



GAGIN: generative adversarial guider imputation network for missing data

Wei Wang^{1,2} · Yimeng Chai^{1,3,4} · Yue Li^{1,2,3}

Received: 23 April 2021 / Accepted: 12 December 2021 / Published online: 8 January 2022
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

Abstract

Missing data imputation aims to accurately impute the unobserved regions with complete data in the real world. Although many current methods have made remarkable advances, the local homogenous regions, especially in boundary, and the reason of the imputed data are still the two most challenging issues. To address these issues, we propose a novel Generative Adversarial Guider Imputation Network (GAGIN) based on generative adversarial network (GAN) for unsupervised imputation, which is composed of a Global-Impute-Net (GIN), a Local-Impute-Net (LIN) and an Impute Guider Model (IGM). The GIN looks at the entire missing regions to generate and impute data as a whole. Considering the reason of the GIN results, IGM is assigned to capture coherent information between global and local and guide the LIN to look only at a small area centered at the missing focused regions. After processing these three modules, the local imputed results are concatenated to those global imputed results, which impute the rational values and refine the local details from rough to accurate. The comprehensive experiments demonstrate our proposed method is significantly superior to the other three state-of-the-art approaches and seven traditional methods, and we achieve the best RMSE surpass the second-best method on both numeric datasets (17.3%) and image dataset (24.1%). Besides, the extensive ablation study validates the superior performance for dealing with missing data imputation.

Keywords Missing data imputation · Imputation guider · Local refine · Deep learning

1 Introduction

Missing data imputation is an important and common topic in the real world, aiming to impute the uncollected and unobserved regions with rational values. Many imputation approaches have been proposed to handle data containing missing observations, such as multivariate time series

imputation [1–5], image imputation [6–9], regression imputation [10, 11], sentence completion [12–14], to cite just a few.

To deal with missing data imputation, various traditional methods can be categorized into two types: (1) The simple statistical imputation methods, (2) The machine learning-based imputation methods. But these methods have the limitations of changing the original data distribution and assuming the data correlate with characteristics [44, 45].

Recently, a few effective methods about missing data imputation based on the lately prevalent generative adversarial network (GAN) [15] have been performed [16–18]. Most of these methods utilize generators and discriminators to learn the information of unobserved regions. The generator frequently generates and imputes missing data to deceive the discriminator, while discriminator discriminate between the imputation and fake data. Although these methods enhance the characteristic expression and follow the data distribution compared to the traditional methods, local homogenous regions especially

✉ Yue Li
liyue80@nankai.edu.cn

¹ College of Computer Science, Nankai University, Tianjin 300350, People's Republic of China

² Key Laboratory for Medical Data Analysis and Statistical Research of Tianjin (KLMDASR), Tianjin 300350, People's Republic of China

³ Trusted AI System Laboratory, College of Cyber Science, Nankai University, Tianjin 300350, People's Republic of China

⁴ Tianjin Key Laboratory of Network and Data Security Technology, Tianjin 300350, People's Republic of China

in boundary (as shown in the blue and red boxes in Fig. 1) and the reason of the imputed data are still the challenging issues that negatively impact the missing data imputation results. Essentially, two main reasons are resulting in these issues being challenging to solve. First, GAN-based methods pay more attention to make the distribution of generated data approximate the distribution of real data as a whole, and the details receive insufficient attention [41, 47, 50]. Hence, the detailed local representation of the imputation is still not accurately accessed. Second, the existing methods capture the random noise to feed the model initially, which ignores the guiding results between different levels from global to local [15, 46, 48, 49]. Therefore, we not only focus on the global and local regions and avail the information effectively to combine the global and the local regions well.

In this paper, we propose a novel unsupervised learning of GAN-based imputation model GAGIN to deal with missing data imputation, which consists of a Global-Impute-Net (GIN), a Local-Impute-Net (LIN) and an Impute Guider Model (IGM). GIN captures the global distribution from the entire dataset and initially generates the imputation as a whole. Considering the clutter local area and the improper results, we design IGM to stretch the information from global to local and guide LIN to refine the local regions, especially the boundary of imputation results. Our GAGIN learns the guide information between global and local and refine local regions enhancing the imputation performance. Hence, the proposed GAGIN imputes more rational values and the inadequate local regions from rough to accurate.

To sum up, the significant contributions of our proposed methods can be summarized as follows:

- We propose a novel GAGIN for missing data imputation. This network is designed to equip with three sub-networks, the GIN can generate imputation as a whole,

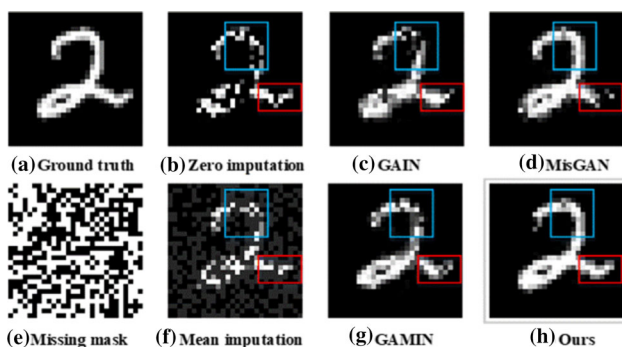


Fig. 1 Some examples of challenges in missing data imputation. **a** Ground truth (GT) and **e** missing mask of which missing components are colored black. Visualization of **b** Zero imputation, **f** Mean imputation, **c** GAIN [18], **g** GAMIN [16], **d** MisGAN [17] and **h** Our proposed GAGIN

while the LIN guided by IGM refines the local region especially for inadequate areas.

- Our method compared with other 10 missing data imputation methods verifies the effectiveness of our GAGIN. The experimental results illustrate that our method outperforms the other state-of-the-art approaches on both numeric and image datasets. Furthermore, the comprehensive ablation study demonstrates that our IGM and LIN perform their effectiveness and superiority for dealing with missing data imputation.

2 Related work

2.1 Traditional methods

Existing traditional missing data imputation methods can be categorized into two classes. The first algorithms are statistical imputation methods such as zero imputation, mean imputation, and the most common value imputation [19]. The second kind of methods is machine learning-based imputation algorithms. Multivariate Imputation by Chained Equations (MICE) [20] fills the missing data by using iterative regression model. MissForest algorithm [21] uses known variables as feature data and the missing data variable as a label and updates the missing values predicted by the random forest. Matrix completion algorithm [22] factorizes the incomplete dataset low-rank matrices and adopts the product of these two matrices to impute the missing data. Expectation–Maximization (EM) algorithm [23] consisting of the “expectation” step and the “maximization” step iteratively updates model parameters and imputed data. K-Nearest Neighbor (KNN) algorithm [24] uses the mean value of k nearest neighbors to fill missing data.

Although these methods are somewhat effective in imputing the missing data, statistical and machine learning-based methods for imputation have various drawbacks. For instance, the main drawbacks of these imputation methods are the lack of the utilization of the information and the change of the original data distribution. Furthermore, MICE, MissForest, EM, KNN, etc., are all based on the assumption that the data are missing at random and having a correlation between characteristics.

