Journal of *Imaging*

MDPI

Article

OW-SLR: Overlapping Windows on Semi-Local Region for Image Super-Resolution

Rishav Bhardwaj [1,*], Janarthanam Jothi Balaji [2] and Vasudevan Lakshminarayanan [1]

[1] School of Optometry and Vision Science, University of Waterloo, Waterloo, ON N2L 3G1, Canada; vengulak@uwaterloo.ca
[2] Department of Optometry, Medical Research Foundation, Chennai 600006, India; jothibalaji@gmail.com
* Correspondence: rishav.bhardwaj@uwaterloo.ca

Abstract: There has been considerable progress in implicit neural representation to upscale an image to any arbitrary resolution. However, existing methods are based on defining a function to predict the Red, Green and Blue (RGB) value from just four specific loci. Relying on just four loci is insufficient as it leads to losing fine details from the neighboring region(s). We show that by taking into account the semi-local region leads to an improvement in performance. In this paper, we propose applying a new technique called Overlapping Windows on Semi-Local Region (OW-SLR) to an image to obtain any arbitrary resolution by taking the coordinates of the semi-local region around a point in the latent space. This extracted detail is used to predict the RGB value of a point. We illustrate the technique by applying the algorithm to the Optical Coherence Tomography-Angiography (OCT-A) images and show that it can upscale them to random resolution. This technique outperforms the existing state-of-the-art methods when applied to the OCT500 dataset. OW-SLR provides better results for classifying healthy and diseased retinal images such as diabetic retinopathy and normals from the given set of OCT-A images.

Keywords: super-resolution; OCT-A; implicit neural representation; retina; diabetic retinopathy; opthalmic images

Citation: Bhardwaj, R.; Jothi Balaji, J.; Lakshminarayanan, V. OW-SLR: Overlapping Windows on Semi-Local Region for Image Super-Resolution. *J. Imaging* **2023**, *9*, 246. https://doi.org/10.3390/jimaging9110246

Academic Editor: Donald Bailey

Received: 28 August 2023
Revised: 24 October 2023
Accepted: 31 October 2023
Published: 8 November 2023

1. Introduction

The primary objective of super resolution (SR) is to obtain a credible high resolution (HR) image from a low resolution (LR) image. The major challenge is to retrieve the information which is too minute or almost non existent, and to extrapolate this information to higher dimensions which is plausible to the human eye. Furthermore, the availability of paired HR-LR image data poses another concern. Typically, an image is downsampled using a specific method in the hope of encountering a real-life LR image that is somewhat similar. The aim of SR models is to fill in the deficient information between the HR and LR images, thereby bridging the gap. Also, for high-dimensional inputs like videos and 3D scans there are quite a few work in the literature [1–6].

Most of the architectures [7–11] proposed for SR of images upsample them by a fixed factor only. This means that a separate architecture needs to be trained for each unseen upscaling factor. However, the real world is continuous in nature, whereas images are represented and stored as discrete values in 2D arrays. Inspired by [12–15] for 3D shape reconstruction using implicit neural representation, ref. [16] proposed Local Implicit Image Function (LIIF) to represent images in a continuous fashion. Some postprocessing is performed to obtain the RGB value of the query point. This approach enables representing and manipulating images in a continuous manner, departing from the traditional discrete representation in 2D arrays.

In our work, we draw partial inspiration from advancements in 3D shape reconstruction, but we extend the approach by considering a semi-local region rather than relying

solely on four specific locations. Our method allows for extrapolation to any random upscaling factor using the same architecture. This architecture takes into account the semi-local region and specifically learns to extract important details related to a query point in the latent space that needs to be upscaled. In this paper, we propose an image representation technique called Overlapping Windows for Semi-Local Representation in a continuous domain and we fine our work as follows: (i) Each image is represented as a set of latent codes, establishing a continuous nature. To determine the RGB value of a point in the HR image within the latent space, we employ a decoding function. (ii) This semi-local region is fed into network as input which generates the embeddings of the intricate details in it which have high probability of getting lost when an entire image is taken into consideration by the networks. (iii) The overlapping window technique allows for effective learning of features within the semi-local region around a point in the latent space using the embeddings. (iv) A decoder takes the features derived from the overlapping window technique and produces the RGB value of the corresponding point in the HR image.

