Exploring Self-Supervised GPR Representation Learning for Building Rooftop Diagnostics

Kevin Lee[1], Wei-Heng Lin[1], Bilal Sher[2,1,*], Talha Javed[2,1,*], Sruti Madhusudhan[2,1,*], Chen Feng[1,2]

[1]Tandon School of Engineering, New York University, USA
[2]Building Diagnostic Robotics, Inc., USA
[*]Indicates equal contributions

k.lee@nyu.edu, whl318@nyu.edu, bas9876@nyu.edu, tj2186@nyu.edu, sm10183@nyu.edu, cfeng@nyu.edu

Abstract -

Ground Penetrating Radar (GPR), known for its applications in diverse domains, demonstrates potential for non-destructive diagnostic assessments on building rooftops. This study delves into the unique characteristics and data structure of GPR, investigating the novel approach of processing GPR as a "contextual neighborhood" of A-scans within their respective B-scans as opposed to the typical pixel-based approach. Given the challenge of obtaining a large corpus of annotated rooftop GPR data, we employ self-supervised deep learning methods for GPR representation learning. Experiments include training a vanilla Autoencoder, Variational Autoencoder, and a Transformer-based Autoencoder on GPR A-scans. Additionally, we extend our analysis by fine-tuning a pre-trained Masked Autoencoder on image based GPR B-Scans to investigate the differences between the conventional pixel-based approach and our proposed A-scan-based approach. Through a meticulous analysis of the learned latent spaces across these methods, we assess the viability of self-supervised deep learning in encoding meaningful GPR representations for downstream tasks. This research contributes to the exploration of GPR's applicability in building rooftop diagnostics and underscores the potential of self-supervised deep learning for efficient representation learning in the absence of annotated data.

Keywords -

GPR; Rooftop; Self-supervised; Transformer; Autoencoder

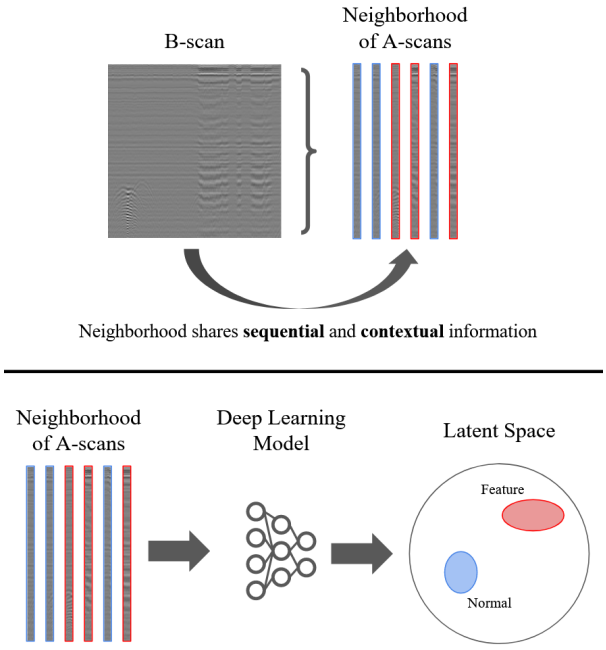

Figure 1. Contextual neighborhood of A-scans within the same B-scan. A-scans from the same B-scan share contextual information that may aid in the identification of features of interest. The blue scans indicate typical GPR readings that may not be pertinent for analysis, and the red scans indicate scans that contain features of interest.

1 Introduction

Ensuring the structural integrity of building rooftops necessitates reliable assessment methods. Ground Penetrating Radar (GPR) emerges as a promising, cost-effective alternative to traditional techniques like infrared thermography and nuclear moisture gauges [1]. Utilizing electromagnetic waves, GPR captures changes in electromagnetic properties within the scanned medium of the propagating waves.

GPR data is typically presented in the form of A-scans and B-scans. An A-scan, a single-point, 1D scan collected at each sampling interval, records amplitude variations of reflected waves over travel time. Despite its 1D nature, an A-scan encapsulates information beyond its precise location due to the non-linear propagation of electromagnetic waves. This distinction can be observed in the appearance of point objects like pipes as hyperbolas in GPR B-scans.