2.2 Generative adversarial networks (GANs)-based methods

In recent years, generative adversarial network (GAN) schemes [15, 25–27] have significantly developed missing data imputation. Yoon et al. [18] proposed an imputation

network (GAIN) that employs a hint vector to complete the missing data. It is trained with a discriminator trying to discriminate which in the fake complete data were imputed or not. GAIN has improved the performance in imputation compared to the traditional methods. Nevertheless, the main drawback of GAIN is the limitation of dealing with high dimensional datasets or high missing rate. Li et al. [17] proposed a GAN architecture (MisGAN) and an imputation method using it. MisGAN, consisting of two generators and two discriminators, learns a complete data. For imputation, another pair of generators and discriminators are used. Although MisGAN has taken advantage of high-dimensional incomplete data, it tends to neglect correspondence between imputations and groundtruth. The above papers did not consider the design of local, so there is a lack of local details. Yoon and Sull [16] proposed a generative adversarial multiple imputation network (GAMIN), which generated candidates of imputation and presented a confidence prediction method to perform reliable multiple imputation. GAMIN has made tremendous advances in high missing rate. However, the missing rate rarely exceeds 80% for dataset. This work has studied local to a certain extent, but due to the limited global-to-local GAN information transmission problem, the solution is not good.

3 The proposed GAGIN

In this paper, we propose a generative adversarial guider imputation network (GAGIN) for missing data imputation. Our design is based on solving the previous two problems:

global-to-local details and conditional guiding. We introduce our model in this section, and in Sect. 4 we provide the theoretical discussion proving our designing. The proposed GAGIN receives a data missing completely at random and outputs the imputation using the guide concept. Figure 2 illustrates the overall architecture of the proposed GAGIN, which involves a Global-Impute-Net (GIN), a Local-Impute-Net (LIN) and an Impute Guider Model (IGM). Section 3.1 describes our GIN, and the detailed explanation of our IGM and LIN is mentioned in Sects. 3.2 and 3.3. The model training for imputation is explained in Sect. 3.4.

3.1 Global-impute-net (GIN)

We design GIN to focus on the entire missing regions to generate and impute data as a whole. After this network simulates the global information, it also needs to pass the information to the local network through the following guider. The design of GIN is as follows:

In the generator, we take the missing data \tilde{X} , the missing mask M , and a noise variable Z as input, while output is a vector of generated data \bar{X} . Obtained by taking the partial observation of \tilde{X} and replacing each missing region with the corresponding value of \bar{X} , \hat{X} corresponds to the completed data vector. Thus, we define \bar{X} and \hat{X} in Eqs. (1) (2) as below:

$$\bar{X} = G_g(\tilde{X}, M, (1 - M) \odot Z) \tag{1}$$

$$\hat{X} = M \odot \tilde{X} + (1 - M) \odot \bar{X} \tag{2}$$

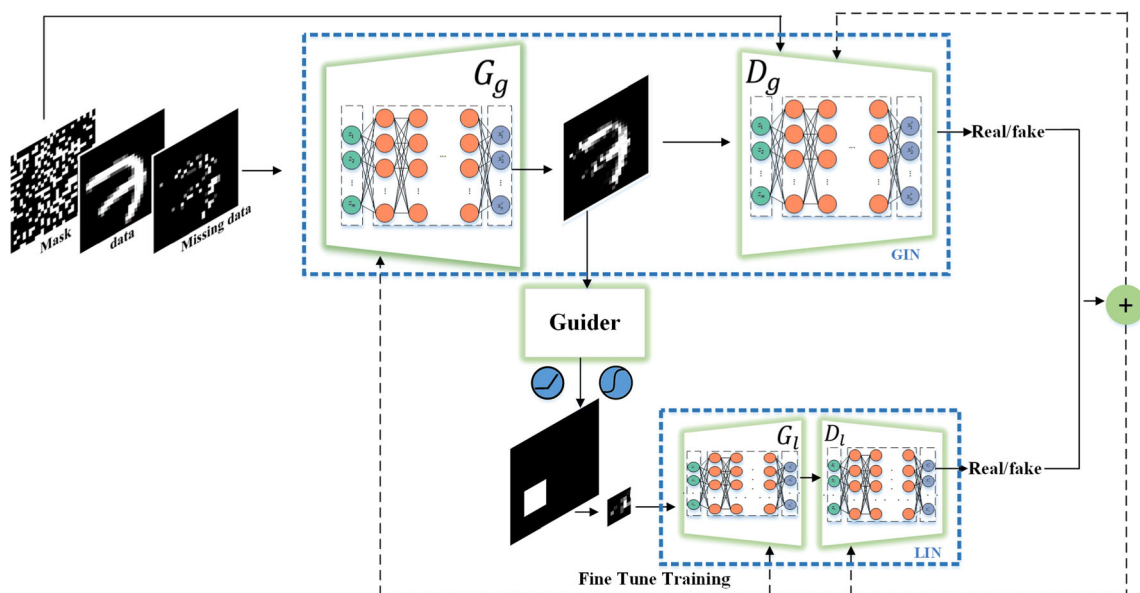


Fig. 2 The overall architecture of Generative Adversarial Guider Imputation Network (GAGIN)

where G_g is defined as a function transforming the unobserved data to generated data for every component and \odot denotes element-wise multiplication.

There is a global discriminator D_g used to train our GIN model, which tries to criticize whether each component of input is imputed or not. The missing mask M and \hat{X} are combined to feed in D_g and output a value in the [0,1] range. Thus, the loss function for the adversarial global impute net is defined in Eq. (3) as follows:

$$L_{adv}^g = \sum_{i=1}^s L_{adv} \left(G_g^i; D_g^i(\hat{X}, M) \right) \tag{3}$$

where the i -th component of D_g^i corresponds to the probability that the i -th component of \hat{X} was observed.

3.2 Impute guider model (IGM)

For missing data imputation, it is critical for imputers to explore the felicitous structure and the appropriate local information especially in smooth homogeneous regions. However, the previous methods learn a variety of global information and treat all characteristics without distinction so that the finer details are ignored. In using GAN to solve such a problem, the difficulty lies in how to use the global information as the pre-condition of the local distribution to guide the generation. As we all know, GANs cannot directly use complex conditions, and the output dimension of GIN is about R^n , which is not suitable as a condition. For this we designed a module IGM.

Based on the above observation, we propose an IGM to guide the LIN according to the GIN result \hat{X} and missing mask to meet both local refinable and values reasonable. Therefore, the global imputation results act as a prior to leading the generation and adjusting the local imputation results. Each local information is extracted from the intermediate imputation guider by the fully connected layer. In order to model the impute guider of the intermediate results $\widehat{x}_{cdd} \in R^{local} \times R^{local}$, our proposed method IGM can be summarized as the following three steps as illustrated in Fig. 3: (1) Dividing the whole GIN results into few partitions and searching candidate local region $f_{search}^{local}(\cdot)$ via Eq. (4); (2) Digging the inter-imputation relationship ε using the extracted x_{cdd} via the multilayer

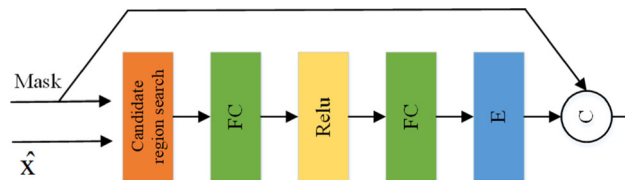


Fig. 3 The workflow of our Impute Guider Model (IGM), where E and C denote the expanding and concatenation operations, respectively

perceptron $f_{FC}(\cdot)$ in Eq. (5); (3) Fusing missing mask and intermediate results via Eq. (6).

$$x_{cdd} = f_{search}^{local}(\hat{X}, M) \tag{4}$$

$$\varepsilon = f_{FC}(x_{cdd}) \tag{5}$$

$$\widehat{x}_{cdd} = (C(E(\varepsilon), M)) \tag{6}$$

where $f_{FC}(\cdot)$ is two fully connected layers with activation function $relu$, $E(\cdot)$ represents expanding the spatial dimension of ε to that of M , and $C(\cdot)$ is the concatenation operation.

