Paper ID: 71, Paper Title: Unsupervised Real Image Super-Resolution via Generative Variational AutoEncoder, Track Name: Challenge Papers

Authors'Response

View Meta-Reviews

Paper ID 71

Paper Title: Unsupervised Real Image Super-Resolution via Generative Variational AutoEncoder Track Name Challenge Papers

META-REVIEWER #1

META-REVIEW QUESTIONS

1. Decision. Note that in case of conditional accept, the submitted camera ready paper will be checked carefully once more. If the paper does not include the required changes then it will be removed from publication. Conditional Accept (Changes are necessary for acceptance)

2. Consolidated report the reviewers find merits and the chair agrees. However, there are still issues that need to be addressed before acceptance.

Therefore, we are inviting the authors to make sure that they:

Comment:

(a) address the reviewers' comments in their improved camera ready paper that will be checked again

Response: Modified. Please check the following response to the reviewers.

Comment:

(b) Cite the following related work: Fritsche et al "Frequency Separation for Real-World Super-Resolution", ICCVW 2019, the winner of the AIM 2019 challenge on real-world SR.

Response: Added. The modification is as follow.

"Convolutional Neural Network (CNN) works better than most machine learning approaches because it can digest huge amount of data to learn different filters for feature extraction via backpropagation. Many CNN based SR approaches [6, 12, 14, 16, 34, 10, 21, 20, 9, 19, 1, 5, 15, 28, 4, 24, 23, 22] have successfully boosted up the image super-resolution performance in both computation and quality."

[9] Manuel Fritsche, Shuhang Gu and Radu Timofte. Frequency separation for real-world super-resolution. 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), pages 3599–3608, 2019.

(Please see p.1, column2, line1 from bottom to p.2, column1, line 6, and ref. [9] in Reference section of the revised paper.)

Comment

(c) define real-world super-resolution and cite and discuss the introductory works [22,23] in Introduction or Related work sections.

Response: Reference added. The modification is as follows.

"*Though researchers came up with different simulations to model the down-sampling process, it still targets on one specific applications. Real-world super-resolution is far* *more complicated. As investigated in [23, 24], there is no available ground-truth LR-HR image pairs. Most supervised image SR approaches have the overfitting problem.*"

(Please see p.2, column1, paragraph3, lines 1-6)

Comment

(d) provide a bibtex of this paper to the challenge organizers for inclusion with the challenge report.

Response: The bibtext of this paper is as follows.

```
@ARTICLE{dSRVAE,
    author       = {Zhi{-}Song Liu and
    Wan{-}Chi Siu and
    Li{-}Wen Wang and
    Chu{-}Tak Li and
    Marie{-}Paule Cani and
    Yui{-}Lam Chan},
    title         = {Unsupervised Real Image Super-Resolution via Generative Variational AutoEncoder},
    journal       = {2020 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW2020)},
    year          = {2020},
}
```

Reviewer 1

View Reviews

Paper ID: 71

Paper Title: Unsupervised Real Image Super-Resolution via Generative Variational AutoEncoder Track Name Challenge Papers

Questions 1. Paper Review Summary This method for real-world super-resolution consists of a cleaning the input image and super-resolution step. For the denoising they are using an Auto Encoder.

Strengths - They understood the problem of domain adoption and mitigated the artifacts that come with standard methods.

Weaknesses

- The visual results contain a slight checkerboard pattern. - Cleaning the input for SR instead of shifting the domain during learning has shown to give more blurry results. This is also the case here.

Recommendation The problem of the challenge was understood and tackled in a reasonable way. Therefore I would propose weak accept. For the final version the writing should be improved and the figures should show more details of zoom crops for the final version.

Response: Modified. We checked figures carefully and replaced some of them as shown in the following figures of the revised paper.

