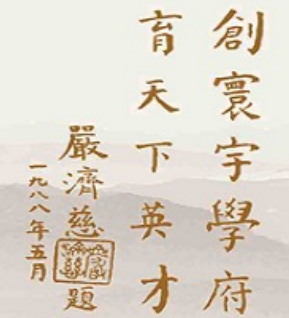
中国科学技术大学
University of Science and Technology of China

CNN-Based Text Image Super-Resolution Tailored for OCR

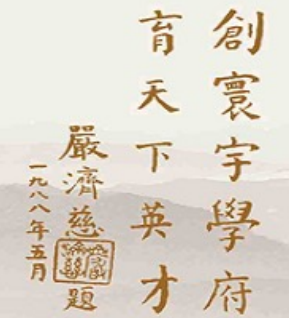
Haochen Zhang, Dong Liu, Zhiwei Xiong

CAS Key Laboratory of Technology in Geo-Spatial Information Processing and Application System, University of Science and Technology of China



中国科学技术大学
University of Science and Technology of China

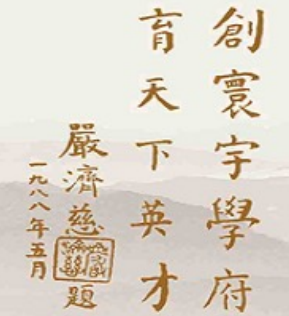
Content



中国科学技术大学
University of Science and Technology of China

Part one

Introduction



OCR & SR

Optical Character Recognition(**OCR**)

20 LOIC REMY

Note： low-resolution images may hamper the performance of OCR.

Text image super-resolution(**SR**)

PSNR vs OCR accuracy

RESULTS OF DIFFERENT METHODS ON THE ICDAR 2015 TEXTSR DATASET

Method	RMSE	PSNR	MSSIM	OCR (%)
Bicubic	19.04	23.5	0.879	60.64
Lanczos3	16.97	24.65	0.902	64.36
Orange Labs	11.27	28.25	0.953	74.12
Zeyde et al.	13.05	27.21	0.941	69.72
A+	10.03	29.5	0.966	73.1
Synchromedia Lab	62.67	12.66	0.623	65.93
ASRS	12.86	26.98	0.95	71.25
SRCNN-1	7.52	31.75	0.98	**77.19**
SRCNN-2	**7.24**	**33.19**	**0.981**	76.1

Two points:

- SR improves the accuracy of OCR.
- The method achieving the highest PSNR does not achieve the highest OCR accuracy.

Contribution

We summarize our attempts of improving CNN-based text image SR method to facilitate OCR.

- A new loss function
- Image padding method
- Model combine

Experimental results:

