Instance weighting-based Knowledge Transfer Network for Seismic Fault Detection

Tiash Ghosh[1*] **Mohammed Fayiz Parappan**[2*] **Mamata Jenamani**[2] **Aurobinda Routray**[1]
[1]Department of Electrical Engineering [2]Department of Industrial and Systems Engineering
IIT Kharagpur
tiashghosh96@gmail.com, fayiz@kgpian.iitkgp.ac.in,
mj@iem.iitkgp.ac.in, aurobinda.routray@gmail.com

Abstract

Geological Fault Detection is a crucial aspect of earthquake prediction and oil exploration. With the advancements in deep learning, the challenging task of accurate fault detection has gained popularity. While the traditional deep learning methods struggle due to the labeling process, training a model solely on synthetic data may not yield satisfactory results due to the disparities between synthetic and real seismic data. To mitigate the impact of these differences, we propose employing an instance weighting-based transfer learning. This allows the model to adapt to only the unique characteristics of the geological data. The proposed method yields satisfying results on the Indian Krishna Godavari Basin dataset.

1 Introduction

Accurate delineation of faults in seismic images serve as a fundamental step in interpretation of subsurface structures, reservoir characterization, and earthquake detection. Traditional methods include discontinuity-based methods like [2, 32, 20, 23, 24, 8, 33]. Over the years, application of machine learning (ML) [5, 4, 29, 35, 1] has significantly revolutionized seismic interpretation.
Subsequently, the task of fault detection transitioned into a segmentation problem, leading to the proposal of encoder-decoder architecture in studies. The study presented in [34] builds upon previous research [12, 35, 17], and others that employed neural network architectures for pixel-level classification. Additionally, other variations such as nested residual U-nets [9], U-net++ [36], and wavelet transform-based CNN [26] architectures, were also introduced. Inspired by the success of attention-based or Transformer networks, a few reports [10, 31] have also been found. Although deep learning techniques have shown promising results, they heavily rely on a huge amount of labeled training data. Transfer learning (TL) is a potential solution to address the above challenge. TL methods in [3] utilize the acquired knowledge to adapt to a similar task [21].
In this study, we aim to overcome the limitations of traditional deep learning methods and the disparities between synthetic and real seismic data, by leveraging transfer learning. Using a base CNN model trained on synthetic seismic data [6], we adopt an appropriate transfer learning strategy to the offshore Indian Krishna Godavari (KG) Basin dataset. Due to the unavailability of data, only a limited amount of literature [25, 27] exists for the Indian KG Basin dataset. To the best of our knowledge, an instance-based domain knowledge transfer has not yet been explored on this dataset. The major contributions of the paper as follows:

1. A quantitative analysis of data distribution differences using suitable measurement parameters, i.e Kullback–Leibler divergence (KLD) and Jensen–Shannon divergences (JSD).
2. Application of an instance weighting-based method for knowledge transfer between source and target domains.

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

3. In the context of geological fault detection, some well-established segmentation metrics may not be suitable for quantitative analysis of the network performance. Hence, a detailed metric-based similarity assessment of ground truth and predicted labels is presented.

2 Methodology

While traditional methods rely on geological attribute calculations, this research adopts a well established computer vision approach of semantic segmentation, where the objective is to assign a class label to each pixel in the image. Let us assume, that the source domain or synthetic data is denoted by $\mathcal{D}_S = (x_{s1}, y_{s1}), (x_{s2}, y_{s2}) \cdots (x_{sn}, y_{sn})$, where x_{si} and y_{si}, denotes the i^{th} instance and the label, respectively. $\mathcal{D}_s$ is employed for training the pre-trained model. Similarly, the target domain data is represented as $\mathcal{D}_T = (x_{t1}, y_{t1}), (x_{t2}, y_{t2}) \cdots (x_{tm}, y_{tm})$. Table 1 shows quantitative evidences confirming that the data from both domains follow different marginal distributions i.e.

