Comparative deep learning studies for indirect tunnel monitoring with and without Fourier pre-processing

Journal Title 0 (0) 1
IOS Press

Comparative deep learning studies for indirect tunnel monitoring with and without Fourier pre-processing

Marco Martino Rosso [a], Angelo Aloisio [b], Vincenzo Randazzo [a], Leonardo Tanzi [a], Giansalvo Cirrincione [c], Giuseppe Carlo Marano [a,*]

[a] *DISEG, Department of Structural, Geotechnical and Building Engineering, Politecnico di Torino, Turin, Italy*
E-mails: marco.rosso@polito.it, vincenzo.randazzo@polito.it, leonardo.tanzi@polito.it, giuseppe.marano@polito.it
[b] *DICEAA, Civil Environmental and Architectural Engineering Department, University of L'Aquila, L'Aquila, Italy*
E-mail: angelo.aloisio1@univaq.it
[c] *Lab. LTI, University of Picardie Jules Verne, Amiens, France*
E-mail: exin@u-picardie.fr

Abstract. In the last decades, the majority of the existing infrastructure heritage is approaching the end of its nominal design life mainly due to aging, deterioration, and degradation phenomena, threatening the safety levels of these strategic routes of communications. For civil engineers and researchers devoted to assessing and monitoring the structural health (SHM) of existing structures, the demand for innovative indirect non-destructive testing (NDT) methods aided with artificial intelligence (AI) is progressively spreading. In the present study, the authors analyzed the exertion of various deep learning models in order to increase the productivity of classifying ground penetrating radar (GPR) images for SHM purposes, especially focusing on road tunnel linings evaluations. Specifically, the authors presented a comparative study employing two convolutional models, i.e. the ResNet-50 and the EfficientNet-B0, and a recent transformer model, i.e. the Vision Transformer (ViT). Precisely, the authors evaluated the effects of training the models with or without pre-processed data through the bi-dimensional Fourier transform. Despite the theoretical advantages envisaged by adopting this kind of pre-processing technique on GPR images, the best classification performances have been still manifested by the classifiers trained without the Fourier pre-processing.

Keywords: Convolutional Neural Networks, Transformer, Fourier Transforms, Ground Penetrating Radar Systems, Nondestructive Examination

1. Introduction

Nowadays, existing strategic infrastructures such as bridges and tunnels are experiencing a substantial reduction in safety levels for deterioration phenomena due to long-term degradation effects of their constitutive materials [1, 2]. To extend the service life of existing heritage, the most widespread approach is monitoring the structural health (SHM) of the systems in order to effectively plan and prioritize preventive maintenance or rehabilitation interventions crucial for lifecycle [3–5]. Since a total replacement of existing infrastructures would be economically unsustainable [6], efficient and innovative monitoring techniques have been developed in the last decades [7]. Periodic direct testing of specimens (e.g. concrete core drilling) is a reliable solution to directly assess the quality, mechanical properties, and temporal changes of the in-situ constitutive structural materials. However, these tests provide punctual, albeit detailed, information, which does not always reflect the actual state of the en-

*Corresponding author. E-mail: giuseppe.marano@polito.it.

tire structure [8]. Moreover, the overall involved direct testing procedures are often lengthy and costly. Therefore, to increase the productivity and quickness of periodical inspections, non-destructive evaluations (NDE), also acknowledged as non-destructive testing (NDT) techniques, have become more prominent, reliable, and adopted methods, lately [9–12]. They are often employed in combination with direct testing to increase the quickness, reduce the expenses, and, in general, mutually overcome the limits of each other [13]. Principally focusing on SHM for road tunnels structures [14, 15], some of the most adopted NDT techniques are e.g. rebound hammer testing [16], ultrasonic pulse testing [17], rebar scanning with pachometer device [18], concrete resistivity [19], acoustic emission passive monitoring for micro-cracks detection [20], thermal imaging thermography with infrared cameras [21, 22], laser scanner and lidar devices to monitoring tunnel linings deformations [21]. Some innovative approaches rely on emerging advanced technologies such as distributed fiber optic sensors [23] or internet of things (IoT) edge devices [24–26]. In the current study, the authors predominantly concentrated on indirect testing with ground penetrating radar (GPR) devices for concrete linings defects detection and annotation [27, 28], even if, in literature, GPR is often adopted to reveal tunnel lining concrete layer thickness [21]. The GPR instrumentation overcomes the limitations of visual inspections, qualified only to catch superficial defects [15]. Similarly to other geophysical methods [29], the GPR device probes the tunnel linings by propagating high-frequency electromagnetic wave impulses (10-2600 MHz) and analyzing the reflected signals [30]. The impulses' penetration level or reflection rate depends on the dielectric features of the inspected material and the possible presence of certain agents (e.g. water, reinforcement bars, the interface between concrete linings and surrounding ground, linings defects). The architecture of a GPR system is composed of emitting and receiver units, a single or dual frequency antenna, display, control, and storage unit [30]. The GPR provides images as output named profiles, where the abscissa represents the progressive distance from the beginning of the probing (i.e. beginning of the tunnel), whereas the ordinate axis represents the GPR examined lining depth. As depicted in Figure 1, in a traditional GPR indirect testing pipeline, specialist staff decodes linings defects from the surveyed profiles with a manual, lengthy and costly post-processing phase [31].

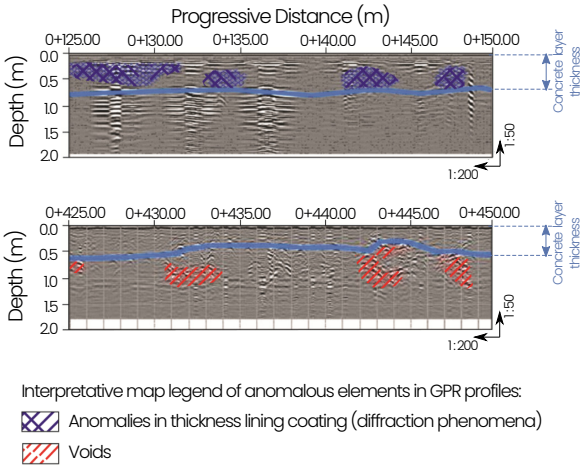

Fig. 1. GPR lining defects recognition by specialist staff.

To improve the efficiency, reliability, and productivity of the traditional GPR monitoring process, artificial intelligence (AI) offers innovative tools to accomplish the above-mentioned task by leveraging computer vision and image processing-based methods [32–36]. Specifically, deep learning (DL) techniques such as convolutional neural networks (CNNs) have been extensively employed for SHM applications [37]. In the existing literature, some innovative DL-based procedures have been introduced in GPR tunnel linings indirect monitoring recently [38–43]. In the review paper [44], the authors evidenced that despite the first adoption of GPR device in the tunnel-related field actually started in the late 1970s, a limited number of research studies have employed deep learning techniques hitherto, motivating the current interest of the present document within this active research field. In [45], the authors adopted deep learning models just to recognize the presence of rebars and to determine the thickness of the concrete layer, without taking into account any other defects or damage. In [46, 47], the authors employed a region proposal CNN named Faster R-CNN for specific target detection in tunnel lining GPR images. Specifically, in [47] the DL models have been trained for very limited purposes, i.e for detecting only rebars in tunnel linings structures. GPR tunnel liner dielectric properties (permittivity maps) inversion and objects identification tasks have been addressed in [48] through a CNN model combined with a recurrent neural network (RNN) composed of bidirectional convolutional long short-term memory (LSTM) blocks. In [49], tunnel linings defects automatic classification has been accomplished with two convolutional models, i.e. the visual geometry group (VGG)

network, i.e. the VGG-16, and the residual neural network (ResNet) with 34 convolutional layers, i.e. the ResNet-34. This study is very limited because the DL model simply divides healthy sample images from the ones with any defects, additionally without explicating which types of considered defects. Furthermore, this study does not present any generality possibilities because the models were trained only on the GPR data coming from the same tunnel, strongly restricting any direct exportation of the trained model to different tunnels. Similarly, another CNN-based automatic defects classification has been proposed in [50] by adopting the rotational region deformable convolutional neural network (R^2DCNN). Since their scarce availability of real data, the authors artificially created synthetic GPR images. However, they admitted the limited quality of synthetic data, unable to capture all the real-world complex conditions. They attempted to apply their proposed synthetically-trained model to two real tunnels. However, fine-tuning was strictly required for both tunnels. Consequently, the main restriction of their methodology is the inability to directly export the trained models to real tunnels GPR data. This strongly limits the direct application of their model to real-world scenarios since it always requires fine-tuning for every tunnel's specific conditions. Instead, as reported later in the present document, the current dataset is composed of various real Italian tunnels. This provides a greater generality of our models, directly exportable to different real-world conditions without virtually requiring any further adaptation. The same authors of [50], in [48], presented the same limitations because they employed only synthetic data and validated their neural model for permittivity map reconstruction only on a laboratory sandbox test with buried objects, without testing on real-world tunnels. In [51], the authors employed a CNN encoder-decoder structure leveraging the similarities with the geophysical seismic inversion procedure to reconstruct permittivity maps. Even in this case, the main limitation is the adoption of synthetic data only. A different approach based on generative adversarial network (GAN) has been used in [52] attempting to attenuate the GPR wave echoes and reflections produced by the reinforcement bars presence in tunnel linings.

In the present study, the authors compared three different DL models, two convolutional models, i.e. *ResNet-50* and *EfficientNet*, and a recent transformer model in the version suited for working with image data, i.e. *Vision Transformer* (ViT). To the authors' knowledge, the present work introduced for the very first time these advanced neural models, i.e. the transformer, for the GPR tunnel linings defects classification task. Peculiarly, to provide reliable, automatic, and AI-aided GPR profiles post-processing, the authors employed the hierarchical multi-level classification tree proposed in [42]. The main goal of the present work is to compare the effects on the classification performances of the three DL analyzed models with and without a prior pre-processing phase of the GPR image dataset through the bi-dimensional Fourier transform, acting as a compressive sensing tool. Compressing information permits reducing data transmission and computational efforts [53], critical aspects for future real-time implementations. The present document is organized as follows. Section 2 briefly describes the image processing with the bi-dimensional Fourier transform technique. Section 3 illustrates the AI-aided tunnel linings investigation methodology with DL-based automatic defects classification. Eventually, section 4 provides the comparative analysis among the various DL-trained models with and without Fourier pre-processing.

