



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2020년08월04일
(11) 등록번호 10-2141302
(24) 등록일자 2020년07월29일

(51) 국제특허분류(Int. Cl.)
G06T 7/11 (2017.01) G06F 17/18 (2006.01)
G06T 5/00 (2019.01) G06T 7/194 (2017.01)
(52) CPC특허분류
G06T 7/11 (2017.01)
G06F 17/18 (2013.01)
(21) 출원번호 10-2019-0024740
(22) 출원일자 2019년03월04일
심사청구일자 2019년03월04일
(56) 선행기술조사문헌
"segnetres-crf: a deep convolutional
encoder-decoder architecture for semantic
image segmentation"*
segnet: a deep convolutional encoder-decoder
architecture for image segmentation*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
이화여자대학교 산학협력단
서울특별시 서대문구 이화여대길 52 (대현동, 이
화여자대학교)
(72) 발명자
김정태
서울특별시 서초구 서운로 197, 102동 402호(서초
동, 롯데캐슬클래식아파트)
조희연
서울특별시 구로구 경인로11길 8-1, 202호(
오류동)
김경실
인천광역시 연수구 신송로6번길 7, 103동 202호(
송도동, 송도 성지리벨루스)
(74) 대리인
특허법인(유한)아이시스

전체 청구항 수 : 총 5 항

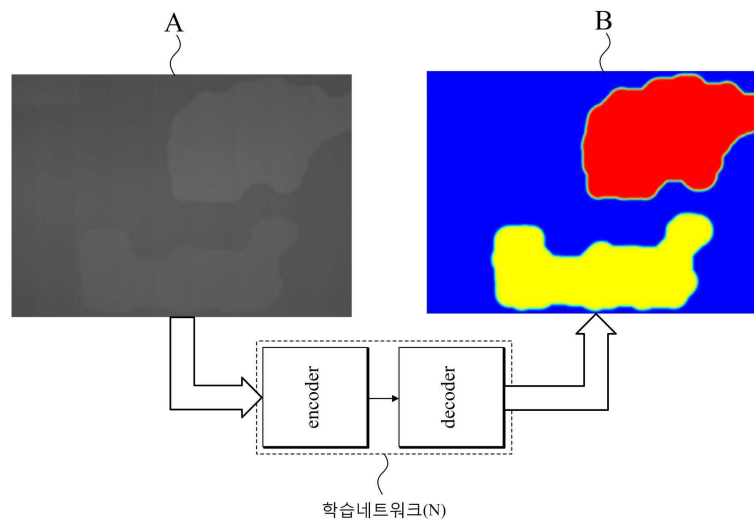
심사관 : 박금옥

(54) 발명의 명칭 회귀 딥러닝 모델 기반의 영상 객체 검출 방법 및 영상처리장치

(57) 요약

회귀 딥러닝 모델 기반의 영상 객체 검출 방법은 영상처리장치가 회귀 딥러닝 모델에 타깃 객체 및 배경이 포함 된 소스 영상을 입력하는 단계 및 상기 영상처리장치가 상기 회귀 딥러닝 모델을 이용하여 상기 소스 영상에서 상기 타깃 객체가 구분된 출력 영상을 생성하는 단계를 포함한다. 상기 회귀 딥러닝 모델은 컨볼루션 인코더-디 코더(Convolutional encoder-decoder) 구조이다. 상기 회귀 딥러닝 모델은 픽셀 단위의 회귀(regression) 방법 으로 상기 타깃 객체와 배경을 구분한다.

대표도 - 도4



(52) CPC특허분류

G06T 5/002 (2013.01)

G06T 7/194 (2017.01)

G06T 2207/20081 (2013.01)

G06T 2207/20182 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호 1415157926

부처명 산업통상자원부

연구관리전문기관 한국산업기술평가관리원

연구사업명 차세대디스플레이기술개발

연구과제명 디스플레이 패널 품질 검사를 위한 학습 기반 영상 시스템

기 여 율 70/100

주관기관 이화여자대학교 산학협력단

연구기간 2018.03.01 ~ 2018.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호 1711066866

부처명 과학기술정보통신부

연구관리전문기관 한국연구재단

연구사업명 개인기초연구(과기정통부)(R&D)

연구과제명 스마트 팩토리를 위한 지능형 머신 비전 시스템

기 여 율 30/100

주관기관 이화여자대학교 산학협력단

연구기간 2018.03.01 ~ 2019.02.28

명세서

청구범위

청구항 1

영상처리장치가 회귀 딥러닝 모델에 타깃 객체 및 배경이 포함된 소스 영상을 입력하는 단계; 및

상기 영상처리장치가 상기 회귀 딥러닝 모델을 이용하여 상기 소스 영상에서 상기 타깃 객체가 구분된 출력 영상을 생성하는 단계를 포함하되,

상기 회귀 딥러닝 모델은 컨볼루션 인코더-디코더(Convolutional encoder-decoder) 구조이고, 상기 회귀 딥러닝 모델은 픽셀 단위의 회귀(regression) 방법으로 상기 타깃 객체와 배경을 구분하고, 상기 인코더의 손실함수는 아래의 수학식으로 정의되는 회귀 딥러닝 모델 기반의 영상 객체 검출 방법.

$$L = \frac{1}{ndcr} \sum_{l=1}^n \sum_{k=1}^d \sum_{j=1}^c \sum_{i=1}^r (Y_{i,j,k,l} - T_{i,j,k,l})^2 + \lambda R$$

$$R = \frac{1}{ndcr} \sum_{l=1}^n \sum_{k=1}^d \sum_{j=1}^c \sum_{i=1}^r \sqrt{(Y_{i,j,k,l} - Y_{i-1,j,k,l})^2 + (Y_{i,j,k,l} - Y_{i,j-1,k,l})^2 + \epsilon^2}$$

(여기서, L은 손실함수, n은 학습에 사용한 데이터 수, d는 채널 수, c는 열 개수, r은 행 개수, Y는 출력 데이터, T는 정답 데이터, R은 정규화 항, λ는 가중치, ε는 임의의 작은 양수)

청구항 2

제1항에 있어서,

상기 인코더는 픽셀 단위로 특징을 추출하여 특징 맵을 생성하고, 상기 디코더는 상기 특징맵을 입력으로 상기 출력 영상을 출력하는 회귀 딥러닝 모델 기반의 영상 객체 검출 방법.