In contrast, B-scans are a series of sequentially mea-

sured A-scans often treated as images, offering a human-interpretable view of subsurface features. In rooftop diagnostics, many applicable downstream tasks appreciate higher levels of granularity in analyses. As B-scans may span many meters in length compared to the centimeter-scale offered by A-scans, we investigate approaches that study GPR at the A-scan level. This allows us the opportunity to explore the potential advantages in studying GPR scans through leveraging their inherent data structure – as opposed to the conventional method of processing GPR B-scans as images. Figure 1 illustrates the concept of contextual neighborhood membership of A-scans within the same B-scan. Deep learning models may be able to extract richer semantics in GPR representations through the exploit of intrinsic contextual information shared between A-scans of the same B-scan, especially considering the capability of A-scans to capture information adjacent to the immediate position of the GPR sensor.

Despite GPR's capabilities in producing high-resolution radargrams, interpreting scans remains challenging even for skilled technicians [2]. This challenge underscores the motivation to integrate deep learning for automated interpretation. While prior applications of deep learning in GPR interpretation exist, such as for land mine detection and utility identification [3, 4, 5], they primarily relied on supervised training with annotated labels. Given the scarcity of such labeled data for building rooftops, our investigation centers on self-supervised methods.

Our contributions can be summarized as follows:

1. Exploration of Self-Supervised Deep Learning Models: We assess the efficacy of self-supervised deep learning models in extracting meaningful features within the latent space of GPR data.
2. Analysis of Inherent Data Structure: We investigate the advantages of addressing GPR data through its native data structure (A-scan, B-scan) as opposed to treating it as a conventional image.

These contributions collectively aim to advance the understanding of GPR applications in building diagnostics, particularly in the context of self-supervised learning.

2 Related Works

2.1 GPR for Subsurface Analysis

GPR has historically found great utility in imaging and analyzing subsurface features in various outdoor environments. For example, the sensor has found use in monitoring the structural integrity of river embankments and levees [6]. The non-destructive nature of GPR enabled the visualization and detection of underground animal burrows without disrupting the site. This advantage has also been leveraged in the inspection of concrete bridges [7]. Similarly to [6], GPR was used to detect and identify the different layers present in the composition of bridges, as well as the thickness of each layer. We consider the success in the use of GPR in these varied environments as an indication of the potential viability in the use of GPR in the building rooftop setting.

2.2 Use of GPR for Building Diagnostics

Though the use of GPR has predominantly existed outside of building diagnostics, investigations on the use of GPR in various applications on buildings and similar structures exist. The technology has been used to study and identify cracks and signs of moisture damage in historical and heritage buildings [8, 9], and has even been used to detect moisture within building walls in conjunction with a thermal camera in [10]. However, many of these methods rely on manual interpretation or an ensemble of processing methods. These limitations affect the scalability and generalizability of the methods, which are key factors to consider when addressing the sheer variability in the types of buildings, not to mention the number of buildings that may need to be serviced.

2.3 Learning-Based GPR Analysis

Several works [1, 3, 4, 5, 11] have delved into learning-based analysis of GPR scan data for feature detection. For instance, [11] utilizes the mask R-CNN [12] architecture to automatically detect and segment cracks in asphalt pavement at the pixel-level. However, these approaches predominantly rely on supervised learning methods, necessitating manual annotation of GPR scans. This reliance on labeled data poses a challenge in terms of scalability due to the labor-intensive nature of annotation. Additionally, these works predominantly study GPR feature detection from the perspective of GPR B-scans as images. To address these limitations, our study explores the potential of self-supervised learning, aiming to extract meaningful features from GPR scans without the need for extensive manual annotations. We also hope to differentiate our contributions through a comprehensive analysis addressing the utility of A-scans in context of B-scan neighborhoods over the typical image-based approach.

2.4 Autoencoders for Self-Supervised Learning

The Autoencoder (AE) architecture is a popular approach to self-supervised learning. This architecture typically involves an Encoder block that learns to compress the input into the latent space, and a Decoder block that learns to extract the original input from the compressed representation [13]. Over the years, many derivatives have emerged from the original Autoencoder idea including the Variational Autoencoder (VAE), which aims to encode the