$$P_S(x) \neq P_T(x) \tag{1}$$

where, $P_S(x)$ and $P_T(x)$ are marginal probability distributions of source and target domain, respectively. Hence, this research adopts a strategy to fine-tune the network based on the dissimilarity between source and target domains. In order to accomplish the fault detection task, the proposed

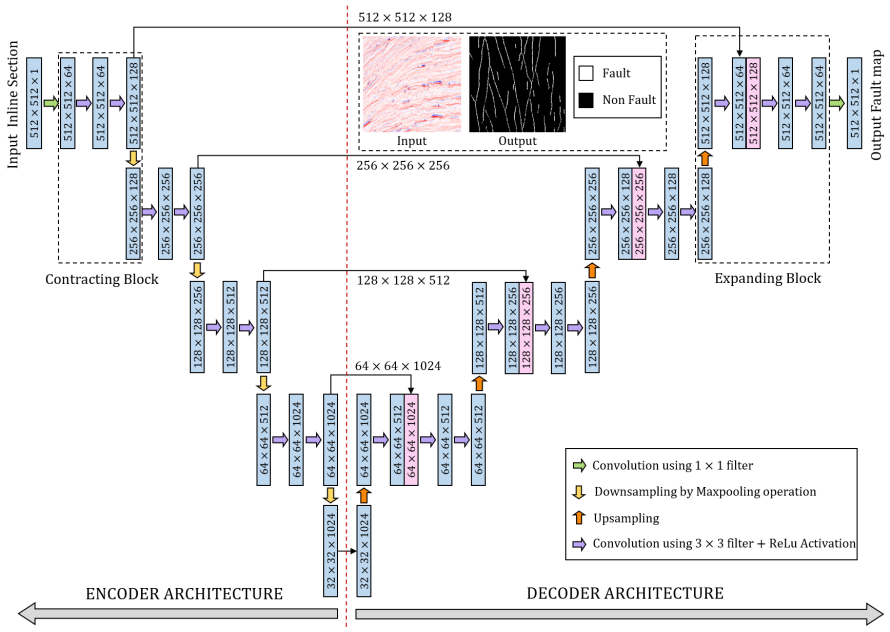

Figure 1: The Encoder-Decoder CNN architecture for 2-D geological fault detection. Note that the black values denote non-fault pixels while the white ones indicate faulty ones.

framework employs a U-net architecture (as depicted in Figure 1) as the encoder-decoder model. In the context of geological fault detection, there is a significant class imbalance between positive labels (faults) and negative labels (no-faults). To tackle this task, we employ three different loss fuctions, namely, the Binary Cross Entropy (BCE) [19], Weighted BCE [22] and the Binary Focal (BFL) [18] loss functions. It has been observed that different components of a CNN learn distinct features [3]. A combination of pre-training and fine-tuning is a prominent trend [30] in deep learning. However, the performance of TL methods depend on the dissimilarity of source and domain datasets. Inspired from [16], we quantify the distribution difference between the target and source domain instances and use it for instance weighting (IW) during model training. We introduce the KL and JS divergences to measure the similarity quantitatively [15, 11, 7].

$$JS(P_1, P_2) = \frac{1}{2} KL\left(P_1 || P\right) + \frac{1}{2} KL\left(P_2 || P\right) \quad \text{and} \quad KL(P_1, P_2) = \sum P_1(x) \log \frac{P_1(x)}{P_2(x)} \tag{2}$$

Also note that P in Equation 2 can be expressed as,

$$P = \frac{P_1 + P_2}{2} \tag{3}$$

The JS divergence between the i^{th} target domain instance and entire source domain instance is:

$$JS(x_{ti}, X_S) = \frac{1}{n} \sum_{j=1}^{n} JS(x_{ti}, x_{sj}) \tag{4}$$

In this research, we fine tuned the network with initial weights equal to the pre-trained model weights instead of random weight initialization. The instances used for final model training are weighted using the normalized value JS divergence values.

$$w_i = JS_{normalized}(x_{ti}, X_S) \tag{5}$$

where, w_i is the weight associated with the i^{th} target domain instance. Thus, the target instances that are similar to the source domain are asigned low training weights, so as their influence is reduced. Therefore, the model focuses on adapting to only the unique characteristics of the real data.