청구항 3

제1항에 있어서,

상기 인코더는 픽셀 단위 회귀를 사용하여 상기 배경에서 왜곡 성분을 평탄화(smoothing)하고, 상기 왜곡 성분이 제거된 배경과 상기 타깃 객체에 대하여 특징을 추출하는 회귀 딥러닝 모델 기반의 영상 객체 검출 방법.

청구항 4

제1항에 있어서,

상기 인코더는 복수의 인코더 블록과 하나의 풀링 계층을 포함하는 회귀 딥러닝 모델 기반의 영상 객체 검출 방법.

청구항 5

제1항에 있어서,

상기 인코더의 손실함수는 특정 픽셀이 주변 픽셀과 일정 범위 내의 유사한 값으로 조절하는 정규화 항을 포함하는 회귀 딥러닝 모델 기반의 영상 객체 검출 방법.

청구항 6

삭제

청구항 7

삭제

청구항 8

삭제

청구항 9

삭제

청구항 10

삭제

청구항 11

삭제

청구항 12

삭제

발명의 설명

기술 분야

[0001] 이하 설명하는 기술은 영상에서 특정 객체를 검출하는 기법에 관한 것이다. 특히 이하 설명하는 기술은 픽셀 단위 회귀 모델을 사용한 객체 검출 기법에 관한 것이다.

배경 기술

[0002] 머신 비전(machine vision)은 결함 검사, 분류, 인식 등의 응용 분야에서 수동검사자를 대체하여 반복적으로 고정도의 검사를 수행할 수 있어서 널리 연구되고 있다.

[0003] 최근 주목받고 있는 머신 러닝 기술은 수동 검사자의 판단 형태를 학습할 수 있어서 기존 머신 비전 기술이 적용되기 어려웠던 응용 분야에서 활발히 연구되고 있다.

선행기술문헌

비특허문헌

[0004] (비특허문헌 0001) Matthias Haselmann, Dieter P. Gruber, Paul Tabatabai, Anomaly Detection using Deep Learning based Image Completion, 2018 17th IEEE International Conference on Machine Learning and Applications

발명의 내용

해결하려는 과제

[0005] 유사한 패턴의 배경을 갖는 영상에서 특정 객체(불량 부위 등)를 검출하기가 쉽지 않다. 종래 기술은 소스(source) 영상에서 특정 객체를 배제한 영상과 소스 영상을 비교하여 차이값을 결정하는 방식을 사용하였다. 다만 유사한 패턴의 배경이나 잡음이 많은 배경에서 특정 객체를 검출하는 과정 자체가 쉽지 않아 제품 불량 판단의 정확도가 떨어진다.

[0006] 이하 설명하는 기술은 픽셀 단위 회귀 방식의 딥러닝 모델을 사용하여 소스 영상에 포함된 특정 객체를 검출하는 기법을 제공하고자 한다. 이하 설명하는 기술은 왜곡 성분이 제거된 배경과 객체를 비교하여 특정 객체를 검출하는 기법을 제공하고자 한다.

과제의 해결 수단

- [0007] 회귀 딥러닝 모델 기반의 영상 객체 검출 방법은 처리장치가 회귀 딥러닝 모델에 타깃 객체 및 배경이 포함된 소스 영상을 입력하는 단계 및 상기 영상처리장치가 상기 회귀 딥러닝 모델을 이용하여 상기 소스 영상에서 상기 타깃 객체가 구분된 출력 영상을 생성하는 단계를 포함한다.
- [0008] 회귀 딥러닝 모델 기반의 객체 검출 장치는 객체 및 배경이 포함된 소스 영상을 입력받는 입력장치, 상기 소스 영상에서 상기 타깃 객체가 구분된 출력 영상을 생성하는 회귀 딥러닝 모델을 저장하는 저장장치 및 상기 소스 영상을 상기 회귀 딥러닝 모델에 입력하여 상기 출력 영상을 생성하는 연산장치를 포함한다.
- [0009] 상기 회귀 딥러닝 모델은 컨볼루션 인코더-디코더(Convolutional encoder-decoder) 구조이고, 상기 회귀 딥러닝 모델은 픽셀 단위의 회귀(regression) 방법으로 상기 타깃 객체와 배경을 구분한다.

발명의 효과

- [0010] 이하 설명하는 기술은 픽셀 단위 회귀 방식을 사용하여 왜곡 성분이 많은 소스 영상에서도 특정 객체를 정확하게 검출한다. 이하 설명하는 기술은 특정 객체가 식별된(구분된) 영상을 생성하여 제품 불량 검출이 용이하다.

도면의 간단한 설명

- [0011] 도 1은 회귀 딥러닝 모델을 이용한 불량 검출 시스템에 대한 예이다.
- 도 2는 컨볼루션 인코더-디코더에 대한 예이다.
- 도 3은 회귀 딥러닝 모델의 학습 네트워크 구조에 대한 예이다.
- 도 4는 소스 영상과 출력 영상에 대한 예이다.
- 도 5는 객체 검출 장치의 구성에 대한 예이다.