77.19% VS 78.10% VS 78.80%

state-of-art

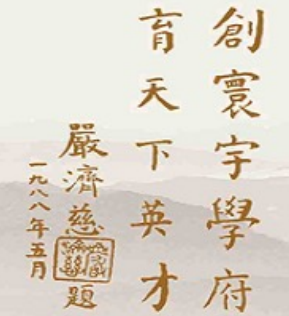
our proposed method

the original HR images performance

中国科学技术大学
University of Science and Technology of China

Part two

Method



New Loss Function

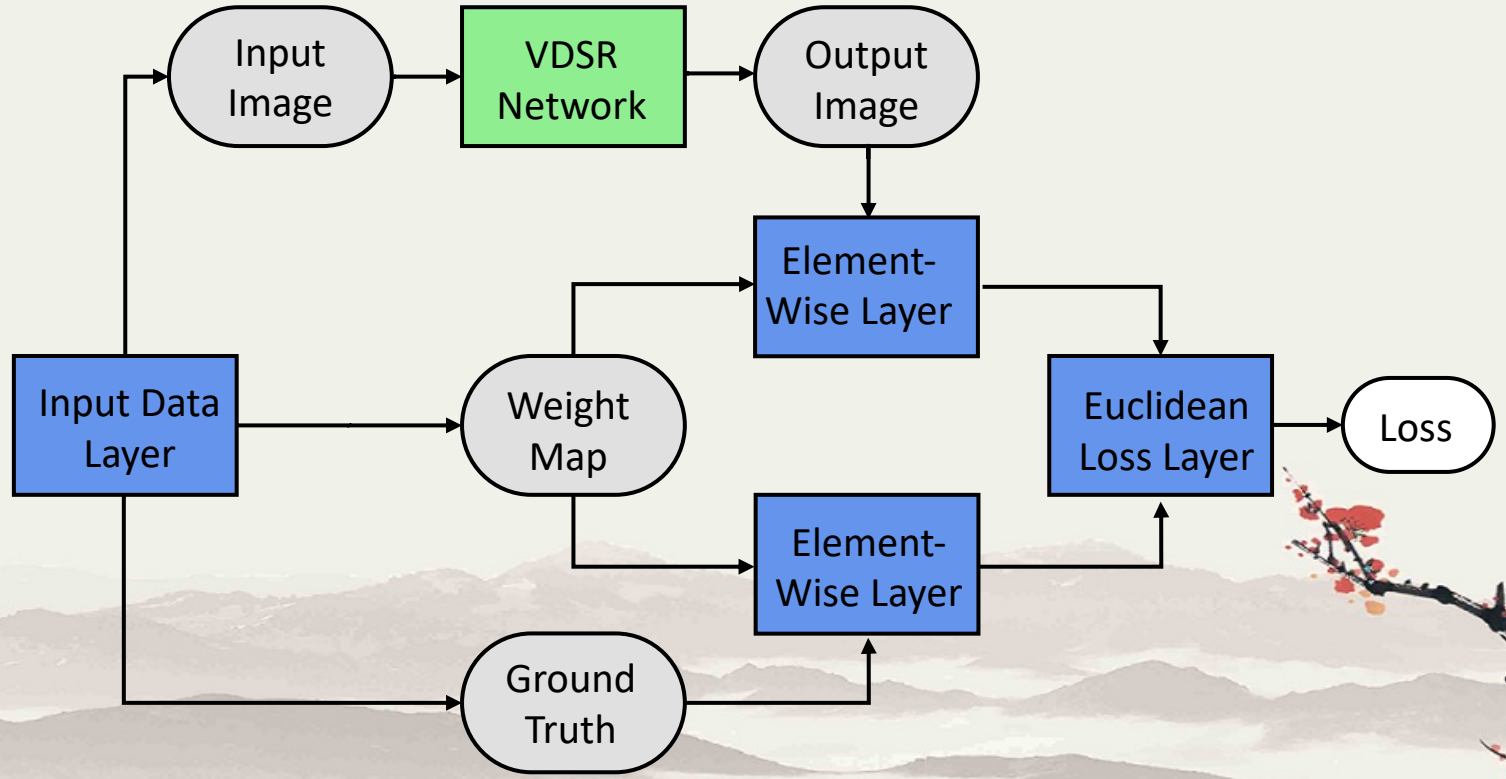
We use a weighted MSE (WMSE) to emphasize high contrast edges more than others to guide CNN to concern more on the high-frequency image details .

$$\mathrm{WMSE} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\left\|I(i,j)-\hat{I}(i,j)\right\|^2 \times f[\mathrm{grad}(i,j)]}{mn}$$

- I and $\hat{I}$ are the original and super-resolved images, respectively
- $grad$ is the gradient magnitude map of the original image.
- $f[\cdot]$ is a certain function to convert gradient magnitude into weight

New Loss Function

$$\mathrm{WMSE} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\left\| I(i,j) \times g[\mathrm{grad}(i,j)] - \hat{I}(i,j) \times g[\mathrm{grad}(i,j)] \right\|^2}{mn}, g[x] = \sqrt{f[x]}$$



Model combine

Suppose we have

- Multiple trained networks N_i , $i = 1, \ldots, M$
- A set of combination weights w_i , $i = 1, \ldots, M$

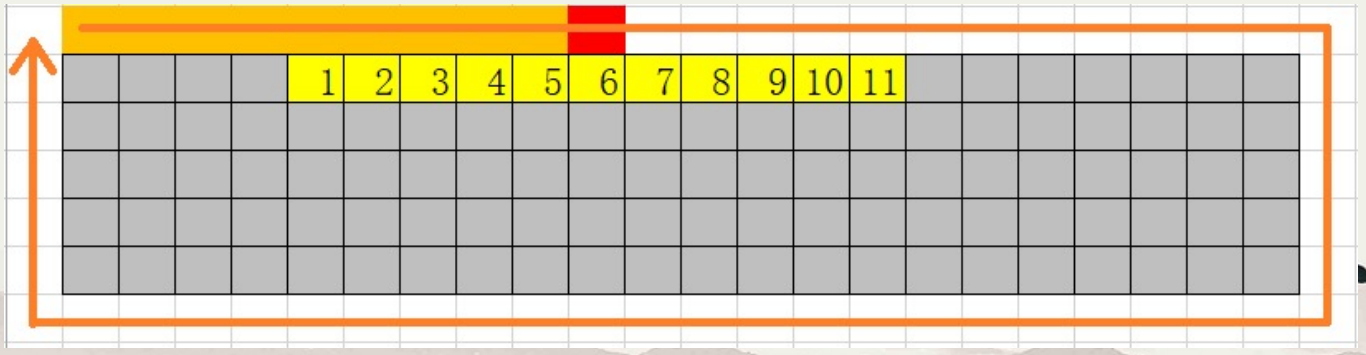
Combine as $$\hat{I} = \sum_{i=1}^{M} w_i \times N_i(I_{LR})$$

