

딥러닝 기반 visual tracking

(Deep Learning for Visual Tracking: A Comprehensive Survey, 2021)

김동원

목차

CNN auto encoder 적용

CNN base visual method

SNN base visual method

RNN & GAN base visual method

GAN base visual method

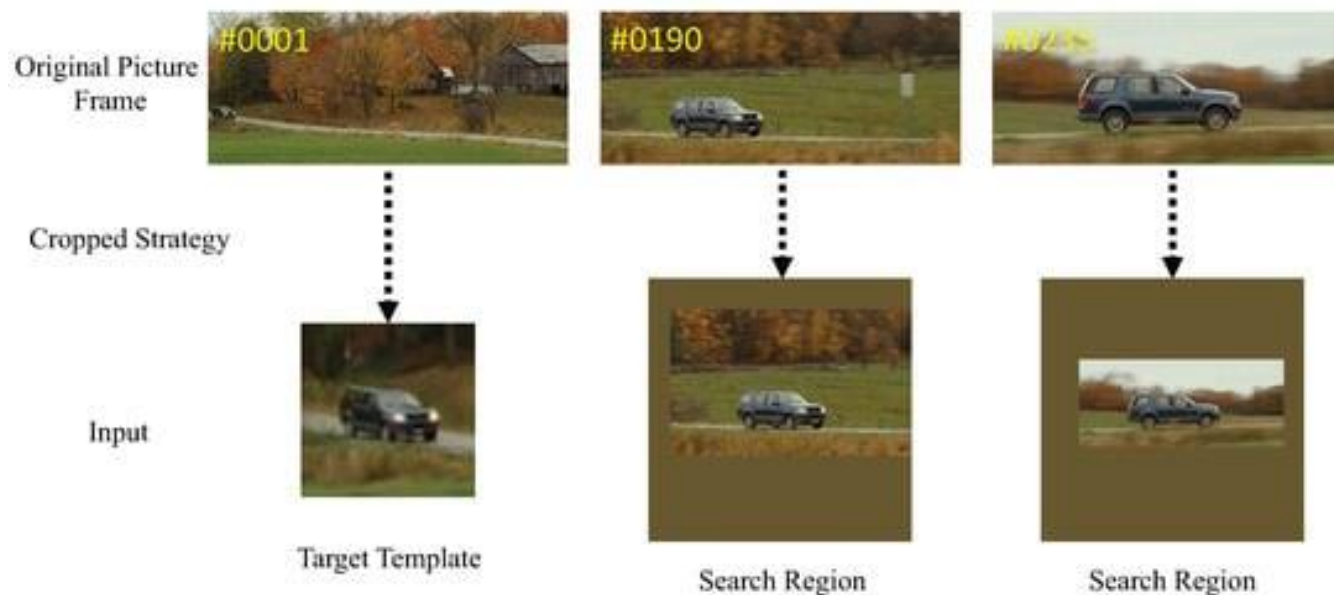
Custom network base method

CNN base visual method

CNN은 비주얼 tracking에서 많이 사용되었지만 최근에는 다른 아키텍처 들도 효율성과 robustness를 향상시켜 개발되고 있다.

CNN 모델은 대용량 지도학습의 의존하고 시간 종속성을 무시하는 것, & 온라인 적용에 의한 계산 복잡도가 존재하는 것이 한계로 가진다.

Target exemplar과 search region의 유사성을 측정하는 방법으로 접근



Image(Siamese Network Tracker Based on Multi-Scale Feature Fusion)

Online learning이란 : 초기에는 추적 대상의 특징을 학습하고 추적 도중에는 추적 대상의 특징을 업데이트 하여 실시간으로 객체 추적을 수행하는 방법

