(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0122628
(43) 공개일자 2024년08월13일

(51) 국제특허분류(Int. Cl.)
G06T 11/00 (2006.01) *G06N 3/0455* (2023.01)
G06N 3/094 (2023.01) *G06T 3/40* (2024.01)
G06T 5/50 (2024.01)
(52) CPC특허분류
G06T 11/001 (2024.01)
G06N 3/0455 (2023.01)
(21) 출원번호 10-2023-0014792
(22) 출원일자 2023년02월03일
심사청구일자 2023년02월03일

(71) 출원인
조선대학교산학협력단
광주광역시 동구 조선대길 126(서석동, 산학협력단)
(72) 발명자
이범식
광주광역시 남구 제석로80번길 5, 102동 903호 (봉선동, 제일풍경채 ELITE PARK)
샤피크 무함마드 함자
광주광역시 동구 지산로52번길 11-3, 105호 (지산동, 쌍용하이빌)
(74) 대리인
특허법인지원

전체 청구항 수 : 총 6 항

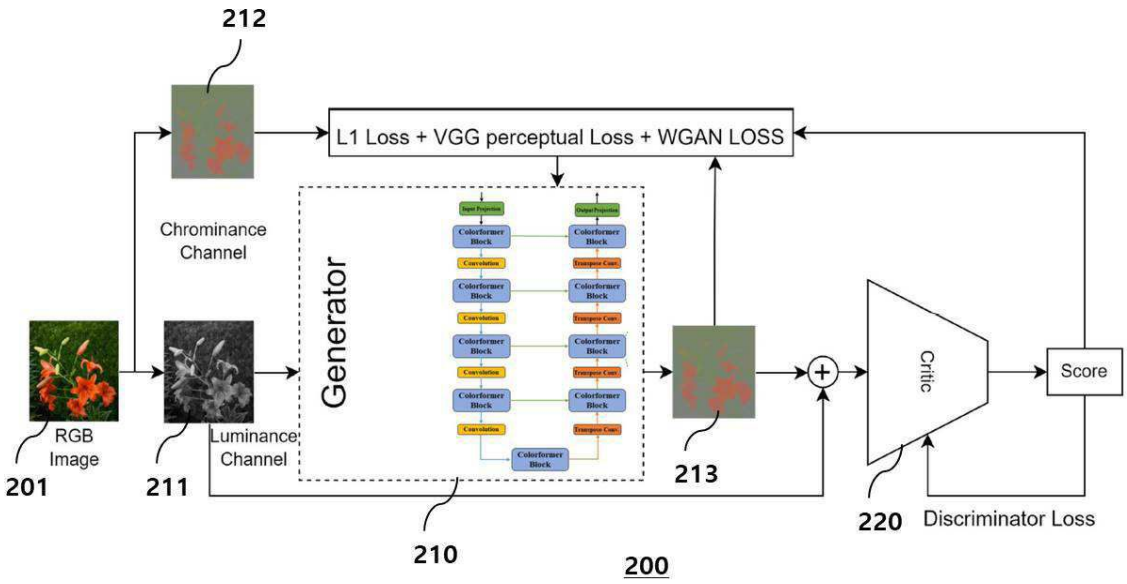
(54) 발명의 명칭 **트랜스포머 블록을 포함하는 GAN 기반의 모델을 이용하여 흑백 이미지를 컬러링하기 위한 전자 장치 및 그 동작 방법**

(57) 요 약

다양한 실시예들에 따라서, 트랜스포머 블록을 포함하는 GAN(Generative Adversarial Network) 기반의 모델을 이용하여 흑백 이미지를 컬러링하기 위한 전자 장치는, 프로세서를 포함하고, 상기 프로세서는, 휘도 채널(luminance channel)에 관한 제1 정보만 포함하는 흑백 이미지를 획득하고, 및 상기 흑백 이미지를 상기 GAN 기

(뒷면에 계속)

대 표 도 - 도2



반의 모델에 적용함으로써, 색차 채널(chrominance channel)에 관한 제2 정보만 포함하는 슈도 컬러 이미지(pseudo color image)를 생성하도록 설정되고, 상기 GAN 기반의 모델은, 색상 변환을 위한 복수의 트랜스포머 블록들(transformer blocks), 복수의 컨볼루션(convolution) 레이어들, 및 복수의 트랜스포스 컨볼루션(transpose convolution) 레이어들을 포함하는 생성기 네트워크(generator network)를 포함하고, 상기 복수의 트랜스포머 블록들 각각은, DWC(Depth Wise Convolution) 레이어, 제1 LN(Layer Normalization) 레이어, W-MSA(Window-based Multi-head Self Attention) 레이어, 제2 LN 레이어, 및 CFF(Colorization Feed Forward) 블록을 포함하고, 상기 W-MSA 레이어는, 제1 그룹의 MSA 및 제2 그룹의 MSA를 포함하고, 제1 그룹의 MSA에는 제1 형태의 윈도우로 분할된 특징 맵(feature map)이 제공되고, 제2 그룹의 MSA에는 상기 제1 형태의 윈도우가 이동된(shifted) 제2 형태의 윈도우로 분할된 특징 맵이 제공될 수 있다. 그 밖의 다양한 실시예들도 가능하다.

본 과제(결과물)는 2022년도 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 지역대학우수과학자지원사업의 결과이다. (2022R1I1A3065473)

This work was supported by the National Research Foundation of Korea (NRF) funded by the Korean Government under Grant. (2022R1I1A3065473)

(52) CPC특허분류
G06N 3/094 (2023.01)
G06T 3/4046 (2024.01)
G06T 5/50 (2024.01)
G06T 2207/20081 (2013.01)
G06T 2207/20084 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1345354483
과제번호	2022R1I1A3065473
부처명	교육부
과제관리(전문)기관명	한국연구재단
연구사업명	지역대학우수과학자지원사업
연구과제명	과거 영상 복원을 위한 영상 및 비디오 디올디파잉 연구
기 여 율	1/1
과제수행기관명	조선대학교
연구기간	2022.06.01 ~ 2025.05.31

명 세 서

청구범위

청구항 1

트랜스포머 블록을 포함하는 GAN(Generative Adversarial Network) 기반의 모델을 이용하여 흑백 이미지를 컬러링하기 위한 전자 장치에 있어서,

프로세서를 포함하고,

상기 프로세서는,

휘도 채널(luminance channel)에 관한 제1 정보만 포함하는 흑백 이미지를 획득하고, 및

상기 흑백 이미지를 상기 GAN 기반의 모델에 적용함으로써, 색차 채널(chrominance channel)에 관한 제2 정보만 포함하는 슈도 컬러 이미지(pseudo color image)를 생성하도록 설정되고,

상기 GAN 기반의 모델은, 색상 변환을 위한 복수의 트랜스포머 블록들(transformer blocks), 복수의 컨볼루션(convolution) 레이어들, 및 복수의 트랜스포스 컨볼루션(transpose convolution) 레이어들을 포함하는 생성기 네트워크(generator network)를 포함하고,

상기 복수의 트랜스포머 블록들 각각은, DWC(Depth Wise Convolution) 레이어, 제1 LN(Layer Normalization) 레이어, W-MSA(Window-based Multi-head Self Attention) 레이어, 제2 LN 레이어, 및 CFF(Colorization Feed Forward) 블록을 포함하고,

상기 W-MSA 레이어는, 제1 그룹의 MSA 및 제2 그룹의 MSA를 포함하고,

제1 그룹의 MSA에는 제1 형태의 윈도우로 분할된 특징 맵(feature map)이 제공되고, 제2 그룹의 MSA에는 상기 제1 형태의 윈도우가 이동된(shifted) 제2 형태의 윈도우로 분할된 특징 맵이 제공되는,

전자 장치.

