24th European Conference on Artificial Intelligence - ECAI 2020
Santiago de Compostela, Spain

SI-AGAN: Spatial Interpolation with Attentional Generative Adversarial Networks for Environment Monitoring

Yujia Gao and **Liang Liu** and **Chi Zhang** and **Xiao Wang** and **Huadong Ma**[1]

Abstract. Monitoring the status of urban environmental phenomena is of great significance for urban research and management. While The monitoring sites are often insufficient and unevenly, interpolation values vary in urban spaces non-linearly. It is difficult to find a method that fulfills the requirements of accuracy, robustness, and flexibility for various types of phenomena. In this paper, we present a new kind of deep learning driven spatial interpolation method which works on the grid data that can be applied on the unevenly distributed sites. To generate better accurate spatial continuous data, we design the S^2 attention structure and incorporate it with the GAN to turn it into SI-AGAN which can model spatial dependencies across different regions via sparsely and unevenly distributed sampling. It can directly learn an end-to-end mapping between low- and high-quality environmental signals without in-depth knowledge of the phenomenon. Experiments on two real-world air-pollution datasets demonstrate that our training strategy effectively makes the GAN work for the interpolation of uneven data and our proposed SI-AGAN significantly outperforms previous state-of-the-art spatial interpolation methods.

1 INTRODUCTION

Monitoring the status of urban environmental phenomena, such as air quality, noise, and meteorological, is of great significance for urban research and management. Currently, most of the environmental phenomena are characterized by measured point data from monitoring sites, while urban researchers and managers often require spatially continuous data over the whole city to make justified interpretations and effective decisions [16, 29]. Therefore, *spatial interpolation* is developed to predict the values of spatial phenomena in unsampled locations. The existing spatial interpolation methods are mainly classified into three categories: local neighborhood methods, geostatistical methods, and variational methods. However, it is difficult to find a method that fulfills the requirements of accuracy, robustness, and flexibility for various types of phenomena. Different methods can produce quite different spatial representations and in-depth knowledge of the phenomenon is needed to evaluate which one is the closest to reality [3, 22].

On the other hand, because of high construction and maintenance costs, the monitoring sites are often insufficient. Take air quality monitoring in Beijing as an example, only 21 sites cover the main city region with about 50×50 km^2 (see Figure 1), i.e., about 125 km^2 per site. The use of traditional spatial interpolation methods on over sparse measured point data results in a distorted model of spatial distribution, leading to potentially wrong decisions based on misleading spatial information. As shown in Figure 1 (b) and (c), for the central zone of Beijing (about 20×20 km^2), the spatial distribution of Particulate Matter under $2.5\mu m$ ($PM_{2.5}$) generated by 4 sites (following the average site density in the main city of Beijing) losses a lot of details, compared with the spatial distribution generated by all the 13 sites in the area. This implies that for the most area of Beijing (except for the central zone), the obtained $PM_{2.5}$ distribution may be heavily distorted and the predict values are unreliable.

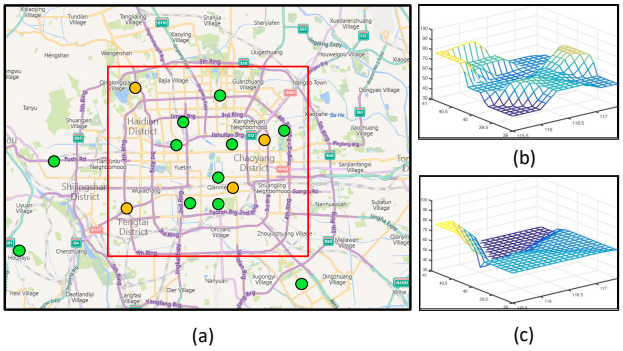


Figure 1. (a) Distribution of Beijing air quality monitoring sites. The region within the red frame is the central area with the highest site density. (b) The high quality signal of $PM_{2.5}$ generated by all the 13 sites in the red frame area. (c) The low quality signal generated by the sites (yellow color).

Towards this end, we explore a new route of *deep learning driven spatial interpolation* that introduces the deep neural network to learn the inner structure and spatial representations of target environmental phenomenon and generate the spatially continuous data with high accuracy. This route is inspired by recent deep neural network based super-resolution work in the image processing field. The authors of [9] first use CNN to learn an end-to-end mapping between low- and high-resolution images. The authors of [21] further present a generative adversarial network based super-resolution framework capable of recovering photo-realistic natural images from 4 times downsampling. In this paper, we try to learn an end-to-end mapping between low- and high-quality spatial signals, generated by sparse sites and dense sites respectively. However, compared with image resolution, deep learning driven spatial interpolation still encounters two specific challenges:

[1] Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia, Beijing University of Posts and Telecommunications, China, email: {gaoyujia, liangliu, zhangchi, byr_wx, mhd}@bupt.edu.cn
*Corresponding author: Liang Liu

