3주차 스터디

텍스트	Deep Learning-based Image and Video Inpainting: A Survey

초록

추상 이미지 및 비디오 인페인팅은 이미지 및 비디오의 누락된 영역에서 그럴듯하고 사실적인 콘텐츠를 채우는 것을 목표로 한다.

이 논문의 목표는 이미지 및 비디오 인페인팅을 위한 딥 러닝 기반 방법을 종합적으로 검토하는 것이다. CNN, VAE, GAN, 확산 모델 등을 다루며, 낮은 수준의 픽셀과 높은 수준의 센서당 유사성에 대한 평가 지표를 제시하고, 성능 평가를 수행하며, represen tative inpainting 방법의 강점과 약점에 대해 논의한다.

비디오는 시간적 일관성을 나타내는 여러 이미지로 구성되므로 비디오 인페인팅은 이미지 인페인팅과 밀접한 관련이 있으며, 전자는 종종 후자로부터 배우거나 확장한다.

GPT를 통해 정리한 내용

🧩 2. 분류 체계 (기술 분류 기준)

● ① 학습 방식

- **지도 학습** (supervised)
- **비지도 / 약지도 학습** (self-supervised, weakly supervised)

● ② 모델 아키텍처

- **CNN 기반 인페인팅**
  - 구조: U-Net, Partial Conv 등
  - 특징: 지역 정보 강점
- **GAN 기반 인페인팅**
  - 구조: Context Encoder, EdgeConnect, CRFill 등
  - 특징: 리얼리즘, 시각적 일관성 우수

- **Transformer 기반 인페인팅**
  - 구조: IPT, MAT, TransFill
  - 특징: 전역 관계 처리, 문맥 보존
- **Diffusion 기반 인페인팅**
  - 구조: RePaint, DALL-E 2 기반 모델
  - 특징: 고품질 결과, 점진적 복원

● ③ 공간/시간 차원 확장

- **이미지 인페인팅 ↔ 비디오 인페인팅**
  - 비디오 인페인팅은 시간적 일관성 필요 (Optical Flow 등 활용)

🔧 3. 주요 기술 요소

● 마스킹 전략

- **불규칙(mask irregularity)**, **전경 제거(foreground masking)** 등 다양
- 모델의 일반화 성능에 큰 영향을 미침

● 손실 함수

- 재구성 손실: L1/L2, Perceptual Loss
- 적대적 손실: GAN Loss
- 경계/텍스처 유지 손실 등

● 포스트프로세싱

- Refinement network, Guided filtering 등으로 품질 개선

📊 4. 평가 지표 및 벤치마크

● 대표 지표

- **PSNR**: 평균 복원 품질
- **SSIM**: 구조적 유사성
- **FID/LPIPS**: 시각적 유사도 및 다양성 평가

● 데이터셋

- **이미지**: CelebA-HQ, Places2, Paris StreetView
- **비디오**: DAVIS, YouTube-VOS, Vimeo90K

🎯 5. 응용 분야

- **객체 제거 / 편집** (object removal)
- **손상 이미지 복원**
- **콘텐츠 생성** (novel view synthesis)
- **의료 영상 보정** (artifact correction)

🔮 6. 도전 과제 및 향후 연구

1. **고해상도 복원**과 더불어 세부 구조 정밀도 향상
2. **시간적 일관성 확보** (비디오)
3. **범용 인페인팅 모델** 개발 (domain generalization)
4. **비지도 학습 기반 인페인팅** 확대

위의 내용을 토대로 추가 정리

기존의 inpainting: 복구해야 하는 부분의 주변을 토대로 낮은 유사성을 가진 것을 생성함 (한계점)

반대로 딥 러닝은 넓은 영역을 칠하고 더 큰 이미지 세트에서 학습한 새로운 그럴듯한 콘텐트를 칠할 수 있는 가능성을 가지고 있다.