3 Description and Analysis of Datasets

In this study, the Equinor synthetic model [6] is utilized as the synthetic data source. The 3-D seismic cube consists of 151 inline sections and 589 crosslines. The data acquisition process involved recording for a duration of 3.6s with a sampling rate of 4ms. Further, the research focuses on a specific area in the offshore Krishna-Godavari (KG) Basin. The seismic data consists of inlines ranging from 1982 to 4517, with a step size of 1, and crosslines ranging from 8800 to 14000, with a step size of 2. During acquisition, the recorded data of 4s was sampled with a vertical sampling rate of 4ms. We quantitatively evaluate the similiarity between the seismic data distributions in order to illustrate the requirement of transfer learning. Table 1 provides a clear distinction between synthetic (S_D) and real data (R_D), revealing that the datasets exhibit dissimilar data distributions. Note that, in Table 1, we consider consecutive inline sections for calculation of KLD and JSD values.

Table 1: Data distribution of different seismic sections

Seismic data type	Similarity parameters	
	KLD	JSD
S_D 90 and S_D 91	10.886	1.234
R_D 2072 and R_D 2073	243.852	113.4
R_D 2073 and S_D 91	1231.154	270.066

To further study the dissimilarity between the distributions, we present an Analysis of Variance (ANOVA) test to establish the divergence in the data distributions. We intially assumes that D_S and D_T come from the same distribution [7]. The sample size for all three populations in the ANOVA test was chosen to be 60. Table 2 highlights a lower p value, indicating that the hypothesis of synthetic and real data belonging to the same distribution is not true. The significant disparity between these data distributions, serves as a strong motivation for the adoption of TL strategies.

Table 2: Statistical parameters obtained from ANOVA test

Test	KL Divergence	JS Divergence
F-statistic	12.606	105.209
p-value	7.614×10^{-6}	7.794×10^{-31}

4 Results and Discussions

In this section, we evaluate the performance of the proposed IW framework on the KG Basin dataset, which reveal the superior performance of the method as compared to state-of-the-art (SOTA) algorithm, namely, Faultseg3D [34]. The synthetic and real data patches exhibit differences in the amplitude patterns between the two datasets. In order to attain maximum effectiveness, we extract patches with maximum similarity between the real dataset and the base model training data. In [13], the authors highlight that an optimal ratio for adopting transfer learning is $2:1$. In our experiments we adhere to these findings by initially train the model with 151 inline sections of the synthteic data [6]. Further, we randomly select a few inline sections of the KG Basin dataset and extracted 85 patches of dimension 512×512 for knowledge transfer. To train the network, we utilized the Adam optimizer [14], using a learning rate of $2e-4$. The model was trained for 50 epochs.
Figure 2 presents the qualitative results of the proposed algorithm with the comparison method. Note

that, the ground truth fault maps are generated through expert's annotation. The visual comparison confirms the effectiveness in accurately delineating faults as compared to the results obtained from [34]. The comparison method exhibits a considerable number of spurious faults, which are clearly not present in our segmentation results. It can be observed that BCE loss demonstrates notable performance even in presence of high class imbalance. The Weighted BCE and BFL loss functions exhibit enhanced continuity and thickness of fault lines.