2. Image processing with Fourier transform

Within the signal processing field, the discrete Fourier transform (DFT) represents the most acknowledged and widespread tool to investigate real-world propagation phenomena and more [6, 54]. The generality of the Fourier analysis provides the ability to analyze and decompose also higher dimensional signals, and thus any digital image which is actually a discrete ordered spatial bi-dimensional distribution of tensors of pixels [55, 56]. Considering a digital image in the spatial domain A of size $n \times m$ with components a_{rs}, with $0 \leq r \leq n-1, 0 \leq s \leq m-1$, the bi-dimensional discrete Fourier transform (2D-DFT) is a matrix F in the Fourier domain of size $n \times m$ with components [57]:

$$f(k,l) = \sum_{r=0}^{n-1} \sum_{s=0}^{m-1} a(r,s) e^{-2\pi i \left(\frac{kr}{m} + \frac{ls}{n}\right)} \tag{1}$$

where: $0 \leq k \leq n-1, 0 \leq l \leq m-1$. Consequently, the 2D-DFT provides a new representation of the digital image as a double sum of the products of the input spatial image and the sinusoidal basis waveform. The average brightness of the input image is summarized by the DC component $f(0,0)$ in the Fourier

domain [58]. On the other hand, the last realization $f(n-1, m-1)$ corresponds to the highest retrievable frequency component according to the Nyquist-Shannon theorem [58]. The inverse mapping is carried out through the bi-dimensional discrete Fourier transform (2D-IDFT):

$$a(r,s) = \frac{1}{n \cdot m} \sum_{k=0}^{n-1} \sum_{l=0}^{m-1} f(k,l) e^{2\pi i \left(\frac{kr}{m} + \frac{ls}{n}\right)} \tag{2}$$

The outcomes of digital image Fourier analysis are assembled into a complex matrix, whose components are usually expressed in terms of phase ($\phi_{k,l}$) and modulus magnitude ($M_{k,l}$). Since, this latter assumes extremely dispersed values of several orders of magnitude, the following logarithmic manipulation is employed:

$$\tilde{f}(k,l) = c \log(1 + |M_{k,l}|) \tag{3}$$

in which

$$M_{k,l} = \sqrt{\mathrm{Re}(f(k,l))^2 + \mathrm{Im}(f(k,l))^2} \tag{4}$$

The factor c of equation (3) is a scale parameter, set the to unity in the present study. Since in many practical applications, the phase $\phi_{k,l}$ is apparently useless, only the information contained in the magnitude is often retained. However, to guarantee a successful inverse 2D-IDFT mapping, this information is mandatory to avoid a corrupted image [58]. Computing the 2D-DFT as a series of $2 \cdot n$ one-dimensional DFTs considerably helped to save computational effort leading to an overall complexity of $O(N^2)$ [58], being N the number of operations to compute computational complexity [59]. To further improve the convergence speed of discrete bi-dimensional signals Fourier analyses, the efficient fast Fourier transform (2D-FFT) algorithm drastically reduces the computational complexity to $O(N \cdot \log_2(N))$ [58, 60].

The DL models denoted as convolutional neural networks (CNN) are essentially based on convolution, correlation, and in general filtering operations. A thorough understating of these operations within Fourier analysis of digital images revealed to the authors the possible advantages of adopting the bi-dimensional Fourier pre-processing technique. Within the present study, the authors mainly focused on the *convolution theorem*, which states that convolving two functions $h(t) * x(t)$ in the input (time or spatial) domain is a simple product in the Fourier domain [61]:

$$h(t) * x(t) = \int_{-\infty}^{+\infty} x(\tau) h(t-\tau) d\tau \;\Leftrightarrow\; H(\omega) X(\omega) \tag{5}$$

Since the correlation operation is closely related to the convolutional one, a correlation theorem holds [61]:

$$\int_{-\infty}^{+\infty} x(\tau) h(t+\tau) d\tau \;\Leftrightarrow\; H(\omega) X^*(\omega) \tag{6}$$

being X^* the transform complex conjugate of $x(t)$. The convolution operation is employed for image filtering [58], e.g. to detect edges, smoothing operations, etc. Digital filter kernel transfer function $h(r,s)$ correlates with the image $a(r,s)$ on a certain receptive field:

$$g(r,s) = h(r,s) * a(r,s) \tag{7}$$

For the duality property, the convolution operation is substantially a correlation in which the filter mask is rotated with a straight angle, i.e. using a flipped kernel $h(-r,-s)$ [62]. Fundamentally, since the CNNs make extensive use of the discrete convolution operations during the initial feature extraction part, the prior adoption of the bi-dimensional Fourier analysis as an image pre-processing technique may provide a more efficient convolution operation. As a matter of fact, the Fourier domain mapping delivers a synthesized and more compact version of the information contained in the original image, as illustrated in Figure 2. On the contrary, a possible drawback may virtually be excessive information compression, which delivers overly similar images, thus threatening the global accuracy of a data-driven classifier. Moreover, since the Fourier domain enhances the components with the higher frequency content, the Fourier pre-processing method permits actually removing the periodic and non-periodic noise or disturbance patterns [63] in the GPR profiles, which are inherent in the heterogeneous reflectivity properties of the inspected material mean with GPR tool.

2.1. Dataset preparation with and without Fourier pre-processing

The dataset used in the current study is based on a series of NDT campaigns conducted by the authors on several tunnel linings with the GPR device. The data have been collected on tunnels spread through-

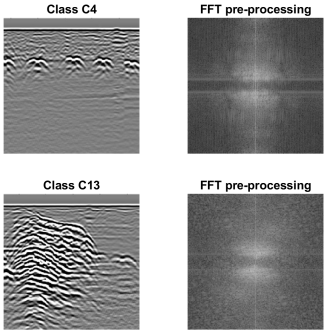

Fig. 2. Resulting magnitude pre-processed images with bi-dimensional Fourier transform of two samples belonging to class C4 (reinforcement bars) and C13 (excavation) respectively.

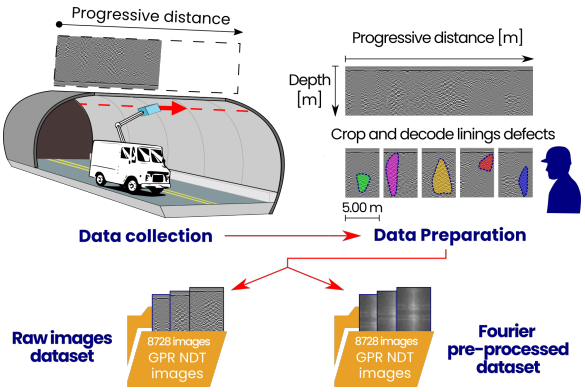

Fig. 3. Data collection, preliminary preparations and the final obtained dataset with and without Fourier pre-processing.

out Italy, whose construction era is between the 1960s and 1980s. To provide a proper dataset to feed a subsequent DL classifier, some basic data preparations were needed after collecting GPR profiles. Firstly, every long output image generated by the GPR testing was interpreted by specialist staff to decode linings defects as the current traditional GPR post-processing workflow [31]. The long images were subsequently cropped with constant pixels step along the abscissa, which represents the progressive distance from the beginning to the end of the tunnel lining profile. This constant pixels step was calibrated in order to provide that each image sample width generally corresponds to about five meters on the real scale length of the tunnel progressive distance. However, in order to avoid some defects that were only placed across the cropping line and consequently end up on different images, the cropping line was occasionally manually adjusted. This latter operation was done on occasion with the minimum invasive intervention, providing a new defect-centered sample image, acting as a sort of local data augmentation. Nevertheless, all the sample images will be subjected to a resizing operation to homogeneously feed the DL models always with the same resolution images. In this way, a total number of 8728 GPR sample images were obtained for the subsequent innovative AI-based paradigm based on DL tunnel lining defects hierarchical classifiers.

Afterward, to further assess the envisaged effects of the bi-dimensional Fourier transform as an image pre-processing tool, the entire dataset of 8728 GPR sample images was pre-processed adopting the 2D-FFT algorithm from the Matlab environment [64]. Specifically, after computing the bi-dimensional FFT as equation (1), the modulus magnitude of each pixel was computed from the resulting complex matrix with the equation (4), followed by the logarithmic transformation exposed in equation (3). Only magnitude information was retained [58], thus producing the final pre-processed reconstructed GPR sample image. Two sample image examples are presented in Figure 2 showing the bi-dimensional Fourier pre-processing effects compared with original GPR raw images. On the left side, two raw GPR sample images illustrate the presence of two different defects evidenced by interpreting the specific pattern, in a similar way to Figure 1. On the right column, the same images undergo to Fourier pre-processing procedure, delivering images of the magnitude of complex terms with the logarithmic manipulation of equation (3). In the following of the present document, for the sake of clearness, whenever the authors refer to the dataset of sample images without any Fourier pre-processing, the adjective *raw* will be explicitly stated, e.g. raw dataset, raw images, etc.

3. Methodology and neural models description

The previously described datasets of GPR sample images with and without bi-dimensional Fourier pre-processing have been classified by adopting three different DL models, briefly described in the current section. As summarized in Figure 4, each dataset of 8728 images in total has been rearranged in a series of 14 folders in order to construct a classification tree composed of six main levels, noting that the total number of available samples gradually decreases from level 1 to level 6. To accurately classify every single defect, this procedure is based on a cascade sequence of binary classifications to produce both a first skimming division in the first levels between healthy and dam-

aged samples, whilst accurately classifying the typology (class) of the identified defect in the other next levels. Specifically, binary classification in level 1 distinguishes between class C1, i.e. healthy samples (4130 images), and class C2, i.e. damaged samples (4598 images). Level 2 is subdivided into levels 2a and level 2b. Level 2a is devoted to categorizing between class C3, i.e. healthy samples without reinforcement bars (3638 images), and class C4, i.e. samples with the presence of reinforcement bars (492 images). Level 2b is devoted to categorizing between class C5, i.e. samples with generic possible warning mix (574 images), and class C6, i.e. samples with more specific warnings which can be further accurately categorized (4024 images). In particular, class C6 contains specific patterns that permit further automatic classifying into specific defect typologies typical in tunnel linings assessment, as evidenced in Figure 4. This means that samples in class C6 may be later categorized as cracks, or anomalies, simple voids, excavations, and detachments. On the other hand, samples belonging to class C5 may contain multiple overlayed defects or other specific patterns that are not directly interpretable with respect to the above-mentioned standard tunnel lining defects. In those cases, the current GPR approach produces warnings that require special care from the tunnel managers. Consequently, the inspectors have to further improve the investigation level to identify which kind of defect, or a mix of defects, is occurring in those critical areas, e.g. providing in situ direct testing or other indirect testing inspections. Binary classification in level 3 distinguishes between class C7, i.e. samples with linings crack presence (900 images), and class C8, i.e. samples with other types of damage (3124 images). Level 4 is devoted to categorizing between class C9, i.e. samples with the presence of anomalies in linings (936 images), and class C10, i.e. samples with other types of defects (2188 images). Binary classification in level 5 distinguishes between class C11, i.e. samples with a simple void in the linings (1108), and class C12, i.e. samples with other types of voids (1080 images). Eventually, level 6 is devoted to categorizing between class C13, i.e. samples with excavation defect (408 images), and class C14, i.e. samples with detachment between the linings and surrounding ground (672 images). For the adopted convolutional models, a balanced training approach was forced by the class with the minimum number of samples. To avoid a biased training of the CNNs toward the class with a higher number of samples, the training set of that class was forced to a smaller set. The size of this set was defined

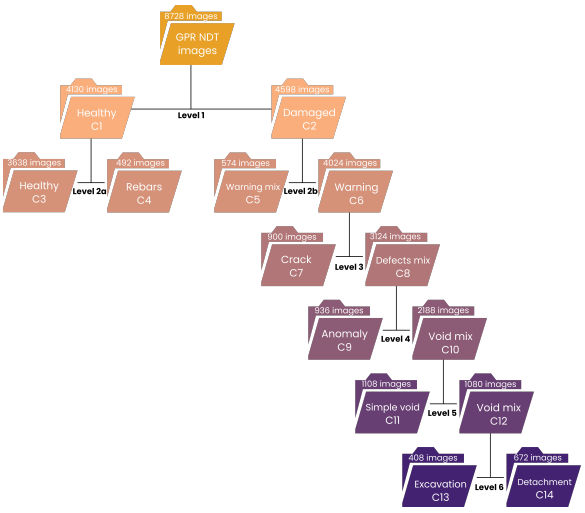

Fig. 4. Dataset folder organization both for raw images and Fourier pre-processing ones representing the adopted hierarchical multi-level classification tree.

according to the number of samples of the class with the minimum data size. This was done to guarantee fair training for the classification model, avoiding a biased classification due to the unbalanced number of images considered at every single level.

