Reconstruction of Missing Samples in Antepartum and Intrapartum FHR Measurements Via Mini-Batch-Based Minimized Sparse Dictionary Learning

Yefei Zhang [iD], Zhidong Zhao [iD], *Member, IEEE*, Yanjun Deng [iD], Xiaohong Zhang [iD], and Yu Zhang [iD]

Abstract—**Fetal Heart Rate (FHR), an important recording in Cardiotocography (CTG)-based fetal health status monitoring, is the only information that clinical obstetricians can directly obtain and use. A challenge, however, is that missing samples are very common in FHR due to various causes such as fetal movements and sensor malfunctions. The aim is the development of an inpainting tool which is suitable for different missing lengths q and various total missing percentages Q, as well as for use in online mode. This study focused on two major impediments to existing inpainting methods: the longer the missing length, the more difficult it is to recover with mathematical methods; the reliance on tens of thousands of training samples, and the computational burden caused by full batch-based dictionary learning algorithms. We present a regularized minimization approach to signal recovery, which combines a $L_{0.6}-\text{norm}$ minimized sparse dictionary learning algorithm (MSDL) and a model optimization strategy for using a mini-batch version for signal recovery. Using 100 FHR recordings with 2 protocols designed to simulate missing clinical data scenarios, the combined method performed favorably in terms of 5 data analysis metrics and 3 clinical indicators. Comparing 4 inpainting methods, we were able to prove the superiority of the proposed algorithm for both large q and large Q. The experimental results showed the lowest values (2.64 (MAE), 4.68 (RMSE)) when $Q = 5\%$ with short interval lengths. The developed architecture provides a reference value for the practical application of recovering missing samples online.**

Index Terms—**Missing sample, reconstruction of FHR recording, sparse dictionary learning, mini-batch version.**

I. INTRODUCTION

FETAL distress and hypoxia are the main causes of adverse events such as neonatal asphyxia and disability [1], which

often occur in late pregnancy and in labor. It is crucial, therefore, to monitor fetal safety in the womb in late pregnancy. Cardiotocography (CTG) is a widely used and non-invasive prenatal diagnosis technique for clinical evaluation of fetal condition in the maternal uterus. A typical CTG recording consists of two simultaneously acquired signals, namely Fetal Heart Rate (FHR) and Uterine Contraction (UC), in which FHR is usually obtained by Doppler ultrasound probe or fetal scalp electrode. The former is a non-invasive prenatal diagnosis technique, while the latter is an invasive technique. On the basis of the guidelines provided by the International Federation of Gynecology and Obstetrics (FIGO) [2], FHR signals can provide valuable information about fetal homeostasis during the critical period of late pregnancy and labor, and are the only information directly available to clinicians to make a professional diagnosis by naked-eye inspection. Therefore, analysis of FHR recordings is of crucial importance. A very challenging objective is development of artificial intelligence auxiliary diagnosis (AIAD) tools for the extraction of meaningful features from FHR recordings that could be reliably used to point out possible fetal and neonatal pathologic conditions.

AIAD tools such as information fusion-based online heart monitoring systems [3], smart health systems for ambulatory maternal and fetal monitoring [4] and smart supervision of cardiomyopathy [5], the ambiguity of which depends in part on the quality of the measured data. Unfortunately, artifacts can interfere with FHR recording due to the gravida's heartbeat, and suffer from frequent invalid or missing samples due to fetal or maternal movements, sensor malfunctions, or misplaced electrodes. This is particularly critical during the labor stage since fetal condition may suddenly change. Labor stage represents a very stressful period; FHR is the sole measurement acquired directly from the fetus during delivery, and there is no way of repeating acquisition as is usually done in the antepartum stage. Therefore, any distortion leading to a corrupted signal means difficulties for fetal surveillance.

Usually, for internal direct measurements of FHR, the percentage of missing samples ranges from 0-10%, and for external ultrasound measurements, the percentage varies from 0-40% [6]. We should note that there are still no guidelines on how large a percentage of missing samples will disqualify an FHR recording from automatic analysis or visual inspection, although
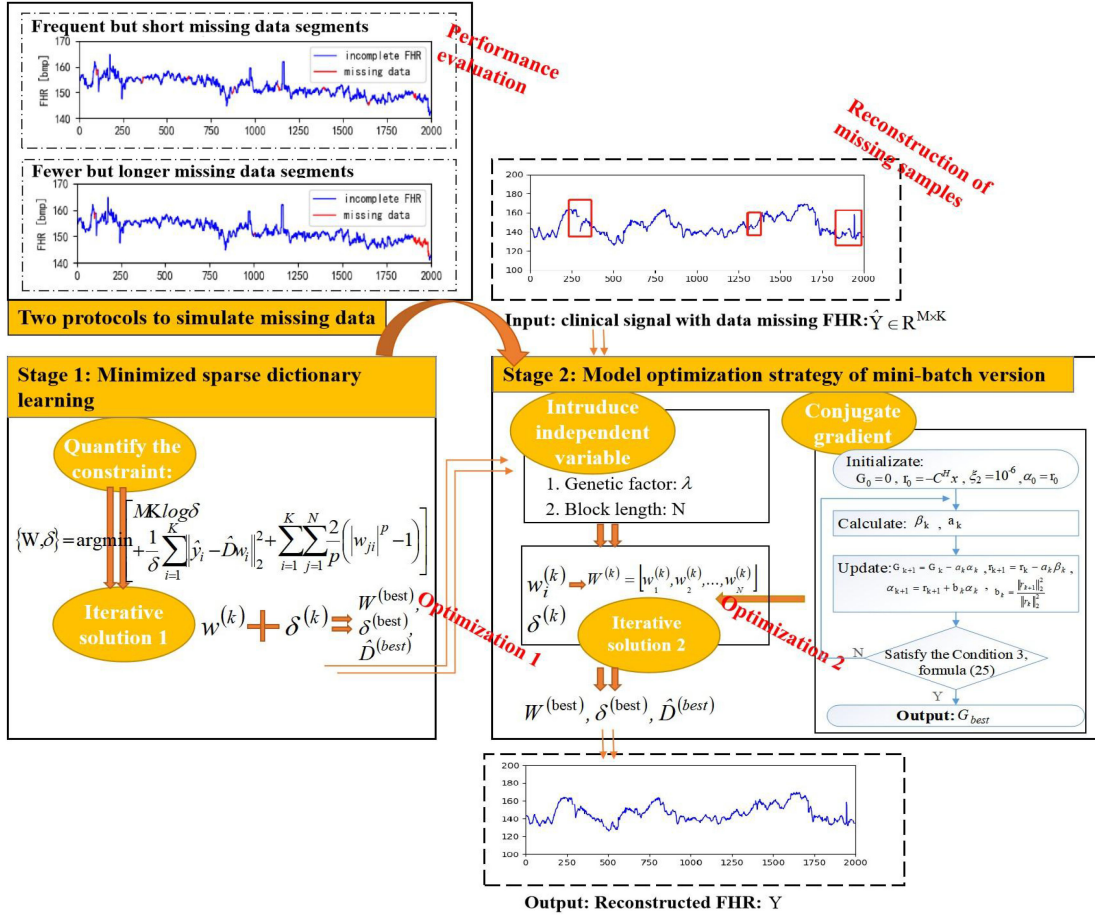
Fig. 1. Architecture of the mini-batch-based minimized sparse dictionary learning approach.

an empirical value given by clinicians for visual inspection, is 50% [7]. Clinicians can tolerate such a high percentage of missing samples in FHR, because: 1) their inspections, unlike those of AIAD systems, are mainly focused on morphological features, and their rich clinical experience provides a large number of empirical samples for morphology-based diagnosis; and 2) human visual perception is far less sensitive to missing samples than AIAD systems' quantitative / qualitative assisted diagnosis. Currently, the development premise of most AIAD tools is to provide a complete data set [8]. Therefore, the reconstruction of FHR values for missing samples has a **twofold purpose**: 1) to improve trace readability when used for naked-eye inspection; and 2) to provide a valuable and complete sampled signal that could be used for AIAD.

