(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0057770
(43) 공개일자 2020년05월26일

(51) 국제특허분류(Int. Cl.)
G06N 3/08 (2006.01) *G06K 9/00* (2006.01)
G06K 9/46 (2006.01) *G06N 3/04* (2006.01)
G06T 11/00 (2006.01) *G06T 7/11* (2017.01)
(52) CPC특허분류
G06N 3/08 (2013.01)
G06K 9/00248 (2013.01)
(21) 출원번호 10-2020-7012540
(22) 출원일자(국제) 2018년10월24일
심사청구일자 2020년04월29일
(85) 번역문제출일자 2020년04월29일
(86) 국제출원번호 PCT/CA2018/051345
(87) 국제공개번호 WO 2019/079895
국제공개일자 2019년05월02일
(30) 우선권주장
62/576,180 2017년10월24일 미국(US)
62/597,494 2017년12월12일 미국(US)

(71) 출원인
로레알
프랑스공화국, 파리 F-75008, 뤼 르와이얄 14
(72) 발명자
레빈쉬타인 알렉스
캐나다 엘4제이 8엑스7 온타리오주 쏜힐 카버넷 로드 67
창 쳉
캐나다 엠5비 2에이치9 온타리오주 토론토 캘턴 스트리트 1412
(뒷면에 계속)
(74) 대리인
양영준, 백만기

전체 청구항 수 : 총 82 항

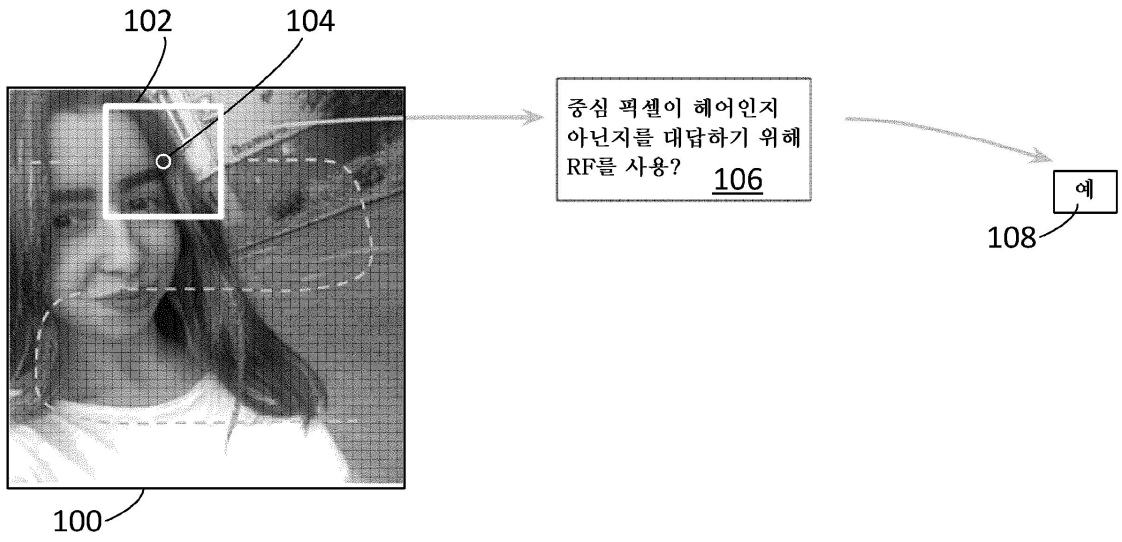
(54) 발명의 명칭 **심층 신경망을 사용하는 이미지 처리를 위한 시스템 및 방법**

(57) 요 약

시스템 및 방법은, 예를 들어, 색 모발에 대한 비디오의 실시간 처리를 위해 컨볼루션 신경망(CNN)을 제공하기 위해 모바일 디바이스 상에서 딥 러닝을 구현한다. 이미지들은 CNN을 이용하여 처리되어 모발 픽셀들의 제각기 모발 매트를 정의한다. 제각기 객체 매트들은 색, 조명, 텍스처 등을 변경하기 위해 그런 것처럼 픽셀 값들을

(뒷면에 계속)

대 표 도



조정할 때 어느 픽셀들을 조정할지를 결정하기 위해 사용될 수 있다. CNN은 세그먼테이션 마스크를 생성하도록 적응된 이미지 분류를 위한 (사전 훈련된) 네트워크를 포함할 수 있다. CNN은 마스크 이미지 그래디언트 일관성 손실을 최소화하기 위해 (예를 들어, 거친 세그먼테이션 데이터를 이용하여) 이미지 세그먼테이션을 위해 훈련될 수 있다. CNN은 인코더 스테이지와 디코더 스테이지의 대응하는 계층들 사이의 스킵 연결들을 추가로 사용할 수 있는데, 여기서 고 해상도이지만 약한 피처들을 포함하는 인코더에서의 더 얕은 계층들은 더 깊은 디코더 계층들로부터의 저 해상도이지만 강한 피처들과 조합된다.

(52) CPC특허분류

G06K 9/00281 (2013.01)

G06K 9/4628 (2013.01)

G06N 3/0454 (2013.01)

G06T 11/001 (2013.01)

G06T 7/11 (2017.01)

(72) 발명자

펑 에드먼드

캐나다 엠6이 3엔2 온타리오주 토론토 웨스트마운트 애버뉴 352

케젤 이리나

캐나다 엠2엔 7에이치5 온타리오주 토론토 클레어트렐 로드 202-2

구오 웬장지

캐나다 엘5씨 3에이5 온타리오주 미시소거 스테인톤 드라이브 854

엘모즈니노 에릭

캐나다 엠5티 1에스6 온타리오주 토론토 칼리지 스트리트 301-356

지앙 루오웨이

캐나다 엠4더블유 3와이1 온타리오주 토론토 블로어 스트리트 이스트 805-85

아아라비 파르함

캐나다 엘4비 2씨7 온타리오주 리치몬드 힐 스트래스언 애버뉴 7

명 세 서

청구범위

청구항 1

이미지를 처리하는 컴퓨팅 디바이스로서:

상기 이미지의 픽셀들을 분류하여 상기 픽셀들 각각이 객체 픽셀인지 객체 픽셀이 아닌지를 결정하여 상기 이미지에서의 객체에 대한 객체 세그먼테이션 마스크를 정의하도록 구성된 컨볼루션 신경망(CNN)을 저장하고 제공하는 저장 유닛 - 상기 CNN은 상기 객체 세그먼테이션 마스크를 정의하도록 적응되는 이미지 분류를 위한 사전 훈련된 네트워크를 포함하고, 상기 CNN은 세그먼테이션 데이터를 사용하여 추가로 훈련됨 -; 및

상기 저장 유닛에 결합되어 상기 CNN을 이용하여 상기 이미지를 처리하여 상기 객체 세그먼테이션 마스크를 생성함으로써 새로운 이미지를 정의하도록 구성된 처리 유닛을 포함하는 컴퓨팅 디바이스.

청구항 2

제1항에 있어서,

상기 CNN은 마스크 이미지 그래디언트 일관성 손실을 최소화하도록 추가로 훈련되는 컴퓨팅 디바이스.

청구항 3

제1항 또는 제2항에 있어서,

상기 마스크 이미지 그래디언트 일관성 손실 L_c 는 다음과 같이 정의되고:

$$L_c = \frac{\sum M_{mag}\left[1-\left(I_x M_x + I_y M_y\right)^2\right]}{\sum M_{mag}},$$

여기서, I_x, I_y는 정규화된 이미지 그래디언트이고, M_x, M_y는 정규화된 마스크 그래디언트이고, M_{mag}은 마스크 그래디언트 크기인 컴퓨팅 디바이스.

청구항 4

제3항에 있어서,

상기 마스크 이미지 그래디언트 일관성 손실 L_c 는, 훈련할 때 전체 손실 L을 최소화하기 위해, 가중치 w를 가지고 바이너리 크로스 엔트로피 손실 L_M과 조합될 수 있고, L은 $L = L_M + wL_C$ 로서 정의되는 컴퓨팅 디바이스.

청구항 5

제1항 내지 제4항 중 어느 한 항에 있어서,

상기 세그먼테이션 데이터는 잡음이 있고 거친 세그먼테이션 데이터인 컴퓨팅 디바이스.

청구항 6

제1항 내지 제5항 중 어느 한 항에 있어서,

상기 세그먼테이션 데이터는 크라우드 소싱된 세그먼테이션 데이터인 컴퓨팅 디바이스.

청구항 7

제1항 내지 제4항 중 어느 한 항에 있어서,

인코더 스테이지에서의 계층들 및 디코더 스테이지에서의 대응하는 계층들 사이의 스킵 연결들을 사용하여 상기 디코더 스테이지에서 업샘플링할 때 저해상도이지만 강한 피처들 및 고 해상도이지만 약한 피처들을 조합하여 상기 객체 세그먼테이션 마스크를 정의하도록 적응되는 컴퓨팅 디바이스.

청구항 8

제1항 내지 제7항 중 어느 한 항에 있어서,

상기 객체는 모발인 컴퓨팅 디바이스.

청구항 9

제1항 내지 제8항 중 어느 한 항에 있어서,

상기 처리 유닛은 상기 객체 세그먼테이션 마스크를 사용하여 선택된 상기 이미지에서의 픽셀들에 대한 변경을 적용함으로써 상기 이미지로부터 상기 새로운 이미지를 정의하도록 구성된 컴퓨팅 디바이스.

청구항 10

제9항에 있어서,

상기 변경은 색의 변경인 컴퓨팅 디바이스.

청구항 11

제9항 또는 제10항에 있어서,

상기 CNN은 실시간으로 상기 새로운 이미지를 비디오 이미지로서 생성하기 위해 제한된 계산 능력을 갖는 모바일 디바이스 상에서 실행되도록 구성된 훈련된 네트워크를 포함하는 컴퓨팅 디바이스.

청구항 12

제9항 내지 제11항 중 어느 한 항에 있어서,

상기 처리 유닛은 상기 이미지 및 상기 새로운 이미지를 디스플레이하기 위해 대화형 그래픽 사용자 인터페이스(GUI)를 제공하도록 구성되고, 상기 대화형 GUI는 상기 새로운 이미지를 정의하기 위해 상기 변경을 결정하기 위한 입력을 수신하도록 구성된 컴퓨팅 디바이스.

청구항 13

제12항에 있어서,

상기 처리 유닛은 상기 객체 세그먼테이션 마스크를 이용하여 상기 이미지에서의 상기 객체의 픽셀들을 분석하여 상기 변경을 위한 하나 이상의 후보를 결정하도록 구성되고, 상기 대화형 GUI는 상기 변경으로서 적용할 상기 후보들 중 하나를 선택하기 위한 입력을 수신하기 위해 상기 하나 이상의 후보를 제시하도록 구성된 컴퓨팅 디바이스.

청구항 14

제1항 내지 제13항 중 어느 한 항에 있어서,

상기 이미지는 셀피 비디오인 컴퓨팅 디바이스.

청구항 15

이미지를 처리하는 컴퓨팅 디바이스로서:

상기 이미지의 픽셀들을 분류하여 상기 픽셀들 각각이 모발 픽셀인지 모발 픽셀이 아닌지를 결정하도록 구성된 컨볼루션 신경망(CNN)을 저장하고 제공하는 저장 유닛 - 상기 CNN은 세그먼테이션 데이터로 훈련될 때 마스크

이미지 그래디언트 일관성 손실을 최소화하여 모발 세그먼테이션 마스크를 정의하도록 훈련됨 -; 및

상기 저장 유닛에 결합된 처리 유닛 - 상기 처리 유닛은 상기 모발 세그먼테이션 마스크를 이용하여 상기 이미지에서의 모발 픽셀들에 새로운 모발 색을 적용함으로써 채색된 모발 이미지를 정의하고 제시하도록 구성됨 - 을 포함하는 컴퓨팅 디바이스.

청구항 16

제15항에 있어서,

상기 CNN은 실시간으로 상기 채색된 모발 이미지를 비디오 이미지로서 생성하기 위해 제한된 계산 능력을 갖는 모바일 디바이스 상에서 실행되도록 구성된 훈련된 네트워크를 포함하는 컴퓨팅 디바이스.

청구항 17

제15항 또는 제16항에 있어서,

상기 CNN은 상기 모발 세그먼테이션 마스크를 생성하도록 적응되는 이미지 분류를 위한 사전 훈련된 네트워크를 포함하고, 상기 사전 훈련된 네트워크는 인코더 스테이지에서의 계층들과 디코더 스테이지에서의 대응하는 계층들 사이의 스킵 연결들을 사용하여 상기 디코더 스테이지에서 업샘플링할 때 저 해상도이지만 강한 피처들과 고 해상도이지만 약한 피처들을 조합하도록 적응된 컴퓨팅 디바이스.

청구항 18

제17항에 있어서,

상기 인코더 스테이지로부터의 더 낮은 해상도의 계층은 상기 디코더 스테이지의 들어오는 더 높은 해상도의 계층과 양립가능한 출력 깊이를 만들기 위해 포인트와이즈 컨볼루션을 먼저 적용함으로써 처리되는 컴퓨팅 디바이스.

청구항 19

제15항 내지 제18항 중 어느 한 항에 있어서,

상기 CNN은 디코더 스테이지에서 업샘플링할 때 대응하는 계층들로부터의 저 해상도이지만 강한 피처들과 고 해상도이지만 약한 피처들을 조합하도록 적응된 컴퓨팅 디바이스.

청구항 20

제15항 내지 제19항 중 어느 한 항에 있어서,

상기 CNN은 상기 마스크 이미지 그래디언트 일관성 손실을 최소화하기 위해 거친 세그먼테이션 데이터를 이용하여 훈련되는 컴퓨팅 디바이스.

청구항 21

제15항 내지 제20항 중 어느 한 항에 있어서,

상기 마스크 이미지 그래디언트 일관성 손실 L_c는 다음과 같이 정의되고:

$$L_c = \frac{\sum M_{mag}\left[1-\left(I_x M_x + I_y M_y\right)^2\right]}{\sum M_{mag}},$$

여기서, I_x,I_y는 정규화된 이미지 그래디언트이고, M_x,M_y는 정규화된 마스크 그래디언트이고, M_{mag}은 마스크 그래디언트 크기인 컴퓨팅 디바이스.

청구항 22

제20항 또는 제21항에 있어서,

상기 마스크 이미지 그래디언트 일관성 손실 L_C는, 훈련할 때 전체 손실 L을 최소화하기 위해, 가중치 w를 가지고 바이너리 크로스 엔트로피 손실 L_M과 조합되고, L은 $L = L_M + wL_C$로서 정의되는 컴퓨팅 디바이스.

청구항 23

제1항 내지 제22항 중 어느 한 항에 있어서,

상기 CNN은 처리 동작들을 최소화하고 제한된 계산 능력을 갖는 모바일 디바이스 상에서 상기 CNN의 실행을 가능하게 하기 위해 뎁스와이즈 컨볼루션(depthwise convolution) 및 포인트와이즈 컨볼루션(pointwise convolution)을 포함하는 뎁스와이즈 분리 가능 컨볼루션(depthwise separable convolution)을 이용하도록 구성된 컴퓨팅 디바이스.

청구항 24

이미지를 처리하는 컴퓨팅 디바이스로서:

처리 유닛, 상기 처리 유닛에 결합된 저장 유닛, 및 상기 처리 유닛 및 상기 저장 유닛 중 적어도 하나에 결합된 입력 디바이스를 포함하고, 상기 저장 유닛은 명령어들을 저장하고 상기 명령어들은 상기 처리 유닛에 의해 실행될 때 상기 컴퓨팅 디바이스가:

상기 입력 디바이스를 통해 상기 이미지를 수신하고;

상기 이미지에서의 모발 픽셀들을 식별하는 모발 세그먼테이션 마스크를 정의하고;

상기 모발 세그먼테이션 마스크를 이용하여 상기 이미지에서의 모발 픽셀들에 새로운 모발 색을 적용함으로써 채색된 모발 이미지를 정의하고;

상기 채색된 모발 이미지를 디스플레이 디바이스에 출력하기 위해 제공하도록 구성하고;

상기 모발 세그먼테이션 마스크는:

상기 저장 유닛에 의해 저장된 컨볼루션 신경망(CNN)을 이용하여 상기 이미지를 처리하여 컨볼루션 계층들의 연속으로 복수의 컨볼루션(conv) 필터를 적용하여 제각기 피처들을 검출함으로써 정의되고, 연속하는 제1 세트의 컨볼루션 계층들은 상기 이미지를 제1 이미지 해상도로부터 최소 해상도까지 다운샘플링하는 출력을 제공하고, 연속하는 제2 세트의 컨볼루션 계층들은 상기 출력을 상기 제1 이미지 해상도로 되돌려 업샘플링하고, 상기 CNN은 마스크 이미지 그래디언트 일관성 손실을 최소화하여 상기 모발 세그먼테이션 마스크를 정의하도록 훈련되는 컴퓨팅 디바이스.

청구항 25

제24항에 있어서,

상기 CNN은 상기 제1 세트 및 상기 제2 세트로부터 대응하는 컨볼루션 계층들 사이의 스킵 연결들을 사용하도록 적응되는 컴퓨팅 디바이스.

청구항 26

제24항 또는 제25항에 있어서,

상기 CNN은 상기 제2 세트의 컨볼루션 계층들의 초기 계층 이전에 그리고 상기 제2 세트의 컨볼루션 계층들의 제각기 후속하는 계층들 이전에 산재된 업샘플링 기능을 가져서 출력을 상기 제1 이미지 해상도로 업샘플링하는 컴퓨팅 디바이스.

청구항 27

제26항에 있어서,

상기 업샘플링 기능은 제각기 스킵 연결들을 사용하고, 상기 제각기 스킵 연결들 각각은:

연속하는 인접한 컨볼루션 계층으로부터 상기 제2 세트의 컨볼루션 계층들의 다음 컨볼루션 계층으로의 입력으로서 출력되는 제1 활성화 맵; 및

상기 제1 세트의 컨볼루션 계층들에서의 더 이른 컨볼루션 계층으로부터 출력되는 제2 활성화 맵 - 상기 제2 활성화 맵은 상기 제1 활성화 맵보다 더 큰 이미지 해상도를 가짐 - 을 조합하는 컴퓨팅 디바이스.

청구항 28

제27항에 있어서,

상기 제각기 스킵 연결들 각각은, 상기 제2 활성화 맵에 적용되는 컨볼루션 1 x 1 필터의 출력을 상기 제1 활성화 맵에 적용되는 업샘플링 기능의 출력에 더하여 상기 제1 활성화 맵의 해상도를 상기 더 큰 이미지 해상도로 증가시키도록 정의되는 컴퓨팅 디바이스.

청구항 29

제24항 내지 제28항 중 어느 한 항에 있어서,

상기 명령어들은 상기 디스플레이 디바이스를 통해 그래픽 사용자 인터페이스(GUI)를 제시하도록 상기 컴퓨팅 디바이스를 구성하고, 상기 GUI는 상기 이미지를 보기 위한 제1 부분 및 상기 채색된 모발 이미지를 동시에 보기 위한 제2 부분을 포함하는 컴퓨팅 디바이스.

청구항 30

제29항에 있어서,

상기 명령어들은 상기 컴퓨팅 디바이스가 상이한 조명 조건에서 상기 새로운 모발 색을 보여주기 위해 상기 채색된 모발 이미지에서의 상기 새로운 모발 색에 조명 조건 처리를 적용하도록 구성하는 컴퓨팅 디바이스.

청구항 31

제24항 내지 제30항 중 어느 한 항에 있어서,

상기 새로운 모발 색은 제1 새로운 모발 색이고, 상기 채색된 모발 이미지는 제1 채색된 모발 이미지이고, 상기 명령어들은 상기 컴퓨팅 디바이스가:

상기 모발 세그먼테이션 마스크를 이용하여 상기 이미지에서의 모발 픽셀들에 제2 새로운 모발 색을 적용함으로써 그리고 상기 디스플레이 디바이스로의 출력을 위해 상기 제2 채색된 모발 이미지를 제공함으로써 제2 채색된 모발 이미지를 정의하고 제시하도록 구성하는 컴퓨팅 디바이스.

청구항 32

제31항에 있어서,

상기 명령어들은 상기 컴퓨팅 디바이스가 상기 디스플레이 디바이스를 통해 2개의 새로운 색 GUI를 제시하도록 구성하고, 상기 2개의 새로운 색 GUI는 상기 제1 채색된 모발 이미지를 보기 위한 제1 새로운 색 부분 및 상기 제2 채색된 모발 이미지를 동시에 보기 위한 제2 새로운 색 부분을 포함하는 컴퓨팅 디바이스.

청구항 33

제24항 내지 제32항 중 어느 한 항에 있어서,

상기 명령어들은 상기 컴퓨팅 디바이스가:

상기 이미지에서의 모발 픽셀들의 현재 모발 색을 포함하는 색에 대해 상기 이미지를 분석하고;

하나 이상의 제안된 새로운 모발 색을 결정하고;

상기 채색된 모발 이미지를 정의하기 위해 상기 새로운 모발 색을 선택하도록 상기 GUI의 대화형 부분을 통해 상기 제안된 새로운 모발 색을 제시하도록 구성하는 컴퓨팅 디바이스.

청구항 34

제24항 내지 제33항 중 어느 한 항에 있어서,

상기 처리 유닛은 상기 CNN을 실행하기 위한 그래픽 처리 유닛(GPU)을 포함하고 상기 컴퓨팅 디바이스는 스마트폰 및 태블릿 중 하나를 포함하는 컴퓨팅 디바이스.

청구항 35

제24항 내지 제34항 중 어느 한 항에 있어서,

상기 이미지는 비디오의 복수의 비디오 이미지 중 하나이고, 상기 명령어들은 상기 컴퓨팅 디바이스가 상기 복수의 비디오 이미지를 처리하여 모발 색을 변경하도록 구성하는 컴퓨팅 디바이스.

청구항 36

제24항 내지 제35항 중 어느 한 항에 있어서,

상기 CNN은 상기 이미지의 픽셀들을 분류하여 상기 픽셀들 각각이 모발 픽셀인지를 결정하도록 적응되는 MobileNet 기반 모델을 포함하는 컴퓨팅 디바이스.

청구항 37

제24항 내지 제36항 중 어느 한 항에 있어서,

상기 CNN은 개별 표준 컨볼루션들이 뎁스와이즈 컨볼루션(depthwise convolution) 및 포인트와이즈 컨볼루션(pointwise convolution)으로 인수분해되는 컨볼루션들을 포함하는 뎁스와이즈 분리가능 컨볼루션 신경망으로서 구성되고, 상기 뎁스와이즈 컨볼루션은 각각의 입력 채널에 단일 필터를 적용하는 것으로 제한되고, 상기 포인트와이즈 컨볼루션은 상기 뎁스와이즈 컨볼루션의 출력들을 조합하는 것으로 제한되는 컴퓨팅 디바이스.

