RESEARCH ARTICLE

COMPUTER-AIDED CIVIL AND INFRASTRUCTURE **ENGINEERING** WILEY

Uncertainty-guided U-Net for soil boundary segmentation using Monte Carlo dropout

X. Zhou[1,2] | B. Sheil[2] | S. Suryasentana[3] | P. Shi[1]

[1]First School of Rail Transportation, Soochow University, Suzhou, P.R. China

[2]Second Laing O'Rourke Centre and Construction Engineering and Technology, Department of Engineering, University of Cambridge, Cambridge, UK

[3]Department of Civil and Environmental Engineering, University of Strathclyde, Glasgow, UK

Correspondence
Peixin Shi, No. 8 Jixue Road, Soochow University, Suzhou 215131, China.
Email: pxshi@suda.edu.cn

Funding information
National Natural Science Foundation of China, Grant/Award Number: 52278405

Abstract

Accurate soil stratification is essential for geotechnical engineering design. Owing to its effectiveness and efficiency, the cone penetration test (CPT) has been widely applied for subsurface stratigraphy, which relies heavily on empiricism for correlations to soil type. Recently, deep learning techniques have shown great promise in learning the relationship between CPT data and soil boundaries automatically. However, the segmentation of soil boundaries is fraught with model and measurement uncertainty. This paper introduces an uncertainty-guided U((-Net (UGU-Net) for improved soil boundary segmentation. The UGU-Net consists of three parts: (a) a Bayesian U-Net to predict a pixel-level uncertainty map, (b) reinforcement of original labels on the basis of the predicted uncertainty map, and (c) a traditional deterministic U-Net, which is applied to the reinforced labels for final soil boundary segmentation. The results show that the proposed UGU-Net outperforms the existing methods in terms of both high accuracy and low uncertainty. A sensitivity study is also conducted to explore the influence of key model parameters on model performance. The proposed method is validated by comparing the predicted subsurface profile with benchmark profiles. The code for this project is available at github.com/Xiaoqi-Zhou-suda/UGU-Net.

1 | INTRODUCTION

1.1 | Algorithm background

Deep learning (DL) methods have achieved state-of-the-art (SOTA) performance in segmentation tasks across a wide variety of practical domains, including pavement distress (Tong et al., 2023), defects (Midwinter et al., 2023), and cracks (Zhou et al., 2023) detection and far-field monitoring (Chern et al., 2023). However, the DL models have several well-documented weaknesses such as: (i) lack of interpretability, expressiveness, and transparency (Ahmed & Lin, 2023); (ii) confusion between in- and out-of-domain samples and sensitivity to domain shifts (Kendall, 2019); (iii) unreliable uncertainty estimates and frequent overconfident predictions (Hariri et al., 2019).

Uncertainty can be categorized as either epistemic or aleatoric (Hüllermeie & Waegeman, 2021; Gawlikowski et al., 2022). Epistemic uncertainty, also referred to as model uncertainty, accounts for uncertainty in model parameters and the model appropriateness itself. This uncertainty type is often caused by shortcomings in model structures, which can then amplify predictive uncertainty by marginalizing over the posterior distribution of model weights. However, it can be potentially reduced through the observation

of more data (Angelelli et al., 2024). Aleatoric (data) uncertainty captures noise inherent in the observations, which are not readily reduced or eliminated from deep neural networks (DNNs; Zhang et al., 2024). Aleatoric uncertainty is usually caused by information loss when digitalizing real-world phenomena using numerical data, such as interpreting information from images with different pixel resolutions or by manual data annotation errors due to cognition insufficiency (Blundell et al., 2015; Lakshminarayanan et al., 2016). Uncertainty estimation is important for safe decision-making in DL application of high-risk fields since it helps avoid potential problems such as bias and overfitting.

An increasing number of DL applications have sought to identify the level of uncertainty associated with output predictions for more informed decision-making. For example, Diao et al. (2022) proposed uncertainty cross-entropy loss to guide the network to directly output the prediction uncertainty instead of executing several times in the prediction phase. Based on uncertainty prediction results, Chen et al. (2023) established a source-free domain adaptation framework and a neighborhood-guided evidence-based contrastive learning (CL) approaches adapted to multiple tasks. Yao et al. (2024) proposed an uncertainty-guided CL framework that utilizes a negative CL to explore pseudo-labels of prediction with low liability and a positive CL to enhance feature learning.

In their survey of methods to quantify uncertainty in DNNs, Monte Carlo (MC) dropout is identified as one of the most popular methods. Dropout is a classical regularization technique used in many SOTA DNNs and is typically applied in fully connected layers (Hinton et al., 2012). Dropout was first introduced to prevent neurons from overfitting on training data by randomly omitting certain neural connections during each iteration (Srivastava et al., 2014). MC dropout approximates the DNN training process as performing variational inference and re-formulates the dropout layers using binary variables, which follow the Bernoulli distribution (Gal & Ghahramani, 2015). Gal (2016) also noted that dropout approximates Bayesian inference, ultimately allowing the estimation of model uncertainty. Leibig et al. (2017) measured uncertainty using MC dropout and demonstrated that incorporating uncertainty information in decision-making could enhance model performance.

Two relevant works using MC dropout for DL-based segmentation are conducted by Tang et al. (2022) and Zhang et al. (2022). Tang et al. (2022) took the uncertainty map generated by a coarse segmentation module as input to obtain the discriminative features. Despite its impressive results in highly confident segmentation,

it failed to explore influencing factors of the uncertainty and lacked sufficient analysis of uncertainty mitigation. Zhang et al. (2022) quantified structural and parametric uncertainties using dropout-based Bayesian neural network (BNN) through a 2D distribution of pixel-wise feature importance. Despite its novelty in exploring the correlation between input features and uncertainty, the method is only tested on two datasets and is not fully generalized to more complex scenarios.

While uncertainty quantification has been the subject of a large body of literature, DNNs have so far failed to fully leverage internal model uncertainty to improve model predictions for more informed data-driven decision-making. Furthermore, there is a notable lack of real-world validation of DL uncertainty estimation. This paper addresses this gap by fully leveraging Bayesian inference for uncertainty estimation, which is then embedded in the DL process for more reliable decision-making in a real-world application.

1.2 | Application background

An accurate model representation of the ground is essential for a vast majority of civil engineering problems (FHWA, 2002). Arguably, the most important element of ground models is the soil stratification, which is typically highly uncertain due to the difficulty in identifying different soil layers and lithological units (Li et al., 2016). This uncertainty arises from the limited and sparse subsurface soil samples that are available, as well as the heterogeneity and anisotropy of natural soil materials.

To estimate subsurface profiles, the cone penetration test (CPT) is typically used to predict soil behavior type and, in turn, identify boundaries (i.e., the stratigraphy). The CPT device measures the sleeve friction f_s and the tip cone resistance q_c, acting on a conical-ended penetrometer that is penetrated into the soil at a constant displacement rate. These measurements are typically recorded in penetration intervals of 20–50 mm (Mayne et al., 2002). Many scholars have conducted machine learning (ML) based research on CPT data. For example, Erharter et al. (2021) attempted to re-define decision boundaries of CPT-based data interpretation charts using a randomly generated artificial-neural network. Padmapriya and Sasilatha (2023) developed a multi-stacking ensemble ML model with feature selection algorithm to determine soil types. However, identification of soil layers from CPT data is uncertain due to: (i) measurement errors arising from the CPT equipment and/or the expertise of operators/engineers; (ii) suitability of the CPT in different soils; and (iii) classification errors arising from empirical relationships to soil type. Therefore,
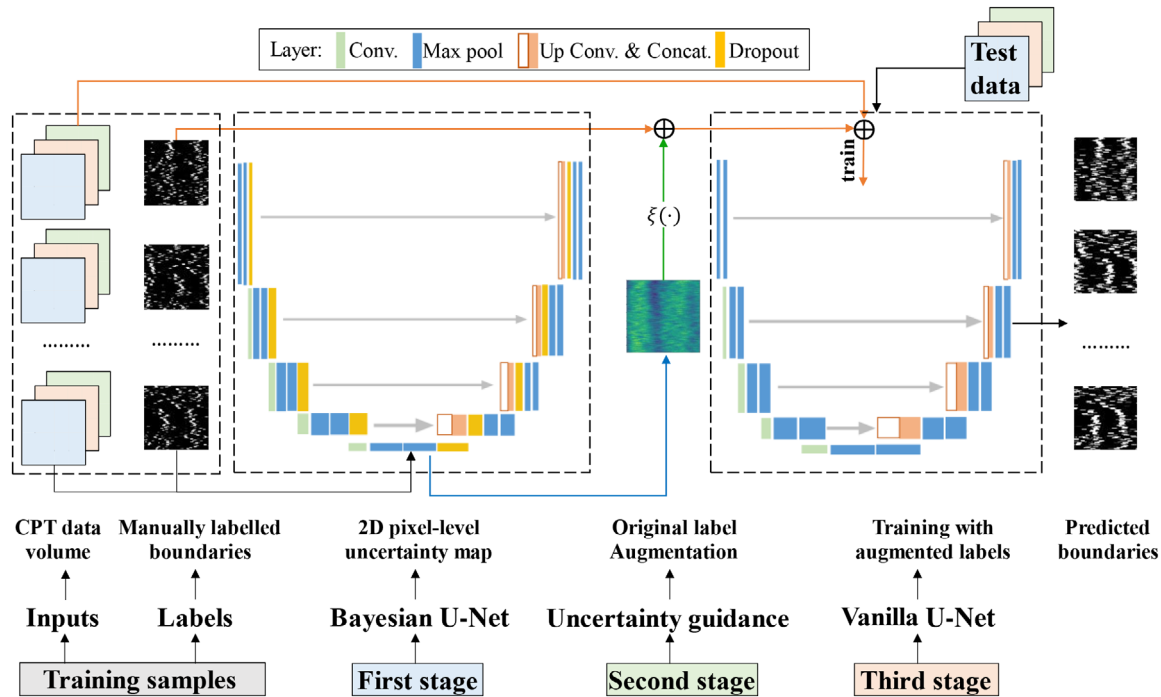
FIGURE 1　Framework of the proposed uncertainty-guided U-Net (UGU-Net) for uncertainty estimation and soil boundary segmentation.

uncertainty inherently exists when applying DL techniques to CPT data for soil layer interpretation.

