Journal of Fiber Bioengineering and Informatics 19:1 (2026) 13–30

Cite as: X. Guo et al.
doi:10.3993/jfbim26008

A YOLOv8-Based Clothing Detection Framework Incorporating Semantic Uncertainty Modeling

Xianfei Guo, Chen Ting, Hong Yan*

College of Textile and Clothing Engineering, Soochow University, Suzhou 215021, China

Abstract

Garment style recognition is critical for intelligent retail but suffers from semantic uncertainties caused by high intra-class similarity and non-rigid deformations. Existing detection and attention-based methods often apply uniform attention across all feature levels. This approach limits fine-grained discrimination in complex scenarios. To address this gap, this study proposes a hierarchical attention-enhanced YOLOv8 framework to optimize recognition precision. By integrating a Convolutional Block Attention Module (CBAM) at the mid-level and an Efficient Channel Attention (ECA) module at the high-level, the model effectively strengthens structural perception and suppresses semantic channel dispersion. Experimental results show that the proposed method achieves 84.99% mAP@0.5, an 11.82% improvement over the baseline, while maintaining real-time performance at 126 FPS. This framework improves fine-grained garment recognition and provides a practical solution for intelligent retail applications in complex scenarios.

Keywords: Garment style recognition; YOLOv8; Hierarchical attention mechanism; Semantic uncertainty; Intelligent fashion retail

1 Introduction

1.1 Background

Garment style is a high-level semantic attribute characterized by structural contours, local textures, proportional patterns, and fine-grained design details. Garment style recognition aims to identify these attributes automatically and supports intelligent retail and personalized recommendation systems. Due to its high sensitivity to subtle visual features and the stringent requirements for robustness in complex environments, garment style recognition remains a key research topic in fashion informatics [1]. Unlike general object detection, garment style recognition requires precise differentiation of subtle variations in structural patterns, material textures, and garment silhouettes. This imposes greater demands on models' feature representation capabilities [2].

*Corresponding author.
Email address: hongyan@suda.edu.cn (Hong Yan).

1940–8676 /
Mar. 2026

Existing studies face two major challenges. First, fine-grained attribute representation is difficult due to the high intra-class similarity among garment categories. It is essential to extract highly discriminative features from localized structural cues [3]. Second, semantic uncertainty arising from non-rigid deformations and real-world conditions significantly degrades model performance. Human pose variation, occlusion, illumination changes, and fabric folds can alter the appearance of garments, leading to misclassification, localization errors, and reduced robustness. [4, 5]. Previous research has shown that accurate garment structural modeling and estimation of fabric mechanical parameters are critical for reliable analysis under these variations [5]. Such techniques have been widely applied in virtual fitting and personalized garment design systems [6, 7].

Although YOLOv8 offers a favorable balance between speed and accuracy, it still exhibits limitations in garment-style recognition, including the loss of local details and insufficient modeling of global context [8]. Attention mechanisms such as CBAM and ECA have been introduced for feature recalibration [9-11]. However, prior studies often apply a single attention module uniformly across hierarchical levels, overlooking the inherent heterogeneity of features. Mid-level features retain rich spatial structures that are crucial for accurate localization. In contrast, high-level features are more abstract, where modeling inter-channel semantic dependencies is essential for distinguishing fine-grained categories. Related research has further explored garment style recognition by integrating image processing and machine learning techniques, with preliminary applications in e-commerce scenarios [12, 13].

To address these limitations, this study proposes a hierarchical attention-enhanced YOLOv8 framework for garment style recognition. Specifically, CBAM is introduced at the mid-level to strengthen structural representation, while ECA is incorporated at the high-level to model local cross-channel dependencies. This design is intended to improve feature discrimination under semantic uncertainty.

The aim of this study is as follows:

(1) To develop a hierarchical attention-enhanced YOLOv8 framework for garment style recognition through the integration of adaptive attention modules at different feature levels.

(2) Construct a specialized garment image dataset characterized by complex backgrounds and fine-grained category challenges, to provide an effective evaluation platform for research in garment style recognition.

(3) Test the effectiveness of the hierarchical attention-enhanced strategy on fine-grained recognition performance while maintaining real-time inference efficiency.

1.2 Related Work

To address challenges such as fine-grained feature discrimination and complex background interference in garment style recognition, extensive research has been conducted in recent years focusing on detection framework design, feature enhancement mechanisms, and attention modelling strategies. This section systematically reviews relevant literature across three dimensions, analysing the strengths and limitations of existing methodologies and establishing the research foundation for the proposed hierarchical attention-enhanced YOLOv8 framework.