In summary, our work makes two key contributions. Firstly, we introduce a novel technique called overlapping windows, which enables efficient learning of features within the semi-local region around a point. This approach allows for more effective representation and extraction of important details. Secondly, our architecture is capable of upscaling an image to any arbitrary factor, providing flexibility and versatility without the need for separate architectures for different upscaling factors. This contribution enables seamless and consistent image upscaling using a unified framework. The project page is available at https://rishavbb.github.io/ow-slr/index.html.

2. Related Work

During the early stages of SR research, images were typically upsampled by a certain factor using simple interpolation techniques, and the network was trained to learn the extrapolation of the LR images [17,18]. However, this approach presents some issues. Firstly, the pre-upsampling process introduces more parameters compared to the post-upsampling process. Pre-upsampling is defined as upscaling the input image and then passing it through the network, whereas post-upsampling is defined as passing the image through the network and then upscaling the feature map. Secondly, due to the higher requirement of parameters more training time becomes a requisite. The network needed to learn the intricacies of the pre-upsampling method, which added to the overall training complexity. Finally, the pre-upsampling process using traditional bicubic interpolation does not yield realistic results during testing. Since it is the first step of the SR pipeline, the network often attempts to mimic this interpolation, which limits the realism of the output images. On the other hand, post-upsampling approaches, where the LR image is downscaled in the very first step, typically involve the use of bicubic interpolation for resizing. However, downscaling an image, even with bicubic interpolation, tends to yield more realistic results compared to upscaling. As a result, the research focus has shifted towards post-upsampling techniques, which provides more efficient and realistic SR results by leveraging downscaling with appropriate interpolation methods in the very first step.

As already mentioned, downscaling of images happens as the initial step in post-upsampling process. The network learns features from the downscaled image and the upsamples the learned features towards the very end. A technique proposed by Shi et al. in their work [8] is known as sub-pixel convolution. Sub-pixel convolution handles the extrapolation of each pixel by accumulating the features along the channel of that pixel. By rearranging the feature channels, sub-pixel convolution enables the network to effectively upscale the LR image to a higher resolution. While sub-pixel convolution provides a practical solution for upsampling by integral factors ($\times 1$, $\times 2$, $\times 3$, etc.), it does not support fractional upsampling factors ($\times 1.4$, $\times 2.9$, etc.). However, for cases where fixed integral upsampling factors are sufficient, sub-pixel convolution offers an efficient approach to achieving high-quality upsampling. The work by Ledig et al. [19] introduced the use of multiple residual blocks for feature extraction in super-resolution (SR) tasks. Their approach

demonstrated the effectiveness of residual blocks in capturing and enhancing image details. Building upon Ledig et al.'s work, Lim et al. [11] proposed an enhanced SR model that incorporated insights regarding batch normalization. They postulated that removing batch normalization from the residual blocks could lead to improved performance for SR tasks. This is because batch normalization tends to normalize the input, which may reduce the network's ability to capture and amplify the fine details required for SR. Removing batch normalization not only results in a reduction in memory requirements but also makes the network faster. Additionally, the work by Shi et al. [8] contributed to the development of various approaches for SR using CNNs. These approaches include methods proposed by [9,19–21]. These methods aimed to enhance feature extraction capabilities specifically tailored for SR problems, further advancing the state-of-the-art in SR research.

After the success of CNNs in SR tasks, researchers explored the use of generative adversarial networks (GANs) to further improve SR performance. Several works, such as [19,22,23], introduced different GAN architectures for extrapolating low-resolution (LR) images to higher resolution. ESRGAN (Enhanced Super-Resolution Generative Adversarial Network) proposed by Wang et al. [24] introduced a perceptual loss function and modified the generator network to produce HR images. This perceptual loss function aimed to align the visual quality of the generated HR images with that of the ground truth HR images, improving the perceptual realism of the results.

In Real-ESRGAN [25], the authors addressed the issue of using LR images downsampled with simple techniques like bicubic interpolation during training. They note that real-world LR images undergo various types of degradations, compressions, and noise, unlike the simple interpolation-based downsampling. To simulate realistic LR images during training, they proposed a novel technique that subjected the training images to various degradation processes, mimicking real-life scenarios. Additionally, Real-ESRGAN introduced an U-Net discriminator to enhance the adversarial training process and improve the quality of the generated HR images.

