



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0130498
(43) 공개일자 2022년09월27일

(51) 국제특허분류(Int. Cl.)
G06T 5/00 (2019.01) G06N 3/08 (2006.01)
G06T 3/40 (2006.01) G06T 7/13 (2017.01)
(52) CPC특허분류
G06T 5/005 (2013.01)
G06N 3/08 (2013.01)
(21) 출원번호 10-2021-0035476
(22) 출원일자 2021년03월18일
심사청구일자 없음

(71) 출원인
서강대학교산학협력단
서울특별시 마포구 백범로 35 (신수동, 서강대학교)
(72) 발명자
강석주
서울특별시 마포구 백범로 35(신수동)
김경훈
서울특별시 동작구 성대로39길 26, 301호(상도동)
(74) 대리인
유미특허법인

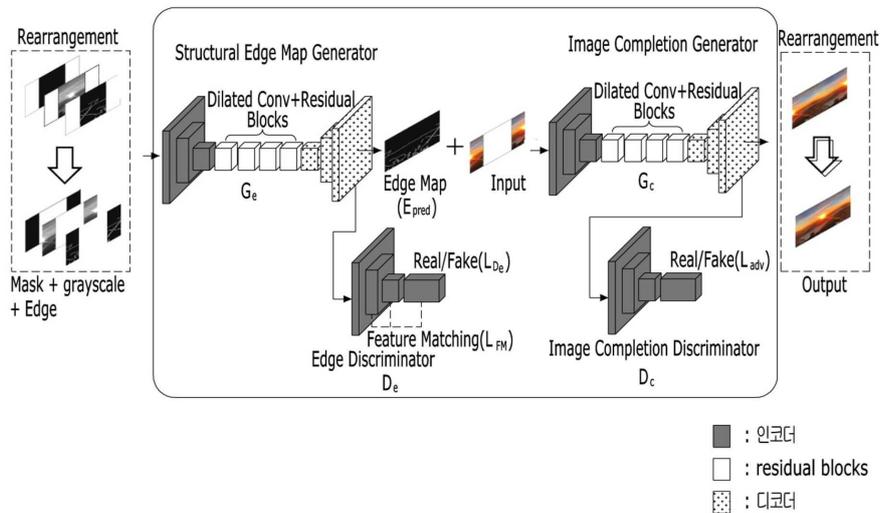
전체 청구항 수 : 총 8 항

(54) 발명의 명칭 딥뉴얼 네트워크 기반 이미지 아웃페인팅 방법 및 그 장치

(57) 요약

딥뉴얼 네트워크 기반 이미지 아웃페인팅 방법 및 그 장치가 개시된다. 이 방법은 적어도 하나의 프로세서에 의해 동작하는 컴퓨팅 장치의 이미지 아웃페인팅 방법으로서, 대상 이미지의 가장자리의 양방향 정보를 이용하여 생성한 마스크 영역을 포함하는 마스크 이미지를 입력받는 단계, GAN(Generative Adversarial Network)을 이용하여 상기 마스크 영역에 대한 엣지 맵을 예측하는 단계, 상기 엣지 맵 및 상기 대상 이미지를 이용하여, 상기 마스크 영역이 좌우에 배치된 컬러 이미지를 생성하는 단계, 그리고 상기 좌우를 교차시켜 완성된 컬러 이미지를 생성하는 단계를 포함한다.

대표도 - 도2



(52) CPC특허분류

G06T 3/40 (2013.01)

G06T 7/13 (2017.01)

G06T 2207/20081 (2013.01)

G06T 2207/20084 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711119713
과제번호	2020M3H4A1A02084899
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	나노미래소재원천기술개발사업
연구과제명	2축 신축성 AMLED용 TFT 및 LED 집적 기술
기여율	1/2
과제수행기관명	경희대학교 산학협력단
연구기간	2020.07.01~2024.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711116161
과제번호	2018-0-01421-003
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	대학ICT연구센터지원사업
연구과제명	인공지능 서비스 실현을 위한 지능형 반도체 설계 핵심기술 개발
기여율	1/2
과제수행기관명	서강대학교 산학협력단
연구기간	2020.01.01~2021.12.31

명세서

청구범위

청구항 1

적어도 하나의 프로세서에 의해 동작하는 컴퓨팅 장치의 이미지 아웃페인팅 방법으로서,
대상 이미지의 가장자리의 양방향 정보를 이용하여 생성한 마스크 영역을 포함하는 마스크 이미지를 입력받는 단계,
GAN(Generative Adversarial Network)을 이용하여 상기 마스크 영역에 대한 엣지 맵을 예측하는 단계,
상기 엣지 맵 및 상기 대상 이미지를 이용하여, 상기 마스크 영역이 좌우에 배치된 컬러 이미지를 생성하는 단계, 그리고
상기 좌우를 교차시켜 완성된 컬러 이미지를 생성하는 단계를 포함하는, 방법.

청구항 2

제1항에서,
원본 이미지를 엣지 처리한 엣지 이미지 및 상기 원본 이미지의 마스크 이미지를 토대로, 상기 GAN을 학습시키는 단계를 더 포함하고,
상기 마스크 이미지는,
상기 원본 이미지에서 정해진 비율이 마스크 영역으로 선정되는, 방법.

청구항 3

제2항에서,
상기 GAN을 학습시키는 단계는,
상기 정해진 비율에 따라 점진적으로 상기 마스크 영역을 증가시킨 복수의 마스크 이미지를 토대로, 상기 학습을 반복하는, 방법.

청구항 4

제1항에서,
상기 입력받는 단계 이전에,
구조적 엣지 맵 생성자 및 엣지 판별자를 포함하는 상기 GAN을 학습시키는 단계를 더 포함하고,
상기 구조적 엣지 맵 생성자는,
상기 마스크 영역에 대한 엣지 맵을 예측하도록 손실을 학습하며,
상기 손실은,
상기 구조적 엣지 맵 생성자의 힌지 손실과, 상기 엣지 판별자의 피쳐-매칭 손실을 포함하는, 방법.

청구항 5

제4항에서,
상기 GAN은,
이미지 완료 생성자 및 이미지 완료 판별자를 더 포함하고,

상기 이미지 완료 생성자는,
 상기 구조적 엷지 맵 생성자에 의해 예측된 엷지 맵으로부터 컬러 이미지를 생성하도록 손실을 학습하고,
 상기 이미지 완료 생성자의 손실은,
 상기 이미지 완료 판별자의 적대적 손실을 포함하는, 방법.

청구항 6

이미지 아웃페인팅을 수행하는 프로그램이 저장된 메모리, 그리고
 상기 프로그램을 실행하는 프로세서를 포함하고,
 상기 프로그램은,
 대상 이미지의 가장자리의 양방향 정보를 이용하여 누락된 영역을 생성하고, 상기 대상 이미지를 상기 누락된
 영역만큼 확장하는 이미지 아웃페인팅을 실행하는 명령어들(Instructions)을 포함하는, 컴퓨팅 장치.

