Residual Attention Mechanism for Remote Sensing Target Hiding

Hao Yuan *, Yongjian Shen, Ning Lv, Yuheng Li, Chen Chen *, Zhouzhou Zhang

Posted Date: 27 July 2023

doi: 10.20944/preprints202307.1811.v1

Keywords: remote sensing mapping; image inpainting; residual attention mechanism; target hiding

Article

Residual Attention Mechanism for Remote Sensing Target Hiding

Hao Yuan [1,*], Yongjian Shen [2], Ning Lv [3], Yuheng Li [1], Chen Chen [3,4,*] and Zhouzhou Zhang [1]

[1] Beijing Research Institute of Telemetry, Beijing, 100076, China
[2] Beijing University of Aeronautics and Astronautics, Beijing, 100191, China
[3] Xidian University, Xi'an 710071, China
[4] Xidian Guangzhou Institute of Technology, Guangzhou 510555, China.
* Correspondence: Hao Yuan and Chen Chen

Abstract: Remote sensing imagery is of great significance for policy decisions, especially for disaster assessment and disaster relief. To ensure the privacy and inviolability of personal buildings, the information containing these buildings must be anonymized during the remote sensing mapping process. Traditional processing methods for these targets in remote sensing mapping are mainly based on manual retrieval and image editing tools, which are inefficient. Deep learning provides a new direction for target hiding. Although the image inpainting method based on deep learning is faster than the manual method, the cost of training calculation is a disadvantage. And the element-wise product operation used in the model increases the risk of vanished or exploded gradients. We propose a Residual Attention Target Hiding (RATH) model for remote sensing target hiding based on deep learning. RATH uses residual attention modules to replace gated convolutions, reducing parameters and mitigating gradient issues. The residual attention module preserves gated convolution performance but provides an adjustable kernel size. RATH retains gated convolutions for dynamic feature selection and balances model depth and width. Furthermore, this paper modifies the contextual attention layer by adjusting the fusion process to enlarge the fusion patch size. Finally, we extend the edge-guided function to preserve the original target information and confound viewers. Ablation studies on an open dataset prove RATH's efficiency for image inpainting and target hiding. RATH achieves state-of-the-art results with lower complexity. And it has the highest similarity for edge-guided target hiding. RATH enables robust, efficient target hiding for privacy protection in remote sensing imagery while balancing performance and complexity. Experiments show RATH's superiority over existing methods in hiding arbitrary-shaped targets.

Keywords: remote sensing mapping, image inpainting, residual attention mechanism, target hiding

1. Introduction

Remote sensing images contain abundant surface features which can support governments and rescue agencies in emergency decision-making, disaster assessment, and rescue deployment[1]. There are many research and application directions of remote sensing mapping for urban development planning and emergency disaster response. Building detection[2] and construction disturbance detection rely on remote sensing interpretation. Generative Adversarial Networks(GANs) supplement labeled data due to inefficient manual labeling. These methods solve issues like low timeliness and unstable effects of traditional artificial remote sensing image processing. While semantic segmentation [5] greatly improves remote sensing image interpretation speed.

To protect privacy, especially of personal buildings, sensitive targets must be processed before public release and use. Current methods mainly rely on manual or semi-automatic labeling of sensitive targets and image editing tools to cover and fill target areas. These methods cannot meet the timeliness requirements of mapping tasks. Moreover, the results depend on operators' skills and thus lack control. Therefore, an automatic method for sensitive target hiding is needed.

As deep learning has developed, much research based on remote sensing data has taken a new direction. To automatically capture the position and contour of sensitive targets, semantic segmentation

 doi:10.20944/preprints202307.1811.v1

was introduced[6]. Its process is the same as that of remote sensing interpretation tasks[39]. After detecting sensitive targets, Qiu et al. proposed an image inpainting model to remove and fill target areas[6]. This method combines object detection with image inpainting to achieve fast, automatic hiding of sensitive targets. The image inpainting model Cont Atten(CA)[7], used in the combined method, aims to hide sensitive targets. However, the model is limited by using regular masks in the training step, which depends on the missing area of the image. For this reason, a new model based on Gated Conv[8] training with free-form masks was proposed. Yu et al. used masks combining irregular and regular masks[8] to reduce the computation of hard-gating masks proposed by Partial Conv[9]. Gated convolution is another innovation of their research. It provides a learnable, dynamic feature selection mechanism at both channel and spatial location levels. However, the extensive use of element-wise products causes unstable gradients. The method of generating two branches with one convolution operation also has the problem of too many parameters.

Therefore, this paper proposes a Residual Attention Target Hiding(RATH) model based on the residual attention mechanism and tuning of the contextual attention module. The model is used to hide targets in emergency remote sensing maps. By bifurcating the generation process into two branches using two concatenated convolutions, our model employs an adjustable kernel size in gated convolutions, endowing greater flexibility. The residual attention mechanism obviates issues of gradient vanishing and explosion. Consequently, our model can be trained extensively without overfitting and achieves enhanced performance. In addition, this paper adjusts the fusion process in the contextual attention module to simplify the operation. We choose a matrix with all elements equal to 1 as the convolution kernel to replace the complex operation containing two transpose operations and two convolutions with an identity matrix as kernels. Finally, this paper extends the edge-guided[3] function to synthesize fabricated targets with a more realistic distribution. We perform ablation experiments using Partial Conv[9], Cont Atten[7], Gated Conv[8], and our proposed RATH model on datasets adapted from the Mnih Massachusetts Building Dataset. Results verify the superiority of our method in remote sensing image inpainting, target hiding, and edge-guided target hiding. The proposed method has better evaluation in SSIM and UQI. It reduces about 1MB in size and speeds up training by 0.045s per batch compared to GatedConv. Our method also has the highest similarity in the edge-guided[4] target hiding task.

The remainder of this paper is organized as follows. Section 1 introduces the development of target hiding in emergency remote sensing mapping and the contributions of our research. Section 2reviews advancements in image inpainting using deep learning which the target hiding is based on. Section 3 elaborates on the principal framework and methodology of the proposed approach. Section 4 presents extensive experiments on diverse datasets to evaluate image inpainting, target hiding, and edge-guided inpainting capabilities. Section 5 introduces our automated application for target hiding utilizing the proposed techniques. Finally, Section 6 concludes the paper and discusses directions for future work.

2. Relate Work

Image inpainting refers to the process of reconstructing lost or damaged parts of images and videos. Early image inpainting methods are relatively simple. Nitzberg et al. proposed an algorithm using image segmentation to remove the objects in front of the foreground[10]. Kokaram et al. used motion estimation and an autoregressive model to fill defects in adjacent frames[11]. The combined frequency with location information, Hirani and Totsuka select a similar texture to fill the target areas[12]. This simple technology produced incredibly good results at that time. But this technology is only responsible for analyzing image texture. Whether the texture is used or not depends on the users. And the target area needs to be segmented by users, which is complex and time-consuming. In 1998, an algorithm based on Nitzberg's was proposed by Masnou and Morel[13]. The main idea of it is to perform the repair by connecting the points of equal rays (lines with equal gray values) that reach the boundary of the area to be repaired, while the area must have a simple topology. Then, a new

 doi:10.20944/preprints202307.1811.v1

static image restoration algorithm was introduced by Colomba Ballister and Marcelo Bertalmio[14]. After the user selects the areas to be restored, the algorithm will automatically fill these areas with the information around them, which has achieved considerable success.

With the development of deep learning, research in the image field has made breakthroughs. GAN[15–18] is the basic structure of image generation which affect the image inpainting, though there have other structures like variational autoencoder(VAE)[19–21] and diffusion model[22–24]. Image generation refers to generating new images from existing datasets. And there are two types of image generation models: unconditional generation and conditional generation(CGAN). GAN is the classical structure of unconditional generation, and conditional generation[25–28] generates new images with condition limited.

