The 8th European Congress on Computational Methods in Applied Sciences and Engineering
ECCOMAS Congress 2022
5-–9 June 2022, Oslo, Norway

ADVANCED DEEP LEARNING COMPARISONS FOR NON-INVASIVE TUNNEL LINING ASSESSMENT FROM GROUND PENETRATING RADAR PROFILES

M. M. Rosso[1*], G. Marasco[1], L. Tanzi[1], S. Aiello[1], A. Aloisio[2], R. Cucuzza[1], B. Chiaia[1], G. Cirrincione[3] and G.C. Marano[1]

[1] Politecnico di Torino, DISEG, Dipartimento di Ingegneria Strutturale, Edile e Geotecnica, Corso Duca Degli Abruzzi, 24, Turin 10128, Italy

[2] Università degli Studi dell'Aquila, DICEAA, Department of Civil, Construction-Architectural and Environmental Engineering, Via G. Gronchi, 18, L'Aquila, 67100 Abruzzo, Italy

[3] University of Picardie Jules Verne, Lab. LTI, Amiens, France.

[*] Corresponding Author: marco.rosso@polito.it

Key words: Tunnel Linings, Ground Penetrating Radar, Deep Learning, Convolutional Neural Network, Vision Transformer, Structural Health Monitoring

Abstract. Innovative, automated, and non-invasive techniques have been developed by scientific community to indirectly assess structural conditions and support the decision-making process for a worthwhile maintenance schedule. Nowadays, machine learning tools are in the spotlight because of their outstanding capabilities to deal with data coming from even heterogeneous sources and their ability to extract information from the structural systems, providing highly effective, reliable, and efficient damage classification tools. In the current study, a supervised multi-level damage classification strategy has been developed regarding Ground Penetrating Radar (GPR) profiles for the assessment of tunnel lining conditions. In previous research, the authors firstly considered a convolutional neural network (CNN), adopting the quite popular ResNet-50, initialized through transfer learning. In the present work, further enhancements have been attempted by adopting two configurations of the newest state-of-art advanced neural architectures: the neural transformers. The foremost is the original Vision Transformer (ViT), whose core is an encoder entirely based on the innovative self-attention mechanism and does not rely on convolution at all. The second is an improvement of ViT which merges convolution and self-attention, the Compact Convolution Transformer (CCT). In conclusion, a critical discussion of the different pros and cons of adopting the above-mentioned different architectures is finally provided, highlighting the actual powerfulness of these technologies in the future civil engineering paradigm nevertheless.

1 INTRODUCTION

Nowadays, the existing infrastructure heritage is experiencing a reduction in the safety levels, especially due to aging and degradation effects [1, 2, 3]. Bridges [4, 5, 6] and tunnels [7, 8]

represents the most important large civil engineering infrastructure items. They accomplish strategic functions, permitting the connectivity of communications routes to overcome natural obstacles. For this reason, in the last decade, a noticeably growing interest has been paid to the research within the structural health monitoring (SHM) field. In particular, indirect and non-invasive monitoring techniques gained particular notoriety due to their increased reliability, less invasive investigations, and reduction of the investigation time, which results in lowering the global maintenance and rehabilitation costs. Automated procedures have been developed by researchers both to improve the efficiency and productivity of the surveys and even to increase the objectivity of the obtained results without requiring procedures strongly affected by the personnel experience [9, 10]. In the present work, a non-destructive structural testing (NDT) technique based on ground penetrating radar (GPR) has been explored. The adoption of GPR relies on the emission of electromagnetic (EM) wave impulses inside the material under study. A receiver antenna collects the reflected back signals to inspect the material in-depth [11]. The GPR provides therefore an image as output, evidencing the presence of anomalies, defects, fractures, etc. overcoming the criticalities of a direct visual inspection. When GPR is adopted to monitor the health status of a tunnel, the GPR linings require experienced personnel to identify and classify the presence of any defects. Machine learning (ML) and, especially, deep learning (DL) methods provide modern and powerful tools for automatic image processing and classification tasks [12, 13]. Convolutional networks (CNNs) have been demonstrated as effective tools to accomplish those tasks, and they represent the most nowadays widespread adopted techniques [14, 15]. In the present work, two other advanced DL methods based on neural transformers have been tested in the road tunnels SHM paradigm. The main goal is to show how recent progress in artificial intelligence (AI) may further improve or not the contemporary CNNs models [16], and if these novel approaches may replace CNNs soon. In next section 2, the neural vision transformer (ViT) [17] and the compact convolution transformer (CCT) [18] are briefly described. In section 3, the AI-based GPR defects automatic classification procedure and the adopted dataset has been described. Eventually, in section 4, the obtained results from the first 3 classification levels are presented and critically discussed.

