Distress Detection in Road Pavements using Neural Networks

Luís Filipe Gomes[1], Cesar Analide[1], and Elisabete Freitas[2]

[1] Centro ALGORITMI, Universidade do Minho
luisgomes24.97@gmail.com; analide@di.uminho.pt,
[2] Centro ISISE, Universidade do Minho
efreitas@civil.uminho.pt

Abstract. Combining Computer Vision (CV) and Anomaly Detection (AD), there is a convergence of methodologies using convolutional layers in AD architectures, which we consider an innovation in the field. The main goal of this work is to present different Artificial Neural Networks (ANN) architectures, applying them to distress detection in road pavements and comparing the results obtained in each approach. The experimented methods for AD in images include a binary classifier as a baseline, an Autoencoder (AE) and a Variational Autoencoder (VAE). Supervised and unsupervised practises are also compared, proving their utility in scenarios where there is no labelled data available. Using the VAE model in a supervised setting, it presents an excellent distinction between good and bad pavement. When labelled data is not available, using the AE model and the distribution of similarities of good pavement reconstructions to calculate the threshold is the best option with accuracy and precision above 94%. The development of these models shows that it is possible to develop an alternative solution to reduce operating costs compared to expensive commercial systems and to improve the usability compared to conventional methods of classifying road surfaces.

Keywords: Artificial Neural Networks, Computer Vision, Anomaly Detection, Autoencoders, Variational Autoencoders, Automatic Pavement Monitoring

1 Introduction

Highways are one of the most important assets in the daily life of modern society, increasing the economic gains of many activity sectors, citizens quality of life, and countries' development, with special impact in urban areas [10].

After the highway construction, the pavement develops distress due to different factors, such as meteorological conditions, materials self-deterioration, and traffic wear. Bad pavement condition impacts the drivers' comfort, road safety [4] and increases travel costs. The accident rate is correlated with the pavement condition, where higher values of roughness and rut depth increase crash rate [15]. In terms of asset management, accurate distress identification is essential. In this context, it is imperative to monitor pavements' condition periodically, as cracking, rutting, reveling, potholes, unevenness. This information is used to deliver condition indicators serving as inputs in road maintenance optimization models to help to select the best maintenance strategy.

The background and related work are presented in Sections 2 and 3, respectively. The experiments are shown in Section 4 and conclusions in Section 5.

2 Background

2.1 Road Pavements Monitoring

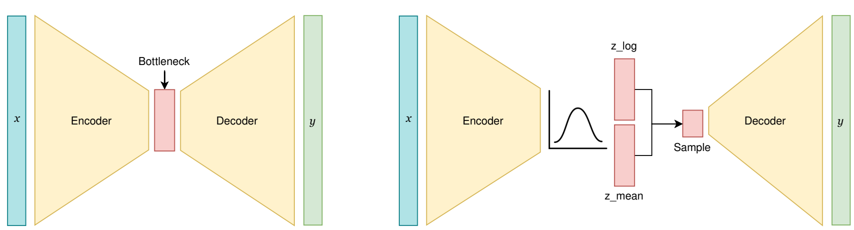
The traditional methods to monitor road pavements include the direct observation of the road, with manual annotations, which is a rudimentary method. The data can be stored in paper format, and it needs to be processed afterwards.

Other approaches use complex systems with 3D image capturing and laser profiling sensors to provide a more detailed report of the pavement condition [5]. These systems require special equipment and trained operators. The results provided by these mechanisms are accurate but represent high cost solutions.

2.2 Artificial Neural Networks and Autoencoders

The use of Artificial Neural Networks (ANN) to solve AD problems is attractive due to the good results they present in other fields. One of the most popular tasks performed by these models is pattern recognition. For this purpose, it is given to the network input-output pairs, and then it will try to find a function that correctly approximates the real relations between them [11].

