UDBNET: Unsupervised Document Binarization Network via Adversarial Game

Amandeep Kumar[a*], Shuvozit Ghose[b*], Pinaki Nath Chowdhury[c], Partha Pratim Roy[d], Umapada Pal[c]

(*Equal Contribution) [a]Techno Main Salt Lake, Sector V, Kolkata, India.
[b]Institute of Engineering and Management, Kolkata India.
[c]Indian Statistical Institute, Kolkata, India. [d]Indian Institute of Technology, Roorkee, India.

Presenter: Pinaki Nath Chowdhury

Venue: International Conference on Pattern Recognition, 2020

Contents:

- Challenges in Document Image Binarization
- Supervised Binarization Methods
- Unsupervised Binarization method Bhunia et al. [1]
- Limitations of Bhunia et al. [1]
- Motivation
- An overview of our Solution
- Adversarial Texture Augmentation Network
- Unsupervised Document Binarization Network
- Results
- Conclusion

Challenges in Document Image Binarization

- Degradation due to faint characters, bleed-through background, clutter and artifacts, dark patches, creases, faded ink, non-uniform variation of intensity, inadequate maintenance, aging effect, ink stains, lighting conditions, warping effect during acquisition etc.
- Faded ink creates difficulty during distinguishing light text from background
- Bleed through occurs when content from the back of a page becomes visible or leaks through
- Dark patches are quite difficult to remove due to varying sizes, intensities and shapes

Supervised Binarization Methods

- Binarization methods works for supervised setup.
- In the supervised setup, we need ground truth binarized image along with the degraded image.
- But, it is difficult to get the corresponding ground truth binary image in many scenarios like in case of historical document image.

Unsupervised Binarization method Bhunia et al. [1]

- Although state-of-the-art binarization methods works for supervised setup, Bhunia et al. [1] first attempts to introduce unsupervised setup in the domain of document image binarization.
- In the supervised setup, we need ground truth binarized image along with the degraded image.
- But, it is difficult to get the corresponding ground truth binary image in many scenarios like in case of historical document image.

Limitations of Bhunia et al. [1]

- The TANet is completely unaware about the content at which it is conditioned on. Thus, the corresponding discriminator can not verify if the content of the generated noisy image remain consistent or not.
- There exist no performance quantifier that validates the performance of the BiNet on real degraded noisy image.
- the Binarization Network (BiNet) has dataset bias towards generated noisy images. But, to address the dataset bias, BiNet does not use any kind of formulation or other techniques.

Motivation

In our observation, these limitations are due to the fact that the TANet and BiNet both employ straight-forward two-player Generative Adversarial Network (GAN) objectives and model two different uncorrelated conditional distributions. We address these limitations by introducing adversarial minmax game in the domain of unsupervised document image binarization.

An overview of our Solution

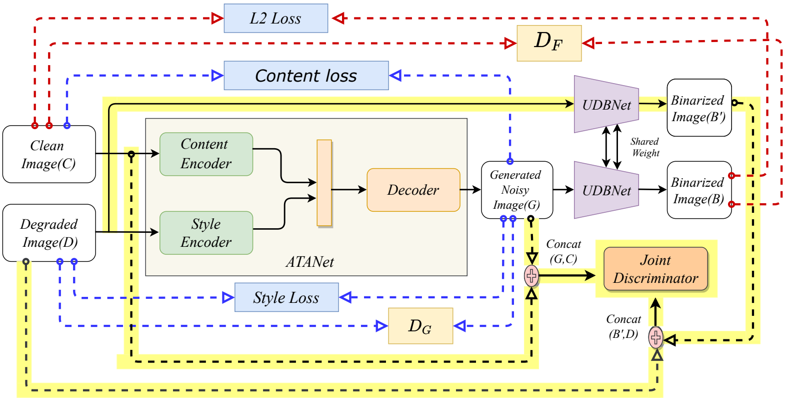

Figure: Illustration of Our proposed Framework. The yellow highlighted region highlights our contribution over Bhunia et al. [1].

An overview of our Solution (continued...)

- We introduce adversarial minmax game in the domain of unsupervised document image binarization proposing Adversarial Texture Augmentation Network (ATANet) and Unsupervised Documenet Binarization Network (UDBNet) which utilize three-player GAN objectives.
- The proposed third player is a joint discriminator tries to couple both the Adversarial Texture Augmentation Network (ATANet) and Unsupervised Document Binarization Network (UDBNet).

