



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0134643
(43) 공개일자 2024년09월10일

(51) 국제특허분류(Int. Cl.)
G06T 11/60 (2006.01) G06N 3/08 (2023.01)
G06T 7/11 (2017.01)
(52) CPC특허분류
G06T 11/60 (2013.01)
G06N 3/08 (2023.01)
(21) 출원번호 10-2023-0027952
(22) 출원일자 2023년03월02일
심사청구일자 2023년03월02일

(71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
김현우
서울시 성북구 북악산로 813 (정릉동, 정릉우성아파트) 101동 809호
고주연
경기도 안양시 동안구 귀인로 213 향촌현대5차아파트 102동 1302호
공인호
경기도 용인시 수지구 정평로 61, 501동 1003호
(74) 대리인
특허법인위솔

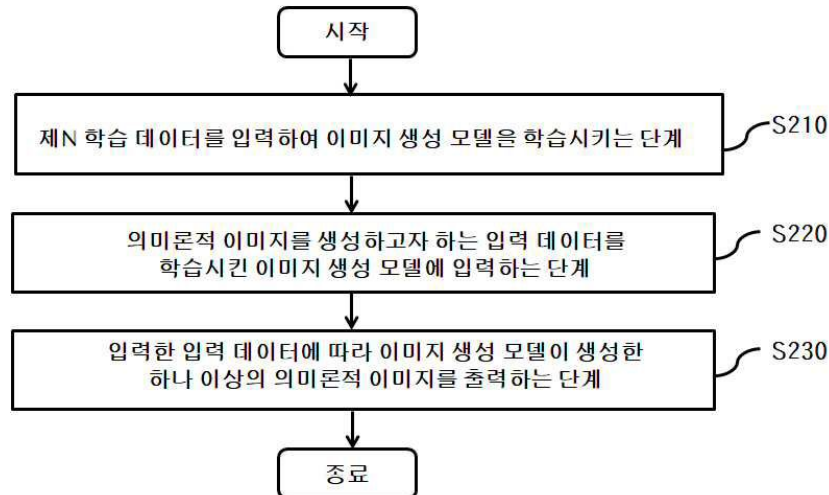
전체 청구항 수 : 총 12 항

(54) 발명의 명칭 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법 및 이를 위한 장치

(57) 요약

본 발명의 일 실시 예에 따른 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법은 (a) 제N(N은 임의의 양의 정수) 학습 데이터를 입력하여 이미지 생성 모델을 학습시키는 단계, (b) 의미론적 이미지를 생성하고자 하는 입력 데이터를 상기 학습시킨 이미지 생성 모델에 입력하는 단계 및 (c) 상기 입력한 입력 데이터에 따라 상기 이미지 생성 모델이 생성한 하나 이상의 의미론적 이미지를 출력하는 단계를 포함하고, 상기 입력 데이터는, 입력 조건 데이터인 조건 이미지 틀(Layout)을 포함하며, 상기 조건 이미지 틀은, 생성하고자 하는 의미론적 이미지가 포함하는 하나 이상의 대상인 클래스(Class)가 대상별로 각각의 카테고리로 분류되며, 상기 클래스가 차지하고 있는 픽셀 영역에 상기 분류된 카테고리 별로 상이한 숫자가 부여된 이미지 틀이다.

대표도 - 도2



(52) CPC특허분류

G06T 7/11 (2017.01)

G06T 2207/20081 (2013.01)

G06T 2207/20084 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1345360941
과제번호	4199990214667
부처명	교육부
과제관리(전문)기관명	한국연구재단
연구사업명	4단계두뇌한국21사업(R&D)
연구과제명	데이터주도과학 미래인재 교육연구단
기 여 율	1/1
과제수행기관명	고려대학교
연구기간	2022.03.01 ~ 2023.02.28

명세서

청구범위

청구항 1

프로세서 및 메모리를 포함하는 장치가 의미론적 이미지(Semantic Image)를 생성하는 방법에 있어서,

- (a) 제N(N은 임의의 양의 정수) 학습 데이터를 입력하여 이미지 생성 모델을 학습시키는 단계;
- (b) 의미론적 이미지를 생성하고자 하는 입력 데이터를 상기 학습시킨 이미지 생성 모델에 입력하는 단계; 및
- (c) 상기 입력한 입력 데이터에 따라 상기 이미지 생성 모델이 생성한 하나 이상의 의미론적 이미지를 출력하는 단계;

를 포함하는 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법에 있어서,

상기 입력 데이터는,

입력 조건에 해당하는 조건 이미지 틀(Layout)이며,

상기 조건 이미지 틀은,

생성하고자 하는 의미론적 이미지가 포함하는 하나 이상의 대상인 클래스(Class)가 대상별로 각각의 카테고리로 분류되며, 상기 클래스가 차지하고 있는 픽셀 영역에 상기 분류된 카테고리 별로 상이한 숫자가 부여된,

조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법.

청구항 2

제1항에 있어서,

상기 (a) 단계는,

(a-1) 상기 제N 학습 데이터에 제1 노이즈 추가 방식 내지 제4 노이즈 추가 방식 중, 어느 하나의 선택을 통해 노이즈를 추가한 제N' 학습 데이터를 생성하는 단계;

(a-2) 상기 생성한 제N' 학습 데이터를 이미지 생성 모델에 입력하여 상기 제N' 학습 데이터에 적용된 노이즈를 예측하는 단계; 및

(a-3) 상기 (a-1) 단계에서 추가한 노이즈와 상기 (a-2) 단계에서 예측한 노이즈를 비교하여 그 차이를 손실함수로 설정하고, 경사 하강법(Gradient Descent)를 적용하여 학습하는 단계;

를 포함하는 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법.

청구항 3

제2항에 있어서,

상기 (a-3) 단계 이후에,

(a-4) 상기 설정한 손실함수가 최소값에 수렴하는지 판단하는 단계;

(a-5) 상기 (a-4) 단계의 판단 결과 NO라면 상기 (a-1) 단계로 회귀하는 단계; 및

(a-6) 상기 (a-4) 단계의 판단 결과 YES라면, 상기 이미지 생성 모델에 대한 학습을 종료하는 단계;

를 더 포함하는 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법.

청구항 4

제2항에 있어서,

상기 제N 학습 데이터는,

제N 실제 이미지이며,

이 경우, 상기 제1 노이즈 추가 방식은,

하기 수학적 식 1에 따라 이전 시간(Timestep)과 연관성이 부존재하는 노이즈 추가 방식인,

조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법.

수학적 식 1: $q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1 - \bar{\alpha}_t)\mathbf{I})$

여기서 $\mathbf{x}_t$ 는 임의의 시점 t 에서 노이즈가 추가된 제N 실제 이미지, $\mathbf{x}_0$ 은 노이즈가 추가되기 전, 최초 상태의 제N 실제 이미지, $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t := \prod_{s=1}^t \alpha_s$, β_t 는 임의의 시점 t 에서의 노이즈 스케줄, $\mathbf{I}$ 는 단위 행렬을 의미함

청구항 5

제2항에 있어서,

상기 제N 학습 데이터는,

제N 실제 이미지이며,

이 경우, 상기 제2 노이즈 추가 방식은,

하기 수학적 식 2에 따라 이전 시간(Timestep)과 직접적인 연관성이 존재하는 노이즈 추가 방식인,

조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법.

수학적 식 2: $q(\mathbf{x}_t|\mathbf{x}_{t-1}) := \mathcal{N}(\mathbf{x}_t; \sqrt{1 - \beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I})$

여기서 $\mathbf{x}_t$ 는 임의의 시점 t 에서 노이즈가 추가된 제N 실제 이미지, $\mathbf{x}_{t-1}$ 은 임의의 시점 t 직전의 시점인 $t-1$ 에서 노이즈가 추가된 제N 실제 이미지, β_t 는 임의의 시점 t 에서의 노이즈 스케줄, $\mathbf{I}$ 는 단위 행렬을 의미함 을 의미함

청구항 6

제2항에 있어서,

상기 제N 학습 데이터는,

제N 실제 이미지이며,

이 경우, 상기 제3 노이즈 추가 방식은,

하기 수학적 식 3에 따라 이전 시간(Timestep)과 직접적인 연관성이 존재하는 노이즈 추가 방식인,

조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법.

수학적 식 3: $q(\mathbf{x}_t|\mathbf{x}_{t-1}) := \mathcal{N}(\mathbf{x}_t; \sqrt{1 - \beta_t}\mathbf{x}_{t-1}, 0)$

여기서 $\mathbf{x}_t$ 는 임의의 시점 t 에서 노이즈가 추가된 제N 실제 이미지, $\mathbf{x}_{t-1}$ 은 임의의 시점 t 직전의 시점인 $t-1$ 에서 노이즈가 추가된 제N 실제 이미지, β_t 는 임의의 시점 t 에서의 노이즈 스케줄, $\mathbf{I}$ 는 단위 행렬을 의미함 을 의미함

청구항 7

제2항에 있어서,

상기 제N 학습 데이터는,

제N 실제 이미지이며,

이 경우, 상기 제4 노이즈 추가 방식은,

상기 수학적 식 4에 따라 이전 시간(Timestep)과 직접적인 연관성이 존재하는 노이즈 추가 방식인,

조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법.

수학적식4:
$$q_{\sigma}(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) = \mathcal{N}(\sqrt{\alpha_{t-1}}\mathbf{x}_0 + \sqrt{1 - \alpha_{t-1} - \sigma_t^2} \cdot \frac{\mathbf{x}_t - \sqrt{\alpha_t}\mathbf{x}_0}{\sqrt{1 - \alpha_t}}, \sigma_t^2\mathbf{I})$$

여기서 $\mathbf{x}_{t-1}$ 은 임의의 시점 t 직전의 시점인 $t-1$ 에서 노이즈가 추가된 제N 실제 이미지, $\mathbf{x}_t$ 는 임의의 시점 t 에서

노이즈가 추가된 제N 실제 이미지, $\mathbf{x}_0$ 은 노이즈가 추가되기 전, 최초 상태의 제N 실제 이미지, $\alpha_t = 1 - \beta_t$,

$\bar{\alpha}_t := \prod_{s=1}^t \alpha_s$, β_t 는 임의의 시점 t 에서의 노이즈 스케줄, σ 는 각 시간 t 마다 설정한 표준편차, $\mathbf{I}$ 는 단위 행렬을 의미함

청구항 8

제2항에 있어서,

상기 제N 학습 데이터는,

제N 실제 이미지이며,

이 경우, 상기 (a-2) 단계는,

(a-2-1) 상기 생성한 제N' 학습 데이터를 이미지 생성 모델에 입력하는 단계;

(a-2-1) 상기 제N 실제 이미지에 대응되는 제N 조건 이미지 틀에 제1 노이즈 추가 방식 내지 제4 노이즈 추가 방식 중, 어느 하나를 통해 노이즈를 추가한 제N 입력 조건 데이터를 상기 이미지 생성 모델에 입력하는 단계; 및

(a-2-3) 상기 입력한 제N 입력 조건 데이터를 반영하여 상기 제N' 학습 데이터에 적용된 노이즈를 예측하는 단계;

를 포함하는 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법.