청구항 7

제6항에서,
 상기 프로그램은,
 GAN(Generative Adversarial Network)을 이용하여 상기 누락된 영역에 대한 엷지 맵 영상을 예측하고, 상기 엷
 지 맵 영상을 통해 상기 확장을 수행하는 명령어들을 추가로 포함하는, 컴퓨팅 장치.

청구항 8

제7항에서,
 상기 프로그램은,
 상기 누락된 영역을 정해진 비율만큼 점진적으로 증가시켜 상기 엷지 맵 영상의 예측을 반복하는 명령어들을 추
 가로 포함하는, 컴퓨팅 장치.

발명의 설명

기술 분야

[0001] 본 발명은 딥뉴럴 네트워크 기반 이미지 아웃페인팅(Outpainting) 방법 및 그 장치에 관한 것이다.

배경 기술

[0002] 최근 딥러닝의 발전과 함께 영상 완성에 대한 연구가 활발히 진행되고 있다. 영상 완성 분야의 대표적인 예로는
 저해상도에서 고해상도 영상으로 복원하는 슈퍼 레졸루션(super-resolution)과 영상 내부의 훼손된 영역을 의미
 적, 인지적으로 자연스럽게 생성하는 이미지 인페인팅(Inpainting), 그리고 영상 외부를 확장하여 생성하는 이
 미지 아웃페인팅이 있다.

[0003] 이미지 아웃페인팅은 주어진 이미지를 훼손시키지 않고 이미지의 맥락을 고려하여 이미지의 외부를 지속적으로
 채우고 확장할 수 있다는 점에서 매우 흥미로운 기술이다. 이러한 이미지 아웃페인팅 기술은 이미지 리타겟팅
 (Re-targeting) 기술에 활용될 수 있다.

[0004] 이미지 리타겟팅이란 각기 다른 디스플레이의 종횡비에 맞춰 이미지를 리사이징하는 기술을 말하는데 이미지 아
 웃페인팅 기술을 활용하면 원본 영상을 왜곡시키지 않고 부족한 영역만큼만 이미지를 생성하여 채울 수 있으며
 리사이징으로 인한 영상의 시각적 품질 저하를 경감시킬 수 있다.

[0005] 고전적인 이미지 아웃페인팅 방법은 딥러닝을 사용하지 않고 patch의 similarity를 보고 후보들을 고르고 가장
 유사한 걸로 채워 넣는 방법이다. 이러한 방법은 생성해야 하는 영역이 커지면 좋은 성능을 내지 못하는 한계가
 있다. 출력 화질을 향상시키기 위해 GAN(Generative Adversarial Network)을 사용하는 몇가지 방법이 이후 제
 안되었다. 논문(Mark Sabini and Gili Rusak. Painting outside the box: Im-age outpainting with gans.
 arXiv preprint arXiv:1808.08483, 2018)는 처음으로 GAN 기반의 방법을 제안했다. 이 방법은 간단한 인코더-

디코더 구조로 구성되며 평균 제곱 오차(MSE, Mean Squared Error)와 GAN 손실만 사용한다. 후처리를 사용하여 출력 이미지를 부드럽게 하지만 생성된 이미지에서 반복되는 픽셀을 표시하는 한계가 있다. 논문(Teterwak, Piotr and Sarna, Aaron and Krishnan, Dilip and Maschinot, Aaron and Belanger, David and Liu, Ce and Freeman, William T. Boundless: Generative adversarial networks for image extension. In Proceedings of the IEEE International Conference on Computer Vision, pages 10521-10530, 2019)은 사전 훈련된 딥뉴럴 네트워크의 의미 정보를 기반으로 판별자의 동작을 조절하는 훈련을 위한 안정화 기법인 판별기에서 의미적 조건화를 사용했다. 논문(Yang, Zongxin and Dong, Jian and Liu, Ping and Yang, Yi and Yan, Shuicheng. Very long natural scenery image prediction by outpainting. In Proceedings of the IEEE International Conference on Computer Vision, pages 10561-10570, 2019)은 장단기 기억(Long Short-Term Memory models, LSTM)을 기반으로 한 RCT(Recurrent Content Transfer) 모델을 제안했다. RCT는 입력 영역 특징을 예측 영역으로 전송하여 입력 영역과 예측 영역 간의 자연스러운 연결을 개선한다. 그러나 생성된 영역은 인접한 입력 영역의 특징을 비교적 크게 반영한다. 따라서, 생성된 전체 영역이 입력 이미지의 중요한 특징인 색조와 일치하지 않는 경향이 있다.

[0006] 이처럼, 종래기술들은 넓은 영역을 생성하는 영상에서의 품질 성능이 좋지 못하다. 특히, 색감과 구조 등이 원본 영상과 일치하지 않으며, 단순히 픽셀을 반복해서 생성하는 경향이 있다. 또한, 아웃페인팅에서 사용하는 GANs 네트워크들은 미지 영역(unknown region)에 대한 예측이 어려워 모드 붕괴, 비-수렴, 기울기 소실 등 다양한 학습 문제가 발생하며 이로 인해 성능 개선에 어려움이 있다.

발명의 내용

해결하려는 과제

[0007] 해결하고자 하는 과제는 양방향 경계 영역 재배열을 사용한 이미지 아웃페인팅 방법 및 그 장치를 제공하는 것이다.

과제의 해결 수단

[0008] 한 특징에 따르면, 적어도 하나의 프로세서에 의해 동작하는 컴퓨팅 장치의 이미지 아웃페인팅 방법으로서, 대상 이미지의 가장자리의 양방향 정보를 이용하여 생성한 마스크 영역을 포함하는 마스크 이미지를 입력받는 단계, GAN(Generative Adversarial Network)을 이용하여 상기 마스크 영역에 대한 엣지 맵을 예측하는 단계, 상기 엣지 맵 및 상기 대상 이미지를 이용하여, 상기 마스크 영역이 좌우에 배치된 컬러 이미지를 생성하는 단계, 그리고 상기 좌우를 교차시켜 완성된 컬러 이미지를 생성하는 단계를 포함한다.

[0009] 원본 이미지를 엣지 처리한 엣지 이미지 및 상기 원본 이미지의 마스크 이미지를 토대로, 상기 GAN을 학습시키는 단계를 더 포함하고, 상기 마스크 이미지는, 상기 원본 이미지에서 정해진 비율이 마스크 영역으로 선정될 수 있다.

[0010] 상기 GAN을 학습시키는 단계는, 상기 정해진 비율에 따라 점진적으로 상기 마스크 영역을 증가시킨 복수의 마스크 이미지를 토대로, 상기 학습을 반복할 수 있다.

