

A Gated and Bifurcated Stacked U-Net for Document Image Dewarping

Hmrishav Bandyopadhyay, Tanmoy Dasgupta, Dr Nibaran Das, Dr Mita Nasipuri

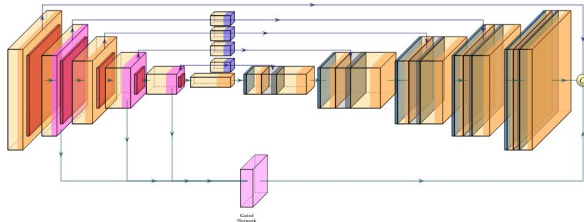
Abstract

Capturing images of documents is one of the easiest and most used methods of recording them. These images however, being captured with the help of handheld devices, often lead to undesirable distortions that are hard to remove. We propose a supervised Gated and Bifurcated Stacked U-Net module to predict a dewarping grid and create a distortion free image from the input.

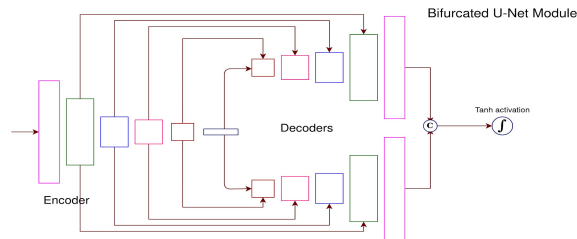
Novelty:

- A) The parameters are learned efficiently and as such the model learns in just 8 thousand images, less than one-tenth of the dataset sizes in previous end-to-end methods like DocUNet[1] and DewarpNet [2].
- B) We propose a bifurcated U-Net as the secondary U-Net of our stacked U-Net system to help in channel level segregation while predicting dense grid unwarps.
- C) A gated branch of the primary U-Net is proposed following [3] that enables the secondary U-Net to recognize overall structure in the warped document image.

Primary U-Net:



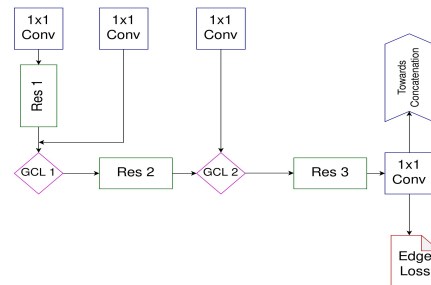
Split Decoder



Motivation behind split

The general CNN works by summing up computed data across all input channels for specific window sizes. This summation results in a merging of the data from multiple channels together into a single 2-dimensional vector and then using the merged data in the later stages. We, however, observe that the dense grid does not favor to the intermingling of channel data, resulting in poor model performance. To get round this issue, we came up with the usage of multiple decoder blocks for the single secondary U-Net encoder.

Gated Network:



A Gated and Bifurcated Stacked U-Net for Document Image Dewarping

Hmrishav Bandyopadhyay, Tanmoy Dasgupta, Dr Nibaran Das, Dr Mita Nasipuri

Results



SSIM:

Level	Our Method	DewarpNet [2]
Original Resolution	0.548915	0.493146
2	0.467136	0.433653
3	0.39162	0.387747
4	0.332977	0.369569
5	0.302610	0.464170
6	0.387984	0.575128
7	0.504144	0.607561
8	0.560574	0.586102
9	0.541162	0.546075

MS-SSIM & LD

Method	MS-SSIM	LD
Tian <i>et. al.</i> [4]	0.13*	33.69
DocUNet [1]	0.410*	14.08
Our Method	0.415	13.2
DewarpNet[2]	0.437	8.98

SSIM (graphical)

References

- [1] X. B. J. W. D. S. Ke Ma, Zhixin Shu, "Docunet: Document image unwarping via a stacked u-net," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2018..
- [2] S. Das, K. Ma, Z. Shu, D. Samaras, and R. Shilkrot, "Dewarper: Single-image document unwarping with stacked 3d and 2d regression networks," in Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 131–140
- [3] R. Liu, J. Lehman, P. Molino, F. P. Such, E. Frank, A. Sergeev, and J. Yosinski, "An intriguing failing of convolutional neural networks and the coordconv solution," CoRR, vol. abs/1807.03247, 2018
- [4] Y. Tian and S. G. Narasimhan, "Rectification and 3d reconstruction of curved document images," in CVPR 2011. IEEE, 2011, pp. 377–384.

