



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2024년07월30일
(11) 등록번호 10-2689642
(24) 등록일자 2024년07월25일

(51) 국제특허분류(Int. Cl.)
G06T 3/40 (2024.01) G06N 3/0464 (2023.01)
G06T 5/00 (2024.01)
(52) CPC특허분류
G06T 3/4046 (2024.01)
G06N 3/0464 (2023.01)
(21) 출원번호 10-2023-0033408
(22) 출원일자 2023년03월14일
심사청구일자 2023년03월14일
(56) 선행기술조사문헌
Jie-En Yao et al., "Local Implicit Normalizing Flow for Arbitrary-Scale Image Super-Resolution", FlowarXiv: 2303.05156v1 [cs.CV], (2023.03.09.)*
Mingi Kwon et al., "Diffusion Models already have a Semantic Latent Space", arXiv:2210.10960v1 [cs.CV], (2022.10.20.)*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
고려대학교산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
김현우
서울시 성북구 북악산로 813 (정릉동, 정릉우성아파트) 101동 809호
박도균
경기도 화성시 동탄원천로 348 (능동) 1001호
(74) 대리인
특허법인위솔

전체 청구항 수 : 총 9 항

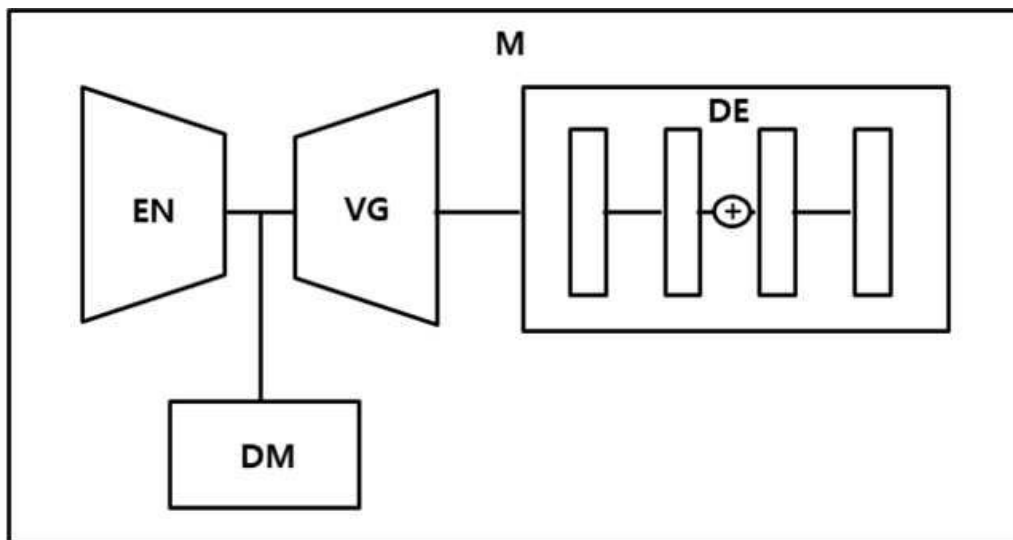
심사관 : 이정은

(54) 발명의 명칭 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 방법 및 이를 위한 장치

(57) 요약

본 발명의 일 실시 예에 따른 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 방법은 (a) 학습 데이터를 입력하여 이미지 생성 모델을 학습시키는 제1 단계 및 (b) 입력 데이터를 상기 학습시킨 이미지 생성 모델에 입력하여 상기 이미지 생성 모델이 생성한 임의의 해상도와 스케일의 이미지를 출력하는 제2 단계를 포함하며, 상기 제1 단계는, 상기 이미지 생성 모델이 포함하는 컨볼루션 신경망(CNN, Convolutional Neural Network) 기반의 인코더부, 컨볼루션 신경망 기반의 잠재 변수 생성부 및 암시적 신경망(INN, Implicit Neural Network) 기반의 디코더부를 학습시키는 제1-1 단계 및 상기 이미지 생성 모델이 포함하는 확산 모델부를 학습시키는 제1-2 단계를 포함한다.

대표도 - 도2



(52) CPC특허분류

G06T 3/4092 (2024.01)

G06T 5/70 (2024.01)

G06T 2207/20081 (2013.01)

G06T 2207/20084 (2013.01)

(72) 발명자

김시현

서울특별시 성북구 고려대로 28길 6 (안암동 5가)
304호

이소진

서울특별시 도봉구 덕릉로53길 10 (창동 대봉아트
빌) 302호

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 임의의 해상도와 스케일의 이미지를 생성하는 방법에 있어서,

(a) 학습 데이터를 입력하여 이미지 생성 모델을 학습시키는 제1 단계; 및

(b) 입력 데이터를 상기 학습시킨 이미지 생성 모델에 입력하여 상기 이미지 생성 모델이 생성한 임의의 해상도와 스케일의 이미지를 출력하는 제2 단계;

를 포함하는 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 방법에 있어서,

상기 제1 단계는,

상기 이미지 생성 모델이 포함하는 컨볼루션 신경망(CNN, Convolutional Neural Network) 기반의 인코더부, 컨볼루션 신경망 기반의 잠재 변수 생성부 및 암시적 신경망(INN, Implicit Neural Network) 기반의 디코더부를 학습시키는 제1-1 단계; 및

상기 이미지 생성 모델이 포함하는 확산 모델부를 학습시키는 제1-2 단계;

를 포함하는 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 방법에 있어서,

상기 제1-1 단계는,

상기 학습 데이터를 상기 인코더부에 입력해 잠재 변수 공간(Image Latent space)으로 매핑시켜 잠재 변수로 변환하는 제1-1-1 단계;

상기 변환한 잠재 변수를 상기 잠재 변수 생성부에 입력해 주파수 정보를 가지고 있는 복수 개의 피쳐맵(Feature Map)으로 분리하는 제1-1-2 단계;

이미지 생성을 위해 필요한 모든 좌표의 좌표값을 상기 분리한 복수 개의 피쳐맵에 쿼리하고, 상기 쿼리한 좌표에 대응되는 피쳐를 상기 디코더부에 입력하는 제1-1-3 단계;

상기 디코더부의 출력값과 학습 데이터 상의 상기 좌표의 RGB값의 차이가 최소화되도록 상기 인코더부, 잠재 변수 생성부 및 디코더부를 학습시키는 제1-1-4 단계;

를 포함하는 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 방법.

청구항 2

삭제

청구항 3

제1항에 있어서,

상기 잠재 변수 공간의 차원은 상기 학습 데이터의 차원보다 작은,

확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 방법.

청구항 4

제1항에 있어서,

상기 복수 개의 피쳐맵은,

하나 이상의 고주파 피쳐맵 및 하나 이상의 저주파 피쳐맵을 포함하는,

확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 방법.

청구항 5

제1항에 있어서,
상기 디코더부의 출력값은,
상기 쿼리한 좌표에 해당하는 이미지의 RGB값인,
확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 방법.

청구항 6

제1항에 있어서,
상기 제1-2 단계는,
상기 변환한 잠재 변수에 노이즈를 추가하여 잠재 변수 z' 를 생성하는 제1-2-1 단계;
상기 생성한 잠재 변수 z' 를 상기 확산 모델에 입력하여 상기 잠재 변수 z' 에 적용된 노이즈를 예측하는 제1-2-2 단계; 및
상기 추가한 노이즈와 예측한 노이즈의 차이가 최소화되도록 상기 확산 모델부를 학습시키는 제1-2-3 단계;
를 포함하는 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 방법.

청구항 7

제1항에 있어서,
상기 제2 단계는,
상기 학습시킨 확산 모델부가 상기 입력 데이터에 적용된 노이즈를 예측하는 제2-1 단계;
상기 입력 데이터에서 상기 예측한 노이즈를 제거하여 잠재 변수로 변환하는 제2-2 단계; 및
상기 변환한 잠재 변수를 상기 잠재 변수 생성부에 입력해 주파수 정보를 가지고 있는 복수 개의 피쳐맵으로 분리하는 제2-3 단계; 및
이미지 생성을 위해 필요한 모든 좌표의 좌표값을 상기 분리한 복수 개의 피쳐맵에 쿼리하고, 이를 상기 학습시킨 디코더부에 입력하여 출력값으로 임의의 해상도와 스케일의 이미지를 출력하는 제2-4 단계;
를 포함하는 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 방법.

청구항 8

제7항에 있어서,
상기 제2-4 단계에서 쿼리하는 좌표의 개수와 위치를 조절하여 상기 출력하는 이미지의 해상도와 스케일을 조절하는,
확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 방법.

청구항 9

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) 학습 데이터를 입력하여 이미지 생성 모델을 학습시키는 제1 오퍼레이션; 및
(B) 입력 데이터를 상기 학습시킨 이미지 생성 모델에 입력하여 상기 이미지 생성 모델이 생성한 임의의 해상도

와 스케일의 이미지를 출력하는 제2 오퍼레이션;

을 실행하는 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 장치에 있어서,
상기 제1 오퍼레이션은,

상기 이미지 생성 모델이 포함하는 컨볼루션 신경망(CNN, Convolutional Neural Network) 기반의 인코더부, 컨볼루션 신경망 기반의 잠재 변수 생성부 및 암시적 신경망(INN, Implicit Neural Network) 기반의 디코더부를 학습시키는 제1-1 오퍼레이션; 및

상기 이미지 생성 모델이 포함하는 확산 모델부를 학습시키는 제1-2 오퍼레이션;

을 포함하는 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 장치에 있어서,
상기 제1-1 오퍼레이션은,

상기 학습 데이터를 상기 인코더부에 입력해 잠재 변수 공간(Image Latent space)으로 매핑시켜 잠재 변수로 변환하는 제1-1-1 오퍼레이션;

상기 변환한 잠재 변수를 상기 잠재 변수 생성부에 입력해 주파수 정보를 가지고 있는 복수 개의 피쳐맵(Feature Map)으로 분리하는 제1-1-2 오퍼레이션;

이미지 생성을 위해 필요한 모든 좌표의 좌표값을 상기 분리한 복수 개의 피쳐맵에 쿼리하고, 상기 쿼리한 좌표에 대응되는 피쳐를 상기 디코더부에 입력하는 제1-1-3 오퍼레이션; 및

상기 디코더부의 출력값과 학습 데이터 상의 상기 좌표의 RGB값의 차이가 최소화되도록 상기 인코더부, 잠재 변수 생성부 및 디코더부를 학습시키는 제1-1-4 오퍼레이션;

를 포함하는 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 장치.

