J. Inst. Eng. India Ser. B (June 2023) 104(3):569–578
https://doi.org/10.1007/s40031-023-00875-2

ORIGINAL CONTRIBUTION

An Efficient Approach for Semantic Segmentation of Salt Domes in Seismic Images Using Improved UNET Architecture

Jyostna Devi Bodapati[1] · **RamaKrishna Sajja**[2] · **Veeranjaneyulu Naralasetti**[3]

Received: 7 July 2021 / Accepted: 16 March 2023 / Published online: 6 April 2023

Abstract Many areas of Earth's surface with large accumulations of gas and oil even have huge deposits of salt under the surface. Exploring such deposits helps many countries to increase the storage capacity of their Petroleum reserves and explore new ones. But finding such deposits is a herculean task. Expert seismic imaging requires human interpretation of salt bodies. But this leads to very biased and highly variable translations. So the idea behind this paper is to build an approach that accurately and automatically identifies if the seismic image contains any region of salt deposit or not. If a surface is found to have salt deposits, then it may contain the accumulations of oil or gas and even the salt domes or caverns can be used as a storage site for already available petroleum or oil. Since semantic segmentation classifies every pixel in the given image to its class label, this can be used to segment the salt deposits from the provided seismic images. In this paper, we introduce a variation of UNet, a popular segmentation model, for seismic image segmentation. We have added a batch normalization layer following every convolution layer as a deeper network helps extract better features which turned out to be true. Some interesting findings of this work are that the augmentation works well as this avoids over-fitting of the model. Sharpening as a post-processing technique has come up with a considerable amount of rise in the performance. Here, we prefer to use the metric as Intersection over Union (IoU) instead of accuracy as it is not as affected by the class imbalances that are inherent in foreground/background segmentation tasks. With the proposed methodology, we achieve an averaged IoU of 85.6 which is far better compared to the IoU achieved with the Segnet approach which stands at 77.

Keywords Salt deposits · Segmentation · Semantic segmentation · Convolutional neural network (ConvNet) · UNet · SegNet · Seismography · Salt domes · Intersection over Union (IoU)

Introduction

Several areas of the Earth that are rich in oil and natural gas also have massive salt deposits beneath the surface [1]. Due to the extremely low permeability of the salt, reservoirs of hydrocarbons such as natural gas or crude oil are trapped by overlying rock-salt formations. Hence, oil and gas exploration companies rely on the precise location of large salt deposits to identify hydrocarbon reservoirs such as crude oil or natural gas [2]. This task necessitates the visualization of underground structures, which is possible with seismic imaging [3]. Seismic imaging involves the emission of sound waves into underground structures of the Earth. The reflection of those sound waves from the various structures will be detected using Geo-phones and saved for the subsequent process of generating the 3-dimensional representation of the underground structures [4]. The process involved in seismic imaging is depicted in Fig. 1. Seismic images provide

✉ Jyostna Devi Bodapati
bjd_cse@vignan.ac.in

RamaKrishna Sajja
bobysajja@gmail.com

Veeranjaneyulu Naralasetti
drnvn_it@vignan.ac.in

[1] Department of Advanced Computer Science and Engineering, Vignan's Foundation for Science Technology and Research, Vadlamudi, India

[2] IAENG Life Member, Guntur, Andhra Pradesh, India

[3] Department of Information Technology and Computer Applications, Vignan's Foundation for Science Technology and Research, Vadlamudi, India

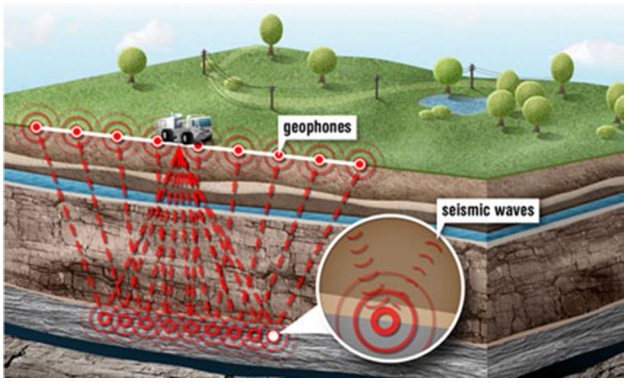

Fig. 1 Depicting the process involved seismic imaging [6]

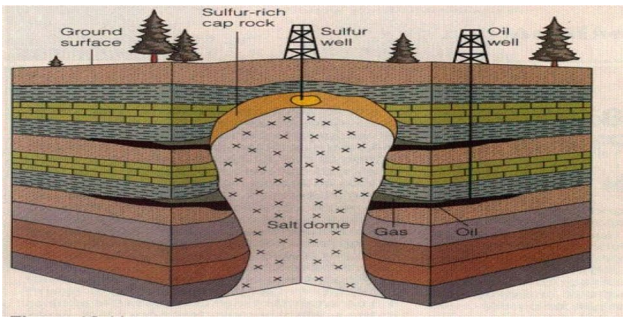

Fig. 2 Visualization of oil trapped between Salt Dome and surrounding rocks [7]

more information about the edges of various rocks, which can be used to determine whether or not a specific rock is in contact with a salt dome [5].

