CT Super-resolution GAN Constrained by the Identical, Residual, and Cycle Learning Ensemble (GAN-CIRCLE) - Supplementary Material

Chenyu You, Guang Li, Yi Zhang*, *Senior Member, IEEE*, Xiaoliu Zhang, Hongming Shan, Mengzhou Li, Shenghong Ju, Zhen Zhao, Zhuiyang Zhang, Wenxiang Cong, Michael W. Vannier, *Member, IEEE*, Punam K. Saha, *Senior Member, IEEE*, Eric A. Hoffman, *Member, IEEE*, and Ge Wang*, *Fellow, IEEE*

***Abstract*—In this supplementary material, we first provide the detailed network architecture of the generative networks and the discriminator in Section I-A. The details of implementations are illustrated in Section I-B. We then introduce the datasets used in our training and testing in Section II. Next, the analysis of the influence of filter size, number of layers, and training patch size is depicted in Section III. Finally, we provide more visual comparisons with state-of-the-art SR algorithms in Section IV.**

I. Methods

A. Network Architecture

*1) **Generative Networks**:* Although more layers and larger model size usually result in the performance gain, for real application we designed a lightweight model to validate the effectiveness of GAN-CIRCLE. The two generative networks G and F are shown in Fig. 1. The network architecture has been optimized for SR CT imaging. It consists of two processing streams: the feature extraction network and the reconstruction network.

In the ***feature extraction network***, we concatenate 12 sets of non-linear SR feature blocks composed of 3×3 Convolution (Conv) kernels, bias, Leaky ReLU, and a dropout layer. We utilize Leaky ReLU to prevent the 'dead ReLU' problem thanks to the nature of leaky rectified linear units (Leaky ReLU): $\max(0, \boldsymbol{x}) - \alpha \max(0, -\boldsymbol{x})$. Applying the dropout layer is to prevent overfitting. The number of filters is shown in Table I. In practice, we avoid normalization which is not suitable for SR, because we observe that it discards the range flexibility of the features. Then, to capture both local and the global image features, all outputs of the hidden layers are concatenated before the reconstruction network through skip connection. The skip connection helps prevent training saturation and overfitting. Diverse features which represent different details of the HRCT components can be constructed in the end of feature extraction network.

In the ***reconstruction network***, we stack two reconstruction branches and integrate the information flows. Because all the outputs from the feature extraction network are densely connected, we propose a parallelized CNNs (Network in Network) [1] which utilize shallow multilayer perceptron (MLP) to perform a nonlinear projection in the spatial domain. There are several benefits with the Network in Network strategy. First, the 1×1 Conv layer can significantly reduce the dimensionality of the filter space for faster computation with less information loss [1]. Second, the 1×1 Conv layer can increase the non-linearity of the network to learn a complex mapping better at the finer levels. For up-sampling, we adopt the transposed convolutional (up-sampling) layers [2] by a scale of 2. The last Conv layer fuses all the feature maps, resulting in an entire residual image containing mostly high-frequency details. In the supervised setting, the up-sampled image by the bicubic interpolation layer is combined (via element-wise addition) with the residual image to produce a HR output. In the unsupervised and semi-supervised settings, no interpolation is involved across the skip connection.

It should be noted that the generator F shares the same architecture as G in both the supervised and unsupervised scenarios. The default stride size is 1. However, for unsupervised feature learning, the stride of the Conv layers is 2 in the 1^{st} feature blocks. Also, for supervised feature learning, the stride of the Conv layers is 2 in the 1^{st} and 2^{nd} feature blocks of F. We refer to the forward generative network G as G-Forward.

*2) **Discriminative Networks**:* As shown in Fig. 2, in reference to the recent successes with GANs [3], [4], D is designed to have 4 stages of Conv, bias, instance norm [5] (IN) and Leaky ReLU, followed by two fully-connected layers, of which the first has 1024 units and the other has a single output. In addition, inspired by [6] no sigmoid cross entropy layer is applied by the end of D. We apply 4×4 filter size for the Conv layers which had different numbers of filters, which are $64, 64, 128, 128, 256, 256, 512, 512$ respectively.