청구항 10

컴퓨팅 장치와 결합하여,

(AA) 학습 데이터를 입력하여 이미지 생성 모델을 학습시키는 제1 단계; 및

(BB) 입력 데이터를 상기 학습시킨 이미지 생성 모델에 입력하여 상기 이미지 생성 모델이 생성한 임의의 해상도와 스케일의 이미지를 출력하는 제2 단계;

를 포함하는 컴퓨터로 판독 가능한 매체에 저장된 컴퓨터 프로그램에 있어서,

상기 제1 단계는,

상기 이미지 생성 모델이 포함하는 컨볼루션 신경망(CNN, Convolutional Neural Network) 기반의 인코더부, 컨볼루션 신경망 기반의 잠재 변수 생성부 및 암시적 신경망(INN, Implicit Neural Network) 기반의 디코더부를 학습시키는 제1-1 단계; 및

상기 이미지 생성 모델이 포함하는 확산 모델부를 학습시키는 제1-2 단계;

를 포함하는 컴퓨터로 판독 가능한 매체에 저장된 컴퓨터 프로그램에 있어서,

상기 제1-1 단계는,

상기 학습 데이터를 상기 인코더부에 입력해 잠재 변수 공간(Image Latent space)으로 매핑시켜 잠재 변수로 변환하는 제1-1-1 단계;

상기 변환한 잠재 변수를 상기 잠재 변수 생성부에 입력해 주파수 정보를 가지고 있는 복수 개의 피쳐맵(Feature Map)으로 분리하는 제1-1-2 단계;

이미지 생성을 위해 필요한 모든 좌표의 좌표값을 상기 분리한 복수 개의 피쳐맵에 쿼리하고, 상기 쿼리한 좌표에 대응되는 피쳐를 상기 디코더부에 입력하는 제1-1-3 단계; 및

상기 디코더부의 출력값과 학습 데이터 상의 상기 좌표의 RGB값의 차이가 최소화되도록 상기 인코더부, 잠재 변수 생성부 및 디코더부를 학습시키는 제1-1-4 단계;

를 포함하는 컴퓨터로 판독 가능한 매체에 저장된 컴퓨터 프로그램.

발명의 설명

기술 분야

- [0001] 본 발명은 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 방법 및 이를 위한 장치에 관한 것이다. 보다 자세하게는 인공지능 기반의 이미지 생성 모델이 학습 데이터에 대한 분포를 학습한 후, 해당 분포로부터 학습 데이터가 아닌 임의의 해상도와 스케일의 새로운 이미지의 다양한 생성이 가능한 방법 및 이를 위한 장치에 관한 것이다.

배경 기술

- [0002] 데이터셋의 분포를 학습하여 해당 분포로부터 학습에 이용되지 않은 새로운 이미지의 샘플링이 가능한 모델을 인공지능 기반의 생성 모델(Generative Model)이라 하며, 생성 모델을 통해 생성된 이미지의 품질과 다양성을 향상시키고자 하는 연구가 활발하게 진행되고 있다.
- [0003] 한편, 대표적인 생성 모델로서 적대적 생성 신경망(GAN, Generative Adversarial Network)를 들 수 있는데, 이는 생성 모델(Generator)과 판별 모델(Discriminator)를 둬으로써 판별 모델은 생성 모델로부터 생성된 가상 이미지와 실제 입력 이미지를 정확하게 구별할 수 있도록 학습되고, 생성 모델은 판별 모델이 구별할 수 없도록 고품질의 가상 이미지를 생성하게 학습되는바, 이를 적대적 학습이라 하며, 이를 통해 적대적 생성 신경망은 고품질의 선명한 이미지를 생성할 수 있다.
- [0004] 그러나 적대적 생성 신경망은 컨볼루션 네트워크(Convolution Network)를 기반으로 설계되어 있기에 학습한 데이터의 해상도 이외의 다른 해상도의 이미지를 생성할 수 없다는 문제점이 있으며, 적대적 학습은 학습이 불안정하여 굉장히 많은 학습 테크닉이 요구되기에 적대적 생성 신경망이 임의의 데이터 도메인에 용이하게 적용되는 것을 저해하고 있고, 생성 모델이 학습한 데이터의 분포를 완벽하게 파악하지 못하고 일부 모드에만 집중함으로써 다양한 이미지가 아니라 유사한 이미지만을 생성하게 되는 모드 콜랩스(Mode Collapse)는 전체적인 다양성을 저해하는 요소로 작용하고 있다.
- [0005] 최근의 인공지능 기반의 생성 모델은 다양한 해상도의 이미지가 다양하게 요구되는 게임, 애니메이션 제작 등과 같은 분야에까지 폭 넓게 적용되려 하고 있으며, 종래의 적대적 생성 신경망은 앞선 문제점들로 인해 적용이 어려운 상황이다. 본 발명은 적대적 생성 신경망의 문제점을 해결하기 위해 제안된 것으로서 학습한 데이터의 해상도뿐만 아니라 임의의 해상도의 이미지를 생성할 수 있으면서 학습 안정도가 높고 생성하는 이미지의 다양성까지 확보할 수 있는 새롭고 진보적인 기술에 관한 것이다.

선행기술문헌

특허문헌

- [0006] (특허문헌 0001) 대한민국 공개특허공보 제 10-2020-0048032호(2020.05.08)

발명의 내용

해결하려는 과제

- [0007] 본 발명이 해결하고자 하는 기술적 과제는 인공지능 기반의 생성 모델이 학습한 데이터의 해상도뿐만 아니라 임의의 해상도의 이미지까지 높은 품질로 생성할 수 있는 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 방법 및 이를 위한 장치를 제공하는 것이다.
- [0008] 본 발명이 해결하고자 하는 또 다른 기술적 과제는 인공지능 기반의 생성 모델의 학습 안정도를 높임으로써 생성하는 이미지의 다양성까지 확보할 수 있는 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 방법 및 이를 위한 장치를 제공하는 것이다.
- [0009] 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

