Edge-Informed Single Image Super-Resolution

Kamyar Nazeri, Harrish Thasarathan, and Mehran Ebrahimi
University of Ontario Institute of Technology, Canada
kamyar.nazeri@uoit.ca harrish.thasarathan@uoit.net mehran.ebrahimi@uoit.ca
http://www.ImagingLab.ca

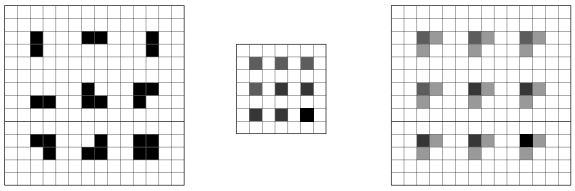
Abstract

The recent increase in the extensive use of digital imaging technologies has brought with it a simultaneous demand for higher-resolution images. We develop a novel "edge-informed" approach to single image super-resolution (SISR). The SISR problem is reformulated as an image inpainting task. We use a two-stage inpainting model as a baseline for super-resolution and show its effectiveness for different scale factors ($\times 2$, $\times 4$, $\times 8$) compared to basic interpolation schemes. This model is trained using a joint optimization of image contents (texture and color) and structures (edges). Quantitative and qualitative comparisons are included and the proposed model is compared with current state-of-the-art techniques. We show that our method of decoupling structure and texture reconstruction improves the quality of the final reconstructed high-resolution image.

1 Introduction

Super-Resolution (SR) is the task of inferring a high-resolution (HR) image from one or more given low-resolution (LR) images. SR plays an important role in various image processing tasks with direct applications in medical imaging, face recognition, satellite imaging, and surveillance [7]. Many existing SR methods reconstruct the HR image by fusing multiple instances of a LR image with different perspectives. These are called Multi-Frame Super-Resolution methods [8]. However, in most applications, only a single instance of the LR image is available from which missing HR information needs to be recovered. Single-Image Super-Resolution (SISR) is a challenging ill-posed inverse problem [6] that normally requires prior information to restrict the solution space of the problem [37].

We take inspiration from a recent image inpainting technique introduced by Nazeri *et al*. [29] to propose a novel approach to Single-Image Super-Resolution by reformulating the problem as an in-between pixels inpainting task. Increasing the resolution of a given LR image requires recovery of pixel intensities in between every two adjacent pixels. The missing pixel intensities can be considered as missing regions of an image inpainting problem. Our inpainting task is modelled as a two stage process that separates structural inpainting and textural inpainting to ensure high frequency information is preserved in the recovered HR image. The pipeline involves first creating a mask for every extra row and column that needs to be filled in the reconstruction of the HR image. The edge generation stage then focuses on "hallucinating" edges in missing regions, and the image completion stage uses the hallucinated edges as prior information to estimate pixel intensities in the missing regions.

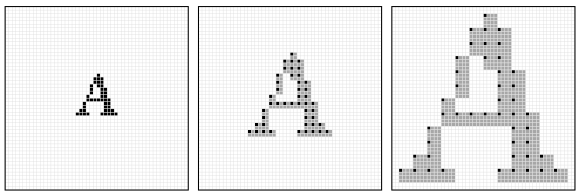


(a) Ground Truth (b) LR Image (c) HR Estimate

Figure 1: Schematic illustration of the super-resolution problem. (a) The ground truth image, (b) The image downsampled by a factor of two. Each four-pixel segment of information on the left turn into one pixel in the middle, as a result, the structure and orientation of edges are not distinguished anymore as the problem is ill-posed. (c) The reconstruction of a high-resolution image from one-pixel segments of information using bilinear interpolation. Most distinctive features in the original image are lost and the result is blurry around the edges.

2 Related Work

Many approaches to SISR have been presented in literature. These methods have been extensively organized by type ac-



(a) LR image (b) Upsample 2× (c) Upsample 4×

Figure 2: An illustration of the proposed inpainting-based method for SISR. (a) The original LR image. (b) Upsampling by a factor of two corresponds to interpolating one pixel between every two adjacent pixels. We add an extra empty row and column for every row and column in the ground truth image (shown in gray) which we fill by an inpainting process. (c) Upsampling by a factor of four corresponds to interpolating three pixels between every two adjacent pixels where we can add three extra empty rows and columns for every row and column in the ground truth image to be inpainted.

cording to their image priors in a study by Yang *et al*. [42]. **Prediction models** generate HR images through predefined mathematical functions. Examples include bilinear interpolation and bicubic interpolation [3], and Lanczos resampling [5]. **Edge-based methods** learn priors from features such as width of an edge [9], or parameter of a gradient profile [39] to reconstruct the HR image. **Statistical methods** exploit different image properties such as gradient distribution [36] to predict HR images. Patch-based methods use exemplar patches from external datasets [2, 11] or the image itself [19, 10] to learn mapping functions from LR to HR.