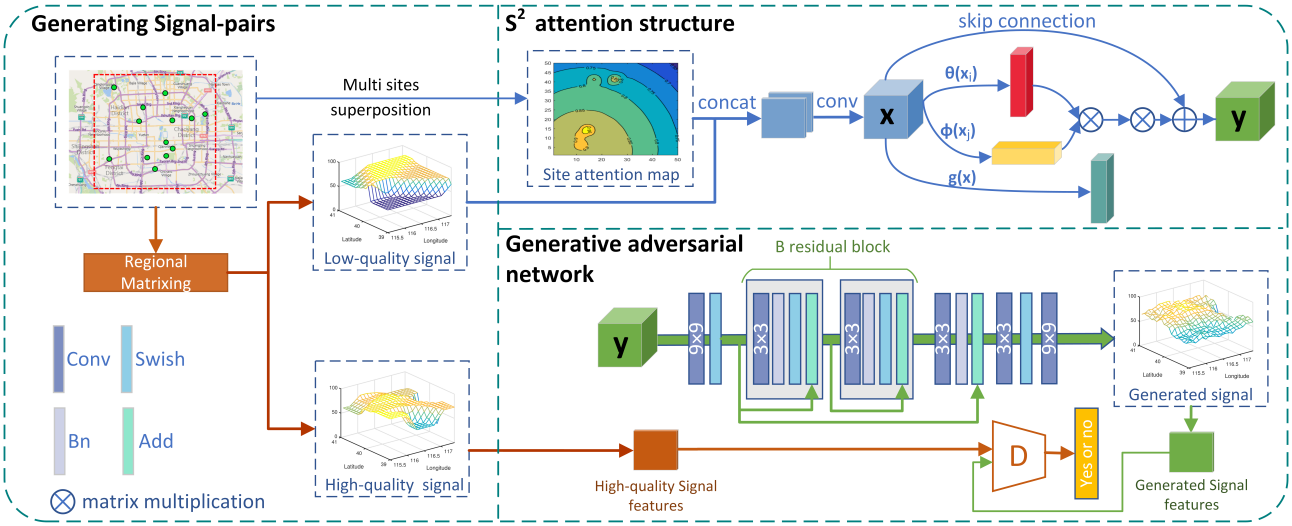


Figure 2. The architecture of the proposed SI-AGAN. Observed values from monitoring stations are generated for Low- and High-quality Signal-pairs that are used for network training. Two different deep residual convolutional networks are used as the generator and discriminator. The S^2 attention structure is shown in the figure, which consists of site attention map and soft-attention layer. The blocks of different color in the figure represent different types of layers, as shown in the lower-left corner of the legend.

1) *Collecting low- and high-quality signal-pairs*. Modeling the mapping relationship is a supervised machine learning process that needs numbers of low-/high-quality signal-pairs as the training and testing samples. However, as mentioned above, monitoring sites are sparsely distributed over the urban area in most cases. It is difficult to obtain high-quality signals as the ground truth for learning the mapping. We observe that although the sites are very sparse over the whole city, there still exist some regions with relatively high site-density. In this paper, we utilize the mean shift algorithm to select the highest site-density region, and then generate the high-quality signal by all sites in the selected region. We further randomly remove some sites in this region and generate the corresponding low-quality signal. Then, the low-/high-quality signal-pairs are obtained.

2) *Uneven distribution of sampling points*. Compared with the super-resolution problem in which the sampling points (pixels) are gridded, the sampling points of environmental signal (measured data from sites) are spatially scattered. This greatly increases the difficulty of learning the mapping relationship via CNN. Existing models rely heavily on convolution to model the dependencies across different image regions, but the convolution processes the information in a local neighborhood and long-range dependencies can only be processed after passing through several convolutional layers that could prevent learning about long-term dependencies [36]. In some spatial interpolation tasks, uneven distribution of sampling points makes long-range dependencies play a more important role in modeling the dependencies across regions around different sites. In this paper, we construct the $\boldsymbol{S^2}$ **attention structure**, which contains two types of attention mechanisms to find the impact relationship among sites. The first is **site attention map** that is based on the sites' distribution, marking the site location and representing the sites' impact on the surrounding grid. The second is a non-local operation based **soft-attention layer**, which is complementary to convolutions, enabling our model can generate signals in which fine details at every location are carefully coordinated with fine details in distant portions of the target region.

Our contributions are summarized as follows:

- We present a new kind of deep learning driven spatial interpolation method, that directly learn an end-to-end mapping between low- and high-quality environmental signals without in-depth knowledge of the phenomenon. It may play a significant role in spatial interpolation in the future because it can fulfill the requirements of robustness and flexibility for a large variety of environmental monitoring tasks.
- To comprehensively utilizing the uneven data, we design a competent spatial interpolation neural network **SI-AGAN** (Attentional Generative Adversarial Network for Spatial Interpolation), as shown in Figure 2. A novel S^2 attention structure is designed to model spatial dependencies across different regions and produce more accurate results of spatial interpolation via sparsely and unevenly distributed sampling.
- We apply our method to Beijing and San Francisco air-pollution datasets. Different quantities and distributions of air quality data are utilized to empirically evaluate the proposed SI-AGAN. Experimental results show that our framework significantly outperforms previous state-of-the-art spatial interpolation methods.

2 RELATED WORKS

Existing methods for spatial interpolation of environmental data include physical model based methods and statistical model based methods. The former includes Street Canyon models [18], Gaussian Plume models [2, 12], and Computational Fluid Dynamics [26], etc. These models are based on several empirical assumptions to simulate the diffusion of air pollutants. Since some empirical assumption may not be consistent with the actual situation, and are affected by site distribution or street geometry, the interpolation results are not satisfactory. The statistical methods [11] are widely used in environmental monitoring inference, such as fuzzy genetic linear membership kriging [27], Land-use regression [17, 31], and autoregressive integrated moving average model [20]. However, many of these models rely on local features from the target location, without taking care of the site distribution between adjacent areas.

Recently, many researchers have developed deep learning based approaches to challenging tasks in urban computing. [39] proposed a semi-supervised learning approach based on a co-training framework that consists of two separate classifiers to estimate fine-grained air quality. [37] exploited an ST-ResNet-based approach to collectively forecast the inflow and outflow of crowds of the city. [5] proposed a deep multi-task learning (MTL) based method to solve spatially fine-grained AQI level estimation and forecasting tasks jointly. [7] used a neural attention model for urban air quality inference, which by adding an attention-based pooling layer that automatically learns the weights of features from different monitoring stations. [41] proposed CEDGANs combining the encoder-decoder structure with adversarial learning to capture deep representations of sampled spatial data, which formalized spatial interpolation as a conditional generative task. These methods have the following problems: 1) These supervised methods usually cannot achieve good performance, due to the lack of training samples; 2) features are defined and extracted artificially, which are often over-specified and incomplete; 3) the effect of site sparsity and heterogeneity on training cannot be considered, especially when interpolating in the low site-density region.

3 PROBLEM FORMULATION

The general formulation of the spatial interpolation problem in 2-dimensional space can be defined as follows:

Given a certain region R, there exist N measured points $p_i(x_i, y_i, z_i), i = 1, 2, \ldots, N$, where (x_i, y_i) and z_i denote the location and value of p_i, respectively. Finding a 2-variate function $f(x, y)$ which passes through the given points, that means, fulfils the condition $f(x_i, y_i) = z_i, i = 1, 2, \ldots, N$.