3.1. ResNet-50

The CNNs are essentially based on the convolution operation to provide an automatic hierarchical feature extraction procedure [65]. Depicted in Figure 5, the ResNet-50 model [66] is based on a deep residual learning process that relies on identity mapping, i.e. skip or shortcut connections throughout the convolutional layer blocks. The positive impact of these shortcut paths is to improve training speed and avoid vanishing gradients [67], mitigating excessive network depth issues [65]. In the present workflow, the dataset of GPR sample images have been priory resized to a resolution of 224x224 pixels [42] and, subsequent to the input, they have processed from five convolutional stages [66–68] also acknowledged as bottleneck blocks [69]. In the first stage, a first convolution layer is followed by batch normalization, activation with the rectified linear unit (ReLU), and max pooling layers. The subsequent stages are arranged in convolutional blocks, i.e. sequences of three convolutional layers on one branch and a residual connection on the other branch joined in a final step beforehand the ReLU activation. Specifically, at the second stage, a first convolution before the residual connection is necessary to en-

sure that the first identity map correctly fits proper tensor dimensions for the adding layer before the ReLU activation. Eventually, the head of the CNN is composed of an average pooling, followed by a flattening layer and a final fully connected layer with a number of units equal to the number of output classes. The last softmax layer converts the numerical output into probabilities belonging to a certain class. It is worth noting that the number 50 of the ResNet model's name represents the total number of convolutional layers jointly to the final fully connected layer. The originally proposed model in literature was arranged with 152 layers [66], however [70] demonstrated that limiting the total number of layers, e.g. to 50, provides beneficial effects in the network learning whilst containing the computational required effort. The current ResNet-50 has been implemented in MATLAB2021a [64], which provided the pre-trained model on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) dataset [71, 72]. The ImageNet pre-trained model represents a starting sub-optimal solution of the training process for the current GPR profile defects classification problem. The model as-is provides general-purpose features which could be useful for different classification applications with 1000 output classes [65]. Starting from this sub-optimal model, the authors modified the final fully connected layer in the CNN head to accomplish a binary classification with 2 output classes and re-trained the model to specialize it for the current application, considering one time the raw GPR images dataset, and another time the bi-dimensional Fourier pre-processed GPR images dataset. A proper definition of hyperparameters is crucial to find the best trade-off between the accuracy levels of the DL classifiers and the computational efficiency. In recent years, different valuable approaches have been proposed for proper hyperparameters tuning, e.g. it is worth mentioning the random search or the grid search in the hyperparameters space, even combined with cross-validation procedures [67, 73]. However, the manual trial-and-error tuning still represents an extensively adopted method for engineering purposes and, sometimes, it is the only presumably possible path because of a prohibitively computational cost for a consistently refined search [74]. In the current study, for all the DL-adopted models, the authors employed an empirical trial-and-error approach to achieving the best hyperparameters set reported in Figure 5 to reach the best found classification results.

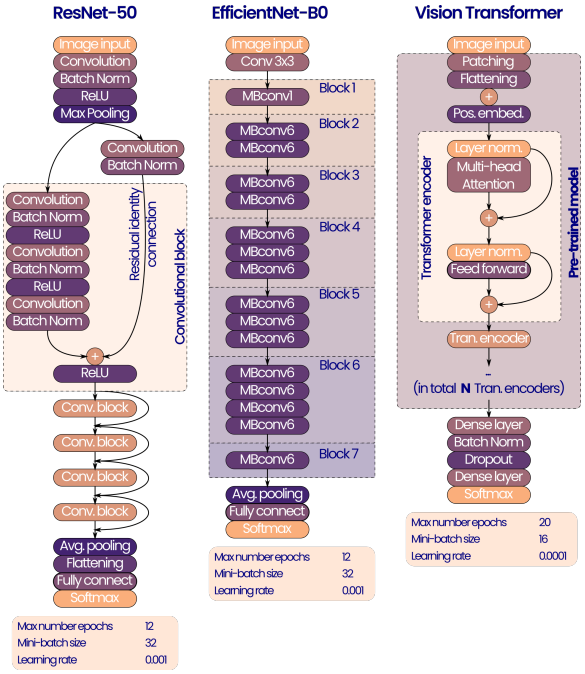

Fig. 5. Graphical illustrative representation of the neural models with hyperparameters adopted in the present study.

3.2. *EfficientNet-B0*

For the sake of comparisons, the authors adopted the contemporary convolutional state-of-art EfficientNet. Presented in 2019 [75], it effectively incorporates multiple techniques and previous existing strategies in an innovative way. A still ongoing widespread methodology to achieve the best accuracy results and contain the required computational effort in CNN is the depth network scaling, i.e. varying the number of layers. The base model ResNet-152 was developed with 152 layers [66], however [70] demonstrated that limiting the total number of layers, e.g. to 50 (ResNet-50), provides comprehensive beneficial effects both in terms of accuracy and computational effort [76]. On the contrary, scaling up CNN models permit enlarging the receptive field [75]. Alternative scaling approaches can be found in [75, 77, 78]. [75] developed a uniform and balanced scaling aiming to optimize the computational effort in terms of floating-point operations per second (FLOPS), thus providing the EfficientNet family models. For the current tunnel defects classification, the authors adopted the base model EfficientNet-B0 [75] provided in MATLAB2021a environment [64]. As illustrated in Figure 5, this implementation relies on 7 building blocks

which employs the inverted residual blocks of MobileNetV2 [75, 79], resulting in a less connected network than ResNet models. Indeed, the residual shortcuts connect only those layers in which the number of inputs and outputs are the same [80]. For the record, MobileNet denotes smaller and more efficient neural models initially developed specifically for the limited resources of mobile hardware [81]. Their efficiency lies in the depthwise separable convolutions operation also acknowledged as spatial-separable convolution, denoted as MBconv in Figure 5, which effectively parallelizes the convolution computing exploiting the three-channel colors (RGB), i.e. the tensor depth, of image data. A deeper insight into the MBconv1 and MBconv6 modules is detailed in [82]. Furthermore, the EfficientNet building blocks adopts the swish activation function, an improved ReLU which is also slightly negative around zero [80, 83], in combination with squeeze-and-excitation block units [80, 84]. Figure 5 illustrates the empirical trial-and-error hyperparameters set adopted to train the current EfficientNet-B0 model.

3.3. Vision transformer

To address the tunnel defects classification problem, the authors also focused on the neural transformers. Firstly presented in [85] for natural language processing (NLP) tasks, they represent a major breakthrough in the DL field with a completely different structure from the CNNs. Transformers are encoder-decoder structures that completely entrust to self-attention and multi-head attention mechanisms, without requiring convolutional layers, and adopting positional embedding to account for token positions [73]. Attention bestows the network the ability to focus on specific parts of the input embedding [85]. The multi-head attention leverages the self-attention to parallel process each embedded sequence input token and concatenates the heads outcomes with a projection layer in order to compute the scored output [73]. [86] analyzed the relationship between the convolution operation and the self-attention mechanism, evidencing the ability of this latter to capture even long-range relationships in the sequence, whereas the foremost is mainly limited to its receptive field. Recent developments have fostered the adoption of the sole encoder part of transformers [87], thus the authors in [88] proposed the Vision Transformer (ViT) to deal with image-data type. In the current study, the ViT large model with 307M parameters and 16 patches (ViT-L16) has been employed. To properly feed the transformer encoder, each input image resized to 224x224 pixels has been subdivided into 16 ordered patches of 14x14 pixels. Each patch is vectorized into a single vector of total length corresponding to the three dimensions product of the tensor patch. A trainable dense layer with linear activation and shared parameters converts the 16 vectors to embedded representations, which are subsequently flattened through a linear projection matrix [89]. To preserve the location of each patch in the original image, the positional embedding may be added element-wise to the flattened representation [85]. Subsequently, the transformer encoders with multi-head attention with 16 heads are fed and this transformer encoder block is repeated $N = 16$ times in the ViT-L16 model. As illustrated in Figure 5, each transformer encoder block employs both layer normalization [90] and residual connections. Specifically, the current adopted implementation relies on the pre-trained ViT-L16 python model based on the public ImageNet-21k [87, 91], in which the head of the network has been replaced with a new dense layer of 2048 units, batch normalization layer, dropout with a probability of 0.5 and a final dense layer. A final softmax layer delivers the classification probability to belong to a certain class, based on a *class* token similarly to [87]. Since ViT training is significantly computationally expensive, the authors decided on a transfer learning solution [37, 89], by adopting a fine-tuning approach of the network's head to accomplish the GPR tunnel defects classification task whilst freezing the training of the rest of the ViT model. Figure 5 illustrates the empirical trial-and-error hyperparameters set adopted to train the current ViT-L16 model.

4. Results and Discussion

In order to investigate and compare the Fourier preprocessing effects on DL-based classification for indirect tunnel monitoring, the three previously described DL models have been trained with both the datasets illustrated in section 2.1, i.e. with raw GPR sample images and with bi-dimensional Fourier GPR sample preprocessed images. In the following, the obtained results are extensively discussed for each DL model individually and, in the final part, the closing section 4.4 argues the results across the various employed techniques.

Table 1
Confusion matrices and classification metrics for ResNet-50 model trained with raw image data.

