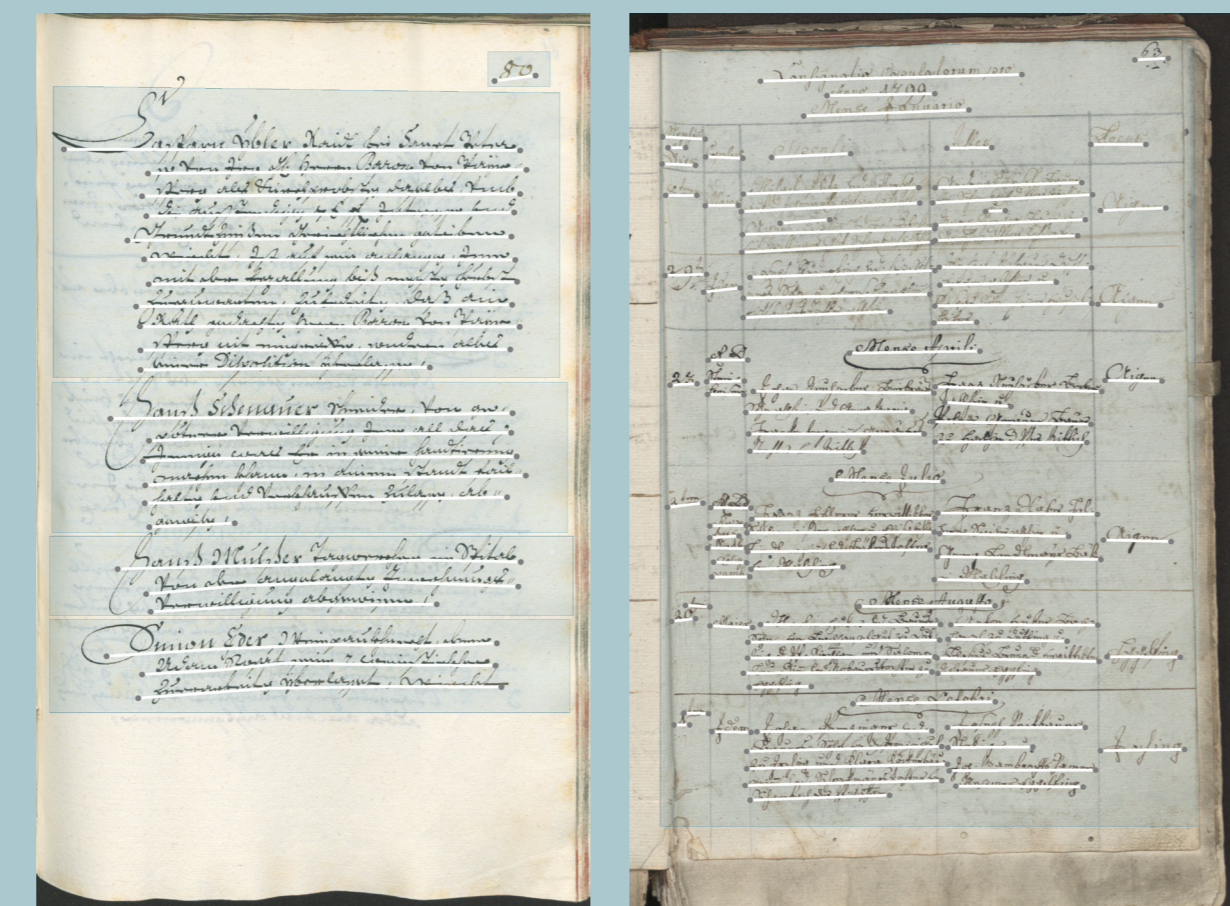


cBAD

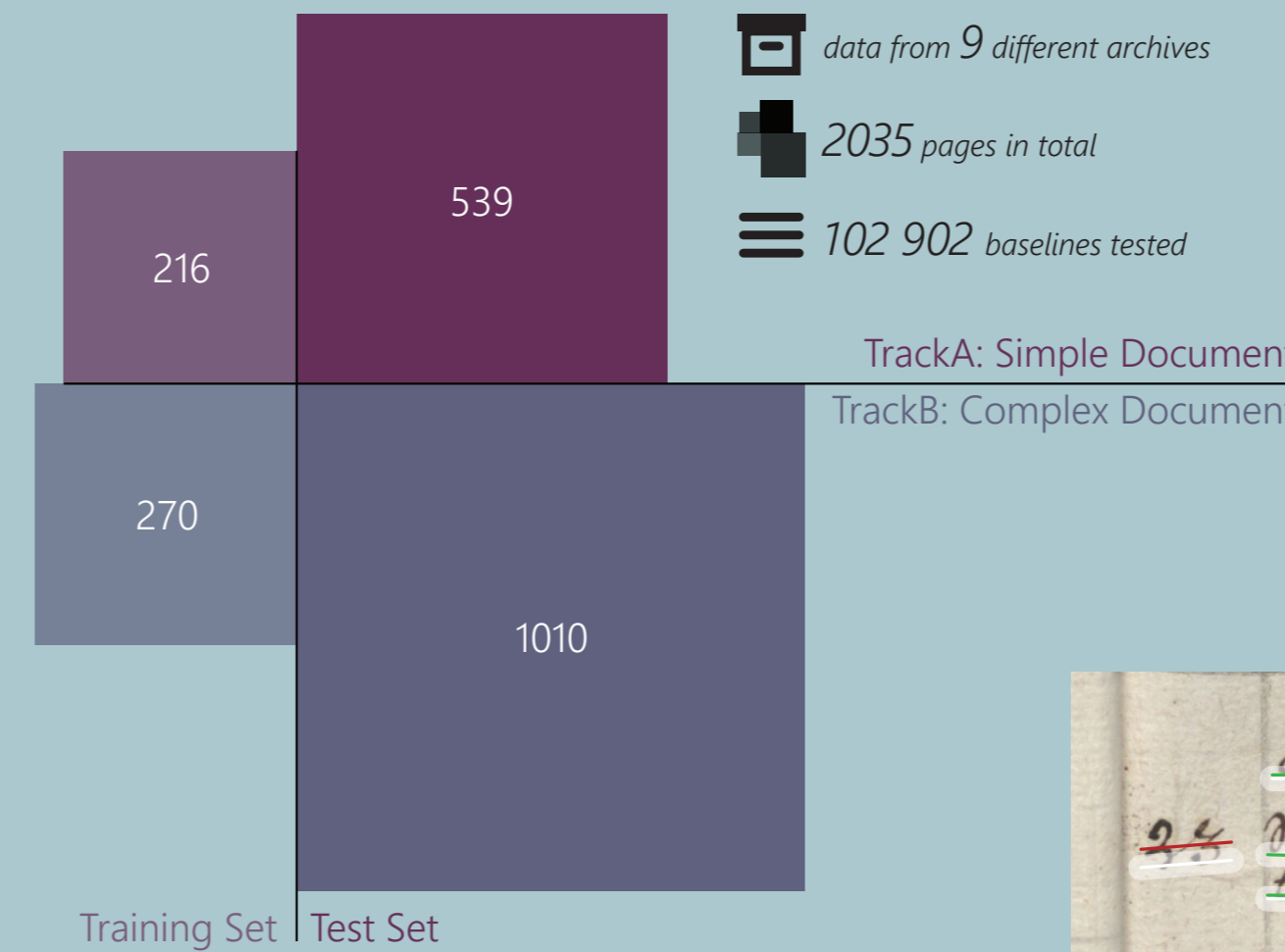
THE COMPETITION

The ICDAR2017 competition on Baseline Detection (cBAD) dataset consists of 2035 document page images that were collected from 9 different archives. Since traditional evaluation schemes are not applicable to the size and modality of this dataset, we present a new one, that introduces baselines to measure performance.

