

## 生成 AI による視覚的確認機能を備えたファッション画像推薦システム

Development of Fashion Image Recommendation System  
with Visual Confirmation Using Generative AI

奥村 侑生\*

Yuki Okumura

黒木 修隆\*

Nobutaka Kuroki

沼 昌宏\*

Masahiro Numa

## 1. はじめに

2023 年度のアパレル EC 市場規模は前年比 4.76% 増の 2 兆 6712 億円となっており [1], ファッション EC 市場への関心が一層高まっている。一方, ファッション EC サイトでは非常に膨大な数の商品が取り扱われているため, 顧客が全ての商品を確認することは困難である。このような背景から, EC サイト上で顧客が自分にあった商品を効率的に選べるような機能が必要となっており, AI を用いたファッション推薦や仮想試着の研究・開発が行われている [2], [3].

## 2. 提案システム

## 2.1 提案システムの概要

提案システムの概要を図 1 に示す。本手法はユーザーの好みの衣服画像(上衣, 下衣)とユーザー自身の写真を入力とし, 出力として推薦商品およびそれらを着用した際のユーザーの姿が得られる。提案システムは以下の 4 つの機能を持つ。

- 衣服の切り抜き機能
- 特徴量抽出・定量化機能
- 類似検索・推薦機能
- 仮想試着機能

(a) のセグメンテーションは YOLOv9 [4] を使用し, (b) の特徴量抽出はマルチモーダルな AI である CLIP で学習された EVA-02 Transformer を使用している。(d) の仮想試着は CatVTON を使

用して行われる。これによりユーザーは自分に似合っているか視覚的に確認できる。

## 2.2 衣服特徴量の抽出・定量化方法

入力画像  $I_{query}$  から背景情報を除去するため, YOLOv9 を用いて自動で上衣または下衣を切り抜き,

$$S_{query} = \text{YoloSegmentation}(I_{query}) \quad (2.1)$$

を得る。この切り抜き画像  $S_{query}$  を学習済みモデルである clip-japanese-base の EVA-02 Transformer に入力し, 特徴ベクトルを,

$$A_{query} = \text{EVA02 Transformer}(S_{query}) \quad (2.2)$$

のように得る。

また, 商品画像  $I_i$  の特徴ベクトルについても同様に計算し,

$$A_i = \text{EVA02 Transformer}(I_i) \quad (2.3)$$

を得る。 $A_i$  はあらかじめ図 2 のようにデータベースに登録しておく。以上の方法で, 衣服の特徴量を定量化する。

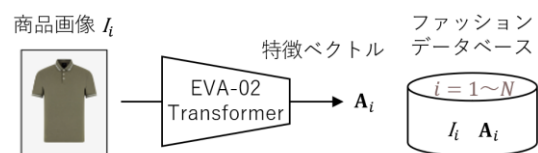


図 2 商品画像の特徴をデータベースに登録する流れ

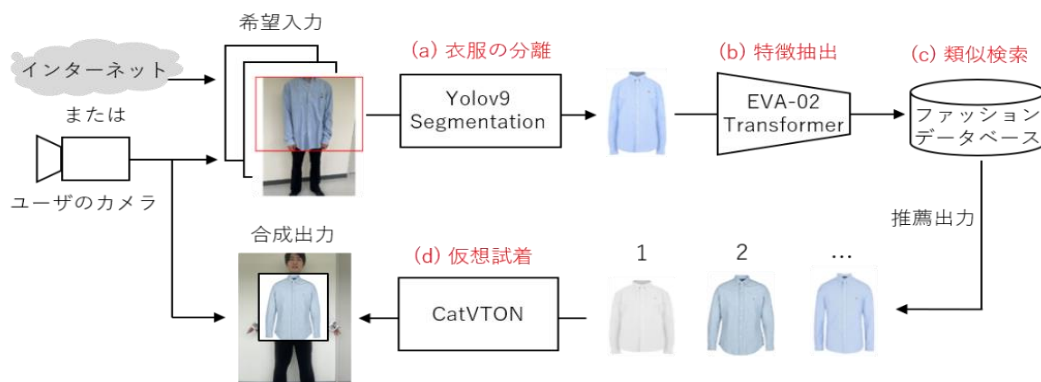


図 1 提案するファッション推薦の流れ

\*神戸大学大学院工学研究科,

Graduate School of Engineering, Kobe University

表 1 使用する AI モデルおよび学習条件

機能	使用する AI モデル	学習条件
(a) 衣服の切り抜き	Yolov9 Segmentation	学習画像 477 枚
(b) 特徴量抽出	EVA-02 Transformer	事前学習済み
(d) 仮想試着	Cat VTON	事前学習済み

## 2.3 ファッション推薦

提案システムはユーザの好みに近い商品を検索して推薦する。クエリ画像と商品画像との類似度は、両者の特徴ベクトル  $\mathbf{A}_{query}$  および  $\mathbf{A}_i$  のコサイン類似度によって算出する。すなわち、クエリ画像と  $i$  番目の商品との類似度  $c_i$  を

$$c_i = \frac{\mathbf{A}_{query} \cdot \mathbf{A}_i}{|\mathbf{A}_{query}| \cdot |\mathbf{A}_i|} \quad (2.4)$$

