



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2024년04월19일

(11) 등록번호 10-2659399

(24) 등록일자 2024년04월17일

- (51) 국제특허분류(Int. Cl.)  
G06V 20/70 (2022.01) G06V 10/44 (2022.01)  
G06V 20/60 (2022.01)
- (52) CPC특허분류  
G06V 20/70 (2023.08)  
G06V 10/454 (2023.08)
- (21) 출원번호 10-2021-0165598
- (22) 출원일자 2021년11월26일  
심사청구일자 2021년11월26일
- (65) 공개번호 10-2023-0078134
- (43) 공개일자 2023년06월02일
- (56) 선행기술조사문헌  
KR1020210069467 A\*  
KR1020200130759 A\*  
\*는 심사관에 의하여 인용된 문헌

- (73) 특허권자  
연세대학교 산학협력단  
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
- (72) 발명자  
함범섭  
서울특별시 강남구 압구정로61길 37, 72동 506호 (압구정동, 한양아파트)  
백동현  
서울특별시 서대문구 신촌로7안길 28, 310호(창천동)  
오영민  
서울특별시 마포구 성미산로32길 10, 201호(연남동, 그린하우스)
- (74) 대리인  
민영준

전체 청구항 수 : 총 14 항

심사관 : 정용택

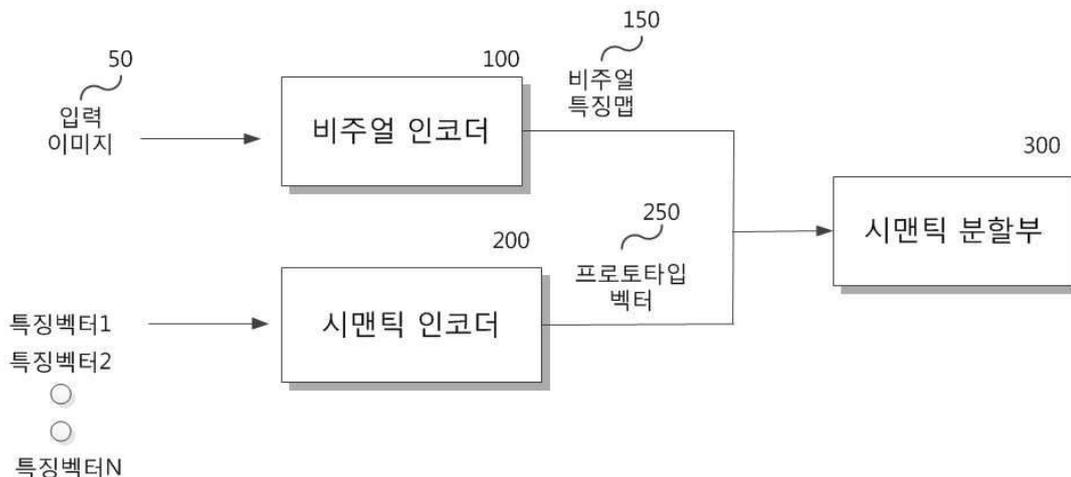
(54) 발명의 명칭 제로샷 시맨틱 분할 장치 및 방법

(57) 요약

제로샷 시맨틱 분할 장치 및 방법이 개시된다. 개시된 장치는, 입력 이미지를 입력받아 신경망 연산을 통해 비주얼 특징맵을 출력하는 비주얼 인코더; 클래스별 특징 벡터를 입력받아 신경망 연산을 통해 클래스별 프로토타입 벡터를 출력하는 시맨틱 인코더; 상기 클래스별 프로토타입 벡터와 상기 비주얼 특징맵의 픽셀별 채널 벡터를 비

(뒷면에 계속)

대표도 - 도1



교하여 상기 비주얼 특징맵의 픽셀 각각에 대해 클래스를 지정하는 시맨틱 분할부를 포함하되, 상기 시맨틱 분할부는 특정 픽셀의 채널 벡터와 가장 유사한 프로토타입 벡터에 상응하는 클래스를 해당 픽셀의 클래스로 지정하며, 상기 프로토타입 벡터와 상기 채널 벡터는 동일한 길이로 설정되고, 상기 비주얼 인코더와 시맨틱 인코더는 적어도 하나의 동일한 손실을 공유하여 동시에 학습된다. 개시된 장치 및 방법에 의하면, 학습되지 않은 클래스에 대해 판별적 방식으로 시맨틱 분할을 수행하여 지속적인 분류기 학습을 요구하지 않으며, 학습되지 않은 클래스를 기 학습된 클래스로 분류하는 편향 문제를 저감시킬 수 있는 장점이 있다.

(52) CPC특허분류

*G06V 20/60* (2023.08)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711126179
과제번호	2016-0-00197-006
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	디지털콘텐츠원천기술개발(R&D, 정보화)
연구과제명	스마트카 다중 센서와 딥러닝을 이용한 초정밀 내츄럴 3D 뷰 생성 기술 개발
기 여 율	1/1
과제수행기관명	연세대학교 산학협력단
연구기간	2021.01.01 ~ 2021.12.31

공지예외적용 : 있음

---

## 명세서

### 청구범위

#### 청구항 1

입력 이미지를 입력받아 신경망 연산을 통해 비주얼 특징맵을 출력하는 비주얼 인코더;

클래스별 특징 벡터를 입력받아 신경망 연산을 통해 클래스별 프로토타입 벡터를 출력하는 시맨틱 인코더;

상기 클래스별 프로토타입 벡터와 상기 비주얼 특징맵의 픽셀별 채널 벡터를 비교하여 상기 비주얼 특징맵의 픽셀 각각에 대해 클래스를 지정하는 시맨틱 분할부를 포함하되,

상기 시맨틱 분할부는 특정 픽셀의 채널 벡터와 가장 유사한 프로토타입 벡터에 상응하는 클래스를 해당 픽셀의 클래스로 지정하며,

상기 프로토타입 벡터와 상기 채널 벡터는 동일한 길이로 설정되고, 상기 비주얼 인코더와 시맨틱 인코더는 적어도 하나의 동일한 손실을 공유하여 동시에 학습되며,

상기 비주얼 인코더와 시맨틱 인코더가 공유하는 손실은 프로토타입 손실을 포함하고, 상기 프로토타입 손실은 상기 시맨틱 인코더에서 출력되는 특정 클래스의 프로토타입 벡터와 상기 비주얼 특징맵의 해당 클래스의 채널 벡터들의 중간값 사이의 손실에 상응하며,

상기 프로토타입 손실을 연산하기 위해 상기 시맨틱 인코더로는 상기 입력 이미지에 클래스별 특징 벡터를 적용하여 생성되는 제1 시맨틱 분할맵을 입력하는 것을 특징으로 하는 시맨틱 분할 장치.

#### 청구항 2

삭제

#### 청구항 3

제1항에 있어서,

상기 프로토타입 손실에 기초하여 상기 비주얼 인코더에서 출력되는 특정 클래스의 채널 벡터들의 중간값은 상기 시맨틱 인코더에서 출력하는 해당 클래스의 프로토타입 벡터와 동일해지는 방향으로 학습되는 것을 특징으로 하는 시맨틱 분할 장치.

#### 청구항 4

제1항에 있어서,

상기 비주얼 인코더와 시맨틱 인코더가 공유하는 손실은 크로스 엔트로피 손실을 포함하며,

상기 크로스 엔트로피 손실에 의해 상기 비주얼 인코더는 동일한 클래스의 채널 벡터들은 임베딩 공간상에서 상대적으로 가까이 위치하고 다른 클래스의 채널 벡터들은 임베딩 공간상에서 상대적으로 멀리 위치하도록 학습되는 것을 특징으로 하는 시맨틱 분할 장치.

#### 청구항 5

제1항에 있어서,

상기 시맨틱 인코더는 상기 시맨틱 인코더로 입력되는 클래스별 특징 벡터들 사이의 거리와 상기 시맨틱 인코더에서 출력되는 클래스별 프로토타입 벡터들 사이의 거리가 동일해지도록 시맨틱 손실을 이용하여 학습이 이루어지는 것을 특징으로 하는 시맨틱 분할 장치.

**청구항 6**

삭제

**청구항 7**

제1항에 있어서,

상기 프로토타입 손실은 다음의 수학적식과 같이 연산되는 것을 특징으로 하는 시맨틱 분할 장치.

$$\mathcal{L}_{\text{center}} = \frac{1}{\sum_{c \in \mathcal{S}} |\mathcal{R}_c|} \sum_{c \in \mathcal{S}} \sum_{\mathbf{p} \in \mathcal{R}_c} d(v(\mathbf{p}), \mu(\mathbf{p}))$$

위 수학적식에서,  $\mathcal{L}_{\text{center}}$ 는 프로토타입 손실,  $c$ 는 클래스,  $\mathcal{S}$ 는 클래스 총 집합,  $\mathbf{p}$ 는 픽셀,  $\mathcal{R}_c$ 는 특정 클래스의 픽셀 집합,  $v(\mathbf{p})$ 는 비주얼 특징맵에서 픽셀 위치  $\mathbf{p}$ 에서의 채널 벡터,  $\mu(\mathbf{p})$ 는 제1 시맨틱 분할맵에서 픽셀 위치  $\mathbf{p}$ 의 특징 벡터를 입력하여 시맨틱 인코더에서 출력되는 프로토타입 벡터이고,  $d()$ 는 두 변수의 거리를 출력하는 함수임.