Adversarial Texture Augmentation Network

- a generator T that characterizes the conditional distribution $P_T(G|C, D)$ and generates noisy image G;
- a discriminator $\mathrm{D_T}$ that discriminates the output image G from the degraded reference image D;
- a joint discriminator $\mathrm{J_D}$ that distinguishes whether a pair of data (G, C) comes from $P_T(C, D)$ or $P_B(C, D)$.

we define adversarial loss of our ATANet as:

$$\begin{aligned} \underset{\mathrm{D_T}}{min}\, \underset{\mathrm{T,J_D}}{max}\, \mathcal{L}_T^{Adv}(\mathrm{D_T}, \mathrm{T}, \mathrm{J_D}) = \mathbb{E}_{D\sim P_D}[\log \mathrm{D_T}(D)] + \\ \mathbb{E}_{(C)\sim P(C),(D)\sim P(D)}[\log(1 - \mathrm{D_T}(\mathrm{T}(C, D))] + \\ \mathbb{E}_{(C)\sim P(C),(D)\sim P(D)}[\log(1 - \mathrm{J_D}(\mathrm{T}(C, D), C)] + \\ \mathbb{E}_{(D)\sim P(D)}[\log(\mathrm{J_D}(\mathrm{F}(D), D)] \end{aligned} \tag{1}$$

Adversarial Texture Augmentation Network (Continued....)

The overall objective of the ATANet is defined as:

$$\mathcal{L}^{ATANet} = \mathcal{L}_T^{Adv}(D_T, T, J_T) + \lambda_s \mathcal{L}^s(T) + \lambda_c \mathcal{L}^c(T) \qquad (2)$$

Where, $\mathcal{L}^s(T)$ and $\mathcal{L}^c(T)$ are style loss and content loss similar to our base mode, λ_s and λ_c are the tunable hyper-parameters to balance multiple objectives.

Unsupervised Document Binarization Network

- a generator F that characterizes the conditional distribution $P_B(B|G)$ and $P_B(B'|D)$ generates binarized clean image B and B' corresponding to G and D respectively;
- a discriminator $\mathrm{D_F}$ determines how good the generator is in generating binarized images B;
- a joint discriminator $\mathrm{J_D}$ that distinguishes whether a pair of data (B', D) comes from distribution $P_B(C, D)$ or $P_T(C, D)$.

We define adversarial loss of our UDBNet as:

$$\begin{aligned} \min_{\mathrm{D_F}} \max_{\mathrm{F,J_D}} \mathcal{L}_F^{Adv}(\mathrm{D_F}, \mathrm{F}, \mathrm{J_F}) = \mathbb{E}_{C\sim P_C}[\log \mathrm{D_F}(C)]+ \\ \mathbb{E}_{G\sim P(D|C)}[\log(1-\mathrm{D_F}(\mathrm{F}(G))]+ \\ \mathbb{E}_{(D)\sim P(D)}[\log(1-\mathrm{J_D}(\mathrm{F}(D), D)]+ \\ \mathbb{E}_{(C)\sim P(C),(D)\sim P(D)}[\log(\mathrm{J_D}(\mathrm{T}(C, D), C)] \end{aligned} \tag{3}$$

Unsupervised Document Binarization Network (Continued....)

The overall objective of the UDBNet is defined as:

$$\mathcal{L}^{UDBNet} = \mathcal{L}_T^{Adv}(D_T, T, J_T) + \lambda_{L2}\mathcal{L}^{L2}(F) \tag{4}$$

Where, $\mathcal{L}^{L2}(F)$ is L_2 loss, λ_{L2} is a tunable hyper-parameter.

Results

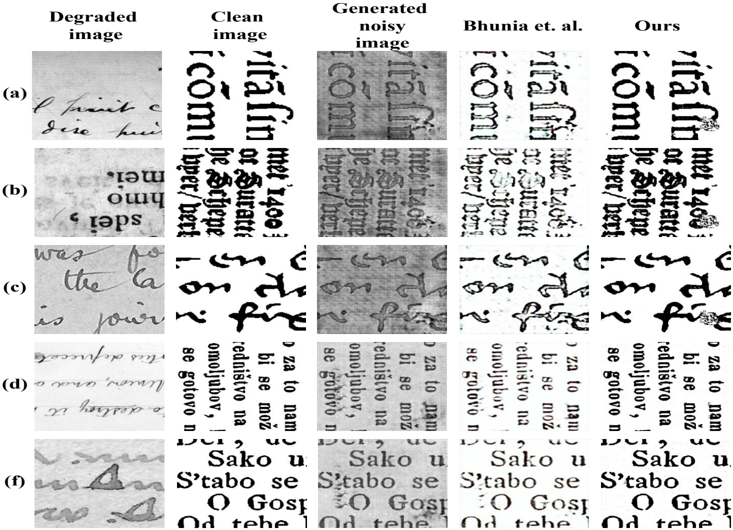

Figure: Comparison of the qualitative results of predicted binarized images by Bhunia [1] and our framework on the evaluation set

Results (continued...)

