

Dual Shot Face Detector Approach 성능 분석

이경민, 인치호, 김제우, 이영한, 조충상, 송혁
전자부품연구원, 세명대학교

leegis0904@keti.re.kr, ich410@semyung.ac.kr, jwkim@keti.re.kr,
yhlee@keti.re.kr, ideafisher@keti.re.kr, hsong@keti.re.kr

A Performance Analysis of Dual Shot Face Detector Method

KyungMin Lee, ChiHo Lin, JeWoo Kim, YoungHan Lee, ChoongSang Cho, Hyok Song
Korea Electronics Technology Institute, Semyung University

요 약

본 논문에서는 얼굴 검출에서 좋은 성능을 보이는 Dual Shot Face Detector (DSFD)을 WIDER FACE 데이터 기반으로 네트워크의 성능 및 특성을 분석한다. 얼굴 검출 엔진의 동작 및 학습을 확인하기 위해 WIDER FACE 데이터를 기반으로 DSFD 를 학습한 후 기존 논문에서 발표되었던 성능을 보이는지 확인한다. 성능 확인된 얼굴 검출기 DSFD 를 세부적으로 분석하기 위하여 테스트 데이터를 10 개 그룹으로 분할하여 AP 측정 후 성능 열화가 높은 그룹의 이미지를 집중적으로 분석하였다. 분석 결과 얼굴 영역 주변에 모자 헬멧, 방패 등에 의해 가려진 경우 얼굴 검출에 많은 어려움이 발생했다. 또한 이미지에 있는 상황을 61 개로 구분하여 정량적 분석을 수행하며 분석 결과 Car Accident 와 People Marching 상황에서 성능이 열화되는 것을 확인하였다. 본 논문에서는 DSFD 네트워크의 성능 열화 부분 및 경향 등을 분석하였다. 이러한 내용은 DSFD 얼굴 검출기의 성능을 개선하거나 특정 도메인에 맞는 검출기를 개발하는데 매우 중요한 자료가 될 것으로 생각한다.

1. 서론

객체 검출 관련한 연구를 지속적으로 관심을 받고 있으며, 사람의 영역에 특화되어 Faster R-CNN[2], SSD[3] 등 딥러닝 네트워크를 활용한 얼굴 검출 관련 연구가 많이 진행되고 있다[1]. 최근엔 WIDER FACE[4] 셋을 사용한 얼굴 검출 평가에서 높은 성능을 보이는 Dual Shot Face Detector (DSFD) [5] 네트워크가 발표되었다. 이러한 연구들을 다양한 응용분야에 활용하기 위해서는 알고리즘의 특징 및 성능을 정확히 분석하는 것이 매우 중요하다. 본 논문에서는 얼굴 검출에서 좋은 성능을 보이는 DSFD를 데이터 셋으로 학습하고 평가하여 네트워크의 특징과 성능을 분석하고자 한다. 이러한 분석은 얼굴 검출기 DSFD 네트워크의 성능을 개선하거나 다양한 응용 도메인에 적용하기 위한 매우 중요한 자료가 될 것으로 생각한다. 본 논문은 서론, DSFD 알고리즘에 대한 설명, 얼굴 검출 네트워크를 학습하고 평가한 결과를 분석하면서 논문을 마친다.

2. DSFD(Dual Shot Face Detector)

얼굴 검출 연구 중 WIDER FACE 데이터 셋 조건에서 높은 성능을 보이는 DSFD(J. Li, Y. Wang et al CVPR 2019) 네트워크가 발표되었다. DSFD 는 두 개의 다른 Anchor 셋을