Because there exist an infinite number of functions that fulfill this requirement, additional conditions have to be imposed, defining the character of various interpolation techniques. In this paper, we try to exploit the deep learning method to learn additional conditions. Therefore, the spatial interpolation problem can be redefined as follows:

Given a low-quality spatial signal, denoted by $\mathbf{S}^L$, which is generated by $p_i(x_i, y_i, z_i), i = 1, 2, \ldots, N$ with a traditional spatial interpolation method, our goal is to recover from $\mathbf{S}^L$ a signal $F(\mathbf{S}^L)$ that is as similar as possible to the ground truth high quality signal $\mathbf{S}^H$. Thus, the spatial interpolation problem becomes how to learn a mapping $F : \mathbf{S}^L \rightarrow \mathbf{S}^H$.

To solve this problem, we need to build a sample set of $(\mathbf{S}^L, \mathbf{S}^H)$ pairs and design an appropriate learning method. The next sections will give details of our solutions.

4 GENERATION OF LOW- AND HIGH-QUALITY SIGNAL-PAIRS

Modeling the mapping relationship is a supervised machine learning process that needs numbers of low- and high-quality signal-pairs. However, it is difficult to obtain high-quality signals as the ground truth for training the model, because of sparsely distributed monitoring sites. We subtly exploit the non-uniformity of site distribution – find the highest site-density subregion $R' \subset R$ and regard the signal generated by all points in R' with spatial interpolation as the high-quality signal $\mathbf{S}^H$. We further randomly remove several sites in R' for decreasing the site-density of R' to the average density of the whole region R, and regard the signal generated by remaining points in R' as the low-quality signal $\mathbf{S}^L$.

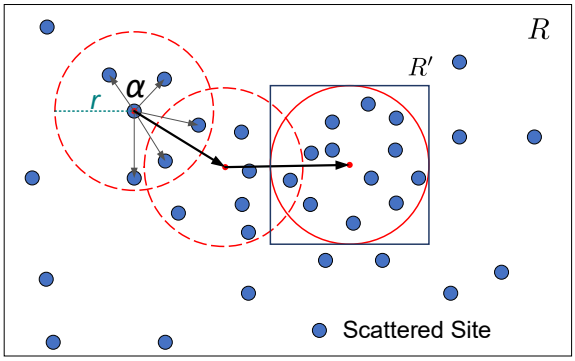


Figure 3. The process of finding the highest site-density region using meanshift algorithm.

To find R', we utilize the mean shift algorithm [8], which is a procedure for locating the maxima of a density function given discrete data sampled from that function. The procedure is shown in Figure 3. For a randomly selected point $\alpha \in R$, all sites in the circle centering on α with radius r respectively generate a vector according to the kernel function K, which is defined as a Gaussian kernel $K(p_i - \alpha) = e^{-c||p_i - \alpha||^2}$ in this paper. Then, we can calculate the sample mean with kernel K at α as

$$M(\alpha) = \frac{\sum_{p_i \in S_r(\alpha)} K(p_i - \alpha) p_i}{\sum_{p_i \in S_r(\alpha)} K(p_i - \alpha)}, \tag{1}$$

where $S_r(\alpha)$ denotes the circular region.

The evolution of α in the form of iterations $\alpha \leftarrow m(\alpha)$ halts when it reaches a fixed point $(m(\alpha) = \alpha)$, i.e., the mean shift algorithm converges to the area with the highest site-density (see the solid circle in Figure 3). To conveniently generate the spatial signal, we circumscribe a square on the circle as the selected region R'.

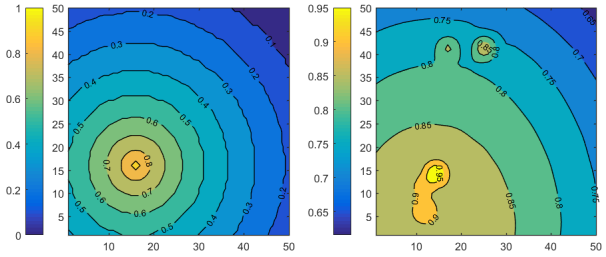
Measured points from all sites in R' are used to generate the high-quality signal $\mathbf{S}^H$ via traditional interpolation methods [2]. Let N' denote the number of measured points in R'. For decreasing the site-density of R' to the site-density of R, we remove $N - \left\lceil \frac{4r^2 N}{\|R\|} \right\rceil$ sites, and then generate $\mathbf{S}^L$ by using measured points from the rest sites in R'.

Obviously, the circle radius r will affect the generation of $(\mathbf{S}^H, \mathbf{S}^L)$ pairs. When R' is too large, the site-density may be insufficient for generating the high-quality signal $\mathbf{S}^H$; when R' is too small, the convolution processes cannot learn enough spatial dependencies for generating the spatial signal over the whole region R of high quality. The selection of r will be discussed in the application section.

5 S^2 ATTENTIONAL GENERATIVE ADVERSARIAL NETWORK

Generative Adversarial Networks (GANs) [13] have shown great performance for signal generation, therefore we exploit GANs to model the mapping relationship between $\mathbf{S}^H$ and $\mathbf{S}^L$, i.e., considering the

[2] In this paper, we exploit the Cubic interpolation to generate $\mathbf{S}^H$ and $\mathbf{S}^L$. The application section will illustrate the performance comparison between Cubic and some other typical interpolation methods.



(a) (b)

Figure 4. (a) Site attention map generated by one site. (b) Site attention map generated by multiple sites.

spatial interpolation problem as an adversarial min-max problem:

$$\min_G \max_D \; \mathrm{E}_{\mathbf{S}^H \sim p_{train}(\mathbf{S}^H)}[\log D(\mathbf{S}^H)] + \mathrm{E}_{\mathbf{S}^L \sim p_G(\mathbf{S}^L)}[\log(1 - D(G(\mathbf{S}^L)))]. \tag{2}$$

The goal is to train a generative model G and make the discriminator D unable to distinguish between generated signal $G(\mathbf{S}^L)$ and high quality signal $\mathbf{S}^H$.

At the same time, the attention mechanism has become a hotspot recently and been successfully used in modeling multi-level dependencies in image captioning [6, 33, 38], image question answering [35], machine translation [4], and image generation [34]. The latest process [15] found that the self-attention module [30], presented for machine translation, can be viewed as a form of the non-local mean in computer vision. This makes non-local operations applicable to image and video problems [40] for effectively capturing long-range dependencies.

