

파노라마 치아 X-ray 합성을 위한 의미 정렬 및 구조 제어 기반 조건부 Diffusion 모델

정대원, 엄세경*
동국대학교 산업시스템 공학과

2025120165@dgu.ac.kr, *skyoum@dgu.edu

A Conditional Diffusion Model Based on Semantic Alignment and Structural Control for Panoramic Dental X-ray Synthesis

Daewon Jung, Sekyoung Youm*

*Dongguk Univ., Department of Industrial Systems Engineering.

요약

파노라마 치아 X-ray는 곡선 투영과 비균일 확대가 결합된 왜곡 영상으로, 범용 Diffusion 모델을 그대로 적용하면 치열 아치와 치아 형태가 쉽게 붕괴한다. 본 연구는 Projected MedCLIP-Structural Diffusion(PMSD)을 제안한다. PMSD는 Stable Diffusion XL(SDXL) 기반 MedCLIP 의미 임베딩을 프로젝터로 정렬해 주입하고, ControlNet-Canny로 구조 조건을 결합하며, Low-Rank Adaptation(LoRA) 기반의 어텐션 전용 경량 미세조정을 수행한다. 23,000장의 파노라마 데이터에서 PMSD는 MedCLIP 제거 변형 대비 관심 영역 구조 유사도(ROI-SSIM)를 0.293에서 0.449로 개선하였고, 정성 비교에서도 치아 경계와 하악 윤곽의 일관성이 향상됨을 보였다.

I. 서론

의료 영상은 개인정보 규제와 라벨링 비용으로 데이터 확보가 제한되며, 합성 데이터의 효용은 해부학적 형태 보존과 인공물 억제로 좌우된다. 파노라마 치아 X-ray는 투영 왜곡이 심하므로, Denoising Diffusion Probabilistic Model (DDPM), Latent Diffusion Model (LDM), Stable Diffusion XL (SDXL)과 같은 범용 사전 학습 모델을 적용할 경우 텍스트 조건만으로 치열 구조를 안정적으로 제어하기 어렵다[1-3]. 이를 보완하기 위해 구조 조건을 주입하는 ControlNet이 활용되며[4], 제한된 자원에서도 도메인 적응을 가능하게 하는 Low-Rank Adaptation(LoRA)이 널리 사용된다[5]. 또한 의료 텍스트-이미지 의미 정합을 위한 의료 영상-텍스트 정렬 모델인 Med-CLIP이 제안되었다[6]. 본 연구는 SDXL 기반에서 MedCLIP 의미 임베딩 정렬, ControlNet-Canny 구조 제어, Attention-only LoRA 경량 적응을 통합한 Projected MedCLIP-Structural Diffusion (PMSD) 모델을 제안한다.

II. 방법론

입력 파노라마 영상 I_{in} 과 텍스트 프롬프트 p 가 주어졌을 때, 의미 조건과 구조 조건을 동시에 반영한 합성 영상 \hat{I} 를 생성하는 조건부 분포 $p_{\theta}(\hat{I}|I_{in}, p)$ 를 학습한다. 구조 조건은 입력 영상에서 추출한 Canny edge 맵 $S = Canny(I_{in})$ 로 정의한다. 학습은 조건부

self-reconstruction 설정으로 수행하며, 목표 영상 $x_0 = I_{in}$ 로 동일하게 둔다.

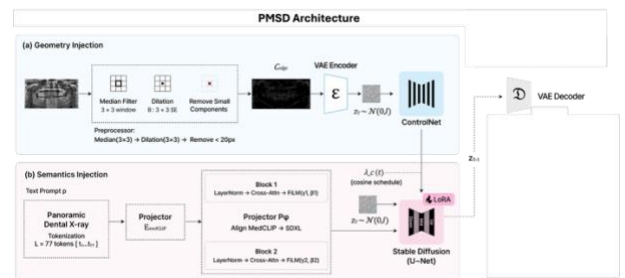


그림 1. 제안 모델(PMSD)의 전체 아키텍처 개요

PMSD는 SDXL 기반 Diffusion 기반 모델에 의료 의미 조건, 구조 조건, 경량 적응을 결합한다. 의료 의미 조건은 MedCLIP 텍스트 임베딩을 경량 프로젝터로 SDXL 조건 공간에 정렬해 주입하며, 구조 조건 S 는 ControlNet에서 인코딩되어 U-Net의 각 해상도 블록에 잔차(residual)로 합산되며, 이를 통해 치열 아치 및 하악 윤곽의 전역 형태 붕괴를 억제한다.

