

IMPACT OF DEEP LEARNING-BASED SUPER-RESOLUTION ON BUILDING FOOTPRINT EXTRACTION

H. He¹, K. Gao¹, W. Tan¹, L. Wang¹, S. N. Fatholahi¹, N. Chen^{1,2}, M. A. Chapman³, J. Li^{*1,4}

¹ Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada

² College of Geological Engineering and Geomatics, Chang'an University, Xi'an, SX710054, China

³ Department of Civil Engineering, Ryerson University, Toronto, ON M5B 2K3, Canada

⁴ Department of Systems Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada

Commission I, WG I/2

KEY WORDS: Deep Learning, Super-resolution, Building Footprint Extraction, Impact, HRNet v2, the Massachusetts Building Dataset.

ABSTRACT

Automated building footprints extraction from High Spatial Resolution (HSR) remote sensing images plays important roles in urban planning and management, and hazard and disease control. However, HSR images are not always available in practice. In these cases, super-resolution, especially deep learning (DL)-based methods, can provide higher spatial resolution images given lower resolution images. In a variety of remote sensing applications, DL based super-resolution methods are widely used. However, there are few studies focusing on the impact of DL-based super-resolution on building footprint extraction. As such, we present an exploration of this topic. Specifically, we first super-resolve the Massachusetts Building Dataset using bicubic interpolation, a pre-trained Super-Resolution CNN (SRCNN), a pre-trained Residual Channel Attention Network (RCAN), a pre-trained Residual Feature Aggregation Network (RFANet). Then, using the dataset under its original resolution, as well as the four different super-resolutions of the dataset, we employ the High-Resolution Network (HRNet) v2 to extract building footprints. Our experiments show that super-resolving either training or test datasets using the latest high-performance DL-based super-resolution method can improve the accuracy of building footprints extraction. Although SRCNN based building footprint extraction gives the highest Overall Accuracy, Intersection of Union and F₁ score, we suggest using the latest super-resolution method to process images before building footprint extraction due to the fixed scale ratio of pre-trained SRCNN and low speed of convergence in training.

1. INTRODUCTION

1.1 Motivation

With rapid urbanization, automated mapping of buildings became important for urban planning and management (Rastogi et al., 2020). In natural hazard and disease control, it also plays a key role (Thomas et al., 2013). With the development of imaging techniques, remote sensing-based methods provide promising ways for building footprint extraction and building mapping. High spatial resolution (HSR) images can be used to extract detailed information and map buildings accurately. However, HSR images are sometimes unavailable in practice. In such circumstances, the super-resolution of existing data is both economical and practical compared to acquiring new data or waiting for a new sensor (Shermeyer and Van Etten, 2019).

When considering existing super-resolution methods, deep learning (DL) based super-resolution show higher performance compared to traditional methods (Dong et al., 2015). There are two classes of super-resolution techniques: single-image super-resolution (SISR) and joint-image super-resolution methods. The former super-resolves images by taking only low spatial resolution target images as input, while the latter takes auxiliary data to aid the super-resolution process. In most cases, the first one is preferred given limited availability of data. In remote sensing, DL-based SISR methods are widely used to super-resolve remote sensing images. However, from literature, we are only aware of few works exploring the impact of single-image super-resolution on downstream tasks. As such, we would like to

explore the impact of using cutting-edge DL-based SISR on building footprints extraction.

1.2 Objectives of the study

The paper aims at exploring the impact of super-resolution on building footprints extraction. We applied a bicubic interpolation, and a pre-trained Super-Resolution CNN (SRCNN) (Dong et al., 2015), Residual Channel Attention Network (RCAN) (Zhang et al., 2018), Residual Feature Aggregation Network (RFANet) (Liu et al., 2020) to super-resolve the Massachusetts building dataset (Mnih, 2013). A comparative study was then conducted using High Resolution Network (HRNet) v2 (Sun et al., 2019) for building footprint extraction from the original, the bicubically interpolated and the super-resolved Massachusetts building datasets. The contribution of this paper is the exploration of the DL-based SISR impact on building footprints extraction.

