**IMAGE SUPER RESOLUTION USING VARIATIONAL METHODS**

ARIHANT M HULLOLI

Dept of Electronics and Communication Engineering

KLS Gogte Institute of Technology

Belagavi Karnataka

**Abstract**:

The goal of a Super Resolution (SR) approach is to create a high-resolution image from a succession of low-resolution images of the same scene. One main impediment to SR restoration is reducing noise and blur without harming edges. Using a controllable weighting parameter, we offer a unique multi frame picture SR technique based on a convex combination of Bilateral Total Variation and a non-smooth second order variational regularization. We demonstrate the existence of a minimizer of the proposed SR model in the space of functions with bounded Hessian, and we validate the approach's efficacy in avoiding undesired artefacts.

**Introduction**

Single Image Super-Resolution (SISR) consists in producing a high-resolution image from its low-resolution counterpart. Image super-resolution has long been considered one of the most arduous challenges in image processing. This is yet another computer vision task that was transformed by the deep learning revolution and has potential applications including but not limited to medical imaging, security, computer graphics, and surveillance.

Since the VAE is an unconditional generative model, in order to perform image super-resolution it has to be turned into a conditional generative model which generates data depending on additional conditioning data. This can be achieved by using the framework of Conditional Variational Autoencoders (CVAE) [39], where the prior is conditioned on an additional random variable and parameterized by a neural network. In this work, we introduce VDVAE-SR, a VDVAE conditioned on low-resolution images by adding a new component that we call LR-encoder as it resembles the encoder of the original VDVAE. This component is connected to the decoder, passing information on each layer in the top-down path both to the prior and the approximate posterior. During training, the latent distributions of the low- and high-resolution images are matched using the KL divergence term in the evidence lower bound (ELBO).

A drawback of deep models such as the VDVAE is that they require a large amount of computing and training time. One way to compensate in that regard is to apply transfer learning and utilize a pre-trained model in order to speed up the process. . However, this is not always straightforward in practice as presenting a pre-trained model with new data could lead to exploding gradients training and can be sensitive to hyperparameters changes.

**2.Related work**

In this section, we give a brief review on related previous works on single image super-resolution. To introduce our proposed RefVAE method, we also revisit some representative works on generative learning approaches for image SR. Finally, we also introduce related works on reference based image super-resolution.

**2.1 Single image super-resolution**

Single image super-resolution uses one single LR image to produce the corresponding SR image. It is a classic topic in image processing and a lot of works have been proposed to resolve it. With the development of learning approaches, learning based image SR dominates this field. We can categorize learning based SR into two groups: conventional learning based approaches and deep learning based approaches

Let us focus on some representative deep learning based approaches. Dongetal. proposed the first CNN based image SR using only three layers of convolutions to learn the mapping relations between LR and HR images. Later on, VDSR [18], LapSRN, SRResNet, EDSR and many other works use a deeper and wider (number of filters) CNN model to learn the mapping functions for super resolution.

**2.2 Generative learning approaches for SR**

Using generative learning approaches for image SR is a popular topic. We can consider using a single CNN model for SR as discriminative learning that does not explore the data distribution for modelling. On the other hand, generative learning approaches explicitly or implicitly study the data distribution for modelling. Generative Adversarial Network is one of the popular implicit generative approaches. It is also used in image SR. For example, SRGAN uses GAN to train 4× SR. In order to generate more photo-realistic features, VGG based feature loss is used to minimize the deep feature

distances between SR and HR images. ESRGAN [39] proposes to use relativistic GAN [17] to let the discriminator predict relative realness instead of the absolute value. ESRGAN+ [33] further improves ESRGAN by introducing noise for to explore stochastic variations. Vartiaonal AutoEncoder is another choice for SR.

**2.3 Reference based Super-Resolution (RefSR**)

The development of reference based image superresolution (RefSR) is not a surprise to researchers. Originally, it comes from the non-local filtering. The conventional learning based approaches are patch based process. In other words, the LR patches can be reconstructed by learning the mapping models from images/videos with similar contents . For example, k Nearest Neighbor (kNN) can be used for online search so it limits the big data search. Random Forests is a fast algorithm that can classify training patches into different groups for diverse regression modelling.we can formulate the SR problem as learning a stochastic mapping, capable of sampling from the space of plausible high-resolution images given a low-resolution image. . Depending on different external or internal images as reference, we can transfer reference features to fill out the missing information for LR image enlargement. As described in the Learning the Super-Resolution Space Challenge [By using an arbitrary reference image to expand data diversity, RefSR is one way to resolve SR space. The advantage is that 1) multiple predictions can be generated and compared and 2) we allow users to choose references with desired patterns for interactive SR.

**3.METHODS**

In this section, we will give detailed introduction on the proposed Reference based image super-resolution approach via Variational AutoEncoder (RefVAE). Let us formally define the image SR.

The complete architecture of the proposed RefVAE is shown in Figure , which includes the training (upper half of the figure) and testing (lower half of the figure) stages. Our proposed RefVAE takes arbitrary references and LR images X for training and testing. It includes three components: 1) VGG Encoder, 2) CVAE (in the pink box), and 3) Image Decoder.

