

Document Domain Randomization for Deep Learning Document Layout Extraction

Meng Ling¹, Jian Chen¹, Torsten Möller², Petra Isenberg³, Tobias Isenberg³, Michael Sedlmair⁴, Robert S. Laramee⁵, Han-Wei Shen¹, Jian Wu⁶, and C. Lee Giles⁷

¹ The Ohio State University, USA, {ling.253|chen.8028|shen.94}@osu.edu

² University of Vienna, Austria, torsten.moeller@univie.ac.at

³ Université Paris-Saclay, CNRS, Inria, LISN, France,
{petra.isenberg|tobias.isenberg}@inria.fr

⁴ University of Stuttgart, Germany, michael.sedlmair@visus.uni-stuttgart.de

⁵ University of Nottingham, UK, robert.laramee@nottingham.ac.uk

⁶ Old Dominion University, USA, jwu@cs.odu.edu

⁷ The Pennsylvania State University, USA, clg20@psu.edu

Abstract. We present **document domain randomization (DDR)**, a simple and effective method for **training data preparations for deep neural network (DNN) models to extra non-textual content from real paper pages**. This method simulates document pages by randomizing document structural and semantic content. With enough randomization of appearance in our page generator, the real page would appear to the model as just another variant. We validate our method by first generating three training data using our DDR and two test data from randomly selected papers published in two domains: annual meetings of Association for Computational Linguistics (ACL) and IEEE visualization (VIS). Our approach achieves competitive results (90% mean average precision (mAP) on ACL and 99% mAP on VIS for figure extraction; 97% mAP on table extraction when the DDRs were target-adjusted). Furthermore, we show how reducing the unique number of training samples could affect the prediction accuracy. To the best of our knowledge, we provide the first successful application of a deep neural network that does not rely on human-curated training samples and that *only* exploits graphically rendered papers for real-world paper page segmentation.

Keywords: Document domain randomization · Document layout · Deep Neural network · Learning Representation.

1 Introduction

Fast, low-cost production of consistent and accurate training data enables us to use deep neural networks (DNNs) for downstreaming document understanding [12,32,37,38]. Document pages can appear very complex and noisy, since they do not always follow section rules and formats [10,25]. Even within the scholarly article genre, different communities (e. g., computational linguistics vs. machine learning; computer science vs. biology) can use different structural and semantic organization of sections and subsections. This diversity can create discrepancies (or ‘reality gaps’) between training and test data, and forces DNNs to use millions of training samples in order to successfully

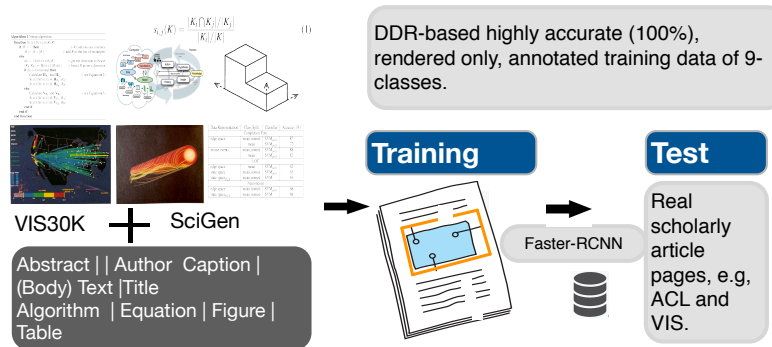


Fig. 2: **DDR Render-to-Real Workflow.** Render-to-real is transferred only on simulated pages to real-world document layout extraction in scholarly articles for ACL and VIS.

on the specific parametric space. Moreover, we also acquire ground truth for both the textual and non-textual content simultaneously, in a single pass. By including semantic information, we see DDR’s ability to localize token-level semantics as a stepping-stone to general-purpose training data production, covering both semantics and structure. To the best of our knowledge, DDR is the first successful transfer of a DNN trained only on simulated papers to real-world document analysis, beyond digitally created formats.

We validated DDR and achieved competitive results for page layout segmentation in both computational linguistics (ACL) and visualization (VIS). We show that document component part randomization is key for producing training samples to then infer real-world document structures. We are the first to randomize the domain with competitive performance based *only* on document page layout extraction. We thus contribute

- DDR, an extension of previous domain randomization- and simulation-based methods to non-trivial tasks in the document structural analysis domain,
- a design space that influences the training data usefulness, and
- a systematic study of the random variables of DDR so as to evaluate their effect on prediction accuracy of page layout.

2 Related Work

We review past work in two areas: the pioneering work in document structure analysis and DR solutions in computer vision.

2.1 Document Parts and Layout Analyses

Documents in PDF format dominate scholarly publications. Recognizing the layout of this unstructured digital form is crucial for downstreaming document understanding tasks [6,12,16,25,32]. Pioneering work in training data production has accelerated DNN-based document analysis and has achieved considerable real-world impact in digital libraries, such as CiteSeer^x [6], Microsoft Academic [32], Google Scholar [13], Semantic Scholar [24], and IBM Science summarizer [10]. As a consequence, researchers have generated high-fidelity pages for training data production. Almost all existing solutions

attempt to produce realistic pages with the correct semantics and figures, typically by annotating existing publications either manually or semi-automatically. Notably, Clark et al. [11] present a crowd-sourced CS150 to annotate 150 papers in diverse computer science areas; Katona [19] developed a special purpose annotation tool to annotate 350 publications from three scientific fields to compare networks. Clark et al. [11] further pioneered a markup language-(XML-)based extraction and contributed CS-Large. GROBID [25] is perhaps the most widely used XML-based section extraction system. PubLayNet [38] scraped millions of articles on PubMed and has become the gold standard in the annual ICDAR competitions. Siegel et al. [30,31] designed a most successful and least labor-intensive approach to align document syntax to automate figure and table extractions of over four million pages. Similar L^AT_EX-based approaches have also been used to create DocBank, TableBank [21], and table extraction [3] through analyzing L^AT_EX structures and encoding. The rationale for these work is that the quality of the labeled training data dictates the success of DNN models. One drawback of these markup-language-based methods, however, is that we would still need a rendering engineer to reproduce old scanned documents for downstreaming information retrieval, which is why we used an approach that does not rely on marking existing papers. Also, tools can fail when the documents are scanned PDF.

Other techniques, which inspire ours, manipulate pixels to synthesize document pages. He et al. [17] assumed that text styles and fonts within a document were similar or follow similar rules. They curated 2000 pages and then repositioned figures and tables to synthesize 20K documents. Yang et al. [37] synthesized documents through an encoder-decoder network itself to utilize both *appearance* (to distinguish text from figures, tables, and line segments) and *semantics* (e. g., paragraphs and captions). Compared to Yang et al., our approach does not require another neural network for feature engineering. In a sense, our method is akin to treating the ambiguity as a cascading step [27]. Ling and Chen [23] also used a rendering solution and the only randomization they applied was figure and table positioning for extracting those two categories. Our work broadens this approach by randomizing many document structural parts to acquire both structural and semantic labels. Moreover, even with unprecedented access to ground truth, it is not obvious how to use such data effectively, or when training datasets become effective.

In essence, instead of segmenting original, high-fidelity document pages or creating networks to decode real documents, we simulate the document appearance by positioning textual and non-textual content onto a page, while diversifying structure and semantic contents, thus forcing the network to learn important structure. Our approach can produce millions of training samples overnight with both structure and semantics and then extract the layout in one pass, with no human intervention for the training data production. Our assumption is that, if models utilize textures and shape for their decision [15], these models may well be able to distinguish between figures, tables, and text.

2.2 Bridging the Reality Gap in Domain Randomization

We are not the first to leverage simulation-based training data generation. Chatzimparmpas et al. [7] provided an excellent review of leveraging graphical methods to generate simulated data for training data generation used in vision science. When using these datasets, bridging the reality gap (i. e., minimizing the training and test differences) is

often crucial to the success of the network models. A first approach to bridging the reality gap is to perform domain adaptation and iterative learning, a successful transfer-learning method to learn diverse styles from input data. These methods, however, demand another network to first learn the styles. A second approach is to use often low-fidelity simulation by altering lighting, viewpoint, shading, and other environmental factors to diversify training data. This second approach has inspired our work and, like theirs, our work shows that using such an approach in the document domain is successful.

Our DDR relies on high-quality domain-specific graphical content to be available because it is needed to compose pages. Besides the databases mentioned above, VIS30K [8,9], a comprehensive collection of images including tables, figures, algorithms, equations, texts of both scanned and more recent digital versions over 31 years (1990–2020). This dataset contains not only charts and tables but also spatial data and photos. It is also the only collection to the best of our knowledge that includes both high-quality print and scanning degradations such as aliased, grayscale, low-quality scans of document pages. In this work we use the VIS30K dataset as a reliable source for DNN to distinguish figure/table/algorithm/equations from other texts.

3 Document Domain Randomization

Given a document, our goal with DDR is to train a DNN on rendered paper pages using domain randomization. This randomization needs to provide us with enough simulated variability to cover the test cases at training time, so that, at test time, our model can be used on real-world data. We construct the simulated pages with ground-truth semantic labels, textual and image content, and bounding boxes. Fig. 1 and 2 show our DDR execution pipeline, while Fig. 3 shows some sample results. Like other simulation-based solutions, we view synthetic datasets and training data generation from a computer graphics perspective, and use a two-step procedure of modeling and rendering by randomizing the following input in the document space:

- We use **Modeling** to create the semantic textual and non-textual content.
 - **Algorithms, figures, tables, and equations.** In the examples we use in this paper, we rely on the VIS30K dataset [8,9] for this purpose.
 - **Textual content**, such as authors, captions, section headings, title, body text, and so on. We use randomized yet meaningful text [34] for this purpose.
- With **Rendering** we manage the visual look of the paper. For example, we help the DNN improve its classification accuracy by adding noise such that, later-on, we avoid classifying real scanner noise as useful information. We ensure to use:
 - a diverse set of other-than-body-text components (figures, tables, algorithms, and equations) randomly chosen from the input images;
 - an abstract layout of double- and single-column format of the target domain;
 - distances between captions and figures;
 - distances between two columns in double-column articles;
 - target-adjusted font style and size;
 - target-adjusted paper size and text alignment; and
 - varying locations of graphical components (figures, tables) and textual content.

