On-Device Text Image Super Resolution

Authors: Dhruval Jain, Arun D Prabhu, Gopi Ramena, Manoj Goyal, Debi Prasanna Mohanty, Sukumar Mohrana, Naresh Purre

Introduction

- Compression techniques used by social media platforms significantly reduce the resolution of images.
- Low Resolution text images poses a challenge for existing text detection and recognition frameworks on device.
- We can employ a super resolution (SR) network to increase the resolution of the images and consequently, OCR confidence. This will make the image more readable to users.





Compressed Image



Original Image

Image after SR

Motivation for Text Super Resolution

- The current (SR) frameworks leverage deep neural networks which are infeasible on-device.
- Deeper networks are required for learning textures in low resolution (LR) images and reproducing it in high resolution (HR) images.
- Increasing the sharpness of outline edges is much more important than reconstruction of texture details to improve OCR confidence.

Network Architecture



Network Architecture

Our SR model consists of the following components:

- 1. <u>Feature Extraction</u>: The feature extractor consists of four convolutional layers. All the feature maps except the third layer are concatenated via skip connections to generate a total of 64 feature maps. We also extract features from sobel edge maps using a conv layer in tandem to improve the edges in HR image.
- 2. <u>Upsampling</u>: The extracted features are passed through 2 convolutional layers followed by a pixel shuffler.

Edge Reconstruction

Input Canny Edge Map



Our result



Bicubic Upsampling









Results

Method	Scale	Metrics	Datasets							
			TextSR	IC13	IC15	IC03	SVT	IIIT5K	CUTE80	MLT
Bicubic	2	PSNR	23.50	31.91	30.58	27.57	36.60	23.78	34.2	36.55
		SSIM	0.879	0.9017	0.8983	0.8588	0.9570	0.8087	0.9351	0.9579
Ours	2	PSNR	31.38	37.12	35.16	33.37	40.83	29.25	39.5	42.76
		SSIM	0.9784	0.9432	0.9065	0.9083	0.9761	0.8872	0.9639	0.9846
Ours (w/o edge features)	2	PSNR	31.32	37.05	35.12	33.33	40.75	29.2	39.46	42.64
		SSIM	0.9712	0.9407	0.9058	0.9078	0.9740	0.8865	0.9617	0.9837

On-Device Performance

- We achieve an OCR accuracy of 75.89% on TextSR dataset using Tesseract whereas the original HR images give an accuracy of 78.1%.
- Number of parameters: 57k
- Inference time : 11.7ms per crop on S10+
- Our model works on both scene and document type images.

THANK YOU!

Please reach out to us in case of any queries.