발명을 실시하기 위한 구체적인 내용

- [0012] 이하 설명하는 기술은 다양한 변경을 가할 수 있고 여러 가지 실시례를 가질 수 있는 바, 특정 실시례들을 도면에 예시하고 상세하게 설명하고자 한다. 그러나, 이는 이하 설명하는 기술을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 이하 설명하는 기술의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.
- [0013] 제1, 제2, A, B 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 해당 구성요소들은 상기 용어들에 의해 한정되지는 않으며, 단지 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 이하 설명하는 기술의 권리 범위를 벗어나지 않으면서 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다. 및/또는 이라는 용어는 복수의 관련된 기재된 항목들의 조합 또는 복수의 관련된 기재된 항목들 중의 어느 항목을 포함한다.
- [0014] 본 명세서에서 사용되는 용어에서 단수의 표현은 문맥상 명백하게 다르게 해석되지 않는 한 복수의 표현을 포함하는 것으로 이해되어야 하고, "포함한다" 등의 용어는 실시된 특징, 개수, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것이 존재함을 의미하는 것이지, 하나 또는 그 이상의 다른 특징들이나 개수, 단계 동작 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 배제하지 않는 것으로 이해되어야 한다.
- [0015] 도면에 대한 상세한 설명을 하기에 앞서, 본 명세서에서의 구성부들에 대한 구분은 각 구성부가 담당하는 주기능 별로 구분한 것에 불과함을 명확히 하고자 한다. 즉, 이하에서 설명할 2개 이상의 구성부가 하나의 구성부로 합쳐지거나 또는 하나의 구성부가 보다 세분화된 기능별로 2개 이상으로 분화되어 구비될 수도 있다. 그리고 이하에서 설명할 구성부 각각은 자신이 담당하는 주기능 이외에도 다른 구성부가 담당하는 기능 중 일부 또는 전부의 기능을 추가적으로 수행할 수도 있으며, 구성부 각각이 담당하는 주기능 중 일부 기능이 다른 구성부에 의해 전담되어 수행될 수도 있음은 물론이다.
- [0016] 또, 방법 또는 동작 방법을 수행함에 있어서, 상기 방법을 이루는 각 과정들은 문맥상 명백하게 특정 순서를 기재하지 않은 이상 명기된 순서와 다르게 일어날 수 있다. 즉, 각 과정들은 명기된 순서와 동일하게 일어날 수도 있고 실질적으로 동시에 수행될 수도 있으며 반대의 순서대로 수행될 수도 있다.
- [0018] 이하 설명하는 기술은 기계학습(machine learning)모델을 사용하여 제품의 불량을 검출하는 기법이다. 이하 설

명하는 기술은 제품에 대한 영상을 분석하여 제품 품질을 검사하는 머신 비전 기법에 적용될 수 있다. 기계학습 모델은 널리 알려진 바와 같이 다양한 모델이 있다. 이하 설명하는 기술은 인공신경망(artificial neural network)을 사용하여 영상을 분석한다고 가정한다.

- [0019] 이하 인공신경망을 이용하여 영상을 분석하는 주체는 영상처리장치라고 명명한다. 영상처리장치는 일정한 데이터 처리 및 연산이 가능한 장치에 해당한다. 영상처리장치는 연산 기능이 있는 컴퓨팅 장치이다. 예컨대, 영상처리장치는 PC, 스마트기기, 서버 등과 같은 장치로 구현될 수 있다. 영상처리장치는 사전에 학습된 인공신경망 모델을 이용하여 입력 영상을 처리한다.
- [0020] 영상처리장치는 제품(객체)을 포함하는 영상을 분석할 수 있다. 분석 대상이 되는 입력 영상을 이하 소스 영상이라고 명명한다. 영상 처리 장치는 인공신경망에 소스 영상을 입력하여 분석 결과를 생성한다. 영상처리장치는 분석 결과를 일정한 영상으로 생성할 수 있다. 영상처리장치가 소스 영상에서 검출한 특정 객체를 타깃 객체라고 명명한다. 예컨대, 타깃 객체는 불량 부위, 특정 부품, 특정 종류의 객체 등이 될 수 있다. 소스 영상은 타깃 객체와 배경을 포함한다.
- [0022] 도 1은 회귀 딥러닝 모델을 이용한 불량 검출 시스템(100)에 대한 예이다. 도 1은 회귀 딥러닝 모델을 사용하여 불량 부위 객체를 검출하는 시스템이다. 회귀 딥러닝 모델은 후술한다. 회귀 딥러닝 모델은 소스 영상에서 특정 객체를 검출하는 신경망 모델이다. 따라서 불량 검출 시스템(100)은 회귀 딥러닝 모델을 적용한 하나의 애플리케이션에 해당한다.
- [0023] 도 1의 불량 검출 시스템(100)은 제품 공정 라인에 설치된 시스템일 수 있다. 카메라(111, 112, 113)는 검사 대상인 제품의 외관을 촬영한다. 도 1은 3개의 카메라(111, 112 및 113)를 예로 도시하였다. 복수의 공정 라인에서 개별 카메라를 사용하여 불량 검사를 할 수 있다. 나아가 하나의 제품을 서로 다른 관점에서 촬영하는 복수의 카메라를 사용하여 불량 검사를 할 수도 있다. 카메라(111, 112 내지 113)는 소스 영상을 캡처한다.
- [0024] 도 1에서 불량 검출하는 영상처리장치로 검출 서버(150) 및 검출 PC(180)를 예로 도시하였다. 불량 검출 시스템(100)은 검출 서버(150) 또는 검출 PC(180) 중 어느 하나를 포함할 수 있다. 나아가 불량 검출 시스템(100)은 검출 서버(150) 및 검출 PC(180)를 모두 포함할 수 있다.
- [0025] 검출 서버(150)는 카메라(111, 112 및 113 중 적어도 하나)로부터 소스 영상을 전달받는다. 검출 서버(150)는 유선 또는 무선 네트워크를 통해 소스 영상을 수신할 수 있다. 검출 서버(150)는 제품의 불량 여부를 나타내는 분석 결과 또는 분할 영상을 생성하여 사용자(10)에게 전달한다. 사용자(10)는 사용자 단말을 통해 검출 서버(150)의 분석 결과 또는 분할 영상을 확인할 수 있다. 사용자 단말은 PC, 스마트 기기, 디스플레이 장치 등과 같은 장치를 의미한다.
- [0026] 검출 PC(180)는 카메라(111, 112 및 113 중 적어도 하나)로부터 소스 영상을 전달받는다. 검출 PC(180)는 유선 또는 무선 네트워크를 통해 소스 영상을 수신할 수 있다. 검출 PC(180)는 제품의 불량 여부를 나타내는 분석 결과 또는 분할 영상을 생성하여 사용자(10)에게 전달한다. 사용자(20)는 검출 PC(180)를 통해 제품의 불량 여부를 확인할 수 있다.
- [0028] 영상처리장치는 회귀 딥러닝 모델을 사용하여 소스 영상에서 타깃 객체를 검출할 수 있다. 회귀 딥러닝 모델은 컨볼루션 인코더-디코더(convolutional encoder/decoder) 구조의 모델일 수 있다. 먼저 컨볼루션 인코더-디코더에 대하여 설명한다.
- [0029] 도 2는 컨볼루션 인코더-디코더에 대한 예이다. 도 2는 일반적인 컨볼루션 인코더-디코더에 대한 예이다. 컨볼루션 인코더-디코더는 컨볼루션 인코더 및 컨볼루션 디코더로 구성된다. 컨볼루션 인코더-디코더는 서로 거울상 구조를 갖는다. 컨볼루션 인코더-디코더는 소스 영상에서 추출되는 특징 맵(feature map)의 크기를 줄였다가 다시 소스 영상 크기만큼 크게 만들어서, 소스 영상의 각 픽셀에 대해 분류 결과로 클래스를 분류한다.
- [0030] 컨볼루션 인코더는 복수의 계층을 갖는다. 도 2는 5개의 계층을 갖는 컨볼루션 인코더를 예로 도시하였다. 하나의 계층은 컨볼루션 계층(convolutional layer)과 풀링 계층(pooling layer)을 갖는다. 도 2는 컨볼루션 계층이라고만 표시하였으나, 해당 계층은 컨볼루션 계층 외에도 배치 표준화 계층(batch normalization layer) 및 비선형화 계층(non linear activation layer)을 더 포함할 수 있다. 이 경우 하나의 계층은 컨볼루션 계층, 배치 표준화 계층, 비선형화 계층 순서를 가질 수 있다. 해당 계층은 복수 회 반복될 수 있다. 도 2는 2회 반복되는 계층을 도시하였다. 즉 하나의 계층은 (i) (컨볼루션 계층, 배치 표준화 계층 및 비선형화