Here we provide a linear/nonlinear module after the guider, allowing the model to perform linear/nonlinear conversion. Traditional theory believes that nonlinear kernel function will increase the representation ability after conversion. We have discussed this module in theoretical research and experiment, respectively.

3.3 Local-impute-net (LIN)

Benefit from the results of IGM, the proposed LIN pays more attention to partial regions relating to the missing location. The GIN takes the full data as input to recognize global consistency of the scene, while the LIN guided by IGM focuses on a small region around the inadequate imputation area to refine the quality of more detailed appearance. Similar to Eq. (3), the adversarial local impute net loss is defined as follows:

$$L_{adv}^l = \sum_{i=1}^s L_{adv} \left(G_l^i(\widehat{x}_{cdd}); D_l^i(\hat{X}, M) \right) \tag{7}$$

Finally, the outputs of the global and the local discriminators are concatenated together into a single vector, which is then processed by a single fully connected layer, to output a continuous value. A sigmoid transfer function

[40] is used so that this value is in the [0,1] range and represents the probability that the data are real, rather than imputed.

3.4 Model training for imputation

We design and joint two training loss functions of the proposed GAGIN: the observation loss for training stability and the adversarial loss for improving the imputation performance.

The objective of imputation is to minimize the difference between the imputed values of the observed components and the real observed values. For GIN, the observation loss is given by:

$$L_{obs}^g = d(M \odot \tilde{X}, M \odot G_g(\tilde{X}, M, (1 - M) \odot Z)) \tag{8}$$

where d represents distance between the imputed data and real data.

Similarly, we obtain the observation loss for the LIN and Eq. (9) shows the whole observation loss.

$$L_{obs} = L_{obs}^g + L_{obs}^l \tag{9}$$

As mentioned above, GIN and the LIN are both adversarial networks and the missing data imputation is well trained between generators and discriminators. To obtain global consistency and finer details, we define the adversarial loss as follows:

$$\begin{aligned} \mathcal{L}_{adv} &= L_{adv}^g + L_{adv}^l \\ &= \sum_{i=1}^s L_{adv} \left(G_g^i; D_g^i(\hat{X}, M) \right) \\ &\quad + \sum_{i=1}^s L_{adv} \left(G_l^i(\widehat{x_{cdd}}); D_l^i(\hat{X}, M) \right) \end{aligned} \tag{10}$$

The overall imputation losses jointing these two functions are defined as below:

$$\mathcal{L}_{imp} = \mathcal{L}_{adv} + \mathcal{L}_{obs} \tag{11}$$

The pseudo-code is presented in Algorithm 1.

Algorithm 1 Pseudo-code of our proposed algorithm GAGIN.

while training loss has not converged **do**

(1) Global-Impute-Net optimization

Draw k_g samples from the dataset $\{(\tilde{x}(j), m(j))\}_{j=1}^{k_g}$

Draw k_g i.i.d samples, $\{z(j)\}_{j=1}^{k_g}$, of \mathbf{Z}

for $j=1, \dots, k_g$ **do**

$\bar{x}(j) \leftarrow G_g(\tilde{x}(j), m(j), z(j))$

$\hat{x}(j) \leftarrow m(j) \odot \tilde{x}(j) + (1 - m(j)) \odot \bar{x}(j)$

end for

Update D_g and G_g (for fixed D_g) using Adaptive Moment Estimation (Adam)

$$\nabla D_g - \sum_{j=1}^{k_g} \mathcal{L}_D(\tilde{x}(j), G_g(\tilde{x}(j), m(j), z(j)))$$

$$\nabla G_g - \sum_{j=1}^{k_g} \mathcal{L}_G(\tilde{x}(j), m(j), z(j))$$

(2) Local-Impute-Net optimization

Draw k_l samples from the dataset $\{(\widehat{x_{cdd}}, m_l(j))\}_{j=1}^{k_l}$

Draw k_g samples from the dataset $\{(\tilde{x}(j), m(j))\}_{j=1}^{k_g}$

Draw k_l i.i.d samples, $\{z(j)\}_{j=1}^{k_l}$, of \mathbf{Z}

for $j=1, \dots, k_l$ **do**

$\bar{x}_l(j) \leftarrow G_l(\widehat{x_{cdd}}(j), m_l(j), z(j))$

$\tilde{x}_l(j) \leftarrow m_l(j) \odot \tilde{x}(j)$

end for

Update D_l and G_l (for fixed D_l) using Adaptive Moment Estimation (Adam)

$$\nabla D_l - \sum_{j=1}^{k_l} \mathcal{L}_D(\tilde{x}_l(j), G_l(\widehat{x_{cdd}}(j), m_l(j), z(j)))$$

$$\nabla G_l - \sum_{j=1}^{k_l} \mathcal{L}_G(\widehat{x_{cdd}}(j), m_l(j), z(j))$$

end while

4 Theoretical analysis

In this section, we discuss the global-to-local problem and conditional guiding problem for GANs as the following points via theoretical analysis: 1) the imputation problem can be transformed and solved by the simulation of the global distribution from GAN; 2) introducing localization could result in a better imputation with a smaller imputation risk function, so as to prove the advantages of global-to-local imputation; 3) our designing could solve the region

distribution simulation in local GAN with the global condition.

4.1 Risk function for imputation

We first define the risk function of the overall imputation. Let the observed value be $x \in X$, the missing value be $y \in Y$, and the target value be $y_0 \in Y$. The goal of imputation is to estimate the missing value through the observed value, which means to evaluate the conditional distribution function $F(y|x)$. The objective can be expressed as the following formula with the conditional mathematical expectation function:

$$r(x) = \int y dF(y|x) \quad (12)$$

Given the functional set $f(x, \alpha)$, $\alpha \in \Lambda_\alpha$ ($f(x, \alpha) \in L_2(P)$), if the regression $r(x)$ belongs to $f(x, \alpha)$, $\alpha \in \Lambda_\alpha$, then the problem of regression function is transformed into solving the following functional. The risk function for imputation can be expressed as:

$$R(\alpha) = \int (y_0 - f(x, \alpha))^2 dF(x, y) \quad (13)$$

And the solution for such risk function is to minimize the above risk function so that

$$f(x, \alpha^*) = \arg \min_{\alpha^*} R(\alpha) \quad (14)$$

Equation (14) is a traditional regression problem [42] that can be represented and solved by the risk function on the functional space. And also be a theoretical description for our imputation task.

Via the definition of empirical risk function, solving (13) is a problem to find out α^* in Λ_α as in (14). However, in the rather complicated scenes, such as the imputation study, solving (13) by empirical risk function is not a quite good option [2] [6] [9]. The difficulty of the above problem becomes the problem of turning α^* into $F(x, y)$, which provides a theoretical basis for using GAN to solve imputation, because GAN can only calculate joint distribution [15] [41].

4.2 Estimation of $F(x, y)$ in GAN Imputation

In this part, we discuss that GAN can well simulate $F(x, y)$, and the simulated $F(x, y)$ can further solve imputation.

