



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2021년12월06일
(11) 등록번호 10-2335702
(24) 등록일자 2021년12월01일

(51) 국제특허분류(Int. Cl.)
G06K 9/00 (2006.01) G06K 9/48 (2006.01)
G06N 20/00 (2019.01) G06N 3/08 (2006.01)
(52) CPC특허분류
G06K 9/00718 (2013.01)
G06K 9/481 (2013.01)
(21) 출원번호 10-2020-0176963
(22) 출원일자 2020년12월17일
심사청구일자 2020년12월17일
(30) 우선권주장
1020200158446 2020년11월24일 대한민국(KR)
(56) 선행기술조사문헌
Y. Li et al., 'Bidirectional Learning for
Domain Adaptation of Semantic Segmentation,'
IEEE/CVF Conference on Computer Vision and
Pattern Recognition (CVPR), 2019,
pp.6936-6945 (2019.06.16.) 1부.*
KR1020200112646 A
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대
학교)
(72) 발명자
손광훈
서울특별시 서대문구 연세로 50, 연세대학교 제3
공학관 C129호(신촌동)
김민수
서울특별시 서대문구 연세로 50, 연세대학교 제3
공학관 C129호(신촌동)
(74) 대리인
민영준

전체 청구항 수 : 총 18 항

심사관 : 황승희

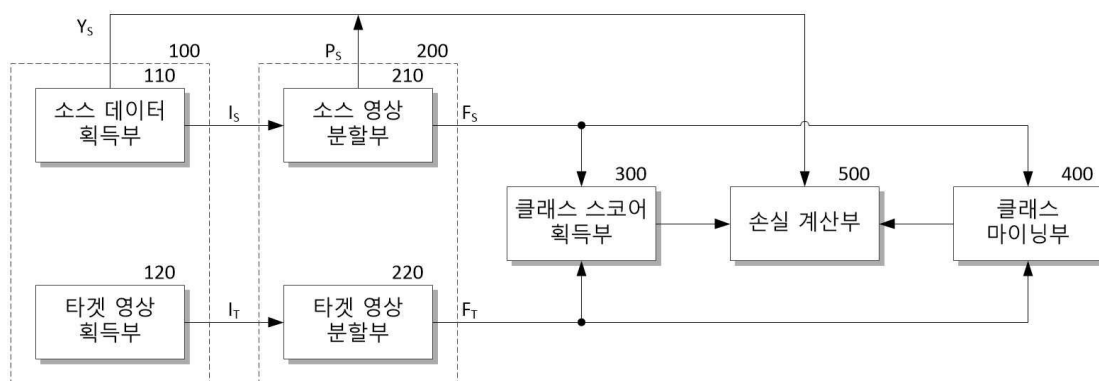
(54) 발명의 명칭 도메인 적응형 의미론적 영상 분할 장치 및 방법

(57) 요약

본 발명은 미리 학습되는 인공 신경망으로 구성되어, 입력 영상을 인가받아 학습된 방식에 따라 인코딩 및 디코딩하여, 입력 영상의 각 픽셀이 기지정된 다수의 클래스에 대응하는 확률을 나타내는 클래스 확률맵을 획득하고, 클래스 확률맵으로부터 입력 영상을 의미론적으로 영역 분할한 의미론적 분할 영상을 출력하며, 학습 시에 소스

(뒷면에 계속)

대표도 - 도1



영상과 입력 영상에 대응하는 타겟 영상 각각을 학습되는 방식에 따라 인코딩 및 디코딩하여, 소스 영상과 타겟 영상의 각 픽셀이 기지정된 다수의 클래스에 대응하는 확률을 나타내는 소스 클래스 확률맵과 타겟 클래스 확률맵을 획득하고, 소스 클래스 확률맵과 타겟 클래스 확률맵 각각에서 유사 픽셀들을 그룹화하여 구분된 다수의 소스 클러스터맵과 다수의 타겟 소스 클러스터맵을 획득하며, 소스 클래스 확률맵으로부터 소스 영상을 의미론적으로 영역 분할하여 소스 분할 영상을 획득하는 의미론적 영상 분할부를 포함하여, 다른 도메인의 학습 데이터를 이용하여 생성적 적대 학습 기법으로 교차 학습될 수 있어, 대량의 학습 데이터를 저비용으로 용이하게 획득할 수 있을 뿐만 아니라 정확하게 의미론적 영상 분할을 수행할 수 있는 의미론적 영상 분할 장치 및 방법을 제공할 수 있다.

(52) CPC특허분류

G06N 20/00 (2021.08)

G06N 3/08 (2013.01)

G06T 7/11 (2017.01)

G06T 2207/20081 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호 1711117053

과제번호 2020-0-00056-001

부처명 과학기술정보통신부

과제관리(전문)기관명 정보통신기획평가원(한국연구재단부설)

연구사업명 정보통신방송연구개발사업

연구과제명 (1단계)(1/2) 현실 세계에서 변화하는 상황에 따라 지속적으로 자가 개선하는 인공지능 기술 개발

기 여 율 1/1

과제수행기관명 연세대학교 산학협력단

연구기간 2020.04.01 ~ 2020.12.31

명세서

청구범위

청구항 1

미리 학습되는 인공 신경망으로 구성되어, 입력 영상을 인가받아 학습된 방식에 따라 인코딩 및 디코딩하여, 상기 입력 영상의 각 픽셀이 기지정된 다수의 클래스에 대응하는 확률을 나타내는 클래스 확률맵을 획득하고, 상기 클래스 확률맵의 각 픽셀에 가장 높은 확률을 갖는 클래스를 할당하고 할당된 클래스에 따라 분할하여, 상기 입력 영상이 의미론적으로 영역 분할된 의미론적 분할 영상을 출력하는 의미론적 영상 분할부를 포함하되,

상기 의미론적 영상 분할부는

인공 신경망을 포함하고, 학습 시에 구비되어 상기 입력 영상과 동일 또는 상이한 도메인에서 획득되는 영상으로 클래스별로 영역이 구분된 진리값이 미리 레이블된 소스 영상을 인가받아 인코딩 및 디코딩하여 상기 소스 클래스 확률맵을 획득하고, 상기 소스 클래스 확률맵의 각 픽셀에 가장 높은 확률을 갖는 클래스를 할당하고 분할하여, 상기 소스 영상이 의미론적으로 영역 분할된 소스 분할 영상을 획득하며, 상기 소스 클래스 확률맵의 다수의 픽셀 사이의 채널 방향 유사도를 기준으로 유사 픽셀들을 그룹화하여 상기 소스 영상의 각 클래스 포함 여부 및 다수의 클래스 각각에 대응하는 수준을 판별하기 위한 다수의 소스 클러스터맵을 획득하는 소스 영상 분할부; 및

상기 소스 영상 분할부와 동일한 구조 및 가중치를 갖도록 학습된 인공 신경망으로 구현되어, 상기 입력 영상을 인가받아 인코딩 및 디코딩하여 상기 클래스 확률맵을 획득하고, 상기 클래스 확률맵으로부터 상기 의미론적 분할 영상을 획득하고, 학습 시에는 상기 입력 영상과 동일한 도메인에서 획득되는 영상으로 진리값이 레이블되지 않은 타겟 영상을 인가받아 인코딩 및 디코딩하여 상기 타겟 클래스 확률맵을 획득하고, 상기 타겟 클래스 확률맵의 각 픽셀에 가장 높은 확률을 갖는 클래스를 할당하고 분할하여 상기 타겟 영상이 의미론적으로 영역 분할된 타겟 분할 영상을 획득하며, 상기 타겟 클래스 확률맵의 다수의 픽셀 사이의 채널 방향 유사도를 기준으로 유사 픽셀들을 그룹화하여 상기 타겟 영상의 각 클래스 포함 여부 및 다수의 클래스 각각에 대응하는 수준을 판별하기 위한 다수의 타겟 클러스터맵을 획득하는 타겟 영상 분할부를 포함하며,

상기 소스 영상 분할부는

상기 소스 영상을 인가받아 학습되는 방식에 따라 인코딩하여 소스 특징맵을 출력하는 소스 인코더;

학습되는 방식에 따라 상기 소스 특징맵을 디코딩하여 상기 소스 클래스 확률맵을 획득하는 소스 확률 추정부;

상기 소스 클래스 확률맵의 각 픽셀에 가장 높은 확률을 갖는 클래스를 할당하여, 상기 소스 분할 영상을 획득하는 소스 분할 출력부;

학습되는 방식에 따라 상기 소스 클래스 확률맵의 다수의 픽셀 사이의 채널 방향 유사도를 기준으로 유사한 픽셀들을 클래스 개수보다 작은 기지정된 K 개의 그룹으로 그룹화하여, K 개의 소스 그룹 특징맵을 획득하는 소스 그룹화부; 및

상기 K 개의 소스 그룹 특징맵 각각과 상기 소스 클래스 확률맵을 원소곱하여 K 개의 소스 클러스터맵을 획득하는 소스 클러스터맵 획득부를 포함하며,

상기 소스 클러스터맵과 상기 타겟 클러스터맵을 이용하여 계산되는 적대적 손실을 통해 상기 인공 신경망을 학습시키는 의미론적 영상 분할 장치.

청구항 2

삭제

청구항 3

삭제

청구항 4

제1항에 있어서, 상기 타겟 영상 분할부는

상기 입력 영상을 인가받아 미리 학습된 방식에 따라 인코딩하여 특징맵을 출력하고, 학습 시에 상기 타겟 영상을 학습되는 방식에 따라 인코딩하여 타겟 특징맵을 출력하는 타겟 인코더;

미리 학습된 방식에 따라 상기 특징맵을 디코딩하여 상기 클래스 확률맵을 획득하고, 학습시에 학습되는 방식에 따라 상기 타겟 특징맵을 디코딩하여 상기 타겟 클래스 확률맵을 획득하는 타겟 확률 추정부;

상기 클래스 확률맵의 각 픽셀에 가장 높은 확률을 갖는 클래스를 할당하여, 상기 의미론적 분할 영상을 획득하는 타겟 분할 출력부;

학습 시에 구비되어 학습되는 방식에 따라 상기 타겟 클래스 확률맵의 다수의 픽셀 사이의 채널 방향 유사도를 기준으로 유사한 픽셀들을 K개의 그룹으로 그룹화하여, K개의 타겟 그룹 특징맵을 획득하는 타겟 그룹화부; 및

상기 K개의 타겟 그룹 특징맵 각각과 상기 타겟 클래스 확률맵을 원소곱하여 K개의 타겟 클러스터맵을 획득하는 타겟 클러스터맵 획득부를 포함하는 의미론적 영상 분할 장치.

청구항 5

제4항에 있어서, 상기 의미론적 영상 분할 장치는

학습 시에 의미론적 영상 분할부를 학습시키기 위한 학습부를 더 포함하고,

상기 학습부는

상기 K개의 소스 클러스터맵과 상기 K개의 타겟 클러스터맵 각각의 구분된 영역에서 각 클래스의 분포 비율을 분석하여, K개의 소스 클래스 스코어 벡터와 K개의 타겟 클래스 스코어 벡터를 획득하는 클래스 스코어 획득부;

상기 K개의 소스 클러스터맵과 상기 K개의 타겟 클러스터맵 각각의 구분된 영역에서 각 클래스의 포함 여부를 분석하여, K개의 소스 클래스 마이닝 벡터와 K개의 타겟 클래스 마이닝 벡터를 획득하는 클래스 마이닝부; 및

상기 소스 영상과 함께 미리 획득되는 상기 소스 영상에 대한 진리값과 상기 소스 분할 영상, 상기 K개의 소스 클러스터맵과 상기 K개의 타겟 클러스터맵, 상기 K개의 소스 클래스 스코어 벡터와 상기 K개의 타겟 클래스 스코어 벡터 및 상기 K개의 소스 마이닝 벡터와 상기 K개의 타겟 마이닝 벡터를 기반으로 기설정된 방식으로 총 손실을 계산하여 역전파하는 손실 계산부를 포함하는 의미론적 영상 분할 장치.

청구항 6

제5항에 있어서, 상기 클래스 스코어 획득부는

상기 K개의 소스 클러스터맵 각각에서 클래스별 구분을 나타내는 채널 축을 제외한 나머지 2축으로 구성되는 평면들 각각에 대해 평균값 풀링을 수행하여, 상기 K개의 소스 클래스 스코어 벡터를 획득하는 소스 클래스 스코어 획득부; 및

상기 K개의 타겟 클러스터맵 각각에서 클래스별 구분을 나타내는 채널 축을 제외한 나머지 2축으로 구성되는 평면들 각각에 대해 평균값 풀링을 수행하여, 상기 K개의 타겟 클래스 스코어 벡터를 획득하는 타겟 클래스 스코어 획득부를 포함하는 의미론적 영상 분할 장치.

청구항 7

제6항에 있어서, 상기 클래스 마이닝부는

상기 K개의 소스 클러스터맵 각각에서 클래스별 구분을 나타내는 채널 축을 제외한 나머지 2축으로 구성되는 평면들 각각에 대해 최대값 풀링을 수행하여, 상기 K개의 소스 클래스 마이닝 벡터를 획득하는 소스 클래스 마이닝부; 및

상기 K개의 타겟 클러스터맵 각각에서 클래스별 구분을 나타내는 채널 축을 제외한 나머지 2축으로 구성되는 평면들 각각에 대해 최대값 풀링을 수행하여, 상기 K개의 타겟 클래스 마이닝 벡터를 획득하는 타겟 클래스 마이닝부를 포함하는 의미론적 영상 분할 장치.

청구항 8

제7항에 있어서, 상기 손실 계산부는

상기 소스 영상에 대한 진리값과 상기 소스 분할 영상 사이의 크로스 엔트로피를 기지정된 방식으로 계산하여 분할 손실을 획득하는 분할 손실 계산부;

상기 K개의 소스 클래스 스코어 벡터와 상기 K개의 타겟 클래스 스코어 벡터에서 서로 대응하는 클래스 스코어 벡터 사이의 차이를 누적 계산하여 일관성 손실을 계산하는 일관성 손실 계산부;

상기 K개의 소스 클래스 스코어 벡터 각각 사이와 상기 K개의 타겟 클래스 스코어 벡터 각각 사이의 클래스 분포가 나타내는 직교성을 계산하고, 계산된 직교성 기반하여 기지정된 방식으로 직교성 손실을 계산하는 직교성 손실 계산부;

인공 신경망으로 구현되어 상기 K개의 소스 클래스 스코어 벡터와 상기 K개의 소스 클러스터맵 중 서로 대응하는 소스 클래스 스코어 벡터와 소스 클러스터맵을 외적하고, 상기 K개의 타겟 클래스 스코어 벡터와 상기 K개의 타겟 클러스터맵 중 서로 대응하는 타겟 클래스 스코어 벡터와 타겟 클러스터맵을 외적한 결과가 소스 영상에서 획득된 값인지 판별하고, 판별 결과에 따라 적대적 도메인 손실을 계산하는 적대적 도메인 판별부;

상기 K개의 소스 클래스 마이닝 벡터와 상기 K개의 타겟 클래스 마이닝 벡터 중 서로 대응하는 마이닝 벡터에서 서로 대응하는 클래스의 존재 여부를 반영하여 클래스 균형 도메인 적응 손실을 계산하는 클래스 손실 계산부; 및

분할 손실과 일관성 손실, 직교성 손실, 적대적 도메인 손실 및 클래스 균형 도메인 적응 손실을 기지정된 방식으로 합산하여 총 손실을 계산하고, 계산된 총 손실을 역전파하는 총 손실 계산부를 포함하는 의미론적 영상 분할 장치.

