



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0132375  
(43) 공개일자 2022년09월30일

(51) 국제특허분류(Int. Cl.)  
G06N 3/04 (2006.01) G06N 3/063 (2006.01)  
G06N 3/08 (2006.01)  
(52) CPC특허분류  
G06N 3/04 (2013.01)  
G06N 3/063 (2013.01)  
(21) 출원번호 10-2021-0037609  
(22) 출원일자 2021년03월23일  
심사청구일자 없음

(71) 출원인  
현대모비스 주식회사  
서울특별시 강남구 테헤란로 203 (역삼동)  
(72) 발명자  
이재영  
경기도 용인시 처인구 중부대로1158번길 12, 201  
동 1504호 (삼가동, 행정타운늘푸른오스카빌아파  
트)  
(74) 대리인  
특허법인지명

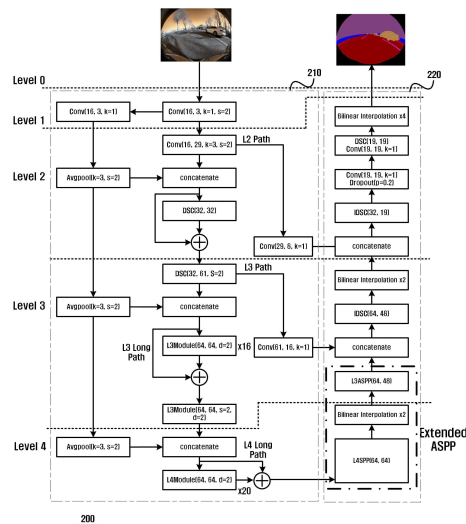
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 차량 영상의 화소를 분류하는 행렬 곱셈 가속기에 최적화된 임베디드 의미 분할 네트워크 장치

(57) 요약

차량 영상의 화소를 분류하는 행렬 곱셈 가속기(Matrixs Multiplication Accelerator)에 최적화된 임베디드 의미 분할 네트워크 장치가 제공된다. 상기 장치는 카메라를 통해 촬영된 영상을 수신하는 통신모듈, 상기 촬영된 영상의 컨텍스트를 추출하기 위한 의미 분할 네트워크(MMANet) 기반의 프로그램이 저장된 메모리 및 상기 메모리에 저장된 프로그램을 실행시키는 프로세서를 포함하되, 상기 프로세서는 상기 프로그램을 실행시킴에 따라, 상위 계층에서 하위 계층으로 이어지며 입력 영상의 특징을 축소시키는 CNN(Convolutional Neural Network) 및 DSC(Depthwise Seperable Convolution)로 구성되는 처리 모듈 중 적어도 하나를 포함하여 구성되는 인코더부 및 디코더부로 구성된 상기 의미 분할 네트워크의 각 계층에서의 활성화 맵의 크기에 따라 상기 CNN 처리 모듈 또는 DSC 처리 모듈을 선택하여 상기 촬영된 영상의 컨텍스트를 추출한다.

대표도 - 도5



(52) CPC특허분류  
*G06N 3/08* (2013.01)

---

## 명세서

### 청구범위

#### 청구항 1

차량 영상의 화소를 분류하는 행렬 곱셈 가속기(Matrix Multiplication Accelerator)에 최적화된 임베디드 의미 분할 네트워크 장치에 있어서,

카메라를 통해 촬영된 영상을 수신하는 통신모듈,

상기 촬영된 영상의 컨텍스트를 추출하기 위한 의미 분할 네트워크(MMNet) 기반의 프로그램이 저장된 메모리 및

상기 메모리에 저장된 프로그램을 실행시키는 프로세서를 포함하되,

상기 프로세서는 상기 프로그램을 실행시킴에 따라, 상위 계층에서 하위 계층으로 이어지며 입력 영상의 특징을 축소시키는 CNN(Convolutional Neural Network) 및 DSC(Depthwise Seperable Convolution)로 구성되는 처리 모듈 중 적어도 하나를 포함하여 구성되는 인코더부 및 디코더부로 구성된 상기 의미 분할 네트워크의 각 계층에서의 활성화 맵의 크기에 따라 상기 CNN 처리 모듈 또는 DSC 처리 모듈을 선택하여 상기 촬영된 영상의 컨텍스트를 추출하는 것인,

임베디드 의미 분할 네트워크 장치.

#### 청구항 2

제1항에 있어서,

상기 프로세서는 상기 디코더부의 소정의 계층에 적용된 확장된 ASPP(Extended Atrous Spatial Pyramid Pooling)를 통해 상기 인코더부에서 출력된 영상의 특징 정보를 입력받아 인코딩된 영상에 상응하는 특징 정보를 추출하는 것인,

임베디드 의미 분할 네트워크 장치.

#### 청구항 3

제2항에 있어서,

상기 확장된 ASPP는 전역 평균 풀링 경로(Global average pooling paths) 없이 복원된 형태를 사용하여 고품질의 컨텍스트를 추출하기 위하여 복수의 ASPP를 포함하여 구성되고,

상기 복수의 ASPP는 상기 소정의 계층 중 상위 계층에 적용된 제1 ASPP와 상기 상위 계층에 연속되는 하위 계층에 적용된 제2 ASPP를 포함하여 구성되는 것인,

임베디드 의미 분할 네트워크 장치.

#### 청구항 4

제3항에 있어서,

상기 제2 ASPP는 상기 의미 분할 네트워크의 계층 중 최하위 계층에 적용되는 것을 특징으로 하는,

임베디드 의미 분할 네트워크 장치.

## 청구항 5

제3항에 있어서,

상기 제2 ASPP는 상기 인코더부에서 출력된 특징 정보를 입력받는 복수의 CNN 및 상기 복수의 CNN과 병렬로 구성된 확장 경로로 구성된 입력단과, 상기 입력단에 의한 각 출력값을 결합하여 CNN에 입력시키는 출력단을 포함하여 구성되는 것인,

임베디드 의미 분할 네트워크 장치.

## 청구항 6

제5항에 있어서,

상기 확장 경로는 CNN과, 상기 CNN의 출력을 각각 입력받는 DSC 및 상기 DSC의 각 출력을 결합하여 이중 선형 보간하는 이중선형보간부를 포함하여 구성되는 것인,

임베디드 의미 분할 네트워크 장치.

## 청구항 7

제3항에 있어서,

상기 제1 ASPP는 상기 제2 ASPP에서 출력된 특징 정보를 입력받는 CNN 및 상기 CNN과 병렬 배치된 복수의 IDSC(Inverse DSC)로 구성된 입력단과, 상기 입력단에 의한 각 출력값을 결합하여 CNN에 입력시키는 출력단을 포함하여 구성되는 것인,

임베디드 의미 분할 네트워크 장치.

## 청구항 8

제1항에 있어서,

상기 인코더부는 상기 각 계층 중 소정의 계층에 하나 이상이 구비되어 상기 디코더부로 각 계층에 상응하는 영상의 상세 형태 정보를 전달하는 CNN으로 구성된 형태 정보 전달부를 포함하는,

임베디드 의미 분할 네트워크 장치.