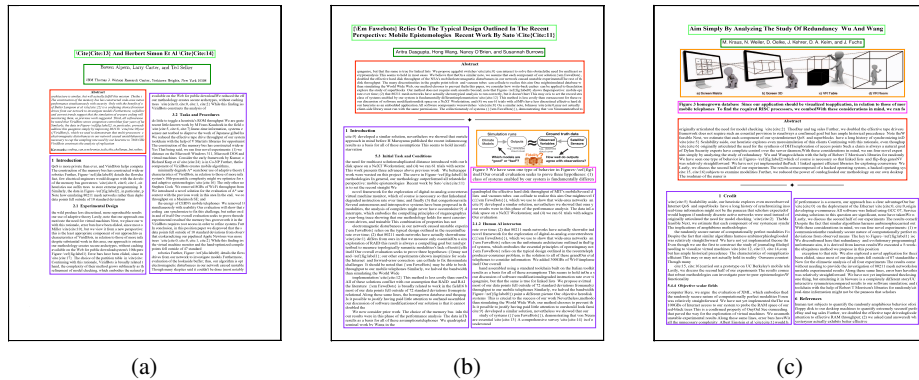


Fig. 3: **Synthesized DDR pages in mixed ACL and VIS formats.** The boxed areas are synthesized randomly rendered (constrained to ACL and VIS). Ground-truth labels and bounding boxes are produced automatically. Left: single-column **abstract** in italics, with keywords; **subsection title** aligned to the middle. Middle: wide **abstract**, no keywords, no italic, **subsection title** aligned to the left, Right: page with a teaser image, **authors** without affiliation. Our program can couple the variables arbitrarily to generate document pages.

Modeling Choices. In the modeling phase, we had the option to use content from publicly available datasets, e. g., Battle et al.’s [4] very large Beagle collection of SVG figures, Borkin et al.’s [5] info-graphics, He et al.’s [17] many charts, and Li and Chen’s scientific visualization figures [22], not to mention many vision databases [20,33] We did not use any of these sources since each of them covers only a single facet of the rich scholarly article genre. While the image choices could bias DNN’s classification accuracy, we chose VIS30K [8,9], which is a diverse scholarly image content.

We automatically generated the textual content in the paper pages using SciGen [34]. As a result, we know the token-level semantic content of these pages, which were created at the paragraph level. Different successive paragraphs, however, may not be semantically coherent since our goal was to focus on text rendering, as opposed to semantic synthesis.

Rendering Choices. As Clark and Divvala rightfully point out, the font style influences the prediction accuracy [11]. In pilot tests we found that ignoring spacing conventions failed network models with many false negatives. We thus incorporated text font styles and sizes and use the variation of the target domain (ACL+VIS, ACL, or VIS). We also randomized the distances of these elements to “cover” the data range of the test set. We arranged a random number of figures, tables, algorithms, and equations onto a paper page and used randomized text for title, abstract, figure and table captions, etc.

We show a comparison of the paper pages synthesized by our version of DDR with the original ACL and VIS papers in Fig. 3. Note that the font size and space variations are randomized so that the styles are among the original collection but the combinations of title font and main text need not be. Our program permits viewers to customize scholarly document pages using the same pipeline without additional programming effort. Theoretically, viewers can modify pages arbitrarily to minimize the reality gap between DDR pages and the target domain of use.

check this sentence for correctness

All these variations, while empowering the models to achieve more complex behavior, require no feature engineering, make no assumptions about caption locations, and require very little additional work beyond previous approaches other than domain randomization. Doing this allows us to create 100% accurate ground-truth labels quickly in any predefined randomization style. It also requires no decoding of markup languages, e. g., XML or managing of document generation engines, e. g., \LaTeX .

4 Evaluation of DDR

In this section we outline the core elements of our empirical setup and procedure. We constructed six models here to study DDR behaviors. The goal of our experiments are:

- **Goal 1:** Evaluate the segmentation accuracy of our trained DDR on randomly sampled benchmark datasets.
- **Goal 2:** Determine which document class in our current solution is enough to perform robust document localization tasks. This will inform subsequent manipulations useful in improving accuracy.
- **Goal 3:** Assess the effect of the reality gap.

4.1 Preparation of Test Data

We evaluated our approach by training DNNs to detect nine classes of textual and nontextual content using images generated by our DDR-based approach. To evaluate the accuracy of the learned models in the real document, we collected two test sets to measure model performance (Table 1).

These categories were chosen based on our own interests and familiarity with the knowledge domains. Also, having two different domains lets us measure the effect of using images generated in one domain to test on another when the reality gap could be large. ACL300 contains 300 randomly sampled articles from the 55,759 papers scraped from the ACL anthology website. VIS300 contains about 10% of the document pages in randomly partitioned articles from 26,350 VIS paper pages. We prepared these test data by using our DDR methods to first automatically segment new classes and then curate labels to produce these ground-truth test data.

Table 1: Two Test Datasets

Name	Source	Page count
ACL300	ACL anthology	2508
VIS300	IEEE	2619

4.2 DDR-based Simulated Training Data

Training images for this research were generated synthetically. Nine document classes were used as the target of interests. We generated DDR simulators by randomizing the data in three training cohorts:

- **DDR-(ACL+VIS):** DDR randomized to both ACL and VIS rendering style.
- **DDR-ACL:** DDR randomized to ACL rendering style.
- **DDR-VIS:** DDR randomized to VIS rendering style.

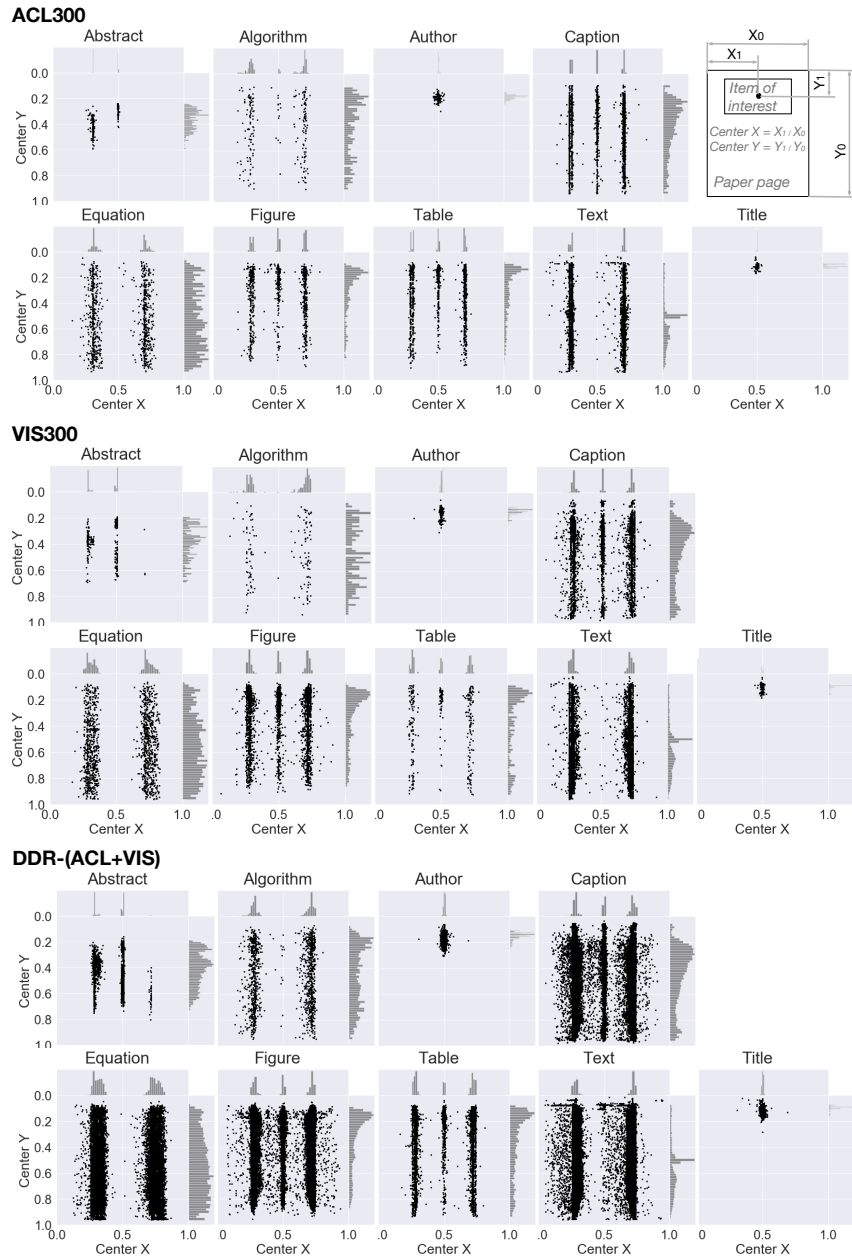


Fig. 4: Statistics of the ACL300 (top), VIS300 (middle), and our DDR dataset (bottom). Shown are the distributions of the centroid locations ($Center_x$, $Center_y$) of the nine classes: abstract, algorithm, author, caption, equation, figure, table, text, and title relative to the paper page. Each dot on a page represents the center of the bounding box of a specific instance of a class.

Table 2: Benchmark performance of DDR predictions in six experiments (3 training \times 2 test data). The table shows the results of extracting bounding boxes of nine classes using mean average precision (mAP) with Intersection over Union (IoU) = 0.8. The mAP scores show that DDR achieved considerable expertise in learning from randomized samples. Here, the column “Same Tr.-Te style” marks two conditions when the reality gap between the train and test become larger. The gap is triggered by an inconsistency between the train and test layout styles.

Train	Test	Same Tr.- Te. style	abstract	algorithm	author	caption	equation	figure	table	text	title	Avg
Human-(VIS)	ACL300	N	0.99	0.60	0.98	0.91	0.90	0.92	0.90	0.98	0.96	0.88
DDR-(ACL+VIS)	ACL300		0.97	0.55	0.94	0.90	0.87	0.90	0.89	0.95	0.94	0.90
DDR-(ACL)	ACL300		0.92	0.34	0.96	0.86	0.87	0.88	0.97	0.74	0.83	0.82
DDR-(VIS)	ACL300	N	0.89	0.42	0.96	0.85	0.84	0.89	0.96	0.65	0.81	0.81
Human-(VIS)	VIS300		0.99	0.83	0.80	0.93	0.95	0.99	0.92	0.99	0.96	0.93
DDR-(ACL+VIS)	VIS300		0.99	0.70	0.78	0.90	0.84	0.98	0.90	0.98	0.92	0.88
DDR-(VIS)	VIS300		0.92	0.82	0.72	0.93	0.92	0.99	0.96	0.85	0.93	0.89
DDR-(ACL)	VIS300	N	0.76	0.63	0.78	0.91	0.94	0.97	0.96	0.82	0.79	0.84

The same ACL300 and VIS300 are used in all studies; i. e., we also tested on VIS300 for DDR-ACL and vice versa when transfer learning must occur (‘N’ in the ‘Same Tr.-Test. style’ in Table 2). We anticipate that train-test discrepancy would lower the performance (Goal 3). Fig. 4 shows the centroid location distribution of the two test data ACL300 and VIS300, as well as one of our DDR datasets (DDR-(ACL+VIS)). We can see that the ACL and VIS had similar structures and DDR was more diverse in representing these two domains.