However, in imputation, the distribution of $F(x, y)$ is unknown, and we can only observe the empirical samples of $(y|x)$ and to estimate the $F(x, y)$. We define a density functional $p(x, y, \beta)$, where $\beta \in \Lambda_\beta$, and if β^* is found, we have

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y p(u, v, \beta^*) du dv \quad (15)$$

$$R_{prop}(\beta) = - \int \ln p(x, y, \beta) dF(x, y) \quad (16)$$

$$p(x, y, \beta^*) = \arg \min_{\beta^*} R_{prop}(\beta) \arg \min \left(- \int \ln p(x, y, \beta) dF(x, y) \right) \quad (17)$$

Based on Bretagnolle–Huber inequality, there is

$$\int |p(x, y, \beta) - p(x, y, \beta_0)| dx \leq 2\sqrt{1 - e^{R(\beta_0) - R(\beta)}} \\ s.t. R(\beta^*) - \inf_{\beta \in \Lambda_\beta} R(\beta) < \epsilon \quad (18)$$

By letting the $\beta = D \in \Lambda_D$ as a discriminator neural network in a functional space, we can obtain from Eq. (18) as follows:

$$\int |p_g(x, y, D) - p(x, y, D_0)| dx \quad (19)$$

And it coincidentally turns to be a GAN loss function as follows:

$$\int (p_g(x, y, D)) dx - \int (p_{real}(x, y, D)) dx \\ = E_{g(z) \sim p_g}(D(g(z))) - E_{(X, Y) \sim p_{real}}(D(X, Y)) \quad (20)$$

We explain below that using $p_g(x, y, D)$ to fit imputation, the fitting result is equivalent to the original fitting Eq. (13), so the problem of imputation can be solved by GAN.

Plugging Eq. (20) into Eq. (13), we find that fitting an imputation function $f(x, \alpha)$ by GAN with an estimated probability $p(x, y, D)$ is to:

$$R(\alpha) = \int (y_0 - f(x, \alpha))^2 dF(x, y) \\ = \int (y_0 - f(x, \alpha))^2 p(x, y, \beta) d(x, y) \\ = \int (y_0 - f(x, \alpha))^2 p(x, y, D_0) d(x, y) \\ = \int (y_0 - f(x, \alpha))^2 [p(x, y, D_0) - p(x, y, D) \\ + p(x, y, D)] d(x, y) \\ = \int (y_0 - f(x, \alpha))^2 [p(x, y, D_0) - p(x, y, D)] d(x, y) \\ + \int (y_0 - f(x, \alpha))^2 [p(x, y, D)] d(x, y) \quad (21)$$

The first term is the error term caused by the inaccurate distribution estimation, which can be seen from Eq. (18),

$$\int (y_0 - f(x, \alpha))^2 [p(x, y, D_0) - p(x, y, D)] d(x, y) \rightarrow 0$$

So, we obtain:

$$R(\alpha) = \int (y_0 - f(x, \alpha))^2 [p(x, y, D)] d(x, y) \tag{22}$$

We use the simulated distribution generated by GAN to fit the imputation, which is the same as the original fit. It

$$\begin{aligned} R_{all}(\alpha) &= \int R_{A_i} dA_i = \int_i L_2(y_0 - f(x, \alpha)) dF_{A_i}(x, y) dA_i \\ &= \int_i \left[\int_{i \neq j} L_2(y_0 - f(x, \alpha)) F(x, y) dA_j \right] dA_i \end{aligned} \tag{24}$$

Given the fact that L_2 is a concave function and from *Jensen Inequality*, we have

$$\begin{aligned} R_{all}(\alpha) &= \int_i \left[\int_{i \neq j} L_2(y_0 - f(x, \alpha)) dF_{A_i}(x, y) \right] dA_i \leq \int_i L_2 \left[\int_{i \neq j} (y_0 - f(x, \alpha)) dF_{A_i}(x, y) \right] dA_i \\ &= \int_i L_2 \left[\int_{all} (y_0 - f(x, \alpha)) F(x, y) dA_{all} - \int_{all} (y_0 - f(x, \alpha)) F(x, y) dA_i \right] dA_i \\ &= \int_i L_2 \left[\int_{all} (y_0 - f(x, \alpha)) F(x, y) dA_{all} \right] dA_i - \int_i L_2 \left[\int_{all} (y_0 - f(x, \alpha)) F(x, y) dA_i \right] dA_i \end{aligned} \tag{25}$$

can also solve the problem of $F(x, y)$ super-dimensional and unknown.

4.3 Advantages of localization imputation

The above conclusion shows that using GAN to simulate a global probability distribution is a reasonable method to solve the imputation. Then we further discuss that if the local simulation method is used, the final overall loss will be smaller.

For the risk function of Eq. (22), suppose we define the region we want to impute as A_i , and let the regions except A_i be $A_{j \neq i}$. s.t. $A_i \cap A_{j \neq i} = \phi$ and $\cup_i A_i = A_{all} = I$. If this area regionalized and imputation performed, the risk function of each area R_{A_i} is defined as:

$$R_{A_i}(\alpha) = \int L_2(y_0 - f(x, \alpha)) dF_{A_i}(x, y) \tag{23}$$

where L_2 is the norm-2 distance and is a concave function, $F_{A_i}(x, y)$ is a margin distribution of $F(x, y)$, $F_{A_i}(x, y) = \int_{i \neq j} F(x, y) dA_j$. Here we use dA_j to simplify $d(x, y)$, s.t. $(x, y) \in A_j$ without confusion.

Hence with a localization imputation, the entire risk function is

Given the fact that $A_i \in A_{all}$ and $L_2(\cdot) \geq 0$, we have

$$R_{all}(\alpha) \leq \int (y_0 - f(x, \alpha))^2 dF(x, y) = R(\alpha) \tag{26}$$

Equation (26) proves that when the regionalization strategy is used, the overall risk of regionalized imputation is less than or equal to the risk of global imputation.

4.4 Localization imputation in GAN with the global Information

We have discussed that localization usually has a smaller risk function. But practically, we have globally observed samples. Compared with Eq. (26) in C, the problem we need to solve is fitting $F(x, y|x_0)$, which is a simulation problem of conditional probability.

As well known, GAN is not good at conditional probability simulation (for example, the classical CGAN can only satisfy one-dimensional condition as the label [43]), while the condition of imputation design is a high-dimensional observation sample set. Next, we support our innovative design through theoretical discussions that how to solve the above difficulty fitting $F(x, y|x_0)$, we make the following design.

We define the result of global imputation net (GIN) as $\hat{y} \in Y$:

$$\hat{y} = r(x) = \int y dF(x, y | x = x_0) \quad (27)$$

\hat{y} is not the optimal solution for the fitting since there is a more minor risk function from localization. An optimization result from localization risk function could be a better result. Hence we will discuss how to get a \hat{y}_{A_i} while satisfying a condition density function of $p(y|x_{A_i}, x_{A_j \neq i})$.

For the local imputation net (LIN), taking into account global-to-local information, we let the input noise as an encoder representation of \hat{y}_{A_i} . Without loss of generality, take the encoding as an operator T , an imputation guide model (IGM) in our framework.

$$Z = f(\hat{y}_{A_i} * T) \quad (28)$$

where Z is the input noise for the local GAN. f can be linear/nonlinear function. Actually, when we use a nonlinear function as activation, the guider is closer to an encoder. Interestingly, through subsequent heuristic experiments we find a linear function has better characterization ability. We suspect this is because a linear transformation can retain more “original information” as a priori, while a nonlinear transformation will bring more disturbances. The relevant results will also be given under Sect. 5.4.

If the local generator in LIN is considered as an encoder T' , then we will find that

$$\bar{y}_{A_i} = Z * T' = \hat{y}_{A_i} * T * T' \quad (29)$$

Given the fact that

$$\hat{y}_{A_i} \sim F(y_{A_i} | x_{A_i}, x_{A_j \neq i}) \quad (30)$$

So we finally have

$$\bar{y}_{A_i} \sim F[(y_{A_i} | x_{A_i}, x_{A_j}) * T * T'] \quad (31)$$

It is a conditional probability from x_{A_i} and x_{A_j} to the localized imputation result \bar{y}_{A_i} . \bar{y}_{A_i} uses the global information x_{A_j} .

The above (31) theoretically supports how the method proposed in this paper uses the global imputation (GIN) result as a condition to generate the process of local imputation (LIN). Hence, we have successfully solved the problem of the conditional simulation in a local imputation.

