

LLM 기반 대화형 스마트 옷장 시스템

백지민^{1*}

LLM-powered Interactive Smart Closet System

Baek-Ji Min^{1*}

^{1*}Undergraduate Student, Department of Artificial Intelligence Software, Dongyang Mirae University, Seoul, 08221 Korea

요 약

최근 IoT와 인공지능 기술의 발전으로 스마트 옷장이 다양한 기능을 수행하는 지능형 시스템으로 발전하고 있으나 대부분은 정적인 의류 인식과 추천 기능에 그치고 있다. 본 논문에서는 사용자의 자연어 질의와 외부 환경 정보(날씨, 일정)를 분석하여 상황에 맞는 스타일을 대화형으로 추천하는 LLM 기반 스마트 옷장 시스템을 제안한다. 제안된 시스템은 YOLO와 ViT를 통한 의류 등록, API 기반 정보 수집, LLM 기반 룩 생성과 질의 처리 기능을 통합하여 설계되었으며 사용자 의도를 반영한 개인 맞춤형 스타일링을 실현한다. 본 연구는 스마트홈 환경에서의 사용자 경험 향상과 패션 추천 시스템의 확장 가능성을 보여준다.

ABSTRACT

With recent advancements in IoT and Artificial Intelligence technologies, smart closets are evolving into intelligent systems capable of performing diverse functions. However, most existing solutions remain limited to static clothing recognition and recommendation capabilities. This paper proposes a Large Language Model (LLM)-based smart closet system that analyzes users' natural language queries and external environmental data, such as weather and schedules, to provide conversational style recommendations tailored to specific contexts. The proposed system is designed by integrating YOLO and ViT for clothing registration, API-based information acquisition, and LLM-driven look generation and query processing. This architecture realizes personalized styling that accurately reflects user intent. This study demonstrates potential improvements in user experience within smart home environments and highlights the scalability of advanced fashion recommendation systems.

키워드 : 대규모 언어 모델, YOLO모델, ViT모델, 어플리케이션 프로그래밍 인터페이스, 스마트홈, 사용자 경험

Keywords : LLM, YOLO, ViT, API, Smart Home, User Experience

Received 29 January 2025,

Revised 29 March 2025,
(출판사에서작성)

Accepted 21 April 2025

* Corresponding Author Baek-Ji Min(E-mail:baekjimin0301@gmail.com, Tel:+82-10-9948-4710)

Undergraduate Student, Department of Artificial Intelligence Software, Dongyang Mirae University, Seoul, 08221 Korea

Open Access <http://doi.org/10.6109/jkiice.2025.29.1.399>

pISSN:2234-4772

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Copyright © The Korea Institute of Information and Communication Engineering.

I. 서 론

스마트홈 시장의 급속한 성장과 함께 사용자 중심의 스마트 가전 제품이 다양한 형태로 진화하고 있다. 특히 IoT(Internet of Things) 기술과 인공지능(AI)의 결합은 일상생활의 효율성과 편의성을 높이는 데 핵심적인 역할을 하고 있으며, 최근에는 GPT, LLaMA, Gemini 등과 같은 대규모 언어 모델(LLM, Large Language Model) 기술까지 상용화되며 인간의 언어를 이해하고 응답하는 자연어 인터페이스가 널리 적용되고 있다.

이러한 기술적 발전에도 불구하고 스마트 옷장 시스템은 여전히 제한적인 기능을 제공하고 있다. 대부분의 기존 스마트 옷장은 단순히 옷을 인식하거나 기온이나 계절과 같은 정적인 요소를 기반으로 옷차림을 추천하는 수준에 그친다. 사용자와의 상호작용은 대부분 메뉴 기반의 입력으로만 제한되며 사용자의 상황, 감정, 일정 등을 고려한 대화형 코드 추천 기능은 부재한 실정이다. 이러한 한계는 "오늘 중요한 회의가 있어. 너무 격식 있게는 말고 깔끔하게 입고 싶어"와 같은 문맥 기반의 질의를 이해하고 해석할 수 있는 기능의 필요성을 부각시킨다.

본 논문에서는 이와 같은 기존 기술의 한계를 극복하고자 옷 인식 시스템 + 외부 환경 정보(날씨, 일정) + 대화형 LLM 인터페이스를 통합한 개인 스타일리스트형 스마트 옷장 시스템을 제안한다. 본 시스템은 이미지 인식을 통해 옷을 등록하고 날씨 데이터 및 일정 데이터를 실시간 분석하여 현재 상황에 적합한 옷차림을 제안한다. 동시에 사용자의 질의에 대해 LLM 기반 언어 모델을 활용하여 자연스럽게 대화하며 스타일을 함께 고민하는 대화형 패션 비서로서 기능한다. 이러한 스마트 옷장은 기존 시스템과 비교하여 정적인 룰 기반 추천을 넘어서 사용자의 의도와 감정 그리고 상황까지 고려한 능동적이고 개인화된 의류 추천 시스템으로 발전할 수 있으며 향후 스마트홈 환경의 핵심 모듈로 확장될 가능성을 가진다.

본 논문의 구성은 다음과 같다. 2장에서는 기존 스마트 옷장의 의류 인식 및 외부 요인 기반 추천 기술의 현황을 검토하고 본 시스템의 기반 기술인 대규모 언어 모델(LLM)의 프롬프트 엔지니어링(Prompt Engineering) 기법과 객체 인식을 위한 YOLO 및 ViT 모델에 대해 살펴본다. 3장에서는 날씨와 패션 룩 데이터의 수집 및

전처리 과정을 설명하고 이를 바탕으로 사용자의 자연어 질의를 처리하여 맞춤형 코드를 제안하는 스마트 옷장 시스템의 아키텍처와 전체 동작 구조를 상세히 기술한다. 4장에서는 제안된 시스템의 실제 구현 내용을 다루며 시스템의 성능과 효율성을 검증하기 위한 평가 방법과 결과를 제시한다. 마지막으로 5장에서는 본 연구의 결론을 기술하고 향후 시스템의 확장과 고도화를 위한 발전 방향을 논의한다.

II. 관련 연구

본 장에서는 제안하는 스마트 옷장 시스템 구현을 위해 적용된 핵심 기술 요소들을 종합적으로 검토한다. 먼저 기존 스마트 옷장에서 사용되어 온 의류 인식 방식과 관련 선행 연구들을 정리하고 날씨 및 일정과 같은 외부 환경 데이터를 기반으로 의류를 추천하는 시스템들의 구조와 활용 사례를 살펴본다. 이어서 최근 다양한 분야에서 활발히 활용되고 있는 대규모 언어 모델(LLM)의 기술적 특성을 살펴보고 모델의 입력 텍스트의 구조 설계를 통해 목적에 맞는 결과를 유도하는 프롬프트 엔지니어링(Prompt Engineering) 기법의 원리와 적용 사례들을 소개한다. 마지막으로 본 시스템의 주요 구성 요소로 사용되는 YOLO(You Only Look Once)와 ViT(Vision Transformers)모델의 구조와 기능에 대해 구체적으로 살펴본다.