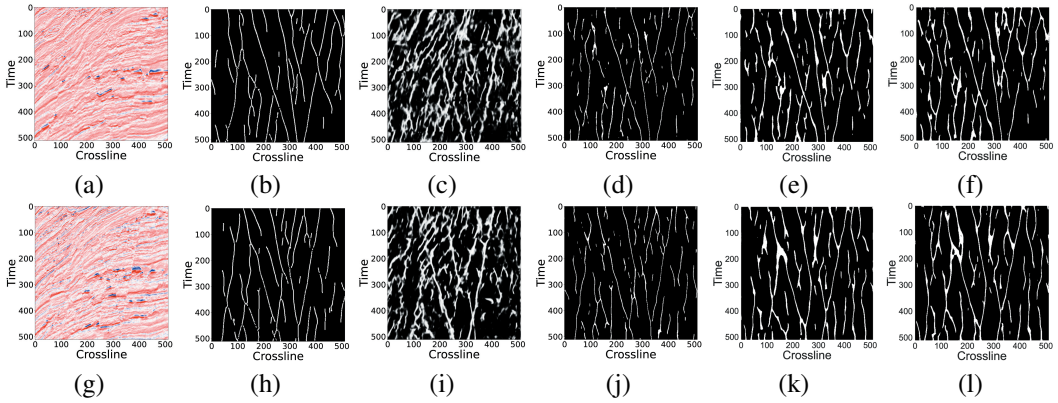

Figure 2: Illustration of the qualitative results using the proposed framework. The first column showcases the seismic patches, while the subsequent column depicts the ground-truth labels. (c), and (i) showcase the performance of [34] in inline sections 1983, and 2013, respectively. The prediction results of the proposed method are illustrated in (d)-(f), and (j)-(l). The 4^{th}, 5^{th}, and 6^{th} columns depict the performance achieved using BCE, weighted BCE, BFL loss functions, respectively.

Table 3: Quantitative analysis of the fault detection methods on KG Basin Dataset

Method	**F1-score**	**Precision**	**Recall**	**Mean FDA**	**Mean ROC-AUC**	**Mean SSIM**	**Mean FSMI**
FaultSeg3D [34]	0.031	0.016	0.352	0.352	0.522	0.381	0.200
Our Method w. BCE	0.149	0.082	*0.811*	**0.811**	**0.856**	**0.816**	*0.453*
Our Method w. Weighted BCE	*0.169*	*0.096*	0.714	0.714	0.807	0.714	0.423
Our Method w. BFL	**0.185**	**0.105**	0.756	*0.756*	*0.830*	0.718	0.448
Our Method w. IW strategy	0.119	0.064	**0.842**	0.644	0.780	*0.731*	**0.500**

In order to assess the performance of the proposed method, we incorporate relevant metrics from the existing literature like Precision, Recall, F1-score, Fault similarity index (FSMI) [37], Fault detection accuracy (FDA) index [17], Structural similarity (SSIM) index [28], Area under the Receiver Operating Characteristic curve (ROC-AUC) to comprehensively evaluate the performance. Moreover, the quantitative performance is illustrated in Table 3. Note that, the mean value of the metrics over all inline sections are reported. The highest value for each metric is marked in bold face, while the second best ones are marked in italics. Notably, we observe that our proposed algorithm achieves higher values, indicating its superior performance. Furthermore, it is important to note that in the context of geological fault detection, some well-established segmentation metrics may not be applicable. In this research, we identify that Recall, FDA, ROC-AUC score, SSIM, and FSMI serve as effective metrics for evaluating the fault detection performance. The optimal result achieved from the variation of loss function is considered for investigating the efficiency of the proposed IW strategy. The quantitative results, both with and without weighting, are documented in Table 3.

5 Conclusions

To overcome the limitations due to limited data availability, we propose the use of TL for geological fault detection. By leveraging a TL framework, the model is allowed to capture general fault patterns of seismic data. The model is fine-tuned using a smaller amount of labeled real seismic data, to adapt to the actual geological conditions and improve fault detection performance in practical applications. Despite class imbalance challenge in the Indian KG Basin dataset, the method demonstrates satisfying results, showcasing its potential for effective geological fault detection. Application of TL proves to be a valuable strategy in mitigating data limitations and enhancing fault detection accuracy.