Different from the existing methods, our S^2 attention structure takes into account 1) the influence of the geographical distribution of the sites on the interpolation performance and 2) the intrinsic relationship between the monitoring value and the spatial distribution. It consists of two parts: a site attention map and a soft-attention layer. Moreover, we design the loss function to make the network more suitable for the spatial interpolation task. The overall framework of SI-AGAN is shown in Figure 2.

5.1 Site Attention Map

In general, the effect of the site on the surrounding grid interpolation values decreases with the added distance. To reflect this relationship, we constructed the site attention map (SA map for short) as the network input. The SA map, denoted by a $2r \times 2r$ matrix $A \in \mathbb{R}^{2r \times 2r}$, which records the grid's attention of the surrounding sites. We assume that site a has the highest attention value of 1 on its grid. A power function $D(x) = x^\beta$ is used to fit the attenuation of the attention value. For the convenience of quantizing attention value of other grid, we propose a function $A_{map}(x)$ based on the summarized $D(x)$ by:

$$A_{map}(x) = max\left(0, 1 - \frac{D(x)}{D(x_{max})}\right). \tag{3}$$

We set x_{max} as the farthest grid from the site in the entire matrix. Figure 4(a) shows the SA map of a single site.

Usually, there are several monitoring sites in the training region. To reflect the interaction among multiple sites, we superimpose and standardize all sites' SA maps. Figure 4(b) shows the SA map generated by 4 sites. It can be seen that the closer the grid is to the site, the higher the attention value. Finally, the SA map is merged with the low-quality signal as the input of the soft-attention layer.

5.2 Soft-Attention Layer

For enabling G can generate signals in which fine details at every location are carefully coordinated with fine details in distant portions of R, we introduce a non-local operation [32] based attention module to the GAN framework. The generic non-local operation in deep neural networks is defined as

$$\mathbf{o}_i = \frac{1}{\mathcal{C}(\mathbf{x})} \sum_{\forall j} f(\mathbf{x}_i, \mathbf{x}_j) g(\mathbf{x}_j), \tag{4}$$

where $\mathbf{x}$ is the input signal features from the previous hidden layer, i is the index of an output position, and j is the index that enumerates all possible positions. The normalization factor $\mathcal{C}(\mathbf{x})$ is set as $\sum_{\forall j} f(\mathbf{x}_i, \mathbf{x}_j)$. The choice of f is the embedded Gaussian function for computing similarity in an embedding space, i.e.,

$$f(\mathbf{x}_i, \mathbf{x}_j) = e^{\theta(\mathbf{x}_i)^T \phi(\mathbf{x}_j)}, \tag{5}$$

where $\theta(\mathbf{x}_i) = \mathbf{w}_\theta \mathbf{x}_i$ and $\phi(\mathbf{x}_j) = \mathbf{w}_\phi \mathbf{x}_j$ are two embeddings. The function g, computing a representation of $\mathbf{x}$ at the position j, is defined as $g(\mathbf{x}_j) = \mathbf{w}_g \mathbf{x}_j$. The three weight matrixs, $\mathbf{w}_\theta$, $\mathbf{w}_\phi$, and $\mathbf{w}_g$ can be learned by 1×1 convolution.

We can see that for a given i, $\frac{1}{\mathcal{C}(\mathbf{x})} f(\mathbf{x}_i, \mathbf{x}_j)$ becomes the *softmax* computation along j, that indicates the extent to which the model attends to the j^{th} location when synthesizing the i^{th} location. The final output is given by $\mathbf{y}_i = \gamma \mathbf{o}_i + \mathbf{x}_i$ where γ is a scale parameter, $\mathbf{y}$ is the output signal of the same size as $\mathbf{x}$.

5.3 Network Structure and Loss Function

A low-quality signal is sent to the generator after passing the S^2 attention structure. A deep residual convolutional network is used as the generator, which is illustrated in Figure 2. Different colors represent different layers. The number on the convolutional layer represents the size of the filter kernel. The signal passes through a convolutional layer with 9×9 filter kernel and an active layer, and then passes through 16 residual blocks with an identical layout. Inspired by [14], each block contains a convolutional layer with 3×3 kernel and 64 out channels, a batch-normalization layer, an activation layer, and a superimposed layer. A generated signal is obtained by the generator.

The discriminator network is trained to solve the maximization problem in Eq.(9). Since the space limitation, the discriminator is not drawn in Figure 2. It contains eight blocks with the same layout as the generator, and there are an average pooling layer and a sigmoid layer before output.

The definition of the loss function is critical for the performance of the SI-AGAN. We formulate the generator loss as the weighted sum of a grid loss l_{grid} and a rival loss l_{rival} component as:

$$L_G = l_{grid} + \lambda \cdot l_{rival}, \tag{6}$$

where λ is a hyper-parameter that adjusts two loss relationships.

Since the environmental signals are continuous and spatially similar, grid loss can minimize the grid-wise error, making the generated signal more precise. It is used by a mean square error function, which is a widely used optimization target for measuring matrix differences:

$$l_{grid} = \frac{1}{w \times h} \sum_{x=1}^{w} \sum_{y=1}^{h} (\mathbf{S}^H_{x,y} - G(\mathbf{S}^L)_{x,y})^2, \tag{7}$$

where w and h are the sizes of environmental signal.

Rival loss is a generator loss based on the traditional GAN network. The purpose is to generate a data distribution that the discriminator cannot distinguish, so that the generated result is closer to the natural environmental signal. It is defined based on the cross-entropy for binary classification:

$$l_{rival} = -\sum_{n=1}^{N} \log D(G(\mathbf{S}^L)), \tag{8}$$

where $D(G(\mathbf{S}^L))$ is the probability that the generated signal is identified as the high-quality signal.

Discriminator loss is defined as follow:

$$L_D = -\sum_{n=1}^{N} \log D(\mathbf{S}^H) - \sum_{n=1}^{N} \log(1 - D(G(\mathbf{S}^L))). \tag{9}$$

The purpose of the discriminator loss is to let discriminator classify the natural high-quality signal into 1 and the generated signal into 0.

