

단일 이미지 기반 기하 일관 다중 시점 합성과 생성적 보안을 이용한 데이터 증강

임채운, 박현희*

명지대학교

{dlacodbs0627, hhpark*}@mju.ac.kr

Geometry-guided Multi-View Data Augmentation from a Single Image with Generative Completion

Lim Chae Yun, Park Hyun Hee*

Department of Applied Artificial Intelligence, Myongji University

요약

최근 컴퓨터 비전에서는 대규모 사전학습 모델이 확산되고 있으나, 도메인 특화 환경에서는 데이터 수집 제약으로 관점 다양성이 제한된다. 기존 데이터 증강 기법은 외형적 변형이나 생성 모델 의존으로 인해 시점 변화 시 기하 구조의 일관성을 유지하기 어렵다. 본 논문은 단일 이미지로부터 기하 정보를 보존한 다중 시점 이미지를 생성하는 데이터 증강 파이프라인을 제안한다. 포인트 맵과 카메라 포즈를 활용한 객체 중심 시점 변환과 결손 영역의 제한적 생성 보안을 통해 기하 구조의 일관성을 유지한다. ImageNet-LT 기반 데이터 불균형 환경에서의 회소 클래스 분류 실험에서 제안 기법은 회소 클래스 성능과 전체 정확도를 모두 향상시켜, 데이터 불균형 환경에서 일반화 성능 개선 효과를 확인하였다.

I. 서론

대규모 사전학습 모델은 비전 응용 분야에서 폭넓게 활용되고 있다. 하지만 국방·산업 등 도메인 특화 환경에서는 데이터 수집과 공개의 제약으로 데이터 다양성이 부족하고, 이로 인해 학습 일반화 성능이 저하된다. 이에 대한 대응 방안으로 데이터 증강이 활용되지만, 전통적 데이터 증강 기법은 외형적 변형에 국한된다. 최근 연구에는 이미지 생성 모델을 통해 관점 다양성을 증가시키고 있지만, 생성 모델의 특성상 시점 변화 시 기하 구조의 일관성을 보장하기 어렵다. 다중 시점 합성(novel view synthesis, NVS) 분야는 관점 다양성 확장에 효과적이나, 대부분 다각도 입력과 장면별 최적화가 필요해 데이터 제한 환경에 적용이 어렵다. 단일 이미지 기반 NVS 또한 생성 모델 중심 접근으로 기하 일관성 확보가 어렵다[1].

본 연구는 이러한 한계를 극복하기 위해 단일 이미지의 기하 구조 일관성을 유지하면서 생성적 보안을 적용한 증강 기법을 제안한다. 추가로, 데이터 불균형이 심한 회소 클래스 분류 환경에서의 실험을 통해 제안하는 기법이 데이터 증강 관점에서 실질적인 다운스트림 성능 향상에 기여할 수 있음을 정량적으로 검증한다.

II. 제안 기법

그림 2는 제안하는 기법의 전체 파이프라인을 보여준다. 가장 먼저, 단일 입력 이미지로부터 프롬프트 기반 분할 모델을 이용해 증강 대상 객체의 마스크를 추출하며, 동일 클래스 객체가 다수 존재하는 경우 가장 넓은 영역을 갖는 객체를 선택한다. 그리고 기하 예측 모델로부터 획득한 포인트 맵을 활용하여 객체 중심 회전 기준점을 추정한다. 이를 위해 객체 마스크의 중심점 방향으로 광선을 투사하고, 포인트 맵 상에서 해당 광선과 가장 가까운 3차원 포인트를 기준점으로 선택함으로써, 별도의 3차원 객체 탐지 없이도 효율적인 객체 중심 추정이 가능하다. 회전 기준점을 추정했다면, 해당 점을 기준으로 객체가 시점 변환 과정에서 시야 밖으로 벗어나는 것을 방지하기 위해 카메라 시점을 객체 방향으로 정렬한다. 이후 기준점과의 거리를 유지한 채 카메라를 수평 또는 수직 방향으로 회전시켜 새로운 시점의 이미지를 렌더링한다. 마지막으로, 시점 변환 이후 포인트 맵이 존재하지 않는 영역을 결손 영역으로 정의하고, 해당 영역에 대해서만 인페인팅 기반 생성 모델을 적용함으로써 신뢰 가능한 기하 정보를 최대한 보존하면서 생성 모델로 인한 구조 왜곡과 환각 현상을 최소화한다.

