Comparison of Object Detection and Instance Segmentation Methods in Chest X-rays

Terry Griffin
Dec 15, 2020

Introduction

- This project investigates object detection and instance segmentation methods for identifying abnormalities in chest X-rays (CXR). The aim is to determine if there is a method which will outperform Mask R-CNN [1] and Faster R-CNN [2] for this specific task.
- We test the performance of several state of the art models on CXR datasets with bounding box or mask annotations.
- The results include an evaluation of the best methods for the task of object detection and segmentation in CXRs. Any differences in performance across the available datasets was also analyzed.

Motivation

- Recent progress in object detection and segmentation has produced methods which report to be superior to Faster R-CNN or Mask R-CNN.
- The task of identifying objects in natural images is different from identifying diseased areas in CXRs. We would like to find the best method for our task, which may be different than the best method for natural images.
- CXR Datasets with object level annotations have become available recently. These datasets can be used to verify and enhance the results from our own UML-Peru datatset.

Related Work

- Object detection and instance segmentation on natural images has had much attention in the past five years, with many methods being proposed. Many of these methods are evaluated on the COCO [17] dataset.
- There have been relatively few papers on object detection of lung issues on CXRs [3][4][5] .
- We have evaluated Mask R-CNN and Faster R-CNN on the UML-Peru dataset previously. One aim is to improve on these results.

Approach

- Test available models for object detection and instance segmentation.
- Measure the mean average precision (mAP), average precision (AP) at 0.5 intersection over union (IoU), and average recall with 300 detections per image (AR)
- Compare the performance of the object detection methods versus the instance segmentation methods.
- Compare the class specific task with the class agnostic task.
- Determine the best model architecture for this task.

Datasets

- Our own dataset, UML-Peru, contains masks for 1,320 images using 11 categories. Of these, four categories contain enough images for classification: Airspace Consolidation, Cavitation, Lymphadenopathy, and Pleural Effusion. There are 1,186 images annotated with these classes. This is the only dataset which can be used for the instance segmentation task.
- TBX11K [3] contains 800 images with bounding box annotations for two classes: Active and Latent TB.
- ChestX-14 (NIH) [4] contains 800 images with bounding box annotation for 8 classes: Atelectasis, Effusion, Cardiomegaly, Infiltrate, Pneumonia, Pneumothorax, Mass, and Nodule. There are between 79 and 180 instances of each class.
- ChestX-Det [5] adds bounding box annotations for 3,543 images from ChestX-14, using 10 classes: Atelectasis, Calcification, Consolidation, Effusion, Emphysema, Fibrosis, Fracture, Mass, Nodule, Pneumothorax

Models

- The models fall into two categories: those for object detection, producing bounding boxes, and those for instance segmentation, producing masks. Models for both tasks are evaluated as three of the four datasets contain only bounding box annotations. We also want to compare the performance of these two tasks on the UML/Peru dataset.
- Many object detection and segmentation methods use a ResNet [6] backbone. Investigating some newer architectures may provide improvement.
- The exact mix of models and backbones is limited by time for the course project.

Implementation Details

- All models use the PyTorch framework
- Three libraries were used:
  - MMDetection (latest)
  - SOLOv2 version based on a fork of MMDetection
  - AdelaiDet, which is based on Detectron2
- All models were trained on a Linux server with two nVidia GTX 1080 Ti GPUs.
- Hyperparameters were tuned individually for each model for the best performance.

Object Detection Models

- Faster R-CNN [2]
- RetinaNet [7]
- Fully Convolutional One-Stage Object Detection (FCOS) [8]
- Single Shot Multibox Detector (SSD) [9]
- You Only Look Once (YOLOv4) [10]
- All models were tested using the latest MMDetection library.

Instance Segmentation Models

- Mask R-CNN [1] – using latest MMDetection library
- Segmenting Objects by Locations (SOLO v2) [11] – using a SOLOv2 implementation based on a fork of MMDetection.
- BlendMask [12] – using the AdelaiDet library, which is based on Detectron2.

Backbone Networks

- Faster R-CNN, Mask R-CNN, RetinaNet, and SOLOv2 were tested with different size ResNet and ResNeXt backbones.
- BlendMask was tested with ResNet-50 and ResNet-101 backbones.
- SSD and YOLOv3 were tested only at the largest standard sizes, as we are more concerned with performance over speed.

Class Specific Object DetectionResults

CLASS-SPECIFIC OBJECT DETECTION RESULTS

	UML/Perú			ChestX-ray14			ChestX-Det10			TBX11K		
	mAP	AP50	AR	mAP	AP50	AR	mAP	AP50	AR	mAP	AP50	AR
Faster R-CNN	0.207	0.534	0.451	0.180	0.457	0.393	0.201	0.515	0.447	0.190	0.501	0.411
SSD512	0.186	0.386	0.459	0.147	0.327	0.396	0.179	0.377	0.436	0.161	0.400	0.401
RetinaNet	0.216	0.559	0.477	0.198	0.487	0.412	0.223	0.537	0.470	0.192	0.436	0.478
YOLOv3-608	0.196	0.395	0.390	0.168	0.371	0.363	0.188	0.400	0.396	0.181	0.411	0.423
FCOS	0.189	0.383	0.455	0.143	0.325	0.381	0.182	0.379	0.435	0.158	0. 401	0.410