For target hiding, image inpainting needs to fill the missing area after the target is removed. In 2016, the first image inpainting model Context Encoder(CE)[29] based on GAN was proposed. The core idea is the channel-wise fully-connected layer, which is similar to the standard fully-connected layer, but each channel handles its characteristics separately. Next, Multi-Scale Neural Patch Synthesis(MSNPS)[30] was regarded as the enhanced CE. They introduced local texture loss to ensure that the fine details of the missing area are similar to other parts. Then another classical model of image inpainting is Globally and Locally Consistent Image Completion(GLCIC)[31]. It used dilated convolution to instead the fully-connected layer for a larger receptive field, then the global discriminator and local discriminator were introduced in the training process. PGGAN[32] embedded residual mechanism and PatchGAN[27] in GLCIC to enhance the performance. Different from the discriminator of GAN, the output of the PatchGAN discriminator is the matrix of prediction labels. Shift-Net[33] introduced the shift-connection layer to U-Net for filling in missing regions of any shape with sharp structures and fine-detailed textures. In 2018, Contextual Attention(CA)[7] calculates the contribution of all outside features to each location in the missing region by matching the generated features and the outside features. In the same year, Partial Convolutions(PartialConv)[9] used partial convolution to predict the area, and the prediction does not depend on the initial value of the hole. And it is the first model training in irregular masks. The results prove the effectiveness of the irregular mask training strategy.

Recently, a model repairing images with adversarial edge learning EdgeConnect[34] was proposed. It divided the image inpainting task into two steps: edge prediction and image inpainting based on edge. The second step is similar to CGAN. Then GatedConv[8] follows the idea of EdgeConnect. The authors of GatedConv extend the image inpainting to user-guided inpainting. By providing a sketch, the model will generate an image that has the same edge as the sketch. They also proposed a gated convolution to replace the convolution operation to learn the effectiveness of each feature in each location. It provides a new direction for target hiding in emergency remote sensing maps.

3. Method and Materials

3.1. Mainframework

While target hiding and image inpainting are similar in nature, they produce distinct outputs. Image inpainting imposes no constraints on the modality of the generated image except for textural consistency. In contrast, target hiding treats the object as foreground and the rest as background. Its objective is to synthesize new images containing only the background. Therefore, training a target hiding model primarily involves learning the textural characteristics of the background. In practice, however, the core capability of target hiding is filling missing regions with surrounding pixels. As such, when the missing area in image inpainting corresponds to the target region in target hiding, the two methods share the same goal.

Compared to Partial Conv[9] and Gated Conv[8] (where Gated Conv builds upon CA[7]), this paper augments Gated Conv by incorporating the proposed residual attention layer and novel contextual attention layer. Additionally, the training methodology for the edge-guided functionality

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 July 2023 doi:10.20944/preprints202307.1811.v1

deviates from CA. This paper also revisits the PatchGAN[27] discriminator and SN-PatchGAN loss[8] from PatchGAN. The architecture of the proposed Residual Attention Target Hiding(RATH) model is summarized in Figure 1.

The authors of Gated Conv argued that an encoder-decoder architecture is better suited for image inpainting compared to the U-Net[35] employed in Partial Conv, especially when masks are centrally located. Although the skip connections in U-Net provide shallow features to deep features, image inpainting primarily involves filling missing regions using surrounding pixels. Therefore, a pure encoder-decoder architecture is more appropriate for this task. There is a skip connection layer in the proposed model to incorporate the original image information. The mask region in the coarse network output is retained while the remaining region is replaced with the original image. The overall framework comprises two encoder-decoder networks representing the progression from coarse to refined results. The discriminator in RATH is patch-based. The outputs are prediction matrices where each element indicates the probability of the patch being "real" or "fake".

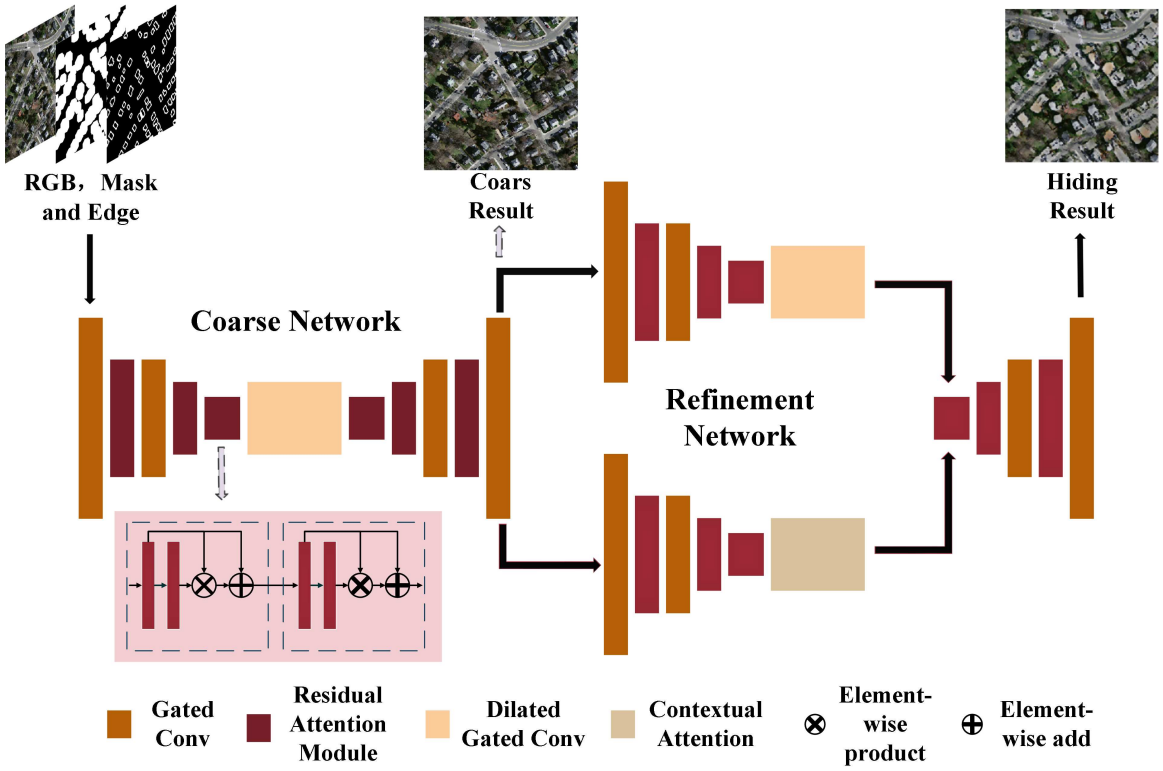


Figure 1. The structure of generator. The model is a two-stage network. Refinement Network refines the preliminary repair results of the Coarse Network.

3.2. Methodology

3.2.1. The Proposed Residual Attention Moudle

Convolutional modules are critical to the overall performance of the model. Compared to Partial Conv [9] and Gated Conv [8], gated convolution offers the following advantages: (1) The parameters in gated convolution are meaningful and trainable. Partial Conv heuristically classifies all spatial locations as either valid or invalid. This means all spatial locations within the mask region are assigned weights by the partial mechanism, and these weights cannot be trained. (2) Gated convolution is more flexible since the effectiveness of pixels can be learned during training without being limited to the weight set specifically designed for image restoration. However, we observed substantially prolonged training time and significant instability in the loss function during our experiments applying Gated Conv for image inpainting. Gated convolution achieves two branches with the same channel number

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 July 2023 doi:10.20944/preprints202307.1811.v1

through a single convolution operation. One branch acts as a control gate for the other branch. The parameters in both the "gate" and those it controls are fully trainable since they are updated during training. Let I denote the input to the module. This operation can be formulated as

$$\begin{aligned} H_G &= \sigma(Gating(I)) \odot \phi(Feature(I)) \\ &= \sigma(\sum\sum W_g \cdot I) \odot \phi(\sum\sum W_f \cdot I) \end{aligned} \tag{1}$$

, where H_G represents the output of gated convolution, $Gating(I)$ represents the "gate" and $Feature(I)$ represents the features waiting to be selected by $Gating(I)$. The parameters W_g and W_f have the same value and denote the kernels in gated convolution. The function ϕ can be any activation function, while σ is confined to the sigmoid function to limit the output within $(0, 1)$. The core of gated convolution lies in the element-wise product operation which consumes substantial computational resources. At the same time, the complexity of the model increases with its depth, leading to overfitting when trained for an extended period. Therefore, we propose a residual attention module for the image inpainting network. An illustration of our module and other alternatives is provided in Figure 2.