청구항 9

제2항에 있어서,

상기 (b) 단계는,

(b-1) 상기 생성하고자 하는 의미론적 이미지와 동일한 사이즈의 노이즈 데이터를 랜덤하게 생성하여 상기 입력 데이터로 상기 이미지 생성 모델에 입력하는 단계; 및

(b-2) 상기 의미론적 이미지를 생성하고자 하는 조건 이미지 틀에 제1 노이즈 추가 방식 내지 제4 노이즈 추가 방식 중, 어느 하나를 통해 노이즈를 추가한 입력 조건 데이터'를 생성하여 상기 이미지 생성 모델에 입력하는 단계;

를 포함하는 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법.

청구항 10

제1항에 있어서,

상기 이미지 생성 모델은,

인코더부 및 디코더부를 포함하며,

조건 이미지 틀로부터 특성값을 획득하고, 이를 통과시켜 스케일값 및 바이어스값을 산출하는 하나 이상의 컨볼

루션 레이어를 포함하는 조건 이미지 틀 활용부;

를 더 포함하고,

상기 산출한 스케일값 및 바이어스값은,

상기 입력 데이터로부터 획득하고 배치 정규화를 진행한 특성값에 입력되어 연산되는,

조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법.

청구항 11

하나 이상의 프로세서;

네트워크 인터페이스;

상기 프로세서에 의해 수행되는 컴퓨터 프로그램을 로드(Load)하는 메모리; 및

대용량 네트워크 데이터 및 상기 컴퓨터 프로그램을 저장하는 스토리지를 포함하되,

상기 컴퓨터 프로그램은 상기 하나 이상의 프로세서에 의해,

(A) 제N(N은 임의의 양의 정수) 학습 데이터를 입력하여 이미지 생성 모델을 학습시키는 오퍼레이션;

(B) 의미론적 이미지를 생성하고자 하는 입력 데이터를 상기 학습시킨 이미지 생성 모델에 입력하는 오퍼레이션; 및

(C) 상기 입력한 입력 데이터에 따라 상기 이미지 생성 모델이 생성한 하나 이상의 의미론적 이미지를 출력하는 오퍼레이션;

을 포함하는 의미론적 이미지 생성 장치에 있어서,

상기 입력 데이터는,

입력 조건 데이터인 조건 이미지 틀(Layout)을 포함하며,

상기 조건 이미지 틀은,

생성하고자 하는 의미론적 이미지가 포함하는 하나 이상의 대상인 클래스(Class)가 대상별로 각각의 카테고리 분류되며, 상기 클래스가 차지하고 있는 픽셀 영역에 상기 분류된 카테고리 별로 상이한 숫자가 부여된,

조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 장치.

청구항 12

컴퓨팅 장치와 결합하여,

(AA) 제N(N은 임의의 양의 정수) 학습 데이터를 입력하여 이미지 생성 모델을 학습시키는 단계;

(BB) 의미론적 이미지를 생성하고자 하는 입력 데이터를 상기 학습시킨 이미지 생성 모델에 입력하는 단계; 및

(CC) 상기 입력한 입력 데이터에 따라 상기 이미지 생성 모델이 생성한 하나 이상의 의미론적 이미지를 출력하는 단계;

를 실행하며,

상기 입력 데이터는,

입력 조건 데이터인 조건 이미지 틀(Layout)을 포함하며,

상기 조건 이미지 틀은,

생성하고자 하는 의미론적 이미지가 포함하는 하나 이상의 대상인 클래스(Class)가 대상별로 각각의 카테고리 분류되며, 상기 클래스가 차지하고 있는 픽셀 영역에 상기 분류된 카테고리 별로 상이한 숫자가 부여된,

컴퓨터로 판독 가능한 매체에 저장된 컴퓨터 프로그램.

발명의 설명

기술 분야

[0001] 본 발명은 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법 및 이를 위한 장치에 관한 것이다. 보다 자세하게는 이미지 틀만을 입력 받은 인공지능 기반의 이미지 생성 모델이 해당 이미지 틀에 부합하는 고품질 이미지를 다양하게 생성할 수 있는 방법 및 이를 위한 장치에 관한 것이다.

배경 기술

[0002] 의미론적 이미지 생성 기술이란 이미지가 포함하는 대상이 차지하고 있는 영역에 픽셀 단위로 해당 대상의 정보가 주어진 경우, 쉽게 이야기하면 일종의 이미지 틀이 주어졌을 때, 해당 이미지 틀에 부합하는 적절한 이미지를 생성(여기서 생성은 합성의 의미를 포함함)하는 기술을 일컫는다.

[0003] 이에 대하여 종래의 의미론적 이미지 생성 기술은 이미지 생성 모델로서 대부분 적대적 생성 신경망(Generative Adversarial Network, GAN)을 적용한바, 적대적 생성 신경망은 생성 모델(Generator)과 분류 모델(Discriminator)로 이루어진 신경망이며, 생성 모델은 분류 모델이 실제 이미지인지 아니면 이를 모방한 가짜 이미지인지 구별하기 어렵도록 진짜와 같은 가짜 이미지를 생성하도록 학습하고, 분류 모델은 생성 모델이 만들어낸 진짜와 같은 가짜 이미지와 실제 이미지를 정확하게 구별할 수 있도록 각자가 학습을 진행함으로써 전체적인 성능이 향상될 수 있는 모델이다. 이와 같은 적대적 생성 신경망을 통해 만들어지는 이미지는 그 품질이 좋고 성능 역시 우수하나, 학습 과정이 불안정하고 최빈값 붕괴 현상(Mode Collapse)에 취약하다는 단점이 존재한다.

[0004] 한편, 최근에는 적대적 생성 신경망의 단점을 극복하기 위해 새로운 모델들이 다수 개발된바, 이 중에서 디퓨전 모델(Diffusion Model)이 특히 주목받고 있다. 디퓨전 모델이란 데이터에 노이즈(Noise)를 조금씩 추가해가면서 모델을 학습시킨 뒤, 노이즈를 역으로 없애가면서 새로운 데이터를 생성할 수 있는 모델인바, 그 성능이 적대적 생성 신경망보다 우수하기에 다양한 분야에서 이를 적용하고자 하는 시도가 계속되고 있다.

[0005] 본 발명은 의미론적 이미지 생성 기술의 이미지 생성 모델로 디퓨전 모델을 적용하되, 단순 채용의 수준을 넘어 응용의 수준까지 개선하여 고품질 이미지를 다양하게 생성할 수 있는 방법 및 이를 위한 장치를 제안하고자 한다.

선행기술문헌

특허문헌

[0006] (특허문헌 0001) 대한민국 공개특허공보 제 10-2021-0040881호(2021.04.14)

발명의 내용

해결하려는 과제

[0007] 본 발명이 해결하고자 하는 기술적 과제는 의미론적 이미지 생성 기술의 이미지 생성 모델로 디퓨전 모델을 적용함으로써 안정적인 학습이 가능한 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법 및 이를 위한 장치를 제공하는 것이다.

[0008] 본 발명이 해결하고자 하는 또 다른 기술적 과제는 의미론적 이미지 생성 기술의 이미지 생성 모델로 디퓨전 모델을 적용함으로써 고품질 이미지를 다양하게 생성할 수 있는 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법 및 이를 위한 장치를 제공하는 것이다.

[0009] 본 발명이 해결하고자 하는 또 다른 기술적 과제는 인공지능 기반의 이미지 생성 모델로 디퓨전 모델을 적용하되, 모델이 처음 접하는 이미지 틀에 대해서도 이미지를 생성할 수 있는 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법 및 이를 위한 장치를 제공하는 것이다.

[0010] 본 발명의 기술적 과제들은 이상에서 언급한 기술적 과제들로 제한되지 않으며, 언급되지 않은 또 다른 기술적 과제들은 아래의 기재로부터 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

[0011] 상기 기술적 과제를 달성하기 위한 본 발명의 일 실시 예에 따른 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법은 (a) 제N(N은 임의의 양의 정수) 학습 데이터를 입력하여 이미지 생성 모델을 학습시키는 단계, (b) 의미론적 이미지를 생성하고자 하는 입력 데이터를 상기 학습시킨 이미지 생성 모델에 입력하는 단계 및 (c) 상기 입력한 입력 데이터에 따라 상기 이미지 생성 모델이 생성한 하나 이상의 의미론적 이미지를 출력하는 단계를 포함하고, 상기 입력 데이터는, 입력 조건 데이터인 조건 이미지 틀(Layout)을 포함하며, , 상기 조건 이미지 틀은, 생성하고자 하는 의미론적 이미지가 포함하는 하나 이상의 대상인 클래스(Class)가 대상별로 각각의 카테고리 고리로 분류되며, 상기 클래스가 차지하고 있는 픽셀 영역에 상기 분류된 카테고리 별로 상이한 숫자가 부여된 이미지 틀이다.

[0012] 일 실시 예에 따르면, 상기 (a) 단계는, (a-1) 상기 제N 학습 데이터에 제1 노이즈 추가 방식 내지 제4 노이즈 추가 방식 중, 어느 하나의 선택을 통해 노이즈를 추가한 제N' 학습 데이터를 생성하는 단계, (a-2) 상기 생성한 제N' 학습 데이터를 이미지 생성 모델에 입력하여 상기 제N' 학습 데이터에 적용된 노이즈를 예측하는 단계 및 (a-3) 상기 (a-1) 단계에서 추가한 노이즈와 상기 (a-2) 단계에서 예측한 노이즈를 비교하여 그 차이를 손실함수로 설정하고, 경사 하강법(Gradient Descent)를 적용하여 학습하는 단계를 포함할 수 있다.

[0013] 일 실시 예에 따르면, 상기 (a-3) 단계 이후에, (a-4) 상기 설정한 손실함수가 최소값에 수렴하는지 판단하는 단계, (a-5) 상기 (a-4) 단계의 판단 결과 NO라면 상기 (a-1) 단계로 회귀하는 단계 및 (a-6) 상기 (a-5) 단계의 판단 결과 YES라면, 상기 이미지 생성 모델에 대한 학습을 종료하는 단계를 더 포함할 수 있다.

[0014] 일 실시 예에 따르면, 상기 제N 학습 데이터는, 제N 실제 이미지이며, 이 경우, 상기 제1 노이즈 추가 방식은, 하기 수학적 식 1에 따라 이전 시간(Timestep)과 연관성이 부존재하는 노이즈 추가 방식일 수 있다.

[0015] 수학적 식 1: $q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\alpha_t}\mathbf{x}_0, (1 - \alpha_t)\mathbf{I})$

[0016] 여기서 $\mathbf{x}_t$ 는 임의의 시점 t에서 노이즈가 추가된 제N 실제 이미지, $\mathbf{x}_0$ 은 노이즈가 추가되기 전, 최초 상태의 제N 실제 이미지, $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t := \prod_{s=1}^t \alpha_s$, β_t 는 임의의 시점 t에서의 노이즈 스케줄, $\mathbf{I}$ 는 단위 행렬을 의미함

[0017] 일 실시 예에 따르면, 상기 제N 학습 데이터는, 제N 실제 이미지이며, 이 경우, 상기 제2 노이즈 추가 방식은, 하기 수학적 식 2에 따라 이전 시간(Timestep)과 직접적인 연관성이 존재하는 노이즈 추가 방식일 수 있다.

