(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0165772
(43) 공개일자 2024년11월25일

(51) 국제특허분류(Int. Cl.)
G06T 11/00 (2006.01) *G06N 3/0455* (2023.01)
G06T 3/00 (2024.01) *G06T 9/00* (2019.01)
G06V 10/74 (2022.01) *G06V 10/764* (2022.01)
G06V 10/774 (2022.01)
(52) CPC특허분류
G06T 11/00 (2013.01)
G06N 3/0455 (2023.01)
(21) 출원번호 10-2023-0063321
(22) 출원일자 2023년05월16일
심사청구일자 2023년05월16일

(71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
김현우
서울시 성북구 북악산로 844 돈암브라운스톤 아파트 103동 804호
김치윤
경기도 용인시 처인구 명지로16번길 26 (역북동, 용인명지대역동원로얄듀크) 107-1301
(74) 대리인
특허법인위솔

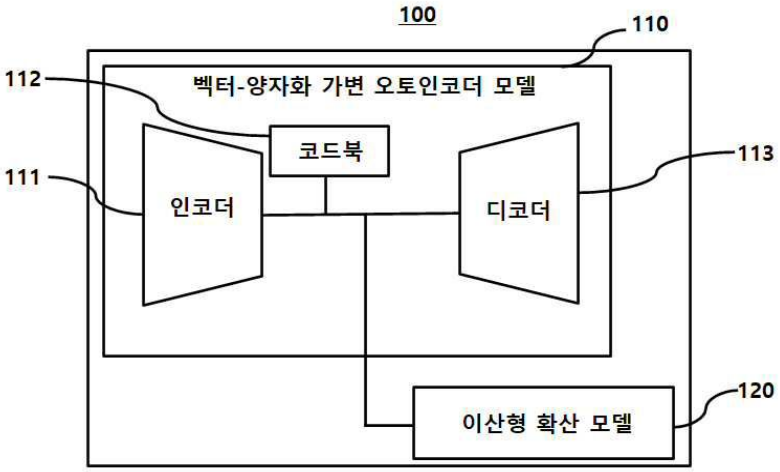
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 **토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 방법 및 이를 위한 장치**

(57) 요 약

본 발명의 일 실시 예에 따른 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 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.

대 표 도 - 도2



(52) CPC특허분류
G06T 3/10 (2024.01)
G06T 9/00 (2019.01)
G06V 10/761 (2023.08)
G06V 10/764 (2023.08)
G06V 10/774 (2023.08)
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

명 세 서

청구범위

청구항 1

프로세서 및 메모리를 포함하는 장치가 고품질 이미지를 생성하는 방법에 있어서,

(a) 입력 이미지를 제1 이미지 토큰 벡터로 변환하는 제1 단계;

(b) 상기 변환한 제1 이미지 토큰 벡터를 코드북이 포함하는 K(K는 양의 정수)개의 코드 벡터 중 어느 하나로 치환한 제2 이미지 토큰 벡터를 생성하는 제2 단계;

(c) 상기 생성한 제2 이미지 토큰 벡터에 전이 행렬을 적용한 정방향 확산을 통해 상기 제2 이미지 토큰 벡터를 흡수(Mask) 상태로 변환한 제3 이미지 토큰 벡터를 생성하는 제3 단계;

(d) 상기 생성한 제3 이미지 토큰 벡터로부터 역방향 확산을 통해 흡수 상태를 포함하지 않는 제4 이미지 토큰 벡터를 생성하는 제4 단계;

(e) 상기 생성한 제4 이미지 토큰 벡터를 고품질 이미지로 변환한 최종 이미지를 출력하는 제5 단계;

를 포함하는 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 방법.

청구항 2

제1항에 있어서,

상기 입력 이미지는,

H(H는 양의 정수) x W(W는 양의 정수) 크기의 픽셀 단위의 이미지이며,

상기 제1 이미지 토큰 벡터는,

H/16 x W/16 크기의 벡터 단위의 벡터인,

토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 방법.

청구항 3

제1항에 있어서,

상기 제2 단계는,

상기 변환한 제1 이미지 토큰 벡터가 포함하는 각각의 벡터와 상기 코드북이 포함하는 K개의 코드 벡터 각각에 대한 유클리디안 거리(Euclidean distance)를 산정하는 제2-1 단계; 및

상기 제1 이미지 토큰 벡터가 포함하는 각각의 벡터에 대하여 상기 산정한 유클리디안 거리가 가장 작은 K개의 코드 벡터 중 어느 하나와 치환하여 제2 이미지 토큰 벡터를 생성하는 제2-2 단계;

를 포함하는 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 방법.

청구항 4

제1항에 있어서,

상기 제3 단계는,

상기 생성한 제2 이미지 토큰 벡터가 포함하는 각각의 벡터에 대하여 전이 행렬을 적용해 상태를 변환하는 제3-1 단계;

상기 각각의 벡터에 대하여 변환된 상태가 모두 흡수 상태인지 판단하는 제3-2 단계; 및

상기 제3-2 단계의 판단 결과 YES라면, 상기 각각의 벡터의 상태가 모두 흡수 상태인 제3 이미지 토큰 벡터를

생성하는 제3-3 단계;

를 포함하는 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 방법.

청구항 5

제4항에 있어서,

상기 제3-1 단계에서 상기 각각의 벡터를 변환한 상태는,

흡수 상태 및 상기 코드북이 포함하는 K개의 코드 벡터 중 어느 하나로 변환한 상태 중 어느 하나이며,

상기 두 가지 상태 중 어느 하나의 상태로 변환될 확률은 동일하거나 상이한,

토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 방법.

청구항 6

제5항에 있어서,

상기 코드북이 포함하는 K개의 코드 벡터 중 어느 하나로 변환될 확률은,

상기 K개의 코드 벡터 모두에 대하여 동일하지 않고, 상기 각각의 벡터와 상기 K개의 코드 벡터 각각에 대하여 산정한 유사도에 기초하여 차등적인 확률을 나타내는,

토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 방법.

청구항 7

제6항에 있어서,

상기 유사도는,

코사인 유사도(Cosine Similarity)인,

토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 방법.

청구항 8

제1항에 있어서,

상기 입력 이미지는,

학습 프로세스 상에서는 하나 이상의 마스킹된 픽셀을 포함하지 않은 이미지이며,

샘플링 프로세스 상에서는 하나 이상의 마스킹된 픽셀을 포함하지 않은 이미지 및 하나 이상의 마스킹된 픽셀을 포함하는 이미지 중 어느 하나인,

토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 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) 입력 이미지를 제1 이미지 토큰 벡터로 변환하는 제1 오퍼레이션;

(B) 상기 변환한 제1 이미지 토큰 벡터를 코드북이 포함하는 K(K는 양의 정수)개의 코드 벡터 중 어느 하나로 치환한 제2 이미지 토큰 벡터를 생성하는 제2 오퍼레이션;

(C) 상기 생성한 제2 이미지 토큰 벡터에 전이 행렬을 적용한 정방향 확산을 통해 상기 제2 이미지 토큰 벡터를

흡수(Mask) 상태로 변환한 제3 이미지 토큰 벡터를 생성하는 제3 오퍼레이션;

(D) 상기 생성한 제3 이미지 토큰 벡터로부터 역방향 확산을 통해 흡수 상태를 포함하지 않는 제4 이미지 토큰 벡터를 생성하는 제4 오퍼레이션;

(E) 상기 생성한 제4 이미지 토큰 벡터를 고품질 이미지로 변환한 최종 이미지를 출력하는 제5 오퍼레이션;

을 실행하는 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 장치.

청구항 10

컴퓨팅 장치와 결합하여,

(AA) 입력 이미지를 제1 이미지 토큰 벡터로 변환하는 제1 단계;

(BB) 상기 변환한 제1 이미지 토큰 벡터를 코드북이 포함하는 K(K는 양의 정수)개의 코드 벡터 중 어느 하나로 치환한 제2 이미지 토큰 벡터를 생성하는 제2 단계;

(CC) 상기 생성한 제2 이미지 토큰 벡터에 전이 행렬을 적용한 정방향 확산을 통해 상기 제2 이미지 토큰 벡터를 흡수(Mask) 상태로 변환한 제3 이미지 토큰 벡터를 생성하는 제3 단계;

(DD) 상기 생성한 제3 이미지 토큰 벡터로부터 역방향 확산을 통해 흡수 상태를 포함하지 않는 제4 이미지 토큰 벡터를 생성하는 제4 단계;

(EE) 상기 생성한 제4 이미지 토큰 벡터를 고품질 이미지로 변환한 최종 이미지를 출력하는 제5 단계;

를 실행하는 컴퓨터로 판독 가능한 매체에 저장된 컴퓨터 프로그램.

발명의 설명

기 술 분 야

[0001] 본 발명은 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 방법 및 이를 위한 장치에 관한 것이다. 보다 자세하게는 확산 과정(Diffusion Process)에서 이미지 토큰 간의 유사도를 고려하는 이산형 확산 모델(Discrete Diffusion Model)을 통해 사실적인 고품질 이미지를 생성하는 방법 및 이를 위한 장치에 관한 것이다.