2 NEURAL TRANSFORMERS: BRIEF OVERVIEW

Since 2017, the study entitled "*Attention is all you need*" [19] echoed as a revolution in DL field. The authors introduced for the first time the novel architecture of neural transformers for Natural Language Processing (NLP) tasks. This paradigm is based on the self-attention mechanism which analyses the entire sequence (sequence-to-sequence mechanism) and retrieves the relationships among the elements, even long-range ones. Transformers' blocks may be parallelized for computational efficiency and may also deal with large-scale datasets. Since these models are remarkably computational demanding, transfer learning processes has revealed to be the most promising direction to effectively exploit them [14]. Recently, impactful and fruitful studies have been conducted such as, e.g., the introduction of BERT model (Bidirectional Encoder Representations from Transformers) [20], or the adaptation of neural transformers to deal with images data types, as further illustrated in the next subsections.

2.1 Vision transformer

Vision Transformer (ViT) is a new state-of-art approach in deep learning for image processing and image classification. Firstly appearing in October 2020, the ViT sounds to beat the best CNN models such as Res-Net for image classification for a sufficiently large dataset for pretraining [21]. ViT is based solely on the encoder network part of the neural transformer architecture, born for natural language processing [19]. Foremost, the ViT model requires partitioning the input image into n patches of the same shape, which can overlap or not, e.g. produced by a sliding window that slides on the input images based on a user-defined stride. Each patch in general is a size 3 tensor of shape $d_1 \times d_2 \times d_3$ corresponding to the Red-Green-Blue (RGB) digital image encoding. A vectorization procedure involves each i-th patch producing a column vector $\boldsymbol{x}_i$, with $i = 1, 2, ..., n$, of dimension $d_1 d_2 d_3 \times 1$. All the n vectors are then fed into a learnable dense layer with shared parameters and with a linear activation function to produce an embedding $\boldsymbol{z}_i$ for each $\boldsymbol{x}_i$. The weights parameters are learned from training data during the training phase. This procedure is acknowledged as flattening using a linear projection matrix [17]. To consider the actual position of patches with respect to the initial input image, a positional encoding [19] is applied and summed to the $\boldsymbol{z}_i$. In this way, the new representation of the input information $\boldsymbol{z}_i$ captures both the content and the position of the i-th patch. Furthermore, the [CLASS] token for classification of BERT model [20] is fed to an embedding layer producing the vector $\boldsymbol{z}_0$ of the same shape as other embeddings. The sequence of vectors $\{\boldsymbol{z}_i\}_{i=0}^{n}$ are subsequently fed to the neural transformer encoder block, composed of a stack of a multi-head self-attention and dense fully-connected layer blocks, actually employing normalization and skip connections. The output of the neural transformer encoder is a new representation of the input vectors $\{\boldsymbol{z}_i\}_{i=0}^{n}$ mapped to a new representations $\{\boldsymbol{c}_i\}_{i=0}^{n}$ which integrates the scaled dot-product attention (the self-attention) [19]. In any case, only the $\boldsymbol{c}_0$ is considered for the classification task, and the others are usually ignored. This vector represents the feature vector from the input image. This vector is thus fed to a softmax classifier which outputs into a column vector $\boldsymbol{p}$ of size equal to the number of output classes. Each element of $\boldsymbol{p}$ presents a probability associated with each output class. In the current study, a pre-trained ViT model has been considered and only a fine-tuning of the last classification layers has been performed on the GPR profiles, within a transfer learning approach, likewise in [17].