3. Method

We illustrate the three main components of our approach in this section along with its pictorial representation in Figure 1. In Section 3.1, we introduce the backbone of our framework. We represent the LR image as a feature map, which serves as the basis for subsequent processing and analysis. In Section 3.2, we demonstrate how we find the semi-local region of an arbitrary point in the HR image. This region contains valuable information that helps determine the corresponding RGB value. In Section 3.3, we highlight the Overlapping Windows technique, which plays a crucial role in predicting the RGB value of a point in the HR image. We accomplish this by leveraging the semi-local region extracted around the sampling points of the feature map. These three parts collectively form the foundation of our approach, allowing for accurate prediction of RGB values.

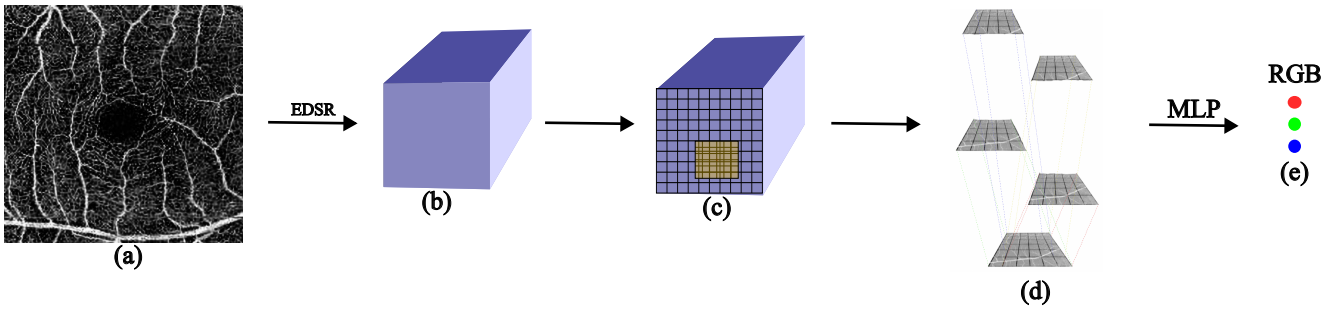

Figure 1. (**a**) An LR image is taken. (**b**) It is passed through EDSR [11] and a feature map is produced. (**c**) Locating the semi-local region (M = 6) around a random selected point from HR image. (**d**) Semi-local region is passed through the proposed Overlapping Windows. (**e**) This output is passed through the MLP to give out the RGB value of a randomly selected point. Steps (**c**–**e**) are performed for all the points in the HR image.

3.1. Backbone Framework

To extract features from the LR image, we employ the enhanced deep residual networks (EDSR) [11]. Specifically, we utilize the baseline architecture of EDSR, which consists of 16 residual blocks.

$$\psi = EDSR(I_{LR}) \tag{1}$$

Given an LR image denoted as $I_{LR} \in \mathbb{R}^{H \times W \times C}$, we express it in the form of a feature map $\psi \in \mathbb{R}^{P \times Q \times D}$. Here, H and W represent the height and width of the LR image, respectively, and C signifies the number of channels. P and Q represent the spatial dimensions of the feature map, and D denotes the depth of the feature map.

3.2. Locating the Semi-Local Region

In our scenario, we aim to predict the RGB value at any random point in a continuous HR image of arbitrary dimensions. Let $I_{HR} \in \mathbb{R}^{X \times Y \times C}$ represent the HR image. To predict the RGB value at a specific point, we first select a point of interest. Then, we identify its corresponding spatially equivalent point in the feature map ψ obtained from the LR image using bilinear interpolation denoted as $\mho_{BI}$.

$$\hat{\boldsymbol{x}} = \mho_{BI}(x, \psi) \tag{2}$$

where $\hat{\boldsymbol{x}}$ and x are the 2D coordinates of the ψ and I_{HR} respectively.

