Fashion Classification Using Transfer Learning

Umnia Hameed
Mount Holyoke College
South Hadley, MA
hamee22u@mtholyoke.edu

Hema Motiani
Mount Holyoke College
South Hadley, MA
motia22h@mtholyoke.edu

***Abstract*—We worked on an image classification pipeline built using transfer learning on the DeepFashion dataset using EfficientNet-B0 and two training strategies. Our project demonstrated a considerable boost in performance from 16.89 percent to 52.05 percent accuracy. Our results show the potential of domain-specific data augmentation and model tuning in enhancing classification accuracy for real-world fashion images.**

I. Introduction

The fashion industry increasingly relies on AI to automate and enhance tasks such as product tagging, recommendation systems, and visual search. This project builds a classification model that can identify high-level fashion categories from real-world images, serving e-commerce platforms and fashion retailers. Automated fashion classification not only streamlines inventory management but also enables more personalized shopping experiences, ultimately benefiting both consumers and retailers. We chose to work on this project because it blends technical challenge with creative application—an ideal opportunity to apply deep learning techniques to a real-world, high-impact domain. The intersection of fashion and AI intrigued us due to its rapidly growing importance in e-commerce and the complexity of recognizing nuanced clothing types in diverse settings. Additionally, building a model that classifies fashion items from real-world images allowed us to explore critical deep learning concepts like transfer learning, data preprocessing, and model generalization. This project provided a meaningful and technically rich way to enhance our machine learning skills while contributing to something we were interested in outside the tech world.

II. Previous Work

Prior to beginning our project, we examined several studies in the area of fashion classification and recommendation systems. Roy et al. [1] proposed a deep learning-based fashion recommendation system tailored to different body types, highlighting the importance of personalized clothing suggestions. Liu et al. [2] introduced the DeepFashion dataset and demonstrated its effectiveness in fine-grained attribute recognition and retrieval. Kiapour et al. [3] explored matching street-style photos with online product listings, emphasizing the role of diverse, annotated datasets in improving real-world applicability. Initially, we experimented with a small fashion dataset and trained a CNN model from scratch. While it achieved good accuracy, the dataset's limited diversity and coarse labels resulted in poor generalization. To overcome this, we adopted the DeepFashion dataset, which offers a wider range of clothing styles, detailed annotations, and greater visual variability. These enhancements enable more robust learning and improved performance on complex classification tasks, making it a better fit .

III. DataSet

We used the DeepFashion dataset [4], developed by the CUHK Multimedia Lab, which contains over 800,000 fashion images annotated with category labels, bounding boxes, landmarks, and clothing attributes. The dataset includes both consumer and shop images, capturing a wide range of real-world visual conditions such as varying poses, lighting, and backgrounds. These characteristics made DeepFashion particularly well-suited for our project, as it provided the diversity and annotation granularity necessary to train robust models capable of handling complex fashion classification tasks. To tailor the dataset to our needs, we applied several restructuring steps. First, we mapped the original fine-grained categories into 50 higher-level fashion classes to simplify the classification task while retaining meaningful distinctions between item types. We then selected the top 20 most represented classes to ensure class balance and reduce model bias. Finally, we applied an 80/20 stratified split to divide the data into training and testing sets while preserving class distribution. Since we were working in TensorFlow, all images were converted into tensors, and stratified sampling ensured that each class remained proportionally represented throughout the dataset pipeline. In addition, we performed standard preprocessing procedures such as image resizing, normalization, and data augmentation to enhance model generalization and training robustness. The size, diversity, and annotation quality of DeepFashion, along with these preprocessing steps and TensorFlow integration, provided a strong foundation for building an effective fashion classification system.

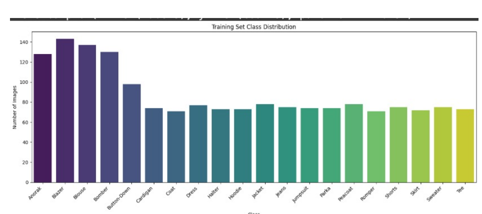

Fig. 1. Distribution of the top 20 fashion classes used in our experiments from the DeepFashion dataset.

