

L1 목적 함수와 채널 프루닝을 이용한 얼굴 검출기 경량화

*이석희 장영균 조남익

서울대학교 뉴미디어통신공동연구소

*seokheel@ispl.snu.ac.kr

Compression and Acceleration of Face Detector using L1 Loss and Channel Pruning

*Lee, Seok Hee Jang, Young Kyun Cho, Nam Ik

Seoul National University Institute of New Media and Communications

요약

본 논문에서는 합성곱 기반의 얼굴 검출기 Dual Shot Face Detector (DSFD)에 대하여, 특징점 맵의 희소화와 채널 프루닝 목적 함수를 사용하여 네트워크 경량화를 수행하였다. 특징점 맵을 희소화하기 위해 L1 목적 함수를 사용했고, 특징점 맵의 채널 프루닝을 하기 위해 채널 최대값이 가장 낮은 채널들의 합을 최소화 시키는 목적함수를 적용했다. 기존의 신경망은 특징점 맵 희소화 비율이 45%였고 두 목적 함수를 적용했을 때 69.67%로 희소화 비율이 높아진 것을 확인했다. 얼굴 검출 성능을 다양한 조명, 크기, 환경, 각도, 표정의 얼굴들을 포함하는 영상들로 이뤄진 Wider Face 데이터 셋으로 실험한 결과, average precision은 하락 했고 easy validation set에서 0.9257, hard validation set에서 0.8363 였다.

1. 서론

심층 신경망은 최근 몇 년간 여러 분야에서 좋은 성능을 보이고 있는 방법 중 하나이지만 좋은 성능을 보이는 만큼 높은 계산량, 메모리량과 전력량을 필요로 한다. 예를 들어, AlexNet[3]은 7억 2천만 개의 multiply-accumulate (MAC) operations이 필요하고 6천만 개의 학습 가능한 파라미터를 갖고 있다. 이런 높은 계산량은 데스크탑이나 서버에서 그래픽 처리 장치(GPU)의 높은 대역폭 메모리 및 병렬 계산으로 처리 가능하다. 하지만 낮은 메모리, GPU의 부재, 한정적인 전력 소비 등과 같은 제한된 리소스를 가진 휴대용 장치 또는 임베디드 시스템에서 딥 러닝 알고리즘을 수행할 때 높은 성능을 기대하기 힘들다. 따라서 딥 러닝 모델을 그대로 사용하기보다는 실제 응용 프로그램에 원활하게 수행 할 수 있도록 신경망을 압축하고 가속화하는 것이 매우 중요하다.

본 논문에서는 얼굴 검출기 Dual Shot Face Detector (DSFD)[1]의 심층 신경망을 압축하고 가속화하기 위해 L1 목적 함수를 적용했다. 특징점 맵을 희소화하기 위해 특징점 맵 전체에 L1 목적 함수와 채널을 프루닝하기 위해 채널의 최대값에 L1 목적 함수를 각각 적용하였다. 기존 얼굴 검출기와 경량화된 얼굴 검출기의 희소화 비율을 비교하고 얼굴 검출 성능을 WIDER FACE[2] 데이터셋으로 비교하였다.

2. L1 목적 함수와 채널 프루닝

네트워크를 압축하고 가속화하는 방법에 프루닝이 있다. 프루닝의 핵심 아이디어는 신경망의 학습된 파라미터의 채널, 필터 등에 중복성이 존재하는데 모델 성능에 크게 기여 하지 않는 중복 채널 또는 파라미터

를 제거하는 것이다. 기존의 간단한 신경망을 입력 행렬 X 에 가중치 행렬 W 를 곱했을 때의 출력 행렬 Y 로 표현할 수 있는데, 프루닝은 일부의 연결을 제거하여 입력 행렬 X' 나 가중치 행렬 W' 을 만들고, 이들로부터 얻어지는 출력 행렬을 Y' 이라 할 때 이를 가능한 한 Y 에 가깝게 만드는 새로운 가중치 행렬 W' 을 학습하는 것이다. 이 때 제거 할 채널, 파라미터나 연결을 선택하는 방법은 알고리즘에 따라 다르다.

