

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구  
국제사무국

(43) 국제공개일  
2025년 3월 13일 (13.03.2025)



(10) 국제공개번호  
WO 2025/053602 A1

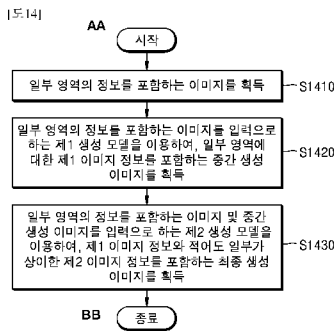
- (51) 국제특허분류:  
G06N 3/0475 (2023.01) G06T 5/77 (2024.01)  
G06N 3/0455 (2023.01) G06T 5/70 (2024.01)  
G06N 3/094 (2023.01) G06T 5/60 (2024.01)  
G06N 3/088 (2023.01) G06F 40/279 (2020.01)
- (21) 국제출원번호: PCT/KR2024/013315
- (22) 국제출원일: 2024년 9월 4일 (04.09.2024)
- (25) 출원언어: 한국어
- (26) 공개언어: 한국어
- (30) 우선권정보:  
10-2023-0117238 2023년 9월 4일 (04.09.2023) KR  
10-2023-0182370 2023년 12월 14일 (14.12.2023) KR  
10-2024-0006753 2024년 1월 16일 (16.01.2024) KR
- (71) 출원인: 삼성전자 주식회사 (SAMSUNG ELECTRONICS CO., LTD.) [KR/KR]; 16677 경기도 수원시 영통구 삼성로 129, Gyeonggi-do (KR).
- (72) 발명자: 조호철 (CHO, Hochul); 16677 경기도 수원시 영통구 삼성로 129, Gyeonggi-do (KR). 공내진 (KONG, Naejin); 16677 경기도 수원시 영통구 삼성로 129, Gyeonggi-do (KR). 이태미 (LEE, Tammy); 16677 경기도

수원시 영통구 삼성로 129, Gyeonggi-do (KR). 안범진 (AHN, Beomjin); 16677 경기도 수원시 영통구 삼성로 129, Gyeonggi-do (KR). 이유정 (LEE, Yujeong); 16677 경기도 수원시 영통구 삼성로 129, Gyeonggi-do (KR). 정진기 (JUNG, Jinki); 16677 경기도 수원시 영통구 삼성로 129, Gyeonggi-do (KR).

- (74) 대리인: 리앤록 특허법인 (Y.P.LEE, MOCK & PARTNERS); 06292 서울특별시 강남구 언주로30길 13 대림아크로텔 12층, Seoul (KR).
- (81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.
- (84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 역내 권리의 보호를 위하여): ARIPO (BW, CV, GH, GM,

(54) Title: METHOD FOR GENERATING PARTIAL REGION OF IMAGE BY USING GENERATIVE MODEL, AND ELECTRONIC DEVICE FOR PERFORMING SAME

(54) 발명의 명칭: 생성 모델을 이용하여 이미지의 일부 영역을 생성하는 방법 및 이를 수행하는 전자 장치



S1410 ... Acquire image including information on partial region  
 S1420 ... Acquire intermediate generated image including first image information on partial region by using first generative model that takes, as input, image including information on partial region  
 S1430 ... Acquire final generated image including second image information, which is at least partially different from first image information, by using second generative model that takes, as input, image including information on partial region and intermediate generated image  
 AA ... Start  
 BB ... End

(57) Abstract: In an embodiment of the present disclosure, provided are a method for generating a partial region of an image by using a generative model, and an electronic device for performing same. The method for generating a partial region of an image by using a generative model may comprise the steps of: acquiring an image including information on a partial region; acquiring an intermediate generated image including first image information on the partial region by inputting the image into a first generative model; and acquiring a final generated image including second image information, which is at least partially different from the first image information, by inputting the image and the intermediate generated image into a second generative model.

(57) 요약서: 본 개시의 일 실시예에 있어서, 생성 모델을 이용하여 이미지의 일부 영역을 생성하는 방법 및 이를 수행하는 전자 장치가 제공된다. 생성 모델을 이용하여 이미지의 일부 영역을 생성하는 방법은, 일부 영역의 정보를 포함하는 이미지를 획득하는 단계, 상기 이미지를 제1 생성 모델에 입력함으로써, 상기 일부 영역에 대한 제1 이미지 정보를 포함하는 중간 생성 이미지를 획득하는 단계, 및 상기 이미지 및 상기 중간 생성 이미지를 제2 생성 모델에 입력함으로써, 제1 이미지 정보와 적어도 일부가 상이한 제2 이미지 정보를 포함하는 최종 생성 이미지를 획득하는 단계를 포함할 수 있다.

KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ,  
UG, ZM, ZW), 유라시아 (AM, AZ, BY, KG, KZ, RU, TJ,  
TM), 유럽 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE,  
ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC,  
ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,  
TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,  
KM, ML, MR, NE, SN, TD, TG).

공개:

— 국제조사보고서와 함께 (조약 제21조(3))

## 명세서

### 발명의 명칭: 생성 모델을 이용하여 이미지의 일

### 부 영역을 생성하는 방법 및 이를 수행하는 전자 장치

#### 기술분야

- [1] 본 개시는 이미지를 처리하는 방법에 관한 것이며, 상세하게는 생성 모델을 이용하여 이미지를 생성하는 방법 및 이를 수행하는 전자 장치에 관한 것이다.

#### 배경기술

- [2] 생성형 AI (generative AI) 기술은 방대한 훈련용 데이터의 패턴과 구조를 학습하고, 그에 기초하여 입력 데이터와 유사한 새로운 데이터를 생성하는 기술을 의미한다. 생성형 AI 기술을 이용하면 텍스트에 대응되는 이미지를 얻거나, 원본 이미지에는 포함되지 않았던 영역까지 이미지를 확장할 수도 있다.
- [3] 생성형 AI 기술은 이미지 처리 분야에 적용되어 아웃페인팅(outpainting) 또는 인페인팅(inpainting)을 지원할 수 있다. 이미지의 스타일과 내용을 유지하면서 이미지를 확장하는 것을 아웃페인팅이라고 하고, 이미지 내 특정 영역에 채워질 이미지를 생성하는 것을 인페인팅이라고 한다.
- [4] 한편, 최근 사용자 경험을 중시하는 이미지 처리 기술의 발전으로 인해 일부 장치나 프로그램은 사용자가 이미지에 포함된 객체들의 위치나 크기를 변환시킬 수 있도록 하는 기능을 지원하기도 한다. 그런데, 객체가 변환된 후의 이미지에서 주변과 어울리지 않게 객체가 표시된다면 사용자의 만족도가 떨어질 수 있다.

#### 발명의 상세한 설명

##### 과제 해결 수단

- [5] 본 개시의 일 실시예에 있어서, 생성 모델을 이용하여 이미지의 일부 영역을 생성하는 방법이 제공될 수 있다. 상기 방법은, 상기 일부 영역의 정보를 포함하는 이미지를 획득하는 단계, 상기 이미지를 제1 생성 모델에 입력함으로써, 상기 일부 영역에 대한 제1 이미지 정보를 포함하는 중간 생성 이미지를 획득하는 단계, 및 상기 이미지 및 상기 중간 생성 이미지를 제2 생성 모델에 입력함으로써, 상기 제1 이미지 정보와 적어도 일부가 상이한 제2 이미지 정보를 포함하는 최종 생성 이미지를 획득하는 단계를 포함할 수 있다.
- [6] 본 개시의 일 실시예에 있어서, 전자 장치가 제공될 수 있다. 상기 전자 장치는, 적어도 하나의 인스트럭션을 저장하는 메모리, 및 적어도 하나의 프로세서를 포함하고, 상기 적어도 하나의 프로세서가 상기 메모리에 저장된 적어도 하나의 인스트럭션을 실행함으로써 전자 장치는, 일부 영역의 정보를 포함하는 이미지를 획득하고, 상기 이미지를 제1 생성 모델에 입력함으로써, 상기 일부 영역에 대한 제1 이미지 정보를 포함하는 중간 생성 이미지를 획득하고, 상기 이미지 및 상기 중간 생성 이미지를 제2 생성 모델에 입력함으로써, 상기 제1 이미지 정보와 적

어도 일부가 상이한 제2 이미지 정보를 포함하는 최종 생성 이미지를 획득할 수 있다.

- [7] 본 개시의 일 실시예에 있어서, 생성 모델을 이용하여 이미지의 일부 영역을 생성하는 방법을 컴퓨터에서 수행하기 위한 프로그램이 기록된 컴퓨터로 읽을 수 있는 기록매체가 제공될 수 있다.

### 도면의 간단한 설명

- [8] 본 개시는, 다음의 자세한 설명과 그에 수반되는 도면들의 결합으로 쉽게 이해될 수 있으며, 참조 번호(reference numerals)들은 구조적 구성요소(structural elements)를 의미한다.
- [9] 도 1은 본 개시의 일 실시예에 따른 생성 모델을 이용하여 이미지의 일부 영역을 생성하는 전자 장치를 설명하기 위한 개념도이다.
- [10] 도 2는 본 개시의 일 실시예에 따른 인코더 및 디코더를 설명하기 위한 개념도이다.
- [11] 도 3은 본 개시의 일 실시예에 따른 마스크 맵을 설명하기 위한 개념도이다.
- [12] 도 4는 본 개시의 일 실시예에 따른 제2 생성 모델의 동작을 설명하기 위한 개념도이다.
- [13] 도 5a는 본 개시의 일 실시예에 따른 제2 생성 모델에 입력되는 텍스트 가이드를 설명하기 위한 개념도이다.
- [14] 도 5b는 본 개시의 일 실시예에 따른 제2 생성 모델에 입력되는 이미지 가이드를 설명하기 위한 개념도이다.
- [15] 도 5c는 본 개시의 일 실시예에 따른 제2 생성 모델에 입력되는 텍스트 가이드 및 이미지 가이드를 설명하기 위한 개념도이다.
- [16] 도 5d는 본 개시의 일 실시예에 따른 제1 생성 모델의 출력이 이미지 가이드로 이용되는 실시예를 설명하기 위한 개념도이다.
- [17] 도 5e는 본 개시의 일 실시예에 따른 제1 생성 모델의 출력이 이미지 가이드로 이용되는 실시예를 설명하기 위한 개념도이다.
- [18] 도 5f는 본 개시의 일 실시예에 따른 텍스트 가이드와 함께, 제1 생성 모델의 출력이 이미지 가이드로 이용되는 실시예를 설명하기 위한 개념도이다.
- [19] 도 5g는 본 개시의 일 실시예에 따른 텍스트 가이드와 함께, 제1 생성 모델의 출력이 이미지 가이드로 이용되는 실시예를 설명하기 위한 개념도이다.
- [20] 도 6은 본 개시의 일 실시예에 따른 제1 생성 모델의 학습 방법을 설명하기 위한 개념도이다.
- [21] 도 7a 및 7b는 본 개시의 일 실시예에 따른 제2 생성 모델의 구성을 설명하기 위한 개념도이다.
- [22] 도 8은 본 개시의 일 실시예에 따른 인터프리터의 구성을 설명하기 위한 개념도이다.

- [23] 도 9a 및 9b는 본 개시의 일 실시예에 따른 중간 생성 이미지에 노이즈를 추가하는 실시예를 설명하기 위한 개념도이다.
- [24] 도 10a 내지 10c는 본 개시의 일 실시예에 따른 중간 생성 이미지의 퀄리티에 따라 디노이징 강도를 결정하는 실시예를 설명하기 위한 개념도이다.
- [25] 도 11은 본 개시의 일 실시예에 따른 생성 모델을 이용하여 이미지의 일부 영역을 생성하는 전자 장치를 설명하기 위한 개념도이다.
- [26] 도 12a 및 12b는 본 개시의 일 실시예에 따른 제5 인코더의 학습 방법을 설명하기 위한 개념도이다.
- [27] 도 13a 및 13b는 본 개시의 일 실시예에 따른 제5 인코더의 학습 방법을 설명하기 위한 개념도이다.
- [28] 도 14는 본 개시의 일 실시예에 따른 생성 모델을 이용하여 이미지의 일부 영역을 생성하는 방법을 설명하기 위한 흐름도이다.
- [29] 도 15는 도 14의 S1410 단계의 세부 단계들을 설명하기 위한 흐름도이다.
- [30] 도 16은 도 14의 S1430 단계의 세부 단계들을 설명하기 위한 흐름도이다.
- [31] 도 17은 도 14의 S1430 단계의 세부 단계들을 설명하기 위한 흐름도이다.
- [32] 도 18은 도 14의 S1430 단계의 세부 단계들을 설명하기 위한 흐름도이다.
- [33] 도 19는 도 14의 S1430 단계의 세부 단계들을 설명하기 위한 흐름도이다.
- [34] 도 20은 본 개시의 일 실시예에 따른 이미지를 인코딩하는 인코더를 학습하는 방법을 설명하기 위한 흐름도이다.
- [35] 도 21은 본 개시의 일 실시예에 따른 생성 모델을 이용하여 이미지의 일부 영역을 생성하는 방법을 설명하기 위한 흐름도이다.
- [36] 도 22는 본 개시의 일 실시예에 따른 전자 장치의 구성을 설명하기 위한 블록도이다.
- [37] 도 23은 본 개시의 일 실시예에 따른 전자 장치 및 서버 장치의 구성을 설명하기 위한 블록도이다.

### 발명의 실시를 위한 형태

- [38] 본 명세서의 실시예들에서 사용되는 용어는 본 개시의 기능을 고려하면서 가능한 현재 널리 사용되는 일반적인 용어들을 선택하였으나, 이는 당 분야에 종사하는 기술자의 의도 또는 판례, 새로운 기술의 출현 등에 따라 달라질 수 있다. 또한, 특정한 경우는 출원인이 임의로 선정한 용어도 있으며, 이 경우 해당되는 실시예의 설명 부분에서 상세히 그 의미를 기재할 것이다. 따라서 본 명세서에서 사용되는 용어는 단순한 용어의 명칭이 아닌, 그 용어가 가지는 의미와 본 개시의 전반에 걸친 내용을 토대로 정의되어야 한다.
- [39] 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함할 수 있다. 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 용어들은 본 명세서에 기재된 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가질 수 있다.

- [40] 본 개시 전체에서 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있음을 의미한다. 또한, 본 명세서에 기재된 "..부", "..모듈" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어 또는 소프트웨어로 구현되거나 하드웨어와 소프트웨어의 결합으로 구현될 수 있다.
- [41] 본 개시에서 사용된 표현 "~하도록 구성된(또는 설정된)(configured to)"은 상황에 따라, 예를 들면, "~에 적합한(suitable for)", "~하는 능력을 가지는(having the capacity to)", "~하도록 설계된(designed to)", "~하도록 변경된(adapted to)", "~하도록 만들어진(made to)", 또는 "~를 할 수 있는(capable of)"과 바꾸어 사용될 수 있다. 용어 "~하도록 구성된(또는 설정된)"은 하드웨어적으로 "특별히 설계된(specifically designed to)" 것만을 반드시 의미하지 않을 수 있다. 대신, 어떤 상황에서는, "~하도록 구성된 시스템"이라는 표현은, 그 시스템이 다른 장치 또는 부품들과 함께 "~할 수 있는" 것을 의미할 수 있다. 예를 들면, 문구 "A, B, 및 C를 수행하도록 구성된(또는 설정된) 프로세서"는 해당 동작을 수행하기 위한 전용 프로세서(예: 임베디드 프로세서), 또는 메모리에 저장된 하나 이상의 소프트웨어 프로그램들을 실행함으로써, 해당 동작들을 수행할 수 있는 범용 프로세서(generic-purpose processor)(예: CPU 또는 application processor)를 의미할 수 있다.
- [42] 또한, 본 개시에서 일 구성요소가 다른 구성요소와 "연결된다" 거나 "접속된다" 등으로 언급된 때에는, 상기 일 구성요소가 상기 다른 구성요소와 직접 연결되거나 또는 직접 접속될 수도 있지만, 특별히 반대되는 기재가 존재하지 않는 이상, 중간에 또 다른 구성요소를 매개하여 연결되거나 또는 접속될 수도 있다고 이해되어야 할 것이다.
- [43] 본 개시에서, '인공지능(Artificial Intelligence)'과 관련된 기능은 프로세서와 메모리를 통해 동작된다. 프로세서는 하나 또는 복수의 프로세서로 구성될 수 있다. 이때, 하나 또는 복수의 프로세서는 CPU, AP, DSP(Digital Signal Processor) 등과 같은 범용 프로세서, GPU, VPU(Vision Processing Unit)와 같은 그래픽 전용 프로세서 또는 NPU와 같은 인공지능 전용 프로세서일 수 있다. 하나 또는 복수의 프로세서는, 메모리에 저장된 기 정의된 동작 규칙 또는 인공지능 모델에 따라, 입력 데이터를 처리하도록 제어한다. 또는, 하나 또는 복수의 프로세서가 인공지능 전용 프로세서인 경우, 인공지능 전용 프로세서는, 특정 인공지능 모델의 처리에 특화된 하드웨어 구조로 설계될 수 있다.
- [44] 기 정의된 동작 규칙 또는 인공지능 모델은 학습을 통해 만들어진 것을 특징으로 한다. 여기서, 학습을 통해 만들어진다는 것은, 기본 인공지능 모델이 학습 알고리즘에 의하여 다수의 학습 데이터들을 이용하여 학습됨으로써, 원하는 특성(또는, 목적)을 수행하도록 설정된 기 정의된 동작 규칙 또는 인공지능 모델이 만들어짐을 의미한다. 이러한 학습은 본 개시에 따른 인공지능이 수행되는 기기 자체에서 이루어질 수도 있고, 별도의 서버 및/또는 시스템을 통해 이루어 질 수도

있다. 학습 알고리즘의 예로는, 지도형 학습(supervised learning), 비지도형 학습(unsupervised learning), 준지도형 학습(semi-supervised learning) 또는 강화 학습(reinforcement learning)이 있으나, 전술한 예에 한정되지 않는다.

- [45] 본 개시의 일 실시예에 있어서, '인공지능 모델'은 신경망 모델을 포함할 수 있다. 신경망 모델은 복수의 신경망 레이어들로 구성될 수 있다. 복수의 신경망 레이어들 각각은 복수의 가중치들(weight values)을 갖고 있으며, 이전(previous) 레이어의 연산 결과와 복수의 가중치들 간의 연산을 통해 신경망 연산을 수행한다. 복수의 신경망 레이어들이 갖고 있는 복수의 가중치들은 인공지능 모델의 학습 결과에 의해 최적화될 수 있다. 예를 들어, 학습 과정 동안 인공지능 모델에서 획득한 손실(loss) 값 또는 비용(cost) 값이 감소 또는 최소화되도록 복수의 가중치들이 갱신될 수 있다. 신경망 모델은 심층 신경망(DNN: Deep Neural Network)를 포함할 수 있으며, 예를 들어, CNN(Convolutional Neural Network), RNN(Recurrent Neural Network), RBM(Restricted Boltzmann Machine), DBN(Deep Belief Network), BRDNN(Bidirectional Recurrent Deep Neural Network) 또는 심층 Q-네트워크(Deep Q-Networks) 등이 있으나, 전술한 예에 한정되지 않는다.
- [46] 본 개시에서, '생성 모델'은 주어진 입력 데이터에 기초하여 새로운 데이터를 생성하는 모델을 나타낼 수 있다. 예를 들어, 생성 모델은 다양한 형태의 데이터(예컨대, 텍스트, 이미지, 비디오, 사운드, 무작위 벡터 데이터 등)를 입력받을 수 있다. 예를 들어, 생성 모델은 새로운 텍스트, 이미지, 비디오, 사운드, 또는 이들의 조합을 생성할 수 있다. 본 개시의 일 실시예에 있어서, 생성 모델은 인공지능 모델일 수 있다. 예를 들어, 생성 모델은 GAN(generative adversarial network) 모델, VAE(variational autoencoder) 모델, 확산 모델, 트랜스포머(transformer) 등을 포함할 수 있으나, 전술한 예에 한정되지 않는다.
- [47] 본 개시에서, '인페인팅'(inpainting)은 원본 이미지의 내부 영역에 대한 추론을 수행하는 작업을 나타낼 수 있다.
- [48] 본 개시에서, '아웃페인팅'(outpainting)은 원본 이미지의 외부 영역에 대한 추론을 수행하는 작업을 나타낼 수 있다.
- [49] 본 개시에서, '마스크'(mask)는 특정 이미지의 전체 영역 중에서 추론이 필요한 영역(또는 미지(unknown) 영역)을 나타낼 수 있다. 예를 들어, '마스킹된 이미지'는 마스킹되지 않은 영역에 대응하는 픽셀 정보와, 마스크에 대응하는 영역에 대한 경계 정보를 포함할 수 있다. 픽셀 정보는 픽셀의 위치 정보(예컨대, 좌표 값) 및 색상 정보(예컨대, RGB 값)를 포함할 수 있다.
- [50] 본 개시에서, '마스크 맵'(mask map)은 이미지의 전체 영역 중에서, 마스킹된 영역과 마스킹되지 않은 영역을 구분하는 데이터를 나타낼 수 있다. 예를 들어, 마스크 맵은 이진화 맵일 수 있다. 예를 들어, 마스킹된 영역은 제1 값으로 표현되고, 마스킹되지 않은 영역은 제2 값으로 표현될 수 있다.
- [51] 본 개시에서, '디노이징'은 제2 생성 모델에 입력되어 출력되는 한 번의 동작으로 정의될 수 있다. 본 개시에서, '디노이징 차수'는 디노이징을 반복한 횟수로 정

의될 수 있다. 본 개시에서, '디노이징 총 차수'는 디노이징을 얼마나 반복할 것인지에 대한 하이퍼파라미터로 정의될 수 있다. 본 개시의 일 실시예에 있어서, 디노이징 총 차수는 사용자 또는 제조사의 설정에 따라 상이할 수 있다.

[52] 본 개시에서, '가이던스'는 추론되어야 할 이미지를 나타내는 정보를 나타낼 수 있다. 예를 들어, 해당 정보는 추론될 이미지를 나타내는 임의의 정보일 수 있다. 예를 들어, 가이던스는 텍스트로부터 생성된 텍스트 가이던스 또는 이미지로부터 생성된 이미지 가이던스를 포함할 수 있으나, 가이던스의 종류는 이에 한정되지 않는다. 가이던스 또는 가이던스 정보는 디노이징 동작을 수행하는 동안 제2 생성 모델에 입력될 수 있다.

[53] 도 1은 본 개시의 일 실시예에 따른 생성 모델을 이용하여 이미지의 일부 영역을 생성하는 전자 장치를 설명하기 위한 개념도이다. 도 1을 참조하면, 전자 장치(1000)는 제1 생성 모델(1100) 및 제2 생성 모델(1200)을 이용하여 마스크된 이미지로부터 최종 생성 이미지를 획득할 수 있다. 본 개시의 일 실시예에 있어서, 제1 생성 모델(1100)과 제2 생성 모델(1200) 각각은 미지 영역이 존재하는 이미지에 기초하여 미지 영역을 추론한 전체 이미지를 생성할 수 있다. 예를 들어, 제1 생성 모델(1100)은 입력 이미지에 기초하여 제1 이미지를 생성할 수 있으며, 제2 생성 모델(1200)은 입력 이미지에 기초하여 제2 이미지를 생성할 수 있다. 여기서, 미지 영역은 입력 이미지에서 이미지 정보가 포함되지 않거나 사용자가 의도한 이미지 정보가 포함되지 않은 영역일 수 있다. 제1 생성 모델(1100)과 제2 생성 모델(1200) 각각은 입력 이미지에 대해 인페인팅 또는 아웃페인팅을 수행할 수 있다. 본 개시의 일 실시예에 따르면, 제1 생성 모델의 출력을 제2 생성 모델의 추론 동작에 활용함으로써 이미지 추론의 속도 및/또는 성능을 획기적으로 개선할 수 있다.

[54] 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 다양한 형태로 구현될 수 있다. 예를 들어, 전자 장치(1000)는 퍼스널 컴퓨터(personal computer; PC), 단말기(terminal), 휴대폰(portable telephone), 스마트폰(smart phone), 태블릿(tablet) PC, 휴대 장치(handheld device), 착용 장치(wearable device), 서버(server) 장치 등을 포함할 수 있다.

[55] 전자 장치(1000)는 일부 영역의 정보를 포함하는 이미지를 획득할 수 있다. 이미지의 일부 영역은 사용자 또는 제조사의 설정에 따라 미리 정의될 수 있다. 일부 실시예에서, 일부 영역은 이미지 내에서 대상 영역 또는 후보 영역으로 지칭될 수 있다. 일부 실시예에서, 일부 영역의 정보는 일부 영역에 대한 정보, 일부 영역과 관련된 정보, 또는 일부 영역에 해당하는 정보로도 지칭될 수 있다. 예를 들어, 일부 영역의 정보는 이미지의 일부 영역에 대응하는 위치 정보 및/또는 이미지의 일부 영역과 일부 영역을 제외한 영역 간의 경계 정보 등을 포함할 수 있다. 예를 들어, 일부 영역의 정보는 이미지 상의 바운딩 박스(bounding box)의 위치 정보를 포함할 수 있다. 예를 들어, 일부 영역의 정보는 일부 영역에 대응하는

이미지 상의 픽셀 좌표 값들을 포함할 수 있다. 본 개시의 일 실시예에 있어서, 일부 영역은 마스크로 지칭될 수도 있다.

- [56] 본 개시의 일 실시예에 있어서, 전자 장치(100)는 센서 및 디스플레이를 포함할 수 있다. 전자 장치(100)는 센서를 이용하여 객체를 촬영한 원본 이미지를 획득할 수 있다. 센서는 전자 CCD (Charge-Coupled Device) 센서 및 CMOS (Complementary Metal-Oxide-Semiconductor) 센서를 포함하되, 이에 국한되지 않는다. 일부 실시예에서, 센서는 카메라로 지칭될 수 있다. 원본 이미지는 RGB 이미지를 포함하되, 이에 국한되지 않을 수 있다. 전자 장치(100)는 디스플레이를 통해 원본 이미지를 표시할 수 있다. 전자 장치(100)는 사용자 인터페이스를 통해 원본 이미지 중 일부 영역에 대한 사용자 입력을 획득할 수 있다. 예를 들어, 일부 영역은 원본 이미지의 내부 영역 및 외부 영역 중 적어도 하나를 포함할 수 있다. 전자 장치(100)는 사용자 입력에 기초하여 일부 영역의 정보를 포함하는 이미지를 생성할 수 있다. 일부 영역의 정보를 포함하는 이미지는 마스크된 이미지로도 지칭될 수 있다. 예를 들어, 일부 영역의 정보를 포함하는 이미지는 원본 이미지에 마스크가 추가된 형태로 구성될 수 있다. 본 개시의 일 실시예에 있어서, 전자 장치(100)는 일부 영역의 정보를 포함하는 이미지 및 사용자 입력 중 적어도 하나에 기초하여 마스크 맵을 생성할 수 있다. 본 개시의 일 실시예에 있어서, 전자 장치(100)는 외부 서버로부터 원본 이미지, 일부 영역의 정보, 및 마스크 맵 중 적어도 하나를 수신할 수 있다.
- [57] 전자 장치(100)는 일부 영역의 정보를 포함하는 이미지를 제1 생성 모델(1100)에 입력할 수 있다. 예를 들어, 제1 생성 모델(1100)은 기 학습된 GAN 모델을 포함할 수 있다. 예를 들어, 제1 생성 모델(1100)은 기 학습된 GAN 모델 중 생성기(generator) 모델을 포함할 수 있다. 본 개시의 일 실시예에 있어서, 전자 장치(100)는 마스크된 이미지와 마스크 맵으로 구성되는 이미지 쌍을 제1 생성 모델에 입력할 수 있다.
- [58] 전자 장치(100)는 제1 생성 모델(1100)로부터 일부 영역에 대한 제1 이미지 정보를 포함하는 중간 생성 이미지를 획득할 수 있다. 중간 생성 이미지는 일부 영역에 대응하는 적어도 하나의 픽셀의 색상 정보를 포함할 수 있다. 예를 들어, 중간 생성 이미지는 일부 영역에 대응하는 적어도 하나의 픽셀의 색상 정보와 함께, 전체 영역에 대응하는 픽셀들의 색상 정보를 포함할 수 있다. 그러나 본 개시는 이에 한정되지 않으며, 중간 생성 이미지는 일부 영역에 대응하는 적어도 하나의 픽셀의 색상 정보를 포함하고, 일부 영역을 제외한 픽셀들의 색상 정보를 포함하지 않을 수 있다.
- [59] 전자 장치(100)는 제2 생성 모델(1200)에 일부 영역의 정보를 포함하는 이미지 및 중간 생성 이미지를 입력할 수 있다. 제2 생성 모델(1200)은 노이즈로부터 이미지를 복원하는 인공지능 모델을 포함할 수 있다. 예를 들어, 제2 생성 모델(1200)은 기 학습된 확산 모델을 포함할 수 있다.

- [60] 본 개시의 일 실시예에 있어서, 제1 생성 모델(1100)은 제2 생성 모델(1200)보다 더 적은 레이어 및/또는 더 적은 가중치 값으로 구성될 수 있다. 본 개시의 일 실시예에 있어서, 제1 생성 모델(1100)의 처리 속도는 제2 생성 모델의 처리 속도보다 빠를 수 있다. 본 개시의 일 실시예에 있어서, 제1 생성 모델(1100)이 차지하는 메모리 용량은 제2 생성 모델(1200)이 차지하는 메모리 용량보다 작을 수 있다.
- [61] 본 개시의 일 실시예에 있어서, 중간 생성 이미지는 제2 생성 모델(1200)에 입력되기 전에 전처리될 수 있다. 예를 들어, 중간 생성 이미지는 이미지 임베딩 및/또는 이미지에 대응하는 텍스트 임베딩으로 변환될 수 있다. 본 개시에서, '임베딩'은 고차원의 데이터로부터 변환된 저차원의 데이터를 나타낼 수 있다. 예를 들어, 임베딩은 임베딩 벡터, 특징 벡터, 특징 표현, 잠재적 벡터, 또는 잠재적 표현 등으로도 지칭될 수 있다. 그러나 본 개시는 이에 국한되지 않으며, 이와 같이 임베딩은 고차원 데이터를 다른 방식으로 변환한 저차원 데이터를 포함할 수 있다.
- [62] 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 일부 영역의 정보와 노이즈 정보를 포함하는 이미지를 기반으로 연결된 이미지를 생성할 수 있다. 예를 들어, 전자 장치(1000)는 기 정의된 초기 노이즈에 일부 영역의 정보를 포함하는 이미지를 연결(concatenating)할 수 있다. 전자 장치(1000)는 연결된 이미지를 제2 생성 모델(1200)에 입력할 수 있다. 전자 장치(1000)는 제2 생성 모델(1200)의 출력에 일부 영역의 정보를 포함하는 이미지를 연결할 수 있다. 전자 장치(1000)는 연결된 이미지를 제2 생성 모델(1200)에 입력할 수 있다. 전자 장치(1000)는 연결된 이미지를 제2 생성 모델(1200)에 입력하는 동작을 반복할 수 있다. 예를 들어, 전자 장치(1000)는 미리 정의된 디노이징(denoising) 총 차수만큼, 연결된 이미지를 제2 생성 모델(1200)에 입력하는 동작을 반복적으로 수행할 수 있다.
- [63] 전자 장치(1000)는 기 정의된 디노이징 총 차수만큼 동작을 반복했는지를 결정할 수 있다. 기 정의된 디노이징 총 차수만큼 동작을 반복한 것으로 결정한 것에 기초하여, 전자 장치(1000)는 제2 생성 모델(1200)으로부터 최종 생성 이미지를 획득할 수 있다.
- [64] 본 개시의 일 실시예에 있어서, 제2 생성 모델(1200)은 적어도 하나의 레이어를 포함할 수 있다. 전자 장치(1000)는 중간 생성 이미지 또는 중간 생성 이미지에 대응하는 이미지 정보(예컨대, 이미지 임베딩)를 제2 생성 모델(1200)의 적어도 하나의 레이어에 입력할 수 있다.
- [65] 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 중간 생성 이미지에 대한 디노이징 강도(denoising strength)을 획득할 수 있다. 디노이징 강도는 이미지에 노이즈를 얼마나 강하게 추가할 것인지에 대한 수치를 나타낼 수 있다. 예를 들어, 디노이징 강도가 0~1 사이의 값을 갖는 것을 가정할 때, 디노이징 강도가 0일 경우 이미지에 노이즈가 추가되지 않으며, 디노이징 강도가 1일 경우 이미지가 완전한 무작위 노이즈로 변경될 수 있다. 예를 들어, 디노이징 강도가 0일 경우 이미지에 노이즈가 추가되지 않을 수 있다. 반면, 디노이징 강도가 1일 경우 이미지가 무작위 노이즈를 포함하도록 변경될 수 있다. 전자 장치(1000)는 디노이징 강

도에 기초하여 중간 생성 이미지에 추가할 노이즈 량을 결정할 수 있다. 노이즈 량은 중간 생성 이미지에 추가할 노이즈의 정도를 나타낼 수 있다. 본 개시의 일 실시예에 있어서, 디노이징 강도는 사용자 또는 제조사의 설정에 따라 기 정의될 수 있다. 예를 들어, 전자 장치(1000)는 사용자 인터페이스를 통해 디노이징 강도에 대응하는 사용자 입력을 획득할 수 있다. 전자 장치(1000)는 사용자 입력에 기초하여 디노이징 강도를 결정할 수 있다. 전자 장치(1000)는 디노이징 강도에 기초하여 중간 생성 이미지에 노이즈를 추가할 수 있다.

- [66] 본 개시의 일 실시예에 있어서, 중간 생성 이미지에 대한 디노이징 강도를 획득하는 기능과, 디노이징 강도에 기초하여 중간 생성 이미지에 노이즈를 추가하는 기능 중 적어도 일부는 제2 생성 모델(1200)에 의해 수행될 수도 있다. 예를 들어, 제2 생성 모델(1200)은 중간 생성 이미지를 수신할 수 있다. 제2 생성 모델(1200)은 중간 생성 이미지로부터 디노이징 강도를 획득할 수 있다. 제2 생성 모델(1200)은 디노이징 강도에 기초하여 중간 생성 이미지에 노이즈를 추가할 수 있다. 노이즈가 추가된 중간 생성 이미지는 제2 생성 모델(1200)의 적어도 하나의 레이어에 입력될 수 있다.
- [67] 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 중간 생성 이미지의 품질을 측정할 수 있다. 예를 들어, 전자 장치(1000)는 중간 생성 이미지에 기초하여 신뢰 값을 획득할 수 있다. 예를 들어, 전자 장치(1000)는 중간 생성 이미지에 기초하여 예측 신뢰 값을 획득할 수 있다. 예를 들어, 예측 신뢰 값은 제1 생성 모델(1100)이 출력(또는 예측, 추론)한 중간 생성 이미지를 신뢰할 수 있는 정도를 나타낼 수 있다. 전자 장치(1000)는 예측 신뢰 값에 기초하여 디노이징 강도를 결정할 수 있다. 예를 들어, 복수의 임계 범위들에 대응하는 디노이징 강도가 매핑될 수 있다. 예를 들어, 제1 임계 범위에 제1 디노이징 강도가 매핑되고, 제2 임계 범위에 제2 디노이징 강도가 매핑될 수 있다. 전자 장치(1000)는 복수의 임계 범위들 중 예측 신뢰 값에 대응하는 임계 범위를 결정할 수 있다. 전자 장치(1000)는 결정된 임계 범위에 매핑된 디노이징 강도로 중간 생성 이미지에 노이즈를 추가할 수 있다.
- [68] 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 디노이징 강도에 기초하여, 기 정의된 디노이징 총 차수 중 노이즈가 추가된 중간 생성 이미지에 대응하는 대상 디노이징 차수를 결정할 수 있다. 전자 장치(1000)는 현재 디노이징 차수를 대상 디노이징 차수로 설정할 수 있다.
- [69] 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 제2 생성 모델(1200)이 출력하는 이미지를 후처리하여 최종 생성 이미지를 획득할 수 있다. 예를 들어, 전자 장치(1000)는 제2 생성 모델(1200)이 출력하는 이미지 중 일부 영역(즉, 미지 영역 또는 대상 영역)에 대응하는 이미지 정보와, 최초 입력된 이미지(일부 영역의 정보를 포함하는 이미지로 지칭될 수도 있음) 중 일부 영역을 제외한 영역(즉, 기지 영역)에 대응하는 이미지 정보에 기초하여, 최종 생성 이미지를 획득할 수 있다.

