



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2017년12월06일
 (11) 등록번호 10-1805318
 (24) 등록일자 2017년11월29일

(51) 국제특허분류(Int. Cl.)
 G06K 9/32 (2006.01) G06K 9/34 (2006.01)
 G06K 9/62 (2006.01)
 (52) CPC특허분류
 G06K 9/3258 (2013.01)
 G06K 9/342 (2013.01)
 (21) 출원번호 10-2016-0144708
 (22) 출원일자 2016년11월01일
 심사청구일자 2016년11월01일
 (56) 선행기술조사문헌
 KR101114135 B1*
 *는 심사관에 의하여 인용된 문헌

(73) 특허권자
 포항공과대학교 산학협력단
 경상북도 포항시 남구 청암로 77 (지곡동)
 (72) 발명자
 이상준
 서울특별시 양천구 목동중앙남로4길 11-17, 203호
 김상우
 경상북도 포항시 남구 지곡로 155, 7동 1003호
 (뒷면에 계속)
 (74) 대리인
 특허법인이상

전체 청구항 수 : 총 20 항

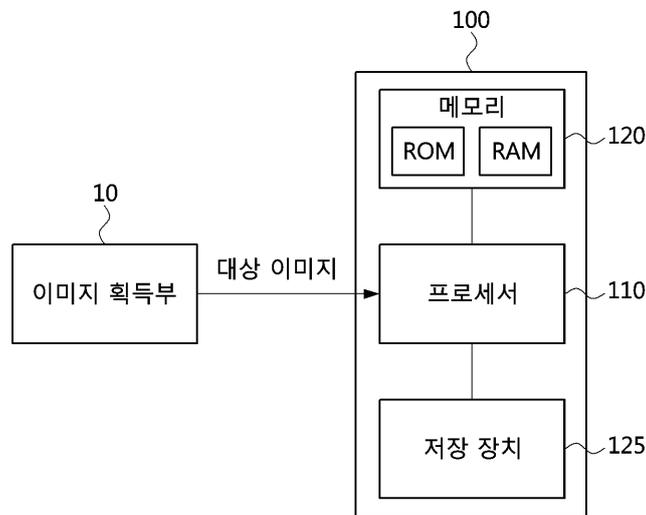
심사관 : 이종경

(54) 발명의 명칭 **텍스트 영역 식별 방법 및 장치**

(57) 요약

텍스트 영역 식별 방법 및 장치가 개시된다. 개시된 텍스트 영역 식별 방법은, 대상 이미지로부터 복수의 패치 이미지들을 추출하는 단계, 복수의 패치 이미지들 각각이 텍스트 이미지로 분류될 확률을 계산하는 단계, 및 대상 이미지의 픽셀들 각각이 텍스트 영역으로 분류될 확률을 계산하는 단계를 포함한다.

대표도 - 도1



(52) CPC특허분류

G06K 9/6232 (2013.01)

G06K 2209/01 (2013.01)

(72) 발명자

반재필

경상북도 포항시 남구 청암로 77, 4동 803호

최혜연

부산광역시 사하구 윤공단로 23-3

공지예외적용 : 있음

명세서

청구범위

청구항 1

텍스트 영역 식별장치가, 대상 이미지에서 텍스트 영역을 식별하는 방법에 있어서,

상기 대상 이미지로부터 복수의 패치 이미지들을 추출하는 단계;

상기 복수의 패치 이미지들 각각이 텍스트 이미지로 분류될 확률을 계산하는 단계; 및

상기 대상 이미지의 픽셀들 각각이 텍스트 영역으로 분류될 확률을 계산하는 단계;를 포함하며,

상기 픽셀들 각각이 텍스트 영역으로 분류될 확률을 계산하는 단계는, 상기 복수의 패치 이미지들 가운데, 미리 정의된 픽셀을 포함하는 패치 이미지들이 상기 텍스트 이미지로 분류될 확률 값들에 기초하여, 상기 미리 정의된 픽셀이 텍스트 영역으로 분류될 확률을 계산하는 텍스트 영역을 식별하는 방법.

청구항 2

청구항 1에 있어서,

상기 복수의 패치 이미지들 각각은 복수의 픽셀들을 포함하도록 추출되는, 텍스트 영역을 식별하는 방법.

청구항 3

청구항 2에 있어서,

상기 복수의 패치 이미지들 각각은, 서로 인접한 패치 이미지들이 적어도 하나의 픽셀을 공유하도록 추출되는 텍스트 영역을 식별하는 방법.

청구항 4

청구항 1에 있어서,

상기 복수의 패치 이미지들 각각은, 상기 패치 이미지들 각각이 $a \times b$ 개의 픽셀을 포함하고, 패치 이미지들 각각의 시작 픽셀들이 각각 $k \times 1$ 개의 픽셀만큼 이격되도록 추출되는 텍스트 영역을 식별하는 방법.

(a , b 은 각각 2보다 큰 자연수, k 는 $k < a$ 을 만족하는 임의의 자연수, 1 는 $1 < b$ 을 만족하는 임의의 자연수)

청구항 5

청구항 1에 있어서,

상기 복수의 패치 이미지들 각각이 텍스트 이미지로 분류될 확률은 심층 컨볼루션 신경망(Deep Convolutional Neural Networks)을 이용하여 계산되는 텍스트 영역을 식별하는 방법.

청구항 6

청구항 5에 있어서,

필터 커널을 이용하여 상기 패치 이미지에 대한 컨볼루션 이미지를 추출하는 단계; 및

상기 컨볼루션 이미지에 대한 풀링(pooling)을 실시하는 단계를 포함하는 텍스트 영역을 식별하는 방법.

청구항 7

청구항 5에 있어서,

복수의 텍스트 이미지들을 이용하여, 상기 심층 컨볼루션 신경망을 학습시키는 단계;를 더 포함하는 텍스트 영역을 식별하는 방법.

청구항 8

청구항 7에 있어서,

상기 심층 컨볼루션 신경망을 학습시키는 단계는,

상기 테스트 이미지에서 복수의 패치 이미지들을 추출하는 단계;

상기 테스트 이미지로부터 추출된 복수의 패치 이미지들 각각이 텍스트 이미지인지 여부에 대한 분류 정보를 수신하는 단계; 및

상기 분류 정보를 이용하여, 상기 심층 컨볼루션 신경망을 학습시키는 단계를 포함하는 텍스트 영역을 식별하는 방법.

청구항 9

청구항 1에 있어서,

상기 미리 정의된 픽셀이 텍스트 영역으로 분류될 확률을 계산하는 단계는,

상기 미리 정의된 픽셀을 포함하는 패치 이미지들 각각이 상기 텍스트 이미지로 분류될 확률들의 제곱 합을 계산하는 단계; 및

상기 확률들의 제곱 합을 정규화(normalize) 시키는 단계를 포함하는 텍스트 영역을 식별하는 방법.

청구항 10

청구항 1에 있어서,

상기 대상 이미지의 상기 픽셀들 각각이 상기 텍스트 영역으로 분류될 확률 값들로부터, 확률 분포맵을 생성하는 단계;

상기 확률 분포맵의 행별로, 상기 픽셀들이 텍스트 영역으로 분류될 확률 값들을 합산함으로써, 수평 프로젝션 프로파일을 계산하는 단계;

상기 확률 분포맵의 열별로, 상기 픽셀들이 텍스트 영역으로 분류될 확률 값들을 합산함으로써, 수직 프로젝션 프로파일을 계산하는 단계; 및

상기 수평 프로젝션 프로파일 및 상기 수직 프로젝션 프로파일로부터 텍스트 영역을 결정하는 단계;를 더 포함하는 텍스트 영역을 식별하는 방법.

청구항 11

대상 이미지에서 텍스트 영역을 식별하는 장치에 있어서,

프로세서(processor); 및

상기 프로세서를 통해 실행되는 적어도 하나의 명령이 저장된 메모리(memory);를 포함하고,

상기 적어도 하나의 명령은,

상기 대상 이미지로부터 복수의 패치 이미지들을 추출하고; 상기 복수의 패치 이미지들 각각이 텍스트 이미지로 분류될 확률을 계산하고; 상기 대상 이미지의 픽셀들 각각이 텍스트 영역으로 분류될 확률을 계산하도록 수행되며,

상기 각각의 픽셀이 텍스트 영역으로 분류될 확률의 계산은, 상기 복수의 패치 이미지들 가운데, 미리 정의된 픽셀을 포함하는 패치 이미지들 각각이 상기 텍스트 이미지로 분류될 확률 값들로부터 상기 미리 정의된 픽셀이 텍스트 영역으로 분류될 확률을 계산함으로써 수행되는 텍스트 영역 식별 장치.

청구항 12

청구항 11에 있어서,

상기 적어도 하나의 명령은,

상기 패치 이미지들 각각이 복수의 픽셀을 포함하도록 상기 패치 이미지들을 추출하도록 실행되는 텍스트 영역 식별 장치.

청구항 13

청구항 12에 있어서,

상기 적어도 하나의 명령은,

서로 인접한 패치 이미지들이 적어도 하나의 픽셀을 공유하도록 상기 패치 이미지들을 추출하도록 실행되는 텍스트 영역 식별 장치.

청구항 14

청구항 11에 있어서,

상기 적어도 하나의 명령은,

상기 패치 이미지들 각각이 $a \times b$ 개의 픽셀을 포함하고, 패치 이미지들의 은 $k \times l$ 픽셀만큼 이격 되도록 상기 패치 이미지들을 추출하도록 실행되는 텍스트 영역 식별 장치.

(a, b은 각각 2보다 큰 자연수, k는 $k < a$ 을 만족하는 임의의 자연수, l는 $l < b$ 을 만족하는 임의의 자연수)

청구항 15

청구항 11에 있어서,

상기 적어도 하나의 명령은,

심층 컨볼루션 신경망(Deep Convolutional Neural Networks)을 이용하여 상기 복수의 패치 이미지들 각각이 텍스트 이미지로 분류될 확률을 계산하도록 실행되는 텍스트 영역 식별 장치.

청구항 16

청구항 15에 있어서,

상기 적어도 하나의 명령은,

필터 커널을 이용하여 상기 패치 이미지에 대한 컨볼루션 이미지를 추출하고, 상기 컨볼루션 이미지에 대한 풀링(pooling)을 실시하도록 실행되는 텍스트 영역 식별 장치.

청구항 17

청구항 15에 있어서,

상기 메모리는 복수의 테스트 이미지를 저장하며,

상기 적어도 하나의 명령은,

복수의 테스트 이미지를 이용하여, 상기 심층 컨볼루션 신경망을 학습시키고, 학습 결과를 상기 메모리에 저장하도록 실행되는 텍스트 영역 식별 장치.