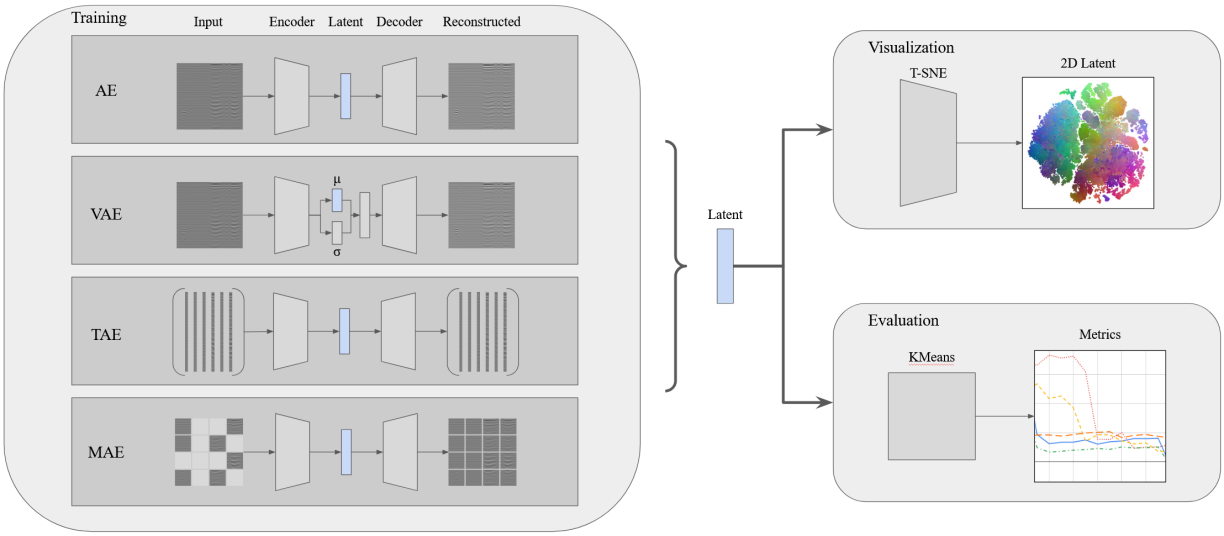

Figure 2. A simplified diagram describing the end-to-end training and testing procedure for each proposed method.

input as a distribution over the latent space [14], and the Masked Autoencoder (MAE), which motivates effective self-supervised learning through the nontrivial task of uncorrupting masked inputs [15]. Transformers [16] have also been largely successful in challenging natural language and computer vision tasks. Recalling the unique property of A-scans sharing contextual information between their neighbors within the same B-scan, we consider a Transformer-based Autoencoder (TAE) that takes a window of sequential A-scans as input. We aim to evaluate these approaches to determine their capabilities in learning meaningful features from GPR data without the need for signals from manually annotated labels.

2.5 Challenges in the Use of Synthetic GPR Data

gprMax [17] is a GPR simulation software that has been leveraged in works such as [11, 18] to generate a large corpus of synthetic scans for training deep learning models. While effective in certain applications, particularly those requiring controlled environments, the use of synthetic scans presents challenges in the context of building rooftop diagnostics. Capturing the inherent variability in rooftop composition and the diverse range of surface features, such as pipes, vents, and walls, becomes complex. The fidelity of synthetic scans may struggle to emulate the nuanced characteristics of real-world GPR data from building rooftops. Due to the present obstacles, we decide against the use of synthetic GPR data in our analysis and instead collect data from real-world commercial and retail building rooftops.

3 Methodology

Our objective is to investigate various methods for learning representations of Ground Penetrating Radar data, specifically focusing on:

1. Vanilla Autoencoder
2. Variational Autoencoder
3. Transformer-based Autoencoder
4. Image-level Masked Autoencoder

These methods are compared against a baseline approach, involving the analysis of preprocessed data without a deep learning model. The Autoencoder, Variational Autoencoder, and Transformer-based Autoencoder are trained from scratch on GPR A-scans, while the ImageNet pretrained Masked Autoencoder will be finetuned on GPR B-scans treated as images. This approach aims to highlight any distinctions between leveraging the inherent data structure of GPR and treating GPR scans as holistic images. Evaluation of each method includes the presentation of reconstruction samples, visualizations of latent spaces using T-SNE for qualitative analysis, and quantitative metrics utilizing KMeans clustering. Figure 2 provides a simplified illustration of the described procedure.

3.1 Dimensionality Reduction

T-SNE, a common dimensionality reduction technique, is employed to visualize high-dimensional latent spaces. For the deep learning models, T-SNE is applied to feature vectors generated from the encoder block's forward pass. As is typically recommended to reduce the computation of T-SNE, models that have encoders that output to relatively

high dimensions use PCA to reduce the feature dimension to 50 prior to the application of T-SNE. This procedure applies to the TAE and MAE, as well as the baseline. Feature vectors for AE, VAE, and TAE represent encoded A-scans, while those for MAE represent the 16x16 patches created from resized B-scans. Despite differences in representation, patches are reasonably analogous to A-scans as they capture local features from the entire B-scan. Indicators of meaningful encoding include visible clustering of feature vectors and separation between clusters.