- [70] 도 2는 본 개시의 일 실시예에 따른 인코더 및 디코더를 설명하기 위한 개념도이다. 도 1에서 설명한 내용과 중복되는 내용은 생략한다.
- [71] 도 2를 참조하면, 전자 장치(1000)는 제1 인코더(1310) 및 디코더(1400)를 포함할 수 있다. 제1 인코더(1310)는 이미지를 인코딩할 수 있다. 디코더(1400)는 이미지 임베딩을 디코딩함으로써 이미지를 출력할 수 있다. 본 개시의 일 실시예에 있어서, 제1 인코더(1310) 및 디코더(1400)는 기 학습된 인공지능 모델일 수 있다.
- [72] 본 개시의 일 실시예에 있어서, 제1 인코더(1310) 및 디코더(1400)는 변이형 오토인코더(Variational Autoencoder; VAE) 모델 또는 벡터 양자화 생성적 적대 신경망(VQGAN; Vector Quantized Generative Adversarial Network)으로 구현될 수 있다. 예를 들어, 제1 인코더(1310)는 변이형 오토인코더 또는 벡터 양자화 생성적 적대 신경망의 인코더 부분일 수 있다. 예를 들어, 디코더(1400)는 변이형 오토인코더 또는 벡터 양자화 생성적 적대 신경망의 디코더 부분일 수 있다. 제1 인코더(1310) 및 디코더(1400)는 변이형 오토인코더 모델 또는 벡터 양자화 생성적 적대 신경망의 학습 방식으로 학습될 수 있다. 예를 들어, 제1 인코더(1310)의 출력은 디코더(1400)에 입력될 수 있다. 예를 들어, 제1 인코더(1310)의 입력과 디코더(1400)의 출력을 비교함으로써 제1 인코더(1310) 및 디코더(1400)가 함께 학습될 수 있다. 그러나 본 개시는 이에 국한되지 않으며, 따라서 다른 실시예에 따르면, 제1 인코더(1310) 및/또는 디코더(1400)는 다른 방식으로 구현될 수도 있다.
- [73] 본 개시의 일 실시예에 있어서, 제1 인코더(1310)는 입력 이미지에 대응하는 잠재(latent) 벡터를 출력할 수 있다. 예를 들어, 잠재 벡터는 평균과 분산으로 표현되는 가우시안(Gaussian) 확률 분포에 기반한 확률 값을 나타낼 수 있다. 제1 인코더(1310)는 제2 생성 모델에 잠재 벡터를 전달할 수 있다.
- [74] 본 개시의 일 실시예에 있어서, 제2 생성 모델(1200)은 잠재 벡터 (이하에서, 최종 노이즈로도 지칭될 수 있음)를 출력할 수 있다. 디코더(1400)는 제2 생성 모델(1200)의 출력인 잠재 벡터를 디코딩함으로써 최종 생성 이미지를 출력할 수 있다.
- [75] 도 3은 본 개시의 일 실시예에 따른 마스크 맵을 설명하기 위한 개념도이다. 도 1 및 2에서 설명한 내용과 중복되는 내용은 생략한다.
- [76] 도 3을 참조하면, 전자 장치(1000)는 이미지의 일부 영역(예컨대, 추론하고자 하는 미지 영역 또는 대상 영역)의 위치 정보를 포함하는 마스크 맵을 획득할 수 있다. 예를 들어, 전자 장치(1000)는 외부 서버로부터 이미지에 대응하는 마스크 맵을 획득할 수 있다. 예를 들어, 전자 장치(1000)는 이미지에 기초하여 마스크 맵을 생성할 수 있다. 예를 들어, 전자 장치(1000)는 일부 영역(예컨대, 미지 영역 또는 대상 영역)을 제1 값(예컨대, 도 3의 흰 색 영역)으로 표현하고, 일부 영역을 제외한 영역(예컨대, 보존 영역)을 제2 값(예컨대, 도 3의 검은색 영역)으로 표현한 마스크 맵을 생성할 수 있다. 예를 들어, 마스크 맵은 제1 값 또는 제2 값으로 구성되는 이진 이미지일 수 있다. 그러나 본 개시는 이에 국한되지 않으며, 따라서 마스크 맵은 다른 형식으로 정의될 수 있다.

- [77] 전자 장치(1000)는 이미지와 마스크 맵을 연결(concatenating)할 수 있다. 예를 들어, 이미지는 RGB 값의 R 값, G 값, 및 B 값 각각을 하나의 채널로 하여 3 채널의 이미지로 표현될 수 있다. 예를 들어, 마스크 맵은 1 채널의 이미지로 표현될 수 있다. 마스크 맵이 연결된 이미지는 총 4개의 채널을 가질 수 있다. 그러나 본 개시는 이에 한정되지 않으며, 이미지의 채널의 개수 또는 마스크 맵의 채널의 개수는 이에 한정되지 않는다. 본 개시에서, 채널은 3차원 입력 데이터 중 하나의 차원을 나타낼 수 있다.
- [78] 전자 장치(1000)는 연결된 이미지를 제1 생성 모델(1100)에 전달할 수 있다. 전자 장치(1000)는 연결된 이미지를 제2 생성 모델(1200)에 전달할 수 있다.
- [79] 도 4는 본 개시의 일 실시예에 따른 제2 생성 모델의 동작을 설명하기 위한 개념도이다. 도 1 내지 3에서 설명한 내용과 중복되는 내용은 생략한다.
- [80] 도 4를 참조하면, 전자 장치(1000)는 이미지를 제1 인코더(1310)에 전달할 수 있다. 제1 인코더(1310)는 이미지를 인코딩할 수 있다. 전자 장치(1000)는 인코딩된 이미지(예컨대, 잠재 벡터)( $Z_{\text{image}}$ )를 획득할 수 있다.
- [81] 전자 장치(1000)는 마스크 맵을 조절할 수 있다. 예를 들어, 전자 장치(1000)는 마스크 맵을 특정 크기로 조절할 수 있다. 특정 크기는 기 정의된 크기일 수 있다. 예를 들어, 기 정의된 크기는 인코딩된 이미지( $Z_{\text{image}}$ )와 동일할 수 있다. 인코딩된 이미지( $Z_{\text{image}}$ )의 채널 수(예컨대, 3)와 조절된 마스크 맵( $M_r$ )의 채널 수(예컨대, 1)는 상이할 수 있다.
- [82] 전자 장치(1000)는 현재 노이즈 정보( $Z_t$ )를 획득할 수 있다. 현재 노이즈 정보( $Z_t$ )의 채널 수 및 크기는 인코딩된 이미지( $Z_{\text{image}}$ )의 채널 수 및 크기와 동일할 수 있다. 그러나, 본 개시는 이에 한정되지 않으며, 현재 노이즈 정보( $Z_t$ )의 채널 수 및 크기와 인코딩된 이미지( $Z_{\text{image}}$ )의 채널 수 및 크기는 상이할 수도 있다. 여기서,  $t$ 는 제2 생성 모델(1200)의 현재 디노이징 차수로 정의된다. 본 개시에서, 현재 디노이징 차수는 제2 생성 모델(1200)이 입력과 출력을 반복할 횟수를 나타낼 수 있다. 예를 들어,  $t$ 는 0 이상 디노이징 총 차수(예컨대,  $T$ ) 이하의 정수로 표현될 수 있다. 예시적으로,  $Z_t$ 는 처음으로 제2 생성 모델(1200)에 입력되는 초기 노이즈 정보를 나타낼 수 있다. 초기 노이즈 정보는 무작위 값들로 구성될 수 있다. 예를 들어, 초기 노이즈 정보는 가우시안 분포를 따르는 가우시안 노이즈로 구성될 수 있으나, 본 개시는 이에 한정되지 않는다.
- [83] 전자 장치(1000)는 인코딩된 이미지( $Z_{\text{image}}$ ), 조절된 마스크 맵( $M_r$ ), 및 현재 노이즈 정보( $Z_t$ )를 연결할 수 있다. 전자 장치(1000)는 인코딩된 이미지( $Z_{\text{image}}$ ), 조절된 마스크 맵( $M_r$ ), 및 현재 노이즈 정보( $Z_t$ )가 연결된 데이터(이하에서, 입력 데이터)를 제2 생성 모델(1200)에 전달할 수 있다. 제2 생성 모델(1200)은 입력 데이터에 기초하여 디노이징 동작을 수행할 수 있다. 본 개시에서, 디노이징 동작은 입력 노이즈로부터 소정의 노이즈를 제거하는 동작을 나타낼 수 있다. 제2 생성 모델

(1200)은 디노이징 동작을 수행함으로써 다음 노이즈 정보( $Z_{t-1}$ )를 생성할 수 있다.

- [84] 전자 장치(1000)는 디노이징 총 차수만큼 디노이징 동작을 수행했는지를 판단할 수 있다. 예를 들어, 전자 장치(1000)는 다음 노이즈 정보( $Z_{t-1}$ )가 최종 노이즈 정보( $Z_0$ )인지를 판단할 수 있다. 디노이징 총 차수만큼 디노이징 동작을 수행한 것으로 판단한 경우, 전자 장치(1000)는 최종 노이즈 정보( $Z_0$ )를 디코더(1400)에 전달할 수 있다. 디코더(1400)는 최종 노이즈 정보( $Z_0$ )에 기초하여 최종 생성 이미지를 생성할 수 있다. 디노이징 총 차수만큼 디노이징 동작을 수행하지 않은 것으로 판단한 경우, 전자 장치(1000)는 인코딩된 이미지( $Z_{image}$ ), 조절된 마스크 맵( $M_r$ ), 및 다음 노이즈 정보( $Z_{t-1}$ )를 연결할 수 있다. 전자 장치(1000)는 연결된 인코딩된 이미지( $Z_{image}$ ), 조절된 마스크 맵( $M_r$ ), 및 다음 노이즈 정보( $Z_{t-1}$ )를 제2 생성 모델(1200)에 입력함으로써 디노이징 동작을 반복할 수 있다.
- [85] 본 개시의 일 실시예에 있어서, 전자 장치(1000)에서 제1 인코더(1310) 및 디코더(1400)가 생략될 수 있다. 이 경우, 전자 장치(1000)는 이미지, 마스크 맵, 및 현재 노이즈 정보( $Z_t$ )를 연결할 수 있다. 전자 장치(1000)는 연결된 이미지, 마스크 맵, 및 현재 노이즈 정보( $Z_t$ )를 제2 생성 모델(1200)에 입력함으로써 디노이징 동작을 반복할 수 있다. 제2 생성 모델(1200)은 최종 노이즈 정보( $Z_0$ )를 출력할 수 있다. 최종 노이즈 정보( $Z_0$ )는 최종 생성 이미지를 포함할 수 있다. 예를 들어, 전자 장치(1000)의 하나 이상의 프로세서는 이미지, 마스크 맵, 및 현재 노이즈 정보( $Z_t$ )를 결합하여 제2 생성 모델(1200)에 입력함으로써 디노이징 동작을 반복적으로 수행할 수 있다. 제2 생성 모델(1200)은 최종 노이즈 정보( $Z_0$ )를 출력할 수 있다. 최종 노이즈 정보( $Z_0$ )는 최종 생성 이미지를 포함할 수 있다.
- [86] 도 5a는 본 개시의 일 실시예에 따른 제2 생성 모델에 입력되는 텍스트 가이던스를 설명하기 위한 개념도이다. 도 1 내지 4에서 설명한 내용과 중복되는 내용은 생략한다.
- [87] 도 5a를 참조하면, 전자 장치(1000)는 제2 인코더(1320)를 포함할 수 있다. 전자 장치(1000)는 텍스트 입력을 획득할 수 있다. 텍스트 입력은 이미지 및/또는 최종 생성 이미지에 대응하는 텍스트를 나타낼 수 있다. 예시적으로, 텍스트 입력은 "A table with a tumbler on in a cafeteria."와 같은 문장으로 표현될 수 있다. 텍스트 입력이 영문 문장인 것으로 설명하였으나, 본 개시는 이에 한정되지 않으며 텍스트 입력은 임의의 언어로 표현될 수 있다.
- [88] 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 외부 서버로부터 텍스트 입력을 획득할 수 있다. 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 사용자 인터페이스를 통해 사용자로부터 텍스트 입력을 획득할 수 있다. 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 사운드-텍스트 변환기(sound-to-text converter)를 포함할 수 있다. 사운드-텍스트 변환기는 음성-텍스트 변환기(speech-to-text converter)를 포함할 수 있다. 전자 장치(1000)는 사용자 인터페이스

스를 통해 사용자로부터 사용자 음성 입력을 획득할 수 있다. 전자 장치(1000)는 사운드-텍스트 변환기를 이용하여 사용자 음성 입력을 텍스트 입력으로 변환할 수 있다. 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 이미지-텍스트 변환기를 포함할 수 있다. 전자 장치(1000)는 이미지-텍스트 변환기에 이미지를 입력함으로써, 이미지를 표현하는 텍스트 입력을 획득할 수 있다.

- [89] 제2 인코더(1320)는 텍스트 입력을 인코딩할 수 있다. 본 개시의 일 실시예에 있어서, 제2 인코더(1320)는 텍스트 입력을 인코딩하도록 기 학습된 인공지능 모델일 수 있다. 제2 인코더(1320)는 인코딩된 텍스트 입력을 제2 생성 모델(1200)에 전달할 수 있다. 인코딩된 텍스트 입력은 텍스트 가이드런스, 텍스트 임베딩 또는 가이드런스 정보로도 지칭될 수 있다. 전자 장치(1000)는 제2 생성 모델(1200)의 적어도 하나의 레이어에 인코딩된 텍스트 입력을 전달할 수 있다. 제2 생성 모델(1200)은 이미지, 중간 생성 이미지 및 인코딩된 텍스트 입력에 기초하여 최종 생성 이미지를 출력할 수 있다.
- [90] 도 5b는 본 개시의 일 실시예에 따른 제2 생성 모델에 입력되는 이미지 가이드런스를 설명하기 위한 개념도이다. 도 1 내지 5a에서 설명한 내용과 중복되는 내용은 생략한다.
- [91] 도 5b를 참조하면, 전자 장치(1000)는 제3 인코더(1330)를 포함할 수 있다. 전자 장치(1000)는 이미지를 제3 인코더(1330)에 전달할 수 있다. 제3 인코더(1330)는 이미지를 인코딩할 수 있다. 본 개시의 일 실시예에 있어서, 제3 인코더(1330)는 이미지를 인코딩하도록 기 학습된 인공지능 모델일 수 있다. 제3 인코더(1330)는 인코딩된 이미지를 제2 생성 모델(1200)에 전달할 수 있다. 인코딩된 이미지는 이미지 가이드런스, 이미지 임베딩 또는 가이드런스 정보로도 지칭될 수 있다. 전자 장치(1000)는 제2 생성 모델(1200)의 적어도 하나의 레이어에 인코딩된 이미지를 전달할 수 있다. 제2 생성 모델(1200)은 이미지, 중간 생성 이미지 및 인코딩된 이미지에 기초하여 최종 생성 이미지를 출력할 수 있다.
- [92] 본 개시의 일 실시예에 있어서, 제3 인코더(1330)는 후술하는 도 11, 12A, 12B, 13A 및 13b의 제5 인코더(1350)와 동일한 방식으로 기 학습될 수 있다.
- [93] 도 5c는 본 개시의 일 실시예에 따른 제2 생성 모델에 입력되는 텍스트 가이드런스 및 이미지 가이드런스를 설명하기 위한 개념도이다. 도 1 내지 5b에서 설명한 내용과 중복되는 내용은 생략한다.
- [94] 도 5c를 참조하면, 전자 장치(1000)는 제2 인코더(1320) 및 제3 인코더(1330)를 포함할 수 있다. 제2 인코더(1320)의 기능 및 동작은 도 5a의 제2 인코더(1320)의 기능 및 동작에 대응되므로 중복되는 내용은 생략한다. 제3 인코더(1330)의 기능 및 동작은 도 5b의 제3 인코더(1330)의 기능 및 동작에 대응되므로 중복되는 내용은 생략한다. 제2 인코더(1320)는 텍스트 입력에 기초하여 텍스트 가이드런스 정보를 출력할 수 있다. 제3 인코더(1330)는 이미지에 기초하여 이미지 가이드런스 정보를 출력할 수 있다.

- [95] 제2 생성 모델(1200)은 제1 뉴럴 네트워크(1210) 및 제2 뉴럴 네트워크(1220), 및 노이즈 블렌더(1230)를 포함할 수 있다. 제1 뉴럴 네트워크(1210)는 이미지, 중간 생성 이미지, 및 이미지 가이드 정보에 기초하여 제1 노이즈(N1)를 출력할 수 있다. 제2 뉴럴 네트워크(1220)는 이미지, 중간 생성 이미지, 및 텍스트 가이드 정보에 기초하여 제2 노이즈(N2)를 출력할 수 있다. 노이즈 블렌더(1230)는 제1 노이즈 및 제2 노이즈를 블렌딩할 수 있다. 노이즈 블렌더(1230)는 제1 노이즈 및 제2 노이즈를 가중합함으로써 블렌딩된 노이즈를 출력할 수 있다. 제2 생성 모델(1200)은 블렌딩된 노이즈에 기초하여 최종 생성 이미지를 출력할 수 있다.
- [96] 본 개시의 일 실시예에 있어서, 제1 노이즈와 제2 노이즈 간의 가중치는 사용자 또는 제조사의 설정에 따라 상이할 수 있다. 예를 들어, 전자 장치(1000)는 사용자 인터페이스를 통해 사용자로부터 가중치를 획득할 수 있다.
- [97] 도 5d는 본 개시의 일 실시예에 따른 제1 생성 모델의 출력이 이미지 가이드 정보로 이용되는 실시예를 설명하기 위한 개념도이다. 도 1 내지 5c에서 설명한 내용과 중복되는 내용은 생략한다.
- [98] 도 5d를 참조하면, 전자 장치(1000)는 제4 인코더(1340)를 포함할 수 있다. 전자 장치(1000)는 중간 생성 이미지를 제4 인코더(1340)에 전달할 수 있다. 제4 인코더(1340)는 중간 생성 이미지를 인코딩할 수 있다. 본 개시의 일 실시예에 있어서, 제4 인코더(1340)는 이미지를 인코딩하도록 학습된 인공지능 모델일 수 있다. 본 개시의 일 실시예에 있어서, 제4 인코더(1340)의 구성, 동작, 및 기능은 제3 인코더(1330)의 구성, 동작, 및 기능에 대응할 수도 있다.
- [99] 제4 인코더(1340)는 인코딩된 중간 생성 이미지를 제2 생성 모델(1200)에 전달할 수 있다. 인코딩된 중간 생성 이미지는 이미지 가이드 정보, 이미지 임베딩 또는 가이드 정보로도 지칭될 수 있다. 전자 장치(1000)는 제2 생성 모델(1200)의 적어도 하나의 레이어에 인코딩된 중간 생성 이미지를 전달할 수 있다. 제2 생성 모델(1200)은 이미지 및 인코딩된 중간 생성 이미지에 기초하여 최종 생성 이미지를 출력할 수 있다.
- [100] 본 개시의 일 실시예에 따르면, 알려진 영역(또는 기지 영역으로도 지칭될 수도 있음)의 이미지 정보를 비교적 많이 참조하는 성질을 가진 제1 생성 모델의 출력을 제2 생성 모델의 가이드 정보로 활용함으로써, 이미지 맥락에 맞는 적절한 이미지를 출력하는 제2 생성 모델을 제공할 수 있다.
- [101] 도 5e는 본 개시의 일 실시예에 따른 제1 생성 모델의 출력이 이미지 가이드 정보로 이용되는 실시예를 설명하기 위한 개념도이다. 도 1 내지 5c에서 설명한 내용과 중복되는 내용은 생략한다.
- [102] 도 5e를 참조하면, 전자 장치(1000)는 제4 인코더(1340)를 포함할 수 있다. 제4 인코더(1340)의 기능 및 동작은 도 5d의 제4 인코더(1340)의 기능 및 동작에 대응되므로 중복되는 내용은 생략한다. 제4 인코더(1340)는 중간 생성 이미지에 기초하여 이미지 가이드 정보를 출력할 수 있다.

- [103] 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 제1 생성 모델(1100)의 출력인 중간 생성 이미지를 제2 생성 모델(1200)에 전달할 수 있다. 제2 생성 모델(1200)은 중간 생성 이미지와 이미지 가이드스 정보에 기초하여 최종 생성 이미지를 추론할 수 있다. 제2 생성 모델(1200)이 중간 생성 이미지에 기초하여 최종 생성 이미지를 추론하는 실시예에 대한 설명은 도 1 내지 5c에서 설명한 내용과 중복되므로 생략한다. 제2 생성 모델(1200)이 이미지 가이드스 정보에 기초하여 최종 생성 이미지를 추론하는 실시예에 대한 설명은 도 5d에서 설명한 내용과 중복되므로 생략한다.
- [104] 본 개시의 일 실시예에 있어서, 도 5e에 도시된 바와 다르게, 제2 생성 모델(1200)은 인코딩된 텍스트 입력을 전달받을 수 있다. 예를 들어, 전자 장치(1000)는 텍스트 입력을 수신하고, 도 5a의 제2 인코더(1320)와 같은 인코더를 이용하여 텍스트 입력을 인코딩할 수 있다. 제2 생성 모델(1200)은 인코딩된 텍스트 입력(텍스트 가이드스), 이미지, 중간 생성 이미지, 인코딩된 중간 생성 이미지(이미지 가이드스)를 입력으로 하여 최종 생성 이미지를 출력할 수 있다.
- [105] 본 개시의 일 실시예에 따르면, 제2 생성 모델(1200)은 이미지 가이드스를 참조할 뿐만 아니라, 중간 생성 이미지에 기초하여 최종 생성 이미지를 추론함으로써 인페인팅 및/또는 아웃페인팅의 성능이 향상될 수 있다.
- [106] 도 5f는 본 개시의 일 실시예에 따른 텍스트 가이드스와 함께, 제1 생성 모델의 출력이 이미지 가이드스로 이용되는 실시예를 설명하기 위한 개념도이다. 도 1 내지 5d에서 설명한 내용과 중복되는 내용은 생략한다.
- [107] 도 5e를 참조하면, 전자 장치(1000)는 제2 인코더(1320), 및 제4 인코더(1340)를 포함할 수 있다. 제2 인코더(1320)의 기능 및 동작은 도 5c의 제2 인코더(1320)의 기능 및 동작에 대응되므로 중복되는 내용은 생략한다. 제4 인코더(1340)의 기능 및 동작은 도 5d의 제4 인코더(1340)의 기능 및 동작에 대응되므로 중복되는 내용은 생략한다. 제2 인코더(1320)는 텍스트 입력에 기초하여 텍스트 가이드스 정보를 출력할 수 있다. 제4 인코더(1340)는 중간 생성 이미지에 기초하여 이미지 가이드스 정보를 출력할 수 있다.
- [108] 제2 생성 모델(1200)은 제1 뉴럴 네트워크(1210), 제2 뉴럴 네트워크(1220), 및 노이즈 블렌더(1230)를 포함할 수 있다. 제1 뉴럴 네트워크(1210), 제2 뉴럴 네트워크(1220), 및 노이즈 블렌더(1230)의 구성, 기능, 및 동작은 도 5c의 제1 뉴럴 네트워크(1210), 제2 뉴럴 네트워크(1220), 및 노이즈 블렌더(1230)의 구성, 기능, 및 동작에 대응되므로 중복되는 내용은 생략한다. 전자 장치(1000)는 제4 인코더(1340)로부터 제1 뉴럴 네트워크(1210)로 이미지 가이드스 정보를 전달할 수 있다. 전자 장치(1000)는 제2 인코더(1340)로부터 제2 뉴럴 네트워크(1220)로 텍스트 가이드스 정보를 전달할 수 있다.
- [109] 본 개시의 일 실시예에 따르면, 이미지 가이드스 뿐만 아니라, 텍스트 가이드스를 참조하여 최종 생성 이미지를 추론함으로써 인페인팅 및/또는 아웃페인팅의 성능이 향상될 수 있다.

- [110] 도 5g는 본 개시의 일 실시예에 따른 텍스트 가이드스와 함께, 제1 생성 모델의 출력이 이미지 가이드스로 이용되는 실시예를 설명하기 위한 개념도이다.
- [111] 도 5g를 참조하면, 전자 장치(1000)는 제2 인코더(1320), 및 제4 인코더(1340)를 포함할 수 있다. 제2 인코더(1320) 및 제4 인코더(1340)의 기능 및 동작은 도 5f의 제2 인코더(1320) 및 제4 인코더(1340)의 기능 및 동작에 대응되므로 중복되는 내용은 생략한다. 제2 인코더(1320)는 텍스트 입력에 기초하여 텍스트 가이드스 정보를 출력할 수 있다. 제4 인코더(1340)는 중간 생성 이미지에 기초하여 이미지 가이드스 정보를 출력할 수 있다.
- [112] 제2 생성 모델(1200)은 제1 뉴럴 네트워크(1210), 제2 뉴럴 네트워크(1220), 및 노이즈 블렌더(1230)를 포함할 수 있다. 제1 뉴럴 네트워크(1210), 제2 뉴럴 네트워크(1220), 및 노이즈 블렌더(1230)의 구성, 기능, 및 동작은 도 5c 및 도 5f의 제1 뉴럴 네트워크(1210), 제2 뉴럴 네트워크(1220), 및 노이즈 블렌더(1230)의 구성, 기능, 및 동작에 대응되므로 중복되는 내용은 생략한다. 전자 장치(1000)는 제4 인코더(1340)로부터 제1 뉴럴 네트워크(1210)로 이미지 가이드스 정보를 전달할 수 있다. 전자 장치(1000)는 제2 인코더(1340)로부터 제2 뉴럴 네트워크(1220)로 텍스트 가이드스 정보를 전달할 수 있다.
- [113] 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 제1 생성 모델(1100)의 출력인 중간 생성 이미지를 제2 생성 모델(1200)에 전달할 수 있다. 제2 생성 모델(1200)은 중간 생성 이미지, 이미지 가이드스 정보, 및 텍스트 가이드스 정보에 기초하여 최종 생성 이미지를 추론할 수 있다. 제2 생성 모델(1200)이 중간 생성 이미지에 기초하여 최종 생성 이미지를 추론하는 실시예에 대한 설명은 도 1 내지 5c, 및 5e에서 설명한 내용과 중복되므로 생략한다. 제2 생성 모델(1200)이 이미지 가이드스 정보 및 텍스트 가이드스 정보에 기초하여 최종 생성 이미지를 추론하는 실시예에 대한 설명은 도 5c 및 5f에서 설명한 내용과 중복되므로 생략한다.
- [114] 본 개시의 일 실시예에 따르면, 제2 생성 모델(1200)은 이미지 가이드스 및/또는 텍스트 가이드스를 참조할 뿐만 아니라, 중간 생성 이미지에 기초하여 최종 생성 이미지를 추론함으로써 인페인팅 및/또는 아웃페인팅의 성능이 향상될 수 있다.
- [115] 도 5a 내지 5g에서는 이미지가 제1 생성 모델(1100) 및 제2 생성 모델(1200)에 입력되는 것으로 도시되었지만, 도시된 바와 다르게, 이미지에 마스크 맵이 연결되어 제1 생성 모델(1100) 및 제2 생성 모델(1200)에 입력될 수 있다.
- [116] 도 5a 내지 5g에서는 도 2 및 도 4의 제1 인코더(1310) 및 디코더(1400)가 생략된 형태로 도시되었지만, 도시된 바와 다르게, 제2 생성 모델(1200)의 입력인 이미지 및/또는 마스크 맵이 제1 인코더(1310)에 의해 인코딩되고, 제2 생성 모델(1200)의 최종 출력은 디코더(1400)에 의해 디코딩될 수 있다.
- [117] 도 6은 본 개시의 일 실시예에 따른 제1 생성 모델의 학습 방법을 설명하기 위한 개념도이다. 도 1 내지 5g에서 설명한 내용과 중복되는 내용은 생략한다.
- [118] 도 6을 참조하면, 모델 학습 시스템(10)은 제1 생성 모델(1100), 판별기(discriminator)(12), 및 손실 함수(loss function)(13)를 포함할 수 있다. 예를 들어,

모델 학습 시스템(10)은 적대적 생성 신경망 모델을 학습하는 시스템일 수 있다. 일부 실시예들에서는 도 6에 도시된 바와 같이, 제1 생성 모델(1100)은 생성기(generator)(1100)로 지칭될 수도 있다. 본 개시에서, 적대적 생성 신경망 모델은 학습을 통해 생성기(1100)와 판별기(12)가 서로의 성능을 개선해 적대적으로 경쟁해 나가는 모델을 나타낼 수 있다. 생성기(1100) 및 판별기(12) 각각은 적어도 하나의 레이어를 포함할 수 있다. 레이어는 입력 데이터로부터 특징을 추출하기 위한 가중치 값들로 구성된 필터를 포함할 수 있다.

- [119] 생성기(1100)는 데이터 셋(DS)을 입력으로 하여, 가짜 데이터(FD)를 출력하도록 학습될 수 있다. 데이터 셋(DS)은 복수의 이미지들을 포함하는 데이터의 집합일 수 있다. 데이터 셋(DS)은 복수의 이미지들 각각에 대한 마스크 맵을 포함할 수 있다. 가짜 데이터(FD)는 가짜 이미지 데이터를 나타낼 수 있다. 실제 데이터 DB(11)는 실제 데이터(RD)의 집합을 포함할 수 있다.
- [120] 판별기(12)는 가짜 데이터(FD) 또는 실제 데이터(RD)를 입력으로 하여, 가짜 데이터(FD) 또는 실제 데이터(RD)의 가짜 여부를 판별하도록 학습될 수 있다.
- [121] 손실 함수(13)는 판별 결과(DR)에 기초하여 손실 함수 값을 계산할 수 있다. 손실 함수 값은 역전파를 통해 판별기(12) 및 생성기(1100)에 전달될 수 있다. 판별기(12) 및 생성기(1100)에 포함되는 적어도 하나의 레이어의 가중치는 손실 함수 값에 기초하여 업데이트될 수 있다.
- [122] 본 개시의 일 실시예에 있어서, 모델 학습 시스템(10)의 기능의 적어도 일부는 도 1 내지 5g에서 설명한 전자 장치(1000)에서 수행될 수 있으나, 본 개시는 이에 한정되지 않으며, 전자 장치(1000)가 아닌 외부 서버 장치에서 수행될 수도 있다.
- [123] 본 개시의 일 실시예에 있어서, 모델 학습 시스템(10)을 통해 학습된 생성기(1100)는 도 1 내지 5g에서 설명한 제1 생성 모델(1100)에 대응할 수 있다.
- [124] 도 7a 및 7b는 본 개시의 일 실시예에 따른 제2 생성 모델의 구성을 설명하기 위한 개념도이다. 도 1 내지 6에서 설명한 내용과 중복되는 내용은 생략한다.
- [125] 도 7a를 참조하면, 제2 생성 모델(1200)은 제1 뉴럴 네트워크(1210)를 포함할 수 있다. 제1 뉴럴 네트워크(1210)는 적어도 하나의 레이어를 포함할 수 있다. 예를 들어, 적어도 하나의 레이어는 크로스 어텐션(cross attention)을 수행할 수 있다. 그러나 본 개시는 이에 한정되지 않으며, 적어도 하나의 레이어는 셀프 어텐션(self-attention) 레이어 또는 레지듀얼 블록(residual block) 등을 포함할 수 있다. 적어도 하나의 레이어 중 크로스 어텐션을 수행하는 레이어는 크로스 어텐션 레이어로 지칭될 수도 있다.
- [126] 전자 장치(1000)는 제1 뉴럴 네트워크(1210)의 크로스 어텐션 레이어에 가이드된 정보를 전달할 수 있다. 제1 뉴럴 네트워크(1210)는 입력 이미지와 가이드된 정보 간의 상관 관계에 기초하여 가이드된 정보에 가중치를 반영할 수 있다.
- [127] 본 개시의 일 실시예에 있어서, 제1 뉴럴 네트워크(1210)는 쿼리(query), 키(key), 및 밸류(value)를 피연산자로 하는 크로스 어텐션(cross-attention) 연산을 수행할 수 있다. 예를 들어, 쿼리는 현재 노이즈 정보를 포함할 수 있고, 키와 밸류는 가

이던스 정보(guidance information)를 포함할 수 있다. 그러나 본 개시는 이에 국한되지 않으며, 쿼리, 키, 밸류가 다른 정보를 포함할 수도 있다. 크로스 어텐션 연산은 크로스 어텐션 레이어에서 수행될 수 있다. 제1 뉴럴 네트워크(1210)는 크로스 어텐션 연산 결과를 다음 레이어에 전달할 수 있다.

- [128] 본 개시의 일 실시예에 있어서, 도 7a에 도시된 바와 달리, 제1 뉴럴 네트워크(1210)는 셀프 어텐션 레이어를 포함할 수 있다. 쿼리, 키, 및 밸류는 현재 노이즈 정보를 포함할 수 있다. 제1 뉴럴 네트워크(1210)는 쿼리, 키, 및 밸류를 피연산자로 하는 셀프 어텐션(self attention) 연산을 수행할 수 있다. 셀프 어텐션 연산은 셀프 어텐션 레이어에서 수행될 수 있다. 제1 뉴럴 네트워크(1210)는 셀프 어텐션 연산 결과를 다음 레이어에 전달할 수 있다.
- [129] 본 개시의 일 실시예에 있어서, 제2 생성 모델(1200)은 인터프리터(1240)를 포함할 수 있다. 예를 들어, 제2 생성 모델(1200)은 텍스트로부터 획득된 텍스트가 이던스 정보를 크로스 어텐션 레이어에 입력하는 방식으로 기 학습된 모델일 수 있다. 인터프리터(1240)는 텍스트가 아닌 임의의 데이터 형태로부터 획득된 가이던스 정보가 크로스 어텐션 레이어에 입력되더라도 제2 생성 모델(1200)의 성능이 유지되도록, 가이던스 정보를 텍스트 가이던스 정보의 속성 및/또는 형태를 갖도록 변환할 수 있다. 예를 들어, 인터프리터(1240)는 이미지 가이던스 정보를 텍스트 가이던스 정보의 속성 및/또는 형태를 갖도록 변환할 수 있다.
- [130] 본 개시의 일 실시예에 있어서, 인터프리터(1240)는 도 5b 내지 5g에 도시된 제3 인코더(1330) 및 제4 인코더(1340)가 출력한 이미지 가이던스 정보를 텍스트 가이던스 정보의 속성 및/또는 형태를 갖도록 변환할 수 있다.
- [131] 본 개시의 일 실시예에 따르면, 학습 비용이 높은 제2 생성 모델을 새로 학습하거나 미세 조정하지 않고 인터프리터 구성만을 추가함으로써 다양한 데이터를 가이던스로 활용할 수 있다.
- [132] 본 개시의 일 실시예에 있어서, 인터프리터(1240)는 생략될 수도 있다. 이 경우, 제2 생성 모델(1200)은 임의의 가이던스 정보(예컨대, 이미지 가이던스 정보)가 크로스 어텐션 레이어에 입력되는 과정을 통해 기 학습될 수 있다.
- [133] 도 7b를 참조하면, 기 학습된 제1 뉴럴 네트워크(1210)를 미세 조정(fine-tuning)함으로써 태스크의 성능을 높일 수 있다. 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 기 학습된 제1 뉴럴 네트워크(1210)의 파라미터들(예컨대, 가중치)을 모두 고정된 상태로, 추가 학습 데이터 입력에 따른 기 정의된 레이어들 중 적어도 일부의 레이어의 가중치 변화량( $\Delta W$ )을 계산할 수 있다.
- [134] 전자 장치(1000)는 고정된 파라미터들에 대응하는 기 정의된 레이어들 중 적어도 일부의 레이어의 고정된 가중치에 가중치 변화량( $\Delta W$ )을 합산함으로써 새로운 가중치를 결정할 수 있다. 본 개시의 일 실시예에 따르면, 제1 뉴럴 네트워크(1210)가 미세 조정됨으로써, 추가 학습 데이터 입력에 대응하는 태스크의 성능이 향상될 수 있다.

- [135] 도 8은 본 개시의 일 실시예에 따른 인터프리터의 구성을 설명하기 위한 개념도이다. 도 1 내지 7b에서 설명한 내용과 중복되는 내용은 생략한다.
- [136] 도 8을 참조하면, 인터프리터(1240)는 적어도 하나의 단층 퍼셉트론(perceptron)을 포함할 수 있다. 본 개시의 일 실시예에 있어서, 인터프리터(1240)는 제1 단층 퍼셉트론(1241) 및 제2 단층 퍼셉트론(1242)을 포함할 수 있다. 인터프리터(1240)는 제1 단층 퍼셉트론(1241) 및 제2 단층 퍼셉트론(1242)을 포함하는 다층 퍼셉트론으로 구성될 수 있다. 본 개시의 일 실시예에 있어서, 제1 단층 퍼셉트론(1241) 및 제2 단층 퍼셉트론(1242)은 서로 연결되어 학습될 수 있다.
- [137] 제1 단층 퍼셉트론(1241)은 제1 임베딩을 입력으로 하여 제2 임베딩을 출력할 수 있다. 제1 임베딩은 가이드نس 정보 또는 대상 이미지 임베딩으로도 지칭될 수 있다. 제2 임베딩은 중간 이미지 임베딩으로도 지칭될 수 있다. 예를 들어, 제1 임베딩은  $k \times m$ 의 크기를 갖는 행렬일 수 있다. ( $k$ 와  $m$ 은 자연수) 제1 단층 퍼셉트론(1241)은 제1 레이어와 제2 레이어를 포함할 수 있다. 제1 레이어는  $m$  개의 노드들을 포함할 수 있다. 제2 레이어는  $n$  개의 노드들을 포함할 수 있다. ( $n$ 은 자연수) 본 개시의 일 실시예에 있어서,  $m$ 은  $n$ 보다 크거나 같은 자연수일 수 있으나, 본 개시는 이에 한정되지 않으며,  $m$ 은  $n$ 보다 작거나 같을 수도 있다. 예를 들어, 제2 임베딩은  $k \times n$ 의 크기를 갖는 행렬일 수 있다.
- [138] 제2 단층 퍼셉트론(1242)은 제2 임베딩을 입력으로 하여 제3 임베딩을 출력할 수 있다. 제3 임베딩은 결과 이미지 임베딩으로도 지칭될 수 있다. 제2 단층 퍼셉트론(1242)은 제3 레이어와 제4 레이어를 포함할 수 있다. 제3 레이어는  $k$  개의 노드들을 포함할 수 있다. 제4 레이어는 1 개의 노드들을 포함할 수 있다. ( $l$ 은 자연수) 본 개시의 일 실시예에 있어서,  $k$ 는 1보다 작거나 같을 수 있으나, 본 개시는 이에 한정되지 않으며,  $k$ 는 1보다 크거나 같을 수도 있다. 예를 들어, 제3 임베딩은  $1 \times n$ 의 크기를 갖는 행렬일 수 있다. 제3 임베딩은 제2 생성 모델(7a, 7b, 1200) (또는 제1 뉴럴 네트워크(도 7a, 7b, 1210))의 크로스 어텐션 레이어와 같은 적어도 하나의 레이어에 입력될 수 있다.
- [139] 본 개시의 일 실시예에 따르면, 인터프리터(1240)는 2개 이상의 단층 퍼셉트론을 포함할 수 있다. 도 8에서 두 개의 단층 퍼셉트론(제1 단층 퍼셉트론(1241), 제2 단층 퍼셉트론(1242))만을 도시하였으나, 본 개시는 이에 한정되지 않으며, 인터프리터(1240)는 3개 이상의 단층 퍼셉트론을 포함할 수 있다. 이전 단층 퍼셉트론이 출력한 임베딩을 입력으로 하여 다음 단층 퍼셉트론이 다음 임베딩을 출력할 수 있다. 이 경우, 마지막 단층 퍼셉트론이 출력한 임베딩은 결과 이미지 임베딩으로 지칭될 수 있다. 결과 이미지 임베딩은 제2 생성 모델(7a, 7b, 1200)(또는 제1 뉴럴 네트워크(도 7a, 7b, 1210))의 크로스 어텐션 레이어와 같은 적어도 하나의 레이어에 입력될 수 있다.
- [140] 본 개시의 일 실시예에 따르면, 도 8에 도시된 바와 달리, 인터프리터(1240)는 유일한 단층 퍼셉트론을 포함할 수 있다. 이 경우, 유일한 단층 퍼셉트론이 출력한 임베딩은 결과 이미지 임베딩으로 지칭될 수 있다. 결과 이미지 임베딩은 제2

생성 모델(7a, 7b, 1200)(또는 제1 뉴럴 네트워크(도 7a, 7b, 1210))의 크로스 어텐션 레이어와 같은 적어도 하나의 레이어에 입력될 수 있다.