1.3 Related work

1.3.1 Super-resolution methods: Super-resolution refers to recovering high spatial resolution images from low spatial resolution images (Dong et al., 2005). According to the data used as model input, there are single-image super-resolution and joint-image super-resolution methods. The former does not require multiple modalities (in terms of resolutions and spectral characteristics) in the data, which makes it easier to implement. Therefore, it is widely used in different disciplines.

Starting from SRCNN, CNNs were introduced into single-image super-resolution, which outperformed conventional methods (Dong et al., 2005). To further improve the performance of single-image super-resolution, different methods were proposed with more complex architectures and more parameters (He et al., 2021). Nowadays, the state-of-the-art methods, including RCAN and RFANet for single-image super-resolution can directly super-resolve low spatial resolution images to higher spatial resolution without any pre-processing steps (such as the bicubic interpolation in SRCNN).

1.3.2 The application of CNN-based single-image super-resolution in Geoscience. In Geoscience, single-image super-resolution also showed its high performance in multiple applications. Ducouman and Fablet (2016) applied SRCNN in downscaling satellite-based sea surface temperature data. In 2017, Lei et al. (2017) proposed a Local-Global Combined Network (LGCNet) for super-resolving remote sensing images. Ma et al. (2018) introduced transfer learning in super-resolving remote sensing images and proposed a network named Transferred Generative Adversarial Network (TGAN). The proposed network achieved the highest performance with respect to the benchmark methods used in their experiment. In Tuna et al. (2018), SRCNN and Very Deep Super-Resolution (VDSR) were adapted to super-resolve remote sensing images. Shermeyer and Van Etten (2019) explored the effect of super-resolution in object detection from satellite images. Ma et al. (2019) combined wavelet transform and the recursive Res-Net (Tai et al., 2017) for super-resolving remote sensing images, which obtained highest performance in their comparative study. In He et al. (2021), SRCNN, Very Deep Super-Resolution (VDSR) and RCAN were used and compared in downscaling Gravity Recovery and Climate Experiment (GARCE) Terrestrial Water Storage data. In a recent work, Lei and Shi. (2021) proposed a Hybrid-scale Self-similarity Exploitation Network (HSENet), which applied a Single-scale Self-similarity Exploitation Module (SSEM) and a Cross-scale Connection Structure (CCS). The experiment demonstrated the high performance of HSENet compared to the state-of-the-art methods.

Apart from these supervised super-resolution networks, unsupervised super-resolution networks were also proposed and used in remote sensing. Haut et al. (2018) developed an hourglass generative network for super-resolution which did not require high-resolution images in the training process. In 2019, Wang et al. (2019) combined two generative CNNs and proposed the Cycle-CNN for super-resolving remote sensing images in an unsupervised manner. Their experiments showed the robustness and competitive performance of their methods compared to other methods.

As we mentioned above, single-image super-resolution were widely used in remote sensing images super-resolution. However, we are only aware of a few works exploring the impact of super-resolution on the down-stream tasks, especially for semantic segmentation and instance segmentation. Among research we summarized above, only Shermeyer and Van Etten (2019), and Lei and Shi (2021) explored the effect of super-resolution on object detection and remote sensing scene classification. Further exploration of the impact of single-image super-resolution on semantic segmentation is necessary. In this work, we selected building footprints extraction as the downstream task of single-image super-resolution and the effects thereof.

1.3.3 CNNs-based building footprints extraction. With the rapid development of the Fully Convolutional Networks (FCN), they were widely used in different remote sensing applications

(Zhu et al., 2017). For building footprint extraction, methods were proposed for extracting building masks and building boundaries using CNNs. For masks extraction, from ConvNet (Yuan et al., 2017), Res-U-Net (Xu et al., 2018), Deep Encoding Network (DE-Net) (Liu et al., 2019), Capsule Feature Pyramid Network (CapsFPN) (Yu et al., 2020) and Multipath Hybrid Attention Network (MHA-Net) (Cai et al., 2021) to the Coarse-to-fine Boundary Refinement Network (CBR-Net) (Guo et al., 2022), newly developed deep learning techniques were gradually introduced in this task, which includes the skip connection, the capsule Network, the attention scheme, the multi-scale feature extraction, etc. It can be noticed that with sophisticatedly designed architecture, the performance of building footprint extraction was improved gradually.