청구항 9

제8항에 있어서, 상기 일관성 손실 계산부는

상기 일관성 손실(L_{co})을 상기 K개의 소스 클래스 스코어 벡터($Q_S = [Q_S^1, \dots, Q_S^K]$)와 상기 K개의 타겟 클래스 스코어 벡터($Q_T = [Q_T^1, \dots, Q_T^K]$)로부터 수학적식

$$\mathcal{L}_{co}(G, C) = \sum_{k \in \{1, \dots, K\}} \|Q_S^k - Q_T^k\|^2.$$

에 따라 계산하고,

직교성 손실 계산부는

상기 K개의 소스 클래스 스코어 벡터(Q_S) 각각 사이와 상기 K개의 타겟 클래스 스코어 벡터(Q_T) 각각 사이의 직교 성분을 수학적식

$$\cos(Q_l^{j_1}, Q_l^{j_2}) = \frac{Q_l^{j_1} \cdot Q_l^{j_2}}{\|Q_l^{j_1}\|_2 \|Q_l^{j_2}\|_2}, \quad j_1, j_2 \in \{1, \dots, K\}.$$

(여기서 $l = (S, T)$ 이다.)

에 따라 계산하고, 상기 직교성 손실(L_{orth})을 수학적식

$$\mathcal{L}_{orth}(G, C) = \sum_l \sum_{j_1, j_2} \cos(Q_l^{j_1}, Q_l^{j_2}),$$

에 따라 계산하는 의미론적 영상 분할 장치.

청구항 10

제8항에 있어서, 상기 적대적 도메인 판별부는

상기 K개의 소스 클래스 스코어 벡터(Q_S)와 상기 K개의 소스 클러스터맵($F_S = [F_S^1, \dots, F_S^K]$) 중 서로 대응하는 소스 클래스 스코어 벡터와 소스 클러스터맵을 외적한 결과($F_S^k \otimes Q_S^k$)가 소스 영상에서 획득된 값인지 판별하고, 상기 K개의 타겟 클래스 스코어 벡터(Q_T)와 상기 K개의 타겟 클러스터맵($F_T = [F_T^1, \dots, F_T^K]$) 중 서로 대응하는 타겟 클래스 스코어 벡터와 타겟 클러스터맵을 외적한 결과($F_T^k \otimes Q_T^k$)가 소스 영상에서 획득된 값인지 판별하여 적대적 도메인 손실(\mathcal{L}_{adv})을 수학식

$$\mathcal{L}_{adv}(G, C, D) = - \sum_k [\log(D(F_S^k \otimes Q_S^k))] - \sum_k [\log(1 - D(F_T^k \otimes Q_T^k))],$$

으로 계산하고,

상기 클래스 손실 계산부는

K개의 소스 클래스 마이닝 벡터($M_S = [M_S^1, \dots, M_S^K]$)와 K개의 타겟 클래스 마이닝 벡터($M_T = [M_T^1, \dots, M_T^K]$)를 인가받아 상기 클래스 균형 도메인 적응 손실(\mathcal{L}_{cl})을 수학식

$$\mathcal{L}_{cl}(G, C) = - \sum_k \sum_u [m_{S,u}^k \geq \tau] \log(m_{T,u}^k),$$

(여기서 $m_{S,u}^k$ 와 $m_{T,u}^k$ 는 각각 제k 소스 클래스 마이닝 벡터(M_S^k)와 제k 타겟 클래스 마이닝 벡터(M_T^k)에서 특정 클래스 값을 나타내는 값이고, τ 는 기지정된 문턱값을 나타낸다.)에 따라 계산하는 의미론적 영상 분할 장치.

청구항 11

입력 영상을 인가받아 의미론적 영상 분할 영상을 출력하는 의미론적 영상 분할 장치에서 수행되는 의미론적 영상 분할 방법으로서,

미리 학습되는 인공 신경망을 이용하여 입력 영상을 인가받아 학습된 방식에 따라 인코딩 및 디코딩하여, 상기 입력 영상의 각 픽셀이 기지정된 다수의 클래스에 대응하는 확률을 나타내는 클래스 확률맵을 획득하고, 상기 클래스 확률맵의 각 픽셀에 가장 높은 확률을 갖는 클래스를 할당하고 할당된 클래스에 따라 분할하여, 상기 입력 영상이 의미론적으로 영역 분할된 의미론적 분할 영상을 출력하는 단계; 및

상기 의미론적 분할 영상을 출력하는 단계 이전에 인공 신경망의 학습을 위해 수행되는 학습 단계를 포함하고,

상기 학습 단계는

상기 입력 영상과 동일 또는 상이한 도메인에서 획득되는 영상으로 클래스별로 영역이 구분된 진리값이 미리 레이블된 소스 영상과 상기 입력 영상과 동일한 도메인에서 획득되는 영상으로 진리값이 레이블되지 않은 타겟 영상을 각각 학습되는 방식에 따라 인코딩 및 디코딩하여, 상기 소스 영상과 상기 타겟 영상의 각 픽셀이 기지정된 다수의 클래스에 대응하는 확률을 나타내는 소스 클래스 확률맵과 타겟 클래스 확률맵을 획득하고, 상기 소스 클래스 확률맵과 상기 타겟 클래스 확률맵 각각에서 다수의 픽셀 사이의 채널 방향 유사도를 기준으로 유사 픽셀들을 그룹화하여 상기 소스 영상과 상기 타겟 영상의 각 클래스 포함 여부 및 다수의 클래스 각각에 대응하는 수준을 판별하기 위한 다수의 소스 클러스터맵과 다수의 타겟 클러스터맵을 획득하며, 상기 소스 클래스 확률맵과 상기 타겟 클래스 확률맵 각각의 각 픽셀에 가장 높은 확률을 갖는 클래스를 할당하고 분할하여 상기 소스 영상과 상기 타겟 영상 각각이 의미론적으로 영역 분할된 소스 분할 영상과 타겟 분할 영상을 획득하는

단계; 및

K개의 소스 클러스터맵과 K개의 타겟 클러스터맵 각각의 구분된 영역에서 각 클래스의 분포 비율과 각 클래스의 포함 여부를 분석하여, K개의 소스 클래스 스코어 벡터와 K개의 타겟 클래스 스코어 벡터 및 K개의 소스 클래스 마이닝 벡터와 K개의 타겟 클래스 마이닝 벡터를 획득하며, 상기 소스 영상과 함께 미리 획득되는 상기 소스 영상에 대한 진리값과 상기 소스 분할 영상, 상기 K개의 소스 클러스터맵과 상기 K개의 타겟 클러스터맵, 상기 K개의 소스 클래스 스코어 벡터와 상기 K개의 타겟 클래스 스코어 벡터 및 상기 K개의 소스 마이닝 벡터와 상기 K개의 타겟 마이닝 벡터를 기반으로 기지정된 방식으로 총 손실을 계산하여 역전파하는 단계를 포함하는 의미론적 영상 분할 방법.

청구항 12

제11항에 있어서, 상기 의미론적 분할 영상을 출력하는 단계는

상기 입력 영상을 인가받아 미리 학습된 방식에 따라 인코딩하여 특징맵을 출력하는 단계;

미리 학습된 방식에 따라 상기 특징맵을 디코딩하여 상기 클래스 확률맵을 획득하는 단계; 및

상기 클래스 확률맵의 각 픽셀에 가장 높은 확률을 갖는 클래스를 할당하여, 상기 의미론적 분할 영상을 획득하는 단계를 포함하는 의미론적 영상 분할 방법.

청구항 13

제12항에 있어서, 상기 소스 분할 영상과 타겟 분할 영상을 획득하는 단계는

학습 단계에서 구비되는 인공 신경망을 이용하여 상기 입력 영상과 동일 또는 상이한 도메인에서 획득된 상기 소스 영상을 인가받아 인코딩 및 디코딩하여 상기 소스 클래스 확률맵을 획득하고, 상기 소스 클래스 확률맵으로부터 상기 소스 분할 영상과 상기 다수의 소스 클러스터맵을 획득하는 단계; 및

상기 입력 영상과 동일한 도메인에서 획득되는 상기 타겟 영상을 인가받아 인공 신경망을 이용하여 인코딩 및 디코딩하여 상기 타겟 클래스 확률맵을 획득하고, 상기 타겟 클래스 확률맵으로부터 상기 다수의 타겟 클러스터맵을 획득하는 단계를 포함하는 의미론적 영상 분할 방법.

청구항 14

제13항에 있어서, 상기 다수의 소스 클러스터맵을 획득하는 단계는

상기 소스 영상을 인가받아 학습되는 방식에 따라 인코딩하여 소스 특징맵을 출력하는 단계;

학습되는 방식에 따라 상기 소스 특징맵을 디코딩하여 상기 소스 클래스 확률맵을 획득하는 단계;

상기 소스 클래스 확률맵의 각 픽셀에 가장 높은 확률을 갖는 클래스를 할당하여, 상기 소스 분할 영상을 획득하는 단계;

상기 소스 클래스 확률맵의 다수의 픽셀 사이의 채널 방향 유사도를 기준으로 유사한 픽셀들을 클래스 개수보다 작은 기지정된 K개의 그룹으로 그룹화하여, K개의 소스 그룹 특징맵을 획득하는 단계; 및

상기 K개의 소스 그룹 특징맵 각각과 상기 소스 클래스 확률맵을 원소곱하여 K개의 소스 클러스터맵을 획득하는 단계를 포함하는 의미론적 영상 분할 방법.

청구항 15

제14항에 있어서, 상기 다수의 타겟 클러스터맵을 획득하는 단계는

상기 소스 영상으로부터 상기 소스 클래스 확률맵을 획득하는 동일한 구조 및 가중치를 갖는 인공 신경망을 이용하여, 상기 타겟 영상으로부터 상기 타겟 클래스 확률맵을 획득하는 의미론적 영상 분할 방법.

청구항 16

제15항에 있어서, 상기 다수의 타겟 클러스터맵을 획득하는 단계는

상기 타겟 영상을 학습되는 방식에 따라 인코딩하여 타겟 특징맵을 출력하는 단계;

학습되는 방식에 따라 상기 타겟 특징맵을 디코딩하여 상기 타겟 클래스 확률맵을 획득하는 단계;

상기 타겟 클래스 확률맵의 다수의 픽셀 사이의 채널 방향 유사도를 기준으로 유사한 픽셀들을 K개의 그룹으로 그룹화하여, K개의 타겟 그룹 특징맵을 획득하는 단계; 및

상기 K개의 타겟 그룹 특징맵 각각과 상기 타겟 클래스 확률맵을 원소곱하여 K개의 타겟 클러스터맵을 획득하는 단계를 포함하는 의미론적 영상 분할 방법.

청구항 17

제16항에 있어서, 상기 역전파하는 단계는

상기 K개의 소스 클러스터맵과 상기 K개의 타겟 클러스터맵 각각에서 클래스별 구분을 나타내는 채널 축을 제외한 나머지 2축으로 구성되는 평면들 각각에 대해 평균값 풀링을 수행하여, K개의 소스 클래스 스코어 벡터와 K개의 타겟 클래스 스코어 벡터를 획득하는 단계;

상기 K개의 소스 클러스터맵과 상기 K개의 타겟 클러스터맵 각각에서 클래스별 구분을 나타내는 채널 축을 제외한 나머지 2축으로 구성되는 평면들 각각에 대해 최대값 풀링을 수행하여 K개의 소스 클래스 마이닝 벡터와 K개의 타겟 클래스 마이닝 벡터를 획득하는 단계; 및

상기 소스 영상과 함께 미리 획득되는 상기 소스 영상에 대한 진리값과 상기 소스 분할 영상, 상기 K개의 소스 클러스터맵과 상기 K개의 타겟 클러스터맵, 상기 K개의 소스 클래스 스코어 벡터와 상기 K개의 타겟 클래스 스코어 벡터 및 상기 K개의 소스 마이닝 벡터와 상기 K개의 타겟 마이닝 벡터를 기반으로 기지정된 방식으로 총 손실을 계산하여 계산된 총 손실을 역전파하는 단계를 포함하는 의미론적 영상 분할 방법.

청구항 18

제17항에 있어서, 상기 총 손실을 역전파하는 단계는

상기 소스 영상에 대한 진리값과 상기 소스 분할 영상 사이의 크로스 엔트로피를 기지정된 방식으로 계산하여 분할 손실을 획득하는 단계;

상기 K개의 소스 클래스 스코어 벡터와 상기 K개의 타겟 클래스 스코어 벡터에서 서로 대응하는 클래스 스코어 벡터 사이의 차이를 누적 계산하여 일관성 손실을 계산하는 단계;

상기 K개의 소스 클래스 스코어 벡터 각각 사이와 상기 K개의 타겟 클래스 스코어 벡터 각각 사이의 클래스 분포가 나타내는 직교성을 계산하고, 계산된 직교성 기반하여 기지정된 방식으로 직교성 손실을 계산하는 단계;

학습되는 인공 신경망을 이용하여 상기 K개의 소스 클래스 스코어 벡터와 상기 K개의 소스 클러스터맵 중 서로 대응하는 소스 클래스 스코어 벡터와 소스 클러스터맵을 외적하고, 상기 K개의 타겟 클래스 스코어 벡터와 상기 K개의 타겟 클러스터맵 중 서로 대응하는 타겟 클래스 스코어 벡터와 타겟 클러스터맵을 외적한 결과가 소스 영상에서 획득된 값인지 판별하고, 판별 결과에 따라 적대적 도메인 손실을 계산하는 단계;

상기 K개의 소스 클래스 마이닝 벡터와 상기 K개의 타겟 클래스 마이닝 벡터 중 서로 대응하는 마이닝 벡터에서 서로 대응하는 클래스의 존재 여부를 반영하여 클래스 균형 도메인 적응 손실을 계산하는 단계; 및

분할 손실과 일관성 손실, 직교성 손실, 적대적 도메인 손실 및 클래스 균형 도메인 적응 손실을 기지정된 방식으로 합산하여 총 손실을 계산하여 역전파하는 단계를 포함하는 의미론적 영상 분할 방법.