- [141] 본 개시의 일 실시예에 따르면, 인터프리터(1240)는 이미지 가이드의 차원을 텍스트 가이드의 차원으로 변환하는 기능을 수행할 수 있다. 본 개시의 일 실시예에 따르면, 인터프리터(1240)는 이미지 가이드와 텍스트 가이드 간의 성질 차이를 해석하거나 전환하는 기능을 수행할 수 있다. 본 개시의 일 실시예에 따르면, 인터프리터(1240)가 상술한 기능들을 수행함으로써, 텍스트 가이드만 이해하도록 학습된 제2 생성 모델이 이미지 가이드와 같은 다양한 가이드 정보 이해할 수 있게 된다.
- [142] 도 9a 및 9b는 본 개시의 일 실시예에 따른 중간 생성 이미지에 노이즈를 추가하는 실시예를 설명하기 위한 개념도이다. 도 1 내지 8에서 설명한 내용과 중복되는 내용은 생략한다.
- [143] 도 9a를 참조하면, 전자 장치(1000)는 디노이징 강도(denoising strength)를 획득할 수 있다. 디노이징 강도는 중간 생성 이미지에 추가될 노이즈 량에 대응할 수 있다. 본 개시의 일 실시예에 있어서, 디노이징 강도는 특정 값일 수 있다. 예를 들어, 특정 값은 사용자 또는 제조사의 설정에 따라 미리 설정된 값일 수 있다. 예를 들어, 디노이징 강도는 제1 생성 모델의 성능 지표를 고려하여 결정될 수 있다.
- [144] 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 노이즈 생성기(1500)를 포함할 수 있다. 노이즈 생성기(1500)는 디노이징 강도에 기초하여 중간 생성 이미지에 노이즈를 추가할 수 있다. 전자 장치(1000)는 디노이징 강도에 매핑된 디노이징 차수를 식별할 수 있다. 전자 장치(1000)는 제2 생성 모델(1200)의 현재 노이즈 정보의 디노이징 차수를 식별된 디노이징 차수로 설정할 수 있다. 노이즈가 추가된 중간 생성 이미지( $Z_n$ )는 제2 생성 모델의 식별된 디노이징 차수(예컨대,  $n$ )를 갖는 현재 노이즈 정보로 활용될 수 있다. 제2 생성 모델(1200)은 이미지(예컨대, 마스크된 이미지), 노이즈가 추가된 중간 생성 이미지( $Z_n$ ), 및 가이드 정보를 입력으로 하여 다음 노이즈 정보( $Z_{n+1}$ )를 획득할 수 있다.
- [145] 도 9b를 참조하면, 전자 장치(1000)는 제1 인코더(1310)를 이용하여 이미지를 인코딩할 수 있다. 제1 인코더(1310)는 인코딩된 이미지( $Z_{i2}$ )를 출력할 수 있다. 전자 장치(1000)는 마스크 맵을 기 정의된 크기로 조절할 수 있다. 전자 장치(1000)는 제1 인코더(1310)를 이용하여 중간 생성 이미지를 인코딩할 수 있다. 제1 인코더(1310)는 인코딩된 중간 생성 이미지( $Z_{i1}$ )를 출력할 수 있다. 노이즈 생성기(1500)는 디노이징 강도에 기초하여 인코딩된 중간 생성 이미지( $Z_{i1}$ )에 노이즈를 추가할 수 있다. 전자 장치(1000)는 디노이징 강도에 대응하는 디노이징 차수( $n$ )를 결정할 수 있다. 노이즈 생성기(1500)는 노이즈가 추가된 중간 생성 이미지( $Z_n$ )를 현재 노이즈 정보로 설정할 수 있다.

- [146] 전자 장치(1000)는 인코딩된 이미지( $Z_{i2}$ ), 조절된 마스크 맵( $M_r$ ), 및 노이즈가 추가된 중간 생성 이미지( $Z_n$ )를 연결할 수 있다. 인코딩된 이미지( $Z_{i2}$ ), 조절된 마스크 맵( $M_r$ ), 및 노이즈가 추가된 중간 생성 이미지( $Z_n$ )를 연결하는 순서는 임의로 정할 수 있으나, 제2 생성 모델(1200) 학습 시 확정된 순서와 디노이징(또는 추론) 시 순서는 동일해야 할 수 있다. 전자 장치(1000)는 인코딩된 이미지( $Z_{i2}$ ), 조절된 마스크 맵( $M_r$ ), 및 노이즈가 추가된 중간 생성 이미지( $Z_n$ )가 연결된 데이터(이하에서, 입력 데이터)를 제2 생성 모델(1200)에 입력할 수 있다. 제2 생성 모델(1200)은 입력 데이터에 기초하여 다음 노이즈 정보( $Z_{n-1}$ )를 출력할 수 있다. 전자 장치(1000)는 인코딩된 이미지( $Z_{i2}$ ), 조절된 마스크 맵( $M_r$ ), 및 다음 노이즈 정보( $Z_{n-1}$ )를 연결하고, 연결된 데이터를 제2 생성 모델(1200)에 입력할 수 있다. 전자 장치(1000)는 제2 생성 모델(1200)의 출력이 최종 노이즈 정보( $Z_0$ )가 될 때까지 디노이징 동작을 반복할 수 있다. 전자 장치(1000)는 최종 노이즈 정보를 디코더(1400)에 입력함으로써 최종 생성 이미지를 획득할 수 있다.
- [147] 본 개시의 일 실시예에 따르면, 제1 생성 모델은 제2 생성 모델 대비 연산 비용이 낮고 추론 속도가 빠르다. 본 개시의 일 실시예에 따르면, 제1 생성 모델의 출력 이미지에 노이즈를 추가한 데이터를 제2 생성 모델의 특정 디노이징 차수의 노이즈로 활용함으로써, 제2 생성 모델의 연산 비용을 절감하고 추론 속도를 향상시킬 수 있다.
- [148] 도 10a 내지 10c는 본 개시의 일 실시예에 따른 중간 생성 이미지의 퀄리티에 따라 디노이징 강도를 결정하는 실시예를 설명하기 위한 개념도이다. 도 1 내지 9b에서 설명한 내용과 중복되는 내용은 생략한다.
- [149] 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 디노이징 강도 결정기(1550)를 포함할 수 있다. 디노이징 강도 결정기(1550)는 중간 생성 이미지에 기초하여 중간 생성 이미지의 예측 신뢰 값을 획득할 수 있다. 예를 들어, 예측 신뢰 값은 제1 생성 모델(1100)이 입력된 이미지에 대한 컨피던스 스코어(confidence score) 값에 기초하여 결정될 수 있다. 그러나 본 개시는 이에 한정되지 않으며, 이미지의 퀄리티를 측정하는 임의의 기법으로도 예측 신뢰 값을 결정할 수 있다. 디노이징 강도 결정기(1550)는 예측 신뢰 값에 기초하여 디노이징 강도를 결정할 수 있다.
- [150] 본 개시의 일 실시예에 있어서, 디노이징 강도 결정기(1550)는 입력된 이미지에 기초하여 일부 영역(예컨대, 미지 영역 또는 마스크된 영역)의 크기 및/또는 모양을 식별할 수 있다. 디노이징 강도 결정기(1550)는 식별된 크기 및/또는 모양에 기초하여 디노이징 강도를 결정할 수 있다. 예를 들어, 미지 영역의 크기가 클수록, 제1 생성 모델(1100)의 예측 성능은 낮아질 수 있다. 디노이징 강도 결정기(1550)는 미지 영역의 크기가 클수록 디노이징 강도를 높일 수 있다. 예를 들어, 미지 영역의 모양이 특정 모양인 경우, 제1 생성 모델(1100)의 예측 성능은 낮아질 수 있다. 디노이징 강도 결정기(1550)는 식별된 모양에 따라 디노이징 강도를 상이하

게 결정할 수 있다. 본 개시의 일 실시예에 있어서, 미지 영역의 모양을 식별하기 위해, 영역의 모양을 분류하는 기 학습된 분류 모델을 이용할 수 있다.

- [151] 도 10b를 참조하면, 전자 장치(1000)는 제4 인코더(1340)를 포함할 수 있다. 제4 인코더(1340)의 구성, 기능, 및 동작은 도 5d 내지 5f의 제4 인코더(1340)의 구성, 기능, 및 동작에 대응하므로, 중복되는 내용은 생략한다. 제2 생성 모델(1200)은 제1 뉴럴 네트워크(1210), 제2 뉴럴 네트워크(1220), 및 노이즈 블렌더(1230)를 포함할 수 있다. 제1 뉴럴 네트워크(1210), 제2 뉴럴 네트워크(1220), 및 노이즈 블렌더(1230)의 구성, 기능, 및 동작은 도 5f 및 5g의 제1 뉴럴 네트워크(1210), 제2 뉴럴 네트워크(1220), 및 노이즈 블렌더(1230)의 구성, 기능, 및 동작에 대응하므로, 중복되는 내용은 생략한다. 제4 인코더(1340)는 중간 생성 이미지를 인코딩할 수 있다. 인코딩된 중간 생성 이미지는 이미지 가이드스 정보 또는 이미지 임베딩으로도 지칭될 수 있다. 전자 장치(1000)는 이미지 가이드스 정보를 제1 뉴럴 네트워크(1210)의 적어도 하나의 레이어에 입력할 수 있다. 전자 장치(1000)는 텍스트 가이드스 정보(또는 텍스트 임베딩으로도 지칭될 수 있음)를 제2 뉴럴 네트워크(1220)의 적어도 하나의 레이어에 입력할 수 있다.
- [152] 노이즈 생성기(1500)는 디노이징 강도에 기초하여 중간 생성 이미지에 노이즈를 추가할 수 있다. 노이즈가 추가된 중간 생성 이미지( $Z_n$ )는 제1 뉴럴 네트워크(1210) 및/또는 제2 뉴럴 네트워크(1220)의 식별된 디노이징 차수(예컨대,  $n$ )를 갖는 현재 노이즈 정보로 활용될 수 있다. 도 10b에서는 추가된 중간 생성 이미지( $Z_n$ )는 제1 뉴럴 네트워크(1210) 및 제2 뉴럴 네트워크(1220) 모두에 입력되는 것으로 도시되었으나, 본 개시는 이에 한정되지 않는다. 추가된 중간 생성 이미지( $Z_n$ )는 제1 뉴럴 네트워크(1210) 및 제2 뉴럴 네트워크(1220) 중 적어도 하나에 입력될 수 있다. 제1 뉴럴 네트워크(1210) 및/또는 제2 뉴럴 네트워크(1220)는 이미지(예컨대, 마스킹된 이미지), 노이즈가 추가된 중간 생성 이미지( $Z_n$ ), 및 가이드스 정보(예컨대, 이미지 가이드스 정보 및/또는 텍스트 가이드스 정보)를 입력으로 하여 다음 노이즈 정보(예컨대,  $Z_{n+1}$ )를 획득할 수 있다.
- [153] 도 10c를 참조하면, 디노이징 강도 결정기(1550)는 마스크 맵으로부터 이미지에서 마스킹된 영역의 크기 및/또는 모양을 식별할 수 있다. 디노이징 강도 결정기(1550)는 예측 신뢰 값, 일부 영역의 크기, 일부 영역의 모양 중 적어도 하나에 기초하여 디노이징 강도를 결정할 수 있다.
- [154] 도 11은 본 개시의 일 실시예에 따른 생성 모델을 이용하여 이미지의 일부 영역을 생성하는 전자 장치를 설명하기 위한 개념도이다. 도 1 내지 10c에서 설명한 내용과 중복되는 내용은 생략한다.
- [155] 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 제5 인코더(1350) 및 제2 생성 모델(1200)을 포함할 수 있다. 전자 장치(1000)는 기 설정된 일부 영역의 정보를 포함하는 이미지를 획득할 수 있다. 예를 들어, 일부 영역은 미지의 영역 또는

마스킹된 영역으로도 지칭될 수도 있다. 이미지는 일부 영역을 제외한 영역에 대한 색상 정보를 포함할 수 있다.

- [156] 본 개시의 일 실시예에 따르면, 이미지는 도 3에서 설명한 바와 같이 마스크 맵을 포함할 수 있다. 전자 장치(1000)는 이미지의 전체 영역에서 일부 영역 (또는 대상 영역)을 구분하는 마스크 맵을 획득할 수 있다. 이때, 일부 영역은 기 설정된 영역일 수 있다. 예를 들어, 전자 장치(1000)는 이미지에 기초하여 마스크 맵을 생성할 수 있다. 전자 장치(1000)는 이미지에 마스크 맵을 연결할 수 있다.
- [157] 전자 장치(1000)는 일부 영역의 정보를 포함하는 이미지를 제5 인코더(1350)에 전달할 수 있다. 제5 인코더(1350)는 이미지를 인코딩하도록 기 학습된 인공지능 모델일 수 있다. 인코딩된 이미지는 제2 생성 모델(1200)에 가이던스 정보로 활용될 수 있다. 제2 생성 모델(1200)은 일부 영역의 정보를 포함하는 이미지 및 제5 인코더(1350)에 의해 인코딩된 이미지를 입력으로 하여 최종 생성 이미지를 출력할 수 있다.
- [158] 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 제1 인코더(1310) 및 디코더(1400)를 포함할 수 있다. 전자 장치(1000)는 일부 영역의 정보를 포함하는 이미지를 제1 인코더(1310)를 이용하여 인코딩할 수 있다. 전자 장치(1000)는 제2 생성 모델의 출력을 디코더(1400)를 이용하여 디코딩할 수 있다.
- [159] 도 12a 내지 12b는 본 개시의 일 실시예에 따른 제5 인코더의 학습 방법을 설명하기 위한 개념도이다. 도 13a 내지 13b는 본 개시의 일 실시예에 따른 제5 인코더의 학습 방법을 설명하기 위한 개념도이다. 도 1 내지 11에서 설명한 내용과 중복되는 내용은 생략한다.
- [160] 도 12a 및 12b를 참조하면, 인코더 학습 시스템(20)은 제5 인코더(1350), 제6 인코더(21), 및 제7 인코더(22)를 포함할 수 있다. 제5 인코더(1350)는 학습이 완료되지 않은 인공지능 모델일 수 있다. 인코더 학습 시스템(20)은 제5 인코더(1350)를 학습시킬 수 있다. 본 개시의 일 실시예에 있어서, 제5 인코더(1350)는 기 학습된 인공지능 모델일 수 있다. 인코더 학습 시스템(20)은 미세 조정과 같은 추가 학습 기법을 통해 제5 인코더(1350)를 학습시킬 수 있다.
- [161] 도 12a를 참조하면, 인코더 학습 시스템(20)은 전체 영역 중 일부 영역에 대한 정보(예컨대, 이미지 정보)를 포함하는 학습 이미지(제1 이미지로도 지칭될 수 있음)를 획득할 수 있다. 예를 들어, 학습 이미지는 전체 영역 중 일부 영역을 제외한 영역에 대한 이미지 정보를 포함할 수 있다. 인코더 학습 시스템(20)은 제5 인코더(1350)에 학습 이미지를 입력할 수 있다. 제5 인코더(1350)는 학습 이미지를 인코딩함으로써 제1 이미지 임베딩을 출력할 수 있다.
- [162] 인코더 학습 시스템(20)은 전체 영역에 대한 이미지 정보를 포함하는 그라운드 트루스 이미지(제2 이미지로도 지칭될 수 있음)를 획득할 수 있다. 본 개시에 있어서, 그라운드 트루스 이미지는 학습 이미지로부터 추론하고자 하는 정답 이미지를 나타낼 수 있다. 인코더 학습 시스템(20)은 제6 인코더(21)에 그라운드 트루스(ground truth) 이미지를 입력할 수 있다. 제6 인코더(21)는 그라운드 트루스 이

미지를 인코딩함으로써 제2 이미지 임베딩을 출력할 수 있다. 본 개시의 일 실시예에 있어서, 제6 인코더(21)는 기 학습된 인공지능 모델일 수 있다. 예를 들어, 제6 인코더(21)는 기 학습된 대비적 언어-이미지 사전학습(Contrastive Language-Image Pretraining; CLIP) 모델의 이미지 인코더 부분일 수 있다.

- [163] 본 개시의 일 실시예에 있어서, 제6 인코더(21)는 기 학습된 상태에서 더 이상 학습되지 않을 수 있다. 예를 들어, 제6 인코더(21)는 기 학습되어 고정된 파라미터를 가질 수 있다. 제6 인코더(21)의 파라미터는 더 이상 갱신되지 않을 수 있다.
- [164] 인코더 학습 시스템(20)은 제1 이미지 임베딩과 제2 이미지 임베딩에 기초하여 제1 손실을 획득할 수 있다. 예를 들어, 인코더 학습 시스템(20)은 제1 이미지 임베딩과 제2 이미지 임베딩 간의 유사도를 계산할 수 있다. 인코더 학습 시스템(20)은 유사도에 기초하여 제1 손실을 획득할 수 있다. 예를 들어, 유사도가 클수록 제1 손실은 작을 수 있다.
- [165] 인코더 학습 시스템(20)은 그라운드 트루스 이미지를 나타내는 그라운드 트루스 텍스트를 획득할 수 있다. 본 개시에서, 그라운드 트루스 텍스트는 학습 이미지로부터 추론하고자 하는 정답 이미지를 설명하는 문장을 나타낼 수 있다. 인코더 학습 시스템(20)은 제7 인코더(22)에 그라운드 트루스 텍스트를 입력할 수 있다. 제7 인코더(22)는 그라운드 트루스 텍스트를 인코딩함으로써 텍스트 임베딩을 출력할 수 있다. 본 개시의 일 실시예에 있어서, 제7 인코더(22)는 기 학습된 인공지능 모델일 수 있다. 예를 들어, 제7 인코더(22)는 기 학습된 대비적 언어-이미지 사전학습(Contrastive Language-Image Pretraining; CLIP) 모델의 텍스트 인코더 부분일 수 있다.
- [166] 본 개시의 일 실시예에 있어서, 제7 인코더(22)는 기 학습된 상태에서 더 이상 학습되지 않을 수 있다. 예를 들어, 제7 인코더(22)는 기 학습되어 고정된 파라미터를 가질 수 있다. 제7 인코더(22)의 파라미터는 더 이상 갱신되지 않을 수 있다.
- [167] 인코더 학습 시스템(20)은 제1 이미지 임베딩과 텍스트 임베딩에 기초하여 제2 손실을 획득할 수 있다. 예를 들어, 인코더 학습 시스템(20)은 제1 이미지 임베딩과 텍스트 임베딩 간의 유사도를 계산할 수 있다. 인코더 학습 시스템(20)은 유사도에 기초하여 제2 손실을 획득할 수 있다. 예를 들어, 유사도가 클수록 제2 손실은 작을 수 있다. 본 개시의 일 실시예에 있어서, 제2 손실을 대비 손실(contrastive loss) 기법을 이용하여 획득할 수 있으나, 본 개시는 이에 한정되지 않으며, 제2 손실은 제1 이미지 임베딩과 텍스트 임베딩과의 유사도를 최대화하기 위한 임의의 손실 계산 기법(또는 손실 함수)을 이용할 수 있다.
- [168] 본 개시의 일 실시예에 있어서, 인코더 학습 시스템(20)은 제1 손실 및 제2 손실에 기초하여 제5 인코더(1350)의 적어도 하나의 파라미터(예컨대, 가중치 및/또는 편향)를 갱신할 수 있다.
- [169] 본 개시의 일 실시예에 있어서, 제6 인코더(21)의 적어도 하나의 파라미터는 갱신되지 않을 수 있으나, 본 개시는 이에 한정되지 않는다. 예를 들어, 인코더 학습

시스템(20)은 제1 손실 및/또는 제2 손실에 기초하여 제6 인코더(21)의 적어도 하나의 파라미터를 갱신할 수도 있다.

- [170] 본 개시의 일 실시예에 있어서, 제7 인코더(22)의 적어도 하나의 파라미터는 갱신되지 않을 수 있으나, 본 개시는 이에 한정되지 않는다. 예를 들어, 인코더 학습 시스템(20)은 제1 손실 및/또는 제2 손실에 기초하여 제7 인코더(22)의 적어도 하나의 파라미터를 갱신할 수도 있다.
- [171] 도 12b를 참조하면, 제5 인코더(1350)는 학습 이미지에 마스크 맵이 연결된 데이터를 입력으로 하여 제1 이미지 임베딩을 출력할 수 있다. 본 개시의 일 실시예에 있어서, 제5 인코더(1350)는 학습 이미지에 대응하는 3 채널 및 마스크 맵에 대응하는 1 채널을 갖는 총 4 채널의 입력 데이터 형식을 지원할 수 있다. 제6 인코더(21)는 그라운드 트루스 이미지에 대응하는 총 3 채널의 입력 데이터 형식을 지원할 수 있다. 그러나 본 개시는 이에 한정되지 않으며, 제5 인코더(1350)는 제6 인코더(21)보다 마스크 맵의 채널 수만큼 더 많은 채널의 입력 데이터 형식을 지원할 수 있다.
- [172] 도 13a 및 13b를 참조하면, 인코더 학습 시스템(20)은 제5 인코더(1350) 및 제7 인코더(22)를 포함할 수 있다. 본 개시의 일 실시예에 있어서, 도 12a 및 12b의 제6 인코더(21)는 제5 인코더(1350)와 동일한 인코더일 수 있다.
- [173] 도 13a를 참조하면, 제5 인코더(1350)는 임의의 학습 단계(예컨대, 이터레이션(iteration))에서 학습 이미지를 입력으로 하여 제1 이미지 임베딩을 출력할 수 있다. 제1 이미지 임베딩이 출력된 동일한 학습 단계에서, 제5 인코더(1350)는 그라운드 트루스 이미지를 입력으로 하여 제2 이미지 임베딩을 출력할 수 있다. 인코더 학습 시스템(20)은 제1 이미지 임베딩 및 제2 이미지 임베딩에 기초하여 제1 손실을 획득할 수 있다. 제7 인코더(22)는 그라운드 트루스 텍스트를 입력으로 하여 텍스트 임베딩을 출력할 수 있다. 인코더 학습 시스템(20)은 제1 이미지 임베딩 및 텍스트 임베딩에 기초하여 제2 손실을 획득할 수 있다. 인코더 학습 시스템(20)은 제1 손실 및 제2 손실에 기초하여 제5 인코더(1350)의 적어도 하나의 파라미터(예컨대, 가중치 및/또는 편향)를 갱신할 수 있다.
- [174] 도 13b를 참조하면, 제5 인코더(1350)는 학습 이미지에 제1 마스크 맵이 연결된 데이터를 입력으로 하여 제1 이미지 임베딩을 출력할 수 있다. 제1 마스크 맵은 전체 영역에서 일부 영역(예컨대, 미지 영역 또는 마스크된 영역)을 구분하는 이진 값 및 좌표 값으로 구성될 수 있다. 제5 인코더(1350)는 그라운드 트루스 이미지에 제2 마스크 맵이 연결된 데이터를 입력으로 하여 제2 이미지 임베딩을 출력할 수 있다. 제2 마스크 맵은 미지 영역이 없으므로, 전체 영역이 하나의 값으로 구성될 수 있다.
- [175] 도 12a 내지 13b과 함께, 도 11을 참조하면, 인코더 학습 시스템(20)에 의해 학습된 제5 인코더(1350)는 제3 이미지(예컨대, 일부 영역의 정보를 포함하는 이미지)를 입력 받아 인코딩할 수 있다. 학습된 제5 인코더(1350)는 제3 이미지에 대응하는 제1 이미지 임베딩을 출력할 수 있다. 제1 이미지 임베딩은 가이던스 정보 또

는 이미지 가이드스 정보로도 지칭될 수 있다. 제1 이미지 임베딩을 제3 이미지를 입력으로 하여 제3 이미지와 적어도 일부와 상이한 제4 이미지(예컨대, 최종 생성 이미지)를 추론하는 제2 생성 모델(1200)의 적어도 하나의 레이어에 입력될 수 있다.

- [176] 본 개시의 일 실시예에 따르면, 도 12a 및 12b에 도시된 제6 인코더(21)가 생략됨으로써, 제6 인코더(21)의 메모리 용량 만큼을 더 확보할 수 있으며, 제5 인코더의 학습 비용이 낮아지고 학습 속도가 향상될 수 있다.
- [177] 도 14는 본 개시의 일 실시예에 따른 생성 모델을 이용하여 이미지의 일부 영역을 생성하는 방법을 설명하기 위한 흐름도이다. 도 1 내지 13b에서 설명한 내용과 중복되는 내용은 생략한다. 설명의 편의를 위해, 도 1을 참조하여, 도 14를 설명한다.
- [178] 도 14를 참조하면, 생성 모델을 이용하여 이미지의 일부 영역을 생성하는 방법은 단계 S1410 내지 S1430을 포함할 수 있다. 본 개시의 일 실시예에 있어서, 단계 S1410 내지 S1430은 전자 장치(1000) 또는 전자 장치(1000) 프로세서(미도시)에 의해 수행될 수 있다. 그러나 본 개시는 이에 한정되지 않으며, 단계 S1410 내지 S1430은 임의의 전자 장치에 의해 수행될 수 있다. 본 개시의 일 실시예에 따르면, 생성 모델을 이용하여 이미지의 일부 영역을 생성하는 방법은 도 14에 도시된 바에 한정되지 않으며, 도 14에 도시된 단계 중 어느 하나를 생략할 수도 있고, 도 14에 도시되지 않은 단계를 더 포함할 수도 있다.
- [179] 단계 S1410에서, 전자 장치(1000)는 일부 영역의 정보를 포함하는 이미지를 획득할 수 있다. 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 일부 영역에 대응하는 마스크 맵을 획득할 수 있다. 전자 장치(100)는 일부 영역의 정보를 포함하는 이미지에 마스크 맵을 연결할 수 있다.
- [180] 단계 S1420에서, 전자 장치(1000)는 일부 영역의 정보를 포함하는 이미지를 입력으로 하는 제1 생성 모델을 이용하여, 일부 영역(예컨대, 미지 영역 또는 마스크된 영역)에 대한 제1 이미지 정보를 포함하는 중간 생성 이미지를 획득할 수 있다. 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 제1 생성 모델(1100)로부터 제1 픽셀 정보를 획득할 수 있다. 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 제1 픽셀 정보와 일부 영역 외의 영역에 대한 이미지의 원본 픽셀 정보 간의 블렌딩(blending) 작업을 수행함으로써 중간 생성 이미지를 획득할 수 있다.
- [181] 단계 S1430에서, 전자 장치(1000)는 일부 영역의 정보를 포함하는 이미지 및 중간 생성 이미지를 입력으로 하는 제2 생성 모델(1200)을 이용하여, 제1 이미지 정보와 적어도 일부가 상이한 제2 이미지 정보를 포함하는 최종 생성 이미지를 획득할 수 있다. 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 제2 생성 모델(120)로부터 제2 픽셀 정보를 획득할 수 있다. 전자 장치(1000)는 제2 픽셀 정보와 일부 영역 외의 영역에 대한 이미지의 원본 픽셀 정보 간의 블렌딩(blending) 작업을 수행함으로써 최종 생성 이미지를 획득할 수 있다. 본 개시의 일 실시예에 있

어서, 제2 생성 모델(1200)은 기 정의된 디노이징 총 차수만큼 디노이징 동작을 반복하여 수행함으로써 최종 생성 이미지를 획득할 수 있다.

- [182] 도 15는 도 14의 S1410 단계의 세부 단계들을 설명하기 위한 흐름도이다. 도 1 내지 14에서 설명한 내용과 중복되는 내용은 생략한다. 설명의 편의를 위해, 도 3을 참조하여, 도 15를 설명한다.
- [183] 도 15를 참조하면, 도 14의 단계 S1410은 단계 S1510 내지 S1520을 포함할 수 있다. 본 개시의 일 실시예에 있어서, 단계 S1510 내지 S1520은 전자 장치(1000) 또는 전자 장치(1000)의 프로세서(미도시)에 의해 수행될 수 있다. 그러나 본 개시는 이에 한정되지 않으며, 단계 S1510 내지 S1520은 임의의 전자 장치에 의해 수행될 수 있다. 본 개시에 따른 단계 S1410의 세부 단계들은 도 15에 도시된 바에 한정되지 않으며, 도 15에 도시된 단계 중 어느 하나를 생략할 수도 있고, 도 15에 도시되지 않은 단계를 더 포함할 수도 있다.
- [184] 단계 S1510에서, 전자 장치(1000)는 일부 영역의 정보를 포함하는 이미지의 전체 영역에서 일부 영역을 구분하는 마스크 맵을 획득할 수 있다. 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 외부 서버로부터 마스크 맵을 획득할 수 있다. 본 개시의 일 실시예에 있어서, 이미지의 마스크된 영역을 제1 값으로, 이미지의 마스크되지 않은 영역을 제2 값으로 구분함으로써 마스크 맵을 생성할 수 있다.
- [185] 단계 S1520에서, 전자 장치(1000)는 일부 영역의 정보를 포함하는 이미지에 마스크 맵을 연결할 수 있다. 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 일부 영역의 정보를 포함하는 이미지를 인코딩할 수 있다. 전자 장치(1000)는 인코딩된 이미지에 마스크 맵을 연결할 수 있다. 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 연결된 이미지에 현재 노이즈 정보를 더 연결할 수 있다. 전자 장치(1000)는 현재 노이즈 정보까지 연결된 데이터를 제2 생성 모델(1200)에 입력할 수 있다.
- [186] 도 16은 도 14의 S1430 단계의 세부 단계들을 설명하기 위한 흐름도이다. 도 1 내지 15에서 설명한 내용과 중복되는 내용은 생략한다. 설명의 편의를 위해, 도 5d 및 5e를 참조하여, 도 16을 설명한다.
- [187] 도 16을 참조하면, 도 14의 단계 S1430은 단계 S1610 내지 S1620을 포함할 수 있다. 본 개시의 일 실시예에 있어서, 단계 S1610 내지 S1620은 전자 장치(1000) 또는 전자 장치(1000)의 프로세서(미도시)에 의해 수행될 수 있다. 그러나 본 개시는 이에 한정되지 않으며, 단계 S1610 내지 S1620은 임의의 전자 장치에 의해 수행될 수 있다. 본 개시에 따른 단계 S1430의 세부 단계들은 도 16에 도시된 바에 한정되지 않으며, 도 16에 도시된 단계 중 어느 하나를 생략할 수도 있고, 도 16에 도시되지 않은 단계를 더 포함할 수도 있다.
- [188] 단계 S1610에서, 전자 장치(1000)는 중간 생성 이미지를 인코딩할 수 있다. 전자 장치(1000)는 기 학습된 인코더를 이용하여 중간 생성 이미지를 인코딩할 수 있다. 본 개시의 일 실시예에 있어서, 인코딩된 중간 생성 이미지는 제2 생성 모델

(1200)의 적어도 하나의 레이어에 입력될 수 있다. 인코딩된 중간 생성 이미지는 제2 생성 모델(1200)에서 가이드스 정보로 활용될 수 있다.

- [189] 단계 S1620에서, 전자 장치(1000)는 일부 영역의 정보를 포함하는 이미지 및 인코딩된 중간 생성 이미지를 입력으로 하는 제2 생성 모델을 이용하여, 제1 이미지 정보와 적어도 일부가 상이한 제2 이미지 정보를 포함하는 최종 생성 이미지를 획득할 수 있다. 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 인코딩된 중간 생성 이미지를 인터프리터에 입력함으로써 변환 데이터를 획득할 수 있다. 전자 장치(1000)는 변환 데이터를 제2 생성 모델의 적어도 하나의 레이어에 입력할 수 있다.
- [190] 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 중간 생성 이미지를 제2 생성 모델(1200)에 전달할 수 있다. 전자 장치(1000)는 일부 영역의 정보를 포함하는 이미지, 중간 생성 이미지, 인코딩된 중간 생성 이미지를 입력으로 하는 제2 생성 모델을 이용하여, 제1 이미지 정보와 적어도 일부가 상이한 제2 이미지 정보를 포함하는 최종 생성 이미지를 획득할 수 있다.
- [191] 도 17은 도 14의 S1430 단계의 세부 단계들을 설명하기 위한 흐름도이다. 도 1 내지 16에서 설명한 내용과 중복되는 내용은 생략한다. 설명의 편의를 위해, 도 5a를 참조하여, 도 17을 설명한다.
- [192] 도 17을 참조하면, 도 14의 단계 S1430은 단계 S1710 내지 S1730을 포함할 수 있다. 본 개시의 일 실시예에 있어서, 단계 S1710 내지 S1730은 전자 장치(1000) 또는 전자 장치(1000)의 프로세서(미도시)에 의해 수행될 수 있다. 그러나 본 개시는 이에 한정되지 않으며, 단계 S1710 내지 S1730은 임의의 전자 장치에 의해 수행될 수 있다. 본 개시에 따른 단계 S1430의 세부 단계들은 도 17에 도시된 바에 한정되지 않으며, 도 17에 도시된 단계 중 어느 하나를 생략할 수도 있고, 도 17에 도시되지 않은 단계를 더 포함할 수도 있다.
- [193] 단계 S1710에서, 전자 장치(1000)는 텍스트 입력을 획득할 수 있다. 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 외부 서버로부터 텍스트 입력을 획득할 수 있다. 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 사용자 인터페이스로부터 텍스트 입력을 획득할 수 있다. 예를 들어, 텍스트 입력은 일부 영역의 정보를 포함하는 이미지 및/또는 최종 생성 이미지를 설명하는 문장을 포함할 수 있다.
- [194] 단계 S1720에서, 전자 장치(1000)는 텍스트 입력을 인코딩할 수 있다. 전자 장치(1000)는 기 학습된 인코더를 이용하여 텍스트 입력을 인코딩할 수 있다. 본 개시의 일 실시예에 있어서, 인코딩된 텍스트 입력은 제2 생성 모델(1200)의 적어도 하나의 레이어에 입력될 수 있다. 인코딩된 텍스트 입력은 제2 생성 모델(1200)에서 가이드스 정보로 활용될 수 있다.
- [195] 단계 S1730에서, 전자 장치(1000)는 인코딩된 텍스트 입력, 일부 영역의 정보를 포함하는 이미지, 및 중간 생성 이미지를 입력으로 하는 제2 생성 모델(1200)을 이용하여, 제1 이미지 정보와 적어도 일부가 상이한 제2 이미지 정보를 포함하

는 최종 생성 이미지를 획득할 수 있다. 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 중간 생성 이미지에 기초하여 현재 노이즈 정보를 생성할 수 있다. 전자 장치(1000)는 일부 영역의 정보를 포함하는 이미지 및 현재 노이즈 정보를 제2 생성 모델(1200)에 입력할 수 있다. 전자 장치(1000)는 인코딩된 텍스트 입력을 제2 생성 모델(1200)의 적어도 하나의 레이어에 입력할 수 있다. 전자 장치(1000)는 제2 생성 모델(1200)로부터 다음 노이즈 정보를 획득할 수 있다.

- [196] 도 18은 도 14의 S1430 단계의 세부 단계들을 설명하기 위한 흐름도이다. 도 1 내지 17에서 설명한 내용과 중복되는 내용은 생략한다. 설명의 편의를 위해, 도 9a 및 9b를 참조하여, 도 18을 설명한다.
- [197] 도 18을 참조하면, 도 14의 단계 S1430은 단계 S1810 내지 S1830을 포함할 수 있다. 본 개시의 일 실시예에 있어서, 단계 S1810 내지 S1830은 전자 장치(1000) 또는 전자 장치(1000)의 프로세서(미도시)에 의해 수행될 수 있다. 그러나 본 개시는 이에 한정되지 않으며, 단계 S1810 내지 S1830은 임의의 전자 장치에 의해 수행될 수 있다. 본 개시에 따른 단계 S1430의 세부 단계들은 도 18에 도시된 바에 한정되지 않으며, 도 18에 도시된 단계 중 어느 하나를 생략할 수도 있고, 도 18에 도시되지 않은 단계를 더 포함할 수도 있다.
- [198] 단계 S1810에서, 전자 장치(1000)는 중간 생성 이미지에 대한 디노이징 강도를 획득할 수 있다. 본 개시의 일 실시예에 있어서, 디노이징 강도는 중간 생성 이미지에 추가할 노이즈 량에 대응할 수 있다. 본 개시의 일 실시예에 있어서, 디노이징 강도는 미리 정의될 수 있다. 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 중간 생성 이미지에 기초하여 예측 신뢰 값을 획득할 수 있다. 전자 장치(1000)는 예측 신뢰 값에 기초하여 디노이징 강도를 결정할 수 있다. 전자 장치(1000)는 마스크 맵에 기초하여 일부 영역(예컨대, 마스크된 영역)의 크기 및/또는 모양을 식별할 수 있다. 전자 장치(1000)는 일부 영역(예컨대, 마스크된 영역)의 크기 및/또는 모양에 기초하여 디노이징 강도를 결정할 수 있다.
- [199] 단계 S1820에서, 전자 장치(1000)는 디노이징 강도에 기초하여 중간 생성 이미지에 노이즈를 추가할 수 있다. 본 개시의 일 실시예에 있어서, 디노이징 강도가 높을수록 추가되는 노이즈 량이 많을 수 있다.
- [200] 단계 S1830에서, 전자 장치(1000)는 일부 영역의 정보를 포함하는 이미지 및 노이즈가 추가된 중간 생성 이미지를 입력으로 하는 제2 생성 모델(1200)을 이용하여, 제1 이미지 정보와 적어도 일부가 상이한 제2 이미지 정보를 포함하는 최종 생성 이미지를 획득할 수 있다.
- [201] 도 19는 도 14의 S1430 단계의 세부 단계들을 설명하기 위한 흐름도이다. 도 1 내지 18에서 설명한 내용과 중복되는 내용은 생략한다. 설명의 편의를 위해, 도 9a 및 9b를 참조하여, 도 19를 설명한다.
- [202] 도 19를 참조하면, 도 14의 단계 S1430은 단계 S1910 내지 S1950을 포함할 수 있다. 본 개시의 일 실시예에 있어서, 단계 S1910 내지 S1950은 전자 장치(1000) 또는 전자 장치(1000)의 프로세서(미도시)에 의해 수행될 수 있다. 그러나 본 개시

는 이에 한정되지 않으며, 단계 S1910 내지 S1950은 임의의 전자 장치에 의해 수행될 수 있다. 본 개시에 따른 단계 S1430의 세부 단계들은 도 19에 도시된 바에 한정되지 않으며, 도 19에 도시된 단계 중 어느 하나를 생략할 수도 있고, 도 19에 도시되지 않은 단계를 더 포함할 수도 있다.