As an initial inpainting method, researchers such as SK Lee et al. [9] and M.Cesarelli et al. [17] have manually reconstructed missing FHR data based on linear/non-linear mathematical models. For example, SK Lee et al. applied a first-order autoregressive conditional heteroscedasticity (ARCH (1)) in 2019; and ZD Zhao et al. (2019) [15] integrated spline interpolation and cubic spline interpolation to to recover FHR data. However, these mathematical inpainting methods are not suitable for FHR with long missing samples. As more and longer missing data are expected when using external measurements devices such as wearable monitors, better estimations are desired.

In addition, to the best of our knowledge, there is no reported study adopting the relevant techniques of deep learning to recover missing FHR recordings. But in some other application scenarios such as image processing and geo-statistics, recovery of missing data has been extensively explored [10]–[14]. These deep learning algorithms can effectively reconstruct regular missing data and minimize the impact of artifacts. However, they rely heavily on learning and training with a large amount of sample data, and cannot be adapted for timely data retrieval because of the time required to process the data.

In this work, we propose an effective mini-batch-based minimized sparse dictionary learning (MSDL) approach (called miniMSDL), to rebuild the FHR values for missing samples, with architecture as shown in Fig. 1. The main contributions and novelty of this work can be summarized as follows:

1) The longer the missing length, the more difficult it is to recover. In addition, the location of missing data is highly random, so adaptive learning of missing information is particularly important. This paper introduces a sparse learning-based dictionary construction and optimization method. Compared with the mathematical methods that rely on a few specified formulas or a sample system [7]–[9], [15]–[17], the right choice of an adaptive learned dictionary can lead to better representation of the signal, i.e., a more sparse representation with less residual.

2) This paper focuses on addressing the problem of online mode, i.e timely data recovery for both antepartum and intrapartum FHR measurements. On the one hand, most previous approaches are actually designed for an offline mode since the reconstruction relies on tens of thousands of training samples and is time-consuming and sometimes quite repetitive [10]–[12]. But on the other, traditional dictionary learning is usually computationally expensive to train as well as to use [18],[19], especially for an inpainting problem with long missing sample length. To reduce the training load, we adopted a model optimization strategy of the mini-batch version, and simultaneously used a conjugate gradient to solve the complex dictionary and sparse coefficient matrix. This two-step operation reduced the number of free variables and alleviated the computational burden, which clears the way for greater clinical / practical application in most online AIAD tools.

The proposed method not only achieves state-of-the-art results on different lengths of missing data, but also provides complete data for research on FHR-based AIAD systems, and makes a modest contribution to the biomedical signal processing knowledge base.

II. RECONSTRUCTION MODEL OF MISSING FHR SAMPLES

Sparse dictionary learning, also known as sparse representation and dictionary learning, is based on the idea of sparsely representing a signal in a certain domain where the learned dictionary represents this domain. Therefore, the algorithm consists of two stages: dictionary generation and sparse coding with a precomputed dictionary.

With $y = [\alpha_1, \alpha_2, \ldots, \alpha_L] \in R^{L \times 1}$ being a 1D signal of length L, it can be approximately expressed as $y = Dw$. Where $D \in R^{L \times N}$ is a dictionary, its columns represent N feature atoms (thus the block length is N), and $w \in R^{N \times 1}$ is a sparse coefficient vector. In the case of $L << N$, $y = Dw$ becomes an underdetermined equation system and the current D is an overcomplete dictionary.

With regard to the reconstruction of missing FHR samples, a similar relationship exists:

$$\hat{y} = Ry \qquad (1)$$

where $\hat{y} \in R^{L \times 1}$ is an incomplete FHR signal of a pregnant woman and $y \in R^{L \times 1}$ corresponds to her complete FHR signal. $R \in R^{L \times L}$ is a random sampling diagonal matrix which represents the position of data missing, $r_{ii} = 0$: data missing of the current position while $r_{ii} = 1$ is data integrity, i.e., if $\hat{y} = [a_1, 0, 0, 0, a_5, \ldots, a_L]^T$, $R = diag(1, 0, 0, 0, 1, \ldots, 1)$, in which $diag\{\cdot\}$ means diagonalization.

Due to the influence of maternal/fetal displacement and probe movement, which create a certain amount of noise and error, formula (1) can be rewritten as follows:

$$\hat{y} = Ry + \varepsilon \qquad (2)$$

in which $\varepsilon \in R^{L \times 1}$ is a noise term. Formula (2) is aimed at reconstructing y based on $\hat{y}$. Combined with $y = Dw$, formula (2) can be expressed as follows:

$$\hat{y} = RDw + \varepsilon = \hat{D}w + \varepsilon, \ \hat{D} = RD \qquad (3)$$

Since R and $\hat{y}$ are known to us, the key to to reconstructing a FHR recording y lies in the solution of $\hat{D}$ and w. It is, obviously, a standard sparse coding problem [20], and can be transformed into:

$$\{\hat{D}, w\} = \underset{\hat{D}, w}{\operatorname{argmin}} \left\| \hat{y}_i - \hat{D}w \right\|_2^2 \ , \quad s.t. \ w \ is \ \text{sparse} \qquad (4)$$

In which w is sparse and can be expressed as $\|w\|_F \leq \delta$, where $\| \cdot \|_F$ is the regularization term F-norm. δ is a regularization term error which is related to the noise term ε and is not known a priori in most applications.

We transformed the reconstruction of missing FHR samples into a sparse coding problem because of: 1) the advantage of dictionary learning. One can represent a large sample with random missing positions by dictionary learning, which is a dimension reduction representation. By learning the key features of the missing location, dictionary learning makes the learned dictionary adaptive to each pregnant woman. 2) the value of sparse representation, which expresses as much information as possible with fewer resources; this can accelerate the calculation speed of the online mode.

If a pregnant woman collects multiple sets of missing FHR signals in fetal monitoring, that is, the incomplete FHR signal is $\hat{Y} = [\hat{y}_1, \ldots, \hat{y}_K] \in R^{M \times K}$, then the reconstruction problem can be expressed as follows:

$$\hat{Y} = [\hat{y}_1, \ldots, \hat{y}_K] = [R_1 D_1 w_1 + \varepsilon, \ldots, R_K D_K w_K + \varepsilon] \qquad (5)$$

Let $\overline{R} = [R_1, \ldots, R_K] \in R^{M \times MK}$, $\overline{D} = [D_1, \ldots, D_K]^T \in R^{MK \times N}$ and $W = [w_1, w_2, \ldots, w_K]$, and formula (5) can be determined as follows:

$$\hat{Y} = \overline{R}\overline{D}W + \varepsilon \qquad (6)$$

Similarly, the key to reconstructing $Y = [y_1, \ldots, y_K] \in R^{M \times K}$ lies in the solution of $\overline{D}$ and W, and the above formula (6) can be transformed into:

$$\{\overline{D}, W\} = \underset{\overline{D}, W}{\operatorname{argmin}} \left\| \hat{Y} - \overline{R}\overline{D}W \right\|_2^2 \ , \quad s.t. \ w_i \ is \ \text{sparse} \qquad (7)$$

Using equation (4) to resolve formula (7), we obtain:

$$\{\hat{D}, W\} = \underset{\hat{D}, W}{\operatorname{argmin}} \sum_{i=1}^{K} \left\| \hat{y}_i - \hat{D}w_i \right\|_2^2, \ s.t. \ \|w_i\|_F < \delta \qquad (8)$$

Overall, the most important and effective constraint imposed on W is L_0-norm, as the following formula (9) shows. It is, however, a nondeterministic polynomial (NP) problem, and rarely applied in practice. Further, though both L_1- and L_2-norm can reduce the risk of overfitting, the former can more easily obtain a "sparse" solution. That is, the sparse coefficient matrix W obtained by formula (10) will have more zero components. Formula (10) can be calculated by iterative convergence or gradient descent, but its shortcoming lies in its large computational burden and high complexity.