Limited previous studies have explored the uncertainties involved in subsurface stratigraphy development. To estimate geological uncertainty, Provenzano (2003) proposed a fuzzy-neural network method to predict the behavior of structures built on complex cohesionless soils and quantified related uncertainties. Qi et al. (2016) proposed a practical method to estimate the horizontal transition probability matrix as the key output of a coupled Markov chain. Shi and Wang (2022) proposed a method to infer the most probable geological domain and to quantify the associated stratigraphic uncertainty.

The aforementioned methods rely heavily on statistical inference or empirical correlations to estimate the stratigraphic uncertainties. In this way, uncertainty is simply an evaluation indicator and has no contribution to model performance. Such uncertainty is commonly denoted as a one-dimensional representation, which fails to provide detailed distributions for specific areas of interest. There has been a distinct lack of published literature that has explored uncertainties when applying DL methods to soil boundary segmentation based on CPT data.

Motivated by these existing drawbacks, this paper develops a "uncertainty-guided U-Net" (UGU-Net) to fully leverage the uncertainty estimation for optimal soil boundary segmentation. A novel DL framework is proposed comprising: (a) a Bayesian U-Net to predict the 2D pixel-

level uncertainty map, (b) reinforcement of original labels on the basis of the predicted uncertainty map, and (c) a traditional deterministic U-Net, which is applied to the reinforced labels for final soil boundary segmentation. The main contributions of this paper are:

1. Task parallel network: A Bayesian U-Net is utilized for uncertainty quantification, joint with a standard U-Net for segmentation. Model configuration is optimized through stochastic sampling of hyper-parameters, as evaluated using key metrics.
2. Data augmentation: A novel domain adaptive data augmentation strategy is adopted to build training samples through cropping, shuffling, and concatenation, toward alleviating the insufficient model training caused by sparsity of datasets.
3. Uncertainty guidance: The uncertainty-guided model training is enabled through reinforcing soil boundaries at highly uncertain pixels. Sensitivity analyses are conducted to explore the influences of the uncertainty threshold and the soil boundary thickness on the overall model performance.

2 │ METHODS

Figure 1 shows the framework of the proposed UGU-Net, which comprises three main stages. First, the positive
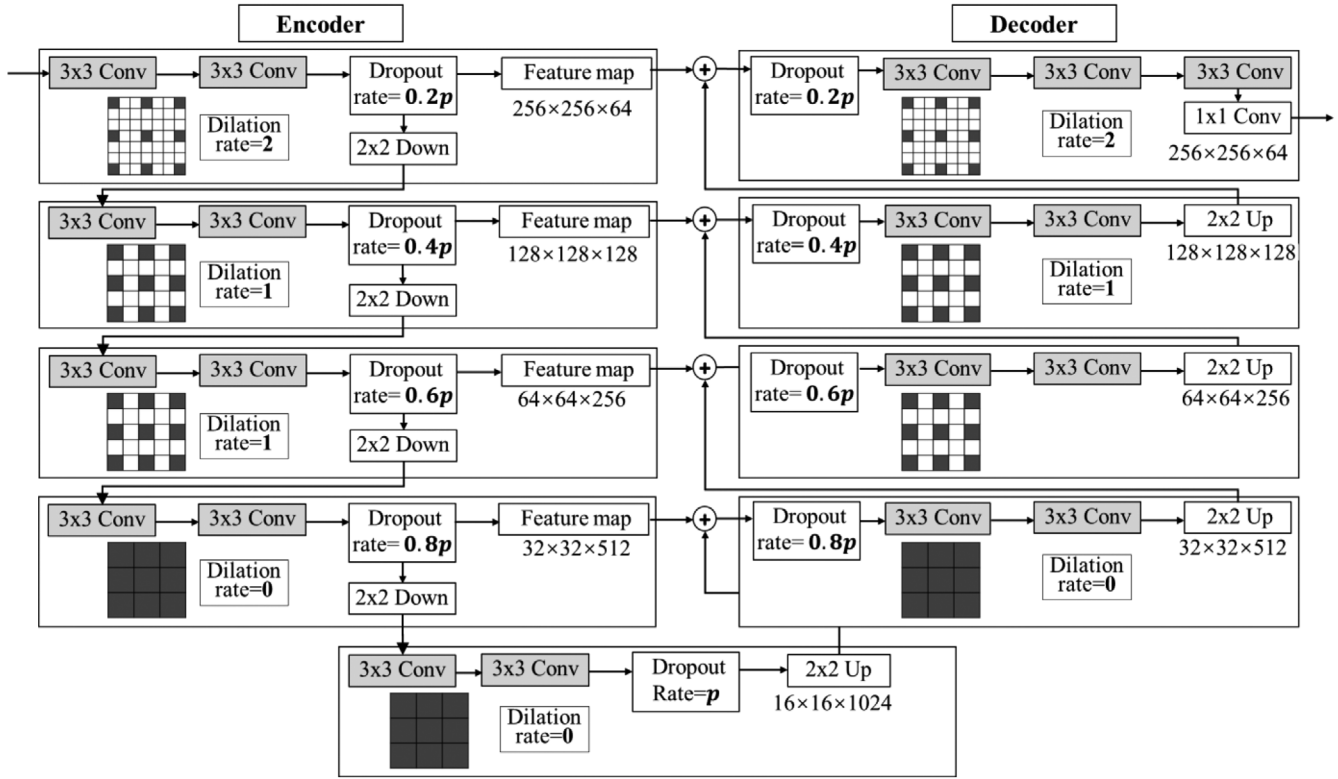
FIGURE 2 Proposed Bayesian U-Net structure with dilated convolution and dynamic dropout rate.

predictive value (PPV) is calculated using a Bayesian U-Net to generate a 2D pixel-wise uncertainty map. Note that highly uncertain areas in the predicted outputs also present large errors. The spatial distribution of uncertainty can be considered a latent variable to represent unobservable error. In this way, the 2D PPV map is then queried using a customized function that extracts the pixels where the PPV is larger than a pre-defined threshold T_u. The training labels are reinforced by thickening the soil boundaries within the identified pixels, which are then supplied to a traditional deterministic U-Net for final soil boundary segmentation.

2.1 | Bayesian U-Net structure

The proposed Bayesian U-Net architecture is designed to recover model uncertainty for the present problem as illustrated in Figure 2. A dropout layer is incorporated into each level of the network to perform Bayesian inference via MC dropout, which is computationally efficient and reduces the risk of overfitting. Inputs of size (224, 224, 3; width [in pixels], height [in pixels], number of channels) are passed into the encoder and the corresponding outputs from the decoder have a size of (224, 224, 1).

The encoder structure on the left-hand side of the U-Net comprises four levels of convolutions. Each level con-

tains two 3 × 3 dilated convolutions ("Conv.") with same padding where the dilation rate is set large at high-level convolution layers to enlarge the reception field for better feature extraction and is subsequently narrowed in later stages. The dilated convolutions are followed by a rectified linear unit activation function and a dropout layer. After that, a max pooling ("Max pool") operation with stride 2 is used to down-sample the feature map. Connection to the convolution layer in the next level doubles the number of feature maps. The decoder has a symmetrical structure but performs the up-sampling required to expand the feature map back to the desired output size. Feature maps from the same level on the encoder are also concatenated as indicated by the gray arrows in Figure 1. The bottleneck bridges the encoder and decoder with a double convolution layer followed by a dropout layer and an up-sampling convolutional layer ("Up Conv").

The dropout rate p ($\in [0, 1]$) defines the probability of dropping out a given neuron at each iteration. A dynamic p is adopted here which increases from $0.2p$ at the initial level to p at the lowest level as shown in Figure 2. This approach seeks to preserve as much information as possible. According to Baldi and Sadowski (2013), regularization can be maximized when $p = 0.5$; this value is not exceeded in this work to avoid excessive information loss.

2.2 | Uncertainty estimates

2.2.1 | Probabilistic modeling

Given a set of training inputs $\mathbf{X}$ (in our case, CPT data) and corresponding outputs $\mathbf{Y}$ (segmentation class label), the objective is to estimate the function $\mathbf{Y} = f(\mathbf{X})$. The posterior distribution over the space of training data is $p(f|\mathbf{X},\mathbf{Y})$, capturing the most probable functions given the observed data. The output y^* for a new input point x^* can be predicted through integrating over all probable functions f and calculating $p(y^*|x^*,\mathbf{X},\mathbf{Y})$ (Gal, 2016; Abdar et al., 2021). To approximate this value, the model is conditioned on a set of random variables $\boldsymbol{\omega}$ where it is assumed that the model depends on these variables alone. The posterior distribution $p(\boldsymbol{\omega}|\mathbf{X},\mathbf{Y})$ is intractable, so a variational distribution $q(\boldsymbol{\omega})$ is used as an approximation. The objective is to obtain $q(\boldsymbol{\omega})$ that minimizes the divergence from $p(\boldsymbol{\omega}|\mathbf{X},\mathbf{Y})$, resulting in the following approximate predictive distribution:

$$q(y^*|x^*) = \int p(y^*|f^*)\, p(f^*|x^*,\omega)q(\omega)\, df^* \quad (1)$$

2.2.2 | BNN

To capture epistemic uncertainty in a neural network, the weight matrix $\mathbf{W}_i$ for each hidden layer $i\,(i = 1,..L)$ is assumed to follow a normal distribution where $\mathbf{W}_i \sim \mathcal{N}(0,\mathbf{I})$ and $\mathbf{I}$ are the identity matrix. A joint Gaussian distribution over all function values $\mathbf{F} = [f_1, f_2,..f_N]$ is assumed to generate observations from a normal distribution centered on $\mathbf{F}$. This exercise produces a BNN by replacing weight parameters in a deterministic network with distributions over these parameters (Kendall and Gal, 2017). A generative model can be represented as

$$\mathbf{F}|\mathbf{X} \sim \mathcal{N}(0,\mathbf{K}(\mathbf{X},\mathbf{X}))$$
$$y_n|\,f_n \sim \mathcal{N}\left(y_n, f_n, \tau^{-1}\right) \quad (2)$$

where $n = 1,...N$; $\mathbf{K}(\cdot,\cdot)$ is the covariance function and τ is the precision coefficient to measure the observation noise (aleatoric uncertainty).