CNN(Convolutional Neural Network)

CNN은 컨볼루션(convolutional), 활성화(activation) 및 다운/업샘플링 레이어(layer)로 구성된 피드포워드 신경망의 한 종류이다.

input 이미지에서 output 이미지로의 매우 비선형적인 매핑을 학습한다.

GAN(Generative Adversarial Network)

GAN은 적대적 프로세스를 통해 데이터 분포를 추정하는 생성기와 판별 입력자로 구성된 생성 모델의 한 유형입니다.

트랜스포머 아키텍처(transformer architecture)와 생성형 디프 융합 모델(Generative dif fusion model)

트랜스포머는 병렬 다중 헤드 주의 모듈을 기반으로 하는 널리 사용되는 네트워크 아키텍처이다.

CNN의 지역성과 비교했을 때, 트랜스포머는 맥락을 이해하는 능력이 더 뛰어나다.

확산 확률 모델은 주로 순방향 프로세스, 역방향 프로세스 및 샘플링 절차를 포함하는 잠재 변수 모델의 한 유형입니다. 확산 모델은 노이즈를 추가하여 데이터를 점진적으로 파괴하는 확률적 프로세스(즉, 확산 프로세스)를 역전시키는 방법을 학습함.

image inpainting의 과정

single-shot framework usu 동맹

손상된 image를 입력으로, inpainted 이미지를 output으로 사용하는 생성기 네트워크를 채택한다.

mask-aware design

누락된 영역(이진 마스크로 표시)은 모양이 다르며 이러한 누락된 영역을 사용한 겹치기에 대한 컨볼루션 연산이 시각적 아티팩트의 원인이 될 수 있습니다. 따라서 일부 연구자들은 고전적인 합성곱 연산 및 정규화를 위한 마스크 인식 솔루션을 제안했습니다. Ren et al. (2015)은 이미지 인페인팅의 고유한 스파가 다양하게 변하는 속성에서 영감을 받아 피처 맵과 마스크가 모두 동일한 컨보레이션 작업을 수행하는 셰퍼드 보간 레이어를 설계했습니다.

마스크 콘솔 볼륨은 마스크를 동시에 업데이트할 수 있습니다. 다양한 불규칙한 구멍을 처리하고 마스크 업데이트 중에 구멍을 진화시키기 위해 Liu et al. (2018)은 컨볼루션 창에서 유효한 영역과 구멍을 구별하는 마스크 유도 컨볼루션 연산, 즉 부분 컨볼루션을 제안했습니다. Xie et al. (2019)은 부분 컨볼루션을 경향을 보이기 위해 훈련 가능한 양방향 어텐션 맵을 제안했으며(Liu et al., 2018), 이는 기능 재정규화 및 마스크 업데이트를 적응적으로 학습할 수 있습니다.

Attention mechanism

시각적 일관성과 의미론적 일관성을 모두 향상시키기 위해 높은 수준의 기능에서 계산된 주의 점수가 낮은 수준의 기능 업데이트에 사용되는 attention 전달 방법을 사용하여 피라미드 컨텍스트 인코더 네트워크를 제안했습니다.

Multi-scale aggregation

다중 스케일 기능 표현을 융합하기 위해 필터 크기가 다른 3개의 컨볼루션 분기로 구성된 생성 다중 열 인페인팅 네트워크를 설계했습니다.

transform domain

Haar 웨이블릿 변환(Haar wavelet transform)을 통해 손상된 이미지를 다른 주파수 성분으로 분해하고(Mallat, 1989), 누락된 영역의 주파수 성분을 예측하기 위해 다중 주파수 확률적 추론 모델을 설계하고, 이미지 공간으로 다시 역 변환했습니다.

Two-stage framework

주로 두 개의 생성기로 구성되며, 첫 번째 생성기는 대략적인 결과를 달성한 다음 두 번째 생성기가 이를 개선한다.

장기적인 상관관계를 생성에 이용한다.

Progressive frameworks

이진 마스크와 손상된 이미지의 연결을 입력으로 사용하고 완료된 이미지를 출력한다.