



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0136705
(43) 공개일자 2024년09월19일

(51) 국제특허분류(Int. Cl.)

G06T 5/50 (2024.01) G06T 3/40 (2024.01)
G06T 5/00 (2024.01) G06T 5/20 (2024.01)
G06T 7/10 (2021.01)

(52) CPC특허분류

G06T 5/50 (2024.01)
G06T 3/40 (2024.01)

(21) 출원번호 10-2023-0029990

(22) 출원일자 2023년03월07일

심사청구일자 2023년03월07일

(71) 출원인

울산과학기술원

울산광역시 울주군 언양읍 유니스트길 50

주식회사 피씨엠티

서울특별시 동대문구 청계천로 433, 13층 1309호
1310호(신설동, 미우골든오피스텔)

주식회사 시이안

서울특별시 종로구 북촌로12길 41 (가회동)

(72) 발명자

이지민

울산광역시 울주군 언양읍 유니스트길 50 (반연리, 울산과학기술)

조형주

경기도 성남시 분당구 내정로166번길 7-6 (수내동)

(뒷면에 계속)

(74) 대리인

특허법인 수

전체 청구항 수 : 총 30 항

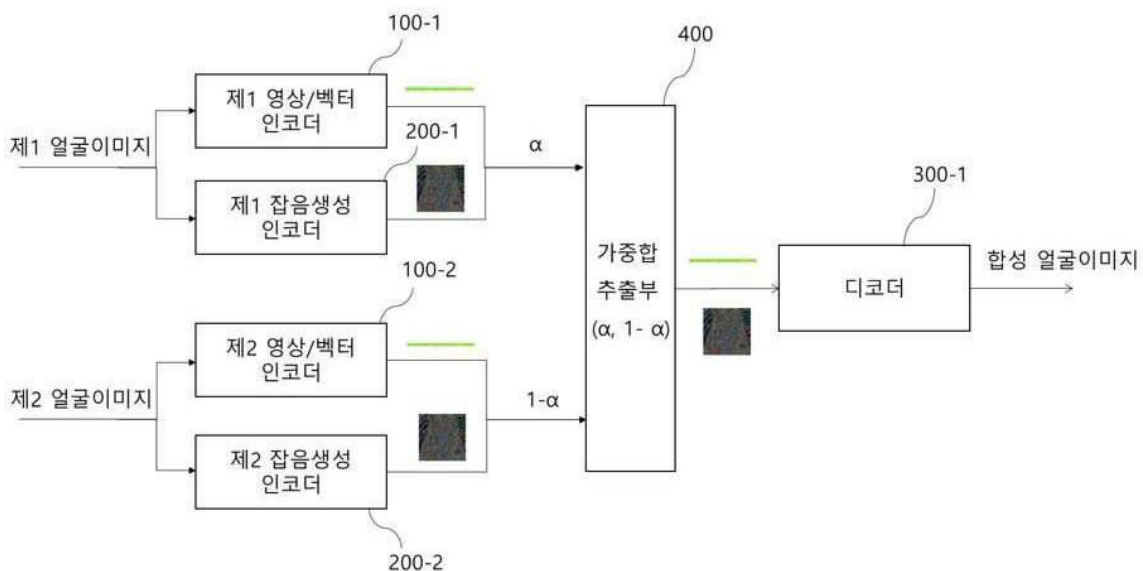
(54) 발명의 명칭 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 장치 및 방법

(57) 요약

본 발명에 따른 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 장치는 제1 얼굴이미지를 입력받아 제1 얼굴이미지로부터 제1 잠재 특징 벡터를 추출하는 제1 영상/벡터 인코더, 제1 영상/벡터 인코더, 제1 얼굴이미지를 입력받아 이를 잠재 공간에 사상시켜 제1 잠재 생성 영상을 추출하는 제1 잠재생성 인코더, 제2 얼굴이미지를 입력받아 제2 얼굴이미

(뒷면에 계속)

대표도



지로부터 제2 잠재 특징 벡터를 추출하는 제2 영상/벡터 인코더, 제2 얼굴이미지를 입력받아 이를 잠재공간에 사상시켜 제2 잡음 생성 영상을 추출하는 제2 잡음생성 인코더, 제1 얼굴이미지 및 제2 얼굴이미지에 대한 가중치를 토대로 제1 잠재 특징 벡터와 제2 잠재 특징 벡터에 대한 가중합 및 제1 잡음 생성 영상과 제2 잡음 생성 영상에 대한 가중합을 통해 신규 잠재공간(latent space) 분포를 생성함으로써 가중합 잠재 특징 벡터와 가중합 잡음 생성 영상을 추출하는 가중합 추출부, 및 추출된 가중합 잠재 특징 벡터와 가중합 잡음 생성 영상으로부터 가상의 합성 얼굴이미지를 생성하는 디코더를 포함할 수 있다.

(52) CPC특허분류

G06T 5/20 (2024.01)

G06T 5/70 (2024.01)

G06T 7/10 (2021.01)

G06T 2207/20081 (2013.01)

G06T 2207/20084 (2013.01)

G06T 2207/20221 (2013.01)

G06T 2207/30201 (2013.01)

(72) 발명자

여은영

서울특별시 종로구 난계로 259, 1101호 (송인동, 경일오피스텔)

이시안

서울특별시 종로구 북촌로12길 41 (가회동)

이부형

서울특별시 광진구 능동로24길 100 (능동)

명세서

청구범위

청구항 1

제1 얼굴이미지를 입력받아 제1 얼굴이미지로부터 제1 잠재 특징 벡터를 추출하는 제1 영상/벡터 인코더;

제1 얼굴이미지를 입력받아 이를 잠재공간에 사상시켜 제1 잡음 생성 영상을 추출하는 제1 잡음생성 인코더;

제2 얼굴이미지를 입력받아 제2 얼굴이미지로부터 제2 잠재 특징 벡터를 추출하는 제2 영상/벡터 인코더;

제2 얼굴이미지를 입력받아 이를 잠재공간에 사상시켜 제2 잡음 생성 영상을 추출하는 제2 잡음생성 인코더;

상기 제1 얼굴이미지 및 상기 제2 얼굴이미지에 대한 가중치를 토대로 상기 제1 잠재 특징 벡터와 상기 제2 잠재 특징 벡터에 대한 가중합 및 상기 제1 잡음 생성 영상과 상기 제2 잡음 생성 영상에 대한 가중합을 통해 신규 잠재공간(latent space) 분포를 생성함으로써 가중합 잠재 특징 벡터와 가중합 잡음 생성 영상을 추출하는 가중합 추출부; 및

추출된 상기 가중합 잠재 특징 벡터와 상기 가중합 잡음 생성 영상으로부터 가상의 합성 얼굴이미지를 생성하는 디코더;를 포함하는 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 장치.

청구항 2

제 1 항에 있어서,

상기 제1 영상/벡터 인코더 및 상기 제2 영상/벡터 인코더는 딥러닝 기반 영상 생성모델을 학습시키는 방법을 통해 학습된 영상/벡터 인코더 이고,

상기 제1 잡음생성 인코더 및 상기 제2 잡음생성 인코더는 상기 딥러닝 기반 영상 생성모델을 학습시키는 방법을 통해 학습된 잡음생성 인코더 이고,

상기 디코더는 상기 딥러닝 기반 영상 생성모델을 학습시키는 방법을 통해 학습된 디코더 인 것을 특징으로 하는 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 장치.

청구항 3

제 1 항에 있어서,

상기 제1 얼굴이미지 및 상기 제2 얼굴이미지에 대해 부여되는 가중치는 0과 1 사이의 값을 부여하는 것을 특징으로 하는 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 장치.

청구항 4

제 2 항에 있어서,

상기 딥러닝 기반 영상 생성모델은,

소정의 얼굴이미지를 상기 영상/벡터 인코더에 입력시켜 상기 소정의 얼굴이미지로부터 분할 영상과 잠재 특징 벡터를 추출하는 단계;

상기 소정의 얼굴이미지를 상기 잡음생성 인코더에 입력시키고 이를 잠재공간으로 사상시켜 잡음 생성 영상을 추출하는 단계; 및

추출된 상기 잠재 특징 벡터와 상기 잡음 생성 영상을 토대로 상기 디코더가 상기 소정의 얼굴이미지를 다시 생성하는 단계

에 기초하여 학습된 딥러닝 기반 영상 생성모델인 것을 특징으로 하는 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 장치.

청구항 5

제 4 항에 있어서,

상기 영상/벡터 인코더는,

입력된 상기 소정의 얼굴이미지에 대한 데이터 전처리를 수행하는 데이터 전처리부를 구비하는 것을 특징으로 하는 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 장치.

청구항 6

제 4 항에 있어서,

상기 영상/벡터 인코더는,

U-Net 기반의 인공지능 모델로서, 특성 맵(feature map)의 크기가 축소되었다가 확장되면서 상기 소정의 얼굴이미지에 대한 상기 분할 영상을 만들어내도록 학습되는 것을 특징으로 하는 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 장치.

청구항 7

제 4 항에 있어서,

상기 영상/벡터 인코더는,

상기 잠재 특징 벡터를 추출하기 위해 가장 작게 축소된 특성 맵(feature map)에 특정 연산(one linear layer)을 수행하여 상기 잠재 특징 벡터를 추출하는 것을 특징으로 하는 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 장치.

청구항 8

제 7 항에 있어서,

상기 영상/벡터 인코더는,

상기 잠재 특징 벡터를 추출하기 위해 VAE (Variational Auto Encoder) 모델을 사용하는 것을 특징으로 하는 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 장치.