References

[1] Mauricio Araya-Polo, Taylor Dahlke, Charlie Frogner, Chiyuan Zhang, Tomaso Poggio, and Detlef Hohl. Automated fault detection without seismic processing. *The Leading Edge*, 36(3):208–214, 2017.

[2] Mike Bahorich and Steve Farmer. 3-d seismic discontinuity for faults and stratigraphic features: The coherence cube. *The leading edge*, 14(10):1053–1058, 1995.

[3] Augusto Cunha, Axelle Pochet, Hélio Lopes, and Marcelo Gattass. Seismic fault detection in real data using transfer learning from a convolutional neural network pretrained with synthetic seismic data. *Computers & Geosciences*, 135:104344, 2020.

[4] Haibin Di, Muhammad Shafiq, and Ghassan AlRegib. Patch-level mlp classification for improved fault detection. In *SEG Technical Program Expanded Abstracts 2018*, pages 2211–2215. Society of Exploration Geophysicists, 2018.

[5] Haibin Di, Muhammad Amir Shafiq, and Ghassan AlRegib. Seismic-fault detection based on multiattribute support vector machine analysis. In *2017 SEG International Exposition and Annual Meeting*. OnePetro, 2017.

[6] Equinor Synthetic Model. Equinor synthetic model.

[7] Basura Fernando, Amaury Habrard, Marc Sebban, and Tinne Tuytelaars. Unsupervised visual domain adaptation using subspace alignment. In *Proceedings of the IEEE international conference on computer vision*, pages 2960–2967, 2013.

[8] Dengliang Gao. Integrating 3d seismic curvature and curvature gradient attributes for fracture characterization: Methodologies and interpretational implications. *Geophysics*, 78(2):O21–O31, 2013.

[9] Kai Gao, Lianjie Huang, and Yingcai Zheng. Fault detection on seismic structural images using a nested residual u-net. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2021.

[10] Kai Gao, Lianjie Huang, Yingcai Zheng, Rongrong Lin, Hao Hu, and Trenton Cladohous. Automatic fault detection on seismic images using a multiscale attention convolutional neural network. *Geophysics*, 87(1):N13–N29, 2021.

[11] Dimitris Glotsos, Spiros Kostopoulos, Panagiota Ravazoula, and Dionisis Cavouras. Image quilting and wavelet fusion for creation of synthetic microscopy nuclei images. *Computer Methods and Programs in Biomedicine*, 162:177–186, 2018.

[12] Bowen Guo, Lu Li, and Yi Luo. A new method for automatic seismic fault detection using convolutional neural network. In *2018 SEG International Exposition and Annual Meeting*. OnePetro, 2018.

[13] Philip Kenneweg, Dominik Stallmann, and Barbara Hammer. Novel transfer learning schemes based on siamese networks and synthetic data. *Neural Computing and Applications*, 35(11):8423–8436, 2023.

[14] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[15] Roland Kwitt and Andreas Uhl. Image similarity measurement by kullback-leibler divergences between complex wavelet subband statistics for texture retrieval. In *IEEE International Conference on Image Processing, ICIP*, pages 933–936, 2008.

[16] Kihoon Lee, Soonyoung Han, Van Huan Pham, Seungyon Cho, Hae-Jin Choi, Jiwoong Lee, Inwoong Noh, and Sang Won Lee. Multi-objective instance weighting-based deep transfer learning network for intelligent fault diagnosis. *Applied Sciences*, 11(5), 2021.

[17] Shengrong Li, Changchun Yang, Hui Sun, and Hao Zhang. Seismic fault detection using an encoder–decoder convolutional neural network with a small training set. *Journal of Geophysics and Engineering*, 16(1):175–189, 2019.

[18] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection, 2018.

