

속성 기반 객체 재식별 및 보행자 속성 인식(PAR) 모델에 대한 포괄적 심층 연구 보고서

1. 서론: 객체 재식별의 패러다임 전환과 속성 인식의 부상

컴퓨터 비전 분야에서 객체 재식별(Object Re-identification, Re-ID), 특히 사람 재식별(Person Re-ID)은 스마트 시티 구축, 지능형 영상 보안, 그리고 범죄 수사 등 공공 안전 영역에서 핵심적인 역할을 수행해 왔다. 전통적인 Re-ID 시스템은 서로 겹치지 않는(non-overlapping) 다중 카메라 네트워크에서 특정 보행자의 시각적 특징(Visual Features)을 매칭하여 동일 인물을 검색하는 것을 목표로 한다.¹ 그러나 조명 변화, 자세(Pose)의 다양성, 카메라 시점의 차이, 그리고 물체에 의한 가려짐(Occlusion)과 같은 환경적 제약은 순수 시각 기반 매칭 성능의 한계를 드러내었다.³

이러한 한계를 극복하기 위해 등장한 개념이 바로 **보행자 속성 인식(Pedestrian Attribute Recognition, PAR)**과의 융합이다. PAR은 이미지 내 보행자의 성별, 연령, 의복 색상, 가방 유무 등 고차원적인 의미론적 속성(Semantic Attributes)을 예측하는 기술이다.⁵ 시각적 특징 벡터(Feature Vector)가 이미지의 픽셀 수준 정보에 의존하는 반면, 속성 정보는 인간이 이해할 수 있는 명시적(Explicit) 정보를 제공하며, 이는 시점이나 조명이 변하더라도 비교적 불변하는(Invariant) 특성을 가진다.

본 보고서는 2023년부터 2025년 초까지 발표된 최신 연구 논문들을 망라하여, 객체 재식별과 속성 인식이 결합된 최신 모델 아키텍처, 학습 방법론, 그리고 벤치마크 성능을 심층 분석한다. 특히, 합성곱 신경망(CNN)에서 **비전 트랜스포머(Vision Transformer, ViT)**로의 아키텍처 전환, **그래프 합성곱 신경망(Graph Convolutional Network, GCN)**을 이용한 속성 상관관계 모델링, 그리고 최근 대규모 언어 모델(Large Language Model, LLM) 및 **비전-언어 모델(Vision-Language Model, VLM)**의 도입을 통한 '대화형 재식별(ChatReID)' 및 '프롬프트 학습(Prompt Learning)'의 등장을 중점적으로 다룬다.

2. 딥러닝 기반 특징 추출의 진화: CNN에서 트랜스포머로

오랫동안 Re-ID와 PAR 분야는 ResNet-50과 같은 심층 CNN 아키텍처가 주도해 왔다.⁷ CNN은 이미지의 국소적(Local) 특징을 추출하는 데 탁월하지만, 이미지 전체를 아우르는 전역적(Global) 문맥 정보를 포착하는 데에는 수용 영역(Receptive Field)의 한계가 있었다.

2.1 트랜스포머(Transformer) 기반 방법론의 대두

최근 연구 동향은 CNN의 귀납적 편향(Inductive Bias)을 넘어서 수 있는 트랜스포머 기반 모델로 급격히 이동하고 있다.¹ 트랜스포머의 핵심인 자기 주의(Self-Attention) 메커니즘은 이미지 내 모든 패치 간의 관계를 동시에 모델링할 수 있어, 신체 부위가 가려지거나 정렬되지

않은 상황에서도 강건한 성능을 발휘한다.

2.1.1 TransReID: 순수 트랜스포머의 한계 극복

TransReID는 객체 재식별을 위해 설계된 최초의 순수 트랜스포머 기반 아키텍처 중 하나이다.⁸ 이 모델은 비전 트랜스포머(ViT)를 Re-ID 작업에 적용할 때 발생하는 두 가지 주요 문제를 해결했다.

1. 국소 정보 손실: ViT는 이미지를 패치 단위로 처리하므로 CNN에 비해 미세한 국소 특징(예: 옷의 로고, 신발의 패턴)을 놓칠 수 있다. TransReID는 **Jigsaw Patch Module (JPM)**을 도입하여 패치들을 섞어 학습시킴으로써 모델이 위치 정보에 과도하게 의존하지 않고 강력한 국소 특징을 학습하도록 유도한다.
2. 카메라 도메인 격차: 카메라마다 다른 시점과 조명 특성을 반영하기 위해 **사이드 정보 임베딩(Side Information Embedding, SIE)**을 도입했다. 이는 위치 임베딩(Positional Embedding)에 카메라 및 뷰포인트 정보를 추가하여 인코딩하는 방식이다.

