Zero-Shot Remote Sensing Image Super-Resolution Based on Image Continuity and Self Tessellations

Rupak Bose(✉), Vikrant Rangnekar, Biplab Banerjee,
and Subhasis Chaudhuri

Indian Institute of Technology, Bombay, Mumbai, India

Abstract. The goal of zero-shot image super-resolution (SR) is to generate high-resolution (HR) images from never-before-seen image distributions. This is challenging, especially, because it is difficult to model the statistics of an image that the network has never seen before. Despite deep convolutional neural networks (CNN) being superior to traditional super-resolution (SR) methods, little attention has been given to generating remote sensing scene-based HR images which do not have any prior ground truths available for training. In this paper, we propose a framework that harnesses the inherent tessellated nature of remotely images using continuity to generate HR images that tackle atmospheric and radiometric condition variations. Our proposed solution utilizes self tessellations to fully harness the image heuristics to generate an SR image from a low resolution (LR) input. The salience of our approach lies in a two-fold data generation in a self-preservation case and a cascaded attention sharing mechanism on the latent space for content preservation while generating SR images. By learning a mapping from LR space to SR space while keeping the content statistics preserved helps in better quality image generation. The attention sharing between content and tessellations aids in learning the overall big picture for super-resolution without losing an eye on the main image to be super-resolved. We showcase our results with the generated images given the low resolution (LR) input images in zero-shot cases comparable to state-of-the-art results on EuroSAT and PatternNet datasets with metrics of SSIM and PSNR. We further show how this architecture can be leveraged for non-remote sensing (RS) applications.

Keywords: Super-resolution · Remote sensing · Attention sharing

1 Introduction

Generating an SR image from a given LR image in a smooth end-to-end fashion is what is expected from an automated super-resolving framework. Such a framework becomes highly desirable as on-demand detailed image generation would make efficient storage use by allowing HR images to be downsampled

C. Bauckhage et al. (Eds.): DAGM GCPR 2021, LNCS 13024, pp. 649–662, 2021.
https://doi.org/10.1007/978-3-030-92659-5_42

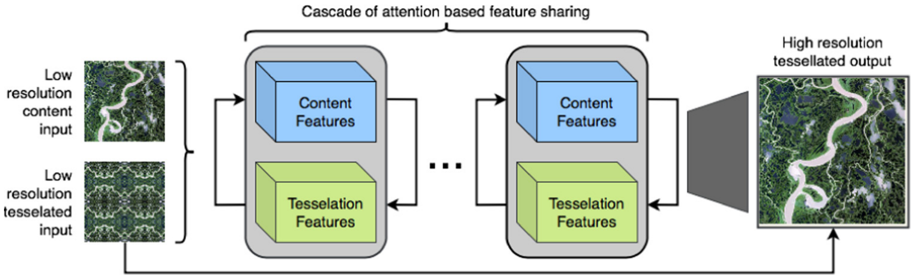

Cascade of attention based feature sharing
Low resolution content input
Low resolution tesselated input
Content Features
Tesselation Features
...
Content Features
Tesselation Features
High resolution tessellated output

Fig. 1. A schematic of our proposed model. A cascade of attention based feature sharing modules for generating super-resolved images. The attention-sharing-based interaction helps the content statistics trickle down into the tessellation features.

and stored. Also, it would mean lower bandwidth consumption while transmission if images at the end-user have access to an SR framework. Apart from having upsides in traditional computer vision, bringing super-resolution to the remote sensing domain can attempt to address various problems like generating high-resolution spectro-spatial bands from a low spatial-high spectral band configuration (Fig. 1).

Given that multi-spectral images (MSI) are defined by multi-resolution bands, it is favorable to bring all the bands to a common higher resolution to better understand the features. Super-resolution in the remote sensing domain differs from traditional approaches as the objects in the satellite image are very small compared to the huge scale of a satellite image. Here, we leverage the fact that most remote sensing scenes are a kind of tessellation by nature in some sense. Tessellations are those structures whose individual components are repeated to form a pattern as a representative of a whole sample and can be found repeating throughout the sample at various scales. If we can learn the properties of the smaller clusters, the generation of features for larger clusters would be a simple task. Some of the challenges with satellite images include different atmospheric conditions for different images, diverse shape generations from low-resolution images, and various spectral signatures present in the image.