B. Implementation Details

In the proposed GAN-CIRCLE, we initialized the weights of the Conv layer based on [7]. We computed *std* in the manner of $\sqrt{2/m}$ where *std* is the standard deviation, $m = f_s^2 \times n_f$, f_s the filter size, and n_f the number of filters. *i.e.*, given $f_s = 3$ and $n_f = 16$, *std* = 0.118 and all bias were initialized to 0. In the training process, we empirically set λ_1, λ_2, λ_3 to 1, 0.5, 0.001. Dropout regularization [8] with $p = 0.8$ was applied to each Conv layer. All the Conv and transposed Conv layers were followed by Leaky ReLu with a slope $\alpha = 0.1$. To make the size of all feature maps the same as that of the input, we padded zeros around the boundaries before the convolution. We utilized the Adam optimizer [9] with $\beta_1 = 0.5, \beta_2 = 0.9$ to minimize the loss function of the proposed network. We set the learning rate to 10^{-4} for all layers and then decreased by

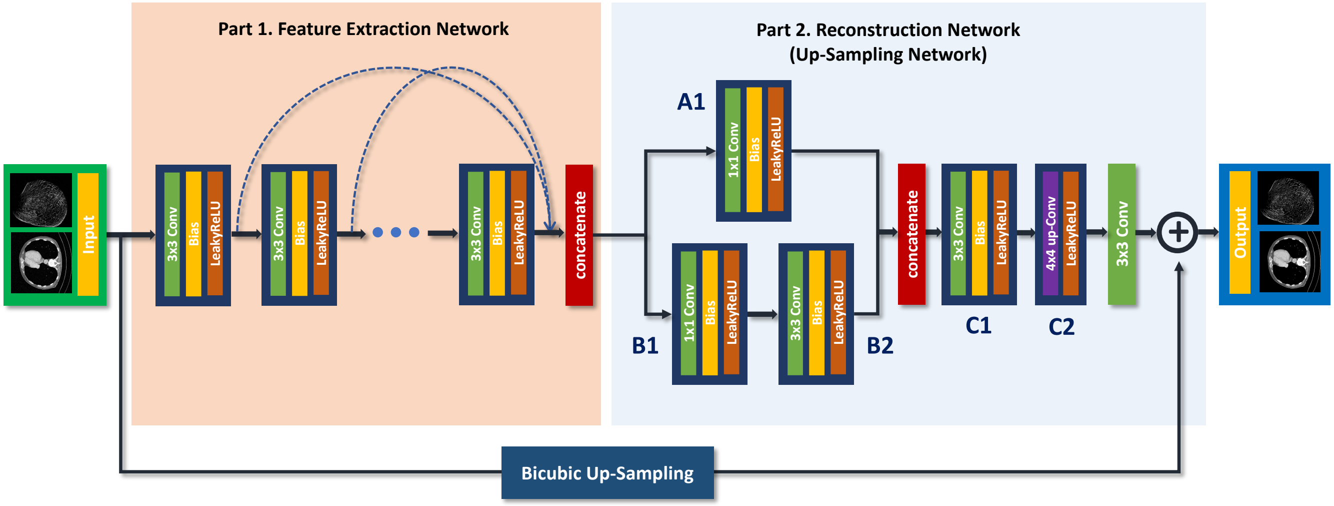


Fig. 1. Architecture of the SR generators. The generator is composed of feature extraction and reconstruction networks. The default stride is 1, except for the 1^{st} feature blocks in which the stride for the conv layers is 2. Up-scaling is performed to embed the residual layer for supervised training, and no interpolation method is used in the network for unsupervised feature learning.

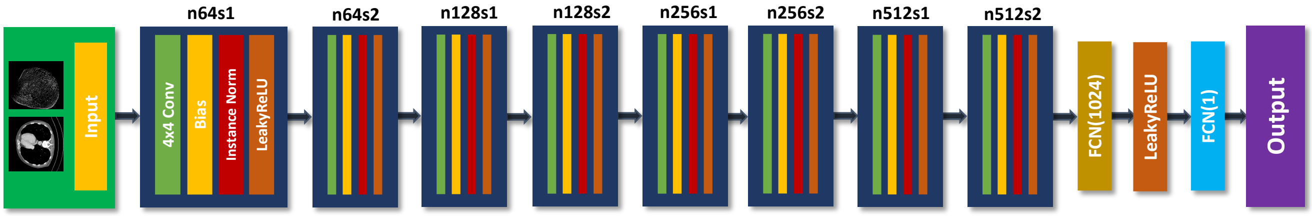


Fig. 2. A architecture of the discriminators. n stands for the number of convolutional kernels, and s stands for stride. *i.e.*, $n32s1$ means the convolutional layer of 32 kernels with stride 1.

