Data Science and Engineering (DSE) Record, Volume 6, issue 1.

Clothes Styling Assistant Based on Celebrities' Styles Using Computer Vision and Deep Learning

Tinnapat Jaimunt [1] and Varin Chouvatut [2]

[1] Master's Degree Program in Data Science, Chiang Mai University, Chiang Mai, Thailand
[2] Department of Computer Engineering, Faculty of Engineering, Chiang Mai University, Chiang Mai, Thailand
tinnapat_jai@cmu.ac.th

Abstract. This study presents a virtual fitting room application that recommends outfits based on celebrity fashion trends using computer vision and deep learning. The system applies hybrid segmentation (semantic and instance) for accurate clothing detection and uses CLAHE preprocessing to enhance image quality. An ensemble model combining CNN, plain NN, SVM, and Random Forest is used for style classification. The application provides users with a style similarity score compared to celebrity outfits. Evaluation through cross-validation and accuracy metrics shows improved performance, highlighting the potential of this approach for intelligent fashion recommendation systems and future Metaverse applications.

Keyword: Computer Vision, Fashion, Styling, Virtual Fitting Room, Deep Learning, Machine Learning, Image Processing.

1 Introduction

Fashion is a reflection of personal identity and influences social perception. However, selecting suitable outfits remains a challenge, especially for those lacking fashion knowledge. This research introduces a Clothes Styling Assistant based on celebrity fashion, supported by deep learning and computer vision.

The proposed system addresses the inefficiencies of traditional shopping and online fitting by offering to upload their photos and receive outfit recommendations based on the style of selected celebrities. By using hybrid segmentation techniques (semantic + instance), the system accurately identifies clothing regions in the image. It then uses a deep learning ensemble approach—combining CNNs, Support Vector Machines (SVM), Random Forest, and a plain Neural Network—to classify the detected clothing and match it with a curated database of celebrity fashion styles.

Additionally, the system incorporates Contrast Limited Adaptive Histogram Equalization (CLAHE) during image preprocessing, enhancing the clarity of patterns and textures in the image, which improves classification performance.

The main goals of this research are to:

- Enhance the efficiency and accessibility of fashion recommendation systems.

- Improve user confidence in outfit selection by referencing popular celebrity looks.

- Leverage hybrid segmentation and ensemble learning to maximize accuracy in clothing classification.

2 Literature Review

In developing a personalized fashion recommendation system grounded in computer vision and deep learning, it is essential to review the key technological components that underpin the system's architecture and methodology. This includes prior studies and technologies in image segmentation, style classification, fashion attribute extraction, and ensemble learning, as well as the influence of celebrity fashion trends in recommendation systems.

2.1 Computer Vision and Fashion Recognition

Computer vision techniques have long been applied to the fashion domain, particularly in object detection, segmentation, and classification of clothing items. Kiapour et al. [5] introduced the concept of fashion recognition through the "Hipster Wars" study, which demonstrated the feasibility of using visual features to classify clothing into specific fashion styles. Similarly, Bossard et al. [3] and Chen et al. [6] highlighted the role of semantic attributes in recognizing clothing types and patterns.

Recent improvements in object detection architectures, particularly Mask R-CNN, Faster R-CNN, and Detectron2, have significantly improved the precision of clothing localization and segmentation, especially when applied to datasets such as DeepFashion2. The segmentation techniques have enabled researchers to distinguish overlapping or layered garments—an important feature for real-world fashion applications.

2.2 Image Segmentation Techniques in Fashion Applications

Segmentation is critical to accurately isolating clothing items from background noise or other objects. Two main approaches—Semantic Segmentation and Instance Segmentation—offer different capabilities. Semantic segmentation labels each pixel according to category but cannot differentiate between object instances. In contrast, instance segmentation identifies each individual object separately, which is especially useful when garments overlap.

To address the limitations of each approach, this study adopts a Hybrid Segmentation technique, combining both methods to improve garment recognition and localization. The use of Detectron2, a state-of-the-art platform developed by Facebook AI Research, enables flexible implementation of such hybrid pipelines using COCO-format annotations.

2.3 Image Enhancement via CLAHE

Image preprocessing plays a vital role in improving model performance, especially in scenarios involving visual texture recognition. CLAHE (Contrast Limited Adaptive Histogram Equalization) has been widely adopted in medical imaging and facial recognition, and recent works (e.g., Chu-Hui et al. [4]) have demonstrated its effectiveness in fashion detection as well. CLAHE enhances the contrast of clothing textures and patterns, which are often subtle and easily lost in low-light or low-resolution conditions.

