



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0082284  
(43) 공개일자 2022년06월17일

(51) 국제특허분류(Int. Cl.)  
G06N 3/04 (2006.01) G06F 16/732 (2019.01)  
G06F 16/738 (2019.01) G06F 16/74 (2019.01)  
G06F 16/75 (2019.01) G06F 16/78 (2019.01)  
G06N 3/08 (2006.01) G06N 5/02 (2006.01)  
(52) CPC특허분류  
G06N 3/0454 (2013.01)  
G06F 16/732 (2019.01)  
(21) 출원번호 10-2020-0172035  
(22) 출원일자 2020년12월10일  
심사청구일자 2020년12월10일

(71) 출원인  
포항공과대학교 산학협력단  
경상북도 포항시 남구 청암로 77 (지곡동)  
(72) 발명자  
조민수  
경상북도 포항시 남구 청암로 77, 포항공과대학교  
컴퓨터공학과 (지곡동)  
정든솔  
경상북도 포항시 남구 청암로 77, 포항공과대학교  
남자기숙사 10동 211호 (지곡동)  
강다현  
전라북도 전주시 덕진구 사근1길 11, 307동 707호  
(송천동2가, 현대3차아파트)  
(74) 대리인  
특허법인(유한)아이시스

전체 청구항 수 : 총 15 항

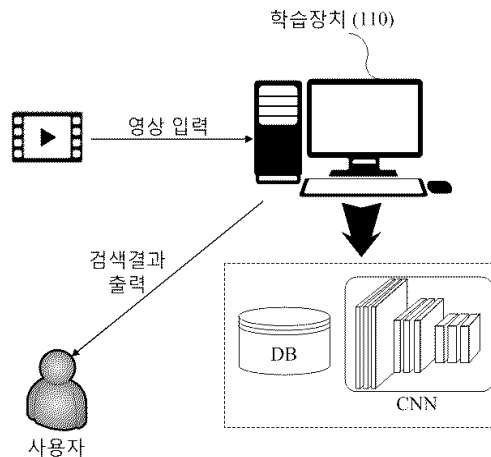
(54) 발명의 명칭 **퓨-샷을 이용한 신경망 모델 구축 방법 및 장치**

**(57) 요약**

개시된 기술은 퓨-샷을 이용한 신경망 모델 구축 방법 및 장치에 관한 것으로, 학습장치가 제 1 훈련용 데이터셋의 제 1 메타데이터셋을 이용하여 신경망 모델을 1차 훈련하는 단계; 및 상기 학습장치가 제 2 훈련용 데이터셋의 제 2 메타데이터셋을 이용하여 상기 1차 훈련된 신경망 모델에 대한 파라미터를 조정하는 단계;를 포함하되, 상기 제 1 훈련용 데이터셋은 상기 신경망 모델을 이용하여 검색하고자 하는 클래스의 동일 여부와 관계 없는 복수의 영상들이고, 상기 제 2 훈련용 데이터셋은 상기 신경망 모델을 이용하여 검색하고자 하는 클래스와 동일한 복수의 영상들이고, 상기 제 1 메타데이터셋 및 상기 제 2 메타데이터셋은 상기 복수개의 영상들로부터 각각 추출한 특징값의 거리정보를 포함한다. 따라서 적은 수의 데이터를 이용하여 검색하고자 하는 클래스에서의 영상 검색 성능을 향상시키는 효과가 있다.

**대표도** - 도1

**100**



(52) CPC특허분류

*G06F 16/738* (2019.01)

*G06F 16/74* (2019.01)

*G06F 16/75* (2019.01)

*G06F 16/78* (2019.01)

*G06N 3/08* (2013.01)

*G06N 5/022* (2019.01)

이 발명을 지원한 국가연구개발사업

|             |                           |
|-------------|---------------------------|
| 과제고유번호      | 1711116284                |
| 과제번호        | 2017M3C4A7069369          |
| 부처명         | 과학기술정보통신부                 |
| 과제관리(전문)기관명 | 한국연구재단                    |
| 연구사업명       | 차세대정보·컴퓨팅기술개발(R&D)        |
| 연구과제명       | 비전 모델 기반 공간 상황 인지 원천기술 연구 |
| 기여율         | 1/1                       |
| 과제수행기관명     | 한양대학교                     |
| 연구기간        | 2020.04.01 ~ 2020.12.31   |

---

## 명세서

### 청구범위

#### 청구항 1

학습장치가 제 1 훈련용 데이터셋의 제 1 메타데이터셋을 이용하여 신경망 모델을 1차 훈련하는 단계; 및  
상기 학습장치가 제 2 훈련용 데이터셋의 제 2 메타데이터셋을 이용하여 상기 1차 훈련된 신경망 모델에 대한 파라미터를 조정하는 단계;를 포함하되,

상기 제 1 훈련용 데이터셋은 상기 신경망 모델을 이용하여 검색하고자 하는 클래스의 동일 여부와 관계 없는 복수의 영상들이고, 상기 제 2 훈련용 데이터셋은 상기 신경망 모델을 이용하여 검색하고자 하는 클래스와 동일한 복수의 영상들이고,

상기 제 1 메타데이터셋 및 상기 제 2 메타데이터셋은 상기 복수개의 영상들로부터 각각 추출한 특징값의 거리 정보를 포함하는 퓨-샷을 이용한 신경망 모델 구축 방법.

#### 청구항 2

제 1 항에 있어서, 상기 파라미터를 조정하는 단계는,

상기 학습장치가 상기 제 2 메타데이터셋에 포함된 거리정보를 이용하여 상기 신경망 모델의 거리함수를 조절하는 퓨-샷을 이용한 신경망 모델 구축 방법.

#### 청구항 3

제 1 항에 있어서,

상기 학습장치는 상기 신경망 모델이 상기 제 1 메타데이터셋을 학습한 결과와 상기 제 2 메타데이터셋을 학습한 결과가 같아지도록 상기 거리함수의 조절을 반복하는 퓨-샷을 이용한 신경망 모델 구축 방법.