[0018] 수학적 식 2: $q(\mathbf{x}_t|\mathbf{x}_{t-1}) := \mathcal{N}(\mathbf{x}_t; \sqrt{1 - \beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I})$

[0019] 여기서 $\mathbf{x}_t$ 는 임의의 시점 t에서 노이즈가 추가된 제N 실제 이미지, $\mathbf{x}_{t-1}$ 은 임의의 시점 t 직전의 시점인 t-1에서 노이즈가 추가된 제N 실제 이미지, β_t 는 임의의 시점 t에서의 노이즈 스케줄, $\mathbf{I}$ 는 단위 행렬을 의미함

[0020] 일 실시 예에 따르면, 상기 제N 학습 데이터는, 제N 실제 이미지이며, 이 경우, 상기 제3 노이즈 추가 방식은, 하기 수학적 식 3에 따라 이전 시간(Timestep)과 직접적인 연관성이 존재하는 노이즈 추가 방식일 수 있다.

[0021] 수학적 식 3: $q(\mathbf{x}_t|\mathbf{x}_{t-1}) := \mathcal{N}(\mathbf{x}_t; \sqrt{1 - \beta_t}\mathbf{x}_{t-1}, 0)$

[0022] 여기서 $\mathbf{x}_t$ 는 임의의 시점 t에서 노이즈가 추가된 제N 실제 이미지, $\mathbf{x}_{t-1}$ 은 임의의 시점 t 직전의 시점인 t-1에서 노이즈가 추가된 제N 실제 이미지, β_t 는 임의의 시점 t에서의 노이즈 스케줄, $\mathbf{I}$ 는 단위 행렬을 의미함 을 의미함

[0023] 일 실시 예에 따르면, 상기 제N 학습 데이터는, 제N 실제 이미지이며, 이 경우, 상기 제4 노이즈 추가 방식은, 하기 수학적 식 4에 따라 이전 시간(Timestep)과 직접적인 연관성이 존재하는 노이즈 추가 방식일 수 있다.

[0024] 수학적식4: $q_{\sigma}(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) = \mathcal{N}(\sqrt{\alpha_{t-1}}\mathbf{x}_0 + \sqrt{1 - \alpha_{t-1} - \sigma_t^2} \cdot \frac{\mathbf{x}_t - \sqrt{\alpha_t}\mathbf{x}_0}{\sqrt{1 - \alpha_t}}, \sigma_t^2\mathbf{I})$

- [0025] 여기서 x_{t-1} 은 임의의 시점 t 직전의 시점인 $t-1$ 에서 노이즈가 추가된 제N 실제 이미지, x_t 는 임의의 시점 t 에서 노이즈가 추가된 제N 실제 이미지, x_0 은 노이즈가 추가되기 전, 최초 상태의 제N 실제 이미지,
- $$\alpha_t = 1 - \beta_t, \bar{\alpha}_t := \prod_{s=1}^t \alpha_s, \beta_t \text{는 임의의 시점 } t \text{에서의 노이즈 스케줄, } \sigma \text{ 는 각 시간 } t \text{마다 설정한 표준 편차, } I \text{는 단위 행렬을 의미함}$$
- [0026] 일 실시 예에 따르면, 상기 제N 학습 데이터는, 제N 실제 이미지이며, 이 경우, 상기 (a-2) 단계는, (a-2-1) 상기 생성한 제N' 학습 데이터를 이미지 생성 모델에 입력하는 단계, (a-2-1) 상기 제N 실제 이미지에 대응되는 제N 조건 이미지 틀에 제1 노이즈 추가 방식 내지 제4 노이즈 추가 방식 중, 어느 하나를 통해 노이즈를 추가한 제N 입력 조건 데이터를 상기 이미지 생성 모델에 입력하는 단계 및
- [0027] (a-2-3) 상기 입력한 제N 입력 조건 데이터를 반영하여 상기 제N' 학습 데이터에 적용된 노이즈를 예측하는 단계를 포함할 수 있다.
- [0028] 일 실시 예에 따르면, 상기 (b) 단계는, (b-1) 상기 생성하고자 하는 의미론적 이미지와 동일한 사이즈의 노이즈 데이터를 랜덤하게 생성하여 상기 입력 데이터로 상기 이미지 생성 모델에 입력하는 단계 및 (b-2) 상기 의미론적 이미지를 생성하고자 하는 조건 이미지 틀에 제1 노이즈 추가 방식 내지 제4 노이즈 추가 방식 중, 어느 하나를 통해 노이즈를 추가한 입력 조건 데이터'를 생성하여 상기 이미지 생성 모델에 입력하는 단계를 포함할 수 있다.
- [0029] 일 실시 예에 따르면, 상기 이미지 생성 모델은, 인코더부 및 디코더부를 포함하며, 조건 이미지 틀로부터 특성값을 획득하고, 이를 통과시켜 스케일값 및 바이어스값을 산출하는 하나 이상의 컨볼루션 레이어를 포함하는 조건 이미지 틀 활용부를 더 포함하고, 상기 산출한 스케일값 및 바이어스값은, 상기 입력 데이터로부터 획득하고 배치 정규화를 진행한 특성값에 입력되어 연산될 수 있다.
- [0030] 상기 기술적 과제를 달성하기 위한 본 발명의 또 다른 실시 예에 따른 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 장치는 하나 이상의 프로세서, 네트워크 인터페이스, 상기 프로세서에 의해 수행되는 컴퓨터 프로그램을 로드(Load)하는 메모리 및 대용량 네트워크 데이터 및 상기 컴퓨터 프로그램을 저장하는 스토리지를 포함하되, 상기 컴퓨터 프로그램은 상기 하나 이상의 프로세서에 의해, (A) 제N(N은 임의의 양의 정수) 학습 데이터를 입력하여 이미지 생성 모델을 학습시키는 오퍼레이션, (B) 의미론적 이미지를 생성하고자 하는 입력 데이터를 상기 학습시킨 이미지 생성 모델에 입력하는 오퍼레이션 및 (C) 상기 입력한 입력 데이터에 따라 상기 이미지 생성 모델이 생성한 하나 이상의 의미론적 이미지를 출력하는 오퍼레이션을 포함하고, 상기 입력 데이터는, 입력 조건에 해당하는 조건 이미지 틀(Layout)이며, 상기 조건 이미지 틀은, 생성하고자 하는 의미론적 이미지가 포함하는 하나 이상의 대상인 클래스(Class)가 대상별로 각각의 카테고리로 분류되며, 상기 클래스가 차지하고 있는 픽셀 영역에 상기 분류된 카테고리 별로 상이한 숫자가 부여된 이미지 틀이다.
- [0031] 상기 기술적 과제를 달성하기 위한 본 발명의 또 다른 실시 예에 따른 매체에 저장된 컴퓨터 프로그램은 컴퓨팅 장치와 결합하여, (AA) 제N(N은 임의의 양의 정수) 학습 데이터를 입력하여 이미지 생성 모델을 학습시키는 단계, (BB) 의미론적 이미지를 생성하고자 하는 입력 데이터를 상기 학습시킨 이미지 생성 모델에 입력하는 단계 및 (CC) 상기 입력한 입력 데이터에 따라 상기 이미지 생성 모델이 생성한 하나 이상의 의미론적 이미지를 출력하는 단계를 실행하며, 상기 입력 데이터는, 입력 조건에 해당하는 조건 이미지 틀(Layout)이고, 상기 조건 이미지 틀은, 생성하고자 하는 의미론적 이미지가 포함하는 하나 이상의 대상인 클래스(Class)가 대상별로 각각의 카테고리로 분류되며, 상기 클래스가 차지하고 있는 픽셀 영역에 상기 분류된 카테고리 별로 상이한 숫자가 부여된 이미지 틀이다.
- 발명의 효과**
- [0032] 상기와 같은 본 발명에 따르면, 이미지 생성 기술의 이미지 생성 모델로 디퓨전 모델을 적용하되, 메모리 자원의 부담 경감, 연산 시간 등의 요소를 실시간으로 고려하여 제1 노이즈 추가 방식 내지 제4 노이즈 추가 방식 중 어느 하나를 적절하게 선택하여 학습을 진행함으로써 안정적인 학습이 가능해짐과 동시에 고품질 이미지를 생성할 수 있다는 효과가 있다.
- [0033] 또한, 최빈값 붕괴 현상으로 인해 생성된 의미론적 이미지의 다양성이 떨어지는 종래 기술의 문제점을 해결하여 하나의 입력 데이터에 대하여 다양한 의미론적 이미지를 생성할 수 있는바, 결과물의 다양성을 확보할 수 있다는 효과가 있다.

- [0034] 또한, 입력 조건에 해당하는 조건 이미지 틀에 대해서도 노이즈를 추가하며 학습을 진행하는 디퓨전 과정을 수행하는바, 이미지 생성 모델이 입력 조건을 보다 효과적으로 받아들일 수 있음과 동시에 처음 접하는 이미지 틀에 대해서도 고품질의 의미론적 이미지를 다양하게 생성할 수 있다는 효과가 있다.
- [0035] 본 발명의 효과들은 이상에서 언급한 효과들로 제한되지 않으며, 언급되지 않은 또 다른 효과들은 아래의 기재로부터 통상의 기술자에게 명확하게 이해 될 수 있을 것이다.

