Pattern Recognition Lab

Friedrich-Alexander-Universität
Technische Fakultät

A Fair Evaluation of Various Deep Learning-based Document Image Binarization Approaches

Richin Sukesh, Mathias Seuret, Anguelos Nicolaou, Martin Mayr, Vincent Christlein

Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

Basic Principle

Motivation

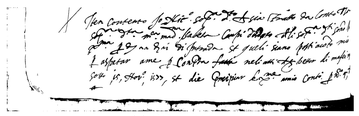
- Binarization of document images is an important pre-processing step in the field of document analysis
- Learning-based approaches often incorporate extra data for training

➔ No fair comparison

Idea: Evaluation of different deep learning-based binarization methods using

- same evaluation protocol
- same training data

Failure Examples



(a) DP-LinkNet

(b) 2-Stage GAN

Experimental Setup

Datasets

- Training: DIBCO2009, DIBCO2010, DIBCO2011, DIBCO2012, DIBCO2014 and DIBCO2016
- Evaluation: DIBCO2013, DIBCO2017, DIBCO2018 and DIBCO2019 [1]

Metrics

- F-measure: $\mathrm{FM} = \frac{2\times\mathrm{Recall}\times\mathrm{Precision}}{\mathrm{Recall}\times\mathrm{Precision}}$
- Pseudo F-measure: $\mathrm{pFM} = \frac{2\times\mathrm{pRecall}\times\mathrm{pPrecision}}{\mathrm{pRecall}\times\mathrm{pPrecision}}$
- Peak signal to noise ratio: $\mathrm{PSNR} = \log_{10}\left(\frac{C^2}{\mathrm{MSE}}\right)$
- Distance reciprocal distortion: $\mathrm{DRD} = \frac{\sum_k \mathrm{DRD}_k}{\mathrm{NUBN}}$

Training Protocol

- All training datasets are used
- Augmentation: horizontal & vertical flipping and rotations
- Hyper-parameters optimized using “optuna” [2]
- Early stopping based on validation performance

Code and models available below:
`https://github.com/RichSu95/Document_Binarization_Collection`

Evaluation

Table 1: Average over (a) all metrics and (b) all ranks. Runtime: NVIDIA RTX 2060 GPU (12 GB RAM), note: DeepOtsu and 2-Stage GAN were limited by available memory.

Model	PSNR↑	FM↑	pFM↑	DRD↓	img/sec↑
DE-GAN	18.37	85.25	87.54	4.64	0.67
Robin (U-Net)	18.24	82.87	84.08	5.95	1.99
DeepOtsu	16.69	79.97	81.89	13.96	0.01
2-Stage GAN	19.07	**87.20**	**88.39**	**4.54**	0.01
DP-LinkNet	**19.10**	83.70	84.36	5.79	0.49
SAE	16.15	79.14	80.77	9.41	0.68
SauvolaNet	18.67	84.91	86.83	4.86	0.37

(a) Average metrics

Avg. rank↓
2.44
4.19
5.50
3.25
3.38
6.63
2.63

(b) Average ranks

- Performance dependent on different datasets
- Best average rank: DE-GAN [3] followed by SauvolaNet [4]
- Best metrics: 2-Stage GAN [5] followed by DE-GAN

Conclusion

- No clear winner
- Optimal binarization approach is dependent on the underlying data, amount, and quality of training samples

Qualitative Results

(a) Input image (b) Ground Truth (c) DE-GAN (d) Robin (U-Net) (e) DeepOtsu (f) 2-Stage GAN (g) DP-LinkNet (h) SAE (i) SauvolaNet

Figure 2: Illustration of some results for an image from DIBCO-2017. Pixels in cyan are false positives. The few pixels in orange are false negatives. Pixels in white or black match the ground truth.

References

[1] I. Pratikakis, K. Zagoris, X. Karagiannis, L. Tsochatzidis, T. Mondal, and I. Marthot-Santaniello, “Icdar 2019 competition on document image binarization (dibco 2019),” in *2019 International Conference on Document Analysis and Recognition (ICDAR)*, pp. 1547–1556, 2019.

[2] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, “Optuna: A next-generation hyperparameter optimization framework,” in *25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD ’19, p. 2623–2631, 2019.

[3] M. A. Souibgui and Y. Kessentini, “De-gan: A conditional generative adversarial network for document enhancement,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, pp. 1180–1191, March 2022.

[4] D. Li, Y. Wu, and Y. Zhou, “Sauvolanet: Learning adaptive sauvola network for degraded document binarization,” in *The 16th International Conference on Document Analysis and Recognition (ICDAR)*, p. 538–553, 2021.

[5] S. Suh, J. Kim, P. Lukowicz, and Y. O. Lee, “Two-stage generative adversarial networks for document image binarization with color noise and background removal,” *CoRR*, vol. abs/2010.10103, 2020.

Contact

Vincent Christlein
Computer Vision
Friedrich-Alexander-Universität Erlangen-Nürnberg
Erlangen, Germany
☎ +49 9131 85-20281
✉ vincent.christlein@fau.de