[0011] 상기 입력받는 단계 이전에, 구조적 엣지 맵 생성자 및 엣지 판별자를 포함하는 상기 GAN을 학습시키는 단계를 더 포함하고, 상기 구조적 엣지 맵 생성자는, 상기 마스크 영역에 대한 엣지 맵을 예측하도록 손실을 학습하며, 상기 손실은, 상기 구조적 엣지 맵 생성자의 힌지 손실과, 상기 엣지 판별자의 피쳐-매칭 손실을 포함할 수 있다.

[0012] 상기 GAN은, 이미지 완료 생성자 및 이미지 완료 판별자를 더 포함하고, 상기 이미지 완료 생성자는, 상기 구조적 엣지 맵 생성자에 의해 예측된 엣지 맵으로부터 컬러 이미지를 생성하도록 손실을 학습하고, 상기 이미지 완료 생성자의 손실은, 상기 이미지 완료 판별자의 적대적 손실을 포함할 수 있다.

[0013] 다른 특징에 따르면, 컴퓨팅 장치는 이미지 아웃페인팅을 수행하는 프로그램이 저장된 메모리, 그리고 상기 프로그램을 실행하는 프로세서를 포함하고, 상기 프로그램은, 대상 이미지의 가장자리의 양방향 정보를 이용하여 누락된 영역을 생성하고, 상기 대상 이미지를 상기 누락된 영역만큼 확장하는 이미지 아웃페인팅을 실행하는 명령어들(Instructions)을 포함한다.

[0014] 상기 프로그램은, GAN(Generative Adversarial Network)을 이용하여 상기 누락된 영역에 대한 엣지 맵 영상을

예측하고, 상기 옛지 맵 영상을 통해 상기 확장을 수행하는 명령어들을 추가로 포함할 수 있다.

[0015] 상기 프로그램은, 상기 누락된 영역을 정해진 비율만큼 점진적으로 증가시켜 상기 옛지 맵 영상의 예측을 반복하는 명령어들을 추가로 포함할 수 있다.

발명의 효과

[0016] 실시예에 따르면, 양방향 경계 영역 재배열을 통하여 이미지의 가장자리의 양방향 정보를 이용하여 누락된 영역을 생성할 수 있도록 함으로써, 단방향 정보를 사용하는 기존 이미지 아웃페인팅 방식보다 높은 품질의 이미지를 생성할 수 있다.

[0017] 또한, 마스크의 크기를 점진적으로 증가시켜 이미지 아웃페인팅 절차를 반복함으로써, 구조적 옛지 맵 생성자와 이미지 완료 생성자를 안정적으로 학습시킬 수 있다.

도면의 간단한 설명

[0018] 도 1은 본 발명의 실시예에 따른 주요 개념을 나타내는 예시도이다.

도 2는 실시예에 따른 이미지 아웃페인팅 구조를 도시한다.

도 3은 실시예에 따른 예지 맵 생성기에서 점진적으로 생성된 예지 맵을 기반으로 하는 이미지 완료 프로세스의 예시도이다.

도 4는 실시예에 따른 학습 절차를 나타낸 순서도이다.

도 5는 실시예에 따른 이미지 아웃페인팅 절차를 나타낸 순서도이다.

도 6은 본 발명의 다른 실시예에 따른 컴퓨팅 장치의 구조도이다.

발명을 실시하기 위한 구체적인 내용

[0019] 아래에서는 첨부한 도면을 참고로 하여 본 발명의 실시예에 대하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

[0020] 명세서 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 있는 것을 의미한다.

[0021] 또한, 명세서에 기재된 "...부", "...기", "...모듈" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.

[0022] 본 발명에서 설명하는 장치들은 적어도 하나의 프로세서, 메모리 장치, 통신 장치 등을 포함하는 하드웨어로 구성되고, 지정된 장소에 하드웨어와 결합되어 실행되는 프로그램이 저장된다. 하드웨어는 본 발명의 방법을 실행할 수 있는 구성과 성능을 가진다. 프로그램은 도면들을 참고로 설명한 본 발명의 동작 방법을 구현한 명령어(instructions)를 포함하고, 프로세서와 메모리 장치 등의 하드웨어와 결합하여 본 발명을 실행한다.

[0023] 본 명세서에서 "전송 또는 제공"은 직접적인 전송 또는 제공하는 것 뿐만 아니라 다른 장치를 통해 또는 우회 경로를 이용하여 간접적으로 전송 또는 제공도 포함할 수 있다.

[0024] 본 명세서에서 단수로 기재된 표현은 "하나" 또는 "단일" 등의 명시적인 표현을 사용하지 않은 이상, 단수 또는 복수로 해석될 수 있다.

[0025] 본 명세서에서 도면에 관계없이 동일한 도면번호는 동일한 구성요소를 지칭하며, "및/또는"은 언급된 구성요소들의 각각 및 하나 이상의 모든 조합을 포함한다.

[0026] 본 명세서에서, 제1, 제2 등과 같이 서수를 포함하는 용어들은 다양한 구성요소들을 설명하는데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되지는 않는다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를들어, 본 개시의 권리 범위를 벗어나지 않으면서 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다.

