



- (51) 국제특허분류:
G06K 9/00 (2006.01) G06N 3/08 (2006.01)
- (21) 국제출원번호: PCT/KR2020/000647
- (22) 국제출원일: 2020년 1월 14일 (14.01.2020)
- (25) 출원언어: 한국어
- (26) 공개언어: 한국어
- (30) 우선권정보:
10-2019-0005188 2019년 1월 15일 (15.01.2019) KR
10-2019-0005200 2019년 1월 15일 (15.01.2019) KR
- (71) 출원인: 포항공과대학교 산학협력단 (POSTECH RESEARCH AND BUSINESS DEVELOPMENT FOUNDATION) [KR/KR]; 37673 경상북도 포항시 남구 청암로 77, Gyeongsangbuk-do (KR).
- (72) 발명자: 이영주 (LEE, Young Joo); 37673 경상북도 포항시 남구 청암로 77, Gyeongsangbuk-do (KR). 김영석 (KIM, Young Seok); 13595 경기도 성남시 분당구 황새울로200번길 9-7, 2101호, Gyeonggi-do (KR). 박군호

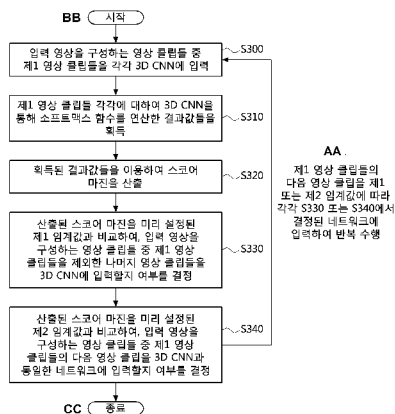
(PARK, Gun Ho); 47168 부산시 부산진구 개금본동로 42, 101동 405호, Busan (KR). 이현훈 (LEE, Hyun Hoon); 38834 경상북도 영천시 모란길 18, 201동 809호, Gyeongsangbuk-do (KR).

(74) 대리인: 특허법인 이상 (E-SANG PATENT & TRADE-MARK LAW FIRM); 06747 서울시 서초구 바우피로 188, 3층, Seoul (KR).

(81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(54) Title: METHOD AND DEVICE FOR HIGH-SPEED IMAGE RECOGNITION USING 3D CNN

(54) 발명의 명칭: 3D CNN을 이용한 고속 영상 인식 방법 및 장치



S300 ... Input, to 3D CNN, each of first image clips among image clips constituting input image
S310 ... Acquire result values obtained by calculating softmax function via 3D CNN with respect to each of first image clips
S320 ... Calculate score margin by using acquired result values
S330 ... Compare calculated score margin with preconfigured first threshold value, and determine whether to input image clips constituting input image, which remain after excluding first image clips, to 3D CNN
S340 ... Compare calculated score margin with preconfigured second threshold value, and determine whether to input, to same network as 3D CNN, image clip subsequent to first image clips among image clips constituting input image
AA ... Input image clip subsequent to first image clips to network determined according to first or second threshold value in S330 or S340, respectively, so as to perform steps repeatedly
BB ... Start
CC ... End

(57) Abstract: Disclosed is a method and device for high-speed image recognition using a 3-dimension convolutional neural network (3D CNN). A method for high-speed image recognition using a 3-dimension convolutional neural network (3D CNN) comprises the steps of: inputting, to a 3D CNN, each of first image clips among image clips constituting an input image; acquiring result values obtained by calculating a softmax function via the 3D CNN with respect to each of the first image clips; calculating a score margin by using the acquired result values; comparing the calculated score margin with a preconfigured threshold value; and in response to the step of comparing, determining whether to input the image clips constituting the input image, which remain after excluding the first image clips, to the 3D CNN. Therefore, a calculation speed for image recognition can be improved.

(57) 요약서: 3D CNN(3-dimension Convolutional Neural Network)을 이용한 고속 영상 인식 방법 및 장치가 개시된다. 3D CNN(3-dimension Convolutional Neural Network)을 이용한 고속 영상 인식 방법은, 입력 영상을 구성하는 영상 클립들 중 제1 영상 클립들을 각각 3D CNN에 입력하는 단계, 상기 제1 영상 클립들 각각에 대하여 상기 3D CNN을 통해 소프트맥스 함수(softmax function)를 연산한 결과값들을 획득하는 단계, 획득된 결과값들을 이용하여 스코어 마진(score margin)을 산출하는 단계, 산출된 스코어 마진을 미리 설정된 임계값과 비교하는 단계 및 상기 비교하는 단계에 대한 응답으로, 상기 입력 영상을 구성하는 영상 클립들 중 상기 제1 영상 클립들을 제외한 나머지 영상 클립들을 상기 3D CNN에 입력할지 여부를 결정하는 단계를 포함한다. 따라서, 영상 인식을 위한 연산 속도를 향상시킬 수 있다.



(84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 역내 권리의 보호를 위하여): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 유라시아 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 유럽 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

공개:

— 국제조사보고서와 함께 (조약 제21조(3))

명세서

발명의 명칭: 3D CNN을 이용한 고속 영상 인식 방법 및 장치 기술분야

- [1] 본 발명은 3D CNN을 이용한 고속 영상 인식 방법 및 장치에 관한 것으로, 더욱 상세하게는 입력 영상 클립 중 일부에 대하여 3D CNN을 이용한 영상 인식을 위한 네트워크 연산을 수행하고, 수행 결과를 기초로 후속 영상 클립에 대한 네트워크 연산을 일부 생략함으로써 연산 속도를 고속화하는 기술에 관한 것이다.

배경기술

- [2] 인공지능 기술이 발전함에 따라 높은 수준의 추상화(abstraction)를 통해 사람의 사고를 컴퓨터가 수행할 수 있도록 학습시키는 기계학습 알고리즘인 딥러닝 기법이 연구되고 있다. 이러한 딥러닝 기법은 심층 신경망(Deep Neural Network), 합성곱 신경망(Convolutional Neural Network, CNN), 순환 신경망(Recurrent Neural Network, RNN) 등과 같은 다양한 인공신경망을 사용하여 트레이닝셋을 훈련시키고 입력된 데이터에 대한 추론을 수행한다.
- [3] 특히, 합성곱 신경망은 이미지 분류에서 뛰어난 성능을 나타내는 것으로 주목받고 있는 네트워크로서, 하나 이상의 합성곱 계층(Convolutional layer)을 포함한다.
- [4] 합성곱 신경망을 이용한 영상 인식 기술은 영상에 포함된 객체를 식별하거나 객체(또는 사람)의 행동을 인식하는 데 주로 사용된다. 그런데, 객체의 행동을 인식하기 위해서 주로 사용되는 3차원 합성곱 신경망(3dimension Convolutional Neural Network)은 단일한 2차원 이미지가 아니라 복수의 2차원 이미지로 구성되는 3차원 영상 이미지를 입력으로 사용한다.
- [5] 종래의 3차원 합성곱 신경망은 깊은 네트워크를 사용함에 따른 많은 연산량과 변수를 처리하기 위하여 많은 자원을 필요로 하기 때문에 IoT(Internet of Things) 기기를 포함한 소형 기기들의 제한적인 자원으로는 구현이 어려운 문제가 있다.

발명의 상세한 설명

기술적 과제

- [6] 상기와 같은 문제점을 해결하기 위한 본 발명의 목적은, 3D CNN을 이용한 고속 영상 인식 방법을 제공하는 데 있다.
- [7] 상기와 같은 문제점을 해결하기 위한 본 발명의 다른 목적은, 3D CNN을 이용한 고속 영상 인식 장치를 제공하는 데 있다.

과제 해결 수단

- [8] 상기 목적을 달성하기 위한 본 발명의 일 측면은, 3D CNN을 이용한 고속 영상 인식 방법을 제공한다.
- [9] 상기 3D CNN을 이용한 고속 영상 인식 방법은, 입력 영상을 구성하는 영상

클립들 중 제1 영상 클립들을 각각 3D CNN(3-dimension Convolutional Neural Network)에 입력하는 단계, 상기 제1 영상 클립들 각각에 대하여 상기 3D CNN을 통해 소프트맥스 함수(softmax function)를 연산한 결과값들을 획득하는 단계, 획득된 결과값들을 이용하여 스코어 마진(score margin)을 산출하는 단계, 산출된 스코어 마진을 미리 설정된 임계값과 비교하여, 상기 입력 영상을 구성하는 영상 클립들 중 상기 제1 영상 클립들을 제외한 적어도 하나의 영상 클립의 추가 입력 여부를 결정하는 단계를 포함할 수 있다.

- [10] 상기 스코어 마진은 상기 결과값들 중 가장 큰 값과 두번째로 큰 값 사이의 차분값일 수 있다.
- [11] 상기 적어도 하나의 영상 클립의 추가 입력 여부를 결정하는 단계는, 상기 제1 영상 클립들을 제외한 나머지 영상 클립들을 상기 3D CNN에 입력할지 여부를 결정하는 단계를 포함할 수 있다. 이때, 상기 나머지 영상 클립들을 상기 3D CNN에 입력할지 여부를 결정하는 단계는, 상기 스코어 마진이 상기 임계값보다 클 경우 상기 제1 영상 클립들 이후의 영상 클립을 상기 3D CNN에 입력하지 않고, 상기 결과값들만으로 상기 입력 영상에 대한 영상 인식을 수행하는 단계 및 상기 스코어 마진이 상기 임계값보다 작을 경우 상기 제1 영상 클립들 이후의 영상 클립을 상기 3D CNN에 입력하는 단계를 포함할 수 있다.
- [12] 한편, 상기 적어도 하나의 영상 클립의 추가 입력 여부를 결정하는 단계는, 상기 제1 영상 클립들의 다음 영상 클립을 상기 3D CNN과 동일한 네트워크에 입력할지 여부를 결정하는 단계를 포함할 수 있다.
- [13] 이때, 상기 다음 영상 클립을 상기 3D CNN과 동일한 네트워크에 입력할지 여부를 결정하는 단계는, 상기 스코어 마진이 상기 임계값보다 클 경우 상기 제1 영상 클립들의 다음 영상 클립을 상기 3D CNN과 동일하거나 상기 3D CNN보다 더 얇은 네트워크에 입력하는 단계 및 상기 스코어 마진이 상기 임계값보다 작을 경우 상기 제1 영상 클립의 다음 영상 클립을 상기 3D CNN보다 더 깊은 네트워크에 입력하는 단계를 포함할 수 있다.
- [14] 또한, 상기 적어도 하나의 영상 클립의 추가 입력 여부를 결정하는 단계는, 상기 스코어 마진을 제1 임계값과 비교하여, 상기 제1 영상 클립들을 제외한 나머지 영상 클립들을 상기 3D CNN에 입력할지 여부를 결정하는 단계 및 상기 스코어 마진을 상기 제1 임계값과 다른 제2 임계값과 비교하여, 상기 제1 영상 클립들의 다음 영상 클립을 상기 3D CNN과 동일한 네트워크에 입력할지 여부를 결정하는 단계를 포함할 수 있다.
- [15] 상기 결과값들을 획득하는 단계는, 상기 소프트맥스 함수를 연산하여 획득된 결과값들을 메모리(memory)에 누적하여 저장하는 단계를 더 포함할 수 있다.
- [16] 상기 임계값은, 영상 인식을 수행하는 단말의 종류, 연산 능력, 입력 영상의 종류, 입력 영상의 해상도, 입력 영상을 구성하는 프레임 수 중 적어도 하나에 따라 결정될 수 있다.
- [17] 상기 입력 영상을 구성하는 영상 클립들 각각은, 상기 입력 영상을 구성하는

복수의 프레임 중에서 미리 설정된 개수의 시간적으로 연속한 프레임들로 구성될 수 있다.