2.1.2 AAformer: 자동 정렬 메커니즘

보행자 이미지에서 사람의 신체 부위와 배경, 혹은 소지품(가방 등)을 명확히 구분하는 것은 매우 중요하다. **AAformer (Auto-Aligned Transformer)**는 트랜스포머 아키텍처 내에 정렬(Alignment) 스킴을 도입하여, 패치 수준에서 인간의 신체 부위와 비신체 부위를 자동으로 위치시키는 방법을 제안했다.⁸ 이는 'Part Token'이라는 학습 가능한 벡터를 사용하여, 이미지 패치들을 의미론적 부위(예: 머리, 상체, 하체)별로 군집화하고, 이를 통해 미정렬(Misalignment) 문제를 해결한다.

2.2 하이브리드 아키텍처와 어텐션 메커니즘

CNN의 공간적 구조 정보와 트랜스포머의 전역적 문맥 정보를 결합하려는 시도 또한 활발하다.

- **HAT (Hierarchical Aggregation Transformer):** 저수준의 구조적 특징과 고수준의 의미적 문맥을 계층적으로 결합하여 성능을 극대화한다.⁸
- **PTCR (Pyramidal Transformer with Conv-Patchify):** CNN의 합성곱 연산을 패치화(Patchify) 과정에 도입하여, 트랜스포머 입력 단계에서부터 풍부한 공간 정보를 보존하도록 설계되었다.⁸

3. 속성 간 상관관계 모델링: 그래프 합성곱 신경망(GCN)의 적용

보행자 속성 인식(PAR)에서 각 속성은 독립적으로 존재하지 않는다. 예를 들어, '치마'를 입은 사람은 '여성'일 확률이 높고, '운동화'를 신은 사람은 '정장'을 입을 확률이 낮다. 이러한 속성 간의 공기(Co-occurrence) 관계와 위상적 구조를 모델링하기 위해 **그래프 합성곱 신경망(GCN)**이 핵심 기술로 부상했다.

3.1 Tran-GCN: 트랜스포머와 GCN의 결합

2025년 4월 *IET Image Processing*에 게재된 **Tran-GCN**은 감시 비디오 환경에서 발생하는 자세 변화와 가려짐 문제를 해결하기 위해 제안된 **최신 SOTA 모델**이다.¹⁰

3.1.1 아키텍처 상세 분석

Tran-GCN은 네 가지 핵심 구성 요소의 **유기적 결합**으로 이루어져 있다.

1. **자세 추정 학습 분기 (Pose Estimation Learning Branch):** VGG-16 기반의 백본을 사용하여 보행자의 **관절(Key points)** 위치를 추정한다. 여기서 생성된 **신뢰도 맵(Certainty Maps)**과 **친화도 필드(Affinity Fields)**는 보행자의 구조적 정보를 제공하며, 이는 단순한 시각적 특징이 놓칠 수 있는 **신체 구조의 불변성**을 포착한다.
2. **트랜스포머 학습 분기 (Transformer Learning Branch):** 이미지 패치 간의 **전역적 의존성**을 학습하여, **국소적으로 분산된 특징들을 의미론적으로 연결**한다.
3. **합성곱 학습 분기 (Convolution Learning Branch):** ResNet-50(마지막 풀링 레이어 제거)을 사용하여 **보행자의 미세한 질감 및 색상 특징**을 추출한다.
4. **그래프 합성곱 모듈 (Graph Convolutional Module, GCM):** 이것이 **모델의 핵심**이다. GCM은 위 세 가지 분기에서 추출된 특징들을 **그래프의 노드(Node)**로 정의한다. **노드 간의 연결(Edge)**은 학습된 **인접 행렬(Adjacency Matrix)**에 의해 정의되는데, 이 행렬은 **자세 추정 분기**에서 얻은 **구조적 정보**를 바탕으로 **가중치가 부여**된다. 이를 통해 모델은 **시각적 특징, 전역적 문맥, 그리고 신체 구조 정보**를 통합하여 **매우 강력한 식별력(Discriminative Power)**을 갖게 된다.

