



(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2020년12월16일  
(11) 등록번호 10-2192015  
(24) 등록일자 2020년12월10일

(51) 국제특허분류(Int. Cl.)  
G06T 11/00 (2006.01) G06F 40/20 (2020.01)  
(52) CPC특허분류  
G06T 11/001 (2013.01)  
G06F 40/205 (2020.01)  
(21) 출원번호 10-2019-0003634  
(22) 출원일자 2019년01월11일  
심사청구일자 2019년01월11일  
(65) 공개번호 10-2020-0092491  
(43) 공개일자 2020년08월04일  
(56) 선행기술조사문헌  
KR1020090040014 A\*  
JP60008981 A\*  
KR1020180049642 A\*  
김민정, et al. "텍스트 매핑을 이용한 스케치 기  
반의 얼굴 이미지 생성." 한국정보과학회 학술발  
표논문집 (2017)\*  
\*는 심사관에 의하여 인용된 문헌

(73) 특허권자  
연세대학교 산학협력단  
서울특별시 서대문구 연세로 50 (신촌동, 연세대  
학교)  
(72) 발명자  
김선주  
서울특별시 서대문구 연세로 50, 제4공학관 723호  
(신촌동)  
남성현  
서울특별시 서대문구 연세로 50, 제4공학관 707호  
(신촌동)  
김윤지  
서울특별시 서대문구 연세로 50, 제4공학관 707호  
(신촌동)  
(74) 대리인  
특허법인우인

전체 청구항 수 : 총 13 항

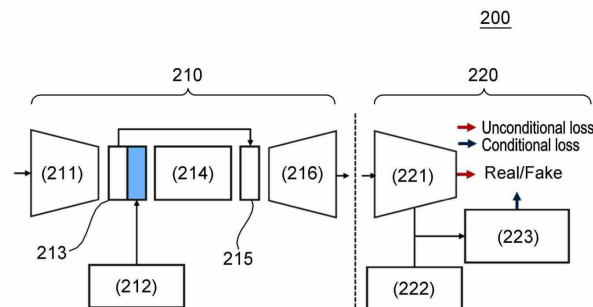
심사관 : 이병우

(54) 발명의 명칭 자연어 문장을 기반으로 하는 변환 이미지 생성 장치, 방법 및 이를 이용하는 변환 이미지 생  
성 시스템

(57) 요약

본 발명의 실시예에 따른 자연어 문장을 기반으로 하는 변환 이미지 생성 장치는, 변환하고자 하는 적어도 하나  
의 객체를 포함하는 입력 이미지를 입력 받아, 상기 입력 이미지를 인코딩하는 이미지 인코더, 상기 입력 이미지  
와 관련된 자연어 문장을 입력 받아, 상기 자연어 문장을 인코딩하는 텍스트 인코더, 상기 입력된 자연어 문장에  
따라 상기 입력 이미지의 적어도 하나의 객체를 변환하는 이미지-자연어 변환부 및 상기 변환된 객체를  
포함하되, 상기 입력 이미지에서 상기 변환된 객체의 영역을 제외한 나머지 영역은 보존된 변환 이미지를 생성하  
는 변환 이미지 생성부를 포함할 수 있다.

대표도



(52) CPC특허분류

**G06T 11/60** (2013.01)

G06T 2207/20081 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711076589
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기술진흥센터
연구사업명	인공지능산업원천기술개발(R&D)
연구과제명	가짜영상 판별성능 향상을 위한 동영상 조작 및 문장 기반 이미지 편집 알고리즘 개
발	
기 여 율	1/1
과제수행기관명	연세대학교 산학협력단
연구기간	2018.08.01 ~ 2019.01.31
공지예외적용	: 있음

---

## 명세서

### 청구범위

#### 청구항 1

변환하고자 하는 적어도 하나의 객체를 포함하는 입력 이미지를 입력 받아, 상기 입력 이미지를 인코딩하는 이미지 인코더;

상기 입력 이미지와 관련된 자연어 문장을 입력 받아, 상기 자연어 문장을 인코딩하는 텍스트 인코더;

상기 입력된 자연어 문장에 따라 상기 입력 이미지의 적어도 하나의 객체를 변환하는 이미지-자연어 변환부; 및  
상기 변환된 객체를 포함하되, 상기 입력 이미지에서 상기 변환된 객체의 영역을 제외한 나머지 영역은 보존된 변환 이미지를 생성하는 변환 이미지 생성부;를 포함하며

상기 텍스트 인코더는, 순환 신경망(RNN) 학습을 통해 상기 자연어 문장에 대한 자연어 특징 값을 포함하는 자연어 특징 블록을 생성하며,

상기 이미지 인코더는, 컨벌루션 레이어를 이용하여 상기 입력 이미지에 대한 이미지 특징 맵을 포함하는 이미지 특징 블록을 생성하고,

상기 이미지-자연어 변환부는, 상기 이미지 특징 맵의 스케일(scale)을 고려하여 상기 자연어 특징 블록을 확장시킨 확장된 자연어 특징 블록을 생성하고, 상기 이미지 특징 블록과 상기 확장된 자연어 특징 블록을 결합하여 이미지-자연어 특징 블록을 생성하는 특징 값 결합부를 포함하는 것을 특징으로 하는 자연어 문장을 기반으로 하는 변환 이미지 생성 장치.

#### 청구항 2

제1항에 있어서,

상기 이미지 인코더는,

컨벌루션 레이어(Convolution layer)를 이용하여 상기 입력 이미지에 대한 이미지 특징 맵을 포함하는 이미지 특징 블록을 생성하는 것을 특징으로 하는 자연어 문장을 기반으로 하는 변환 이미지 생성 장치.

#### 청구항 3

삭제

#### 청구항 4

제1항에 있어서, 상기 텍스트 인코더는,

상기 자연어 문장을 의미론적 분석을 통해 단어 요소들로 분절하는 단어 요소 분절부; 및

상기 분절된 단어 요소 각각에 대한 단어 특징 값들을 생성하는 단어 특징 값 생성부;를 더 포함하고,

상기 텍스트 인코더는, 상기 단어 특징 값 생성부로부터 생성된 단어 특징 값들을 포함하는 단어 특징 블록들을 생성하고, 상기 자연어 특징 블록은 상기 단어 특징 블록들을 포함하는 것을 특징으로 하는 자연어 문장을 기반으로 하는 변환 이미지 생성 장치.

#### 청구항 5

삭제

#### 청구항 6

제1항에 있어서,

상기 이미지-자연어 변환부는, 컨벌루션 연산을 통한 컨벌루션 특징을 추출하는 복수의 컨벌루션 레이어들을 이

용하는 레지듀얼 블록부를 더 포함하고,

상기 레지듀얼 블록부는 상기 이미지-자연어 특징 블록을 입력 받아 레지듀얼 변환을 위한 컨벌루션 레이어를 적용하여 레지듀얼 변환 특징 블록을 생성하는 것을 특징으로 하는 자연어 문장을 기반으로 하는 변환 이미지 생성 장치.

## 청구항 7

제6항에 있어서,

상기 이미지-자연어 변환부, 상기 이미지 특징 블록과 상기 레지듀얼 변환 특징 블록을 합산하여 합산된 합산 특징 블록을 생성하는 합산부;를 더 포함하고,

상기 변환 이미지 생성부는, 상기 합산부에 의해 생성된 합산 특징 블록을 디코딩하여 상기 변환 이미지를 생성하는 것을 특징으로 하는 자연어 문장을 기반으로 하는 변환 이미지 생성 장치.

## 청구항 8

이미지 인코더가 변환하고자 하는 적어도 하나의 객체를 포함하는 입력 이미지를 입력 받아, 컨벌루션 레이어(Convolution layer)를 이용하여 상기 입력 이미지에 대한 이미지 특징 맵을 포함하는 이미지 특징 블록을 생성함에 따라 상기 입력 이미지를 인코딩하는 단계;

텍스트 인코더가 상기 입력 이미지와 관련된 자연어 문장을 입력 받아, 순환 신경망(RNN) 학습을 통해 상기 자연어 문장에 대한 자연어 특징 값을 포함하는 자연어 특징 블록을 생성함에 따라 상기 자연어 문장을 인코딩하는 단계;

이미지-자연어 변환부가 상기 입력된 자연어 문장에 따라 상기 입력 이미지의 적어도 하나의 객체를 변환하는 단계; 및

변환 이미지 생성부가 상기 변환된 객체를 포함하되, 상기 입력 이미지에서 상기 변환된 객체의 영역을 제외한 나머지 영역은 보존된 변환 이미지를 생성하는 단계;를 포함하며,

상기 입력 이미지의 적어도 하나의 객체를 변환하는 단계는, 상기 이미지 특징 맵의 스케일(scale)을 고려하여 상기 자연어 특징 블록을 확장시킨 확장된 자연어 특징 블록을 생성하고, 상기 이미지 특징 블록과 상기 확장된 자연어 특징 블록을 결합하여 이미지-자연어 특징 블록을 생성하는 단계를 포함하는 것을 특징으로 하는 자연어 문장을 기반으로 하는 변환 이미지 생성 방법.