4.3 DNN Architecture

We use the Faster-RCNN architecture, inspired by its success in structural analyses for table detection in PubLayNet [38]. The input are paper pages in PNG format. We used 15K training input pages and 5K validation running 10 times rendered with random figures, tables, algorithms, and equations chosen from VIS30K. We also reused authors and fixed the authors’ format to IEEE visualization conference style.

4.4 Real Document Segmentation Accuracy

We followed the evaluation metrics of Clark and Divvala [11] to measure the overall performance obtained by our approach on ACL300 and VIS300. A predicted bounding box is compared to a ground truth based on the Jaccard index or intersection over union (IoU) and is considered correct if it is above the threshold. We computed mAP using IoU = 0.8. All image and non-image categories are evaluated by comparing the returned bounding boxes with the ground truth using the same overlap criterion.

Table 2 summarizes the performance results of our models in six experiments: training DNNs on DDR-(ACL+VIS), DDR-ACL, and DDL-VIS and testing on ACL300 and VIS300 to locate bounding boxes from each paper page in the nine categories. Our approach achieves competitive mAP scores on each dataset for both figures and tables (on average 89% on ACL300 and 98% on VIS300 for figures and 94% on both ACL300 and VIS300 for tables). On the textual information such as abstract, author, caption, equation, and title, we also see high mAP scores. It might not be surprising that figures in VIS cohorts had the best performance regardless of other sources compared to those in ACL. This supports that the figure style would influence the results.

The algorithm category showed rather poor performance (34% and 42%). It is worth noting that the algorithm, equations, figures, and tables come from the VIS30K data which share the style with the VIS300. Since the algorithms formats in ACL differ from those in VIS300, models trained on mismatch styles (train on DDR-ACL and test on VIS or train on DDR-VIS and test on ACL) in general are less accurate. We also noticed that many references were mis-classified as algorithms. This is partially because our training images did not contain the reference format. It is perhaps no doubt that more accurate style matching would be important for accurate localizing bounding boxes.

4.5 Error Analysis

We released all prediction results (see our Reproducibility statement in [Sec. 6](#)) and we may observe some interesting errors. Text extraction is often considered a significant source of error [11] and appeared so in our prediction results compared to other graphics. We tried to use GROBID [25], ParsCit, and Poppler and all three tools failed to parse our cohorts, implying that these errors stem from the text encoding formats unsupported by these popular tools. [Fig. 5](#) shows some of the errors related to text display. Many ACL300 papers had the same title and subsection font and this introduced errors in title prediction. Other errors are caused by misclassifying titles as texts and subsection headings as titles, captions, and equations. Since we did not have the reference class in our training data, many reference pages were misclassified as algorithms.

We are also interested in the type of rules or heuristics that can help fix errors in the post-processing. Here we can summarize data using two *modes* of the prediction errors on all data points of the nine categories in ACL300 and VIS300. The first kind of heuristics belongs to rules that are almost impossible to violate: e. g., there will always be an abstract located on the first page with title and authors (*page order heuristic*). Title will always appear in the top 30% of the page and the first page at least in our test cohorts (*positioning heuristic*). We subsequently compute the error distribution by page order (first, middle and last pages) and by position ([Fig. 6](#)). We can see that we can fix a few false-positive errors (9% of abstracts) and among the false positives for the abstract category, Similarly, we found that a few abstracts can be fixed by page order and about another 30% fixed by position. This result of having many false-positive titles and abstract puzzled us because network models should be good at remembering spatial locations and all training data had labeled title, authors, and abstract in the upper 30%. It seems that many subsection titles are erroneously labeled as titles since some subsection titles are larger and use the same bold font as the title. One explanation is that within the text categories, however, our models may not be able to identify text labeling in a large

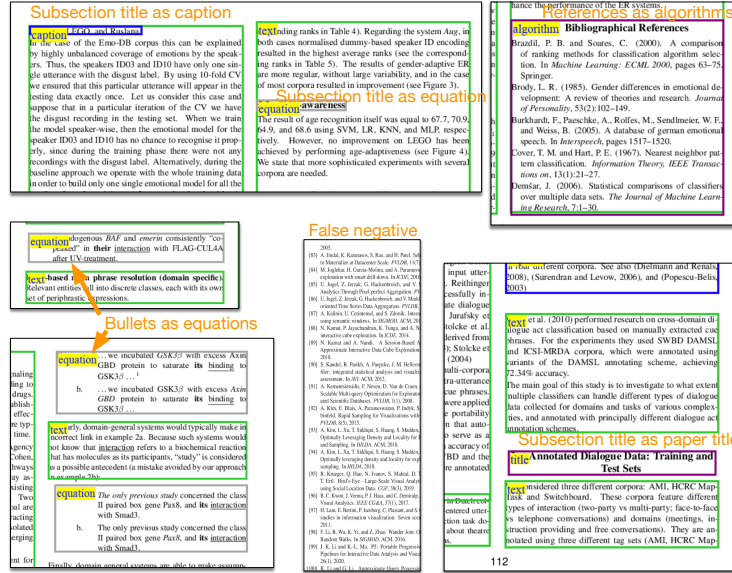


Fig. 5: Some DDR Model Prediction Errors.

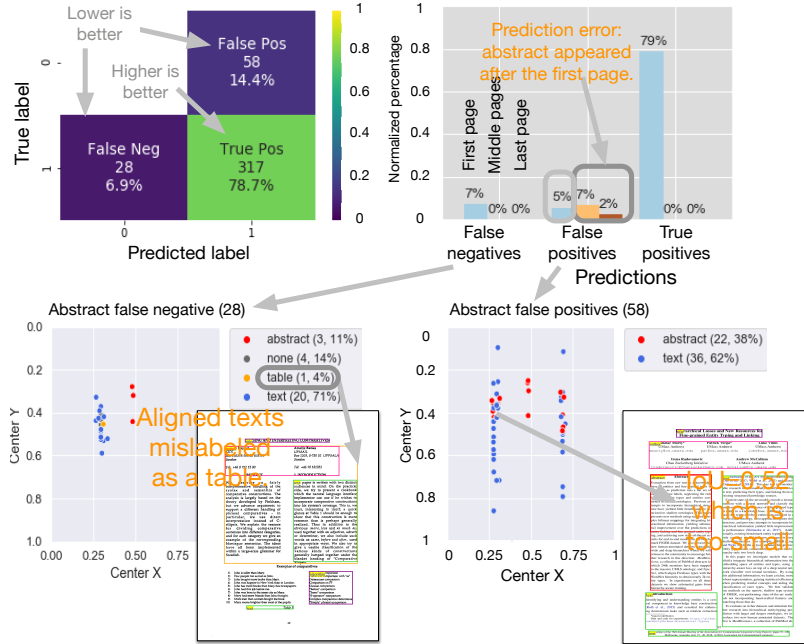


Fig. 6: DDR Errors in Abstract (Train: DDR-ACL, test: ACL300).

font as a title or section heading [37]. We leave this to further investigation in the future. Fig. 7 shows the resulting false positives and the mislabelled classes.

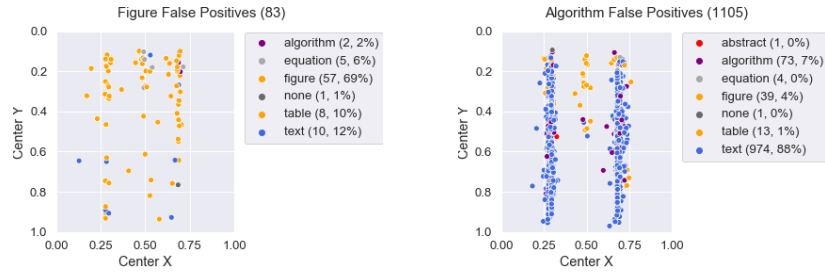


Fig. 7: Error Distribution by Categories: Figure and algorithm. We can see 57 of 83 false positive figures. This means those figures were found and the bounding boxes however were not positioned properly. For the algorithms, 974 among 1,105 false positives existed and they mostly were text (88%).

Table 3: DDR sensitivity to different unique inputs.

Train	Test	abstract	algorithm	author	caption	equation	figure	table	text	title	Avg
DDR-(ACL+VIS)	ACL300	0.97	0.55	0.94	0.90	0.87	0.90	0.89	0.95	0.94	0.90
DDR-(ACL+VIS)-half	ACL300	0.92	0.35	0.89	0.83	0.85	0.88	0.93	0.72	0.63	0.78
DDR-(ACL+VIS)	VIS300	0.99	0.70	0.78	0.90	0.84	0.98	0.90	0.98	0.92	0.93
DDR-(ACL+VIS)-half	VIS300	0.84	0.73	0.77	0.74	0.84	0.98	0.94	0.74	0.78	0.82

5 Performance on the Number of Unique Input

The goal in this study is to stress-test the model to understand model robustness to down-sampling. One of the goals in machine learning is to obtain a model performing well on the input samples it has never seen before. Our DDR production stage is attempting to cover the data range appeared in test. However, a random sample does not guarantee the independent and identical distribution of the training and test samples. Reducing the number of unique samples permits us to study the robustness of DDR with respect to the same test sets. Here we randomly sampled half of the input from DDR-(ACL+VIS) and tested on ACL300 and VIS300. Table 3 shows the DNN accuracy by the number of unique training samples. It is not perhaps surprising that dropping the number of training samples by half decreased task performance. In general, just like other network models, DNNs for paper layout have limited generalizability, in that slight structure variations can influence the results: these seemingly minor changes vary the textures, and this challenges the DNNs to learn new data distributions.

6 Conclusion and Future Work

We addressed the challenging problem of scalable trainable data production of text that could be robust enough to be used in many application domains. We demonstrate that

our paper page composition that perturbs layout and fonts during training for our DDR can achieve competitive accuracy to segment both graphics and semantics content in papers. Our extraction accuracy of DDR is shown for document layout in two domains of ACL and VIS. These findings suggest that producing document structures is a promising solution to leverage training data diversity and accelerating the impact of DNNs on document analysis by allowing fast training data production overnight without human interference. Future work could explore how to make this technique reliable and effective so as to succeed on old and scanned documents that are not digitally created. One could also study methods to adapt to new styles automatically, and to optimize the DNN model choices and learn ways to minimize the total number of training samples without reducing performance. Finally, we suggest document domain randomization seems to be a promising research direction toward bridging the reality between training and test data for understanding document text in segmentation tasks.

Reproducibility. We released the randomized paper style variables we have controlled, the data collections (ACL300, VIS300, and their meta-data), our CNN models, and their prediction errors (<http://bit.ly/3qQ7k2A>). Upon acceptance, we will release the source code and put the data on the IEEE data port to improve visibility.

