



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0026996
(43) 공개일자 2021년03월10일

(51) 국제특허분류(Int. Cl.)
G06T 7/11 (2017.01) G06K 9/00 (2006.01)
G06K 9/62 (2006.01) G06N 3/04 (2006.01)
G06T 11/60 (2006.01)
(52) CPC특허분류
G06T 7/11 (2017.01)
G06K 9/00718 (2013.01)
(21) 출원번호 10-2019-0153388
(22) 출원일자 2019년11월26일
심사청구일자 2019년11월26일
(30) 우선권주장
1020190107334 2019년08월30일 대한민국(KR)

(71) 출원인
울산과학기술원
울산광역시 울주군 언양읍 유니스트길 50
주식회사 인이지
울산광역시 울주군 언양읍 유니스트길 50, 106동
701-9호
(72) 발명자
최재식
울산광역시 울주군 언양읍 유니스트길 50
알리, 토우시
울산광역시 울주군 언양읍 유니스트길 50
(74) 대리인
특허법인 무한

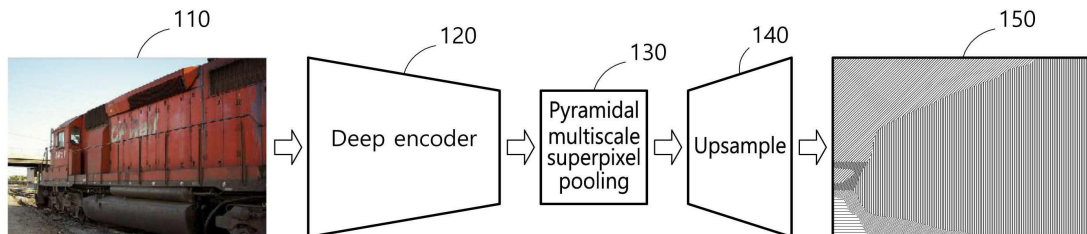
전체 청구항 수 : 총 15 항

(54) 발명의 명칭 컨텍스트 정보 추출 방법 및 장치

(57) 요약

컨텍스트 정보 추출 방법 및 장치가 개시된다. 일 실시예에 따른 컨텍스트 정보 추출 방법은 입력 영상에 관한 로컬 특징맵의 스케일에 따른 복수의 초해상맵을 획득하는 단계, 복수의 초해상맵 각각에 기초하여, 로컬 특징맵을 풀링(pooling)하여 복수의 글로벌 특징맵을 획득하는 단계, 복수의 글로벌 특징맵을 로컬 특징맵을 연결(concatenate)하는 단계 및 연결된 특징맵에 기초하여, 입력 영상의 컨텍스트 정보를 추출하는 단계를 포함한다.

대표도



(52) CPC특허분류

G06K 9/6223 (2013.01)

G06N 3/0454 (2013.01)

G06T 11/60 (2013.01)

G06T 2207/20084 (2013.01)

명세서

청구범위

청구항 1

입력 영상에 관한 로컬 특징맵의 스케일에 따른 복수의 초해상맵을 획득하는 단계;

상기 복수의 초해상맵 각각에 기초하여, 상기 로컬 특징맵을 풀링(pooling)하여 복수의 글로벌 특징맵을 획득하는 단계;

상기 복수의 글로벌 특징맵을 상기 로컬 특징맵을 연결(concatenate)하는 단계; 및

상기 연결된 특징맵에 기초하여, 상기 입력 영상의 컨텍스트 정보를 추출하는 단계를 포함하는, 컨텍스트 정보 추출 방법.

청구항 2

제1항에 있어서,

상기 글로벌 특징맵을 상기 로컬 특징맵 크기에 대응하도록 업샘플링하는 단계

를 더 포함하고,

상기 연결하는 단계는

상기 업샘플링된 글로벌 특징맵과 상기 로컬 특징맵을 연결하는 단계

를 포함하는, 컨텍스트 정보 추출 방법.

청구항 3

제2항에 있어서,

상기 업샘플링하는 단계는

상기 복수의 초해상맵에 기초하여, 상기 글로벌 특징맵을 상기 특징맵 크기에 대응하도록 업샘플링하는 단계

를 포함하는, 컨텍스트 정보 추출 방법.

청구항 4

제1항에 있어서,

상기 초해상맵을 획득하는 단계는

상기 입력 영상에 기초하여, 상기 초해상맵의 모양 및 크기를 결정하는 단계

를 포함하는, 컨텍스트 정보 추출 방법.

청구항 5

제1항에 있어서,

상기 초해상맵을 획득하는 단계는

SLIC(Simple linear iterative clustering) 알고리즘에 기초하여, 상기 초해상맵을 획득하는 단계

를 포함하는, 컨텍스트 정보 추출 방법.

청구항 6

제1항에 있어서,

1X1 컨볼루션 레이어를 이용하여, 상기 복수의 글로벌 특징맵의 채널 수를 조정하는 단계
를 더 포함하는, 컨텍스트 정보 추출 방법.