청구항 9

제 4 항에 있어서,

상기 잠재 특징 벡터는,

상기 소정의 얼굴이미지에 내재되어 있는 특징을 1차원 공간에 담고 있으며, 하이레벨 레프리젠테이션(high-level representation)의 성격을 갖는 것을 특징으로 하는 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 장치.

청구항 10

제 4 항에 있어서,

상기 잡음생성 인코더는,

U-Net 기반의 인공지능 모델로서, 특성 맵(feature map)의 크기가 축소되었다가 확장되면서 상기 소정의 얼굴이미지에 대한 상기 잡음 생성 영상을 만들어내도록 학습되는 것을 특징으로 하는 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 장치.

청구항 11

제 4 항에 있어서,

상기 잡음 생성 영상은,

상기 소정의 얼굴이미지에 내재되어 있는 특징을 2차원 공간으로 담고 있으며, 로우레벨 레프리젠테이션(low-level representation)의 성격을 갖는 것을 특징으로 하는 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 장치.

성 장치.

청구항 12

제 4 항에 있어서,

상기 디코더는,

인공지능 모델 중 컨디셔널 디퓨전 모델(Conditional Diffusion model)로서, 추출된 상기 잠재 특징 벡터와 상기 잡음 생성 영상을 입력으로 받아 얼굴이미지를 생성해내도록 학습되는 것을 특징으로 하는 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 장치.

청구항 13

제 4 항에 있어서,

상기 디코더는,

이미지를 변환하는 과정에서 Conditional Generative Adversarial Networks(cGAN), Unpaired Image-to-Image Translation, Conditional Variational Autoencoder (cVAE), 및 Diffusion Probabilistic Model (DPM) 중 적어도 어느 하나의 기술이 활용되는 것을 특징으로 하는 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 장치.

청구항 14

제 13 항에 있어서,

상기 디코더는,

GAN 또는 Diffusion Inversion 기술로 잠재 공간(latent space)에서 특정 속성을 변환하는 것이 활용되는 것을 특징으로 하는 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 장치.

청구항 15

제 4 항에 있어서,

상기 영상/벡터 인코더는,

상기 분할 영상을 생성하도록 학습되는 과정에서 얼굴을 표현할 수 있는 특성(feature)들을 학습하도록 이루어지는 것을 특징으로 하는 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 장치.

청구항 16

제1 영상/벡터 인코더가 제1 얼굴이미지를 입력받아 상기 제1 얼굴이미지로부터 제1 잠재 특징 벡터를 추출하는 한편, 제1 잡음생성 인코더가 상기 제1 얼굴이미지를 입력받아 이를 잠재공간에 사상시켜 제1 잡음 생성 영상을 추출하는 제1 잠재특징벡터 및 잡음생성영상 추출단계;

제2 영상/벡터 인코더가 제2 얼굴이미지를 입력받아 상기 제2 얼굴이미지로부터 제2 잠재 특징 벡터를 추출하는 한편, 제2 잡음생성 인코더가 상기 제2 얼굴이미지를 입력받아 이를 잠재공간에 사상시켜 제2 잡음 생성 영상을 추출하는 제2 잠재특징벡터 및 잡음생성영상 추출단계;

가중합 산출부가 상기 제1 얼굴이미지 및 상기 제2 얼굴이미지에 대한 가중치를 토대로 상기 제1 잠재 특징 벡터와 상기 제2 잠재 특징 벡터에 대한 가중합 및 상기 제1 잡음 생성 영상과 상기 제2 잡음 생성 영상에 대한 가중합을 통해 신규 잠재공간(latent space) 분포를 생성함으로써 가중합 잠재 특징 벡터와 가중합 잡음 생성 영상을 추출하는 가중합 잠재특징벡터 및 잡음생성영상 추출단계; 및

디코더가 추출된 상기 가중합 잠재 특징 벡터와 상기 가중합 잡음 생성 영상으로부터 가상의 합성 얼굴이미지를 생성하는 합성 얼굴이미지 생성단계;를 포함하는 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 방법.

청구항 17

제 16 항에 있어서,

상기 제1 영상/벡터 인코더 및 상기 제2 영상/벡터 인코더는 딥러닝 기반 영상 생성모델을 학습시키는 방법을 통해 학습된 영상/벡터 인코더 이고,

상기 제1 잡음생성 인코더 및 상기 제2 잡음생성 인코더는 상기 딥러닝 기반 영상 생성모델을 학습시키는 방법을 통해 학습된 잡음생성 인코더 이고,

상기 디코더는 상기 딥러닝 기반 영상 생성모델을 학습시키는 방법을 통해 학습된 디코더 인 것을 특징으로 하는 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 방법.

청구항 18

제 16 항에 있어서,

상기 제1 얼굴이미지 및 상기 제2 얼굴이미지에 대해 부여되는 가중치는 0과 1 사이의 값을 부여하는 것을 특징으로 하는 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 방법.

청구항 19

제 17 항에 있어서,

상기 딥러닝 기반 영상 생성모델은,

소정의 얼굴이미지를 상기 영상/벡터 인코더에 입력시켜 상기 소정의 얼굴이미지로부터 분할 영상과 잠재 특징 벡터를 추출하는 단계;

상기 소정의 얼굴이미지를 상기 잡음생성 인코더에 입력시키고 이를 잠재공간으로 사상시켜 잡음 생성 영상을 추출하는 단계; 및

상기 잠재 특징 벡터와 상기 잡음 생성 영상을 토대로 상기 디코더가 상기 소정의 얼굴이미지를 다시 생성하는 단계

에 기초하여 학습된 딥러닝 기반 영상 생성모델인 것을 특징으로 하는 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 방법.

청구항 20

제 19 항에 있어서,

상기 영상/벡터 인코더는,

데이터 전처리부를 구비하여, 입력된 상기 소정의 얼굴이미지에 대한 데이터 전처리를 수행하는 것을 특징으로 하는 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 방법.

청구항 21

제 19 항에 있어서,

상기 영상/벡터 인코더는,

U-Net 기반의 인공지능 모델로서, 특성 맵(feature map)의 크기가 축소되었다가 확장되면서 상기 소정의 얼굴이미지에 대한 상기 분할 영상을 만들어내도록 학습되는 것을 특징으로 하는 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 방법.

청구항 22

제 19 항에 있어서,

상기 영상/벡터 인코더는,

상기 잠재 특징 벡터를 추출하기 위해 가장 작게 축소된 특성 맵(feature map)에 특정 연산(one linear layer)을 수행하여 상기 잠재 특징 벡터를 추출하는 것을 특징으로 하는 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 방법.

청구항 23

제 22 항에 있어서,

상기 영상/벡터 인코더는,

상기 잠재 특징 벡터를 추출하기 위해 VAE (Variational Auto Encoder) 모델을 사용하는 것을 특징으로 하는 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 방법.

청구항 24

제 19 항에 있어서,

상기 잠재 특징 벡터는,

상기 소정의 얼굴이미지에 내재되어 있는 특징을 1차원 공간에 담고 있으며, 하이레벨 레프리젠테이션(high-level representation)의 성격을 갖는 것을 특징으로 하는 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 방법.

청구항 25

제 19 항에 있어서,

상기 잡음생성 인코더는,

U-Net 기반의 인공지능 모델로서, 특성 맵(feature map)의 크기가 축소되었다가 확장되면서 상기 소정의 얼굴이미지에 대한 상기 잡음 생성 영상을 만들어내도록 학습되는 것을 특징으로 하는 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 방법.

청구항 26

제 19 항에 있어서,

상기 잡음 생성 영상은,

상기 소정의 얼굴이미지에 내재되어 있는 특징을 2차원 공간으로 담고 있으며, 로우레벨 레프리젠테이션(low-level representation)의 성격을 갖는 것을 특징으로 하는 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 방법.

청구항 27

제 19 항에 있어서,

상기 디코더는,

인공지능 모델 중 컨디셔널 디퓨전 모델(Conditional Diffusion model)로서, 추출된 상기 잠재 특징 벡터와 상기 잡음 생성 영상을 입력으로 받아 얼굴이미지를 생성해내도록 학습되는 것을 특징으로 하는 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 방법.

청구항 28

제 19 항에 있어서,

상기 디코더는,

이미지를 변환하는 과정에서 Conditional Generative Adversarial Networks(cGAN), Unpaired Image-to-Image Translation, Conditional Variational Autoencoder (cVAE), 및 Diffusion Probabilistic Model (DPM) 중 적어도 어느 하나의 기술이 활용되는 것을 특징으로 하는 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 방법.

청구항 29

제 28 항에 있어서,

상기 디코더는,

GAN 또는 Diffusion Inversion 기술로 잠재 공간(latent space)에서 특정 속성을 변환하는 것이 활용되는 것을 특징으로 하는 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 방법.

청구항 30

제 19 항에 있어서,

상기 영상/벡터 인코더는,

상기 분할 영상을 생성하도록 학습되는 과정에서 얼굴을 표현할 수 있는 특성(feature)들을 학습하도록 이루어 지는 것을 특징으로 하는 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 방법.

발명의 설명

기술 분야

[0001] 본 발명은 얼굴이미지를 합성하는 장치 및 방법에 관한 것으로서, 상세하게는, 인공지능(Artificial Intelligence) 기술 중 딥러닝(Deep learning)을 활용해 2개의 얼굴이미지를 합성하여 가상의 얼굴이미지를 생성하는 장치 및 방법에 관한 것이다.

배경 기술

[0003] 인공지능의 한 분야인 딥러닝(deep learning)을 이용한 이미지 합성 기술은 고해상도의 이미지를 생성해 낼 수 있을 정도로 높은 수준의 발전을 보여왔다.