2.2 Compact convolutional transformer

In [18], the authors proposed a novel variant of the standard ViT model, attempting to provide a compact version of this model in order to overcome the common problem of transformers "*data-hungry*". Since they have millions of parameters, it appeared impossible to train a ViT from scratch for most common applications where datasets are usually limited in size [17]. The Compact Convolution Transformer (CCT) starts from the ViT architecture and provides a few essential changes. Foremost, the CCT introduced a convolutional tokenization method, with a customizable number of blocks, each of them based on the conventional convolution blocks:

$$\boldsymbol{x}_0 = \text{MaxPool}[\text{ReLU}(\text{Conv2d}(\boldsymbol{x}))] \tag{1}$$

where the convolution operation Conv2d is performed by d filters applied to the input image $\boldsymbol{x}$, coincident with the embedding dimension of the encoder block. In this way, the CCT bypasses

the patching problem, which requires an image whose sizes are exactly divisible by the patch size. Moreover, the convolution, followed by rectified linear unit (ReLU) activation function and MaxPool pooling operation, provide a more efficient compact image embedding representation [18]. CNN desirable properties, such as efficiency, learnable weight sharing, local information preservation, and equivariant representations, have been embedded in this way into the transformer paradigm. Furthermore, by exploiting the self-attention mechanism, CTT overcomes the CNN shortcoming when dealing with long-sequence relationships, besides preserving local information among patches. Unlike the ViT model, CCT removes the BERT-based [CLASS] token to perform the final classification. CTT rather introduces a so-called sequence pooling mechanism, to compress the information of the output results from the multi-head self-attention blocks, improving the efficiency of the final multi-layer perceptron (MLP) which actually performs the classification task [18].

3 GPR FOR ROAD TUNNEL SHM

The adoption of GPR permits indirect NDT for road tunnel health inspections with quite rapid and highly efficient surveys [22]. Based on a geophysical technique [11], the GPR reconstructs an image of the internal conditions of the material investigated according to its dielectric characteristics. A GPR tunnel lining profile image presents the explored depth on the vertical axis and the longitudinal progressive distance from the beginning of the tunnel until its end on the abscissa axis. The high-frequency EM impulses (frequency in the range of 10-2600 MHz) are reflected when the EM waves encounter sudden dielectric feature changes, possibly induced by anomalies, voids, frontiers with different materials, cracks, etc. In the reconstructed images, each of these causes provides appreciably characteristic patterns, which can be virtually reconducted to its specific cause. Nowadays, expert and qualified staff provide a manual interpretation of GPR tunnel linings profiles for SHM purposes, resulting in a labeled image as depicted in Figure 1. The main goal of the current study is to explore AI and DL-based methods, providing more reliable, less subjective, and automatic image processing and tunnel defects classification [12, 13].

3.1 Hierarchical multi-level road tunnels defect classification

The GPR road tunnel linings dataset of the present work refers to a dataset collected by a GPR testing campaign performed on Italian tunnels dated between the 1960s and 1980s. The size of the available dataset is depicted in Figure 2. Moreover, two types of GPR have been adopted for the NDT surveys based on the frequency range, and thus on the type of antenna [15]. Starting from human experts' labelled profiles, those images have been cropped with a constant step of 5.00 m each. These cropped images may contain well recognizable defects, therefore they represented the actual input data for the DL models. When required, the cutting step has been manually adjusted to avoid the same defect straddles between two consecutive cropped samples. In this way, the quality of the samples improved, affecting also the DL learning performances. In the current work, the classification task is organized as depicted in Figure 2. It involves a hierarchical multi-level classification procedure to better identify which kind of defect is observed in the GPR tunnel linings profiles. In general, seven DL classification models should be trained to accomplish the GPR tunnel defects classification tasks for SHM purposes. In the

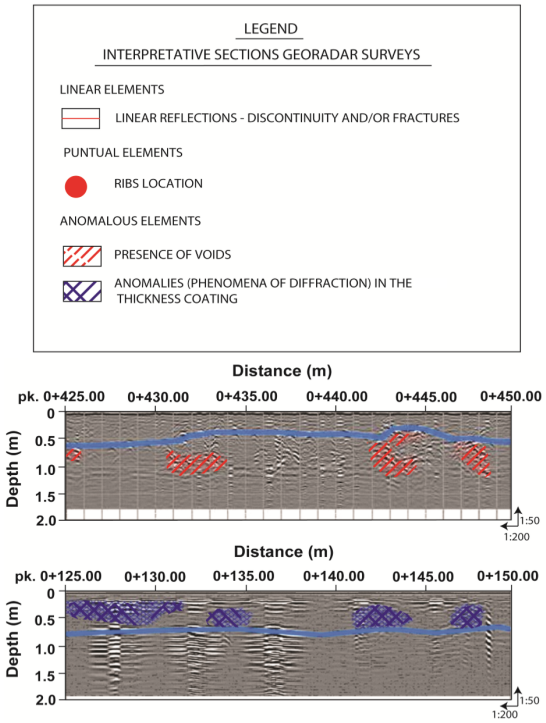


Figure 1: An example of a GPR profile with defect patterns interpretation by human experts.

present work, only the first three models have been analysed, i.e. the level 6, the level 5, and the level 4 illustrated in Figure 2. On these three analysed levels, the ViT and the CCT have been trained and tested.

