

# Deep Learning for Cloth Recognition and Gender Prediction

Preethy Ann Thomas  
PG Scholar  
Amal Jyothi College of Engineering  
Kottayam, Kerala  
preethyanthomas@gmail.com

Sruthimol Kurian  
Assistant Proffessor  
Department of Computer Applications  
Amal Jyothi College of Engineering  
Kottayam, Kerala  
sruthimolkurian@amaljyothi.ac.in

**Abstract**— *The swift advancement of computer vision and machine learning methodologies has stimulated inventive applications in a variety of fields. This work presents a novel method that redefines the field of fashion product detection and description inside photographs by utilizing cutting-edge technologies such as deep learning and image processing. With important ramifications for sectors including e-commerce, content creation, and fashion research, the algorithm provides a comprehensive approach to automating product detection and description. The methodology uses pre-trained deep learning models—ResNet-50 in particular—for transfer learning, which gives the algorithm the ability to understand and recognize a wide variety of items and their characteristics. With the use of a multi-phase procedure that includes data collection, preprocessing, modeling, and iterative improvements, the algorithm performs very well in the accurate and naturalistic identification and representation of fashion items. This expertise not only makes purchasing more enjoyable for the consumer, but it also offers priceless insights into the always changing fashion industry.*

**Keywords**— *Deep Learning, ResNet-50, Image Processing*

## I. INTRODUCTION

The convergence of computer vision, deep learning, and image analysis technologies has ushered in an era of revolution across many sectors in the modern digital world, where visual material predominates. This research presents a novel algorithm that reinvents how we interact with fashion items by utilizing this state-of-the-art technology. Inherently visual and constantly changing, the fashion industry offers both interesting difficulties and promising opportunities for technology advancement. In addition to being a technological accomplishment, the ability to automatically recognize, classify, and describe fashion goods from photos offers up a wide range of possibilities for e-commerce, content development, fashion analysis, and other uses

This technique exemplifies the convergence of deep learning, a branch of machine learning that has advanced image analysis, and computer vision, a field dedicated to allowing robots to comprehend visual information. ResNet-50, a pre-trained deep learning model that uses transfer learning to give it the ability to detect a wide range of fashion goods and their intrinsic characteristics, is at the core of the algorithm's capabilities. The approach is made more effective by the seamless integration of image processing techniques by improving picture quality, performing data augmentation, and extracting crucial information from visual data. The ability of the algorithm to accurately identify prominent garment colors while ignoring background and shadow colors is a key component of its competence.

## II. RELATED WORKS

Hidayati, W.-H. Cheng, and K.-L. Hua provide a novel approach for automated identification of clothing genre in full-body images. [1] Their methodology uses supervised learning to construct genre classifiers by utilizing certain visual cues from garment design theory. It demonstrated an average accuracy of around 88% for upperwear and lowerwear genres when tested on a dataset of 3250 full-body photos. In order to identify genres more precisely, the authors want to improve face recognition and add novel components like attractiveness evaluation. The real-world applications of this study in e-commerce and fashion appear promising.

Bossard et al. present a pipeline for recognizing clothing in natural scenes, with applications in e-commerce, event recognition, and advertising. [2] Their approach combines upper body detection, random forests for clothing type, and SVMs for style attributes. It leverages web-crawled data and achieves 41.38% average accuracy on an 80,000-image benchmark dataset. This work fills a vital role in clothing recognition for diverse applications.

A technique for creating realistic cloth models from drawings using composite templates is presented in the study by Hong Chen et al. (2006). [3] This creative method allows for a wide variety of fabric patterns and textures by bridging the gap between drawing and computer graphics. With real-time textile synthesis and modification, it has important applications in computer graphics and fashion design. The creation of models and drawing of clothing has advanced significantly with this effort.

The authors of [4] convey DeepFashion, an enormous apparel dataset with an annotation richness and scale that surpasses previous collections. With more than 800,000 photos featuring detailed classification, many qualities, locations, and cross-domain picture correspondences, DeepFashion is a useful tool. The authors provide FashionNet, a deep learning network that efficiently learns garment attributes by utilizing these annotations. Robust garment representations are produced by FashionNet by combining large attribute learning with landmark predictions. In order to demonstrate the effectiveness of the model and the significance of DeepFashion for further study, the authors offer benchmark datasets for apparel recognition and retrieval.

A framework for automatically identifying clothing genres in full-body person images is presented in the study [5]. Feature extraction, tag assignment, face and upper-body detection, and online shop photos are used in the training phase. The system recognizes and categorizes different clothing genres in input photos during the runtime phase. On

a dataset of more than 1,000 photos, the method attained a high accuracy of 92.04%, recall of 92.45%, and an F-score of 92.25%, demonstrating its efficacy in clothing genre recognition. The field of automated garment recognition benefits from this work[5].

