RGB 영상과 열 적외선 영상 기반 객체 탐지 알고리즘 수행 및 성능 비교

김신(1), 이예지, 윤경로, *임한신(2), *이희경, *추현곤, *서정일
건국대학교, *ETRI
new.xin22@gmail.com(1), hslim@etri.re.kr(2)

Object Detection and Performance Comparison based on RGB image and thermal infrared radiation

Kim Shin, Lee Yegi, Yoon Kyoungro, *Lim Hanshin, *Lee Hee Kyoung, *Choo Hyon-gon, *Seo Jeongil
Konkuk University, *ETRI

요 약

현재 대부분의 객체 탐지 알고리즘은 RGB 영상을 기반으로 개발되고 있다. 하지만 안개가 끼거나 비가 오는 날 또는 밤중에 촬영한 RGB 영상은 흐리거나 잘 보이지 않아 높지 않은 객체 탐지 결과를 보여줄 수 있다. 열 적외선 영상은 열 센서로 인해 만들어지는 영상으로 RGB 영상에 비해 기상조건이나 촬영 시간대에 상관없이 취득 될 수 있다. 본 논문에서는 RGB 영상과 열 적외선 영상을 기반으로 객체 탐지 알고리즘을 수행하고 각 영상에 따른 객체 탐지 성능을 비교한다. 야간에 취득한 RGB 영상과 열 적외선 영상에 객체 탐지를 수행하였으며, 열 적외선 영상 기반 결과가 RGB 영상 기반일 때 보다 더 높은 정확도를 보여주었다. 추가적으로 밤 시간대의 RGB 영상과 열 적외선 영상을 선정하여 객체 탐지 네트워크를 튜닝하였으며, fine-tuned 네트워크를 이용하여 객체 탐지한 실험 결과 역시 열 적외선 영상이 RGB 영상보다 더 높은 객체 탐지 정확도를 보이는 것을 확인할 수 있었다.

1. 서론

수 년간 딥 러닝 기술이 극도로 발전하면서, 객체 탐지, 얼굴 탐지, 객체 세그먼테이션 등 다양한 인공 지능 영역에서 정확도의 향상을 보였다. 객체 탐지 분야의 경우, 이미 기술이 고도화되어 객체 인식 기반으로 교통량을 분석하는 솔루션을 만드는 등 실제 산업에 적용되는 사례도 많아졌다.

대부분의 객체 인식 알고리즘은 RGB 영상을 기반으로 훈련되었고 실 생활에서 사용되고 있다. 하지만 RGB 영상은 야간에 취득하거나 비가 오거나 안개가 심하는 등 기상 조건이 안 좋을 경우 고품질의 영상이 어렵다는 단점이 있다.

본 논문에서는 RGB 영상 뿐만 아니라 열 적외선 영상을 기반으로 한 객체 탐지 알고리즘 수행, 객체 탐지 네트워크 튜닝 및 그에 따른 성능 비교하고자 한다. 열 적외선 영상은 야간이나 기상 조건이 안 좋을 때 영상을 취득하여도 RGB 영상과 달리 객체의 모습이 비교적 뚜렷하게 나와 좋은 질의 영상을 취득할 수 있기 때문이다.

본 논문의 구성은 다음과 같다. 제 2 절에서는 영상 특성 비교를 서술하며 제 3 절에서는 영상 특성에 따른 객체 탐지 실험 결과에 대해 기술한다. 제 4 절에서는 Fine-tuning 한 객체 탐지 알고리즘 실험 및 실험 결과에 대해 서술하며 마지막으로 5 절에서 본 논문의 결론을 짓는다.

2. 영상 특성 비교

RGB 영상은 RGB 센서 카메라를 통해 얻어진 영상으로 흔히 알고 있는 카메라로 취득한 영상을 말한다. RGB 센서는 센서에 들어오는 빛을 기반으로 영상을 생성하기 때문에 빛이 적은 밤에는 고화질의 영상을 얻기 힘들며, 또한 안개가 심한 때나 비가 많이 오는 때에도 좋은 화질의 영상을 취득하기 어려울 수 있다.

