목표

CRM 모델의 설계, 구현 및 성능 평가를 통해 단일 이미지에서 3D 텍스처 메시를 효율적으로 생성하는 방법론을 분석.

배경 및 문제점

1. Transformer 기반 모델이 언어, 이미지, 비디오 생성 등 다양한 분야에서 높은 성능을 보여주지만, 해당 모델은 Geometric 영역을 구조화하지 못함.
2. Feed-forward 3D generative models (e.g. LRM) 은 주로 triplane을 사용하여 결과를 생성하지만, 기하학적 우선순위를 잘 활용하지 못해 최적의 결과를 도출해내지 못함. 또한 학습 시간이 매우 길어 비효율적임.

환경 설정

1. **데이터셋**: Objaverse에서 필터링된 고품질 376k 데이터를 사용.
2. **학습 세팅**:
   A. 8개의 NVIDIA A800 GPU를 사용해 6일간 11만회 반복을 통해 학습 진행.
   B. 학습 중 Gaussian 노이즈와 랜덤 리사이즈(random resizing)로 모델의 강건성을 향상.
3. **적용 기술**:
   A. Zero-SNR Training
   B. Random Resizing
   C. Contour Augmentation

방법론

1. 6방위 orthographic view 이미지를 single 이미지에서 생성
2. 1.에서 생성한 이미지를 Convolutinal U-Net에 Feeding
3. 고해상도 triplane 생성을 위해 pixel-level 정렬 및 변환 대역 개선

특이점

1. 실루엣 + 텍스처로 triplane 구조체 생성
2. Convolutional U-Net과 Canonical Coordinate Map을 적용하여 reconstruction network 생성
3. **Orthographic views, CCMS**: not directly available
   - Multi-view diffusion model을 선행 학습시켜 생성
4. NeRF, Gaussian Splatting과 달리, textured meshes를 얻기 위해 추가 step 진행
5. **Loss Function**: MSE, LPIPS loss

기술적 기여

- CCM을 추가하여 기하학적 디테일을 향상.
- U-Net 기반 설계를 통해 Transformer보다 더 효율적이고 상세한 결과 생성.
- Flexicubes로 메시의 품질과 학습 속도를 개선.

결과

- **정량적 평가**: Chamfer Distance, Volume IoU, F-Score 및 PSNR 등의 지표에서 기존 모델보다 높은 성능을 달성.
- **정성적 평가**: CRM이 기존 방법보다 더 부드럽고 디테일한 메시를 생성.
- **한계**: 불규칙한 입력 조건에서는 품질이 저하될 가능성 존재.

결론

CRM은 기하학적 사전 정보와 효율적인 네트워크 설계를 결합하여 3D 콘텐츠 생성의 새로운 기준을 제시합니다. 엔드투엔드 학습으로 높은 품질과 빠른 생성 속도를 제공하며, 향후 3D 생성 분야에서 다양한 응용 가능성을 보입니다. 하지만 입력이미지의 높은 각도 또는 다른 시야각에서는 결과가 만족스럽지 않을 수 있으며 다중 뷰가 항상 완전히 일관된 결과를 생성하기 어렵기 때문에 품질을 저하 시킬수 있습니다.

추가로 다른 생성 모델과 마찬가지로 CRM은 악의적 혹은 허위의 3d 컨텐츠를 생성하는데 사용될 가능성이 있어 주의가 필요합니다.