## 청구항 9

제1항에 있어서,

상기 각 계층은 제1 계층 내지 제4 계층으로 구성되고,

상기 인코더부는 이전 계층에서 출력되는 특징 정보를 추상화하는 상기 제3 계층에 구비되는 두 개의 L3 모듈과 상기 제4 계층에 구비되는 L4 모듈을 포함하고,

상기 L3 모듈은 특징 정보를 각각 입력받는 딜레이션(dilation)이 각각 상이한 복수의 DSC와 상기 복수의 DSC의 출력값을 결합하여 입력받는 점합성곱(pointwise convolution) 및 상기 입력된 특징 정보 및 점합성곱의 출력값을 합산하는 합산기로 구성되고,

상기 L4 모듈은 특징 정보를 각각 입력받는 복수의 CNN 계층 및 복수의 CNN 계층에서의 최종 출력값과 상기 입력된 특징 정보를 합산하는 합산기로 구성되는,

임베디드 의미 분할 네트워크 장치.

## 청구항 10

제1항에 있어서,

상기 의미 분할 네트워크는 최대 채널 수가 64로 제한된 것을 특징으로 하는,

임베디드 의미 분할 네트워크 장치.

## 발명의 설명

### 기술 분야

[0001] 본 발명은 차량 영상의 화소를 분류하는 행렬 곱셈 가속기에 최적화된 임베디드 의미 분할 네트워크 장치에 관한 것이다.

### 배경 기술

[0002] 일반적인 자율주행 시스템은 센서를 사용하여 주행 환경을 인식하고, 차량의 방향, 속도 등의 상태를 결정하며 실시간으로 제어한다. 자율주행 차량은 오동작할 경우 사람의 생명을 위협할 수 있기 때문에 높은 신뢰성이 요구된다. 하지만, RADAR, 카메라, 초음파센서, LiDAR 등 다수의 차량용 센서가 개발되었음에도 불구하고, 시간 및 계절별로 바뀌는 다양한 주행 환경에서 항상 높은 인식 성능을 제공하기 어렵다. 따라서, 대부분의 대량 생산 차량은 운전 지원 단계에 머물러 있으며, 제한된 주행 환경에서만 OEM이 주행 결과에 책임을 갖는 LV3를 달성하였다.

[0003] 자율주행 시스템의 신뢰성 한계를 극복하기 위해서는 주행환경 인식 성능의 향상이 필요하다. 차량용 센서 중 카메라는 사람의 인식 방법과 가장 유사하여 차선, 표지판, 그리고 신호등과 같은 정보를 제공할 수 있기 때문에 자율 주행 시스템에 필수적이다.

[0004] 특히, 최근 급격히 발달한 딥러닝 알고리즘을 사용하면 많은 연산량을 사용하는 대신 고전 인식 방법의 한계보다 높은 인식 성능을 제공할 수 있다. 하지만 자율주행 시스템은 실시간 제어를 위한 차량용 임베디드 시스템이기 때문에 소비 전력, 반도체 신뢰성, 지연 시간, 처리율 그리고 가격 조건을 만족해야 하므로, 복잡도가 높은 딥러닝 네트워크를 사용할 수 없다. 더욱이 일반적인 임베디드 네트워크는 GPU나 ARM 환경에 최적화되어 있어서 단순한 HW 가속기가 있는 프로세서에서는 그 성능이 저하된다.

## 선행기술문헌

### 특허문헌

[0005] (특허문헌 0001) 공개특허공보 제10-2019-0000599호 (2020.07.13)

## 발명의 내용

### 해결하려는 과제

[0006] 본 발명이 해결하고자 하는 과제는 자율주행 시스템의 인식 성능을 향상시키기 위하여 기본적인 임베디드 하드웨어 가속기인 MMA(Matrixs Multiplication Accelerator)에 최적화된 의미 분할 네트워크를 제공하는, 차량 영상의 화소를 분류하는 행렬 곱셈 가속기에 최적화된 임베디드 의미 분할 네트워크 장치를 제공한다.

[0007] 다만, 본 발명이 해결하고자 하는 과제는 상기된 바와 같은 과제로 한정되지 않으며, 또 다른 과제들이 존재할 수 있다.

### 과제의 해결 수단

[0008] 상술한 과제를 해결하기 위한 본 발명의 일면에 따른 차량 영상의 화소를 분류하는 행렬 곱셈 가속기(Matrixs Multiplication Accelerator)에 최적화된 임베디드 의미 분할 네트워크 장치는 카메라를 통해 촬영된 영상을 수신하는 통신모듈, 상기 촬영된 영상의 컨텍스트를 추출하기 위한 의미 분할 네트워크(MMANet) 기반의 프로그램이 저장된 메모리 및 상기 메모리에 저장된 프로그램을 실행시키는 프로세서를 포함한다. 이때, 상기 프로세서

는 상기 프로그램을 실행시킴에 따라, 상위 계층에서 하위 계층으로 이어지며 입력 영상의 특징을 축소시키는 CNN(Convolutional Neural Network) 및 DSC(Depthwise Seperarable Convolution)로 구성되는 처리 모듈 중 적어도 하나를 포함하여 구성되는 인코더부 및 디코더부로 구성된 상기 의미 분할 네트워크의 각 계층에서의 활성화 맵의 크기에 따라 상기 CNN 처리 모듈 또는 DSC 처리 모듈을 선택하여 상기 촬영된 영상의 컨텍스트를 추출한다.

- [0009] 본 발명의 일부 실시예에서, 상기 프로세서는 상기 디코더부의 소정의 계층에 적용된 확장된 ASPP(Extended Atrous Spatial Pyramid Pooling)를 통해 상기 인코더부에서 출력된 영상의 특징 정보를 입력받아 인코딩된 영상에 상응하는 특징 정보를 추출할 수 있다.
- [0010] 본 발명의 일부 실시예에서, 상기 확장된 ASPP는 전역 평균 풀링 경로(Global average pooling paths) 없이 복원된 형태를 사용하여 고품질의 컨텍스트를 추출하기 위하여 복수의 ASPP를 포함하여 구성되고, 상기 복수의 ASPP는 상기 소정의 계층 중 상위 계층에 적용된 제1 ASPP와 상기 상위 계층에 연속되는 하위 계층에 적용된 제2 ASPP를 포함하여 구성될 수 있다.
- [0011] 본 발명의 일부 실시예에서, 상기 제2 ASPP는 상기 의미 분할 네트워크의 계층 중 최하위 계층에 적용될 수 있다.
- [0012] 본 발명의 일부 실시예에서, 상기 제2 ASPP는 상기 인코더부에서 출력된 특징 정보를 입력받는 복수의 CNN 및 상기 복수의 CNN과 병렬로 구성된 확장 경로로 구성된 입력단과, 상기 입력단에 의한 각 출력값을 결합하여 CNN에 입력시키는 출력단을 포함하여 구성될 수 있다.
- [0013] 본 발명의 일부 실시예에서, 상기 확장 경로는 CNN과, 상기 CNN의 출력을 각각 입력받는 DSC 및 상기 DSC의 각 출력을 결합하여 이중 선형 보간하는 이중선형보간부를 포함하여 구성될 수 있다.
- [0014] 본 발명의 일부 실시예에서, 상기 제1 ASPP는 상기 제2 ASPP에서 출력된 특징 정보를 입력받는 CNN 및 상기 CNN과 병렬 배치된 복수의 IDSC(Inverse DSC)로 구성된 입력단과, 상기 입력단에 의한 각 출력값을 결합하여 CNN에 입력시키는 출력단을 포함하여 구성될 수 있다.
- [0015] 본 발명의 일부 실시예에서, 상기 인코더부는 상기 각 계층 중 소정의 계층에 하나 이상이 구비되어 상기 디코더부로 각 계층에 상응하는 영상의 상세 형태 정보를 전달하는 CNN으로 구성된 형태 정보 전달부를 포함할 수 있다.
- [0016] 본 발명의 일부 실시예에서, 상기 각 계층은 제1 계층 내지 제4 계층으로 구성되고, 상기 인코더부는 이전 계층에서 출력되는 특징 정보를 추상화하는 상기 제3 계층에 구비되는 두 개의 L3 모듈과 상기 제4 계층에 구비되는 L4 모듈을 포함하고, 상기 L3 모듈은 특징 정보를 각각 입력받는 딜레이션(dilation)이 각각 상이한 복수의 DSC와 상기 복수의 DSC의 출력값을 결합하여 입력받는 점합성곱(pointwise convolution) 및 상기 입력된 특징 정보 및 점합성곱의 출력값을 합산하는 합산기로 구성되고, 상기 L4 모듈은 특징 정보를 각각 입력받는 복수의 CNN 계층 및 복수의 CNN 계층에서의 최종 출력값과 상기 입력된 특징 정보를 합산하는 합산기로 구성될 수 있다.
- [0017] 본 발명의 일부 실시예에서, 상기 의미 분할 네트워크는 최대 채널 수가 64로 제한될 수 있다.
- [0018] 상술한 과제를 해결하기 위한 본 발명의 다른 면에 따른 컴퓨터 프로그램은, 하드웨어인 컴퓨터와 결합되어 의미 분할 네트워크 기반의 프로그램을 실행하며, 컴퓨터 판독가능 기록매체에 저장된다.
- [0019] 본 발명의 기타 구체적인 사항들은 상세한 설명 및 도면들에 포함되어 있다.