4 RESULTS AND DISCUSSION

In the present work, the first three out of seven DL classification models have been analysed to accomplish the GPR tunnel defects classification tasks for SHM purposes with state-of-art image-processing advanced DL approaches. Specifically, referring to Figure 2, the authors focused on level 6 (1080 images in total), level 5 (2188 images in total), and level 4 (3124 images in total). On these three analysed levels, the ViT and the CCT have been trained and tested. The dataset has been split considering 90% of the total images per level to belong to the training set and the resulting 10% belonging to the test set. For the ViT model, the pre-trained model provided by [23] has been imported based on available python Keras implementation [24]. Fine-tuning final blocks have been added to the ViT model, the input images have been resized to 224x224 pixels, batch size has been fixed to 16, a validation set of 10% of the training set has been used during the training process, the categorical cross-entropy loss [25] has been adopted, and the model for each level has been trained for 20 epochs. The results obtained on the test set are illustrated in Table 1 in terms of confusion matrices, precision, recall, f1-score metrics [26]. For the ViT model, an example of self-attention maps has also been provided in Figure 3.

For the CCT model, model training from scratch has been performed according to [18]. Two

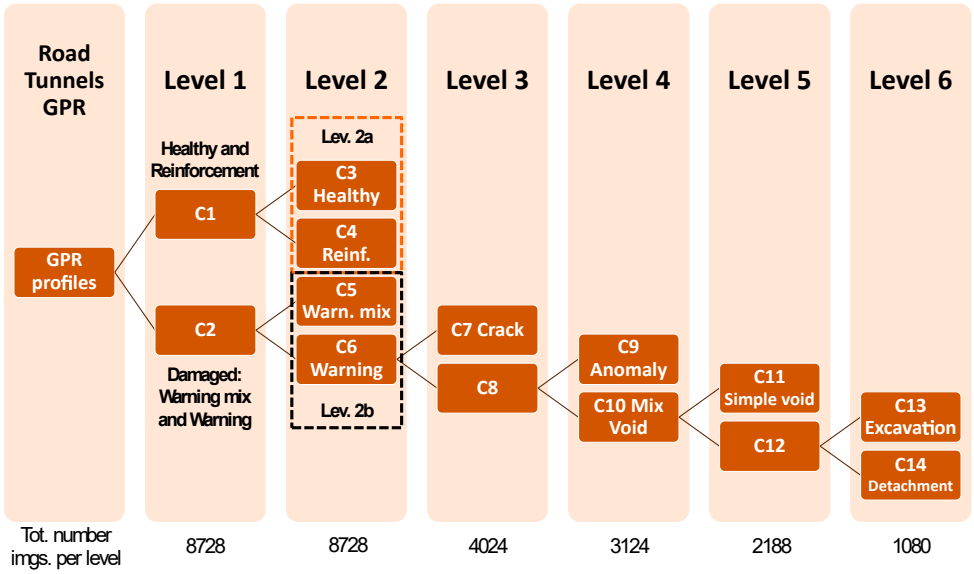
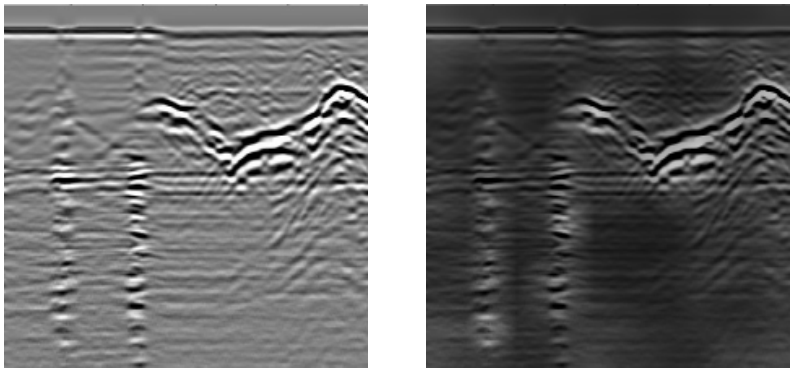


Figure 2: Hierarchical tree multi-level classification representation.