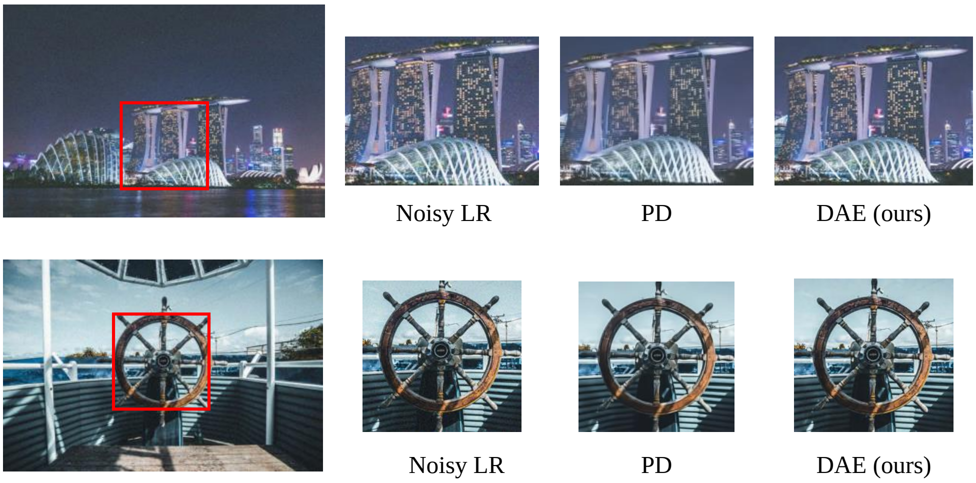


Figure 4. Visualization of image denoising on NTIRE2020 validation images. Enlarged red boxes are included for better comparison.

(Please see page 6 of the revised paper)

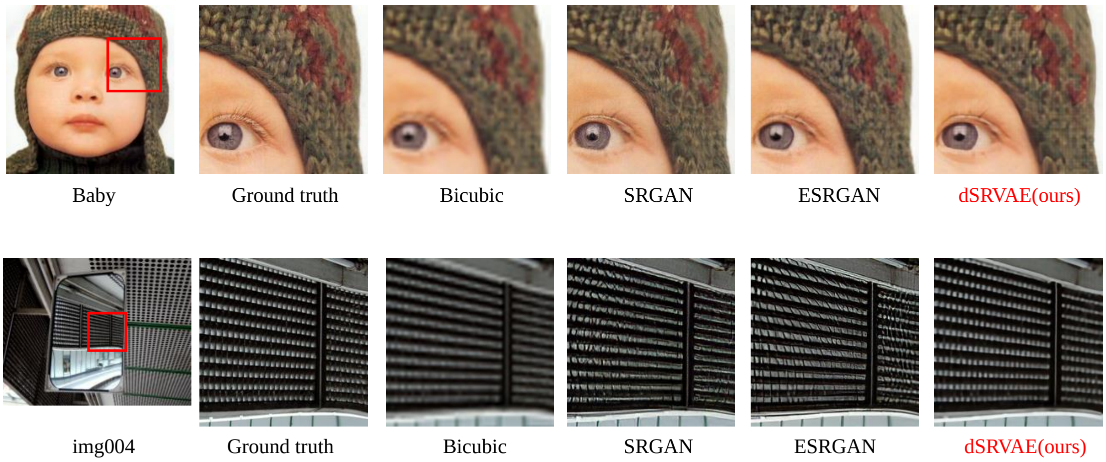


Figure 5. Visualization of 4× image super-resolution on Set5 and Urban100 images. Enlarged red boxes are included for better comparison.

