Automated Footwear Classification

Baila Ly 4027963 Sanjay Thambithrurai 4018440 Benjamin Zitella 4021138 Jia Hao To 40263401 Rasel Abdul Samad 4020992

1. Problem Statement

The intent of footwear classification is to automate identification of footwear types from images, enabling efficient visual understanding for downstream applications. Footwear type classification from images is an important computer vision problem for applications in e-commerce, visual search, and product categorization. However, the task is challenging due to large visual variations of the same footwear categories, and diverse imaging conditions present.

The objective of this project is to evaluate the performance and scalability of various convolutional neural networks (CNN) architectures for footwear images classification across datasets of different complexity. We aim for the results to provide insights into the relationship between the dataset scale and model capacity, such their strengths and limitations of various CNNs when applied to the real-world footwear classification task.

2. Dataset Selection

We selected three shoe classification datasets downloadable from Kaggle with varying complexity. All use JPEG format with RGB color space.

Dataset	Cls	Total	Img/Cls	Res.
Footwear 3K [1]	3	3,000	1,000	136×102
Shoes Classification [2]	5	13,715	2,600-2,900	Variable
UT Zappos50K [3]	12	50,025	1,500-2,000	136×102

Table 1. Footwear datasets with varying class granularity.

The pre-defined train/validation/test folder splits provided by the Shoes Classification Dataset will be used exactly as is. There are no pre-defined splits in the Footwear: Shoe vs. Sandal vs. Boot and UT Zappos50K datasets. During preprocessing, we will define stratified 70/15/15 train/validation/test divides for the Footwear dataset. Using *stratified random sampling*, we will first subsample the 50,025 images for UT Zappos50K. We will extract the 12 most common functional subcategories from the metadata (e.g., running shoes, high heels, loafers, and oxfords), then randomly sample 1,500–2,000 images per subcategory without replacement (seed=42), and create 70/15/15 splits.

3. Possible Methodology

3.1. Pipeline

Our pipeline focuses on comparing models trained from scratch to those utilizing transfer learning .Three architecture will be implemented: ResNet-50[4], MobileNetV2[5], and VGG-16[6]. All three models will be trained from scratch accross three different datasets of increasing class granularity with a total of nine models. These models will be evaluated against two transfer learning versions (ResNet-50 and MobileNetV2) to assess efficiency pretrained weigths in the retail imagery domain.

3.2. Data Processing and Training

Images will be resized to 224 x 224[7] and normalized using ImageNet mean and standard deviation to guarantee compatibility with pre-trained weights[8]. A 70/15/15 train/validation/test split will be utilized in order to prevent overfitting. Real-time augmentations such as horizontal flips, ±20°rotations and zooming will be implemented to improve generalization and reduce overfitting and further mitigate overfitting. Training will be performed using the loss algorithim Adam optimizer with a learning rate of $1 \cdot 10^{-4}$[9] and the loss function $CrossEntropyLoss$[10] loss over 50 epochs. The MobileNetV2 will be will be specifically targetted during hyperparameter tuning, optimizing batch size(16,32,64)[11] and dropout rates to improve performance.

4. Gantt Chart

Figure 1. Gantt Chart for Project Timeline

5. Bibliography

References

[1] https://www.kaggle.com/datasets/adityakadam1/footwear. 1

[2] https://www.kaggle.com/datasets/utkarshsaxenadn/shoes-classification-dataset-13k-images. 1

[3] https://vision.cs.utexas.edu/projects/finegrained/utzap50k/. 1