도메인 적응은 전체 U-Net을 재학습하지 않고, Attention 선형 Projection 계층에만 LoRA를 적용하여 파라미터 효율적으로 수행한다. 실질적으로 학습되는 파라미터는 프로젝터 P_{θ} , ControlNet, LoRA 파라미터이며, SDXL 기반의 기본 가중치는 고정하여 학습 안정성과 비용을 동시에 확보한다.

전방 확산 과정은 다음과 같이 정의된다.

$$q(x_t | x_0) = \mathcal{N}(x_t; \sqrt{\bar{a}_t}x_0, (1 - \bar{a}_t)\mathbf{I}) \quad (1)$$

여기서 $x_0 = I_{in}$ 이며, 모델은 x_t , 시간 t , 의미 조건 \tilde{e} , 구조 조건 S 를 입력받아 노이즈/신호 혼합 변수인 v 를 예측한다. $\alpha'_t = \sqrt{\bar{a}_t}$, $\sigma_t = \sqrt{(1 - \bar{a}_t)}$ 로 두면 다음과 같이 정의된다.

$$x_t = \alpha'_t x_0 + \sigma_t \epsilon, \quad \epsilon \sim \mathcal{N}(0, \mathbf{I}), \quad v = \alpha'_t \epsilon - \sigma_t x_0 \quad (2)$$

학습손실은 다음과 같다.

$$\mathcal{L}_{denoise}(\theta) = \mathbb{E}[\|v - \hat{v}_\theta(x_t, t, \tilde{e}, S)\|_2^2] \quad (3)$$

LoRA는 Attention 선형 계층의 가중치 W 에 저랭크 갱신을 추가하며, 원래 가중치 W 는 고정하고 ΔW 만 학습한다. 추론 시에는 프롬프트 p 로부터 \tilde{e} 를 계산하고 입력 영상에서 얻은 S 를 ControlNet에 제공한 뒤, 표준 샘플링 절차로 \hat{I} 를 생성한다.

III. 실험 및 결과

표 1. 영상 품질 및 임상 유용성 정량 비교

Metric	PMSD	SDXL (pretrained)	SDXL w/o MedCLIP
PSNR	11.913	13.291	10.491
SSIM	0.478	0.488	0.375
ROI-SSIM	0.449	0.507	0.293
전문가 수용 진단성	52.2%	5.6%	20%