#### 청구항 4

제 2 항에 있어서,

상기 제 2 메타데이터셋은 상기 훈련용 데이터셋에서 추출한 서포트 데이터셋 및 쿼리 데이터셋을 포함하고,

상기 학습장치는 상기 서포트 데이터셋을 이용하여 상기 신경망 모델의 거리함수를 미세조정하고 상기 쿼리 데이터셋을 이용하여 상기 거리함수의 손실값을 최소화하는 퓨-샷을 이용한 신경망 모델 구축 방법.

#### 청구항 5

제 4 항에 있어서,

상기 서포트 데이터셋은 상기 신경망 모델의 학습을 위한 기준치를 초과하는 레이블 개수를 포함하고 상기 쿼리 데이터셋은 상기 기준치 미만의 레이블 개수를 포함하는 퓨-샷을 이용한 신경망 모델 구축 방법.

#### 청구항 6

제 1 항에 있어서, 상기 신경망 모델 구축 방법은,

상기 거리함수를 조절하는 단계 이후에 상기 신경망 모델의 검색 성능을 평가하는 단계를 더 포함하고,

상기 검색 성능을 평가하는 단계는, 상기 신경망 모델에 상기 검색하고자 하는 영상을 입력하고 데이터베이스에 저장된 복수개의 영상들 중 상기 검색하고자 하는 영상의 클래스와 동일한 클래스를 갖는 일부의 영상들을 출력하고 상기 출력된 일부의 영상들에 대한 피드백에 따라 검색 정확도를 판단하는 퓨-샷을 이용한 신경망 모델 구축 방법.

#### 청구항 7

제 6 항에 있어서,

상기 학습장치는 상기 데이터베이스에 저장된 복수개의 영상들 중 상기 검색하고자 하는 영상의 특징점과 거리가 가까운 순서대로 상기 일부의 영상들을 출력하는 퓨-샷을 이용한 신경망 모델 구축 방법.

#### 청구항 8

제 1 항에 있어서,

상기 학습장치는 상기 데이터베이스에 저장된 복수의 영상들 중 상기 검색하고자 하는 영상과 동일한 클래스의 영상은 거리가 가까운 것으로 분류하고 서로 다른 클래스의 영상은 거리가 먼 것으로 분류하는 퓨-샷을 이용한 신경망 모델 구축 방법.

#### 청구항 9

검색하고자 하는 영상을 입력받는 입력장치;

복수개의 영상 및 상기 복수개의 영상 각각에 대한 특징점의 거리정보를 포함하는 복수개의 메타데이터를 포함하는 훈련용 데이터셋 및 상기 검색하고자 하는 영상에 대한 결과를 출력하는 신경망 모델을 저장하는 저장장치; 및

상기 훈련용 데이터셋에서 추출한 제 1 메타데이터를 상기 신경망 모델에 입력하여 상기 신경망 모델을 훈련하고 상기 훈련된 신경망 모델에 상기 훈련용 데이터셋에서 추출한 제 2 메타데이터를 입력하여 상기 신경망 모델의 거리함수를 조정하는 연산장치;를 포함하는 퓨-샷을 이용한 신경망 모델 구축 장치.

#### 청구항 10

제 9 항에 있어서,

상기 연산장치는 상기 신경망 모델이 상기 제 1 메타데이터셋을 학습한 결과와 상기 제 2 메타데이터셋을 학습한 결과가 같아지도록 상기 거리함수의 조정을 반복하는 퓨-샷을 이용한 신경망 모델 구축 장치.

#### 청구항 11

제 10 항에 있어서,

상기 제 2 메타데이터셋은 상기 훈련용 데이터셋에서 추출한 서포트 데이터셋 및 쿼리 데이터셋을 포함하고,

상기 연산장치는 상기 서포트 데이터셋을 이용하여 상기 신경망 모델의 거리함수를 미세조정하고 상기 쿼리 데이터셋을 이용하여 상기 거리함수의 손실값을 최소화하는 퓨-샷을 이용한 신경망 모델 구축 장치.

#### 청구항 12

제 11 항에 있어서,

상기 서포트 데이터셋은 상기 신경망 모델의 학습을 위한 기준치를 초과하는 레이블 개수를 포함하고 상기 쿼리 데이터셋은 상기 기준치 미만의 레이블 개수를 포함하는 퓨-샷을 이용한 신경망 모델 구축 장치.

#### 청구항 13

제 9 항에 있어서, 상기 신경망 모델 구축 장치는,

상기 검색하고자 하는 영상에 대한 검색 결과를 출력하는 출력장치를 더 포함하고,

상기 연산장치는 상기 저장장치에 저장된 복수개의 영상들 중 상기 제 2 메타데이터셋과 클래스가 동일한 일부의 영상들을 상기 출력장치를 통해 출력하는 퓨-샷을 이용한 신경망 모델 구축 장치.

#### 청구항 14

제 13 항에 있어서,

상기 연산장치는 상기 복수개의 영상들 중 상기 검색하고자 하는 영상의 특징점과 거리가 가까운 순서대로 상기 일부의 영상들을 출력하는 퓨-샷을 이용한 신경망 모델 구축 장치.

**청구항 15**

제 9 항에 있어서,

상기 연산장치는 상기 저장장치에 저장된 복수의 영상들 중 상기 검색하고자 하는 영상과 동일한 클래스의 영상은 거리가 가까운 것으로 분류하고 서로 다른 클래스의 영상은 거리가 먼 것으로 분류하는 퓨-샷을 이용한 신경망 모델 구축 장치.

**발명의 설명**

**기술 분야**

[0001] 개시된 기술은 적은 수의 데이터(Few-shot)를 이용하여 신경망 모델을 적응시켜서 영상 검색 성능을 높이는 방법 및 장치에 관한 것이다.