청구항 38

제24항 내지 제37항 중 어느 한 항에 있어서,

상기 마스크 이미지 그래디언트 일관성 손실 L_c 은 다음과 같이 정의되고:

$$L_c = \frac{\sum M_{mag}\left[1-\left(I_x M_x + I_y M_y\right)^2\right]}{\sum M_{mag}}$$

여기서, I_x,I_y는 정규화된 이미지 그래디언트이고, M_x,M_y는 정규화된 마스크 그래디언트이고, M_{mag}은 마스크 그래디언트 크기인 컴퓨팅 디바이스.

청구항 39

이미지들을 처리하여 객체 세그먼테이션 마스크를 정의하도록 훈련된 CNN을 생성하도록 구성된 컴퓨팅 디바이스로서:

이미지 분류를 위해 구성되고 또한 제한된 계산 능력을 갖는 런타임 컴퓨팅 디바이스 상에서 실행되도록 구성된 CNN을 수신하는 저장 유닛;

입력 및 디스플레이 출력을 수신하기 위해 대화형 인터페이스를 제공하도록 구성된 처리 유닛을 포함하고, 상기 처리 유닛은:

객체 세그먼테이션 마스크를 정의하기 위해 상기 CNN을 적응시키고 적응된 대로의 상기 CNN을 상기 저장 유닛에 저장하기 위해 입력을 수신하고;

객체 세그먼테이션을 위해 라벨링된 세그먼테이션 훈련 데이터를 사용하여 적응된 대로의 상기 CNN을 훈련하여 상기 CNN을 생성하여 이미지들을 처리하여 상기 객체 세그먼테이션 마스크를 정의하기 위해 입력을 수신하도록 구성된 컴퓨팅 디바이스.

청구항 40

제39항에 있어서,

상기 처리 유닛은:

훈련할 때 마스크 이미지 그래디언트 일관성 손실을 최소화하도록 정의된 상기 마스크 이미지 그래디언트 일관성 손실 함수 최소화를 사용하고;

인코더 스테이지에서의 계층들과 디코더 스테이지에서의 대응하는 계층들 사이의 스킵 연결들을 사용하여 상기 디코더 스테이지에서 업샘플링할 때 저 해상도이지만 강한 피처들과 고 해상도이지만 약한 피처들을 조합하고;

적응된 대로의 상기 CNN을 상기 저장 유닛에 저장하는 것 중 적어도 하나를 하도록 상기 CNN을 적응시키기 위해 입력을 수신하도록 구성된 컴퓨팅 디바이스.

청구항 41

이미지들을 처리하여 객체 세그먼테이션 마스크를 정의하도록 훈련된 CNN을 생성하는 방법으로서:

이미지 분류를 위해 구성되고 또한 제한된 계산 능력을 갖는 컴퓨팅 디바이스 상에서 실행되도록 구성되는 CNN을 획득하는 단계;

상기 CNN을 적응시켜 객체 세그먼테이션 마스크를 정의하는 단계; 및

객체 세그먼테이션을 위해 라벨링된 세그먼테이션 훈련 데이터를 사용하여 적응된 대로의 상기 CNN을 훈련하여 상기 객체 세그먼테이션 마스크를 정의하는 단계를 포함하는 방법.

청구항 42

제41항에 있어서,

상기 CNN은 이미지 분류를 위해 사전 훈련되고, 상기 훈련 단계는 사전 훈련된 대로의 상기 CNN을 추가로 훈련하는 방법.

청구항 43

제41항 또는 제42항에 있어서,

상기 CNN을 생성하여 이미지들을 처리하기 위해, 마스크 이미지 그래디언트 일관성 손실을 최소화하도록 정의된 마스크 이미지 그래디언트 일관성 손실 함수를 이용하여 상기 CNN을 적응시키는 단계를 포함하는 방법.

청구항 44

제41항 내지 제43항 중 어느 한 항에 있어서,

상기 세그먼테이션 훈련 데이터는 크라우드 소싱된 데이터로부터 정의된 잡음이 있고 거친 세그먼테이션 훈련 데이터인 방법.

청구항 45

제41항 내지 제44항 중 어느 한 항에 있어서,

디코더 스테이지에서 업샘플링할 때 저 해상도이지만 강한 피처들과 고 해상도이지만 약한 피처들을 조합하기 위해 인코더 스테이지에서의 계층들과 상기 디코더 스테이지에서의 대응하는 계층들 사이의 스킵 연결들을 사용하도록 상기 CNN을 적응시키는 단계를 포함하는 방법.

청구항 46

제41항 내지 제45항 중 어느 한 항에 있어서,

상기 마스크 이미지 그래디언트 일관성 손실 함수는 조합된 손실을 최소화하기 위해 바이너리 크로스 엔트로피 손실 함수와 조합되는 방법.

청구항 47

제41항 내지 제46항 중 어느 한 항에 있어서,

상기 마스크 이미지 그래디언트 일관성 손실 L_c 는 다음과 같이 정의되고:

$$L_c = \frac{\sum M_{mag}\left[1-\left(I_x M_x + I_y M_y\right)^2\right]}{\sum M_{mag}}$$

여기서, I_x,I_y는 정규화된 이미지 그래디언트이고, M_x,M_y는 정규화된 마스크 그래디언트이고, M_{mag}은 마스크 그래디언트 크기인 방법.

청구항 48

제41항 내지 제47항 중 어느 한 항에 있어서,

모바일 디바이스 상에 저장하기 위한 상기 객체 세그먼테이션 마스크를 정의하기 위해 상기 CNN을 제공하여 이미지들을 처리하는 단계를 포함하는 방법.

청구항 49

방법으로서:

컴퓨팅 디바이스의 저장 유닛에, 이미지의 픽셀들을 분류하여 상기 픽셀들 각각이 객체 픽셀이거나 객체 픽셀이 아닌지를 결정하여 상기 이미지에서의 객체에 대한 객체 세그먼테이션 마스크를 정의하도록 구성된 컨볼루션 신경망(CNN)을 저장하는 단계; 및

상기 컴퓨팅 디바이스의 처리 유닛을 통해 상기 CNN을 이용하여 이미지를 처리하여 상기 객체 세그먼테이션 마스크를 생성하여 새로운 이미지를 정의하는 단계를 포함하고;

상기 CNN은 상기 객체 세그먼테이션 마스크를 정의하도록 적응되고 또한 세그먼테이션 데이터로 훈련되는 이미지 분류를 위한 사전 훈련된 네트워크를 포함하는 방법.

청구항 50

제49항에 있어서,

상기 CNN은:

세그먼테이션 데이터를 사용하여 훈련될 때 마스크 이미지 그래디언트 일관성 손실을 최소화하기 위해 마스크 이미지 그래디언트 일관성 손실 함수를 사용하는 것; 및

인코더 스테이지에서의 계층들 및 디코더 스테이지에서의 대응하는 계층들 사이의 스킵 연결들을 사용하여 상기 디코더 스테이지에서 업샘플링할 때 저 해상도이지만 강한 피처들과 고 해상도이지만 약한 피처들을 조합하여 상기 객체 세그먼테이션 마스크를 정의하는 것 중 적어도 하나를 하도록 적응되는 방법.

청구항 51

제50항에 있어서,

상기 마스크 이미지 그래디언트 일관성 손실 L_c 는 다음과 같이 정의되고:

$$L_c = \frac{\sum M_{mag}\left[1-\left(I_x M_x + I_y M_y\right)^2\right]}{\sum M_{mag}},$$

여기서, I_x,I_y는 정규화된 이미지 그래디언트이고, M_x,M_y는 정규화된 마스크 그래디언트이고, M_{mag}은 마스크 그래

디언트 크기인 방법.

청구항 52

제51항에 있어서,

상기 마스크 이미지 그래디언트 일관성 손실은, 훈련할 때 전체 손실 L을 최소화하기 위해, 가중치 w를 가지고 바이너리 크로스 엔트로피 손실과 조합되고, L은 $L = L_M + wL_C$ 로서 정의되는 방법.

청구항 53

제49항 내지 제52항 중 어느 한 항에 있어서,

상기 객체는 모발인 방법.

청구항 54

제49항 내지 제53항 중 어느 한 항에 있어서,

상기 객체 세그먼테이션 마스크를 사용하여 선택된 이미지에서의 픽셀들에 대한 변경을 적용함으로써 상기 이미지로부터 상기 새로운 이미지를 상기 처리 유닛을 통해 정의하는 단계를 포함하는 방법.

청구항 55

제54항에 있어서,

상기 변경은 색의 변경인 방법.

청구항 56

제54항 또는 제55항에 있어서,

상기 CNN은 실시간으로 상기 새로운 이미지를 비디오 이미지로서 생성하기 위해 제한된 계산 능력을 갖는 모바일 디바이스 상에서 실행되도록 구성된 훈련된 네트워크를 포함하는 방법.

청구항 57

제54항 내지 제56항 중 어느 한 항에 있어서,

상기 처리 유닛은 상기 이미지 및 상기 새로운 이미지를 디스플레이하기 위해 대화형 그래픽 사용자 인터페이스(GUI)를 제공하도록 구성되고, 상기 대화형 GUI는 상기 새로운 이미지를 정의하기 위해 상기 변경을 결정하기 위한 입력을 수신하도록 구성된 방법.

청구항 58

제57항에 있어서,

상기 처리 유닛에 의해, 상기 객체 세그먼테이션 마스크를 이용하여 상기 이미지에서의 상기 객체의 픽셀들을 분석하여 상기 변경을 위한 하나 이상의 후보를 결정하는 단계를 포함하고, 상기 대화형 GUI는 상기 변경으로서 적용할 후보들 중 하나를 선택하기 위한 입력을 수신하기 위해 상기 하나 이상의 후보를 제시하도록 구성된 방법.

청구항 59

제49항 내지 제58항 중 어느 한 항에 있어서,

상기 이미지는 셀피 비디오인 방법.

청구항 60

이미지를 처리하는 방법으로서:

컴퓨팅 디바이스의 저장 유닛에, 상기 이미지의 픽셀들을 분류하여 상기 픽셀들 각각이 모발 픽셀인지 모발 픽

셀이 아닌지를 결정하도록 구성된 컨볼루션 신경망(CNN)을 저장하는 단계 - 상기 CNN은 모발 세그먼테이션 마스크를 정의하기 위해 세그먼테이션 데이터로 훈련될 때 마스크 이미지 그래디언트 일관성 손실을 최소화하도록 훈련됨 -; 및

상기 저장 유닛에 결합된 상기 컴퓨팅 디바이스의 처리 유닛에 의해 상기 모발 세그먼테이션 마스크를 이용하여 상기 이미지에서의 모발 픽셀들에 새로운 모발 색을 적용함으로써 채색된 모발 이미지를 정의하고 제시하는 단계를 포함하는 방법.

청구항 61

제60항에 있어서,

상기 CNN은 실시간으로 상기 채색된 모발 이미지를 비디오 이미지로서 생성하기 위해 제한된 계산 능력을 갖는 모바일 디바이스 상에서 실행되도록 구성된 훈련된 네트워크를 포함하는 방법.

청구항 62

제60항 또는 제61항에 있어서,

상기 CNN은 상기 모발 세그먼테이션 마스크를 생성하도록 적응되는 이미지 분류를 위한 사전 훈련된 네트워크를 포함하고, 상기 사전 훈련된 네트워크는 상기 디코더 스테이지에서 업샘플링할 때 저 해상도이지만 강한 피처들과 고 해상도이지만 약한 피처를 조합하기 위해 인코더 스테이지에서의 계층들과 디코더 스테이지에서의 대응하는 계층들 사이의 스킵 연결들을 사용하도록 적응되는 방법.

청구항 63

제62항에 있어서,

상기 인코더 스테이지로부터의 더 낮은 해상도의 계층은 상기 디코더 스테이지의 들어오는 더 높은 해상도의 계층과 양립가능한 출력 깊이를 만들기 위해 포인트와이즈 컨볼루션을 먼저 적용함으로써 처리되는 방법.

청구항 64

제60항 내지 제63항 중 어느 한 항에 있어서,

상기 CNN은 거친 세그먼테이션 데이터를 이용하여 훈련되어 상기 마스크 이미지 그래디언트 일관성 손실을 최소화하는 방법.

청구항 65

제60항 내지 제64항 중 어느 한 항에 있어서,

상기 마스크 이미지 그래디언트 일관성 손실 L_c 는 다음과 같이 정의되고:

$$L_c = \frac{\sum M_{mag}\left[1-\left(I_x M_x + I_y M_y\right)^2\right]}{\sum M_{mag}},$$

여기서, I_x, I_y는 정규화된 이미지 그래디언트이고, M_x, M_y는 정규화된 마스크 그래디언트이고, M_{mag}은 마스크 그래디언트 크기인 방법.

청구항 66

제64항 또는 제65항에 있어서,

상기 마스크 이미지 그래디언트 일관성 손실 L_c 는, 훈련할 때 전체 손실 L을 최소화하기 위해, 가중치 w를 가

지고 바이너리 크로스 엔트로피 손실 L_M과 조합되고, L은 $L = L_M + wL_C$로서 정의되는 방법.

청구항 67

제41항 내지 제66항 중 어느 한 항에 있어서,

상기 CNN은 처리 동작들을 최소화하고 또한 제한된 계산 능력을 갖는 모바일 디바이스 상에서 상기 CNN의 실행을 가능하게 하기 위해 뎁스와이즈 컨볼루션(depthwise convolution) 및 포인트와이즈 컨볼루션(pointwise convolution)을 포함하는 뎁스와이즈 분리 가능 컨볼루션(depthwise separable convolution)을 이용하도록 구성되는 방법.

청구항 68

이미지를 처리하는 방법으로서:

처리 유닛을 통해 상기 이미지를 수신하는 단계;

상기 처리 유닛을 통해, 상기 이미지에서의 모발 픽셀들을 식별하는 모발 세그먼테이션 마스크를 정의하는 단계;

상기 처리 유닛을 통해 상기 모발 세그먼테이션 마스크를 이용하여 상기 이미지에서의 모발 픽셀들에 새로운 모발 색을 적용함으로써 채색된 모발 이미지를 정의하는 단계; 및

상기 처리 유닛을 통해 디스플레이 디바이스로의 출력을 위해 상기 채색된 모발 이미지를 제공하는 단계를 포함하고;

상기 모발 세그먼테이션 마스크는:

상기 처리 유닛에 결합된 저장 유닛에 의해 저장된 컨볼루션 신경망(CNN)을 이용하여 상기 이미지를 처리하여 제각기 피처들을 검출하도록 컨볼루션 계층들의 연속으로 복수의 컨볼루션(conv) 필터를 적용하고 - 상기 CNN은 잡음이 있고 거친 세그먼테이션 데이터로 훈련됨-, 및 모발 세그먼테이션 마스크를 정의하기 위해 마스크 이미지 그래디언트 일관성 손실을 최소화함으로써 정의되는 방법.

청구항 69

제68항에 있어서,

연속하는 제1 세트의 컨볼루션 계층들은 상기 이미지를 제1 이미지 해상도로부터 최소 해상도까지 다운샘플링한 출력을 제공하고, 연속하는 제2 세트의 컨볼루션 계층들은 상기 출력을 상기 제1 이미지 해상도로 되돌려 업샘플링하고, 상기 CNN은 상기 제1 세트 및 제2 세트로부터의 대응하는 컨볼루션 계층들 사이의 스킵 연결들을 이용하는 방법.

청구항 70

제69항에 있어서,

상기 CNN은 출력을 상기 제1 이미지 해상도로 업샘플링하기 위해 상기 제2 세트의 컨볼루션 계층들의 초기 계층 이전에 그리고 상기 제2 세트의 컨볼루션 계층들의 제각기 후속 계층들 이전에 산재된 업샘플링 함수를 포함하는 방법.

청구항 71

제70항에 있어서,

상기 업샘플링 함수는 제각기 스킵 연결들을 사용하고, 상기 제각기 스킵 연결들 각각은:

연속하는 인접한 컨볼루션 계층으로부터 상기 제2 세트의 컨볼루션 계층들의 다음 컨볼루션 계층으로의 입력으로서 출력되는 제1 활성화 맵; 및

상기 제1 세트의 컨볼루션 계층들에서의 더 이른 컨볼루션 계층으로부터 출력되는 제2 활성화 맵 - 상기 제2 활성화 맵은 상기 제1 활성화 맵보다 더 큰 이미지 해상도를 가짐 - 을 조합하는 방법.

청구항 72

제71항에 있어서,

상기 제각기 스킵 연결들 각각은, 상기 제2 활성화 맵에 적용되는 컨볼루션 1 x 1 필터의 출력을 상기 제1 활성화 맵에 적용되는 업샘플링 함수의 출력에 더하여 상기 제1 활성화 맵의 해상도를 더 큰 이미지 해상도로 증가시키도록 정의되는 방법.

청구항 73

제68항 내지 제72항 중 어느 한 항에 있어서,

상기 처리 유닛에 의해, 상기 디스플레이 디바이스를 통해 그래픽 사용자 인터페이스(GUI)를 제시하는 단계를 포함하고, 상기 GUI는 상기 이미지를 보기 위한 제1 부분 및 상기 채색된 모발 이미지를 동시에 보기 위한 제2 부분을 포함하는 방법.

청구항 74

제73항에 있어서,

상이한 조명 조건에서 상기 새로운 모발 색을 보여주기 위해 상기 채색된 모발 이미지에서 상기 새로운 모발 색에 대한 조명 조건 처리를 적용하는 단계를 포함하는 방법.

청구항 75

제68항 내지 제74항 중 어느 한 항에 있어서,

상기 새로운 모발 색은 제1 새로운 모발 색이고 상기 채색된 모발 이미지는 제1 채색된 모발 이미지이고, 상기 방법은:

상기 모발 세그먼테이션 마스크를 이용하여 상기 이미지에서의 모발 픽셀들에 제2 새로운 모발 색을 적용함으로써 그리고 상기 디스플레이 디바이스로의 출력을 위해 상기 제2 채색된 모발 이미지를 제공함으로써 상기 처리 유닛에 의해 제2 채색된 모발 이미지를 정의하고 제시하는 단계를 포함하는 방법.

청구항 76

제75항에 있어서,

상기 처리 유닛에 의해, 상기 디스플레이 디바이스를 통해 2개의 새로운 색 GUI를 제시하는 단계를 포함하고, 상기 2개의 새로운 색 GUI는 상기 제1 채색된 모발 이미지를 보기 위한 제1 새로운 채색된 부분 및 상기 제2 채색된 모발 이미지를 동시에 보기 위한 제2 새로운 채색된 부분을 포함하는 방법.

청구항 77

제68항 내지 제76항 중 어느 한 항에 있어서,

상기 처리 유닛에 의해, 상기 이미지에서의 모발 픽셀들의 현재 모발 색을 포함하는 색을 위한 상기 이미지를 분석하는 단계;

하나 이상의 제안된 새로운 모발 색을 결정하는 단계; 및

상기 채색된 모발 이미지를 정의하기 위해 상기 새로운 모발 색을 선택하도록 상기 GUI의 대화형 부분을 통해 상기 제안된 새로운 모발 색들을 제시하는 단계를 포함하는 방법.

청구항 78

제68항 내지 제77항 중 어느 한 항에 있어서,

상기 처리 유닛은 상기 CNN을 실행하기 위한 그래픽 처리 유닛(GPU)을 포함하고, 상기 처리 유닛 및 상기 저장 유닛은 스마트폰 및 태블릿 중 하나를 포함하는 컴퓨팅 디바이스에 의해 제공되는 방법.

청구항 79

제68항 내지 제78항 중 어느 한 항에 있어서,

상기 이미지는 비디오의 복수의 비디오 이미지 중 하나이고, 상기 방법은 모발 색을 변경하기 위해 상기 복수의 비디오 이미지를 상기 처리 유닛에 의해 처리하는 단계를 포함하는 방법.

청구항 80

제68항 내지 제79항 중 어느 한 항에 있어서,

상기 CNN은 상기 이미지의 픽셀들을 분류하여 상기 픽셀들 각각이 모발 픽셀인지를 결정하도록 적응된 MobileNet 기반 모델을 포함하는 방법.

청구항 81

제68항 내지 제80항 중 어느 한 항에 있어서,

상기 CNN은 개별 표준 컨볼루션들이 뎁스와이즈 컨볼루션(depthwise convolution) 및 포인트와이즈 컨볼루션(pointwise convolution)으로 인수분해되는 컨볼루션들을 포함하는 뎁스와이즈 분리가능 컨볼루션 신경망으로서 구성되고, 상기 뎁스와이즈 컨볼루션은 각각의 입력 채널에 단일 필터를 적용하는 것으로 제한되고, 상기 포인트와이즈 컨볼루션은 상기 뎁스와이즈 컨볼루션의 출력들을 조합하는 것으로 제한되는 방법.

청구항 82

제68항 내지 제81항 중 어느 한 항에 있어서,

상기 마스크 이미지 그래디언트 일관성 손실 L_c는 다음과 같이 정의되고:

$$L_c = \frac{\sum M_{mag}\left[1-\left(I_x M_x + I_y M_y\right)^2\right]}{\sum M_{mag}}$$

여기서, I_x, I_y는 정규화된 이미지 그래디언트이고, M_x, M_y는 정규화된 마스크 그래디언트이고, M_{mag}은 마스크 그래디언트 크기인 방법.

발명의 설명

기 술 분 야

[0001] [교차 참조]

[0002] 본 출원은, 다음과 같은 출원들에 대해, 미국에 대하여는 그의 국내적 이익을, 및 다른 관할권들에 대하여는 파리 조약 우선권을 주장한다: 1) 2017년 10월 24일자로 출원되고 발명의 명칭이 "A System and Method for Video Hair Coloration Using Deep Neural Networks"인 미국 가출원 번호 제62/576,180호; 및 2) 2017년 12월 12일자로 출원되고, 발명의 명칭이 "A System and Method for Real-time Deep Hair Matting on Mobile Devices"인 미국 가출원 번호 제62/597,494호; 각각의 출원의 전체 내용은 이러한 통합이 허용되는 임의의 관할권에 관하여 참조에 의해 본 명세서에 통합된다.

[0003] 본 출원은 소스 이미지로부터 새로운 이미지를 정의하기 위한 이미지 처리에 관한 것이고, 보다 구체적으로는 컨볼루션 신경망(convolutional neural network, CNN)과 같은 심층 신경망(deep neural network)을 사용하여 이미지를 처리하는 것에 관한 것이다.