- [0010] 상기 기술적 과제를 달성하기 위한 본 발명의 일 실시 예에 따른 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 방법은 (a) 학습 데이터를 입력하여 이미지 생성 모델을 학습시키는 제1 단계 및 (b) 입력 데이터를 상기 학습시킨 이미지 생성 모델에 입력하여 상기 이미지 생성 모델이 생성한 임의의 해상도와 스케일의 이미지를 출력하는 제2 단계를 포함하며, 상기 제1 단계는, 상기 이미지 생성 모델이 포함하는 컨볼루션 신경망(CNN, Convolutional Neural Network) 기반의 인코더부, 컨볼루션 신경망 기반의 잠재 변수 생성부 및 암시적 신경망(INN, Implicit Neural Network) 기반의 디코더부를 학습시키는 제1-1 단계 및 상기 이미지 생성 모델이 포함하는 확산 모델부를 학습시키는 제1-2 단계를 포함한다.
- [0011] 일 실시 예에 따르면, 상기 제1-1 단계는, 상기 학습 데이터를 상기 인코더부에 입력해 잠재 변수 공간(Image Latent space)으로 매핑시켜 잠재 변수로 변환하는 제1-1-1 단계, 상기 변환한 잠재 변수를 상기 잠재 변수 생성부에 입력해 주파수 정보를 가지고 있는 복수 개의 피쳐맵(Feature Map)으로 분리하는 제1-1-2 단계, 이미지 생성을 위해 필요한 모든 좌표의 좌표값을 상기 분리한 복수 개의 피쳐맵에 쿼리하고, 상기 쿼리한 좌표에 대응되는 피쳐를 상기 디코더부에 입력하는 제1-1-3 단계, 상기 디코더부의 출력값과 학습 데이터 상의 상기 좌표의 RGB값의 차이가 최소화되도록 상기 인코더부, 잠재 변수 생성부 및 디코더부를 학습시키는 제1-1-4 단계를 포함할 수 있다.
- [0012] 일 실시 예에 따르면, 상기 잠재 변수 공간의 차원은 상기 학습 데이터의 차원보다 작을 수 있다.
- [0013] 일 실시 예에 따르면, 상기 복수 개의 피쳐맵은, 하나 이상의 고주파 피쳐맵 및 하나 이상의 저주파 피쳐맵을 포함할 수 있다.
- [0014] 일 실시 예에 따르면, 상기 디코더부의 출력값은, 상기 쿼리한 좌표에 해당하는 이미지의 RGB값일 수 있다.
- [0015] 일 실시 예에 따르면, 상기 제1-2 단계는, 상기 변환한 잠재 변수에 노이즈를 추가하여 잠재 변수'를 생성하는 제1-2-1 단계, 상기 생성한 잠재 변수'를 상기 확산 모델에 입력하여 상기 잠재 변수'에 적용된 노이즈를 예측하는 제1-2-2 단계 및 상기 추가한 노이즈와 예측한 노이즈의 차이가 최소화되도록 상기 확산 모델부를 학습시키는 제1-2-3 단계를 포함할 수 있다.
- [0016] 일 실시 예에 따르면, 상기 제2 단계는, 상기 학습시킨 확산 모델부가 상기 입력 데이터에 적용된 노이즈를 예측하는 제2-1 단계, 상기 입력 데이터에서 상기 예측한 노이즈를 제거하여 잠재 변수로 변환하는 제2-2 단계 및 상기 변환한 잠재 변수를 상기 학습시킨 잠재 변수 생성부에 입력해 주파수 정보를 가지고 있는 복수 개의 피쳐맵으로 분리하는 제2-3 단계 및 이미지 생성을 위해 필요한 모든 좌표의 좌표값을 상기 분리한 복수 개의 피쳐맵에 쿼리하고, 이를 상기 학습시킨 디코더부에 입력하여 출력값으로 임의의 해상도와 스케일의 이미지를 출력하는 제2-4 단계를 포함할 수 있다.
- [0017] 일 실시 예에 따르면, 상기 제2-4 단계에서 쿼리하는 좌표의 개수와 위치를 조절하여 상기 출력하는 이미지의 해상도와 스케일을 조절할 수 있다.
- [0018] 상기 기술적 과제를 달성하기 위한 본 발명의 또 다른 실시 예에 따른 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 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) 학습 데이터를 입력하여 이미지 생성 모델을 학습시키는 제1 오퍼레이션 및 (B) 입력 데이터를 상기 학습시킨 이미지 생성 모델에 입력하여 상기 이미지 생성 모델이 생성한 임의의 해상도와 스케일의 이미지를 출력하는 제2 오퍼레이션을 실행하며, 상기 제1 오퍼레이션은, 상기 이미지 생성 모델이 포함하는 컨볼루션 신경망(CNN, Convolutional Neural Network) 기반의 인코더부, 컨볼루션 신경망 기반의 잠재 변수 생성부 및 암시적 신경망(INN, Implicit Neural Network) 기반의 디코더부를 학습시키는 제1-1 오퍼레이션 및 상기 이미지 생성 모델이 포함하는 확산 모델부를 학습시키는 제1-2 오퍼레이션을 포함한다.
- [0019] 상기 기술적 과제를 달성하기 위한 본 발명의 또 다른 실시 예에 따른 매체에 저장된 컴퓨터 프로그램은 컴퓨팅 장치와 결합하여, (AA) 학습 데이터를 입력하여 이미지 생성 모델을 학습시키는 제1 단계 및 (BB) 입력 데이터를 상기 학습시킨 이미지 생성 모델에 입력하여 상기 이미지 생성 모델이 생성한 임의의 해상도와 스케일의 이미지를 출력하는 제2 단계를 포함하며, 상기 제1 단계는, 상기 이미지 생성 모델이 포함하는 컨볼루션 신경망(CNN, Convolutional Neural Network) 기반의 인코더부, 컨볼루션 신경망 기반의 잠재 변수 생성부 및 암시적

신경망(INN, Implicit Neural Network) 기반의 디코더부를 학습시키는 제1-1 단계 및 상기 이미지 생성 모델이 포함하는 확산 모델부를 학습시키는 제1-2 단계를 포함한다.

발명의 효과

- [0020] 상기와 같은 본 발명에 따르면, 인공지능 기반의 이미지 생성 모델이 암시적 신경망 기반의 디코더를 포함하는 바, 쿼리하는 좌표의 개수와 위치를 조절함으로써 학습한 데이터의 해상도뿐만 아니라 임의의 해상도의 이미지까지 높은 품질로 생성할 수 있다는 효과가 있다.
- [0021] 또한, 인공지능 기반의 이미지 생성 모델이 확산 모델 기반의 확산 모델부를 포함하는바, 적대적 생성 신경망 대비 안정적인 학습이 가능하며 생성하는 이미지의 다양성까지 확보할 수 있다는 효과가 있다.
- [0022] 또한, 인공지능 기반의 이미지 생성 모델이 계층적 임베딩을 활용하여 이미지의 저주파수 특징과 고주파수 특징을 분리하는 잠재 변수 생성부를 포함하는바, 고해상도 이미지에서 나타나는 고주파수 특징을 보다 뚜렷하게 생성할 수 있다는 효과가 있다.
- [0023] 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

- [0024] 도 1은 본 발명의 제1 실시 예에 따른 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 장치가 포함하는 전체 구성을 예시적으로 도시한 도면이다.
- 도 2는 본 발명의 제1 실시 예에 따른 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 장치가 포함하는 프로세서에 구현된 인공지능 기반의 이미지 생성 모델의 세부 구성을 예시적으로 도시한 도면이다.
- 도 3은 본 발명의 제2 실시 예에 따른 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 방법의 대표적인 단계를 나타낸 순서도이다.
- 도 4는 본 발명의 제2 실시 예에 따른 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 방법에 있어서, 이미지 생성 모델을 학습시키는제1-1 단계인 S310-1 단계를 구체화한 순서도이다.
- 도 5는 학습 데이터의 분포가 인코더부를 통해 잠재 변수 공간에서의 분포로 변환되는 모식도를 예시적으로 도시한 도면이다.
- 도 6은 잠재 변수가 잠재 변수 생성부를 통해 주파수 정보를 가지고 있는 복수 개의 피쳐맵으로 분리되는 모식도를 예시적으로 도시한 도면이다.
- 도 7은 이미지 생성을 위해 필요한 특정 좌표의 좌표값을 하나의 피쳐맵에 쿼리하고, 이를 디코더부에 입력하는 모식도를 예시적으로 도시한 도면이다.
- 도 8은 앞선 도 7에서 디코더부의 출력값인 쿼리한 모든 좌표에 해당하는 이미지의 RGB 값을 취합한 모습을 예시적으로 도시한 도면이다.
- 도 9는 학습 데이터 입력에 따라 학습 과정에서의 출력 이미지가 출력되는 모습을 예시적으로 도시한 도면이다.
- 도 10은 본 발명의 제2 실시 예에 따른 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 방법에 있어서, 이미지 생성 모델을 학습시키는제1-2 단계인 S310-2 단계를 구체화한 순서도이다.
- 도 11은 노이즈 스케줄에 따라 잠재 변수에 노이즈를 추가함으로써 잠재 변수'인 데이터의 특징 분포가 점차 단순화되는 과정을 예시적으로 도시한 도면이다.
- 도 12는 쿼리하는 좌표의 개수와 위치를 조절함으로써 출력하는 이미지의 해상도와 스케일을 조절하는 모습을 예시적으로 도시한 도면이다.
- 도 13은 본 발명의 제2 실시 예에 따른 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 방법 학습 데이터와 이를 기초로 실제 실제 생성하는 이미지를 나타낸 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0025] 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.
- [0026] 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.
- [0027] 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.
- [0028] 또한, 어떤 구성요소들을 포함한다는 표현은 "개방형"의 표현으로서 해당 구성요소들이 존재하는 것을 단순히 지칭할 뿐이며, 추가적인 구성요소들을 배제하는 것으로 이해되어서는 안 된다.
- [0029] 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.
- [0030] 이하에서는 도면들을 참조하여 본 발명의 세부적인 실시 예들에 대해 살펴보도록 한다.
- [0031] 도 1은 본 발명의 제1 실시 예에 따른 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 장치(100)가 포함하는 전체 구성을 예시적으로 도시한 도면이다.
- [0032] 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.
- [0033] 본 발명의 제1 실시 예에 따른 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 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.
- [0034] 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 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 방법을 수행하기 위한 적어도 하나의 애플리케이션 또는 프로그램에 대한 연산을 수행할 수 있는바, 이를 위해 프로세서(10)에는 소정의 구조를 나타내는 확산 모델과 암시적 신경망 기반의 이미지 생성 모델을 포함할 수 있으나, 이에 대해서는 후술 하도록 한다.
- [0035] 네트워크 인터페이스(20)는 본 발명의 제1 실시 예에 따른 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 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.
- [0036] 메모리(30)는 각종 정보, 명령 및/또는 정보를 저장하며, 본 발명의 제2 실시 예에 따른 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 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.
- [0037] 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.
- [0038] 컴퓨터 프로그램(41)은 메모리(30)에 로드되어, 하나 이상의 프로세서(10)에 의해, (A) 학습 데이터를 입력하여

이미지 생성 모델을 학습시키는 제1 오퍼레이션 및 (B) 입력 데이터를 상기 학습시킨 이미지 생성 모델에 입력하여 상기 이미지 생성 모델이 생성한 임의의 해상도와 스케일의 이미지를 출력하는 제2 오퍼레이션을 실행할 수 있다.

