생성형 AI 광고 이미지 평가를 위한 벤치마크 데이터셋 및 자동 평가 지표 개발 연구
- 광고사진 표현 형식을 중심으로

A Study on the Development of a Benchmark Dataset and Automatic Evaluation Metrics for Evaluating AI-Generated Advertising Images
- Focusing on the Expressive Forms of Advertising Photography

선행연구

광고사진 표현 형식 기반 생성형 AI 이미지 벤치마크 설계 및 상용 모델 성능 비교 (심현준, 2026)
생성형 AI 시대 광고사진 제작 방식 변화에 대한 탐색적 연구 (심현준, 2025)
광고사진 산업의 생성형 AI 활용에 관한 전략적 분석 연구 (심현준, 2025)
광고사진의 표현 방식과 시각적 은유 유형에 따른 광고효과 연구 (심현준, 2022)
광고사진의 표현 형식이 광고효과에 미치는 영향_국내 가전제품 인쇄광고를 중심으로 (심현준, 2021)

KEYWORD
생성형 AI 광고 이미지, 광고사진 표현 형식, 2AFC, 전문가 평가, 도메인 특화 자동 평가

연 구 자 : 심현준 홍익대학교 대학원 AI·실감미디어콘텐츠학과 박사과정
지도교수 : 한정엽 홍익대학교 대학원 AI·실감미디어콘텐츠학과

목차

제1장 서론

1.1 연구배경
1.2 연구문제
1.3 연구목적
1.4 연구범위 및 방법
1.4.1 연구의 범위
1.4.2 연구의 방법

제2장 이론적 배경

2.1 광고사진의 표현과 평가
2.1.1 광고사진의 표현 형식
2.1.2 광고사진 산업과 제작 방식의 변화
2.1.3 광고 이미지 평가의 지각 차이와 품질 차원
2.2 벤치마크 데이터셋과 전문가 판단
2.2.1 벤치마크 데이터셋의 지형과 성능
2.2.2 광고 도메인 벤치마크와 기준 이미지 구조
2.2.3 전문가 판단 자료와 신뢰도
2.3 생성형 AI 이미지 평가 지표
2.3.1 지각 분포 기반 이미지 평가 지표
2.3.2 비전-언어 모델과 의미 기반 평가
2.3.3 도메인 특화 자동 평가 지표의 조정과 예측 원리

제3장 연구 설계

3.1 연구 개요
3.2 연구 1 설계
3.2.1 AD-ImageBench 자료의 구성
3.2.2 전문가 평가 설계
3.2.3 AD-LPIPS 파인튜닝 설계
3.2.4 AD-Score 학습 설계
3.3. 연구 2 설계
3.3.1 독립 테스트셋의 구성
3.3.2 역설계 프롬프트 작성과 AI 이미지 생성
3.3.3 평가 지표 및 통계 분석

제4장 연구 결과

4.1 연구 1 결과
4.1.1 AD-ImageBench의 구성과 표현 형식 분류
4.1.2 전문가 평가의 신뢰도
4.1.3 AD-LPIPS의 학습 결과와 성능 비교
4.1.4 AD-Score의 학습 결과와 전문가 평정값 예측 성능
4.2 연구 2 결과
4.2.1 네 가지 자동 평가 지표의 산출과 기술 통계
4.2.2 상용 생성 모델의 통계적 비교
4.2.3 표현 형식별 하위 집합 분석
4.2.4 전문가 다수결 판단과의 일치 및 불일치 분석
4.3 종합 결과
4.3.1 AD-LPIPS와 AD-Score의 결과 교차 확인
4.3.2 두 도구의 역할 분화와 해석 범위
4.3.3 표현 형식 분류의 실증적 역할
4.3.4 전문가 다수결 판단과 자동 지표의 관계

제5장 결론

5.1 연구의 요약 및 결론
5.2 연구의 시사점
5.2.1 이론적 함의
5.2.2 방법론적 함의
5.2.3 실무적 함의
5.3 연구의 한계 및 제언

연구 배경

생성형 AI

- 텍스트·이미지·영상 제작 전반에 확산, 시각 콘텐츠의 생산·소비 방식을 변화 (Ramesh et al., 2022; Rombach et al., 2022; 윤병권, 2024)
- 디지털 전환은 인간 창작자의 도구적 진화에 머물렀으나, 생성형 AI는 촬영 이전 단계에서 직접 이미지를 산출 (Manovich, 2018; 심은혜·현혜연, 2023)
- AI 기반 제작은 프롬프트 설계, 반복 생성, 후보 이미지 선별, 후반 조정이 결합된 방식 (장유진·이종근, 2023; 김활빈·박민아, 2024)
- 상용 생성 모델의 불투명성, 잦은 업데이트는 광고사진 평가의 재현 안정성, 검증 가능성을 약화 (Hartwig et al., 2025; Walkowiak & Potts, 2024)

생성형 AI이미지

- 상용 모델 산출물은 인간 전문가 결과물에 근접한 수준에 도달 (Hartmann et al., 2025)
- 광고 이미지 제작의 시안 제작, 콘셉트 도출, 배경 합성 등 초기 제작 단계에 활용 (장유진·이종근, 2023; 김활빈·박민아, 2024)
- 광고사진으로 활용은 제품 재현성, 표현 형식 적합성, 메시지 전달력의 검토 필요 (심현준·최종인, 2021; Hartwig et al., 2025)
- 광고사진가는 이미지 설계자, 프롬프트 기획자 이자 생성 결과물의 큐레이터 (심은혜·현혜연, 2023; 정현목, 2024)

전통 광고사진 제작 → 제작 방식의 확장 → **생성형 AI 기반 제작**

AI로제작된광고이미지

연구 배경

생성 이미지 평가

- LPIPS, CLIPScore, FID, IS 등 범용 지표를 중심으로 발전
  (Zhang et al., 2018; Hessel et al., 2021; Salimans et al., 2016; Heusel et al., 2017)
- 지각적 유사도, 의미 정합성, 이미지 집합의 분포적 품질을 측정
  (Zhang et al., 2018; Hessel et al., 2021; Heusel et al., 2017)
- 도메인에 맞는 측정을 하기 위해서는 특화된 준거를 자동 평가 지표에 반영할 필요
  (Hartwig et al., 2025)

광고사진 평가

- 광고사진은 설득 커뮤니케이션으로 단일 기준이나 점수로 설명될 수 없음
  (Ye & Kovashka, 2019)
- 광고 이미지는 창의성, 의미 전달력, 제품 설명력, 시각적 완성도 등 다차원 평가가 필요
  (MacKenzie et al., 1986)
- 표현 형식별 자료 구성, 전문가 판단 수집, 자동 지표 개발, 독립 테스트셋 적용 절차가 필요
  (Paullada et al., 2021; Raji et al., 2021; Zhang et al., 2018)

생성 이미지 평가 범용 지표

이미지 간 지각 차이

텍스트-이미지 의미 정합성

실제 이미지와 생성 이미지 분포 차이

생성 이미지 품질

…

광고사진 평가는 광고사진 전문가의 판단 기준이 중요

- 광고사진 표현 형식
- 다차원판단
- 생성이미지의 상대판단
- 표현형식별 품질 차이

연구문제

광고사진의 표현 형식에 기반하여 생성형 AI 광고 이미지를 자동 평가할 수 있는가?

연구문제 1 광고사진 전문가의 판단은 자동 평가 지표의 학습 신호로 어떻게 구성될 수 있는가?

1-1 전문가의 쌍대 비교 판단은 지각 차이 기반 변별 지표의 학습 신호로 어떻게 구성될 수 있는가?

1-2 전문가의 다차원 품질 평정값은 품질 예측 보조 지표의 학습 신호로 어떻게 구성될 수 있는가?

연구문제 2 개발된 도메인 특화 자동 평가 지표는 광고사진 전문가의 판단과 어떤 대응 관계를 보이는가?

