Relief Well Identification from Sattelite Imagery Using Dual Kernel Filters Unets

Linbin Zhang, Joe Tom and Zachary Nick

US Army Engineer Research and Development Center
3909 Halls Ferry Road, Vicksburg, MS 39180
linbin.zhang@erdc.dren.mil

ABSTRACT

Precise identification of levee relief well locations from satellite imagery can be challenging and time-consuming. In this research, we propose a novel approach utilizing a U-Net architecture to identify relief wells along flood control levees, as a proof of concept of combining freely available imagery with machine learning techniques. The goal of this research is to develop techniques to make infrastructure asset management more efficient by automating the process. Our study highlights the crucial role of the convolution kernel size in the U-Net architecture, which significantly influences the accuracy of the results. Larger convolution kernel sizes excel in capturing extensive contextual information from the input image, potentially leading to superior outcomes. However, training with larger kernels is computationally intensive. Conversely, smaller convolution kernel sizes excel in capturing local features. To strike a balance these considerations, we introduce a Dual Kernel Filter U-Net, which combines two U-Nets with distinct convolution kernel sizes, 3x3 and 11x11. This innovative approach aims to harness the strengths of both convolution kernel sizes to improve accuracy and overall performance. The proposed Dual kernel filter U-Net was trained and evaluated on a real dataset of relief wells from Google Earth imagery. Our evaluation results demonstrate that this model achieves an accuracy rate of 99.76% with training dataset 2626 satellite images. Notably, this significant accuracy enhancement is achieved without substantially increasing computational time, making it a promising advancement in satellite image analysis for object location identification and asset, and disaster management.

Keywords: U-Net; satellite images; dual kernel filter; machine learning.

1. Instruction

Deep learning has demonstrated remarkable success across various domains. The UNet architecture (Ronneberger et al., 2015) holds a significant position in deep learning. Originally introduced for medical image segmentation, UNet has evolved over years, finding widespread applications in diverse fields and spawning numerous variants tailored for different tasks. Notably, UNets have proven to excel in tasks demanding detailed delineation of intricate structures, including their outstanding performance in medical imaging and various segmentation applications.

The classic UNet architecture consists of two key components: the encoder and the decoder. In UNet, global features are obtained through convolution. The encoder extracted global features via multiple convolutional layers, culminating in the deepest global feature representation in the final encoder layer. The decoder, comprising multiple upsampling modules, transforms the abstract representation into a semantic segmentation mask. The decoder begins with a 2x2 transpose convolution, upsampling low-resolution feature maps to the original input image size from encoder. Subsequently, it concatenates with the corresponding skip connection feature map from the encoder block, preserving features from earlier layers often lost due to network depth. The output of the last decoder undergoes a 1x1 convolution with sigmoid activation, generating a segmentation mask for pixel-wise classification.

Since the inception of UNet in 2015, numerical variants have emerged in the literature (Zhou etal., 2018; Zhang, et al., 2023; Jin etal., 2019; Dolz et al, 2019). In the ResUnet, the encoding and decoding processes are connected through residual connections to avoid information loss during the encoding process. The integration of attenuation mechanism into UNet has gained prominence in recent years due to their notable performance across various fields. Notable UNet variants include:

a) CBAM UNet: This variant incorporates concurrent spatial and channel attention blocks into UNet.
b) ASCU-Net: A UNet variant featuring a tripartite attention mechanism.
c) Depthwise Separable UNet: This version employs a UNet architecture with multiscale filtering through depthwise separable convolutions.
d) Multiscale ResUNet: An extension of UNet that incorporates multiscale features for improved performance.

In recent years, UNets have found application in crack detection with levee systems. This underscores their versatility and adaptability to diverse domains, showcasing their efficacy beyond their original medical imaging applications.

This paper introduces a model designed to identify relief wells from satellite images, considering both local detail features and global features. While most UNets use small kernel 3x3, recent research (Ding, et al., 2022) suggests that large kernel sizes can yield superior results, especially in shape-biased scenarios. However, large kernels are computational expensive. To strike a balance between accuracy and computational costs, we propose the Dual kernel Size UNet. Built on the UNet model, it combines 3 x 3 and 11x11 kernel sizes from ResUNet. The smaller kernel captures features missed by the large one, enhancing performance. This novel UNet variant has been applied to identify relief wells in satellite images for levee system management.

