(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2024년08월05일
(11) 등록번호 10-2691078
(24) 등록일자 2024년07월29일

(51) 국제특허분류(Int. Cl.)
*G06T 5/00* (2024.01) *G06N 20/00* (2019.01)
*G06T 7/50* (2017.01)
(52) CPC특허분류
*G06T 5/70* (2024.01)
*G06N 20/00* (2021.08)
(21) 출원번호 10-2023-0043503
(22) 출원일자 2023년04월03일
심사청구일자 2023년04월03일
(30) 우선권주장
1020230025534 2023년02월27일 대한민국(KR)
(56) 선행기술조사문헌
CN110060215 A*
CN114820344 A*
CN115409733 A*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
**고려대학교 산학협력단**
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
**김승룡**
서울특별시 성북구 오패산로 46, 103동 1804호(하월곡동, 월곡두산위브아파트)
**김경년**
경상북도 안동시 강남1길 133-3(정하동)
*(뒷면에 계속)*
(74) 대리인
**송인호, 윤형근, 최관락**

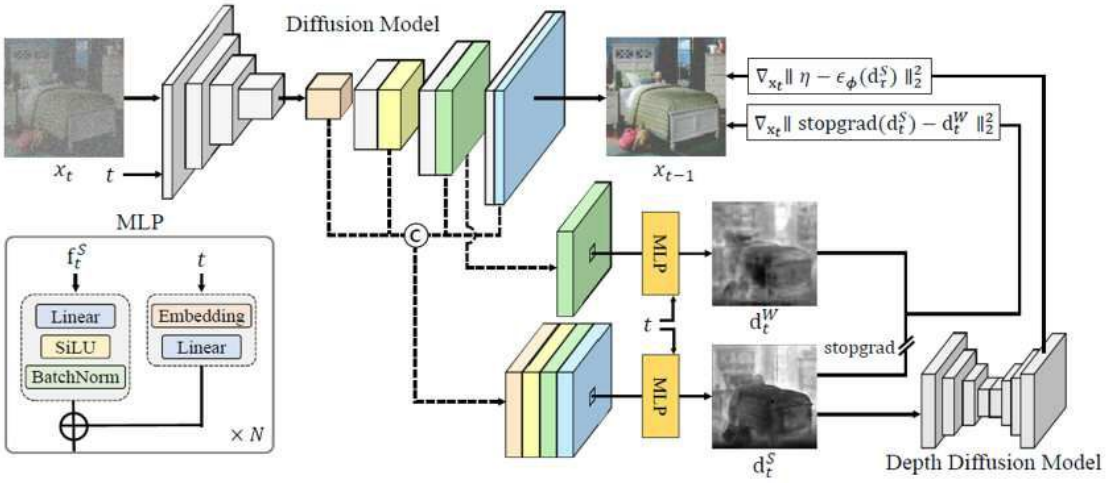
전체 청구항 수 : 총 6 항 심사관 : 김광식

(54) 발명의 명칭 **기하학적 정보를 반영한 노이즈 제거 확산 모델 기반 이미지 생성 방법 및 장치**

## (57) 요 약

본 발명은 기하학적 정보를 반영한 노이즈 제거 확산 모델 기반 이미지 생성 방법 및 장치를 개시한다. 본 발명에 따르면, 프로세서; 및 상기 프로세서에 연결되는 메모리를 포함하되, 상기 메모리는, 인코더-디코더 기반의 노이즈 제거 확산 모델을 이용하여 가우시안 노이즈를 반복적으로 제거하여 이미지를 생성하고, 제1 시간 단계에서, 상기 노이즈 제거 확산 모델의 하나 이상의 디코더 레이어로부터 중간 특징을 추출하고, 미리 학습된 다층 퍼셉트론 기반 깊이 예측기에 상기 중간 특징을 입력하여 깊이 맵을 생성하고, 상기 깊이 맵을 통한 깊이 정보를 상기 노이즈 제거 확산 모델이 상기 제1 시간 단계에서 출력하는 샘플링 이미지에 주입하여 깊이 인식 이미지를 생성하도록, 상기 프로세서에 의해 실행되는 프로그램 명령어들을 저장한 노이즈 제거 확산 모델 기반 이미지 생성 장치가 제공된다.

*대 표 도* - 도1



(52) CPC특허분류
*G06T 7/50* (2017.01)
*G06T 2207/10028* (2013.01)
*G06T 2207/20081* (2013.01)
*G06T 2207/20182* (2013.01)
(72) 발명자
**이규성**
경기도 안양시 동안구 관악대로 121, 108동 2204호 (비산동, 비산 삼성 래미안)
**장우석**
서울특별시 서초구 신반포로 45, 63동 502호 (반포동, 반포아파트)
**홍수성**
서울특별시 영등포구 의사당대로 127, 15층 101-1502호 (여의도동, 롯데캐슬엠파이어)
**서준영**
경기도 안양시 동안구 학의로 20, 118동 102호 (비산동, 관악아파트)

이 발명을 지원한 국가연구개발사업

| | |
|---|---|
| 과제고유번호 | 1711159611 |
| 과제번호 | 2020-0-01819-003 |
| 부처명 | 과학기술정보통신부 |
| 과제관리(전문)기관명 | 정보통신기획평가원 |
| 연구사업명 | 정보통신방송혁신인재양성 |
| 연구과제명 | ICT명품인재양성(고려대학교) |
| 기 여 율 | 1/1 |
| 과제수행기관명 | 고려대학교산학협력단 |
| 연구기간 | 2022.01.01 ~ 2022.12.31 |