In summary, we have discussed that GAN can be used to simulate $F(x, y)$ to solve the imputation problem and localized imputation can theoretically bring better solutions. Finally, the proposed method can use global information as a condition to complete the localized imputation.

5 Experiments

5.1 Datasets and experimentation details

5.1.1 Datasets

We evaluate our method both on numeric datasets and on image dataset available for missing data imputation tasks: UCI Machine Learning Repository [28] and MNIST [29]. The UCI maintains 559 datasets as a service to the machine learning community. Like [18], we select four real-world datasets (Breast, Spam, Credit, Letter) to evaluate the imputation performance quantitatively, as shown in Table 1. MNIST is a dataset of handwritten digits images of size 28×28 containing 70,000 images. We use the provided 50,000 as training set, 10,000 as validation set, and the remaining 10,000 images as testing set. Tenfold cross-validation is applied.

5.1.2 Experimentation details

For our training set, the range values of each numeric dataset and image dataset are rescaled to $[0, 1]$. We simulate that each missing value is independent of the missing rate. The dropout missing rates of our experiments are set from 10 to 80% with a step of 10%. During the training process, our GAGIN parameters are initialized by Xavier [30, 37]. The dimensional vector Z is the same as the input data. Our whole network is trained by Adam optimizer [31] with learning rate $1e-3$ and the batch size is set to 128. Then, we stop the whole learning process at 10 k iterations. We implement our model based on TensorFlow [32] framework using python 3.7 [38] and our experiments run on a Nvidia RTX 2080Ti GPU [39]. Due to demonstrating the proposed model’s performance fairly, as to all the compared methods, we implement with the same FC architecture that only fully connected layers for both the generators and the discriminators.

5.2 Evaluation metrics

To be fair to all methods, we use a unified evaluation metric to quantitatively analyze the results of all missing

Table 1 UCI Datasets used in the experiments

	No. of samples	No. of attributes
Breast	569	32
Spam	4601	57
Credit	30,000	24
letter	20,000	16

data imputation methods. Through the inspiration of the papers [16–18], we choose root mean square error (RMSE) [33, 36] and Frechet inception distance (FID) [34, 35] as the evaluation metrics, to compare the performances with the state-of-the-art imputation methods.

The RMSE between the real data and imputed data of missing data imputation is defined as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\text{real}_i - \text{imputed}_i)^2} \tag{32}$$

where N represents the number of samples.

The FID is a measure of the similarity between the generated images and real ones, which is defined as below:

$$FID(i, r) = \|\mu_i - \mu_r\|_2^2 + Tr \left(\sum_i + \sum_r - 2 \left(\sum_i \sum_r \right)^{\frac{1}{2}} \right) \tag{33}$$

where (i, r) represents the imputed image and the real image. μ_i and \sum_i are the mean-value and covariance matrix of the imputed image’s vectors, the same as μ_r and \sum_r . Tr denotes the trace of the matrix.

5.3 Performance comparison on UCI and MNIST missing dataset

In the experiments, our GAGIN is compared against 10 methods, including three state-of-the-art methods (GAIN, MisGAN, GAMIN) and seven traditional methods (zero-imputation, mean-imputation, MICE, MissForest, Matrix, EM, KNN). Fairly, as to all the compared methods, we directly use the author’s released codes or repeat according to the author’s idea to perform the evaluations. Before the starting of the experiments, we have standardized the input datasets.

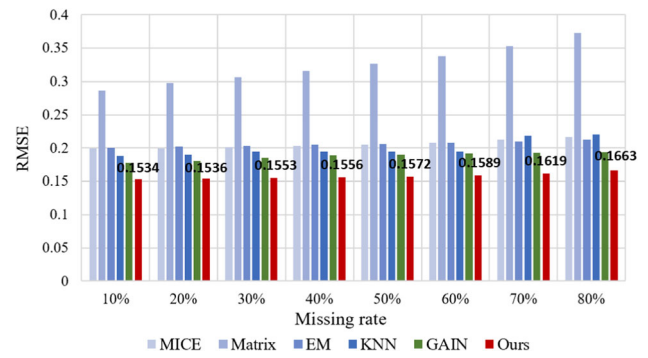


Fig. 4 Comparison of the RMSE performance by the different methods on Credit dataset for various missing rates from 10 to 80% with a step of 10%. The smaller the RMSE, the better the results

5.3.1 Evaluation on various missing rates and dimensions for UCI missing dataset

Table 2 shows the comparison results between some other imputation methods and our proposed method GAGIN in the last row. We conduct two experiments, in which we vary the missing rate and the number of dimensions on the Credit dataset. Figure 4 shows the RMSE performance for various missing rates from 10 to 80% with a step of 10%. The blue columns show the traditional methods for imputation, while the green polylines represent the GAN-based imputation methods. Even though the RMSE of each algorithm decreases as missing rates increase, our GAGIN consistently outperforms the benchmarks across the entire range of missing rates as the proposed GAGIN can capture the global information and learn the local relationship of the unobserved values guided by IGM. Besides, our method can impute the missing data with more accurate values.

We also investigate the influence of the comparison for various dimensions by the different methods as shown in Fig. 5. The RMSE performances decrease with the increasing of dimensions. The red line shows the superiority of RMSE with different dimensional datasets. We can

Table 2 Comparison of the RMSE of the different methods on UCI dataset with 20% missing rate(Average ± Std of RMSE)

	Breast	Spam	Credit	letter
MICE	0.1626 ± 0.0044	0.0691 ± 0.0007	0.2574 ± 0.0028	0.1537 ± 0.0029
MissForest	0.1667 ± 0.0067	0.0564 ± 0.0011	0.1991 ± 0.0005	0.1542 ± 0.0018
Matrix	0.2010 ± 0.0023	0.0604 ± 0.0008	0.2533 ± 0.0007	0.1448 ± 0.0004
EM	0.2286 ± 0.0054	0.0754 ± 0.0009	0.2419 ± 0.0026	0.1575 ± 0.0006
KNN	0.1383 ± 0.0045	0.0581 ± 0.0022	0.1923 ± 0.0015	0.1302 ± 0.0007
GAIN18	0.0881 ± 0.0089	0.0529 ± 0.0011	0.1801 ± 0.0018	0.1296 ± 0.0005
Ours-tanh	0.0864 ± 0.0021	0.0531 ± 0.0015	0.1797 ± 0.0019	0.1281 ± 0.0009
Ours-sigmoid	0.0838 ± 0.0032	0.0527 ± 0.0013	0.1805 ± 0.0013	0.1273 ± 0.0006
Ours-linear	0.0789 ± 0.0019	0.0500 ± 0.0008	0.1536 ± 0.0010	0.1229 ± 0.0004

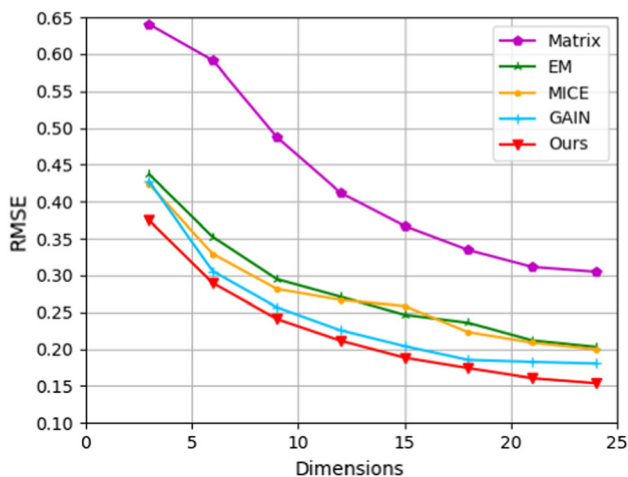


Fig. 5 Comparison of the RMSE performance by the different methods on the Credit dataset for various dimension numbers

conclude that the proposed GAGIN is also robust to the number of dimensions.