$$\{\hat{D}, W\} = \underset{\hat{D}, W}{\operatorname{argmin}} \sum_{i=1}^{K} \left\| \hat{y}_i - \hat{D}w_i \right\|_2^2, \ s.t. \ \|w_i\|_0 < \delta \qquad (9)$$

$$\{\hat{D}, W\} = \underset{\hat{D}, W}{\arg\min} \sum_{i=1}^{K} \left\| \hat{y}_i - \hat{D} w_i \right\|_2^2, \quad s.t. \ \|w_i\|_1 < \delta \quad (10)$$

Further analysis shows that replacing the L_1-norm with the L_p-norm $(0 < p < 1)$ can produce a much more accurate reconstruction of sparse signals under the same acquisition conditions. The theoretical analysis leading to this conclusion has been supported by detailed theoretical research and experiments in the literature [21]. Therefore, we used the L_p-norm to reconstruct the sparse coding problem, and obtained the objective function as follows:

$$\{\hat{D}, W\} = \underset{\hat{D}, W}{\arg\min} \sum_{i=1}^{K} \|\hat{y}_i - \hat{D} w_i\|_2^2, \quad s.t. \ \|w_i\|_p < \delta, \ 0 < p < 1 \quad (11)$$

III. MINI-BATCH-BASED MINIMIZED SPARSE DICTIONARY LEARNING FOR DATA RECOVERY

A. Optimization to Objective Function: MSDL

Analyzing the objective function in formula (11), $\hat{D}$ and W are free variables, and neither L_p-norm nor the regularization term error δ is a prior parameter, which makes it difficult to choose in practice. We developed an L_p-norm-based minimized sparse dictionary learning algorithm with a mini-batch version and combined it with a conjugate gradient solution for data recovery and noise term estimation. To determine L_p-norm automatically, the bayesian information criterion (BIC) is adopted.

1) *First, quantify the constraints of formula (11)*, and obtain the minimized calculation formula below:

$$\{W, \delta\} = \arg\min \left[MK \log \delta + \frac{1}{\delta} \sum_{i=1}^{K} \left\| \hat{y}_i - \hat{D} w_i \right\|_2^2 + \sum_{i=1}^{K} \sum_{j=1}^{N} \frac{2}{p} \left(|w_{ji}|^p - 1 \right) \right] \quad (12)$$

in which w_i is the ith sparse coefficient vector of sparse coefficient matrix W and w_{ji} is the value of W in the row j and column i. The first part $MK \log \delta$ is a fitting degree of regularization term error, the second part $\frac{1}{\delta} \sum_{i=1}^{K} \|\hat{y}_i - \hat{D} w_i\|_2^2$ is a fitting degree of a noise term, and thus these two parts can be collectively defined as a fitting term. The third part $\sum_{i=1}^{K} \sum_{j=1}^{N} \frac{2}{p} (|w_{ji}|^p - 1)$ is a penalty term: if $p = 1$, it can be resolved as $2(\sum_{i=1}^{K} \|w_i\|_1 - KN)$, which is similar to L_1-norm; If instead $p \to 0$, it does not show similarity to L_0-norm and can be resolved as $2 \sum_{i=1}^{K} \sum_{j=1}^{N} \log|w_{ji}|$.

2) *Second, find an iterative solution of the independent variable W:* Let $w_i^{(k)}$ be the update result of the kth iteration of the ith sparse coefficient vector and $\delta^{(k)}$ be the regularization term error of the kth iteration. If one marks formula (12) as $f(W, \delta)$, it can also be expressed as follows:

$$f(W, \delta) = \sum_{i=1}^{K} f(w_i, \delta)$$

$$= \sum_{i=1}^{K} \left(M \log \delta + \frac{1}{\delta} \left\| \hat{y}_i - \hat{D} w_i \right\|_2^2 + \sum_{j=1}^{N} \frac{2}{p} \left(|w_{ji}|^p - 1 \right) \right) \quad (13)$$

Calculate the partial derivative of $f(w_i, \delta^{(k)})$ with respect to w_i^{H}, in which $(\cdot)^{H}$ represents a conjugate transpose operation, and set it equal to 0, to get:

$$\frac{1}{\delta^{(k)}} \hat{D}^{H} \hat{D} w_i - \frac{1}{\delta^{(k)}} \hat{D}^{H} \hat{y}_i + P^{-1} w_i = 0,$$

$$P = \text{diag} \left\{ \left[|w_{1i}|^{2-p}, |w_{2i}|^{2-p}, \ldots, |w_{Ni}|^{2-p} \right]^{T} \right\} \quad (14)$$

Since P is a nonlinear function of $w_i^{(k)}$, it is hard to solve the iterative update expression $w_i^{(k)}$ from (14). Therefore, we use a heuristic method here. Let $P \cong P^{(k)}$, i.e., $P^{(k)} = \text{diag}\{[|w_{1i}^{(k)}|^{2-p}, |w_{2i}^{(k)}|^{2-p}, \ldots, |w_{Ni}^{(k)}|^{2-p}]^{T}\}$, to get:

$$w_i = \left(\hat{D}^{H} \hat{D} + \delta^{(k)} P^{-1} \right)^{-1} \hat{D}^{H} \hat{y}_i \quad (15)$$

In conclusion, we can obtain $w_i^{(k+1)}$ of the ith sparse coefficient vector at the (k+1)th iteration:

$$w_i^{(k+1)} = \left(\hat{D}^{H} \hat{D} + \delta^{(k)} \left(P^{(k)} \right)^{-1} \right)^{-1} \hat{D}^{H} \hat{y}_i$$

$$= P^{(k)} \hat{D}^{H} \left(\hat{D}^{H} P^{(k)} \hat{D} + \delta^{(k)} I \right)^{-1} \hat{y}_i \quad (16)$$

Thus, the sparse coefficient matrix is determined: $W^{(k+1)} = [w_1^{(k+1)}, w_2^{(k+1)}, \ldots, w_K^{(k+1)}]$

3) *Finally, find an iterative solution of the independent variable δ:* δ is a regularization term error in formula (11), which is the quantization formula of noise term ε. Calculate the partial derivative of $f(w_i^{(k)}, \delta)$ with respect to δ, and set it equal to 0, to get:

$$\delta = \frac{\sum_{i=1}^{K} \left\| \hat{y} - \hat{D} w_i^{(k)} \right\|_2^2}{M} \quad (17)$$

Then, obtain the iterative update formula $\delta^{(k+1)}$ at the (k+1)th iteration:

$$\delta^{(k+1)} = \frac{\sum_{i=1}^{K} \left\| \hat{y} - \hat{D} w_i^{(k+1)} \right\|_2^2}{M} \quad (18)$$

Based on the above analysis, we can update the sparse vector matrix W and the regularization term error δ with formulas (16) and (18), so as to further realize the iterative solution of the reconstruction model.

Algorithm 1: Full-batch-based MSDL
1   **Input data:** $\hat{y}$ (this is the same as $\hat{Y} \in \mathrm{R}^{M \times K}, K = 1$)
2   **Initialization:** set $p = 0.6$, $\xi_1 = 10^{-3}$; R : Random sampling matrix which measures the position of missing data; D: Fourier transform matrix; $\hat{D} = RD$; $w^{(0)} \cong \hat{D}^{-1}\hat{y} = \hat{D}\hat{y} \big/ \lVert \hat{D} \rVert_2^2$, $\delta^{(0)} = \lVert \hat{y} - \hat{D}w^{(0)} \rVert_2^2 \big/ \mathrm{L}$
3   **While Condition 1:**
4   **For k in range(n_iter):**   % n_iter: number of iterations
5   **Sparse coding step:**
6   Find $w^{(k)}$, using formula (16)
7   **Dictionary update step:**
8   Find $\delta^{(k)}$, using formula (18).
9   Update the dictionary $\hat{D}$ via SVD.
10   **Output:** $w^{(\text{best})}, \delta^{(\text{best})}, \hat{D}^{(best)}$
11   **Calculate:** $\mathrm{D} = R^{-1}\hat{D}^{(best)}$, $y = \mathrm{D}w^{(\text{best})}$

Note: since $\hat{y} \in \mathrm{R}^{L \times 1}$, $i = 1$ ($w_i^{(k)}$) and we simplified it to $w^{(k)}$; n_iter will be set and introduced in Section IV.B; the changing trend of the BIC for various $\mathrm{L}_p - \mathrm{norm}$ is in Fig. 5(c).