**청구항 8**

제3항에 있어서,

상기 프로토타입 손실을 연산하기 위해 상기 시맨틱 인코더로는 상기 입력 이미지의 클래스별로 특징 벡터를 적용하고, 상기 클래스별 특징 벡터가 적용된 이미지를 축소하고, 상기 축소된 이미지를 선형 보간을 통해 본래의 이미지 크기로 확대한 제2 시맨틱 분할맵이 입력되는 것을 특징으로 하는 시맨틱 분할 장치.

**청구항 9**

제8항에 있어서,

상기 프로토타입 손실은 다음의 수학적식과 같이 연산되는 것을 특징으로 하는 시맨틱 분할 장치.

$$\mathcal{L}_{\text{bar}} = \frac{1}{\sum_{c \in \mathcal{S}} |\mathcal{R}_c|} \sum_{c \in \mathcal{S}} \sum_{\mathbf{p} \in \mathcal{R}_c} d(v(\mathbf{p}), \tilde{\mu}(\mathbf{p}))$$

위 수학적식에서,  $\mathcal{L}_{\text{bar}}$ 는 프로토타입 손실,  $c$ 는 클래스,  $\mathcal{S}$ 는 클래스 총 집합,  $\mathbf{p}$ 는 픽셀,  $\mathcal{R}_c$ 는 특정 클래스의 픽셀 집합,  $v(\mathbf{p})$ 는 비주얼 특징맵에서 픽셀 위치  $\mathbf{p}$ 에서의 채널 벡터,  $\tilde{\mu}(\mathbf{p})$ 는 제2 시맨틱 분할맵에서 픽셀 위치  $\mathbf{p}$ 의 특징 벡터를 입력하여 시맨틱 인코더에서 출력되는 프로토타입 벡터이고,  $d()$ 는 두 변수의 거리를 출력하는 함수임.

**청구항 10**

입력 이미지를 입력받아 신경망 연산을 통해 비주얼 특징맵을 비주얼 인코더를 통해 출력하는 단계(a);

클래스별 특징 벡터를 입력받아 신경망 연산을 통해 클래스별 프로토타입 벡터를 시맨틱 인코더를 통해 출력하는 단계(b);

상기 클래스별 프로토타입 벡터와 상기 비주얼 특징맵의 픽셀별 채널 벡터를 비교하여 상기 비주얼 특징맵의 픽셀 각각에 대해 클래스를 지정하는 단계(c)를 포함하되,

상기 단계 (c)는 특정 픽셀의 채널 벡터와 가장 유사한 프로토타입 벡터에 상응하는 클래스를 해당 픽셀의 클래스로 지정하며,

상기 프로토타입 벡터와 상기 채널 벡터는 동일한 길이로 설정되고, 상기 비주얼 인코더와 시맨틱 인코더는 적어도 하나의 동일한 손실을 공유하여 동시에 학습되고,

상기 비주얼 인코더와 시맨틱 인코더가 공유하는 손실은 프로토타입 손실을 포함하며, 상기 프로토타입 손실은 상기 시맨틱 인코더에서 출력되는 특정 클래스의 프로토타입 벡터와 상기 비주얼 특징맵의 해당 클래스의 채널 벡터들의 중간값 사이의 손실에 상응하고,

상기 프로토타입 손실을 연산하기 위해 상기 시맨틱 인코더로는 상기 입력 이미지에 클래스별 특징 벡터를 적용하여 생성되는 제1 시맨틱 분할맵을 입력하는 것을 특징으로 하는 시맨틱 분할 방법.

### 청구항 11

삭제

### 청구항 12

제10항에 있어서,

상기 프로토타입 손실에 기초하여 상기 비주얼 인코더에서 출력되는 특정 클래스의 채널 벡터들의 중간값은 상기 시맨틱 인코더에서 출력하는 해당 클래스의 프로토타입 벡터와 동일해지는 방향으로 학습되는 것을 특징으로 하는 시맨틱 분할 방법.

### 청구항 13

제10항에 있어서,

상기 비주얼 인코더와 시맨틱 인코더가 공유하는 손실은 크로스 엔트로피 손실을 포함하며,

상기 크로스 엔트로피 손실에 의해 상기 비주얼 인코더는 동일한 클래스의 채널 벡터들은 임베딩 공간상에서 상대적으로 가까이 위치하고 다른 클래스의 채널 벡터들은 임베딩 공간상에서 상대적으로 멀리 위치하도록 학습되는 것을 특징으로 하는 시맨틱 분할 방법.

### 청구항 14

제10항에 있어서,

상기 시맨틱 인코더는 상기 시맨틱 인코더로 입력되는 클래스별 특징 벡터들 사이의 거리와 상기 시맨틱 인코더에서 출력되는 클래스별 프로토타입 벡터들 사이의 거리가 동일해지도록 시맨틱 손실을 이용하여 학습이 이루어지는 것을 특징으로 하는 시맨틱 분할 방법.

### 청구항 15

삭제

### 청구항 16

제10항에 있어서,

상기 프로토타입 손실은 다음의 수학적식과 같이 연산되는 것을 특징으로 하는 시맨틱 분할 방법.

$$\mathcal{L}_{\text{center}} = \frac{1}{\sum_{c \in \mathcal{S}} |\mathcal{R}_c|} \sum_{c \in \mathcal{S}} \sum_{\mathbf{p} \in \mathcal{R}_c} d(v(\mathbf{p}), \mu(\mathbf{p}))$$

위 수학적식에서,  $\mathcal{L}_{\text{center}}$ 는 프로토타입 손실,  $c$ 는 클래스,  $\mathcal{S}$ 는 클래스 총 집합,  $\mathbf{p}$ 는 픽셀,  $\mathcal{R}_c$ 는 특정 클래스의 픽셀 집합,  $v(\mathbf{p})$ 는 비주얼 특징맵에서 픽셀 위치  $\mathbf{p}$ 에서의 채널 벡터,  $\mu(\mathbf{p})$ 는 제1 시맨틱 분할맵에서 픽셀 위치  $\mathbf{p}$ 의 특징 벡터를 입력하여 시맨틱 인코더에서 출력되는 프로토타입 벡터이고,  $d()$ 는 두 변수의 거리를 출력하는 함수임.

**청구항 17**

제12항에 있어서,

상기 프로토타입 손실을 연산하기 위해 상기 시맨틱 인코더로는 상기 입력 이미지에 클래스별 특징 벡터를 적용하고, 상기 클래스별 특징 벡터가 적용된 이미지를 축소하고, 상기 축소된 이미지를 선형 보간을 통해 본래의 이미지 크기로 확대한 제2 시맨틱 분할맵이 입력되는 것을 특징으로 하는 시맨틱 분할 방법.

**청구항 18**

제17항에 있어서,

상기 프로토타입 손실은 다음의 수학적식과 같이 연산되는 것을 특징으로 하는 시맨틱 분할 방법.

$$\mathcal{L}_{\text{bar}} = \frac{1}{\sum_{c \in \mathcal{S}} |\mathcal{R}_c|} \sum_{c \in \mathcal{S}} \sum_{\mathbf{p} \in \mathcal{R}_c} d(v(\mathbf{p}), \tilde{\mu}(\mathbf{p}))$$

위 수학적식에서,  $\mathcal{L}_{\text{bar}}$ 는 프로토타입 손실,  $c$ 는 클래스,  $\mathcal{S}$ 는 클래스 총 집합,  $\mathbf{p}$ 는 픽셀,  $\mathcal{R}_c$ 는 특정 클래스의 픽셀 집합,  $v(\mathbf{p})$ 는 비주얼 특징맵에서 픽셀 위치  $\mathbf{p}$ 에서의 채널 벡터,  $\tilde{\mu}(\mathbf{p})$ 는 제2 시맨틱 분할맵에서 픽셀 위치  $\mathbf{p}$ 의 특징 벡터를 입력하여 시맨틱 인코더에서 출력되는 프로토타입 벡터이고,  $d()$ 는 두 변수의 거리를 출력하는 함수임.

**발명의 설명**

**기술 분야**

[0001] 본 발명은 시맨틱 분할 장치 및 방법에 관한 것으로서, 더욱 상세하게는 학습되지 않은 클래스에 대해서도 시맨틱 분할이 가능한 제로샷 시맨틱 분할 장치 및 방법에 관한 것이다.

