

Learning Wear Patterns on Footwear Outsoles Using Convolutional Neural Networks

Xavier Francis*, Hamid Sharifzadeh*, Angus Newton†, Nilufar Baghaei‡, and Soheil Varastehpour*

*School of Computing
Unitec Institute of Technology
Auckland, New Zealand
{xfrancis, hsharifzadeh, spour}@unitec.ac.nz

†Physical Evidence Team
Institute of Environmental Science and Research (ESR)
Auckland, New Zealand
angus.newton@esr.cri.nz

‡Department of Information Technology
Otago Polytechnic Auckland Campus (OPAIC)
Auckland, New Zealand
nilufar.baghaei@op.ac.nz

Abstract—Footwear outsoles acquire characteristics unique to the individual wearing them over time. Forensic scientists largely rely on their skills and knowledge—gained through years of experience—to analyse such characteristics on a shoeprint. In this work, we present a deep learning model that, for the first time, can predict the wear pattern on a unique dataset of shoeprints that captures the life and wear of a pair of shoes. We also present an additional architecture able to reconstruct the outsole back to its original state on a given week, and provide empirical evaluations of the performance of both models.

Index Terms—Shoeprints, Deep Learning, CNN, Predictive Modeling, Wear-and-Tear

I. INTRODUCTION

Among the many forms of physical evidence found at crime scenes, shoeprints are one of the most frequently seen, with a high degree of evidential value attributed to them. Marks and prints formed by the footwear worn by the criminal(s) are frequently found at scenes-of-crime and their study was being recorded as early as 1786 [1]. This is in part due to the ability of a shoeprint to uniquely identify an individual, by evaluation of the combination of tears, nicks, cuts, scratches and other abrasions that form on the outsole as a function of wear. This ‘wear pattern’ is influenced by biomechanics such as the weight and gait of the wearer, environmental stressors, and additional factors like the material of construction. Bodziak defined wear as “the erosion of the outsole due to abrasive forces that occur between the outsole and the ground” [2]. By considering the wear pattern, in addition to the pattern of the outsole introduced in the manufacturing process, one is able to ascertain if the shoe of a suspect formed the print found at the crime scene.

In spite of their uniquely identifiable nature and their frequency of appearance at scenes-of-crime, shoeprints are not

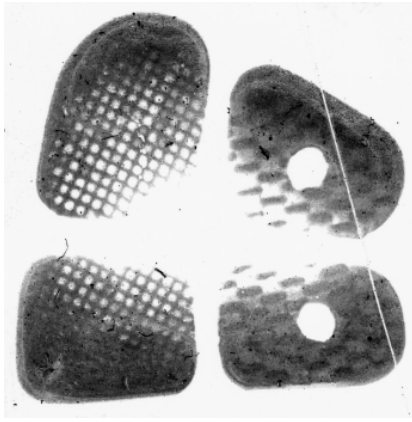
often used as evidence in a court of law. This is in part due to the variation in quality of scene-of-crime impressions, which are often incomplete or degraded. Another challenge is the large search space of potential outsoles; arising from the number of outsoles being designed and manufactured.

Consider the scenario where a substantial period of time elapses between the perpetration of a crime and the identification of suspect(s). In such situations, it falls on the forensic scientist to evaluate the outsole and determine if it matches the scene print while accounting for the formation of additional wear features. This task involves the careful analysis of the outsole and requires intimate knowledge of the breadth of factors and variables that influence wear patterns.

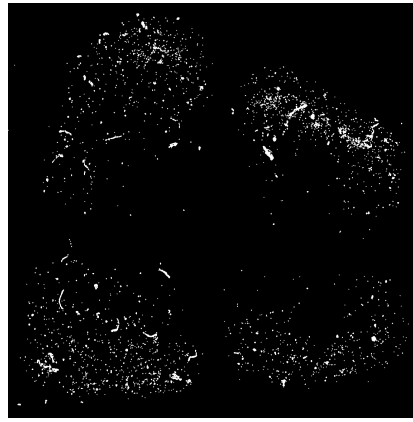
The forensic examiner’s interpretation of the shoeprint and its admissibility as evidence is built through their years of experience in studying shoeprints and the individualising characteristics that contribute to the wear pattern. Such knowledge is notoriously hard to quantify and explain. Deep learning models have made large strides in developing representations of domains like these.

In this work, we adapt a convolutional neural network (CNN) architecture for the task of pixel-wise prediction of shoeprint wear. Our core contributions are as follows—(i) we describe a methodology that utilises a CNN to predict outsole wear formation on a unique dataset of shoeprints, and (ii) an alternate architecture that is able to reconstruct the outsole back to its original state on a given week within a timeframe of one year.

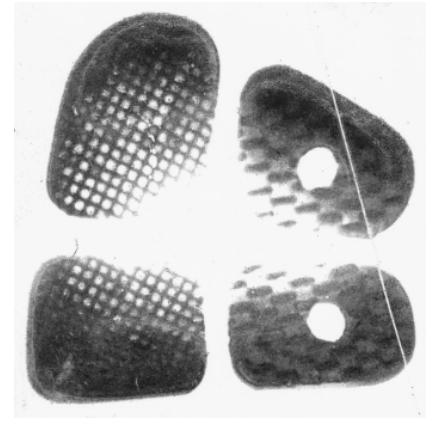
In the following sections we first survey the related literature in the domains of forensics and shoeprints, and deep learning; followed by a description of our novel dataset. We proceed to detail our methodology, analyse the results of our experiments, and finally we conclude the paper.



