

Supplementary Materials

Anonymous Authors

1 MORE VISUAL RESULTS

Fig. 1-2 and Fig. 3 show more qualitative results of the downsampled and upsampled images at a scale of $4\times$ on both ACDC-snow and our RS dataset, respectively. We present more comparisons with LDL and IRN. Our method can achieve blurred informative downsampled images and favorably better reconstruction results, as it eliminates redundancy between images at the downscaling stage and utilizes reference images to reconstruct the structures and details at the upscaling stage.

2 MORE ABLATION STUDIES

2.1 Sampling of Latent Variable

The latent variable z is randomly sampled from standard Gaussian distribution, thus we would like to explore some discussions on the influence of different samples of z in this subsection.

We calculate and visualize the difference of reconstructed HR images on different samples of z . The results in Fig. 4 show that the differences are only slight meaningless random high-frequency noise without typical textures, indicating that the proposed RefScale reconstructs meaningful high-frequency contents with the assistance of reference and embedding senseless noise into randomness. We speculate that the reason for this could be that one of our optimization objectives is to encourage pixel-level similarity (L1 loss) between the reconstructed image and the original HR image, thus the diversity of different z shows the randomness of imperceptible high-frequency details.

In addition, we further verify the influence of the deviation degree of z from the center of the standard Gaussian distribution. We take downsampled image y and scaled latent variable αz as input, where α is the scaling factor, to explore changes in reconstructed images. Note that the probability distribution density of sampling αz decreases as α increases. As illustrated in Fig. 5, fidelity images can still be reconstructed with a smaller α , and more distortion is introduced by the larger deviation, meaning that the proposed RefScale faithfully embeds mutual information to the specified distribution and is robust to a slight distribution deviation.

2.2 Robustness to Affine Transformations

Although the preprocessing technology of image registration can align the geographical coordinates between remote sensing images, considering the fact that the reference images may not be perfectly aligned in practical application scenarios, we examine the robustness of the proposed RefScale to geometric deformation. We perform translation and rotation transformations to the reference image. As depicted in Fig. 6, we perform 2 scales of translation transformation, 20-pixel, 40-pixel, and 2 scales of rotation transformation 5° , 10° . The results show excellent stability for the small-scale affine transformation of the reference image, as our method enable to discard the information adaptively and the transformed texture can still contribute to the HR image reconstruction.

Table 1: Computational complexity comparison of various rescaling methods

Model Size						
	Bicubic	EDSR	RCAN	NLSN	LDL	LTE
$\times 2$	/	40.7M	15.4M	41.8M	16.7M	12.1M
$\times 4$	/	43.1M	15.6M	44.2M	16.7M	12.1M
	TTSR [†]	C^2 Matching [†]	TAD & TAU	IRN	HCFlow [†]	Ours
$\times 2$	-	-	0.7M	1.7M	-	2.8M
$\times 4$	6.4M	8.9M	0.7M	4.4M	23.2M	3.5M
Computational Cost						
	Bicubic	EDSR	RCAN	NLSN	LDL	LTE
$\times 2$	-	127.7G	88.5G	267.5G	14.0G	61.6G
$\times 4$	-	157.6G	92.0G	328.4G	56.2G	100.4G
	TTSR [†]	C^2 Matching [†]	TAD & TAU	IRN	HCFlow [†]	Ours
$\times 2$	-	-	0.53G	10.4G	-	68.3G
$\times 4$	76.0G	132.5G	10.7G	58.6G	318.7G	80.6G

[†] denotes the method can only implement $4\times$ rescaling.

Table 2: Computational complexity of compression w/o proposed RefScale

	Rescaling		Compression			Sum	
	Size	FLOPS	Input Scale	Size	FLOPS	Size	FLOPS
GMM	-	-	1	31.7M	108.2G	31.7M	108.2G
GMM+ RefScale $2\times$	2.8M	68.3G	1/2	31.7M	27.0G	34.5M	95.3G
GMM+ RefScale $4\times$	3.5M	80.6G	1/4	31.7M	6.7G	35.2M	87.3G

3 COMPUTATIONAL COMPLEXITY

3.1 Model Complexity of Rescaling

We compare the detailed computational complexity between the proposed RefScale and other image rescaling methods with open-source code on our RS test set. We demonstrate the results of $2\times$ and $4\times$ here. As shown in Table 1, our method consumes fewer parameters than most methods, benefiting from the lightweight property of INNs. In addition, we calculate the FLOPs for models to downscale and upscale images, and our method demonstrates moderates computational complexity in terms of FLOPS.