Figure 2 depicts a seismic image of a salt dome deforming the surrounding rocks, forming a trap for natural gas or oil. Unfortunately, the exact identification of large salt deposits from seismic images is notoriously difficult and often requires manual interpretation of seismic images by the domain experts [8]. Manual interpretation is not only time-consuming and expensive but also involves human bias, which can put oil exploration companies in potentially dangerous situations. In recent years, a number of machine learning tools have been introduced for the automatic interpretation of seismic images to speed-up the interpretation process and, to some extent, reduce the human bias [9]. Salt region detection from the seismic images can be posed as a classification, localization or semantic segmentation problems as shown in Fig. 3. If the problem is posed as a classification task, then the model simply needs to check whether the given seismic image contains salt deposits or not. Several machine learning models have been introduced in the literature for the classification of seismic images based on the existence of salt domes in the images [10]. Convolutional neural networks (CNNs) proved to be successful for solving variety of real-world problems in computer vision such as object detection [11] and image classification [12, 13] problems even with small number of training data [14, 15]. Successful application of CNNs piqued the interest of the research community in the application of CNNs for analyzing seismic images [16].

In addition to identifying the presence of salt deposits, sometimes it is essential to draw a boundary box representing the region of salt domes in the given seismic images, and hence this problem can be posed as a segmentation task. Separating the salt body from the rest of the background is especially important, which can be accomplished by classifying every pixel in the seismic image as salt region or not, and the task can be better posed as semantic segmentation [16]. Semantic segmentation task on seismic imaging has drawn focus of the research community by the launch of the competition held by TGS [17]. Semantic segmentation task is much more challenging than image classification [18], detection [19] and localization [20] tasks, and hence, the conventional machine learning approaches may not be appropriate for such complex tasks [4]. Standard convolutional neural network focuses its task on image classification, while segmentation requires producing the entire input image with the predicted mask. Encoder–decoder models such as UNet, SegNet, DeconvNet and FCN have become

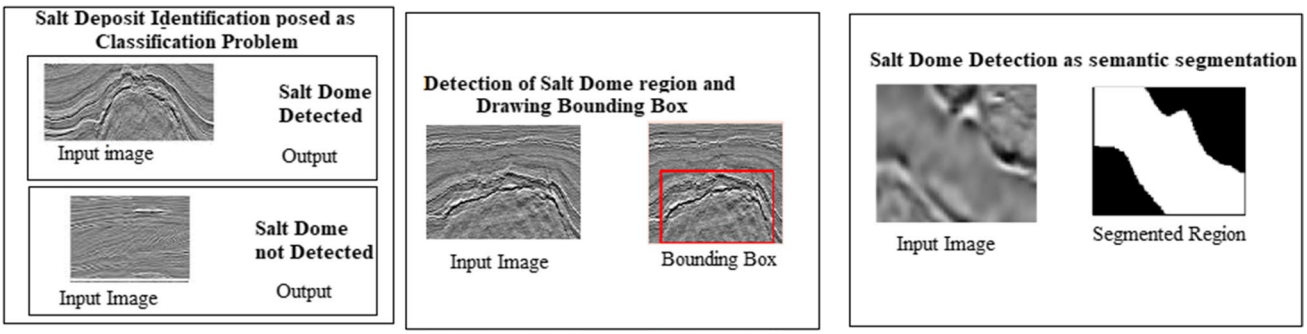

Fig. 3 Posing salt dome detection from seismic images as **a** classification **b** localization and **c** semantic segmentation problems

popular for salt dome segmentation from seismic imaging [21]. UNets are a variant of convolutional networks basically designed for semantic segmentation and are popular in the medical imaging domain. Numerous works have come up applying these UNet architectures for semantic segmentation of salt domes from seismic images. However, their performance is sub-optimal and still there is huge scope for improvement in the generalization of these models.

In this work, we present an efficient approach for segmenting salt deposit regions from seismic images. We approach the problem using semantic segmentation, in which each pixel of the seismic image is identified as belonging to a salt deposit region or not. We propose an improved version of the UNet architecture for segmenting salt regions from seismic images. Initially, all the images in the dataset are pre-processed, and the training images are used to train the proposed improved UNet architecture. The contractive path of the UNet is designed with a sequence of convolution operations interleaved with max pooling layers, where the expansion path is designed with transpose convolution operations. The UNet layers are designed in such a way that they promote better generalization and the masked regions are further post-processed for improving the scores. A high-pass filter is applied to the masks generated by the UNet to smooth out the segmentation. Augmentation is applied on the training dataset to avoid over-fitting of the proposed model and improve generalization. The proposed approach is evaluated on the publicly available benchmark TGS dataset. The proposed UNet architecture along with the sharpening operation as a post-processing technique resulted in significant increase in performance. The Intersection of Union (IoU) scores obtained by the proposed model are much better compared to the baseline models.

The method presented in this paper is based on the author's participation and it relies on training a deep convolutional neural network (CNN) for semantic segmentation. The architecture of the proposed network is inspired by the UNet model in combination with ResNet and DenseNet architectures. To

Literature Survey

Many researchers are drawn to the task of identifying and segmenting salt deposit regions in seismic images because it reveals the presence of hydrocarbon reserves. We intend to present a review of existing machine learning and deep learning approaches for salt deposit region segmentation in this section.