B. Optimization Solution: Model Optimization Strategy of Mini-Batch Version

Using formulas (16) and (18) derived from MSDL, we can perform sparse coding and dictionary learning with a greedy method such as matching pursuit [22] along with its optimization, called least absolute shrinkage and selection operator (LASSO) [23]. Since these methods are iteratively updated based on all training samples, they are also called batch-based solutions. For example, if the incomplete FHR signal of a pregnant woman is $\hat{y} \in \mathrm{R}^{L \times 1}$, i.e., the length of the training FHR sample is L, full batch-based MSDL can be performed using Algorithm 1 (Table I), where Condition 1 represents the following inequality:

$$\frac{\lVert w^{(k)} - w^{(k-1)} \rVert_2}{\lVert w^{(k)} \rVert_2} \leq \xi_1 \tag{19}$$

which measures the difference degree between $w^{(k)}$ and $w^{(k-1)}$. The smaller the value, the smaller the difference degree between these two iterations. ξ_1 is a small positive number.

As seen in algorithm 1, the sparse coefficient matrix and dictionary depend on all the previous information in each iteration, and thus all training samples must be processed in one iteration. In this way, most of the time will be spent on gradient calculation, which is very time-consuming. Specifically, if L training samples are iterated a total of n_iter times, the calculation amount is $L \times$ n_iter, which is relatively large. Therefore, we further expanded the model optimization strategy of a mini-batch version to MSDL (hence the name mini-batch-based MSDL).

Let us take as an example the training set $\hat{y} \in \mathrm{R}^{L \times 1}$, in which L training samples are divided into K parts (that is, K subsets), each with $^L/_K$ samples. We will call each subset a mini batch, and the subset size "batch size". Then traverses each mini batch, calculates its gradient, and updates the information. In this way, traversing all mini batches is just like performing K gradient descents. Thus, Condition 1 can be implemented with a smaller n_iter. Mini-batches decline faster than the traditional

full-batch-based solution, which improves the memory utilization and parallelization efficiency of the algorithm. Marking $^L/_K$ as M, the training set becomes $\hat{Y} \in \mathrm{R}^{M \times K}$, with a total of K mini batches $\hat{y} \in \mathrm{R}^{M \times 1}$; and each $\hat{y} \in \mathrm{R}^{M \times 1}$ should be trained in sequence.

In this study, we implemented such a variant with a genetic factor λ, thus obtaining the iterative updating formula of $w_i^{(k+1)}$ and $\delta^{(k+1)}$ as in the follow equations (20)-(21). The old data from the previous iteration is partially forgotten.

$$w_i^{(k+1)} = \left(P^{(k)}\hat{D}^{\mathrm{H}} \left(\hat{D}^{\mathrm{H}} P^{(k)}\hat{D} + \delta^{(k)}I \right)^{-1} \hat{y}_i \right)_{\text{new}} + \lambda w_i^{(k)} \tag{20}$$

$$\delta^{(k+1)} = \left(\sum_{i=1}^{} \lVert \hat{y} - \hat{D}w_i^{(k+1)} \rVert_2^2 \big/ \mathrm{M} \right)_{new} + \lambda \delta^{(k)} \tag{21}$$

The miniMSDL can be seen in Algorithm 2 (Table II) where Condition 2 satisfies the constraints shown in equation (22), which is a variant of formula (19) and calculates the difference degree of W in the iterative update process. Using a mini-batch solution without adapting the entire data set would mean using the learned dictionary from the previous iteration as the initial dictionary for the current iteration. Doing so would make the MSDL capable of processing large amounts of data and also suitable for online mode.

$$\left(\sum_{i=1}^{} \lVert w_i^{(k)} - w_i^{(k-1)} \rVert_2 \right) \Big/ \left(\sum_{i=1}^{} \lVert w_i^{(k)} \rVert_2 \right) \leq \xi_1 \tag{22}$$

C. Optimization Solution: Conjugate Gradient Solution

When applying the mini-batch-based MSDL algorithm to reconstruct the missing FHR samples, the iterative update of the sparse coefficient matrix W involves calculation of

TABLE II
MINI-BATCH-BASED MSDL ALGORITHM

Algorithm 2: Mini-batch-based MSDL.
1   **Input data:** $\hat{Y}$ ($\hat{Y} \in R^{M \times K}$)
2   **Initialization:** set $p = 0.6$, $\xi_1 = 10^{-3}$, $\lambda = 0.6$, D: Fourier transform matrix, $\hat{D} = RD$
$$w_i^{(0)} \cong \hat{D}^{-1}\hat{y} = \hat{D}(:,i)\hat{y} \big/ \left\|\hat{D}(:,i)\right\|_2^2 , \quad W^{(0)} = \left[w_1^{(0)}, w_2^{(0)}, ..., w_K^{(0)}\right], \quad \delta^{(0)} = \sum_{i=1}^{N}\left\|\hat{y} - \hat{D}w_i^{(0)}\right\|_2^2 \Big/ M$$
3   **While Condition 2:**
4   **For i in range(n_iter):**   % n_iter: number of iterations
5   **For k in range(K):**   % K: number of mini-batches to complete training once
6   Get the new set of training set $\hat{y}_i$, and R_i
7   **Sparse coding step:**
8   Find $w_i^{(k)}$, using formula (20)
9   **Dictionary update step:**
10   Find $\delta^{(k)}$, using formula (21).
11   Update the dictionary $\hat{D}$ via SVD.
12   Calculate: $D_i = R_i^{-1}\hat{D}$
13 $W^{(k)} = \left[w_1^{(k)}, w_2^{(k)}, ..., w_K^{(k)}\right], \quad \overline{D} = \left[D_1, ..., D_K\right]^T$
14   **Output:** $W^{(best)}, \delta^{(best)}, \overline{D}^{(best)}$
15   **Calculate:** $y_i = D_i^{(best)}w_i^{(best)}$, $Y = \left[y_1, ..., y_K\right]$

Note: the changing trend of test time for various λ is in Fig. 5(d); training time for various K is in Fig. 5(e).

the inverse of the matrix. Specifically, the following part of $(\hat{D}^H P^{(k)}\hat{D} + \delta^{(k)}I)^{-1}$ equation (20) has a high computational complexity. Therefore, we converted it to a least squares problem solved by a conjugate gradient, thus reducing the computational complexity of the sparse coding update process.

Conjugate gradient is a nonlinear optimization algorithm that falls between steepest descent and the Newton method. Theoretically, we can obtain an optimal solution with only few steps, which is very suitable for the iterative solution of a sparse matrix.