(Please see page 7 of the revised paper)

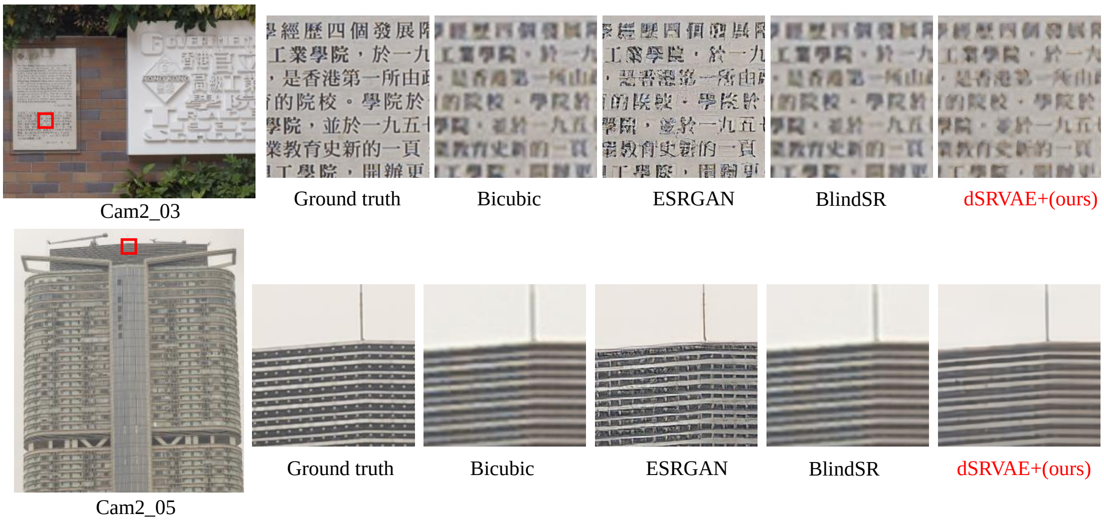


Figure 6. Visualization of 4× image super-resolution on NTIRE2019 validation. Enlarged red boxes are included for better comparison.

(Please see page 7 of the revised paper)