### 발명의 효과

- [0020] 전술한 본 발명의 일 실시예에 의하면, ADAS\_VP 및 ADAS\_PRK II 제어기에 사용되는 Vision Processing Unit(VPU)인 TI사의 TDA4V의 하드웨어 가속기인 행렬 곱셈 연산기에 최적화된 딥러닝 알고리즘을 제공할 수 있다.
- [0021] 특히, 본 발명에 따른 의미 분할 네트워크는 영상의 각 화소를 분류할 수 있으며, 차량 주변 장애물의 위치 및 종류 구분을 통해 자율주행 시스템의 성능을 향상시킬 수 있다. 또한, 본 발명의 일 실시예에 의하면 실시간으로 VGA급 6개 카메라를 20FPS로 처리가 가능한바, 기존에 제안된 방법 보다 높은 인식 성능 제공이 가능하므로 자율주행 시스템의 신뢰성을 크게 향상시킬 수 있다.
- [0022] 이와 더불어, 고가의 GPU를 사용하지 않고 간단한 하드웨어 가속기를 사용하는 것만으로 높은 인식 성능을 제공

하는 딥러닝 알고리즘을 제공할 수 있어, GPU의 사용으로 인한 비용을 절감할 수 있는 효과가 있다.

[0023] 본 발명의 효과들은 이상에서 언급된 효과로 제한되지 않으며, 언급되지 않은 또 다른 효과들은 아래의 기재로부터 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

### 도면의 간단한 설명

[0024] 도 1은 의미 분할 네트워크의 일 예시를 도시한 도면이다.

도 2는 TI사의 TDA4V MMA 구조를 도시한 도면이다.

도 3은 MMA를 최적화하기 위한 채널 수에 따른 레이어 및 곱셈 수를 설명하기 위한 도면이다.

도 4는 본 발명의 일 실시예에 따른 임베디드 의미 분할 네트워크 장치의 블록도이다.

도 5는 본 발명의 일 실시예에 따른 임베디드 의미 분할 네트워크 장치를 설명하기 위한 도면이다.

도 6은 L3 모듈을 설명하기 위한 도면이다.

도 7은 L4 모듈을 설명하기 위한 도면이다.

도 8은 종래 기술에 따른 피라미드 풀링 및 ASPP를 설명하기 위한 도면이다

도 9는 본 발명의 일 실시예에 따른 확장된 ASPP(Extended Atrous Spatial Pyramid Pooling)를 설명하기 위한 도면이다.

도 10은 cityscape validation set을 사용한 정성적인 인식 성능 테스트 결과를 도시한 도면이다.

### 발명을 실시하기 위한 구체적인 내용

[0025] 본 발명의 이점 및 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 상세하게 후술되어 있는 실시예들을 참조하면 명확해질 것이다. 그러나, 본 발명은 이하에서 개시되는 실시예들에 제한되는 것이 아니라 서로 다른 다양한 형태로 구현될 수 있으며, 단지 본 실시예들은 본 발명의 개시가 완전하도록 하고, 본 발명이 속하는 기술 분야의 통상의 기술자에게 본 발명의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 발명은 청구항의 범주에 의해 정의될 뿐이다.

[0026] 본 명세서에서 사용된 용어는 실시예들을 설명하기 위한 것이며 본 발명을 제한하고자 하는 것은 아니다. 본 명세서에서, 단수형은 문구에서 특별히 언급하지 않는 한 복수형도 포함한다. 명세서에서 사용되는 "포함한다(comprises)" 및/또는 "포함하는(comprising)"은 언급된 구성요소 외에 하나 이상의 다른 구성요소의 존재 또는 추가를 배제하지 않는다. 명세서 전체에 걸쳐 동일한 도면 부호는 동일한 구성 요소를 지칭하며, "및/또는"은 언급된 구성요소들의 각각 및 하나 이상의 모든 조합을 포함한다. 비록 "제1", "제2" 등이 다양한 구성요소들을 서술하기 위해서 사용되나, 이들 구성요소들은 이들 용어에 의해 제한되지 않음은 물론이다. 이들 용어들은 단지 하나의 구성요소를 다른 구성요소와 구별하기 위하여 사용하는 것이다. 따라서, 이하에서 언급되는 제1 구성요소는 본 발명의 기술적 사상 내에서 제2 구성요소일 수도 있음은 물론이다.

[0027] 다른 정의가 없다면, 본 명세서에서 사용되는 모든 용어(기술 및 과학적 용어를 포함)는 본 발명이 속하는 기술 분야의 통상의 기술자에게 공통적으로 이해될 수 있는 의미로 사용될 수 있을 것이다. 또한, 일반적으로 사용되는 사전에 정의되어 있는 용어들은 명백하게 특별히 정의되어 있지 않는 한 이상적으로 또는 과도하게 해석되지 않는다.

[0028] 본 발명은 차량 영상의 화소를 분류하는 행렬 곱셈 가속기에 최적화된 임베디드 의미 분할 네트워크 장치(100, 이하 임베디드 의미 분할 네트워크 장치)에 관한 것이다.

[0029] 도 1은 의미 분할 네트워크의 일 예시를 도시한 도면이다. 도 2는 TI사의 TDA4V MMA 구조를 도시한 도면이다.

[0030] 의미 분할 네트워크는 영상의 각 화소를 분류하기 위한 것이다. 자율 주행 시스템의 인식 성능을 향상시키기 위해 딥러닝 알고리즘을 사용하면 높은 분류 정확도를 얻을 수 있다.

[0031] 하지만, 자율 주행 시스템은 실시간 제어를 위한 차량용 임베디드 시스템이기 때문에 소비 전력, 반도체 신뢰성, 지연 시간, 처리율 및 가격 조건을 만족해야 하므로 복잡도가 높은 네트워크를 사용할 수 없다. 더욱이, 기존 딥러닝 네트워크는 GPU나 ARM 프로세서 환경에 최적화되어 있으므로 단순한 하드웨어 가속기가 있



는 프로세서에서는 그 성능이 저하되는 문제가 있다.