계층) $\times$ n회 + (ii) 풀링 계층 구조를 갖는다.

- [0031] 컨볼루션 계층은 입력 이미지에 대한 컨볼루션 연산을 통해 특징맵(feature map)을 출력한다. 이때 컨볼루션 연산을 수행하는 필터(filter)를 커널(kernel) 이라고도 부른다. 필터의 크기를 필터 크기 또는 커널 크기라고 한다. 커널을 구성하는 연산 파라미터(parameter)를 커널 파라미터(kernel parameter), 필터 파라미터(filter parameter), 또는 가중치(weight)라고 한다. 컨볼루션 계층에서는 하나의 입력에 서로 다른 종류의 필터를 사용할 수 있다.
- [0032] 컨볼루션 계층은 입력이미지의 특정 영역을 대상으로 컨볼루션 연산을 수행한다. 연산 영역을 윈도우(window)라고 부른다. 윈도우는 영상의 좌측 상단에서 우측 하단까지 한 칸씩 이동할 수 있고, 한 번에 이동하는 이동 크기를 조절할 수 있다. 이동 크기를 스트라이드(stride)라고 한다. 컨볼루션 계층은 입력이미지에서 윈도우를 이동하면서 입력이미지의 모든 영역에 대하여 컨볼루션 연산을 수행한다. 한편 컨볼루션 계층은 영상의 가장 자리에 패딩(padding)을 하여 컨볼루션 연산 후 입력 영상의 차원을 유지시킨다.
- [0033] 풀링 계층은 컨볼루션 계층의 연산 결과로 얻은 특징맵을 서브 샘플링(sub sampling)한다. 풀링 연산은 최대 풀링(max pooling)과 평균 풀링(average pooling) 등이 있다. 최대 풀링은 윈도우 내에서 가장 큰 샘플 값을 선택한다. 평균 풀링은 윈도우에 포함된 값의 평균 값으로 샘플링한다. 일반적으로 풀링은 스트라이드와 윈도우의 크기가 같도록 하는 것일 일반적이다.
- [0034] 비선형 연산 계층(nonlinear operation layer)은 뉴런(노드)에서 출력값을 결정하는 계층이다. 비선형 연산 계층은 전달 함수(transfer function)를 사용한다. 전달 함수는 Relu, sigmoid 함수 등이 있다.
- [0035] 컨볼루션 인코더는 소스 영상에 대한 특징 맵을 생성한다.
- [0036] 컨볼루션 디코더는 컨볼루션 인코더가 생성한 특징 맵을 이용하여 일정한 영상을 생성한다. 컨볼루션 디코더는 복수의 계층으로 구성된다. 도 2는 5개의 계층을 갖는 컨볼루션 디코더를 예로 도시하였다.
- [0037] 하나의 계층은 업샘플링 계층(upsampling layer) 및 역컨볼루션 계층(deconvolutional layer)을 갖는다. 역컨볼루션 계층은 컨볼루션 인코더의 컨볼루션 계층의 역동작을 수행한다. 역컨볼루션 계층은 컨볼루션 계층, 배치 표준화 계층, 비 선형화 계층의 구조가 반복될 수 있다. 다만 컨볼루션 디코더는 컨볼루션 인코더와 대칭적인 구조를 갖기에, 컨볼루션 인코더와 동일한 개수의 컨볼루션 계층, 동일한 컨볼루션 필터 크기 및 개수를 갖는다.
- [0038] 업샘플링 계층은 풀링 계층의 역동작을 수행한다. 업샘플링 계층은 업샘플링(upsampling)을 진행한다. 업샘플링 계층은 풀링 계층과 다르게 반대로 차원을 확대하는 역할을 한다.
- [0039] 역컨볼루션 계층은 컨볼루션 계층의 역동작을 수행한다. 역컨볼루션 계층은 컨볼루션 계층과 반대 방향으로 컨볼루션 연산을 수행한다. 역컨볼루션 계층은 입력으로 특징맵을 받아 커널을 이용한 컨볼루션 연산으로 출력 영상을 생성한다. 스트라이드를 1로 하면 역컨볼루션 계층은 특징맵의 가로, 세로 크기가 출력의 가로, 세로와 동일한 영상을 출력한다. 스트라이드를 2로 하면 역컨볼루션 계층은 특징맵의 가로, 세로 크기 대비 절반 크기의 영상을 출력한다.
- [0041] 도 3은 회귀 딥러닝 모델의 학습 네트워크 구조에 대한 예이다. 회귀 딥러닝 모델은 컨볼루션 인코더-디코더 구조를 갖는다. 도 3에서 학습 네트워크(N)은 회귀 딥러닝 모델인 학습 네트워크를 의미한다. 학습 네트워크(N)은 인코더(N1) 및 디코더(N2)를 포함한다.
- [0042] 인코더(N1)는 소스 영상을 입력받아 특징 맵을 출력한다. 디코더(N2)는 입력받은 특징 맵을 기준으로 특정 영상을 출력한다. 디코더(N2)가 출력하는 영상을 출력 영상이라고 명명한다.
- [0043] 인코더(N1)는 소스 영상의 특징 맵을 생성하고, 특징 맵의 크기를 감소시키면서 일부 유효한 특징값만을 추출한다. 디코더(N2)는 인코더(N1)가 출력하는 유효한 특징값을 기준으로 특징 맵의 크기를 증가시킨다. 디코더(N2)는 최종적으로 소스 영상과 동일한 크기의 출력 영상을 출력한다.
- [0044] 인코더(N1)는 복수의 인코더 단을 포함한다. 도 3은 3개의 인코더 단(encoder 1, encoder 2 및 encoder 3)을 예로 도시하였다.
- [0045] 인코더 단은 복수의 컨볼루션 블록과 하나의 풀링 계층으로 구성될 수 있다. 하나의 컨볼루션 블록은 컨볼루션 계층, 비선형화 계층 및 표준화 계층을 포함한다. 비선형화 계층은 relu layer를 사용할 수 있다. 표준화 계층은 배치 표준화 계층을 사용할 수 있다. 각 계층의 동작은 전술한 바와 같다.