5.3.2 Quantitative evaluation on MNIST missing dataset

Our comparison of evaluation metrics is shown in Table 3, and our method’s evaluation scores are shown in the last column. For RMSE, our GAGIN surpasses the second-best method by 24.1% on MNIST dataset with 50% missing rate. Moreover, our method presents the lowest score of FID and is 33.2% lower than the second-best method. To intuitively show the comparison, we calculate the FID score of all the methods under independent dropout with missing rates from 10 to 80%. The FID results are illustrated in Fig. 6. The blue polylines show the traditional methods for imputation, while the green polylines represent the GAN-based imputation methods. It can be observed that, in all cases, our proposed method evidently gains the best imputation FID than the others.

When the missing rate is low (10%-60%), the overall distribution is retained more, and the GANs method performs well owing to simulating the global information well. However, when the missing rate increases, the global

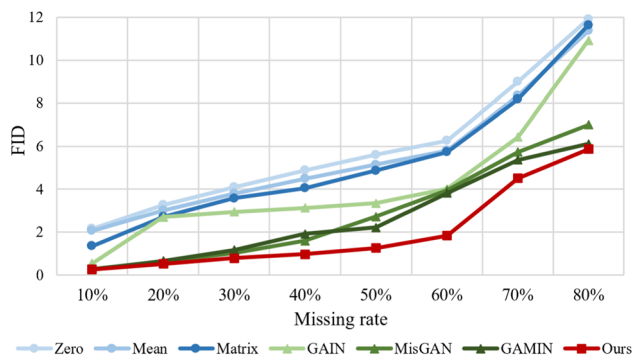


Fig. 6 Comparison of the FID score by the different methods trained on MNIST dataset for various missing rates from 10 to 80% with a step of 10%. The smaller the FID, the better the results

information becomes less, and the traditional algorithm is greatly affected at this time (the observation surrounding is more missing points). When the missing rate is greater than 60%, the RMSE rises sharply. It is undeniable that our method has the best performance, as the global-to-local method gives a better solution and optimizes the details. In addition, this problem can be derived from the subsequent test experiment. Consequently, our GAGIN outperforms all the other methods on missing data imputation.

5.3.3 Qualitative comparison on MNIST missing dataset

Figure 7 shows the visible imputation results generated from different methods on MNIST dataset of 50% missing rate. The (c) to (e) columns are the traditional methods, i.e. zero-imputation, mean-imputation and matrix-based imputation, while the (f) to (h) columns are the GAN-based methods for imputation recently as GAIN, MisGAN and GAMIN. Intuitively, the zero-imputation method and the mean-imputation method cannot produce valuable imputations. The matrix-based imputation generates the blurred images with stars. Furthermore, images imputed by GAN-based methods possess unclear and insufficient boundary in detail. It is apparent from Fig. 7 that our imputation results of the proposed method present the clearest and the most precise boundary particularly in visual performance.

Table 3 Comparison of the evaluation metrics of the different methods on MNIST dataset with 50% missing rate

Metric	Zero imputation	Mean imputation	Matrix imputation	GAIN18	MisGAN19	GAMIN20	Ours-tanh	Ours-sigmoid	Ours-linear
FID	5.6061	5.1386	4.8681	3.3416	2.1124	1.6602	1.3880	1.3714	1.2467
RMSE	0.2522	0.2402	0.2387	0.2028	0.1825	0.1805	0.1518	0.1537	0.1455

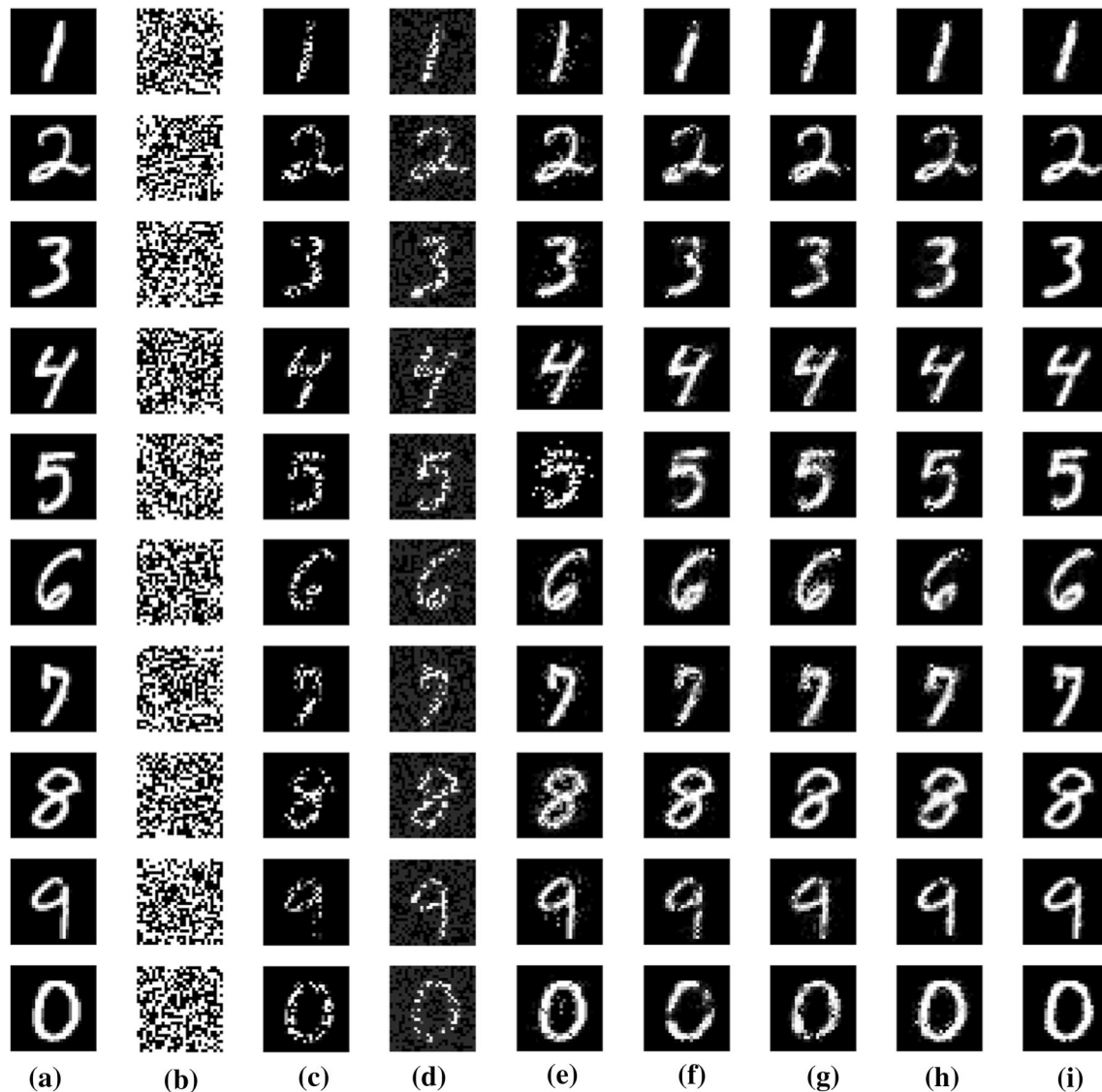


Fig. 7 Imputation results of 50% dropout missing: **a** groundtruth, **b** missing mask of which missing components are colored black, **c** impute with zero value, **d** impute with mean value, **e** impute with Matrix algorithm, **f** GAIN based imputation, **g** MisGAN based

imputation, **h** GAMIN based imputation, **i** Ours GAGIN based imputation. The **c** to **e** and **f** to **h** columns respectively show the traditional methods and GAN based methods for imputation