1) First, let $G = (\hat{D}^H P\hat{D} + \delta I)^{-1}\hat{y}$, rewritten as follows:

$$G = \left(C^H C\right)^{-1} C^H x, \; in \; which \; C = \begin{bmatrix} P^{\frac{1}{2}}\hat{D}^H \\ \delta^{\frac{1}{2}}I \end{bmatrix},$$

$$x = \begin{bmatrix} 0 \\ \delta^{-\frac{1}{2}}\hat{y} \end{bmatrix} \tag{23}$$

2) Second, analyze the above formula (23) and transform it into the least squares problem:

$$\min_{G} \left\|CG - x\right\|_2^2 \tag{24}$$

3) Finally, perform the above formula (24) in Algorithm 3 (Table III), where Condition 3 satisfies the following inequality:

$$\frac{\left\|r_{k+1}\right\|_2^2}{\left\|\hat{y}\right\|_2^2} \leq \xi_2 \tag{25}$$

Based on the above analysis, this study realizes the reconstruction of FHR values for missing samples, using miniMSDL

TABLE III
DETAILED STEPS OF CONJUGATE GRADIENT FOR LEAST SQUARE PROBLEM
(PROVIDE COMPUTING SERVICES FOR MINIMSDL)

Algorithm 3: Conjugate gradient-based least square problem
1   **Input data:** Iteration initial value $G_0 = 0$, residual $r_0 = -C^H x$
2   **Initialization:** set $\xi_2 = 10^{-6}$, $\alpha_0 = r_0$
3   **While Condition 3:**
4   **Calculate:** $\beta_k = C^H C\alpha$
5 $a_k = \dfrac{\left\|r_k\right\|_2^2}{\alpha_k^H \beta_k}$
6   **Update:** $G_{k+1} = G_k - a_k\alpha_k$
7 $r_{k+1} = r_k - a_k\beta_k$
8 $\alpha_{k+1} = r_{k+1} + b_k\alpha_k$, in which $b_k = \dfrac{\left\|r_{k+1}\right\|_2^2}{\left\|r_k\right\|_2^2}$
9   **Output data:** Solution of equation G_{best}

and the conjugate gradient solution. The detailed and complete process is shown in Fig. 2.

IV. EXPERIMENTS AND RESULTS

A. Data Description

We carried out a large number of experiments using the Czech Technical University-University Hospital in Brno (CTU-UHB) Intrapartum Cardiotocography Database [24], [25]. This
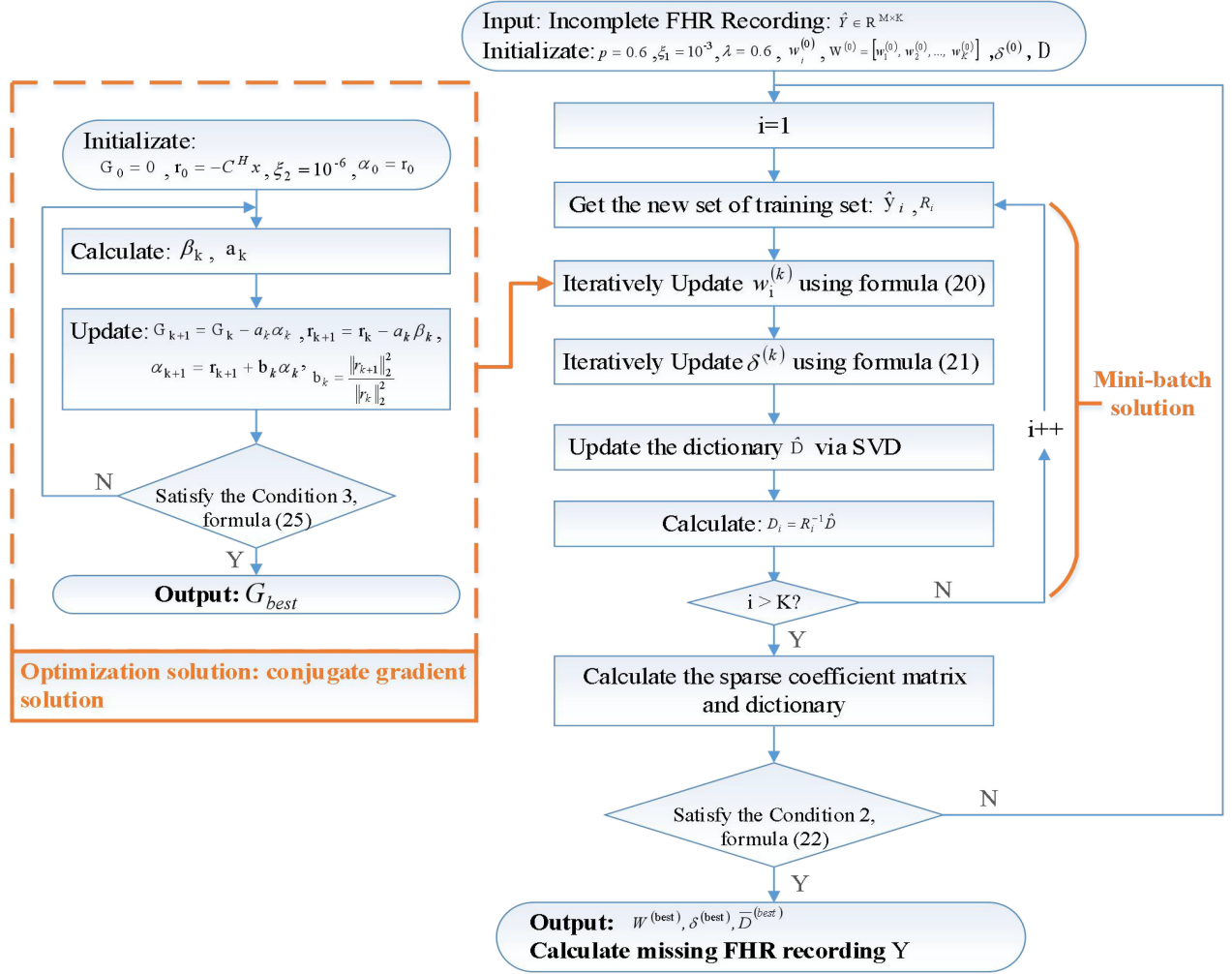
Fig. 2.   Proposed reconstruction model of missing FHR samples, which combines conjugate gradient with miniMSDL.

database, from the obstetrics ward of the UHB, Czech Republic, contains 552 CTG recordings, which were carefully selected from 9164 recordings collected between April 2010 and August 2012. Each CTG contains a FHR time series and a UC signal, each sampled at 4 Hz. These 552 recordings were screened by Chudacek et al [26] using many clinical and technical criteria, for instance, a maximum of 60 minutes for the first stage of labor and a maximum of 30 minutes for the second stage of labor.

The original length of each FHR signal is greater than 10000. In order to obtain a more intuitive and reliable recovery result for missing signals, we intercepted complete FHR data without missing samples and obvious artifacts from each FHR signal, each with length L = 2000, as shown in Fig. 3(a). It should be noted that for reconstruction of missing FHR in practice, there is no required signal length; FHR signals of any length are suitable. The data missing from FHR is highly random, and there is currently no gold standard that can truly simulate fetal/ maternal movements and other scenarios. So, we designed two protocols as described below to simulate a missing FHR sample, which is representative of a rather comprehensive scenario of challenges and issues for regression techniques. More precisely, we defined Q, q and num as total missing
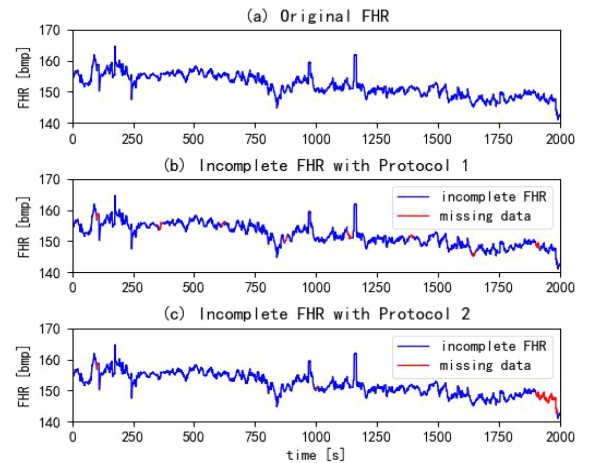

Fig. 3.   Complete FHR segment of the 2nd pregnant woman (top); Protocol 1 for missing data (middle), in which $Q = 5\%$, $q = \{9, 10, 11, 12, 13, 14, 15, 16\}$, and $num = 8$; Protocol 2 for missing data (bottom), in which $Q = 5\%$, $num = 3$, $q = \{5, 7, 88\}$. The red dotted line represents the original missing samples. bpm: beat per minute. Note: these two examples are not the only options for simulation protocols.
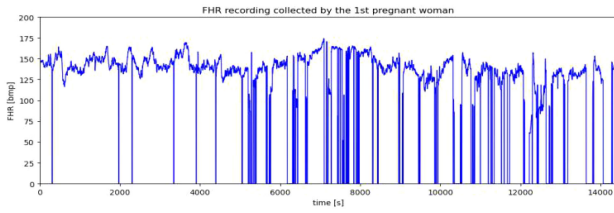
Fig. 4. FHR recordings collected by the 1st pregnant woman, with frequently missing data. Each abnormal waveform (of the value to zero) signifies different degrees of missing data.

percentage, length of each missing, and total number of missing times, respectively, with the following relationship: $Q = \sum_{i=1}^{num} q_i / L$. The first indicator measures the missing degree of an entire FHR, while the latter two measure the number and frequency of data missing in a pregnant woman throughout fetal monitoring.