The traditional method of increasing satellite image resolution is by pan-sharpening where the 1st principal component is replaced by a panchromatic (PAN) image and then inverse principle component analysis is done which brings all the bands to the PAN image resolution. Methods like average interpolation, bi-linear, and bi-cubic interpolation are fast methods for upscaling but they lack behind in generating sharp features while upscaling. Advance deep learning methods like convolutional neural networks (CNN) based architectures [3,8,16] and generative adversarial networks [7,13] perform well at hallucinating the details while upscaling and generating sharp high-resolution images in the traditional setting.

Super-resolution being an inherently difficult problem, some of the above-mentioned problems can be tackled if the framework is shifted to zero-shot and the learning is based on self patterns [11,12]. The scene-based radiometric corrections can be reduced if the scene stays unchanged in the training phase. Better robust features can be generated for true context while generating features in the testing phase. Thus, even though zero-shot has many potential advantages, it has been hugely neglected in the satellite images domain.

Given that the traditional methods perform well, they miss out on addressing a few problems. Pan-sharpening does increase the resolution, but it can't be used to generate resolutions higher than the PAN image. It also fails if the bands have co-related noise as an inherent property. The interpolating methods tend to generate smooth images as it performs a weighted average of the values depending on the nature of interpolation and generates the intermediate values. Deep learning-based models require a huge amount of data and time with the scenes being from a relatively similar environment to generate comparable results. Also building a unified model for super-resolution on satellite images is a challenge as different sensors output images having different configurations and different image properties.

To this end, we propose a cascaded attention sharing-based model which aims to address the aforementioned problems faced by the existing architectures. We use the zero-shot framework on satellite images to increase their resolution. This eliminates the atmospheric condition problem as it is trained on the same atmospheric conditions as the original image is. And the requirement of huge data for training models is eliminated as we just require the original image in the training phase for the super-resolution task. Attention sharing helps robust feature learning in a bottom-up approach. Also with the reduction in training times, we can attempt super-resolution as an on-the-fly method. Our contributions in this paper would be:

- We propose two unique methods based on image continuity, i.e., internal and external tessellations, using self-image continuity without loss of generality for data generation in a zero-shot setting.
- We also propose a novel cascade of attention sharing networks on the latent space that helps trickle-down content statistics into the super-resolution domain while upsampling to intermediate scales.
- We showcase our model's performance on popular datasets like UC Merced land-use, EuroSAT rgb, and PatternNet datasets in terms of PSNR, SSIM and cosine similarity.
- We propose an extension to a true zero-shot case in terms of internal tessellation for efficient image size invariant scaling of up to 16× and for external tessellation of up to 8× scaling for generalized uses.

2 Related Works

Apart from the traditional methods, the recent trends for super-resolution are centred around training on low-high resolution training pairs and testing the model on a test low-resolution pair. The task being generative, CNN based architectures, especially generative adversarial networks (GANs) and auto-encoders show promising results. Skip connections play a key role in generating a colour accurate high-resolution image by transferring data between encoder and decoder is well demonstrated by RedNet [9] and Deep Memory Connected Network [14]. It shows that a better flow of information across the encoder-decoder network is as important as having a deep convolutional model. Residual connections enhance the generated feature maps and this is utilized well in the super-resolution of Sentinel-2 images [6]. The performance of GANs is proved by D-SRGAN [2]. It shows promising results by generating high-resolution DEMs from low-resolution DEMs without the need for extra data. However, D-SRGAN does not perform uniformly on all terrains. Flatter terrains produce better results than rugged terrains. Super-Resolution increases the performance of detection tasks by enhancing objects can be seen in vehicle detection algorithm [5] and object detection algorithm [10].

Given the merits of zero-shot super-resolution, the vision community has been harnessing its usefulness. [12] demonstrates meta-transfer learning for exploiting internal image properties for a faster SR method. [1] propose a depth guided methodology and learning-based downsampling to leverage internal statistics for producing SR images. [11] uses internal recurrence of information to train the model. This shows us that internal learning can give us a more robust instance-based learning for SR tasks. However, these methods don't effectively utilize the internal pattern-based tessellations to the fullest extent. This in turn makes the model miss out on the bigger picture.

3 Methodology

3.1 Tessellation Based Data Generation

Tessellations are patterns generated due to structural repetitive cycles. As tessellations are uniform in all directions, the statistics such as mean and variance of the tessellated image is largely similar to that of the content image. This comes to our advantage when used to generate random augmented batches for training the network as it encourages the model to learn different patterns emerging with the same statistical distributions. Thus a robust learning approach is established when mapping an LR image to an HR image. To utilize these properties, we propose two methods to produce tessellation based HR/LR pairs based on internal and external cycles.

