

바이오시스템

Biosystem

상명대학교 융합공과대학 지능데이터융합학부

휴먼지능정보공학전공

일반대학원 지능정보공학과

일반대학원 감성공학과

일반대학원 스포츠ICT융합학과

디지털 신기술 바이오헬스케어 혁신공유대학 사업단

상명대학교 지능정보기술연구소 (ai.smu.ac.kr)

강의목차

- 바이오시스템 기초
 - 바이오헬스
- 생체정보데이터 종류와 특성
 - 생체신호(뇌파, 심전도, 근전도)
 - 영상정보
 - 동작정보
- 생체정보데이터 처리
 - 신호시스템 및 신호처리 기초
 - 영상처리 기초
 - 동작(모션)정보 처리 기초
- 생체정보데이터 분석
 - 시계열 분석
 - 주파수 분석
 - 기계학습(머신러닝)
 - 인공지능

강의목차

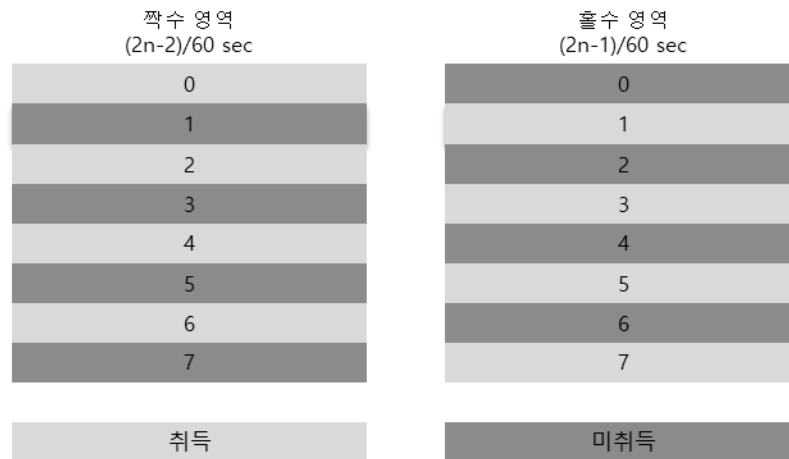
- **생체정보데이터 소프트웨어 및 하드웨어**
 - 생체정보데이터 소프트웨어 종류와 특성
 - 헬스케어 소프트웨어 (의료소프트웨어)
 - 생체정보데이터 하드웨어 종류와 특성
 - 헬스케어 하드웨어(의료기기)
- **연구프로젝트 관리 방법론**
 - 연구프로젝트 계획서 작성법
 - 연구프로젝트 개발 방법론 작성법
 - 연구프로젝트 사업비 작성법
- **연구논문 분석 및 작성**
 - 연구논문 초록 작성법
 - 연구논문 작성법
 - 연구윤리 기초
- **프로젝트 발표**
 - 생체신호 / 영상정보 / 동작정보 적용

학습목표

- 동영상기초
 - 동영상영상(움직임)

동영상기초

- 동영상
 - 물체 또는 카메라 이동에 의해 연속적으로 변하는 영상 집합
- 아날로그 동영상
 - 비월 주사
 - 3차원 시공간을 1차원 아날로그 신호로 스캔
 - 영상의 짝수 선과 홀수 선을 번갈아 가며 출력
 - NTSC TV, CRT 디스플레이에서 사용



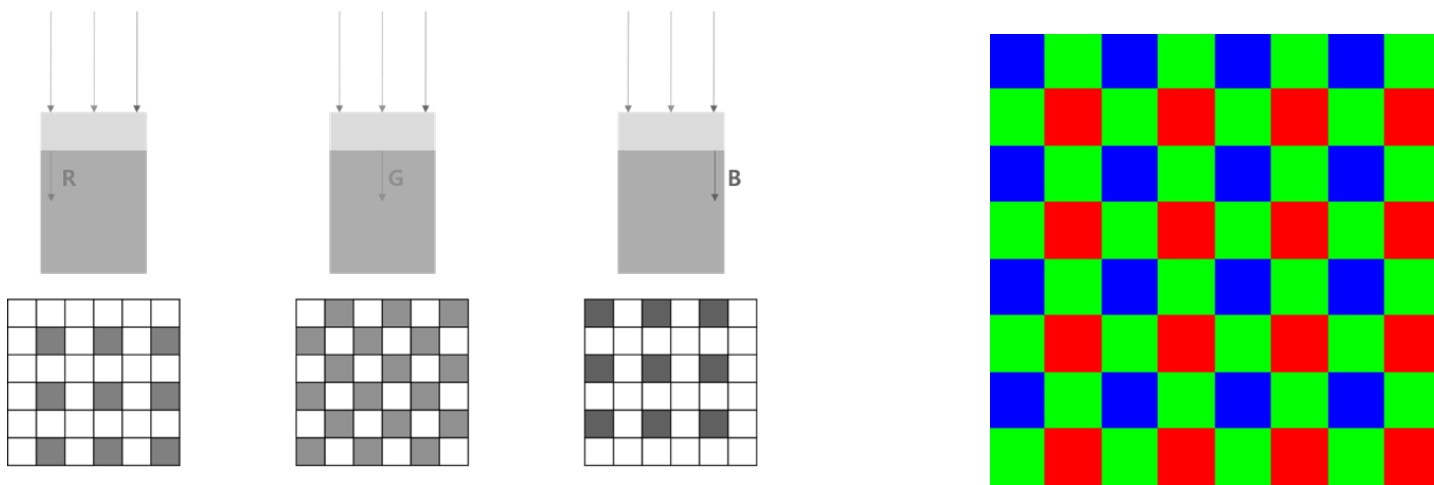
동영상기초

- 디지털 동영상

- 디지털 데이터로 표현한 동영상 (취득 및 표현)

- Bayer 필터

- 인지 특성에 기반한 RGB 색상 성분 취득
 - R 필터는 red만을, G 필터는 green만을, B 필터는 blue 만을 투과시키는 광학필터
 - 실제 이미지 센서는 각 화소에서 R, G, B 중 어느 한 색만을 감지할 수 있지만, 우리가 보는 카메라 영상에서는 각 화소마다 R,G,B 전체 색상을 보여줌
 - 칼라 이미지를 받으면 화소별로 3 바이트가 필요하지만 원래의 raw Bayer pattern 데이터를 그대로 받은 후 PC 단에서 이미지를 보간하면 데이터 전송량을 크게 줄임으로써 프레임수를 증가 가능