배 경 기 술

[0004] 소스 이미지를 분석하여 그 안의 특정 주제를 식별하는데 이미지 처리가 유용한 많은 시나리오들이 있는데, 적어도 그러한 시나리오들의 서브세트에서 새로운 이미지를 생성하기 위해 정정들 또는 다른 변경들을 행하는 데 이미지 처리가 유용한 많은 시나리오들이 있다. 이미지 처리는 이미지에서 표현된 객체를 분류하고 및/또는 이미지에서 객체의 위치를 식별하기 위해 사용될 수 있다. 색, 텍스처, 조명, 밝기, 콘트라스트 및 다른 속성들

에 대한 변경들을 위해 그런 것처럼 이미지의 속성들(예를 들어, 픽셀의 제각기 값)을 정정하거나 변경하기 위해 이미지 처리가 사용될 수 있다. 다른 변경들은 이미지에 객체들을 추가하거나 또는 이미지로부터 객체들을 빼는 것, 객체들의 형상들을 변경하는 것 등을 포함할 수 있다.

[0005] 일 예에서, 이미지 처리는 소스 이미지에서의 대상(subject)의 모발(hair)을 채색하여 채색된 모발 이미지를 생성하는데 사용될 수 있다.

[0006] 이미지 처리는 종종 컴퓨팅 디바이스들, 특히, 스마트폰들, 태블릿들 등과 같은 흔한 모바일 디바이스들에 대해 리소스 집약적이다. 이는 일련의 이미지들을 포함하는 비디오를 실시간으로 처리할 때 특히 사실이다.

발명의 내용

과제의 해결 수단

[0007] 이하의 설명은 딥 러닝(deep learning)을 구현하는 것에 관한 것이고, 특히 모바일 디바이스 상에서 딥 러닝을 구현하는 것에 관한 것이다. 본 개시내용의 목적은, 라이브 비디오를 처리하기 위해, 예를 들어, 모발과 같은 객체를 세그먼팅하고 모발 색을 변경하기 위해 딥 러닝 환경을 제공하는 것이다. 본 기술분야의 통상의 기술자는 모발 이외의 객체들이 검출될 수 있고 색 또는 다른 제각기 속성들이 변경될 수 있다는 것을 이해할 것이다. 비디오 이미지들(예를 들어, 프레임들)은 각각의 비디오 이미지로부터 모발 픽셀들(예를 들어, 객체 픽셀들)의 제각기 모발 매트(matte)(예를 들어, 객체 마스크)를 정의하기 위해 심층 신경망을 사용하여 처리될 수 있다. 제각기 객체 매트들은 색, 조명, 텍스처 등과 같은 비디오 이미지의 속성을 조절할 때 어느 픽셀들을 조절할지를 결정하기 위해 사용될 수 있다. 모발 채색을 위한 일 예에서, 딥 러닝 신경망, 예를 들어, 컨볼루션 신경망은 소스 이미지의 픽셀들을 분류하여 각각이 모발 픽셀인지를 결정하고 모발 마스크를 정의하도록 구성된다. 그 다음, 마스크는 소스 이미지의 속성을 변경하여 새로운 이미지를 생성하기 위해 사용된다. CNN은 세그먼테이션 마스크를 생성하도록 적응된, 이미지 분류를 위한 사전 훈련된 네트워크를 포함할 수 있다. CNN은 거친 세그먼테이션 데이터(coarse segmentation data)를 이용하여, 그리고 훈련될 때 마스크 이미지 그래디언트 일관성 손실(mask-image gradient consistency loss)을 최소화하도록 추가로 훈련될 수 있다. CNN은 인코더 스테이지와 디코더 스테이지의 대응하는 계층들 사이의 스킵 연결(skip connection)들을 추가로 사용할 수 있는데, 여기서 고 해상도이지만 약한 피처들을 포함하는 인코더에서의 더 얕은 계층들은 디코더에서의 더 깊은 계층들로부터의 저 해상도이지만 강한 피처들과 조합된다.

[0008] 이러한 마스크는 직접적으로 또는 간접적으로 다른 객체들을 구별하기 위해 이용될 수 있다. 모발 마스크는 개인의 마진 또는 에지를 정의할 수 있고, 여기서 모발 마스크 외부의 주제는, 예를 들어, 배경일 수 있다. 검출될 수 있는 다른 객체들은 피부 등을 포함한다.

[0009] 이미지를 처리하는 컴퓨팅 디바이스가 제공되는데, 이 컴퓨팅 디바이스는: 이미지의 픽셀들을 분류하여 픽셀들 각각이 객체 픽셀인지 객체 픽셀이 아닌지를 결정하여 이미지에서의 객체에 대한 객체 세그먼테이션 마스크를 정의하도록 구성되는 컨볼루션 신경망(CNN)을 저장하고 제공하는 저장 유닛 - CNN은 객체 세그먼테이션 마스크를 정의하도록 적응된, 이미지 분류를 위한 사전 훈련된 네트워크를 포함하고, CNN은 세그먼테이션 데이터를 사용하여 추가로 훈련됨 -; 및 저장 유닛에 결합되어 CNN을 이용하여 이미지를 처리하여 객체 세그먼테이션 마스크를 생성하여 새로운 이미지를 정의하도록 구성된 처리 유닛을 포함한다.

[0010] CNN은 세그먼테이션 데이터를 사용하여 훈련될 때 마스크 이미지 그래디언트 일관성 손실을 최소화하도록 적응될 수 있다.

[0011] CNN은, 인코더 스테이지에서의 계층들 및 디코더 스테이지에서의 대응하는 계층들 사이의 스킵 연결들을 사용하여 디코더 스테이지에서 업샘플링할 때 저해상도이지만 강한 피처들 및 고 해상도이지만 약한 피처들을 조합하여 객체 세그먼테이션 마스크를 정의하도록 적응될 수 있다.

[0012] 마스크 이미지 그래디언트 일관성 손실 L_c는 다음과 같이 정의될 수 있다:

[0013] $$L_c = \frac{\sum M_{mag}\left[1-\left(I_x M_x + I_y M_y\right)^2\right]}{\sum M_{mag}},$$

[0014] 여기서, I_x,I_y는 정규화된 이미지 그래디언트이고, M_x,M_y는 정규화된 마스크 그래디언트이고, M_{mag}은 마스크 그래디언트 크기이다.

[0015] 마스크 이미지 그래디언트 일관성 손실은, 훈련할 때 전체 손실 L을 최소화하기 위해, 가중치 w를 가지고 바이너리 크로스 엔트로피 손실(binary cross entropy loss)과 조합될 수 있고, 여기서 L은 $L = L_M + wL_C$로서 정의된다.

[0016] CNN은 잡음이 있고 거친 세그먼테이션 데이터를 사용하여 훈련될 수 있다. 세그먼테이션 데이터는 크라우드 소싱된 세그먼테이션 데이터(crowd-sourced segmentation data)일 수 있다.

[0017] 객체는 모발일 수 있다. 처리 유닛은 객체 세그먼테이션 마스크를 사용하여 선택된 이미지에서의 픽셀들에 대한 변경을 적용함으로써 이미지로부터 새로운 이미지를 정의하도록 구성될 수 있다. 변경은 색의 변경일 수 있다(예를 들어, 객체가 모발 색의 변경을 시뮬레이팅하는 모발일 때).

[0018] CNN은 실시간으로 새로운 이미지를 비디오 이미지로서 생성하기 위해 제한된 계산 능력을 갖는 모바일 디바이스 상에서 실행되도록 구성된 훈련된 네트워크를 포함할 수 있다.

[0019] 처리 유닛은 이미지 및 새로운 이미지를 디스플레이하기 위해 대화형 그래픽 사용자 인터페이스(GUI)를 제공하도록 구성될 수 있다. 대화형 GUI는 새로운 이미지를 정의하기 위해 변경을 결정하기 위한 입력을 수신하도록 구성될 수 있다. 처리 유닛은 객체 세그먼테이션 마스크를 이용하여 이미지에서의 객체의 픽셀들을 분석하여, 변경에 대한 하나 이상의 후보를 결정하도록 구성될 수 있다. 대화형 GUI는 하나 이상의 후보를 제시하여 변경으로서 적용할 후보들 중 하나를 선택하기 위한 입력을 수신하도록 구성될 수 있다.

[0020] 이미지는 셀피 비디오(selfie video)일 수 있다.

[0021] 이미지를 처리하는 컴퓨팅 디바이스가 제공되는데, 컴퓨팅 디바이스는: 이미지의 픽셀들을 분류하여 픽셀들 각각이 모발 픽셀인지 모발 픽셀이 아닌지를 결정하도록 구성된 컨볼루션 신경망(CNN)을 저장하고 제공하는 저장 유닛 - CNN은 모발 세그먼테이션 마스크를 정의하기 위해 세그먼테이션 데이터로 훈련될 때 마스크 이미지 그래디언트 일관성 손실을 최소화하도록 훈련됨 -; 및 저장 유닛에 결합된 처리 유닛 - 처리 유닛은 모발 세그먼테이션 마스크를 이용하여 이미지에서의 모발 픽셀들에 새로운 모발 색을 적용함으로써 채색된 모발 이미지를 정의하고 제시하도록 구성됨 - 을 포함한다.

[0022] 이미지를 처리하는 컴퓨팅 디바이스가 제공되는데, 컴퓨팅 디바이스는: 처리 유닛, 처리 유닛에 결합된 저장 유닛, 및 처리 유닛 및 저장 유닛 중 적어도 하나에 결합된 입력 디바이스를 포함하고, 저장 유닛은 명령어들을 저장하고, 명령어들은 처리 유닛에 의해 실행될 때 컴퓨팅 디바이스가 입력 디바이스를 통해 이미지를 수신하고; 이미지에서의 모발 픽셀들을 식별하는 모발 세그먼테이션 마스크를 정의하고; 모발 세그먼테이션 마스크를 이용하여 이미지에서의 모발 픽셀들에 새로운 모발 색을 적용함으로써 채색된 모발 이미지를 정의하고; 및 채색된 모발 이미지를 디스플레이 디바이스에 출력하기 위해 제공하도록 구성하고; 및 모발 세그먼테이션 마스크는: 제각기 피처들을 검출하기 위해 컨볼루션 계층들의 연속으로 복수의 컨볼루션(conv) 필터들을 적용하기 위해 저장 유닛에 의해 저장된 컨볼루션 신경망(CNN)을 이용하여 이미지를 처리함으로써 정의되고, 연속하는 제1 세트의 컨볼루션 계층들은 이미지를 제1 이미지 해상도로부터 최소 해상도까지 다운샘플링하는 출력을 제공하고, 연속하는 제2 세트의 컨볼루션 계층들은 출력을 제1 이미지 해상도로 되돌려 업샘플링하고, 및 CNN은 모발 세그먼테이션 마스크를 정의하기 위해 마스크 이미지 그래디언트 일관성 손실을 최소화하도록 훈련된다.

[0023] 이미지들을 처리하여 객체 세그먼테이션 마스크를 정의하도록 훈련된 CNN을 생성하도록 구성된 컴퓨팅 디바이스가 제공되고, 컴퓨팅 디바이스는: 이미지 분류를 위해 구성된 CNN을 수신하고 제한된 계산 능력을 갖는 런타임 컴퓨팅 디바이스 상에서 실행하도록 구성된 저장 유닛; 및 입력을 수신하고 및 출력을 디스플레이하기 위해 대화형 인터페이스를 제공하도록 구성된 처리 유닛 - 처리 유닛은: 입력을 수신하여 CNN을 적응시켜 객체 세그먼테이션 마스크를 정의하고 적응된 대로의 CNN을 저장 유닛에 저장하고; 및 입력을 수신하여 객체 세그먼테이션을 위해 라벨링된 세그먼테이션 훈련 데이터를 사용하여 적응된 대로의 CNN을 훈련하여 CNN을 생성하여 이미지들을 처리함으로써 객체 세그먼테이션 마스크를 정의하도록 구성됨 - 을 포함한다. 처리 유닛은 입력을 수신하여 CNN을 적응시켜 훈련할 때 마스크 이미지 그래디언트 일관성 손실을 최소화하도록 정의된 마스크 이미지 그래디언트 일관성 손실 함수를 최소화하는 것을 사용하는 것; 및 인코더 스테이지에서의 계층들과 디코더 스테이지에서의 대응하는 계층들 사이의 스킵 연결들을 사용하여 디코더 스테이지에서 업샘플링할 때 저 해상도이지만 강한 피처들과 고 해상도이지만 약한 피처들을 조합하고; 및 적응된 대로의 CNN을 저장 유닛에 저장하는 것 중

적어도 하나를 하도록 구성될 수 있다.

[0024] 이미지들을 처리하여 객체 세그먼테이션 마스크를 정의하도록 훈련된 CNN을 생성하는 방법이 제공된다. 방법은: 이미지 분류를 위해 구성되고 및 제한된 계산 능력을 갖는 컴퓨팅 디바이스 상에서 실행되도록 구성된 CNN을 획득하는 단계; CNN을 적응시켜 객체 세그먼테이션 마스크를 정의하는 단계; 및 객체 세그먼테이션을 위해 라벨링된 세그먼테이션 훈련 데이터를 사용하여 적응된 대로의 CNN을 훈련하여 객체 세그먼테이션 마스크를 정의하는 단계를 포함한다. CNN은, 마스크 이미지 그래디언트 일관성 손실을 최소화하도록 정의된 마스크 이미지 그래디언트 일관성 손실 함수를 이용하여 훈련되어 CNN을 생성하여 이미지를 처리하여 객체 세그먼테이션 마스크를 정의할 수 있다.

[0025] CNN은 이미지 분류를 위해 사전 훈련될 수 있고, 훈련하는 단계는 사전 훈련된 대로의 CNN을 추가로 훈련한다. 세그먼테이션 훈련 데이터는 잡음이 있고 거친 세그먼테이션 훈련 데이터(예를 들어, 크라우드 소싱된 데이터로부터 정의된 것)일 수 있다.

[0026] 방법은, 디코더 스테이지에서 업샘플링할 때, 저 해상도이지만 강한 피처들과 고 해상도이지만 약한 피처들을 조합하기 위해, 인코더 스테이지에서의 계층들과 디코더 스테이지에서의 대응하는 계층들 사이의 스킵 연결들을 사용하도록 CNN을 적응시키는 단계를 포함할 수 있다.

[0027] 마스크 이미지 그래디언트 일관성 손실 함수는 조합된 손실을 최소화하기 위해 바이너리 크로스 엔트로피 손실 함수와 조합될 수 있다. 방법은, 모바일 디바이스 상에 저장하기 위한 객체 세그먼테이션 마스크를 정의하도록 이미지들을 처리하기 위해 CNN을 제공하는 단계를 포함할 수 있다.

[0028] 방법이 제공되는데, 방법은 이미지의 픽셀들을 분류하여 픽셀들 각각이 객체 픽셀인지 또는 객체 픽셀이 아닌지를 결정하여 이미지에서의 객체에 대한 객체 세그먼테이션 마스크를 정의하도록 구성된 CNN을 컴퓨팅 디바이스의 저장 유닛에 저장하는 단계를 포함하고, CNN은 객체 세그먼테이션 마스크를 정의하도록 적응되고 세그먼테이션 데이터로 훈련된, 이미지 분류를 위한 사전 훈련된 네트워크를 포함한다.

[0029] CNN은, 세그먼테이션 데이터를 사용하여 훈련될 때 마스크 이미지 그래디언트 일관성 손실을 최소화하도록 추가로 훈련되었고; 및 CNN을 이용하여 컴퓨팅 디바이스의 처리 유닛을 통해 이미지를 처리하여 객체 세그먼테이션 마스크를 생성하여 새로운 이미지를 정의한다.

[0030] 이미지를 처리하는 방법이 제공된다. 방법은: 이미지의 픽셀들을 분류하여 픽셀들 각각이 모발 픽셀인지 모발 픽셀이 아닌지를 결정하도록 구성된 컨볼루션 신경망(CNN)을 컴퓨팅 디바이스의 저장 유닛에 저장하는 단계 - CNN은 모발 세그먼테이션 마스크를 정의하기 위해 세그먼테이션 데이터로 훈련될 때 마스크 이미지 그래디언트 일관성 손실을 최소화하도록 훈련됨 -; 및 저장 유닛에 결합된 컴퓨팅 디바이스의 처리 유닛에 의해 모발 세그먼테이션 마스크를 이용하여 이미지에서의 모발 픽셀들에 새로운 모발 색을 적용함으로써 채색된 모발 이미지를 정의하고 제시하는 단계를 포함한다.

[0031] 이미지를 처리하는 방법이 제공되는데, 이 방법은: 이미지를 처리 유닛을 통해 수신하는 단계; 처리 유닛을 통해, 이미지에서의 모발 픽셀들을 식별하는 모발 세그먼테이션 마스크를 정의하는 단계; 처리 유닛을 통해 모발 세그먼테이션 마스크를 이용하여 이미지에서의 모발 픽셀들에 새로운 모발 색을 적용함으로써 채색된 모발 이미지를 정의하는 단계; 및 처리 유닛을 통해 채색된 모발 이미지를 디스플레이 디바이스로의 출력을 위해 제공하는 단계를 포함한다. 모발 세그먼테이션 마스크는: 처리 유닛에 결합된 저장 유닛에 의해 저장된 컨볼루션 신경망(CNN)을 이용하여 이미지를 처리하여 컨볼루션 계층들의 연속으로 복수의 컨볼루션(conv) 필터를 적용하여 제각기 피처들을 검출하고 - CNN은 잡음이 있고 거친 세그먼테이션 데이터로 훈련됨 - 및 모발 세그먼테이션 마스크를 정의하기 위해 마스크 이미지 그래디언트 일관성 손실을 최소화함으로써 정의된다.

[0032] 연속하는 제1 세트의 컨볼루션 계층들은 이미지를 제1 이미지 해상도로부터 최소 해상도까지 다운샘플링하는 출력을 제공할 수 있고 연속하는 제2 세트의 컨볼루션 계층들은 출력을 제1 이미지 해상도로 되돌려 업샘플링하고, CNN은 제1 세트 및 제2 세트로부터의 대응하는 컨볼루션 계층들 간의 스킵 연결들을 이용한다.

[0033] CNN은 제2 세트의 컨볼루션 계층들의 초기 계층 이전에 그리고 제2 세트의 컨볼루션 계층들의 제각기 후속 계층 이전에 산재(intersperse)된 업샘플링 함수를 포함하여 출력을 제1 이미지 해상도로 업샘플링할 수 있다.

[0034] 업샘플링 함수는 제각기 스킵 연결들을 사용할 수 있고, 제각기 스킵 연결들 각각은: 제2 세트의 컨볼루션 계층들의 다음 컨볼루션 계층에 대한 입력으로서 연속하는 인접한 컨볼루션 계층으로부터 출력되는 제1 활성화 맵; 및 제1 세트의 컨볼루션 계층들에서의 더 초기의 컨볼루션 계층으로부터 출력되는 제2 활성화 맵 - 제2 활성화

맵은 제1 활성화 맵보다 더 큰 이미지 해상도를 가짐 - 을 조합한다. 제각기 스킵 연결들 각각은, 제1 활성화 맵에 적용되는 업샘플링 함수의 출력에 제2 활성화 맵에 적용되는 컨볼루션 1 x 1 필터의 출력을 더하여 제1 활성화 맵의 해상도를 더 큰 이미지 해상도로 증가시키도록 정의될 수 있다.

[0035] 방법은 처리 유닛에 의해 디스플레이 디바이스를 통해 그래픽 사용자 인터페이스(GUI)를 제시하는 단계를 포함할 수 있고, GUI는 이미지를 보기 위한 제1 부분 및 채색된 모발 이미지를 동시에 보기 위한 제2 부분을 포함한다.

[0036] 방법은 상이한 조명 상태에서 새로운 모발 색을 보여주기 위해 조명 조건 처리를 채색된 모발 이미지에서의 새로운 모발 색에 적용하는 단계를 포함할 수 있다.

[0037] 새로운 모발 색은 제1 새로운 모발 색일 수 있고 채색된 모발 이미지는 제1 채색된 모발 이미지일 수 있다. 이러한 경우에, 방법은: 모발 세그먼테이션 마스크를 이용하여 이미지에서의 모발 픽셀들에 제2 새로운 모발 색을 적용함으로써 및 디스플레이 디바이스로의 출력을 위해 제2 채색된 모발 이미지를 제공함으로써 제2 채색된 모발 이미지를 처리 유닛에 의해 정의하고 제시하는 단계를 포함할 수 있다. 방법은 처리 유닛에 의해 디스플레이 디바이스를 통해 2개의 새로운 색 GUI를 제시하는 단계를 포함할 수 있고, 2개의 새로운 색 GUI는 제1 채색된 모발 이미지를 보기 위한 제1 새로운 색 부분 및 제2 채색된 모발 이미지를 동시에 보기 위한 제2 새로운 색 부분을 포함한다.

[0038] 방법은: 처리 유닛에 의해 이미지에서의 모발 픽셀들의 현재 모발 색을 포함하는 색에 대한 이미지를 분석하는 단계; 하나 이상의 제안된 새로운 모발 색을 결정하고; 및 채색된 모발 이미지를 정의하도록 새로운 모발 색을 선택하기 위해 GUI의 대화형 부분을 통해 제안된 새로운 모발 색들을 제시하는 단계를 포함할 수 있다.

[0039] 처리 유닛은 CNN을 실행하기 위한 그래픽 처리 유닛(GPU)을 포함할 수 있고, 처리 유닛 및 저장 유닛은 스마트폰 및 태블릿 중 하나를 포함하는 컴퓨팅 디바이스에 의해 제공될 수 있다. 방법 이미지는 비디오의 복수의 비디오 이미지 중 하나일 수 있고, 방법은 모발 색을 변경하기 위해 처리 유닛에 의해 복수의 비디오 이미지를 처리하는 단계를 포함할 수 있다.

[0040] CNN은 이미지의 픽셀들을 분류하여 픽셀들 각각이 모발 픽셀인지를 결정하도록 적응되는 MobileNet 기반 모델을 포함할 수 있다. CNN은 개별 표준 컨볼루션들이 뎁스와이즈 컨볼루션(depthwise convolution) 및 포인트와이즈 컨볼루션(pointwise convolution)으로 인수분해되는 컨볼루션들을 포함하는 뎁스와이즈 분리가능 컨볼루션 신경망으로서 구성될 수 있으며, 뎁스와이즈 컨볼루션은 각각의 입력 채널에 단일 필터를 적용하는 것으로 제한되고, 포인트와이즈 컨볼루션은 뎁스와이즈 컨볼루션의 출력들을 조합하는 것으로 제한된다.