Furthermore, we extract a square semi-local region around this corresponding point. The size of this region is determined by a length parameter M units, where each unit dimension of the square region corresponds to the inverse of the dimensions P and Q of the feature map ψ along its length and breadth respectively defined in Equation (5) which is used to find the discrete positions in the semi-local region. Once we have identified the square semi-local region around the corresponding point in the feature map ψ, we proceed to extract $M \times M$ depth features from this region using Equation (3). These depth features capture the important information necessary for predicting the RGB value at the desired point in the HR image. To extract these features, we employ a closest Euclidean distance approach denoted by $\eth_{ED}$. Each point within the $M \times M$ region in ψ is mapped to the nearest point in the latent space, which represents the extracted depth feature. Figure 2 illustrates the working of selecting of features from the feature map. This mapping ensures that we capture the most relevant information from the semi-local region.

$$\hat{\boldsymbol{X}} = (\hat{\boldsymbol{x}}_x - \psi_x * i, \hat{\boldsymbol{x}}_y - \psi_y * j) \tag{3}$$

$$i = \{\frac{-M}{2}, \frac{-M}{2} + 1, \ldots \frac{+M}{2} - 1, \frac{+M}{2}\}, j = \{\frac{-M}{2}, \frac{-M}{2} + 1, \ldots \frac{+M}{2} - 1, \frac{+M}{2}\} \tag{4}$$

$$\psi_x = \frac{1}{P}, \psi_y = \frac{1}{Q} \tag{5}$$

Thus $\hat{\boldsymbol{X}}$ holds the 2D coordinates of all the $M \times M$ points.

$$S = \eth_{ED}(\hat{\boldsymbol{X}}, \psi) \tag{6}$$

Figure 3 illustrates how the semi-local region is identified and used to extract the $M \times M$ depth features from the feature map ψ. This depiction helps to visualize the steps involved in the feature extraction process.

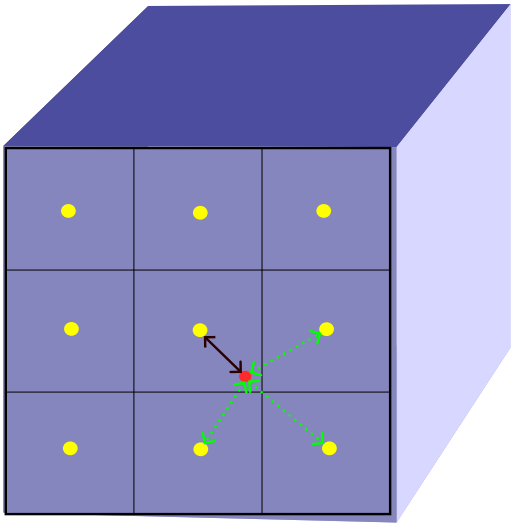

Figure 2. To extract features from a feature map of size 3×3, we focus on a specific query point represented by a red dot. In order to determine which pixel locations in the feature map correspond to this query point, we compute the Euclidean distance between the query point and the center points of each pixel location. In the provided image, the black line represents the closest pixel location in the feature map to the query point.