- [18] 상기 목적을 달성하기 위한 본 발명의 다른 측면은, 3D CNN을 이용한 고속 영상 인식 장치를 제공한다.
- [19] 3D CNN을 이용한 고속 영상 인식 장치는, 적어도 하나의 프로세서(processor), 및 상기 적어도 하나의 프로세서가 적어도 하나의 단계를 수행하도록 지시하는 명령어들(instructions)을 저장하는 메모리(memory)를 포함할 수 있다.
- [20] 상기 적어도 하나의 단계는, 입력 영상을 구성하는 영상 클립들 중 제1 영상 클립들을 각각 3D CNN(3-dimension Convolutional Neural Network)에 입력하는 단계, 상기 제1 영상 클립들 각각에 대하여 상기 3D CNN을 통해 소프트맥스 함수(softmax function)를 연산한 결과값들을 획득하는 단계, 획득된 결과값들을 이용하여 스코어 마진(score margin)을 산출하는 단계, 산출된 스코어 마진을 미리 설정된 임계값과 비교하여, 상기 입력 영상을 구성하는 영상 클립들 중 상기 제1 영상 클립들을 제외한 적어도 하나의 영상 클립의 추가 입력 여부를 결정하는 단계를 포함할 수 있다.
- [21] 상기 스코어 마진은 상기 결과값들 중 가장 큰 값과 두번째로 큰 값 사이의 차분값일 수 있다.
- [22] 상기 적어도 하나의 영상 클립의 추가 입력 여부를 결정하는 단계는, 상기 제1 영상 클립들을 제외한 나머지 영상 클립들을 상기 3D CNN에 입력할지 여부를 결정하는 단계를 포함할 수 있다.
- [23] 이때, 상기 나머지 영상 클립들을 상기 3D CNN에 입력할지 여부를 결정하는 단계는, 상기 스코어 마진이 상기 임계값보다 클 경우 상기 제1 영상 클립들 이후의 영상 클립을 상기 3D CNN에 입력하지 않고, 상기 결과값들만으로 상기 입력 영상에 대한 영상 인식을 수행하는 단계 및 상기 스코어 마진이 상기 임계값보다 작을 경우, 상기 제1 영상 클립들 이후의 영상 클립을 상기 3D CNN에 입력하는 단계를 포함할 수 있다.
- [24] 한편, 상기 적어도 하나의 영상 클립의 추가 입력 여부를 결정하는 단계는, 상기 제1 영상 클립들의 다음 영상 클립을 상기 3D CNN과 동일한 네트워크에 입력할지 여부를 결정하는 단계를 포함할 수 있다.
- [25] 여기서, 상기 다음 영상 클립을 상기 3D CNN과 동일한 네트워크에 입력할지 여부를 결정하는 단계는, 상기 스코어 마진이 상기 임계값보다 클 경우 상기 제1 영상 클립들의 다음 영상 클립을 상기 3D CNN과 동일하거나 상기 3D CNN보다 더 얇은 네트워크에 입력하는 단계 및 상기 스코어 마진이 상기 임계값보다 작으면, 상기 제1 영상 클립의 다음 영상 클립을 상기 3D CNN보다 더 깊은 네트워크에 입력하는 단계를 포함할 수 있다.
- [26] 상기 결과값들을 획득하는 단계는, 상기 소프트맥스 함수를 연산하여 획득된 결과값들을 메모리(memory)에 누적하여 저장하는 단계를 더 포함할 수 있다.
- [27] 상기 임계값은, 영상 인식을 수행하는 단말의 종류, 연산 능력, 입력 영상의

종류, 입력 영상의 해상도, 입력 영상을 구성하는 프레임 수 중 적어도 하나에 따라 결정될 수 있다.

- [28] 상기 입력 영상을 구성하는 영상 클립들 각각은, 상기 입력 영상을 구성하는 복수의 프레임 중에서 미리 설정된 개수의 시간적으로 연속한 프레임들로 구성될 수 있다.
- [29] 상기 목적을 달성하기 위한 본 발명의 또다른 측면은, 3D CNN을 이용한 고속 영상 인식 장치를 제공한다.
- [30] 3D CNN을 이용한 고속 영상 인식 장치는, 적어도 하나의 프로세서(processor), 및 상기 적어도 하나의 프로세서가 적어도 하나의 단계를 수행하도록 지시하는 명령어들(instructions)을 저장하는 메모리(memory)를 포함할 수 있다.
- [31] 상기 적어도 하나의 단계는, 입력 영상을 구성하는 영상 클립들 중 제1 영상 클립들을 각각 3D CNN(3-dimension Convolutional Neural Network)에 입력하는 단계, 상기 제1 영상 클립들 각각에 대하여 상기 3D CNN을 통해 소프트맥스 함수(softmax function)를 연산한 결과값들을 획득하는 단계, 획득된 결과값들을 이용하여 스코어 마진(score margin)을 산출하는 단계, 산출된 스코어 마진을 미리 설정된 제1 임계값과 비교하여, 상기 입력 영상을 구성하는 영상 클립들 중 상기 제1 영상 클립들을 제외한 나머지 영상 클립들을 상기 3D CNN과 동일한 네트워크에 입력할지 여부를 결정하는 단계 및 상기 스코어 마진을 미리 설정된 제2 임계값과 비교하여, 상기 제1 영상 클립들의 다음 영상 클립을 상기 3D CNN과 동일한 네트워크에 입력할지 여부를 결정하는 단계를 포함할 수 있다.

발명의 효과

- [32] 상기와 같은 본 발명에 따른 3D CNN을 이용한 고속 영상 인식 방법 및 장치를 이용할 경우에는 스코어 마진에 따라 후속 영상 클립에 대한 연산을 생략함으로써 연산 속도를 향상시키고 시스템 자원 요구사항을 낮출 수 있다.
- [33] 또한, 제한적인 자원을 갖는 각종 장치에서도 3D CNN을 이용하여 영상 인식을 수행할 수 있는 장점이 있다.

도면의 간단한 설명

- [34] 도 1은 본 발명의 일 실시예에 따른 2차원 합성곱 신경망을 설명하기 위한 예시도이다.
- [35] 도 2는 본 발명의 일 실시예에 따른 3D CNN을 설명하기 위한 예시도이다.
- [36] 도 3a은 본 발명의 일 실시예에 따른 스코어 마진값을 설명하기 위한 히스토그램이다.
- [37] 도 3b는 본 발명의 다른 실시예에 따른 스코어 마진값을 설명하기 위한 히스토그램이다.
- [38] 도 4는 본 발명의 제1 실시예에 따른 3D CNN을 이용한 고속 영상 인식 방법에 대한 흐름도이다.
- [39] 도 5는 본 발명의 제2 실시예에 따른 3D CNN을 이용하는 고속 영상 인식

방법에 대한 흐름도이다.

[40] 도 6은 본 발명의 제3 실시예에 따른 3D CNN을 이용한 고속 영상 인식 방법에 대한 흐름도이다.

[41] 도 7은 본 발명의 제1 내지 제3 실시예에 따른 3D CNN을 이용한 고속 영상 인식 장치에 대한 구성도이다.

발명의 실시를 위한 형태

[42] 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. 각 도면을 설명하면서 유사한 참조부호를 유사한 구성요소에 대해 사용하였다.

[43] 제1, 제2, A, B 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 본 발명의 권리 범위를 벗어나지 않으면서 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다. 및/또는 이라는 용어는 복수의 관련된 기재된 항목들의 조합 또는 복수의 관련된 기재된 항목들 중의 어느 항목을 포함한다.

[44] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소가 다른 구성요소에 "직접 연결되어" 있다거나 "직접 접속되어" 있다고 언급된 때에는, 중간에 다른 구성요소가 존재하지 않는 것으로 이해되어야 할 것이다.

[45] 본 출원에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 출원에서, "포함하다" 또는 "가지다" 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[46] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가지고 있다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥 상 가지는 의미와 일치하는 의미를 가지는 것으로 해석되어야 하며, 본 출원에서 명백하게

- 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.
- [47] 이하, 본 발명에 따른 바람직한 실시예를 첨부된 도면을 참조하여 상세하게 설명한다.
- [48]
- [49] 도 1은 본 발명의 일 실시예에 따른 2차원 합성곱 신경망을 설명하기 위한 예시도이다.
- [50] 도 1을 참조하면, 2차원 합성곱 신경망(Convolutional Neural Network, 이하 CNN)의 기본적인 계층 구조를 확인할 수 있다. 구체적으로, 2차원 합성곱 신경망은 입력 이미지를 입력으로 받아 합성곱 연산을 수행하여 특징맵(feature map)을 출력하는 컨볼루션 계층(Convolutional layer, 10), 활성화 함수(activation function)를 이용하여 컨볼루션 계층(10)의 출력값을 정규화하는 활성화 계층(activation layer, 11), 활성화 계층(11)의 출력에 대하여 샘플링 또는 풀링을 수행하여 대표적 특징을 추출하는 풀링 계층(pooling layer, 12)를 포함할 수 있다. 이때, 컨볼루션 계층(10), 활성화 계층(11), 풀링 계층(12)의 연결 구조는 여러 세트 반복적으로 구성될 수 있다. 또한, 합성곱 신경망은 상기 연결 구조의 후단에 풀링 계층(12)을 통해 추출된 여러 개의 특징들을 결합하는 전결합층(Fully-Connected layer, 13)과 연결되고, 소프트맥스 함수(softmax function)를 이용하여 전결합층(13)의 출력을 정규화하는 소프트 맥스 계층(Softmax layer, 14)으로 연결될 수 있다.
- [51] 컨볼루션 계층(10)은 입력 이미지와 필터 사이의 컨볼루션 연산을 수행할 수 있다. 필터는 입력 이미지의 각 픽셀들(pixels)과 합성곱 연산을 수행하기 위한 성분값을 갖는 픽셀 단위의 영역으로 정의될 수 있다. 이때, 픽셀 단위의 영역을 필터의 크기로 지칭할 수 있으며, 필터는 일반적으로 행렬로서 표현될 수 있다. 컨볼루션 계층(10)은, 필터를 입력 이미지의 가로, 세로 방향으로 이동(sliding)시키면서, 필터와 입력 이미지 사이의 컨볼루션 연산을 반복할 수 있다. 이때, 필터가 한번에 이동하는 간격을 스트라이드(stride)로 정의할 수 있다. 예를 들어 스트라이드 값이 2라면, 2개의 픽셀 간격만큼 필터가 이동하면서 입력 이미지와의 합성곱 연산을 수행할 수 있다. 또한, 컨볼루션 계층(convolutional layer, 10)이 반복됨에 따라 출력되는 이미지(또는 특징맵)의 크기가 작아질 수 있는데, 컨볼루션 계층은 출력되는 특징맵의 크기를 조절하기 위하여 패딩 과정을 수행할 수 있다. 여기서 패딩 과정은 입력 이미지의 바깥쪽 영역에 특정 값(예를 들면 0)을 채우는 과정일 수 있다.
- [52] 활성화 계층(11)에서 활성화 함수는 어떠한 값(또는 행렬)으로 추출된 특징을 비선형 값으로 바꾸는 함수로서, 시그모이드(sigmoid) 함수, ReLU 함수 등이 사용될 수 있다. 도 1에서 활성화 계층(11)은 설명의 편의를 위해 컨볼루션 계층(10)과 별도로 도시하였으나 활성화 계층(11)이 컨볼루션 계층(10)에 포함되는 것으로 해석할 수도 있다.
- [53] 풀링 계층(12)은 추출된 특징맵에 대하여 서브 샘플링(subsampling) 또는

풀링(pooling)을 수행하여 특징맵을 대표하는 특징을 선정하는 계층으로서, 특징맵의 일정 영역에 대하여 가장 큰 값을 추출하는 맥스 풀링(max pooling), 평균값을 추출하는 에버리지 풀링(average pooling) 등이 수행될 수 있다. 이때, 풀링 계층(12)은 활성화 계층(11) 이후에 반드시 수행되는 것이 아니라 선택적으로 수행될 수도 있다.

[54] 전결합층(13)은 일반적으로 CNN의 마지막에 위치하며, 전결합층(13)에서는 컨볼루션 계층(10), 활성화 계층(11), 풀링 계층(12)을 통해 추출된 특징들을 결합하여 어떤 클래스(class)에 해당하는지를 판단할 수 있다.

[55] 구체적으로, 전결합층(13)은 입력된 특징맵의 모든 픽셀을 벡터화하고, 각각의 파라미터 값들을 곱한 후, 연산 결과들을 종합하여 가장 큰 값을 갖는 클래스를 결과로 출력할 수 있다. 소프트맥스 계층(14)은, 소프트맥스 함수를 사용하여 전결합층(13)에서의 연산 결과 값을 0과 1 사이의 확률 값으로 표현할 수 있다. 예를 들어, 소프트맥스 함수는 입력된 값을 0~1사이의 값으로 모두 정규화하며 출력 값들의 총합은 항상 1이 되는 특성을 가진 함수일 수 있다. 도 1에서, 소프트맥스 계층(14)은 설명의 편의를 위해 전결합층(13)과 별도로 도시하였으나 전결합층(13)에 포함되는 것으로 해석할 수도 있다.

