Computer Vision and Deep Learning for Cloth Recognition and Gender Prediction

Preethy Ann Thomas
PG Scholar
Amal Jyothi College of Engineering
Kottayam, Kerala
preethyannthomas@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, image processing, and natural language processing (NLP). 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— ***NLP, ResNet-50***

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 colours while ignoring background and shadow colours is a key component of its competence.

II. RELATED WORKS

Early work in clothing recognition focused on clothing segmentation and localization, with the goal of defining clothing regions in images. Clothing retrieval methods were developed to find similar clothing items in response to input clothing images. Street-to-shop image retrieval addressed the discrepancy between user photographs and clothing images from online stores. Vittayakorn et al. introduced an approach to learning outfit similarity, based on descriptors for colour, texture, and shape, aiming to measure the similarity of clothing images.

Several studies have been conducted in the field of clothing recognition. Zhang et al. developed a clothing recognition system, focusing on upperwear in controlled fitting room environments. Yang and Yu introduced a video content analysis system to locate human figures, segment clothing regions, and recognize clothing categories. Hidayati et al. presented an algorithm to discover visual style elements representing fashion trends.

Fashion parsing studies also played a role in clothing recognition. Yamaguchi et al. assigned semantic labels to each pixel in images, recognizing clothing items using a retrieval-based approach. Simo-Serra et al. proposed segmenting different garments worn by individuals, considering the dependencies between clothing and human pose. Liang et al. addressed human parsing tasks, predicting 18 labels for each pixel.

Recent advances in clothing attributes recognition were made by researchers such as Yamaguchi et al. and Chen et al. They introduced style rules models based on conditional random fields (CRF) and the combination of clothing attributes. Importantly, they highlighted the importance of considering pose information in clothing recognition.

Chen et al. performed human pose estimation but were limited to straight-frontal pose images. In contrast, the proposed approach in this paper accommodates non-straight-frontal poses, converting them into straight-frontal poses. Features representing clothing attributes are extracted from specific human body parts to ensure consistent representations.

III. METHADOLOGY

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, colour 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 ±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. Colour Jitter:

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

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. Colour 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 colour 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 colour difference formula to make accurate colour differences is a crucial part of the research. Moreover, garment colours are separated from the backdrop using clustering algorithms and the LAB (L*, a*, b*) colour space. The algorithm optimizes efficiency in the context of fashion analysis by deftly handling shadows and complementary colours. It also calculates and displays the clothing's prevalent colours, offering insightful information about colour trends.

G. Caption Generation

The algorithm uses a flexible function to create engaging product descriptions depending on apparel colours and product categories. Writing interesting text descriptions for different fashion goods is the main goal. The program carefully creates colour-based captions, such 'A [colour] [product category],' given a certain product category and a range of apparel colours, 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 colour-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

In conclusion, the algorithm represents a pioneering solution that combines computer vision, deep learning, image processing, and NLP 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, colour of cloth and caption generated for cloth.

V. 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.