청구항 19

제18항에 있어서, 상기 일관성 손실을 계산하는 단계는

상기 일관성 손실(L_{co})을 상기 K개의 소스 클래스 스코어 벡터($Q_s = [Q_s^1, \dots, Q_s^K]$)와 상기 K개의 타겟 클래스 스코어 벡터($Q_t = [Q_t^1, \dots, Q_t^K]$)로부터 수학식

$$\mathcal{L}_{co}(G, C) = \sum_{k \in \{1, \dots, K\}} \|Q_S^k - Q_T^k\|^2.$$

에 따라 계산하고,

직교성 손실을 계산하는 단계는

상기 K개의 소스 클래스 스코어 벡터(Q_S) 각각 사이와 상기 K개의 타겟 클래스 스코어 벡터(Q_T) 각각 사이의 직교 성분을 수학식

$$\cos(Q_l^{j_1}, Q_l^{j_2}) = \frac{Q_l^{j_1} \cdot Q_l^{j_2}}{\|Q_l^{j_1}\|_2 \|Q_l^{j_2}\|_2}, \quad j_1, j_2 \in \{1, \dots, K\}.$$

(여기서 $l = (S, T)$ 이다.)

에 따라 계산하고, 상기 직교성 손실(\mathcal{L}_{orth})을 수학식

$$\mathcal{L}_{orth}(G, C) = \sum_l \sum_{j_1, j_2} \cos(Q_l^{j_1}, Q_l^{j_2}),$$

에 따라 계산하는 의미론적 영상 분할 방법.

청구항 20

제19항에 있어서, 상기 적대적 도메인 손실을 계산하는 단계는

상기 K개의 소스 클래스 스코어 벡터(Q_S)와 상기 K개의 소스 클러스터맵($F_S = [F_S^1, \dots, F_S^K]$) 중 서로 대응하는 소스 클래스 스코어 벡터와 소스 클러스터맵을 외적한 결과($F_S^k \otimes Q_S^k$)가 소스 영상에서 획득된 값인지 판별하고, 상기 K개의 타겟 클래스 스코어 벡터(Q_T)와 상기 K개의 타겟 클러스터맵($F_T = [F_T^1, \dots, F_T^K]$) 중 서로 대응하는 타겟 클래스 스코어 벡터와 타겟 클러스터맵을 외적한 결과($F_T^k \otimes Q_T^k$)가 소스 영상에서 획득된 값인지 판별하여 적대적 도메인 손실(\mathcal{L}_{adv})을 수학식

$$\begin{aligned} \mathcal{L}_{adv}(G, C, D) = & - \sum_k [\log(D(F_S^k \otimes Q_S^k))] \\ & - \sum_k [\log(1 - D(F_T^k \otimes Q_T^k))], \end{aligned}$$

으로 계산하고,

상기 클래스 균형 도메인 적응 손실을 계산하는 단계는

K개의 소스 클래스 마이닝 벡터($M_S = [M_S^1, \dots, M_S^K]$)와 K개의 타겟 클래스 마이닝 벡터($M_T = [M_T^1, \dots, M_T^K]$)를 인가받아 상기 클래스 균형 도메인 적응 손실(\mathcal{L}_{cl})을 수학식

$$\mathcal{L}_{cl}(G, C) = - \sum_k \sum_u [m_{S,u}^k \geq \tau] \log(m_{T,u}^k),$$

(여기서 $m_{S,u}^k$ 와 $m_{T,u}^k$ 는 각각 제k 소스 클래스 마이닝 벡터(M_S^k)와 제k 타겟 클래스 마이닝 벡터(M_T^k)에서 특정

클래스 값을 나타내는 값이고, τ 는 기지정된 문턱값을 나타낸다.)에 따라 계산하는 의미론적 영상 분할 방법.

발명의 설명

기술 분야

[0001] 본 발명은 의미론적 영상 분할 장치 및 방법에 관한 것으로, 도메인 적응형 의미론적 영상 분할 장치 및 방법에 관한 것이다.

배경 기술

[0002] 의미론적 영상 분할(semantic segmentation)은 입력 영상을 식별 가능한 클래스에 대응하는 영역별로 분할하는 것을 말한다. 입력 영상에는 각 클래스로 지정될 수 있는 다수의 객체가 포함될 수 있으며, 의미론적 영상 분할은 입력 영상에 포함된 다수의 객체 중 동일 클래스로 식별되는 객체들이 위치하는 영역과 다른 클래스로 식별되는 객체들이 위치하는 영역을 구분한다.

[0003] 최근의 지도학습 방식으로 학습된 합성곱 신경망(CNN)은 의미론적 영상 분할에서 우수한 성능을 나타내고 있다. 다만 지도 학습의 경우, 학습용 영상과 학습용 영상의 각 객체별 클래스가 식별되어, 클래스별로 영상의 영역이 구분된 진리값(ground truth)이 레이블된 대량의 학습 데이터를 필요로 한다. 그러나 지도 학습을 위한 학습 데이터의 진리값은 대부분 수작업에 의해 획득되므로, 대량의 학습데이터를 생성하기 위해서는 많은 시간과 비용 및 노력이 필요하다.

[0004] 이에 최근에는 의미론적 영상 분할이 수행되어야 하는 실제 영상과 상이한 도메인인 가상 환경(i.e. 게임)에서 진리값이 레이블된 학습 데이터를 용이하게 획득하여 지도 학습을 수행하고자 하는 연구가 활발하게 진행되고 있다. 그러나 실제 영상과 다른 도메인인 가상 환경에서 만들어진 영상은 실제 영상과 데이터 분포가 상이하기 때문에, 가상 환경 영상을 통해 학습된 네트워크가 실제 환경에서는 잘 동작하지 않는 문제가 발생한다. 즉 가상 환경이라는 실제 환경과 상이한 도메인에서 생성된 학습 데이터를 이용하여 지도 학습을 수행하는 경우, 실제 환경에서 요구되는 성능에 도달하지 못한다는 한계가 있다.

선행기술문헌

특허문헌

[0005] (특허문헌 0001) 한국 공개 특허 제10-2020-0112646호 (2020.10.05 공개)

발명의 내용

해결하려는 과제

[0006] 본 발명의 목적은 적용 대상 영상이 획득되는 도메인과 상이한 도메인에서 저비용으로 용이하게 획득되는 학습 데이터를 이용하여 학습될 수 있는 의미론적 영상 분할 장치 및 방법을 제공하는데 있다.

[0007] 본 발명의 다른 목적은 다른 도메인의 학습 데이터를 이용하여 학습되는 의미론적 영상 분할 방식을 생성적 적대 학습 기법으로 다른 도메인에 교차 학습되어 정확하게 의미론적 영상 분할을 수행할 수 있는 의미론적 영상 분할 장치 및 방법을 제공하는데 있다.

과제의 해결 수단

[0008] 상기 목적을 달성하기 위한 본 발명의 일 실시예에 따른 의미론적 영상 분할 장치는 미리 학습되는 인공 신경망으로 구성되어, 입력 영상을 인가받아 학습된 방식에 따라 인코딩 및 디코딩하여, 상기 입력 영상의 각 픽셀이 기지정된 다수의 클래스에 대응하는 확률을 나타내는 클래스 확률맵을 획득하고, 상기 클래스 확률맵으로부터 상기 입력 영상을 의미론적으로 영역 분할한 의미론적 분할 영상을 출력하며, 학습 시에 소스 영상과 상기 입력 영상에 대응하는 타겟 영상 각각을 학습되는 방식에 따라 인코딩 및 디코딩하여, 상기 소스 영상과 상기 타겟 영상의 각 픽셀이 기지정된 다수의 클래스에 대응하는 확률을 나타내는 소스 클래스 확률맵과 타겟 클래스 확률맵을 획득하고, 상기 소스 클래스 확률맵과 상기 타겟 클래스 확률맵 각각에서 유사 픽셀들을 그룹화하여 구분된 다수의 소스 클러스터맵과 다수의 타겟 소스 클러스터맵을 획득하며, 상기 소스 클래스 확률맵으로부터 상기

소스 영상을 의미론적으로 영역 분할하여 소스 분할 영상을 획득하는 의미론적 영상 분할부를 포함한다.

- [0009] 상기 의미론적 영상 분할부는 인공 신경망을 포함하고, 학습 시에 구비되어 상기 입력 영상과 동일 또는 상이한 도메인에서 획득된 상기 소스 영상을 인가받아 인코딩 및 디코딩하여 상기 소스 클래스 확률맵을 획득하고, 상기 소스 클래스 확률맵으로부터 상기 소스 분할 영상과 상기 다수의 소스 클러스터맵을 획득하는 소스 영상 분할부; 및 상기 소스 영상 분할부와 동일한 구조 및 가중치를 갖도록 학습된 인공 신경망으로 구현되어, 상기 입력 영상을 인가받아 인코딩 및 디코딩하여 상기 클래스 확률맵을 획득하고, 상기 클래스 확률맵으로부터 상기 의미론적 분할 영상을 획득하고, 학습 시에는 상기 입력 영상과 동일한 도메인에서 획득되는 상기 타겟 영상을 인가받아 인코딩 및 디코딩하여 상기 타겟 클래스 확률맵을 획득하고, 상기 타겟 클래스 확률맵으로부터 상기 다수의 타겟 클러스터맵을 획득하는 타겟 영상 분할부를 포함할 수 있다.
- [0010] 상기 소스 영상 분할부는 상기 소스 영상을 인가받아 학습되는 방식에 따라 인코딩하여 소스 특징맵을 출력하는 소스 인코더; 학습되는 방식에 따라 상기 소스 특징맵을 디코딩하여 상기 소스 클래스 확률맵을 획득하는 소스 확률 추정부; 상기 소스 클래스 확률맵의 각 픽셀에 가장 높은 확률을 갖는 클래스를 할당하여, 상기 소스 분할 영상을 획득하는 소스 분할 출력부; 학습되는 방식에 따라 상기 소스 클래스 확률맵의 다수의 픽셀 사이의 유사도를 추정하여 유사한 픽셀들을 클래스 개수보다 작은 기지정된 K개의 그룹으로 그룹화하여, K개의 소스 그룹 특징맵을 획득하는 소스 그룹화부; 및 상기 K개의 소스 그룹 특징맵 각각과 상기 소스 클래스 확률맵을 원소곱하여 K개의 소스 클러스터맵을 획득하는 소스 클러스터맵 획득부를 포함할 수 있다.
- [0011] 상기 타겟 영상 분할부는 상기 입력 영상을 인가받아 미리 학습된 방식에 따라 인코딩하여 특징맵을 출력하고, 학습 시에 상기 타겟 영상을 학습되는 방식에 따라 인코딩하여 타겟 특징맵을 출력하는 타겟 인코더; 미리 학습된 방식에 따라 상기 특징맵을 디코딩하여 상기 클래스 확률맵을 획득하고, 학습시에 학습되는 방식에 따라 상기 타겟 특징맵을 디코딩하여 상기 타겟 클래스 확률맵을 획득하는 타겟 확률 추정부; 상기 클래스 확률맵의 각 픽셀에 가장 높은 확률을 갖는 클래스를 할당하여, 상기 의미론적 분할 영상을 획득하는 타겟 분할 출력부; 학습 시에 구비되어 학습되는 방식에 따라 상기 타겟 클래스 확률맵의 다수의 픽셀 사이의 유사도를 추정하여 유사한 픽셀들을 K개의 그룹으로 그룹화하여, K개의 타겟 그룹 특징맵을 획득하는 타겟 그룹화부; 및 상기 K개의 타겟 그룹 특징맵 각각과 상기 타겟 클래스 확률맵을 원소곱하여 K개의 타겟 클러스터맵을 획득하는 타겟 클러스터맵 획득부를 포함할 수 있다.
- [0012] 상기 의미론적 영상 분할 장치는 학습 시에 의미론적 영상 분할부를 학습시키기 위한 학습부를 더 포함하고, 상기 학습부는 상기 K개의 소스 클러스터맵과 상기 K개의 타겟 소스 클러스터맵 각각의 구분된 영역에서 각 클래스의 분포 비율을 분석하여, K개의 소스 클래스 스코어 벡터와 K개의 타겟 클래스 스코어 벡터를 획득하는 클래스 스코어 획득부; 상기 K개의 소스 클러스터맵과 상기 K개의 타겟 소스 클러스터맵 각각의 구분된 영역에서 각 클래스의 포함 여부를 분석하여, K개의 소스 클래스 마이닝 벡터와 K개의 타겟 클래스 마이닝 벡터를 획득하는 클래스 마이닝부; 및 상기 소스 영상과 함께 미리 획득되는 상기 소스 영상에 대한 진리값과 상기 소스 분할 영상, 상기 K개의 소스 클러스터맵과 상기 K개의 타겟 클러스터맵, 상기 K개의 소스 클래스 스코어 벡터와 상기 K개의 타겟 클래스 스코어 벡터 및 상기 K개의 소스 마이닝 벡터와 상기 K개의 타겟 마이닝 벡터를 기반으로 기 지정된 방식으로 총 손실을 계산하여 역전파하는 손실 계산부를 포함할 수 있다.
- [0013] 상기 클래스 스코어 획득부는 상기 K개의 소스 클러스터맵 각각에서 클래스별 구분을 나타내는 채널 축을 제외한 나머지 2축으로 구성되는 평면들 각각에 대해 평균값 풀링을 수행하여, 상기 K개의 소스 클래스 스코어 벡터를 획득하는 소스 클래스 스코어 획득부; 및 상기 K개의 타겟 클러스터맵 각각에서 클래스별 구분을 나타내는 채널 축을 제외한 나머지 2축으로 구성되는 평면들 각각에 대해 평균값 풀링을 수행하여, 상기 K개의 타겟 클래스 스코어 벡터를 획득하는 타겟 클래스 스코어 획득부를 포함할 수 있다.
- [0014] 상기 클래스 마이닝부는 상기 K개의 소스 클러스터맵 각각에서 클래스별 구분을 나타내는 채널 축을 제외한 나머지 2축으로 구성되는 평면들 각각에 대해 최대값 풀링을 수행하여, 상기 K개의 소스 클래스 마이닝 벡터를 획득하는 소스 클래스 마이닝부; 및 상기 K개의 타겟 클러스터맵 각각에서 클래스별 구분을 나타내는 채널 축을 제외한 나머지 2축으로 구성되는 평면들 각각에 대해 최대값 풀링을 수행하여, 상기 K개의 타겟 클래스 마이닝 벡터를 획득하는 타겟 클래스 마이닝부를 포함할 수 있다.
- [0015] 상기 손실 계산부는 상기 소스 영상에 대한 진리값과 상기 소스 분할 영상 사이의 크로스 엔트로피를 기 지정된 방식으로 계산하여 분할 손실을 획득하는 분할 손실 계산부; 상기 K개의 소스 클래스 스코어 벡터와 상기 K개의 타겟 클래스 스코어 벡터에서 서로 대응하는 클래스 스코어 벡터 사이의 차이를 누적 계산하여 일관성 손실을 계산하는 일관성 손실 계산부; 상기 K개의 소스 클래스 스코어 벡터 각각 사이와 상기 K개의 타겟 클래스 스코어

어 벡터 각각 사이의 클래스 분포가 나타내는 직교성을 계산하고, 계산된 직교성 기반하여 기지정된 방식으로 직교성 손실을 계산하는 직교성 손실 계산부; 인공 신경망으로 구현되어 상기 K개의 소스 클래스 스코어 벡터와 상기 K개의 소스 클러스터맵 중 서로 대응하는 소스 클래스 스코어 벡터와 소스 클러스터맵을 외적하고, 상기 K개의 타겟 클래스 스코어 벡터와 상기 K개의 타겟 클러스터맵 중 서로 대응하는 타겟 클래스 스코어 벡터와 타겟 클러스터맵을 외적한 결과가 소스 영상에서 획득된 값인지 판별하고, 판별 결과에 따라 적대적 도메인 손실을 계산하는 적대적 도메인 판별부; 상기 K개의 소스 클래스 마이닝 벡터와 상기 K개의 타겟 클래스 마이닝 벡터 중 서로 대응하는 마이닝 벡터에서 서로 대응하는 클래스의 존재 여부를 반영하여 클래스 균형 도메인 적응 손실을 계산하는 클래스 손실 계산부; 및 분할 손실과 일관성 손실, 직교성 손실, 적대적 도메인 손실 및 클래스 균형 도메인 적응 손실을 기지정된 방식으로 합산하여 총 손실을 계산하고, 계산된 총 손실을 역전파하는 총 손실 계산부를 포함할 수 있다.