[56]

[57] 도 2는 본 발명의 일 실시예에 따른 3D CNN을 설명하기 위한 예시도이다.

[58] 3차원 컨볼루션 신경망(3-dimension Convolutional Neural Network, 이하 3D CNN)은 도 1에 따른 2차원 합성곱 신경망을 시간축으로 한 차원 확장시킨 인공 신경망으로 해석할 수 있다. 도 1에 따른 2차원 합성곱 신경망은 일반적으로 이미지를 입력으로 받고, 입력된 이미지 상의 공간적인 특성을 통해 입력된 이미지를 분류하거나 입력된 이미지 내부의 객체를 식별하는 등과 같은 용도로 주로 사용될 수 있다.

[59] 그러나, 2차원 합성곱 신경망은 시간 정보가 포함된 동영상 데이터를 처리할 수 없는 한계를 가진다. 반면 3D CNN은 동영상 데이터의 시간 성분까지 고려하여 합성곱 연산과 풀링 연산 등을 수행하므로, 동영상 데이터의 시간적 속성을 고려하여 특징을 추출할 수 있다.

[60] 구체적으로, 도 2를 참조하면, 먼저 시간축에 따른 복수의 프레임(또는 픽처)으로 구성되는 동영상 데이터인 입력 영상(20)을 복수의 영상 클립(21)으로 분류하고, 각각의 영상 클립을 3D CNN(22)에 대한 입력으로 사용할 수 있다. 이때, 영상 클립(21)은 미리 설정된 개수(3D CNN이 한번에 처리할 수 있는 프레임의 수)의 프레임으로 구성되는데, 예를 들어 영상 클립(21)은 시간축상으로 연속된 프레임들로 구성될 수 있다. 또한, 각 프레임(도 2에 따른 예시에서 $f=0, f=1$)은 K개의 채널로 구성되고 각 채널은 W·H의 해상도를 갖는 이미지로 구성될 수 있다. 예를 들어 각 프레임이 RGB 성분의 이미지라면, 채널은 R(Red), G(Green), B(Blue) 각각의 성분에 따라 3개일 수 있다.

[61] 3D CNN(22)의 구조는 기본적으로 도 1에 따른 2차원 합성곱 신경망과

동일하거나 유사하지만, 시간축에 따른 영상 데이터들을 모두 이용하는 점에서 차이가 있을 수 있다. 예를 들어, 3D CNN(22)의 컨볼루션 계층은 2차원 컨볼루션과 마찬가지로 필터가 이미지를 스캔하듯이 움직이면서 컨볼루션 연산을 하는데, 시간축으로도 스트라이드(stride) 값만큼 이동하여 컨볼루션 연산을 수행할 수 있다. 또한, 3D CNN(22)에 따른 풀링 계층은 도 1에서 설명한 풀링 계층(12)을 시간축으로 한 차원 확장된 형태로, 시간축에 따른 픽셀값들을 모두 이용할 수 있다. 3D CNN(22)에 따른 전결합 계층은 도 1에 따른 전결합 계층(13)과 마찬가지로 마지막 특성맵에 존재하는 모든 픽셀을 벡터화하여 파라미터와의 가중합을 구하며, 3D CNN(22)에 따른 소프트맥스 계층은 도 1에 따른 소프트맥스 계층(14)과 마찬가지로 동작할 수 있다.

[62] 이처럼, 3D CNN(22)은 시간 축상의 영상 데이터를 함께 고려하여 학습을 수행하기 때문에 시간적으로 변화하는 사람의 동작을 학습하는데 유리할 수 있다. 다만, 시간 축상의 영상 데이터들을 함께 고려해야 하므로, 2차원 합성곱 신경망보다 더 많은 파라미터와 연산량이 필요한 문제가 있다.

[63] 따라서, 본 발명에서는 3D CNN에 따른 연산량을 줄이고, 고속으로 영상 인식을 수행할 수 있는 방법을 제안한다.

[64]

[65] 일반적인 3차원 CNN에서는 도 2와 같이 입력 영상을 구성하는 모든 영상 클립들에 대하여 동일한 3D CNN을 통해 소프트맥스 값을 산출하고, 산출된 소프트맥스 값을 이용하여 영상을 인식한다. 그러나, 모든 영상 클립에 대해서 3D CNN을 통한 소프트맥스 값을 산출할 경우 연산량이 많기 때문에 연산속도가 저하되는 문제가 있다. 특히, 제한적인 연산 자원만을 사용할 수 있는 소형 단말에서는 과도한 연산량을 감당하기 어렵기 때문에 연산량을 줄이고 고속으로 영상을 인식할 수 있는 방법이 요구된다.

[66] 본 발명의 일 실시예에서는 이러한 문제점을 해결하기 위한 수단으로서, 스코어 마진(score margin)이라는 개념을 정의할 수 있다. 스코어 마진(score margin)은 다음의 수학적 식 1로 정의할 수 있다.

[67] [수식 1]

$$score\ margin = V_{softmax1} - V_{softmax2}$$

[68] 상기 수학적 식 1을 참조하면, 스코어 마진(score margin)은 지금까지 3D CNN을 통해 각 영상 클립에 대하여 소프트맥스 함수를 산출한 결과값들 중에서 가장 큰 값($V_{softmax1}$)과 두 번째로 큰 값($V_{softmax2}$) 사이의 차분값으로 정의할 수 있다. 이때, 소프트 맥스 함수를 통하여 산출된 값은 0과 1 사이의 값을 가지므로, 스코어 마진도 0과 1 사이의 값을 가질 수 있다.

[69] 수학적 식 1에 따른 스코어 마진이 영상 인식의 성공과 실패에 얼마나 영향이 있는지 파악하기 위하여, 도 3a 및 도 3b와 같이, UCF101 데이터 셋을 대상으로 스코어 마진의 결과 그래프를 산출하였다.

[70]

[71] 도 3a은 본 발명의 일 실시예에 따른 스코어 마진값을 설명하기 위한 히스토그램이다.

[72] 도 3a를 참조하면, 영상 인식이 성공한 경우에 따른 영상 데이터들(세로축)의 스코어 마진값(가로축)에 대한 분포를 확인할 수 있는데, 스코어 마진값이 0.9와 1사이의 값을 갖는 데이터들이 월등히 많은 것을 알 수 있다.

[73]

[74] 도 3b는 본 발명의 다른 실시예에 따른 스코어 마진값을 설명하기 위한 히스토그램이다.

[75] 도 3b를 참조하면, 영상 인식이 실패한 경우에 따른 영상 데이터들(세로축)의 스코어 마진값(가로축)에 대한 분포를 확인할 수 있는데, 스코어 마진값이 작은 쪽에 더 많은 데이터가 분포해 있는 것을 확인할 수 있다.

[76]

[77] 따라서, 도 3a와 도 3b를 종합하면, 스코어 마진값이 충분히 크다면 현재 3D CNN을 통해 분석한 영상 클립들만으로도 입력 영상에 대한 영상 인식이 성공한 것으로 판단할 수 있어 이후의 영상 클립들에 대해 추가로 영상 인식을 수행할 필요성이 낮다. 이하에서는, 현재까지 3D CNN에 영상 클립들을 입력하여 산출한 스코어 마진값을 평가하여 영상 인식이 성공한 것으로 판단된 경우, 후속 영상 클립들을 3D CNN에 입력하는 과정을 생략하거나, 연산 복잡도가 낮은 3D CNN을 이용하여 후속 영상 클립에 대한 분석을 수행하는 방법을 제안한다.

[78]

[79] 도 4는 본 발명의 제1 실시예에 따른 3D CNN을 이용한 고속 영상 인식 방법에 대한 흐름도이다.

[80] 도 4를 참조하면, 본 발명의 제1 실시예에 따른 3D CNN을 이용한 고속 영상 인식 방법은, 입력 영상을 구성하는 영상 클립들 중 제1 영상 클립들을 각각 3D CNN(3-dimension Convolutional Neural Network)에 입력하는 단계(S100), 상기 제1 영상 클립들 각각에 대하여 상기 3D CNN을 통해 소프트맥스 함수(softmax function)를 연산한 결과값들을 획득하는 단계(S110), 획득된 결과값들을 이용하여 스코어 마진(score margin)을 산출하는 단계(S120), 산출된 스코어 마진을 미리 설정된 임계값과 비교하는 단계(S130) 및 상기 비교하는 단계에 대한 응답으로, 상기 입력 영상을 구성하는 영상 클립들 중 상기 제1 영상 클립들을 제외한 나머지 영상 클립들을 상기 3D CNN에 입력할지 여부를 결정하는 단계(S140)를 포함할 수 있다.

[81] 여기서 제1 영상 클립들은, 3D CNN에 입력할 최초 하나의 영상 클립을 의미할 수도 있고, 최초 영상 클립부터 복수 개의 영상 클립을 의미할 수도 있다.

[82] 상기 스코어 마진은 상기 결과값들 중 가장 큰 값과 두번째로 큰 값 사이의 차분값일 수 있다. 예를 들어 스코어 마진은 수학식 1에 따라 정의할 수 있다.

[83] 상기 나머지 영상 클립들을 상기 3D CNN에 입력할지 여부를 결정하는

- 단계(S140)는, 상기 스코어 마진이 상기 임계값보다 크면, 상기 제1 영상 클립들 이후의 영상 클립을 상기 3D CNN에 입력하지 않고, 상기 결과값들만으로 상기 입력 영상에 대한 영상 인식을 수행하는 단계를 포함할 수 있다. 따라서, 제1 영상 클립들에 대한 3D CNN의 분석만으로 영상 인식 결과를 최종적으로 도출하고 제1 영상 클립들 이후의 영상 클립에 대한 분석은 생략할 수 있다.
- [84] 상기 나머지 영상 클립들을 상기 3D CNN에 입력할지 여부를 결정하는 단계(S140)는, 상기 스코어 마진이 상기 임계값보다 작으면, 상기 제1 영상 클립들 이후의 영상 클립을 상기 3D CNN에 입력하는 단계를 포함할 수 있다.
- [85] 따라서, 제1 영상 클립들 이후의 영상 클립들이 입력될 때마다 반복하여 스코어 마진을 구하고 임계값 비교를 수행함으로써, 다음 영상 클립을 입력할지 현재 단계에서 영상 인식 결과를 최종적으로 도출하고 영상 인식을 종료할지 여부를 결정할 수 있다.
- [86] 상기 결과값들을 획득하는 단계(S110)는, 상기 소프트맥스 함수를 연산하여 획득된 결과값들을 메모리(memory)에 누적하여 저장하는 단계를 더 포함할 수 있다. 즉, 결과값들은 계속하여 누적하여 저장함으로써, 저장된 결과값들에 다음 영상 클립이 3D CNN에 입력되어 연산된 소프트맥스 함수의 결과값을 추가로 포함하여 단계 S120에 따른 스코어 마진을 산출할 수 있다.
- [87] 상기 임계값은, 영상 인식을 수행하는 단말의 종류, 연산 능력, 입력 영상의 종류, 입력 영상의 해상도, 입력 영상을 구성하는 프레임 수 중 적어도 하나에 따라 결정될 수 있다.
- [88] 상기 입력 영상을 구성하는 영상 클립들 각각은, 상기 입력 영상을 구성하는 복수의 프레임 중에서 미리 설정된 개수의 시간적으로 연속한 프레임들로 구성될 수 있다.
- [89]
- [90] 도 5는 본 발명의 제2 실시예에 따른 3D CNN을 이용하는 고속 영상 인식 방법에 대한 흐름도이다.
- [91] 도 5를 참조하면, 본 발명의 제2 실시예에 따른 3D CNN(3-dimension Convolutional Neural Network)을 이용하는 고속 영상 인식 방법은, 입력 영상을 구성하는 영상 클립들 중 제1 영상 클립들을 각각 3D CNN(3-dimension Convolutional Neural Network)에 입력하는 단계(S200), 상기 제1 영상 클립들 각각에 대하여 상기 3D CNN을 통해 소프트맥스 함수(softmax function)를 연산한 결과값들을 획득하는 단계(S210), 획득된 결과값들을 이용하여 스코어 마진(score margin)을 산출하는 단계(S220), 산출된 스코어 마진을 미리 설정된 임계값과 비교하는 단계(S230) 및 상기 비교하는 단계에 대한 응답으로, 상기 입력 영상을 구성하는 영상 클립들 중 상기 제1 영상 클립들의 다음 영상 클립을 상기 3D CNN과 동일한 네트워크에 입력할지 여부를 결정하는 단계(S240)를 포함할 수 있다.
- [92] 여기서 제1 영상 클립들은, 3D CNN에 입력할 최초 하나의 영상 클립을 의미할

수도 있고, 최초 영상 클립부터 복수 개의 영상 클립을 의미할 수도 있다.