도면의 간단한 설명

- [0036] 도 1은 본 발명의 제1 실시 예에 따른 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 장치가 포함하는 전체 구성을 예시적으로 도시한 도면이다.
- 도 2는 본 발명의 제2 실시 예에 따른 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법의 대표적인 단계를 나타낸 순서도이다.
- 도 3은 본 발명의 제2 실시 예에 따른 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법에 있어서, 이미지 생성 모델을 학습시키는 S210 단계를 구체화한 순서도이다.
- 도 4는 제1 노이즈 추가 방식의 모식도를 예시적으로 도시한 도면이다.
- 도 5는 제2 노이즈 추가 방식의 모식도를 예시적으로 도시한 도면이다.
- 도 6은 제3 노이즈 추가 방식의 모식도를 예시적으로 도시한 도면이다.
- 도 7은 제4 노이즈 추가 방식(σ 가 0이 아닌 경우)의 모식도를 예시적으로 도시한 도면이다.
- 도 8은 제4 노이즈 추가 방식(σ 가 0인 경우)의 모식도를 예시적으로 도시한 도면이다.
- 도 9는 디퓨전 모델에 입력되는 입력 조건인 조건 이미지 틀을 예시적으로 도시한 도면이다.
- 도 10은 도 9에 도시된 조건 이미지 틀을 이미지 생성 모델에 입력하여 생성한 의미론적 이미지를 조건 이미지 틀과 함께 예시적으로 도시한 도면이다.
- 도 11는 종래의 이미지 생성 모델인 U-Net의 구조를 예시적으로 도시한 도면이다.
- 도 12은 본 발명에서 제안하는 이미지 생성 모델의 구조인 Y-Net의 구조를 예시적으로 도시한 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0037] 본 발명의 목적과 기술적 구성 및 그에 따른 작용 효과에 관한 자세한 사항은 본 발명의 명세서에 첨부된 도면에 의거한 이하의 상세한 설명에 의해 보다 명확하게 이해될 것이다. 첨부된 도면을 참조하여 본 발명에 따른 실시 예를 상세하게 설명한다.
- [0038] 본 명세서에서 개시되는 실시 예들은 본 발명의 범위를 한정하는 것으로 해석되거나 이용되지 않아야 할 것이다. 이 분야의 통상의 기술자에게 본 명세서의 실시 예를 포함한 설명은 다양한 응용을 갖는다는 것이 당연하다. 따라서, 본 발명의 상세한 설명에 기재된 임의의 실시 예들은 본 발명을 보다 잘 설명하기 위한 예시적인 것이며 본 발명의 범위가 실시 예들로 한정되는 것을 의도하지 않는다.
- [0039] 도면에 표시되고 아래에 설명되는 기능 블록들은 가능한 구현의 예들일 뿐이다. 다른 구현들에서는 상세한 설명의 사상 및 범위를 벗어나지 않는 범위에서 다른 기능 블록들이 사용될 수 있다. 또한, 본 발명의 하나 이상의 기능 블록이 개별 블록들로 표시되지만, 본 발명의 기능 블록들 중 하나 이상은 동일 기능을 실행하는 다양한 하드웨어 및 소프트웨어 구성들의 조합일 수 있다.
- [0040] 또한, 어떤 구성요소들을 포함한다는 표현은 "개방형"의 표현으로서 해당 구성요소들이 존재하는 것을 단순히 지칭할 뿐이며, 추가적인 구성요소들을 배제하는 것으로 이해되어서는 안 된다.
- [0041] 나아가 어떤 구성요소가 다른 구성요소에 "연결되어" 있다거나 "접속되어" 있다고 언급될 때에는, 그 다른 구성요소에 직접적으로 연결 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 한다.
- [0042] 이하에서는 도면들을 참조하여 본 발명의 세부적인 실시 예들에 대해 살펴보도록 한다.
- [0043] 도 1은 본 발명의 제1 실시 예에 따른 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 장치(100)가 포함하

는 전체 구성을 예시적으로 도시한 도면이다.

- [0044] 그러나 이는 본 발명의 목적을 달성하기 위한 바람직한 실시 예일 뿐이며, 필요에 따라 일부 구성이 추가되거나 삭제될 수 있고, 어느 한 구성이 수행하는 역할을 다른 구성이 함께 수행할 수도 있음은 물론이다.
- [0045] 본 발명의 제1 실시 예에 따른 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 장치(100)는 프로세서(10), 네트워크 인터페이스(20), 메모리(30), 스토리지(40) 및 이들을 연결하는 데이터 버스(50)를 포함할 수 있으며, 기타 본 발명의 목적을 달성함에 있어 요구되는 부가적인 구성들을 더 포함할 수 있음은 물론이라 할 것이다.
- [0046] 프로세서(10)는 각 구성의 전반적인 동작을 제어한다. 프로세서(10)는 CPU(Central Processing Unit), MPU(Micro Processor Unit), MCU(Micro Controller Unit) 또는 본 발명이 속하는 기술 분야에서 널리 알려져 있는 형태의 인공지능 프로세서 중 어느 하나일 수 있다. 아울러, 프로세서(10)는 본 발명의 제2 실시 예에 따른 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법을 수행하기 위한 적어도 하나의 애플리케이션 또는 프로그램에 대한 연산을 수행할 수 있는바, 이를 위해 프로세서(10)에는 소정의 구조를 나타내는 디퓨전 모델 기반 이미지 생성 모델을 포함할 수 있으나, 이에 대해서는 후술하도록 한다.
- [0047] 네트워크 인터페이스(20)는 본 발명의 제1 실시 예에 따른 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 장치(100)의 유무선 인터넷 통신을 지원하며, 그 밖의 공지의 통신 방식을 지원할 수도 있다. 따라서 네트워크 인터페이스(20)는 그에 따른 통신 모듈을 포함하여 구성될 수 있다.
- [0048] 메모리(30)는 각종 정보, 명령 및/또는 정보를 저장하며, 본 발명의 제2 실시 예에 따른 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법을 수행하기 위해 스토리지(40)로부터 하나 이상의 컴퓨터 프로그램(41)을 로드할 수 있다. 도 1에서는 메모리(30)의 하나로 RAM을 도시하였으나 이와 더불어 다양한 저장 매체를 메모리(30)로 이용할 수 있음은 물론이다.
- [0049] 스토리지(40)는 하나 이상의 컴퓨터 프로그램(41) 및 대용량 네트워크 정보(42)를 비임시적으로 저장할 수 있다. 이러한 스토리지(40)는 ROM(Read Only Memory), EPROM(Erasable Programmable ROM), EEPROM(Electrically Erasable Programmable ROM), 플래시 메모리 등과 같은 비휘발성 메모리, 하드 디스크(HDD), 보조 저장 매체(SSD), 착탈형 디스크, 또는 본 발명이 속하는 기술 분야에서 널리 알려져 있는 임의의 형태의 컴퓨터로 읽을 수 있는 기록 매체 중 어느 하나일 수 있다.
- [0050] 컴퓨터 프로그램(41)은 메모리(30)에 로드되어, 하나 이상의 프로세서(10)에 의해, (A) 제N(N은 임의의 양의 정수) 학습 데이터를 입력하여 이미지 생성 모델을 학습시키는 오퍼레이션, (B) 의미론적 이미지를 생성하고자 하는 입력 데이터를 상기 학습시킨 이미지 생성 모델에 입력하는 오퍼레이션 및 (C) 상기 입력한 입력 데이터에 따라 상기 이미지 생성 모델이 생성한 하나 이상의 의미론적 이미지를 출력하는 오퍼레이션을 실행할 수 있다.
- [0051] 이상 간단하게 언급한 컴퓨터 프로그램(41)이 수행하는 오퍼레이션은 컴퓨터 프로그램(41)의 일 기능으로 볼 수 있으며, 보다 자세한 설명은 본 발명의 제2 실시 예에 따른 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법에 대한 설명에서 후술하도록 한다.
- [0052] 데이터 버스(50)는 이상 설명한 프로세서(10), 네트워크 인터페이스(20), 메모리(30) 및 스토리지(40) 사이의 명령 및/또는 정보의 이동 경로가 된다.
- [0053] 이상 간단하게 설명한 본 발명의 제1 실시 예에 따른 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 장치(100)는 독립된 디바이스의 형태, 예를 들어 전자 기기나 서버(클라우드 포함)의 형태일 수 있으며, 여기서 전자 기기는 한 장소에 고정 설치되어 사용하는 데스크톱 PC, 서버 디바이스 등과 같은 기기 뿐만 아니라, 스마트폰, 태블릿 PC, 노트북 PC, PDA, PMP 등과 같이 휴대가 용이한 포터블 기기 등이라도 무방한바, 프로세서(10)에 해당하는 CPU 등이 설치되고 네트워크 기능만 보유하고 있는 전자 기기라면 어떠한 것이라도 무방하다 할 것이다.
- [0054] 이하, 본 발명의 제1 실시 예에 따른 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 장치(100)가 독립된 디바이스인 전자 기기 형태 중, "서버"의 형태임을 전제로 이미지 틀을 입력하여 이에 부합하는 의미론적 이미지를 생성하고자 하는 사용자의 사용자 단말(미도시)에 설치된 전용 어플리케이션을 통해 본 발명의 제2 실시 예에 따른 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법을 제공하는 과정에 대하여 도 2 내지 도 11을 참조하여 설명하도록 한다.
- [0055] 도 2는 본 발명의 제2 실시 예에 따른 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법의 대표적인 단

계를 나타낸 순서도이다.

- [0056] 그러나 이는 본 발명의 목적을 달성함에 있어서 바람직한 실시 예일 뿐이며, 필요에 따라 일부 단계가 추가 또는 삭제될 수 있음은 물론이고, 어느 한 단계가 다른 단계에 포함되어 수행될 수도 있다.
- [0057] 한편, 각 단계는 본 발명의 제1 실시 예에 따른 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 장치(100)를 통해 이루어지는 것을 전제로 하며, 본 발명의 제1 실시 예에 따른 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 장치(100)가 "서버"의 형태임을 전제로 하였기 때문에 사용자 단말(미도시)에 설치된 전용 어플리케이션을 본 발명의 제1 실시 예에 따른 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 장치(100)와 동일한 의미로 볼 것이고, 설명의 편의를 위해 이들 모두를 "장치(100)"로 명명하도록 한다.
- [0058] 우선, 장치(100)가 제N(은 임의의 양의 정수) 학습 데이터를 입력하여 이미지 생성 모델을 학습시킨다(S210).
- [0059] 여기서 이미지 생성 모델은 디퓨전 모델 기반의 이미지 생성 모델을 의미하며, 이하 도 3을 참조하여 S210 단계에 대해 보다 자세히 설명하도록 한다.
- [0060] 도 3은 본 발명의 제2 실시 예에 따른 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법에 있어서, 이미지 생성 모델을 학습시키는 S210 단계를 구체화한 순서도이다.
- [0061] 그러나 이는 본 발명의 목적을 달성함에 있어서 바람직한 실시 예일 뿐이며, 필요에 따라 일부 단계가 추가 또는 삭제될 수 있음은 물론이고, 어느 한 단계가 다른 단계에 포함되어 수행될 수도 있다.
- [0062] 우선, 장치(100)가 제N 학습 데이터에 제1 노이즈 추가 방식 내지 제4 노이즈 추가 방식 중, 어느 하나의 선택을 통해 노이즈를 추가한 제N' 학습 데이터를 생성한다(S210-1).
- [0063] 여기서 제N 학습 데이터는 실제 이미지인 제N 실제 이미지일 수 있으며, 예를 들어, 이미지 생성 모델을 통해 생성하고자 하는 의미론적 이미지가 사람 얼굴 이미지라고 한다면, 제N 실제 이미지는 사람 얼굴 이미지일 수 있다.
- [0064] 한편, 본 발명의 제2 실시 예에 따른 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법은 보편적인 디퓨전 모델에서 학습을 위해 노이즈를 추가하는 방식 대비 노이즈 추가 방식에서 특이성을 보유하고 있는바, 이것이 제1 노이즈 추가 방식 내지 제4 노이즈 추가 방식이다.
- [0065] 이에 대한 구체적인 설명에 앞서 설명의 전제 조건을 설정하면, 제1 내지 제4 노이즈 추가 방식은 제N 실제 이미지에 대하여 임의의 시간 t 에서 노이즈를 추가하는 방식, 보다 구체적으로 임의의 중간 시간(Timestamp) t 에서의 데이터(x_t 또는 y_t)를 얻기 위한 수학적식을 어떻게 설계하느냐에 따라 상이해질 수 있는 것이며, 학습 과정에서는 시간이 0, 1, 2편 ($T-2$), ($T-1$), T 와 같이 설정되고, 추론 과정에서는 T , ($T-1$), ($T-2$)편2, 1, 0과 같이 시간이 변화한다. 또한 제N 실제 이미지는 시간 0에서 원본(또는 샘플링, 추론 과정에서라면 - 생성된 이미지)의 형태이고, 시간 T 에서는 완전한 노이즈 형태이며, 매 시간(Timestamp) 마다 노이즈를 얼마나 추가할지에 대한 노이즈 스케줄을 β_1 , β_2 , β_3 편 β_t 편 β_{T-1} , β_T 라 하도록 한다.
- [0066] 제1 노이즈 추가 방식은 하기 수학적식 1에 따라 이전 시간(Timestep)과 연관성이 부존재하는 노이즈 추가 방식이다.
- [0067] 수학적식 1:
$$q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1 - \bar{\alpha}_t)\mathbf{I})$$
- [0068] 여기서 x_t 는 임의의 시점 t 에서 노이즈가 추가된 제N 실제 이미지, x_0 은 노이즈가 추가되기 전, 최초 상태의 제N 실제 이미지, $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t := \prod_{s=1}^t \alpha_s$, β_t 는 임의의 시점 t 에서의 노이즈 스케줄, $\mathbf{I}$ 는 단위 행렬을 의미한다.
- [0069] 이와 같은 제1 노이즈 추가 방식은 보편적인 디퓨전 모델에서 학습을 진행하는 경우 노이즈를 추가하는 방식과 유사하나, 본 발명에서 이와 같이 새롭게 정의한 것은 의미론적 이미지 생성을 위한 이미지 생성 모델의 학습 과정에서 노이즈를 추가하는 것과 달리, 제N 조건 이미지 틀에 대한 설명에서 후술하겠지만 학습이 완료된 이미지 생성 모델에 추론을 진행하면서 조건 이미지 틀에 노이즈를 추가하는 조건 이미지 디퓨전 방식은 종래에 존재하지 않았기 때문이다.
- [0070] 한편, 수학적식 1에서 확인 가능하듯이 제1 노이즈 추가 방식은 이전 시간과 직접적인 연관성이 부존재하며, 오직