During the early stages, researchers considered using a variety of hand-crafted feature extraction approaches for the analysis of seismic images [4]. The work of Halpert and Clapp revealed that features extracted by a single feature extraction algorithm may not lead to accurate region segmentation. They also proposed using additional features such as dip variability, amplitude and frequency content of seismic image and demonstrated that using these additional features improves the segmentation results when compared to the existing approaches [9]. Following this, many other works were presented, including SalSi, which develops novel seismic attributes for salt mine identification [22]. In the subsequent methods, study of the 3D seismic images became a subject of interest [23]. The work of Amin and Deriche [1], which uses a multi directional edge detector, is worth discussing. Instead of focusing solely on the edges, their method computes gradients along the edges in combination with the diagonals. This makes the algorithm efficient although the dips along the salt edges are not so certain. Since this method is not affected by amplitude variations, it outperforms existing texture and gradient-based techniques. This method includes calculating the edge map of 3D seismic images and normalizing the data to improve the quality of the salt dome edges [24]. This is followed by the computation of gradient magnitudes in all the x, y, z and diagonal directions. The skeleton of the salt boundary is then obtained by blending and thresholding all of the available edge maps.

Research on salt deposit region segmentation has grown at an exponential rate since the introduction of neural network models and their success in real-time applications. Convolutional neural network (CNN) architectures, which were originally designed for classification, were successfully extended to perform segmentation. Among these, encoder–decoder models such as fully convolutional network (FCN), UNet, SegNet and DeconvNet have become popular for salt dome segmentation from seismic imaging [12]. FCNs are the simple extension to CNNs which are designed by replacing the final dense layers of CNN with convolutional and upsampling layers [16]. Finally, after upsampling to the required size, layers required for pixel level classification are attached to obtain the region prediction. The efforts toward comparing the performance of FCN with traditional approaches revealed that FCNs consider the local seismic patterns while learning the target region from seismic images, and hence avoid the coherent noises [5]. Another intriguing aspect is that, unlike traditional machine learning models, these deep neural network models are capable of directly learning features required to construct mapping between the input seismic signals and the target salt bodies, avoiding the need for manual attribute selection.

Among the encoder–decoder models, UNet-based architectures are popular due to their success in the medical imaging domain. UNet is a variant of convolutional networks basically designed for segmentation of regions of interest from the images. Unlike the conventional segmentation approach, UNet applies pixel-wise classification to obtain the region of interest. The encoder component of these

models is designed to reduce the spatial dimensions of the input image while increasing the number of channels. The output tensor produced by the encoder, called as the bottleneck layer and is passed as input to the decoder [25]. The decoder then enhances the spatial dimensions while decreasing the number of channels. The output layer of the model is designed to restore the spatial dims and to make a prediction for each pixel in the input image. In recent years, more complicated architectures of CNN like SegNet [21] and DeconvNet [25] have been successfully applied to segment interested objects from the rest of the images. SegNet and DeconvNet models follow similar architectures except with some slight variations. Both of these networks consist of the encoder and decoder kind of structure in which the encoder is a classification CNN without the top layer while the decoder consists of same layers in the reverse order usually referred to as the transposed convolution or upsampling followed by the convolutional layers. In the proposed approach below, we introduce a variation of the UNet [26] such that it fits best to the current application. Here, 2D images are used unlike in the existing approaches [27] that process 3D data. The proposed methodology is novel in the process of training along with the variations made to the UNet.

Several deep convolutional neural networks (CNNs) are introduced for semantic segmentation of salt bodies from seismic images. A squeeze-extraction feature pyramid networks (Se-FPNs) achieve high-quality segmentation effect [28]. The squeeze-extraction approach of the model learns to suppress irrelevant regions in seismic images while highlighting salient features useful for the task. The architecture inspired by the UNet model in combination with ResNet and DenseNet architectures is comparable and, in most cases, results in better results [29]. An improved encoder–decoder deep neural network model was introduced for salt domes and faults identification that alleviates the scarcity issue of labeled seismic data by following transfer learning approach [30]. The success and applicability of these DNN models are subject to the availability of labeled data, which incurs costs associated with manual annotations. Addressing this limited labeled data issues, augmentation approaches are widely used to generate synthetic data. A data augmentation method based on training two generative models was used to augment the number of samples in a seismic image dataset [31]. This method employs two generative models: a variational autoencoder to generate salt body mask patches, and the conditional normalizing flow model to receive the generated masks and generate the associated seismic image patches. Semi-supervised approaches utilize unlabeled data to address issues associated with limited data resources. A semi-supervised approach for segmentation of salt bodies in seismic images outperforms state-of-the-art on the TGS Salt Identification Challenge dataset and is ranked first among 3234 competing methods [8].

Methodology

Main aim of this work is to segment the salt deposit regions from the rest of the given seismic images. The architecture of the proposed model, training approach and the other technical details are discussed in this section. The proposed architecture is derived from the UNet architecture [26]. Our proposed UNet model is entirely different the original UNet architecture in terms of the number of blocks, number of filters at each layer, introduction of batch normalization [32] layers, and many others to make it suitable to the segmentation of the salt deposit regions from the seismic images.

Pre-processing

As a preprocessing technique, the training images of size 101×101 are loaded and resized to 128×128 to make them suitable for passing through the proposed UNET model. For image resizing, the nearest-neighbor interpolation method is used. The images in the dataset have pixel values ranging from 0 to 255, making the computation of these large numeric values more complex using deep neural network models. Images are normalized such that each pixel is rescaled to a range of 0 to 1 by dividing each pixel by 255 to reduce the computation complexity of the deep neural network models that are used for further segmentation.

Proposed Network Architecture

The proposed network is derived from the UNet architecture. UNet, being a popular approach to be used for segmentation tasks, applies classification on each and every pixel in the given input image and thereby produces a mask of the same size as input.

