Automated Footwear Classification

Baila Ly 4027963 Sanjay Thambithrurai 4018440 Benjamin Zitella 4021138 Jia Hao To 40263401 Rasel Abdul Samad 40209924

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 as its strength 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. Methodology

Our pipeline utilizes a transfer learning approach to handle the diverse resolutions and backgrounds of the aggregated data.All images will be standardized to 224 x 224 pixels and normalized. To prevent overfitting on smaller brand-specific sets,we will apply real-time data augmentation (flips, rotations, and brightness shifts). We will benchmark a custom baseline CNN against a pre-trained ResNet-50 architecture to leverage its residual learning capabilities for complex feature extraction. [To be changed]

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