6 APPLICATION AND EVALUATION

We apply the proposed SI-AGAN to air quality monitoring in Beijing and San Francisco. The goal is to generate high-quality signals of different pollutants, over the whole city via measured points from sparse monitoring sites.

6.1 Settings

6.1.1 *Datesets*

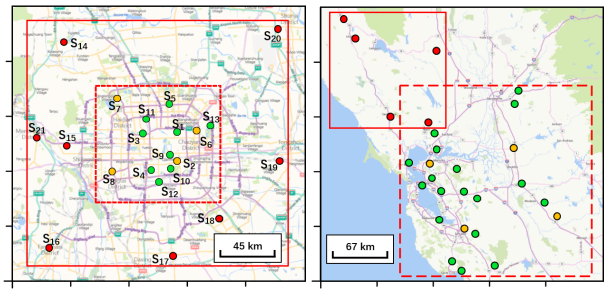
- **Beijing air-pollution dataset** (or Beijing-AP for short), which is released by Beijing municipal environmental monitoring center[3]. It includes the data from all the 21 monitoring sites in the main city of Beijing. Every site reports a record, containing site ID, timestamp, and concentration values of six pollutants ($PM_{2.5}$, PM_{10}, NO_2, CO, O_3 and SO_2), every hour. The unit of pollutant concentration is $\mu g/m^3$. There are 162,568 records from 2017/1/1 to 2018/9/1 in our dataset.
- **San Francisco air-pollution dataset** (or San Francisco-AP for short). It is released by the United States Environmental Protection Agency[4], which are collected by totally 29 sites every hour from 2017/1/1 to 2018/6/30 in San Francisco, America, with a total of 381,060 records. Each record contains 5 feature elements, which are site ID, timestamp, and the concentration of $PM_{2.5}$, NO_2, and O_3. The unit of pollutant concentration is $\mu g/m^3$.

6.1.2 *Preprocessing*

Figure 5 illustrates the sites' distribution of two datasets. For the Beijing dataset (Figure 5(a)), We find a 20×20km^2 area as the highest site-density region R', and the black solid line frame is the main city of Beijing (50×50 km^2). There are 13 sites (S_1-S_{13}) located in R'. The high-quality signal $\mathbf{S}^H$ is generated via the records from the 13 sites, and the low-quality signal $\mathbf{S}^L$ is generated via the records from the randomly selected 4 sites in the 13 sites. That means each record will generate 6 $(\mathbf{S}^H, \mathbf{S}^L)$ pairs for 6 pollutants, respectively.

[3] Official website of Beijing municipal environmental monitoring center: http://zx.bjmemc.com.cn/getAqiList.shtml

[4] Official website of United States Environmental Protection Agency: https://aqs.epa.gov/aqsweb/airdata/download_files.html



(a) Beijing site map (b) San Francisco site map

Figure 5. The red dotted line is the high site-density area that is found by the clustering algorithm. The yellow sites are used to generate low-quality signals, and the red sites are used to verify the effect of the model in the non-training region.

Records from 4 randomly selected days per month are used to generate the testing set, and the rest is used to generate the training set. We also select 10% of training data as the validation set and the training will stop according to the validation score. For the San Francisco dataset (Figure 5(b)), the evaluation process is the same as above. To improve the reliability, we used the average of three identical experiments as the final experimental result for the deep learning method.

Min-max normalization is used for data preprocessing, which is mapped to the [0, 1] range according to the maximum and minimum values. In the evaluation phase, we restored the predicted values to the normal scale.

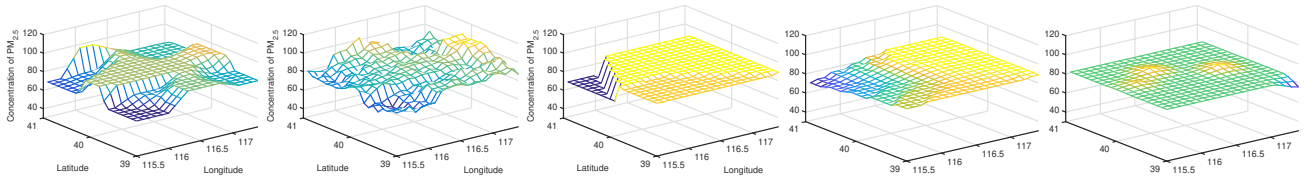
6.1.3 *Hyper-parameters*

Adam optimizer [19] is applied to train the network. The structural parameters of generative net and adversarial net are shown in Figure 2, and the hyper-parameter λ in Eq.(6) is set to $1 \times e^{-4}$. During the training phase, the batchsize of generator network is 21 and the learning rate of both generator and discriminative network are 0.0001. In training, to reduce the computational overhead, two epoch pre-trainings are performed first, meaning that only the generator network is trained. Then, generator and discriminator are trained in an alternating fashion by minimizing the hinge version of the min-max problem until both parties reach a dynamic balance. Swish activation [25] is used to increase the ability of the neural network and express the model by adding nonlinear factors.

6.1.4 *Baselines*

We compare the proposed SI-AGAN with five baselines: the first three are the representative methods of the three traditional categories of spatial interpolation, and the last two are deep learning methods.

- **Nearest Neighbor Interpolation** [1]: A method for determining the value of an unknown point based on nearby points, which is often used for image scaling.
- **Kriging Interpolation** [28]: A widely used method of spatial interpolation for value inference in geostatistics.
- **Cubic Interpolation** [10]: A method based on Delaunay triangles, which can get a smooth fitted surface.
- **SR-CNN** [9]: A method for applying the deep learning method to the image super-resolution field for the first time, which contains a three-layer convolution with different functions.



(a) High-quality signal (b) SI-AGAN interpolation (c) Nearest interpolation (d) Cubic interpolation (e) Kriging interpolation

Figure 6. Interpolation performance of $PM_{2.5}$ in highest site-density subregion.

- **SR-GAN** [21]: A generative adversarial network is used to achieve the super-resolution reconstruction of a single image, which is currently the most accurate method.

Cubic interpolation and Nearest interpolation method use the interpolate function in the scipy package. The Kriging method is implemented by PyKrige toolkit [23]. Our model, as well as the SR-CNN and SR-GAN, are implemented with PyTorch [24] on the GPU server with Tesla K80 and Intel Xeon E5. Similar to SI-AGAN, we test different hyper-parameters for them all, finding the best setting for each.