(a) Cropped region of heel on week 4. Noise is evident.



(b) Noise map of image 1a obtained via thresholding.



(c) Fully filtered image, showing mitigated noise.

Fig. 1: Intermediate stages of the denoising methodology developed for our dataset.

II. RELATED WORK

A. Shoeprint Classification

The responsibilities of the forensic footwear examiner are: (i) to identify the make and model of a given shoeprint, by comparing it against a large set of known prints and (ii) to consider the individualising characteristics of the print to assign the print to an owner. The first task is largely objective in nature; by comparing the scene-of-crime print against a database of reference shoeprints, one is able to find a match and retrieve the relevant metadata. Numerous computational methods have been developed over the years to assist the examiner in this task.

Automated approaches to shoeprint retrieval and classification have seen a multitude of approaches — Fourier features [3], fractals [4], power spectral density [5], Hu moment invariants [6], Harris points and SIFT descriptors [7], Mahalanobis distance as feature descriptors [8], wavelets as an edge detector and neural networks for recognition [9], and transforms like Radon and Gabor [10] [11].

In 2017, Richetelli et al. [12] postulated that the recent advances in deep learning could carry over to field of footwear classification. Kong et al. [13] and Zhang et al. [14] were some of the first to apply CNNs to this task. However, they have not considered wear patterns and our work can be seen as a new contribution to the literature on deep learning applications in forensic science.

B. Shoeprint Wear

While the above research considers the challenge of using computational methods to aid in the task of shoeprint identification, our focus is on using computational methods to model shoeprint wear; specifically, we consider how outsole features change over time. Research in this domain is sparse, with a few considering wear formation manually [15] [16], and fewer using pattern recognition techniques [17] [18]. All of the above mentioned studies vary in scale, time, and ambition.

Understandably, controlling the variables that influence outsole wear is in itself a challenge.

C. Image-to-Image Regression

Given our dataset of 52 shoeprints, described in III, we wish to learn a model of the wear pattern captured within. Once trained, this model should be capable of extrapolating the wear pattern on seeing a new shoeprint. Fundamentally, we approach this as an image-to-image regression task. The literature contains many successful applications of deep learning to these types of dense prediction tasks; such as image inpainting [19] [20], super-resolution [21], denoising [22], and image recovery from compressed representations [23]. Deep neural networks (DNNs) and their convolutional variants have established state-of-the-art performance over nearly all facets of computer vision tasks. One of the primary advantages of using DNNs is their ability to learn end-to-end mappings without the use of image priors, or the explicit engineering of features.

Our dataset shows the life and wear of a pair of shoes through impressions captured at evenly spaced intervals of time. To the best of our knowledge, this is the first time such a dataset has been used in the literature of deep learning. A closely related problem is video frame generation/prediction [24] that involves operations on inputs in the spatial domain, while simultaneously capturing correlations in the temporal domain. Notably, in video frame prediction, one has access to an extensive amount of data by using each frame in the video sequence as a datapoint. Finn et al. [25] use a combination of convolutional and LSTM layers to model pixel motion and optical flow. They introduce a dataset with 1.5 million video frames and a model that predicts video sequences up to 1 second in the future. Our dataset, described in the next section, is significantly smaller in size.

III. DATASET

For the collection of our dataset, we limit the influence of environmental variables and consider the formation of

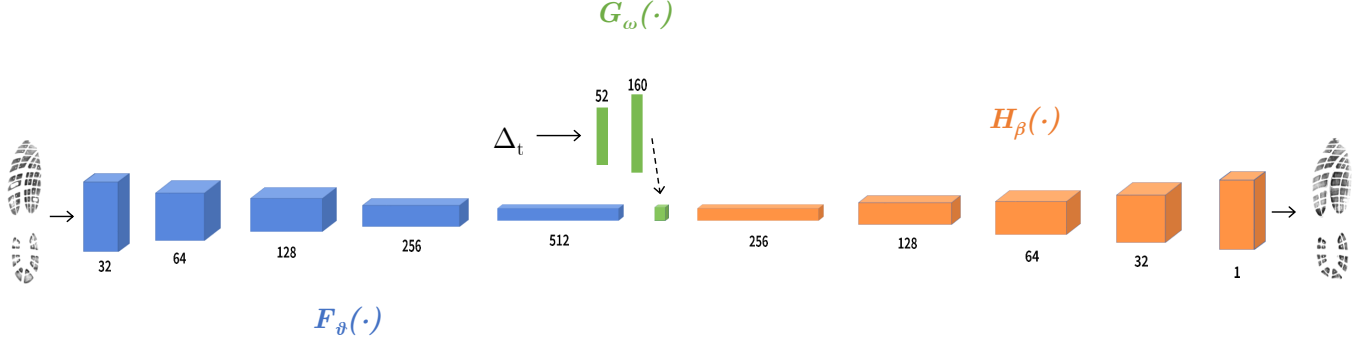


Fig. 2: A visualisation of the architecture of our CNN.

wear characteristics on a single pair of shoes, worn daily by one individual, over the course of one year. A pair of *Asics*-brand men’s sneakers were purchased and worn by the forensic specialist everyday for a period of 52 weeks in an urban environment. Impressions of both outsoles were then captured every fortnight using *BVDA* gels. These impressions were scanned into high-resolution digital negative *TIFF* files, yielding a total of 52 files, including week 0 (unfortunately, impressions were not captured on week 6). Each file is a 256-level grayscale image.