For boundaries delineation, two ways were developed. The first way was introducing a regularization process after the mask extraction. The second way was embedding active contour model or recurrent neural network in CNN based networks for end-to-end building delineation. For the first type of methods, the most representative of its kind was proposed by Zhao et al. (2018). They used a regularization process to generate the boundaries of building footprint from mask extraction results from Mask R-CNN. For end-to-end building delineation, Marcos et al. (2018) developed a Deep Structured Active Contours (DSAC) framework to extract building boundaries directly from input images. In recent work, Zhao et al. (2021) proposed a new network based on PolyMapper (Li et al., 2019), which combined recurrent neural network and convolutional neural network together to extract building boundaries directly from images. The new method proposed by Zhao et al. (2021) showed the highest performance compared to other building delineation methods in their experiments.

Although CNNs-based building footprint extraction methods were quite successful in both mask extraction and building delineation, the spatial resolution of input images limited its applications. As we mentioned in the previous section, single-images super-resolution provided a promising way to overcome the spatial resolution related issues in practice. However, the impact of super-resolution on building footprint extraction was still under explored, Therefore, in this work, we would like to mitigate the gap

2. METHOD

2.1 Data

Among all publicly available building datasets, we selected the Massachusetts Building Dataset in this work as it has a 1 m spatial resolution, which limits detailed information extraction. The dataset consists of 137, 4 and 10 1500×1500 aerial images of Boston area in the training, validation, and test datasets, respectively. Due to computational requirements, we cropped and padded the images and masks into small patches of size 512×512.

To explore the impact of DL-based SISR, we super-resolved the processed images to 30 cm per pixel using bicubically interpolation and pretrained SRCNN, RCAN and RFANet. We used bicubic interpolation masks as ground truth for both super-resolved images as we found the interpolation is sufficient to up-sample the simple geometric shapes in the masks. To explore the impact of super-resolution on building footprint extraction, we also applied High Resolution Network v2 (HRNet) (Sun et al., 2019) to extract building footprint from original and super-resolved images, which is a high-performance semantic

segmentation method. As we only care about the impact of super-resolution on building footprint extraction, we did not apply the innovative mask extraction methods or boundaries delineation methods in this work. Although, in theory, they are affected in the same way by super-resolution.

2.2 Methods

2.2.1 Super-resolution methods

RFANet is composed of a head part, a trunk part, and a reconstruction part. The first- and second-part perform shallow and deep feature extraction. The reconstruction part is an Efficient Sub-Pixel Convolutional Neural Network (ESPCNN) (Shi et al., 2016), which works to recover high spatial resolution images from feature maps generated from the trunk part. The trunk part consists of 30 Residual Feature Aggregation modules which are composed of 4 Enhanced Spatial Attention (ESA) blocks (Liu et al., 2020).

The architecture of RCAN is similar to RFANet. It is also composed of a head part, a trunk part, and a reconstruction part (including an upscale module and a convolutional layer). Each part serves a similar purpose as its counterpart in RFANet. The reconstruction part also has an ESPCNN. The trunk part for RCAN includes 10 residual groups, each of which has 20 residual channel attention blocks (Zhang et al., 2018).

The SRCNN was the first CNNs-based single-image super-resolution method, which is composed of three convolutional layers. These layers handle patch extraction and representation, non-linear mapping, and reconstruction, with kernel size of 9×9 , 1×1 and 5×5 , respectively (Dong et al., 2015).

For detailed architecture information, such as the number of filters and each kernel size, we direct the readers to the original works.

2.2.2 Building footprint extraction method

Given high performance of HRNet v2, we adopted it in this work for building footprint extraction. In HRNet v2, a combination of multi-resolution group convolutions and multi-resolution convolutions is embedded in four-stage networks, which results in the strong representation of features and high performance for semantic segmentation (Sun et al., 2019). For detailed architecture information, we direct the readers to the original work.

