(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0103956
(43) 공개일자 2024년07월04일

(51) 국제특허분류(Int. Cl.)
G06V 10/764 (2022.01) *G06V 10/10* (2023.01)
G06V 10/26 (2022.01) *G06V 10/774* (2022.01)
(52) CPC특허분류
G06V 10/764 (2023.08)
G06V 10/10 (2023.08)
(21) 출원번호 10-2023-0107891
(22) 출원일자 2023년08월17일
심사청구일자 2023년08월17일
(30) 우선권주장
1020220186054 2022년12월27일 대한민국(KR)

(71) 출원인
포항공과대학교 산학협력단
경상북도 포항시 남구 청암로 77 (지곡동)
(72) 발명자
조민수
경상북도 포항시 남구 청암로 77
강다현
경상북도 포항시 남구 청암로 77
(74) 대리인
인비전 특허법인

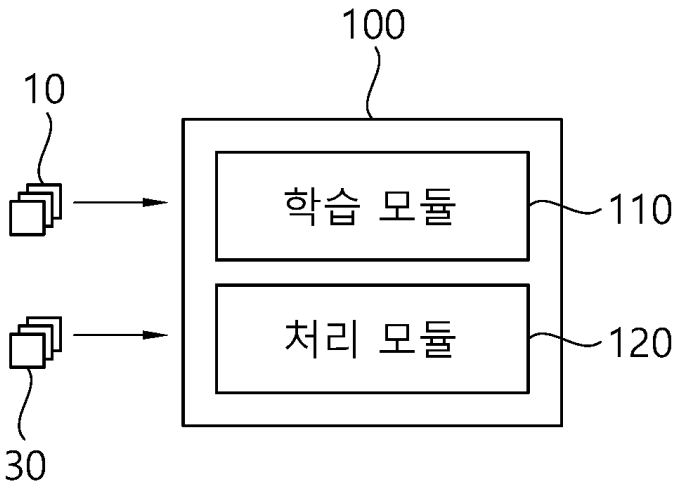
전체 청구항 수 : 총 18 항

(54) 발명의 명칭 **퓨-샷 학습 방법 및 이를 이용한 영상 처리 시스템**

(57) 요 약

본 발명에 따른 퓨-샷 학습 방법은 이미지 및 상기 이미지에 대한 분할 예측값(segmentation prediction)을 획득하는 단계 및 기학습된 모델로 쿼리 이미지가 제공될 때에 상기 쿼리 이미지로부터 특정 영역의 분류 및 분할을 동시에 수행하도록 상기 이미지 및 분할 지표를 기반으로 상기 모델을 학습하는 단계를 포함하고, 상기 모델은 ASNet(Attentive Squeeze Network)을 포함하며, 상기 쿼리 이미지에 관련성이 낮은 물체가 존재할 때에 분류를 수행하지 않고 관련성이 높은 물체가 존재할 때에 분류 및 분할을 수행할 수 있다.

대 표 도 - 도1



(52) CPC특허분류
G06V 10/26 (2023.08)
G06V 10/774 (2023.08)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711160525
과제번호	2022-0-00959-001
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	사람중심인공지능핵심원천기술개발(R&D)
연구과제명	(2세부) 의사결정 지원을 위한 퓨샷 학습 기반 시각및 언어에 대한 인과관계 추론기술개발
기 여 율	1/1
과제수행기관명	서울대학교 산학협력단
연구기간	2022.04.01 ~ 2022.12.31

명 세 서

청구범위

청구항 1

이미지 및 상기 이미지에 대한 분할 예측값(segmentation prediction)을 획득하는 단계; 및

기학습된 모델로 쿼리 이미지가 제공될 때에 상기 쿼리 이미지로부터 특정 영역의 분류 및 분할을 동시에 수행하도록 상기 이미지 및 분할 지표를 기반으로 상기 모델을 학습하는 단계를 포함하고,

상기 모델은

ASNet(Attentive Squeeze Network)을 포함하며, 상기 쿼리 이미지에 관련성이 낮은 물체가 존재할 때에 분류를 수행하지 않고 관련성이 높은 물체가 존재할 때에 분류 및 분할을 수행하는 것을 특징으로 하는 퓨-샷 학습 방법.

청구항 2

제1 항에 있어서,

상기 모델의 학습에서는

통합적 퓨-샷 학습(Integrative few-shot learning, iFSL) 방법이 적용되고,

상기 통합적 퓨-샷 학습은

상기 쿼리 이미지 내 등장하는 하위 집합을 식별하고 클래스에 해당하는 문제 분할 마스크 집합을 예측하도록 상기 모델을 학습하는 것을 특징으로 하는 퓨-샷 학습 방법.

청구항 3

제1 항에 있어서,

상기 모델은

복수 개의 이미지 사이의 상관 텐서를 계산하고, 상기 상관 텐서를 Strided self-attention layer에 통과시켜 분류 맵을 생성하는 것을 특징으로 하는 퓨-샷 학습 방법.

청구항 4

제3 항에 있어서

상기 ASNet은

AS Layer(Attentive Squeeze Layer)를 포함하며,

상기 AS Layer는

고차 자기주의(High-order self-attention) 레이어로 마련되어 상기 상관 텐서를 기반으로 다른 수준의 상관 표현을 반환하는 것을 특징으로 하는 퓨-샷 학습 방법.

청구항 5

제4 항에 있어서,

상기 ASNet은

상기 쿼리 이미지와 서포트 이미지 사이의 피라미드형 교차 상관관계 텐서인 초상관관계(Hypercorrelation)를 입력으로 하는 것을 특징으로 하는 퓨-샷 학습 방법.

청구항 6

제2 항에 있어서,

상기 통합적 퓨-샷 학습에서는

맥스 풀링(Max pooling)을 이용하여 추론을 수행하는 것을 특징으로 하는 퓨-샷 학습 방법.

청구항 7

제2 항에 있어서,

상기 통합적 퓨-샷 학습에서는

분류 손실과 분할 손실을 사용하며, 클래스 태그 또는 분할 주석을 사용하여 학습자를 훈련하는 것을 특징으로 하는 퓨-샷 학습 방법.

청구항 8

제7 항에 있어서,

상기 분류 손실은

공간적으로 평균 풀링된 클래스 점수와 그 정답 클래스 레이블 사이의 평균 이진 교차 엔트로피인 것을 특징으로 하는 퓨-샷 학습 방법.

청구항 9

제7 항에 있어서,

상기 분할 손실은

개별 위치의 클래스 분포와 실제 분할 주석 사이의 평균 교차 엔트로피인 것을 특징으로 하는 퓨-샷 학습 방법.

청구항 10

외부로부터 제공되는 이미지를 기학습된 모델에 입력하여 상기 이미지로부터 특정 영역의 분류 및 분할을 동시에 수행하는 처리모듈을 포함하고,

상기 모델의 학습에서는

이미지 및 상기 이미지에 대한 분할 예측값(segmentation prediction)을 획득하고,

기학습된 모델로 쿼리 이미지가 제공될 때에 상기 쿼리 이미지로부터 특정 영역의 분류 및 분할을 동시에 수행하도록 상기 이미지 및 분할 지표를 기반으로 상기 모델을 학습하며,

상기 모델은

ASNet(Attentive Squeeze Network)을 포함하며, 상기 쿼리 이미지에 관련성이 낮은 물체가 존재할 때에 분류를 수행하지 않고 관련성이 높은 물체가 존재할 때에 분류 및 분할을 수행하는 것을 특징으로 하는 영상 처리 시스템.