[0016] 상기 목적을 달성하기 위한 본 발명의 다른 실시예에 따른 의미론적 영상 분할 방법은 미리 학습되는 인공 신경망을 이용하여 입력 영상을 인가받아 학습된 방식에 따라 인코딩 및 디코딩하여, 상기 입력 영상의 각 픽셀이 기지정된 다수의 클래스에 대응하는 확률을 나타내는 클래스 확률맵을 획득하고, 상기 클래스 확률맵으로부터 상기 입력 영상을 의미론적으로 영역 분할한 의미론적 분할 영상을 출력하는 단계; 및 상기 의미론적 분할 영상을 출력하는 단계 이전에 인공 신경망의 학습 시에 수행되는 학습 단계를 포함하고, 상기 학습 단계는 소스 영상과 상기 입력 영상에 대응하는 타겟 영상 각각을 학습되는 방식에 따라 인코딩 및 디코딩하여, 상기 소스 영상과 상기 타겟 영상의 각 픽셀이 기지정된 다수의 클래스에 대응하는 확률을 나타내는 소스 클래스 확률맵과 타겟 클래스 확률맵을 획득하고, 상기 소스 클래스 확률맵과 상기 타겟 클래스 확률맵 각각에서 유사 픽셀들을 그룹화하여 구분된 다수의 소스 클러스터맵과 다수의 타겟 소스 클러스터맵을 획득하며, 상기 소스 클래스 확률맵으로부터 상기 소스 영상을 의미론적으로 영역 분할하여 소스 분할 영상을 획득하는 단계; 및 상기 K개의 소스 클러스터맵과 상기 K개의 타겟 소스 클러스터맵 각각의 구분된 영역에서 각 클래스의 분포 비율과 각 클래스의 포함 여부를 분석하여, K개의 소스 클래스 스코어 벡터와 K개의 타겟 클래스 스코어 벡터 및 K개의 소스 클래스 마이닝 벡터와 K개의 타겟 클래스 마이닝 벡터를 획득하며, 상기 소스 영상과 함께 미리 획득되는 상기 소스 영상에 대한 진리값과 상기 소스 분할 영상, 상기 K개의 소스 클러스터맵과 상기 K개의 타겟 클러스터맵, 상기 K개의 소스 클래스 스코어 벡터와 상기 K개의 타겟 클래스 스코어 벡터 및 상기 K개의 소스 마이닝 벡터와 상기 K개의 타겟 마이닝 벡터를 기반으로 기지정된 방식으로 총 손실을 계산하여 역전파하는 단계를 포함한다.

발명의 효과

[0017] 따라서, 본 발명의 실시예에 따른 의미론적 영상 분할 장치 및 방법은 다른 도메인의 학습 데이터를 이용하여 생성적 적대 학습 기법으로 교차 학습될 수 있어, 대량의 학습 데이터를 저비용으로 용이하게 획득할 수 있을 뿐만 아니라 정확하게 의미론적 영상 분할을 수행할 수 있다.

도면의 간단한 설명

[0018] 도 1은 본 발명의 일 실시예에 따른 의미론적 영상 분할 장치의 개략적 구조를 나타낸다.
 도 2는 도 1의 의미론적 영상 분할부의 상세 구조를 나타낸다.
 도 3은 도 1의 의미론적 영상 분할부의 동작을 설명하기 위한 도면이다.
 도 4는 도 2의 그룹화부에서 획득되는 그룹 특징맵의 일 예를 나타낸다.
 도 5는 도 1의 클래스 스코어 획득부와 클래스 마이닝부의 상세 구조를 나타낸다.
 도 6은 도 5의 클래스 스코어 획득부의 동작을 설명하기 위한 도면이다.
 도 7은 도 5의 클래스 마이닝부의 동작을 설명하기 위한 도면이다.
 도 8은 도 1의 손실 계산부의 상세 구조를 나타낸다.
 도 9는 도 8의 손실 계산부의 각 구성별 동작을 설명하기 위한 도면이다.
 도 10은 본 실시예에 따른 의미론적 영상 분할 장치의 의미론적 영상 분할 결과의 일 예를 나타낸다.
 도 11은 본 발명의 일 실시예에 따른 의미론적 영상 분할 방법을 나타낸다.
 도 12는 도 11의 손실 역전파 단계를 상세하게 나타낸 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0019] 본 발명과 본 발명의 동작상의 이점 및 본 발명의 실시에 의하여 달성되는 목적을 충분히 이해하기 위해서는 본 발명의 바람직한 실시예를 예시하는 첨부 도면 및 첨부 도면에 기재된 내용을 참조하여야만 한다.
- [0020] 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 부재임을 나타낸다.
- [0021] 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.
- [0022] 도 1은 본 발명의 일 실시예에 따른 의미론적 영상 분할 장치의 개략적 구조를 나타내고, 도 2는 도 1의 의미론적 영상 분할부의 상세 구조를 나타내며, 도 3은 도 1의 의미론적 영상 분할부의 동작을 설명하기 위한 도면이다. 그리고 도 4는 도 2의 그룹화부에서 획득되는 그룹 특징맵의 일 예를 나타낸다.
- [0023] 도 1을 참조하면, 본 실시예에 따른 의미론적 영상 분할 장치는 영상 획득부(100), 의미론적 영상 분할부(200), 클래스 스코어 획득부(300), 클래스 마이닝부(400) 및 손실 계산부(500)를 포함한다.
- [0024] 우선 영상 획득부(100)는 소스 데이터 획득부(110)와 타겟 영상 획득부(120)를 포함할 수 있다. 여기서 타겟 영상 획득부(120)는 의미론적으로 분할되어야 하는 타겟 영상(I_T)을 획득한다. 타겟 영상(I_T)은 진리값이 레이블되지 않은 영상으로, 이후 의미론적 영상 분할 장치가 실제로 영상을 의미론적으로 분할하기 위해 이용될 때, 입력되는 입력 영상으로 볼 수 있다. 즉 타겟 영상(I_T)은 실제 이용될 도메인에서 진리값이 포함되지 않은 영상이다.
- [0025] 그리고 소스 데이터 획득부(110)는 타겟 영상(I_T)과 동일하거나 상이한 도메인에서 획득된 소스 영상(I_S)과 소스 영상(I_S)에서 클래스별로 영역이 구분된 의미론적 분할 영상인 진리값(Y_S)이 함께 포함된 데이터이다. 일반적으로 인공 신경망을 포함하여 구현되는 의미론적 영상 분할 장치가 입력되는 영상을 의미론적으로 분할하기 위해서는, 입력 영상과 동일한 도메인의 데이터를 이용하여 학습되어야 한다. 그러나 상기한 바와 같이, 입력 영상과 동일한 도메인의 영상에 대한 진리값을 획득하기는 용이하지 않은 경우가 많다.
- [0026] 이에 본 실시예에서 의미론적 영상 분할 장치는 학습 시에 입력 영상의 도메인에 무관하게 획득된 소스 영상(I_S)과 소스 영상(I_S)에 대해 의미론적 영상 분할된 진리값(Y_S)이 포함된 소스 데이터와 진리값이 포함되지 않지만 입력 영상과 동일한 도메인에서 획득된 타겟 영상(I_T)을 함께 이용하여 학습을 수행할 수 있도록 한다. 따라서 학습 데이터의 부족으로 인한 문제를 해소하면서도 학습 데이터로 이용되는 소스 데이터가 다른 도메인에서 획득되어 발생하는 성능 한계를 극복할 수 있도록 한다. 그러나 경우에 따라서 소스 데이터의 소스 영상(I_S)은 입력 영상과 동일한 도메인에서 획득된 영상이어도 무방하다. 여기서 소스 영상(I_S)이 입력 영상과 동일한 도메인에서 획득된 영상인 경우에도 타겟 영상(I_T)이 함께 이용되는 것은 적은 양의 소스 데이터로도 안정적인 학습이 수행될 수 있도록 학습 데이터의 부족을 보완하기 위해서이다.
- [0027] 도 3에서는 일 예로 소스 영상(I_S)과 타겟 영상(I_T)이 동일한 도메인의 실제 환경에서 획득된 영상인 것으로 도시하였으나, 소스 영상(I_S)은 가상 환경에서 획득된 영상이어도 무방하다.
- [0028] 다만, 소스 영상(I_S)과 타겟 영상(I_T)은 유사한 대상에 대한 영상으로 유사 구조를 갖는 것이 바람직하다. 즉 도 2에서와 같이, 영상 분할 대상이 되는 타겟 영상(I_T)이 도로 영상인 경우, 소스 영상(I_S) 또한 유사하게 도로 영상으로 획득되는 것이 바람직하다.
- [0029] 도 1에서는 영상 획득부(100)가 소스 데이터 획득부(110)와 타겟 영상 획득부(120)를 포함하는 것으로 도시하였으나, 소스 데이터는 학습 시에만 요구되므로, 학습이 완료된 이후에는 타겟 영상 획득부(120)만을 구비하여, 의미론적 영상 분할이 되어야하는 타겟 도메인의 영상을 인가받을 수 있다.

- [0030] 의미론적 영상 분할부(200)는 인공 신경망을 포함하여 구현된다. 인공 신경망을 포함하는 의미론적 영상 분할부(200)는 영상 획득부(100)에서 획득된 소스 영상(I_S)과 타겟 영상(I_T)을 인가받고, 학습되는 방식에 따라 인가된 소스 영상(I_S)과 타겟 영상(I_T)의 각 픽셀의 클래스를 추정하고, 추정된 클래스에 따라 동일 클래스의 픽셀들을 그룹화하여 영역을 구분함으로써, 의미론적으로 영역이 분할된 소스 분할 영상(S_{seg})과 타겟 분할 영상(T_{seg})을 출력한다. 의미론적 영상 분할부(200)는 학습 시, 소스 영상(I_S)과 타겟 영상(I_T)을 인가받아 소스 분할 영상(S_{seg})과 타겟 분할 영상(T_{seg})을 출력한다. 여기서 소스 분할 영상(S_{seg})은 학습을 위해 획득되는 영상이다.
- 즉 의미론적 영상 분할 장치가 실제 이용되는 경우에, 의미론적 영상 분할부(200)는 입력 영상인 타겟 영상(I_T)만을 인가받아 타겟 분할 영상(T_{seg})을 획득하고, 획득된 타겟 분할 영상(T_{seg})을 의미론적 영상 분할 장치의 출력인 의미론적 분할 영상으로서 출력한다.
- [0031] 또한 의미론적 영상 분할부(200)는 학습 시에 소스 분할 영상(S_{seg})과 타겟 분할 영상(T_{seg}) 이외에 소스 영상(I_S)과 타겟 영상(I_T)의 각 픽셀이 기지정된 다수의 클래스 각각에 대응할 확률에 기초하여 획득되는 소스 클러스터맵(F_S)과 타겟 클러스터맵(F_T)을 더 획득한다.
- [0032] 도 2를 참조하면, 의미론적 영상 분할부(200)는 각각 인공 신경망으로 구현되어 소스 영상(I_S)을 인가받아 의미론적으로 분할하여 소스 분할 영상(S_{seg})과 소스 클러스터맵(F_S)을 획득하는 소스 영상 분할부(210)와 타겟 영상(I_T)을 인가받아 의미론적으로 분할하여 타겟 분할 영상(T_{seg})과 타겟 클러스터맵(F_T)을 획득하는 타겟 영상 분할부(220)를 포함할 수 있다.
- [0033] 우선 소스 영상 분할부(210)는 소스 인코더(211), 소스 확률 추정부(212), 소스 분할 출력부(213), 소스 그룹화부(214) 및 소스 클러스터맵 획득부(215)를 포함할 수 있다.
- [0034] 도 3에 도시된 바와 같이, 소스 인코더(211)는 다수의 히든 레이어(hidden layer)를 포함하는 인공 신경망으로 구현되어, 학습되는 방식에 따라 인가되는 소스 영상(I_S)을 인코딩하여 소스 특징맵(f_S)을 획득한다.
- [0035] 그리고 소스 확률 추정부(212) 또한 인공 신경망으로 구현되어, 인가되는 소스 특징맵(f_S)을 학습되는 방식에 따라 디코딩하여 소스 영상(I_S)의 다수의 픽셀 각각이 기지정된 다수의 클래스 각각에 대응할 확률을 추정하여, 소스 클래스 확률맵(P_S)을 획득한다. 여기서는 소스 확률 추정부(212)가 소스 영상(I_S)의 다수의 픽셀 각각에 대해 학습에 의해 식별 가능한 개수(cls, 여기서는 일 예로 cls = 19)) 클래스 각각에 대응하는 확률을 추정하여, 소스 클래스 확률맵(P_S)을 획득하는 것으로 가정한다. 즉 소스 클래스 확률맵(P_S)은 소스 영상(I_S)과 동일한 폭(W)과 높이(H)를 갖고 기지정된 클래스 개수에 대응하는 채널 길이($c = \text{cls}$)를 갖는 3차원 특징맵으로 획득될 수 있다. 또한 다수의 클래스는 의미론적 영상 분할 장치가 학습에 의해 식별 가능한 객체의 종류를 나타내며, 일 예로 다수의 클래스는 각각 차량, 나무, 빌딩, 도로, 자동차 및 자전거 등의 객체를 나타낼 수 있다.
- [0036] 소스 분할 출력부(213)는 소스 클래스 확률맵(P_S)을 인가받아 각 픽셀에서 채널축 방향에서의 채널값(즉 클래스별 확률값)을 분석하여 최대값을 갖는 채널을 해당 픽셀의 클래스로 할당하여 소스 분할 영상(S_{seg})을 획득한다. 즉 다수의 픽셀 각각에 가장 높은 확률로 추정된 클래스를 할당하고, 픽셀별로 할당된 클래스에 따라 영역을 구분하여 소스 영상(I_S)에 대한 2차원 의미론적 분할 영상인 소스 분할 영상(S_{seg})을 출력한다.
- [0037] 여기서는 설명의 편의를 위하여 소스 확률 추정부(212)와 소스 분할 출력부(213)를 별도의 구성으로 도시하였으나, 소스 분할 출력부(213)는 완전 연결 레이어(Fully Connected Layer) 등으로 구현되어 소스 확률 추정부(212)에 결합될 수 있으며, 이 경우, 소스 확률 추정부(212)와 소스 분할 출력부(213)는 소스 디코더로 통합될 수 있다.
- [0038] 소스 인코더(211)와 소스 디코더가 소스 영상(I_S)으로부터 소스 클래스 확률맵(P_S)을 생성하기 위해 수행하는 신경망 연산을 생성 연산자(G)로 표현할 때, 소스 클래스 확률맵(P_S)과 소스 영상(I_S) 사이의 관계는 $P_S = G(I_S)$ 로 표현될 수 있다.
- [0039] 소스 그룹화부(214)도 인공 신경망으로 구현될 수 있으며, 소스 클래스 확률맵(P_S)을 인가받아 학습되는 방식에

따라 각 픽셀 중 유사한 픽셀들을 그룹화하여 별도의 특징맵으로 구분함으로써 다수의 소스 그룹 특징맵(H_S)을 획득한다. 여기서 소스 그룹화부(214)는 소스 클래스 확률맵(P_S)에서 각 픽셀들의 채널 방향 유사도를 추정하고, 유사한 것으로 추정된 픽셀을 그룹화하여, 다수의 그룹으로 그룹화된 픽셀들의 영역을 나타내는 다수의 소스 그룹 특징맵(H_S)을 획득한다. 여기서 다수의 소스 그룹 특징맵(H_S) 각각은 동일한 그룹으로 그룹화된 픽셀들의 영역이 표시되는 2차원 분할 영역맵의 형태로 획득되며, 소스 그룹화부(214)는 식별 가능한 클래스 개수(ccls)보다 적은 K개(여기서는 일 예로 $K = 8$)로 그룹화하여, K개의 소스 그룹 특징맵(H_S)을 획득할 수 있다.