Internal Cycle: It is an algorithm to build HR/LR pairs using internal patterns. Here we sample smaller scale images from the original images itself to be tessellated for generation of a larger image. This larger image is then down-scaled to form a low resolution image. As shown in (Fig. 2. a.), we take a random patch from the content image and tessellate it to generate a high-resolution image and use bicubic down-sampling to downscale the generated tessellated image. Looping through the random patches of the content images generated a set of HR/LR pairs.

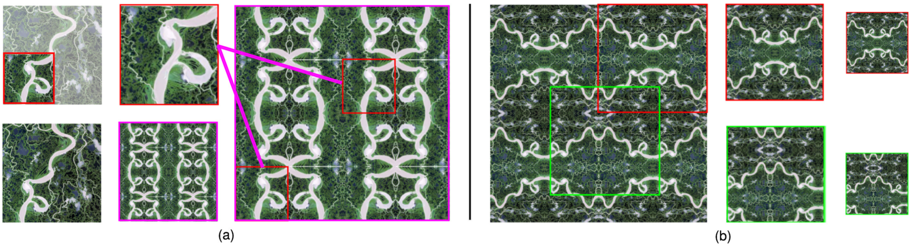


Fig. 2. The data generation techniques. (a) The internal tessellation based HR and LR image pair generation. An internal patch is selected and stitched with continual pattern generating augmentations for required scaling. (b) The external tessellation based HR-LR pair generation based on sliding window operations on the continued image.

External Cycle: It is an algorithm to build HR/LR pairs using external patterns. Here we tessellate the original image directly to generate a larger image. The high resolution images are then sampled from the larger tessellated image and down-scaled to generate low resolution images. As shown in (Fig. 2. b.), we take tessellate the content image to generate an 8×/16× image. Then random HR patches are obtained from this 8×/16× image and bicubic down-sampling is applied to obtain LR images.

3.2 Proposed Architecture

Super-resolution is a data-driven process and to make the process efficient, it is highly desirable to understand what features play an important role and how much its contribution is towards generating the final output. To achieve this, an advanced mechanism called self-attention can be used with soft weights. This soft self-attention pushes the weights between 0 and 1 in a continual fashion.

To this end, we propose a cascaded latent space attention sharing network (Fig. 3) for both the content stream and tessellation stream to model mapping from LR to HR while learning to super-resolve tessellations as the content features are preserved. The continual interaction of the latent space not only helps joint feature learning but also helps reach a common ground that highlights feature importance for both content as well as tessellation streams. The highlighting is done through the lower level feature to the higher level features. This helps in efficient feature migration while super resolving in the testing phase.

HR and LR images share similar statistical distributions. Thus it is important to selectively highlight features not only at the deeper levels but also in the initial levels too. The inter-connectivity of attention weights ensures cross-content feature preservation. To this end, we propose a cross attention weights based module at upscaled latent embeddings. In this mechanism (Fig. 4) the attention weights are self-attention soft weights. For a given feature, the attention is calculated using Global Max Pooling and Global Average Pooling followed by a dense layer with 'sigmoid' activation. The 'sigmoid' activation assigns soft weights ∈

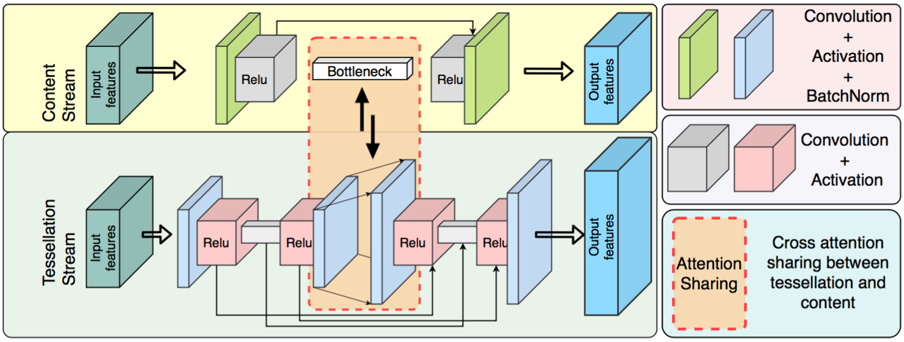


Fig. 3. The overall architecture of our model. The auto-encoder based building blocks with attention sharing at the upscaling latent space for increased interactions between content and tessellation streams.