The proposed network contains four convolution blocks each in the contraction path and expansion paths. The proposed model uses batch normalization with every convolution layer.

The network shown in Fig. 4 takes a seismic image of size $(128 \times 128 \times 1)$ as input. It follows an encoder–decoder kind of structure, the encoder part encodes the given seismic image using different pooling techniques. Both the encoder and decoder modules of the UNET contain four blocks. The encoder extracts the information of "what" is present in the image and produces a representation of $128 \times 128 \times 64$. The decoder part of the architecture increases the size of encoded representation produced by the encoder and extracts the information of "where" in the given image. In the contracting path, each block contains two consecutive convolution layers followed by a max pooling layer.

The first block contains two convolution layers followed by a pooling and batch normalization layers which increases the input number of channels from 1 to 8. There are four

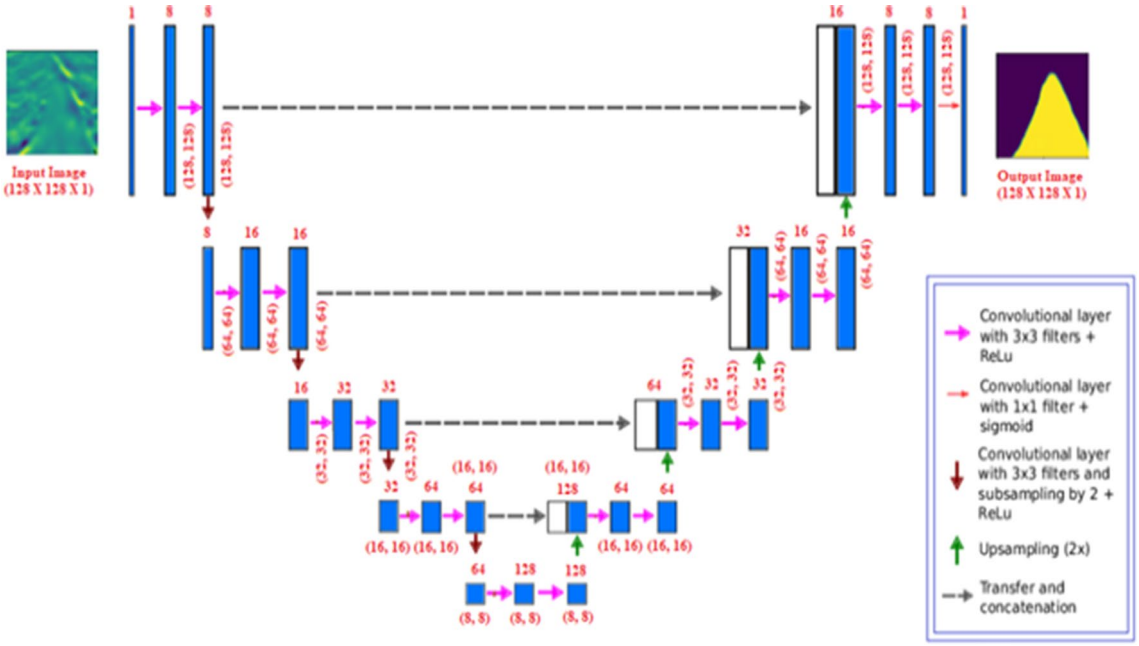

Fig. 4 Proposed UNET-based architecture for segmentation of the salt deposit regions from seismic images

such blocks in the entire contraction path and by the end of all the four blocks, the channel count gets increased to the 64 channels since the process of convolution increases the depth of the image. Each convolution layer filter size is fixed to 3×3. In Fig. 4, the arrow pointing down indicates a max pooling layer basically used to subsample the output of the convolution layer. The encoder module of the proposed model is designed with four convolution blocks at the end of the contraction path, the image size becomes $8 \times 8 \times 64$. The operation involved at every convolution layer can be represented mathematical as below. Consider a 3D sensor $X \in \mathbb{R}^{c \times w \times h}$ a 3D tensor, where c is the number of channels, w, h are the width and height, respectively. The convolution operation performed between layers l and $l+1$, denoted by $X^{l+1} = X^l * W^l$ and can be defined as follows:

$$X^{l+1} = \sum_{c' \in [1,c], i' \in [1,z], j' \in [1,z]} W^l_{k,c',i',j'} * X^l_{c',i+i'-1,j+j'-1} \quad (1)$$

$$\text{where} \quad k \in [1, c^{l+1}], i \in [1, w^{l+1}], j \in [1, h^{l+1}], \quad (2)$$

Here, $X^l in \mathbb{R}^{c^l \times w^l \times h^l}$ is the 3D tensor received at layer l; $W^l in \mathbb{R}^{c^{l+1} \times c^l \times z \times z}$ is a set of c^{l+1} filters, with each filter of shape $c^l \times z \times z$; $X^{l+1} in \mathbb{R}^{c^{l+1} \times w^{l+1} \times h^{l+1}}$ is the 3D tensor output after the convolution operation and will serve as input to the $l+1$ layer in the network. The spatial dimensions of the output tensor w^{l+1}, h^{l+1} are by default $w^l + z - 1$ and $h^l + z - 1$, respectively, but one can also pad a number of zeros at the borders of X^l to achieve spatial dimensions unchanged.

Batch normalization is used for avoiding the over-fitting problem in the model training Here onwards, the expansive path starts where the model uses upsampling techniques to increase the size of the representation produced by the contraction blocks.