In the following part, we utilize the mean squared error (MSE) between the generated signals by SI-AGAN/baselines and $\mathbf{S}^H$ as the evaluation metric.

6.2 Case Illustration

To visually show the effect of SI-AGAN, we first choose the $PM_{2.5}$ records at 19:00 on Nov. 12, 2017, in Beijing as an example. Figure 6 illustrates the $PM_{2.5}$ signals over R'. The left one is $\mathbf{S}^H$ (ground truth) generated by 13 measured points, and the right 4 signals are generated by 4 measured points. The signal generated by SI-AGAN is close to $\mathbf{S}^H$, while the other 3 signals generated by the three spatial interpolation methods are heavily distorted because of the sparse measured points.

We also try to use the trained model to generate the $PM_{2.5}$ signal over the whole region R of Beijing. The signal generated by Cubic interpolation on 21 sites is shown in the left part of Figure 7. This is also used as the input $\mathbf{S}^L$ of SI-AGAN, and then the signal is shown in the right part of Figure 7 is generated. We can also observe that the signal generated by SI-AGAN contains much more details than Cubic interpolation, especially in the area with very sparse measured points.

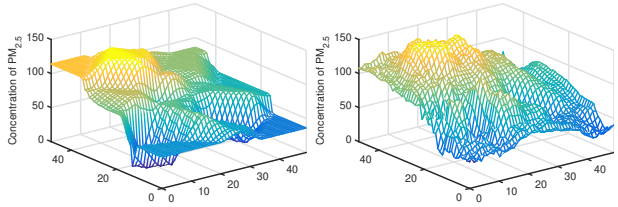


Figure 7. Apply the generated model to the whole city of Beijing which contains low-density region. The left side is network input, and the right side is the reconstructed environmental signal of $PM_{2.5}$.

6.3 Comparison of Statistical Results

We further conduct some quantitative comparisons between SI-AGAN and baselines on two datasets. The comparisons include the three following parts. The evaluation in the training region and non-training region results are listed in Table 1. The effect of r on SI-AGAN performance is listed in Table 2.

6.3.1 *Comparison in Training Region (R')*

For two datasets, SI-AGAN achieves the lowest MSE among the different pollutants. Compared with Kriging, the Nearest neighbor, and Cubic (traditional interpolation methods), SI-AGAN decreases the MSE by 50.5%, 44.3%, 38.7% in Beijing dataset, and 67.2%, 78.0%, 72.5% in San Francisco dataset. In the training region, Kriging performs the weakest, because too sparse site distribution is not conducive to spatial modeling. For the cubic method, the site distribution of the edge region cannot meet the requirements of the Delaunay triangle construction, there will be some blank points at the edge of the grid. In the experiment, we performed the neighborhood completion of the blank point, but the existence of the blank point brings a lot of limitations in the real scene.

On the other hand, deep learning methods achieve better performance than traditional methods, and it has great potential in spatial interpolation tasks. SI-AGAN by adding site attention map and non-local operation to make the network pay attention to the global characteristics, and can utilize the geographical features between the sites to speed up the calculation speed and improve the accuracy. In the Beijing dataset, the average MSE of SI-AGAN decreased by 61.8% and 21.7% compared with SR-CNN and SR-GAN. In the San Francisco dataset, SI-AGAN decreased by 12.0% and 9.1%. Since the average pollutant concentration in America is much smaller than that in China, the performance results of the two datasets on the same pollutant have a large gap, but the trends in different interpolation methods are basically the same.

6.3.2 *Comparison in Non-training Region ($R - R'$)*

To further illustrate the universality of the model, we evaluated on 8 sites (S_{14}-S_{21}) and 5 sites in the non-training region of two datasets respectively. We alternately remove one site in the non-training sites, using SI-AGAN and baselines to predict the values and comparing it with the ground truth from the site. The statistical MSEs are listed in Table 1. All of the pollutants achieved the best interpolation effect in the SI-AGAN method.

For the Beijing dataset in the non-training region, the average MSE of SI-AGAN is reduced by 8.4%, 17.4%, 21.3%, 20.6%, and 17.3% compared to five baseline methods respectively. For the San Francisco dataset, the MSEs have decreased by 3.8%, 26.6%, 24.7%,

Table 1. Comparison between SI-AGAN and baselines in training region and non-training region. The best performance are written in bold.

		Method	Kriging	Nearest	Cubic	SR-CNN	SR-GAN	SI-AGAN
R'	Beijing-AP	$PM_{2.5}$	325.4	265.2	193.4	160.9	151.3	**147.4**
		PM_{10}	3180	2413	2741	2387	2366	**2224**
		NO_2	330.9	319.9	324.3	125.9	121.4	**120.7**
		CO	0.162	0.163	0.181	0.091	0.081	**0.080**
		O_3	492.9	478.1	319.0	192.1	187.5	**184.3**
		SO_2	30.62	28.98	24.98	18.07	17.94	**17.78**
	San Francisco-AP	PM25	34.33	47.34	29.67	10.36	9.87	**9.70**
		NO2	16.44	30.21	26.82	6.39	5.28	**5.26**
		O3	4.11E-05	5.57E-05	5.15E-05	1.77E-5	1.57E-05	**1.56E-05**
$R - R'$	Beijing-AP	$PM_{2.5}$	652.1	544.4	647.7	563.1	533.6	**465.6**
		PM_{10}	3117	5014	3787	3031	4084	**2991**
		NO_2	401.9	468.4	397.9	623.9	592.2	**394.4**
		CO	0.567	0.509	0.637	0.579	0.566	**0.501**
		O_3	710.1	721.4	718.9	807.4	753.4	**660.6**
		SO_2	42.74	54.76	80.39	65.58	44.75	**41.71**
	San Francisco-AP	$PM_{2.5}$	58.60	68.57	67.73	56.13	55.11	**51.64**
		NO_2	23.13	29.68	27.82	23.49	22.25	**21.54**
		O_3	1.19E-4	1.62E-4	1.62E-4	1.20E-4	1.28E-4	**1.17E-4**

62.7%, and 60.3%. It is worth noting that, unlike the training region, the performance of kriging in the non-training region is better. We believe that because only one site is removed during the test, the site density becomes larger and the advantage of kriging is highlighted. Even so, SI-AGAN still has an advantage over other baselines.