공지예외적용 : 있음

# 명 세 서

## 청구범위

### 청구항 1

기하학적 정보를 반영한 노이즈 제거 확산 모델 기반 이미지 생성 장치로서,

프로세서; 및

상기 프로세서에 연결되는 메모리를 포함하되,

상기 메모리는,

인코더-디코더 기반의 노이즈 제거 확산 모델을 이용하여 가우시안 노이즈를 반복적으로 제거하여 이미지를 생성하고,

제1 시간 단계에서, 상기 노이즈 제거 확산 모델의 하나 이상의 디코더 레이어로부터 중간 특징을 추출하고,

미리 학습된 다층 퍼셉트론 기반 깊이 예측기에 상기 중간 특징을 입력하여 깊이 맵을 생성하고,

상기 깊이 맵을 통한 깊이 정보를 상기 노이즈 제거 확산 모델이 상기 제1 시간 단계에서 출력하는 샘플링 이미지에 주입하여 깊이 인식 이미지를 생성하도록,

상기 프로세서에 의해 실행되는 프로그램 명령어들을 저장하되,

상기 프로그램 명령어들은,

k번째 디코더 레이어로부터 제1 중간 특징을 추출하고,

상기 제1 중간 특징을 미리 학습된 다층 퍼셉트론 기반 제1 깊이 예측기에 입력하여 제1 깊이 맵을 생성하고,

복수의 디코더 레이어의 집합으로부터 제2 중간 특징을 추출하고,

상기 제2 중간 특징을 미리 학습된 다층 퍼셉트론 기반 제2 깊이 예측기에 입력하여 제2 깊이 맵을 생성하고,

상기 제2 깊이 맵을 의사 레이블로 하고, 상기 제1 깊이 맵을 예측으로 하여 제1 및 제2 깊이 맵 사이의 일관성 손실을 이용하여 제1 손실을 계산하고,

상기 제1 손실에 따른 깊이 정보를 상기 샘플링 이미지에 주입하고,

상기 제2 깊이 맵에 스탑 그래디언트가 적용되는 노이즈 제거 확산 모델 기반 이미지 생성 장치.

### 청구항 2

삭제

### 청구항 3

삭제

### 청구항 4

삭제

### 청구항 5

제1항에 있어서,

상기 깊이 예측기는 시간 임베딩 블록을 포함하여 각 시간 단계에서 깊이를 측정하는 노이즈 제거 확산 모델 기반 이미지 생성 장치.

### 청구항 6

제1항에 있어서,

상기 프로그램 명령어들은,

상기 깊이 맵을 미리 학습된 깊이 확산 모델에 입력하고,

상기 깊이 확산 모델에서 출력하는 깊이 정보를 상기 노이즈 제거 확산 모델이 상기 제1 시간 단계에서 출력하는 샘플링 이미지에 주입하여 깊이 인식 이미지를 생성하는 노이즈 제거 확산 모델 기반 이미지 생성 장치.

**청구항 7**

제6항에 있어서,

상기 깊이 정보는, 상기 깊이 맵에 추가된 노이즈와 상기 깊이 맵으로부터 예측된 노이즈 사이의 평균 제곱 오차의 그래디언트인 노이즈 제거 확산 모델 기반 이미지 생성 장치.

**청구항 8**

프로세서 및 메모리를 포함하는 장치에서 기하학적 정보를 반영한 노이즈 제거 확산 모델 기반으로 이미지를 생성하는 방법으로서,

인코더-디코더 기반의 노이즈 제거 확산 모델을 이용하여 가우시안 노이즈를 반복적으로 제거하여 이미지를 생성하는 단계;

제1 시간 단계에서, 상기 노이즈 제거 확산 모델의 하나 이상의 디코더 레이어로부터 중간 특징을 추출하는 단계;

미리 학습된 다층 퍼셉트론 기반 깊이 예측기에 상기 중간 특징을 입력하여 깊이 맵을 생성하는 단계; 및

상기 깊이 맵을 통한 깊이 정보를 상기 노이즈 제거 확산 모델이 상기 제1 시간 단계에서 출력하는 샘플링 이미지에 주입하여 깊이 인식 이미지를 생성하는 단계를 포함하되,

상기 중간 특징을 추출하는 단계는,

k번째 레이어로부터 제1 중간 특징을 추출하는 단계; 및

복수의 디코더 레이어의 집합으로부터 제2 중간 특징을 추출하는 단계를 포함하고,

상기 깊이 맵을 생성하는 단계는,

상기 제1 중간 특징을 미리 학습된 다층 퍼셉트론 기반 제1 깊이 예측기에 입력하여 제1 깊이 맵을 생성하는 단계; 및

상기 제2 중간 특징을 미리 학습된 다층 퍼셉트론 기반 제2 깊이 예측기에 입력하여 제2 깊이 맵을 생성하는 단계를 포함하고,

상기 깊이 맵을 생성하는 단계는,

상기 제2 깊이 맵을 의사 레이블로 하고, 상기 제1 깊이 맵을 예측으로 하여 제1 및 제2 깊이 맵 사이의 일관성 손실을 이용하여 제1 손실을 계산하는 단계를 포함하고,

상기 제1 손실에 따른 깊이 정보가 상기 샘플링 이미지에 주입되고,

상기 제2 깊이 맵에 스탑 그래디언트가 적용되는 노이즈 제거 확산 모델 기반 이미지 생성 방법.

**청구항 9**

삭제

**청구항 10**

제8항에 따른 방법을 수행하는 컴퓨터 판독 가능한 기록매체에 저장된 컴퓨터 프로그램.

## 발명의 설명

### 기 술 분 야

[0001] 본 발명은 기하학적 정보를 반영한 노이즈 제거 확산 모델 기반 이미지 생성 방법 및 장치에 관한 것이다.

### 배 경 기 술

[0002] 확산 모델을 통한 이미지 생성 기술은 잠재 변수에서 가우시안 노이즈를 점진적으로 제거하는 역-확산 과정을 통해 무작위 혹은 조건부로 원하는 영역의 이미지를 생성하는 기술이다.

[0003] 이러한 확산 모델에서는 크게 학습과 이미지 추출 과정을 수행하는데, 모델의 성능을 높이기 위한 이미지 추출 과정은 크게 분류기의 유무로 나뉠 수 있다. 그 중 분류기를 활용한 확산 모델은 외부의 분류기의 그래디언트를 활용하여 생성 이미지의 다양성과 품질의 균형을 제어할 수 있다.