- [93] 상기 스코어 마진은 상기 결과값들 중 가장 큰 값과 두번째로 큰 값 사이의 차분값일 수 있다. 예를 들어 스코어 마진은 수학식 1에 따라 정의할 수 있다.
- [94] 상기 다음 영상 클립을 상기 3D CNN과 동일한 네트워크에 입력할지 여부를 결정하는 단계(S240)는, 상기 스코어 마진이 상기 임계값보다 크면, 상기 제1 영상 클립들의 다음 영상 클립을 상기 3D CNN과 동일하거나 상기 3D CNN보다 더 얇은 네트워크에 입력하는 단계를 포함할 수 있다. 즉, 스코어 마진이 임계값보다 크다면 현재 입력된 영상 클립들로 추론한 영상 인식 결과가 옳았을 가능성이 높기 때문에, 다음 영상 클립은 현재 추론에 사용한 3D CNN과 동일하거나 더 얇은 네트워크에 입력함으로써 연산 속도를 향상시킬 수 있다. 이때, 얇은 네트워크라는 의미는 컨볼루션 계층의 개수가 작거나 연산 복잡도가 낮은 네트워크를 의미할 수 있다.
- [95] 상기 다음 영상 클립을 상기 3D CNN과 동일한 네트워크에 입력할지 여부를 결정하는 단계(S240)는, 상기 스코어 마진이 상기 임계값보다 작으면, 상기 제1 영상 클립들의 다음 영상 클립을 상기 3D CNN보다 더 깊은 네트워크에 입력하는 단계를 포함할 수 있다. 즉, 스코어 마진이 임계값보다 작다면 현재 입력된 영상 클립들로 추론한 영상 인식 결과가 틀렸을 가능성이 높기 때문에, 다음 영상 클립은 현재 추론에 사용한 3D CNN보다 더 깊은 네트워크에 입력함으로써 연산 속도를 향상시킬 수 있다. 이때, 깊은 네트워크라는 의미는 컨볼루션 계층의 개수가 많거나 연산 복잡도가 높은 네트워크를 의미할 수 있다.
- [96] 상기 결과값들을 획득하는 단계(S210)는, 상기 소프트맥스 함수를 연산하여 획득된 결과값들을 메모리(memory)에 누적하여 저장하는 단계를 더 포함할 수 있다. 즉, 결과값들은 계속하여 누적하여 저장하고, 기존에 저장한 결과값들에 다음 영상 클립에 대한 소프트맥스 함수의 결과값을 추가로 포함함으로써, 단계 S220에 따른 스코어 마진을 산출할 수 있다.
- [97] 상기 임계값은, 영상 인식을 수행하는 단말의 종류, 연산 능력, 입력 영상의 종류, 입력 영상의 해상도, 입력 영상을 구성하는 프레임 수 중 적어도 하나에 따라 결정될 수 있다.
- [98] 상기 입력 영상을 구성하는 영상 클립들 각각은, 상기 입력 영상을 구성하는 복수의 프레임 중에서 미리 설정된 개수의 시간적으로 연속한 프레임들로 구성될 수 있다.
- [99] 단계 S240에서 다음 영상 클립을 입력할 네트워크가 결정되었다면, 제1 영상 클립들의 다음 영상을 단계 S240에서 결정된 네트워크에 입력하여 단계 S210부터 단계 S240에 따른 과정을 반복함으로써, 입력 영상을 구성하는 영상 클립 모두에 대한 네트워크를 동적으로 결정할 수 있다.
- [100] 또한, 단계 S240에서 다음 영상 클립이 마지막 영상 클립인 경우에는 마지막 영상 클립을 단계 S240에서 결정된 네트워크에 입력하여 소프트맥스 함수를

연산하고, 그동안 연산한 소프트맥스 함수의 결과값들을 종합하여 최종적으로 영상 인식 결과를 도출함으로써 영상 인식을 종료할 수 있다.

[101] 한편, 도 4 및 도 5에 따른 제1 실시예와 제2 실시예는 서로 결합되어 제3 실시예로 구현될 수 있다.

[102] 실시예에 따라 보다 구체적으로 설명하면, 제1 실시예와 제2 실시예 모두 본 발명에서 정의하는 스코어 마진을 기반으로 이후의 연산을 생략하거나 적용할 네트워크를 달리할 수 있다. 따라서, 제1 실시예에 따른 단계 S100 내지 S120를 통해 산출된 스코어 마진을 제1 임계값과 비교함으로써 제1 실시예에 따른 단계 S140을 적용하고, 단계 S100 내지 S120를 통해 산출된 스코어 마진을 제2 임계값과 비교함으로써 제2 실시예에 따른 단계 S240을 적용하는 제3 실시예를 수행할 수 있다. 이때, 제1 임계값과 제2 임계값은 서로 다른 값으로 설정할 수 있으나, 동일한 값으로 설정하는 것을 배제하지 않는다. 상기 제3 실시예는 하기도 6을 참조하여 보다 구체적으로 기술하겠다.

[103]

[104] 도 6은 본 발명의 제3 실시예에 따른 3D CNN을 이용한 고속 영상 인식 방법에 대한 흐름도이다.

[105] 도 6을 참조하면, 본 발명의 제3 실시예에 따른 3D CNN을 이용한 고속 영상 인식 방법은, 입력 영상을 구성하는 영상 클립들 중 제1 영상 클립들을 각각 3D CNN(3-dimension Convolutional Neural Network)에 입력하는 단계(S300), 상기 제1 영상 클립들 각각에 대하여 상기 3D CNN을 통해 소프트맥스 함수(softmax function)를 연산한 결과값들을 획득하는 단계(S310), 획득된 결과값들을 이용하여 스코어 마진(score margin)을 산출하는 단계(S320), 산출된 스코어 마진을 미리 설정된 제1 임계값과 비교하여, 상기 입력 영상을 구성하는 영상 클립들 중 상기 제1 영상 클립들을 제외한 나머지 영상 클립들을 상기 3D CNN에 입력할지 여부를 결정하는 단계(S330) 및 산출된 스코어 마진을 미리 설정된 제2 임계값과 비교하여, 상기 입력 영상을 구성하는 영상 클립들 중 상기 제1 영상 클립들의 다음 영상 클립을 상기 3D CNN과 동일한 네트워크에 입력할지 여부를 결정하는 단계(S340)를 포함할 수 있다.

[106] 여기서 제1 영상 클립들은, 3D CNN에 입력할 최초 하나의 영상 클립을 의미할 수도 있고, 최초 영상 클립부터 복수 개의 영상 클립을 의미할 수도 있다.

[107] 또한, 상기 스코어 마진은 상기 결과값들 중 가장 큰 값과 두번째로 큰 값 사이의 차분값일 수 있다. 예를 들어, 스코어 마진은 앞서 개시된 수학식 1에 따라 정의할 수 있다.

[108] 상기 나머지 영상 클립들을 상기 3D CNN에 입력할지 여부를 결정하는 단계(S330) 시 상기 스코어 마진이 상기 제1 임계값보다 크면, 상기 제1 영상 클립들 이후의 영상 클립을 상기 3D CNN에 입력하지 않고, 상기 결과값들만으로 상기 입력 영상에 대한 영상 인식을 수행하는 단계를 포함할 수 있다. 따라서, 제1 영상 클립들에 대한 3D CNN의 분석만으로 영상 인식 결과를

최종적으로 도출하고 제1 영상 클립들 이후의 영상 클립에 대한 분석은 생략할 수 있다.

- [109] 반면, 상기 나머지 영상 클립들을 상기 3D CNN에 입력할지 여부를 결정하는 단계(S330) 시 상기 스코어 마진이 상기 제1 임계값보다 작으면, 상기 제1 영상 클립들 이후의 영상 클립을 상기 3D CNN에 입력하는 단계를 포함할 수 있다. 따라서, 제1 영상 클립들 이후의 영상 클립들이 입력될 때마다 반복하여 스코어 마진을 구하고 임계값 비교를 수행함으로써, 다음 영상 클립을 입력할지 현재 단계에서 영상 인식 결과를 최종적으로 도출하고 영상 인식을 종료할지 여부를 결정할 수 있다.
- [110] 또한, 상기 다음 영상 클립을 상기 3D CNN과 동일한 네트워크에 입력할지 여부를 결정하는 단계(S340) 시, 상기 스코어 마진이 상기 제2 임계값보다 크면, 상기 제1 영상 클립들의 다음 영상 클립을 상기 3D CNN과 동일하거나 상기 3D CNN보다 더 얇은 네트워크에 입력하는 단계를 포함할 수 있다. 즉, 앞서 도 5에서 설명한 바와 같이, 스코어 마진이 상기 제2 임계값보다 크다면 현재 입력된 영상 클립들로 추론한 영상 인식 결과가 옳았을 가능성이 높기 때문에, 다음 영상 클립은 현재 추론에 사용한 3D CNN과 동일하거나 더 얇은 네트워크에 입력함으로써 연산 속도를 향상시킬 수 있다. 이때, 얇은 네트워크라는 의미는 컨볼루션 계층의 개수가 작거나 연산 복잡도가 낮은 네트워크를 의미할 수 있다.
- [111] 반면, 상기 다음 영상 클립을 상기 3D CNN과 동일한 네트워크에 입력할지 여부를 결정하는 단계(S340) 시, 상기 스코어 마진이 상기 제2 임계값보다 작으면, 상기 제1 영상 클립들의 다음 영상 클립을 상기 3D CNN보다 더 깊은 네트워크에 입력하는 단계를 포함할 수 있다. 앞서 도 5에서 설명한 바와 마찬가지로, 스코어 마진이 상기 제2 임계값보다 작다면 현재 입력된 영상 클립들로 추론한 영상 인식 결과가 틀렸을 가능성이 높기 때문에, 다음 영상 클립은 현재 추론에 사용한 3D CNN보다 더 깊은 네트워크에 입력함으로써 연산 속도를 향상시킬 수 있다. 이때, 깊은 네트워크라는 의미는 컨볼루션 계층의 개수가 많거나 연산 복잡도가 높은 네트워크를 의미할 수 있다.
- [112] 상기 결과값들을 획득하는 단계(S110)는, 상기 소프트맥스 함수를 연산하여 획득된 결과값들을 메모리(memory)에 누적하여 저장하는 단계를 더 포함할 수 있다. 즉, 결과값들은 계속하여 누적하여 저장함으로써, 저장된 결과값들에 다음 영상 클립이 3D CNN에 입력되어 연산된 소프트맥스 함수의 결과값을 추가로 포함하여 단계 S120에 따른 스코어 마진을 산출할 수 있다.
- [113] 상기 임계값은, 영상 인식을 수행하는 단말의 종류, 연산 능력, 입력 영상의 종류, 입력 영상의 해상도, 입력 영상을 구성하는 프레임 수 중 적어도 하나에 따라 결정될 수 있다.
- [114] 상기 입력 영상을 구성하는 영상 클립들 각각은, 상기 입력 영상을 구성하는 복수의 프레임 중에서 미리 설정된 개수의 시간적으로 연속한 프레임들로

구성될 수 있다.

[115]

[116] 도 7은 본 발명의 제1 내지 제3 실시예에 따른 3D CNN을 이용한 고속 영상 인식 장치에 대한 구성도이다.

[117] 도 7을 참조하면, 본 발명의 제1 내지 제3 실시예에 따른 3D CNN을 이용한 고속 영상 인식 장치(100)는, 적어도 하나의 프로세서(processor, 110), 및 상기 적어도 하나의 프로세서(110)가 적어도 하나의 단계를 수행하도록 지시하는 명령어들(instructions)을 저장하는 메모리(memory, 120)를 포함할 수 있다.

[118] 여기서 적어도 하나의 프로세서(110)는 중앙 처리 장치(central processing unit, CPU), 그래픽 처리 장치(graphics processing unit, GPU), 또는 본 발명의 실시예들에 따른 방법들이 수행되는 전용의 프로세서를 의미할 수 있다.