3.1.2 성능 평가

Market-1501, DukeMTMC-ReID, MSMT17 데이터셋에서 수행된 실험 결과, **Tran-GCN**은 기존의 **CNN 기반 방법론**이나 **순수 트랜스포머 모델 대비 월등한 정확도**를 보였다. 특히 **복잡한 배경**이나 **심한 가려짐**이 존재하는 **MSMT17 데이터셋**에서 **그 효용성이 입증**되었다.¹²

3.2 MCGCN: 다중 상관관계 그래프

****MCGCN (Multi-Correlation Graph Convolutional Network)****은 **속성 인식의 정확도**를 높이기 위해 세 가지 유형의 그래프를 통합적으로 운용한다.¹³

1. **의미 그래프 (Semantic Graph):** 속성 간의 **논리적 관계**(예: 온톨로지 지식)를 기반으로 구축되며, **데이터셋의 사전 지식(Prior Knowledge)**을 활용해 초기화된다.
2. **시각 그래프 (Visual Graph):** 이미지에서 추출된 **2D 특징 맵**을 **노드로 투영**하여, 각 속성이 이미지의 **어느 영역(Region)**과 관련이 있는지를 **공간적으로 모델링**한다.
3. **합성 그래프 (Synthesis Graph):** 의미 그래프와 시각 그래프를 **연결(Inter-graph edges)**하여, **시각적 정보가 의미적 추론을 보정**하고 반대로 **의미적 지식이 시각적 주의(Attention)**를 유도하는 **상호 보완적 학습**을 수행한다.

이러한 구조는 데이터 불균형 문제를 완화하기 위한 **suit_polyloss 전략**과 결합되어, **소수 클래스**(예: 모자 착용, 안경 착용 등)의 인식률을 크게 향상시켰다.

4. 비전-언어 모델(VLM)의 융합과 프롬프트 학습의 혁신

2024년의 가장 큰 기술적 도약은 **CLIP (Contrastive Language-Image Pre-training)**과 같은 대규모 비전-언어 모델을 **Re-ID** 및 **PAR** 작업에 적용한 것이다. 이는 시각 정보만을 이용하던 기존 방식에서 탈피하여, 텍스트(속성 설명)와 이미지를 동일한 임베딩 공간에 매핑함으로써 인식 성능과 일반화 능력을 획기적으로 개선했다.

4.1 PromptPAR: 파라미터 효율적 전이 학습

PromptPAR는 **CLIP**과 같은 거대 모델을 전체 미세 조정(Full Fine-tuning)하는 데 드는 막대한 비용과 데이터 요구량을 줄이기 위해 제안되었다.¹⁴

- 핵심 메커니즘: **PromptPAR**는 사전 학습된 **CLIP**의 이미지 및 텍스트 인코더를 고정(Freeze)하고, 오직 **프롬프트 벡터(Prompt Vectors)**와 분류 헤드(Classification Head)만을 학습시킨다. 학습되는 파라미터는 전체 모델의 0.75%에 불과하다.
- 영역 인지 프롬프트(**Region-Aware Prompting**): 단순한 텍스트 프롬프트 대신, 이미지의 특정 영역(예: 상체, 하체)에 특화된 프롬프트를 학습하여, 모델이 속성을 예측할 때 해당 신체 부위에 집중하도록 유도한다.
- 성과: 이 방법론은 **PA-100K**, **RAPv2** 등 주요 벤치마크에서 새로운 **SOTA(State-of-the-Art)**를 기록했으며, 특히 **훈련 데이터에 없는 속성을 인식해야 하는 제로샷(Zero-shot) 설정**에서도 탁월한 성능을 보였다.¹⁶

4.2 ViTA-PAR: 시각-텍스트 정렬의 고도화

ViTA-PAR는 "Open-Set" 속성 인식을 목표로 한다.¹⁴ 기존 모델들이 미리 정의된 속성 클래스(Closed-Set)만 인식할 수 있었던 반면, **ViTA-PAR**는 자연어 설명을 통해 새로운 속성을 유연하게 인식한다.

- 시각적 속성 프롬프트(**Visual Attribute Prompts**): 전역적 특징에서 국소적 특징으로 이어지는 계층적 의미를 포착한다.
- 문맥 프롬프팅(**Context Prompting**): "사람"과 "속성" 간의 문맥적 관계를 학습하는 템플릿을 도입하여 텍스트 임베딩을 풍부하게 만든다.

5. 대규모 언어 모델(LLM)과 생성형 Re-ID의 시대

텍스트-이미지 매칭을 넘어, **LLM**의 추론 능력을 직접적으로 활용하는 연구들이 2025년을 기점으로 폭발적으로 증가하고 있다. 이는 단순히 "매칭"하는 것을 넘어 **시스템과 "대화"하며 대상을 찾는 패러다임**을 제시한다.