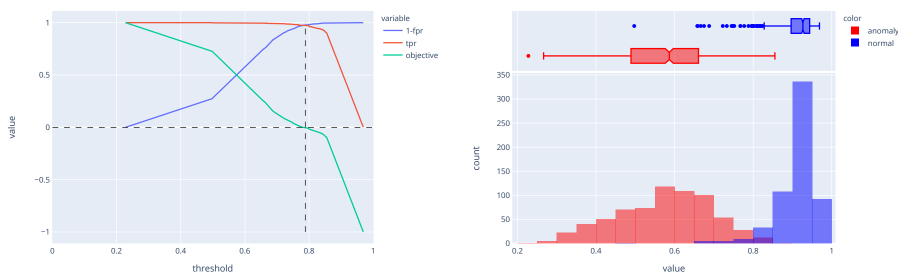
There are two main components in the Autoencoder (AE) architecture (Figure 1a): the *encoder*, that provide a dimension reduction over the input; and the decoder that makes the reverse process. The input and output in the autoencoder training process is the same [3].



(a) Simple bottleneck AE architecture. (b) VAE architecture.

Fig. 1: Autoencoders architectures.

The behaviour of the network consists of reconstructing the original data. In Figure 1a, the original data is represented by x and the reconstructions by y. During the training process, the AE learns to reconstruct only the *normal* class instances. In this scenario the network overfitting to that class is desirable. When the network is trained, the anomalies are found by comparing the original and reconstructed data. It is considered an anomaly if the reconstruction similarity is low (or the reconstruction error is high), according to a defined threshold. The threshold is calculated using different methods, depending on data availability. When a labelled dataset is available it is possible to calculate the threshold in a supervised way, using an objective function that maximizes the True Positives Rate (TPR) and minimize the False Positive Rate (FPR): $TPR-(1-FPR)=0$ (Figure 2a). In cases where only the *Normal* class is present to train the models, the threshold can be inferred by selecting the lower fence of the reconstruction similarities distribution: $t = Q1 - 1.5 \cdot IQR$ (Figure 2b). Since novelties are poorly reproduced by the models, their reconstruction similarities are lower and out of that distribution.



(a) Supervised method (objective function). (b) Unsupervised method (*Normal* class distribution).

Fig. 2: Example of threshold calculation methods.

The VAE (Figure 1b) behaves similarly to the AE architecture. Using the same principle of detecting the anomalies through the reconstruction errors, it differs from the AE in the latent space representation. Instead of using a simple tensor to code the inputs, this architecture uses a distribution and a sample from it to reconstruct each data point. Thereby, the reconstructions are not expected to be so similar since a random sample is performed. However, this approach has the advantage of grouping similar inputs in the latent space. The *z_mean* and *z_log* variables represent the mean and variance of the latent distribution, respectively. The *Sample* is a random selected point from that distribution.

3 Related Work

Different research approaches are being continually explored. One of the most used strategies is the application of smartphone devices to collect different kinds of data that is posteriorly used in data mining processes [13]. The major problems in those methods are related to difficulties found in the devices sensors. The sensors are heterogeneous depending on the device brand and model. The GPS data is not accurate in some cases and depending on the accelerometer sensitivity, it can detect activities that are not related with the pavement conditions. The use of inertial sensors is also dependent on some vehicle characteristics like the suspension system, which introduces even more noise to the data acquisition [6]. Furthermore, anomalies like cracks cannot be detected with accelerometer data since they don't interfere in the car stability.

Instead of vibration-based methods, vision-based ones can avoid the mentioned problems. This approach has the advantage of providing a visual understanding of the observations, that can be used to classify each instance. Several approaches using imaging methods present the distinction between different methods according to the level of detail: presence, detection and measurement [9]. The *presence* is the distinction between good and bad pavement, the *detection* focus on distinguishing between different types of distresses and the *measurement* works on a more specific level to identify their severity. The used data can be 3D images [18] with highly gathering costs or 2D images [17] where budget cameras can reduce the solutions costs.