The outsole itself consists of approximately 63 ‘block features’ of varying size and shape. Some of these block features contain other features within — such as the *Asics* brand logo, which is not visible in the impression from week 0, but reveals itself over time; a text pattern that says ‘*GEL*’; and circular ‘holes’ present in many of the block features. These are all patterns imprinted by the manufacturer and their appearance and disappearance can be observed as they wear.

During the course of recording these impressions, many forms of unwanted features were captured in addition to the shoeprint. These included air bubbles, fingerprints, dust, debris overlapping the shoeprint, ghosting of the impression, and areas of missing detail. Of these, the most egregious was determined to be the debris which appeared to consist of fibres and other objects that were transferred from the outsole and onto the gel in the process of imprinting. The debris was particularly problematic due to it obscuring regions of interest in the shoeprint. We refer to these unwanted features that obscured the object of interest as ‘noise’.

To address this, we developed a denoising method capable of maintaining the high and low-level wear patterns, while simultaneously mitigating the noise present in the dataset. This methodology operates by using a local adaptive thresholding to obtain a binary mask, ROI filtering and morphological operations to process the mask, creating a noise map which allows for processing each block feature of the outsole independently, and using an averaging filter to mitigate the noise. Figure 1 highlights a few stages of this denoising methodology. The shoeprint images were cropped, denoised, and registered in

the data preparation stage.

Next, we articulate the CNN architectures developed to model the wear pattern on this dataset.

IV. METHODOLOGY

For the task of modelling wear patterns, we implement a CNN architecture in the style of an auto-encoder, inspired by the work in Tatarchenko et al. [26] and Vukoti et al. [27]. This architecture consists of three branches—an encoder $F_\theta(\cdot)$ that takes as input a shoeprint image X , a delta branch $G_\omega(\cdot)$ that encodes a representation of time from a parameter Δ_t , and a decoder $H_\beta(\cdot)$ that learns an upsampling function to predict the wear pattern. These branches take the form shown in (1).

$$\begin{aligned} F_\theta &= \sigma(X * \theta) \equiv f, \\ G_\omega &= \sigma(\Delta_t \cdot \omega) \equiv g, \\ H_\beta &= \sigma(f \# g *' \beta) \end{aligned} \quad (1)$$

where σ represents an activation function—ReLU or sigmoid, $*$ represents the convolution operation, $\#$ the concatenation of two tensors, and $*'$ the transpose convolution. Bias terms are omitted for notational convenience.

The encoder is made up of 5 convolutional layers that act as feature extractors by performing discrete convolutions over the input image, with an increasing depth of feature maps. We double feature maps with each layer, going from 32 in the first layer, to 512 in the last convolutional layer. The delta branch consists of 2 fully connected layers; the output of this branch is reshaped and concatenated with the output of the last convolutional layer. This tensor is then fed into the 5 transpose convolutional layers of the decoder, which successively upsample the extracted feature maps and the output of the delta branch to produce an output of the same dimensions as the input.

Transpose convolutional layers are used here as a learnable upsampling function, as opposed to a fixed upsampling function (such as bilinear) in combination with 2D convolutions, as frequently seen in the literature. We discard pooling layers