2.4 Fashion Style Classification and Ensemble Learning

Style classification models traditionally rely on supervised learning approaches such as **CNNs** for feature extraction and classification. However, ensemble methods, which combine the strengths of multiple classifiers, have shown to improve accuracy and robustness. Yulin Chen [10] demonstrated that combining models such as CNNs, Random Forests, and SVMs leads to better predictive performance in fashion trend classification tasks on social media.

In this study, an **ensemble model** integrates four classifiers: CNN, SVM, Plain Neural Network, and Random Forest, using **Weighted Soft Voting**. Each model contributes to the final decision based on its predictive confidence and validation accuracy, allowing for a balanced, reliable classification system.

2.5 Role of Celebrity Fashion in Recommendation Systems

The influence of celebrities on fashion decisions has been well-documented in both marketing and consumer behavior literature. Public figures often act as trendsetters, and their outfits frequently influence buying choices, especially among younger audiences. Integrating celebrity fashion databases into recommendation systems adds a socially relevant and aspirational dimension to personalized styling, as seen in prior works such as Veit et al. [7].

This study curates a celebrity dataset with strict criteria for influence and fashion relevance (e.g., presence in TIME100 or participation in global fashion events), creating a meaningful and targeted reference base for recommendation matching

3 Data and Methodology

3.1 Data

- Source: 16,000 fashion images (400 images each from 40 celebrities) curated from Pinterest and DeepFashion2.
- Celebrity Criteria: Selection based on influence (e.g., TIME100, BoF 500) and fashion relevance (e.g., Vogue, CFDA awards).
- Annotation Format: Converted DeepFashion2 annotations into COCO format for use with Detectron2.

3.2 Methodology

The development of a clothes styling assistant using computer vision and deep learning requires a comprehensive methodology that spans from dataset construction and preprocessing to model development, training, and evaluation. This section outlines the data sources, data preparation steps, model architecture, and experimental design used in this research as illustrated in Figure 1.

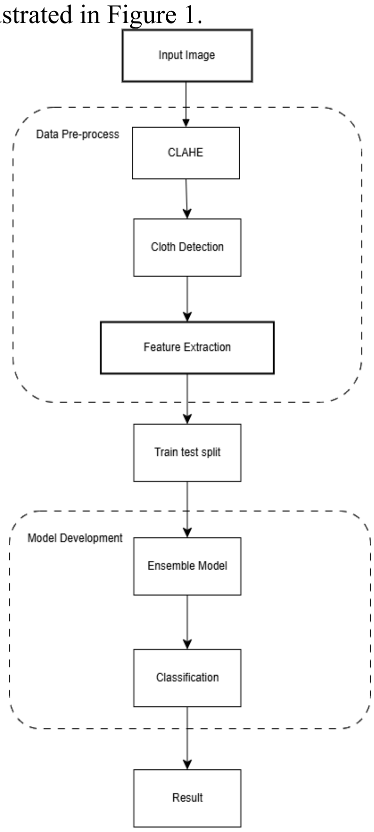

Fig. 1. Data architecture.

3.2.1. CLAHE Enhancement

To enhance image contrast and texture visibility—especially crucial for pattern recognition—CLAHE (Contrast Limited Adaptive Histogram Equalization) was applied. This technique improves fine-grained details such as fabric patterns and edge clarity. As figure 2 showing is the result of image that applied CLAHE

Fig. 2. Applied CLAHE

3.2.2. Image Normalization and Resizing

All images were resized to 224×224 pixels to conform with the input dimensions of CNN architectures. Pixel values were normalized to a [0, 1] range to stabilize training and accelerate convergence.