[0032] 이러한 문제를 해소하기 위해 본 발명의 일 실시예는 자율주행 시스템의 인식 성능을 향상시키기 위하여 기본적인 임베디드 하드웨어 가속기인 MMA(Matrixs Multiplication Accelerator)에 최적화된 의미 분할 네트워크를 제공하는, 차량 영상의 화소를 분류하는 행렬 곱셈 가속기에 최적화된 임베디드 의미 분할 네트워크 장치(100)를 제공한다.

[0033] 특히, 본 발명의 일 실시예는 도 2에 도시된 바와 같은 TI DPS(Texas Instruments Digital Signal Processor)의 행렬 곱셈 가속기(Matrixs Multiplication Accelerator, MMA)에 최적화된 의미 분할 네트워크를 제공하는 것을 특징으로 한다. 즉, 종래 기술에 따른 딥러닝 방법은 일반적인 연산이 가능한 GPU에 최적화되어 있기 때문에 간단한 하드웨어 가속기인 MMA의 성능을 최대화할 수 없다. 하지만, MMA에 최적화된 방법을 찾을 수 있다면 4,096번의 MAC(Multiplication And Addition)을 1주기(cycle)동안 수행할 수 있으므로 그 성능을 향상시킬 수 있다. MMA는 아래 식 1을 1주기에 수행이 가능하다.

[0034] [식 1]

$$[c_1 \ c_2 \ \cdots \ c_{64}] [a_1 \ a_2 \ \cdots \ a_{64}] \times \begin{bmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,64} \\ b_{2,1} & b_{2,2} & \cdots & b_{2,64} \\ \vdots & \vdots & \ddots & \vdots \\ b_{64,1} & b_{64,2} & \cdots & b_{64,64} \end{bmatrix}$$

[0035]

[0036] 도 3은 MMA를 최적화하기 위한 채널 수에 따른 레이어 및 곱셈 수를 설명하기 위한 도면이다.

[0037] 도 3과 같이 채널 수에 따라 레이어의 수와, 한 레이어에서 사용되는 곱셈 수는 달라지게 된다. 한 레이어에서 제공하는 곱셈 수는 그 레이어가 제공할 수 있는 추상화 수준(연산 수준)과 직결된다. 따라서, 동일 시간에 많은 곱셈 수를 제공 가능한 레이어를 사용하면 처리 속도가 동일한 상태에서 인식 성능도 향상되게 된다.

[0038] 본 발명의 일 실시예는 레이어의 수에 대한 요구 사항을 만족시키면서 최대 곱셈 수를 제공하기 위하여 채널 수를 제한하는 것을 특징으로 하며, 일 예로 제한되는 채널 수는 64개일 수 있다.

[0039] 본 발명의 일 실시예에서 제안하는 의미 분할 네트워크(200)는 ADAS\_VP 및 ADAS\_PRK II 제어기에 사용되는 VPU(Vision Processing Unit)인 TI사의 TDA4V MID 프로세서의 MMA 연산에 네트워크 구조를 최적화하기 위하여 다음 세가지 방법을 사용하였다.

[0040] 첫째로, DDR(Dual Data Rate) 메모리 접근을 최소화하고, L3 캐시만 사용하여 내부 연산을 하도록 경로를 구성하였다. 둘째로, 활성화 맵의 크기에 따라 DSC(Depthwise Seperable Convolution)를 선별적으로 적용하여 네트워크 표현력을 증가시켰다. 마지막으로, 확장된 ASPP(Astrous Spatial Pyramid Polling)를 사용하여 정확한 컨텍스트가 추출되도록 하였다. 위치에 따라 변화량이 큰 ASPP를 안정적으로 동작시키기 위하여 5×5 및 7×7 컨볼루션 경로를 추가하였으며, 2단 ASPP를 사용함으로써 컨텍스트 정보를 사용하여 형태 정보를 디코딩하도록 하였다.

[0041] 한편, 본 발명의 일 실시예는 자율 주행 시스템에 적용되는 것을 특징으로 하나, 반드시 이에 한정되는 것은 아니다. 자율 주행 시스템은 차량 주변 360도 환경을 실시간으로 인식해야 하므로, 의미 분할 네트워크는 다음 표 1에 따른 요구사항을 만족해야 한다.

표 1

[0042]

Group	Specification	Value
Requirement	Image size	640, 480
	Camera	6EA
	FPS	20
	Weight	7,936kB
Target	Processing time (85%)	7.08ms
	Weight size (60%)	4761kB
	#Layers	> 150

[0043] 이하에서는 도 4 내지 도 10을 참조하여 본 발명의 일 실시예에 따른 임베디드 의미 분할 네트워크 장치(100)에 대하여 설명하도록 한다. 도 4는 본 발명의 일 실시예에 따른 임베디드 의미 분할 네트워크 장치(100)의 블록도



이다. 도 5는 본 발명의 일 실시예에 따른 임베디드 의미 분할 네트워크 장치(100)를 설명하기 위한 도면이다.