2.1 기존 스마트 옷장에서 사용된 의류 인식 방식

스마트 옷장 시스템에서 가장 핵심이 되는 기술 중 하나는 의류 인식 기술이다. 이 기술은 사용자가 보유한 옷을 디지털 정보로 변환하여 저장하고 이후 추천, 스타일링, 관리 등의 기능을 수행하는 기반이 된다. 다양한 방식의 의류 인식 기술이 제안되어 왔으며 각각 구현 편의성과 정확도, 자동화 가능성 등의 측면에서 다양한 특성을 가진다. 알아본 다른 스마트 옷장 시스템은 RFID(Radio Frequency Identification) 기술을 기반으로 구현되었다. 의류에 RFID 태그를 부착하고 옷장 내부 또는 외부에 설치된 리더기를 통해 비접촉 방식으로 의류 정보를 인식하는 구조이다. 이 방식은 빠른 속도와 높은 인식률을 장점으로 하며 대량의 의류를 자동으로 관리할 수 있는 시스템에 적합하다. 하지만

태그 비용, 금속이나 수분 환경에서의 인식 오류 등의 문제로 인해 일반 가정용 옷장에는 적용이 제한적이다 [1].

QR 코드 또는 NFC 태그를 활용한 방식은 RFID에 비해 비교적 저렴하고 구현이 간단하다는 점에서 많이 활용되고 있다. 특히 QR 코드는 시각적으로 스캔이 가능하고 스마트폰을 통한 접근성이 뛰어나다는 장점이 있다[2]. 하지만 이 방식은 의류 제조 단계에서 QR 코드가 사전 부착되어 있어야 의미가 있기 때문에 현실적으로는 모든 의류에 적용되기 어렵고 일부 제품에 한정되어 사용되거나 아예 적용이 불가능한 경우도 많을 것으로 예상된다.

최근에는 딥러닝 기반의 이미지 인식 기술을 활용한 스마트 옷장이 주목받고 있다. 이 방식은 카메라를 통해 촬영한 옷의 이미지를 분석하여 색상, 패턴, 형태, 카테고리 등의 정보를 자동으로 추출하고 분류한다[3]. YOLO, CNN 기반 모델 등의 객체 탐지 알고리즘이 주로 사용되며 의류에 태그를 부착하지 않아도 되는 점에서 사용자 접근성과 확장성이 뛰어나다. 해당 시스템은 카메라로 촬영한 이미지를 데이터베이스로 전송하고 학습된 딥러닝 모델을 통해 색상 및 종류 등을 분류한 후 사용자에게 정보를 전달하는 구조를 갖는다.

2.2 외부 요인을 기반으로 한 의류 추천 기술

최근 의류 추천 시스템은 단순히 스타일의 유사성이나 과거 선택 기록에 기반한 추천을 넘어 사용자 개인의 생활 맥락과 실시간 환경 정보를 반영하는 지능형 시스템으로 발전하고 있다. 특히 기상 정보, 시간대, 일정, 위치 등 다양한 외부 요인을 반영한 추천 기술은 개인화 및 상황 적합성을 강화하는 방향으로 주목받고 있다. 그중에서도 날씨 정보는 사용자들이 옷을 선택하는데 있어 가장 직접적인 외부 요인으로 작용한다[4]. 기온, 강수량, 습도, 풍속 등은 계절뿐 아니라 일별, 시간별로 의류 착용의 기준을 변화시키기 때문이다. 이러한 정보를 활용하는 대표적인 방식은 외부 기상 API를 통해 실시간 데이터를 수집하고 이에 따라 보온성, 통기성, 방수 여부 등의 속성을 가진 의류를 필터링하거나 우선 추천하는 구조이다[5].

또한 최근에는 날씨뿐 아니라 웹 크롤링 기반 트렌드 분석 기술을 활용하여 추천 정확도를 높이려는 시도도 활발히 이루어지고 있다. SNS, 패션 블로그, 뉴스 기사,

쇼핑몰 데이터 등에서 특정 기간 내 자주 언급된 아이템이나 코드 정보를 수집한 후 날씨 조건과 매칭된 스타일 패턴을 학습하는 방식이 대표적이다[6]. 뿐만 아니라 일정 기반 추천 기술도 함께 주목받고 있다. 사용자의 스마트폰 캘린더나 일정 앱과 연동하여 일정의 목적과 시간에 따라 적절한 복장을 제안하는 구조이다. 예를 들어 "오전 9시 사무실 회의"가 있는 날은 포멀한 셔츠와 슬랙스를 추천하고, "휴일 오후 가족 외출" 시에는 캐주얼한 코디를 우선적으로 제안하는 방식이다. 해당 기술은 주로 스마트홈, 스마트워치, 개인 비서 시스템 등과 결합되어 사용자의 맥락을 실시간으로 파악하고 이에 따른 적시성 높은 추천을 가능케 한다. 이러한 다차원 외부 요인을 통합적으로 고려하는 추천 시스템은 고도화된 사용자 경험을 가능하게 하며 단순한 추천을 넘어선 패션 가이드 역할을 수행할 수 있다. 본 연구에서도 이러한 접근을 반영하여 기상 API 데이터, 일정 정보를 사용하여 구현한다.

2.3 대규모 언어 모델과 prompt engineering 기법

대규모 언어 모델(LLM, Large Language Models)은 최근 다양한 응용 분야에서 두각을 나타내며 자연어 처리(NLP) 기술의 패러다임을 혁신적으로 변화시키고 있다. 특히 GPT, LLaMA, Gemini와 같은 LLM들은 수십억 개 이상의 파라미터를 기반으로 대규모 데이터를 학습하여 인간 수준에 가까운 언어 이해 및 생성 능력을 확보하고 있다.

최근에는 모델의 파라미터를 직접 업데이트하는 미세 조정(Fine-tuning) 대신 적절한 입력 텍스트(Prompt)를 설계하여 모델의 잠재된 능력을 이끌어내는 프롬프트 엔지니어링(Prompt Engineering) 기법이 주목받고 있다. 이는 모델의 가중치를 고정된 상태에서 예시를 제공하거나 지시사항을 구체화하여 원하는 출력을 유도하는 인컨텍스트 러닝(In-context Learning) 방식에 기반한다. 이 접근법은 모델을 별도로 재학습시킬 필요가 없어 구축 및 유지보수 비용이 낮고 프롬프트 수정만으로 다양한 도메인과 태스크에 유연하게 대응할 수 있다는 장점이 있다.

특히 추천 시스템 분야에서는 사용자의 선호도와 아이템 정보를 텍스트 형태의 프롬프트로 변환하여 LLM에 입력함으로써 문맥을 반영한 정교한 추천을 수행하는 연구가 활발하다. 사용자 프로필과 상호작용 기록을

프롬프트로 구성하여 대화형으로 추천을 제공하는 시스템을 제안한 연구도 있으며 이를 통해 별도의 학습 없이도 높은 수준의 개인화 추천과 설명 가능한 답변이 가능함을 보였다[7]. 또한 사고 사슬(Chain-of-Thought) 기법은 모델이 중간 추론 단계를 명시적으로 생성하도록 유도하여 복잡한 문제에 대해 보다 정확하고 일관된 해답을 도출할 수 있도록 한다[8].

