

2025 인하 인공지능 챌린지

언어 정보 기반 이미지 색채화

Posterior

22251450 김준희 22232264 이정인
12203290 조수하 12191892 박성혁

Contents

01 Introduction

02 Method

03 Conclusion



01 Introduction

대회 개요 : 언어 정보 기반 이미지 색채화

2025
인하 인공지능
챌린지 개최



1. 자연어로 제공된 캡션을 기반으로 흑백 이미지를 적절한 색상으로 색채화
2. 텍스트 + 이미지 멀티모달 융합을 통한 고품질 색채화
3. 평가 방식

$$\text{Score} = 0.6 \cdot \text{HSV_Similarity} + 0.4 \cdot \text{CLIP_Score}$$

- 1) HSV_Similarity: 실제 이미지와 생성 이미지 HSV 히스토그램 유사도
- 2) CLIP_Score: 캡션과 생성 이미지의 코사인 유사도 (ViT-L-14)

데이터 개요

1. Train Data Set

- 학습용 흑백 이미지 65,260개 샘플
- 학습용 색채 이미지 65,260개 샘플
- 흑백이미지 색채 관련 캡션



what is the chair to the right of the child made of? pink cup on table.
what is in the silver bowl? what color is the pizza cutter? pink and blue plastic cup.
yellow design on shirt. on which side of the picture is the basil?
pizza with extra basil on top.

2. Test Data Set

- 평가용 흑백 이미지 200개 샘플
- 흑백이미지 색채 관련 캡션



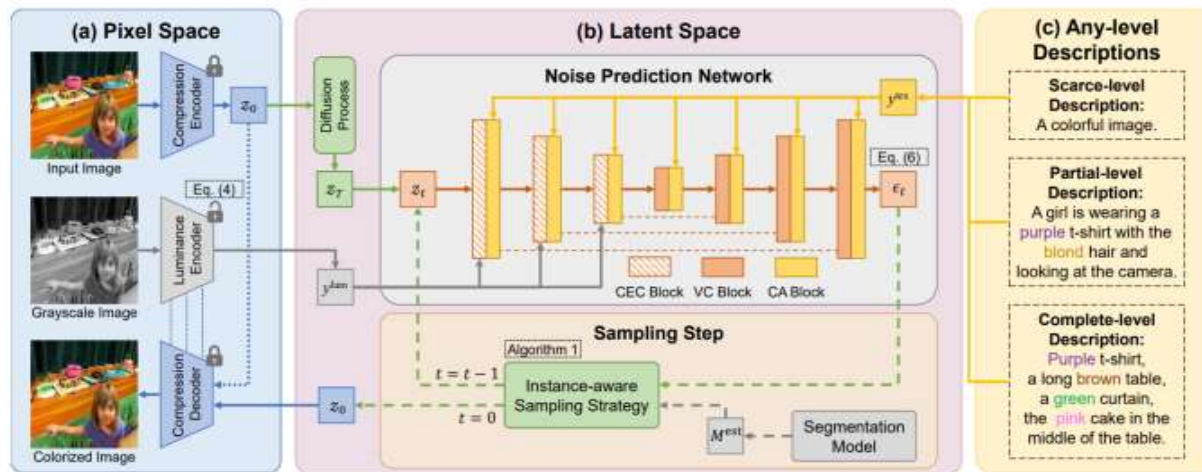
tennis racquet with white taped handle. yellow tennis ball in motion.
red polo shirt with cross logo. line judge sitting behind player. a white wrist band.
red nike head band. white t-shirt under a red polo shirt.
white cross on a red background.

A large, solid dark blue triangle is positioned on the left side of the slide, pointing towards the bottom right.