Here, transposed convolution is used as an upsampling technique which expands the size of images. Here, a padding operation is performed on the initial image followed by a convolution operation. Just like in the contraction path, here also there are four such blocks and by the end of these blocks, the image of original size is obtained. For the final prediction, a 1D convolution layer with a 1×1 kernel using sigmoid activation function is applied on the output of the last block. The 1D convolution helps in reduction of the number of channels to the required network output, while sigmoid activation function maps every pixel in the output block to the range of 0 to 1. The obtained values will be rounded to 0 or 1 to get the final predicted mask of the same size as the input image. At the end of the training, the model weights are saved and are used during inference (Fig. 5).

Generation of Random Mini-batches from a Sampled Batch for Training

In the proposed methodology, we use random batch training [33] so that the trained model will not be biased toward the final set of samples from training data. In every iteration, we select a random batch of 400 samples from the available 3200 images and from that 400, again another batch of 20

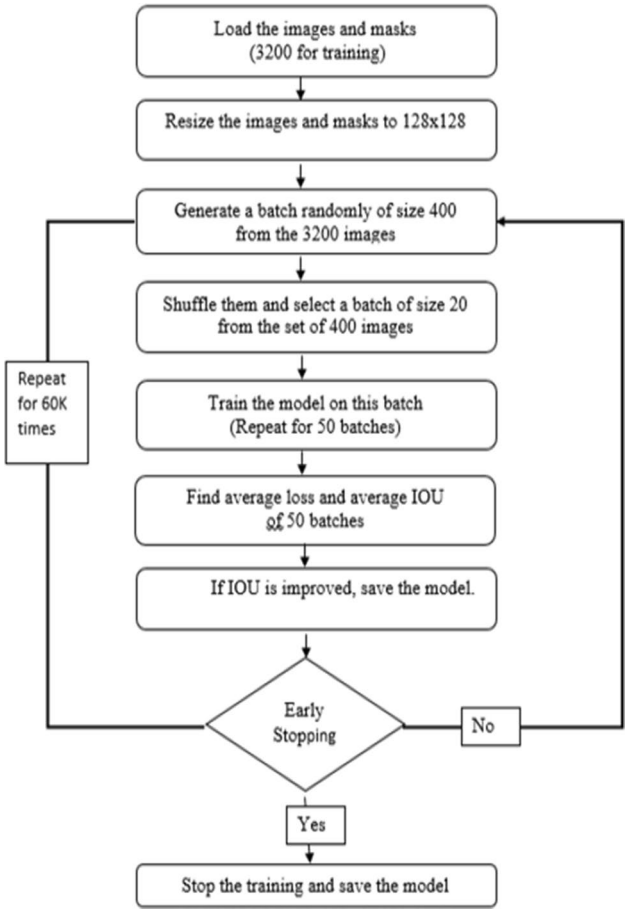

Fig. 5 Generation of Random mini-batches from a sampled training data

samples is taken at random and is passed to the model for training.

Inferences by the Model and Post-processing

During the inference stage, we apply post-processing on the outputs produced by the model. For each seismic image, after getting the prediction of the output mask, a sharpening technique is applied. This sharpening as a post-processing technique leads to a better visualization of the salt deposits present in the predicted mask thereby leading to a better IoU score. There are usually two filters known as low pass and high pass that are applied on the images to improve their visualization capacity. Low-pass filter application is referred to as smoothing, whereas the high-pass filter application is considered as the sharpening technique. A high-pass filter usually attenuates the low frequencies and allows high frequencies to pass through them. This makes the pixels of salt in the predicted mask to pass through the filter and produces better scores. The following high-pass filter is applied for

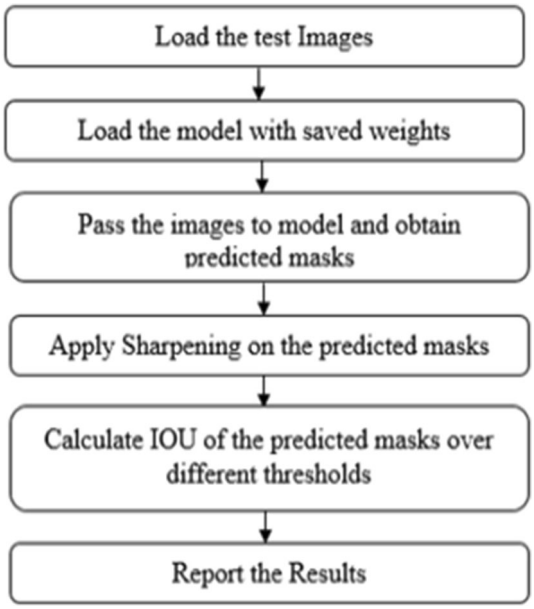

Fig. 6 Steps followed during the inference stage of the model

sharpening of the output masks as part of the post-processing of the output masks produced by the model.

$$\begin{bmatrix} -2 & -2 & -2 \\ -2 & 17 & -2 \\ -2 & -2 & -2 \end{bmatrix}$$

The details of each step followed during the inference stage and post-processing are outlined in Fig. 6

Experimental Studies

This section presents the experimental studies carried out to understand the performance of the proposed model for the identification of the regions with salt deposits in the given seismic images.