2.2.3 | Dropout as approximate variational inference

To achieve approximate variational inference in BNN, Bernoulli distributions are adopted to implement dropout during the training process (Bernardo & Smith, 2009; Osawa et al., 2019). Dropout is applied by sampling a binary vector $\boldsymbol{z_i} = \{z_{i,q}|q = 1....K_{i-1}\}$ of dimension K_{i-1} for each

hidden layer $i\,(i = 1,..L)$, where $\boldsymbol{z_{i,q}} \sim Bernoulli(p_i), q = 1..K_{i-1}\,(i = 1,...L)$, parameterized by $p_i \in [0,1]$. Suppose that $\hat{y}$ is the output of the neural network, $E(\cdot)$ is the loss function, and $\mathbf{W}_i$ and $\mathbf{b}_i$ denote the weight and bias matrices for each hidden layer i, respectively. The minimization objective can be defined as

$$\mathcal{L}_{dropout} = -\frac{1}{N}\sum_{i=1}^{N}E(y_i,\hat{y}_i) + \frac{1-p}{\tau N}\sum_{i=1}^{L}\left(||\mathbf{W}_i||_2^2\,|\,||\mathbf{b}_i||_2^2\right) \quad (3)$$

where the second term on the right-hand side introduces L2 regularization with weight decay during optimization. Given a new data input x^*, the probability of possible output y^* can be calculated using the predictive probability $p(y^*|x^*,\mathbf{X},\mathbf{Y})$. The negative predictive log-likelihood (NPL) can be approximated by MC integration with T (number of performing MC dropouts):

$$-\log p(y^*|x^*,\mathbf{X},\mathbf{Y}) \approx -\log\left(\frac{1}{T}\sum_{t=1}^{T}p(y^*|x^*,\hat{\omega}_t)\right)$$
$$\approx -\text{logsumexp}\left(-\frac{1}{2}\tau||y-\hat{y}_t||^2\right)$$
$$+\log T + \frac{1}{2}\log 2\pi\frac{1}{2}\log\tau^{-1} \quad (4)$$

with $\hat{\omega}_t \sim q(\omega)$ representing the tth sampling from distribution $q(\omega)$. A common value of $T = 1000$ is selected to approximate the uncertainty.

Combining both epistemic and aleatoric uncertainty, the PPV in this network can be approximated as

$$\text{PPV}(y^*|x^*,\mathbf{X},\mathbf{Y}) \approx \frac{1}{T}\sum_{t=1}^{T}\hat{y}_t^2 - \left(\frac{1}{T}\sum_{t=1}^{T}\hat{y}_t\right)^2 + \tau^{-1} \quad (5)$$

The mean predictive variance (MPV) is also proposed to represent image-based uncertainty on a global scale. Let s^* be the total number of pixels in the output such that the MPV can be defined as

$$\text{MPV} = \frac{1}{|s^*|}\sum_{y^*\in s^*}\text{PPV}(y^*|x^*) \quad (6)$$

2.3 | Uncertainty-guided reinforcement

The PPV is calculated to visualize the 2D representation of the uncertainty atlas using the trained Bayesian U-Net. The pixels are extracted where the PPV exceeds the threshold T_u, written as r^*. Let $r_1^* = \{x = 1|x \in r^*\}$ denote those pixels representing soil boundaries in set r^*. The parameter R_u is introduced to represent the degree to which

soil boundaries account for uncertainty as defined in Equation (11). For the extracted pixels, the original labels are reinforced by thickening the soil boundaries to denote uncertainty around their exact location. The detailed workflow of UGU-Net can be described by pseudo code.

$$\xi(\text{PPV}) = r^* = \{x \,|\, PPV(x) \rangle T_u, x \in s^*\} \tag{7}$$

$$R_u = \frac{|r_1^*|}{|r^*|} \tag{8}$$

Pseudo Code: UGU-Net for reliable soil segmentation based on MC dropout.

Input: An input image x_i $(i = 1, ..N)$ with size (224, 224, 3), output y_i with size (224,224,1), dropout rate p, precision coefficient τ; threshold value T_u

Initialization: "He" initialization of parameters

Process:

1: **for** 1 to T:
2:      Apply dropout to the Bayesian U-Net with dropout rate p and precision coefficient τ and use Adam optimizer in conjunction with the loss function in Equation (6)
3:          **Output**: $\hat{y}_i$
4:      **end for**
5: Calculate the PPV
6: **for** pixels PPV(m, n) in PPV:
7:      **if** $PPV(m,n) > T_u$:
8:          **add** (m,n) in r^*
9:      **end for**
10: **for** pixels in r^*:
11:     **if** $y_i(m,n) = 1$:
12:         **add** (m,n) in r_1^*
13:     **end for**
14: **for** pixels $y_i(m,n)$ in y_i:
15:     **if** $y_i(m,n) = 1$ and $y_i(m,n) \in r_1^*$:
16:         **update** $y_i(m-1 : m+1, n) = 1$
17:     **end for**
18: Denote updated y_i as y_i'
19: Feed x_i and y_i' into traditional deterministic U-Net for model training
20: Apply the trained U-Net to test x^*
21: **Output**: y^*

---

3 | CASE STUDY

3.1 | Data samples with adaptive data augmentation

The training data in this paper are gathered from 515 CPTs undertaken as part of the Suzhou No. 6 metro line con-

struction project. The CPT data can be mathematically represented as a matrix $D^c (c = 1, ..n_{cpt})$ comprising measurements q_c and f_s and their corresponding depths where n_{cpt} is the total number of CPT points. Each data record (q_c, f_s, Depth) in D^c could be processed using empirical correlations to acquire the soil type. Figure 3 presents an exemplar CPT curve and the interpretated soil type corresponding to each data record. Based on soil classification results, soil boundaries are determined via grouping the soil of the same type and similar depth into the same stratum. The CPT data are denoted by matrix D^c, while the interpreted soil types are represented by a binary matrix O^c, with 0 and 1 representing background and soil boundaries, respectively. The matrix D^c has a size of $N^c \times 3$ where N^c represents the number of data records for the cth CPT point and the matrix O^c has a size of $1 \times N^c$. Each D^c is paired with a O^c.

A binary image is adopted to further clarify the annotation of soil boundaries, where the white pixel corresponds to 1 in O^c and represents soil boundaries, while the black pixel corresponds to 0 in O^c and represents the background. In matrix O^c, consecutive pixels (resolution of 0.1 m) are annotated as 1 to indicate a potential range for the existence of soil boundaries. The term "soil thickness" does not represent the actual physical thickness of soil boundaries but instead indicates the range of potential boundaries in the data. For example, a thickness of 1 m means that 10 consecutive pixels are likely soil boundaries. This parameter provides more flexibility for the inherent variations in real-world soil boundary detection. In practice, even experienced engineers may not pinpoint the exact boundary locations with high resolution, so this range represents the most probable boundary region.

An important assumption needs to be made for this domain application. Note that the matrix O^c is acquired through manually annotating soil boundaries by experienced engineers. With reference to geological exploration reports, borehole logs adjacent to CPTs, and laboratory testing data, such annotation is usually considered as accurate for geotechnical engineering design and thus assumed as the ground truth (GT) labels in this application.

D^c is not suitable for training the neural network due to intractable gradient descent since the input size of $3 \times N^c$ is not computation-friendly owing to asymmetrical feature distribution along two dimensions. It may cause gradient vanishment or explosion and lead to sub-optimal results. To solve this issue, a domain adaptive data augmentation method is adopted for a more explainable representation of training samples. Data augmentation is essential to prevent overfitting (Krizhevsky et al., 2012), which improves the associated network training through a carefully designed augmentation scheme (Paschali et al., 2019). For CPT data, the model should be robust to distribution shift of the CPT data among different CPT locations. These
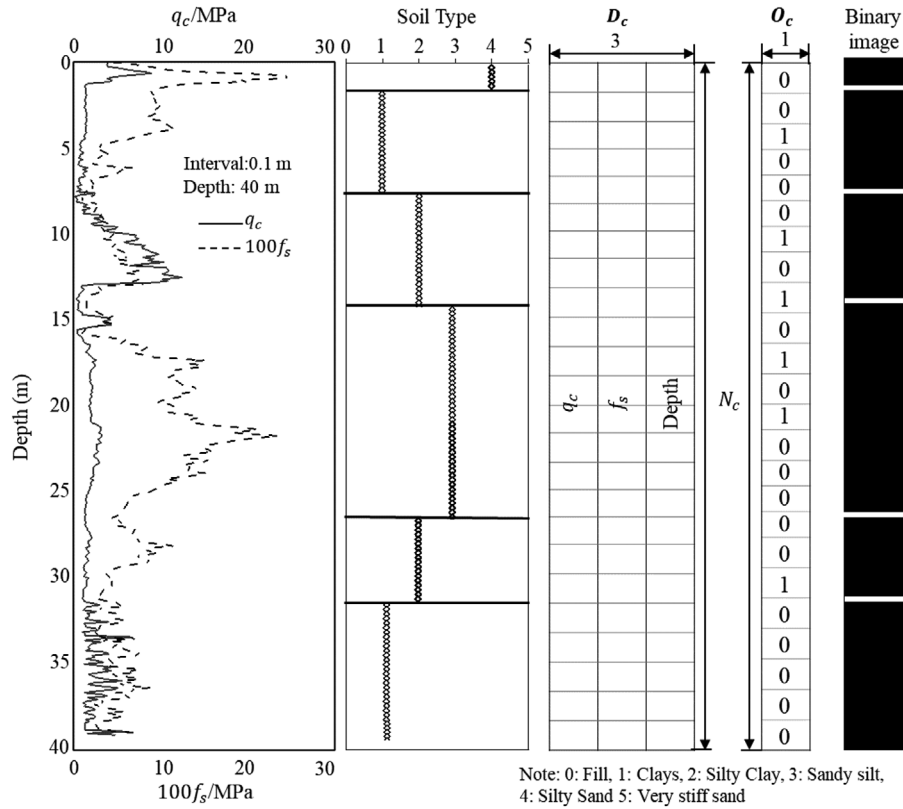
FIGURE 3 An exemplar cone penetration test (CPT) curve and interpreted soil type with determined soil boundaries represented through white lines in binary image.
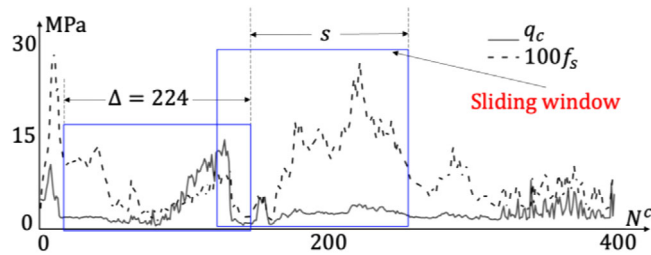

FIGURE 4 A CPT curve extracted from the gathered database showing the adopted sliding window and stride.

issues motivate a novel adaptive augmentation method following two main steps:

Step 1: For each CPT dataset, a sliding window of width Δ is adopted with a pre-defined stride (s) as shown in Figure 4. The length of the window is set as 224 to crop 224 consecutive pieces of records from the original curve. In this way, n_c number of input sub-matrices $\boldsymbol{d_c}$ with size of 3×224 can be acquired for one CPT curve (considering all three measurements/channels). Such input size aligns with the common image dimensions used in image classification tasks. It is a compromise between the level

of details captured in the images and the memory requirements during model training and inference. Larger input sizes may offer finer details, but they increase computational costs, while smaller sizes may lose important information, impacting model performance.

$$n_i = \left\lceil \frac{N^i - 224}{s} \right\rceil \tag{9}$$

Note that if $N^c - 224$ cannot be exactly divided by s, n_i should be rounded up to ensure the last sub-matrix covering the end of the curve. In total, $\Sigma_c = \sum_{i=1}^{515} n_c$ sub-matrixes are created for the present database comprising 515 CPTs. The same operation is performed on $\boldsymbol{O^c}$ to acquire the paired annotated output $\boldsymbol{o^c}$ for each $\boldsymbol{d_c}$. In total, Σ_c number of pairs constitute a large database $\mathbb{D}$.