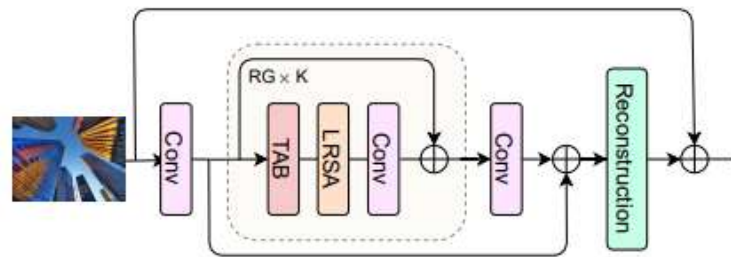
02 Method

2.0 전체 파이프라인

Step1. 색채화 (512×512 → 256×256, L-CAD)



Step3. 후보정



(a) Overall Architecture

Step2. 초해상화 (256×256 → 512×512, CATANet)



2.1 색채화 모델 : L-CAD

1. 모형 개요

- Stable Diffusion 모델을 색채화 태스크에 맞게 파인튜닝한 사전학습 모델 사용
- 그레이 스케일 정보(구조, 윤곽)를 보존하면서 색상 정보만 추가하는 방법론

L-CAD: Language-based Colorization with Any-level Descriptions using Diffusion Priors

Zheng Chang^{#1} Shuchen Weng^{#2,3} Peixuan Zhang¹ Yu Li⁴ Si Li¹ Boxin Shi^{2,3}

¹ School of Artificial Intelligence, Beijing University of Posts and Telecommunications

² National Key Laboratory for Multimedia Information Processing

School of Computer Science, Peking University

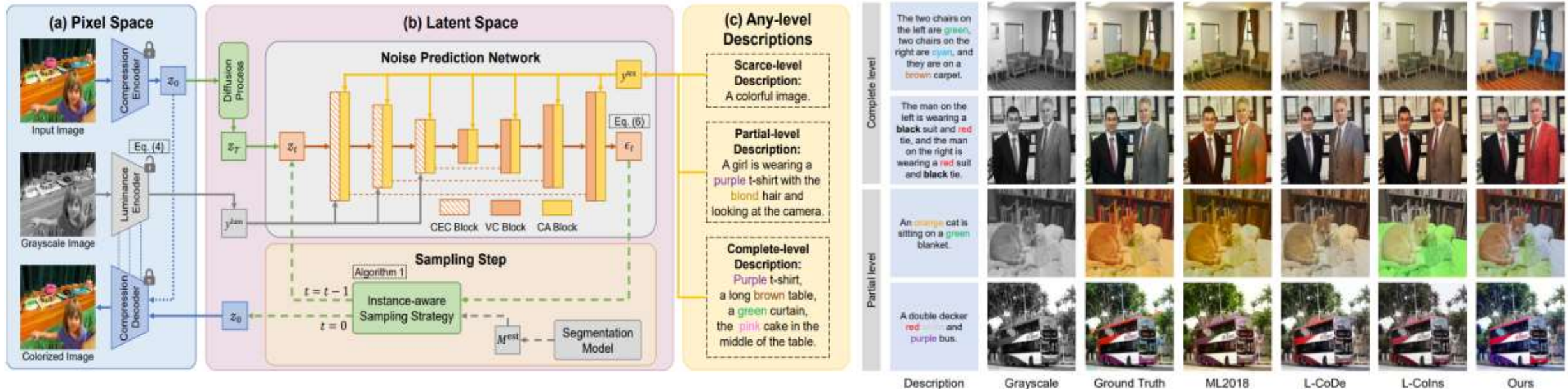
³ National Engineering Research Center of Visual Technology