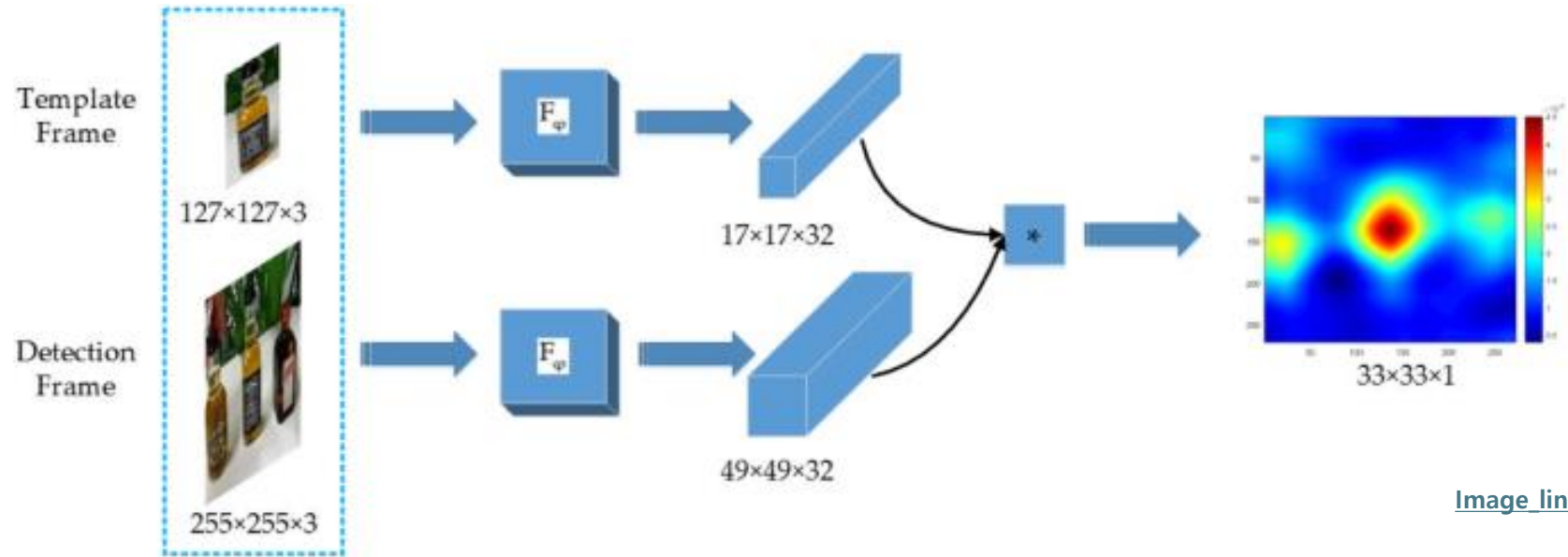
태그를 태워달라야 하는 경우에 유용하다

SNN base visual method

SNN(Siamese neural network)

SNN은 두개 이상의 CNN layer/block/networks를 가지로 사용하여 유사성을 학습하는 방법이다.

Online 적용과 물체 occlusion의 분야에서 아직 연구 중 이다.



RNN & GAN base visual method

RNN과 GAN은 tracking 분야에서 제한적인 연구가 진행중

RNN은 동영상 프레임에서 시간적 정보를 포착하는데 사용하지만 long-term learning dependencies와 학습의 비용이 크다는 한계가 존재

GAN을 활용한 모델은 다양한 외형변화를 캡처하여 모델의 성능을 개선한다

GAN, generator로 positive data imbalance문제를 해결방안으로 제시 되었지만 일반화와 학습에 문제가 있다.

Long term learning dependency란 과거의 정보가 현재와 미래의 결정에 중요한 영향을 미치는 현상

Custom network base visual method

여러 네트워크 레이어와 아키텍처를 결합하여 추적 성능을 향상시키고, 계산 복잡도를 줄이는 방법을 위해 사용됨

메타학습을 통해 빠르게 적응할 수 있는 모델을 구축하고 다양한 추적 문제를 해결하는데 유용

메타학습이란 학습을 학습하는 접근 방식으로 새로운 작업에 대해 빠른 적응하는 능력을 학습하는 것 (새로운 대상이나 환경에 대한 변화)

Tracking 에서 중요한 문제점

Robust target representation: 추적 중 변할 수 있는 물체의 외관에 대해 안정적으로 추적하는 것

Balancing training data: 주어진 이미지(positive data)이 매우 적어 과적합 발생 위험이 증가하고, 학습데이터의 노이즈가 결과에 큰 영향

Computational complexity problem: CNN과 같은 구조는 높은 연산 복잡도를 가진다

CNNs 모델의 해결방안

사전학습 네트워크를 사용하고 대규모 데이터 셋을 사용,
시간과 공간의 정보를 모델 일반화 향상을 위해 통합한다.