열 적외선 영상은 열 센서로부터 받은 열 정보를 영상화한 것을 의미한다. 빛과 상관없이 열 정보를 이용하여 영상을 생성하기 때문에 기후가 좋지 않거나 빛이 적은 밤일 때에도 RGB 영상에 비해 고화질의 영상을 얻을 수 있다. 그림 1 은 FLIR 데이터[1]의 샘플로서, 밤에 취득한 RGB 영상과 열 적외선 영상을 보여주며 열 적외선 영상이 RGB 영상에 비해 조금 더 선명하게 객체를 나타내는 것을 확인할 수 있다.

A. RGB 영상 샘플

B. 열 적외선 영상 샘플

그림 1. FLIR 데이터 세트 야간 이미지 샘플

본 논문에서는 두 영상에 대한 객체 탐지 알고리즘 성능을 평가하기 위해 FLIR 데이터 세트[1]를 선택하였다. FLIR 데이터 세트는 FLIR 사에서 제공하는 이미지 데이터 세트로 RGB 영상과 annotation 이 있는 열 적외선 영상 모두 제공하기 때문에 RGB 영상과 열 적외선 영상의 객체 탐지 알고리즘 수행 성능 비교가 가능하다.

3. 객체 탐지 수행 및 성능 평가

RGB 영상과 열 적외선 영상에 대한 객체 탐지 알고리즘 성능을 비교하기 위해 Facebook AI Research 의 Faster R-CNN X101-FPN[2] 네트워크를 선택하였다. 압축을 하지 않은 원본 영상 데이터와 더불어 기계를 위한 압축된 비트스트림 전송 실험을 위해 VTM(VVC Test Model) 으로 인/디코딩을 수행하고 QP(Quantization Parameter) 및 해상도 스케일 요소에 따라 압축된 이미지에 대해서도 객체 탐지 알고리즘도 실험하였다. 본 실험에서 사용한 QP 는 17, 22, 27, 32, 37, 42, 47 이며 스케일 요소는 해상도 100%, 75%, 50%, 25%이다.

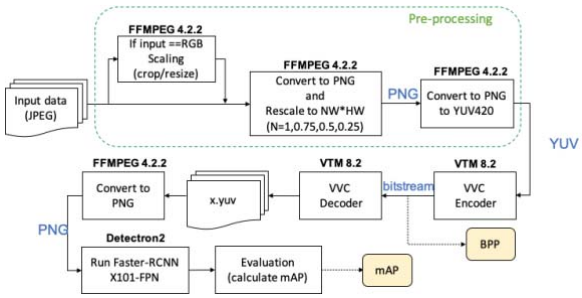


그림 2. Pre-processing, VTM 인/디코딩 및 객체 인식 수행 과정

객체 탐지 및 정확도 측정을 위해 FLIR 데이터 세트에서 야간에 촬영된 RGB 영상과 열 적외선 영상 각 300 장씩 추출하였다. 해당 데이터에서 제공하는 열 적외선 영상 (640 x 512)과 RGB 영상 (1800 x 1600)은 해상도가 상이하기 때문에 열 적외선에 맞게 RGB 영상의 해상도를 조절하였으며, 열 적외선 영상의 annotations 만 제공하므로 RGB 영상에 대한 annotations 을 제작하여 실험을 진행하였다. 사람, 자동차, 자전거 총 3 가지 레이블에 제한하여 COCO AP metric 을 계산하였고, 정확도 비교 결과 RGB 영상에 비해 열 적외선 영상이 mAP 기준 최대 약 9% 높은 성능을 보여주는 것을 확인하였다.