Level 1	Predicted		**Accuracy**	**92.60%**	
True	C1	C2	Precision	Recall	f1-score
C1	93.30%	6.70%	92.01%	93.30%	92.65%
C2	8.10%	91.90%	93.20%	91.90%	92.55%
Level 2a	Predicted		**Accuracy**	**97.25%**	
True	C3	C4	Precision	Recall	f1-score
C3	98.40%	1.60%	96.19%	98.40%	97.28%
C4	3.90%	96.10%	98.36%	96.10%	97.22%
Level 2b	Predicted		**Accuracy**	**90.40%**	
True	C5	C6	Precision	Recall	f1-score
C5	90.90%	9.10%	90.00%	90.90%	90.45%
C6	10.10%	89.90%	90.81%	89.90%	90.35%
Level 3	Predicted		**Accuracy**	**95.90%**	
True	C7	C8	Precision	Recall	f1-score
C7	92.70%	7.30%	99.04%	92.70%	95.76%
C8	0.90%	99.10%	93.14%	99.10%	96.03%
Level 4	Predicted		**Accuracy**	**91.80%**	
True	C9	C10	Precision	Recall	f1-score
C9	94.90%	5.10%	89.36%	94.90%	92.05%
C10	11.30%	88.70%	94.56%	88.70%	91.54%
Level 5	Predicted		**Accuracy**	**98.30%**	
True	C11	C12	Precision	Recall	f1-score
C11	98.80%	1.20%	97.82%	98.80%	98.31%
C12	2.20%	97.80%	98.79%	97.80%	98.29%
Level 6	Predicted		**Accuracy**	**95.35%**	
True	C13	C14	Precision	Recall	f1-score
C13	96.60%	3.40%	94.24%	96.60%	95.41%
C14	5.90%	94.10%	96.51%	94.10%	95.29%

Table 2
Confusion matrices and classification metrics for ResNet-50 model trained with bi-dimensional Fourier pre-processed image data.

Level 1	Predicted		**Accuracy**	**88.25%**	
True	C1	C2	Precision	Recall	f1-score
C1	87.90%	12.10%	88.52%	87.90%	88.21%
C2	11.40%	88.60%	87.98%	88.60%	88.29%
Level 2a	Predicted		**Accuracy**	**83.15%**	
True	C3	C4	Precision	Recall	f1-score
C3	79.30%	20.70%	85.92%	79.30%	82.48%
C4	13.00%	87.00%	80.78%	87.00%	83.77%
Level 2b	Predicted		**Accuracy**	**76.30%**	
True	C5	C6	Precision	Recall	f1-score
C5	73.50%	26.50%	77.86%	73.50%	75.62%
C6	20.90%	79.10%	74.91%	79.10%	76.95%
Level 3	Predicted		**Accuracy**	**85.90%**	
True	C7	C8	Precision	Recall	f1-score
C7	97.80%	22.00%	91.57%	81.64%	86.32%
C8	9.00%	91.00%	80.53%	91.00%	85.45%
Level 4	Predicted		**Accuracy**	**85.15%**	
True	C9	C10	Precision	Recall	f1-score
C9	83.90%	16.10%	86.05%	83.90%	84.96%
C10	13.60%	86.40%	84.29%	86.40%	85.33%
Level 5	Predicted		**Accuracy**	**89.90%**	
True	C11	C12	Precision	Recall	f1-score
C11	85.70%	14.30%	93.56%	85.70%	89.46%
C12	5.90%	94.10%	86.81%	94.10%	90.31%
Level 6	Predicted		**Accuracy**	**90.55%**	
True	C13	C14	Precision	Recall	f1-score
C13	92.40%	7.60%	89.10%	92.40%	90.72%
C14	11.30%	88.70%	92.11%	88.70%	90.37%

4.1. Classification results for ResNet-50

Concerning the ResNet-50 model described in 3.1, the authors have split the dataset with a proportion of 80% for the training set and 20% for the test set. Furthermore, the authors adopted the k-fold cross-validation method with $k = 10$ folds, representing a good choice to avoid both significant variance and biased values according to [92]. Specifically, in [73], the authors recommend a higher value of k when the dataset variance is high, whilst a smaller value for datasets with low variance. In the current study, the entire dataset has been subdivided into k similar parts named folds. Subsequently, every single model at the various levels depicted in Figure 4 has been trained ten times considering always different training sets composed of $k - 1$ folds. A distinct test set has been employed to compute the ten resulting classification results, i.e. the one-fold left out from the various training phases. Finally, these ten results have been averaged for every single level of the tunnels GPR defects classification tree.

Table 1 reports the confusion matrices of the averaged classification results expressed in percentages for the models trained with the raw GPR samples dataset. The table also illustrates the level of overall accuracies and the class metrics precision, recall, and f1-score. It is worth noting that every level has revealed a good accuracy above 90% in all the cases, reaching a peak of 98.30% in level 5 and a minimum value of 90.40% in level 2b. Averaging all the levels of accuracies, the ResNet-50 model trained with the raw dataset, i.e. without any Fourier pre-processing, reached a global classification accuracy of 94.51%. On the other hand, Table 2 reports the confusion matrices of the averaged classification results expressed in percentages for the

Table 3
Confusion matrices and classification metrics for EfficientNet model trained with raw image data.

Level 1	Predicted		**Accuracy**	**94.55%**	
True	C1	C2	Precision	Recall	f1-score
C1	95.50%	4.50%	93.73%	95.50%	94.60%
C2	6.39%	93.61%	95.41%	93.61%	94.50%
Level 2a	Predicted		**Accuracy**	**91.07%**	
True	C3	C4	Precision	Recall	f1-score
C3	89.24%	10.76%	92.63%	89.24%	90.91%
C4	7.10%	92.90%	89.62%	92.90%	91.23%
Level 2b	Predicted		**Accuracy**	**81.01%**	
True	C5	C6	Precision	Recall	f1-score
C5	81.71%	18.29%	80.58%	81.71%	81.14%
C6	19.69%	80.31%	81.45%	80.31%	80.87%
Level 3	Predicted		**Accuracy**	**94.94%**	
True	C7	C8	Precision	Recall	f1-score
C7	99.00%	1.00%	91.57%	99.00%	95.14%
C8	9.11%	90.89%	98.91%	90.89%	94.73%
Level 4	Predicted		**Accuracy**	**90.70%**	
True	C9	C10	Precision	Recall	f1-score
C9	88.56%	11.44%	92.52%	88.56%	90.50%
C10	7.16%	92.84%	89.03%	92.84%	90.90%
Level 5	Predicted		**Accuracy**	**93.47%**	
True	C11	C12	Precision	Recall	f1-score
C11	90.65%	9.35%	96.07%	90.65%	93.28%
C12	3.70%	96.30%	91.15%	96.30%	93.65%
Level 6	Predicted		**Accuracy**	**96.08%**	
True	C13	C14	Precision	Recall	f1-score
C13	96.33%	3.67%	95.85%	96.33%	96.09%
C14	4.17%	95.83%	96.31%	95.83%	96.07%

Table 4
Confusion matrices and classification metrics for EfficientNet model trained with bi-dimensional Fourier pre-processed image data.

Level 1	Predicted		**Accuracy**	**87.43%**	
True	C1	C2	Precision	Recall	f1-score
C1	87.65%	12.35%	87.27%	87.65%	87.46%
C2	12.78%	87.22%	87.60%	87.22%	87.41%
Level 2a	Predicted		**Accuracy**	**84.15%**	
True	C3	C4	Precision	Recall	f1-score
C3	82.75%	17.25%	85.14%	82.75%	83.93%
C4	14.44%	85.56%	83.22%	85.56%	84.37%
Level 2b	Predicted		**Accuracy**	**73.87%**	
True	C5	C6	Precision	Recall	f1-score
C5	72.47%	27.53%	74.55%	72.47%	73.50%
C6	24.73%	75.27%	73.22%	75.27%	74.23%
Level 3	Predicted		**Accuracy**	**93.06%**	
True	C7	C8	Precision	Recall	f1-score
C7	98.78%	1.22%	88.63%	98.78%	93.43%
C8	12.67%	87.33%	98.62%	87.33%	92.63%
Level 4	Predicted		**Accuracy**	**0.8215**	
True	C9	C10	Precision	Recall	f1-score
C9	81.95%	18.05%	82.29%	81.95%	82.12%
C10	17.64%	82.36%	82.02%	82.36%	82.19%
Level 5	Predicted		**Accuracy**	**88.66%**	
True	C11	C12	Precision	Recall	f1-score
C11	83.70%	16.30%	92.91%	83.70%	88.07%
C12	6.39%	93.61%	85.17%	93.61%	89.19%
Level 6	Predicted		**Accuracy**	**92.28%**	
True	C13	C14	Precision	Recall	f1-score
C13	94.85%	5.15%	90.21%	94.85%	92.47%
C14	10.29%	89.71%	94.57%	89.71%	92.07%

models trained with the bi-dimensional Fourier pre-processed GPR samples dataset. In this circumstance, level 2b stands out for its worst accuracy value stacked to 76.30%. However, in the other levels, the ResNet-50 has revealed a good accuracy above 85% in virtually all the cases, reaching a peak value of 90.55% in level 6. Averaging all the levels of accuracies, the ResNet-50 model trained with the bi-dimensional Fourier pre-processed dataset reached a global classification accuracy of 85.60%, about 8.91% below the global accuracy of the ResNet-50 model trained with the raw dataset. These results demonstrated that, notwithstanding the envisaged advantages of adopting the Fourier pre-processing technique on the GPR sample images for the convolution operation, the ResNet-50 model is not able to reach the accuracy levels of the previous case, i.e. trained with the raw GPR dataset. Downstream of the obtained results, the authors suppose that the Fourier pre-processing probably introduced an exaggerated information compression, thus providing too similar images with such detrimental effects on the classification accuracy.

In an effort to demonstrate the contingent presence of overfitting during the training phase of all the ResNet-50 trained models with and without the Fourier pre-processed dataset, the convergence curves have been reported in appendix A in Figure A.1. These graphs show the trend of the loss, the accuracy, the validation loss, and the validation accuracy during the training epochs or iterations. Since each level accounts for 10 different trained models because of the k-fold cross-validation procedure, the authors represented the average curves among the 10 considered models. However, for the purpose of not losing the variability information among the ten different models, the shaded area around the average curve rep-

Table 5
Confusion matrices and classification metrics for ViT model trained with raw image data.

Level 1	Predicted		**Accuracy**			**95.42%**		
True	C1	C2	Class	Nr img/class	Test support	Precision	Recall	f1-score
C1	380	21	C1	408	401	95.24%	94.76%	95.00%
C2	19	453	C2	672	472	95.57%	95.97%	95.77%
Level 2a	Predicted		**Accuracy**			**99.03%**		
True	C3	C4	Class	Nr img/class	Test support	Precision	Recall	f1-score
C3	359	0	C3	408	359	98.90%	100.00%	99.45%
C4	4	50	C4	672	54	100.00%	92.59%	96.15%
Level 2b	Predicted		**Accuracy**			**94.57%**		
True	C5	C6	Class	Nr img/class	Test support	Precision	Recall	f1-score
C5	45	11	C5	408	56	76.27%	80.36%	78.26%
C6	14	390	C6	672	404	97.26%	96.53%	96.89%
Level 3	Predicted		**Accuracy**			**100.00%**		
True	C7	C8	Class	Nr img/class	Test support	Precision	Recall	f1-score
C7	95	0	C7	408	95	100.00%	100.00%	100.00%
C8	0	308	C8	672	308	100.00%	100.00%	100.00%
Level 4	Predicted		**Accuracy**			**99.04%**		
True	C9	C10	Class	Nr img/class	Test support	Precision	Recall	f1-score
C9	94	2	C9	408	96	98.95%	97.92%	98.43%
C10	1	216	C10	672	217	99.08%	99.54%	99.31%
Level 5	Predicted		**Accuracy**			**99.54%**		
True	C11	C12	Class	Nr img/class	Test support	Precision	Recall	f1-score
C11	115	0	C11	408	115	99.14%	100.00%	99.57%
C12	1	103	C12	672	104	100.00%	99.04%	99.52%
Level 6	Predicted		**Accuracy**			**99.07%**		
True	C13	C14	Class	Nr img/class	Test support	Precision	Recall	f1-score
C13	52	1	C13	408	53	100.00%	98.11%	99.05%
C14	0	55	C14	672	55	98.21%	100.00%	99.10%

resents the envelope among the maximum and minimum curves among the 10 considered models. Excluding level 1 in which a slightly increasing trend of the average validation loss manifests around iteration 400, the ResNet-50 with raw dataset presents a comprehensive excellent behavior without any evidence of overfitting issues. Concerning the convergence curves of the ResNet-50 model with Fourier pre-processed GPR images dataset, a noticeable overfitting problem is evidenced in the level 2b from iteration around 50, thus explaining the poor classification accuracy of that level, as illustrated in table 2. Moreover, slightly overfitting phenomena are tangible in levels 1 from iteration around 400 and level 4 from iteration around 80.