### III. METHADODOLOGY

#### A. Data Collection

In the data collection phase, we curate a diverse dataset of fashion product images from multiple sources, including e-commerce platforms, fashion magazines, publicly available datasets, and collaborations with fashion retailers. Each image is meticulously annotated with detailed information, encompassing product type and gender category.

#### B. Data Preprocessing

The data preprocessing step involves a series of critical procedures to enhance the quality and consistency of raw images. Techniques include noise reduction, background removal, color correction, resolution and size standardization, and metadata handling. These steps ensure that the dataset is well-prepared for machine learning model training.

#### C. Data Augmentation

Data augmentation is a critical step in enhancing the performance and generalization capability of machine learning models, particularly in computer vision tasks like fashion product recognition. This process involves applying various transformations to the original images in the dataset, creating new training samples with slight variations. The goal is to expose the model to a more diverse range of images during training, making it more robust and less prone to overfitting. Here are the key data augmentation techniques used in the algorithm:

##### a. Image Rotation:

Rotation involves changing the orientation of the image by a random degree, typically within a specified range, such as  $\pm 10$  degrees. By rotating images, we simulate different angles at which fashion products might appear in photographs. This helps the model become invariant to the rotation of objects within the images.

##### b. Horizontal and Vertical Flipping:

Flipping images horizontally and vertically introduces variations in the orientation of fashion products. For example, a left-facing clothing item becomes right-facing when flipped horizontally. This technique ensures that the model learns from images with different orientations.

##### c. Random Cropping:

Random cropping involves taking random crops of the original images. These crops can occur at various positions within the image, simulating different perspectives and zoom levels. Cropping allows the model to learn from different portions of the clothing items and adapt to variations in image composition.

##### d. Brightness and Contrast Adjustments:

Adjusting brightness and contrast levels randomly within a certain range helps the model handle different lighting conditions in real-world images. This augmentation technique ensures that the model can recognize fashion products in various lighting environments.

##### e. Gaussian Noise:

Adding Gaussian noise to images simulates noise that may be present in real photographs. This augmentation makes the model more robust to noisy input data, helping it generalize better to real-world scenarios.

##### f. Color Jitter:

Color jitter involves changing the colors in images by adjusting hue, saturation, and brightness randomly. This technique diversifies the color palette the model is exposed to during training, making it more adaptable to variations in clothing colors.

##### g. Scaling and Resizing:

Scaling and resizing images to different dimensions helps the model learn from clothing items of varying sizes and resolutions. This augmentation ensures that the model can recognize products regardless of their scale in images.

##### h. Shearing:

Shearing introduces distortions that mimic perspective changes in images. This technique is valuable for training the model to recognize clothing items photographed from different angles.

#### D. Model Development

The machine learning model at the core of the algorithm is constructed using a Convolutional Neural Network (CNN) architecture, specifically ResNet-50. ResNet-50 is a well-established CNN architecture known for its remarkable performance in various image recognition tasks.

##### a. Model Architecture:

The chosen model architecture, ResNet-50, is a deep neural network consisting of multiple layers, including convolutional layers, pooling layers, and fully connected layers. This architecture is designed to automatically learn hierarchical features and patterns from input images, making it highly suitable for image classification tasks.

1. **Convolutional Layers:** Convolutional filters are applied to input images to detect various visual patterns and features, such as edges, textures, and shapes.
2. **Pooling Layers:** Pooling layers reduce the spatial dimensions of feature maps, helping to reduce

computational complexity and making the model more robust to variations in object position.

3. Fully Connected Layers: These layers flatten the output from the convolutional layers and pass it through one or more fully connected (dense) layers for the final classification.

*b. Loss Function:*

To train the model effectively for clothing product classification, the categorical cross-entropy loss function is employed. This loss function quantifies the dissimilarity between the predicted class probabilities and the true class labels. By minimizing this loss during training, the model learns to make accurate predictions and classify fashion products into their respective categories.

*c. Optimization Algorithm*

Stochastic Gradient Descent (SGD) is utilized as the optimization algorithm in the training process. SGD iteratively updates the model's weights to minimize the loss function, allowing the model to converge towards an optimal solution. The choice of learning rate, a crucial hyperparameter in SGD, is carefully considered and tuned to ensure efficient and effective training.