표 1. 열 적외선 영상에 대한 사전 훈련 네트워크 수행 결과

Resolution	QP	mAP	AP50	AP75	APs	APm	APl	BPP
-	original	29.726	56.819	27.147	13.767	41.325	66.916	6.861
100%	17	27.279	52.523	23.700	11.607	38.955	65.254	2.476
	22	29.379	56.420	26.961	13.348	40.703	68.281	1.807
	27	29.097	55.862	26.433	13.560	39.895	67.340	1.224
	32	27.842	52.786	25.331	13.550	38.494	61.553	0.306
	37	20.631	39.496	18.665	7.968	28.961	56.310	0.131
	42	10.134	19.142	8.534	2.433	15.266	33.322	0.065
	47	2.910	5.946	2.623	0.766	4.626	8.487	0.030
75%	17	26.803	50.344	24.710	11.343	37.228	67.781	1.376
	22	25.848	49.477	23.997	10.629	36.277	65.736	0.888
	27	25.589	48.771	23.731	10.709	35.309	65.799	0.400
	32	22.407	42.920	21.199	9.534	30.622	61.071	0.189
	37	15.725	30.582	14.592	5.091	23.357	46.597	0.098
	42	5.581	11.964	4.819	1.047	7.681	23.066	0.049
	47	1.550	2.905	1.239	0.231	2.584	5.176	0.022
50%	17	21.306	38.021	20.906	6.252	31.422	64.418	0.578
	22	21.179	37.641	20.084	6.156	30.488	65.533	0.352
	27	18.962	34.040	17.752	5.610	27.652	58.932	0.193
	32	15.799	28.684	16.198	4.119	22.250	54.384	0.107
	37	8.487	15.791	8.975	1.431	11.968	33.569	0.056
	42	2.244	5.192	1.716	0.342	3.434	8.515	0.027
	47	0.503	0.660	0.660	0.033	0.286	1.611	0.012
25%	17	9.300	16.327	8.928	0.726	12.924	47.183	0.149
	22	8.632	15.698	8.191	0.747	11.268	45.601	0.100
	27	7.474	13.721	6.433	0.485	10.404	37.545	0.064
	32	5.405	10.353	5.322	0.323	6.377	30.369	0.038
	37	2.163	4.415	1.102	0.160	2.870	10.167	0.020
	42	0.309	0.594	0.259	0.066	0.240	2.251	0.010
	47	0.000	0.000	0.000	0.000	0.000	0.000	0.005

A. 사전 훈련 네트워크 기반 객체 탐지 결과 (RGB)

B. 사전 훈련 네트워크 기반 객체 탐지 결과 (열 적외선)

그림 3. 객체 탐지 결과 (QP: 22, Scale Factor : 100%)

표 2. RGB 영상에 대한 사전 네트워크 수행 결과

Resolution	QP	mAP	AP50	AP75	APs	APm	APl	BPP
-	original	22.236	41.968	20.068	7.293	32.984	61.818	15.403
100%	17	21.768	42.431	19.105	7.086	32.012	63.684	1.556
	22	22.160	43.726	19.582	7.298	33.406	60.975	0.576
	27	20.705	41.126	17.702	5.997	31.981	57.737	0.236
	32	17.536	34.515	13.875	4.872	25.889	54.186	0.116
	37	11.552	22.220	10.237	2.802	17.023	36.155	0.064
	42	3.551	8.073	2.715	0.799	5.139	15.482	0.036
	47	0.847	1.740	0.675	0.187	0.628	6.455	0.019
75%	17	22.708	42.815	20.373	6.905	33.586	65.245	0.881
	22	22.046	41.932	19.423	6.877	32.857	62.175	0.448
	27	18.719	36.426	15.716	5.744	28.773	51.095	0.194
	32	15.403	31.546	11.888	4.557	22.220	51.172	0.092
	37	8.294	17.360	6.958	2.286	11.729	27.750	0.050
	42	2.353	5.435	1.699	0.592	3.243	10.192	0.028
	47	0.408	0.692	0.519	0.243	0.369	2.374	0.015
50%	17	20.159	38.470	17.780	5.671	29.838	63.045	0.373
	22	17.823	34.587	16.257	5.133	26.080	55.828	0.189
	27	14.658	28.621	12.008	3.928	21.501	46.816	0.097
	32	9.429	20.043	7.767	2.320	14.019	31.328	0.055
	37	3.607	8.174	2.363	0.889	5.251	13.623	0.031
	42	0.912	1.935	0.673	0.125	1.034	5.644	0.017
	47	0.186	0.330	0.165	0.000	0.132	0.833	0.009
25%	17	10.162	20.302	7.382	1.944	14.393	43.026	0.101
	22	8.394	16.374	6.708	0.688	12.012	36.783	0.061
	27	4.754	10.454	3.134	0.398	6.557	20.199	0.038
	32	2.002	4.216	1.574	0.231	1.937	10.775	0.023
	37	0.614	1.141	0.550	0.350	0.217	4.002	0.013
	42	0.231	0.330	0.330	0.000	0.000	0.783	0.008
	47	0.040	0.132	0.000	0.000	0.000	0.092	0.004