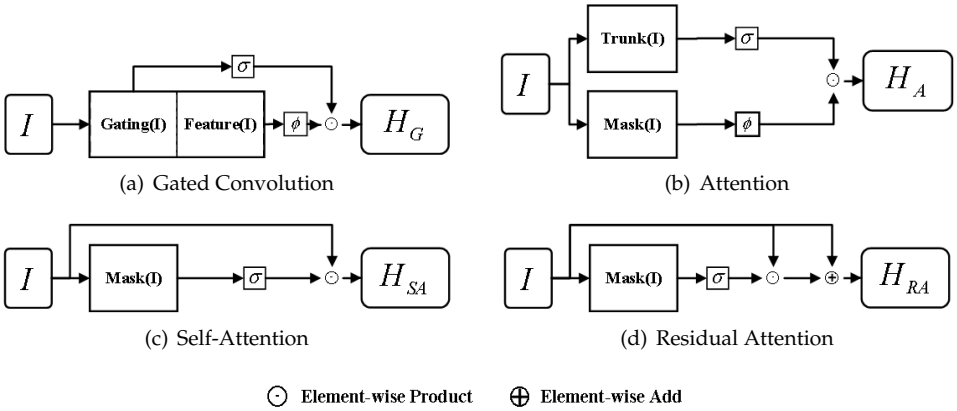


Figure 2. Illustration of four modules. The gated convolution could be regarded as a special attetnion mechanism.

The "gate" mechanism is proposed early in attention mechanism[36]. The output of attention module H_A can be formulated as

$$H_A = \sigma(Mask(I)) \odot \phi(Trunk(I)) \tag{2}$$

, where $Mask(I)$ represents the gate controlled the output of $Trunk(I)$. In the application of attention modules, the $Mask(I)$ and $Trunk(I)$ have various structures. Although gated convolution and attention modules have the same expression, their implementation methods are different. The $Mask(I)$ and $Trunk(I)$ could be obtained with the same convolutional kernels in gated convolution. Inspired by it, we replaced $Trunk(I)$ in the attention mechanism, or $Feature(I)$ in gated convolution, with the inputs, which is called the self-attention mechanism. It was formulated as:

$$H_{SA} = \sigma(Mask(I)) \odot I \tag{3}$$

As shown in Figure 2, the inputs are connected directly to the element-wise product. In this paper, we insert one convolutional layer before the attention modules to extract input features during the forward pass of training. The kernel size of the self-attention modules is 1 to obtain channel-wise features. This significantly decreases the overall number of model parameters and accelerates training.

However, the gradient transfer problem still exists. Let θ denote the parameters of the mask branch and ϕ denote the parameters used in the previous layer to extract information I. The gradient of the self-attention module can be computed as:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 July 2023 doi:10.20944/preprints202307.1811.v1

$$\frac{\partial H_{SA}(I,\theta,\phi)}{\partial \phi} = \sigma(Mask(I,\theta))\frac{\partial I(\phi)}{\partial \phi} \tag{4}$$

Due to the sigmoid function, the values in $\sigma(Mask(I,\theta))$ always stay belong to $(0,1)$. Then the gradient value has the risk of developing towards near zero, which is also called the gradient disappearance. Therefore, this paper introduced a residual mechanism to mitigate the problem of gradient disappearance. The output of the residual attention mechanism H_{RA} could be calculated by formula 5. And the calculation of the gradient of H_{RA} followed the formula 6.

$$H_{RA} = \sigma(Mask(I)) \odot I + I = (1 + \sigma(Mask(I))) \odot I \tag{5}$$

$$\frac{\partial H_{RA}(I,\theta,\phi)}{\partial \phi} = (1 + \sigma(Mask(I,\theta)))\frac{\partial I(\phi)}{\partial \phi} \tag{6}$$

Given that 1 exists in the gradient, the weights of ∂I remain confined between $(1,2)$. To mitigate the risk of gradient explosion, we apply rectified linear unit (ReLU) activation functions subsequent to each convolution operation. Additionally, we insert one convolutional layer prior to the residual attention module. The kernel size of the residual attention module is set to 1 to extract channel-wise features. As depicted in Figure 2, if n denotes the number of channels of $Trunk(I)$, then $Gating(I)$, $Feature(I)$ and $Mask(I)$ share the same value. However, by factorizing the convolution operation in gated convolution into two successive steps, the residual attention module affords greater flexibility. The total number of parameters can be modulated contingent upon available computational resources. The element-wise product enables learning of a dynamic feature selection mechanism for each channel and spatial location, conferring the same benefits as gated convolution. For a substantial quantity of convolution kernels, the self-attention mechanism and gated convolution yield nearly equivalent outcomes.

In this work, we substitute the intermediate layers with residual attention modules. Due to the more parameters, gated convolution exhibits superior performance in extracting global and local information. And the gated mechanism can support soft mask training strategy in the initial stage of network. Consequently, we retain gated convolution in the upsampling layer, downsampling layer, and initial and final convolution layers of the coarse and refinement networks. This strategy aims to balance the computational cost across layers, thereby reducing the number of weight parameters and simplifying the model. Relative to Gated Conv, our model contains 1 MB fewer parameters and achieves improved performance.

3.2.2. Contextual Attention Layer

The primary innovation of Cont Atten [7] resides in the contextual attention layer. To synthesize more photorealistic images, contextual attention extracts image patches from the foreground and background to serve as convolution kernels. This process is illustrated in Figure 3.

The contextual attention layer departs from the conventional approach of employing direct convolution and element-wise multiplication in the attention module. To mitigate the substantial computational burden, this work resizes the input dimensionality by half while maintaining the original input shape for the output. Two image extraction operations are performed on the input, the first to obtain the original information and the second to acquire information subsequent to downsampling. The extraction mechanism can preserve the native input values. With these primordial input values serving as kernels, the convolution operation can be regarded as a self-attention mechanism. However, the contextual attention layer employs the group convolution approach. It partitions kernels and inputs into groups, performing convolution within each group. The outputs are then concatenated channel-wise. In this layer, the group unit is 1 batch, indicating that a tensor of shape $(batch, height, width, channel)$ is partitioned into $batch$ tensors of shape $(1, height, width, channel)$. These tensors undergo convolution with information extracted from themselves. This work normalizes

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 July 2023 doi:10.20944/preprints202307.1811.v1

the outputs of the downsampling operation. In the mask branch, the dimensionality of extracted matrices is reduced to one through averaging. The shape of features H_{SC} subsequent to self-convolution is $(1, height, width, height * width)$. The fuse module then conducts two transform operations and two convolutions with a unit matrix of size 3.

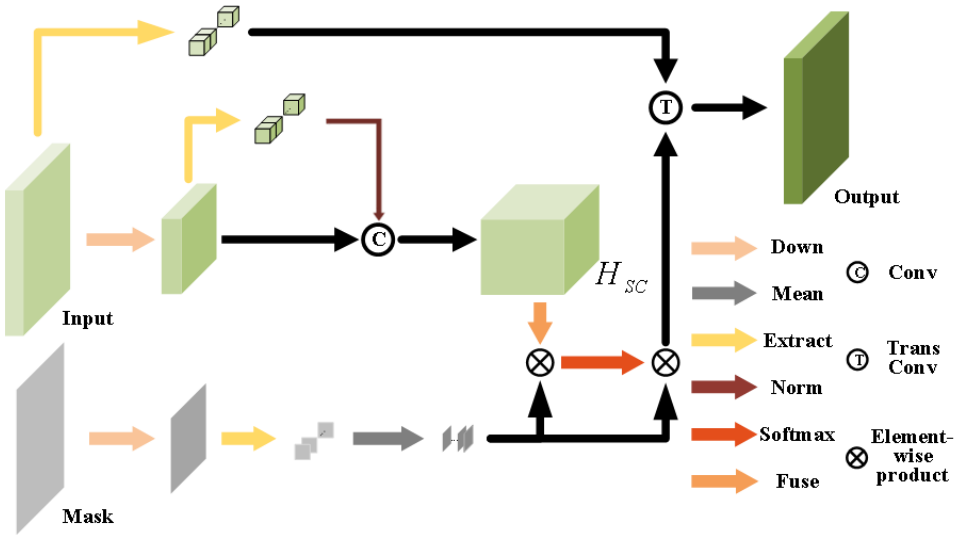


Figure 3. The structure of Contextual Attention layer. To mitigate the substantial computational burden, this layer resizes the input dimensionality by half while maintaining the original input shape for the output.

In this layer, the product operation is utilized to obtain incomplete information within the masked area. The probability is calculated using the softmax function, which requires two product operations for its computation. The initial product operation preceding the softmax function serves to extract features within a designated mask area, while the subsequent operation following the softmax function is intended to suppress probabilities outside the same mask area. Upon completion of this second multiplication process, the extracted features are utilized as transformation convolution kernels in order to realign the input matrix to its original shape. The resulting output is then merged with the other branch in a channel-wise fashion.