3.2 Clustering

KMeans clustering is utilized to quantitatively evaluate the quality of the learned latent spaces. The algorithm is similarly applied to the generated feature vectors for each method. Two metrics, the Silhouette Score [19] and the average distance between cluster centroids, are employed for comparison. The Silhouette Score assesses cluster well-definition, while the average distance measures cluster separation. These metrics are calculated over a range of $k = [2, 3, ..., 15]$ clusters, as the appropriate number of clusters needed to capture GPR representations is unknown. The value of the Silhouette Score can range between -1 and 1, inclusive, with a value closer to 1 indicating high-quality clusters. A score greater than 0.5 is generally interpreted as reasonable.

3.3 Data Collection

The dataset comprises diverse scans collected from real-world building rooftops using the Proceq GP8800 SFCW handheld GPR sensor. SFCW, or Stepped Frequency Continuous Wave, indicates that this sensor is capable of broadcasting a wide range of frequencies to scan at both shallow and deep depths. While capable of a broader range of sampling rates, the sensor was set to record one scan per centimeter traveled, focusing on an approximately 9 x 9 cm area. The GPR is mounted on a user-controlled differential drive robot, seen in Figure 3 approximately 3cm from the ground surface. The GPR employed a servo motor to facilitate scanning by matching the GPR scanning rate with the robot's speed. During the data collection process, the robot navigated the roof in a series of straight segments, limiting the GPR data to those linear paths. The scanning process was intentionally paused during turns. Multiple scans were performed per rooftop. The straight segments were positioned approximately 0.9m apart from one another, ensuring comprehensive coverage of the entire roof surface. Multiple paths across each rooftop were scanned, resulting in a dataset of over 1400 B-scans or over 1.3 million A-scans. Each B-scan contains a variable number of A-scans – from a few tens of scans to a few thousand – and each A-scan records 655ns in two-way travel time. The dataset covers a diverse range of roofing surfaces, including built-up roofing with gravel overburden, Styrene-Butadiene-Styrene (SBS) modified asphalt roofing, Ethylene Propylene Diene Monomer (EPDM) roofing, Thermoplastic Polyolefin (TPO) roofing, and Polyvinyl Chloride (PVC) roofing. In addition, scans were acquired on both dry and moisture-saturated surfaces. The total area scanned exceeds 50k square meters.

3.4 Preprocessing

The dataset is partitioned into train, validation, and test sets (containing 1181, 144, and 144 B-scans, or over 1.1m, 136k, and 130k A-scans, respectively). While standard preprocessing techniques including signal gain and denoising are employed for better manual interpretation of GPR data, we have found that normalization by mean and standard deviation across training A-scans accomplishes the same as well as enhancing training performance.

Figure 3. Robot used for GPR Data Collection

3.5 Training

Each model is trained using varied regiments that are best suited for the model. The AE and VAE is trained for 100 epochs using a starting learning rate of 1.5e-4 and a batch size of 128. The batches are comprised of A-scans that are randomly sampled from the training dataset. The TAE is trained for 600 epochs using a starting learning rate of 1.5e-4 and a batch size of 16. The batches for the TAE are comprised of a random window of 64 sequential A-scans from a randomly sampled B-scan. If the B-scan is shorter than 64 in length, the scan is padded with the mean A-scan to make up the difference. The MAE is finetuned over 50 epochs using a starting learning rate of 1e-3 and a batch size of 16, with the masking ratio set to the default 0.75 as described in the original paper. During inference, the masking ratio is set to 0 to ensure each token is encoded, as the original architecture discards all

masked tokens immediately after masking. Each batch is comprised of randomly sampled B-scans that are resized to (224 x 224). The resized scans are additionally normalized using the mean and standard deviation calculated from ImageNet. All models are trained using the AdamW optimizer with decay = 0.05 and betas = (0.9, 0.95) and a cosine decay learning rate schedule. Training is conducted on a machine with an Nvidia RTX 8000 GPU and the code is written using PyTorch.