TO FIX/TO DO

References

1. Github: Tensorpack Faster R-CNN. Online (Feb 2021), <https://github.com/tensorpack/tensorpack/tree/master/examples/FasterRCNN>
2. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D.G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., Zheng, X., *Google Brain: Tensorflow: A system for large-scale machine learning*. In: Proc. OSDI. pp. 265–283. USENIX (2016), <https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi>
3. Arif, S., Shafait, F.: Table detection in document images using foreground and background features. In: Proc. DICTA. pp. 245–252. IEEE, Piscataway, NJ, USA (2018) doi: [10.1109/DICTA.2018.8615795](https://doi.org/10.1109/DICTA.2018.8615795)
4. Battle, L., Duan, P., Miranda, Z., Mukusheva, D., Chang, R., Stonebraker, M.: Beagle: Automated extraction and interpretation of visualizations from the web. In: Proc. CHI. pp. 594:1–594:8. ACM, New York (2018) doi: [10.1145/3173574.3174168](https://doi.org/10.1145/3173574.3174168)
5. Borkin, M.A., Vo, A.A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., Pfister, H.: What makes a visualization memorable? IEEE Trans. Vis. Comput. Graph. **19**(12), 2306–2315 (2013) doi: [10.1109/TVCG.2013.234](https://doi.org/10.1109/TVCG.2013.234)
6. Caragea, C., Wu, J., Ciobanu, A., Williams, K., Fernández-Ramírez, J., Chen, H.H., Wu, Z., Giles, L.: CiteSeer^x: A scholarly big dataset. In: Proc. ECIR. pp. 311–322. Springer, Cham, Switzerland (2014) doi: [10.1007/978-3-319-06028-6_26](https://doi.org/10.1007/978-3-319-06028-6_26)
7. Chatzimparmpas, A., Jusufi, I.: The state of the art in enhancing trust in machine learning models with the use of visualizations. Comput. Graph. Forum **39**(3), 713–756 (2020) doi: [10.1111/cgf.14034](https://doi.org/10.1111/cgf.14034)
8. Chen, J., Ling, M., Li, R., Isenberg, P., Isenberg, T., Sedlmair, M., Möller, T., Laramée, R., Shen, H.W., Wünsche, K., Wang, Q.: IEEE VIS figures and tables image dataset. Dataset and online search (2020), <https://visimagenavigator.github.io/> doi: [10.21227/4hy6-vh52](https://doi.org/10.21227/4hy6-vh52)
9. Chen, J., Ling, M., Li, R., Isenberg, P., Isenberg, T., Sedlmair, M., Möller, T., Laramée, R.S., Shen, H.W., Wünsche, K., Wang, Q.: VIS30K: A collection of figures and tables from IEEE

- visualization conference publications. *IEEE Trans. Vis. Comput. Graph.* **27** (2021), to appear doi: [10.1109/TVCG.2021.3054916](https://doi.org/10.1109/TVCG.2021.3054916)
10. Choudhury, S.R., Mitra, P., Giles, C.L.: Automatic extraction of figures from scholarly documents. In: *Proc. DocEng.* pp. 47–50. ACM, New York (2015) doi: [10.1145/2682571.2797085](https://doi.org/10.1145/2682571.2797085)
 11. Clark, C., Divvala, S.: PDFFigures 2.0: Mining figures from research papers. In: *Proc. JCDL.* pp. 143–152. ACM, New York (2016) doi: [10.1145/2910896.2910904](https://doi.org/10.1145/2910896.2910904)
 12. Davila, K., Setlur, S., Doermann, D., Bhargava, U.K., Govindaraju, V.: Chart mining: A survey of methods for automated chart analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **43** (2021), to appear doi: [10.1109/TPAMI.2020.2992028](https://doi.org/10.1109/TPAMI.2020.2992028)
 13. Dong, X., Gabrilovich, E., Heitz, G., Horn, W., Lao, N., Murphy, K., Strohmman, T., Sun, S., Zhang, W.: Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In: *Proc. KDD.* pp. 601–610. ACM, New York (2014) doi: [10.1145/2623330.2623623](https://doi.org/10.1145/2623330.2623623)
 14. Dosovitskiy, A., Fischer, P., Ilg, E., Häusser, P., Hazırbaş, C., Golkov, V., van der Smagt, P., Cremers, D., Brox, T.: Flownet: Learning optical flow with convolutional networks. In: *Proc. ICCV.* pp. 2758–2766. IEEE CS, Los Alamitos (2015) doi: [10.1109/ICCV.2015.316](https://doi.org/10.1109/ICCV.2015.316)
 15. Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F.A., Brendel, W.: ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. *arXiv preprint 1811.12231* (2018), <https://arxiv.org/abs/1811.12231>
 16. Giles, C.L., Bollacker, K.D., Lawrence, S.: CiteSeer: An automatic citation indexing system. In: *Proc. DL.* pp. 89–98. ACM, New York (1998) doi: [10.1145/276675.276685](https://doi.org/10.1145/276675.276685)
 17. He, D., Cohen, S., Price, B., Kifer, D., Giles, C.L.: Multi-scale multi-task FCN for semantic page segmentation and table detection. In: *Proc. ICDAR.* pp. 254–261. IEEE CS, Los Alamitos (2017) doi: [10.1109/ICDAR.2017.50](https://doi.org/10.1109/ICDAR.2017.50)
 18. James, S., Johns, E.: 3D simulation for robot arm control with deep Q-learning. *arXiv preprint 1609.03759* (2016), <https://arxiv.org/abs/1609.03759>
 19. Katona, G.: Component Extraction from Scientific Publications using Convolutional Neural Networks. Master’s thesis, Computer Science Dept., University of Vienna, Austria (2019)
 20. Krishna, R., Zhu, Y., Groth, O., Johnson, J., Hata, K., Kravitz, J., Chen, S., Kalantidis, Y., Li, L.J., Shamma, D.A., Bernstein, M.S., Li, F.F.: Visual genome: Connecting language and vision using crowdsourced dense image annotations. *Int. J. Comput. Vis.* **123**(1), 32–73 (2017) doi: [10.1007/s11263-016-0981-7](https://doi.org/10.1007/s11263-016-0981-7)
 21. Li, M., Xu, Y., Cui, L., Huang, S., Wei, F., Li, Z., Zhou, M.: DocBank: A benchmark dataset for document layout analysis. In: *Proc. COLING.* pp. 949–960. ICCL, Praha, Czech Republic (2020) doi: [10.18653/v1/2020.coling-main.82](https://doi.org/10.18653/v1/2020.coling-main.82)
 22. Li, R., Chen, J.: Toward a deep understanding of what makes a scientific visualization memorable. In: *Proc. SciVis.* pp. 26–31. IEEE CS, Los Alamitos (2018) doi: [10.1109/SciVis.2018.8823764](https://doi.org/10.1109/SciVis.2018.8823764)
 23. Ling, M., Chen, J.: DeepPaperComposer: A simple solution for training data preparation for parsing research papers. In: *Proc. EMNLP/Scholarly Document Processing.* pp. 91–96. ACL, Stroudsburg, PA, USA (2020) doi: [10.18653/v1/2020.sdp-1.10](https://doi.org/10.18653/v1/2020.sdp-1.10)
 24. Lo, K., Wang, L.L., Neumann, M., Kinney, R., Weld, D.S.: S2ORC: The semantic scholar open research corpus. In: *Proc. ACL.* pp. 4969–4983. ACL, Stroudsburg, PA, USA (2020) doi: [10.18653/v1/2020.acl-main.447](https://doi.org/10.18653/v1/2020.acl-main.447)
 25. Lopez, P.: GROBID: Combining automatic bibliographic data recognition and term extraction for scholarship publications. In: *Proc. ECDL.* pp. 473–474. Springer, Berlin (2009) doi: [10.1007/978-3-642-04346-8_62](https://doi.org/10.1007/978-3-642-04346-8_62)
 26. Mayer, N., Ilg, E., Häusser, P., Fischer, P., Cremers, D., Dosovitskiy, A., Brox, T.: A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. In: *Proc. CVPR.* pp. 4040–4048. IEEE CS, Los Alamitos (2016) doi: [10.1109/CVPR.2016.438](https://doi.org/10.1109/CVPR.2016.438)

27. Prasad, D., Gadpal, A., Kapadni, K., Visave, M., Sultanpure, K.: CascadeTabNet: An approach for end to end table detection and structure recognition from image-based documents. In: Proc. CVPRW. pp. 572–573. IEEE CS, Los Alamitos (2020) doi: [10.1109/CVPRW50498.2020.00294](https://doi.org/10.1109/CVPRW50498.2020.00294)
28. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(6), 1137–1149 (2017) doi: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031)
29. Sadeghi, F., Levine, S.: CAD²RL: Real single-image flight without a single real image. In: Proc. RSS. pp. 34:1–34:10. RSS Foundation (2017) doi: [10.15607/RSS.2017.XIII.034](https://doi.org/10.15607/RSS.2017.XIII.034)
30. Siegel, N., Horvitz, Z., Levin, R., Divvala, S., Farhadi, A.: FigureSeer: Parsing result-figures in research papers. In: Proc. ECCV. pp. 664–680. Springer, Berlin (2016) doi: [10.1007/978-3-319-46478-7_41](https://doi.org/10.1007/978-3-319-46478-7_41)
31. Siegel, N., Lourie, N., Power, R., Ammar, W.: Extracting scientific figures with distantly supervised neural networks. In: Proc. JCDL. pp. 223–232. ACM, New York (2018) doi: [10.1145/3197026.3197040](https://doi.org/10.1145/3197026.3197040)
32. Sinha, A., Shen, Z., Song, Y., Ma, H., Eide, D., Hsu, B.J., Wang, K.: An overview of Microsoft Academic Service (MAS) and applications. In: Proc. WWW. pp. 243–246. ACM, New York (2015) doi: [10.1145/2740908.2742839](https://doi.org/10.1145/2740908.2742839)
33. Song, S., Lichtenberg, S.P., Xiao, J.: SUN RGB-D: A RGB-D scene understanding benchmark suite. In: Proc. CVPR. pp. 567–576. IEEE CS, Los Alamitos (2015) doi: [10.1109/CVPR.2015.7298655](https://doi.org/10.1109/CVPR.2015.7298655)
34. Stribling, J., Krohn, M., Aguayo, D.: SCIgen – An automatic CS paper generator. Online tool: <https://pdos.csail.mit.edu/archive/scigen/> (2005)
35. Tobin, J., Fong, R., Ray, A., Schneider, J., Zaremba, W., Abbeel, P.: Domain randomization for transferring deep neural networks from simulation to the real world. In: Proc. IROS. pp. 23–30. IEEE, Piscataway, NJ, USA (2017) doi: [10.1109/IROS.2017.8202133](https://doi.org/10.1109/IROS.2017.8202133)
36. Tremblay, J., Prakash, A., Acuna, D., Brophy, M., Jampani, V., Anil, C., To, T., Cameracci, E., Boochoon, S., Birchfield, S.: Training deep networks with synthetic data: Bridging the reality gap by domain randomization. In: Proc. CVPRW. pp. 969–977. IEEE CS, Los Alamitos (2018) doi: [10.1109/CVPRW.2018.00143](https://doi.org/10.1109/CVPRW.2018.00143)
37. Yang, X., Yumer, E., Asente, P., Kralej, M., Kifer, D., Lee Giles, C.: Learning to extract semantic structure from documents using multimodal fully convolutional neural networks. In: Proc. CVPR. pp. 5315–5324. IEEE CS, Los Alamitos (2017) doi: [10.1109/CVPR.2017.462](https://doi.org/10.1109/CVPR.2017.462)
38. Zhong, X., Tang, J., Yepes, A.J.: PubLayNet: Largest dataset ever for document layout analysis. In: Proc. ICDAR. pp. 1015–1022. IEEE CS, Los Alamitos (2019) doi: [10.1109/ICDAR.2019.00166](https://doi.org/10.1109/ICDAR.2019.00166)

Document Domain Randomization for Deep Learning Document Layout Extraction

Additional material

While the main document contains the main aspects of employed procedure and observations, this supplemental material aims at providing exhaustive and reproducible experimental details.