- [0039] 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 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 방법에 대한 설명에서 후술하도록 한다.
- [0040] 데이터 버스(50)는 이상 설명한 프로세서(10), 네트워크 인터페이스(20), 메모리(30) 및 스토리지(40) 사이의 명령 및/또는 정보의 이동 경로가 된다.
- [0041] 이상 간단하게 설명한 본 발명의 제1 실시 예에 따른 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 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.
- [0042] 도 2에는 이상 간단하게 설명한 본 발명의 제1 실시 예에 따른 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 장치(100)가 포함하는 프로세서(10)에 구현된 인공지능 기반의 이미지 생성 모델(M)의 세부 구성을 예시적으로 도시하였다. 도 2를 참조하면, 이미지 생성 모델(M)은 컨볼루션 신경망(CNN, Convolutional Neural Network) 기반의 인코더부(EN), 컨볼루션 신경망(CNN, Convolutional Neural Network) 기반의 잠재 변수 생성부(VG), 암시적 신경망(INN, Implicit Neural Network) 기반의 디코더부(DE) 및 확산 모델부(DM)를 포함하고 있음을 확인할 수 있으며, 이에 대한 자세한 설명은 본 발명의 제2 실시 예에 따른 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 방법에 대한 설명에서 후술하도록 한다.
- [0043] 이하, 본 발명의 제1 실시 예에 따른 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 장치(100)가 독립된 디바이스인 전자 기기 형태 중, "서버"의 형태임을 전제로 임의의 해상도와 스케일의 이미지를 생성하고자 하는 사용자의 사용자 단말(미도시)에 설치된 전용 어플리케이션을 통해 본 발명의 제2 실시 예에 따른 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 방법을 제공하는 과정에 대하여 도 3 내지 도 12를 참조하여 설명하도록 한다.
- [0044] 도 3은 본 발명의 제2 실시 예에 따른 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 방법의 대표적인 단계를 나타낸 순서도이다.
- [0045] 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.
- [0046] 한편, 각 단계는 본 발명의 제1 실시 예에 따른 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 장치(100)를 통해 이루어지는 것을 전제로 하며, 본 발명의 제1 실시 예에 따른 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 장치(100)가 "서버"의 형태임을 전제로 하였기 때문에 사용자 단말(미도시)에 설치된 전용 어플리케이션을 본 발명의 제1 실시 예에 따른 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 장치(100)와 동일한 의미로 볼 것이고, 설명의 편의를 위해 이들 모두를 "장치(100)"로 명명하도록 한다.
- [0047] 우선, 장치(100)가 학습 데이터를 입력하여 이미지 생성 모델(M)을 학습시키며(S310), 이를 제1 단계라 한다.
- [0048] 여기서 이미지 생성 모델(M)은 확산 모델과 암시적 신경망 기반의 이미지 생성 모델을 의미하며, 제1 단계는 이미지 생성 모델(M)이 포함하는 컨볼루션 신경망(CNN, Convolutional Neural Network) 기반의 인코더부(EN), 컨볼루션 신경망 기반의 잠재 변수 생성부(VG) 및 암시적 신경망(INN, Implicit Neural Network) 기반의 디코더부(DE)를 학습시키는 제1-1 단계(S310-1) 및 이미지 생성 모델(M)이 포함하는 확산 모델부(DM)를 학습시키는 제1-2 단계(S310-2)를 포함할 수 있다. 이 중에서 제1-1 단계가 암시적 신경망을 활용하여 임의의 해상도의 이미지를 생성할 수 있음에 기여하는 단계이며, 제1-2 단계가 안정적인 학습을 통해 생성하는 이미지의 다양성을 확보할 수 있음에 기여하는 단계인바, 이하 제1-1 단계인 S310-1 단계 부터 자세히 설명하도록 한다.
- [0049] 도 4는 본 발명의 제2 실시 예에 따른 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를

생성하는 방법에 있어서, 이미지 생성 모델(M)을 학습시키는 제1-1 단계인 S310-1 단계를 구체화한 순서도이다.

- [0050] 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.
- [0051] 우선, 학습 데이터를 인코더부(EN)에 입력해 잠재 변수 공간(Image Latent space)으로 매핑시켜 잠재 변수로 변환하며(S310-1-1), 이를 제1-1-1 단계라 한다.
- [0052] 여기서 학습 데이터는 이미지 생성 모델(M)을 학습시키기 위한 임의의 입력 이미지를 가정으로 설정하며, 잠재 변수는 생성 모델(Generator)이 사용하는 입력값으로 데이터의 특징(피쳐, Feature)을 분포로 나타낸 것 또는 학습 데이터인 임의의 입력 이미지에 대한 이미지 코드로 이해할 수 있고, 이를 통해 새로운 데이터를 생성할 수 있으며, 잠재 변수 공간은 잠재 변수가 가질 수 있는 모든 값을 가진 공간, 보다 구체적으로 잠재 변수를 축으로 하는 고차원 공간을 의미하는바, 모두 생성 모델에서 보편적으로 사용하는 의미 그대로 해석하면 충분하다 할 것이다.
- [0053] 학습 데이터는 인코더부(EN)에 입력됨으로써 잠재 변수로 변환되는바, 이는 입력 이미지인 학습 데이터의 분포를 인코더부(EN)를 통해 학습 데이터의 유의미한 정보를 보존하고 있는 잠재 변수 공간에서의 분포로 변환시키는 것이며, 잠재 변수 공간의 차원은 학습 데이터의 차원보다 작다.
- [0054] 도 5는 학습 데이터의 분포가 인코더부(EN)를 통해 잠재 변수 공간에서의 분포로 변환되는 모식도를 예시적으로 도시한바, 인코더부(EN)에 입력되는 학습 데이터의 분포가 다소 복잡하게 이루어져 있음에 반해, 인코더부(EN)에서 출력된 잠재 변수 공간에서의 분포는 이보다 훨씬 단순한 형태, 즉 작은 차원의 분포임을 확인할 수 있으며, 이는 입력 이미지인 학습 데이터를 표현하기에 필요한 최소한의 특징에 대한 분포로 이해할 수 있을 것이다.
- [0055] 이후, 변환한 잠재 변수를 잠재 변수 생성부(VG)에 입력해 주파수 정보를 가지고 있는 복수 개의 피쳐맵(Feature Map)으로 분리하며(S310-1-2), 이를 제1-1-2 단계라 한다.
- [0056] 잠재 변수 생성부(VG)는 컨볼루션 신경망 기반의 네트워크로 구성되어 있기에 변환된 잠재 변수를 다양한 주파수 정보를 가지고 있는 복수 개의 피쳐맵으로 분리 가능하며, 이는 학습 데이터인 입력 이미지의 고주파수(High-frequency)와 저주파수(Low-frequency) 특징을 잠재 변수 공간으로 분리시키는 것이다. 그에 따라 복수 개의 피쳐맵은 하나 이상의 고주파 피쳐맵 및 하나 이상의 저주파 피쳐맵을 포함할 수 있다.
- [0057] 도 6은 잠재 변수가 잠재 변수 생성부(VG)를 통해 주파수 정보를 가지고 있는 복수 개의 피쳐맵으로 분리되는 모식도를 예시적으로 도시한바, 컨볼루션 네트워크의 계층적 특징을 이용하여 네트워크의 중간 단계와 마지막 단계에서 피쳐맵을 산출할 수 있으며, 각각의 피쳐맵을 고주파수 특징과 저주파수 특징을 담당하는 잠재 변수로 사용할 수 있다.
- [0058] 이와 같이 본 발명의 제2 실시 예에 따른 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 방법은 하나의 입력 이미지에 대하여, 보다 구체적으로 하나의 입력 이미지를 변환한 잠재 변수를 복수 개의 피쳐맵으로 분리하기에 다양한 해상도에서 나타내는 다양한 정보를 다채롭게 표현하여 이미지를 생성할 수 있으며, 고해상도의 이미지를 생성할 때 필요한 고주파수 특징을 독립적으로 학습 및 생성시킬 수 있기 때문에 고품질의 고해상도 이미지까지 생성할 수 있다.
- [0059] 다음으로 이미지 생성을 위해 필요한 모든 좌표의 좌표값을 분리한 복수 개의 피쳐맵에 쿼리하고, 쿼리한 좌표에 대응되는 피쳐를 디코더부(DE)에 입력하며(S310-1-3), 이를 제1-1-3 단계라 한다.
- [0060] 이는 쿼리된 좌표에 대응되는 하나의 피쳐를 피쳐맵 내에서 쌍선형 보간법(Bilinear Interpolation)을 통해 추출하고, 이를 암시적 신경망 기반의 디코더부(DE)에 입력하는 것이다.
- [0061] 여기서 쌍선형 보간법이란 두 개의 변수에 대한 함수값을 알고 있을 때, 이를 이용하여 다른 변수에 대한 함수값을 추정하는 보간법 중 하나, 쉽게 이야기하면 직사각형의 네 꼭지점에서의 값이 주어져 있을 때 이 사각형의 변 및 내부의 임의의 점에서의 값을 추정하는 것인바, 1차원 공간에서 사용하는 선형 보간법을 2차원 공간에서도 사용하기 위해 확장한 보간법이다.
- [0062] 도 7은 이미지 생성을 위해 필요한 특정 좌표의 좌표값을 하나의 피쳐맵에 쿼리하고, 이를 디코더부(DE)에 입력하는 모식도를 예시적으로 도시한바, 서로 인접한 네 개의 좌표는 하나의 사각형으로 볼 수 있기에 쌍선형 보간법을 적용하여 해당 사각형의 변 및 내부의 모든 점에서의 좌표값 역시 산출하여 피쳐맵에 쿼리하고 디코더부

(DE)에 입력할 수 있을 것이다.