TABLE I
NUMBER OF FILTERS ON EACH CONVOLUTION (CONV) LAYER OF THE GENERATIVE NETWORK.

	Feature extraction network											Reconstruction network					
	1	2	3	4	5	6	7	8	9	10	11	12	A1	B1	B2	C1	C2	Output
G/F	64	54	48	43	39	35	31	28	25	22	18	16	24	8	8	32	16	1

a factor of 2 for every 50 epochs and terminated the training after 100 epochs. All experiments were conducted using the TensorFlow library on a NVIDA TITAN XP GPU.

C. Parameter Setting

Tuning the hypermeters is a significant way to understand the model performance. In this study, we have performed a grid search on the hyperparameters λ_1, λ_2, λ_3, which are selected from the set $\{0, 0.01, 0.05, 0.1, 0.5, 1.0, 10, 20, \infty\}$. First, we selected the parameter $\lambda_1 \in \{0, 0.01, 0.05, 0.1, 0.5, 1.0, 10, 20, \infty\}$ when $\lambda_2 = 0, \lambda_3 = 0$. Note that the hyperparameter $\lambda_1 = \infty$ ($\lambda_1 = 0$) denotes as the SR model was only optimized with respect to the cycle-consistency loss (the adversarial loss). In this scheme, we investigated the performance of the SR model on the validation dataset. The results demonstrate that the parameter $\lambda_1 = 1$ achieved the highest PSNR value. Then, the hyperparameter λ_2 was chosen based on the performance on the validation set, and then the retrained model was used for computing the corresponding PSNR values when $\lambda_3 = 0$. Next, we investigate the performance on the validation set with different hyperparameter λ_3 values. It was observed that the $\lambda_1 = 1, \lambda_2 = 0.5, \lambda_3 = 0.01$ achived the best PSNR values. Also, it was also observed that the cycle consistency loss, identity loss, and joint sparsifying transform loss can improve the average PSNR values. We have reported the average PSNR values using GAN-CIRCLE in *main context* Fig. 9d.

II. DATASETS

A. Data Preprocessing

We perform image pre-processing for all CT images through the following workflow. The original CT images were first

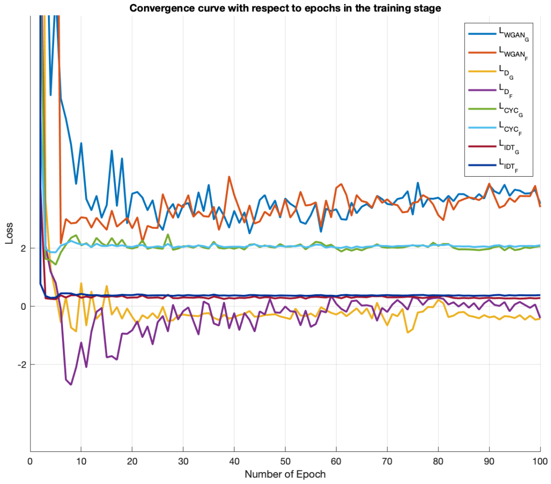


Fig. 3. Convergence curve with respect to the epochs in the training stage.

scaled from the CT Hounsfield Value (HU) to the unit interval [0,1], and treated as the ground-truth HRCT images. In addition, we followed the convention in [10], [11] to generate LR images by adding noise to the original images and then lowering the spatial resolution by a factor of 2. For convenience in training our proposed network, we up-sampled the LR image via proximal interpolation to ensure that $\boldsymbol{x}$ and $\boldsymbol{y}$ are of the same size.

Since the amount of training data plays a significant role in training neural networks [12], we extracted overlapping patches from LRCT and HRCT images instead of directly feeding the entire CT images to the training pipeline. The overlapped patches were obtained with a predefined sliding size. This strategy preserves local anatomical details, and boost the number of samples. We randomly cropped HRCT images into patches of 64×64, along with their corresponding LRCT patches of size 32×32 at the same center point for supervised learning. With the unsupervised learning methods, the size of the HRCT and LRCT patches are 64×64 in batches of size 64.