The contextual attention layer serves as the core component of both the Cont Atten [7] and our Refinement Network. To enhance the network's optimization, we have replaced the Fuse module with two convolutional operations that employ unit matrices consisting entirely of 1s. The original function of the fuse module was to create pixel patches via convolution with a unit matrix. By adopting this approach, we can generate larger patches using 3-dimensional matrices filled with 1s, which can act as an effective substitution for the Fuse module.

3.2.3. Free-Form Mask

Prior to Partial Conv [9], the conventional method for image inpainting involved the use of a rectangular mask at the center. However, this approach lacked flexibility and controllability, as the masks were not considered a pivotal component of the overall model. Despite attempts at introducing random rotations, dilations, and cropping, the resulting irregular masks remained mere transformations of the original mask, with limited effectiveness in image inpainting. Furthermore, when applied to target hiding, the unpredictability and uncontrollable nature of the target shapes posed a significant challenge. Therefore, it became necessary to devise a more sophisticated algorithm that could effectively address these limitations.

To address the limitations of the previous approaches, this paper proposes a novel algorithm that generates randomized masks with irregular shapes during training. The shape of the mask can cover various forms such as lines, circles, and rectangles, and by limiting the available range, the algorithm

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 July 2023 doi:10.20944/preprints202307.1811.v1

randomly applies these graphics onto a zero board of the same size as the original images. The shape of the resulting mask is illustrated in Figure 4. By introducing unique masks for each training instance, overfitting is avoided, and the random irregular masks significantly improve the model's ability to handle target shapes with non-conventional geometries. Experimental results demonstrate the efficacy of the proposed free-form mask training strategy for image inpainting.

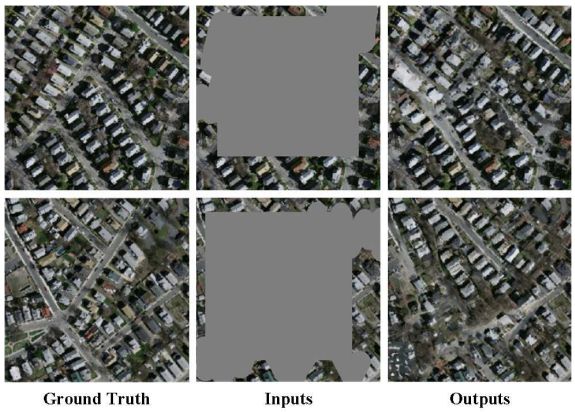


Figure 4. The examples in the training process. The inputs are images with masks covered, and the outputs are results repaired by models.

3.2.4. Edge-guided Target Hiding

The proposed model also includes edge-guided target hiding as one of its functions. In previous research, Yu et al. [8] introduced sketches or edges into the image inpainting model using gated convolutions, which guides the model in repairing the image based on the edges. In this paper, the sketch mainly represents the edge of the target, serving as the conditional label in CGAN. However, traditional edge extraction algorithms often perform poorly on remote sensing images. These images contain more refined details compared to ordinary images[40], making it difficult for traditional methods which rely on contour continuity and gradient changes. Targets within remote sensing images often have various colors with large gradients in both value and channel differences, causing edge extraction algorithms based on these factors to lose their effectiveness. To address this issue, Figure 5 demonstrates a comparison between edges extracted from Canny operator and the proposed method.

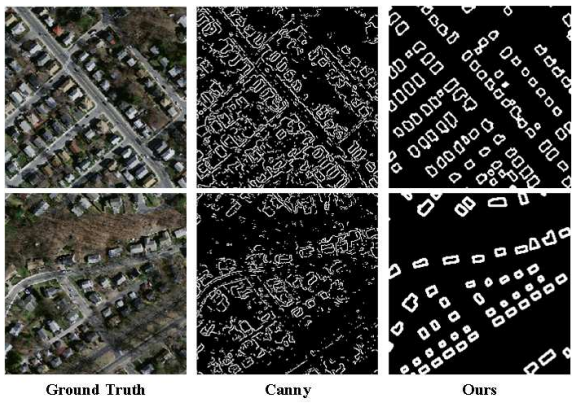


Figure 5. Comparision of edge extraction methods. The edge extraction results using traditional Canny operator exhibit many noise points, while the edges extracted by proposed method are significantly clearer and more accurate.

Therefore, we leverage image dilation and erosion techniques to accurately extract building edges from the results of semantic segmentation. This methodology is notably more straightforward and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 July 2023 doi:10.20944/preprints202307.1811.v1

precise compared to using either the HED edge detector [37] or the Canny operator. Moreover, by contouring these edges with a semantic segmentation network, we enrich the adaptability of our image inpainting approach towards automatic target hiding. As illustrated in Figure 5, the edges produced by the Canny operator suffer from many noise points that cannot be utilized for training an image inpainting network, while our method is much more suitable. The edge channel is incorporated into the discriminator's loss calculation. Ultimately, the edge-guided[41] image inpainting model requires three inputs of equal shape (image, mask, and edge channel).

3.3. Materials

All ablation experiments were performed on a 64-bit Linux system that was equipped with a single NVIDIA GeForce RTX 2080 Ti graphics card and 12 GB of memory. All models were trained using Tensorflow v1.14. We utilized the Adam optimizer in both the generator and discriminator, with a learning rate of $1 \times e^{-4}$. The batch size for all experiments was set to 16, and the iterations were limited to 10^8. We saved the model at every 2000 iterations. The loss function used for our experiments remained unchanged and was the same as Gated Conv.

4. Experiment and Result

The Mnih Massachusetts Building Dataset [38], consisting of 151 aerial images with a size of 1500×1500, was chosen for evaluation in this paper. The main foreground in the dataset is buildings, and the images were cut into 256×256 for analysis purposes. Two different input types were utilized for the image inpainting and target hiding experiments in this paper. In the case of image inpainting, the focus was on repairing random missing regions across the input image. Alternatively, for target hiding, repairs were made specifically to regions in the input image where targets were located. Additionally, the paper sought to extend its work to include edge-guided target hiding tasks, which required a different training process than that used for image inpainting and traditional target hiding tasks. The datasets used in following experiments contain 4464 pictures, with 3960 pictures for the training set, 144 pictures for the validation set, and 360 pictures for the test set.

4.1. Experimental Comparison for the Image Inpainting Task

The main objective of this experiment was to evaluate the performance of various models in repairing images. Specifically, the free-form mask strategy was applied only to the Gated Conv with attention mechanism and residual attention mechanism, while the other models employed their own respective training strategies. In both the training and testing phases, each sample consisted of a single image and an irregular mask. During testing, researchers drew random masks to assess the proposed strategy's ability to repair irregular regions. Training speed and weight parameters for each model are presented in Table 1. Notably, the self-attention network and residual attention network had reduced parameters by 1 MB and exhibited faster training speeds compared to other models.

Table 1. Comparison of two image inpainting methods for calculation cost.

Methods	Gated Conv	Self-Atten	Res Atten(Ours)
Parameters	9M548K958B	**8M400K414B**	**8M400K414B**
Training speed (sec/batch)	0.705	**0.66**	**0.66**

The results of various inpainting models are presented in Figure 6. All models demonstrate the ability to repair missing image regions, but their performance varies. Notably, the distribution of pixels around the boundary of the missing region remains inconsistent, particularly in the case of Partial Conv. Additionally, while the objects in the generated images appear valid, their contours are imperfect. Small objects such as houses are adequately reconstructed by all models, but larger

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 July 2023 doi:10.20944/preprints202307.1811.v1

objects suffer from varying levels of distortion. In particular, the buildings repaired by Partial Convo and Cont Atten exhibit suboptimal performance. Another significant difference among the various models is their ability to learn object relationships, notably the correlation between buildings and roads. Typically, there exists a well-established relationship between these two objects, whereby buildings are situated along the edges of roads on either the left or right side. Our proposed models exhibit consistent and remarkable performance in generating accurately positioned roads flanked by buildings that conform to this known relationship. In terms of overall performance, our models demonstrate superiority compared to the other approaches evaluated in this study.

Figure 6. Example cases of qualitative comparison on image inpainting. The images from left to right represent the inputs, original images, and the result of PartialConv, Cont Atten, Gated Conv, and ours.

This paper also assesses the performance of image restoration using similarity indicators, which are presented in Table 2. Our proposed method exhibits the smallest values on ℓ_1Sim and ℓ_2Sim indices, indicating that images generated by the residual attention mechanism have the highest similarity in pixel distribution with the original images. Furthermore, our method achieves higher values on both PSNR and UQI, verifying its efficacy for image restoration tasks.