5.1 ChatReID: 텍스트 주도형 검색 패러다임

2025년 *arXiv*에 공개된 **ChatReID**는 기존의 **VLM** 기반 방법론들이 가진 한계(고정된 템플릿 의존, 복잡한 추론 불가)를 극복하기 위해 제안되었다.¹⁷

- 계층적 점진적 튜닝 (**Hierarchical Progressive Tuning, HPT**): **ChatReID**는 세 단계의 튜닝 전략을 사용한다.
 1. **1단계 - 보행자 속성 이해**: 이미지 캡셔닝(Image Captioning) 작업을 통해 모델이 보행자의 시각적 특징을 텍스트로 상세하게 기술하는 능력을 학습한다. 이는 모델이

단순히 이미지를 벡터로 압축하는 것이 아니라, "빨간 옷", "배낭" 등의 의미를 명확히 이해하도록 강제한다.

2. **2단계 - 미세 이미지 검색:** 텍스트-이미지 및 이미지-이미지 간의 정밀한 매칭 능력을 학습한다.
 3. **3단계 - 다중 모달 추론:** 복잡한 사용자 질의(예: "파란 옷을 입은 사람 중 배낭을 멘 남자를 찾아줘")를 처리할 수 있는 논리적 추론 능력을 강화한다.
- 성과: ChatReID는 10개의 벤치마크에서 SOTA를 달성했으며, 특히 사용자가 반복적으로 질의를 수정하며 검색을 좁혀가는 대화형 검색 시나리오에서 독보적인 성능을 보여준다.¹⁸

5.2 LLM-PAR: 언어 모델 증강 프레임워크

LLM-PAR는 LLM을 활용해 시각적 분류 성능을 높이는 증강(Augmentation) 프레임워크이다.²⁰

- **AGFA 모듈:** 속성 그룹별 특징을 집계(Aggregate)하고, 이를 Q-former를 통해 LLM에 입력한다. LLM은 이를 바탕으로 풍부한 보행자 묘사 텍스트를 생성하고, 이 텍스트는 다시 시각적 특징 학습의 보조 신호(Auxiliary Supervision)로 사용된다.
- 결과: PETA 데이터셋에서 기존 최고 모델인 SSPNet 대비 mA(mean Accuracy)를 3.52% 향상시키는 등, LLM의 언어적 지식이 시각적 모호성을 해소하는 데 크게 기여함을 입증했다.

5.3 MP-ReID: 다중 프롬프트 학습

MP-ReID는 하나의 이미지에 대해 다양하고 풍부한 텍스트 설명을 생성하여 검색 성능을 높이는 전략을 취한다.²²

- 명시적/암시적 프롬프트: ChatGPT와 VQA(Visual Question Answering) 모델을 앙상블하여 "안경을 쓴 남자", "체크무늬 셔츠"와 같은 명시적 속성 문장을 생성하고, 동시에 학습 가능한 암시적 토큰을 통해 문맥 정보를 보완한다.
- 교차 모달 정렬: 이렇게 생성된 다중 프롬프트는 이미지 특징과 교차 어텐션(Cross-Attention)을 통해 융합되어, 단일 레이블 기반 학습보다 훨씬 강력한 임베딩 공간을 형성한다.

6. 설명 가능한 AI (XAI)와 속성 중요도 분석

딥러닝 모델, 특히 트랜스포머와 같은 거대 모델의 가장 큰 단점은 '블랙박스'라는 점이다. 모델이 왜 두 이미지를 같다고 판단했는지 알기 어렵다. MoSAIC-ReID는 이러한 불투명성을 해결하기 위해 제안되었다.²³

6.1 MoSAIC-ReID: 전문가 혼합(MoE) 방법론

이 연구는 전문가 혼합(Mixture of Experts, MoE) 아키텍처를 도입하여 Re-ID 모델이 실제로 어떤 속성에 의존하는지를 정량적으로 분석했다.

- **LoRA 기반 전문가:** 각 전문가는 LoRA(Low-Rank Adaptation) 모듈로 구성되며, 특정 속성(예: 상의 색상, 가방 유무)에만 반응하도록 특화된다.
- **오라클 라우터(Oracle Router):** 분석 단계에서 정답(Ground-truth) 속성 정보를 기반으로

특정 전문가만을 활성화(Activate)하거나 비활성화하여, 해당 속성이 최종 식별 성능에 미치는 인과적 영향력(Causal Impact)을 측정한다.