- [0063] 디코더부(DE)에 입력까지 했다면, 디코더부(DE)의 출력값과 학습 데이터 상의 상기 좌표의 RGB값의 차이가 최소화되도록 인코더부(EN), 잠재 변수 생성부(VG) 및 디코더부(DE)를 학습시키며(S310-1-4), 이를 제1-1-4 단계라 한다.
- [0064] 앞서, 디코더부(DE)에 입력한 것이 쿼리한 좌표에 대한 추출된 피처이므로 디코더부(DE)의 출력값은 쿼리한 좌표에 해당하는 이미지의 RGB 값일 수 있으며, 이를 이미지 생성을 위해 필요한 모든 좌표의 좌표값에 대하여 수행하기 때문에 디코더부(DE)의 출력값 전부를 취합하면 하나의 완성된 이미지가 될 수 있다.
- [0065] 도 8은 앞선 도 7에서 디코더부(DE)의 출력값인 쿼리한 모든 좌표에 해당하는 이미지의 RGB 값을 취합한 모습을 예시적으로 도시한바, 모든 좌표에 대하여 RGB값이 부여되기 때문에 이를 취합하여 강아지 얼굴 이미지가 생성된 것을 확인할 수 있다.
- [0066] 이처럼 S310-1-3 단계까지 진행됨으로써 생성하고자 하는 이미지가 생성되는바, 이를 출력 이미지라 한다면 S310-1-4 단계에서의 학습은 학습 데이터인 임의의 입력 이미지와 출력 이미지의 차이가 최소화될 수 있는, 보다 구체적으로 동일한 좌표에 대하여 디코더부(DE)의 출력값(RGB값)과 학습 데이터 상의 RGB값의 차이가 최소화되도록 인코더부(EN), 잠재 변수 생성부(VG) 및 디코더부(DE)를 학습시키는 것으로볼 수 있는바, 도 9는 학습 데이터인 임의의 입력 이미지(강아지 얼굴 이미지)를 인코더부(EN)에 입력하고, 잠재 변수 생성부(VG)를 지나 디코더부(DE)에 의해 출력 이미지인 강아지 얼굴 이미지가 출력되는 모습을 예시적으로 도시한 도면이며, 맨 좌측의 입력 이미지인 강아지 얼굴 이미지와 맨 우측의 출력 이미지인 강아지 얼굴 이미지의 차이가 최소화될 수 있도록 학습시키는 것이 S310-1-4 단계에서의 학습이다.
- [0067] 한편, 본 발명의 제2 실시 예에 따른 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 방법은 인코더부(EN)를 통해 변환된 잠재 변수의 분산이 지나치게 높아지지 않고 정규 분포에 가까워지도록 제약을 주는 추가적인 손실 함수를 통해 잠재 변수 공간이 매끄러운(Smooth) 공간이 되도록 하여 학습의 효율성을 향상시킬 수도 있을 것이다.
- [0068] 이상 제1-1-1 단계(S310-1-1 단계) 내지 제1-1-4 단계(S310-1-4 단계)에 대한 설명을 통해 제1-1 단계(S310-1)에서의 학습에 대한 설명을 마무리하며, 제1-1 단계에서의 학습이 완료되면 확산 모델(DM)을 학습시키는 제1-2 단계인 S310-2 단계에서의 학습이 개시된다. 이하 설명하도록 한다.
- [0069] 도 10은 본 발명의 제2 실시 예에 따른 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 방법에 있어서, 이미지 생성 모델(M)을 학습시키는제1-2 단계인 S310-2 단계를 구체화한 순서도이다.
- [0070] 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.
- [0071] 우선, 변환한 잠재 변수에 노이즈를 추가하여 잠재 변수'를 생성하며(S310-2-1), 이를 제1-2-1 단계라 한다.
- [0072] 여기서 노이즈는 사전에 기 설정된 노이즈 스케줄에 따른 노이즈일 수 있으며, 도 11에 노이즈 스케줄에 따라 잠재 변수에 노이즈를 추가함으로써 잠재 변수'인 데이터의 특징 분포가 점차 단순화되는 과정을 예시적으로 도시하였다.
- [0073] 이후, 생성한 잠재 변수'를 확산 모델부(DM)에 입력하여 잠재 변수'에 적용된 노이즈를 예측하며(S310-2-2), 이를 제1-2-2 단계라 한다.
- [0074] 이는 잠재 변수'에 적용된 노이즈를 예측하고, 예측한 만큼을 잠재 변수'로부터 제거한다면 노이즈 추가 전의 데이터인 잠재 변수를 산출할 수 있기 때문이며 확산 모델부(DM) 학습 과정에서의 특징에 해당하는바, 확산 모델에서 노이즈를 추가하는, 즉 디퓨전(Diffusion)을 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 위해서이며, 데이터에 노이즈를 추가하여 적정량 퍼트리게 되면 확률값이 낮았던 데이터가 줄어들게 되고, 해당 부분에 대한 학습 문제점이 완화될 수 있기 때문이다.
- [0075] 마지막으로, 추가한 노이즈와 예측한 노이즈의 차이가 최소화되도록 확산 모델부(DM)를 학습시키며(S310-2-3), 이를 제1-2-3 단계라 한다.
- [0076] 장치(100)는 제1-2-1 단계에서 노이즈 스케줄에 따라 잠재 변수에 노이즈를 추가하였으므로 얼마만큼의 노이즈

를 추가했는지에 대한 정보를 보유하고 있으며, 이를 정답 노이즈로, 제1-2-3 단계에서 예측한 노이즈를 예측 노이즈라 한다면, 그 차이를 손실함수 손실함수는 $|(\text{정답 노이즈}) - (\text{예측 노이즈})|^2$ 로 설정하고 경사 하강법 등의 학습 방법을 적용해 차이가 최소화되도록 학습시키는 것이다. 그에 따라 제1-2 단계에서 확산 모델부(DM)의 역할은 잠재 변수 공간으로 변환된 특징 분포를 사전 정의된 샘플링이 간단한 특징 분포로 변환시키는 것이라 할 것이다.

[0077] 확산 모델부(DM)의 학습이 완료되면, 노이즈가 추가된 단순한 특징 분포를 갖는 잠재 변수 또는 입력 데이터가 입력되었을 때, 얼마만큼의 노이즈가 추가된 상태인지를 예측하고, 예측한 노이즈를 제거함으로써 노이즈가 추가되기 전 복잡한 특징 분포를 갖는의 잠재 변수 또는 입력 데이터를 생성해낼 수 있는바, 이를 앞서 설명한 제1-1-2 단계 내지 제1-1-3 단계에 따라 잠재 변수 생성부(VG)에 입력해 복수 개의 피쳐맵으로 분리하고, 좌표의 좌표값을 피쳐맵에 쿼리하여 쿼리한 좌표에 대응되는 피쳐를 디코더부(DE)에 입력함으로써 출력 이미지를 생성할 수 있을 것이다.

[0078] 이상 제1-2-1 단계(S310-2-1 단계) 내지 제1-2-3 단계(S310-2-3 단계)에 대한 설명을 통해 제1-2 단계(S310-2)에서의 학습에 대한 설명까지 마무리하며, 이미지 생성 모델(M)의 학습이 완료되었다는 전제 하에, 다시 도 2에 대한 설명으로 돌아가도록 한다.

[0079] 이미지 생성 모델을 학습시켰다면, 장치(100)가 입력 데이터를 학습시킨 이미지 생성 모델(M)에 입력하여 이미지 생성 모델(M)이 생성한 임의의 해상도와 스케일의 이미지를 출력하며(S320), 이를 제2 단계라 한다.

[0080] 이러한 제2 단계는 학습이 완료된 이미지 생성 모델(M)을 이용하여 실제 출력 이미지를 얻는 과정, 즉 샘플링 과정을 의미하는바, 앞서 제1 단계인 학습 과정에서의와 상이하게 샘플링 과정에서는 인코더부(EN)의 동작이 요구되지 않는다.

[0081] 제2 단계는 출력 이미지를 생성하는 과정이기에 기본적으로 앞서 설명한 제1-1 단계와 유사하나, 확산 모델부(DM)의 학습이 완료된 상태이기에 학습시킨 확산 모델부(DM)가 입력 데이터에 적용된 노이즈를 예측하는 S320-1 단계인 제2-1 단계가 시작 단계가 되며, 정규 분포를 따르는 임의의 분포인 사전 정의된 분포로부터 노이즈를 샘플링하는 단계로 볼 수 있고, 여기서 사전 정의된 분포는 앞서 제1-2-3 단계에서 언급한 사전 정의된 샘플링이 간단한 특징 분포라 할 것이다.

[0082] 그 이후 입력 데이터에서 예측한 노이즈를 제거하여 잠재 변수로 변환하는 S320-2 단계인 제2-2 단계가 수행되며, 노이즈를 제거하기 때문에 확산 모델부(DM)가 동작하며, 임의의 노이즈에서 순차적으로 노이즈를 제거하여 노이즈가 포함되지 않은 잠재 변수를 변환시키는 과정이라 할 것이다.

[0083] 그 이후에 S310-1-2 단계인 제1-1-2 단계에 대응되는 S320-3 단계인 제2-3 단계(변환한 잠재 변수를 학습시킨 잠재 변수 생성부(VG)에 입력해 주파수 정보를 가지고 있는 복수 개의 피쳐맵으로 분리하는 단계)와 S310-1-3 단계인 제1-1-3 단계에 대응되는 S320-4 단계인 제2-4 단계(이미지 생성을 위해 필요한 모든 좌표의 좌표값을 분리한 복수 개의 피쳐맵에 쿼리하고, 이를 학습시킨 디코더부(DE)에 입력하여 출력값으로 임의의 해상도와 스케일의 이미지를 출력하는 단계)가 수행됨으로써 출력 이미지를 출력할 수 있으며, 중복 서술을 방지하기 위해 자세한 설명은 생략하도록 한다.

[0084] 한편, 본 발명의 제2 실시 예에 따른 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 방법은 S320-4 단계인 제2-4 단계에서 쿼리하는 좌표의 개수와 위치를 조절함으로써 출력하는 이미지의 해상도와 스케일을 조절할 수 있는바, 도 12에 이를 예시적으로 도시해 놓았다.