청구항 11

제1 항에 있어서,

상기 모델의 학습에서는

통합적 퓨-샷 학습(Integrative few-shot learning, iFSL) 방법이 적용되고,

상기 통합적 퓨-샷 학습은

상기 쿼리 이미지 내 등장하는 하위 집합을 식별하고 클래스에 해당하는 문제 분할 마스크 집합을 예측하도록 상기 모델을 학습하는 것을 특징으로 하는 영상 처리 시스템.

청구항 12

제10 항에 있어서,

상기 모델은

복수 개의 이미지 사이의 상관 텐서를 계산하고, 상기 상관 텐서를 Strided self-attention layer에 통과시켜 분류 맵을 생성하는 것을 특징으로 하는 영상 처리 시스템.

청구항 13

제12 항에 있어서

상기 ASNet은

AS Layer(Attentive Squeeze Layer)를 포함하며,

상기 AS Layer는

고차 자기주의(High-order self-attention) 레이어로 마련되어 상기 상관 텐서를 기반으로 다른 수준의 상관 표현을 반환하는 것을 특징으로 하는 영상 처리 시스템.

청구항 14

제13 항에 있어서,

상기 ASNet은

상기 쿼리 이미지와 서포트 이미지 사이의 피라미드형 교차 상관관계 텐서인 초상관관계(Hypercorrelation)를 입력으로 하는 것을 특징으로 하는 영상 처리 시스템.

청구항 15

제11 항에 있어서,

상기 통합적 퓨-샷 학습에서는

맥스 풀링(Max pooling)을 이용하여 추론을 수행하는 것을 특징으로 하는 영상 처리 시스템.

청구항 16

제11 항에 있어서,

상기 통합적 퓨-샷 학습에서는

분류 손실과 분할 손실을 사용하며, 클래스 태그 또는 분할 주석을 사용하여 학습자를 훈련하는 것을 특징으로 하는 영상 처리 시스템.

청구항 17

제16 항에 있어서,

상기 분류 손실은

공간적으로 평균 풀링된 클래스 점수와 그 정답 클래스 레이블 사이의 평균 이진 교차 엔트로피인 것을 특징으로 하는 영상 처리 시스템.

청구항 18

제16 항에 있어서,

상기 분할 손실은

개별 위치의 클래스 분포와 실제 분할 주석 사이의 평균 교차 엔트로피인 것을 특징으로 하는 영상 처리 시스템.

발명의 설명

기 술 분 야

[0001] 본 발명은 퓨-샷 학습 방법 및 이를 이용한 영상 처리 시스템에 관한 것으로, 보다 상세하게는 쿼리(Query) 이미지의 처리를 수행하는 모델의 퓨-샷 학습 방법 및 이를 이용한 영상 처리 시스템에 관한 것이다.

배 경 기 술

[0002] 일반적으로 퓨-샷 학습은 적은 학습 데이터(Support set)로 쿼리 이미지를 올바르게 예측하기 위한 학습 방법이다. 퓨-샷 학습은 적은 학습 데이터로 높은 성능을 얻을 수 있다는 장점으로 인해 컴퓨터 비전 분야에서 다양한 연구가 수행되고 있다. 이러한 퓨-샷 학습은 쿼리 이미지를 타깃 클래스로 분류하는 것을 목표로 한다. 즉, 퓨-샷 학습은 쿼리 이미지가 어떠한 타깃 클래스에 속하는지를 학습하는 개념보다 어떠한 타켓 클래스와 같은 클래스인지를 학습하는 방법일 수 있다.

[0003] 일례로, 퓨-샷 학습은 퓨-샷 분류 기술 및 퓨-샷 분할 기술로 구분할 수 있다. 퓨-샷 분류 기술은 쿼리 이미지를 타깃 클래스로 분류하는 것을 목표로 한다. 즉, 타깃 클래스에 대해 몇 가지 예제의 서포트 세트(Support set)가 주어질 때 쿼리 이미지를 타깃 클래스로 분류할 수 있다. 그리고 퓨-샷 분할 기술은 쿼리 이미지의 타깃 영역을 타깃 클래스와 유사한 설정으로 분할하는 것을 목표로 한다. 그러나 종래의 퓨-샷 학습은 퓨-샷 분류 기술 및 퓨-샷 분할 기술이 서로 밀접한 관련이 있음에도 불구하고, 지금까지 개별적으로 연구 개발이 수행되고 있다.

[0004] 또한, 퓨-샷 분류 기술 및 퓨-샷 분할 기술은 현실적으로 함께 반영하기 어려운 문제점이 있었다. 퓨-샷 분류 기술은 쿼리가 타깃 클래스 중 하나를 포함한다고 가정한다. 그러나 퓨-샷 분할 기술은 여러 클래스를 허용하지만 분할에서 타깃 클래스가 없는 경우 처리를 수행하지 못하는 문제점이 있었다.

선행기술문헌

특허문헌

[0005] (특허문헌 0001) 대한민국 공개특허공보 제10-2022-0084632호(클러스터링 기능을 구비한 퓨샷 분류 장치 및 이의 메타 학습 방법, 2022.06.21.)

발명의 내용

해결하려는 과제

[0006] 본 발명의 목적은 쿼리 이미지가 주어질 때에, 모델이 쿼리 이미지 내 등장하는 하위 집합을 식별하고 클래스에 해당하는 문제 분할 마스크 집합을 예측하도록 하는 퓨-샷 학습 방법 및 이를 이용한 영상 처리 시스템을 제공하기 위한 것이다.

과제의 해결 수단

[0007] 본 발명에 따른 퓨-샷 학습 방법은 이미지 및 상기 이미지에 대한 분할 예측값(segmentation prediction)을 획득하는 단계 및 기학습된 모델로 쿼리 이미지가 제공될 때에 상기 쿼리 이미지로부터 특정 영역의 분류 및 분할을 동시에 수행하도록 상기 이미지 및 분할 지표를 기반으로 상기 모델을 학습하는 단계를 포함하고, 상기 모델은 ASNet(Attentive Squeeze Network)을 포함하며, 상기 쿼리 이미지에 관련성이 낮은 물체가 존재할 때에 분류를 수행하지 않고 관련성이 높은 물체가 존재할 때에 분류 및 분할을 수행한다.

[0008] 상기 모델의 학습에서는 통합적 퓨-샷 학습(Integrative few-shot learning, iFSL) 방법이 적용되고, 상기 통합적 퓨-샷 학습은 상기 쿼리 이미지 내 등장하는 하위 집합을 식별하고 클래스에 해당하는 문제 분할 마스크 집합을 예측하도록 상기 모델을 학습할 수 있다.

[0009] 상기 모델은 복수 개의 이미지 사이의 상관 텐서를 계산하고, 상기 상관 텐서를 Strided self-attention layer에 통과시켜 분류 맵을 생성할 수 있다.

[0010] 상기 ASNet은 AS Layer(Attentive Squeeze Layer)를 포함하며, 상기 AS Layer는 고차 자기주의(High-order self-attention) 레이어로 마련되어 상기 상관 텐서를 기반으로 다른 수준의 상관 표현을 반환할 수 있다.

[0011] 상기 ASNet은 상기 쿼리 이미지와 서포트 이미지 사이의 피라미드형 교차 상관관계 텐서인 초상관관계(Hypercorrelation)를 입력으로 할 수 있다.

[0012] 상기 통합적 퓨-샷 학습에서는 맥스 풀링(Max pooling)을 이용하여 추론을 수행할 수 있다.