This paper is organized as follows:
a). Introduction of Dual Kernel size ResUNet: In the section, we provide an introduction to the proposed Dual Kernel Size ResUNets, highlighting its unique architecture and the motivation behind incorporating dual kernel sizes. b). Testing results and experiments with satellite images: the subsequent section delves into the testing results obtained through our experiments with satellite images. This includes insights into mask generation and the image preprocessing techniques employed to enhance the performance of the model. c). Conclusion: the final section encapsulates our conclusions drawn from the study, summarizing key findings, and potentially suggesting avenues for future research and application of the proposed Dual Kernel Size ResUNet in satellite image analysis.

2. Method

The methodology employed in this study introduces the Dual Kernel Filter Unet, a novel architecture derived from the Unet model. UNet , an enahnced network based on the fully convolutional network (FCN), has demonstrated exceptional performance with minimal data in biomedical image segmentation applications. The proposed Dual Kernel Filter Unet architecture is depicted in Figure 1.

The Dual Kernel Filter Unet integrates two Residual Unets (ResUNet), each operating with distinct convolutional kernel sizes. A ResUNet, a three-level encoder and decoder structure with skip connections, processes satellite images as inputs and produces masked images as outputs. One ResUNet utilizes a large 11x11 kernel size, while the other employed a smaller 3x3 kernel size. The rationale behind incoporating different kernel sizes lies in the ability to larger kernels to capture more contextual information, thereby producing superior results. However, it is noted that larger kernels tend to be more shape-biased compared to their smaller counterparts. Each ResUNet, with its specific kernel size, extracts feature maps at different levels. Combining these two ResUNets aims to enhance accuracy and overall performance.

Within each ResUNet, the contracting path on the left side can be viewed as an encoder responsible for feature extraction, capturing semantic information from the satellite images. The symmetric expanding path on the right side serves as a decoder, generating the corresponding mask image. This dual-path architecture facilitates the extraction of intricate features and synthesis of accurate segmentation masks.

The input data size for this study is set at 256x256 pixels. To efficiently capture principal features while increasing the receptive field and reducing parameter count, downsampling is employed. The input images traverse through three downsampling layers, constracting the spatial dimensions of the data tensor from 128x128 to 16x16.

The decoding layer mirrors theencoding layer structure. However, instead of average pooling, a transposed convolution with a stride of 2 is used to gradually restrore the resolution of feature maps through the upsampling operation for mask image reconstruction. The skip connection operation concatenates the feature map from the encoder with the feature map in the corresponding decoder. This fusion of shallow and deep features aids in reconstructing the image. The decoding path automatically learns relevant features lost during downsampling, compensating for low-level semantic information in the high-resolution feature map restored by upsampling. Addtionally, concatenation facilities backward propagation of model weight gradients during training.

The outputs of the two ResUNets are concatenated and connected together. One layer is added to this concatenated structure. During the training, dropout layers with a dropout rate of 20% are inserted after each layer to mitigate the overfitting problem.

In the training process, binary crossentropy is employed as the loss function, with the Adam optimizer used to minimized this loss function. This comprehensive approach ensures effective training and optimization of the proposed Dual Kernel Filter Unet model.

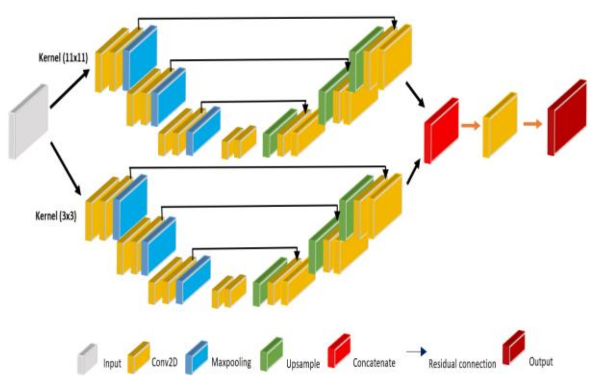


Figure 1. The network structure of the dual kernel size consists of two Unets. Each UNet has encoder, decoder and ressidual connection. One convolution kernel size is 11x11 and the other one is 3x3.

3. Machine learning application

The novel Dual Kernel Size UNet is subjected to testing using satellite images. The dataset comprises relief well images obtained from google maps, initially resized to 265x265 pixels. However, with only 470 images acquired, the dataset was deemed insufficient for robust model training. To address this limitation, augmentation techniques were applied to augment the dataset and enhance diversity.

Various augmentation methods were employed, including Gassian blur, salt image, horizontal and vertical shifts, flip and zoom. These techniques effectively generated additional images, resulting in a total of 3283 images for comprehensive machine learning testing. This augmented dataset ensures a more robust and diverse training process for the Dual Kernel Size UNet, allowing it to generalize better and capture a wider range of features during training and testing phases.