[0085] 지금까지 본 발명의 제2 실시 예에 따른 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 방법에 대하여 설명하였다. 본 발명에 따르면, 인공지능 기반의 이미지 생성 모델(M)이 암시적 신경망 기반의 디코더(DE)를 포함하는바, 쿼리하는 좌표의 개수와 위치를 조절함으로써 학습한 데이터의 해상도뿐만 아니라 임의의 해상도의 이미지까지 높은 품질로 생성할 수 있다. 또한, 인공지능 기반의 이미지 생성 모델(M)이 확산 모델 기반의 확산 모델부(DM)를 포함하는바, 적대적 생성 신경망 대비 안정적인 학습이 가능하며 생성하는 이미지의 다양성까지 확보할 수 있다. 더 나아가 인공지능 기반의 이미지 생성 모델(M)이 계층적 임베딩을 활용하여 이미지의 저주파수 특징과 고주파수 특징을 분리하는 잠재 변수 생성부(VG)를 포함하는바, 고해상도 이미지에서 나타나는 고주파수 특징을 보다 뚜렷하게 생성할 수 있다.

[0086] 도 13은 본 발명의 제2 실시 예에 따른 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 방법에 대하여 학습 데이터인 고양이 얼굴 이미지(256 * 256)로 학습시킨 후, 출력 이미지이자 해상도

와 스케일이 변경된 또 다른 고양이 얼굴 이미지(1024 * 1024)를 도시한바, 학습 데이터와 상이한 해상도의 이미지를 생성할 수 있을 뿐만 아니라 생성하는 이미지의 다양성까지 확보함을 확인할 수 있다.

[0087] 마지막으로, 본 발명의 제1 실시 예에 따른 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 장치(100)와 본 발명의 제2 실시 예에 따른 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 방법은 본 발명의 제3 실시 예에 따른 컴퓨터로 판독 가능한 매체에 저장된 컴퓨터 프로그램으로 구현할 수도 있는바, 이 경우 컴퓨팅 장치와 결합하여 (AA) 학습 데이터를 입력하여 이미지 생성 모델을 학습시키는 제1 단계 및 (BB) 입력 데이터를 상기 학습시킨 이미지 생성 모델에 입력하여 상기 이미지 생성 모델이 생성한 임의의 해상도와 스케일의 이미지를 출력하는 제2 단계를 포함하되, 상기 제1 단계는, 상기 이미지 생성 모델이 포함하는 컨볼루션 신경망(CNN, Convolutional Neural Network) 기반의 인코더부, 컨볼루션 신경망 기반의 잠재 변수 생성부 및 암시적 신경망(INN, Implicit Neural Network) 기반의 디코더부를 학습시키는 제1-1 단계 및 상기 이미지 생성 모델이 포함하는 확산 모델을 학습시키는 제1-2 단계를 실행할 수 있을 것이며, 중복 서술을 위해 자세히 기재하지는 않았지만 본 발명의 제1 실시 예에 따른 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 장치(100) 및 본 발명의 제2 실시 예에 따른 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 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.

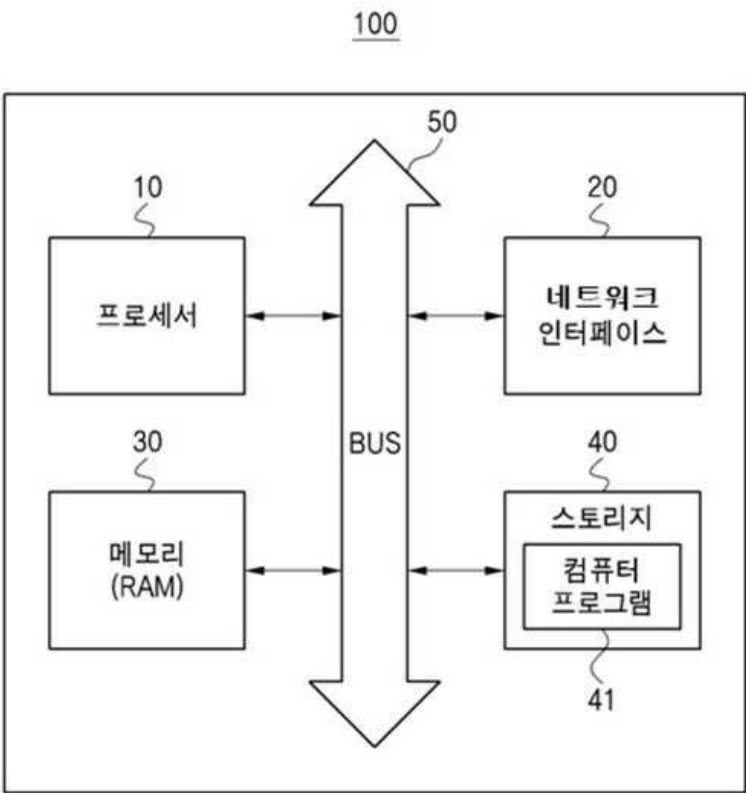
[0088] 이상 첨부된 도면을 참조하여 본 발명의 실시 예들을 설명하였지만, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 본 발명이 그 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 실시될 수 있다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시 예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다.

부호의 설명

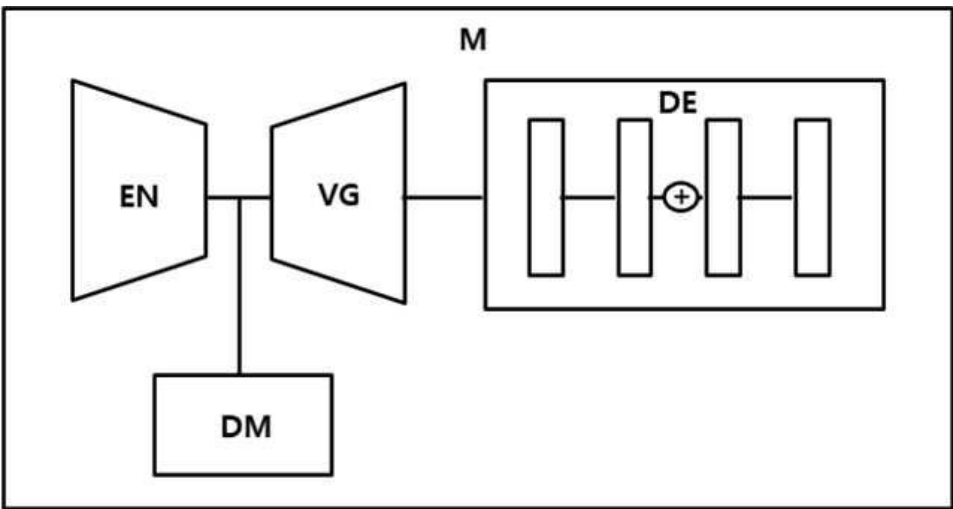
[0089] 10: 프로세서
20: 네트워크 인터페이스
30: 메모리
40: 스토리지
41: 컴퓨터 프로그램
50: 정보 버스
100: 확산 모델과 암시적 신경망 기반의 임의의 해상도와 스케일의 이미지를 생성하는 장치
M: 이미지 생성 모델
EN: 인코더부
VG: 잠재 변수 생성부
DE: 디코더부
HM: 확산 모델부