[0046] 풀링 계층은 최대 풀링(max pooling)을 수행할 수 있다. 풀링 계층은 최댓값을 갖는 픽셀 위치를 저장해두고, 대응되는 디코더 단의 업 샘플링 계층에 전달한다(도 3에서 점선 화살표로 표시함).

[0047] 디코더(N2)는 복수의 디코더 단을 포함한다. 도 3은 3개의 디코더 단(decoder 1, decoder 2 및 decoder 3)을 예로 도시하였다. 전술한 바와 같이 디코더(N1)는 인코더(N1)에 대칭되는 구조(거울상)를 갖는다. 따라서 디코더 단은 하나의 업샘플링 계층 및 복수(컨볼루션 블록의 개수와 동일한)의 역컨볼루션 블록으로 구성될 수 있다. 디코더 단은 인코더 단의 동작을 역으로 수행한다.

[0048] 업샘플링 계층은 대응되는 인코더 단의 최대 풀링 계층으로부터 전달 받은 최댓값 픽셀 위치에 입력 특징 맵의 값을 출력하고, 최댓값 픽셀 위치 외에는 '0'의 값을 출력한다. 디코더 단은 대응되는 인코더 단의 컨볼루션 블록 및 컨볼루션 계층과 필터 개수, 필터 크기 등이 동일하다.

[0050] 딥 러닝 모델은 파라미터(parameter)는 손실함수를 감소시키는 방향으로 학습을 수행한다. 손실함수는 학습 과정을 통해 신경망의 가중치를 최적화할 수 있다. 예컨대, 가중치 최적화는 경사 하강법(gradient descent method)을 이용할 수 있다. 영상처리장치는 정의된 손실함수에 대하여 학습하고자 하는 각 파라미터의 기울기(gradient) $d_{\text{parameter}}$ 를 계산한다. 영상처리장치는 기울기를 최소화하는 값을 찾아 가중치를 최적화할 수 있다. 영상처리장치는 각 파라미터의 기울기를 이용하여 각 파라미터의 값 업데이트할 수 있다. 업데이트되는 파라미터 $\text{parameter}_{\text{new}}$ 는 다음과 같이 결정될 수 있다. $\text{parameter}_{\text{new}} = \text{parameter}_{\text{prev}} - \text{learnRate} * d_{\text{parameter}}$ 이다. 여기서 학습률(learnRate)은 모델 학습 시에 직접 설정하는 설정 값이다. 도 3의 학습 네트워크는 아래의 수학적 식 1과 같은 손실함수 L을 이용할 수 있다.

수학적 식 1

$$L = \frac{1}{ndcr} \sum_{l=1}^n \sum_{k=1}^d \sum_{j=1}^c \sum_{i=1}^r (Y_{i,j,k,l} - T_{i,j,k,l})^2 + \lambda R$$

[0054] 수학적 식 1에서 n은 학습에 사용한 데이터 수, d는 채널 수, c는 열 개수, r은 행 개수이다. 열 및 행은 영상에서의 픽셀 위치를 나타낸다. Y는 출력 데이터, T는 정답 데이터이다. R은 정규화 항이다. 손실함수는 출력 데이터와 정답 데이터 간의 MSE(Mean Squared Error)와 정규화 항을 합한 형태를 갖는다. MSE는 결국 출력 데이터가 정답 데이터와 같아지도록 한다. λ 는 하이퍼 파라미터(hyper parameter)로 사전에 설정되는 값이다. 손실함수의 정규화항에 대한 가중치 결정한다.

[0055] 아래 수학적 식 2는 정규화 항을 나타낸다.

수학적 식 2

$$R = \frac{1}{ndcr} \sum_{l=1}^n \sum_{k=1}^d \sum_{j=1}^c \sum_{i=1}^r \sqrt{(Y_{i,j,k,l} - Y_{i-1,j,k,l})^2 + (Y_{i,j,k,l} - Y_{i,j-1,k,l})^2 + \epsilon^2}$$

[0059] 수학적 식 2에서 ϵ 는 제곱근 내부의 값이 0이 되는 상황을 방지하기 위한 임의의 작은 양수를 의미한다.