**배경 기술**

[0003] 시맨틱 분할(semantic segmentation)은 입력 영상을 식별 가능한 클래스 각각에 대응하는 영역별로 분할하는 것을 의미하며, 자율 주행, 의료 영상, 영상 편집 등 다양한 응용 분야에 적용될 수 있다. 이러한 시맨틱 영상 분할은 입력 영상의 다수의 픽셀 각각을 사람, 자동차, 자전거 등과 같은 객체를 지정된 클래스로 분류하여 레이블링하는 것을 목표로 한다.

[0004] CNN(convolutional neural networks) 과 같은 인공 신경망을 이용하는 딥 러닝 기반 시맨틱 영상 분할 기술들은 우수한 성능을 나타내지만, 학습을 위해서는 각 객체의 클래스가 픽셀 단위로 레이블되어 클래스별 객체 영역이 정확하게 표현된 픽셀 수준(pixel-level)의 학습 데이터가 대량으로 필요하다.

- [0005] 한편, 제로샷 시맨틱 기술은 학습 과정에서 학습된 클래스뿐만 아니라 학습되지 않은 클래스에 대해서도 추가 정보를 이용하여 시맨틱 분할을 수행하는 기법이다.
- [0006] 한편, 제로샷 시맨틱 분할은 학습 시에 학습한 클래스뿐만 아니라 학습되지 않은 클래스에 대해서도 시맨틱 분할을 할 수 있는 기법을 의미한다. 기존의 제로샷 시맨틱 분할은 생성적 방법(Generative method)을 이용하여 시맨틱 분할을 수행하였다. 생성적 방법은 다단의 스테이지를 통해 시맨틱 분할을 수행하는 기법으로서 학습되지 않은 클래스에 대해서는 특징자를 별도로 생성하여 제로샷 시맨틱 분할을 수행하는 기법이다. 생성적 방법은 최종 스테이지에서 학습되지 않은 클래스에 대한 특징자를 생성한 후 이를 분류기에 입력하여 시맨틱 분할을 수행하도록 한다.
- [0007] 이러한, 생성적 방법은 의미 특징을 고려하지 않고 시맨틱 분할을 수행하기에 편향 문제를 발생시키며, 편향 문제는 학습되지 않은 클래스를 학습된 클래스로 분류하는 문제를 의미한다.
- [0008] 아울러, 기존의 생성적 방법은 새로운 클래스가 새롭게 등장하거나 없어지는 경우 분류기를 매번 새롭게 학습시켜야 하는 문제가 있어 현실적으로 사용하기 어려운 문제점이 있었다.

**발명의 내용**

**해결하려는 과제**

- [0010] 본 발명은 학습되지 않은 클래스에 대해 판별적 방식으로 시맨틱 분할을 수행하여 지속적인 분류기 학습을 요구하지 않은 제로샷 시맨틱 분할 장치 및 방법을 제안한다.
- [0011] 또한, 본 발명은 학습되지 않은 클래스를 기 학습된 클래스로 분류하는 편향 문제를 저감시킬 수 있는 제로샷 시맨틱 분할 장치 및 방법을 제안한다.

**과제의 해결 수단**

- [0013] 본 발명의 일 측면에 따르면, 입력 이미지를 입력받아 신경망 연산을 통해 비주얼 특징맵을 출력하는 비주얼 인코더; 클래스별 특징 벡터를 입력받아 신경망 연산을 통해 클래스별 프로토타입 벡터를 출력하는 시맨틱 인코더; 상기 클래스별 프로토타입 벡터와 상기 비주얼 특징맵의 픽셀별 채널 벡터를 비교하여 상기 비주얼 특징맵의 픽셀 각각에 대해 클래스를 지정하는 시맨틱 분할부를 포함하되, 상기 시맨틱 분할부는 특정 픽셀의 채널 벡터와 가장 유사한 프로토타입 벡터에 상응하는 클래스를 해당 픽셀의 클래스로 지정하며, 상기 프로토타입 벡터와 상기 채널 벡터는 동일한 길이로 설정되고, 상기 비주얼 인코더와 시맨틱 인코더는 적어도 하나의 동일한 손실을 공유하여 동시에 학습되는 시맨틱 분할 장치가 제공된다.
- [0014] 상기 비주얼 인코더와 시맨틱 인코더가 공유하는 손실은 프로토타입 손실을 포함하며, 상기 프로토타입 손실은 상기 시맨틱 인코더에서 출력되는 특정 클래스의 프로토타입 벡터와 상기 비주얼 특징맵의 해당 클래스의 채널 벡터들의 중간값 사이의 손실에 상응한다.
- [0015] 상기 프로토타입 손실에 기초하여 상기 비주얼 인코더에서 출력되는 특정 클래스의 채널 벡터들의 중간값은 상기 시맨틱 인코더에서 출력하는 해당 클래스의 프로토타입 벡터와 동일해지는 방향으로 학습된다.
- [0016] 상기 비주얼 인코더와 시맨틱 인코더가 공유하는 손실은 크로스 엔트로피 손실을 포함하며, 상기 크로스 엔트로피 손실에 의해 상기 비주얼 인코더는 동일한 클래스의 채널 벡터들은 임베딩 공간상에서 상대적으로 가까이 위치하고 다른 클래스의 채널 벡터들은 임베딩 공간상에서 상대적으로 멀리 위치하도록 학습된다.
- [0017] 상기 시맨틱 인코더는 상기 시맨틱 인코더로 입력되는 클래스별 특징 벡터들 사이의 거리와 상기 시맨틱 인코더에서 출력되는 클래스별 프로토타입 벡터들 사이의 거리가 동일해지도록 시맨틱 손실을 이용하여 학습이 이루어진다.
- [0018] 상기 프로토타입 손실을 연산하기 위해 상기 시맨틱 인코더로는 상기 입력 이미지에 클래스별 특징 벡터를 적용하여 생성되는 제1 시맨틱 분할맵을 입력한다.

[0019] 상기 제1 시맨틱 분할맵을 이용한 프로토타입 손실은 다음의 수학적식과 같이 연산된다.

$$\mathcal{L}_{\text{center}} = \frac{1}{\sum_{c \in \mathcal{S}} |\mathcal{R}_c|} \sum_{c \in \mathcal{S}} \sum_{\mathbf{p} \in \mathcal{R}_c} d(v(\mathbf{p}), \mu(\mathbf{p}))$$

[0020]

[0021] 위 수학적식에서,  $\mathcal{L}_{\text{center}}$ 는 프로토타입 손실,  $c$ 는 클래스,  $\mathcal{S}$ 는 클래스 총 집합,  $\mathbf{p}$ 는 픽셀,  $\mathcal{R}_c$ 는 특정 클래스의 픽셀 집합,  $v(\mathbf{p})$ 는 비주얼 특징맵에서 픽셀 위치  $\mathbf{p}$ 에서의 채널 벡터,  $\mu(\mathbf{p})$ 는 제1 시맨틱 분할맵에서 픽셀 위치  $\mathbf{p}$ 의 특징 벡터를 입력하여 시맨틱 인코더에서 출력되는 프로토타입 벡터이고,  $d()$ 는 두 변수의 거리를 출력하는 함수임.

[0022] 상기 프로토타입 손실을 연산하기 위해 상기 시맨틱 인코더로는 상기 입력 이미지의 클래스별로 특징 벡터를 적용하고, 상기 클래스별 특징 벡터가 적용된 이미지를 축소하고, 상기 축소된 이미지를 선형 보간을 통해 본래의 이미지 크기로 확대한 제2 시맨틱 분할맵이 입력된다.

[0023] 상기 제2 시맨틱 분할맵을 이용한 프로토타입 손실은 다음의 수학적식과 같이 연산된다.

$$\mathcal{L}_{\text{bar}} = \frac{1}{\sum_{c \in \mathcal{S}} |\mathcal{R}_c|} \sum_{c \in \mathcal{S}} \sum_{\mathbf{p} \in \mathcal{R}_c} d(v(\mathbf{p}), \tilde{\mu}(\mathbf{p}))$$

[0024]

[0025] 위 수학적식에서,  $\mathcal{L}_{\text{bar}}$ 는 프로토타입 손실,  $c$ 는 클래스,  $\mathcal{S}$ 는 클래스 총 집합,  $\mathbf{p}$ 는 픽셀,  $\mathcal{R}_c$ 는 특정 클래스의 픽셀 집합,  $v(\mathbf{p})$ 는 비주얼 특징맵에서 픽셀 위치  $\mathbf{p}$ 에서의 채널 벡터,  $\tilde{\mu}(\mathbf{p})$ 는 제2 시맨틱 분할맵에서 픽셀 위치  $\mathbf{p}$ 의 특징 벡터를 입력하여 시맨틱 인코더에서 출력되는 프로토타입 벡터이고,  $d()$ 는 두 변수의 거리를 출력하는 함수임.