5.4 Ablation analysis

5.4.1 Linear or nonlinear GIN

In our framework, we use a linear/nonlinear module after IGM module. Here we heuristically choose two nonlinear functions as the kernel transform after the Guider. Compared to the linear transform, the result is listed in the last three columns of Tables 2 and 3. Interestingly, people generally consider adding nonlinear kernel could increase the representation ability of the module. However, in our study, we can see that the effect of using nonlinear functions has become worse. This is because the guide here obtains information from global random variables as a

priori to generate the local part. To avoid potential overfitting, a permutation is rational enough to introduce some uncertainty/variance from a solid global-to-local relationship. Nonlinear functions extremely disturb this situation, especially in the case of digital data, which excessively weaken the priors passed by the guider. On the other hand, this proves the effectiveness of the guider conduction prior we designed. To sum up, this part of the work is enlightening. It is not ruled out that knowing the nonlinear function that can simultaneously have better representation and more prior information is in the future research. Therefore, we still retain this module in the final model design.

Table 4 Ablation analysis of our proposed GAGIN using FID and RMSE

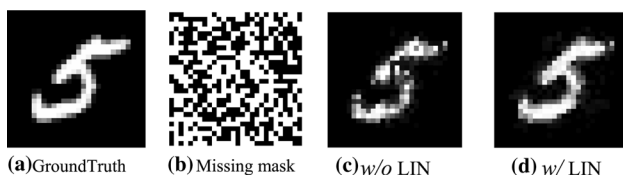
Methods	Credit	MNSIT	
	RMSE	RMSE	FID
GAGIN	0.1536	0.1455	1.2467
Effectiveness of LIN			
w/o LIN	0.1792	0.1803	1.6247
Effectiveness of IGM			
w/o \hat{X}	0.2176	0.2293	3.4219
w/o $f_{search}^{local} + w/o f_{FC}$	0.2035	0.2177	3.1907
w/o f_{search}^{local}	0.1904	0.2075	2.8585
w/o f_{FC}	0.1848	0.1987	2.2786

5.4.2 Effectiveness of local-impute-net (LIN)

As mentioned, our LIN's goal is to optimize the local boundary details and refine the inadequate imputation area. To validate the effectiveness of LIN, we get rid of the complete LIN and then directly obtain the results of IGM and GIN concatenated together as the final outputs. The RMSE with (Row GAGIN) and without LIN (denoted as "w/o LIN") are reported in Table 4. It can be observed that the proposed GAGIN with LIN works better than that without LIN. In addition, our method without LIN gains the best performance compared with other state-of-the-art approaches and traditional-based methods combining Table 3, which also highlights the effectiveness of our subnetworks (*i.e.* GIN and IGM). Intuitively, we show the visual results generated by our method without LIN and with LIN (denoted as "w/ LIN") in Fig. 8. In terms of visual results, (d) is more concrete than (c), illustrating that our LIN can refine the boundary details and effectively promote the final imputation outputs.

5.4.3 Analysis of impute guider model (IGM)

To explore the importance of IGM (*i.e.* $\hat{X}, f_{search}^{local}$ and f_{FC}), we conduct our experiments with four different instances. Considering 1) why feed \hat{X} generated by GIN into our IGM as input; 2) how to choose different candidate local region

**Fig. 8** Visual comparison of missing data imputation results without and with LIN**Table 5** Statistical t-test of our proposed GAGIN using t-value and p-value

Missing rate	t-value	p-value
10%	0.0778	0.9380
20%	0.2365	0.8130
30%	0.3403	0.7337
40%	0.4722	0.6368
50%	0.8799	0.3791
60%	2.0391	0.0416
70%	2.8428	0.0045
80%	6.3088	3.89E-10

for our impute guider model; 3) how to dig the inter-imputation relationship using the extracted region. Based on above considerations, our ablation studies about the input $\hat{X}, f_{search}^{local}$ and f_{FC} are shown in the effectiveness of IGM part in Table 4. To investigate the importance of GIN's information, we first replace the input with random noise (denoted as "w/o \hat{X} "). To verify the effectiveness of impute guider model, we get the RMSE and FID results without f_{search}^{local} and f_{FC} (denoted as "w/o $f_{search}^{local} + w/o f_{FC}$ ") by replacing these modules with the corresponding fully connected layers. As for the local region choice, the impute guide model is only equipped with f_{FC} (denoted as "w/ f_{FC} "). Then, as for the digging the relationship of imputing guider model, we only add f_{search}^{local} (denoted as "w/ f_{search}^{local} "). Compared with basic model (w/o $f_{search}^{local} + w/o f_{FC}$), our IGM can decrease FID up to 1.944 and reduce RMSE up to 0.0722. Hence, our $\hat{X}, f_{search}^{local}$ and f_{FC} significantly improve the results.

5.4.4 Statistical T test to ensure the superiority of GAGIN

T test is a common method testing two independent samples in statistics. Considering whether imputed data and real data are similar, we used the t test method to quantify the significance of the difference between the two type samples in missing data imputation. The test results with 10–80% missing rate of MNIST data set are presented in Table 5. From the table, we see that as the missing rate increases, the p value gets smaller, but the t value gets larger. When p value > 0.05 , accept the null hypothesis and consider that the difference between the two samples is not significant, and vice versa. When the missing rate is between 10 and 60%, the p value is more significant than 0.05, indicating that the generated imputed data are similar to the real data, and the lower the missing rate, the more similar. When missing rate $> 70\%$, the p value < 0.05 , and it is considered that there is a certain gap between the

generated sample and the real sample. The t test further proves the significance of our proposed method in statistics.

6 Conclusions

In this paper, we propose a novel generative adversarial guider imputation network (GAGIN) for missing data imputation. To solve the interference of local clutter and the inaccurate imputation boundary details, we design a Global-Impute-Net (GIN), a Local-Impute-Net (LIN) and an Impute Guider Model (IGM).

After the GIN generating and imputing data on the whole, the LIN is assigned to capture and refine local details guided by the IGM. Comprehensive experiments indicate our proposed method has the superiority of missing data imputation. However, we need to improve our method for all realistic settings. Future work will investigate the performance of GAGIN in other missing data types (MAR, MNAR). Furthermore, we plan to make an additional absolute guide imputation to enhance the performance of our method.

Acknowledgements This work was supported by Qian Xuesen Laboratory of Space Technology, CAST(GZZKFJ2020002), National Key Research and Development Program of China under the grant number (2018hjyzkft-002).

Author contribution Wei Wang was involved in supervision and project administration. Yimeng Chai was involved in methodology, software, and writing—original draft. Yue Li was involved in conceptualization, methodology, and writing—review & editing.

Declarations

Conflict of Interest We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, and there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled.