청구항 7

제1항에 있어서,

상기 복수의 글로벌 특징맵을 획득하는 단계는

상기 복수의 초해상맵 각각에 기초하여, 상기 로컬 특징맵에 평균 풀링(average pooling)하는 단계
를 포함하는, 컨텍스트 정보 추출 방법.

청구항 8

하드웨어와 결합되어 제1항 내지 제7항 중 어느 하나의 항의 방법을 실행시키기 위하여 매체에 저장된 컴퓨터 프로그램.

청구항 9

적어도 하나의 프로그램이 저장된 메모리; 및

상기 적어도 하나의 프로그램을 실행함으로써 뉴럴 네트워크를 구동하는 프로세서를 포함하고,

상기 프로세서는,

입력 영상에 관한 로컬 특징맵의 스케일에 따른 복수의 초해상맵을 획득하고,

상기 복수의 초해상맵 각각에 기초하여, 상기 로컬 특징맵을 풀링(pooling)하여 복수의 글로벌 특징맵을 획득하
고,

상기 복수의 글로벌 특징맵을 상기 로컬 특징맵을 연결(concatenate)하고,

상기 연결된 특징맵에 기초하여, 상기 입력 영상의 컨텍스트 정보를 추출하는 컨텍스트 정보 추출 장치.

청구항 10

제9항에 있어서,

상기 프로세서는

상기 글로벌 특징맵을 상기 특징맵 크기에 대응하도록 업샘플링하고,

상기 업샘플링된 글로벌 특징맵과 상기 로컬 특징맵을 연결하는, 컨텍스트 정보 추출 장치.

청구항 11

제10항에 있어서,

상기 프로세서는

상기 복수의 초해상맵에 기초하여 상기 글로벌 특징맵을 상기 특징맵 크기에 대응하도록 업샘플링하는, 컨텍스트 정보 추출 장치.

청구항 12

제9항에 있어서,

상기 프로세서는

상기 입력 영상에 기초하여 상기 초해상맵의 모양 및 크기를 결정하는, 컨텍스트 정보 추출 장치.

청구항 13

제9항에 있어서,

상기 프로세서는

SLIC 알고리즘에 기초하여 상기 초해상맵을 획득하는, 컨텍스트 정보 추출 장치.

청구항 14

제9항에 있어서,

상기 프로세서는

1X1 컨볼루션 레이어를 이용하여, 상기 복수의 글로벌 특징맵의 채널 수를 조정하는, 컨텍스트 정보 추출 장치.

청구항 15

제9항에 있어서,

상기 프로세서는

상기 복수의 초해상맵 각각에 기초하여, 상기 로컬 특징맵에 평균 풀링하는, 컨텍스트 정보 추출 장치.

발명의 설명

기술 분야

[0001] 아래 실시예들은 컨텍스트 정보 추출 방법 및 장치에 관한 것이다.

배경 기술

[0003] 영상인식 기술들은 최근 딥러닝(deep learning), 특히 영상과 같은 공간 정보를 다루는데 적합한 합성곱 신경망(Convolutional Neural Network, CNN)의 발전에 힘입어 큰 성능 향상을 이루었다. CNN은 데이터로부터 학습되는 영상표현을 제공하는데, 이 영상표현이 사람이 직접 설계한 영상표현 대비 계층적이고 더욱 풍부한 정보를 제공하기 때문에 딥러닝 이전 시기의 영상인식 기술들의 여러 한계들을 뛰어넘을 수 있었다.

[0004] 특히, 여러 영상인식 기술 중 의미론적 분할(semantic segmentation)은 이미지에 나타나는 객체들을 자동으로 검출 또는 추출하기 위해 복수의 상이한 객체 클래스들 중 하나 이상에 따라 이미지들의 요소들을 분류하는 기술을 의미할 수 있다.

[0005] 종래의 의미론적 분할(semantic segmentation)을 위해 FCN를 기본으로 설계된 모델은 특징맵에 존재하는 로컬 컨텍스트 정보(예를 들어, 모양, 형상, 재질의 특성 등이 포함될 것이다)을 활용했었다. 이에, 비슷한 형상 또는 재질을 가진 물체들(예를 들어, 보트와 자동차, 침대 시트와 베게, 빌딩과 고층빌딩 등)을 잘 구별하지 못하는 문제점들을 가지고 있었다.

발명의 내용

해결하려는 과제

[0007] 실시예들은 로컬 특징맵 뿐만 아니라 글로벌 특징맵을 의미론적 분할의 기준으로 활용하고자 한다.

[0008] 실시예들은 초해상 풀링을 통해 입력 영상에 적응성이 높은 글로벌 특징맵을 생성하고자 한다.

