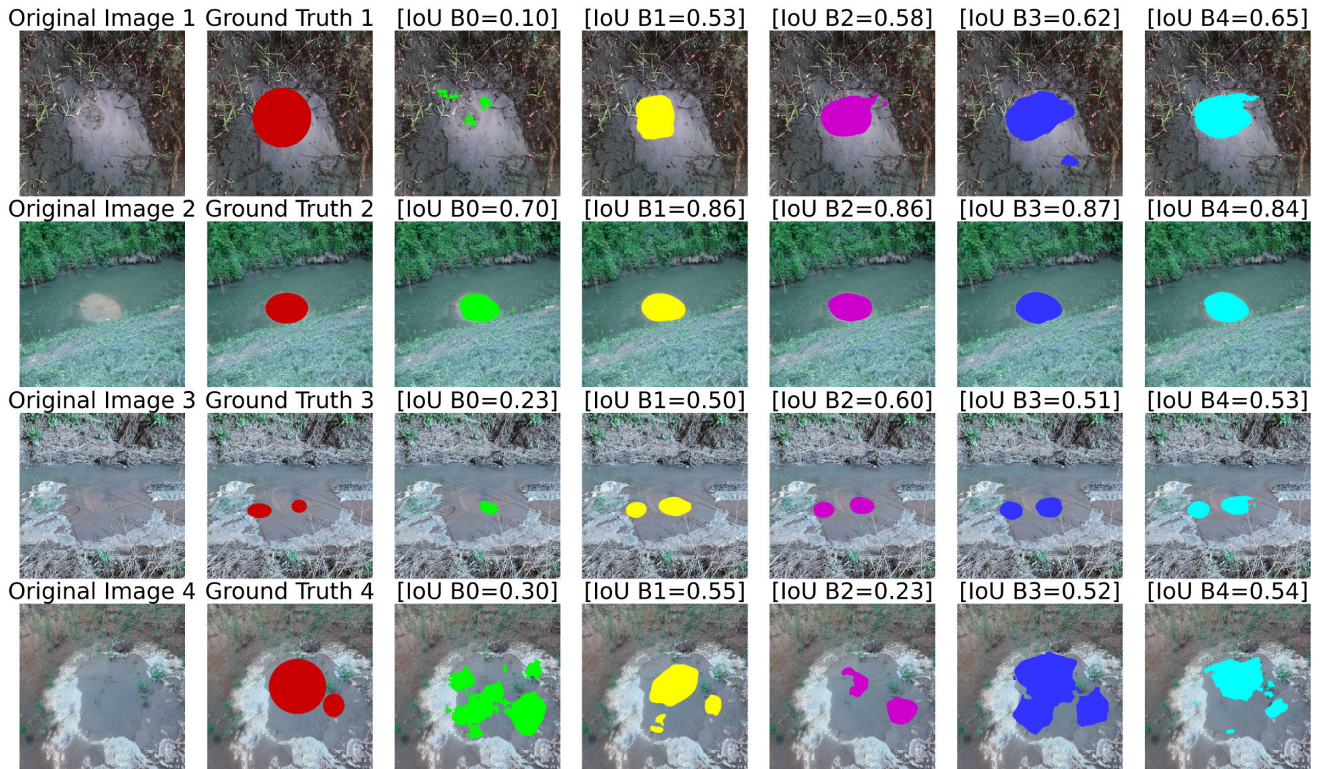
FIGURE 13. Example inferences on test images with grassland as background. A comparison of baseline architecture, SandBoilNet with CBAM or SE or no PCA with proposed PPCSAttention block or best performing proposed SandBoilNet model represented by B0, B1, B2, B3, and B4, respectively.
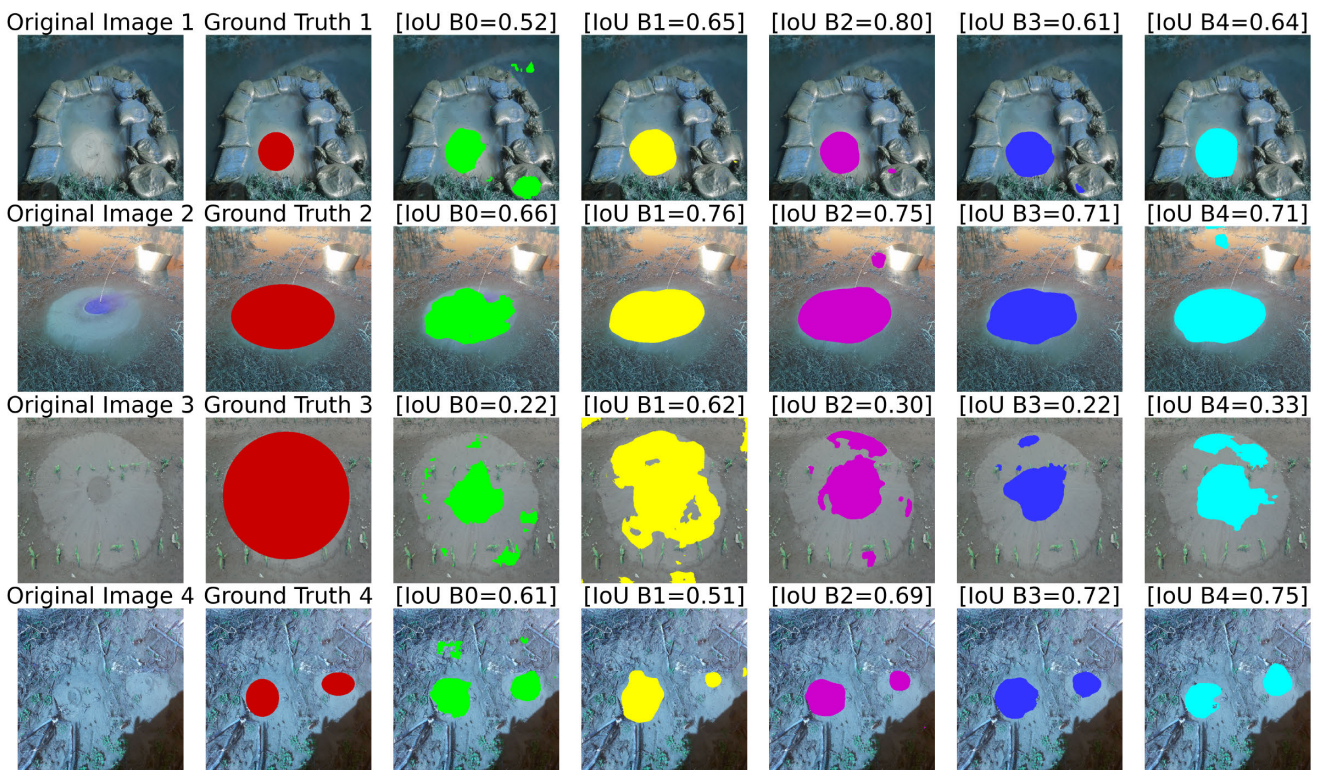