이러한 기술적 흐름은 스마트 옷장 시스템에도 효과적으로 적용될 수 있다. 날씨, 일정, 사용자의 기분 등 비정형화된 문맥 정보를 처리해야 하는 스타일링 추천 작업에서 프롬프트 엔지니어링을 적용한 LLM은 사용자의 의도를 정확히 파악하고 논리적인 근거를 갖춘 코디를 제안할 수 있다. 본 연구에서는 이러한 프롬프트 기반 접근 방식을 통해 복잡한 사용자 질의에 유연하게 대응하고 개인화된 스타일링 경험을 제공하는 시스템을 구현한다.

2.4 You Only Look Once (YOLO)

객체 탐지는 입력된 이미지 내에서 객체의 위치(Bounding Box)를 찾고 해당 객체의 클래스를 분류하는 컴퓨터 비전의 핵심 기술이다. 초기 딥러닝 기반 모델인 R-CNN 계열은 후보 영역 추출과 분류를 분리한 2 단계(2-stage) 방식을 사용하여 속도 측면에서 한계가 있었다[9].

반면 YOLO는 객체 탐지를 하나의 회귀 문제로 재정의하여 처리 속도를 비약적으로 향상시켰다.

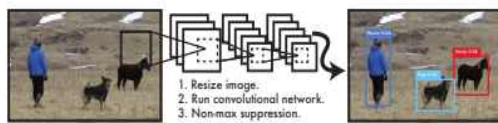


Fig. 1 YOLO Detection System

그림 1은 YOLO의 전체적인 탐지 시스템을 보여준다. YOLO는 복잡한 파이프라인 대신 입력 이미지를 고정된 크기로 리사이즈(Resize)하고 단일 컨볼루션 신경망(Convolutional Network)을 통과시킨 후 비최대 억제(Non-max suppression) 과정을 거쳐 최종 객체를 검출하는 단순하고 효율적인 구조를 가진다.

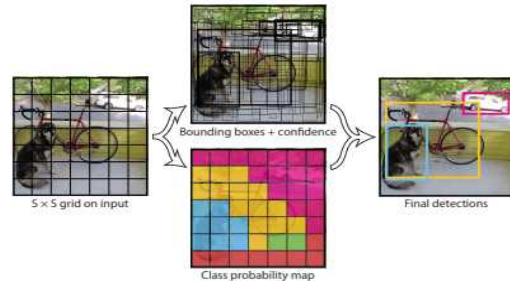


Fig. 2 Principles of the YOLO Algorithm

구체적인 작동 원리는 그림 2와 같다. 모델은 입력 이미지를 $S \times S$ 크기의 그리드로 분할하고 각 그리드 셀마다 B 개의 바운딩 박스와 해당 박스에 객체가 포함될 신뢰도 그리고 클래스 확률을 동시에 예측한다. 이 두 가지 정보는 최종적으로 결합되어 각 객체의 정확한 위치와 종류를 나타내는 바운딩 박스로 도출된다.

이러한 구조적 특징 덕분에 YOLO는 배경 오류가 적고 실시간 처리가 가능하다. 본 연구에서 제안하는 스마트 옷장 시스템은 사용자가 촬영한 의류를 지연 없이 즉각적으로 인식하고 등록해야 하므로 빠른 추론 속도와 준수한 정확도를 동시에 만족하는 YOLO 모델을 핵심 기술로 채택하였다.

2.5 Vision Transformer (ViT)

컴퓨터 비전 분야에서는 오랫동안 합성곱 신경망(CNN)이 이미지의 지역적 특징을 추출하는데 표준으로 자리 잡아왔다. 그러나 자연어 처리(NLP) 분야에서 Transformer가 문맥 정보 학습에 혁신적인 성과를 거두면서 이미지를 패치 기반 토큰 시퀀스로 변환하여 Transformer만으로 학습하는 Vision Transformer(ViT)가 제안되었다[10].

ViT는 CNN의 핵심인 컨볼루션 연산을 배제하고 이미지를 단어의 시퀀스처럼 처리하는 방식을 택했다. 그림 3은 본 연구에서 의류 속성 분석을 위해 참조한 ViT의 전체적인 아키텍처를 보여준다.

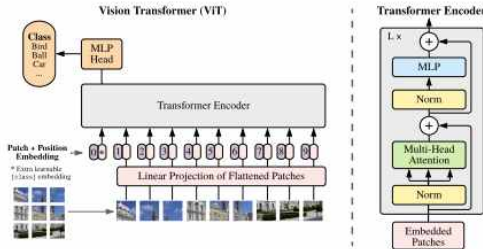


Fig. 3 Architecture of ViT Model

그림 3과 같이 ViT는 입력 이미지를 $P \times P$ 크기의 작은 패치들로 분할한다. 분할된 각 패치는 선형 투영(Linear Projection of Flattened Patches) 과정을 거쳐 1차원 벡터로 변환되며 이미지 내 위치 정보를 보존하기 위해 위치 임베딩(Position Embedding)이 더해진다.

이렇게 준비된 패치 시퀀스는 Transformer 인코더(Encoder)에 입력된다. 인코더 내부의 멀티 헤드 셀프 어텐션(Multi-Head Attention) 메커니즘은 이미지의 특정 부분뿐만 아니라 전체 영역 간의 상호작용을 학습한다. 출력단의 MLP Head는 학습된 특징을 바탕으로 최종적인 분류를 수행한다. 이와 같은 ViT 구조는 합성곱 연산에 의존하지 않고도 이미지 전반의 관계를 유연하게 모델링할 수 있다는 점에서 복잡한 시각적 패턴과 속성이 혼재된 의류 이미지 분석에 특히 적합한 접근 방식이라 할 수 있다.

III. 스마트 옷장 시스템

본 장에서는 스마트 옷장 시스템의 전체적인 동작 구조와 핵심 아키텍처, 데이터 수집 및 전처리를 상세히 기술한다. 본 시스템은 크게 사용자의 의류를 등록하고 속성을 분석하는 '지능형 인식 모듈', 상황과 맥락에 맞는 스타일을 제안하는 '대화형 추천 모듈', 그리고 추천된 룩을 가상으로 시착해 볼 수 있는 '가상 피팅 모듈'의 세 가지 핵심 요소로 구성된다.

3.1 데이터 수집 및 전처리

스마트 옷장 시스템은 사용자 맞춤형 추천 기능을 구현하기 위해 다양한 종류의 데이터를 통합적으로 수집하고 분석해야 한다. 본 절에서는 시스템에서 사용되는

데이터의 유형과 수집 방식, 전처리 절차에 대해 구체적으로 설명한다. 특히 기상 정보, 일정 정보, 패션 룩 데이터 등 다양한 도메인에서 수집된 데이터를 통합하여 상황 인식 기반의 의류 추천 시스템을 구성하는 부분에 초점을 둔다.

