Proceedings of the International Conference on Sustainable Expert Systems (ICSES-2024)
IEEE Xplore Part Number: CFP24VS6-ART; ISBN: 979-8-3315-4036-4

FootwearFusion: Convolutional Neural Network for Shoe, Sandal, and Boot Classification

Poonam Shourie[1]
Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India
Poonam.shourie@chitkara.edu.in

Vatsala Anand[2]
Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India
vatsala.anand@chitkara.edu.in

Sheifali Gupta[3]
Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India
sheifali.gupta@chitkara.edu.in

***Abstract*— A fundamental task in computer vision, footwear categorization finds many applications in the retail, security, and fashion industries. This paper employs Convolutional Neural Networks (CNNs) to compare and categorize three popular footwear types: shoes, boots, and sandals. The collection is made up of photos that were gathered from several sources and feature a range of footwear brands, colors, and styles. High accuracy rates on both training and validation sets show how effective CNNs are at differentiating between shoes, boots, and sandals based on experimental results. It also examines the effect of class imbalance and dataset size on model performance, offering solutions and insights into possible problems. The present study serves as a standard for assessing future developments in this field and delivers insightful information about the use of deep learning algorithms for footwear categorization in manufacturing technology. Promising outcomes and practical deployment in real-world applications, such as automated inventory management, e-commerce product categorization, and security screening systems, are possible with the created model.**

Keywords—security, CNN, Shoe, sandals, boots

I. INTRODUCTION

Classifying footwear is an important task in computer vision and pattern recognition, with applications in surveillance systems and e-commerce. The practical applications of automatically classifying various shoes, boots, and sandals kinds include safety inspections, personalized recommendation systems, and commercial management of stock. Conventional techniques for classifying footwear mostly depended on feature engineering and manual inspection, both of which have problems with limited scalability and generalization to other brands and styles. But because to recent developments in deep learning, especially the discipline has undergone a revolution as automatic feature learning from unprocessed image data has become possible [1-3].

In this work, shoes, boots, and sandals are the three common forms of footwear that are classified using CNNs. These categories cover a broad spectrum of forms, textures, and styles; their innate visual similarities and minor distinctions make the classification issue difficult to solve. From input photos, deep learning models—particularly CNNs—are skilled at automatically deriving hierarchical representations of visual information [4-6]. CNNs can acquire discriminative features straight from the pixel values, eliminating the need for human feature creation and enabling the capture of minute details and patterns that are critical for differentiating between various shoe types. When it comes to handling big datasets and challenging categorization problems, deep learning models can scale well.

Deep learning models are highly scalable and adaptable to the expanding dataset size as labeled footwear photos become more widely available, which could result in further gains in accuracy. Deep learning—and CNNs in particular—revolutionizes the cataloguing of shoes, boots, and sandals by providing automated feature learning, end-to-end training, flexibility to variations, and scalability, all leading to more precise and effective classification systems [7].

This article will include related work in the section below, followed by the method used and the assessment of the model on various parameters.

II. LITERATURE REVIEW

In a literature review on shoe classification, previous studies, publications, and articles about shoe classification will be summarized and examined. It explains the importance of shoe classification in various fields such as fashion, sports, and ergonomics.

Shoe categorization and recommendation were improved with the application of CNN image identification technology. The CNN model was trained on a large dataset to identify distinct shoe attributes and styles with an accuracy of 95.5% The technique is included in an easy-to-use web platform that provides real-time image recognition. Users can take a picture of a shoe they want to identify instantly, along with information about its brand, price, and availability. Additionally, the CNN-powered recommendation engine enhances the shopping experience by offering tailored recommendations based on client preferences, style, and color [8].

It is now essential to categorize shoes in a way that allows industry experts to meet consumer requests while taking health and medical considerations into account. Attempts have been made to develop a credible and practical framework for shoe categorization. The proposed model in this study shows results of the shoe class as used in the data, with an accuracy of 88.8% using the EfficientNetB3 model [9].