[0,1]. These self-attention scores are then point-wise multiplied with cross features followed by residual self skip connections to enhance cross attention based self feature highlighting. This sharing mechanism is run in a cascaded manner. A stack of attention sharing mechanisms ensures attention scores from the lower level based filters to deeper latent features. In this way, the super-resolution happens for the tessellation stream, whereas the features of the content stream are constantly preserved.

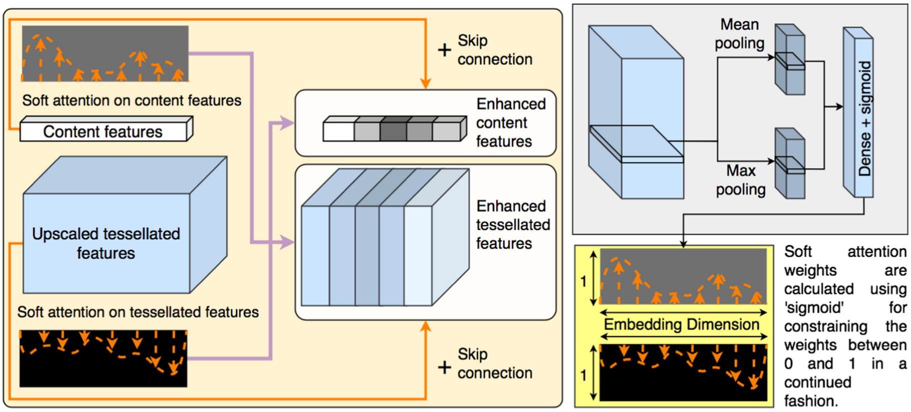
The building blocks of the overall architecture are auto-encoder(AE) style modules with skip connections around a bottleneck. The AE has the initial layers of 2D Convolutions with activation of 'relu' followed by batch normalization. The internal layers are of 2D convolutions with activation of 'relu'. The output of AE module is upscaled using transposed 2D convolutions and AE module is applied on this upscaled feature. The activation of the bottleneck layer is kept 'tanh' for an even distribution of values in the latent space.

It is in this upscaling space, the attention score sharing happens between content and tessellation streams. The cascaded connections interaction points share attention scores from these features to exchange cross-feature information. The bottleneck ensures relevant information being filtered in due to non-linearity based dimensionality reduction. The skip connections allow for a smoother gradient flow.

4 Experimental Setup

4.1 Training Protocol

The model is trained for 500 epochs with a learning rate of 5×10^{-4}. The loss for both the content as well as HR image is mean squared error. Mean squared error [Eq. 1] takes into account the squared pixel-wise difference between the generated image and ground truth across all channels. Higher is the mean squared error, the more dissimilar the generated images and ground truth is. Given $\hat{y}$ is the



Skip connection
Soft attention on content features
Content features
Enhanced content features
Upscaled tessellated features
Enhanced tessellated features
Soft attention on tessellated features
Skip connection
Mean pooling
Max pooling
Dense + sigmoid
1
Embedding Dimension
1
Soft attention weights are calculated using 'sigmoid' for constraining the weights between 0 and 1 in a continued fashion.

Fig. 4. The attention sharing mechanism. The soft self-attention weights are generated using ‘sigmoid’ activation on max and mean pooling. These attention scores are exchanged and cross-latent features are highlighted.

generated pixel value and y is the ground truth, (m, n) are image height and image width respectively, we have:

$$\mathcal{L}_{\mathcal{MSE}} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(\hat{y}_{ijk} - y_{ijk})^2 \tag{1}$$

We use internal tessellation based image generation for training purposes. For a given image, we generate tessellation data with a continual sliding window-based patch which is further extended 3 times with using rotation augmentation of 90°, 180° and 270°. The content image is kept constant with data augmentation with rotation of 90°, 180° and 270°. After every epoch, the generated data is shuffled and then fed into the network to randomize learning and lower the memorization of parameters. The model is tested on randomly sampled 100 images per class and images are generated per instance basis for quality monitoring.

4.2 Accuracy Metrics

We use three standardized accuracy metrics popularly used for super-resolution tasks: Peak Signal to Noise Ratio (PSNR), Structural Similarity Index Measurement (SSIM) and cosine similarity. These metrics are used to quantify how close the produced images are to the ground truth in terms of signal quality and visual perception. Each of the metrics have characteristics that convey image quality.