**배경 기술**

[0002] 영상 검색은 이미지 인식과 마찬가지로 다양한 산업 분야에 응용 가능하다. 최근 인터넷의 검색엔진에서 손쉽게 찾아 볼 수 있는 기술이다. 이 뿐만 아니라, 의료 영상 진단, 자율 주행 등의 분야에 응용될 수 있다. 미리 정해진 클래스(class)를 알고 있어야 하는 영상 인식과는 다르게 클래스와 상관 없이 주어진 쿼리 영상과 비슷한 영상을 찾는 기술이므로 범용성이 높다.

[0003] 일반적인 영상 검색 기술은 각 영상의 특징점(feature)을 추출해 특징점 간의 유클리드 거리를 측정하여 영상의 비슷한 정도를 계산한다. 전통적으로는 사람이 직접 설계한 특징점 추출 방식인 SIFT(Scale-Invariant Feature Transform)나 HOG(Histogram of Oriented Gradient) 등을 사용한다.

[0004] 신경망 기술의 발전에 따라, 컨볼루션 신경망(Convolutional Neural Network, CNN)을 이용해 영상의 특징점을 학습하는 방식이 사용되고 있다. CNN은 같은 클래스의 영상들의 특징점 간 거리는 가까워지도록 하고, 다른 클래스를 가진 영상들의 특징점 간 거리는 멀어지도록 하는 거리 학습(Metric learning)으로 학습한다. 이 때 두 영상의 거리만을 고려하는 방법, 세 영상의 거리를 복합적으로 고려하는 방법, 그것보다 더 많은 영상들의 거리를 복합적으로 고려하는 방법 등이 있다. 또한, 영상들 간의 절대 거리뿐만 아니라 상대적인 거리를 고려하여 학습하는 방법 등이 일반적으로 활용되고 있다. 거리 학습으로 학습된 깊은 신경망을 통해 쿼리 영상과 데이터 베이스 영상들의 특징점을 추출해 유클리드 거리를 측정하고, 이 거리가 가까운 영상이 검색 결과가 된다.

[0005] 그러나 이러한 기법들은 일반화에 방점을 두어, 학습할 때 사용한 영상들의 클래스와 실제로 검색하고자 하는 영상의 클래스가 상이하다는 문제점이 있다. 이에 따라, 학습할 때 사용한 영상들의 클래스와 실제로 검색하고자 하는 영상의 클래스의 차이가 크다면 일반화가 잘 되지 않고 검색에 실패하게 된다는 한계가 있다. 또한, 실제로 검색하고자 하는 영상을 거리 학습에 사용하기 위해 많은 학습 데이터를 만들어 학습에 사용하면 검색의 성능이 향상되지만, 어떤 영상을 검색하고 싶은지 미리 알기 어려우며 많은 학습 데이터를 만드는 것은 어렵고 노동집약적인 일이다.

[0006] 한편, 영상 인식의 경우, 인식하고자 하는 클래스의 영상의 레이블(Label)이 적을 경우, 이 적은 레이블에 과적합(Over fitting)하는 문제를 방지하면서 좋은 인식 성능을 내는 퓨-샷 학습 기법이 제안되고 있다. 퓨-샷 학습은 레이블이 충분히 많은 모종의 데이터베이스를 이용해 메타 학습을 진행한다. 일반적인 메타 학습에서는 적은 수의 레이블을 이용해 미세 조정(Fine-tuning)하고 미세 조정에 사용하지 않은 영상을 잘 인식하는지를 평가하기 위해 에피소드를 사용해 학습한다. 이 에피소드를 이용해 네트워크 변수의 초기값(Initialization)이나 학습율(Learning rate)을 학습한다. 최근에는 코사인 유사도 분류기(Cosine-similarity classifier)를 이용해 레이블이 많은 데이터베이스에서 사전 학습한 이후, 적은 레이블을 이용해 미세 적응하는 방식을 사용하는 방식이 제안되었다.

**선행기술문헌**

**특허문헌**

[0007] (특허문헌 0001) 한국 등록특허 제10-2171178호

**발명의 내용**

**해결하려는 과제**

[0008] 개시된 기술은 적은 수의 데이터(Few-shot)를 이용하여 신경망 모델을 적응시켜서 영상 검색 성능을 높이는 방법 및 장치를 제공하는데 있다.

**과제의 해결 수단**

[0009] 상기의 기술적 과제를 이루기 위하여 개시된 기술의 제 1 측면은 학습장치가 제 1 훈련용 데이터셋의 제 1 메타 데이터셋을 이용하여 신경망 모델을 1차 훈련하는 단계 및 상기 학습장치가 제 2 훈련용 데이터셋의 제 2 메타 데이터셋을 이용하여 상기 1차 훈련된 신경망 모델에 대한 파라미터를 조정하는 단계를 포함하되, 상기 제 1 훈련용 데이터셋은 상기 신경망 모델을 이용하여 검색하고자 하는 클래스의 동일 여부와 관계 없는 복수의 영상들이고, 상기 제 2 훈련용 데이터셋은 상기 신경망 모델을 이용하여 검색하고자 하는 클래스와 동일한 복수의 영상들이고, 상기 제 1 메타데이터셋 및 상기 제 2 메타데이터셋은 상기 복수개의 영상들로부터 각각 추출한 특징값의 거리정보를 포함하는 퓨-샷을 이용한 신경망 모델 구축 방법을 제공하는데 있다.

[0010] 상기의 기술적 과제를 이루기 위하여 개시된 기술의 제 2 측면은 검색하고자 하는 영상을 입력받는 입력장치, 복수개의 영상 및 상기 복수개의 영상 각각에 대한 특징점의 거리정보를 포함하는 복수개의 메타데이터를 포함하는 훈련용 데이터셋 및 상기 검색하고자 하는 영상에 대한 결과를 출력하는 신경망 모델을 저장하는 저장장치 및 상기 훈련용 데이터셋에서 추출한 제 1 메타데이터를 상기 신경망 모델에 입력하여 상기 신경망 모델을 훈련하고 상기 훈련된 신경망 모델에 상기 훈련용 데이터셋에서 추출한 제 2 메타데이터를 입력하여 상기 신경망 모델의 거리함수를 조정하는 연산장치를 포함하는 퓨-샷을 이용한 신경망 모델 구축 장치를 제공하는데 있다.

