
ISyE 6740 - Summer 2023
Project Proposal

Team Member Names: Vivi Banh & Chukwuemeka Okoli

Project Title: Deep Neural Network for Shoe Model Identification

Project Goal

The goal of this project is to build a deep neural network that can identify shoe models from images. The network will be trained on a dataset of images of shoes, and it will be able to classify new images with high accuracy.

Problem Statement

In today's fashion-forward society, social media plays a pivotal role in shaping consumer habits and driving trends. As people browse through fashion photos online, they seek convenient ways to identify and purchase the products showcased in those images. However, traditional methods of searching for unknown products on search engines can be frustrating and time-consuming.

Our project aims to address this challenge by developing an image classification system that can accurately identify Nike shoe models worn by models in photos and provide consumers with the name of the shoe models. With an upload of an image, users can swiftly access the exact name of the desired Nike shoe model, and this potential streamlined approach significantly enhances the shopping experience, making it easier for consumers to find and purchase the specific shoe models they are interested in.

Integrating our image classification system into popular platforms like Instagram or a web application can bring significant benefits to retailers and e-commerce platforms. By seamlessly incorporating a product identification feature, businesses can captivate potential buyers, enhance customer engagement, and drive conversions. This transformative capability bridges the gap between inspiration and purchase, transforming social media platforms into highly effective sales channels. The integration of our image classification system can create new opportunities for both retailers and consumers, maximizing sales potential and delivering an exceptional shopping experience.

Project Scope

The scope of the project will involve the following tasks:

- Collecting dataset of images of shoes.
- Preprocessing the images to prepare them for training.
- Designing and training a deep neural network using the Convolutional Neural Network (CNN) architecture.
- Testing the performance of the CNN model.
- Discussing the deployment of the trained model to a production environment.

Data Source

The data utilized for this project is the UT Zappos50K Shoe Dataset (ver 1.2), curated by Yu and Grauman (2014) and available on Kaggle.com. This comprehensive shoe dataset comprises 50,025 catalog images obtained from Zappos.com. The images have dimensions of 136 x 102 pixels, showcasing shoes centered on a white background and consistently oriented for ease of analysis. The dataset encompasses various shoe models, including the renowned brand, Nike. Specifically, our focus will be on images of Air Jordan 1, Air Force 1, and Air Max 1, which will serve as the foundation for designing and training our model. Furthermore, an independent set of images featuring models wearing both Nike and non-Nike shoes will be used to evaluate the performance of our trained model.

Methodology

We will utilize a Convolutional Neural Network (CNN) model, a deep learning architecture well-suited for image classification tasks. The CNN model will be trained using the labeled shoe images to learn the distinctive features of each shoe model. We will create our model using Python 3.10. The process can be summarized as follows:

1. Data preprocessing: The shoe images will undergo preprocessing steps, including resizing and normalization, to ensure they are in a suitable format for training. Data augmentation techniques such as vertical and horizontal reflections, rotation up to 90 degrees, and vertical and horizontal shifting of the images up to 20% of their original size will be applied to enhance the model's ability to generalize.

2. Model architecture: A CNN model will be designed, consisting of convolutional layers, pooling layers, and fully connected layers. The architecture will be tailored to capture the distinctive visual features of Nike shoe models. Techniques such as dropout and batch normalization may be incorporated to improve the model's generalization and prevent overfitting.
3. Model training: The designed CNN model will be trained using the labeled shoe images as input and their corresponding shoe model names as target labels. The model will learn to recognize the visual patterns associated with each shoe model during the training process. Training will utilize optimization algorithms such as stochastic gradient descent (SGD) or Adam optimization. The number of iterations over the dataset and batch size will be specified for efficient training.
4. Model evaluation: To assess the accuracy of the trained CNN model, we will evaluate its performance using a separate set of test images that feature models wearing various Nike shoes. Evaluation metrics such as categorical cross-entropy loss, accuracy, precision, recall, and F1 score will be calculated to measure the model's ability to correctly identify the shoe models.
5. Fine-tuning: If necessary, we will fine-tune the model by adjusting hyperparameters such as the learning rate, batch size, or the number of layers. Each configuration change will be evaluated for accuracy to identify the best-performing model for our final results.
6. Deployment: Considerations for deploying the trained model into a production environment (e.g., potential integration with a user-facing application) will be discussed if time permits.

Evaluation and Final Results

The accuracy of the trained model will be a key metric in evaluating its performance. We will compare the model's predictions for identifying shoe models with the true labels of the test images. To measure the accuracy of the model's predictions, we will use evaluation metrics such as categorical cross-entropy, which measures the difference between predicted probabilities and true labels, and accuracy, which calculates the percentage of correctly classified images. Additionally, we will also evaluate the precision, recall, and F1 score of the model. Precision represents the proportion of true positive predictions among all positive predictions, while recall measures the proportion of true positive predictions among all actual positive instances. The F1 score combines both precision and recall into a single metric, providing a balanced evaluation of the model's performance.

The final result of the project will be a well-trained model capable of accurately identifying Nike shoe models worn by models in photos. The success of the project will be determined by the model's ability to achieve a high level of accuracy in identifying the correct shoe models from the images, providing a solid foundation for further advancements in image classification and consumer-driven shopping experiences.

For future enhancements, the trained model can be deployed as a user-facing application to enable a user-friendly experience. Consumers will be able to upload images and receive the corresponding shoe model names as the output. This application will streamline the process of identifying and purchasing specific Nike shoe models, enhancing the shopping experience for consumers.

References

1. Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). Understanding of a Convolutional Neural Network. 2017 International Conference on Engineering and Technology (ICET), 1–6. <https://doi.org/10.1109/icengtechnol.2017.8308186>
2. Sharma, N., Jain, V., & Mishra, A. (2018). An Analysis Of Convolutional Neural Networks For Image Classification. *Procedia Computer Science*, 132, 377–384. <https://doi.org/10.1016/j.procs.2018.05.198>
3. Yu, A., & Grauman, K. (2014). Fine-Grained Visual Comparisons with Local Learning. *Computer Vision and Pattern Recognition*. <https://doi.org/10.1109/cvpr.2014.32>