원본 데이터와 노이즈 스케줄만으로 산정되기 때문에 실제 이미지에 노이즈가 추가되는 과정에서 각 과정의(노이즈된) 실제 이미지 간의 명시적인 관계성이 부족할 수 있으므로 생성된 의미론적 이미지 품질의 저하로 이어질 가능성이 있으나, 수학적 1에 원본 데이터에 해당하는 실제 이미지를 대입하기만 하면 바로 해당 시간에서 노이즈가 추가된 조건 이미지 틀을 산정할 수 있기 때문에 장치(100)의 시간과 메모리 활용에서 매우 큰 이점이 있다 할 것이다.

[0071] 도 4에 제1 노이즈 추가 방식의 모식도를 예시적으로 도시해 놓았으며, y_0 , y_{t-1} , y_t , y_T 모두 연결됨이 없이 각자 개별적인 방향을 향하고 있기에 이전 시간과 직접적인 연관성이 부존재함을 확인할 수 있으며, 그에 따라 제1 노이즈 추가 방식을 Jump Noising(점프 방식)으로 명명하도록 한다.

[0072] 제2 노이즈 추가 방식은 하기 수학적 2에 따라 이전 시간(Timestep)과 직접적인 연관성이 존재하는 노이즈 추가 방식이다.

[0073] 수학적 2: $q(\mathbf{x}_t|\mathbf{x}_{t-1}) := \mathcal{N}(\mathbf{x}_t; \sqrt{1 - \beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I})$

[0074] 여기서 x_t 는 임의의 시점 t 에서 노이즈가 추가된 제N 실제 이미지, x_{t-1} 은 임의의 시점 t 직전의 시점인 $t-1$ 에서 노이즈가 추가된 제N 실제 이미지, β_t 는 임의의 시점 t 에서의 노이즈 스케줄, $\mathbf{I}$ 는 단위 행렬을 의미함을 의미한다.

[0075] 수학적 2에서 확인 가능하듯이 제2 노이즈 추가 방식은 인접한 시간과 직접적인 연관성이 존재하는, 보다 구체적으로 이전 시간의 데이터를 직접 활용하여 다음 시간의 데이터를 생성하는 방식이며, 각 시간에서의 데이터가 이전 및 이후 시간과 긴밀한 관계를 보유하고 있기 때문에 생성되는 의미론적 이미지가 노이즈가 추가된 실제 이미지에 의해 보다 정교하고 정확하게 안내를 받을 수 있기에 더욱 양질의 의미론적 이미지가 생성될 수 있다.

[0076] 그러나 원본 데이터, 즉 최초 상태의 제N 실제 이미지만 있으면 되는 제1 노이즈 추가 방식과는 달리, 각 시간에 맞는 노이즈가 추가된 실제 이미지를 순서대로 이미지 생성 모델에 입력해줘야 하기 때문에 각 시간에 따른 노이즈가 추가된 데이터를 모두 메모리에 저장해두어야 하며, 이로 인해 장치(100)의 메모리 자원에서 일정 수준 이상의 부담이 생길 수 있다.

[0077] 도 5에 제2 노이즈 추가 방식의 모식도를 예시적으로 도시해 놓았으며, y_0 , y_{t-1} , y_t , y_T 모두 향하고 있는 방향은 상이하나, 하나로 연결되어 있기에 이전 시간과 직접적인 연관성이 존재함을 확인할 수 있으며, 그에 따라 제2 노이즈 추가 방식을 Ancestral Noising(순차적 방식)으로 명명하도록 한다.

[0078] 제3 노이즈 추가 방식은 하기 수학적 3에 따라 이전 시간(Timestep)과 직접적인 연관성이 존재하는 노이즈 추가 방식이다.

[0079] 수학적 3: $q(\mathbf{x}_t|\mathbf{x}_{t-1}) := \mathcal{N}(\mathbf{x}_t; \sqrt{1 - \beta_t}\mathbf{x}_{t-1}, 0)$

[0080] 여기서 x_t 는 임의의 시점 t 에서 노이즈가 추가된 제N 실제 이미지, x_{t-1} 은 임의의 시점 t 직전의 시점인 $t-1$ 에서 노이즈가 추가된 제N 실제 이미지, β_t 는 임의의 시점 t 에서의 노이즈 스케줄, $\mathbf{I}$ 는 단위 행렬을 의미함을 의미한다.

[0081] 제3 노이즈 추가 방식은 제2 노이즈 추가 방식과 동일하게 수학적 2를 기초로 하는데, 수학적 2가 나타내는 정규 분포에서 표준편차(공분산) 항을 제외하고(0으로 처리) 평균 항만을 사용하는 방식인바, 제2 노이즈 추가 방식의 장점인 인접한 시간 간에 보다 긴밀한 연관성을 그대로 유지하면서도 표준편차로 인해 발생하는 확률적 연산 부분을 생략하고 평균항, 즉 가장 확률이 높을 것으로 기대되는 방향으로 실제 이미지에 노이즈를 추가해주는 방식인바, 평균 항만을 사용하기에 제2 노이즈 추가 방식과 같이 각 시간에 따른 노이즈가 추가된 데이터를 모두 메모리에 저장해둘 필요가 없으므로 장치(100)의 메모리 자원에서 발생할 수 있는 부담을 제거할 수 있다.

[0082] 도 6에 제3 노이즈 추가 방식의 모식도를 예시적으로 도시해 놓았으며, 평균 항만을 이용하기에 y_0 , y_{t-1} , y_t , y_T 모두 향하고 있는 방향이 동일하며, 이들 모두가 하나로 연결되어 있기에 이전 시간과 직접적인 연관성이 존재함을 확인할 수 있고, 그에 따라 제3 노이즈 추가 방식을 Mean Noising(평균 방식)으로 명명하도록 한다.

[0083] 마지막으로 제4 노이즈 추가 방식은 하기 수학적 4에 따라 이전 시간(Timestep)과 직접적인 연관성이 존재하는

노이즈 추가 방식이다.

[0084] 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법.

[0085] 수학식4:

$$q_{\sigma}(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) = \mathcal{N}(\sqrt{\alpha_{t-1}}\mathbf{x}_0 + \sqrt{1 - \alpha_{t-1} - \sigma_t^2} \cdot \frac{\mathbf{x}_t - \sqrt{\alpha_t}\mathbf{x}_0}{\sqrt{1 - \alpha_t}}, \sigma_t^2 \mathbf{I})$$

[0086] 여기서 $\mathbf{x}_{t-1}$ 은 임의의 시점 t 직전의 시점인 $t-1$ 에서 노이즈가 추가된 제N 실제 이미지, $\mathbf{x}_t$ 는 임의의 시점 t 에서 노이즈가 추가된 제N 실제 이미지, $\mathbf{x}_0$ 은 노이즈가 추가되기 전, 최초 상태의 제N 실제 이미지, $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t := \prod_{s=1}^t \alpha_s$, β_t 는 임의의 시점 t 에서의 노이즈 스케줄, σ 는 각 시간 t 마다 설정한 표준편차, $\mathbf{I}$ 는 단위 행렬을 의미한다.

[0087] 이와 같은 수학식 4는 $q_{\sigma}(\mathbf{x}_T|\mathbf{x}_0) = \mathcal{N}(\sqrt{\alpha_T}\mathbf{x}_0, (1 - \alpha_T)\mathbf{I})$ 으로부터 파생한 수학식인바, 보편적인 디퓨전 모델에서와 같이 마르코브 과정은 아니며(Non-Markovian), σ 는 각 시간 t 마다 설정한 표준편차이므로 σ 가 클 수록 디퓨전 과정(Forward Process)이 확률적이게 되고, 0에 가까워질수록 결정론적으로 된다.

[0088] 즉, 수학식 4에서 σ 는 장치(100)의 관리자가 사용자가 개별적으로 설정할 수 있으며, 각 시간 t 에 대한 σ 를 모두 0으로 설정하여 각 시간 별 변환 과정을 결정론적으로 만드는 경우 제3 노이즈 추가 방식과 유사하게 확률적 연산 부분을 생략하기 때문에 제2 노이즈 추가 방식과 같이 각 시간에 따른 노이즈가 추가된 데이터를 모두 메모리에 저장해둘 필요가 없으므로 장치(100)의 메모리 자원에서 발생할 수 있는 부담을 제거할 수 있다.

[0089] 한편, 각 시간 t 에 대한 σ 를 개별적으로 설정하는 경우 각 시간 별 변환 과정이 확률적으로 만들어지는데, 제2 노이즈 추가 방식과 유사하게 생성되는 의미론적 이미지가 노이즈가 추가된 실제 이미지에 의해 보다 정교하고 정확하게 안내를 받을 수 있기에 더욱 양질의 의미론적 이미지가 생성될 수 있다. 즉, 제4 노이즈 추가 방식은 각 시간 t 에 대한 σ 의 설정에 따라 제2 노이즈 추가 방식의 장점 또는 제3 노이즈 추가 방식의 장점을 취사 선택하여 제공 받을 수 있는 방식이다.

[0090] 도 7에 제4 노이즈 추가 방식의 모식도, 보다 구체적으로 σ 가 0이 아닌 경우를 예시적으로 도시해 놓았으며, y_T 를 먼저 산정한 후, y_t 와 y_0 으로부터 y_{t-1} 을 산정하는 것을 반복함으로써 y_T 부터 y_1 까지 산정한다. σ 가 0이 아니기에 확률적이게 되는데, 도 6의 경우와 같이 y_T , y_t , y_{t-1} , y_0 모두 향하고 있는 방향은 상이하나, 하나로 연결되어 있기에 이전 시간과 직접적인 연관성이 존재함을 확인할 수 있다.