청구항 2

제1 항에 있어서,

상기 제1 그룹의 MSA 및 상기 제2 그룹의 MSA는 각각 4개의 헤드로 구성되는,

전자 장치.

청구항 3

제2 항에 있어서,

상기 W-MSA 레이어는 상기 제1 그룹의 MSA의 결과 값 및 상기 제2 그룹의 MSA의 결과 값을 연결(concatenation)한 값을 다음 레이어로 전달하는,

전자 장치.

청구항 4

제3 항에 있어서,

상기 생성기 네트워크는, 상기 복수의 트랜스포머 블록들을 이용하는 인코더(encoder)-디코더(decoder) 기반의 아키텍처를 가지는,

전자 장치.

청구항 5

제4 항에 있어서,

상기 GAN 기반의 모델은 pixel wise(L1) 손실 함수(L_{L1}), VGG 손실 함수(L_{VGG}), 및 WGAN 손실 함수(L_{wgan})를 모두 고려한 총 손실(L_{total})을 이용하여 훈련되고,

상기 pixel wise(L1) 손실 함수(L_{L1}), 상기 VGG 손실 함수(L_{VGG}), 상기 WGAN 손실 함수(L_{wgan}), 및 상기 총 손실(L_{total})은 하기의 [수학식 1]에 의하여 산출되는,

전자 장치.

[수학식 1]

$$① L_{wgan} \begin{cases} L_G = -E_{\tilde{y}}[D(\tilde{y},x)] \\ L_D = E_y[D(y,x)] - E_{\tilde{y}}[D(\tilde{y},x)] + \lambda \times GP \end{cases}$$

$$② L_{L1} = \left|J - G(I)\right|_1$$

$$③ L_{VGG} = \left|\left|\varphi(J) - \varphi(G(I))\right|\right|_2^2$$

$$④ L_{total} = L_{wgan} + \lambda_1 L_{L1} + \lambda_2 L_{VGG}$$

J는 ground truth에 해당하는 컬러 이미지를 나타내고, G(I)는 슈도 컬러 이미지를 나타내며, λ, λ_1, λ_2는 가중치를 나타냄.

청구항 6

제5 항에 있어서,

상기 프로세서는,

상기 GAN 기반 모델을 통하여 상기 흑백 이미지와 상기 슈도 컬러 이미지를 합성함으로써, 최종 컬러 이미지를 생성하도록 설정된,

전자 장치.

발명의 설명

기 술 분 야

[0001] 본 발명의 다양한 실시예는, 트랜스포머 블록을 포함하는 GAN(Generative Adversarial Network) 기반의 모델을 이용하여 흑백 이미지를 컬러링하기 위한 전자 장치 및 그 동작 방법에 관한 것이다.

배 경 기 술

[0002] GAN(Generative Adversarial Network, 생성적 대립 신경망)은 생성자와 분별자가 서로 경쟁하면서 실제 이미지와 거의 유사한 이미지를 자동으로 생성할 수 있는 기계 학습(Machine Learning) 모델이다.

[0003] GAN은 가짜 이미지를 생성하여, 분별자를 최대한 속일 수 있도록 학습된 생성자 및 가짜 이미지 및 실제 이미지를 최대한 정확하게 분류할 수 있도록 학습된 분별자로 구성된다. GAN은 생성자 및 분별자를 순차적으로 반복 학습시키는 대립적 프로세스를 통해 발전시키는 과정을 통해 실제 이미지와 매우 유사한 가짜 이미지를 생성할 수 있다.

[0004] 또한, Transformer는 자연어 처리(NLP, Natural Language Processing) 분야에서 활발히 연구되었고, 현재는 컴

퓨터 비전 분야에서도 다양한 문제에 응용되어 사용되고 있다.

선행기술문헌

특허문헌

[0005] (특허문헌 0001) 대한민국 공개특허 제10-2020-0120970호(2020.10.23.공개) "Gan 기반 딥러닝 모델을 이용한 이미지 생성 방법 및 장치"

발명의 내용

해결하려는 과제

[0006] 본 발명은 트랜스포머 블록을 포함하는 GAN 기반의 모델을 이용하여 한 개의 휘도 채널로 구성된 흑백 이미지로부터 두 개의 색차 채널로 구성된 컬러 이미지를 생성하는 전자 장치를 제공한다.

과제의 해결 수단

[0007] 다양한 실시예들에 따라서, 트랜스포머 블록을 포함하는 GAN(Generative Adversarial Network) 기반의 모델을 이용하여 흑백 이미지를 컬러링하기 위한 전자 장치는, 프로세서를 포함하고, 상기 프로세서는, 휘도 채널(luminance channel)에 관한 제1 정보만 포함하는 흑백 이미지를 획득하고, 및 상기 흑백 이미지를 상기 GAN 기반의 모델에 적용함으로써, 색차 채널(chrominance channel)에 관한 제2 정보만 포함하는 슈도 컬러 이미지(pseudo color image)를 생성하도록 설정되고, 상기 GAN 기반의 모델은, 색상 변환을 위한 복수의 트랜스포머 블록들(transformer blocks), 복수의 컨볼루션(convolution) 레이어들, 및 복수의 트랜스포스 컨볼루션(transpose convolution) 레이어들을 포함하는 생성기 네트워크(generator network)를 포함하고, 상기 복수의 트랜스포머 블록들 각각은, DWC(Depth Wise Convolution) 레이어, 제1 LN(Layer Normalization) 레이어, W-MSA(Window-based Multi-head Self Attention) 레이어, 제2 LN 레이어, 및 CFF(Colorization Feed Forward) 블록을 포함하고, 상기 W-MSA 레이어는, 제1 그룹의 MSA 및 제2 그룹의 MSA를 포함하고, 제1 그룹의 MSA에는 제1 형태의 윈도우로 분할된 특징 맵(feature map)이 제공되고, 제2 그룹의 MSA에는 상기 제1 형태의 윈도우가 이동된(shifted) 제2 형태의 윈도우로 분할된 특징 맵이 제공될 수 있다.

발명의 효과

[0008] 본 발명은 트랜스포머 블록을 포함하는 GAN 기반의 모델을 이용하여 흑백 이미지를 컬러링함으로써, 사실에 가까운 컬러 이미지를 생성하는 효과를 제공할 수 있고, 특히, 트랜스포머 블록 내의 W-MSA 레이어에서 shifted window를 통하여 특징 맵을 처리함으로써, 다른 색상화 모델에 비하여 고품질의 컬러링 효과를 제공할 수 있다.

도면의 간단한 설명

[0009] 도 1은 본 발명의 다양한 실시예에 따른 전자 장치 및 네트워크의 블록도를 도시한다.

도 2는 다양한 실시예들에 따른, 전자 장치가 트랜스포머를 포함하는 GAN 기반의 모델을 이용하여 흑백 이미지를 컬러링하는 동작을 설명하기 위한 도면이다.

도 3은 다양한 실시예들에 따른, GAN 기반의 모델에 포함된 트랜스포머 블록의 구조를 설명하기 위한 도면이다.

도 4는 다양한 실시예들에 따른, 트랜스포머 블록 내의 W-MSA 레이어의 구조를 설명하기 위한 도면이다.