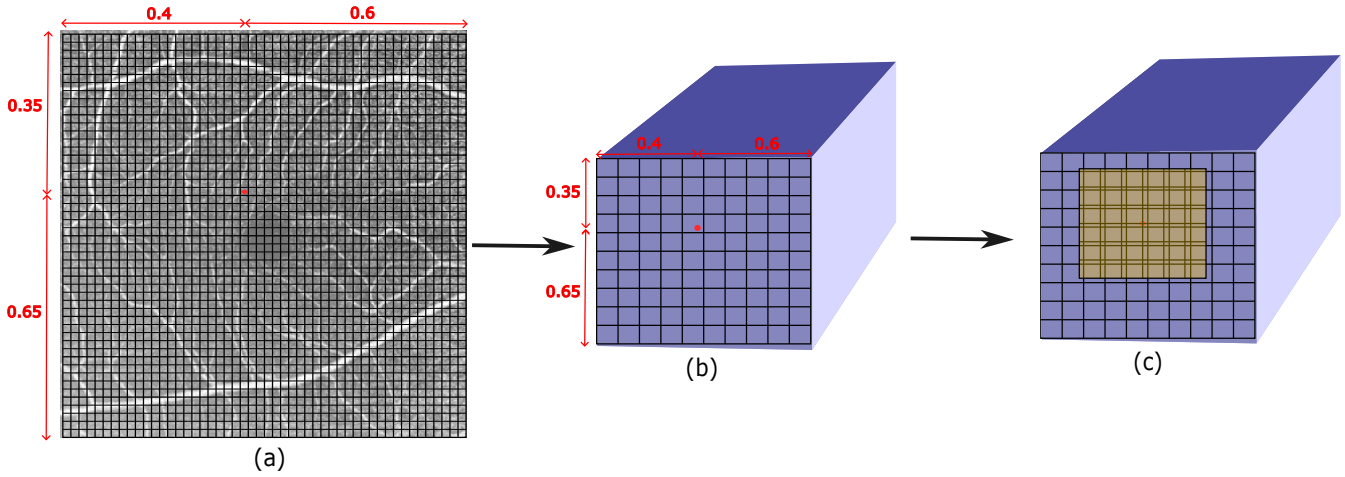

Figure 3. (**a**) Given an HR image, a point of interest (red dot) is selected to predict its RGB value. (**b**) Its corresponding spatially equivalent 2D coordinate is selected from the feature map. (**c**) Locating the semi-local region (M = 6) around the calculated 2D coordinate.

3.3. Overlapping Windows

After extracting the semi-local region $S \in \mathbb{R}^{M \times M \times D}$, our objective is to obtain the RGB value of the center point using this region. To achieve this, we employ a overlapping window-based approach. We start with four windows, each with a size of $M - 1$, positioned at the four corners of S. Each window extracts information from its respective region and passes it on to the next subsequent window in the process. With each iteration, the size of the window decreases by 1 until it reaches a final size of $\frac{M}{2}$. This iterative process ensures that information is progressively gathered and refined towards the center point. This approach allows us to effectively capture and utilize the information from the semi-local region while focusing on the features that are most relevant for determining the RGB value.

$$\Gamma = s_i * w_i \tag{7}$$

In each iteration i, where the window size decreases by 1 for the next step, we utilize weights w_i for combining the features from all four corners. This ensures that the information from each corner is properly incorporated and made available for the subsequent iteration. In the last step, we take a final window size of 2, but instead of being positioned at the corners as in previous iterations, it is centered around the target point of interest. The features extracted from this final window are then passed through a Multi-Layer Perceptron (MLP) to make the final prediction.

By adapting the window positions and sizes throughout the iterations, we effectively capture and aggregate the relevant information from the semi-local region. This approach allows us to make accurate predictions at the target point, utilizing the combined features from all iterations and the final MLP-based processing. Figure 4 shows the working of the overlapping windows.

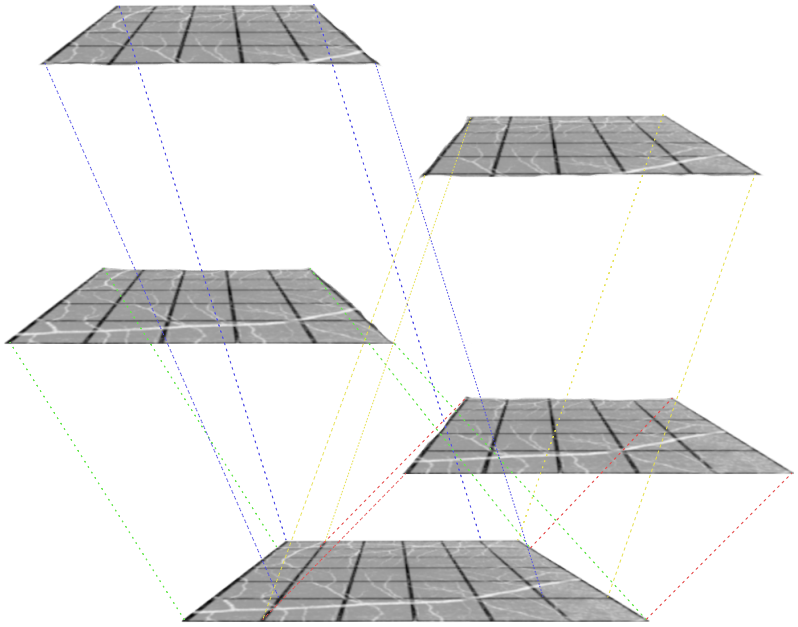

Figure 4. The first iteration of overlapping windows, where the window size = M − 1 (M = 6). Assuming the feature map is of negligible depth and four windows are positioned at the four corners of the feature map.

