Image Super Resolution Using Generative Adversarial Networks and Local Saliency Maps for Retinal Image Analysis

Dwarikanath Mahapatra(✉), Behzad Bozorgtabar, Sajini Hewavitharanage, and Rahil Garnavi

IBM Research Australia, Melbourne, Australia
{dwarim,sydb,sajinihe,rahilgar}@au1.ibm.com

Abstract. We propose an image super resolution (ISR) method using generative adversarial networks (GANs) that takes a low resolution input fundus image and generates a high resolution super resolved (SR) image upto scaling factor of 16. This facilitates more accurate automated image analysis, especially for small or blurred landmarks and pathologies. Local saliency maps, which define each pixel's importance, are used to define a novel saliency loss in the GAN cost function. Experimental results show the resulting SR images have perceptual quality very close to the original images and perform better than competing methods that do not weigh pixels according to their importance. When used for retinal vasculature segmentation, our SR images result in accuracy levels close to those obtained when using the original images.

1 Introduction

Normal retinal fundus images have high resolution to detect and segment prominent landmarks and pathologies, but not sufficient for small and indistinct pathologies (microaneurysms, haemorrhages) and smaller vessel branches. Image super resolution (ISR) produces highly accurate super resolved (SR) images from single field of view (FOV) images that are comparable with the original HR images. This is particularly relevant for tele-ophthalmology requiring transmission of acquired LR images. SR images improves detection of neovascularization, segmentation of small vessel branches and small pathologies not visible in the original LR images. Ophthalmologist may also use it to closely analyze suspicious regions with minute abnormalities.

Medical ISR methods using examples [12] and self similarity [9] were reliant on external data which put them at a disadvantage. Subsequently, parametric generative models learned the mapping between the original and LR version to upscale MR brain [13] and cardiac [3] images. These approaches are computationally demanding as candidate patches are searched in the training dataset to find the most suitable HR sample. Other methods using random forest regressors [17] convolutional neural networks (CNNs) [4] produce high quality images.

M. Descoteaux et al. (Eds.): MICCAI 2017, Part III, LNCS 10435, pp. 382–390, 2017.
DOI: 10.1007/978-3-319-66179-7_44

Recent work using Fourier burst accumulation [7], generative adversarial networks (GANs) [8] and CNNs [10] also highlight the importance of ISR for medical image analysis problems.

GANs [8] are state-of-the-art for ISR primarily due to the ResNet based generator architecture, but are less effective with retinal images for scaling factors greater than 4 due to cost functions that do not explicitly include local structure information. To overcome this limitation we propose a image SR method based on GANs that has the following novelties in its cost function: (1) using local saliency maps computed from curvature maps (that highlight local structures) and (2) entropy filtering (to highlight compact regions). Together they outperform [8] for $4-16$x magnification and the resulting SR images preserve information content and perceptual information of the LR image. Our method's effectiveness is demonstrated in segmenting the retinal vasculature of SR images.

2 Saliency Map Calculation

Existing saliency methods highlight a globally salient region while 'local' saliency maps are essential to compute individual pixel importance values. Inspired by Perazzi et al. [11] we combining abstraction, element distribution and uniqueness to generate a local saliency map for retinal images.

Abstraction: Using superpixels for abstraction (as in [11]) provides high level global information. Instead curvature maps (I_{curv}) capture local structural information based on vessel curvature and other small elements.

$$I_{Curv} = \frac{f_{xx}f_y^2 + f_{yy}f_x^2 - 2f_{xy}f_xf_y}{\left(f_x^2 + f_y^2\right)^{3/2}}, \tag{1}$$

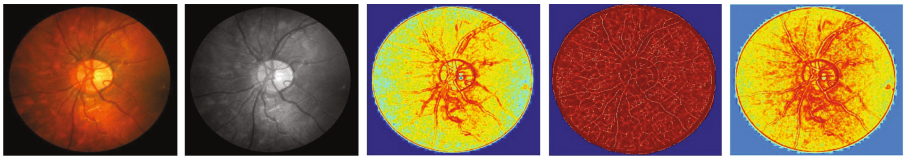
where f_x, f_y are image gradients and f_{xx}, f_{yy} indicate second derivatives.