도면

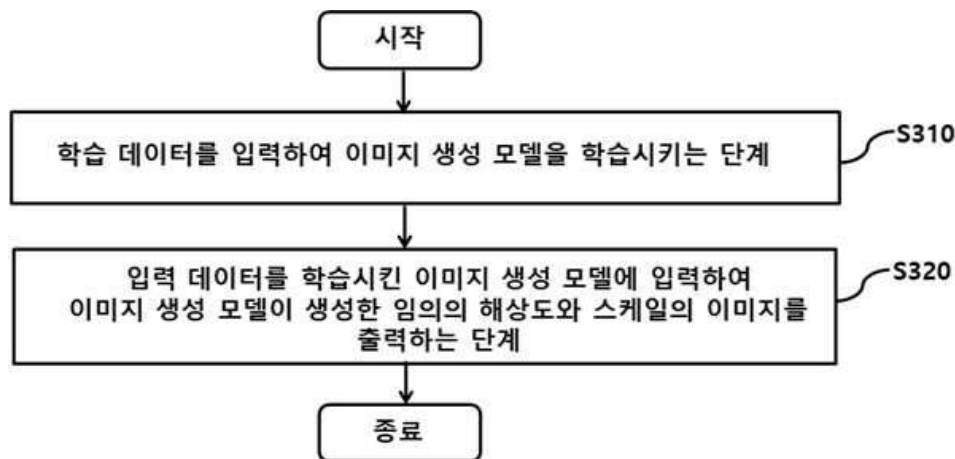
도면1



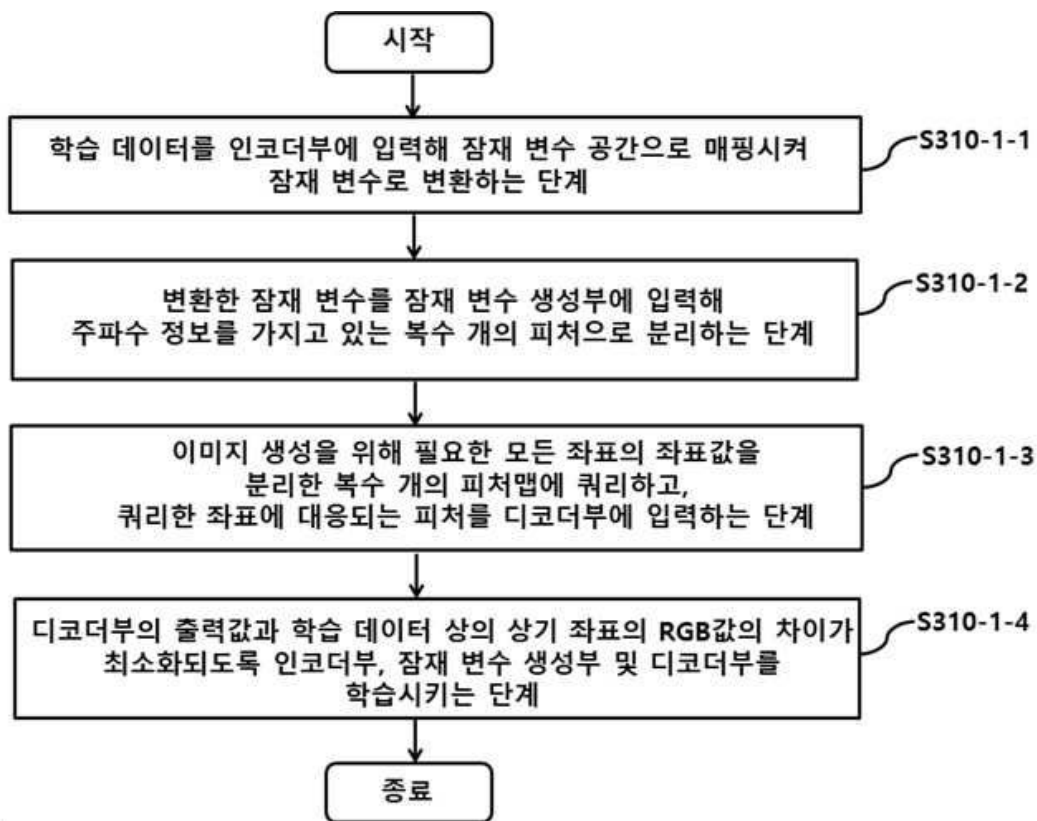
도면2



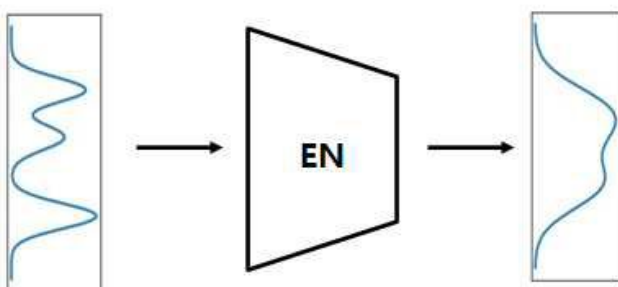
도면3



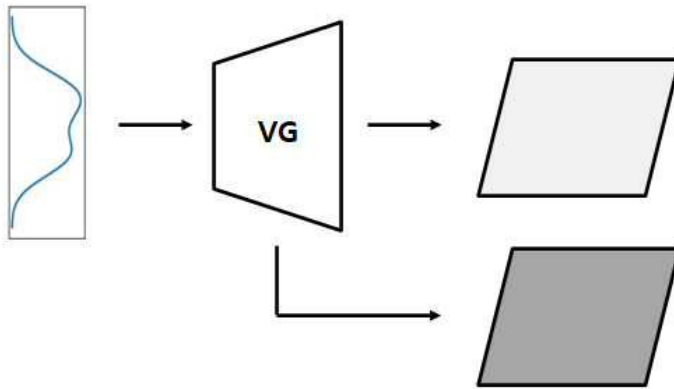
도면4



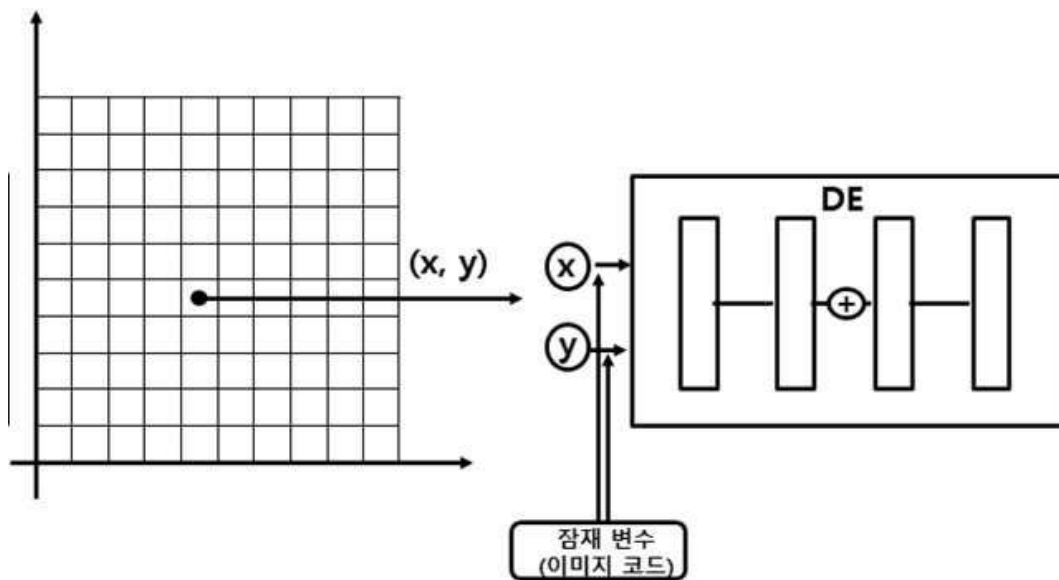
도면5



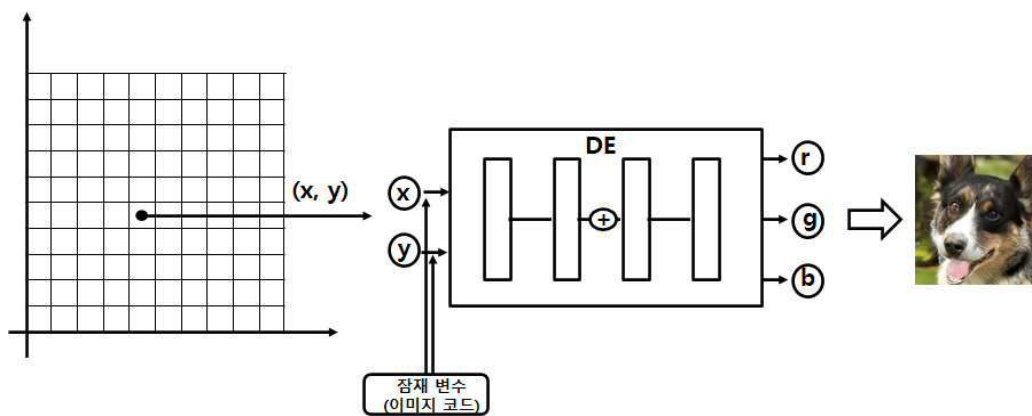
도면6



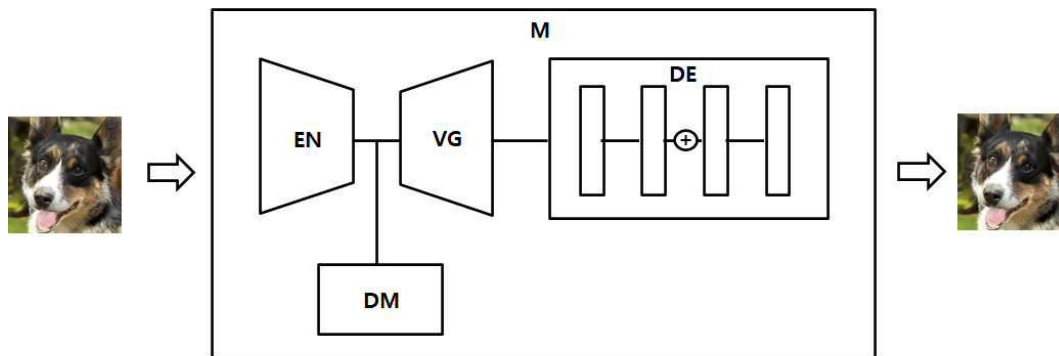
도면7



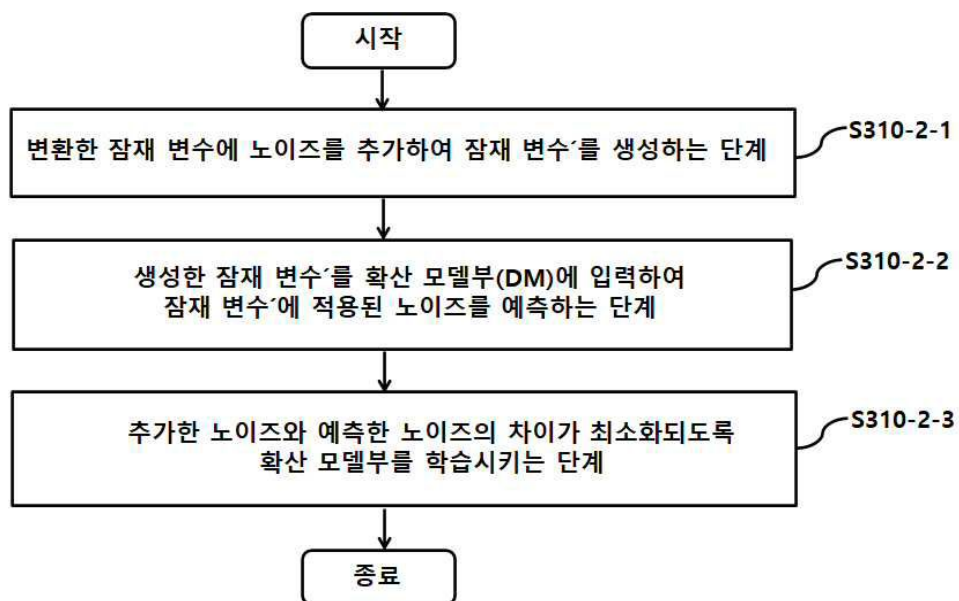
도면8



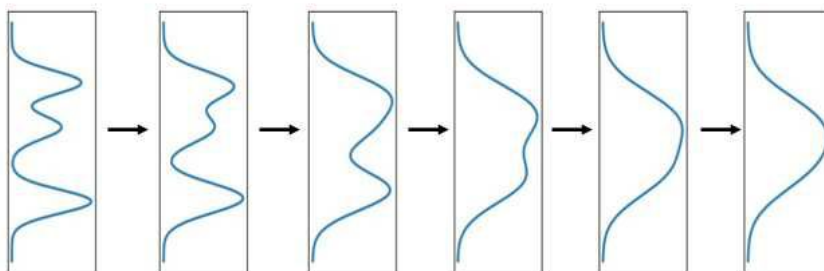
도면9



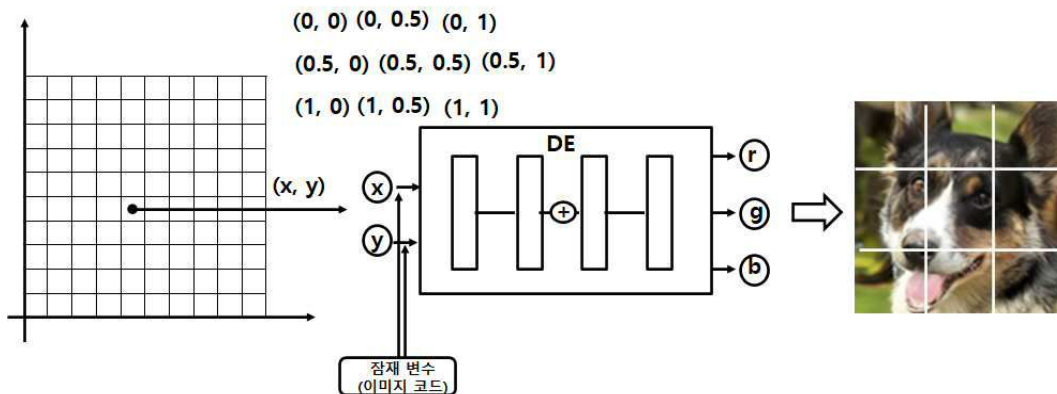
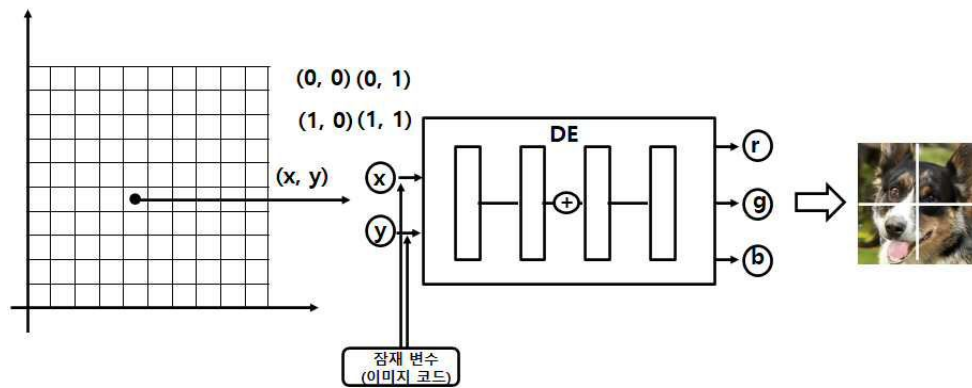
도면10



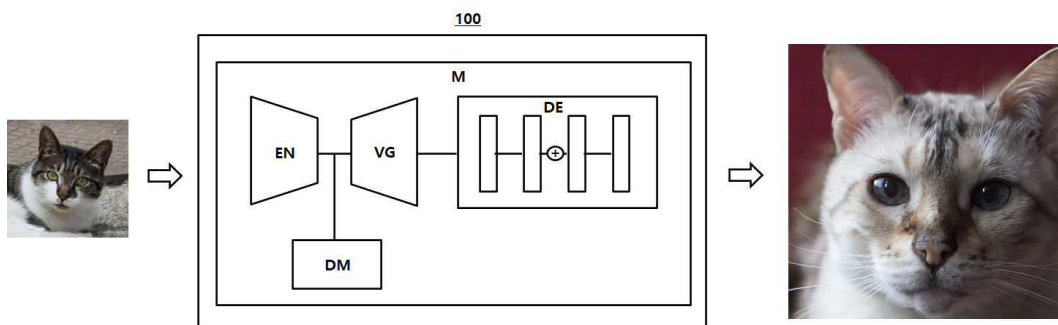
도면11



도면12



도면13



【심사관 직권보정사항】

【직권보정 1】

【보정항목】 청구범위

【보정세부항목】 청구항 10

【변경전】

컴퓨팅 장치와 결합하여,

(AA) 학습 데이터를 입력하여 이미지 생성 모델을 학습시키는 제1 단계; 및

(BB) 입력 데이터를 상기 학습시킨 이미지 생성 모델에 입력하여 상기 이미지 생성 모델이 생성한 임의의 해상도와 스케일의 이미지를 출력하는 제2 단계;

를 포함하는 컴퓨터로 판독 가능한 매체에 저장된 컴퓨터 프로그램에 있어서,

상기 제1 단계는,

상기 이미지 생성 모델이 포함하는 컨볼루션 신경망(CNN, Convolutional Neural Network) 기반의 인코더부, 컨볼루션 신경망 기반의 잠재 변수 생성부 및 암시적 신경망(INN, Implicit Neural Network) 기반의 디코더부를 학습시키는 제1-1 단계; 및

상기 이미지 생성 모델이 포함하는 확산 모델부를 학습시키는 제1-2 단계;

를 포함하는 컴퓨터로 판독 가능한 매체에 저장된 컴퓨터 프로그램 컴퓨터로 판독 가능한 매체에 저장된 컴퓨터 프로그램에 있어서,

상기 제1-1 단계는,

상기 학습 데이터를 상기 인코더부에 입력해 잠재 변수 공간(Image Latent space)으로 매핑시켜 잠재 변수로 변환하는 제1-1-1 단계;