2-1 지각 차이 기반 변별 지표는 전문가 판단과 어떤 대응 관계를 보이며, 범용 지각 지표와 비교할 때 어떤 차이를 보이는가?

2-2 다차원 품질 예측 보조 지표는 전문가 평정값과 어떤 대응 관계를 보이며, 입력 특징은 예측 성능에 어떻게 기여하는가?

연구문제 3 개발된 도메인 특화 자동 평가 지표는 독립 테스트셋에서 어떤 적용 가능성과 해석 한계를 보이는가?

3-1 독립 테스트셋에서 두 지표는 두 상용 생성 모델의 차이를 어떤 방향과 크기로 산출하는가?

3-2 두 지표의 결과는 광고사진 표현 형식에 따라 어떻게 달라지는가?

3-3 두 지표의 비교 세트별 판정은 전문가 다수결 판단과 어느 정도 일치하거나 불일치하는가?

연구 목적

**광고사진 표현 형식에 기반한 초기 벤치마크 데이터셋과 자동 평가 지표를 개발
적용 가능성과 해석 한계를 실증적으로 검토**

광고사진 전문가의 판단을 학습 가능한 평가 신호로 구성하고,
자동 지표가 생성 이미지 비교와 표현 형식별 모델 차이 분석에서 보조적 정량 기준으로 활용될 수 있는 조건을 확인

광고사진 표현 형식 기반 데이터셋 부족	▶	**광고사진 표현 형식 기반 초기 벤치마크 데이터셋 AD-ImageBench 구축**
광고사진 전문가 판단 기반 자동 지표 부족	▶	**전문가 2AFC 쌍대 비교 판단을 학습 신호로 활용한 AD-LPIPS 개발 다차원 품질 평정값 예측 보조 지표 AD-Score 개발**
표현 형식별 모델 성능 해석 부족	▶	**독립 테스트셋으로 적용 가능성과 해석을 검토**

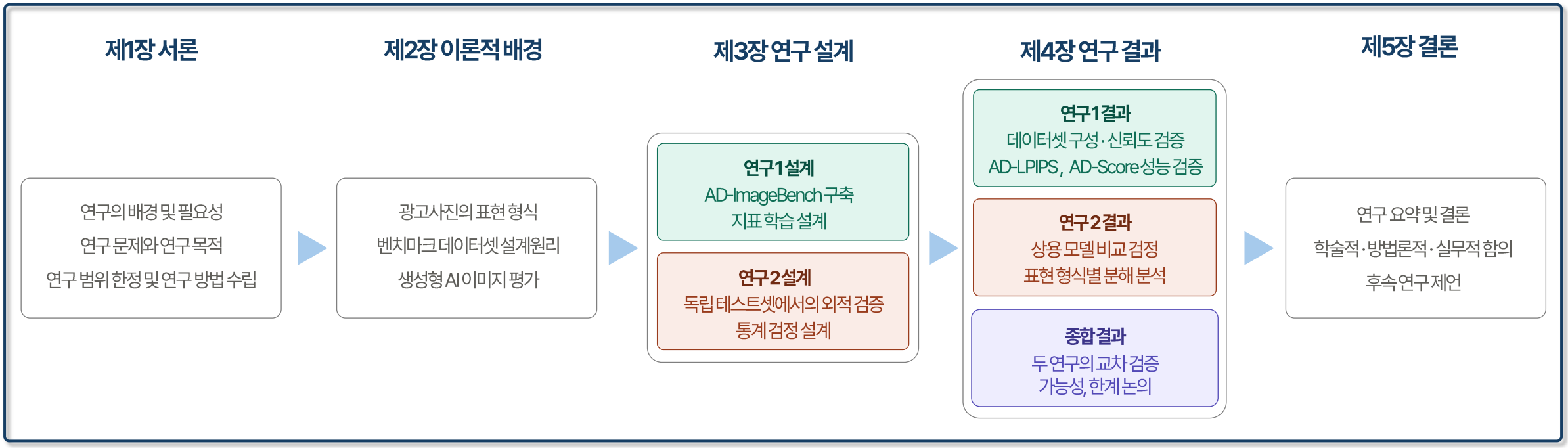
연구 범위와 방법

형식적 범위

- 실제 집행된 광고사진 (1,000장)과 생성 AI 이미지
- 국제 광고사진 공모전 수상작 (180장)과 생성 AI 이미지
- 미드저니 V7, 나노 바나나 프로
- 광고사진 전문가 평정

내용적 범위

- 광고사진 표현 형식 분류
- 2AFC 쌍대 비교 평가
- 다차원 품질 평정
- 자동 평가 지표 학습 및 독립 테스트셋 적용



광고사진의 표현과 평가

- 광고사진은 사진의 재현 능력과 광고의 설득적 기능이 결합된 시각 산물 (Wells, 2004)
- 광고사진은 제품·서비스 제시에 그치지 않고 소비자 주목, 브랜드 메시지, 정서적 분위기를 시각 관습에 따라 구성 (Scott, 1994; McQuarrie & Mick, 1996)
- 광고 이미지는 해석 가능한 구조적 단위와 관습에 기반한 유형으로 분류 가능 (Ye & Kovashka, 2019)

광고사진의 표현 형식은 시각적 관습을 표현하는 커뮤니케이션 전략이기에 (Scott, 1994; McQuarrie & Mick, 1996)
같은 광고 이미지라도 어떤 표현 형식에 속하는 가에 따라 품질 평가 기준이 달라질 수 있다.

광고사진 표현 분류 연구

1989 — **김문진**
7개 표현 형식 체계화
→ 본 연구의 모태

1998-2006 — **김우형·유영복·김희수**
표현 기법·구성 요소 축으로 확장

2014 — **김서윤 외**
8개 형식

2021 — **심현준·최종인**
6가지 표현 형식
(AFF/DAY/DRM/FUL/PRD/SEN)
→ 본 연구의 분류 체계

6가지 표현 형식 (심현준·최종인, 2021)

인물·관계중심
정서적유대와 일상맥락

정감표현형 (AFF)
- 감정, 관계, 정서를 중심으로 메시지를 전달하는 표현 형식
- 인간 감정 중심, 관계 강조, 서사적 분위기

일상표현형 (DAY)
- 제품이 일상상황 속에서 자연스럽게 사용되는 모습을 표현하는 형식
- 현실적 상황, 자연스러운 연출, 생활 맥락 강조

시각적강도중심
주목과 인상을 유도하는 형식

극적표현형 (DRM)
- 비현실적, 과장된, 상징적 연출을 통해 메시지를 전달하는 표현 형식
- 비현실성, 과장/충격, 상징적 이미지

감각표현형 (SEN)
- 색채, 질감, 형태 등 감각적 요소를 강조하는 표현 형식
- 질감 강조, 색채 중심, 형태/패턴 중심

제품중심
제품을 시각의 중심에 두는 형식

제품표현형 (PRD)
- 제품을 중심으로 정보 전달 및 기능 강조를 목적으로 하는 표현 형식
- 제품이 화면 중심, 정보 전달 중심

충족표현형 (FUL)
- 문제 해결 또는 욕구 충족의 결과를 강조하는 표현 형식
- Before/After 구조, 결과 강조, 기능적 효과 표현

광고사진의 표현과 평가

- 생성형 AI는 자연어 프롬프트만으로 사실적 고해상도 이미지 생성 가능 (Ramesh et al., 2022; Rombach et al., 2022; 윤병권, 2024)
- 시안·콘셉트·톤앤매너·배경 합성 등 제작 초기 단계에서 광범위 활용 (장유진·이종근, 2023; 김활빈·박민아, 2024)
- 모델·제품·핵심 피사체는 실제 촬영, 배경·보조 연출은 AI로 보완하는 혼합 방식이 확산 (장유진·이종근, 2023)