[0026] 본 발명의 다른 측면에 따르면, 입력 이미지를 입력받아 신경망 연산을 통해 비주얼 특징맵을 비주얼 인코더를 통해 출력하는 단계(a); 클래스별 특징 벡터를 입력받아 신경망 연산을 통해 클래스별 프로토타입 벡터를 시맨틱 인코더를 통해 출력하는 단계(b); 상기 클래스별 프로토타입 벡터와 상기 비주얼 특징맵의 픽셀별 채널 벡터를 비교하여 상기 비주얼 특징맵의 픽셀 각각에 대해 클래스를 지정하는 단계(c)를 포함하되, 상기 단계 (c)는 특정 픽셀의 채널 벡터와 가장 유사한 프로토타입 벡터에 대응하는 클래스를 해당 픽셀의 클래스로 지정하며, 상기 프로토타입 벡터와 상기 채널 벡터는 동일한 길이로 설정되고, 상기 비주얼 인코더와 시맨틱 인코더는 적어도 하나의 동일한 손실을 공유하여 동시에 학습되는 시맨틱 분할 방법이 제공된다.

### 발명의 효과

[0028] 따라서, 본 발명에 의하면, 학습되지 않은 클래스에 대해 관별적 방식으로 시맨틱 분할을 수행하여 지속적인 분류기 학습을 요구하지 않으며, 학습되지 않은 클래스를 기 학습된 클래스로 분류하는 편향 문제를 저감시킬 수 있는 장점이 있다.

### 도면의 간단한 설명

- [0030] 도 1은 본 발명의 일 실시예에 따른 제로샷 시맨틱 분할 장치의 전체적인 구조를 나타낸 블록도.  
 도 2는 본 발명의 일 실시예에 따라 비주얼 인코더에서 출력되는 채널 벡터들의 임베딩 공간의 일례를 나타낸 도면.  
 도 3은 본 발명의 일 실시예에 따라 학습되지 않은 클래스에 대해 시맨틱 분할을 수행하는 원리를 나타낸 도면.  
 도 4는 본 발명의 일 실시예에 따른 비주얼 인코더 및 시맨틱 인코더의 학습 구조를 나타낸 도면.  
 도 5는 본 발명의 일 실시예에 따른 시맨틱 분할 장치의 학습을 위해 생성되는 제1 시맨틱 분할맵을 생성하는 원리를 나타낸 도면.  
 도 6은 본 발명의 일 실시예에 따른 시맨틱 분할 장치의 학습을 위해 생성되는 제2 시맨틱 맵을 생성하는 원리를 나타낸 도면.

도 7은 본 발명의 일 실시예에 따른 프로토타입 손실을 개념적으로 설명하기 위한 도면.

도 8은 본 발명의 일 실시예에 따른 시맨틱 손실을 설명하기 위한 도면.

도 9는 본 발명의 일 실시예에 따른 제로샷 시맨틱 분할 장치의 학습 방법을 나타낸 순서도.

**발명을 실시하기 위한 구체적인 내용**

- [0031] 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.
- [0032] 이하, 첨부한 도면을 참조하여 본 발명의 바람직한 실시예를 설명함으로써, 본 발명을 상세히 설명한다. 그러나, 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며, 설명하는 실시예에 한정되는 것이 아니다. 그리고, 본 발명을 명확하게 설명하기 위하여 설명과 관계없는 부분은 생략되며, 도면의 동일한 참조부호는 동일한 부재를 나타낸다.
- [0033] 명세서 전체에서, 어떤 부분이 어떤 구성요소를 “포함” 한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라, 다른 구성요소를 더 포함할 수 있는 것을 의미한다. 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.
- [0034] 도 1은 본 발명의 일 실시예에 따른 제로샷 시맨틱 분할 장치의 전체적인 구조를 나타낸 블록도이다.
- [0035] 도 1을 참조하면, 본 발명의 일 실시예에 따른 제로샷 시맨틱 분할 장치는 비주얼 인코더(100), 시맨틱 인코더(200) 및 시맨틱 분할부(300)를 포함한다.
- [0036] 비주얼 인코더(100)로는 시맨틱 분할을 위한 이미지(50)가 입력된다. 비주얼 인코더(100)는 신경망 연산을 통해 입력 이미지(50)에 대한 비주얼 특징맵(150)을 출력한다. 일례로, 비주얼 인코더(100)는 CNN 등과 같은 신경망 연산을 통해 비주얼 특징맵을 출력할 수 있다. 입력 이미지(50)가  $H_1 \times W_1$ 의 사이즈의 이미지일 경우, 비주얼 인코더(100)를 통해 출력되는 비주얼 특징맵(150)은  $H_2 \times W_2 \times C$ 의 사이즈를 가질 수 있다. 여기서, H는 높이를 의미하고 W는 폭을 의미한다. 입력 이미지의 사이즈와 비주얼 특징맵의 사이즈(높이 및 폭)은 동일할 수도 있으며 다르게 설정될 수도 있을 것이다. C는 비주얼 특징맵의 각 픽셀이 가지는 채널 벡터의 길이를 의미한다. 비주얼 특징맵(150)은 각 픽셀별로 채널 벡터를 가지게 되며, 비주얼 특징맵(150)을 출력하는 비주얼 인코더(100)는 입력 이미지(50)에 포함된 객체 클래스별로 유사한 채널 벡터를 출력하도록 학습된다.
- [0037] 도 2는 본 발명의 일 실시예에 따라 비주얼 인코더에서 출력되는 채널 벡터들의 임베딩 공간의 일례를 나타낸 도면이다.
- [0038] 도 2에서, 파란색 포인트들은 자동차를 나타내는 픽셀들의 채널 벡터들이며, 녹색의 포인트들은 자전거를 나타내는 픽셀들의 채널 벡터들이다. 도 2에 도시된 바와 같이, 자동차를 나타내는 픽셀들은 임베딩 공간에서 서로 인접하여 위치하도록 비주얼 인코더(100)가 학습되고, 자전거를 나타내는 픽셀들은 임베딩 공간에서 서로 인접하여 위치하도록 비주얼 인코더(100)가 학습된다.
- [0039] 한편, 시맨틱 인코더(200)로는 분할 대상 클래스의 특징 벡터가 입력된다. 예를 들어, 분할 대상 클래스가 자동차 및 자전거일 경우 시맨틱 인코더(200)로는 자동차의 특징 벡터 및 자전거의 특징 벡터가 입력된다. 분할 대상 클래스의 특징 벡터를 입력받은 시맨틱 인코더(200)는 신경망 연산을 통해 각 클래스의 특징 벡터의 상응하는 프로토타입 벡터(250)를 출력한다. 예를 들어, 자동차에 대한 특징 벡터 및 자전거에 대한 특징 벡터가 입력될 경우, 시맨틱 인코더(200)는 자전거에 대한 프로토타입 벡터 및 자동차에 대한 프로토타입 벡터를 각각 출력한다. 본 발명의 일 실시예에 따르면, 시맨틱 인코더(200)는 FC(Fully Connected) 신경망일 수 있으나 이에 한정되는 것은 아니다.
- [0040] 여기서, 시맨틱 인코더(200)로 입력되는 특징 벡터는 상용의 데이터베이스로부터 획득할 수 있는 벡터이다. 예를 들어, 위키피디아와 같은 데이터베이스는 각 클래스별로 특징 벡터를 제공하고 있으며, 이와 같이 상업적으로 획득 가능한 특징 벡터들을 시맨틱 인코더(200)에 입력하는 것이다.
- [0041] 본 발명의 일 실시예에 따르면, 시맨틱 인코더(200)로부터 출력되는 프로토타입 벡터의 길이와 비주얼 인코더(100)를 통해 출력되는 픽셀별 채널 벡터의 길이는 동일하게 설정된다.