**발명의 효과**

[0011] 개시된 기술의 실시 예들은 다음의 장점들을 포함하는 효과를 가질 수 있다. 다만, 개시된 기술의 실시 예들이 이를 전부 포함하여야 한다는 의미는 아니므로, 개시된 기술의 권리범위는 이에 의하여 제한되는 것으로 이해되어서는 아니 될 것이다.

[0012] 개시된 기술의 일 실시예에 따르면 퓨-샷을 이용한 신경망 모델 구축 방법 및 장치는 적은 수의 데이터를 이용하여 검색하고자 하는 클래스에서의 영상 검색 성능을 향상시키는 효과가 있다.

[0013] 또한, 학습에 사용한 데이터셋과 큰 차이가 나는 클래스의 영상을 검색하더라도 원활하게 검색 결과를 출력하므로 의료영상이나 군사영상 등의 전문분야에 대한 영상 검색에 있어서 높은 정확도를 나타내는 효과가 있다.

**도면의 간단한 설명**

[0014] 도 1은 개시된 기술의 일 실시예에 따른 퓨-샷을 이용하여 적응된 신경망 모델을 이용하여 영상을 검색하는 과정을 나타낸 도면이다.

도 2는 개시된 기술의 일 실시예에 따른 퓨-샷을 이용한 신경망 모델 구축 방법에 대한 순서도이다.

도 3은 개시된 기술의 일 실시예에 따른 퓨-샷을 이용한 신경망 모델 적응 장치에 대한 블록도이다.

도 4는 신경망 모델을 적응시키는 과정을 나타낸 도면이다.

**발명을 실시하기 위한 구체적인 내용**

[0015] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세한 설명에 상세하게 설명하고자 한다. 그러나, 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.

[0016] 제 1, 제 2, A, B 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 해당 구성요소들은 상기 용어들에 의해 한정되지는 않으며, 단지 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 본 발명의 권리 범위를 벗어나지 않으면서 제 1 구성요소는 제 2 구성요소로 명명될 수 있고, 유사하게 제 2 구성요소도 제 1 구성요소로 명명될 수 있다. 및/또는 이라는 용어는 복수의 관련된 기재된 항목들의

조합 또는 복수의 관련된 기재된 항목들 중의 어느 항목을 포함한다.

- [0017] 본 명세서에서 사용되는 용어에서 단수의 표현은 문맥상 명백하게 다르게 해석되지 않는 한 복수의 표현을 포함하는 것으로 이해되어야 한다. 그리고 "포함한다" 등의 용어는 실시된 특징, 개수, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것이 존재함을 의미하는 것이지, 하나 또는 그 이상의 다른 특징들이나 개수, 단계 동작 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 배제하지 않는 것으로 이해되어야 한다.
- [0018] 도면에 대한 상세한 설명을 하기에 앞서, 본 명세서에서의 구성부들에 대한 구분은 각 구성부가 담당하는 주기능 별로 구분한 것에 불과함을 명확히 하고자 한다. 즉, 이하에서 설명할 2개 이상의 구성부가 하나의 구성부로 합쳐지거나 또는 하나의 구성부가 보다 세분화된 기능별로 2개 이상으로 분화되어 구비될 수도 있다.
- [0019] 그리고 이하에서 설명할 구성부 각각은 자신이 담당하는 주기능 이외에도 다른 구성부가 담당하는 기능 중 일부 또는 전부의 기능을 추가적으로 수행할 수도 있으며, 구성부 각각이 담당하는 주기능 중 일부 기능이 다른 구성부에 의해 전담되어 수행될 수도 있음은 물론이다. 따라서, 본 명세서를 통해 설명되는 각 구성부들의 존재 여부는 기능적으로 해석되어야 할 것이다.
- [0020] 도 1은 개시된 기술의 일 실시예에 따른 퓨-샷을 이용하여 적용된 신경망 모델을 이용하여 영상을 검색하는 과정을 나타낸 도면이다. 도 1을 참조하면 학습장치(110)는 데이터베이스와 신경망 모델을 저장한다. 데이터베이스에는 사용자의 검색 결과로 출력하기 위한 복수개의 영상이 저장되어 있으며 신경망 모델은 이러한 복수개의 영상들 중 사용자의 검색에 부합하는 영상을 출력하기 위해서 훈련하는 과정을 거친다. 학습장치(110)는 영상 검색 기능을 탑재하는 장치와 연결될 수 있으며 하나의 학습장치에서 영상 검색을 위한 학습과 검색 결과를 출력하는 기능을 모두 포함할 수 있다. 사용자는 학습장치에 검색하고자 하는 영상을 입력할 수 있으며 학습장치는 입력된 영상과 유사한 영상을 화면에 출력함으로써 사용자에게 제공할 수 있다.
- [0021] 신경망 모델을 훈련하기 위해서 학습장치(110)는 2가지 메타학습을 수행한다. 여기에서 메타학습은 데이터에 대한 거리정보를 이용하여 수행되는 학습을 의미한다. 즉, 메타데이터는 영상에서 추출한 특징값의 거리정보를 의미한다. 학습장치(110)는 각각 서로 다른 복수의 영상을 포함하는 2개의 훈련용 데이터셋을 이용하여 두 번의 메타학습을 수행할 수 있다. 예컨대, 제 1 훈련용 데이터셋의 제 1 메타데이터셋을 이용하여 신경망 모델을 훈련할 수 있다. 그리고 제 2 훈련용 데이터셋의 제 2 메타데이터셋을 이용하여 훈련된 신경망 모델을 검증할 수 있다.
- [0022] 한편, 학습장치(110)는 신경망 모델은 2가지 메타데이터셋을 이용한 학습의 결과가 서로 동일해질 때까지 또는 일정 이상의 유사도를 가질 때까지 훈련을 반복하게 된다. 예컨대, 제 1 메타데이터셋에 대한 출력 결과와 제 2 메타데이터셋에 대한 출력 결과를 비교하여 두 결과값의 거리가 동일한지 혹은 유사한지를 비교할 수 있다. 여기에서 제 1 메타데이터셋을 추출한 제 1 훈련용 데이터셋은 신경망 모델을 이용하여 검색하고자 하는 영상의 클래스와 서로 다른 클래스의 복수의 영상들을 포함한다. 그리고 제 2 메타데이터셋을 추출하는 제 2 훈련용 데이터셋은 검색하고자 하는 영상의 클래스와 동일한 클래스의 복수의 영상들을 포함한다. 학습장치(110)는 검색하고자 하는 영상과 동일한 클래스를 갖는 영상으로 신경망 모델을 적용시킨다. 적용을 위한 영상은 적은 수의 데이터를 이용한다. 예컨대, 적은 수의 레이블과 영상을 포함하는 퓨-샷(Few shot) 데이터일 수 있다. 이와 같이 훈련과 검증을 반복하여 결과가 일치해지도록 학습이 완료된 뒤 적용하는 과정을 거치면 영상 검색을 테스트하여 성능을 평가할 수 있다.
- [0023] 상술한 바와 같이 학습장치(110)는 영상 검색을 위해 적은 양의 레이블에 적용하도록 하는 퓨-샷 적용 기법을 이용하여 신경망 모델을 훈련 및 적용시킨다. 개시된 기술에서 제안하는 퓨-샷 적용 방식은 거리정보 기반의 메타학습 방식을 따르며 제 1 메타데이터셋과 제 2 메타데이터셋의 두 가지 메타데이터를 이용한다. 제 1 메타데이터셋은 많은 레이블을 포함하며, 제 2 메타데이터셋은 훈련용 데이터셋에서 일부를 추출하여 적은 수의 레이블과 영상으로 진행되는 메타학습을 의미한다. 예컨대, 제 1 메타데이터셋을 이용하는 훈련은 단순 미세조정(Simple fine-tuning) 방식의 메타학습일 수 있고, 제 2 메타데이터셋을 이용하는 검증은 모델에 구애 받지 않는 메타 학습(Model-agnostic Meta Learning, MAML)일 수 있다. 학습장치는 이와 같이 훈련과 검증을 통해 신경망 모델의 학습을 완료한다.
- [0024] 한편, 학습장치는 신경망 모델을 훈련하는 과정에서 충분히 많은 수의 제 1 메타데이터셋을 이용하여 훈련을 수행한다. 신경망 모델은 입력된 제 1 메타데이터셋에 포함된 다수의 거리정보와 영상에 따라 학습된다. 여기에서 제 1 메타데이터셋은 나중에 실제로 검색하고자 하는 영상의 클래스와 구분될 수 있다. 즉, 훈련하는 과정에서 이용되는 제 1 메타데이터셋은 일반적인 영상들을 포함하는 훈련용 데이터셋을 이용하는 것이며 의료영상이나