[0041] (비일시적) 저장 유닛이, 처리 유닛에 의해 실행될 때, 컴퓨팅 디바이스의 동작들을 구성하여 본 명세서에서의 컴퓨터 구현 방법 양태들 중 임의의 것을 수행하도록 하는 명령어들을 저장하는 컴퓨터 프로그램 제품 양태들을 포함하는 이들 및 다른 양태들이 본 기술분야의 통상의 기술자에게 명백할 것이다.

도면의 간단한 설명

[0042] 도 1은 본 명세서에서의 예에 따라 이미지의 픽셀들을 분류하기 위한 랜덤 포레스트 기반 분류기(random forest based classifier)의 적용의 스케치이다.

도 2는 Google Inc.로부터의 MobileNet 모델에 따른 심층 신경망 아키텍처의 예시이다.

도 3은 본 명세서에서의 예에 따라 이미지를 처리하고 이미지 마스크를 출력으로서 생성하도록 구성된 심층 신경망 아키텍처의 예시이다.

도 4는 본 명세서에서의 예에 따라 이미지를 처리하고 이미지 마스크를 출력으로서 생성하기 위해 스킵 연결들을 사용하도록 구성된 심층 신경망 아키텍처의 예시이다.

도 5는 본 명세서에서의 예에 따라 이미지를 처리하고 이미지 마스크를 출력으로서 생성하도록 훈련하기 위해 마스크 이미지 그래디언트 일관성 손실 측도에 의해 스킵 연결들을 사용하도록 구성된 심층 신경망 아키텍처의 예시이다.

도 6은 도 3 내지 도 5의 신경망 모델들에 대한 정성적 결과들의 이미지 테이블이다.

도 7은 도 3 및 도 5의 신경망 모델들에 대한 정성적 결과들의 이미지 테이블이다.

도 8은 본 명세서에서의 예에 따라 이미지들을 처리하기 위해 심층 신경망을 갖도록 구성된 컴퓨팅 디바이스의

블록도이다.

도 9는 본 명세서에 제공된 예들에 따른 도 8의 컴퓨팅 디바이스의 동작들의 흐름도이다.

도 10은 본 명세서의 교시에 따라 예시적 CNN을 정의하는 단계들의 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0043] 본 발명의 개념은 그의 특정 실시예들을 통해 최상으로 설명되는데, 실시예들은 첨부 도면들을 참조하여 본 명세서에서 설명되고, 첨부 도면들에서는 유사한 참조 번호들이 그 전체에 걸쳐 유사한 특징들을 가리킨다. 본 명세서에서 사용될 때, 발명이라는 용어는 단지 실시예들 자체가 아닌, 이하에 설명되는 실시예들의 기초에 깔린 발명 개념을 암시하도록 의도된다는 점이 이해되어야 한다. 일반적인 발명 개념은 이하 설명되는 예시적인 실시예들로만 제한되지 않고, 이하의 설명들은 이러한 관점에서 읽어야 한다는 것을 또한 이해해야 한다.

[0044] 추가적으로, 단어 예시적이라는 것은 본 명세서에서 예, 사례 또는 예시로서 역할함을 의미하기 위해 사용된다. 본 명세서에서 예시적인 것으로서 지정되는 구성, 프로세스, 설계, 기술 등의 임의의 실시예가 반드시 다른 그러한 실시예들에 비해 바람직하거나 유리한 것으로 해석되어야 하는 것은 아니다. 예시적인 것으로서 본 명세서에 표시된 예들의 특정 품질 또는 적합성은 의도되지도 않고 추론되어야 하지도 않는다.

[0045] 딥 러닝 및 세그먼테이션

[0046] 실시간 이미지 세그먼테이션은 다수의 응용을 갖는 컴퓨터 비전에서 중요한 문제이다. 이들 중에는 미용 애플리케이션에서의 생생한 색 증강을 위한 모발의 세그먼테이션이 있다. 그러나, 이러한 사용 경우는 추가적인 도전 과제들을 제시한다. 첫째, 간단한 형상을 갖는 많은 물체와는 달리, 모발은 매우 복잡한 구조를 갖는다. 현실적인 색 증강을 위해, 거친 모발 세그먼테이션 마스크는 불충분하다. 그 대신에 모발 매트가 필요하다. 둘째, 많은 미용 애플리케이션들이 모바일 디바이스들 상에서 또는 웹 브라우저들에서 실행되는데, 여기서는 강력한 컴퓨팅 자원들이 이용가능하지 않다. 이는 실시간 성능을 달성하는 것을 더 어렵게 만든다. 모바일 디바이스 상에서 30fps를 초과하여 정확하게 모발을 세그먼팅하기 위한 시스템 및 방법 등이 본 명세서에 기술되어 있다.

[0047] 모발 세그먼테이션 시스템 및 방법은 컨볼루션 신경망(CNN)들에 기초한다. 대부분의 현대 CNN들은 강력한 GPU들 상에서도 실시간으로 실행될 수 없고, 대량의 메모리를 점유할 수 있다. 본 명세서에서의 시스템 및 방법의 목표는 모바일 디바이스 상의 실시간 성능이다. 제1 기여에서, 모바일 디바이스 상에서 사용될 만큼 충분히 빠르고 콤팩트한, 모발 세그먼테이션을 위한 Google Inc.의 최근에 제안된 MobileNets™ 아키텍처를 적응시키는 방법이 도시되어 있다. MobileNets에 관한 상세 사항은 2017년 4월 17일의 arXiv:1704.04861v1[cs:CV]에서의 하워드(Howard) 등에 의해 저술된 "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications"에서 발견할 수 있고, 이 문건은 참조에 의해 본 명세서에 통합된다.

[0048] 상세한 모발 세그먼테이션 실측(ground truth)이 없는 경우, 네트워크는 잡음이 있는 거친 크라우드 소싱 데이터(라벨링이 픽셀 레벨에서 미세하게 정확하지 않은 거친 세그먼테이션 데이터) 상에서 훈련된다. 그러나, 거친 세그먼테이션 결과는 모발 색 증강 목적을 위해 심미적으로 불만족스럽다. 현실적인 색 증강을 위해서는, 더 정확한 모발 매트가 개선된 결과들을 산출한다. 제2 기여에서, 정확한 모발 매트 훈련 데이터에 대한 필요 없이 실시간으로 더 정확한 모발 매트들을 획득하기 위한 방법을 제안한다. 먼저, 미세 레벨 상세 사항들을 캡처하기 위한 용량을 갖도록 베이스라인 네트워크 아키텍처를 수정하는 방법이 보여진다. 다음으로, 지각적으로 매력적인 매팅(matting) 결과들을 촉진하는 보조 손실 함수를 추가함으로써, 네트워크가 거친 모발 세그먼테이션 훈련 데이터만을 사용하여 상세한 모발 매트들을 산출하도록 훈련될 수 있다는 것이 보여진다. 이 접근법을 간단한 가이드된(guided) 필터(이미지 크기에 대한 라이너 런타임 복잡성을 갖는 에지 보존 필터) 후처리에 비교해 보면, 보다 정확하고 더 선명한 결과를 산출한다는 것을 보여준다.

[0049] 딥 러닝 및 세그먼테이션을 상세히 설명하기 전에, 모바일 디바이스들 상에서 비디오를 위한 모발 채색 솔루션을 개발하는 초기의 접근법들이 본 출원인에 의해 착수되고 평가되었다. 예로서, 색 히스토그램, 위치 및 그래디언트 인자들의 피처들에 기초하여 랜덤 포레스트(RF) 모델을 통합하는 분류기가 개발되었다. 분류기는 픽셀들을 연속적으로 처리하여, 필터에서의 중심 픽셀이 모발 픽셀인지 모발 픽셀이 아닌지를 결정하기 위해, 잘 알려진 바와 같이, 이미지 주위의 필터 또는 커널을 슬라이딩시킨다. 도 1에는, 필터(102)의 중심 픽셀(106)이 모발 픽셀인지 아닌지를 출력하는 RF 분류기(104)를 이용하여 픽셀들을 연속적으로 처리하기 위해 이미지 위에서 필터(102)를 스캐닝함으로써 이미지(100)가 처리되는 스케치가 도시된다. 출력(108)은 모발 매트(도시되지

않음)를 정의하기 위해 사용될 수 있고, 여기서 출력은 모발 매트에서의 대응하는 픽셀의 데이터 값이다. 도 1의 스케치에서, 필터(102), 중심 픽셀(106), 및 예시된 경로는 일정한 비율로 되어 있지 않다는 것이 이해된다. 결과는 속도에서도 정확도에서도 유망하지 않았으며, 따라서 딥 러닝을 선호하여 이 접근법은 폐기되었다.

[0050] 처음부터 딥 러닝이 선택되지 않았던 이유는, 그것이 모바일 디바이스들 상에서 실시간으로 실행되게 하는 것이 여전히 상당히 어렵다는 것 때문이다. 대부분의 딥 러닝 아키텍처는 심지어 강력한 GPU들 상에서도 실시간으로 실행되지 않는다. 초기 접근법은 비주얼 그룹 기하학적(Visual Group Geometry) 신경망 아키텍처, 즉 VGG16을 적응시켰다. ImageNet(객체 인식 소프트웨어 검색과 함께 사용하도록 설계된 대규모 시각적 데이터베이스(오픈 소스 데이터세트)) 상에서 사전 훈련된 VGG16 기반 분류 네트워크는 마지막 3개의 계층(예를 들어, 완전 연결된 계층들 및 출력 계층)을 제거하고 여러 개의 컨볼루션(종종 본 명세서에서 "conv"로 축약됨) 전치(transpose) 계층들을 추가하여 시맨틱 세그먼테이션 네트워크로 변환함으로써 적응되었다. 결과들(출력)이 매우 양호하지만, 처리는 느리고, 특히 모바일 디바이스 상에서 (프레임 당 초에 걸쳐) 그러했다. 따라서, 이 접근법은 크기가 더 작고 또한 처리 속도 등을 향상시키기 위해 더 적은 동작들을 수행하는 더 경량의 아키텍처를 찾기 위해 바뀌었다.

[0051] MobileNet 아키텍처

[0052] Google Inc.의 MobileNet 아키텍처는 모바일 디바이스들에 대해 구현되는 경량의 사전 훈련된 딥 러닝 신경망 아키텍처이다. 도 2는 입력으로서의 소스 이미지(100), MobileNet의 사전 훈련된 CNN에 따른 네트워크 계층들/계층 그룹들(202), 및 MobileNet 네트워크(200)의 출력으로서의 클래스 라벨들(204)을 보여주는 개략도이다. MobileNet에 의해 처리되는 입력 이미지들은 224 x 224 픽셀이고 3개의 색 값의 해상도(즉, 224 x 224 x 3)를 갖는다.

[0053] MobileNet 아키텍처는 처리 동작들(즉, 부동 소수점 연산들, 곱셈들 및/또는 덧셈들 등)을 최소화하기 위해 뎁스와이즈 분리가능 컨볼루션들(인수분해된 컨볼루션들의 형태)을 채택한다. 뎁스와이즈 분리가능 컨볼루션들은, 요구되는 동작들의 수를 감소 또는 최소화함으로써 처리를 더 고속화하기 위한 관점에서, 표준 컨볼루션을 뎁스와이즈 컨볼루션 및 1 x 1 컨볼루션(또한 "포인트와이즈 컨볼루션"으로 지칭됨)으로 인수분해한다(예를 들어, 그것의 함수들을 토해 낸다). 뎁스와이즈 컨볼루션은 각각의 입력 채널에 단일 필터를 적용한다. 이어서, 포인트와이즈 컨볼루션은 뎁스와이즈 컨볼루션의 출력들을 조합하기 위해 1 x 1 컨볼루션을 적용하고, 필터링 및 조합 기능들/동작들을, 표준 컨볼루션들에 의해 수행되는 단일 필터링 및 조합 동작이 아니라 2개의 단계로 분리한다. 따라서 아키텍처(200)에서의 구조들은, "계층 그룹"을 정의하거나 예시하기 위해 구조 당 2개의 컨볼루션 계층, 1개의 뎁스와이즈 컨볼루션 계층 및 1개의 포인트와이즈 컨볼루션 계층을 포함할 수 있다.

[0054] 표 1은 좌측에서 시작하여 우측으로 Softmax 동작까지의 17개의 계층/계층 그룹(202) 각각에 대한 활성화 맵 크기 정보 및 처리 동작(들) 정보를 보여준다. 엄격하게는, MobileNet는 그의 제1 풀 컨볼루션 계층 및 그의 완전 연결층으로부터 28개의 컨볼루션 계층을 갖는다.

표 1

계층들/계층 그룹들	맵 크기	처리 동작(들)				
		Conv 3X3	뎁스와이즈 컨볼루션 3X3+BN+ReLU+Conv 1X1+BN+ReLU	평균 풀 7x7	완전 연결	Softmax
1	112x112x32	X				
2	112x112x64		X			
3-4	56x56x128		X			
5-6	28x28x256		X			
7-12	14x14x512		X			
13-14	7x7x1024		X			
15	1x1x1024			X		
16	1x1x1024				X	
17	1x1x1024					X

[0055]

[0056] 표 1에서, BN은 활성화들(하나의 계층/동작들로부터 다음의 것으로 제공되는 활성화 맵에서의 개별 값들)을 조정 및 스케일링함으로써 (후속 동작에 대한) 입력을 정규화하기 위한 배치 정규화(batch normalization, batchnorm) 함수를 나타낸다. ReLU는 정류기이고, 정류된 선형 유닛 함수를 나타낸다(예를 들어, 입력 X에 대해 max function(X, 0)이어서, X의 모든 네거티브 값들이 0에 설정되도록 한다). 다운샘플링은 뎁스와이즈 컨볼루션들에서뿐만 아니라 제1 계층에서 스트라이드된 컨볼루션으로 처리된다. 평균 풀 7 x 7에 의한 최종 다운샘플은 7 x 7 어레이에서의 평균 값들에 기초한 다운샘플링 함수를 사용한다. Softmax, 또는 정규화된 지수 함수는 임의의 실수 값들의 K-차원 벡터 z를 실수 값들의 K-차원 벡터 $\sigma(z)$ 으로 "밀어넣고(squash)", 여기서 각각의 엔트리는 범위(0, 1)에 있고, 모든 엔트리들은 합하여 1이 된다(예를 들어, 스케일링 및 정규화 함수). 유용하게는, 출력은 범주적 분포(categorical distribution) - K개의 상이한 가능한 결과(카테고리들 또는 클래스들) 상에서의 확률 분포 - 를 나타내기 위해 사용될 수 있고, 따라서 K개의 클래스로 분류하는 신경망 분류기들에 의해 빈번하게 사용된다. 제각기 17개의 계층은 도 2에서 처리 동작(들)에 의해 그레이스케일 및 패턴 코딩된다.

[0057] 도 3은 계층들/계층 그룹들(302)을 갖는 적응된 딥 러닝 네트워크(300)를 예시하는 도면이다. 네트워크(300)는 또한 처리를 위한 입력으로서 소스 이미지(100)를 수신한다. 네트워크(300)(예를 들어, 그 계층들/계층 그룹들(302))는 모발 마스크(304)를 출력하도록 적응된다. 표 2는 좌측에서 시작하여 우측으로 진행하는 22개의 계층/계층 그룹(302) 각각에 대한 활성화 맵 크기 정보 및 처리 동작(들) 정보를 보여준다.

표 2

계층/그룹	맵 크기	처리 동작(들)					
		Conv 3X3	뎁스와이즈 컨볼루션 3X3+BN+ReLU+Conv 1X1+BN+ReLU	업 샘플링	뎁스와이즈 컨볼루션 3x3+Conv 1X1+BN+ReLU	Conv 1x1 +ReLU	Softmax
1	112x112x32	X					
2	112x112x64		X				
3-4	56x56x128		X				
5-6	28x28x256		X				
7-12	28x28x512		X				
13-14	28x28x1024		X				
15	56x56x1024			X			
16	56x56x64				X		
17	112x112x64			X			
18	112x112x64				X		
19	224x224x64			X			
20	224x224x64				X		
21	224x224x2					X	
22	224x224x2						X

[0058]

[0059] 네트워크(300)는 네트워크(200)와 유사하지만 적응된다. 네트워크(200)의 14 x 14 해상도에의 및 그 후 7x7 해상도에의 다운샘플링이 회피되고, 최소 해상도는 계층들 5-14에서 28 x 28이다. 최종 3개의 계층이 제거된다(즉, 2개의 완전 연결 계층 및 Softmax 계층, 그렇지만 최종 Softmax 계층이 또한 사용된다). 미세 상세사항들을 보존하기 위해, 2 대 1의 스텝 크기로 최종 2개의 계층의 스텝 크기를 변경함으로써 출력 피처 해상도가 증가된다. MobileNet의 베이스 아키텍처에 통합된 ImageNet에 대한 사전 훈련된 가중치들의 사용으로 인해, 업데이트된 해상도를 갖는 계층들에 대한 커널들은 그들의 원래 해상도에 대해 그들의 스케일링 인자에 의해 팽창된다. 즉, 2의 인자만큼 증가된 계층들에 대한 커널들은 2만큼 팽창되고 4의 인자만큼 증가된 계층들에 대한 커널들은 4만큼 팽창된다. 이는 인코더 스테이지에서 28 x 28의 최종 최소 해상도를 산출한다. 계층들/계층 그룹들 15 및 그 이후의 것들은 디코더 스테이지를 정의할 수 있다.

[0060] 계층들 2-14, 16, 18 및 20은 뎁스와이즈 분리가능 컨볼루션들, 즉 인수분해된 표준 컨볼루션들을 포함하는데, 여기서 뎁스와이즈 컨볼루션은 단일 필터를 각각의 입력 채널에 적용하고, 포인트와이즈 컨볼루션은 뎁스와이즈 컨볼루션의 출력들을 조합한다. 뎁스와이즈 분리 컨볼루션들은 계산 및 모델 사이즈를 감소시키는 효과를 가지며, 둘 모두는 모바일 디바이스 환경에서의 처리에 대해 도움이 된다.

[0061] 디코더 페이즈는 인코더 페이즈로부터의 상기 CNN 피처들을 입력으로서 취하고, 이들을 원래의 224 x 224 해상도에서의 모발 마스크로 업샘플링한다. 업샘플링은 계층들 15, 17 및 19에서 수행되고, 계층들 16, 18 및 20에서의 추가 피처 분석으로 교번된다. 업샘플링은 반전된 MobileNet 아키텍처의 단순화된 버전에 의해 수행된다. 각각의 스테이지에서, 동작들은 2 x 2 이웃에서 각각의 픽셀을 복제함으로써 이전 계층을 2의 인자만큼 업샘플링한다. 이어서, 표 2에 도시된 바와 같이 분리가능 뎁스와이즈 컨볼루션이 적용되고, 이어서 64개의 필터를 갖는 포인트와이즈 1 x 1 컨볼루션들이 뒤따르고, ReLU가 뒤따른다. 동작들은, Softmax 활성화 및 모발/비-모발에 대한 2개의 출력 채널로 1 x 1 컨볼루션을 더함으로써 계층/계층 그룹 21에서 종료한다.

[0062] 도시되지는 않았지만, 네트워크는 예측된 마스크들과 실측 마스크들 사이의 바이너리 크로스 엔트로피 손실 L_M을 최소화함으로써 훈련된다. 바이너리 크로스 엔트로피는 도 5와 관련하여 이하에 더 논의된다.

[0063] 따라서, 도 3은 모발 세그먼테이션을 위한 완전 컨볼루션 MobileNet 아키텍처를 도시한다. 따라서, 이와 같이 정의되고 훈련된 모델은 제각기 피처들을 검출하기 위해 컨볼루션 계층들의 연속으로 복수의 컨볼루션(conv) 필터를 적용하고, 여기서 연속적인 제1 세트의 컨볼루션 계층들은 이미지를 제1 이미지 해상도로부터 최소 해상도로 다운샘플링(및 인코딩)하는 출력을 제공하고, 연속적인 제2 세트의 컨볼루션 계층들은 제1 이미지 해상도로 되돌려 출력을 업샘플링한다. 모델은 제2 세트의 컨볼루션 계층들의 초기 계층 이전에 그리고 제2 세트의 컨볼루션 계층들의 제각기 후속하는 계층들 이전에 산재된 업샘플링 동작들을 가져서 출력을 제1 이미지 해상도로 업샘플링한다. 제1 세트는 인코더 스테이지를 정의하고 제2 세트는 디코더 스테이지를 정의할 수 있다.

[0064] 심층 신경망들을 훈련하는 것은 많은 양의 데이터를 요구한다. 일반적인 시맨틱 세그먼테이션을 위한 큰 데이터세트들이 있지만, 이들 데이터세트들은 모발 세그먼테이션에 대해 훨씬 인기가 덜하다. 더욱이, 비교적 단순한 형상을 갖는 자동차와 같은 일부 물체와는 달리, 모발 형상은 매우 복잡하다. 따라서, 모발에 대한 정확한 실측 세그먼테이션를 획득하는 것은 훨씬 더 어려운 과제이다.

[0065] 이 과제에 대처하기 위해, ImageNet 상의 사전 훈련된 네트워크가 사용되었다. 이것은 모발 세그먼테이션 데이터에 대해 더 미세 조정되었다. 그럼에도 불구하고, 수천 개의 훈련 이미지가 여전히 필요하다. 사용자들이 그들의 모발을 수동으로 마킹해야만 하는 모발 채색 앱을 사용하여 데이터가 크라우드 소싱되었다. 이 데이터를 얻는 것은 저렴하지만, 결과적인 모발 세그먼테이션 라벨들은 매우 잡음이 많고 거칠다. 이 소스 데이터는 충분히 양호한 모발 마스크들을 가진 사람 얼굴들의 이미지들만을 유지함으로써 수동으로 클리닝될 수 있다. 이것은 스크래치로부터 모발을 마킹하거나 잘못된 세그먼테이션들을 고치는 것보다 상당히 빠르다. 2개의 인 하우스 테스트 데이터 세트가 유사하게 정의된다.

[0066] 상기 네트워크(300)는 애플 사(Apple Corporation)로부터의 코어 ML™ 라이브러리를 포함하는 Apple iPad Pro 12.9(2015) 상에서 구현되었다. 코어 ML은 MobileNet 모델로부터 MobileNet 클래스를 자동으로 생성하고, 설명된 바와 같이 적응될 수 있다. 병렬화를 이용하기 위해, 모델은 아이패드의 GPU 및 그 관련 메모리를 사용하여 처리되었다. 일부 구현들에 대해서는 원하는 처리를 달성하기 위해 CNN이 GPU에 의해 처리(예를 들어, 실행)될 수 있고, 다른 경우들에서는 컴퓨팅 디바이스의 CPU를 사용하여 처리하는 것으로 충분할 수 있다는 점에 유의한다.

