TABLE V: Comparison of the performance of fine tuning different pre-trained F-RCNN and M-RCNN models on the SPD documents. The performance scores are MAP @ IOU [0.50:0.95] evaluated via a 5-fold cross-document-validation on 20 SPD documents. The results demonstrate the advantage of PubLayNet over general image datasets in domain adaptation for document layout analysis.

Category	Method	Initialization	F-RCNN	M-RCNN
Text	zero-shot	PubLayNet	0.482	0.468
	fine tuning	PubLayNet	**0.701**	**0.708**
	fine tuning	COCO	0.651	0.661
	fine tuning	ImageNet	0.622	0.629
List	zero-shot	PubLayNet	0.508	0.510
	fine tuning	PubLayNet	**0.681**	**0.684**
	fine tuning	COCO	0.622	0.611
	fine tuning	ImageNet	0.603	0.603
Table	zero-shot	PubLayNet	0.422	0.419
	fine tuning	PubLayNet	0.541	**0.596**
	fine tuning	COCO	**0.560**	0.588
	fine tuning	ImageNet	0.528	0.573
Macro average	zero-shot	PubLayNet	0.470	0.465
	fine tuning	PubLayNet	**0.641**	**0.663**
	fine tuning	COCO	0.611	0.620
	fine tuning	ImageNet	0.584	0.602

to train deep learning models that can accurately recognize the layout of unseen journals. For documents in a distant domain, e.g., government documents and SPD documents, we demonstrated the value of using PubLayNet in a transfer learning setting.

VI. Conclusion

We automatically generated the PubLayNet dataset, which is the largest ever available document layout annotation dataset exploiting redundancy in PCMOA. This dataset allows state-of-the-art object detection algorithms to be trained delivering high performance layout recognition on biomedical articles. Furthermore, this dataset is shown to be helpful to pre-train object detection algorithms to identify tables and different document layout objects in health insurance documents. These results are encouraging since the developed dataset is potentially helpful for document layout annotation of other domains. PubLayNet is available from https://github.com/ibm-aur-nlp/PubLayNet.

As future work, we plan to exploit PMCOA for the automatic generation of large datasets to solve other document analysis problems with deep learning models. For example, PubLayNet does not contain relationships between the layout elements, e.g., a paragraph and a section title. Such information is available in the XML representation and can be exploit to automatically create a dataset of the logical structure of documents.

Acknowledgment

We thank Manoj Gambhir for relevant feedback on this work and for his work in the development of the SPD data set and Shaila Pervin for her work in the development of the SPD data set.

References

[1] P. W. J. Staar, M. Dolfi, C. Auer, and C. Bekas, "Corpus conversion service: A machine learning platform to ingest documents at scale," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ser. KDD '18. New York, NY, USA: ACM, 2018, pp. 774–782. [Online]. Available: http://doi.acm.org/10.1145/3219819.3219834

[2] R. Cattoni, T. Coianiz, S. Messelodi, and C. M. Modena, "Geometric layout analysis techniques for document image understanding: a review," *ITC-irst Technical Report*, vol. 9703, no. 09, 1998.

[3] T. M. Breuel, "Two geometric algorithms for layout analysis," in *International workshop on document analysis systems*. Springer, 2002, pp. 188–199.

[4] ——, "High performance document layout analysis," in *Proceedings of the Symposium on Document Image Understanding Technology*, 2003, pp. 209–218.

[5] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Computer Vision (ICCV), 2017 IEEE International Conference on*. IEEE, 2017, pp. 2980–2988.

[6] S. Schreiber, S. Agne, I. Wolf, A. Dengel, and S. Ahmed, "Deepdesrt: Deep learning for detection and structure recognition of tables in document images," in *Document Analysis and Recognition (ICDAR), 2017 14th IAPR International Conference on*, vol. 1. IEEE, 2017, pp. 1162–1167.

[7] A. Antonacopoulos, D. Bridson, C. Papadopoulos, and S. Pletschacher, "A realistic dataset for performance evaluation of document layout analysis," in *Document Analysis and Recognition, 2009. ICDAR'09. 10th International Conference on*. IEEE, 2009, pp. 296–300.

[8] C. Clausner, C. Papadopoulos, S. Pletschacher, and A. Antonacopoulos, "The enp image and ground truth dataset of historical newspapers," in *Document Analysis and Recognition (ICDAR), 2015 13th International Conference on*. IEEE, 2015, pp. 931–935.

[9] C. Clausner, A. Antonacopoulos, and S. Pletschacher, "Icdar2017 competition on recognition of documents with complex layouts-rdcl2017," in *Document Analysis and Recognition (ICDAR), 2017 14th IAPR International Conference on*, vol. 1. IEEE, 2017, pp. 1404–1410.

[10] A. Shahab, F. Shafait, T. Kieninger, and A. Dengel, "An open approach towards the benchmarking of table structure recognition systems," in *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*. ACM, 2010, pp. 113–120.

[11] J. Fang, X. Tao, Z. Tang, R. Qiu, and Y. Liu, "Dataset, ground-truth and performance metrics for table detection evaluation," in *Document Analysis Systems (DAS), 2012 10th IAPR International Workshop on*. IEEE, 2012, pp. 445–449.

[12] B. Yildiz, K. Kaiser, and S. Miksch, "pdf2table: A method to extract table information from pdf files," in *IICAI*, 2005, pp. 1773–1785.

[13] D. N. Tran, T. A. Tran, A. Oh, S. H. Kim, and I. S. Na, "Table detection from document image using vertical arrangement of text blocks," *International Journal of Contents*, vol. 11, no. 4, pp. 77–85, 2015.

[14] D. He, S. Cohen, B. Price, D. Kifer, and C. L. Giles, "Multi-scale multi-task fcn for semantic page segmentation and table detection," in *Document Analysis and Recognition (ICDAR), 2017 14th IAPR International Conference on*, vol. 1. IEEE, 2017, pp. 254–261.

[15] L. Hao, L. Gao, X. Yi, and Z. Tang, "A table detection method for pdf documents based on convolutional neural networks," in *Document Analysis Systems (DAS), 2016 12th IAPR Workshop on*. IEEE, 2016, pp. 287–292.

[16] A. Gilani, S. R. Qasim, I. Malik, and F. Shafait, "Table detection using deep learning," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, vol. 01, Nov 2017, pp. 771–776.

[17] I. Kavasidis, S. Palazzo, C. Spampinato, C. Pino, D. Giordano, D. Giuffrida, and P. Messina, "A saliency-based convolutional neural network for table and chart detection in digitized documents," *arXiv preprint arXiv:1804.06236*, 2018.

[18] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals," in *Soviet physics doklady*, vol. 10, no. 8, 1966, pp. 707–710.