[0009] 실시예들은 멀티스케일 초해상 풀링을 통해 다양한 글로벌 특징맵을 의미론적 분할의 기준으로 활용하고자 한다.

[0010] 실시예들은 보다 높은 정확도를 갖는 컨텍스트 정보를 추출하고자 한다.

과제의 해결 수단

[0012] 일 실시예에 따른 컨텍스트 정보 추출 방법은 입력 영상에 관한 로컬 특징맵의 스케일에 따른 복수의 초해상맵을 획득하는 단계; 상기 복수의 초해상맵 각각에 기초하여, 상기 로컬 특징맵을 풀링(pooling)하여 복수의 글로벌 특징맵을 획득하는 단계; 상기 복수의 글로벌 특징맵을 상기 로컬 특징맵을 연결(concatenate)하는 단계; 및 상기 연결된 특징맵에 기초하여, 상기 입력 영상의 컨텍스트 정보를 추출하는 단계를 포함한다.

[0013] 일 실시예에 따른 컨텍스트 정보 추출 방법은 상기 글로벌 특징맵을 상기 로컬 특징맵 크기에 대응하도록 업샘플링하는 단계를 더 포함하고, 상기 연결하는 단계는 상기 업샘플링된 글로벌 특징맵과 상기 로컬 특징맵을 연결하는 단계를 포함할 수 있다.

[0014] 상기 업샘플링하는 단계는 상기 복수의 초해상맵에 기초하여, 상기 글로벌 특징맵을 상기 특징맵 크기에 대응하도록 업샘플링하는 단계를 포함할 수 있다.

[0015] 상기 초해상맵을 획득하는 단계는 상기 입력 영상에 기초하여, 상기 초해상맵의 모양 및 크기를 결정하는 단계를 포함할 수 있다.

[0016] 상기 초해상맵을 획득하는 단계는 SLIC(Simple linear iterative clustering) 알고리즘에 기초하여, 상기 초해상맵을 획득하는 단계를 포함할 수 있다.

[0017] 일 실시예에 따른 컨텍스트 정보 추출 방법은 1X1 컨볼루션 레이어를 이용하여, 상기 복수의 글로벌 특징맵의 채널 수를 조정하는 단계를 더 포함할 수 있다.

[0018] 상기 복수의 글로벌 특징맵을 획득하는 단계는 상기 복수의 초해상맵 각각에 기초하여, 상기 로컬 특징맵에 평균 풀링(average pooling)하는 단계를 포함할 수 있다.

[0019] 일 실시예에 따른 컨텍스트 정보 추출 장치는 적어도 하나의 프로그램이 저장된 메모리; 및 상기 적어도 하나의 프로그램을 실행함으로써 뉴럴 네트워크를 구동하는 프로세서를 포함하고, 상기 프로세서는 입력 영상에 관한 로컬 특징맵의 스케일에 따른 복수의 초해상맵을 획득하고, 상기 복수의 초해상맵 각각에 기초하여, 상기 로컬 특징맵을 풀링(pooling)하여 복수의 글로벌 특징맵을 획득하고, 상기 복수의 글로벌 특징맵을 상기 로컬 특징맵을 연결(concatenate)하고, 상기 연결된 특징맵에 기초하여, 상기 입력 영상의 컨텍스트 정보를 추출한다.

[0020] 상기 프로세서는 상기 글로벌 특징맵을 상기 특징맵 크기에 대응하도록 업샘플링하고, 상기 업샘플링된 글로벌 특징맵과 상기 로컬 특징맵을 연결할 수 있다.

[0021] 상기 프로세서는 상기 복수의 초해상맵에 기초하여 상기 글로벌 특징맵을 상기 특징맵 크기에 대응하도록 업샘플링할 수 있다.

[0022] 상기 프로세서는 상기 입력 영상에 기초하여 상기 초해상맵의 모양 및 크기를 결정할 수 있다.

- [0023] 상기 프로세서는 SLIC 알고리즘에 기초하여 상기 초해상맵을 획득할 수 있다.
- [0024] 상기 프로세서는 1X1 컨볼루션 레이어를 이용하여, 상기 복수의 글로벌 특징맵의 채널 수를 조정할 수 있다.
- [0025] 상기 프로세서는 상기 복수의 초해상맵 각각에 기초하여, 상기 로컬 특징맵에 평균 풀링할 수 있다.

발명의 효과

- [0027] 실시예들은 로컬 특징맵 뿐만 아니라 글로벌 특징맵을 의미론적 분할의 기준으로 활용할 수 있다.
- [0028] 실시예들은 초해상 풀링을 통해 입력 영상에 적응성이 높은 글로벌 특징맵을 생성할 수 있다.
- [0029] 실시예들은 멀티스케일 초해상 풀링을 통해 다양한 글로벌 특징맵을 의미론적 분할의 기준으로 활용할 수 있다.
- [0030] 실시예들은 보다 높은 정확도를 갖는 컨텍스트 정보를 추출할 수 있다.