군사영상과 같은 특수한 클래스의 영상과는 다른 클래스를 갖는다. 신경망 모델은 단순 거리 학습 기법을 활용하여 학습이 진행된다. 그 이후 실제로 검색하기 원하는 클래스의 영상들을 이용하여 미세조정을 시행할 수 있다. 미세조정 역시 거리 학습 기법을 활용하며, 적은 데이터에 과적합하는 문제를 방지하기 위하여 적은 횟수의 학습만을 진행한다.

[0025] 제 2 메타데이터셋을 이용하여 신경망 모델을 검증하는 과정에서 학습장치는 에피소드 학습 기법을 활용한다. 학습장치는 훈련용 데이터셋에서 적은 수의 서포트 데이터셋(Support Dataset)과 쿼리 데이터셋(Query Dataset)을 제 2 메타데이터셋으로 추출한다. 이 때 서포트 데이터셋과 쿼리 데이터셋은 서로소이다. 즉, 서포트 데이터셋 및 쿼리 데이터셋은 서로 다른 클래스의 영상을 포함한다. 서포트 데이터셋은 신경망 모델의 학습을 위한 기준치를 초과하는 레이블 개수를 포함하고 쿼리 데이터셋은 기준치 미만의 레이블 개수를 포함할 수 있다.

[0026] 학습장치(110)는 신경망 모델을 서포트 데이터셋으로 미세조정 한 이후, 쿼리 데이터셋으로 손실(Loss)을 계산한다. 이를 통해 미세조정 이전의 각 변수의 초기화를 학습한다. 이는 현재의 초기화를 이용해 미세조정을 했을 때 신경망 모델이 과적합되는 문제가 발생하지 않도록 잘 학습되는지를 평가하고 미세조정이 잘 되는 방향으로 학습되도록 한다.

[0027] 이와 같이 두 가지 메타학습 방식을 신경망 모델에 적용한 후 신경망 모델을 영상 검색에 이용하기 위해, 신경망 모델을 통해 출력된 영상의 특징점을 광역 평균 풀링(Global Average Pooling)한 후 유클리드 정규화(L2 normalization)을 거칠 수 있다. 위 과정을 거쳐 쿼리 영상의 특징점과 데이터베이스 내 영상들의 특징점들을 추출하여 유클리드 거리를 비교함으로써 쿼리 영상과 거리가 가장 가까운 순으로 일부의 영상을 출력할 수 있다.

[0028] 한편, 학습장치(110)는 출력 결과가 정확한지를 평가할 수 있다. 예컨대, 사용자로부터 출력된 영상에 대한 정확도를 확인하기 위한 피드백을 수신할 수 있다. 학습장치(110)는 피드백에 따라 출력 결과를 조정할 수 있도록 신경망 모델을 조정하여 보다 정확한 영상 검색을 수행할 수 있다.