청구항 18

청구항 17에 있어서,

상기 적어도 하나의 명령은,

상기 테스트 이미지에서 복수의 패치 이미지들을 추출하고, 상기 테스트 이미지로부터 추출된 복수의 패치 이미지들 각각이 텍스트 이미지인지 여부에 대한 분류 정보를 입력 받고, 상기 분류 정보를 이용하여 상기 심층 컨볼루션 신경망을 학습시키도록 실행되는 텍스트 영역 식별 장치.

청구항 19

청구항 11에 있어서,

상기 적어도 하나의 명령은,

상기 미리 정의된 픽셀을 포함하는 패치 이미지들 각각이 상기 텍스트 이미지로 분류될 확률들의 제곱 합을 계산하고, 상기 확률들의 제곱 합을 정규화(normalize) 시키도록 실행되는 텍스트 영역 식별 장치.

청구항 20

청구항 11에 있어서,

상기 적어도 하나의 명령은,

상기 픽셀들 각각이 상기 텍스트 영역으로 분류될 확률 값들로부터, 확률 분포맵을 생성하고, 상기 확률 분포맵의 행 별로, 상기 픽셀들이 텍스트 영역으로 분류될 확률 값들을 합산함으로써, 수평 프로젝션 프로파일을 계산하고, 상기 확률 분포맵의 열 별로, 상기 픽셀들이 텍스트 영역으로 분류될 확률 값들을 합산함으로써, 수직 프로젝션 프로파일을 계산하고, 상기 수평 프로젝션 프로파일 및 상기 수직 프로젝션 프로파일로부터 텍스트 영역을 결정하도록 실행되는 텍스트 영역 식별 장치.

발명의 설명

기술 분야

[0001] 본 발명은 텍스트 영역 식별 방법 및 장치에 관한 것으로, 더욱 상세하게는 대상 이미지에서 텍스트가 존재하는 영역을 식별하는 방법 및 장치에 관한 것이다.

배경 기술

[0002] 최근 무인화 기술의 발전에 따라 촬영 영상에서 텍스트 이미지를 인식하는 기술이 요구되고 있다. 예를 들어, 제철소와 같은 산업 현장에서는 많은 양의 슬라브 관리를 자동화 시키기 위해, 관리번호를 자동으로 인식하는 기술이 요구되고 있다.

[0003] 영상 이미지에서 텍스트를 인식하기 위해 여러 가지 영상 처리 기법들이 개발되고 있다. 그 중 하나로 규칙 기반의 영상처리 기법이 있다. 하지만, 규칙 기반의 영상 처리 기법은 기반이 되는 수치 설계가 요구되며, 수치들을 설계하는 전문가의 영향을 많이 받는다. 또한, 영상 이미지의 가변성을 반영하지 못하여 정확도가 떨어지는 문제가 있다.

[0004] 이러한 문제점을 보완하기 위해 최근 기계학습 기반의 영상처리 기법이 개발되고 있다. 하지만, 이 경우도 영상 이미지에서 바로 텍스트를 인식하기 위해서는 과도하게 많은 데이터 처리량이 요구된다. 따라서, 효율적인 영상 처리를 위해 대상 이미지에서 텍스트가 존재하는 텍스트 영역을 먼저 식별하는 기술이 요구된다.

[0005] 일반적인 텍스트 인식에 있어서 개별 문자의 분류는 기존의 영상처리 및 기계학습 방법으로도 수행 가능하며, 복잡한 영상에서 문자열 인식의 정확도 하락은 주로 문자 영역을 찾아내는 과정에서 발생한다. 이에, 제철소와 같은 고온 환경에서 이미지의 품질이 좋지 않은 경우에는 텍스트 영역을 정확하고 빠르게 식별하는 기술이 요구된다.

발명의 내용

해결하려는 과제

[0006] 상기와 같은 문제점을 해결하기 위한 본 발명의 목적은 대상 이미지에서 높은 신뢰도로 텍스트 영역을 식별하는 방법 및 장치를 제공하는 데 있다.

과제의 해결 수단

[0007] 일 실시예에 따르면,

[0008] 텍스트 영역 식별장치가, 대상 이미지에서 텍스트 영역을 식별하는 방법에 있어서,

[0009] 상기 대상 이미지로부터 복수의 패치 이미지들을 추출하는 단계;

[0010] 상기 복수의 패치 이미지들 각각이 텍스트 이미지로 분류될 확률을 계산하는 단계; 및

- [0011] 상기 대상 이미지의 픽셀들 각각이 텍스트 영역으로 분류될 확률을 계산하는 단계;를 포함하며,
- [0012] 상기 픽셀들 각각이 텍스트 영역으로 분류될 확률을 계산하는 단계는, 상기 복수의 패치 이미지들 가운데, 미리 정의된 픽셀을 포함하는 패치 이미지들이 상기 텍스트 이미지로 분류될 확률 값들에 기초하여, 상기 미리 정의된 픽셀이 텍스트 영역으로 분류될 확률을 계산하는 텍스트 영역을 식별하는 방법이 개시된다.
- [0013] 여기서, 상기 복수의 패치 이미지들 각각은 복수의 픽셀들을 포함하도록 추출될 수 있다.
- [0014] 여기서, 상기 복수의 패치 이미지들 각각은, 서로 인접한 패치 이미지들이 적어도 하나의 픽셀을 공유하도록 추출될 수 있다.
- [0015] 여기서, 상기 복수의 패치 이미지들 각각은, 상기 패치 이미지들 각각이 $a \times b$ 개의 픽셀을 포함하고, 패치 이미지들 각각의 시작 픽셀들이 각각 $k \times 1$ 개의 픽셀만큼 이격되도록 추출될 수 있다.
- [0016] (a , b 은 각각 2보다 큰 자연수, k 는 $k < a$ 을 만족하는 임의의 자연수, 1 는 $1 < b$ 을 만족하는 임의의 자연수)
- [0017] 여기서, 상기 복수의 패치 이미지들 각각이 텍스트 이미지로 분류될 확률은 심층 컨볼루션 신경망(Deep Convolutional Neural Networks)을 이용하여 계산될 수 있다.
- [0018] 상기 텍스트 영역을 식별하는 방법은, 필터 커널을 이용하여 상기 패치 이미지에 대한 컨볼루션 이미지를 추출하는 단계; 및
- [0019] 상기 컨볼루션 이미지에 대한 풀링(pooling)을 실시하는 단계를 포함할 수 있다.
- [0020] 상기 텍스트 영역을 식별하는 방법은, 복수의 테스트 이미지들을 이용하여, 상기 심층 컨볼루션 신경망을 학습시키는 단계;를 더 포함할 수 있다.
- [0021] 상기 심층 컨볼루션 신경망을 학습시키는 단계는,
- [0022] 상기 테스트 이미지에서 복수의 패치 이미지들을 추출하는 단계;
- [0023] 상기 테스트 이미지로부터 추출된 복수의 패치 이미지들 각각이 텍스트 이미지인지 여부에 대한 분류 정보를 수신하는 단계; 및
- [0024] 상기 분류 정보를 이용하여, 상기 심층 컨볼루션 신경망을 학습시키는 단계를 포함할 수 있다.
- [0025] 상기 미리 정의된 픽셀이 텍스트 영역으로 분류될 확률을 계산하는 단계는,
- [0026] 상기 미리 정의된 픽셀을 포함하는 패치 이미지들 각각이 상기 텍스트 이미지로 분류될 확률들의 제곱 합을 계산하는 단계; 및
- [0027] 상기 확률들의 제곱 합을 정규화(normalize) 시키는 단계를 포함할 수 있다.
- [0028] 상기 텍스트 영역을 식별하는 방법은, 상기 대상 이미지의 상기 픽셀들 각각이 상기 텍스트 영역으로 분류될 확률 값들로부터, 확률 분포맵을 생성하는 단계;
- [0029] 상기 확률 분포맵의 행별로, 상기 픽셀들이 텍스트 영역으로 분류될 확률 값들을 합산함으로써, 수평 프로젝션 프로파일을 계산하는 단계;
- [0030] 상기 확률 분포맵의 열별로, 상기 픽셀들이 텍스트 영역으로 분류될 확률 값들을 합산함으로써, 수직 프로젝션 프로파일을 계산하는 단계; 및
- [0031] 상기 수평 프로젝션 프로파일 및 상기 수직 프로젝션 프로파일로부터 텍스트 영역을 결정하는 단계;를 더 포함할 수 있다.
- [0032] 다른 실시예에 따르면,
- [0033] 대상 이미지에서 텍스트 영역을 식별하는 장치에 있어서,
- [0034] 프로세서(processor); 및
- [0035] 상기 프로세서를 통해 실행되는 적어도 하나의 명령이 저장된 메모리(memory);를 포함하고,
- [0036] 상기 적어도 하나의 명령은,
- [0037] 상기 대상 이미지로부터 복수의 패치 이미지들을 추출하고; 상기 복수의 패치 이미지들 각각이 텍스트 이미지로