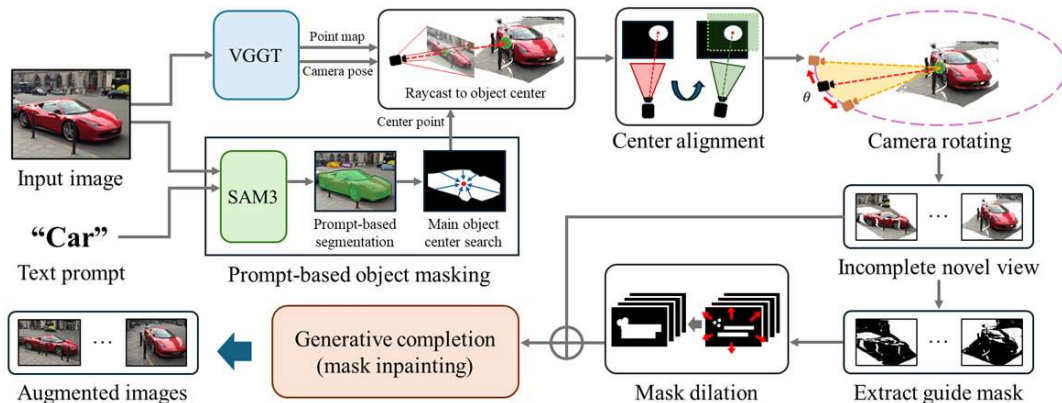


그림 1. 기하 변환을 활용한 다중 시점 합성 데이터 증강 기법 전체 아키텍처

III. 실험 및 결과

본 연구에서는 제안한 증강 파이프라인의 구현 설정과 long-tailed 분류 환경에서의 실효성을 실험을 통해 검증한다. 가상 시점은 Open3D 기반 렌더링과 point size 4를 사용하며, 시점 변환으로 발생한 결손 영역 보완에는 LaMa[2] 인페인팅 모델을 적용한다. 결손 영역이 매우 작은 경우에는 마스크 증폭을 통해 인페인팅 안정성을 확보한다. 시점 회전 각도는 기하 구조 보존과 결손 영역 비율 간 균형을 고려하여 15도와 30도로 설정하고, 단일 이미지당 수평 및 수직 방향으로 총 네 장의 이미지를 증강한다.

최소 클래스 분류 성능 평가는 ImageNet-LT[3] 데이터셋에서 수행하며, 클래스 분포에 따라 many, medium, few로 구분한다. 분류 모델은 ResNet-50을 사용하고, 증강 미적용 baseline 및 전통적 증강 기법과 비교한다. 제안 기법은 few 클래스에 한해 오프라인으로 적용되며, many, medium, few 클래스별 및 전체 top-1 정확도를 기준으로 평가한다. 모든 실험은 동일한 하이퍼파라미터로 90 epoch 학습을 수행하였다.