상기 변환한 잠재 변수를 상기 잠재 변수 생성부에 입력해 주파수 정보를 가지고 있는 복수 개의 피처맵 (Feature Map)으로 분리하는 제1-1-2 단계;

이미지 생성을 위해 필요한 모든 좌표의 좌표값을 상기 분리한 복수 개의 피처맵에 쿼리하고, 상기 쿼리한 좌표에 대응되는 피처를 상기 디코더부에 입력하는 제1-1-3 단계; 및

상기 디코더부의 출력값과 학습 데이터 상의 상기 좌표의 RGB값의 차이가 최소화되도록 상기 인코더부, 잠재 변수 생성부 및 디코더부를 학습시키는 제1-1-4 단계;

를 포함하는 컴퓨터로 판독 가능한 매체에 저장된 컴퓨터 프로그램 컴퓨터로 판독 가능한 매체에 저장된 컴퓨터 프로그램.

【변경후】

컴퓨팅 장치와 결합하여,

(AA) 학습 데이터를 입력하여 이미지 생성 모델을 학습시키는 제1 단계; 및

(BB) 입력 데이터를 상기 학습시킨 이미지 생성 모델에 입력하여 상기 이미지 생성 모델이 생성한 임의의 해상도와 스케일의 이미지를 출력하는 제2 단계;

를 포함하는 컴퓨터로 판독 가능한 매체에 저장된 컴퓨터 프로그램에 있어서,

상기 제1 단계는,

상기 이미지 생성 모델이 포함하는 컨볼루션 신경망(CNN, Convolutional Neural Network) 기반의 인코더부, 컨볼루션 신경망 기반의 잠재 변수 생성부 및 암시적 신경망(INN, Implicit Neural Network) 기반의 디코더부를 학습시키는 제1-1 단계; 및

상기 이미지 생성 모델이 포함하는 확산 모델부를 학습시키는 제1-2 단계;

를 포함하는 컴퓨터로 판독 가능한 매체에 저장된 컴퓨터 프로그램에 있어서,

상기 제1-1 단계는,

상기 학습 데이터를 상기 인코더부에 입력해 잠재 변수 공간(Image Latent space)으로 매핑시켜 잠재 변수로 변환하는 제1-1-1 단계;

상기 변환한 잠재 변수를 상기 잠재 변수 생성부에 입력해 주파수 정보를 가지고 있는 복수 개의 피처맵 (Feature Map)으로 분리하는 제1-1-2 단계;

이미지 생성을 위해 필요한 모든 좌표의 좌표값을 상기 분리한 복수 개의 피처맵에 쿼리하고, 상기 쿼리한 좌표에 대응되는 피처를 상기 디코더부에 입력하는 제1-1-3 단계; 및

상기 디코더부의 출력값과 학습 데이터 상의 상기 좌표의 RGB값의 차이가 최소화되도록 상기 인코더부, 잠재 변수 생성부 및 디코더부를 학습시키는 제1-1-4 단계;

를 포함하는 컴퓨터로 판독 가능한 매체에 저장된 컴퓨터 프로그램.