[0067] 아키텍처(300)의 콤팩트화 및 코어 ML 라이브러리의 사용으로 인해, 단일 이미지에 걸친 순방향 패스는 60ms만을 취한다. 네트워크는 또한 Tensorflow™(Google Inc.에 의해 원래 개발된 머신 러닝 및 심층 학습을 위한 지원을 갖는 고성능 수치 계산을 위한 오픈 소스 소프트웨어 라이브러리)를 사용하여 구현되었다. 필적할만한 처리는 ~300ms에서 더 느렸다. Tensorflow는 NEON™ 최적화를 갖지만(NEON 기술은 ARM 사(Arm Limited)의 특정 프로세서들에 대한 SIMD(single instruction multiple data) 아키텍처 확장에서의 진보임), 이것은 그래픽 처리를 위해 최적화되지 않았다. 현대의 폰들 및 태블릿들 상의 GPU들은 상당한 전력을 소비하는 것으로 인식된다.

[0068] 이 모델은 인 하우스(In-house) 세트 1이 350개의 얼굴 크로핑된 이미지(face cropped image)를 포함하는, 표 3에 도시된 대로의 이미 매우 양호한 정성적 및 정량적 결과를 산출한다. 이는 크라우드 소싱된 데이터로부터 수동으로 주석화되고, 사용되는 훈련 데이터와 유사하다. 인 하우스 세트 2는 (소스 입력 디바이스로부터) 3:4 종횡비의 108개의 얼굴 이미지이고, 수동으로 라벨링된다. 인 하우스 세트 1은 더 거친 수동 라벨링을 갖고, 인 하우스 세트 2는 더 미세한 수동 라벨링이나, 어느 세트도 미세한 라벨링을 갖지 않는다.

표 3

[0069]

테스트 세트	정밀도	리콜	F1 점수	IoU	성능	정확도
인 하우스 세트 1	0.89406	0.93423	0.91045	0.84302	0.77183	0.95994
인 하우스 세트 2	0.904836	0.92926	0.914681	0.84514	0.76855	0.95289

[0070] 그러나, 이러한 접근법은 여전히 모든 모발를 캡처하지 않고, 정확한 알파 매트가 아니라 단지 거칠고 및 얼룩진 마스크만을 제공한다. 이것을 시각적으로 더 매력적이게 하기 위해 가이드된 필터링을 이용하여 결과적인 마스크를 후처리하는 것은 이하에 더 설명되는 바와 같이 사소한 오류들만을 정정한다.

[0071] CNN들을 이용하여 더 진짜(및 바람직하게는 진짜) 매팅을 획득하기 위해 결과를 개선하라는 요구가 있다. 이 프레임워크에 2개의 과제가 존재한다 - CNN(300)은 인코딩 스테이지에서 이미지를 아주 많이 다운샘플링하고, 따라서 결과적인 마스크들은 매우 높은 해상도의 상세 사항을 포함할 것으로 예상될 수 없다. 마찬가지로, 훈련 데이터도 테스트 데이터도 매팅 방법을 훈련 및 평가하는 데에 언급된 바와 같이, 충분한 정확도로 이용가능하지 않다.

[0072] 다운샘플링의 첫 번째 문제를 해결하기 위해, 스킵 연결들이 업샘플링 동작들을 재정의하기 위해 아키텍처에 추가된다. 스킵 연결들을 추가함으로써, 강하지만 저 해상도의 피처들이 더 약하지만 더 고 해상도의 피처들과 조합된다. 추가된 스킵 연결들로 인해, 다운샘플링을 제한할 필요가 더 이상 없기 때문에, 아키텍처는 원래의 인코더 아키텍처로 복귀하여 7 x 7 해상도까지 진행하는 것을 유의한다. 도 3의 아키텍처가 적응되었고 해상도가 예를 들어 28 x 28에서 중단되었다면, 더 적은 스킵 연결들이 이용되었을 것이다. 이러한 방식으로, 고 해상도이지만 약한 피처들을 포함하는 인코더에서의 얕은 계층들은 더 깊은 계층들로부터의 저 해상도이지만 강한 피처들과 조합된다. 이 계층들은, 들어오는 인코더 계층들에 1 x 1 컨볼루션을 먼저 적용하여 들어오는 디코더 계층들과 양립되는 출력 깊이(3개의 외부 스킵 연결에 대한 64 및 내부 스킵 연결에 대한 1024)를 만들고 그 다음에 더하기를 사용하여 계층들을 병합함으로써 조합된다. 각각의 해상도에 대해, 해당 해상도에서 가장 깊은 인코더 계층이 스킵 연결을 위해 취해진다.

[0073] 도 4는 계층들/계층 그룹들(402)을 갖는 적응된 딥 러닝 네트워크(400)를 도시하는 예시이다. 네트워크(400)는 또한 처리를 위한 입력으로서 소스 이미지(100)를 수신한다. 네트워크(400)(예를 들어, 그의 계층들/그룹들(402))는 모발 마스크(404)를 출력하도록 적응된다.

[0074] 표 4는 좌측에서 시작하여 우측으로의 26개의 계층/계층 그룹(402) 각각에 대한 활성화 맵 크기 정보 및 처리 동작(들) 정보를 나타낸다.

[0075] 네트워크(300) 및 네트워크(400) 모두의 모델에서, 복수의 컨볼루션 필터 각각은 활성화 맵을 생성하고, 하나의 컨볼루션 계층에 의해 생성된 활성화 맵은 연속하여 다음의 컨볼루션 계층에 입력을 제공하기 위해 출력된다. 복수의 컨볼루션 필터는 제1 세트의 컨볼루션 필터들 및 제2 세트의 컨볼루션 필터들을 포함하여서, 인코더를 포함하는 것과 같이 제1 세트의 컨볼루션 필터들이 제1 세트의 계층들에서 (예를 들어, 네트워크(400)의 계층들 2-14에서) 이미지를 처리하도록 하고; 및 제2 세트의 컨볼루션 필터들은 디코더를 포함하는 것과 같이 제2 세트의 계층들에서(예를 들어, 네트워크(400)에서의 계층들 16, 18, 20, 22 및 24에서) 이미지를 처리한다. 모발 세그먼테이션 마스크는 연속하는 최종 컨볼루션 계층으로부터 (예를 들어, 계층 25로부터) 출력된 최종 활성화 맵으로부터 정의된다.

[0076] 네트워크(300) 또는 네트워크(400)의 모델은 출력을 정규화하고 선형으로 정류하기 위해 제1 세트의 컨볼루션 필터들로 산재된 연속하는 정규화 함수 및 정류기 함수를 포함할 수 있다. 모델은 출력을 선형으로 정류하기 위해 제2 세트의 컨볼루션 필터들로 산재된 정류기 함수를 포함할 수 있다. 네트워크(300) 또는 네트워크(400)의 모델에서, 제1 세트의 계층들은: 초기 컨볼루션 3 x 3 필터에 의해 정의된 초기 계층; 및 연속하는 제각기 뎁스와이즈 컨볼루션 3 x 3 필터, 배치 정규화(batch normalization) 함수 및 정류된 선형 유닛 함수, 및 컨볼루션 1 x 1 필터와 그에 뒤따르는 배치 정규화 및 정류된 선형 유닛 함수에 의해 각각 정의되는 복수의 후속하는 뎁스와이즈 분리 가능 컨볼루션을 포함한다.

[0077] 네트워크(300) 또는 네트워크(400)의 모델에서, 제2 세트의 계층들은: 제각기 뎁스와이즈 컨볼루션 3 x 3 필터, 컨볼루션 1 x 1 필터 및 정류된 선형 유닛 함수에 의해 각각 정의되는 연속하는 복수의 초기 계층; 및 최종 컨볼루션 1 x 1 필터 및 정류된 선형 유닛 함수에 의해 정의되는 최종 계층을 포함한다.

표 4

계층/그룹	맵 크기	처리 동작(들)					
		Conv 3X3	뎁스와이즈 컨볼루션 3X3+BN+ReLU+Conv 1X1+BN+ReLU	업 샘플링	뎁스와이즈 컨볼루션 3x3+Conv 1X1+BN+ReLU	Conv 1x1 +ReLU	Softmax
1	112x112x32	X					
2	112x112x64		X				
3-4	56x56x128		X				
5-6	28x28x256		X				
7-12	14x14x512		X				
13-14	7x7x1024		X				
15	14x14x1024			X			
16	14x14x64				X		
17	28x28x64			X			
18	28x28x64				X		
19	56x56x64			X			
20	56x56x64				X		
21	112x112x64			X			
22	112x112x64				X		
23	224x224x64			X			
24	224x224x64				X		
25	224x224x2					X	
26	224x224x2						X

[0078]

[0079] 도 4는 계층들 15, 17, 19, 21 및 23에서의 업샘플링 동작들을 도시한다. 계층들 15, 17, 19 및 21에서의 업샘플링은 이전 계층의 출력 맵과 스킵 연결을 수행하는 것을 포함한다. 이전 계층의 맵은 타겟 해상도와 동일한 해상도를 갖는다. 이러한 계층들에서의 업샘플링 기능은 채널 정보를 조합하기 위해 인코더 스테이지의 더 이른 대응하는 계층의 맵에 대해 컨볼루션 1 x 1 동작을 수행하고, 제각기 계층에 대한 입력으로서 통상적으로 제공되는 인접 계층의 활성화 맵에 대해 업샘플링 동작(예를 들어, 설명된 바와 같은 2 x 2)을 수행하여, 이것이 더 이른 계층으로부터의 맵과 동일한 해상도를 갖도록 한다. 컨볼루션 1 x 1 동작 및 업샘플링 동작의 출력이 더해진다.

[0080] 따라서, 업샘플링 함수는 제각기 스킵 연결을 사용하는데, 여기서 제각기 스킵 연결들 각각은 제2 세트의 컨볼루션 계층들의 다음 컨볼루션 계층에 대한 입력으로서 연속하는 인접한 컨볼루션 계층으로부터 출력된 제1 활성화 맵 ; 및 제1 세트의 컨볼루션 계층들에서의 더 이른 컨볼루션 계층으로부터 출력된 제2 활성화 맵 - 제2 활성화 맵은 제1 활성화 맵보다 더 큰 이미지 해상도를 가짐 - 을 조합한다. 제각기 스킵 연결들 각각은, 제2 활성화 맵에 적용되는 컨볼루션 1 x 1 필터의 출력을 제1 활성화 맵에 적용되는 업샘플링 함수의 출력에 더하여 제1 활성화 맵의 해상도를 더 큰 이미지 해상도로 증가시키도록 정의된다. 마지막으로, 모발 세그먼테이션 맵은 Softmax(정규화된 지수 함수)를 최종 활성화 맵에 적용하여 모발 세그먼테이션 맵의 각각의 픽셀에 대해 0과 1 사이의 값들을 정의함으로써 정의된다.

[0081] 이 접근법의 정량적 결과들은 표 5에 도시되어 있다:

표 5

[0082]

테스트 세트	정밀도	리콜	F1 점수	IoU	성능	정확도
In-house Set 1	0.88237	0.94727	0.91024	0.84226	0.770518	0.95817
In-house Set 2	0.908239	0.94196	0.92320	0.85971	0.79548	0.95699

[0083] 더욱이, 최종 인코더 해상도에서의 감소로 인해, 상기 아키텍처는 이것이 추가적인 디코더 계층들을 포함한다 하더라도 훨씬 더 빠르다. 단일 이미지에 걸쳐 코어 ML을 이용하는 순방향 패스는 Apple iPad Pro 12.9(2015) 상에서 30ms를 취한다.

[0084] 정확도 관점에서, 스킵 연결들이 없는 모델보다 제2 세트에 대해 이것이 더 양호하게 보이기는 하지만, 제1 세트에 대한 결과는 결론이 나지 않는다. 이는 거친 세그먼테이션 데이터가 제한된 정확도를 갖는다는 상기 제2 관점을 예시하며, 이는 훈련을 방해할 뿐만 아니라 테스팅도 방해한다.

[0085] 이러한 데이터에 의한 정량적 평가는 현재의 성능 레벨에서 그 능력에 어느 정도 도달했다고 주장된다. 그러나, 정성적으로, 이 아키텍처는 단지 약간 더 좋은 것으로 또한 보인다. 하나의 가능한 설명은 스킵 연결 아키텍처가 이제 미세 레벨 상세 사항을 학습하기 위한 능력을 갖는 반면, 이러한 상세 사항들은 우리의 현재 훈련 데이터에 존재하지 않아서, 훈련 세트에서의 것들이 거친 것처럼 바로 그만큼 거친 결과적인 네트워크 출력 마스크들을 만든다는 것이다.

[0086] 마스크 이미지 그래디언트 일관성 손실의 평가 및 최소화

[0087] 이용가능한 훈련 및 테스트 데이터가 주어지면, CNN은 거친 세그먼테이션 훈련 데이터만을 사용하여 모발 매팅을 학습하는 것에 제한된다. 레만(Rhemann) 등의 작업에 의해 동기를 부여받아, 마스크 정확성의 지각적으로 고무된 측도가 추가된다. 이를 위해, 마스크와 이미지 그래디언트들 사이에 일관성 측도가 추가된다. 거리(손실) 측도는 다음과 같다.

[0088] 마스크 이미지 그래디언트 일관성 손실은 수학식 1에 도시된다.

수학식 1

[0089]
$$L_c = \frac{\sum M_{mag}\left[1-\left(I_x M_x + I_y M_y\right)^2\right]}{\sum M_{mag}}$$

[0090] 여기서, I_x, I_y는 정규화된 이미지 그래디언트이고, M_x, M_y는 정규화된 마스크 그래디언트이고, M_{mag}은 마스크 그래디언트 크기이다. 손실(L_c)의 값은 이미지와 마스크 그래디언트들 사이에 합의가 있을 때 작다. 이 손실은 가중치 w로 원래의 바이너리 크로스 엔트로피 손실에 추가되어, 전체 손실을 만든다.

수학식 2

[0091]
$$L = L_M + wL_C$$

[0092] 2개의 손실의 조합은 이미지 에지들에 부착되는 마스크들을 생성하는 동안 훈련 마스크들에 대해 참인 것 사이

의 균형을 유지한다. 이 마스크 이미지 그래디언트 일관성 손실 측도는 기존의 모델들을 평가하고 새로운 모델을 훈련하는 것 둘 다를 위해 사용되며, 여기서 바이너리 크로스 엔트로피(손실) 측도는 수학식 2에 표시된 바와 같이 수학식 1의 이 새로운 측도와 조합된다.

[0093] 크로스 엔트로피 손실 또는 로그 손실은 그 출력이 0과 1 사이의 확률값인 분류 모델의 성능을 측정한다. 크로스 엔트로피 손실은 예측된 확률이 실제(참) 라벨로부터 발산할수록 증가한다. 본 예에서, 모델은 제각기 관측들 o에서의 각각의 픽셀을 2개의 클래스 c -모발, 모발이 아님- 중 하나로서 분류하고, 따라서 바이너리 크로스 엔트로피 손실 L_M은 수학식 3에서와 같이 계산될 수 있다.

수학식 3

[0094] $$L_M = -(y \log(p)+(1-y) \log(1-p))$$

[0095] 여기서 y는 관측 o에 대한 올바른 분류에 대한 바이너리 라벨(0 또는 1)이고 p는 관측 o가 클래스 c의 것인 예측된 확률이다.

[0096] 도 5는 이미지(100(I))를 수신하고 마스크(504(M))를 생성하는 스킵 연결들을 갖는 네트워크(500)를 포함하는 훈련 아키텍처를 도시한다. 네트워크(500)는 구조에 있어서 네트워크(400)에 필적하지만 논의된 손실 측도들을 사용하여 훈련된다.

[0097] 또한, 이미지(100)로부터의 I_x,I_y(정규화된 이미지 그래디언트)를 포함하는 입력(502) 및 훈련을 위한 마스크 이미지 그래디언트 일관성 손실 판정기 컴포넌트(508)에의 M_x,M_y(정규화된 마스크 그래디언트)을 포함하는 입력(506)이 도시되어 있다. 또한 바이너리 크로스 엔트로피 손실 판정기 컴포넌트(510)가 도시된다. 상술한 바와 같이 네트워크(500)를 훈련시키기 위해 손실 L 파라미터를 정의하도록 마스크 이미지 그래디언트 일관성 손실 L_C 및 바이너리 크로스 엔트로피 손실 L_M 이 조합된다(도시되지 않음).

[0098] 따라서, 네트워크(500)는 오픈 소스 이미지 훈련 데이터를 사용하여 사전 훈련된 것과 같은 이미지 분류를 위한 사전 훈련된 네트워크를 포함하는 CNN이다. 사전 훈련된 네트워크는 이미지 그 자체를 분류하기보다는 모발 세그먼테이션 마스크와 같은 객체 세그먼테이션 마스크를 정의하도록 적응된다. CNN은 훈련될 때 마스크 이미지 그래디언트 일관성 손실을 최소화하도록 더 훈련되고, 거친 세그먼테이션 데이터를 이용하여 그렇게 훈련될 수 있다. 이 마스크 이미지 그래디언트 일관성 손실은 바이너리 크로스 엔트로피 손실과 조합될 수 있다. CNN은, 인코더 스테이지에서의 계층들 및 디코더 스테이지에서의 대응하는 계층들 사이의 스킵 연결들을 사용하여 객체 세그먼테이션 마스크를 정의하기 위해 디코더 스테이지에서 업샘플링할 때 저해상도이지만 강한 피처들 및 고해상도이지만 약한 피처들을 조합하도록 더 적응될 수 있다.

[0099] 결과적인 마스크들은 훨씬 더 매트들처럼 보이고, 훨씬 더 상세화된다. 도 6은 정성적 결과를 묘사하는 이미지 테이블(600)을 도시한다. 제각기 도 3 내지 도 5의 모델들을 이용하여 생성된 제각기 마스크들(마스크들(304, 404 및 504)의 예들)뿐만 아니라 처리될 이미지(602)(이미지(100)의 예)가 도시되어 있다. 도 3에 따른 스킵 연결들이 없는 모델(모델 1)을 이용하여 마스크(604)가 생성되었고, 도 4에 따른 스킵 연결들을 갖는 모델(모델 2)을 사용하여 마스크(606)가 생성되었고, 마스크들(608)이 도 5에 따른 훈련(모델 3)에 뒤이은 스킵 연결들 및 마스크 이미지 그래디언트 일관성 손실을 갖는 모델을 사용하여 생성되었다.

[0100] 정량적으로, 이 방법은 새로운 일관성 측도에 따라 더 양호하게 수행하지만, 측도들의 나머지(실측과 유사)에 기초해 약간 더 악화된다. 그러나, 앞서 언급된 바와 같이, 이용가능한 현재 실측 정확도가 주어지면, 특정 레벨을 넘어서 실측과의 예측의 일치를 최대화하는 것이 바람직하지 않을 수 있다. 표 6은 동일한 테스트 데이터 세트들을 사용하는 모든 모델들의 정성적 결과들을 도시한다.

표 6

Mod.	테스트 세트	정밀도	리콜	F1 점수	IoU	성능	정확도	마스크 이미지 그래디언트 일관성 손실	실측 이미지 그래디언트 일관성 손실
1	인 하우스 세트 1	0.89406	0.93423	0.91045	0.84302	0.77183	0.95994	0.2539	0.175
	인 하우스 세트 2	0.904836	0.92926	0.914681	0.84514	0.76855	0.95289	0.275	0.171
2	인 하우스 세트 1	0.88237	0.94727	0.91024	0.84226	0.770518	0.95817	0.2306	0.175
	인 하우스 세트 2	0.908239	0.94196	0.92320	0.85971	0.79548	0.95699	0.2414	0.171
3	인 하우스 세트 1	0.93495	0.887203	0.89963	0.83444	0.748409	0.96119	0.095	0.175
	인 하우스 세트 1	0.95587	0.87826	0.91198	0.842757	0.74259	0.95407	0.0912	0.171

[0101]

[0102] 도 5의 아키텍처의 모델 3의 매트 출력은 도 3의 아키텍처의 모델 1의 더 거친 마스크 출력에 비교되고, 가이드된 필터(guided filter)가 추가되어 모델 1의 해당 출력과 비교되었다. 가이드된 필터는 에지 보존 필터이고, 이미지 크기에 대한 선형 런타임 복잡성을 갖는다. 이는 iPad Pro 상에서 224 x 224 이미지를 처리하기 위해 5ms만이 걸린다. 도 7은 도 3 및 5의 모델들의 정성적 결과들을 묘사하는 이미지 테이블(700)을 도시한다. 처리될 이미지(702)(이미지(100)의 예)가 도시된다. 이미지(704)는 가이드된 필터 후처리 없이, 도 3의 모델 1로부터의 마스크이다. 이미지(706)는, 추가된 가이드된 필터 후처리를 갖는 모델 1로부터의 마스크이다. 이미지(708)는 도 5의 모델 3으로부터 출력되는 마스크(또는 매트)이다. 언급된 바와 같이, 모델 3은 스킵 연결들을 사용하고, 크라우드 소스 데이터로부터의 잡음 있는 그리고 거친 세그먼테이션 데이터로 훈련되고, 마스크 이미지 그래디언트 일관성 손실 함수로 훈련된다.

[0103] 가이드된 필터를 이용하는 이미지(706)는 개개의 모발 가닥들이 더 분명해지면서 더 상세한 것을 포착하는 것을 보여준다. 그러나, 가이드된 필터는 마스크의 에지들 근처에 국부적으로만 상세 사항들을 추가한다. 더욱이, 리파인된 마스크들의 에지들은 이들 주위에 가시적인 후광을 가지며, 이것은 모발 색이 그 주변과의 더 낮은 콘트라스트를 가질 때 훨씬 더 분명해진다. 이 후광은 모발 재채색 동안 색 번짐(color bleeding)을 야기한다. 도 5의 아키텍처는 (이미지(708)에서 보여지는 바와 같이) 더 선명한 에지들을 산출하고, 가이드된 필터 후처리에서 보이는 원하지 않는 후광 효과 없이 더 긴 모발 가닥들을 캡처한다.