(a) (b)

Figure 3: ViT processed output: example of attention map visualization. (a) resized image belonging to C11 class; (b) attention map provided by ViT model.

convolutional layers have been adopted in place of the image patching of ViT, with kernel size equal to 3 and pooling stride set to 2. The optional positional embedding has been maintained in the current implementation. The input images have been resized to 224x224 pixels, the number of encoder layer blocks has been set equal to 2, the batch size has been set to 16, the categorical cross-entropy loss [25] has been adopted, and each model has been trained for 20 epochs. The results obtained on the test set are illustrated in Table 2 in terms of confusion matrices, precision, recall, f1-score metrics [26].

Table 1: Results of ViT model for levels 6, 5 and 4.

Level 6								
	Predicted		Accuracy	Tot. images/class		Test set support	Precision	Recall	f1-score
True	C13	C14		C13	408	53	1.000	0.981	0.990
C13	52	1	0.99074	C14	672	55	0.982	1.000	0.991
C14	0	55		Total	1080	108			

Level 5								
	Predicted		Accuracy	Tot. images/class		Test set support	Precision	Recall	f1-score
True	C11	C12		C11	1108	115	0.991	1.000	0.996
C11	115	0	0.99543	C12	1080	104	1.000	0.990	0.995
C12	1	103		Total	2188	219			

Level 4								
	Predicted		Accuracy	Tot. images/class		Test set support	Precision	Recall	f1-score
True	C9	C10		C9	936	96	0.989	0.979	0.984
C9	94	2	0.99042	C10	2188	217	0.991	0.995	0.993
C10	1	216		Total	3124	313			

Table 2: Results of CCT model for levels 6, 5 and 4.

Level 6								
	Predicted		Accuracy	Tot. images/class		Test set support	Precision	Recall	f1-score
True	C13	C14		C13	408	53	0.875	0.660	0.753
C13	35	18	0.78704	C14	672	55	0.735	0.909	0.813
C14	5	50		Total	1080	108			

Level 5								
	Predicted		Accuracy	Tot. images/class		Test set support	Precision	Recall	f1-score
True	C11	C12		C11	1108	115	0.821	0.800	0.811
C11	92	23	0.80365	C12	1080	104	0.785	0.808	0.796
C12	20	84		Total	2188	219			

Level 4								
	Predicted		Accuracy	Tot. images/class		Test set support	Precision	Recall	f1-score
True	C9	C10		C9	936	96	0.774	0.750	0.762
C9	72	24	0.85623	C10	2188	217	0.891	0.903	0.897
C10	21	196		Total	3124	313			

4.1 Discussion

The advanced DL VIT and CCT models have been trained for automatic defects classification of road tunnels with NDT based on indirect GPR tunnel linings profiles. Table 1 illustrates the advantages of adopting a pre-trained model in the spirit of a transfer learning approach. Because of this approach, even with a reduced number of epochs, the results in terms of accuracy are noticeably for the so far analyzed levels. The overall accuracy for the three analyzed levels is obtained by the product of the three level's accuracy, resulting in about 97.7%. The self-

attention mechanism provides the ViT of a powerful feature extractor, as depicted in Figure 3. In this Figure, it is possible to notice how the attention map enhances the characteristic parts which help the DL model in the classification task, and how the useless parts of the image have been darkened. Both precision and recall are close to unity, and their harmonic mean, i.e. the f1-score, demonstrates remarkable high performances for the fine-tuned ViT model for automated GPR defect classification. On the other hand, despite CCT being declared to be a transformer compact version able to work with small datasets, the performances obtained in this study are mildly satisfactory. The accuracy for each level stalled at about 80%, which virtually appears quite acceptable. However, if the overall accuracy is computed, this results in about 54.2%, thus quite poor performances. Furthermore, a deeper insight into the confusion matrix evidence how one class is often better recognized with respect to the other class at each binary classification level. This fact evidences that the CCT exhibits insufficient classification ability. Moreover, in this case, the accuracy does not represent the best metric to evaluate CCT performance, but the f1-score may represent a more reliable metric. Probably, these poor performances are referred to the fact that training CCT from scratch involves millions of parameters, however, the dataset is generally small compared to the number of learnable parameters. This produces the CCT model to be essentially prone behaviour towards overfitting issues. Future improvements in CCT performances may be related to increase the training epochs to further reduce the loss, or increase the batch size to 32. This may virtually lead to higher computational efforts, but it virtually represents the first attempt to improve the performance of the model, always with special consideration for overfitting issues.