Deep Learning-based methods have achieved great performance on SISR using deep convolutional neural networks (CNN) with a per-pixel Euclidean loss [37, 4, 23]. Euclidean loss, however, is less effective to reconstruct high-frequency structures such as edges and textures. Recently Johnson *et al*. [21] proposed feed-forward CNN using a perceptual loss. In particular, they used a pre-trained VGG network [38] to extract high-level features from an image effectively separating content and style. Their model was trained with a joint optimization of *Feature reconstruction loss* and *Style reconstruction loss* and achieved state-of-the-art results on SISR for challenging $\times 8$ magnification factor. To encourage spatial smoothness and mitigate the checkerboard artifact [31] of using feature reconstruction loss, they introduced *total variation regularization* [33] to their model objective. Sajjadi *et al*. [35] proposed to use style loss in a patch-wise fashion to reduce the checkerboard artifact and enforce locally similar textures between the HR and ground truth images. They also used an adversarial loss to produce sharp results and further improve SISR results. Adversarial loss has also shown to be very effective in producing realistically synthesized high-frequency textures for SISR [25, 16, 32], however, the results of these GAN-based approaches tend to include less meaningful high-frequency noise around the edges that is unrelated to the input image [32]. Our work herein is inspired by the model proposed by Liu *et al*. [27] which extended their image inpainting framework to image super-resolution tasks by offsetting pixels and inserting holes. We present a SISR model that simultaneously improves structure, texture, and color to generate a photo-realistic high-resolution image.

3 Model

We propose a Single Image Super-Resolution framework based on a two stage adversarial model [15] consisting of an edge enhancement step and an image completion step. Both the edge enhancement and image completion steps consist of their own generator/discriminator pair that decouples SISR into two separate problems *i.e.* structure and texture. Let G_1 and D_1 be the generator and discriminator for the edge enhancement step, and G_2 and D_2 be the generator and discriminator for the image completion step. Our edge enhancement and image completion generators are built from encoders that downsample twice, followed by eight residual blocks [17], and decoders that upsample to the original input size. We use dilated convolutions in our residual layers. Our generators follow similar architectures to the method proposed by Johnson *et al*. [21] shown to achieve superior results for super-resolution [35, 14], image-to-image translation [45], and style transfer. Our discriminator follows the architecture of a 70×70 PatchGAN [20, 45] that classifies overlapping 70×70 image patches as real or fake. We use instance normalization [40] across all layers of the network, which normalizes across the spatial dimension to generate qualitatively superior images during training and at test time.

3.1 Edge Enhancement

Our edge enhancement stage boosts the edges obtained from a low-resolution image to yield a high-resolution edge map. Let $\mathbf{I}^{LR}$ and $\mathbf{I}^{HR}$ be the low-resolution and high-resolution images. Their corresponding edge maps will be denoted as $\mathbf{C}^{LR}$ and $\mathbf{C}^{HR}$ respectively and $\mathbf{I}^{LR}_{gray}$ is a grayscale counterpart of the low-resolution image. We add a nearest-neighbor interpolation module at the beginning of the network to resize the low-resolution image and its Canny edge-map to the same size as the HR image. The edge enhancement network G_1 predicts the high-resolution edge map

$$\mathbf{C}_{pred} = G_1(\mathbf{I}^{LR}_{gray}, \mathbf{C}^{LR}), \tag{1}$$

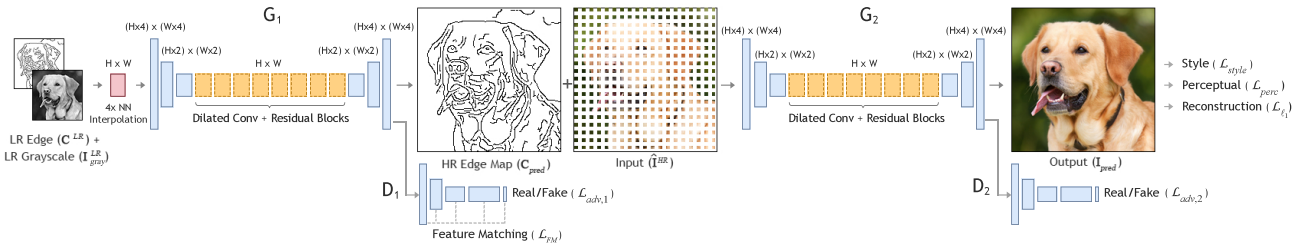


Figure 3: Summary of our proposed method. G_1 takes a low resolution greyscale image $\mathbf{I}^{LR}_{gray}$ and its corresponding low resolution edge map $\mathbf{C}^{LR}$ interpolated to the desired high resolution image size and outputs a high resolution edge map $\mathbf{C}_{pred}$. G_2 takes the high resolution edge map generated by G_1 as well as an incomplete HR image $\mathbf{I}_{gt}$ created by offsetting the pixels of the ground truth LR image using a fixed fractionally strided convolution kernel. The output is the high resolution image $\mathbf{I}_{pred}$.