[0013] 상기 통합적 퓨-샷 학습에서는 분류 손실과 분할 손실을 사용하며, 클래스 태그 또는 분할 주석을 사용하여 학습자를 훈련할 수 있다.

[0014] 상기 분류 손실은 공간적으로 평균 풀링된 클래스 점수와 그 정답 클래스 레이블 사이의 평균 이진 교차 엔트로피일 수 있다.

[0015] 상기 분할 손실은 개별 위치의 클래스 분포와 실제 분할 주석 사이의 평균 교차 엔트로피일 수 있다.

[0016] 한편, 본 발명에 따른 영상 처리 시스템은 외부로부터 제공되는 이미지를 기학습된 모델에 입력하여 상기 이미지로부터 특정 영역의 분류 및 분할을 동시에 수행하는 처리모듈을 포함하고, 상기 모델의 학습에서는 이미지 및 상기 이미지에 대한 분할 예측값(segmentation prediction)을 획득하고, 기학습된 모델로 쿼리 이미지가 제공될 때에 상기 쿼리 이미지로부터 특정 영역의 분류 및 분할을 동시에 수행하도록 상기 이미지 및 분할 지표를 기반으로 상기 모델을 학습하며, 상기 모델은 ASNet(Attentive Squeeze Network)을 포함하며, 상기 쿼리 이미지에 관련성이 낮은 물체가 존재할 때에 분류를 수행하지 않고 관련성이 높은 물체가 존재할 때에 분류 및 분할을 수행할 수 있다.

[0017] 상기 모델의 학습에서는 통합적 퓨-샷 학습(Integrative few-shot learning, iFSL) 방법이 적용되고, 상기 통합적 퓨-샷 학습은 상기 쿼리 이미지 내 등장하는 하위 집합을 식별하고 클래스에 해당하는 문제 분할 마스크 집합을 예측하도록 상기 모델을 학습할 수 있다.

[0018] 상기 모델은 복수 개의 이미지 사이의 상관 텐서를 계산하고, 상기 상관 텐서를 Strided self-attention layer에 통과시켜 분류 맵을 생성할 수 있다.

[0019] 상기 ASNet은 AS Layer(Attentive Squeeze Layer)를 포함하며, 상기 AS Layer는 고차 자기주의(High-order self-attention) 레이어로 마련되어 상기 상관 텐서를 기반으로 다른 수준의 상관 표현을 반환할 수 있다.

[0020] 상기 ASNet은 상기 쿼리 이미지와 서포트 이미지 사이의 피라미드형 교차 상관관계 텐서인 초상관관계(Hypercorrelation)를 입력으로 할 수 있다.

[0021] 상기 통합적 퓨-샷 학습에서는 맥스 풀링(Max pooling)을 이용하여 추론을 수행할 수 있다.

[0022] 상기 통합적 퓨-샷 학습에서는 분류 손실과 분할 손실을 사용하며, 클래스 태그 또는 분할 주석을 사용하여 학습자를 훈련할 수 있다.

[0023] 상기 분류 손실은 공간적으로 평균 풀링된 클래스 점수와 그 정답 클래스 레이블 사이의 평균 이진 교차 엔트로피일 수 있다.

[0024] 상기 분할 손실은 개별 위치의 클래스 분포와 실제 분할 주석 사이의 평균 교차 엔트로피일 수 있다.

발명의 효과

[0025] 본 발명에 따른 퓨-샷 학습 방법 및 이를 이용한 영상 처리 시스템은 FS-CS에 효과적이며, iFSL가 약지표 또는 강지표로 학습 가능하기 때문에 확장 가능성이 높은 효과가 있다.

[0026] 이상과 같은 본 발명의 기술적 효과는 이상에서 언급한 효과로 제한되지 않으며, 언급되지 않은 또 다른 기술적 효과들은 아래의 기재로부터 당업자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

[0027] 도 1은 본 실시예에 따른 영상 처리 시스템을 개략적으로 나타낸 개념도이고,

도 2는 본 실시예에 따른 영상 처리 시스템에서 이미지를 처리하는 방법을 나타낸 개념도이고,

도 3은 본 실시예에 따른 통합적 퓨-샷 학습 방법에서의 ASNet을 나타낸 개념도이고,

도 4는 본 실시예에 따른 영상 처리 시스템과 다른 방법들을 FS-CS문제에서 Pascal 데이터세트를 이용하여 비교한 분류 결과이고,

도 5는 본 실시예에 따른 영상 처리 시스템과 다른 방법들을 FS-CS문제에서 Pascal 데이터세트를 이용하여 비교한 분할 결과이고,

도 6은 본 실시예에 따른 영상 처리 시스템의 ASNet의 분할 결과를 나타낸 도면이고,

도 7은 클래스 개수 N을 변경하며 평가한 네가지 방법의 성능을 나타낸 도면이고,

도 8은 문제 환원성을 나타내기 위해 A에서 학습되고 B에서 평가된 모델을 나타낸 도면이다.

발명을 실시하기 위한 구체적인 내용

[0028] 이하 첨부된 도면을 참조하여 본 발명의 실시예를 상세히 설명한다. 그러나 본 실시예는 이하에서 개시되는 실시예에 한정되는 것이 아니라 서로 다양한 형태로 구현될 수 있으며, 단지 본 실시예는 본 발명의 개시가 완전하도록 하며, 통상의 지식을 가진 자에게 발명의 범주를 완전하게 알려주기 위해 제공되는 것이다. 도면에서의 요소의 형상 등은 보다 명확한 설명을 위하여 과장되게 표현된 부분이 있을 수 있으며, 도면 상에서 동일 부호로 표시된 요소는 동일 요소를 의미한다.

[0029] 도 1은 본 실시예에 따른 영상 처리 시스템을 개략적으로 나타낸 개념도이고, 도 2는 본 실시예에 따른 영상 처리 시스템에서 이미지를 처리하는 방법을 나타낸 개념도이다.

[0030] 도 1 및 도 2에 도시된 바와 같이, 본 실시예에 따른 영상 처리 시스템은 극소수 영상 자료를 활용한 영상 분류 및 분할 문제의 통합적 접근 (Integrative Few-Shot Learning for Classification and Segmentation, FS-CS)이 적용된다.

[0031] 일례로, 영상 처리 시스템(100)은 학습 모듈(110) 및 처리 모듈(120)을 포함할 수 있다. 학습모듈(110)은 극소수의 영상(10)을 기반으로 학습하여, 이하 설명될 FS-CS 문제를 해결하기 위한 모델을 구축한다. 그리고 처리모듈(120)은 학습 모델(110)에 의해 학습된 모델을 기반으로 처리용 영상(30)이 제공될 때에 영상 분류 및 분할을 수행할 수 있다.

[0032] 학습 모듈(110)은 이미지 및 상기 이미지에 대한 분할 예측값(Segmentation Prediction)를 획득하여 학습을 수행한다. 이에, 학습 모듈(110)은 퓨-샷 학습을 통해 학습된 모델이 분류와 분할 두 문제를 동시에 해결할 수 있게 한다. 즉, 쿼리 이미지와 서포트 이미지가 주어지면, 학습된 모델은 각 클래스에 해당하는 물체의 존재를 식별하고, 물체 위치의 분할 마스크를 예측하도록 훈련될 수 있다.