4.2. Classification results for EfficientNet-B0

Regarding the EfficientNet-B0 model described in 3.2, similarly to before, the authors have split the dataset with a proportion of 80% for the training set and 20% for the test set. In a similar manner, the authors adopted the k-fold cross-validation method also for this convolutional model with $k = 10$ folds. Table 3 reports the confusion matrices of the averaged classification results expressed in percentages of the EfficientNet-B0 models trained with the raw GPR samples dataset for each binary classification level of Figure 4. As before, the table also illustrates the level of overall accuracies and the class metrics precision, recall, and f1-score. It is worth noting that every level has revealed a fairly good accuracy above 90% in virtually all the cases, except for level 2b in which the worst value of 81.01% is reported. Level 2b was likewise observed with the lowest accuracy also for The ResNet-50 model. On the contrary, the best accuracy of 96.08% was obtained in level 6. Averaging all the levels of accuracies, the EfficientNet-B0 model trained

Table 6
Confusion matrices and classification metrics for ViT model trained with bi-dimensional Fourier pre-processed image data.

Level 1	Predicted		**Accuracy**			**86.14%**		
True	C1	C2	Class	Nr img/class	Test support	Precision	Recall	f1-score
C1	302	99	C1	408	401	93.21%	75.31%	83.31%
C2	22	450	C2	672	472	81.97%	95.34%	88.15%
Level 2a	Predicted		**Accuracy**			**92.98%**		
True	C3	C4	Class	Nr img/class	Test support	Precision	Recall	f1-score
C3	358	1	C3	408	359	92.75%	99.72%	96.11%
C4	28	26	C4	672	54	96.30%	48.15%	64.20%
Level 2b	Predicted		**Accuracy**			**90.87%**		
True	C5	C6	Class	Nr img/class	Test support	Precision	Recall	f1-score
C5	29	27	C5	408	56	65.91%	51.79%	58.00%
C6	15	389	C6	672	404	93.51%	96.29%	94.88%
Level 3	Predicted		**Accuracy**			**98.76%**		
True	C7	C8	Class	Nr img/class	Test support	Precision	Recall	f1-score
C7	95	0	C7	408	95	95.00%	100.00%	97.44%
C8	5	303	C8	672	308	100.00%	98.38%	99.18%
Level 4	Predicted		**Accuracy**			**94.57%**		
True	C9	C10	Class	Nr img/class	Test support	Precision	Recall	f1-score
C9	85	11	C9	408	96	93.41%	88.54%	90.91%
C10	6	211	C10	672	217	95.05%	97.24%	96.13%
Level 5	Predicted		**Accuracy**			**93.15%**		
True	C11	C12	Class	Nr img/class	Test support	Precision	Recall	f1-score
C11	103	12	C11	408	115	97.17%	89.57%	93.21%
C12	3	101	C12	672	104	89.38%	97.12%	93.09%
Level 6	Predicted		**Accuracy**			**99.07%**		
True	C13	C14	Class	Nr img/class	Test support	Precision	Recall	f1-score
C13	52	1	C13	408	53	100.00%	98.11%	99.05%
C14	0	55	C14	672	55	98.21%	100.00%	99.10%

with the raw dataset, i.e. without any Fourier pre-processing, reached a global classification accuracy of 91.69%.

Conversely, Table 4 reports the confusion matrices of the averaged classification results expressed in percentages for the EfficientNet-B0 models trained with the bi-dimensional Fourier pre-processed GPR samples dataset. In the present case, level 2b pointed out, once again, the worst accuracy value stacked to 73.87%, i.e. 7.14% below than the counterpart EfficientNet-B0 trained with the raw dataset. However, in the other levels, the EfficientNet-B0 has revealed a good accuracy above 80% in virtually all the cases, except for level 2b, with an average reduction of 5.75% with respect to the counterpart EfficientNet-B0 trained with the raw dataset. The maximum accuracy value of 93.06% was realized in level 3. Averaging all the levels of accuracies, the EfficientNet-B0 model trained with the bi-dimensional Fourier pre-processed dataset reached a global classification accuracy of 85.94%, about 5.75% below the global accuracy of the same models trained with the raw dataset. Even in these circumstances, the obtained results proved that the bi-dimensional Fourier pre-processing provided detrimental effects in terms of classification accuracy. Both ResNet-50 and EfficientNet-B0 models exhibit a worse classification behavior with the bi-dimensional Fourier pre-processed dataset despite the envisaged beneficial effects in computing the convolution operation.

To demonstrate any potential presence of overfitting during the training phase of all the EfficientNet-B0 trained models with and without the Fourier pre-processed dataset, the convergence curves during the training iterations have been reported in appendix A in Figure A.2. Although the EfficientNet-B0 models trained with raw dataset apparently do not manifest any sign of overfitting issue presence, level 2b revealed a barely noticeable slightly increasing trend

of the average validation loss manifests around iteration 100. Concerning the convergence curves of the EfficientNet-B0 model with the Fourier pre-processed GPR images dataset, slightly overfitting issues are evidenced in level 1 from iteration around 400, in level 2b from iteration around 80, and in level 4 from iteration around 150.

4.3. Classification results for ViT

Concerning the ViT model described in 3.3, on this occasion, the authors have split the dataset with a proportion of 90% for the training set and 10% for the test set. Furthermore, due to the quite prohibitive computational costs for training the ViT model from scratch, the authors adopted a pre-trained model and provided the fine-tuning training of the head of the network only, as illustrated in Figure 5. For the same reason of computational demanding resources, the k-fold cross-validation method has not been employed with the transformers models of the present study. Table 5 reports the confusion matrices of the averaged classification results expressed in absolute terms, i.e. the number of samples from the test set of the raw GPR samples dataset which has been predicted for each class. The table illustrates the level of overall accuracies and the class metrics precision, recall, and f1-score. It is worth noting that every level has revealed excellent accuracy results above 94% in all the cases, even reaching a peak value of 100.00% in level 3 and with a minimum accuracy value of 95.42% in correspondence of level 2b, just like the worst levels of the above-mentioned convolutional models. Averaging all the levels of accuracies, the ViT model trained with the raw dataset, i.e. without any Fourier pre-processing, reached a global classification accuracy of 98.10%. On the other hand, Table 6 reports the confusion matrices of the averaged classification results expressed in percentages for the ViT models trained with the bi-dimensional Fourier pre-processed GPR samples dataset. In this case, the worst level is the first one, presenting the worst accuracy value of 86.14%. In the other levels, the ViT has still revealed a good accuracy greater than 90% in virtually all the cases nonetheless, still reaching a noticeable maximum accuracy value of 99.07% in level 6. However, averaging all the levels of accuracies, the ViT model trained with the bi-dimensional Fourier pre-processed dataset reached a less global classification accuracy of 93.65%, with an average reduction of 4.45% with respect to the counterpart ViT trained with the raw dataset. Again, the above-mentioned results

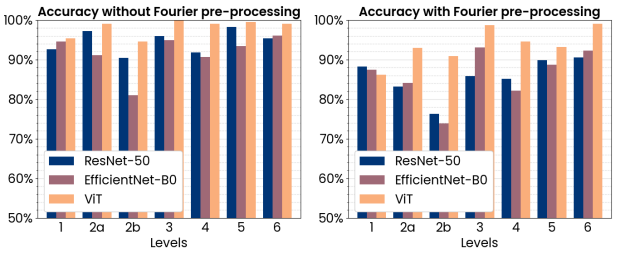

Fig. 6. Comparative analysis of the various DL models' classification accuracy with and without Fourier pre-processing among the classification levels.

demonstrated that, notwithstanding the envisaged advantages of adopting the Fourier pre-processing technique on the GPR sample images, also the ViT model is not able to reach the accuracy levels of the training with the raw GPR dataset. Since ViT is not essentially based on the convolution operation likewise CNNs, the obtained results strengthen the authors' suppositions of an excessive information compression produced with the Fourier pre-processing procedure, resulting in fairly deleterious effects on the classification capacity of the analyzed DL models.

For the purpose of demonstrating a possible presence of overfitting during the training phase of all the ViT trained models with and without the Fourier pre-processed dataset, the convergence curves have been reported in appendix A in Figure A.3. These graphs show the trend of the loss, the accuracy, the validation loss, and the validation accuracy during the training epochs. The convergence curves do not always reach the maximum of 20 epochs because of the adoption of the early-stopping criterion. This means that the training phase is early interrupted when no further improvements occur to both save computational resources and avoid overfitting training. Despite the validation loss curves appearing quite noisy during the training epochs, their global descending trends proved that ViT model trained with raw GPR images dataset does not incur any overfitting phenomena at every level. Focusing on the ViT models trained with the Fourier pre-processed dataset, the validation curve trends revealed overfitting occurrence in level 1, level 4, and slight evidence in level 3, besides they appeared to be noisier than the previous case.

4.4. Comparative analysis of the classification results

In the current closing section, the authors compared the results among the various DL trained mod-

Table 7
Global average accuracy for the three analyzed neural models.