In network training, we found that NO_2 concentration in Beijing is more difficult to train than other pollutants. Its convergence curve is down slower. To analyze the cause, we output the interpolation result of SI-AGAN and ground truth. It was found that SI-AGAN is sensitive to the rate of change of the interpolation point, but the value of the change is not certain. In other words, it can get a good inflection point, but there will be larger or smaller than the actual value. The possible reason is that NO_2 has a low concentration in the atmosphere and is greatly affected by the geographical environments such as factories and automobile exhausts. In actual use, for a more mutated data type, to get a better interpolation effect, it is necessary to collect as much data as possible for network training.

6.3.3 Effect of r on SI-AGAN Performance

The value of r should make the number of sites in the selected high site-density area large, and the distance between sites should be as small as possible. This is closely related to the city area and station distribution. We compare the interpolation effects of using different radius to train the model in two air-pollution datasets. Table 2 shows the performance of different r between observed value in official monitoring sites and SI-AGAN interpolation result, using the MSE metric. The smaller the value, the more accurate the interpolation. The best performances are written in bold.

For the Beijing air-pollution dataset, when $\| R' \| = 10 \times 10$ km^2, SI-AGAN cannot learn the regional characteristics very well. However, if the radius is too large ($\| R' \| = 30 \times 30$ km^2), the site density may become insufficient for generating high-quality data signals. For PM_{10} and O_3, the model performance when $\| R' \| = 30 \times 30$ km^2 is significantly lower than $\| R' \| = 20 \times 20$ km^2, although $\| R' \| = 30 \times 30$ km^2 has a bit advantage for CO. Therefore, we use the 20×20 km^2 as the training region in this application.

The same experimental results are presented in the San Francisco dataset. Due to the smaller size of the San Francisco area, we chose the smaller training region than Beijing. When $\| R' \| = 15 \times 15$ km^2, the best results were obtained for the interpolation of the three pollutants. When $\| R' \| = 5 \times 5$ km^2, the interpolation result is significantly deteriorated. When $\| R' \| > 10 \times 10$ km^2, the interpolation result tends to be stable. It can be seen that the value of r is too small to affect the result more seriously than when r is large.

Table 2. Comparison the effect of using different radius r to train the SIAGAN model in two air-pollution datasets. For each dataset, experimental results were performed in the same non-training region.

	Training Region	$PM_{2.5}$	PM_{10}	NO_2	CO	O_3	SO_2
BJ-AP	10×10 km^2	940	5933	1032	0.569	1754	54
	20×20 km^2	**466**	**2991**	**394**	0.501	**661**	**42**
	30×30 km^2	478	4947	477	**0.484**	906	48
SF-AP	5×5 km^2	136.2	-	35.90	-	2.94E-4	-
	10×10 km^2	56.16	-	24.61	-	1.18E-4	-
	15×15 km^2	**51.64**	-	**21.54**	-	**1.17E-4**	-
	20×20 km^2	53.79	-	27.02	-	1.30E-4	-

7 CONCLUSIONS

In this paper, we propose a novel attentional generative adversarial framework, SI-AGAN, for interpolating signals via a sparsely and unevenly distributed monitoring site. Without any additional information, the SI-AGAN model can learn the correspondence between low/high-quality signals through numerous instances. Adding the S^2 attention structure to the generator can extract global features without increasing the number of convolution layers or increasing the convolution kernel. Extensive experiments on two real-world air-pollution datasets demonstrated the great potential of the SI-AGAN model for signal spatial interpolation.

ACKNOWLEDGEMENTS

This work is supported by the National Natural Science Foundation of China (61722201, 61632008), Science Fund for Creative Research Groups of the National Natural Science Foundation of China (61921003), and Fund for International Cooperation and Exchange of NSFC (61720106007).

REFERENCES

[1] Christophe Accadia, Stefano Mariani, Marco Casaioli, Alfredo Lavagnini, and Antonio Speranza, 'Sensitivity of precipitation forecast skill scores to bilinear interpolation and a simple nearest-neighbor average method on high-resolution verification grids', *Weather and forecasting*, **18**(5), 918–932, (2003).

[2] N Kh Arystanbekova, 'Application of gaussian plume models for air pollution simulation at instantaneous emissions', *Mathematics and Computers in Simulation*, **67**(4-5), 451–458, (2004).

[3] Pierre Aumond, Arnaud Can, Vivien Mallet, Bert De Coensel, Carlos Ribeiro, Dick Botteldooren, and Catherine Lavandier, 'Kriging-based spatial interpolation from measurements for sound level mapping in urban areas', *The Journal of the Acoustical Society of America*, **143**(5), 2847–2857, (2018).

[4] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio, 'Neural machine translation by jointly learning to align and translate', in *3rd International Conference on Learning Representations, ICLR 2015*, (2015).

[5] Ling Chen, Yifang Ding, Dandan Lyu, Xiaoze Liu, and Hanyu Long, 'Deep multi-task learning based urban air quality index modelling', *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, **3**(1), 2, (2019).

[6] Long Chen, Hanwang Zhang, Jun Xiao, Liqiang Nie, Jian Shao, Wei Liu, and Tat-Seng Chua, 'Sca-cnn: Spatial and channel-wise attention in convolutional networks for image captioning', in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5659–5667, (2017).

[7] Weiyu Cheng, Yanyan Shen, Yanmin Zhu, and Linpeng Huang, 'A neural attention model for urban air quality inference: Learning the weights of monitoring stations', in *Thirty-Second AAAI Conference on Artificial Intelligence*, (2018).

[8] Yizong Cheng, 'Mean shift, mode seeking, and clustering', *IEEE transactions on pattern analysis and machine intelligence*, **17**(8), 790–799, (1995).

[9] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang, 'Image super-resolution using deep convolutional networks', *IEEE transactions on pattern analysis and machine intelligence*, **38**(2), 295–307, (2016).