[0004] 또한, 최근 들어 다양한 정보를 조건으로 이미지 생성 과정에서 주입시켜 사용자가 원하는 방향의 이미지를 생성하는 연구가 활발히 이루어지고 있으며, 확산 모델의 기학습된 네트워크의 중간 층 표현 정보를 통해 이미지의 의미론적인 추론이 가능하다는 점을 발견한 연구 주제도 주목을 받고 있다.

[0005] 확산 모델은 텍스트 안내 이미지 합성, 이미지 복원, 의미론적 세분화를 포함한 다양한 분야에 적용되고 있다. 그러나 지금까지의 확산 모델은 일반적으로 시각적인 구조적 측면에서만 초점을 맞추고 이미지 생성 중에 형상 기하학에 대한 여부는 고려하지 않는다.

[0006] 실제로 조건 없이 확산 모델을 통해 이미지를 생성했을 때 눈으로 보기에는 그럴듯하지만, 실제 생성 이미지의 깊이 값이나 3차원 특성을 측정해보면 그 영상 속 사물의 원근감이 모호하고 이미지 내의 배치 정보가 망가진 이미지가 생성되는 문제가 있다. 이러한 한계는 생성된 이미지를 활용할 수 있는 다양한 분야에서 안정성 및 신뢰도 측면에서 문제를 야기할 수 있다.

[0007] 특히, 실제 3D 기하학이 필요한 로봇 공학 및 자율 주행과 같은 응용 분야에서는 그 문제가 사고로 이어지거나, 그 외의 다양한 분야에 활용될 수 있는 여지를 저해할 수 있는 요소가 될 것이다.

### 선행기술문헌

#### 특허문헌

[0008] (특허문헌 0001) KR 등록특허 10-2102182

## 발명의 내용

### 해결하려는 과제

[0009] 상기한 종래기술의 문제점을 해결하기 위해, 본 발명은 신뢰성을 높일 수 있는 기하학적 정보를 반영한 노이즈 제거 확산 모델 기반 이미지 생성 방법 및 장치를 제안하고자 한다.

### 과제의 해결 수단

[0010] 상기한 바와 같은 목적을 달성하기 위하여, 본 발명의 일 실시예에 따르면, 기하학적 정보를 반영한 노이즈 제거 확산 모델 기반 이미지 생성 장치로서, 프로세서; 및 상기 프로세서에 연결되는 메모리를 포함하되, 상기 메모리는, 인코더-디코더 기반의 노이즈 제거 확산 모델을 이용하여 가우시안 노이즈를 반복적으로 제거하여 이미지를 생성하고, 제1 시간 단계에서, 상기 노이즈 제거 확산 모델의 하나 이상의 디코더 레이어로부터 중간 특징을 추출하고, 미리 학습된 다층 퍼셉트론 기반 깊이 예측기에 상기 중간 특징을 입력하여 깊이 맵을 생성하고, 상기 깊이 맵을 통한 깊이 정보를 상기 노이즈 제거 확산 모델이 상기 제1 시간 단계에서 출력하는 샘플링 이미지에 주입하여 깊이 인식 이미지를 생성하도록, 상기 프로세서에 의해 실행되는 프로그램 명령어들을 저장한 노이즈 제거 확산 모델 기반 이미지 생성 장치가 제공된다.

[0011] 상기 프로그램 명령어들은, k번째 디코더 레이어로부터 제1 중간 특징을 추출하고, 상기 제1 중간 특징을 미리 학습된 다층 퍼셉트론 기반 제1 깊이 예측기에 입력하여 제1 깊이 맵을 생성하고, 복수의 디코더 레이어의 집합

으로부터 제2 중간 특징을 추출하고, 상기 제2 중간 특징을 미리 학습된 다층 퍼셉트론 기반 제2 깊이 예측기에 입력하여 제2 깊이 맵을 생성할 수 있다.

[0012] 상기 프로그램 명령어들은, 상기 제2 깊이 맵을 의사 레이블로 하고, 상기 제1 깊이 맵을 예측으로 하여 제1 및 제2 깊이 맵 사이의 일관성 손실을 이용하여 제1 손실을 계산하고, 상기 제1 손실에 따른 깊이 정보를 상기 샘플링 이미지에 주입할 수 있다.

[0013] 상기 제2 깊이 맵에 스탑 그래디언트가 적용될 수 있다.

[0014] 상기 깊이 예측기는 시간 임베딩 블록을 포함하여 각 시간 단계에서 깊이를 측정할 수 있다.

[0015] 상기 프로그램 명령어들은, 상기 깊이 맵을 미리 학습된 깊이 확산 모델에 입력하고, 상기 깊이 확산 모델에서 출력하는 깊이 정보를 상기 노이즈 제거 확산 모델이 상기 제1 시간 단계에서 출력하는 샘플링 이미지에 주입하여 깊이 인식 이미지를 생성할 수 있다.

[0016] 상기 깊이 정보는, 상기 깊이 맵에 추가된 노이즈와 상기 깊이 맵으로부터 예측된 노이즈 사이의 평균 제곱 오차의 그래디언트일 수 있다.

[0017] 본 발명의 다른 측면에 따르면, 프로세서 및 메모리를 포함하는 장치에서 기하학적 정보를 반영한 노이즈 제거 확산 모델 기반으로 이미지를 생성하는 방법으로서, 인코더-디코더 기반의 노이즈 제거 확산 모델을 이용하여 가우시안 노이즈를 반복적으로 제거하여 이미지를 생성하는 단계; 제1 시간 단계에서, 상기 노이즈 제거 확산 모델의 하나 이상의 디코더 레이어로부터 중간 특징을 추출하는 단계; 미리 학습된 다층 퍼셉트론 기반 깊이 예측기에 상기 중간 특징을 입력하여 깊이 맵을 생성하는 단계; 및 상기 깊이 맵을 통한 깊이 정보를 상기 노이즈 제거 확산 모델이 상기 제1 시간 단계에서 출력하는 샘플링 이미지에 주입하여 깊이 인식 이미지를 생성하는 단계를 포함하는 노이즈 제거 확산 모델 기반 이미지 생성 방법이 제공된다.