Class Agnostic Object Detection Results

CLASS-AGNOSTIC OBJECT DETECTION RESULTS

	UML/Perú			ChestX-ray14			ChestX-Det10			TBX11K		
	mAP	AP50	AR	mAP	AP50	AR	mAP	AP50	AR	mAP	AP50	AR
Faster R-CNN	0.247	0.575	0.512	0.190	0.462	0.421	0.222	0.534	0.468	0.210	0.521	0.431
SSD512	0.193	0.393	0.471	0.158	0.331	0.402	0.186	0.395	0.456	0.173	0.421	0.409
RetinaNet	0.265	0.595	0.525	0.217	0.490	0.433	0.241	0.553	0.483	0.205	0.448	0.495
YOLOv3-608	0.201	0.452	0.493	0.170	0.358	0.402	0.193	0.400	0.462	0.193	0.437	0.425
FCOS	0.195	0.396	0.475	0.156	0.328	0.399	0.190	0.397	0.458	0.177	0.420	0.412

Instance Segmentation Results

INSTANCE SEGMENTATION RESULTS

	UML/Perú		
	mAP	AP50	AR
Mask R-CNN	0.197	0.534	0.461
BlendMask	0.112	0.287	0.389
SOLOv2	0.135	0.325	0.383

Results

- RetinaNet outperformed Faster R-CNN.
- Other object detection networks did not do as well.
- Mask R-CNN has the best instance segmentation performance.
- The class-agnostic test has only slightly better performance than the class specific task. This indicates that identifying the correct class is not the source of most errors.
- The best object detection results were better than the best instance segmentation task results. Finding bounding boxes is easier than finding masks.

Results (cont'd)

- All models showed consistent behavior across the four datasets.
- The UML/Peru dataset had the strongest performance. This could be due to having annotations by a single expert, which may be more consistent than other datasets.
- The smallest dataset, ChestX-ray14 had the weakest performance.

Future Work

- Include other networks
  - YOLOv3, including an ablation study of special features
  - Efficient-Det
- Investigate using different backbones, such as DenseNet

References

[1] K. He, G. Gkioxari, P. Dollar, and R. Girshick, “Mask R-CNN,” in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PP, pp.1–1, 06 2018.

[2] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks” in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.

[3] Y. Liu, Y. H. Wu, Y. Ban, H. Wang, and M. M. Cheng, “Rethinking computer-aided tuberculosis diagnosis,” in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 2643–2652.

[4] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. Summers, “Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 3462–3471.

[5] J. Liu, J. Lian, and Y. Yu, “ChestX-Det10: Chest x-ray dataset on detection of thoracic abnormalities,” 2020. [Online]. Available:https://arxiv.org/abs/2006.10550

[6] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016

References (cont'd)

[7] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PP, pp. 1–1, 07 2018.

[8] Z. Tian, C. Shen, H. Chen, and H. Tong, "Fcos: Fully convolutional one-stage object detection," in *IEEE International Conference on Computer Vision*, 10 2019, pp. 9626–9635.

[9] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. Berg, "SSD: Single shot multibox detector," in *European Conference on Computer Vision*, vol. 9905, 10 2016, pp. 21–37

[10] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," 2020. [Online]. Available:https://arxiv.org/abs/2004.10934

[11] X.Wang, R.Zhang, T.Kong, L.Li, and C.Shen,"Solov2: Dynamic, faster and stronger," 2020. [Online]. Available:https://arxiv.org/abs/2003.10152

[12] H. Chen, K. Sun, Z. Tian, C. Shen, Y. Huang, and Y. Yan, "Blendmask: Top-down meets bottom-up for instance segmentation," 2020. [Online].Available: https://arxiv.org/abs/2001.00309

References (cont'd)

[13] H. Zhang, C. Wu, Z. Zhang, Y. Zhu, Z. Zhang, H. Lin, Y. Sun, T. He,J. Mueller, R. Manmatha, M. Li, and A. Smola, "Resnest: Split-attention networks," 2020. [Online]. Available: https://arxiv.org/abs/2004.08955

[14] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei, "Deformable convolutional networks," in *2017 IEEE International Conference on Computer Vision*. Los Alamitos, CA, USA:IEEE Computer Society, oct 2017, pp. 764–773. [Online]. Available:https://doi.ieeecomputersociety.org/10.1109/ICCV.2017.89

[15] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving into high quality object detection," in *2018 IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[16] K. Chen, J. Pang, J. Wang, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu,J. Shi, W. Ouyang, C. C. Loy, and D. Lin, "Hybrid task cascade for instance segmentation," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4969–4978.

[17] T. Lin, M. Maire, S. J. Belongie, L. D. Bourdev, R. B. Girshick, J. Hays, P. Perona, D. Ramanan, P. Dollar, and C. L. Zitnick, "Microsoft COCO: common objects in context," *CoRR*, vol. abs/1405.0312, 2014.[Online]. Available: http://arxiv.org/abs/1405.0312