Markus Diem, Florian Kleber, Stefan Fiel, Tobias Grüning, and Basilis Gatos



Track A: Simple Documents Track B: Complex Documents



data from 9 different archives
2035 pages in total
102 902 baselines tested

TrackA: Simple Documents
TrackB: Complex Documents

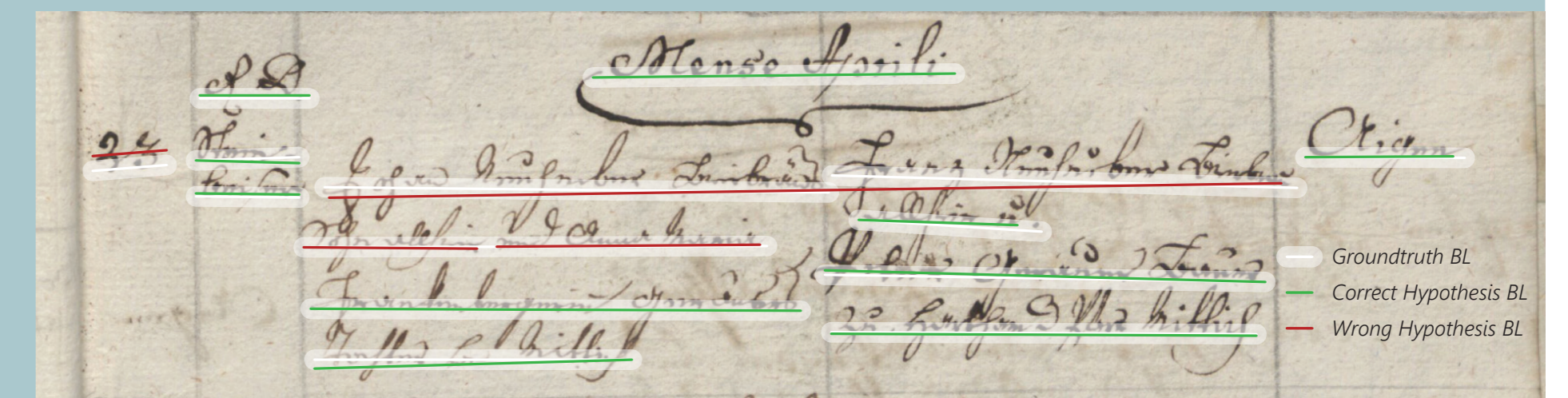
NEW EVALUATION SCHEME

$$P = \frac{\sum_{(g,h) \in \mathcal{M}(G,H)} COV(h,g)}{|\mathcal{H}|}$$

$$R = \frac{\sum_{g \in G} COV(g,H)}{|G|}$$

$$F = \frac{2RP}{R+P}$$

G Groundtruth BL
 H Hypothesis BL
 COV Coverage Function
 $\mathcal{M}(G,H)$ Alignment Function



THE DATASET

PARTICIPANTS

The competition was carried out using the ScriptNet platform [4]. Teams could download the training set along with GT and the images of the test set. For evaluation, participants uploaded the resulting XMLs (one per image) which were directly evaluated on ScriptNet. Methods of five different teams were submitted for Track A and four teams submitted to Track B.

BYU

Chris Tensmeyer, Brian Davis, and Curtis Wigington
Dept. of Computer Science, Brigham Young University, Provo, USA

IRISA

Aurélie Lemaitre, Jean Camillerapp, and Bertrand Couasnon
IRISA - University Rennes 2 and Insa Rennes, France

DMRZ

Georg Mackenbrock, Michael Fink, Thomas Layer, and Michael Sprinzl
Deutsches Medizinrechenzentrum GmbH & Co KG, Vienna, Austria

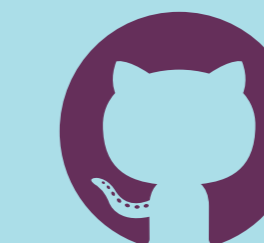
LITIS

Guillaume Renton, Clément Chatelain, Sébastien Adam, Christopher Kermorvant, and Thierry Paquet
Normandie Univ, UNIROUEN, UNIHAVRE, INSA Rouen, LITIS, 76000 Rouen, France

UPVLC

Moisés Pastor and Lorenzo Quirós
PRHLT research centre. Universitat Politècnica de València

DATASET
The annotated dataset is publicly available at Zenodo.
<https://zenodo.org/search?page=1&size=20&q=cbad>



EVALUATION TOOL
The benchmarking tool and its source code are available on github.
<https://github.com/Transkribus/TranskribusBaselineEvaluationScheme>



EVALUATION SCHEME
T. Grüning, R. Labahn, M. Diem, F. Kleber, and S. Fiel, "READ-BAD: A new dataset and evaluation scheme for baseline detection in archival documents," CoRR, vol. abs/1705.03311, 2017.