2.3 Evaluation metrics

For single-image super-resolution, we applied commonly used Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) (Liu et al., 2020) to evaluate the performance of pre-trained models on SWOOP dataset¹. To quantitatively evaluate building footprint extraction results, we applied Overall Accuracy (or average pixel accuracy), Intersection of Union, mean IoU (mIoU), Precision, Recall and F_1 score, which are widely used to evaluate CNNs-based building footprint extraction method (He et al., 2022). Details about these metrics are omitted here to conserve space.

2.4 Implementation details

For CNNs-based super-resolution, we set epochs as 20, batch size as 16, loss function as mean absolute error and optimizer as Adam. The learning rate was initialized as $1e-4$ and decreased to half every $1e5$ iterations. For HRNet v2, we used the Adam optimizer, $1e-4$, Jaccard loss and 100 as the optimizer, learning rate, loss function and epoch, respectively. We conducted the experiment using TensorFlow framework 2.4.0 on a single GeForce RTX 3090 GPU and CUDA 11.2.

3. RESULTS AND DISCUSSION

3.1 Qualitative evaluation

To qualitatively evaluate the impact of super-resolution on building footprint extraction, we selected two patches of images. We compared super-resolved images and extracted building footprints generated using different super-resolution methods by using HRNet v2 trained on the super-resolve images. As shown in Fig.1 to 5, in the first row are the original images and ground truth label of two examples scenes. From the second row to the fifth row², the super-resolved images and extracted building footprint were tabulated. In Fig.1, the extraction model HRNet v2 was trained on original images. In Fig.2 to Fig.5 the HRNet v2 models were trained using bicubically interpolated images, SRCNN super-resolved images, RCAN super-resolved images and RFANet super-resolved images. For simplification, in the rest of the paper, we denote the extraction model trained on bicubically interpolated, SRCNN super-resolved, RCAN super-resolved, RFANet super-resolved Massachusetts training building dataset as “BI-HRNet v2”, “SRCNN-HRNet v2”, “RCAN-HRNet v2” and “RFANet-HRNet v2”, respectively. The model trained on the original Massachusetts Building Dataset is denoted as “HRNet v2”.

As shown in Fig.2 to Fig.5, we found that the models trained on images super-resolved by the newer method achieves higher performance on all test datasets, with the model trained on RFANet super-resolved images obtaining the best results

Any particular trained model performs best when the test set is processed/super-resolved in the same way as the training set. SRCNN and bicubic interpolation are approximately the same type of super-resolution (bicubic-like). When super-resolving the test-set using the same method as the training set, the newer method generated higher quality images and more accurate extraction results. For example, with the model RCAN-HRNet v2, the extraction results of RFANet super-resolved images were more accurate than that of RCAN super-resolved images.



¹ Images are collected from Southwestern Ontario Orthoimagery Project (SWOOP) 2010. Produced by the Ontario Ministry of Natural Resources under License with the Ontario Ministry of Natural Resources © Queen's Printer for Ontario, 2010-2011.

² For aesthetics and clarity purpose, we enlarged the pictures in the first row of each figure.

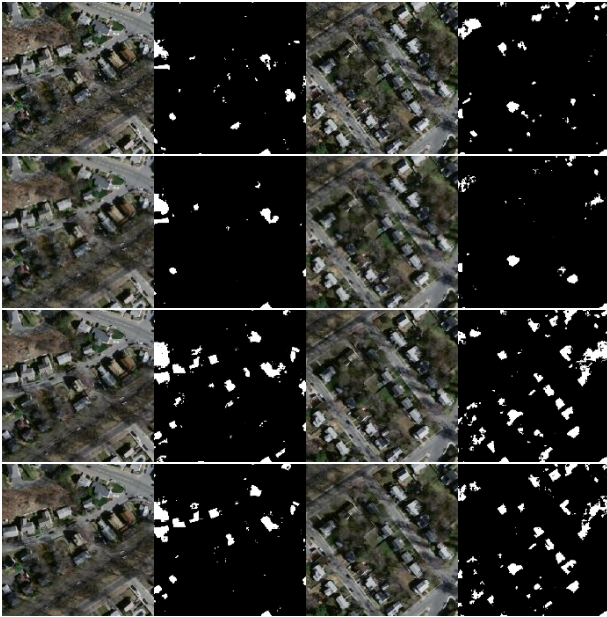


Fig. 1. Examples of the extraction results using "HRNet v2"