1) *Protocol One: Frequent but short missing data segments:* In this protocol, each missing length q is randomly selected, $q \in (0, 25]$. The total missing percentage Q is increased from 5% to 40%, i.e., $Q \in [5\%, 10\%, 20\%, 30\%, 40\%]$, and num is adjusted successively as follows: increased from 8 to 64, i.e., $num \in [8, 16, 32, 48, 64]$. The reason why $Q \leq 40\%$ is the existing rule of clinical collection: when Q exceeds 40%, this signal is not available [6], [26]. This rule is also applied in the screening of CTU-UHB. The main goal of this protocol was to evaluate: 1) whether the proposed algorithm could achieve reliable reconstruction with multiple data missing (a large num) in actual acquisition. In other words, the maximum value of num that the proposed algorithm could accept; and 2) the dependence of miniMSDL on total missing percentage Q.

2) *Protocol Two: Fewer but longer missing data segments:* In this protocol, Q is set at 5%, i.e., the total length of data missing is 100. Define one of the missing lengths as a long time missing, $q_{max} \in (25, 100]$, and adjust num with the mathematical rules of formula (26). The main goal of this protocol is to evaluate the dependence of miniMSDL on the length of each missing q.

$$q_{max} = \begin{cases} \in (25, 50], \ num = 5, \ q \\ \quad = \{ q_1, \ldots, q_4, q_{max} \}; \\ \in (50, 98], \ num = 3, q = \{ q_1, q_2, q_{max} \}; \\ = 99, \quad num = 2, \ q = \{ 99, 1 \}; \\ = 100, \quad num = 1, \ q = \{ 100 \}. \end{cases} \tag{26}$$

We adopted this probabilistic approach to reproduce missing data segments which did not exhibit any periodicity or correlation with the FHR time-series. Examples of these two simulation protocols are shown in Fig. 3(b) and (c). Significantly, there are more than 300 groups of FHR signals in CTU-UHB which have multiple data missing, including the 1th, 8th, 9th, 12th, …, 549th, and 552th maternal recordings collected. Fig. 4 shows the signal of the 1st pregnant woman, which frequently has data missing, to the extent that it was difficult to intercept a signal segment

satisfying the conditions. Therefore, we sequentially screened these 552 groups of recordings and selected 100 groups for the missing data simulation, for example the 2nd recording, shown in Fig. 3(a).

B. Experimental Setup

1) Our experiments were carried out using Python 3.7 + scikit-learn 0.23.1 + scipy 1.4.1 on a personal server with an Intel(R) Xeon(R) Silver 4110 GHz and an NVIDIA Quadro P4000 GPU. The following basic parameters were used for the sparse coding and dictionary update: the number of iterations (n_iter in Table III) was set to 15; the algorithm used to solve the lasso problem was "lars"; the algorithm used to transform the data was "omp" and conjugate gradient algorithm proposed above. For optimal accuracy, the four most important parameters (block length N, batch size M, L_p-norm and genetic factor λ) were set based on 100 FHR recordings as described in Section IV.C.

2) *The primary indicators:* In order to provide a quantitative evaluation of the recovery of missing FHR, we first compared the mean absolute error (MAE) and the root mean square error (RMSE) of the reconstructed FHR and the original complete signal, which are well-known metrics typically used in data analysis. Concurrently, we increased the BIC [27] as well as the training and test time to study the optimal parameters of L_p-norm, M, and λ for the proposed miniMSDL approach. In data analysis, the BIC is the criterion for model selection for a limited number of training samples. With the increase of model complexity, the BIC value will increase, and conversely, as the likelihood increases, the BIC value will decrease. Therefore, the model with the lowest BIC value is the best.

3) *The secondary indicators:* The above algorithms were next evaluated by three clinical characteristics, as shown in the following formulas (26)–(28) [28]. These three indicators comprehensively described the shape and changes of the FHR baseline, and were the most important morphological time-domain features.

Short-term variability (STV):

$$\text{STV} = \frac{1}{24} \sum_{i=1}^{24} |s\hat{y}_i - sy_i| \tag{26}$$

Long-term variability (LTV):

$$\text{LTV} = IQR \left(\sqrt{\hat{y}_i^2 + \hat{y}_{i+1}^2} \right) \tag{27}$$

Interval index (INT):

$$\text{INT} = \frac{STV}{std(\hat{y}_i)} \tag{28}$$

where $s\hat{y}_i = \hat{y}_{(10(i-1)+1)}$ is the value of FHR recording $\hat{y}_i$ for each period of 2.5s; IQR denotes the inter-quartile range [0.25, 0.75]; and std represents the standard deviation.
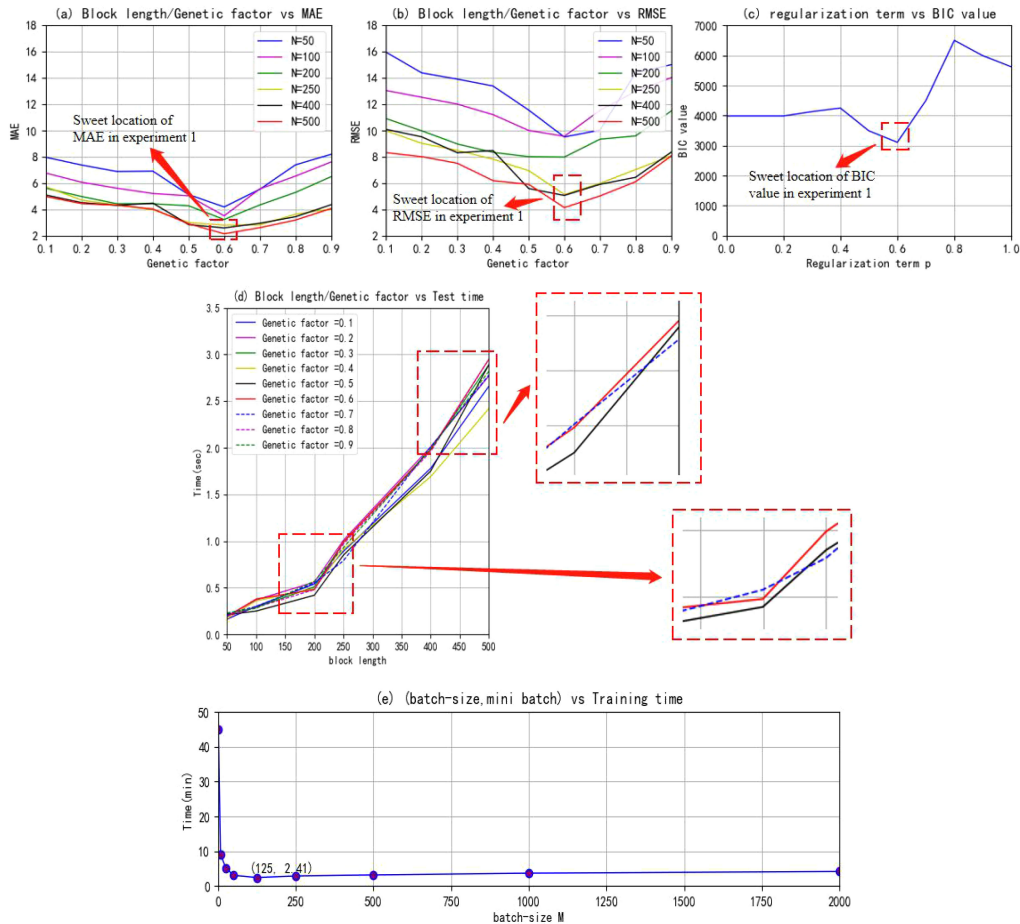
Fig. 5. Comparison of the primary indicators for different parameter combinations. Experimental data: using Protocol 1 to simulate the data missing: $Q = 10\%$, $num = 16$: (a)–(b), (d) set $p = 0.6$, $M = 125$, (c) set $\lambda = 0.6$, $N = 400$, $M = 125$, and (e) set $\lambda = 0.6$, $p = 0.6$, $N = 400$.