[0060] 정규화항 R은 출력 데이터의 값이 산발적으로 분포하지 않고, 주변 픽셀과 비슷한 값으로 출력되게 한다. 정규화항 R은 출력 데이터 Y의 행 방향 및 열 방향으로의 주변 픽셀 간의 TV(total variation)를 정규화한다. TV 정규화는 영상에서 급격하게 값이 변하는 경계 부분에서 경계를 비교적 보존하면서, 주변 픽셀 간에 비슷한 값을 갖게 한다. 정규화항은 특정 픽셀을 인접한 주변 픽셀과 일정한 범위 내에서 유사한 값을 갖게 한다.

[0061] 소스 영상의 배경은 노이즈 또는 불균일 조명 성분 등을 가질 수 있다. 정규화항은 이러한 배경에서 왜곡 성분을 제거 내지 평탄화(smoothing)하는 역할을 한다.

[0062] 도 3에 도시한 학습된 회귀 딥러닝 모델은 픽셀 단위 회귀 방법으로 각 픽셀 별로 배경과의 차이 값을

계산한다. 인코더(N1)가 픽셀 단위로 특징을 추출하는 과정에서 회귀 방식으로 배경의 노이즈를 제거하고, 노이즈가 제거된 배경과 현재 픽셀(예컨대, 특정 객체 일부)의 차이 값을 계산한다. 이를 통해 배경이 복잡하거나, 배경이 특정 패턴을 갖는 경우에도 배경과 타깃 객체를 구분한다. 디코더(N2)는 이와 같이 구분된 특징을 갖는 특징 맵에 기초하여 출력 영상을 생성한다.

[0064] 도 4는 소스 영상과 출력 영상에 대한 예이다. 도 4에서 A는 소스 영상의 예를 나타내고, B는 출력 영상의 예를 나타낸다. 출력 영상은 결과를 직관적으로 나타내기 위해 차이값을 붉은 색상 값으로 나타낸 예이다. 여기서 차이값은 해당 픽셀과 배경과의 차이를 의미한다. 붉은색 픽셀(영역)은 차이값이 가장 높은 영역을 나타내고, 노란색 픽셀은 붉은 색보다는 차이값이 낮은 영역을 나타내고, 파란색은 차이값이 없는 배경을 나타낸다.

[0066] 도 5는 객체 검출 장치(200)의 구성에 대한 예이다. 객체 검출 장치(200)는 전술한 영상처리장치에 해당한다. 객체 검출 장치(200)는 도 1의 검출 서버(150) 내지 검출 PC(180)에 해당하기도 한다.

[0067] 객체 검출 장치(200)는 회귀 딥러닝 모델을 이용하여 소스 영상에서 객체가 식별된 출력 영상을 생성한다. 객체 검출 장치(200)는 물리적으로 다양한 형태로 구현될 수 있다. 예컨대, 객체 검출 장치(200)는 PC와 같은 컴퓨터 장치, 네트워크의 서버, 영상 처리 전용 칩셋 등의 형태를 가질 수 있다. 컴퓨터 장치는 스마트 기기 등과 같은 모바일 기기를 포함할 수 있다.

[0068] 객체 검출 장치(200)는 저장 장치(210), 메모리(220), 연산장치(230), 인터페이스 장치(240) 및 통신 장치(250)를 포함한다.

[0069] 저장 장치(210)는 영상 처리를 위한 신경망 모델을 저장한다. 예컨대, 저장 장치(210)는 도 3과 같은 회귀 딥러닝 모델을 저장할 수 있다. 회귀 딥러닝 모델은 사전에 학습된 상태라고 가정한다. 나아가 저장 장치(210)는 영상 처리에 필요한 프로그램 내지 소스 코드 등을 저장할 수 있다. 저장 장치(210)는 입력된 소스 영상 및 생성된 분할 영상을 저장할 수 있다.

[0070] 메모리(220)는 객체 검출 장치(200)가 수신한 소스 영상 및 출력 영상 생성과정에서 생성되는 데이터 및 정보 등을 저장할 수 있다.

[0071] 인터페이스 장치(240)는 외부로부터 일정한 명령 및 데이터를 입력받는 장치이다. 인터페이스 장치(240)는 물리적으로 연결된 입력 장치 또는 외부 저장 장치로부터 소스 영상을 입력받을 수 있다. 인터페이스 장치(240)는 영상 처리를 위한 각종 신경망 모델을 입력받을 수 있다. 인터페이스 장치(240)는 신경망 모델 생성을 위한 학습데이터, 정보 및 파라미터값을 입력받을 수도 있다.

[0072] 통신 장치(250)는 유선 또는 무선 네트워크를 통해 일정한 정보를 수신하고 전송하는 구성을 의미한다. 통신 장치(250)는 외부 객체로부터 소스 영상을 수신할 수 있다. 통신 장치(250)는 각종 신경망 모델 및 모델 학습을 위한 데이터도 수신할 수 있다. 통신 장치(250)는 생성한 출력 영상을 외부 객체로 송신할 수 있다.

[0073] 통신 장치(250) 내지 인터페이스 장치(240)는 외부로부터 일정한 데이터 내지 명령을 전달받는 장치이다. 통신 장치(250) 내지 인터페이스 장치(240)를 입력장치라고 명명할 수 있다.

[0074] 연산 장치(230)는 저장장치(210)에 저장된 신경망 모델 내지 프로그램을 이용하여 타깃 객체가 검출된 출력 영상을 생성한다. 연산 장치(230)는 주어진 학습 데이터를 이용하여 영상 처리 과정에 사용되는 회귀 딥러닝 모델을 학습할 수 있다. 연산 장치(230)는 전술한 과정을 통해 구축된 회귀 딥러닝 모델을 이용하여 소스 영상에 대한 출력 영상을 생성할 수 있다. 연산 장치(230)는 데이터를 처리하고, 일정한 연산을 처리하는 프로세서, AP, 프로그램이 임베디드된 칩과 같은 장치일 수 있다.