잘 알려진 프루닝 방법 중 하나는 채널 프루닝[4]이다. 이 방법은 특징점 맵에서 성능 향상에 크게 영향을 미치지 않는 중복된 채널을 제거하는 것이 목표이다. 중복된 채널을 제거하면 해당 채널에 곱해지던 가중치 파라미터가 더 이상 필요하지 않게 된다. 따라서, 특징점 맵을 구하기 위해 곱해졌던 가중치 파라미터와 다음 층에서의 특징점 맵을 계산하기 위해 다음 단계의 가중치 파라미터는 공통적으로 일부의 파라미터가 필요 없어지게 되고 신경망에서 제거 될 수 있다.

본 논문에서 네트워크 경량화를 위해 두 가지 목적 함수를 사용하였다. L1 함수는 희소화에 직접적으로 영향이 있는 L0 함수의 tightest relaxation으로 알려져있다 [5]. 본 논문에서도 L0 목적 함수가 아닌 gradient를 계산할 수 있는 L1 목적 함수를 대신 사용하였다. <수식 1>에서 전체 layer가 L 개 있고 특징점 맵의 크기가 channel은 C 개, height는 H , width는 W 라고 하고 특징점 맵의 픽셀을 $x_{l,h,w,c}$ 라고 했을 때, 첫 번째 목적 함수는 특징점 맵의 희소화를 위해 L1 목적 함수를 특징점 맵의 각각의 픽셀들에 직접적으로 적용하였다. 두 번째 목적 함수는 채널 프루닝을 위해 설계된 L1 목적 함수를 사용하였는데, 각 레이어에서 채널마다 최대값을 뽑은 후 L1 목적 함수를 적용하였고 모든 채널 C 가 아닌 값이 가장 낮은 하위 20%의 채널들 C' 에 대해서만 계산했다. 즉, 반응도가 낮은 채널은 결과에 영향을 적게 미치고 반응도가 높은

채널은 결과에 영향을 높게 미친다고 가정을 하고 반응도가 낮은 채널에 대해서만 희소화를 유도하도록 목적 함수를 적용했다. 마지막으로 α 라는 두 번째 목적 함수에 목적 함수간의 가중치를 주는 하이퍼 파라미터를 사용했다.

$$\sum_l \sum_h \sum_w \sum_c |x_{l,h,w,c}| + \alpha \sum_l \sum_{c \in C'} |\max_{h,w} (x_{l,h,w,c})| \quad (1)$$

3. 실험

본 실험에서는 제안한 방법을 이용하여 Dual Shot Face Detector (DSFD)[1]라는 얼굴 검출기에 대한 경량화를 진행했다. 기존의 검출기와는 달리 DSFD[1]는 얼굴을 두 개의 검출 단계로 나눠서 검출을 하므로 성능이 뛰어나다. 첫 번째 검출 단계에서 검출기는 크기가 큰 얼굴이나 배경과 쉽게 구별 되는 쉬운 경우에 대하여 얼굴을 찾는다. 두 번째 검출 단계에서는 첫 번째 단계에서 생성한 특징점 맵을 재사용하는데 다양한 스케일의 특징점 맵을 결합하여 유의미한 특징점 맵을 재 생성한다. 그리고 생성된 특징점 맵을 활용하여 첫 번째 검출 단계에서 놓쳤던 작은 얼굴이나 배경과 구별하기 어려운 얼굴을 검출한다.

DSFD[1]를 학습 할 때 WIDER FACE[2] 데이터셋이라는 공개적으로 개방 된 데이터셋을 사용했다. WIDER FACE[2]는 393,703 개의 얼굴을 가진 32,203 개의 영상이 있고 스케일, 포즈, 겹침, 흐릿함, 표현, 화장, 조명, 포즈 등이 매우 다양하게 구성되어있다. 이렇게 다양한 환경의 영상으로 학습하면 검출기가 강인해서 다양한 환경의 얼굴을 찾을 수 있다.