- [0042] 시맨틱 분할부(300)는 시맨틱 인코더(200)에서 출력되는 클래스별 프로토타입 벡터와 비주얼 인코더(100)에서 출력되는 비주얼 특징맵을 이용하여 시맨틱 분할을 수행한다. 시맨틱 분할부(300)는 비주얼 특징맵의 각 채널 벡터와 클래스별 프로토타입 벡터를 비교하여 비주얼 특징맵의 픽셀별로 어느 하나의 클래스를 지정하여 시맨틱 분할을 수행한다. 구체적으로 시맨틱 분할부(300)는 특정 픽셀의 채널 벡터와 시맨틱 인코더(200)로부터 출력된 다수의 클래스별 프로토타입 벡터와의 유사도를 연산한다. 시맨틱 분할부(300)는 해당 픽셀의 채널 벡터와 가장 높은 유사도를 가지는 특정 클래스의 프로토타입 벡터를 판단하고, 해당 프로토타입 벡터의 클래스를 해당 픽셀의 클래스를 지정한다. 이와 같은 픽셀별 클래스 지정 작업이 모든 픽셀에 대해 이루어지면 시맨틱 인코더(200)로 입력된 클래스에 대한 시맨틱 분할이 이루어진다. 예를 들어, 시맨틱 인코더(200)로 입력된 클래스가, 자전거, 자동차 및 백그라운드라고 할 경우, 각 픽셀별로 자전거, 자동차 및 백그라운드 중 어느 하나가 지정되는 것이다.
- [0043] 결국, 시맨틱 분할부(300)는 각 픽셀의 채널 벡터와 가장 유사한 클래스별 프로토타입 벡터를 탐색하고, 탐색된 프로토타입 벡터의 클래스를 해당 픽셀의 클래스로 지정하는 작업을 통해 시맨틱 분할을 수행하는 것이다.
- [0044] 도 1을 참조하여 설명한 본 발명의 시맨틱 분할 장치는 비주얼 인코더(100) 및 시맨틱 인코더(200)의 학습이 완료되고 시맨틱 분할 장치를 추론 단계에서 이용할 경우의 동작을 설명한 것이다.
- [0045] 도 1에 도시된 시맨틱 분할 장치는 미리 학습되지 않은 클래스에 대해서도 시맨틱 분할을 가능하도록 한다.
- [0046] 비주얼 인코더(100) 및 시맨틱 인코더(200)에서 미리 학습된 클래스가 자전거, 자동차 및 백그라운드라고 가정한다. 그런데, 신경망 사용자가 사람에 대해서도 시맨틱 분할을 수행할 필요가 있다고 판단할 경우, 신경망 사용자는 사람에 대한 특징 벡터를 시맨틱 인코더(200)에 입력하여 미리 학습되지 않은 클래스에 대해서도 제로샷 시맨틱 분할이 가능하다.
- [0047] 이 경우, 신경망 사용자는 자전거, 자동차 및 사람에 대한 특징 벡터를 시맨틱 인코더(200)에 함께 입력한다. 시맨틱 인코더(200)는 미리 학습된 방식에 따라 자전거에 대한 프로토타입 벡터, 자동차에 대한 프로토타입 벡터 및 사람에 대한 프로토타입 벡터를 출력한다.
- [0048] 도 3은 본 발명의 일 실시예에 따라 학습되지 않은 클래스에 대해 시맨틱 분할을 수행하는 원리를 나타낸 도면이다.
- [0049] 도 3을 참조하면, 비주얼 인코더(100)로는 입력 이미지(50)가 입력된다. 또한, 시맨틱 인코더(200)로는 미리 학습된 클래스인 자동차(car) 및 자전거(bicycle)에 대한 특징벡터와 미리 학습되지 않은 클래스인 사람(person)에 대한 특징 벡터가 입력된다.
- [0050] 시맨틱 분할부(300)는 시맨틱 인코더(200)가 출력하는 자전거에 대한 프로토타입 벡터, 자동차에 대한 프로토타입 벡터 및 사람에 대한 프로토타입 벡터를 이용하여 시맨틱 분할을 수행한다.
- [0051] 도 3에 도시된 바와 같이, 임베딩 공간에서 사람에 대한 프로토타입 벡터와 유사한 채널 벡터들의 픽셀에 대해서는 사람 클래스로 분류하고, 자동차에 대한 프로토타입 벡터와 유사한 채널 벡터들의 픽셀에 대해서는 자동차 클래스로 분류하며, 자전거에 대한 프로토타입 벡터와 유사한 채널 벡터들의 픽셀에 대해서는 자전거 클래스로 분류한다.
- [0052] 도 4는 본 발명의 일 실시예에 따른 비주얼 인코더 및 시맨틱 인코더의 학습 구조를 나타낸 도면이다.
- [0053] 앞서 설명한 바와 같이 비주얼 인코더(100) 및 시맨틱 인코더(200)는 인공 신경망이며, 비주얼 인코더(100) 및 시맨틱 인코더(200)의 신경망 가중치는 학습을 통해 설정된다.
- [0054] 본 발명의 학습 구조에서 주요한 특징 중 하나는 비주얼 인코더(100) 및 시맨틱 인코더(200)가 동일한 손실(loss)을 공유하면서 학습된다는 것이며, 이러한 학습 구조를 통해 시맨틱 인코더(200)에서 출력되는 프로토타입 벡터를 이용하여 비주얼 인코더에서 출력하는 특징맵에 대한 시맨틱 분할을 수행하는 것이 가능해지도록 한다.
- [0055] 도 4를 참조하면, 본 발명의 일 실시예에 따른 비주얼 인코더(100) 및 시맨틱 인코더(200)는 프로토타입 손실 및 크로스엔트로피 손실을 공유하면서 함께 학습이 이루어진다.
- [0056] 프로토타입 손실은 시맨틱 인코더(200)에서 출력하는 특정 클래스의 프로토타입 벡터와 비주얼 인코더(100)에서 출력하는 특정 클래스의 채널 벡터들의 중간값 사이의 손실을 의미한다. 즉, 프로토타입 손실에 의해 비주얼 인코더(100)에서 출력하는 특정 클래스의 채널 벡터들의 중간값과 시맨틱 인코더(200)에서 출력하는 특정 클래스

의 프로토타입 벡터가 동일해지는 방향으로 비주얼 인코더(100) 및 시맨틱 인코더(200)가 학습되는 것이다.

[0057] 프로토타입 손실을 통해, 비주얼 인코더(100)의 출력인 비주얼 특징맵에서 자동차를 나타내는 채널 벡터들의 중간값과 시맨틱 인코더(200)에서 출력하는 자동차에 대한 프로토타입 벡터가 동일해지도록 비주얼 인코더(100) 및 시맨틱 인코더(200)가 학습되는 것이다.

[0058] 크로스 엔트로피 손실은 동일한 클래스의 채널 벡터 및 프로토타입 벡터는 서로 가까워지도록하고 상이한 클래스의 채널 벡터 및 프로토타입 벡터는 서로 멀어지도록 학습하기 위한 손실이다.

[0059] 클래스의 총 집합을 S라고하고, p를 픽셀 좌표, R을 비주얼 특징맵, c는 클래스, R<sub>c</sub>는 비주얼 특징맵에서 특정 클래스를 나타내는 픽셀들의 채널 벡터 집합이라고 정의할 때, 크로스 엔트로피 손실은 다음의 수학적 식 1과 같이 연산될 수 있다.

**수학적 식 1**

$$\mathcal{L}_{ce} = - \frac{1}{\sum_{c \in S} |R_c|} \sum_{c \in S} \sum_{p \in R_c} \log \frac{e^{\omega_c \cdot v(p)}}{\sum_{j \in S} e^{\omega_j \cdot v(p)}}$$

[0060] 위 수학적 식 1에서, ω<sub>c</sub>는 특정 클래스 c의 가중치 벡터이고 ω<sub>j</sub>는 모든 클래스의 가중치 벡터이며, v(p)는 픽셀 위치 p에서의 채널 벡터를 의미한다. 가중치 벡터들은 학습을 통해 설정될 수 있으며, 동일한 클래스의 채널 벡터들은 해당 클래스의 가중치 벡터와 큰 내적값을 가지도록 학습이 이루어지고, 다른 클래스의 가중치 벡터와는 작은 내적값을 가지도록 학습되는 것을 통해 서로 다른 클래스의 채널 벡터 및 프로토타입 벡터는 임베딩 공간에서 멀어지고 같은 클래스의 채널 벡터 및 프로토타입 벡터는 가까워지도록 학습되는 것이 가능하다.

[0061] 한편, 본 발명의 바람직한 실시예에 따르면, 프로토타입 손실을 학습하기 위해 별도의 시맨틱 분할맵을 생성하여 학습을 수행할 수 있다. 도 1을 참조하여 살펴본 바와 같이, 추론 단계에서는 시맨틱 인코더(200)로 상업적으로 입수 가능한 클래스의 특징 벡터가 입력된다. 그러나, 학습 단계에서는 클래스의 특징 벡터 자체가 아닌 시맨틱 분할맵을 이용하여 학습을 수행하는 것이 바람직하며, 시맨틱 분할맵이 시맨틱 인코더(200)로 입력되어 프로토타입 손실 및 크로스 엔트로피 손실에 의한 비주얼 인코더(100) 및 시맨틱 인코더(200)의 학습이 이루어진다.

[0062] 도 5는 본 발명의 일 실시예에 따른 시맨틱 분할 장치의 학습을 위해 생성되는 제1 시맨틱 분할맵을 생성하는 원리를 나타낸 도면이다.

[0063] 도 5의 (a)는 입력 이미지를 나타내고, 도 5의 (b)는 입력 이미지로부터 제1 시맨틱 분할맵을 생성하는 원리를 나타낸 도면이다.

[0064] 도 5의 (a)에 도시된 바와 같이, 입력 이미지는 테이블과 사람이 촬영된 이미지이다. 도 5(a)의 입력 이미지는 3개의 클래스로 구분되며, 테이블, 사람 및 백그라운드이다.

[0065] 도 5의 (b)를 참조하면, 1차적으로 입력 이미지에서 각 클래스 영역을 구분한다. 입력 이미지는 클래스에 대한 정답을 이미 알고 있는 이미지이므로 이에 대한 클래스 영역 구분이 가능하다.