도면의 간단한 설명

- [0032] 도 1은 일 실시예에 따른 컨텍스트 정보 추출 모델을 설명하기 위한 도면이다.
- 도 2는 일 실시예에 따른 컨텍스트 정보 추출 방법을 설명하기 위한 순서도이다.
- 도 3은 일 실시예에 따른 피라미드 멀티스케일 초해상 네트워크를 설명하기 위한 도면이다.
- 도 4는 일 실시예에 따른 컨텍스트 정보 추출 장치의 블록도이다.

발명을 실시하기 위한 구체적인 내용

- [0033] 본 명세서에서 개시되어 있는 특정한 구조적 또는 기능적 설명들은 단지 기술적 개념에 따른 실시예들을 설명하기 위한 목적으로 예시된 것으로서, 실시예들은 다양한 다른 형태로 실시될 수 있으며 본 명세서에 설명된 실시예들에 한정되지 않는다.
- [0034] 제1 또는 제2 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 이런 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 이해되어야 한다. 예를 들어 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소는 제1 구성요소로도 명명될 수 있다.
- [0035] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소가 다른 구성요소에 "직접 연결되어" 있다거나 "직접 접속되어" 있다고 언급된 때에는, 중간에 다른 구성요소가 존재하지 않는 것으로 이해되어야 할 것이다. 구성요소들 간의 관계를 설명하는 표현들, 예를 들어 "~간의"와 "바로~간의" 또는 "~에 이웃하는"과 "~에 직접 이웃하는" 등도 마찬가지로 해석되어야 한다.
- [0036] 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 실시된 특징, 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것이 존재함으로써 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.
- [0037] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 해당 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가진다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 의미를 갖는 것으로 해석되어야 하며, 본 명세서에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.
- [0038] 실시예들은 퍼스널 컴퓨터, 랩톱 컴퓨터, 태블릿 컴퓨터, 스마트 폰, 텔레비전, 스마트 가전 기기, 지능형 자동차, 키오스크, 웨어러블 장치 등 다양한 형태의 제품으로 구현될 수 있다. 이하, 실시예들을 첨부된 도면을 참조하여 상세하게 설명한다. 각 도면에 제시된 동일한 참조 부호는 동일한 부재를 나타낸다.