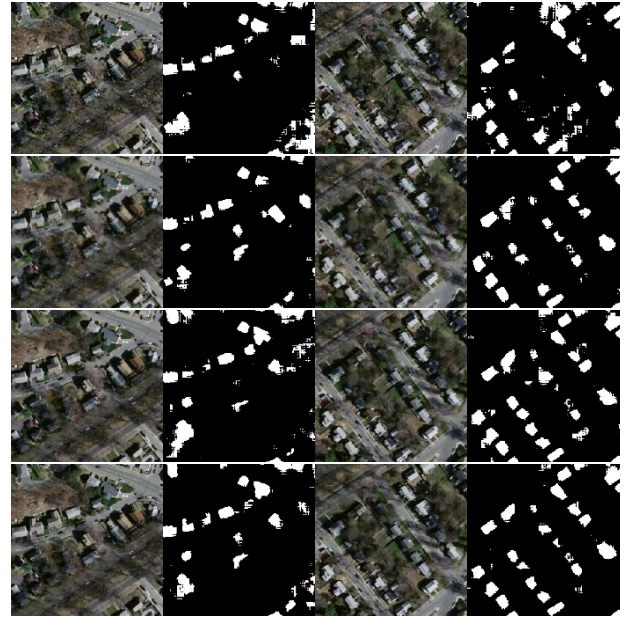


Fig. 3. Examples of the extraction results using "SRCNN-HRNet v2"

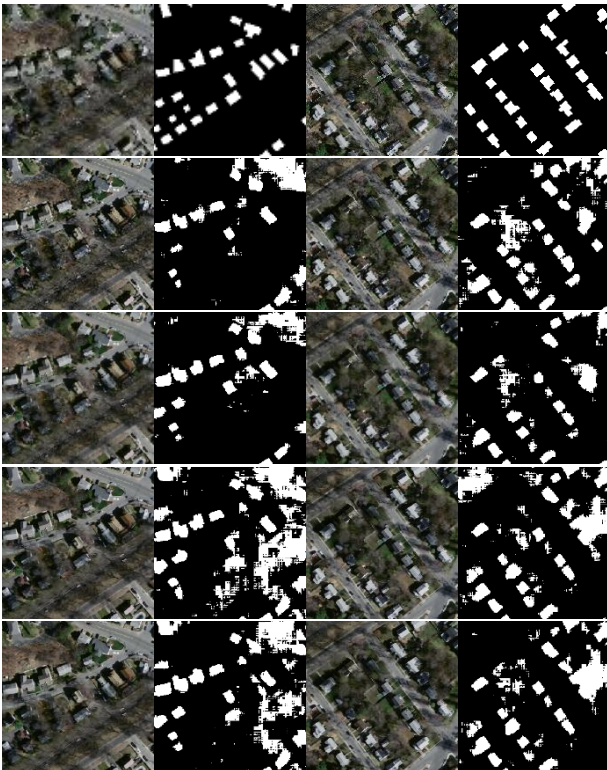


Fig. 2. Examples of the extraction results using "BI-HRNet v2"



Fig. 4. Examples of the extraction results using "RCAN-HRNet v2"



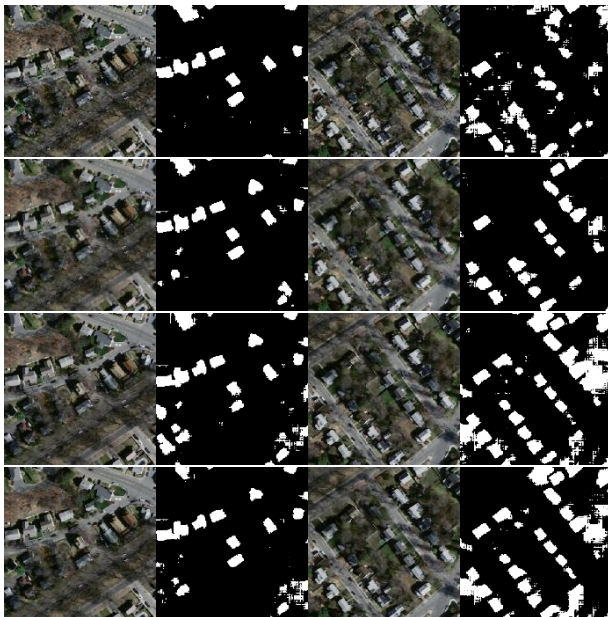


Fig. 5. Examples of the extraction results using “RFANet-HRNet v2”