3.1.1 날씨 데이터

사용자에게 상황에 맞는 적절한 옷을 추천하기 위해서는 현재의 기온과 날씨 상태에 대한 정확한 정보가 필수적이다. 이에 본 시스템은 OpenWeatherMap API를 활용하여 실시간 기상 데이터를 수집하였다. 사용자의 위치 좌표(위도, 경도)를 입력하면 시스템은 그림 4와 같은 JSON 형식의 응답 데이터를 받아온다.

```
{
  "lat": 33.44,
  "lon": -94.04,
  "timezone": "America/Chicago",
  "timezone_offset": -18000,
  "current": {
    "dt": 1684929490,
    "sunrise": 1684926645,
    "sunset": 1684977332,
    "temp": 292.55,
    "feels_like": 292.87,
    "pressure": 1014,
    "humidity": 89,
    "dew_point": 290.69,
    "uvi": 0.16,
    "clouds": 53,
    "visibility": 10000,
    "wind_speed": 3.13,
    "wind_deg": 93,
    "wind_gust": 6.71,
    "weather": [
      {
        "id": 803,
        "main": "Clouds",
        "description": "broken clouds",
        "icon": "04d"
      }
    ]
  }
},
```

Fig. 4 Structure of Weather Data

수신된 데이터 중에서 본 연구는 의류 선택에 직접적인 영향을 주는 **current** 항목의 현재 기온(temp)과 체감 온도(feels_like) 값을 추출하여 사용하였다. 또한 **weather** 항목에서 날씨의 상태를 설명하는 정보(main, description)를 가져와 현재 비가 오는지 맑은지 등의 구체적인 기상 상황을 파악하였다. 이렇게 구성된 날씨 데이터는 시스템의 환경 변수로 저장되며 이후 스타일 추천 단계에서 사용자에게 가장 적합한 룩을 제안하는 근거 자료로 활용된다.

3.1.2 패션 룩 데이터

본 시스템의 핵심인 의류 인식 및 스타일 분석 모델을 학습시키기 위해 AI Hub에서 제공하는 'K-Fashion' 이미지 데이터셋을 활용하였다. 해당 데이터셋은 실제 패션 모델의 전신 이미지와 이에 대한 상세한 메타데이터가 그림 5와 같이 JSON 형식으로 매핑되어 있다.

```
{
  "파일번호": 1,
  "파일이름": "REGION_001_04.jpg", //파일명
  "텍스트좌표": {
    "아우터": [0], //좌측 X,Y좌표/ 우측 X,Y좌표
    "하의": [0],
    "원피스": [0],
    "상의": [0]
  },
  "폴리곤좌표": {
    "아우터": [0], //좌측 X,Y좌표/ 우측 X,Y좌표
    "하의": [0],
    "원피스": [0],
    "상의": [0]
  },
  "라벨링": {
    "스타일": [
      {
        "스타일": "스트리트",
        "아우터": [
          {
            "가장": "플",
            "카테고리": "점퍼", //분류항목
            "디테일": [
              "스트링", "지퍼",
              "프린트": [
                "무지",
                "꽃": "오버사이즈"
              ],
            "하의": [
              {
                "가장": "발목",
                "카테고리": "청바지",
                "디테일": [
                  "플임",
                  "소재": "데님",
                  "핏": "노멀"
                ],
                "원피스": [0],
                "상의": [
                  {
                    "카테고리": "티셔츠",
                    "소재": "저지",
                    "프린트": [
                      "무지",
                      "네라인": "라운드넥",
                      "꽃": "루즈"
                    ]
                  }
                ]
              }
            ]
          }
        ]
      }
    ]
  }
}
```

Fig. 5 Structure of K-Fashion Data

우리는 컴퓨팅 자원의 부족으로 인해 전체 데이터셋 중 '클래식, 로맨틱, 모던, 힙합, 스트리트'의 5가지 핵심 스타일 카테고리를 선정하였으며 각 스타일별로 약 10,000장씩 선별하여 학습 데이터셋을 구축하였다. 이 원천 데이터는 시스템의 두 가지 핵심 모델인 YOLO와 ViT의 학습 목적에 맞춰 다음과 같이 두 가지 형태로 전처리되었다.

첫째, YOLO 기반의 객체 탐지 및 세그멘테이션 학습을 위해 위치 정보를 추출하였다. JSON 데이터 내의 '폴리곤 좌표(Polygon Coordinates)'를 파싱하여 의류의 정확한 영역 정보를 확보하고 이를 원본 이미지와 매칭하여 모델이 이미지 내에서 의류의 위치와 형상을 정확히 인식하도록 데이터셋을 구성하였다.

둘째, ViT 기반의 속성 분석 학습을 위해 개별 의류 이미지를 생성하였다. JSON 데이터의 '텍스트 좌표(Rect Coordinates, Bounding Box)'를 기준으로 원본 룩 이미지에서 상의, 하의, 아우터, 원피스 등 각 아이템 영역을

잘라내어 그림 6에서의 Modified Data와 같은 개별 이미지를 생성하였다.



Fig. 6 Cropping Process for Individual Clothing Item Identification

또한 통합되어 있던 JSON 메타데이터를 분해하여 잘라낸 각 이미지(상의, 하의 등)에 해당하는 속성 정보(디테일, 소재, 핏 등)만을 담은 개별 JSON 파일을 생성하였다. 이렇게 정제된 데이터셋은 향후 3.2절에서 기술할 ViT 모델이 의류의 미세한 특징을 학습하는 데 사용된다.

3.2 시스템 아키텍처 및 전체 동작 구조

스마트 옷장 시스템은 단순히 보유한 옷을 분류하거나 리스트업하는 정적인 관리 도구를 넘어 사용자의 의도, 날씨, 일정, 문맥 등 다양한 외부 요인을 통합적으로 고려하여 동적인 스타일 추천을 제공하는 지능형 대화 시스템을 지향한다. 이러한 목표를 달성하기 위해 본 연구에서는 YOLO와 ViT를 활용한 비전 인식 기술, 프롬프트 엔지니어링 기반의 LLM 추론 기술, 그리고 RTM Pose와 MediaPipe를 결합한 포즈 추정 및 생성 기술을 유기적으로 통합한 멀티모달(Multi-modal) 아키텍처를 설계하였다. 전체 시스템의 동작 프로세스는 크게 '지능형 의류 인식', '맥락 기반 스타일 추천', '관절 기반 가상 피팅'의 3단계 순환 구조로 이루어진다. 이어지는 절에서는 이러한 전체 흐름을 구성하는 각 핵심 모듈의 세부적인 기술 구현과 학습 방법에 대해 구체적으로 기술한다.