III. Additional Analysis

In this section, we draw the convergence of the proposed network. Note that the test data were randomly selected from the Tibia dataset (average values from 150 images). Finally, we provide more visualization examples on cadaver spine CT data.

A. Network Convergence

Training GANs is not easy, suffering from instability and model collapse problems, which were theoretically investigated in [6], [13]. To facilitate the convergence of the GAN training process, we adopted WGAN-GP [13] and the cycle-consistent loss. As depicted by the convergence curve in Fig. 3, the proposed network converged stably, where $\mathcal{L}_{\mathrm{WGAN}_G}$ denotes the first term of the adversarial loss termed as the generator loss, and $\mathcal{L}_{\mathrm{D}_G}$ refers to Eq. 5 termed as the discriminator loss. It can be seen that the cycle-consistent loss smoothly converged. These data indicate the stability of the overall training process.

IV. More qualitative comparisons

In this section, we provide more visual comparisons of the difference images on Tibia and Abdominal datasets and on cadaver spine examples.

A. More examples under the GAN-CIRCLE framework

We first obtained 6 LRCT and HRCT image pairs using cadaver spines on the SIEMENS CT scanner. The parameters are as follows: X-ray source circular scanning, 120 kVp, 680 mAs, and the HRCT and LRCT images were reconstructed using the filtered backprojection algorithm (FBP) with different kernels: HRCT image size 512×512, 962 slices at 0.1172×0.1172 mm^2 pixel size, and the LRCT image size 256×256, 481 slices at 0.2344×0.2344 mm^2 pixel size.

1) Examples on simulated Cadaver spine dataset: We followed the previous data preprocessing method to obtain matched HRCT and LRCT images pairs. More simulated visual examples are shown in Fig. 4.

2) Examples on real Cadaver spine dataset: Since the real data are unmatched, we accordingly evaluated our proposed GAN-CIRCLE^s and GAN-CIRCLE^u networks for 1X resolution improvement. In this section, we provide more visual examples in Fig. 5.

References

[1] M. Lin, Q. Chen, and S. Yan, "Network in network," *Int. Conf. Learn. Representations. (ICLR)*, 2014.

[2] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, "Deconvolutional networks," in *Proc. IEEE Conf. Comp. Vis. Patt. Recogn. (CVPR)*, 2010, pp. 2528–2535.

[3] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Int. Conf. Learn. Representations. (ICLR)*, 2015.

[4] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. P. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network." in *Proc. IEEE Conf. Comp. Vis. Patt. Recogn. (CVPR)*, vol. 2, no. 3, 2017, p. 4.

[5] D. Ulyanov, A. Vedaldi, and V. S. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," *CoRR*, vol. abs/1607.08022, 2016.

[6] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2017, pp. 214–223.