광고사진 제작 방식의 변화

전통적 촬영 ~1990s

기획·세트·조명·연출·물리적 촬영·후반 리터칭의 순차적 공정
광고사진가는 카메라를 매개로 현실을 포착하는 창작자

디지털 전환 1990s–2010s

디지털 카메라와 후반 작업 소프트웨어의 보편화
합성·리터칭·색 보정의 유연성 확장
촬영 결과의 보완·강화에 한정

생성형 AI 도입 2020s–

DALL·E 2, Midjourney, Stable Diffusion
자연어 프롬프트만으로 촬영 이전 단계에서
이미지를 직접 산출 (Ramesh et al., 2022; Rombach et al., 2022)

광고사진가 역할의 확장

촬영자 → **프롬프트 설계자 · 이미지 큐레이터** →

생성 이미지 판단 확장

- 표현 형식 적합성
- 재현 안정성
- 브랜드 톤 일관성
- 후속 제작·집행 위험 관리

전문성이 사라지는 것이 아니라 **촬영 현장의 물리적 통제에서 이미지 선택·조율로 이동**
(심은혜·현혜연, 2023; 정현목, 2024)

광고사진의 표현과 평가

- 광고사진은 미적 산물·설득적 도구·정보 매체·브랜드 태도 형성 매개체의 결합 (MacKenzie et al., 1986)
- 사회·문화 관습 속에서 해석되는 시각적 텍스트로 규정 (Scott, 1994)
- 광고 이미지는 지각적 유사도나 단일 점수로 환원되기 어렵기에, 기준 이미지 기반의 상대적 비교와 단일 이미지의 다차원 품질 평정이 필요

광고 이미지 평가의 두 영역과 두 가지 도메인 특화 접근

기준 이미지 기반 비교

원본 광고사진과 생성 이미지 사이의 지각 차이 변별

원본 광고사진 ↔ 생성 이미지

도메인 특화 변별 점수의 필요

전문가의 상대적 선호 판단에 맞춰 재보정된 지각 차이 척도

단일 이미지 다차원 품질 평정

개별 이미지에 대한 4개 품질 차원의 전문가 평정

창의성 | 의미 전달력 | 제품 설명력 | 시각적 완성도

도메인 특화 품질 예측 점수의 필요

4개 평가 차원의 전문가 평정값을 예측하는 회귀 척도

벤치마크 데이터셋과 전문가 판단

- 벤치마크는 단순 이미지 모음이 아닌, 동일 조건에서 모델별 결과를 비교·해석할 수 있는 공통 평가 장치
- 후속 연구는 비교와 축적을 수행할 수 있는 연구 기반으로 기능 (Deng et al., 2009; Lin et al., 2014; Wang et al., 2019b)
- 1~2%p의 성능 향상이나 소수점 단위의 지표 개선은 후속 연구의 방향을 조정하는 공학적 의미 (Recht et al., 2019; Beyer et al., 2020)

벤치마크 설계의 5가지 원칙 (Raji et al., 2021)

대표성	실제 자료 분포와 적용 맥락의 반영
포괄성	도메인 주요 변이와 하위 범주의 누락 없는 포함
균형성	범주 간 표본 수의 분석 가능한 구성
재현 가능성	자료·라벨·평가 절차의 독립적 검토
접근성	자료·메타데이터·평가 절차의 합리적 접근

주요 벤치마크의 계보와 광고 도메인의 등장 (Deng et al., 2009; Lin et al., 2014; Zhang et al., 2018; Li et al., 2024)

벤치마크	주요 과제	규모	연도
MNIST	손글씨 숫자 분류	학습 60,000장/ 테스트 10,000장	1998
ImageNet	대규모 이미지 분류	약 120만 학습 이미지, 1,000개 범주	2009
COCO	객체 검출·분할	약 33만 이미지, 80개 객체 범주	2014
BAPPS	지각 차이 변별 (2AFC)	패치 쌍	2018
AGIQA-3K	AI 생성 이미지 품질	약 3,000장	2023
AIGI-VC	AI 생성 광고 이미지	약 2,500장	2025

일반 이미지 인식 지각 품질 평가 광고 도메인

벤치마크 데이터셋과 전문가 판단

- 학습 기반 지각 지표와 생성형 AI 평가 연구에서 인간 판단은 단순 사후 검증을 넘어 모델 학습 신호로 직접 활용 (Zhang et al., 2018)
- 일반 사용자 선호로는 광고사진 품질 측정 부족 → 제품 재현·조명·질감·구도·표현 형식 적합성·브랜드 메시지 전달은 전문가 판단 영역
- 인간이 절대 점수를 부여하는 것보다 두 자극의 상대적 차이를 판단하는 방식이 더 안정적 (Thurstone, 1927; Bradley & Terry, 1952)

광고 도메인 벤치마크의 세 구성 요소

비교 조건의 고정 → 기준이미지 구조

광고 의도·제품 제시 방식·표현 형식의 통제. 정답 이미지가 아닌 참조 조건

도메인 변이의 포함 → 표현형식별 자료 구성

AFF·DAY·DRM·SEN·PRD·FUL의 균형. 형식별 조건부 해석의 가능성

실무적 평가 기준 → 전문가 판단 자료

제품 재현·표현 형식 적합성·브랜드 메시지 등 전문가 판단의 학습 신호화

전문가 판단 자료의 두 유형과 학습 신호로서의 기능

쌍대 비교 자료 (2AFC)

품질 평정 자료 (단일 이미지)

생성형 AI 이미지 평가 지표

• 각 유형은 모두 일반 이미지 평가의 강점을 갖지만, 광고사진의 표현 형식·제품 설명력·브랜드 메시지·설득 효과를 직접 반영하도록 학습된 지표는 부재
→ 범용 지표를 광고 도메인 판단에 맞춰 조정하는 구조가 필요

지각 기반 지표

Perceptual

원본과 생성 이미지 간 지각 차이

MSE · PSNR
픽셀 단위 차이
(Wang & Bovik, 2002)

SSIM · MS-SSIM · VIF
구조 정보 비교
(Wang et al., 2003, 2004; Sheikh & Bovik, 2006)

LPIPS
BAPPS 데이터셋 기반 인간 지각 학습
(Zhang et al., 2018)

도메인 한계
LPIPS는 자연 이미지 도메인에 한정된 학습으로, 광고사진의 표현 형식·제품 재현·브랜드 메시지를 직접 반영하지 못함

분포 기반 지표

Distribution-based

이미지 집합과 실제 이미지 집합의 통계적 분포 비교

IS (Inception Score)
범주 예측 명확성과 다양성
(Salimans et al., 2016)

FID
두 이미지 집합의 특징 분포 비교
(Heusel et al., 2017)

도메인 한계
비교 세트 단위 상대 판단을 다루지 못하며, 표현 형식별 평가 기준 차이를 반영하기 어려움

비전-언어 의미 기반 지표

Vision-Language

이미지·텍스트·인간 선호 사이의 의미 적합성

CLIPScore
이미지-텍스트 임베딩 유사도
(Hessel et al., 2021)

TIFA · VQAScore
VQA 기반 구성적 의미 평가
(Hu et al., 2023; Lin et al., 2024)

PickScore · ImageReward
인간 선호 학습 기반
(Kirstain et al., 2023; Xu et al., 2023)

도메인 한계
광고사진의 표현 형식 관습·정서적 분위기·브랜드 톤은 일반 의미 공간과 일반 선호 데이터로는 충분히 반영되지 않음

생성형 AI 이미지 평가 지표

• 범용 LPIPS는 모든 조건의 인간 판단을 만족시키지 못하며, 식별된 불일치 조건에 맞춘 재보정은 학술적으로 확립된 방법론
→ 광고사진 전문가의 도메인 판단 기반 재보정의 근거

LPIPS 계열 재보정 연구

R-LPIPS (Ghazanfari et al., 2023)