where $\mathbf{I}^{LR}_{gray}$ and $\mathbf{C}^{LR}$ are the inputs to the network. The hinge variant [28] of the adversarial loss objective over the generator and discriminator are defined as

$$\mathcal{L}_{G_1} = -\mathbb{E}_{\mathbf{I}_{gray}}\left[D_1(\mathbf{C}_{pred}, \mathbf{I}_{gray})\right], \tag{2}$$

$$\begin{aligned}\mathcal{L}_{D_1} = {} & \mathbb{E}_{(\mathbf{C}_{gt}, \mathbf{I}_{gray})}\left[\max(0, 1 - D_1(\mathbf{C}_{gt}, \mathbf{I}_{gray}))\right] \\ & + \mathbb{E}_{\mathbf{I}_{gray}}\left[\max(0, 1 + D_1(\mathbf{C}_{pred}, \mathbf{I}_{gray}))\right].\end{aligned} \tag{3}$$

We also include a feature matching loss objective $\mathcal{L}_{FM}$ [41] to our edge enhancement generator which compares activation maps in the intermediate layers of the discriminator. This stabilizes the training process by forcing the generator to produce results with representations that are similar to real images. Perceptual loss [21, 13, 12] has also been known to accomplish this same task using a pretrained VGG network. However, since the VGG network is not trained to produce edge information, it fails to capture the result that we seek in the initial stage. The feature matching loss is defined as

$$\mathcal{L}_{FM} = \mathbb{E}\left[\sum_i \frac{1}{N_i}\left\|D_1^{(i)}(\mathbf{C}_{gt}) - D_1^{(i)}(\mathbf{C}_{pred})\right\|_1\right], \tag{4}$$

where N_i is the number of elements in the i'th activation layer, and $D_1^{(i)}$ is the activation in the i'th layer of the discriminator. Spectral normalization (SN) [28] further stabilizes training by scaling down weight matrices by their respective largest singular values, effectively restricting the Lipschitz constant of the network to one. Although this was originally proposed to be used only on the discriminator, recent works [43, 30] suggest that the generator can also benefit from SN by suppressing sudden changes of parameter and gradient values. We apply SN to both the generator and discriminator. The final joint loss objective for G_1 with regularization parameters λ_{G_1} and λ_{FM} thus becomes

$$\mathcal{J}_{G_1} = \lambda_{G_1}\mathcal{L}_{G_1} + \lambda_{FM}\mathcal{L}_{FM}, \tag{5}$$

where we choose $\lambda_{G_1} = 1$ and $\lambda_{FM} = 10$ for all experiments.

3.2 Image Completion

The image completion stage upscales the LR image to an incomplete HR image as input to G_2 using a fixed fractionally strided convolution kernel. This has the effect of adding empty rows and columns in-between pixels. To offset the pixels and increase the size of an image by a factor of s we use an $s \times s$ convolution kernel with stride of $1/s$. Let K denote a fixed strided convolution kernel and $\hat{\mathbf{I}}^{HR}$ represent the high-resolution image being constructed by offsetting the pixels from the LR image.

$$K_2 = \begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix} \qquad K_4 = \begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{bmatrix}$$

Figure 4: Fixed fractionally strided convolution kernels to offset the pixels of the LR image and create an incomplete HR image for $\times 2$ and $\times 4$ SISR factors.

$$\hat{\mathbf{I}}^{HR} = \mathbf{I}^{LR} * K. \tag{6}$$

The HR image is then generated using G_2:

$$\mathbf{I}_{(pred)} = G_2(\hat{\mathbf{I}}^{HR}, \mathbf{C}_{(pred)}). \tag{7}$$

We proceed to train G_2 with another joint loss consisting of an l_1 loss, hinge loss, perceptual loss, and style loss. The hinge variant of the adversarial loss follows equations 2 and 3

$$\mathcal{L}_{G_2} = -\mathbb{E}_{\mathbf{C}_{pred}}\left[D_2(\mathbf{I}_{pred}, \mathbf{C}_{pred})\right], \tag{8}$$

$$\begin{aligned}\mathcal{L}_{D_2} = {} & \mathbb{E}_{(\mathbf{I}_{gt}, \mathbf{C}_{pred})}\left[\max(0, 1 - D_2(\mathbf{I}_{gt}, \mathbf{C}_{pred}))\right] \\ & + \mathbb{E}_{\mathbf{C}_{pred}}\left[\max(0, 1 + D_2(\mathbf{I}_{pred}, \mathbf{C}_{pred}))\right].\end{aligned} \tag{9}$$