분류될 확률을 계산하고; 상기 대상 이미지의 픽셀들 각각이 텍스트 영역으로 분류될 확률을 계산하도록 수행되며,

- [0038] 상기 각각의 픽셀이 텍스트 영역으로 분류될 확률의 계산은, 상기 복수의 패치 이미지들 가운데, 미리 정의된 픽셀을 포함하는 패치 이미지들 각각이 상기 텍스트 이미지로 분류될 확률 값들로부터 상기 미리 정의된 픽셀이 텍스트 영역으로 분류될 확률을 계산함으로써 수행되는 텍스트 영역 식별 장치가 개시된다.
- [0039] 상기 적어도 하나의 명령은,
- [0040] 상기 패치 이미지들 각각이 복수의 픽셀을 포함하도록 상기 패치 이미지들을 추출하도록 실행될 수 있다.
- [0041] 여기서, 상기 적어도 하나의 명령은,
- [0042] 서로 인접한 패치 이미지들이 적어도 하나의 픽셀을 공유하도록 상기 패치 이미지들을 추출하도록 실행될 수 있다.
- [0043] 여기서, 상기 적어도 하나의 명령은,
- [0044] 상기 패치 이미지들 각각이 $a \times b$ 개의 픽셀을 포함하고, 패치 이미지들의 은 $k \times l$ 픽셀만큼 이격 되도록 상기 패치 이미지들을 추출하도록 실행될 수 있다.
- [0045] (a , b 은 각각 2보다 큰 자연수, k 는 $k < a$ 을 만족하는 임의의 자연수, l 는 $l < b$ 을 만족하는 임의의 자연수)
- [0046] 여기서, 상기 적어도 하나의 명령은,
- [0047] 심층 컨볼루션 신경망(Deep Convolutional Neural Networks)을 이용하여 상기 복수의 패치 이미지들 각각이 텍스트 이미지로 분류될 확률을 계산하도록 실행될 수 있다.
- [0048] 여기서, 상기 적어도 하나의 명령은,
- [0049] 필터 커널을 이용하여 상기 패치 이미지에 대한 컨볼루션 이미지를 추출하고, 상기 컨볼루션 이미지에 대한 풀링(pooling)을 실시하도록 실행될 수 있다.
- [0050] 여기서, 상기 메모리는 복수의 테스트 이미지를 저장하며,
- [0051] 상기 적어도 하나의 명령은,
- [0052] 복수의 테스트 이미지를 이용하여, 상기 심층 컨볼루션 신경망을 학습시키고, 학습 결과를 상기 메모리에 저장하도록 실행될 수 있다.
- [0053] 여기서, 상기 적어도 하나의 명령은,
- [0054] 상기 테스트 이미지에서 복수의 패치 이미지들을 추출하고, 상기 테스트 이미지로부터 추출된 복수의 패치 이미지들 각각이 텍스트 이미지인지 여부에 대한 분류 정보를 입력 받고, 상기 분류 정보를 이용하여 상기 심층 컨볼루션 신경망을 학습시키도록 실행될 수 있다.
- [0055] 여기서, 상기 적어도 하나의 명령은,
- [0056] 상기 미리 정의된 픽셀을 포함하는 패치 이미지들 각각이 상기 텍스트 이미지로 분류될 확률들의 제곱 합을 계산하고, 상기 확률들의 제곱 합을 정규화(normalize) 시키도록 실행될 수 있다.
- [0057] 여기서, 상기 적어도 하나의 명령은,
- [0058] 상기 픽셀들 각각이 상기 텍스트 영역으로 분류될 확률 값들로부터, 확률 분포맵을 생성하고, 상기 확률 분포맵의 행 별로, 상기 픽셀들이 텍스트 영역으로 분류될 확률 값들을 합산함으로써, 수평 프로젝션 프로파일을 계산하고, 상기 확률 분포맵의 열 별로, 상기 픽셀들이 텍스트 영역으로 분류될 확률 값들을 합산함으로써, 수직 프로젝션 프로파일을 계산하고, 상기 수평 프로젝션 프로파일 및 상기 수직 프로젝션 프로파일로부터 텍스트 영역을 결정하도록 실행될 수 있다.

발명의 효과

- [0059] 개시된 실시예들에 따르면, 미리 정의된 픽셀이 텍스트 영역으로 분류될 확률을 계산할 때, 인접한 픽셀들의 특성이 함께 고려될 수 있다. 따라서, 텍스트 영역 식별 결과의 신뢰도가 높아질 수 있다. 또한, 대상 이미지 전체가 아닌, 대상 이미지의 부분 이미지에 해당하는 패치 이미지를 이용하여 심층 컨볼루션 신경망을 학습 시

키기 때문에, 적은 양의 데이터를 이용하여 심층 컨볼루션 신경망 학습이 가능하다.

도면의 간단한 설명

[0060]

- 도 1은 예시적인 실시예에 따른 텍스트 영역 식별 장치를 나타낸 블록도이다.
- 도 2는 도 1에서 나타낸 텍스트 영역 식별 장치가 대상 이미지에서 텍스트 영역을 식별하는 방법을 나타낸 흐름도이다.
- 도 3은 프로세서가 대상 이미지에서 제1 패치 이미지를 추출하는 것을 나타낸 개념도이다.
- 도 4는 프로세서가 대상 이미지에서 제2 패치 이미지를 추출하는 것을 나타낸 개념도이다.
- 도 5는 심층 컨볼루션 신경망의 컨볼루션 레이어 의해 이미지의 특성을 추출하는 것을 나타내는 개념도이다.
- 도 6은 프로세서가 도 5에서 나타낸 이미지로부터 컨볼루션 이미지를 추출한 것을 나타낸 개념도이다.
- 도 7은 프로세서가 심층 컨볼루션 신경망을 이용해서 패치 이미지가 텍스트 이미지로 분류될 확률을 계산하는 것을 나타낸 개념도이다.
- 도 8은 도 7에서 나타낸 심층 컨볼루션 신경망의 변형예이다.
- 도 9는 도 7 및 도 8에서 나타낸 심층 컨볼루션 신경망의 변형예이다.
- 도 10은 프로세서가 픽셀이 텍스트 영역으로 분류될 확률을 계산하는 것을 개념적으로 나타낸 개념도이다.
- 도 11은 제철소에서 촬영된 대상 이미지를 나타낸다.
- 도 12는 추가적인 실시예에 따른 텍스트 영역 식별 방법을 나타낸 흐름도이다.
- 도 13은 대상 이미지와 대상 이미지로부터 생성된 확률 분포맵을 예시적으로 나타낸 도면이다.
- 도 14는 도 12에서 나타낸 확률 분포맵으로부터 계산된 수평 프로젝션 프로파일과 수직 프로젝션 프로파일을 나타낸 개념도이다.
- 도 15는 텍스트 영역 식별 방법에 대한 제2 실시예를 나타낸 흐름도이다.
- 도 16은 도 15의 S110 단계를 보다 상세하게 나타낸 흐름도이다.
- 도 17은 테스트 이미지의 일 예이다.
- 도 18은 도 17의 테스트 이미지에 포함된 패치 이미지들을 확대하여 나타낸 도면이다.

발명을 실시하기 위한 구체적인 내용

[0061]

본 발명은 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세하게 설명하고자 한다. 그러나, 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.

[0062]

제1, 제2 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 본 발명의 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.

[0063]

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.

[0064]

본 출원에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 출원에서, "포함하다" 또는 "가지다" 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조

합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

- [0065] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가지고 있다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥 상 가지는 의미와 일치하는 의미를 가진 것으로 해석되어야 하며, 본 출원에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.
- [0066] 이하, 첨부한 도면들을 참조하여, 본 발명의 바람직한 실시예를 보다 상세하게 설명하고자 한다. 본 발명을 설명함에 있어 전체적인 이해를 용이하게 하기 위하여 도면상의 동일한 구성요소에 대해서는 동일한 참조부호를 사용하고 동일한 구성요소에 대해서 중복된 설명은 생략한다.
- [0067] 본 명세서에서 대상 이미지는 분석 대상이 되는 이미지를 의미하며, 영상 이미지 혹은 사진 이미지를 포함할 수 있다. 대상 이미지의 텍스트 영역이란 대상 이미지에서 텍스트가 나타난 영역을 의미한다. 테스트 이미지란 프로세서(110)에 의해 수행되는 딥러닝 알고리즘을 학습시키기 위한 이미지를 의미한다. 메모리(120)는 테스트 이미지에 대해 미리 부여된 라벨링 정보를 저장할 수 있다. 패치 이미지란, 대상 이미지 또는 테스트 이미지에서 추출된 이미지를 의미한다. 대상 이미지에서 추출된 패치 이미지는 대상 이미지의 부분 이미지로서 대상 이미지보다 작은 수의 픽셀을 포함한다. 테스트 이미지에서 추출된 패치 이미지는 테스트 이미지의 부분 이미지로서 테스트 이미지보다 작은 수의 픽셀을 포함한다. 패치 이미지는 패치 이미지 내에 텍스트 영역을 포함하는 정도에 따라서 텍스트 이미지와 비-텍스트 이미지로 분류된다.
- [0068] 도 1은 예시적인 실시예에 따른 텍스트 영역 식별 장치(100)를 나타낸 블록도이다.
- [0069] 도 1을 참조하면, 예시적인 실시예에 따른 텍스트 영역 식별 장치(100) 적어도 하나의 프로세서(110), 메모리(120) 및 저장 장치(125) 등을 포함할 수 있다.
- [0070] 프로세서(110)는 메모리(120) 및/또는 저장 장치(125)에 저장된 프로그램 명령(program command)을 실행할 수 있다. 프로세서(110)는 중앙 처리 장치(central processing unit; CPU), 그래픽 처리 장치(graphics processing unit; GPU) 또는 본 발명에 따른 방법들이 수행되는 전용의 프로세서를 의미할 수 있다. 메모리(120)와 저장 장치(160)는 휘발성 저장 매체 및/또는 비휘발성 저장 매체로 구성될 수 있다. 예를 들어, 메모리(120)는 읽기 전용 메모리(read only memory; ROM) 및/또는 랜덤 액세스 메모리(random access memory; RAM)로 구성될 수 있다.
- [0071] 메모리(120)는 프로세서(110)를 통해 실행되는 적어도 하나의 명령을 저장하고 있을 수 있다. 적어도 하나의 명령은 대상 이미지로부터 복수의 패치 이미지들을 추출하는 명령과, 복수의 패치 이미지들 각각이 텍스트 이미지로 분류될 확률을 계산하는 명령, 및 대상 이미지의 픽셀들 각각이 텍스트 영역으로 분류될 확률을 계산하는 명령을 포함할 수 있다.
- [0072] 메모리(120)에 저장된 명령들은 프로세서(110)의 기계 학습에 의해 업데이트될 수 있다. 프로세서(110)는 기계 학습에 의해 메모리에 저장된 명령을 변경할 수 있다. 프로세서(110)가 수행하는 기계 학습은 지도 학습 방식에 의해 이루어질 수 있다.
- [0073] 프로세서(110)는 메모리(120)에 저장된 명령에 따라 이미지 획득부(10)로부터 대상 이미지를 수신할 수 있다. 이미지 획득부(10)는 촬영 장치일 수 있다. 이미지 획득부(10)는 촬영한 이미지를 대상 이미지로 프로세서(110)에 전송할 수 있다. 다른 예로, 이미지 획득부(10)는 사용자로부터 이미지를 수신하고, 수신된 이미지를 대상 이미지로 프로세서(110)에 전송할 수 있다.
- [0074] 프로세서(110)는 대상 이미지로부터 복수의 패치 이미지들을 추출할 수 있다. 패치 이미지는 대상 이미지의 부분 이미지로서 대상 이미지보다 작은 수의 픽셀을 포함할 수 있다. 서로 인접한 패치 이미지들은 적어도 하나의 픽셀을 공유할 수 있다. 프로세서(110)는 복수의 패치 이미지들 각각이 텍스트 이미지로 분류될 확률을 계산할 수 있다. 즉, 프로세서(110)는 패치 이미지들 각각이 유효한 텍스트 영역을 포함할 확률을 계산할 수 있다. 프로세서(110)는 후술하는 심층 컨볼루션 신경망(Deep Convolutional Neural Network)을 이용하여 패치 이미지들 각각이 텍스트 이미지로 분류될 확률을 계산할 수 있다. 프로세서(110)는 패치 이미지들 각각이 텍스트 이미지로 분류될 확률 값들에 기초하여, 대상 이미지의 픽셀들 각각이 텍스트 영역으로 분류될 확률을 계산할 수 있다. 상술한 과정들에 의해 프로세서(110)는 대상 이미지에서 텍스트가 존재하는 텍스트 영역을 결정할 수 있다. 그리고, 프로세서(110)는 텍스트 영역에 대해서만 텍스트 인식 메커니즘을 적용하여, 대상 이미지에 포함된 문

자열이 무엇인지를 알아낼 수 있다.