[0004] 딥러닝 알고리즘은 인간 뇌의 뉴런들이 모여 신호를 전달하는 프로세스를 모델링하여 만든 알고리즘으로 인간의 뇌에서 진행되는 의사결정 과정을 모방한 인공 신경망(Artificial Neural Network) 구조를 통해 학습을 한다.

[0005] 딥러닝은 수집된 수많은 데이터를 사용하여 주어진 입력에 대해 올바른 출력을 위해 데이터 자체의 중요한 특징을 기계 스스로 학습을 한다. 따라서, 딥러닝은 성능 향상을 위해서 풍부한 학습 데이터를 확보하는 것이 중요한 요소 중 하나다.

[0006] 이전에는 사용자가 얼굴 이미지에 대한 보정을 직접 수행하였으므로 사용자에게 따라 보정 이미지의 퀄리티가 다른 것이 일반적이었다. 그러나, 수많은 데이터 학습을 통해 모델링된 딥러닝 기술을 이용하게 됨으로써 보다 자연스러운 얼굴이미지의 합성이 가능하게 되었으며, 이러한 딥러닝 기술을 이용하여 원하는 가상의 얼굴을 생성하기 위한 다양한 방법들이 제안되어 왔다.

[0007] 그러나, 종래의 경우 무작위 생성 방식에 기반하여 얼굴이미지를 생성하기 때문에 정해진 횟수 내에 원하는 특징을 갖는 얼굴이미지를 생성하는 것이 쉽지 않았고, 또한 사용자가 원하는 얼굴이미지가 생성될 것이라는 보장도 되지 않았다.

[0008] 이로 인해, 사용자가 원하는 얼굴 특징을 갖고 있는 가상의 얼굴이미지가 생성되기까지 상당한 그래픽 처리장치(Graphics Processing Unit)의 연산량이 요구되고, 많은 시간이 소요되는 문제점이 있다.

선행기술문헌

특허문헌

[0010] (특허문헌 0001) 대한민국 등록특허공보 제10-2417425호 (2022.07.01)

발명의 내용

해결하려는 과제

[0011] 본 발명은 이러한 종래기술의 문제점을 극복하기 위해 딥러닝 기반 영상 생성모델을 이용하여 2개의 얼굴 이미지로부터 사용자가 원하는 얼굴 특징을 갖는 가상의 얼굴이미지를 보다 적은 시도로 신속하게 생성할 수 있는 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 장치 및 방법을 제공하는 것을 목적으로 한다.

과제의 해결 수단

[0013] 본 발명의 일 실시예에 따른 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 장치는, 제1 얼굴이미지를 입력받아 제1 얼굴이미지로부터 제1 잠재 특징 벡터를 추출하는 제1 영상/벡터 인코더; 제1 얼굴이미지를 입력받아 이를 잠재공간에 사상시켜 제1 잡음 생성 영상을 추출하는 제1 잡음생성 인코더; 제2 얼굴이미지를 입력받아 제2 얼굴이미지로부터 제2 잠재 특징 벡터를 추출하는 제2 영상/벡터 인코더; 제2 얼굴이미지를 입력받아 이를 잠재공간에 사상시켜 제2 잡음 생성 영상을 추출하는 제2 잡음생성 인코더; 상기 제1 얼굴이미지 및 상기 제2 얼굴이미지에 대한 가중치를 토대로 상기 제1 잠재 특징 벡터와 상기 제2 잠재 특징 벡터에 대한 가중합 및 상기 제1 잡음 생성 영상과 상기 제2 잡음 생성 영상에 대한 가중합을 통해 신규 잠재공간(latent space) 분포를 생성함으로써 가중합 잠재 특징 벡터와 가중합 잡음 생성 영상을 추출하는 가중합 추출부; 및 추출된 상기 가중합 잠재 특징 벡터와 상기 가중합 잡음 생성 영상으로부터 가상의 합성 얼굴이미지를 생성하는 디코더;를 포함할 수 있다.

[0014] 여기서, 상기 제1 영상/벡터 인코더 및 상기 제2 영상/벡터 인코더는 딥러닝 기반 영상 생성모델을 학습시키는 방법을 통해 학습된 영상/벡터 인코더 이고, 상기 제1 잡음생성 인코더 및 상기 제2 잡음생성 인코더는 상기 딥러닝 기반 영상 생성모델을 학습시키는 방법을 통해 학습된 잡음생성 인코더 이고, 상기 디코더는 상기 딥러닝 기반 영상 생성모델을 학습시키는 방법을 통해 학습된 디코더 일 수 있다.

[0015] 이때, 상기 제1 얼굴이미지 및 상기 제2 얼굴이미지에 대해 부여되는 가중치는 0과 1 사이의 값을 부여할 수 있다.

[0016] 한편, 상기 딥러닝 기반 영상 생성모델은, 소정의 얼굴이미지를 영상/벡터 인코더에 입력시켜 상기 소정의 얼굴 이미지로부터 분할 영상과 잠재 특징 벡터를 추출하는 단계; 상기 소정의 얼굴이미지를 잡음생성 인코더에 입력시키고 이를 잠재공간으로 사상시켜 잡음 생성 영상을 추출하는 단계; 및 디코더가 추출된 상기 잠재 특징 벡터와 상기 잡음 생성 영상을 토대로 상기 소정의 얼굴이미지를 다시 생성하는 단계에 기초하여 학습된 딥러닝 기반 영상 생성모델일 수 있다.

[0017] 여기서, 상기 영상/벡터 인코더는, 입력된 상기 소정의 얼굴이미지에 대한 데이터 전처리를 수행하는 데이터 전처리부를 구비할 수 있다.

[0018] 이때, 상기 영상/벡터 인코더는, U-Net 기반의 인공지능 모델로서, 특성 맵(feature map)의 크기가 축소되었다가 확장되면서 상기 소정의 얼굴이미지에 대한 상기 분할 영상을 만들어내도록 학습될 수 있다.

[0019] 또한, 상기 영상/벡터 인코더는, 상기 잠재 특징 벡터를 추출하기 위해 가장 작게 축소된 특성 맵(feature map)에 특정 연산(one linear layer)을 수행하여 상기 잠재 특징 벡터를 추출할 수 있다.

[0020] 또한, 상기 영상/벡터 인코더는, 상기 잠재 특징 벡터를 추출하기 위해 VAE (Variational Auto Encoder) 모델을 사용할 수 있다.

[0021] 이때, 상기 잠재 특징 벡터는, 상기 소정의 얼굴이미지에 내재되어 있는 특징을 1차원 공간에 담고 있으며, 하イレ벨 레프리젠테이션(high-level representation)의 성격을 갖는 것일 수 있다.

[0022] 한편, 상기 잡음생성 인코더는, U-Net 기반의 인공지능 모델로서, 특성 맵(feature map)의 크기가 축소되었다가 확장되면서 상기 소정의 얼굴이미지에 대한 상기 잡음 생성 영상을 만들어내도록 학습될 수 있다.

[0023] 여기서, 상기 잡음 생성 영상은, 상기 소정의 얼굴이미지에 내재되어 있는 특징을 2차원 공간으로 담고 있으며, 로우레벨 레프리젠테이션(low-level representation)의 성격을 갖는 것일 수 있다.

[0024] 한편, 상기 디코더는, 인공지능 모델 중 컨디셔널 디퓨전 모델(Conditional Diffusion model)로서, 추출된 상기 잠재 특징 벡터와 상기 잡음 생성 영상을 입력으로 받아 얼굴이미지를 생성해내도록 학습될 수 있다.

[0025] 또한, 상기 디코더는, 이미지를 변환하는 과정에서 Conditional Generative Adversarial Networks(cGAN), Unpaired Image-to-Image Translation, Conditional Variational Autoencoder (cVAE), 및 Diffusion

Probabilistic Model (DPM) 중 적어도 어느 하나의 기술이 활용될 수 있다.