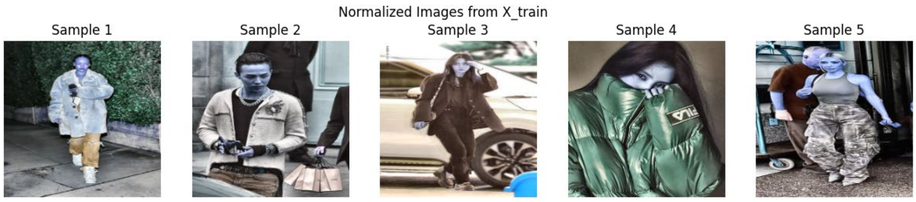

Fig. 3. Image Normalize and Resizing

3.2.3. Annotation Conversion

Annotations in the DeepFashion2 dataset were converted to COCO JSON format, with metadata fields including:

- image_id, bbox, segmentation, category_id
- Categories defined for 13 garment types (e.g., shirts, trousers, skirts)

Category	Number of Instances
short_sleeve_top	12556
long_sleeve_top	2011
Shorts	4167
short_sleeve_outerwear	3127
Sling_dress	1149
Long_sleeve_outerwear	5966
vest	2113
trouser	9586
long_sleeve_dress	1477
short_sleeve_dress	142
sling	322
skirt	6522
Vest_dress	3352
Total	52490

Table 1Deepfashion2 Dataset

This allowed seamless integration with Detectron2 for both instance and semantic segmentation tasks

3.3 Model Architecture and Development

The proposed system consists of two major pipelines: Clothing Detection (Segmentation) and Clothing Classification.

3.3.1. Segmentation Pipeline

To accurately identify garments in user images, a **Hybrid Segmentation** approach was implemented:

- **Semantic Segmentation** for pixel-level garment identification
- **Instance Segmentation** for distinguishing overlapping or multi-layered garments

Mask R-CNN, integrated via **Detectron2**, was trained on the processed DeepFashion2 dataset. This enabled the system to isolate clothing items from complex backgrounds and human features.

3.3.2. Feature Extraction

A **pre-trained ResNet50** was used as a feature extractor, outputting a **2048-dimensional feature vector** for each image. These vectors were then used as input to various classifiers.

3.4 Clothing Classification Models

Four different classifiers were implemented and compared:

- CNN (End-to-End) : Trained directly on CLAHE-enhanced images using a sequential Conv2D architecture with increasing filter depths (32 → 128), followed by dense classification layers.
- Plain Neural Network (NN) : Trained on ResNet50-generated features with dense layers and dropout, optimized using Adam and early stopping.
- Random Forest : Used RandomForestClassifier with 100 trees and default Gini impurity, trained on the ResNet50 feature set.
- Support Vector Machine (SVM) : Used SVC with RBF kernel and probability=True for integration into the ensemble system.

3.5 Ensemble Learning

To improve classification robustness and accuracy, a Weighted Soft Voting Ensemble was adopted. Each model contributed a probability vector over classes, weighted based on individual accuracy performance. The final prediction was the class with the highest weighted average probability.

Weighting order (by individual accuracy):

$$W_{1rf} = 0.4 \, , W_{cnn} = 0.3 \, , W_{nn} = 0.2 \, , W_{svm} = 0.1 \tag{1}$$

1. Random Forest
2. CNN
3. Plain Neural Network
4. SVM

This strategy leveraged the diversity in model architecture and learning behavior to enhance overall generalization.

3.6 Model Training and Evaluation

Training Procedure

- ResNet50 was frozen during feature extraction (no fine-tuning).
- CNN and NN models were trained using 80/10/10 data splits (train/val/test).
- Early stopping was used to prevent overfitting.

Model performance was evaluated using Accuracy, Precision, Recall and F-1 score. Cross-validation and side-by-side comparison between CLAHE-enhanced and raw images were performed to evaluate the preprocessing impact.

3.7 System Deployment

The final system was deployed as a **web-based application**, comprising:

- **Frontend:** HTML/CSS (user uploads image)
- **Backend:** Python Flask API (runs detection, classification)
- **Output:** Top-1 and Top-3 celebrity style matches with similarity percentages

The system provides real-time feedback and supports image uploads from local files.