식별한계
적대적 섭동에 대한 LPIPS의 취약성

조정방식
적대적 학습 기반 견고성 강화

Shift-tolerant (Ghildyal & Liu, 2022)

식별한계
미세한 공간 이동에 대한 과잉 민감성

조정방식
이동 내성 가중치 재학습

CLPIPS (Trinh et al., 2026)

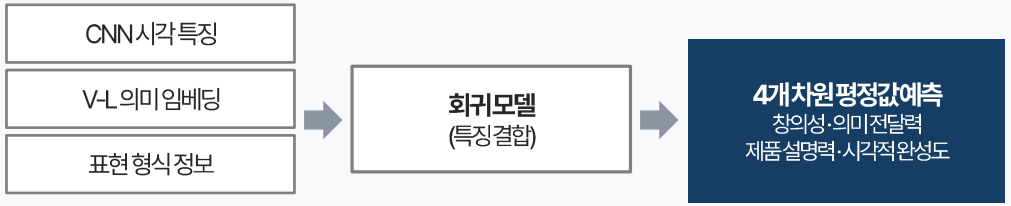
식별한계
개인·맥락별 유사성 판단의 차이

조정방식
순위 판단 기반 개인화 조정

기술적 구조

지각 차이 기반 재보정

다차원 품질 예측



연구 개요

연구 1·연구 2의 흐름

연구 1

지표 개발·내부 성능 검증

자료	**AD-ImageBench** (원본 광고사진 1,000장 + 생성 이미지 2,000장 + 표현 형식 라벨)
전문가	**2AFC 쌍대 비교** + **Q1 라벨 타당도** + **Q2–Q5 다차원 품질 평정**
산출물	**AD-LPIPS** (2AFC 학습) + **AD-Score** (Q2~Q5 학습)

연구 2

독립 테스트셋 적용·외부 검증

자료	**광고사진 공모전 수상작** (학습 미참여) (수상작 180장 + 생성 이미지 360장)
적용지표	**AD-LPIPS·AD-Score** + 범용 LPIPS·CLIPScore
검증	**광고사진 전문가 3인의 2AFC 다수결** (연구 1 미참여)

연구 진행의 4단계 절차

연구1

원본 수집·분류

광고사진 1,000장 수집
6가지 표현 형식 분류
코더 간 신뢰도 검증

연구1

전문가 평가

Q1 라벨 타당도 점검
2AFC 쌍대 비교
Q2~Q5 다차원 품질 평정

연구1

지표 학습

AD-LPIPS 학습 (2AFC 신호)
AD-Score 학습 (Q2~Q5 신호)
내부 성능 측정

연구2

독립 테스트셋 적용

180장에 4개 지표 적용
전문가 3인 2AFC 비교
표현 형식별 하위 분석

연구 1 설계

AD-ImageBench 자료의 구성

원본자료 TVCF 광고 포털 (2023.1–2025.12)에서 1,704장 → **1,000장 선정**

표현형식분포 PRD 250 · AFF/DAY/DRM/FUL/SEN 각 150 = 1,000장 (코더 2인 + 조정자 1인)

비교세트 원본 + **Midjourney V7** + **Nano Banana Pro** = 3,000장 (역설계 프롬프트 1,000개)

학습 720	80	테스트 200

표현형식별 비율 유지 (72% / 8% / 20%) · random seed 42 고정

전문가 평가 설계

평가자 광고사진 스튜디오 대표 · 광고제작회사 대표 · 사진학과 교수 등 (경력 7–35년)

2AFC 쌍대 비교
1,000세트 × 7인 = 7,000건
→ **AD-LPIPS 학습 신호**

Q1 · Q2–Q5 평정 (5점)
1,000장 × 5문항 × 7인 = 35,000건
→ **AD-Score 학습 신호 (Q2–Q5)**

AD-LPIPS 파인튜닝 설계

구조 사전 학습 **VGG16** (5층 conv) 고정 → **lin층 채널별 가중치 학습**

(원본, MJ, Nano) **비교세트** → VGG16 특징 차이 + lin 층 결합 → 변별 점수 차이 → **BCE loss**

학습신호 전문가 7인의 **선택비율 $y_i \in [0,1]$** (다수결만이 아닌 합의 강도까지 입력)

초매개변수 Adam · lr 0.001 · batch 8 · max epoch 50 · 5-epoch 조기 종료

AD-Score 학습 설계

입력특징 세 종류의 특징을 단일 입력 벡터로 결합

VGG16 시각	CLIP 의미	표현형식	
1,472	512	6	= 1,990차원

회귀모델 Multi-output **Ridge 회귀** (λ=1.0) → Q2–Q5 4차원 동시 예측

학습자료 원본 광고사진 1,000장 + 전문가 7인 평균 평정값 (Q2–Q5)

연구 1 설계

정감 표현형 (AFF)

Midjournet V7 | Original | Nano Banana Pro

극적 표현형 (DRM)

Midjournet V7 | Original | Nano Banana Pro

제품 표현형 (PRD)

Midjournet V7 | Original | Nano Banana Pro

일상 표현형 (DAY)

충족 표현형 (FUL)

감각 표현형 (SEN)

원본 광고사진과 AI 생성 이미지

제3장 연구 설계

연구 2 설계

독립 테스트셋의 구성

자료출처 한국광고사진가협회(KAPA) 주최 Global Vision Awards (2017–2025)

출품작 1,640점 (전체 모집단) → 수상작 334장 (본상 + 입선) → **180장 선정** (6 형식 × 30 균등)

균등구성이유 특정 표현 형식의 표본 비중이 모델 비교 결과를 좌우하지 않도록 통제

역설계 프롬프트 · AI 이미지 생성

절차일관성 연구 1의 **5요소 표준화 구조**와 작성 원칙을 그대로 유지

원본 180장 + **Midjourney V7** 180장 + **Nano Banana Pro** 180장

= 540장의 이미지가 **180개 비교 세트**를 형성

자동 지표 산출과 모델 간 비교

AD-LPIPS	범용 LPIPS	CLIPScore	AD-Score
낮을수록	낮을수록	높을수록	높을수록

주검정 대응표본 t-검정 + Wilcoxon 부호순위 검정 (보조)

효과크기 Cohen d_z (95% bootstrap CI, seed 42)

다중검정보정 4개 지표 동시검정 → **Bonferroni $\alpha = 0.0125$**

비교세트별판정 정확 이항 검정 · McNemar (지표 간) · 형식별 하위 분석 (탐색적)

전문가 다수결 판단 비교

신규전문가 3인 광고사진 스튜디오 대표 (경력 17~25년) · **연구 1 평가미참여**

전문가 3인 2AFC → 3:0 또는 2:1 다수결 → **자동지표 판정과 대조**

일치도지표 단순 일치율 · **Cohen's κ** · **균형 정확도** · 범주별 재현율 · **다수 범주 기준선**

전문가신뢰도 Fleiss' κ (주) + Krippendorff's α + 쌍별 Cohen's κ

보조분석 표현 형식별 · 합의 강도별 (3:0 vs 2:1) · 다수결 방향별

연구 1 결과

표현 형식 분류신뢰도와 평가자 간 신뢰도

코더 일치도와 전문가 라벨 적합성 평정 모두 학습 신호로 활용할 수준의 신뢰도에 도달

코더 2인의 분류 일관성

단순 일치율	Cohen's κ	Cohen's κ
92.80%	**0.913**	**0.919**
928/1,000건	전체 신뢰도	6형식 한정(n=995)

전문가 7인의 라벨 적합성 평정 (Q1)