도 5는 다양한 실시예들에 따른, CFF 레이어의 구조를 설명하기 위한 도면이다.

도 6은 다양한 실시예들에 따른, 전자 장치가 GAN 기반의 모델을 이용하여 흑백 이미지로부터 최종 컬러 이미지를 생성하는 실시예를 나타낸다.

도 7a 및 7b는 다양한 실시예들에 따른, 본 발명이 적용된 GAN 기반의 모델과 다른 색상화 모델들 사이의 차이를 나타내는 도면이다.

발명을 실시하기 위한 구체적인 내용

[0010] 이하, 본 문서의 다양한 실시예들이 첨부된 도면을 참조하여 기재된다. 실시예 및 이에 사용된 용어들은 본 문서에 기재된 기술을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 해당 실시예의 다양한 변경, 균등물, 및/또는 대체물을 포함하는 것으로 이해되어야 한다. 도면의 설명과 관련하여, 유사한 구성요소에 대해서는 유사한 참조 부호가 사용될 수 있다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함할 수 있다. 본 문서에서, "A 또는 B" 또는 "A 및/또는 B 중 적어도 하나" 등의 표현은 함께 나열된 항목들의 모든 가능한 조합을 포함할 수 있다. "제 1," "제 2," "첫째," 또는 "둘째,"등의 표현들은 해당 구성요소들을, 순서 또는 중요도에 상관없이 수식할 수 있고, 한 구성요소를 다른 구성요소와 구분하기 위해 사용될 뿐 해당 구성요소들을 한정하지 않는다. 어떤(예: 제 1) 구성요소가 다른(예: 제 2) 구성요소에 "(기능적으로 또는 통신적으로) 연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 상기 어떤 구성요소가 상기 다른 구성요소에 직접적으로 연결되거나, 다른 구성요소(예: 제 3 구성요소)를 통하여 연결될 수 있다.

[0011] 본 문서에서, "~하도록 구성된(또는 설정된)(configured to)"은 상황에 따라, 예를 들면, 하드웨어적 또는 소프트웨어적으로 "~에 적합한," "~하는 능력을 가지는," "~하도록 변경된," "~하도록 만들어진," "~를 할 수 있는," 또는 "~하도록 설계된"과 상호 호환적으로(interchangeably) 사용될 수 있다. 어떤 상황에서는, "~하도록 구성된 장치"라는 표현은, 그 장치가 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 메모리 장치에 저장된 하나 이상의 소프트웨어 프로그램들을 실행함으로써, 해당 동작들을 수행할 수 있는 범용 프로세서(예: CPU 또는 application processor)를 의미할 수 있다.

[0012] 본 문서의 다양한 실시예들에 따른 전자 장치는, 예를 들면, 스마트폰, 태블릿 PC, 데스크탑 PC, 랩탑 PC, 넷북 컴퓨터, 워크스테이션, 서버 중 적어도 하나를 포함할 수 있다.

[0013] 도 1을 참조하여, 다양한 실시예에서의, 네트워크 환경(100) 내의 전자 장치(101)가 기재된다. 전자 장치(101)는 버스(110), 프로세서(120), 메모리(130), 입출력 인터페이스(150), 디스플레이(160), 및 통신 인터페이스(170)를 포함할 수 있다. 어떤 실시예에서는, 전자 장치(101)는, 구성요소들 중 적어도 하나를 생략하거나 다른 구성요소를 추가적으로 구비할 수 있다. 버스(110)는 구성요소들(110-170)을 서로 연결하고, 구성요소들 간의 통신(예: 제어 메시지 또는 데이터)을 전달하는 회로를 포함할 수 있다. 프로세서(120)는, 중앙처리장치, 어플리케이션 프로세서, 또는 커뮤니케이션 프로세서(communication processor(CP)) 중 하나 또는 그 이상을 포함할 수 있다. 프로세서(120)는, 예를 들면, 전자 장치(101)의 적어도 하나의 다른 구성요소들의 제어 및/또는 통신에 관한 연산이나 데이터 처리를 실행할 수 있다.

[0014] 메모리(130)는, 휘발성 및/또는 비휘발성 메모리를 포함할 수 있다. 메모리(130)는, 예를 들면, 전자 장치(101)의 적어도 하나의 다른 구성요소에 관계된 명령 또는 데이터를 저장할 수 있다. 한 실시예에 따르면, 메모리(130)는 소프트웨어 및/또는 프로그램(140)을 저장할 수 있다. 프로그램(140)은, 예를 들면, 커널(141), 미들웨어(143), 어플리케이션 프로그래밍 인터페이스(API)(145), 및/또는 어플리케이션 프로그램(또는 "어플리케이션")(147) 등을 포함할 수 있다. 커널(141), 미들웨어(143), 또는 API(145)의 적어도 일부는, 운영 시스템으로 지칭될 수 있다. 커널(141)은, 예를 들면, 다른 프로그램들(예: 미들웨어(143), API(145), 또는 어플리케이션 프로그램(147))에 구현된 동작 또는 기능을 실행하는 데 사용되는 시스템 리소스들(예: 버스(110), 프로세서(120), 또는 메모리(130) 등)을 제어 또는 관리할 수 있다. 또한, 커널(141)은 미들웨어(143), API(145), 또는 어플리케이션 프로그램(147)에서 전자 장치(101)의 개별 구성요소에 접근함으로써, 시스템 리소스들을 제어 또는 관리할 수 있는 인터페이스를 제공할 수 있다.

[0015] 미들웨어(143)는, 예를 들면, API(145) 또는 어플리케이션 프로그램(147)이 커널(141)과 통신하여 데이터를 주고받을 수 있도록 중개 역할을 수행할 수 있다. 또한, 미들웨어(143)는 어플리케이션 프로그램(147)으로부터 수신된 하나 이상의 작업 요청들을 우선 순위에 따라 처리할 수 있다. 예를 들면, 미들웨어(143)는 어플리케이션 프로그램(147) 중 적어도 하나에 전자 장치(101)의 시스템 리소스(예: 버스(110), 프로세서(120), 또는 메모리(130) 등)를 사용할 수 있는 우선 순위를 부여하고, 상기 하나 이상의 작업 요청들을 처리할 수 있다. API(145)는 어플리케이션(147)이 커널(141) 또는 미들웨어(143)에서 제공되는 기능을 제어하기 위한 인터페이스로, 예를 들면, 파일 제어, 창 제어, 영상 처리, 또는 문자 제어 등을 위한 적어도 하나의 인터페이스 또는 함수(예: 명령어)를 포함할 수 있다. 입출력 인터페이스(150)는, 예를 들면, 사용자 또는 다른 외부 기기로부터 입력된 명령 또는 데이터를 전자 장치(101)의 다른 구성요소(들)에 전달하거나, 또는 전자 장치(101)의 다른 구성요소(들)로부터 수신된 명령 또는 데이터를 사용자 또는 다른 외부 기기로 출력할 수 있다.