- [0026] 또한, 상기 디코더는, GAN 또는 Diffusion Inversion 기술로 잠재 공간(latent space)에서 특정 속성을 변환하는 것이 활용될 수 있다.
- [0027] 여기서, 상기 영상/벡터 인코더는, 상기 분할 영상을 생성하도록 학습되는 과정에서 얼굴을 표현할 수 있는 특성(feature)들을 학습하도록 이루어질 수 있다.
- [0028] 한편, 본 발명의 일 실시예에 따른 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 방법은, 제1 영상/벡터 인코더가 제1 얼굴이미지를 입력받아 상기 제1 얼굴이미지로부터 제1 잠재 특징 벡터를 추출하는 한편, 제1 잡음생성 인코더가 상기 제1 얼굴이미지를 입력받아 이를 잠재공간에 사상시켜 제1 잡음 생성 영상을 추출하는 제1 잠재특징벡터 및 잡음생성영상 추출단계; 제2 영상/벡터 인코더가 제2 얼굴이미지를 입력받아 상기 제2 얼굴이미지로부터 제2 잠재 특징 벡터를 추출하는 한편, 제2 잡음생성 인코더가 상기 제2 얼굴이미지를 입력받아 이를 잠재공간에 사상시켜 제2 잡음 생성 영상을 추출하는 제2 잠재특징벡터 및 잡음생성영상 추출단계; 가중합 산출부가 상기 제1 얼굴이미지 및 상기 제2 얼굴이미지에 대한 가중치를 토대로 상기 제1 잠재 특징 벡터와 상기 제2 잠재 특징 벡터에 대한 가중합 및 상기 제1 잡음 생성 영상과 상기 제2 잡음 생성 영상에 대한 가중합을 통해 신규 잠재공간(latent space) 분포를 생성함으로써 가중합 잠재 특징 벡터와 가중합 잡음 생성 영상을 추출하는 가중합 잠재특징벡터 및 잡음생성영상 추출단계; 및 디코더가 추출된 상기 가중합 잠재 특징 벡터와 상기 가중합 잡음 생성 영상으로부터 가상의 합성 얼굴이미지를 생성하는 합성 얼굴이미지 생성단계;를 포함할 수 있다.
- [0029] 여기서, 상기 제1 영상/벡터 인코더 및 상기 제2 영상/벡터 인코더는 딥러닝 기반 영상 생성모델을 학습시키는 방법을 통해 학습된 영상/벡터 인코더 이고, 상기 제1 잡음생성 인코더 및 상기 제2 잡음생성 인코더는 상기 딥러닝 기반 영상 생성모델을 학습시키는 방법을 통해 학습된 잡음생성 인코더 이고, 상기 디코더는 상기 딥러닝 기반 영상 생성모델을 학습시키는 방법을 통해 학습된 디코더 일 수 있다.
- [0030] 이때, 상기 제1 얼굴이미지 및 상기 제2 얼굴이미지에 대해 부여되는 가중치는 0과 1 사이의 값을 부여할 수 있다.
- [0031] 한편, 상기 딥러닝 기반 영상 생성모델은, 소정의 얼굴이미지를 영상/벡터 인코더에 입력시켜 상기 소정의 얼굴이미지로부터 분할 영상과 잠재 특징 벡터를 추출하는 단계; 상기 소정의 얼굴이미지를 잡음생성 인코더에 입력시키고 이를 잠재공간으로 사상시켜 잡음 생성 영상을 추출하는 단계; 및 디코더가 추출된 상기 잠재 특징 벡터와 상기 잡음 생성 영상을 토대로 상기 소정의 얼굴이미지를 다시 생성하는 단계에 기초하여 학습된 딥러닝 기반 영상 생성모델일 수 있다.
- [0032] 여기서, 상기 영상/벡터 인코더는, 입력된 상기 소정의 얼굴이미지에 대한 데이터 전처리를 수행하는 데이터 전처리부를 구비할 수 있다.
- [0033] 이때, 상기 영상/벡터 인코더는, U-Net 기반의 인공지능 모델로서, 특성 맵(feature map)의 크기가 축소되었다가 확장되면서 상기 소정의 얼굴이미지에 대한 상기 분할 영상을 만들어내도록 학습될 수 있다.
- [0034] 또한, 상기 영상/벡터 인코더는, 상기 잠재 특징 벡터를 추출하기 위해 가장 작게 축소된 특성 맵(feature map)에 특정 연산(one linear layer)을 수행하여 상기 잠재 특징 벡터를 추출할 수 있다.
- [0035] 또한, 상기 영상/벡터 인코더는, 상기 잠재 특징 벡터를 추출하기 위해 VAE (Variational Auto Encoder) 모델을 사용할 수 있다.
- [0036] 이때, 상기 잠재 특징 벡터는, 상기 소정의 얼굴이미지에 내재되어 있는 특징을 1차원 공간에 담고 있으며, 하イレ벨 레프리젠테이션(high-level representation)의 성격을 갖는 것일 수 있다.
- [0037] 한편, 상기 잡음생성 인코더는, U-Net 기반의 인공지능 모델로서, 특성 맵(feature map)의 크기가 축소되었다가 확장되면서 상기 소정의 얼굴이미지에 대한 상기 잡음 생성 영상을 만들어내도록 학습될 수 있다.
- [0038] 여기서, 상기 잡음 생성 영상은, 상기 소정의 얼굴이미지에 내재되어 있는 특징을 2차원 공간으로 담고 있으며, 로우레벨 레프리젠테이션(low-level representation)의 성격을 갖는 것일 수 있다.
- [0039] 한편, 상기 디코더는, 인공지능 모델 중 컨디셔널 디퓨전 모델(Conditional Diffusion model)로서, 추출된 상기 잠재 특징 벡터와 상기 잡음 생성 영상을 입력으로 받아 얼굴이미지를 생성해내도록 학습될 수 있다.

- [0040] 또한, 상기 디코더는, 이미지를 변환하는 과정에서 Conditional Generative Adversarial Networks(cGAN), Unpaired Image-to-Image Translation, Conditional Variational Autoencoder (cVAE), 및 Diffusion Probabilistic Model (DPM) 중 적어도 어느 하나의 기술이 활용될 수 있다.
- [0041] 또한, 상기 디코더는, GAN 또는 Diffusion Inversion 기술로 잠재 공간(latent space)에서 특정 속성을 변환하는 것이 활용될 수 있다.
- [0042] 여기서, 상기 영상/벡터 인코더는, 상기 분할 영상을 생성하도록 학습되는 과정에서 얼굴을 표현할 수 있는 특성(feature)들을 학습하도록 이루어질 수 있다.

발명의 효과

- [0044] 본 발명에 따른 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 장치 및 방법은 딥러닝 기반 영상 생성 모델을 이용하여 2개의 얼굴 이미지로부터 사용자가 원하는 얼굴 특징을 갖는 가상의 얼굴이미지를 보다 적은 시도로 신속하게 생성할 수 있다.

도면의 간단한 설명

- [0046] 도 1은 본 발명의 일 실시예에 따른 신경망(neural network)을 나타낸 개략도이다.
- 도 2는 본 발명의 일 실시예에 따른 딥러닝 기반 영상 생성모델을 학습시키는 방법을 설명하기 위한 도면이다.
- 도 3은 본 발명의 일 실시예에 따른 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 장치를 도시한 도면이다.
- 도 4는 본 발명의 다른 실시예에 따른 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 방법을 도시한 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0047] 다양한 실시예들이 이제 도면을 참조하여 설명된다. 본 명세서에서, 다양한 설명들이 본 발명의 이해를 제공하기 위해서 제시된다. 그러나, 이러한 실시예들은 이러한 구체적인 설명 없이도 실행될 수 있음이 명백하다.
- [0048] 본 명세서에서 사용되는 용어 "컴포넌트", "모듈", "시스템" 등은 컴퓨터-관련 엔티티, 하드웨어, 펌웨어, 소프트웨어, 소프트웨어 및 하드웨어의 조합, 또는 소프트웨어의 실행을 지칭한다. 예를 들어, 컴포넌트는 프로세서 상에서 실행되는 처리과정(procedure), 프로세서, 객체, 실행 스레드, 프로그램, 및/또는 컴퓨터일 수 있지만, 이들로 제한되는 것은 아니다. 예를 들어, 컴퓨팅 장치에서 실행되는 애플리케이션 및 컴퓨팅 장치 모두 컴포넌트일 수 있다. 하나 이상의 컴포넌트는 프로세서 및/또는 실행 스레드 내에 상주할 수 있다. 일 컴포넌트는 하나의 컴퓨터 내에 로컬화 될 수 있다. 일 컴포넌트는 2개 이상의 컴퓨터들 사이에 분배될 수 있다. 또한, 이러한 컴포넌트들은 그 내부에 저장된 다양한 데이터 구조들을 갖는 다양한 컴퓨터 판독가능한 매체로부터 실행할 수 있다. 컴포넌트들은 예를 들어 하나 이상의 데이터 패킷들을 갖는 신호(예를 들면, 로컬 시스템, 분산 시스템에서 다른 컴포넌트와 상호작용하는 하나의 컴포넌트로부터의 데이터 및/또는 신호를 통해 다른 시스템과 인터넷과 같은 네트워크를 통해 전송되는 데이터)에 따라 로컬 및/또는 원격 처리들을 통해 통신할 수 있다.
- [0049] 더불어, 용어 "또는"은 배타적 "또는"이 아니라 내포적 "또는"을 의미하는 것으로 의도된다. 즉, 달리 특정되지 않거나 문맥상 명확하지 않은 경우에, "X는 A 또는 B를 이용한다"는 자연적인 내포적 치환 중 하나를 의미하는 것으로 의도된다. 즉, X가 A를 이용하거나; X가 B를 이용하거나; 또는 X가 A 및 B 모두를 이용하는 경우, "X는 A 또는 B를 이용한다"가 이들 경우들 어느 것으로도 적용될 수 있다. 또한, 본 명세서에 사용된 "및/또는"이라는 용어는 열거된 관련 아이템들 중 하나 이상의 아이템의 가능한 모든 조합을 지칭하고 포함하는 것으로 이해되어야 한다.
- [0050] 또한, "포함한다" 및/또는 "포함하는"이라는 용어는, 해당 특징 및/또는 구성요소가 존재함을 의미하는 것으로 이해되어야 한다. 다만, "포함한다" 및/또는 "포함하는"이라는 용어는, 하나 이상의 다른 특징, 구성요소 및/또는 이들의 그룹의 존재 또는 추가를 배제하지 않는 것으로 이해되어야 한다. 또한, 달리 특정되지 않거나 단수 형태를 지시하는 것으로 문맥상 명확하지 않은 경우에, 본 명세서와 청구범위에서 단수는 일반적으로 "하나 또는

는 그 이상"을 의미하는 것으로 해석되어야 한다.