[0018] 본 발명의 또 다른 측면에 따르면, 상기한 방법을 수행하는 컴퓨터 판독 가능한 기록매체에 저장된 컴퓨터 프로그램이 제공된다.

## *발명의 효과*

[0019] 본 발명에 따르면, 기존에 존재하는 다양한 확산 모델의 네트워크와 손실함수는 그대로 유지하되, 깊이 정보를 이미지 추출 과정에서 주입시켜 이미지가 기하학적으로 신뢰도 있게 생성 가능한 장점이 있다.

[0020] 또한, 본 발명에 따르면, 확산 모델을 통한 다양성을 가진 이미지를 3차원 재구성 및 가상 데이터를 생성하는데 있어서 보다 안정성 있고 믿을 수 있는 영상의 생성이 가능하다는 장점이 있다.

## *도면의 간단한 설명*

[0021] 도 1은 본 발명의 바람직한 일 실시예에 따른 기하학적 정보를 반영한 노이즈 제거 확산 모델 기반 이미지 생성 프레임워크를 도시한 도면이다.

도 2는 본 실시예에 따른 프레임워크의 샘플링 과정을 시각화한 도면이다.

도 3은 본 실시예에 따른 깊이 예측 성능의 정량적 비교를 나타낸 도면이다.

도 4는 LSUN 침실 데이터세트에 대한 정성적 비교를 나타낸 도면이다.

## *발명을 실시하기 위한 구체적인 내용*

[0022] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세한 설명에 상세하게 설명하고자 한다. 그러나 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.

[0023] 본 명세서에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0024] 또한, 각 도면을 참조하여 설명하는 실시예의 구성 요소가 해당 실시예에만 제한적으로 적용되는 것은 아니며, 본 발명의 기술적 사상이 유지되는 범위 내에서 다른 실시예에 포함되도록 구현될 수 있으며, 또한 별도의 설명이 생략될지라도 복수의 실시예가 통합된 하나의 실시예로 다시 구현될 수도 있음은 당연하다.

[0025] 또한, 첨부 도면을 참조하여 설명함에 있어, 도면 부호에 관계없이 동일한 구성 요소는 동일하거나 관련된 참조 부호를 부여하고 이에 대한 중복되는 설명은 생략하기로 한다. 본 발명을 설명함에 있어서 관련된 공지 기술에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우 그 상세한 설명을 생략한다.

[0026] 본 실시예는 기하학적 인식의 부족으로 이미지가 왜곡되는 문제를 해결하기 위해, 이미지 생성 과정에서 기하학적 정보를 노이즈 제거 확산 모델에 통합하는 새로운 프레임워크를 제안한다.

[0027] 본 실시예에서는 깊이 레이블이 지정된 데이터에 대한 학습을 통해 깊이 맵을 사용하여 깊이 인식 이미지(depth-aware image) 생성을 유도한다. 또한, 본 실시예에서는 깊이 인식 이미지 생성을 더욱 향상시키기 위한 샘플링 방법을 제안한다.

[0029] 이하에서는 도면을 참조하여 본 실시예에 따른 깊이 인식 이미지 생성 과정을 상세하게 설명한다.

[0030] 노이즈 제거 확산 모델은 가우시안 노이즈를 반복적으로 디노이징하여 이미지를 생성하는 생성 모델이다.

[0031] 구체적으로, 입력 이미지 $\mathbf{x}_0$ 와 입력 노이즈 $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ 가 주어지면 확산 모델은 시간 단계 t에 따라 점진적으로 노이즈를 추가한다.

[0032] $t \in \{T, T-1, \ldots, 1\}$ 에 대한 미리 정의된 분산 스케줄 $\beta_t$ 에 대해, $\alpha_t$ 와 $1-\beta_t$ 로서 $\bar{\alpha}_t$ 및 $\Pi_{k=1}^{t}\alpha_k$ 를 각각 나타내고, 임의의 시간 t에서 순방향 프로세스는 닫힌 형식으로 표현할 수 있다.

*수학식 1*

[0033] $$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon.$$

[0034] 확산 모델의 훈련 목표를 다음과 같이 재가중할 수 있다.

*수학식 2*

[0035] $$\mathcal{L}_{\text{simple}} = \|\epsilon - \epsilon_\theta(\mathbf{x}_t)\|_2^2,$$

[0036] 여기서, $\epsilon_\theta$ 는 $\theta$ 로 파라미터화된 노이즈 제거 확산 모델(U-Net)이다. 결과적으로 확산 모델의 샘플링 과정은 순방향 프로세스의 역순으로 정의된다.

*수학식 3*

[0037] $$\mathbf{x}_{t-1} \sim \mathcal{N}(\mu_\theta(\mathbf{x}_t), \sigma_t^2),$$

[0038] 여기서, $\sigma_t^2$ 는 t에만 의존하는 상수이며, $\mu_\theta$ 는 재-파라미터화에서 계산된 평균의 예측이다.

[0039] 분산도 다음의 수학식을 통해 예측될 수 있고, 수학식 3은 다음과 같이 변경된다.

*수학식 4*

[0040] $$\mathbf{x}_{t-1} \sim \mathcal{N}(\mu_\theta(\mathbf{x}_t), \Sigma_\theta(\mathbf{x}_t)),$$

[0041] 여기서, $\Sigma_\theta$ 는 모델 출력에서 역 과정의 분산 예측이다.

[0042] 확산 모델을 사용하여 조건부 이미지 생성을 위해서는 유도(guidance) 방법의 필요성이 증가하고 있다.