[0016] 디스플레이(160)는, 예를 들면, 액정 디스플레이(LCD), 발광 다이오드(LED) 디스플레이, 유기 발광 다이오드

(OLED) 디스플레이, 또는 마이크로 전자기계 시스템 (MEMS) 디스플레이, 또는 전자종이(electronic paper) 디스플레이를 포함할 수 있다. 디스플레이(160)는, 예를 들면, 사용자에게 각종 콘텐츠(예: 텍스트, 이미지, 비디오, 아이콘, 및/또는 심볼 등)을 표시할 수 있다. 디스플레이(160)는, 터치 스크린을 포함할 수 있으며, 예를 들면, 전자 펜 또는 사용자의 신체의 일부를 이용한 터치, 제스쳐, 근접, 또는 호버링 입력을 수신할 수 있다. 통신 인터페이스(170)는, 예를 들면, 전자 장치(101)와 외부 장치(예: 제 1 외부 전자 장치(102), 제 2 외부 전자 장치(104), 또는 서버(106)) 간의 통신을 설정할 수 있다. 예를 들면, 통신 인터페이스(170)는 무선 통신 또는 유선 통신을 통해서 네트워크(162)에 연결되어 외부 장치(예: 제 2 외부 전자 장치(104) 또는 서버(106))와 통신할 수 있다.

[0017] 무선 통신은, 예를 들면, LTE, LTE-A(LTE Advance), CDMA(code division multiple access), WCDMA(wideband CDMA), UMTS(universal mobile telecommunications system), WiBro(Wireless Broadband), 또는 GSM(Global System for Mobile Communications) 등 중 적어도 하나를 사용하는 셀룰러 통신을 포함할 수 있다. 한 실시예에 따르면, 무선 통신은, 예를 들면, WiFi(wireless fidelity), 블루투스, 블루투스 저전력(BLE), 지그비(Zigbee), NFC(near field communication), 자력 시큐어 트랜스미션(Magnetic Secure Transmission), 라디오 프리퀀시(RF), 또는 보디 에어리어 네트워크(BAN) 중 적어도 하나를 포함할 수 있다. 한실시예에 따르면, 무선 통신은 GNSS를 포함할 수 있다. GNSS는, 예를 들면, GPS(Global Positioning System), Glonass(Global Navigation Satellite System), Beidou Navigation Satellite System(이하 "Beidou") 또는 Galileo, the European global satellite-based navigation system일 수 있다. 이하, 본 문서에서는, "GPS"는 "GNSS"와 상호 호환적으로 사용될 수 있다. 유선 통신은, 예를 들면, USB(universal serial bus), HDMI(high definition multimedia interface), RS-232(recommended standard232), 전력선 통신, 또는 POTS(plain old telephone service) 등 중 적어도 하나를 포함할 수 있다. 네트워크(162)는 텔레커뮤니케이션 네트워크, 예를 들면, 컴퓨터 네트워크(예: LAN 또는 WAN), 인터넷, 또는 텔레폰 네트워크 중 적어도 하나를 포함할 수 있다.

[0018] 제 1 및 제 2 외부 전자 장치(102, 104) 각각은 전자 장치(101)와 동일한 또는 다른 종류의 장치일 수 있다. 다양한 실시예에 따르면, 전자 장치(101)에서 실행되는 동작들의 전부 또는 일부는 다른 하나 또는 복수의 전자 장치(예: 전자 장치(102,104), 또는 서버(106)에서 실행될 수 있다. 한 실시예에 따르면, 전자 장치(101)가 어떤 기능이나 서비스를 자동으로 또는 요청에 의하여 수행해야 할 경우에, 전자 장치(101)는 기능 또는 서비스를 자체적으로 실행시키는 대신에 또는 추가적으로, 그와 연관된 적어도 일부 기능을 다른 장치(예: 전자 장치(102, 104), 또는 서버(106))에게 요청할 수 있다. 다른 전자 장치(예: 전자 장치(102, 104), 또는 서버(106))는 요청된 기능 또는 추가 기능을 실행하고, 그 결과를 전자 장치(101)로 전달할 수 있다. 전자 장치(101)는 수신된 결과를 그대로 또는 추가적으로 처리하여 요청된 기능이나 서비스를 제공할 수 있다. 이를 위하여, 예를 들면, 클라우드 컴퓨팅, 분산 컴퓨팅, 또는 클라이언트-서버 컴퓨팅 기술이 이용될 수 있다.

[0020] 도 2는 다양한 실시예들에 따른, 전자 장치(예: 도 1의 전자 장치(101))가 트랜스포머를 포함하는 GAN 기반의 모델을 이용하여 흑백 이미지를 컬러링하는 동작을 설명하기 위한 도면이다.

[0021] 도 3은 다양한 실시예들에 따른, GAN 기반의 모델에 포함된 트랜스포머 블록의 구조를 설명하기 위한 도면이다.

[0022] 도 4는 다양한 실시예들에 따른, 트랜스포머 블록 내의 W-MSA 레이어의 구조를 설명하기 위한 도면이다.

[0023] 도 5는 다양한 실시예들에 따른, CFF 레이어의 구조를 설명하기 위한 도면이다.

[0024] 다양한 실시예들에 따르면, 전자 장치(예: 도 1의 프로세서(120))는 휘도 채널(luminance channel)에 관한 제1 정보만 포함하는 흑백 이미지를 획득할 수 있다. 일 실시예에 따라, 도 2를 참조하면, 전자 장치(101)는 트랜스포머 블록(transformer block)을 포함하는 GAN 기반의 모델(200)을 훈련시키기 위하여, RGB 이미지(201)를 이용하여 휘도 채널(luminance channel)에 관한 제1 정보만 포함하는 흑백 이미지(211) 및 색차 채널(chrominance channel)에 관한 제2 정보만 포함하는 정답 컬러 이미지(212)를 생성할 수 있다. 예를 들어, 전자 장치(101)는 RGB 이미지(201)를 CIELAB 색 공간(color space)의 이미지로 변환시킬 수 있고, CIELAB 색 공간의 L값, A값, B값 중에서 L값을 추출하여 휘도 채널에 관한 정보를 포함하는 흑백 이미지(211)를 생성하고, L값 및 A값을 추출하여 색차 채널에 관한 정보를 포함하는 정답 컬러 이미지(212)를 생성할 수 있다. 휘도 채널은 1개의 채널로 구성될 수 있고, 색차 채널은 두 개의 색상 채널들로 구성될 수 있다. 일 실시예에 따르면, 전자 장치(101)는 훈련된 GAN 기반의 모델(200)을 이용하여 이미지 컬러링을 수행하기 위하여, 외부 장치(예: 도 1의 전자 장치(102) 또는 서버(106))로부터 컬러링을 수행할 흑백 이미지(211)를 획득할 수 있다.

[0025] 다양한 실시예들에 따르면, 전자 장치(예: 도 1의 프로세서(120))는 흑백 이미지(211)를 GAN 기반의 모델(200)에 적용함으로써, 색차 채널에 관한 제2 정보만 포함하는 슈도 컬러 이미지(pseudo color image)(213)를 생성할 수 있다. 일 실시예에 따라, 도 2를 참조하면, 전자 장치(101)는 흑백 이미지(211) 이미지를 GAN 기반의 모델(200)에 포함된 생성기 네트워크(210)에 적용함으로써, 색차 채널에 관한 제2 정보만 포함하는 슈도 컬러 이미지(213)를 생성할 수 있다. 슈도 컬러 이미지(213)는 흑백 이미지(211)에 대하여 컬러링을 수행한 이미지를 의미할 수 있다.

[0026] 다양한 실시예들에 따라, 도 3을 참조하면, GAN 기반의 모델(200)은 색상 변환을 위한 복수의 트랜스포머 블록들(301), 복수의 컨볼루션 레이어들(302), 및 복수의 트랜스포스 컨볼루션 레이어들(303)을 포함하는 생성기 네트워크(210)를 포함할 수 있다. 일 실시예에 따르면, 생성기 네트워크(210)는, 복수의 트랜스포머 블록들을 이용하는 인코더(encoder)-디코더(decoder) 기반의 아키텍처를 가질 수 있다.