Details of the Seismic Image Dataset

The dataset used for the experimental studies is released by the world's leading geoscience data company, TGS-NOPEC Geophysical Company (TGS), and is publicly available in Kaggle website [17]. This dataset comprises 4000 seismic images of size 101×101 in which each pixel tells you whether the particular pixel area is salt or not. Out of the 4000 images, 80% of the images were used for the training purpose and the rest of the images (20%) were used for the testing purpose. From the train split, 10% of the images are reserved for validation split, based on which hyper parameters are tuned and finalized. Few sample images from the

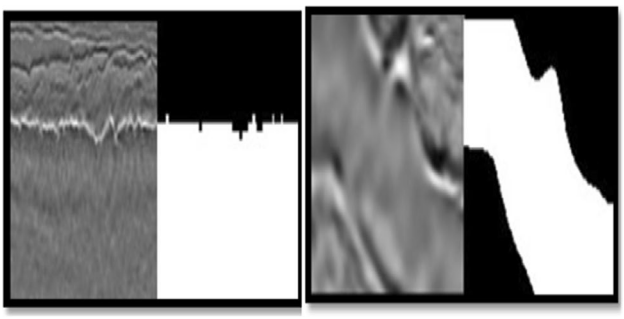

Fig. 7 Samples from TGS dataset showing the seismic images along with their corresponding masks for Salt Deposit Regions

Table 1 General confusion matrix

	$Y = 1$	$Y = 0$
$\hat{Y} = 1$	TP	FN
$\hat{Y} = 0$	FP	TN

dataset along with the corresponding masks are shown in Fig. 7.

Metrics Used

The performance of the model is monitored in the form of the metric Intersection of Union (IoU) [34], which was the most common metric used for the segmentation tasks. During the inference stage, an image is passed to the model and the model produces the classification value for every pixel of the input image. Now we calculate the IoU (Intersection over Union) value, a popular metric for segmentation tasks, by comparing both the predicted and original masks.

The interpretation of IoU of a particular image is the ratio between the intersection region and union regions of the ground truth mask and the prediction mask. Let Y be a 2D ground truth mask and $\hat{Y}$ be the 2D predicted mask and let the vectorized representations of Y and $\hat{Y}$ are $Y_1, Y_2, Y_3, ...Y_n$ and $\hat{Y}_1, \hat{Y}_2, \hat{Y}_3, ...\hat{Y}_n$, respectively. Let n represent the number of pixels and Y_i be the i^{th} pixel of the ground truth mask and $\hat{Y}_i$ be the i^{th} pixel of the prediction mask. As both ground truth and predicted masks are binary, $Y_i, \hat{Y}_i \in \{0, 1\}, \forall i \in [1, n]$. Mathematical computation of IoU can be represented as follows:

$$IOU(Y, \hat{Y}) = \frac{Y \cap \hat{Y}}{Y \cup \hat{Y}} = \frac{\Sigma_{i=1}^{n} min(Y_i, \hat{Y}_i)}{\Sigma_{i=1}^{n} max(Y_i, \hat{Y}_i)} \tag{3}$$

Equation 3 can be rewritten as a form of the confusion matrix between Y and $\hat{Y}$ as shown in table 1

In Table 1, TP, FP, TN and FN indicate true positives, false positives, true negatives and false negatives, respectively. IOU can be computed as:

$$IoU(Y, \hat{Y}) = \frac{TP}{TP + FN + FP} \tag{4}$$

Hyper Parameter Details of Model Training

While training the model the model is trained for 200 epochs using early stopping with a patience of 20 epochs. RelU activation function is used with each convolution layer of both the contraction and expansion modules and sigmoid activation is used with the final layer of the model. In both contraction and expansion paths, the same padding is used to make the outputs of the layers remain the same as those of input. The convolution layers of the encoder use kernels of size 3×3, whereas the pooling layer uses a window of size 2×2. Transpose convolution layers of the decoder use kernels of size 2×2 with row wise and column wise strides of 2,2. The learning rate is set to 0.01. Binary cross entropy is used as the loss function and the model is optimized using Adam optimizer. Data augmentation is applied to increase the training set which in turn avoids model over-fitting. Though we have experimented with different architectures, this proposed model proved to be more successful in terms of both complexity and performance.

Inference from the Model

In the testing phase, as shown in Fig. 6, we load the remaining 20% data with images and masks. The weights that are saved earlier are loaded into the model and the model is used for the prediction of the masks for the given images. As a post-processing technique, we have used sharpening on the predicted masks which has shown some improvement. Figure 8 illustrates the effect of applying sharpening operation on the output masks. Compared to the masks without sharpening those with sharpening looks more smoother and helps to improve the scores of the model.

Results and Discussion

In this experiment, we present the segmentation result of the proposed model with different thresholds. For every seismic image, IOU value is computed using Eq. (3). If the IoU value of an image is greater than a certain threshold then its IOU value contributes to the computation of the IOU value of the dataset.

$$IoU(D) = \sum_{i=1}^{m} \frac{\mathbb{1}(D_i) * IoU(D_i)}{m} \tag{5}$$

In equation 5, m is the number of images in the dataset, D_i be the i^{th} image and D is the dataset with m images. $\mathbb{1}(a)$ is the indicator function and is computed as follows:

$$\mathbb{1}(a) = \begin{cases} a & \geq \gamma \\ 0, & \text{otherwise} \end{cases}$$

where γ is the threshold applied on IoU value. IoU on a threshold, γ, indicates that a particular IoU value has crossed that threshold. For Example, a predicted output mask is considered to be valid over a threshold of 0.7 if the value of IoU is above 0.7 for that particular mask. Once the IoU of the dataset is computed, we vary the thresholds and compute the average IoU value for the dataset (Table 2). The average of IoUs over different thresholds varying from 0.7 to 0.95 with a step of 0.05 is computed and is reported in Table 3. The proposed model is compared with the SegNet model [21] as we use SegNet as the baseline method.