3.2.1 YOLO와 ViT기반 의류 인식

본 시스템의 의류 인식 모듈은 비정형의 사용자 이미지로부터 의류 영역을 정밀하게 추출하고 카테고리 및 세부 속성을 심층 분석하여 정형화된 데이터로 변환하

는 핵심 기능을 수행한다. 이를 구현하기 위해 본 연구에서는 객체 탐지 중에서도 세그멘테이션(Segmentation) 모델 선정 단계에서 다양한 버전의 모델 성능을 비교 분석하였다. 그림 7은 YOLO 시리즈별 객체 탐지 성능을 비교한 선행 연구 결과를 보여준다[11].

| Model | F1 Score | PR Curve (Car) | PR Curve (Dog) | PR Curve (Person) |
|----------|----------|----------------|----------------|-------------------|
| YOLOv5s | 73% | 86 | 58 | 77 |
| YOLOv8s | 86% | 92 | 75 | 87 |
| YOLOv9s | 76% | 85 | 72 | 77 |
| YOLOv10s | 70% | 83 | 50 | 76 |
| YOLOv11s | 79% | 91 | 64 | 85 |
| YOLOv12s | 60% | 80 | 44 | 74 |

Fig. 7 Performance Comparison of YOLO Model Variants

비교 결과 YOLOv8s 모델이 F1 Score 86%를 기록하며 YOLOv5s(73%), YOLOv9s(76%) 등 타 모델 대비 가장 우수한 성능을 보였다. 특히 의류와 같이 형태가 다양하고 복잡한 객체를 정확히 인식하기 위해서는 정밀도(Precision)와 재현율(Recall)의 조화 평균인 F1 Score가 중요하므로 본 연구에서는 YOLOv8 기반의 세그멘테이션 모델을 최종 채택하였다. 이를 통해 구축된 전체 인식 파이프라인은 그림 8과 같이 YOLO를 통한 영역 검출과 ViT를 통한 계층적 속성 분석으로 이어진다.



Fig. 8 Architecture of the Clothing Recognition Module

먼저 1차적으로 YOLOv8 모델이 입력 이미지 내 의류 영역(Bounding Box)을 검출하고 배경을 분리하는 세그멘테이션을 수행한다. 모델 학습을 위해 앞서 3.1.2

절에서 설명한 패션 룩 데이터셋을 8:2 비율로 학습(Training) 및 검증(Validation) 세트로 분할하였으며 NVIDIA RTX 3060 GPU 환경에서 총 100 Epoch 동안 학습을 진행하였다. 학습 결과는 그림 9와 같다.

| | | |
|-------------------------|-------------------------|-----------------------|
| Epoch 1 | Epoch 100 | •Box(P): 0.958 |
| train/box_loss: 0.90595 | train/box_loss: 0.48438 | •Box(R): 0.939 |
| train/seg_loss: 3.33039 | train/seg_loss: 2.54747 | •Box(mAP50): 0.978 |
| train/cls_loss: 1.78572 | train/cls_loss: 0.28457 | •Box(mAP50-95): 0.888 |

Fig. 9 Training Results and Performance Metrics of the YOLO Models

학습 초기(Epoch 1) 대비 종료 시점(Epoch 100)에서 Box Loss는 0.906에서 0.484로 Seg Loss는 3.330에서 2.547로 현저히 감소하며 모델이 안정적으로 수렴함을 확인하였다. 최종 성능 평가에서 객체 위치 탐지 정확도를 나타내는 mAP50은 0.978 보다 엄격한 기준인 mAP50-95에서도 0.888이라는 매우 높은 수치를 기록하였다. 이는 모델이 복잡한 배경 속에서도 97.8%의 정확도로 의류 영역을 식별해냄을 의미한다.

YOLO를 통해 배경이 제거(Background Removal)된 의류 이미지는 다음 단계인 ViT 분석 모듈로 전달된다. ViT 분석은 정확도를 극대화하기 위해 '카테고리 분류'와 '속성 분석'의 2단계 계층적 구조로 설계되었다. 1단계에서는 입력 이미지를 상의, 하의, 아우터, 원피스의 4대 대분류로 식별하며 약 40 Epoch의 학습을 통해 97%의 분류 정확도를 달성하였다.

이어지는 2단계에서는 1차 분류 결과에 따라 각 카테고리에 특화된 전용 ViT 모델로 분기되어 입력된다. 각 모델은 해당 카테고리 고유의 세부 속성(패턴, 소재, 핏 등)을 추론하며 다수의 모델을 학습시키기 때문에 시간 관계상 대부분 20 Epoch를 기준으로 학습을 진행하였다. 그림 10은 각 카테고리별 모델의 학습 진행에 따른 Loss 감소 추이를 보여주며 모든 모델이 안정적으로 학습되었음을 나타낸다.

결론적으로 성능 비교를 통해 검증된 YOLOv8 모델과 계층적 ViT 구조의 결합은 이미지 내 의류의 위치 정보와 의미론적 속성 정보를 효과적으로 추출하여 이어지는 맥락 기반 스타일 추천 시스템에 고품질의 메타데이터를 제공하는 기반이 된다.

| |
|-----------------------|
| category: '상의' |
| epoch: 40 |
| initial loss: 11.1909 |
| final loss: 0.1017 |
| category: '하의' |
| epoch: 20 |
| initial loss: 6.8652 |
| final loss: 0.1128 |
| category: '아우터' |
| epoch: 20 |
| initial loss: 12.9713 |
| final loss: 0.1238 |
| category: '원피스' |
| epoch: 20 |
| initial loss: 14.9024 |
| final loss: 0.0768 |

Fig. 10 Training Results of ViT Models by Clothing Category

3.2.2 맥락 기반 스타일 추천

본 절에서는 사용자의 음성 질의를 이해하고 실시간 날씨 정보와 보유한 의류 데이터를 종합하여 최적의 스타일을 제안하기 위한 LLM 기반의 추천 프로세스를 설명한다.

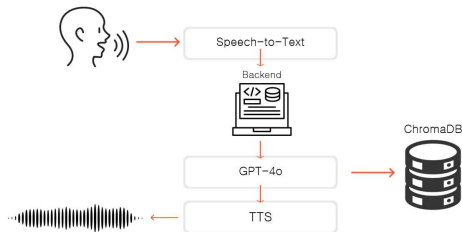


Fig. 11 Architecture of the Context-Based Style Recommendation Module

전체적인 시스템의 동작 흐름은 그림 11과 같이 음성 인식에서 시작하여 추론을 거쳐 다시 음성 및 시각 정보로 피드백되는 구조를 가진다. 사용자가 자연어로 코디를 요청하면 시스템은 우선 STT(Speech-to-Text) 모듈을 통해 음성 신호를 텍스트 데이터로 변환한다. 변환된 텍스트 질의는 단독으로 LLM에 입력되는 것이 아니라 추천의 정확도와 개인화를 보장하기 위해 랭체인

(LangChain) 프레임워크를 통해 다양한 문맥(Context) 정보와 결합되어 하나의 완성된 프롬프트 체인(Prompt Chain)을 구성한다.

```
result = chain.invoke({
    "input": user_text,
    "context": context,
    "format_instructions": self.parser.get_format_instructions(),
    "gender": user['gender'],
    "cloth_attr": cloth_attr
})
```