동영상기초

- 디지털 동영상

- 저장 포맷

- YUV 컬러 공간

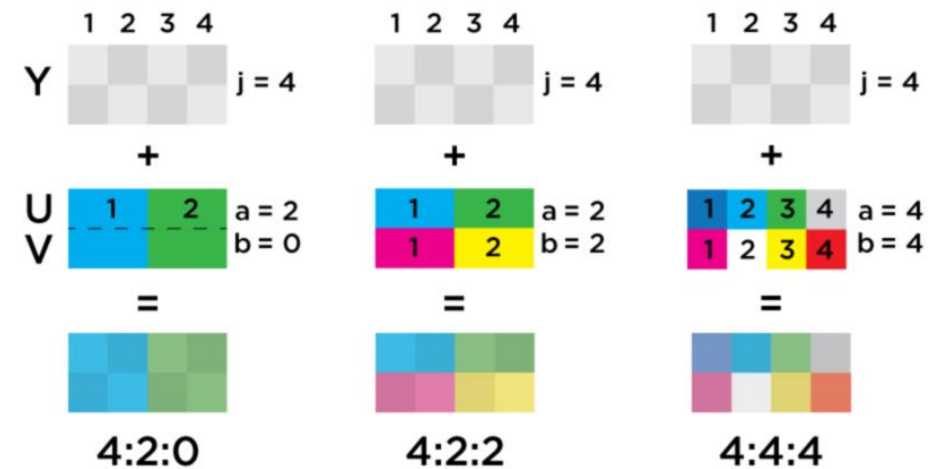
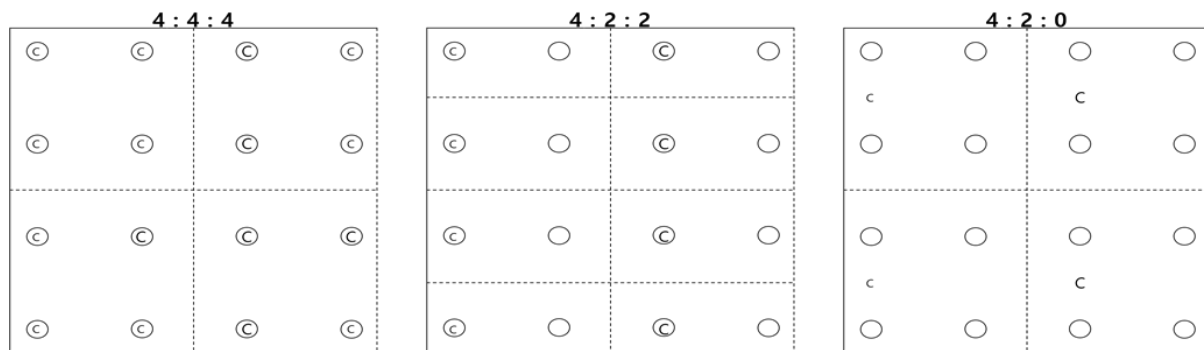
- 밝기와 컬러 신호를 분리하여 처리.

- Y : 밝기 값, UV : 컬러 정보

- $Y = 0.257R + 0.504G + 0.095B + 16$

- $U = -0.148R - 0.291G + 0.499B + 128$

- $V = 0.439R - 0.368G - 0.071B + 128$



동영상기초

- 움직임 추정
 - 한 장면과 다음 장면 간의 물체의 움직임 추정
 - 물체의 움직임 추정
 - 물체와 배경을 분리 후 물체의 움직임을 추정.
 - 화소 또는 블록 단위의 움직임 추정
 - 두 단위 화소 간 3차원 공간의 움직임을 추정.
 - 강체 움직임
 - 물체 내 거리가 변하지 않고 고정 된 거리를 가진 것을 기반으로 함.
 - 물체 모양에 변형이 없음.
 - 이동, 회전만을 고려
 - 비 강체 움직임
 - 시간에 따라 모양과 크기가 변하는 물체의 움직임

동영상기초

- 움직임 표현
 - 점 P 의 3차원 공간 상 움직임

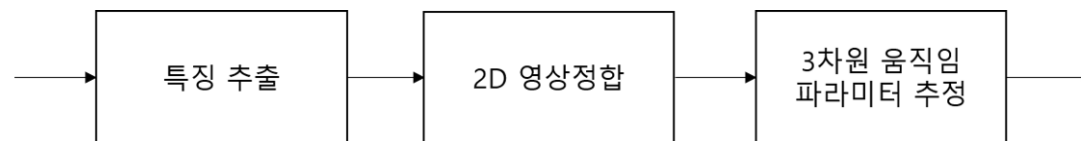
$$P' = RP + T$$

- $R : 3 \times 3$ 회전 변환
- $T : \text{이동 변환 벡터}$

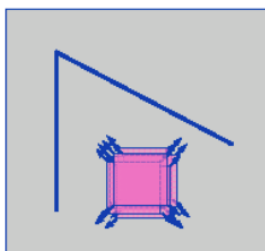
- $P = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$ and $P' = \begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix}, T = \begin{bmatrix} T_1 \\ T_2 \\ T_3 \end{bmatrix}$

동영상기초

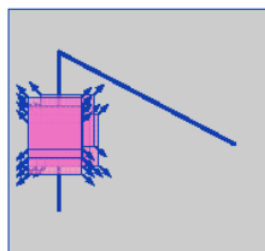
- 움직임 추정
 - 3차원 움직임 추정



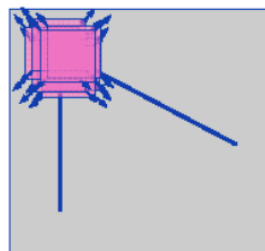
- 특징 추출 : 2D 영상 정합을 정확히 할 수 있는 점 추출
 - 해리스 코너 검출, 경계선 검출 방법
 - 코너는 영상의 지역적 특성중 하나로써, 에지의 방향이 급격하게 변하는 부분
 - 코너는 에지나 직선과 같은 다른 지역적 특성에 비해 분별력이 높음



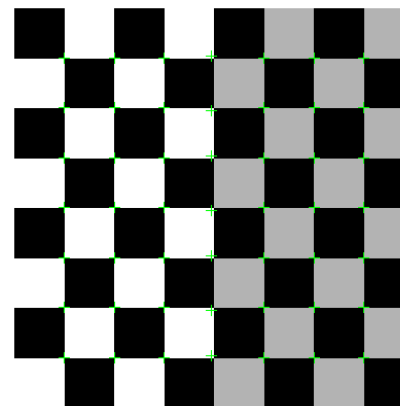
“flat” region:
no change in
all directions



“edge”:
no change along
the edge direction



“corner”:
significant change
in all directions



- 2D 영상정합 : 추출한 특징을 이용하여 영상 정합 수행

동영상기초

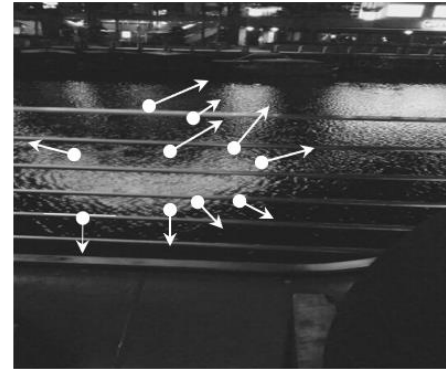
- 움직임을 위한 영상 정합 알고리즘
 - 전역 움직임
 - 영상 전체를 하나의 움직임 모델로 가정
 - 카메라의 이동만 고려
 - 화소 기반 움직임
 - 점, 선분 등의 특징 점에서 움직임을 추정
 - 블록 기반 움직임
 - 영상을 일정 단위 블록으로 구분하여 블록 단위로 움직임 파라미터를 추정
 - 물체 기반 움직임
 - 사전에 물체가 분할 되어 있을 때 적용 가능

동영상기초

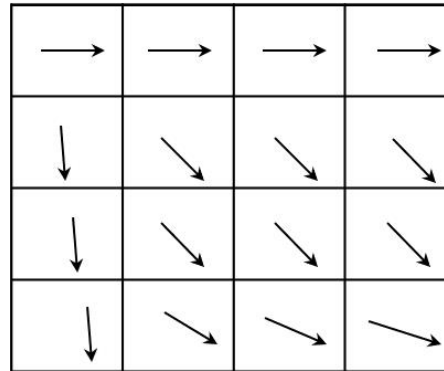
- 움직임 추정



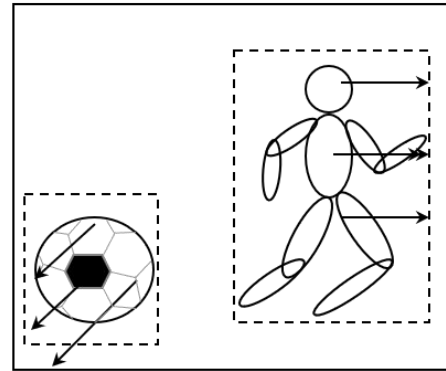
(a) 전역 움직임



(b) 화소 기반 움직임



(c) 블록 기반 움직임



(d) 물체 기반 움직임

동영상기초

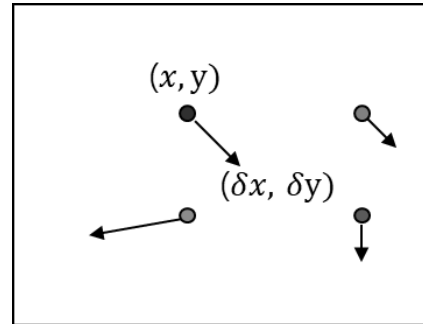
- 옵티컬 플로 기반의 움직임 추정
 - 옵티컬 플로(optical flow)
 - James J. Gibson, 1940
 - 움직임에 관련 된 시각적 자극의 표현
 - 가정
 - 움직임이 있기 전과 후의 밝기값에 변화가 없다.
- 영상처리에서의 옵티컬 플로
 - 관찰자와 화면 사이 상대적으로 발생하는 물체, 표면 혹은 모서리의 움직임 패턴을 표현 가능