FIGURE 14. Example inference on test images with minimal or no grassland as background. A comparison of baseline architecture, SandBoilNet with CBAM or SE or no PCA with proposed PPCSAttention block or best performing proposed SandBoilNet model represented by B0, B1, B2, B3, and B4, respectively.
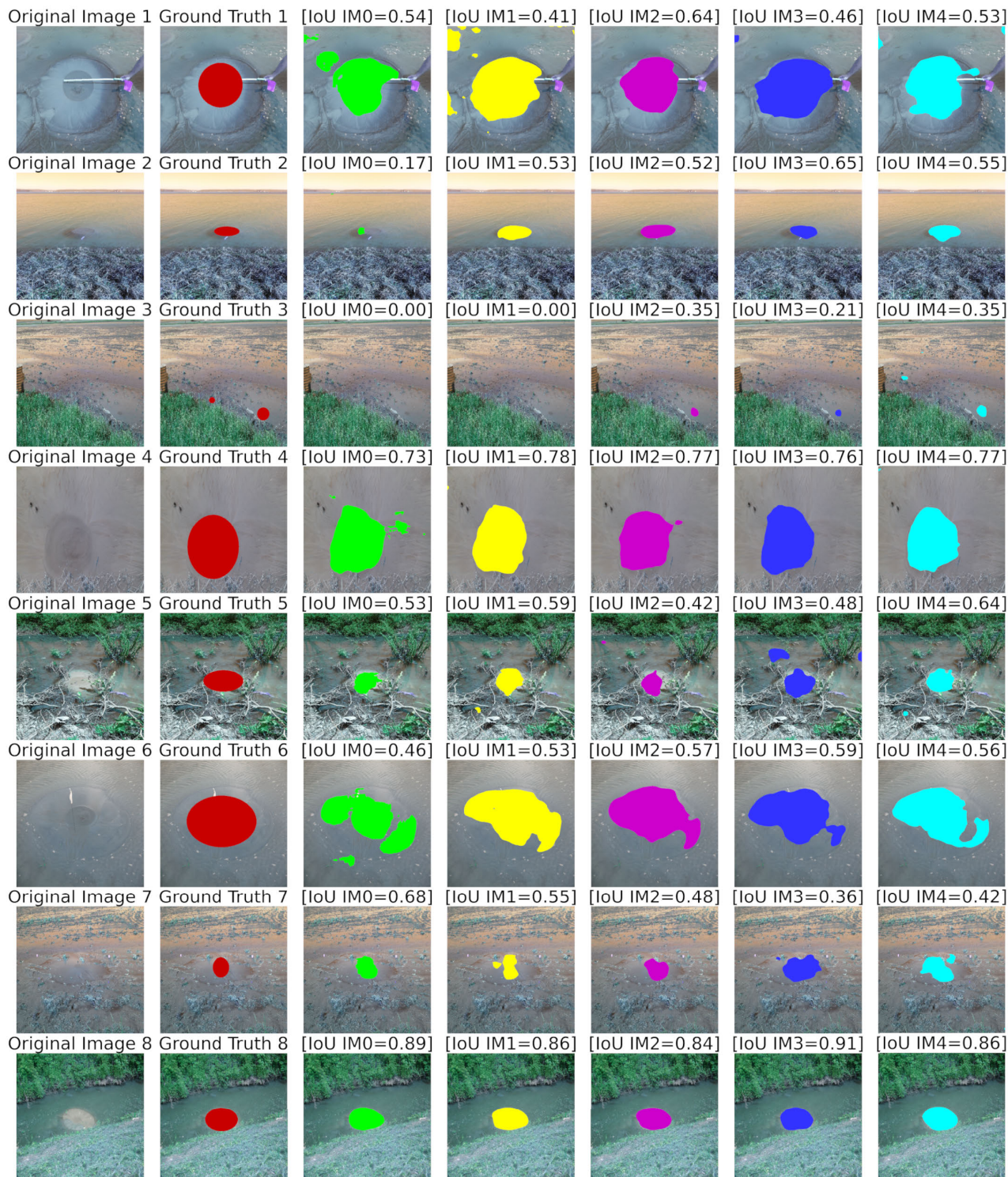
FIGURE 15. Example inference on overall test images. Visual comparison of the adopted version of SandBoilNet where LeakyRI Module is included in the bottleneck layer represented by the Model along with various attention modules for comparison, represented by IM0 for Conv Block instead of LeakyRI, IM1 with LeakyRI, IM2 with CBAM and LeakyRI, IM3 with SE and LeakyRI, and IM4 with proposed PPCSAttention and LeakyRI.