본 논문의 경량화 신경망 학습 과정은 2단계 파이프 라인으로 진행했다. 먼저 일반적인 신경망처럼 기존의 신경망을 WIDER FACE 학습 데이터셋으로 학습해 파라미터간의 연결성을 찾았다. 두 번째 단계에서 경량화 목적 함수를 추가로 적용하여 신경망의 성능에 크게 영향을 미치지 않는 일부 연결을 제거하며 기존의 신경망에 사용했던 목적 함수를 그대로 사용하여 검출 성능을 유지시켰다. 하이퍼 파라미터에 대해서 수식 1의 α 는 0.1로 설정하고 C' 은 최대값이 가장 낮은 하위 20%의 채널들로 설정하고 학습했다.

4. 실험 결과

WIDER FACE[2]의 validation set에 대해 <표 1, 2>와 같이 희소화 비율과 average precision(AP)를 계산해 성능을 평가했다. 특징점 맵 희소화 비율을 비교했을 때 두 목적 함수를 적용한 경우가 69.67%로 가장 높은 비율을 보였다. 채널 프루닝만 적용했을 때 채널 희소화 비율이 20.00 %로 가장 높았지만 특징점 맵 희소화 비율은 가장 떨어지는 것을 확인했다. 그리고, 표 2에서 검출 성능을 나타내는 지표인 AP가 기존 신경망의 성능에 가장 가까웠고 작은 성능 하락을 보였다. 두 목적 함수를 모두 적용했을 때 특징점 맵 희소화 비율이 가장 높은 만큼 검출 성능도 가장 떨어졌다.

표 1 WIDER FACE 데이터셋으로 실험했을 때 신경망의 희소화 비율 비교

	채널 희소화 비율	특징점 맵 희소화 비율
Original	4.86%	45%
Both	12.95%	69.67%
L1 on feature map	8.32%	68.74%
L1 on channel pruning	20.00%	50.67%

표 2 WIDER FACE validation 데이터셋에서 average precision

	easy AP	medium AP	hard AP
Original	0.9431	0.9320	0.8569
Both	0.9257	0.9135	0.8363
L1 on feature map	0.9309	0.9191	0.8455
L1 on channel pruning	0.9402	0.9263	0.8521

5. 결론

본 논문은 합성곱을 기반으로 한 얼굴 검출기 Dual Shot Face Detector (DSFD)[1]를 특징점 맵과 채널을 희소화 하는 방법으로 경량화 시켰다. 특징점 맵을 희소화 하기 위해 학습 시 L1 목적 함수를 적용했고 특징점 맵의 채널 프루닝을 하기 위해 채널 최대값이 가장 낮은 채널들에 L1 목적함수를 적용했다. 제안한 방법의 성능을 평가하기 위하여 기존 신경망과 경량화한 신경망에 대해 특징점 맵의 희소화 비율, 채널 희소화 비율과 검출 성능을 비교했다. 두 목적 함수를 적용했을 때 특징점 맵의 희소화 비율이 가장 높았지만, 경량화 신경망의 특징점 맵의 희소화 비율이 높아질수록 검출 성능이 떨어지는 것을 확인했다.

감사의 글

이 논문은 2020년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 1711075689, AI 어플리케이션을 지원하는 IoT 연동 분산 Edge 클라우드 기술 개발)

참고문헌

[1] Li, Jian, Yabiao Wang, Changan Wang, Ying Tai, Jianjun Qian, Jian Yang, Chengjie Wang, Jilin Li, and Feiyue Huang. "DSFD: dual shot face detector." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5060-5069. 2019.

[2] Yang, Shuo, Ping Luo, Chen-Change Loy, and Xiaoou Tang. "Wider face: A face detection benchmark." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 5525-5533. 2016.

[3] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton.

"Imagenet classification with deep convolutional neural networks." In Advances in neural information processing systems, pp. 1097-1105. 2012.

[4] He, Yihui, Xiangyu Zhang, and Jian Sun. "Channel pruning for accelerating very deep neural networks." In Proceedings of the IEEE International Conference on Computer Vision, pp. 1389-1397. 2017.

[5] Georgiadis, Georgios. "Accelerating Convolutional Neural Networks via Activation Map Compression." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7085-7095. 2019.