[0043] 훈련된 분류기 $p_\psi(\mathbf{y}|\mathbf{x}_t)$ 는 노이즈 이미지 $\mathbf{x}_t$ 를 입력으로 하여 클래스 $\mathbf{y}$ 를 분류하고, $\mathbf{x}_t$ 에 대한 그래디언트는 다음과 같이 계산된다.

*수학식 5*

[0044] $$\nabla_{\mathbf{x}_t} \log(p_\psi(\mathbf{y}|\mathbf{x}_t)).$$

[0045] 그런 다음, 분류기 유도가 있는 샘플링 과정은 다음과 같이 정의된다.

*수학식 6*

[0046] $$\mathbf{x}_{t-1} \sim \mathcal{N}(\mu_\theta(\mathbf{x}_t) + w\Sigma\nabla_{\mathbf{x}_t} \log(p_\psi(\mathbf{y}|\mathbf{x}_t)), \Sigma_\theta(\mathbf{x}_t)).$$

[0047] 여기서, $w > 0$ 는 유도 척도이다.

[0048] 상기한 방법이 샘플링 품질을 향상시키는 하나, 기하학적 인식이 부족하여 원근감이 없거나 레이아웃 왜곡을 포함하는 경우가 많다.

[0049] 본 실시예에 따르면, 기하학적 정보를 확산 모델에 명시적으로 통합하기 위해 새로운 프레임워크를 제안한다.

[0050] 도 1은 본 발명의 바람직한 일 실시예에 따른 기하학적 정보를 반영한 노이즈 제거 확산 모델 기반 이미지 생성 프레임워크를 도시한 도면이다.

[0051] 도 1에 전체적인 과정은 프로세서 및 메모리를 포함하는 장치에서 수행될 수 있다.

[0052] 여기서, 프로세서는 컴퓨터 프로그램을 실행할 수 있는 CPU(central processing unit)나 그 밖에 가상 머신 등을 포함할 수 있다.

[0053] 메모리는 고정식 하드 드라이브나 착탈식 저장 장치와 같은 불휘발성 저장 장치를 포함할 수 있다. 착탈식 저장 장치는 컴팩트 플래시 유닛, USB 메모리 스틱 등을 포함할 수 있다. 메모리는 각종 랜덤 액세스 메모리와 같은 휘발성 메모리도 포함할 수 있다.

[0054] 본 실시예에 메모리에는 프로그램 명령어들이 저장되며, 프로그램 명령어들은, 인코더-디코더 기반의 노이즈 제거 확산 모델을 이용하여 가우시안 노이즈를 반복적으로 제거하여 이미지를 생성하고, 제1 시간 단계에서, 상기 노이즈 제거 확산 모델의 디코더에 포함되는 하나 이상의 레이어로부터 중간 특징을 추출하고, 미리 학습된 다층 퍼셉트론 기반 깊이 예측기에 상기 중간 특징을 입력하여 깊이 맵을 생성하고, 상기 깊이 맵을 통한 깊이 정보를 상기 노이즈 제거 확산 모델이 상기 제1 시간 단계에서 출력하는 샘플링 이미지에 주입하여 깊이 인식 이미지를 생성한다.

[0055] 구체적으로 깊이 인식 이미지(depth-aware image)를 생성하기 위해 소량의 깊이 레이블 데이터를 갖는 내부 표현(internal representation)을 사용하여 다층 퍼셉트론 기반 깊이 예측기(depth predictor)를 훈련한다. 그런

다음 샘플링 과정에서 얻은 깊이 맵을 사용하여 중간 이미지를 유도한다.

[0056] 이하에서는, 레이블-효율 방식(label-efficient way)으로 깊이 맵을 획득하는 방법과 생성된 이미지가 깊이 인식을 갖도록 유도하는 샘플링 방법을 설명한다.

[0057] 먼저, 깊이 예측기의 레이블-효율(label-efficient) 방식을 설명한다.

[0058] 간단한 방식으로 확산 모델을 통해 깊이 인식 이미지를 생성하려면 많은 양의 이미지 깊이 쌍 또는 노이즈가 있는 이미지에 대해 훈련된 외부 대규모 깊이 추정 네트워크가 필요하다.

[0059] 이러한 문제를 해결하기 위해 본 실시예에서는 깊이를 추정하기 위해 이미지의 깊이 정보를 포함할 수 있는 노이즈 제거 확산 모델로 학습한 풍부한 표현을 재사용한다.

[0060] 확산 모델로 훈련된 네트워크의 내부 특징이 의미론적 정보를 인코딩할 수 있음을 고려하여, 프레임워크에 깊이 정보를 통합한다.

[0061] 이하에서는, 본 실시예에 따른 확산 모델이 U-Net인 것으로 하여 설명한다.

[0062] 레이블-효율 방식으로 깊이 추정을 수행하기 위해 U-Net으로부터 중간 특징(intermediate feature)을 수신하고 노이즈가 많은 입력 이미지의 깊이 값을 추정하는 픽셀 단위의 얕은 다층 퍼셉트론 (shallow Multi-Layer Perceptron) 리그레서(regressor)를 사용한다.

[0063] 특히 확산 U-Net의 k번째 디코더 레이어로부터 중간 특징 $\mathbf{f}_t(k) \in \mathbb{R}^{C(k)\times H(k)\times W(k)}$ 을 추출한다.

[0064] 여기서 $C(k)$ 는 채널 차원, $H(k)\times W(k)$ 는 U-Net k번째 디코더 레이어의 공간 해상도를 나타낸다.

[0065] 그런 다음 깊이 예측기 내의 MLP 블록을 픽셀 단위로 쿼리하여 깊이 맵을 생성한다.

[0066] 여기서 깊이 맵은 다음과 같이 공식화할 수 있다.

*수학식 7*

[0067] $$\mathbf{d}_t(k) = \mathrm{MLP}(\mathbf{f}_t(k))$$

[0068] 또한 일반적으로 한 레이어의 특징만 사용하는 것보다 다른 U-Net 레이어의 특징을 더 많이 사용하는 것이 좋다. 따라서 본 실시예에서는 많은 레이어에서 더 많은 특징을 추출하고 다음과 같이 공식화할 수 있는 채널 차원에서 연결한다.