multiscale features, leading to improved performance of models as shown in Table 4.

The utilization of various advanced techniques, including the inception module, attention block, PCA-based feature compression, GroupNormalization, and LeakyReLU activation function, has significantly enhanced the network architecture's capability to perform controlled transfer learning from a pretrained model. Additionally, fine-tuning the

bottleneck layer on the sand boil dataset has facilitated the acquisition of specific features related to the target domain dataset. As depicted in Fig. 15, incorporating a simple yet powerful, pyramidal pooling-based attention module has had an instrumental impact in boosting the model's ability to learn sand boil segmentation during training. These findings highlight the efficacy of our proposed architecture and fine-tuning approach.

VI. CONCLUSION

This paper presents a deep neural network that uses controlled transfer learning to locate and segment sand boil in images. The model combines CNN with depthwise convolution for local information extraction. At the same time, the pyramidal pooling channel spatial attention module is used for global contextual information extraction to segment sand boil precisely. In addition, the proposed model, SandBoilNet, effectively addresses the vanishing gradient problem during training through residual connection comprising of PCA-based transformation of feature maps from the pretrained model and attention module. Our proposed model, SandBoilNet, outperformed CNN-based state-of-the-art methods in sand boil segmentation, confirming its suitability for better levee system monitoring tasks. Furthermore, this work contributes significantly to the field, showcasing the power of deep learning techniques in solving complex real-world problems. Code and data can be found **here**.