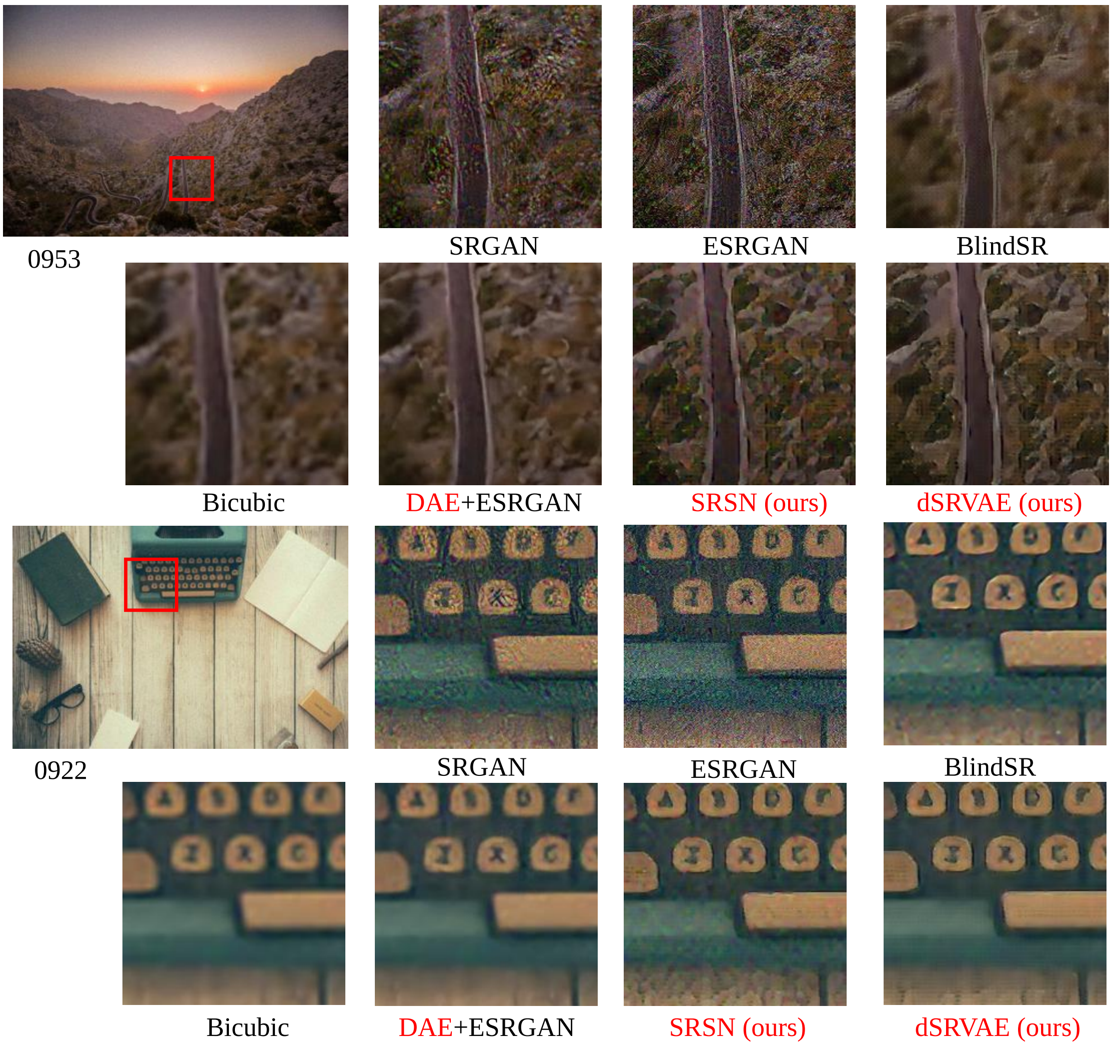


Figure 7. Visualization of 4× image super-resolution on NTIRE2020 validation.

Enlarged red boxes are included for better comparison.

(Please see page 8 of the revised paper)

Furthermore, we highly recommend readers to check the GitHub link provided in the paper
https://github. com/Holmes-Alan/dSRVAE
and download the original SR images for comparison. We also provide supplementary results on NTIRE2020 validation at the end of this response.
Note that our proposed method is an unsupervised approach which was not trained with ground-truth LR-HR images. Compared with other state-of-the-art supervised approaches, it is reasonable to have SR images with a little bit of blur effect. From the results on Figure 5 and 6, we can find that ESRGAN and SRGAN generate sharper "fake" patterns while the proposed dSRVAE can restore the details with less errors. Our main focus is on real world image with noises. That is why we propose a joint denoising and super-resolution model. From the results on Figure 7, we can find that our approach can handle real noises and generate SR images with better visual quality. It demonstrates the efficiency of our proposed work.

Comment: They do not cite other papers that use denoising for super-resolution

Response: References added. The modification is as follows.
"*On the other hand, with the huge learning capacity of deep neural network, we can assume degradation factors in low-resolution image generation, like adding different noise levels, forming blur kernels with some combinations of scale factors, etc. and then combine various of these factors for a general image super-resolution. For example, we can have joint demosaicing and super-resolution [37,33], joint denoising and super-resolution [31] and joint deblurring, denoising and super-resolution [5, 29, 30].*"
(Please see p.2, column2, line 7 from the bottom to p.3 to p.3, column1, line 3 of the revised paper)

Reviewer 2:
Questions 1. Paper Review This paper proposes a real-world image super-resolution based on using a denoising VAE to first clean the input image from noise. A light-weight network is then applied to super resolve the image. The propose method has some interesting elements. The particular solution is novel. Results on the NTIRE2019 dataset and other datasets are favorable compared to the baselines.
There are some weaknesses:

Comment: Regarding writing, the paper suffers from a bit childish language and awkward formulations. In particular in the introduction.

Response: Further improved. We have carefully checked the paper, rewritten some parts and chosen better words for the improvement of the presentation.

Comment: The denoising strategy has been explored in previous works in this context. In particular, the authors does not refer to the Cycle-in-cycle GAN presented in the CVPR 2018 NTIRE workshop. The authors must better refer to prior works.

Response: Reference added. The modification is as follows.