Step 2: To construct training samples with size of $224 \times 224 \times 3$, 224 input–output pairs (of CPT data and corresponding soil boundary labels) are randomly selected from $\mathbb{D}$ and then concatenated as shown in Figure 5. In total, $\mathbf{C}_{\boldsymbol{n}}^{224} A_{224}^{224}$ (see Notation for definition) training samples are created.
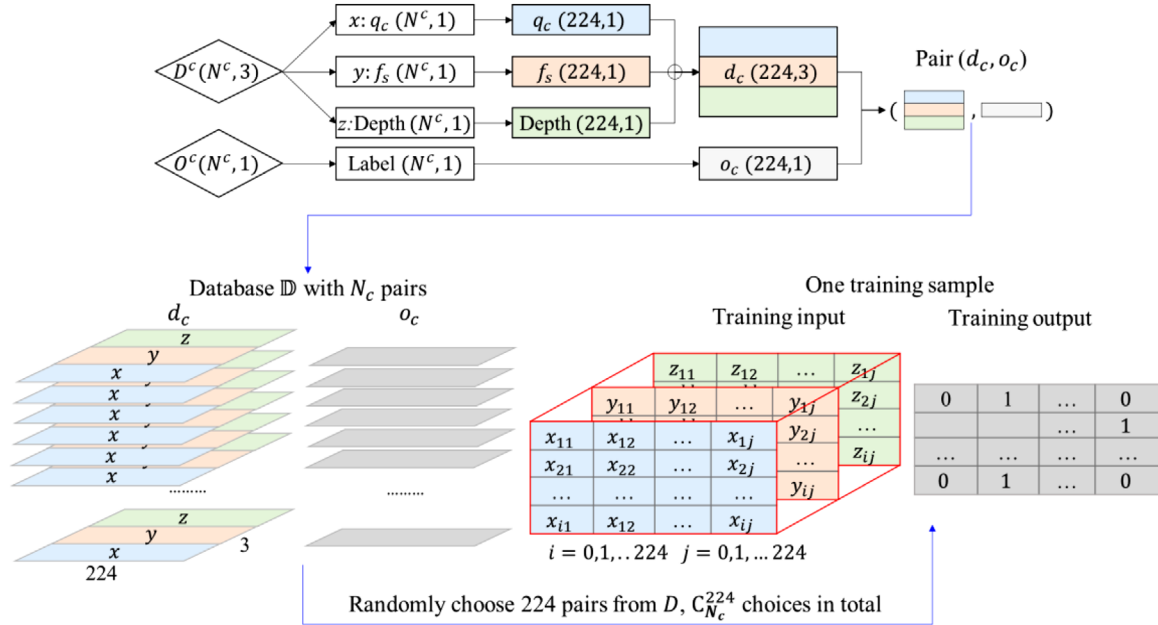
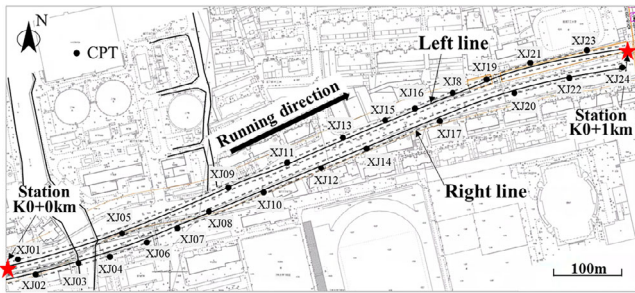**FIGURE 5** Reinforced training dataset and an exemplar training sample.

FIGURE 6 Typical CPT layout for the Suzhou No.6 metro line project; section from Station K0 + 0 km to Station K0 + 1km shown.

This stochastic selection and concatenation process can be considered as data augmentation, which can potentially increase model robustness by introducing more data distributional uncertainties. In this paper, 1000 samples are selected for network training, which are divided into training and validation datasets with a split ratio of 80% to 20%.

3.2 | Network uncertainty identification

Figure 6 shows the typical layout of CPTs within the section from K0 + 0 m to K0 + 1000 m, which are staggered around the metro line. The spacing between two adjacent CPTs ranges between 30 and 50 m, which is too large to construct a high-resolution profile if using CPT data alone. The depth for each curve ranges from 40 to 70 m. In that

case, the uncertainty should be identified before applying the DL to soil boundary segmentation. Four sources of uncertainty is identified here as

- Parameter uncertainty: Parametric uncertainty arises due to a lack of knowledge or insufficient understanding of the system. The subsurface is naturally uncertain and the soil classification relies heavily on empiricism. For these reasons, inaccuracy of personal judgements can lead to epistemic uncertainty as shown in Figure 7a.
- Distribution shift: The quality of CPT data depends on the skill of the operator and the accuracy of the instrument, which introduces aleatoric uncertainty. Moreover, distribution shift may occur between training, validation, and testing datasets since each CPT curve is distinct and independent from the others as illustrated in Figure 7b.
- Model inadequacy: The model requires pre-defined hyperparameters that may not lead to the optimal solution. The model may also suffer from inadequacies such as data leaks, overfitting, or insufficient data. These issues introduce epistemic uncertainties, as depicted in Figure 7c, and necessitate hyperparameter tuning for better model performance.
- Code uncertainty: Code uncertainty is recognized as epistemic uncertainty in data structures and design patterns, which occurs because of numerical approximations and errors during the training process, as shown in Figure 7d.
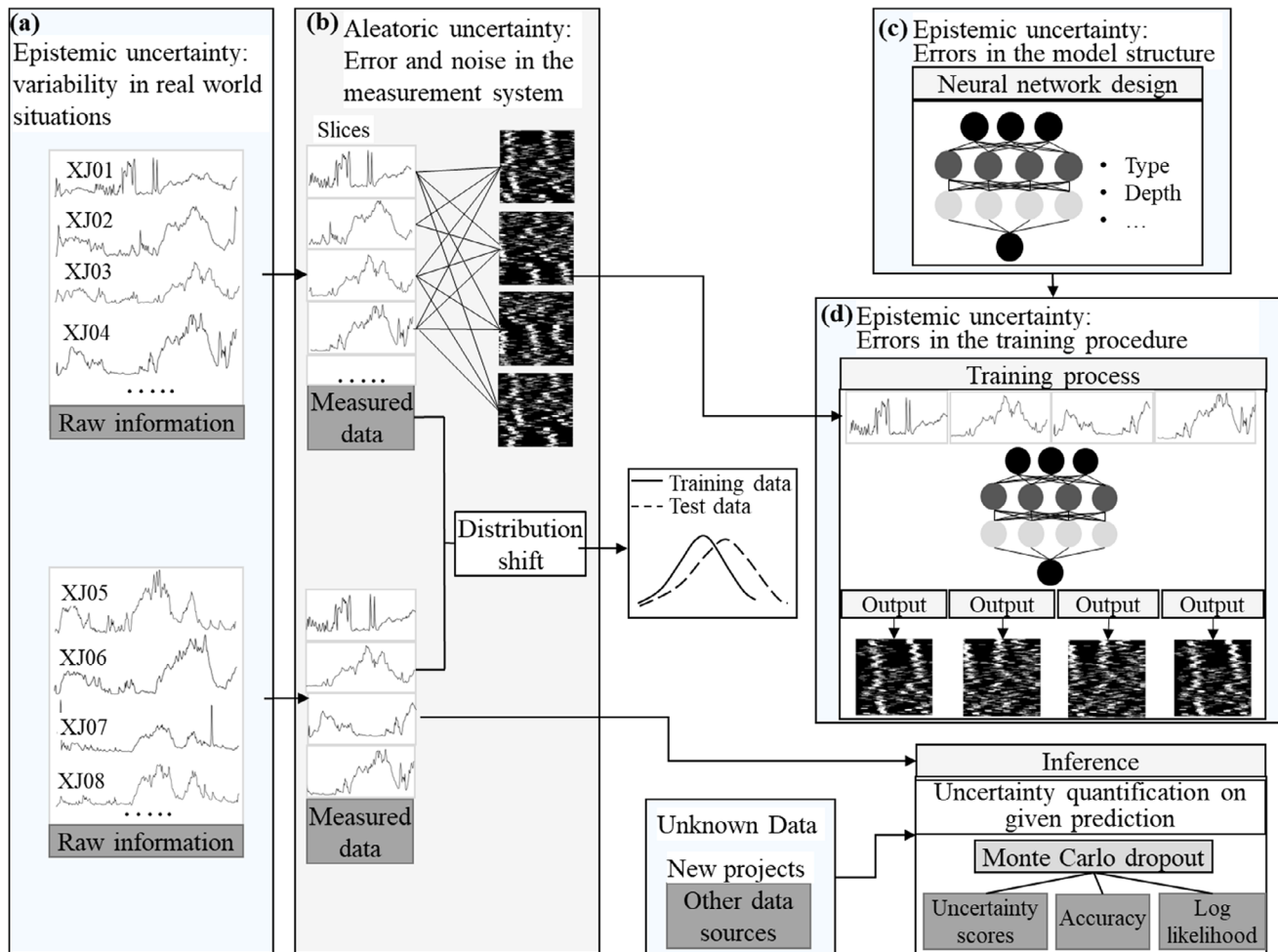
FIGURE 7 Source of epistemic and aleatoric uncertainty considered in the present Bayesian U-Net for application to soil boundary segmentation using CPT data.
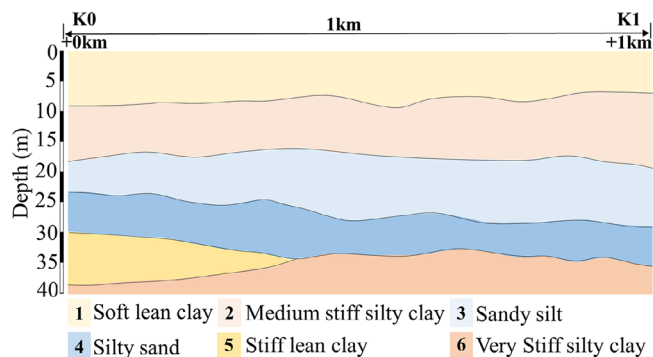
FIGURE 8 Typical geological cross-section in the Suzhou area.