ACKNOWLEDGMENT

REFERENCES

[1] H. L. Ellis, C. B. Groves, and G. R. Fischer, "Rapid levee assessment for reliability and risk analysis," in *Proc. GeoCongress, Geosustainability Geohazard Mitigation*, Mar. 2008, pp. 170–177.

[2] K. Richards, B. Doerge, M. Pabst, D. Hanneman, and T. O'Leary, *Evaluation and Monitoring of Seepage and Internal Erosion*, S. Leffel, Ed. Washington, DC, USA: FEMA, 2015.

[3] J. A. Schaefer, T. M. O'Leary, and B. A. Robbins, "Assessing the implications of sand boils for backward erosion piping risk," in *Proc. Geo-Risk*, Jun. 2017, pp. 124–136.

[4] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3523–3542, Jul. 2022.

[5] L. Torrey and J. Shavlik, "Transfer learning," in *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*. Hershey, PA, USA: IGI Global, 2010, pp. 242–264.

[6] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *J. Big Data*, vol. 3, no. 1, pp. 1–40, Dec. 2016.

[7] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[8] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[9] D. Feng, C. Haase-Schütz, L. Rosenbaum, H. Hertlein, C. Gläser, F. Timm, W. Wiesbeck, and K. Dietmayer, "Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 3, pp. 1341–1360, Mar. 2021.

[10] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS J. Photogramm. Remote Sens.*, vol. 152, pp. 166–177, Jun. 2019.

[11] A. Xu, L. Wang, S. Feng, and Y. Qu, "Threshold-based level set method of image segmentation," in *Proc. 3rd Int. Conf. Intell. Netw. Intell. Syst.*, Nov. 2010, pp. 703–706.

[12] C. Cigla and A. A. Alatan, "Region-based image segmentation via graph cuts," in *Proc. 15th IEEE Int. Conf. Image Process.*, Oct. 2008, pp. 2272–2275.

[13] Z. Yu-Qian, G. Wei-Hua, C. Zhen-Cheng, T. Jing-Tian, and L. Ling-Yun, "Medical images edge detection based on mathematical morphology," in *Proc. IEEE Eng. Med. Biol. 27th Annu. Conf.*, Jan. 2005, pp. 6492–6495.

[14] 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.

[15] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6230–6239.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[17] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, Munich, Springer, Oct. 2015, pp. 234–241.

[18] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Proc. 4th Int. Workshop Deep Learn. Med. Image Anal. Multimodal Learn. Clin. Decis. Support (DLMIA), 8th Int. Workshop ML-CDS*, Granada, Spain: Springer, Sep. 2018, pp. 3–11.

[19] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 565–571.

[20] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[21] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.

[22] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*.

[23] N. Ibtehaz and M. S. Rahman, "MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation," *Neural Netw.*, vol. 121, pp. 74–87, Jan. 2020.

[24] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proc. IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021.

[25] C. Kolb, "Geologic control of sand boils along Mississippi river levees," USACE Waterways Exp. Station, Vicksburg, MS, USA, Tech. Rep., Miscellaneous Paper S-75-22, 1975.

[26] C. S. P. Ojha, V. P. Singh, and D. D. Adrian, "Determination of critical head in soil piping," *J. Hydraulic Eng.*, vol. 129, no. 7, pp. 511–518, Jul. 2003.

[27] S. N. Semmens and W. Zhou, "Evaluation of environmental predictors for sand boil formation: Rhine–Meuse delta, Netherlands," *Environ. Earth Sci.*, vol. 78, no. 15, pp. 1–11, Aug. 2019.

[28] P. Maragos, A. Sofou, G. B. Stamou, V. Tzouvaras, E. Papatheodorou, and G. P. Stamou, "Image analysis of soil micromorphology: Feature extraction, segmentation, and quality inference," *EURASIP J. Adv. Signal Process.*, vol. 2004, no. 6, Dec. 2004, Art. no. 356937.

[29] J. L. Hernández-Hernández, G. García-Mateos, J. M. González-Esquiva, D. Escarabajal-Henarejos, A. Ruiz-Canales, and J. M. Molina-Martínez, "Optimal color space selection method for plant/soil segmentation in agriculture," *Comput. Electron. Agricult.*, vol. 122, pp. 124–132, Mar. 2016.

[30] R. A. D. A. Nóbrega, J. Aanstoos, B. Gokaraju, M. Mahrooghy, L. Dabirru, and C. G. O'Hara, "Mapping weaknesses in the Mississippi river levee system using multi-temporal UAVSAR data," *Revista Brasileira de Cartografia*, vol. 65, no. 4, pp. 681–694, Aug. 2013.