Another article, a two-step deep learning-based technique that uses RGB photos of the shoe to find errors in printed shoes. The part of the shoe in the picture is noted in the first step. Next, using an autoencoder technique, this region is examined for abnormalities in the second phase. The technique requires little integration work into current and ongoing process processes

2024 4th International Conference on Sustainable Expert Systems (ICSES) | 979-8-3315-4036-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICSES63445.2024.10763197

because it employs standard RGB images. A promising result of around 85% is shown by this research [10].

For fast retrieval and discriminative shoe feature expression, this research provides a three-level feature representation for the semantic hierarchy of attribute convolutional neural network. The shoe images are successfully matched across many domains by the features that were derived from the image, region, and part levels. With its newly created loss function, the methodically combines semantic features of nearer visual appearances to avoid shoe images with clear visual differences from being mistaken for one another. To train our network and assess our system, we gather an extensive shoe data set consisting of 12652 related online domain photos and 14341 street domain images with fine-grained features. The accuracy of retrieving the top-20 results is much higher than when using pre-trained CNN features [11].

The article provides a summary of the literature review's main conclusions. Stress the significance of ongoing studies on CNN-based shoe categorization and its possible effects on a range of applications.

III. Materials and Methodology

A. Input dataset

There are 15,000 photos of shoes, sandals, and boots in this dataset. Five thousand photos in every category. The photos are in the RGB color type with a resolution of 136x102 pixels as shown in figure 1.

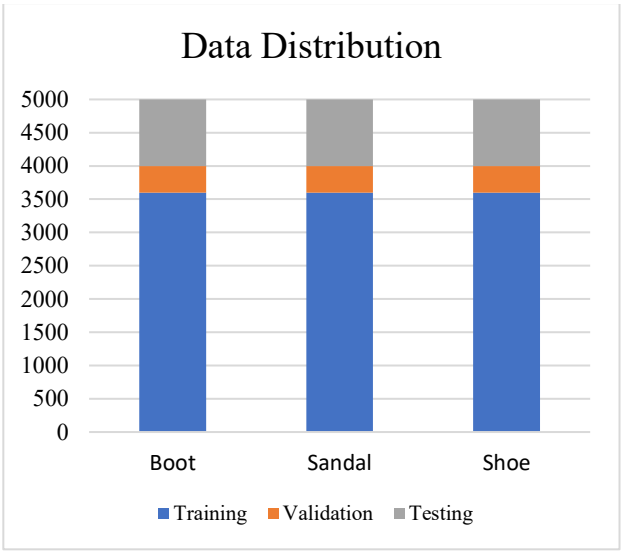

Fig 1. Dataset division for training, validation, and testing [12]

Here, there are three different classes: sandals, Shoes, and boots, samples are shown in Figure 2. With deep neural networks, such as CNNs, this dataset is perfect for multiclass classification.10800 training photos,3000 test photos, and 1200 validation photos in total.

(a) (b) (c)

Fig 2. Samples of shoes, sandals, and boot [12]

B. Methodology

The proposed model has many convolutional layers, max-pooling ,dropout for regularisation, and finally fully linked layers.

Input Layer (Conv2D): This layer uses ReLU activation to apply 64 5x5 filters to the input picture. MaxPooling2D (M1): This layer uses a 2x2 pool size to conduct maximum pooling. Dropout: Twenty percent of the neurons will be arbitrarily fall off during training when this layer's dropout regularisation is applied at 0.2.

Conv2D (C2): This layer activates ReLU and applies 32 5x5 filters. MaxPooling2D (M2): Maximum pooling in a 2x2 pool.

Conv2D (C3): An additional convolutional layer using 32 5x5 filters with ReLU activation. MaxPooling2D (M3):

Maximum pooling in a 2x2 pool. Dropout: An additional layer of dropouts with a 0.2 rate.

Conv2D: An additional convolutional layer using 32 5x5 filters activated by ReLU.MaxPooling2D (M4): Maximum pooling in a 2x2 pool.

Dropout: An additional layer of dropouts with a 0.2 rate. Flatten: The output of the prior layer is flattened to create a one-dimensional array.

Dense: A layer with 256 neurons with ReLU activation that is fully linked.