[10] Stephen A Dyer and Justin S Dyer, 'Cubic-spline interpolation. 1', *IEEE Instrumentation & Measurement Magazine*, **4**(1), 44–46, (2001).

[11] Jo Eidsvik, Tapan Mukerji, and Debarun Bhattacharjya, *Value of information in the earth sciences: Integrating spatial modeling and decision analysis*, Cambridge University Press, 2015.

[12] Thad Godish, Wayne T Davis, and Joshua S Fu, *Air quality*, CRC Press, 2014.

[13] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, 'Generative adversarial nets', in *Advances in neural information processing systems*, pp. 2672–2680, (2014).

[14] Sam Gross and Michael Wilber, 'Training and investigating residual nets', *Facebook AI Research*, (2016).

[15] Jiatao Gu, James Bradbury, Caiming Xiong, Victor OK Li, and Richard Socher, 'Non-autoregressive neural machine translation', *arXiv preprint arXiv:1711.02281*, (2017).

[16] François Ingelrest, Guillermo Barrenetxea, Gunnar Schaefer, Martin Vetterli, Olivier Couach, and Marc Parlange, 'Sensorscope: Application-specific sensor network for environmental monitoring', *ACM Transactions on Sensor Networks (TOSN)*, **6**(2), 17, (2010).

[17] Saori Kashima, Takashi Yorifuji, Toshihide Tsuda, and Hiroyuki Doi, 'Application of land use regression to regulatory air quality data in japan', *Science of the Total Environment*, **407**(8), 3055–3062, (2009).

[18] Minjoong J Kim, Rokjin J Park, and Jae-Jin Kim, 'Urban air quality modeling with full o3–nox–voc chemistry: Implications for o3 and pm air quality in a street canyon', *Atmospheric Environment*, **47**, 330–340, (2012).

[19] Diederik P Kingma and Jimmy Ba, 'Adam: A method for stochastic optimization', *arXiv preprint arXiv:1412.6980*, (2014).

[20] Ujjwal Kumar and VK Jain, 'Arima forecasting of ambient air pollutants (o 3, no, no 2 and co)', *Stochastic Environmental Research and Risk Assessment*, **24**(5), 751–760, (2010).

[21] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew P Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al., 'Photo-realistic single image super-resolution using a generative adversarial network.', in *CVPR*, volume 2, p. 4, (2017).

[22] Jin Li and Andrew D Heap, 'Spatial interpolation methods applied in the environmental sciences: A review', *Environmental Modelling & Software*, **53**, 173–189, (2014).

[23] BS Murphy, 'Pykrige: development of a kriging toolkit for python', in *AGU fall meeting abstracts*, (2014).

[24] Adam Paszke, Sam Gross, Soumith Chintala, and Gregory Chanan. Pytorch: Tensors and dynamic neural networks in python with strong gpu acceleration, 2017.

[25] Prajit Ramachandran, Barret Zoph, and Quoc V Le, 'Swish: a self-gated activation function', *arXiv preprint arXiv:1710.05941*, **7**, (2017).

[26] H Scaar, T Teodorov, T Ziegler, and J Mellmann, 'Computational fluid dynamics (cfd) analysis of air flow uniformity in a fixed-bed dryer for medicinal plants', in *Ist International Symposium on CFD Applications in Agriculture 1008*, pp. 119–126, (2012).

[27] Rouzbeh Shad, Mohammad Saadi Mesgari, Arefeh Shad, et al., 'Predicting air pollution using fuzzy genetic linear membership kriging in gis', *Computers, environment and urban systems*, **33**(6), 472–481, (2009).

[28] Michael L Stein, *Interpolation of spatial data: some theory for kriging*, Springer Science & Business Media, 2012.

[29] PV Sundareshwar, R Murtugudde, G Srinivasan, S Singh, KJ Ramesh, R Ramesh, SB Verma, D Agarwal, D Baldocchi, CK Baru, et al., 'Environmental monitoring network for india', *Science*, **316**(5822), 204–205, (2007).

[30] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin, 'Attention is all you need', in *Advances in Neural Information Processing Systems*, pp. 5998–6008, (2017).

[31] Rongrong Wang, Sarah B Henderson, Hind Sbihi, Ryan W Allen, and Michael Brauer, 'Temporal stability of land use regression models for traffic-related air pollution', *Atmospheric Environment*, **64**, 312–319, (2013).

[32] Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He, 'Non-local neural networks', in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7794–7803, (2018).

[33] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio, 'Show, attend and tell: Neural image caption generation with visual attention', in *International conference on machine learning*, pp. 2048–2057, (2015).

[34] Tao Xu, Pengchuan Zhang, Qiuyuan Huang, Han Zhang, Zhe Gan, Xiaolei Huang, and Xiaodong He, 'Attngan: Fine-grained text to image generation with attentional generative adversarial networks', in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1316–1324, (2018).

[35] Zichao Yang, Xiaodong He, Jianfeng Gao, Li Deng, and Alex Smola, 'Stacked attention networks for image question answering', in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 21–29, (2016).

[36] Han Zhang, Ian Goodfellow, Dimitris Metaxas, and Augustus Odena, 'Self-attention generative adversarial networks', in *International Conference on Machine Learning*, pp. 7354–7363, (2019).

[37] Junbo Zhang, Yu Zheng, and Dekang Qi, 'Deep spatio-temporal residual networks for citywide crowd flows prediction', in *Thirty-First AAAI Conference on Artificial Intelligence*, (2017).

[38] Zizhao Zhang, Yuanpu Xie, Fuyong Xing, Mason McGough, and Lin Yang, 'Mdnet: A semantically and visually interpretable medical image diagnosis network', in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 6428–6436, (2017).

[39] Yu Zheng, Furui Liu, and Hsun-Ping Hsieh, 'U-air: When urban air quality inference meets big data', in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1436–1444. ACM, (2013).

[40] Luowei Zhou, Yingbo Zhou, Jason J Corso, Richard Socher, and Caiming Xiong, 'End-to-end dense video captioning with masked transformer', in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8739–8748, (2018).

[41] Di Zhu, Ximeng Cheng, Fan Zhang, Xin Yao, Yong Gao, and Yu Liu, 'Spatial interpolation using conditional generative adversarial neural networks', *International Journal of Geographical Information Science*, 1–24, (2019).