- [203] 단계 S1910에서, 전자 장치(1000)는 현재 노이즈 정보를 획득할 수 있다. 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 제2 생성 모델(1200)의 출력으로부터 현재 노이즈 정보를 획득할 수 있다. 본 개시의 일 실시예에 있어서, 초기 디노이징 차수의 경우, 제2 생성 모델(1200)의 출력이 존재하지 않는다. 이 경우, 전자 장치(1000)는 무작위 값들로 구성되는 현재 노이즈 정보를 생성할 수 있다. 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 가우시안 노이즈에 따른 무작위 값들로 구성되는 현재 노이즈 정보를 생성할 수 있다.
- [204] 단계 S1920에서, 전자 장치(1000)는 현재 노이즈 정보 및 일부 영역의 정보를 포함하는 이미지 (또는 인코딩된 이미지)를 연결할 수 있다. 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 마스크 맵 또는 일부 영역의 정보를 포함하는 이미지의 위치 및 폭을 갖도록 크기가 조절된 마스크 맵을 현재 노이즈 정보 및 일부 영역의 정보를 포함하는 이미지에 함께 연결할 수 있다.
- [205] 단계 S1930에서, 전자 장치(1000)는 연결된 이미지를 제2 생성 모델(1200)에 입력할 수 있다. 본 개시의 일 실시예에 있어서, 제2 생성 모델(1200)는 디노이징 동작을 수행함으로써 연결된 이미지에 기초하여 다음 노이즈 정보를 출력할 수 있다.
- [206] 단계 S1940에서, 전자 장치(1000)는 제2 생성 모델(1200)의 출력인 다음 노이즈 정보를 획득할 수 있다.
- [207] 단계 S1950에서, 전자 장치(1000)는 기 정의된 디노이징 총 차수 만큼 디노이징 동작을 반복하였는지를 결정할 수 있다. 기 정의된 디노이징 총 차수만큼 디노이징 동작을 반복한 것으로 결정한 것에 기초하여(예), 전자 장치(1000)는 다음 노이즈 정보에 기초하여 최종 생성 이미지를 생성할 수 있다. 기 정의된 디노이징 총 차수만큼 디노이징 동작을 반복한 것으로 결정한 것에 기초하여(아니오), 절차는 단계 S1910으로 이동한다. 다음 노이즈 정보는 다음 디노이징 차수의 현재 노이즈 정보가 될 수 있다.
- [208] 도 20은 본 개시의 일 실시예에 따른 이미지를 인코딩하는 인코더를 학습하는 방법을 설명하기 위한 흐름도이다. 도 1 내지 19에서 설명한 내용과 중복되는 내용은 생략한다. 설명의 편의를 위해, 도 12a 및 13b를 참조하여, 도 20을 설명한다. 도 12a 및 13b의 제5 인코더(1350), 제6 인코더(21), 및 제7 인코더(22)의 구성, 기능, 및 동작은 각각 도 20에서 제1 인코더, 제2 인코더, 및 제3 인코더의 구성, 기능, 및 동작에 대응할 수 있다.
- [209] 도 20을 참조하면, 이미지를 인코딩하는 인코더를 학습하는 방법은 단계 S2010 내지 S2070을 포함할 수 있다. 본 개시의 일 실시예에 있어서, 단계 S2010 내지 S2070은 인코더 학습 시스템(20)에 의해 수행될 수 있고, 인코더 학습 시스템(20)

의 기능 중 적어도 일부는 임의의 전자 장치 또는 임의의 전자 장치의 프로세서에 의해 수행될 수 있다. 본 개시의 일 실시예에 따른, 이미지를 인코딩하는 인코더를 학습하는 방법은 도 20에 도시된 바에 한정되지 않으며, 도 20에 도시된 단계 중 어느 하나를 생략할 수도 있고, 도 20에 도시되지 않은 단계를 더 포함할 수도 있다.

- [210] 단계 S2010에서, 인코더 학습 시스템(20)은 전체 영역 중 일부 영역에 대한 정보를 포함하는 제1 이미지, 전체 영역에 대한 이미지 정보를 포함하는 제2 이미지, 및 제2 이미지를 나타내는 텍스트를 획득할 수 있다. 본 개시의 일 실시예에 있어서, 인코더 학습 시스템(20)은 제1 이미지에, 전체 영역에서 일부 영역을 구분하는 제1 마스크 맵을 연결할 수 있다.
- [211] 단계 S2020에서, 인코더 학습 시스템(20)은 제1 이미지를 입력으로 하는 제1 인코더를 이용하여, 제1 이미지 임베딩을 획득할 수 있다.
- [212] 단계 S2030에서, 인코더 학습 시스템(20)은 제2 이미지를 입력으로 하는 제2 인코더를 이용하여 제2 이미지 임베딩을 획득할 수 있다. 본 개시의 일 실시예에 있어서, 제2 인코더는 기 학습된 인공지능 모델일 수 있다. 본 개시의 일 실시예에 있어서, 제2 인코더는 기 학습된 상태에서 더 이상 학습되지 않을 수 있다. 예를 들어, 제2 인코더는 고정된 파라미터를 갖되, 더 이상 갱신되지 않을 수 있다. 본 개시의 일 실시예에 있어서, 제1 인코더와 제2 인코더는 동일한 인코더일 수 있다. 이 경우, 제1 인코더 및 제2 인코더에 대응하는 인코더는 학습이 완료되지 않은 인코더일 수 있다. 예를 들어, 제1 인코더 및 제2 인코더에 대응하는 인코더는 학습이 수행되지 않을 수 있으나, 본 개시는 이에 한정되지 않으며, 인코더는 기 학습된 인공지능 모델일 수 있다. 기 학습된 인코더는 미세 조정과 같은 추가 학습 기법을 통해 추가로 학습될 수 있다. 본 개시의 일 실시예에 있어서, 인코더 학습 시스템(20)은 제2 이미지에, 전체 영역을 하나의 값으로 구성한 제2 마스크 맵을 연결할 수 있다.
- [213] 단계 S2040에서, 인코더 학습 시스템(20)은 텍스트를 입력으로 하는 제3 인코더를 이용하여, 텍스트 임베딩을 획득할 수 있다. 본 개시의 일 실시예에 있어서, 제3 인코더는 기 학습된 인공지능 모델일 수 있다. 본 개시의 일 실시예에 있어서, 제3 인코더는 기 학습된 상태에서 더 이상 학습되지 않을 수 있다. 예를 들어, 제3 인코더는 고정된 파라미터를 갖되, 더 이상 갱신되지 않을 수 있다.
- [214] 단계 S2050에서, 인코더 학습 시스템(20)은 제1 이미지 임베딩과 제2 이미지 임베딩에 기초하여 제1 손실을 획득할 수 있다.
- [215] 단계 S2060에서, 인코더 학습 시스템(20)은 제1 이미지 임베딩과 텍스트 임베딩에 기초하여 제2 손실을 획득할 수 있다.
- [216] 단계 S2070에서, 인코더 학습 시스템(20)은 제1 손실 및 제2 손실에 기초하여 제1 인코더의 적어도 하나의 파라미터를 갱신할 수 있다. 본 개시의 일 실시예에 있어서, 인코더 학습 시스템(20)은 제1 손실 및 제2 손실에 기초하여 제2 인코더 및/또는 제3 인코더의 적어도 하나의 파라미터를 갱신할 수 있다. 본 개시의 일 실시

예에 있어서, 인코더 학습 시스템(20)은 기 정의된 학습 차수 동안 제1 인코더, 제2 인코더, 및/또는 제3 인코더의 적어도 하나의 파라미터를 반복하여 갱신할 수 있다.

- [217] 도 21은 본 개시의 일 실시예에 따른 생성 모델을 이용하여 이미지의 일부 영역을 생성하는 방법을 설명하기 위한 흐름도이다. 도 1 내지 20에서 설명한 내용과 중복되는 내용은 생략한다. 설명의 편의를 위해, 도 11을 참조하여, 도 21을 설명한다. 도 11의 제5 인코더(1350) 및 제2 생성 모델(1200)의 구성, 기능, 및 동작은 도 21의 제1 인코더 및 생성 모델의 구성, 기능, 및 동작에 대응할 수 있다.
- [218] 도 21을 참조하면, 생성 모델을 이용하여 이미지의 일부 영역을 생성하는 방법은 단계 S2110 내지 S2130을 포함할 수 있다. 본 개시의 일 실시예에 있어서, 단계 S2110 내지 S2130은 전자 장치(1000) 또는 전자 장치(1000) 프로세서(미도시)에 의해 수행될 수 있다. 그러나 본 개시는 이에 한정되지 않으며, 단계 S2110 내지 S2130은 임의의 전자 장치에 의해 수행될 수 있다. 본 개시의 일 실시예에 따른, 생성 모델을 이용하여 이미지의 일부 영역을 생성하는 방법은 도 21에 도시된 바에 한정되지 않으며, 도 21에 도시된 단계 중 어느 하나를 생략할 수도 있고, 도 21에 도시되지 않은 단계를 더 포함할 수도 있다.
- [219] 단계 S2110에서, 전자 장치(1000)는 일부 영역의 정보를 포함하는 이미지를 획득할 수 있다.
- [220] 단계 S2120에서, 전자 장치(1000)는 일부 영역의 정보를 포함하는 이미지를 입력으로 하는 제1 인코더를 이용하여, 대상 이미지 임베딩을 획득할 수 있다. 본 개시의 일 실시예에 있어서, 제1 인코더는, 전체 영역 중 일부 영역에 대한 위치 정보를 포함하는 제1 학습 이미지, 전체 영역에 대한 이미지 정보를 포함하는 제2 학습 이미지, 및 제2 학습 이미지를 나타내는 학습 텍스트를 획득하고, 제1 학습 이미지를 입력으로 하는 제1 인코더를 이용하여, 제1 이미지 임베딩을 획득하고, 제2 학습 이미지를 입력으로 하는 제2 인코더를 이용하여 제2 이미지 임베딩을 획득하고, 학습 텍스트를 입력으로 하는 제3 인코더를 이용하여 텍스트 임베딩을 획득하고, 제1 이미지 임베딩과 제2 이미지 임베딩에 기초하여 제1 손실을 획득하고, 제1 이미지 임베딩과 텍스트 임베딩에 기초하여 제2 손실을 획득하고, 제1 손실 및 제2 손실에 기초하여 제1 인코더의 적어도 하나의 파라미터를 갱신함으로써 학습될 수 있다. 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 제1 손실 및/또는 제2 손실에 기초하여 제2 인코더 및/또는 제3 인코더의 적어도 하나의 파라미터를 갱신할 수 있다.
- [221] 단계 S2130에서, 전자 장치(1000)는 일부 영역의 정보를 포함하는 이미지, 및 대상 이미지 임베딩을 입력으로 하는 생성 모델을 이용하여, 최종 생성 이미지를 획득할 수 있다.
- [222] 본 개시의 일 실시예에 있어서, 생성 모델은 일부 영역의 정보를 포함하는 이미지에 기초하여 최종 생성 이미지를 출력하는 제1 뉴럴 네트워크를 포함할 수 있다. 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 대상 이미지 임베딩을 제1

뉴럴 네트워크의 적어도 하나의 레이어에 입력할 수 있다. 본 개시의 일 실시예에 있어서, 생성 모델은 대상 이미지 임베딩을 변환하여 제1 뉴럴 네트워크의 적어도 하나의 레이어에 전달하는 인터프리터를 포함할 수 있다. 본 개시의 일 실시예에 있어서, 생성 모델의 인터프리터는 제1 단층 퍼셉트론 및 제2 단층 퍼셉트론을 포함할 수 있다. 본 개시의 일 실시예에 있어서, 생성 모델의 인터프리터는 3개 이상의 단층 퍼셉트론들을 포함할 수 있다. 본 개시의 일 실시예에 있어서, 생성 모델의 인터프리터는 유일한 단층 퍼셉트론을 포함할 수 있다.

- [223] 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 대상 이미지 임베딩을 입력으로 하는 제1 단층 퍼셉트론을 이용하여 중간 이미지 임베딩을 획득할 수 있다. 전자 장치(1000)는 중간 이미지 임베딩을 입력으로 하는 제2 단층 퍼셉트론을 이용하여 결과 이미지 임베딩을 획득할 수 있다. 전자 장치(1000)는 결과 이미지 임베딩을 제1 뉴럴 네트워크의 적어도 하나의 레이어에 입력할 수 있다.
- [224] 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 이전 단층 퍼셉트론이 출력한 이미지 임베딩을 입력으로 하는 다음 단층 퍼셉트론을 다중으로 구성하고, 마지막 단층 퍼셉트론을 이용하여 결과 이미지 임베딩을 획득할 수 있다. 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 제1 단층 퍼셉트론을 이용하여 곧바로 결과 이미지 임베딩을 획득할 수 있다.
- [225] 본 개시의 일 실시예에 있어서, 생성 모델은 일부 영역의 정보를 포함하는 이미지에 기초하여 제1 노이즈를 출력할 수 있다. 생성 모델은 일부 영역의 정보를 포함하는 이미지에 기초하여 제2 노이즈를 출력할 수 있다. 생성 모델은 대상 이미지 임베딩을 변환하여 제1 뉴럴 네트워크의 적어도 하나의 레이어에 전달하는 인터프리터를 포함할 수 있다.
- [226] 본 개시의 일 실시예에 있어서, 생성 모델은 제1 뉴럴 네트워크 및 제2 뉴럴 네트워크를 포함할 수 있다. 제1 뉴럴 네트워크는 일부 영역의 정보를 포함하는 이미지, 중간 생성 이미지, 및 이미지 가이드스 정보 중 적어도 하나를 이용하여 제1 노이즈를 출력할 수 있다. 제2 뉴럴 네트워크는 일부 영역의 정보를 포함하는 이미지, 중간 생성 이미지, 및 텍스트 가이드스 정보 중 적어도 하나를 이용하여 제2 노이즈를 출력할 수 있다.
- [227] 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 일부 영역의 정보를 포함하는 이미지에 대응하는 대상 텍스트를 획득할 수 있다. 전자 장치(1000)는 대상 텍스트에 기초하여 대상 텍스트 임베딩을 획득할 수 있다. 전자 장치(1000)는 대상 텍스트 임베딩을 제2 뉴럴 네트워크의 적어도 하나의 레이어에 입력할 수 있다.
- [228] 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 제1 노이즈 및 제2 노이즈에 기초하여 최종 생성 이미지를 획득할 수 있다. 본 개시의 일 실시예에 있어서, 전자 장치(1000)는 제1 노이즈 및 제2 노이즈의 가중합을 통해 다음 노이즈 정보를 획득할 수 있다.
- [229] 도 22는 본 개시의 일 실시예에 따른 사용자 장치의 구성을 설명하기 위한 블록도이다. 도 1 내지 21에서 설명한 내용과 중복되는 내용은 생략한다. 도 1 내지 11

의 전자 장치(1000)의 구성, 기능, 및 동작은 도 22의 사용자 장치(2000)의 구성, 기능, 및 동작에 대응할 수 있다.

- [230] 도 22를 참조하면, 사용자 장치(2000)는 통신 인터페이스(2100), 사용자 인터페이스(2200), 카메라(2300), 프로세서(2400), 및 메모리(2500)를 포함할 수 있다. 그러나 도시된 구성요소 모두가 필수구성요소인 것은 아니다. 도시된 구성요소보다 많은 구성요소에 의해 사용자 장치(2000)가 구현될 수도 있고, 그보다 적은 구성요소에 의해서도 사용자 장치(2000)는 구현될 수 있다.
- [231] 통신 인터페이스(2100)는, 사용자 장치(2000)와 서버 장치(미도시), 사용자 장치(2000)와 임의의 전자 장치(미도시), 및 사용자 장치(2000)와 다른 사용자 장치(미도시) 간의 통신을 수행하는 하나 이상의 구성요소를 포함할 수 있다.
- [232] 본 개시의 일 실시예에 있어서, 사용자 장치(2000)는 통신 인터페이스(2100)를 통해 서버 장치로부터 일부 영역(또는 마스크된 영역)을 포함하는 이미지를 수신할 수 있다. 사용자 장치(2000)는 통신 인터페이스(2100)를 통해 서버 장치로부터 마스크 맵을 수신할 수 있다. 사용자 장치(2000)는 통신 인터페이스(2100)를 통해 서버 장치로부터 텍스트 입력을 수신할 수 있다. 사용자 장치(2000)는 통신 인터페이스(2100)를 통해 다른 전자 장치로부터 최종 생성 이미지를 추론하기 위해 필요한 다양한 하이퍼파라미터들(예컨대, 디노이징 총 차수 등)을 수신할 수 있다. 사용자 장치(2000)는 통신 인터페이스(2100)를 통해 서버 장치로부터 기 학습된 생성 모델 및/또는 기 학습된 인코더를 수신할 수 있다.
- [233] 사용자 인터페이스(2200)는 입력 인터페이스 및 출력 인터페이스를 포함할 수 있다.
- [234] 입력 인터페이스는, 사용자로부터의 입력(이하에서, 사용자 입력)을 수신하기 위한 것이다. 입력 인터페이스는 키 패드(key pad), 돔 스위치 (dome switch), 터치 패드(접촉식 정전 용량 방식, 압력식 저항막 방식, 적외선 감지 방식, 표면 초음파 전도 방식, 적분식 장력 측정 방식, 피에조 효과 방식 등), 조그 휠, 조그 스위치, 마이크 중 적어도 하나일 수 있으나, 이에 한정되는 것은 아니다.
- [235] 본 개시의 일 실시예에 있어서, 사용자 장치(2000)는 입력 인터페이스를 통해 사용자가 설정한 하이퍼파라미터 등을 수신할 수 있다. 사용자 장치(2000)는 입력 인터페이스를 통해 이미지 및/또는 텍스트를 수신할 수 있다. 본 개시의 일 실시예에 있어서, 사용자 장치(2000)는 마이크를 통해 사용자의 음성을 처리한 오디오 신호를 획득할 수 있다. 사용자 장치(2000)는 오디오 신호를 텍스트로 변환할 수 있다.
- [236] 출력 인터페이스는 오디오 신호 또는 비디오 신호의 출력을 위한 것으로, 예컨대 디스플레이 또는 스피커 등을 포함할 수 있다.
- [237] 본 개시의 일 실시예에 있어서, 사용자 장치(2000)는 디스플레이를 통해서 영상을 표시할 수 있다. 예를 들어, 사용자 장치(2000)는 입력 인터페이스에 대응하는 GUI를 디스플레이를 통해 표시할 수 있다. 예를 들어, 사용자 장치(2000)는 이미지를 디스플레이를 통해 표시할 수 있다. 사용자 장치(2000)는 입력 인터페이스

를 통해 디스플레이에 표시된 이미지 중 일부 영역을 지정하는 사용자 입력을 수신할 수 있다. 사용자 장치(2000)는 사용자 입력에 기초하여 이미지의 일부 영역을 마스킹할 수 있다.

- [238] 본 개시의 일 실시예에 있어서, 사용자 장치(2000)는 입력 인터페이스를 통해 디스플레이에 표시된 이미지를 회전시키거나 및/또는 크기를 조절하는 사용자 입력을 수신할 수 있다. 사용자 장치(2000)는 사용자 입력에 기초하여 기 정의된 이미지 크기 내에 이미지 정보가 없는 영역을 마스킹할 수 있다.
- [239] 본 개시의 일 실시예에 있어서, 사용자 장치(2000)는 입력 인터페이스를 통해 디스플레이에 표시된 이미지 상에 그려진 사용자 입력(예컨대, 임의의 선이나 도형)을 수신할 수 있다. 사용자 장치(2000)는 사용자 입력에 대응하는 영역을 마스킹할 수 있다.
- [240] 본 개시의 일 실시예에 있어서, 사용자 장치(2000)는 객체 세그멘테이션(object segmentation)을 수행하는 인공지능 모델을 이용하여, 이미지 내의 적어도 하나의 객체 영역을 세그멘팅할 수 있다. 인공지능 모델은 사용자 장치(2000)의 메모리(2500)에 저장될 수 있다. 이 경우, 사용자 장치(2000)의 프로세서(2400)는 인공지능 모델에 이미지를 입력하여 세그멘테이션 결과를 출력할 수 있다. 또는, 사용자 장치(2000) 외부의 장치(예컨대, 서버)로부터 인공지능 모델의 세그멘테이션 결과를 수신할 수도 있다. 사용자 장치(2000)는 디스플레이에 이미지와 세그멘테이션 결과를 함께 표시할 수 있다. 사용자 장치(2000)는 입력 인터페이스를 통해 디스플레이에 표시된 적어도 하나의 객체를 선택하는 사용자 입력을 수신할 수 있다. 사용자 장치(2000)는 세그멘테이션 결과에 따른 객체들 중 사용자 입력에 대응하는 객체를 결정할 수 있다. 사용자 장치(2000)는 결정된 객체의 영역을 마스킹할 수 있다. 그러나 본 개시는 이에 한정되지 않으며, 사용자 장치(2000)는 결정된 객체의 영역과, 디스플레이에 표시된 이미지 상에 그려진 사용자 입력(예컨대, 임의의 선이나 도형)에 대응하는 영역을 함께 병합하여 마스킹할 수 있다. 본 개시의 일 실시예에 있어서, 사용자 장치(2000)는 디스플레이를 통해 이미지를 표시할 수 있다. 사용자 장치(2000)는 입력 인터페이스를 통해 디스플레이에 표시된 이미지의 적어도 하나의 위치 값에 대응하는 사용자 입력을 수신할 수 있다. 예를 들어, 적어도 하나의 위치 값은 이미지 픽셀의 좌표 값으로 구성될 수 있다. 예를 들어, 적어도 하나의 위치 값은 이미지 내의 특정 영역의 경계에 대응하는 이미지 픽셀의 좌표 값들로 구성될 수 있다. 사용자 장치(2000)는 적어도 하나의 위치 값을 입력으로 하는 인공지능 모델을 이용하여, 적어도 하나의 위치 값에 대응하는 객체의 영역과 그 외의 영역을 구분하는 세그멘테이션 맵을 획득할 수 있다. 사용자 장치(2000)는 세그멘테이션 맵에 기초하여 사용자 입력에 대응하는 객체의 영역을 마스킹할 수 있다.
- [241] 예를 들어, 디스플레이는 액정 디스플레이(liquid crystal display), 박막 트랜지스터 액정 디스플레이(thin film transistor-liquid crystal display), 발광 다이오드(LED, light-emitting diode), 유기 발광 다이오드(organic light-emitting diode), 플렉시블 디

스플레이(flexible display), 3차원 디스플레이(3D display), 전기영동 디스플레이(electrophoretic display) 중에서 적어도 하나를 포함할 수 있다. 그리고 사용자 장치(2000)의 구현 형태에 따라 디스플레이를 2개 이상 포함할 수도 있다.

- [242] 스피커는 통신 인터페이스(2100)로부터 수신되거나 메모리(2500)에 저장된 오디오 신호를 출력할 수 있다.
- [243] 카메라(2300)는 주변 공간을 촬영하여 이미지를 생성할 수 있다. 카메라(2300)는 이미지 센서를 포함할 수 있다. 본 개시의 일 실시예에 있어서, 사용자 장치(2000)는 카메라(2300)로부터 촬영된 이미지에 기초하여 제1 생성 모델(1100) 및/또는 제2 생성 모델(1200)을 학습시킬 수 있다. 본 개시의 일 실시예에 있어서, 사용자 장치(2000)는 카메라(2300)로부터 촬영된 이미지를 제1 생성 모델(1100) 및/또는 제2 생성 모델(1200)에 입력시킴으로써 최종 생성 이미지를 획득할 수 있다.
- [244] 프로세서(2400)는, 메모리(2500)에 저장된 프로그램 또는 정보를 이용하여 사용자 장치(2000)의 전반적인 동작을 제어할 수 있다. 프로세서(2400)는 AP(application processor), CPU(central processing unit) 또는 GPU(graphic processing unit)와 같은 범용 프로세서와 소프트웨어의 조합을 통해 구현될 수도 있다. 전용 프로세서의 경우, 본 개시의 실시예를 구현하기 위한 메모리를 포함하거나, 외부 메모리를 이용하기 위한 메모리 처리부를 포함할 수 있다. 프로세서(2400)는 복수의 프로세서로 구성될 수도 있다. 이 경우, 전용 프로세서들의 조합으로 구현될 수도 있고, AP, CPU 또는 GPU와 같은 다수의 범용 프로세서들과 소프트웨어의 조합을 통해 구현될 수도 있다.
- [245] 본 개시의 일 실시예에 있어서, 프로세서(2400)는, 인공지능(AI; artificial intelligence) 전용 프로세서를 포함할 수 있다. 인공지능 전용 프로세서는, 인공지능(AI)을 위한 전용 하드웨어 칩 형태로 제작될 수도 있고, 또는 기존의 범용 프로세서(예: CPU 또는 application processor) 또는 그래픽 전용 프로세서(예: GPU)의 일부로 제작되어 사용자 장치(2000)에 탑재될 수도 있다. 인공지능 전용 프로세서는 제1 생성 모델(1100) 제2 생성 모델(1200), 인코더(1300), 및 디코더(1400) 중 적어도 하나와 관련한 추론 및/또는 학습 연산을 수행할 수 있다.
- [246] 본 개시의 일 실시예에 있어서, 프로세서(2400)는 제1 생성 모델(1100) 및/또는 제2 생성 모델(1200)을 이용하여, 이미지의 기 지정된 영역을 제외한 이미지 정보에 기초하여 이미지의 기 지정된 영역을 추론할 수 있다. 여기서, 기 지정된 영역은 미지 영역을 의미할 수 있다.
- [247] 본 개시의 일 실시예에 있어서, 프로세서(2400)는 메모리(2500)에 저장된 학습 데이터셋을 이용하여 제1 생성 모델(1100), 제2 생성 모델(1200), 인코더(1300), 및 디코더(1400) 중 적어도 하나를 학습시킬 수 있다. 프로세서(2400)는 학습된 제1 생성 모델(1100) 제2 생성 모델(1200), 인코더(1300), 및/또는 디코더(1400)를 메모리(2500)에 저장할 수 있다.
- [248] 메모리(2500)는, 프로세서(2400)의 처리를 위한 프로그램을 저장할 수도 있고, 입/출력되는 데이터들을 저장할 수도 있다. 본 개시의 일 실시예에 있어

서, 메모리(2500)는 플래시 메모리 타입(flash memory type), 하드디스크 타입(hard disk type), 멀티미디어 카드 마이크로 타입(multimedia card micro type), 카드 타입의 메모리(예를 들어 SD 또는 XD 메모리 등), 램(RAM, Random Access Memory) SRAM(Static Random Access Memory), 롬(ROM, Read-Only Memory), EEPROM(Electrically Erasable Programmable Read-Only Memory), PROM(Programmable Read-Only Memory), 자기 메모리, 자기 디스크, 광디스크 중 적어도 하나의 타입의 저장매체를 포함할 수 있다. 메모리(2500)에 저장된 프로그램들은 그 기능에 따라 복수 개의 모듈들로 분류할 수 있다.

- [249] 본 개시의 일 실시예에 있어서, 메모리(2500)는 제1 생성 모델(1100), 제2 생성 모델(1200), 인코더(1300), 디코더(1400), 노이즈 생성기(1500), 및 디노이징 강도 결정기(1550)를 포함할 수 있다.
- [250] 제1 생성 모델(1100), 제2 생성 모델(1200), 디코더(1400), 노이즈 생성기(1500), 및 디노이징 강도 결정기(1550)의 구성, 기능, 및 동작은 도 1 내지 11의 제1 생성 모델(1100), 제2 생성 모델(1200), 디코더(1400), 노이즈 생성기(1500), 및 디노이징 강도 결정기(1550)의 구성, 기능, 및 동작에 대응할 수 있다. 인코더(1300)는 복수의 인코더들을 포함할 수 있다. 복수의 인코더들의 구성, 기능, 및 동작은 도 1 내지 13b의 제1 인코더 내지 제7 인코더(1310, 1320, 1330, 1340, 1350, 21, 22)의 구성, 기능, 및 동작에 대응할 수 있다. 복수의 인코더들 각각은 서로 다른 오토인코더의 인코더 부분일 수 있다.
- [251] 도 23은 본 개시의 일 실시예에 따른 사용자 장치 및 서버 장치의 구성을 설명하기 위한 블록도이다. 도 1 내지 21에서 설명한 내용과 중복되는 내용은 생략한다. 도 1 내지 11의 전자 장치(1000)의 구성, 기능, 및 동작은 도 23의 서버 장치(3000)의 구성, 기능, 및 동작에 대응할 수 있다. 도 22의 사용자 장치(2000)의 구성, 기능, 및 동작은 도 23의 사용자 장치(2000)의 구성, 기능, 및 동작에 대응할 수 있다.
- [252] 도 23을 참조하면, 서버 장치(3000)는 통신 인터페이스(3100), 프로세서(3200), 및 메모리(3300)를 포함할 수 있다. 그러나 도시된 구성요소 모두가 필수구성요소인 것은 아니다. 도시된 구성요소보다 많은 구성요소에 의해 서버 장치(3000)가 구현될 수도 있고, 그보다 적은 구성요소에 의해서도 서버 장치(3000)는 구현될 수 있다.
- [253] 통신 인터페이스(3100)는 서버 장치(3000)와 사용자 장치(2000), 서버 장치(3000)와 임의의 전자 장치(미도시), 및 서버 장치(3000)와 외부 서버 장치(미도시) 간의 통신을 수행하는 하나 이상의 구성요소를 포함할 수 있다.
- [254] 본 개시의 일 실시예에 있어서, 서버 장치(3000)는 통신 인터페이스(3100)를 통해 사용자 장치(2000)로부터 일부 영역(또는 마스킹된 영역)을 포함하는 이미지를 수신할 수 있다. 서버 장치(3000)는 통신 인터페이스(3100)를 통해 사용자 장치(2000)로부터 마스크 맵을 수신할 수 있다. 서버 장치(3000)는 통신 인터페이스(3100)를 통해 사용자 장치(2000)로부터 텍스트 입력을 수신할 수 있다. 서버 장치(3000)는 통신 인터페이스(3100)를 통해 사용자 장치(2000)로부터 최종 생성

이미지를 추론하기 위해 필요한 다양한 하이퍼파라미터들(예컨대, 디노이징 총 차수 등)을 수신할 수 있다.

- [255] 프로세서(3200)는, 메모리(3300)에 저장된 프로그램 또는 정보를 이용하여 서버 장치(3000)의 전반적인 동작을 제어할 수 있다. 프로세서(3200)는 AP(application processor), CPU(central processing unit) 또는 GPU(graphic processing unit)와 같은 범용 프로세서와 소프트웨어의 조합을 통해 구현될 수도 있다. 전용 프로세서의 경우, 본 개시의 실시예를 구현하기 위한 메모리를 포함하거나, 외부 메모리를 이용하기 위한 메모리 처리부를 포함할 수 있다. 프로세서(3200)는 복수의 프로세서로 구성될 수도 있다. 이 경우, 전용 프로세서들의 조합으로 구현될 수도 있고, AP, CPU 또는 GPU와 같은 다수의 범용 프로세서들과 소프트웨어의 조합을 통해 구현될 수도 있다.
- [256] 본 개시의 일 실시예에 있어서, 프로세서(3200)는, 인공지능(AI; artificial intelligence) 전용 프로세서를 포함할 수 있다. 인공지능 전용 프로세서는, 인공지능(AI)을 위한 전용 하드웨어 칩 형태로 제작될 수도 있고, 또는 기존의 범용 프로세서(예: CPU 또는 application processor) 또는 그래픽 전용 프로세서(예: GPU)의 일부로 제작되어 서버 장치(3000)에 탑재될 수도 있다. 인공지능 전용 프로세서는 제1 생성 모델(1100) 제2 생성 모델(1200), 인코더(1300), 및 디코더(1400) 중 적어도 하나와 관련한 추론 및/또는 학습 연산을 수행할 수 있다.
- [257] 본 개시의 일 실시예에 있어서, 프로세서(3200)는 통신 인터페이스(3100)를 통해 사용자 장치(2000)로부터 이미지 및 이미지의 일부 영역 생성을 위한 요청 신호를 수신할 수 있다. 프로세서(3200)는 요청 신호에 응답하여 이미지를 학습된 제1 생성 모델(1100) 및/또는 제2 생성 모델(1200)에 입력함으로써 이미지의 일부 영역이 생성된 최종 생성 이미지를 추론할 수 있다. 프로세서(3200)는 통신 인터페이스(3100)를 통해 최종 생성 이미지를 사용자 장치(2000)에 전달할 수 있다. 사용자 장치(2000)는 최종 생성 이미지를 수신할 수 있다. 사용자 장치(2000)는 사용자 인터페이스(2200)를 통해 최종 생성 이미지를 표시할 수 있다.
- [258] 본 개시의 일 실시예에 있어서, 프로세서(3200)는 제1 생성 모델(1100) 및/또는 제2 생성 모델(1200)을 이용하여, 이미지의 기 지정된 영역을 제외한 이미지 정보에 기초하여 이미지의 기 지정된 영역을 추론할 수 있다.
- [259] 본 개시의 일 실시예에 있어서, 프로세서(3200)는 메모리(3300)에 저장된 학습 데이터셋을 이용하여 제1 생성 모델(1100), 제2 생성 모델(1200), 인코더(1300), 및 디코더(1400) 중 적어도 하나를 학습시킬 수 있다. 프로세서(2400)는 학습된 제1 생성 모델(1100) 제2 생성 모델(1200), 인코더(1300), 및/또는 디코더(1400)를 메모리(3300)에 저장할 수 있다.
- [260] 메모리(3300)는, 프로세서(3200)의 처리를 위한 프로그램을 저장할 수도 있고, 입/출력되는 데이터들을 저장할 수도 있다. 본 개시의 일 실시예에 있어서, 메모리(2500)는 플래시 메모리 타입(flash memory type), 하드디스크 타입(hard disk type), 멀티미디어 카드 마이크로 타입(multimedia card micro type),

카드 타입의 메모리(예를 들어 SD 또는 XD 메모리 등), 램(RAM, Random Access Memory) SRAM(Static Random Access Memory), 롬(ROM, Read-Only Memory), EEPROM(Electrically Erasable Programmable Read-Only Memory), PROM(Programmable Read-Only Memory), 자기 메모리, 자기 디스크, 광디스크 중 적어도 하나의 타입의 저장매체를 포함할 수 있다. 메모리(2500)에 저장된 프로그램들은 그 기능에 따라 복수 개의 모듈들로 분류할 수 있다.

- [261] 본 개시의 일 실시예에 있어서, 메모리(3300)는 제1 생성 모델(1100), 제2 생성 모델(1200), 인코더(1300), 디코더(1400), 노이즈 생성기(1500), 및 디노이징 강도 결정기(1550)를 포함할 수 있다.
- [262] 제1 생성 모델(1100), 제2 생성 모델(1200), 디코더(1400), 노이즈 생성기(1500), 및 디노이징 강도 결정기(1550)의 구성, 기능, 및 동작은 도 1 내지 11의 제1 생성 모델(1100), 제2 생성 모델(1200), 디코더(1400), 노이즈 생성기(1500), 및 디노이징 강도 결정기(1550)의 구성, 기능, 및 동작에 대응할 수 있다. 인코더(1300)는 복수의 인코더들을 포함할 수 있다. 복수의 인코더들의 구성, 기능, 및 동작은 도 1 내지 13b의 제1 인코더 내지 제7 인코더(1310, 1320, 1330, 1340, 1350, 21, 22)의 구성, 기능, 및 동작에 대응할 수 있다. 복수의 인코더들 각각은 서로 다른 오토인코더의 인코더 부분일 수 있다.
- [263] 도 23에 도시된 바와 달리, 서버 장치(3000)의 기능의 적어도 일부는 사용자 장치(2000)에서 수행될 수 있다. 예를 들어, 사용자 장치(2000)는 메모리(2500)에 저장된 제1 생성 모델(1100)을 이용하여 중간 생성 이미지를 추론할 수 있다. 서버 장치(3000)는 통신 인터페이스(3100)를 통해 중간 생성 이미지를 수신할 수 있다. 서버 장치(3000)는 메모리(3300)에 저장된 제2 생성 모델(1200)을 이용하여 중간 생성 이미지에 기초하여 최종 생성 이미지를 추론할 수 있다. 그러나 이는 일 예시일 뿐, 제1 생성 모델(1100), 제2 생성 모델(1200), 인코더(1300), 디코더(1400), 노이즈 생성기(1500), 및 디노이징 강도 결정기(1550) 중 적어도 일부는 사용자 장치(2000)에서 실행되거나, 서버 장치(3000)에서 실행될 수 있다.
- [264] 본 개시의 일 실시예에 있어서, 생성 모델을 이용하여 이미지의 일부 영역을 생성하는 방법이 제공될 수 있다. 상기 방법은, 상기 일부 영역의 정보를 포함하는 이미지를 획득하는 단계를 포함할 수 있다. 상기 방법은, 상기 일부 영역의 정보를 포함하는 이미지를 입력으로 하는 제1 생성 모델을 이용하여, 상기 일부 영역에 대한 제1 이미지 정보를 포함하는 중간 생성 이미지를 획득하는 단계를 포함할 수 있다. 상기 방법은, 상기 일부 영역의 정보를 포함하는 이미지 및 상기 중간 생성 이미지를 입력으로 하는 제2 생성 모델을 이용하여, 상기 제1 이미지 정보와 적어도 일부가 상이한 제2 이미지 정보를 포함하는 최종 생성 이미지를 획득하는 단계를 포함할 수 있다.
- [265] 본 개시의 일 실시예에 있어서, 상기 일부 영역의 정보를 포함하는 이미지를 획득하는 단계는, 상기 일부 영역의 정보를 포함하는 이미지의 전체 영역에서 상기 일부 영역을 구분하는 마스크 맵(mask map)을 획득하는 단계를 포함할 수 있다.