- [0051] 그리고, "A 또는 B 중 적어도 하나"이라는 용어는, "A만을 포함하는 경우", "B 만을 포함하는 경우", "A와 B의 구성으로 조합된 경우"를 의미하는 것으로 해석되어야 한다.
- [0052] 본 발명에서 네트워크 함수와 인공 신경망 및 뉴럴 네트워크(neural network)는 상호교환 가능하게 사용될 수 있다.
- [0053] 이하, 첨부된 도면을 참조하여 본 발명에 따른 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 장치 및 방법에 대해 상세히 설명한다.
- [0055] 먼저, 도 1은 본 발명의 일 실시예에 따른 신경망(neural network)을 나타낸 개략도이다.
- [0056] 본 명세서에 걸쳐, 연산 모델, 신경망, 네트워크 함수, 뉴럴 네트워크(neural network)는 동일한 의미로 사용될 수 있다. 신경망은 일반적으로 노드라 지칭될 수 있는 상호 연결된 계산 단위들의 집합으로 구성될 수 있다. 이러한 노드들은 뉴런(neuron)들로 지칭될 수도 있다. 신경망은 적어도 하나 이상의 노드들을 포함하여 구성된다. 신경망들을 구성하는 노드(또는 뉴런)들은 하나 이상의 링크에 의해 상호 연결될 수 있다.
- [0057] 신경망 내에서, 링크를 통해 연결된 하나 이상의 노드들은 상대적으로 입력 노드 및 출력 노드의 관계를 형성할 수 있다. 입력 노드 및 출력 노드의 개념은 상대적인 것으로서, 하나의 노드에 대하여 출력 노드 관계에 있는 임의의 노드는 다른 노드와의 관계에서 입력 노드 관계에 있을 수 있으며, 그 역도 성립할 수 있다. 상술한 바와 같이, 입력 노드 대 출력 노드 관계는 링크를 중심으로 생성될 수 있다. 하나의 입력 노드에 하나 이상의 출력 노드가 링크를 통해 연결될 수 있으며, 그 역도 성립할 수 있다.
- [0058] 하나의 링크를 통해 연결된 입력 노드 및 출력 노드 관계에서, 출력 노드의 데이터는 입력 노드에 입력된 데이터에 기초하여 그 값이 결정될 수 있다. 여기서 입력 노드와 출력 노드를 상호 연결하는 링크는 가중치(weight)를 가질 수 있다. 가중치는 가변적일 수 있으며, 신경망이 원하는 기능을 수행하기 위해, 사용자 또는 알고리즘에 의해 가변 될 수 있다. 예를 들어, 하나의 출력 노드에 하나 이상의 입력 노드가 각각의 링크에 의해 상호 연결된 경우, 출력 노드는 상기 출력 노드와 연결된 입력 노드들에 입력된 값들 및 각각의 입력 노드들에 대응하는 링크에 설정된 가중치에 기초하여 출력 노드 값을 결정할 수 있다.
- [0059] 상술한 바와 같이, 신경망은 하나 이상의 노드들이 하나 이상의 링크를 통해 상호 연결되어 신경망 내에서 입력 노드 및 출력 노드 관계를 형성한다. 신경망 내에서 노드들과 링크들의 개수 및 노드들과 링크들 사이의 연관관계, 링크들 각각에 부여된 가중치의 값에 따라, 신경망의 특성이 결정될 수 있다. 예를 들어, 동일한 개수의 노드 및 링크들이 존재하고, 링크들의 가중치 값이 상이한 두 신경망이 존재하는 경우, 두 개의 신경망들은 서로 상이한 것으로 인식될 수 있다.
- [0060] 신경망은 하나 이상의 노드들의 집합으로 구성될 수 있다. 신경망을 구성하는 노드들의 부분 집합은 레이어(layer)를 구성할 수 있다. 신경망을 구성하는 노드들 중 일부는, 최초 입력 노드로부터의 거리에 기초하여, 하나의 레이어(layer)를 구성할 수 있다. 예를 들어, 최초 입력 노드로부터 거리가 n 인 노드들의 집합은, n 레이어를 구성할 수 있다. 최초 입력 노드로부터 거리는, 최초 입력 노드로부터 해당 노드까지 도달하기 위해 거쳐야 하는 링크들의 최소 개수에 의해 정의될 수 있다. 그러나, 이러한 레이어의 정의는 설명을 위한 임의적인 것으로서, 신경망 내에서 레이어의 차수는 상술한 것과 상이한 방법으로 정의될 수 있다. 예를 들어, 노드들의 레이어는 최종 출력 노드로부터 거리에 의해 정의될 수도 있다.
- [0061] 최초 입력 노드는 신경망 내의 노드들 중 다른 노드들과의 관계에서 링크를 거치지 않고 데이터가 직접 입력되는 하나 이상의 노드들을 의미할 수 있다. 또는, 신경망 네트워크 내에서, 링크를 기준으로 한 노드 간의 관계에 있어서, 링크로 연결된 다른 입력 노드들을 가지지 않는 노드들을 의미할 수 있다. 이와 유사하게, 최종 출력 노드는 신경망 내의 노드들 중 다른 노드들과의 관계에서, 출력 노드를 가지지 않는 하나 이상의 노드들을 의미할 수 있다. 또한, 히든 노드는 최초 입력 노드 및 최종 출력 노드가 아닌 신경망을 구성하는 노드들을 의미할 수 있다.
- [0062] 본 발명의 일 실시예에 따른 신경망은 입력 레이어의 노드의 개수가 출력 레이어의 노드의 개수와 동일할 수 있으며, 입력 레이어에서 히든 레이어로 진행됨에 따라 노드의 수가 감소하다가 다시 증가하는 형태의 신경망일 수 있다. 또한, 본 발명의 다른 일 실시예에 따른 신경망은 입력 레이어의 노드의 개수가 출력 레이어의 노드의 개수 보다 적을 수 있으며, 입력 레이어에서 히든 레이어로 진행됨에 따라 노드의 수가 감소하는 형태의 신경망

일 수 있다. 또한, 본 발명의 또 다른 일 실시예에 따른 신경망은 입력 레이어의 노드의 개수가 출력 레이어의 노드의 개수보다 많을 수 있으며, 입력 레이어에서 히든 레이어로 진행됨에 따라 노드의 수가 증가하는 형태의 신경망일 수 있다. 본 발명의 또 다른 일 실시예에 따른 신경망은 상술한 신경망들의 조합된 형태의 신경망일 수 있다.

[0063] 딥 뉴럴 네트워크(DNN: deep neural network, 심층신경망)는 입력 레이어와 출력 레이어 외에 복수의 히든 레이어를 포함하는 신경망을 의미할 수 있다. 딥 뉴럴 네트워크를 이용하면 데이터의 잠재적인 구조(latent structures)를 파악할 수 있다. 즉, 사진, 글, 비디오, 음성, 음악의 잠재적인 구조(예를 들어, 어떤 물체가 사진에 있는지, 글의 내용과 감정이 무엇인지, 음성의 내용과 감정이 무엇인지 등)를 파악할 수 있다. 딥 뉴럴 네트워크는 컨볼루션 뉴럴 네트워크(CNN: convolutional neural network), 리커런트 뉴럴 네트워크(RNN: recurrent neural network), 오토 인코더(auto encoder), GAN(Generative Adversarial Networks), 제한 볼츠만 머신(RBM: restricted boltzmann machine), 심층 신뢰 네트워크(DBN: deep belief network), Q 네트워크, U 네트워크, 삼 네트워크, 적대적 생성 네트워크(GAN: Generative Adversarial Network) 등을 포함할 수 있다. 전술한 딥 뉴럴 네트워크의 기재는 예시일 뿐이며 본 발명은 이에 제한되지 않는다.

[0064] 본 발명의 일 실시예에서 네트워크 함수는 오토 인코더(autoencoder)를 포함할 수도 있다. 오토 인코더는 입력 데이터와 유사한 출력 데이터를 출력하기 위한 인공 신경망의 일종일 수 있다. 오토 인코더는 적어도 하나의 히든 레이어를 포함할 수 있으며, 홀수 개의 히든 레이어가 입출력 레이어 사이에 배치될 수 있다. 각각의 레이어의 노드의 수는 입력 레이어의 노드의 수에서 병목 레이어(인코딩)라는 중간 레이어로 축소되었다가, 병목 레이어에서 출력 레이어(입력 레이어와 대칭)로 축소와 대칭되어 확장될 수도 있다. 오토 인코더는 비선형 차원 감소를 수행할 수 있다. 입력 레이어 및 출력 레이어의 수는 입력 데이터의 전처리 이후에 차원과 대응될 수 있다. 오토 인코더 구조에서 인코더에 포함된 히든 레이어의 노드의 수는 입력 레이어에서 멀어질수록 감소하는 구조를 가질 수 있다. 병목 레이어(인코더와 디코더 사이에 위치하는 가장 적은 노드를 가진 레이어)의 노드의 수는 너무 작은 경우 충분한 양의 정보가 전달되지 않을 수 있으므로, 특정 수 이상(예를 들어, 입력 레이어의 절반 이상 등)으로 유지될 수도 있다.

[0065] 뉴럴 네트워크는 교사 학습(supervised learning), 비교사 학습(unsupervised learning), 반교사학습(semi supervised learning), 또는 강화 학습(reinforcement learning) 중 적어도 하나의 방식으로 학습될 수 있다. 뉴럴 네트워크의 학습은 뉴럴 네트워크가 특정한 동작을 수행하기 위한 지식을 뉴럴 네트워크에 적용하는 과정일 수 있다.