[0104] 추가 보너스로서, 도 5의 아키텍처는 모바일 디바이스 상의 프레임당 30ms만을 취하고 추가 후처리 매팅 단계에 대한 필요 없이 도 4의 아키텍처와 비교하여 2배 빨리 실행된다. 고해상도 상세 사항을 캡처하는 것을 돕는 스킵 연결들의 사용으로 인해, 도 5의 아키텍처는 7 x 7 해상도를 갖는 가장 깊은 계층들을 갖는 원래의 MobileNet 인코더 구조를 유지한다. 이러한 계층들은 많은 깊이 채널들(1024)을 가지며, 증가된 해상도로 처리하기에 매우 비싸게 된다. 7 x 7 해상도를 갖는 것은 도 4의 아키텍처에서의 28 x 28 해상도에 비해 훨씬 더 빠르게 처리를 이룬다.

[0105] 실험

[0106] 방법은 3개의 데이터세트에 대해 평가되었다. 먼저, 9000회 훈련, 380회 유효성 검사, 및 282개 테스팅 이미지로 구성되는 크라우드 소싱된 데이터세트가 있다. 모든 3개의 서브세트는 원래 이미지들 및 그들의 플립(flip)된 버전들을 포함한다. 타겟은 모바일 디바이스 상에서의 모발 매팅이기 때문에, 데이터의 전처리는 전형적인 셀피에 대해 예상되는 스케일에 기초하여 얼굴을 검출하고 그것 주위의 영역을 크로핑함으로써 수행된다.

[0107] 방법을 기존 접근법들과 비교하기 위해, 2개의 공개 데이터세트가 평가된다: 캐(Kae) 등의 LFW 파트 데이터세트 및 구오(Guo) 및 아라비(Aarabi)의 모발 데이터세트. 전자는 2927개의 250 x 250 이미지, 1500회 훈련, 500회 유효성 검사, 및 927개의 테스트 이미지로 구성된다. 픽셀들은 슈퍼픽셀 레벨에서 생성된 모발, 피부 및 배경의 3개 카테고리로 라벨링된다. 후자는 115개의 고해상도 이미지들로 구성된다. 이것은 너무 적은 이미지들을 가져서 훈련할 수가 없기 때문에, 우리는 이 세트에 대해 평가할 때 우리의 크라우드 소싱된 훈련 데이터를 사

용한다. 이 데이터세트를 우리의 훈련 데이터에 일치하게 하기 위해, 유사한 방식으로 (얼굴 검출 및 크로핑을 이용하여) 전처리가 수행되어, 플립된 이미지들을 또한 추가한다. 소수의 경우에 얼굴들이 검출되지 않았기 때문에, 결과적인 데이터세트는 212개의 이미지로 구성된다.

[0108] 훈련은, 학습율 1:0, ρ=0:95, 및 ε=1e-7을 사용하여 케라스(F. Chollet et al., https://github.com/keras-team/keras, 2015)에서의 아다델타(Adadelta: an adaptive learning rate method," arXiv preprint arXiv:1212.5701, 2012) 방법을 사용하여 4의 배치(batch) 크기를 이용하여 행해진다. L2 정규화가 컨볼루션 계층들에 대해서만 가중치 $2 \cdot 10^{-5}$와 함께 사용된다. 뎁스와이즈 컨볼루션 계층들 및 마지막 컨볼루션 계층은 정규화되지 않는다. 손실 균형 가중치는 (수학식 3)에서의 ω = 0:5에 설정된다.

[0109] 3 클래스 LFW 데이터에 있어서, 모발 클래스만이 마스크 이미지 그래디언트 일관성 손실에 기여한다. 이 모델은 50세(epoch)에 대해 훈련되고, 최상의 수행 세는 유효성 검사 데이터를 이용하여 선택된다. 크라우드 소싱된 데이터세트에 대한 훈련은, Nvidia GeForce GTX 1080 Ti™(Nvidia, GeForce 및 GTX 1080 Ti는 Nvidia사의 상표임) GPU 상에서 5시간이 걸리고, 훨씬 더 작은 훈련 세트 크기로 인해 LFW 파트들 상에서 1시간 미만으로 걸린다.

[0110] A. 정량적 평가

[0111] 정량적 성능 분석을 위해, F1 점수, 성능, IoU, 및 정확도가 측정되어 모든 테스트 이미지들에 걸쳐 평균화된다. 이미지 및 모발 마스크 에지들의 일관성을 측정하기 위해, 마스크 이미지 그래디언트 일관성 손실(수학식 1)이 또한 보고된다. 크라우드 소싱된 이미지들(이미지 데이터)의 수동 클린 업 동안 이미지들은 마스크들에 대해 정정되기보다는 단지 필터링되었다는 것을 상기한다. 그 결과, 모발 주석(hair annotation)의 품질은 여전히 열악하다. 따라서, 크라우드 소싱된 데이터에 대한 평가 이전에, 테스트 마스크들의 수동 보정이 착수되었고, 주석당 2분 이하를 소비한다. 이는 약간 더 좋은 실측을 산출하였다. 방법의 3개의 변형(모델 1, 가이드된 필터링을 갖는 모델 1 및 모델 3)이 이러한 재라벨링된 데이터에 대해 평가된다. 3개의 방법 모두는 실측 비교 측도들과 관련하여 유사하게 수행되지만, 모델 3은 그래디언트 일관성 손실 카테고리에서 분명한 승자이며, 이는 그 마스크들이 이미지 에지들에 훨씬 더 양호하게 부착되는 것을 나타낸다.

[0112] LFW 파트 데이터세트에 대해, 킨(Qin) 등에서 최상의 수행 방법을 가지며 온 파(on-par) 성능이 보고되지만, 이것은 모바일 디바이스 상에서 실시간으로 달성된다. 정확도 측도만이, 이것이 킨 등에서 사용되는 유일한 측도이기 때문에, 평가를 위해 사용된다. 아마 틀림없이, 특히 LFW 파트들이 슈퍼픽셀 레벨에서 주석되기 때문에, 거기에서의 실측은 높은 정확도의 분석을 위해 충분히 양호할 수 없다. 거기의 구오 및 아라비의 데이터세트에 대해서, 0:9376의 F1 점수 및 0:8253의 성능이 보고된다. HNN은 이 후처리된 데이터세트에 대해 재실행되었고, 0:7673의 F1 점수 및 0:4674의 성능으로, 저자들에 의해 보고된 것과 유사한 성능을 획득하였다.

[0113] B. 정성적 평가

[0114] 방법은 정성적 분석을 위해 공개적으로 이용가능한 셀피 이미지들에 대해 평가된다. 모델 1은 양호하지만 거친 마스크들을 산출한다. 가이드된 필터를 갖는 모델 1은 더 양호한 마스크들을 생성하지만 모발 경계 주위에 바람직하지 않은 흐릿함(blur)을 갖는다. 가장 정확하고 가장 날카로운 결과는 모델 3에 의해 달성된다. 가이드된 필터 후처리 모델 3 양쪽 모두의 실패 모드는, 어두운 모발의 경우에 눈썹들 또는 라이트 모발(light hair)에 대한 밝은 배경의 경우에서와 같이 모발 부근에서의 모발 유사 객체들의 그들의 언더 세그먼테이션(under-segmentation)이다.

[0115] 또한, 모발 내부의 하이라이트는 모델 3으로부터의 모발 마스크가 불균일하게 되도록 야기할 수 있다.

[0116] C. 네트워크 아키텍처 실험

[0117] 유효성 확인 데이터를 이용하여, 다수의 디코더 계층 채널들에 의한 실험들이 착수되었지만, 이것이 정확도에 큰 영향을 미치지 않는 것으로 관찰되었고, 64개 채널은 대부분의 측도들에 따라 최상의 결과들을 산출하였다. 이러한 실험들은 그래디언트 일관성 손실을 사용하지 않고 도 4의 스킵 연결 아키텍처를 이용하여 행해졌다.

[0118] 하워드(Howard) 등은 MobileNets가 더 높은 이미지 해상도가 주어졌을 때 더 잘 수행하는 것을 관찰하였다. 정확한 모발 매팅의 목표가 주어지면, 모델 3을 이용하여 실험들이 착수되었고, MobileNet이 ImageNet에 대해 훈련되었던 최고 해상도인 224 x 224를 넘어 해상도를 증가시켰다. 224 x 224 이미지 대 480 x 480 이미지로부터 모델 3을 이용하여 추론된 마스크들의 정성적 비교는, 480 x 480 결과가 모발 에지들 주위에서 더 정확하게 보

이는 것을 나타내며, 더 긴 모발 가닥들이 얼굴 위의(예를 들어, 코 상의) 것들을 포함하여 캡처된다. 그러나, 이전 섹션에서 언급된 문제들이 또한 강조되고, 비 모발 영역들로의 모발 마스크의 번짐이 더 많고, 마스크의 내부는 모발 하이라이트들로 인해 비균일하게 된다. 또한, 더 큰 이미지를 처리하는 것은 상당히 더 비싸다.

[0119] 위에서 언급된 바와 같이, CNN은 모바일 디바이스와 같은 사용자의 컴퓨팅 디바이스 상에서의 런타임 실행을 위해 구성된다. 그것은 CNN의 실행이 처리의 이점(예를 들어, 이러한 GPU들에서의 병렬화)을 이용하기 위해 그러한 디바이스의 GPU에 적어도 부분적으로 기초하도록 구성될 수 있다. 일부 구현들에서, 실행은 CPU 상에 있을 수 있다. 거친 세그먼테이션 데이터(훈련 데이터)를 사용하여 훈련된 네트워크를 정의하기 위한 훈련 환경들은 런타임 환경들과는 다를 수 있다는 것이 이해된다. 훈련 환경들은 훈련 동작들을 촉진하기 위해 더 높은 처리 능력 및/또는 더 많은 저장소를 가질 수 있다.

[0120] 도 8은, 핸드헬드 모바일 디바이스(예를 들어, 스마트폰 또는 태블릿)와 같은, 본 개시내용의 하나 이상의 양태에 따른, 예시적인 컴퓨팅 디바이스(800)의 블록도이다. 그러나, 이것은 랩톱, 데스크톱, 워크스테이션 등과 같은 또 다른 컴퓨팅 디바이스일 수 있다. 앞서 언급한 바와 같이, 조사 및 개발 노력의 목적은, 처리된 이미지들(예를 들어, 비디오)을 효과적으로 그리고 효율적으로 생성하기 위해 제한된 자원을 갖는 핸드헬드 모바일 디바이스 상에서 동작하기 위한 심층 신경망을 생성하는 것이다.

[0121] 컴퓨팅 디바이스(800)는, 예를 들어, 비디오와 같은 하나 이상의 이미지를 획득하고 이미지들을 처리하여 하나 이상의 속성을 변경하고 새로운 이미지들을 제시하는 사용자 디바이스를 포함한다. 일 예에서, 이미지들은 이미지들에서의 모발의 색을 변경하도록 처리된다. 컴퓨팅 디바이스(800)는 하나 이상의 프로세서(802), 하나 이상의 입력 디바이스(804), 제스처 기반 I/O 디바이스(806), 하나 이상의 통신 유닛(808) 및 하나 이상의 출력 디바이스(810)를 포함한다. 컴퓨팅 디바이스(800)는 또한 하나 이상의 모듈 및/또는 데이터를 저장하는 하나 이상의 저장 디바이스(812)를 포함한다. 모듈들은 심층 신경망 모델(814), 그래픽 사용자 인터페이스(GUI 818)를 위한 컴포넌트들을 갖는 애플리케이션(816), 색 예측(820) 및 이미지 획득(822)을 포함할 수 있다. 데이터는 처리를 위한 하나 이상의 이미지(예를 들어, 이미지(824)), 하나 이상의 이미지로부터 생성된 하나 이상의 마스크(예를 들어, 이미지(824)로부터 생성된 특수 마스크(826)), 및 하나 이상의 마스크 및 하나 이상의 이미지를 사용하여 생성된 하나 이상의 새로운 이미지(예를 들어, 새로운 이미지(828))를 포함할 수 있다.

[0122] 애플리케이션(816)은 비디오와 같은 하나 이상의 이미지를 획득하고, 이미지들을 처리하여 하나 이상의 속성을 변경하고 새로운 이미지들을 제시하는 기능을 제공한다. 일 예에서, 이미지들은 이미지들에서의 모발의 색을 변경하도록 처리된다. 애플리케이션은 신경망 모델(814)에 의해 제공되는 심층 신경망을 이용하여 이미지 처리를 수행한다. 네트워크 모델은 도 3, 도 4 및 도 5에 도시된 모델들 중 임의의 것으로서 구성될 수 있다.

[0123] 애플리케이션(816)은 이미지의 하나 이상의 속성을 변경하기 위해 색 데이터(830)와 같은 특정 속성 데이터와 연관될 수 있다. 속성들을 변경하는 것은 픽셀 값들을 변경하여 새로운 이미지를 생성하는 것에 관한 것이다. 이미지 관련 데이터(예를 들어, 이미지들을 저장, 인쇄 및/또는 디스플레이하기 위한 것)가 다양한 색 모델들 및 데이터 포맷들을 사용하여 표현될 수 있고 애플리케이션(816)이 그에 따라 구성될 수 있다는 것이 이해된다. 다른 예들에서, 속성 데이터는 조명 조건들, 텍스처, 형상 등과 같은 효과를 변경하는 것과 관련될 수 있다. 애플리케이션(816)은, 예를 들어, 원하는 위치에서의 이미지(예를 들어, 심층 신경망에 의해 식별되는 이미지에서의 관심 있는 객체 또는 그의 부분)에 효과를 적용하기 위해 속성(들)(도시되지 않음)을 변경하기 위한 하나 이상의 기능으로 구성될 수 있다.

[0124] 저장 디바이스(들)(212)는 운영 체제(832) 및 통신 모듈들을 포함하는 다른 모듈들(도시되지 않음)과 같은 추가 모듈들을 저장할 수 있다; 그래픽 처리 모듈들(예를 들어, 프로세서들(802)의 GPU에 대한 것); 맵 모듈; 연락처 모듈; 캘린더 모듈; 사진/갤러리 모듈; 사진(이미지/미디어) 편집기; 미디어 플레이어 및/또는 스트리밍 모듈; 소셜 미디어 애플리케이션들; 브라우저 모듈 등을 포함하는 다른 모듈들(도시되지 않음)과 같은 추가 모듈들을 저장할 수 있다. 저장 디바이스들은 본 명세서에서 저장 유닛들로서 참조될 수 있다.

[0125] 통신 채널들(838)은, 통신가능하게, 물리적으로, 및/또는 동작 가능하든 간에, 컴포넌트 간 통신을 위해 컴포넌트들(802, 804, 806, 808, 810, 812) 및 임의의 모듈들(814, 816 및 826) 각각을 결합할 수 있다. 일부 예들에서, 통신 채널들(838)은 시스템 버스, 네트워크 접속, 프로세스간 통신 데이터 구조, 또는 데이터를 통신하기 위한 임의의 다른 방법을 포함할 수 있다.

[0126] 하나 이상의 프로세서(802)는 기능을 구현하고 및/또는 컴퓨팅 디바이스(800) 내에서 명령어들을 실행할 수 있다. 예를 들어, 프로세서들(802)은 무엇보다도(예를 들어, 운영 체제, 애플리케이션들 등) 도 8에 도시된 모듈

들의 기능성을 실행하기 위해 저장 디바이스들(812)로부터 명령어들 및/또는 데이터를 수신하도록 구성될 수 있다. 컴퓨팅 디바이스(800)는 저장 디바이스들(812)에 데이터/정보를 저장할 수 있다. 기능 중 일부가 이하 본 명세서에서 더 설명된다. 하나의 모듈이 다른 모듈의 기능을 지원할 수 있는 식으로, 동작들은 도 8의 모듈들(814, 816 및 826) 내에 정확히 속하지 않을 수 있다는 것이 이해된다.

[0127] 동작들을 수행하기 위한 컴퓨터 프로그램 코드는 하나 이상의 프로그래밍 언어, 예를 들어, Java, Smalltalk, C++ 등과 같은 객체 지향 프로그래밍 언어, 또는 "C" 프로그래밍 언어 또는 유사한 프로그래밍 언어들과 같은 종래의 절차적 프로그래밍 언어의 임의의 조합으로 작성될 수 있다.

[0128] 컴퓨팅 디바이스(800)는 제스처 기반 I/O 디바이스(806)의 스크린 상에 디스플레이하기 위해, 또는 일부 예들에서 프로젝터, 모니터 또는 다른 디스플레이 디바이스에 의한 디스플레이를 위해 출력을 생성할 수 있다. 제스처 기반 I/O 디바이스(806)는 다양한 기술들을 사용하여 구성될 수 있다(예를 들어, 입력 능력에 관해: 저항성 터치스크린, 표면 탄성파 터치스크린, 용량성 터치스크린, 투영 용량성 터치스크린, 압력 감지 스크린, 음향 펄스 인식 터치스크린, 또는 또 다른 존재 감지 스크린 기술; 및 출력 능력과 관련하여: 액정 디스플레이(LCD), 발광 다이오드(LED) 디스플레이, 유기 발광 다이오드(OLED) 디스플레이, 도트 매트릭스 디스플레이, e-잉크, 또는 유사한 단색 또는 컬러 디스플레이).

[0129] 본 명세서에 설명된 예들에서, 제스처 기반 I/O 디바이스(806)는 터치스크린과 상호작용하는 사용자로부터의 촉각 상호작용 또는 제스처들을 입력으로서 수신할 수 있는 터치스크린 디바이스를 포함한다. 이러한 제스처들은 탭 제스처들, 드래깅 또는 스와이핑 제스처들, 플리킹 제스처들, 제스처들을 일시중지하는 것(예를 들어, 사용자가 적어도 임계 시간 기간 동안 스크린의 동일한 위치를 터치하는 경우)을 포함할 수 있으며, 여기서 사용자는 제스처 기반 I/O 디바이스(806)의 하나 이상의 위치를 터치하거나 가리킨다. 제스처 기반 I/O 디바이스(806)는 또한 비 탭 제스처들을 포함할 수 있다. 제스처 기반 I/O 디바이스(806)는 그래픽 사용자 인터페이스와 같은 정보를 사용자에게 출력하거나 디스플레이할 수 있다. 제스처 기반 I/O 디바이스(806)는, 예를 들어, 무엇보다도 이미지들을 보고, 이미지들을 처리하고 새로운 이미지들을 디스플레이하기 위한 애플리케이션(818), 메시징 애플리케이션들, 전화 통신들, 연락처 및 캘린더 애플리케이션들, 웹 브라우징 애플리케이션들, 게임 애플리케이션들, 전자책 애플리케이션들 및 금융, 지불 및 다른 애플리케이션들 또는 기능들을 포함하는, 컴퓨팅 디바이스(800)의 다양한 애플리케이션들, 기능들 및 능력들을 제시할 수 있다.

[0130] 본 개시내용이 주로 I/O 능력을 갖는 디스플레이 스크린 디바이스(예를 들어, 터치스크린)의 형태인 제스처 기반 I/O 디바이스(806)를 예시하고 논의하지만, 제스처 기반 I/O 디바이스들의 다른 예들이 이용될 수 있으며, 이것은 움직임을 검출할 수 있고 스크린 자체를 포함하지 않을 수 있다. 이러한 경우에, 컴퓨팅 디바이스(800)는 디스플레이 스크린을 포함하거나 또는 새로운 이미지들을 제시하기 위해 디스플레이 디바이스에 결합된다. 컴퓨팅 디바이스(800)는 트랙 패드/터치 패드, 하나 이상의 카메라, 또는 또 다른 존재 또는 제스처 감지 입력 디바이스로부터 제스처 기반 입력을 수신할 수 있고, 여기서 존재는, 예를 들어, 사용자의 전부 또는 일부의 모션을 포함하는 사용자의 양태들을 의미한다.

[0131] 하나 이상의 통신 유닛(808)은, 예를 들어 새로운 속성 데이터 또는 애플리케이션 기능을 수신하기 위해, 하나 이상의 네트워크 상에서 네트워크 신호들을 송신 및/또는 수신하는 것에 의해 하나 이상의 통신 네트워크(도시되지 않음)를 통해 또 다른 컴퓨팅 디바이스, 인쇄 디바이스 또는 디스플레이 디바이스(모두 도시되지 않음)와 새로운 이미지들을 공유하기 위해 외부 디바이스들(도시되지 않음)과 통신할 수 있다. 통신 유닛들은 무선 및/또는 유선 통신을 위한 다양한 안테나들 및/또는 네트워크 인터페이스 카드들, 칩들(예를 들어, GPS(Global Positioning Satellite)) 등을 포함할 수 있다.

[0132] 입력 디바이스들(804) 및 출력 디바이스들(810)은 하나 이상의 버튼, 스위치, 포인팅 디바이스, 카메라, 키보드, 마이크로폰, 하나 이상의 센서(예를 들어, 생체인식 등), 스피커, 벨, 하나 이상의 라이트, 햅틱(진동) 디바이스 등 중 임의의 것을 포함할 수 있다. 동일한 것 중 하나 이상이 범용 직렬 버스(USB) 또는 다른 통신 채널(예를 들어, 838)을 통해 결합될 수 있다. 카메라(입력 디바이스(804))는 사용자가 "셀피"를 취하기 위해 제스처 기반 I/O 디바이스(806)를 보면서 카메라를 사용하여 이미지(들)를 캡처하는 것을 허용하기 위해 정면 배향으로 (즉, 동일한 측 상으로) 될 수 있다.

[0133] 하나 이상의 저장 디바이스(812)는, 예를 들어, 단기 메모리 또는 장기 메모리로서 상이한 형태들 및/또는 구성들을 취할 수 있다. 저장 디바이스들(812)은, 전력이 제거될 때 저장된 콘텐츠를 유지하지 않는 휘발성 메모리로서 정보의 단기 저장을 위해 구성될 수 있다. 휘발성 메모리 예들은 랜덤 액세스 메모리(RAM), 동적 랜덤 액세스 메모리(DRAM), 정적 랜덤 액세스 메모리(SRAM) 등을 포함한다. 저장 디바이스들(812)은, 일부 예들에서,

예를 들어, 휘발성 메모리보다 더 많은 양의 정보를 저장하고 및/또는 장기 정보를 저장하여 전력이 제거될 때 정보를 유지하는 하나 이상의 컴퓨터 판독가능 저장 매체를 또한 포함한다. 비휘발성 메모리 예들은 자기 하드 디스크들, 광학 디스크들, 플로피 디스크들, 플래시 메모리들, 또는 EPROM(electrically programmable memory) 또는 EEPROM(electrically erasable and programmable) 메모리의 형태들을 포함한다.

[0134] 도시되지는 않았지만, 컴퓨팅 디바이스는 예를 들어, 적절한 훈련 및/또는 테스팅 데이터와 함께 도 5에 도시된 바와 같은 네트워크를 사용하여 신경망 모델(814)을 훈련시키기 위한 훈련 환경으로서 구성될 수 있다.