6.2 분석 결과 및 통찰

MoSAIC-ReID를 통한 대규모 분석 결과는 다음과 같은 중요한 통찰을 제공한다.²³

- 1. **색상의 지배력**: 현재의 SOTA 모델들은 여전히 의복의 색상(특히 하의 색상)에 과도하게 의존한다. 이는 **모델이** 사람이 인지하는 '신원'보다는 색상 히스토그램 유사도에 편향되어 있음을 시사한다.
- 2. **액세서리의 미미한 영향**: 가방이나 모자와 같은 액세서리는 보행자를 구별하는 데 중요한 단서임에도 불구하고, 실제 모델의 판단에는 제한적인 영향만 미치는 것으로 나타났다. 이는 데이터셋 내에서 해당 속성의 빈도수가 낮기 때문(Long-tail 분포)으로 분석된다.

7. 데이터셋 벤치마크: 포화 상태와 새로운 도전

모델의 발전은 데이터셋의 진화와 함께 이루어졌다. 2024년 이후의 연구들은 기존 데이터셋의 성능 포화를 지적하며, 더 현실적이고 가혹한 환경을 반영하는 새로운 벤치마크를 제안하고 있다.

7.1 주요 데이터셋 현황 및 성능 비교

7.1.1 보행자 속성 인식 (PAR) 벤치마크

PAR 분야에서는 PA-100K와 PETA가 표준으로 자리 잡았다.

데이터셋	이미지 수	속성 수	특징	최신 SOTA 모델	성능 (F1 / mA)
PA-100K ²⁷	100,000	26	최대 규모 야외 감시 카메라 데이터	LLM-PAR ²⁰	91.09% / 88.xx%
PETA ²⁸	19,000	61	다양한 데이터셋의 혼합, 해상도 다양	LLM-PAR ²⁰	90.02% / 92.20%
RAPv2 ²⁹	84,928	72	실내/실외 혼합, 장기간 수집	PromptPA R ¹⁶	81.00% (F1)

데이터 해석: LLM-PAR와 PromptPAR의 등장은 F1 스코어 90%의 벽을 넘어서게 했다. 특히 LLM-PAR는 PETA 데이터셋에서 기존 방법론 대비 mAP를 3.52%나 끌어올리며 LLM 증강의 위력을 증명했다.

7.1.2 객체 재식별 (Person Re-ID) 벤치마크

Re-ID 분야는 **Market-1501**과 **DukeMTMC-reID**가 표준이나, 최근 성능 포화로 인해 **MSMT17**이나 **Occluded-Duke**와 같은 난이도 높은 데이터셋이 선호된다.

데이터셋	특징	최신 SOTA 모델	Rank-1 정확도	mAP
Market-1501 <small>30</small>	1,501 IDs, 6 Cameras	ChatReID ¹⁹	96.4%	93.5%
DukeMTMC-reID	야외 트래킹 시나리오	Tran-GCN ¹¹	> 90%	High
MSMT17	복잡한 배경, 시간대 변화	ChatReID ¹⁹	87.5%	72.4%

데이터 해석: Market-1501에서의 Rank-1 정확도가 96%를 상회함에 따라, 단순한 매칭 정확도보다는 ****mINP (mean Inverse Negative Penalty)****와 같은 지표를 통해 가장 찾기 어려운 샘플(Hardest Positive)을 얼마나 잘 찾는지가 더 중요해지고 있다.⁸

7.2 차세대 데이터셋: 현실 세계의 복잡성 반영

기존 데이터셋의 한계(제한된 환경, 깨끗한 이미지)를 극복하기 위해 2024~2025년에 공개된 데이터셋들은 다음과 같은 특징을 가진다.

1. MSP60K (Multi-Scenario Pedestrian 60K):

- ²⁰에 의해 제안됨.
- 규모: 60,122장 이미지, 57개 속성, 8개 시나리오.
- 혁신점: 단순히 이미지를 수집한 것이 아니라, 블러(Blur), 가려짐, 노이즈, 압축 손실 등 현실 세계 CCTV에서 발생하는 **열화 요인을 인위적으로 합성**하여 **"파괴적인(Destructive)"** 테스트 환경을 제공한다. 이는 실험실 환경과 실제 배포 환경 간의 **성능 격차(Domain Gap)**를 줄이기 위함이다.

2. EventPAR:

- ¹⁵에 의해 공개된 최초의 대규모 **RGB-Event** 데이터셋.
- 이벤트 카메라(**Event Camera**): **픽셀의 밝기 변화만을 비동기적으로 기록하는 뉴로모픽 센서를 활용**한다. 이는 **120dB** 이상의 다이내믹 레인지와 마이크로초 단위의 반응 속도를 가져, 기존 RGB 카메라가 실패하는 야간이나 고속 이동 상황에서도

보행자를 식별할 수 있게 한다.