사용하여 듀얼 샷 검출 방법으로 하나의 이미지 속에 객체의 크기 차이가 큰 경우에도 높은 성능을 갖는 얼굴 검출기이다. DSFD 네트워크는 단일 샷에서 듀얼 샷을 확장하기 위해 Feature Pyramid Network(FPN) [6]와 Receive Field Block(RFB) [7]의 장점을 결합하여 Feature Map의 분별력과 견고성을 향상시키는 Feature Enhance Module(FEM)을 제안하였으며, 듀얼 샷의 검출 성능을 더욱 효율적으로 검출하기 위해 두 개의 다른 Anchor 셋을 이용한 Progressive Anchor Loss(PAL)을 채택하였다. 또한, 새로운 Anchor 할당의 전략은 데이터 확대를 통합하여 Regressor의 초기화를 개선함으로써 얼굴 검출 성능을 향상을 위해 Improved Anchor Matching(IAM)을 제안하였다. 얼굴 검출기 DSFD의 framework는 그림 1과 같으며, 원본 특징 샷 (a)에서 향상된 (c)을 만들기 위해 VGG/ResNet 기본 특징 추출 구조에 FEM (b)을 사용하여 향상된 PAL을 얻는 구조를 나타낸다.

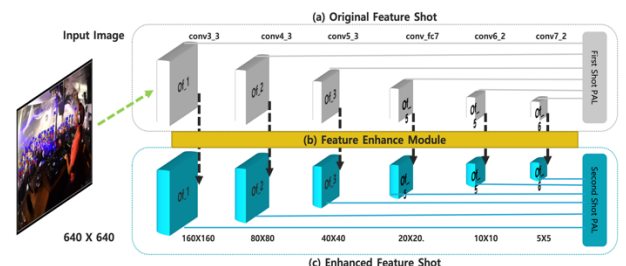


그림 1 DSFD 네트워크 Framework

3. DSFD 성능 및 특성 분석

본 논문에서는 얼굴 검출 네트워크 DSFD 을 학습하고 얻어진 모델 파라미터를 기반으로 도출된 결과 및 성능을 분석하는 과정을 수행한다. 얼굴 검출 네트워크를 학습하기 위하여 WIDER FACE 셋을 사용했으며, 데이터 셋은 총 32,203 개의 이미지에 397,703 개의 얼굴 경계 박스와 61 개의 상황 별로 이미지로 구성되며, 검출의 난이도를 고려한 Easy, Medium, Hard Label 정보를 포함하고 있다.

위와 같이 구성된 데이터 셋을 Train, Validation, Test 데이터 셋으로 나눠 제공되며, 예측 사각 박스와 Ground-truth(GT) 사각 박스 간의 Intersection over Union(IoU) 의 임계 값 0.5 기반으로 해당 예측 사각 박스의 신뢰도 점수를 평가하는 방식을 채택하여 성능을 평가한다[8]. 실험을 위해서 DSFD 네트워크를 WIDER FACE Train 셋을 통해 학습했으며, WIDER FACE Test 셋에는 GT 사각 박스 정보가 없기 때문에 정량적 평가를 위해 WIDER FACE Validation 셋의 10%에 해당하는 322 장을 이미지를 랜덤으로 선택하여 테스트 입력으로 사용하고, 나머지 이미지를 학습의 Validation 으로 사용하였다.

기존 DSFD 네트워크를 제안한 논문[5]의 성능을 다시 확인하기 위해서 VGG16 구조의 CNN 을 Feature Extractor 로 사용하였으며 논문에서 제시되었던 FEM/PAL/IAM 을 순차적으로 적용하면서 성능 측정하였다. FEM 을 적용한 실험 결과 표 1 과 같이 Easy, Medium, Hard set 에 대해서 93.3%, 90.6%, 79.4%을 보였으며 이는 논문[5]에서 발표된 결과와 유사하다. 또한, 기존 논문[5]에서는 VGG16 을 사용한 경우 PAL 과 IAM 모듈의 영향을 분석한 결과는 없지만, 본 논문에서는 관련 실험을 수행했으며 표 2 와 같이 94.9%, 93.7%, 88.0%로 PAL 과 IAM 적용에 의한 성능 향상을 VGG16 조건에서도 확인하였다.