동영상기초

- 옵티컬 플로 예측 모델

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t)$$

- 특정 시점 t 에서 점 (x, y) 가 짧은 시간 명암의 변화 없이 $(\delta x, \delta y)$ 만큼 이동



$I(x, y)$

동영상기초

- 옵티컬 플로 예측 모델
 - $\frac{\delta x}{\delta t}, \frac{\delta y}{\delta t}$: 각 축에 대한 속도 항
 - V_x, V_y 로 치환
 - $\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}, \frac{\partial I}{\partial t}$: 밝기 값 함수에 대한 편미분
 - I_x, I_y, I_t 로 치환
- 옵티컬 플로 제한 조건식

$$I_x V_x + I_y V_y = -I_t$$

동영상기초

- 블록 기반의 움직임 추정

- 정의

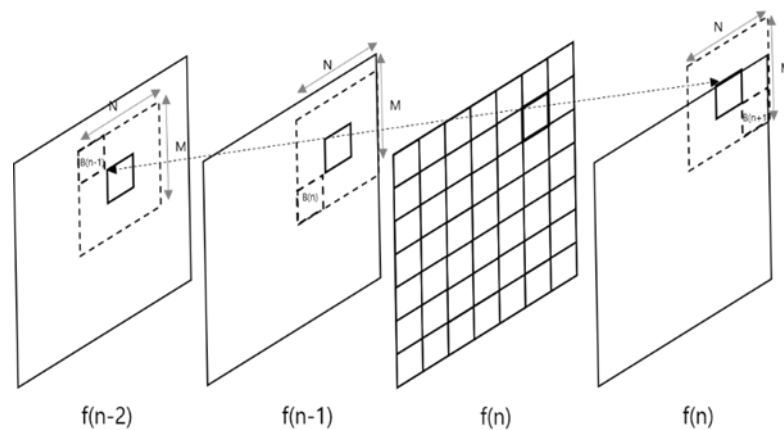
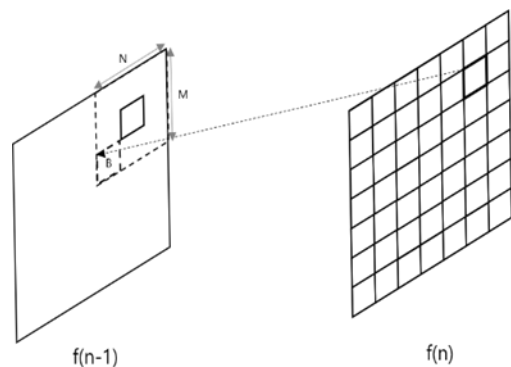
- 영상을 다양한 크기의 블록으로 분할하여 시간상 연속된 참조 영상으로부터 움직임을 추정하는 방법

- 가정

- 한 블록 내의 모든 화소가 같은 움직임 벡터를 가진다

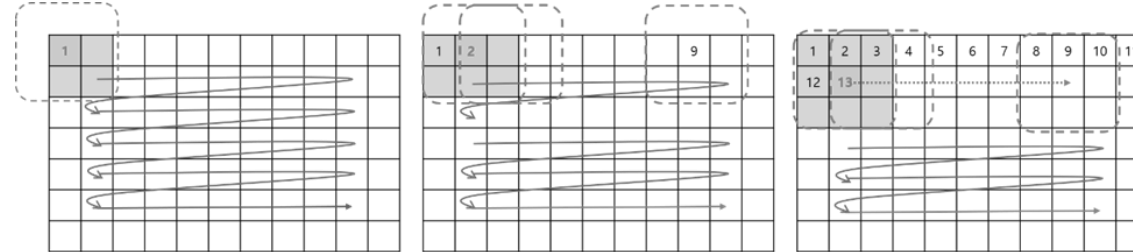
- 분류

- 순방향 움직임 예측
 - 역방향 움직임 예측
 - 양방향 움직임 예측

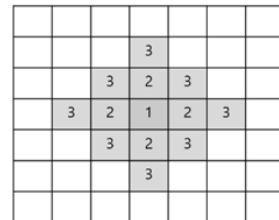


동영상기초

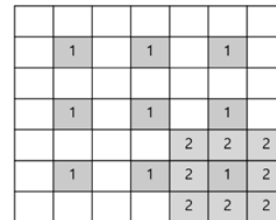
- 블록 기반의 움직임 추정
 - 움직임 벡터 추정
 - 현재 블록과 가장 유사한 목표 블록 탐색
- 방법
 - 순차 탐색
 - 다이아몬드 탐색
 - 2단계 탐색



(a) 순차 탐색



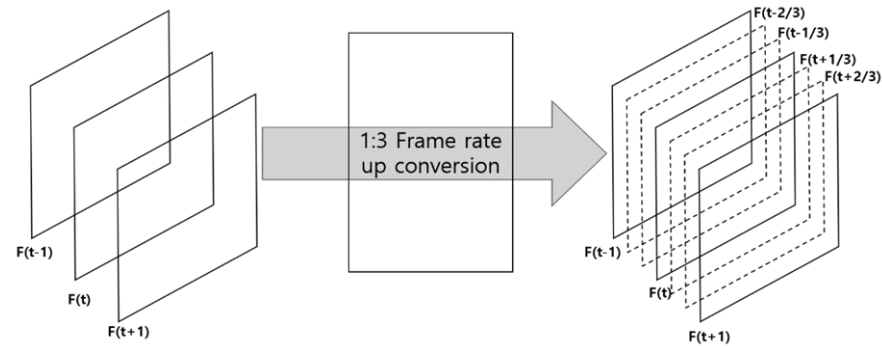
(b) 다이아몬드형 탐색



(c) 2단계 탐색

동영상기초

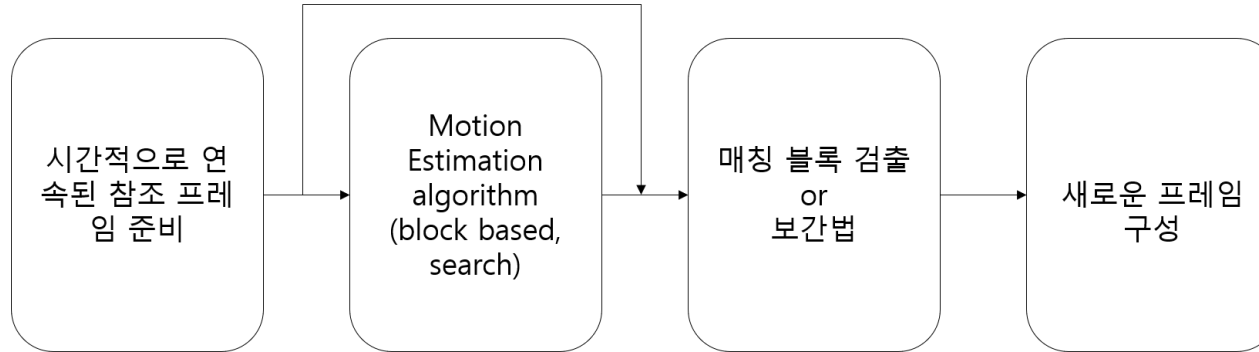
- Frame Rate Up-Conversion
 - 영상 시간 해상도 개선 기술



