Siamese 네트워크의 템플릿 업데이트를 통한 적외선 영상 표적 추적 기법 연구

김태윤[1]*, 이인호[2], 박찬국[2]

서울대학교 지능형우주항공시스템/자동화시스템공동연구소[1]
서울대학교 항공우주공학과/자동화시스템공동연구소[2]

Infrared Target Tracking Method Based on Siamese Network with Template update

Tae Yoon Kim[1]*, In Ho Lee[2], Chan Gook Park[2]

Key Words : Infrared Target Tracking(적외선 표적 추적), Siamese Network(샴 네트워크), Attention Module (어텐션 모듈), Region Proposal Network(영역 제안 네트워크)

서 론

최근 적외선 영상에서 표적 추적은 딥러닝 기법, 특히 Siamese 네트워크[(1)]의 도입으로 성능이 크게 향상되었다. Siamese 네트워크는 초기 표적 이미지와 입력 영상을 대칭적인 구조로 비교하여 유사도를 기반으로 추적을 수행하며, 이는 온라인 추적기보다 성능과 속도에서 우수하다고 평가받는다. 그러나 기본적인 Siamese 네트워크는 첫 번째 프레임의 표적 이미지만을 사용하기 때문에 표적의 외형 변화나 주변 방해 요소에 취약한 단점이 있다.

이러한 한계를 극복하기 위해 템플릿을 추적 중에 업데이트하거나, 네트워크에 어텐션 모듈을 도입하여 중요한 특징에 집중하는 접근법[(3)]들이 제안되었다. 템플릿 업데이트는 선형적 방식이나 사전 학습된 신경망을 활용하는 UpdateNet[(4)] 등이 있으며, 어텐션 모듈은 특징 맵에서 채널, 공간 정보와 같은 중요한 특징을 더욱 강조해 복잡한 배경이나 주변 방해 요소를 억제한다. 하지만 적외선 영상에서 이러한 기법들의 성능 평가는 충분히 이루어지지 않았다.

따라서 본 연구는 사전학습 기반 템플릿 업데이트와 어텐션 모듈을 포함한 Siamese 네트워크를 적외선 영상에서 평가하고, 두 기법을 통합한 네트워크를 제안하여 표적의 외형 변화와 방해 요소가 많은 환경에서의 성능 향상을 목표로 한다.

Siamese의 템플릿 업데이트 및 어텐션

UpdateNet은 기존 템플릿 업데이트 방식의 한계를 극복하기 위해 제안된 네트워크로, 입력 템플릿 사이에서 비선형적인 관계를 학습하여 최적의 템플릿을 생성한다.

예측 과정에서는 현재의 프레임 T_i, 이전까지 누적된 템플릿 $\tilde{T}_{i-1}$, 초기 템플릿 T_{GT}을 결합하여 다음 프레임을 위한 업데이트된 템플릿 $\tilde{T}_i$을 예측한다. 이는 다음과 같은 수식으로 나타낼 수 있다.

$$\tilde{T}_i = \phi(T_{GT}, \tilde{T}_{i-1}, T_i) \quad (1)$$

ϕ는 현재 프레임 입력을 통해 업데이트된 템플릿을 다음 프레임의 실제 템플릿 T_{GT}^{i+1}과의 유클리드 거리를 최소화하도록 학습되어, 추적 환경의 변화를 반영하여 템플릿 업데이트를 수행하게 된다.

$$L_2 = \left\| \phi(T_{GT}, \tilde{T}_{i-1}, T_i - T_{GT}^{i+1} \right\|_2^2 \quad (2)$$

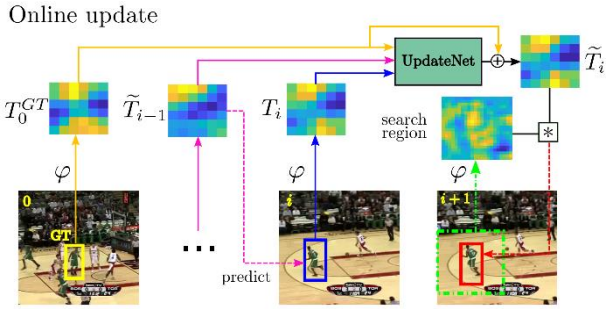

Fig. 1. Framework of UpdateNet

한편, SiamAtt는 기존의 SiamRPN의 구조에 어텐션 모듈을 추가하여 표적의 위치를 더욱 정밀하게 예측하고, SiamRPN의 classification 결과와 가중합을 통해 최종 표적의 위치를 예측한다. 어텐션 모듈은 두 층의 컨볼루션 레이어로 구성되어, 템플릿과 입력 프레임의 Correlation 특징 맵을 입력받아 공간적 특징을 강조한다. 생성되는 어텐션 스코어의 사이즈는 앵커의 사이즈를 k라고 할 때, $W * H * k \times 1$이 된다.