4. 객체 탐지 네트워크 fine-tuning 및 성능 평가

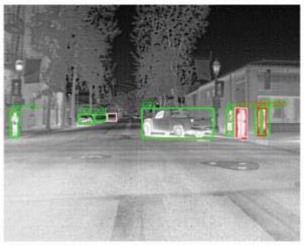
Faster R-CNN X101-FPN 은 COCO 데이터 세트를 기반으로 훈련된 네트워크다. COCO 데이터 세트는 실내 또는 낮에 촬영한 RGB 영상으로 구성되어 있어 FLIR 데이터 세트와는 차이가 있으며, 이에 따라 객체 탐지 정확도가 떨어질 가능성이 있다. 따라서 FLIR 데이터 세트에서 야간 RGB 영상 1,000 장과 열 적외선 영상 1,000 장을 선정하여 fine-tuning 하였으며, fine-tuned 네트워크를 이용하여 객체 탐지의 정확도를 측정하였다. 실험은 이전과 동일하게 원본과 VTM 으로 인/디코딩을 수행한 600 장의 영상들에 대해 객체 탐지를 수행하였으며, 3 가지 레이블에 제한하여 객체 탐지 정확도를 측정하였다. 실험 결과, 사전 훈련 네트워크에 비해 fine-tuned 네트워크로 객체 인식을 수행하는 경우 열 적외선 영상의 탐지 정확도는 mAP 기준 최대 15%, RGB 영상의 탐지 정확도는 2% 정도 탐지 능력이 향상되었다. 게다가, 열 적외선 영상에 대해 RGB 영상보다 mAP 기준 약 15%가 높은 객체 탐지 정확도를 보이며 여전히 열 적외선 영상이 RGB 영상에 비해 높은 탐지 정확도를 보이는 것을 확인할 수 있었다.

표 3. 열 적외선 영상에 대한 Fine-tuned 네트워크 수행 결과

Resolution	QP	mAP	AP50	AP75	APs	APm	APl	BPP
	original	40.557	77.969	36.978	31.156	47.050	64.680	6.861
100%	17	39.279	75.542	35.822	29.911	45.560	66.563	2.476
	22	39.446	76.642	35.151	30.514	45.592	61.689	1.807
	27	39.643	76.692	35.657	30.874	46.037	62.608	1.224
	32	37.690	73.089	31.470	28.526	45.175	58.572	0.306
	37	33.784	69.240	27.716	24.625	41.792	52.623	0.131
	42	23.069	47.960	18.865	15.676	29.358	40.428	0.065
	47	10.634	24.140	7.648	6.130	13.976	24.376	0.030
75%	17	40.229	77.593	35.847	29.634	48.125	65.888	1.376
	22	40.431	77.304	35.431	29.731	48.229	64.498	0.888
	27	39.160	74.827	33.401	28.957	47.273	59.904	0.400
	32	36.423	72.423	30.391	26.601	44.209	57.550	0.189
	37	30.992	62.335	25.447	22.157	38.728	46.858	0.098
	42	19.437	41.373	14.263	12.315	25.536	33.216	0.049
	47	7.456	16.758	5.648	3.465	10.294	20.800	0.022
50%	17	35.955	68.673	31.193	24.267	44.742	63.161	0.578
	22	35.711	69.105	31.565	24.134	44.969	62.860	0.352
	27	34.121	65.454	30.308	23.202	43.282	55.227	0.193
	32	30.406	59.647	25.461	20.049	39.954	49.065	0.107
	37	21.838	44.769	18.297	13.317	29.886	37.600	0.056
	42	9.860	22.821	7.220	5.025	13.902	23.275	0.027
	47	2.501	6.178	1.672	0.883	3.758	7.324	0.012
25%	17	19.172	38.465	16.643	7.689	28.669	47.023	0.149
	22	18.405	36.985	15.587	7.752	27.551	43.892	0.100
	27	16.175	33.780	13.305	6.431	24.377	40.549	0.064
	32	11.847	25.995	9.113	4.249	17.686	33.756	0.038
	37	6.142	14.007	4.104	2.156	8.952	22.072	0.020
	42	1.371	3.433	0.983	0.403	1.589	7.125	0.010
	47	0.126	0.334	0.083	0.154	0.144	0.244	0.005