Consequently, we can summarize our findings as 1) newer super-resolution methods facilitated higher performance extraction models allowing them to achieve more accurate extraction results, 2) with a given extraction model, when the training dataset and test dataset shared similar super-resolving process, more complicated super-resolution processing generated more accurate extraction results. These effects are quantified in the next section.

3.2 Quantitative evaluation

Before analyzing the impact of super-resolution on building footprint extraction, we first explored the performance of super-resolution methods we used in our experiment. We used the super-resolution methods to super-resolve the SWOOP dataset from 1 m to 0.25 m and evaluated the quality of super-resolved images using PSNR and SSIM.

Table 1. Performance of different super-resolution methods

Models	Parameters	PSNR (dB)	SSIM
Bicubic	-	29.13	0.69
SRCNN	18,497	29.59	0.70
RCAN	16,406,409	30.41	0.73
RFANet	10,692,489	30.66	0.75

Table 2. Quantitative evaluation of building footprint extraction results (in %)

Model	Test data	OA	IoU	mIoU	Precision	Recall	F ₁ score
BI-HRNet v2	Bicubic interpolation (0.3)	81.48	44.88	61.54	52.14	76.32	61.95
	SRCNN (0.3)	84.28	47.40	64.54	58.28	71.73	64.31
	RCAN (0.3)	79.43	43.56	59.56	48.75	80.38	60.69
	RFANet (0.3)	80.98	45.20	61.32	51.19	79.45	62.26
	Original (1)	52.04	24.76	33.90	25.95	84.38	39.69
SRCNN-HRNet v2	Bicubic interpolation (0.3)	82.43	44.50	62.02	54.20	71.31	61.59
	SRCNN (0.3)	86.63	50.65	67.58	65.17	69.43	67.24
	RCAN (0.3)	83.53	47.79	64.19	56.10	76.33	64.67
	RFANet (0.3)	84.16	48.40	64.90	57.59	75.21	65.23
	Original (1)	62.91	24.23	41.08	28.17	63.41	39.01
RCAN-HRNet v2	Bicubic interpolation (0.3)	83.59	41.37	61.41	58.44	58.61	58.52
	SRCNN (0.3)	86.36	44.98	64.81	68.85	56.47	62.05
	RCAN (0.3)	86.02	49.51	66.65	63.34	69.39	66.23
	RFANet (0.3)	86.47	49.89	67.12	65.00	68.22	66.57
	Original (1)	45.49	23.53	29.02	24.19	89.68	38.10
RFANet-HRNet v2	Bicubic interpolation (0.3)	81.75	42.83	60.85	52.91	69.21	59.97
	SRCNN (0.3)	85.44	47.84	65.52	62.05	67.63	64.72
	RCAN (0.3)	84.20	49.13	65.24	57.44	77.24	65.89
	RFANet (0.3)	84.57	49.49	65.66	58.34	76.54	66.21
	Original (1)	53.51	23.29	34.58	25.20	75.47	37.79
HRNet v2	Bicubic interpolation (0.3)	79.85	8.94	44.19	45.45	10.01	16.41
	SRCNN (0.3)	80.34	10.99	45.42	50.97	12.29	19.81
	RCAN (0.3)	81.62	22.70	51.63	57.29	27.32	36.99
	RFANet (0.3)	81.78	21.28	51.06	59.23	24.93	35.10
	Original (1)	87.46	47.68	66.76	68.49	61.08	64.57

As shown in Table 1, RFANet surpassed other methods in both PSNR and SIIM on the SWOOP dataset.