C. Optimization of Different Parameters

In this experiment, we looked at the effect of different parameters on the performance of miniMSDL. More specifically, the study mainly involved four training parameters: block length N in the dictionary $\hat{D} \in R^{M \times N}$ (also known as feature atoms); batch size M in mini-batch version training, which determined the number of mini-batches to complete training once (namely K); regularization term L_p-norm; and the genetic factor λ in the mini-batch version.

Based on 100 groups of complete FHR signals, we used Protocol 1 to simulate missing data, and set $Q = 10\%$, $num = 16$. Sub-graphs (a)-(e) in Fig. 5 show the changes in signal recovery as the block length N / batch size M/ genetic factor λ/ regularization term L_p-norm increased.

Sub-graphs (a) and (b) show that the proposed miniMSDL reached a sweet spot at around $\lambda = 0.6$. After partially magnifying three curves of $\lambda = 0.5$, $\lambda = 0.6$, and $\lambda = 0.7$ in sub-graph (d), we found that when compared with the other two curves, the test time at $\lambda = 0.6$ was relatively long, but the overall difference was within an acceptably small range. Further, the BIC values for different p are shown in sub-graph (c) and reached a sweet spot at around $p = 0.6$. With the increase of N, the values of MAE and RMSE showed a decreasing trend and reached the

ideal location around $N = 400$ and $N = 500$, so $N = 400$ or $N = 500$ are ideal values. However, it can be seen from Fig. 5(d) that with the same genetic factor, block length N ultimately affected the test time. The larger the value of N, the longer the test time. Sub-graph (e) shows that increasing batch size M within a reasonable range was conducive to improving the memory utilization and parallelization efficiency of the proposed algorithm. However, it is not advisable to blindly increase it, because this could lead to more time being needed to achieve the same accuracy and slower adjustment of the parameters. When the batch size is increased beyond normal limits, it will be converted to full-batch learning, which means that the proposed miniMSDL algorithm becomes batch-based MSDL (shown in Table I). However, it can be seen from Fig. 5(e) that the effect of full-batch-based MSDL ($M = 2000$, $K = 1$) is not optimal. By studying Fig. 5 we can see that:

– The shorter the block length N, the faster the test time.
– There was a sweet spot for the value of genetic factor λ that was not affected by the value of block length N.
– There was a point of optimal balance between M and K, i.e., $(M, K) = (125, 16)$.
– In order to reduce the testing time as much as possible while achieving the sweet spot, we set $\lambda = 0.6$, $p = 0.6$, $N = 400$, and $M = 125$.

TABLE IV
RECOVERY PERFORMANCE OF SECONDARY INDICATORS WITH VARIABLE TOTAL MISSING PERCENTAGE Q

	STV				LTV				INT (x10⁻³)			
Q	10%	20%	30%	40%	10%	20%	30%	40%	10%	20%	30%	40%
Missing FHR	29.85	29.41	39.88	51.69	4.20	3.81	3.23	3.04	2.21	3.05	3.46	5.72
Reconstructed FHR	**22.39**	**23.84**	28.89	34.10	**4.13**	4.75	3.98	3.36	**1.93**	**1.93**	2.25	7.21
Original complete FHR	23.44				4.61				1.93			
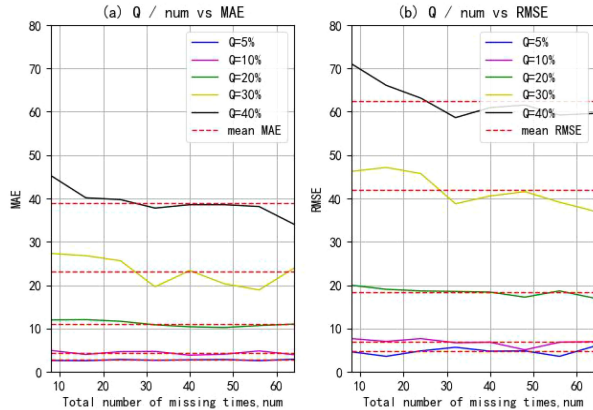
Fig. 6. Recovery performance of the primary indicators with variable Q and num.

how similar the morphological time-domain characteristics of the reconstructed signal were to the original signal. Due to space limitations, we only list the four groups with the best performance when Q = 10%, 20%, 30%, and 40%. As expected, higher deviations were noticeable with large Q. By studying Fig. 6 and Table IV we can see that:

– When the total missing percentage Q was small, the recovery result of the proposed miniMSDL algorithm was independent of the total number of missing times num.
– The higher the total missing percentage Q, the worse the degree of recovery.

Fig. 7 Shows One of the Missing Data Simulations for the 2nd Pregnant woman: $Q = 5\%$, $num = 8$, and the Reconstruction Result Was Based on miniMSDL. It is Worth Nothing That the Overlap of the Red and Blue Curves Indicates Good Performance in Reconstructing FHR.

V. DISCUSSION

A. Discussion 1: Influence of Protocol 1

In this section, we assess the performance of the proposed method in Protocol 1. Specifically, the experiment was designed to evaluate the influence of total missing percentage Q and total number of missing times num.

Performance of the primary and secondary indicators with different Q and various num are shown in Fig. 6 and Table IV, respectively. It can be seen that the changing curve of MSE and the mean line basically overlap when Q is 5%, 10%, and 20%. We can draw a general conclusion that all num values achieved similar good performance for short interval lengths (within $Q \leq 20\%$) and achieved the lowest values of 2.64 (MAE) and 4.68 (RMSE) with Q = 5%. Therefore, when the total missing percentage Q was relatively small, miniMSDL could achieve reliable data recovery no matter how many times missing data occurred in the actual collection. However, the amount of missing data that miniMSDL could withstand was limited. In fact, any regression algorithm has similar problems, and we will give a detailed experimental analysis in Section V.C. It is worth noting that excessive missing data during actual signal acquisition is itself a cause of unqualified samples, and even with external ultrasound measurements, the maximum data missing will be no more than 40% [6]. Nowadays, with the development of semiconductors, microelectronics, etc., the severity of missing signals has been reduced to a certain extent.

In Table IV, we show the computations for three clinical indicators and compare the resulting values with the ones obtained from the original recording. The intent was to look at the clinical characteristics of the reconstructed signals and see

B. Discussion 2: Influence of Protocol 2

We Would Like to Point Out That While the Inpainting Methods Can Reconstruct the Missing Data to Varying degrees, They Might Miss Some Details If the Missing Length q Becomes Too large. Therefore, We Devised an Experiment to Take a Closer Look At the Performance for Different Lengths of Each Missing q.