## 청구항 9

제8항에 있어서,

상기 자연어 문장을 인코딩하는 단계는,

상기 자연어 문장을 의미론적 분석을 통해 단어 요소들로 분절하는 단계; 및

상기 분절된 단어 요소 각각에 대한 단어 특징 값들을 생성하는 단계;를 더 포함하고,

상기 자연어 문장을 인코딩하는 단계는, 상기 생성된 단어 특징 값들을 포함하는 자연어 특징 블록을 생성하고, 상기 자연어 특징 블록은 상기 단어 특징 값들을 포함하는 것을 특징으로 하는 자연어 문장을 기반으로 하는 변환 이미지 생성 방법.

## 청구항 10

제9항에 있어서,

상기 입력 이미지의 적어도 하나의 객체를 변환하는 단계는,

상기 이미지-자연어 특징 블록을 입력 받아 레지듀얼 변환을 위한 컨벌루션 레이어를 적용하여 레지듀얼 변환 특징 블록을 생성하는 단계;

를 더 포함하는 것을 특징으로 하는 자연어 문장을 기반으로 하는 변환 이미지 생성 방법.

#### 청구항 11

변환하고자 하는 적어도 하나의 객체를 포함하는 입력 이미지를 입력 받아, 상기 입력 이미지를 인코딩하는 제1 이미지 인코더; 상기 입력 이미지와 관련된 자연어 문장을 입력 받아, 상기 자연어 문장을 인코딩하는 제1 텍스트 인코더; 상기 입력된 자연어 문장에 따라 상기 입력 이미지의 적어도 하나의 객체를 변환하는 이미지-자연어 변환부; 및 상기 변환된 객체를 포함하되, 상기 입력 이미지에서 상기 변환된 객체의 영역을 제외한 나머지 영역은 보존된 변환 이미지를 생성하는 변환 이미지 생성부;를 포함하는 생성기; 및

상기 생성기로부터 생성된 변환 이미지를 입력 받아, 상기 변환 이미지를 인코딩하는 제2 이미지 인코더; 상기 변환 이미지를 검증하기 위한 상기 자연어 문장을 입력 받아, 상기 자연어 문장을 인코딩하는 제2 텍스트 인코더; 상기 인코딩된 변환 이미지 및 인코딩된 자연어 문장을 이용하여 상기 변환 이미지를 검증하기 위한 국소 매칭 스코어를 산출하는 자연어 기반 구별부; 및 상기 국소 매칭 스코어를 고려하여 상기 변환 이미지를 재변환하기 위한 재변환여부를 결정하는 재변환여부 결정부;를 포함하는 검증기;

를 포함하는 변환 이미지 생성 시스템.

#### 청구항 12

제11항에 있어서,

상기 제2 이미지 인코더는, 컨벌루션 레이어(Convolution layer)를 이용하여 상기 변환 이미지에 대한 변환 이미지 특징 맵을 포함하는 변환 이미지 특징 블록을 생성하고,

상기 변환 이미지 특징 맵은, 상기 컨벌루션 레이어에 의해 컨벌루션 변환된 객체별 이미지 특징 맵들을 포함하는 것을 특징으로 하는 변환 이미지 생성 시스템.

#### 청구항 13

제12항에 있어서,

상기 제2 텍스트 인코더는,

상기 자연어 문장을 의미론적 분석을 통해 단어 요소들로 분절하고, 순환 신경망(RNN) 학습을 통해 상기 분절된 단어 요소 각각에 대한 단어 특징 값을 생성하는 것을 특징으로 하는 변환 이미지 생성 시스템.

#### 청구항 14

제13항에 있어서,

상기 자연어 기반 구별부는, 상기 제2 텍스트 인코더로부터 생성된 각 단어 특징 값과 상기 각 객체별 이미지 특징 맵을 매칭함에 따라 상기 국소 매칭 스코어를 산출하는 것을 특징으로 하는 변환 이미지 생성 시스템.

#### 청구항 15

컴퓨터에 이미지 인코더가 변환하고자 하는 적어도 하나의 객체를 포함하는 입력 이미지를 입력 받아, 컨벌루션 레이어(Convolution layer)를 이용하여 상기 입력 이미지에 대한 이미지 특징 맵을 포함하는 이미지 특징 블록을 생성함에 따라 상기 입력 이미지를 인코딩하는 단계;

텍스트 인코더가 상기 입력 이미지와 관련된 자연어 문장을 입력 받아, 순환 신경망(RNN) 학습을 통해 상기 자연어 문장에 대한 자연어 특징 값을 포함하는 자연어 특징 블록을 생성함에 따라 상기 자연어 문장을 인코딩하는 단계;

이미지-자연어 변환부가 상기 입력된 자연어 문장에 따라 상기 입력 이미지의 적어도 하나의 객체를 변환하는 단계; 및

변환 이미지 생성부가 상기 변환된 객체를 포함하되, 상기 입력 이미지에서 상기 변환된 객체의 영역을 제외한 나머지 영역은 보존된 변환 이미지를 생성하는 단계;를 실행시키며,

상기 입력 이미지의 적어도 하나의 객체를 변환하는 단계는, 상기 이미지 특징 맵의 스케일(scale)을 고려하여 상기 자연어 특징 블록을 확장시킨 확장된 자연어 특징 블록을 생성하고, 상기 이미지 특징 블록과 상기 확장된 자연어 특징 블록을 결합하여 이미지-자연어 특징 블록을 생성하는 단계를 포함하는 것을 특징으로 하는 컴퓨터

판독 가능한 저장 매체에 저장된 컴퓨터 프로그램.

## 발명의 설명

### 기술 분야

[0001] 본 발명은 자연어 문장을 기반으로 하는 변환 이미지 생성 장치, 방법 및 이를 이용하는 변환 이미지 생성 시스템에 관한 것이다.

### 배경 기술

[0002] 스마트 폰이 일상적인 장치로 부상한 이후로 사진을 찍는 것은 사람들의 삶의 중요한 부분이 되었다. 최근에는, 이러한 추세에 따라 이미지를 조작하거나 편집하기 위한 요구가 많아지면서, 사진을 더 보기 좋게 하거나 사용자의 요구를 충족시킬 수 있도록 하는 시스템을 개발하게 되었다.

[0003] 하지만, 개발되어 왔던 종래의 시스템은 상업적으로 사용 가능한 툴을 사용하여 이미지를 편집할 수는 있지만, 이미지 편집 전문가가 아닌 일반 스마트 폰 사용자는 이미지 편집을 하기 위해 이러한 툴을 사용하는 것은 많은 어려움이 있는 것이 현실이다.

[0004] 또한, 스마트 폰과 같은 모바일 장치를 통해 이미지를 편집하는 것은 많은 제약이 있어 보다 정교한 편집을 구현하는데 한계가 있었다.

### 선행기술문헌

#### 특허문헌

[0005] (특허문헌 0001) 한국 등록 특허 제10-189428호 (등록)

### 발명의 내용

#### 해결하려는 과제

[0006] 상기 기술한 종래의 문제점을 해결하기 위한 본 발명의 실시예에 따른 자연어 문장을 기반으로 하는 변환 이미지 생성 장치, 방법 및 시스템은 사용자의 의도에 맞게 고해상도의 이미지로 변환 및 편집하는 것을 목적으로 한다.

#### 과제의 해결 수단

[0007] 상기 목적을 달성하기 위한 본 발명의 실시예에 따른 자연어 문장을 기반으로 하는 변환 이미지 생성 장치는, 변환하고자 하는 적어도 하나의 객체를 포함하는 입력 이미지를 입력 받아, 상기 입력 이미지를 인코딩하는 이미지 인코더, 상기 입력 이미지와 관련된 자연어 문장을 입력 받아, 상기 자연어 문장을 인코딩하는 텍스트 인코더, 상기 입력된 자연어 문장에 따라 상기 입력 이미지의 적어도 하나의 객체를 변환하는 이미지-자연어 변환부 및 상기 변환된 객체를 포함하되, 상기 입력 이미지에서 상기 변환된 객체의 영역을 제외한 나머지 영역은 보존된 변환 이미지를 생성하는 변환 이미지 생성부를 포함할 수 있다.

[0008] 또한, 상기 이미지 인코더는, 컨벌루션 레이어(Convolution layer)를 이용하여 상기 입력 이미지에 대한 이미지 특징 맵을 포함하는 이미지 특징 블록을 생성할 수 있다.

[0009] 또한, 상기 텍스트 인코더는, 순환 신경망(RNN) 학습을 통해 상기 자연어 문장에 대한 자연어 특징 값을 포함하는 자연어 특징 블록을 생성할 수 있다.

[0010] 또한, 상기 텍스트 인코더는, 상기 자연어 문장을 의미론적 분석을 통해 단어 요소들로 분절하는 단어 요소 분절부 및 상기 분절된 단어 요소 각각에 대한 단어 특징 값들을 생성하는 단어 특징 값 생성부;를 더 포함할 수 있고, 상기 텍스트 인코더는, 상기 단어 특징 값 생성부로부터 생성된 단어 특징 값들을 포함하는 단어 특징 블록들을 생성하고, 상기 자연어 특징 블록은 상기 단어 특징 블록들을 포함할 수 있다.