[0066] 입력 이미지에서 클래스 영역 구분이 이루어지면, 각 클래스 영역에 준비된 특징 벡터를 적용한다. 즉, 테이블에 대한 특징 벡터, 사람에 대한 특징 벡터 및 백그라운드에 대한 특징 벡터를 적용하는 것이다. 각 특징 벡터는 상업적으로 이용 가능한 데이터베이스로부터 획득한다.

[0067] 이와 같이 각 클래스 영역의 클래스에 상응하는 특징 벡터를 입력 이미지에 적용하여 제1 시맨틱 분할맵(500)이 생성되며, 제1 시맨틱 분할맵(500)이 시맨틱 인코더에 입력되는 것이다.

[0068] 시맨틱 인코더(200)는 제1 시맨틱 분할맵(500)을 입력받아, 사람에 대한 프로토타입 벡터, 테이블에 대한 프로토타입 벡터 및 백그라운드에 대한 프로토타입 벡터를 출력한다.

[0069] 제1 시맨틱 분할맵(500)이 이용될 경우의 제1 프로토타입 손실(L<sub>center</sub>)은 다음의 수학적 식 2와 같이 연산될 수 있다.

수학식 2

$$\mathcal{L}_{\text{center}} = \frac{1}{\sum_{c \in \mathcal{S}} |\mathcal{R}_c|} \sum_{c \in \mathcal{S}} \sum_{\mathbf{p} \in \mathcal{R}_c} d(v(\mathbf{p}), \mu(\mathbf{p}))$$

[0071]

[0072] 위 수학식 2에서,  $c$ 는 클래스,  $\mathcal{S}$ 는 클래스 총 집합,  $\mathbf{p}$ 는 픽셀,  $\mathcal{R}_c$ 는 특정 클래스의 픽셀 집합,  $v(\mathbf{p})$ 는 비주얼 특징맵에서 픽셀 위치  $\mathbf{p}$ 에서의 채널 벡터,  $\mu(\mathbf{p})$ 는 제1 시맨틱 분할맵에서 픽셀 위치  $\mathbf{p}$ 에서의 프로토타입 벡터를 나타낸다. 또한,  $d(a,b)$ 는  $a$ 와  $b$  사이의 거리를 나타내는 함수로서, 위 수학식2는  $v(\mathbf{p})$ 의 총합과  $\mu(\mathbf{p})$ 의 총합이 동일해지도록 학습되는 것이다.

[0073] 결국, 수학식 2와 같은 제1 프로토타입 손실을 통해 특정 클래스의 채널 벡터의 중간값이 해당 클래스의 프로토타입 벡터와 동일해지도록 학습이 이루어지는 것이다.

[0074] 이와 같은 비주얼 인코더(100)와 채널 인코더(200)의 동시 학습을 통해 채널 인코더(200)의 프로토타입 벡터를 기준으로 비주얼 인코더(100)에서 출력되는 특징맵에 대한 시맨틱 분할이 가능한 것이다.

[0075] 그런데, 제1 시맨틱 분할맵을 이용하여 학습을 수행할 때 편향 문제가 발생할 수 있다. 비주얼 인코더(100)를 통해 출력되는 비주얼 특징맵은 콘볼루션 뉴럴 네트워크(CNN)이므로 픽셀들의 값이 연속적이다. 그런데, 시맨틱 인코더(200)에서 출력되는 값은 도 5의 (b)와 같이 이산적인(discrete)값을 가지게 된다. 이러한 차이는 학습을 방해하고 편향 문제를 야기할 수 있게 된다.

[0076] 이에, 본 발명에서는 제2 시맨틱 분할맵을 이용하여 다른 형태의 프로토타입 손실(제2 프로토타입 손실)을 이용하여 학습하는 구조 역시 제안한다.

[0077] 도 6은 본 발명의 일 실시예에 따른 시맨틱 분할 장치의 학습을 위해 생성되는 제2 시맨틱 맵을 생성하는 원리를 나타낸 도면이다.

[0078] 도 6에서, 입력 이미지는 도 5와 동일하며 입력 이미지를 클래스에 따라 구분하는 작업은 동일하게 이루어진다.

[0079] 입력 이미지에서 클래스별 영역이 구분되면, 입력 이미지를 축소한 축소 이미지(600)를 생성한다. 축소 이미지가 생성되면, 축소 이미지에 각 클래스별 특징 벡터를 적용한다. 제1 시맨틱 분할맵은 축소되지 않은 이미지에 클래스별 특징 벡터를 적용하나 제2 시맨틱 분할맵을 생성할 때에는 축소 이미지(600)에 클래스별 특징 벡터를 적용하는 것이다.

[0080] 축소 이미지(600)에 클래스별 특징 벡터를 적용한 후에는 클래스별 특징 벡터가 적용된 축소 이미지를 선형 보간을 통해 확대하여 제2 시맨틱 분할맵(610)을 생성한다. 선형 보간을 통해 이미지가 확대되면서 경계 영역에서의 특징 벡터는 연속적인 특징을 가질 수 있게 된다.

[0081] 이에 따라 시맨틱 인코더에서 출력하는 프로토타입 벡터들도 경계 영역에서는 연속적인 값을 가질 수 있다. 결국, 제2 시맨틱 분할맵(610)을 이용하여 학습을 수행할 때 비주얼 인코더의 특징맵과 같이 경계 영역에서 연속적인 특징을 가지므로 학습 시 발생하는 편향 문제를 해결할 수 있다.

[0082] 다음의 수학식 3은 제1 시맨틱 분할맵(610)을 이용하여 제2 프로토타입 손실( $\mathcal{L}_{\text{bar}}$ )을 계산하는 방법을 나타낸 것이다.

수학식 3

$$\mathcal{L}_{\text{bar}} = \frac{1}{\sum_{c \in \mathcal{S}} |\mathcal{R}_c|} \sum_{c \in \mathcal{S}} \sum_{\mathbf{p} \in \mathcal{R}_c} d(v(\mathbf{p}), \tilde{\mu}(\mathbf{p}))$$

[0083]

[0084] 제2 프로토타입 손실은 수학적 2의 제1 프로토타입 손실과 비교할 때  $\mu(p)$  대신  $\tilde{\mu}(p)$ 가 사용된다는 점에서 차이가 있다.  $\tilde{\mu}(p)$ 는 제2 시맨틱 분할맵을 시맨틱 인코더에 입력하여 출력되는 프로토타입 벡터들을 의미한다.

[0085] 도 7은 본 발명의 일 실시예에 따른 프로토타입 손실을 개념적으로 설명하기 위한 도면이다.

[0086] 도 7을 참조하면, 시맨틱 인코더(200)로는 자동차(car)에 대한 특징 벡터와 자전거(bicycle)에 대한 특징 벡터가 입력되며, 시맨틱 인코더(200)는 자동차에 대한 프로토타입 벡터(700) 및 자전거에 대한 프로토타입 벡터(710)를 출력한다.

[0087] 한편, 비주얼 인코더(100)로는 입력 이미지가 입력되며, 비주얼 인코더는 각 픽셀별로 클래스에 대한 채널 벡터를 출력한다. 클래스에 대한 채널 벡터는 임베딩 공간(720)상에 나타낼 수 있다. 이때, 자동차(car)를 나타내는 채널 벡터들과 자전거(bicycle)를 나타내는 채널 벡터들은 각각 임베딩 공간(720)상에 투영되는데, 자동차를 나타내는 채널 벡터들의 중간값은 자동차에 대한 프로토타입 벡터와 동일해지도록 프로토타입 손실이 연산되는 것이다. 자전거에 대해서도 동일한 방식으로 프로토타입 손실이 연산된다.

[0088] 결국, 도 7에 도시된 임베딩 공간(720)은 시맨틱 인코더(200)의 프로토타입 벡터와 비주얼 인코더(100)의 채널 벡터가 함께 투영되는 공통 임베딩 공간(Joint Embedding Space)로 표현할 수 있다.

[0089] 한편, 다시 도 4를 참조하면, 시맨틱 인코더(200)에 대해서는 시맨틱 손실을 이용하여 추가적인 학습이 이루어진다. 시맨틱 손실은 입력되는 각 클래스의 특징 벡터들간의 거리와 시맨틱 인코더에서 출력되는 각 클래스별 특징 벡터들간의 거리의 차에 대한 손실이다. 시맨틱 인코더(200)는 시맨틱 손실을 이용하여 입력되는 클래스별 특징 벡터들간의 거리의 차이가 출력되는 클래스별 프로토타입 벡터들간의 거리차가 동일해지도록 학습을 수행한다.

[0090] 예를 들어, 개와 고양이는 서로 다른 클래스이나 형상에서 유사성이 있으므로 특징 벡터들간의 거리 차가 상대적으로 작을 것이다. 그러나, 개와 사람은 형상의 유사성이 낮으므로 특징 벡터들간의 거리차가 상대적으로 클 것이다. 이러한 특징 벡터들간의 거리 차가 프로토타입 벡터들에도 반영되도록 시맨틱 손실을 이용한 학습이 이루어지는 것이다.