- **감정 속성 포함**: 외형뿐만 아니라 6가지의 인간 감정 속성까지 포함하여, 행동 분석(Behavior Analysis)으로의 확장 가능성을 열었다.

3. UPAR (Unified Pedestrian Attribute Recognition) Challenge:

- ³¹에서 주관. PETA, PA-100K, Market-1501 등을 통합(Harmonization)하여, 훈련된 도메인과 전혀 다른 도메인(예: MEVID 비디오 데이터셋)에서 테스트하는 교차 도메인 일반화(**Cross-Domain Generalization**) 능력을 평가한다.

8. 미래 연구 방향 및 도전 과제

현재 기술의 발전 속도를 고려할 때, 향후 2~3년 내에 다음과 같은 분야에서 중요한 진전이 있을 것으로 예상된다.

8.1 개방형 세계(Open-World) 및 제로샷 Re-ID

훈련 데이터에 없는 새로운 속성이나 신원을 식별하는 능력은 필수적이다. **ViTA-PAR**와 **OAPR (Open-Attribute Recognition)** ¹⁴ 연구는 고정된 클래스 분류기 대신 자연어 텍스트 임베딩을 활용하여, "한 번도 본 적 없는" 속성을 가진 사람을 검색하는 방향으로 나아가고 있다.

8.2 엣지 컴퓨팅 및 경량화

드론(UAV)이나 CCTV 엣지 단말에서 실시간으로 Re-ID를 수행하기 위해서는 모델의 경량화가 시급하다. **PromptPAR**와 같은 파라미터 효율적 튜닝 기법은 거대 모델의 지식을 유지하면서도 연산 비용을 최소화하는 해결책이 될 수 있다.³²

8.3 적대적 공격 방어 (Adversarial Robustness)

보안 시스템으로서의 신뢰성을 확보하기 위해 적대적 공격(Adversarial Attack)에 대한 방어 기술이 중요해지고 있다. **OpenPAR** 툴킷에 포함된 **ASL-PAR (Adversarial Semantic and Label Perturbation)** 모듈은 보행자 이미지에 미세한 노이즈를 섞어 속성 인식을 교란시키는 공격을 시뮬레이션하고, 이를 방어하는 연구를 촉진하고 있다.¹⁵

8.4 텍스트-이미지 검색의 표준화

"속성 인식"과 "텍스트 기반 사람 검색"의 경계가 허물어지고 있다. **ChatReID**와 같은 프레임워크는 속성 레이블을 자연어의 하위 집합으로 간주하며, 미래의 Re-ID 시스템은 단순한 검색 엔진이 아니라 "어제 빨간 셔츠를 입고 지나간 사람이 오늘은 무슨 옷을 입었는지 추론"할 수 있는 지능형 대화 에이전트로 발전할 것이다.

9. 결론

본 보고서는 2023년부터 2025년 현재까지의 속성 기반 객체 재식별 및 보행자 속성 인식 기술의 비약적인 발전을 조망하였다. 연구의 중심축은 **CNN 기반의 정적 특징 추출**에서 **트랜스포머와 GCN을 통한 구조적/전역적 관계 모델링**으로, 그리고 더 나아가 **대규모 언어**

모델을 활용한 의미론적 추론 및 대화형 검색으로 이동하고 있다.

특히 **Tran-GCN**과 같은 구조적 모델링 기법은 영상 내의 가려짐과 자세 변화에 강건함을 더했고, **ChatReID**와 **PromptPAR**는 사전 학습된 방대한 지식을 보안 도메인에 효율적으로 전이시키는 방법을 제시했다. 또한, **MSP60K**와 **EventPAR**와 같은 차세대 데이터셋의 등장은 기존 벤치마크의 성능 포화를 넘어, 현실 세계의 극한 환경에서도 작동할 수 있는 시스템 개발을 가속화하고 있다.

이러한 기술적 진보는 향후 스마트 시티의 안전망을 강화하고, 실종자 수색이나 범죄 예방과 같은 사회적 문제를 해결하는 데 결정적인 기여를 할 것으로 기대된다.