IV. Method

We trained two models using the EfficientNet-B0 [5] architecture under different training strategies to evaluate the impact of transfer learning and fine-tuning techniques on classification performance. The first model applies a classic transfer learning approach. It leverages the pretrained EfficientNet-B0 network, freezing all layers except for the final fully connected (FC) layer, which is replaced with a new classification head tailored to the top 20 DeepFashion classes. Moderate data augmentation—including resizing, random cropping, horizontal flipping, and color jittering—is applied to increase generalization. The model is trained for 10 epochs using the Adam optimizer with a learning rate of 0.001. Only the new FC layer is updated during training. No learning rate scheduling or early stopping is used. Evaluation is performed using accuracy and confusion matrix metrics. The second model incorporates a more refined fine-tuning approach. Instead of freezing the entire base model, it unfreezes the last two blocks (blocks 6 and 7), along with the convolutional head and final FC layer, allowing selective adaptation of deeper features. Preprocessing includes lighter but more varied augmentations such as horizontal flips, slight rotations, and color jitter. A learning rate scheduler is applied to reduce the learning rate as training progresses, and L2 regularization is added to minimize overfitting. This model uses the Adam optimizer with a lower learning rate (0.0005), and early stopping is implemented to preserve the best-performing checkpoint based on validation accuracy.

V. Evaluation

Accuracy was the primary evaluation metric, assessed on the held-out test set. Model 1 achieved an accuracy of 16.89 percent, indicating significant underfitting and limited generalization capacity, likely due to its restricted training setup and lack of data augmentation. Model 2, which incorporated fine-tuning of the last two blocks along with regularization and data augmentation, showed a marked improvement with an accuracy of 52.05 percent. In addition to accuracy, Model 2 achieved a precision of 0.62, recall of 0.59, and an F1 score of 0.59, reflecting a more balanced and effective classification performance across the top 20 fashion categories. Despite this progress, the model still exhibited a noticeable number of misclassification errors, particularly among visually similar classes such as blazers and anoraks or cardigans and coats. These errors highlight the need for improved feature differentiation and potentially merging overlapping categories. Addressing these misclassifications remains a key area for future enhancement, whether through refined labeling, additional training data, or more sophisticated model architectures.

VI. Results

Through this project, we gained valuable experience working with a large-scale dataset and learned the importance of effective preprocessing in building scalable and accurate deep learning models. Handling the DeepFashion dataset, which contains over 800,000 images, presented several challenges related to data volume, structure, and class imbalance. To make

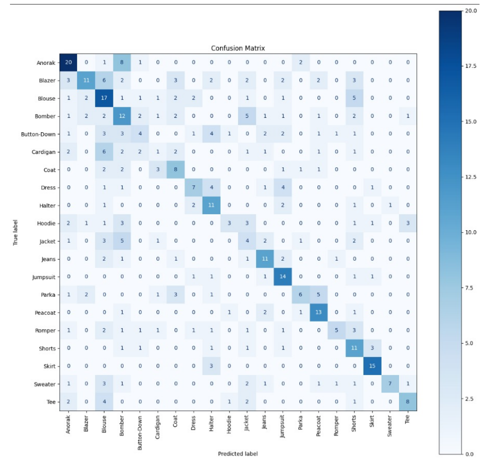

Fig. 2. Confusion matrix for Model 1. Misclassifications are widespread, indicating limited feature learning and underfitting.

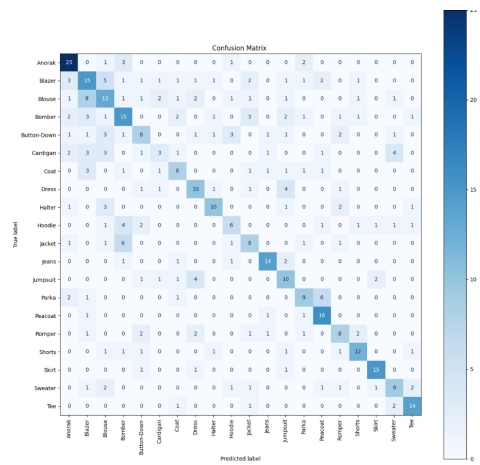

Fig. 3. Confusion matrix for Model 2. Diagonal dominance reflects improved accuracy, though confusion between visually similar categories persists.