[0011] 또한, 상기 이미지 인코더는, 컨벌루션 레이어를 이용하여 상기 입력 이미지에 대한 이미지 특징 맵을 포함하는

이미지 특징 블록을 생성하고, 상기 이미지-자연어 변환부는, 상기 이미지 특징 맵의 스케일(scale)을 고려하여 상기 자연어 특징 블록을 확장시킨 자연어 특징 블록을 생성하고, 상기 이미지 특징 블록과 상기 자연어 특징 블록을 결합하여 이미지-자연어 특징 블록을 생성하는 특징 값 결합부를 더 포함할 수 있다.

[0012] 또한, 상기 이미지-자연어 변환부는, 컨벌루션 연산을 통한 컨벌루션 특징을 추출하는 복수의 컨벌루션 레이어들을 이용하는 레지듀얼 블록부를 더 포함하고, 상기 레지듀얼 블록부는 상기 이미지-자연어 특징 블록을 입력받아 레지듀얼 변환을 위한 컨벌루션 레이어를 적용하여 레지듀얼 변환 특징 블록을 생성할 수 있다.

[0013] 또한, 상기 이미지-자연어 변환부, 상기 영상 특징 블록과 상기 레지듀얼 변환 특징 블록을 합산하여 합산된 합산 특징 블록을 생성하는 합산부;를 더 포함하고, 상기 변환 이미지 생성부는, 상기 합산부에 의해 생성된 합산 특징 블록을 디코딩하여 상기 변환 이미지를 생성할 수 있다.

[0014] 상술한 목적을 달성하기 위한 본 발명의 다른 실시예에 따른 자연어 문장을 기반으로 하는 변환 이미지 생성 방법은, 이미지 인코더가 변환하고자 하는 적어도 하나의 객체를 포함하는 입력 이미지를 입력받아, 컨벌루션 레이어(Convolution layer)를 이용하여 상기 입력 이미지에 대한 이미지 특징 맵을 포함하는 이미지 특징 블록을 생성 생성함에 따라 상기 입력 이미지를 인코딩하는 단계, 텍스트 인코더가 상기 입력 이미지와 관련된 자연어 문장을 입력받아, 순환 신경망(RNN) 학습을 통해 상기 자연어 문장에 대한 자연어 특징 값을 포함하는 자연어 특징 블록을 생성함에 따라 상기 자연어 문장을 인코딩하는 단계, 이미지-자연어 변환부가 상기 입력된 자연어 문장에 따라 상기 입력 이미지의 적어도 하나의 객체를 변환하는 단계 및 변환 이미지 생성부가 상기 변환된 객체를 포함하되, 상기 입력 이미지에서 상기 변환된 객체의 영역을 제외한 나머지 영역은 보존된 변환 이미지를 생성하는 단계를 포함할 수 있다.

[0015] 또한, 상기 자연어 문장을 인코딩하는 단계는, 상기 자연어 문장을 의미론적 분석을 통해 단어 요소들로 분절하는 단계 및 상기 분절된 단어 요소 각각에 대한 단어 특징 값들을 생성하는 단계;를 더 포함하고, 상기 자연어 문장을 인코딩하는 단계는, 상기 생성된 단어 특징 값들을 포함하는 자연어 특징 블록을 생성하고, 상기 자연어 특징 블록은 상기 단어 특징 블록을 포함할 수 있다.

[0016] 또한, 상기 입력 이미지의 적어도 하나의 객체를 변환하는 단계는, 상기 이미지 특징 맵의 스케일(scale)을 고려하여 상기 자연어 특징 블록을 확장시킨 자연어 특징 블록을 생성하고, 상기 이미지 특징 블록과 상기 자연어 특징 블록을 결합하여 이미지-자연어 특징 블록을 생성하는 단계 및 상기 이미지-자연어 특징 블록을 입력받아 레지듀얼 변환을 위한 컨벌루션 레이어를 적용하여 레지듀얼 변환 특징 블록을 생성하는 단계를 포함할 수 있다.

[0017] 상술한 목적을 달성하기 위한 본 발명의 다른 실시예에 따른 자연어 문장을 기반으로 하는 변환 이미지 생성 시스템은 변환하고자 하는 적어도 하나의 객체를 포함하는 입력 이미지를 입력받아, 상기 입력 이미지를 인코딩하는 제1 이미지 인코더; 상기 입력 이미지와 관련된 자연어 문장을 입력받아, 상기 자연어 문장을 인코딩하는 제1 텍스트 인코더; 상기 입력된 자연어 문장에 따라 상기 입력 이미지의 적어도 하나의 객체를 변환하는 이미지-자연어 변환부; 및 상기 변환된 객체를 포함하되, 상기 입력 이미지에서 상기 변환된 객체의 영역을 제외한 나머지 영역은 보존된 변환 이미지를 생성하는 변환 이미지 생성부;를 포함하는 생성기 및 상기 생성기로부터 생성된 변환 이미지를 입력받아, 상기 변환 이미지를 인코딩하는 제2 이미지 인코더; 상기 변환 이미지를 검증하기 위한 상기 자연어 문장을 입력받아, 상기 자연어 문장을 인코딩하는 제2 텍스트 인코더; 상기 인코딩된 변환 이미지 및 인코딩된 자연어 문장을 이용하여 상기 변환 이미지를 검증하기 위한 국소 매칭 스코어를 산출하는 자연어 기반 구별부; 및 상기 국소 매칭 스코어를 고려하여 상기 변환 이미지를 재변환하기 위한 재변환여부를 결정하는 재변환여부 결정부;를 포함하는 검증기를 포함할 수 있다.

[0018] 또한, 상기 제2 이미지 인코더는, 컨벌루션 레이어(Convolution layer)를 이용하여 상기 변환 이미지에 대한 변환 이미지 특징 맵을 포함하는 변환 이미지 특징 블록을 생성하고, 상기 변환 이미지 특징 맵은, 상기 컨벌루션 레이어에 의해 컨벌루션 변환된 객체별 이미지 특징 맵들을 포함할 수 있다.

[0019] 또한, 상기 제2 텍스트 인코더는, 상기 자연어 문장을 의미론적 분석을 통해 단어 요소들로 분절하고, 순환 신경망(RNN) 학습을 통해 상기 분절된 단어 요소 각각에 대한 단어 특징 값을 생성할 수 있다.

[0020] 또한, 상기 자연어 기반 구별부는, 상기 제2 텍스트 인코더로부터 생성된 각 단어 특징 값과 상기 각 객체별 이미지 특징 맵을 매칭함에 따라 상기 국소 매칭 스코어를 산출할 수 있다.

[0021] 또한 본 발명은 상기한 방법에 따른 변환 영상 생성 방법을 실행시키는 컴퓨터로 판독 가능한 기록매체에 저장

된 컴퓨터 프로그램을 제안한다.

### 발명의 효과

[0022] 본 발명의 실시예에 따른 변환 영상 생성 장치, 방법 및 이를 이용하는 시스템은 이미지를 변환하고자 하는 자연어 문장에 맞게 이미지를 변환하되, 자연어 문장과 관련이 없는 이미지 내 영역은 손실되지 않고, 그대로 보존함으로써 높은 성능으로 이미지를 변환할 수 있는 효과가 있다.

### 도면의 간단한 설명

[0023] 도1은 본 발명의 일 실시예에 따른 변환 이미지 생성 장치, 방법 및 시스템에 입력되는 입력 이미지 및 자연어 문장과 그에 따라 출력되는 변환 이미지를 예시하여 도시한 도면이다.

도2는 본 발명의 일 실시예에 따른 자연어 문장을 기반으로 하는 변환 이미지 생성 시스템을 개략적으로 도시한 블록도이다.

도3은 본 발명의 일 실시예에 따른 검증기의 구성을 보다 구체적으로 도시한 도면이다.

도4는 본 발명의 변환 이미지 생성 시스템을 이용한 최종 변환 이미지와 SISGAN 및 AttnGAN을 이용한 최종 변환 이미지를 비교한 도면이다.

도5는 본 발명의 변환 이미지 생성 시스템을 이용한 최종 변환 이미지와 SISGAN 및 AttnGAN을 이용한 최종 변환 이미지를 비교한 표이다.

도6 내지 도7은 자연어 기반 구별부가 단어 요소별 국소 매칭 스코어를 산출하는 과정을 나타낸 것이다.

도8은 본 발명의 일 실시예에 따른 변환 이미지 생성 방법을 시간의 순서에 따라 도시한 흐름도이다.

### 발명을 실시하기 위한 구체적인 내용

[0024] 본 발명과 본 발명의 동작상의 이점 및 본 발명의 실시예에 의하여 달성되는 목적을 충분히 이해하기 위해서는 본 발명의 바람직한 실시예를 예시하는 첨부 도면 및 첨부 도면에 기재된 내용을 참조하여야만 한다. 또한, 명세서에 기재된 "...부", "...기", "모듈", "블록"등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.

[0025] 이하, 본 발명의 일 실시예를 첨부된 도면들을 참조하여 상세히 설명한다. 또한, 본 발명을 설명함에 있어, 관련된 공지 구성 또는 기능에 대한 구체적인 설명이 본 발명의 요지를 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명은 생략할 수 있다.