[부록] 주요 알고리즘 및 기술 요약

기술 구분	대표 모델	핵심 메커니즘	주요 성과 (Dataset)	참고 문헌
GCN	Tran-GCN	Pose, ViT, CNN 특징을 그래프 노드로 통합하여 관계 학습	High Performance (MSMT17)	10
GCN	MCGCN	의미, 시각, 합성의 3중 그래프 구조로 속성 공기 관계 모델링	SOTA (PAR)	13
Prompt	PromptPAR	CLIP의 파라미터를 고정하고 0.75%의 프롬프트 벡터만 학습	F1 90.15% (PA-100K)	16
LLM	ChatReID	3단계 계층적 튜닝(HPT)을 통한 대화형	Rank-1 96.4% (Market-1501)	19

		검색 및 추론		
LLM	LLM-PAR	LLM이 생성한 캡션을 보조 신호로 사용하여 시각 특징 강화	mA +3.52% (PETA)	20
MoE	MoSAIC-ReID	속성별 전문가(LoRA) 모듈과 오라클 라우터를 통한 해석 가능성 확보	속성 중요도 정량화	25
XAI	ALM	약지도 학습 기반의 속성별 영역 자동 탐지 (FPN 활용)	SOTA (PA-100K)	33

참고 자료

1. [2401.06960] Transformer for Object Re-Identification: A Survey - arXiv, 12월 24, 2025에 액세스, <https://arxiv.org/abs/2401.06960>
2. A Generative Approach to Person Reidentification - MDPI, 12월 24, 2025에 액세스, <https://www.mdpi.com/1424-8220/24/4/1240>
3. Identifying Re-identification Challenges: Past, Current and Future Trends, 12월 24, 2025에 액세스, <https://d-nb.info/1352060272/34>
4. AANet: Attribute Attention Network for Person Re-Identifications - CVF Open Access, 12월 24, 2025에 액세스, https://openaccess.thecvf.com/content_CVPR_2019/papers/Tay_AANet_Attribute_Attention_Network_for_Person_Re-Identifications_CVPR_2019_paper.pdf
5. (PDF) Pedestrian Attribute Recognition: A Survey - ResearchGate, 12월 24, 2025에 액세스, https://www.researchgate.net/publication/330553565_Pedestrian_Attribute_Recognition_A_Survey
6. Attribute-Guided Pedestrian Retrieval: Bridging Person Re-ID with ..., 12월 24, 2025에 액세스, https://www.researchgate.net/publication/384217932_Attribute-Guided_Pedestrian_Retrieval_Bridging_Person_Re-ID_with_Internal_Attribute_Variability
7. Transformer for Object Re-Identification: A Survey - arXiv, 12월 24, 2025에 액세스, <https://arxiv.org/html/2401.06960v2>
8. [PDF] Transformer for Object Re-identification: A Survey - Semantic Scholar, 12월

- 24, 2025에 액세스,
<https://www.semanticscholar.org/paper/Transformer-for-Object-Re-Identification%3A-A-Survey-Ye-Chen/536d1582424a51f91c53b442b0b55a65959f98ad>
9. Person Re-Identification via Multi-Dimensional Attention Mechanisms - ResearchGate, 12월 24, 2025에 액세스,
https://www.researchgate.net/publication/394003376_Person_Re-Identification_via_Multi-Dimensional_Attention_Mechanisms
 10. Tran-GCN: A Transformer-Enhanced Graph Convolutional Network for Person Re-Identification in Monitoring Videos - arXiv, 12월 24, 2025에 액세스,
<https://arxiv.org/html/2409.09391>
 11. Tran-GCN: A Transformer-Enhanced Graph Convolutional Network for Person Re-Identification in Monitoring Videos - ResearchGate, 12월 24, 2025에 액세스,
https://www.researchgate.net/publication/391284730_Tran-GCN_A_Transformer-Enhanced_Graph_Convolutional_Network_for_Person_Re-Identification_in_Monitoring_Videos
 12. A Transformer-Enhanced Graph Convolutional Network for Person Re-Identification in Monitoring Videos - arXiv, 12월 24, 2025에 액세스,
<https://www.arxiv.org/pdf/2409.09391>
 13. MCGCN: Multi-Correlation Graph Convolutional Network for ..., 12월 24, 2025에 액세스,
https://www.researchgate.net/publication/378651209_MCGCN_Multi-Correlation_Graph_Convolutional_Network_for_Pedestrian_Attribute_Recognition
 14. Open-Attribute Recognition for Person Retrieval: Finding People Through Distinctive and Novel Attributes - arXiv, 12월 24, 2025에 액세스,
<https://arxiv.org/html/2508.01389v1>
 15. [OpenPAR] An open-source framework for Pedestrian Attribute Recognition, based on PyTorch - GitHub, 12월 24, 2025에 액세스,
<https://github.com/Event-AHU/OpenPAR>
 16. Pedestrian Attribute Recognition via CLIP based Prompt Vision-Language Fusion - arXiv, 12월 24, 2025에 액세스, <https://arxiv.org/html/2312.10692v1>
 17. ChatRelD: Open-ended Interactive Person Retrieval via Hierarchical Progressive Tuning for Vision Language Models - arXiv, 12월 24, 2025에 액세스,
<https://arxiv.org/html/2502.19958v1>
 18. ChatRelD: Open-ended Interactive Person Retrieval via Hierarchical Progressive Tuning for Vision Language Models - arXiv, 12월 24, 2025에 액세스,
<https://arxiv.org/html/2502.19958v2>
 19. arXiv:2502.19958v1 [cs.CV] 27 Feb 2025, 12월 24, 2025에 액세스,
<https://arxiv.org/pdf/2502.19958>
 20. Pedestrian Attribute Recognition: A New Benchmark Dataset and A Large Language Model Augmented Framework - arXiv, 12월 24, 2025에 액세스,
<https://arxiv.org/html/2408.09720v1>
 21. [Quick Review] Pedestrian Attribute Recognition: A New Benchmark Dataset and A Large Language Model Augmented Framework - Liner, 12월 24, 2025에 액세스,
<https://liner.com/review/pedestrian-attribute-recognition-new-benchmark-dataset-and-large-language-model>