Neural model	Without Fourier	With Fourier
ResNet-50	94.51%	85.60%
EfficientNet-B0	91.69%	85.94%
ViT	98.10%	93.65%

els. Figure 6 provides a comparative overview of the obtained accuracy results. The classification outcomes have been organized for the various GPR defects classification levels. The graph is arranged according to the three DL analyzed models, and depicted in two juxtaposed histogram representations related to the training phase with the raw dataset and with the bi-dimensional Fourier pre-processed dataset. At first sight of the diagram, among the various DL models, the ViT architecture delivered the highest accuracy values for virtually all the levels of both cases with and without Fourier pre-processing. However, the ResNet-50 provided an accuracy result of 88.25% with the Fourier pre-processed dataset, thus providing a higher result than ViT model. As evidenced from the convergence curves, the ViT trained with Fourier pre-processed images evidence a slightly overfitting phenomenon in level 1. Jointly with the excessive data compression of the Fourier operation, as visually demonstrated in Figure 2, the ViT model produced the worst accuracy performance in level 1 with Fourier pre-processing concerning other models. The EfficientNet-B0 model globally produced the worst results among almost all the levels for both two cases under comparison. However, with a deeper insight, the ResNet-50 provided the worst results in level 1 focusing on raw images dataset, and in levels 2a and 6 within the Fourier pre-processed case. It is worth mentioning that generally all three DL models struggled to reach high accuracy vale in level 2b. With a deeper inspection of the various convergence curves reported in the appendix, overfitting issues emerged in ResNet-50 with Fourier pre-processed dataset, in EfficientNet-B0 in both the two analyzed cases, and in the ViT model with Fourier pre-processed dataset. The difficulties in level 2b may be related to the critical unbalance in the amount of GPR images samples between classes C5 and C6. It is worth recalling that samples belonging to class C5 may contain multiple overlayed defects or other specific patterns that are not directly interpretable with respect to the standard tunnel lining defects classification of Figure 4. This means that it is not possible to priorly exclude that those special patterns sometimes could present some parts quite similar to specific defects patterns belonging to class C6. Therefore, it could be also reasonable that those parts in samples of class C5 may possibly mislead the neural models, thus providing misclassified results. In addition, another possible reason could also be a quite critical similarity degree among the images of these two specific classes C5 and C6. This may be plausible especially in the Fourier case, which may produce overly similar images due to excessive data information compression.

Eventually, Table 7 reports the global average accuracy results among the various levels. It is worth noting the average accuracy reductions for the three DL models between the raw dataset case and Fourier pre-processed dataset. The ViT model recorded the lowest average accuracy reduction equal to 4.45%, whereas the EfficientNet-B0 exhibited a reduction value of 5.75%. The highest reduction of 8.91% was suffered from the ResNet-50 model. Despite the second-best model in terms of accuracy is the ResNet-50 with the raw dataset, it appeared to be the least robust architecture to the induced effects of the Fourier pre-processed dataset, thus delivering the most consistent average accuracy reduction.

5. Conclusions

This paper focuses on GPR testing of tunnel linings profiles using a DL-based image recognition framework. The authors compare the performance of three DL models for indirect tunnel defects classification. Nowadays, tunnel monitoring with innovative NDT is widespread, demanding more automation from DL methods. The authors adopted a hierarchical binary classification approach to group the types of defects identified in the GPR profiles. The core and main findings of this paper can be summarized as follows:

- Three DL models have been employed, two convolutional models, i.e. the ResNet-50 model and the EfficientNet-B0 model, and a recent transformer architecture, i.e. the ViT model. The authors trained all the models with two different datasets, adequately prepared to compare the induced effects of a common, widespread image processing technique, i.e. the bi-dimensional Fourier transform.
- The Fourier pre-processing of GPR images determined a significant accuracy reduction compared to the raw dataset. Therefore, despite the computational advantages of Fourier pre-processing,

Fourier pre-processing introduced an exaggerated data compression. The related information loss leads to overly similar images, with detrimental effects on the final classification accuracy.
- The ViT model delivered the highest classification accuracy values for virtually all the levels both with and without Fourier pre-processing.

The current AI-aided approach for GPR indirect tunnel monitoring mainly deals with the defects classification and detection task, which is the first level of an ideal SHM paradigm [93]. Future research efforts will be directed towards the remaining three SHM steps, i.e. the damage localization, the damage severity quantification, and the actual safety health state assessment. The primary purpose of SHM is to provide a reliable and exhaustive diagnosis of existing structures and infrastructures [93]. A promising research path in that direction may naturally leverage the attention map provided by transformer models' outputs. Further future developments to address defects localization may leverage also the potentialities offered by the object detection task, e.g. employing a Faster R-CNN. Future improvements may also involve other compressive sensing techniques and transforms, e.g. wavelet [94–96].

Acknowledgments

Computational resources provided by hpc@polito (http://www.hpc.polito.it)

References

[1] Wang J, Mi Z, Zhang T, Wang D. Durability degradation of tunnel-lining-shotcrete exposed to nitric acid: Neutralization and nitrate ion migration. Construction and Building Materials. 2022;336:127554.
[2] Zhang W, Qiu J, Zhao C. Structural behavior degradation of corroded metro tunnel lining segment. Structure and Infrastructure Engineering. 2022:1-17.
[3] Kapoor NR, Kumar A, Arora HC, Kumar A. Structural Health Monitoring of Existing Building Structures for Creating Green Smart Cities Using Deep Learning. In: Recurrent Neural Networks. CRC Press;. p. 203-32.
[4] Tyagi AK, Abraham A. Recurrent Neural Networks: Concepts and Applications. CRC Press; 2022.
[5] Li Z, Park HS, Adeli H. New method for modal identification of super high-rise building structures using discretized synchrosqueezed wavelet and Hilbert transforms. The Structural Design of Tall and Special Buildings. 2017;26(3):e1312.
[6] Farrar CR, Worden K. Structural health monitoring: a machine learning perspective. John Wiley & Sons; 2012.
[7] Strauss A, Neuner H, Rigler M, Polt M, Seywald C, Kostjak V, et al. Verification of the performance of reinforced concrete profiles of alpine infrastructure systems assisted by innovative monitoring. Copernicus Meetings; 2022.
[8] Jiang Y, Zhang X, Taniguchi T. Quantitative condition inspection and assessment of tunnel lining. Automation in Construction. 2019;102:258-69.
[9] Chiaia B, Marasco G, Ventura G, Zannini Quirini C. Customised active monitoring system for structural control and maintenance optimisation. Journal of Civil Structural Health Monitoring. 2020;10(2):267-82.
[10] Marasco G, Chiaia B, Ventura G. AI based bridge health assessment. 9th International Workshop on Reliable Engineering Computing (REC2021) is "Risk and Uncertainty in Engineering Computations"; 2021. .
[11] Thakur A, et al. A Review on Non-destructive Techniques for Corrosion Monitoring in Reinforced Concrete Structures. Recent Advances in Structural Engineering and Construction Management. 2023:951-68.
[12] Zhu M, Zhu H, Guo F, Chen X, Ju JW. Tunnel condition assessment via cloud model-based random forests and self-training approach. Computer-Aided Civil and Infrastructure Engineering. 2021;36(2):164-79.
[13] Aloisio A, Pasca DP, Battista L, Rosso MM, Cucuzza R, Marano G, et al. 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. 2022 01;37:686-97.
[14] Bhalla S, Yang YW, Zhao J, Soh CK. Structural health monitoring of underground facilities – Technological issues and challenges. Tunnelling and Underground Space Technology. 2005;20(5):487-500.
[15] Dawood T, Zhu Z, Zayed T. Deterioration mapping in subway infrastructure using sensory data of GPR. Tunnelling and Underground Space Technology. 2020;103:103487.
[16] Menz N, Gerasimidis S, Civjan S, Czach J, Rigney J. Review of post-fire inspection procedures for concrete tunnels. Transportation research record. 2021;2675(9):1304-15.
[17] Mohamad FAJ, Rahim RA, Ahmad N, Jamaludin J, Ibrahim S, Rahiman MHF, et al. NDT-Defect Detection on Concrete using Ultrasonic: A Review. Journal of Tomography System & Sensors Application Vol. 2021;4(1).
[18] Chen R, Tran KT, Dinh K, Ferraro CC. Evaluation of Ultrasonic SH-Waveform Tomography for Determining Cover Thickness and Rebar Size in Concrete Structures. Journal of Nondestructive Evaluation. 2022;41(2):1-16.
[19] Geng Q, Ye Y, Wang X. Identifying void defects behind Tunnel composite lining based on transient electromagnetic radar method. NDT & E International. 2022;125:102562.
[20] Behnia A, Chai HK, Shiotani T. Advanced structural health monitoring of concrete structures with the aid of acoustic emission. Construction and Building Materials. 2014;65:282-302.
[21] Hartbower PE, Stolarski PJ. Structural Materials Technology: An NDT Conference (1996). CRC Press; 1996.
[22] Sirca Jr GF, Adeli H. Infrared thermography for detecting defects in concrete structures. Journal of Civil Engineering and Management. 2018;24(7):508-15.
[23] Monsberger CM, Bauer P, Buchmayer F, Lienhart W. Large-scale distributed fiber optic sensing network for short and long-term integrity monitoring of tunnel linings. Journal of Civil Structural Health Monitoring. 2022:1-11.

