YOLOv8 을 위한 다중 스케일 Ghost 컨볼루션 기반 경량 키포인트 검출 모델

이자호 [1], 조인휘 [2]
[1] 한양대학교 컴퓨터 소프트웨어학과
[2] 한양대학교 컴퓨터 소프트웨어학과 교수

ijaho1997@hanyang.ac.kr, iwjoe@hanyang.ac.kr

Lightweight Key Point Detection Model Based on Multi-Scale Ghost Convolution for YOLOv8

Zihao Li [1], Inwhee Joe [2]
[1, 2] Dept. of Computer Science, Hanyang University

요 약

컴퓨터 비전 응용은 우리 생활에서 중요한 역할을 한다. 현재, 대규모 모델의 등장으로 딥 러닝의 훈련 및 운행 비용이 급격히 상승하고 있다. 자원이 제한된 환경에서는 일부 AI 프로그램을 실행할 수 없게 되므로, 경량화 연구가 필요하다. YOLOv8 은 현재 주요 목표 검출 모델 중 하나이며, 본 논문은 다중 스케일 Ghost 컨볼루션 모듈을 사용하여 구축된 새로운 YOLOv8-pose-msg 키포인트 검출 모델을 제안한다. 다양한 사양에서 새 모델의 매개변수 양은 최소 34% 감소할 수 있으며, 최대 59%까지 감소할 수 있다. 종합적인 검출 성능은 비교적 대규모 데이터셋에서 원래의 수준을 유지할 수 있으며, 소규모 데이터셋에서의 키포인트 검출은 30% 이상 증가할 수 있다. 동시에 최대 25%의 훈련 및 추론 시간을 절약할 수 있다.

1. 서론

인공 지능 기술의 발전으로 자율 주행에서부터 지능형 어시스턴트까지, AI 기술은 우리 일상 생활의 일부가 되었다. 그러나 Transformer[1] 기반의 대규모 딥 러닝 모델, 예를 들어 목표 검출기 DETR[2]의 등장이 늘어나면서 훈련 및 운행 비용이 급격히 증가하였다, 특히 자원이 제한된 장치나 환경에서 더욱 그러하다. 따라서 모델 경량화는 중요한 연구 분야가 되었다.

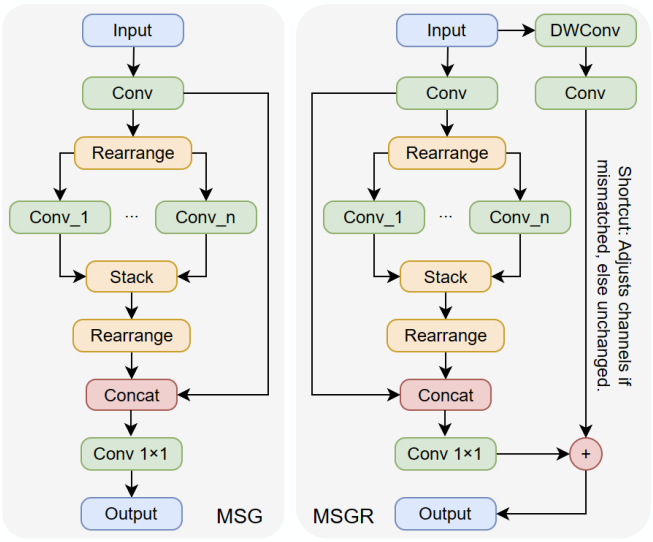
YOLO 시리즈 모델은 목표 검출의 주요 선택으로 여러 번의 업그레이드를 거쳐, YOLOv8[3]은 우수한 성능과 다양한 기능으로 현재 주요 목표 검출기 중 하나가 되었다. 또한, GhostNet[4]은 계산 비용과 모델 크기를 줄이면서 성능을 유지하거나 향상시키기 위해 고안된 효율적인 컨볼루션 네트워크 설계이다.

본 논문은 GhostNet 의 GhostConv 모듈을 기반으로, 다중 스케일 고스트 컨볼루션(Multi-scale GhostConv, 이하 MSG) 모듈을 적용한 YOLOv8-pose-msg 라는 새로운 모델을 설계하였다. 이 작업은 자원이 제한된 환경에서 효율적인 AI 모델을 배치할 가능성을 탐구하며, 모델 경량화 방향에 새로운 시각을 제공한다.