Fig. 12 Configuration of Key Input Variables for the LLM Chain

그림 12는 실제 LLM 호출 시 주입되는 핵심 변수들의 구성을 보여준다. 여기에는 사용자의 발화 내용인 input뿐만 아니라 3.1.1절에서 수집한 기온 및 날씨 상태를 포함하는 context, 3.2.1절에서 분석된 사용자의 보유 의류 리스트와 속성 정보를 담은 cloth_attr, 그리고 사용자 성별 정보인 gender가 포함된다. 또한, LLM의 불확실한 출력을 시스템이 처리 가능한 형태로 제어하기 위해 format_instructions 변수를 통해 JSON 출력 포맷을 강제한다.

```
self.final_prompt = ChatPromptTemplate.from_messages([
    ("system", """나는 한국어 패션 비서야.

    다음 지침을 따라줘:
    1) 사용자의 질문을 이해하고, 현재 날씨와 상황을 고려해.
    2) 핵심 키워드는 5-6개(장소, 상황)와 스타일은 (cloth_attr) 에 있는 정보로 3-5개 키워드를 뽑아.
    3) 키워드와 스타일은 중복 없이 다양하게 선택해.
    4) 짧고 자연스러운 존댓말 추천 문구를 작성해.
    5) 사용자의 성별은 (gender) 이
    6) 현재 날씨를 고려하여 옷 조합을 추천해줘

    {context}

    {format_instructions}

    출력은 반드시 json 형식으로만 해줘."""),
    ("human", "{input}")
])
```

Fig. 13 Structure of the System Prompt for the AI Fashion Assistant

이러한 변수들은 사전에 정교하게 설계된 시스템 프롬프트(System Prompt) 템플릿에 삽입되어 모델의 페르소나(Persona)와 동작 방식을 정의한다. 그림 13에서 볼 수 있듯이 시스템은 LLM에게 "한국어 패션 비서"라는 역할을 부여하고 구체적인 행동 지침을 내린다. 지침에는 사용자의 질문과 날씨 상황을 고려하여 cloth_attr에 존재하는 의류 중에서 중복 없이 키워드와 스타일을 선별할 것 그리고 이를 바탕으로 자연스러운 추천

문구를 작성할 것 등의 요구사항이 포함된다. 이는 범용 언어 모델이 불필요한 대화를 생성하는 것을 방지하고 패션 전문가로서의 목적 지향적인 답변을 생성하도록 유도한다. 또한 단순히 지시사항만으로 복잡한 스타일링을 수행하는 제로샷(Zero-shot) 방식의 한계를 극복하기 위해 해당 시스템에서는 퓨샷 러닝(Few-Shot Learning) 기법을 적용하여 추론 성능을 고도화하였다.

```
# Few-shot 예제
self.examples = [
    {
        "input": "비 오는 날 데이트룩 추천.",
        "output": "'{
            '키워드': ['비', '데이트', '트렌치 코트', '블라우스', '미디 스커트', '멜로루즈'],
            '스타일': ['로맨틱', '모던', '시크'],
            '추천문구': '비 오는 날에는 트렌치 코트와 미디 스커트로 세련되게 연출해 보세요.'
        }'"
    },
    {
        "input": "겨울 회사 회의인데 깔끔하게",
        "output": "'{
            '키워드': ['겨울', '회사 회의', '니트', '슬랙스', '체스터 코트', '데바슈즈'],
            '스타일': ['클래식', '미니멀', '모던'],
            '추천문구': '회식에는 니트와 슬랙스에 체스터 코트를 매치하시면 깔끔해 보이실 거예요.'
        }'"
    },
    {
        "input": "합피츠 개강파티 옷 추천",
        "output": "'{
            '키워드': ['합피츠', '개강파티', '세츠', '데님', '스니커즈', '크로스백'],
            '스타일': ['캐주얼', '스마트캐주얼', '내추럴'],
            '추천문구': '개강파티에는 세츠와 데님에 스니커즈로 편안하면서도 단정하게 연출해 보세요.'
        }'"
    }
]
```

Fig. 14 Examples of Few-shot Learning for Intent Classification

그림 14와 같이 "비 오는 날 데이트룩 추천"이나 "겨울 회사 회식" 등 구체적인 상황 예시와 이에 대해 작성한 모범 답안을 프롬프트에 예제로 함께 제공하였다. 이러한 퓨샷 예제들은 LLM이 추상적인 사용자 요청을 구체적인 의류 조합으로 변환하는 논리적 사고의 패턴(Chain of Thought)을 학습하게 한다. 최종적으로 생성된 추천 멘트는 TTS(Text-to-Speech) 모듈을 통해 음성으로 변환되어 사용자에게 전달된다. 이를 통해 사용자는 마치 실제 스타일리스트와 대화하는 듯한 직관적이고 상호작용적인 경험을 얻게 된다.

3.2.3 관절 기반 가상 피팅

앞선 3.2.2절에서 LLM이 제안한 스타일링 조합은 텍스트 및 음성 정보의 형태로 사용자에게 전달되지만 실제 착용 시의 핏이나 실루엣을 확인하기에는 한계가 있다. 그래서 본 절에서는 그런 한계점을 해결하고자 구현했던 추천된 의류를 사용자의 신체 위에 가상으로 합성하여 시각적 결과를 제공하는 가상 피팅(Virtual Tr

y-On) 모듈에 대해서 설명하고자 한다.

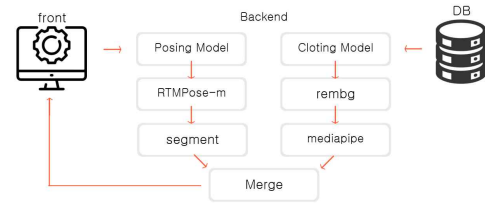


Fig. 15 Architecture of the Pose-based Virtual Fitting Module

이 모듈은 실시간성을 보장하면서도 자연스러운 합성을 위해 그림 15와 같이 사용자 포즈 추정과 의류 전처리가 병렬적으로 수행되는 이원화된 아키텍처를 채택하였다.

가상 피팅 프로세스는 프론트엔드(Frontend)로부터 사용자의 실시간 영상이 입력되면서 시작된다. 입력된 영상은 'Posing Model'을 통해 사람의 영역이 1차적으로 검출되며 이어지는 RTMPose-m 모델이 핵심적인 역할을 수행한다. RTMPose-m은 경량화된 고성능 포즈 추정 모델로 영상 내 사용자의 어깨, 팔꿈치, 손목, 골반 등 주요 관절 포인트를 실시간으로 예측하여 의류가 입혀질 정확한 위치 좌표를 산출한다. 동시에 'Segment' 모듈은 의류가 덮여서는 안 되는 영역(얼굴, 목, 손 등)에 대한 마스크(Mask)를 생성하여 합성 시 의류가 피부를 부자연스럽게 침범하는 폐색(Occlusion) 현상을 방지한다.