4 Expirements and Results

For the present work, the main focus is to detect the *presence* of distresses in the pavement, using 2D images to achieve this objective. Similar to the approach presented by Nan Wang for detecting brain tumor anomalies [16], AE and VAE models are used in this context to detect distresses.

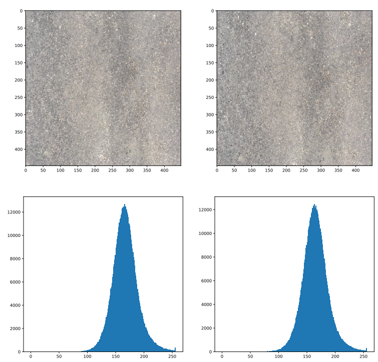
4.1 Dataset and Preprocessing

The input data used for training the Machine Learning (ML) models is a public dataset of road pavement images [1]. The set is originally divided into 2 groups: *Non-Crack* with pavements images in good condition; and *Crack* - with images of cracked pavement.

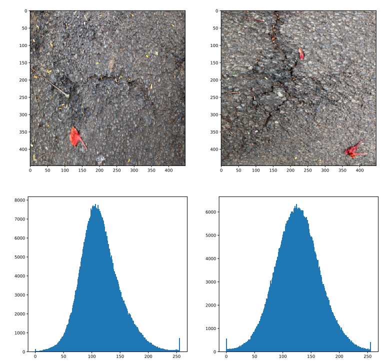
The images are homogeneous in terms of measurements, having a size of 448x448 pixels with 3 color channels (RGB). The dataset is balanced with a total size of 400 images, 200 images in each group.

Studying the images it is clear that the *Anomaly* class (Figure 3b) presents darker values than the *Normal* cases (Figure 3a).

The treatment of the images was carried out so the most relevant aspects for detecting the distresses were highlighted. The transformations that have the best results are the following: downsize the inputs to 256x256 black and white images; use of the *bilateralFilter* since it blurs the image preserving its contours, as the distresses are evidenced at the same time that the pixels corresponding to the asphalt are blurred, removing patterns that could create confusion in the training of the network [14]; and *threshold* filter since anomalies in general have



(a) *Normal* class histograms.



(b) *Anomaly* class histograms.

Fig. 3: Examples of histograms from both classes.

a darker color than the rest of the asphalt, due to differences in light incidence. Examples of these transformations are presented in Figure 4, where the first and third rows show the *Normal* and *Anomaly* original images, respectively and the second and fourth rows show the same images after being processed. The used filters are available in the *OpenCV* library [2].

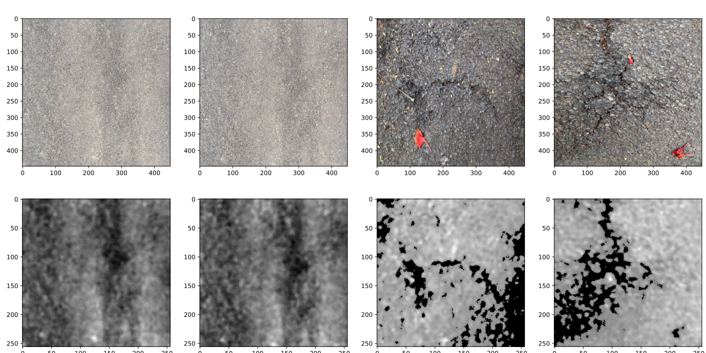


Fig. 4: Examples of the preprocessing step.

This treatment facilitate the work of the ANN in learning the patterns, while minimizing the computational capacity required, since the number of operations to be performed is proportional to the size of the image and the number of color channels is reduced from 3 (RGB) to 1 (*gray-scale*). In the processed images, the distresses stand out clearly. The image set is divided with a proportion of 80/20% for training and testing, respectively.