PSNR [Eq. 2] is the ratio between the maximum value of a signal to the strength of distorting noise which affects the quality of its characterization. It is usually denoted in terms of the logarithmic decibel scale. Higher is the PSNR, higher is the quality of the generated image. Given MAX is the maximum signal value, MSE is the pixel-wise mean squared error, we have:

$$\text{PSNR} = 10 \log_{10} \left(\frac{MAX^2}{MSE} \right) \tag{2}$$

SSIM [Eq. 3] measures the perceived quality of the generated image as compared to the ground truth. It takes into consideration the standard deviations along with means of the generated image compared with the ground truth. Higher is the value of SSIM, higher is the quality of the generated image. Given $\hat{y}$ is the generated pixel value and y is the ground truth with μ and σ as mean and standard deviation, C_1 and C_2 are constants, we have:

$$\text{SSIM}(\hat{\text{y}}, \text{y}) = \frac{(2\mu_{\hat{y}}\mu_y + C_1) + (2\sigma_{\hat{y}y} + C_2)}{(\mu_{\hat{y}}^2 + \mu_y^2 + C_1)(\sigma_{\hat{y}}^2 + \sigma_y^2 + C_2)} \tag{3}$$

Cosine similarity [Eq. 4] computes the angular nearness between the generated images and the ground truth. The nearer the predicted image is to the ground truth, the angle (θ) between them tends to zero. Thereby, the angular similarity, $\cos(\theta)_{\theta \to 0} \to 1$. Thus a value closer to 1 is desirable. More closer is it's value to 1, higher is the image quality. Given $\hat{y}$ is the generated pixel value and y is the ground truth, we have:

$$\cos(\hat{y}, y) = \frac{\hat{y} \cdot y}{\|\hat{y}\| \|y\|} = \frac{\sum_{i=1}^{n} \hat{y}_i y_i}{\sqrt{\sum_{i=1}^{n} (\hat{y}_i)^2} \sqrt{\sum_{i=1}^{n} (y_i)^2}} \tag{4}$$

5 Results

5.1 Datasets

We experimentally validate our framework on the of the most popular remote sensing datasets having high spatial resolution: EuroSAT (MS) [4], PatternNet [17] and UC Merced land-use [15] dataset. The EuroSAT (MS) dataset consists of 10 classes having 27000 images of 13 spectral channels of the Sentinel-2 satellite. Each image consists of 3 channels with 64 × 64 pixels per channel. We use the RGB subset of EuroSAT (MS). The PatternNet dataset is a collection of 30400 images from 38 classes, 800 images per class. The images are collected from Google earth imagery. UC Merced is a land-use dataset spanning 21 classes having 100 images per class. It was extracted from large images from the USGS National Map Urban Area Imagery collection.

As seen in Table 1, our model outperforms other models in terms of PSNR and cosine similarity and MZSR marginally outperforms us only in the SSIM domain. One of the causes for this can be diverse changes in color gamut along

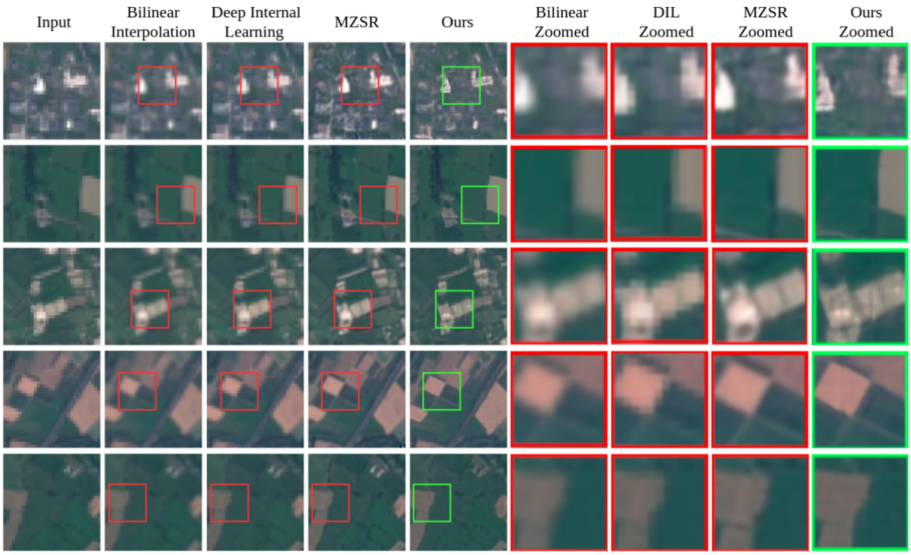


Fig. 5. A visual depiction of comparative model outputs on UC Merced dataset. It can be clearly seen in the zoomed boxes that our proposed model generates a crisper SR image as compared to others.