Element Distribution: Element distribution captures compactness and continuity of retinal structures. Pixel (s) entropy is given by $I_{Ent}(s) = -\sum_{i \in N_s} p_i \log p_i$. I_{Ent} is the entropy image, $N_s = 7 \times 7$ is s's neighborhood and determines compactness, p_i is the probability of intensity i in N_s calculated using a 8 bin histogram (experimentally determined). I_{Ent} assigns low values for regions with compact objects. Hence its values are normalized to $[0, 1]$ and transformed as $1 - I_{Ent}$ to highlight compact regions. I_{Ent} is smoothed using a Gaussian low pass filter of size 3 and standard deviation 0.5 to remove any isolated noisy regions or pixel clusters.

Uniqueness: The sum of weighted difference of pixel feature maps is,

$$D_F(s) = \sum_i \exp\left(-\|s - s_i\|\right) |F(s) - F(s_i)|, \tag{2}$$

where D_F indicates the difference or uniqueness map for feature map F (I_{Curv} or $1 - I_{Ent}$); $\|s - s_i\|$ is the Euclidean distance between s and its ith neighbor s_i.



(a) (b) (c) (d) (e)

Fig. 1. (a) Original fundus image (b) grayscale image; (c) $1 - I_{Ent}$; (d) I_{Curv}; (e) local saliency map I_{Sal} at scale 7×7. Warmer colours indicate higher values

Squared difference of feature values in [11] introduces blur which is undesirable for image super resolution. Instead we use the *absolute difference* and also normalize the difference map to $[0, 1]$. The final saliency map is,

$$I_{Sal} = w_1 \times D_{Curv} \; + \; (1 - w_1) \times D_{1-Ent}. \tag{3}$$

where w_1 balances the relative contribution of each feature. $w_1 = 0.4$ was experimentally set by varying it between $[0, 1]$ in steps of 0.01 and comparing the quality of the resulting SR images on a subset of 50 images. The resulting saliency map in Fig. 1(e) clearly highlights the local retinal structures and thus justifies its use in the GAN cost function.

3 Generative Adversarial Networks

ISR estimates a high-resolution, super resolved image I^{SR} from a low-resolution input image I^{LR}. For training, I^{LR} is the low-resolution version of the high resolution counterpart I^{HR}, obtained by applying a Gaussian filter to I_{HR} followed by downsampling with factor $\mathbf{r}$. The generator network is a feed-forward CNN (G_{θ_G}) whose parameters $\theta_G = W_{1:L}; b_{1:L}$ are obtained by,

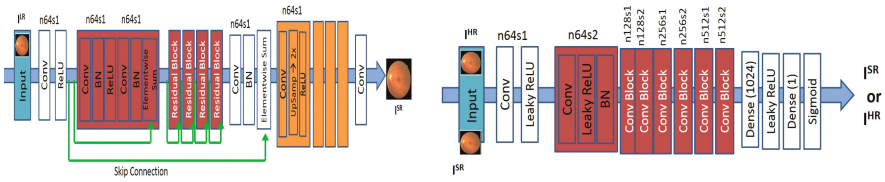
$$\widehat{\theta} = \arg\min_{\theta_G} \frac{1}{N} \sum_{n=1}^{N} l^{SR} \left(G_{\theta_G}(I_n^{LR}), I_n^{HR} \right), \tag{4}$$

where l^{SR} is the loss function and I_n^{HR}, I_n^{LR} are HR and LR images. The adversarial min-max problem is defined by,

$$\min_{\theta_G} \max_{\theta_D} \mathop{\mathbb{E}}_{I^{HR} \; p_{train}(I^{HR})} [\log D_{\theta_D}(I^{HR})] + \mathop{\mathbb{E}}_{I^{LR} \; p_G(I^{LR})} [\log(1 - D_{\theta_D}(G_{\theta_G}(I^{LR}))] \tag{5}$$

This trains a generative model G with the goal of fooling a differentiable discriminator D that is trained to distinguish SR images from real images. G creates solutions that are very similar to real images and thus difficult to classify by D. This encourages perceptually superior solutions than obtained by minimizing pixel-wise mean square error (MSE). G employs residual blocks (Fig. 2(a)). Each block has two convolutional layers with 3×3 filters and 64 feature maps, followed by batch normalization and ReLU activation.