4.2 Baseline - Binary Classifier

The baseline classifier model is a Convolutional Network, where the inputs are images with dimensions 256x256x1 (width x height x color channels). The output variable is a value between 0 and 1 that will represent one of the two classes, with the division point of the classes being 0.5. The architecture is composed by 4 *Conv2D* Keras layers (16,32,64 and 64 3x3 filters), interspersed by 3 *MaxPooling* layers (2x2). All the layers except the last use the *ReLU* activation function. The last one uses the *sigmoid* function to give the result in the desired domain - [0,1].

In the class "Anomaly", 85% of the real cases of distress are identified by the model. In turn, the model guarantees with a 97% success rate that a case predicted to be anomalous is in fact corresponding to a pavement image in poor condition. The prediction capacity of the model in the two classes presents an accuracy of 91%, which means that in the vast majority of cases the image of the pavement will be classified correctly (Table 1).

4.3 AE and VAE models

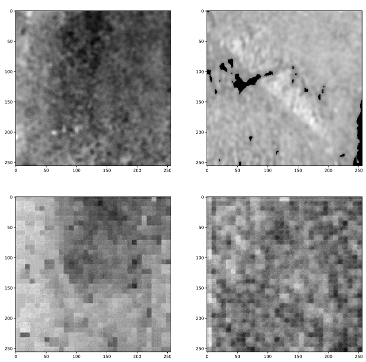
The AE model reconstructs the good pavement images and it is expected that images representing bad condition are poorly reproduced. The implemented structure is as follows: *Encoder* - 3 *Conv2D* (16, 64 and 32 3x3 filters) + *MaxPooling2D* layers (2x2 window); *Bottleneck* - 2 *Dense* layers (40 and 1024 neurons); *Decoder* - 3 *Conv2DTranspose* (all with 64 3x3 filters) intercalated by 2 *BatchNormalization* layers. The *Conv2D* layer is the output of the model that returns an image comparable to the input. The activation functions used in all layers is the *ReLU*. The output layer uses the *sigmoid* function to retrieve the reconstructed pixel darkness (Figure 5a).

With the supervised threshold, it is possible to correctly predict the test cases. The unsupervised threshold is not so accurate in the classes division (Table 1).

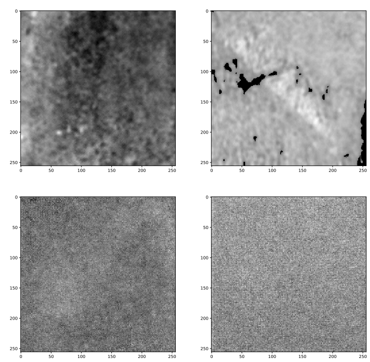
In the VAE, the encoder architecture is similar to the AE encoder structure, with a slight difference. In this case, before the *Bottleneck*, there are two forked *Dense* layers, that represent the mean and standard deviation of the latent distribution. From this distribution, a sample is acquired using a *Lambda* layer that uses a previous specified sampling function.

The output of this layer is the VAE *Bottleneck* and therefore, the output of the encoder part. The decoder has the same behaviour of the AE decoder, building an image similar to the input from the latent representation.

For the training process, a special loss function needs to be defined, using the sum of two losses: the Kullback–Leibler divergence measure to approximate two distributions and the MSE to minimize the difference between the original and reconstructed pixel values. The *Adam* optimizer is used with a learning rate of 0.0005, preventing exploding and vanishing gradients introduced by the intermediate representation. In this case, the reconstructions (Figure 5b) are not so similar as the AE, since they are not a perfect correspondence to the input.



(a) AE reconstructions.



(b) VAE reconstructions.

Fig. 5: Image reconstructions performed by the models.

5 Conclusion

5.1 Discussion

On total, more than 250 experimental models were trained, using different architecture and building decisions. Based on the training events, some discussion questions can be appointed:

1. *How are the results so good if the reconstructions do not appear to be similar to the original images?* The important in reconstructions for AD is not the absolute quality of the reconstructions, but the difference between both classes. The bigger the difference, the better. Even if the reconstructions are not similar when looking at the images, the important pixel level similarities are captured by the model, being also reflected when calculating the similarity between the original and reconstructed images.
2. *What were the criteria to select the image preprocessing filters?* Initially, a group of filters were selected to preserve the main characteristics of the pavement and the distresses in the images. From this group, using the classifier model with a fixed architecture, multiple tests were performed varying only the different filters combinations and parameters. The select filters are the ones that lead to the best results.