Table 2. Evaluation index of models on image repairing.

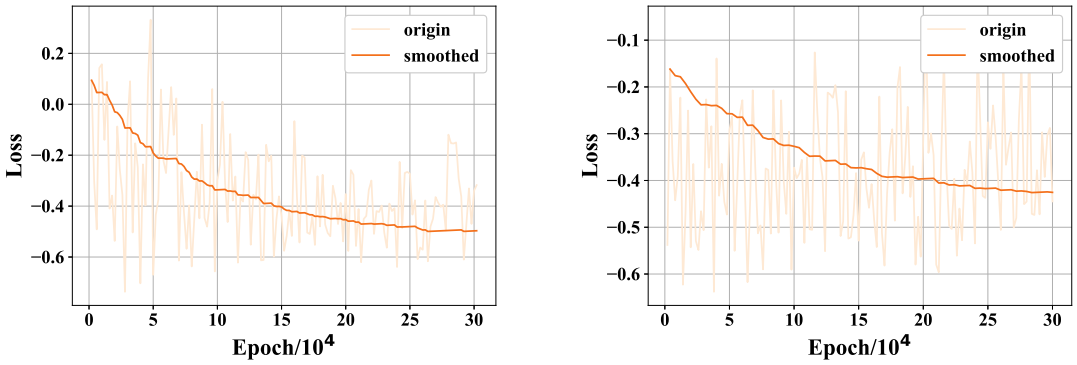
Methods	Cont Atten	Partial Conv	Gated Conv	RATH (Ours)
ℓ_1 **Sim/(%)**	98.47	98.54	98.59	**98.61**
ℓ_2 **Sim/(%)**	87.98	88.30	88.50	**88.62**
PSNR	18.81	19.10	19.29	**19.43**
SSIM/(%)	91.86	**92.04**	81.49	91.72
UQI/(%)	90.98	91.38	91.62	**91.70**

In Figure 7 and 8, we compare the loss curves of the generator and discriminator with those of Gated Conv.The $x-$axis represents the number of epochs, measured in units of 10^4, and the $y-$axis corresponds to the loss value. We apply a smoothing technique to the loss curve in order to present

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 July 2023 doi:10.20944/preprints202307.1811.v1

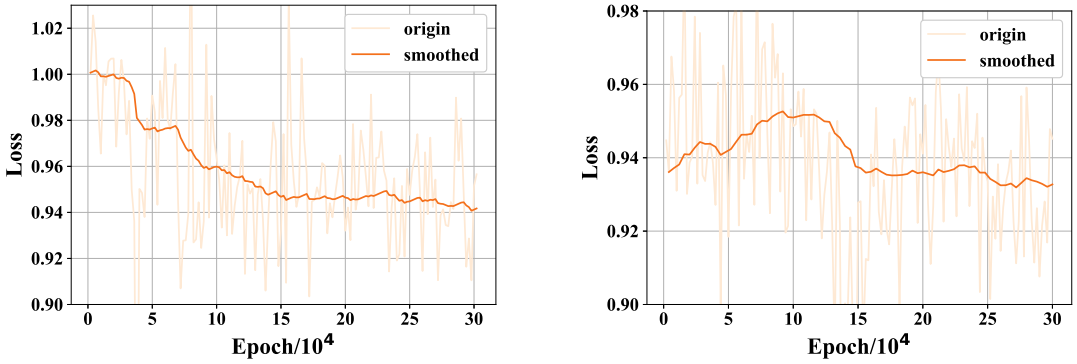
a more consistent representation of the training trend. The initial loss curve exhibited significant divergence, which may have obscured the underlying pattern. By employing this approach, we aim to enhance the clarity of the displayed training trend. The G_{loss} values of both models show a decrease over time. While, our D_{loss} curve demonstrates an initial increase followed by a subsequent decline.

To further demonstrate the convergence properties of our model, Figure 9 depicts the discriminator loss curve over an extended training period. As observed, the loss consistently decreases and eventually plateaus, indicating the model reaches a stable equilibrium. Our proposed methodology exhibits smaller fluctuations in both generator and discriminator losses compared to Gated Conv, as evidenced by the loss curves. This demonstrates a more stable and effective training process. The superior convergence properties of our framework, with narrower loss ranges, indicate it is better optimized and outperforms Gated Conv for image inpainting. Our loss trajectory analysis provides quantitative verification that the training stability afforded by our approach translates to improved model performance.



(a) Gated Conv (b) Ours

Figure 7. The loss curves of generators. Our method demonstrates narrower ranges of change on G_{loss}.



(a) Gated Conv (b) Ours

Figure 8. The loss curve of our discriminator in larger epochs. Our D_{loss} curve exhibits a trend of initial growth followed by decline.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 July 2023 doi:10.20944/preprints202307.1811.v1

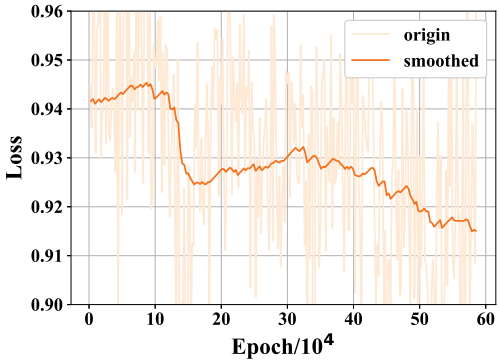


Figure 9. Targets hiding case study with comparison. The images from left to right represent the inputs, original images, and the result of Partial Conv, Cont Atten, Gated Conv, and ours.

4.2. Experimental Comparison for the Targets Hiding Task

In comparison to the method presented in [6] for auto-detection and hiding of targets, we further enhance the effectiveness of the results obtained through semantic segmentation. While object detection methods like MASK-RCNN produce results that cover the entire object surface area, their output may be too rough or simple to facilitate the subsequent step of object restoration using object contours. On the other hand, predictions made by a semantic segmentation network yield more accurate contours, but may not cover objects entirely, leading to exposed regions that reveal information about hidden objects. In light of this drawback, we dilated the labels of our datasets, utilizing the resulting masks as missing portions. The input of the network consisted of images containing four channels (R, G, B, and mask) with hiding applied to the outputs. Figure 10 displays the results of our hiding models.

Figure 10. Targets hiding case study with comparison. The images from left to right represent the inputs, original images, and the result of Partial Conv, Cont Atten, Gated Conv, and ours.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 July 2023 doi:10.20944/preprints202307.1811.v1

In this task, we consider several crucial aspects that influence the effectiveness of our approach. The first point is similar to image inpainting, where the focus is on maintaining continuity between the repaired and original regions. Inconsistency in pixel distribution can cause discontinuities, making it difficult to distinguish differences in pixel values visually, especially for high-density target areas. However, evaluating the mean image pixel values can help address these challenges. The second aspect concerns whether objects are entirely hidden, and although Gated Conv and our model demonstrate good hiding performance overall, large object hiding remains an issue. Lastly, attention must be paid to the characteristics of the generated objects, such as repairing a region covering part of a road, which must be restored while preserving continuity with the original road. As depicted in Figure 10, the original characteristics of roads remain intact when there are only a few missing parts. Our model successfully maintains the width and high continuity of the original roads, highlighting its effectiveness in this regard.

Table 3 shows the evaluation indexes of these models. Due to the lack of accepted evaluation methods for target hiding, we selected similarities as our evaluation indexes, including ℓ_1Sim, ℓ_2Sim, PSNR, SSIM, and UQI. The ℓ_1Sim and ℓ_2Sim reflect the overall pixel distribution difference between the original and repaired images. The similarity values for these methods were relatively close, indicating they could remove targets with a similar pixel distribution as the original images. However, since buildings have different colors than the background, the similarity between results and the original will generally be lower when removing targets. Our model had the minimum value in both indicators compared to other models, proving its efficacy in target hiding.

Table 3. Evaluation index of models on target hiding.

Methods	Cont Atten	Partial Conv	Gated Conv	RATH (Ours)
ℓ_1 **Sim/(%)**	**97.45**	97.51	97.68	97.52
ℓ_2 **Sim/(%)**	85.32	**85.26**	86.02	85.54
PSNR	**18.19**	18.32	18.60	18.33
SSIM/(%)	88.92	88.71	88.71	**88.18**
UQI/(%)	86.41	86.74	87.31	**86.40**

4.3. Experimental Comparison for the Edge-guided Target Hiding Task

Although the above method could to some extent, effectively conceal targets, the location and contour information of the repaired images were still discernible. To address this issue, we propose an extension to our model by incorporating edge guidance for image generation. Specifically, given that our primary objective is to hide objects, we train the model using only object edges. By changing the original object edge, the proposed model effectively obscures an object's location information to mislead the viewer. In addition, the target-hiding model's robust restoration ability enables the preservation of valid detail characteristics of the targets, thereby concealing their location information.