3.1. Annotation and Preprocessing

To facilitate the training process, we utilized the VGGImage Annotator tool for creating segmentation masks for all 3283 satellite images. The tool generates a JSON file containing coordinates for manually labeled relief well regions. A Python script was developed to convert these coordinates from the JSON file into corresponding masks for the input images.

In Figure 2(a), an example of a satellite input image is displayed, while Figure 2(b) showcases its corresponding mask, depicting the relief well regions. This annotated dataset provides crucial ground truth information for training the Dual Kernel Size UNet.

Before initiating the training, a critical step involved normalizing the input images. Each feature was normalized independently by subtracting the mean and dividing by the standard deviation. Our tests demonstrated that this normalization significantly contributes to improving accuracy and convergence during the training process. This step ensures that the model is trained on a standardized input, enhancing its performance and enabling it to effectively learn relevant features for relief well identification.

(a)

(b)

Figure 2. (a) Original satellite image. Two red arrows point to relief wells. (b) mask created by the VGGImage Annotator tool.

3.2. Training and Evaluation

The network was trained using the provided training data with a configuration of 200 epochs and early stopping to prevent model overfitting. The early stopping criteria were set to monitor “validation accuracy” with a patience value of 5, and the restoration of the best weights was enabled. The Adam optimizer was employed with a learning rate of 0.001, and a batch size of 10.
The binary cross-entropy loss function was utilized during model training.

The training dataset consisted of 2626 images, while 657 images were reserved for testing during the training process. The training concluded after 33 epochs. Due to the small size of relief wells in the input images, intersection over Union (IoU) was not computed.

3.3. Training Parameters

- Epochs: 200.
- Early stopping: Monitor- “validation accuracy”, Patience – 5, Restore best weights – True.
- Optimizer: Adam, learning rate: 0.001
- Batch size: 10.
- Loss function: Binary cross-entropy.

3.4. Results

- Accuracy: 99.97% on the training dataset (2626 satellite images).
- Loss: Decreased by more than 90%.

3.5. Visual Evaluation

- Figure 3: Accuracy curves.
- Figure 4: Loss function
- Figure 5: Visual comparison of randomly selected segmentation outputs.
- Figure 6(a): Satellite image with relief wells marked by red arrows
- Figure 6(b): Predicted results with red dots accurately identifying relief well locations.

The network demonstrated exceptional accuracy, successfully predicting relief well locations in the satellite images. It effectively distinguished relief wells from features resembling them, such as roofs, as evidenced in Figure 6(b). The achieved accuracy rate of 99.97% underscores the model's robust performance on the training dataset.

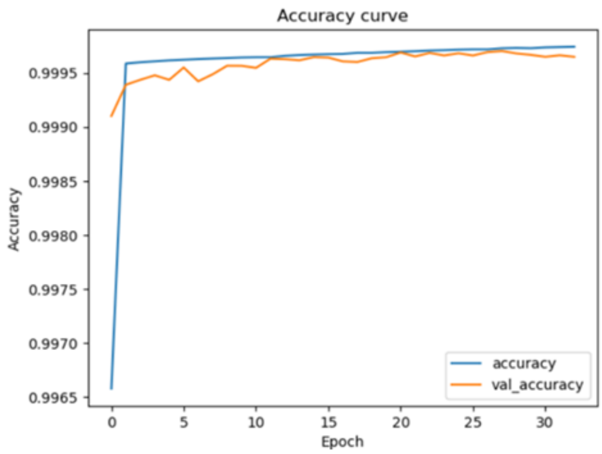


Figure 3. Accuracy curve vs epochs.

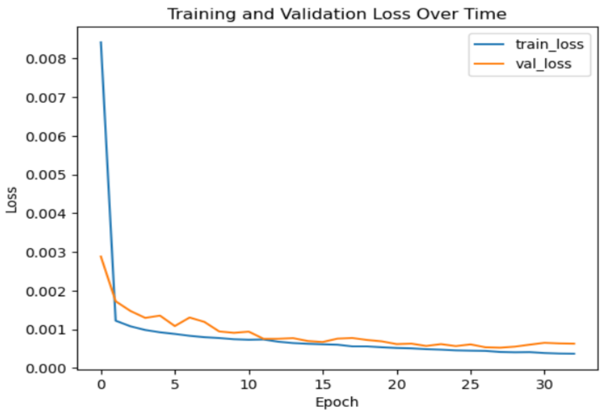


Figure 4. Loss function vs epoch.