표 1. DSFD 네트워크 학습 후 성능 분석 결과 (AP 측정)

비교 항목	Component	Easy	Medium	Hard
DSFD 논문[5]	FSSD+VGG16+FEM	93.0%	90.2%	79.1%
DSFD 테스트	FSSD+VGG16+FEM	93.3%	90.6%	79.4%

표 2. VGG16 Feature 추출 모듈에 PAL 과 IAM 을 적용한 실험 결과

Component	Easy	Medium	Hard
FSSD+VGG16+FEM+IAM+PAL	94.9%	93.7%	88.0%

테스트 데이터의 Easy 기준으로 94.9%를 보이는 높은 성능의 DSFD 네트워크가 어떠한 경우에 성능 열화를 보이는지 분석하기 위하여 테스트 데이터를 10 개 그룹으로 분할하고 정략적 성능을 측정하였다. 측정 결과 표 3 와 같이 10 개 그룹별로 성능을 측정한 결과 1,2 번 그룹에서 평균 성능보다 매우 낮은 성능을 보였다. 이를 바탕으로 우리는 1,2 번 그룹에 해당되는 이미지 중 얼굴 검출 성능이 매우 낮은 샘플에 대해서 분석을 진행하였다.

1, 2 번 그룹에서 성능을 열화 시키는 이미지는 그림 2 에서 보여진다. 그림 2 의 첫번째 Column 는 GT, 두번째 Column 은 검출기에 의한 결과, 마지막 Column 은 검출기에 의한 오류 부분을 표기한 그림이다.

표 3. 테스트 데이터를 분할하여 측정한 성능 결과

Data Set	1 (32장)	2 (32장)	3 (32장)	4 (32장)	5 (32장)	6 (32장)	7 (32장)	8 (32장)	9 (32장)	10 (34장)
AP	0.865	0.905	0.957	0.975	0.957	0.999	0.98	0.996	0.979	0.934
GT face	1372	1109	477	253	233	117	199	116	120	192
PR face	5606	4058	1720	1983	2167	941	1169	1115	835	1073
Detection face	1059	981	419	241	196	100	186	110	120	185

그림 2 에서 파란색 사각 박스는 Ground-truth(GT), 검은색 사각 박스는 GT 에는 없으나 검출기에서 출력된 결과, 붉은색 사각 박스는 IoU 0.5 이상의 결과 박스, 노란색 사각 박스는 IoU 값이 0.3 이상 0.5 미만의 결과, 주황색 박스는 GT 에서 검출하지 못한 박스를 나타낸다.