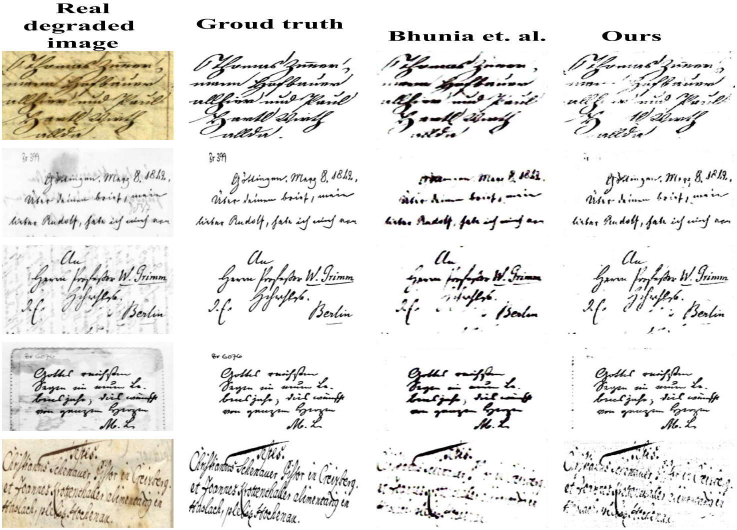

Figure: Binarization results on real test images by passing through UDBNet.

Results (continued...)

Table: Comparsion of Our method with Baseline Methods

Methods	**F-Measure**	F_{PS}	**PSNR**	**DRD**
UDBNet-CL	92.7	95.8	19.9	2.6
UDBNet-GRL	93.2	96.0	20.1	2.4
Ours	**93.4**	**96.2**	**20.1**	**2.2**

Results (continued...)

Table: Quantative results on DIBCO 2011 dataset

Methods	**DIBCO 2011 Dataset**			
	F-Measure	F_{PS}	**PSNR**	**DRD**
Otsu [2]	82.1	84.8	15.7	9.0
Sauvola [3]	82.1	87.7	15.6	8.5
Howe [4]	91.7	92.0	19.3	3.4
Su [5]	87.8	90.0	17.6	4.8
Jia [6]	91.9	95.1	19.0	2.6
Vo [7]	88.2	90.3	20.1	2.9
Vo [8]	93.3	96.4	20.1	2.0
DeepOtsu [9]	93.4	95.8	19.9	1.9
Bhunia [1]	93.7	96.8	20.1	1.8
Ours	**95.2**	**97.9**	**20.4**	**1.5**

Results (continued...)

Table: Quantative results on H-DIBCO 2016

Methods	**H-DIBCO 2016 Dataset**			
	F-Measure	F_{PS}	**PSNR**	**DRD**
Otsu [2]	86.6	89.9	17.8	5.6
Sauvola [3]	84.6	88.4	17.1	6.3
Howe [4]	87.5	92.3	18.1	5.4
Su [5]	84.8	88.9	17.6	5.6
Jia [6]	90.5	93.3	19.3	3.9
Vo [7]	87.3	90.5	17.5	4.4
Vo [8]	90.1	93.6	19.0	3.5
Westphal [10]	88.8	92.5	18.4	3.9
DeepOtsu [9]	91.4	94.3	19.6	2.9
Bhunia [1]	92.3	95.4	19.9	2.7
Ours	**93.4**	**96.2**	**20.1**	**2.2**

Conclusion

- We have proposed a novel approach towards document binarization by introducing three-player min-max adversarial game.
- A joint discriminator which tries to couple the Adversarial Texture Augmentation Network (ATANet) and Unsupervised Document Binarization Network (UDBNet) so that it can tackle the dataset bias problem and perform well on the real degraded document image.
- The effectiveness of our system by conducting experiments on publicly available DIBCO datasets.

References I

- A. K. B. et al., “Improving document binarization via adversarial noise-texture augmentation,” in *ICIP*, 2019.
- N. Otsu, “A threshold selection method from gray-level histograms,” *IEEE-TSMC*, 1979.
- J. Sauvola and M. Pietikäinen, “Adaptive document image binarization,” *Pattern recognition*, 2000.
- N. Howe, “Document binarization with automatic parameter tuning,” *IJDAR*, 2013.
- B. S. et al., “Binarization of historical document images using the local maximum and minimum,” in *DAS*, 2010.
- F. J. et al., “Degraded document image binarization using structural symmetry of strokes,” *Pattern Recognition*, 2018.

References II

- G. D. Vo and C. Park, “Robust regression for image binarization under heavy noise and nonuniform background,” *Pattern Recognition*, 2018.
- Q. V. et al., “Binarization of degraded document images based on hierarchical deep supervised network,” *Pattern Recognition*, 2018.
- S. He and L. Schomaker, “Deepotsu: Document enhancement and binarization using iterative deep learning,” *Pattern Recognition*, 2019.
- F. W. et al., “Document image binarization using recurrent neural networks,” in *DAS*, 2018.

Thank you
Questions?

Source Code is available at:
https://github.com/VIROBO-15/UDBNET