- 이미 존재하는 프레임 $F(t-1), F(t), F(t+1)$ 을 참조 프레임으로 사용
- 새로운 프레임 $F(t-2/3), F(t-1/3), F(t+1/3), F(t+2/3)$ 을 생성

동영상기초

- Frame Rate Up-Conversion
 - 방법



- 영상 내 객체의 움직임을 고려하지 않는 방법
 - 다양한 필터링을 통한 보간 프레임 생성
 - 장점 : 상대적 낮은 복잡도
 - 단점 : 객체 영역에서의 블러링 발생
- 영상 내 객체의 움직임을 고려하는 방법
 - 움직임 벡터 기반으로 보간 프레임 생성
 - 장점 : 객체 영역에서의 블러링 문제 해결 가능
 - 단점 : 상대적 높은 복잡도

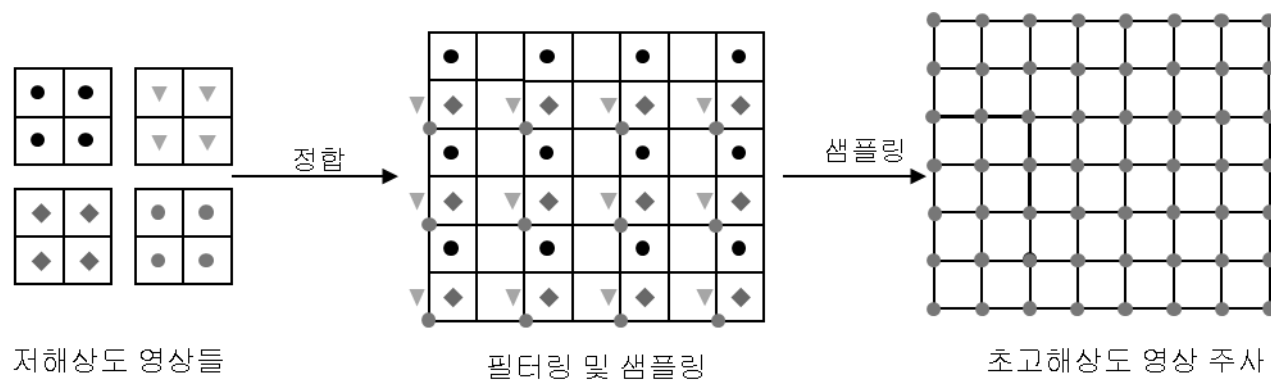
동영상기초

- Super Resolution

- 저해상도 영상으로부터 고해상도 영상 생성.

저해상도 영상 → 영상 정합 → 역 필터링 → 고해상도 샘플링 → 초고해상도 영상

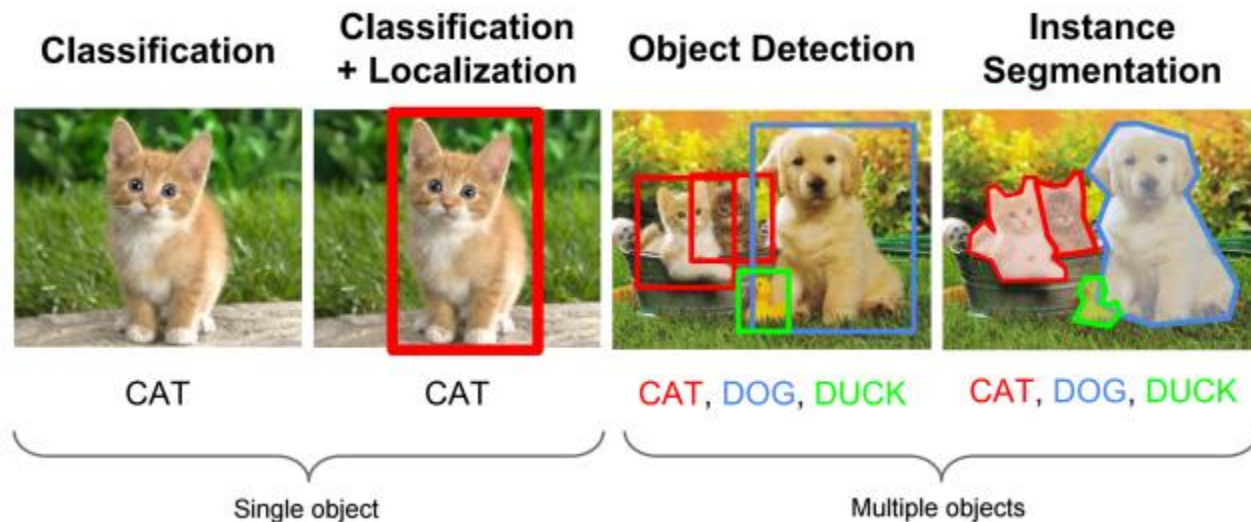
- 영상 정합 및 움직임 예측을 통한 저해상도 영상의 공통 특징 추출
- 영상 보간을 이용한 고해상도 영상 샘플링



동영상기초

- 물체 감지

- 물체 감지는 이미지 프레임에 존재하는 모든 물체를 식별하고 정확하게 라벨링하는 방법으로 크게 두 단계로 구성
- 1 : Object Localization : 여기서는 이미지에서 객체의 정확한 위치를 찾기 위해 경계 상자 또는 둘러싸는 영역을 최대한 조밀하게 결정
- 2 : 이미지 분류 : 현지화 된 객체는 객체에 라벨을 지정하는 분류기로 공급



동영상기초

- **Semantic Segmentation**

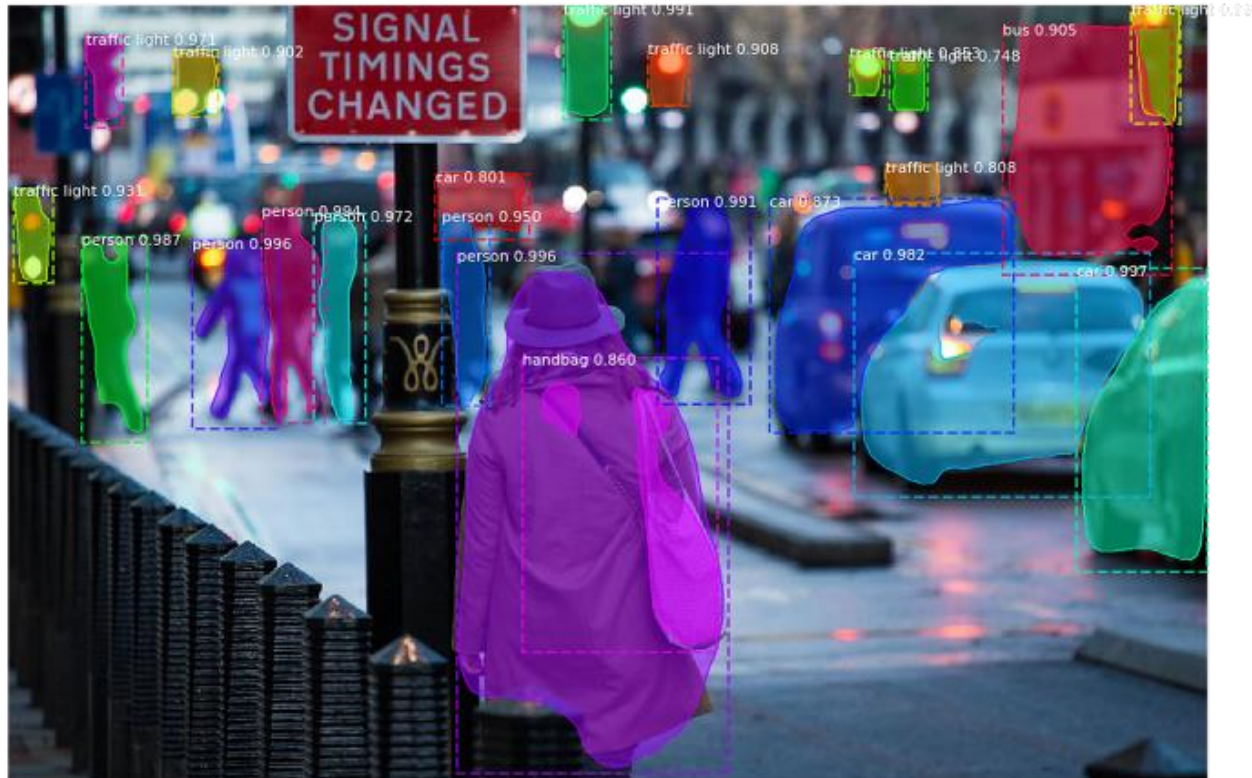
- 주어진 이미지의 각 픽셀을 특정 클래스 레이블에 연결하는 과정
- 다음 이미지에서 픽셀은 자동차, 나무, 보행자 등으로 레이블이 지정된 다음 이러한 세그먼트를 사용하여 다양한 개체 간의 상호 작용 / 관계를 찾음