*수학식 8*

[0069] $$\mathbf{g}_t = [\mathbf{f}_t(1); \mathbf{f}_t(2); \cdots; \mathbf{f}_t(d)],$$

[0070] 여기서 $[\cdot;\cdot]$ 는 중간 특징 사이의 채널별 연결(concatenation operation)이며, d는 선택한 디코더 레이어의 총 개수이다. 그런 다음 이들을 픽셀별 다층 퍼셉트론 기반 깊이 예측기에 전달하고 수학식 8의 관점에서 수학식 7의 일반화를 수행한다.

*수학식 9*

[0071] $$\mathbf{d}_t = \mathrm{MLP}(\mathbf{g}_t).$$

[0072] 임의의 시간 단계에서 깊이 맵을 예측하기 위해 추가 시간 포함 블록(시간 임베딩 모듈)을 추가하여 픽셀별 깊이 예측기를 수정한다. 따라서 본 실시예에서는 전체 샘플링 과정에서 깊이 예측기를 사용할 수 있다. 본 실시

예에서는 확산 U-Net의 시간 임베딩 모듈을 이용하며, 수학식 7을 다음과 같이 표현할 수 있다.

*수학식 10*

[0073] $$\mathbf{d}_t = \mathrm{MLP}(\mathbf{g}_t, t).$$

[0074] 본 실시예서는 아래와 같은 L1 손실이 있는 ground-truth 깊이 맵 y를 사용하여 확산 U-Net의 고정 특징(frozen feature)으로만 깊이 예측기를 훈련한다.

*수학식 11*

[0075] $$\mathcal{L}_{\text{depth}} = \|\mathbf{d}_t - \mathbf{y}\|_1.$$

[0076] 생성된 이미지가 그럴듯한 깊이 맵을 생성하도록 하기 위해 샘플링 프로세스 중에 예측된 깊이 맵이 정확해야 하며, 이를 위해 본 실시예에서는 두 가지 유도 기법을 사용한다.

[0077] 유도 방정식의 일반적인 형태는 다음과 같이 공식화된다.

*수학식 12*

[0078] $$\mathbf{x}_{t-1} \sim \mathcal{N}(\mu_\theta(\mathbf{x}_t) - \omega\nabla_{\mathbf{x}_t}\mathcal{L}_{\text{depth}}, \Sigma_\theta(\mathbf{x}_t)).$$

[0079] 하나는, 일관성 정규화(consistency regularization)를 통한 강한 깊이 예측을 약한 예측이 따라가도록 하는 학습 방향을 통해 이미지에 깊이 정보를 주입하는 것이고, 다른 하나는 깊이 정보를 학습한 확산 모델의 출력을 우선 정보로 활용하는 것이다.

[0080] 이를 통해 깊이 정보를 이미지 샘플링 과정에서 깊이 정보를 주입할 수 있게 된다.

[0081] 일관성 정규화 기법은 강한 깊이 예측을 약한 예측이 따라가도록 하는 것으로서, 더 많은 특징 블록의 집합에서 얻은 예측이 적은 특징 블록에서 얻은 예측보다 더 유익한 것으로 간주한다.

[0082] 예를 들어, 복수의 디코더 레이어 중 6번째 레이어의 특징, $\mathbf{g}^W = [\mathbf{f}_t(6)]$ 을 사용하여 얻어지는 예측을 약한 예측이라 정의하고, 2, 4, 5, 6 및 7번째 레이어 집합의 특징, $\mathbf{g}^S = [\mathbf{f}_t(2); \mathbf{f}_t(4); \mathbf{f}_t(5); \mathbf{f}_t(6); \mathbf{f}_t(7)]$ 을 사용하여 얻어지는 예측을 강한 예측이라 정의한다.

[0083] 상기한 같이 서로 다른 특징의 서로 다른 채널 차원을 설명하기 위해 MLP-S 및 MLP-W라는 두 가지 비대칭 깊이 예측기를 제안한다.

[0084] 즉, 제1 깊이 예측기(MLP-S, 강한 분기 예측기)는 U-Net 블록에서 더 많은 중간 특징을 입력으로 하는 반면 제2 깊이 예측기(MLP-W, 약한 분기 예측기)는 더 적은 중간 특징을 입력으로 하여 훈련한다.

*수학식 13*

[0085] $$\mathbf{d}_t^S = \text{MLP-S}(\mathbf{g}_t^S, t), \quad \mathbf{d}_t^W = \text{MLP-W}(\mathbf{g}_t^W, t).$$

[0086] 본 실시예에 따르면, $\mathbf{d}_t^S$ 를 의사 레이블(pseudo-label)로 하고 $\mathbf{d}_t^W$ 를 예측으로 하여 두 개의 예측된

밀도 맵 사이의 일관성 손실을 이용하여 손실을 계산한다.

[0087] 도 2는 본 실시예에 따른 프레임워크의 샘플링 과정을 시각화한 도면이다.

[0088] 도 2에서 위에서 아래로 강한 분기 예측과 약한 분기 예측을 나타내며, 도 2를 참조하면, 강한 분기 예측기는 초기 단계에서 강건한 깊이 예측을 제공하며, 이에 따라 의사 레이블로 사용할 수 있다.

[0089] 본 실시예에서는 강한 예측이 약한 예측을 학습하는 것을 방지하기 위해 강한 특징에 의한 깊이 맵에 스탑 그래디언트(stopgradient)를 적용한다.

[0090] x에 대한 손실 그래디언트는 확산 모델을 통해 이동하고 해당 시간 단계에서의 샘플링 이미지에 주입되어 깊이 인식 이미지가 생성되도록 한다.

[0091] 이러한 과정은 다음과 같이 공식화될 수 있다.