[0033] 일례로, 관심있는 N 클래스를 포함하는 타깃 클래스 세트 C와 각 N 타깃 클래스에 대한 K개씩의 예시 이미지를 가정할 때에, 정답 지표 y는 클래스의 존재 유무(Weak label, 약지표) 또는 분할 마스크 정답(Strong label, 강지표)이고, 주어진 지표의 상황에 맞게 선택될 수 있다. 이에, 쿼리 이미지 x가 주어질 때에 모델은 쿼리 이미지 x 내 등장하는 대상 클래스 y의 하위 집합을 식별하는 동시에, 그 클래스에 해당하는 물체 분할 마스크 Y 집합을 예측해야 한다.

[0034] 이에, 학습 모듈(110)은 모델로 쿼리 이미지가 주어질 때에, 모델이 쿼리 이미지 내 등장하는 하위 집합을 식별하고 클래스에 해당하는 문제 분할 마스크 집합을 예측하도록 통합적 퓨-샷 학습(Integrative few-shot learning, iFSL) 방법이 적용된다. 이러한 통합적 퓨-샷 학습 방법은 모델이 적은 수의 이미지로 분류 및 분할을 동시에 수행하도록 한다.

[0035] 일례로, 통합 퓨-샷 학습자 f는 쿼리 이미지 x와 서포트 이미지 S를 입력받고, 클래스별 물체 분할 마스크 Y를 출력한다. 분할 마스크 Y의 집합은 아래 수학식 1과 같이 N개의 클래스에 대해 Y(n) ∈ RH×W로 구성된다.

수학식 1

[0036] $$Y=f(x,S;\theta)\in R^{H\times W\times N}$$

[0037] 여기서, $H^{x}W$는 각 마스크 맵의 크기를 나타내며, θ는 메타 학습을 위한 매개변수이다. 그리고 지도상의 각 위치의 출력은 해당 클래스의 물체 영역에 위치할 확률을 나타낸다.

[0038] 통합적 퓨-샷 학습 방법은 클래스별 존재 유무와 클래스 분할 마스크 모두에 대해 공유 분할 마스크 Y 위에 추

론을 수행한다. 클래스별 발생의 멀티 핫 벡터(Multi-hot vector)는 아래 수학식 2와 같이 예측된다.

수학식 2

[0039] $$\hat{\mathbf{y}}_C^{(n)} = \begin{cases} 1 \text{ if } \max_{\mathbf{p}\in[H]\times[W]} \mathbf{Y}^{(n)}(\mathbf{p}) \geq \delta, \\ 0 \text{ otherwise,} \end{cases}$$

[0040] 여기서, p는 2D 위치를 나타낸다. 그리고 δ는 임계값이며, [k]는 1에서 k까지의 정수 집합, 즉 [k]={1, 2,…, k}을 나타낸다.

[0041] 일반적으로 평균 풀링(Average pooling)을 사용한 추론은 멀티 라벨 분류에서 작은 객체를 놓치기 쉽기 때문에, 통합적 퓨-샷 학습 방법은 맥스 풀링(Max pooling)을 사용하여 추론을 수행한다. 공유 분할 마스크 내의 임의의 위치에서 탐지된 클래스는 클래스의 존재를 나타낸다.

[0042] 한편, 분할 클래스 확률 마스크는 픽셀 클래스 배타적 속성 아래의 클래스별 물체 예측 마스크에서 파생된다. 픽셀은 항상 N개의 물체 클래스와 전경 클래스 사이에서 고유한 클래스로 분류되는 성질을 이용하였다. 전경 클래스는 명시적으로 주어지지 않기 때문에 별도의 전경 예제 클래스가 요구된다. 이에, 통합적 퓨-샷 학습 방법은 전경 클래스 맵을 추정할 때에 n개의 물체 클래스 맵을 평균하여, 아래의 수학식 3 및 4와 같이 에피소딕 전경 맵 Ybg를 계산하고 이를 클래스별 전경 맵과 연결하여 분할 확률 텐서 Ys를 취득한다.

수학식 3

[0043] $$\mathbf{Y}_{\text{bg}} = \frac{1}{N}\sum_{n=1}^{N}(1-\mathbf{Y}^{(n)}),$$

수학식 4

[0044] $$\mathbf{Y}_S = [\mathbf{Y}|\mathbf{Y}_{\text{bg}}] \in \mathbb{R}^{H\times W\times(N+1)}$$

[0045] 최종 분할 마스크 $\hat{\mathbf{Y}}_S \in \mathbb{R}^{H\times W}$ 는 아래의 수학식 5와 같이 확률 분포 중 가장 높은 확률값의 클래스를 선택하여 예측한다.

수학식 5

[0046] $$\hat{\mathbf{Y}}_S = \arg\max_{n\in[N+1]} \mathbf{Y}_S.$$

[0047] 한편, 통합적 퓨-샷 학습 방법에서는 분류 손실과 분할 손실을 사용하며, 클래스 태그 또는 분할 주석을 사용하여 학습자를 훈련할 수 있다.

[0048] 분류 손실은 공간적으로 평균 풀링된 클래스 점수와 그 정답 클래스 레이블 사이의 평균 이진 교차 엔트로피로 아래의 수학식 6과 같이 공식화된다. 여기서, $\mathbf{y}_{\text{gt}}$ 는 멀티 핫 벡터를 나타낸다.

[0049]

수학식 6

[0050] $$\mathcal{L}_C = -\frac{1}{N}\sum_{n=1}^{N} \mathbf{y}_{gt}^{(n)} \log \frac{1}{HW} \sum_{\mathbf{p}\in[H]\times[W]} \mathbf{Y}^{(n)}(\mathbf{p}),$$

[0051] 그리고 분할 손실은 개별 위치의 클래스 분포와 실제 분할 주석 사이의 평균 교차 엔트로피로 아래의 수학식 7과 같이 공식화된다. 여기서, $\mathbf{y}_{gt}$ 는 실측 분할 마스크를 나타낸다.

수학식 7

[0052] $$\mathcal{L}_S = -\frac{1}{(N+1)}\frac{1}{HW}\sum_{n=1}^{N+1}\sum_{\mathbf{p}\in[H]\times[W]} \mathbf{Y}_{gt}^{(n)}(\mathbf{p}) \log \mathbf{Y}_S^{(n)}(\mathbf{p}),$$

[0053] 분류 손실과 분할 손실은 분류라는 유사한 목표를 공유한다. 그러나 각 이미지 또는 각 픽셀을 분류할지 여부에서 차이가 있다. 따라서 학습 목표는 주어진 훈련 감독 수준에 따라 손실 중 하나가 사용될 수 있다. 즉, 약한 레이블을 사용할 수 있는지 강한 레이블을 사용할 수 있는지에 따라 선택될 수 있다.

[0054] 도 3은 본 실시예에 따른 통합적 퓨-샷 학습 방법에서의 ASNet을 나타낸 개념도이다.

[0055] 도 3에 도시된 바와 같이, 본 실시예에 따른 통합적 퓨-샷 학습 방법에서는 ASNet(Attentive Squeeze Network)을 이용한다.

[0056] ASNet(Attentive Squeeze Network)은 쿼리 이미지에 관련성이 없는 물체가 존재할 때에 해당 이미지를 '관련 없음'으로 분류하여 분류를 수행하지 않고 관련성이 높은 물체가 존재할 때에 '관련 있음'으로 분류하여 해당 물체에 대한 분류 및 분할을 수행할 수 있다.

[0057] 이러한 ASNet은 복수의 이미지 사이의 상관 텐서를 계산하고, 계산된 상관 텐서를 Strided self-attention layer에 통과시켜 분류 맵을 생성하는 형태로 마련될 수 있다. 이러한 ASNet의 주요 구성은 AS Layer(Attentive Squeeze Layer)이다. AS Layer는 고차 자기주의(High-order self-attention) 레이어로, 상관 텐서를 기반으로 다른 수준의 상관 표현을 반환한다. 이러한 ASNet은 쿼리 이미지와 서포트 이미지 사이의 피라미드형 교차 상관관계 텐서, 즉 초상관관계(Hypercorrelation)를 입력으로 한다.