동영상기초

- 인스턴스 분할

동일한 클래스의 여러 객체를 개별 객체 / 개별 엔티티로 취급한다는 점을 제외하고는 의미론적 분할과 유사한 각 픽셀에 클래스 레이블을 연결함



동영상기초

- **Panoptic Segmentation**

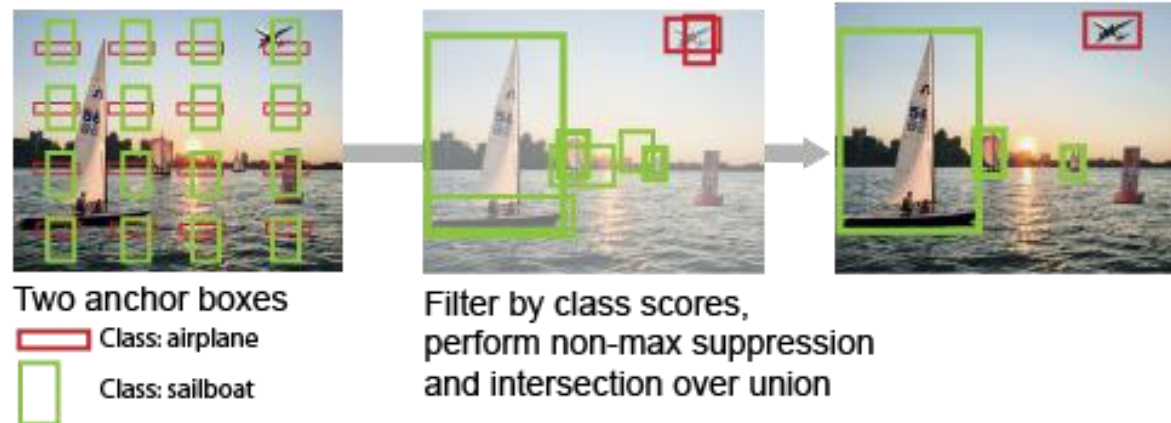
- 각 픽셀과 두 값을 연관시키는 방식으로 인스턴스와 시맨틱 분할의 조합
- 클래스 레이블과 인스턴스 번호. 또한 하늘, 도로 및 총체적으로 사물로 알려진 기타 배경 요소를 인식



동영상기초

• 앵커 상자

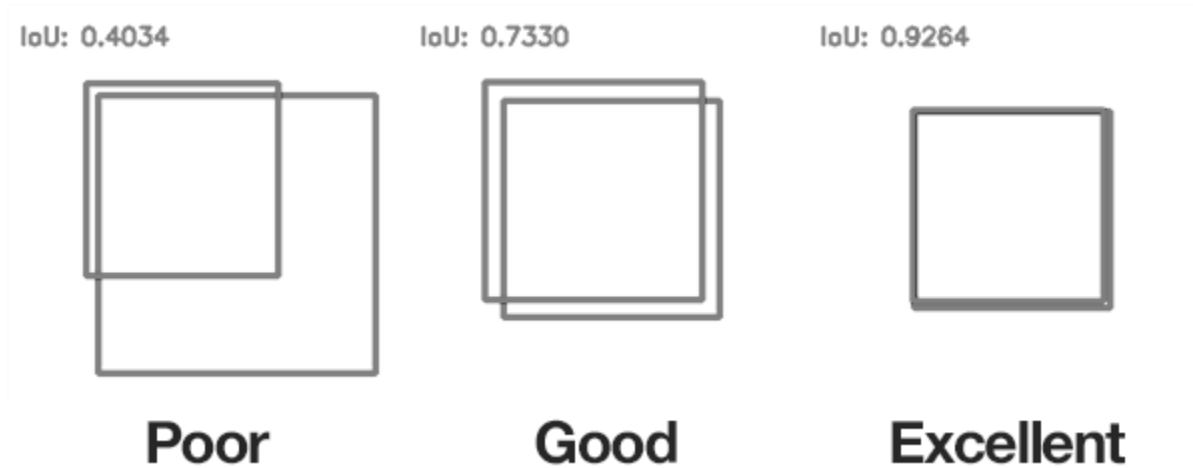
- 특정 높이와 너비의 미리 정의 된 경계 상자 집합
- 상자는 감지하려는 특정 객체 클래스의 배율 및 종횡비를 캡처하도록 정의되며 훈련 데이터 세트의 객체 크기에 따라 선택
- 감지 중에 미리 정의 된 앵커 상자가 이미지 전체에 바둑판 식으로 배열
- 네트워크는 확률과 배경, IoU (Intersection over Union) 및 모든 타일 앵커 상자의 오프셋과 같은 기타 속성을 예측함
- 예측은 각 개별 앵커 상자를 구체화하는 데 사용되며 각각 다른 개체 크기에 대해 여러 앵커 상자를 정의함
- 네트워크는 이러한 앵커 상자를 구체화하여 최종적으로 단단한 경계 상자를 출력(스케일 및 종횡 비로 정의)