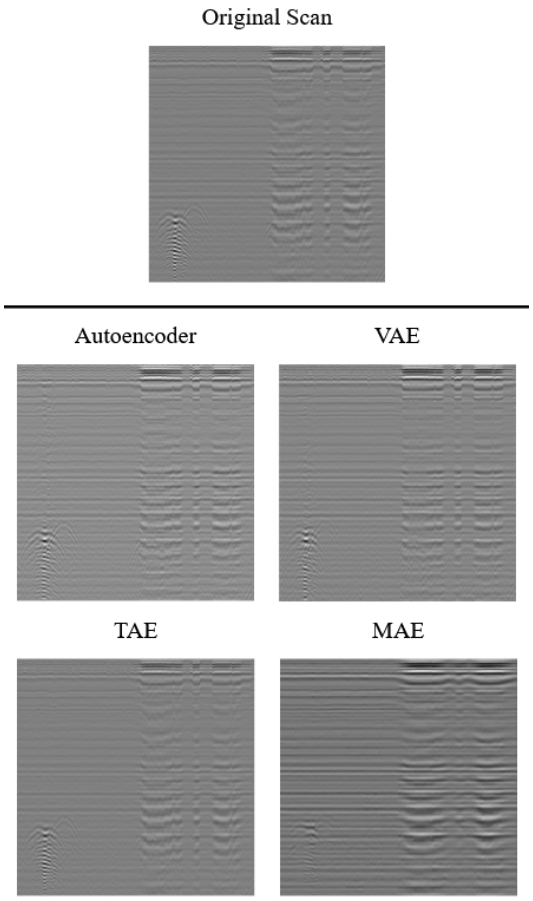

Figure 4. **Sample reconstruction outputs** from all trained deep learning methods. This scan is specifically chosen to study the diverse set of features that may be present in GPR scan data.

reading of some sort of point object, possibly a pipe, and on the right there is a response that is typical of "ponding" or moisture on the surface. All models are successful in reconstructing these features as well as the surrounding responses without excessive artifacts or noise. This is to demonstrate that the models are effective in encoding and reconstructing GPR signals, which allows us to further our investigation in determining whether anything meaningful is being encoded in the latent spaces of each model.

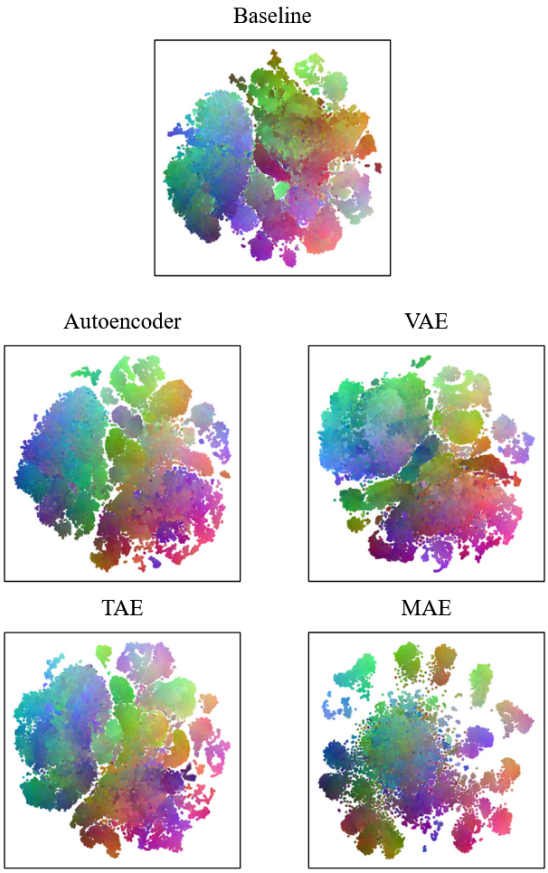

Figure 5. **Illustration of the learned latent spaces** generated from each method, with the inclusion of the baseline for reference. We note the similarities present in the baseline, AE, VAE, and TAE. This figure is best viewed in color.

4 Results

4.1 Reconstruction

As shown in Figure 4, every deep learning model trained is successful in reconstructing GPR scans. To highlight the various features that may be present in GPR scans, a specific scan was chosen from the test dataset for visualization. On the bottom left corner of the scan, there is a

4.2 Qualitative Evaluation

T-SNE visualizations are created with perplexity set to 50 to account for the large number of A-scans. Two visualizations are provided, one displaying the latent space using the entire test dataset and the other showing the encoded features of a single B-scan. This is done to examine how the models are organizing the latent space, as well as to provide a relevant visualization for interpretation.