D solves the maximization problem in Eq. 5. It has eight convolutional layers with the kernels increasing by a factor of 2 from 64 to 512 (Fig. 2(b)). Leaky ReLU is used and strided convolutions reduce the image dimension when the number of features is doubled. The resulting 512 feature maps are followed by two dense layers and a final sigmoid activation to obtain a probability map.



(a) (b)

Fig. 2. (a) Generator Network; (b) Discriminator network. $n64s1$ denotes 64 feature maps (n) and stride (s) 1 for each convolutional layer.

3.1 Loss Function

l^{SR} is a combination of content loss (l^{SR}_{Cont}) and adversarial or generative loss (l^{SR}_{Gen}), balanced by a factor $\alpha = 0.01$, and is given by:

$$l^{SR} = l^{SR}_{Cont} + \alpha l^{SR}_{Gen} \tag{6}$$

Content Loss: MSE results in smooth SR images that are perceptually unsatisfying and lack high frequency content. Perceptually important details in SR images is preserved by our saliency weighted MSE loss (l_{w-MSE}),

$$l_{w-MSE} = \frac{1}{WH} \sum_{x=1}^{W} \sum_{y=1}^{H} \times (w^{HR}_I I^{HR}_{x,y} - w^{SR}_I G_{\theta_G}(I^{LR})_{x,y})^2, \tag{7}$$

where w^{HR}_I, w^{SR}_I are saliency values of HR (I^{HR}) and SR ($G_{\theta_G}(I^{LR})$) images. A CNN loss [8] is calculated as the $L2$ distance between SR image and ground-truth HR image using all 512 feature maps of Relu $4-1$ layer of a pre-trained $VGG-16$ [15].

$$l^{SR}_{CNN} = \frac{1}{W_{i,j} H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}(G_{\theta_G}(I^{LR}))_{x,y})^2 \tag{8}$$

$\phi_{i,j}$ the feature map obtained by the $j-$th convolution (after activation) before the $i-$th max pooling layer and $W_{i,j}$ and $H_{i,j}$ are the dimensions of ϕ.

Local Saliency Loss: Our novel saliency loss measures the difference in saliency maps of $G_{\theta_G}(I^{LR})$) and I^{HR} by comparing their local landmarks. This enables inclusion of important structural information in the cost function.

$$l^{SR}_{Sal} = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} ((I^{HR}_{Sal})_{x,y} - (G_{\theta_G}(I^{LR})_{Sal})_{x,y})^2 \tag{9}$$

I^{HR}_{Sal} and $G_{\theta_G}(I^{LR})_{Sal}$ denote the saliency maps of I^{HR} and $G_{\theta_G}(I^{LR})$.

Adversarial Loss: The generative loss l^{SR}_{Gen} [8] over all training samples is

$$l^{SR}_{Gen} = \sum_{n=1}^{N} -\log D_{\theta_D}(G_{\theta_G}(I^{LR})) \tag{10}$$

$D_{\theta_D}(G_{\theta_G}(I^{LR}))$ is probability that $G_{\theta_G}(I^{LR})$ is a natural HR image. This network favours solutions in the manifold of retinal images. Convergence is facilitated by minimizing $-\log D_{\theta_D}(G_{\theta_G}(I^{LR}))$ instead of $-\log[1 - D_{\theta_D}(G_{\theta_G}(I^{LR}))]$.

Training: The model in Fig. 2(a) generates images upscaled by 2×. For higher scale factors we feed the output of this network to another identical network to get 4× images. Consecutive networks can be combined to get 8×, 16×, etc. image resolution. In each subsequent upsampling step the corresponding networks are trained on images of size 4×, 8×, etc.