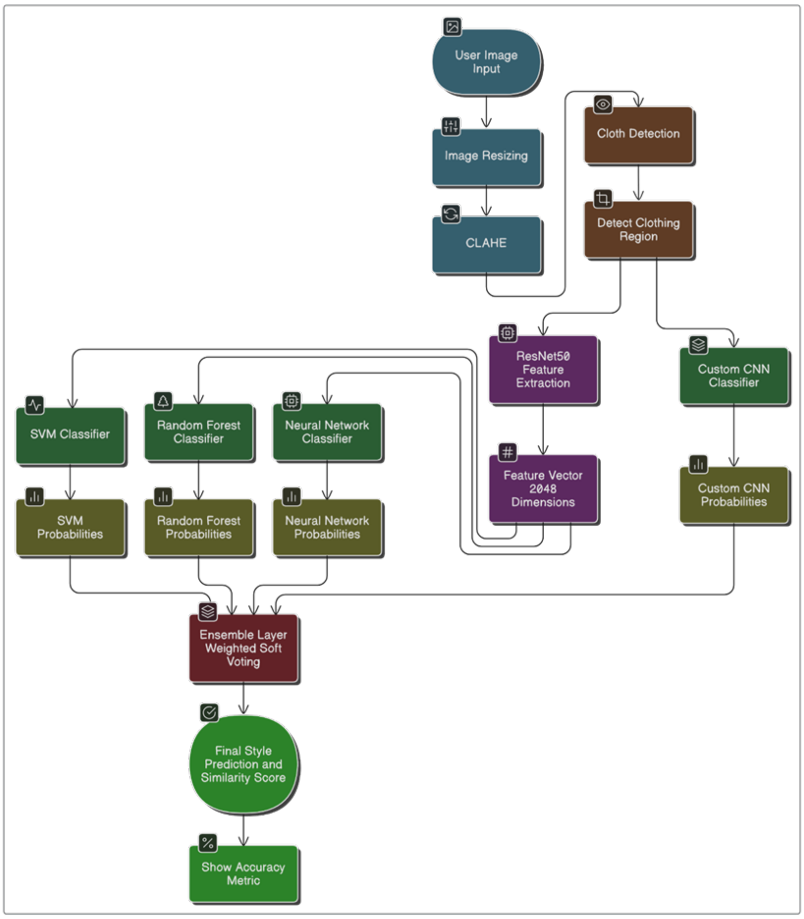

Fig 4 System Architecture

4 Result

This section presents the experimental setup, evaluation criteria, and comparative results from the segmentation and classification tasks. It also analyzes the contribution of various models and preprocessing techniques, particularly the effectiveness of CLAHE and ensemble learning in enhancing system performance.

4.1 Segmentation Results

To segment clothing items, a hybrid approach using **Semantic** Segmentation and Instance Segmentation was applied through Mask R-CNN in Detectron2.

- **Hybrid Segmentation** successfully combined pixel-level accuracy from semantic segmentation with object-level distinction from instance segmentation as Figure 5.

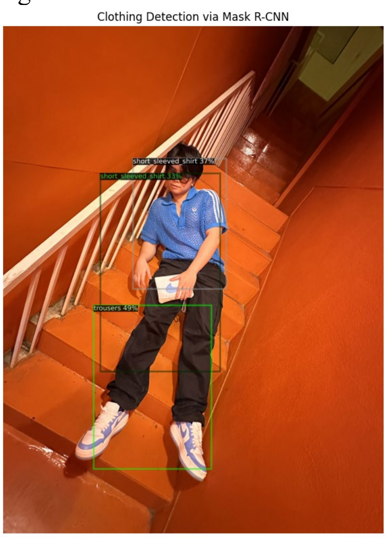

Fig 5 Segmentation Result

- **Qualitative Outputs:** Visual inspections of output masks showed improved garment boundary clarity with hybrid methods compared to using semantic or instance techniques in isolation.

4.2 Classification Results

Table 2 summarizes the performance of all models using CLAHE-enhanced images.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN	71.68	76.40	66.64	68.51
SVM	47.06	45.09	37.76	36.78
Random Forest	72.33	84.35	66.62	71.91
Plain NN	71.68	69.36	67.50	67.22
Ensemble	**73.86**	**79.15**	**68.54**	**70.86**

Table 2Model Comparison (With CLAHE)

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN	72.11	77.39	67.29	68.98
SVM	39.21	38.53	29.77	29.97
Random Forest	71.02	86.01	65.87	71.38
Plain NN	69.06	66.31	62.27	62.59
Ensemble	72.98	78.47	67.79	70.03

Table 3Model Comparison (Without CLAHE)

The **Ensemble Model** achieved the highest overall performance, confirming the benefit of combining multiple learning perspectives. We can see that CLAHE provided a **notable accuracy boost for SVM (+8%)** and a moderate improvement for Plain NN (+2.6%). Ensemble learning consistently outperformed individual models across both conditions, further validating the architecture.

4.3 System Demonstration

The system was tested on real-world user inputs. Users uploaded their own photos, and the system returned a similarity score ranking top 3 celebrity style matches as figure 5.