[31] J. Dyson, A. Mancini, E. Frontoni, and P. Zingaretti, "Deep learning for soil and crop segmentation from remotely sensed data," *Remote Sens.*, vol. 11, no. 16, p. 1859, Aug. 2019.

[32] A. G. Smith, J. Petersen, R. Selvan, and C. R. Rasmussen, "Segmentation of roots in soil with U-Net," *Plant Methods*, vol. 16, no. 1, pp. 1–15, 2020.

[33] A. Kuchi, M. T. Hoque, M. Abdelguerfi, and M. C. Flanagin, "Machine learning applications in detecting sand boils from images," *Array*, vols. 3–4, Sep. 2019, Art. no. 100012.

[34] A. Kuchi, M. Panta, M. T. Hoque, M. Abdelguerfi, and M. C. Flanagin, "A machine learning approach to detecting cracks in levees and floodwalls," *Remote Sens. Appl., Soc. Environ.*, vol. 22, Apr. 2021, Art. no. 100513.

[35] M. Panta, M. T. Hoque, M. Abdelguerfi, and M. C. Flanagin, "Pixel-level crack detection in levee systems: A comparative study," in *Proc. IEEE Int. Geosci. Remote Sens. Symp (IGARSS)*, Jul. 2022, pp. 3059–3062.

[36] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep learning techniques for medical image segmentation: Achievements and challenges," *J. Digit. Imag.*, vol. 32, no. 4, pp. 582–596, Aug. 2019.

[37] M. Panta, M. T. Hoque, M. Abdelguerfi, and M. C. Flanagin, "IterLUNet: Deep learning architecture for pixel-wise crack detection in levee systems," *IEEE Access*, vol. 11, pp. 12249–12262, 2023.

[38] R. Alshawi, M. T. Hoque, and M. C. Flanagin, "A depth-wise separable U-Net architecture with multiscale filters to detect sinkholes," *Remote Sens.*, vol. 15, no. 5, p. 1384, Feb. 2023.

[39] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, "U-Net and its variants for medical image segmentation: A review of theory and applications," *IEEE Access*, vol. 9, pp. 82031–82057, 2021.

[40] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proc. 14th Eur. Conf. Comput. Vis. (ECCV)*, Amsterdam, The Netherlands: Springer, Oct. 2016, pp. 630–645.

[41] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[42] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

[43] Y. Wu and K. He, "Group normalization," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.

[44] A. K. Dubey and V. Jain, "Comparative study of convolution neural network's Relu and leaky-Relu activation functions," in *Applications of Computing, Automation and Wireless Systems in Electrical Engineering*. Singapore: Springer, 2019, pp. 873–880.

[45] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[46] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.

[47] H. Abdi and L. J. Williams, "Principal component analysis," *WIREs Comput. Statistic*, vol. 2, no. 4, pp. 433–459, Jul./Aug. 2010

[48] M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, "Transfusion: Understanding transfer learning for medical imaging," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 3347–3357.

[49] H. Li, P. Chaudhari, H. Yang, M. Lam, A. Ravichandran, R. Bhotika, and S. Soatto, "Rethinking the hyperparameters for fine-tuning," 2020, *arXiv:2002.11770*.

[50] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[51] A. Dutta and A. Zisserman, "The VIA annotation software for images, audio and video," in *Proc. 27th ACM Int. Conf. Multimedia*, Oct. 2019, pp. 2276–2279.

[52] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, "Albumentations: Fast and flexible image augmentations," *Information*, vol. 11, no. 2, p. 125, Feb. 2020. [Online]. Available: https://www.mdpi.com/2078-2489/11/2/125

[53] R. Fan, H. Wang, M. J. Bocus, and M. Liu, "We learn better road pothole detection: From attention aggregation to adversarial domain adaptation," in *Computer Vision—ECCV 2020 Workshops*, Glasgow, U.K.: Springer, Aug. 2020, pp. 285–300.

[54] S. Jadon, "A survey of loss functions for semantic segmentation," in *Proc. IEEE Conf. Comput. Intell. Bioinf. Comput. Biol. (CIBCB)*, Oct. 2020, pp. 1–7.