A. 사전 훈련 네트워크 기반 객체 탐지 결과 (RGB)

B. Fine-tuned 네트워크 기반 객체 탐지 결과(RGB)



C. 사전 훈련 네트워크 기반 객체 탐지 결과(열 적외선)

D. Fine-tuned 네트워크 기반 객체 탐지 결과(열 적외선)

그림 4. 사전 훈련 네트워크 및 Fine-tuned 네트워크 객체

탐지 결과 (QP: 22, Scale Factor : 100%)

표 4. RGB 영상에 대한 Fine-tuned 네트워크 수행 결과

Resolution	QP	mAP	AP50	AP75	APs	APm	APl	BPP
	original	23.752	50.275	18.982	9.818	37.050	64.969	15.403
100%	17	23.678	49.267	19.422	9.494	36.696	63.023	1.556
	22	23.998	50.818	20.545	9.573	37.427	67.511	0.576
	27	21.695	46.393	17.381	8.205	34.091	63.118	0.236
	32	17.535	37.936	14.126	5.882	27.648	53.864	0.116
	37	12.380	26.113	9.588	3.241	18.586	45.408	0.064
	42	6.171	13.817	4.304	1.045	8.373	34.405	0.036
	47	2.068	4.612	1.410	0.346	2.171	15.478	0.019
75%	17	24.004	51.281	19.727	9.543	37.680	63.611	0.881
	22	22.941	48.576	18.918	9.003	35.997	65.580	0.448
	27	20.758	42.844	17.899	7.335	32.820	63.111	0.194
	32	16.957	36.547	12.878	5.487	25.917	50.419	0.092
	37	11.173	24.690	8.216	2.703	16.619	39.169	0.050
	42	4.531	11.547	2.882	0.754	6.791	23.477	0.028
	47	1.416	3.200	1.097	0.340	1.222	9.598	0.015
50%	17	22.131	45.383	17.836	8.156	35.163	65.426	0.373
	22	20.895	43.475	16.716	7.267	32.145	62.527	0.189
	27	18.156	38.132	14.761	5.497	29.239	54.215	0.097
	32	13.161	29.120	8.887	3.403	19.525	44.815	0.055
	37	6.218	14.059	4.131	1.028	8.745	30.239	0.031
	42	2.194	5.448	1.643	0.349	2.964	13.009	0.017
	47	0.702	1.590	0.435	0.041	0.520	5.163	0.009
25%	17	13.317	28.412	10.700	2.640	20.471	54.820	0.101
	22	11.577	24.603	9.864	1.873	17.976	49.067	0.061
	27	8.240	17.843	7.325	1.241	12.091	35.586	0.038
	32	4.688	10.841	3.233	0.626	5.960	26.491	0.023
	37	1.460	3.197	1.244	0.148	1.324	13.403	0.013
	42	0.494	0.960	0.384	0.014	0.286	3.678	0.008
	47	0.113	0.248	0.165	0.000	0.127	0.547	0.004

4. 결론

본 논문에서는 열 적외선 영상과 RGB 영상의 특성에 대해 비교하고, 각 영상에 대해 객체 탐지 알고리즘인 Faster R-CNN X101-FPN 네트워크를 수행하고 객체 탐지 정확도를 비교하였다. RGB 영상으로 사전 훈련된 네트워크임에도 불구하고 열 적외선 영상에 대한 객체 탐지 정확도가 mAP 기준으로 약 9% 높은 것을 보였다. 추가적으로 열 적외선 영상과 RGB 영상 각 1,000 장씩을 선정하여 네트워크를 fine-tuning 하였으며, fine-tuned 네트워크로 객체 탐지를 실험한 결과 두 영역의 영상에 대해 모두 정확도가 상승한 것과 더하여 여전히 열 적외선 영상 기반 객체 탐지 정확도가 높은 것을 확인할 수 있었다.

본 연구 논문은 과학기술정보통신부 및 정보통신기획평가원의 출연금으로 수행되고 있는 한국전자통신연구원 “기계를 위한 영상 부호화 기술 개발”(2020-0-00011)의 위탁연구과제의 연구결과 입니다.

참고문헌

[1] “FREE FLIR Thermal Dataset for Algorithm Training” https://www.flir.com/oem/adas/adas-dataset-form/

[2] “Detectron2”,

https://github.com/facebookresearch/detectron2