동영상기초

- **IOU (Intersection over Union)**

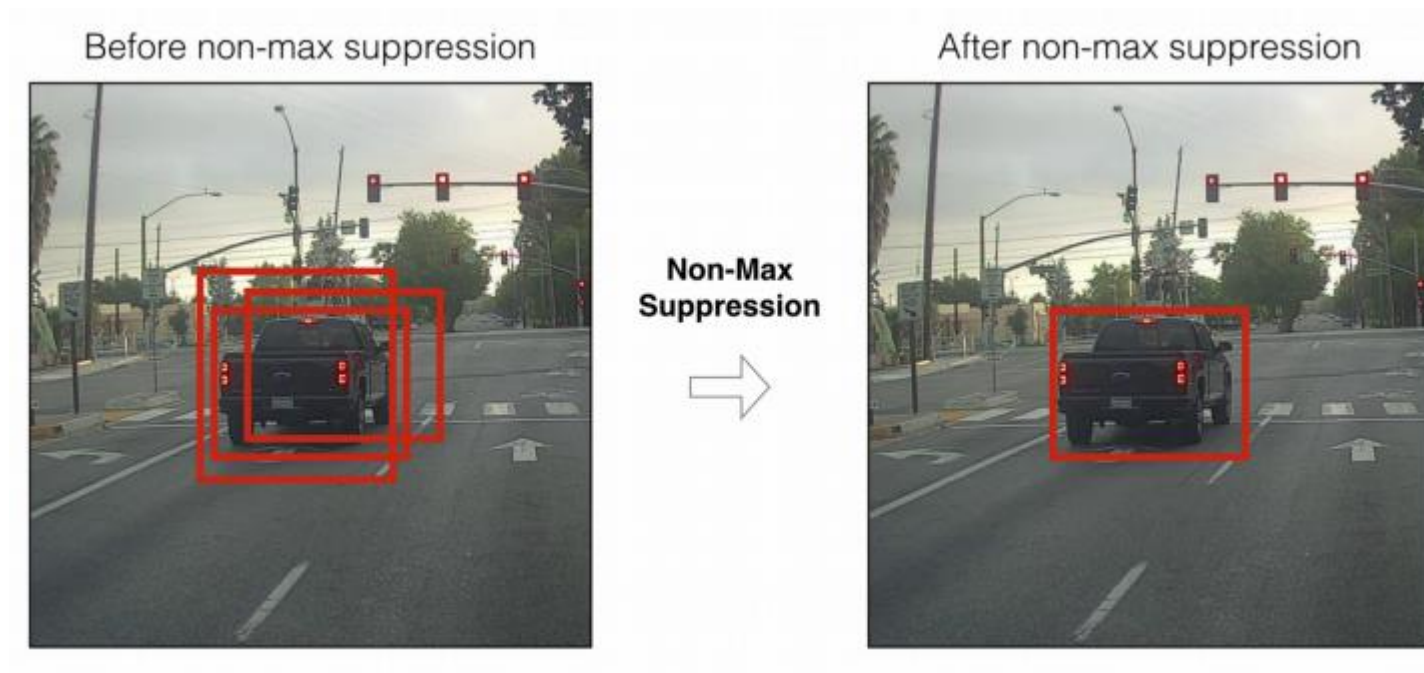
- 실제 Ground Truth에서 예측 된 경계 상자의 정확도를 확인하는 데 사용되는 평가 메트릭



동영상기초

- 비 최대 억제

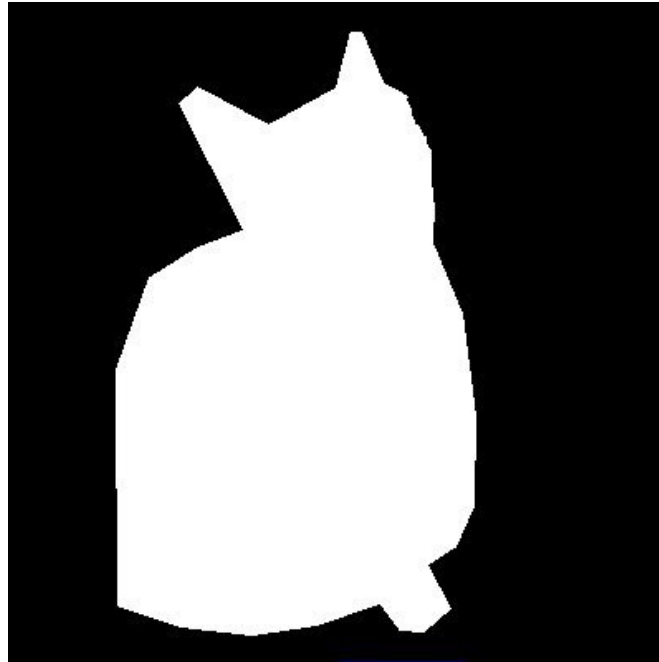
- 주어진 개체에 대해 여러 개의 상자가 있는 경우 이름에서 알 수 있듯이 기술은 최대 IOU를 가진 상자를 제외한 모든 상자를 삭제



동영상기초

- **Binary Mask**

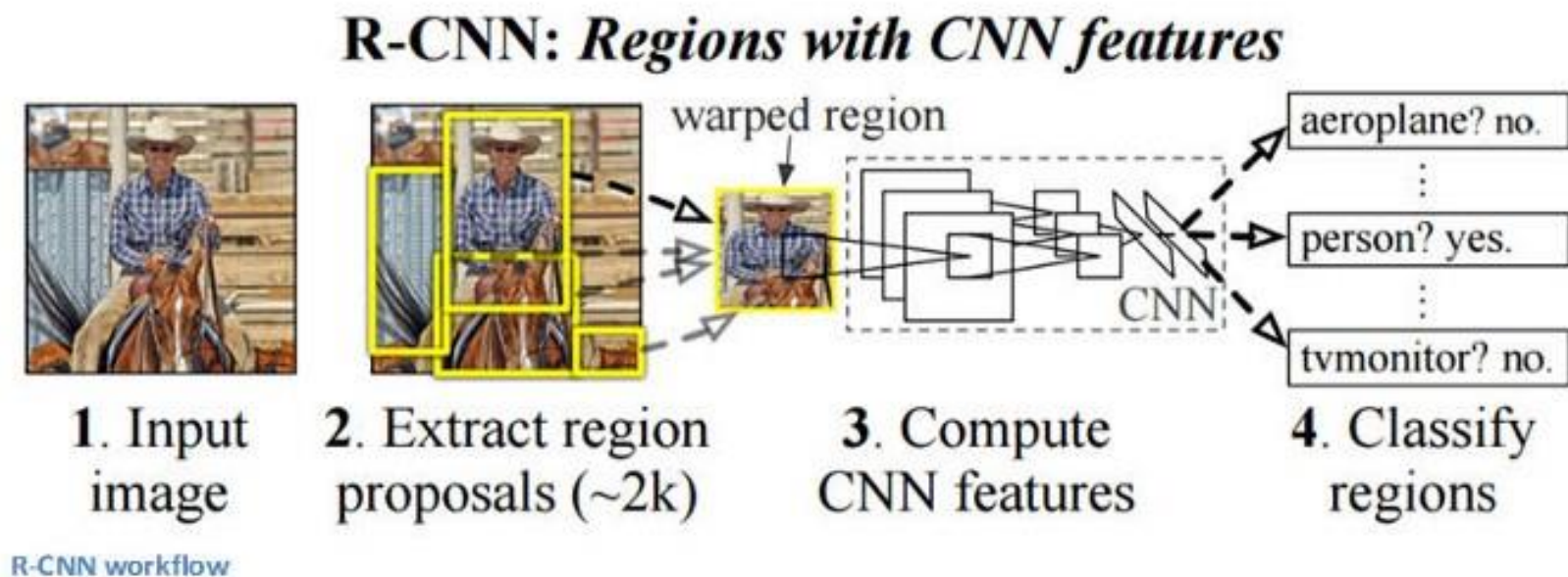
- 이미지의 동일한 픽셀 너비와 높이를 나타내는 데이터 포인트를 갖는 2D 배열
- 마스크의 각 픽셀 은 예측 된 인스턴스에 속하는지 여부에 대해 a 1또는 0(true또는 false) 레이블이 지정



동영상기초

- RCNN

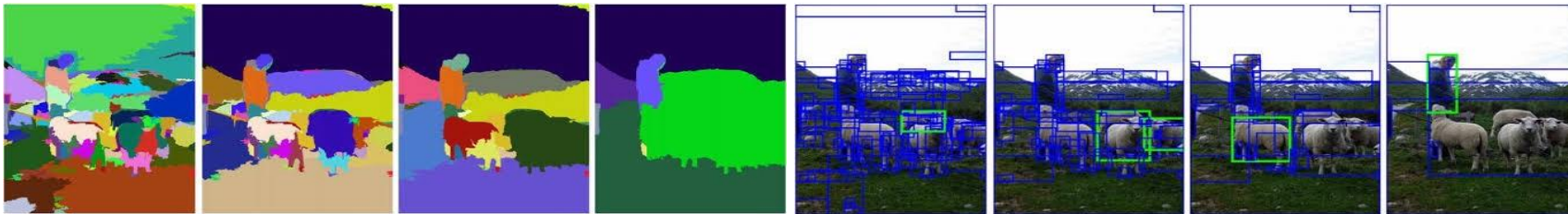
- RCNN은 지역 제안 기반 객체 감지 알고리즘으로 지역 기반 컨볼루션 신경망