Q1 평균	4점 이상 비율	ICC(A,k)
4.079	**64.2%**	**0.916**
5점 척도	이미지 기준	7인 평균 신뢰도

평가자 간 신뢰도 : 2AFC · Q1 · Q2–Q5

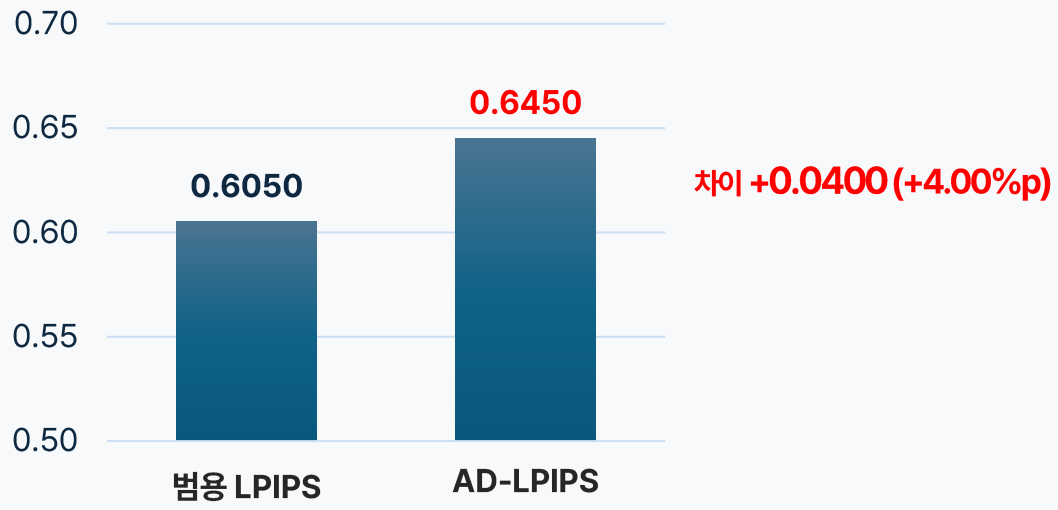
평가 유형	주 지표	보조 지표	해석	활용
2AFC 평가	Fleiss κ = **0.641**	Krippendorff α = **0.641**	상당한 일치 (Landis & Koch, 1977)	AD-LPIPS 학습 신호
Q1 라벨 타당도	ICC(A,k) = **0.916**	Krippendorff α = **0.651**	매우 우수 (7인 평균 평정의 안정성)	표현 형식 라벨 적합성 점검
Q2–Q5 다차원 품질 (통합)	ICC(A,k) = **0.969**	Krippendorff α = **0.819**	매우 높음 (4문항 통합)	AD-Score 학습 신호

연구 1 결과

AD-LPIPS의 학습 결과와 성능 비교

+4.00%p 의 정확도 개선 관측 단, McNemar 검정에서 통계적 우월성은 미확정

테스트 세트 200개 : 2AFC 정확도



차이 +0.0400 (+4.00%p)

McNemar 검정 : 통계적 유의성

	범용 LPIPS 일치	범용 LPIPS 불일치
AD-LPIPS 일치	103	18
AD-LPIPS 불일치	26	53

$\chi^2(1) = 1.114 \cdot p = .291$

Spearman ρ : 0.144 → 0.180

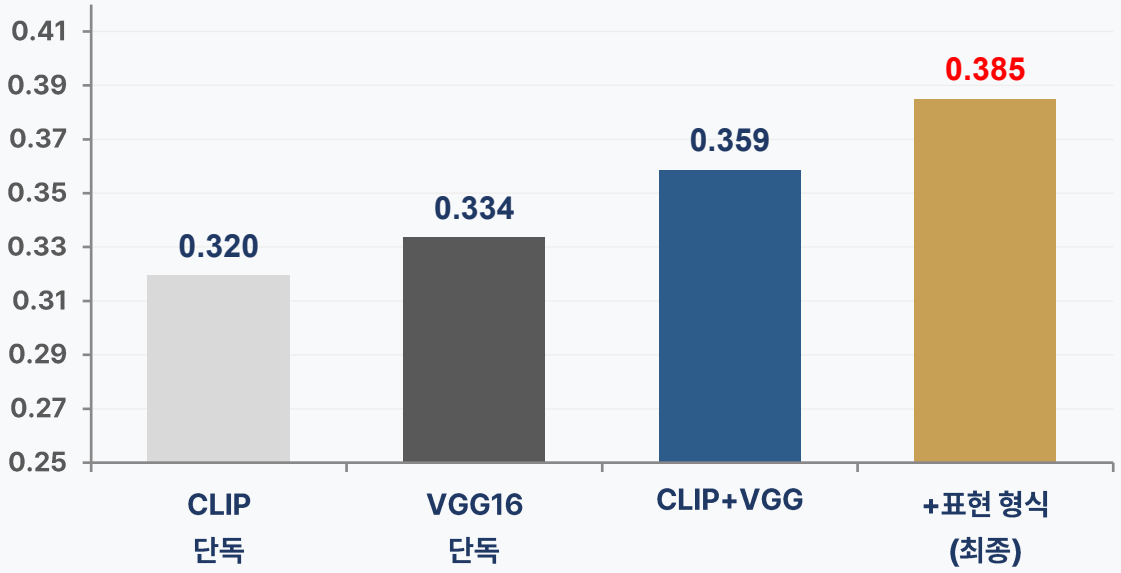
lin층가중치평균 cos(θ) = 0.549 (BAPPS대비 변화)