- [0040] 도 1은 일 실시예에 따른 컨텍스트 정보 추출 모델을 설명하기 위한 도면이다.
- [0041] 도 1을 참조하면, 일 실시예에 따른 컨텍스트 정보 추출 모델은 입력 영상(110)에 대한 의미론적 분할을 수행할 수 있다. 의미론적 분할은 입력 영상(110)에 포함된 객체의 픽셀 단위로 구분하고, 각 픽셀이 어떤 클래스(class)인지 추론하는 것을 의미할 수 있다.
- [0042] 종래의 FCN(Fully Convolutional Network)을 이용한 의미론적 분할 모델은 특징맵에 존재하는 로컬 컨텍스트 정보(예를 들어, 모양, 형상, 재질의 특성 등이 포함될 것이다)을 활용했었다. 이에, 비슷한 형상 또는 재질을 가진 물체들을 잘 구별하지 못하는 문제점들을 가지고 있었다. FCN을 이용한 의미론적 분할 모델은 이미지 전체에서 컨텍스트 정보를 뽑아내는 능력이 부족했기 때문에 다양한 풍경에서의 일반화 능력이 떨어진다는 단점이 있다. 예를 들어, FCN을 이용한 의미론적 분할 모델은 보트와 자동차를 구별하지 못할 수 있다. 이미지 전반에 걸쳐서 정보를 뽑아낼 수 있다면, '물 위에 있기 때문에 이 물체는 보트다.'는 추론이 가능할 수 있다.
- [0043] 일반적인 FCN 을 이용한 의미론적 분할 모델은 컨볼루션 레이어를 통해 인코딩한 정보를 점진적으로 보간(interpolation)하고, 보간된 정보는 다시 컨볼루션 레이어를 거치며 보다 풍부해질 수 있다. 이런 과정을 거쳐 입력 영상에 상응하는 사이즈의 스코어 맵(score map)을 출력하고, 이를 통해 각 픽셀별로 어떤 클래스에 속하는지 스코어를 표시할 수 있다.
- [0044] 이와 달리, 일 실시예에 따른 컨텍스트 정보 추출 모델은 로컬 특징맵 뿐만 아니라 글로벌 특징맵을 의미론적 분할의 기준으로 활용할 수 있다. 글로벌 특징맵은 입력 영상(110)의 전역적 특징 정보에 관한 특징맵일 수 있고, 로컬 특징맵은 입력 영상(110)의 지역적 특징 정보에 관한 특징맵일 수 있다. 글로벌 특징맵을 사용하면 이미지 전반에 걸쳐서 정보를 뽑아낼 수 있다.
- [0045] 예를 들어, 일 실시예에 따른 컨텍스트 정보 추출 모델은 자동차 또는 보트 모양의 객체가 물 근처에 있는 경우, 객체에 대응하는 로컬 특징맵과, 물의 특징이 포함되어 평균을 구한 글로벌 특징맵을 통해 형상과 주변 상황을 모두 고려할 수 있다.
- [0046] 나아가, 일 실시예에 따른 컨텍스트 정보 추출 모델은 하나의 글로벌 특징 맵이 아닌, 스케일에 따른 복수의 글로벌 특징맵을 사용하여 여러 측면에서 입력 영상(110)의 전체적인 특징을 추출할 수 있다.
- [0047] 컨텍스트 정보 추출 모델은 입력 영상(110)을 딥 인코더(120)(예를 들어, 컨볼루션 레이어)에 입력하여 로컬 특징맵을 추출할 수 있다. 나아가, 컨텍스트 정보 추출 모델은 피라미드 멀티스케일 초해상(super pixel) 풀링 레이어(130)를 통하여 복수의 글로벌 특징맵을 획득할 수 있다. 이 후, 풀링을 통해 크기가 줄어든 피라미드 멀티스케일 초해상 풀링 레이어(130)의 출력을 업샘플링 레이어(140)를 통해 다시 원래 크기로 키워 입력 영상(110)의 컨텍스트 정보를 포함하는 최종 출력(150)을 획득할 수 있다.
- [0049] 도 2는 일 실시예에 따른 컨텍스트 정보 추출 방법을 설명하기 위한 순서도이다.
- [0050] 도 2를 참조하면, 단계들(210 내지 240)은 일 실시예에 따른 컨텍스트 정보 추출 장치에 의해 수행될 수 있다. 컨텍스트 정보 추출 장치는 하나 또는 그 이상의 하드웨어 모듈, 하나 또는 그 이상의 소프트웨어 모듈, 또는 이들의 다양한 조합에 의하여 구현될 수 있다. 나아가, 도 2에 도시된 동작은 도시된 순서 및 방식으로 수행될 수 있지만, 도시된 실시예의 사상 및 범위를 벗어나지 않으면서 일부 동작의 순서가 변경되거나 일부 동작이 생략될 수 있다. 도 2에 도시된 다수의 동작은 병렬로 또는 동시에 수행될 수 있다.
- [0051] 단계(210)에서, 컨텍스트 정보 추출 장치는 입력 영상에 관한 로컬 특징맵의 스케일에 따른 복수의 초해상맵을 획득한다. 영상을 특징이 비슷한 작은 균일한 영역으로 잘게 나누고 이 작은 영역들을 기본단위로 하여 영상처리를 할 수 있는데, 여기서 작게 나누어 진 균일영역을 초해상(super pixel)이라 할 수 있고, 특징이 비슷한 작은 균일한 영역으로 잘게 나누는 동작을 초해상 처리라고 할 수 있다. 초해상맵은 영상 또는 영상의 특징맵에 초해상 처리를 한 결과물일 수 있다.
- [0052] 컨텍스트 정보 추출 장치는 로컬 특징맵에 스케일에 따라 복수의 초해상 처리를 수행하여 복수의 초해상맵을 획득할 수 있다. 예를 들어, 컨텍스트 정보 추출 장치는 초해상 100개, 200개, 300개 및 400개로 구성된 복수의 초해상맵을 획득할 수 있다.
- [0053] 컨텍스트 정보 추출 장치는 SLIC(Simple linear iterative clustering) 알고리즘에 기초하여, 초해상맵을 획득

할 수 있다. 해당 분야에서 통상의 지식을 가진 자라면 SLIC 알고리즘을 명확하게 이해할 수 있는 바, 보다 상세한 설명은 생략한다. 다만, SLIC 알고리즘은 초해상맵을 획득하는 하나의 실시예일 뿐, 초해상맵을 획득하는 방법은 이에 한정되지 않고 다양한 방법을 통해 초해상맵을 획득할 수 있다.

[0054] 단계(220)에서, 컨텍스트 정보 추출 장치는 복수의 초해상맵 각각에 기초하여, 로컬 특징맵을 풀링(pooling)하여 복수의 글로벌 특징맵을 획득한다. 컨텍스트 정보 추출 장치는 복수의 초해상맵을 풀링을 위한 각각의 구획을 정의하기 위해 활용할 수 있다. 컨텍스트 정보 추출 장치는 로컬 특징맵으로부터 초해상 단위의 특징 벡터를 얻고, 그 뒤의 개별 초해상을 분류할 수 있다.

[0055] 컨텍스트 정보 추출 장치는 복수의 초해상맵 각각에 기초하여 복수의 글로벌 특징맵을 획득하기 때문에 입력 영상의 전체적인 특징을 다양한 측면에서 추출할 수 있다. 예를 들어, 초해상 100개로 구성된 초해상맵은 가장 조악하지만 가장 넓은 범위의 정보를 포함할 수 있고, 초해상 400개로 구성된 초해상맵은 가장 정밀하지만 가장 좁은 범위의 정보를 포함할 수 있다.