[0058] 피라미드 상관관계는 서포터 이미지의 공간 차원을 점진적으로 압축하는 피라미드 AS Layer에 공급되며 피라미드 출력은 상향식 경로를 통해 최종 전경맵에 병합된다.

[0059] 도 2와 같이, N-방향 출력 맵은 병렬로 계산되고 수집되어 수학식 1의 클래스별 전경 맵을 준비하여 통합적 퓨-샷 학습 방법에 적용한다.

[0060] 보다 구체적으로 ASNet의 구조를 살펴보면, ASNet은 쿼리(도2의 적색)와 서포트(도 2의 파란색) 사이의 이미지 특징 맵으로 초상관관계를 구축한다. 여기서, 4D 상관관계는 두 개의 2D 사각형으로 표시되어 있다.

[0061] 이러한 ASNet은 Global Self attention을 통해 각 쿼리 차원에 대한 서포트 차원을 점진적으로 압축하여 상관관계를 전경 맵으로 변환하는 방법을 학습할 수 있다. 다만, 도 2에서는 입력 상관관계(Input correlation), 중간 특징(intermediate feature) 및 출력 전경 맵(Output foreground map)의 채널 차원은 생략하였다.

[0062] 한편, 초상관관계 구조에 대해 살펴보면, ASNet은 쿼리 이미지와 서포트 이미지 간에 초상관관계를 구성하고, 각 서포트 입력에 대해 전경 분할 마스크를 생성하는 방법을 학습할 수 있다.

[0063] 일례로, 입력 초상관관계를 준비하기 위해 에피소드, 즉, 쿼리 이미지와 서포트 이미지의 집합을 쿼리 이미지, 서포트 이미지 및 서포트 레이블의 쌍으로 된 목록으로 열거한다.

[0064] 여기서, 입력 이미지는 CNN(Convolutional Neural Network)의 Stacked Convolutional layer로 공급되고, 중간 및 상위 수준의 출력 특징 맵이 수집되어 특징 피라미드 $\{\mathbf{F}^{(l)}\}_{l=1}^{L}$ 을 구축한다. 여기서, l은 단위 레이어의 인덱스(ResNet50의 병목 레이어)를 나타낸다. 그리고 쿼리 및 서포트 특징 피라미드 쌍에서 특징 맵 간의 코사인

유사도를 계산하여 $H_q^{(l)} \times W_q^{(l)} \times H_s^{(l)} \times W_s^{(l)}$ 사이즈의 4D 상관 텐서를 얻고, ReLU(Rectified Linear Unit)를 아래의 수학식 8과 같이 이용한다.

수학식 8

[0065] $$\mathbf{C}^{(l)}(\mathbf{p}_q, \mathbf{p}_s) = \mathrm{ReLU}\left(\frac{\mathbf{F}_q^{(l)}(\mathbf{p}_q) \cdot \mathbf{F}_s^{(l)}(\mathbf{p}_s)}{||\mathbf{F}_q^{(l)}(\mathbf{p}_q)|| \, ||\mathbf{F}_s^{(l)}(\mathbf{p}_s)||}\right).$$

[0066] 이러한 L 상관 텐서는 동일한 공간 크기의 P 그룹으로 그룹화되고, 각 그룹의 텐서를 새로운 채널 차원을 따라 연결하여 초상관 피라미드를 구현한다.

[0067] $\{\mathbf{C}^{(p)} | \mathbf{C}^{(p)} \in \mathbb{R}^{H_q^{(p)} \times W_q^{(p)} \times H_s^{(p)} \times W_s^{(p)} \times C_{in}^{(p)}}\}_{p=1}^{P}$ 에서 채널 크기 $C_{in}^{(p)}$ 는 P번째 그룹의 연결된 텐서 수에 해당한다. 상관 텐서의 처음 두 공간 $\mathbb{R}^{H_q \times W_q}$ 은 쿼리 차원으로 사용하고, 마지막 두 공간 $\mathbb{R}^{H_s \times W_s}$ 은 서포트 차원으로 사용한다.

[0068] 한편, AS Layer에 대하여 살펴보면, AS Layer는 Strided self-attention를 통해 상관 텐서를 더 작은 서포트 차원을 가진 다른 텐서로 변환한다. 여기서, 텐서는 각 요소가 서포트 패턴을 나타내는 행렬로 다시 캐스팅될 수 있다. 일례로, 초상관 피라미드에서 상관 텐서 $\mathbf{C}(\mathbf{x}_q) \in \mathbb{R}^{H_s \times W_s \times C_{in}}$ 이 주어지면, 상관 텐서를 $H_q \times W_q$ 사이즈의 블록 행렬로 재구성하고, 각 요소는 쿼리 위치 $\mathbf{x}_q$ 에서 $\mathbf{C}(\mathbf{x}_q) \in \mathbb{R}^{H_s \times W_s \times C_{in}}$ 의 상관 텐서에 대응하는 수학식 9와 같은 크기의 블록 행렬로 재구성한다.

수학식 9

[0069] $$\mathbf{C}^{\mathrm{block}} = \begin{bmatrix} \mathbf{C}((1,1)) & \cdots & \mathbf{C}((1,W_q)) \\ \vdots & \ddots & \vdots \\ \mathbf{C}((H_q,1)) & \cdots & \mathbf{C}((H_q,W_q)) \end{bmatrix}.$$

[0070] AS Layer의 목표는 각 서포트 지원 텐서의 Global context를 분석하고, 쿼리 차원은 유지하면서 서포트 지원 차원의 줄인 상관관계 표현을 추출하는 것이다($\mathbb{R}^{H_q \times W_q \times H_s \times W_s \times C_{in}}$ -> $\mathbb{R}^{H_q \times W_q \times H_s' \times W_s' \times C_{out}}$). 여기서, H_s' 는 H_s 보다 작거나 같고, W_s' 는 W_s 보다 작거나 같다.

[0071] 그리고 AS Layer는 서포트 상관관계의 전체적 패턴을 학습하기 위해 상관관계에 대한 Global Self attention 매커니즘을 적용한다. 여기서, Self attention 가중치는 모든 쿼리 위치에서 공유되며 병렬로 처리될 수 있다.

[0072] 이때, 모든 위치가 아래의 계산을 공유함으로 모든 쿼리 위치 $\mathbf{x}_q$ 에 대한 서포트 상관 텐서를 $\mathbf{C}^s = \mathbf{C}^{\mathrm{block}}(\mathbf{x}_q)$ 로 표시할 수 있다. Self attention 연산에서는 서포트 상관 텐서 $\mathbf{C}^s$ 를 Target, key, value triplet에 $\mathbf{T}, \mathbf{K}, \mathbf{V} \in \mathbb{R}^{H_s' \times W_s' \times C_{hd}}$ 와 같이 임베딩한다. 여기서, strides가 1보다 크거나 같은 3개의 컨볼루션을 사용하여 출력 크기를 제어할 수 있다.

[0073] 이후, 결과물인 타깃과 주요 상관관계 표현인 $\mathbf{T}$ 및 $\mathbf{K}$ 를 이용하여 attention context를 계산한다. attention context는 아래의 수학식 10의 행렬 곱셈으로 계산될 수 있다.