4 | EXPERIMENTS

4.1 | Benchmark profile

Figure 8 shows the expert design subsurface profile between stations K0 and K1 developed for the Suzhou No.

6 metro line project and is considered here as a benchmark for comparison with model predictions. The upper stratum is soft grayish yellow medium plasticity lean clay, with thickness ranging between 8 and 10 m, underlain by medium stiff gray-brown medium plasticity silty clay with thickness ranging from 10 to 12 m. This is followed by 5 to 10 m thick medium dense to dense gray sandy silt, underlain with dense silty sand at a depth of approximately 25 to 35 m. Very stiff grayish yellow medium plasticity silty clay exists below that depth, occasionally interlaced with stiff low to medium plasticity lean clay.

4.2 | Random sampling for hyperparameters tuning

Among the hyperparameters considered for tuning, p is varied from 0 to 0.5. The parameter τ is varied between 0 and 1 to introduce an appropriate degree of data uncertainty considering that all data are standardized between

TABLE 1 Model metrics under different sampling scale of (p, τ).

	Metrics	Min	Max	Mean	SD
Sample = 100	ACC	0.7897	0.9979	0.9226	0.0979
	MPV	0.0019	0.2318	0.0837	0.0983
	NPL	0.9307	3.2749	1.4139	0.4945
	ER	0.0011	0.1722	0.0794	0.0753
Sample = 200	ACC	0.7897	0.9979	0.9379	0.0915
	MPV	0.0011	0.2319	0.0707	0.0919
	NPL	0.9248	3.6800	1.4237	0.4805
	ER	0.0012	0.1722	0.0731	0.0724
Sample = 300	ACC	0.7897	0.9979	0.9410	0.0898
	MPV	0.0005	0.2319	0.0684	0.0905
	NPL	0.9204	3.6800	1.4569	0.5286
	ER	0.0010	0.1755	0.0724	0.0719

Abbreviations: ACC, accuracy; ER, error; MPV, mean predictive variance; NPL, negative predictive log-likelihood; SD, standard deviation.

0 and 1 during training. Rather than an exhaustive and expensive grid search approach, this paper approximates the uncertainty distribution using stochastic sampling of p and τ from $(0, 0.5)$ and $(0, 1)$, respectively. For each selected pair, the following metrics over model prediction on validation datasets are calculated: accuracy (ACC), MPV, NPL and error (ER), where ACC and ER can be defined as

$$\text{ACC} = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

$$\text{ER} = \sqrt{\frac{1}{N_T} \sum_{n=1}^{N_T} ||y_n - y_n^*||^2} \tag{11}$$

where TN and TP represent the true negative and true positive cases, respectively, and FN and FP represent the false negative and false positive cases, respectively. y_n and y_n^* represent the real and predicted output, respectively, corresponding to the nth test input. N_T represents the total number of inputs in the test dataset. The weights are initialized with "He" normal distributions, and the network is trained using adaptive moment estimation for 100 epochs at a learning rate of 10^{-4}. The batch size is 155.

Table 1 shows the influence of the sampling scale for various (p, τ) pairs. Only image-wise MPV and NPL are adopted as the uncertainty estimate for efficient hyperparameter tuning. As the sampling scale increases, the mean and standard deviation of the uncertainty decrease and finally stabilize at a sample scale of 300. Figure 9 presents the contour maps of the MPV, ACC, and ER in $p - \tau$ space under sampling scale of 100, 200, and 300. The MPV is reduced with a lower p, resulting from the reduced variations (dropout) in the network hidden layers. The NPL is small when a small τ is obtained, resulting in

less penalty regularization in the last term of the NPL. The ACC reaches a maximum value of 0.9979 when $(p, \tau) = (0.441, 0.507)$, which gives the most accurate model configuration.

The development of training and validation accuracy (loss) over 100 epochs under optimal model configuration is shown in Figure 10. There is a notable improvement in both training accuracy and loss over the first 20 epochs before gradually plateauing at 0.9976 and 0.0275, respectively. For some qualitative examination of model performance, eight random predictive outputs obtained by the optimal model are shown in Figure 11. Upon visual inspection, the generated images are quite similar to the real outputs. This indicates that the tuned optimal model could well extract the image features and locate the soil boundaries accurately, which should lay the foundation for the following analyses.

4.3 | Visualization and benchmarking

To validate the model robustness and visualize the predicted subsurface profile, 24 consecutive 40-m deep CPTs are selected to construct a testing dataset following the domain adaptive augmentation strategy. The predicted boundaries are concatenated and elongated for clearer illustration as shown in Figure 12a. The figure shows that the soil boundaries stretch continuously across the section. The result is benchmarked against the GT in Figure 12b. Comparisons of key features are highlighted using green boxes from a to g. Based on visual inspection, the UGU-Net shows more similar results with the real output. The soil strata within boundaries are assigned with soil types based on previous exploration tests and soil classification results as shown in Figure 13. The upper soil stratum is soft lean clay, underlain with medium stiff silty clay. Below that, stratum of sandy silt to silty sand occurs, followed by stiff lean clay and very stiff silty clay by sequence. Visual comparison between Figures 13 and 8 indicates that the predicted soil profile matches reasonably well with the benchmark profile (cover the same CPT points).

4.4 | Extensive experiments

This experiment extends the prediction to a differently distributed open-source dataset, thus providing further evidence for model robustness and generalization ability. The CPT data collected at the NGES (National Geotechnical Experimentation Site) at Texas A&M University (clay site; Briaud, 2000) are used to construct the testing dataset and investigate the model performance. Figure 14 shows the curve (q_c and f_s) of an exemplar

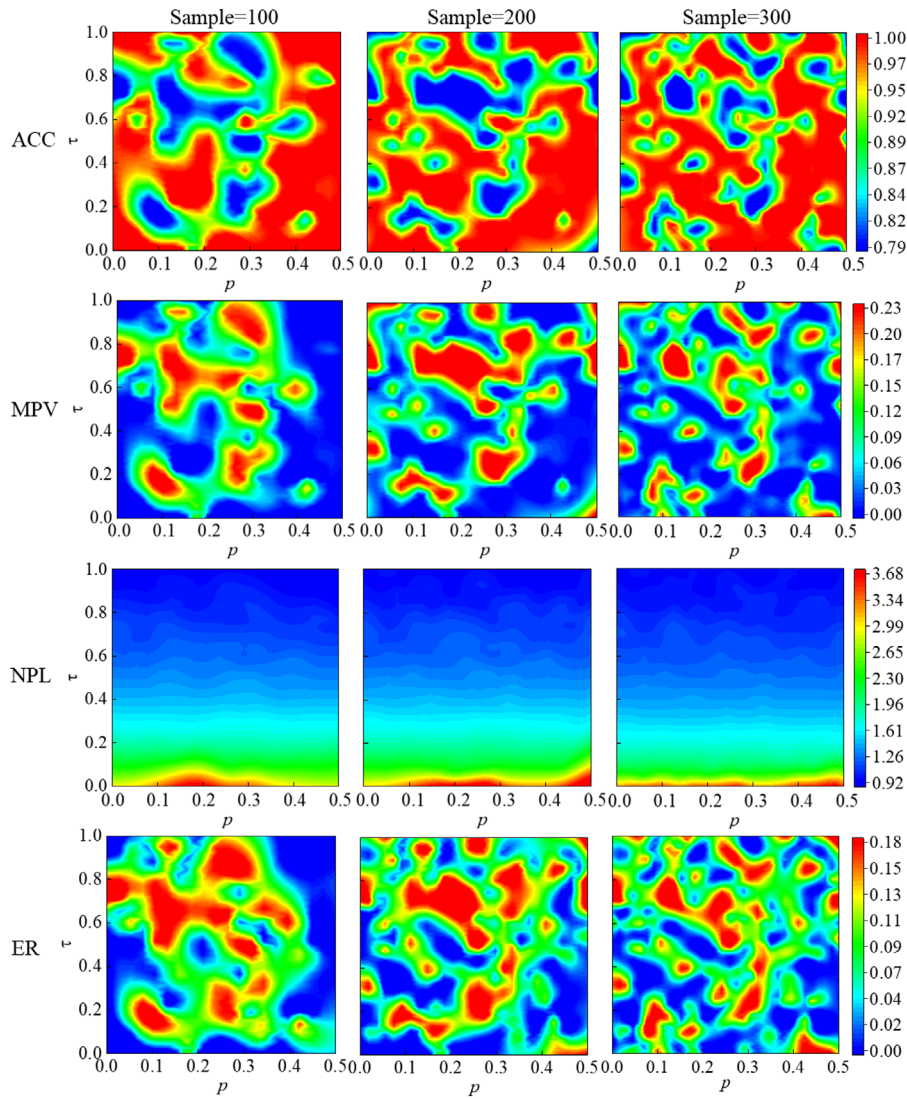**FIGURE 9** Contour map of different metrics with different p and τ.
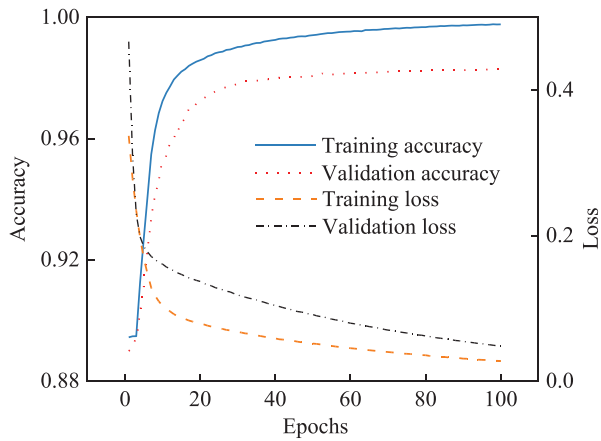

FIGURE 10 Development of accuracy and loss during training of the proposed UGU-Net on 100 epochs for $(p, \tau) = (0.441, 0.507)$.
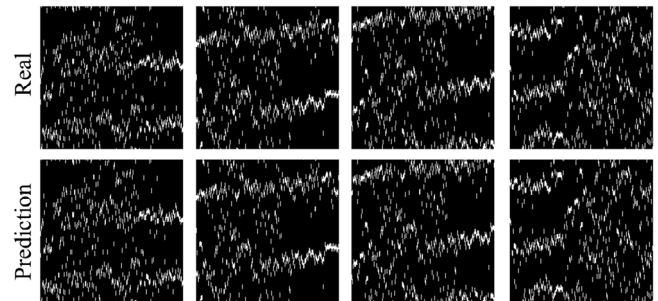

FIGURE 11 Exemplar UGU-Net-predicted outputs of soil boundaries and comparison with the ground truth (GT) measurements from the Suzhou metro line project; $(p, \tau) = (0.441, 0.507)$.