"Since there is no ground truth HR images to calculate the reconstruction loss (e.g. L1-norm loss), we propose a novel cycle training strategy that comes from the back-projection theory, which is different from the previous related works [30, 38]."

[30] Yuan Yuan, Siyuan Liu, Jiawei Zhang, Yongbing Zhang, Chao Dong, and Liang Lin. Unsupervised image super resolution using cycle-in-cycle generative adversarial networks. CoRR, abs/1809.00437, 2018

[38] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle consistent adversarial networks. In Computer Vision (ICCV), 2017 IEEE International Conference on, 2017

(Please see p.4, column2, lines 1-4 of the 2nd paragraph of the revised paper.)

Comment: The results are clean but a bit blurry.

Response: Clarified, a new figure 7 is modified and added in the revised version.

Note that our proposed method is an unsupervised approach which was not trained with ground-truth LR-HR images. Compared with other state-of-the-art supervised approaches, it is reasonable to have SR images with a little bit of blur effect. From the results on Figure 5 and 6, we can find that ESRGAN and SRGAN generate sharper "fake" patterns while the proposed dSRVAE can restore the details with less errors. Our main focus is on real world image with noises. That is why we propose a joint denoising and super-resolution model. From the results on Figure 7, we can find that our approach can handle real noises and generate SR images with better visual quality. It demonstrates the efficiency of our proposed work.