계산 복잡도를 위해 단순한 네트워크 사용, 여러 작은 네트워크로 세분화

SNNs 모델의 해결방안

깊은 여러 레벨의 feature들을 결합, 여러 loss함수를 사용 등 약간의 방해를 통해 robustness를 향상시키는 방법
오프라인 학습만 진행하면 일반화에 약한 모습을 확인 하여 온라인 업데이트 전략 사용

Correlation filter를 사용

Correlation filter 이란 물체 트래킹에서 많이 사용되는 방법으로 이미지와 템플릿 간의 유사성을 측정하는 방법, search region과 목표 물체의 템플릿과 입력 이미지 패치 간의 상관 연산을 통해 유사도가 높은 위치의 물체를 새로운 물체로 인식

모델의 학습

딥러닝 기반 추적 모델은 기울기 기반 최적화 방법을 사용하는데 오프라인, 온라인 파인튜닝에 따라 연산 복잡도의 차이를 가진다.

데이터 셋은 LaSOT, TrackingNET과 같은 대규모 데이터 셋을 활용한다.

오프라인 학습만 진행된 경우: 일반적인 물체의 특징들을 제공하고 불균형한 훈련 데이터와 고정된 가정을 인한 과적합 위험을 줄이다.

온라인 학습만 진행된 경우: 평가 비디오에서 보지 못한 대상을 구별하기 위해 온라인에서 훈련시켜 파라미터를 적응시킨다. 대규모 훈련데이터에서 오프라인 훈련이 시간소모가 크고 사전 훈련된 모델들이 추적할 특정 대상을 잘 표현하지 못하기 때문에 온라인과 오프라인을 교대로 활용한다.

온라인 & 오프라인 : 최대 성능을 위해 온라인과 오프라인 모두 학습되는데, 이는 오프라인과 온라인에서 학습된 특성들은 도메인에 특화되고, 주로 전경 정보, 장애물 구별을 잘한다.

데이터 augmentation: 기하학적 변환과 색상 공간 증강은 tracking에서 많이 사용되는 방법이고 GANs를 사용하는 방법은 물체의 여러 외형변화를 캡처하여 추적하는 것의 성능이 좋은 영향을 미친다.

모델의 objective function

목적함수에 따라 분류, 회귀, 분류 & 회귀

분류기반 목표 함수: 제안된 바운딩 박스를 물체와 배경으로 분류하는 것을 목표로 한다. 동일한 레이블로 지정된 대상이 여럿 있을 경우 문제가 발생할 수 있고, 임의의 외형으로 클래스를 분류하고 추적하면 다른 대상을 인식할 수도 있다.

회귀 기반 목적함수: 프레임에서 직접 대상을 로컬라이징하고 l_2 , l_1 함수를 최소화 시키고, 신경망을 효율적으로 최적화 할 수 있다.

분류 & 회귀 : 배경 분류와 회귀 모두 사용하는 방법은 많은 추적 모델이 사용하고, 추적 , 탐지 및 연속적 로컬라이징과 이 과정간 프레임 격차를 해소할 수 있다. 대상의 영역은 분류점수를 통해 사용되고, 회귀 매칭 함수를 통해 최적화의 효율성과 정확성을 향상 시킨다.

모델의 metric

Short term과 long term에 따라 측정 metric이 다르다

Short term metric: CLE(중심위치 오차), 정규화된 정밀도(유클리디안 거리로 계산), Accuracy, 실제와 예측 바운딩박스 교차비율, robustness, failure score(추적 모델이 물체를 놓쳤을 경우 다시 초기화되는 횟수) 기대평균 겹침(EAO)정확도와 강건성 점수를 결합, AOS(average overlap score), AUC

Long term metric: precision, 재현률, F-score, MaxGM(추적기가 얼마나 잘 추적하는지)

모델의 metric

Short term과 long term에 따라 측정 metric이 다르다

Short term metric: CLE(중심위치 오차), 정규화된 정밀도(유클리디안 거리로 계산), Accuracy, 실제와 예측 바운딩박스 교차비율, robustness, failure score(추적 모델이 물체를 놓쳤을 경우 다시 초기화되는 횟수) 기대평균 겹침(EAO)정확도와 강건성 점수를 결합, AOS(average overlap score), AUC

Long term metric: precision, 재현률, F-score, MaxGM(추적기가 얼마나 잘 추적하는지)