[0026] 이하에서는 본 발명의 실시예에 따른 자연어 문장을 기반으로 하는 변환 이미지 생성 장치, 방법 및 이를 이용하는 시스템의 구성을 관련된 도면을 참조하여 상세히 설명한다.

[0027] 본 발명의 자연어 문장을 기반으로 하는 변환 이미지 생성 장치, 방법 및 이를 이용하는 시스템은 도1에 도시된 바와 같이 입력되는 입력 이미지(100)와 자연어 문장(110)을 기반으로, 입력된 자연어 문장(110)에 맞게 입력 이미지(100)를 변환하여 변환 이미지(120)를 출력할 수 있다. 이와 같은 본 발명의 실시예에 따른 자연어 문장을 기반으로 하는 변환 이미지 생성 장치, 방법 및 시스템은 생성적 적대 신경망(GAN; Generative Adversarial Network)을 기반으로 동작하는 것으로서, 구체적인 설명을 위해 도1 내지 도3을 참조한다.

[0028] 도1은 본 발명의 일 실시예에 따른 변환 이미지 생성 장치, 방법 및 시스템에 입력되는 입력 이미지 및 자연어 문장과 그에 따라 출력되는 변환 이미지를 예시하여 도시한 도면이고, 도2는 본 발명의 일 실시예에 따른 자연어 문장을 기반으로 하는 변환 이미지 생성 시스템을 개략적으로 도시한 블록도이며, 도3은 본 발명의 일 실시예에 따른 검증기의 구성을 보다 구체적으로 도시한 도면이다.

[0029] 먼저, 도2를 참조하면, 본 발명의 자연어 문장을 기반으로 하는 변환 이미지 생성 시스템(200)은 생성기(210) 및 검증기(220)를 포함한다.

[0030] 그리고, 본 발명의 일 실시예에 따른 생성기(210)는 제1 이미지 인코더(211), 제1 텍스트 인코더(212), 이미지-자연어 변환부(213, 214, 215) 및 디코더(216)를 포함하여 구성될 수 있다. 보다 구체적으로 본 발명의 이미지-자연어 변환부는 특징 값 결합부(213), 레지듀얼 블록부(214), 및 합산부(215)로 구성될 수 있다.

[0031] 먼저, 본 발명의 실시예에 따른 제1 이미지 인코더(211)는 도1에 도시된 바와 같은, 변환하고자 하는 적어도 하



나의 객체를 포함하는 입력 이미지(100)를 입력 받아, 입력된 입력 이미지를 인코딩 한다. 설명의 편의를 위하여 도1에 도시된 입력 이미지(100)의 객체는 새(101)와 나무(102)인 것으로 가정하여 설명하도록 한다.

- [0032] 제1 이미지 인코더(211)는 컨벌루션 레이어(convolution layer)를 이용하여 입력 이미지(100)에 대한 이미지 특징 맵을 포함하는 이미지 특징 블록을 생성함에 따라 입력 이미지를 인코딩할 수 있다.
- [0033] 제1 이미지 인코더(211)가 이용하는 신경망에 포함된 컨벌루션 레이어는 컨벌루션 연산을 통해 이미지 특징 맵을 생성한다. 컨벌루션 레이어는 예를 들어, 에지 검출을 위한 레이어, 특정의 객체를 검출하기 위한 레이어등을 포함할 수 있다.
- [0034] 본 실시예에서 이미지 특징 맵은 입력영상에 특정 레이어를 컨벌루션하여 얻어지며, 차원이 축소된 영상으로 예시될 수 있다. 즉, 이미지 인코더에서는 복수개의 이미지 특징맵을 생성하고, 복수개의 이미지 특징맵들을 결합하여 이미지 특징블록을 생성한다
- [0035] 보다 상세하게는, 컨벌루션 레이어는 미리 정의된 컨벌루션 필터를 이용하여 입력 이미지상에서 컨벌루션 연산을 하고, 그 결과 입력 이미지 내의 객체별 특징값들이 맵 상에서 표현되는 특징 맵(feature map)들을 생성할 수 있다. 여기서, 특징 블록이란 생성된 특징 맵들의 집합의 개념으로 볼 수 있고, 특징값이란 벡터나 행렬을 의미한다.
- [0036] 그리고, 본 발명의 실시예에 따른 제1 텍스트 인코더(212)는 입력 이미지(100)와 관련된 자연어 문장(110)을 입력 받아, 자연어 문장을 인코딩 한다. 제1 텍스트 인코더(212)는 순환 신경망(RNN; Recurrent Neural Network) 학습을 통해 자연어 문장에 대한 자연어 특징 값을 포함하는 자연어 특징 블록을 생성함에 따라 자연어 문장을 인코딩 할 수 있다.
- [0037] 본 발명의 제1 텍스트 인코더(212)는 도면에 따로 도시하지는 않았으나, 단어 요소 분절부 및 단어 특징 값 생성부를 포함하여 구성될 수 있다. 여기서, 본 발명의 실시예에 따른 단어 요소 분절부는, 입력된 자연어 문장을 의미론적(semantic) 분석을 통해 단어 요소들로 분절할 수 있고, 단어 특징 값 생성부는, 양방향 순환 신경망(RNN) 학습을 통해 분절된 단어 요소 각각에 대한 단어 특징 값을 생성하고, 생성된 단어 특징 값들을 합친 단어 특징 블록을 생성할 수 있다.
- [0038] 즉, 단어 요소 분절부에 의해 입력된 자연어 문장이 단어 요소로 분절되면, 본 발명의 단어 특징 값 생성부는 각 단어 요소 각각에 대한 단어 특징 값을 생성할 수 있다. 여기서, 단어 특징 값 생성부가 추출하는 단어 특징 값은 단어 특징 값들이 맵 상에 표현된 특징 맵 또는 특징 레이어일 수 있고, 각 단어별 특징 값은 벡터 또는 행렬로 표현될 수 있다. 즉, 제1 텍스트 인코더(212)가 생성하는 상기 자연어 특징 맵은 각 단어 요소별 단어 특징 맵들을 포함하는 개념일 수 있다.
- [0039] 보다 구체적으로 설명하면, 단어 특징 값 생성부는 분절된 복수개의 단어 요소들 각각을 순차적으로 RNN 학습시킨다.
- [0040] 일 예인 도1의 (a)를 참조하면, 단어 요소 분절부는 입력된 자연어 문장을 의미론적으로 분석 함에 따라 "bird", "red", "head", "breast", "grey", "wings"와 같은 단어 요소들로 분절할 수 있다. 이에 따라, 단어 특징 값 생성부는, "bird" 단어 요소를 RNN 학습시켜 단어 특징 값을 생성하여 출력하고, 상기 생성된 "bird" 단어 요소에 따른 특징 값을 포함하여 다음 단어 요소인 "red" 단어 요소를 RNN 학습시켜 이에 따른 단어 특징 값을 생성한다. 이렇게, 단어 특징 값 생성부는 복수개의 단어 요소들 각각을 순차적으로 RNN 학습시킴에 따라 복수개의 단어 특징 값들을 생성할 수 있다.
- [0041] 이와 같은 원리로, 제1 텍스트 인코더(212)는 복수개의 단어 특징 값들을 생성하고, 상기 복수개의 단어 특징 값들을 포함하는 단어 특징 블록을 생성할 수 있다. 여기서, 제1 텍스트 인코더(212)에 의해 생성된 단어 특징 블록의 차원은 제1 이미지 인코더(211)에 의해 생성된 이미지 특징 블록의 차원과 다른 차원으로 생성될 수 있다. 예컨대, 단어 특징 블록은 이미지 특징 블록보다 더 작은 차원으로 생성될 수 있다.
- [0042] 본 발명의 특징 값 결합부(213)는 제1 이미지 인코더(211)로부터 생성된 이미지 특징 맵의 스케일(scale)을 고려하여 제1 텍스트 인코더(212)로부터 생성된 상기 자연어 특징 블록을 확장시킴에 따라 확장된 자연어 특징 블록을 생성하고, 이에 따라 상기 이미지 특징 블록과 상기 확장된 자연어 특징 블록을 결합하여 이미지-자연어 특징 블록을 생성할 수 있다.
- [0043] 그리고, 컨벌루션 연산을 통한 컨벌루션 특징을 추출하는 복수의 컨벌루션 레이어들을 이용하되, 인접한 컨벌루션 레이어를 연결하는 경로 및 상기 인접한 레이어를 적어도 하나 이상 건너 뛰어 다른 컨벌루션 레이어로 연결

되는 레지듀얼(residual) 경로를 형성하는 본 발명의 레지듀얼 블록부(214)는, 상기 특징 값 결합부(213)로부터 생성된 이미지-자연어 특징 블록을 입력 받아 레지듀얼 변환하여 레지듀얼 변환 특징 블록을 생성할 수 있다.