**3.1 VGG Encoder**

The VGG Encoder follows the structure of VGG-19 by keeping all convolution layers and discarding the fully connection layers. We directly use pre-trained VGG19 to extract feature maps from references and the bicubic up sampled LR images, where G stands for the process of VGG feature extraction. Inside the VGG Encoder, there are three max pooling layers to down sample the input image by 8×. Note that we resize arbitrary reference images to 256 × 256 before passing it through the : 1) VGG was trained for general image classification that the extracted feature maps are generalized to images with different contents and 2) we want to project the LR and reference images to a same feature domain such that we can fuse their features together for super-resolution.

**3.2 Conditional Variational AutoEncoder**

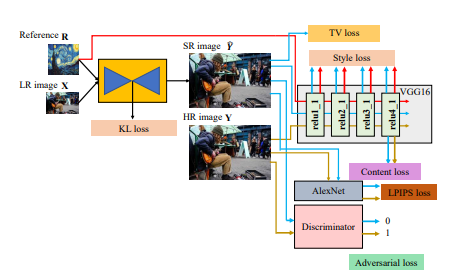
We have the Conditional Variational AutoEncoder (CVAE) that projects the reference feature maps to a latent space to learn the hidden distribution via Feature Encoder. The Feature Decoder learns to transfer the reference features as conditions for LR feature maps. The combination of Feature Encoder and Decoder (for detailed structures: see the right of Figure 2) forms the Variational Inference process. The idea of Variational Inference is to learn the generative model for the reference images that can be represented by a Gaussian model as the mean and variance of the learned Gaussian model. In order to transfer the conditional features to the LR feature map, we use a convolution block to learn the mean and variance (note that the mean and variance are the spatial statistics of the feature maps, rather than the variables of the Gaussian distribution) for the LR feature maps.

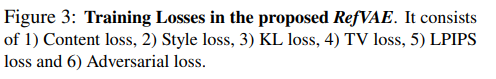
**3.3 Image Decoder**

Finally, the Image Decoder learns to reconstruct the SR image from the fused feature maps . The structure of Image Decoder (dark green boxes in Figure 2) has a similar structure as the VGG Encoder stacking three convolution layers followed by a simple bilinear interpolation.

**3.4. Training losses**

To train the proposed RefVAE to generate SR results with photo-realistic visual quality, we suggest to use a discriminator to reduce the perceptual distance between SR and ground truth images. We design the same discriminator as PatchGAN [38] and the adversarial loss.

****

****

where mean and std are the operations for calculating the mean and variance of the feature maps. In order to have SR images visually close to the HR image, the LPIPS loss is used to measure perceptual difference. The Total Variation loss is used to encourage smooth quality, for which we calculate the first-order horizontal and vertical pixel. where the weighting parameters for content loss, style loss, LPIPS loss, TV loss and KL loss. Figure 3 shows a summary of all the loss terms for readers’ reference.

**4. Evaluation**

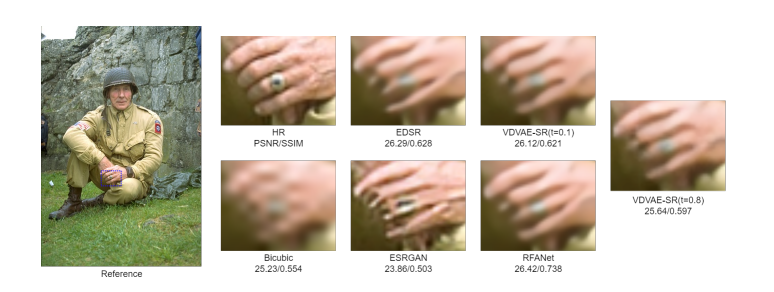
In terms of evaluation metrics, we use the traditional PSNR and SSIM quality measures, both widely used as metrics for image restoration tasks. While PSNR (Peak Signal to Noise Ratio) is calculated based on the mean squared error of the pixel-to-pixel difference, the SSIM (Structural Similarity Method) is considered to have a closer correlation with human perception by calculating distortion levels based on comparisons of structure, luminance, and contrast. Additional to the traditional PSNR and SSIM metrics, we evaluate the produced images using the DISTS [9] score, which has showed evidence that the metric matches closer to human perception.

We thus adopt the same approach, following prior work. Finally, note that the YCbCr space is used during the testing phase exclusively, while the training and validation are still performed in the RGB colour space.