상기 일부 영역의 정보를 포함하는 이미지를 획득하는 단계는, 상기 일부 영역의 정보를 포함하는 이미지에 상기 마스크 맵을 연결(concatenating)하는 단계를 포함할 수 있다.

- [266] 본 개시의 일 실시예에 있어서, 상기 최종 생성 이미지를 획득하는 단계는, 상기 중간 생성 이미지를 인코딩하는 단계를 포함할 수 있다. 상기 최종 생성 이미지를 획득하는 단계는, 상기 일부 영역의 정보를 포함하는 이미지 및 상기 인코딩된 중간 생성 이미지를 입력으로 하는 제2 생성 모델을 이용하여, 상기 제1 이미지 정보와 적어도 일부가 상이한 제2 이미지 정보를 포함하는 최종 생성 이미지를 획득하는 단계를 포함할 수 있다.
- [267] 본 개시의 일 실시예에 있어서, 상기 최종 생성 이미지를 획득하는 단계는, 상기 일부 영역의 정보를 포함하는 이미지, 상기 인코딩된 중간 생성 이미지, 및 상기 중간 생성 이미지를 입력으로 하는 제2 생성 모델을 이용하여, 상기 제1 이미지 정보와 적어도 일부가 상이한 제2 이미지 정보를 포함하는 최종 생성 이미지를 획득하는 단계를 포함할 수 있다.
- [268] 본 개시의 일 실시예에 있어서, 상기 최종 생성 이미지를 획득하는 단계는, 텍스트 입력을 획득하는 단계를 포함할 수 있다. 상기 최종 생성 이미지를 획득하는 단계는, 상기 텍스트 입력을 인코딩하는 단계를 포함할 수 있다. 상기 최종 생성 이미지를 획득하는 단계는, 상기 인코딩된 텍스트 입력, 상기 일부 영역의 정보를 포함하는 이미지, 및 상기 중간 생성 이미지를 입력으로 하는 제2 생성 모델을 이용하여, 상기 제1 이미지 정보와 적어도 일부가 상이한 제2 이미지 정보를 포함하는 최종 생성 이미지를 획득하는 단계를 포함할 수 있다.
- [269] 본 개시의 일 실시예에 있어서, 상기 최종 생성 이미지를 획득하는 단계는, 상기 인코딩된 텍스트 입력, 상기 일부 영역의 정보를 포함하는 이미지, 상기 중간 생성 이미지, 및 상기 인코딩된 중간 생성 이미지를 입력으로 하는 제2 생성 모델을 이용하여, 상기 제1 이미지 정보와 적어도 일부가 상이한 제2 이미지 정보를 포함하는 최종 생성 이미지를 획득하는 단계를 포함할 수 있다.
- [270] 본 개시의 일 실시예에 있어서, 상기 최종 생성 이미지를 획득하는 단계는, 상기 중간 생성 이미지에 대한 디노이징 강도를 획득하는 단계를 포함할 수 있다. 상기 최종 생성 이미지를 획득하는 단계는, 상기 디노이징 강도에 기초하여 상기 중간 생성 이미지에 노이즈를 추가하는 단계를 포함할 수 있다. 상기 최종 생성 이미지를 획득하는 단계는, 상기 일부 영역의 정보를 포함하는 이미지 및 상기 노이즈가 추가된 중간 생성 이미지를 입력으로 하는 제2 생성 모델을 이용하여, 상기 제1 이미지 정보와 적어도 일부가 상이한 제2 이미지 정보를 포함하는 최종 생성 이미지를 획득하는 단계를 포함할 수 있다.
- [271] 본 개시의 일 실시예에 있어서, 상기 중간 생성 이미지에 대한 디노이징 강도를 획득하는 단계는, 상기 중간 생성 이미지에 기초하여 예측 신뢰 값을 획득하는 단계를 포함할 수 있다. 상기 중간 생성 이미지에 대한 디노이징 강도를 획득하는 단계는, 상기 예측 신뢰 값, 상기 일부 영역의 크기, 및 상기 일부 영역의 모양

중 적어도 하나에 기초하여 상기 디노이징 강도를 결정하는 단계를 포함할 수 있다.

- [272] 본 개시의 일 실시예에 있어서, 상기 최종 생성 이미지를 획득하는 단계는, 현재 노이즈 정보를 획득하는 단계를 포함할 수 있다. 상기 최종 생성 이미지를 획득하는 단계는, 상기 현재 노이즈 정보 및 상기 일부 영역의 정보를 포함하는 이미지를 연결하는 단계를 포함할 수 있다. 상기 최종 생성 이미지를 획득하는 단계는, 상기 연결된 이미지를 상기 제2 생성 모델에 입력하는 단계 포함할 수 있다. 상기 최종 생성 이미지를 획득하는 단계는, 상기 제2 생성 모델로부터 다음 노이즈 정보를 획득하는 단계를 포함할 수 있다.
- [273] 본 개시의 일 실시예에 있어서, 상기 현재 노이즈 정보는 상기 노이즈가 추가된 중간 생성 이미지에 대응할 수 있다.
- [274] 본 개시의 일 실시예에 있어서, 상기 최종 생성 이미지를 획득하는 단계는, 상기 디노이징 강도에 기초하여, 기 정의된 디노이징 총 차수 중 상기 노이즈가 추가된 중간 생성 이미지에 대응하는 대상 디노이징 차수를 결정하는 단계를 포함할 수 있다. 상기 최종 생성 이미지를 획득하는 단계는, 상기 현재 노이즈 정보의 디노이징 차수를 상기 결정된 대상 디노이징 차수로 설정하는 단계를 포함할 수 있다.
- [275] 본 개시의 일 실시예에 있어서, 상기 제1 생성 모델은 GAN(generative adversarial network) 모델일 수 있다. 본 개시의 일 실시예에 있어서, 상기 제2 생성 모델은 확산(diffusion) 모델일 수 있다.
- [276] 본 개시의 일 실시예에 있어서, 전자 장치가 제공될 수 있다. 상기 전자 장치는, 적어도 하나의 인스트럭션을 저장하는 메모리를 포함할 수 있다. 상기 전자 장치는, 상기 적어도 하나의 인스트럭션을 실행하는 적어도 하나의 프로세서를 포함할 수 있다. 상기 적어도 하나의 프로세서는, 일부 영역의 정보를 포함하는 이미지를 획득할 수 있다. 상기 적어도 하나의 프로세서는, 상기 일부 영역의 정보를 포함하는 이미지를 입력으로 하는 제1 생성 모델을 이용하여, 상기 일부 영역에 대한 제1 이미지 정보를 포함하는 중간 생성 이미지를 획득할 수 있다. 상기 적어도 하나의 프로세서는, 상기 일부 영역의 정보를 포함하는 이미지 및 상기 중간 생성 이미지를 입력으로 하는 제2 생성 모델을 이용하여, 상기 제1 이미지 정보와 적어도 일부가 상이한 제2 이미지 정보를 포함하는 최종 생성 이미지를 획득할 수 있다.
- [277] 본 개시의 일 실시예에 있어서, 이미지를 인코딩하는 인코더를 학습하는 방법이 제공될 수 있다. 상기 방법은, 전체 영역 중 일부 영역에 대한 정보를 포함하는 제1 이미지, 전체 영역에 대한 이미지 정보를 포함하는 제2 이미지, 및 상기 제2 이미지를 나타내는 텍스트를 획득하는 단계를 포함할 수 있다. 상기 방법은, 상기 제1 이미지를 입력으로 하는 제1 인코더를 이용하여, 제1 이미지 임베딩(embedding)을 획득하는 단계를 포함할 수 있다. 상기 방법은, 상기 제2 이미지를 입력으로 하는 제2 인코더를 이용하여, 제2 이미지 임베딩을 획득하는 단계를

포함할 수 있다. 상기 방법은, 상기 텍스트를 입력으로 하는 제3 인코더를 이용하여, 텍스트 임베딩을 획득하는 단계를 포함할 수 있다. 상기 방법은, 상기 제1 이미지 임베딩과 상기 제2 이미지 임베딩에 기초하여 제1 손실을 획득하는 단계를 포함할 수 있다. 상기 방법은, 상기 제1 이미지 임베딩과 상기 텍스트 임베딩에 기초하여 제2 손실을 획득하는 단계를 포함할 수 있다. 상기 방법은, 상기 제1 손실 및 상기 제2 손실에 기초하여 상기 제1 인코더의 적어도 하나의 파라미터를 갱신하는 단계를 포함할 수 있다.

- [278] 본 개시의 일 실시예에 있어서, 상기 제2 인코더 및 상기 제3 인코더는 기 학습되어 고정된 파라미터를 갖는 인공지능 모델들일 수 있다.
- [279] 본 개시의 일 실시예에 있어서, 상기 방법은, 상기 제1 손실 및 상기 제2 손실에 기초하여, 상기 제2 인코더 및 상기 제3 인코더의 적어도 하나의 파라미터를 갱신하는 단계를 포함할 수 있다.
- [280] 본 개시의 일 실시예에 있어서, 상기 방법은, 상기 제1 이미지에, 상기 전체 영역에서 상기 일부 영역을 구분하는 제1 마스크 맵(mask map)을 연결하는 단계를 포함할 수 있다.
- [281] 본 개시의 일 실시예에 있어서, 상기 제1 인코더와 상기 제2 인코더는 동일한 인코더일 수 있다.
- [282] 본 개시의 일 실시예에 있어서, 상기 방법은, 상기 제2 이미지에, 상기 전체 영역을 하나의 값으로 구성한 제2 마스크 맵을 연결하는 단계를 포함할 수 있다.
- [283] 본 개시의 일 실시예에 있어서, 생성 모델을 이용하여 이미지의 일부 영역을 생성하는 방법이 제공될 수 있다. 상기 방법은, 상기 일부 영역의 정보를 포함하는 이미지를 획득하는 단계를 포함할 수 있다. 상기 방법은, 상기 일부 영역의 정보를 포함하는 이미지를 입력으로 하는 제1 인코더를 이용하여, 대상 이미지 임베딩을 획득하는 단계를 포함할 수 있다. 상기 방법은, 상기 일부 영역의 정보를 포함하는 이미지 및 상기 대상 이미지 임베딩을 입력으로 하는 생성 모델을 이용하여, 최종 생성 이미지를 획득하는 단계를 포함할 수 있다.
- [284] 본 개시의 일 실시예에 있어서, 상기 제1 인코더는, 상기 전체 영역 중 일부 영역에 대한 위치 정보를 포함하는 제1 학습 이미지, 전체 영역에 대한 이미지 정보를 포함하는 제2 학습 이미지, 및 상기 제2 학습 이미지를 나타내는 학습 텍스트를 획득하고, 상기 제1 학습 이미지를 입력으로 하는 상기 제1 인코더를 이용하여, 제1 이미지 임베딩(embedding)을 획득하고, 상기 제2 학습 이미지를 입력으로 하는 제2 인코더를 이용하여 제2 이미지 임베딩을 획득하고, 상기 학습 텍스트를 입력으로 하는 제3 인코더를 이용하여 텍스트 임베딩을 획득하고, 상기 제1 이미지 임베딩과 상기 제2 이미지 임베딩에 기초하여 제1 손실을 획득하고, 상기 제1 이미지 임베딩과 상기 텍스트 임베딩에 기초하여 제2 손실을 획득하고, 상기 제1 손실 및 상기 제2 손실에 기초하여 상기 제1 인코더의 적어도 하나의 파라미터를 갱신함으로써 학습될 수 있다.

- [285] 본 개시의 일 실시예에 있어서, 상기 제2 인코더 및 상기 제3 인코더는 기 학습되어 고정된 파라미터를 갖는 인공지능 모델들일 수 있다.
- [286] 본 개시의 일 실시예에 있어서, 상기 제1 학습 이미지는 상기 전체 영역에서 상기 일부 영역을 구분하는 제1 마스크 맵(mask map)에 연결된, 방법.
- [287] 본 개시의 일 실시예에 있어서, 상기 제1 인코더와 상기 제2 인코더는 동일한 인코더일 수 있다.
- [288] 본 개시의 일 실시예에 있어서, 상기 제2 학습 이미지는 상기 전체 영역을 하나의 값으로 구성한 제2 마스크 맵에 연결될 수 있다.
- [289] 본 개시의 일 실시예에 있어서, 상기 생성 모델은, 상기 일부 영역의 정보를 포함하는 이미지에 기초하여 상기 최종 생성 이미지를 출력하는 제1 뉴럴 네트워크를 포함할 수 있다.
- [290] 본 개시의 일 실시예에 있어서, 상기 방법은, 상기 대상 이미지 임베딩을 상기 제1 뉴럴 네트워크의 적어도 하나의 레이어에 입력하는 단계를 포함할 수 있다.
- [291] 본 개시의 일 실시예에 있어서, 상기 생성 모델은, 상기 대상 이미지 임베딩을 변환하여 상기 제1 뉴럴 네트워크의 적어도 하나의 레이어에 전달하는 인터프리터를 포함할 수 있다. 상기 생성 모델은, 상기 인터프리터는 적어도 하나의 단층 퍼셉트론으로 구성될 수 있다.
- [292] 본 개시의 일 실시예에 있어서, 적어도 하나의 단층 퍼셉트론은 제1 단층 퍼셉트론 및 제2 단층 퍼셉트론을 포함할 수 있다. 본 개시의 일 실시예에 있어서, 상기 인터프리터는 3개 이상의 단층 퍼셉트론들로 구성될 수 있다. 본 개시의 일 실시예에 있어서, 상기 인터프리터는 유일한 단층 퍼셉트론으로 구성될 수 있다.
- [293] 본 개시의 일 실시예에 있어서, 상기 방법은, 상기 대상 이미지 임베딩을 입력으로 하는 상기 제1 단층 퍼셉트론을 이용하여 중간 이미지 임베딩을 획득하는 단계를 포함할 수 있다. 상기 방법은, 상기 중간 이미지 임베딩을 입력으로 하는 상기 제2 단층 퍼셉트론을 이용하여 결과 이미지 임베딩을 획득하는 단계를 포함할 수 있다. 상기 방법은, 상기 결과 이미지 임베딩을 상기 제1 뉴럴 네트워크의 적어도 하나의 레이어에 입력하는 단계를 포함할 수 있다.
- [294] 본 개시의 일 실시예에 있어서, 상기 방법은, 이전 단층 퍼셉트론이 출력한 이미지 임베딩을 입력으로 하는 다음 단층 퍼셉트론이 다음 이미지 임베딩을 출력하는 단계를 포함할 수 있다. 상기 방법은, 상기 이전 단층 퍼셉트론이 출력한 이미지 임베딩을 입력으로 하는 다음 단층 퍼셉트론이 다음 이미지 임베딩을 출력하는 단계를 반복할 수 있다. 상기 방법은, 상기 적어도 하나의 단층 퍼셉트론 중 마지막 단층 퍼셉트론을 이용하여 결과 이미지 임베딩을 획득하는 단계를 포함할 수 있다. 상기 방법은, 상기 결과 이미지 임베딩을 상기 제1 뉴럴 네트워크의 적어도 하나의 레이어에 입력하는 단계를 포함할 수 있다.
- [295] 본 개시의 일 실시예에 있어서, 적어도 하나의 단층 퍼셉트론은 유일한 단층 퍼셉트론을 포함할 수 있다. 상기 방법은, 유일한 단층 퍼셉트론을 이용하여 결과 이미지 임베딩을 획득하는 단계를 포함할 수 있다. 상기 방법은, 상기 결과 이미

지 임베딩을 상기 제1 뉴럴 네트워크의 적어도 하나의 레이어에 입력하는 단계를 포함할 수 있다.

- [296] 본 개시의 일 실시예에 있어서, 상기 생성 모델은, 상기 일부 영역의 정보를 포함하는 이미지에 기초하여 제1 노이즈를 출력하는 제1 뉴럴 네트워크를 포함할 수 있다. 상기 생성 모델은, 상기 일부 영역의 정보를 포함하는 이미지에 기초하여 제2 노이즈를 출력하는 제2 뉴럴 네트워크를 포함할 수 있다. 상기 생성 모델은, 상기 대상 이미지 임베딩을 변환하여 상기 제1 뉴럴 네트워크의 적어도 하나의 레이어에 전달하는 인터프리터를 포함할 수 있다.
- [297] 본 개시의 일 실시예에 있어서, 제1 뉴럴 네트워크는 상기 일부 영역의 정보를 포함하는 이미지 및 상기 대상 이미지 임베딩을 입력으로 하여 제1 노이즈를 출력할 수 있다.
- [298] 본 개시의 일 실시예에 있어서, 제1 뉴럴 네트워크는 상기 일부 영역의 정보를 포함하는 이미지, 상기 대상 이미지 임베딩, 및 중간 생성 이미지를 입력으로 하여 제1 노이즈를 출력할 수 있다.
- [299] 본 개시의 일 실시예에 있어서, 상기 방법은, 상기 일부 영역의 정보를 포함하는 이미지에 대응하는 대상 텍스트를 획득하는 단계를 포함할 수 있다. 상기 방법은, 상기 대상 텍스트에 기초하여 대상 텍스트 임베딩을 획득하는 단계를 포함할 수 있다. 상기 방법은, 상기 대상 텍스트 임베딩을 상기 제2 뉴럴 네트워크의 적어도 하나의 레이어에 입력하는 단계를 포함할 수 있다.
- [300] 본 개시의 일 실시예에 있어서, 제2 뉴럴 네트워크는 상기 일부 영역의 정보를 포함하는 이미지 및 대상 텍스트 임베딩을 입력으로 하여 제2 노이즈를 출력할 수 있다.
- [301] 본 개시의 일 실시예에 있어서, 제2 뉴럴 네트워크는 상기 일부 영역의 정보를 포함하는 이미지, 대상 텍스트 임베딩, 및 중간 생성 이미지를 입력으로 하여 제2 노이즈를 출력할 수 있다.
- [302] 본 개시의 일 실시예에 있어서, 상기 방법은, 상기 제1 노이즈 및 상기 제2 노이즈에 기초하여 상기 최종 생성 이미지를 획득하는 단계를 포함할 수 있다.
- [303] 본 개시의 일 실시예에 있어서, 전자 장치가 제공될 수 있다. 전자 장치는, 적어도 하나의 인스트럭션을 저장하는 메모리를 포함할 수 있다. 전자 장치는, 상기 적어도 하나의 인스트럭션을 실행하는 적어도 하나의 프로세서를 포함할 수 있다. 상기 적어도 하나의 프로세서는, 전체 영역 중 일부 영역에 대한 정보를 포함하는 제1 이미지, 전체 영역에 대한 이미지 정보를 포함하는 제2 이미지, 및 상기 제2 이미지를 나타내는 텍스트를 획득하고, 상기 적어도 하나의 프로세서는, 상기 제1 이미지를 입력으로 하는 제1 인코더를 이용하여, 제1 이미지 임베딩 (embedding)을 획득하고, 상기 적어도 하나의 프로세서는, 상기 제2 이미지를 입력으로 하는 제2 인코더를 이용하여, 제2 이미지 임베딩을 획득하고, 상기 적어도 하나의 프로세서는, 상기 텍스트를 입력으로 하는 제3 인코더를 이용하여, 텍스트 임베딩을 획득하고, 상기 적어도 하나의 프로세서는, 상기 제1 이미지 임베

당과 상기 제2 이미지 임베딩에 기초하여 제1 손실을 획득하고, 상기 적어도 하나의 프로세서는, 상기 제1 이미지 임베딩과 상기 텍스트 임베딩에 기초하여 제2 손실을 획득하고, 상기 적어도 하나의 프로세서는, 상기 제1 손실 및 상기 제2 손실에 기초하여 상기 제1 인코더의 적어도 하나의 파라미터를 갱신하는, 전자 장치.

- [304] 본 개시의 일 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 매체에 기록되는 프로그램 명령은 본 개시를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.
- [305] 본 개시의 일부 실시예는 컴퓨터에 의해 실행되는 프로그램 모듈과 같은 컴퓨터에 의해 실행가능한 명령어를 포함하는 기록 매체의 형태로도 구현될 수 있다. 컴퓨터 판독 가능 매체는 컴퓨터에 의해 액세스될 수 있는 임의의 가용 매체일 수 있고, 휘발성 및 비휘발성 매체, 분리형 및 비분리형 매체를 모두 포함한다. 또한, 컴퓨터 판독가능 매체는 컴퓨터 저장 매체 및 통신 매체를 모두 포함할 수 있다. 컴퓨터 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현된 휘발성 및 비휘발성, 분리형 및 비분리형 매체를 모두 포함한다. 통신 매체는 전형적으로 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈, 또는 반송파와 같은 변조된 데이터 신호의 기타 데이터, 또는 기타 전송 매커니즘을 포함하며, 임의의 정보 전달 매체를 포함한다. 또한, 본 개시의 일부 실시예는 컴퓨터에 의해 실행되는 컴퓨터 프로그램과 같은 컴퓨터에 의해 실행가능한 명령어를 포함하는 컴퓨터 프로그램 또는 컴퓨터 프로그램 제품(computer program product)으로도 구현될 수 있다.
- [306] 본 개시의 일 실시예에 있어서, 기기로 읽을 수 있는 저장매체는, 비일시적(non-transitory) 저장매체의 형태로 제공될 수 있다. 여기서, '비일시적 저장매체'는 실재(tangible)하는 장치이고, 신호(signal)(예: 전자기파)를 포함하지 않는다는 것을 의미할 뿐이며, 이 용어는 데이터가 저장매체에 반영구적으로 저장되는 경우와 일시적으로 저장되는 경우를 구분하지 않는다. 예로, '비일시적 저장매체'는 데이터가 일시적으로 저장되는 버퍼를 포함할 수 있다.

- [307] 본 개시의 실시예에 따르면, 본 문서에 개시된 다양한 실시예들에 따른 방법은 컴퓨터 프로그램 제품(computer program product)에 포함되어 제공될 수 있다. 컴퓨터 프로그램 제품은 상품으로서 판매자 및 구매자 간에 거래될 수 있다. 컴퓨터 프로그램 제품은 기기로 읽을 수 있는 저장 매체(예: compact disc read only memory (CD-ROM))의 형태로 배포되거나, 또는 어플리케이션 스토어를 통해 또는 두개의 사용자 장치들(예: 스마트폰들) 간에 직접, 온라인으로 배포(예: 다운로드 또는 업로드)될 수 있다. 온라인 배포의 경우에, 컴퓨터 프로그램 제품(예: 다운로드 가능한 앱(downloadable app))의 적어도 일부는 제조사의 서버, 어플리케이션 스토어의 서버, 또는 중계 서버의 메모리와 같은 기기로 읽을 수 있는 저장 매체에 적어도 일시 저장되거나, 임시적으로 생성될 수 있다.
- [308] 이상에서 본 개시의 실시예들에 대하여 상세하게 설명하였지만 본 개시의 권리 범위는 이에 한정되는 것은 아니고 다음의 청구범위에서 정의하고 있는 본 개시의 기본 개념을 이용한 당업자의 여러 변형 및 개량 형태 또한 본 개시의 권리 범위에 속한다.

## 청구범위

- [청구항 1] 생성 모델을 이용하여 이미지의 일부 영역을 생성하는 방법에 있어서, 상기 일부 영역의 정보를 포함하는 이미지를 획득하는 단계; 상기 이미지를 제1 생성 모델(1100)에 입력함으로써, 상기 일부 영역에 대한 제1 이미지 정보를 포함하는 중간 생성 이미지를 획득하는 단계; 및 상기 이미지 및 상기 중간 생성 이미지를 제2 생성 모델(1200)에 입력함으로써, 상기 제1 이미지 정보와 적어도 일부가 상이한 제2 이미지 정보를 포함하는 최종 생성 이미지를 획득하는 단계를 포함하는, 방법.
- [청구항 2] 제1항에 있어서, 상기 일부 영역의 정보를 포함하는 이미지를 획득하는 단계는: 상기 이미지의 전체 영역에서 상기 일부 영역을 구분하는 마스크 맵(mask map)을 획득하는 단계; 및 상기 이미지에 상기 마스크 맵을 연결(concatenating)하는 단계를 더 포함하는, 방법.
- [청구항 3] 제1항 및 제2항 중 어느 한 항에 있어서, 상기 최종 생성 이미지를 획득하는 단계는: 상기 중간 생성 이미지를 인코딩하는 단계; 및 상기 이미지 및 상기 인코딩된 중간 생성 이미지를 상기 제2 생성 모델에 입력함으로써, 상기 최종 생성 이미지를 획득하는 단계를 포함하는, 방법.
- [청구항 4] 제1항 내지 제3항 중 어느 한 항에 있어서, 상기 최종 생성 이미지를 획득하는 단계는: 텍스트 입력을 획득하는 단계; 상기 텍스트 입력을 인코딩하는 단계; 및 상기 인코딩된 텍스트 입력, 상기 이미지, 및 상기 중간 생성 이미지를 상기 제2 생성 모델에 입력함으로써, 상기 최종 생성 이미지를 획득하는 단계를 포함하는, 방법.
- [청구항 5] 제1항 내지 제4항 중 어느 한 항에 있어서, 상기 최종 생성 이미지를 획득하는 단계는: 상기 중간 생성 이미지에 대한 디노이징 강도를 획득하는 단계; 상기 디노이징 강도에 기초하여 상기 중간 생성 이미지에 노이즈를 추가하는 단계; 및 상기 이미지 및 상기 노이즈가 추가된 중간 생성 이미지를 상기 제2 생성 모델에 입력함으로써, 상기 최종 생성 이미지를 획득하는 단계를 포함하는, 방법.
- [청구항 6] 제1항 내지 제5항 중 어느 한 항에 있어서, 상기 중간 생성 이미지에 대한 디노이징 강도를 획득하는 단계는: 상기 중간 생성 이미지에 기초하여 예측 신뢰 값을 획득하는 단계; 및

상기 예측 신뢰 값, 상기 일부 영역의 크기, 및 상기 일부 영역의 모양 중 적어도 하나에 기초하여 상기 디노이징 강도를 결정하는 단계를 포함하는, 방법.

- [청구항 7] 제1항 내지 제6항 중 어느 한 항에 있어서, 상기 최종 생성 이미지를 획득하는 단계는: 현재 노이즈 정보를 획득하는 단계; 상기 현재 노이즈 정보 및 상기 이미지를 연결하는 단계; 상기 연결된 이미지를 상기 제2 생성 모델에 입력하는 단계; 및 상기 제2 생성 모델로부터 다음 노이즈 정보를 획득하는 단계를 포함하는, 방법.
- [청구항 8] 제1항 내지 제7항 중 어느 한 항에 있어서, 상기 최종 생성 이미지를 획득하는 단계는: 상기 디노이징 강도에 기초하여, 기 정의된 디노이징 총 차수 중 상기 노이즈가 추가된 중간 생성 이미지에 대응하는 대상 디노이징 차수를 결정하는 단계; 및 상기 현재 노이즈 정보의 디노이징 차수를 상기 결정된 대상 디노이징 차수로 설정하는 단계를 포함하는, 방법.
- [청구항 9] 제1항 내지 제8항 중 어느 한 항에 있어서, 상기 제1 생성 모델은 GAN(generative adversarial network) 모델이고, 상기 제2 생성 모델은 확산(diffusion) 모델인, 방법.
- [청구항 10] 적어도 하나의 인스트럭션을 저장하는 메모리(2500; 3300); 및 적어도 하나의 프로세서(2400; 3200)를 포함하고, 상기 적어도 하나의 프로세서(2400; 3200)가 상기 메모리에 저장된 적어도 하나의 인스트럭션을 실행함으로써 전자 장치는, 일부 영역의 정보를 포함하는 이미지를 획득하고, 상기 이미지를 제1 생성 모델(1100)에 입력함으로써, 상기 일부 영역에 대한 제1 이미지 정보를 포함하는 중간 생성 이미지를 획득하고, 상기 이미지 및 상기 중간 생성 이미지를 제2 생성 모델(1200)에 입력함으로써, 상기 제1 이미지 정보와 적어도 일부가 상이한 제2 이미지 정보를 포함하는 최종 생성 이미지를 획득하는, 전자 장치.
- [청구항 11] 제10항에 있어서, 상기 적어도 하나의 프로세서(2400; 3200)가 상기 메모리에 저장된 적어도 하나의 인스트럭션을 실행함으로써 전자 장치는: 상기 이미지의 전체 영역에서 상기 일부 영역을 구분하는 마스크 맵(mask map)을 획득하고, 상기 이미지에 상기 마스크 맵을 연결(concatenating)하는, 전자 장치.
- [청구항 12] 제10항 및 제11항 중 어느 한 항에 있어서,

상기 적어도 하나의 프로세서(2400; 3200)가 상기 메모리에 저장된 적어도 하나의 인스트럭션을 실행함으로써 전자 장치는:

상기 중간 생성 이미지를 인코딩하고,

상기 이미지 및 상기 인코딩된 중간 생성 이미지를 상기 제2 생성 모델에 입력함으로써, 상기 최종 생성 이미지를 획득하는, 전자 장치.

[청구항 13]

제10항 내지 제12항 중 어느 한 항에 있어서,

상기 적어도 하나의 프로세서(2400; 3200)가 상기 메모리에 저장된 적어도 하나의 인스트럭션을 실행함으로써 전자 장치는:

텍스트 입력을 획득하고,

상기 텍스트 입력을 인코딩하고,

상기 인코딩된 텍스트 입력, 상기 이미지, 및 상기 중간 생성 이미지를 상기 제2 생성 모델에 입력함으로써, 상기 최종 생성 이미지를 획득하는, 전자 장치.

[청구항 14]

제10항 내지 제13항 중 어느 한 항에 있어서,

상기 적어도 하나의 프로세서(2400; 3200)가 상기 메모리에 저장된 적어도 하나의 인스트럭션을 실행함으로써 전자 장치는:

상기 중간 생성 이미지에 대한 디노이징 강도를 획득하고,

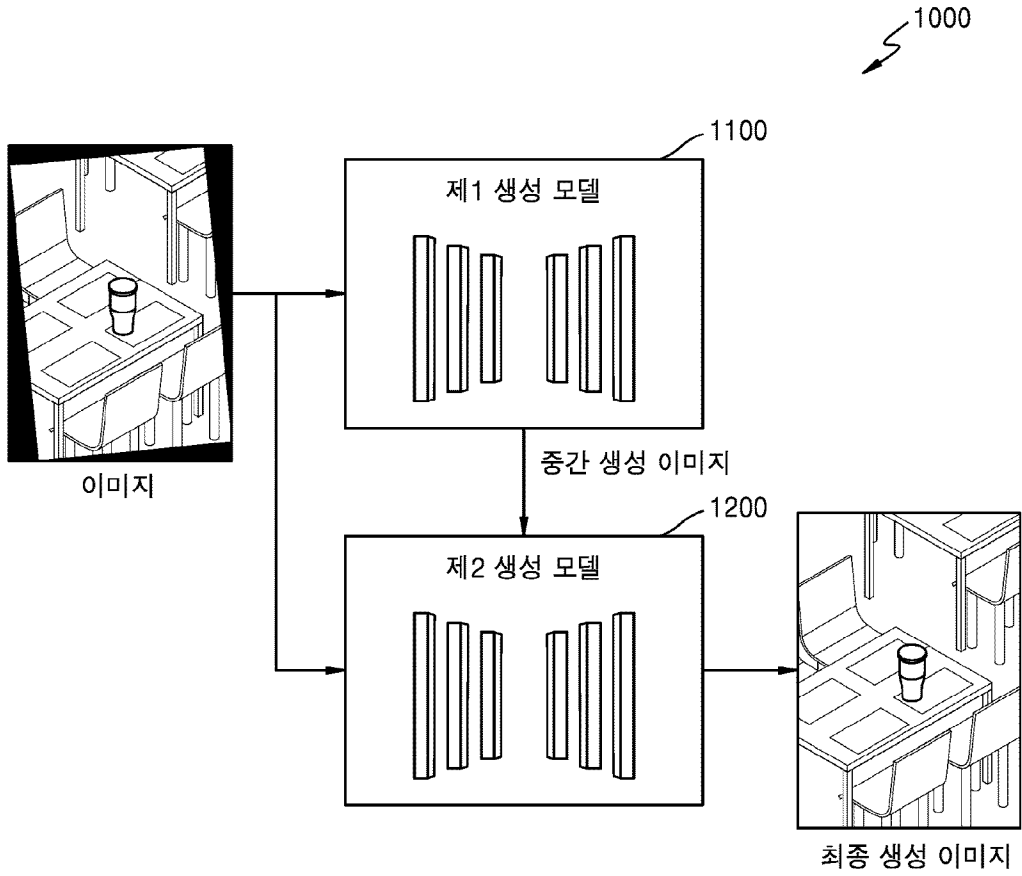
상기 디노이징 강도에 기초하여 상기 중간 생성 이미지에 노이즈를 추가하고,

상기 이미지 및 상기 노이즈가 추가된 중간 생성 이미지를 상기 제2 생성 모델에 입력함으로써, 상기 최종 생성 이미지를 획득하는 전자 장치.

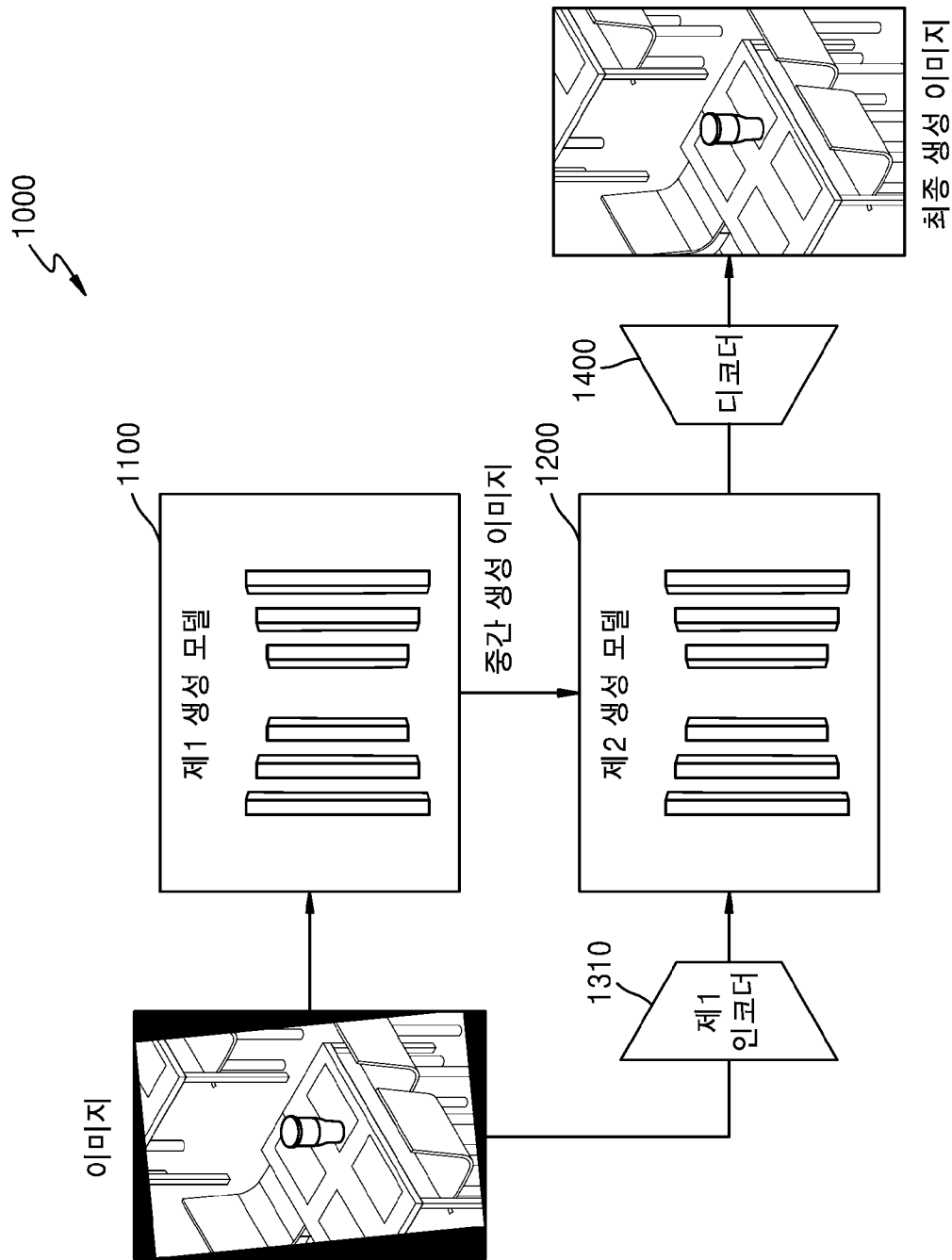
[청구항 15]

제1항 내지 제9항 중 어느 한 항의 방법을 컴퓨터에서 수행하기 위한 프로그램이 기록된 컴퓨터로 읽을 수 있는 기록매체.