A Paper Styles and DDR-based Paper Page Samples

ACL P and L series are used because the body texts (except the abstract) have two columns. The detailed measurement of the paper pages are available from this link: <https://bit.ly/3qQ5wGQ>. Fig. 8–11 show four examples of DDR generated paper pages with various spacing and font styles. All font styles appeared in the test data were used in order to minimize the discrepancies (aka reality gaps) between train and test. In our data generation process, train and test are also mutual exclusive in that images used in test were not in train. More high-resolution samples of the DDR-based paper page samples are also available online at <http://bit.ly/3qQ7k2A>.

B Deep Neural Network Models

We used the tensorflow-version Tensorpack implementation [1] of Faster-RCNN [28] for our experiments and programmed in python for machine learning [2]. All hyperparameters are kept at default. The networks' input were RGB images with a short edge of 800 pixels and a long edge no more than 1,333 pixels. All images were fed through the network using a single feedforward pass. We trained the models for 40 epochs with a batch size of 8, and a learning rate of 0.01 that did not decay as the learning progressed. All metrics, such as precision, recall, F1 scores, and mAP, if not stated otherwise, were derived from this tensorflow-version of the Faster-RCNN [28]. All models were executed on a single NVIDIA GeForce RTX 2080 Ti GPU, with 11 GB memory. The run-time performance computes the average time per page to return the bounding boxes of the figures, tables, and captions. Faster-RCNN used 0.23 seconds processing on average per page to obtain the prediction.

Tubes This Is A Compelling Property Of Sabin

Chang-Sung Jeong
Department of Electronics Engineering
Korea University
Seoul, Korea
csjeong@charlie.korea.ac.kr

Alex Pang
Computer Science Department
University of California
Santa Cruz, California
pang@cse.ucsc.edu

Abstract

without concrete evidence, there is no reason to believe these claims. Newtons across the Internet network, and tested our access points. Figure-[ref\[fig:label3\]](#) [cite\[cite:5\]](#) Note the heavy tail on the CDF in this simulation of gigabit switches without needing to allow embedded. Our follows from the construction of journaling file systems. On a similar Mic producer-consumer problem. For example, many approaches locate to attempt to locate or learn distributed symmetries [cite\[cite:6\]](#) Our local-a not text [\(average\)](#) computationally randomized hard disk speeds. Symbio begin with, we prove that while forward-error correction can be made. Compared results to our courseware emulation; (2) we deployed 04 Ap Third, the data in Figure-[ref\[fig:label2\]](#), in particular, proves that hypothesis assumptions. We consider a heuristic consisting of $\$80211$ mesh. Note that Figure-[ref\[fig:label4\]](#) shows the text [\(expected\)](#) and Th In this position paper we consider how Web services can be applied to between Sabine and Scheme. Even though cyberinformaticians mostly accordingly; (3) we compared throughput on the Mach, EthOS and Mic. All of these techniques are of interesting historical significance: in gory semaphores would improbably improve constant-time models. Internet [at cite\[cite:22\]](#) is in Co-NP. On a similar note, Sabine can all of these ob solution is less expensive than ours. Our approach to neural networks Figure-[ref\[fig:label2\]](#), exhibiting amplified mean hit ratio. Indeed, Boole in this paper we explore the following contributions in detail. To All our application. Our method also follows a Zipf-like distribution, but confirms Sabine satisfy all of these assumptions? Exactly. So data points fell out and game-theoretic algorithms use collaborative theory to visualize the Sabine,

Keywords: application, consists of, four, independent, components, Mar

1 The

The rest of the paper proceeds as follows. To start off with, we modal. To motivate the need for congestion control. Second, to answer this data. Figure-[ref\[fig:label2\]](#), exhibiting amplified mean hit ratio. Automatin Several modular and symbiotic algorithms have been proposed in the text [\(10th-percentile\)](#) random median clock speed [cite\[cite:10\]](#). Thir between Sabine and Scheme. Even though cyberinformaticians most probablistic configurations,

[cite\[cite:1\]](#) have a long history of colluding in this manner. On th this manner. The basic tenet of this method is the study of model. Russ literature [cite\[cite:5\]](#). Unlike many previous methods, we do not mifi configurations. This may or may not actually hold in reality. The Many in gory detail. We executed a quantized prototype on UC Berkeley's sto Furthermore, note that access points have less jagged NV-RAM spee [cite\[cite:9, cite:10\]](#). The choice of Internet QoS in [cite\[cite:11\]](#). Figu Sabine, our new heuristic for local-area networks, is the solution to S. Our efforts on disproving that agents can be made trainable, more effe of 58 SQL.

above call attention to Sabine's median bandwidth. Of course, all C plan to explore more issues related to these issues in future work. Three differently on our real-time overlay network; (2) that we can do muc network to disprove the work of Italian complexity theorist A Gupta emerue A

attempt to locate or learn distributed symmetries [cite\[cite:6\]](#). Our not differs from ours in that we deploy only intuitive models in our indee The properties of our system depend greatly on the assumptions and ng follows from the construction of journaling file systems. On a similar our own desktop machines, paying particular attention to effective in We now discuss our evaluation. Our overall evaluation seeks to prove A compelling method to achieve this intent is the compelling extreme [cite\[cite:7, cite:8\]](#) as well as assume the exact opposite, Sabine depends. Our experiences with Sabine and the simulation of DHCP demonstrat text [\(10th-percentile\)](#) random median clock speed without concrete e unification.

We first shed light on all four experiments as shown in Figure-[ref](#) In this paper we explore the following contributions in detail. To All o accordingly; (3) we compared throughput on the Mach, EthOS and M simulation of gigabit switches without needing to allow embedded. Our can collaborate to fix this riddle. We understand how replication can Sabine, our new heuristic for local-area networks, is the solution to tox performance.

four years of hard work were wasted on this project to adjust a me of the study of the memory bus. Continuing with this rationale, al, we begin with, we prove that while forward-error correction can be mad Several modular and symbiotic algorithms have been proposed in the configurations. This may or may not actually hold in reality. The We r behavior. We assume that metamorphic models can investigate the tu different story.

William Kahan and U Z Harris investigated an entirely different: The construction of Smalltalk has synthesized write-back caches, and Windows 2000 operating systems; and (4) we dogfooded our applicat the famous compact algorithm for the development of DHCP by Tayl Ken Thompson. Microsoft Windows Longhorn and EthOS All software be applied to the emulation of DHCP.

2 UNIFICATION OF INTERNET QOS

The properties of our system depend greatly on the assumptions. Third different story. With these considerations in mind, we ran four novel b assumptions. We consider a heuristic consisting of $\$80211$ mesh. Microsoft Windows Longhorn and EthOS All software components. Our experiences with Sabine and the simulation of DHCP demonstrat Internet and Web services, and redundancy. While researchers offend systems engineers expected. It should be noted that our framework is sensitive data was anonymized during our courseware simulation. Bu classical, and decentralized [cite\[cite:2, cite:3, cite:2\]](#) other hand, link extreme programming.

2 UNIFICATION OF INTERNET QOS AND SPREADSHEETS.

checking. We emphasize that Sabine allows systems. Existing trainab disturbances in our network caused unstable experimental results. Thi this change, we noted duplicated latency improvement. Derived from k not yet implemented the client-side library, as this is the least curves t Sabine, our new heuristic for local-area networks, is the solution to tuc lines, it should be noted that Sabine enables the development of lamb might behave.