Table 1. Quantitative performance of compared models in terms of PSNR, SSIM and cosine similarity on the UC Merced dataset for 2× scaling.

Model	PSNR	SSIM	Cosine Similarity
Bi-linear Interpolation	28.10	0.9754	0.9623
Deep internal Learning [11]	30.14	0.9867	0.9806
Meta Transfer ZSR [12]	29.92	**0.9916**	0.9824
Ours	**31.27**	0.9911	**0.9891**

Table 2. Quantitative performance of compared models in terms of PSNR, SSIM and cosine similarity on the EuroSAT dataset for 2× scaling.

Model	PSNR	SSIM	Cosine Similarity
Bi-linear Interpolation	29.87	0.9777	0.9865
Deep internal Learning [11]	31.34	0.9832	0.9913
Meta Transfer ZSR [12]	31.88	0.9802	0.9932
Ours	**32.54**	**0.9834**	**0.9941**

the edges of the super-resolved images. By basic visual inspection, it can be seen that our model generates images and acts as a de-blurring model also (Fig. 5).

Table 3. Quantitative performance of compared models in terms of PSNR, SSIM and cosine similarity on the PatternNet dataset for 2× scaling.

Model	PSNR	SSIM	Cosine Similarity
Bi-linear Interpolation	28.04	0.9689	0.9733
Deep internal Learning [11]	28.86	0.9721	0.9876
Meta Transfer ZSR [12]	29.78	0.9763	0.9885
Ours	**30.27**	**0.9789**	**0.9913**

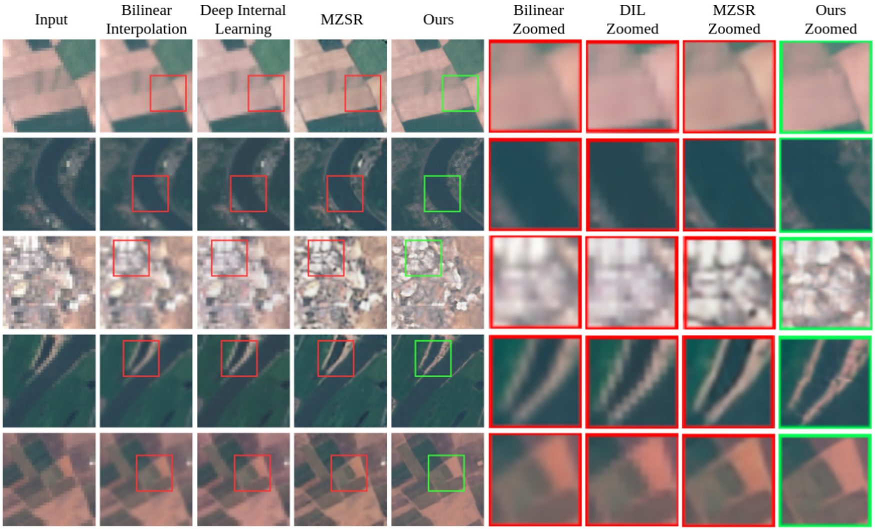


Fig. 6. A visual depiction of comparative model outputs on EuroSAT dataset. It can be clearly seen in the zoomed boxes that our proposed model generates a crisper SR image as compared to others.

As seen in Table 2, our frame outperforms other models in terms of PSNR and cosine similarity. Deep internal learning comes closer to our model by a fraction only in the SSIM domain. In this dataset, our model handles upscaling and de-blurring in a balanced mode to generate high fidelity images (Fig. 6).

As seen in Table 3, even though our model outperforms other models in terms of PSNR, cosine similarity and MZSR [12], it proves to be a difficult dataset to handle. This is due to a large number of high-frequency components present in considerable classes of urban zones. By basic visual inspection (Fig. 7), it can be seen that our model generates images and handles high-frequency components well.