4. Results and Discussion

4.1. Dataset

We used the OCT500 [26] dataset and randomly sampled 524 images from it to train our network. It consists of 300 3 × 3 OCTA images and 224 6 × 6 OCTA images. We use For evaluation, 80 images were selected and we report the results using peak signal-to-noise ratio (PSNR) metric.

4.2. Implementation Details

During the training process, we apply downsampling to each image using bicubic interpolation in PyTorch [27]. This downsampling is performed by selecting a random factor, which introduces the desired level of degradation to the images. For training, we utilize a batch size of 16 images. From each high-resolution (HR) image, we randomly select 1500 points for which we aim to calculate the RGB values. These points serve as the targets for our network during the optimization process.

To optimize the network, we employ the L1 loss function and use the Adam optimizer [28]. The learning rate is initialized as 1×10^{-4} and is decayed by a factor of 0.3 at specific epochs, namely [40, 60, 70]. We train the network for a total of 100 epochs, allowing it to learn the necessary representations and refine its predictions over time.

Furthermore, each LR image is converted into a feature map of size 48 × 48 with a depth of 64 using the EDSR-baseline architecture. This conversion process ensures that the LR images are properly represented and aligned with the architecture used in the training process.

4.3. Quantitative Results

In Figure 5, we present a comparison of the performance of our proposed OW-SLR method against existing works. The original image patch is first downsampled using bicubic interpolation to a lower resolution. It is evident that there is a significant loss of image quality in the LR patches compared to the ground truth (GT) image. However, our model outperforms the other existing methods, demonstrating a significant improvement when the LR image is extrapolated to a higher scale. The results obtained by our model show better preservation of details and higher fidelity compared to the other approaches when the given image is extrapolated to higher scale. The PSNR results of each image are shown in Table 1.

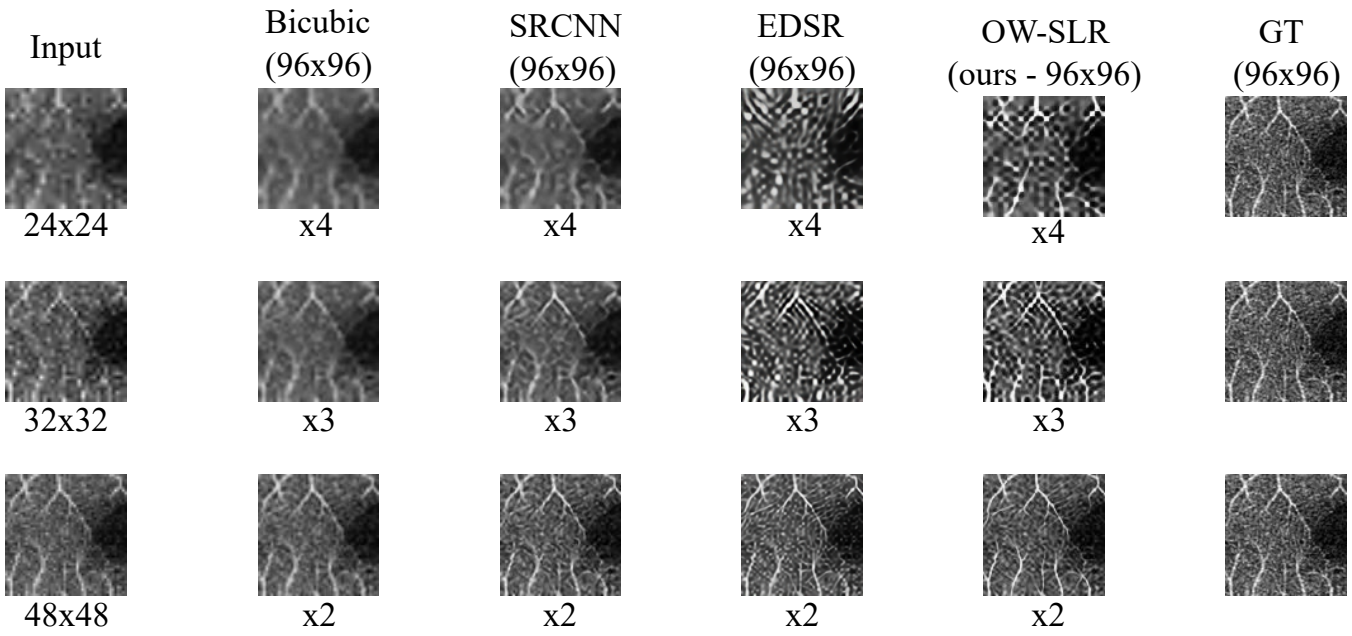

Figure 5. A 96 × 96 patch is taken and its size is reduced to 24 × 24 (first row), 32 × 32 (second row) and 48 × 48 (third row) using bicubic interpolation. Our architecture uses the same set to weights reproduce the given results. However, others require different set of weights for a newer scale to be trained on. The PSNR results of each image are shown in Table 1.