We do have

- N_i : Trained by different random initializations
- w_i : Exhaustive search from a finite set

Image padding

In each recursion



Part three

Experiment

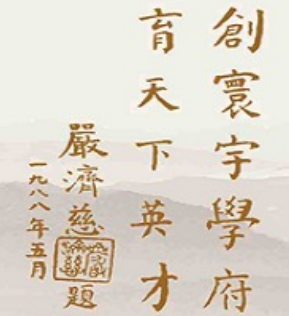


Image padding

HR image

No padding

Duplication

Our padding

RESULTS OF DIFFERENT PADDING METHODS

Method	PSNR	MSSIM
No padding	32.18	0.9841
Duplication padding	32.31	0.9842
Our proposed padding	32.42	0.9842

New Loss Function

loss fuction	Experiment	OCR accuracy	SSIM	PSNR（dB）
1）f(x)=x^2	Net 1.1	76.34%	0.9795	31.65
	Net 1.2	76.54%	0.98	31.71
	Net 1.3	76.41%	0.9796	31.57
	Net 1.4	75.86%	0.9795	31.59
	average	76.29%	0.97965	31.63
2）f(x)=x	Net 2.1	76.54%	0.9815	31.87
	Net 2.2	76.82%	0.9812	31.80
	Net 2.3	75.15%	0.9813	31.91
	Net 2.4	76.44%	0.9811	31.74
	average	76.24%	0.981275	31.83
3）f(x)=x^1/2	Net 3.1	77.23%	**0.9828**	**32.00**
	Net 3.2	75.15%	0.9821	31.94
	Net 3.3	76.72%	0.9826	31.94
	Net 3.4	**77.98%**	0.9827	31.98
	average	76.77%	0.98255	31.97

Table 1 WMSE

Table 2 State-of-art

Method	**RMSE**	**PSNR**	**MSSIM**	**OCR (%)**
SRCNN-1	7.52	31.75	0.98	**77.19**
SRCNN-2	**7.24**	**33.19**	**0.981**	76.1

Table 3 Standard MSE

Method	PSNR	MSSIM
No padding	32.18	0.9841

* The OCR accuracy is measured by using the Tesseract-OCR software.

Model combine

$$\hat{I} = \alpha \cdot N_1(I_{LR}) + (1-\alpha) \cdot N_2(I_{LR})$$

$N_1 = Net\ 3.4$
$N_2 = Net\ 3.2$

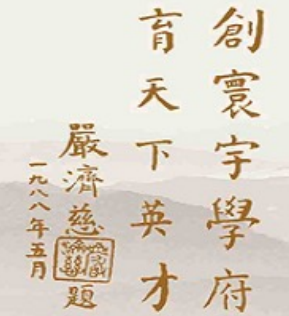
dataset	method	OCR (%)	OCR of combine(%)
training set	Net 3.2	73.58	74.25
	Net 3.4	73.93	
testing set	Net 3.2	75.15	78.10
	Net 3.4	77.98	

α = 0.67

中国科学技术大学
University of Science and Technology of China

Part four

Conclusion



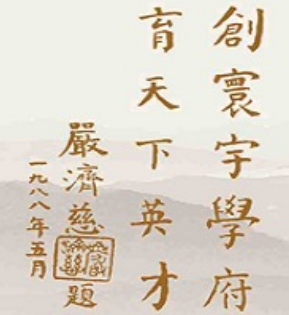
中国科学技术大学
University of Science and Technology of China

Conclusion

First A new loss function to guide the CNN to focus on high-frequency image details.

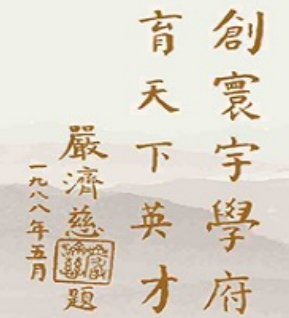
Second Image padding method to refine the image boundaries during CNN based SR.

Third Model combination to further improve the performance.



中国科学技术大学
University of Science and Technology of China

Thank you for listening !

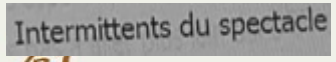
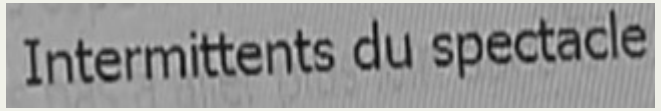


Why the results of using VDSR on the text images are not satisfactory, especially at image boundaries ?

- Text images in the TextSR dataset are much smaller compared with natural images.
- VDSR is very deep (20 layers), the receptive field of the last convolutional layer is 41 $\times$41 .