School of Computer Science, Peking University

⁴ International Digital Economy Academy

{zhengchang98, pzhang, lisi}@bupt.edu.cn

{shuchenveng, shiboxin}@pku.edu.cn liyu@idea.edu.cn



2.1 색채화 모델 : L-CAD

2. 모형 특징

1) Luminance-encoder

휘도 인코더를 통해 그레이스케일 정보를 추출

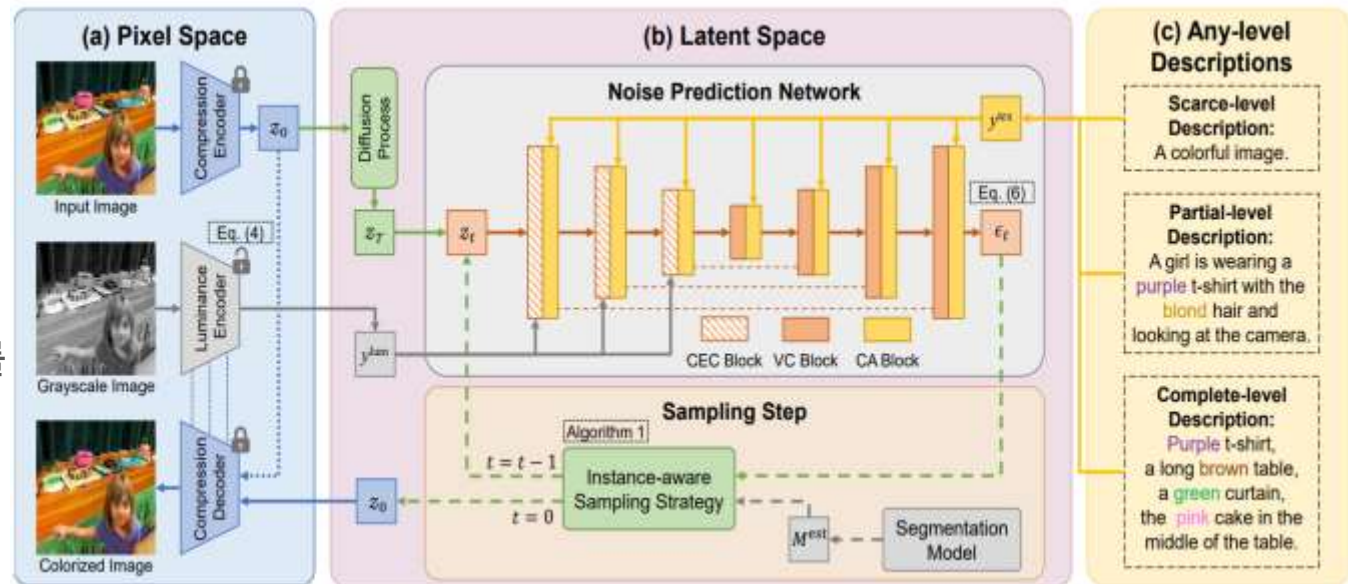
2) Semantic-aligned latent representation

다운 샘플링의 일반 합성곱 블록을 휘도 인코더 정보를
함께 통합하는 CEC 블록으로 교체
잠재 공간 내 이미지 실제 구조가 대응되도록 함.

3) Any-level description

다양한 색채 주석 레벨 (Any-level Description)로 학습 범용성 확보

- Scarce-level: 단순 색감 설명
- Partial-level: 특정 객체와 색 정보
- Complete-level: 객체 색+위치+속성 등의 통합적 정보



2.2 초해상화 모델 : CATANet

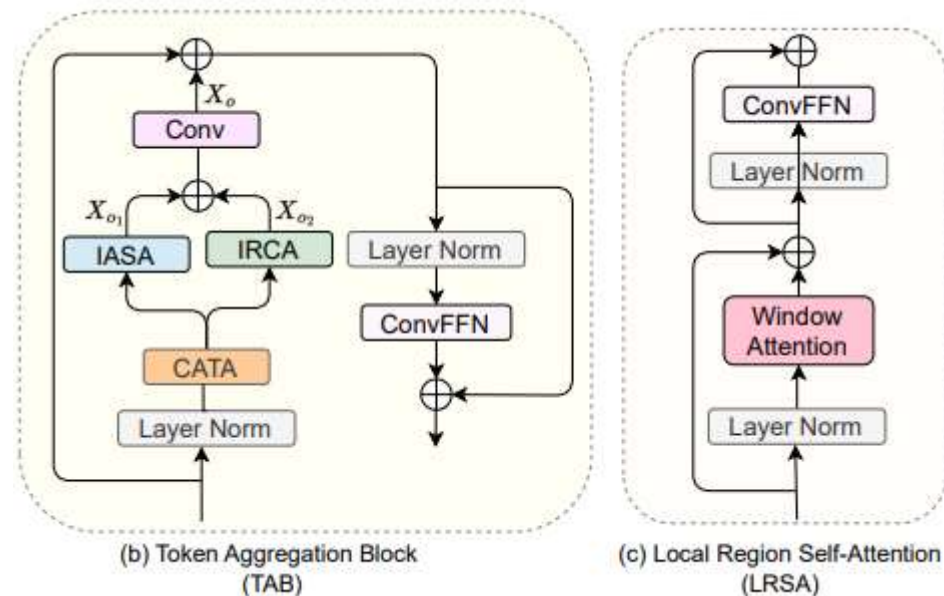
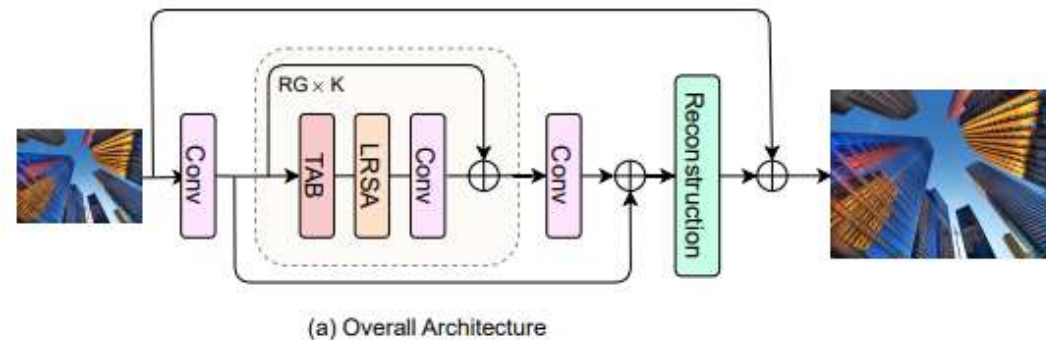
2. 모형 핵심 구조