We include style loss $\mathcal{L}_{style}$ and perceptual loss $\mathcal{L}_{perc}$ [13, 21] in our joint loss objective to further supplement training. Perceptual loss minimizes the Manhattan distance between feature maps generated from intermediate layers of VGG-19 trained on the ImageNet dataset [34]. This has the effect of encouraging perceptually similar predictions with ground truth labels. Perceptual loss is defined as

$$\mathcal{L}_{perc} = \mathbb{E}\left[\sum_i \frac{1}{N_i} \left\|\phi_i(\mathbf{I}_{gt}) - \phi_i(\mathbf{I}_{pred})\right\|_1\right], \quad (10)$$

where N_i is the number of elements in the i'th activation of VGG-19. While perceptual loss encourages perceptual similarities between ground truth images and predictions, style loss encourages texture similarities by minimizing the Manhattan distance between the Gram matrices of the intermediate feature maps. The Gram matrix of feature map ϕ_i is represented by G_j^ϕ [13] and distributes spatial information of texture, shape, and style. Style loss is defined as

$$\mathcal{L}_{style} = \mathbb{E}\left[\sum_j \|G_j^\phi(\mathbf{I}_{gt}) - G_j^\phi(\mathbf{I}_{pred})\|_1\right]. \quad (11)$$

Style loss was shown by Sajjadi *et al*. [35] to successfully mitigate the "checkerboard" artifact caused by transpose convolutions [31]. For both style and perceptual loss we extract feature maps from `relu1_1`, `relu2_1`, `relu3_1`, `relu4_1` and `relu5_1` of VGG-19. We do not use feature matching loss in the image completion stage. While the feature matching loss is a regularizer to the adversarial loss in the edge generator, the perceptual loss used in this stage has the same effect while it is shown to be more effective loss for image generation tasks [29, 35, 21, 21]. Thus the complete joint loss objective is

$$\mathcal{J}_{G_2} = \lambda_{\ell_1}\mathcal{L}_{\ell_1} + \lambda_{G_2}\mathcal{L}_{G_2} + \lambda_p\mathcal{L}_{perc} + \lambda_s\mathcal{L}_{style}. \quad (12)$$

In all of our experiments we choose to train with parameters $\lambda_{\ell_1} = 1$, $\lambda_{G_2} = \lambda_p = 0.1$, and $\lambda_s = 250$ to effectively minimize the reconstruction, style, perceptual, and adversarial loss to generate a photo-realistic high-resolution image.

4 Experiments

4.1 Training Setup

To train G_1, we generate edge maps using Canny edge detector [1]. We can control the level of detail in the LR edge map by changing the Gaussian filter smoothing parameter σ. For our purposes, we found $\sigma \approx 2$ yields the best results. All of our experiments are implemented in PyTorch, with the HR images fixed at 512×512 and the LR input scaled accordingly based on the zooming factor. We choose a batch size of eight during training. The models of both stages were optimized using Adam optimizer [24] with $\beta_1 = 0$ and $\beta_2 = 0.9$. In our experiments, we didn't find any improvement by jointly optimizing G_1 and G_2, also we are limited to a smaller batch size due to the large memory footprint of the joint optimization, hence the generators from each stage are trained separately. We train G_1 using a learning rate of 10^{-4} with Canny edges until the loss plateaus. We lower the learning rate to 10^{-5} and continue training until convergence. We then freeze the weights of G_1 and continue to train G_2 with the same learning rates.

4.2 Datasets

Our proposed models are evaluated on the following publicly available datasets.

- Celeb-HQ [22]. High-quality version of the CelebA dataset with 30K images.
  https://github.com/tkarras/progressive_growing_of_gans
- Places2 [44]. More than 10 million images comprising 400+ unique scene categories.
  http://places2.csail.mit.edu/
- Set5, Set14, BSDS100, Urban100 [18]. Standard SISR evaluation datasets.
  http://vllab.ucmerced.edu/wlai24/LapSRN/

Results are compared against the current state-of-the-art methods both qualitatively and quantitatively.

4.3 Qualitative Evaluation

Figures 5 and 6 show results of the proposed SISR method for scale factors of $\times 4$ and $\times 8$ respectively. For visualization purposes, the LR image is resized using nearest-neighbor interpolation. All HR images are cropped at 512×512, which means the LR images are 128×128 and 64×64 for scale factors of $\times 4$ and $\times 8$ respectively. We obtain the LR images by blurring the HR with a Gaussian kernel of width $\sigma = 1$ followed by downsampling with the corresponding zooming scale factor. The results are compared against bicubic interpolation and our proposed model without the edge generation network as a baseline. Despite having almost high PSNR/SSIM, the baseline model produces blurry results around the edges while our full model (with edge-maps) remains faithful to the high-frequency edge data and produces sharp photorealistic images.