[0135] 컴퓨팅 디바이스(800)는 처리 유닛 및 처리 유닛에 결합된 저장 유닛을 포함한다. 저장 유닛은 명령어들을 저장하고, 이 명령어들은, 처리 유닛에 의해 실행될 때, 컴퓨팅 디바이스가, 픽셀들 각각이 관심 대상 객체의 멤버인지(예를 들어, 모발 픽셀인지)를 결정하기 위해 이미지의 픽셀들을 분류하고; 및 관심 대상 객체의 멤버인 픽셀들의 하나 이상의 속성을 변경함으로써 새로운 이미지(예를 들어, 채색된 모발 이미지)를 정의 및 제시하도록 구성된 딥 러닝 신경망 모델(예를 들어, 컨볼루션 신경망을 포함함)을 저장하고 제공하도록 구성한다. 변경 동작들은 심층 신경망 모델을 이용하여 이미지로부터 정의된 마스크를 사용한다. 일 예에서, 관심 대상 객체는 모발이고 속성은 색이다. 따라서, 하나 이상의 속성을 변경하는 것은 모발 세그먼테이션 마스크를 이용하여 이미지에서의 모발 픽셀들에 새로운 모발 색을 적용하는 것이다.

[0136] 심층 신경망은 랩톱, 데스크톱, 워크스테이션, 서버 또는 다른 비슷한 생성 컴퓨팅 디바이스와 같은 "더 큰" 디바이스보다 더 적은 처리 자원들을 갖는 모바일 디바이스(예를 들어, 스마트폰 또는 태블릿)인 컴퓨팅 디바이스를 위한 경량 아키텍처에 적응된다.

[0137] 심층 신경망 모델은 개별 표준 컨볼루션들이 뎁스와이즈 컨볼루션 및 포인트 컨볼루션으로 인수분해되는 컨볼루션들을 포함하는 뎁스와이즈 분리 가능 컨볼루션 신경망으로서 구성될 수 있다. 뎁스와이즈 컨볼루션은 각각의 입력 채널에 단일 필터를 적용하는 것으로 제한되고, 포인트 컨볼루션은 뎁스와이즈 컨볼루션의 출력들을 조합하는 것으로 제한된다.

[0138] 심층 신경망 모델은, 업샘플링할 때와 같이, 인코더에서의 계층들과 디코더에서의 대응하는 계층들 사이에서, 스킵 연결들을 수행하는 동작들을 포함하도록 추가로 구성될 수 있다.

[0139] 심층 신경망 모델은 마스크 이미지 그래디언트 일관성 손실 측도를 이용하여 훈련될 수 있고, 그에 의해, 마스크 이미지 그래디언트 일관성 손실이 처리되는 이미지에 대해 결정되고, 생성된 마스크 및 사용된 손실 측도는 모델을 훈련시킨다. 마스크 이미지 그래디언트 일관성 손실은 수학식 1에 따라 결정될 수 있다.

[0140] 도 9는 일 예에 따른 컴퓨팅 디바이스(800)의 동작들(900)의 흐름도이다. 동작들(900)은 사용자의 모발을 포함하는 사용자의 머리의 이미지들을 정의하는 비디오 프레임들을 포함하는 비디오를 포함하는 셀피를 취하기 위해 애플리케이션(816)과 같은 애플리케이션을 사용하고, 새로운 모발 색을 선택하고, 및 사용자의 모발의 모발 색이 새로운 모발 색인 비디오를 시청하는 컴퓨팅 디바이스(800)의 사용자에 관한 것이다. (902)에서, 동작들은 GUI(818) 및 이미지 획득 컴포넌트(822)를 호출할 수 있는 카메라와 같은 입력 디바이스 또는 유닛을 통해 이미지를 수신한다. 이미지는 제스처 기반 I/O 디바이스(806) 또는 또 다른 디스플레이 디바이스와 같은 디스플레이(904)를 위해 제공된다. 모바일 디바이스 상에서의 모발에 대한 처리를 돕기 위해, 이미지(데이터)가 전처리될 수 있다(도시되지 않음). 이미지들은 얼굴을 검출하고 전형적인 셀피들에 대해 예상되는 스케일에 기초하여 그것 주위의 영역을 크로핑함으로써 전처리될 수 있다.

[0141] (906)에서, 제스처 기반 I/O 디바이스(806) 및 GUI(818)(대화형 GUI)를 통해 입력이 수신되어 새로운 모발 색을 선택한다. GUI(818)는 선택용 대화형 인터페이스를 통해 모발 색 데이터(814)를 제시하도록 구성될 수 있다. 일부 예들에서, 애플리케이션(816)은 색을 제안하도록 구성될 수 있다. 도시되지는 않았지만, 동작들은 수신된 이미지 및 선택적으로 다른 색(예를 들어, 피부색) 또는 광 정보 등으로부터 기존의 또는 현재의 모발 색을 결정하는 것을 포함할 수 있다. 데이터(도시되지 않음)로서 표현되는 사용자 선호도들은 GUI(818)를 통해 요청될 수 있다. 동작들은 색 예측 컴포넌트(820)에 동일한 것을 제공하는 것을 추가로 포함할 수 있다. 색 예측 컴포넌트(820)는 기존의 모발 색, 피부 또는 다른 색 및/또는 광 정보, 사용자 선호들, 추세 등 중 하나 이상에 응답하여 적절한 색(예를 들어, 색 데이터(830)로부터의 새로운 모발 색들에 대한 하나 이상의 후보)를 제안하는 기능을 가질 수 있다.

[0142] (908)에서, 동작들은, 본 명세서에 기술되는 바와 같은 심층 신경망 모델(814)에 의해 식별되는 제2 이미지에서의 모발의 픽셀들에 새로운 속성, 즉 새로운 모발 색을 적용하기 위한 처리를 위해 제2 이미지를 수신한다. 카메라가 계속해서 이미지들을 캡처함에 따라, 기존의 모발 색을 정의하기 위해 사용되는 제1 이미지는 더 이상

현재의 것이 아니다.

[0143] (910)에서, 동작들은 모델(814)을 사용하여 (제2) 이미지에서의 모발 픽셀들을 식별하는 모발 세그먼테이션 마스크를 정의한다. (912)에서, 동작들은 모발 세그먼테이션 마스크를 이용하여 이미지에서의 모발 픽셀들에 새로운 모발 색을 적용함으로써 새로운 이미지(예를 들어, 채색된 모발 이미지)를 정의한다. (914)에서, 동작들은 GUI 컴포넌트(818)에 의해 제공되는 GUI에서 제스처 기반 I/O 디바이스(806)에 출력하기 위한 새로운 이미지를 제공한다. 추가의 이미지들이 카메라로부터 수신됨에 따라, 추가의 제각기 마스크들이 정의되고, 채색된 모발을 갖는 추가의 제각기 새로운 이미지가 정의되고 제시된다.

[0144] 다른 속성 변경들, 기존의 및 새로운 모발 색들 또는 2개의 새로운 모발 색들의 라이브 비교들을 촉진하기 위해 또는 새로운 모발 색을 보여주는 정지 이미지들 또는 비디오 세그먼트들을 저장하기 위해 추가적인 또는 대안적인 GUI들 또는 GUI 기능들이 제공될 수 있다. 동작들은 GUI를 제스처 기반 I/O 디바이스를 통해 제시할 수 있고, 여기서 GUI는, 분할된 스크린 배열에서 그런 것처럼, 이미지를 보기 위한 제1 부분 및 채색된 모발 이미지를 동시에 볼 수 있는 제2 부분을 포함한다.

[0145] 동작들은 상이한 조명 조건에서 모발을 보여주기 위해 조명 조건들 처리를 모발 색(기존의 또는 새로운 색)에 적용할 수 있다. 동작들은 제각기 새로운 이미지들에서 제1 새로운 모발 색 및 제2 새로운 모발 색을 보여주도록 구성될 수 있다. 단일 마스크가 정의되고 2개의 새로운 모발 색을 적용하기 위해 2개의 별도의 채색 동작에 제공될 수 있다. 제1 및 제2 색들에 대한 제각기 새로운 색 이미지들이 순차적으로 또는 동시에 제공될 수 있다.

[0146] 논의된 임의의 GUI 인터페이스 옵션들 또는 제어들에 부가적으로 또는 대안적으로, 음성 활성화 제어들이 제공될 수 있다.

[0147] 다른 광 아키텍처들이 인코더 및 디코더의 대응하는 계층들 사이의 스킵 연결들을 이용하여 모발 세그먼테이션 마스크를 생성하기 위해 유사한 방식으로 적응될 수 있고 마스크 이미지 그래디언트 일관성 손실 함수를 이용하여 훈련될 수 있다. 이러한 아키텍처의 일 예는, 장(Zhang) 등 및 Megvii Technology 사의 포인트와이즈 그룹 컨볼루션 및 채널 셔플을 사용하여 매우 제한된 계산 능력(예를 들어, 10-150 MFLOPs)을 갖는 모바일을 위해 특별히 설계된 계산 효율적 CNN인 ShuffleNet™이다. ShuffleNet에 관한 상세 사항들은 본 명세서에 참조에 의해 포함되는, ShuffleNet: An Extremely Efficient Convolutional Neural Network for Moble Devices of Zhang et al. arXiv:1707.01083v2 [cs:CV] 7 Dec. 2017에 제공된다.

[0148] 도 10은 본 명세서의 교시에 따라 예시적 CNN을 정의하는 단계들(1000)의 흐름도이다. (1002)에서, 이미지 분류에 대해 사전 훈련되고, 제한된 계산 능력을 갖는 컴퓨팅 디바이스 상에서 실행되도록 구성된 CNN이 획득된다(예를 들어, MobileNet, ShuffleNet 등).

[0149] (1004)에서, 단계는 CNN을 적응시켜, 객체 세그먼테이션 마스크를 정의하고, 예를 들어, 완전 연결된 계층들을 제거하고, 업샘플링 동작들을 정의하는 등을 한다. (1006)에서, 단계는 CNN을 적응시켜 인코더 스테이지에서의 계층 등과 디코더 스테이지에서의 대응하는 계층들 간의 스킵 연결들을 사용하여, 디코더 스테이지에서 업샘플링할 때 저 해상도이지만 강한 피처들과 고 해상도이지만 약한 피처들을 조합한다. (1008)에서, 객체 세그먼테이션을 위한 라벨링된 데이터를 포함하는 세그먼테이션 훈련 데이터를 획득하기 위한 단계가 수행된다. 언급된 바와 같이, 이는 크라우드 소싱되고 그리고 객체 세그먼테이션 마스크(라벨링)가 미세하지 않은 잡음이 있고 거친 세그먼테이션 데이터일 수 있다.

[0150] 단계(1010)는 훈련할 때 최소화하기 위해 마스크 이미지 그래디언트 일관성 손실 함수를 파라미터로서 정의하기 위해 수행된다. (1012)에서, 단계가 수행되어, 세그먼테이션 훈련 데이터 및 마스크 이미지 그래디언트 일관성 손실 함수를 사용하여 적응된 대로의 사전 훈련된 CNN을 추가로 훈련시켜서 마스크 이미지 그래디언트 일관성 손실을 최소화하여 추가 훈련된 CNN을 생성한다. (1014)에서, 추가 훈련된 CNN은 객체 세그먼테이션을 위해 라벨링된 데이터를 포함하는 세그먼테이션 테스팅 데이터로 테스트된다. 도 10의 단계들 중 일부는 선택적이라는 것이 명백할 것이다. 제한이 아닌 예로서, 스킵 연결들을 사용하도록 적응시키는 것은 선택적일 수 있다. 그래디언트 일관성 손실을 최소화하도록 적응시키는 것(및 따라서 이러한 손실에 대해 유사하게 훈련하는 것)은 선택적일 수 있다. 훈련 데이터는 크라우드 소스 데이터로부터 획득되는 것과 같이 잡음이 있고 거친 세그먼테이션(예를 들어, 미세한 라벨링 없음)일 수 있다.

[0151] 이와 같이 훈련된 CNN은 설명된 바와 같이 모바일 디바이스 상에 저장하고 그 상에서 사용하기 위해 제공될 수

있다.

[0152] 컴퓨팅 디바이스 양태들에 부가하여, 통상의 기술자는 컴퓨터 프로그램 제품 양태들이 개시된 것을 이해할 것인데, 여기서, 명령어들은 컴퓨팅 디바이스를 구성하여 본 명세서에 저장된 방법 양태들 중 임의의 것을 수행하도록 비일시적 저장 디바이스(예를 들어, 메모리, CD-ROM, DVD-ROM, 디스크 등)에 저장된다.

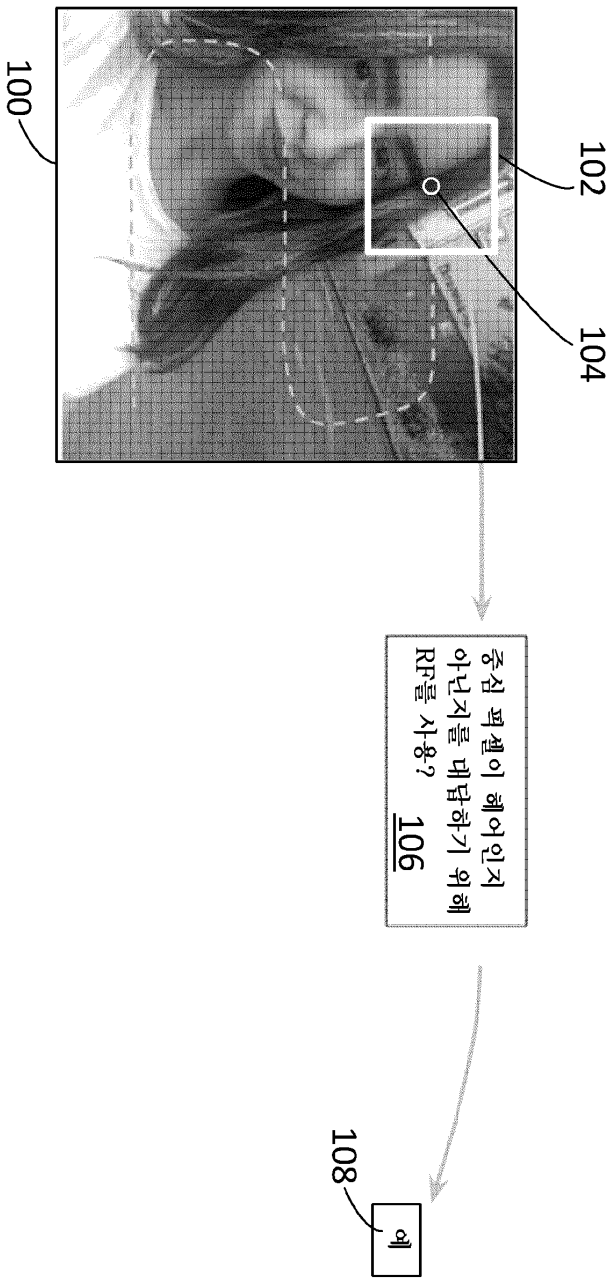
[0153] 실제 구현은 본 명세서에 설명된 특징들 중 임의의 것 또는 전부를 포함할 수 있다. 이들 및 다른 양태들, 특징들 및 다양한 조합들은 방법들, 장치들, 시스템들, 기능들을 수행하기 위한 수단, 프로그램 제품들, 및 본 명세서에 설명된 특징들을 조합하여 다른 방식들로 표현될 수 있다. 많은 실시예들이 설명되었다. 그럼에도 불구하고, 본 명세서에 설명된 프로세스들 및 기술들의 사상 및 범위로부터 벗어나지 않고 다양한 수정들이 이루어질 수 있다는 것을 이해할 것이다. 또한, 다른 단계들이 제공될 수 있고, 또는 단계들이 설명된 프로세스로부터 제거될 수 있고, 다른 컴포넌트들이 설명된 시스템들에 부가 또는 그로부터 제거될 수 있다. 따라서, 다른 실시예들은 이하의 청구항들의 범위 내에 있다.

[0154] 본 명세서의 설명 및 청구범위 전체에 걸쳐, "포함한다" 및 "포함하다" 및 이들의 변형들은 "포함하지만 이들로 제한되지는 않음"을 의미하고, 이들은 다른 컴포넌트들, 정수들 또는 단계들을 배제하도록 의도되지 않는다(그리고 그렇지 않다). 본 명세서 전체에 걸쳐, 단수는 문맥이 달리 요구하지 않는 한 복수를 포함한다. 특히, 부정 관사가 사용되는 경우, 본 명세서는 문맥이 달리 요구하지 않는 한, 복수뿐만 아니라 단수를 고려하는 것으로 이해되어야 한다.

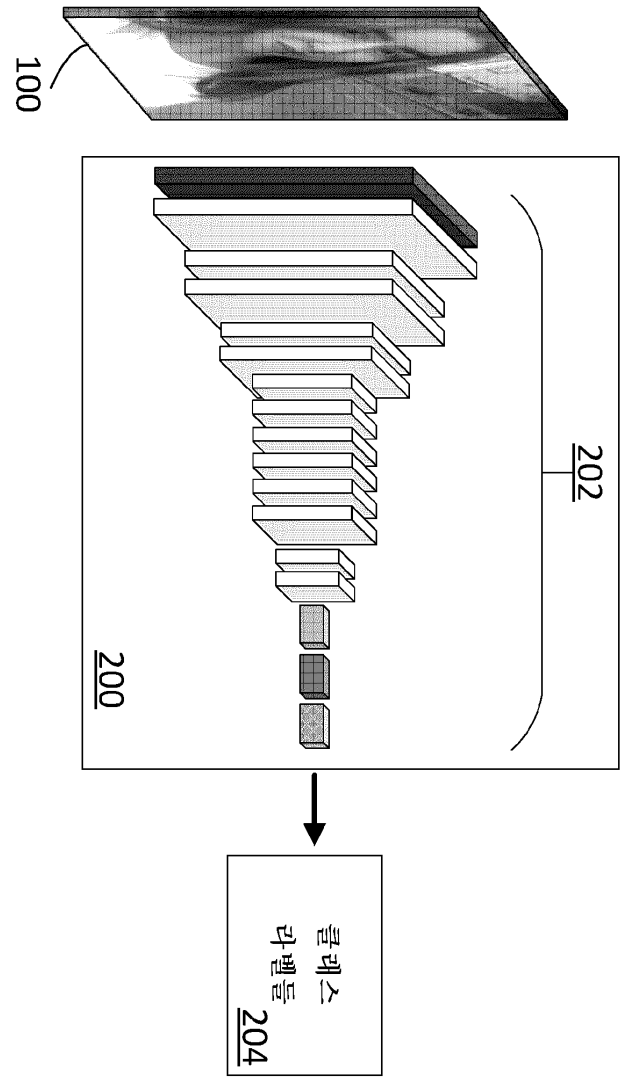
[0155] 본 발명의 특정 양태, 실시예 또는 예와 관련하여 설명된 특징들, 정수 특성들, 화합물들, 화학적 반족들(moieties) 또는 그룹들은 그와 양립하지 못하는 것이 아닌 한, 임의의 다른 양태, 실시예 또는 예에 적용가능한 것으로 이해되어야 한다. 본 명세서에 개시된 모든 특징들(임의의 첨부된 청구항들, 요약서 및 도면들을 포함함), 및/또는 이와 같이 개시된 임의의 방법 또는 프로세스의 모든 단계들은, 이러한 특징들 및/또는 단계들 중 적어도 일부가 상호 배타적인 조합들을 제외하고는, 임의의 조합으로 조합될 수 있다. 본 발명은 임의의 전술한 예들 또는 실시예들의 상세 사항들로만 제한되지는 않는다. 본 발명은 본 명세서(임의의 첨부된 청구항, 요약서 및 도면을 포함함)에 개시된 특징들의 임의의 신규한 하나 또는 임의의 신규한 조합, 또는 개시된 임의의 방법 또는 프로세스의 단계들 중 임의의 신규한 하나 또는 임의의 신규한 조합으로 확장된다.