[55] F. Chollet et al., "Keras," 2015. [Online]. Available: https://keras.io/getting_started/faq/#how-should-i-cite-keras and https://keras.io

[56] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.

[57] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1800–1807.

MANISHA PANTA received the bachelor's degree in computer engineering from Tribhuvan University, in 2016, the master's degree in computer science from The University of New Orleans, LA, USA, in 2019, and the Ph.D. degree in engineering and applied sciences with a concentration in computer science from The University of New Orleans, successfully defending her dissertation, in July 2023. Her passion lies in teaching software engineering, researching applied science, and actively contributing to women in tech communities, such as the ACM-Women Chapter at UNO, Nepali Women in Computing (NWiC), and Women in Data Science Nepal. Through these engagements, she advocates for technology and data science while empowering women in the field.

MD. TAMJIDUL HOQUE received the Ph.D. degree in information technology from Monash University, Melbourne, VIC, Australia, in 2008. He is currently an Associate Professor with the Computer Science Department, The University of New Orleans, New Orleans, LA, USA. From 2011 to 2012, he was a Post-doctoral Fellow with Indiana University–Purdue University Indianapolis, Indianapolis, IN, USA. From 2007 to 2011, he was a Research Fellow with Griffith University, Brisbane, QLD, Australia. His current research interests include deep/machine learning, evolutionary computation, and artificial intelligence, applying them to complex optimization problems, especially for bioinformatics problems, such as protein structure prediction, disorder predictor, and energy function.

KENDALL N. NILES received the Bachelor of Science degree (Hons.) in computer science. She is currently a Computer Scientist with the U.S. Army Corps of Engineers. In this role, she employs the deep understanding of edge computing, machine learning, the Internet of Things (IoT), and data engineering to develop innovative solutions for flood control systems. A prolific contributor to the field, she has four publications to her credit, further demonstrating her expertise and commitment to advancing the field of computer science. Her hard work and dedication have been recognized with two professional awards, highlighting the impact of her contributions. She maintains active memberships in several professional organizations, including the Institute of Electrical and Electronics Engineers, the Association for Computing Machinery, the Society of American Military Engineers, and the Society of Women Engineers. In addition to her professional responsibilities, she dedicates her time to mentorship and community service. She volunteers as a Counselor for the Society of American Military Engineers and acts as a Mentor for the Base Camp Coding Academy. These roles reflect her dedication to fostering the growth and development of the next generation of technology professionals.

JOE TOM received the B.S. degree from Mississippi State University, the M.S. degree from the University of California at Davis, Davis, and the Ph.D. degree from The University of Western Australia (UWA), in 2018. He is currently a Research Civil Engineer with the U.S. Army Engineer Research and Development Center (ERDC). Prior to joining ERDC, he was an Assistant Professor in geotechnical engineering with the University of Illinois at Urbana-Champaign, a Research Fellow with the Centre for Offshore Foundation Systems, UWA, and a Senior Geotechnical Engineer with Advanced Geomechanics (now Fugro), Perth, Australia. His research interests include sediment transport, the scour of bridge and coastal infrastructure, the geotechnical and hydraulic performance of levee and dam infrastructure, and the inspection and monitoring of civil infrastructure.

MAIK FALANAGIN received the Master of Engineering degree in electrical engineering and computer science from MIT and the Ph.D. degree in engineering and applied sciences from The University of New Orleans (UNO). He is currently a Louisiana's First Professionally-Licensed Software Engineer. He has been with the Corps of Engineers, since 2001, including his time as a Contractor. He is also the GIS Lead for the New Orleans District and the Technical Lead of their Software Development Team.

• • •

MAHDI ABDELGUERFI is currently a Professor in computer science with The University of New Orleans (UNO) and the Founder and the Executive Director of the Interdisciplinary Canizaro-Livingston Gulf States Center for Environmental Informatics (GulfSCEI—pronounced Gulf Sea), whose primary goal is to assist state, federal agencies, and non-government organizations solve environmental problems of the coastal margin of the State of Louisiana and other Gulf States. He is a recipient of the 2022 UNO's Career Research Award. He was awarded the 2003 and 2016 Alan Berman Annual Research Publication Award from the Naval Research Laboratory (NRL), Department of the Navy for research performed jointly with scientists from NRL. He received the 2016 UNO's Competitive Funding Prize in recognition of outstanding and innovative work in support of UNO's research mission. He was also a recipient of UNO's Early Career Achievement Award for Excellence in Research.