References

- Fortuin V, Baranchuk D, Rätsch G, et al. (2020) Gp-vae: Deep probabilistic time series imputation[C]//International Conference on artificial intelligence and statistics. PMLR, pp 1651–1661
- Yonghong Luo, Ying Zhang, Xiangrui Cai, and Xiaojie Yuan. (2019) EGAN: End-to-end generative adversarial network for multivariate time series imputation. In: 12th International joint conference on artificial intelligence IJCAI-19
- Rubanova Y, Chen R T Q, Duvenaud D. 2019 Latent odes for irregularly-sampled time series[J]. arXiv preprint [arXiv:1907.03907](https://arxiv.org/abs/1907.03907)
- Liu Y, Yu R, Zheng S, et al. Naomi 2019 Non-auto regressive multiresolution sequence imputation[J]. arXiv preprint [arXiv:1901.10946](https://arxiv.org/abs/1901.10946)
- Fedus W, Goodfellow I, Dai A M. Maskgan 2018 better text generation via filling in the_[J]. arXiv preprint [arXiv:1801.07736](https://arxiv.org/abs/1801.07736)
- Lee D, Kim J, Moon W J, et al. 2019 CollaGAN: Collaborative GAN for missing image data imputation[C] In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp 2487–2496
- Becker P, Pandya H, Gebhardt G, et al. 2019 Recurrent kalman networks: Factorized inference in high-dimensional deep feature spaces[C]//International conference on machine learning. PMLR pp 544–552
- Dalca AV, Bouman KL, Freeman WT et al (2018) Medical image imputation from image collections[J]. IEEE Trans Med Imaging 38(2):504–514
- Lee D, Moon W J, Ye J C. 2019 Which contrast does matter? towards a deep understanding of MR contrast using collaborative GAN[J]. arXiv preprint [arXiv:1905.04105](https://arxiv.org/abs/1905.04105)
- Khosravi P, Liang Y, Choi Y J, et al. 2019 What to expect of classifiers? reasoning about logistic regression with missing features[J]. arXiv preprint [arXiv:1903.01620](https://arxiv.org/abs/1903.01620)
- Cortes D. 2019 Imputing missing values with unsupervised random trees[J]. arXiv preprint [arXiv:1911.06646](https://arxiv.org/abs/1911.06646)
- Brown T B, Mann B, Ryder N, et al. 2020 Language models are few-shot learners[J]. arXiv preprint [arXiv:2005.14165](https://arxiv.org/abs/2005.14165)
- Tran K, Bisazza A, Monz C. 2016 Recurrent memory networks for language modeling[J]. arXiv preprint [arXiv:1601.01272](https://arxiv.org/abs/1601.01272)
- Zhang X, Lu L, Lapata M. 2015 Top-down tree long short-term memory networks[J]. arXiv preprint [arXiv:1511.00060](https://arxiv.org/abs/1511.00060)
- Goodfellow IJ, Pouget-Abadie J, Mirza M (2014) Generative Adversarial Networks. Adv Neural Inf Process Syst 3:2672–2680
- Seongwook Yoon, and Sanghoon Sull. 2020 GAMIN: Generative adversarial multiple imputation network for highly missing data. 2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR). IEEE
- Steven Cheng-Xian Li, Bo Jiang, and Benjamin Marlin. 2019 Misgan: Learning from incomplete data with generative adversarial networks
- Jinsung Yoon, James Jordon, and Mihaela Schaar. 2018 Gain: Missing data imputation using generative adversarial nets. In International conference on machine learning, pp 5675–5684
- Kantardzic Mehmed. 2011 Data mining: concepts, models, methods, and algorithms
- White I R, Royston P, Wood A M (2011) Multiple imputation using chained equations: issues and guidance for practice. Statistic Med 30(4):377–399
- Stekhoven DJ, Bühlmann P (2011) Missforest—non-parametric missing value imputation for mixed-type data. Bioinformatics 28(1):112–118
- Evrin Acar, Daniel M Dunlavy, and Tamara G Kolda, and Morten Mørup. 2010 Scalable tensor factorizations with missing data. In Proceedings of the 2010 SIAM international conference on data mining, pp 701–712. SIAM
- García-Laencina PJ, Sancho-Gómez J-L, Figueiras-Vidal AR (2010) Pattern classification with missing data: a review[J]. Neural Comput Appl 19(2):263–282
- Hudak A T, Crookston N L, Evans J S, Hall D E, Falkowski M J (2008) Nearest neighbor imputation of species-level, plot-scale forest structure attributes from lidar data. Remote Sens Environ 112(5):2232–2245
- Li M, Lin J, Ding Y, et al. 2020 Gan compression: Efficient architectures for interactive conditional gans[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp 5284–5294
- Shen Y, Gu J, Tang X, et al. 2020 Interpreting the latent space of gans for semantic face editing[C]//Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition. pp 9243–9252

27. Daras G, Odena A, Zhang H, et al. 2020 Your local GAN: Designing two dimensional local attention mechanisms for generative models[C]//Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition. pp 14531–14539
28. Lichman M. 2013 UCI machine learning repository. URL <http://archive.ics.uci.edu/ml>.
29. LeCun Y, and Cortes C. 2010 MNIST handwritten digit database. URL <http://yann.lecun.com/exdb/mnist/>.
30. Glorot X, Bengio Y (2010) Understanding the difficulty of training deep feedforward neural networks. *J Mach Learn Res* 9:249–256
31. Diederik P. Kingma, and Jimmy Lei Ba. 2014 Adam: A method for stochastic optimization. *Computer Science*
32. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M, et al. 2016 Tensorflow: a system for large-scale machine learning
33. Suhrud Balakrishnan and S. Chopra. 2012 Collaborative ranking. *WSDM '12*, pp 143–152. ACM
34. Heusel M, Ramsauer H, Unterthiner T, Nessler B, Hochreiter S. 2017 Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Neural Information Processing Systems*, pp 6626–6637
35. XU, Qiantong, et al. 2018 An empirical study on evaluation metrics of generative adversarial networks. arXiv preprint [arXiv:1806.07755](https://arxiv.org/abs/1806.07755)
36. Chai T, Draxler RR (2014) Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geoscientific model development* 7(3):1247–1250
37. Kumar S K. 2017 On weight initialization in deep neural networks[J]. arXiv preprint [arXiv:1704.08863](https://arxiv.org/abs/1704.08863)
38. Pajankar A (2021) Useful unix commands and tools[M]//Practical Linux with Raspberry Pi OS. Apress, Berkeley, CA, pp 81–89
39. Kumar N. 2019 Neural network implementation using CUDA[D]
40. Yin X et al (2003) A flexible sigmoid function of determinate owth. *Annal Botany* 91(3):361–371
41. Gulrajani I, Ahmed F, Arjovsky M, et al. 2017 Improved Training of Wasserstein GANs[J]. arXiv preprint [arXiv:1704.00028v3](https://arxiv.org/abs/1704.00028v3)
42. Vapnik V. 2013 The nature of statistical learning theory[M]. Springer science & business media
43. Mirza M, Osindero S. 2014 Conditional generative adversarial nets[J]. arXiv preprint [arXiv:1411.1784](https://arxiv.org/abs/1411.1784)
44. Liu Y, Gopalakrishnan V (2017) An overview and evaluation of recent machine learning imputation methods using cardiac imaging data. *Data* 2(1):8
45. Jerez José M et al (2010) Missing data imputation using statistical and machine learning methods in a real breast cancer problem. *Artifi Intell Med* 50(2):105–115
46. Li L, Fu H, Xu X (2021) Active learning with sampling by joint global-local uncertainty for salient object detection. *Neural Comput Applic*. <https://doi.org/10.1007/s00521-021-06395-8>
47. Ma X, Li X, Zhou Y et al (2021) Image smoothing based on global sparsity decomposition and a variable parameter. *Comp Visual Media* 7:483–497
48. Wang Q, Hu X, Gao Q et al (2014) Global–local fisher discriminant approach for face recognition. *Neural Comput Applic* 25:1137–1144
49. Cheng Y, Song F, Qian K (2021) Missing multi-label learning with non-equilibrium based on two-level autoencoder. *Appl Intell* 51:6997–7015
50. Raja PS, Sasirekha K, Thangavel K (2020) A novel fuzzy rough clustering parameter-based missing value imputation. *Neural Comput Applic* 32:10033–10050

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.