- [0027] 본 명세서에서 도면을 참고하여 설명한 실시예들에서, 임의의 실시예로 단독 구현될 수도 있고, 여러 실시예가 병합되거나 분할될 수도 있고, 각 실시예에서 특정 동작은 수행되지 않을 수 있다.
- [0028] 도 1은 본 발명의 실시예에 따른 이미지 아웃페인팅의 개념을 나타내는 예시도이다.
- [0029] 도 1의 (a)를 참조하면, 양방향 또는 좌우 화살표 방향으로 이미지가 생성되며 이미지의 가운데 절취선을 중심으로 이미지가 분할된다. 도 1의 (b)를 참조하면, 이미지의 가운데 절취선을 중심으로 왼쪽 이미지와 오른쪽 이미지가 좌우로 교차된다. 그러면, 이미지 생성 방향은 절취선, 즉, 안쪽을 향한다. 도 1의 (c)를 참조하면, 양방향 화살표 방향으로 이미지가 생성된 후, 이미지의 가운데 절취선을 중심으로 왼쪽과 오른쪽 이미지가 교차되면, 도 1의 (d)와 같이 확장 이미지가 생성된다.
- [0030] 이처럼, 실시예에 따르면, 양방향 엣지 영역 재배열을 사용하여 이미지 아웃페인팅을 구현한다. 양방향 엣지 영역 재배열은 이미지의 가장자리의 양방향 정보를 이용하여 누락된 영역을 생성할 수 있도록 함으로써, 단방향 정보를 사용하는 기존 이미지 아웃페인팅 방식보다 높은 품질의 이미지를 생성할 수 있다.
- [0031] 본 발명의 실시예는, 양방향 엣지 영역 재배열을 사용하는 이미지 아웃페인팅을 수행하는 프로그램이 저장된 메모리 및 상기 프로그램을 실행하는 적어도 하나의 프로세서에 의해 동작하는 컴퓨팅 장치에 의해 구현된다. 이미지 아웃페인팅에 대해 설명하면, 다음과 같다.
- [0032] 도 2는 실시예에 따른 이미지 아웃페인팅 구조를 도시한다.
- [0033] 도 2를 참조하면, 이미지 아웃페인팅 구조는 엣지 맵 생성 네트워크(edge map generation network)와 이미지 완성 네트워크(image completion network)로 구성된다.
- [0034] 각 네트워크는 GAN(Generative Adversarial Network) 기반 인공지능망으로서, 생성자(Generator)와 판별자(Discriminator) 쌍(Pair)으로 구성된다. 즉, 엣지 맵 생성 네트워크는 구조적 엣지 맵 생성자(structural edge map generator)와 엣지 판별자(Edge Discriminator)로 구성되고, 이들을 각각 G_e 와 D_e 로 정의한다. 이미지 완성 네트워크는 이미지 완료 생성자(Image Completion Generator)와 이미지 완료 판별자(Image Completion Discriminator)로 구성되고, 이들을 각각 G_c 와 D_c 로 정의한다.
- [0035] 구조적 엣지 맵 생성자(G_e)는 재배열(Rearrangement)된 각각의 마스크 이미지, 마스크된 그레이스케일(Grayscale) 이미지, 마스크된 엣지(Edge) 이미지를 입력받는다.
- [0036] 구조적 엣지 맵 생성자(G_e)의 인코더는 입력 이미지를 압축한다. 구조적 엣지 맵 생성자(G_e)의 residual block은 블록간의 정보 잔차를 계산한다. 구조적 엣지 맵 생성자(G_e)의 디코더는 압축된 정보를 다시 이미지로 압축 해제한다.
- [0037] 이미지 완료 생성자(G_c)는 구조적 엣지 맵 생성자(G_e)를 통해 생성된 글로벌엣지 맵(Edge Map)(E_{pred})와 재배열된 마스크(Mask) 이미지를 입력받는다.
- [0038] 이미지 완료 생성자(G_c)의 인코더는 입력 이미지를 압축한다. 이미지 완료 생성자(G_c)의 residual block은 블록간의 정보 잔차를 계산한다. 이미지 완료 생성자(G_c)의 디코더는 압축된 정보를 다시 이미지로 압축 해제한다.
- [0039] 이때, 마스크 이미지는 원본 이미지를 마스킹 처리하는데 사용되는 2진 이미지(binary image)이다. 마스크된 이미지는 도 1(a)의 원본 이미지를 마스킹 처리한 이미지이다. 여기서, 마스크된 이미지는 마스크 이미지를 통해 마스킹한 결과 이미지에 해당한다. 마스크된 그레이스케일 이미지는 도 1(a)의 원본 이미지를 그레이스케일로 처리하고 마스크 처리를 한 이미지이다. 마스크된 엣지 이미지는 도 1(a)의 원본 이미지를 엣지 처리하고 마스크 처리를 한 이미지이다.
- [0040] 컴퓨팅 장치는 도 1에서 설명한 바와 같이, 마스크 이미지, 그레이스케일 이미지, 엣지 이미지 각각에 대하여 도 1(b)와 같이 이미지의 가운데 절취선을 중심으로 왼쪽과 오른쪽 이미지를 교차시키는 재배열을 한다.
- [0041] 이처럼, 양방향 엣지 영역 재배열로 전처리된 마스크 이미지, 그레이스케일 이미지, 엣지 이미지를 입력받은 구조적 엣지 맵 생성자(G_e)는 누락된 영역(missing region)의 엣지들을 환각(hallucinate)한다.
- [0042] I_{gt} 는 원본 이미지(ground-truth image)이고, E_{gt} 는 엣지 이미지이며, I_{gray} 는 그레이스케일 이미지이다. 구조적

엣지 맵 생성자(G_e)는 마스크된 그레이스케일 이미지(\tilde{I}_{gray})와 마스크된 엣지 이미지(\tilde{E}_{gt})를 입력으로 사용하며, 이들을 수학적식으로 나타내면 다음과 같다.

[0043] [수학식 1]

$$[0044] \quad \tilde{I}_{gray} = I_{gray} \odot (1 - M)$$

[0045] [수학식 2]

$$[0046] \quad \tilde{E}_{gt} = E_{gt} \odot (1 - M)$$

[0047] 수학식 1과 2에서, \odot 은 아다마르 곱(Hadamard product)을 나타내는 연산 기호이다. M 은 마스크를 나타내며, 누락 영역은 1로, 그리고 배경은 0으로 표시하는 이진 이미지이다.

[0048] 구조적 엣지 맵 생성자(G_e)는 확장된 컨볼루션(dilated convolution)과 잔차 블록(residual blocks)으로 구성된다. 구조적 엣지 맵 생성자(G_e)는 마스크된 영역의 엣지 맵(E_{pred})을 예측하며, 이를 수학적식으로 나타내면 수학식 3과 같다.

[0049] [수학식 3]

$$[0050] \quad E_{pred} = G_e(\tilde{I}_{gray}, \tilde{E}_{gt}, M).$$

[0051] 즉, 구조적 엣지 맵 생성자(G_e)는 마스크된 영역에 대하여 예측한 엣지 맵(E_{pred})을 출력한다.

[0052] 엣지 판별자(D_e)는 I_{gray} 에 조건화된 E_{gt} 와 E_{pred} 를 입력받아 엣지 맵(E_{pred})이 실제(real)인지 가짜(fake)인지 예측한다.

[0053] 구조적 엣지 맵 생성자(G_e)는 GAN 손실 및 피쳐(Feature)-매칭(Matching) 손실의 힙지 변형(hinge variant)으로 구성된 목표로 훈련된다. 힙지 손실(hinge loss)은 이진 분류에서 효과적으로 사용되는데, 이는 구조적 엣지 맵 생성자(G_e)를 바이너리 엣지 맵을 사용하여 훈련하기 때문이다. 이를 수학적식으로 나타내면, 수학식 4와 같다.

[0054] [수학식 4]

$$[0055] \quad \mathcal{L}_{G_e} = \lambda_{hinge} \mathcal{L}_{hinge} + \lambda_{FM} \mathcal{L}_{FM},$$

[0056] 여기서, λ_{hinge} 와 λ_{FM} 는 정규화 매개 변수(regularization parameters)이다. \mathcal{L}_{G_e} 는 엣지 맵 생성자(G_e)의 최종 손실이다. \mathcal{L}_{G_e} 는 엣지 맵 생성자(G_e)의 손실과 피쳐 매칭 손실의 합으로 계산될 수 있다. \mathcal{L}_{hinge} 는 힙지 손실이 적용된 엣지 맵 생성자(G_e)의 손실이다. \mathcal{L}_{hinge} 는 그레이 스케일 원본 이미지와 엣지 맵 생성자(G_e)에 의해 예측한 이미지를 토대로, 엣지 판별자(D_e)에 의해 크로스 엔트로피(cross entropy) 손실로 계산될 수 있다. \mathcal{L}_{FM} 은 피쳐 매칭 손실이다. \mathcal{L}_{FM} 은 엣지 맵 생성자(G_e)의 예측값과 엣지 맵 생성자(G_e)에 입력된 실제값을 엣지 판별자(D_e)에 입력하고, 엣지 판별자(D_e)의 중간층 결과물 피쳐맵 사이의 차이값에 절대값을 취해 오차 합을 통해 계산될 수 있다.