3. *There are other available datasets containing pavement images. Why aren't they used?* The selected dataset provides the images in similar settings of the real use case relative to the preferred camera positioning. Other datasets, even though with more images, show different perspectives of the pavement, that are not suitable for this case.

5.2 Conclusions

The main goal of the present study is to prove the possibility of creation of a low-cost, automated pavement monitoring system, using image data to detect distresses. Regarding the ANN models, it is also proved that any of the models are liable to be used, depending on the context and on the importance given to each metric. In Table 1 are shown the overall results in order to compare the different models and approaches, from a distress detection perspective (recall and precision for the anomaly class).

Table 1: Pavement models comparison by metric.

	Classifier	**AE**		**VAE**	
	Supervised	**Supervised**	**Unsupervised**	**Supervised**	**Unsupervised**
AUROC	0.91	0.99		1.00	
Accuracy	0.91	0.94	0.81	1.00	0.72
Precision	0.97	0.94	1.00	1.00	1.00
Recall	0.85	0.94	0.62	1.00	0.44

The model that performs better in a generic view is the VAE architecture using the supervised threshold calculation method, which presents a perfect distinction between both classes. This is the best option to take when there is training data for both classes. When looking for scenarios where there are no labelled data available (e.g., when the data is acquired from a road that is known to be in good conditions *à priori*), the models to be used are the AE and VAE. All the presented methodologies are a first approach to solve the problem and do not represent a fully functional solution. They can be further explored to acquire even better results.

5.3 Future Work

The present article opens a wide spectrum of future work. Regarding the studied research topics, the following work directions are proposed:

1. Improve the studied methodologies scores, specially in a unsupervised scenario, where there is no labelled data to train the models. This can be done exploring the existent AE and VAE models, that use only normal data to be trained, and modifying the threshold calculation algorithms. Examples of

alternatives to calculate the threshold is the use of z-scores and the Empirical Rule, when the SSIM data is a Gaussian distribution, or the Chebyshev's Theorem otherwise [12].

2. The GAN framework is also pointed as an alternative solution when dealing with AD. It is suggested the analysis of this approach to detect the presence of distresses in images [7]. Also, generative models, such as VAE and GAN architectures, provide a reduced representation of the images. Some characteristics are encoded in that representation as feature vectors. It is proposed to train similar networks with pavement images to discover feature vectors that can reflect the pavement conditions spectrum.
3. Transfer learning refers to the use of pretrained models, applying them to a different domain, *transferring* the already learned information [8]. This is a common practice in CV problems, freezing some convolutional layers already trained with images and training only a part of the network that will be specific for that domain [3]. It can also be used in this case, reducing the training times and eventually improving the results.

Acknowledgments

This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the R&D Units Project Scope: UIDB/00319/2020.

References

[1] A. Jayanth Balaji et al. *Asphalt Crack Dataset*. https://data.mendeley.com/datasets/xnzhj3x8v4/2. 2019. DOI: 10.17632/XNZHJ3X8V4.2. URL: https://data.mendeley.com/datasets/xnzhj3x8v4/2.

[2] G. Bradski. "The OpenCV Library". In: *Dr. Dobb's Journal of Software Tools* (2000).

[3] Francois Chollet. *Deep Learning with Python*. 1st. USA: Manning Publications Co., 2017. ISBN: 1617294438.

[4] Fabrizio D'Amico et al. "Evaluation of the impact of pavement degradation on driving comfort and safety using a dynamic simulation model". In: *Advances in Transportation Studies* 1 (July 2018), pp. 109–120. DOI: 10.4399/978882551688310.