*수학식 14*

[0092] $$\mathcal{L}_{\mathrm{dc}} = \|\mathrm{stopgrad}(\mathbf{d}_t^S) - \mathbf{d}_t^W\|_2^2,$$

[0093] 여기서, stopgrad는 스탑 그래디언트 적용을 나타내고, 수학식 12와 함께 상기한 그래디언트를 사용하여 깊이 인식 이미지 생성을 유도할 수 있다.

*수학식 15*

[0094] $$\mathbf{x}_{t-1} \sim \mathcal{N}(\mu_\theta(\mathbf{x}_t) - \omega_{\mathrm{dc}}\nabla_{\mathbf{x}_t}\mathcal{L}_{\mathrm{dc}}, \Sigma_\theta(\mathbf{x}_t))$$

[0095] 여기서, $\omega_{\mathrm{dc}}$ 는 유도 척도를 나타낸다.

[0096] 도 3은 본 실시예에 따른 깊이 예측 성능의 정량적 비교를 나타낸 도면이다.

[0097] 도 3에서 x축은 시간 단계를 나타내고, y축은 깊이 추정의 정확도를 나타낸다.

[0098] 도 3의 (a)는 서로 다른 U-Net 블록에 따른 깊이 예측 정확도를 비교한 것이고, 각 라인은 중간 특징이 추출된 디코더 레이어를 나타내고, (b)는 훈련에 표시된 이미지 수와 깊이 예측 정확도를 나타낸 것이다.

[0099] 이하에서는 두 번째 유도 기법인 깊이 사전 유도(Depth prior guidance)를 설명한다.

[0100] 사전 훈련된 확산 모델은 노이즈가 있는 분포를 현실적인 분포로 효과적으로 정제할 수 있거나 확산 모델의 지식을 활용하여 실제 데이터와 일치하도록 데이터의 노이즈가 있는 초기화를 최적화하는 데 도움이 될 수 있다.

[0101] 본 실시예에서는 깊이 도메인에서 작은 해상도 확산 모델(U-Net)을 훈련하고 두 번째 안내 방법에 대한 우선 정보로 사용한다.

[0102] 이미지 생성을 위한 확산 모델 U-Net의 하나 이상의 디코더 레이어에서 중간 특징을 추출하여 다층 퍼셉트론 기반 깊이 예측기를 사용하여 해당 깊이 맵을 추정할 수 있다.

[0103] 이전 확산 모델을 활용하여 다음과 같은 순방향 프로세스를 사용하여 깊이 예측 $\mathbf{d}_0^S$ 에 노이즈를 주입한다.

*수학식 16*

[0104] $$\mathbf{d}_\tau^S = \sqrt{\bar{\alpha}_\tau}\mathbf{d}_0^S + \sqrt{1-\bar{\alpha}_\tau}\eta, \quad \eta \sim \mathcal{N}(\mathbf{0}, \mathbf{I}),$$

[0105] 여기서, $\tau$ 는 이전 확산 모델에서 사용되는 시간 단계이다.

[0106] 깊이 예측에 노이즈를 추가한 후, 추가된 노이즈를 추정하기 위해 이전 네트워크에 공급한다.

[0107] 그런 다음 추가된 노이즈와 x에 대한 예측된 노이즈 사이의 평균 제곱 오차의 그래디언트를 계산한다.

[0108] 이는 다음과 같이 정의된다.

*수학식 17*

[0109] $$\mathcal{L}_{\mathrm{dp}} = \|\eta - \epsilon_{\phi}(\mathbf{d}_{\tau}^{S}))\|_{2}^{2}.$$

[0110] 상기한 손실의 그래디언트를 외부 분류기 그래디언트로 간주하면 생성된 이미지가 이전의 깊이와 일치하도록 유도할 수 있다.

*수학식 18*

[0111] $$\mathbf{x}_{t-1} \sim \mathcal{N}(\mu_{\theta}(\mathbf{x}_{t}) - \omega_{\mathrm{dp}}\nabla_{\mathbf{x}_{t}}\mathcal{L}_{\mathrm{dp}}, \Sigma_{\theta}(\mathbf{x}_{t})),$$

[0112] 여기서, $\omega_{\mathrm{dp}}$ 는 유도 척도를 나타낸다.

[0113] 제안된 DCG와 DPG를 통합함으로써 두 가지를 모두 사용하여 샘플링된 이미지를 유도할 수 있다. 수학식 15 및 18에서 전체 샘플링은 다음과 같이 나타낼 수 있다.

*수학식 19*

[0114] $$\mathbf{x}_{t-1} \sim \mathcal{N}(\mu_{\theta}(\mathbf{x}_{t}) - \omega_{\mathrm{dc}}\nabla_{\mathbf{x}_{t}}\mathcal{L}_{\mathrm{dc}} - \omega_{\mathrm{dp}}\nabla_{\mathbf{x}_{t}}\mathcal{L}_{\mathrm{dp}}, \Sigma_{\theta}(\mathbf{x}_{t})).$$

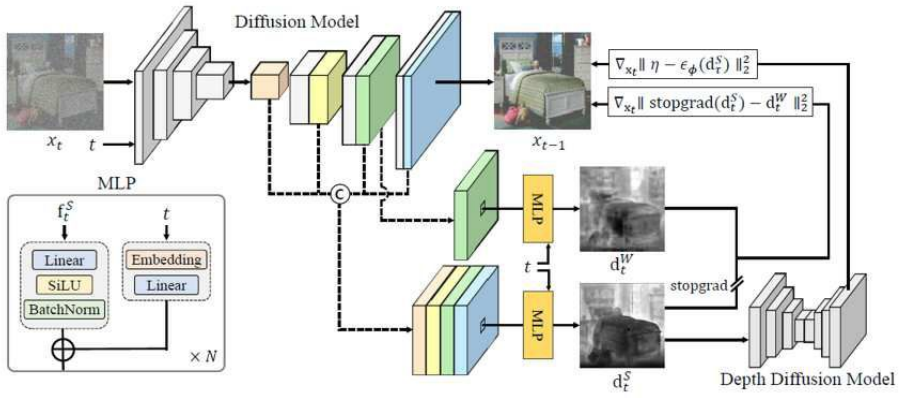
[0115] 도 4는 LSUN 침실 데이터세트에 대한 정성적 비교를 나타낸 도면이다.