FIGURE 12 One predicted output by some state-of-the-art (SOTA) methods and their comparison with that by UGU-Net and GT.

FIGURE 13 Visualization of the subsurface cross-section based on predicted soil boundaries by UGU-Net.

CPT, where the soil behavioral type (SBT) classification proposed by Robertson (2016) is achieved, primarily including 3, 4, 5, 6, 8, and 9, with each number explained in footnotes. This site consists of a sequence of very stiff to hard clays with silt seams as illustrated in the borehole log. The predicted binary image by UGU-Net is aligned with the SBT classification and borehole log, where white lines represent predicted locations of soil boundaries. The depth-dependent soil strata distribution is achieved by filling the soil type within the soil boundaries. Compared with borehole log, the SBT results present denser soil boundaries. For example, corresponding to the stratum of clay with silt seams in borehole log, the SBT results further divide it into clayey silt (SBT 4) and sand mixtures (SBT 5).

Comparison shows that the predictive soil boundaries are generally consistent with the SBT classification results.

FIGURE 14 Prediction on NGES and comparison with soil behavioral type (SBT) results and borehole log.

Compared with borehole log, it provides a more detailed stratigraphic distribution and shows the presence of the boundary between silt mixtures and sand mixtures. However, the boundary between clay and silty clay is not figured out, probably because the data variation is not drastic enough for the model to detect the boundary between these two adjacent strata. This extensive experiment further validates the generalization ability of the network to heterogeneous datasets. Moreover, it shows limitations in terms of discriminating soil types of similar characteristics and properties. This could potentially be alleviated if trained on the wider dataset.

5 | SENSITIVITY ANALYSES

Testing datasets similar as that in previous experiments are utilized throughout the sensitivity analyses. The common segmentation evaluation metrics intersection over union (IoU) and dice coefficient (DC) score are used to evaluate the segmentation performance:

$$DC = \frac{2TP}{2TP + FP + FN} \quad (12)$$

$$IoU = \frac{TP}{TP + FP + FN} \quad (13)$$

5.1 | Threshold value

To reinforce the training outputs based on the uncertainty estimate, the soil boundaries are thickened by 0.2 m for pix-

TABLE 2 Influence of T_u on uncertainty degree.

T_u/PPV_{max}	0.5	0.6	0.7	0.8	0.9	0.95	0.99
R_u	0.912	0.911	0.911	0.911	**0.912**	0.914	0.918
ACC	0.9938	0.9938	0.9939	0.9942	**0.9961**	0.9950	0.9951
DC	0.9800	0.9790	0.9780	0.9776	**0.9809**	0.9781	0.9791
ER	0.0763	0.0763	0.0724	0.0679	**0.0542**	0.0601	0.0551
Δ ACC	0	0	0.0001	0.0003	**0.0011**	0.0008	−0.001

Abbreviations: ACC, accuracy; DC, dice coefficient; ER, error; PPV, positive predictive value.
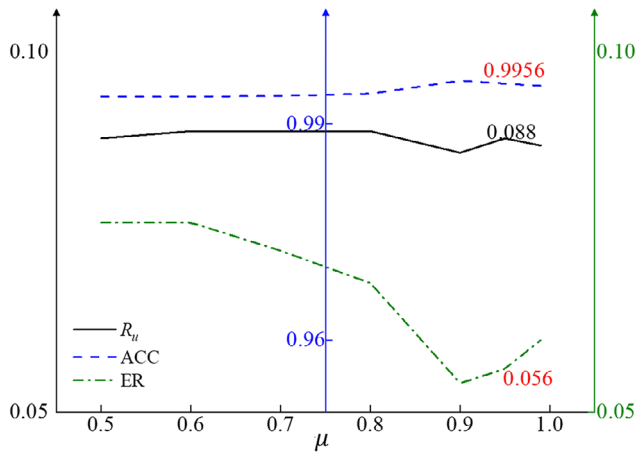
FIGURE 15 Variation of key performance with the increase of μ; "black line": variation of R_u; "blue dashed line": variation of ACC; "green dotted line": variation of ER.

FIGURE 16 2D distribution of the positive predictive value (PPV) greater than a given threshold where $\mu = T_u/\text{PPV}_{max}$.

els in set r_1^*. Let PPV_{max} represent the maximum value of PPV, fine-tuning of $\mu = T_u/\text{PPV}_{max}$ is explored from 0.5 to 0.99 to extract the uncertain pixel locations. Figure 15 and Table 2 show the influence of μ on key performance metrics. The degree of uncertainty (R_u) determined by μ reaches a maximum value of 0.088 when $\mu = 0.9$. There is a significant increase in ACC when μ is increased from 0.8 to 0.9.

Figure 16 shows the distribution of uncertainty in the PPV map for different μ, where the most effective uncertainty reduction can be observed when μ increases from 0.8 to 0.9, indicating that the PPV value is mostly distributed within $[0.8\text{PPV}_{max}, 0.9\text{PPV}_{max}]$. Therefore, it can be reasonably assumed that $\mu = 0.9$ delivers the most effective extraction of uncertain pixels.

An exemplar prediction of the soil profile following reinforcement under different μ is shown in Figure 17. To present how different μ influence the prediction performance, specific attention to the areas marked in green box shows the comparison before ($\mu = 1.0$) and after reinforcement ($0.5 \leq \mu < 1.0$). When $\mu > 0.9$, most of the soil boundaries are "ignored" and remain unchanged. By comparison, if $\mu < 0.9$, over reinforcement occurs and results in some adjacent boundaries being stuck, which may cause confusion for modelers. Thus, augmentation with $\mu =$

0.9 is shown to be effective in improving segmentation performance.

5.2 | Soil boundary "thickness"

In the previous experiment, the soil boundary "thickness" was set to 0.8 m (eight consecutive pixels were used to represent the boundary) as a default value; clearly, this is an indication of location uncertainty rather than an attempt to reflect the physical reality. Here, sensitivity analyses are conducted to explore the role of the thickness on the segmentation results. To this end, the boundary thickness is varied from 0.4 to 1.0 m as illustrated in Figure 18, which shows the cropped small section of a training label. The value of μ is set as 0.9 for these comparisons based on the results from the previous analyses.

Table 3 shows the evaluation metrics obtained under different soil thicknesses. Although the ACC is the highest, while the ER is the lowest for a thickness of 0.4 m, it does not lead to the best segmentation performance since there are more background (black) pixels. Usually, the ACC quantifies the proportion of correctly predicted labels out of the total number of predictions. If the dataset is imbalanced (e.g., containing substantial background pixels), a
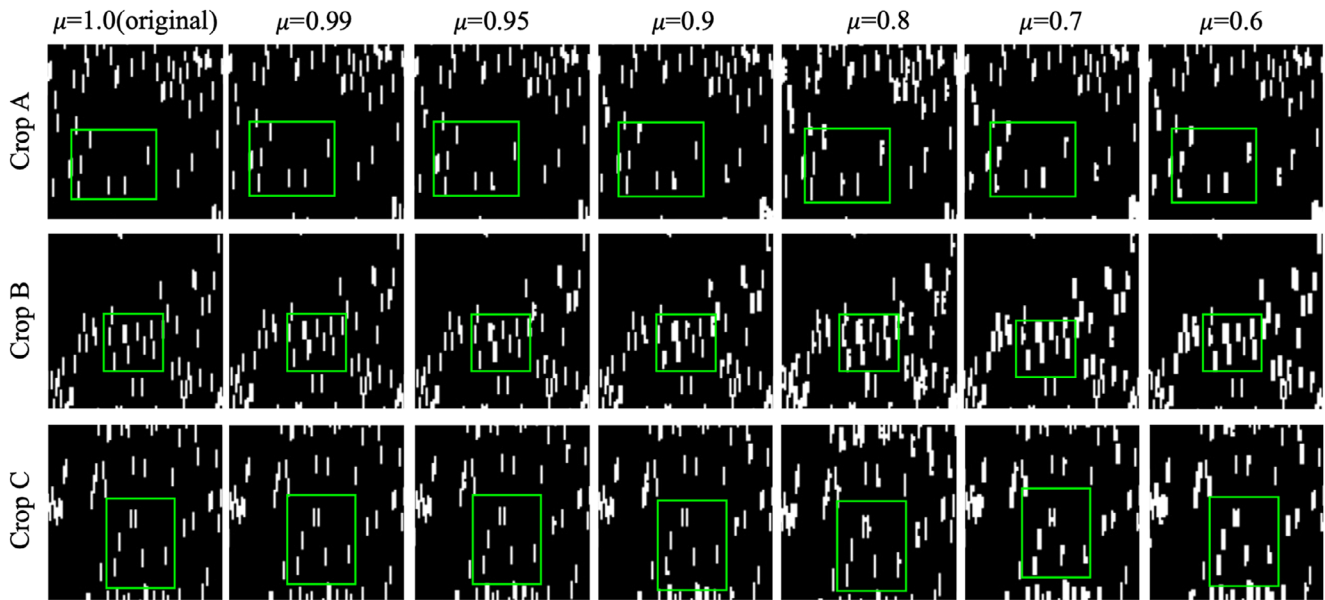
FIGURE 17 Different cropped sections in the predicted output by UGU-Net for illustration of soil boundary variation for different μ.
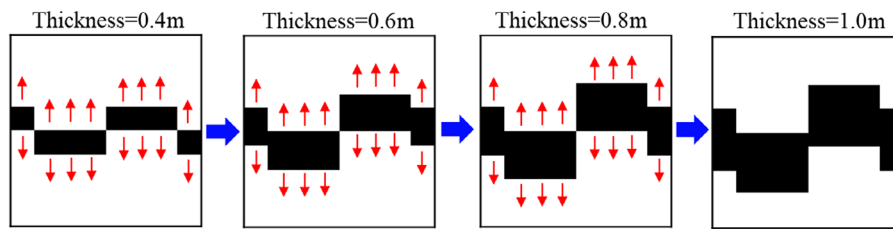

FIGURE 18 Illustration of thickness increment by one pixel in a fraction of one training output.