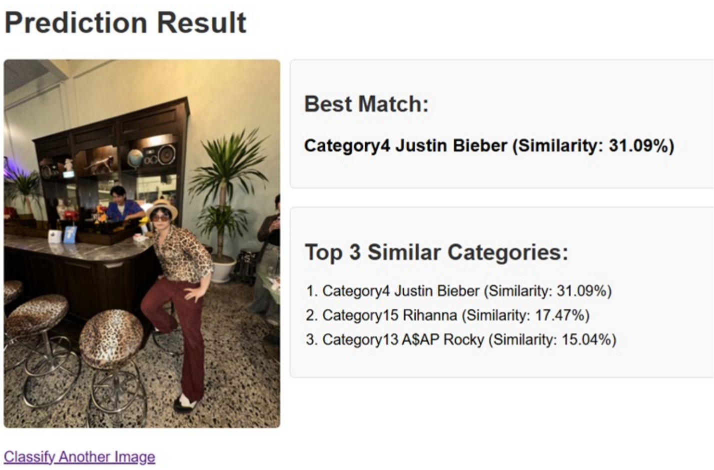

Fig 5 Application Output

5 Conclusion and Future Work

This study presents an integrated system for **celebrity-style-based virtual fashion recommendation** using deep learning and computer vision. The contributions include:

- **Hybrid Segmentation (Semantic + Instance):** Improved garment boundary detection in varied contexts.
- **CLAHE Image Enhancement:** Positively impacted models sensitive to contrast, especially SVM and NN.
- **Ensemble Classification (CNN, SVM, RF, NN):** Outperformed individual models by combining predictive strengths, achieving **73.86% accuracy**.
- **User-Facing Web Application:** Enabled real-time style similarity scoring using backend inference pipelines.

The system demonstrated both **technical feasibility** and **user-centric value** in facilitating fashion decisions based on well-curated celebrity references.

5.1 Limitations

The system developed in this study is a prototype designed for basic functionality using limited computational resources. As a result, the system's outputs may not yet be optimized for deployment in real-world applications. The fashion dataset employed, such as DeepFashion2, includes a limited number of clothing categories and may not comprehensively represent global fashion trends. Consequently, the system's performance may fall short when exposed to a wider diversity of clothing styles. Furthermore, the machine learning models used in this project function as black-box systems, meaning they are difficult to interpret or explain in detail. This limitation hampers the ability to provide transparent justifications for specific predictions or recommendations generated by the models.

5.2 Challenges and Obstacles

One of the major challenges encountered was the use of DeepFashion2, which is a publicly available but outdated dataset. The dataset has not received recent updates, and access to the latest version requires direct communication and approval from the dataset creators. Additionally, the large file size makes it difficult to train models locally or deploy them on cloud platforms without significant computational support. Another obstacle was the availability of celebrity images. For some individuals, especially from regions where public image sharing is culturally limited, it was difficult to gather a sufficient number of high-quality images. This constraint limited the diversity of representative figures used as fashion references in terms of race, gender, and cultural background. Lastly, the training process for deep learning models demanded extensive computational resources, including high memory capacity and GPU-based processing, which presented a challenge given the limitations of the available infrastructure.

5.3 Recommendations and Future Work

The inspiration for this study stemmed from the researcher's personal interest in fashion and the common challenge of purchasing clothing online. In many cases, size discrepancies and the inability to visualize how an outfit would appear in real life can result in dissatisfaction or a lack of confidence. However, seeing a well-known figure wearing a similar style often increases a wearer's confidence. This insight led to the concept of a celebrity-based recommendation system. In the future, the system could be extended into a fully functional virtual fitting room or avatar-based dressing application. As the fashion-tech space evolves, there is increasing use of video-based clothing analysis and real-time fashion recommendation platforms. The current project could serve as a foundational module for such advanced applications, provided that sufficient resources and expanded datasets are available to support deeper training and system scaling.