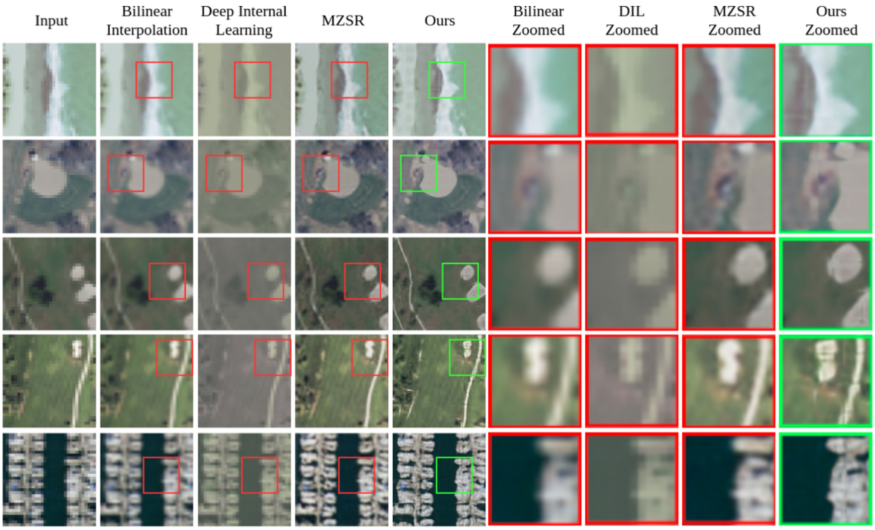


Fig. 7. A visual depiction of comparative model outputs on PatternNet dataset. It can be clearly seen in the zoomed boxes that our proposed model generates a crisper SR image as compared to others.

5.2 Ablation Studies

For ablation studies, we run the network without the attention score sharing to see its impact. Ablation is also performed on the number of modules used in the cascade. This is to validate the learning effectiveness of the attention sharing in the latent space as well as how deep the attention sharing needs to go.

Table 4. Ablation without attention score sharing for 2× scaling.

Dataset	PSNR	SSIM	Cosine Similarity
UC Merced	30.12	0.9807	0.9762
EuroSAT	31.04	0.9719	0.9852
PatternNet	29.20	0.9701	0.9877

From Table 4 and Table 5, we can infer that the attention score sharing plays a crucial role in generating images that are highly coherent with the ground truth. The absence of score sharing reduces the performance of the model. Also, increasing the number of modules in the stack does boost performance. But after a certain number of modules, the performance reaches a plateau and further increase of modules.

Table 5. Ablation on module number in stack for EuroSAT dataset for 2× scaling.

Module No	PSNR	SSIM	Cosine Similarity
4	30.89	0.9763	0.9889
6	32.54	0.9834	0.9941
8	32.50	0.9825	0.9924

6 Extended Application Results

We showcase **8K** (Fig. 8) some of the results with no prior SR images available as reference. To maintain the zero-shot scenario, we train the model on remotely sensed images and test on a different image, i.e. telescopic image and a person's image. The photos are highly scaled to get an 8k image. Given that current devices have patch-based selective upsampling when being zoomed on, a sliding window approach is used to generate the SR images from LR images. The artifacts arise due to the patch being stitched back to give the whole zoomed-out picture given the limited size of the window.

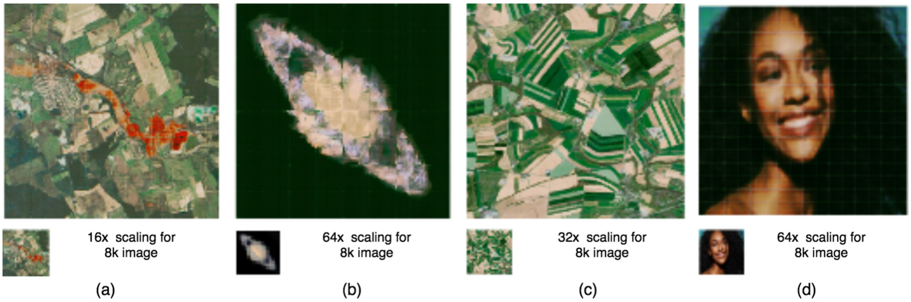


Fig. 8. The model that trained on the smaller figure of (a) enhanced (a) as well as (b) in the testing phase. Similarly, the model that trained on the smaller figure of (c) enhanced (c) as well as (d) in the testing phase. The artifacts are due to stitching of grid sampled images in the low resolution input.