[24] Mishra M, Lourenço PB, Ramana GV. Structural health monitoring of civil engineering structures by using the internet of things: A review. Journal of Building Engineering. 2022;48:103954.
[25] Qian H. Design of Tunnel Automatic Monitoring System Based on BIM and IOT. In: Journal of Physics: Conference Series. vol. 1982. IOP Publishing; 2021. p. 012073.
[26] Żarski M, Wójcik B, Książek K, Miszczak JA. Finicky transfer learning—A method of pruning convolutional neural networks for cracks classification on edge devices. Computer-Aided Civil and Infrastructure Engineering. 2022;37(4):500-15.
[27] Dwivedi SK, Vishwakarma M, Soni PA. Advances and Researches on Non Destructive Testing: A Review. Materials Today: Proceedings. 2018;5(2, Part 1):3690-8. 7th International Conference of Materials Processing and Characterization, March 17-19, 2017.
[28] Tosti F, Ferrante C. Using ground penetrating radar methods to investigate reinforced concrete structures. Surveys in Geophysics. 2020;41(3):485-530.
[29] Cardarelli E, Marrone C, Orlando L. Evaluation of tunnel stability using integrated geophysical methods. Journal of Applied Geophysics. 2003;52(2):93-102.
[30] Gopalakrishnan K, Ceylan H, Kim S, Yang S. Wireless MEMS for transportation infrastructure health monitoring. In: Wireless MEMS Networks and Applications. Elsevier; 2017. p. 53-76.
[31] Al-Nuaimy W, Huang Y, Nakhkash M, Fang MTC, Nguyen VT, Eriksen A. Automatic detection of buried utilities and solid objects with GPR using neural networks and pattern recognition. Journal of Applied Geophysics. 2000;43(2):157-65.
[32] Lei M, Liu L, Shi C, Tan Y, Lin Y, Wang W. A novel tunnel-lining crack recognition system based on digital image technology. Tunnelling and Underground Space Technology. 2021;108:103724.
[33] Hsieh YA, Yang Z, James Tsai YC. Convolutional neural network for automated classification of jointed plain concrete pavement conditions. Computer-Aided Civil and Infrastructure Engineering. 2021;36(11):1382-97.
[34] Guo J, Wang Q, Li Y. Semi-supervised learning based on convolutional neural network and uncertainty filter for façade defects classification. Computer-Aided Civil and Infrastructure Engineering. 2021;36(3):302-17.
[35] Macias-Garcia E, Galeana-Perez D, Medrano-Hermosillo J, Bayro-Corrochano E. Multi-stage deep learning perception system for mobile robots. Integrated Computer-Aided Engineering. 2021;28(2):191-205.
[36] Dao DV, Adeli H, Ly HB, Le LM, Le VM, Le TT, et al. A sensitivity and robustness analysis of GPR and ANN for high-performance concrete compressive strength prediction using a Monte Carlo simulation. Sustainability. 2020;12(3):830.
[37] Feng C, Zhang H, Wang S, Li Y, Wang H, Yan F. Structural damage detection using deep convolutional neural network and transfer learning. KSCE Journal of Civil Engineering. 2019;23(10):4493-502.
[38] Puntu JM, Chang PY, Lin DJ, Amania HH, Doyoro YG. A Comprehensive Evaluation for the Tunnel Conditions with Ground Penetrating Radar Measurements. Remote Sensing. 2021;13(21):4250.
[39] Chiaia B, Marasco G, Aiello S. Deep convolutional neural network for multi-level non-invasive tunnel lining assessment. Frontiers Of Structural And Civil Engineering. 2022.
[40] Guo L, Cai L, Chen D. Research on tunnel lining image target recognition method based on YOLOv3. In: 2022 IEEE 10th Joint International Information Technology and Artificial Intelligence Conference (ITAIC). vol. 10. IEEE; 2022. p. 1800-5.
[41] Rosso MMR, Marasco GM, Tanzi LT, Aiello SA, Aloisio AA, Cucuzza RC, et al. Advanced deep learning comparisons for non-invasive tunnel lining assessment from ground penetrating radar profiles. In: ECCOMAS Congress 2022-8th European Congress on Computational Methods in Applied Sciences and Engineering; 2022. .
[42] Marasco G, Rosso MM, Aiello S, Aloisio A, Cirrincione G, Chiaia B, et al. Ground Penetrating Radar Fourier Preprocessing for Deep Learning Tunnel Defects' Automated Classification. In: International Conference on Engineering Applications of Neural Networks. Springer; 2022. p. 165-76.
[43] Zhou Z, Zhang J, Gong C. Automatic detection method of tunnel lining multi-defects via an enhanced You Only Look Once network. Computer-Aided Civil and Infrastructure Engineering. 2022;37(6):762-80.
[44] Tong Z, Gao J, Yuan D. Advances of deep learning applications in ground-penetrating radar: A survey. Construction and Building Materials. 2020;258:120371.
[45] Qin H, Zhang D, Tang Y, Wang Y. Automatic recognition of tunnel lining elements from GPR images using deep convolutional networks with data augmentation. Automation in Construction. 2021;130:103830.
[46] Li C, Cai L, Guo L, Chen D. Research on Target Recognition Method of Tunnel Lining Image Based on Deep Learning. In: 2022 IEEE 6th Information Technology and Mechatronics Engineering Conference (ITOEC). vol. 6. IEEE; 2022. p. 1833-7.
[47] FENG Ds, YANG Zl. Automatic recognition of ground penetrating radar image of tunnel lining structure based on deep learning. Progress in Geophysics. 2020;35(4):1552-6.
[48] Wang J, Liu H, Jiang P, Wang Z, Sui Q, Zhang F. GPRI2Net: A Deep-Neural-Network-Based Ground Penetrating Radar Data Inversion and Object Identification Framework for Consecutive and Long Survey Lines. IEEE Transactions on Geoscience and Remote Sensing. 2021;60:1-20.
[49] Zhu A, Chen S, Lu F, Ma C, Zhang F. Recognition Method of Tunnel Lining Defects Based on Deep Learning. Wireless Communications and Mobile Computing. 2021;2021.
[50] Wang J, Zhang J, Cohn AG, Wang Z, Liu H, Kang W, et al. Arbitrarily-oriented tunnel lining defects detection from Ground Penetrating Radar images using deep Convolutional Neural networks. Automation in Construction. 2022;133:104044.
[51] Liu B, Ren Y, Liu H, Xu H, Wang Z, Cohn AG, et al. GPRInvNet: Deep learning-based ground-penetrating radar data inversion for tunnel linings. IEEE Transactions on Geoscience and Remote Sensing. 2021;59(10):8305-25.
[52] Wang Y, Qin H, Tang Y, Zhang D, Wang Z, Pan S. A deep learning network to improve tunnel lining defect identification using ground penetrating radar. In: IOP Conference Series: Earth and Environmental Science. vol. 861. IOP Publishing; 2021. p. 042057.
[53] Bao Y, Tang Z, Li H. Compressive-sensing data reconstruction for structural health monitoring: a machine-learning approach. Structural Health Monitoring. 2020;19(1):293-304.
[54] Canuto C, Tabacco A. Mathematical Analysis II. vol. 85. Springer; 2015.
[55] Lim JS. Two-dimensional signal and image processing. Englewood Cliffs. 1990.

[56] Solomon C, Breckon T. Fundamentals of Digital Image Processing: A practical approach with examples in Matlab. John Wiley & Sons; 2011.
[57] Broughton SA, Bryan K. Discrete Fourier analysis and wavelets: applications to signal and image processing. John Wiley & Sons; 2018.
[58] Fisher R, Perkins S, Walker A, Wolfart E. Hypermedia image processing reference. England: John Wiley & Sons Ltd. 1996:118-30.
[59] Neri F. An introduction to computational complexity. In: Linear Algebra for Computational Sciences and Engineering. Springer; 2019. p. 419-32.
[60] Thompson M. Digital Image Processing by Rafael C. Gonzalez and Paul Wintz. Leonardo. 1981;14(3):256-7.
[61] Brigham EO. The fast Fourier transform and its applications. Prentice-Hall, Inc.; 1988.
[62] Gonzalez RC, Woods RE, Eddins SL. Digital image processing using MATLAB. Pearson Education India; 2004.
[63] Woods RE, Gonzalez RC. Digital image processing third edition; 2021.
[64] Inc. The Mathworks. MATLAB version 9.10.0.1649659 (R2021a) Update 1. Natick, Massachusetts; 2021.
[65] Aggarwal CC, et al. Neural networks and deep learning. Springer. 2018;10:978-3.
[66] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016. p. 770-8.
[67] Géron A. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. " O'Reilly Media, Inc."; 2019.
[68] Rawat W, Wang Z. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. Neural Computation. 2017 09;29(9):2352-449.
[69] Koonce B. ResNet 50. In: Convolutional neural networks with swift for tensorflow. Springer; 2021. p. 63-72.
[70] Zagoruyko S, Komodakis N. Wide residual networks. arXiv preprint arXiv:160507146. 2016.
[71] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. Imagenet large scale visual recognition challenge. International journal of computer vision. 2015;115(3):211-52.
[72] Markoff J. For web images, creating new technology to seek and find. New York Times. 2012.
[73] Ekman M. Learning Deep Learning: Theory and Practice of Neural Networks, Computer Vision, NLP, and Transformers Using TensorFlow. Addison-Wesley Professional; 2021.
[74] Anitescu C, Atroshchenko E, Alajlan N, Rabczuk T. Artificial Neural Network Methods for the Solution of Second Order Boundary Value Problems. Computers, Materials & Continua. 2019;59(1):345-59.
[75] Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In: International conference on machine learning. PMLR; 2019. p. 6105-14.
[76] Bianco S, Cadene R, Celona L, Napoletano P. Benchmark analysis of representative deep neural network architectures. IEEE access. 2018;6:64270-7.
[77] Andrew G, Menglong Z, et al. Efficient convolutional neural networks for mobile vision applications. Mobilenets. 2017.
[78] Huang Y, Cheng Y, Bapna A, Firat O, Chen D, Chen M, et al. Gpipe: Efficient training of giant neural networks using pipeline parallelism. Advances in neural information processing systems. 2019;32.
[79] Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p. 4510-20.
[80] Koonce B. EfficientNet. In: Convolutional neural networks with swift for tensorflow. Springer; 2021. p. 109-23.
[81] Koonce B. Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization. Springer; 2021.
[82] Tan M, Chen B, Pang R, Vasudevan V, Sandler M, Howard A, et al. Mnasnet: Platform-aware neural architecture search for mobile. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2019. p. 2820-8.
[83] Ramachandran P, Zoph B, Le QV. Searching for activation functions. arXiv preprint arXiv:171005941. 2017.
[84] Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p. 7132-41.
[85] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Advances in neural information processing systems. 2017;30.
[86] Cordonnier JB, Loukas A, Jaggi M. Multi-head attention: Collaborate instead of concatenate. arXiv preprint arXiv:200616362. 2020.
[87] Devlin J, Chang MW, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:181004805. 2018.
[88] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In: International Conference on Learning Representations; 2021. Available from: https://openreview.net/forum?id=YicbFdNTTy.
[89] Tanzi L, Audisio A, Cirrincione G, Aprato A, Vezzetti E. Vision Transformer for femur fracture classification. Injury. 2022.
[90] Ba JL, Kiros JR, Hinton GE. Layer normalization. arXiv preprint arXiv:160706450. 2016.
[91] Morales F, et al.. vit-keras, Keras implementation of ViT (Vision Transformer). GitHub; 2015. Available from: https://github.com/faustomorales/vit-keras.
[92] Gareth J, Daniela W, Trevor H, Robert T. An introduction to statistical learning: with applications in R. Spinger; 2013.
[93] Rytter A. Vibrational based inspection of civil engineering structures. Dept. of Building Technology and Structural Engineering, Aalborg University; 1993.
[94] Zhou Z, Adeli H. Time-frequency signal analysis of earthquake records using Mexican hat wavelets. Computer-Aided Civil and Infrastructure Engineering. 2003;18(5):379-89.
[95] Perez-Ramirez CA, Amezquita-Sanchez JP, Adeli H, Valtierra-Rodriguez M, Camarena-Martinez D, Romero-Troncoso RJ. New methodology for modal parameters identification of smart civil structures using ambient vibrations and synchrosqueezed wavelet transform. Engineering Applications of Artificial Intelligence. 2016;48:1-12.
[96] Li Z, Adeli H. New adaptive robust H∞ control of smart structures using synchrosqueezed wavelet transform and recursive least-squares algorithm. Engineering Applications of Artificial Intelligence. 2022;116:105473.

