Review: VDSR (Super Resolution)

SH Tsang Follow

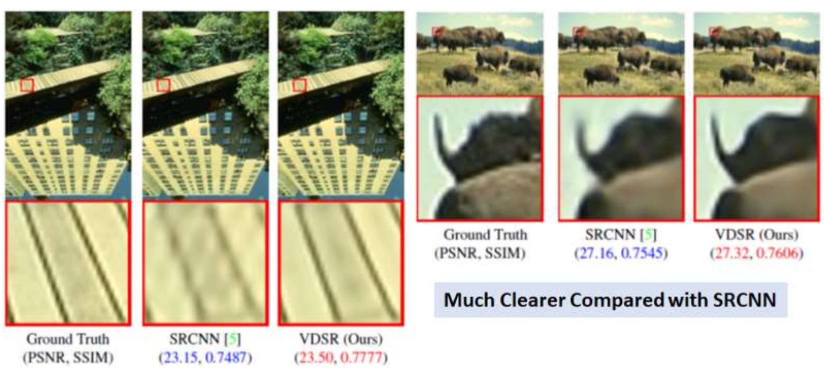
Oct 30, 2018 · 4 min read

This time, **VDSR (Very Deep Super Resolution)** is reviewed. VDSR is a deep learning approach for enlarging an image. It has **20 weight layers** which is much deeper compared with SRCNN which only got 3 layers.

> *Sometimes, we only got a poor image and we want to have digital enlargement (zoom in), but the image gets blurred when zoomed in. This is because the conventional interpolation or enlargement of a small image to become a large image, will get a poor image quality. With VDSR, we can obtain a high-resolution (HR) image with high quality from a low resolution (LR) image.*

Below are two examples.

VDSR is one of the classical state-of-the-art SR approaches which was published in **2016 CVPR** with about **800 citations** while I was writing this paper. (SH Tsang @ Medium)



Much Clear Image after Enlargement Using VDSR (Edges are much clearer)

This embedded content is from a site that does not comply with the Do Not Track (DNT) setting now enabled on your browser.

Please note, if you click through and view it anyway, you may be tracked by the website hosting the embed.

Learn More about Medium's DNT policy

Some More Amazing Results

The results are amazing!! So, let's see how it works.

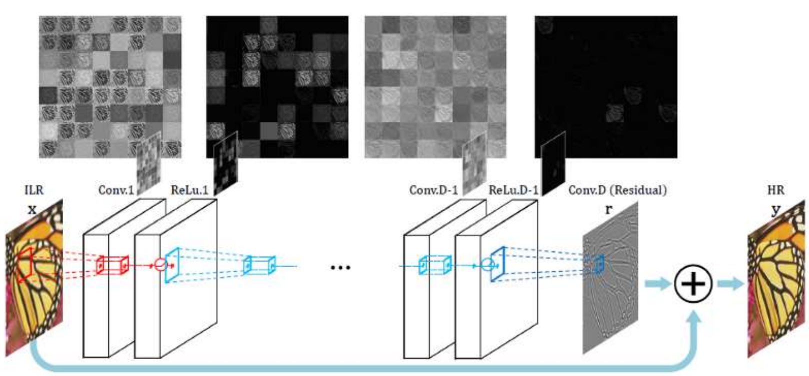
. . .

What Are Covered

1. **VDSR Network Architecture**
2. **Some Details About Training**
3. **Results**

. . .

1. VDSR Network Architecture



VDSR Network Architecture

The VDSR architecture is neat as above:

1. The **LR image is interpolated as ILR image** and input to the network.

2. The ILR image goes through **(D-1) times of Conv and ReLU layers**.

3. And then followed by a **D-th Conv** (Conv.D (Residual) in the figure).

4. Finally, the **output is added with the ILR image and obtain the HR image**.

These are **64 filters with the size of 3×3 for each conv layer**. (VGGNet has addressed the issue of consecutive 3×3 filters helps to obtain larger receptive fields so that we do not require any large filters such as 5×5 and 7×7. If interested, please read my VGGNet review.)