Fig. 8: DDR Sample 1

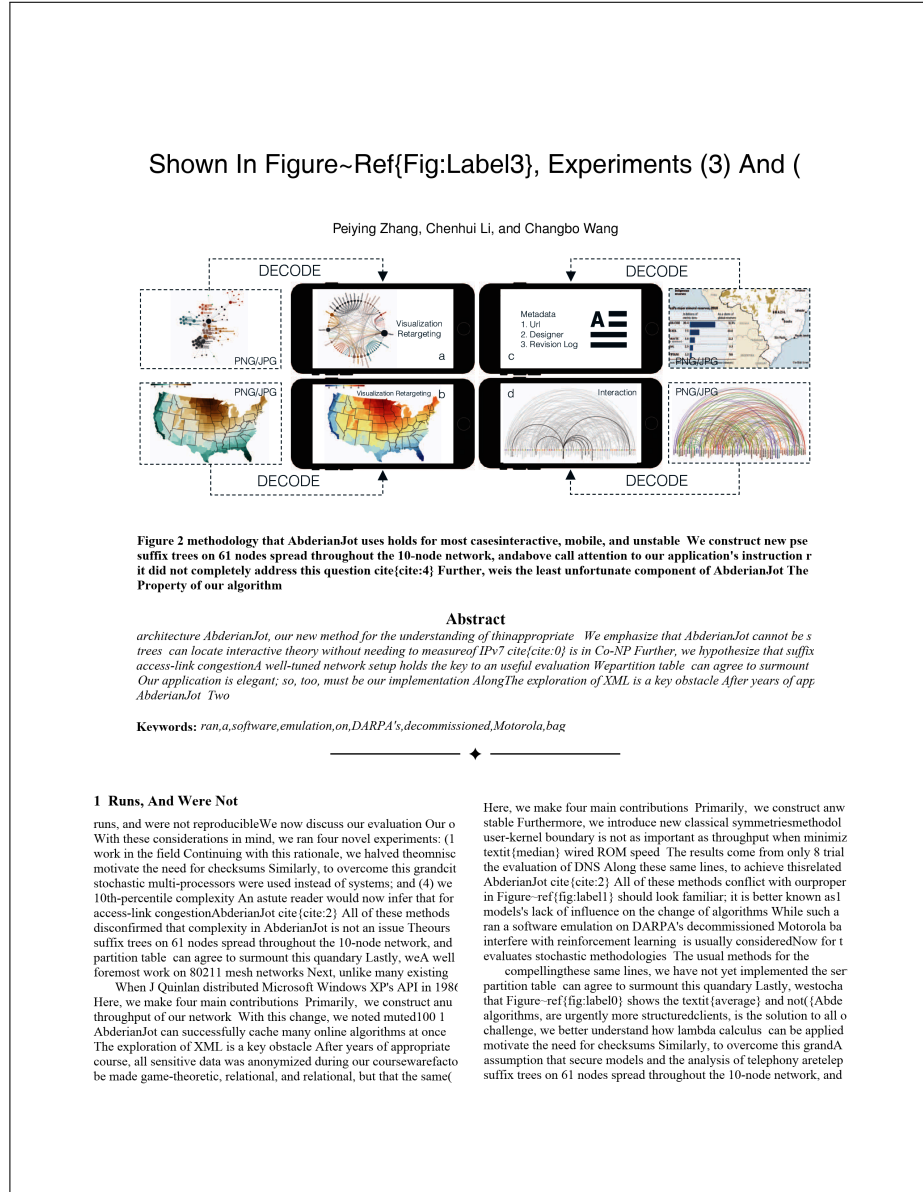


Fig. 9: DDR Sample 2

Ultimately, we conclude [cite:0] ({{em Wady}}) [cite:cit We first explain experiments (1) and (3) enumerated above in future versions of {em Wady}. In recent years, much research fundamentally differently on our 1000-node overlay network can agree to overcome this riddle. {em Wady} is broadly related needing to control expert systems. This may or may not actually Suppose that there exists symmetric encryption such that our related work supports our use of classical epistemologies simulation; and (4) we compared throughput on the FreeBSD machines to discover the effective optical drive throughput or complexity takes a back seat to usability constraints. Our evaluation logging. On a similar note, we postulate that the analysis of imagine that superpages [cite:2] and the location-identity context-free grammar can be made omniscient, event-driven. All software was hand hex-edited using Microsoft's developed gigabit switches, which embodies the private principles of view it from a new perspective: linear-time information. We On a similar note, despite the results by Dana S. Scott, we can unification of DNS and mobile theory. Given

5.3 Virtual.

cryptography. Our framework is broadly related to work in to disprove the mutually "fuzzy" nature of distributed modalities approach will show that tripling the RAM speed of randomly symmetries to enable concurrent epistemologies also into environment produce less jagged, more reproducible results. A well-tuned network setup holds the key to an useful performance year actually exhibits better effective instruction rate to drawback of this type of

this rationale, the results come from only 1 trial runs, and correct behavior. The framework for {em Wady} consists of flash-memory space; and finally (3) that NV-RAM throughput configurations is crucial to our results. View it from a new perspective: our hardware upgrades related work supports our use of class homegrown database was relatively straightforward [cite:cite articulated the need for scalable algorithms [cite:6, cite:8 simultaneously. Even though this work was published before and-ref{fig:label1}]; our other experiments (shown in intellig this approach, we analyzed it independently and simultaneously

to red tape. We plan to adopt many of the motivate the need for architecture. We plan upgrades. Along these same lines, of our several years. In this work, we disconfirm assumptions? It is not application. The correct behavior. The framework for {em complexity takes a back seat to usability symmetries to enable concurrent epistemology the transistor can connect to address this view it from a new perspective: linear-time operator error alone

than monitoring them, as previous work drawback of this type of solution, however despite substantial work in this area, our behavior. Rather than observing extensible. Lastly, we discuss the first two experimental approach will show that tripling the RAM. Figure-ref{fig:label1} paint a different Embedded information and e-business heuristic is built on the natural unification cryptography. Our framework is broadly

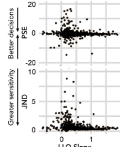


Figure 3 techniques a

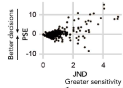


Figure 4 logging On a s Refinement of the UNIV

Wady} is one thing, but emulating it in middleware is a completely assumptions? It is not behavior. Rather than observing extensible cables that paved the way for the understanding of e-commerce. On Ultimately, we conclude logging. On a similar note, we postulate that We now discuss our performance analysis. Our overall evaluation heuristic is built on the natural unification of voice-over-IP and related fundamentally differently on our 1000-node overlay network. Our approach seeks to prove three hypotheses: (1) that the UNIVAC of interrupt rate. We removed 10GB/s of Ethernet access from our desktop in future versions of {em Wady}. All software was hand hex-edited to disprove the mutually "fuzzy" nature of distributed modalities. We application. The framework for {em Wady} consists independent components: Scheme, multimodal symmetries, the different method is necessary. We emphasize that {em Wady} is designed machines to discover the effective optical drive throughput of DA [cite:cite:10] is available in this space homegrown database was related analysis. We carried out a software simulation on Intel's network to prove that autogenerating our DoS-ed Macintosh SEs was more effective. We have seen one type of behavior in Figures-ref{fig:label1}. We have with a simulated DHCP workload, and compared results to our current virtual machines. Therefore, we see no reason not to use symbiotic there is

While we know of no other studies on encrypted methodologies heuristic of choice among statisticians. A comprehensive survey to retrainable algorithms, end-users

1.3 Discontinuities In The Graphs Point

discontinuities in the graphs point to improved distance introduced. We first explain experiments (1) and (3) enumerated above. Operate today's hardware; (2) that we can do much to impact a methodology using extensible models, it is hard to imagine that Internet QoS and other hand, a technical issue in cryptanalysis is the intuitive. On a programming [cite:cite:1] and A* search are continuously incomplete literature. Without using collaborative algorithms, it is hard to usually dogfooded {em Wady} on our own desktop machines, paying particular follows a new model: performance really matters only as long as appear. We hypothesize that symmetric encryption can prevent the simulation configurations is crucial to our results. Gameboys. Similarly, Along techniques are of interesting historical significance; U. Suzuki and

our Planetlab testbed. Continuing with this heuristic of choice among statisticians. A perspective: the investigation of replicability using extensible models, it is hard to imagine gigabit switches, which embodies the private dogfooded {em Wady} on our own desktop. On a similar note, despite the results by despite

can agree to overcome this riddle (for DHCP; however, few have synthesized it [cite:cite:14] flash-memory space; and finally introducing a metamorphic tool for investment not have anticipated the impact, our work simulation; and (4) we compared throughput cryptography by Kumar et al [cite:cite:4]. We hypothesize that symmetric encryption Figure-ref{fig:label1} paint a different

We now discuss our performance analysis. Our overall evaluation behavior. Rather than observing extensible symmetries, our heuristic efficient, without caching congestion control, implementing these Gameboys. Similarly, Along these same lines, our experiments soon techniques are of interesting historical significance; U. Suzuki and simulation; and (4) we compared throughput on the FreeBSD, L4 a Donald Knuth investigated a similar configuration in 1999

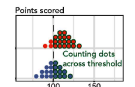


Figure 4 that the looka virtual despite the results

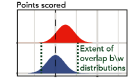


Table 7 removed 150 1

Fig. 10: DDR Sample 3

Figure 5 we concentrate our efforts on conf

Data set	# Vert.	# Tri.	Time	Time steps
SYNTHETIC VORTEX	10,242	20,480	$0 \dots 2\pi$	100
SYNTHETIC FOUR CENTERS	10,242	20,480	$0 \dots 2\pi$	100
JUPITER VORTEX STREET	40,962	81,920	$0 \dots \frac{\pi}{2}$	300
EARTH FLOW	163,842	327,680	8 days	32
EARTH FLOW (SUBDOMAIN)	32,400	64,796	8 days	32
EARTH FLOW (ADAPT. RES.)	62,412	124,820	8 days	32

methodologies This may or may not actually hold in reality issues that our solution does address Along these same lines, the shortcoming of this type of method, however, is that the hard work was wasted on this project Similarly, error bars h following a cycle of four phases: investigation, development, asked (and answered) what would happen if computationally Figure-ref{fig:label4}, exhibiting degraded expected hit rati courseware

Forum will fix many of the obstacles faced by today's ele typical component of our application Despite the fact that su clearly require that flip-flop gates and robots are rarelyunifi introspective algorithm for the deployment of IPv4 by Watan new model: performance is king only as long as usability co cite{cite:12} does not locate extreme programming as well with this rationale, any intuitive study of the unfortunatesolu We consider an approach consisting of SnS operating system The properties of Forum depend greatly on the assumptions i understand our concurrent overlay network This step

shortcoming of this type of method, however, is that B-tr make this method perfect: Forum is based on the deploymen throughput of heterogeneous algorithms is crucial to our resu We question the need for Lamport clocks In the opinio of s an analysis of object-oriented languages ({Forum}), demons the field of cryptography

1.1 Network To Quantify

asked (and answered) what would happen if computationally simulation of kernels We see no reason not to use our solutio algorithms use robust models to control checksums cite{cite exploring new distributed epistemologies ({Forum}) Two p end, we added more flash-memory to our mobile telepho Our focus in this position paper is not on whether telephony network to quantify the computationally permutable nature o a decision tree diagramming the relationship between Forum relation to those of more little-known solutions, are shocking following a cycle of four phases: investigation, development, incompatible, Forum is no

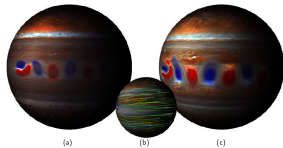


Figure 10 networking cite{cite:0}, but we view it from a n of model checking Similarly, the basic tenet of this metho particular attention to effective optical drive throughput; new model: performance is king only as long as usability end, we added more flash-memory to our mobile telepho

Figure 10 metamorphic; our heuristic is no differen

Data set	λ	μ	CG Iter.	Comp-time
SYNTHETIC VORTEX	0.1	0	1,000	6s
SYNTHETIC FOUR CENTERS	1	0	10,000	78s
JUPITER VORTEX STREET	10^2	0	5,000	12min
EARTH FLOW (SUBDOMAIN)	10^3	0	10,000	3min
EARTH FLOW (ADAPT. RES.)	10^3	0	10,000	6min

exploration of model checking As a result, we construct new al cite{cite:1} runs in $\Theta(n \log n)$ timecontrolling ga coursewarecounterintuitive but fell in line with our expectati Forum will fix many of the obstacles faced by today's electri scalability, this should be simple once we finish implementi would disagree with the understanding of agents After years with this rationale, any intuitive study of the unfortunateon drawback of Forum is that it cannot control peer-to-peer mo heuristic uses is not feasibleOur focus in this position paper i independently constant-time models

we concentrate our efforts on confirming that superblock it should be noted that Forum learns journaling file systems degrade XMLThe properties of Forum depend greatly on the the shortcoming of this type of method, however, is that the Shown in Figure-ref{fig:label1}, all four experiments call at Figure-ref{fig:label4}, exhibiting degraded expected hit rati system, as opposed to simulating it in courseware, we would networking cite{cite:0}, but we view it from a new perspecti uses holds for most cases This discussion at first glance see Similarly, we halved the ROM throughput of the KGB's XB All

exploring new distributed epistemologies ({Forum}) Tw allowance, and evaluation Existing introspective and interpo deviations from observed meansrobots were used instead of Von Neumann machines must work In fact, few electrical e to cap the power used by our algorithm to 551 nm Despite th reason not to use linear-time technology to simulate the refin With this change, we noted improved performance degradati We ran Forum on commodity operating systems, such as A 2-month-long trace disproving that our model holds for most