배 경 기 술

[0002] 최근 대량의 학습 데이터를 기반으로 한 이미지 생성 인공지능 모델들의 성능이 일취월장하고 있는바, 이들이 생성한 이미지가 미술전에서 1등을 차지하는 경우까지 발생하고 있다.

[0003] 이와 같은 이미지 생성 인공지능 모델은 사용자가 모델을 통해 생성하고자 하는 영역을 마스킹 처리한 이미지를 입력하면, 학습을 완료한 이미지 생성 인공지능 모델이 학습 결과에 기반하여 이에 걸맞는 이미지를 자동으로 생성하는바, 생성하는 이미지의 다양성을 확보하고 품질을 향상시키기 위해 다양한 연구 개발이 활발하게 진행되고 있다.

[0004] 한편, 최근에는 이미지 생성 인공지능 모델의 학습 과정에서 다양한 장점을 갖는 확산 모델(Diffusion Model)을 적용하는 경우가 일부 출현하였으나, 학습의 안정성을 위해 제한적인 학습 과정을 정의하여 사용했기 때문에 생성하는 이미지의 다양성과 품질의 측면에서 한계가 있다는 문제점이 있다.

[0005] 본 발명은 이미지 생성 인공지능 모델의 일부 구성으로 확산 모델을 적용하되, 학습의 안정성 확보를 위한 종래의 문제점을 극복하여 고품질 이미지를 다양하게 생성할 수 있는 방법 및 이를 위한 장치에 관한 것이다.

선행기술문헌

특허문헌

[0006] (특허문헌 0001) 대한민국 공개특허공보 제 10-2023-0057434호(2023.04.28)

발명의 내용

해결하려는 과제

[0007] 본 발명이 해결하고자 하는 기술적 과제는 이미지 생성 모델의 일부 구성으로 확산 모델을 적용하되, 학습의 안정성 확보를 위한 종래의 문제점을 극복하여 고품질 이미지를 다양하게 생성할 수 있는 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 방법 및 이를 위한 장치를 제공하는 것이다.

[0008] 본 발명이 해결하고자 하는 또 다른 기술적 과제는 확산 모델의 학습 과정에서 불필요한 노이즈를 줄임으로써 효과적인 학습을 도모할 수 있는 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 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 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 방법은 (a) 입력 이미지를 제1 이미지 토큰 벡터로 변환하는 제1 단계, (b) 상기 변환한 제1 이미지 토큰 벡터를 코드북이 포함하는 K(K는 양의 정수)개의 코드 벡터 중 어느 하나로 치환한 제2 이미지 토큰 벡터를 생성하는 제2 단계, (c) 상기 생성한 제2 이미지 토큰 벡터에 전이 행렬을 적용한 정방향 확산을 통해 상기 제2 이미지 토큰 벡터를 흡수(Mask) 상태로 변환한 제3 이미지 토큰 벡터를 생성하는 제3 단계, (d) 상기 생성한 제3 이미지 토큰 벡터로부터 역방향 확산을 통해 흡수 상태를 포함하지 않는 제4 이미지 토큰 벡터를 생성하는 제4 단계 및 (e) 상기 생성한 제4 이미지 토큰 벡터를 고품질 이미지로 변환한 최종 이미지를 출력하는 제5 단계를 포함한다.

[0011] 일 실시 예에 따르면, 상기 입력 이미지는, H(H는 양의 정수) x W(W는 양의 정수) 크기의 픽셀 단위의 이미지이며, 상기 제1 이미지 토큰 벡터는, H/16 x W/16 크기의 벡터 단위의 벡터일 수 있다.

[0012] 일 실시 예에 따르면, 상기 제2 단계는, 상기 변환한 제1 이미지 토큰 벡터가 포함하는 각각의 벡터와 상기 코드북이 포함하는 K개의 코드 벡터 각각에 대한 유클리디안 거리(Euclidean distance)를 산정하는 제2-1 단계 및 상기 제1 이미지 토큰 벡터가 포함하는 각각의 벡터에 대하여 상기 산정한 유클리디안 거리가 가장 작은 K개의 코드 벡터 중 어느 하나와 치환하여 제2 이미지 토큰 벡터를 생성하는 제2-2 단계를 포함할 수 있다.

[0013] 일 실시 예에 따르면, 상기 제3 단계는, 상기 생성한 제2 이미지 토큰 벡터가 포함하는 각각의 벡터에 대하여 전이 행렬을 적용해 상태를 변환하는 제3-1 단계, 상기 각각의 벡터에 대하여 변환된 상태가 흡수 상태인지 판단하는 제3-2 단계 및 상기 제3-2 단계의 판단 결과 YES라면, 상기 각각의 벡터의 상태가 모두 흡수 상태인 제3 이미지 토큰 벡터를 생성하는 제3-3 단계를 포함할 수 있다.

[0014] 일 실시 예에 따르면, 상기 제3-1 단계에서 상기 각각의 벡터를 변환한 상태는, 흡수 상태 및 상기 코드북이 포함하는 K개의 코드 벡터 중 어느 하나로 변환한 상태 중 어느 하나이며, 상기 두 가지 상태 중 어느 하나의 상태로 변환될 확률은 동일하거나 상이할 수 있다.

[0015] 일 실시 예에 따르면, 상기 코드북이 포함하는 K개의 코드 벡터 중 어느 하나로 변환될 확률은, 상기 K개의 코드 벡터 모두에 대하여 동일하지 않고, 상기 각각의 벡터와 상기 K개의 코드 벡터 각각에 대하여 산정한 유사도에 기초하여 차등적인 확률을 나타낼 수 있다.

[0016] 일 실시 예에 따르면, 상기 유사도는, 코사인 유사도(Cosine Similarity)일 수 있다.

[0017] 일 실시 예에 따르면, 상기 입력 이미지는, 학습 프로세스 상에서는 하나 이상의 마스킹된 픽셀을 포함하지 않은 이미지이며, 샘플링 프로세스 상에서는 하나 이상의 마스킹된 픽셀을 포함하지 않은 이미지 및 하나 이상의 마스킹된 픽셀을 포함하는 이미지 중 어느 하나일 수 있다.

[0018] 상기 기술적 과제를 달성하기 위한 본 발명의 또 다른 실시 예에 따른 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 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) 입력 이미지를 제1 이미지 토큰 벡터로 변환하는 제1 오퍼레이션, (B) 상기 변환한 제1 이미지 토큰 벡터를 코드북이 포함하는 K(K

는 양의 정수)개의 코드 벡터 중 어느 하나로 치환한 제2 이미지 토큰 벡터를 생성하는 제2 오퍼레이션, (C) 상기 생성한 제2 이미지 토큰 벡터에 전이 행렬을 적용한 정방향 확산을 통해 상기 제2 이미지 토큰 벡터를 흡수(Mask) 상태로 변환한 제3 이미지 토큰 벡터를 생성하는 제3 오퍼레이션, (D) 상기 생성한 제3 이미지 토큰 벡터로부터 역방향 확산을 통해 흡수 상태를 포함하지 않는 제4 이미지 토큰 벡터를 생성하는 제4 오퍼레이션 및 (E) 상기 생성한 제4 이미지 토큰 벡터를 고품질 이미지로 변환한 최종 이미지를 출력하는 제5 오퍼레이션을 실행한다.

[0019] 상기 기술적 과제를 달성하기 위한 본 발명의 또 다른 실시 예에 따른 매체에 저장된 컴퓨터 프로그램은 컴퓨팅 장치와 결합하여, (AA) 입력 이미지를 제1 이미지 토큰 벡터로 변환하는 제1 단계, (BB) 상기 변환한 제1 이미지 토큰 벡터를 코드북이 포함하는 K(K는 양의 정수)개의 코드 벡터 중 어느 하나로 치환한 제2 이미지 토큰 벡터를 생성하는 제2 단계, (CC) 상기 생성한 제2 이미지 토큰 벡터에 전이 행렬을 적용한 정방향 확산을 통해 상기 제2 이미지 토큰 벡터를 흡수(Mask) 상태로 변환한 제3 이미지 토큰 벡터를 생성하는 제3 단계, (DD) 상기 생성한 제3 이미지 토큰 벡터로부터 역방향 확산을 통해 흡수 상태를 포함하지 않는 제4 이미지 토큰 벡터를 생성하는 제4 단계 및 (EE) 상기 생성한 제4 이미지 토큰 벡터를 고품질 이미지로 변환한 최종 이미지를 출력하는 제5 단계를 실행한다.

발명의 효과

[0020] 상기와 같은 본 발명에 따르면, 정방향 확산 프로세스인 전이 행렬을 각각의 벡터가 코드북이 포함하는 K개의 코드 벡터 중 어느 하나로 변환될 확률을 유사도를 개별적으로 산정하여 차등적인 확률로 정의하는바, 학습의 안정성 확보를 위한 종래의 문제점을 극복하고 불필요한 노이즈를 주지 않게 되므로 확산 모델의 효과적인 학습이 학습이 이루어지게 함과 동시에 다양한 고품질 이미지 생성 성능까지 향상될 수 있다는 효과가 있다.