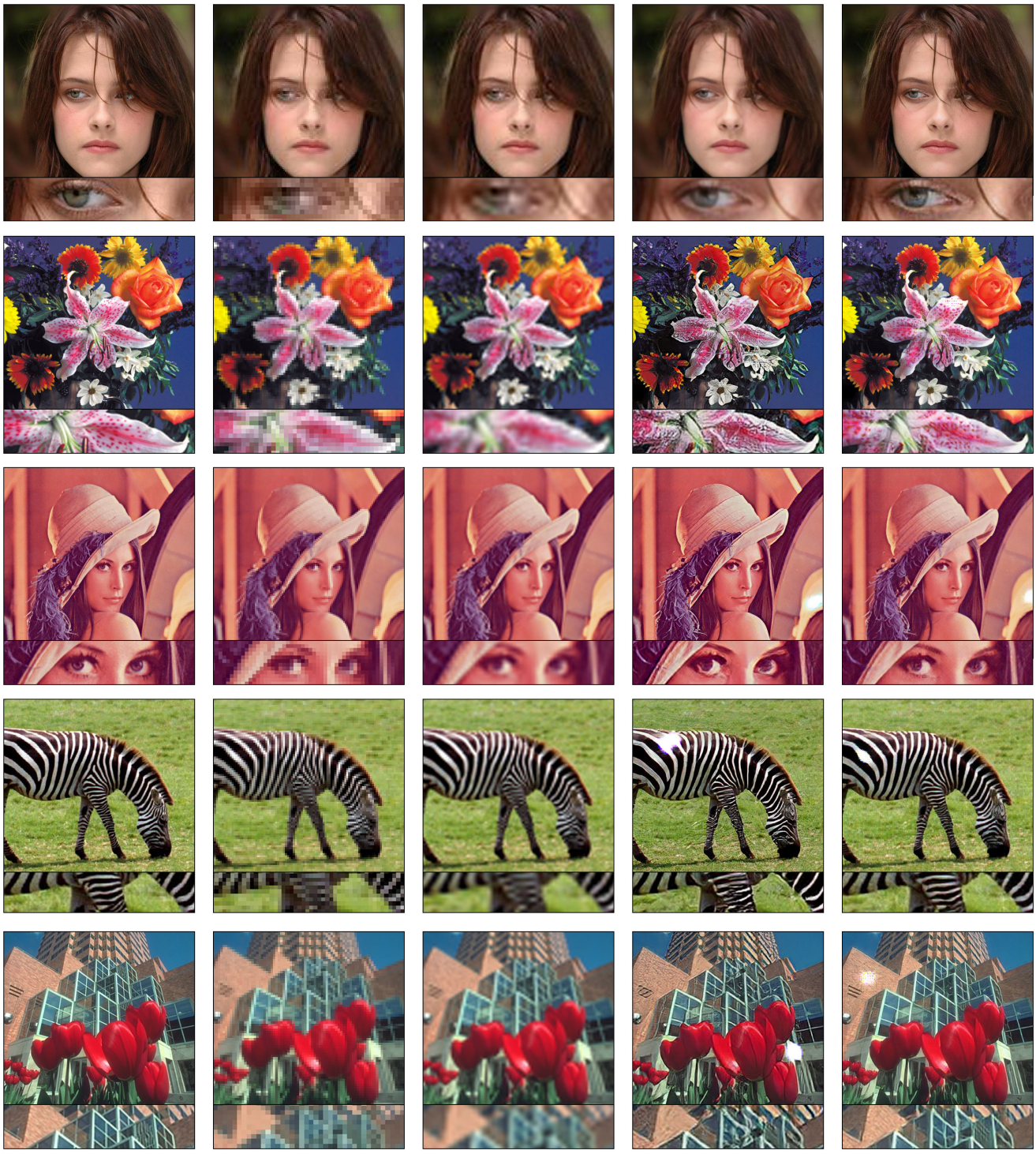


Figure 5: Comparison of qualitative results of images for $\times 4$ scale factor SISR cropped at 512×512. Left to right: Ground Truth HR, LR image upscaled using nearest-neighbor interpolation, SISR using bicubic interpolation, Baseline (no edge data), Ours (Full Model)