3.2 Model Complexity of Compression

We compare compression model size and computational cost with and without rescaling. We employ a learning-based codec, GMM [2], as baseline, which is an entropy coding model using discretized Gaussian mixture likelihoods for estimating latent representation. As shown in Table 2, the model size of rescaling is about one-tenth of the compression model, which does not bring significant model size improvement. The computational complexity of compression model decreases quadratically as a result of the reduction in image spatial resolution caused by rescaling. The overall computational complexity in the last raw (including encoding and decoding) dropped slightly with $2\times$ and $4\times$ rescaling. Coupled with our method, the size of bitstream could be reduced manyfold with only a small increase in model size and no increase in computational complexity.

4 APPLICATION

The proposed Refscale is a suitable technique for remote sensing transmission with limited bandwidth and periodic satellite revisits. Limited bandwidth prompts us to consider the transmission overhead of low-resolution images, and periodic satellite revisits bring practicality to the introduction of reference imagery. The whole process can be described by the following five parts:

1) **Selecting reference image.** According to the location and the time of the mission, a historical image is selected at ground station as reference image.

2) **Uploading to satellite.** The reference image is uploaded through uplink to the satellite along with the task command.

3) **Downscaling image on-board.** The captured image is down-scaled on the satellite by the proposed RefScale based on the uploaded reference image.

4) **Compressed transmission.** The generated low-resolution image is transmitted through the downlink to the ground station with compression algorithms.

5) **Reference-guided reconstructing.** At the ground station, the low-resolution image is recovered to original resolution using the same reference as the uploaded one.

Based on the dual-link transmission [1] between satellites and the ground, uploading the reference images by uplink does not consume the bandwidth of the downlink for transmitting captured images, making this process practical. A similar process flow is also mentioned in [3].

REFERENCES

- [1] Francesc Auli-Llinàs, Michael W Marcellin, Victor Sanchez, Joan Bartrina-Rapesta, and Miguel Hernández-Cabronero. 2018. Dual link image coding for earth observation satellites. *IEEE Trans. Geosci. Remote. Sens.* 56, 9 (2018), 5083–5096.
- [2] Zhengxue Cheng, Heming Sun, Masaru Takeuchi, and Jiro Katto. 2020. Learned image compression with discretized gaussian mixture likelihoods and attention modules. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 7939–7948.
- [3] Huiwen Wang, Liang Liao, Jing Xiao, Weisi Lin, and Mi Wang. 2023. Uplink-Assist Downlink Remote-Sensing Image Compression via Historical Referencing. *IEEE Transactions on Geoscience and Remote Sensing* 61 (2023), 3315725.