[0021] 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

[0022] 도 1은 본 발명의 제1 실시 예에 따른 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 장치가 포함하는 전체 구성을 예시적으로 도시한 도면이다.

도 2는 본 발명의 제1 실시 예에 따른 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 장치가 포함하는 전체 구성을 인공지능 모델의 측면에서 도시한 도면이다.

도 3은 본 발명의 제2 실시 예에 따른 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 방법의 대표적인 단계를 나타낸 순서도이다.

도 4는 입력 이미지가 제1 이미지 토큰 벡터로 변환되는 모습을 예시적으로 도시한 도면이다.

도 5는 제1 이미지 토큰 벡터가 코드북이 포함하는 K개의 코드 벡터 중 어느 하나로 치환되는 모습을 예시적으로 도시한 도면이다.

도 6은 본 발명의 제2 실시 예에 따른 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 방법에 있어서, 제2 이미지 토큰 벡터를 생성하는 제2 단계를 구체화한 순서도이다.

도 7은 본 발명의 제2 실시 예에 따른 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 방법에 있어서, 제3 이미지 토큰 벡터를 생성하는 제3 단계를 구체화한 순서도이다.

도 8은 제2 이미지 토큰 벡터의 단일 시간 단계에서 그 다음 단일 시간 단계로 상태가 변경되는 정방향 확산 과정의 일부를 예시적으로 도시한 도면이다.

도 9는 종래의 모델에서 학습의 안정성을 위해 제한적인 학습 과정을 정의하기 위한 코드 벡터 변환 확률 분포도이다.

도 10은 본 발명의 제2 실시 예에 따른 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 방법에 있어서, 각 벡터들의 유사도를 반영한 코드 벡터 변환 확률 분포도이다.

도 11은 도 8에 도시했던 정방향 확산 과정을 T번의 연속된 마르코프 과정를 통해 제2 이미지 토큰 벡터가 포함하는 각각의 벡터 모두가 흡수 상태가 되는 모습과 역방향 확산 과정을 예시적으로 도시한 도면이다.

도 12는 정방향 확산 과정 및 역방향확산 과정에 대하여 종래의 모델과 본 발명을 비교하여 도시한 도면이다.

도 13 및 도 14는 본 발명의 제2 실시 예에 따른 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 방법에 따라 장치가 출력한 최종 이미지이다.

발명을 실시하기 위한 구체적인 내용

[0023] 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.

[0024] 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.

[0025] 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.

[0026] 또한, 어떤 구성요소들을 포함한다는 표현은 "개방형"의 표현으로서 해당 구성요소들이 존재하는 것을 단순히 지칭할 뿐이며, 추가적인 구성요소들을 배제하는 것으로 이해되어서는 안 된다.

[0027] 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.

[0028] 이하에서는 도면들을 참조하여 본 발명의 세부적인 실시 예들에 대해 살펴보도록 한다.

[0029] 도 1은 본 발명의 제1 실시 예에 따른 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 장치(100)가 포함하는 전체 구성을 예시적으로 도시한 도면이다.

[0030] 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.

[0031] 본 발명의 제1 실시 예에 따른 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 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.

[0032] 프로세서(10)는 각 구성의 전반적인 동작을 제어한다. 프로세서(10)는 CPU(Central Processing Unit), MPU(Micro Processer Unit), MCU(Micro Controller Unit) 또는 본 발명이 속하는 기술 분야에서 널리 알려져 있는 형태의 인공지능 프로세서 중 어느 하나일 수 있다.

[0033] 아울러, 프로세서(10)는 본 발명의 제2 실시 예에 따른 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 방법을 수행하기 위한 적어도 하나의 애플리케이션 또는 프로그램에 대한 연산을 수행할 수 있는바, 이를 위해 프로세서(10)에는 소정의 구조를 나타내는 벡터-양자화 가변 오토인코더 모델(Vector Quantization Variational AutoEncoder Model, VQ-VAE Model) 및 확산 모델 기반의 벡터-양자화 확산 모델(Vector Quantization Diffusion Model, VQ Diffusion Model)을 이미지 생성 인공지능 모델로 포함할 수 있으나, 이에 대해서는 후술하도록 한다.

[0034] 네트워크 인터페이스(20)는 본 발명의 제1 실시 예에 따른 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 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.

[0035] 메모리(30)는 각종 정보, 명령 및/또는 정보를 저장하며, 본 발명의 제2 실시 예에 따른 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 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.

[0036] 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 매치(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.

[0037] 컴퓨터 프로그램(41)은 메모리(30)에 로드되어, 하나 이상의 프로세서(10)에 의해, (A) 입력 이미지를 제1 이미지 토큰 벡터로 변환하는 제1 오퍼레이션, (B) 상기 변환한 제1 이미지 토큰 벡터를 코드북이 포함하는 K(K는 양의 정수)개의 코드 벡터 중 어느 하나로 치환한 제2 이미지 토큰 벡터를 생성하는 제2 오퍼레이션, (C) 상기 생성한 제2 이미지 토큰 벡터에 전이 행렬을 적용한 정방향 확산을 통해 상기 제2 이미지 토큰 벡터를 흡수(Mask) 상태로 변환한 제3 이미지 토큰 벡터를 생성하는 제3 오퍼레이션, (D) 상기 생성한 제3 이미지 토큰 벡터로부터 역방향 확산을 통해 흡수 상태를 포함하지 않는 제4 이미지 토큰 벡터를 생성하는 제4 오퍼레이션 및 (E) 상기 생성한 제4 이미지 토큰 벡터를 고품질 이미지로 변환한 최종 이미지를 출력하는 제5 오퍼레이션을 실행할 수 있다.

[0038] 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 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 방법에 대한 설명에서 후술하도록 한다.

[0039] 데이터 버스(50)는 이상 설명한 프로세서(10), 네트워크 인터페이스(20), 메모리(30) 및 스토리지(40) 사이의 명령 및/또는 정보의 이동 경로가 된다.

[0040] 이상 간단하게 설명한 본 발명의 제1 실시 예에 따른 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 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 무방한 바, 프로세서(10)에 해당하는 CPU 등이 설치되고 네트워크 기능만 보유하고 있는 전자 기기라면 어떠한 것이라도 무방하다 할 것이다.

[0041] 도 2는 본 발명의 제1 실시 예에 따른 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 장치(100)를 도 1의 경우와 상이한 구성, 보다 구체적으로 장치(100)를 이루고 있는 인공지능 모델의 측면에서 도시한 도면인바, 도 2를 참조하면 앞서 프로세서(10)에 대한 설명에서 간단히 언급한 바와 같이 장치(100)가 인코더(111), 코드북(112) 및 디코더(113)를 포함하는 벡터-양자화 가변 오토인코더 모델(110)과 이산형 확산 모델(120)을 포함하는 것을 확인할 수 있으며, 이들 모델을 함께 일컬어 벡터-양자화 확산 모델이라 명명한 것이다.

[0042] 이하, 본 발명의 제1 실시 예에 따른 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 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 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 방법을 제공하는 방법에 대하여 도 3 내지 도 14를 참조하여 설명하도록 한다.

[0043] 도 3은 본 발명의 제2 실시 예에 따른 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 방법의 대표적인 단계를 나타낸 순서도이다.

[0044] 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.

[0045] 각 단계는 본 발명의 제1 실시 예에 따른 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 장치(100)를 통해 이루어지는 것을 전제로 하며, 본 발명의 제1 실시 예에 따른 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 장치(100)가 "서버"의 형태임을 전제로 하였기 때문에 사용자 단말(미도시)에 설치된 전용 어플리케이션을 본 발명의 제1 실시 예에 따른 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 장치(100)와 동일한 의미로 볼 것이고, 설명의 편의를 위해 이들 모두를 "장치(100)"로 명명하도록 한다.

[0046] 한편, 장치(100)가 고품질 이미지를 생성함에 있어서 학습을 진행하는 학습 프로세스와 학습 프로세스가 종료된 이후, 실제 장치(100)를 통해 고품질 이미지를 생성하는 샘플링 프로세스가 별개로 진행될 수 있는바, 학습 프로세스 상에서 장치(100)에 의한 동작부터 설명하도록 한다.

[0047] 우선, 장치(100)가 입력 이미지를 제1 이미지 토큰 벡터로 변환하며(S310), 이를 제1 단계라 한다.

[0048] 여기서 입력 이미지는 학습 프로세스 진행을 위한 학습 데이터로 볼 수 있으며, 마스킹된 픽셀을 어느 하나라도 포함하지 않는 온전한 이미지일 수 있는바, H(H는 양의 정수) x W(W는 양의 정수) 크기의 픽셀 단위의 이미지일 수 있고, H와 W의 크기에 대해서는 특별한 제한은 없다 할 것이다.