[0057] 힙지 손실이 있는 구조적 엣지 맵 생성자(G_e)와 엣지 판별자(D_e)는 다음 수학식 5를 통해 업데이트된다.

[0058] [수학식 5]

$$\mathcal{L}_{hinge} = -\mathbb{E}_{I_{gray}} [D_e(E_{pred}, I_{gray})],$$

$$\begin{aligned} \mathcal{L}_{D_e} = & \mathbb{E}_{(E_{gt}, I_{gray})} [max(0, 1 - D_e(E_{gt}, I_{gray}))] \\ & + \mathbb{E}_{I_{gray}} [max(0, 1 + D_e(E_{pred}, I_{gray}))]. \end{aligned}$$

[0059]

[0060] 여기서, $-\mathbb{E}_{I_{gray}}$ 는 입력 데이터 즉, 그레이스케일 이미지에 대한 확률 분포 기대값을 의미한다.

[0061] 구조적 엣지 맵 생성자(G_e)는 구조적 엣지 맵 생성자(G_e)를 통해 생성한 E_{pred} 와 I_{gray} 를 구조적 엣지 맵 판별자인 D_e 를 통해 손실(차이)을 구하고 그 손실만큼 구조적 엣지 맵 생성자(G_e)를 업데이트한다.

[0062] 엣지 판별자(D_e)는 힌지(hinge)구조로 이루어진 손실 형태로, 원본 이미지(E_{gt})를 판별할 때는 1보다 작은 손실과 생성 이미지(E_{pred})를 판별할 때는 1보다 큰 손실을 얻도록 하여, 결과적으로 더해진 형태로 판별자를 업데이트한다.

[0063] 피쳐-매칭 손실인 \mathcal{L}_{FM} 은 엣지 판별자(D_e)의 중간 계층에서 활성화 맵들을 비교한다. 이는 엣지 판별자(D_e)가 실제 이미지와 유사한 출력을 생성하도록 강제함으로써 훈련 프로세스를 안정화시킬 수 있다. 이는 활성화 맵이 사전 훈련된 VGG 네트워크의 피쳐 맵들과 비교되는 지각 손실(perceptual loss)과 유사하다.

[0064] 피쳐-매칭 손실인 \mathcal{L}_{FM} 은 다음 수학식 6과 같이 정의될 수 있다.

[0065] [수학식 6]

$$\mathcal{L}_{FM} = \mathbb{E} \left[\sum_i^L \frac{1}{N_i} \|D_e^{(i)}(E_{gt}) - D_e^{(i)}(E_{pred})\|_1 \right],$$

[0066]

[0067] 여기서, N_i 는 엘리먼트들의 개수이고, D_e 는 엣지 판별자(D_e)의 i 번째 계층에서 활성화이다.

[0068] 구조적 엣지 맵 생성자(G_e)로부터 엣지 맵(E_{pred})을 획득한 이미지 완료 생성자(G_c)는 완성된 컬러 이미지를 생성한다. 이때, 이미지 완료 생성자(G_c)는 dilated convolution과 residual blocks으로 구성된다.

[0069] 이미지 완료 생성자(G_c)는 마스크된 컬러 이미지(\tilde{I}_{gt})와 조건부 합성 엣지 맵(E_{comp})을 입력으로 사용하고, 완성된 컬러 이미지(I_{pred})를 최종 출력한다.

[0070] 이때, 마스크된 컬러 이미지(\tilde{I}_{gt})와 조건부 합성 엣지 맵(E_{comp})은 다음의 수학식 7 및 8과 같다.

[0071] [수학식 7]

$$\tilde{I}_{gt} = I_{gt} \odot (1 - M)$$

[0072]

[0073] [수학식 8]

$$E_{comp} = E_{gt} \odot (1 - M) + E_{pred} \odot M$$

[0074]

[0075] 완성된 컬러 이미지(I_{pred})는 수학식 9와 같다.

[0076] [수학식 9]

$$I_{pred} = G_c \left(\tilde{I}_{gt}, E_{comp} \right)$$

[0077]

[0078] 이미지 완료 네트워크는 ℓ_1 로스, 적대적(adversarial) 손실, 지각(perceptual) 손실, 스타일 손실로 구성된 공동 손실(joint loss)에 대해 훈련된다. 적절한 스케일링을 보장하기 위해 ℓ_1 로스는 마스크 사이즈로 정규화된다. 현실적인 결과를 생성하기 위하여 이미지 완료 생성자(G_c)에서 적대적 손실을 적용한다.

[0079] 이미지 완료 판별자(D_c)의 적대적 손실(\mathcal{L}_{adv})은 다음 수학식 10과 같이 정의할 수 있다.

[0080] [수학식 10]

$$\mathcal{L}_{adv} = \mathbb{E}_{(I_{gt}, E_{comp})} [\log D_c(I_{gt}, E_{comp})] + \mathbb{E}_{E_{comp}} [\log[1 - D_c(I_{pred}, E_{comp})]].$$

[0081]

[0082] 적대적 손실의 첫번째 항($\mathbb{E}_{(I_{gt}, E_{comp})} [\log D_c(I_{gt}, E_{comp})]$)은 진짜 이미지를 넣었을 때의 이미지 완료 판별자(D_c)의 값이며 두번째 항($\mathbb{E}_{E_{comp}} [\log[1 - D_c(I_{pred}, E_{comp})]$)은 생성자가 만든 가짜 이미지를 넣었을 때의 이미지 완료 판별자(D_c)의 값이다. 결과적으로, 이 두 항의 목적함수를 이용해 진짜 이미지일 때는 큰 값을, 가짜 이미지일 때는 작은 값을 출력하도록 이미지 완료 판별자(D_c)를 학습시킨다.

[0083] 지각 손실은 아래 수학식 11과 같이 정의할 수 있다.

[0084] [수학식 11]

$$\mathcal{L}_{perc} = \mathbb{E} \left[\sum_i \frac{1}{N_i} \|\varphi_i(I_{gt}) - \varphi_i(I_{pred})\|_1 \right].$$

[0085]

[0086] 여기서, φ_i 은 pre-train 된 VGG(Visual Geometry Group) network로 정의한다. 앞서 FM 손실과 마찬가지로 중간층의 피쳐맵간의 거리를 계산하는 것이다. 차이점은 판별자가 아닌 사전 학습된 VGG network를 이용한다는 것이다. \mathcal{L}_{perc} 는 사전 훈련된 네트워크의 활성화 맵 간의 거리 측정을 정의하여 기능과 지각적으로 유사하지 않은 결과에 페널티를 부여한다.