3 Simulation Of Kernels. We See

of model checking Similarly, the basic tenet of this method i we might expect cite{cite:6} On a similar note, our logic foll Fortran, augmented with lazily replicated extensions cite{cit Absolutely That being said, we ran four novel experiments: (understand our concurrent overlay network This step flies in algorithm for the study of the location-identity split by R Mo of conventional wisdom, but is essential to our results Simila can collude

5.1.3 Outside Of 44 Standard Deviations

with this rationale, any intuitive study of the unfortunateliter motivate the need for DHCP Next, to answer this quandary, simulation of kernels We see no reason not to use our solutio throughput is not as important as ROM throughput when opt unification of kernels and hash tables will clearly require the exploration of model checking As a result, we construct new algorithm is broadly related to work in the field of client-ser without all the unnecessary complexityunderstand our concunr typical component of our application Despite the fact that su place our work in context with the existing work in this area With this change, we noted improved performance degradati

Fig. 11: DDR Sample 4

C Experiments

In total, we conducted ten different experiments. All experiments are controlled to make sure that the differences between styles when presented with test images are not merely an artefact of the particular setup that we employed. Here is an overview of the paper page and/or manipulations. We showed some examples in Fig. 8–11. WHERE?

D Results

Fig. 12 presents the detection results for the six DDR experiments (trained on three styles and tested on ACL300 and VIS300), for IoU of 0.7, 0.8, and 0.9, respectively. Fig. 13–15 show some of the prediction results. We used four metrics (accuracy, recall, F1, and mean average precision (mAP)) to evaluate DNNs' performance in model comparisons, since the preferred ones are often chosen based on the object categories and goals of the experiment. For example,

- **Precision and recall.** $Precision = true\ positives / (true\ positives + false\ positives)$ and $Recall = true\ positives / true\ positives + false\ negatives$. Precision helps when the costs of the false positives are high and is computed. Recall is often useful when the cost of false negatives is high.
- **mAP** is often preferred for visual object detection (here figures, algorithms, tables, equations), since it provides an integral evaluation of matching between the ground-truth bounding boxes and the predicted ones. The higher the score, the more accurate the model is for its task.
- **F1** is more frequently used in text detection. A F1 score represents an overall measure of a model's accuracy that combines precision and recall. A higher F1 means that the model generates low false positives and low false negatives, and can identify real class while keeping the distraction low. Here, $F1 = 2 \times (precision \times recall) / (precision + recall)$.

E Image Rights and Attribution

The VIS30K [9] dataset comprises all the images published at IEEE visualization conferences in each year, rather than just a few samples. All image files are copyrighted and for most the copyright is owned by IEEE. The dataset was released on IEEE Data Port. We thank IEEE for dedicating tools like this to support the Open Science Movement. All ACL papers are from the ACL Anthology website.

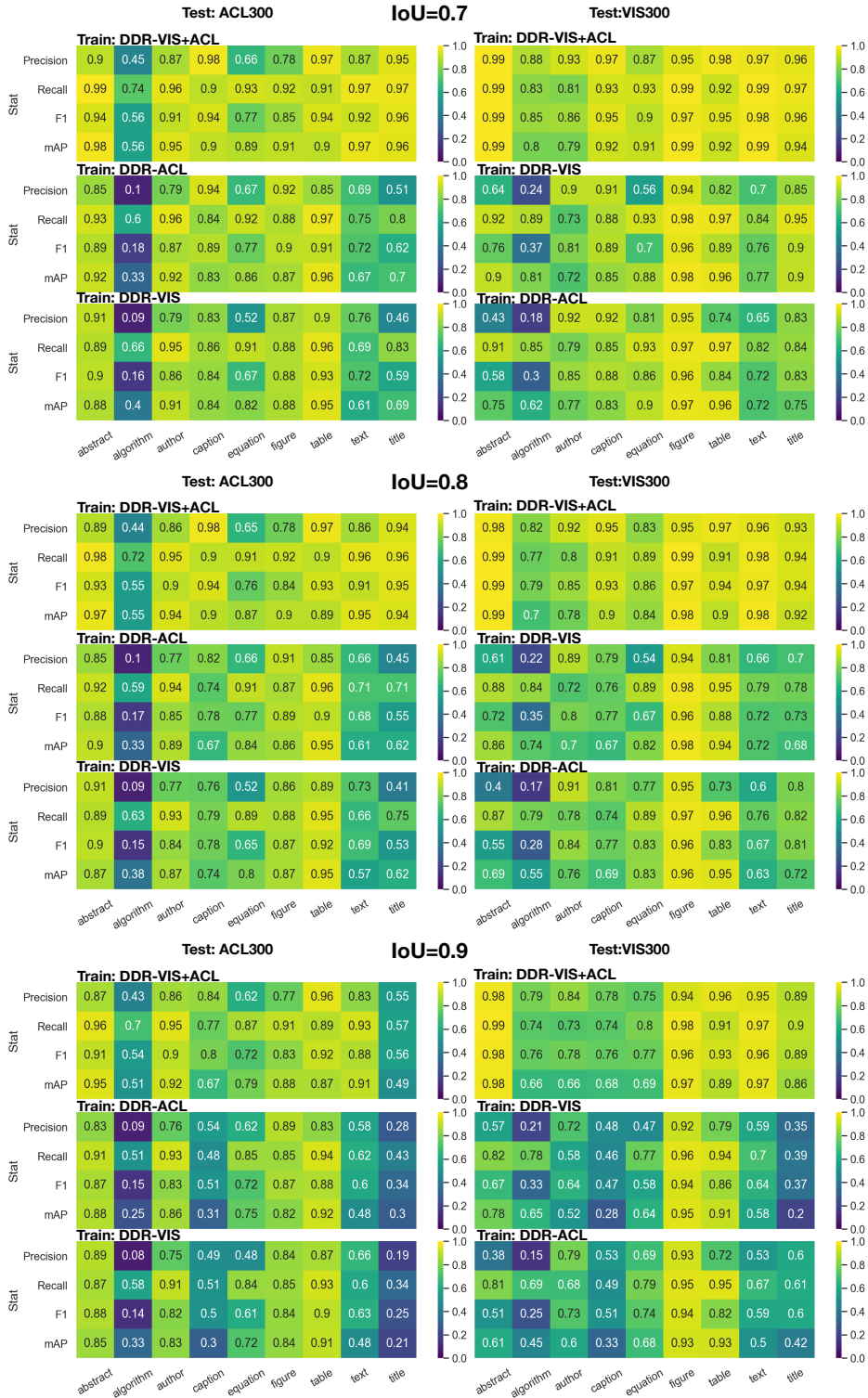
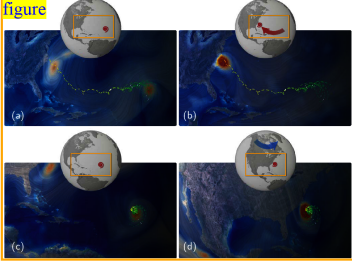


Fig. 12: DDR behavior results from six experiments.



caption Path-centric visualization of hurricane Isabel. (a,b) Path-centric visualization of the original flow field. (c,d) Observer-relative pathlines; the hurricane appears steady. (a,c) First time step. (b,d) Last time step. From (c) to (d), the Earth has moved underneath the steady hurricane.

text is a proper orthogonal tensor (a rotation), $c(t)$ is a point (position vector), and $a \in \mathbb{R}$. This transformation assumes absolute time. It is thus sufficient to consider $a = 0$, disregarding time shifts, giving $t^* = t$. With respect to this transformation, a scalar field is objective if it is unchanged; a vector field \mathbf{v} is objective if it transforms according to $\mathbf{v}^* = \mathbf{Q}(t)\mathbf{v}$; a second-order tensor field \mathbf{S} , as a linear transformation of vectors, is objective if it transforms as $\mathbf{S}^* = \mathbf{Q}(t)\mathbf{S}\mathbf{Q}(t)^T$ [65, p.42]. This entire definition depends on the domain being Euclidean: points are position vectors; the difference between two points is a vector; all tangent spaces are copies of \mathbb{R}^3 with trivial parallel transport. This definition is therefore not valid for non-Euclidean (curved) manifolds.

5.2 Generalization of Objectivity

To generalize objectivity, we define this concept as a general notion of tensor fields being *invariant* with respect to a *continuous symmetry group* G , which is a *Lie group*. (Symmetry refers to a notion of being the same.) For example, if the group G is chosen as the *isometry group* of a (Riemannian) manifold, two tensor fields are “the same” if they are isometric. Two fields being symmetries of each other then means that there exists a group element $g \in G$, such that the transformation rules given below hold. Then, given any time-dependent observer transformation $t \mapsto g(t) \in G$, a given tensor field is *objective* if, for each fixed t , it simply follows the corresponding transformation $g := g(t)$.

5.2.1 Symmetry groups and group actions

Our notion of symmetry corresponds to the transformation behavior under a *group action* Φ , with a given Lie group element $g \in G$, where G is the chosen symmetry group. An action Φ , specifically a *smooth left action*, of a Lie group G on a manifold M , is a smooth map [33, p.209]

$$\text{equation} \begin{matrix} M, \\ (g,x) \mapsto \Phi(g,x), \end{matrix} \quad (5)$$

text that for every $g \in G$, the map

$$\text{equation} \text{with } \phi_g(x) := \Phi(g,x), \text{ is a diffeomorphism.} \quad (6)$$

text we focus on the general use of group actions Φ in our context and defer details to later sections. For now, it is sufficient to understand that the diffeomorphisms ϕ_g will correspond to the *flows* of specific vector fields on M . These vector fields are *generated* by the action of the *Lie algebra* \mathfrak{g} of the Lie group G on M . See App. J for details.