4 Experiments and Results

Dataset: We apply our algorithm on 5000 retinal fundus images from multiple sources with different image dimensions [2], and augmented 100 times by rotation and translation. The dark borders were removed and the images resized to 1024×1024 pixels. Our method was implemented with Python and TensorFlow (for GANs). For GAN optimization we use Adam with $\beta_1 = 0.93$ and batch normalization. The ResNet was trained with a learning rate of 0.001 and 10^5 update iterations. MSE based ResNet was used to initialize G. The final GAN was trained with 10^5 update iterations at learning rate 10^{-3}. The average training time using the augmented version from 4000 images was 14 h for scaling factor(r) 2, 26 h for $r = 4$, and 40 h for $r = 8$. Time to generate a super resolved image is 1 ms for $r = 2$, 1.4 ms for $r = 4$, and 1.9 ms for $r = 8$. Training and test was performed on a NVIDIA Tesla K40 GPU with 12 GB RAM.

4.1 Image Super Resolution Results

The following ISR methods are used for comparison: (1) $SRGAN_{Ledig}$: - the baseline GAN using MSE and CNN loss [8]; (2) $SRGAN_{Sal}$: - our proposed method; (3) $SRCNN$:- CNN based method of [4]; (4) $SR-RF$:- random forest based method of [14]; (5) SSR :- self super resolution method of [7]. Resized

Table 1. Comparative results of different methods for image super resolution.

	Scaling factor(r) = 4					Scaling factor (r) = 8					r = 16
	SSIM	RMSE (10^{-6})	PSNR dB	S3	p	SSIM	RMSE (10^{-6})	PSNR dB	S3	p	SSIM
$SRGAN_{Sal}$	0.89	6.2	44.3	0.83	-	0.84	7.5	39	0.74	-	0.80
$SRGAN_{Ledig}$	0.78	8.1	36.4	0.65	< 0.001	0.73	9.3	31	0.60	< 0.001	0.69
$SRCNN$	0.75	9.1	34.3	0.61	< 0.009	0.67	10.9	28	0.57	< 0.001	0.64
SR-RF	0.71	10.3	30.2	0.57	< 0.009	0.62	12.3	25	0.55	< 0.001	0.59
SSR	0.67	11.2	27.1	0.54	< 0.001	0.60	13.7	22	0.21	< 0.001	0.56

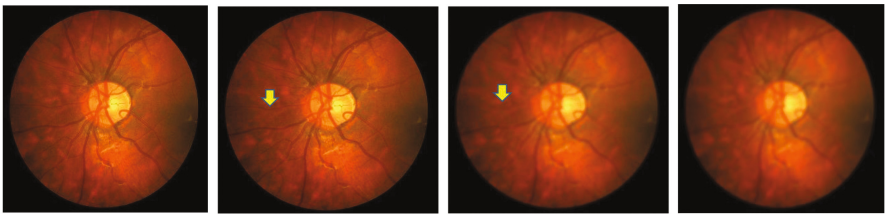
1024×1024 images are ground truth HR images, I^{HR}, which are downsampled by different r to obtain I^{LR} from which I^{SR} are generated. $Y-$channel images of I^{HR} and I^{SR} are used to compute: (1) peak signal to noise ratio (PSNR); (2) structural similarity (SSIM) [19]; (3) $S3$ - the sharpness metric of [18]; and (4) root mean square error (RMSE). Higher values of $1, 2, 3$ and lower values of 4 indicate better performance.

Results of $5-$fold cross validation for $r = 4, 8$ are presented in Table 1. Due to space constraints only SSIM values are shown for $r = 16$. For $r = 2$, performance difference of all methods is small but becomes more pronounced for higher r. $SRGAN_{Sal}$ gives the best results for all r, and the improvement over competing methods is significant as is evident from the $p-$values of Wilcoxon signed-rank tests. Figure 3 shows results of the top 3 methods (due to space constraints) for $r = 4$. $SRGAN_{Sal}$ shows the best performance as is evident from the SR image in Fig. 3(b) where one of the minor retinal branches (indicated by yellow arrow) is clearly visible. On the other hand the SR image by $SRGAN_{Ledig}$ (Fig. 3(c)) is blurry and does not clearly show this retinal branch. Other methods perform much worse, with significant blur visible for the main branches as well. Clearly, $SRGAN_{Sal}$ gives the closest reconstruction to the HR image of Fig. 3(a).