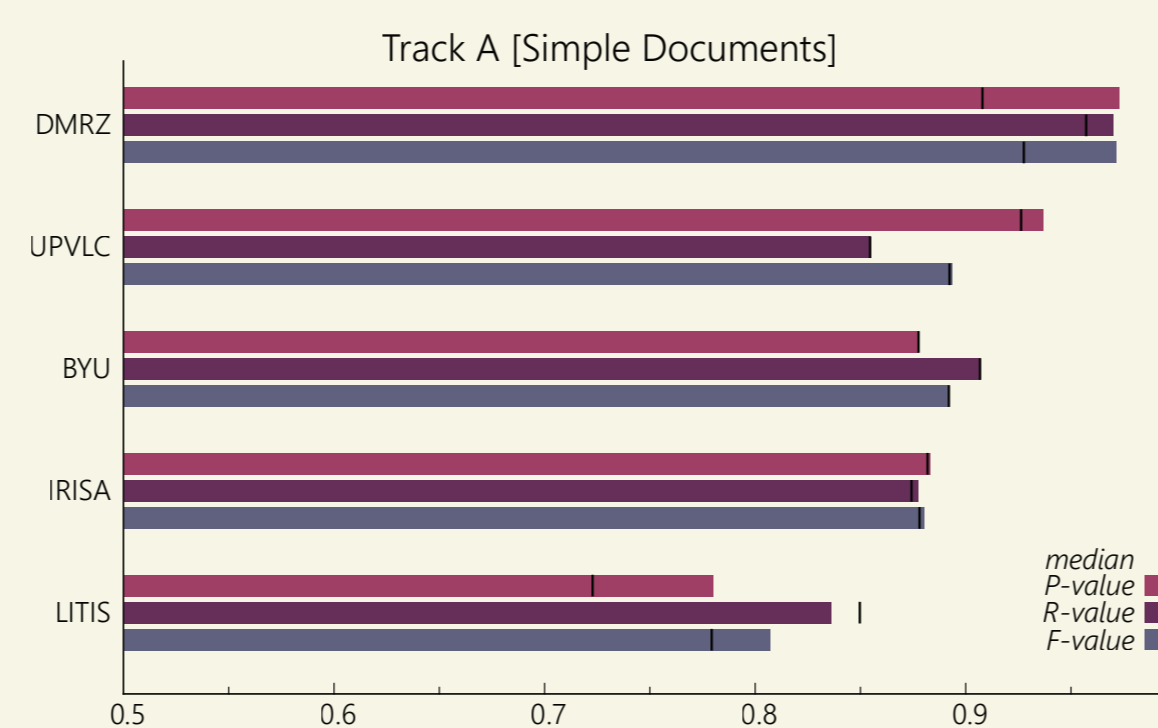


PLATFORM
The competition results are available on ScriptNet.
<https://scriptnet.lit.demokritos.gr/competitions/5/1/>

RESOURCES

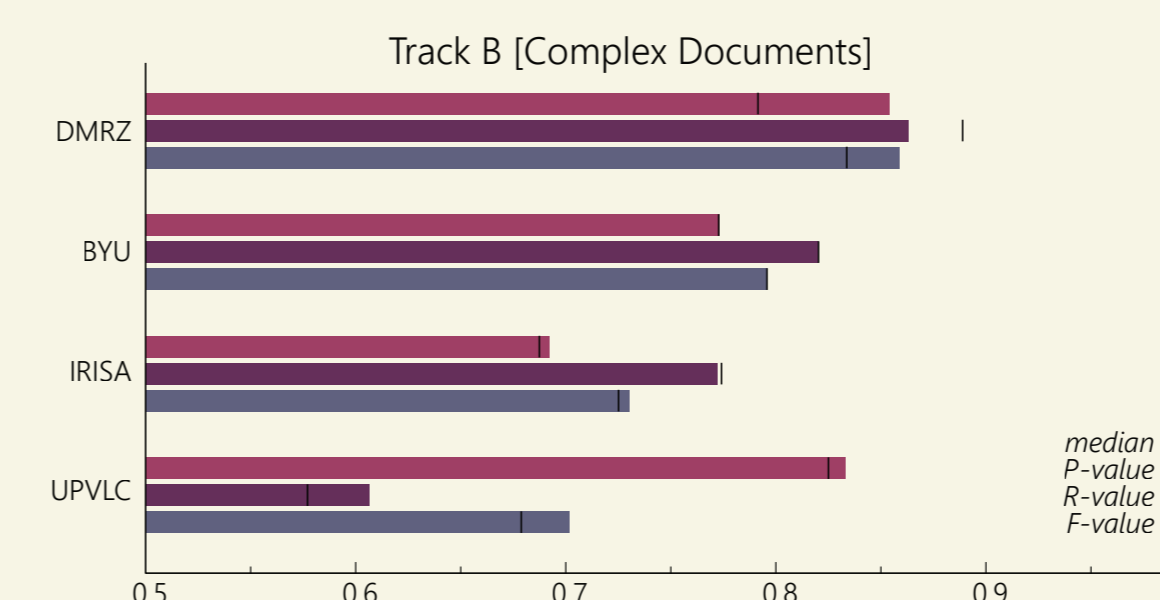
RESULTS

The evaluation was carried out with the new baseline evaluation scheme. The median F-value of all submissions for Track A is 0.89 and 0.76 for Track B. Hence, state-of-the-art baseline detection methods achieve decent results on historical documents given that the text regions are segmented. If complex layout variations are present, baseline detection is still a challenging task.



Method	P-value	R-value	F-value
DMRZ	0.973	0.970	0.971
UPVLC	0.937	0.855	0.894
BYU	0.878	0.907	0.892
IRISA	0.883	0.877	0.880
LITIS	0.780	0.836	0.807

TRACK A SIMPLE DOCUMENTS



TRACK B COMPLEX DOCUMENTS

Method	P-value	R-value	F-value
DMRZ	0.854	0.863	0.859
BYU	0.773	0.820	0.796
IRISA	0.692	0.772	0.730
UPVLC	0.833	0.606	0.702