[0044] 즉, 본 발명의 변환 이미지 생성 시스템의 생성기(210)는 ResNet 구조가 적용된 신경망을 이용하여 변환 이미지를 생성할 수 있다. 레지듀얼 블록부(214)는 ResNet 구조가 적용된 신경망으로서, 신경망 내에 레지듀얼 연결(Residual Connection)을 이용하여 기울기 값의 소실 문제(vanishing gradient)를 해결할 수 있는 장점이 있다. 레지듀얼 연결은 하나의 계층을 뛰어 넘는 연결로써, 실제 설계된 연결대로 움직이는 경로와 그 경로를 뛰어 넘어 다음 단계로 직접 연결되는 경로 등 총 두 가지 경로가 존재한다.

[0045] 본 발명의 레지듀얼 블록부(214)는 복수개의 컨벌루션 레이어로 구성되는 레지듀얼 블록을 포함할 수 있고, 이러한 레지듀얼 블록은 설계자가 구현하고자 하는 신경망의 깊이에 따라 하나로 설계될 수도 있고, 복수개의 레지듀얼 블록으로 설계될 수도 있다. 즉, 레지듀얼 블록의 개수에는 특별한 제한은 없다.

[0046] 또한, 레지듀얼 블록부(214)의 복수개의 컨벌루션 레이어들 사이에 풀링 레이어(pooling layer)가 주기적으로 위치하여 신경망의 파라미터의 수나 연산량을 줄이기 위해 깊이는 유지하면서, 차원을 감소시키는 기능을 수행할 수 있다. 레지듀얼 블록부(214)의 컨벌루션 레이어는 자연어 문장에 포함된 객체에 대한 입력 이미지에서의 특정 부분에 따른 특징 값을 추출하고, 풀링 레이어는 추출한 특징 값들 중 제일 중요한 값만 추려낼 수 있다. 이상 설명한 ResNet의 구조는 이미 공지된 기술로서 보다 구체적인 설명은 이하 생략한다.

[0047] 일 실시예에 따라 레지듀얼 블록부(214)를 구성하는 레지듀얼 블록이 복수개로 구현되는 경우, 하나의 레지듀얼 블록의 출력을 인접한 다른 레지듀얼 블록의 입력으로 참조되도록 하며, 이에 따라 신경망의 깊이가 깊어짐에 따라 발생하는 Vanishing gradient 문제를 해결함으로써 최종적으로 자연어 문장에 맞게 변환 이미지가 보다 높은 성능으로 변환되어 출력될 수 있다.

[0048] 본 발명의 합산부(215)는 제1 이미지 인코더(211)로부터 생성된 이미지 특징 블록과, 레지듀얼 블록부(214)로부터 생성된 레지듀얼 변환 특징 블록을 합산하여 합산 특징 블록을 생성할 수 있다.

[0049] 이때, 본 발명의 생성기(210)의 이미지-자연어 변환부는 입력된 자연어 문장에 포함되어 있지 않은 입력 이미지의 일 영역(예를 들어, 배경부분)이 변환(조작)되어 새롭게 생성되는 것을 방지하기 위해, 이미지-자연어 변환부는 아래 <수학식1>을 이용하여, 자연어 문장에 포함되어 있지 않은 상기 입력 이미지의 일 영역을 보존시킬 수 있도록 한다.

### 수학식 1

[0050] 
$$L_{rec} = \|\mathbf{x} - G(\mathbf{x}, \mathbf{t})\|$$

[0051] 여기서,  $L_{rec}$ 는 보존 손실값,  $\mathbf{x}$ 는 입력 이미지(특징 블록),  $\mathbf{t}$ 는 분절된 단어 요소들 중 입력 이미지에 포함된 객체들 중 적어도 하나와 매칭되는 단어 요소(특징 값)를 의미한다.

[0052] 이렇게, 변환 이미지 생성부(216)는 합산부(215)에 의해 생성된 합산 특징 블록을 디코딩하여, 입력된 자연어 문장에 맞게 변환된 객체를 포함하고, 상기 변환된 객체의 영역을 제외한 나머지 영역은 복원된 변환 이미지를 생성할 수 있다.

[0053] 다음으로, 본 발명의 변환 이미지 생성 시스템(200)의 검증기(220)에 대해 설명한다. 검증기(220)는 생성기(210)로부터 생성된 변환 이미지가 자연어 문장에 맞게 얼마나 잘 표현되었는지 검증하기 위한 것이다. 즉, 본 발명의 생성기(210) 및 검증기(220)는 생성적 적대 신경망(GAN) 구조로 구성된다. 여기서, 검증기(220)는, 생성기(210)가 변환 이미지를 생성하기 이전에 상기 변환 이미지를 검증(예를 들어, Real/Fake를 판별)하기 위해 미리 학습된 상태일 수 있다.

[0054] 도2를 참조하면, 본 발명의 검증기(220)는 제2 이미지 인코더(221), 제2 텍스트 인코더(222) 및 자연어 기반 구별부(223) 및 재변환 여부 결정부(미도시)를 포함하여 구성될 수 있다.

[0055] 제2 이미지 인코더(221)는 생성기(210)에서 생성된 변환 이미지를 입력 받고, 컨벌루션 레이어를 이용하여 변환 이미지에 대한 변환 이미지 특징 맵을 포함하는 변환 이미지 특징 블록을 생성함에 따라 변환 이미지를 인코딩할 수 있다.

- [0056] 그리고, 제2 텍스트 인코더(222)는 상기 생성기(210)로부터 생성된 변환 이미지를 검증하기 위한 자연어 문장을 입력 받아, 순환 신경망(RNN) 학습을 통해, 입력된 상기 자연어 문장에 대한 자연어 특징 값(이하, 특징 벡터)을 생성함에 따라 자연어 문장을 인코딩 할 수 있다. 제2 텍스트 인코더(222)가 입력 받는 자연어 문장은, 생성기(210)의 제1 텍스트 인코더(212)가 입력 받았던 자연어 문장과 동일한 문장이다.
- [0057] 또한, 검증기(220)의 제2 이미지 인코더(221) 및 제2 텍스트 인코더(222)는 생성기(210)의 제1 이미지 인코더(211) 및 제1 텍스트 인코더(212)와 동일한 동작을 수행하고, 입력되는 이미지만 상이한 것으로서, 제2 이미지 인코더 및 제2 텍스트 인코더의 구체적인 동작 및 그 동작에 따라 각각 출력되는 특징 블록 및 특징 벡터에 대한 자세한 설명은 생략한다.
- [0058] 제2 이미지 인코더 및 제2 텍스트 인코더 각각에서 변환 이미지 특징 블록과 자연어 특징 벡터가 생성되면, 본 발명의 자연어 기반 구별부(223)는 상기 변환 이미지 특징 블록과 상기 자연어 특징 벡터를 매칭함에 따라 국소 매칭 스코어를 산출할 수 있다. 자연어 기반 구별부(223, 300)의 구체적인 설명을 위해 도3을 참고한다.
- [0059] 도3은 검증기(300)의 구성을 개략적으로 도시한 것이다.
- [0060] 생성기에서 생성된 변환 이미지가 제2 이미지 인코더(310)로 입력되어, 제2 이미지 인코더(310)는 변환 이미지를 인코딩하여 변환 이미지 특징 블록(331)을 생성한다. 그리고, 생성기에서 이용되었던 자연어 문장이 제2 텍스트 인코더(320)로 입력되어, 제2 텍스트 인코더(320)는 자연어 문장을 의미론적인 단어 요소별로 분절하여, 분절된 단어별 단어 특징 벡터(333)을 생성한다.
- [0061] 이때, 변환 이미지 특징 맵은, 제2 이미지 인코더(310) 내에 마련된 컨벌루션 레이어에 의해 컨벌루션 변환된 객체별 이미지 특징 블록들( $j-1, j, j+1$ )(331)을 포함할 수 있다. 이때, 본 발명의 제2 이미지 인코더(310)는 입력된 변환 이미지를 제2 이미지 인코더 내에 마련된 적어도 하나의 컨벌루션 레이어를 거친 후에 오버피팅(overfitting) 문제를 최소화하기 위해, 생성된 최종 특징 맵에  $1 \times 1$  컨벌루션 연산을 통해 GAP(Global Average Pooling)시켜, 생성되는 특징 맵의 차원을 줄여줌으로써 변환 이미지 벡터(332)를 생성할 수 있다.
- [0062] 자연어 기반 구별부(330)는 상기 생성된 변환 이미지 벡터(332)와 단어 특징 벡터(333)를 매칭시키는 국소 구별자(LD; Local Discriminator)(334)를 더 포함할 수 있고, 국소 구별자(334)는 변환 이미지 벡터(332)와 단어 특징 벡터(333)을 매칭시킴에 따라 국소 매칭 스코어를 산출할 수 있다.
- [0063] 예를 들어, 자연어 기반 구별부(330)는 제2 텍스트 인코더(222)에서 분절된 단어 요소가 3개라고 할 때, 분절된 단어 요소 각각에 상응하는 국소 구별자들은 단어별 단어 특징 벡터들( $w_{i-1}, w_i, w_{i+1}$ )(333) 각각을 이미지 벡터( $v_{j-1}, v_j, v_{j+1}$ )(332) 각각과 매칭시켜 단어별 국소 매칭 스코어를 각각 산출할 수 있다.
- [0064] 국소 매칭 스코어는 국소 구별자가 해당 단어 요소가 변환 이미지에 존재하는지를 판단함에 따라 산출될 수 있으며, 아래 <수학식2> 내지 <수학식5>를 이용하여 국소 매칭 스코어를 산출할 수 있다.