[도 1]

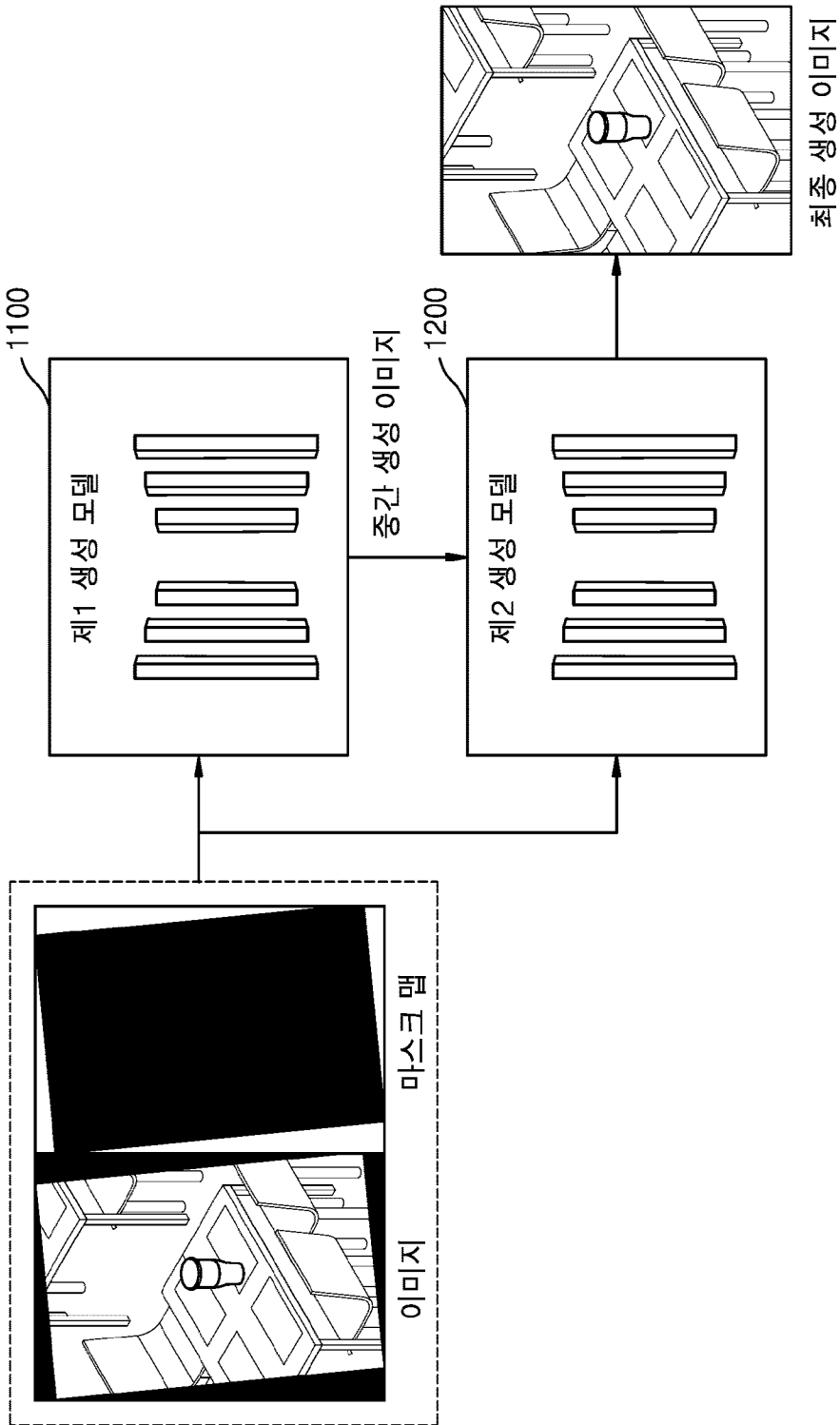


[도2]

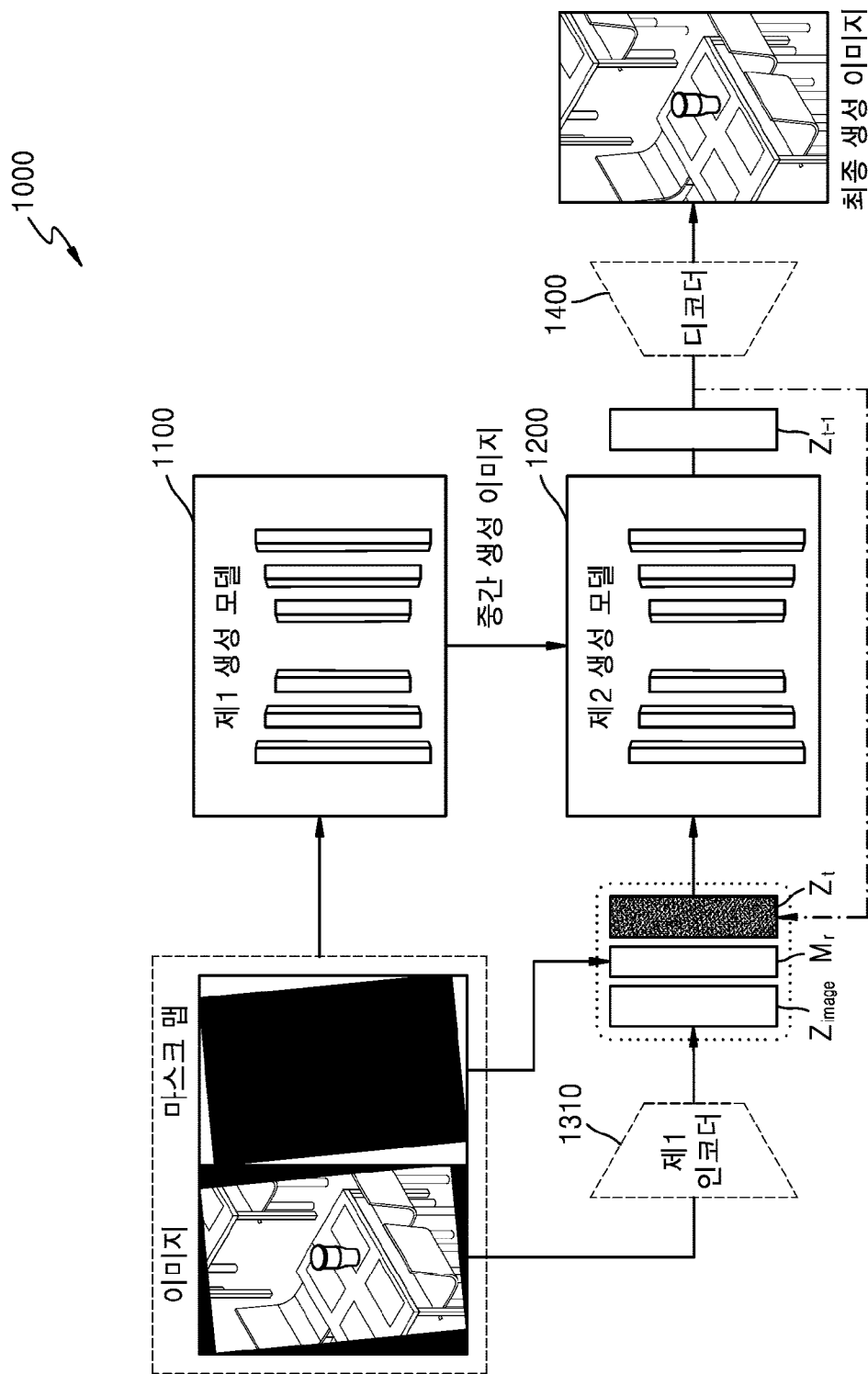


[도3]

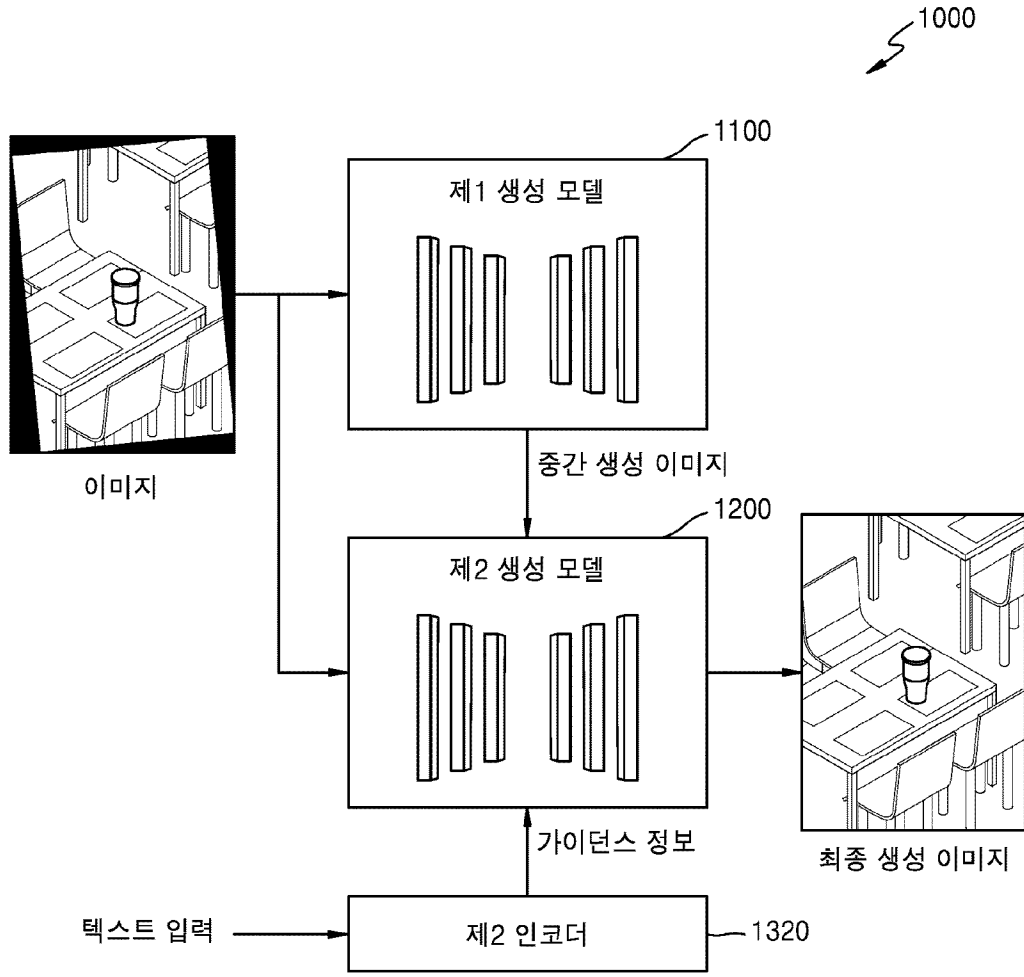
1000 ↘



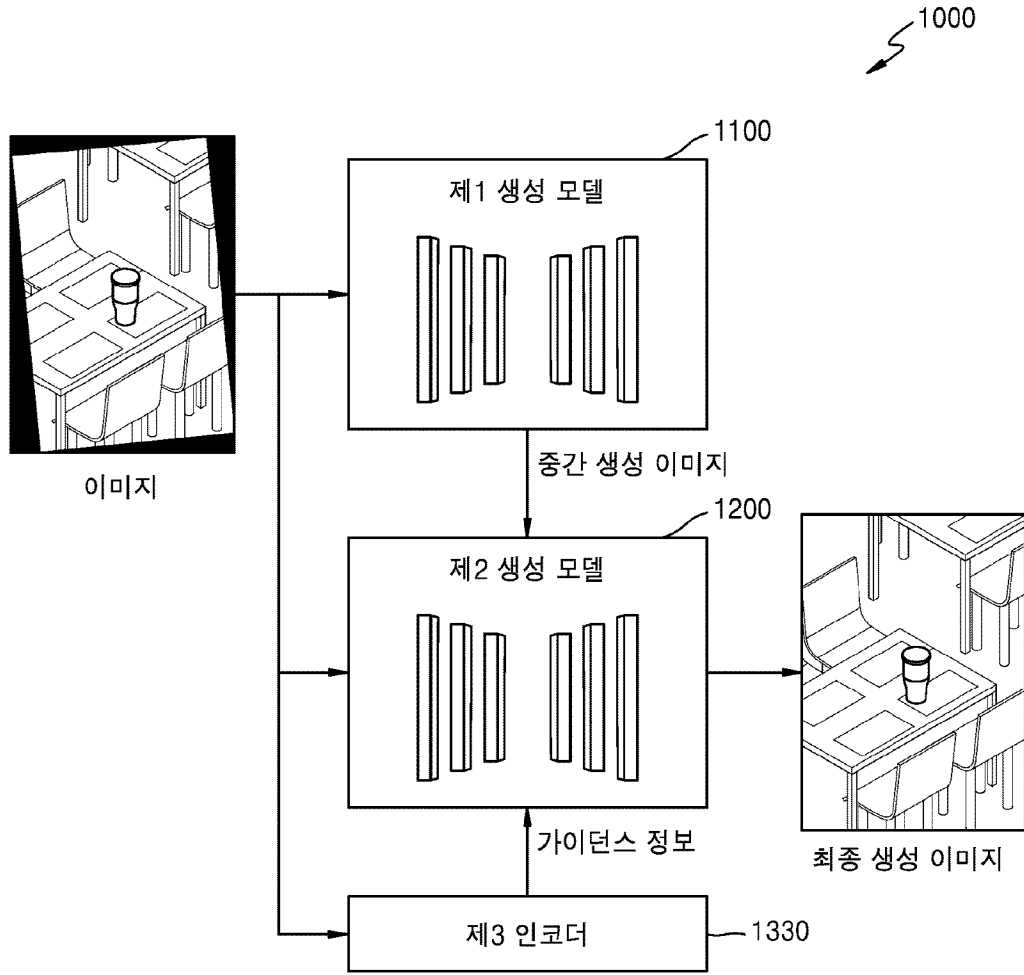
[도4]



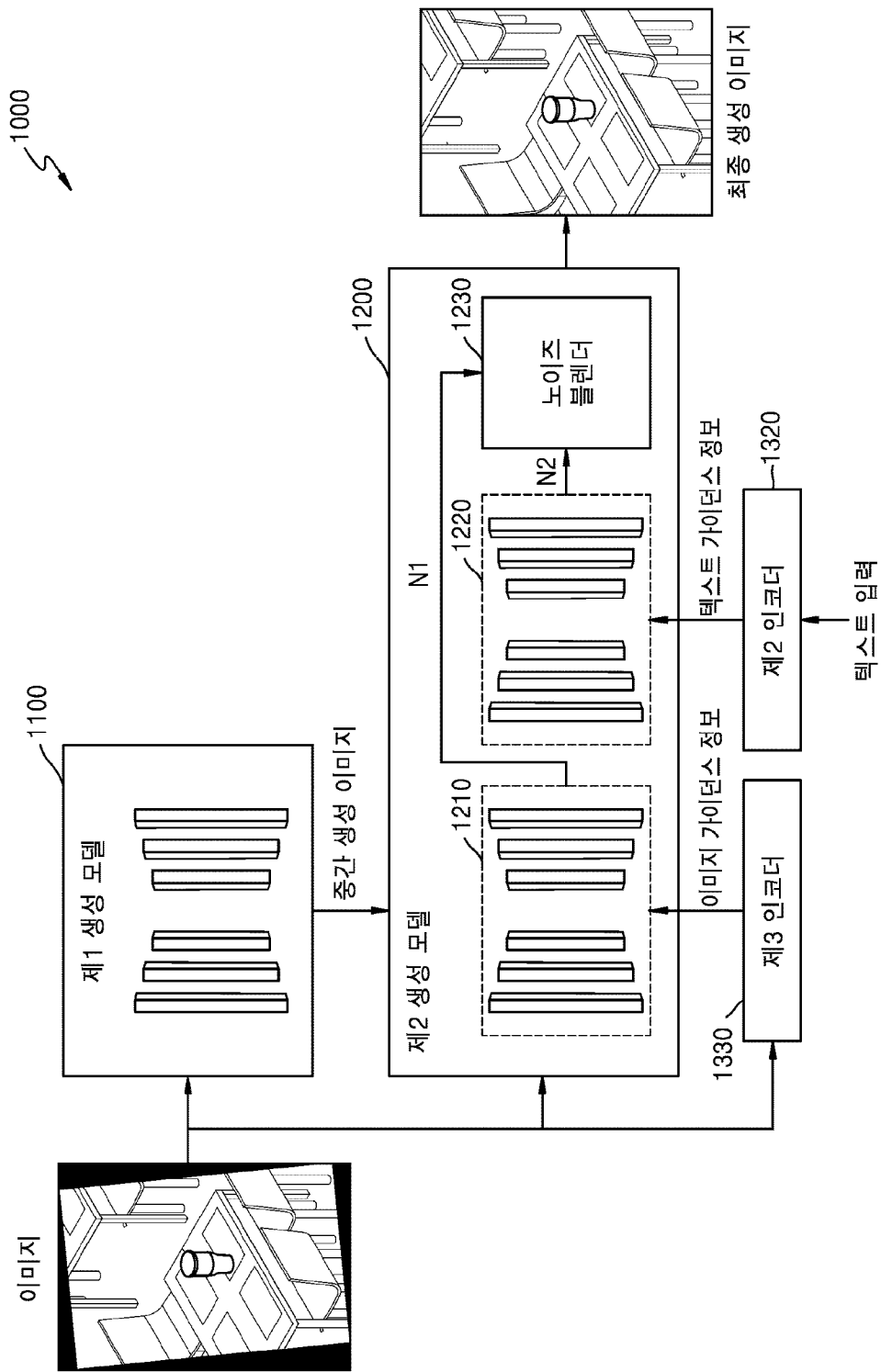
[도5a]



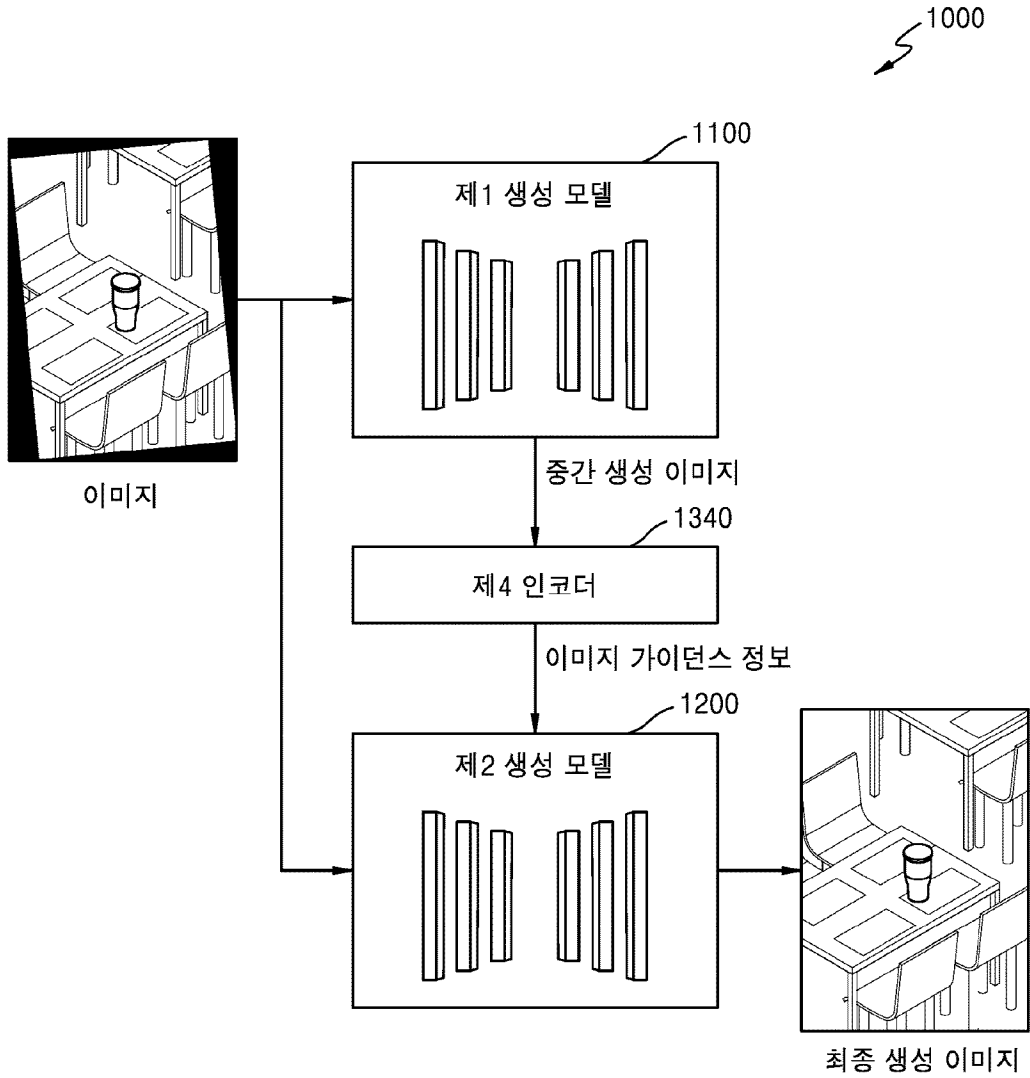
[도5b]



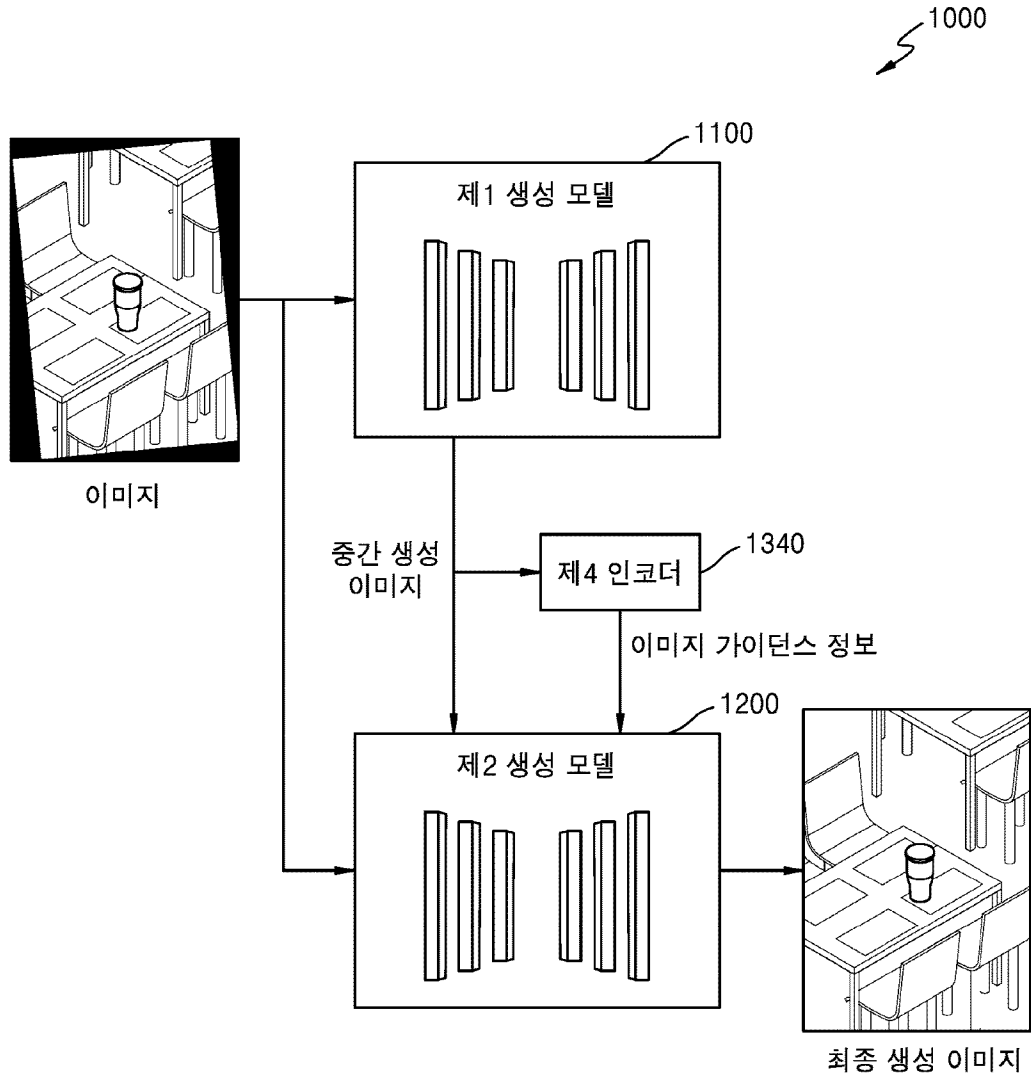
[도5c]



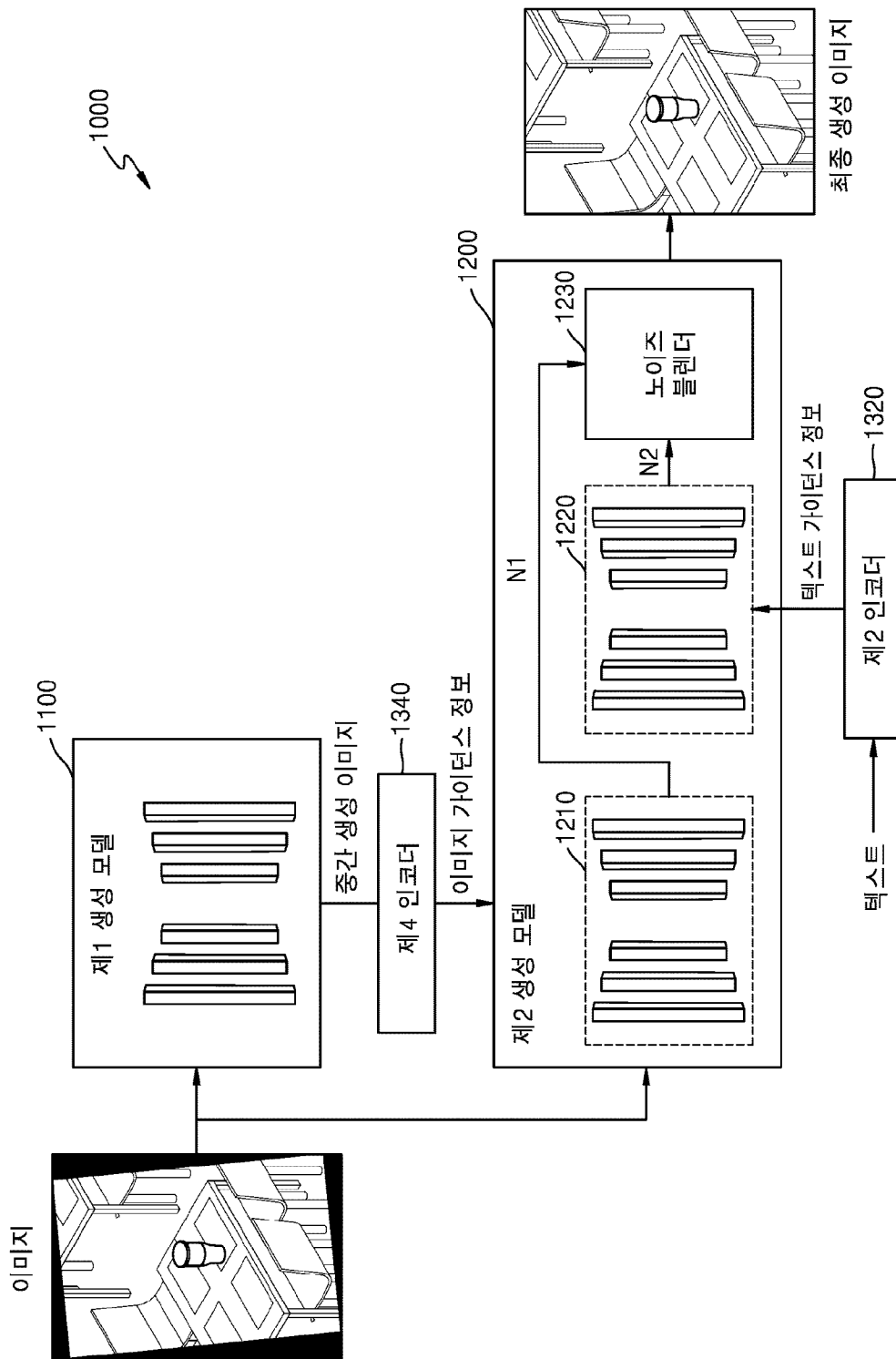
[도5d]



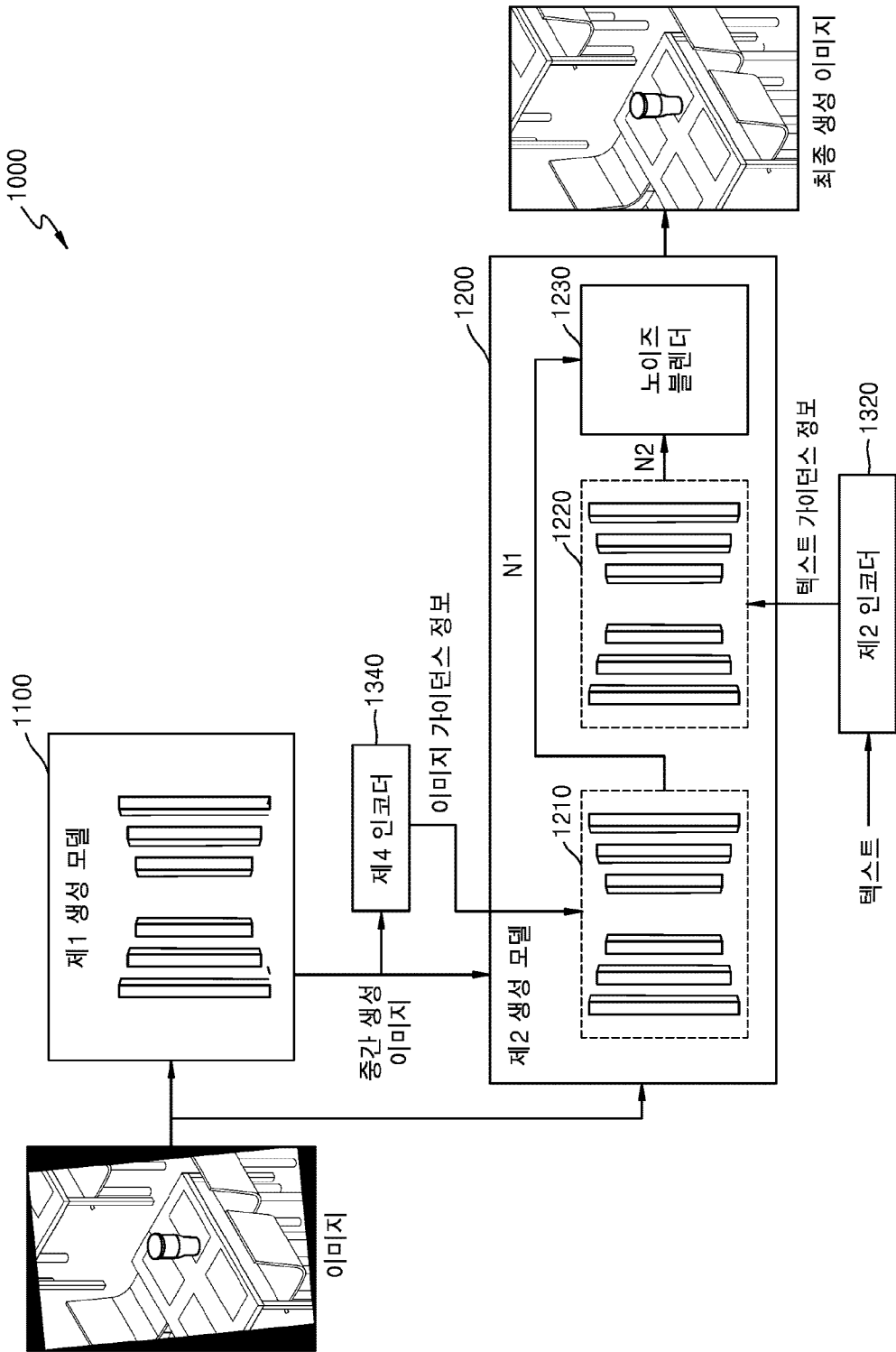
[도5e]



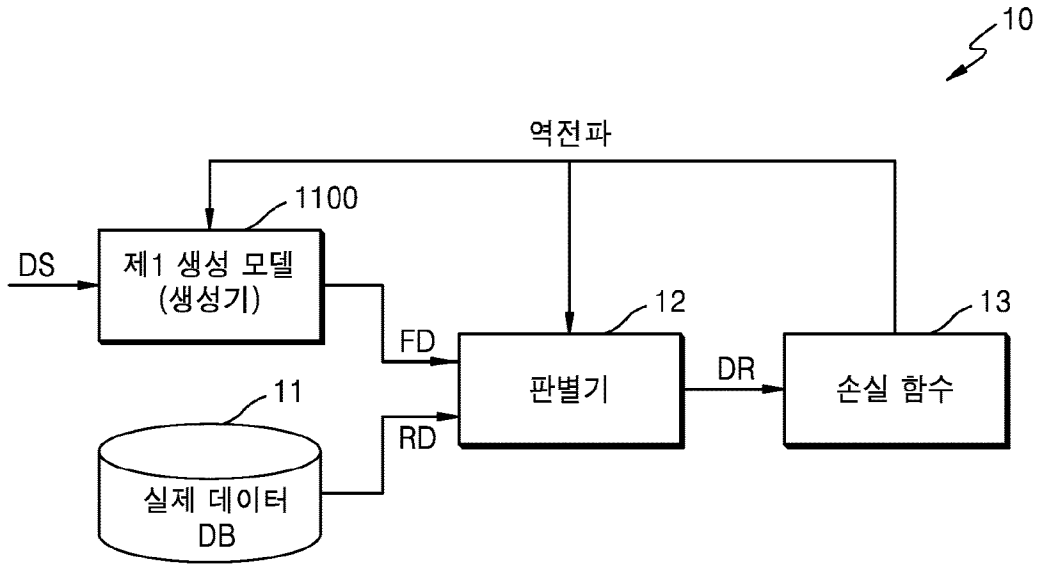
[도5f]



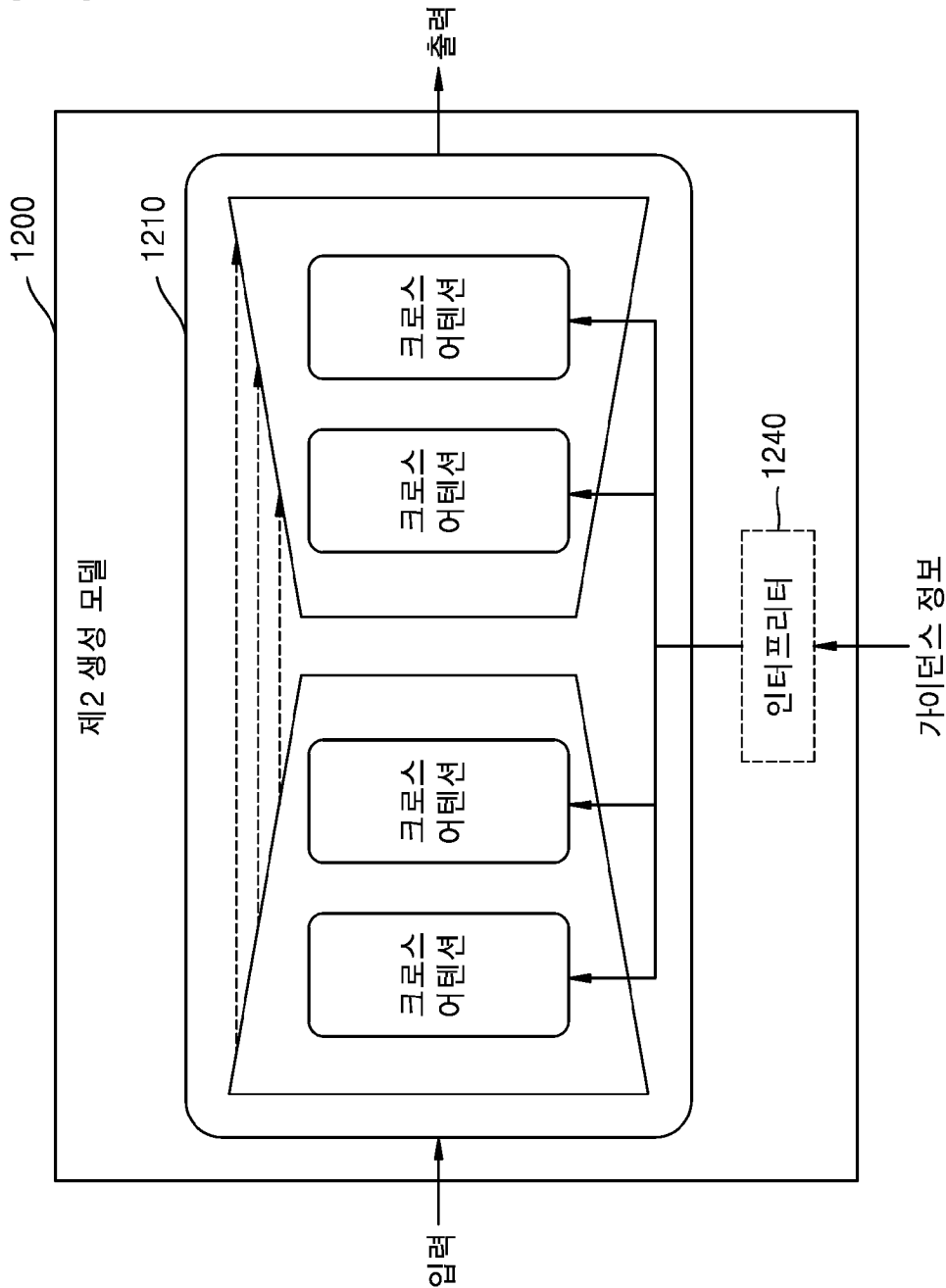
[도5g]