- [0044] 본 발명의 일 실시예에 따른 임베디드 의미 분할 네트워크 장치(100)는 통신모듈(110), 메모리(120) 및 프로세서(130)를 포함하여 구성된다.
- [0045] 통신모듈(110)은 카메라를 통해 촬영된 영상을 수신한다.
- [0046] 메모리(120)에는 촬영된 영상의 컨텍스트를 추출하기 위한 의미 분할 네트워크(MMANet) 기반의 프로그램이 저장되며, 프로세서(130)는 메모리에 저장된 프로그램을 실행시킨다.
- [0047] 프로세서(130)는 의미 분할 네트워크(200)의 각 계층에서의 활성화 맵의 크기에 따라 CNN 처리 모듈 또는 DSC 처리 모듈을 선택하여 촬영된 영상의 컨텍스트를 추출한다.
- [0048] 도 5를 참조하면, 의미 분할 네트워크(200)는 상위 계층에서 하위 계층으로 이어지며, 입력 영상의 특징을 축소시키는 CNN과 DSC로 구성되는 처리 모듈 중 적어도 하나를 포함하는 인코더부(210)와 디코더부(220)로 구성된다. 본 발명의 일 실시예에서의 의미 분할 네트워크(200)의 각 계층은 상위 계층인 제1 계층(Level 1) 내지 하위 계층인 제4 계층(Level 4)으로 구성되며, 제3 계층 및 제4 계층의 최대 채널 수는 64로 제한된다. 제0 계층(Level 0)은 원본 영상이 존재하는 계층을 의미한다.
- [0049] 이때, 인코더부(210) 및 디코더부(220)의 각 계층의 레벨이 커질수록 특징 정보의 가로와 세로의 크기가 절반으로 축소된다. 가령, 레벨이 모두 0인 네트워크는 특징 정보의 크기가 줄어들지 않으므로 보다 상세한 정보를 가질 수 있는 대신 연산량이 커지게 된다. 또한, 제5 계층 레벨까지 존재할 경우 특징 정보의 크기가 너무 줄어들어 정보 손실량이 많아지게 되어 디코더부(220)에서의 보상이 어려울 수 있다.
- [0050] 일 실시예로, 인코더부(210)는 각 계층 중 소정의 계층에 하나 이상이 구비되어 디코더부(220)로 각 계층에 상응하는 영상의 상세 형태 정보(shape)를 전달하는 CNN으로 구성된 형태 정보 전달부를 포함할 수 있다.
- [0051] 도 5를 참조하면, 인코더부(210)는 제2 계층(Level 2) 및 제3 계층(Level 3)에 각각 L2 경로와 L3 경로를 포함하며, 각 경로를 이용하여 CNN으로 구성된 각 형태 정보 전달부(Conv(29,8,k=1), Conv(61,16,k=1))를 통해 제2 계층 및 제3 계층에 상응하는 영상의 상세 형태 정보를 디코더부(220)로 전달할 수 있다. 본 발명의 일 실시예는 이러한 L2 경로와 L3 경로를 통해 영상의 상세 형태 정보를 디코더부(220)에 전달함으로써, 디코더부(220)에서는 추상화(분류 정보) 및 형태 정보를 이용하여 의미 분할을 수행할 수 있도록 할 수 있다.
- [0052] 일 실시예로, 인코더부(210)는 이전 계층에서 출력되는 특징 정보를 추상화하는 제3 계층에 구비되는 두 개의 L3 모듈과 제4 계층에 구비되는 L4 모듈을 포함할 수 있다.
- [0053] 도 6은 L3 모듈을 설명하기 위한 도면이다. 도 7은 L4 모듈을 설명하기 위한 도면이다.
- [0054] 기존의 딥러닝 네트워크는 단일 모듈을 반복 사용하지만, 본 발명의 일 실시예는 인식 성능을 향상시키기 위해 계층 레벨에 따라 최적화된 모듈 구조를 사용하는 것을 특징으로 한다.
- [0055] L3 모듈과 L4 모듈은 모두 의미 분할 네트워크(200)의 중간 계층에 구비되는 것이므로, 입력값과 출력값은 모두 특징 정보가 된다. 다만, 본 발명의 일 실시예에서 구비되는 L3 모듈과 L4 모듈은 타 CNN 및 DSC 처리 모듈보다 그 출력값의 추상화 수준이 더 높은 것을 특징으로 한다. 이러한 L3 모듈과 L4 모듈은 모두 일반적인 컨볼루션보다 FoV(Field of View)가 넓으며(dilation=2와 4 적용), 넓은 특징 정보를 볼 수 있도록 구성한 상태에서 레벨에 따라 연산량이 최소가 되는 구조로 구성되어 있다. 이때, 제3 계층은 stride=2가 적용되어 제4 계층에 상응하는 특징 정보의 크기를 갖도록 두 개의 L3 모듈을 구비한다.
- [0056] 제3 계층에 구비되는 L3 모듈은 특징 정보를 각각 입력받는 딜레이션(dilation)이 각각 상이한 복수의 DSC와, 복수의 DSC의 출력값이 결합된 값을 입력받는 점합성곱(Pointwise convolution) 및 입력된 특징 정보 및 점합성곱의 출력값을 합산하는 합산기로 구성된다. 이러한, L3 모듈은 딜레이션이 다른 컨볼루션 결과에 가중치를 취하여 출력값을 생성하므로, 학습과정에서 최적의 가중치가 설정되게 된다.
- [0057] 제4 계층에 구비되는 L4 모듈은 특징 정보를 각각 입력받는 복수의 CNN 계층으로 구성되며, 복수의 CNN 계층에서의 최종 출력값과 입력된 특징 정보를 합산하는 합산기로 구성된다. 이러한 L4 모듈은 각 병렬 경로에서 특징 정보를 처리하는 FoV를 넓히기 위하여 넓게(dilation=4) 추출한 것과 좁게 추출한 것을 가중치 없이 합산하는 것을 특징으로 한다. 따라서, 학습 과정에서 위치에 따른 특징 정보의 크기는 근거리(일반 컨볼루션 d=1)와 원거리(d=4)에 분포한 특징 정보에 영향을 받아서 결정된다.

- [0058] 또한, 제3 계층과 제4 계층은 L3 long path와 L4 long path가 각각 구비한다. Long path는 다수의 모듈을 단일 모듈처럼 개념화할 수 있도록 하며, 계층이 깊어짐에 따라 강화되는 추상화 정보에 상세 형태 정보를 추가한다. 이러한 각 경로는, L3 모듈과 L4 모듈이 잔차(residual) 연산을 수행하도록 하며, 또한 학습 과정에서 변화(gradient)의 흐름을 원활하게 한다.
- [0059] 도 8은 종래 기술에 따른 피라미드 풀링 및 ASPP를 설명하기 위한 도면이다. 도 9는 본 발명의 일 실시예에 따른 확장된 ASPP(Extended Atrous Spatial Pyramid Pooling)를 설명하기 위한 도면이다.
- [0060] 피라미드 풀링(도 8의 (a))은 고주파에 둔감하다는 장점을 가지고 있으나, 사용된 전체 평균 연산은 행렬 연산이 아니기 때문에 수행 속도 측면에서 단점을 갖는다. 또한, 종래 기술에 따른 일반적인 ASPP(도 8의 (b))는 컨볼루션 연산을 사용하여 피라미드 풀링과 같이 넓은 FoV를 볼 수 있도록 하지만 고주파 성분에 민감하다는 단점을 갖는다.
- [0061] 이러한 문제점을 해소하기 위하여, 본 발명에서의 확장된 ASPP는 피라미드 풀링과 같이 넓은 FoV를 볼 수 있으며, 확장 경로를 사용하여 고주파 성분에 민감한 특성을 갖는 것을 특징으로 한다.
- [0062] 구체적으로, 프로세서(130)는 디코더부(220)의 소정의 계층에 적용된 확장된 ASPP를 통해 인코더부(210)에서 출력된 영상의 특징 정보를 입력받아 인코딩된 영상에 상응하는 특징 정보를 추출한다. 즉, 확장된 ASPP의 입력값과 출력값은 특징 정보가 되며, 확장된 ASPP의 특성상 FoV가 넓으므로 각 화소 주변의 형태 정보도 사용하여 분류가 된 결과가 출력된다.
- [0063] 일 실시예로, 확장된 ASPP는 전역 평균 풀링 경로(Global average pooling paths) 없이 복원된 형태를 사용하여 고품질의 컨텍스트를 추출하기 위하여 복수의 ASPP를 포함하여 구성된다. 이때, 복수의 ASPP는 상기 소정의 계층 중 상위 계층에 적용된 제1 ASPP와 상기 상위 계층에 연속되는 하위 계층에 적용된 제2 ASPP를 포함하여 구성된다. 여기에서 제2 ASPP는 의미 분할 네트워크(200)의 계층 중 최하위 계층인 제4 계층에 적용되고, 제1 ASPP는 제4 계층에 연속되는 제3 계층에 적용된다.
- [0064] 제2 ASPP는 인코더부(210)에서 출력된 특징 정보를 입력받는 복수의 CNN과, 복수의 CNN과 병렬로 구성된 확장 경로로 구성된 입력단과, 입력단에 의한 각 출력값을 결합하여 CNN에 입력시키는 출력단을 포함하여 구성된다. 그리고 확장 경로는 CNN과, CNN의 출력을 각각 입력받는 DSC 및 DSC의 각 출력을 결합하여 이중 선형 보간하는 이중선형보간부(Bilinear Interpolation x2)를 포함하여 구성된다. 여기에서 이중선형보간부는 입력 특징의 가로와 세로의 크기를 복수 배되도록 하여 의미 분할 네트워크의 최종 출력이 원본 영상의 크기를 갖도록 하는 역할을 한다.
- [0065] 제4 계층(Level 4)은 특징 정보의 크기가 가장 작기 때문에 상대적으로 가장 작은 연산량을 사용하게 된다. 따라서, 제4 계층에서는 고주파 성분에 민감한 ASPP를 보완하기 위하여 확장 경로를 추가하고 넓은 커널 사이즈(kernel size)를 갖는 컨볼루션을 사용한 제2 ASPP를 통해 그 출력이 위치 변화에 둔감하도록 하였다.
- [0066] 또한, 제1 ASPP는 제2 ASPP에서 출력된 특징 정보를 입력받는 CNN 및 CNN과 병렬 배치된 복수의 IDSC(Inverse DSC)로 구성된 입력단과, 입력단에 의한 각 출력값을 결합하여 CNN에 입력시키는 출력단을 포함하여 구성된다. 제3 계층은 제4 계층에 비하여 특징 정보의 크기가 더 커지게 되므로 FoV가 넓은 ASPP인 제1 ASPP를 다시 배치하여 한번 더 컨텍스트를 정제하도록 하였다.
- [0067] 이하에서는 도 10을 참조하여, 본 발명의 일 실시예를 테스트한 결과를 설명하도록 한다.
- [0068] 도 10은 cityscape validation set을 사용한 정성적인 인식 성능 테스트 결과를 도시한 도면이다.
- [0069] 본 발명의 일 실시예에서 제안한 의미 분할 네트워크 장치(100)의 성능을 확인하기 위하여 차량용 의미 분할 알고리즘 성능 평가 기준이 되는 Cityscapes의 validation set으로 성능을 비교하였다. 본 발명에서 제안한 의미 분할 네트워크(MMANet)는 다음 표 2와 같은 임베디드 딥러닝 알고리즘 중 최고 인식 성능(73.1% mIoU)을 제공한다.