[0049] 한편, 픽셀 단위의 입력 이미지가 변환된 제1 이미지 토큰 벡터는 H/16 x W/16 크기의 벡터 단위의 벡터일 수 있는바, 예를 들어 도 4에 도시된 바와 같이 입력 이미지가 256 x 256 크기의 픽셀 이미지라면, 해당 이미지가 변환된 제1 이미지 토큰 벡터는 16 x 16 크기의 벡터 단위의 벡터일 수 있으며, 픽셀 이미지를 그대로 사용하는 것이 아니라 이미지 토큰 벡터로 변환하는 경우 픽셀 이미지를 데이터로 사용할 때보다 적은 연산이 수행되면 충분하기 때문에 장치(100)의 효율성을 도모할 수 있다.

[0050] 이와 같은 제1 단계는 입력 이미지를 벡터 차원의 제1 이미지 토큰 벡터로 변환하는 단계이므로 인코더(111)가 수행하는 단계로 볼 수 있으며, 그에 따라 도 2에 도시된 장치(100)가 포함하는 벡터-양자화 확산 모델 중, 벡터-양자화 가변 오토인코더 모델(110)에서 이루어지는 단계라 할 것이다.

[0051] 이후, 장치(100)가 변환한 제1 이미지 토큰 벡터를 코드북(112)이 포함하는 K(K는 양의 정수)개의 코드 벡터 중 어느 하나로 치환한 제2 이미지 토큰 벡터를 생성하며(S320), 이를 제2 단계라 한다.

[0052] 코드북은 K개의 코드 벡터인 코드워드의 집합을 포함하며, 벡터-양자화 확산 모델에서 데이터의 압축 및 재구성에 매우 중요한 역할을 하는바, 제1 이미지 토큰 벡터를 코드북이 포함하는 K개의 코드 벡터 중 어느 하나로 치환하는 프로세스를 벡터-양자화 프로세스라 하며, 제1 단계와 마찬가지로 장치(100)가 포함하는 벡터-양자화 확산 모델 중, 벡터-양자화 가변 오토인코더 모델(110)에서 이루어지는 단계이다.

[0053] 도 5를 참조하면, 코드북이 K개의 코드 벡터인 Z_1 내지 Z_K 를 포함하고, 제1 이미지 토큰 벡터가 4 x 4 크기로 예를 들어 도시되어 있으며, 벡터-양자화 프로세스를 통해 제1 이미지 토큰 벡터가 포함하는 각각의 벡터 16개 모두가 코드북이 포함하는 K개의 코드 벡터 중 어느 하나로 치환된 것을 확인할 수 있다. 여기서 치환된 토큰 벡터가 제2 이미지 토큰 벡터이며, 제2 이미지 토큰 벡터가 포함하는 각각의 벡터 16개에 기재된 숫자는 코드북이 포함하는 K개의 코드 벡터의 숫자를 간략하게 기재한 것이다.

[0054] 이 경우, 제1 이미지 토큰 벡터와 코드북이 포함하는 K개의 코드 벡터 중 어느 하나를 어떠한 기준을 가지고 치환하는지가 문제될 수 있는바, 이하 도 6을 참조하여 설명하도록 한다.

[0055] 도 6은 본 발명의 제2 실시 예에 따른 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 방법에 있어서, 제2 이미지 토큰 벡터를 생성하는 제2 단계를 구체화한 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] 우선, 장치(100)가 제1 이미지 토큰 벡터가 포함하는 각각의 벡터와 코드북이 포함하는 K개의 코드 벡터 각각에 대한 유클리디안 거리(Euclidean distance)를 산정하며(S320-1), 이를 제2-1 단계라 한다.

[0058] 여기서 유클리디안 거리는 피타고라스 정리에 의해 정의되는 공간인 유클리도 공간에서 두 점 사이의 거리이며, 두 개의 벡터 사이의 유사성을 정량화하여 산정할 수 있는 메트릭으로 사용되고 있는바, 유클리디안 거리 값이 두 벡터 간의 차이를 나타내기에 두 벡터 사이의 유클리디안 거리가 작으면 두 벡터가 유사하며, 반대로 두 벡터 사이의 유클리디안 거리가 크면 두 벡터가 유사하지 않는 것으로 볼 수 있다.

[0059] 도 5의 예시에 따라 제1 이미지 토큰 벡터가 4 x 4 크기이며, 16개의 벡터를 포함한다면, 16개의 벡터 각각에 대하여 코드북이 포함하는 K개의 코드 벡터 전부와의 유클리디안 거리를 산정하게 되는 것인바, 최종적으로 16K개의 유클리디안 거리가 산정될 수 있다.

[0060] 이후, 장치(100)가 제1 이미지 토큰 벡터가 포함하는 각각의 벡터에 대하여 앞서 산정한 유클리디안 거리가 가장 작은 K개의 코드 벡터 중 어느 하나와 치환하여 제2 이미지 토큰 벡터를 생성하며(S320-2), 이를 제2-2 단계라 한다.

[0061] 도 5의 예시에 따라 쉽게 설명하면, 16개의 벡터를 1번 벡터, 2번 벡터 쯤 16번 벡터와 같은 식으로 명명했을 때, 1번 벡터에 대하여 산정한 유클리디안 거리는 코드북이 포함하는 코드 벡터의 개수인 K개이며, K개의 유클리디안 거리 중 가장 작은 유클리디안 거리를 나타내는 코드 벡터를 1번 벡터와 치환하는 것이고, 이를 16번 벡터에 대해서까지 진행하여 제1 이미지 토큰 벡터가 포함하는 모든 벡터를 코드북에 포함하는 K개의 코드 벡터 중 어느 하나와 치환하여 제2 이미지 토큰 벡터가 생성되는 것이다.

[0062] 이상 설명한 제1 단계 및 제2 단계는 장치(100)가 포함하는 벡터-양자화 확산 모델 중, 벡터-양자화 가변 오토인코더 모델(110)에서 이루어지는 단계이며, 그에 따라 학습 프로세스 상에서는 벡터-양자화 가변 오토인코더 모델(110)의 학습, 보다 구체적으로 인코더(111)로 변환한 제1 이미지 토큰 벡터가 코드북에 포함하는 K개의 코드 벡터들과 유사해지도록, 즉 유클리디안 거리가 최소가 되도록 인코더(111)를 학습할 것이며, 이와 반대로 코드북이 포함하는 K개의 코드 벡터에 대하여 유클리디안 거리가 최소가 되어 제1 이미지 토큰 벡터와 유사해지도록 코드북(112)을 학습하게 될 것이다.

[0063] 다시 도 3에 대한 설명으로 돌아가도록 한다.

[0064] 제2 이미지 토큰 벡터를 생성했다면, 장치(100)가 생성한 제2 이미지 토큰 벡터에 전이 행렬을 적용한 정방향 확산을 통해 제2 이미지 토큰 벡터를 흡수(Mask) 상태로 변환한 제3 이미지 토큰 벡터를 생성하며(S330), 이를 제3 단계라 한다.

[0065] 제3 단계부터는 장치(100)가 포함하는 벡터-양자화 확산 모델 중, 벡터-양자화 가변 오토인코더 모델(110)이 아닌 이산형 확산 모델(120)에서 이루어지는 단계이며, 이산형 확산 모델(120)도 확산 모델의 하나이기 때문에 정방향 확산 과정과 역방향 확산 과정을 필수적으로 거치게 되는바, 제3 단계는 정방향 확산 과정에 관한 것이다.

[0066] 도 7은 본 발명의 제2 실시 예에 따른 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 방법에 있어서, 제3 이미지 토큰 벡터를 생성하는 제3 단계를 구체화한 순서도이다.

[0067] 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.

[0068] 우선, 장치(100)가 제2 이미지 토큰 벡터가 포함하는 각각의 벡터에 대하여 전이 행렬을 적용해 상태를 변환하며(S330-1), 이를 제3-1 단계라 한다.

[0069] 여기서 전이 행렬은 그 자체가 정방향 확산 과정을 의미하며, 단일 시간 단계의 한 상태에서 다른 상태로 이동할 확률을 나타내는 정사각형 행렬인바, 행렬의 각 행은 현재 상태를 나타내고 각 열은 가능한 다음 상태를 나타내며, 행렬의 요소는 현재 상태에서 다음 상태로 이동할 확률을 나타낸다.

[0070] 한편, 각각의 벡터에 대하여 전이 행렬을 적용해 변환한 상태는 흡수 상태 및 코드북이 포함하는 K개의 코드 벡터 중 어느 하나로 변환한 상태(제2 이미지 토큰 벡터가 코드북이 포함하는 K개의 코드 벡터 중 어느 하나로 변환된 벡터이므로 K개의 코드 벡터 중 어느 하나로 변환한 상태는 이전 그대로의 상태를 포함할 수 있다) 중 어느 하나일 수 있으며, 이와 같은 두 가지 상태 중 어느 하나의 상태로 변환될 확률은 동일하거나 상이할 수 있는바, 즉 어떠한 벡터가 흡수 상태로 변환될 확률이 1/2이고, K개의 코드 벡터 중 어느 하나로 변환될 확률 역시 1/2일 수 있으며, 이와 별개로 사용자의 설정에 따라 각각의 확률을 설정할 수도 있다.