The output layer (dense) is the last output layer using a softmax activation function. The number of neurons in this layer is determined by the variable n classes, which indicates the number of classes in your classification task.

Table 1. Sequential proposed model

Model: "Sequential"		
Layers	Output Shape	Parameters
Input (conv 2D)	(None,124, 124,64)	4864
Max pooling (M1)	(None,62,62,64)	0
Dropout	(None,62,62,64)	0
Conv 2D(C2)	(None,58,58,32)	51232
Max pooling (M2)	(None, 29, 29, 32)	0
Conv 2D(C3)	(None, 25, 25, 32)	25632
Max pooling (M3)	(None, 12, 12, 32)	0
Dropout(D1)	(None,12,12,32)	0
Conv 2D(C4)	(None, 8, 8, 32)	25632
Max pooling (M4)	(None, 4, 4, 32)	0
Dropout(D2)	(None, 4, 4, 32)	0
Flatten_1	(None,512)	0
Dense_1	(None,256)	131328
Output_1	(None,3)	771
Total Parameters:239,459 Trainable Parameters: 239,459 Non-Trainable Parameters:0		

IV. Results and Discussions

This study investigates assessment criteria to train and assess CNN models for shoe classification, including accuracy, precision, recall, and F1-score. Measures have been adjusted to take into consideration application domain requirements.

A. Epochwise performance

The method of evaluating a classification model's performance at several training epochs for shoes, sandals, and boots is known as epoch-wise assessment. The training set is exploited to train the model on the classification model (such as CNN), and after each epoch, the model's performance is tracked on the validation set. The accuracy of the model has reached 97.19% for the test case and 97.08 % for the validation data set.

Table 2. Epoch wise evaluation

Epoch	Loss	Accuracy	Val_Loss	Val_acc
1	0.6235	0.7140	0.3995	0.8950
5	0.2222	0.9177	0.1615	0.9475
10	0.1539	0.9437	0.1263	0.9508
15	0.0941	0.9670	0.0790	0.9692
20	0.0847	0.9702	0.0817	0.9700
23	0.0793	0.9719	0.0794	0.9708

B. Accuracy and loss curves

The paper showed the training and validation metrics across epochs in Figures 4 and 5, which are the accuracy and loss values for both the training and validation sets after each epoch during training, to visualize the accuracy and loss curves for shoe classification using a Convolutional Neural Network (CNN).

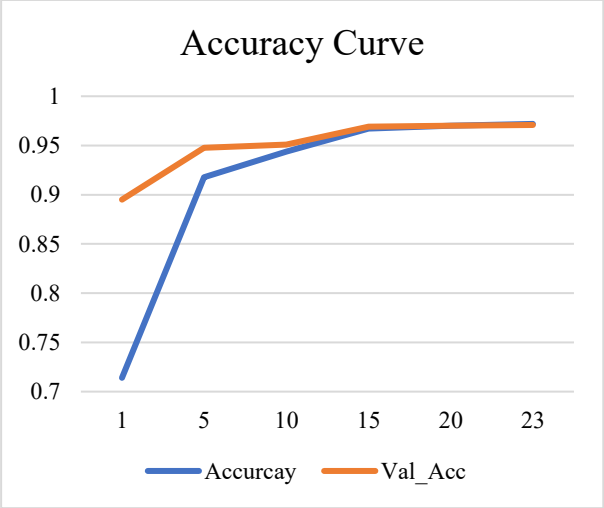

Fig 4. Proposed model accuracy curve

Fig 5. Proposed model loss curve

C. Confusion Matrix

A classification model's confusion matrix helped to determine which classes are confusing one another and where the model might be going wrong. It offers insightful information about how well the model is performing. The correct predictions are represented by the diagonal elements of the matrix (which go from top-left to bottom-right), whereas the wrong guesses are represented by the off-diagonal elements. The boot has 982 accurate predictions,977 sandals, and 983 shoe correct predictions.