2. 본론

A. 기본 모듈

MSG 모듈을 도입함으로써 전통적인 Ghost 컨볼루션을 최적화하였으며, 주요 특징은 다음과 같다: 1) 다중 스케일 처리를 도입하여 다양한 크기의 컨볼루션 커널을 통해 다중 스케일 특징을 포착함으로써 다양한 객체에 대한 적응성을 향상시킨다; 2) 특성 맵(Feature Map)을 먼저 재배열한 다음 그룹으로 처리하고 쌓고 나서 다시 재배열하여 각 스케일의 특성을 통합한다; 3) 모듈화 설계를 통해 Module List 를 사용하여 컨볼루션 층을 동적으로 조정함으로써 확장성을 강화한다. 또한, 다중 스케일 고스트 컨볼루션 잔차(MSG Residual, 이하 MSGR) 모듈을 설계하여 잔차 연결을 통해 네트워크 심화로 인한 기울기 소실(Vanishing Gradient) 등의 문제를 개선한다. 정의된 잔차 연결에서 입력과 출력 채널 수가 일치하지 않을 경우, 모듈은 깊이 분리 가능한 컨볼루션(DWConv)[5] 등의 작업을 통해 특성 맵을 조정할 것이다.

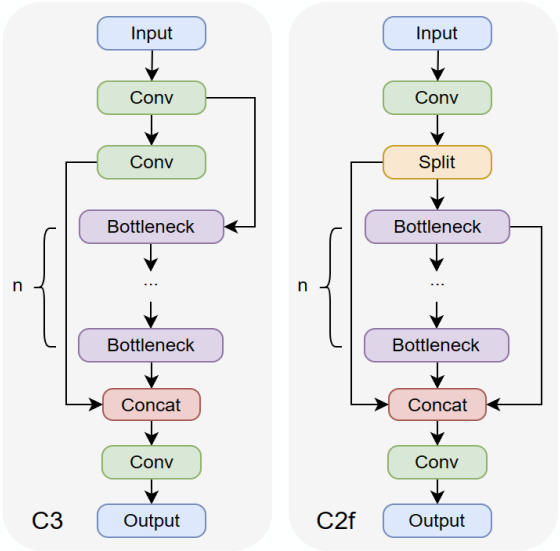


(그림 1) MSG 와 MSGR 모듈 개략도.

B. MSG 기반 모듈

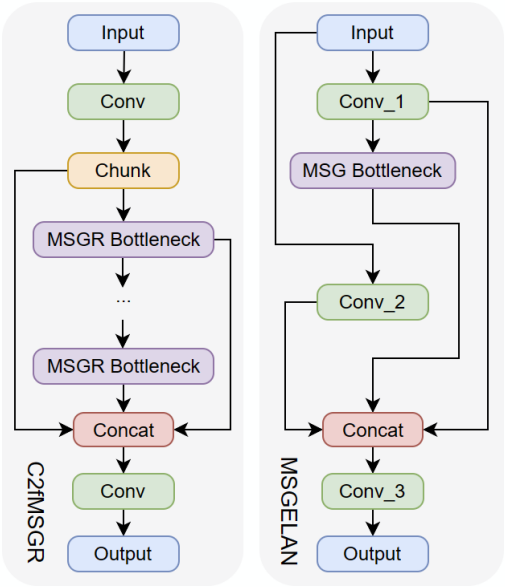
두 개의 동일한 기본 모듈을 연속적으로 연결하여 Bottleneck 모듈을 구성함으로써 보다 깊은 수준의 특징 추출 및 처리를 수행하여, 보다 복잡한 C2fMSGR 모듈과 MSGELAN 모듈을 정의한다. 이 두 가지 모듈은 각각 Backbone 네트워크와 Neck 네트워크 내의 C2f 모듈을 대체할 것이다.

C2f 모듈은 YOLOv5 의 C3 모듈에 기반한 개선된 특별한 CSP Bottleneck 구조이다. 그림 2 에서와 같이, 이 구조는 모델의 규모에 따라 여러 개의 Bottleneck 모듈을 포함하고 있다. 추가적인 분기 교차층(Cross Layers) 연결을 통해 C2f 모듈은 C3 모듈에 비하여 더욱 풍부한 기울기 흐름과 강화된 특성 표현 능력을 가지고 있음을 보여준다.



(그림 2) C3 와 C2f 모듈 개략도.

C2f 모듈과 마찬가지로, C2fMSGR 모듈은 먼저 입력의 채널 수를 확장한 다음, CSP(Cross Stage Partial) 개념[6]을 기반으로 하여, Bottleneck 모듈을 통과할 때마다 출력을 리스트에 추가하고, 마지막에 모든 특성 맵을 합친다. MSGELAN 모듈은 YOLOv7[7]에서 ELAN(Efficient-Layer Aggregation Network) 설계 기반으로 하여, 새로운 구조를 사용하는 동시에 너무 많은 전환층(Transition Layer)을 사용하지 않도록 한다. 이러한 구조는 특성 맵의 크기나 채널 수를 변경하지만, 연구에 따르면 전환층을 통과하는 동안 기울기가 손실되거나 변형될 수 있는 경우가 있다. 따라서 모듈 수정에 대한 최적화 방향은 이러한 설계를 최대한 줄이는 것이다.