- [0075] 이하에서는 도 2를 참조하여 설명한 텍스트 영역 식별 장치(100)가 대상 이미지에서 텍스트 영역을 식별하는 방법에 관하여 설명한다.
- [0076] 도 2는 도 1에서 나타낸 텍스트 영역 식별 장치(100)가 대상 이미지에서 텍스트 영역을 식별하는 방법을 나타낸 흐름도이다.
- [0077] 도 2를 참조하면, 텍스트 영역 식별방법은, 대상 이미지로부터 복수의 패치 이미지들을 추출하는 단계(S120), 복수의 패치 이미지들 각각이 텍스트 이미지로 분류될 확률을 계산하는 단계(S130) 및 대상 이미지의 픽셀들 각각이 텍스트 영역으로 분류될 확률을 계산하는 단계(S140)를 포함할 수 있다.
- [0078] S120 단계에서, 프로세서(110)는 이미지 획득부(10)로부터 대상 이미지를 수신할 수 있다.
- [0079] 대상 이미지가 영상인 경우, 프로세서(110)는 대상 이미지의 프레임 별로 텍스트 영역을 식별할 수 있다. 프로세서(110)는 대상 이미지에서 복수의 패치 이미지들을 추출할 수 있다. 패치 이미지는 복수의 픽셀을 포함할 수 있다. 패치 이미지는 대상 이미지보다 작은 수의 픽셀을 포함할 수 있다.
- [0080] 도 3은 프로세서(110)가 대상 이미지에서 제1 패치 이미지를 추출하는 것을 나타낸 개념도이다.
- [0081] 도 3을 참조하면, 프로세서(110)가 대상 이미지(TI)에서 제1 패치 이미지(PI1)를 추출할 수 있다. 대상 이미지(TI)는 10×10 개의 픽셀을 포함할 수 있다. 제1 패치 이미지(PI1)는 4×6 개의 픽셀을 포함할 수 있다. 상기 수치들은 설명의 용이성을 위해 예시적으로 제시한 것에 불과하며, 실제 픽셀 개수와 상이할 수 있다.
- [0082] 도 4는 프로세서(110)가 대상 이미지에서 제2 패치 이미지를 추출하는 것을 나타낸 개념도이다.
- [0083] 도 4를 참조하면, 프로세서(110)가 대상 이미지(TI)에서 제2 패치 이미지(PI2)를 추출할 수 있다. 대상 이미지(TI)는 10×10 개의 픽셀을 포함할 수 있다. 제2 패치 이미지(PI2)는 4×6개의 픽셀을 포함할 수 있다. 상기 수치들은 설명의 용이성을 위해 예시적으로 제시한 것에 불과하며, 실제 픽셀 개수와 상이할 수 있다.
- [0084] 도 3 및 도 4를 참조하면, 서로 인접하는 제1 패치 이미지(PI1)와 제2 패치 이미지(PI2)는 1×2 픽셀만큼 이격되어 있을 수 있다.
- [0085] 즉, 서로 인접하는 제1 패치 이미지(PI1)의 시작 픽셀과, 제2 패치 이미지(PI2)의 시작 픽셀이 1×2 픽셀의 보폭만큼 이격 되어 있을 수 있다. 패치 이미지들(PI1, PI2) 각각은 보폭 영역(stride region)을 포함할 수 있다. 예를 들어, 제1 패치 이미지(PI1)는 제1 보폭 영역(SR1)을 포함하고, 제2 패치 이미지(PI2)는 제2 보폭 영역(SR2)을 포함할 수 있다. 각각의 패치 이미지에 포함된 보폭 영역은 다른 패치 이미지에 포함된 보폭 영역과 서로 중첩되지 않을 수 있다.
- [0086] 도 3 및 도 4에서 나타낸 바와 같이, 인접한 패치 이미지들(PI1, PI2) 사이의 보폭 크기는 패치 이미지들(PI1, PI2) 각각의 크기보다 작을 수 있다. 예를 들어, 패치 이미지들 각각은 a×b 개의 픽셀을 포함할 수 있다. 여기서, a, b는 각각 2보다 큰 자연수를 의미한다. 보폭 영역은 k×1 개의 픽셀을 포함할 수 있다. 패치 이미지들(PI1, PI2)은 k×1 개의 픽셀만큼 서로 이격 되어 있을 수 있다.
- [0087] 대상 이미지가 m×n 개의 픽셀을 포함할 때(m은 m>a를 만족하는 자연수, n은 n>b를 만족하는 자연수), 프로세서(110)는 대상 이미지로부터 [(m-a+k)/k]×[(n-b+1)/1]개의 패치 이미지들을 추출할 수 있다. 패치 이미지의 크기가 대상 이미지에 비해 상당히 작은 경우, m>>a, n>>b가 만족되어 [(m-a+k)/k]×[(n-b+1)/1] m/k×n/1이 될 수 있다.
- [0088] 다시 도 2를 참조하면, S130 단계에서, 프로세서(110)는 패치 이미지가 텍스트 이미지로 분류될 확률을 계산할 수 있다. 프로세서(110)는 패치 이미지의 특성을 분석함으로써, 패치 이미지가 텍스트 이미지로 분류될 확률을 계산할 수 있다. 프로세서(110)는 딥러닝 방식을 이용하여 패치 이미지의 특성을 분석할 수 있다. 예를 들어, 프로세서(110)는 심층 컨볼루션 신경망을 이용할 수 있다.
- [0089] 도 5는 심층 컨볼루션 신경망의 컨볼루션 레이어(Convolutional Layer)에 의해 이미지의 특성을 추출하는 것을 나타내는 개념도이다.
- [0090] 도 5를 참조하면, 프로세서(110)는 필터 커널(FI)을 이용하여 이미지(OI)에 대한 컨볼루션을 계산할 수 있다. 필터 커널(FI)은 이미지(OI)의 픽셀 개수보다 작은 크기의 매트릭스를 의미한다. 필터 커널(FI)의 각 성분은 가중치에 대한 정보를 포함할 수 있다. 예를 들어, 필터 커널(FI)의 (1,1) 성분은 0이다. 따라서, 컨볼루션을 계

산할 때, 필터 커널(FI)의 (1,1) 성분에 대응하는 이미지(OI)의 픽셀에 대해서 가중치를 0으로 계산할 수 있다. 다른 예로, 필터 커널(FI)의 (2,1) 성분은 1이다. 따라서, 컨볼루션을 계산할 때, 필터 커널(FI)의 (2,1) 성분에 대응하는 이미지(OI)의 픽셀에 대해서 가중치를 1로 계산할 수 있다.

[0091] 프로세서(110)는 이미지(OI) 상에서 필터 커널(FI)의 위치를 변경해가면서, 이미지(OI)에 대한 컨볼루션을 계산할 수 있다. 그리고, 프로세서(110)는 계산된 컨볼루션 값들로부터 컨볼루션 이미지를 추출할 수 있다.

[0092] 도 6은 프로세서(110)가 도 5에서 나타난 이미지(OI)로부터 컨볼루션 이미지를 추출한 것을 나타낸 개념도이다.

[0093] 도 5에서 나타난 필터 커널(FI)이 이미지(OI) 상에서 움직일 수 있는 경우의 수가 $(10-3+1) \times (10-3+1) = 8 \times 8$ 개이기 때문에, 프로세서(110)는 8×8 개의 컨볼루션 값들을 계산할 수 있다. 그리고, 8×8 개의 컨볼루션 값들로부터 도 6에서 나타난 바와 같은 8×8 픽셀 크기의 컨볼루션 이미지를 추출할 수 있다. 컨볼루션 이미지(CI)의 픽셀 개수는 원본 이미지(OI)보다 작아지게 된다. 프로세서(110)는 필터 커널(FI)을 이용함으로써, 원본 이미지의 특성이 반영된 컨볼루션 이미지를 추출할 수 있다. 프로세서(110)는 필터 커널(FI)을 이용하여 원본보다 크기가 작으며 원본의 특성을 반영하는 컨볼루션 이미지를 추출할 수 있다. 컨볼루션 이미지를 추출하는 노드들의 집합을 컨볼루션 레이어(Convolutional Layer)라고 한다.

[0094] 도 7은 프로세서(110)가 심층 컨볼루션 신경망을 이용해서 패치 이미지가 텍스트 이미지로 분류될 확률을 계산하는 것을 나타낸 개념도이다.

[0095] 가장 하위 레이어는 패치 이미지를 입력 받을 수 있다. 패치 이미지는 20×48 개의 픽셀을 포함할 수 있다. 각각의 픽셀 별로 색상 이미지 정보가 포함되어 있으므로, 패치 이미지에 대한 정보는 $20 \times 48 \times 3$ 개의 최하위 노드들에 입력될 수 있다. 제1 컨볼루션 레이어(CL1)는 5×7 크기의 필터 커널을 이용하여 컨볼루션 이미지를 추출할 수 있다. 제1 컨볼루션 레이어(CL1)는 32개의 컨볼루션 이미지를 추출할 수 있다. 각각의 컨볼루션 이미지는 서로 다른 커널 필터에 의해 추출될 수 있으며, 서로 다른 이미지의 특성을 반영할 수 있다. 각각의 컨볼루션 이미지는 $(20-5+1) \times (48-7+1) = 16 \times 42$ 개의 픽셀을 포함할 수 있다. 총 32개의 컨볼루션 이미지가 추출되므로, 제1 컨볼루션 레이어(CL1)에서 총 $16 \times 42 \times 32$ 개의 노드가 형성될 수 있다.