As shown in Table 2, all metrics’ values followed the expected trends found in the previous section. As we were exploring the impact of super-resolution on building footprint extraction. IoU and precision were good indicators. In Table 2, each model achieved its highest test scores when the test set is super-resolved using the same or similar methods as those used for the training

set. For example, extraction results of SRCNN super-resolved images had the highest IoU of 47.4% and 50.65% using extraction models BI-HRNet v2 and SRCNN-HRNet v2. Extraction results of RFANet super-resolved images had the highest IoU of 49.89% and 49.49% using extraction models RCAN-HRNet v2 and RFANet-HRNet v2. These results confirmed the qualitative findings from the previous section. Therefore, we recommend using the latest and most powerful CNNs-based super-resolution to super-resolve building datasets

and then extract building footprints from those super-resolved images. This approach will work if the test set is super-resolved using the same method. Although SRCNN showed high performance in our experiment, we do not recommend use SRCNN in practice due to requiring retraining when super-resolving images with different scales. In addition, without the skip connection included in more modern methods, the convergence speed of training would be slow.

As shown in Table 2, there were results which were hard to explain. The highest OA, mIoU and F₁ score came from the extraction results of SRCNN super-resolved test dataset using the SRCNN-HRNet v2, which is out of our expectation. We plan to investigate the causes and implications of these results in future works.

4. CONCLUSION AND RECOMMENDATION

Despite the wide use of super-resolution methods on remote sensing images, few works explored the impact of super-resolution on downstream remote sensing applications, especially semantic segmentation. Therefore, in this work, we explored the impact of single-image super-resolution on building footprint extraction. We selected pre-trained SRCNN, RCAN, RFANet and bicubic interpolation to super-resolve the Massachusetts Building Dataset. Then, we trained and tested HRNet v2 using the super-resolved images. Our experiment showed that the extraction model performs best when the training set and test set are super-resolved using the same methods. Therefore, we suggest using the latest super-resolution method to process building dataset during both training and testing/deployment to improve the accuracy of building footprint extraction.

REFERENCES

Cai, J. and Chen, Y., 2021. MHA-Net: Multipath Hybrid Attention Network for building footprint extraction from high-resolution remote sensing imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, pp.5807-5817.

Dong, C., Loy, C.C., He, K. and Tang, X., 2015. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2), pp.295-307.

Ducourneau, A. and Fablet, R., 2016, December. Deep learning for ocean remote sensing: an application of convolutional neural networks for super-resolution on satellite-derived SST data. In 2016 9th IAPR Workshop on Pattern Recognition in Remote Sensing (PRRS), pp. 1-6.

Guo, H., Du, B., Zhang, L. and Su, X., 2022. A coarse-to-fine boundary refinement network for building footprint extraction from remote sensing imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 183, pp.240-252.

Haut, J.M., Fernandez-Beltran, R., Paoletti, M.E., Plaza, J., Plaza, A. and Pla, F., 2018. A new deep generative network for unsupervised remote sensing single-image super-resolution. *IEEE Transactions on Geoscience and Remote sensing*, 56(11), pp.6792-6810.

He, H., Jiang, Z., Gao, K., Narges Fatholahi, S., Tan, W., Hu, B., Xu, H., Chapman, M.A. and Li, J., 2022. Waterloo building dataset: a city-scale vector building dataset for mapping building footprints using aerial orthoimagery. *Geomatica*, pp.1-17.

He, H., Yang, K., Wang, S., Petrosians, H.A., Liu, M., Li, J., Marcato Junior, J., Gonçalves, W.N., Wang, L. and Li, J., 2021. Deep Learning Approaches to Spatial Downscaling of GRACE Terrestrial Water Storage Products Using EALCO Model Over Canada. *Canadian Journal of Remote Sensing*, 47(4), pp.657-675.

Lei, S., Shi, Z. and Zou, Z., 2017. Super-resolution for remote sensing images via local-global combined network. *IEEE Geoscience and Remote Sensing Letters*, 14(8), pp.1243-1247.

Lei, S. and Shi, Z., 2021. Hybrid-scale self-similarity exploitation for remote sensing image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 60, pp.1-10.

Li, Z., Wegner, J.D. and Lucchi, A., 2019. Topological map extraction from overhead images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1715-1724.