によって求める。これらの中から上位 3 つの商品について仮想試着画像を生成し、ユーザにファッション推薦として提示する。

## 3. 評価実験

### 3.1 実験内容

提案システムを用いたファッション推薦の評価を行う。使用した AI の学習条件を表 1 に示す。商品画像として DressCode データセット [3] の画像 53,792 枚を使用した。なお、ユーザは好みの衣服および自身の写真を撮影し、提案システムに入力した。

### 3.2 実験結果と考察

実験結果を図 3 に示す。まず推薦された 3 商品に関して、単なる画素レベルの類似性ではなく、「形状」「色合い」「模様」などの視覚的類似性が高い商品であることが分かる。例えば、図 3(a) のように胸元にロゴが入った青色の Y シャツを入力した場合、Best 1, 2, 3 のいずれもロゴが入った商品が推薦された。さらに図 3(b) では縦線が入った太めのジーンズを入力した場合、Best 1 では縦線が入ったジーンズが推薦され、Best 2, 3 においては太めのジーンズが推薦されている。このことから、CLIP の特徴空間における類似度の計算が有効に働いていると考える。

次に仮想試着に関して、ユーザのポーズに合わせて適切な変形が行われているため、実際に着用した際のイメージを把握しやすいことが分かる。この機能はユーザが直感的に商品を比較検討する上で非常に有用である。

一方で、提案システムは類似度の高い商品を連続して提示する傾向があり、推薦結果の多様性に課題が残ることも確認された。また、仮想試着において、図 3(a) では胸元のロゴの解像度が低く、商品の通りに転写されないことや、図 3(b) では商品の形状が着衣中の衣服の形状に引き寄せられる傾向が見受けられた。今後の生成 AI のさらなる精度向上が必要と考える。

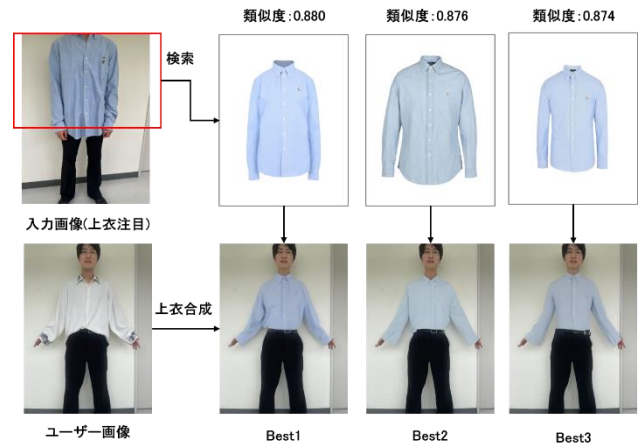
## 4. まとめ

本研究では、生成 AI を活用し、視覚的確認機能を備えたフ

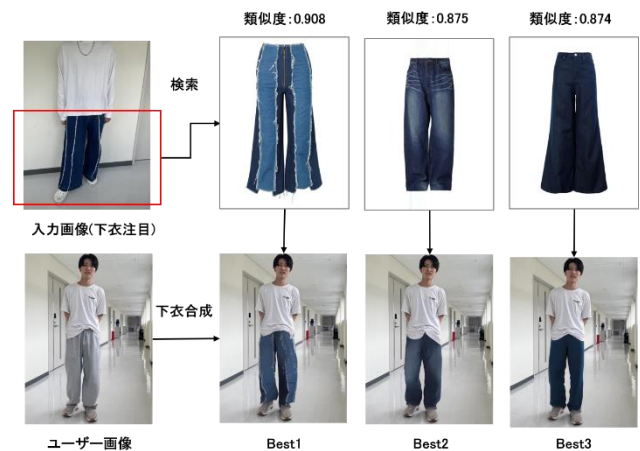
ァッション推薦システムを提案した。実験の結果、推薦商品には「形状」「色合い」「模様」の観点でユーザの好みが正しく反映されることを確認した。また、仮想試着では推薦された商品を着用した際のイメージをユーザが把握できることが確認された。今後は、ファッション推薦の有効性に関する客観的な評価方法の導入と、仮想試着の画像品質の向上に取り組んでいく予定である。

## 5. 参考文献

- [1] 経済産業省 商務情報政策局 情報経済課, “令和 5 年度電子商取引に関する市場調査報告書,” 2024 年 6 月.
- [2] Z. Chong, X. Dong, H. Li, S. Zhang, W. Zhang, X. Zhang, H. Zhao, D. Jiang, X. Liang, “CatVTON: Concatenation Is All You Need for Virtual Try-On with Diffusion Models,” arXiv:2407.15886, 2024.
- [3] D. Morelli, M. Fincato, M. Cornia, F. Landi, F. Gesari, R. Cucchiara, “Dress Code: High-Resolution Multi-Category Virtual Try-On,” arXiv:2204.08532, 2022.
- [4] C. Wang, I. Yeh, H. M. Liao, “YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information,” arXiv:2402.13616, 2024.



(a) 上衣を基に推薦を行った結果



(b) 下衣を基に推薦を行った結果

図 3 ファッション推薦結果