## 수학식 2

$$f_{w_i}(v) = \sigma(W(w_i) \cdot v + b(w_i))$$

- [0065]
- [0066] 여기서,  $f_{w_i}$ 는 국소 구별자가 국소 매칭 스코어를 산출하기 위해 사용하는 함수이고,  $w_i$ 는 단어 특징 벡터( $i$ 번째 단어 요소에 따른 단어 특징 벡터)이며,  $v$ 는 변환 이미지 특징 블록이 GAP(Global Average Pooling) 처리되어 추출된 이미지 벡터이고,  $W$ 는 가중치(weight),  $b$ 는 바이어스(bias)이며,  $\sigma$ 는 시그모이드(sigmoid) 함수를 나타낸다
- [0067] 또한, 본 발명의 자연어 기반 구별부는 분절된 단어 요소별 가중치에 따라서 비교적 주요하다고 판단되지 않는 단어 요소의 영향을 줄이기 위해 아래 <수학식3>을 적용한다.

## 수학식 3

$$\alpha_i = \frac{\exp(\mathbf{u}^T \mathbf{w}_i)}{\sum_i \exp(\mathbf{u}^T \mathbf{w}_i)}$$

[0068]

[0069] 여기서,  $\mathbf{u}^T$ 는 단어 특징 벡터에 대한 시간 평균,  $\mathbf{w}_i$ 는 (i번째) 단어 특징 벡터,  $\alpha_i$ 는 소프트맥스 함수를 통해 구한 i번째 단어의 중요도를 나타낸다.

[0070] 그리고, 본 발명의 자연어 기반 구별부는 아래 <수학식4>를 이용하여 각 단어 요소별 국소 매칭 스코어를 종합하여 최종 이미지-자연어 스코어를 산출할 수 있다.

## 수학식 4

$$D(\mathbf{x}, \mathbf{t}) = \prod_{i=1}^T [f_{\mathbf{w}_i}(\mathbf{v})]^{\alpha_i}$$

[0071]

[0072] 여기서,  $\mathbf{x}$ 는 입력단에 들어가는 이미지이고,  $\mathbf{t}$ 는 입력된 이미지  $\mathbf{x}$ 의 내용과 매칭되는 의미를 담고 있는 단어이며,  $\mathbf{v}$ 는 이미지 인코더를 통해 구한 이미지의 특징 벡터이고,  $f_{\mathbf{w}_i}$ 는 국소 구별자가 국소 매칭 스코어를 산출하기 위해 사용하는 함수이다.

[0073] 또한, 본 발명의 검증기는 아래 <수학식5>를 이용하여, 다중 차원으로 생성된 이미지 특징 맵을 모두 고려하여 최종 이미지-자연어 스코어를 산출할 수 있다.

## 수학식 5

$$D(\mathbf{x}, \mathbf{t}) = \prod_{i=1}^T [\sum_j \beta_{ij} f_{\mathbf{w}_{i,j}}(\mathbf{v}_j)]^{\alpha_i}$$

[0074]

[0075] 여기서,  $\mathbf{v}_j$ 는 j번째 레이어의 이미지 벡터이고,  $\beta_{ij}$ 는 단어 특징 벡터에 대한 j번째 레이어의 이미지 벡터의 중요도를 결정하는 가중치(softmax weight)이다.

[0076] 이에 따라, 재변환 여부 결정부(340)는 산출된 국소 매칭 스코어들을 고려하여 변환 이미지를 재변환하기 위한 재변환 여부를 결정할 수 있다.

[0077] 즉, 재변환 여부 결정부(340)는 각 국소 구별자(LD)에서 산출된 국소 매칭 스코어들을 고려하여 변환 이미지에 대한 최종 스코어가 미리 설정된 기준 스코어 보다 낮으면 생성기의 제1 이미지 인코더, 제1 텍스트 인코더 및 이미지-자연어 변환부의 각 가중치를 조절할 수 있다.

## 수학식 6

$$L_D = \mathbb{E}_{\mathbf{x}, \mathbf{t}, \hat{\mathbf{t}} \sim p_{data}} [\log D(\mathbf{x}) + \lambda_1 (\log D(\mathbf{x}, \mathbf{t}) + \log (1 - D(\mathbf{x}, \hat{\mathbf{t}})))] \\ + \mathbb{E}_{\mathbf{x}, \hat{\mathbf{t}} \sim p_{data}} [\log (1 - D(G(\mathbf{x}, \hat{\mathbf{t}})))]$$

[0078]

$$L_G = \mathbb{E}_{\mathbf{x}, \hat{\mathbf{t}} \sim p_{data}} [\log D(\mathbf{x}) + \lambda_1 \log D(G(\mathbf{x}, \hat{\mathbf{t}}), \hat{\mathbf{t}})] + \lambda_2 L_{rec}$$

[0079]

[0080]

여기서,  $\mathbf{x}$ 는 입력된 변환 이미지의 특징 맵,  $\hat{\mathbf{t}}$ 는 분절된 단어 요소들 중 변환 이미지에 포함된 객체들 중 적어도 하나와 매칭되는 단어 요소의 단어 특징 맵,  $\hat{\mathbf{t}}$ 는 분절된 단어 요소들 중 변환 이미지와 매칭되지 않는 단어 요소의 단어 특징 맵,  $\lambda$ 는 추가 손실의 조절을 위한 가중치,  $L_D$ 는 검증기의 학습을 위해 정의된 가치함수,  $L_G$ 는 생성기의 학습을 위해 정의된 가치함수이다.

[0081]

도4는 본 발명의 변환 이미지 생성 시스템을 이용한 최종 변환 이미지와 SISGAN 및 AttnGAN을 이용한 최종 변환 이미지를 비교한 도면이다. 도4를 참고하면, SISGAN 및 AttnGAN은 입력된 자연어 문장에 포함된 객체에 대한 변환은 이루어졌지만, 자연어 문장과 관련이 없는 객체 및 영역은 보존하지 못한 반면, 본 발명의 변환 이미지 생성 시스템을 이용하였을 때는, 입력 이미지를 자연어 문장에 맞게 변환함과 동시에 자연어 문장과 관련이 없는 부분은 보존할 수 있었다.

[0082]

도5는 본 발명의 변환 이미지 생성 시스템을 이용한 연구 결과와 SISGAN 및 AttnGAN을 이용한 연구 결과를 비교한 결과표이다. 연구 비교는 통상의 연구자들의 평가에 의해 이루어진 결과로, 평가 기준으로는 변환된 변환 이미지의 시각적 속성(색, 텍스처)이 자연어 문장과 일치하는지, 자연어 문장과 관련 없는 객체 및 영역이 보존되었는지를 반영하는 변환 이미지의 정확도 및 자연스러움 등을 바탕으로 평가되었다.

[0083]

도5에 나타난 바와 같이, 통상의 연구자들은 본 발명의 변환 이미지 생성 시스템을 이용하여 변환된 변환 이미지의 정확도 및 자연스러운 정도가 SISGAN 및 AttnGAN을 이용한 변환 이미지보다 우수한 것으로 평가되었다. 또한, 자연어 문장과 관련 없는 객체 및 영역의 보존에 대한 보존 오류( $L_2$  error) 역시 다른 종래의 기술보다 성능이 높게 나타난 것으로 확인되었다.

[0084]

도6은 입력 이미지(610)를 변환 이미지(650)와 같이 변환시키기 위해, 본 발명의 자연어 기반 구별부가 자연어 문장(600)을 의미론적 단어 요소(brown, black, wings)별로 분절함에 따라, 참조번호 620 내지 640과 같이 각 단어 요소별 국소 매칭 스코어를 산출하는 과정을 나타낸 것이다.

[0085]

마찬가지로, 도7은 입력 이미지(710)를 변환 이미지(750)와 같이 변환시키기 위해, 본 발명의 자연어 기반 구별부가 자연어 문장(700)을 의미론적 단어 요소(flower, and, yellow)별로 분절함에 따라, 참조번호 720 내지 740과 같이 각 단어 요소별 국소 매칭 스코어를 산출하는 과정을 나타낸 것이다.

[0086]

다음으로, 도8은 본 발명의 자연어 문장을 기반으로 하는 변환 이미지 생성 방법을 시간의 흐름에 따라 도시한 흐름도이다.

[0087]

먼저, S800 단계에서 생성기의 제1 이미지 인코더가 변환하고자 하는 적어도 하나의 객체를 포함하는 입력 이미지를 입력 받고, 컨벌루션 레이어(Convolution layer)를 이용하여 상기 입력 이미지에 대한 이미지 특징 맵을 추출함에 따라 상기 입력 이미지를 인코딩한다.

[0088]

그리고, S810 단계에서 제1 텍스트 인코더가 상기 입력 이미지와 관련된 자연어 문장을 입력 받고, 순환 신경망(RNN) 학습을 통해 상기 자연어 문장에 대한 자연어 특징 맵을 추출함에 따라 상기 자연어 문장을 인코딩한다.

[0089]

그리고, S820 단계에서, 이미지-자연어 변환부가 입력된 자연어 문장에 따라 상기 입력 이미지의 적어도 하나의 객체를 변환한다.