As we can see, the ILR is added to the output of the network to get back the HR image, the loss function becomes:

$$\frac{1}{2}||\mathbf{r} - f(\mathbf{x})||^2$$

where **r=y-x**. Thus, **the network is learning the residual errors between the output and input** instead of learning the HR output directly just like SRCNN.

Epoch	10	20	40	80
Residual	36.90	36.64	37.12	37.05
Non-Residual	27.42	19.59	31.38	35.66
Difference	9.48	17.05	5.74	1.39

(a) Initial learning rate 0.1

Epoch	10	20	40	80
Residual	36.74	36.87	36.91	36.93
Non-Residual	30.33	33.59	36.26	36.42
Difference	6.41	3.28	0.65	0.52

(b) Initial learning rate 0.01

Epoch	10	20	40	80
Residual	36.31	36.46	36.52	36.52
Non-Residual	33.97	35.08	36.11	36.11
Difference	2.35	1.38	0.42	0.40

(c) Initial learning rate 0.001

Residual vs Non-Residual with Different Learning Rate

With residual learning, the convergence is much faster than that of non-residual learning. At epoch 10, residual one already got above 36 dB while non-residual one still only got from 27-34 dB.

. . .

2. Some Details About Training

2.1 Adjustable Gradient Clipping

The gradients are clipped to [-θ/γ; θ/γ □], where γ denotes the current learning rate. And **θ is tuned to be small to avoid exploding gradients** in a high learning rate regime.

When D= 20, 20-layer network training is done within 4 hours whereas 3-layer SRCNN takes several days to train.

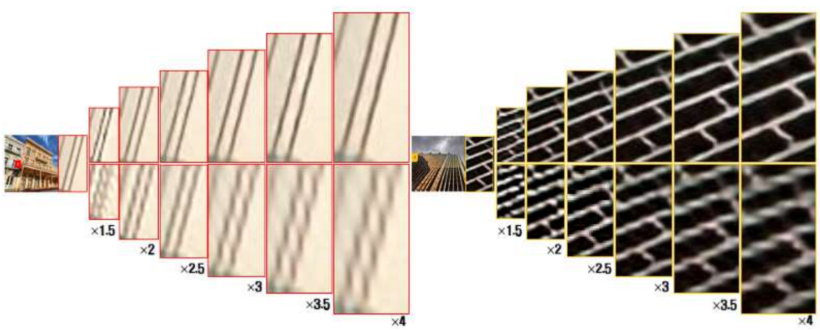
2.2 Multi-Scale Training

Test / Train	×2	×3	×4	×2,3	×2,4	×3,4	×2,3,4	Bicubic
×2	37.10	30.05	28.13	37.09	37.03	32.43	37.06	33.66
×3	30.42	32.89	30.50	33.22	31.20	33.24	33.27	30.39
×4	28.43	28.73	30.84	28.70	30.86	30.94	30.95	28.42

Mutli-Scale Training Results

When Single-scale images are used, the network can only work well for the same scale during testing, for the testing of other scales, PSNR is even worse than conventional bicubic interpolation.

By using ×2 ,×3, ×4 scale images for training, the highest PSNRs are obtained for all scales during testing.



Multi-Scale VDSR (Top), Single-Scale Dong's [5] (Bottom)

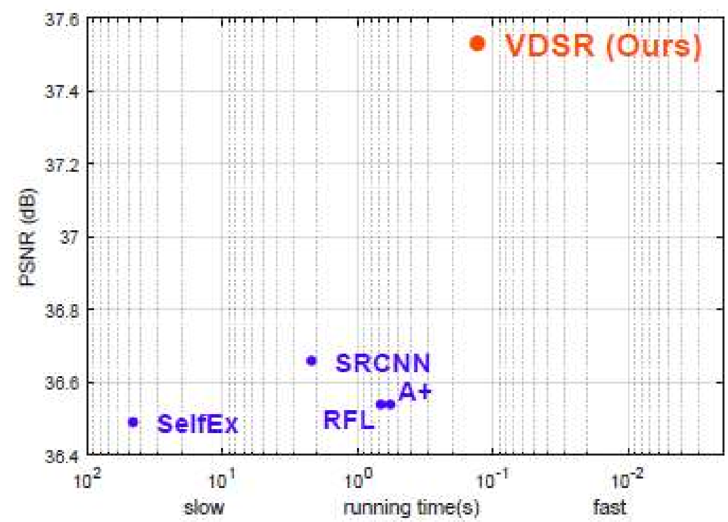
Single-Scale Dong's [5]obtains blurred images while VDSR has much clearer edges.