도면

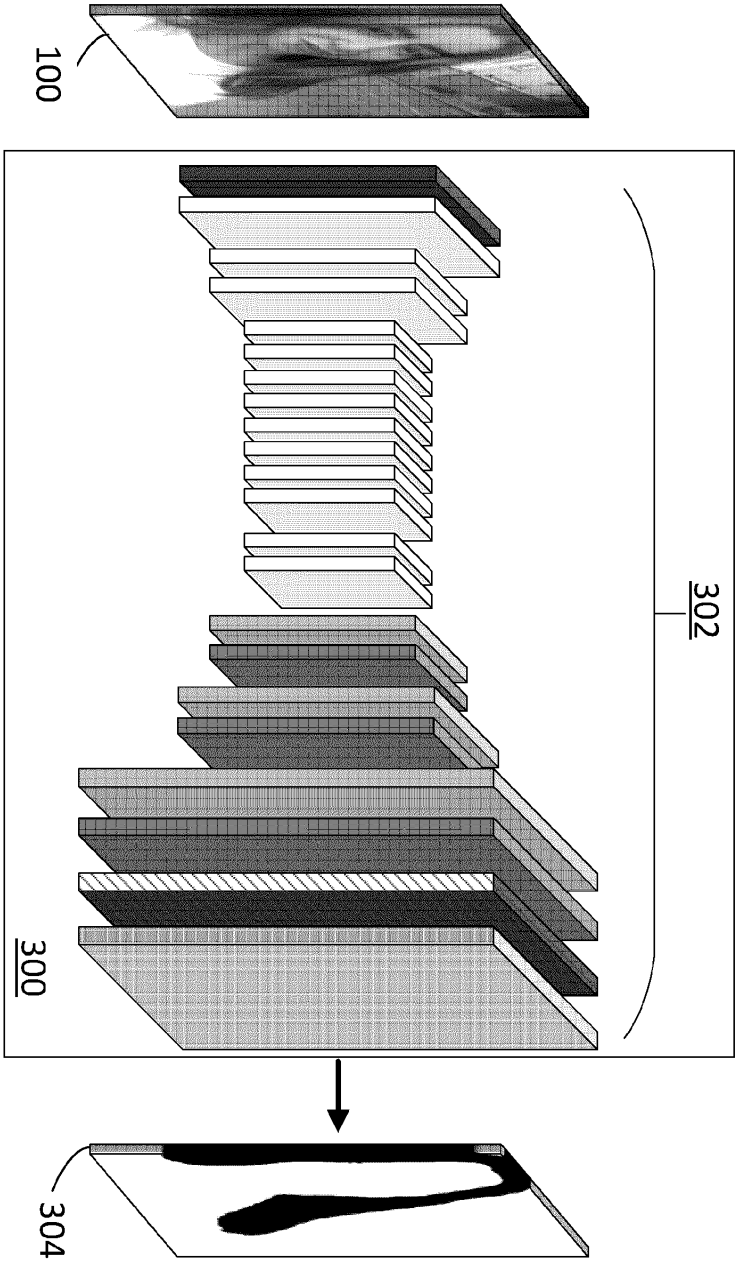
도면1



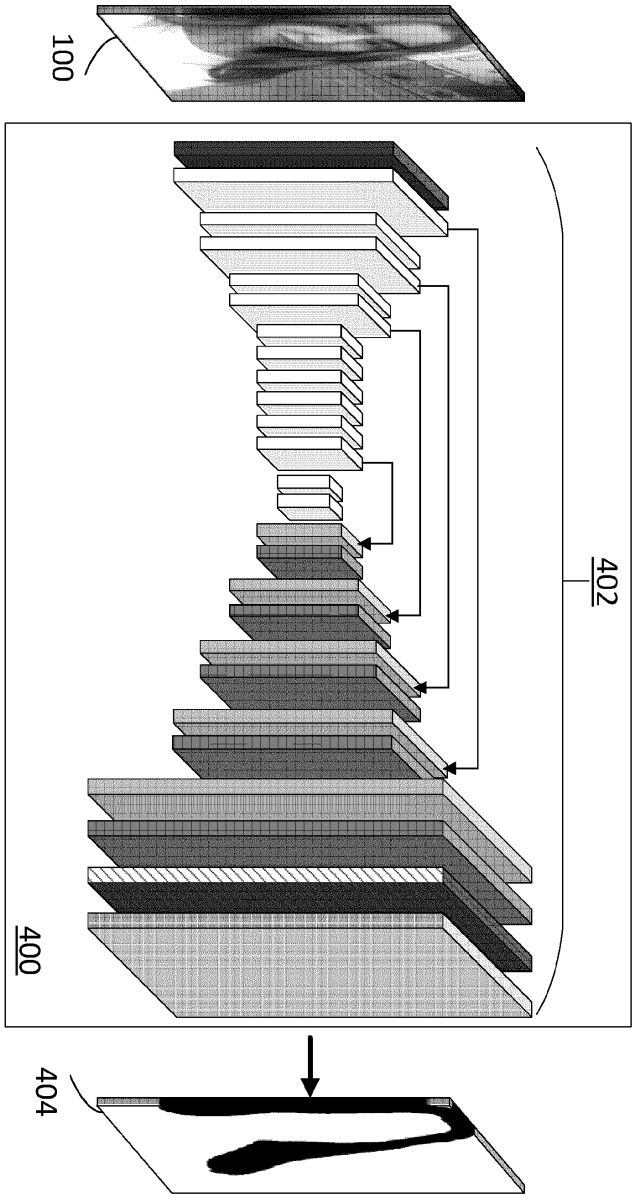
도면2



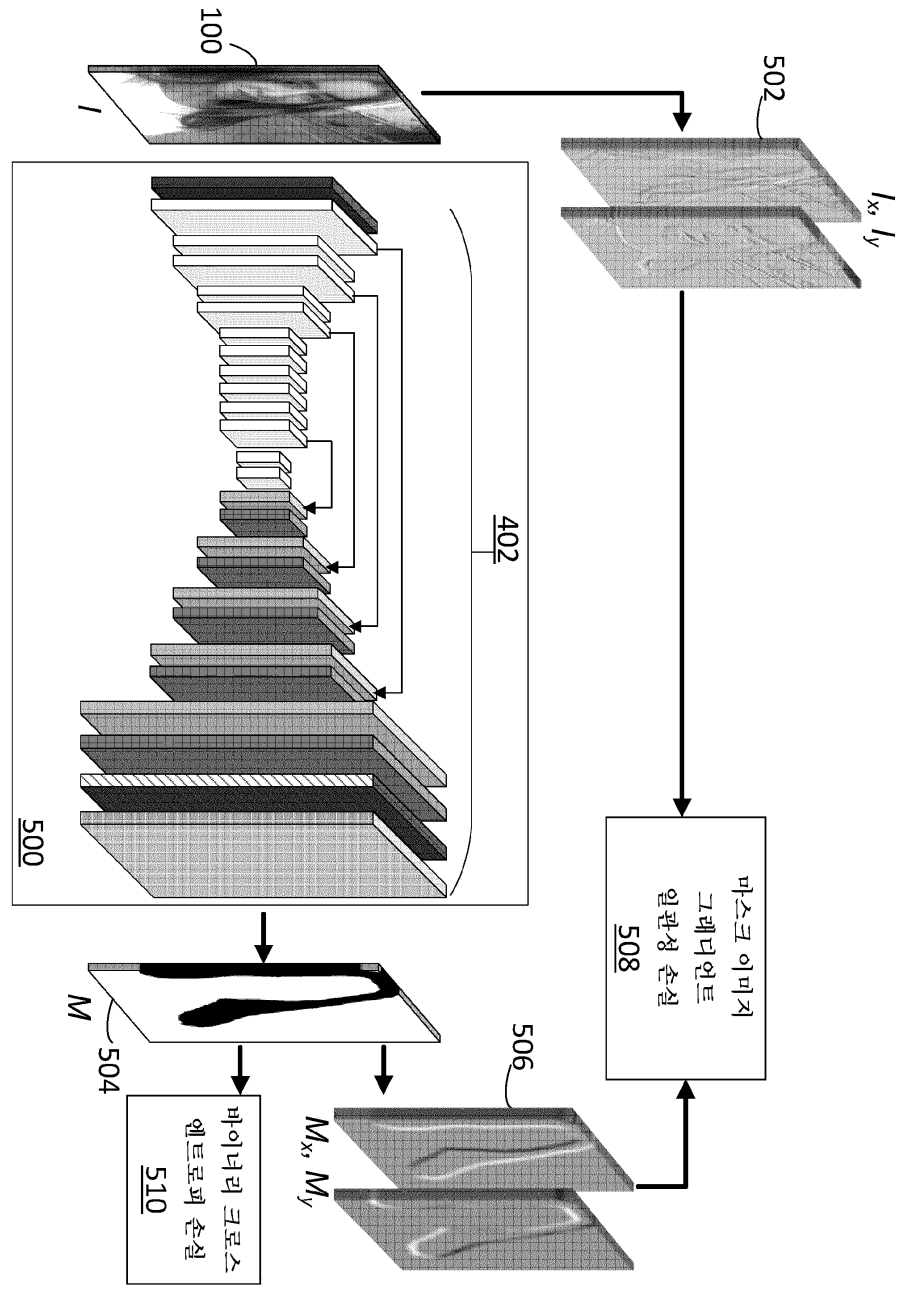
도면3



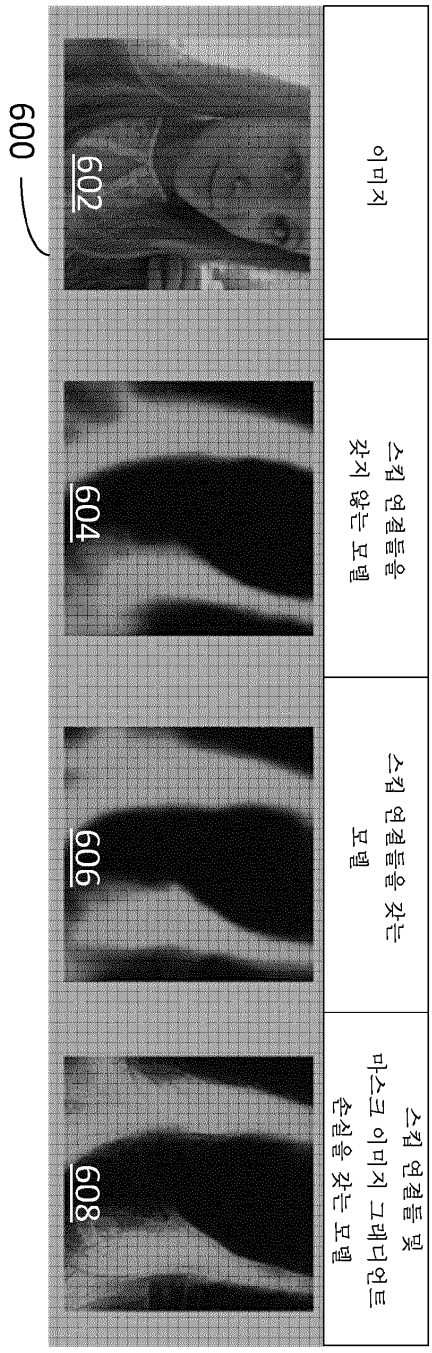
도면4



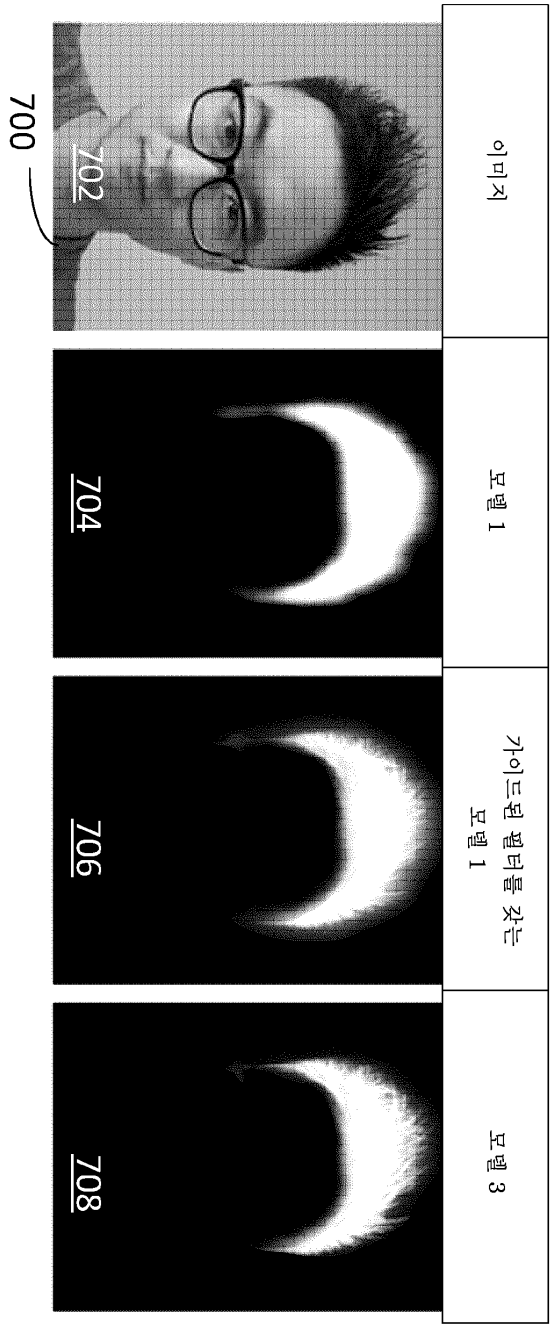
도면5


100
I
502
I_x, I_y
500
402
마스크 이미지
그래디언트
일관성 손실
508
M
504
506
M_x, M_y
바이너리 크로스
엔트로피 손실
510

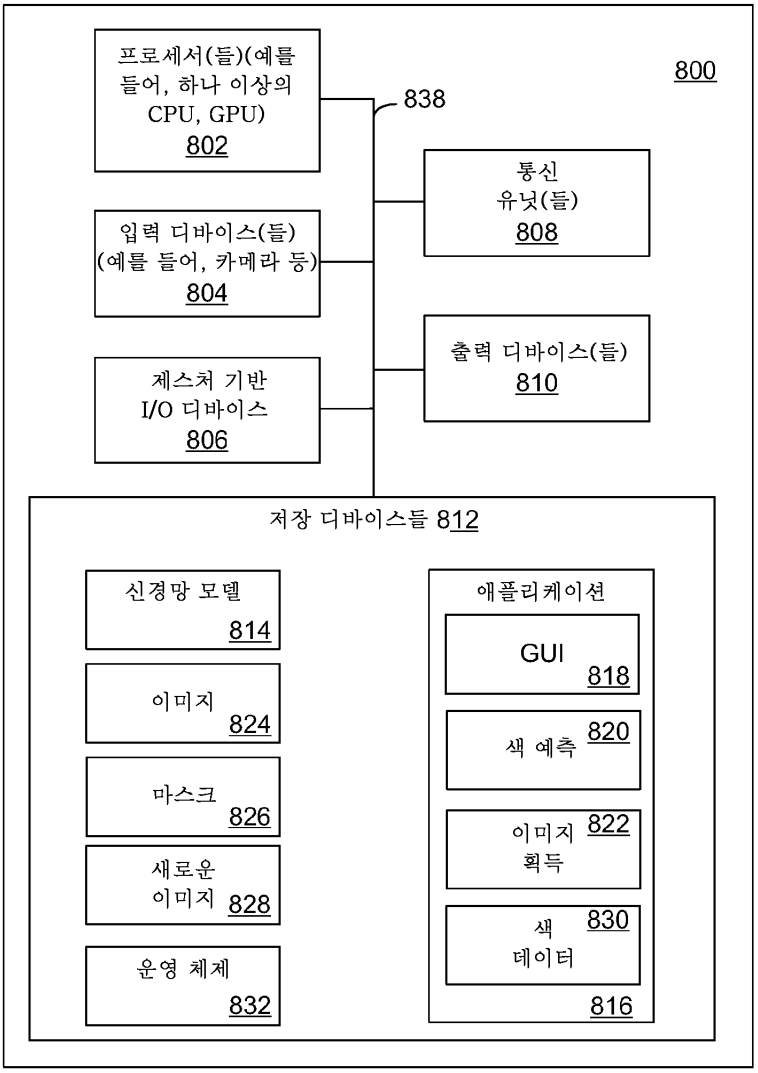
도면6


600
602
604
606
608
이미지
스킵 연결들을 갖지 않는 모델
스킵 연결들을 갖는 모델
스킵 연결들 및 마스크 이미지 그레디언트 손실을 갖는 모델

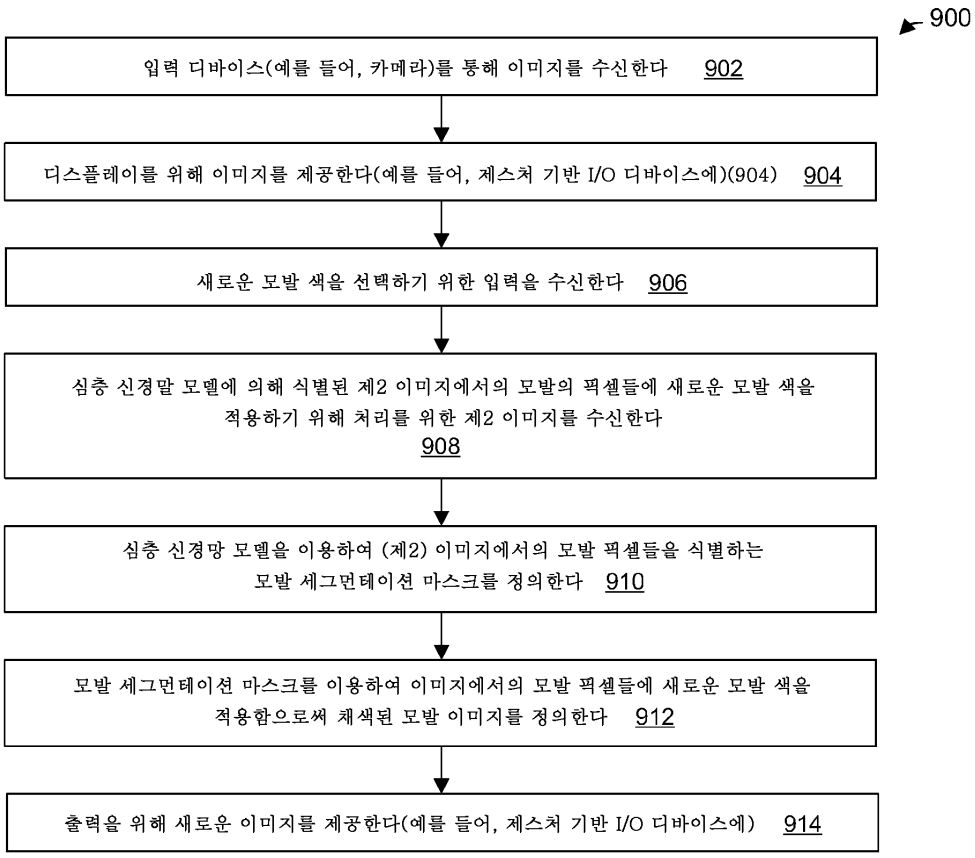
도면7



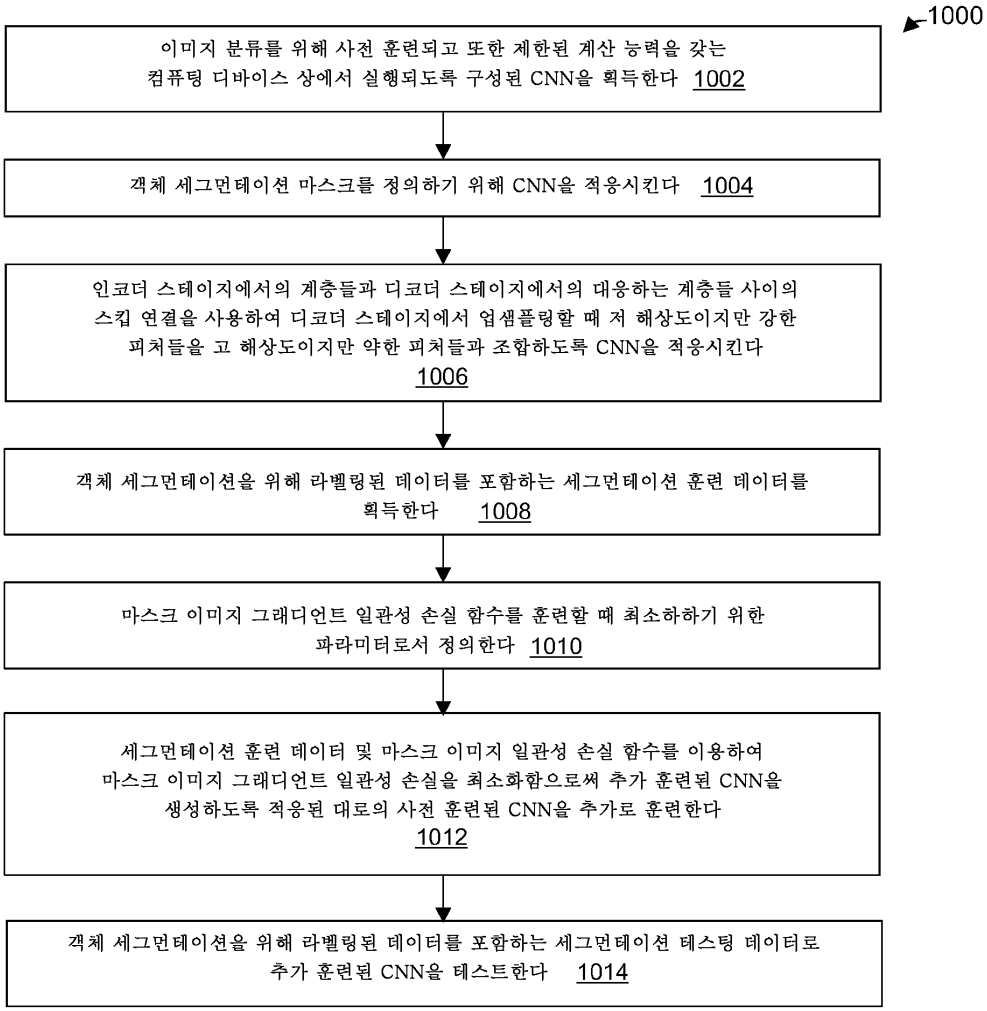
도면8


800
프로세서(들)(예를 들어, 하나 이상의 CPU, GPU)
802
838
통신 유닛(들)
808
입력 디바이스(들)(예를 들어, 카메라 등)
804
출력 디바이스(들)
810
제스처 기반 I/O 디바이스
806
저장 디바이스들 812
신경망 모델
814
이미지
824
마스크
826
새로운 이미지
828
운영 체제
832
애플리케이션
GUI
818
색 예측
820
이미지 획득
822
색 데이터
830
816

도면9


900
입력 디바이스(예를 들어, 카메라)를 통해 이미지를 수신한다 902
디스플레이를 위해 이미지를 제공한다(예를 들어, 제스처 기반 I/O 디바이스에)(904) 904
새로운 모발 색을 선택하기 위한 입력을 수신한다 906
심층 신경망 모델에 의해 식별된 제2 이미지에서의 모발의 픽셀들에 새로운 모발 색을 적용하기 위해 처리를 위한 제2 이미지를 수신한다 908
심층 신경망 모델을 이용하여 (제2) 이미지에서의 모발 픽셀들을 식별하는 모발 세그먼테이션 마스크를 정의한다 910
모발 세그먼테이션 마스크를 이용하여 이미지에서의 모발 픽셀들에 새로운 모발 색을 적용함으로써 채색된 모발 이미지를 정의한다 912
출력을 위해 새로운 이미지를 제공한다(예를 들어, 제스처 기반 I/O 디바이스에) 914

도면10


1000
이미지 분류를 위해 사전 훈련되고 또한 제한된 계산 능력을 갖는 컴퓨팅 디바이스 상에서 실행되도록 구성된 CNN을 획득한다 1002
객체 세그먼테이션 마스크를 정의하기 위해 CNN을 적응시킨다 1004
인코더 스테이지에서의 계층들과 디코더 스테이지에서의 대응하는 계층들 사이의 스킵 연결을 사용하여 디코더 스테이지에서 업샘플링할 때 저 해상도이지만 강한 피처들을 고 해상도이지만 약한 피처들과 조합하도록 CNN을 적응시킨다 1006
객체 세그먼테이션을 위해 라벨링된 데이터를 포함하는 세그먼테이션 훈련 데이터를 획득한다 1008
마스크 이미지 그래디언트 일관성 손실 함수를 훈련할 때 최소하하기 위한 파라미터로서 정의한다 1010
세그먼테이션 훈련 데이터 및 마스크 이미지 일관성 손실 함수를 이용하여 마스크 이미지 그래디언트 일관성 손실을 최소화함으로써 추가 훈련된 CNN을 생성하도록 적응된 대로의 사전 훈련된 CNN을 추가로 훈련한다 1012
객체 세그먼테이션을 위해 라벨링된 데이터를 포함하는 세그먼테이션 테스팅 데이터로 추가 훈련된 CNN을 테스트한다 1014