References

1. [1] A. Sidnev, A. Trushkov, M. Kazakov, I. Korolev and V. Sorokin, "DeepMark: One-Shot Clothing Detection," 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), 2019, pp. 3201-3204, doi: 10.1109/ICCVW.2019.00399.
2. [2] Fashion Statistics. https://fashionunited.com/ global-fashion-industry-statistics.
3. [3] L. Bossard, M. Dantone, C. Leistner, C. Wengert, T. Quack, and L. Van Gool. Apparel Classification with Style. In ACCV, 2012.
4. [4] Lee, C.-H.; Lin, C.-W. A Two-Phase Fashion Apparel Detection Method Based on YOLOv4. Appl. Sci. 2021, 11, 3782. https://doi.org/10.3390/app11093782
5. [5] Kiapour M.H., Yamaguchi K., Berg A.C., Berg T.L. (2014) Hipster Wars: Discovering Elements of Fashion Styles. In: Fleet D., Pajdla T., Schiele B., Tuytelaars T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8689. Springer, Cham. https://doi.org/10.1007/978-3-319-10590-1_31
6. [6] Chen H., Gallagher A., Girod B. (2012) Describing Clothing by Semantic Attributes. In: Fitzgibbon A., Lazebnik S., Perona P., Sato Y., Schmid C. (eds) Computer Vision – ECCV 2012. ECCV 2012. Lecture Notes in Computer Science, vol 7574. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-33712-3_44
7. [7] Veit, Andreas, et al. "Learning visual clothing style with heterogeneous dyadic co occurrences." Proceedings of the IEEE International Conference on Computer Vision. 2015.
8. [8] Z. Song, Meng Wang, Xian-sheng Hua and S. Yan, "Predicting occupation via human clothing and contexts," 2011 International Conference on Computer Vision, 2011, pp. 1084-1091, doi: 10.1109/ICCV.2011.6126355.
9. [9] Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. "Yolov4: Optimal speed and accuracy of object detection." arXiv preprint arXiv:2004.10934 (2020).
10. [10] Chen, Yulin. "A Social Media Mining and Ensemble Learning Model: Application to Luxury and Fast Fashion Brands." Information 12.4 (2021): 149.
11. [11] N. Bonettini, E. D. Cannas, S. Mandelli, L. Bondi, P. Bestagini and S. Tubaro, "Video Face Manipulation Detection Through Ensemble of CNNs," 2020 25th International Conference on Pattern Recognition (ICPR), 2021, pp. 5012-5019, doi: 10.1109/ICPR48806.2021.9412711
12. [12] Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." Advances in neural information processing systems 28 (2015).
13. [13] เรียนรู้การทำ Image Segmentation กับงานด้านแฟชั่นบน Kaggle
14. [14] Bolya, Daniel, et al. "Yolact: Real-time instance segmentation." Proceedings of the IEEE/CVF international conference on computer vision. 2019.
15. [15] Reza, Ali M. "Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement." Journal of VLSI signal processing systems for signal, image and video technology 38.1 (2004): 35-44.
16. [16] Crane, Diana. Fashion and its social agendas: Class, gender, and identity in clothing. University of Chicago Press, 2012.
17. [17] Liu, Yu, et al. "1st Place Solutions for OpenImage2019--Object Detection and Instance Segmentation." arXiv preprint arXiv:2003.07557 (2020).
18. [18] Bueno, Gloria, et al. "Segmentation Techniques." Modern Trends in Diatom Identification. Springer, Cham, 2020. 135-149.
19. [19] Z. Cai and N. Vasconcelos, "Cascade R-CNN: High Quality Object Detection and Instance Segmentation," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 5, pp. 1483-1498, 1 May 2021, doi: 10.1109/TPAMI.2019.2956516.
20. [20] Jiang, Xuefeng, et al. "Capsnet, cnn, fcn: Comparative performance evaluation for image classification." International Journal of Machine Learning and Computing 9.6 (2019): 840-848.

21. [21] Karademir-Hazir, Irmak. "Narratives of ‘getting it right’: Class, culture and good taste in clothing." International Journal of Fashion Studies 7.2 (2020): 147-165.
22. [22] Elihos, A., Alkan, B., Balci, B., & Artan, Y. (2018). Comparison of Image Classification and Object Detection for Passenger Seat Belt Violation Detection Using NIR & RGB Surveillance Camera Images. 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). doi:10.1109/avss.2018.8639447
23. [23] Mallu, Mallu. "Fashion Object Detection and Pixel-Wise Semantic Segmentation: Crowdsourcing framework for image bounding box detection & Pixel-Wise Segmentation." (2018).
24. [24] Qian, Xuelin, et al. "Long-term cloth-changing person re-identification." Proceedings of the Asian Conference on Computer Vision. 2020.
25. [25] Tian, Yujing, et al. "Research on the technology of searching for fashion trend image based on ResNet50 model." Journal of Physics: Conference Series. Vol. 1642. No. 1. IOP Publishing, 2020..