어텐션 모듈은 학습 과정에서 표적의 중심을 양성 샘플, 이외는 모두 음성 샘플로 간주한다. $p_{i,j}$를 (i,j)에서 네트워크의 예측 스코어로, $y_{i,j}$를 참값으로 하며, Focal loss를 통해 학습된다. 또한 데이터셋에서 양성샘플과 음성샘플의 불균형을 조절하기 위해 α와 β를 가중치로 두어 안정적인 학습을 수행한다. 이를

식으로 나타내면 아래와 같다.

$$\mathcal{L}_{att} = \begin{cases} \sum_{i=1,j=1}^{H,W} (1-p_{i,j})^{\alpha} \log(1-p_{i,j}) \; if \; y_{i,j} = 1 \\ (1-y_{i,j})^{\beta} (p_{i,j})^{\alpha} \log(1-p_{i,j}) \quad otherwise \end{cases} \quad (3)$$

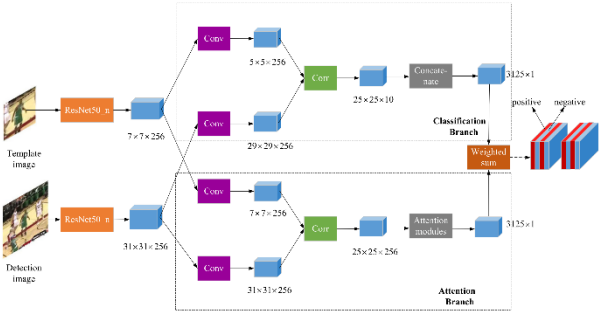

Fig. 2. Framework of SiamAtt

제안하는 통합 네트워크의 구조는 UpdateNet의 템플릿 업데이트 매커니즘과 SiamAtt의 어텐션 모듈을 결합하여 상호 보완적으로 동작한다. UpdateNet은 기존 템플릿과 입력 이미지를 결합하여 대상의 현재 상태를 반영한 최신 템플릿 $\tilde{T}_i$ 를 생성하고, 이는 SiamAtt의 입력으로 사용된다. 이러한 구조는 고정된 템플릿 사용으로 발생할 수 있는 오류를 줄이고, 대상의 외형 변화와 환경 변화에 더욱 유연하게 대응할 수 있도록 한다. 이어서 SiamAtt는 업데이트된 템플릿을 기반으로 어텐션 모듈을 통해 배경 잡음이나 방해 요소를 효과적으로 억제하고, 대상의 위치를 보다 정확하게 예측하게 된다. 이는 궁극적으로 실시간으로 변화를 반영할 수 있는 템플릿 업데이트와 어텐션 모듈을 통한 불필요한 정보의 최소화를 결합하여, 추적 성능을 향상시키게 된다.

시뮬레이션 결과

네트워크는 LSOTB-TIR[2] 데이터셋을 활용하여 학습 되었으며, 성능 확인은 외형 변화와 주변 방해 요소로 인해 일반 모델에서는 추적 성능이 떨어지는 데이터를 선별하여 수행하였다. SiamRPN을 기본 네트워크로 하여 UpdateNet, SiamAtt를 각각 적용했을 때와 모두 적용했을 때의 성능을 비교하였다. Fig. 3.의 그림은 제안하는 기법을 통해 얻어진 적외선 표적 추적 결과이다. 또한 Table 1.의 표는 추적 결과에서 측정된 mIoU를 비교하였으며, 제안된 기법이 가장 높은 성능을 보임을 확인할 수 있다.

Fig. 3. Result of Infrared Target Tracking

Table 1. Score of each methods

Methods	Update	Attention	mIoU
SiamRPN	X	X	0.592
UpdateNet	O	X	0.617
SiamAtt	X	O	0.632
Proposed	O	O	**0.639**

결 론

본 연구에서는 UpdateNet의 템플릿 업데이트와 SiamAtt의 어텐션 모듈을 결합한 통합 Siamese 네트워크를 제안하였다. 실험 결과, 제안된 기법은 표적의 외형 변화와 방해 요소가 많은 환경에서도 성능이 향상되었으며, 템플릿 업데이트와 어텐션 모듈이 효과적으로 성능을 개선함을 확인할 수 있었다.

후 기

이 연구는 인공지능 비행제어 특화연구실 프로그램의 일환으로 국방과학연구소와 방위사업청의 지원으로 수행되었음(UD230014SD).

참고문헌

1) Bertinetto, L., Valmadre, J., Henriques, J. F., Vedaldi, A., and Torr, P. H. S., “Fully-Convolutional Siamese Networks for Object Tracking,” European Conference on Computer Vision (ECCV) Workshops, 2016, pp. 850-865.

2) Zhu, X., Wang, H., Wu, B., and Yan, J., “LSOTB-TIR: A Large-Scale High-Diversity Thermal Infrared Single Object Tracking Benchmark,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2019, pp. 2003-2012.

3) Li, B., Wu, W., Wang, Q., Zhang, F., Xing, J., and Yan, J., “SiamAtt: Siamese Attention Network for Visual Tracking,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 4854-4863.

4) Zhu, Z., Wang, Q., Bo, L., Wu, W., Yan, J., and Hu, W., “UpdateNet: Learning the Model Update for Siamese Trackers,” IEEE International Conference on Computer Vision (ICCV), 2018, pp. 2512-2521.