Importance of Saliency Maps: Excluding l^{SR}_{CNN}, and using l^{SR}_{Sal} and l_{w-MSE} for $r = 4$ gives SSIM = 0.81, RMSE = 6.9, PSNR = 38.6 dB, and S3 = 0.69. They are slightly higher than $SRGAN_{Ledig}$, indicating local saliency maps alone perform better than l^{SR}_{CNN} and MSE in preserving image information. Combining local saliency information with CNN loss significantly improves SR image quality. Using either curvature (SSIM = 0.82) or entropy (SSIM = 0.83) for $r = 4$ lowers performance, thus highlighting their individual importance in the final saliency map. Using saliency maps of [11] gave SSIM $= 0.80, 0.74, 0.70$ for $r = 4, 8, 16$. Our proposed local saliency maps outperforms [11] (a global saliency map) as it does not capture fine structural information in retinal fundus images.

4.2 Retinal Blood Vessel Segmentation Results

We present retinal vessel segmentation results on the DRIVE [16], STARE [6] and CHASE_DB1 [5] datasets with 40, 20 and 28 images respectively. Original images and manual annotations (I^{HR}) are downsampled by $r = 4, 8$ to get I^{LR}



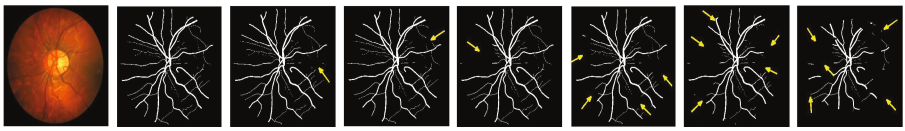
(a) (b) (c) (d)

Fig. 3. Super resolution results for $r = 4$. (a) original HR; SR images from: (b) $SRGAN_{Sal}$; (c) $SRGAN_{Ledig}$; and (d) $SRCNN$.

and 5 sets of I^{SR} from 5 methods trained on [2]. I^{HR} and the 5 sets of I^{SR} were used to train 6 different state-of-the art U-Nets for vasculature segmentation [1]. The average accuracy (Acc) and sensitivity (Sen) for $r = 4, 8$ is summarized in Table 2. Better ISR methods should give higher vessel segmentation accuracy and performance of I^{HR} gives a lower bound on the segmentation error. $SRGAN_{Sal}$'s performance is closest to I^{HR}, and establishes its superiority over all competing methods. Figures 4(a)–(h) show results of vessel segmentation on an example image. $SRGAN_{Sal}$'s performance is most similar to I^{HR} as is evident from the areas of inaccurate segmentation highlighted by yellow arrows. Most of the methods do not segment the finer vasculature structures, while SSR and SR-RF are unable to segment some of the major arteries.

Table 2. Comparative vasculature segmentation results of different super resolution methods. The values are for scaling factor 4 and 8.

	HR		$SRGAN_{Sal}$		$SRGAN_{Ledig}$		$SRCNN$		SR-RF		SSR	
	Acc	Sen	Acc	Sen	Acc	Sen	Acc	Sen	Acc	Sen	Acc	Sen
DRIVE	0.98	0.79	0.96	0.77	0.92	0.74	0.89	0.73	0.87	0.70	0.85	0.69
STARE	0.98	0.90	0.96	0.87	0.91	0.84	0.89	0.81	0.86	0.77	0.83	0.72
CHASE_DB	0.97	0.84	0.95	0.82	0.91	0.76	0.87	0.72	0.85	0.70	0.81	0.68



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 4. Results for retinal vessel segmentation; (a) retinal image; (b) manual ground truth mask; results obtained when training on (c) orginal HR images; SR images by (d) $SRGAN_{Sal}$; (e) $SRGAN_{Ledig}$; (f) $SRCNN$; (g) SR-RF; (h) SSR. Yellow arrows highlight regions of inaccurate segmentation.