[0066] 뉴럴 네트워크는 출력의 오류를 최소화하는 방향으로 학습될 수 있다. 뉴럴 네트워크의 학습에서 반복적으로 학습 데이터를 뉴럴 네트워크에 입력시키고 학습 데이터에 대한 뉴럴 네트워크의 출력과 타겟의 에러를 계산하고, 에러를 줄이기 위한 방향으로 뉴럴 네트워크의 에러를 뉴럴 네트워크의 출력 레이어에서부터 입력 레이어 방향으로 역전파(backpropagation)하여 뉴럴 네트워크의 각 노드의 가중치를 업데이트 하는 과정이다. 교사 학습의 경우 각각의 학습 데이터에 정답이 라벨링되어있는 학습 데이터를 사용하며(즉, 라벨링된 학습 데이터), 비교사 학습의 경우는 각각의 학습 데이터에 정답이 라벨링되어 있지 않을 수 있다. 즉, 예를 들어 데이터 분류에 관한 교사 학습의 경우의 학습 데이터는 학습 데이터 각각에 카테고리가 라벨링 된 데이터 일 수 있다. 라벨링된 학습 데이터가 뉴럴 네트워크에 입력되고, 뉴럴 네트워크의 출력(카테고리)과 학습 데이터의 라벨을 비교함으로써 오류(error)가 계산될 수 있다. 다른 예로, 데이터 분류에 관한 비교사 학습의 경우 입력인 학습 데이터가 뉴럴 네트워크 출력과 비교됨으로써 오류가 계산될 수 있다. 계산된 오류는 뉴럴 네트워크에서 역방향(즉, 출력 레이어에서 입력 레이어 방향)으로 역전파 되며, 역전파에 따라 뉴럴 네트워크의 각 레이어의 각 노드들의 연결 가중치가 업데이트 될 수 있다. 업데이트 되는 각 노드의 연결 가중치는 학습률(learning rate)에 따라 변화량이 결정될 수 있다. 입력 데이터에 대한 뉴럴 네트워크의 계산과 에러의 역전파는 학습 사이클(epoch)을 구성할 수 있다. 학습률은 뉴럴 네트워크의 학습 사이클의 반복 횟수에 따라 상이하게 적용될 수 있다. 예를 들어, 뉴럴 네트워크의 학습 초기에는 높은 학습률을 사용하여 뉴럴 네트워크가 빠르게 일정 수준의 성능을 확보하도록 하여 효율성을 높이고, 학습 후기에는 낮은 학습률을 사용하여 정확도를 높일 수 있다.

[0067] 뉴럴 네트워크의 학습에서 일반적으로 학습 데이터는 실제 데이터(즉, 학습된 뉴럴 네트워크를 이용하여 처리하고자 하는 데이터)의 부분집합일 수 있으며, 따라서, 학습 데이터에 대한 오류는 감소하나 실제 데이터에 대해서는 오류가 증가하는 학습 사이클이 존재할 수 있다. 과적합(overfitting)은 이와 같이 학습 데이터에 과하게 학습하여 실제 데이터에 대한 오류가 증가하는 현상이다. 예를 들어, 노란색 고양이를 보여 고양이를 학습한 뉴럴 네트워크가 노란색 이외의 고양이를 보고는 고양이임을 인식하지 못하는 현상이 과적합의 일종일 수 있다. 과적합은 머신러닝 알고리즘의 오류를 증가시키는 원인으로 작용할 수 있다. 이러한 과적합을 막기 위하여 다양

한 최적화 방법이 사용될 수 있다. 과적합을 막기 위해서는 학습 데이터를 증가시키거나, 레귤라이제이션(regularization), 학습의 과정에서 네트워크의 노드 일부를 비활성화하는 드롭아웃(dropout), 배치 정규화 레이어(batch normalization layer)의 활용 등의 방법이 적용될 수 있다.

- [0068] 강화 학습은 신경망 모델이 상태(state)에 기초하여 보다 나은 행동(action)을 결정할 수 있도록, 신경망 모델이 선택한 행동에 대해 산출되는 보상(reward)에 기초하여 신경망 모델을 학습시키는 학습 방법이다. 상태는 현재 시점에서 상황이 어떠한 지를 나타내는 값의 집합으로써, 신경망 모델의 입력으로 이해될 수 있다. 행동은 신경망 모델이 취할 수 있는 선택지에 따른 결정을 일컫는 말로, 신경망 모델의 출력으로 이해될 수 있다. 보상은 신경망 모델이 어떠한 행동을 수행했을 때 따라오는 이득을 말하며, 현재 상태 및 행동에 대해 평가하는 값을 나타낸다. 강화 학습은 행동에 대해 보상이 주어진다 점에서 시행착오를 통한 학습으로 이해될 수 있다. 강화 학습 과정에서 신경망 모델에게 주어지는 보상은 여러 행동의 결과가 누적된 보상일 수 있다. 강화 학습을 통해 여러가지 상태와 행동에 따른 보상을 고려하여, 보상 그 자체 또는 보상의 총 합과 같은 리턴(return)이 최대가 되도록 하는 신경망 모델을 생성할 수 있다.
- [0070] 다음, 도 2는 본 발명의 딥러닝 기반 영상 생성모델을 학습시키는 방법을 설명하기 위한 도면이다.
- [0071] 본 발명의 딥러닝 기반 영상 생성모델로 사용될 수 있는 알고리즘은 합성곱 신경망(CNN: Convolution Neural Network), 적대적 생성 신경망(GAN: Generative Adversarial Networks), 확산 확률 모형(DPM: Diffusion Probabilistic Model) 등을 적용할 수 있으며, 개시한 알고리즘에 한정되지 않고, 그 외 다양한 딥러닝 알고리즘의 적용이 가능함은 물론이다.
- [0072] 도 2를 참조하면, 본 발명에서는 소정의 얼굴이미지를 영상/벡터 인코더(100)에 입력시켜 소정의 얼굴이미지로부터 분할 영상과 잠재 특징 벡터를 추출한다.
- [0073] 좀더 구체적으로, 영상/벡터 인코더(100)는 U-Net 기반의 인공지능 모델로서, 특성 맵(feature map)의 크기가 축소되었다가 확장되면서 소정의 얼굴이미지에 대한 분할 영상을 만들어내도록 학습된다.
- [0074] 또한, 영상/벡터 인코더(100)는 잠재 특징 벡터를 추출하기 위해 가장 작게 축소된 특성 맵(feature map)에 특정 연산(one linear layer)을 수행하여 잠재 특징 벡터를 추출하며, 이때 VAE (Variational Auto Encoder) 등의 모델을 사용할 수 있다.
- [0075] 이와 같이 추출된 잠재 특징 벡터는 소정의 얼굴이미지에 내재되어 있는 특징을 1차원 공간에 담고 있으며, 하이레벨 레프리젠테이션(high-level representation)의 성격을 갖는다.
- [0076] 한편, 잡음생성 인코더(200)에도 상기 소정의 얼굴이미지를 입력시키고 이를 잠재공간으로 사상시켜 잡음 생성 영상을 추출한다.
- [0077] 좀더 구체적으로, 잡음생성 인코더(200)는 U-Net 기반의 인공지능 모델로서, 특성 맵(feature map)의 크기가 축소되었다가 확장되면서 소정의 얼굴이미지에 대한 잡음 생성 영상을 만들어내도록 학습된다.
- [0078] 이와 같이 추출된 잡음 생성 영상은 소정의 얼굴이미지에 내재되어 있는 특징을 2차원 공간으로 담고 있으며, 로우레벨 레프리젠테이션(low-level representation)의 성격을 갖는다.
- [0079] 이후, 디코더(300)는 추출된 잠재 특징 벡터와 잡음 생성 영상을 토대로 상기 소정의 얼굴이미지를 다시 생성한다.
- [0080] 이때, 디코더(300)는 인공지능 모델 중 컨디셔널 디퓨전 모델(Conditional Diffusion model)로서, 추출된 잠재 특징 벡터와 잡음 생성 영상을 입력으로 받아 얼굴이미지를 생성해내도록 학습된다.
- [0081] 이미지를 변환하는 과정에서 Conditional Generative Adversarial Networks(cGAN), Unpaired Image-to-Image Translation, Conditional Variational Autoencoder (cVAE), Diffusion Probabilistic Model (DPM) 등의 기술이 활용될 수 있으며, 또한 GAN 혹은 Diffusion Inversion 기술로 잠재 공간(latent space)에서 특정 속성을 변환하는 것도 활용될 수 있다.
- [0082] 한편, 본 발명의 딥러닝 기반 영상 생성모델에서, 디코더(300)가 상기 소정의 얼굴이미지를 다시 생성할 때 분할 영상을 사용하지 않음에도 영상/벡터 인코더(100)에서 분할 영상을 생성하는 이유는, 분할 영상을 생성하도록 영상/벡터 인코더(100)가 학습되는 과정에서 얼굴을 표현할 수 있는 적절한 특성(feature)들을 영상/벡터 인

코더(100)가 효과적으로 학습할 수 있기 때문이다. 즉, 영상/벡터 인코더(100)가 분할 영상을 적절하게 생성하게 하는 과정은 영상/벡터 인코더(100)가 잠재 특징 벡터를 적절하게 생성하는데 큰 도움을 준다.