[0087] 스타일 손실(\mathcal{L}_{style})은 수학식 12와 같이 정의되며, 이는 공지되어 있다.

[0088] [수학식 12]

$$\mathcal{L}_{style} = \mathbb{E}_j \left[\|G_j^\varphi(\tilde{I}_{pred}) - G_j^\varphi(\tilde{I}_{gt})\|_1 \right].$$

[0089]

[0090] 최종적으로, 총 손실은 다음 수학식 13과 같이 정의할 수 있다.

[0091] [수학식 13]

$$\mathcal{L}_{G_c} = \lambda_{\ell_1} \mathcal{L}_{\ell_1} + \lambda_{adv} \mathcal{L}_{adv} + \lambda_p \mathcal{L}_{perc} + \lambda_{style} \mathcal{L}_{style}.$$

[0092]

[0093] 이때, 스타일 손실(\mathcal{L}_{style})의 비율을 높임으로써, 색상 톤 불일치 문제를 해결할 수 있다. \mathcal{L}_{G_c} 은 이미지 완

료 생성자(G_c)의 최종 손실로서, 각각의 손실함수에 각각의 가중치를 곱한 값의 합으로 구성된다. \mathcal{L}_{l_1} 은 손실이다. λ_{l_1} 은 로스 가중치 변수로서, 생성 이미지와 실제 이미지와의 차이값에 절대값을 구하고 그 오차들의 합을 구하여 획득된다. 이는 가장 기본적인 손실함수이며 전체적으로 모델이 원본이미지와 비슷한 이미지를 생성할 수 있도록 학습하는데 사용된다. \mathcal{L}_{adv} 은 적대적 손실 가중치 변수로서, 모델이 좀더 선명한 이미지를 생성하도록 학습하는데 효과가 있다. \mathcal{L}_{perc} 은 지각(perceptual) 손실로서, 이미지의 구조적인 면, 즉 뼈대와 같은 정보가 잘 유지될 수 있도록 하는 효과가 있다. \mathcal{L}_{style} 은 style 손실의 가중치 변수로서, texture나 색상과 같은 표면적인 정보, 이미지의 style을 학습하는데 도움을 준다. λ_p 은 지각(perceptual) 손실 가중치 변수이다.

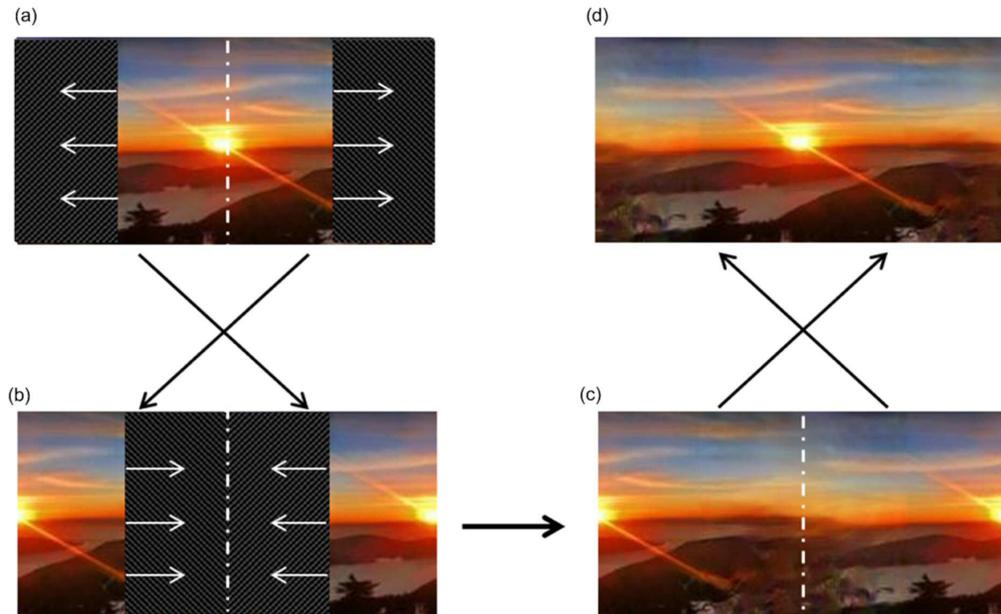
- [0094] 한 예시에 따르면, λ_{l_1} 은 1로, λ_{adv} 은 0.2로, λ_p 은 0.1로, λ_{style} 은 250으로 설정될 수 있다. 이 값들은 기존 연구들의 실험을 토대로 설정될 수 있다.
- [0095] λ_{l_1} , λ_{adv} , λ_p , λ_{style} 은 각각 손실 별로 가중치를 다르게 주기위해서 각 손실 별 가중치를 주기위한 가중치 변수에 해당한다.
- [0096] 그런데, GAN은 모드 붕괴(mode collapse), 비융합(non-convergence), 소멸 그레이디언트(vanishing gradient)와 같은 문제로 인해 훈련하기 어렵다. 이러한 문제는 컨텍스트 없이 예상보다 많은 영역이 생성되는 이미지 출력 작업에도 적용된다.
- [0097] 실시예에 따르면, 마스크 크기를 수평으로 증가시킴으로써 이미지 아웃페인팅의 관점에서 GAN 훈련을 안정시키기 위한 간단하지만 효과적인 훈련 방법을 제안하며 이는 도 3과 같다.
- [0098] 도 3은 실시예에 따른 예지 맵 생성기에서 점진적으로 생성된 예지 맵을 기반으로 하는 이미지 완료 프로세스의 예시도이다.
- [0099] 도 3을 참조하면, 입력 마스크 이미지(Input masked image)의 마스크 영역(공백)이 커질수록 생성된 엣지 맵(Generated edge map)은 GT(Ground Truth) 이미지와 더 차이가 나는 영상인 것을 알 수 있다. 따라서, 구조적 엣지 맵 생성자(G_e)에 입력되는 마스크 영역의 크기를 늘리는 동시에 보다 안정적으로 학습할 수 있도록 학습 단계를 마스크 영역의 크기별로 나눈다. 예를들어, 마스크는 32단계로 나뉘며 각 단계에서 점차 커진다. 초기 마스크 크기는 입력 이미지의 3.125%이며 선형적으로 증가하며 최종 마스크 크기는 입력 이미지의 50%로 설정될 수 있다.
- [0100] 훈련할 때 이미지 내부의 마스크를 설정하여 수평 진행 단계 학습 방식으로 진행한다. 테스트할 때 영상 외부에서 마스크된 영상이 입력으로 제공되며 엣지 영역 재배열 방법을 사용하여 아웃 페인팅 이미지를 생성한다.
- [0101] 도 4는 실시예에 따른 학습 절차를 나타낸 순서도로서, 각 단계에서 구조적 엣지 맵 생성자(G_e)에 의해 생성된 엣지 맵을 기반으로 한 이미지 완료 프로세스를 나타낸다. 이미지 완료 생성자(D_e)의 결과는 엣지 맵에 따라 달라진다.
- [0102] 도 4를 참조하면, 32 단계(Task)로 마스크의 크기를 구분하였다. 앞서 도 2 구조를 32단계로 반복적으로 학습한다. 입력 이미지들의 unknown region(mask) 영역을 점진적으로 늘려가면서 네트워크가 안정적으로 학습할 수 있다.
- [0103] 총 마스크의 크기는 32단계로 나뉘며 초기 마스크 크기는 입력 이미지의 3.125%이고 선형적으로 증가하여 최종 마스크 크기는 입력 이미지의 50%로 설정된다. 이러한 과정을 통하여, 구조적 엣지 맵 생성자(G_e) 및 이미지 완료 생성자(D_e)를 학습시킨다. 이때, 도 4의 입력 이미지들은 도 2의 구조적 엣지 맵 생성자(G_e)의 입력 이미지들과 동일하고, 다만, 마스크 이미지의 경우, 마스크 영역의 크기가 달라진다.
- [0105] 지금까지 설명한 내용을 토대로, 컴퓨팅 장치의 이미지 아웃페인팅 방법에 대해 설명하면 다음과 같다.