[0091] 한편, 도 8에 제4 노이즈 추가 방식의 모식도, 보다 구체적으로 σ 가 0인 경우를 예시적으로 도시해 놓았으며, σ 가 0이기에 결정론적이게 되는데, 도 7의 경우와 같이 y_T , y_t , y_{t-1} , y_0 모두 향하고 있는 방향은 동일하며, 하나로 연결되어 있기에 이전 시간과 직접적인 연관성이 존재함을 확인할 수 있으며, 그에 따라 제4 노이즈 추가 방식을 DDIM Noising(Denoising Diffusion Implicit Models)으로 명명하도록 한다.

[0092] 이상 네 가지 노이즈 추가 방식에 대하여 설명하였다. 본 발명의 제2 실시 예에 따른 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법은 학습 과정에서 생성하고자 하는 의미론적 이미지의 품질 우선, 장치(100)의 메모리 자원의 부담 경감, 연산 시간 등의 요소를 실시간으로 고려하여 제1 노이즈 추가 방식 내지 제4 노이즈 추가 방식 중 어느 하나를 적절하게 선택하여 제N 학습 데이터에 노이즈를 효과적으로 추가할 수 있으며, 여러 가지 노이즈 추가 방식을 활용함으로써 생성하고자 하는 의미론적 이미지의 다양성을 확보할 수 있다. 또한, 제N 학습 데이터에 대하여 시간 t 에서는 제1 노이즈 추가 방식으로, $t+1$ 에서는 제2 노이즈 추가 방식으로, $t+3$ 에서는 제3 노이즈 추가 방식으로 등과 같이 하나의 학습 데이터에 대하여 네 가지 노이즈 추가 방식 간, 실시간 전환 역시 자유롭게 가능하게 함으로써 학습의 완성도까지 향상시킬 수도 있을 것이다.

[0093] 다시 도 3에 대한 설명으로 돌아가도록 한다.

[0094] 제 N' 학습 데이터를 생성했다면, 장치(100)가 생성한 제 N' 학습 데이터를 이미지 생성 모델에 입력하여 제 N' 학습 데이터에 적용된 노이즈를 예측한다(S210-2).

[0095] 이는 제 N' 학습 데이터에 적용된 노이즈를 예측하고, 예측한 만큼을 제 N' 학습 데이터로부터 제거한다면 노이즈 추가 전의 데이터인 제N 학습 데이터를 산출할 수 있기 때문이며, 앞서 설명한 본 발명의 특이성인 제1 노

이즈 추가 방식 내지 제4 노이즈 추가 방식은 S210-2 단계하고도 연관될 수 있는바, 이하 설명하도록 한다.

- [0096] 노이즈를 예측하는 S210-2 단계는 장치(100)가 제N' 학습 데이터를 이미지 생성 모델에 입력하는 단계(S210-2-1), 제N 학습 데이터인 제N 실제 이미지에 대응되는 제N 조건 이미지 틀에 제1 노이즈 추가 방식 내지 제4 노이즈 추가 방식 중, 어느 하나를 통해 노이즈를 추가한 제N 입력 조건 데이터를 이미지 생성 모델에 입력하는 단계(S210-2-2) 및 입력한 제N 입력 조건 데이터를 반영하여 제N' 학습 데이터에 적용된 노이즈를 예측하는 단계(S210-2-3)를 포함할 수 있으며, 본 발명의 제2 실시 예에 따른 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법은 학습 과정에서 제N 실제 이미지뿐만 아니라 제N 실제 이미지에 대응되는 제N 조건 이미지 틀에 대해서도 노이즈를 추가하여 이미지 생성 모델에 입력함으로써 학습 완성도를 높이고 전체 이미지 생성 모델의 성능 향상에 이바지할 수 있다.
- [0097] 여기서 조건 이미지 틀이란 디퓨전 모델의 입력 조건(Conditional Input)에 해당하는 것으로서, 조건 이미지 틀을 통해 생성하고자 하는 의미론적 이미지가 포함하는 하나 이상의 대상인 클래스(Class)가 대상별로 각각의 카테고리로 분류되며, 클래스가 차지하고 있는 픽셀 영역에 분류된 카테고리 별로 상이한 숫자, 예를 들어 0 내지 18 사이의 숫자 또는 그 이상의 숫자까지 부여될 수 있는바, 이 때문에 조건 이미지 틀을 Semantic layout, Semantic map, Semantic condition 등으로도 부르기도 한다.
- [0098] 도 9는 디퓨전 모델에 입력되는 입력 조건인 조건 이미지 틀을 예시적으로 도시한바, 숫자가 부여된 개별적인 픽셀은 크기가 매우 작아 도면 상에 표시하기에 어려우므로 식별의 편의성을 고려하여 숫자 대신 서로 상이한 색상으로 구분해 놓았으며, 가장 좌측 도면에서의 구름(회색), 하늘(하늘색), 나무(갈색), 산(청록색), 바다(파랑색), 잔디(녹색)가 조건 이미지 틀을 통해 생성하고자 하는 의미론적 이미지가 포함하는 대상인 클래스(또는 카테고리)에 해당하는 것으로 볼 수 있다.
- [0099] 이를 기초로 우측 네 개의 도면을 살펴보면(순서대로 1 내지 4 번 도면이라 한다), 1번 도면은 클래스로 구름, 하늘, 나무, 바다를 2번 도면은 클래스로 구름, 하늘, 나무, 바다를, 3번 도면은 클래스로 구름, 하늘, 나무, 잔디를, 4번 도면은 클래스로 구름, 하늘, 나무, 산, 바다를 포함하고 있음을 각각의 클래스가 조건 이미지 틀 상에서 차지하고 있는 픽셀 영역을 통해 확인할 수 있으며, 각각의 클래스에 해당하는 픽셀 영역이 이미지 생성 모델을 통해 생성된 실제 클래스에 대한 이미지가 생성될 영역에 해당한다.
- [0100] 4번 이미지를 예로 하여 설명하면, 회색으로 표시된 구름 클래스에 대한 픽셀 영역에 실제 구름의 이미지가, 하늘색으로 표시된 하늘 클래스에 대한 픽셀 영역에 실제 하늘의 이미지가, 갈색으로 표시된 나무 클래스에 대한 픽셀 영역에 실제 나무의 이미지가, 청록색으로 표시된 산에 대한 픽셀 영역에 실제 산의 이미지가, 파랑색으로 표시된 바다 클래스에 대한 픽셀 영역에 실제 바다의 이미지가 생성되는 것이며, 생성하고자 하는 의미론적 이미지 상의 어느 영역에 어떠한 대상(또는 클래스)이 생성되어야 하는지에 대한 정보를 제공하기에 입력 조건이라 하는 것이다.
- [0101] 이와 같이 조건 이미지 틀은 입력 조건에 해당하기 때문에 보편적인 디퓨전 모델에서 노이즈 추가의 대상이 되지 않고 단순 입력되나, 입력 조건에 해당하는 조건 이미지 틀에도 노이즈를 추가해야 함에는 다양한 이유가 존재함을 본 발명에 대한 연구를 통해 인지하게 되었다. 구체적으로 의미론적 이미지 생성을 위해 입력되는 조건 이미지 틀은 픽셀마다 값 자체는 정수로 되어 있으며, 카테고리 개수만큼 존재하는 이산 데이터지만, 픽셀마다의 값이 조합되어 만들어진 조건 이미지 틀 자체는 존재 가능한 개수가 거의 무한하며, 실제로 의미론적 이미지를 생성할 경우(샘플링을 수행할 경우), 이미지 생성 모델이 한번도 본적이 없는 조건 이미지 틀이 입력으로 들어오게 되어 낮은 조건에 강건한 모델을 구축해야 할 필요성이 높기 때문인바, 조건 이미지 틀에 노이즈를 추가하는 본 발명의 독자적인 기술적 특징을 "조건 이미지 디퓨전"이라 명명하도록 하며, 본 발명의 명칭은 이를 활용하여 작명되었다.
- [0102] 디퓨전 모델에서 노이즈를 추가하는, 즉 디퓨전을 시키는 과정이 도입된 이유는 입력 데이터 분포에서 확률값이 낮은 데이터에 대해 학습이 제대로 이루어지지 않는다는 문제점, 보다 구체적으로 확률값이 낮은 데이터에 대해서는 학습을 위한 손실 함수 값이 낮게 정해지기 때문에 학습이 제대로 이루어지지 않는다는 문제점을 해결하기 위해서인바, 데이터에 노이즈를 추가하여 적정량 퍼트리게 되면 확률값이 낮았던 데이터가 줄어들게 되고, 해당 부분에 대한 학습 문제점이 완화될 수 있기에 본 발명에서는 디퓨전 모델이 효과를 보인 지점과 앞서 설명한 의미론적 이미지 생성에서 조건 이미지 틀의 특성이 동일한 맥락에 놓여져 있다고 판단하여 입력 조건에 해당하는 조건 이미지 틀까지 노이즈 추가 대상에 획기적으로 편입시킨 것이다.
- [0103] 그에 따라 이상 설명한 제N 학습 데이터인 제N 실제 이미지에 노이즈를 추가하는 독자적인 네 가지 방식인 제1

노이즈 추가 방식 내지 제4 노이즈 추가 방식은 제N 조건 이미지 틀에 대해서도 동일하게 적용될 수 있으며, 이때 각 수학적식에서 실제 이미지에 해당하는 부분을 조건 이미지 틀로만 변경하면 충분하므로 중복 서술을 방지하기 위해 자세한 설명은 생략하도록 한다.

[0104] 다시 도 3에 대한 설명으로 돌아가도록 한다.

[0105] 제N' 학습 데이터에 적용된 노이즈를 예측했다면, 장치(100)가 S210-1 단계에서 추가한 노이즈와 S210-2 단계에서 예측한 노이즈를 비교하여 그 차이를 손실함수로 설정하고, 경사 하강법(Gradient Descent)을 적용하여 학습한다(S210-3).

[0106] 장치(100)는 S210-1 단계에서 제N 학습 데이터에 노이즈를 추가하였으므로 얼마만큼의 노이즈를 추가했는지에 대한 정보를 보유하고 있으며, 이를 정답 노이즈로, S210-2 단계에서 예측한 노이즈를 예측 노이즈라 한다면, 여기서의 손실함수는 $|(\text{정답 노이즈}) - (\text{예측 노이즈})|^2$ 로 설정할 수 있다.

[0107] 설정한 손실함수에 적용하는 경사 하강법은 S210-3 단계에서 활용되는 예시적인 딥러닝 학습 이론이며, 보다 구체적으로 손실함수의 기울기 변화를 통해 손실함수를 줄이는 방법으로써, 초기 시점에 대한 기울기를 구하여 기울기의 반대방향으로 이동하는 과정을 통해 기울기를 수렴시키는 것으로 오차를 줄이는 딥러닝 학습 이론인바, 경사 하강법 적용을 통해 손실함수가 점점 최소값에 수렴하게 될 것이며, 그에 따라 정답 노이즈와 예측 노이즈의 차이가 줄어들게 될 것이므로 본 발명의 제2 실시 예에 따른 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법에 최적화된 학습 이론일 것이나, 이에 반드시 한정하는 것은 아니며 다른 공지된 학습 이론을 적용할 수 있음은 물론이라 할 것이다.