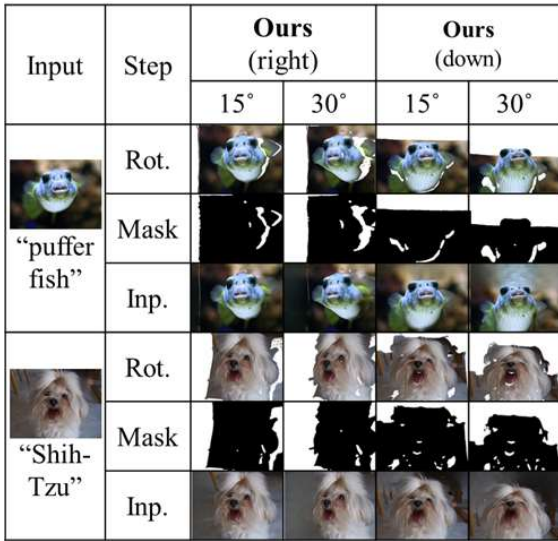


그림 2. ImageNet-LT에서 제안 기법의 단계별 결과 예시

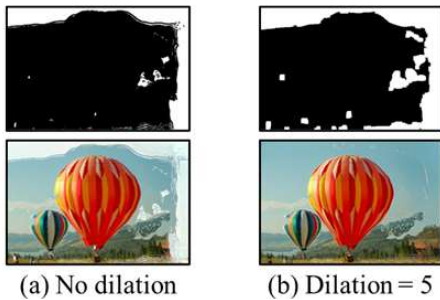


그림 3. 마스크 증폭 적용 전·후 과정 및 결과

그림 2은 단일 이미지로부터 생성된 가상 시점 렌더링과 결손 영역 보완 결과를 보여준다. 제안 기법은 객체 중심 회전을 통해 관점 다양성을 확장하면서도 객체의 형태와 공간적 배치를 안정적으로 유지한다. 회전 각도가 커질수록 관점 다양성은 증가하나 결손 영역 역시 함께 증가하며, 작은 결손 영역의 경우 그림 3과 같이 마스크 증폭을 적용함으로써 생성 안정성을 개선할 수 있다. 표 2의 ImageNet-LT 분류 실험 결과, 증강을 적용하지 않은 경우 few 클래스에서 매우 낮은 성능을 보였으나, 제안 기법을 적용하면 few 클래스 성능이 일관되게 향상되고 전체 정확도 또한 개선된다. 이는 기하 일관 관점 다양성 확장이 최소 클래스의 일반화 성능에 효과적임을 의미한다. 전통적 증강과 결합한 경우 15° 회전이 가장 안정적인 성능을 보였으며, 30° 회전은 단독 적용 시에는 효과적이나 외형

증강과 결합할 경우 학습 노이즈가 증가하는 경향을 확인하였다.

표 1. ImageNet-LT에서의 long-tailed 분류 성능 비교

Augmentation	Accuracy(%) ↑			
	many	medium	few	overall
None	35.522	14.751	3.329	20.222
Ours(15°)	39.643	16.858	6.961	23.260
Ours(30°)	41.084	18.299	6.868	24.468
Trad.	62.110	35.050	11.671	40.886
Ours(15°)+Trad.	62.663	35.956	11.947	41.574
Ours(30°)+Trad.	62.202	34.806	11.842	40.822

한편, 그림 4와 같이 포인트 맵 예측이 부정확한 장면이나 객체가 프레임 가장자리에 위치한 경우에는 기하 왜곡이 발생할 수 있고, 인페인팅 의존도가 증가하여 생성 결과의 안정성이 저하될 수 있다. 이는 과도한 회전 각도에서 더욱 두드러지며, 본 논문은 이를 방지하고자 시점 변환을 15°에서 30°로 제한하여 실험을 진행하였다.



(a) Background-dominant (b) Boundary object (c) Large rotation

그림 4 제안하는 증강 기법의 실패 사례

IV. 결론

본 논문은 단일 이미지로부터 기하 정보를 보존한 다중 시점 이미지를 생성하여 데이터 증강 파이프라인을 제안한다. 제안 기법은 별도 추가 학습이나 장면 최적화 없이도 단일 이미지 입력만으로 증강이 가능하다는 점에서 데이터 수집과 라벨링이 제한된 환경에서 실용적인 활용 가능성을 갖는다. 실험 결과, 제안 기법은 long-tailed 분류 문제에서도 일반화 성능을 일관되게 향상시키며, 전체 분류 성능에도 긍정적인 영향을 미침을 확인하였다. 다만, 제안 기법은 기하 예측 모델 정확도에 의존하는 한계가 존재한다. 본 논문은 이러한 한계를 고려하여 제한적인 회전 범위 내에서 증강 효과를 검증하였다. 향후 연구에서는 증강 결과를 다시 활용한 반복적 증강 전략을 탐구하고자 한다. 궁극적으로는 자유 시점 장면 증강을 달성하는 것을 목표로 한다.

ACKNOWLEDGMENT

본 과제(결과물)는 2025년도 교육부 및 경기도의 재원으로 경기RISE센터의 지원을 받아 수행된 지역혁신중심 대학지원체계(RISE)의 결과입니다. (2025-RISE-09-A15)

참고 문헌

- [1] H. Kim et al., "Single-Image 3D Reconstruction Using Self- and Cross-Attention Mechanisms," *J. Inst. Electron. Eng. Korea*, pp. 112-119, 2024.
- [2] R. Suvorov et al., "Resolution-robust Large Mask Inpainting with Fourier Convolutions," *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, Jun. 2021.
- [3] Z. Liu et al., "Large-Scale Long-Tailed Recognition in an Open World," *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, Jun. 2019.