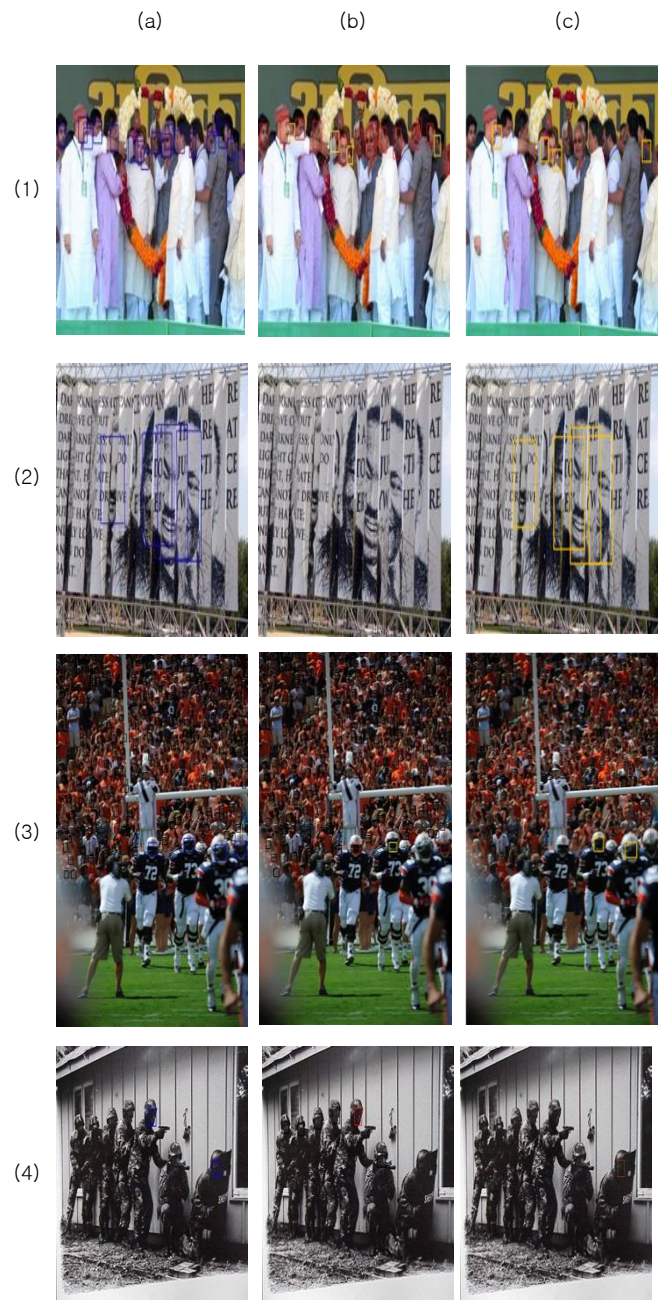


그림 2 성능을 열화 시키는 이미지 예

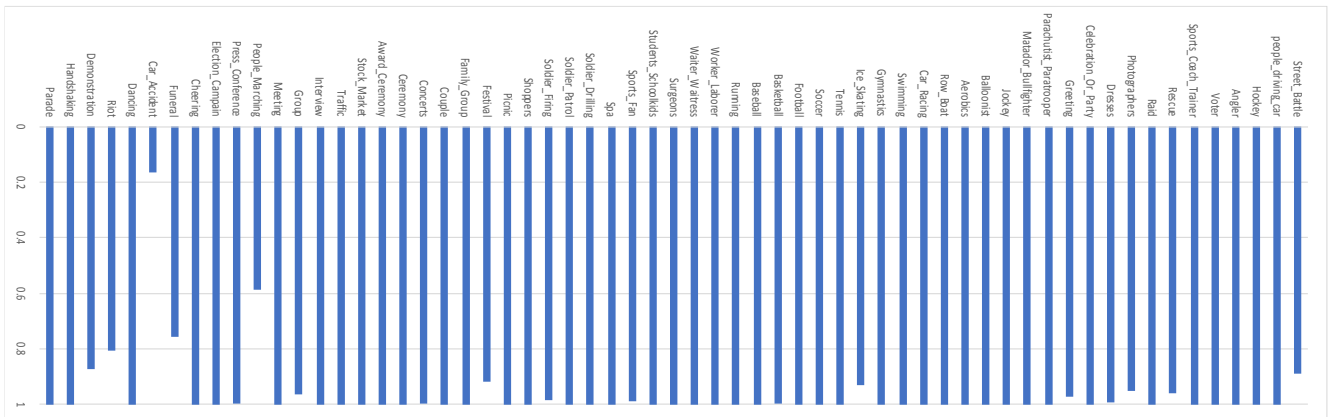


그림 3 WIDER FACE dataset 의 이벤트 61 개의 AP 수치 수정

그림 2-(1)의 경우, 사람의 어깨, 팔, 얼굴 등 일부 가려진 사람 얼굴의 GT 사각 박스는 얼굴이 아닌 영역의 비율이 전체 GT 사각 박스에 30% 이상 차지한다. 이러한 사각 박스는 학습 데이터 셋 내에 적은 수의 박스 데이터이며, 이를 바탕으로 학습된 가중치 모델의 예측으로는 얼굴 영역이 아닌 부분까지 포함하는 박스를 추출할 수 없어 예측 및 GT 사각 박스 간의 IoU 값이 낮게 계산되는 것을 보였다.

그림 2-(2)는 다른 객체에 의해 가려진 객체 검출에 대한 문제점과 조금 다르게 GT 사각 박스는 한 사람의 얼굴을 쪼개서 현수막에 그린 후 현수막을 설치하여 한 사람의 얼굴처럼 보이는 상황의 이미지이다. 이런 이미지의 GT 사각 박스 영역을 정의할 때 한 사람의 얼굴로 봐야 할지 여러 사람 얼굴로 봐야 할지에 기준이 애매하여 검출에 어려움을 보였다.