[0040] 소스 그룹화부(214)가 소스 클래스 확률맵(P_S)으로부터 다수의 소스 그룹 특징맵(H_S)을 획득하기 위해 수행하는 신경망 연산을 그룹화 연산자(C)로 표현할 때, 소스 그룹 특징맵(H_S)과 소스 클래스 확률맵(P_S) 사이의 관계는 $H_S = C(P_S)$ 로 표현될 수 있다.

[0041] 또한 소스 그룹화부(214)가 식별 가능한 클래스 개수(ccls)보다 적은 K개의 소스 그룹 특징맵(H_S)을 획득하는 것은 소스 영상(I_S)에 식별 가능한 모든 클래스의 객체가 포함되지 않는 경우가 많을 뿐만 아니라, 학습 과정 동안 불완전한 학습에 의해 클래스 오식별의 가능성이 높기 때문이다. 특히 후술하는 바와 같이, 타겟 영상 분할부(220)의 타겟 그룹화부(224)가 진리값(Y_T)이 존재하지 않는 타겟 영상(I_T)에 대해 식별되는 모든 클래스 개수에 대응하는 개수의 타겟 그룹 특징맵(H_T)을 획득하는 경우, 정확한 클래스 정보를 획득하기 어려워 잘못된 학습이 수행될 수 있기 때문이다.

[0042] 소스 클러스터맵 획득부(215)는 소스 확률 추정부(212)에서 획득된 소스 클래스 확률맵(P_S)과 소스 그룹화부(214)에서 획득된 K개의 소스 그룹 특징맵(H_S)을 인가받고, 인가된 소스 클래스 확률맵(P_S)을 K개의 소스 그룹 특징맵(H_S) 각각과 원소 곱(Elementwise Multiplication)하여 다수의 소스 클러스터맵(F_S)을 획득한다. 상기한 바와 같이, 소스 클래스 확률맵(P_S)은 클래스 개수(ccls)에 대응하는 채널 길이(c)를 갖는 3차원 특징맵이고, K개의 소스 그룹 특징맵(H_S) 각각은 2차원 특징맵이므로, 소스 클러스터맵(F_S)은 각각 클래스 개수(ccls)에 대응하는 채널 길이(c)를 갖는 K개의 3차원 특징맵으로 획득될 수 있다.

[0043] 한편, 타겟 영상 분할부(220) 또한 소스 영상 분할부(210)와 동일하게 타겟 인코더(221), 타겟 확률 추정부(222), 타겟 분할 출력부(223) 타겟 그룹화부(224) 및 타겟 클러스터맵 획득부(225)를 포함할 수 있다.

[0044] 특히 타겟 인코더(221)와 타겟 확률 추정부(222), 타겟 분할 출력부(223), 타겟 그룹화부(224)는 소스 영상 분할부(210)의 소스 인코더(211), 소스 확률 추정부(212), 소스 분할 출력부(213), 소스 그룹화부(214)와 동일한 구조와 동일한 가중치를 갖는 동일한 인공 신경망으로 구현될 수 있다. 그리고 타겟 클러스터맵 획득부(225)는 소스 클러스터맵 획득부(215)와 동일한 연산을 수행하는 연산 모듈로 구성될 수 있다. 즉 타겟 영상 분할부(220)와 소스 영상 분할부(210)는 동일한 구조를 갖고 함께 학습되어 동일한 연산을 수행하는 삼 신경망(Siamese Network)으로 구현될 수 있다.

[0045] 이에 타겟 인코더(221)는 학습되는 방식에 따라 인가되는 타겟 영상(I_T)을 인코딩하여 타겟 특징맵(f_T)을 획득하고, 타겟 확률 추정부(222)는 타겟 특징맵(f_T)을 학습되는 방식에 따라 디코딩하여 타겟 클래스 확률맵(P_T)을 획득한다. 그리고 타겟 분할 출력부(223)는 타겟 클래스 확률맵(P_T)을 인가받아 각 픽셀에서 채널축 방향에서의 채널값(즉 클래스별 확률값)을 분석하여 최대값을 갖는 채널을 해당 픽셀의 클래스로 할당하여 타겟 영상(I_T)에 대한 2차원 의미론적 분할 영상인 타겟 분할 영상(T_{seg})을 획득하여 출력한다. 본 실시예에 따른 의미론적 영상 분할 장치가 실제 이용되는 경우, 타겟 영상(I_T)이 의미론적 영상 분할될 대상인 입력 영상으로 인가되므로, 의미론적 영상 분할 장치가 학습 완료된 이후 출력되는 타겟 분할 영상(T_{seg})은 의미론적 영상 분할 장치가 입력 영상에 대해 의미론적 영상 분할을 수행한 결과인 의미론적 분할 영상을 나타낸다.

[0046] 여기서도 타겟 확률 추정부(222)와 타겟 분할 출력부(223)는 타겟 디코더로 통합될 수 있다.

[0047] 이때, 타겟 인코더(221)와 타겟 디코더가 소스 인코더(211)와 소스 디코더와 동일한 구조와 동일한 가중치를 갖는 인공 신경망으로 구현되므로, 타겟 인코더(221)와 타겟 디코더에서 수행되는 연산은 소스 인코더(211)와 소스 디코더의 연산자와 동일한 생성 연산자(G)로 표현되며, 타겟 클래스 확률맵(P_T)과 타겟 영상(I_T) 사이의 관계

는 $P_T = G(I_T)$ 로 표현될 수 있다.

- [0048] 타겟 그룹화부(224)는 타겟 클래스 확률맵(P_T)을 인가받아 학습되는 방식에 따라 각 픽셀 중 유사한 픽셀들을 그룹화하여 별도의 특징맵으로 구분함으로써 K개의 타겟 그룹 특징맵(H_T)을 획득한다. 또한 타겟 그룹화부(224)가 타겟 클래스 확률맵(P_T)으로부터 다수의 타겟 그룹 특징맵(H_T)을 획득하기 위해 수행하는 그룹화 연산자(C) 또한 소스 그룹화부(214)의 연산자와 동일하므로, 타겟 그룹 특징맵(H_T)과 타겟 클래스 확률맵(P_T) 사이의 관계는 $H_T = C(P_T)$ 로 표현될 수 있다.
- [0049] 상기한 바와 같이, 타겟 그룹화부(224)가 K개의 타겟 그룹 특징맵(H_T)을 획득하는 것은 진리값(Y_T)이 존재하지 않는 타겟 영상(I_T)에 대해 식별되는 모든 클래스 개수에 대응하는 개수의 타겟 그룹 특징맵(H_T)을 획득하는 경우, 정확한 클래스 정보를 획득하기 어려워 잘못된 학습이 수행될 수 있기 때문이며, 소스 그룹화부(214)와 타겟 그룹화부(224)는 동일한 구조를 갖고 동일한 연산을 수행해야 하므로, 소스 그룹화부(214) 또한 K개의 소스 그룹 특징맵(H_S)을 획득하도록 구성된다.
- [0050] 도 4를 살펴보면, 도 4에서 (a) 및 (b)는 소스 영상(I_S)과 진리값(Y_S)을 나타내고, (c) 내지 (g)는 각각 K개의 소스 그룹 특징맵(H_S) 중 제1, 3, 5, 7 및 8 소스 그룹 특징맵(H_S)을 나타낸다. 그리고 (h) 및 (i)는 타겟 영상(I_T)과 진리값(Y_T)을 나타내고, (j) 내지 (n)는 각각 K개의 타겟 그룹 특징맵(H_T) 중 제1, 3, 5, 7 및 8 타겟 그룹 특징맵(H_T)을 나타낸다. 도 4에 도시된 바와 같이, K개의 소스 그룹 특징맵(H_S)과 K개의 타겟 그룹 특징맵(H_T) 각각은 유사한 특성을 갖는 픽셀들로 구성되는 2차원 특징맵으로 나타난다. 이때 학습이 수행되어 감에 따라 동일한 그룹 특징맵(H_S , H_T)에 포함되는 픽셀은 동일한 클래스에 대응하는 반면, 다른 그룹 특징맵(H_S , H_T)에 포함되는 픽셀은 다른 클래스에 대응해야 한다. 그러나 학습이 완전하게 수행되지 않은 경우, 클래스 식별이 정확하지 않을 수 있으므로, 서로 다른 그룹 특징맵(H_S , H_T)에 포함된 픽셀들이 동일한 클래스를 나타내거나, 동일한 그룹 특징맵(H_S , H_T)에 포함된 픽셀들이 서로 상이한 클래스를 나타낼 수도 있다.
- [0051] 도 4에서 (c)와 (i)에 도시된 제1 그룹 특징맵(H_S^1 , H_T^1)과 (d)와 (k)에 도시된 제3 그룹 특징맵(H_S^3 , H_T^3) 및 (f)와 (m)에 도시된 제7 그룹 특징맵(H_S^7 , H_T^7) 각각 동일하게 차량 클래스와 도로 클래스 및 하늘 클래스에 대응하는 픽셀 영역으로 구성되어 있어 정상적으로 획득된 것으로 볼 수 있다. 그러나 (e)에 도시된 제5 소스 그룹 특징맵(H_S^5)은 하늘을 나타내는 반면, (l)에 도시된 제5 타겟 그룹 특징맵(H_T^5)은 숲을 나타내므로, 서로 상이한 객체를 나타내고 있다. 유사하게 (g)에 도시된 제8 소스 그룹 특징맵(H_S^8)은 벽면 클래스를 나타내는 반면, (n)에 도시된 제8 타겟 그룹 특징맵(H_T^8)은 하늘 클래스를 나타내므로, 서로 상이한 객체를 나타내고 있음을 알 수 있어 학습이 충분하게 수행되지 않았음을 알 수 있다.
- [0052] 타겟 클러스터맵 획득부(225)는 타겟 확률 추정부(222)에서 획득된 타겟 클래스 확률맵(P_T)과 타겟 그룹화부(224)에서 획득된 K개의 타겟 그룹 특징맵(H_T)을 인가받아 원소 곱(Elementwise Multiplication)하여 K개의 타겟 클러스터맵(F_T)을 획득한다.
- [0053] 그리고 클래스 스코어 획득부(300)는 의미론적 영상 분할부(200)에서 획득된 K개의 소스 클러스터맵(F_S)과 K개의 타겟 클러스터맵(F_T) 각각이 다수의 클래스에 대응하는 수준을 나타내는 소스 클래스 스코어 벡터(Q_S)와 타겟 클래스 스코어 벡터(Q_T)를 획득한다.
- [0054] 한편 클래스 마이닝부(400)는 클래스 스코어 획득부(300)와 유사하게 의미론적 영상 분할부(200)에서 획득된 K개의 소스 클러스터맵(F_S)과 K개의 타겟 클러스터맵(F_T) 각각에 다수의 클래스의 특성이 포함되어 있는지 여부를 판별하여 소스 마이닝 벡터(M_S)와 타겟 마이닝 벡터(M_T)를 획득한다.
- [0055] 여기서 클래스 스코어 획득부(300)와 클래스 마이닝부(400)는 입력 영상과 도메인이 상이하거나 양이 적은 소스

영상(I_S)에서 획득된 K개의 소스 클러스터맵(F_S)과 입력 영상과 동일한 도메인이지만 진리값이 존재하지 않는 타겟 영상(I_T)에서 획득된 K개의 타겟 클러스터맵(F_T)을 기반으로 교차 도메인 학습이 수행될 수 있도록 한다.