(그림 3) C2fMSGR 와 MSGELAN 모듈 개략도.

C. 경량화 결과

YOLOv8-pose 의 공식 문서[3]를 참고하고, 최소(n)·중등(m)·최대(x)의 세 가지 규격으로 예를 들어 모든 모델의 매개변수 양과 GFLOPs 의 통계를 통해 경량화의 정도를 확인한다. 표 1 과 같이, msg 는 MSG 기반의 새로운 모델을 나타낸다.

<표 1> 매개변수와 GFLOPs 통계

	n	n-msg	m	m-msg	x	x-msg
Params	3.3M	2.18M	26.4M	12.6M	69.5M	28.5M
GFLOPs	9.3	6.7	81.2	41.0	263.9	112.2

D. 실험 및 성능 평가

이 절에서는 3 가지 상황을 설계하여 실험을 수행하고 성능 분석을 진행할 것이다. 다른 실험별로 관

련 데이터셋은 다음과 같다: 1) 대형 데이터셋 COCO-Pose, 15 만 개 이상의 인체 샘플을 포함한다; 2) 소형 데이터셋 Tiger-Pose[8], 200 여 장의 이미지만을 포함하여 데이터 부족 상황을 시뮬레이션하기 위해 사용된다; 3) 중형 데이터셋 COCO-Pose-n, COCO-Pose 의 1/10 규모로 실제 상황을 시뮬레이션하기 위해 사용된다. 모든 실험은 동일한 실험 조건 및 매개변수를 기반으로 수행되었으며, 실험 결과는 여러 차례 실험의 평균값을 취했다.

<표 2> 훈련 성능

Groups	Models	B(mAP50)	B(mAP50-95)	P(mAP50)	P(mAP50-95)
Exp.1	YOLOv8n-pose	0.889	0.656	0.733	0.399
Exp.1	Ours (n-scale)	0.880	0.649	0.725	0.392
Exp.2	YOLOv8n-pose	0.995	0.925	0.526	0.131
Exp.2	Ours (n-scale)	0.995	0.929	0.805	0.174
Exp.3	YOLOv8x-pose	0.873	0.624	0.665	0.302
Exp.3	Ours (x-scale)	0.875	0.621	0.659	0.296

실험 결과에서 볼 수 있듯이, MSG 기반 모델은 상대적으로 복잡한 데이터셋에서 Baseline[3]과 기본적으로 동일한 성능을 유지하며, 소형 데이터셋에서의 성능은 30% 이상 향상되었다.

<표 3> COCO-Pose-n 데이터셋 훈련 시간과 추론 속도

Models	Training Duration /100 epochs	Avg. Inference Speed (32s 720p@24fps)	Inference Time (32s 720p@24fps)
YOLOv8x-pose	5.203 hours	24.9 FPS	71.35s
Ours (x-scale)	3.889 hours	34.7 FPS	53.13s

실제 상황에 더욱 부합하는 중형 데이터셋을 사용할 경우, MSG 기반 모델은 훈련 시간을 25.3% 줄일 수 있다. 또한, 동영상 클립을 사용한 추론 테스트에서 새 모델은 기존 모델보다 25.5% 빠른 성능을 보인다.

3. 결론

전통적인 컨볼루션 연산은 대량의 특성 맵을 생성하기 때문에 더 많은 계산 자원이 필요하다. 이에 본 논문에서는 MSG 모듈을 기반으로 한 YOLOv8 Pose 검출기를 제안한다. 실험 결과에 따르면, 새 모델은 최대 59%의 매개변수와 57.4%의 GFLOPs 를 줄일 수 있다. 검출 성능 측면에서, 새 모델은 소형 데이터셋에서 우수한 성능을 보이며, 중대형 데이터셋에서는 기존 모델의 성능을 유지할 수 있다. 동시에, 훈련 및 추론 단계에서 최대 25%의 시간을 절약한다. 따라서 이 설계는 특성 추출 능력이 필요하고 경량화를 추구하는 네트워크 아키텍처, 특히 소형 모바일 장치나 에지 컴퓨팅 장치와 같이 자원이 제한된 환경에서 딥러닝 모델을 실행하는 경우에 적합하다.

참고문헌

[1] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.

[2] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020, August). End-to-end object detection with transformers. In European conference on computer vision (pp. 213-229). Cham: Springer International Publishing.

[3] Jocher, G., Chaurasia, A., & Qiu, J. (2023). Ultralytics YOLO (Version 8.1) [Computer software]. https://github.com/ultralytics/ultralytics

[4] Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C., & Xu, C. (2020). Ghostnet: More features from cheap operations. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 1580-1589).

[5] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

[6] Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.

[7] Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 7464-7475).

[8] Ultralytics. (n.d.). Tiger-Pose Dataset. Ultralytics. Retrieved March 5, 2024, from https://docs.ultralytics.com/datasets/pose/tiger-pose