Table 1. PSNR result of each of the input images across different methods shown in Figure 5.

Patch Size	Bicubic	SRCNN [17]	EDSR [11]	OW-SLR (Ours)
24 × 24	11.96	12.87	13.79	13.92
32 × 32	14.18	15.10	16.04	16.26
48 × 48	15.37	16.89	17.66	17.98

It is worth noting that our model achieves these results for different scaling factors using the same set of weights trained once. In contrast, the other models would need to be retrained for each new scale to which the LR image is extrapolated. This highlights the versatility and efficiency of our model in handling various scaling factors without the need for additional training.

In Table 2, we provide the upscaling time taken by the proposed model by different factors, while training it just once.

Table 2. Time taken to extrapolate a 320 × 320 image on a single Nvidia Titan V of 12 Gigabyte size.

Extrapolation Factor	Time Taken (In Seconds)
2×	6.48
2.4×	8.90
3×	12.01
3.9×	19.46
4.5×	26.29
5×	33.75

In Table 3, we present the results of this technique compared to the existing state-of-the-art methods on the OCT500 [26] dataset. The evaluation metric used in this case is the peak signal-to-noise ratio (PSNR). Our work demonstrates superior performance compared to LIIF, highlighting the effectiveness of considering the semi-local region instead of solely focusing on four specific locations. By incorporating the information from the semi-local region, our approach achieves improved results in terms of PSNR, showcasing the benefits of our methodology for super-resolution tasks.

Table 3. PSNR result on the 300 images from OCT500 [26].

Methods	PSNR ↑
Real-ESRGAN [24]	15.66
SRCNN [17]	16.51
EDSR [11]	17.49
LIIF [16]	17.60
OW-SLR (ours)	**17.93**

5. Conclusions and Future Work

OCTA images help us for the diagnosis of retinal diseases. However, due to various reasons like speckle noise, movement of the eye, hardware incapabilities, etc. we lose onto intricate details in the capillaries that play a crucial role for correct diagnosis. We propose this architecture which upscales a given LR image to arbitrary higher dimensions with enhanced image quality. First, we extract the image features using a backbone architecture. We then select a random point in the HR image and calculate its equivalent spatial point in the extracted feature map. We find the semi-local region around this calculated point and pass it through the proposed Overlapping Windows architecture. Finally, an MLP is used to predict the RGB value using the output of the overlapping window architecture. We hope our work will help the people in the medical field in their diagnosis. PSNR 17.93 is achieved for the OCT500 dataset which outperforms the other state-of-the-art work. The technique outperforms the existing methods and allows upscaling images to arbitrary resolution by training the architecture just once.

While effective, it is worth noting that this algorithm does come with a slightly higher computational cost due to its consideration of the semi-local region. There remains potential for further enhancements in both computational efficiency and accuracy while taking the semi-local region into account. This work will provide a stepping stone for future researchers to make strides in this direction.

Author Contributions: R.B. has contributed towards implementation of the entire architecture, making the repository publicly available and writing the manuscript, J.J.B. contributed towards knowledge of the data worked upon from the clinical side and article editing, V.L. contributed to advice, suggestions, funding and article editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by University of Waterloo research grant.