- [0106] 도 5는 실시예에 따른 이미지 아웃페인팅 방법 나타낸 순서도이다.
- [0107] 도 5를 참조하면, 컴퓨팅 장치는 대상 이미지의 unknown region(mask) 영역을 정해진 기준에 따른 크기로 설정하고 도 1(b)에서 설명한 바와 같이 좌우 교차시켜 설정한다(S101).
- [0108] 컴퓨팅 장치는 구조적 엷지 맵 생성자에서 S101에서 설정한 크기를 가진 마스크 이미지, 마스크된 그레이스케일 이미지, 마스크된 엷지 이미지를 입력받아 엷지 맵(E_{pred})을 출력한다(S103).
- [0109] 이미지 완료 생성자는 엷지 맵(E_{pred})과 S101에서 설정한 크기를 가진 컬러 마스크된 이미지를 입력받아 마스크 영역이 컬러로 채워진 이미지를 출력한다(S105). 이렇게 출력한 컬러로 채워진 이미지는 도 1(d)에서 설명한 바와 같이 좌우 교차시켜 완성된 컬러 이미지로 출력된다(S107).
- [0110] 다음, 컴퓨팅 장치는 정해진 단계를 충족하는지 판단(S109)하고, 정해진 단계를 충족하지 않으면, S101에서 설정한 크기를 정해진 비율만큼 증가시킨 마스크 이미지를 생성한다(S111).
- [0111] S111에서 생성한 마스크 이미지를 이용하여 S103 ~ S109를 반복한다. 이처럼, 마스크의 크기를 점진적으로 증가시켜 이미지 아웃페인팅 절차를 반복함으로써, 구조적 엷지 맵 생성자와 이미지 완료 생성자를 안정적으로 학습시킬 수 있다.
- [0113] 한편, 도 6은 본 발명의 다른 실시예에 따른 컴퓨팅 장치의 구조도이다.
- [0114] 도 1 내지 도 5에서 설명한 이미지 아웃페인팅은 컴퓨터 장치, 컴퓨터 판독 가능 매체로 구현될 수 있다.
- [0115] 도 6을 참조하면, 적어도 하나의 프로세서에 의해 동작하는 컴퓨팅 장치(100)에서, 도 1 내지 도 5에서 설명한 본 발명의 동작, 즉, 이미지 아웃페인팅을 실행하도록 기술된 명령들(instructions)이 포함된 프로그램을 실행할 수 있다.
- [0116] 컴퓨팅 장치(100)는 프로세서(110), 메모리(120), 입력 인터페이스 장치(130), 출력 인터페이스 장치(140), 네트워크 인터페이스(150) 및 저장 장치(160)를 포함하며, 이들은 버스(170)를 통해 통신할 수 있다.
- [0117] 프로세서(110)는 위의 도 1 내지 도 5를 토대로 설명한 방법들을 구현하도록 구성될 수 있다. 프로세서(110)는 중앙 처리 장치(Central Processing Unit, CPU)이거나, 또는 메모리(120) 또는 저장 장치(160)에 저장된 명령을 실행하는 반도체 장치일 수 있다.
- [0118] 메모리(120)는 프로세서(110)와 연결되고 프로세서(110)의 동작과 관련한 다양한 정보를 저장한다. 메모리(120)는 프로세서(110)에서 수행하기 위한 명령어를 저장하고 있거나 저장 장치(160)로부터 명령어를 로드하여 일시 저장할 수 있다. 프로세서(110)는 메모리(120)에 저장되어 있거나 로드된 명령어를 실행할 수 있다. 메모리는 ROM(Read Only Memory) 및 RAM(Random Access Memory)를 포함할 수 있다.
- [0119] 본 발명의 실시예에서 메모리(120)는 프로세서(110)의 내부 또는 외부에 위치할 수 있고, 메모리(120)는 이미 알려진 다양한 수단을 통해 프로세서(110)와 연결될 수 있다.
- [0120] 네트워크 인터페이스 장치(150)는 네트워크에 연결되어 신호를 송수신하도록 구성된다.
- [0121] 본 발명의 실시예들에서 개시된 시스템, 장치 및 방법은 다른 방식으로 구현될 수 있음을 이해해야 한다. 예를 들어, 개시된 장치 실시 예는 단지 예일 뿐이다. 예를 들어, 시스템의 처리 장치의 구분은 단지 논리적인 기능 구분이며, 실제로 구현하는 동안 다른 구분이 될 수 있다. 예를 들어, 다수의 장치 또는 구성 요소가 결합되거나 다른 시스템에 통합되거나 일부 기능이 무시되거나 수행되지 않을 수 있다.
- [0122] 또한, 본 발명의 실시예에 따른 방법을 수행하도록 구성된 장치들은 하나의 장치에 통합되거나, 장치들 각각은 물리적으로 단독으로 존재할 수도 있고, 또는 둘 이상의 장치가 하나의 장치로 통합될 수도 있다. 전술한 실시예들에서 기술된 방법을 수행하도록 구성된 장치는 하드웨어의 형태로 구현될 수 있거나 소프트웨어 기능 유닛의 형태로 구현될 수 있다.
- [0124] 이상에서 본 발명의 실시예에 대하여 상세하게 설명하였지만 본 발명의 권리범위는 이에 한정되는 것은 아니고 다음의 청구범위에서 정의하고 있는 본 발명의 기본 개념을 이용한 당업자의 여러 변형 및 개량 형태 또한 본