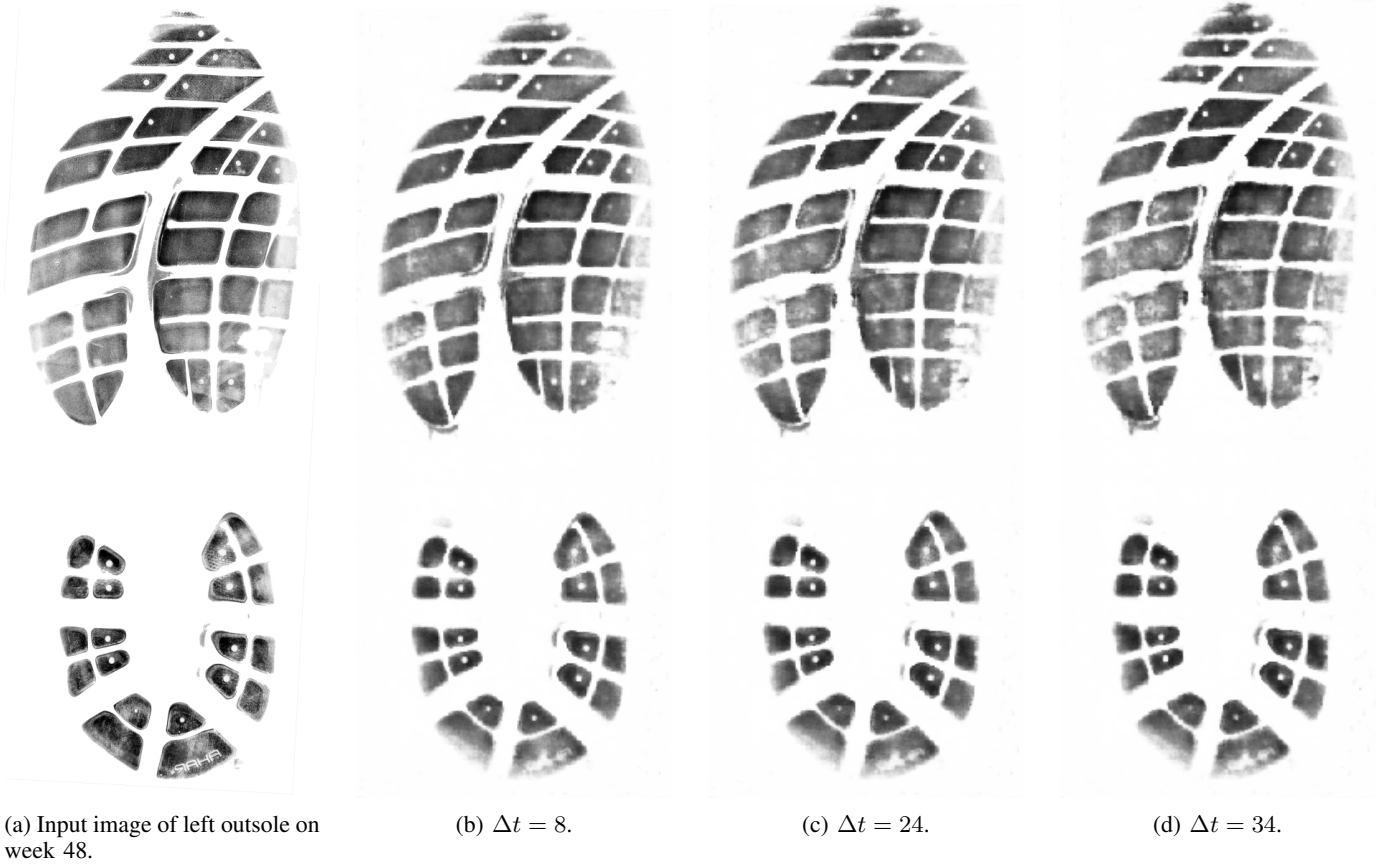


Fig. 3: Predictions of the model described in IV-A, given week 48 as input and a range of values for Δt .

as traditionally seen in convolutional architectures since our denoising method removes redundant information in the image in the pre-processing stage.

The parameters of the network are updated by minimising the squared error loss:

$$L = \frac{1}{n} \sum_{i=1}^n \|H_{\beta}(F_{\theta}(X_i) + G_{\omega}(\Delta t)) - Y_i\|_2^2 \quad (2)$$

where n is the number of training images presented to the network in one epoch, $X_i \in \mathbb{R}^2$ is the i th image, and Y_i the ground-truth image that corresponds to the input.

A visualisation of this architecture is shown in Figure 2.

A. Moving Forward: Outsole Wear Prediction

Our first model is designed to extrapolate wear patterns present in the shoeprint and form a prediction of what they might look like after a given period of time, denoted by Δt . The input image is presented at current relative time $t_0 = 0$. We train this model to predict the appearance of the input, after the elapsed time Δt , where $\Delta t \in [0, 52]$. Δt is incremented in steps of 2 to maintain consistency with the timeframe captured in our dataset. The model then predicts the shoeprint at $t_0 + \Delta t$.

Formally, we train the model by feeding inputs as batches of $\{X, \Delta t, Y\}$ tuples, where X represents the input image centered at a current relative time $t_0 = 0$; Δt represents the

desired temporal displacement; and Y represents the ground truth shoeprint image after the desired temporal displacement. Figure 3 shows a sample of predictions from this model.

B. Moving Backward: Outsole Reconstruction

Our second model is one that reconstructs the input shoeprint back to its state on any given week in a timeframe of one year. For this task, we use the same architecture as in IV-A. The only difference with this model is in how we design the Δt parameter. Here, Δt is represented as a logical vector $\in \mathbb{R}^{52 \times 1}$; wherein each element represents a week of the year, taking a value in $\{0, 1\}$, such that the desired week corresponding to the ground truth Y is represented as 1, and all other weeks represented as 0.

Once again, we train the model by presenting $\{X, \Delta t, Y\}$ tuples, and design the logical Δt vector in increments of 2 to correspond with the fortnightly nature of our captured dataset. Figure 4b shows a reconstruction produced by this model.

V. EXPERIMENTS

A. Parameters

Model training is performed by dividing the dataset of 52 images into an 80/20 training/test split. The first 42 images — both left and right outsoles — are used to train the model in conjunction with a MSE loss and the Adam optimisation