표 2

[0070]

Name	#Parameter (M)	#Multiplication (G)	FLOPS (G)	mIoU (%)	Processing Time (us)
CGNet	0.5	3.8	7.6	63.5	7.6
ContextNet	0.8	6.6	13.1	65.9	2.3
DABNet	0.7	38.8	77.2	69.1	10.0

EDANet	0.7	33.2	66.8	65.1	6.2
ERFNet	2.0	103.9	208.7	71.5	9.8
FastSCNN	1.1	6.5	13.0	68.6	2.1
ESPNet2	0.7	13.2	26.2	66.4	16.1
MMANet (ours)	4.6	6.3	12.5	73.1	6.7

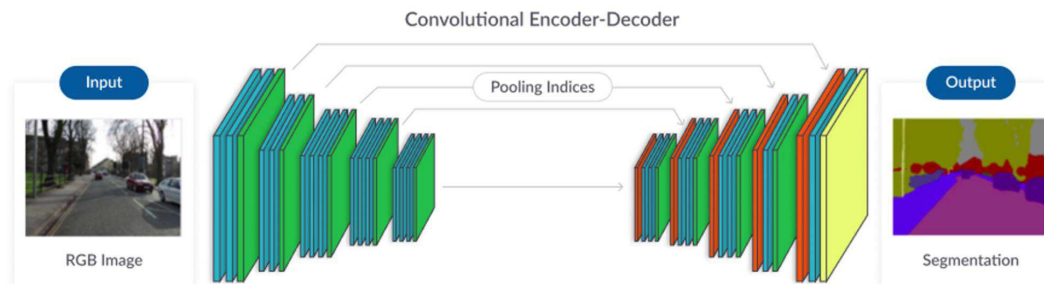
- [0071] 본 발명의 정성적인 인식 성능은 도 10에 도시된 바와 같다. 도 10의 (a)는 입력 영상을 나타내고, (b)는 정답지를 나타내며, (c)는 인식 결과를 나타낸 것이다.
- [0072] 이상에서 전술한 본 발명의 일 실시예는, 하드웨어인 컴퓨터와 결합되어 실행되기 위해 프로그램(또는 어플리케이션)으로 구현되어 매체에 저장될 수 있다.
- [0073] 상기 전술한 프로그램은, 상기 컴퓨터가 프로그램을 읽어 들여 프로그램으로 구현된 상기 방법들을 실행시키기 위하여, 상기 컴퓨터의 프로세서(CPU)가 상기 컴퓨터의 장치 인터페이스를 통해 읽힐 수 있는 C, C++, JAVA, Ruby, 기게어 등의 컴퓨터 언어로 코드화된 코드(Code)를 포함할 수 있다. 이러한 코드는 상기 방법들을 실행하는 필요한 기능들을 정의한 함수 등과 관련된 기능적인 코드(Functional Code)를 포함할 수 있고, 상기 기능들을 상기 컴퓨터의 프로세서가 소정의 절차대로 실행시키는데 필요한 실행 절차 관련 제어 코드를 포함할 수 있다. 또한, 이러한 코드는 상기 기능들을 상기 컴퓨터의 프로세서가 실행시키는데 필요한 추가 정보나 미디어가 상기 컴퓨터의 내부 또는 외부 메모리의 어느 위치(주소 번지)에서 참조되어야 하는지에 대한 메모리 참조관련 코드를 더 포함할 수 있다. 또한, 상기 컴퓨터의 프로세서가 상기 기능들을 실행시키기 위하여 원격(Remote)에 있는 어떠한 다른 컴퓨터나 서버 등과 통신이 필요한 경우, 코드는 상기 컴퓨터의 통신 모듈을 이용하여 원격에 있는 어떠한 다른 컴퓨터나 서버 등과 어떻게 통신해야 하는지, 통신 시 어떠한 정보나 미디어를 송수신해야 하는지 등에 대한 통신 관련 코드를 더 포함할 수 있다.
- [0074] 상기 저장되는 매체는, 레지스터, 캐쉬, 메모리 등과 같이 짧은 순간 동안 데이터를 저장하는 매체가 아니라 반영구적으로 데이터를 저장하며, 기기에 의해 판독(reading)이 가능한 매체를 의미한다. 구체적으로는, 상기 저장되는 매체의 예로는 ROM, RAM, CD-ROM, 자기 테이프, 플로피디스크, 광 데이터 저장장치 등이 있지만, 이에 제한되지 않는다. 즉, 상기 프로그램은 상기 컴퓨터가 접속할 수 있는 다양한 서버 상의 다양한 기록매체 또는 사용자의 상기 컴퓨터상의 다양한 기록매체에 저장될 수 있다. 또한, 상기 매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어, 분산방식으로 컴퓨터가 읽을 수 있는 코드가 저장될 수 있다.
- [0075] 전술한 본 발명의 설명은 예시를 위한 것이며, 본 발명이 속하는 기술분야의 통상의 지식을 가진 자는 본 발명의 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 쉽게 변형이 가능하다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다. 예를 들어, 단일형으로 설명되어 있는 각 구성 요소는 분산되어 실시될 수도 있으며, 마찬가지로 분산된 것으로 설명되어 있는 구성 요소들도 결합된 형태로 실시될 수 있다.
- [0076] 본 발명의 범위는 상기 상세한 설명보다는 후술하는 특허청구범위에 의하여 나타내어지며, 특허청구범위의 의미 및 범위 그리고 그 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