[0027] 일 실시예에 따라, 도 3을 참조하면, 트랜스포머 블록(301)은 DWC(Depth Wise Convolution) 레이어(311), 제1 LN(Layer Normalization) 레이어(312), W-MSA(Window-based Multi-head Self Attention) 레이어(313), 제2 LN 레이어(314), 및 CFF(Colorization Feed Forward) 블록(315)을 포함할 수 있다.

[0028] 일 실시예에 따르면, DWC 레이어(311)는 로컬 정보(local information)를 획득하기 위하여 트랜스포머 블록(301) 내에 구현될 수 있다. 픽셀 색상은 주변 픽셀 색상에 따라 달라질 수 있으므로, 색상화(colorization) 동작에서 로컬 정보는 중요하다. 후술하겠지만, CFF 블록(315) 또한 로컬 정보를 획득하기 위하여 DWC 레이어를 포함할 수 있다.

[0029] 일 실시예에 따르면, 제1 LN 레이어(312) 및 제2 LN 레이어(314)는 기존의 트랜스포머 블록에서 사용되는 LN 레이어를 포함할 수 있고, 이는 이 기술분야의 당업자에 의하여 용이하게 실시 가능한 구성이므로, 이에 관한 설명은 생략하도록 한다.

[0030] 일 실시예에 따라, 도 4를 참조하면, W-MSA 레이어(313)는 제1 그룹의 MSA(401) 및 제2 그룹의 MSA(402)를 포함할 수 있고, 상기 제1 그룹의 MSA(401)에는 제1 형태의 윈도우로 분할된 특징 맵(feature map)(410)이 제공되고, 제2 그룹의 MSA(402)에는 제1 형태의 윈도우가 이동된(shifted) 제2 형태의 윈도우로 분할된 특징 맵(420)이 제공될 수 있다. 일 실시예에 따르면, 제1 그룹의 MSA(401) 및 제2 그룹의 MSA(402)는 각각 4개의 헤드로 구성되어, W-MSA 레이어(313)는 총 8개의 헤드를 포함할 수 있다. 구체적으로, 제1 형태의 윈도우로 분할된 특징 맵(410)의 각 윈도우는 제1 그룹의 MSA(401) 각각으로 제공되고, 제2 형태의 윈도우로 분할된 특징 맵(420)의 각 윈도우는 제2 그룹의 MSA(402) 각각으로 제공될 수 있다. 이 때, W-MSA 레이어(313)는 제1 그룹의 MSA(401) 및 제2 그룹의 MSA(402)의 결과 값을 연결(concatenation)한 값을 다음 레이어(예: 제2 LN 레이어(314))로 전달할 수 있다.

[0031] 일 실시예에 따라, 도 5를 참조하면, CFF 블록(315)는 로컬 정보(local information)를 획득하도록 구현될 수 있고, 제1 선형 레이어(501), DWC 레이어(502), 및 제2 선형 레이어(503)를 포함할 수 있다. 상기 로컬 정보는, 대다수의 트랜스포머 블록에서 MLP(Multi-Layer Perceptron)를 이용하여 전역 정보(global information)를 획득하는 과정만큼 중요할 수 있다. CFF 블록(315)은 특징들(features)이 제1 선형 레이어(501)를 통과한 후에, 상기 특징들은 DWC 레이어(502)로 입력하기 적합한 구조로 reshape되고, complexity를 감소시키기 위하여 DWC 레이어(502)가 적용될 수 있다. CFF 블록(315)는 특징들을 DWC 레이어(502)를 통하여 convolution한 후에, 특징들에 대하여 flatten 동작을 수행하여 제2 선형 레이어(503)에 통과시킬 수 있다.

[0032] 일 실시예에 따라, 도 2를 참조하면, 전자 장치(101)는 GAN 기반 모델(200)을 통하여, 생성기 네트워크(210)에 의하여 생성된 슈도 컬러 이미지(213)를 입력 이미지인 흑백 이미지(211)와 합성(예: concatenation)함으로써, 최종 컬러 이미지를 생성할 수 있다.

[0033] 일 실시예에 따라, 도 2를 참조하면, 판별기 네트워크(220)(예: critic)는 기존의 GAN 기반의 모델에서 사용되는 판별기 네트워크를 포함할 수 있고, 이는 이 기술분야의 당업자에 의하여 용이하게 실시 가능한 구성이므로, 이에 관한 설명은 생략하도록 한다.

[0034] 일 실시예에 따라, 도 2를 참조하면, GAN 기반의 모델(200)은 pixel wise(L1) 손실 함수(L_{L1}), VGG 손실 함수(L_{VGG}), 및 WGAN 손실 함수(L_{wgan})를 모두 고려한 총 손실(L_{total})을 이용하여 훈련될 수 있다. 이 때, 상기 pixel wise(L1) 손실 함수(L_{L1}), 상기 VGG 손실 함수(L_{VGG}), 상기 WGAN 손실 함수(L_{wgan}), 및 상기 총 손실(L_{total})은 하기의 [수학식 1]에 의하여 산출될 수 있다.

수학식 1

[0035]

$$① L_{wgan} \begin{cases} L_G = -E_{\tilde{y}}[D(\tilde{y},x)] \\ L_D = E_y[D(y,x)] - E_{\tilde{y}}[D(\tilde{y},x)] + \lambda \times GP \end{cases}$$

$$② L_{L1} = \left|\left|J - G(I)\right|\right|_1$$

$$③ L_{VGG} = \left|\left|\varphi(J) - \varphi(G(I))\right|\right|_2^2$$

$$④ L_{total} = L_{wgan} + \lambda_1 L_{L1} + \lambda_2 L_{VGG}$$

[0036] 여기서, J는 ground truth에 해당하는 컬러 이미지를 나타내고, G(I)는 슈도 컬러 이미지를 나타내며, λ, λ_1, λ_2는 가중치를 나타낼 수 있다.

[0037] 일 실시예에 따르면, 생성기 네트워크(210)는 pixel wise(L1) 손실 함수(L_{L1}), VGG 손실 함수(L_{VGG}), 및 WGAN 손실 함수(L_{wgan})를 합산한 총 손실에 기초하여 훈련될 수 있다.

[0039] 도 6은 다양한 실시예들에 따른, 전자 장치(예: 도 1의 전자 장치(101))가 GAN 기반의 모델(예: 도 2의 모델(200))을 이용하여 흑백 이미지로부터 최종 컬러 이미지를 생성하는 실시예를 나타낸다.

[0040] 본 발명에 따른 전자 장치(101)는 휘도 채널에 관한 정보만 포함하는 흑백 이미지, 예를 들어 "a) input",를 획득하여, GAN 기반의 모델(200)을 이용하여 최종 컬러 이미지, 예를 들어 "b) colorized",를 생성할 수 있다. 이때, 전자 장치(101)는 학습 이미지, 예를 들어 "c) ground truth",를 이용하여 GAN 기반의 모델(200)을 훈련시킬 수 있다.

[0042] 도 7a 및 7b는 다양한 실시예들에 따른, 본 발명이 적용된 GAN 기반의 모델(예: 도 2의 모델(200))과 다른 색상화 모델들 사이의 차이를 나타내는 도면이다.