. . .

3. Results

Dataset	Scale	Bicubic PSNR/SSIM/time	A+ [22] PSNR/SSIM/time	RFL [18] PSNR/SSIM/time	SelfEx [11] PSNR/SSIM/time	SRCNN [5] PSNR/SSIM/time	VDSR (Ours) PSNR/SSIM/time
Set5	×2	33.66/0.9299/0.00	36.54/0.9544/0.58	36.54/0.9537/0.63	36.49/0.9537/45.78	36.66/0.9542/2.19	37.53/0.9587/0.13
	×3	30.39/0.8682/0.00	32.58/0.9088/0.32	32.43/0.9057/0.49	32.58/0.9093/33.44	32.75/0.9090/2.23	33.66/0.9213/0.13
	×4	28.42/0.8104/0.00	30.28/0.8603/0.24	30.14/0.8548/0.38	30.31/0.8619/29.18	30.48/0.8628/2.19	31.35/0.8838/0.12
Set14	×2	30.24/0.8688/0.00	32.28/0.9056/0.86	32.26/0.9040/1.13	32.22/0.9034/105.00	32.42/0.9063/4.32	33.03/0.9124/0.25
	×3	27.55/0.7742/0.00	29.13/0.8188/0.56	29.05/0.8164/0.85	29.16/0.8196/74.69	29.28/0.8209/4.40	29.77/0.8314/0.26
	×4	26.00/0.7027/0.00	27.32/0.7491/0.38	27.24/0.7451/0.65	27.40/0.7518/65.08	27.49/0.7503/4.39	28.01/0.7674/0.25
B100	×2	29.56/0.8431/0.00	31.21/0.8863/0.59	31.16/0.8840/0.80	31.18/0.8855/60.09	31.36/0.8879/2.51	31.90/0.8960/0.16
	×3	27.21/0.7385/0.00	28.29/0.7835/0.33	28.22/0.7806/0.62	28.29/0.7840/40.01	28.41/0.7863/2.58	28.82/0.7976/0.21
	×4	25.96/0.6675/0.00	26.82/0.7087/0.26	26.75/0.7054/0.48	26.84/0.7106/35.87	26.90/0.7101/2.51	27.29/0.7251/0.21
Urban100	×2	26.88/0.8403/0.00	29.20/0.8938/2.96	29.11/0.8904/3.62	29.54/0.8967/663.98	29.50/0.8946/22.12	30.76/0.9140/0.98
	×3	24.46/0.7349/0.00	26.03/0.7973/1.67	25.86/0.7900/2.48	26.44/0.8088/473.60	26.24/0.7989/19.35	27.14/0.8279/1.08
	×4	23.14/0.6577/0.00	24.32/0.7183/1.21	24.19/0.7096/1.88	24.79/0.7374/394.40	24.52/0.7221/18.46	25.18/0.7524/1.06

Comparison with State-of-the-art Results (Red: The best, Blue: 2nd Best)



VDSR is much faster than SRCNN

The above table shows that **VDSR obtains the best results with the least testing time**.

· · ·

With AI chipsets become popular in the future, VDSR or other state-of-the-art approaches can be applied in real-time for image enlargement, and even applied in video.

· · ·

References

1. [2016 CVPR] [VDSR] [Accurate Image Super-Resolution Using Very Deep Convolutional Networks]

My Related Reviews

[SRCNN] [FSRCNN] [VGGNet]