Appendix A. Convergence curves

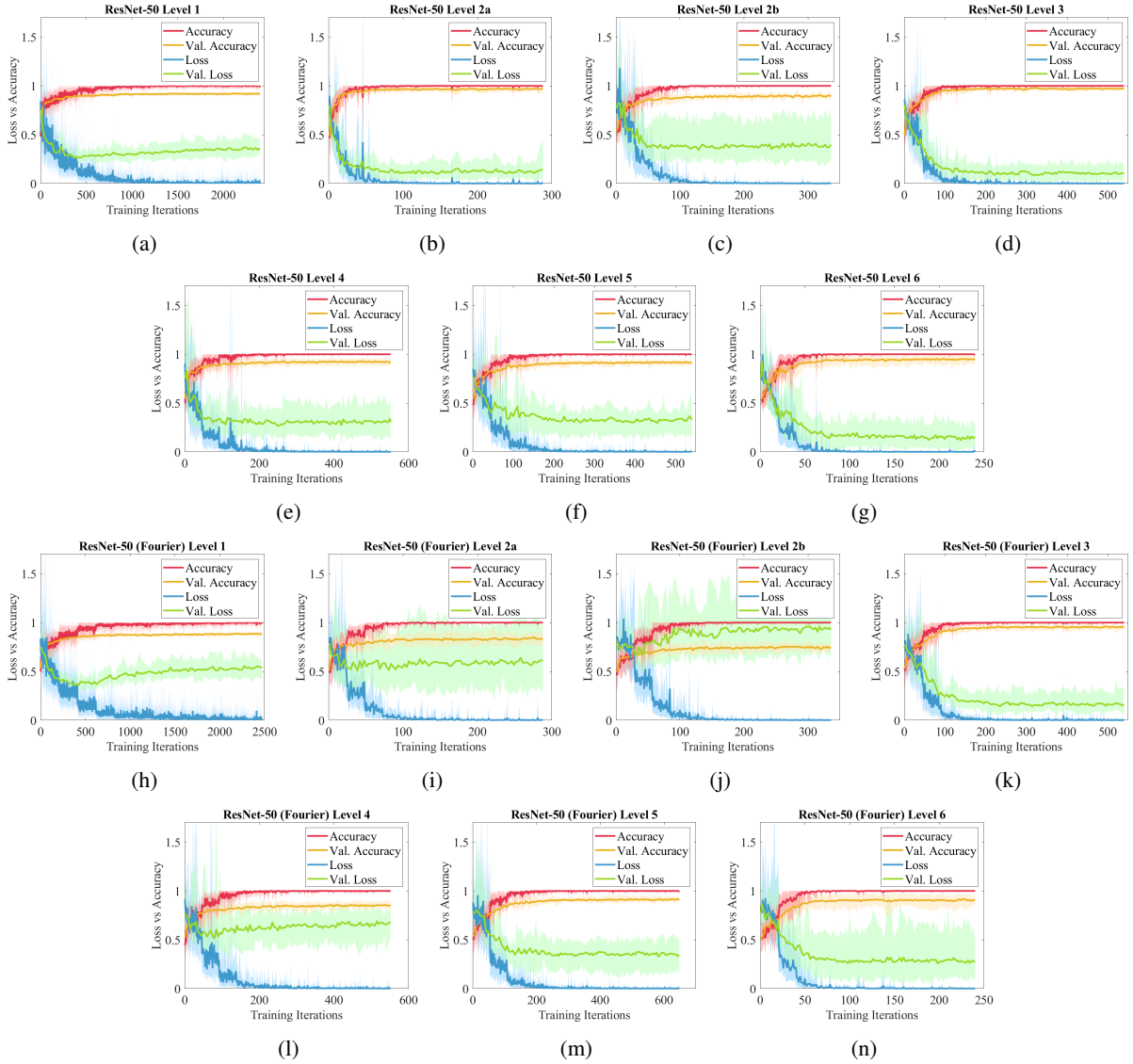

Fig. A.1. Loss versus accuracy during the training iterations. (a-g) ResNet-50 trained with raw images. (h-n) ResNet-50 trained with Fourier pre-processed images.

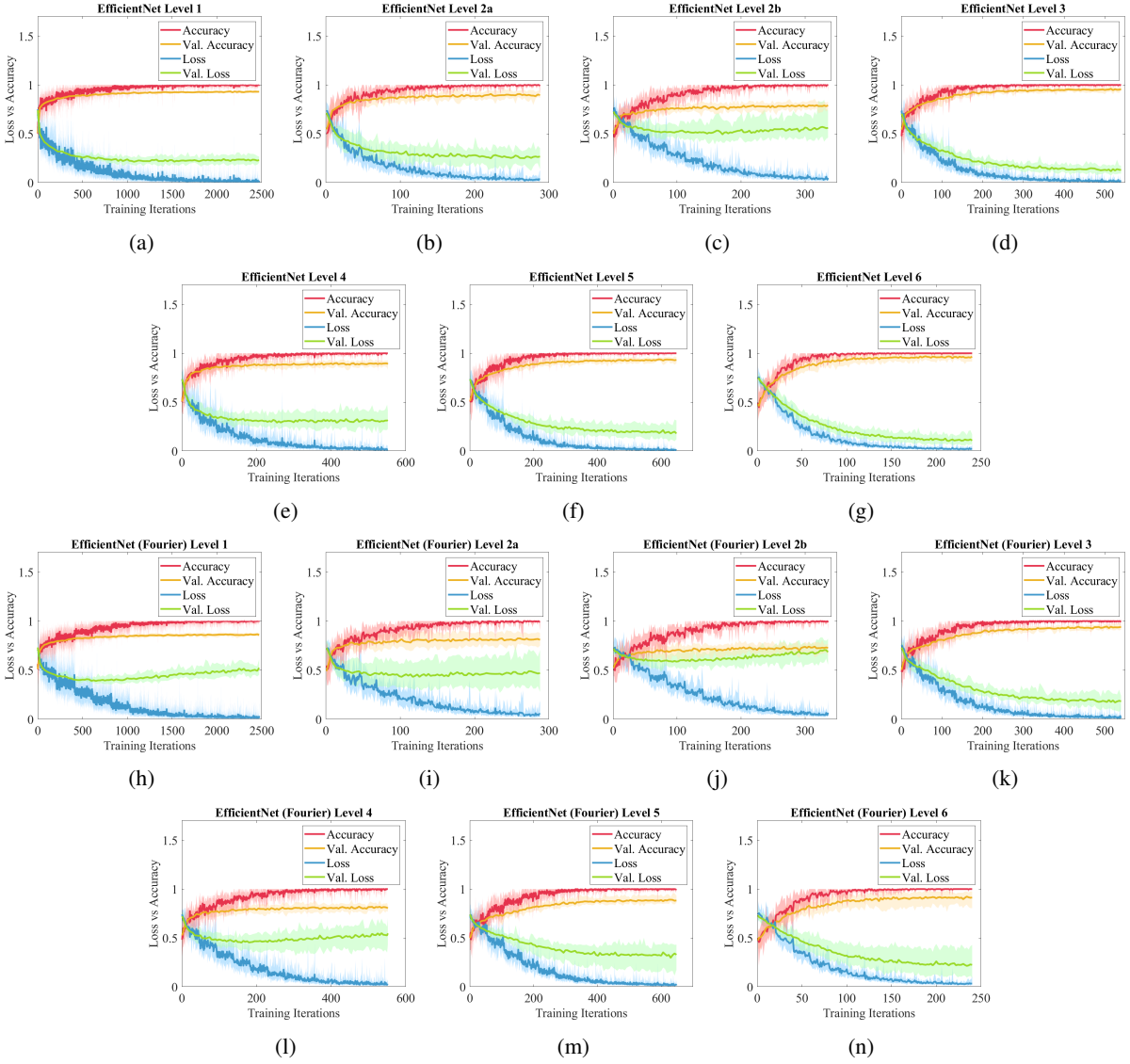

Fig. A.2. Loss versus accuracy during the training iterations. (a-g) EfficientNet-B0 trained with raw images. (h-n) EfficientNet-B0 trained with Fourier pre-processed images.

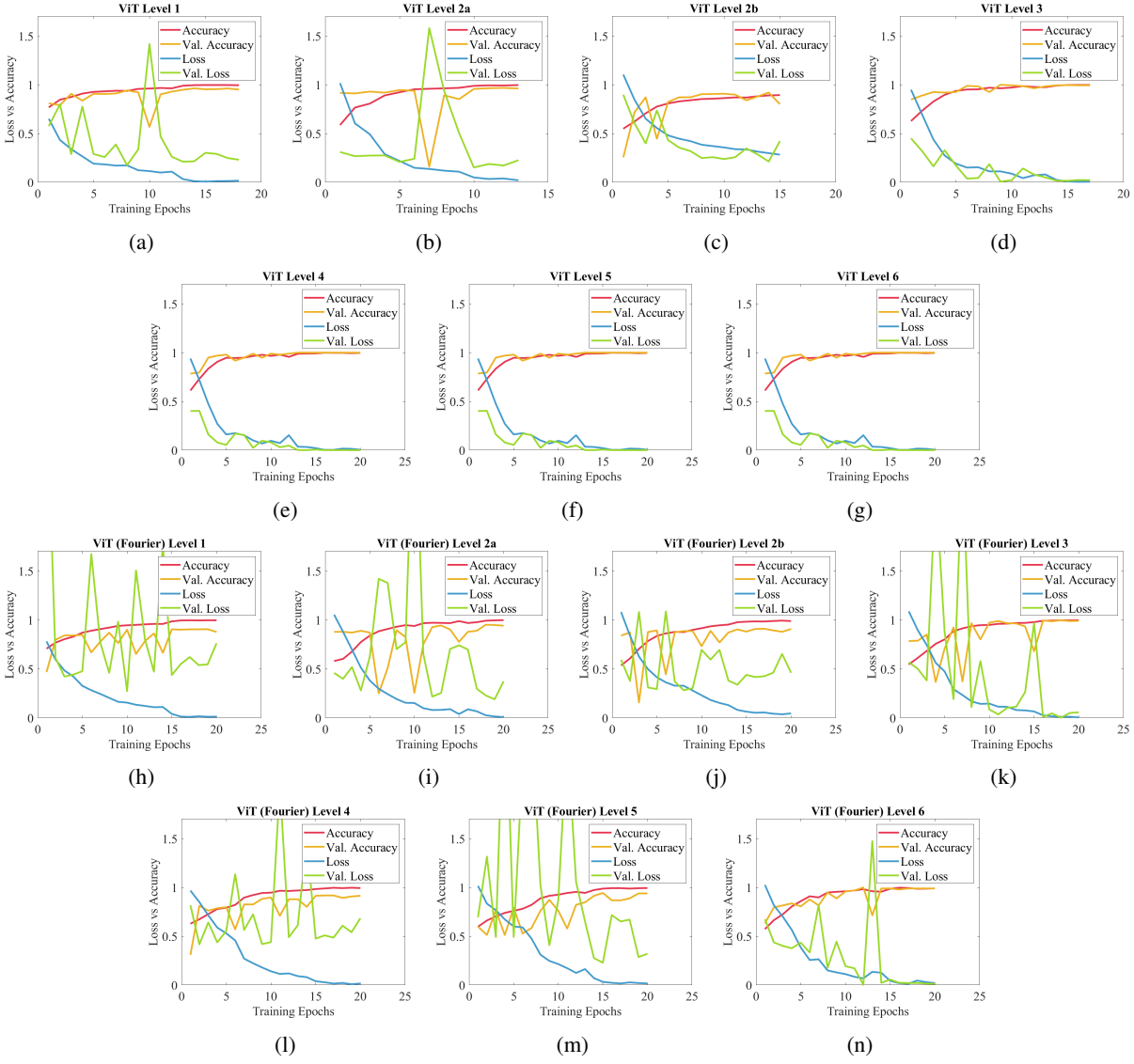

Fig. A.3. Loss versus accuracy during the training iterations. (a-g) ViT trained with raw images. (h-n) ViT trained with Fourier pre-processed images.