연구 1 결과

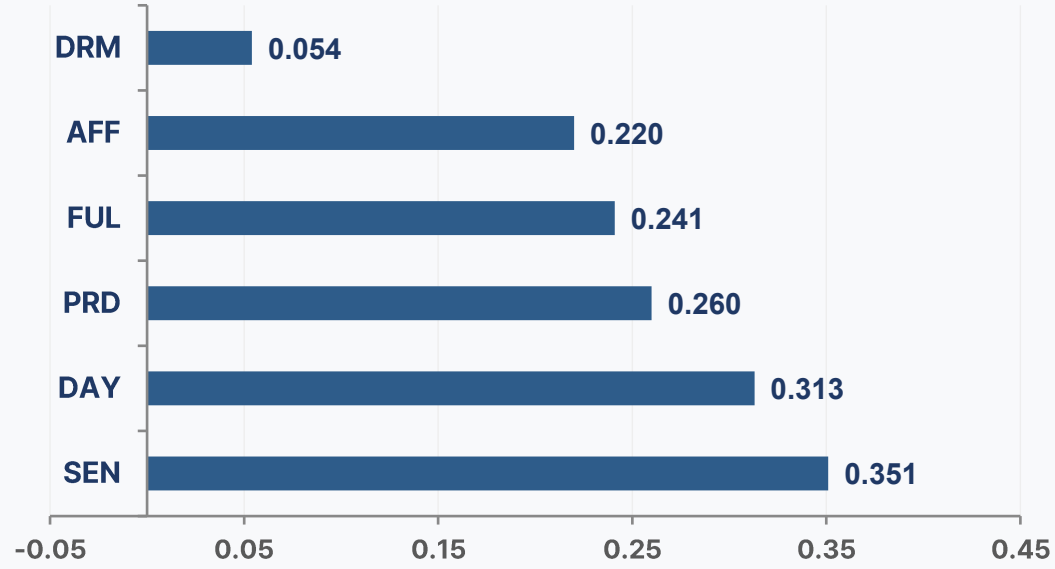
AD-Score 예측 성능과 표현 형식 정보의 기여

5-fold OOF 교차검증 · Pearson r = 0.385 (95% CI [0.333, 0.436]) · 중간 수준 효과 크기

입력 특징 절제 실험 : 시각·의미·표현 형식 정보의 단계적 기여



표현 형식별 종합 점수 Pearson r

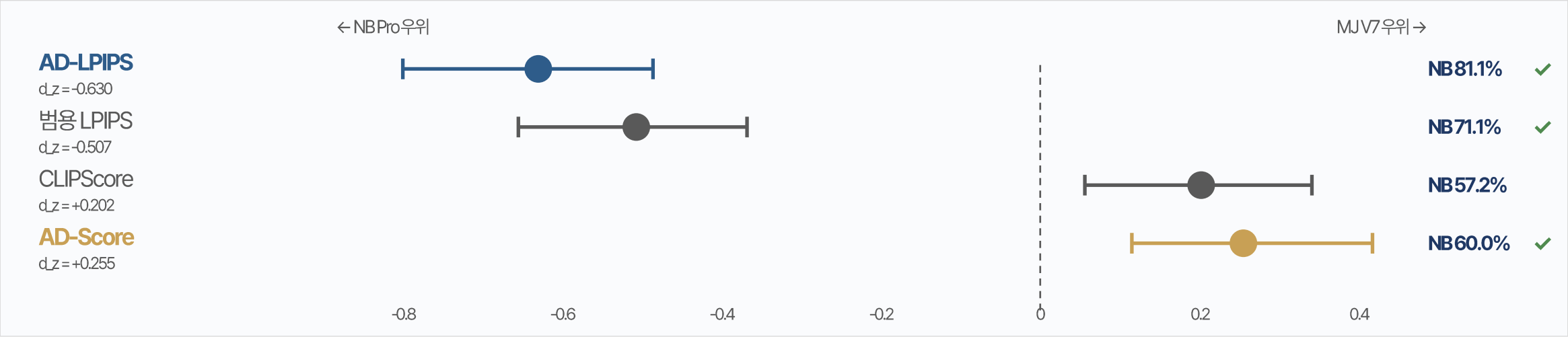


연구 2 결과

독립 테스트셋의 네 가지 자동 지표 통계 비교

4개 지표 모두 NB Pro 우위 방향 · Bonferroni 보정 후에도 t-검정 유의 단, 효과 크기와 우세 비율은 지표마다 상이

Cohen d_z · 효과 크기 (95% CI)



네 가지 자동 지표 우세 모델 판정의 수렴/불일치 양상 (180 비교 세트)

4개 지표 모두 같은 우세 모델 = 32.8% → 약 2/3 비교 세트에서 지표 간 불일치—단일 지표가 광고 평가의 모든 차원을 대표하지 않음

연구 2 결과

표현 형식별 하위집합 분석_조건부 비교

"어떤 모델이 더 우수한가" 가 아니라 "어떤 표현 형식에서, 어떤 평가 관점에 따라, 어떤 모델이 우세한가"

표현형식	AD-LPIPS	범용 LPIPS	CLIPScore	AD-Score
AFF 정감 표현형	**-0.928** ✓NB 86.7%	**-0.700** ✓NB 80.0%	**+0.273** 56.7%	**+0.336** 63.3%
DAY 일상 표현형	**-0.513** ✓NB 76.7%	**-0.426** 73.3%	**-0.091** 40.0%	**+0.514** ✓NB 66.7%
DRM 극적 표현형	**-0.619** ✓NB 86.7%	**-0.606** ✓NB 73.3%	**+0.125** 53.3%	**+0.117** 56.7%
FUL 충족 표현형	**-0.682** ✓NB 76.7%	**-0.519** ✓NB 70.0%	**+0.156** 60.0%	**-0.252** 36.7%
PRD 제품 표현형	**-0.858** ✓NB 90.0%	**-0.767** ✓NB 80.0%	**+0.709** ✓NB 80.0%	**+0.529** ✓NB 73.3%
SEN 감각 표현형	**-0.304** 70.0%	**-0.050** 50.0%	**+0.045** 53.3%	**+0.335** 63.3%

PRD : 4지표 수렴

제품 표현형은 4지표 모두 Bonferroni 후에도 NB 우위.
제품 형태·물성·프롬프트 의미·품질 평정이 같은 방향으로 수렴

SEN : 4지표 모두 미달

감각 표현형은 어느 지표도 Bonferroni 후 유의 X.
색감·질감·분위기의 미세 속성이 모델 차이의 안정적 측정 어려움

FUL : 지표 간 방향 불일치

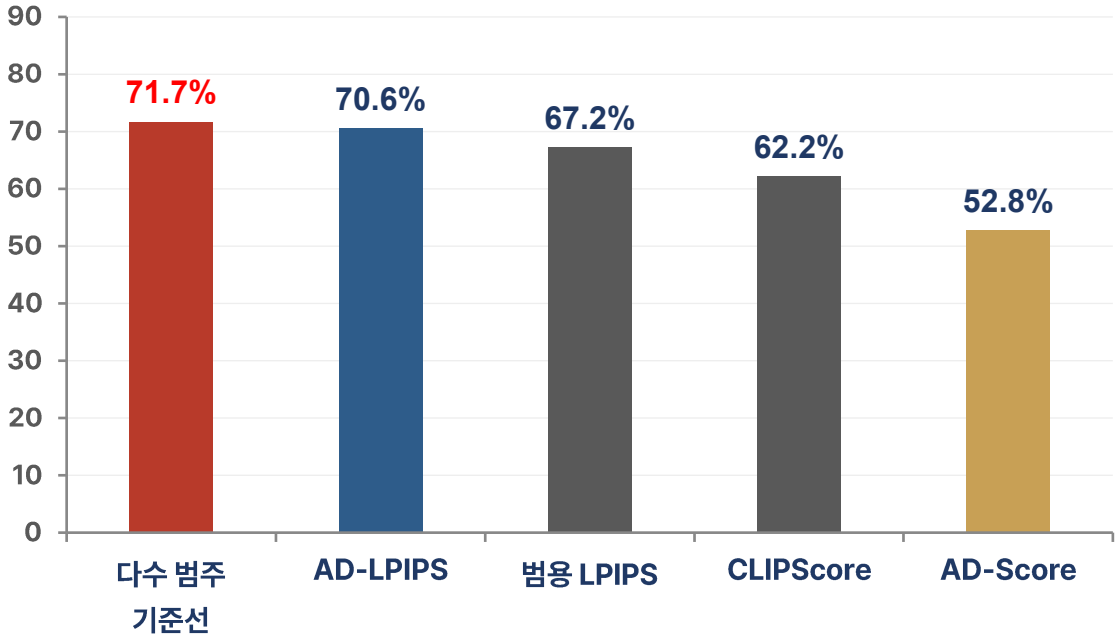
지각 계열은 NB 우위, AD-Score는 MJ 우위 방향.
같은 형식 안에서 지각적 재현과 다차원 품질 예측이 다른 결과

연구 2 결과

자동 지표와 전문가 다수결 판단의 일치도

AD-LPIPS의 단순 일치율 70.6% 는 다수 범주 기준선 71.7%에 미달 · 보정 지표 종합 시 명확한 우월성 미확정

단순 일치율 vs 다수 범주 기준선



보정 지표 종합 비교

지표	균형 정확도	Cohen κ	MCC
기준선	50.0%	0.000	—
AD-LPIPS	58.7%	0.193	0.200
범용 LPIPS	59.9%	0.197	0.197
CLIPScore	61.2%	0.194	0.204
AD-Score	48.1%	−0.034	−0.035

종합 결과

연구 1 · 연구 2 결과의 교차 확인

AD-LPIPS
지각 차이 기반 변별 점수 (2AFC 학습)

연구 1 (테스트 200)
범용 LPIPS 0.605 → **AD-LPIPS 0.645** (+4.00%p)
· McNemar p=.291 통계적 우월성 미확정

연구 2 (독립 테스트셋 180)
Cohen d_z = **−0.630** (범용 −0.507보다 큰 효과)
· 단순 일치율 70.6% < 다수 범주 기준선 71.7%

AD-Score
Q2~Q5 다차원 품질 회귀 예측 (Q2~Q5 학습)

연구 1 (5-fold OOF)
Pearson r = **0.385** · Spearman ρ = 0.374 · τ = 0.253
· 중간 수준 효과 크기 (Cohen 1988) · 보조적 선별 가능성