[0096] 제1 풀링 레이어(P01)는 풀링 방식을 이용하여 제1 컨볼루션 레이어(CL1)보다 작은 개수의 노드들을 형성할 수 있다. 예를 들어, 제1 풀링 레이어(P01)는 컨볼루션 이미지에서 인접한 픽셀들을 머지(merge)하여 하나의 대표 값을 획득함으로써, 노드들의 개수를 줄일 수 있다. 예시적으로, 제1 풀링 레이어(P01)는 2×2 개의 노드들에서 하나의 대표 값을 획득할 수 있다. 이와 같이 복수의 노드에서 하나의 대표 값을 취해 노드 개수를 줄이는 작업을 풀링(Pooling)이라고 한다. 대표 값은 여러 가지 방식으로 획득될 수 있다. 예를 들어, 2×2 개의 노드들에서 하나의 노드 값이 랜덤으로 선택되어 대표 값으로 획득될 수 있다. 다른 예로, 2×2 개의 노드들에서 컨볼루션 값이 최고인 노드가 선택되어 대표 값으로 획득될 수 있다. 제1 풀링 레이어(P01)의 풀링에 의해서 제1 풀링 레이어(P01)에서 계산된 노드들의 숫자는 $(16 \div 2) \times (42 \div 2) \times 32 = 8 \times 21 \times 32$ 개가 될 수 있다.

[0097] 제2 컨볼루션 레이어(CL2)는 3×7 크기의 필터 커널을 이용하여 컨볼루션 이미지를 추출할 수 있다. 제2 컨볼루션 레이어에 의해 64개의 컨볼루션 이미지가 추출될 수 있다. 제2 컨볼루션 레이어(CL2)에 의해 $(8-3+1) \times (21-7+1) \times 32 = 6 \times 15 \times 32$ 개의 노드들이 형성될 수 있다. 제2 풀링 레이어(P02)는 제2 컨볼루션 레이어(CL2)에서 형성된 노드들에 대해 풀링을 할 수 있다. 제2 풀링 레이어(P02)는 3×3 개의 노드들에서 하나의 노드 값을 대표 값으로 획득할 수 있다. 제2 풀링 레이어(P02)는 풀링에 의해 $(6 \div 3) \times (15 \div 3) \times 64 = 2 \times 5 \times 64$ 개의 노드들을 형성할 수 있다.

[0098] 제3 컨볼루션 레이어(CL3)는 2×5 크기의 필터 커널을 이용하여 컨볼루션 이미지를 추출할 수 있다. 제3 컨볼루션 레이어에 의해 64개의 노드들이 형성될 수 있다.

[0099] 완전 연결 레이어(Fully connected layer; FC)는 제1 내지 제3 컨볼루션 레이어(CL1, CL2, CL3)와, 제1 및 제2 풀링 레이어(P01, P02)와 연결될 수 있다. 완전 연결 레이어(FC)는 이전 모든 하위 레이어들에 대해 서로 다른 가중치를 부여할 수 있다. 완전 연결 레이어(FC)가 하위 레이어들에 가중치를 부여하는 방식은 후술하는 지도 학습에 의해 학습될 수 있다. 프로세서(110)는 지도 학습에 의해 완전 연결 레이어(FC)가 하위 레이어들에 가중치를 부여하는 방식을 변경함으로써, 심층 컨볼루션 신경망을 개선할 수 있다.

[0100] 완전 연결 레이어(FC)는 하위 레이어들에 서로 다른 가중치를 부여하고, 가중치가 반영된 값들을 조합하여 최종적으로 2개의 파라미터들을 계산할 수 있다. 그리고, 소프트 맥스 레이어(SM)는 2개의 파라미터들로부터 패치 이미지가 텍스트 이미지로 분류될 확률 및 패치 이미지가 비-텍스트 이미지로 분류될 확률을 각각 계산하여 출력할 수 있다. 상술한 설명들 중 수치들에 관한 설명은 이해를 돕기 위해 예시적으로 제시한 것에 불과할 뿐,

실시예가 이에 제한되는 것은 아니다.

- [0101] 도 8은 도 7에서 나타난 심층 컨볼루션 신경망의 변형예이다. 도 8의 내용을 설명함에 있어서, 도 7과 중복되는 내용의 설명은 생략한다.
- [0102] 도 8을 참조하면, 변형예에 따른 심층 컨볼루션 신경망은 두 개의 완전 연결 레이어(FC1, FC2)를 포함할 수 있다. 레이어가 추가됨에 따라, 계산 시간과 계산에 요구되는 자원은 다소 늘어날 수 있다. 하지만, 하위 레이어들에 대한 가중치를 부여하여 결과 파라미터들을 계산하는 과정이 제1 완전 연결 레이어(FC1) 및 제2 완전 연결 레이어(FC2)를 거쳐 이루어지도록 함으로써, 소프트 맥스 레이어(SM)에서 출력되는 결과의 신뢰도를 높일 수 있다.
- [0103] 도 9는 도 7 및 도 8에서 나타난 심층 컨볼루션 신경망의 변형예이다. 도 9의 내용을 설명함에 있어서, 도 7 및 도 8과 중복되는 내용의 설명은 생략한다.
- [0104] 도 9를 참조하면, 풀링 레이어(P01)는 하나로 줄이고, 컨볼루션 레이어들(CL1, CL2, CL3, CL4, CL5) 5개로 늘릴 수 있다. 그리고, 완전 연결 레이어들(FC1, FC2, FC3, FC4)의 개수를 4개로 늘릴 수 있다. 계산량이 상대적으로 작은 풀링 레이어(P01)를 줄이고, 컨볼루션 레이어들(CL1, CL2, CL3, CL4, CL5)과 완전 연결 레이어들(FC1, FC2, FC3, FC4)의 개수를 늘렸기 때문에, 계산 시간과 계산에 요구되는 자원이 다소 늘어날 수 있다. 하지만, 컨볼루션 과정과 가중치 부여 과정에 정확도가 높아지면서 결과적으로, 소프트 맥스 레이어(SM)에서 출력되는 결과의 신뢰도를 높일 수 있다.
- [0105] 이상에서 도 7 내지 도 9를 참조하여, 프로세서(110)가 심층 컨볼루션 신경망을 이용하여, 패치 이미지가 텍스트 이미지로 분류될 확률을 계산하는 것을 설명하였다. 상술한 심층 컨볼루션 신경망의 예들은 예시적인 것에 불과하며, 심층 컨볼루션 신경망의 구체적인 구조는 통상의 기술자에게 자명한 범위 내에서 변경될 수 있다.
- [0106] 다시 도 2를 참조하면, S140 단계에서, 프로세서(110)는 대상 이미지의 픽셀 별로, 픽셀 각각이 텍스트 영역으로 분류될 확률을 계산할 수 있다. 프로세서(110)는 미리 정의된 픽셀을 포함하는 패치 이미지들이 텍스트 이미지로 분류될 확률 값들에 기초하여, 미리 정의된 픽셀이 텍스트 영역으로 분류될 확률을 계산할 수 있다.
- [0107] 도 10은 프로세서(110)가 픽셀 $x(2,3)$ 이 텍스트 영역으로 분류될 확률을 계산하는 것을 개념적으로 나타낸 개념도이다.
- [0108] 도 10을 참조하면, 패치 이미지들(PI1, PI2, PI3, PI4) 각각은 6×4 개의 픽셀을 포함할 수 있다. 그리고, 인접한 패치 이미지들은 1×2 크기의 보폭 영역만큼 이격 되어 있을 수 있다. 인접한 패치 이미지들이 서로 적어도 하나의 픽셀을 공유할 수 있다. 따라서, 픽셀 $x(2,3)$ 은 복수의 패치 이미지들(PI1, PI2, PI3, PI4)에 포함될 수 있다.
- [0109] 제1 내지 제4 패치 이미지(PI1, PI2, PI3, PI4)가 픽셀 $P(2,3)$ 을 공유할 수 있다. 프로세서(110)는 제1 패치 이미지가 텍스트 이미지로 분류될 확률 $c(P1)$ 을 계산할 수 있다. 프로세서(110)는 제2 패치 이미지가 텍스트 이미지로 분류될 확률 $c(P2)$ 를 계산할 수 있다. 프로세서(110)는 제3 패치 이미지가 텍스트 이미지로 분류될 확률 $c(P3)$ 를 계산할 수 있다. 프로세서(110)는 제4 패치 이미지가 텍스트 이미지로 분류될 확률 $c(P4)$ 를 계산할 수 있다. 프로세서는 확률 $c(P1)$, $c(P2)$, $c(P3)$, $c(P4)$ 의 제곱 합을 계산할 수 있다. 그리고, 확률들의 제곱 합을 정규화(normalize) 함으로써 픽셀 $x(2,3)$ 이 텍스트 영역으로 분류될 확률을 계산할 수 있다.
- [0110] 프로세서(110)는 임의의 픽셀 x 를 포함하는 패치 이미지들 각각이 텍스트 이미지로 분류될 확률들의 제곱 합 $ac(x)$ 를 수학적 식 1에 따라 계산할 수 있다.

수학적 식 1

$$ac(x) = \sum_{P \in S(x)} c^2(P)$$

[0111]

[0112] 수학식 1에서, $S(x)$ 는 픽셀 x 를 포함하는 패치 이미지들의 집합을 의미한다. 또한, $c^2(P)$ 는 패치 이미지 P 가 텍스트 이미지로 분류될 확률의 제곱을 의미한다.

[0113] 프로세서(110)는 확률들의 제곱 합을 정규화하여 픽셀 x 가 텍스트 영역으로 분류될 확률 $ac_n(x)$ 를 계산할 수 있다. $ac_n(x)$ 는 수학식 2에 따라 계산될 수 있다.

수학식 2

$$ac_n(x) = \frac{1}{N(x)} \sum_{P \in S(x)} c^2(P)$$

[0114]

[0115] 수학식 2에서, $S(x)$ 는 픽셀 x 를 포함하는 패치 이미지들의 집합을 의미한다. 또한, $c^2(P)$ 는 패치 이미지 P 가 텍스트 이미지로 분류될 확률의 제곱을 의미한다. $N(x)$ 는 픽셀 x 를 포함하는 패치 이미지들의 개수로, 집합 $S(x)$ 의 원소 개수에 해당한다.

[0116] 상술한 바와 같이, 프로세서(110)는 픽셀이 텍스트 영역에 해당하는지를 판단함에 있어서, 패치 이미지 단위로 텍스트 이미지 분류 확률을 계산하고 이들 확률을 누적하여 픽셀 단위로 텍스트 영역으로 분류될 확률을 계산할 수 있다. 따라서, 하나의 픽셀이 텍스트 영역에 포함되는지 여부를 판단할 때, 인접한 픽셀의 특성도 함께 반영될 수 있다. 인접한 픽셀 특성을 반영함으로써, 텍스트에 인접하였으나 명도나 휘도가 낮은 픽셀도 함께 텍스트 영역으로 분류될 수 있다. 또한, 텍스트에 인접하지 않으면서 명도나 휘도가 높은 픽셀은 비-텍스트 영역으로 분류될 수 있다.

[0117] 도 11은 제철소에서 촬영된 대상 이미지를 나타낸다.