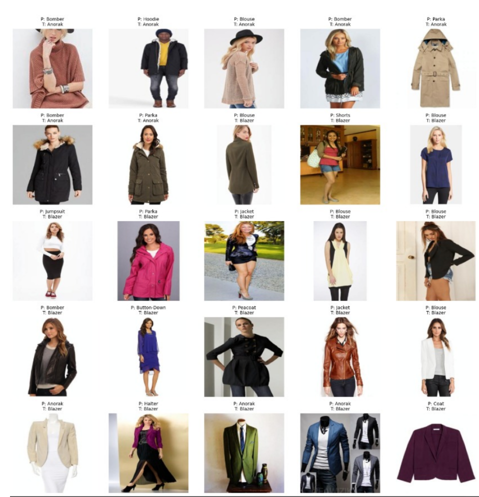

Fig. 4. Screenshot highlighting sample misclassifications made by the model. Common confusions occur between visually similar categories such as blazers vs. anoraks and cardigans vs. coats.

the dataset more manageable and meaningful, we developed a Python script to remap overly specific category labels into 50 broader classes and then selected the top 20 most represented ones for training. We also performed an 80/20 stratified split to ensure balanced representation across both training and test sets. Preprocessing steps such as resizing, normalization using ImageNet statistics, and applying data augmentation were essential for improving generalization and reducing overfitting. Fine-tuning the pretrained EfficientNet-B0 model with regularization techniques and learning rate scheduling resulted in substantial performance improvements over the baseline. This experience highlighted the critical role of model architecture choices, data handling, and training strategies in solving real-world classification problems. Ultimately, we learned that building a robust classification system for complex datasets not only involves selecting the right model, but also requires thoughtful preprocessing, careful evaluation, and iterative improvement based on observed model behavior.

VII. Discussion

The results clearly show that Model 2 significantly outperforms Model 1, emphasizing the importance of fine-tuning and regularization in improving model adaptability and overall performance. While Model 1 struggled to learn meaningful features—likely due to limited training and the lack of augmentation—Model 2 benefited from transfer learning, data augmentation, and targeted fine-tuning of deeper layers, resulting in improved generalization across diverse clothing categories. Because the model still faces challenges in distinguishing between visually similar categories Future improvements could involve merging such confusing categories, enhancing image augmentation strategies to expose the model to greater visual variability, and extending training over more epochs to allow deeper learning of class-specific features. Additionally, introducing more sophisticated architectures or attention mechanisms may help the model focus on subtle discriminative regions within garments. Exploring class-specific metrics and analyzing confusion matrices in more depth could also guide targeted improvements. Overall, while Model 2 marks a significant step forward, continued refinements are necessary to achieve high accuracy and reliability in real-world applications.

Summary

We developed a fashion classification model using transfer learning on the DeepFashion dataset, exploring different architectures and fine-tuning strategies. Our experiments showed that incorporating data augmentation and progressively unfreezing layers during fine-tuning consistently improved model performance, achieving higher accuracy and better generalization. Detailed analyses highlighted challenges with overlapping categories and class imbalances, which impacted certain labels' precision and recall. Through iterative training and validation, we identified that extending training epochs and refining category definitions could further enhance results. Overall, the combination of transfer learning, data augmentation, and careful model tuning proved effective for robust fashion item classification, setting a solid foundation for future improvements.

References

[1] S. Roy and S. Jha, "Advanced Fashion Recommendation System for Different Body Types using Deep Learning Models," 2022. [Online]. Available: https://www.researchgate.net/publication/362086487_Advanced_Fashion_Recommendation_System_for_Different_Body_Types_using_Deep_Learning_Models

[2] Z. Liu, P. Luo, S. Qiu, X. Wang, and X. Tang, "DeepFashion: Powering Robust Clothes Recognition and Retrieval," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1096–1104. [Online]. Available: https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Liu_DeepFashion_Powering_Robust_CVPR_2016_paper.pdf

[3] M. H. Kiapour, X. Han, S. Lazebnik, A. C. Berg, and T. L. Berg, "Where to Buy It: Matching Street Clothing Photos in Online Shops," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2015, pp. 3343–3351. [Online]. Available: https://www.researchgate.net/publication/304408709_Where_to_Buy_It_Matching_Street_Clothing_Photos_in_Online_Shops

[4] CUHK Multimedia Lab, 'DeepFashion Dataset," [Online]. Available: https://mmlab.ie.cuhk.edu.hk/projects/DeepFashion.html

[5] Keras Documentation, "EfficientNetB0," [Online]. Available: https://keras.io/api/applications/efficientnet/#efficientnetb0-function. [Accessed: May 18, 2025].