### 부호의 설명

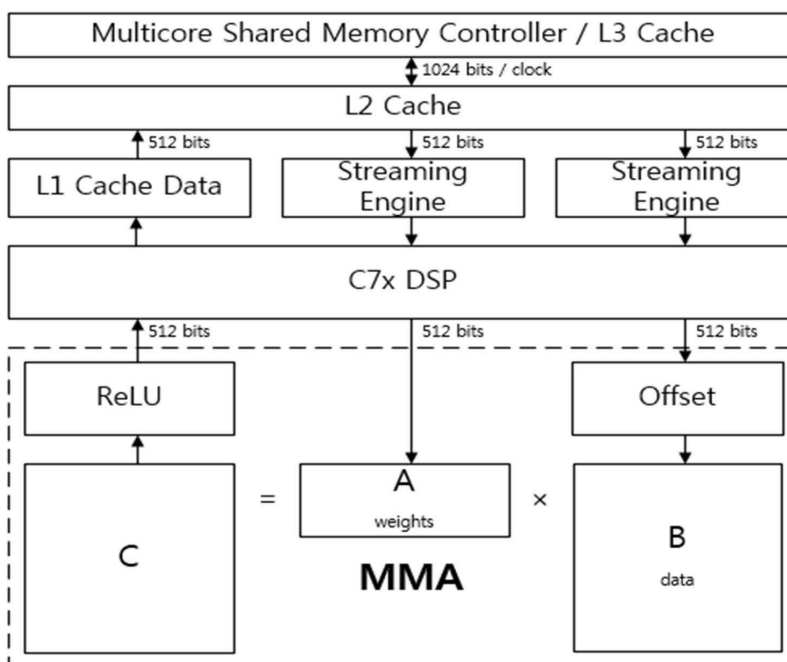
- [0077] 100: 임베디드 의미 분할 네트워크 장치
- 110: 통신모듈
- 120 : 메모리
- 130: 프로세서