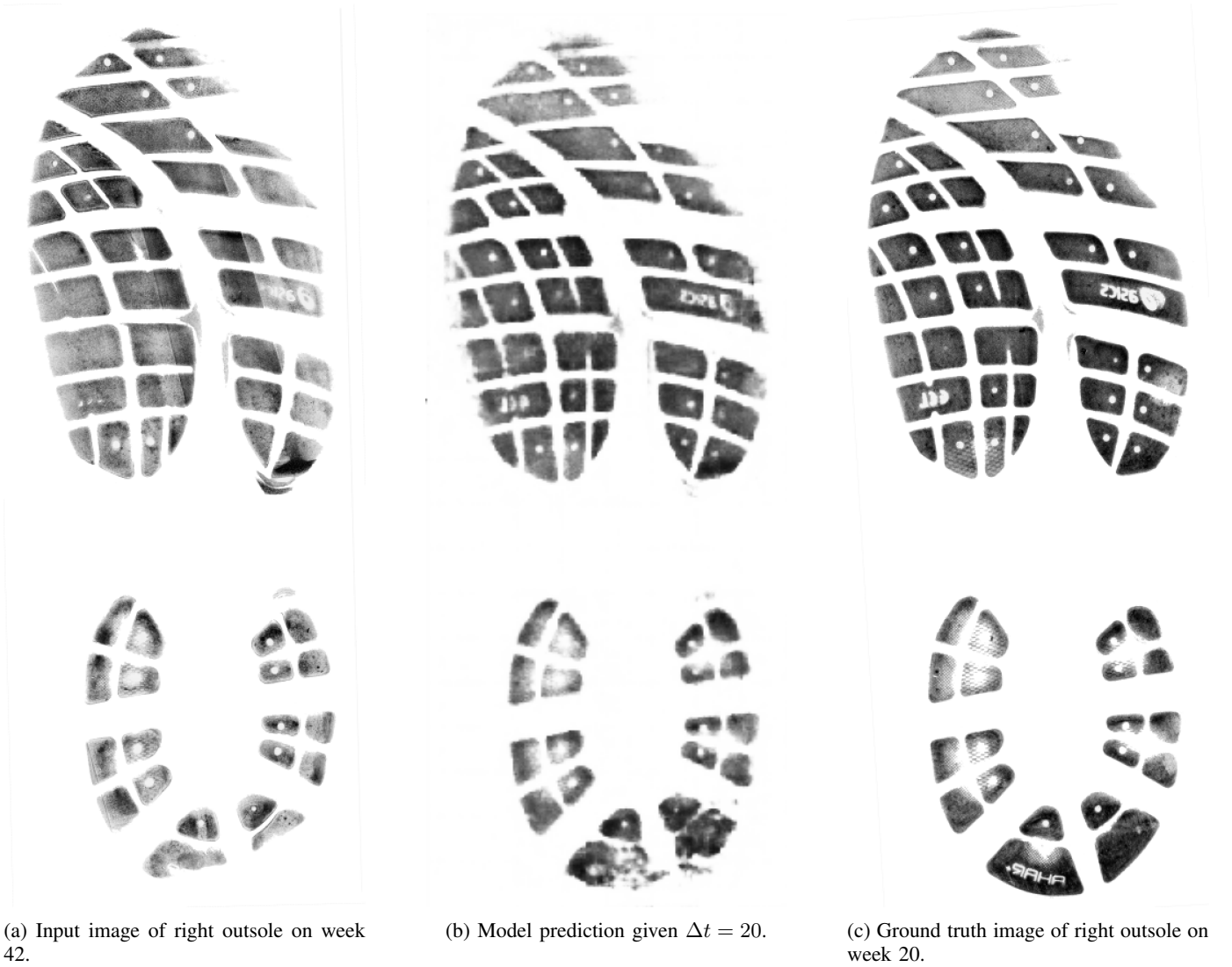


Fig. 4: Outsole reconstruction predicted by the model detailed in IV-B.

algorithm [28]. For training model IV-B, the split is reversed — i.e. we use the last 42 images for training, and test with the remaining images in our dataset that capture the start of our timeline. We use a learning rate of $1e-5$ and train for 10,000 epochs. Activation functions throughout are the ReLU; except for the last layer which uses a sigmoid function to obtain outputs $\in [0, 1]$. We found the same hyperparameters to be effective for both our models.

Our dataset is composed of 52 grayscale images with a resolution of 13750×5500 . Fitting this dataset into memory during training required downsampling it to 640×256 . We train both models end-to-end from random initialisation, to generate the desired output outsole given the image of a shoeprint from our dataset. Alternative learning rates, initialisation schemes, optimisers, and loss functions were evaluated before settling on the above.

B. Results

Our network successfully learns to model the high-level wear pattern embedded in the shoeprints. From observing the outputs, it is evident that the model has formed an internal representation sufficiently capable of predicting the wear pattern found in the dataset. Relevant regions of Figure 3 have been cropped and highlighted in Figure 5. Similarly, Figure 6 consists of crops of Figure 4. We compare the predictions from the model against the ground-truth images from our dataset and note the below observations:

- In Figure 5a we see a cluster of four block features on the right edge of the outsole. In the model’s predictions, we see them degrade and eventually merge in the final prediction, $\Delta_t = 34$. From the ground-truth image of week 52, we confirm that this change has indeed occurred; although clearly the model’s estimation of 20 weeks is far off from the reality of this eventual merger