[5] John Laurent et al. "Using 3D Laser Profiling Sensors for the Automated Measurement of Road Surface Conditions". In: *7th RILEM International Conference on Cracking in Pavements*. Ed. by A. Scarpas et al. Dordrecht: Springer Netherlands, 2012, pp. 157–167. ISBN: 978-94-007-4566-7. DOI: 10.1007/978-94-007-4566-7_16.

[6] Johannes Masino et al. "Learning from the crowd: Road infrastructure monitoring system". In: *Journal of Traffic and Transportation Engineering (English Edition)* 4.5 (Oct. 2017), pp. 451–463. DOI: 10.1016/j.jtte.2017.06.003. URL: https://doi.org/10.1016/j.jtte.2017.06.003.

[7] Federico Di Mattia et al. "A Survey on GANs for Anomaly Detection". In: *CoRR* abs/1906.11632 (2019). arXiv: 1906.11632. URL: http://arxiv.org/abs/1906.11632.

[8] Sinno Jialin Pan and Qiang Yang. "A Survey on Transfer Learning". In: *IEEE Transactions on Knowledge and Data Engineering* 22.10 (Oct. 2010), pp. 1345–1359. DOI: 10.1109/tkde.2009.191. URL: https://doi.org/10.1109/tkde.2009.191.

[9] Stefania C. Radopoulou and Ioannis Brilakis. "Automated Detection of Multiple Pavement Defects". In: *Journal of Computing in Civil Engineering* 31.2 (Mar. 2017), p. 04016057. DOI: 10.1061/(asce)cp.1943-5487.0000623. URL: https://doi.org/10.1061/(asce)cp.1943-5487.0000623.

[10] Terance Rephann and Andrew Isserman. "New highways as economic development tools: An evaluation using quasi-experimental matching methods". In: *Regional Science and Urban Economics* 24.6 (Dec. 1994), pp. 723–751. DOI: 10.1016/0166-0462(94)90009-4. URL: https://doi.org/10.1016/0166-0462(94)90009-4.

[11] J.P.A. Ribeiro, L.F.F. Gomes, and S.M.O. Tavares. "Artificial Neural Networks Applied in Mechanical Structural Design". In: *Journal of Computation and Artificial Intelligence in Mechanics and Biomechanics* 1.1 (Apr. 2021), pp. 14–21. DOI: 10.5281/zenodo.4669797. URL: https://doi.org/10.5281/zenodo.4669797.

[12] Douglas Shafer and Zhiyi Zhang. *Introductory Statistics*. Saylor Foundation, 2012. ISBN: 9781453344873.

[13] Nuno Silva et al. "Road Anomalies Detection System Evaluation". In: *Sensors* 18.7 (June 2018), p. 1984. DOI: 10.3390/s18071984. URL: https://doi.org/10.3390/s18071984.

[14] C. Tomasi and R. Manduchi. "Bilateral filtering for gray and color images". In: *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*. 1998, pp. 839–846. DOI: 10.1109/ICCV.1998.710815.

[15] Mounica Vinayakamurthy. "Effect of Pavement Condition on Accident Rate". MA thesis. ARIZONA STATE UNIVERSITY, 2017.

[16] Nan Wang et al. *Brain Tumor Anomaly Detection via Latent Regularized Adversarial Network*. 2020. arXiv: 2007.04734 [eess.IV].

[17] Fan Yang et al. "Feature Pyramid and Hierarchical Boosting Network for Pavement Crack Detection". In: *CoRR* abs/1901.06340 (2019). arXiv: 1901.06340. URL: http://arxiv.org/abs/1901.06340.

[18] Allen Zhang et al. "Deep Learning–Based Fully Automated Pavement Crack Detection on 3D Asphalt Surfaces with an Improved CrackNet". In: *Journal of Computing in Civil Engineering* 32.5 (Sept. 2018), p. 04018041. DOI: 10.1061/(asce)cp.1943-5487.0000775. URL: https://doi.org/10.1061/(asce)cp.1943-5487.0000775.