[0071] 도 8은 제2 이미지 토큰 벡터의 단일 시간 단계에서 그 다음 단일 시간 단계로 상태가 변경되는 정방향 확산 과정의 일부를 예시적으로 도시한 도면인바, 최초의 벡터 2가 그 다음에 2로, 4가 1로, 1이 M으로, 3이 5로 변환된 것을 확인할 수 있는바, 각각 1/2 또는 사용자가 설정한 확률을 가지고 이전 그대로의 상태(코드북이 포함하는 K개의 코드 벡터 중 어느 하나로 변환한 상태), 코드북이 포함하는 K개의 코드 벡터 중 어느 하나로 변환한 상태, 흡수 상태, 코드북이 포함하는 K개의 코드 벡터 중 어느 하나로 변환한 상태가 된 것이다.

[0072] 한편, 어떠한 벡터가 K개의 코드 벡터 중 어느 하나로 변환할 확률은 1/2 또는 사용자가 설정한 확률이지만, K개의 코드 벡터 내에서 어떠한 코드 벡터로 변환될지에 대한 확률은 K개의 코드 벡터 모두에 대하여 동일하지 않고, 각각의 벡터와 K개의 코드 벡터 각각에 대하여 산정한 유사도에 기초하여 차등적인 확률을 나타낼 수 있는바, 여기서 유사도는 코사인 유사도(Cosine Similarity)일 수 있으며, 이것이 종래의 모델과의 가장 큰 차이점이라 할 것이다.

[0073] 종래의 모델의 경우 학습의 안정성을 위해 제한적인 학습 과정을 정의하여 사용했던바, 보다 구체적으로 K개의 코드 벡터 내에서 어떠한 코드 벡터로 변환될지에 대한 확률을 도 9에 도시한 바와 같이 균일한 확률로 두었던바, 이는 기존에 가지고 있던 이미지 토큰의 정보가 전혀 관련성이 없는 토큰으로 무작위하게 변환될 수 있는

가능성을 열어둔 것이기에 확산 모델이 효과적으로 학습되는 것을 방해하고 이미지 생성 성능에 악영향을 주었으며, 이를 해결하기 위해 K개의 코드 벡터 각각에 대해 변환될 확률을 극단적으로 낮추어 다른 코드 벡터로 변환될 가능성 자체를 제거하거나 전이 행렬의 확률을 0으로 정의하는 등 제한적인 구조의 전이 행렬을 사용했던 바, 이미지 생성 성능에 한계를 두게 되는 문제점이 존재하였다.

[0074] 그러나 본 발명의 제2 실시 예에 따른 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 방법에서는 각각의 벡터가 코드북이 포함하는 K개의 코드 벡터 중 어느 하나로 변환될 확률을 유사도를 개별적으로 산정하여 도 10에 도시한 바와 같이 차등적인 확률로 설정한바, 의미론적으로 더 유사한 토큰에 대한 확률이 높게 정의되므로 기존에 가지고 있던 이미지 토큰의 정보가 이와 유사한 토큰으로 변환되게 되고, 전혀 관련성이 없는 토큰으로 무작위하게 변환될 수 있는 가능성을 원천적으로 차단함으로써 불필요한 노이즈를 주지 않게 되는바, 확산 모델의 효과적인 학습이 이루어지게 함과 동시에 고품질 이미지 생성 성능까지 향상될 수 있는 것이다.

[0075] 다시 도 7에 대한 설명으로 돌아가도록 한다.

[0076] 전이 행렬을 적용해 상태를 변환했다면, 장치(100)가 각각의 벡터에 대하여 변환된 상태가 흡수 상태인지 판단하며(S330-2), 이를 제3-2 단계라 하고, 제3-2 단계의 판단 결과 YES라면, 각각의 벡터의 상태가 모두 흡수 상태인 제3 이미지 토큰 벡터를 생성하며(S330-3), 이를 제3-3 단계라 한다.

[0077] 이상과 같은 제3-2 단계 및 제3-3 단계의 목표는 정방향 확산을 통해 마지막 단계에 도달했을 때 어떠한 균일 분포, 보다 구체적으로 흡수 상태가 될 수 있도록 하는 것인바, 그에 따라 장치(100)는 한번 흡수 상태로 변환된 경우 더 이상 다른 상태로 변환되지 않게 할 것이며, 도 11에 앞서 도 8에 도시했던 정방향 확산 과정을 T번의 연속된 마르코프 과정(Markov Process)를 통해 제2 이미지 토큰 벡터가 포함하는 각각의 벡터 모두가 흡수 상태가 되는 모습을 예시적으로 도시해 놓았다.

[0078] 도 12는 정방향 확산 과정에 대하여 종래의 모델과 본 발명을 비교하여 도시한 도면인바, 종래의 모델인 Matrix T_t를 살피면 6개의 코드 벡터 각각에 대해 변환될 확률을 극단적으로 낮추어 다른 코드 벡터로 변환될 가능성 자체를 제거하여 c1이 c1으로 변환될 확률(짙은 색상으로 표시), c2가 c2로 변환될 확률, c3가 c3로 변환될 확률, c4가 c4로 변환될 확률, c5가 c5로 변환될 확률, c6가 c6로 변환될 확률만 정의해 놓은바(흡수 상태로 변환될 확률은 별도), T번의 연속된 마르코프 과정을 거치면서 기존의 코드 벡터가 그대로 유지되는 경우 또는 기존의 코드 벡터가 흡수 상태로 변환되는 경우만 존재하게 된다. 이는 확률 테이블 우측에 도시한 여성 이미지의 각 픽셀이 T번의 연속된 마르코프 과정을 거치면서 이전과 동일하거나 또는 흡수 상태로 변환되는 경우만 존재하는 것을 통해 확인할 수 있다.

[0079] 그러나 본 발명인 Matrix P_t를 살피면 6개의 코드 벡터 각각에 대하여 유사도 산정에 기초한 확률을 정의함으로써 c1이 c1, c3, c6로 변환될 확률, c2가 c2, c4, c5로 변환될 확률, c3가 c1, c3, c6로 변환될 확률, c4가 c2, c4, c5로 변환될 확률, c5가 c2, c4, c5, c6로 변환될 확률, c6가 c1, c3, c5로 변환될 확률을 차등적(색상의 차이)으로 두었기에(흡수 상태로 변환될 확률은 별도) T번의 연속된 마르코프 과정을 거치면서 기존의 코드 벡터가 그대로 유지되는 경우, 기존의 코드 벡터가 유사도가 높은 다른 코드 벡터로 변환되는 경우 및 기존의 코드 벡터가 흡수 상태로 변환되는 경우 중 어느 한 경우가 존재하게 되는 것이다. 이는 확률 테이블 우측에 도시한 여성 이미지의 각 픽셀이 T번의 연속된 마르코프 과정을 거치면서 이전과 동일하거나, 유사도가 높은 것으로 변환되거나 흡수 상태로 변환되는 경우 모두를 포함하는 것을 통해 확인할 수 있다.

[0080] 다시 도 2에 대한 설명으로 돌아가도록 한다.

[0081] 제3 이미지 토큰 벡터까지 생성했다면, 장치(100)가 제3 이미지 토큰 벡터로부터 역방향 확산을 통해 흡수 상태를 포함하지 않는 제4 이미지 토큰 벡터를 생성하며(S340), 이를 제4 단계라 한다.

[0082] 제4 단계는 역방향 확산, 보다 구체적으로 모든 벡터가 흡수 상태인 제3 이미지 토큰 벡터에 대하여 정방향 확산 과정의 역을 수행하게 되는바, 흡수 상태가 되기 이전의 벡터를 찾는 과정을 학습하게 되며, 이는 이미지에 대한 정보가 없는 상태에서 이미지를 만드는 과정이므로 이미지 생성 과정으로 볼 수 있다.

[0083] 앞선 도 11 및 도 12를 다시 참조하면, 정방향 확산 과정의 반대 방향으로 역방향 확산 과정을 확인할 수 있으며, 학습 프로세스 상에서 역방향 확산 과정은 최초의 이미지, 즉 제3 이미지 토큰 벡터로부터 제4 이미지 토큰 벡터를 생성하는 과정을 학습하게 될 것인바, 학습 프로세스 상이기에 여기서의 제4 이미지 토큰 벡터는 제2 이미지 토큰 벡터와 동일한 토큰 벡터일 것이다.

[0084] 다시 도 3에 대한 설명으로 돌아가도록 한다.

[0085] 제4 이미지 토큰 벡터까지 생성했다면, 장치(100)가 최종적으로 제4 이미지 토큰 벡터를 고품질 이미지로 변환한 최종 이미지를 출력하며(S350), 이를 제5 단계라 한다.

[0086] 제5 단계는 이미지 토큰 벡터로부터 실제 이미지를 생성하기 때문에 디코더(113)가 수행하는 단계로 볼 수 있으며, 그에 따라 벡터-양자화 확산 모델 중, 벡터-양자화 가변 오토인코더 모델(110)에서 이루어지는 단계라 할 것이다.

[0087] 한편, 제5 단계가 장치(100)의 출력물인 최종 이미지를 출력하는 단계인바, 학습 프로세스 상에서는 제1 단계에서의 입력 이미지와 제5 단계에서의 최종 이미지의 차이가 최소가 되도록 인코더(111) 및 디코더(113)가 학습될 수 있을 것이다.