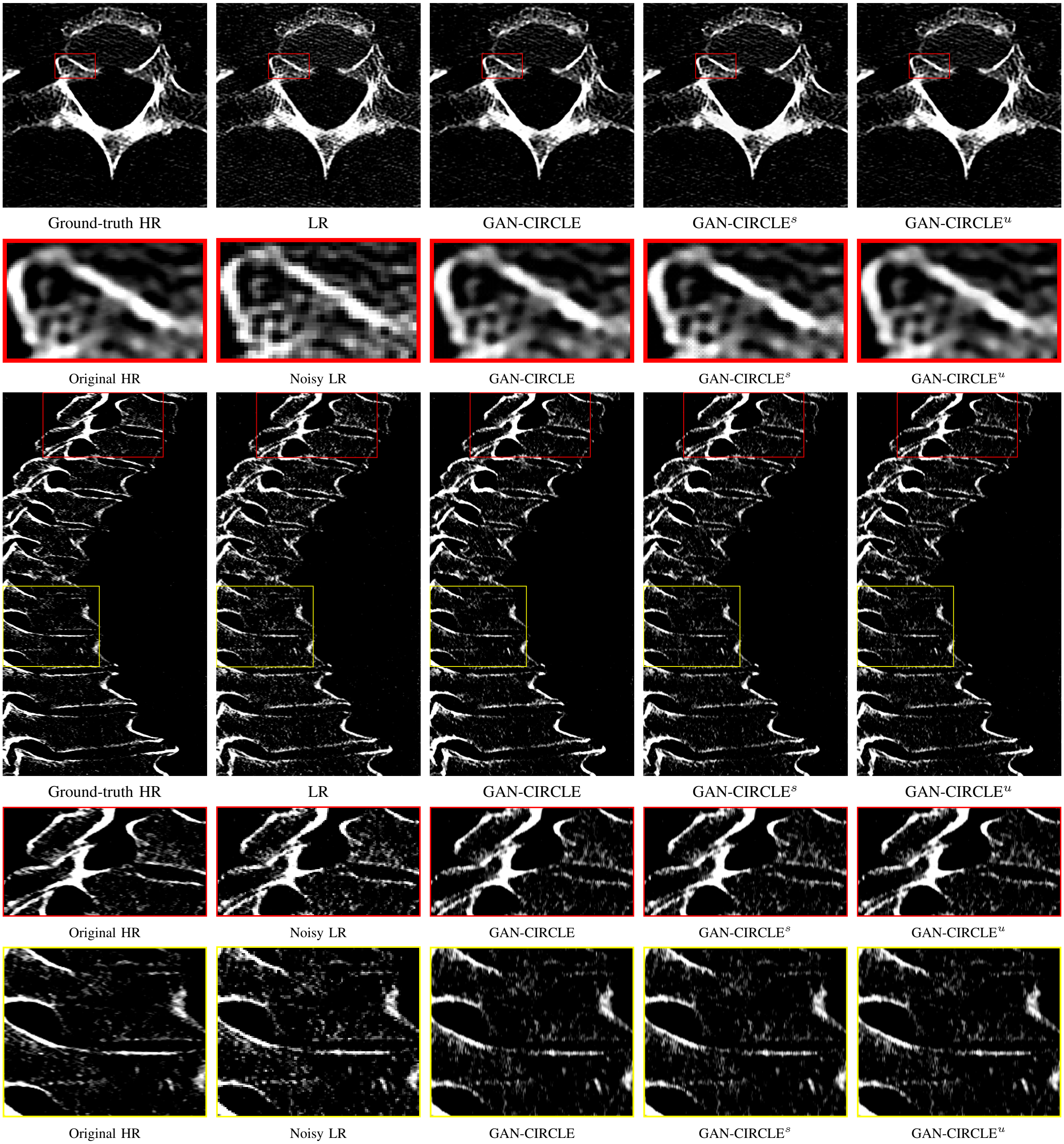
[7] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proc. IEEE Int. Conf. Comp. Vis. (ICCV)*, 2015, pp. 1026–1034.

[8] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *CoRR*, vol. abs/1207.0580, 2012.

[9] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Int. Conf. Learn. Representations. (ICLR)*, 2015.

[10] C. Jiang, Q. Zhang, R. Fan, and Z. Hu, "Super-resolution CT image reconstruction based on dictionary learning and sparse representation," *Sci. Rep.*, vol. 8, no. 1, p. 8799, 2018.

[11] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *Proc. IEEE Int. Conf. Comp. Vis. (ICCV)*, 2009, pp. 349–356.