Data Availability Statement: All generated data are presented in the article body.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bashir, S.M.A.; Wang, Y.; Khan, M.; Niu, Y. A comprehensive review of deep learning based single image super-resolution. *PeerJ Comput. Sci.* **2021**, *7*, 1–56. [CrossRef] [PubMed]
2. Tan, R.; Yuan, Y.; Huang, R.; Luo, J. Video super-resolution with spatial-temporal transformer encoder. In Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), Taipei, Taiwan, 18–22 July 2022; pp. 1–6.
3. Li, H.; Zhang, P. Spatio-temporal fusion network for video super-resolution. In Proceedings of the International Joint Conference on Neural Networks, Online, 18–22 July 2021.
4. Thawakar, O.; Patil, P.W.; Dudhane, A.; Murala, S.; Kulkarni, U. Image and video super resolution using recurrent generative adversarial network. In Proceedings of the 16th IEEE International Conference on Advanced Video and Signal Based Surveillance AVSS 2019, Taipei, Taiwan, 18–21 September 2019.
5. Li, H.; Yang, Y.; Chang, M.; Feng, H.; Xu, Z.; Li, Q.; Chen, Y. SRDiff: Single image super-resolution with diffusion probabilistic models. *Neurocomputing* **2022**, *479*, 47–59. [CrossRef]
6. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Region-based convolutional networks for accurate object detection and segmentation. *IEEE Trans. Pattern. Anal. Mach. Intell.* **2016**, *38*, 142–158. [CrossRef] [PubMed]
7. Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Van Gool, L.; Timofte, R.S. Image restoration using swin transformer. In Proceedings of the IEEE/CVF international Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 1833–1844.
8. Shi, W.; Jose, C.; Ferenc, H.; Johannes, T.; Andrew, A.P.; Rob, B.; Daniel, R.; Wang, Z. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1874–1883.
9. Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; Fu, Y. Image super-resolution using very deep residual channel attention networks. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 286–301.
10. Peng, S.; Niemeyer, M.; Mescheder, L.; Pollefeys, M.; Geiger, A. Convolutional Occupancy Networks. In Proceedings of the Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020; pp. 523–540.
11. Lim, B.; Son, S.; Kim, H.; Nah, S.; Mu Lee, K. Enhanced deep residual networks for single image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 136–144.
12. Saito, S.; Huang, Z.; Natsume, R.; Morishima, S.; Kanazawa, A.; Li, H. Pifu: Pixel-aligned implicit function for high-resolution clothed human digitization. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 2304–2314.
13. Park, J.J.; Florence, P.; Straub, J.; Newcombe, R.; Lovegrove, S. Deepsdf: Learning continuous signed distance functions for shape representation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 165–174.
14. Mescheder, L.; Oechsle, M.; Niemeyer, M.; Nowozin, S.; Geiger, A. Occupancy networks: Learning 3d reconstruction in function space. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 4460–4470.
15. Sitzmann, V.; Martel, J.; Bergman, A.; Lindell, D.; Wetzstein, G. Implicit neural representations with periodic activation functions. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 7462–7473.
16. Chen, Y.; Liu, S.; Wang, X. Learning continuous image representation with local implicit image function. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 8628–8638.
17. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 295–307. [CrossRef] [PubMed]
18. Kim, J.; Lee, J.K.; Lee, K.M. Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1646–1654.
19. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; 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, Honolulu, HI, USA, 21–26 July 2017; pp. 4681–4690.
20. Mei, Y.; Fan, Y.; Zhou, Y. Image super-resolution with non-local sparse attention. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 3517–3526.
21. Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; Fu, Y. Residual dense network for image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2472–2481.
22. Sajjadi, M.S.M.; Scholkopf, B.; Hirsch, M. Enhancenet: Single image super-resolution through automated texture synthesis. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 4491–4500.
23. Wang, X.; Yu, K.; Dong, C.; Loy, C.C. Recovering realistic texture in image super-resolution by deep spatial feature transform. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 606–615.
24. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; Chen, C.L.; Qiao, Y.; Tang, X. Esrgan: Enhanced super-resolution generative adversarial networks. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops, Munich, Germany, 8–14 September 2018.
25. Wang, X.; Xie, L.; Dong, C.; Shan, Y. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 1905–1914.

26. Li, M.; Chen, Y.; Yuan, S.; Chen, Q. OCTA-500. 2019. Available online: https://arxiv.org/ftp/arxiv/papers/2012/2012.07261.pdf (accessed on 23 October 2023).
27. Adam, P.; Sam, G.; Francisco, M.; Adam, L.; James, B.; Gregory, C.; Trevor, K.; Lin, Z.; Natalia, G.; Luca, A.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2019 ; Volume 32.
28. Kingma, D.P.; Ba, J.A. A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.