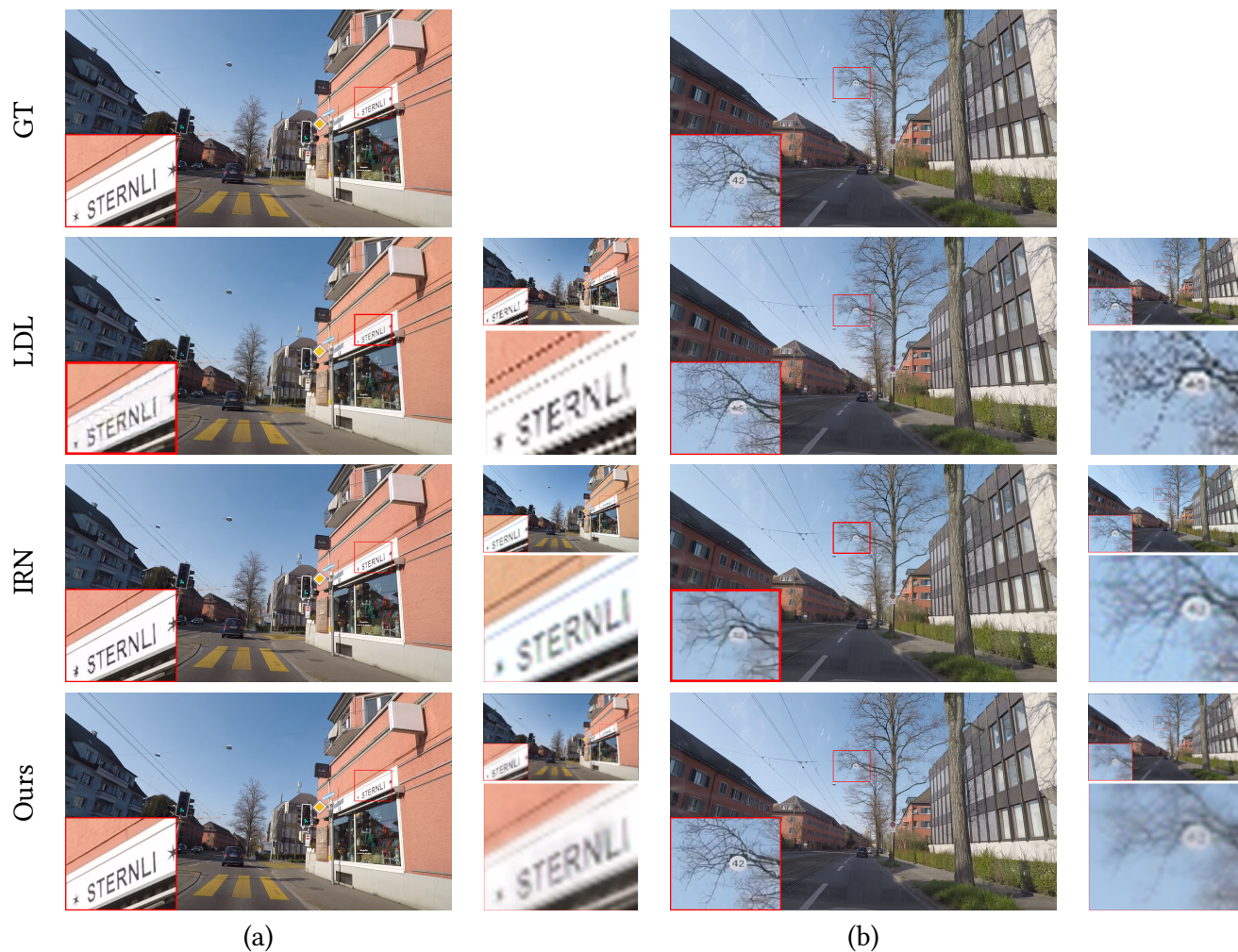


Figure 1: More qualitative results on ACDC-snow dataset. In each group, the reconstructed image is on the left, and the downscaled image with a zoom-in image is on the right for the convenience of observation. Our method outperforms the state-of-the-art on road signs and characters. We reconstructed the fine structure information from more blurry downscaled images, thanks to the assistance of the references.



Figure 2: More qualitative results on ACDC-snow dataset. In each group, the reconstructed image is on the left, and the downscaled image with a zoom-in image is on the right for the convenience of observation. Our method outperforms the state-of-the-art on road signs and characters. We reconstructed the fine structure information from more blurry downscaled images, thanks to the assistance of the references.