1) Token Aggregation Block (TAB)

- IASA (Instance-Aware Spatial Aggregation)
객체 경계와 인스턴스별 구조를 고려해 특징을 집약,
전역 문맥 + 지역 구조 정보를 동시에 반영
- IRCA (Instance-Region Cross Attention)
서로 다른 인스턴스 영역 간의 상호작용 학습, 색·텍스처 누락 방지
- CATA (Cross-Attention Token Aggregation)
글로벌 토큰과 로컬 토큰의 상호 보완적 융합, 세부 질감 + 전역 색조 균형 유지

2) Local Region Self-Attention (LRSA)

- 입력 이미지를 윈도우 단위로 나눠서(Window Attention)
지역적 self-attention 수행



2.3 후보정 적용

1. 감마보정 (Gamma Correction) : 밝기와 대비를 비선형적으로 조정

- 값이 1보다 커질 시 어두운 영역 밝기 증가, 1보다 작아질 시 밝은 영역과 어두운 영역의 대비 강화

2. CLAHE 대비 보정 (CLAHE Contrast) : 전체적인 명암 대비 향상

- LAB 색공간의 L 채널(밝기)에만 적용하여 색정보는 그대로 유지하면서 밝기 대비 조절

3. 채도강화 (Boost Saturation) : 이미지 색상 선명화

- Scale이 1보다 크면 채도 증가, 1보다 작으면 채도 감소

4. 언샤프 마스크 (Unsharp Mask) : 경계 부분을 강조하여 선명도 향상

- 경계부분을 강조하여 전반적인 선명도를 향상

5. 양방향 필터 (Bilateral Filtering) : 경계는 보존한 채 노이즈를 최소화

- 색상 차이에 따른 필터영향, 공간 거리에 따른 영향 범위 조절하여 후보정



03 Conclusion

3.1 실험 결과 정량적 비교 (LB)

	색채화 모델	초해상화 모델	후보정 방식	Public 스코어	Private 스코어
실험1	L-CAD	CATAnet	-	0.686	0.689
실험2			실험1 + 감마보정	0.698	0.699
실험3			실험2 + 대비보정	0.694	0.698
실험4			실험3 + 채도 강화	0.693	0.696
실험5			실험4 + 언샤프, 양방향필터	0.703	0.703

3.2 최종 파이프라인 정성적 결과

1. 색채화

(512×512 → 256×256, L-CAD)



2. 초해상화

(256×256 → 512×512, CATANet)



3. 후보정

(감마, CLAHE 대비, 채도, 언샤프, 양방향 필터)



2025 인하 인공지능 챌린지

언어 정보 기반 이미지 색채화