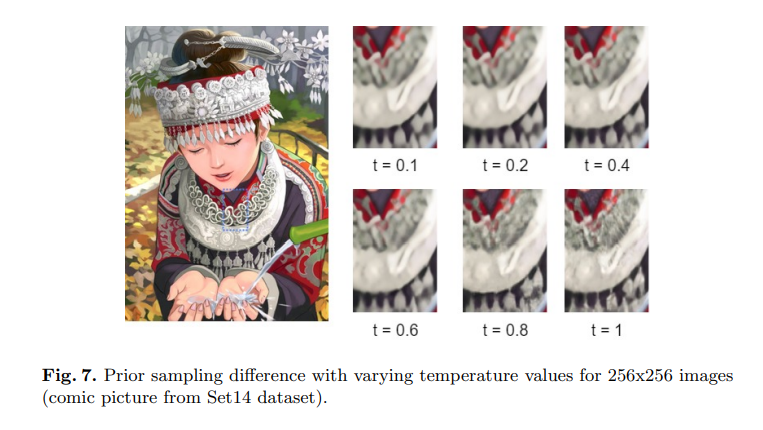
**5.RESULTS**

Quantitative Results. We compare our method to three other super-resolution methods, namely EDSR, ESRGAN and RFANet, based on their official implementation. The quantitative results on PSNR and SSIM are shown in Table 1, where EDSR performs best on both metrics, with our method (t = 0.1) closely following on second place.

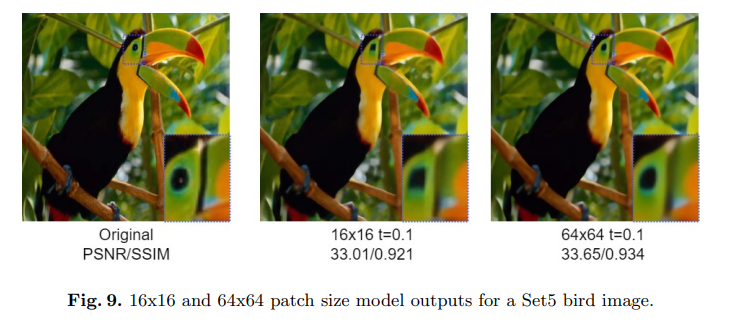
Qualitative Results. Figures show a visual comparison of two pictures from BSD100 dataset between the original HR image, Bicubic, EDSR, ESRGAN, RFANet and our method with both 0.1 and 0.8 temperatures. As for the RFANet, the images are still blurrier, but having more visual similarities with our method than EDSR.

****

**Temperature** The “temperature” parameter t, taking values between 0 and 1, is used in VDVAE when sampling from prior in generative mode, often resulting in higher-quality samples when lowered as observed in previous work [22,43]. Reducing the temperature results in reducing the variance of the Gaussian distributions in the prior and so achieving more regularity in the generated samples. Fig. 7 shows examples of samples with different temperatures.

****

**Patch Size** A crucial parameter in our super-resolution method is the size of patches to which we apply super-resolution addressed also in [46,38]. After experimenting with patches of size 16x16 and 64x64 (i.e., 64x64 and 256x256 after super-resolution), we observed that the 16x16 patch size models were generally performing worse than their counterparts with bigger patch sizes, both in terms of PSNR and SSIM, and in a perceptual sense as the models fail to recreate details that the 64x64 patch models have no problem with.

****

**6.conclusion**

In this paper, we investigated the use of Very Deep Variational Autoencoders (VDVAE) for the purpose of generating super-resolution (SR) images. After the introduction of the proposed VDVAE-SR model, and based on the results presented, we conclude that the introduced model and its quantitative and qualitative results are satisfying as they are comparable to other popular methods, generating images that compensate between image sharpness and visual artifacts. to improve the results even further, as multiple modifications such as changes to training time, layer architecture, or the use of more flexible distributions can be investigated in the future.

**References**

1. Agustsson, E., Timofte, R.: Ntire 2017 challenge on single image super-resolution: Dataset and study. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (7 2017).

2. Bachlechner, T., Majumder, B.P., Mao, H.H., Cottrell, G.W., McAuley, J.: Rezero is all you need: Fast convergence at large depth (2020).

3. Brock, A., Donahue, J., Simonyan, K.: Large scale gan training for high fidelity natural image synthesis. arXiv preprint arXiv:1809.11096 (2018).

4. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016).

5. Huang, J.B., Singh, A., Ahuja, N.: Single image super-resolution from transformed self-exemplars. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (6 2015).

6. Jolicoeur-Martineau, A.: The relativistic discriminator: a key element missing from standard gan. arXiv preprint arXiv:1807.00734 (2018).

7. Kingma, D.P., Salimans, T., Jozefowicz, R., Chen, X., Sutskever, I., Welling, M.: Improved variational inference with inverse autoregressive flow. Advances in neural information processing systems 29 (2016).

8. . Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K.: Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 136–144 (2017).

9. Vahdat, A., Kautz, J.: Nvae: A deep hierarchical variational autoencoder. arXiv preprint arXiv:2007.03898 (2020).

10. N. C. Rakotonirina and A. Rasoanaivo. Esrgan : Further improving enhanced super-resolution generative adversarial network. In ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 3637–3641, 2020.

11. ] M. Haris, G. Shakhnarovich, and N. Ukita. Deep backprojection networks for single image super-resolution. IEEE Transactions on Pattern Analysis and Machine Intelligence, pages 1–1, 2020.

12. Haitian Zheng, Mengqi Ji, Haoqian Wang, Yebin Liu, and Lu Fang. Crossnet: An end-to-end reference-based super resolution network using cross-scale warping. In Proceedings of the European Conference on Computer Vision (ECCV), pages 88–104, 2018.