[0076] 또한, 상술한 바와 같은 영상 처리 방법, 배경에서 타깃 객체를 구분한 영상을 생성하는 방법 내지 분량 검출 방법은 컴퓨터에서 실행될 수 있는 실행가능한 알고리즘을 포함하는 프로그램(또는 어플리케이션)으로 구현될 수 있다. 상기 프로그램은 비일시적 판독 가능 매체(non-transitory computer readable medium)에 저장되어 제공될 수 있다.

[0077] 비일시적 판독 가능 매체란 레지스터, 캐쉬, 메모리 등과 같이 짧은 순간 동안 데이터를 저장하는 매체가 아니라 반영구적으로 데이터를 저장하며, 기기에 의해 판독(reading)이 가능한 매체를 의미한다. 구체적으로는, 상술한 다양한 어플리케이션 또는 프로그램들은 CD, DVD, 하드 디스크, 블루레이 디스크, USB, 메모리카드, ROM 등과 같은 비일시적 판독 가능 매체에 저장되어 제공될 수 있다.

[0079] 본 실시예 및 본 명세서에 첨부된 도면은 전술한 기술에 포함되는 기술적 사상의 일부를 명확하게 나타내고 있

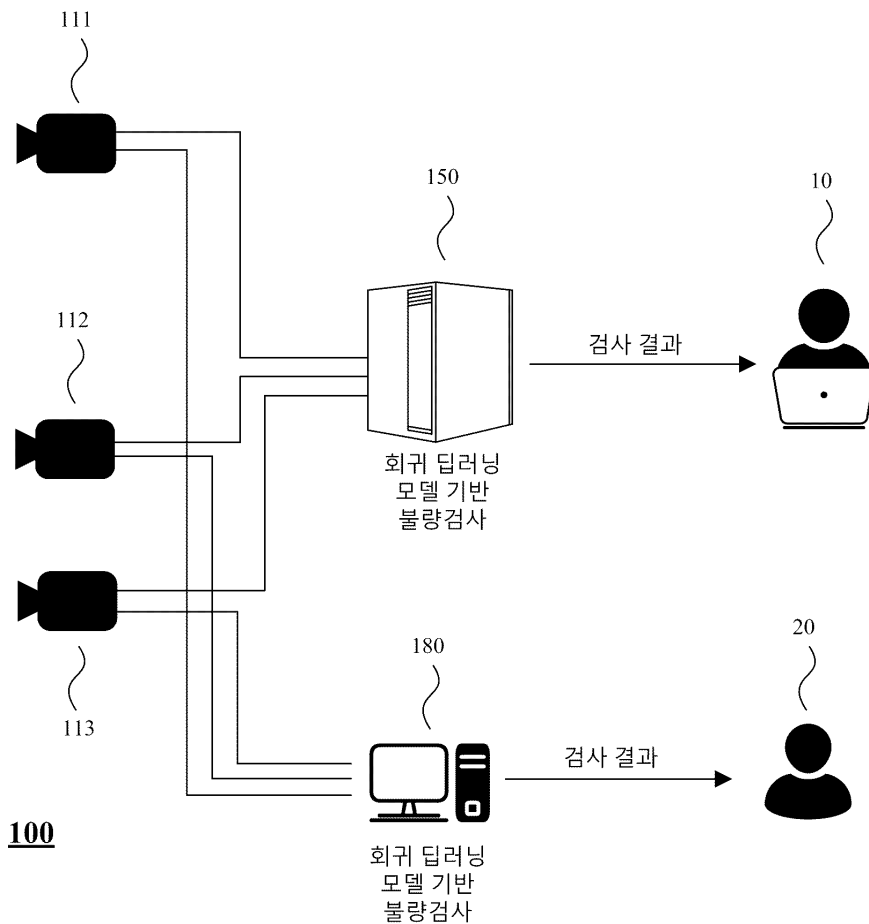
는 것에 불과하며, 전술한 기술의 명세서 및 도면에 포함된 기술적 사상의 범위 내에서 당업자가 용이하게 유추할 수 있는 변형 예와 구체적인 실시례는 모두 전술한 기술의 권리범위에 포함되는 것이 자명하다고 할 것이다.

부호의 설명

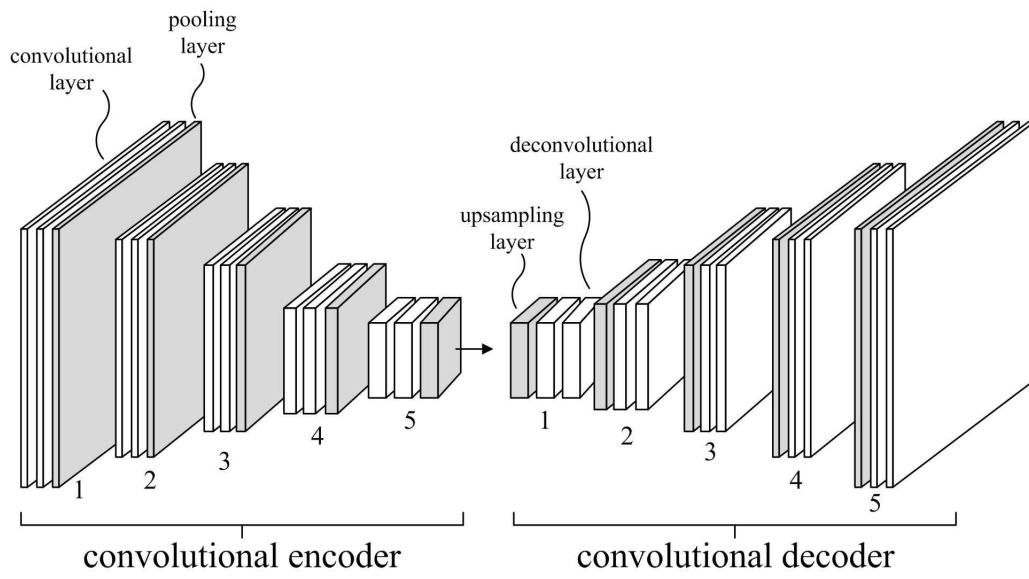
- [0080]
- 100 : 불량 검출 시스템
 - 111, 112, 113 : 카메라
 - 150 : 검출 서버
 - 180 : 검출 PC
 - 10, 20 : 사용자
 - 200 : 객체 검출 장치
 - 210 : 저장장치
 - 220 : 메모리
 - 230 : 연산장치
 - 240 : 인터페이스장치
 - 250 : 통신장치