- [0056] 손실 계산부(500)는 소스 데이터 획득부(110)에서 획득된 소스 영상(I_S)과 소스 분할 영상(S_{seg}), 소스 클러스터맵(F_S)과 타겟 클러스터맵(F_T), 소스 클래스 스코어 벡터(Q_S)와 타겟 클래스 스코어 벡터(Q_T) 및 소스 마이닝 벡터(M_S)와 타겟 마이닝 벡터(M_T)를 인가받아 기지정된 방식으로 의미론적 영상 분할부(200)의 손실(L)을 계산하여 역전파함으로써, 의미론적 영상 분할부(200)를 학습시킨다.
- [0057] 도 5는 도 1의 클래스 스코어 획득부와 클래스 마이닝부의 상세 구조를 나타내고, 도 6은 도 5의 클래스 스코어 획득부의 동작을 설명하기 위한 도면이며, 도 7은 도 5의 클래스 마이닝부의 동작을 설명하기 위한 도면이다.
- [0058] 도 5를 참조하면, 클래스 스코어 획득부(300)는 소스 클러스터맵(F_S)을 인가받아 소스 클래스 스코어 벡터(Q_S)를 획득하는 소스 클래스 스코어 획득부(310)와 타겟 클러스터맵(F_T)을 인가받아 타겟 클래스 스코어 벡터(Q_T)를 획득하는 타겟 클래스 스코어 획득부(320)를 포함할 수 있다.
- [0059] 도 6에 도시된 바와 같이, 소스 클래스 스코어 획득부(310)와 타겟 클래스 스코어 획득부(320) 각각은 대응하는 K개의 클러스터맵($F_l = [F_l^1, \dots, F_l^K]$, 여기서 $l = (S, T)$)을 인가받는다. 즉 K개의 소스 클러스터맵($F_S = [F_S^1, \dots, F_S^K]$)과 K개의 타겟 클러스터맵($F_T = [F_T^1, \dots, F_T^K]$)을 인가받는다. 여기서 K개의 클러스터맵($F_l = [F_l^1, \dots, F_l^K]$) 각각은 다수의 픽셀이 그룹화되고 구분되어 획득된 특징맵으로 클래스의 개수(cls)에 대응하는 채널 길이를 가지며, 각 픽셀에서 채널축 방향의 값은 해당 픽셀이 각 클래스에 대응하는 수준을 나타내는 것으로 볼 수 있다.
- [0060] 소스 클래스 스코어 획득부(310)와 타겟 클래스 스코어 획득부(320) 각각은 K개의 클러스터맵($F_l = [F_l^1, \dots, F_l^K]$) 각각에 대해 $H \times W$ 평면 상에서 평균값 풀링(average pooling)을 수행하여 K개의 클래스 스코어 벡터($Q_l = [Q_l^1, \dots, Q_l^K]$, 여기서 $l = (S, T)$)를 획득할 수 있으며, 이는 소스 영상(I_S)과 타겟 영상(I_T)에서 K개로 그룹화되어 구분된 영역 각각에서 개별 클래스에 대응하는 객체가 분포되는 비율을 나타내는 것으로 볼 수 있다. 즉 구분된 K개의 영역 각각에서 각 클래스에 대응하는 영역의 비율을 나타낸다.
- [0061] 클래스 마이닝부(400)는 소스 클러스터맵(F_S)을 인가받아 소스 마이닝 벡터(M_S)를 획득하는 소스 클래스 마이닝부(410)와 타겟 클러스터맵(F_T)을 인가받아 타겟 클래스 마이닝 벡터(M_T)를 획득하는 타겟 클래스 마이닝부(420)를 포함할 수 있다.
- [0062] 도 7을 참조하면, 소스 클래스 마이닝부(410)와 타겟 클래스 마이닝부(420)는 소스 클래스 스코어 획득부(310)와 타겟 클래스 스코어 획득부(320)와 유사하게 각각 대응하는 K개의 클러스터맵($F_l = [F_l^1, \dots, F_l^K]$)을 인가받고, 인가된 K개의 클러스터맵($F_l = [F_l^1, \dots, F_l^K]$) 각각에 대해 $H \times W$ 평면 상에서 최대값 풀링(max pooling)을 수행하여 K개의 클래스 마이닝 벡터($M_l = [M_l^1, \dots, M_l^K]$, 여기서 $l = (S, T)$)를 획득할 수 있다. K개의 클래스 마이닝 벡터($M_l = [M_l^1, \dots, M_l^K]$) 각각은 K개로 그룹화되어 구분된 영역 각각에서 개별 클래스에 대응하는 객체 영역이 포함되어 있는지 여부를 나타내는 것으로 볼 수 있다.
- [0063] K개의 클래스 스코어 벡터($Q_l = [Q_l^1, \dots, Q_l^K]$)는 K개로 구분된 영역 각각에서 해당 클래스의 객체가 분포되는 비율을 나타내지만, 이와 같은 분포 비율에 따라 객체 영역을 추정하는 경우, 상대적으로 먼거리 등에 위치하여 작은 크기의 객체는 무시되는 경우가 빈번하게 발생할 수 있다. 이에 클래스 마이닝부(400)는 소스 영상(I_S)과 타겟 영상(I_T)에 그룹화되어 구분된 K개의 영역 각각에 각 클래스에 해당하는 객체가 포함되는지 여부를 명확하게 하여 작은 크기의 객체일지라도 정확하게 검출될 수 있도록 한다.
- [0064] 본 실시예에서 클래스 스코어 획득부(300)와 클래스 마이닝부(400)는 클래스 분포 분석부로 통합될 수 있다.

또한 클래스 분포 분석부와 손실 계산부(500)를 통합하여 학습부라 할 수도 있다.

- [0065] 도 8은 도 1의 손실 계산부의 상세 구조를 나타내고, 도 9는 도 8의 손실 계산부의 각 구성별 동작을 설명하기 위한 도면이다.
- [0066] 도 8을 참조하면, 손실 계산부(500)는 분할 손실 계산부(510), 일관성 손실 계산부(520), 직교성 손실 계산부(530), 적대적 도메인 판별부(540), 클래스 손실 계산부(550) 및 총 손실 계산부(560)를 포함할 수 있다.
- [0067] 분할 손실 계산부(510)는 소스 데이터에 소스 영상(I_S)과 함께 포함된 진리값(Y_S)과 소스 영상 분할부(210)의 소스 분할 출력부(213)에서 출력되는 소스 분할 영상(S_{seg})을 인가받아, 기지정된 방식으로 진리값(Y_S)과 소스 분할 영상(S_{seg}) 사이의 크로스 엔트로피(cross entropy)를 계산하여 분할 손실(L_{seg})을 획득한다.
- [0068] 분할 손실 계산부(510)는 소스 영상 분할부(210)의 소스 인코더(211)와 소스 확률 추정부(212) 및 소스 분할 출력부(213)가 소스 영상(I_S)으로부터 정상적인 소스 분할 영상(S_{seg})을 생성할 수 있도록 학습시키기 위한 구성으로 기존에 인공 신경망으로 구성되는 의미론적 영상 분할 장치에서 주로 이용하는 손실이다. 크로스 엔트로피를 기반으로 분할 손실(L_{seg})을 계산하는 방식은 공지된 기술이므로 여기서는 상세하게 설명하지 않는다.
- [0069] 일관성 손실 계산부(520)는 소스 클래스 스코어 벡터(Q_S)와 타겟 클래스 스코어 벡터(Q_T)를 인가받고, 도 9의 (a)에 도시된 바와 같이, 인가된 K개의 소스 클래스 스코어 벡터($Q_S = [Q_S^1, \dots, Q_S^K]$)와 K개의 타겟 클래스 스코어 벡터($Q_T = [Q_T^1, \dots, Q_T^K]$)에서 서로 대응하는 클래스 스코어 벡터 사이의 차이를 누적하여 수학식 1에 따라 일관성 손실(L_{co})을 계산한다.

수학식 1

$$\mathcal{L}_{co}(G, C) = \sum_{k \in \{1, \dots, K\}} \|Q_S^k - Q_T^k\|^2.$$

- [0070]
- [0071] 본 실시예에서 비록 소스 영상(I_S)과 타겟 영상(I_T)이 서로 상이한 영상이고, 서로 다른 도메인에서 획득된 영상일 수도 있으나, 소스 영상(I_S)과 타겟 영상(I_T)을 함께 이용하여 학습을 수행하므로, 소스 영상(I_S)과 타겟 영상(I_T)에서 획득되는 소스 클래스 스코어 벡터(Q_S)와 타겟 클래스 스코어 벡터(Q_T)의 서로 대응하는 클래스에서는 일관되는 공통 성분이 포함되어야 한다. 이에 일관성 손실 계산부(520)는 소스 클래스 스코어 벡터(Q_S)와 타겟 클래스 스코어 벡터(Q_T) 사이의 차이를 일관성 손실(L_{co})로 계산하여 획득한다.
- [0072] 직교성 손실 계산부(530)는 도 9의 (b)에 도시된 바와 같이, 인가된 K개의 소스 클래스 스코어 벡터($Q_S = [Q_S^1, \dots, Q_S^K]$)는 서로 상이해야 하고, 마찬가지로 K개의 타겟 클래스 스코어 벡터($Q_T = [Q_T^1, \dots, Q_T^K]$)또한 서로 상이해야 한다는 점에서 직교성 손실(L_{orth})을 계산한다.
- [0073] 상기한 바와 같이, K개의 소스 클래스 스코어 벡터($Q_S = [Q_S^1, \dots, Q_S^K]$)와 K개의 타겟 클래스 스코어 벡터($Q_T = [Q_T^1, \dots, Q_T^K]$)는 각각 유사 픽셀들이 그룹화되어 그룹별로 구분된 영역에서 획득된 K개의 소스 클러스터맵($F_S = [F_S^1, \dots, F_S^K]$)과 K개의 타겟 클러스터맵($F_T = [F_T^1, \dots, F_T^K]$)으로부터 획득된다. 따라서, K개의 소스 클래스 스코어 벡터($Q_S = [Q_S^1, \dots, Q_S^K]$) 각각은 서로 직교성이 있어야 하며, K개의 타겟 클래스 스코어 벡터($Q_T = [Q_T^1, \dots, Q_T^K]$) 각각 또한 서로 직교성이 있어야 한다.
- [0074] 이에 직교성 손실 계산부(530)는 수학식 2 및 수학식 3을 이용하여 소스 클래스 스코어 벡터(Q_S)와 타겟 클래스

스코어 벡터(Q_T) 각각의 직교성 손실(L_{orth})을 계산할 수 있다.

수학식 2

$$\cos(Q_l^{j_1}, Q_l^{j_2}) = \frac{Q_l^{j_1} \cdot Q_l^{j_2}}{\|Q_l^{j_1}\|_2 \|Q_l^{j_2}\|_2}, \quad j_1, j_2 \in \{1, \dots, K\}.$$

수학식 3

$$\mathcal{L}_{orth}(G, C) = \sum_l \sum_{j_1, j_2} \cos(Q_l^{j_1}, Q_l^{j_2}),$$

수학식 2 및 3에서 $l = (S, T)$ 이다.

직교성 손실 계산부(530)는 도 8에 도시된 바와 같이, 소스 클래스 스코어 벡터(Q_S)에 대한 직교성 손실인 소스 직교성 손실(L_{orthS})을 계산하는 소스 직교성 손실 계산부(531)와 타겟 클래스 스코어 벡터(Q_T)에 대한 직교성 손실인 타겟 직교성 손실(L_{orthT})을 계산하는 타겟 직교성 손실 계산부(532)를 포함할 수 있다.

적대적 도메인 판별부(540)는 소스 클래스 확률맵(P_S)을 생성하는 소스 인코더(211)와 소스 디코더와 타겟 클래스 확률맵(P_T)을 생성하는 타겟 인코더(221)와 타겟 디코더를 생성자로 고려할 때, 생성자에 대항하는 적대적 판별자로서 기능하는 인공 신경망으로 구현될 수 있다.

도 9의 (d)에 도시된 바와 같이, 적대적 도메인 판별부(540)는 K 개의 소스 클러스터맵($F_S = [F_S^1, \dots, F_S^K]$)과 K 개의 소스 클래스 스코어 벡터($Q_S = [Q_S^1, \dots, Q_S^K]$)을 인가받아, 서로 대응하는 소스 클러스터맵과 소스 클래스 스코어 벡터가 외적(Outer Product)된 결과를 인가받아, 학습되는 방식에 따라 인가되는 외적값이 소스에 대한 값인지 판별한다. 또한 적대적 도메인 판별부(540)는 K 개의 소스 클러스터맵 타겟 클러스터맵($F_T = [F_T^1, \dots, F_T^K]$)과 K 개의 타겟 클래스 스코어 벡터($Q_T = [Q_T^1, \dots, Q_T^K]$)을 인가받아, 서로 대응하는 타겟 클러스터맵과 타겟 클래스 스코어 벡터가 외적된 결과를 인가받아, 학습되는 방식에 따라 인가되는 외적값이 소스 영상(I_S)에 대한 값인지 판별한다. 적대적 도메인 판별부(540)는 소스 영상(I_S)에 대한 값으로 판별되면 1의 판별값을 출력하는 반면, 타겟 영상(I_T)에 대한 값으로 판별되면 0의 판별값을 출력하도록 할 수 있다. 따라서 학습이 수행되는 동안 생성자인 소스 인코더(211)와 소스 디코더와 타겟 인코더(221)와 타겟 디코더는 적대적 판별자인 적대적 도메인 판별부(540)가 오판별하도록 학습되는 반면, 적대적 도메인 판별부(540)는 생성자에서 생성된 결과가 소스 영상(I_S)으로부터 추출된 값인지 여부를 정확하게 판별하도록 학습된다.

적대적 도메인 판별부(540)는 수학식 4에 따라 적대적 도메인 손실(L_{adv})을 계산할 수 있다.

수학식 4

$$\mathcal{L}_{adv}(G, C, D) = - \sum_k [\log(D(F_S^k \otimes Q_S^k))] - \sum_k [\log(1 - D(F_T^k \otimes Q_T^k))],$$

[0083] 여기서 $D()$ 는 적대적 도메인 판별부(540)의 판정 함수를 나타내고, \otimes 는 외적 연산자를 나타낸다.

[0084] 클래스 손실 계산부(550)는 K 개의 소스 클래스 마이닝 벡터($M_s = [M_s^1, \dots, M_s^K]$)와 K 개의 타겟 클래스 마이닝 벡터($M_t = [M_t^1, \dots, M_t^K]$)를 인가받아, 도 9의 (c)에 도시된 바와 같이, 서로 대응하는 마이닝 벡터에서 서로 대응하는 클래스의 존재 여부를 반영하여 수학적 식 5에 따라 클래스 균형 도메인 적응 손실(L_{cl})을 계산한다.

수학적 식 5

$$\mathcal{L}_{cl}(G, C) = - \sum_k \sum_u [m_{S,u}^k \geq \tau] \log(m_{T,u}^k),$$

[0085]

[0086] $m_{S,u}^k$ 와 $m_{T,u}^k$ 는 각각 제 k 소스 클래스 마이닝 벡터(M_s^k)와 제 k 타겟 클래스 마이닝 벡터(M_t^k)에서 특정 클래스 값을 나타내는 값이고, τ 는 기지정된 문턱값(여기서는 일 예로 0.3)을 나타낸다.

[0087] 클래스 손실 계산부(550)는 소스 영상(I_s)과 타겟 영상(I_t)이 서로 상이한 도메인에서 획득된 영상일지라도, 동일한 클래스의 객체를 포함해야 한다는 관점에서 클래스 균형 도메인 적응 손실(L_{cl})을 수학적 식 5에 따라 계산한다.

[0088] 총 손실 계산부(560)는 분할 손실 계산부(510), 일관성 손실 계산부(520), 직교성 손실 계산부(530), 적대적 도메인 판별부(540) 및 클래스 손실 계산부(550) 각각에서 계산된 분할 손실(L_{seg})과 일관성 손실(L_{co}), 직교성 손실(L_{orth}), 적대적 도메인 손실(L_{cadv}) 및 클래스 균형 도메인 적응 손실(L_{cl})을 인가받아 수학적 식 6에 따라 총 손실(L)을 계산한다.

수학적 식 6

$$L(G, C, D) = L_{seg}(G) + \lambda_{co}L_{co}(G, C) + \lambda_{cl}L_{cl}(G, C) + \lambda_{orth}L_{orth}(G, C) + \lambda_{cadv}L_{cadv}(G, C, D),$$

[0089]

[0090] 여기서 λ_{co} , λ_{orth} , λ_{cadv} , λ_{cl} 각각은 손실 가중치이다.