[0108] 한편, 경사 하강법 적용에 따라 S210-3 단계 이후에는 설정한 손실함수가 최소값에 수렴하는지 판단하는 단계(S210-4), S210-4 단계의 판단 결과 NO라면, S210-1 단계로 회귀하는 단계 및 S210-4 단계의 판단 결과 YES라면, 이미지 생성 모델에 대한 학습을 종료하는 단계가 수행될 수행될 수 있을 것이다.

[0109] 다시 도 2에 대한 설명으로 돌아가도록 한다.

[0110] 이미지 생성 모델의 학습이 완료되었다면, 장치(100)가 의미론적 이미지를 생성하고자 하는 입력 데이터를 학습시킨 이미지 생성 모델에 입력한다(S220).

[0111] 이와 같은 S220 단계는 이미지 생성 모델의 학습을 완료한 이후에 실제로 의미론적 이미지를 생성할 경우(샘플링, 추론을 수행할 경우)의 시작 단계에 해당하며, 본 발명의 제2 실시 예에 따른 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법은 이를 위해 입력하는 입력 데이터는 생성하고자 하는 의미론적 이미지와 동일한 사이즈의 노이즈 데이터를 랜덤하게 생성한 데이터 및 입력 조건 데이터에 해당하는 조건 이미지 틀을 포함할 수 있다.

[0112] 또한, 본 발명의 제2 실시 예에 따른 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법은 실제로 의미론적 이미지를 생성할 경우 입력 데이터가 포함하는 입력 조건 데이터인 의미론적 이미지를 생성하고자 하는 조건 이미지 틀에도 노이즈를 추가하여 이미지 생성 모델에 입력하는바, 노이즈 추가 방식의 특이성이 존재하므로 입력 데이터에 대하여 학습 과정에서와 같은 상태를 의도적으로 만들어줌으로써 생성하고자 하는 의미론적 이미지의 품질을 향상시키기 위함이다. 그에 따라 S220 단계는 장치(100)가 생성하고자 하는 의미론적 이미지와 동일한 사이즈의 노이즈 데이터를 랜덤하게 생성하여(여기서 노이즈 데이터는 랜덤하되, 완전한 노이즈 데이터(시점 T)이며, 가우시안 노이즈 분포로부터 샘플링할 수 있음) 입력 데이터로 이미지 생성 모델에 입력하는 단계(S220-1) 및 의미론적 이미지를 생성하고자 하는 조건 이미지 틀에 제1 노이즈 추가 방식 내지 제4 노이즈 추가 방식 중, 어느 하나를 통해 노이즈를 추가한 입력 조건 데이터'를 생성하여 이미지 생성 모델에 입력하는 단계(S220-2)를 포함할 수 있다.

[0113] 한편, 입력한 입력 데이터에 따라 이미지 생성 모델은 S210 단계의 학습 과정에서 학습한 바에 따라 입력 데이터에 얼마 만큼의 노이즈가 적용되어 있는지 예측하고, 입력 데이터에서 예측한만큼의 노이즈를 제거하며, 노이즈를 제거한 입력 데이터에 시점 T-1에 해당하는 노이즈를 노이즈 스케줄에 따라 추가하여 시점 T-1에서의 입력 데이터를 생성하여 이미지 생성 모델에 다시 입력하되, 이러한 과정을 시점 T-2, T-3 쯤 2, 1, 0까지 반복하여 최종적으로 생성한 입력 데이터를 최종적인 의미론적 이미지로 출력한다 할 것이다.

[0114] 이를 추론(Inference) 과정이라 하며, 도 10에 도 9에 도시된 조건 이미지 틀을 이미지 생성 모델에 입력하여 생성한 의미론적 이미지를 조건 이미지 틀과 함께 예시적으로 도시해 놓았다. 도 10을 참조하면, 조건 이미지 틀상에서 클래스에 해당하는 구름(회색), 하늘(하늘색), 나무(갈색), 산(청록색), 바다(파랑색), 잔디(녹색)가

차지하고 있는 픽셀 영역에 이미지 생성 모델이 생성한 실제 구름, 하늘, 나무, 산, 바다, 잔디가 이미지가 채워져 있는 것을 확인할 수 있으며, 하나의 조건 이미지 틀에 대하여 복수 개의 의미론적 이미지가 다양하게 생성되어 있는 것을 확인할 수 있다.

[0115] 지금까지 본 발명의 제2 실시 예에 따른 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법에 대하여 설명하였다. 본 발명에 따르면, 이미지 생성 기술의 이미지 생성 모델로 디퓨전 모델을 적용하되, 장치(100)의 메모리 자원의 부담 경감, 연산 시간 등의 요소를 실시간으로 고려하여 제1 노이즈 추가 방식 내지 제4 노이즈 추가 방식 중 어느 하나를 적절하게 선택하여 학습을 진행함으로써 안정적인 학습이 가능해짐과 동시에 고품질 이미지를 생성할 수 있다. 아울러, 최빈값 붕괴 현상으로 인해 생성된 의미론적 이미지의 다양성이 떨어지는 종래 기술의 문제점을 해결하여 하나의 입력 데이터에 대하여 다양한 의미론적 이미지를 생성할 수 있는바, 결과물의 다양성을 확보할 수 있다. 또한, 입력 조건 데이터에 해당하는 조건 이미지 틀에 대해서도 노이즈를 추가하며 학습을 진행하는 디퓨전 과정을 수행하는바, 이미지 생성 모델이 입력 조건을 보다 효과적으로 받아들일 수 있음과 동시에 처음 접하는 이미지 틀에 대해서도 고품질의 의미론적 이미지를 다양하게 생성할 수 있다.

[0116] 도 11에는 종래의 이미지 생성 모델인 U-Net의 구조를 예시적으로, 도 12에는 본 발명의 제1 실시 예에 따른 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 장치(100)와 본 발명의 제2 실시 예에 따른 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법에서 언급한 이미지 생성 모델의 구조를 예시적으로 도시해 놓았다. 도 12를 참조하면, 본 발명에서의 이미지 생성 모델은 인코더부(I)와 디코더부(D)를 포함하되, 조건 이미지 틀로부터 특성값을 획득하고, 이를 통과시켜 스케일값 및 바이어스값을 산출하는 하나 이상의 컨볼루션 레이어를 포함하는 조건 이미지 틀 활용부(U)를 더 포함하고 있음을 확인할 수 있다.

[0117] 인코더부(I)와 디코더부(D)만을 본다면 도 11에 도시된 U-Net의 구조와 동일하나, 본 발명에서의 이미지 생성 모델은 입력 조건에 해당하는 조건 이미지 틀에 대해서까지 디퓨전 과정을 수행하기에 그에 따른 특이성을 반영할 수 있는 이미지 틀 활용부(U)를 더 포함하는바, 그에 따라 전체적인 구조가 Y와 유사하기에 이를 Y-Net라 명명하여 새롭게 제안하며, 이미지 틀 활용부(U)가 산출한 스케일값 및 바이어스값은 입력 데이터로부터 획득하고 배치 정규화를 진행한 특성값에 입력되어 연산될 수 있는바, 이는 SPADE(Spatially-Adaptive Denormalization) 방식, 쉽게 이야기하면 조건 이미지 틀을 활용하여 특성 벡터에 조건 이미지 틀에 대한 의미론적 정보를 정규화 방식으로 주입하는 기술을 본 발명의 특성에 맞추어 응용한 것으로 볼 수 있다 할 것이다.

[0118] 마지막으로, 본 발명의 제1 실시 예에 따른 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 장치(100)와 본 발명의 제2 실시 예에 따른 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법은 본 발명의 제3 실시 예에 따른 컴퓨터로 판독 가능한 매체에 저장된 컴퓨터 프로그램으로 구현할 수도 있는바, 이 경우 컴퓨팅 장치와 결합하여 (AA) 제N(N은 임의의 양의 정수) 학습 데이터를 입력하여 이미지 생성 모델을 학습시키는 단계, (BB) 의미론적 이미지를 생성하고자 하는 입력 데이터를 상기 학습시킨 이미지 생성 모델에 입력하는 단계 및 (CC) 상기 입력한 입력 데이터에 따라 상기 이미지 생성 모델이 생성한 하나 이상의 의미론적 이미지를 출력하는 단계를 실행할 수 있을 것이며, 중복 서술을 위해 자세히 기재하지는 않았지만 본 발명의 제1 실시 예에 따른 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 장치(100) 및 본 발명의 제2 실시 예에 따른 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 방법에 적용된 모든 기술적 특징은 본 발명의 제3 실시 예에 따른 컴퓨터로 판독 가능한 매체에 저장된 컴퓨터 프로그램에 모두 동일하게 적용될 수 있음은 물론이라 할 것이다.

[0119] 이상 첨부된 도면을 참조하여 본 발명의 실시 예들을 설명하였지만, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 본 발명이 그 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 실시될 수 있다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시 예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다.