Table 3 shows the IoU scores obtained by different methods with various thresholds and the average IoU over all the thresholds from 0.7 to 0.95 with a step value of 0.05 along with the loss of training is reported. We can observe that the IoU scores obtained by the proposed model are much superior compared to the scores obtained by the baseline model. The transpose convolution applied in the expansion path makes the segmentation better.

Effect of Post-processing and Augmentation

As a post-processing step, we apply high-pass filters to the masks produced by the models. This experiments are performed to understand the effect of the post-processing.

The Input image along with the true mask and the predicted mask are plotted here for various images before and after the sharpening. The images in Fig. 8 show the effect of sharpening as a post-processing technique on the predicted masks.

From Fig. 8, we can observe that the proposed model with sharpening as a post-processing technique produces good improvement in the average IoU value in comparison with the other methodologies.

Table 2 Comparison of the IoU scores produced by the proposed UNet architecture with SegNet

Parameters		SegNet	UNet (Proposed)
Loss		0.084	0.006
Threshold (γ)	0.70	78.10	85.64
	0.75	77.30	85.56
	0.80	76.35	85.46
	0.85	75.08	82.00
	0.90	73.24	85.21
	0.95	70.77	84.98
Average IoU		77.08	85.60

Table 3 Comparison of the IoU scores produced by the proposed UNet architecture with and without augmentation

Parameters		No augmentation	Augmentation
Loss		0.088	0.006
Threshold	0.70	84.80	85.64
	0.75	83.60	85.56
	0.80	82.17	85.46
	0.85	80.36	82.00
	0.90	77.70	85.21
	0.95	73.05	84.98
Average IoU		82.95	85.60

Comparison Study

This subsection compares the proposed method with the recent literature to demonstrate the effectiveness. Table 4 presents the scores obtained by the proposed model in comparison with the scores obtained by recent models developed for salt segmentation tasks on TGS dataset.

From Table 4, it is clear that the proposed model outperforms various existing models in the literature. The model demonstrates superior performance over several existing models in terms of mean IoU scores.

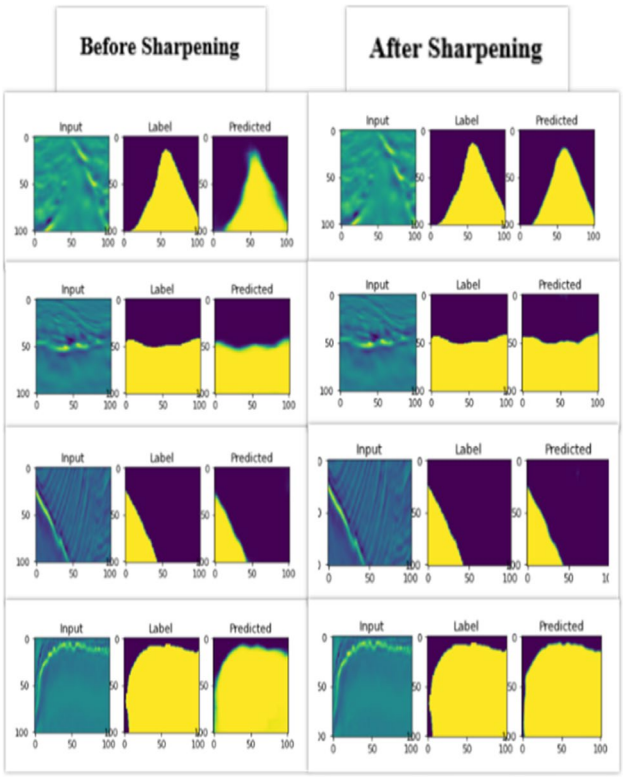

Fig. 8 Comparison of predicted masks with and without applying sharpening for post-processing

Table 4 Performance comparison of the proposed improved UNet with literature considering TGS dataset

Model reference	Mean IoU
FCN [28]	75.79
DeepLab [28]	76.58
Unet [28]	77.06
ResUnet [28]	80.74
UNet [29]	82.44
LinkNet [29]	82.59
UNet + ResNet + DenseNEt (No dropout) [29]	83.10
PSPNet [29]	83.17
UNet + ResNet + DenseNEt [29]	83.18
FPN [29]	83.62
SeResNet152-FPN [28]	84.57
Improved UNet (proposed)	85.60

Conclusion

In this work, we propose a machine learning model to segment salt deposit regions from the seismic images. A variation of the UNet architecture is proposed for the semantic segmentation. The proposed model is validated on the benchmark TGS dataset that is publicly available online. Our experimental results show that the proposed model outperforms the baseline. In addition we apply a high-pass filter on the output masks to make the masks more sharpened. Augmentation helps the model to generalize well by avoiding over-fitting and allows the model to achieve an average IoU of 85.60.

These neural network models are typically designed with deeper layers and offer superior performance compared to traditional approaches; however, they are data hungry and are not appropriate for low resource data scenarios.

Funding No funding.

Declarations

Conflict of interest The authors declare that there is no conflict of interest.