[0056] 나아가, 컨텍스트 정보 추출 장치는 초해상 풀링을 통해 입력 영상에 적응성이 높은 글로벌 특징맵을 생성할 수 있다. 구체적으로, 컨텍스트 정보 추출 장치는 입력 영상에 포함된 객체의 모양 및 구조를 고려하지 않는 PSPNet(Pyramid Scene Parsing Network)과는 달리, 초해상맵에 기초하여 글로벌 특징맵을 획득하기 때문에 입력 영상에 포함된 객체의 모양 및 구조를 고려할 수 있다.

[0057] 단계(230)에서, 컨텍스트 정보 추출 장치는 복수의 글로벌 특징맵을 로컬 특징맵을 연결(concatenate)할 수 있다. 예를 들어, 컨텍스트 정보 추출 장치는 하나의 로컬 특징맵과 4개의 글로벌 특징맵을 연결할 수 있다. 이때, 컨텍스트 정보 추출 장치는 글로벌 컨텍스트 정보와 로컬 컨텍스트 정보의 비율을 동일하도록 로컬 특징맵과 글로벌 특징맵을 조정하여 연결할 수 있다.

[0058] 단계(240)에서, 컨텍스트 정보 추출 장치는 연결된 특징맵에 기초하여, 입력 영상의 컨텍스트 정보를 추출한다. 예를 들어, 컨텍스트 정보 추출 장치는 연결된 특징맵을 컨볼루션 레이어에 통과시켜 입력 영상의 컨텍스트 정보를 픽셀별로 추출할 수 있다.

[0060] 도 3은 일 실시예에 따른 피라미드 멀티스케일 초해상 네트워크를 설명하기 위한 도면이다.

[0061] 도 3을 참조하면, 일 실시예에 따른 피라미드 멀티스케일 초해상 네트워크는 입력 영상에 관한 로컬 특징맵을 생성하는 컨볼루션 레이어, 복수의 초해상맵을 생성하는 초해상 처리 레이어(예를 들어, SLIC 알고리즘 처리 레이어) 및 복수의 초해상맵 각각에 기초하여, 로컬 특징맵을 풀링하여 복수의 글로벌 특징맵을 생성하는 피라미드 멀티스케일 초해상 풀링 레이어를 포함할 수 있다.

[0062] 일 실시예에 따른 피라미드 멀티스케일 초해상 풀링 레이어는 평균 풀링(average pooling) 또는 맥스 풀링(max pooling)을 수행할 수 있다. 피라미드 멀티스케일 초해상 풀링 레이어의 입력 I는 C*H*W(C는 채널의 수, H는 높이, W는 너비)의 크기를 갖을 수 있다. 피라미드 멀티스케일 초해상 풀링 레이어는 복수의 초해상맵에 기초하여 풀링을 수행하는데, 초해상맵 S의 크기는 H*W이고 K개의 초해상으로 구성될 수 있다. 이러한 경우, 출력 벡터 Z는 C*K의 모양을 갖을 수 있다. 평균 풀링을 통해 풀링을 수행할 경우, 피라미드 멀티스케일 초해상 풀링 레이어의 출력은 수학식 1과 같을 수 있다.

수학식 1

$$Z_{C,K} = \frac{1}{N} \sum_h \sum_w I_C * M_K$$

[0063]

[0064] 여기서, M_K 는 초해상맵 S에서 인덱스 k를 사용하여 모든 픽셀을 1로 설정하고 나머지는 0으로 설정하여 생성되는 이진 마스크일 수 있다. N은 생성된 마스크 M에서 1s의 수일 수 있다.

[0065] 피라미드 멀티스케일 초해상 네트워크는 피라미드 멀티스케일 초해상 풀링 레이어 뒤에 1x1 컨볼루션 레이어를

더 포함할 수 있다. 피라미드 멀티스케일 초해상 네트워크는 풀링 처리 후 1X1 컨볼루션 레이어를 통해 복수의 글로벌 특징맵의 채널 수를 조정할 수 있다. 예를 들어, 풀링 레이어의 개수를 N이라고 할 때, 출력 채널의 수는 입력 채널의 수/N일 수 있다.

[0066] 피라미드 멀티스케일 초해상 네트워크는 초해상 프로젝션 레이어(superpixel projection layer)를 더 포함할 수 있다. 초해상 프로젝션 레이어는 이중 선형 업샘플링(bilinear upsampling) 방법 대신에, 복수의 초해상맵을 사용한 컨텍스트 특징 프로젝션을 방법을 사용하여 글로벌 특징맵을 로컬 특징맵 크기에 대응하도록 업샘플링할 수 있다. 예를 들어, 컨텍스트 정보 추출 장치는 복수의 초해상맵에 기초하여 글로벌 특징맵을 특징맵 크기에 대응하도록 업샘플링할 수 있다.