[0029] 도 2는 개시된 기술의 일 실시예에 따른 퓨-샷을 이용한 신경망 모델 구축 방법에 대한 순서도이다. 도 2를 참조하면 퓨-샷을 이용한 신경망 모델 구축 방법(200)은 학습장치가 신경망 모델을 제 1 메타데이터셋을 이용하여 훈련하는 단계(210), 학습장치가 신경망 모델을 제 2 메타데이터셋을 이용하여 파라미터를 조정하는 단계(220), 두 메타데이터셋의 출력 결과를 비교하는 단계(230) 및 영상 검색 성능을 평가하는 단계(240)를 포함한다.

[0030] 신경망 모델은 메타학습 방식에 따라 훈련된다. 일 실시예로, 210 단계에 따라 미세조정 메타학습(Simple fine-tuning Meta learning)이 수행되고 220 단계에 따라 모델에 구애받지 않는 메타학습(Model agnostic Meta learning, MAML)이 수행될 수 있다. 신경망 모델은 230 단계에 따라 두 메타학습의 결과가 동일해질 때까지 반복적으로 훈련될 수 있다. 또는 두 메타학습의 결과가 일정 이상의 유사도를 나타낼 때까지 반복될 수 있다. 동일한 결과 또는 일정 이상의 유사도를 갖는 결과를 나타낸다는 것은 신경망 모델의 거리함수가 조정되기 이전의 각 변수의 초기화 상태를 학습하는 것을 의미한다. 즉, 신경망 모델을 미세조정했을 때 과적합 문제 없이 정확하게 학습되도록 학습할 수 있다.

[0031] 240 단계에서 학습장치는 퓨샷을 이용하여 적응된 신경망 모델을 이용하여 영상 검색을 수행함으로써 검색 성능을 평가할 수 있다. 앞서의 학습 과정에 따라 신경망 모델은 특정 클래스의 영상과 동일한 클래스의 영상들을 정확하게 출력할 수 있도록 학습되었다. 따라서 일반적인 영상데이터에 포함되지 않는 드문 경우의 영상에 대해서도 높은 정확도의 검색 성능을 나타낼 수 있다. 일 실시예로, 의료영상이나 군사영상과 같이 일반적이지 않은 클래스의 영상에 대하여 검색을 수행할 수 있다. 학습장치는 검색하고자 하는 영상과 거리가 가까운 영상을 출력하고 그에 따른 피드백을 수신하여 검색 성능을 평가할 수 있다.

[0032] 도 3은 개시된 기술의 일 실시예에 따른 퓨-샷을 이용한 신경망 모델 구축 장치에 대한 블록도이다. 도 3을 참조하면 신경망 모델 구축 장치(300)는 학습장치가 그 역할을 수행하거나 학습장치에 탑재되는 별도의 장치로 구현될 수 있다. 신경망 모델 구축 장치(300)는 입력장치(310), 저장장치(320) 및 연산장치(330)를 포함한다. 그리고 영상 검색 결과를 출력하는 출력장치(340)를 더 포함한다. 입력장치(310)는 신경망 모델 구축 장치(300)의 마우스나 키보드와 같은 입력수단의 형태로 구현될 수 있다. 저장장치(320)는 신경망 모델 구축 장치(300)의 메모리로 구현될 수 있다. 연산장치(330)는 신경망 모델 구축 장치(300)의 AP나 CPU와 같은 프로세서로 구현될 수 있다. 출력장치(340)는 텍스트나 그래픽적 요소를 출력할 수 있는 디스플레이로 구현될 수 있다.

[0033] 입력장치(310)는 검색하고자 하는 영상을 입력받는다. 저장장치(320)는 복수개의 영상 및 복수개의 영상 각각에 대한 특징점의 거리정보를 포함하는 복수개의 메타데이터를 포함하는 훈련용 데이터셋을 저장한다. 그리고 검색



하고자 하는 영상에 대한 결과를 출력하는 신경망 모델도 함께 저장한다.

[0034] 연산장치(330)는 훈련용 데이터셋에서 추출한 제 1 메타데이터를 상기 신경망 모델에 입력하여 신경망 모델을 훈련한다. 그리고 훈련된 신경망 모델에 훈련용 데이터셋에서 추출한 제 2 메타데이터를 입력하여 신경망 모델의 거리함수를 조정한다. 이러한 과정에 따라 신경망 모델 구축 장치(300)는 신경망 모델이 영상에 대한 정확한 검색을 수행하도록 학습한다.

[0035] 한편, 상술한 신경망 모델 구축 장치(300)는 컴퓨터와 같은 디바이스에서 실행될 수 있는 실행가능한 알고리즘을 포함하는 프로그램(또는 어플리케이션)으로 구현될 수 있다. 상기 프로그램은 일시적 또는 비일시적 판독 가능 매체(non-transitory computer readable medium)에 저장되어 제공될 수 있다.

[0036] 비일시적 판독 가능 매체란 레지스터, 캐쉬, 메모리 등과 같이 짧은 순간 동안 데이터를 저장하는 매체가 아니라 반영구적으로 데이터를 저장하며, 기기에 의해 판독(reading)이 가능한 매체를 의미한다. 구체적으로는, 상술한 다양한 어플리케이션 또는 프로그램들은 CD, DVD, 하드 디스크, 블루레이 디스크, USB, 메모리카드, ROM (read-only memory), PROM (programmable read only memory), EPROM(Erasable PROM, EPROM) 또는 EEPROM(Electrically EPROM) 또는 플래시 메모리 등과 같은 비일시적 판독 가능 매체에 저장되어 제공될 수 있다.

[0037] 일시적 판독 가능 매체는 스태틱 램(Static RAM, SRAM), 다이내믹 램(Dynamic RAM, DRAM), 싱크로너스 디램(Synchronous DRAM, SDRAM), 2배속 SDRAM(Double Data Rate SDRAM, DDR SDRAM), 증강형 SDRAM(Enhanced SDRAM, ESDRAM), 동기화 DRAM(Synclink DRAM, SLDRAM) 및 직접 램버스 램(Direct Rambus RAM, DRRAM) 과 같은 다양한 RAM을 의미한다.