[119] 예를 들어, 본 발명의 제1 실시예 및 제2 실시예에 따른 3D CNN을 이용한 고속 영상 인식 장치의 프로세서(110)가 수행하는 적어도 하나의 단계는, 입력 영상을 구성하는 영상 클립들 중 제1 영상 클립들을 각각 3D CNN(3-dimension Convolutional Neural Network)에 입력하는 단계, 상기 제1 영상 클립들 각각에 대하여 상기 3D CNN을 통해 소프트맥스 함수(softmax function)를 연산한 결과값들을 획득하는 단계, 획득된 결과값들을 이용하여 스코어 마진(score margin)을 산출하는 단계, 산출된 스코어 마진을 미리 설정된 임계값과 비교하여, 상기 입력 영상을 구성하는 영상 클립들 중 상기 제1 영상 클립들을 제외한 적어도 하나의 영상 클립의 추가 입력 여부를 결정하는 단계를 포함할 수 있다.

[120] 여기서, 상기 스코어 마진은 상기 결과값들 중 가장 큰 값과 두번째로 큰 값 사이의 차분값일 수 있다.

[121] 제1 실시예에 따르면, 상기 적어도 하나의 영상 클립의 추가 입력 여부 결정하는 단계는, 상기 제1 영상 클립들을 제외한 나머지 영상 클립들을 상기 3D CNN에 입력할지 여부를 결정하는 단계를 포함할 수 있다.

[122] 상기 나머지 영상 클립들을 상기 3D CNN에 입력할지 여부를 결정하는 단계는 상기 스코어 마진이 상기 임계값보다 클 경우, 프로세서(110)는 상기 제1 영상 클립들 이후의 영상 클립을 상기 3D CNN에 입력하지 않고, 상기 결과값들만으로 상기 입력 영상에 대한 영상 인식을 수행하는 단계를 포함할 수 있다.

[123] 한편, 상기 스코어 마진이 상기 임계값보다 작을 경우, 프로세서(110)는 상기 제1 영상 클립들 이후의 영상 클립을 상기 3D CNN에 입력하는 단계를 포함할 수 있다.

[124] 제2 실시예에 따르면, 상기 적어도 하나의 영상 클립의 추가 입력 여부를 결정하는 단계는, 상기 제1 영상 클립들의 다음 영상 클립을 상기 3D CNN과 동일한 네트워크에 입력할지 여부를 결정하는 단계를 포함할 수 있다.

[125] 여기서, 상기 다음 영상 클립을 상기 3D CNN과 동일한 네트워크에 입력할지 여부를 결정하는 단계는 상기 스코어 마진이 상기 임계값보다 클 경우, 상기 제1

영상 클립들의 다음 영상 클립을 상기 3D CNN과 동일하거나 상기 3D CNN보다 더 얇은 네트워크에 입력하는 단계를 포함할 수 있다.

- [126] 한편, 상기 다음 영상 클립을 상기 3D CNN과 동일한 네트워크에 입력할지 여부를 결정하는 단계는 상기 스코어 마진이 상기 임계값보다 작을 경우, 상기 제1 영상 클립의 다음 영상 클립을 상기 3D CNN보다 더 깊은 네트워크에 입력하는 단계를 포함할 수 있다.
- [127] 본 발명의 제1 실시예 및 제2 실시예에 따른 3D CNN을 이용한 고속 영상 인식 장치의 프로세서(110)의 적어도 하나의 단계 중 상기 결과값들을 획득하는 단계는, 상기 소프트맥스 함수를 연산하여 획득된 결과값들을 메모리(memory)에 누적하여 저장하는 단계를 더 포함할 수 있다.
- [128] 또한, 상기 임계값은, 영상 인식을 수행하는 단말의 종류, 연산 능력, 입력 영상의 종류, 입력 영상의 해상도, 입력 영상을 구성하는 프레임 수 중 적어도 하나에 따라 결정될 수 있다.
- [129] 상기 입력 영상을 구성하는 영상 클립들 각각은, 상기 입력 영상을 구성하는 복수의 프레임 중에서 미리 설정된 개수의 시간적으로 연속한 프레임들로 구성될 수 있다.
- [130] 한편, 본 발명의 제3 실시예에 따른 3D CNN을 이용한 고속 영상 인식 장치의 프로세서(110)가 수행하는 적어도 하나의 단계는, 입력 영상을 구성하는 영상 클립들 중 제1 영상 클립들을 각각 3D CNN(3-dimension Convolutional Neural Network)에 입력하는 단계, 상기 제1 영상 클립들 각각에 대하여 상기 3D CNN을 통해 소프트맥스 함수(softmax function)를 연산한 결과값들을 획득하는 단계, 획득된 결과값들을 이용하여 스코어 마진(score margin)을 산출하는 단계, 산출된 스코어 마진을 미리 설정된 제1 임계값과 비교하여, 상기 입력 영상을 구성하는 영상 클립들 중 상기 제1 영상 클립들을 제외한 나머지 영상 클립들을 상기 3D CNN과 동일한 네트워크에 입력할지 여부를 결정하는 단계 및 상기 스코어 마진을 미리 설정된 제2 임계값과 비교하여, 상기 제1 영상 클립들의 다음 영상 클립을 상기 3D CNN과 동일한 네트워크에 입력할지 여부를 결정하는 단계를 포함할 수 있다.
- [131] 여기서, 상기 스코어 마진은 상기 결과값들 중 가장 큰 값과 두번째로 큰 값 사이의 차분값일 수 있다.
- [132] 또한, 산출된 스코어 마진을 미리 설정된 제1 임계값과 비교하여, 상기 입력 영상을 구성하는 영상 클립들 중 상기 제1 영상 클립들을 제외한 나머지 영상 클립들을 상기 3D CNN과 동일한 네트워크에 입력할지 여부를 결정하는 단계는 상기 스코어 마진이 상기 제1 임계값보다 클 경우, 상기 제1 영상 클립들 이후의 영상 클립을 상기 3D CNN에 입력하지 않고, 상기 결과값들만으로 상기 입력 영상에 대한 영상 인식을 수행하는 단계를 포함할 수 있다.
- [133] 한편, 상기 스코어 마진이 상기 제1 임계값보다 작을 경우, 상기 제1 영상 클립들 이후의 영상 클립을 상기 3D CNN에 입력하는 단계를 포함할 수 있다.

- [134] 또한, 상기 스코어 마진을 미리 설정된 제2 임계값과 비교하여, 상기 제1 영상 클립들의 다음 영상 클립을 상기 3D CNN과 동일한 네트워크에 입력할지 여부를 결정하는 단계 시 상기 제1 영상 클립들의 다음 영상 클립을 상기 3D CNN과 동일한 네트워크에 입력할지 여부를 결정하는 단계를 포함할 수 있다.
- [135] 여기서, 상기 다음 영상 클립을 상기 3D CNN과 동일한 네트워크에 입력할지 여부를 결정하는 단계 시 상기 스코어 마진이 상기 제2 임계값보다 클 경우, 상기 제1 영상 클립들의 다음 영상 클립을 상기 3D CNN과 동일하거나 상기 3D CNN보다 더 얇은 네트워크에 입력하는 단계를 포함할 수 있다.
- [136] 한편, 상기 스코어 마진이 상기 제2 임계값보다 작을 경우, 상기 제1 영상 클립의 다음 영상 클립을 상기 3D CNN보다 더 깊은 네트워크에 입력하는 단계를 포함할 수 있다.
- [137] 메모리(120) 및 저장 장치(160) 각각은 휘발성 저장 매체 및 비휘발성 저장 매체 중에서 적어도 하나로 구성될 수 있다. 예를 들어, 메모리(120)는 읽기 전용 메모리(read only memory, ROM) 및 랜덤 액세스 메모리(random access memory, RAM) 중에서 적어도 하나로 구성될 수 있다.
- [138] 또한, 3D CNN을 이용한 고속 영상 인식 장치(100)는 무선 네트워크를 통해 통신을 수행하는 송수신 장치(transceiver)(130)를 포함할 수 있으며, 입력 인터페이스 장치(140), 출력 인터페이스 장치(150), 저장 장치(160) 등을 더 포함할 수 있다. 3D CNN을 이용한 고속 영상 인식 장치(100)에 포함된 각각의 구성 요소들은 버스(bus)(170)에 의해 연결되어 서로 통신을 수행할 수 있다.
- [139] 3D CNN을 이용한 고속 영상 인식 장치(100)는, 예를 들면, 통신 가능한 데스크탑 컴퓨터(desktop computer), 랩탑 컴퓨터(laptop computer), 노트북(notebook), 스마트폰(smart phone), 태블릿 PC(tablet PC), 모바일폰(mobile phone), 스마트 워치(smart watch), 스마트 글래스(smart glass), e-book 리더기, PMP(portable multimedia player), 휴대용 게임기, 네비게이션(navigation) 장치, 디지털 카메라(digital camera), DMB(digital multimedia broadcasting) 재생기, 디지털 음성 녹음기(digital audio recorder), 디지털 음성 재생기(digital audio player), 디지털 동영상 녹화기(digital video recorder), 디지털 동영상 재생기(digital video player), PDA(Personal Digital Assistant) 등으로 제공될 수 있다.
- [140]
- [141] 본 발명에 따른 방법들은 다양한 컴퓨터 수단을 통해 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 컴퓨터 판독 가능 매체에 기록되는 프로그램 명령은 본 발명을 위해 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다.
- [142] 컴퓨터 판독 가능 매체의 예에는 롬(ROM), 램(RAM), 플래시 메모리(flash

memory) 등과 같이 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함될 수 있다. 프로그램 명령의 예에는 컴파일러(compiler)에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터(interpreter) 등을 사용해서 컴퓨터에 의해 실행될 수 있는 고급 언어 코드를 포함할 수 있다. 상술한 하드웨어 장치는 본 발명의 동작을 수행하기 위해 적어도 하나의 소프트웨어 모듈로 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

[143] 또한, 상술한 방법 또는 장치는 그 구성이나 기능의 전부 또는 일부가 결합되어 구현되거나, 분리되어 구현될 수 있다.

[144] 상기에서는 본 발명의 바람직한 실시예를 참조하여 설명하였지만, 해당 기술 분야의 숙련된 당업자는 하기의 특허 청구의 범위에 기재된 본 발명의 사상 및 영역으로부터 벗어나지 않는 범위 내에서 본 발명을 다양하게 수정 및 변경시킬 수 있음을 이해할 수 있을 것이다.

청구범위

- [청구항 1] 3D CNN(3-dimension Convolutional Neural Network)을 이용한 고속 영상 인식 방법으로,
입력 영상을 구성하는 영상 클립들 중 제1 영상 클립들을 각각 3D CNN에 입력하는 단계;
상기 제1 영상 클립들 각각에 대하여 상기 3D CNN을 통해 소프트맥스 함수(softmax function)를 연산한 결과값들을 획득하는 단계;
획득된 결과값들을 이용하여 스코어 마진(score margin)을 산출하는 단계;
및
산출된 스코어 마진을 미리 설정된 임계값과 비교하여, 상기 입력 영상을 구성하는 영상 클립들 중 상기 제1 영상 클립들을 제외한 적어도 하나의 영상 클립의 추가 입력 여부를 결정하는 단계를 포함하는, 고속 영상 인식 방법.
- [청구항 2] 청구항 1에서,
상기 스코어 마진은 상기 결과값들 중 가장 큰 값과 두번째로 큰 값 사이의 차분값인, 고속 영상 인식 방법.
- [청구항 3] 청구항 1에서,
상기 적어도 하나의 영상 클립의 추가 입력 여부를 결정하는 단계는,
상기 제1 영상 클립들을 제외한 나머지 영상 클립들을 상기 3D CNN에 입력할지 여부를 결정하는 단계를 포함하는, 고속 영상 인식 방법.
- [청구항 4] 청구항 3에서,
상기 나머지 영상 클립들을 상기 3D CNN에 입력할지 여부를 결정하는 단계는,
상기 스코어 마진이 상기 임계값보다 클 경우, 상기 제1 영상 클립들 이후의 영상 클립을 상기 3D CNN에 입력하지 않고, 상기 결과값들만으로 상기 입력 영상에 대한 영상 인식을 수행하는 단계; 및
상기 스코어 마진이 상기 임계값보다 작을 경우, 상기 제1 영상 클립들 이후의 영상 클립을 상기 3D CNN에 입력하는 단계를 포함하는, 고속 영상 인식 방법.
- [청구항 5] 청구항 1에서,
상기 적어도 하나의 영상 클립의 추가 입력 여부를 결정하는 단계는,
상기 제1 영상 클립들의 다음 영상 클립을 상기 3D CNN과 동일한 네트워크에 입력할지 여부를 결정하는 단계를 포함하는, 고속 영상 인식 방법.
- [청구항 6] 청구항 5에서,
상기 다음 영상 클립을 상기 3D CNN과 동일한 네트워크에 입력할지 여부를 결정하는 단계는,

상기 스코어 마진이 상기 임계값보다 클 경우, 상기 제1 영상 클립들의 다음 영상 클립을 상기 3D CNN과 동일하거나 상기 3D CNN보다 더 얇은 네트워크에 입력하는 단계; 및
 상기 스코어 마진이 상기 임계값보다 작을 경우, 상기 제1 영상 클립의 다음 영상 클립을 상기 3D CNN보다 더 깊은 네트워크에 입력하는 단계를 포함하는, 고속 영상 인식 방법.