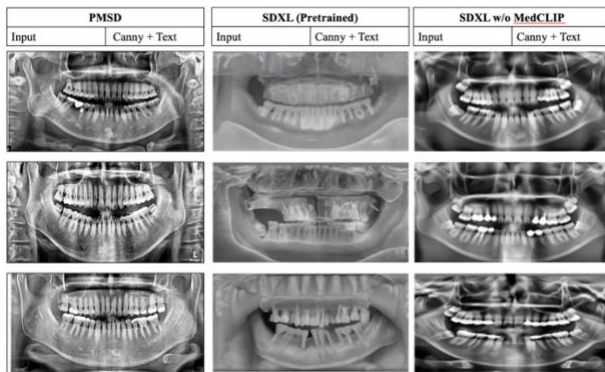


그림 2. 생성 결과 시각적 비교

본 연구는 단일 기관의 비식별화 파노라마 치아 X-ray 23,000 장을 사용한다. 학습 20,000 / 검증 1,000 / 실험 2,000 으로 나누며, Test는 최종 정량, 정성 평가에만 사용한다. ROI 기반 지표는 치열과 하악을 포함하는 고정 사각 ROI를 정의하고, 입력, 생성, 참조 영상에 동일 좌표로 적용해 계산한다. 비교 대상은 diffusion 계열 기준선과 제안 방법이다. SDXL(pretrained)은 사전학습 SDXL에 Canny 구조 조건과 텍스트를 입력하는 체로샷 설정이며, SDXL w/o MedCLIP은 Canny 구조 조건에서 MedCLIP 없이 LoRA로 미세조정을 한 모델이다. 모든 방법은 동일한 1024×512 해상도의 추론 설정으로 비교한다.

평가는 정량 지표와 전문가 기반 정성 요약으로 구성하며, 각 방법당 Test에서 500 샘플을 생성한다. 표 1에서 PMSD의 PSNR은 SDXL보다 다소 낮게 측정되었다. 이는 PSNR 지표가 픽셀 단위의 오차에 기반하기 때문에, 흐릿하게 평균화된 영상일수록 오히려 점수가 높게 나오는 경향이 있기 때문이다. 반면, 치아 영상 진단에서 핵심적인 요소는 치아의 경계와 해부학적 구조의 보존이다.

따라서 구조적 유사도를 측정하는 관심 영역 구조 유사도(ROI-SSIM) 지표가 0.293에서 0.449로 대폭 향상된 점과, 실제 전문가들이 판독 가능하다고 응답한 전문가 수용 진단성이 기준 모델 대비 압도적으로 높다는 점은 PMSD가 임상적으로 더 유의미한 고품질 영상을 생성함을 입증한다. 그림 2에서 SDXL(pretrained)은 윤곽은 유지되지만 명암 분포, 조직 대비, 금속 인공물 표현이 부자연스러워 파노라마 X-ray로서의 시각적 개연성이 낮은 사례가 관찰된다. 이를 보완하기 위해 전문가 수용 진단성(%)을 함께 보고하며, 이는 비식별화된 조건에서 생성 영상이 실제 파노라마 X-ray로 수용 가능한지(Yes/No)를 전문가가 판단한 비율이다. 표 1에서 PMSD는 SDXL w/o MedCLIP 대비 ROI-SSIM과 전문가 수용 진단성을 함께 개선한다.

IV. 결론

본 연구는 파노라마 치아 X-ray 합성에서 텍스트 조건만으로는 불안정한 치아 형태 제어 문제를 완화하기 위해, SDXL 기반 Diffusion 기반에 의료 의미 임베딩, 구조 조건, 경량 적응을 통합한 PMSD를 제안하였다. PMSD는 MedCLIP 의미 priors를 프로젝터로 정렬 및 주입하고, ControlNet-Canny로 구조 조건을 결합하며, LoRA 기반 경량 미세조정으로 도메인 적응을 수행한다. 실험에서 PMSD는 MedCLIP 제거 변형 대비 ROI 기반 품질을 개선하고, 정성 비교에서도 치아 경계 붕괴와 아치 형태 불안정, 인공물 발생이 감소하는 경향을 보였다.

ACKNOWLEDGMENT

본 연구는 2025년도 정부(교육부)의 재원으로 한국연구재단 기초연구사업의 지원을 받아 수행된 연구임 (No. RS-2025-25425542).

참고 문헌

- [1] J. Ho et al., "Denoising Diffusion Probabilistic Models," NeurIPS 2020.
- [2] R. Rombach et al., "High-Resolution Image Synthesis with Latent Diffusion Models" CVPR, 2022.
- [3] D. Podell et al., "SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis", ICLR, 2024.
- [4] L. Zhang et al., "Adding Conditional Control to Text-to-Image Diffusion Models" ICCV, 2023.
- [5] E. Hu et al., "LoRA: Low-Rank Adaptation of Large Language Models", ICLR, 2022.
- [6] Z. Wang et al., "MedCLIP: Contrastive Learning from Unpaired Medical Images and Text", EMNLP, 2022.