도면

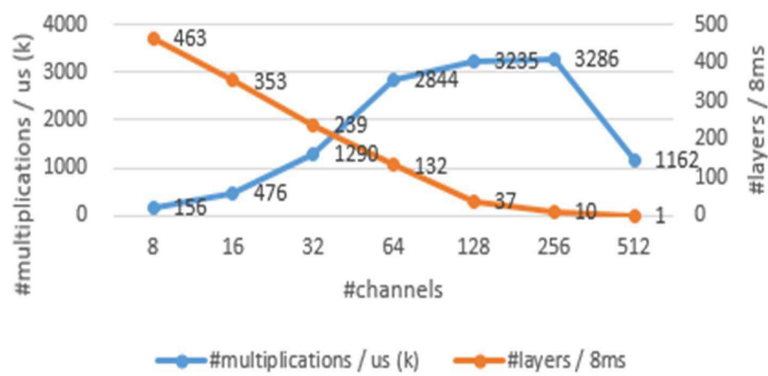
도면1



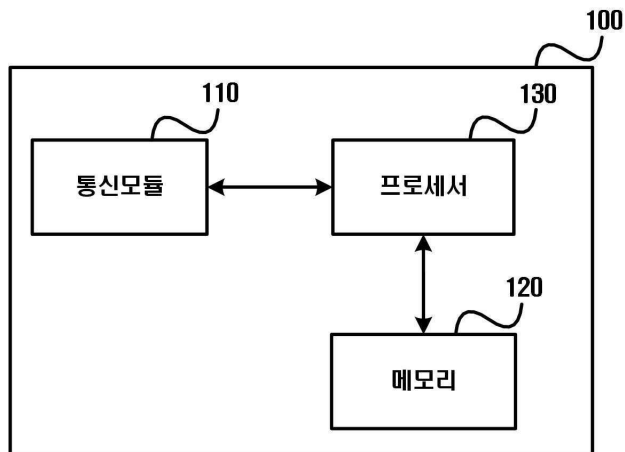
도면2



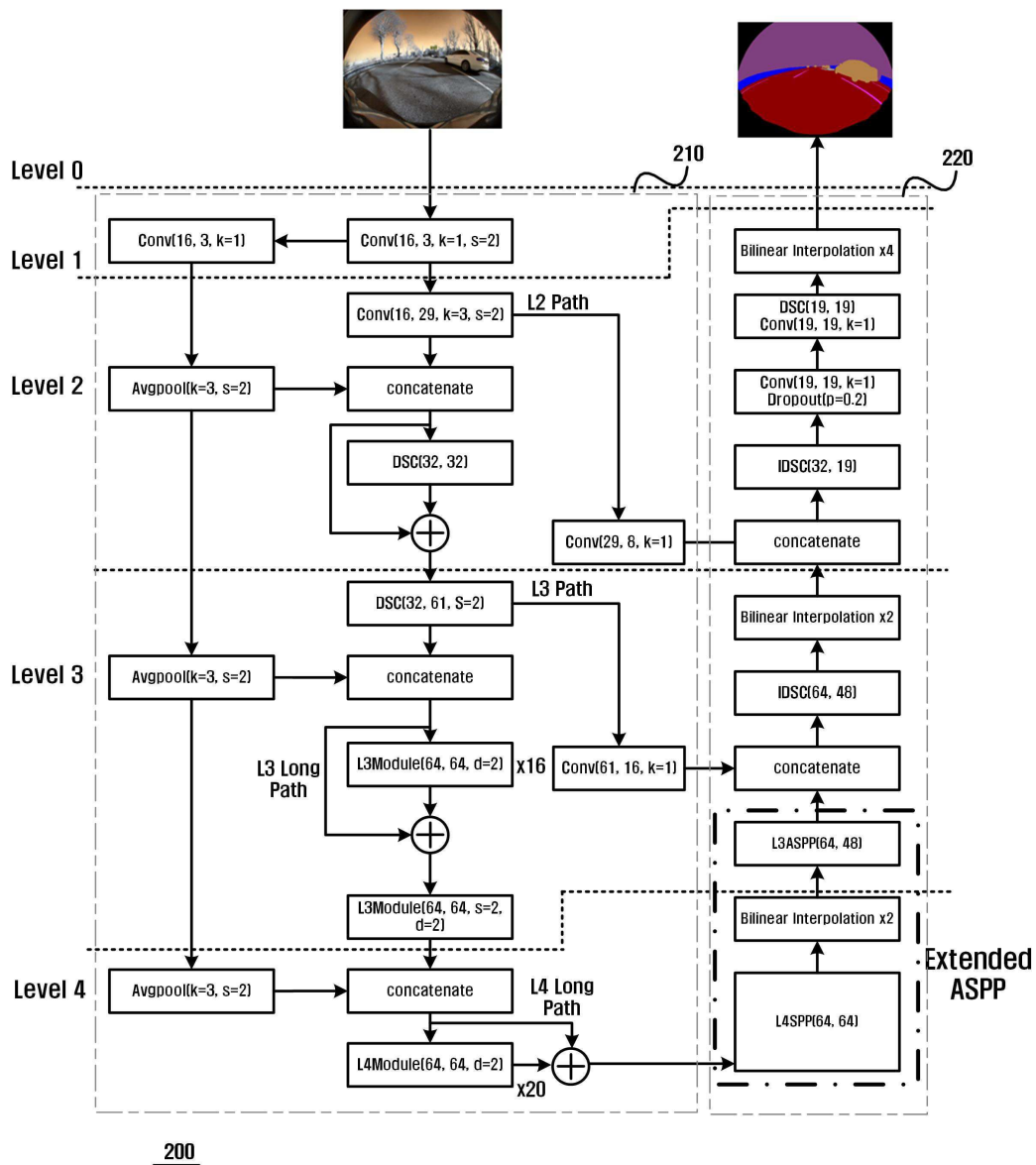
도면3



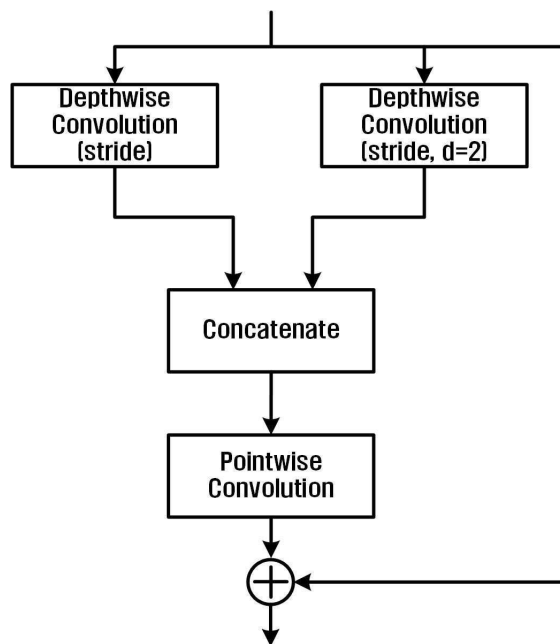
도면4



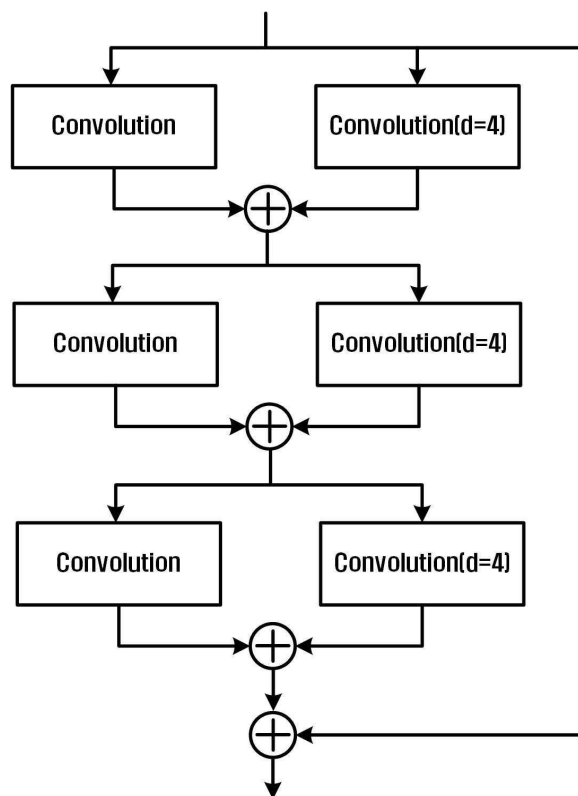
도면5



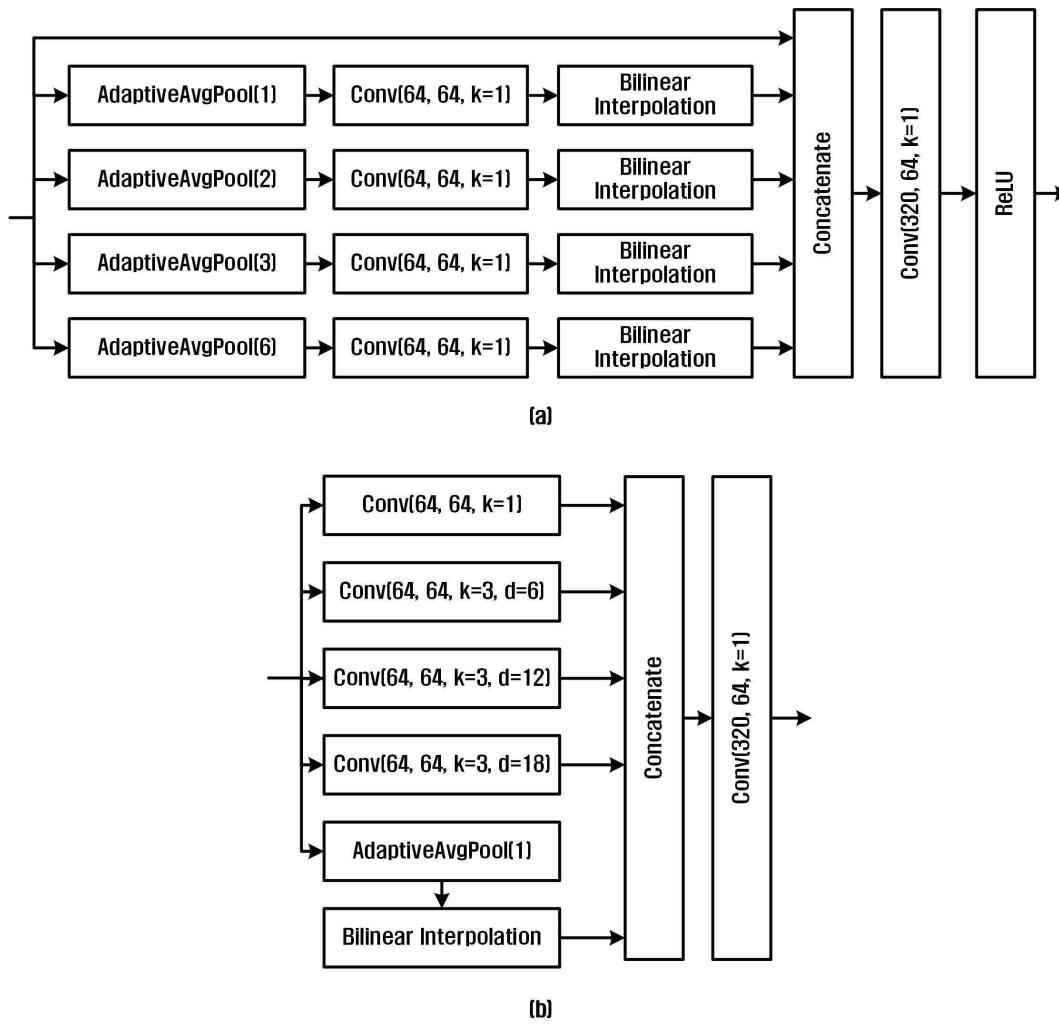
도면6



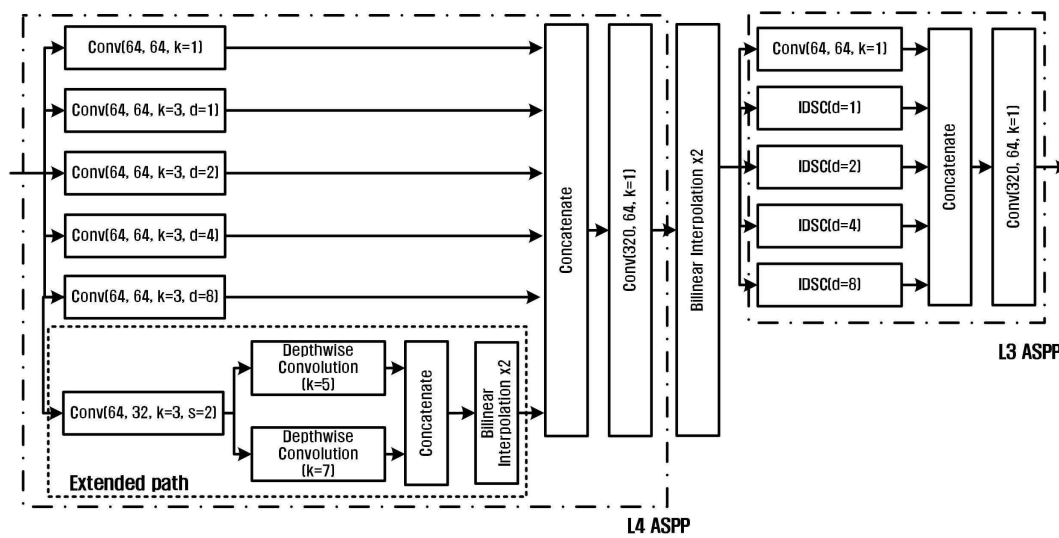
도면7



도면8



도면9





도면10