[청구항 7] 청구항 1항에서,
 상기 적어도 하나의 영상 클립의 추가 입력 여부를 결정하는 단계는, 상기 스코어 마진을 제1 임계값과 비교하여, 상기 제1 영상 클립들을 제외한 나머지 영상 클립들을 상기 3D CNN에 입력할지 여부를 결정하는 단계; 및
 상기 스코어 마진을 상기 제1 임계값과 다른 제2 임계값과 비교하여, 상기 제1 영상 클립들의 다음 영상 클립을 상기 3D CNN과 동일한 네트워크에 입력할지 여부를 결정하는 단계를 포함하는, 고속 영상 인식 방법.

[청구항 8] 청구항 1에서,
 상기 결과값들을 획득하는 단계는,
 상기 소프트맥스 함수를 연산하여 획득된 결과값들을 메모리(memory)에 누적하여 저장하는 단계를 더 포함하는, 고속 영상 인식 방법.

[청구항 9] 청구항 1에서,
 상기 임계값은,
 영상 인식을 수행하는 단말의 종류, 연산 능력, 입력 영상의 종류, 입력 영상의 해상도, 입력 영상을 구성하는 프레임 수 중 적어도 하나에 따라 결정되는, 고속 영상 인식 방법.

[청구항 10] 청구항 1에서,
 상기 입력 영상을 구성하는 영상 클립들 각각은,
 상기 입력 영상을 구성하는 복수의 프레임 중에서 미리 설정된 개수의 시간적으로 연속한 프레임들로 구성되는, 고속 영상 인식 방법.

[청구항 11] 3D CNN(3-dimension Convolutional Neural Network)을 이용한 고속 영상 인식 장치로서,
 적어도 하나의 프로세서(processor); 및
 상기 적어도 하나의 프로세서가 적어도 하나의 단계를 수행하도록 지시하는 명령어들(instructions)을 저장하는 메모리(memory)를 포함하고,
 상기 적어도 하나의 단계는,
 입력 영상을 구성하는 영상 클립들 중 제1 영상 클립들을 각각 3D CNN에 입력하는 단계;
 상기 제1 영상 클립들 각각에 대하여 상기 3D CNN을 통해 소프트맥스 함수(softmax function)를 연산한 결과값들을 획득하는 단계;
 획득된 결과값들을 이용하여 스코어 마진(score margin)을 산출하는 단계;

산출된 스코어 마진을 미리 설정된 임계값과 비교하여, 상기 입력 영상을 구성하는 영상 클립들 중 상기 제1 영상 클립들을 제외한 적어도 하나의 영상 클립의 추가 입력 여부를 결정하는 단계를 포함하는, 고속 영상 인식 장치.

[청구항 12] 청구항 11에서,

상기 스코어 마진은 상기 결과값들 중 가장 큰 값과 두번째로 큰 값 사이의 차분값인, 고속 영상 인식 장치.

[청구항 13] 청구항 11에서,

상기 적어도 하나의 영상 클립의 추가 입력 여부를 결정하는 단계는, 상기 제1 영상 클립들을 제외한 나머지 영상 클립들을 상기 3D CNN에 입력할지 여부를 결정하는 단계를 포함하는, 고속 영상 인식 장치.

[청구항 14] 청구항 13에서,

상기 나머지 영상 클립들을 상기 3D CNN에 입력할지 여부를 결정하는 단계는,

상기 스코어 마진이 상기 임계값보다 클 경우, 상기 제1 영상 클립들 이후의 영상 클립을 상기 3D CNN에 입력하지 않고, 상기 결과값들만으로 상기 입력 영상에 대한 영상 인식을 수행하는 단계; 및

상기 스코어 마진이 상기 임계값보다 작을 경우, 상기 제1 영상 클립들 이후의 영상 클립을 상기 3D CNN에 입력하는 단계를 포함하는, 고속 영상 인식 장치.

[청구항 15] 청구항 11에서,

상기 적어도 하나의 영상 클립의 추가 입력 여부를 결정하는 단계는, 상기 제1 영상 클립들의 다음 영상 클립을 상기 3D CNN과 동일한 네트워크에 입력할지 여부를 결정하는 단계를 포함하는, 고속 영상 인식 장치.

[청구항 16] 청구항 15에서,

상기 다음 영상 클립을 상기 3D CNN과 동일한 네트워크에 입력할지 여부를 결정하는 단계는,

상기 스코어 마진이 상기 임계값보다 클 경우, 상기 제1 영상 클립들의 다음 영상 클립을 상기 3D CNN과 동일하거나 상기 3D CNN보다 더 얇은 네트워크에 입력하는 단계; 및

상기 스코어 마진이 상기 임계값보다 작을 경우, 상기 제1 영상 클립의 다음 영상 클립을 상기 3D CNN보다 더 깊은 네트워크에 입력하는 단계를 포함하는, 고속 영상 인식 장치.

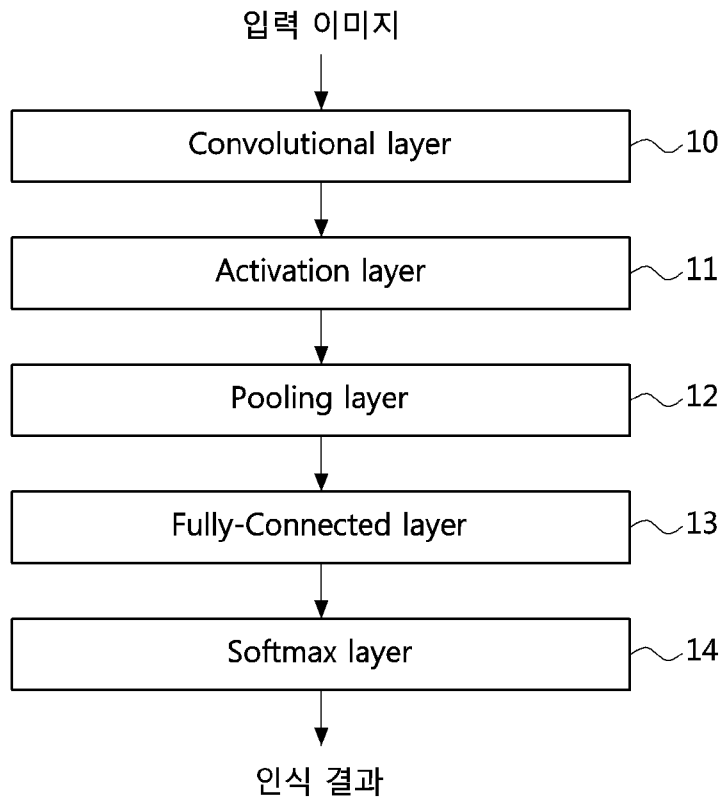
[청구항 17] 청구항 11에서,

상기 결과값들을 획득하는 단계는,

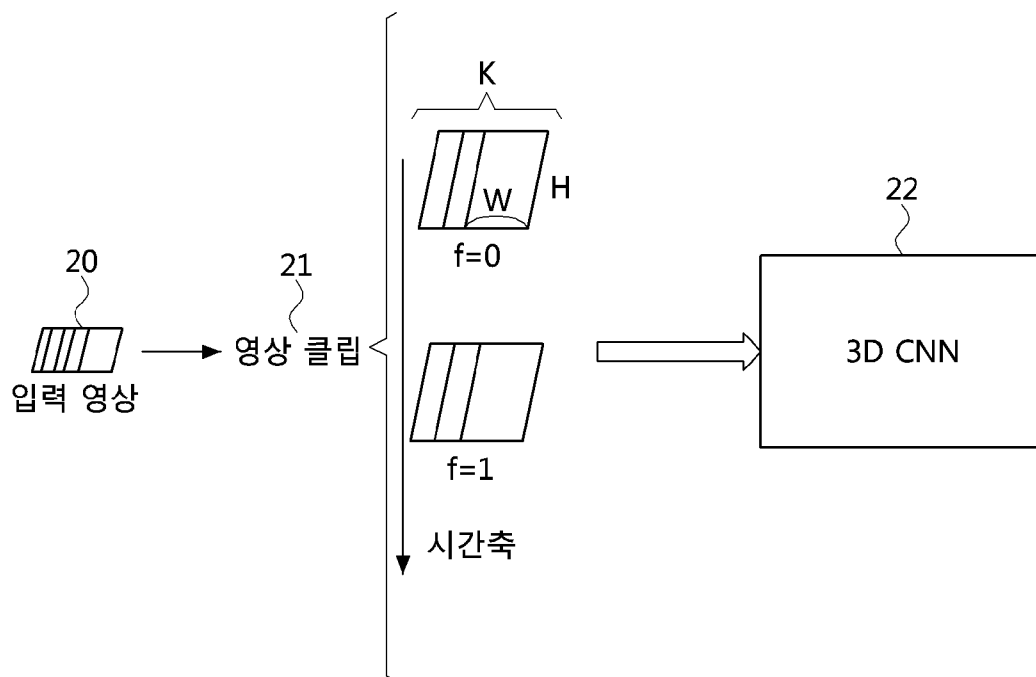
상기 소프트맥스 함수를 연산하여 획득된 결과값들을 메모리(memory)에 누적하여 저장하는 단계를 더 포함하는, 고속 영상 인식 장치.

- [청구항 18] 청구항 11에서,
상기 임계값은,
영상 인식을 수행하는 단말의 종류, 연산 능력, 입력 영상의 종류, 입력
영상의 해상도, 입력 영상을 구성하는 프레임 수 중 적어도 하나에 따라
결정되는, 고속 영상 인식 장치.
- [청구항 19] 청구항 11에서,
상기 입력 영상을 구성하는 영상 클립들 각각은,
상기 입력 영상을 구성하는 복수의 프레임 중에서 미리 설정된 개수의
시간적으로 연속한 프레임들로 구성되는, 고속 영상 인식 장치.
- [청구항 20] 3D CNN(3-dimension Convolutional Neural Network)을 이용한 고속 영상
인식 장치로서,
적어도 하나의 프로세서(processor); 및
상기 적어도 하나의 프로세서가 적어도 하나의 단계를 수행하도록
지시하는 명령어들(instructions)을 저장하는 메모리(memory)를 포함하고,
상기 적어도 하나의 단계는,
입력 영상을 구성하는 영상 클립들 중 제1 영상 클립들을 각각 3D CNN에
입력하는 단계;
상기 제1 영상 클립들 각각에 대하여 상기 3D CNN을 통해 소프트맥스
함수(softmax function)를 연산한 결과값들을 획득하는 단계;
획득된 결과값들을 이용하여 스코어 마진(score margin)을 산출하는 단계;
산출된 스코어 마진을 미리 설정된 제1 임계값과 비교하여, 상기 입력
영상을 구성하는 영상 클립들 중 상기 제1 영상 클립들을 제외한 나머지
영상 클립들을 상기 3D CNN과 동일한 네트워크에 입력할지 여부를
결정하는 단계; 및
상기 스코어 마진을 미리 설정된 제2 임계값과 비교하여, 상기 제1 영상
클립들의 다음 영상 클립을 상기 3D CNN과 동일한 네트워크에 입력할지
여부를 결정하는 단계를 포함하는, 고속 영상 인식 장치.