We compared the proposed method with the image inpainting model, and the results are presented in Figure 11. Our model yielded results that were more similar to the original images, as demonstrated by the results in the yellow region of the figure. In particular, the fake buildings generated by the two methods varied in color, whereas our model reproduced the same color as the original buildings. However, we observed that the existing methods struggled to handle large objects, with particularly poor results for large buildings. Additionally, they were unable to deal with the connection part between foreground and background, such as the red region.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 July 2023 doi:10.20944/preprints202307.1811.v1

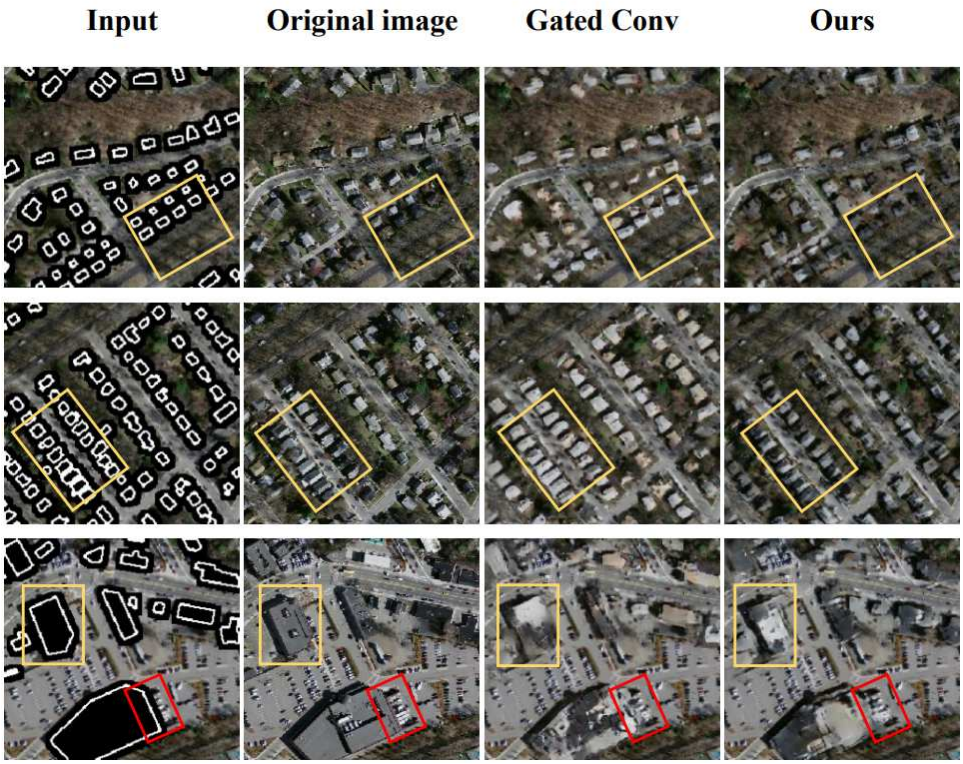


Figure 11. Edge-guided targets hiding case study with comparison. The images from left to right represent the inputs, original images, and the result of Gated Conv, and ours.

The comparison between our proposed method and the Gated Convolution model on the edge-guided target hiding task is presented in Table 4. In order to demonstrate the efficacy of our method, we computed the similarity between the results and the original images. Specifically, we replaced the true targets with fake targets at their respective locations. A higher similarity evaluation index indicates greater ability to perform this task. Our method achieved higher values in all five indexes, which demonstrates its superior suitability for the edge-guided target hiding task.

Table 4. Evaluation index of models on edge-guided target hiding.

Methods	Gated Conv	RATH (Ours)
ℓ_1 **Sim/(%)**	97.45	**97.84**
ℓ_2 **Sim/(%)**	85.31	**86.44**
PSNR	18.19	**18.80**
SSIM/(%)	89.50	**90.44**
UQI/(%)	88.21	**89.01**

Additionally, we drew some edges that differed from the original building distribution. By using masks to cover the buildings that were intended to be concealed, our models were able to generate new objects on the incomplete images. These hidden results have been illustrated in Figure 12. The ability to generate fake objects in both models is undeniable, however, there are still differences regarding the relationship between the foreground and background. We made the building have a regular arrangement and painted edges and masks, giving the impression that there were roads in the images. Our model was able to recognize the regular pattern, and generate the expected roads between the fake buildings, as shown in the second row. Although the new roads in the first row were similar to the original image, the crossover point of two roads was not executed well. The two images did not connect as 'T' roads as we had expected. Our model was highly effective in dealing with the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 July 2023 doi:10.20944/preprints202307.1811.v1

connection points between the generated roads and the original roads, as demonstrated in the third row. As Figure 12 indicates, the finely designed mask and edge images were able to guide the models in generating new and significantly different complete images.

Figure 12. Edge-guided targets hiding case study with comparison. We arranged the buildings in a regular manner and painted edges and masks.

Overall, our proposed method produced significant improvements over existing approaches, effectively concealing target location and contour information while maintaining the visual fidelity of the original images.

5. Application

To handle large-batch data, we integrate semantic segmentation and target hiding into an automated framework. This section describes the workflow of our proposed application for automated target hiding leveraging deep learning approaches. As illustrated in Figure 13, the framework first performs semantic segmentation on the input image to generate semantic maps. Based on the maps, target regions can be identified and concealed through two alternative techniques - direct replacement or edge-guided synthesis - depending on the desired hiding effects. This automated pipeline enables efficient batch processing for target hiding in large datasets.

We utilize Inception-v3 U-Net, which substitutes the decoder with Inception-v3, as the semantic segmentation network. This approach achieves a relatively balanced trade-off between training efficiency and segmentation performance. The dilate operation expands the prediction area, thereby reducing the impact of semantic segmentation errors on target hiding outcomes. Our proposed framework focuses on achieving effective target hiding. We present two techniques to process original images containing targets, depending on desired hiding effects.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 July 2023 doi:10.20944/preprints202307.1811.v1

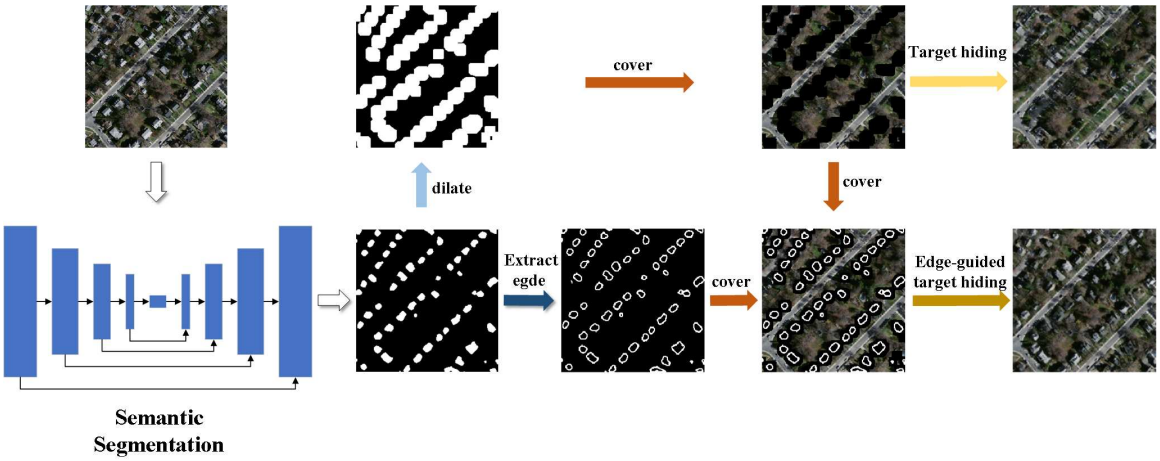


Figure 13. Schematic diagram of the automated target hiding application workflow. Framework illustrating the two techniques for target hiding - direct replacement based on the mask, and edge-guided synthesis.