[0091] 도 8은 본 발명의 일 실시예에 따른 시맨틱 손실을 설명하기 위한 도면이다.

[0092] 도 8을 참조하면, 특징 벡터의 임베딩 공간(800)과 프로토타입 벡터의 임베딩 공간(810)이 도시되어 있다. 클래스로는 테이블, 고양이, 개 및 사람이 표시되어 있다.

[0093] 상용의 데이터베이스로부터 획득되는 각 클래스의 특징 벡터들은 특징 벡터의 임베딩 공간(800)상에 투영되고, 각 특징 벡터들이 시맨틱 인코더(200)에 입력되어 출력되는 프로토타입 벡터들은 프로토타입 벡터의 임베딩 공간(810)에 투영된다.

[0094] 특징 벡터의 임베딩 공간(800)에서의 두 클래스간 거리와 프로토타입 벡터의 임베딩 공간(810)에서의 두 클래스간 거리의 차를 시맨틱 손실로 정의할 수 있으며, 구체적으로 시맨틱 손실은 다음의 수학적 4와 같이 정의될 수 있다.

수학적 4

$$\mathcal{L}_{sc} = - \sum_{i \in S} \sum_{j \in S} r_{ij} \log \frac{\hat{r}_{ij}}{r_{ij}}.$$

[0095]

[0096] 위 수학적 4에서,  $i$ 와  $j$ 는 클래스를 의미하고,  $S$ 는 클래스 집합을 의미하며,  $r_{ij}$ 는 특징 벡터의 임베딩 공간에서  $i$  클래스와  $j$  클래스간 거리를 의미하고,  $\hat{r}_{ij}$ 는 프로토타입 벡터의 임베딩 공간에서  $i$  클래스와  $j$  클래스간 거리를 의미한다.

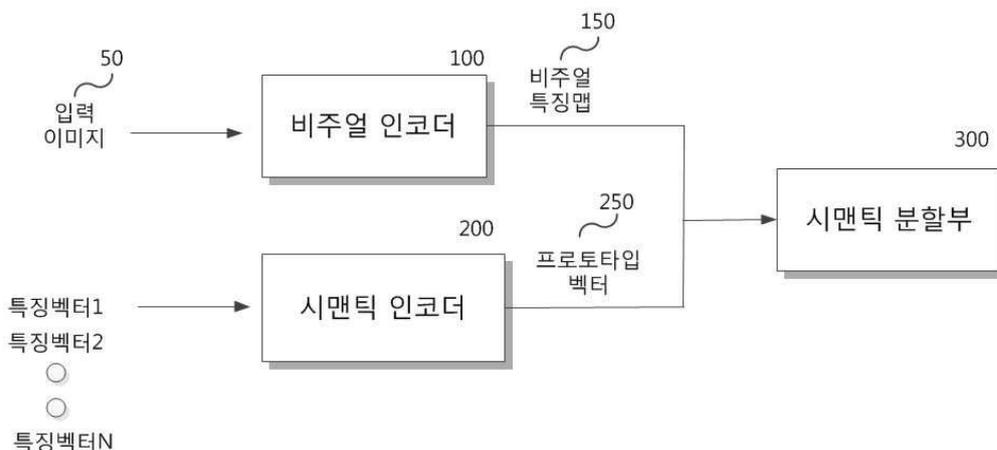
[0097] 도 9는 본 발명의 일 실시예에 따른 제로샷 시맨틱 분할 장치의 학습 방법을 나타낸 순서도이다. 도 9에 도시된

학습 방법은 제2 시맨틱 분할맵을 이용하여 학습을 수행하는 경우를 나타낸 순서도이다.

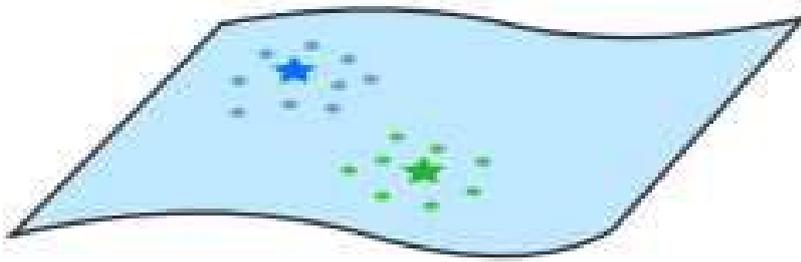
- [0098] 도 9를 참조하면, 입력 이미지를 비주얼 인코더(100)에 입력하여 비주얼 특징맵을 생성한다(단계 900).
- [0099] 입력 이미지에 대해 클래스별 구분을 수행한 후 입력 이미지를 축소시킨다(단계 902).
- [0100] 축소된 입력 이미지에 대해 각 클래스에 상응하는 특징 벡터를 적용한다(단계 904).
- [0101] 특징 벡터가 적용된 축소된 이미지를 선형 보간을 통해 확대시켜 원래의 사이즈로 복원시켜 제2 시맨틱 분할맵을 생성한다(단계 906).
- [0102] 제2 시맨틱 분할맵을 시맨틱 인코더(200)에 입력하여 클래스별 프로토타입 벡터를 출력한다(단계 908).
- [0103] 비주얼 특징맵의 클래스별 채널 벡터들과 시맨틱 인코더(200)에서 출력되는 클래스별 프로토타입 벡터를 이용하여 제2 프로토타입 손실을 연산한다(단계 910). 제2 프로토타입 손실은 수학식 3과 같이 연산될 수 있다.
- [0104] 한편, 비주얼 특징맵에서 출력되는 채널 벡터들을 이용하여 크로스 엔트로피 손실을 연산한다(단계 912). 크로스 엔트로피 손실은 수학식 1과 같이 연산될 수 있다. 앞서 설명한 바와 같이, 크로스 엔트로피 손실은 같은 클래스의 채널 벡터들은 임베딩 공간에서 가까워지고 다른 클래스의 채널 벡터들은 임베딩 공간에서 멀어지도록 학습하기 위해 사용된다.
- [0105] 시맨틱 인코더에 입력되는 클래스별 특징 벡터들간 거리 및 시맨틱 인코더에서 출력되는 클래스별 프로토타입 벡터들간 거리를 이용하여 시맨틱 손실을 연산한다(단계 914). 시맨틱 손실은 수학식 4와 같이 연산될 수 있다.
- [0106] 제2 프로토타입 손실, 크로스 엔트로피 손실을 이용하여 비주얼 인코더(100) 및 시맨틱 인코더(200)를 동시에 학습시킨다(단계 916). 제2 프로토타입 손실과 크로스 엔트로피 손실을 이용하여 비주얼 인코더(100)와 시맨틱 인코더(200)의 가중치를 동시에 갱신하는 것이다. 한편, 시맨틱 인코더(200)는 시맨틱 손실도 함께 반영하여 가중치 갱신이 이루어진다.
- [0107] 본 발명에 따른 방법은 컴퓨터에서 실행시키기 위한 매체에 저장된 컴퓨터 프로그램으로 구현될 수 있다. 여기서 컴퓨터 판독가능 매체는 컴퓨터에 의해 액세스 될 수 있는 임의의 가용 매체일 수 있고, 또한 컴퓨터 저장 매체를 모두 포함할 수 있다. 컴퓨터 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현된 휘발성 및 비휘발성, 분리형 및 비분리형 매체를 모두 포함하며, ROM(판독 전용 메모리), RAM(랜덤 액세스 메모리), CD(컴팩트 디스크)-ROM, DVD(디지털 비디오 디스크)-ROM, 자기 테이프, 플로피 디스크, 광데이터 저장장치 등을 포함할 수 있다.
- [0108] 본 발명은 도면에 도시된 실시예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 것이다.
- [0109] 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 청구범위의 기술적 사상에 의해 정해져야 할 것이다.