TABLE 3 Influence of boundary thickness on model metrics with $T_u/\text{PPV}_{\max} = 0.9$

	Thickness = 0.4	Thickness = 0.6	Thickness = 0.8	Thickness = 1.0
MPV	0.0004	0.0003	0.0002	0.0006
ACC	0.9978	0.9925	0.9972	0.9968
DC	0.9844	0.9586	0.9885	0.9826
ER	0.039	0.0748	0.0648	0.0493
R_u	0.962	0.951	0.9293	0.985

Abbreviations: ACC, accuracy; DC, dice coefficient; ER, error; MPV, mean predictive variance; PPV, positive predictive value.

model can still achieve high accuracy by correctly predicting the dominant class (in this case, the background), while still performing poorly on the minority class (such as the object of interest). By contrast, the DC measures the overlap between the predicted and true regions (or labels) and is sensitive to both false positives and false negatives. In Table 3, the DC score reaches the maximum of 0.9885, while the MPV and the R_u reach the minimum when the thickness equals to 0.8 m, indicating optimal seg-

mentation performance with low uncertainty under such configuration.

Figure 19 shows the predicted outputs which correspond to a plane covering several CPTs to enable comparisons and visual inspection of the result. Based on the predicted output, the subsurface profile could be reconstructed by reversing all previous actions in Section 3.1. The NLP maps show that high uncertainty could be observed when the thickness is between 0.4 and 0.6 m. As the thickness increases, the predicted boundaries get clearer and more continuous.

5.3 | Comparative studies

To quantitively evaluate the predicted outputs by UGU-Net, predictions are benchmarked against the relevant SOTA U-Net methods, including a traditional U-Net, Attention U-Net, U-Net++, and RCNN-Attention U-Net. The same training and validation datasets with that in previous experiments are adopted when exploring other networks. All models are trained over 100 epochs in the
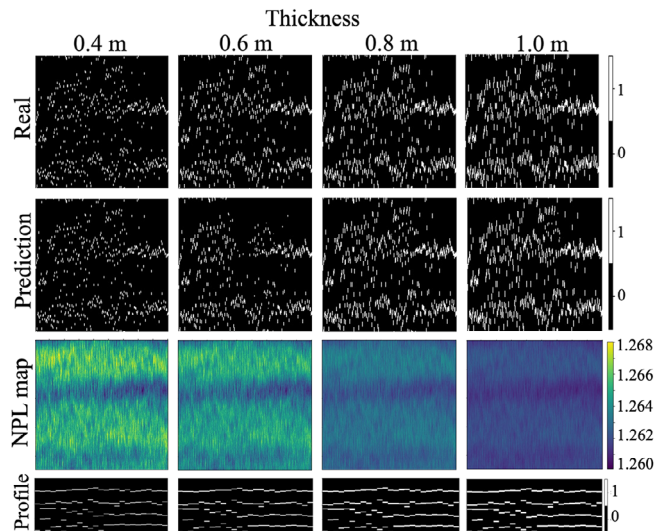
FIGURE 19 UGU-Net-predicted soil boundaries, negative predictive log-likelihood (NPL) distribution and reversed soil profile, compared to the GT for a variation in the annotated boundary thickness; Section K0 + 0 m to K0 + 1000 m.

TABLE 4 Segmentation performance of different SOTA methods.

Model	DC	IoU
UGU-Net (present)	0.9885	0.9884
U-Net (Ronneberger et al., 2015)	0.6051	0.5988
Attention U-Net (Oktay et al., 2018)	0.9015	0.9004
U-Net++ (Zongwei Zhou et al., 2018)	0.7055	0.6996
RCNN-Attention U-Net (Tajbakhsh et al., 2020)	0.7941	0.7846

Abbreviations: DC, dice coefficient; IoU, intersection over union; SOTA, state-of-the-art; UGU-Net, uncertainty-guided U-Net.

same experimental environment at a learning rate of 10^{-4}. The batch size is 155. The testing dataset is constructed through randomly cropping 100 data matrices from $\mathbb{D}$, with a size of $224 \times 224 \times 1$.

The models shown in Table 4 finally predict such testing datasets, and the results are compared. Table 4 shows that the UGU-Net achieves the best performance with the highest DC score of 0.9885. The DC score acquired by the attention-based U-Net is greater than that acquired by U-Net and U-Net++ (non-attention-based), indicating that attention on local image details indeed improves model performance. The predicted outputs by the different methods are presented in Figure 20. The results indicate that the outputs by UGU-Net are most similar to the GT. Attention-based U-Net captures global structures but fails to predict the detailed boundaries, while non-attention-based U-Net models are notably worse for boundary prediction.
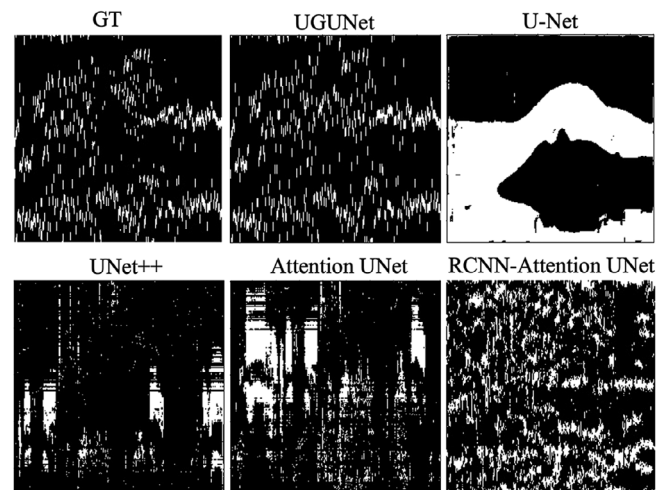
FIGURE 20 One predicted output by some SOTA methods and their comparison with that by UGU-Net and GT.

6 | DISCUSSION

The uncertainty guidance network training increases model robustness and quantifies relevant uncertainties, providing more reliable prediction and fostering more informed decision making. It realizes soil boundary segmentation end to end and relies less on human intervention, compared with traditional manual delineation. It reduces reliance on manual intervention, allowing it to be more efficient than traditional delineation approaches. Training the model takes just a few minutes, and predictions on test data take only seconds. In contrast, geostatistical and probabilistic methods typically require dozens of hours for data inference, thus demonstrating substantial savings in computational time and cost.

This impact could be expanded to other applications in geotechnical engineering. Uncertainty information from the soil profile can be used to inform site investigation campaigns such that test locations are optimized to reduce uncertainty most effectively. In turn, this uncertainty information will also help improve geotechnical design by allowing geotechnical engineers to design according to a prescribed level of risk. This will allow engineers to balance the cost of over-design with the risk of under-design.

Despite the improvements achieved by UGU-Net, several limitations remain. The model is trained on a limited number of samples and may face difficulties when applied to unique soil types, such as expansive soils, loess, or highly organic soils. For example, the failure to discriminate between clay and silty clay as observed in Section 4.3, demonstrates the model's limitation in segmenting adjacent soil strata with minimal variation in CPT measurement. Additionally, geological conditions such as

interlayers can introduce complexities not fully captured by the proposed method.

Furthermore, conducting MC dropout for uncertainty estimation introduces a trade-off. Running multiple forward passes during inference increases computational time, and MC dropout, as an approximation of Bayesian inference, may not always capture the true posterior distribution. The dropout rate also affects the quality of uncertainty estimates—low dropout may lead to overconfident predictions, while high dropout increases randomness and reduces reliability.

Future work should explore the effect of dataset size on uncertainty distribution and how these estimates enhance the robustness on test data. An attention mechanism could be added to focus the model on uncertain regions, and 3D DNN models could be explored to capture spatial soil boundary information. Additionally, improving the computational efficiency of MC dropout remains an important avenue for further development.

7 | CONCLUSION

This paper has described a novel "UGU-Net" DNN for soil boundary segmentation based on CPT data. The network is capable of segmenting soil profiles as well as estimating the associated uncertainty map. A Bayesian U-Net is devised, with varying dilation rate to include an enlarged receptive field during convolution and varying dropout rate to perform Bayesian inference while also alleviating overfitting and regularization. A sensitivity analysis is performed on the model hyperparameters where the hyperparameter pairs (p, τ) are randomly sampled to calculate the MPV and NPL metrics; the optimal model configuration is achieved using 300 samplings and for a value of $(p, \tau) = (0.441, 0.507)$.

Additional sensitivity analyses are conducted on the optimal DNN model to explore the influence of the uncertainty map threshold value and soil thickness on model performance. The DC score is shown to reach a maximum when the soil thickness equals to 0.8 m, representing the best segmentation results. The predicted subsurface is visualized and compared with a benchmark profile, showing that the proposed model can predict soil boundaries with high efficiency and effectiveness.