The first employs a direct target hiding method which takes an image and mask as inputs, where the mask delineates the coverage area and is derived from dilating the segmentation outputs. This method aims to completely remove targets from the original image. The second leverages an edge-guided approach with the image, mask, and edge as inputs. The edge also originates from segmentation but preserves more spatial details. This edge-guided technique generates simulated targets for more natural integration rather than removal. As discussed in subsection 4.2 and subsection 4.3, the two methods achieve very different hiding effects - one eliminates real targets while the other synthesizes fake targets. Our framework provides the flexibility to produce varied results based on desired concealment goals.

Figure 14 compares the outputs of our direct target hiding and edge-guided target hiding techniques. From left to right: original images, semantic segmentation results, target hiding inputs, edge-guided inputs, direct hiding outputs, and edge-guided outputs. The target hiding inputs provide only location information, resulting in random synthesized objects like trees, grass, buildings, or parking lots in the filled regions. Thus, direct hiding is better suited for completely removing any targets from the images.

In contrast, the edge-guided inputs contain contour information delineating foreground objects. Thus, the edge-guided model training is more targeted and learns to generate replacements consistent with the input contours. When the input edges match the typical foreground patterns seen during training, such as buildings, the model synthesizes building-like structures in missing areas. Furthermore, the edge-guided approach captures relationships between foreground and background elements like buildings and roads. This domain-specific knowledge enables more semantically coherent and natural scene completion compared to direct target hiding. The edge guidance provides critical spatial cues to generate plausible foreground objects that blend with original backgrounds.

Furthermore, the edge utilized is a simple binary image of white and black pixels that can be easily generated. As shown in Figure 12, a random mask and hand-drawn outline could also be supplied to our framework. However, the hand-drawn outline should reflect similar spatial distributions as the targets, such as buildings in a regular grid layout. Provided with a reasonable edge, our proposed edge-guided target hiding technique can effectively conceal the desired target regions. The ability to use crude inputs like hand-drawn outlines highlights the robustness and flexibility of our edge-guided approach for target hiding under varied conditions.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 July 2023 doi:10.20944/preprints202307.1811.v1

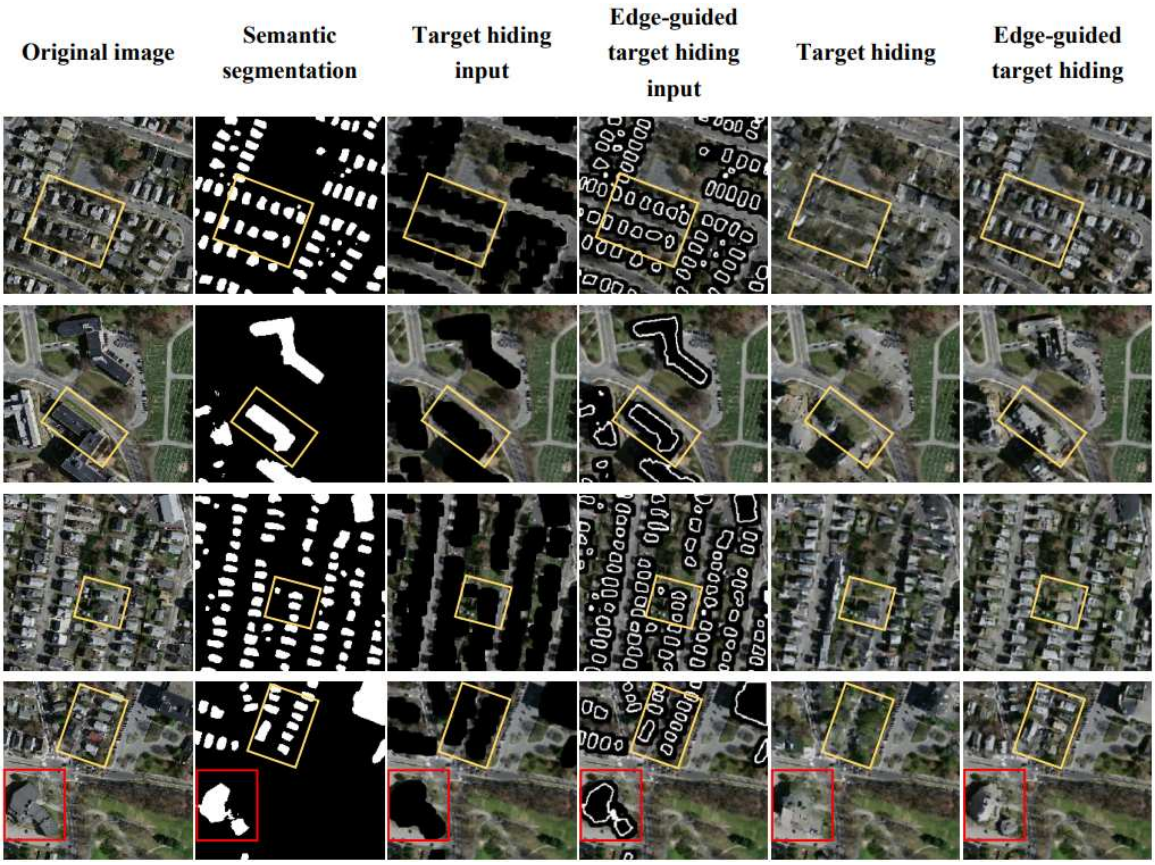


Figure 14. Comparison of target hiding and edge-guided target hiding. From left to right: original images, semantic segmentation results, target hiding inputs, edge-guided inputs, direct hiding outputs, and edge-guided outputs.

6. Conclusion

Target hiding is crucial for emergency mapping using remote sensing data. To obtain a suitable target hiding model, we introduce a residual attention mechanism into image inpainting models and adopt an edge-guided training strategy. The residual attention mechanism renders our model more flexible, economical, and efficient. Specifically, our model reduces the parameter size by 1MB and decreases training time by 0.045 seconds per batch (with a batch size of 16). The residual attention mechanism also resolves issues of gradient instability without compromising the hiding effect. Furthermore, we replace the kernels used for fusing contextual attention layers with full 1 matrices to enlarge the patch size. Our model demonstrates superior performance on image inpainting, target hiding, and edge-guided target hiding tasks. We extend the edge-guided function to preserve target contours and positions, misleading viewers with fabricated targets. Experiments prove our model is well-suited for target hiding tasks. Further, by integrating semantic segmentation, our framework can efficiently process large batches of remote sensing data in an automated manner.

Although our proposed model demonstrates effectiveness for both image inpainting and target hiding tasks, several aspects could be further improved. The free-form mask generation algorithm currently employed is imperfect, as the regularly-centered masks do not account for missing regions located in the corners of images. Moreover, our current model does not incorporate a loss function to explicitly enforce edge coherence between the filled region and original image. This could result in artifacts or discontinuities along boundaries. Designing suitable loss functions to improve boundary smoothness and seamless transition represents an important direction for future refinement.

Author Contributions: Hao Yuan: Conceptualization, Methodology, Software, Writing-Original Draft Preparation, and Visualization. Yongjian Shen: Conceptualization, Resources, Writing-Original Draft, Supervision, Project Administration, and Methodology. Ning Lv: Data curation, and Writing-review. Yuheng Li: Validation, and Resources. Chen Chen: Editing, and Formal Analysis. Zhouzhou Zhang: Investigation and Data curation.