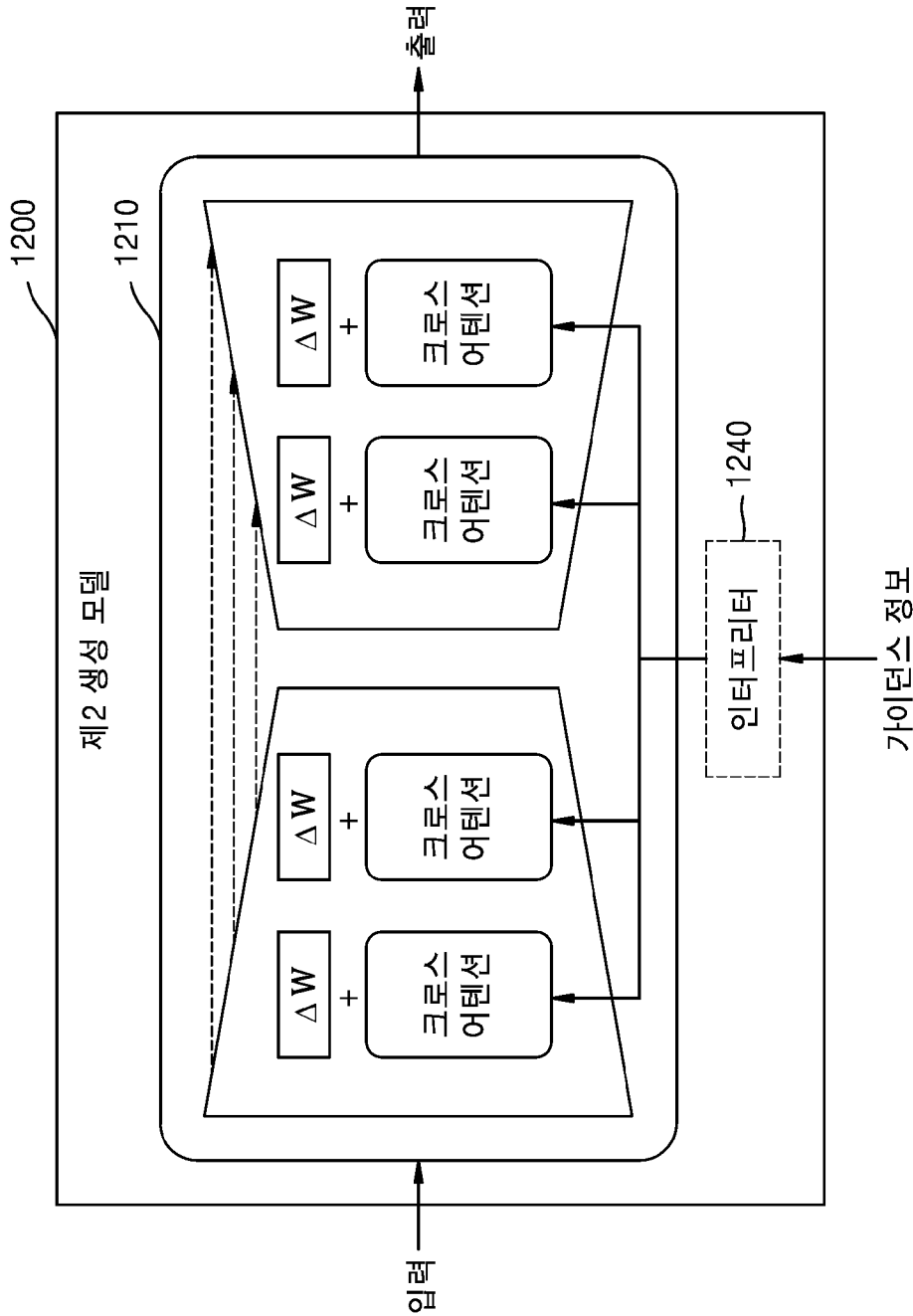
[도6]



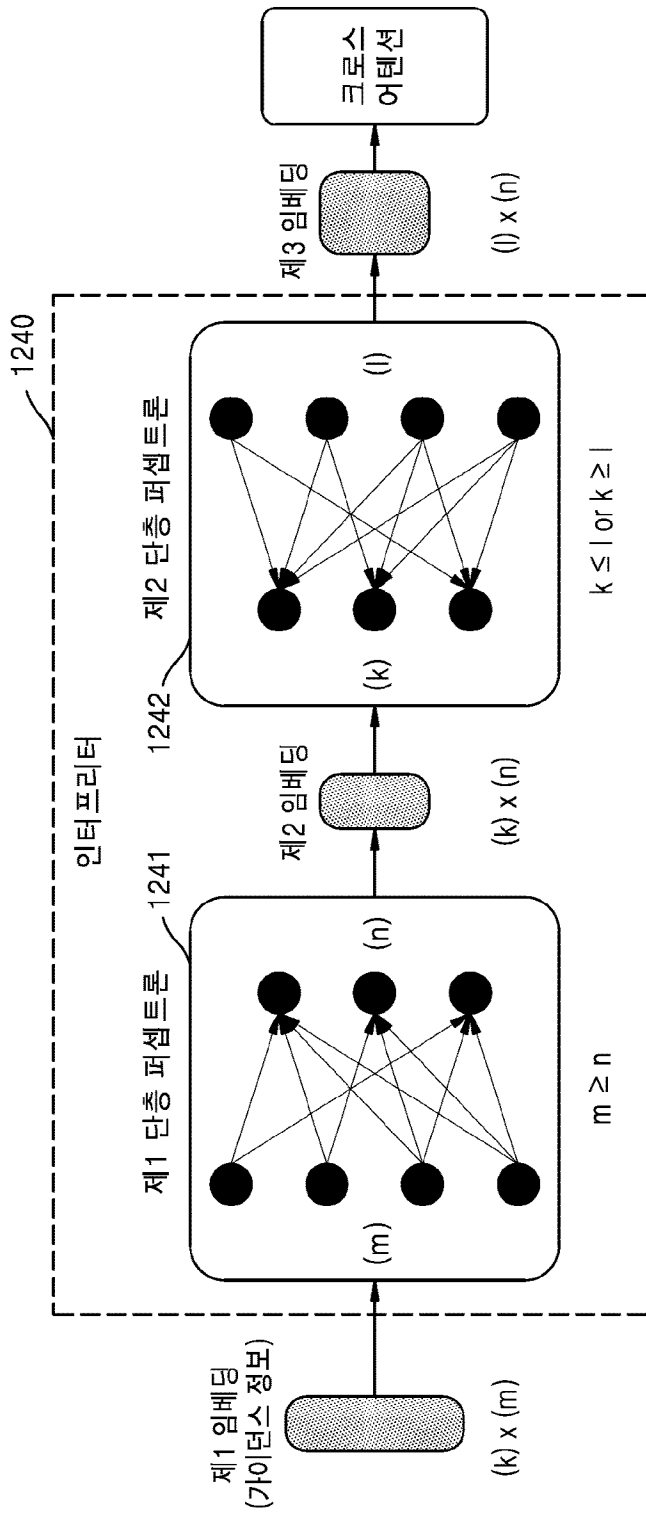
[도7a]



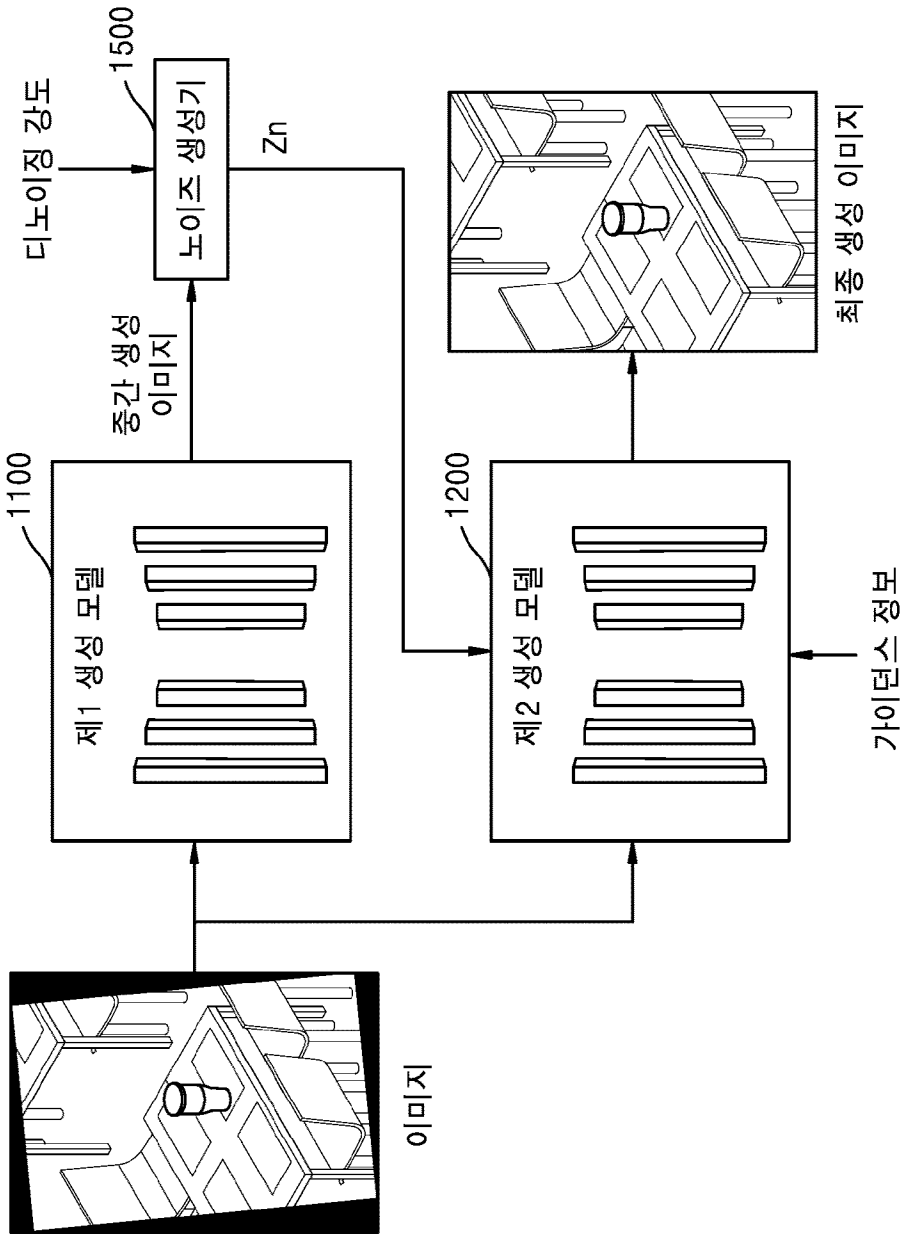
[도7b]



[도8]

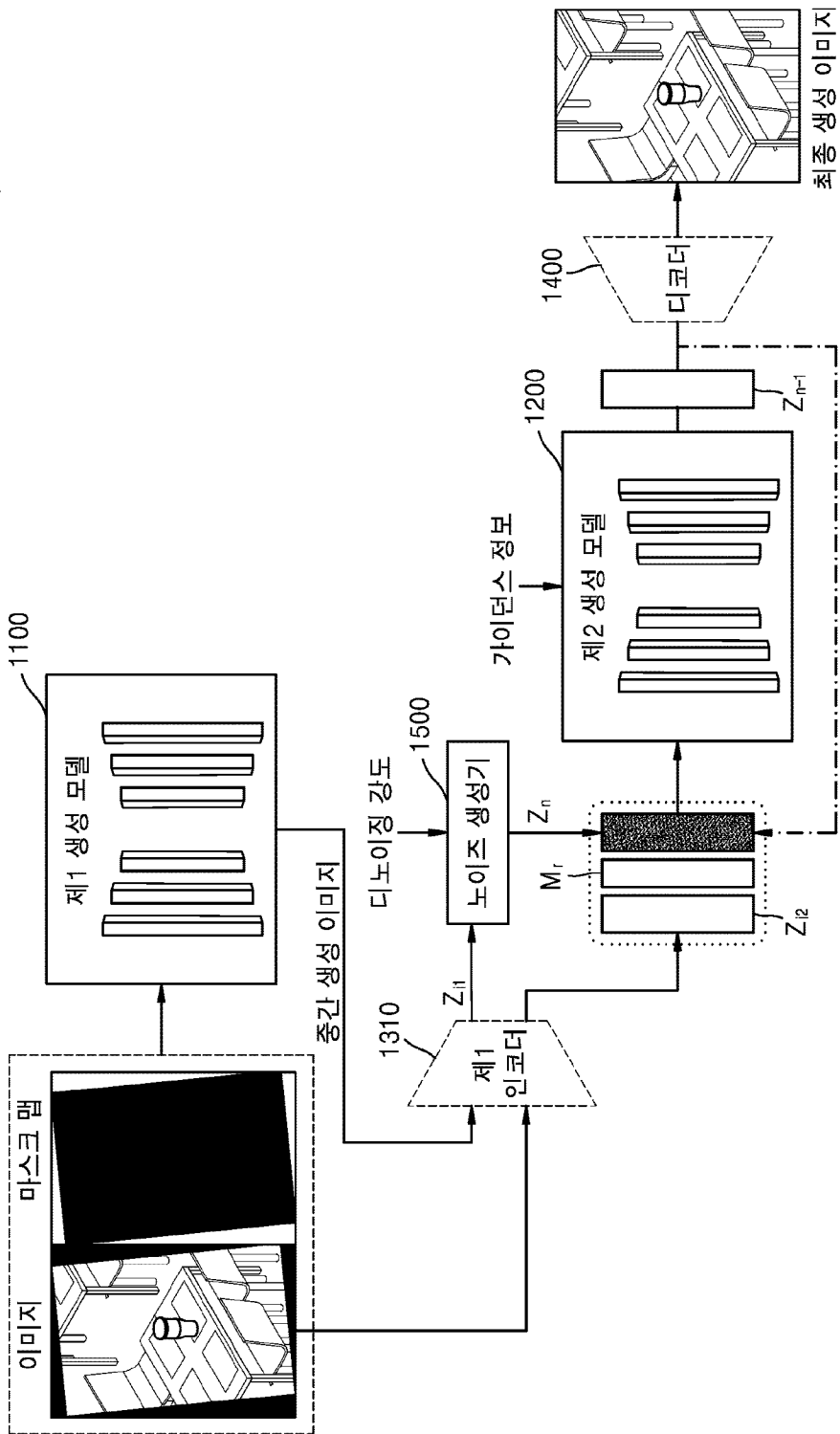


1000 [도 9a]

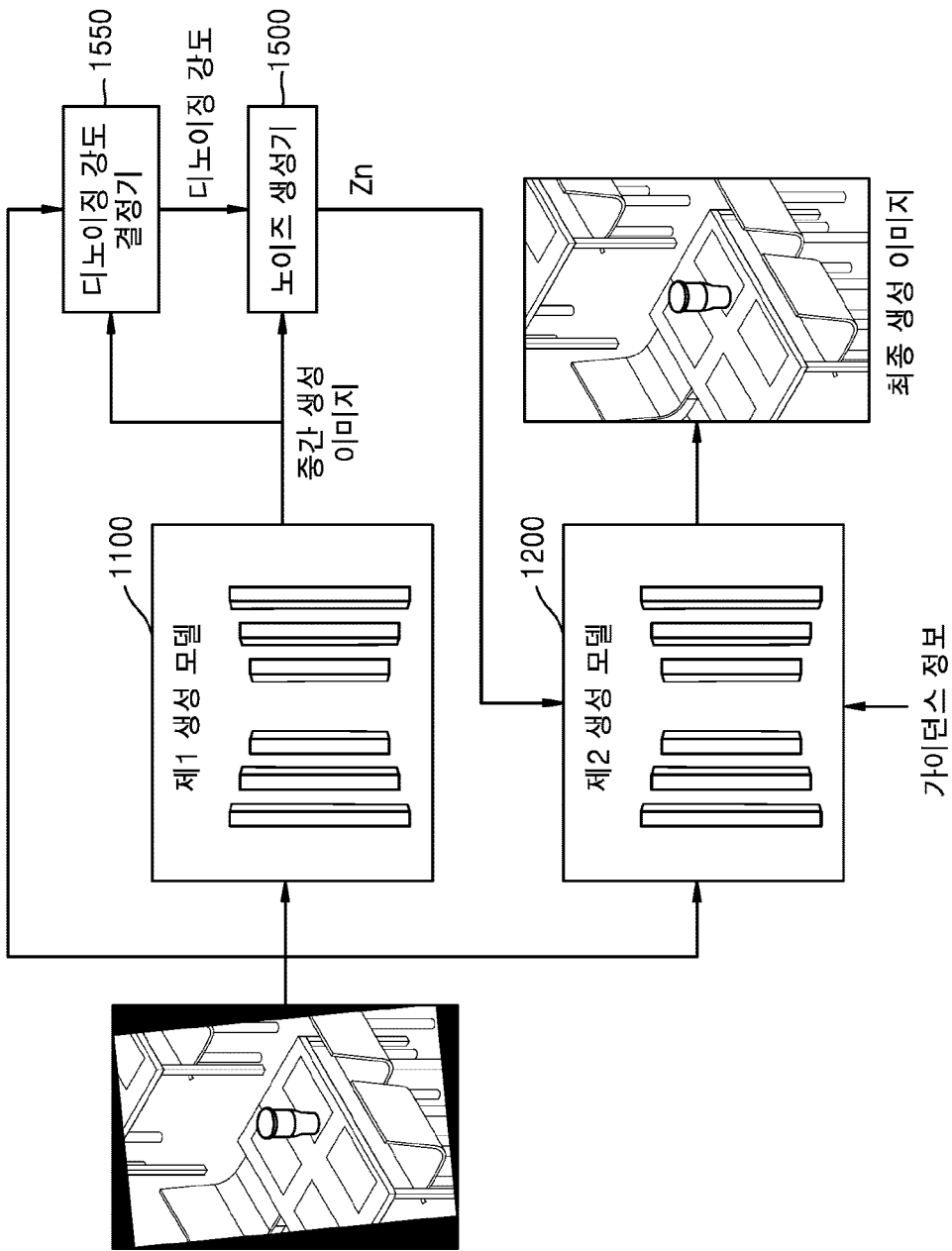


[도9b]

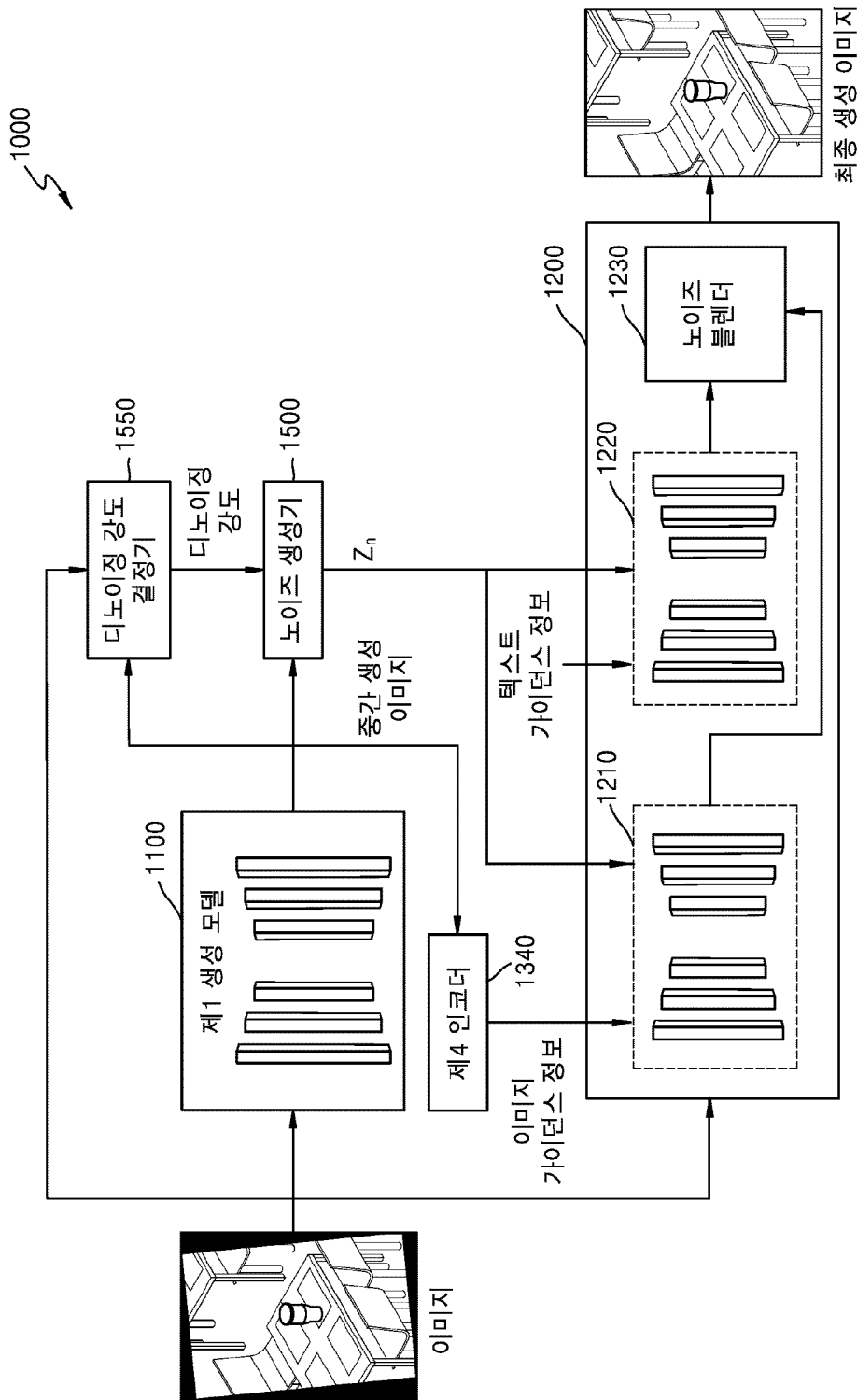
1000 ↙



1000 [도 10a]

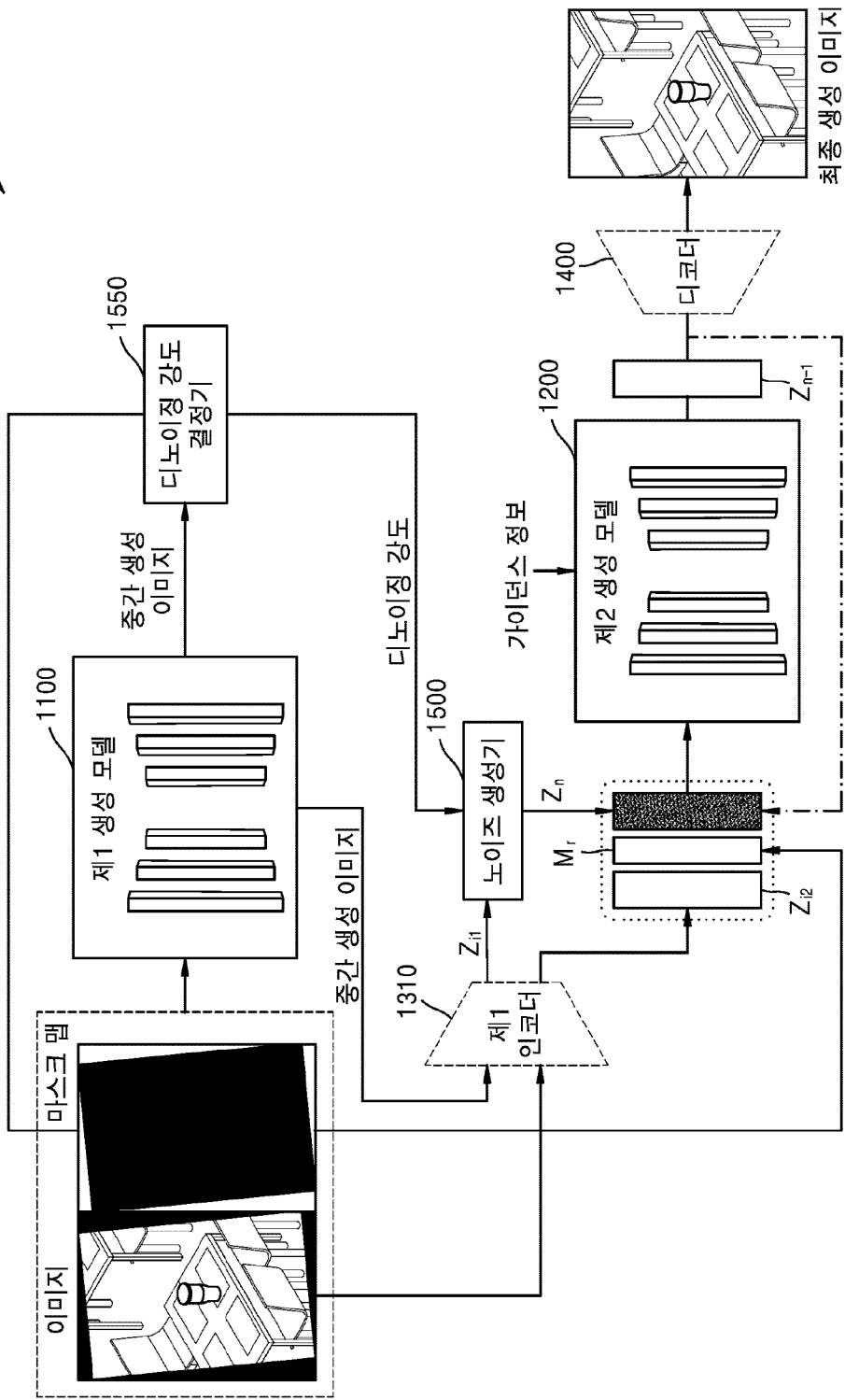


[도 10b]

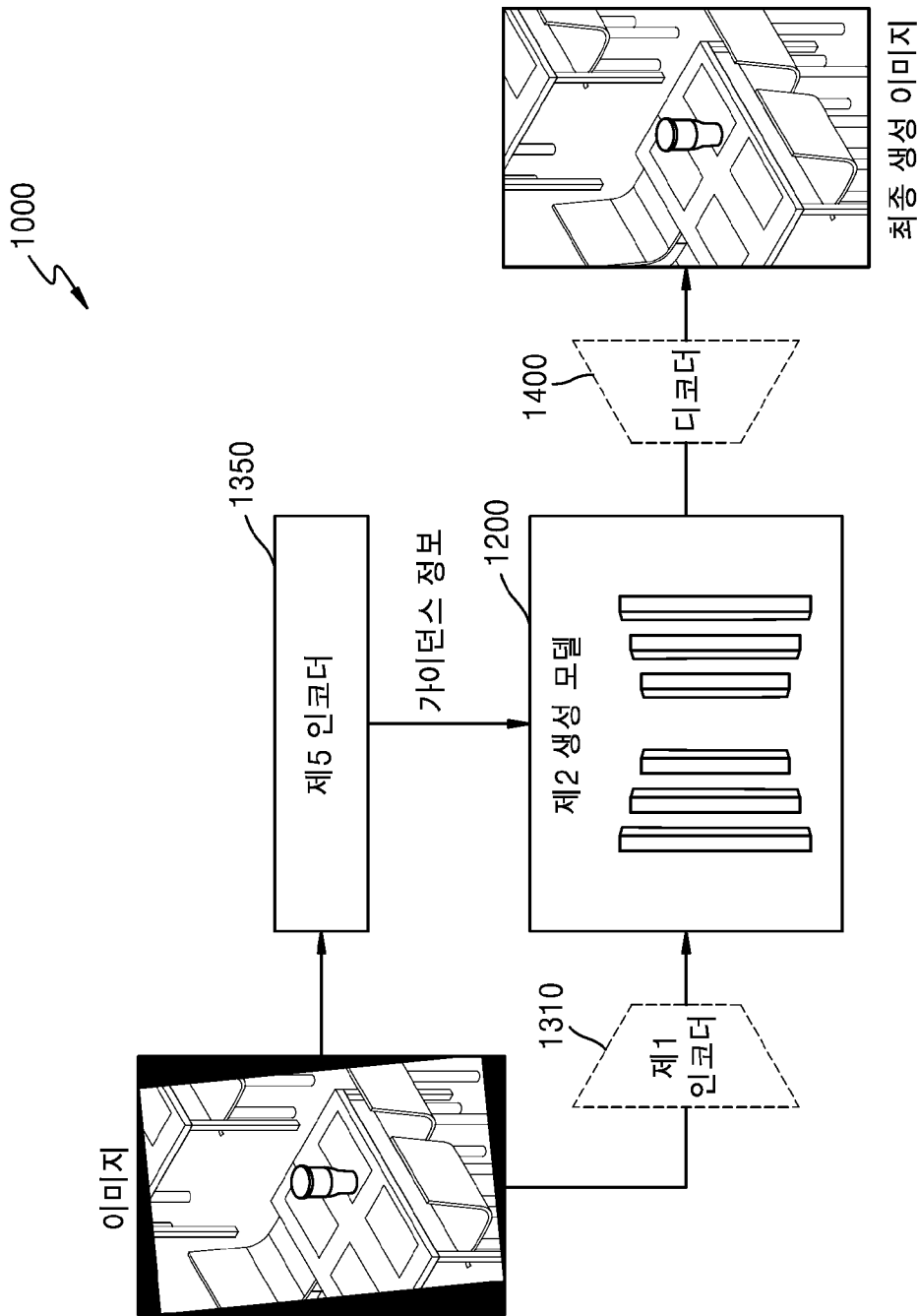


[도10c]

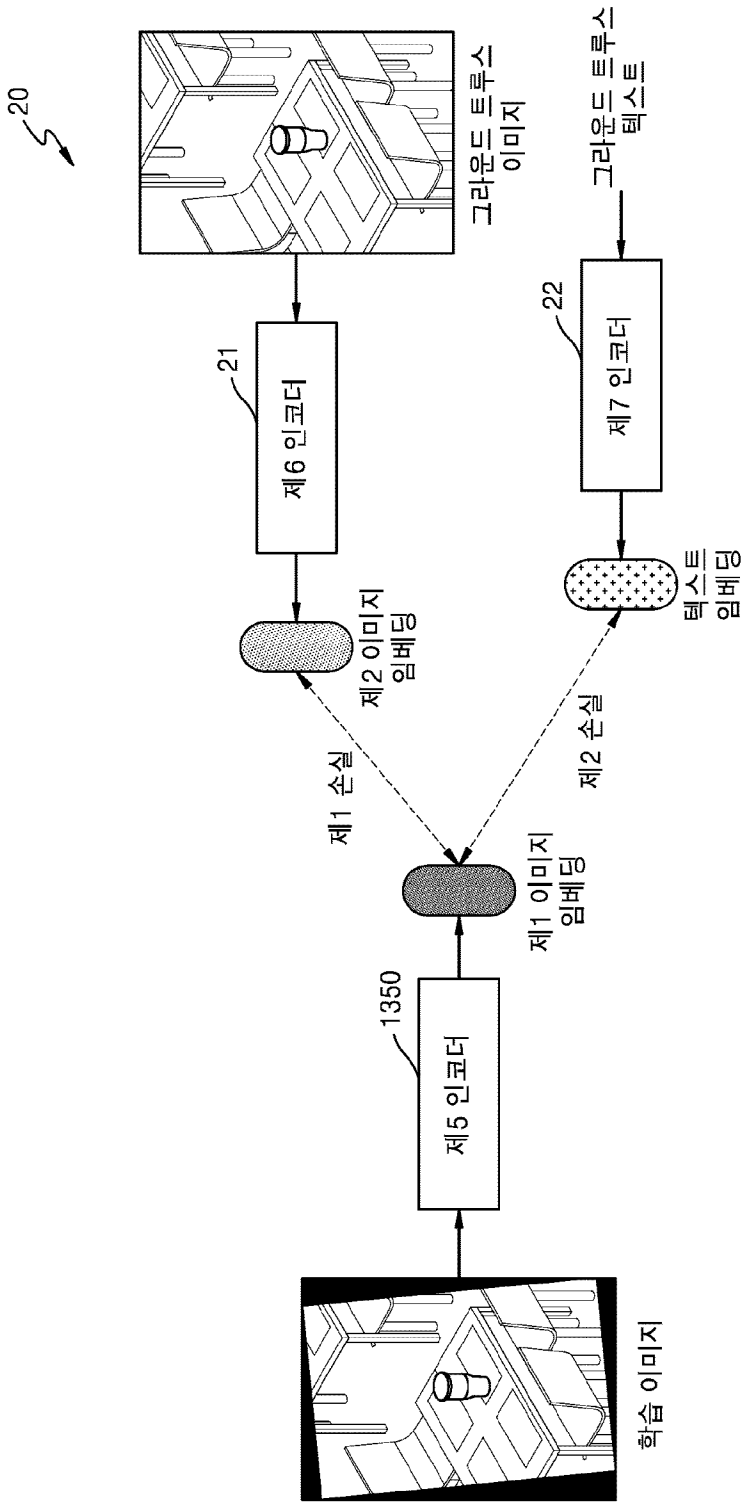
1000 ↘



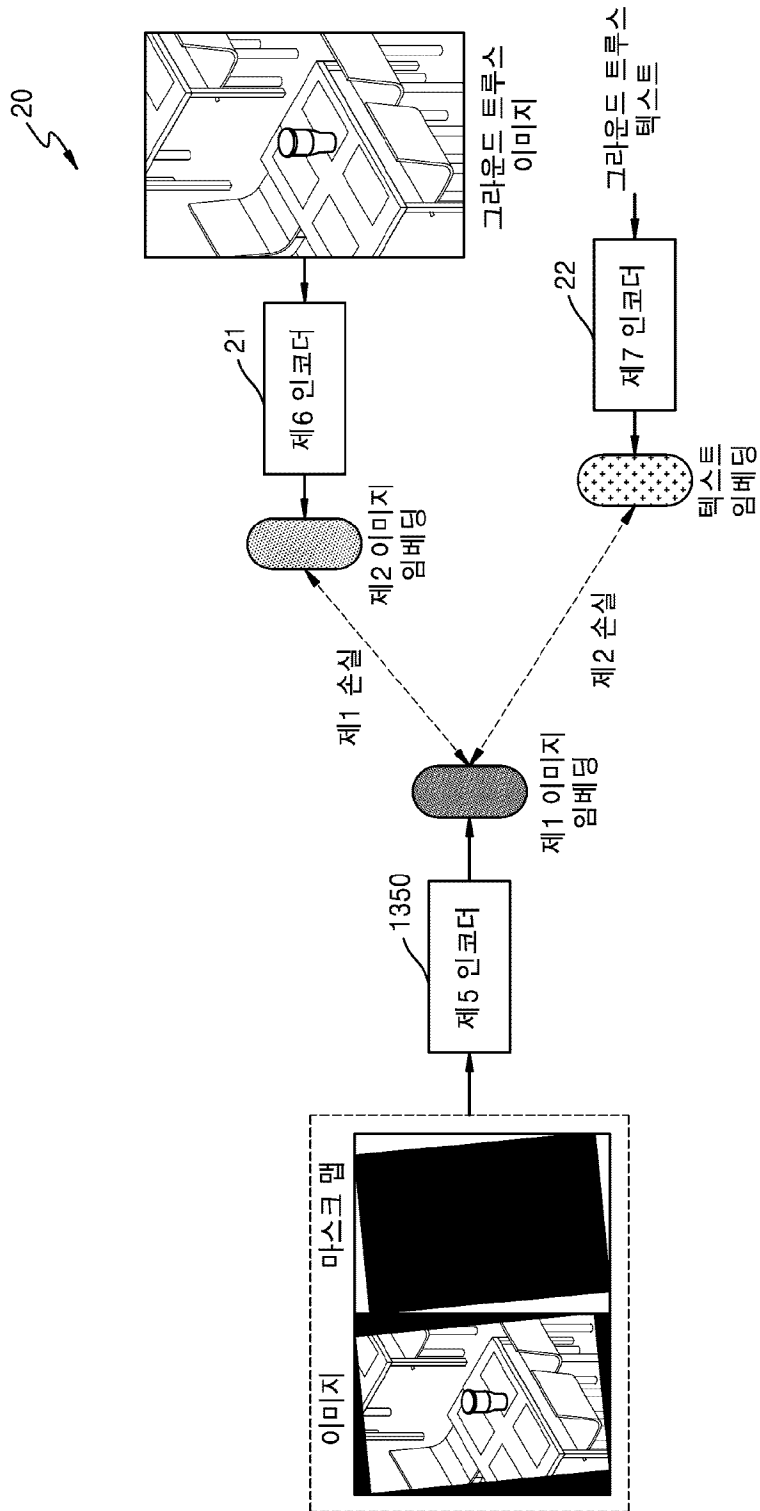
[도11]



[도 12a]