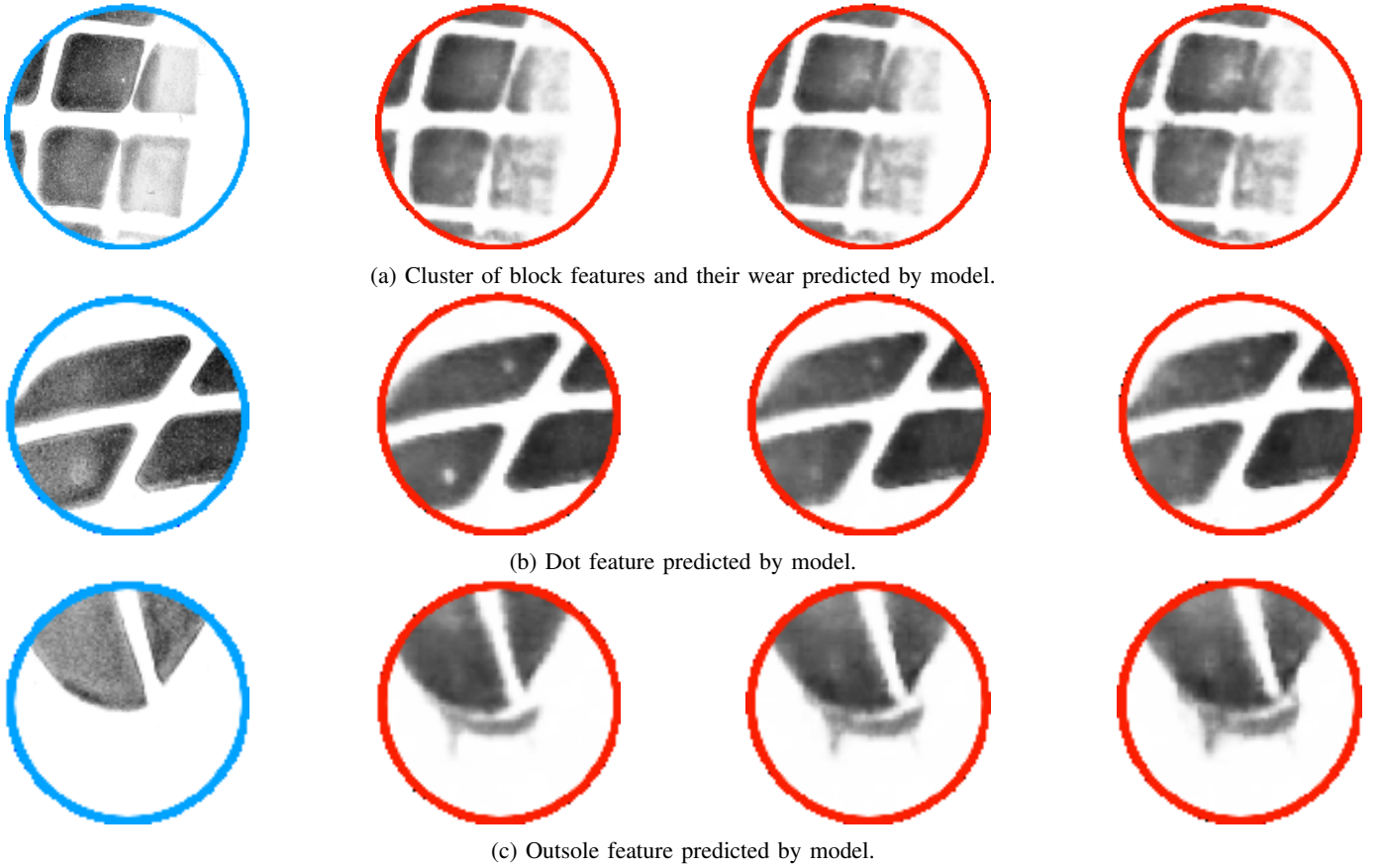


Fig. 5: Highlights of relevant regions of predictions shown in Figure 3. Encircled in blue is the input to the model of the left outsole on week 48. Circled in red are the predicted wear patterns of the model, given Δ_t values of 8, 24, and 34, respectively from left to right.

materialising in 4 weeks.

- Figure 5b shows two ‘dot features’ visible in the first prediction of $\Delta_t = 8$; note that these two features are not present in the input image of week 46. This feature is present throughout the outsole on many of the block features but have disappeared through wear-and-tear. It also happens to be visible in this exact region in all of the training images—weeks 0 through 42—but had eroded from the outsole by the time the input shoeprint was captured. Interestingly, in the model’s latter two predictions— $\Delta_t = 24, 34$ —we see the feature degrade and eventually disappear, in line with the ground-truth; showing that the model has learned the wear development on this and similar features, despite consistently observing the dot features in this region throughout the training images.
- Figure 5c highlights a ‘ridge’ feature seen in all the predictions, but not in the input. We verify through our dataset that this is in fact a feature of the outsole, seen in roughly half of the images in the training set, but is missing in the input image. Note how the model’s predictions show this ridge growing progressively larger in size, as the outsole erodes.

- In the outsole reconstructions of model IV-B, we see the successful reproduction of the *Asics* brand logo (Figure 6a), and the separation of block features that had merged through wear (Figure 6b). Also note the reconstruction of the feature that spells the word ‘GEL’, imprinted by the manufacturer.
- The bottom region of the heel in the prediction seen in Figure 4b is blurry and poorly defined. This is due to the inconsistency of the appearance of this region in the dataset. During the data collection phase this region was either frequently occluded by fingerprints and debris, or ill-formed due to a lack of pressure between the outsole and the gel while collecting the impression. We deduce that this inconsistency in appearance is what has led the model to develop a fuzzy representation of this region.

From our evaluation of the results, we ascertain that our methodology is sufficient to capture the wear pattern from our dataset, and to perform both outsole prediction and reconstruction with reasonable accuracy.