[0116] 도 4의 (a), (c), (e), (g)는 유도 없이 생성한 이미지를 나타낸 것이고, (b), (d), (f), (h)는 깊이 정보의 주입을 통해 생성된 이미지를 나타낸다.

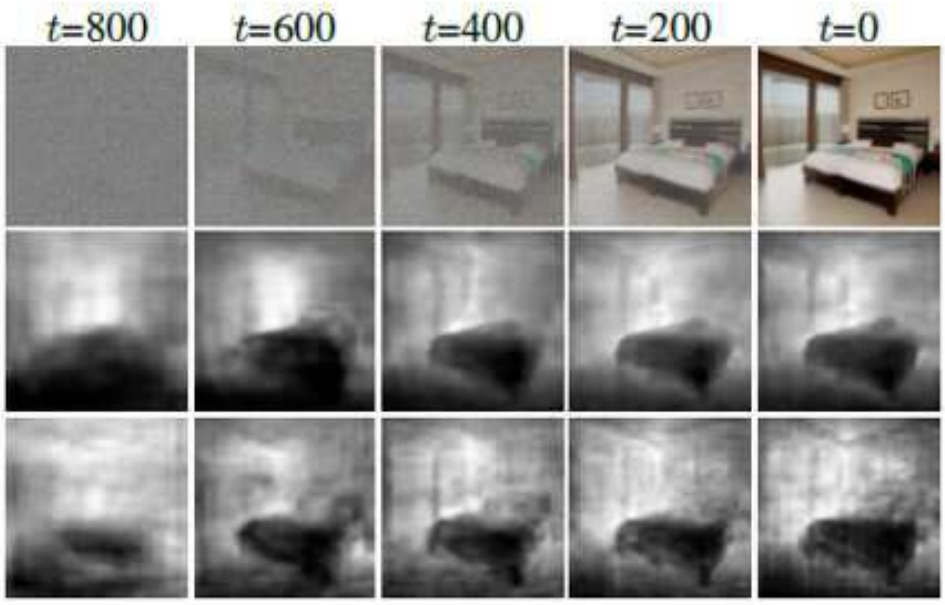
[0117] 상기한 본 발명의 실시예는 예시의 목적을 위해 개시된 것이고, 본 발명에 대한 통상의 지식을 가지는 당업자라면 본 발명의 사상과 범위 안에서 다양한 수정, 변경, 부가가 가능할 것이며, 이러한 수정, 변경 및 부가는 하기의 특허청구범위에 속하는 것으로 보아야 할 것이다.

## 도면

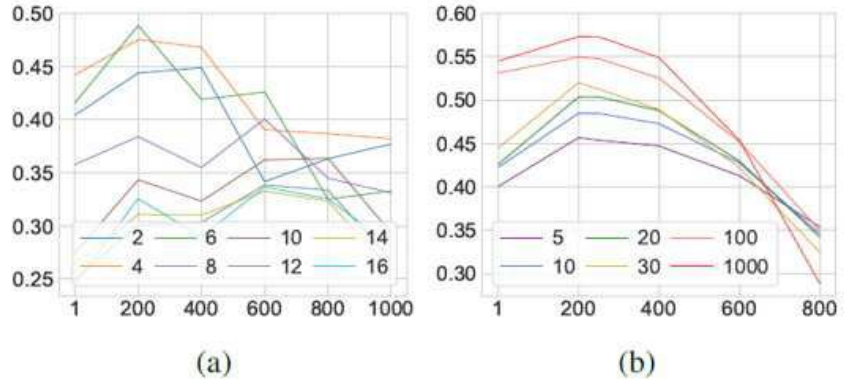
### 도면1


Diffusion Model
$x_t$
$t$
MLP
$f_t^S$
$t$
Linear
SiLU
BatchNorm
Embedding
Linear
× N
$x_{t-1}$
$\nabla_{x_t} \| \eta - \epsilon_\phi(d_t^S) \|_2^2$
$\nabla_{x_t} \| \text{stopgrad}(d_t^S) - d_t^W \|_2^2$
MLP
MLP
$t$
$d_t^W$
$d_t^S$
stopgrad
Depth Diffusion Model

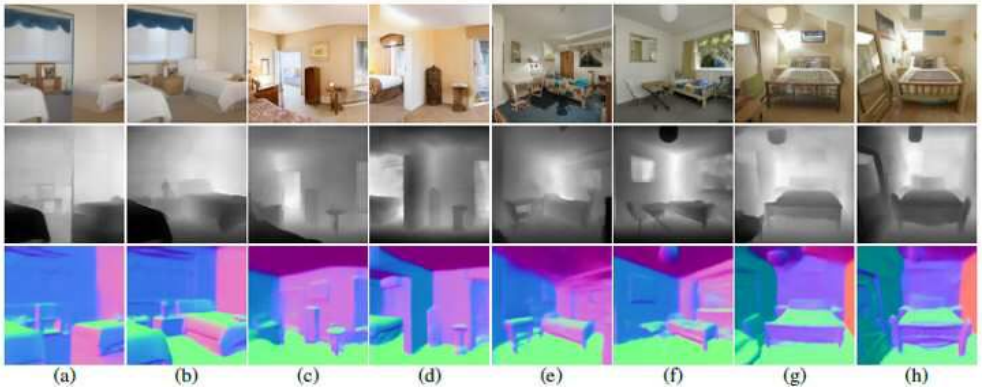
### 도면2


t=800
t=600
t=400
t=200
t=0

도면3


0.50
0.45
0.40
0.35
0.30
0.25
2
4
6
8
10
12
14
16
1
200
400
600
800
1000
(a)
0.60
0.55
0.50
0.45
0.40
0.35
0.30
5
10
20
30
100
1000
1
200
400
600
800
(b)

도면4


(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)