연구 2 (독립 테스트셋 180)
Cohen d_z = +0.255 (모델 비교 유의) · **단순 일치율 52.8%**
· 전문가 다수결 일치 **Cohen κ = −0.034 (무작위 수준)**

두 도구의 역할 분화 : 학습 신호 · 활용 맥락 · 평가 단위

구분	AD-LPIPS	AD-Score
학습 신호	전문가 2AFC 쌍대 비교 판단	전문가 Q2~Q5 다차원 품질 평정값
활용 맥락	두 생성 이미지의 상대적 변별	단일 이미지의 다차원 품질 회귀 예측
평가 단위	비교 세트 (원본 + 생성 이미지 쌍)	단일 이미지
산출 방향	거리 (낮을수록 원본과 가까움)	점수 (높을수록 품질 평정 예측 높음)

종합 결과

표현 형식의 두 실증적 역할

AD-Score 입력 특징의 추가 설명력

CLIP + VGG16
r = 0.3586

→

CLIP + VGG16+ 표현형식 (최종)
r = 0.3850

표현 형식별 분석 단위로서의 차이 발견

PRD	SEN	FUL
수렴	차이 부재	방향불일치
4지표 모두 NB 우위 일관	어느 지표도 Bonf. 후 유의 X	지각 NB ≠ Q2~Q5 MJ

자동 지표의 적용 범위

도구	적용가능 영역	확인된 한계	권장 활용
AD-LPIPS	원본·기준 이미지 있는 비교 세트의 상대적 변별	McNemar 미유의 · 다수 범주 기준선 미달	**비교 보조 도구**
AD-Score	원본 광고사진 Q2~Q5 평정 예측 (중간 수준 r=0.385)	AI 생성 이미지 비교 상황에서 평가 맥락 차이로 제한적	**초기 선별 도구**

연구의 요약 및 결론

연구 1은 **AD-ImageBench와 두 도구의 학습**, 연구 2는 **독립 테스트셋 적용과 전문가 다수결 검토**를 수행하였으며, 세 연구문제의 결과는 다음과 같이 정리된다.

[연구문제1]

전문가 판단의 학습 신호화 원리

광고사진 전문가의 판단을 자동 평가 지표의 학습 신호로 구성하기 위한 원리와 조건

두 평가 영역의 분리 설계

· **상대적 비교 (2AFC)** → AD-LPIPS 학습 신호
· **Q2~Q5 다차원 평정** → AD-Score 학습 신호

AD-LPIPS의 위상

수학적 거리 함수가 아니라, 채널별 지각 차이를 전문가의 상대적 판단 기준 쪽으로 재가중한 **변별 점수**

AD-Score의 위상

Q2~Q5 평정값을 회귀 예측하는 **단일 이미지 품질 예측 지표** (Q1은 라벨 타당도 자료로 분리)

[연구문제2]

두 도구와 전문가 판단의 대응

설계된 두 자동 평가 지표가 광고사진 전문가의 판단과 어떤 대응 관계를 보이는가

AD-LPIPS의 대응

범용 LPIPS 대비 +4.00%p 일치 방향 변화
통계적 우월성은 미확정

AD-Score의 대응

Q2~Q5 평균과 r=0.385의 일정한 대응
표현 형식 정보가 **맥락 정보로 추가 기여** (+0.026)

두 도구의 관계

직렬적 평가 체계가 아닌 **병렬적 보조 도구**
도메인 특화 효과는 가중치 자체가 아니라 **계산 위치와 평가 과제의 결합**에 따라 달라진다

[연구문제3]

독립 테스트셋 적용 조건과 한계

구축된 자동 평가 도구가 독립 테스트셋에서 어떤 조건으로 적용될 수 있으며, 그 한계는 무엇을 의미하는가

전체 평균 · 조건부 차이

4 지표 모두 **NB Pro 우위 방향**이지만, 표현 형식 단위로는 **PRD 수렴 / SEN 부재 / FUL 방향 불일치**의 조건부 차이

전문가 다수결 비교

단순 일치율은 **다수 범주 기준선 71.7% 미달**
보정 지표 종합 시 명확한 우월성은 미확정

보수적 활용 권고

자동 지표는 **전문가 판단을 대체하기보다 초기 선별·비교 보조 도구**로 제한적 활용

광고사진 표현 형식 기반 자동 평가는 **단일한 품질 판정 체계가 아니라, 표현 형식과 평가 관점을 명시한 조건부 비교 절차**로 보는 것이 적절
자동 지표는 광고 이미지의 **최종 품질 확정 기준이 아니라, 비교 가능한 형태로 정리하고 표현 형식별 차이를 해석하는 보조적 정량 기준**으로 활용가능

연구의 시사점

이론적 함의

표현 형식, 도메인 특화 학습, 광고 품질 평가의 다차원성

표현 형식의 분석 단위 확장

광고사진의 여섯 표현 형식은 분류 라벨을 넘어 자료 구성,
모델 입력 특징, 결과 해석을 연결하는 분석 단위로 활용 가능

도메인 특화 학습의 조건부 효과

재보정 가중치는 계산 위치와 평가 과제의 결합에 따라
효과가 달라지며, 단일한 방식으로 보장되지 않음

광고 품질 평가의 다차원성

지각 차이, 의미 적합성, Q2~Q5 다차원 평정이
동일한 차원이 아닌 결합된 평가 과정임을 실증적으로 확인

방법론적 함의

벤치마크 통합 구조, 두 지표의 분리 설계, 보수적 검토 방식

AD-ImageBench의 통합 구조

원본 광고사진, 생성 이미지, 표현 형식 라벨,
전문가 평가 자료를 하나의 벤치마크에 결합한 초기 설계

두 지표의 분리 설계

2AFC와 Q2~Q5 평정값을 통합 점수로 합치지 않고
병렬 도구로 설계한 점이 평가의 구분을 가능하게 함

보수적 검토 방식

단순 일치율뿐 아니라 다수 범주 기준선과 우연 보정 지표를
함께 보는 절차는 주관적 판단 도메인에서 필요한 검토 방식

실무적 함의

비교 보조 도구로의 활용, 표현 형식별 제작 전략

두 도구의 실무 활용 분화

· AD-LPIPS: 비교 쌍의 상대적 변별
· AD-Score: 검토 순서·강약점 탐색

AD-LPIPS의 적합 맥락

원본 또는 기준 이미지가 있는 재제작·리터칭·스타일 전이
맥락에서 도메인 특화 변별 점수로 기능

표현 형식별 모델 선택 전략

제품 중심 형식은 자동 지표 활용도 높고, 감각 표현형은 제한적
형식·관점에 따른 조건부 모델 선택 권장

연구의 한계 및 제언

한계

자료와 모델의 범위

국내 광고 아카이브(연구 1) + KAPA 공모전(연구 2) 자료 + MJ V7·NB Pro 시점 한정
다양한 광고 장르·문화권·제품군을 포괄하지 못함

평가 자료의 범위

광고사진 전문가의 시각 평가에 한정
소비자 반응·브랜드 태도·구매 의향·문화적 수용성·법률적 적합성 미포함

자동 지표의 적용 범위

AD-LPIPS는 기준 이미지 있는 비교 세트에서만 해석 적합
AD-Score는 외부 적용 시 전문가 다수결과의 대응이 제한적

후속 연구 제언

자료 확대

다양한 광고사진 장르, 제품군, 문화권, 생성 모델을 포함
수집 시점, 모델 버전, 프롬프트 조건을 기록하여 재평가 가능 절차 마련

다층 평가 구조

전문가 평가 + 소비자 평가(메시지 이해도·주목도·브랜드 호감도·구매 의향)
+ 브랜드 실무자·아트디렉터·카피라이터 시점 결합

지표 설계 세분화

· AD-LPIPS: 표현 형식별 재보정 비교
· AD-Score: Q2~Q5 개별 차원 분리 분석

결과 제시 방식

점수 순 나열이 아닌 표현 형식별 강약점, 차원별 예측값, 전문가 판단 불일치 사례를 함께 제시

AD-ImageBench, AD-LPIPS, AD-Score는 **완성된 표준 데이터셋이나 범용 평가 지표가 아니라, 생성 이미지 비교와 표현 형식별 해석을 지원하기 위한 초기 연구 도구**이다.
후속 연구에서 자료·평가자가 확대되고 표현 형식별 적용 조건이 보완되면 **방법론적 기반으로 확장**될 수 있다.