*d. Training*

The training process involves multiple epochs, during which the model is exposed to the augmented training dataset. The model iteratively learns from the data, adjusting its internal parameters to improve its ability to recognize and classify fashion products accurately. Training for multiple epochs allows the model to refine its representations and learn complex patterns present in the images.

*E. Gender Prediction*

DeepFace, a potent deep learning library with expertise in accurate gender prediction from facial traits in photos, is the algorithm used in this case. Prominent for its remarkable precision, DeepFace employs sophisticated algorithms to classify gender according to key features of the face. DeepFace is a key technology that's widely utilized in fields where accurate gender analysis from visual data is crucial.

*F. Color Detection*

The algorithm transforms the way that fashion goods are seen and characterized in photos by integrating a multi-step method that combines deep learning, image processing, and color analysis. The system uses transfer learning with pre-trained deep learning models (ResNet-50), which allows it to comprehend and identify a wide variety of items and their characteristics. Using the CIEDE2000 color difference formula to make accurate color differences is a crucial part of the research. Moreover, garment colors are separated from the backdrop using clustering algorithms and the LAB ( $L^*$ ,

$a^*$ ,  $b^*$ ) color space. The algorithm optimizes efficiency in the context of fashion analysis by deftly handling shadows and complementary colors. It also calculates and displays the clothing's prevalent colors, offering insightful information about color trends.

*G. Caption Generation*

The algorithm uses a flexible function to create engaging product descriptions depending on apparel colors and product categories. Writing interesting text descriptions for different fashion goods is the main goal. The program carefully creates color-based captions, such 'A [color] [product category],' given a certain product category and a range of apparel colors, to emphasize the product's aesthetic appeal. Furthermore, it adds variation to the text generating process by arbitrarily producing a wide array of generic captions that encapsulate style and fashion. The function builds a pool of intriguing descriptions by combining the color-based and random captions, from which a random option is chosen. This dynamic caption generating method improves the way fashion goods are presented and marketed by accommodating a wide range of style preferences and events.

*H. Evaluation*

We evaluate the algorithm's performance using quantitative metrics such as accuracy, precision, recall, and F1-score for product recognition.

## IV. RESULTS AND DISCUSSIONS

The algorithm represents a pioneering solution that combines computer vision, deep learning and image processing techniques to revolutionize fashion product recognition and description. Its applications extend to e-commerce, content creation, and fashion analysis, offering users an enhanced shopping experience and valuable insights. Through a comprehensive methodology and iterative improvement process, the algorithm excels in recognizing and describing fashion products accurately and intuitively.



Figure 1: Output screen that displays cloth type, gender, color of cloth and caption generated for cloth.

## V. CONCLUSION

This study offers a novel method that improves the detection and description of fashion products in images by fusing computer vision and deep learning. Its objective is to enhance fashion analysis, content production, and e-commerce. The algorithm is quite good at recognizing and describing fashion products with sophisticated methods.

This research has an extensive and promising future. Firstly, the algorithm presents chances to improve the online purchasing experience through its real-time deployment in mobile applications and e-commerce platforms. Users may obtain fast product descriptions and color analyses by seamlessly integrating this technology, which empowers them to make more educated purchase decisions. In addition, the algorithm may be expanded to include user-generated content, which would enable users to contribute their own fashion photos for evaluation and style suggestions. This would encourage user interaction and customization. Finally, by combining color information and descriptions, the technology's applications reach into the field of fashion trend research, providing insightful information that helps the industry foresee and proactively address changing trends.

## VI. REFERENCES

- [1] S. C. Hidayati, W.-H. Cheng, and K.-L. Hua, "Learning and Recognition of Clothing Genres from full body images" in *IEEE Transactions on Cybernetics*, May 2018.
- [2] Lukas Bossard, Matthias Dantone, Christian Leistner, Christian Wengert, Till Quack, Luc Van Gool, "Apparel Classification with Style", in "Asian Conference on Computer Vision", Daejeon, Korea, November 5-9, 2012.
- [3] Hong Chen, Zi Jian Xu, Zi Qiang Liu, Song Chun Zhu, "Composite Templates for Cloth Modelling and Sketching" in *IEEE Conference*, June 2006.
- [4] Liu, Z., Luo, P., Qiu, S., Wang, X., Tang, X, "Deepfashion: Powering robust clothes recognition and retrieval with rich annotations", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Las Vegas, NV, USA, 2016, pp.1096-1104.
- [5] S. C. Hidayati, W.-H. Cheng, and K.-L. Hua, "Clothing genre classification by exploiting the style elements" in *Proceedings of ACM International conference on Multimedia*, Nara, Japan, 2012, pp. 1137–1140.