[0043] 도 7a를 참조하면, 본 발명이 적용된 GAN 기반의 모델(200)을 이용하여 입력 흑백 이미지로부터 생성된 최종 컬러 이미지(예: proposed method)는 다른 색상화 모델을 이용하여 생성된 컬러 이미지(예: ChromaGAN, Deoldify, ColTrans)에 비하여 원본 이미지(예: Groud Truth)에 더 가까운 것을 알 수 있다.

[0044] 또한 도 7b를 참조하면, 본 발명이 적용된 GAN 기반의 모델(200)을 이용하였을 때, 다른 색상화 모델을 이용할 때보다 품질이 가장 좋거나, 또는 원본 이미지와 가장 유사한 형태의 컬러 이미지를 생성하는 것을 알 수 있다.

[0046] 다양한 실시예들에 따라서, 트랜스포머 블록을 포함하는 GAN(Generative Adversarial Network) 기반의 모델을 이용하여 흑백 이미지를 컬러링하기 위한 전자 장치는, 프로세서를 포함하고, 상기 프로세서는, 휘도 채널(luminance channel)에 관한 제1 정보만 포함하는 흑백 이미지를 획득하고, 및 상기 흑백 이미지를 상기 GAN 기반의 모델에 적용함으로써, 색차 채널(chrominance channel)에 관한 제2 정보만 포함하는 슈도 컬러 이미지(pseudo color image)를 생성하도록 설정되고, 상기 GAN 기반의 모델은, 색상 변환을 위한 복수의 트랜스포머 블록들(transformer blocks), 복수의 컨볼루션(convolution) 레이어들, 및 복수의 트랜스포스 컨볼루션(transpose convolution) 레이어들을 포함하는 생성기 네트워크(generator network)를 포함하고, 상기 복수의 트랜스포머 블록들 각각은, DWC(Depth Wise Convolution) 레이어, 제1 LN(Layer Normalization) 레이어, W-MSA(Window-based Multi-head Self Attention) 레이어, 제2 LN 레이어, 및 CFF(Colorization Feed Forward) 블록을 포함하고, 상기 W-MSA 레이어는, 제1 그룹의 MSA 및 제2 그룹의 MSA를 포함하고, 제1 그룹의 MSA에는 제1 형태의 윈도우로 분할된 특징 맵(feature map)이 제공되고, 제2 그룹의 MSA에는 상기 제1 형태의 윈도우가 이동된(shifted) 제2 형태의 윈도우로 분할된 특징 맵이 제공될 수 있다.

[0047] 다양한 실시예들에 따라서, 상기 제1 그룹의 MSA 및 상기 제2 그룹의 MSA는 각각 4개의 헤드로 구성될 수 있다.

[0048] 다양한 실시예들에 따라서, 상기 W-MSA 레이어는 상기 제1 그룹의 MSA의 결과 값 및 상기 제2 그룹의 MSA의 결과 값을 연결(concatenation)한 값을 다음 레이어로 전달할 수 있다.

[0049] 다양한 실시예들에 따라서, 상기 생성기 네트워크는, 상기 복수의 트랜스포머 블록들을 이용하는 인코더(encoder)-디코더(decoder) 기반의 아키텍처를 가질 수 있다.

[0050] 다양한 실시예들에 따라서, 상기 GAN 기반의 모델은 pixel wise(L1) 손실 함수(L_{L1}), VGG 손실 함수(L_{VGG}), 및 WGAN 손실 함수(L_{wgan})를 모두 고려한 총 손실(L_{total})을 이용하여 훈련되고, 상기 pixel wise(L1) 손실 함수(L_{L1}), 상기 VGG 손실 함수(L_{VGG}), 상기 WGAN 손실 함수(L_{wgan}), 및 상기 총 손실(L_{total})은 하기의 [수학식 1]에 의하여 산출될 수 있다.

[0051] [수학식 1]

$$① L_{wgan} \begin{cases} L_G = -E_{\tilde{y}}[D(\tilde{y},x)] \\ L_D = E_y[D(y,x)] - E_{\tilde{y}}[D(\tilde{y},x)] + \lambda \times GP \end{cases}$$

$$② L_{L1} = \left|\left|J - G(I)\right|\right|_1$$

$$③ L_{VGG} = \left|\left|\varphi(J) - \varphi(G(I))\right|\right|_2^2$$

[0052] $$④ L_{total} = L_{wgan} + \lambda_1 L_{L1} + \lambda_2 L_{VGG}$$

[0053] J는 ground truth에 해당하는 컬러 이미지를 나타내고, G(I)는 슈도 컬러 이미지를 나타내며, λ, λ_1, λ_2는 가중치를 나타냄.

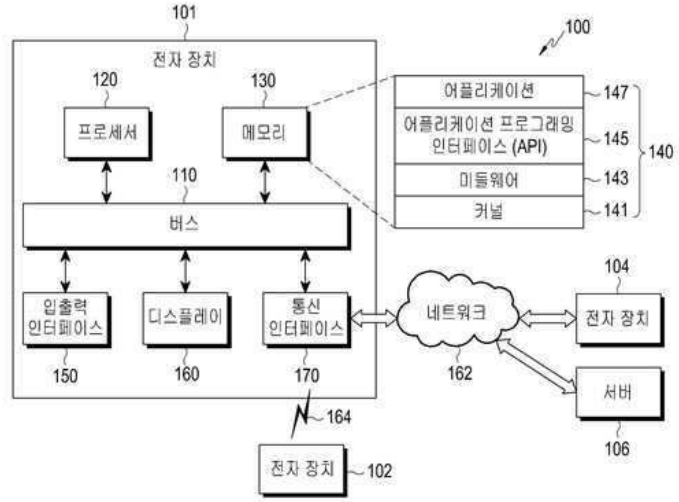
[0054] 다양한 실시예들에 따라서, 상기 프로세서는, 상기 GAN 기반 모델을 통하여 상기 흑백 이미지와 상기 슈도 컬러 이미지를 합성함으로써, 최종 컬러 이미지를 생성하도록 설정될 수 있다.

[0056] 본 문서에서 사용된 용어 "모듈" 또는 "~부"은 하드웨어, 소프트웨어 또는 펌웨어로 구성된 유닛을 포함하며, 예를 들면, 로직, 논리 블록, 부품, 또는 회로 등의 용어와 상호 호환적으로 사용될 수 있다. "모듈" 또는 "~부"는, 일체로 구성된 부품 또는 하나 또는 그 이상의 기능을 수행하는 최소 단위 또는 그 일부가 될 수 있다. "모듈" 또는 "~부"는 기계적으로 또는 전자적으로 구현될 수 있으며, 예를 들면, 어떤 동작들을 수행하는, 알려졌거나 앞으로 개발될, ASIC(application-specific integrated circuit) 칩, FPGAs(field-programmable gate arrays), 또는 프로그램 가능 논리 장치를 포함할 수 있고, 프로세서(120)에 의하여 실행될 수 있다. 다양한 실시예에 따른 장치(예: 모듈들 또는 그 기능들) 또는 방법(예: 동작들)의 적어도 일부는 프로그램 모듈의 형태로 컴퓨터로 판독 가능한 저장 매체(예: 메모리(130))에 저장된 명령어로 구현될 수 있다. 상기 명령어가 프로세서(예: 프로세서(120))에 의해 실행될 경우, 프로세서가 상기 명령어에 해당하는 기능을 수행할 수 있다. 컴퓨터로 판독 가능한 기록 매체는, 하드디스크, 플로피디스크, 마그네틱 매체(예: 자기테이프), 광기록 매체(예: CD-ROM, DVD, 자기-광 매체 (예: 플롭티컬 디스크), 내장 메모리 등을 포함할 수 있다. 명령어는 컴파일러에 의해 만들어지는 코드 또는 인터프리터에 의해 실행될 수 있는 코드를 포함할 수 있다. 다양한 실시예에 따른 모듈 또는 프로그램 모듈은 전술한 구성요소들 중 적어도 하나 이상을 포함하거나, 일부가 생략되거나, 또는 다른 구성요소를 더 포함할 수 있다. 다양한 실시예에 따른, 모듈, 프로그램 모듈 또는 다른 구성요소에 의해 수행되는 동작들은 순차적, 병렬적, 반복적 또는 휴리스틱하게 실행되거나, 적어도 일부 동작이 다른 순서로 실행되거나, 생략되거나, 또는 다른 동작이 추가될 수 있다.