[0118] 도 11을 참조하면, 영역 R1은 텍스트를 포함하지 않음에도 일부 픽셀들이 명도가 밝고, 엷지로 검출되어 종래 방식에 의할 경우, 일부 픽셀들이 텍스트 영역으로 분류될 수 있다. 또한, 영역 R2는 텍스트 영역에 해당함에도 불구하고, 픽셀의 명도나 휘도가 떨어져 비-텍스트 영역으로 분류될 수 있다. 이에 의해 일부 텍스트 영역이 누락되어 정확한 텍스트 인식에 장애가 발생할 수 있다.

[0119] 종래의 방식과 달리 본 발명의 실시예에 따르면, 패치 이미지 기반의 분석 방식에 의해 인접한 픽셀의 특성들이 함께 고려되므로 위와 같은 문제점을 개선할 수 있다. 또한, 종래 대상 이미지 전체를 input data로 하는 방식과 달리, 대상 이미지의 부분 이미지인 패치 이미지를 input data로 하기 때문에, 심층 컨볼루션 신경망의 구동에 요구되는 시간, 전력, 하드웨어 자원을 절감할 수 있다.

[0120] 도 12는 추가적인 실시예에 따른 텍스트 영역 식별 방법을 나타낸 흐름도이다. 도 12의 실시예를 설명함에 있어서, 도 2와 중복되는 내용은 생략한다.

[0121] 도 12를 참조하면, S150 단계에서, 프로세서(110)는 확률 분포맵을 생성할 수 있다. 프로세서(110)는 대상 이미지의 픽셀들 각각이 상기 텍스트 영역으로 분류될 확률 값들로부터, 확률 분포맵을 생성할 수 있다.

[0122] 도 13은 대상 이미지와 대상 이미지로부터 생성된 확률 분포맵을 예시적으로 나타낸 도면이다.

[0123] 도 13을 참조하면, 프로세서(110)는 대상 이미지(TI)에 대해 S120, S130, S140 단계를 거쳐 픽셀 별로 각 픽셀이 텍스트 영역으로 분류될 확률을 계산할 수 있다. 그리고, 각각의 픽셀이 텍스트 영역으로 분류될 확률 값들로부터 확률 분포맵(PM)을 생성할 수 있다. 확률 분포맵(PM)에서 텍스트 영역으로 분류될 확률이 높은 픽셀은 밝게 표시되고, 텍스트 영역으로 분류될 확률이 낮은 픽셀은 어둡게 표시될 수 있다. 상술한 바와 같이, 인접한 픽셀들의 특성이 고려되므로, 텍스트가 존재하는 근처 영역의 픽셀들 만이 밝게 표시될 수 있다.

[0124] 다시 도 12를 참조하면, S160 단계에서, 프로세서(110)는 확률 분포맵의 행(row)별로 픽셀들이 텍스트 영역으로 분류될 확률 값들을 합산하여 수평 프로젝션 프로파일(horizontal projection profile)을 계산할 수 있다. S170 단계에서, 프로세서(110)는 확률 분포맵의 열(column)별로 픽셀들이 텍스트 영역으로 분류될 확률 값들을 합산하여 수직 프로젝션 프로파일(vertical projection profile)을 계산할 수 있다. 그리고, S180 단계에서,

프로세서(110)는 수평 프로젝션 프로파일 및 수직 프로젝션 프로파일로부터 텍스트 영역을 결정할 수 있다.

- [0125] 도 14는 도 12에서 나타난 확률 분포맵(PM)으로부터 계산된 수평 프로젝션 프로파일(HP)과 수직 프로젝션 프로파일(VP1, VP2)을 나타낸 개념도이다.
- [0126] 도 14를 참조하면, 프로세서(110)는 행별로 누적된 확률 값들의 합산 값으로부터 수평 프로젝션 프로파일(HP)을 계산할 수 있다. 프로세서(110)는 수평 프로젝션 프로파일(HP)에서 확률 누적 값이 임계 값(붉은 선으로 표시) 이상인 두 개의 영역을 추출할 수 있다. 프로세서(110)는 수평 프로젝션 프로파일(HP)로부터 추출된 두 개의 영역 각각에 대해 수직 프로젝션 프로파일들(VP1, VP2)를 계산할 수 있다. 프로세서(110)는 수직 프로젝션 프로파일들(VP1, VP2)에서 확률 누적 값이 임계 값(붉은 선으로 표시) 이상이 영역들을 추출할 수 있다. 프로세서(110)는 수평 프로젝션 프로파일(HP)로부터 추출된 영역들과 수직 프로젝션 프로파일(VP)로부터 추출된 영역들이 교집합을 형성하는 영역을 텍스트 영역으로 결정할 수 있다. 텍스트 영역은 직사각형 모양일 수 있다. 프로세서(110)는 수평 프로젝션 프로파일(HP)과 수직 프로젝션 프로파일(VP)을 이용함으로써, 텍스트 영역 결정의 신뢰도를 높일 수 있다.
- [0127] 도 15는 텍스트 영역 식별 방법에 대한 제2 실시예를 나타낸 흐름도이다. 도 15의 실시예를 설명함에 있어서, 도 2와 중복되는 내용은 생략한다.
- [0128] 도 15를 참조하면, S110 단계에서, 프로세서(110)는 심층 컨볼루션 신경망을 학습할 수 있다. 프로세서(110)는 기계학습에 의해 심층 컨볼루션 신경망의 구성을 업데이트할 수 있다. 기계학습은 지도 학습 방식으로 이루어질 수 있다. 지도 학습은 미리 마련된 예제와 그에 대한 분류 정보를 학습 데이터로 제공하고, 학습 데이터에 근거해서 신경망의 구성을 발전시키는 학습 방법을 의미한다.
- [0129] 도 16은 도 15의 S110 단계를 보다 상세하게 나타낸 흐름도이다.
- [0130] 도 16을 참조하면, S112 단계에서, 프로세서(110)는 테스트 이미지를 입력 받아 메모리(120)에 저장할 수 있다. 프로세서(110)는 테스트 이미지들에 포함된 패치 이미지들에 대한 분류 정보를 입력 받아 메모리(120)에 저장할 수 있다. 패치 이미지들에 대한 분류 정보는 패치 이미지들 각각이 텍스트 이미지인지, 비-텍스트 이미지인지 여부에 대한 분류 정보를 포함할 수 있다.
- [0131] 도 17은 테스트 이미지의 일 예이다. 그리고, 도 18은 도 17의 테스트 이미지에 포함된 패치 이미지들을 확대하여 나타낸 도면이다.
- [0132] 도 17 및 도 18을 참조하면, 패치 이미지 (a)는 텍스트가 기재된 부분을 포함하고 있으므로, 사용자는 패치 이미지 (a)를 텍스트 이미지로 분류하여 텍스트 영역 식별 장치(100)에 입력할 수 있다. 프로세서(110)는 패치 이미지(a)의 분류 정보를 메모리(120)에 저장할 수 있다. 패치 이미지(b)는 텍스트 이미지로 미리 분류되고, 패치 이미지 (c), (d)는 비-텍스트 이미지로 분류될 수 있다. 메모리(120)에 패치 이미지들 각각에 대한 분류 정보가 저장될 수 있다.
- [0133] S114 단계에서, 프로세서(110)는 테스트 이미지로부터 패치 이미지를 추출할 수 있다.
- [0134] S116 단계에서, 프로세서(110)는 추출한 패치 이미지들과, 메모리에 저장된 패치 이미지들 각각에 대한 분류 정보를 이용하여, 심층 컨볼루션 신경망을 학습에 의해 발전시킬 수 있다.
- [0135] 예를 들어, 프로세서(110)는 테스트 이미지로부터 패치 이미지 (a)를 추출하고, 패치 이미지 (a)를 심층 컨볼루션 신경망에 입력시킬 수 있다. 그리고, 프로세서(110)는 심층 컨볼루션 신경망의 출력 결과가 메모리(120)에 저장된 패치 이미지(a)의 분류 정보에 대응되도록 심층 컨볼루션 신경망의 작동 방식을 수정할 수 있다. 즉, 프로세서(110)는 패치 이미지 (a)에 대한 분류 정보를 지도 학습 데이터로 활용할 수 있다. 마찬가지로, 프로세서(110)는 다른 패치 이미지들 (b), (c), (d)에 대한 분류 정보들을 지도 학습 데이터로 활용할 수 있다.
- [0136] 프로세서(110)는 지도 학습 데이터로 참조하여, 심층 컨볼루션 신경망의 작동 방식을 수정할 수 있다. 예시적으로, 프로세서(110)는 완전 연결 레이어가 레이어들의 가중치를 주는 방식을 변경해가면서, 어느 방식에서 심층 컨볼루션 신경망의 출력 결과가 지도 학습 데이터와 매칭 비율이 높은 지를 계산할 수 있다. 그리고, 프로세서(110)는 매칭 비율을 가장 높게 하는 가중치 세트를 채용하여 완전 연결 레이어의 연산이 이루어지도록 할 수 있다. 지도 학습에 이용되는 테스트 이미지의 개수가 많아질수록 심층 컨볼루션 신경망의 완성도가 높아질 수 있다.
- [0137] 이상에서 도 1 내지 도 18을 참조하여, 예시적인 실시예들에 따른 텍스트 영역 식별 방법 및 장치에 관하여 설

명하였다. 상술한 실시예들에 따르면, 미리 정의된 픽셀이 텍스트 영역으로 분류될 확률을 계산할 때, 인접한 픽셀들의 특성이 함께 고려될 수 있다. 따라서, 텍스트 영역 식별 결과의 신뢰도가 높아질 수 있다.

[0138] 또한, 대상 이미지 전체가 아닌, 대상 이미지의 부분 이미지에 해당하는 패치 이미지를 이용하여 심층 컨볼루션 신경망을 학습 시키기 때문에, 적은 양의 데이터를 이용하여 심층 컨볼루션 신경망 학습이 가능하다.