Through analyzing the latent space, we may be able to understand whether the models are encoding certain A-scans similarly or if there is some latent structure to the data. For instance, A-scans that indicate the presence of objects like pipes would ideally be encoded closer together in a cluster and separate from A-scans that indicate moisture. Color is provided to further illustrate the location of points in 3D space. In Figure 4 we observe similar behavior to the baseline for the AE, VAE, and TAE models. While not well defined, weak clustering is apparent. Notably, the latent space appears to be organized similarly across the mentioned methods, particularly in the large cluster visible on the left. Despite the visible formation of weak feature clusters, it is not immediately apparent whether these groups are truly distinct or semantically meaningful.

This similarity is continued in Figure 6 where we observe the same pattern of a large cluster towards the left and smaller, tighter clusters on the right in the latent space visualizations. Displaying the latent spaces alongside the corresponding encoded scan provides additional visual cues that aid in understanding how the scans are encoded. We observe the possibility of a grouping of like A-scans which is particularly evident in the TAE. The A-scans belonging to the point object reading are encoded near each other, with the same being observed for the A-scans pertaining to the moisture-saturated surface.

The MAE stands out among the methods tested. The latent space visualization shown in Figure 5 for the MAE describes reasonably well-defined and separated clusters. However, the MAE results in Figure 6 are not as intuitive to interpret. The latent space appears sparse and there is no visible pattern in the encoding visualization. Therefore, it is uncertain whether the clustering exhibited in Figure 5 is representative of meaningful encodings. This result also suggests the validity of studying GPR scans as a collection of A-scans in a neighborhood of their respective B-scans.

4.3 Quantitative Evaluation

The following metrics are calculated over a range of $k = [2, 3, ..., 15]$ clusters, inclusive. This is done to ensure a holistic understanding of method performance despite not knowing intuitively how many clusters are appropriate for GPR data. The number of clusters that adequately describe the GPR signal behavior is unknown. As mentioned prior, a positive Silhouette Score that is close to 1 and a large Average Centroid Distance indicates clusters that are well defined and distinct.

While the results in Figure 7 demonstrate the AE generating higher quality clusters among the methods tested, the scores observed indicate fairly weak performance. Similarly, the results shown in Figure 8 seem to suggest a substantial separation of clusters from the AE. However, the AE performs similarly to the MAE in this metric, which

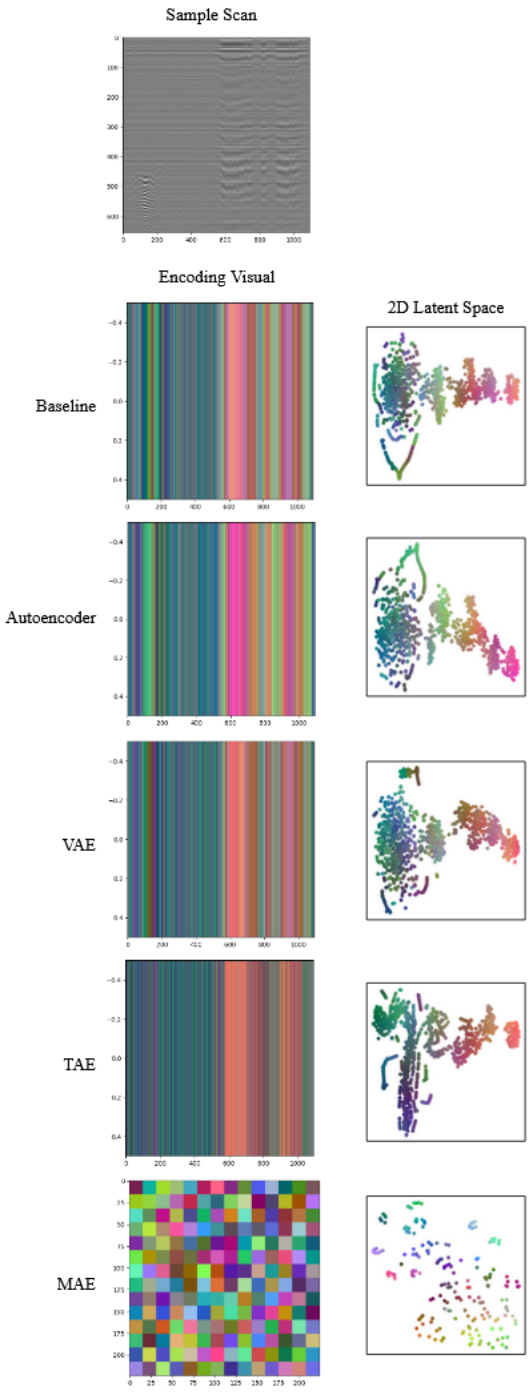

Figure 6. **Encoding visual** of a specific scan through all methods. The colors seen in the encoding visual correspond to the colors of their respective points in the 2D latent space. For the baseline, AE, VAE, and TAE, the columns represent the encoded A-scans from the sample scan. The MAE visual represents the encoded 16 x 16 patches. This figure is best viewed in color.

measured a comparatively low Silhouette Score. Thus, the quantitative results also remain inconclusive.