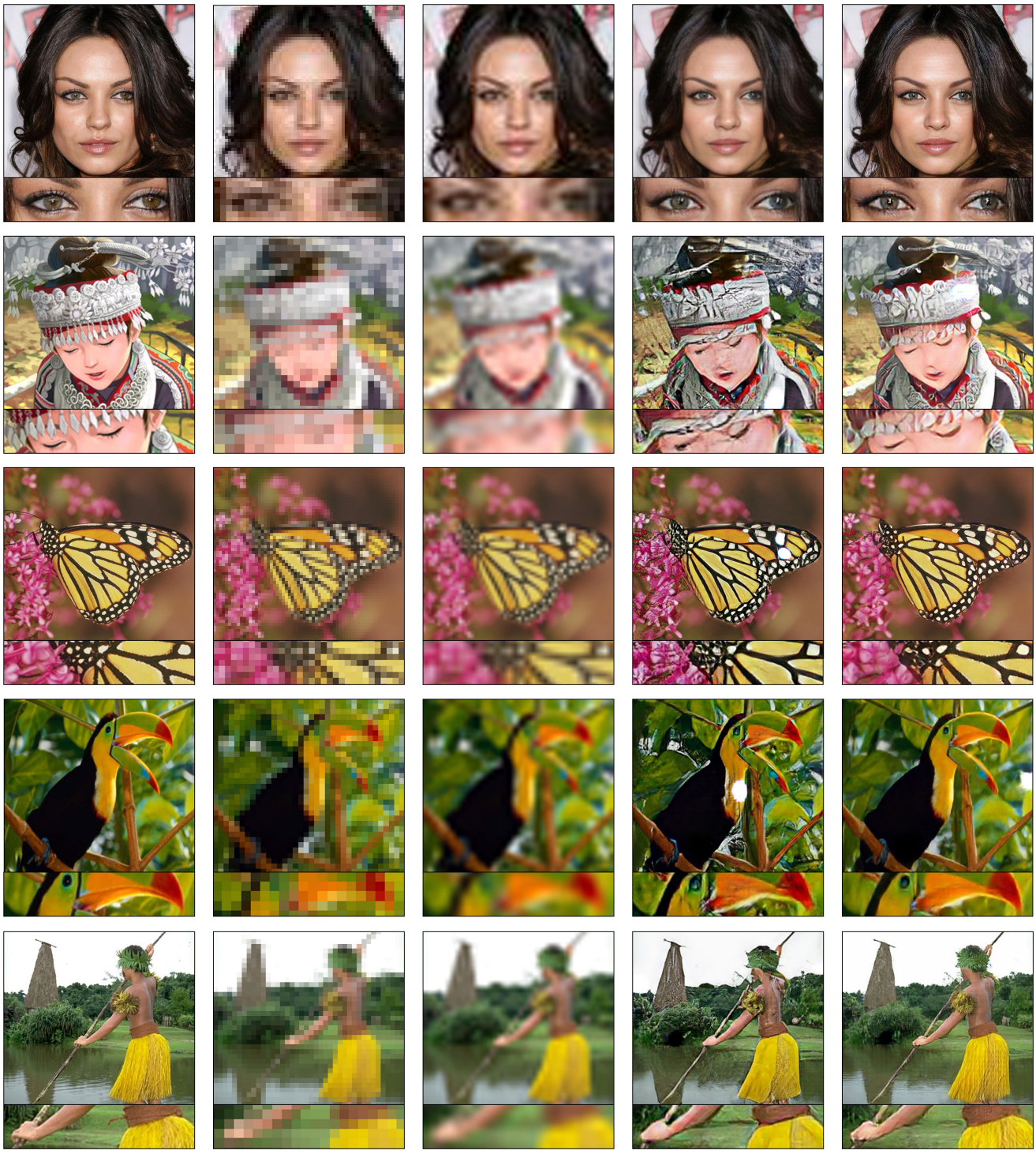


Figure 6: Comparison of qualitative results of images for $\times 8$ scale factor SISR cropped at 512×512. Left to right: Ground Truth HR, LR image upscaled using nearest-neighbor interpolation, SISR using bicubic interpolation, Baseline (no edge data), Ours (Full Model)

		Dataset	Bicubic	ENet	EDSR	Baseline	Ours
PSNR	×2	Set5	33.66	33.89	**38.20**	27.32	33.60
		Set14	30.24	30.45	**34.02**	24.86	29.24
		BSD100	29.56	28.30	**32.37**	23.97	28.12
		Celeb-HQ	**33.25**	-	-	31.33	32.12
	×4	Set5	28.42	28.56	**32.62**	24.22	28.59
		Set14	25.99	25.77	**28.94**	21.56	25.19
		BSD100	25.96	24.93	**27.79**	20.78	24.25
		Celeb-HQ	**29.59**	-	-	27.94	28.23
	×8	Set5	**23.80**	-	-	19.32	23.73
		Set14	**22.37**	-	-	18.47	21.44
		BSD100	**22.11**	-	-	18.65	21.63
		Celeb-HQ	**26.66**	-	-	25.46	25.56
SSIM	×2	Set5	0.930	0.928	0.961	0.974	**0.985**
		Set14	0.869	0.862	0.920	0.930	**0.954**
		BSD100	0.843	0.873	0.902	0.909	**0.932**
		Celeb-HQ	0.967	-	-	0.957	**0.968**
	×4	Set5	0.810	0.809	0.898	0.929	**0.965**
		Set14	0.703	0.678	0.790	0.832	**0.894**
		BSD100	0.668	0.627	0.744	0.773	**0.851**
		Celeb-HQ	0.834	-	-	0.910	**0.912**
	×8	Set5	0.646	-	-	0.801	**0.904**
		Set14	0.552	-	-	0.708	**0.793**
		BSD100	0.532	-	-	0.663	**0752**
		Celeb-HQ	0.782	-	-	0.841	**0.857**

Table 1: Comparison of PSNR and SSIM for $\times 2$, $\times 4$, and $\times 8$ factor SISR over **Set5**, **Set14**, **BSD100**, and **Celeb-HQ** datasets with bicubic interpolation, ENet [35], EDSR [26], and baseline (without edge-data). The best result of each row is boldfaced.