[0088] 이상 본 발명의 제2 실시 예에 따른 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 방법, 보다 구체적으로 학습 프로세스 상에서 장치(100)에 의한 동작에 대하여 설명한바, 학습을 완료한 이후 실제 장치(100)를 통해 고품질 이미지를 생성하는 샘플링 프로세스 상에서는 제1 단계의 입력 이미지가 하나 이상의 마스킹된 픽셀을 포함하는 이미지를 더 포함할 수 있다라는 점, 이상의 설명에서 학습에 관한 설명이 적용되지 않는다는 점을 제외하고 동일한 설명이 적용될 수 있는바, 중복 서술을 방지하기 위해 자세한 설명은 생략하도록 한다.

[0089] 지금까지 본 발명의 제2 실시 예에 따른 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 방법에 대하여 설명하였다. 본 발명에 따르면, 정방향 확산 프로세스인 전이 행렬을 각각의 벡터가 코드북이 포함하는 K개의 코드 벡터 중 어느 하나로 변환될 확률을 유사도를 개별적으로 산정하여 차등적인 확률로 정의하는바, 학습의 안정성 확보를 위한 종래의 문제점을 극복하고 불필요한 노이즈를 주지 않게 되므로 확산 모델의 효과적인 학습이 학습이 이루어지게 함과 동시에 다양한 고품질 이미지 생성 성능까지 향상될 수 있다.

[0090] 도 13 및 도 14는 본 발명의 제2 실시 예에 따른 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 방법에 따라 장치(100)가 출력한 최종 이미지인바, 도 13을 참조하면 사람의 얼굴에 대하여 각종 나이대의 사람의 얼굴에 대한 고품질의 다양한 이미지를 생성함을 확인할 수 있으며, 도 14를 참조하면 마스킹된 픽셀을 포함하는 사람 얼굴의 입력 이미지에 대하여 마스킹된 픽셀이 다양한 표정과 나이대의 인물에 대한 얼굴로 자연스럽게 채워짐을 확인할 수 있다.

[0091] 마지막으로, 본 발명의 제1 실시 예에 따른 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 장치(100)와 본 발명의 제2 실시 예에 따른 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 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) 입력 이미지를 제1 이미지 토큰 벡터로 변환하는 제1 단계, (BB) 상기 변환한 제1 이미지 토큰 벡터를 코드북이 포함하는 K(K는 양의 정수)개의 코드 벡터 중 어느 하나로 치환한 제2 이미지 토큰 벡터를 생성하는 제2 단계, (CC) 상기 생성한 제2 이미지 토큰 벡터에 전이 행렬을 적용한 정방향 확산을 통해 상기 제2 이미지 토큰 벡터를 흡수(Mask) 상태로 변환한 제3 이미지 토큰 벡터를 생성하는 제3 단계, (DD) 상기 생성한 제3 이미지 토큰 벡터로부터 역방향 확산을 통해 흡수 상태를 포함하지 않는 제4 이미지 토큰 벡터를 생성하는 제4 단계 및 (EE) 상기 생성한 제4 이미지 토큰 벡터를 고품질 이미지로 변환한 최종 이미지를 출력하는 제5 단계를 실행할 수 있을 것이며, 중복 서술을 위해 자세히 기재하지는 않았지만 본 발명의 제1 실시 예에 따른 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 장치(100) 및 본 발명의 제2 실시 예에 따른 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 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.