[0090]

그리고, S830 단계에서, 변환 이미지 생성부가 상기 변환된 객체를 포함하되, 상기 입력 이미지에서 상기 변환된 객체의 영역을 제외한 나머지 영역은 보존된 변환 이미지를 생성한다.

[0091]

그리고, S840 단계에서 검증기의 제2 이미지 인코더는 상기 생성기로부터 생성된 변환 이미지를 입력 받고, 컨벌루션 레이어(Convolution layer)를 이용하여 상기 변환 이미지에 대한 변환 이미지 특징 맵을 생성함에 따라 상기 변환 이미지를 인코딩한다.

[0092]

그리고, S850 단계에서 제2 텍스트 인코더는 상기 변환 이미지를 검증하기 위한 상기 자연어 문장을 입력 받아, 순환 신경망(RNN) 학습을 통해 상기 자연어 문장에 대한 자연어 특징 맵을 생성함에 따라 상기 자연어 문장을 인코딩한다.



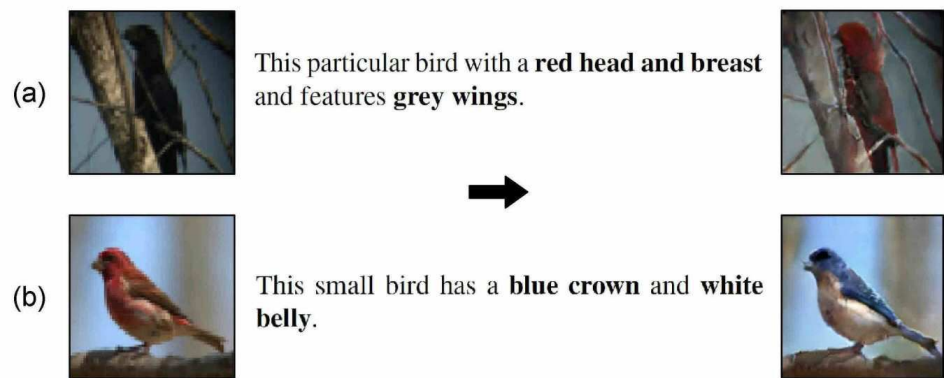
- [0093] 그리고, S860 단계에서 자연어 기반 구별부는 변환 이미지 특징 맵과 상기 자연어 특징 맵을 매칭함에 따라 국소 매칭 스코어를 산출한다.
- [0094] 이에 따라, S870 단계에서 재변환 여부 결정부는 상기 산출된 국소 매칭 스코어를 고려하여 상기 변환 이미지를 재변환하기 위한 재변환 여부를 결정한다.
- [0095] 상술한 바와 같은, 본 발명의 변환 이미지 생성 시스템, 장치 및 방법은 자연어 문장을 의미론적인 단어들로 분절함으로써, 개별적 특징을 먼저 정의하고, 분절된 각각의 단어들의 특징을 이미지와 매칭시키는 방법과, 국소 구별을 통해 각 단어 요소별 국소 매칭 스코어를 산출함으로써 랭킹 손실을 최소화할 수 있다. 또한, 본 발명의 변환 이미지 생성 시스템은 검증기를 보조 손실(auxiliary loss)로 사용하지 않는다.
- [0096] 이상에서 설명한 본 발명의 실시예를 구성하는 모든 구성요소들이 하나로 결합하거나 결합하여 동작하는 것으로 기재되어 있다고 해서, 본 발명이 반드시 이러한 실시예에 한정되는 것은 아니다. 즉, 본 발명의 목적 범위 안에서라면, 그 모든 구성요소들이 하나 이상으로 선택적으로 결합하여 동작할 수도 있다. 또한, 그 모든 구성요소들이 각각 하나의 독립적인 하드웨어로 구현될 수 있지만, 각 구성요소들의 그 일부 또는 전부가 선택적으로 조합되어 하나 또는 복수개의 하드웨어에서 조합된 일부 또는 전부의 기능을 수행하는 프로그램 모듈을 갖는 컴퓨터 프로그램으로서 구현될 수도 있다. 또한, 이와 같은 컴퓨터 프로그램은 USB 메모리, CD 디스크, 플래쉬 메모리 등과 같은 컴퓨터가 읽을 수 있는 기록매체(Computer Readable Media)에 저장되어 컴퓨터에 의하여 읽혀지고 실행됨으로써, 본 발명의 실시예를 구현할 수 있다. 컴퓨터 프로그램의 기록매체로서는 자기 기록매체, 광 기록매체 등이 포함될 수 있다.
- [0097] 이상의 설명은 본 발명의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 발명의 본질적인 특성에서 벗어나지 않는 범위 내에서 다양한 수정, 변경 및 치환이 가능할 것이다. 따라서, 본 발명에 개시된 실시예 및 첨부된 도면들은 본 발명의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이고, 이러한 실시예 및 첨부된 도면에 의하여 본 발명의 기술 사상의 범위가 한정되는 것은 아니다. 본 발명의 보호 범위는 아래의 청구 범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 발명의 권리 범위에 포함되는 것으로 해석되어야 할 것이다.

## 부호의 설명

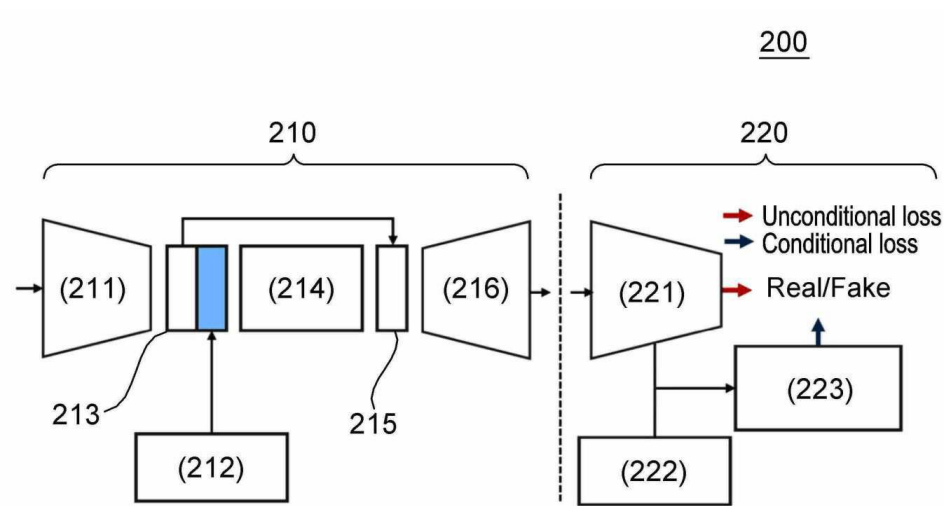
- [0098] 200: 변환 이미지 생성 시스템
- 210: 생성기
- 211: 제1 이미지 인코더
- 212: 제1 텍스트 인코더
- 213: 특징 값 결합부
- 214: 레지듀얼 블록부
- 215: 합산부
- 216: 변환 이미지 생성부
- 220: 검증기
- 221: 제2 이미지 인코더
- 222: 제2 텍스트 인코더
- 223: 자연어 기반 구별부