[0068] 도 4는 일 실시예에 따른 컨텍스트 정보 추출 장치의 블록도이다.

[0069] 도 4를 참조하면, 일 실시예에 따른 컨텍스트 정보 추출 장치(400)는 프로세서(410)를 포함한다. 컨텍스트 정보 추출 장치(400)는 메모리(430), 통신 인터페이스(450), 및 센서들(470)을 더 포함할 수 있다. 프로세서(410), 메모리(430), 통신 인터페이스(450), 및 센서들(470)은 통신 버스(405)를 통해 서로 통신할 수 있다.

[0070] 메모리(430)에는 적어도 하나의 프로그램이 저장될 수 있고, 프로세서(410)는 적어도 하나의 프로그램을 실행함으로써 뉴럴 네트워크를 구동할 수 있다.

[0071] 프로세서(410)는 입력 영상에 관한 로컬 특징맵의 스케일에 따른 복수의 초해상맵을 획득하고, 복수의 초해상맵 각각에 기초하여, 로컬 특징맵을 풀링(pooling)하여 복수의 글로벌 특징맵을 획득하고, 복수의 글로벌 특징맵을 로컬 특징맵을 연결(concatenate)하고, 연결된 특징맵에 기초하여, 입력 영상의 컨텍스트 정보를 추출한다.

[0072] 메모리(430)는 휘발성 메모리 또는 비 휘발성 메모리일 수 있다.

[0073] 센서들(470)은 예를 들어, 입력 영상을 수신하는 영상 센서 등을 포함할 수 있다.

[0074] 실시예에 따라서, 프로세서(410)는 글로벌 특징맵을 특징맵 크기에 대응하도록 업샘플링하고, 업샘플링된 글로벌 특징맵과 로컬 특징맵을 연결할 수 있다.

[0075] 프로세서(410)는 복수의 초해상맵에 기초하여 글로벌 특징맵을 특징맵 크기에 대응하도록 업샘플링할 수 있다. 프로세서(410)는 입력 영상에 기초하여 초해상맵의 모양 및 크기를 결정할 수 있다. 프로세서(410)는 SLIC 알고리즘에 기초하여 초해상맵을 획득할 수 있다. 프로세서(410)는 1X1 컨볼루션 레이어를 이용하여, 복수의 글로벌 특징맵의 채널 수를 조정할 수 있다. 프로세서(410)는 복수의 초해상맵 각각에 기초하여, 로컬 특징맵에 평균 풀링할 수 있다.

[0077] 이상에서 설명된 실시예들은 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치, 방법 및 구성요소는, 예를 들어, 프로세서, 콘트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPGA(field programmable gate array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 콘트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

[0078] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치, 또는 전송되는 신호 파(signal wave)에 영구적으로,

또는 일시적으로 구체화(embody)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

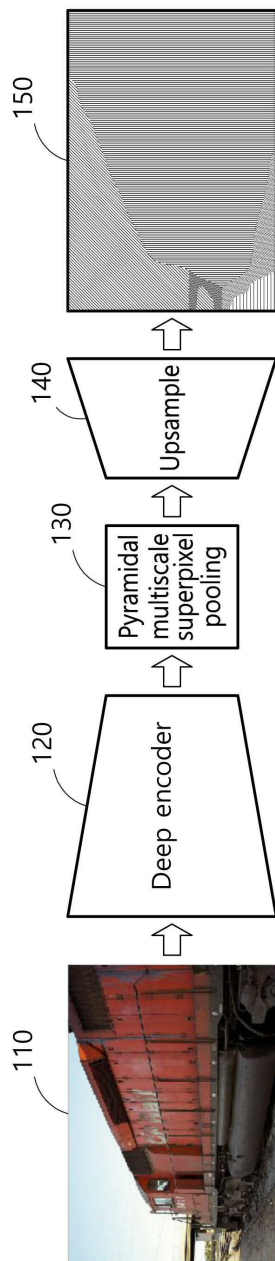
[0079] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.