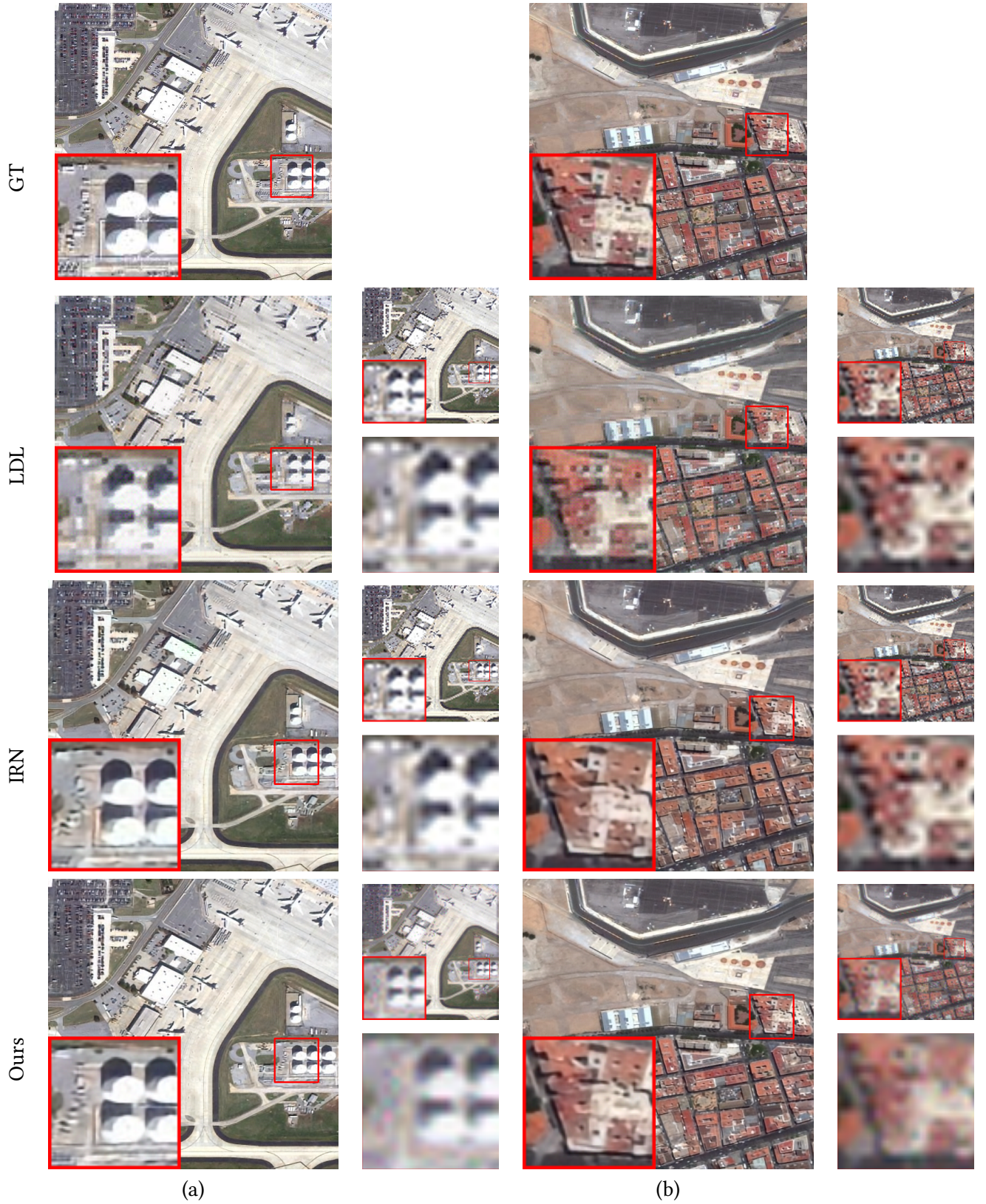


Figure 3: More qualitative results on RS dataset. In each group, the reconstructed image is on the left, and the downscaled image with a zoom-in image is on the right for the convenience of observation. Our method significantly outperforms LDL in reconstruction quality, and generates more informative and blurry downscaled images than IRN.