도면

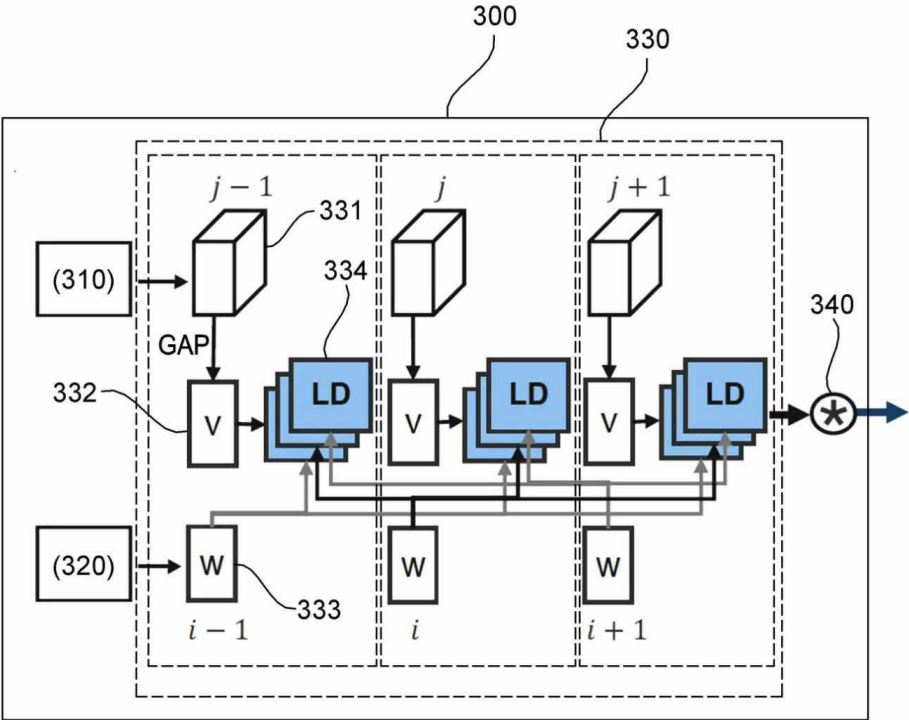
도면1



도면2

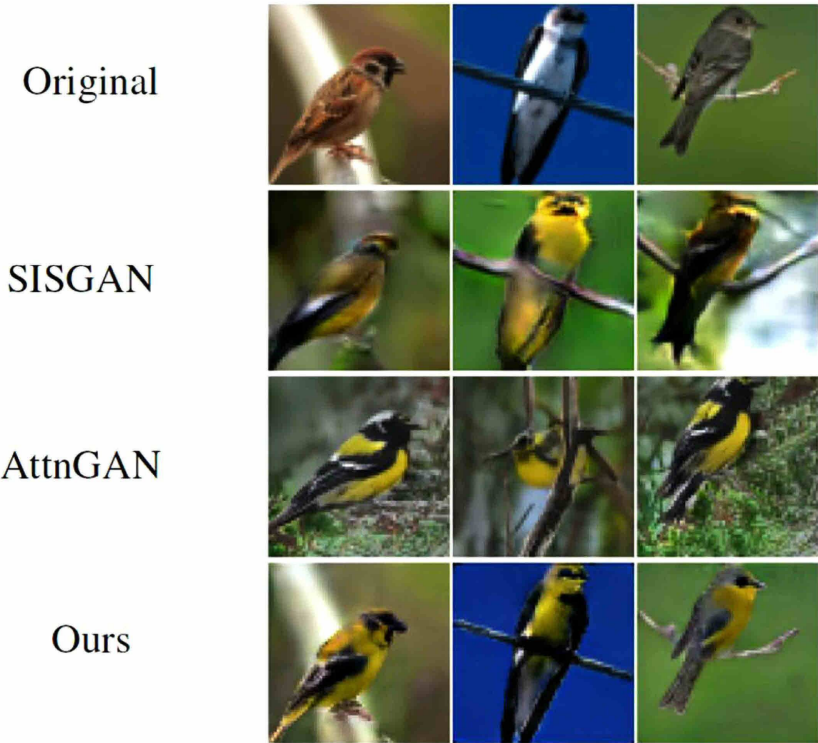


도면3



도면4

This is a **black bird** with **gray**  
**and white wings** and a **bright**  
**yellow belly and chest.**

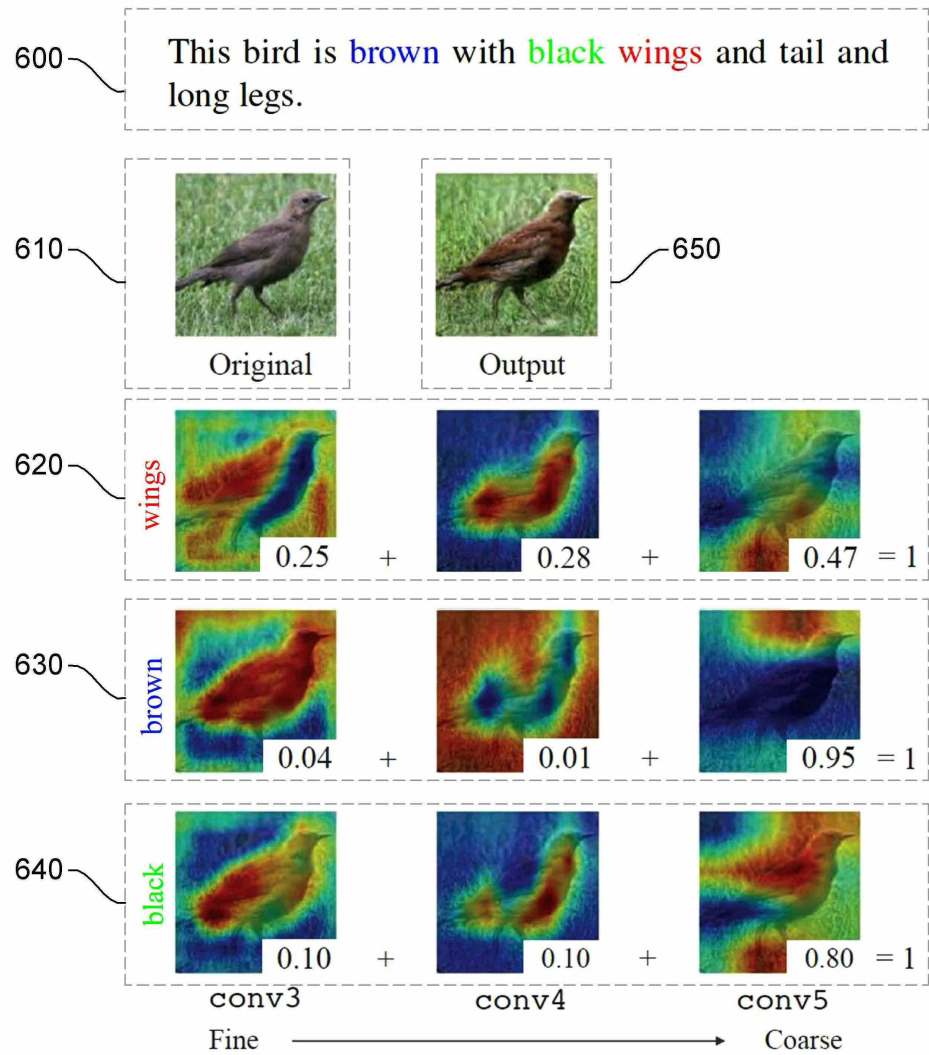




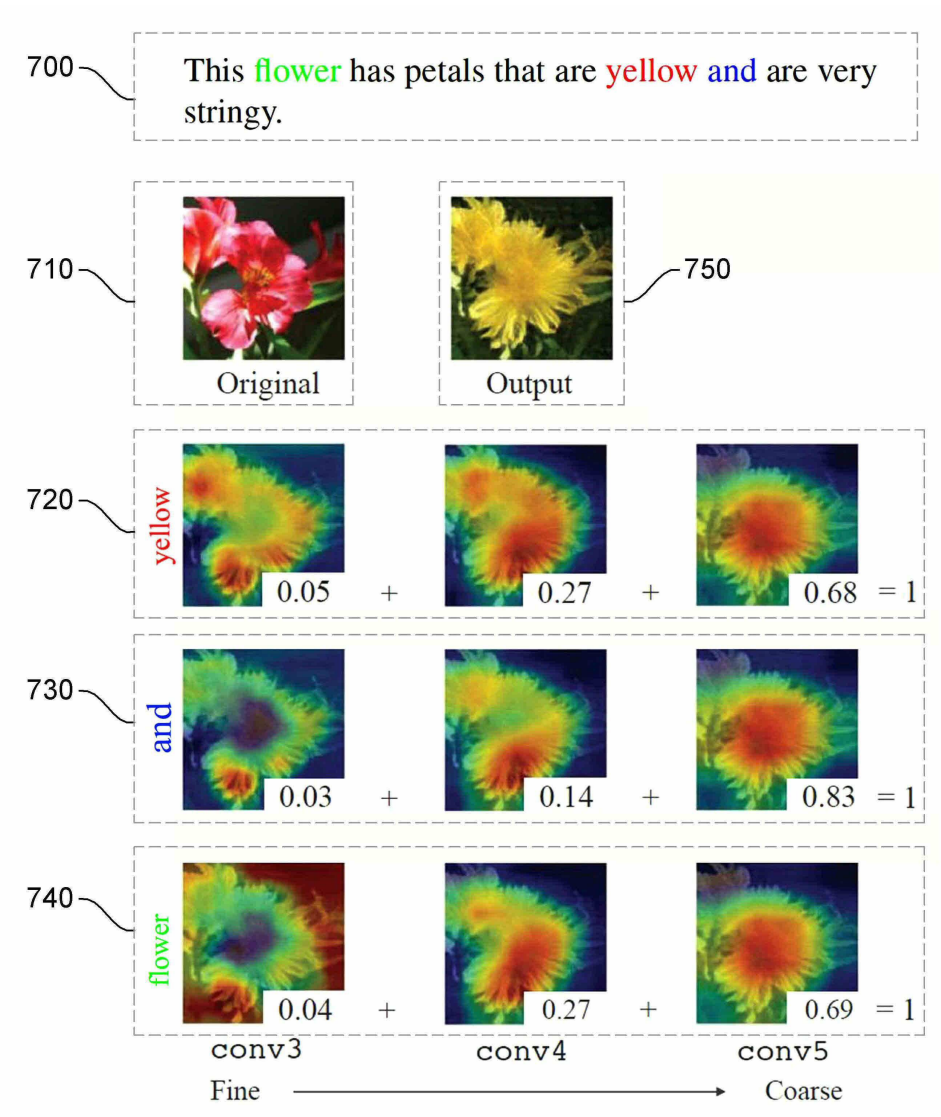
도면5

Method	Accuracy	Naturalness	$L_2$ error
SISGAN	2.33	2.34	0.30
AttnGAN	2.19	2.11	0.25
Ours	<b>1.49</b>	<b>1.56</b>	<b>0.11</b>

도면6



도면7



도면8