동영상기초

• RCNN

- 1) 분할 : Girshick 등의 원본 rcnn 논문 [1] al. 선택 검색 방법을 사용하여 약 2000 개의 지역 제안을 생성
- 1.1) 선택적 검색 : 계층 적 그룹화 알고리즘을 사용하여 지역 제안을 생성
- 1.1.1) 초기 영역 생성 : 먼저 그래프 기반 이미지 분할 알고리즘을 실행하여 아래 이미지의 가장 왼쪽 열에 표시된 초기 영역을 획득
- 1.1.2) 유사성 측정 : 다음 기준에 따라 지역 간의 유사성을 찾습니다. 색깔 / 크기 / 모양 호환성
- 유사성 메트릭은 다음과 같이 획득 $S(r_i, r_j) = a_1 S_{\text{colour}}(r_i, r_j) + a_2 S_{\text{texture}}(r_i, r_j) + a_3 S_{\text{size}}(r_i, r_j) + a_4 S_{\text{fill}}(r_i, r_j)$
- 1.3) 재귀적 그룹화 : 이러한 초기 지역에서 시작하여 유사성 메트릭을 기반으로 이러한 지역을 재귀 적으로 그룹화. 필요한 수 또는 제안이 달성되면 중단
- 1.2) Warping : 각 지역 제안은 Convnet의 필요한 입력 크기에 맞게 크기가 조정 (스케일링)되고 타이트한 상자에 포함
- 1.3) 특징 추출 : 이러한 뒤틀린 영역 각각은 4096 길이의 특징 벡터를 출력하는 Convnet에 1 y 1로 공급
- 1.4) 분류 : 4096 길이의 특징 벡터는 객체의 존재 여부를 분류하고 레이블을 할당하는 SVM에 공급
- 1.5) Bounding Box Regressor : 클래스 레이블 외에도 rcnn은 객체의 경계 상자 좌표를 출력하는 선형회귀자를 사용
- 1.6) IOU 및 non-max 억제 : 중첩되는 경우 가장 높은 점수를받은 영역이 선택되고 나머지는 폐기



동영상기초

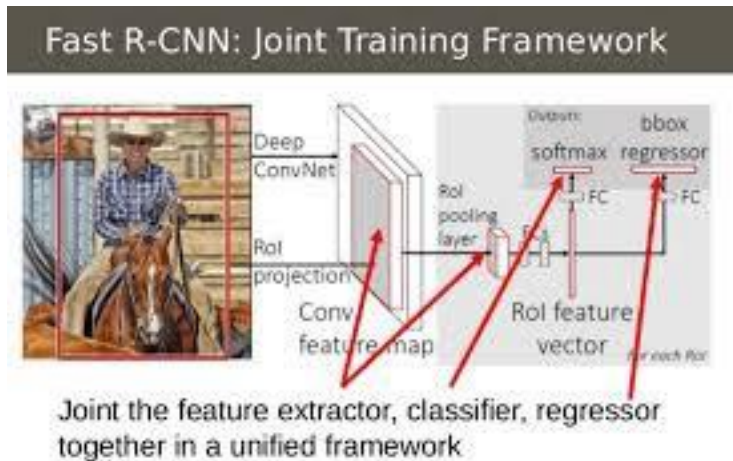
- 빠른 RCNN

- rcnn의 일부 단점을 제거하므로 rcnn의 즉석 버전
- 장점 : 1. R-CNN, SPPnet보다 높은 검출 품질 (mAP) / 2. 단일 단계 프로세스이므로 계산 시간이 단축. / 3. 중간 기능을 캐시하기 위해 추가 디스크 스토리지가 필요하지 않음
- 4. rcnn 및 SPPnet에 비해 더 적은 매개 변수.
- 방법: 1) 기능 맵 생성 : 개체 제안과 함께 전체 이미지가 Convnet에 공급. Conv 레이어와 Max Pooling 레이어를 통과하면 피쳐 맵을 획득 / 2) ROI 풀링 : 기능 맵의 관심 영역 (ROI)에는 $y(r, c, h, w)$ 좌표가 제공됩니다. 이 ROI는 ROI 풀링 레이어를 통과하여 $H \times W$ 기능 맵을 획득 / 3. 완전 연결 계층 : 이 피쳐 맵은 FC 계층으로 추출 된 다음 FC 계층을 통해 클래스 확률 예측을 위한 소프트 맥스와 경계 상자 회귀 출력을위한 회귀 자로 전달

동영상기초

- 빠른 RCNN

- rcnn의 일부 단점을 제거하므로 rcnn의 즉석 버전
- 장점 : 1. R-CNN, SPPnet보다 높은 검출 품질 (mAP) / 2. 단일 단계 프로세스이므로 계산 시간이 단축. / 3. 중간 기능을 캐시하기 위해 추가 디스크 스토리지가 필요하지 않음 4. rcnn 및 SPPnet에 비해 더 적은 매개 변수.
- 방법: 1) 기능 맵 생성 : 개체 제안과 함께 전체 이미지가 Convnet에 공급. Conv 레이어와 Max Pooling 레이어를 통과하면 피쳐 맵을 획득 / 2) ROI 풀링 : 기능 맵의 관심 영역 (ROI)에는 $y(r, c, h, w)$ 좌표가 제공됩니다. 이 ROI는 ROI 풀링 레이어를 통과하여 $H \times W$ 기능 맵을 획득 / 3. 완전 연결 계층 :이 피쳐 맵은 FC 계층으로 추출 된 다음 FC 계층을 통해 클래스 확률 예측을 위한 소프트 맥스와 경계 상자 회귀 출력을위한 회귀 자로 전달



동영상기초

- RCNN 마스크

- Mask RCNN은 병렬 마스크 출력 분기를 추가하여 Faster Rcnن을 확장합니다. 인스턴스 분할에 사용되는 매우 중요한 방법

- 자극:

- 1. 더 빠른 Rcnن, Yolo 및 기타 물체 감지 알고리즘은 경계 상자과 해당 상자과 관련된 클래스확률 레이블을 출력
- 2. 인간으로서 우리는 실물 물체 주변에 상자를 그려서 찾는 것이 아니라 물체를 감지하기 위해 물체의 윤곽과 자세를 감지
- 3. 이와 관련하여 마스크 rcnn은 인간의 객체 인식 스타일에 근접
- 4. 마스크 rcnn에 대한 연구는 우리가 panoptic 세분화, 사람 키폰트 감지, 스포츠 포즈 추정 등의 영역으로 나아가도록 동기를 부여
- 5. 모든 자율 주행 차는 마스크 rcnn의 기본 개념을 사용

- 아키텍처 및 구현 :

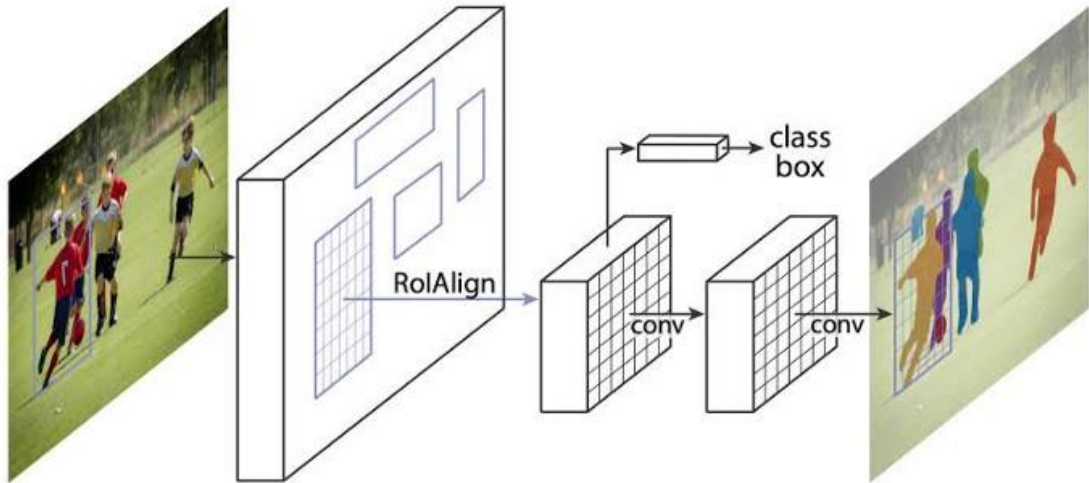
- 1. 마스크 R-CNN은 동일한 첫 번째 단계 (RPN)로 동일한 2 단계 절차를 채택 / 2. 두 번째 단계에서, Mask R-CNN은 클래스 및 박스 오프셋을 예측하는 것과 동시에 각 ROI에 대한 바이너리 마스크를 출력 /