수학식 10

[0074] $$\mathbf{A} = \mathbf{T}\mathbf{K}^{\top} \in \mathbb{R}^{H'_s \times W'_s \times H'_s \times W'_s}.$$

[0075] 이후, 주요 전경 위치의 선출이 전경 영역에 더 많이 유도할 수 있는 경우에 서포트 마스크 주석 $\mathbf{Y}_s$에 의해 주목도를 마스킹하는 1로 합산되도록 소프트맥스에 의해 주목도 context가 아래의 수학식 11과 같이 정규화된다.

수학식 11

[0076] $$\bar{\mathbf{A}}(\mathbf{p}_t, \mathbf{p}_k) = \frac{\exp\left(\mathbf{A}(\mathbf{p}_t, \mathbf{p}_k)\mathbf{Y}_s(\mathbf{p}_k)\right)}{\sum_{\mathbf{p}'_k} \exp\left(\mathbf{A}(\mathbf{p}_t, \mathbf{p}'_k)\mathbf{Y}_s(\mathbf{p}'_k)\right)},$$

$$\mathbf{Y}_s(\mathbf{p}_k) = \begin{cases} 1 & \text{if } \mathbf{p}_k \in [H'_s] \times [W'_s] \text{ is foreground,} \\ -\infty & \text{otherwise.} \end{cases}$$

[0077] 그리고 마스킹된 attention context $\bar{\mathbf{A}}$를 이용하여 V를 포함하는 값을 아래의 수학식 12와 같이 집계할 수 있다.

수학식 12

[0078] $$\mathbf{C}^s_A = \bar{\mathbf{A}}\mathbf{V} \in \mathbb{R}^{H'_s \times W'_s \times C_{hd}}.$$

[0079] 여기서, 어텐딩된 표현은 MLP 레이어 $\mathbf{W}_0$에 공급되고 입력에 추가된다. 입력과 출력 차원이 일치하지 않는 경우에 입력은 선택적으로 컨볼루션 레이어 $\mathbf{W}_1$에 공급된다. 그리고 그룹 정규화 및 ReLU 활성화로 구성된 활성화 레이어 $\varphi(\cdot)$가 수학식 13과 같이 추가된다.

수학식 13

[0080] $$\mathbf{C}^s_0 = \varphi(\mathbf{W}_0(\mathbf{C}^s_A) + \mathbf{W}_1(\mathbf{C}^s)) \in \mathbb{R}^{H'_s \times W'_s \times C_{out}}.$$

[0081] 출력은 AS Layer의 단위 작업을 마무리하는 다른 MLP로 수학식 14와 같이 공급된다.

수학식 14

[0082] $$\mathbf{C}^{s\prime} = \varphi(\mathbf{W}_{FF}(\mathbf{C}^s_0) + \mathbf{C}^s_0) \in \mathbb{R}^{H'_s \times W'_s \times C_{out}},$$

[0083] 여기서, 수학식 9의 블록 행렬에서 해당 위치를 포함시킬수 있다. AS Layer은 스택을 쌓아 서포트 상관 텐서 $H'_s \times W'_s$의 크기를 점진적으로 줄일 수 있다.

[0084] 한편, ASNet의 멀티 레이어 융합(Multi layer fusion)을 살펴보면, 피라미드 상관관계 표현은 쌍 단위 연산을 계산식으로 수행하며 가장 거친 수준에서 가장 세밀한 수준으로 병합될 수 있다. 먼저, 최하위 상관관계 표현을 인접한 이전 쿼리 공간 차원으로 2선형 업샘플링하고 두 표현을 더하여 혼합된 표현 $\mathbf{C}^{mix}$를 획득할 수 있다. 그리고 혼합된 표현은 크기가 $H'_s = W'_s = 1$의 포인트 특징이 될 때까지 두 개의 순차적인 AS Layer에 공급되고 피라미드 융합에 공급된다. 여기서, 가장 빠른 융합 레이어의 출력은 컨볼루션 디코더로 공급되며, 인터리브 2D

컨볼루션과 2선형 업샘플링으로 구성되어 C-차원 채널을 전경과 배경으로 매핑하고 출력 공간 크기를 입력 쿼리 이미지 크기로 매핑한다.

[0085] 그리고 클래스별 전경 맵 계산을 살펴보면, K-샷 출력 전경 활성화 맵의 평균을 계산하여 각 클래스에 대한 마스크 예측을 생성할 수 있다. 평균화된 출력 맵은 바이너리 세분화 맵의 두 채널에 대해 소프트맥스로 정규화된다. 그리고 세분화 맵의 두 채널에 대한 소프트맥스로 정규화로 전경 확률 예측 $\mathbf{Y}^{(n)} \in \mathbb{R}^{H \times W}$ 을 획득할 수 있다.

[0086] 이에, 학습 모듈은 모델을 학습시켜 학습된 모델이 쿼리 이미지가 주어질 때에, 모델이 쿼리 이미지 내 등장하는 하위 집합을 식별하고 클래스에 해당하는 문제 분할 마스크 집합을 예측하도록 한다.

[0087] 이러한 모델은 추후 처리 모듈에 탑재될 수 있으며, 처리 모듈은 처리용 이미지가 제공될 때에 모델이 쿼리 이미지 내 등장하는 하위 집합을 식별하고 클래스에 해당하는 문제 분할 마스크 집합을 예측하여, 특정 영역의 분류 및 분할을 수행할 수 있다.

[0088] 한편, 이하에서는 본 실시예에 따른 퓨-샷 학습 방법을 이용한 모델 실험에 대하여 상세히 설명하도록 한다.

[0089] 도 4는 본 실시예에 따른 영상 처리 시스템과 다른 방법들을 FS-CS문제에서 Pascal 데이터세트를 이용하여 비교한 분류 결과이고, 도 5는 본 실시예에 따른 영상 처리 시스템과 다른 방법들을 FS-CS문제에서 Pascal 데이터세트를 이용하여 비교한 분할 결과이다.

[0090] 도 4 및 도 5에 도시된 바와 같이, 본 실시예에 따른 모델 실험에서는 실험에서는 FS-CS 문제에 대한 통합적 퓨-샷 학습 방법에 대한 iFSL 프레임워크를 평가할 수 있다. 실험에서는 Pascal와 Rsenet을 사용하여 수행되었으며, 달리 지정되지 않는 한 1-way 1-shot 설정으로 평가되었다.

[0091] 실험에서는 FS-CS에 대한 iFSL 프레임워크를 검증하며, 제안된 모델의 성능을 3가지 모델(PASNet, PFENet 및 HSNet)의 성능과 비교하였다. 3가지 모델들은 기존의 FS-S작업에 대해 제안되였으며, 모든 모델은 공정한 비교를 위해 iFSL에서 훈련되었다.

[0092] FS-CS문제에서 iFSL 프레임워크를 정량적으로 검증하며, 여기서 제안된 방식은 분할 성능뿐만 아니라 퓨샷 분류 측면에서 다른 방법을 능가한다는 것을 알 수 있다. 도 4는 영상 처리시스템(100)과 다른 방법들을 FS-CS문제에서 Pascal-5 데이터세트를 이용하여 비교한 분류 결과이다. 이때, 모든 방법은 iFSL의 강지표 학습법을 통하여 학습되었고, 제안방법-약지도 방법만 약지도 방법으로 학습되었다. 그리고 도 5는 영상 처리시스템(100)과 다른 방법들을 FS-CS문제에서 Pascal-5 실험환경을 이용하여 비교한 분할 결과이다.