사용자 정보가 처리되는 동안 반대편 파이프라인에서는 추천된 의류에 대한 전처리가 병렬적으로 진행된다. 데이터베이스(DB)에서 호출된 의류 이미지는 'Cloting Model'을 거쳐 기본적인 특징이 식별된 후 rembg 라이브러리를 통해 배경이 제거된 순수한 의류 객체로 변환된다. 이후 MediaPipe 프레임워크가 적용되어 의류 마스크 위에서 가상의 관절 포인트를 생성한다. 이는 평면적인 의류 이미지를 사용자의 입체적인 관절 움직임에 맞춰 변형하기 위한 기준점으로 활용된다. 최종적으로 'Merge' 단계에서는 RTMPose-m이 추출한 사용자의 신체 관절 좌표와 MediaPipe가 생성한 의류의 가상 좌표를 매핑(Mapping)하여 합성을 수행한다. 이 과정에서 의류 이미지는 사용자의 자세에 따라 자연스럽게 늘어나거나 회전하는 변환을 거치게 되며 앞서 생성

된 피부 영역 마스크와 결합되어 깊이감이 반영된 최종 결과물을 생성한다. 이렇게 완성된 가상 피팅 영상은 다시 프론트엔드로 전송되어 사용자가 마치 거울을 보듯 자신의 움직임에 맞춰 옷을 입어보는 실감 나는 사용자 경험을 제공한다.

IV. 구현 및 결과

본 장에서는 제안한 스마트 옷장 시스템의 실제 구현 결과와 시스템 전반에 대한 평가 방법을 기술한다. 3장에서 검증한 개별 AI 모델들이 유기적으로 결합되어 사용자에게 어떤 인터페이스로 제공되는지 설명하고 완성된 시스템이 실용적인 관점에서 의도한 대로 동작하는지 확인하기 위한 통합 평가 기준을 제시한다.

4.1 시스템 구현

본 연구에서 제안하는 스마트 옷장 시스템은 사용자의 생활 공간에 자연스럽게 녹아드는 인텔리전트 가구를 지향한다.



Fig. 16 Conceptual prototype of the wardrobe system

시스템의 물리적 구현 형태는 그림 16과 같이 기존 옷장의 도어 전면에 대형 디지털 스크린이 부착된 형태를 상정하였다. 이러한 구성은 사용자가 외출 전 거울을 보는 행위와 동시에 스타일 추천 및 가상 피팅 서비스를 직관적으로 이용할 수 있는 환경을 제공한다.

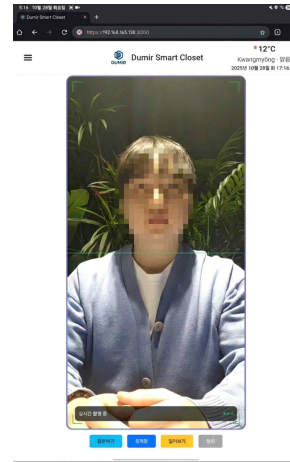


Fig. 17 User Interface of the Implemented Smart Closet System

실제 구현된 시스템의 사용자 인터페이스(UI)는 그림 17과 같다. 화면 중앙에는 실시간 카메라 영상이 배치되어 사용자의 모습과 현재 착장 상태를 보여주며 상단에는 현재 날씨와 온도가 표시되어 오늘의 날씨정보를 알 수 있다. 화면 하단에는 시스템의 핵심 기능을 수행하기 위한 4개의 주요 버튼인 '질문하기', '옷저장', '입어보기', '정지'를 배치하였다. 각 버튼의 기능적 역할은 다음과 같다. 먼저 '질문하기' 버튼은 3.2.2절에서 기술한 LLM 기반의 대화형 추천 모듈을 호출하여 사용자의 음성 질의를 처리한다. '옷저장' 버튼은 YOLO와 ViT 모델을 구동하여 사용자의 의류를 분석하고 개인 데이터베이스에 등록하는 기능을 수행하며 '입어보기' 버튼은 3.2.3절의 RTM Pose 기반 가상 피팅 모듈을 실행하여 추천된 의류를 영상 위에 합성한다. 또한 인터페이스 좌측 상단의 메뉴 버튼을 통해 사용자는 자신이 저장한 의류 리스트를 확인하거나 상세 설정을 관리할 수 있다.

4.2 평가 방법

본 연구에서는 제안한 스마트 옷장 시스템의 객관적인 성능 검증과 실무적 유용성을 평가하기 위해 정량적 평가와 정성적 평가를 병행하는 다각적 평가 체계를 설계하였다. 전체 평가 비중은 시스템의 기술적 완성도를 확인하는 정량적 지표에 40%, 사용자 경험과 실용성을 확인하는 정성적 지표에 60%를 할당하였다. 이는 본 시

시스템이 단순한 기술 구현을 넘어 사용자 중심의 지능형 서비스를 지향하기 때문이다. 정량적 평가는 시스템을 구성하는 핵심 인공지능 모델들의 동작 정확도를 수치화하여 진행한다. 이는 시스템이 제공하는 서비스의 신뢰도를 결정짓는 기초 지표가 된다. 주요 평가 요소로는 3.2.1절에서 상세히 기술한 YOLOv8 모델의 의류 탐지 및 세그멘테이션 정확도(mAP), ViT 모델의 계층적 카테고리 및 속성 분류 정확도가 포함된다. 또한, 3.2.2절에서 다룬 LLM 기반 추천 시스템의 질의 처리 응답 성공률을 통해 기술적 성능을 종합적으로 판단한다. 개별 모델의 세부적인 성능 수치는 앞선 3장에서 충분히 검증되었으므로 본 절에서는 이를 시스템 전체 성능의 정량적 근거로 활용한다.

정성적 평가는 시스템 전체 경험에 대한 사용자 반응을 중심으로 수행된다. 추천 결과에 대한 만족도, 질의 처리의 맥락 적합성, 인터페이스의 사용 편의성 등 시스템의 종합적인 실용성을 확인하는 데 목적이 있다. 평가의 객관성을 확보하기 위해 평가 대상자는 사용자 역할에 따라 세 가지 그룹으로 분류하여 설계하였다.

Table. 1 Detailed Description of User Roles

| Group | User Role | Example User Role | Composition Ratio |
|-------|-------------------------|--|-------------------|
| A | Fashion Influencer | Social media style creator, YouTuber | 35% |
| B | General Consumer | Regular user with an interest in fashion | 50% |
| C | Fashion Designer/Expert | Stylist, brand-affiliated professional | 15% |

표 1은 실제 사용 환경과 타겟 사용자를 반영하여 구성한 평가단 그룹의 상세 명세이다. 평가단은 실제 대중의 특성을 반영한 일반 사용자(Group B)를 중심으로 하되 트렌드 주도층인 패션 인플루언서(Group A)와 전문가인 디자이너(Group C)의 의견을 균형 있게 수렴하도록 설계하였다. 또한 각 그룹 내에서는 성별 및 연령대의 다양성을 고려하여 샘플링을 진행하며 총 30명 내외의 사용자를 대상으로 실험을 수행한다. 사용성 테스트는 실제 사용 시나리오를 반영한 과제 기반 평가 방식

으로 진행된다.

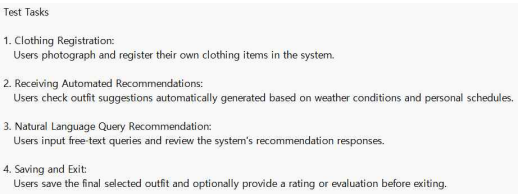


Fig. 18 Overview of the Test Task

참여자는 그림 18에 제시된 4단계의 절차를 수행하며 각 단계에서 인터페이스 반응과 추천 품질을 평가하게 된다.