[0080] 이상과 같이 실시예들이 비록 한정된 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기를 기초로 다양한 기술적 수정 및 변형을 적용할 수 있다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

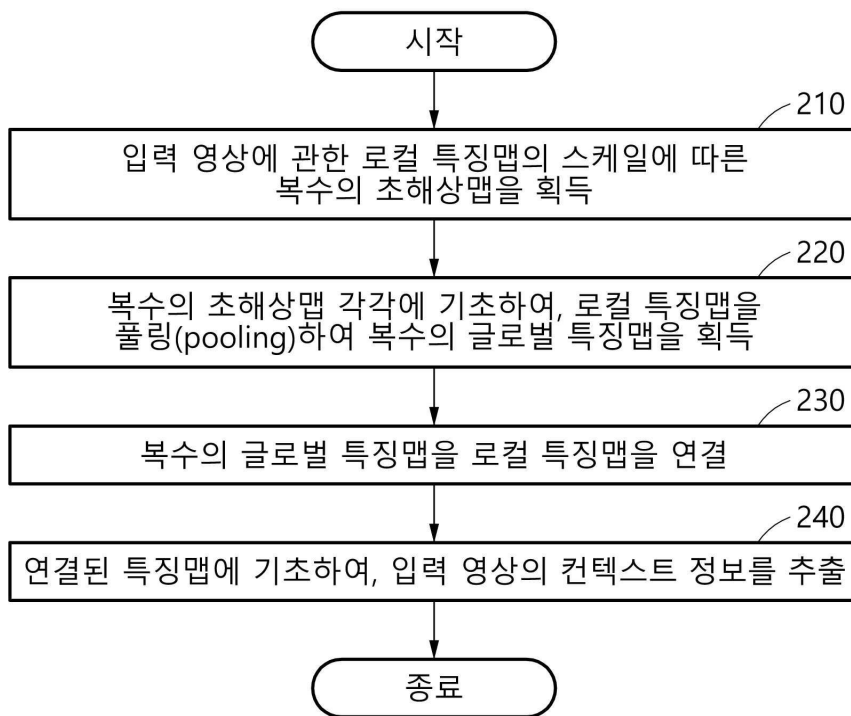
[0081] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

도면

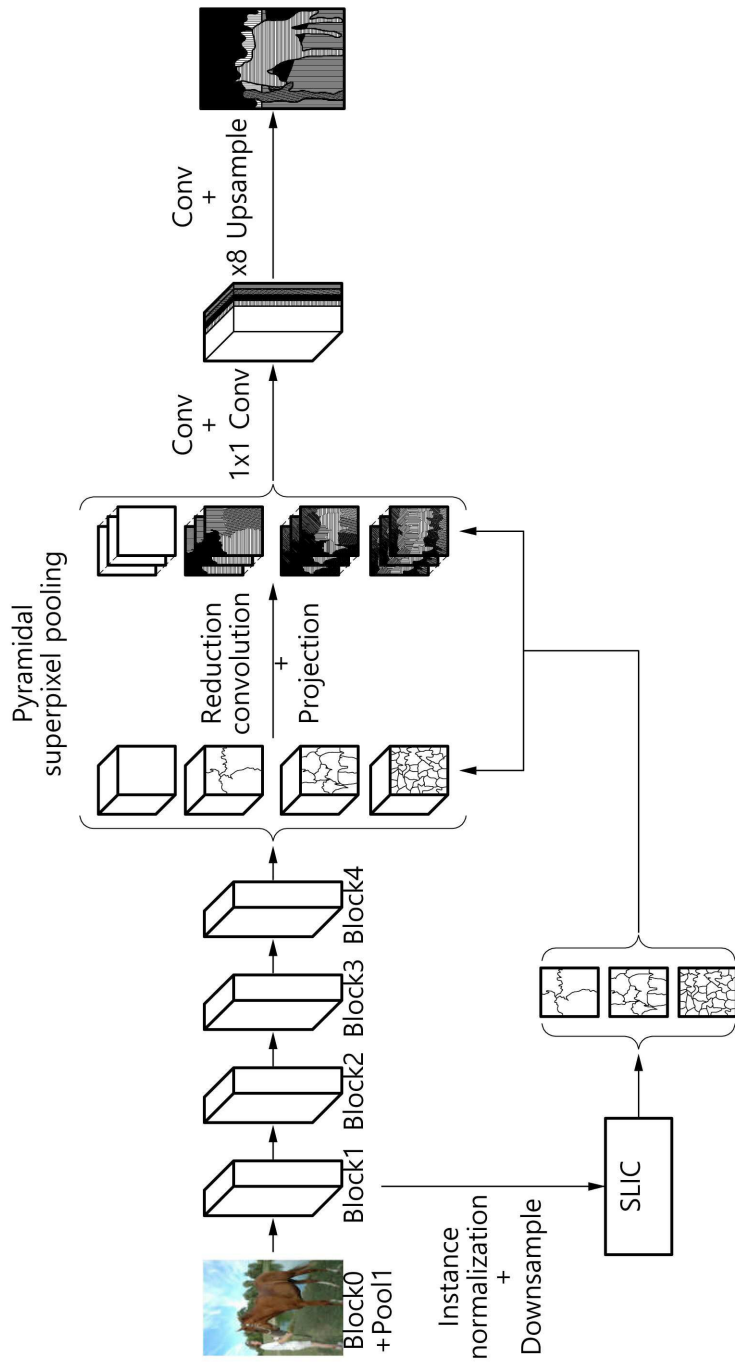
도면1



도면2



도면3



도면4