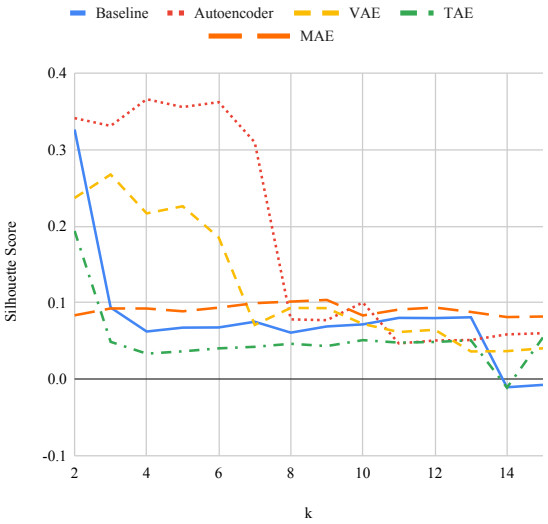

Figure 7. Silhouette Score for 2-15 clusters. This figure is best viewed in color.

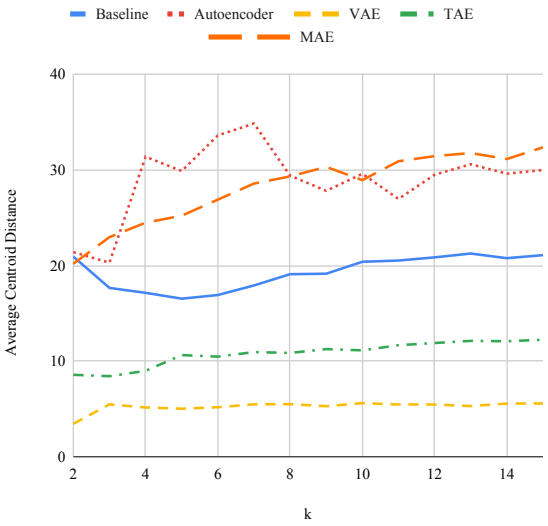

Figure 8. Average Centroid Distance for 2-15 clusters. This figure is best viewed in color.

5 Discussion and Conclusions

Self-supervision enables the possibility of leveraging the vast capabilities of deep learning in GPR processing at a massive scale. We believe that advancements in this area can further refine the performance and generalizability of works such as [1] and [4] through training on far larger and more comprehensive unlabeled GPR datasets. Additionally, enabling self-supervised training on massive real-world datasets reduces the reliance on simulated synthetic data, whose characteristics may not translate accurately to the real world. As such, we recognize the significance of extending the utility of GPR in rooftop diagnostics through self-supervised GPR representation learning. Our study, however, reveals that this task is not without its complexities.

The challenges in learning meaningful representations likely stem from the inherent variability in GPR scan data. Rooftop scans encompass diverse materials, compositions, and features distributed somewhat unpredictably. This intricacy poses a substantial obstacle for models aiming to autonomously derive semantically significant encodings.

Despite the inconclusive results in representation learning, we remain optimistic in the approach of studying GPR as a sequence of A-scans within a shared B-scan context based on the results demonstrated in Figure 6.

Moving forward, addressing these challenges may involve exploring additional pre-processing techniques or alternative model architectures. Future research should investigate strategies to enhance the robustness of self-supervised GPR representation learning, ensuring its adaptability across diverse scenarios in building diagnostics.

Acknowledgements: This work is supported by NSF grant #2322242.

Disclosure: Bilal Sher, Sruti Madhusudhan, Talha Javed, and Chen Feng are founders of Building Diagnostic Robotics, Inc., a startup company that uses AI and robotics for building inspections.

References

[1] Sher B. and Feng C. Deepgpr: Learning to identify moisture defects in building envelope assemblies from ground penetrating radar. In Borja García de Soto, Vicente Gonzalez-Moret, and Ioannis Brilakis, editors, *Proceedings of the 40th International Symposium on Automation and Robotics in Construction*, pages 561–568, Chennai, India, July 2023. International Association for Automation and Robotics in Construction (IAARC). ISBN 978-0-6458322-0-4. doi:10.22260/ISARC2023/0075.