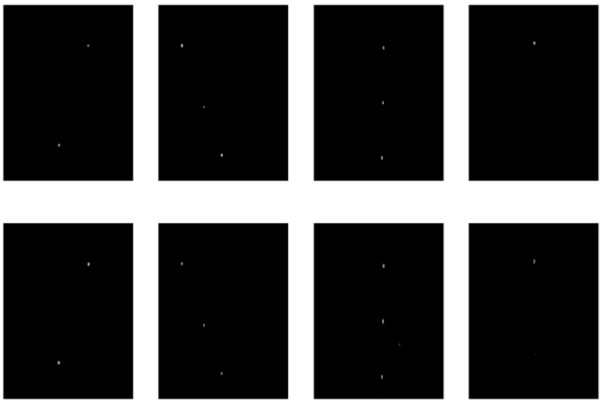
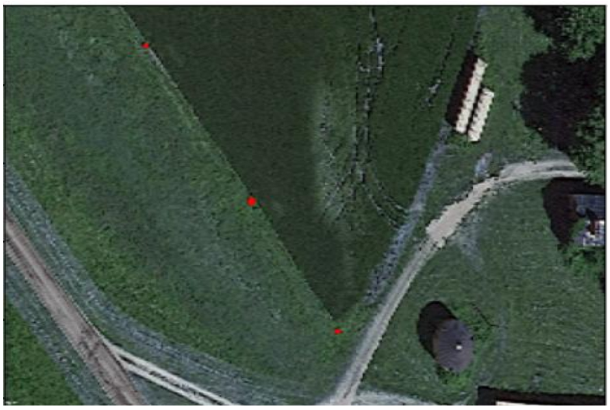


Figure 5. Top row shows the ground true masks. Bottom row shows the predicted relief well locations.

(a)



(b)

Figure 6. (a) input original satellite image. There are 3 relief wells (arrowed). (b) predicted results (read dots) overlaied on satellite images.

4. Conclusions

This study introduces a novel approach, the Dual Kernel Filter ResUNet, for automated identification of relief well locations in satellite images. Achieving an impressive accuracy rate of 99.97%, the model leverages the advantages of both small and large convolution kernels. Comparative analysis with other relevant deep learning models showcases the superior performance of the proposed model. The incorporation of a residual UNet contributes to parameters reduction, facilitating efficient training and faster convergence.

The proposed approach holds promise for various imaging applications, extending beyond relief well identification to areas such as crack detection and sand boil detection in levee systems. The success of this model emphasizes its potential contribution to enhancing the efficiency and accuracy of imaging-based applications in critical infrastructure management.

References

X. Ding, X. Zhang, Y. Zhou, J. Han, G. Ding, and J. Sun, "Scaling up your kernels to 31x31: Revisiting large kernel design in CNNs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 11953–11965.
https://doi.org/10.48550/arXiv.2203.06717

Jose Dolz, Ismail Ben Ayed, and Christian Desrosiers. Dense multi-path u-net for ischemic stroke lesion segmentation in multiple image modalities. In Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part I 4, pages 271–282. Springer, 2019.
https://doi.org/10.48550/arXiv.1810.07003

Dutta, A., Gupta, A., and A. Zissermann, 2016, *VGG Image Annotator (VIA)*. [Online]. Available: https://www.robots.ox.ac.uk/~vgg/software/via/

Jin, Q., Meng, Z., Pham, T., Chen Q., Wei, L., and Su, R., 2019, Dunet: A deformable network 5 for retinal vessel segmentation. Knowledge-Based Systems, 178:149–162.
https://doi.org/10.48550/arXiv.1811.01206

LeCun, Y., Bengio, Y., and Hinton, G., 2015, Deep learning. nature, 521(7553):436–444, 2015.
https://doi.org/10.1038/nature14539

Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. , 2018, Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999.
https://doi.org/10.48550/arXiv.1804.03999

Ronneberger, O.; Fischer, P.; Brox, T., 2015, U-Net: Convolutional Networks for Biomedical Image Segmentation; Springer International Publishing: Berlin, Germany.
https://doi.org/10.48550/arXiv.1505.04597

Jiawei Zhang, Yanchun Zhang, Yuzhen Jin, Jilan Xu, and Xiaowei Xu. Mdu-net: Multi-scale densely connected u-net for biomedical image segmentation. Health Information Science and Systems, 11(1):13, 2023.
https://doi.org/10.48550/arXiv.1812.00352

Xiaohu Zhang and Haifeng Huang. 2022, P-unet: Parallel attention based unet for crack detection. In 2022 7th International Conference on Signal and Image Processing (ICSIP), pages 311–315. IEEE.
https://doi.org/10.1109/ICSIP55141.2022.9886386

Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4, pages 3–11. Springer, 2018.
https://doi.org/10.48550/arXiv.1807.10165