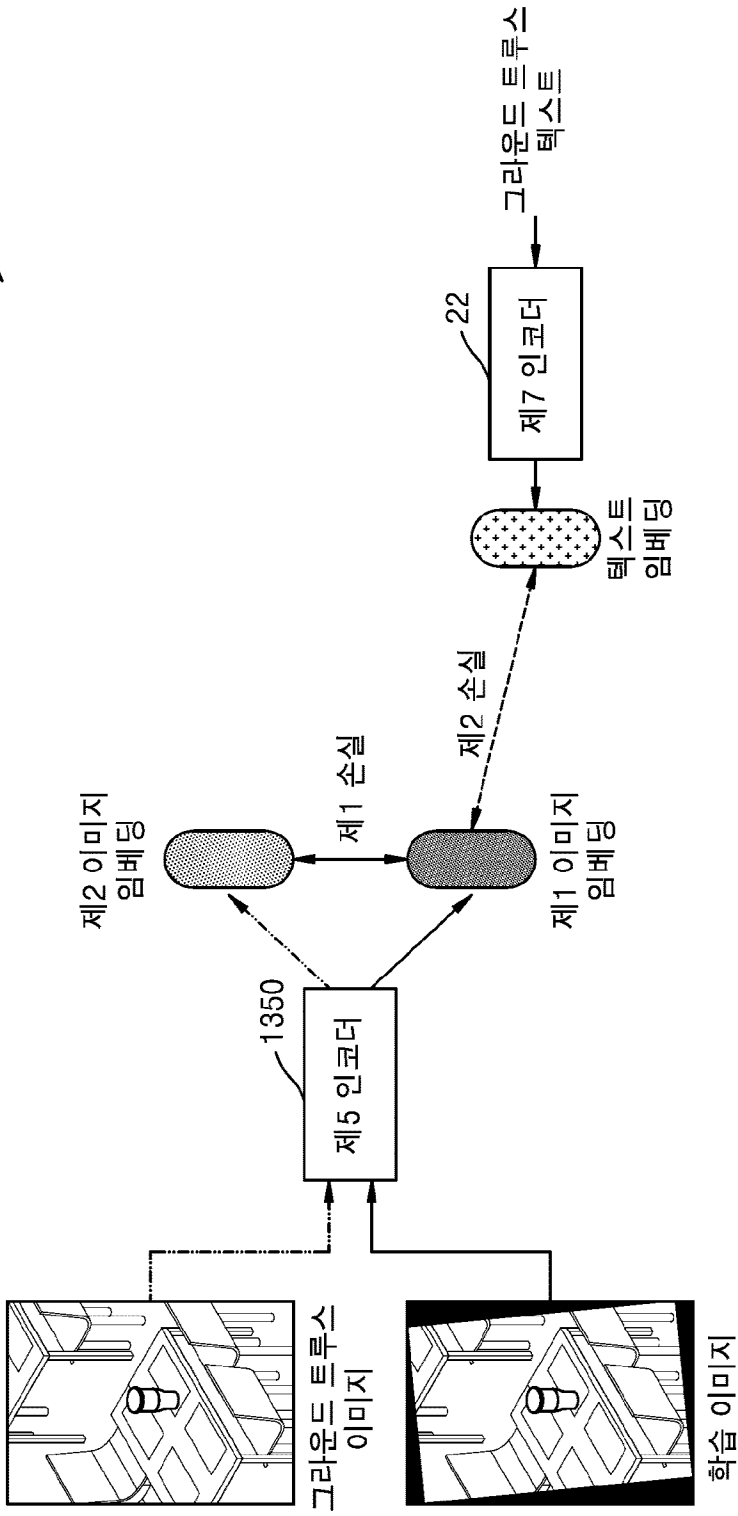


[도 12b]

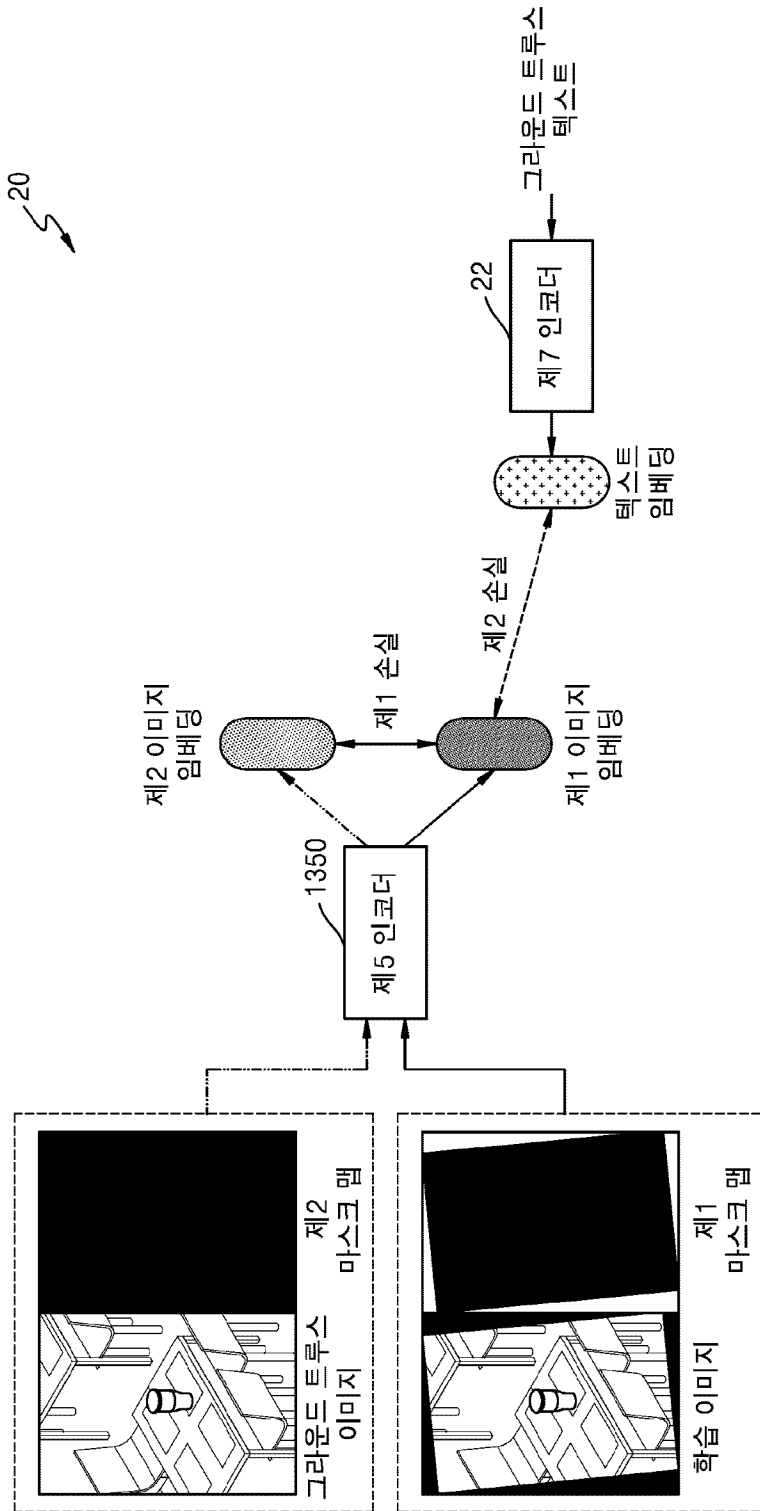


[도 13a]

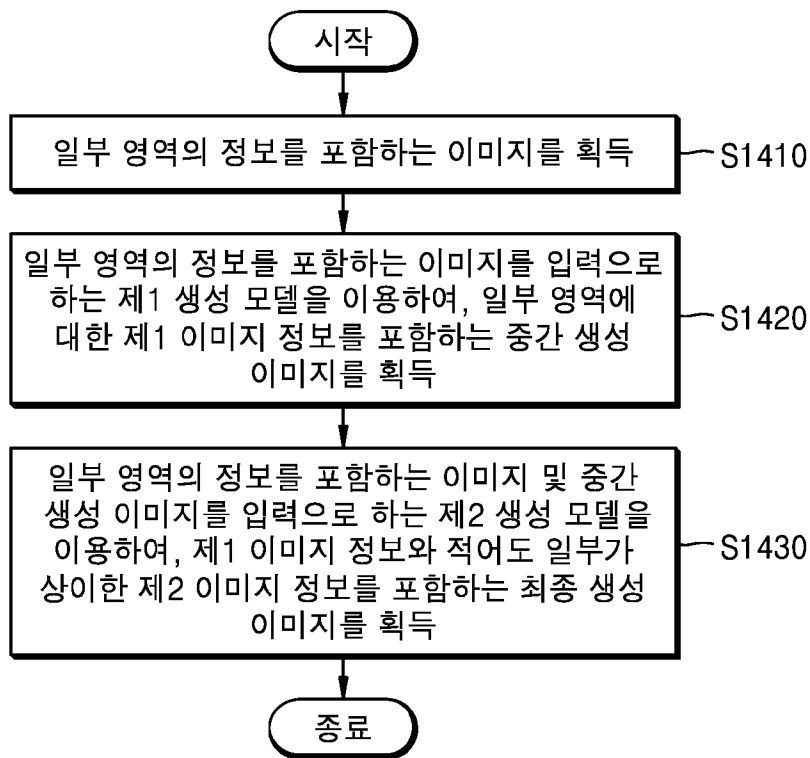
20 ↘



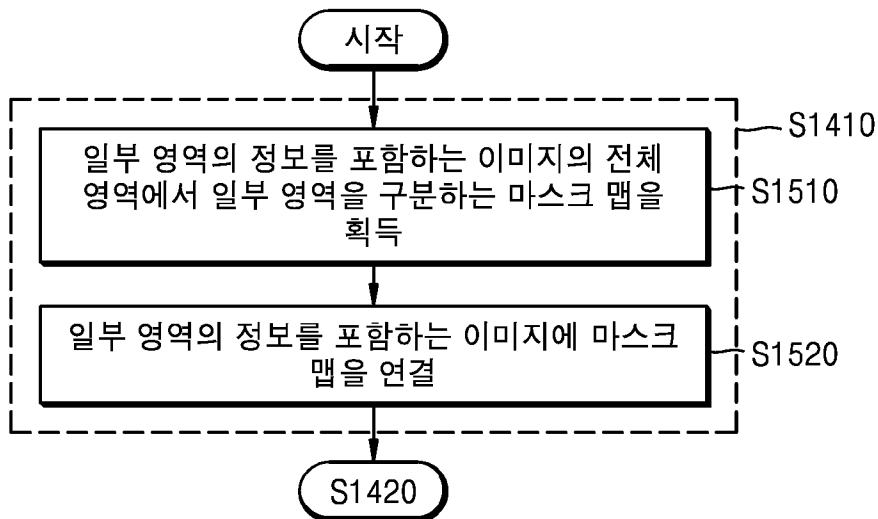
[도 13b]



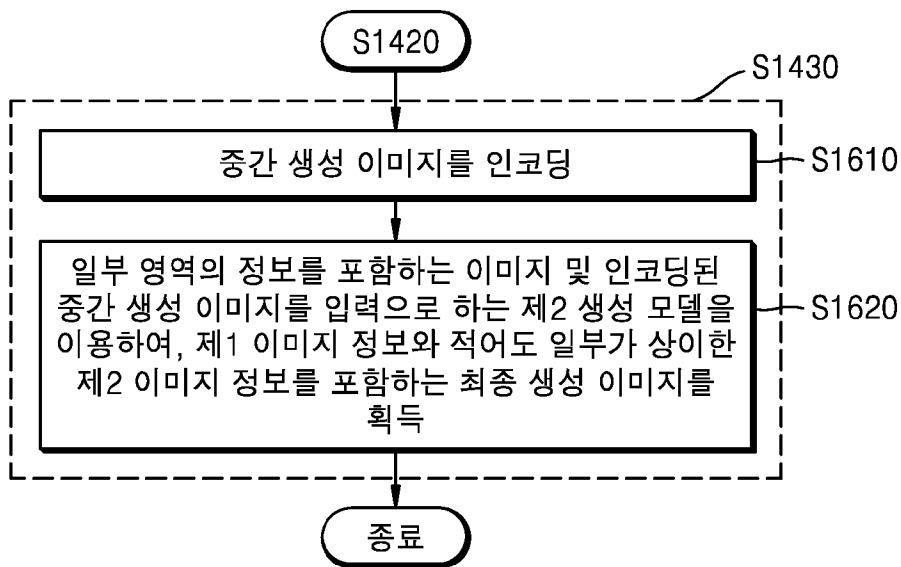
[도14]



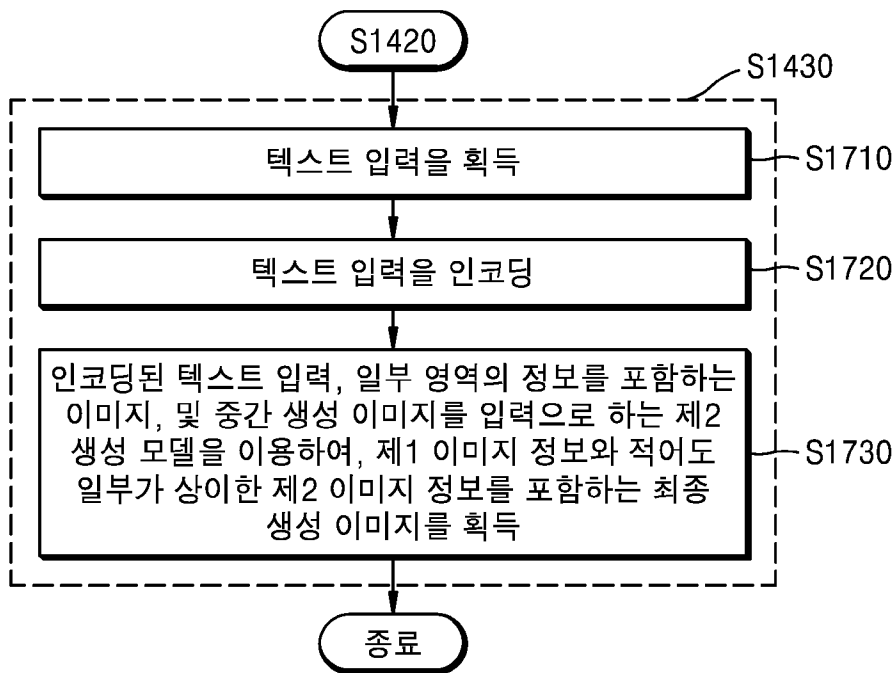
[도15]



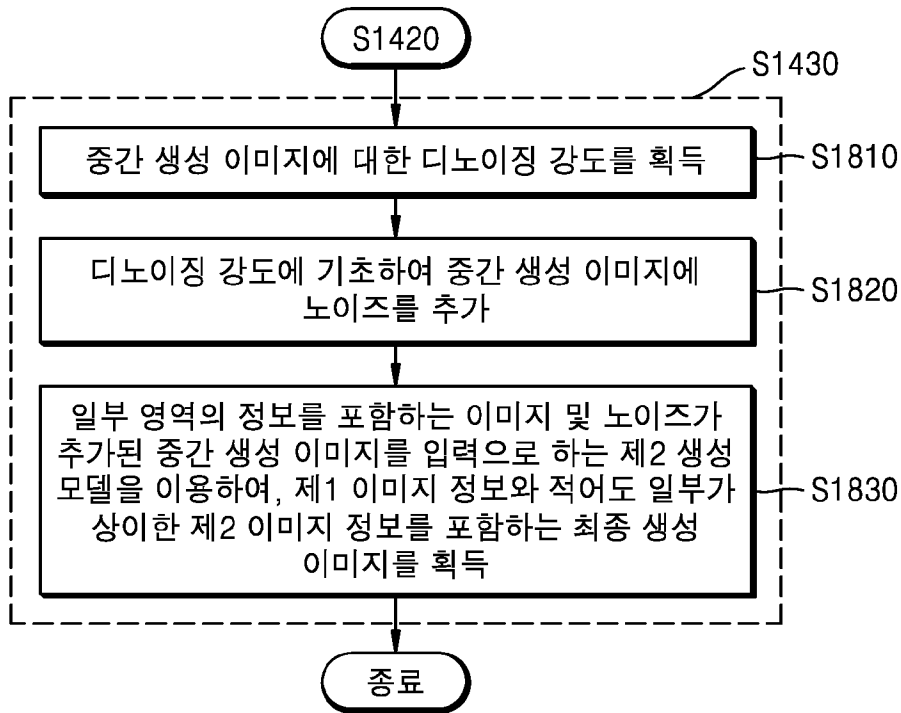
[도16]



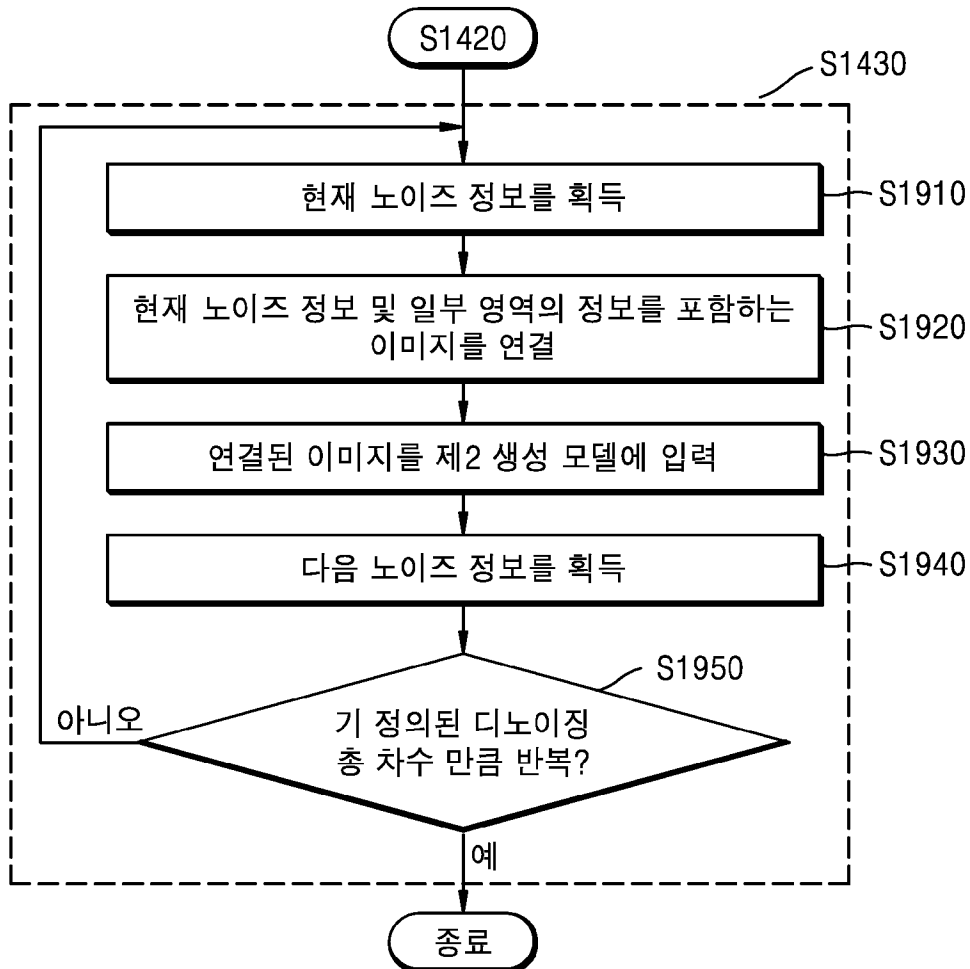
[도17]



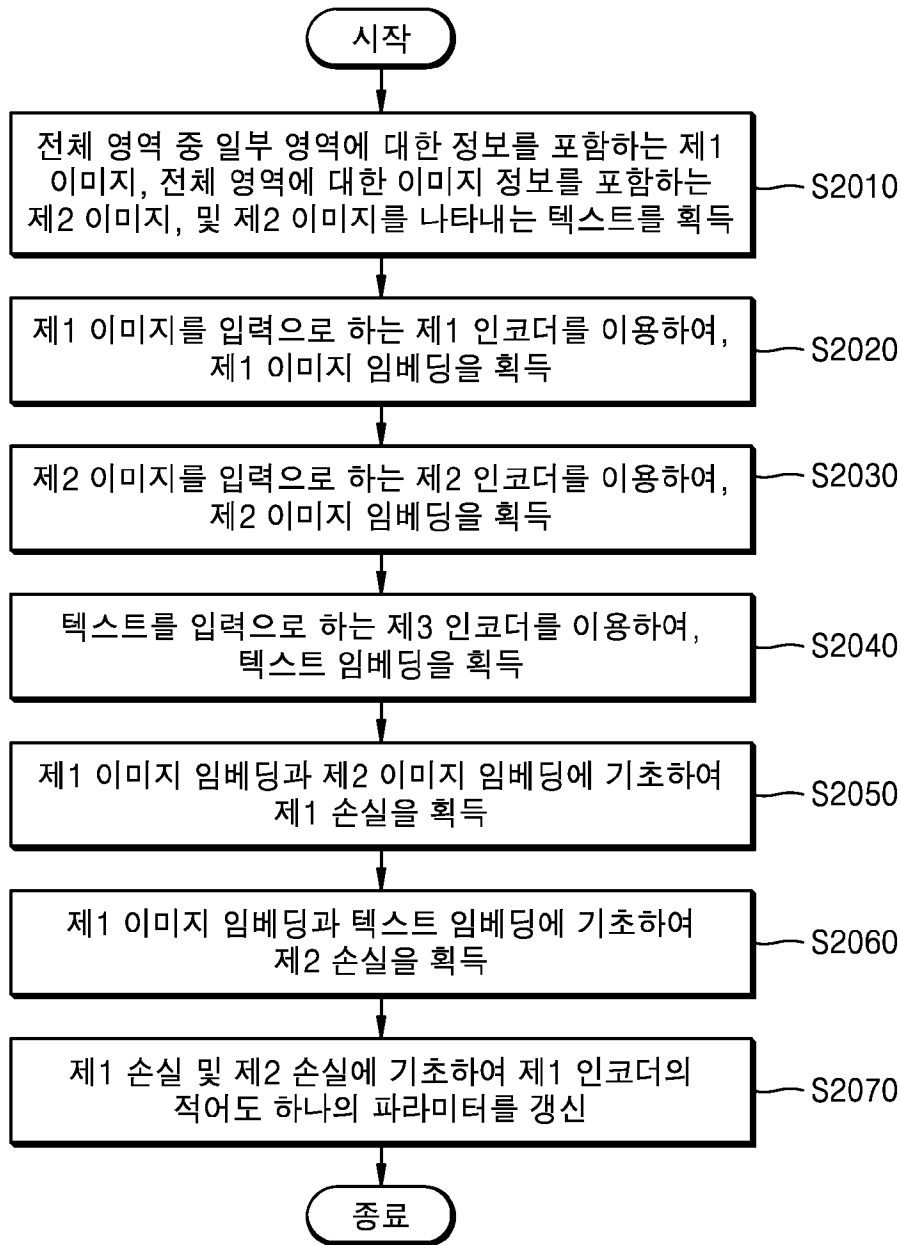
[도18]



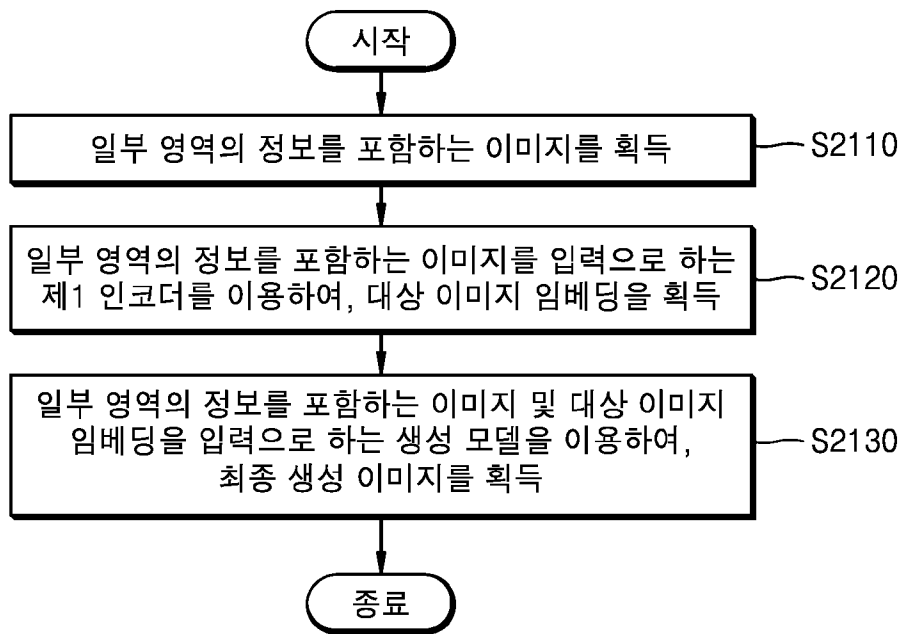
[도19]



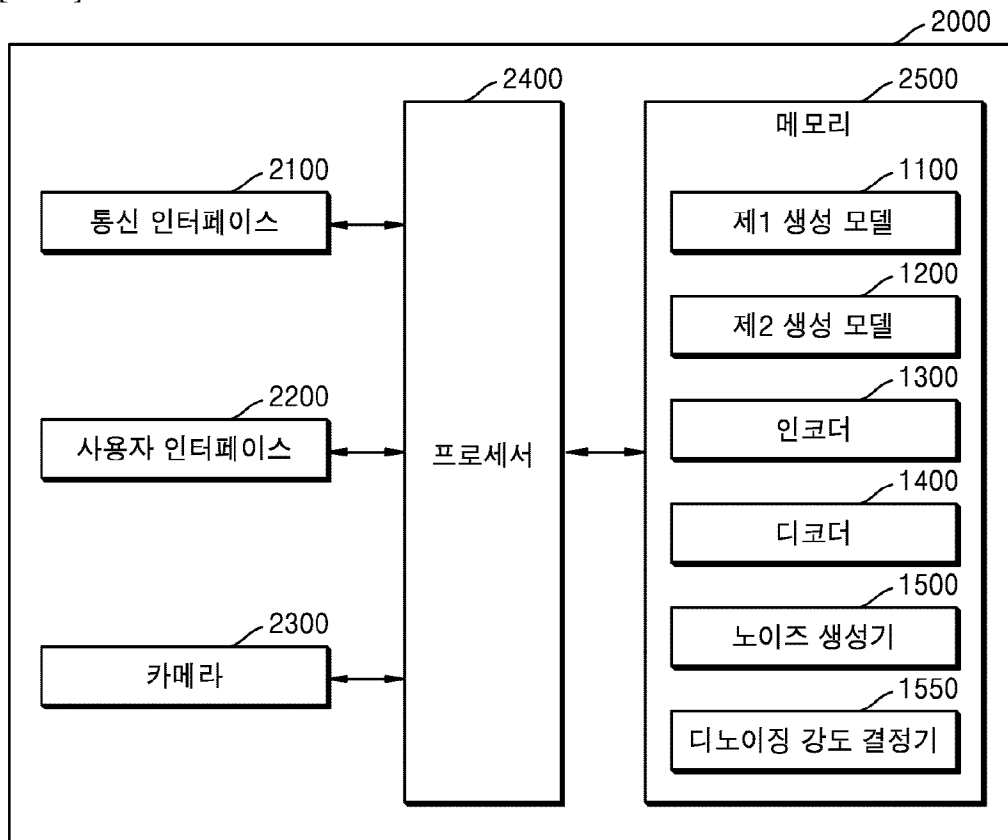
[도20]



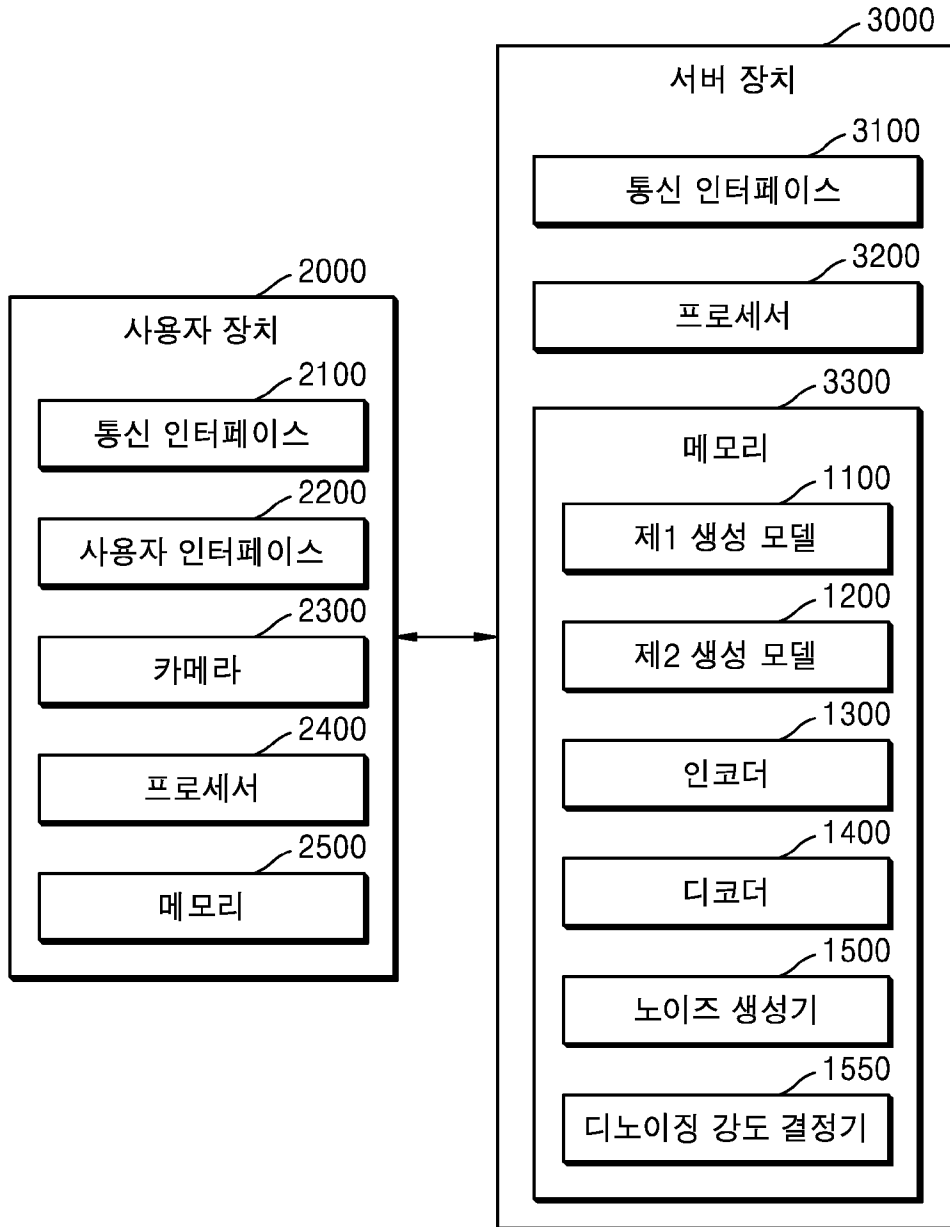
[도21]



[도22]



[도23]



## INTERNATIONAL SEARCH REPORT

International application No.

**PCT/KR2024/013315****A. CLASSIFICATION OF SUBJECT MATTER**

**G06N 3/0475**(2023.01)i; **G06N 3/0455**(2023.01)i; **G06N 3/094**(2023.01)i; **G06N 3/088**(2023.01)i; **G06T 5/77**(2024.01)i;  
**G06T 5/70**(2024.01)i; **G06T 5/60**(2024.01)i; **G06F 40/279**(2020.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

G06N 3/0475(2023.01); G06N 3/0464(2023.01); G06T 3/40(2006.01); G06T 5/00(2006.01); G06T 7/49(2017.01)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean utility models and applications for utility models: IPC as above  
 Japanese utility models and applications for utility models: IPC as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS (KIPO internal) & keywords: 생성 모델(generative model), 이미지(image), 마스크 맵(mask map), 디노이징(denoising), 확산 모델(diffusion model)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	KR 10-2260628 B1 (LEE, In Hyun) 03 June 2021 (2021-06-03) See paragraphs [0063]-[0067] and [0070]; and figures 3-6.	1-15
Y	US 2023-0095092 A1 (NVIDIA CORPORATION) 30 March 2023 (2023-03-30) See paragraphs [0058], [0064]-[0065], [0081] and [0083]; and figures 1-2.	1-15
A	KR 10-2479965 B1 (SAMSUNG ELECTRONICS CO., LTD.) 20 December 2022 (2022-12-20) See claims 1-2.	1-15
A	KR 10-2287407 B1 (INDUSTRY ACADEMIC COOPERATION FOUNDATION OF YEUNGNAM UNIVERSITY) 06 August 2021 (2021-08-06) See paragraphs [0034] and [0039]-[0048].	1-15
A	CN 115908187 A (BEIHANG UNIVERSITY) 04 April 2023 (2023-04-04) See claim 1.	1-15

Further documents are listed in the continuation of Box C.

See patent family annex.

\* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance  
 “D” document cited by the applicant in the international application  
 “E” earlier application or patent but published on or after the international filing date  
 “L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)  
 “O” document referring to an oral disclosure, use, exhibition or other means  
 “P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

**04 December 2024**

Date of mailing of the international search report

**04 December 2024**

Name and mailing address of the ISA/KR

**Korean Intellectual Property Office  
 Government Complex-Daejeon Building 4, 189 Cheongsaro, Seo-gu, Daejeon 35208**

Facsimile No. +82-42-481-8578

Authorized officer

Telephone No.

**INTERNATIONAL SEARCH REPORT**  
**Information on patent family members**

International application No.

**PCT/KR2024/013315**

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
KR	10-2260628	B1	03 June 2021	None			
US	2023-0095092	A1	30 March 2023	None			
KR	10-2479965	B1	20 December 2022	CN	108694700	A	23 October 2018
				CN	108694700	B	01 December 2023
				TW	201837854	A	16 October 2018
				TW	I756378	B	01 March 2022
				US	10489887	B2	26 November 2019
				US	10970820	B2	06 April 2021
				US	11694305	B2	04 July 2023
				US	2018-0293707	A1	11 October 2018
				US	2020-0090305	A1	19 March 2020
				US	2021-0224953	A1	22 July 2021
KR	10-2287407	B1	06 August 2021	US	11869129	B2	09 January 2024
				US	2023-0274479	A1	31 August 2023
				WO	2022-131497	A1	23 June 2022
CN	115908187	A	04 April 2023	None			

<b>A. 발명이 속하는 기술분류(국제특허분류(IPC))</b> G06N 3/0475(2023.01)i; G06N 3/0455(2023.01)i; G06N 3/094(2023.01)i; G06N 3/088(2023.01)i; G06T 5/77(2024.01)i; G06T 5/70(2024.01)i; G06T 5/60(2024.01)i; G06F 40/279(2020.01)i		
<b>B. 조사된 분야</b>		
조사된 최소문헌(국제특허분류를 기재) G06N 3/0475(2023.01); G06N 3/0464(2023.01); G06T 3/40(2006.01); G06T 5/00(2006.01); G06T 7/49(2017.01)		
조사된 기술분야에 속하는 최소문헌 이외의 문헌 한국등록실용신안공보 및 한국공개실용신안공보: 조사된 최소문헌란에 기재된 IPC 일본등록실용신안공보 및 일본공개실용신안공보: 조사된 최소문헌란에 기재된 IPC		
국제조사에 이용된 전산 데이터베이스(데이터베이스의 명칭 및 검색어(해당하는 경우)) eKOMPASS(특허청 내부 검색시스템) & 키워드: 생성 모델(generative model), 이미지(image), 마스크 맵(mask map), 디노이징(denoising), 확산 모델(diffusion model)		
<b>C. 관련 문헌</b>		
카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
Y	KR 10-2260628 B1 (이인현) 2021.06.03 단락 [0063]-[0067], [0070]; 및 도면 3-6	1-15
Y	US 2023-0095092 A1 (NVIDIA CORPORATION) 2023.03.30 단락 [0058], [0064]-[0065], [0081], [0083]; 및 도면 1-2	1-15
A	KR 10-2479965 B1 (삼성전자주식회사) 2022.12.20 청구항 1-2	1-15
A	KR 10-2287407 B1 (영남대학교 산학협력단) 2021.08.06 단락 [0034], [0039]-[0048]	1-15
A	CN 115908187 A (BEIHANG UNIVERSITY) 2023.04.04 청구항 1	1-15
<input type="checkbox"/> 추가 문헌이 C(계속)에 기재되어 있습니다. <input checked="" type="checkbox"/> 대응특허에 관한 별지를 참조하십시오.		
* 인용된 문헌의 특별 카테고리: “A” 특별히 관련이 없는 것으로 보이는 일반적인 기술수준을 정의한 문헌 “D” 본 국제출원에서 출원인이 인용한 문헌 “E” 국제출원일보다 빠른 출원일 또는 우선일을 가지나 국제출원일 이후에 공개된 선출원 또는 특허 문헌 “L” 우선권 주장에 의문을 제기하는 문헌 또는 다른 인용문헌의 공개일 또는 다른 특별한 이유(이유를 명시)를 밝히기 위하여 인용된 문헌 “O” 구두 개시, 사용, 전시 또는 기타 수단을 언급하고 있는 문헌 “P” 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌 “T” 국제출원일 또는 우선일 후에 공개된 문헌으로, 출원과 상충하지 않으며 발명의 기초가 되는 원리나 이론을 이해하기 위해 인용된 문헌 “X” 특별한 관련이 있는 문헌. 해당 문헌 하나만으로 청구된 발명의 신규성 또는 진보성이 없는 것으로 본다. “Y” 특별한 관련이 있는 문헌. 해당 문헌이 하나 이상의 다른 문헌과 조합하는 경우로 그 조합이 당업자에게 자명한 경우 청구된 발명은 진보성이 없는 것으로 본다. “&” 동일한 대응특허문헌에 속하는 문헌		
국제조사의 실제 완료일 2024년12월04일 (04.12.2024)	국제조사보고서 발송일 2024년12월04일 (04.12.2024)	
ISA/KR의 명칭 및 우편주소 대한민국 특허청 (35208) 대전광역시 서구 청사로 189, 4동 (둔산동, 정부대전청사) 팩스 번호 +82-42-481-8578	심사관 남기영 전화번호 +82-42-481-8593	

국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
KR 10-2260628 B1	2021/06/03	없음	
US 2023-0095092 A1	2023/03/30	없음	
KR 10-2479965 B1	2022/12/20	CN 108694700 A	2018/10/23
		CN 108694700 B	2023/12/01
		TW 201837854 A	2018/10/16
		TW I756378 B	2022/03/01
		US 10489887 B2	2019/11/26
		US 10970820 B2	2021/04/06
		US 11694305 B2	2023/07/04
		US 2018-0293707 A1	2018/10/11
		US 2020-0090305 A1	2020/03/19
		US 2021-0224953 A1	2021/07/22
KR 10-2287407 B1	2021/08/06	US 11869129 B2	2024/01/09
		US 2023-0274479 A1	2023/08/31
		WO 2022-131497 A1	2022/06/23
CN 115908187 A	2023/04/04	없음	