도면

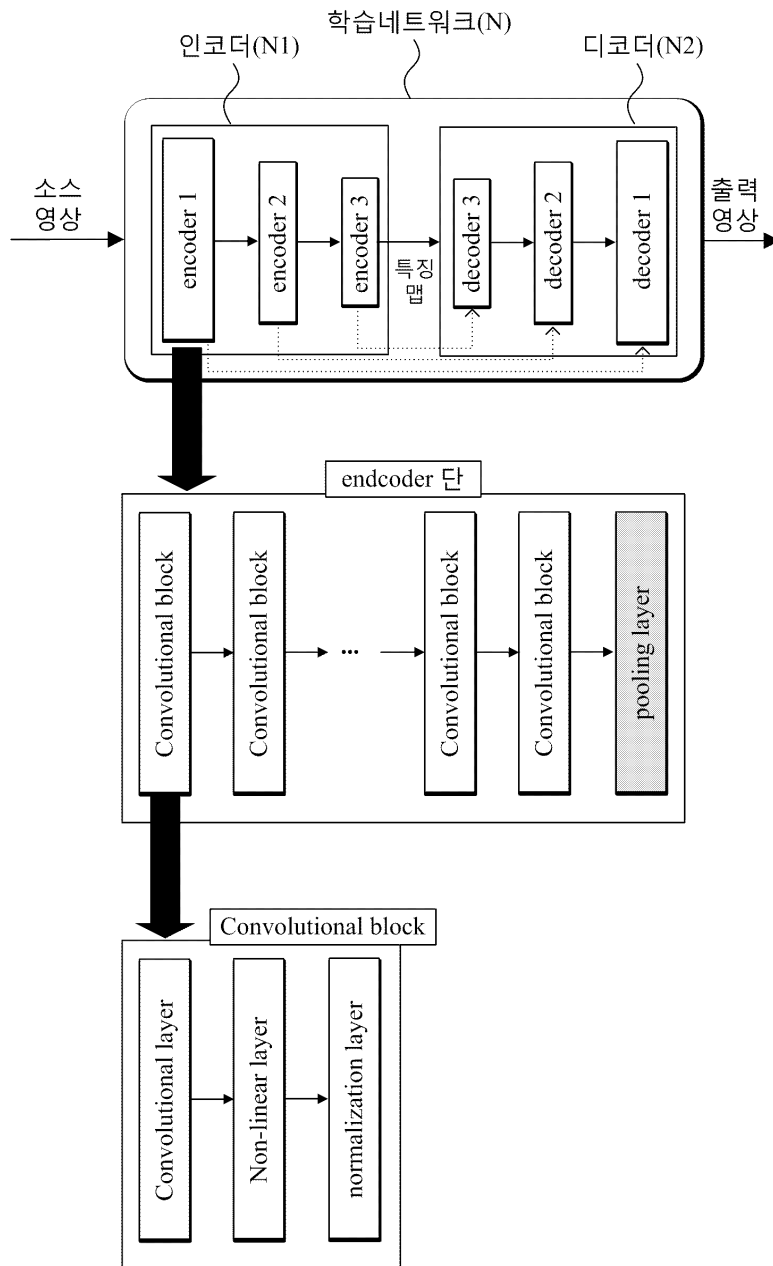
도면1



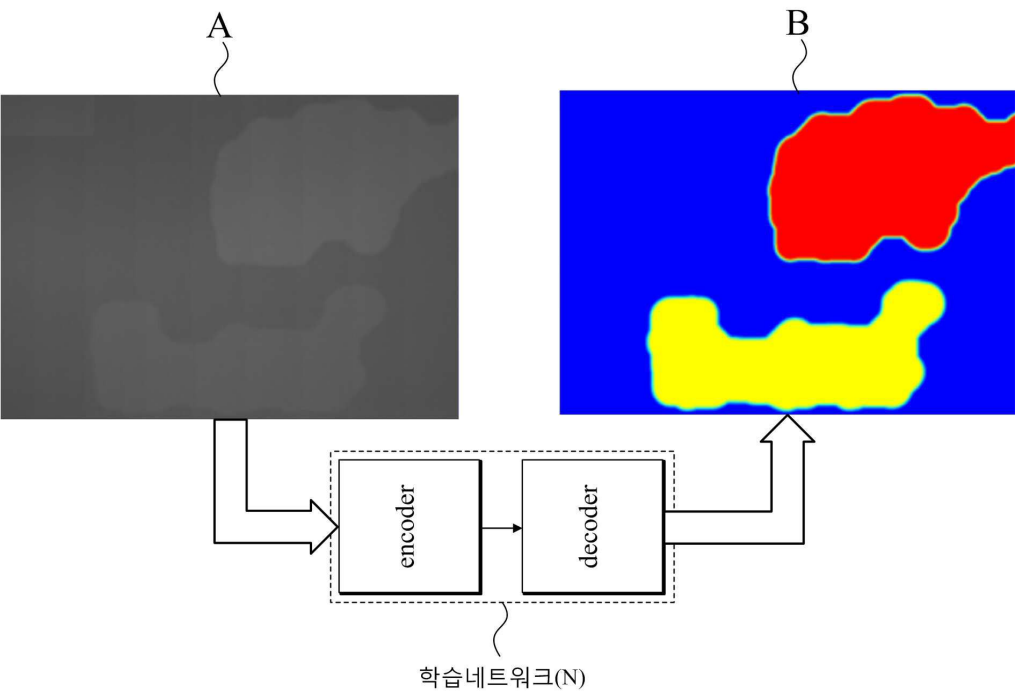
도면2



도면3



도면4



도면5