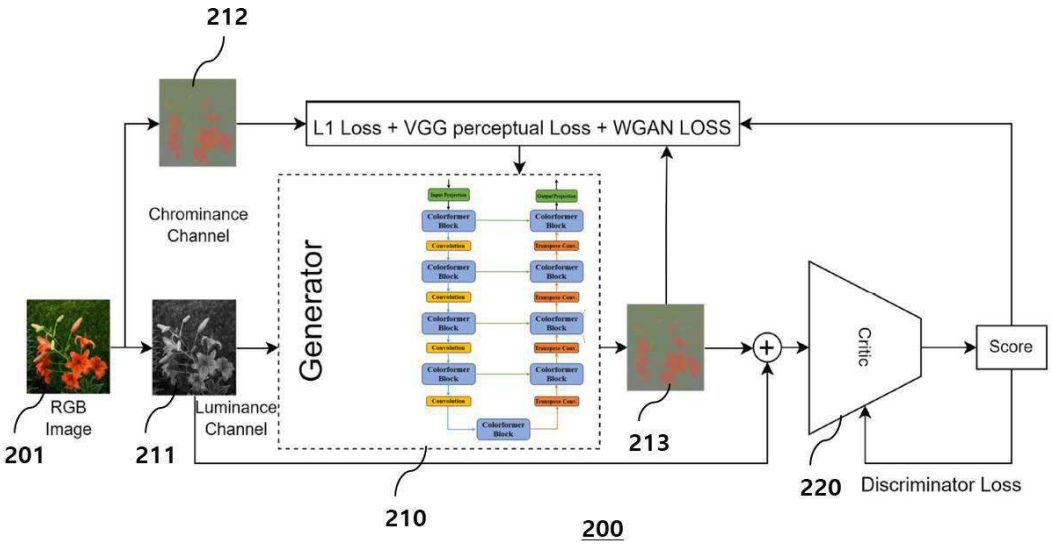
[0057] 그리고 본 문서에 개시된 실시예는 개시된, 기술 내용의 설명 및 이해를 위해 제시된 것이며, 본 개시의 범위를 한정하는 것은 아니다. 따라서, 본 개시의 범위는, 본 개시의 기술적 사상에 근거한 모든 변경 또는 다양한 다른 실시예를 포함하는 것으로 해석되어야 한다.

도면

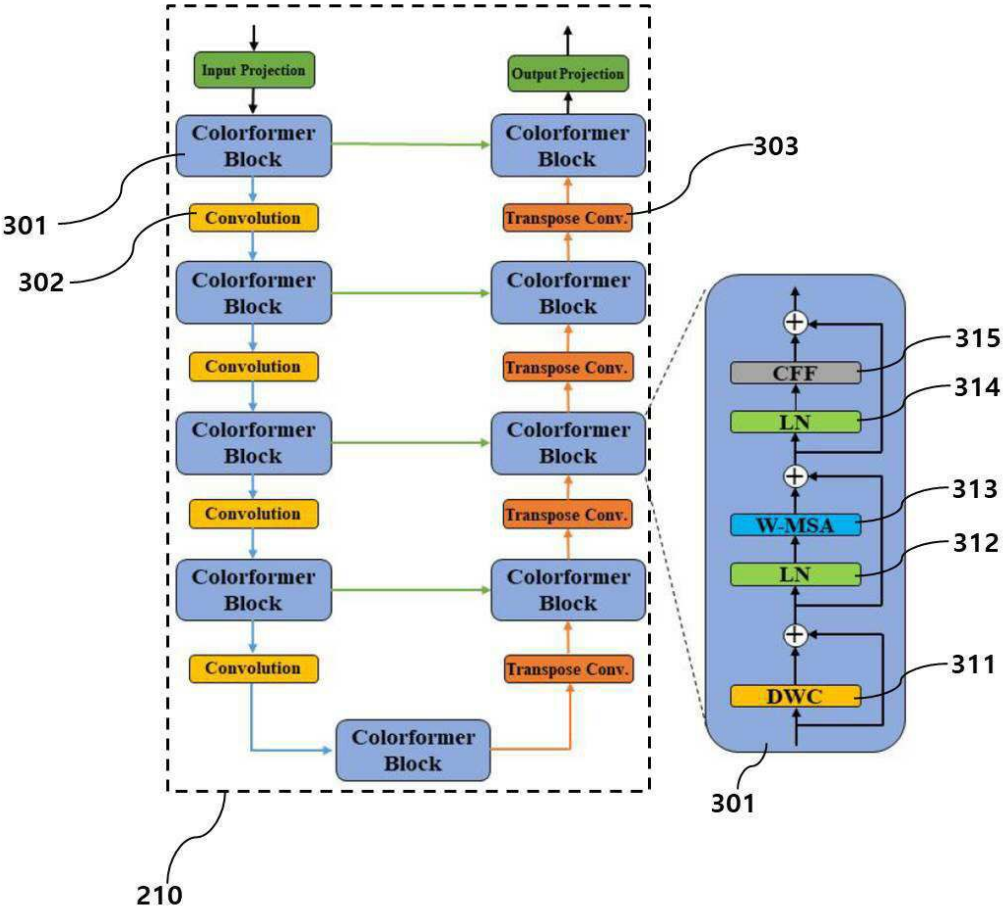
도면1


101
전자 장치
100
120
프로세서
130
메모리
어플리케이션
147
어플리케이션 프로그래밍 인터페이스 (API)
145
140
미들웨어
143
커널
141
110
버스
104
입출력 인터페이스
150
디스플레이
160
통신 인터페이스
170
네트워크
162
전자 장치
서버
106
164
전자 장치
102