[0091] 그리고 총 손실 계산부(560)는 계산된 총 손실(L)을 역전파하여, 의미론적 영상 분할 장치를 학습시키며, 기지정된 횟수 또는 총 손실(L)이 기지정된 손실 문턱값 이하가 되는 것과 같이 기지정된 조건에 도달하면, 학습을 종료한다.

[0092] 한편, 본 실시예에 따른 의미론적 영상 분할 장치는 입력 영상에 대해 의미론적 분할 영상을 획득하는 것을 목적으로 하므로, 의미론적 영상 분할 장치가 실제 이용되는 경우, 소스 데이터 획득부(110)와 소스 영상 분할부(210), 클래스 스코어 획득부(300), 클래스 마이닝부(400) 및 손실 계산부(500)는 생략될 수 있다. 또한 타겟 영상 분할부(220)에서도 타겟 인코더(221)와 타겟 확률 추정부(222) 및 타겟 분할 출력부(223)만을 이용하고, 타겟 그룹화부(224)와 타겟 클러스터맵 획득부(225)는 제거될 수 있다. 즉 실질적으로 의미론적 영상 분할 장치는 타겟 영상 획득부(120)와 타겟 인코더(221) 및 타겟 디코더만으로 구성될 수 있으며, 나머지 구성은 타겟 인코더(221) 및 타겟 디코더가 정상적으로 입력 영상에 대해 의미론적 분할을 수행할 수 있도록 학습을 위해 구비되는 구성으로 볼 수 있다.

[0093] 도 10은 본 실시예에 따른 의미론적 영상 분할 장치의 의미론적 영상 분할 결과의 일 예를 나타낸다.

[0094] 도 10에서 (a)와 (f)는 각각 소스 영상(I_s)과 타겟 영상(I_t)을 나타내고, (b)와 (g)는 소스 영상(I_s)과 타겟 영

상(I_T) 각각에 대한 진리값(Y_S , Y_T)인 의미론적 분할 영상을 나타낸다. 그리고 (c)와 (h)는 교차 도메인 학습되지 않은 의미론적 영상 분할 장치에서 획득된 의미론적 분할 영상을 나타내고, (d) 및 (i)는 기존의 베이스라인 방식으로 학습된 의미론적 영상 분할 장치에서 획득된 의미론적 분할 영상을 나타낸다. 마지막으로 (e)와 (j)는 본 실시예에 따른 의미론적 영상 분할 장치에 의해 획득된 의미론적 분할 영상을 나타낸다. 도 10을 참조하면, 본 실시예에 따른 의미론적 영상 분할 장치에 의해 획득된 의미론적 분할 영상이 기존의 방식으로 획득된 의미론적 분할 영상에 비해 매우 우수한 성능을 나타내어 진리값(Y_S , Y_T)에 유사한 결과를 도출할 수 있음을 알 수 있다.

[0095] 도 11은 본 발명의 일 실시예에 따른 의미론적 영상 분할 방법을 나타내고, 도 12는 도 11의 손실 역전과 단계를 상세하게 나타낸 도면이다.

[0096] 도 11을 참조하면, 본 실시예에 따른 의미론적 영상 분할 방법은 크게 학습 단계(S100)와 의미론적 영상 분할 단계(S200)로 구분될 수 있다. 학습 단계(S200)는 우선 학습을 위한 학습 데이터로서 소스 영상(I_S)과 소스 영상(I_S)에 대해 의미론적 영상 분할된 진리값(Y_S)을 포함하는 소스 데이터와 타겟 영상(I_T)을 획득한다(S110). 여기서 타겟 영상(I_T)은 이후, 의미론적 영상 분할 단계(S200)에서 획득되는 입력 영상과 동일한 도메인의 영상이지만, 소스 영상(I_S)은 다른 도메인에서 획득된 영상이어도 무방하다.

[0097] 소스 데이터와 타겟 영상(I_T)이 획득되면, 인공 신경망을 이용하여 소스 영상(I_S)과 타겟 영상(I_T)을 각각 학습되는 방식에 따라 인코딩하여 소스 특징맵(f_S)과 타겟 특징맵(f_T)을 획득한다(S120). 그리고 소스 특징맵(f_S)과 타겟 특징맵(f_T)을 디코딩하여 소스 영상(I_S)과 타겟 영상(I_T)의 각 픽셀이 기지정된 다수의 클래스 각각에 대응할 확률을 추정하여 소스 클래스 확률맵(P_S)과 타겟 클래스 확률맵(P_T)을 획득한다(S130). 여기서 소스 클래스 확률맵(P_S)과 타겟 클래스 확률맵(P_T)은 각각 소스 영상(I_S) 및 타겟 영상(I_T)과 동일한 폭(W)과 높이(H)를 갖고 기지정된 클래스 개수에 대응하는 채널 길이($c = cls$)를 갖는 3차원 특징맵으로 획득될 수 있다.

[0098] 여기서 소스 영상(I_S)으로부터 소스 클래스 확률맵(P_S)을 획득하는 인공 신경망과 타겟 영상(I_T)으로부터 타겟 클래스 확률맵(P_T)을 획득하는 인공 신경망은 동일한 구조와 동일한 가중치를 갖는 동일 신경망으로 구성된다.

[0099] 소스 클래스 확률맵(P_S)과 타겟 클래스 확률맵(P_T)이 획득되면, 소스 클래스 확률맵(P_S)의 각 픽셀의 채널축 방향에서의 채널값(클래스별 확률값)을 분석하여 최대값을 갖는 채널의 클래스를 해당 픽셀의 클래스로 할당하여 소스 분할 영상(S_{seg})을 획득한다(S140).

[0100] 한편, 학습되는 방식에 따라 소스 클래스 확률맵(P_S)과 타겟 클래스 확률맵(P_T) 각각에서 다수의 픽셀 중 유사한 픽셀들을 클래스 개수(cls)보다 작은 기지정된 K개의 그룹으로 그룹화하여, K개의 소스 그룹 특징맵(H_S)과 K개의 타겟 그룹 특징맵(H_T)을 획득하고, 획득된 소스 클래스 확률맵(P_S)과 K개의 소스 그룹 특징맵(H_S) 각각을 원소 곱하고, 타겟 클래스 확률맵(P_T)과 K개의 타겟 그룹 특징맵(H_T) 각각을 원소 곱하여, K개의 소스 클러스터맵($F_S = [F_S^1, \dots, F_S^K]$)과 K개의 타겟 클러스터맵($F_T = [F_T^1, \dots, F_T^K]$)을 획득한다(S150).

[0101] 이후 획득된 K개의 소스 클러스터맵(F_S)과 K개의 타겟 클러스터맵(F_T) 각각에 대해 평균값 풀링을 수행하여 소스 영상(I_S)과 타겟 영상(I_T)에서 K개로 그룹화되어 구분된 영역 각각에서 개별 클래스에 대응하는 객체가 분포되는 비율을 나타내는 K개의 소스 클래스 스코어 벡터($Q_S = [Q_S^1, \dots, Q_S^K]$)와 K개의 타겟 클래스 스코어 벡터($Q_T = [Q_T^1, \dots, Q_T^K]$)를 획득한다(S160).

[0102] 이와 함께 K개의 소스 클러스터맵(F_S)과 K개의 타겟 클러스터맵(F_T) 각각에 대해 최대값 풀링을 수행하여 소스 영상(I_S)과 타겟 영상(I_T)에서 K개로 그룹화되어 구분된 영역 각각에서 개별 클래스에 대응하는 객체의 존재 여부를 나타내는 K개의 소스 클래스 마이닝 벡터($M_S = [M_S^1, \dots, M_S^K]$)와 K개의 타겟 클래스 마이닝 벡터($M_T = [M_T^1, \dots, M_T^K]$)를 획득한다(S170).

- [0103] 그리고 획득된 소스 영상(I_s)에 대한 진리값(Y_s)과 소스 분할 영상(S_{seg}), K개의 소스 클러스터맵(F_s)과 K개의 타겟 클러스터맵(F_T), K개의 소스 클래스 스코어 벡터(Q_s)와 K개의 타겟 클래스 스코어 벡터(Q_T) 및 K개의 소스 마이닝 벡터(M_s)와 K개의 타겟 마이닝 벡터(M_T)를 기반으로 기지정된 방식으로 손실(L)을 계산하여 역전파함으로써, 인공 신경망을 학습시킨다(S180).
- [0104] 도 12를 참조하면, 손실을 역전파하는 단계(S180)는 우선 진리값(Y_s)과 소스 분할 영상(S_{seg}) 사이의 크로스 엔트로피를 기지정된 방식으로 계산하여 분할 손실(L_{seg})로 획득한다(S181).
- [0105] 그리고 K개의 소스 클래스 스코어 벡터(Q_s)와 K개의 타겟 클래스 스코어 벡터(Q_T)에서 서로 대응하는 클래스 스코어 벡터 사이의 차이를 누적하여 수학적 1에 따라 일관성 손실(L_{co})을 계산한다(S182).
- [0106] 또한 K개의 소스 클래스 스코어 벡터(Q_s)와 K개의 타겟 클래스 스코어 벡터(Q_T) 각각 사이의 직교성 손실(L_{orth})을 수학적 2 및 3에 따라 계산한다(S183).
- [0107] 한편, K개의 소스 클래스 스코어 벡터(Q_s)와 K개의 소스 클러스터맵(F_s) 중 서로 대응하는 소스 클래스 스코어 벡터와 소스 클러스터맵을 외적하고, K개의 타겟 클래스 스코어 벡터(Q_T)와 K개의 타겟 클래스 스코어 벡터(Q_T)를 외적한 결과가 학습되는 방식에 따라 소스에 대한 값인지 판별하고, 판별 결과에 따른 적대적 도메인 손실(L_{adv})을 수학적 4에 따라 계산한다(S184).
- [0108] 마지막으로 K개의 소스 클래스 마이닝 벡터(M_s)와 K개의 타겟 클래스 마이닝 벡터(M_T) 중 서로 대응하는 마이닝 벡터에서 서로 대응하는 클래스의 존재 여부를 반영하여 수학적 5에 따라 클래스 균형 도메인 적응 손실(L_{cl})을 계산한다(S185).
- [0109] 이후 계산된 분할 손실(L_{seg})과 일관성 손실(L_{co}), 직교성 손실(L_{orth}), 적대적 도메인 손실(L_{adv}) 및 클래스 균형 도메인 적응 손실(L_{cl})을 수학적 6과 같이 기지정된 방식으로 합산하여 총 손실(L)을 계산하고, 계산된 총 손실(L)을 역전파한다(S186).
- [0110] 총 손실(L)이 역전파되면, 학습이 기지정된 조건에 도달하였는지 분석하여, 학습을 종료할지 여부를 판별한다(S190). 만일 학습이 더 수행되어야 하는 것으로 판별되면, 학습 단계(S100)를 다시 수행한다. 그러나 학습이 종료된 것으로 판별되면, 의미론적 영상 분할 단계(S200)를 수행한다.
- [0111] 의미론적 영상 분할 단계(S200)에서는 우선 입력 영상을 획득한다(S210). 그리고 미리 학습된 방식에 따라 입력 영상을 인코딩하여, 입력 특징맵(f)을 획득한다(S220). 입력 특징맵(f)이 획득되면, 획득된 입력 특징맵(f)을 미리 학습된 방식에 따라 디코딩하여, 입력 클래스 확률맵(P)을 획득한다(S230). 여기서 입력 영상은 의미론적 영상 분할 대상이 되는 영상으로, 학습 단계(S100)의 타겟 영상(I_T)과 동일한 도메인의 영상이며, 이에 입력 영상을 타겟 영상으로 간주할 때, 타겟 영상과 동일한 방식으로 인코딩 및 디코딩되어 획득되는 입력 특징맵(f)과 입력 클래스 확률맵(P)은 타겟 특징맵(f_T)과 타겟 클래스 확률맵(P_T)으로 볼 수 있다.
- [0112] 그리고 입력 클래스 확률맵(P)의 각 픽셀의 채널축 방향에서의 채널값(클래스별 확률값)을 분석하여 최대값을 갖는 채널의 클래스를 해당 픽셀의 클래스로 할당하여 입력 영상에 대한 의미론적 분할 영상을 획득한다(S240). 이때, 입력 클래스 확률맵(P)으로부터 의미론적 분할 영상을 획득하는 방식은 소스 클래스 확률맵(P_s)으로부터 소스 분할 영상(S_{seg})을 획득하는 방식과 동일하게 구현될 수 있으며, 미리 학습된 인공 신경망을 이용하여 구현될 수도 있다.
- [0113] 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.
- [0114] 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.

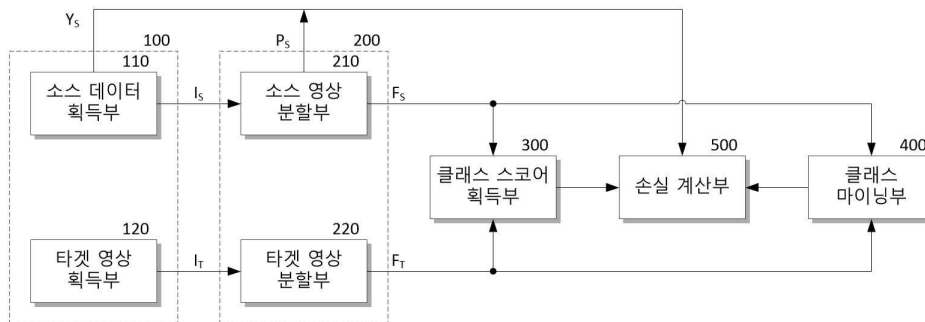
[0115] 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 청구범위의 기술적 사상에 의해 정해져야 할 것이다.