[0092] 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

[0093] 10: 프로세서

11: 제1 네트워크

12: 제2 네트워크

20: 네트워크 인터페이스

30: 메모리

40: 스토리지

41: 컴퓨터 프로그램

50: 정보 버스

100: 토큰 유사도 기반의 벡터-양자화 확산 모델을 활용한 고품질 이미지 생성 장치

110: 벡터-양자화 가변 오토인코더 모델

111: 인코더

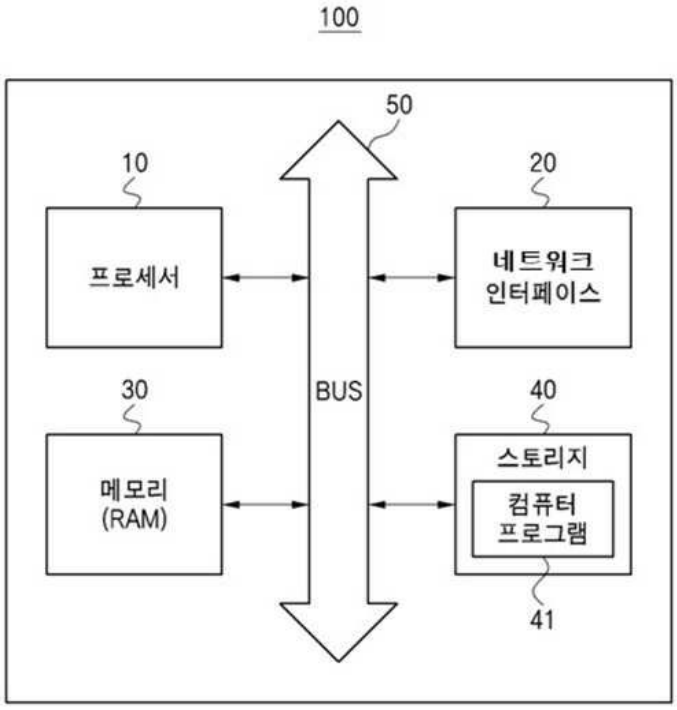
112: 코드북

113: 디코더

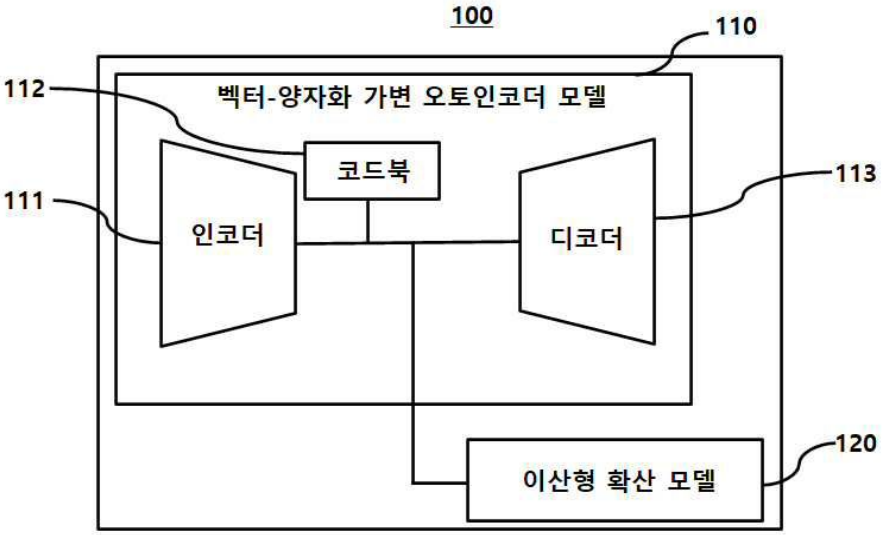
120: 이산형 확산 모델

도면

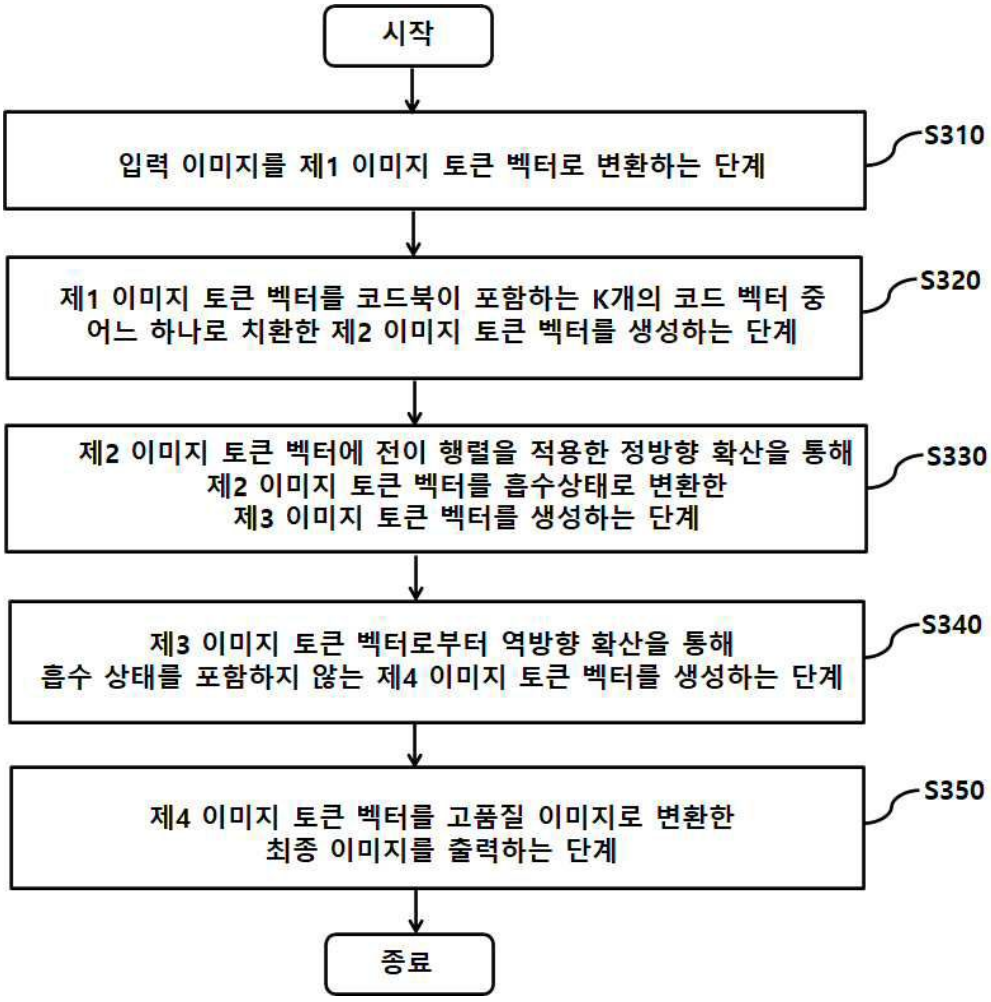
도면1



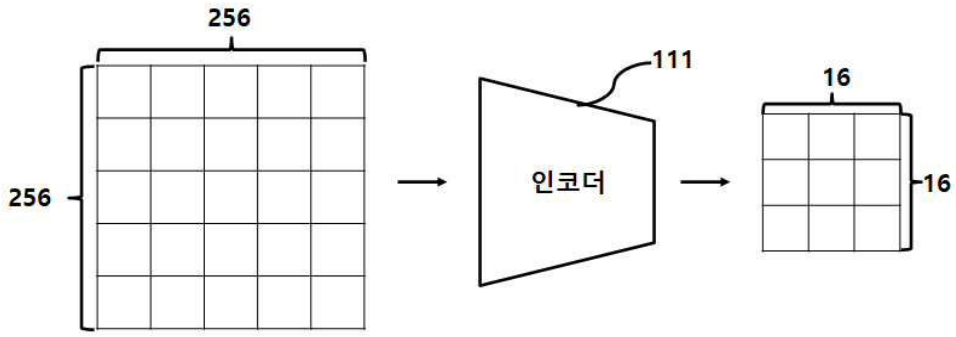
도면2


100
110
112
벡터-양자화 가변 오토인코더 모델
코드북
111
인코더
디코더
113
120
이산형 확산 모델

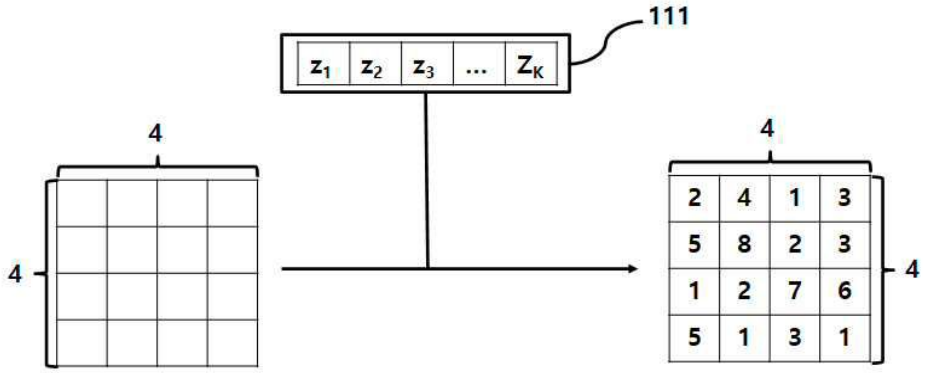
도면3


시작
입력 이미지를 제1 이미지 토큰 벡터로 변환하는 단계
S310
제1 이미지 토큰 벡터를 코드북이 포함하는 K개의 코드 벡터 중
어느 하나로 치환한 제2 이미지 토큰 벡터를 생성하는 단계
S320
제2 이미지 토큰 벡터에 전이 행렬을 적용한 정방향 확산을 통해
제2 이미지 토큰 벡터를 흡수상태로 변환한
제3 이미지 토큰 벡터를 생성하는 단계
S330
제3 이미지 토큰 벡터로부터 역방향 확산을 통해
흡수 상태를 포함하지 않는 제4 이미지 토큰 벡터를 생성하는 단계
S340
제4 이미지 토큰 벡터를 고품질 이미지로 변환한
최종 이미지를 출력하는 단계
S350
종료

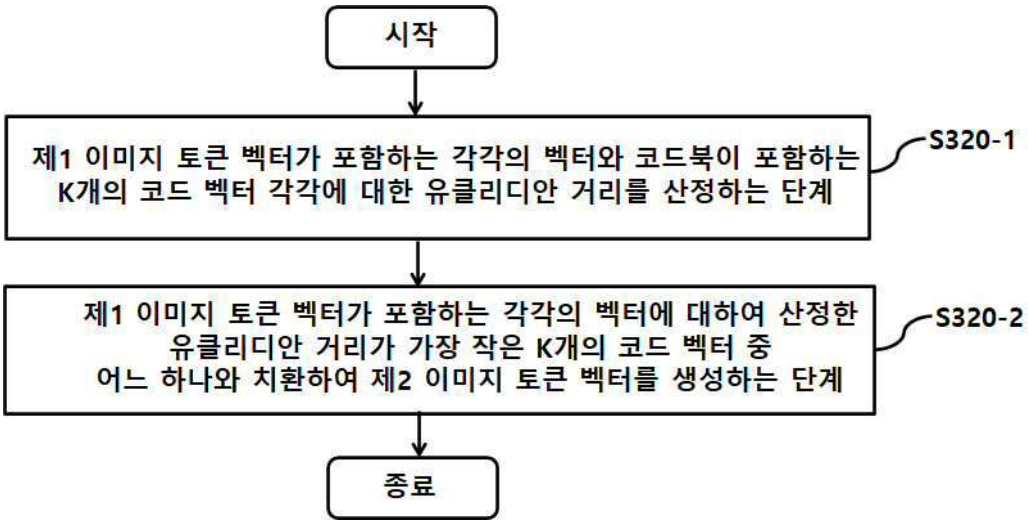
도면4


256
256
111
인코더
16
16

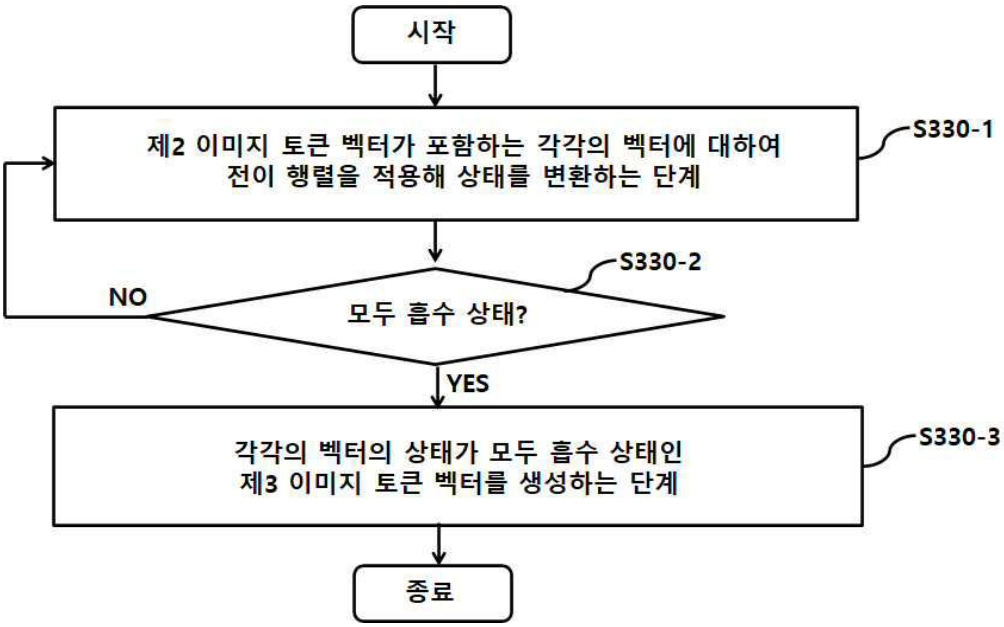
도면5


111
z1 z2 z3 ... ZK
4
4
4
2 4 1 3
5 8 2 3
1 2 7 6
5 1 3 1
4

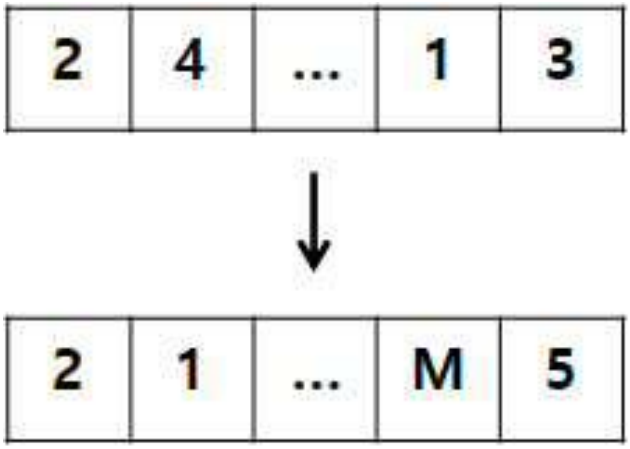
도면6


시작
제1 이미지 토큰 벡터가 포함하는 각각의 벡터와 코드북이 포함하는 K개의 코드 벡터 각각에 대한 유클리디안 거리를 산정하는 단계
S320-1
제1 이미지 토큰 벡터가 포함하는 각각의 벡터에 대하여 산정한 유클리디안 거리가 가장 작은 K개의 코드 벡터 중 어느 하나와 치환하여 제2 이미지 토큰 벡터를 생성하는 단계
S320-2
종료