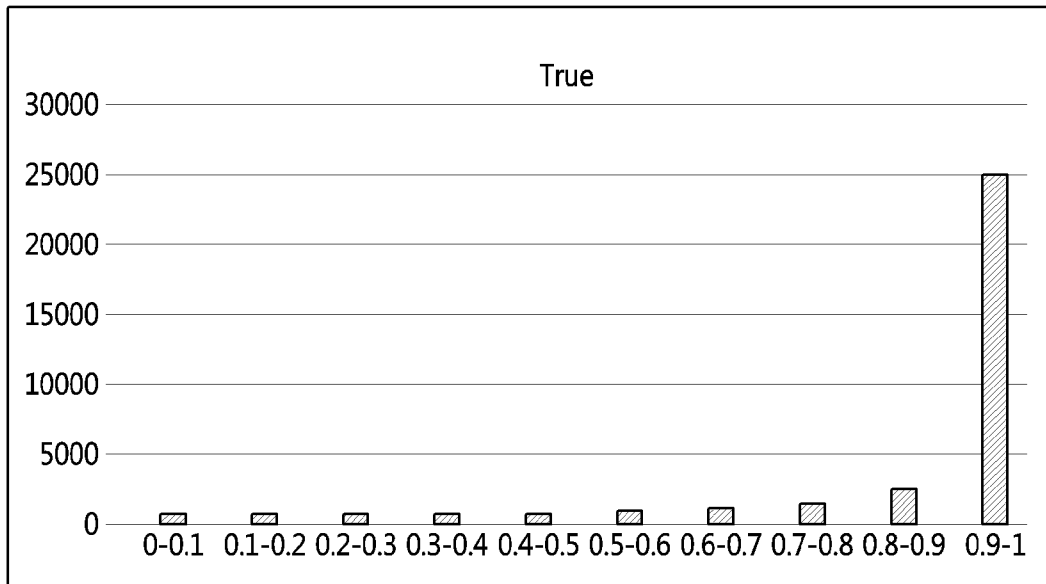
[도1]



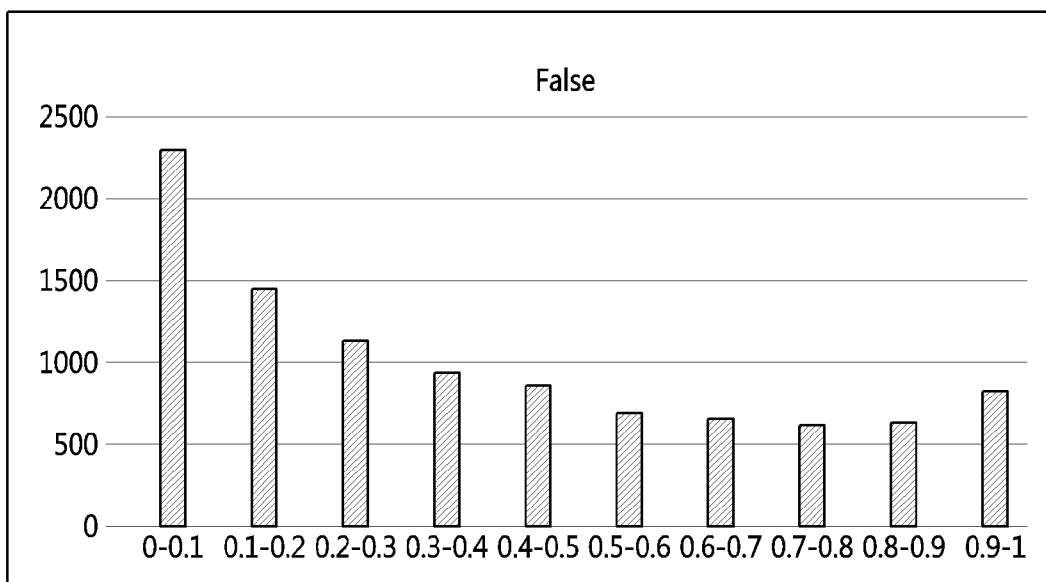
[도2]



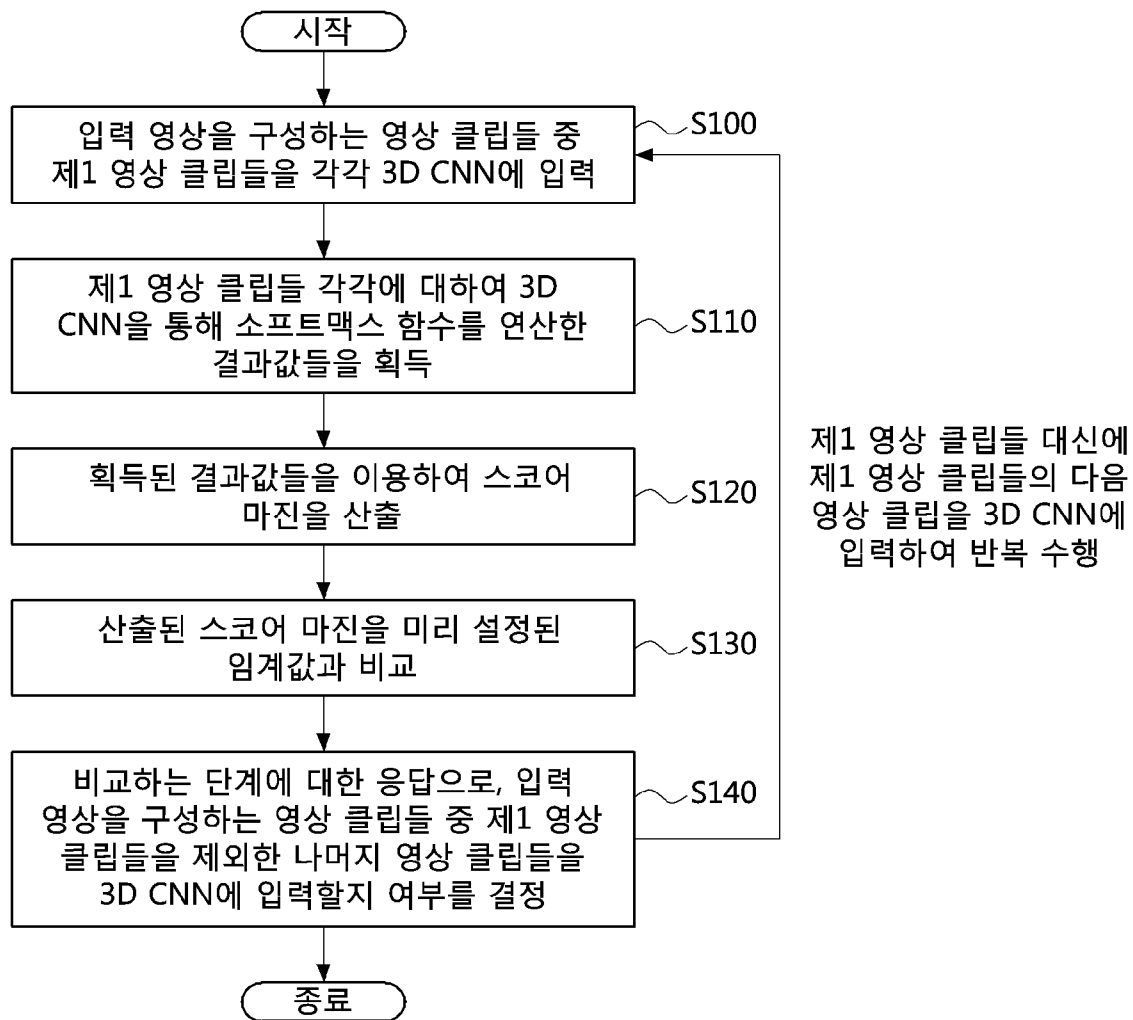
[도3a]



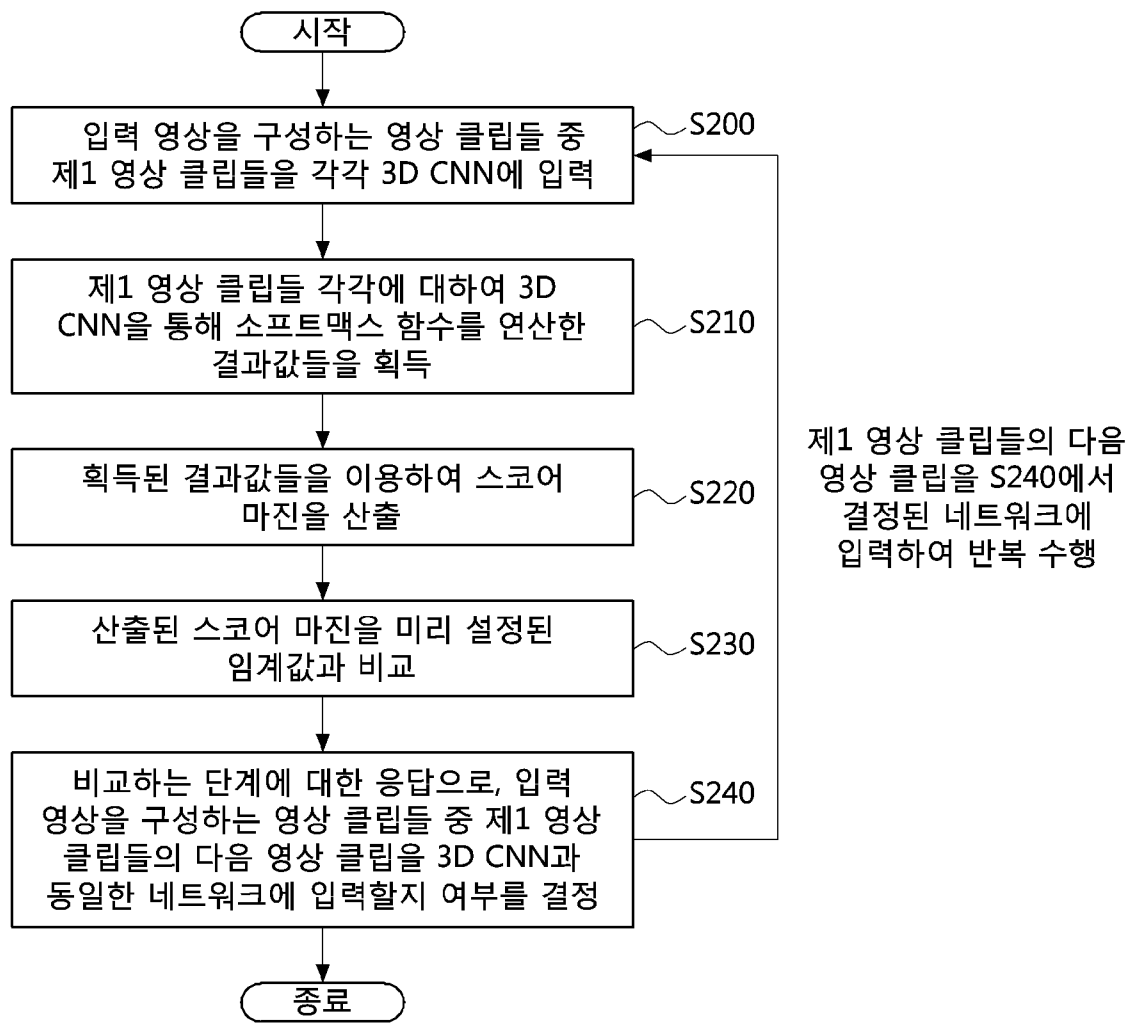
[도3b]