NOTATION

p	dropout rate—probability of dropping out neurons at each iteration
T_u	Threshold for extracting uncertain pixel locations
$\xi(\cdot)$	a function to extract pixels where PPV is smaller than T_u
$\mathbf{X}$	training inputs
$x_i(i = 1 \dots N)$	one element in $\mathbf{X}$
N	number of training inputs
$\mathbf{X}$	training inputs
$y_i(i = 1 \dots N)$	one element in $\mathbf{Y}$
f	a function to describe the relation between x_i and y_i
$p(\cdot)$	probability distribution
x^*	a new input not included in $\mathbf{X}$
y^*	predicted output given the x^*
ω	a set of random variables
$q(\cdot)$	variation distribution
f^*	a function to describe the relation between x^* and y^*
L	the number of hidden layers
$\mathbf{W} = [\mathbf{W}_i]_{i=1}^L$	the weight matrix with $\mathbf{W}_i$ for each hidden layer
$\mathcal{N}(\cdot, \cdot)$	Gaussian distribution
$\mathbf{F} = [\boldsymbol{f}_1, \boldsymbol{f}_2, ..\boldsymbol{f}_N]$	set of all function for each set of input and output
$\mathbf{K}(\mathbf{X}, \mathbf{X})$	covariance function
τ	precision coefficient
$z_i(i = 1, L)$	binary vectors
K_{i-1}	the dimension for each hidden layer i
$Bernoulli(\cdot)$	Bernoulli distribution
$E(\cdot)$	the loss function
$\hat{y}$	the output of the neural network
$\mathbf{b}_i$	bias matrix for each hidden layer i
$\mathcal{L}_{dropout}$	the objective function with dropout layers
$\hat{\boldsymbol{\omega}}_t(t = 1, ..T)$	t-th sampling from the distribution $\boldsymbol{\omega}$
T	the number of samplings
s^*	the total number of pixels in the output
$I^i(i = 1, ..515)$	CPT data matrix for each hole
N^i	the number of records for each hole
D^i	the data matrix denoting CPT data
d^i	cropped sections of D^i
O^i	the annotated label for matrix D^i
o^i	cropped sections of O^i
n_i	the number input sub-matrix cropped from each CPT curve
$\mathbb{D}$	the database for pairs of sub-input and sub-output matrixes
$\mathrm{C}_n^m = \frac{n!}{m!(n-m)!}$	combination function
$\mathrm{A}_n^m = \frac{n!}{(n-m)!}$	permutation function
N_T	the number of test inputs
$\mu = T_u/PPV_{\max}$	
r^*	set of pixels where the PPV is higher than T_u

r_1^* pixels representing soil boundaries in set r^*

R_u uncertain degree

REFERENCES

Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., & Acharya, U. R. (2021). A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion*, 76, 243–297.

Ahmed, U., & Lin, J. C. W. (2023). Robust adversarial uncertainty quantification for deep learning fine-tuning. *Journal of Supercomputing*, 68, 11355–11386.

Angelelli, M., Gervasi, M., & Ciavolino, E. (2024). Representations of epistemic uncertainty and awareness in data-driven strategies. *Soft Computing*, Advance online publication. https://doi.org/10.1007/s00500-024-09661-8

Baldi, P., & Sadowski, P. J. (2013). Understanding dropout. *Advances in neural information processing systems*, 26.

Bernardo, J. M., & Smith, A. F. (2009). *Bayesian theory*. John Wiley & Sons.

Blundell, C., Cornebise, J., Kavukcuoglu, K., & Wiersta, D. (2015). Weight uncertainty in neural networks. In F. Bach & D. Blei (Eds.), *Proceedings of the 32nd International Conference on International Conference on Machine Learning* (Vol. 37, pp. 1613–1622). ACM.

Briaud, J. L. (2000). The National Geotechnical Experimentation Sites at Texas A&M University: Clay and sand. A summary. In J. Benoît & A. J. Lutenegger (Eds.), *National Geotechnical Experimentation Sites* (Vol. 93, pp. 26–51). Geotechnical Special Publication.

Chen, D., Zhu, H., & Yang, S. (2023). UC-SFDA: Source-free domain adaptation via uncertainty prediction and evidence-based contrastive learning. *Knowledge-Based Systems*, 275, 110728.

Chern, W. C., Nguyen, T. V., Asari, V. K., & Kim, H. (2023). Impact of loss functions on semantic segmentation in far-field monitoring. *Computer-Aided Civil and Infrastructure Engineering*, 38(3), 372–390.

Diao, Z., Jiang, H., & Shi, T. (2022). A unified uncertainty network for tumour segmentation using uncertainty cross entropy and prototype similarity. *Knowledge-Based Systems*, 246, 108739.

Erharter, G. H., Oberhollenzer, S., Fankhauser, A., Marte, R., & Marcher, T. (2021). Learning decision boundaries for cone penetration test classification. *Computer-Aided Civil and Infrastructure Engineering*, 36(4), 489–503.

Gal, Y. (2016). *Uncertainty in deep learning* [Doctoral dissection]. University of Cambridge.

Gal, Y., & Ghahramani, Z. (2015). *Bayesian convolutional neural networks with Bernoulli approximate variational inference*. arXiv: 1506.02158, 2015. https://doi.org/10.48550/arXiv.1506.02158

Gawlikowski, J., Yasso, C. R. N., Ali, M., Lee, J., Humt, M., Feng, J. X., Kruspe, A., Triebel, R., Jung, P., Roscher, R., Shahzad, M., Yang, W., Bamler, R., & Zhu, X. X. (2022). *A survey of uncertainty in deep neural network*. arXiv: 2107.03342v3. https://doi.org/10.1007/s10462-023-10562-9

Hariri, R. H., Fredericks, E. M., & Bowers, K. M. (2019). Uncertainty in big data analysis: Survey, opportunities, and challenges. *Journal of Big Data*, 6(1), 1–16.

Hinton, G. E., Srivastava, N., Krizhavsky, A., Sutskever, I., & Salakhutdinov, R. R. (2012). *Improving neural networks by preventing coadaptation of feature detectors*. arXiv: 1207.0580. https://doi.org/10.48550/arXiv.1207.058

Hüllermeie, E., & Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods, *Machine Learning*, 110(3), 457–506.

Kendall, A. (2019). *Geometry and uncertainty in deep learning for computer vision* [Doctoral dissertation]. University of Cambridge.

Kendall, A., & Gal, Y. (2017). What uncertainties do we need in Bayesian deep learning for computer vision? *Advances in Neural Information Processing Systems*, 30, Long Beach, CA.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 1097–1105.

Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2016). *Simple and scalable predictive uncertainty estimation using deep ensembles*. arXiv: 1612.0147v1. https://doi.org/10.48550/arXiv.1612.01474

Leibig, C., Allken, V., Ayhan, M. S., Berens, P., & Wahl, S. (2017). Leveraging uncertainty information from deep neural networks for disease detection. *Scientific Reports*, 7, 1–14.

Li, Z., Wang, X. R., Wang, H., & Liang, R. Y. (2016). Quantifying stratigraphic uncertainties by stochastic simulation techniques based on Markov random field. *Engineering Geology*, 201, 106–122.

Mayne, P. W., Christopher, B. R., & DeJong, J. (2002). *Subsurface investigations—geotechnical site characterization. Reference Manual (No. FHWA-NHI-01-031)*, United States Federal Highway Administration.

Midwinter, M., Al-Sabbag, Z. A., & Yeum, C. M. (2023). Unsupervised defect segmentation with pose priors. *Computer-Aided Civil and Infrastructure Engineering*, 38(17), 2455–2471.

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., & Glocker, B. (2018). *Attention U-Net: Learning where to look for the pancreas*. arXiv preprint arXiv:1804.03999. https://doi.org/10.48550/arXiv.1804.03999

Osawa, K., Swaroop, S., Khan, M. E. E., Jain, A., Eschenhagen, R., Turner, R. E., & Yokota, R. (2019). *Practical deep learning with Bayesian principles*. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, & R. Garnett (Eds.) *Advances in neural information processing systems* (pp. 4287–4299). Curran Associates.

Padmapriya, J., & Sasilatha, T. (2023). Deep learning based multi-labelled soil classification and empirical estimation toward sustainable agriculture. *Engineering Applications of Artificial Intelligence*, 119, 105690.

Paschali, M., Simson, W., Roy, A. G., Naeem, M. F., Göbl, R., Wachinger, C., & Navab, N. (2019). *Data augmentation with manifold exploring geometric transformations for increased performance and robustness*. arXiv preprint arXiv:1901.04420. https://doi.org/10.48550/arXiv.1901.04420

Provenzano, P. (2003). A fuzzy-neural network method for modeling uncertainties in soil-structure interaction problems. *Computer-Aided Civil and Infrastructure Engineering*, *18*(6), 391–411.

Qi, X., Li, D., Phoon, K. K., Cao, Z., & Tang, X. (2016). Simulation of geologic uncertainty using coupled Markov chain. *Engineering Geology*, *207*, 129–140.

Robertson, P. K. (2016). Cone penetration test (CPT)-based soil behaviour type (SBT) classification system—An update. *Canadian Geotechnical Journal*, *53*(12), 1910–1927.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Munich, Germany (pp. 234–241).

Shi, C., & Wang, Y. (2022). Data-driven construction of three-dimensional subsurface geological models from limited site-specific boreholes and prior geological knowledge for underground digital twin. *Tunnelling and Underground Space Technology*, *126*, 104493.

Srivastava, N., Hinton, G., Krizhevasky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research (JMLR)*, *15*(1), 1929–1958.

Tajbakhsh, N., Jeyaseelan, L., Li, Q., Chiang, J. N., Wu, Z., & Ding, X. (2020). Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Medical Image Analysis*, *63*, 101693.

Tang, P., Yang, P., Nie, D., Wu, X., Zhou, J., & Wang, Y. (2022). Unified medical image segmentation by learning from uncertainty in an end-to-end manner. *Knowledge-Based Systems*, *241*, 108215.

Tong, Z., Ma, T., Zhang, W., & Huyan, J. (2023). Evidential transformer for pavement distress segmentation, *Computer-Aided Civil and Infrastructure Engineering*, *38*(16), 2317–2338.

Yao, B., Dong, L., Qiu, X., Song, K., Yan, D., & Peng, C. (2024). Uncertainty-guided contrastive learning for weakly supervised point cloud segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, *62*, 1–13.

Zhang, W., Ma, Z. M., Das, S., Weng, T. W. L., Megretski, A., Daniel, L., & Nguyen, L. M. (2024). One step closer to unbiased aleatoric uncertainty estimation. *Proceedings of the AAAI Conference on Artificial Intelligence*, *38*(15), 16857–16864.

Zhang, X., Chan, F. T., & Mahadevan, S. (2022). Explainable machine learning in image classification models: An uncertainty quantification perspective. *Knowledge-Based Systems*, *243*, 108418.

Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018). UNet++: A nested U-Net architecture for medical image segmentation. In D. Stoyanov (Eds.), Lecture notes in computer Science: *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* (*Vol*. 11045, pp. 3–11). Springer.

Zhou, Z., Zhang, J., & Gong, C. (2023). Hybrid semantic segmentation for tunnel lining cracks based on Swin transformer and convolutional neural network. *Computer-Aided Civil and Infrastructure Engineering*, *38*(17), 2491–2510.