5 CONCLUSIONS

In the present work, two advanced state-of-art DL models for image processing have been adopted for road tunnels GPR linings for automatic multi-level defects classification for SHM purposes. Specifically, the neural transformers adapted to deal with image data have been used. The foremost DL technique is the ViT model adopted with the transfer learning paradigm. The ViT approach exhibited remarkably interesting results for the analyzed classification levels on the test set, providing an overall accuracy of around 97.7%. For the sake of comparisons, another transformer variant, the CCT, has been explored since their interesting compact version which exploits convolutional layer instead of patching methods as starting embedding and tokenization procedure. However, the resulting performances in terms of accuracy on the test set appear mildly satisfactory considering every binary classification level alone (around 80% for every single level), nevertheless resulting in insufficient overall accuracy performances considering the three analysed levels (about just 54.2%). In future studies, the authors will extend the present implementation to the other classification levels and compare the transformer solution with other more simple and widespread techniques, such as CNN.

6 ACKNOWLEDGMENTS

This research was supported by project MSCA-RISE-2020 Marie Skłodowska-Curie Research and Innovation Staff Exchange (RISE) - ADDOPTML (ntua.gr) The authors would like to thank G.C. Marano and the project ADDOPTML for funding supporting this research.
Computational resources provided by hpc@polito (http://www.hpc.polito.it).

References

[1] Mingfeng Lei, Linghui Liu, Chenghua Shi, Yuan Tan, Yuexiang Lin, and Weidong Wang. A novel tunnel-lining crack recognition system based on digital image technology. *Tunnelling and Underground Space Technology*, 108:103724, 2021. ISSN 0886-7798. doi: https://doi.org/10.1016/j.tust.2020.103724.

[2] Raffaele Cucuzza, C. Costi, Marco Martino Rosso, Marco Domaneschi, Giuseppe Marano, and D. Masera. Optimal strengthening by steel truss arches in prestressed girder bridges. *Proceedings of the Institution of Civil Engineers - Bridge Engineering*, pages 1–51, 01 2022. doi: 10.1680/jbren.21.00056.

[3] Fabio Di Trapani, Giovanni Tomaselli, Antonio Pio Sberna, Marco Martino Rosso, Giuseppe Carlo Marano, Liborio Cavaleri, and Gabriele Bertagnoli. Dynamic response of infilled frames subject to accidental column losses. In Carlo Pellegrino, Flora Faleschini, Mariano Angelo Zanini, José C. Matos, Joan R. Casas, and Alfred Strauss, editors, *Proceedings of the 1st Conference of the European Association on Quality Control of Bridges and Structures*, pages 1100–1107, Cham, 2022. Springer International Publishing. ISBN 978-3-030-91877-4.

[4] Bernardino Chiaia, Giulio Ventura, Cristina Zannini Quirini, and Giulia Marasco. Bridge active monitoring for maintenance and structural safety. In *International Conference on Arch Bridges*, pages 866–873. Springer, 2019.

[5] Rebecca Asso, Raffaele Cucuzza, Marco Martino Rosso, Davide Masera, and Giuseppe Carlo Marano. Bridges monitoring: an application of ai with gaussian processes. In *14th International Conference on Evolutionary and Deterministic Methods for Design, Optimization and Control*. Institute of Structural Analysis and Antiseismic Research National Technical University of Athens, 2021. doi: https://doi.org/10.7712/140121.7964.18426.

[6] Angelo Aloisio, Dag Pasquale Pasca, Luca Battista, Marco Martino Rosso, Raffaele Cucuzza, Giuseppe Marano, and Rocco Alaggio. Indirect assessment of concrete resistance from fe model updating and young's modulus estimation of a multi-span psc viaduct: Experimental tests and validation. *Elsevier Structures*, 37:686–697, 01 2022. doi: 10.1016/j.istruc.2022.01.045.