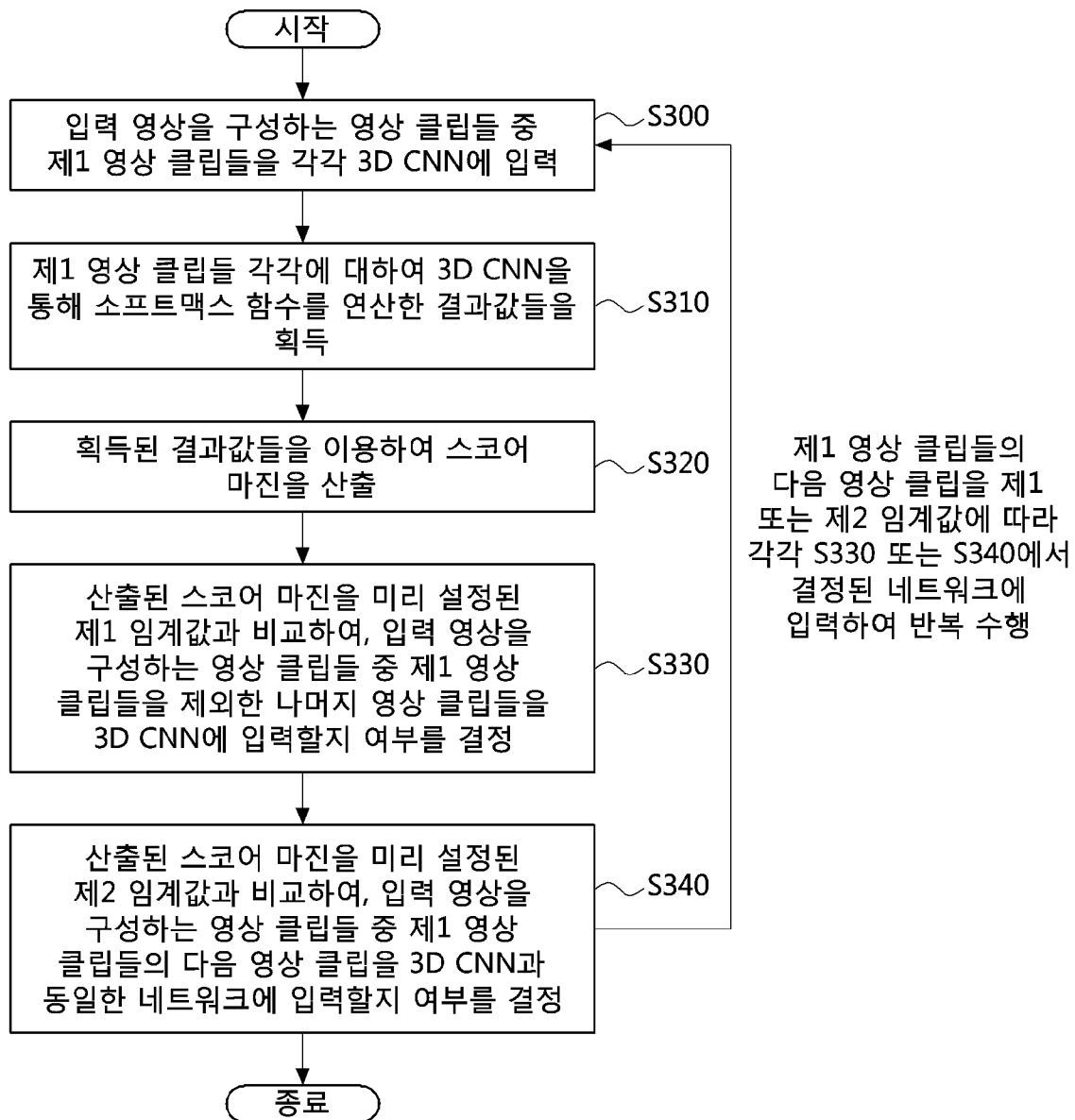
[도4]



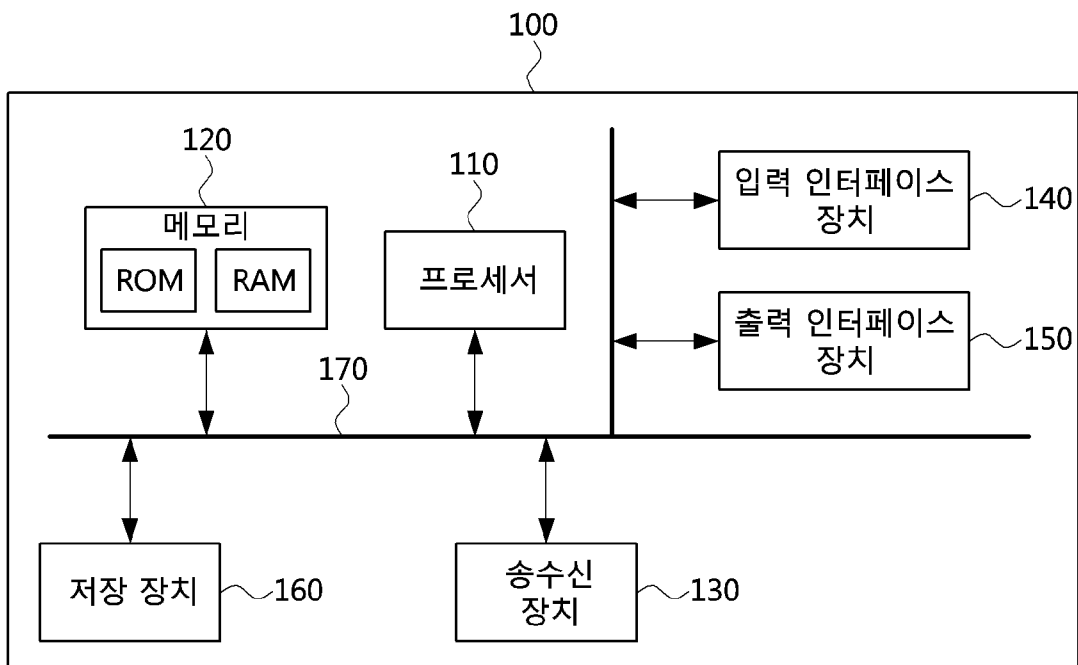
[도5]



[도6]



[도7]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2020/000647

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/00(2006.01)i, G06N 3/08(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06K 9/00; G01S 17/89; G06N 3/04; G06T 15/08; G06T 15/20; G06T 17/05; G16H 30/40; G06N 3/08

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean utility models and applications for utility models: IPC as above

Japanese utility models and applications for utility models: IPC as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS (KIPO internal) & Keywords: 3D, CNN, image clip, softmax function, score margin

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2017-538999 A (NOKIA TECHNOLOGIES OY.) 28 December 2017 See paragraphs [0026]-[0031] and figures 1-6.	1-20
A	KR 10-2017-0000748 A (SAMSUNG ELECTRONICS CO., LTD.) 03 January 2017 See claims 1-2 and figures 1-10.	1-20
A	KR 10-2016-0083900 A (FACEBOOK, INC.) 12 July 2016 See claims 1-3 and figures 1-7.	1-20
A	US 2018-0260997 A1 (SIEMENS HEALTHCARE GMBH.) 13 September 2018 See claim 1 and figures 1-4.	1-20
A	US 2018-0232947 A1 (VAYAVISION, LTD.) 16 August 2018 See claim 1 and figures 1A-8.	1-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

07 MAY 2020 (07.05.2020)

Date of mailing of the international search report

07 MAY 2020 (07.05.2020)

Name and mailing address of the ISA/KR

Korean Intellectual Property Office
Government Complex Daejeon Building 4, 189, Cheongsu-ro, Seo-gu,
Daejeon, 35208, Republic of Korea

Facsimile No. +82-42-481-8578

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/KR2020/000647

Patent document cited in search report	Publication date	Patent family member	Publication date
JP 2017-538999 A	28/12/2017	CN 107004138 A EP 3234867 A1 JP 6488380 B2 MX 2017008086 A US 10275688 B2 US 2017-0351936 A1 WO 2016-095117 A1	01/08/2017 25/10/2017 20/03/2019 31/10/2017 30/04/2019 07/12/2017 23/06/2016
KR 10-2017-0000748 A	03/01/2017	EP 3147827 A1	29/03/2017
KR 10-2016-0083900 A	12/07/2016	AU 2016-341919 A1 CA 2928932 A1 CN 105874474 A JP 2017-501514 A JP 6594329 B2 MX 2016005868 A MX 362983 B US 10095917 B2 US 2015-0125049 A1 US 2019-0171868 A1 WO 2015-066628 A1	19/05/2016 07/05/2015 17/08/2016 12/01/2017 23/10/2019 20/01/2017 28/02/2019 09/10/2018 07/05/2015 06/06/2019 07/05/2015
US 2018-0260997 A1	13/09/2018	EP 3373245 A2 EP 3373245 A3 US 10607393 B2	12/09/2018 14/11/2018 31/03/2020
US 2018-0232947 A1	16/08/2018	US 10445928 B2	15/10/2019

A. 발명이 속하는 기술분류(국제특허분류(IPC))

G06K 9/00(2006.01)i, G06N 3/08(2006.01)i

B. 조사된 분야

조사된 최소문헌(국제특허분류를 기재)

G06K 9/00; G01S 17/89; G06N 3/04; G06T 15/08; G06T 15/20; G06T 17/05; G16H 30/40; G06N 3/08

조사된 기술분야에 속하는 최소문헌 이외의 문헌

한국등록실용신안공보 및 한국공개실용신안공보: 조사된 최소문헌란에 기재된 IPC

일본등록실용신안공보 및 일본공개실용신안공보: 조사된 최소문헌란에 기재된 IPC

국제조사에 이용된 전산 데이터베이스(데이터베이스의 명칭 및 검색어(해당하는 경우))

eKOMPASS(특허청 내부 검색시스템) & 키워드: 3D, CNN, 영상 클립(image clip), 소프트맥스 함수(softmax fuction), 스코어 마진(score margin)

C. 관련 문헌

카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
A	JP 2017-538999 A (NOKIA TECHNOLOGIES OY) 2017.12.28 단락 [0026]-[0031] 및 도면 1-6	1-20
A	KR 10-2017-0000748 A (삼성전자주식회사) 2017.01.03 청구항 1-2 및 도면 1-10	1-20
A	KR 10-2016-0083900 A (페이스북, 인크.) 2016.07.12 청구항 1-3 및 도면 1-7	1-20
A	US 2018-0260997 A1 (SIEMENS HEALTHCARE GMBH) 2018.09.13 청구항 1 및 도면 1-4	1-20
A	US 2018-0232947 A1 (VAYAVISION, LTD.) 2018.08.16 청구항 1 및 도면 1A-8	1-20

☐ 추가 문헌이 C(계속)에 기재되어 있습니다.

☒ 대응특허에 관한 별지를 참조하십시오.

* 인용된 문헌의 특별 카테고리:

“A” 특별히 관련이 없는 것으로 보이는 일반적인 기술수준을 정의한 문헌

“D” 본 국제출원에서 출원인이 인용한 문헌

“E” 국제출원일보다 빠른 출원일 또는 우선일을 가지나 국제출원일 이후 “X”에 공개된 선출원 또는 특허 문헌

“L” 우선권 주장에 의문을 제기하는 문헌 또는 다른 인용문헌의 공개일 또는 다른 특별한 이유(이유를 명시)를 밝히기 위하여 인용된 문헌

“O” 구두 개시, 사용, 전시 또는 기타 수단을 언급하고 있는 문헌

“P” 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌

“T” 국제출원일 또는 우선일 후에 공개된 문헌으로, 출원과 상충하지 않으며 발명의 기초가 되는 원리나 이론을 이해하기 위해 인용된 문헌

“X” 특별한 관련이 있는 문헌. 해당 문헌 하나만으로 청구된 발명의 신규성 또는 진보성이 없는 것으로 본다.

“Y” 특별한 관련이 있는 문헌. 해당 문헌이 하나 이상의 다른 문헌과 조합하는 경우로 그 조합이 당업자에게 자명한 경우 청구된 발명은 진보성이 없는 것으로 본다.

“&” 동일한 대응특허문헌에 속하는 문헌

국제조사의 실제 완료일

2020년 05월 07일 (07.05.2020)

국제조사보고서 발송일

2020년 05월 07일 (07.05.2020)

ISA/KR의 명칭 및 우편주소



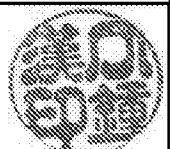
대한민국 특허청
(35208) 대전광역시 서구 청사로 189,
4동 (둔산동, 정부대전청사)

팩스 번호 +82-42-481-8578

심사관

정종한

전화번호 +82-42-481-5642



국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
JP 2017-538999 A	2017/12/28	CN 107004138 A EP 3234867 A1 JP 6488380 B2 MX 2017008086 A US 10275688 B2 US 2017-0351936 A1 WO 2016-095117 A1	2017/08/01 2017/10/25 2019/03/20 2017/10/31 2019/04/30 2017/12/07 2016/06/23
KR 10-2017-0000748 A	2017/01/03	EP 3147827 A1	2017/03/29
KR 10-2016-0083900 A	2016/07/12	AU 2016-341919 A1 CA 2928932 A1 CN 105874474 A JP 2017-501514 A JP 6594329 B2 MX 2016005868 A MX 362983 B US 10095917 B2 US 2015-0125049 A1 US 2019-0171868 A1 WO 2015-066628 A1	2016/05/19 2015/05/07 2016/08/17 2017/01/12 2019/10/23 2017/01/20 2019/02/28 2018/10/09 2015/05/07 2019/06/06 2015/05/07
US 2018-0260997 A1	2018/09/13	EP 3373245 A2 EP 3373245 A3 US 10607393 B2	2018/09/12 2018/11/14 2020/03/31
US 2018-0232947 A1	2018/08/16	US 10445928 B2	2019/10/15