부호의 설명

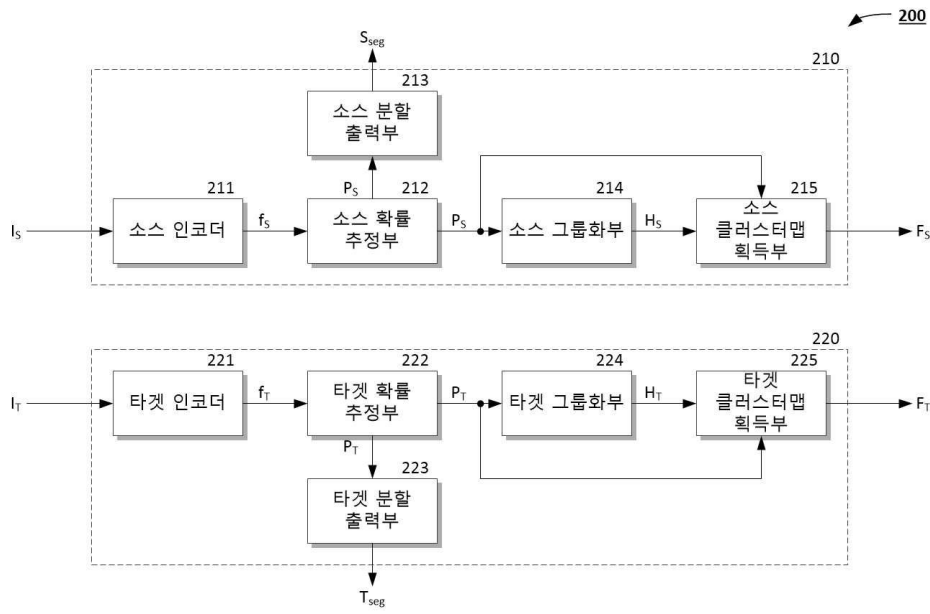
[0116]	100: 영상 획득부	110: 소스 데이터 획득부
	120: 타겟 영상 획득부	200: 의미론적 영상 분할부
	210: 소스 영상 분할부	211: 소스 인코더
	212: 소스 확률 추정부	213: 소스 분할 출력부
	214: 소스 그룹화부	215 소스 클러스터맵 획득부
	220: 타겟 영상 분할부	221: 타겟 인코더
	222: 타겟 확률 추정부	223: 타겟 분할 출력부
	224: 타겟 그룹화부	225: 타겟 클러스터맵 획득부
	300: 클래스 스코어 획득부	310: 소스 클래스 스코어 획득부
	320: 타겟 클래스 스코어 획득부	400: 클래스 마이닝부
	410: 소스 클래스 마이닝부	420: 타겟 클래스 마이닝부
	500: 손실 계산부	510: 분할 손실 계산부
	520: 일관성 손실 계산부	531: 소스 직교성 손실 계산부
	532: 타겟 직교성 손실 계산부	540: 적대적 도메인 판별부
	550: 클래스 손실 계산부	560: 총 손실 계산부

도면

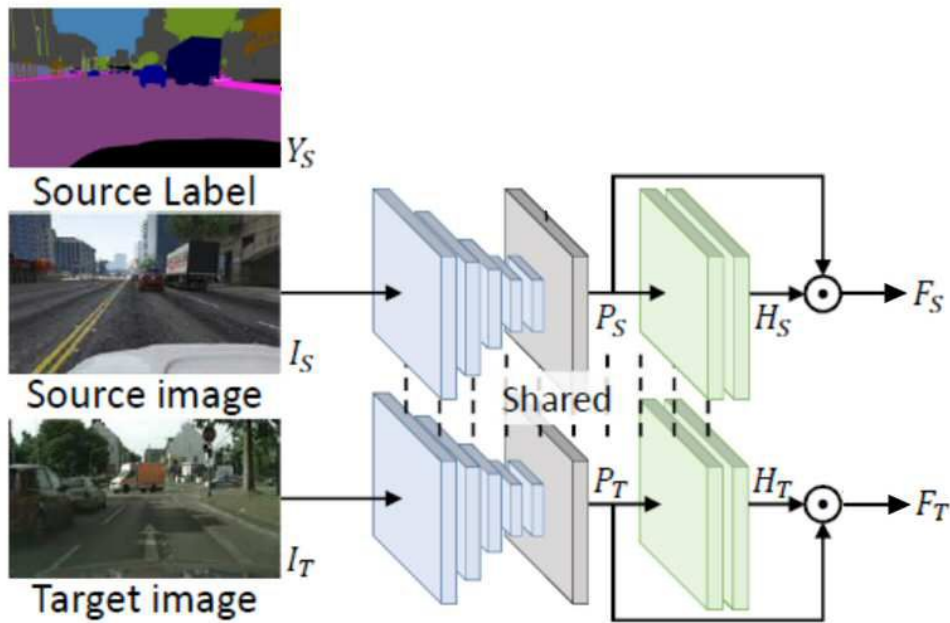
도면1



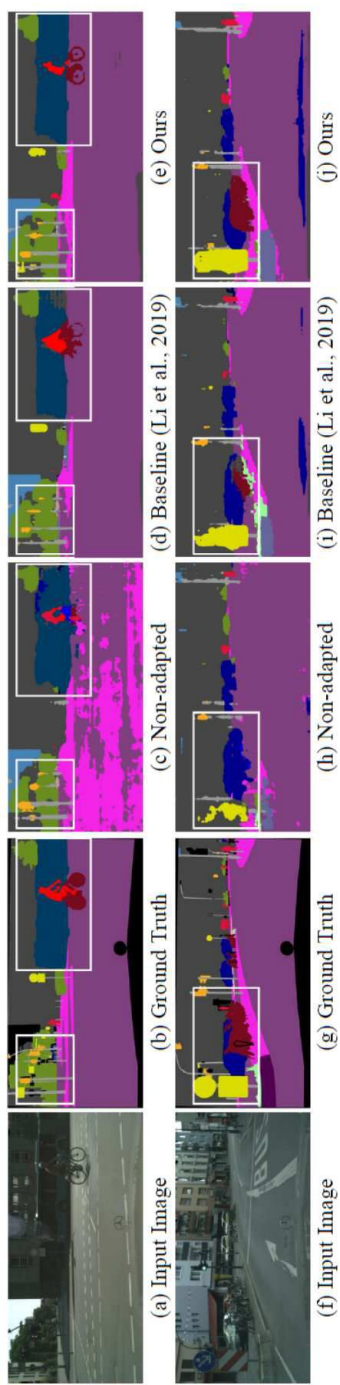
도면2



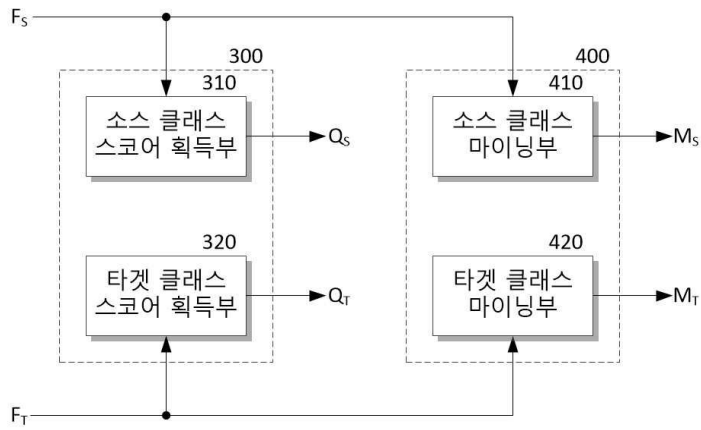
도면3



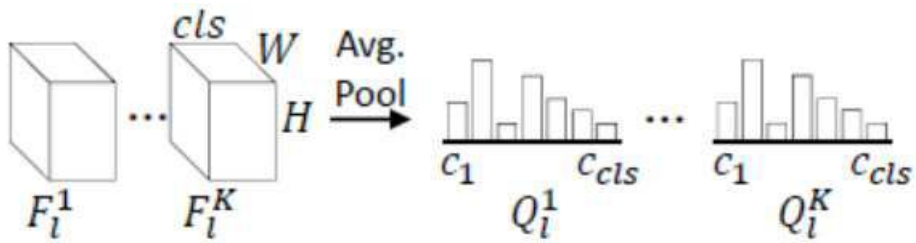
도면4



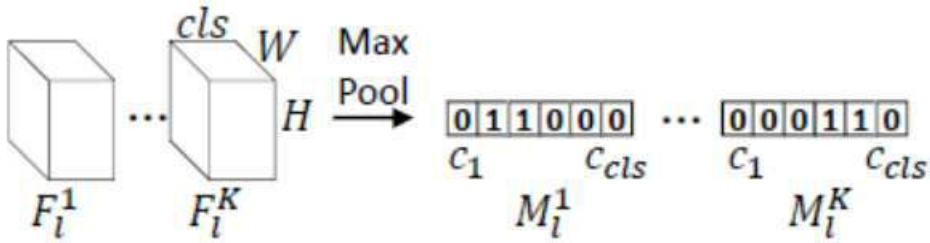
도면5



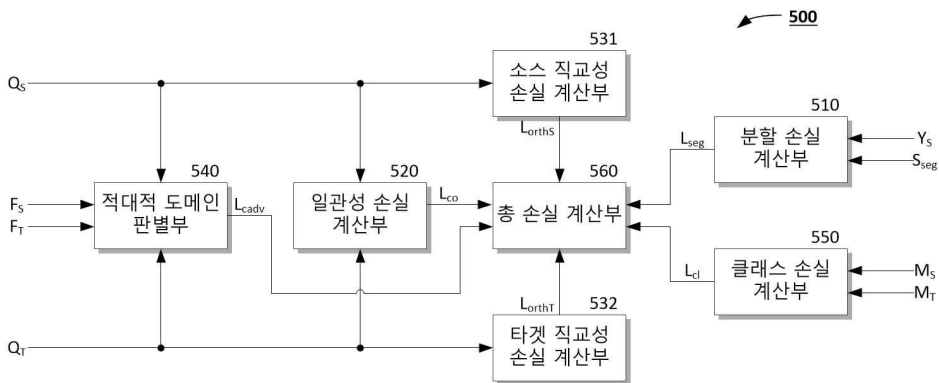
도면6



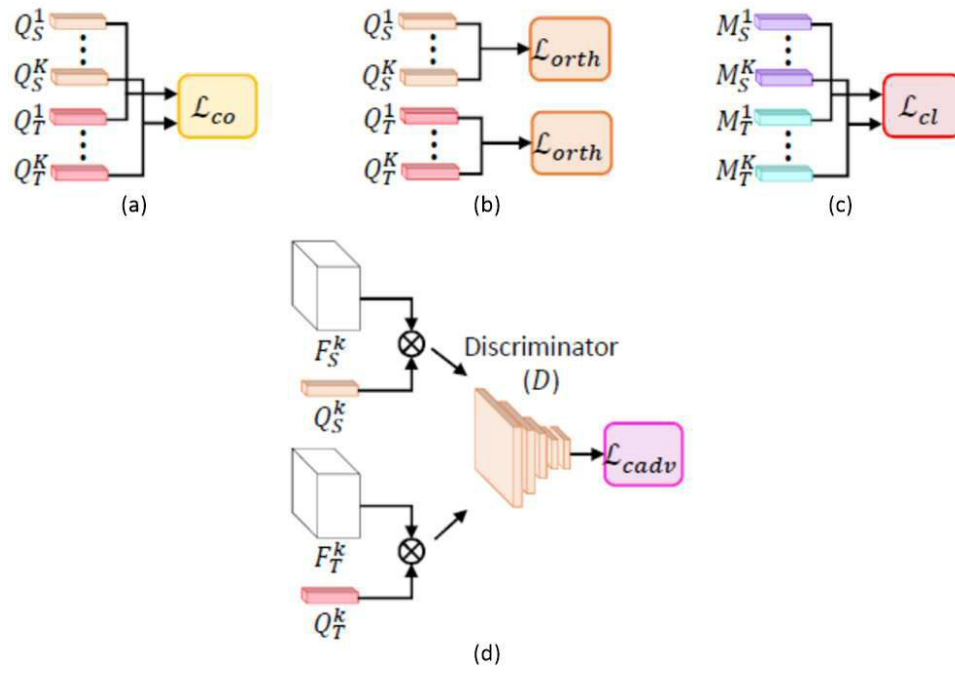
도면7



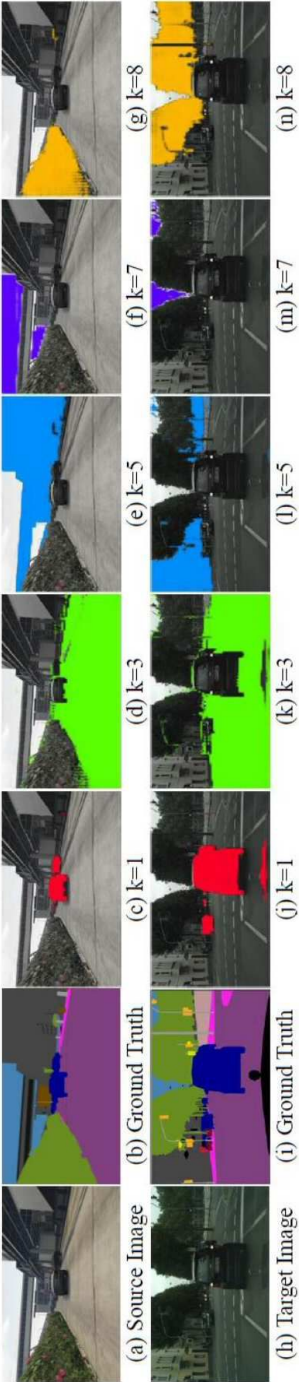
도면8



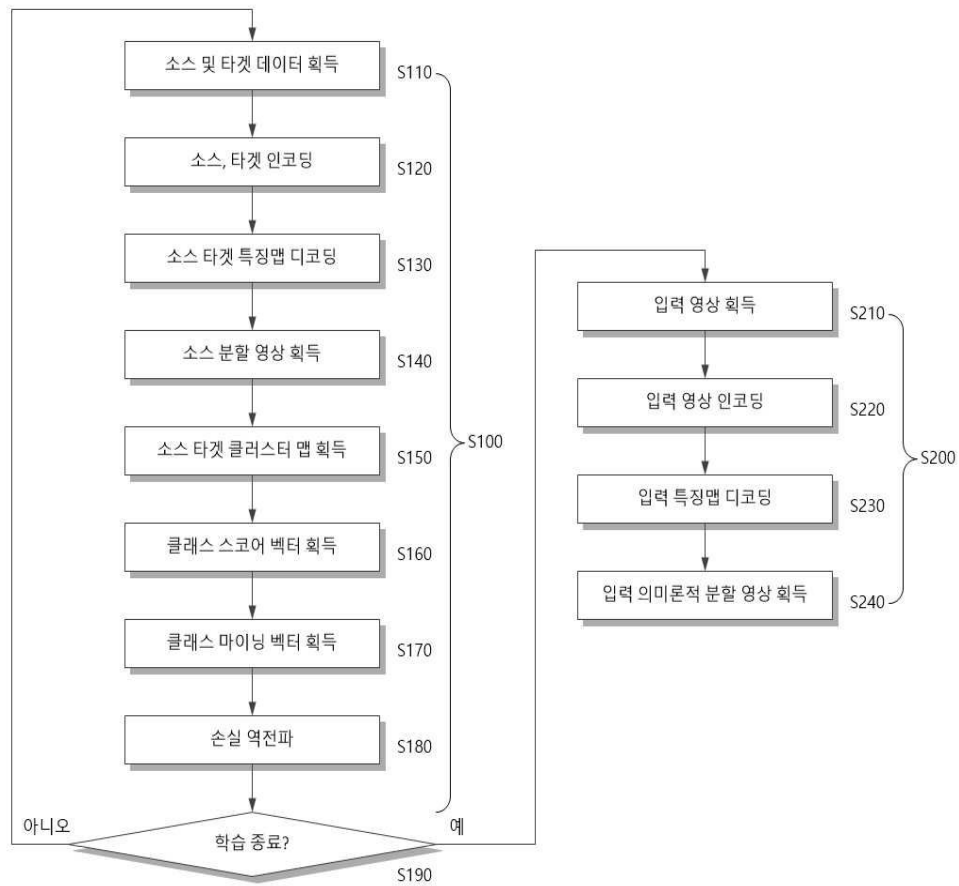
도면9



도면10



도면11



도면12