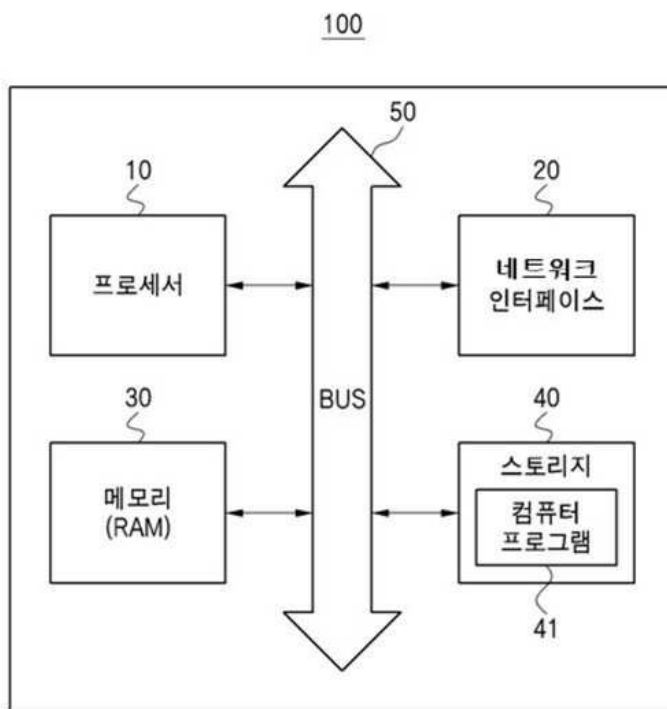
부호의 설명

[0120] 10: 프로세서
11: 제1 네트워크
12: 제2 네트워크
20: 네트워크 인터페이스
30: 메모리

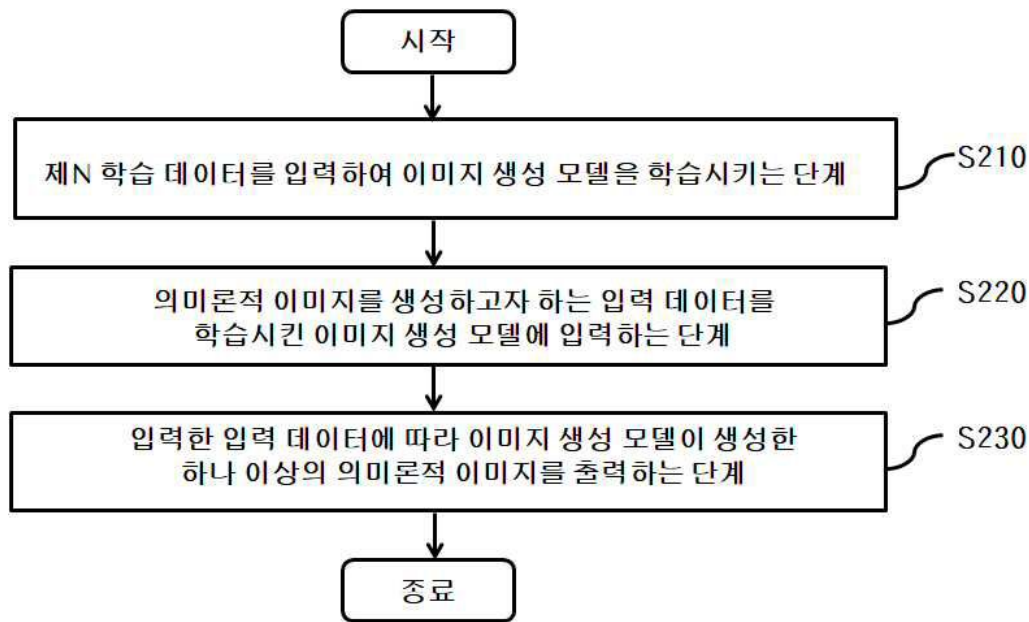
- 40: 스토리지
 41: 컴퓨터 프로그램
 50: 정보 버스
 100: 조건 이미지 디퓨전을 활용한 의미론적 이미지 생성 장치
 I: 인코더부
 D: 디코더부
 U: 조건 이미지 틀 활용부

도면

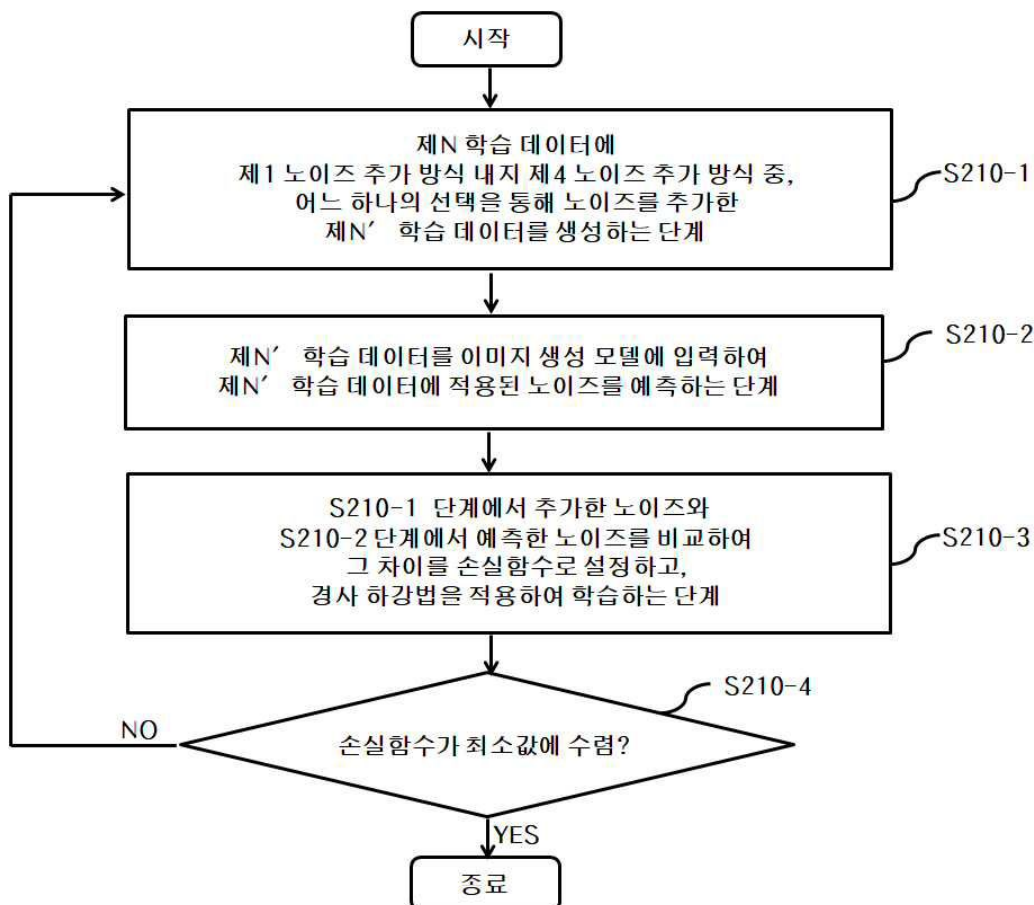
도면1



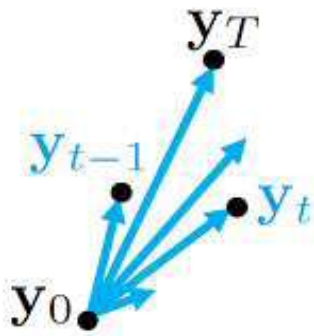
도면2



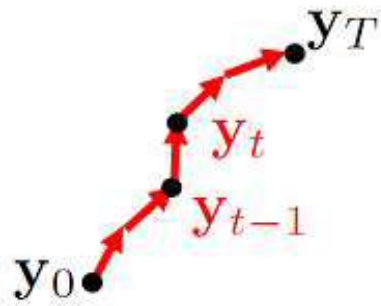
도면3



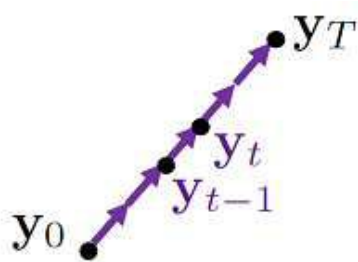
도면4



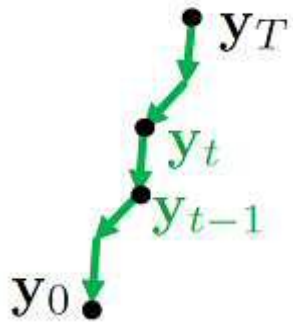
도면5



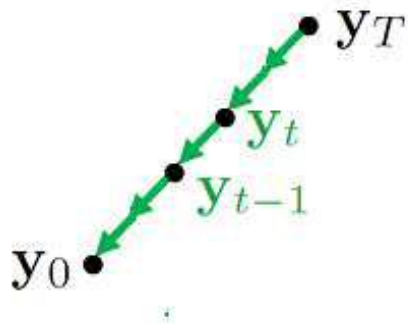
도면6



도면7



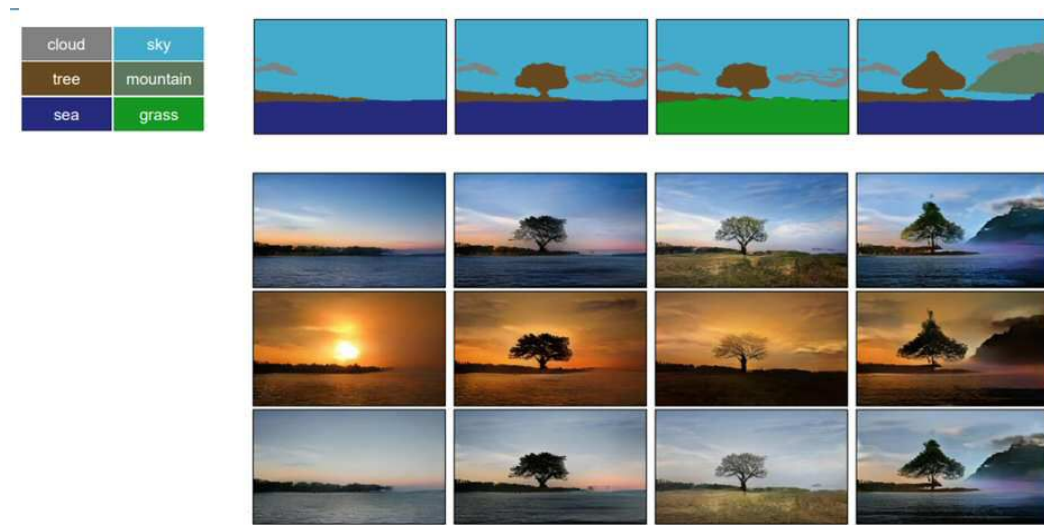
도면8



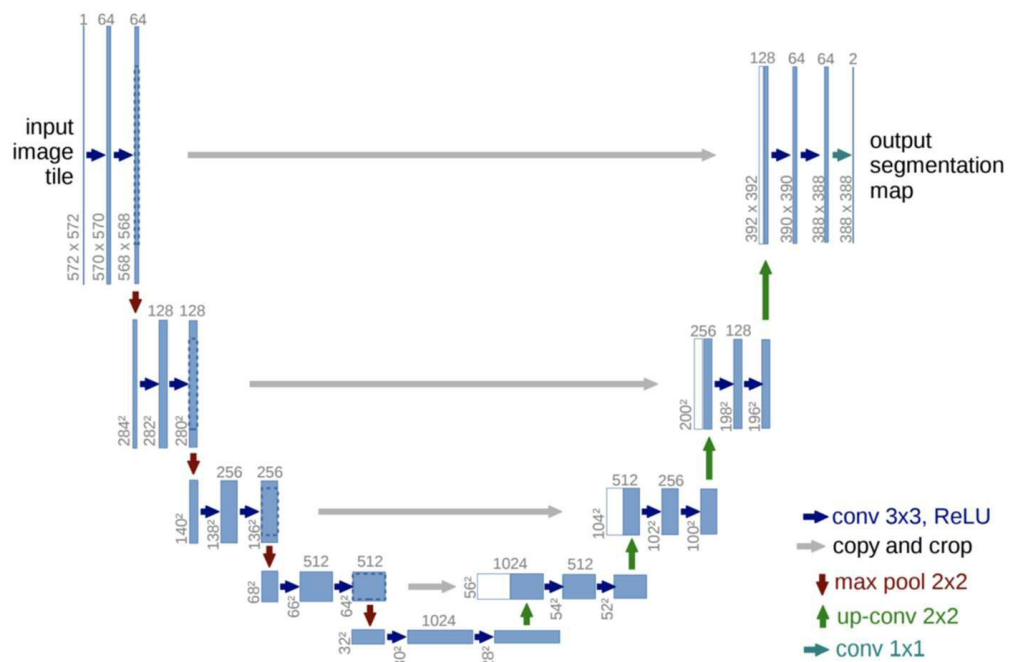
도면9



도면10



도면11



도면12