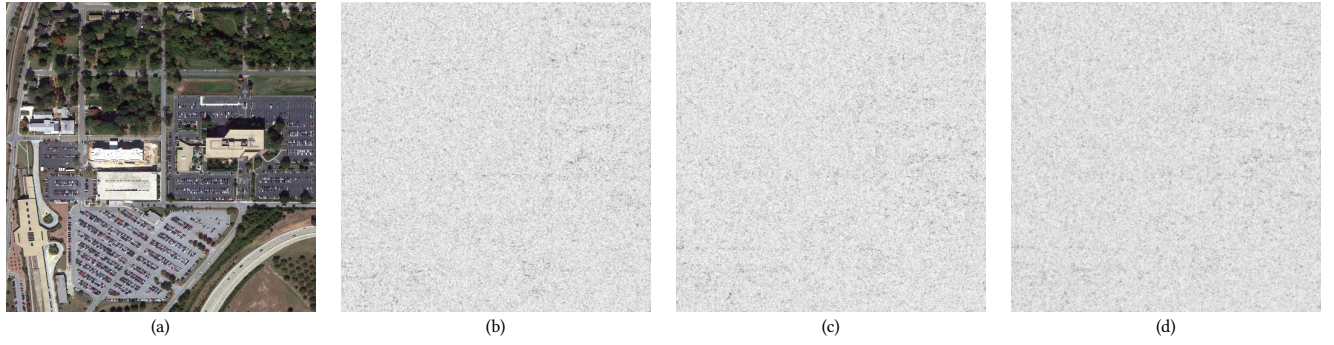


Figure 4: Visualisation of the difference of reconstructed images on different samples of z . (a) is the ground truth and (b-d) are reconstructed differences of three samples of z , and are enhanced for better visualization. The differences are meaningless random noise in high-frequency regions without structure information.

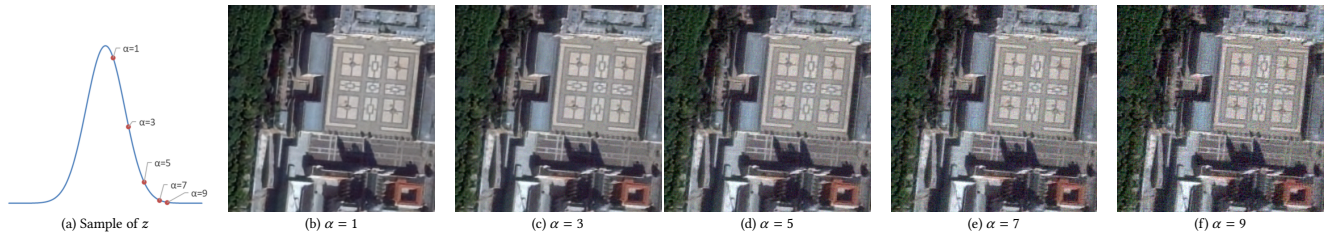


Figure 5: Influence of sampling deviation degree. α is the scaling factor of sampling, and (a) indicates where each case falls in the distribution. (b-f) are the reconstructed images with different sampling of z . A larger deviation to the center of the distribution results in more noisy textures and distortion.

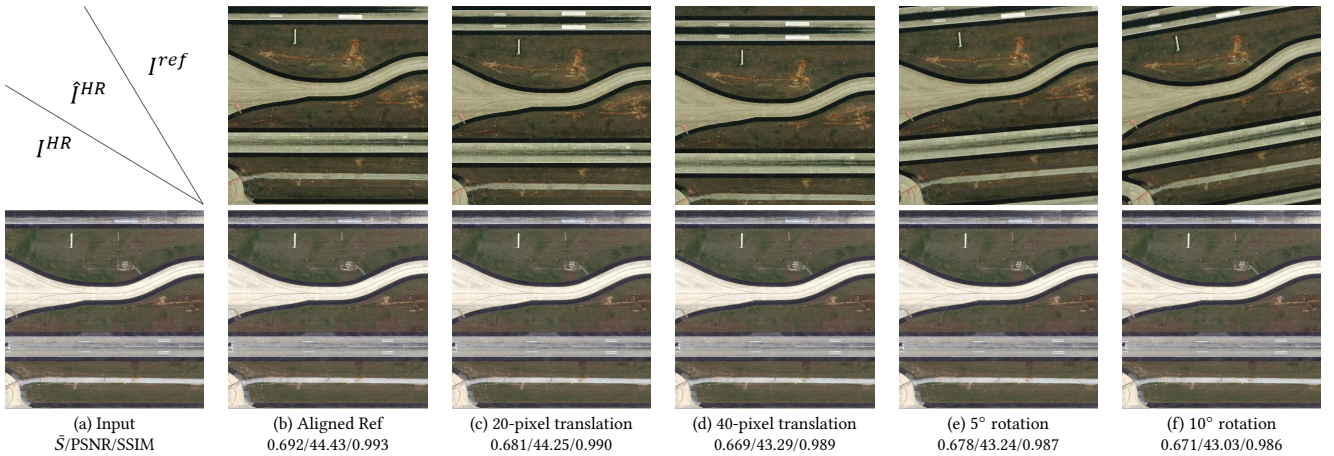


Figure 6: The effects of affine transformation of the reference image. (c-d) and (e-f) show the effect of the translation and rotation transformation, respectively. The first row is the reference images and the second row is the corresponding reconstructed images.