- ROI 정렬 레이어 :

- Faster Rcnن의 ROI 풀 레이어는 부동 소수점 값을 플로팅하는 것과 같은 양자화를 수행하고 Maxpool과 같은 집계 기능을 수행 / ii. 이러한 작업은 인스턴스 분할에 필요한 미세한 픽셀 대 픽셀 배열을 파괴하는 거친 특징을 초래 / 이를 해결하기 위해 Mask Rcnن은 픽셀 정렬을 유지하고 마스크 정확도를 향상시키는 양자화 대신 이중 선형 보간을 사용하는 ROI 정렬 레이어를 사용

동영상기초

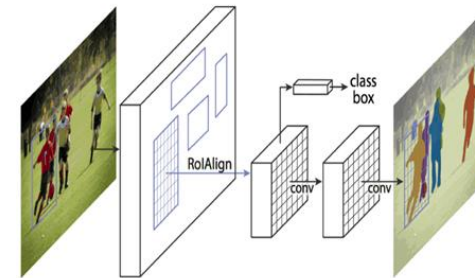
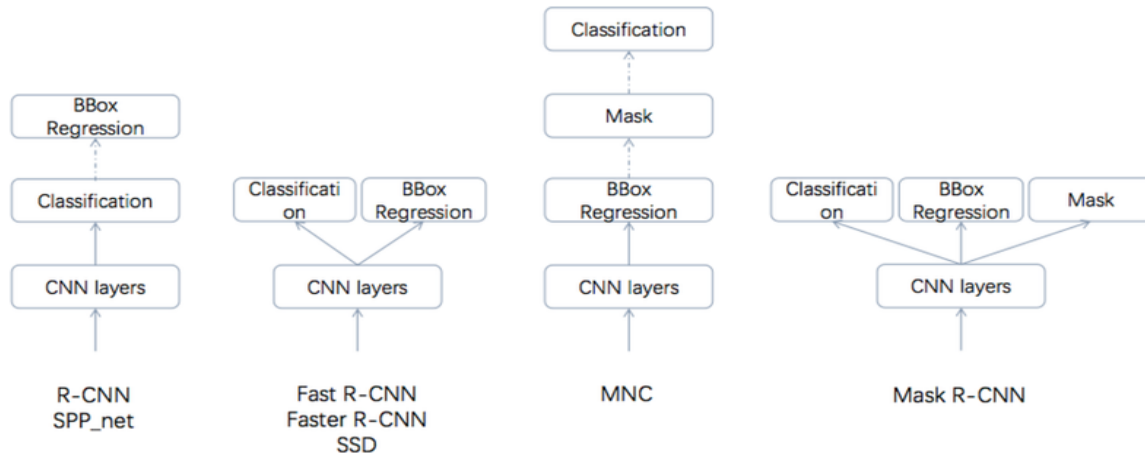
- RCNN 마스크



동영상기초

- 딥러닝 기반 객체 인스턴스 분할 모듈

- Faster R-CNN을 Object Detection 역할을 하도록 하고 각각의 RoI(Region of Interest)에 Mask segmentation을 해주는 작은 FCN(Fully Connected Network)을 추가하는 방식 적용
- 딥러닝 기반 다중 객체 분할 모듈
- CNN 레이어를 사용해 피쳐 맵을 추출 후 RPN을 통해 개체를 포함하고 있을 가능성이 높은 윈도우 생성하고 고정 크기로 조정된 뒤(RoI 풀링) 클래스 확률과 해당 객체에 대한 더욱 정확한 경계 박스를 예측하고 여러 객체 카테고리에 대한 분류 대신 윈도우의 객체 포함 유무에 대한 이진 분류(Binary Classification)로 수행



The Mask R-CNN framework for instance segmentation



동영상기초



사람 인체/자세 3D AI 데이터
<https://aihub.or.kr/aidata/8014>

사람 동작 영상 AI 데이터
<https://aihub.or.kr/aidata/138>

피트니스 자세 이미지 AI 데이터
<https://aihub.or.kr/aidata/8051>

사람 인체.자세 데이터 개요

1. 데이터셋 목적
 - 2D 인체 영상을 3D 모델로 변환 시, 자세(pose)와 형태(shape)를 추론하기 위한 데이터 셋
2. 데이터셋 구성
 - 이미지 및 2차원 관절 정보
 - 3차원으로 표현되는 인체 자세(관절 위치)
 - 3차원으로 표현되는 인체 형태(인체 표면 폴리곤)
3. 데이터셋 구축 범위
 - 사람의 행동 및 동작 70여종 이상 정의
 - 3D 모델 50만 건, 2D 이미지 200만장 이상

데이터셋 활용 AI 임무

2D 2차원 관절 추정(2D Human Pose Estimation)

- 이미지로부터 사람의 2차원 관절 좌표 추정
- Body Box 찾은 후 추정하는 "Top-down" 방식
- 관절을 찾아 연관관계 분석 "Bottom-up" 방식

3D 3차원 관절 추정(3D Human Pose Estimation)

- 이미지로부터 사람의 3차원 관절 좌표 추정
- Single-view 3D Pose Estimation
- Multi-view 3D Pose Estimation

3D 3차원 형태 추정(Human Shape Estimation)

- 이미지로부터 사람의 3차원 형태를 추정
- Model-based(Parametric Human Body Model)
- Model-free

관절 정보

ID	정보	ID	정보
0	오른쪽 발	8	목 상단
1	오른쪽 무릎	9	머리
2	오른쪽 둔부	10	오른쪽 손목
3	왼쪽 둔부	11	오른쪽 팔꿈치
4	왼쪽 무릎	12	오른쪽 어깨
5	왼쪽 발	13	왼쪽 어깨
6	골반	14	왼쪽 팔꿈치
7	흉부	15	왼쪽 손목

헬스장 운동 20여 종

집 운동 10여 종



첫 번째 Annotation

각 비디오마다 사람이 하고 있는 운동의 종류와 미리 정의된 운동 상태의 여부

- ✓ 미리 정의된 운동 상태는 전문 서적들과 전문가들의 의견을 따라 객관화된 규칙을 만들 예정
- ✓ 각 운동마다 약 5개의 운동 상태가 정의되어 있음.

두 번째 Annotation

사람의 3D keypoint coordinate

기존보다 더 dense한 keypoint annotation 제공 예정

MSOCO human keypoint set HumanFit human keypoint set

바이오시스템

- R-Project

- 주제는 무엇인가?
- 필요로 하는 또는 고려하는 디바이스는 무엇인가?
- 필요로 하는 또는 고려하는 소프트웨어 기능은 무엇인가?
- 어떠한 신호를 사용하는가?
- 물리량 검출을 위한 센서는 무엇인가?
- 디지털변환 및 처리 방법은 무엇인가?
- 요구되는 신호처리 분석 방법은 무엇인가?
- 활용하고자 하는 바는 무엇인가?