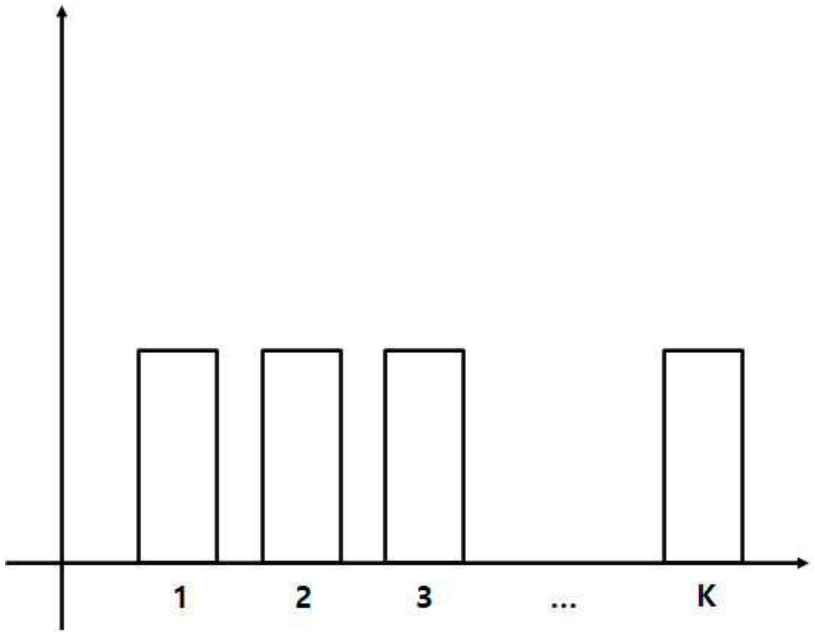
도면7


시작
제2 이미지 토큰 벡터가 포함하는 각각의 벡터에 대하여
전이 행렬을 적용해 상태를 변환하는 단계
S330-1
S330-2
NO
모두 흡수 상태?
YES
각각의 벡터의 상태가 모두 흡수 상태인
제3 이미지 토큰 벡터를 생성하는 단계
S330-3
종료

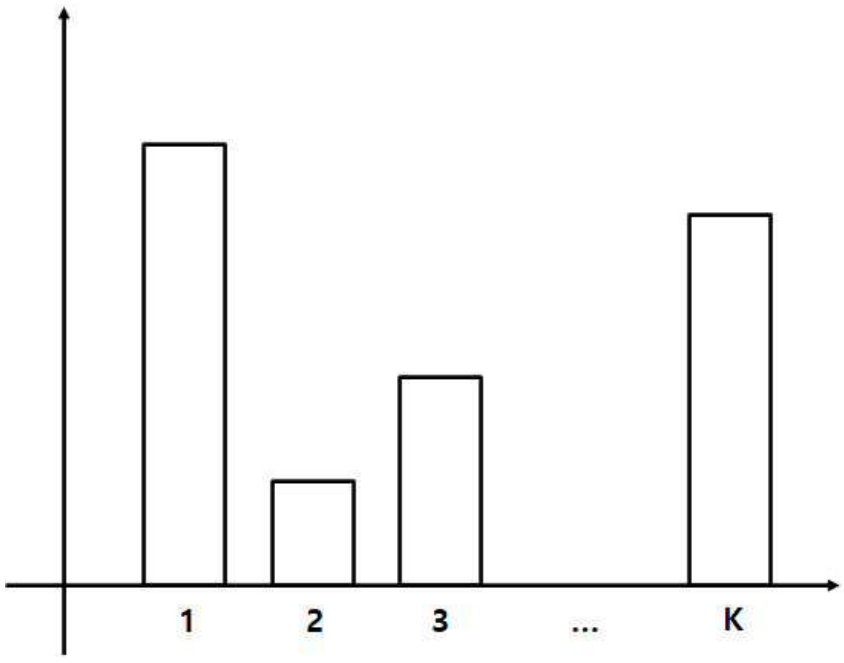
도면8


2
4
...
1
3
2
1
...
M
5

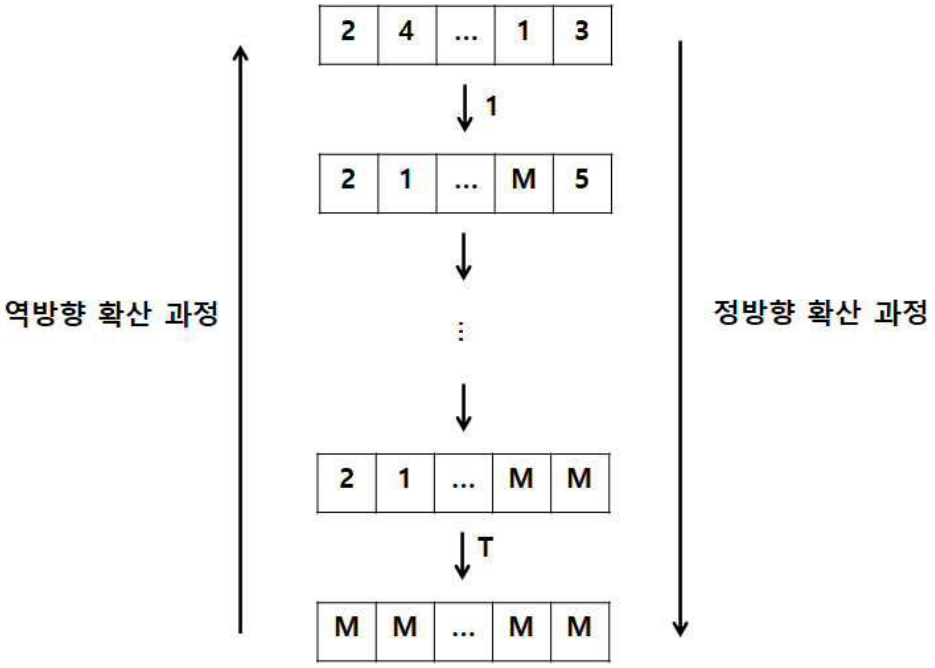
도면9



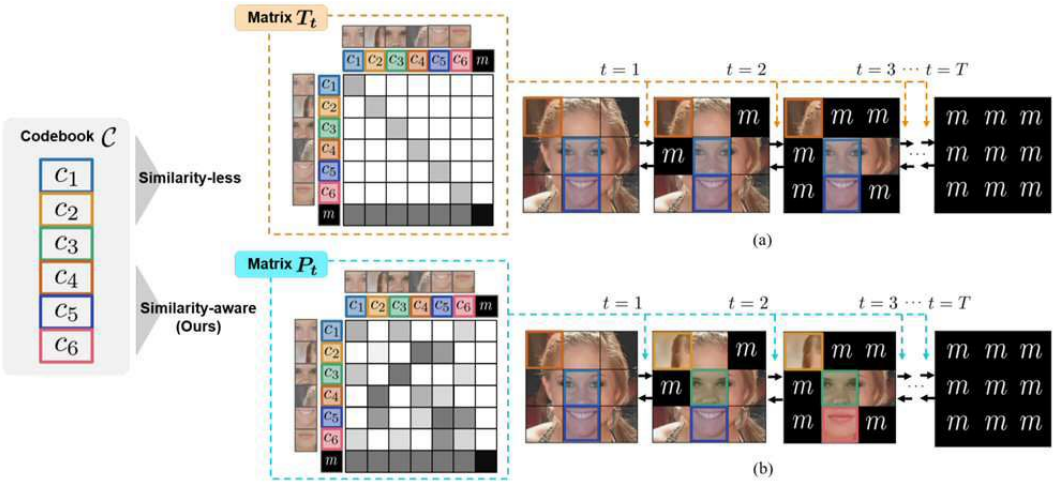
도면10



도면11


2 4 ... 1 3
1
2 1 ... M 5
2 1 ... M M
T
M M ... M M
역방향 확산 과정
정방향 확산 과정

도면12


Codebook $\mathcal{C}$
c_1
c_2
c_3
c_4
c_5
c_6
Similarity-less
Similarity-aware (Ours)
Matrix T_t
Matrix P_t
$t=1$
$t=2$
$t=3 \cdots t=T$
(a)
(b)

도면13



도면14