[7] S. Bhalla, Y.W. Yang, J. Zhao, and C.K. Soh. Structural health monitoring of underground facilities – technological issues and challenges. *Tunnelling and Underground Space Technology*, 20(5):487–500, 2005. ISSN 0886-7798. doi: https://doi.org/10.1016/j.tust.2005.03.003.

[8] Allen G Davis, Malcolm K Lim, and Claus Germann Petersen. Rapid and economical evaluation of concrete tunnel linings with impulse response and impulse radar non-destructive methods. *NDT & E International*, 38(3):181–186, 2005.

[9] Yujing Jiang, Xuepeng Zhang, and Tetsuya Taniguchi. Quantitative condition inspection and assessment of tunnel lining. *Automation in Construction*, 102:258–269, 2019. ISSN 0926-5805. doi: https://doi.org/10.1016/j.autcon.2019.03.001.

[10] Leanne Attard, Carl James Debono, Gianluca Valentino, and Mario Di Castro. Tunnel inspection using photogrammetric techniques and image processing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 144:180–188, 2018. ISSN 0924-2716. doi: https://doi.org/10.1016/j.isprsjprs.2018.07.010.

[11] E. Cardarelli, C. Marrone, and L. Orlando. Evaluation of tunnel stability using integrated geophysical methods. *Journal of Applied Geophysics*, 52(2):93–102, 2003. ISSN 0926-9851. doi: https://doi.org/10.1016/S0926-9851(02)00242-2.

[12] Thikra Dawood, Zhenhua Zhu, and Tarek Zayed. Deterioration mapping in subway infrastructure using sensory data of gpr. *Tunnelling and Underground Space Technology*, 103: 103487, 2020. ISSN 0886-7798. doi: https://doi.org/10.1016/j.tust.2020.103487.

[13] W Al-Nuaimy, Y Huang, M Nakhkash, M.T.C Fang, V.T Nguyen, and A Eriksen. Automatic detection of buried utilities and solid objects with gpr using neural networks and pattern recognition. *Journal of Applied Geophysics*, 43(2):157–165, 2000. ISSN 0926-9851. doi: https://doi.org/10.1016/S0926-9851(99)00055-5.

[14] Chuncheng Feng, Hua Zhang, Shuang Wang, Yonglong Li, Haoran Wang, and Fei Yan. Structural damage detection using deep convolutional neural network and transfer learning. *KSCE Journal of Civil Engineering*, 23(10):4493–4502, 2019.

[15] Bernardino Chiaia, Giulia Marasco, and Salvatore Aiello. Deep convolutional neural network for multi-level non-invasive tunnel lining assessment. *Frontiers Of Structural And Civil Engineering*, 2022. doi: https://doi.org/10.1007/s11709-021-0800-2.

[16] Waseem Rawat and Zenghui Wang. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Computation*, 29(9):2352–2449, 09 2017. ISSN 0899-7667. doi: 10.1162/neco_a_00990. URL `https://doi.org/10.1162/neco_a_00990`.

[17] Leonardo Tanzi, Andrea Audisio, Giansalvo Cirrincione, Alessandro Aprato, and Enrico Vezzetti. *Vision Transformer for femur fracture classification.* 2021.

[18] Ali Hassani, Steven Walton, Nikhil Shah, Abulikemu Abuduweili, Jiachen Li, and Humphrey Shi. *Escaping the Big Data Paradigm with Compact Transformers.* 2021.

[19] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[20] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[21] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021. URL `https://openreview.net/forum?id=YicbFdNTTy`.

[22] Sandeep Kumar Dwivedi, Manish Vishwakarma, and Prof.Akhilesh Soni. Advances and researches on non destructive testing: A review. *Materials Today: Proceedings*, 5(2, Part 1):3690–3698, 2018. ISSN 2214-7853. doi: https://doi.org/10.1016/j.matpr.2017.11.620. 7th International Conference of Materials Processing and Characterization, March 17-19, 2017.

[23] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[24] Fausto Morales et al. vit-keras, keras implementation of vit (vision transformer), 2015. URL `https://github.com/faustomorales/vit-keras`.

[25] Charu C Aggarwal et al. Neural networks and deep learning. *Springer*, 10:978–3, 2018.

[26] Aurélien Géron. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems.* " O'Reilly Media, Inc.", 2019.