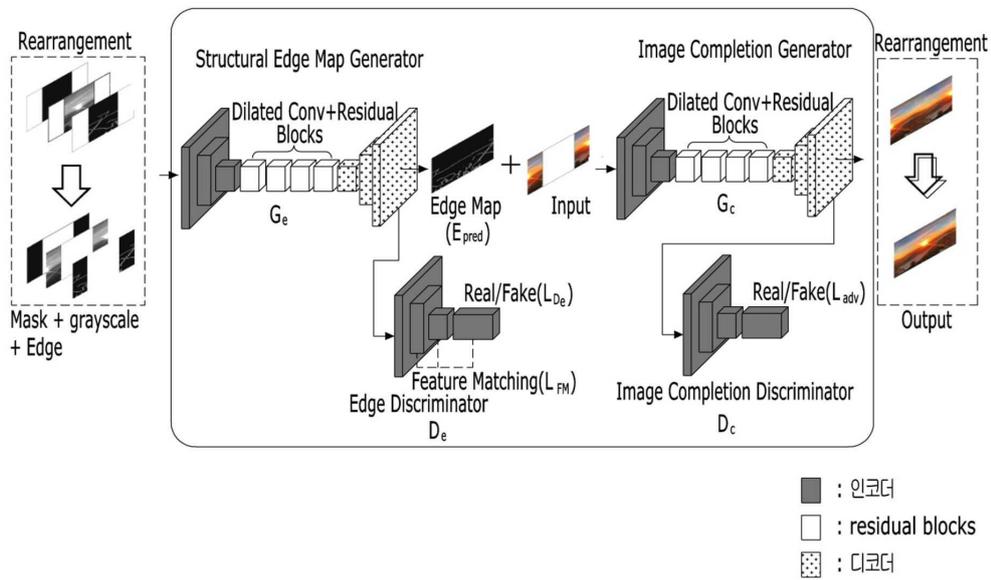
발명의 권리범위에 속하는 것이다.

도면

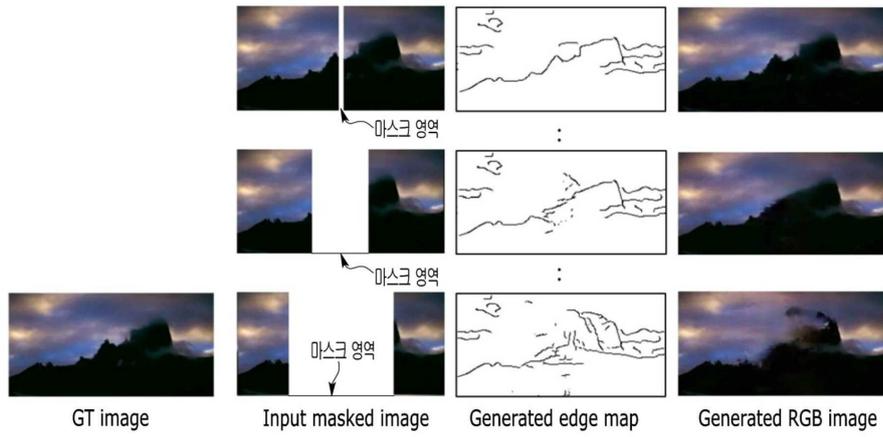
도면1



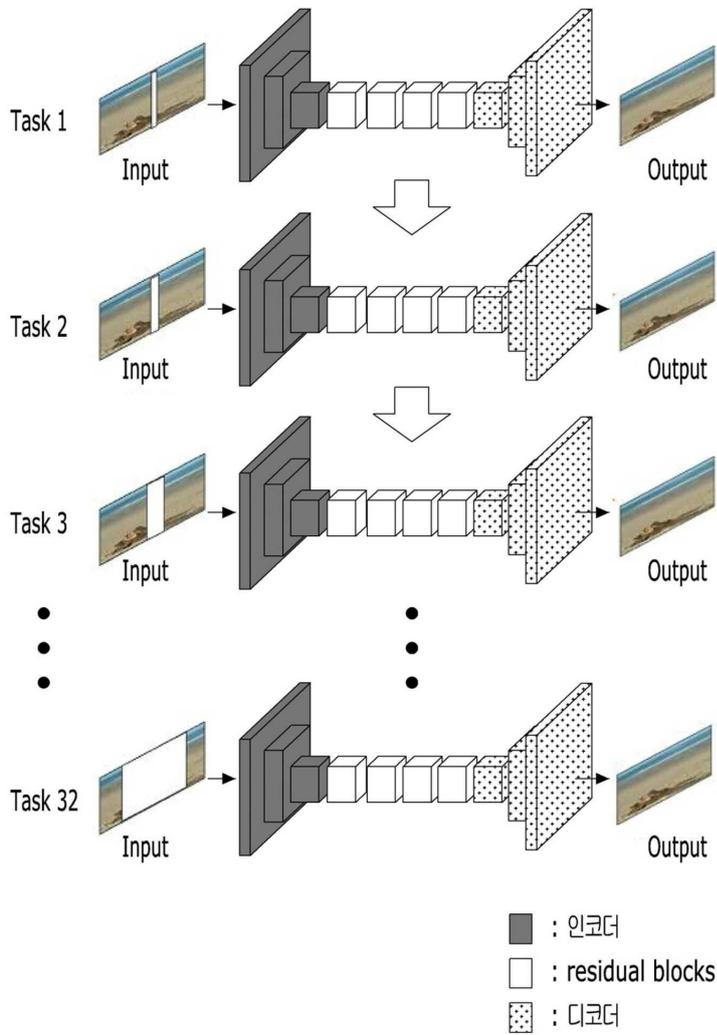
도면2



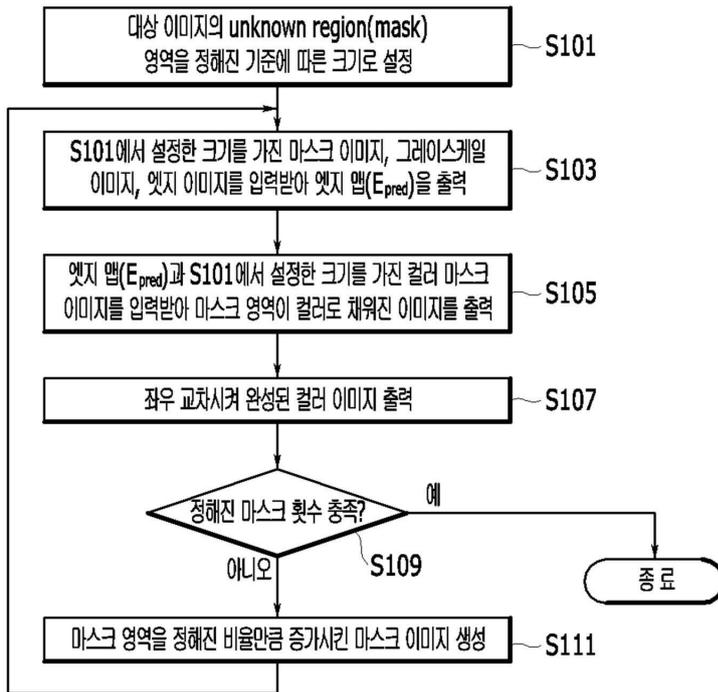
도면3



도면4



도면5



도면6