참고문헌

광고사진·표현 형식 · 시각 수사 은유 관련 주요 참고문헌

김문진. (1989). 상표 이미지 선호도와 패션 광고 전략에 관한 연구: 여대생을 중심으로. 이화여자대학교 대학원 석사학위논문.

김석란. (1998). 광고사진의 시각적 Image에 관한 고찰: Fashion catalogue 사진을 중심으로. 숙명여자대학교 대학원 석사학위논문.

심현준, & 최종인. (2021). 광고사진의 표현 형식이 광고효과에 미치는 영향: 국내 가전제품 인쇄광고를 중심으로. 현대사진영상학회논문집, 24(2), 55–85.

심현준. (2022). 광고사진의 표현 방식과 시각적 은유 유형에 따른 광고효과 연구. 상명대학교 일반대학원 박사학위논문.

Scott, L. M. (1994). Images in advertising: The need for a theory of visual rhetoric. Journal of Consumer Research, 21(2), 252–273.

McQuarrie, E. F., & Mick, D. G. (1996). Figures of rhetoric in advertising language. Journal of Consumer Research, 22(4), 424–438.

McQuarrie, E. F., & Mick, D. G. (1999). Visual rhetoric in advertising. Journal of Consumer Research, 26(1), 37–54.

Phillips, B. J., & McQuarrie, E. F. (2004). Beyond visual metaphor: A new typology of visual rhetoric in advertising. Marketing Theory, 4(1–2), 113–136.

White, A., & Smith, B. L. (2001). Assessing advertising creativity using the Creative Product Semantic Scale. Journal of Advertising Research, 41(6), 27–34.

Ye, K., & Kovashka, A. (2019). Interpreting the rhetoric of visual advertisements. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(7), 1645–1658.

데이터셋 · 벤치마크 · 재현성 · 데이터 편향 관련 주요 참고문헌

Donoho, D. (2017). 50 years of data science. Journal of Computational and Graphical Statistics, 26(4), 745–766.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 248–255.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollar, P., & Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. Proceedings of the European Conference on Computer Vision, 740–755.

Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., & Bowman, S. R. (2019). GLUE: A multi-task benchmark and analysis platform for natural language understanding. Proceedings of the International Conference on Learning Representations.

Park, S., Moon, J., Kim, S., Cho, W. I., Han, J., Park, J., Song, C., Kim, J., Song, Y., Oh, T., Lee, J., Oh, J., Lyu, S., Jeong, Y., Lee, I., Seo, S., Lee, D., Kim, H., Lee, M., Jang, S., Do, S., Kim, S., Lim, K., Lee, J., Park, K., Shin, J., Kim, S., Park, L., Oh, A., Ha, J.-W., & Cho, K. (2021). KLUE: Korean Language Understanding Evaluation. Advances in Neural Information Processing Systems: Datasets and Benchmarks Track.

Paullada, A., Raji, I. D., Bender, E. M., Denton, E., & Hanna, A. (2021). Data and its discontents: A survey of dataset development and use in machine learning research. Patterns, 2(11), Article 100336.

Raji, I. D., Bender, E. M., Paullada, A., Denton, E., & Hanna, A. (2021). AI and the everything in the whole wide world benchmark. Advances in Neural Information Processing Systems: Datasets and Benchmarks Track.

Recht, B., Roelofs, R., Schmidt, L., & Shankar, V. (2019). Do ImageNet classifiers generalize to ImageNet? Proceedings of the 36th International Conference on Machine Learning, 5389–5400.

Bouthillier, X., Delaunay, P., Bronzi, M., Trofimov, A., Nichyporuk, B., Szeto, J., et al. (2021). Accounting for variance in machine learning benchmarks. Proceedings of Machine Learning and Systems, 3, 747–769.

Liao, T., Taori, R., Raji, I. D., & Schmidt, L. (2021). Are we learning yet? A meta review of evaluation failures across machine learning. Advances in Neural Information Processing Systems: Datasets and Benchmarks Track.

생성 이미지 평가 지표 · 품질 · 선호 · 이미지 벤치마크 관련 주요 참고문헌

Zhang, R., Isola, P., Efros, A. A., Shechtman, E., & Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 586–595.

Hessel, J., Holtzman, A., Forbes, M., Le Bras, R., & Choi, Y. (2021). CLIPScore: A reference-free evaluation metric for image captioning. Proceedings of EMNLP, 7514–7528.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). GANs trained by a two time-scale update rule converge to a local Nash equilibrium. Advances in Neural Information Processing Systems, 30, 6626–6637.

Hu, Y., Liu, B., Kasai, J., Wang, Y., Ostendorf, M., Krishna, R., & Smith, N. A. (2023). TIFA: Accurate and interpretable text-to-image faithfulness evaluation with question answering. Proceedings of the IEEE/CVF International Conference on Computer Vision, 20406–20417.

Lin, Z., Pathak, D., Li, B., Li, J., Xia, X., Neubig, G., Zhang, P., & Ramanan, D. (2024). Evaluating text-to-visual generation with image-to-text generation. Proceedings of the European Conference on Computer Vision.

Xu, J., Liu, X., Wu, Y., Tong, Y., Li, Q., Ding, M., Tang, J., & Dong, Y. (2023). ImageReward: Learning and evaluating human preferences for text-to-image generation. Advances in Neural Information Processing Systems, 36.

Li, C., Zhang, Z., Wu, H., Sun, W., Min, X., Liu, X., Zhai, G., & Lin, W. (2024). AGIQA-3K: An open database for AI-generated image quality assessment. IEEE Transactions on Circuits and Systems for Video Technology, 34(8), 6833–6846.

Tian, Y., Li, Y., Chen, B., Zhu, H., Wang, S., & Kwong, S. (2025). AI-generated image quality assessment in visual communication. Proceedings of the AAAI Conference on Artificial Intelligence, 39(7), 7392–7400.

Aghazadeh, A., & Kovashka, A. (2025). CAP: Evaluation of persuasive and creative image generation. Proceedings of the IEEE/CVF International Conference on Computer Vision, 16970–16980.

Hartwig, S., Engel, D., Sick, L., Kniesel, H., Payer, T., Poonam, P., Glöckler, M., Bauerle, A., & Ropinski, T. (2025). A survey on quality metrics for text-to-image generation. IEEE Transactions on Visualization and Computer Graphics, 31(10), 9464–9483.

인간평가 · 통계검증 · 실험설계 관련 주요 참고문헌

Thurstone, L. L. (1927). A law of comparative judgment. Psychological Review, 34(4), 273–286.

Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. Biometrika, 39(3/4), 324–345.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1), 37–46.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. Psychological Bulletin, 76(5), 378–382.

Krippendorff, K. (2004). Content analysis: An introduction to its methodology (2nd ed.). Sage Publications.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin, 86(2), 420–428.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33(1), 159–174.

Cicchetti, D. V., & Feinstein, A. R. (1990). High agreement but low kappa: II. Resolving the paradoxes. Journal of Clinical Epidemiology, 43(6), 551–558.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum Associates.

Bonferroni, C. E. (1936). Teoria statistica delle classi e calcolo delle probabilità. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze.

생성형 AI 광고 이미지 평가를 위한 벤치마크 데이터셋 및 자동 평가 지표 개발 연구
- 광고사진 표현 형식을 중심으로

경청해 주셔서 감사합니다.