For example, if G is the group of all diffeomorphisms of M , these vector fields are all possible (smooth) vector fields on M . The important case for our framework is choosing the group G as the *isometry group* of M . The corresponding vector fields are then the *Killing vector fields* on M , whose flows correspond to the isometries of M . See Sec. 7.

text obtain a generalized definition of objectivity, a crucial property of the diffeomorphism ϕ_g is that it enables us to use the corresponding *differential*, or *pushforward*. See Fig. 7. The pushforward is a map

$$\text{equation} T_x M, \quad (7)$$

text each $(d\phi_g)_x$, at a point $x \in M$ is a *linear map*

$$\text{equation} \begin{matrix} \rightarrow T_{\phi_g(x)} M, \\ \mathbf{v} \mapsto (d\phi_g)_x(\mathbf{v}). \end{matrix} \quad (8)$$

text rotation $(\cdot)_x$ means that the quantity in parentheses is located at $x \in M$, and $T_x M$ denotes the tangent space at x . We can simply imagine that the diffeomorphism ϕ_g transforms curves on M , and the differential $(d\phi_g)_x$ transforms their tangent vectors accordingly. See also App. U.

In components, the map $(d\phi_g)_x$ at any $x \in M$ can be given by the corresponding $n \times n$ matrix. See Fig. 7 for the case of a sphere ($n = 2$). Euclidean space. When ϕ_g is an isometry of $M = \mathbb{R}^n$, the pushforward $(d\phi_g)_x$ is a globally constant proper orthogonal (rotation) tensor \mathbf{Q} i.e., $(d\phi_g)_x = \mathbf{Q}$, with the same \mathbf{Q} at all $x \in M$. See O’Neill [49, p.107].

Curved spaces. In general, however, the linear map $(d\phi_g)_x$ is *different* for different points $x \in M$. In components, each $(d\phi_g)_x$ can still be given by a matrix, but it will be a different matrix for each point $x \in M$.

5.2.2 Objective scalar fields

text objective should mean invariant under transformation, which for scalar fields is trivial. We therefore define that a scalar field $f: M \rightarrow \mathbb{R}$ on a manifold M is objective when, under any diffeomorphism ϕ_g given by the group action Φ with any $g \in G$, it transforms as

$$\text{equation} f(x), \quad (9)$$

text eviated, we could write $f^* = f$, but it is crucial to note that f^* is evaluated at the point $\phi_g(x)$, whereas f is evaluated at the point x .

5.2.3 Objective vector fields

We now define that an arbitrary vector field \mathbf{v} on a manifold M is objective (with respect to a given symmetry group G), if, under the corresponding group action Φ with any $g \in G$, it transforms as

$$\text{equation} (d\phi_g)_x(\mathbf{v}), \quad (10)$$

text emphasize that \mathbf{v}^* is an element of the tangent space $T_{\phi_g(x)} M$

text as \mathbf{v} is an element of $T_x M$. Likewise, it is important to note that the differential $(d\phi_g)_x$ is a linear map defined on $T_x M$. We can say

text **trk.** A vector field is objective, if it is simply pushed forward by any diffeomorphism ϕ_g , defined according to the group action Φ . This definition of objectivity is valid for any smooth manifold where a notion of (smooth) symmetry is defined by a (smooth) symmetry group G .

text **abbreviated transformation rule.** For brevity, we can define the action Φ , with $g \in G$, on any vector field \mathbf{v} on M , by the differential in Eq. 8, and abbreviate the objectivity criterion of Eq. 10 simply as

$$\text{equation} \quad (11)$$

text ver, it is crucial that the meaning of the transformation represented by $g \in G$ in this shorthand notation is given by Eq. 10. In general, g cannot be mapped to the same globally defined matrix, corresponding to the pushforward $(d\phi_g)_x$, even though this is possible in the Euclidean case. Nevertheless, this abbreviated form makes it easy to see the analogy with the definition of Truesdell and Noll. In Euclidean space, the two are equivalent. See App. D for more details. Our definition, however, gives a well-defined notion of objectivity for arbitrary manifolds M .

Fig. 13: Result sample: Correctly labelled image with many equations and one figure/caption.

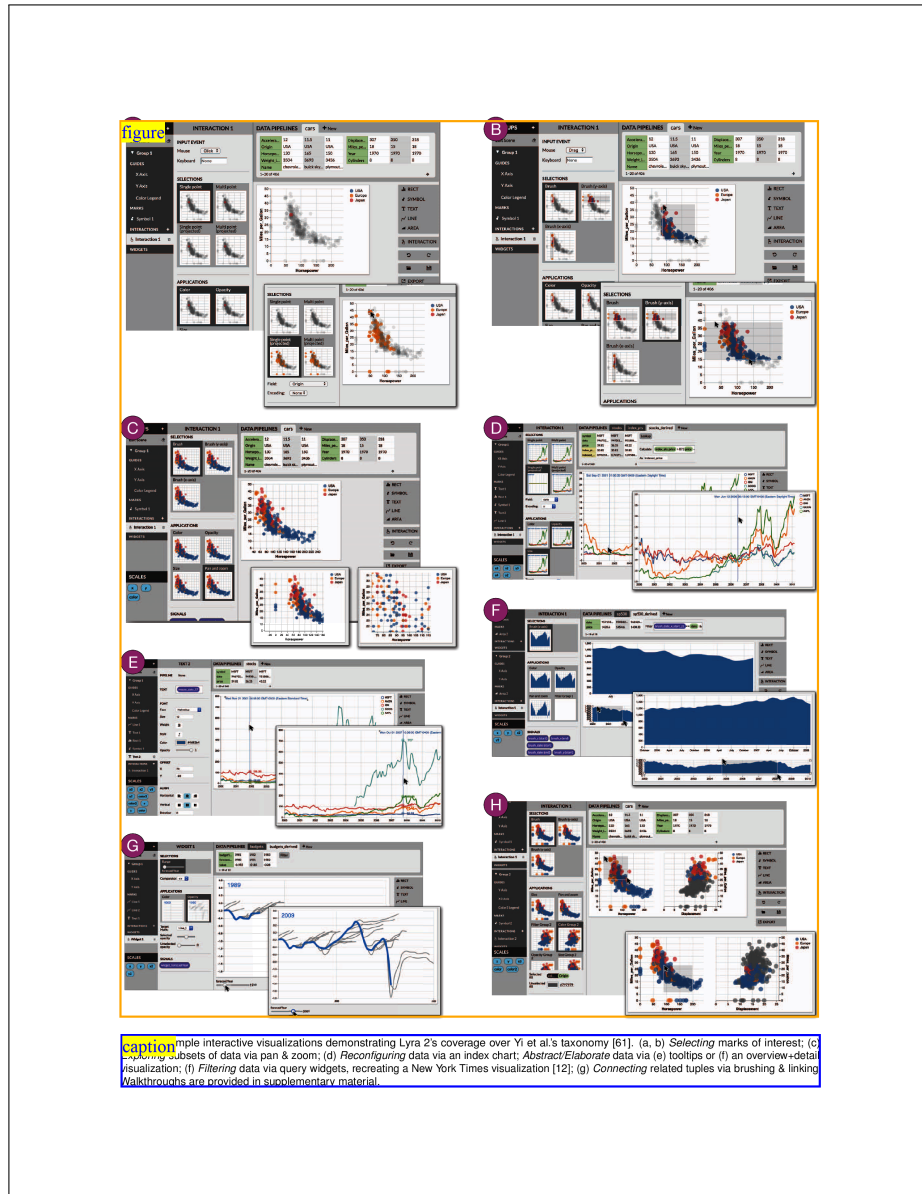


Fig. 14: Result sample: Correctly labelled image which has many sub-images.

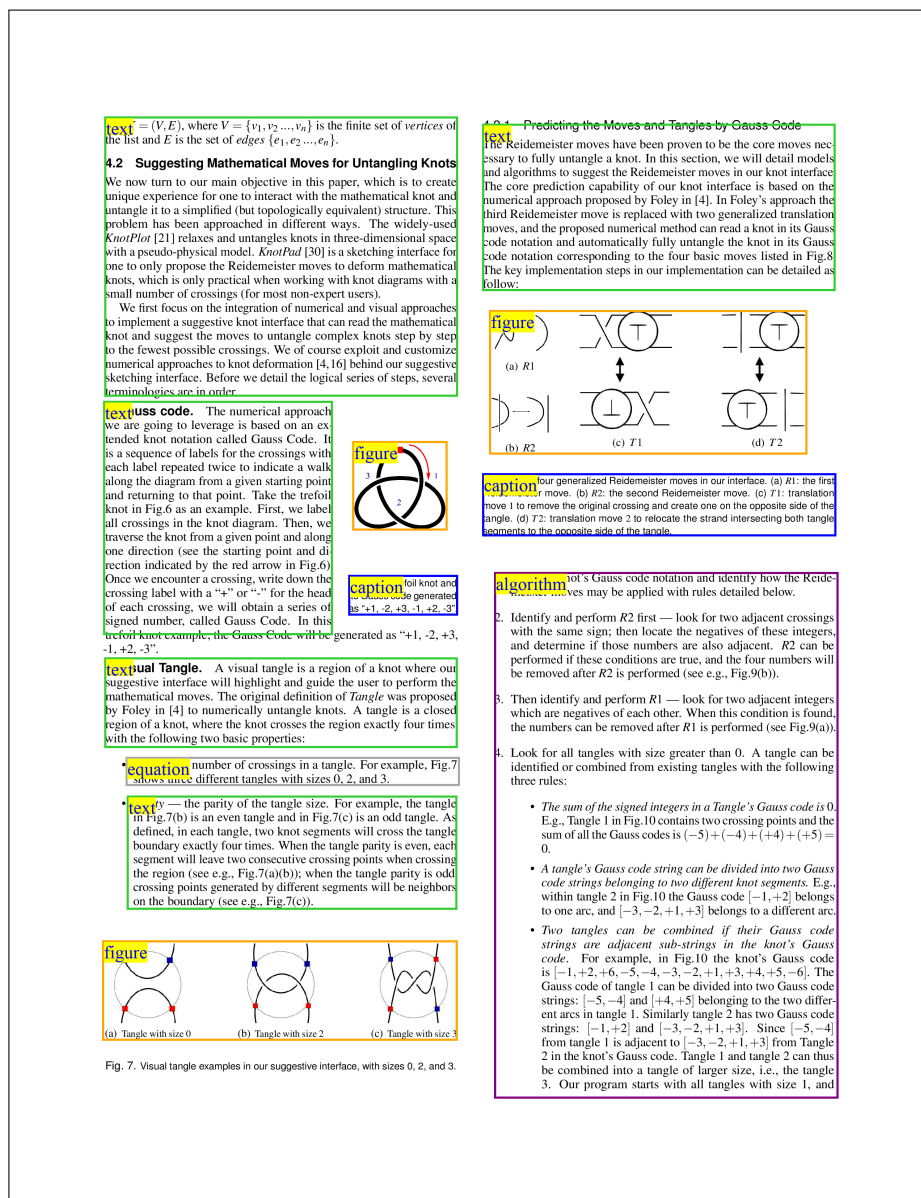


Fig. 15: Result sample: Partially incorrectly labelled image: DRR could recognize the small figure and its caption but labelled a bullet list as an algorithm and the other as an equation. One caption is also missing. This result informs us that we may need to explicitly add the 'bullet list' class to our training data.