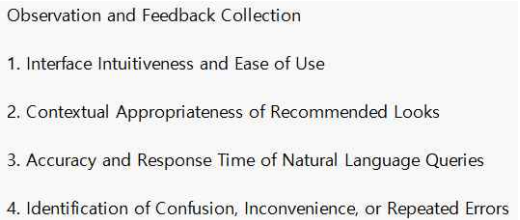


Fig. 19 Process of Observation and Feedback Collection

실험 과정에서 연구자는 참여자의 행동을 관찰하고 그림 19에 제시된 4가지 핵심 평가 기준을 바탕으로 심층 설문 및 인터뷰를 진행하여 문제점과 개선 사항을 도출한다. 결론적으로 본 연구의 평가 체계는 모델의 정확도를 기반으로 한 정량적 지표(40%)와 사용자 경험 중심의 정성적 지표(60%)를 통합하여 운영된다. 특히 다양한 성향을 가진 일반 사용자 외에도 패션 전문가 그룹의 피드백을 반영함으로써 제안하는 추천 시스템이 실제 현장에서 얼마나 실용적이고 적합하게 작동할 수 있는지를 다각도로 검증하도록 설계되었다.

V. 결 론

본 논문에서는 사용자의 보유 의류 데이터와 실시간 환경 정보를 결합하여 최적의 스타일을 제안하고 시각화하는 대화형 스마트 옷장 시스템을 제안하고 구현하였다. 제안된 시스템은 YOLOv8과 ViT를 활용한 계층적 의류 인식, 프롬프트 엔지니어링을 통한 LLM 기반

의 맥락 인식 추천, 그리고 RTM Pose와 MediaPipe를 활용한 관절 기반 가상 피팅 기능을 통합한 멀티모달 아키텍처로 설계되었다. 또한 실제 옷장 외관에 디지털 스크린을 결합한 사용자 인터페이스를 상정하여 기술과 생활 가구 간의 직관적인 상호작용 가능성을 제시하였다.

이를 통해 본 시스템은 기상 환경, 사용자의 일정, 성별 및 개인적 취향 등 복잡한 맥락을 반영한 정교한 스타일 추천이 가능하며 사용자의 의도에 맞는 의류 선택을 효율적으로 지원할 수 있음을 확인하였다. 정량적 평가를 통해 YOLO와 ViT 모델이 의류 탐지 및 속성 분류에서 높은 정확도를 기록함을 입증하였으며 설계된 정성적 평가 프레임워크를 통해 패션 전문가와 일반 사용자 모두에게 실용적인 사용자 경험을 제공할 수 있는 기반을 마련하였다. 특히 자연어 질의를 통해 상황을 묘사하고 추천을 받는 구조는 기존의 경직된 추천 시스템과 비교하여 훨씬 유연하고 인간 중심적인 인터페이스를 제공한다는 점에서 학술적 및 실무적 측면에서 유의미한 시사점을 제공한다.

향후에는 본 시스템의 활용 범위를 확장하고 서비스의 완성도를 높이기 위해 다음과 같은 방향으로 연구를 지속하고자 한다. 첫째, 스타일러, 세탁기, 건조기 등 스마트 홈 가전과의 연동을 통해 의류의 위생 상태나 세탁 주기를 자동으로 관리하는 능동형 의류 케어 시스템으로 고도화할 예정이다. 둘째, 더 나아가 개인화된 패션 데이터 분석 결과를 바탕으로 이커머스 플랫폼과 연동하여 사용자에게 부족한 아이템을 제안하는 등 비즈니스 모델로서의 확장 가능성을 타진할 계획이다. 이러한 시도는 인공지능 소프트웨어 기술이 단순한 편리함을 넘어 사람들의 일상에 실질적인 가치를 더하고 지속 가능한 패션 생태계를 구축하는 데 기여할 것으로 기대된다.

ACKNOWLEDGEMENTS

This paper was written as part of the “AI Graduation Project” course at Dongyang Mirae University. Through this course, an integrated understanding of AI-based service planning and implementation was developed and applied to the design of a smart wardrobe system.

REFERENCES

- [1] S. M. Kim and M. J. Park, “A study on the development of clothing coordination system,” *Journal of Digital Convergence*, vol. 15, no. 4, pp. 505 – 513, Apr. 2014. Available: <https://koreascience.kr/article/JAKO201414938222305.pdf>
- [2] J. H. Kim, “Deep learning-based fashion image analysis technology and service trends,” *Journal of the Korea Contents Association*, vol. 15, no. 11, pp. 818 – 828, Nov. 2015. Available: <https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE06542365>
- [3] H. J. Jang and J. E. Oh, “Trends in deep learning-based fashion image retrieval technology,” *Journal of the Korea Contents Association*, vol. 20, no. 4, pp. 23 – 32, Apr. 2020. Available: <https://koreascience.kr/article/JAKO202002761569551.pdf>
- [4] V. Poncelet and Y. Ren, Virtual fitting rooms: An evaluation of consumer trust and purchase intention, Master’s thesis, Uppsala University, Sweden, 2023. Available: <https://www.diva-portal.org/smash/get/diva2:1816013/FULLTEXT01.pdf>
- [5] H. Sharma, Exploring the impact of AI on the future of fashion retail: Consumer perception and interaction, Master’s thesis, Dublin Business School, Ireland, 2023. Available: <https://esource.dbs.ie/server/api/core/bitstreams/3e9b4b2f-c698-4475-8675-056cd0d2ebb4/content>

- [6] S. Lee and M. Choi, "FashionAI: Generating fashion style comments based on clothing attributes using large language models," in Proc. 2023 IEEE Int. Conf. Consumer Electron. (ICCE), 2023, pp. 1 - 4. DOI: 10.1109/ICCE56470.2023.10675715.
- [7] Y. Gao et al., "Chat-REC: Towards Interactive and Explainable LLMs-Augmented Recommender System," arXiv preprint arXiv:2303.14524, 2023. Available: <https://arxiv.org/pdf/2303.14524>
- [8] J. Wei et al., "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models," in Advances in Neural Information Processing Systems (NeurIPS), vol. 35, 2022, pp. 24824-24837. Available: <https://arxiv.org/pdf/2201.11903>
- [9] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788. DOI: 10.1109/CVPR.2016.91
- [10] A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in International Conference on Learning Representations (ICLR), 2021. Available: <https://arxiv.org/pdf/2010.11929>
- [11] F. Jan, Z. A. Khan, R. Ahmad, Z. Khan, and Z. Mumtaz, "Benchmarking YOLO Variants for Thermal Image Object Detection in Low-Light Environments," Multidisciplinary Surgical Research Annals, vol. 3, no. 4, 2025. Available: <https://zenodo.org/records/17310153>



백지민 (Baek-Ji Min)

동양미래대학교 인공지능소프트웨어학과 학사 과정 재학 중
 ※관심분야 : 인공지능