4.4 Quantitative Evaluation

We evaluate our model using PSNR and SSIM for $\times 2$, $\times 4$ and $\times 8$ SISR scale factors. Table 1 shows the performance of our model against bicubic interpolation and current state of the art SISR models over datasets Set5, Set14, BSD100, and Celeb-HQ. Statistics for competing models for $\times 2$ and $\times 4$ SR were obtained from their respective papers where available. Results for a challenging case of $\times 8$ are only compared against bicubic interpolation. Note that the PSNR in our results is lower than competing models. In particular, EDSR by Lim *et al*. [26] has achieved the best PSNR for every dataset. However, their model is only trained with per-pixel ℓ_1 loss and fails to reconstruct sharp edges despite having higher PSNR. Similar results in recent research [21, 35] show that PSNR favors smooth/blurry results.

4.5 Accuracy of Edge Generator

Table 2 shows the accuracy of our edge enhancer G_1 for Celeb-HQ and Places2 datasets for the Single Image Super-Resolution task. We measure precision and recall for various scale factors of SISR. In all experiments, the width of the Gaussian smoothing filter $\sigma = 2$ for Canny edge detection.

Figure 7: Comparison of edge prediction results for $\times 4$ scale factor SISR cropped at 512×512. Left to right: Ground Truth HR, HR edge-map, LR image upscaled using nearest-neighbor interpolation, LR edge-map upscaled using nearest-neighbor interpolation, $\times 4$ SISR, $\times 4$ predicted edge-map SISR.

	Scale	Precision	Recall
Celeb-HQ	$\times$**2**	74.27	73.21
	$\times$**4**	45.14	43.04
	$\times$**8**	23.23	19.09
Places2	$\times$**2**	79.18	80.24
	$\times$**4**	60.80	58.19
	$\times$**8**	31.06	23.93

Table 2: Quantitative performance of edge enhancer for Single Image Super-Resolution trained on Canny edges with $\sigma = 2$ for 512×512 images. Statistics are calculated over the standard test sets of each dataset.

Figure 7 shows results of the edge prediction stage for $\times 4$ scale factor. HR images are cropped at 512×512 and for visualization purposes, the LR image and its edge-map are resized using nearest-neighbor interpolation.

5 Discussion and Future Work

We propose a new structure-driven deep learning model for Single Image Super-Resolution (SISR) by recasting the problem as an in-between pixels inpainting task. One benefit of this approach over most deep-learning based SISR models is that we only have a unified model that can be used for different SISR zooming scales. Most deep-learning based SISR models take the LR image as input and generate the HR by in-network upsampling layers, given a zooming factor. For each different zooming factor, different network architecture and training is required. On the other hand, our model takes the LR image and adds empty space between pixels before using it as input to the network. Our proposed model learns to fill in the missing pixels by relying on available edge information to create the high-resolution image and effectively applies parameter sharing for different scales of SISR. Quantitative results show the effectiveness of the structure-guided inpainting model for the SISR problem where it achieves state-of-the-art results on standard benchmarks.

One shortcoming of the proposed inpainting-based SISR model is that it requires minimizing two disjoint optimizing algorithms. A better approach is to incorporate the edge generation stage into the inpainting model's objective. This model could be trained using a joint optimization of image contents and structures and potentially outperform the disjoint two-stage optimization algorithm computationally while preserving sharp details of the image.

Our method leads to an interesting direction, which raises the question that what other information could be learned from the original dataset to help the super-resolution process. Our source code is available at:
https://github.com/knazeri/edge-informed-sisr

Acknowledgments

This research was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC). We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan V GPU used for this research.