[2] Utsi E. *Ground Penetrating Radar: Theory and Practice*. Elsevier Science, 2017. ISBN 9780081022177.

[3] Lameri S., Lombardi F., Bestagini P., Lualdi M., and Tubaro S. Landmine detection

from gpr data using convolutional neural networks. In *2017 25th European Signal Processing Conference (EUSIPCO)*, pages 508–512, 2017. doi:10.23919/EUSIPCO.2017.8081259.

[4] Hou F., Lei W., Li S., and Xi J. Deep learning-based subsurface target detection from gpr scans. *IEEE Sensors Journal*, 21(6):8161–8171, 2021. doi:10.1109/JSEN.2021.3050262.

[5] Jaufer R., Ihamouten A., Guilbert D., Todkar S., Yaram T., and Derobert X. Deep learning based automatic hyperbola detection on gpr data for buried utility pipes mapping. In *2021 11th International Workshop on Advanced Ground Penetrating Radar (IWAGPR)*, pages 1–6, 2021. doi:10.1109/IWAGPR50767.2021.9843151.

[6] Di Prinzio M., Bittelli M., Castellarin A., and Pisa P. Application of gpr to the monitoring of river embankments. *Journal of Applied Geophysics*, 71(2):53–61, 2010. ISSN 0926-9851. doi:10.1016/j.jappgeo.2010.04.002.

[7] Hugenschmidt J. and Mastrangelo R. Gpr inspection of concrete bridges. *Cement and Concrete Composites*, 28(4):384–392, 2006. ISSN 0958-9465. doi:10.1016/j.cemconcomp.2006.02.016. Non-Destructive Testing.

[8] Orlando L. and Slob E. Using multicomponent gpr to monitor cracks in a historical building. *Journal of Applied Geophysics*, 67(4):327–334, 2009. ISSN 0926-9851. doi:10.1016/j.jappgeo.2008.09.003. Advanced Applications, Systems and Modelling for GPR.

[9] Perez-Gracia V., Solla M., and Fontul S. Analysis of the gpr signal for moisture detection: application to heritage buildings. *International Journal of Architectural Heritage*, 18(2):230–253, 2024. doi:10.1080/15583058.2022.2139652.

[10] Garrido I., Solla M., Lagüela S., and Fernández N. Irt and gpr techniques for moisture detection and characterisation in buildings. *Sensors*, 20(22), 2020. ISSN 1424-8220. doi:10.3390/s20226421.

[11] Liu Z., Yeoh J., Gu X., Dong Q., Chen Y., Wu W., Wang L., and Wang D. Automatic pixel-level detection of vertical cracks in asphalt pavement based on gpr investigation and improved mask r-cnn. *Automation in Construction*, 146:104689, 2023. ISSN 0926-5805. doi:10.1016/j.autcon.2022.104689.

[12] He K., Gkioxari G., Dollár P., and Girshick R. Mask r-cnn. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2980–2988, 2017. doi:10.1109/ICCV.2017.322.

[13] Ballard D. Modular learning in neural networks. In *Proceedings of the Sixth National Conference on Artificial Intelligence - Volume 1*, AAAI'87, page 279–284. AAAI Press, 1987. ISBN 0934613427.

[14] Kingma D. and Welling M. Auto-encoding variational bayes, 2022.

[15] He K., Chen X., Xie S., Li Y., Dollár P., and Girshick R. Masked autoencoders are scalable vision learners. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15979–15988, 2022. doi:10.1109/CVPR52688.2022.01553.

[16] Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A., Kaiser L., and Polosukhin I. Attention is all you need. In Guyon I., Von Luxburg U., Bengio S., Wallach H., Fergus R., Vishwanathan S., and Garnett R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[17] Warren C., Giannopoulos A., and Giannakis I. gprmax: Open source software to simulate electromagnetic wave propagation for ground penetrating radar. *Computer Physics Communications*, 209:163–170, 2016. ISSN 0010-4655. doi:10.1016/j.cpc.2016.08.020.

[18] Dai Q., Lee Y., Sun H., Ow G., Yusof M., and Yucel A. Dmrf-unet: A two-stage deep learning scheme for gpr data inversion under heterogeneous soil conditions. *IEEE Transactions on Antennas and Propagation*, 70(8):6313–6328, August 2022. ISSN 1558-2221. doi:10.1109/tap.2022.3176386.

[19] Rousseeuw P. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20: 53–65, 1987. ISSN 0377-0427. doi:10.1016/0377-0427(87)90125-7.