- [0083] 이상과 같은 본 발명의 딥러닝 기반 영상 생성모델에 대한 학습 시에는, 분할 영상과 잠재 특징 벡터, 및 잡음 생성 영상을 혼합해 인코더에 입력된 소정의 얼굴이미지를 다시 형상화하는 방식으로 학습이 이루어진다.
- [0084] 딥러닝은 성능 향상을 위해서 풍부한 학습 데이터를 확보하는 것이 중요한 요소이므로, 본 발명에서 이용하는 딥러닝 기반 영상 생성모델을 학습시키기 위한 얼굴이미지는 그 수가 많을수록 바람직하다.
- [0085] 또한, 이러한 얼굴이미지에 대한 데이터 학습 방식은 기본적으로 지도학습(Supervised learning) 기반의 얼굴을 구분하는 분류(classification) 및 벡터를 예측하는 회귀생성(regression) 방식 등으로 수행될 수 있으나, 비지도학습(Unsupervised learning) 기반의 군집화(clustering)를 통해서 학습을 수행할 수도 있다.
- [0086] 구현예에 따라서, 영상/벡터 인코더(100)는 데이터 전처리부(미도시)를 구비하여 입력된 소정의 얼굴이미지에 대한 데이터 전처리를 수행할 수 있으며 얼굴을 추출하여 얼굴 영역, 위치, 색상, 밝기 중 적어도 하나를 전처리하여 특징점 검출을 용이하도록 조정할 수 있다. 또한, 이러한 데이터 전처리부는 합성을 위해 데이터 증폭 및 이미지 크기 정렬을 수행할 수 있으며, 이미지 크기를 기 설정된 크기로 정렬할 수 있다.
- [0087] 이후, 영상/벡터 인코더(100)는 전처리된 데이터에서 잠재 특징 벡터를 추출할 수 있으며, 즉 전처리된 얼굴이미지에서 눈, 코, 입, 귀 중 적어도 하나의 위치를 검출하고 이를 특징점으로 설정한 후 특징점들의 좌표값들을 연결하여 잠재 특징 벡터를 추출할 수 있다.
- [0088] 구현예에 따라서, 눈의 양쪽 끝 지점이나, 눈동자의 가운데 지점, 눈썹의 양쪽 끝 지점, 코의 가운데 지점, 입의 양쪽 끝 지점, 귀의 위아래 끝 지점 등과 같이 기 설정된 지점을 특징점으로 설정할 수 있다.
- [0090] 도 3은 본 발명의 일 실시예에 따른 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 장치를 도시한 도면이다.
- [0091] 도 3에서 볼 수 있는 바와 같이, 본 발명의 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 장치는 제1 얼굴이미지를 입력받아 제1 얼굴이미지로부터 제1 잠재 특징 벡터를 추출하는 제1 영상/벡터 인코더(100-1), 제1 얼굴이미지를 입력받아 이를 잠재공간에 사상시켜 제1 잡음 생성 영상을 추출하는 제1 잡음생성 인코더(200-1), 제2 얼굴이미지를 입력받아 제2 얼굴이미지로부터 제2 잠재 특징 벡터를 추출하는 제2 영상/벡터 인코더(100-2), 제2 얼굴이미지를 입력받아 이를 잠재공간에 사상시켜 제2 잡음 생성 영상을 추출하는 제2 잡음생성 인코더(200-2), 제1 얼굴이미지 및 제2 얼굴이미지에 대한 가중치(α , $1-\alpha$)를 토대로 제1 잠재 특징 벡터와 제2 잠재 특징 벡터에 대한 가중합 및 제1 잡음 생성 영상과 제2 잡음 생성 영상에 대한 가중합을 통해 신규 잠재공간(latent space) 분포를 생성함으로써 가중합 잠재 특징 벡터와 가중합 잡음 생성 영상을 추출하는 가중합 추출부(400), 및 추출된 가중합 잠재 특징 벡터와 가중합 잡음 생성 영상으로부터 가상의 합성 얼굴이미지를 생성하는 디코더(300-1)로 이루어진다.
- [0092] 여기서, 제1 영상/벡터 인코더(100-1) 및 제2 영상/벡터 인코더(100-2)는 도 2에서 설명한 본 발명의 딥러닝 기반 영상 생성모델을 학습시키는 방법을 통해 학습된 영상/벡터 인코더 이고, 제1 잡음생성 인코더(200-1) 및 제2 잡음생성 인코더(200-2)는 도 2에서 설명한 본 발명의 딥러닝 기반 영상 생성모델을 학습시키는 방법을 통해 학습된 잡음생성 인코더 이다. 또한, 디코더(300-1)는 도 2에서 설명한 본 발명의 딥러닝 기반 영상 생성모델을 학습시키는 방법을 통해 학습된 디코더 이다.
- [0093] 이때, 제1 얼굴이미지 및 제2 얼굴이미지에 대해 부여되는 가중치(α , $1-\alpha$)는 0과 1 사이의 값을 부여할 수 있으며, 일례로 제1 얼굴이미지의 가중치(α)가 0.3 이라면, 제1 얼굴이미지가 30%, 제2 얼굴이미지가 70% 반영되게 된다.
- [0094] 본 발명은 이를 통해 사용자가 원하는 얼굴 특징을 갖는 2개의 얼굴이미지를 본 발명의 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 장치의 입력으로 하는 한편, 각 얼굴이미지에 대해 반영하고 싶은 비율을 상기 가중치를 통해 설정하여 해당 비율만큼 반영된 가상의 합성 얼굴이미지를 생성할 수 있게 된다.
- [0095] 구현예에 따라서, 생성된 가상의 합성 얼굴이미지와, 사용자가 원하는 또다른 특징을 갖는 제3 얼굴이미지를 재차 합성할 수도 있으며, 이를 통해 본 발명은 빠른 시간 안에 사용자가 원하는 특징을 갖는 가상의 합성 얼굴이미지를 최종적으로 생성할 수도 있다.

- [0097] 다음, 도 4는 본 발명의 다른 실시예에 따른 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 방법을 도시한 도면이다.
- [0098] 도 4에서와 같이, 본 발명에 따른 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 방법은 제1 영상/벡터 인코더(100-1)가 제1 얼굴이미지를 입력받아 제1 얼굴이미지로부터 제1 잠재 특징 벡터를 추출하는 한편, 제1 잡음생성 인코더(200-1)가 제1 얼굴이미지를 입력받아 이를 잠재공간에 사상시켜 제1 잡음 생성 영상을 추출하는 제1 잠재 특징 벡터 및 잡음생성영상 추출단계(S100), 제2 영상/벡터 인코더(100-2)가 제2 얼굴이미지를 입력받아 제2 얼굴이미지로부터 제2 잠재 특징 벡터를 추출하는 한편, 제2 잡음생성 인코더(200-2)가 제2 얼굴이미지를 입력받아 이를 잠재공간에 사상시켜 제2 잡음 생성 영상을 추출하는 제2 잠재 특징 벡터 및 잡음생성영상 추출단계(S200), 가중합 추출부(400)가 제1 얼굴이미지 및 제2 얼굴이미지에 대한 가중치(α , $1-\alpha$)를 토대로 제1 잠재 특징 벡터와 제2 잠재 특징 벡터에 대한 가중합 및 제1 잡음 생성 영상과 제2 잡음 생성 영상에 대한 가중합을 통해 신규 잠재공간(latent space) 분포를 생성함으로써 가중합 잠재 특징 벡터와 가중합 잡음 생성 영상을 추출하는 가중합 잠재 특징 벡터 및 잡음생성영상 추출단계(S300), 및 디코더(300-1)가 추출된 가중합 잠재 특징 벡터와 가중합 잡음 생성 영상으로부터 가상의 합성 얼굴이미지를 생성하는 합성 얼굴이미지 생성 단계(S400)로 이루어진다.
- [0099] 여기서, 제1 영상/벡터 인코더(100-1) 및 제2 영상/벡터 인코더(100-2)는 도 2에서 설명한 본 발명의 딥러닝 기반 영상 생성모델을 학습시키는 방법을 통해 학습된 영상/벡터 인코더 이고, 제1 잡음생성 인코더(200-1) 및 제2 잡음생성 인코더(200-2)는 도 2에서 설명한 본 발명의 딥러닝 기반 영상 생성모델을 학습시키는 방법을 통해 학습된 잡음생성 인코더 이다. 또한, 디코더(300-1)는 도 2에서 설명한 본 발명의 딥러닝 기반 영상 생성모델을 학습시키는 방법을 통해 학습된 디코더 이다.
- [0100] 이때, 제1 얼굴이미지 및 제2 얼굴이미지에 대해 부여되는 가중치(α , $1-\alpha$)는 0과 1 사이의 값을 부여할 수 있으며, 이를 통해 사용자가 원하는 얼굴 특징을 갖는 2개의 얼굴이미지를 본 발명의 입력으로 하는 한편, 각 얼굴이미지에 대해 반영하고 싶은 비율을 상기 가중치를 통해 설정하여 해당 비율만큼 반영된 가상의 합성 얼굴이미지를 생성할 수 있게 된다.
- [0101] 구현예에 따라서, 생성된 가상의 합성 얼굴이미지와, 사용자가 원하는 또다른 특징을 갖는 제3 얼굴이미지를 재차 합성할 수도 있으며, 이를 통해 본 발명에서는 빠른 시간 안에 사용자가 원하는 특징을 갖는 가상의 합성 얼굴이미지를 최종적으로 생성할 수도 있다.
- [0103] 이상과 같은 본 발명에 따른 2개의 얼굴이미지를 이용한 딥러닝 기반 합성얼굴 생성 장치 및 방법은 딥러닝 기반 영상 생성모델을 이용하여 2개의 얼굴 이미지로부터 사용자가 원하는 얼굴 특징을 갖는 가상의 얼굴이미지를 보다 적은 시도로 신속하게 생성할 수 있다.
- [0105] 본 발명의 기술 분야에서 통상의 지식을 가진 자는 여기에 개시된 실시예들과 관련하여 설명된 다양한 예시적인 논리 블록들, 모듈들, 프로세서들, 수단들, 회로들 및 알고리즘 단계들이 전자 하드웨어, (편의를 위해, 여기에서 소프트웨어로 지칭되는) 다양한 형태들의 프로그램 또는 설계 코드 또는 이들 모두의 결합에 의해 구현될 수 있다는 것을 이해할 것이다. 하드웨어 및 소프트웨어의 이러한 상호 호환성을 명확하게 설명하기 위해, 다양한 예시적인 컴포넌트들, 블록들, 모듈들, 회로들 및 단계들이 이들의 기능과 관련하여 위에서 일반적으로 설명되었다. 이러한 기능이 하드웨어 또는 소프트웨어로서 구현되는지 여부는 특정한 애플리케이션 및 전체 시스템에 대하여 부과되는 설계 제약들에 따라 좌우된다. 본 발명의 기술 분야에서 통상의 지식을 가진 자는 각각의 특정한 애플리케이션에 대하여 다양한 방식으로 설명된 기능을 구현할 수 있으나, 이러한 구현 결정들은 본 발명의 범위를 벗어나는 것으로 해석되어서는 안 될 것이다.
- [0106] 여기서 제시된 다양한 실시예들은 방법, 장치, 또는 표준 프로그래밍 및/또는 엔지니어링 기술을 사용한 제조물품(article)으로 구현될 수 있다. 용어 제조물품은 임의의 컴퓨터-판독가능 저장장치로부터 액세스 가능한 컴퓨터 프로그램, 캐리어, 또는 매체(media)를 포함한다. 예를 들어, 컴퓨터-판독가능 저장매체는 자기 저장 장치(예를 들면, 하드 디스크, 플로피 디스크, 자기 스트립, 등), 광학 디스크(예를 들면, CD, DVD, 등), 스마트 카드, 및 플래쉬 메모리 장치(예를 들면, EEPROM, 카드, 스틱, 키 드라이브, 등)를 포함하지만, 이들로 제한되

는 것은 아니다. 또한, 여기서 제시되는 다양한 저장 매체는 정보를 저장하기 위한 하나 이상의 장치 및/또는 다른 기계-판독가능한 매체를 포함한다.