[19] Yide Ma, Liu Qing, and Qian Zhi-bai. Automated image segmentation using improved pcnn model based on cross-entropy. pages 743 – 746, 11 2004.

[20] Kurt J Marfurt, R Lynn Kirlin, Steven L Farmer, and Michael S Bahorich. 3-d seismic attributes using a semblance-based coherency algorithm. *Geophysics*, 63(4):1150–1165, 1998.

[21] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2010.

[22] Vasyl Pihur, Susmita Datta, and Somnath Datta. Weighted rank aggregation of cluster validation measures: a Monte Carlo cross-entropy approach. *Bioinformatics*, 23(13):1607–1615, 2007.

[23] Trygve Randen, Stein Inge Pedersen, and Lars Sønneland. Automatic extraction of fault surfaces from three-dimensional seismic data. In *2001 SEG Annual Meeting*. OnePetro, 2001.

[24] Andy Roberts. Curvature attributes and their application to 3 d interpreted horizons. *First break*, 19(2):85–100, 2001.

[25] Ratul Kishore Saha, Tiash Ghosh, Sanjai Kumar Singh, Mamata Jenamani, Aurobinda Routray, and Arpita Mondal. Fast and parallel semblance algorithm for detecting faults in large seismic volumes. In *IECON 2022 – 48th Annual Conference of the IEEE Industrial Electronics Society*, pages 1–6, 2022.

[26] Shian Shen, Haishan Li, Wenchao Chen, Xiaokai Wang, and Binke Huang. Seismic fault interpretation using 3-d scattering wavelet transform cnn. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2022.

[27] Gagandeep Singh, Rahul Mahadik, William K. Mohanty, Aurobinda Routray, Deepan Datta, and Sanket Smarak Panda. Seismic multi-attribute approach using visual saliency for subtle fault visualization. *Exploration Geophysics*, 54(4):387–394, 2023.

[28] Syed Zakwan Syed Zaini, Nur Najihah Sofia, Mohd Marzuki, Mohd Firdaus Abdullah, Khairul Azman Ahmad, Iza Sazanita Isa, and Siti Noraini Sulaiman. Image quality assessment for image segmentation algorithms: Qualitative and quantitative analyses. In *2019 9th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, pages 66–71, 2019.

[29] Kristofer M Tingdahl and Matthijs De Rooij. Semi-automatic detection of faults in 3d seismic data. *Geophysical prospecting*, 53(4):533–542, 2005.

[30] Grega Vrbančič and Vili Podgorelec. Transfer learning with adaptive fine-tuning. *IEEE Access*, 8:196197–196211, 2020.

[31] Zhiwei Wang, Jiachun You, Wei Liu, and Xingjian Wang. Transformer assisted dual u-net for seismic fault detection. *Frontiers in Earth Science*, 11:1047626, 2023.

[32] Xinming Wu. Directional structure-tensor-based coherence to detect seismic faults and channels. *Geophysics*, 82(2):A–13, 2017.

[33] Xinming Wu and Dave Hale. 3d seismic image processing for faults. *Geophysics*, 81(2):IM1–IM11, 2016.

[34] Xinming Wu, Luming Liang, Yunzhi Shi, and Sergey Fomel. 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.

[35] Wei Xiong, Xu Ji, Yue Ma, Yuxiang Wang, Nasher M AlBinHassan, Mustafa N Ali, and Yi Luo. Seismic fault detection with convolutional neural network. *Geophysics*, 83(5):O97–O103, 2018.

[36] Dun Yang, Yufei Cai, Guangmin Hu, Xingmiao Yao, and Wen Zou. Seismic fault detection based on 3d unet++ model. In *SEG International Exposition and Annual Meeting*. OnePetro, 2020.

[37] Ruoshui Zhou, Xingmiao Yao, Yaojun Wang, Guangmin Hu, and Fucai Yu. Seismic fault detection with progressive transfer learning. *Acta Geophysica*, 69:2187–2203, 2021.