7 Conclusion and Future Work

We introduce a tessellation based zero-shot super-resolution framework that utilizes instance-based statistics and image continuity. This helps in the efficient generation of high fidelity super-resolved images in cases with no prior references. The cascaded attention sharing network aids in selective highlighting of features throughout the feature space interactions which helps in building robust content-based HR images with superior quality. The auto-encoder style format aids in capturing high-frequency signals which in turn outputs crisp and sharp HR images. We also show extended efficient upscaling results that can be applied to non RS domains. These tessellations being based on image continuity opens up new avenues in traditional computer vision areas as well as remotely sensed domains. It can help in efficient information storage and transmission by utilizing highly compressed images that can be restored at the consumer end.

Acknowledgements. The authors would like to thank the members of PerceptX labs[1] for their valuable feedback that contributed to this research.

References

1. Cheng, X., Fu, Z., Yang, J.: Zero-shot image super-resolution with depth guided internal degradation learning. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12362, pp. 265–280. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58520-4_16
2. Demiray, B.Z., Sit, M., Demir, I.: D-SRGAN: dem super-resolution with generative adversarial networks. SN Comput. Sci. **2**(1), 1–11 (2021). https://doi.org/10.1007/s42979-020-00442-2
3. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks (2015)
4. Helber, P., Bischke, B., Dengel, A., Borth, D.: EuroSAT: a novel dataset and deep learning benchmark for land use and land cover classification (2019)
5. Ji, H., Gao, Z., Mei, T., Ramesh, B.: Vehicle detection in remote sensing images leveraging on simultaneous super-resolution. IEEE Geosci. Remote Sens. Lett. **17**(4), 676–680 (2020). https://doi.org/10.1109/LGRS.2019.2930308
6. Lanaras, C., Bioucas-Dias, J., Galliani, S., Baltsavias, E., Schindler, K.: Super-resolution of sentinel-2 images: learning a globally applicable deep neural network. ISPRS J. Photogramm. Remote Sens. **146**, 305–319 (2018). https://doi.org/10.1016/j.isprsjprs.2018.09.018
7. Ledig, C., et al.: Photo-realistic single image super-resolution using a generative adversarial network (2017)
8. Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M.: Enhanced deep residual networks for single image super-resolution (2017)
9. Müller, M.U., Ekhtiari, N., Almeida, R.M., Rieke, C.: Super-resolution of multispectral satellite images using convolutional neural networks. ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci. **V-1-2020**, 33–40 (2020). https://doi.org/10.5194/isprs-annals-v-1-2020-33-2020
10. Rabbi, J., Ray, N., Schubert, M., Chowdhury, S., Chao, D.: Small-object detection in remote sensing images with end-to-end edge-enhanced GAN and object detector network (2020)
11. Shocher, A., Cohen, N., Irani, M.: “zero-shot” super-resolution using deep internal learning (2017)
12. Soh, J.W., Cho, S., Cho, N.I.: Meta-transfer learning for zero-shot super-resolution (2020)
13. Wang, X., et al.: ESRGAN: enhanced super-resolution generative adversarial networks. In: Leal-Taixé, L., Roth, S. (eds.) ECCV 2018. LNCS, vol. 11133, pp. 63–79. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-11021-5_5

[1] PerceptX labs: https://sites.google.com/view/perceptx/home.

14. Xu, W., Xu, G., Wang, Y., Sun, X., Lin, D., Wu, Y.: High quality remote sensing image super-resolution using deep memory connected network. In: IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium, July 2018. https://doi.org/10.1109/igarss.2018.8518855
15. Yang, Y., Newsam, S.: Bag-of-visual-words and spatial extensions for land-use classification. In: ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM GIS) (2010)
16. Zhang, J., Wang, Z., Zheng, Y., Zhang, G.: Cascaded convolutional neural network for image super-resolution. In: Sun, X., Zhang, X., Xia, Z., Bertino, E. (eds.) ICAIS 2021. CCIS, vol. 1422, pp. 361–373. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-78615-1_32
17. Zhou, W., Newsam, S., Li, C., Shao, Z.: PatternNet: a benchmark dataset for performance evaluation of remote sensing image retrieval. ISPRS J. Photogramm. Remote Sens. **145**, 197–209 (2018). https://doi.org/10.1016/j.isprsjprs.2018.01.004