ICDAR 2015 TextSR dataset

中国科学技术大学
University of Science and Technology of China

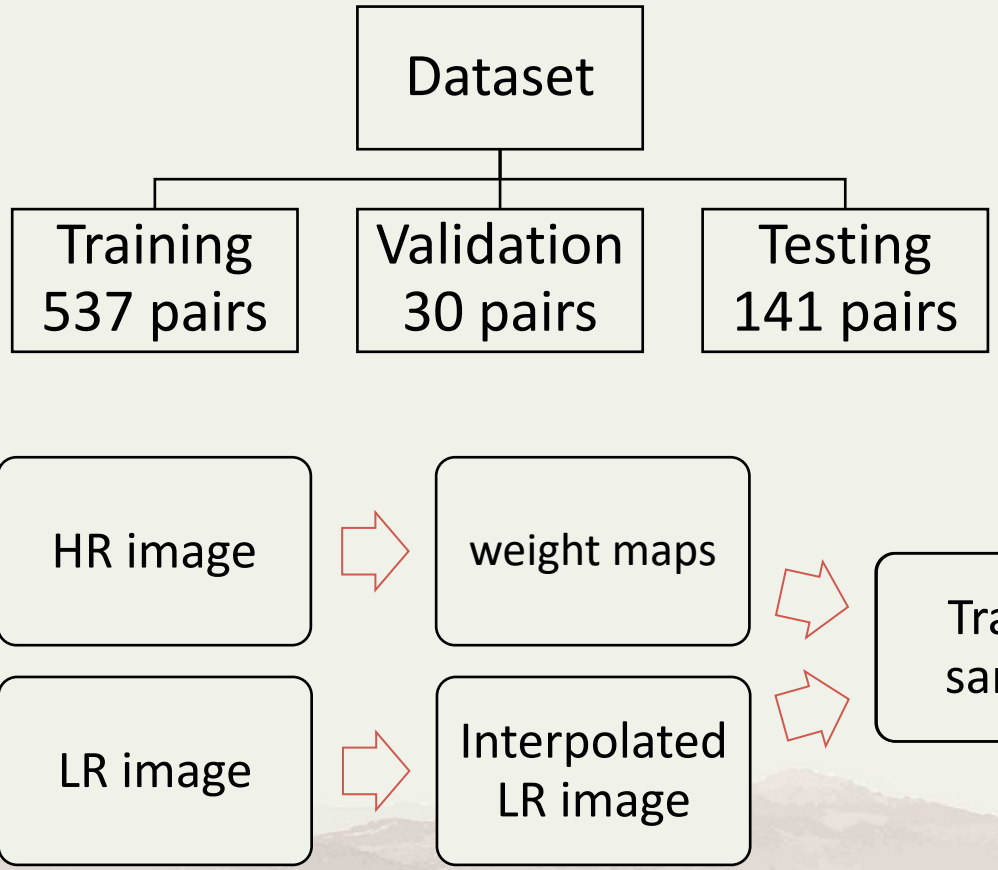
Low resolution	High resolution
 height: 9	 height: 18
 height: 13	 height: 26
 height: 17	 height: 34
 height: 22	 height: 44
 height: 26	 height: 52

The training set consists of 567 pairs of HR-LR grayscale images and the ground-truth OCR results

The test set consists of 141 pairs of HR-LR images as well as their ground-truth OCR results

OCR results include English letters, numbers, and 14 special characters such as “,”.

Implementation Details



Framework Caffe:

- The initial learning rate is 10^{-2} (10^{-3} for biases)
- Learning rate is decreased to 0.1 every 49,160 iterations
- Training is stopped after 196,640 iterations

10 layer vs 20 layer

Table 1 10 layer VDSR

Learning rate	bicubic	VDSR
10^-4	23.5011	30.0127
10^-5	23.5011	29.9136
10^-6	23.5011	29.8435
10^-4-->10^-5	23.5011	**30.0622**
10^-5-->10^-6	23.5011	29.9276
10^-4-->10^-7	23.5011	30.0061
10^-5-->10^-8	23.5011	29.8794

Table 2 20 layer VDSR

Method	PSNR	MSSIM
No padding	**32.18**	0.9841

SR vs original HR

Label: Apres_utilisation_du_shampooing_force_&_densite,_pellicules_visibles,_usage_regulier
SR: Apms_utilisation_du_shampooing_i_&_densite,_pellicules_visibies,_usage_rigulier,
HR: iAprths_utilisation_du_shampooing_IQ_&_densite_peilicuies_visibies,_usage_r&gulier.

Label: MERCREDI
SR: MENCREDI
HR: MENCKEDI

Label: Est-ce_le_Christ
SR: I:st-ce_I6_Christ
HR: I:st-ce_I6_UHFISL

Label: 21_GABRIEL_ACHILIER
SR: 1_21_GABRIEL_ACHILIER
HR: &_21_GABMEL_ACHILIER