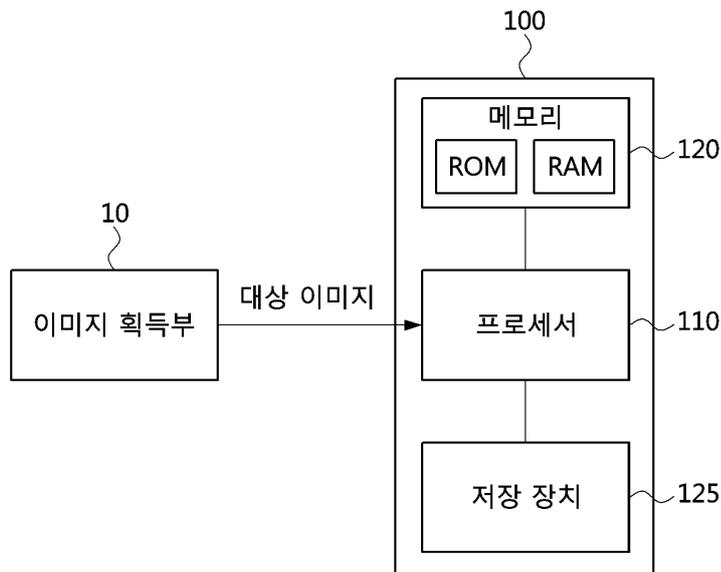
[0139] 이상 실시예를 참조하여 설명하였지만, 해당 기술 분야의 숙련된 당업자는 하기의 특허 청구의 범위에 기재된 본 발명의 사상 및 영역으로부터 벗어나지 않는 범위 내에서 본 발명을 다양하게 수정 및 변경시킬 수 있음을 이해할 수 있을 것이다.

부호의 설명

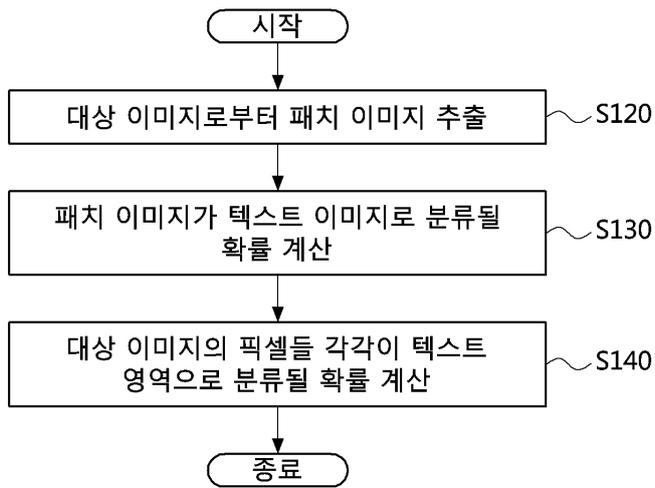
- [0140] 100 : 텍스트 영역 식별 장치
- 110 : 프로세서
- 120 : 메모리
- 125 : 저장 장치
- TI : 대상 이미지
- PI : 패치 이미지