[0093] 도 6은 본 실시예에 따른 영상 처리 시스템의 ASNet의 분할 결과를 나타낸 도면이다. 그리고 도 7은 클래스 개수 N을 변경하며 평가한 네가지 방법의 성능을 나타낸 도면이고, 도 8은 문제 환원성을 나타내기 위해 A에서 학습되고 B에서 평가된 모델을 나타낸 도면이다.

[0094] 도 6 내지 도 8에 나타낸 분할 성능과 같이, iFSL 프레임워크는 쿼리 영상의 물체들을 예서 영상을 참고하여 적절하게 분할하는 것을 알 수 있다.

[0095] 한편, FS-CS는 임의의 클래스 수로 다중 클래스 문제로 확장될 수 있다. 도 5은 예시 영상의 클래스 수를 1에서 5까지 비교시킴으로써 네가지 방법의 FS-CS성능을 비교하였다. 도 7에서 제안된 방식은 클래스 수가 다양하더라더 다른 방법보다 지속적으로 더 나은 성능을 보여준다는 것을 알 수 있다.

[0096] 한편, FS-CS, FS-C 및 FS-S 문제간의 환원성을 평가하면 FS-CS가 기존의 두 문제를 포괄 및 일반화함을 알 수 있다. 도 8과 같이, FS-S->FS-CS는 FS-S문제에서 학습된 모델이 FS-CS설정에서 평가되는 결과를 알 수 있다. 본 실험에서는 FS-C 또는 FS-S에 대한 문제 환경을 구성하기 위해 예 클래스 발생의 제약을 충족하는 자료만 선별하여 평가하였다. FS-C의 경우 클래스 정보(약지표)를 사용하였다. 다른 모든 설정은 동일하게 Pascal-i 자료집에서 ResNet50을 사용하였다.

[0097] 결과는 FS-CS로 학습한 모델들, 즉 FS-CS에 대해 학습되고, 기존의 FS-C로 환원 가능한 동기에 기존의 FS-C의 단점을 극복하는 것을 알 수 있다. 퓨샷 분류 작업, 즉 FS-C와 FS-CS ‴간의 환원성은 도 8a에 제시되어 있다. 이 설정에서 FS-CS ‴ 모델은 다중 지표 분류에 대해 훈련되지만 두 클래스 사이의 더 높은 클래스 응답을 예측하여 평가된다. FS-CS 모델은 분류 정확도 측면에서 FS-C 모델과 밀접하게 경쟁한다. 대조적으로 분할 작업인 FS-S와 FS-CS사이의 환원성은 도 8b와 c에 표시된 것과 같은 비대칭 결과를 초래하는 것을 알 수 있다. FS-CS모

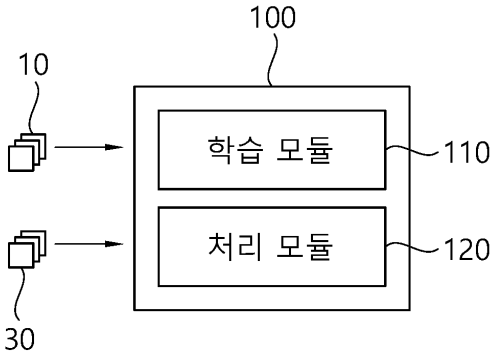
델은 FS-S에서 상대적으로 작은 성능 저하를 보여주는 것을 알 수 있다. 그러나 FS-S학습자는 심각한 성능 저하를 보여준다. 질적인 결과에 따르면 FS-S모델은 거짓 양성자를 예측하며, 대조적은 FS-CS모델은 물체가 예시 영상과 클래스 관련성까지 같이 파악하기에 성공적으로 물체를 분할할 수 있다.

[0098] 이에, 본 발명에 따른 통합적 퓨샷 학습 방법 및 이를 이용한 영상 처리시스템은 FS-CS에 효과적이며, iFSL가 약지표 또는 강지표로 학습 가능하기 때문에 확장 가능성이 높은 효과가 있다.

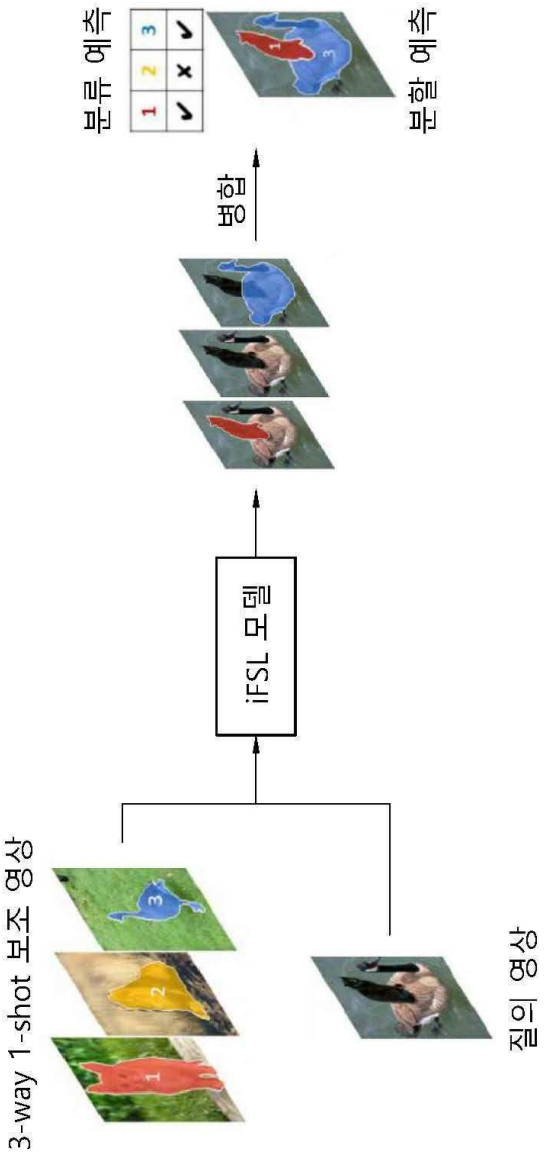
[0099] 앞에서 설명되고, 도면에 도시된 본 발명의 일 실시예는 본 발명의 기술적 사상을 한정하는 것으로 해석되어서는 안 된다. 본 발명의 보호범위는 청구범위에 기재된 사항에 의하여만 제한되고, 본 발명의 기술분야에서 통상의 지식을 가진 자는 본 발명의 기술적 사상을 다양한 형태로 개량 변경하는 것이 가능하다. 따라서 이러한 개량 및 변경은 통상의 지식을 가진 자에게 자명한 것인 한 본 발명의 보호범위에 속하게 될 것이다.