[0038] 도 4는 신경망 모델을 적용시키는 과정을 나타낸 도면이다. 도 4를 참조하면 우선 첫 번째 메타학습을 위한 데이터셋  $S^{mtr}$  을 이용해 신경망 모델  $f$  를 메타 학습한다. 이 과정에서 신경망 모델은 거리 학습을 통해 거리함수를 미세조정하는 방식으로 수행된다. 다음으로 두번째 메타학습의 경우 우선 데이터셋에서 적은 수의 서포트 집합  $S_k^{mtr}$  과 쿼리 집합  $P_k^{mtr}$  을 추출한다. 우선, 학습 중인 초기화  $\theta_0$  로 CNN을 초기화한 후,  $S_k^{mtr}$  을 이용해 미세조정한다. 서포트 집합을 이용하여 미세조정하는 과정은 이하의 수학적 식 1에 따라 수행된다.

**수학적 식 1**

[0039] 
$$\theta_1 = \theta_0 - \alpha \nabla_{\theta} \mathcal{L}(S_k^{mtr}; \theta_0)$$

[0041] 수학적 식 1을 참조하면 신경망 모델을 초기화 상태인  $\theta_0$  부터 시작하여 서포트 집합을 이용해 미세조정이 가능하다. 여기에서  $\mathcal{L}$  은 일반적인 거리 학습의 손실 함수이며,  $\alpha$  는 학습 횟수를 의미한다. 이렇게 연산 된  $\theta_1$  이 잘 동작하는지  $P_k^{mtr}$  에 평가하여 초기화  $\theta_0$  를 메타 갱신한다. 메타 갱신하는 과정은 이하의 수학적 식 2에 따라 수행된다.

**수학적 식 2**

[0042] 
$$\theta'_0 = \theta_0 - \beta \nabla_{\theta} \mathcal{L}(P_k^{mtr}; \theta_1) = \theta_0 - \beta \nabla_{\theta} \mathcal{L}(P_k^{mtr}; \theta_0 - \alpha \nabla_{\theta} \mathcal{L}(S_k^{mtr}; \theta_0))$$

[0044] 수학적 식 2를 참조하면 미세조정이 완료 된  $\theta_1$  이 잘 동작하는지 평가하고, 해당 손실 함수를 통해 초기화  $\theta_0$  를 메타 갱신하는 것이 가능하다. 여기에서  $\beta$  는 메타학습 횟수를 나타낸다.

[0045] 한편, 메타학습이 완료된 이후, 검색을 원하는 클래스에 해당하는 레이블 된 소수의 서포트 집합을  $S^s$ , 실제로 검색을 수행할 데이터베이스 및 쿼리에 해당하는 쿼리 집합을  $P^q$  라 한다. 이 때,  $S^s$  에서 적은 횟수 미세조정을 수

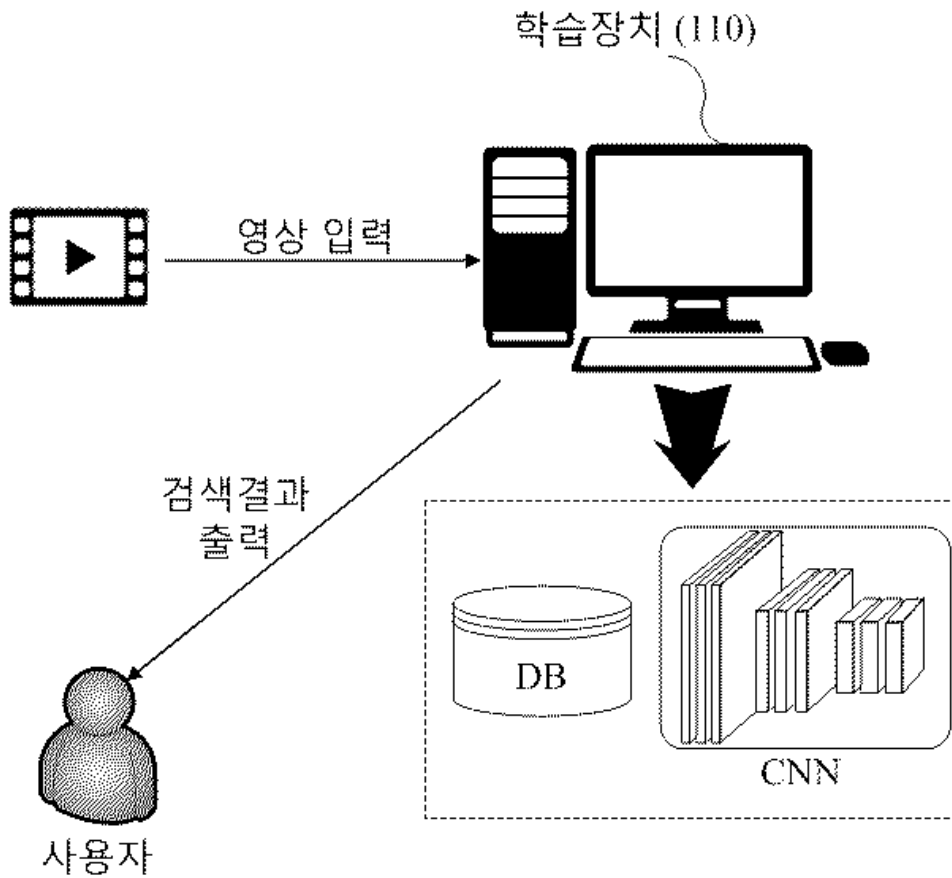
행한 후, P에서 실제 검색을 수행한다. 이러한 과정에 따라 신경망 모델을 퓨-샷으로 적응시켜서 영상 검색 성능의 정확도를 높일 수 있다. 특히 의료영상이나 군사영상과 같이 일반적인 데이터베이스에 저장되는 복수의 영상들과는 클래스 차이가 큰 영상을 검색할 때 높은 정확도를 나타낼 수 있다.