5 Conclusion

We have proposed a novel method for super resolution of retinal fundus images based on GANs. Local saliency maps effectively quantify a pixel's perceptual relevance, and are used to weight each pixel according to its importance and define a novel saliency loss. When incorporated into the GAN loss function, the resulting SR images are better than those obtained using CNN feature loss. Experimental results show combination of saliency and CNN loss significantly outperforms current state of the art GANs and other competing ISR methods. The resulting super resolved images can be used to increase the size and resolution of low dimensional images, and then apply different image analysis algorithms. When using our SR images for retinal vessel segmentation the results are close to those obtained with the original high resolution images. Our method can be applied for other medical images as well.

References

1. https://github.com/orobix/retina-unet
2. http://www.eyepacs.com/
3. Bhatia, K., Price, A., Shi, W., Hajnal, J., Rueckert., D.: Super-resolution reconstruction of cardiac mri using coupled dictionary learning. In: Proceedings of ISBI, pp. 947–950 (2014)
4. Dong, C., Loy, C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE Trans. Patt. Anal. Mach. Intell. **38**(2), 295–307 (2016)
5. Fraz, M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A., Owen, C., Barman, S.: An ensemble classification-based approach applied to retinal blood vessel segmentation. IEEE Trans. Biomed. Engg. **59**(9), 2538–2548 (2012)
6. Hoover, A., Kouznetsova, V., Goldbaum, M.: Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. IEEE Trans. Med. Imag. **19**(3), 203–210 (2000)
7. Jog, A., Carass, A., Prince, J.L.: Self super-resolution for magnetic resonance images. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9902, pp. 553–560. Springer, Cham (2016). doi:10.1007/978-3-319-46726-9_64
8. Ledig, C., et. al.: Photo-realistic single image super-resolution using a generative adversarial network. CoRR abs/1609.04802 (2016)
9. Manjon, J., Coupe, P., Buades, A., Fonov, V., Collins, D., Robles, M.: Non-local MRI upsampling. Med. Image Anal. **14**(6), 784–792 (2010)
10. Oktay, O., et al.: Multi-input cardiac image super-resolution using convolutional neural networks. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9902, pp. 246–254. Springer, Cham (2016). doi:10.1007/978-3-319-46726-9_29
11. Perazzi, F., Kraehenbuhl, P., Pritch, Y., Hornung, A.: Saliency filters: contrast based filtering for salient region detection. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 733–740 (2012)
12. Rousseau, F.: Brain hallucination. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008. LNCS, vol. 5302, pp. 497–508. Springer, Heidelberg (2008). doi:10.1007/978-3-540-88682-2_38

13. Rueda, A., Malpica, N., Romero, E.: Single-image super-resolution of brain mr images using overcomplete dictionaries. Med. Image Anal. **17**(1), 113–132 (2013)
14. Schulter, S., Leistner, C., Bischof., H.: Fast and accurate image upscaling with super-resolution forests. In: Proceedings of CVPR, pp. 3791–3799 (2015)
15. Simonyan, K., Zisserman., A.: Very deep convolutional networks for large-scale image recognition. CoRR abs/1409.1556 (2014)
16. Staal, J., Abramoff, M., Niemeijer, M., Viergever, M., Ginneken, B.V.: Ridge based vessel segmentation in color images of the retina. IEEE Trans. Med. Imag. **23**(4), 501–509 (2004)
17. Tanno, R., Ghosh, A., Grussu, F., Kaden, E., Criminisi, A., Alexander, D.C.: Bayesian image quality transfer. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 265–273. Springer, Cham (2016). doi:10.1007/978-3-319-46723-8_31
18. Vu, C., Phan, T., Chandler, D.: S3: a spectral and spatial measure of local perceived sharpness in natural images. IEEE Trans. Imag. Proc. **21**(3), 934–945 (2012)
19. Wang, Z., et al.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Imag. Proc. **13**(4), 600–612 (2004)