그림 2-(3)와 (4)는 얼굴 영역의 낮은 명암으로 발생하는 검출의 어려움으로 예를 나타낸다. 그림 2-(3)은 모자, 헬멧이 얼굴 영역을 일부 가려졌기 때문에 얼굴에 전체적으로 명암이 낮아져 얼굴의 특징을 찾을 수 없는 문제점이다. 그림 2-(4)는 흑백 이미지로 헬멧으로 발생하는 낮은 명암이 얼굴을 검출에 어려움을 보였다. 또한 그림 2-(4)는 방패와 헬멧으로 얼굴이 완전히 가려진 영역을 얼굴 영역으로 GT 사각 박스로 정의한 애매한 기준으로 검출이 어렵다.

테스트 이미지를 61 개 상황으로 구분하여 평가하였을 경우 그림 3 과 같은 결과를 보인다. 그림 3 에서 세로 축은 AP 값을 의미하고 가로축은 61 개의 상황 종류를 표현한다. 이를 바탕으로 61 개의 상황 종류에 대해서 정량적 분석을 진행하였다. 분석 결과 Car Accident 상황에서 매우 낮은 AP 결과를 보였으며, People Marching, Funeral, Riot 상황 순으로 낮은 AP 값을 보였다.

4. 결론

본 논문에서는 얼굴 검출에서 좋은 성능을 보이는 얼굴 검출기 DSFD 네트워크의 성능을 개선하거나 다양한 응용 도메인에 적용하기 위해 DSFD 을 WIDER FACE 셋으로 학습하고 평가하여 네트워크의 성능 및 특성을 분석하였다. 분석 결과 DSFD 네트워크는 데이터 셋의 GT 사각 박스가 가진 얼굴이 아닌 영역이 30% 이상인 경우, 또는 모자 헬멧, 방패 등에 의해 가려진 경우 그리고 낮은 명암의 얼굴에서

검출에 많은 어려움이 발생되었다. 이러한 분석 자료는 DSFD 네트워크를 개선하는데 많은 도움을 줄 수 있을 것이 기대하며, 향후 다양한 응용 도메인에 적용 가능한 개선된 DSFD 네트워크 연구에 매우 중용한 자료가 될 것으로 생각한다.

Acknowledgement

본 논문은 2018 년도 과학기술정보통신부 정보보호핵심원천기술개발사업의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임.(2018-0-00401, 영상 빅데이터 학습을 위한 De-identification 기술 개발)

References

- [1] Zhaowei Cai, Quanfu Fan, Rogerio S Feris, and Nuno Vasconcelos. A unified multi-scale deep convolutional neural network for fast object detection. In Proceedings of European Conference on Computer Vision (ECCV), 2016.
- [2] Shaoqing Ren et al, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", Computer Vision and Pattern Recognition: arXiv:1506.01497.
- [3] Wei Liu et al, "SSD: Single Shot MultiBox Detector", European Conference on Computer Vision (ECCV), 2016.
- [4] Shuo Yang Ping, Luo Chen Change, and Loy Xiaoou Tang, "WIDER FACE: A Face Detection Benchmark", Computer Vision and Pattern Recognition: arXiv:1511.06523, Nov 2015.
- [5] Jian Li et al, "DSFD: Dual Shot Face Detector", IEEE Conference on Computer Vision and Pattern Recognition, CVPR, Jun,2019.
- [6] Tsung-Yi Lin et al, "Feature Pyramid Networks for Object Detection", Computer Vision and Pattern Recognition: arXiv:1612.03144.
- [7] Songtao Liu, Di Huang and Yunhong Wang, "Receptive Field Block Net for Accurate and Fast Object Detection", Computer Vision and Pattern Recognition: arXiv:1711.07767.

[8] Everingham, Mark, et al. "The pascal visual object classes (voc) challenge." *International journal of computer vision* 88.2 (2010): 303–338.