We also note that our network is handicapped by a lack of training data. In the era of deep learning, where models are routinely trained with millions of datapoints, we have sufficed with a meagre 52 images. Despite the size of the dataset,

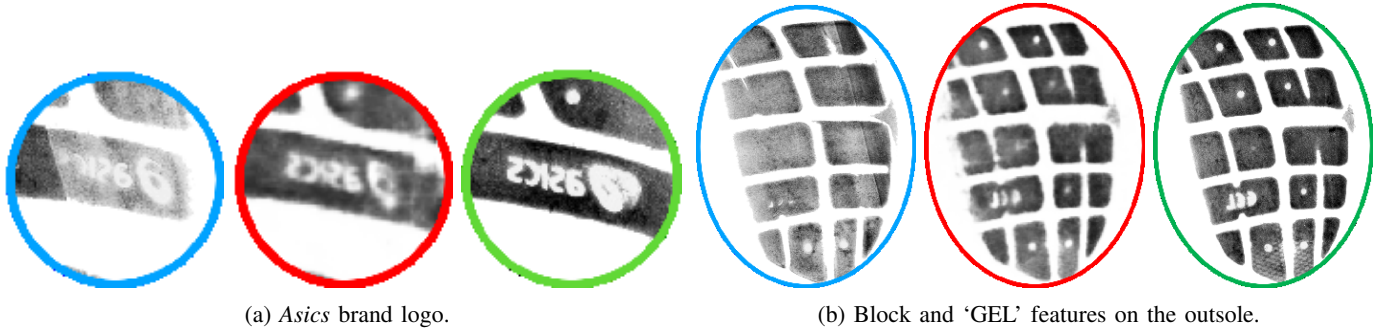


Fig. 6: Highlights of relevant regions of outsoles shown in Figure 4. Encircled in blue is the input to the model of the right outsole on week 42. Circled in red is the predicted reconstruction of the model for week 20, and circled in green is the ground-truth image of week 20 from the dataset.

each pixel in the input image is a feature the model can learn from, and the 640×256 resolution of our training data is purely limited by processing power; allowing for a more robust model to be trained using higher resolution images. The generalisation ability of deep learning models can also benefit from an adequately sized dataset that fully captures the diversity of the problem domain.

Empirical evaluations are given in the next subsection.

TABLE I: Mean and standard deviation of SSIM scores.

	Model IV-A	Model IV-B
Mean	0.8645	0.8596
STD	0.0381	0.0345

C. Evaluation

For an objective evaluation of the performance of our models, we use the standard metric of *Structural Similarity Index* (SSIM) [29], by comparing the predictions of the models against the ground-truth images from the validation dataset. SSIM is defined in (3).

$$SSIM(f, g) = l(f, g)c(f, g)s(f, g) \quad (3)$$

where

$$\begin{cases} l(f, g) = \frac{2\mu_f\mu_g + C_1}{\mu_f^2 + \mu_g^2 + C_1}, \\ c(f, g) = \frac{2\sigma_f\sigma_g + C_2}{\sigma_f^2 + \sigma_g^2 + C_2}, \\ s(f, g) = \frac{\sigma_{fg} + C_3}{\sigma_f\sigma_g + C_3} \end{cases}$$

where l , c , and s denote the luminance, contrast, and structure comparison functions respectively. The term f denotes the ground-truth image and g the predicted image. μ and σ denote mean and standard deviation of image luminance and contrast, respectively. σ_{fg} is the covariance between f and g . C_1 , C_2 , and C_3 are positive constants employed to avoid a null denominator. The SSIM index is a positive value in $[0, 1]$, where 0 denotes no correlation and 1 denotes $f = g$.

The results are given in Table I. As evident, the models have an average accuracy of 86%.

Additionally, we compared *Peak Signal-to-Noise Ratio* (PSNR) (4) scores for our models.

$$PSNR(f, g) = 10 \log_{10}(255^2 / MSE(f, g)) \quad (4)$$

Once again, in (4) as in (3), f and g denote the ground-truth and predicted images, respectively; and MSE is the mean squared error — i.e. MSE between f and g . The PSNR score for model IV-A has an average of 22dB, while model IV-B scores 21.3dB. Results are given in Table II.

TABLE II: Mean and standard deviation of PSNR scores.

	Model IV-A	Model IV-B
Mean	22.0065	21.3681
STD	1.8930	1.8820

VI. CONCLUSION

We present a convolutional neural network architecture in the style of an auto-encoder that, for the first time, can model the wear pattern collected in a unique dataset of shoeprints. We show that the model can learn an accurate representation of the pattern of wear-and-tear found in the shoeprints by applying it to predict the wear pattern on the outsole after a given temporal displacement; and by having it reconstruct the outsole back to its original state at a previous point in time. We address the drawbacks of the model and present objective evaluations of its performance, which show the predictions of both models to be 86% accurate.

This work adds to the scant literature on shoeprint wear patterns by presenting a computational model of outsole wear. The model presented within can be applied to supplement the skills and expertise of the forensic examiner in their analysis of crime scene shoeprints; and additionally to train the novice forensic scientist to hone their skills.

ACKNOWLEDGEMENTS

The authors express their gratitude to the *High Technology Transdisciplinary Research Network* at Unitec, and the *Institute of Environmental Science and Research* (ESR), New Zealand for jointly funding this research.