22. Multi-Prompts Learning with Cross-Modal Alignment for Attribute ..., 12월 24, 2025에 액세스, <https://ojs.aaai.org/index.php/AAAI/article/view/28524/29021>
23. [2512.08697] What really matters for person re-identification? A Mixture-of-Experts Framework for Semantic Attribute Importance - arXiv, 12월 24, 2025에 액세스, <https://arxiv.org/abs/2512.08697>
24. [論文評述] What really matters for person re-identification? A Mixture-of-Experts Framework for Semantic Attribute Importance - Moonlight, 12월 24, 2025에 액세스, <https://www.themoonlight.io/tw/review/what-really-matters-for-person-re-identification-a-mixture-of-experts-framework-for-semantic-attribute-importance>
25. What really matters for person re-identification? A Mixture-of-Experts Framework for Semantic Attribute Importance - arXiv, 12월 24, 2025에 액세스, <https://arxiv.org/html/2512.08697v1>
26. What really matters for person re-identification? A Mixture-of-Experts Framework for Semantic Attribute Importance - ResearchGate, 12월 24, 2025에 액세스, https://www.researchgate.net/publication/398417676_What_really_matters_for_person_re-identification_A_Mixture-of-Experts_Framework_for_Semantic_Attribute_Importance
27. PA-100K - Kaggle, 12월 24, 2025에 액세스, <https://www.kaggle.com/datasets/yuulind/pa-100k>
28. Unified Pedestrian Attribute Recognition (UPAR) Dataset - GitHub, 12월 24, 2025에 액세스, https://github.com/speckean/upar_dataset
29. wangxiao5791509/Pedestrian-Attribute-Recognition-Paper-List - GitHub, 12월 24, 2025에 액세스, <https://github.com/wangxiao5791509/Pedestrian-Attribute-Recognition-Paper-List>
30. Scalable Person Re-identification: A Benchmark - Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung IOSB, 12월 24, 2025에 액세스, https://www.iosb.fraunhofer.de/content/dam/iosb/iosbtest/bilder-videos/kompetenzen/bildauswertung/VID/publikationen/Scalable_Person_Re-identification_A_Benchmark.pdf
31. UPAR Challenge 2024: Pedestrian Attribute Recognition and Attribute-Based Person Retrieval - Dataset, Design, and Results - CVF Open Access, 12월 24, 2025에 액세스, https://openaccess.thecvf.com/content/WACV2024W/RWS/papers/Cormier_UPAR_Challenge_2024_Pedestrian_Attribute_Recognition_and_Attribute-Based_Person_Retrieval_WACVW_2024_paper.pdf
32. A Large Scale Benchmark of Person Re-Identification - MDPI, 12월 24, 2025에 액세스, <https://www.mdpi.com/2504-446X/8/7/279>
33. [Quick Review] Improving Pedestrian Attribute Recognition With Weakly-Supervised Multi-Scale Attribute-Specific Localization - Liner, 12월 24, 2025에 액세스, <https://liner.com/review/improving-pedestrian-attribute-recognition-with-weakly-supervised-multiscale-attributespecific-localization>