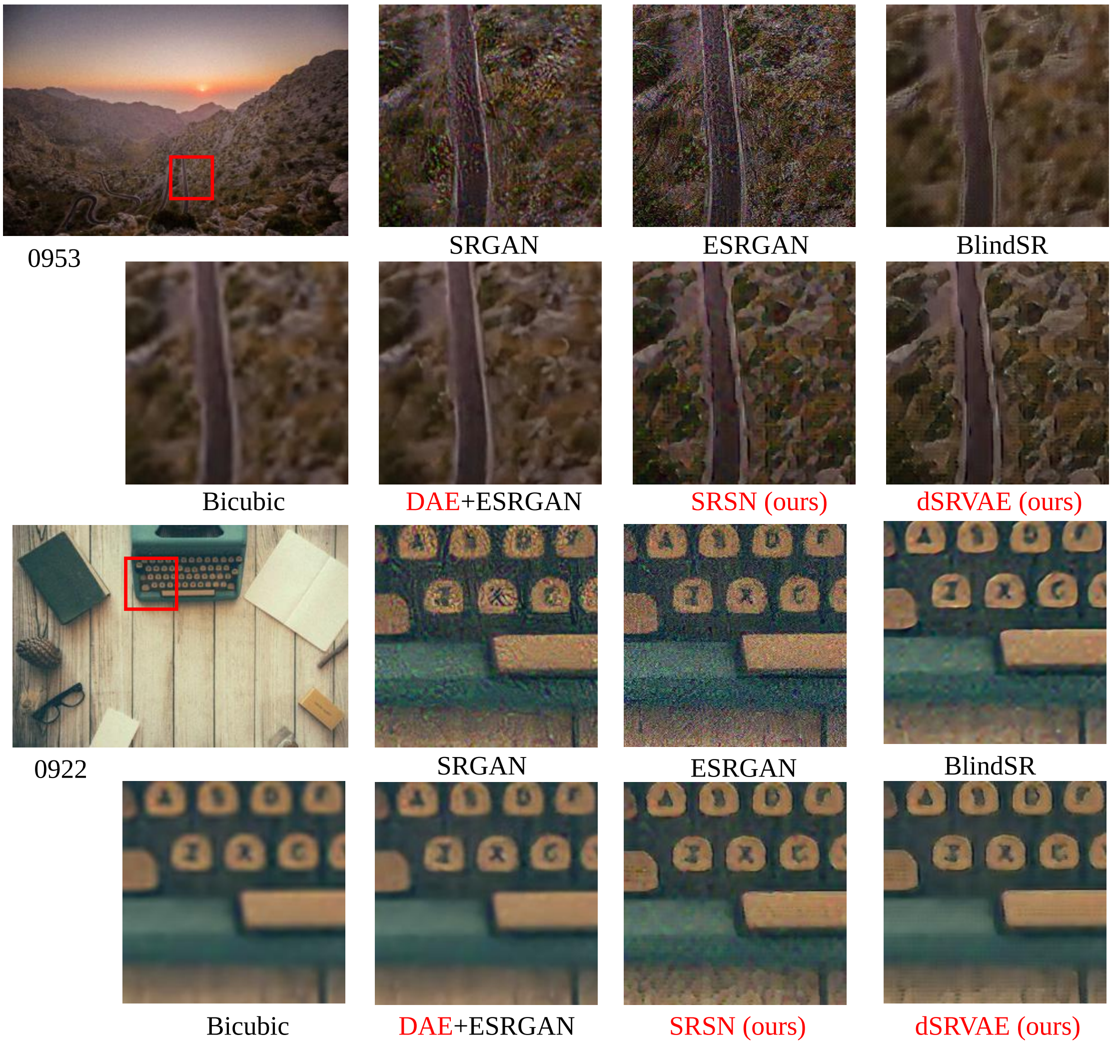


Figure 7. Visualization of 4× image super-resolution on NTIRE2020 validation. Enlarged red boxes are included for better comparison.

(Please see page 8 of the revised paper)

We also provide supplementary results on NTIRE2020 validation at the end of this response.

Comment: This paper has some issues, but is still a relevant contribution for the workshop.
Response: Thanks.

Supplementary results:

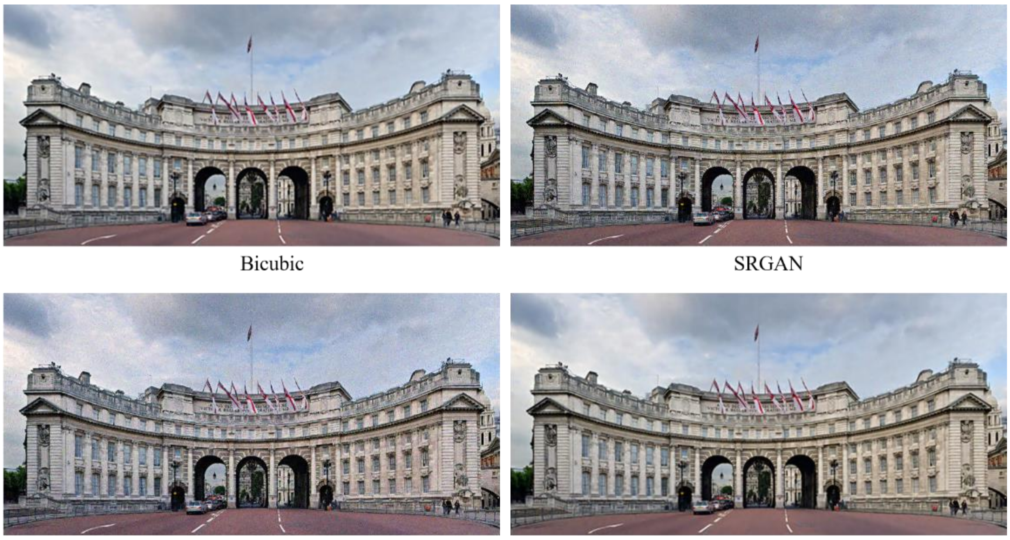
Visualization on NTIRE2020 Real Image SR Validation dataset: We compared with SRGAN and ESRGAN for visual comparison. It can be found that SRGAN and ESRGAN cannot handle real noise. Our proposed dSRVAE can clean and super-resolve the image with better quality, like the sky in all images.

0930.png

Bicubic SRGAN

ESRGAN Ours

0993.png



Bicubic SRGAN

ESRGAN Ours

0959.png

Bicubic SRGAN

ESRGAN Ours