References

[1] J. Canny. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, pages 679–698, 1986. 4

[2] H. Chang, D.-Y. Yeung, and Y. Xiong. Super-resolution through neighbor embedding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2004. 2

[3] C. De Boor. *A practical guide to splines*, volume 27. springer-verlag New York, 1978. 2

[4] C. Dong, C. C. Loy, K. He, and X. Tang. Learning a deep convolutional network for image super-resolution. In *European conference on computer vision*, pages 184–199. Springer, 2014. 2

[5] C. E. Duchon. Lanczos filtering in one and two dimensions. *Journal of applied meteorology*, 18(8):1016–1022, 1979. 2

[6] M. Ebrahimi and E. R. Vrscay. Solving the inverse problem of image zooming using self-examples. In *International Conference Image Analysis and Recognition*, pages 117–130. Springer, 2007. 1

[7] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar. Advances and challenges in super-resolution. *International Journal of Imaging Systems and Technology*, 14(2):47–57, 2004. 1

[8] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar. Fast and robust multiframe super resolution. *IEEE transactions on image processing*, 13(10):1327–1344, 2004. 1

[9] R. Fattal. Image upsampling via imposed edge statistics. *ACM transactions on graphics (TOG)*, 26(3):95, 2007. 2

[10] G. Freedman and R. Fattal. Image and video upscaling from local self-examples. *ACM Transactions on Graphics (TOG)*, 30(2):12, 2011. 2

[11] W. T. Freeman, T. R. Jones, and E. C. Pasztor. Example-based super-resolution. *IEEE Computer graphics and Applications*, (2):56–65, 2002. 2

[12] L. Gatys, A. S. Ecker, and M. Bethge. Texture synthesis using convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 262–270, 2015. 3

[13] L. A. Gatys, A. S. Ecker, and M. Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2414–2423, 2016. 3, 4

[14] M. W. Gondal, B. Schölkopf, and M. Hirsch. The unreasonable effectiveness of texture transfer for single image super-resolution. In *Workshop and Challenge on Perceptual Image Restoration and Manipulation (PIRM) at the 15th European Conference on Computer Vision (ECCV)*, 2018. 2

[15] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014. 2

[16] M. Haris, G. Shakhnarovich, and N. Ukita. Deep back-projection networks for super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1664–1673, 2018. 2

[17] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. 2

[18] J.-B. Huang, A. Singh, and N. Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5197–5206, 2015. 4

[19] D. G. S. B. M. Irani. Super-resolution from a single image. In *Proceedings of the IEEE International Conference on Computer Vision, Kyoto, Japan*, 2009. 2

[20] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 2

[21] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision (ECCV)*, pages 694–711. Springer, 2016. 2, 3, 4, 7

[22] T. Karras, T. Aila, S. Laine, and J. Lehtinen. Progressive growing of GANs for improved quality, stability, and variation. In *International Conference on Learning Representations*, 2018. 4

[23] J. Kim, J. Kwon Lee, and K. Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016. 2

[24] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015. 4

[25] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In

Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 4681–4690, 2017. 2

[26] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017. 7

[27] G. Liu, F. A. Reda, K. J. Shih, T.-C. Wang, A. Tao, and B. Catanzaro. Image inpainting for irregular holes using partial convolutions. In *European Conference on Computer Vision (ECCV)*, September 2018. 2

[28] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida. Spectral normalization for generative adversarial networks. In *International Conference on Learning Representations*, 2018. 3

[29] K. Nazeri, E. Ng, T. Joseph, F. Qureshi, and M. Ebrahimi. Edgeconnect: Generative image inpainting with adversarial edge learning. *arXiv preprint arXiv:1901.00212*, 2019. 1, 4

[30] A. Odena, J. Buckman, C. Olsson, T. B. Brown, C. Olah, C. Raffel, and I. Goodfellow. Is generator conditioning causally related to gan performance? In *Proceedings of the 35th International Conference on Machine Learning*, 2018. 3

[31] A. Odena, V. Dumoulin, and C. Olah. Deconvolution and checkerboard artifacts. *Distill*, 1(10):e3, 2016. 2, 4

[32] S.-J. Park, H. Son, S. Cho, K.-S. Hong, and S. Lee. Srfeat: Single image super-resolution with feature discrimination. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 439–455, 2018. 2

[33] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60:259–268, 1992. 2

[34] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015. 4

[35] M. S. M. Sajjadi, B. Scholkopf, and M. Hirsch. Enhancenet: Single image super-resolution through automated texture synthesis. In *The IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2017. 2, 4, 7

[36] Q. Shan, Z. Li, J. Jia, and C.-K. Tang. Fast image/video upsampling. In *ACM Transactions on Graphics (TOG)*, volume 27, page 153. ACM, 2008. 2

[37] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Conference on Computer Vision and Pattern Recognition*, 2016. 1, 2

[38] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014. 2

[39] J. Sun, Z. Xu, and H.-Y. Shum. Image super-resolution using gradient profile prior. In *Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008. 2

[40] D. Ulyanov, A. Vedaldi, and V. Lempitsky. Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 2

[41] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, page 5, 2018. 3

[42] C.-Y. Yang, C. Ma, and M.-H. Yang. Single-image super-resolution: A benchmark. In *European Conference on Computer Vision*, pages 372–386. Springer, 2014. 2

[43] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena. Self-attention generative adversarial networks. *arXiv preprint arXiv:1805.08318*, 2018. 3

[44] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba. Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017. 4

[45] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *The IEEE International Conference on Computer Vision (ICCV)*, 2017. 2