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

References

1. Zhu, Q.; Cao, Z.; Lin, H.; Xie, W.; Ding, Y. Key technologies of emergency surveying and mapping service system. *Wuhan Daxue Xuebao (Xinxi Kexue Ban)/Geomatics and Information Science of Wuhan University* **2014**, *39*, 551–555. doi:10.13203/j.whugis20130351.
2. Ding, W.; Zhang, L. Building Detection in Remote Sensing Image Based on Improved YOLOV5. 2021 17th International Conference on Computational Intelligence and Security (CIS), 2021, pp. 133–136. doi:10.1109/CIS54983.2021.00036.
3. Chen, C.; Yao, G.; Liu, L.; Pei, Q.; Song, H.; Dustdar, S. A cooperative vehicle-infrastructure system for road hazards detection with edge intelligence. *IEEE Transactions on Intelligent Transportation Systems* **2023**.
4. C. Chen, G. Yao, C. Wang, S. Goudos, and S. Wan, "Enhancing the robustness of object detection via 6g vehicular edge computing," *Digital Communications and Networks*, vol. 8, no. 6, pp. 923–931, 2022.
5. Sui, B.; Cao, Y.; Bai, X.; Zhang, S.; Wu, R. BIBED-Seg: Block-in-Block Edge Detection Network for Guiding Semantic Segmentation Task of High-Resolution Remote Sensing Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2023**, *16*, 1531–1549. doi:10.1109/JSTARS.2023.3237584.
6. Qiu, T.; Liang, X.; Du, Q.; Ren, F.; Lu, P.; Wu, C. Techniques for the Automatic Detection and Hiding of Sensitive Targets in Emergency Mapping Based on Remote Sensing Data. *ISPRS International Journal of Geo-Information* **2021**, *10*. doi:10.3390/ijgi10020068.
7. Yu, J.; Lin, Z.; Yang, J.; Shen, X.; Lu, X.; Huang, T.S. Generative Image Inpainting with Contextual Attention. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 5505–5514. doi:10.1109/CVPR.2018.00577.
8. Yu, J.; Lin, Z.; Yang, J.; Shen, X.; Lu, X.; Huang, T. Free-Form Image Inpainting With Gated Convolution. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 4470–4479. doi:10.1109/ICCV.2019.00457.
9. Liu, G.; Reda, F.A.; Shih, K.J.; Wang, T.C.; Tao, A.; Catanzaro, B. Image Inpainting for Irregular Holes Using Partial Convolutions, 2018, [arXiv:cs.CV/1804.07723].
10. Nitzberg, M.; Mumford, D.; Shiota, T. *Filtering, Segmentation and Depth*; Vol. 662, 1993. doi:10.1007/3-540-56484-5.
11. Kokaram, A.; Morris, R.; Fitzgerald, W.; Rayner, P. Interpolation of missing data in image sequences. *IEEE Transactions on Image Processing* **1995**, *4*, 1509–1519. doi:10.1109/83.469932.
12. Hirani, A.N.; Totsuka, T. Combining frequency and spatial domain information for fast interactive image noise removal. *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques* **1996**.
13. Masnou, S.; Morel, J.M. Level lines based disocclusion. Proceedings 1998 International Conference on Image Processing. ICIP98 (Cat. No.98CB36269), 1998, pp. 259–263 vol.3. doi:10.1109/ICIP.1998.999016.
14. Bertalmío, M.; Sapiro, G.; Caselles, V.; Ballester, C. Image inpainting. *Proceedings of the 27th annual conference on Computer graphics and interactive techniques* **2000**.
15. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks. *Advances in Neural Information Processing Systems* **2014**, *3*. doi:10.1145/3422622.
16. Radford, A.; Metz, L.; Chintala, S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks **2015**.
17. Mehralian, M.; Karasfi, B. RDCGAN: Unsupervised Representation Learning With Regularized Deep Convolutional Generative Adversarial Networks. 2018 9th Conference on Artificial Intelligence and Robotics and 2nd Asia-Pacific International Symposium, 2018, pp. 31–38. doi:10.1109/AIAR.2018.8769811.
18. Brock, A.; Donahue, J.; Simonyan, K. Large Scale GAN Training for High Fidelity Natural Image Synthesis, 2019, [arXiv:cs.LG/1809.11096].

19. Wang, Z. Using Gaussian Process in Clockwork Variational Autoencoder for Video Prediction. 2022 International Conference on Information Technology Research and Innovation (ICITRI), 2022, pp. 6–11. doi:10.1109/ICITRI56423.2022.9970241.
20. Yílmaz, M.A.; Keleşş, O.; Güven, H.; Tekalp, A.M.; Malik, J.; Kíranyaz, S. Self-Organized Variational Autoencoders (Self-Vae) For Learned Image Compression. 2021 IEEE International Conference on Image Processing (ICIP), 2021, pp. 3732–3736. doi:10.1109/ICIP42928.2021.9506041.
21. Liu, X.; Gherbi, A.; Wei, Z.; Li, W.; Cheriet, M. Multispectral Image Reconstruction From Color Images Using Enhanced Variational Autoencoder and Generative Adversarial Network. *IEEE Access* **2021**, *9*, 1666–1679. doi:10.1109/ACCESS.2020.3047074.
22. Nichol, A.; Dhariwal, P.; Ramesh, A.; Shyam, P.; Mishkin, P.; McGrew, B.; Sutskever, I.; Chen, M. GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models, 2022, [arXiv:cs.CV/2112.10741].
23. Ramesh, A.; Dhariwal, P.; Nichol, A.; Chu, C.; Chen, M. Hierarchical Text-Conditional Image Generation with CLIP Latents, 2022, [arXiv:cs.CV/2204.06125].
24. Saharia, C.; Chan, W.; Saxena, S.; Li, L.; Whang, J.; Denton, E.; Ghasemipour, S.K.S.; Ayan, B.K.; Mahdavi, S.S.; Lopes, R.G.; Salimans, T.; Ho, J.; Fleet, D.J.; Norouzi, M. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding, 2022, [arXiv:cs.CV/2205.11487].
25. Mirza, M.; Osindero, S. Conditional Generative Adversarial Nets, 2014, [arXiv:cs.LG/1411.1784].
26. Zhang, H.; Xu, T.; Li, H.; Zhang, S.; Wang, X.; Huang, X.; Metaxas, D. StackGAN: Text to Photo-Realistic Image Synthesis with Stacked Generative Adversarial Networks. 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 5908–5916. doi:10.1109/ICCV.2017.629.
27. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-Image Translation with Conditional Adversarial Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 5967–5976. doi:10.1109/CVPR.2017.632.
28. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2242–2251. doi:10.1109/ICCV.2017.244.
29. Pathak, D.; Krahenbuhl, P.; Donahue, J.; Darrell, T.; Efros, A.A. Context Encoders: Feature Learning by Inpainting. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); IEEE Computer Society: Los Alamitos, CA, USA, 2016; pp. 2536–2544. doi:10.1109/CVPR.2016.278.
30. Yang, C.; Lu, X.; Lin, Z.; Shechtman, E.; Wang, O.; Li, H. High-Resolution Image Inpainting Using Multi-scale Neural Patch Synthesis. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 4076–4084. doi:10.1109/CVPR.2017.434.
31. Iizuka, S.; Simo-Serra, E.; Ishikawa, H. Globally and Locally Consistent Image Completion. *ACM Trans. Graph.* **2017**, *36*. doi:10.1145/3072959.3073659.
32. Karras, T.; Aila, T.; Laine, S.; Lehtinen, J. Progressive Growing of GANs for Improved Quality, Stability, and Variation, 2018, [arXiv:cs.NE/1710.10196].
33. Yan, Z.; Li, X.; Li, M.; Zuo, W.; Shan, S. Shift-Net: Image Inpainting via Deep Feature Rearrangement. Computer Vision – ECCV 2018; Ferrari, V.; Hebert, M.; Sminchisescu, C.; Weiss, Y., Eds.; Springer International Publishing: Cham, 2018; pp. 3–19.
34. Nazeri, K.; Ng, E.; Joseph, T.; Qureshi, F.; Ebrahimi, M. EdgeConnect: Generative Image Inpainting with Adversarial Edge Learning. 2019.
35. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation, 2015, [arXiv:cs.CV/1505.04597].
36. Srivastava, R.K.; Greff, K.; Schmidhuber, J. Training Very Deep Networks. Advances in Neural Information Processing Systems; Cortes, C.; Lawrence, N.; Lee, D.; Sugiyama, M.; Garnett, R., Eds. Curran Associates, Inc., 2015, Vol. 28, pp. 2377–2385.
37. Xie, S.; Tu, Z. Holistically-Nested Edge Detection. 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1395–1403. doi:10.1109/ICCV.2015.164.
38. Mnih, V. Mnih Massachusetts Building Dataset **2013**.
39. N. Lv, Z. Zhang, C. Li, J. Deng, T. Su, C. Chen, and Y. Zhou, "A hybrid-attention semantic segmentation network for remote sensing interpretation in land-use surveillance," *International Journal of Machine Learning and Cybernetics*, vol. 14, no. 2, pp. 395–406, 2023.

 doi:10.20944/preprints202307.1811.v1

40. Z. Liao, C. Chen, Y. Ju, C. He, J. Jiang, and Q. Pei, "Multi-controller deployment in sdn-enabled 6g space–air–ground integrated network," *Remote Sensing*, vol. 14, no. 5, p. 1076, 2022.
41. C. Chen, C. Wang, B. Liu, C. He, L. Cong, and S. Wan, "Edge intelligence empowered vehicle detection and image segmentation for autonomous vehicles," *IEEE Transactions on Intelligent Transportation Systems*, 2023.