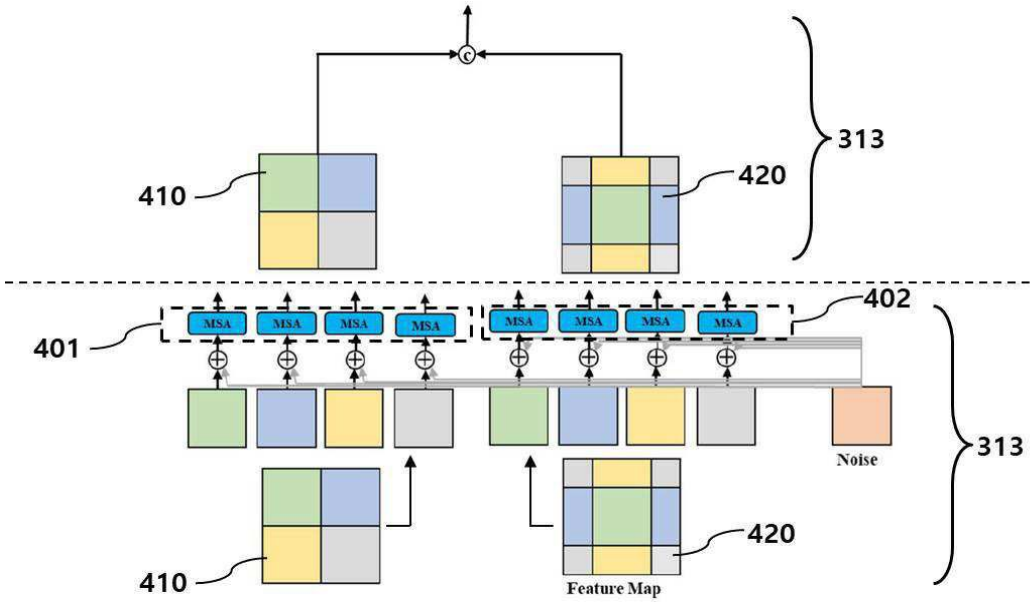
도면2


212
L1 Loss + VGG perceptual Loss + WGAN LOSS
Chrominance Channel
Generator
Critic
Score
RGB Image
201
Luminance Channel
211
213
220
Discriminator Loss
210
200

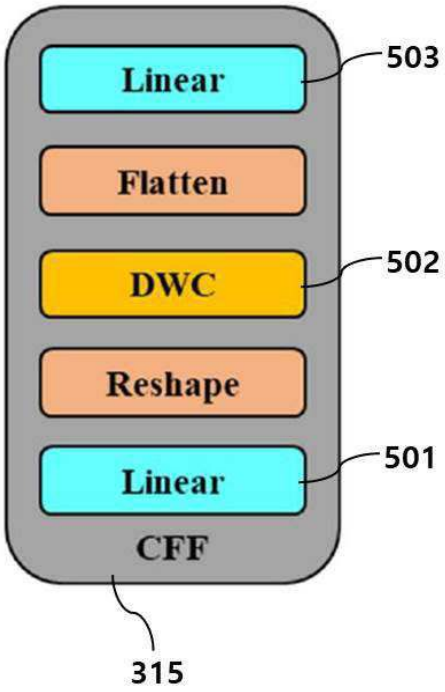
도면3


Input Projection
Output Projection
Colorformer Block
Convolution
Transpose Conv.
301
302
303
210
CFF
LN
W-MSA
DWC
311
312
313
314
315

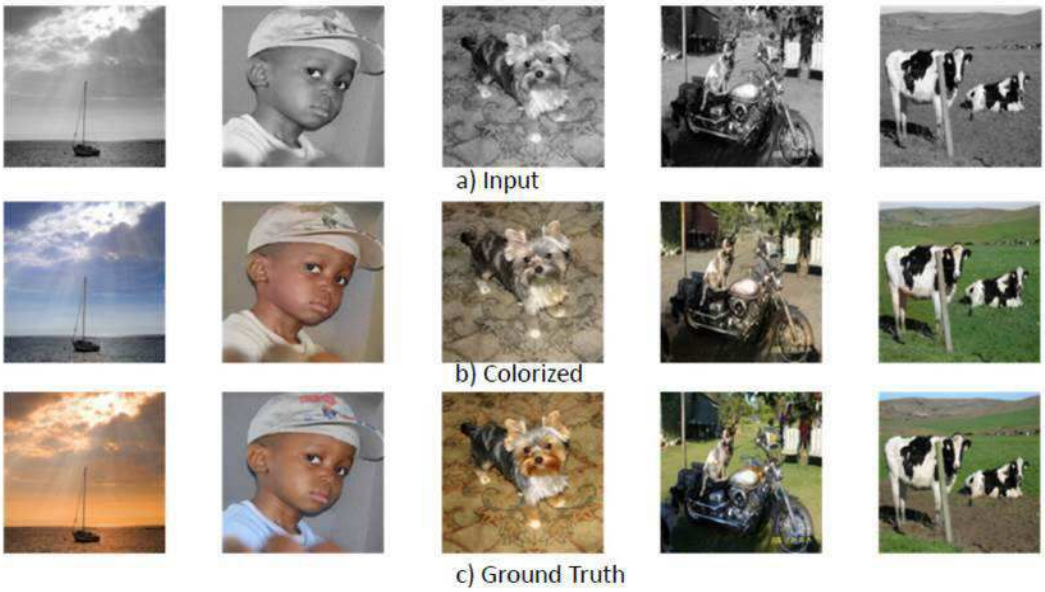
도면4


410
420
313
401
402
MSA
Noise
Feature Map

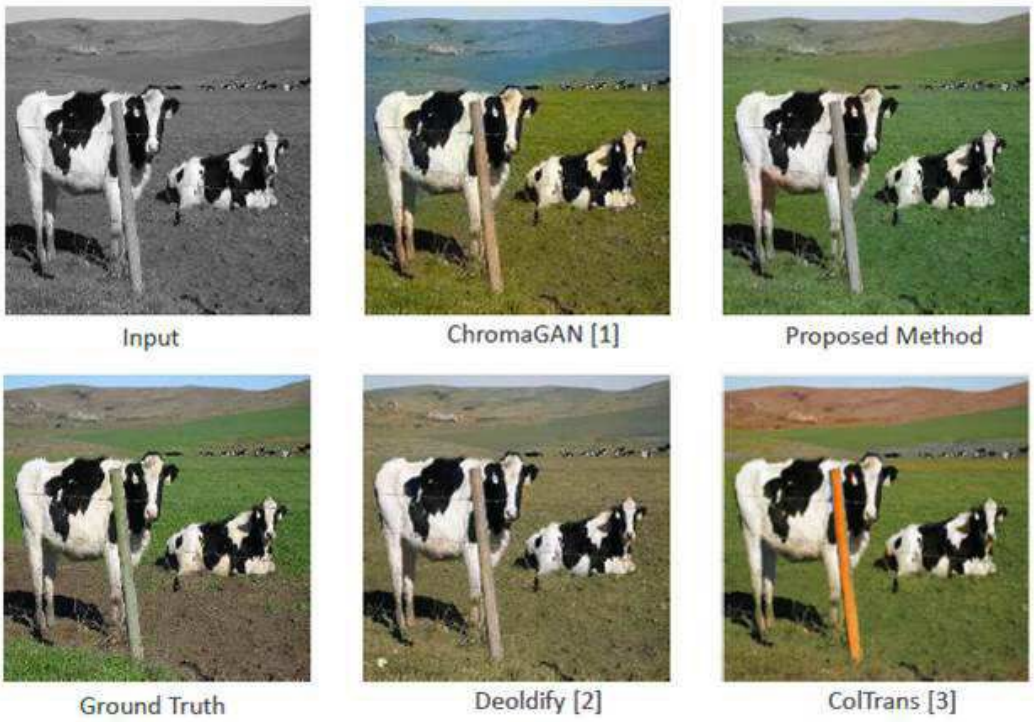
도면5


Linear
503
Flatten
DWC
502
Reshape
Linear
501
CFF
315

도면6


a) Input
b) Colorized
c) Ground Truth

도면7a



도면7b

Methods	PSNR [dB]	SSIM
Input	27.8	0.06
ChromaGAN [1]	29.0	0.77
Deoldify [2]	29.2	0.75
ColTran [3]	29.2	0.810
Proposed Network	29.59	0.819