Liu, H., Luo, J., Huang, B., Hu, X., Sun, Y., Yang, Y., Xu, N. and Zhou, N., 2019. DE-Net: Deep Encoding Network for Building Extraction from High-Resolution Remote Sensing Imagery. *Remote Sensing*, 11(20), p.2380.

Liu, J., Zhang, W., Tang, Y., Tang, J. and Wu, G., 2020. Residual Feature Aggregation Network for Image Super-Resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2359-2368.

Ma, W., Pan, Z., Guo, J. and Lei, B., 2019. Achieving super-resolution remote sensing images via the wavelet transform combined with the recursive res-net. *IEEE Transactions on Geoscience and Remote Sensing*, 57(6), pp.3512-3527.

Ma, W., Pan, Z., Guo, J. and Lei, B., 2018, July. Super-resolution of remote sensing images based on transferred generative adversarial network. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pp. 1148-1151. IEEE.

Marcos, D., Tuia, D., Kellenberger, B., Zhang, L., Bai, M., Liao, R. and Urtasun, R., 2018. Learning deep structured active contours end-to-end. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8877-8885.

Mnih, V. 2013. Machine learning for aerial image labeling. PhD Thesis, University of Toronto.

Rastogi, K., Bodani, P. and Sharma, S.A., 2020. Automatic building footprint extraction from very high-resolution imagery using deep learning techniques. *Geocarto International*, pp.1-13.

Shermeyer, J. and Van Etten, A., 2019, June. The Effects of Super-Resolution on Object Detection Performance in Satellite Imagery. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1432-1441.

- Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D. and Wang, Z., 2016. 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, pp. 1874-1883.
- Sun, K., Zhao, Y., Jiang, B., Cheng, T., Xiao, B., Liu, D., Mu, Y., Wang, X., Liu, W. and Wang, J., 2019. High-resolution representations for labeling pixels and regions. arXiv preprint arXiv:1904.04514.
- Tai, Y., Yang, J. and Liu, X., 2017. Image super-resolution via deep recursive residual network. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3147-3155.
- Thomas, J., Kareem, A., and Bowyer, K.W. 2013. Automated post-storm damage classification of low-rise building roofing systems using high-resolution aerial imagery. *IEEE Transactions on Geoscience and Remote Sensing*. 52(7): pp. 3851-3861. doi:10.1109/TGRS.2013.2277092.
- Tuna, C., Unal, G. and Sertel, E., 2018. Single-frame super resolution of remote-sensing images by convolutional neural networks. *International journal of remote sensing*, 39(8), pp.2463-2479.
- Wang, P., Zhang, H., Zhou, F. and Jiang, Z., 2019, July. Unsupervised remote sensing image super-resolution using cycle CNN. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, pp. 3117-3120.
- Xu, Y., Wu, L., Xie, Z. and Chen, Z., 2018. Building extraction in very high resolution remote sensing imagery using deep learning and guided filters. *Remote Sensing*, 10(1), p.144.
- Yu, Y., Ren, Y., Guan, H., Li, D., Yu, C., Jin, S. and Wang, L., 2020. Capsule feature pyramid network for building footprint extraction from high-resolution aerial imagery. *IEEE Geoscience and Remote Sensing Letters*, 18(5), pp.895-899.
- Yuan, J., 2017. Learning building extraction in aerial scenes with convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 40(11), pp.2793-2798.
- Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B. and Fu, Y., 2018. Image super-resolution using very deep residual channel attention networks. In Proceedings of the European conference on computer vision (ECCV), pp. 286-301.
- Zhao, K., Kang, J., Jung, J. and Sohn, G., 2018, June. Building Extraction From Satellite Images Using Mask R-CNN With Building Boundary Regularization. In *CVPR Workshops*, pp. 247-251.
- Zhao, W., Persello, C. and Stein, A., 2021. Building outline delineation: From aerial images to polygons with an improved end-to-end learning framework. *ISPRS Journal of Photogrammetry and Remote Sensing*, 175, pp.119-131.
- Zhu, X.X., Tuia, D., Mou, L., Xia, G.S., Zhang, L., Xu, F. and Fraundorfer, F., 2017. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4), pp.8-36.