도면

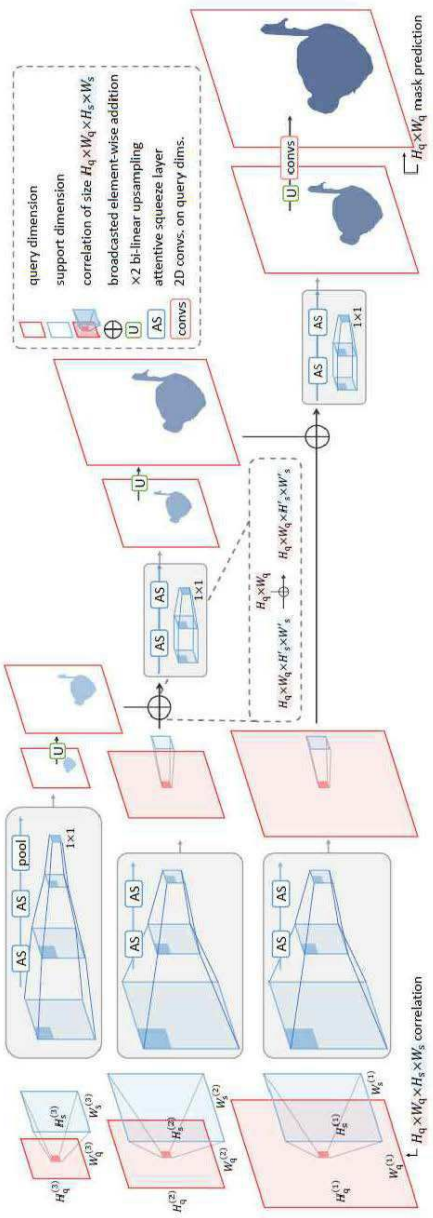
도면1



도면2



도면3


query dimension
support dimension
correlation of size $H_q \times W_q \times H_s \times W_s$
broadcasted element-wise addition
×2 bi-linear upsampling
attentive squeeze layer
2D convs. on query dims.
convs
AS
U
pool
1×1
$H_q \times W_q$ mask prediction
$H_q \times W_q \times H_s \times W_s$ correlation

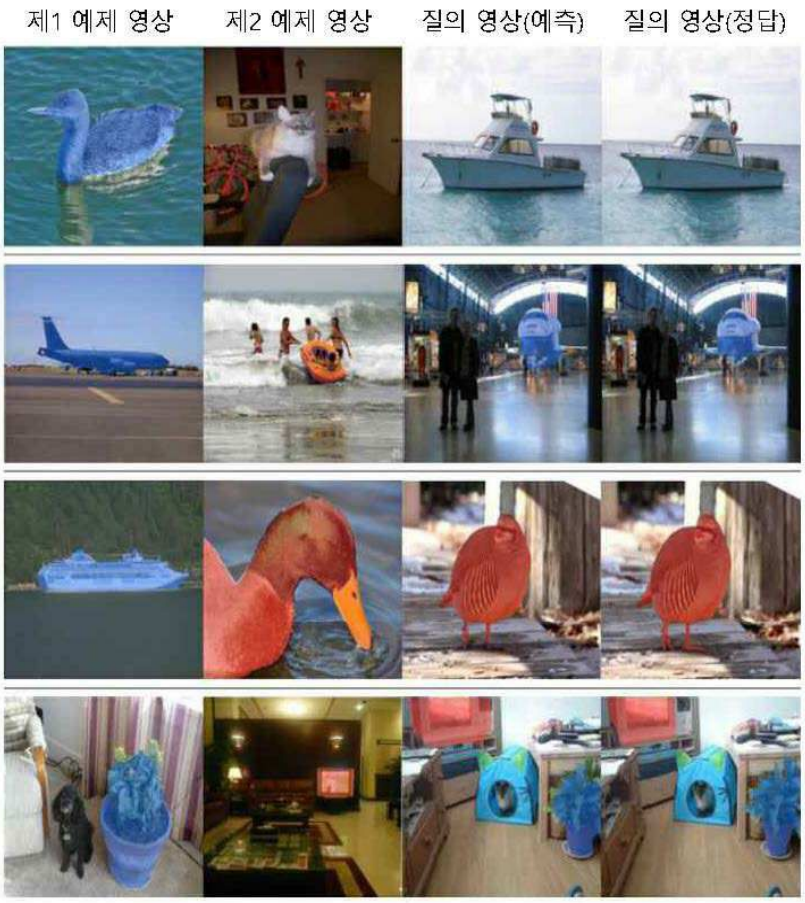
도면4

	1-way 1-shot				
	fold0	fold1	fold2	fold3	avg.
PANet	69.9	67.7	68.8	69.4	69.0
PFENet	69.8	82.4	68.1	77.9	74.6
HSNet	86.6	84.8	76.9	86.3	83.7
제안 방법 약지도	84.2	83.8	69.9	82.3	80.1
제안 방법	85.7	87.5	76.2	87.8	84.3

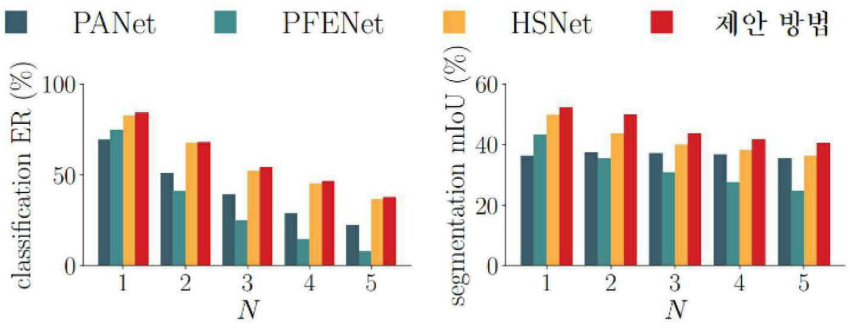
도면5

	1-way 1-shot				
	fold0	fold1	fold2	fold3	avg.
PANet	32.8	45.8	31.0	35.1	36.2
PFENet	38.3	54.7	35.1	43.8	43.0
HSNet	49.1	59.7	41.0	49.08	49.7
제안 방법 약지도	11.5	20.7	12.9	16.4	15.4
제안 방법	51.7	61.2	42.5	53.7	52.3

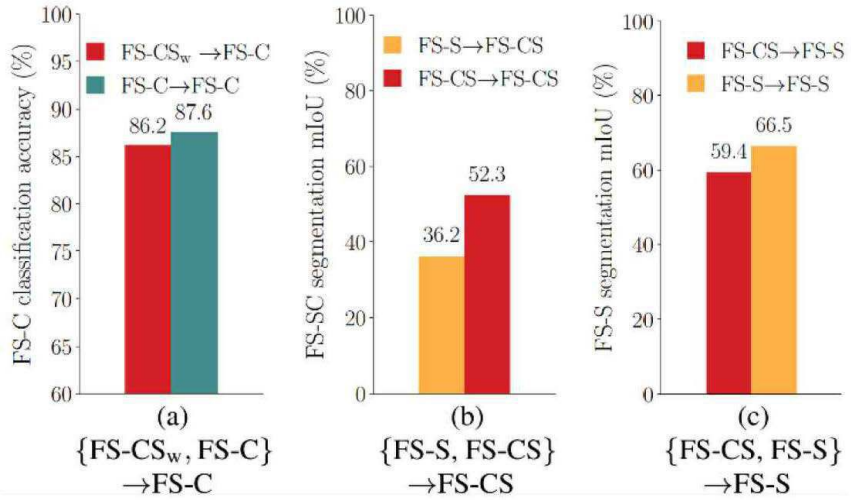
도면6


제1 예제 영상
제2 예제 영상
질의 영상(예측)
질의 영상(정답)

도면7


PANet
PFENet
HSNet
제안 방법
classification ER (%)
100
50
0
1
2
3
4
5
N
segmentation mIoU (%)
60
40
20
0
1
2
3
4
5
N

도면8


FS-C classification accuracy (%)
FS-CSw →FS-C
FS-C→FS-C
86.2
87.6
(a)
{FS-CSw, FS-C}
→FS-C
FS-SC segmentation mIoU (%)
FS-S→FS-CS
FS-CS→FS-CS
36.2
52.3
(b)
{FS-S, FS-CS}
→FS-CS
FS-S segmentation mIoU (%)
FS-CS→FS-S
FS-S→FS-S
59.4
66.5
(c)
{FS-CS, FS-S}
→FS-S