도면

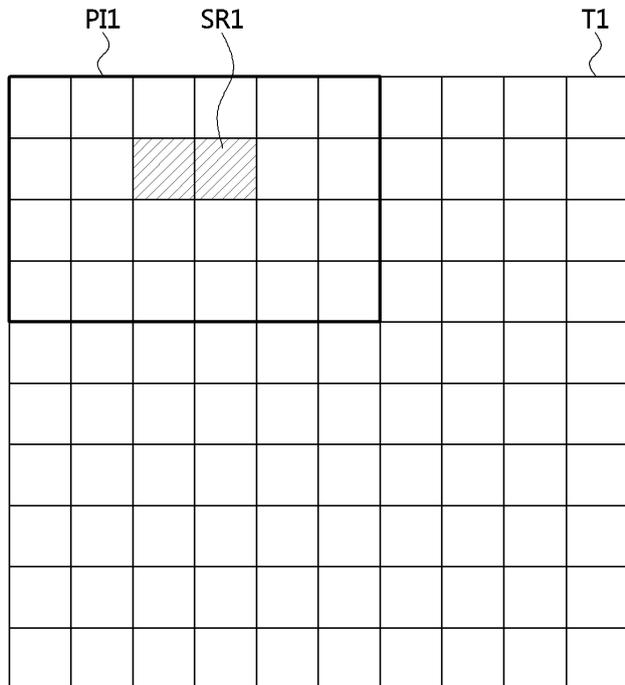
도면1



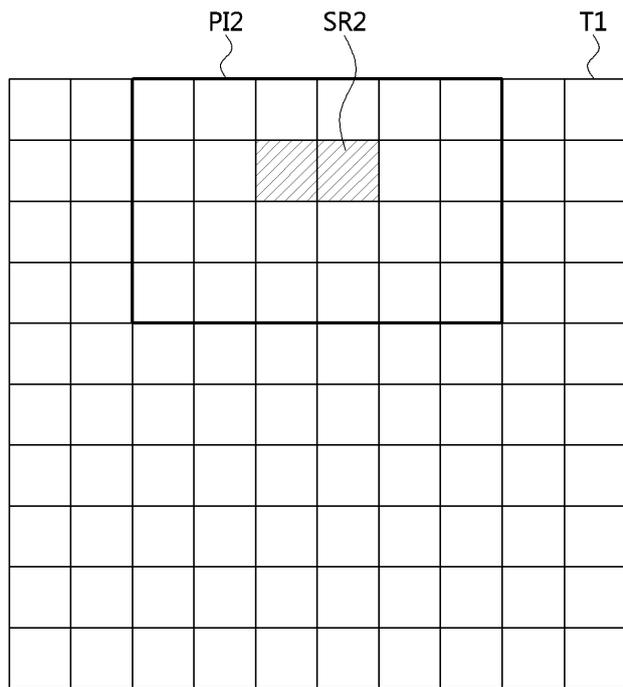
도면2



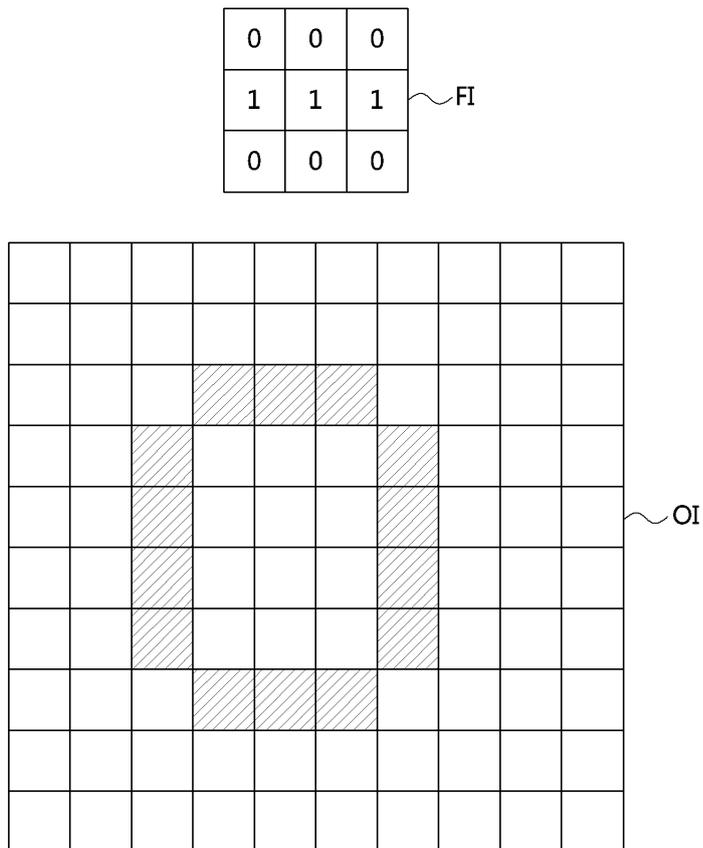
도면3



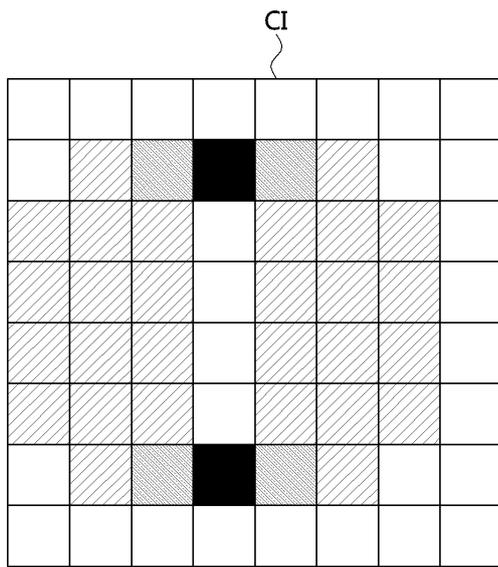
도면4



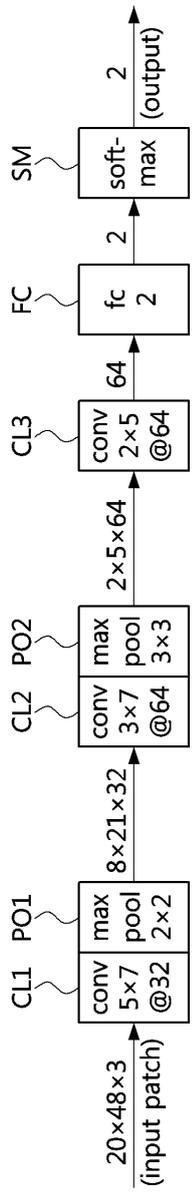
도면5



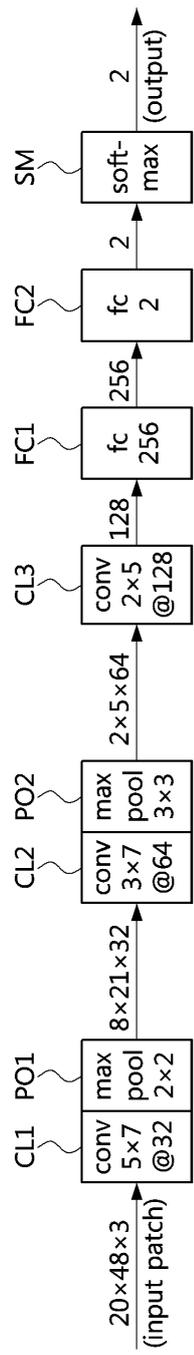
도면6



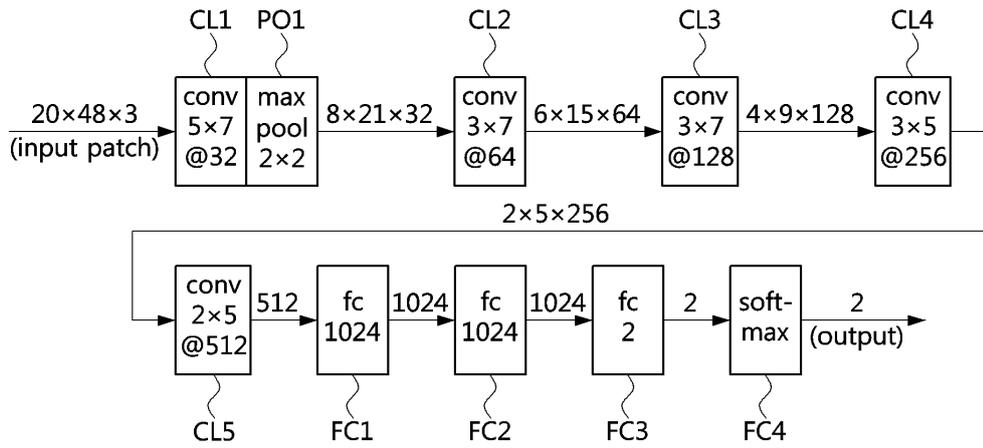
도면7



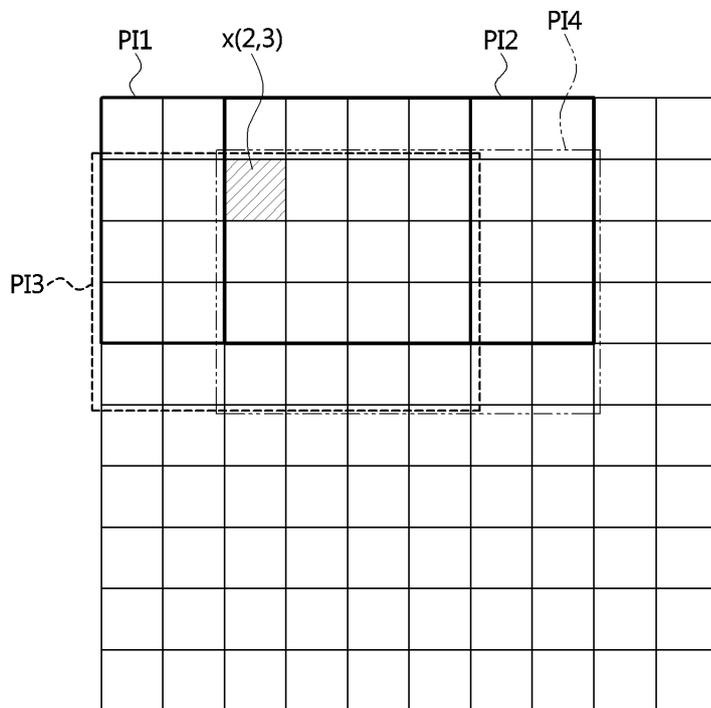
도면8



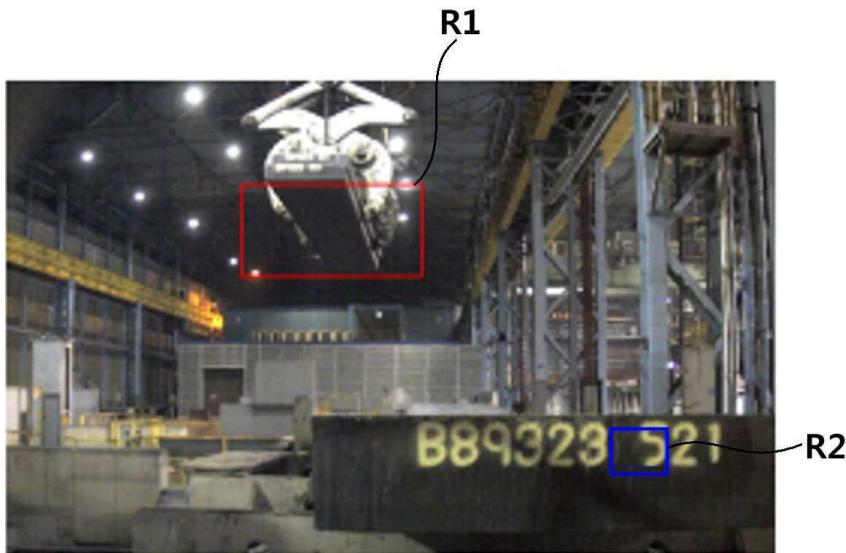
도면9



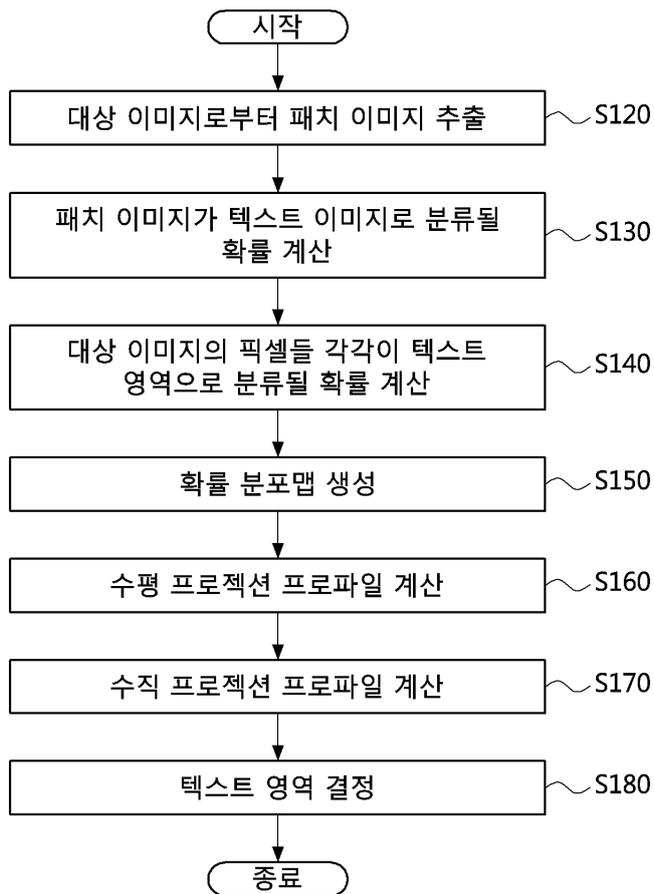
도면10



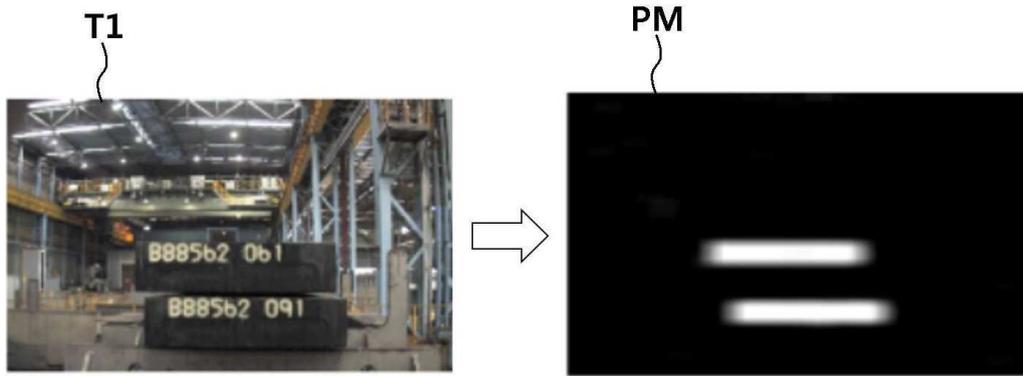
도면11



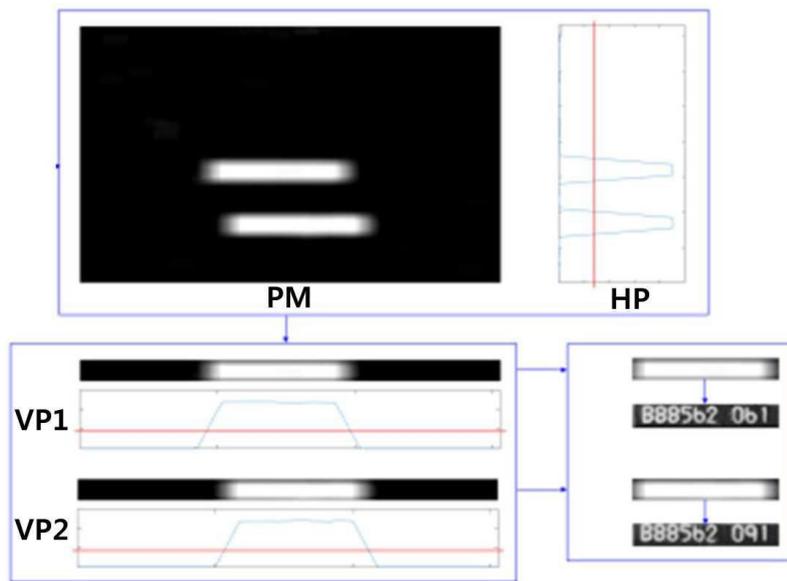
도면12



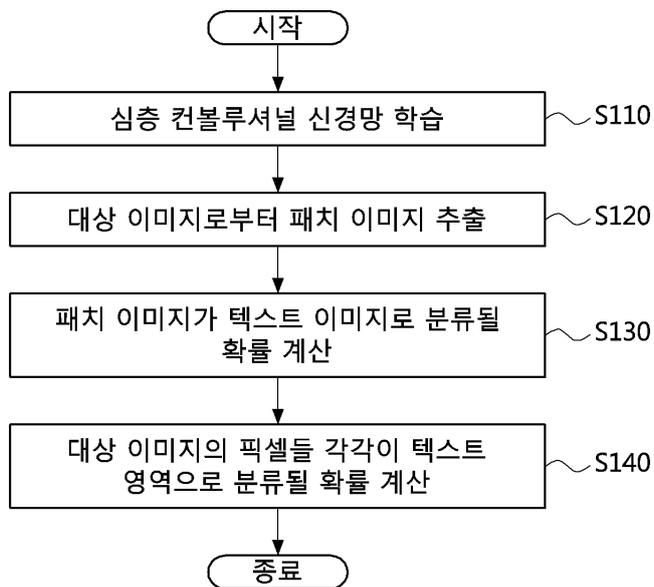
도면13



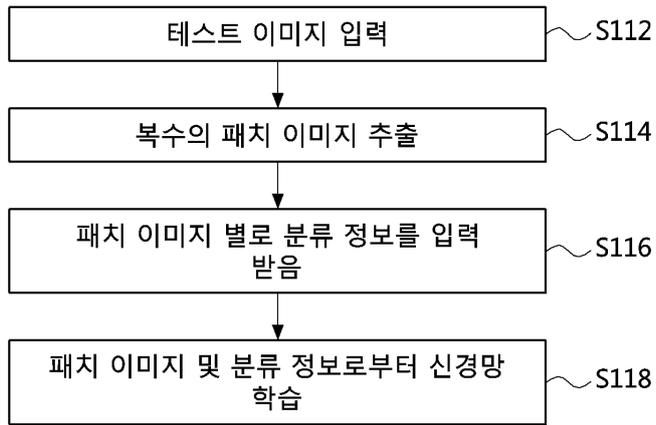
도면14



도면15



도면16



도면17



도면18