[0107] 제시된 프로세스들에 있는 단계들의 특정한 순서 또는 계층 구조는 예시적인 접근들의 일례임을 이해하도록 한다. 설계 우선순위들에 기반하여, 본 발명의 범위 내에서 프로세스들에 있는 단계들의 특정한 순서 또는 계층 구조가 재배열될 수 있다는 것을 이해하도록 한다. 첨부된 방법 청구항들은 샘플 순서로 다양한 단계들의 엘리먼트들을 제공하지만 제시된 특정한 순서 또는 계층 구조에 한정되는 것을 의미하지는 않는다.

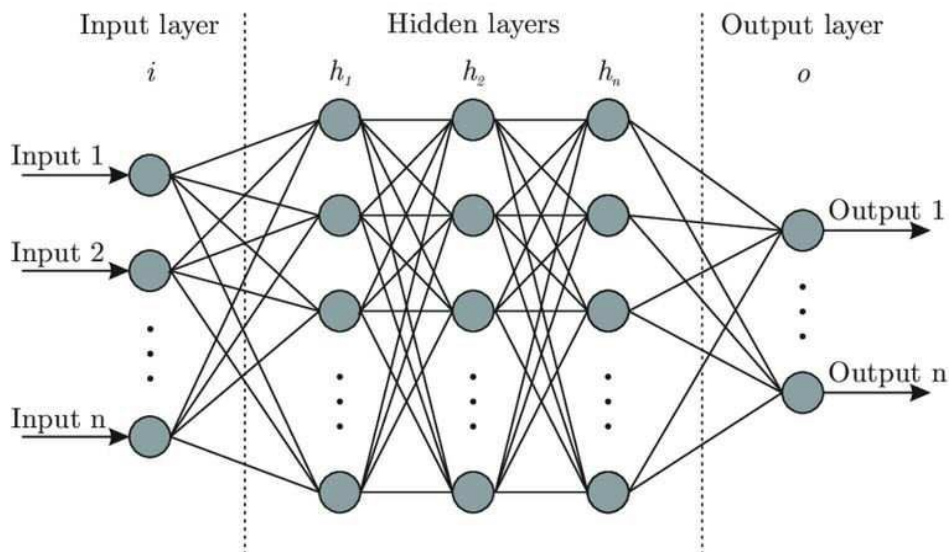
[0108] 제시된 실시예들에 대한 설명은 임의의 본 발명의 기술 분야에서 통상의 지식을 가진 자가 본 발명을 이용하거나 또는 실시할 수 있도록 제공된다. 이러한 실시예들에 대한 다양한 변형들은 본 발명의 기술 분야에서 통상의 지식을 가진 자에게 명백할 것이며, 여기에 정의된 일반적인 원리들은 본 발명의 범위를 벗어남이 없이 다른 실시예들에 적용될 수 있다. 그리하여, 본 발명은 여기에 제시된 실시예들로 한정되는 것이 아니라, 여기에 제시된 원리들 및 신규한 특징들과 일관되는 최광의의 범위에서 해석되어야 할 것이다.

부호의 설명

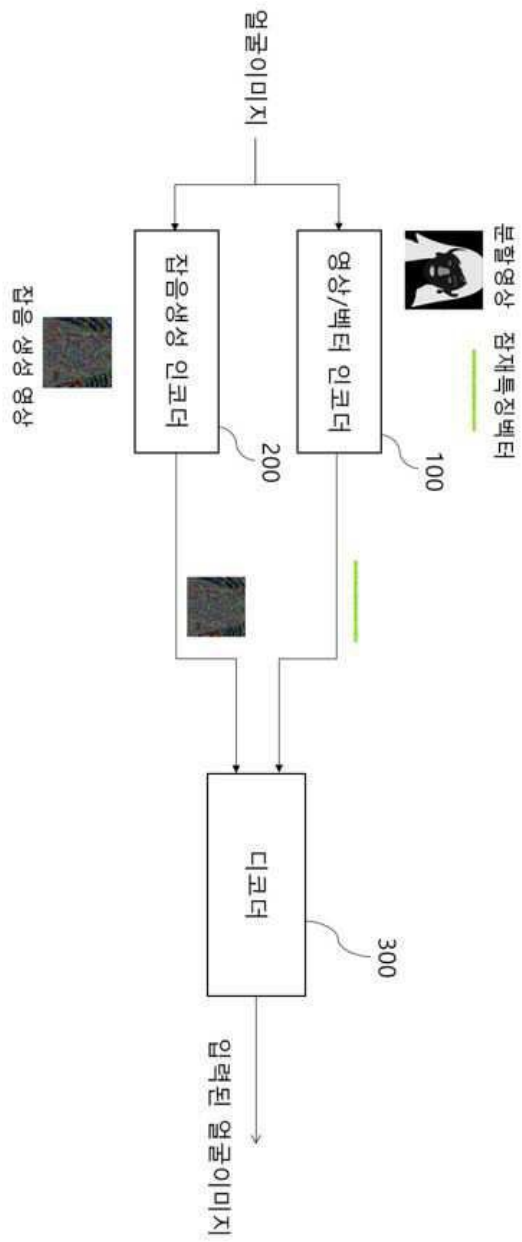
[0110] 100: 영상/벡터 인코더
100-1: 제1 영상/벡터 인코더
100-2: 제2 영상/벡터 인코더
200: 잡음생성 인코더
200-1: 제1 잡음생성 인코더
200-2: 제2 잡음생성 인코더
300,300-1: 디코더
400: 가중합 추출부

도면

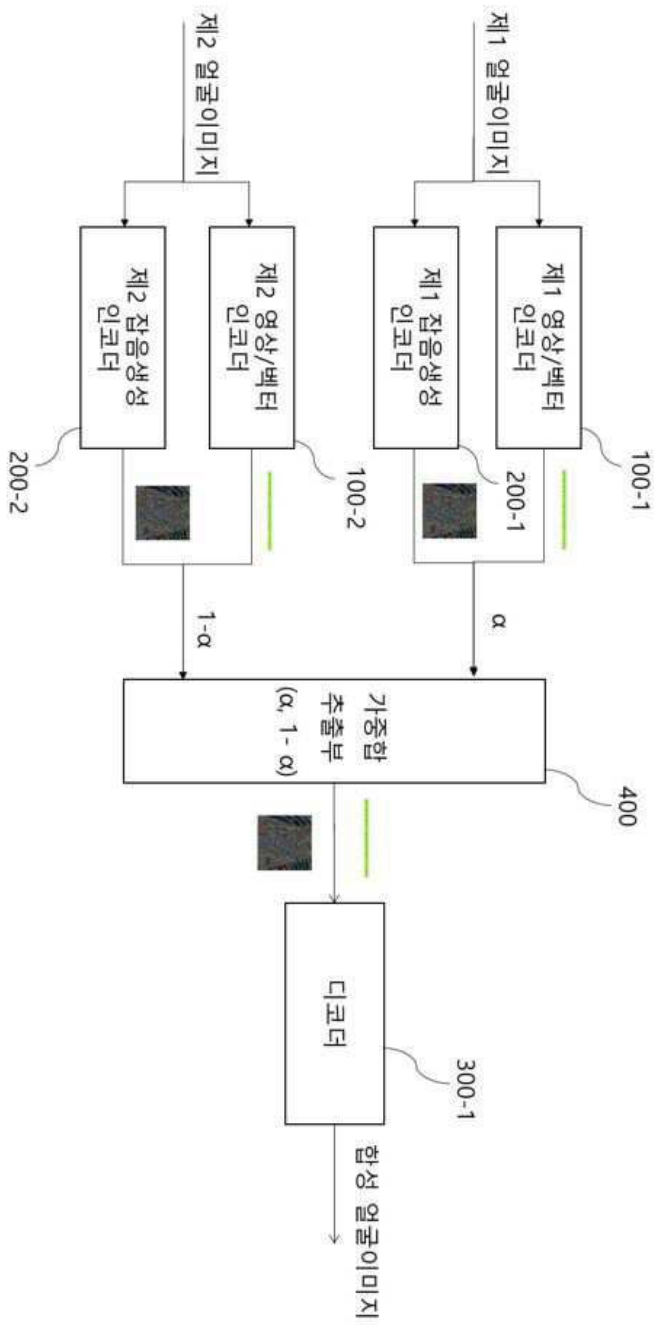
도면1



도면2



도면3



도면4