[0046] 개시된 기술의 일 실시예에 따른 퓨-샷을 이용한 신경망 모델 구축 방법 및 장치는 이해를 돕기 위하여 도면에 도시된 실시 예를 참고로 설명되었으나, 이는 예시적인 것에 불과하며, 당해 분야에서 통상적 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시 예가 가능하다는 점을 이해할 것이다. 따라서, 개시된 기술의 진정한 기술적 보호범위는 첨부된 특허청구범위에 의해 정해져야 할 것이다.

도면

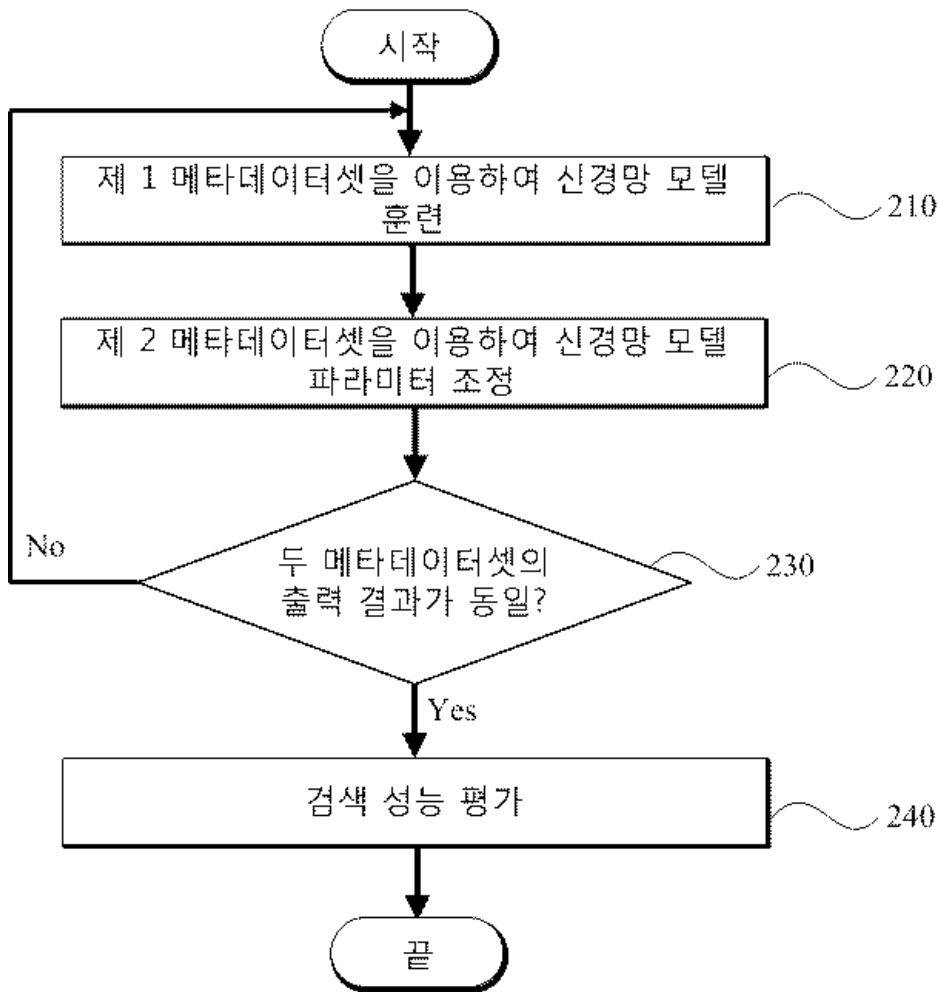
도면1

**100**

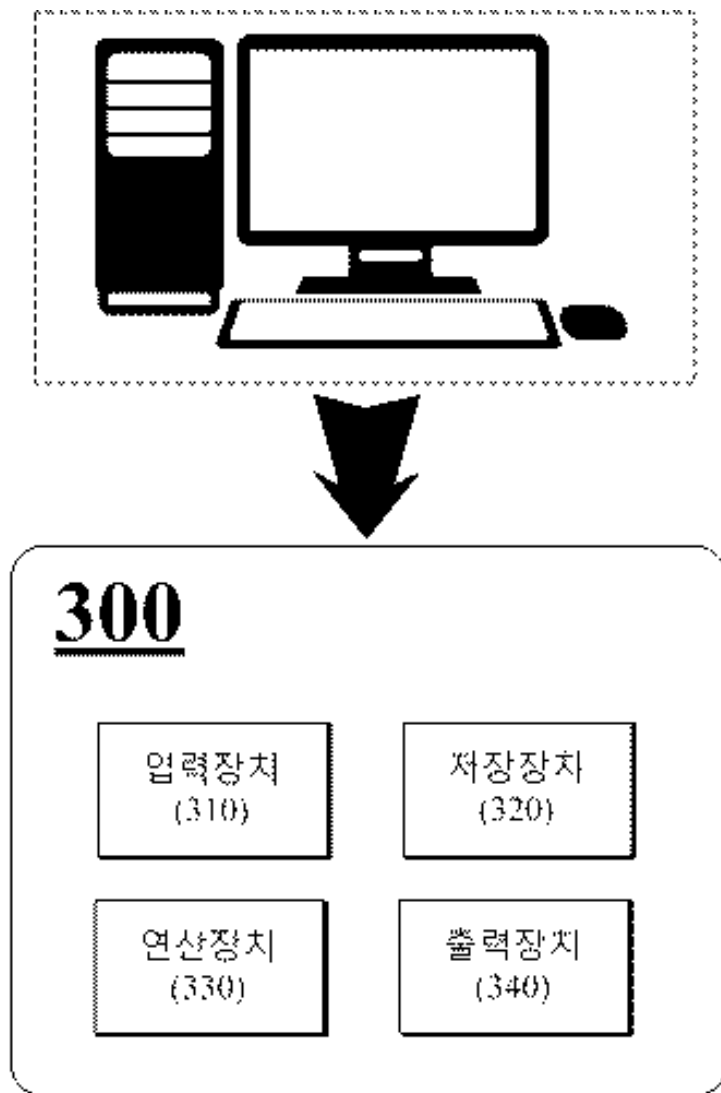


도면2

**200**



도면3



도면4

**400**