Ground-truth HR LR GAN-CIRCLE GAN-CIRCLE^s GAN-CIRCLE^u

Original HR Noisy LR GAN-CIRCLE GAN-CIRCLE^s GAN-CIRCLE^u

Ground-truth HR LR GAN-CIRCLE GAN-CIRCLE^s GAN-CIRCLE^u

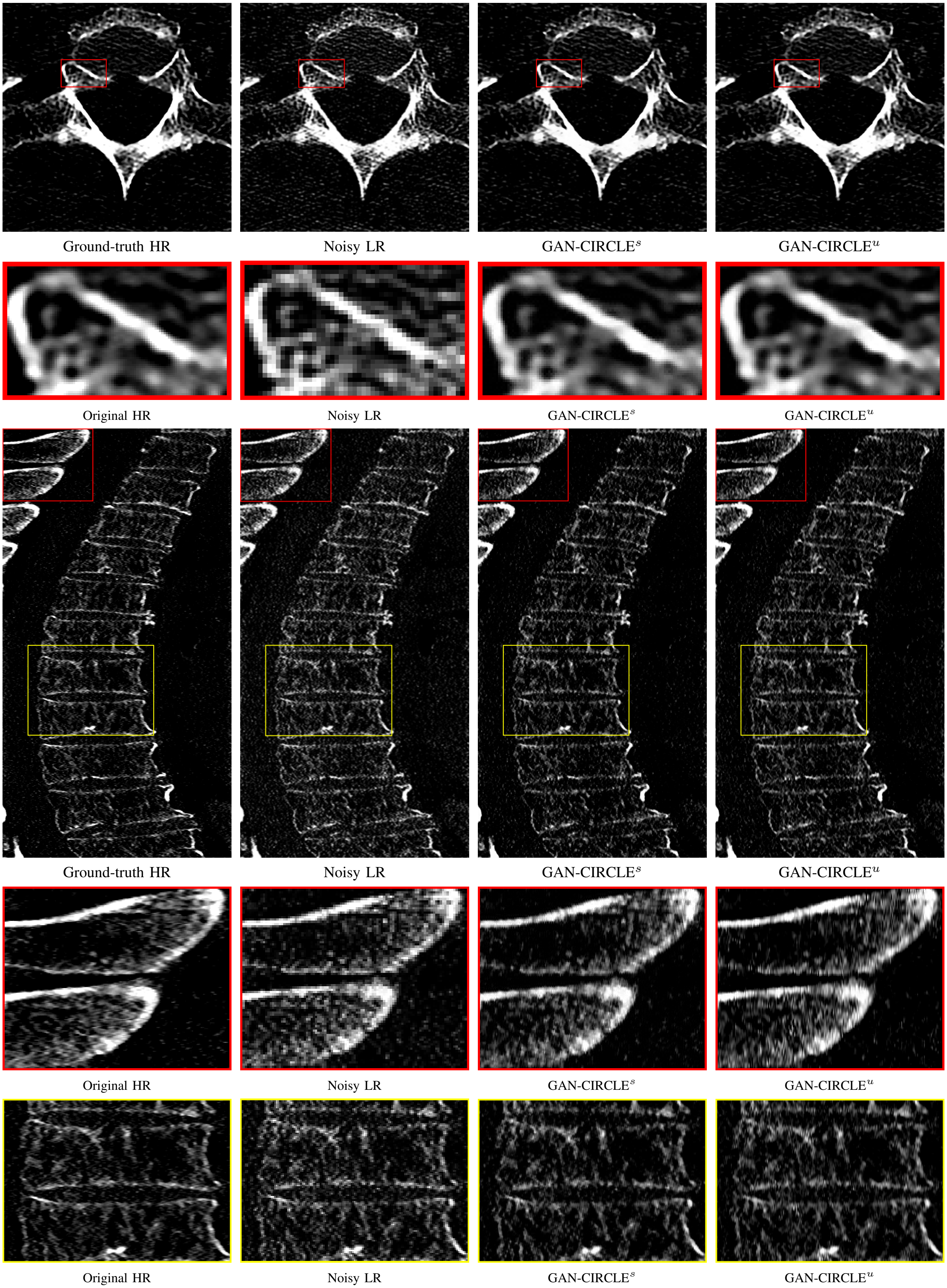
Original HR Noisy LR GAN-CIRCLE GAN-CIRCLE^s GAN-CIRCLE^u

Original HR Noisy LR GAN-CIRCLE GAN-CIRCLE^s GAN-CIRCLE^u

Fig. 4. Visual comparsion of SRCT Cases from the simulated Cadaver Spine dataset. The display window is [-120, 920] HU. The restored bony structures are shown in the red and yellow boxes. (**Zoomed for visual clarity**).

[12] D. Liu, Z. Wang, Y. Fan, X. Liu, Z. Wang, S. Chang, X. Wang, and T. S. Huang, "Learning temporal dynamics for video super-resolution: A deep learning approach," *IEEE Trans. Image Process.*, vol. 27, no. 7, pp. 3432–3445, 2018.

[13] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5769–5779.



Ground-truth HR
Noisy LR
GAN-CIRCLE^s
GAN-CIRCLE^u

Original HR
Noisy LR
GAN-CIRCLE^s
GAN-CIRCLE^u

Ground-truth HR
Noisy LR
GAN-CIRCLE^s
GAN-CIRCLE^u

Original HR
Noisy LR
GAN-CIRCLE^s
GAN-CIRCLE^u

Original HR
Noisy LR
GAN-CIRCLE^s
GAN-CIRCLE^u

Fig. 5. Visual comparsion of SRCT Cases from the real Cadaver Spine dataset. The display window is [-120, 920] HU. The restored bony structures are shown in the red and yellow boxes. (**Zoomed for visual clarity**).