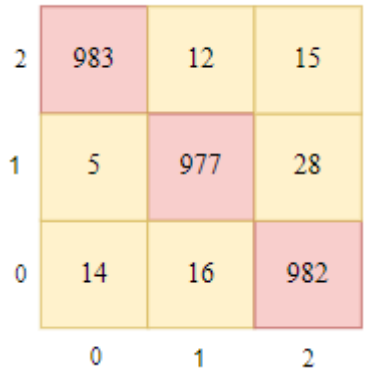

Fig 6. Confusion matrix of the proposed model

D. Performance matrix

Metrics called performance metrics are employed to assess a machine learning model's performance. These metrics offer information about the model's performance and are used to evaluate how useful it is for a classification task as shown below in table 2.

Table 3. Performance of different classes

	Precision	**Recall**	**F1-score**
0	0.98	0.97	0.98
1	0.97	0.97	0.97
2	0.96	0.97	0.96

V. Conclusion

Based on visual characteristics, the CNN model distinguished between shoes, boots, and sandals with an outstanding accuracy rate. This shows that the model is resilient in correctly identifying footwear items and that it has successfully learned the discriminative patterns and traits specific to each category. There are several chances for additional development and exploration. It is possible to expand the model's usefulness and applicability to handle a wider range of possibilities and difficulties by emphasizing user feedback integration, multimodal techniques, transfer learning, fine-tuning, dataset extension, and cross-domain applications that can eventually result in increased user happiness and performance.

References

[1] Shalihah, M., 2015. A look at the world through a word" Shoes": A componential analysis of meaning. English Language Studies, 15(1), pp.81-90.

[2] Khosla, N. and Venkataraman, V., 2015. Building image-based shoe search using convolutional neural networks. CS231n course project reports, pp.1-7.

[3] Trivedi, N. K., Gautam, V., Anand, A., Aljahdali, H. M., Villar, S. G., Anand, D., ... & Kadry, S. (2021). Early detection and classification of tomato leaf disease using high-performance deep neural network. Sensors, 21(23), 7987.

[4] A. Kumar, S. Sharma, N. Goyal, A. Singh, X. Cheng, and P. Singh, "Secure and energy-efficient smart building architecture with emerging technology IoT," Comput. Commun., vol. 176, pp. 207–217, 2021.

[5] R. Dogra, S. Rani, and B. Sharma, "A review to forest fires and its detection techniques using wireless sensor network," in Lecture Notes in Electrical Engineering, Singapore: Springer Nature Singapore, 2021, pp. 1339–1350.

[6] C. Kaushal, S. Bhat, D. Koundal, and A. Singla, "Recent trends in computer assisted diagnosis (CAD) system for breast cancer diagnosis using histopathological images," IRBM, vol. 40, no. 4, pp. 211–227, 2019.

[7] N. K. Trivedi, S. Simaiya, U. K. Lilhore, and S. K. Sharma, "An efficient credit card fraud detection model based on machine learning methods," International Journal of Advanced Science and Technology, vol. 29, no. 5, pp. 3414–3424, 2020.

[8] C.-C. Chang, C.-H. Wei, C.-S. Chen, J.-A. Chen, S.-H. Yeh, and S. Hsiao, "A shoe shopping system based on convolutional neural network image recognition," in 2023 IEEE 5th Eurasia Conference on IOT, Communication and Engineering (ECICE), 2023, pp. 305–309.

[9] K. S. Gill, A. Sharma, V. Anand, and R. Gupta, "Smart shoe classification using artificial intelligence on EfficientnetB3 model," in 2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT), 2023, pp. 254–258.

[10] Kreutz, M., Böttjer, A., Trapp, M., Lütjen, M. and Freitag, M., 2022. Towards individualized shoes: Deep learning-based fault detection for 3D printed footwear. Procedia CIRP, 107, pp.196-201.

[11] Zhan, H., Shi, B. and Kot, A.C., 2017. Cross-domain shoe retrieval with a semantic hierarchy of attribute classification network. IEEE Transactions on Image Processing, 26(12), pp.5867-5881.

[12] https://www.kaggle.com/datasets/hasibalmuzdadid/shoe-vs-sandal-vs-boot-dataset-15k-images