References

1. A. Asjad, D. Mohamed, A new approach for salt dome detection using a 3D multidirectional edge detector. Appl. Geophys. **12**(3), 334–342 (2015)
2. X. Wu et al., FaultSeg3D: using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation. Geophysics **84**(3), IM35–IM45 (2019)
3. T. Maćkowski, A. Sowiżdżał, and A. Wachowicz-Pyzik. Seismic methods in geothermal water resource exploration: case study from Łóodź trough, central part of Poland. Geofluids 2019 (2019)
4. M.A. Shafiq et al., Detection of salt-dome boundary surfaces in migrated seismic volumes using gradient of textures. in *SEG Technical Program Expanded Abstracts 2015. Society of Exploration Geophysicists*, 2015, pp. 1811–1815
5. H. Di, Z. Wang, and G. AlRegib, Deep convolutional neural networks for seismic salt-body delineation. in c*AAPG Annual Convention and Exhibition* (2018)
6. earthsky.org. Using seismic technologies in oil and gas exploration. Accessed 15 March 2020, http://www.chevron.com/stories/seismic-imaging (2020)
7. setterfield.org. Salt Dome Analysis. Accessed 20 March 2020. http://www.setterfield.org/salt deposits/salt dome analysis text.html (2020)
8. Y. Babakhin, A. Sanakoyeu, H. Kitamura, Semi-supervised segmentation of salt bodies in seismic images using an ensemble of convolutional neural networks. in *German Conference on Pattern Recognition*. Springer, pp. 218–231 (2019)
9. A. Halpert, R.G. Clapp, Salt body segmentation with dip and frequency attributes. in *SEPReport* vol. 136, pp. 113–124 (2008)
10. D. Koroteev, Z. Tekic, Artificial intelligence in oil and gas upstream: trends, challenges, and scenarios for the future. Energy AI **3**, 100041 (2021)
11. J.D. Bodapati, Stacked convolutional autoencoder representations with spatial attention for efficient diabetic retinopathy diagnosis, in *Multimedia Tools and Applications*, pp. 1–24 (2022)
12. J.D. Bodapati, V.N. Rohith, ChxCapsNet: deep capsule network with transfer learning for evaluating pneumonia in paediatric chest radiographs. Measurement **188**, 110491 (2022)
13. J.D. Bodapati, V.N. Rohith, V. Dondeti, Ensemble of deep capsule neural networks: an application to pediatric pneumonia prediction. Phys. Eng. Sci. Med. **45**(3), 949–959 (2022)
14. J.D. Bodapati, Modified self-training based statistical models for image classification and speaker identification. Int. J. Speech Technol. **24**(4), 1007–1015 (2021)
15. J.D. Bodapati, SAE-PD-Seq: sequence autoencoder-based pretraining of decoder for sequence learning tasks. Signal Image Video Process. **15**(7), 1453–1459 (2021)
16. J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440 (2015)
17. Kaggle.com. TGS Salt Identification Challenge, Segment Salt Deposits Beneath the Earth's Surface. Accessed on January 2020, http://www.kaggle.com/c/tgssalt- identification-challenge (2020)
18. D. Lu, Q. Weng, A survey of image classification methods and techniques for improving classification performance. Int. J. Remote Sens. **28**(5), 823–870 (2007)
19. A.R. Zamir, M. Shah, Accurate image localization based on google maps street view, in *European Conference on Computer Vision*. Springer, pp. 255–268 (2010)
20. Z.-Q. Zhao et al., Object detection with deep learning: a review. IEEE Trans. Neural Netw. Learn. Syst. **30**(11), 3212–3232 (2019)
21. V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **39**(12), 2481–2495 (2017)
22. M.A. Shafiq et al., Salsi: a new seismic attribute for salt dome detection. in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE*, pp. 1876–1880 (2016)
23. H. Di, M. Shafiq, G. Al- Regib, Multi-attribute k-means clustering for salt-boundary delineation from three-dimensional seismic data. Geophys. J. Int. **215**(3), 1999–2007 (2018)

24. X. Wu, Methods to compute salt likelihoods and extract salt boundaries from 3D seismic images. Geophysics **81**(6), IM119–IM126 (2016)
25. H. Noh, S. Hong, and B. Han, Learning deconvolution network for semantic segmentation. in *Proceedings of the IEEE International Conference on Computer Vision*. pp. 1520–1528 (2015)
26. O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation. in *International Conference on Medical Image Computing and Computer-assisted Intervention*. Springer, pp. 234–241 (2015)
27. A.U. Waldeland et al., Convolutional neural networks for automated seismic interpretation. Leading Edge **37**(7), 529–537 (2018)
28. B. Liu et al., Image segmentation of salt deposits using deep convolutional neural network, in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). IEEE*, pp. 3304–3309 (2019)
29. A. Milosavljević, Identification of salt deposits on seismic images using deep learning method for semantic segmentation. ISPRS Int. J. Geo-Information **9**(1), 24 (2020)
30. M. Alfarhan, M. Deriche, A. Maalej, Robust concurrent detection of salt domes and faults in seismic surveys using an improved UNet architecture. in *IEEE Access* (2020)
31. L.F. Henriques et al. Generating data augmentation samples for semantic segmentation of salt bodies in a synthetic seismic image dataset. arXiv preprint arXiv:2106.08269 (2021)
32. S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167 (2015)
33. N. Bjorck et al. Understanding batch normalization. in *Advances in Neural Information Processing Systems*, pp. 7694–7705 (2018)
34. H. Rezatofighi et al., Generalized intersection over union: a metric and a loss for bounding box regression. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 658–666 (2019)