REFERENCES

- [1] M. Liukkonen, H. Majamaa, and J. Virtanen, "The role and duties of the shoeprint/toolmark examiner in forensic laboratories," *Forensic Science International*, vol. 82, no. 1, pp. 99–108, 1996.
- [2] W. J. Bodziak, *Footwear impression evidence: detection, recovery and examination*, CRC Press, 1999.
- [3] Z. Geradts and J. Keijzer, "The image-database REBEZO for shoeprints with developments on automatic classification of shoe outsole designs," *Forensic Science International*, vol. 82, no. 1, pp. 21–31, 1996.
- [4] A. Alexander, A. Bouridane, and D. Crookes, "Automatic classification and recognition of shoeprints," in *Seventh International Conference on Image Processing and Its Applications*, 1999, pp. 638–641.
- [5] P. de Chazal, J. Flynn, and R. B. Reilly, "Automated processing of shoeprint images based on the Fourier transform for use in forensic science," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 341–350, Mar. 2005.
- [6] G. Al Garni and M. Hamiane, "A novel technique for automatic shoeprint image retrieval," *Forensic Science International*, vol. 181, no. 1–3, pp. 10–14, 2008.
- [7] O. Nibouche, A. Bouridane, M. Gueham, and M. Laadjel, "Rotation Invariant Matching of Partial Shoeprints," in *13th International Machine Vision and Image Processing Conference*, Sept. 2009, pp. 94–98.
- [8] F. Dardi, F. Cervelli, and S. Carrato, "A combined approach for footwear retrieval of crime scene shoe marks," in *3rd International Conference on Imaging for Crime Detection and Prevention*, Dec. 2009, pp. 1–6.
- [9] R. Wang, W. Hong, and N. Yang, "The research on footprint recognition method based on wavelet and fuzzy neural network," in *2009 Ninth International Conference on Hybrid Intelligent Systems*. IEEE, 2009, pp. 428–432.
- [10] P. M. Patil and J. V. Kulkarni, "Rotation and intensity invariant shoeprint matching using Gabor transform with application to forensic science," *Pattern Recognition*, vol. 42, no. 7, pp. 1308–1317, 2009.
- [11] W. Pei, Y. Y. Zhu, Y. N. Na, and X. G. He, "Multiscale Gabor Wavelet for Shoeprint Image Retrieval," in *2nd International Congress on Image and Signal Processing*, Oct. 2009, pp. 1–5.
- [12] N. Richetelli, M. C. Lee, C. A. Lasky, M. E. Gump, and J. A. Speir, "Classification of footwear outsole patterns using Fourier transform and local interest points," *Forensic Science International*, vol. 275, pp. 102–109, 2017.
- [13] B. Kong, D. Ramanan, and C. Fowlkes, "Cross-Domain Forensic Shoeprint Matching," in *British Machine Vision Conference*, London, Sept. 2017.
- [14] Y. Zhang, H. Fu, E. Dellandréa, and L. Chen, "Adapting Convolutional Neural Networks on the Shoeprint Retrieval for Forensic Use," in *Chinese Conference on Biometric Recognition*, J. Zhou, Y. Wang, Z. Sun, Y. Xu, L. Shen, J. Feng, S. Shan, Y. Qiao, Z. Guo, and S. Yu, Eds., Shenzhen, China, 2017, pp. 520–527, Springer International Publishing.
- [15] J. M. Wyatt, K. Duncan, and M. A. Trimpe, "Aging of shoes and its effect on shoeprint impressions," *Journal of Forensic Identification*, vol. 55, no. 2, pp. 181, 2005.
- [16] T. W. Adair, J. Lemay, A. McDonald, R. Shaw, and R. Tewes, "The Mount Bierstadt study: An experiment in unique damage formation in footwear," *Journal of Forensic Identification*, vol. 57, no. 2, pp. 199, 2007.
- [17] N. D. K. Petraco, C. Gambino, T. A. Kubic, D. Olivio, and N. Petraco, "Statistical Discrimination of Footwear: A Method for the Comparison of Accidentals on Shoe Outsoles Inspired by Facial Recognition Techniques," *Journal of Forensic Sciences*, vol. 55, no. 1, pp. 34–41, 2010.
- [18] H. D. Sheets, S. Gross, G. Langenburg, P. J. Bush, and M. A. Bush, "Shape measurement tools in footwear analysis: A statistical investigation of accidental characteristics over time," *Forensic Science International*, vol. 232, no. 1, pp. 84–91, 2013.
- [19] D. Pathak, P. Kr. Ahenbuhl, J. Donahue, T. Darrell, and A. Efros, "Context Encoders: Feature Learning by Inpainting," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.
- [20] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Generative image inpainting with contextual attention," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5505–5514.
- [21] C. Dong, C. C. Loy, K. He, and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, Feb. 2016.
- [22] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, July 2017.
- [23] C. Dong, Y. Deng, C. Change Loy, and X. Tang, "Compression artifacts reduction by a deep convolutional network," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 576–584.
- [24] M. Mathieu, C. Couprie, and Y. LeCun, "Deep multi-scale video prediction beyond mean square error," in *International Conference on Learning Representations*, 2015.
- [25] C. Finn, I. Goodfellow, and S. Levine, "Unsupervised Learning for Physical Interaction through Video Prediction," in *Advances in Neural Information Processing Systems* 29, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, Eds., pp. 64–72. Curran Associates, Inc., 2016.
- [26] M. Tatarchenko, A. Dosovitskiy, and T. Brox, "Multi-view 3d models from single images with a convolutional network," in *European Conference on Computer Vision*. Springer, 2016, pp. 322–337.
- [27] V. Vukotić, S.-L. Pintea, C. Raymond, G. Gravier, and J. Van Gemert, "One-Step Time-Dependent Future Video Frame Prediction with a Convolutional Encoder-Decoder Neural Network," in *International Conference on Image Analysis and Processing*. Springer, 2017, pp. 140–151.
- [28] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations*, 2015.
- [29] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.