**도면**

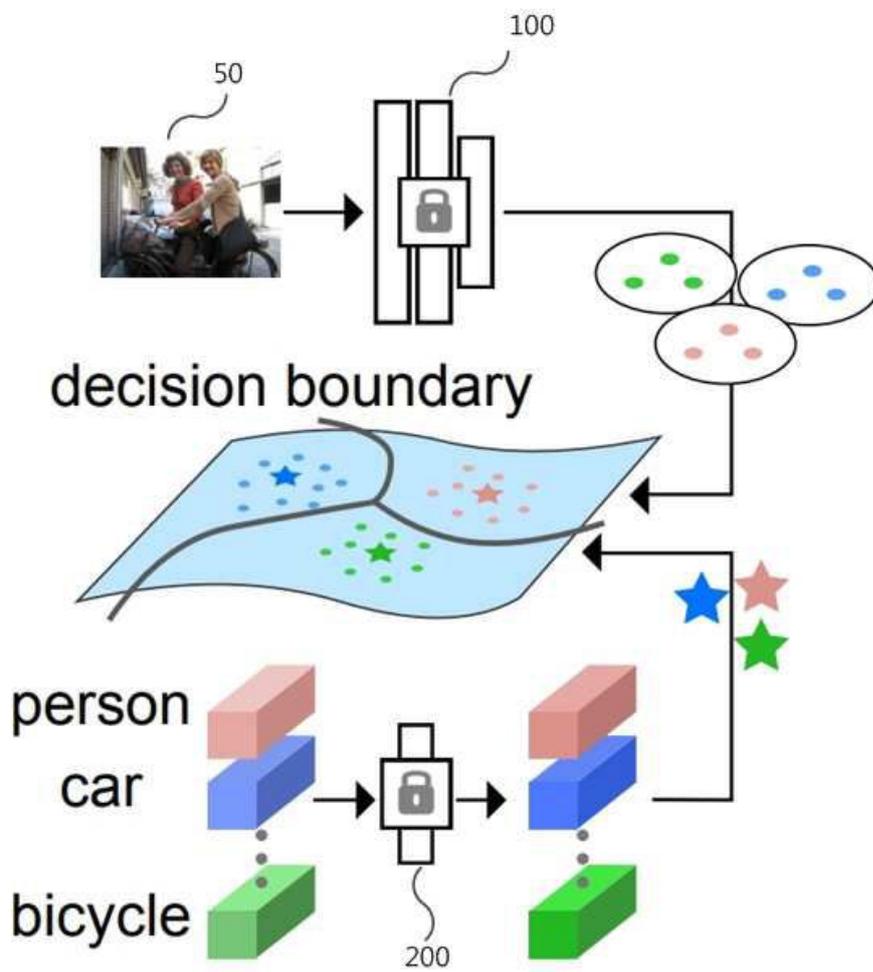
**도면1**



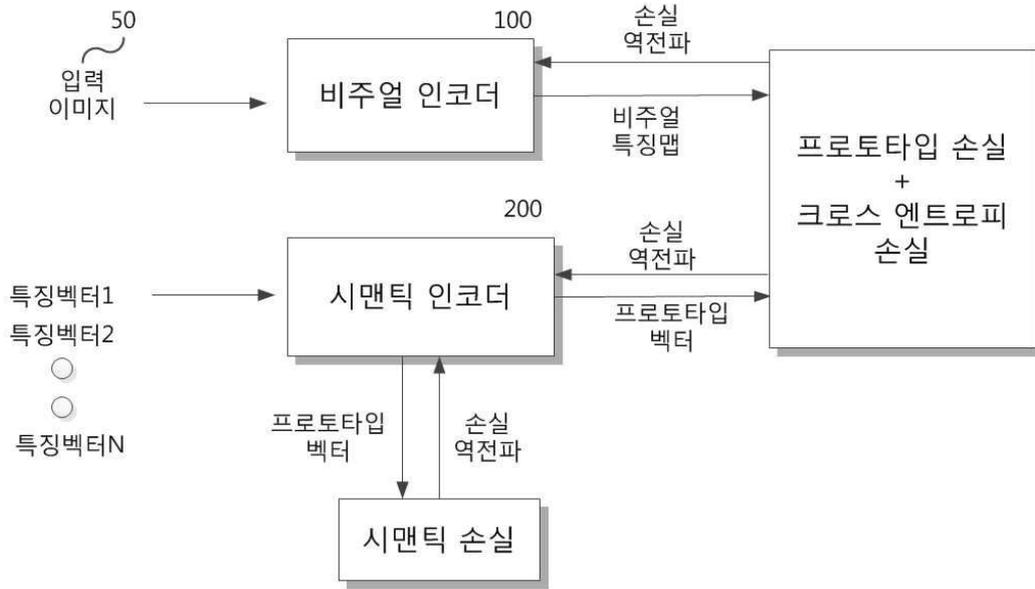
도면2



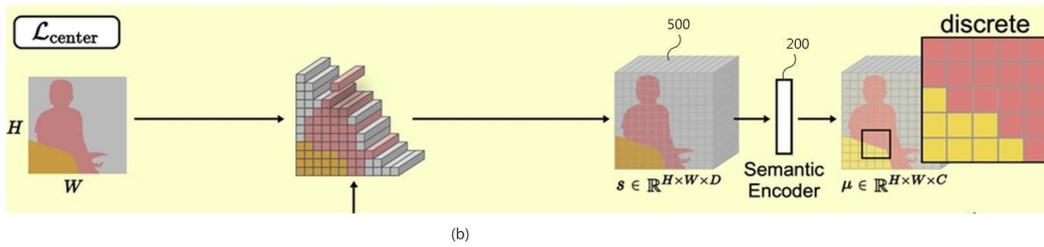
도면3



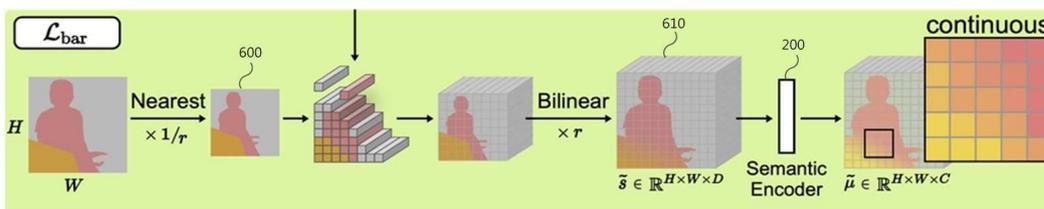
도면4



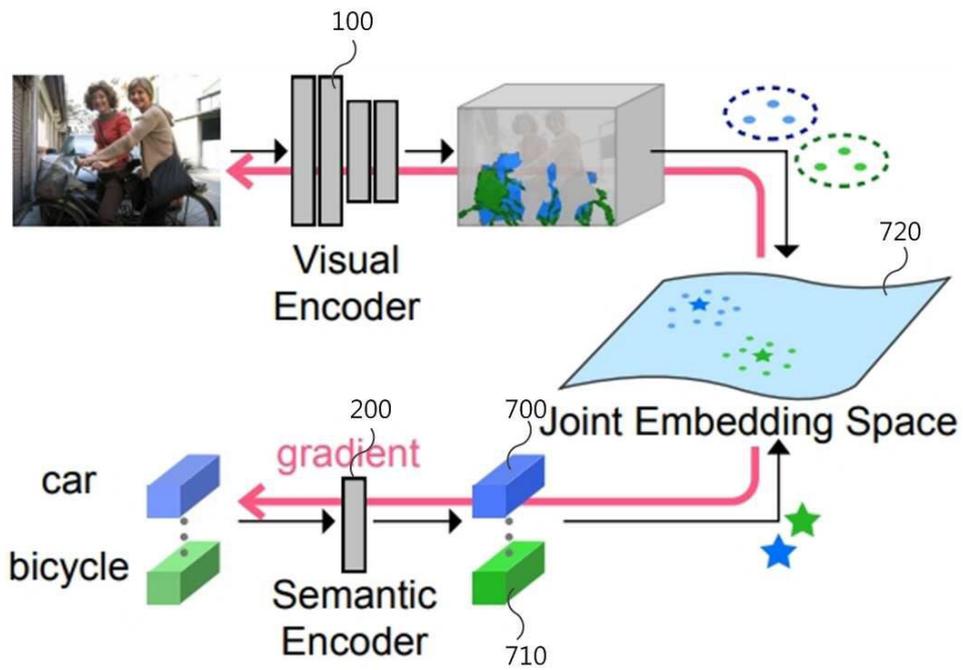
도면5



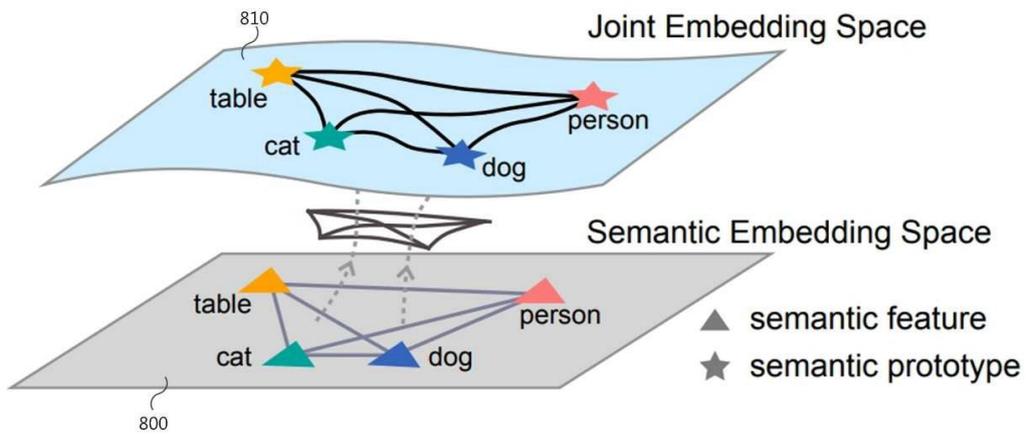
도면6



도면7



도면8



도면9