Similar to Discussion 1, the experiment was based on 100 groups of complete FHR signals in CTU-UHB and used Protocol 2 to simulate missing data, with Q set at 5%. Fig. 8 shows the change in MAE and RMSE with the increase of missing length q. For example, when $q = 100$, the proposed miniMSDL achieved achieved a MAE of 45.91 and RMSE of 75.91. It is evident that the waveform properties for signals with short missing lengths were restored better than those for signals with long missing lengths. Therefore, a long missing length affected data retrieval ability. However, when the length of each missing q satisfied q ≤ 60, the proposed miniMSDL could achieve fairly good reconstruction performance.

Actually, in the case of FHR recordings, temporary increases or decreases are important details when determining fetal well-being, and indicate accelerations and decelerations in the heart rate. Therefore, a long missing length in itself creates an unqualified collection. In an abrupt acceleration or deceleration the FHR has a change of 15 beats per minute with a time from onset to extremum of 30 seconds and total duration of less than 2 minutes. Based on this, it is safe to reconstruct segments with a maximum length of 25 seconds [19]. Each FHR of CTU-UHB dataset was sampled at 4 Hz, corresponding to 100 sample point. This is why we set the maximum value of
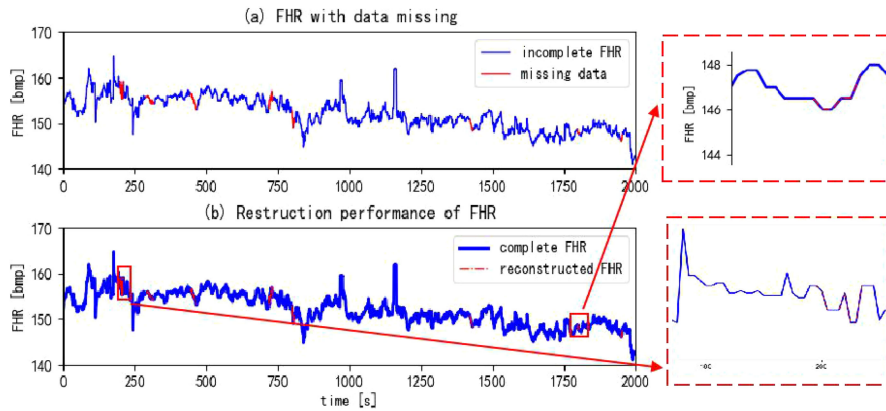
Fig. 7.   Protocol 1 for missing data (top), reconstruction of missing data (bottom).
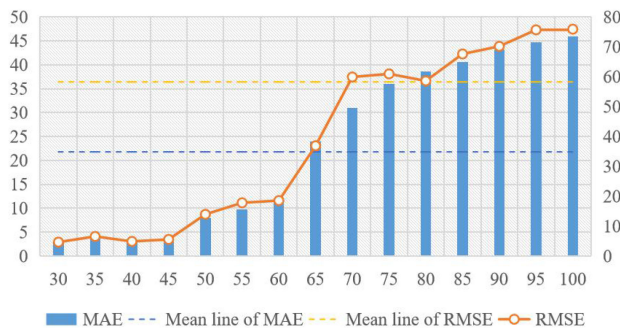

Fig. 8.   Recovery performance of the first indicators with variable q.

q_{max} to 100, and the total missing percentage as $Q = 5\%$ in this protocol.

C. Discussion 3: Comparison With Related Works

Over the years, multiple studies on the recovery of missing data have been conducted in many application scenarios. In order to carry out a more objective and comparative performance evaluation, we first evaluated the proposed miniMSDL through the reproduction of three classic methods, and then made a comparison with state-of-the-art works.

1) Evaluation Based on Reproducing the Classical Algorithm: The reproduced methods were three-step interpolation that integrated spline interpolation and cubic spline interpolation to to recover FHR data, which our team proposed in a previous study [15]; cubic spline interpolation [16] and K-Singular value decomposition (K-SVD) [18], which represent typical mathematical algorithms or sparse dictionary learning. As explained in Section IV. C, the recovery performance on missing FHR had a strong relationship with the total missing percentage Q and length of each missing q. Therefore, we reproduced these methods through the following two simulations, in which the experimental data in Figs. 9 and 10 was based on 100 groups from the CTU-UHB database:

1) In the case of Protocol 1, Qincreased from 5% to 40%, with a step of 5; and q ranged from 8 to 64, increasing in steps of 8. In Fig. 9(a)(b), all algorithms have achieved the ideal reconstruction with a small total missing percentage

Q. When the total missing percentage Qincreased to 30%, the performance of Algorithms 2 and 3 decreased notably. Fig. 10 shows the best performance of these four methods in Protocol 1 ($Q = 5\%$, num $= 8$). The closer the secondary indicators of each algorithm are to the original complete FHR, the better the recovery of the algorithm. It can be seen that Algorithm 1 and the proposed miniMSDL provided optimal performance, nearly independent from the missing data pattern.

2) In the case of Protocol 2, $Q = 5\%$, while q ranged from 25 to 100. From Fig. 9(c), (d) we can see that Algorithms 1 and 2 were the most sensitive methods to length of each missing $q$values. This was no surprise since they were partially or completely calculated based on the mean value of the sample points before and after the current missing position. The longer the missing length, the more difficult it is to recover. Sub-graphs (a) -(c) in Fig. 10 show that all algorithms are quite different from the original complete FHR in STV and LTV, especially Algorithms 1 and 2. By studying Fig. 10 we can see that the challenge of using Algorithm 1 and 2, meanwhile, was that it introduced artifacts which increased the need for the subsequent denoising step.

The obtained results show that the total missing percentage that all regression methods could withstand was limited. To some extent, this is indeed a common drawback in regression algorithms. However, both the FHR acquisition requirements and equipment ensure that the Qwill not be too large. Therefore, the impact of such drawbacks is acceptable.

2) Comparison With State-Of-The-Art Inpainting Methods: In addition to the comparison with classical methods, we compared the performance of the proposed miniMSDL with state-of-the-art methodologies. Table V summarizes representative research that has been conducted in the past few years. Since the datasets used and the amount of missing data in each study are not in complete accord, it is difficult to evaluate directly from indicators such as MAE. We analyzed them and compared them with the following three aspects.

Compared with statistics-based inpainting like that proposed by G. Feng et al. [7] and SK Lee *et al.* [9], an important advantage of the method outlined in this paper is that through
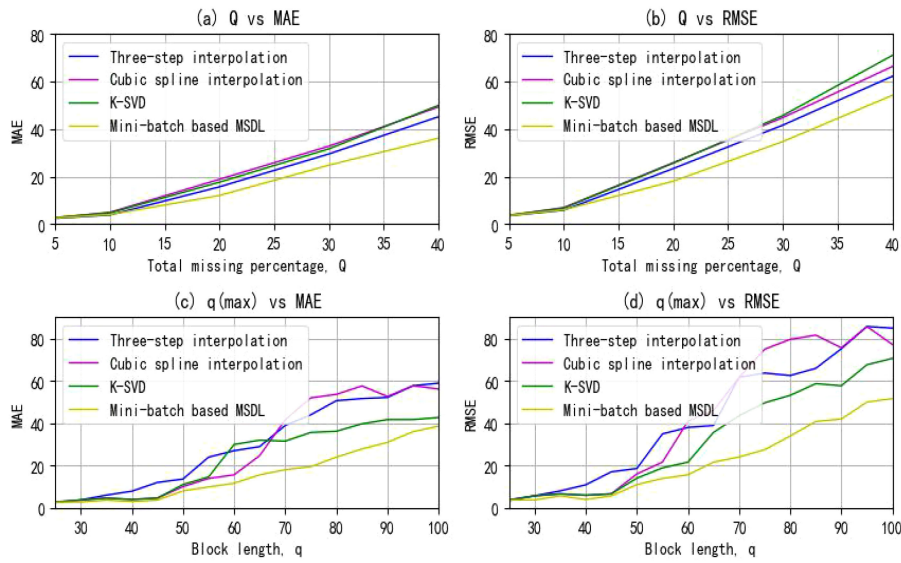
Fig. 9. Comparison of four classical approaches: Recovery performance of the primary indicators, with variable Q and q.
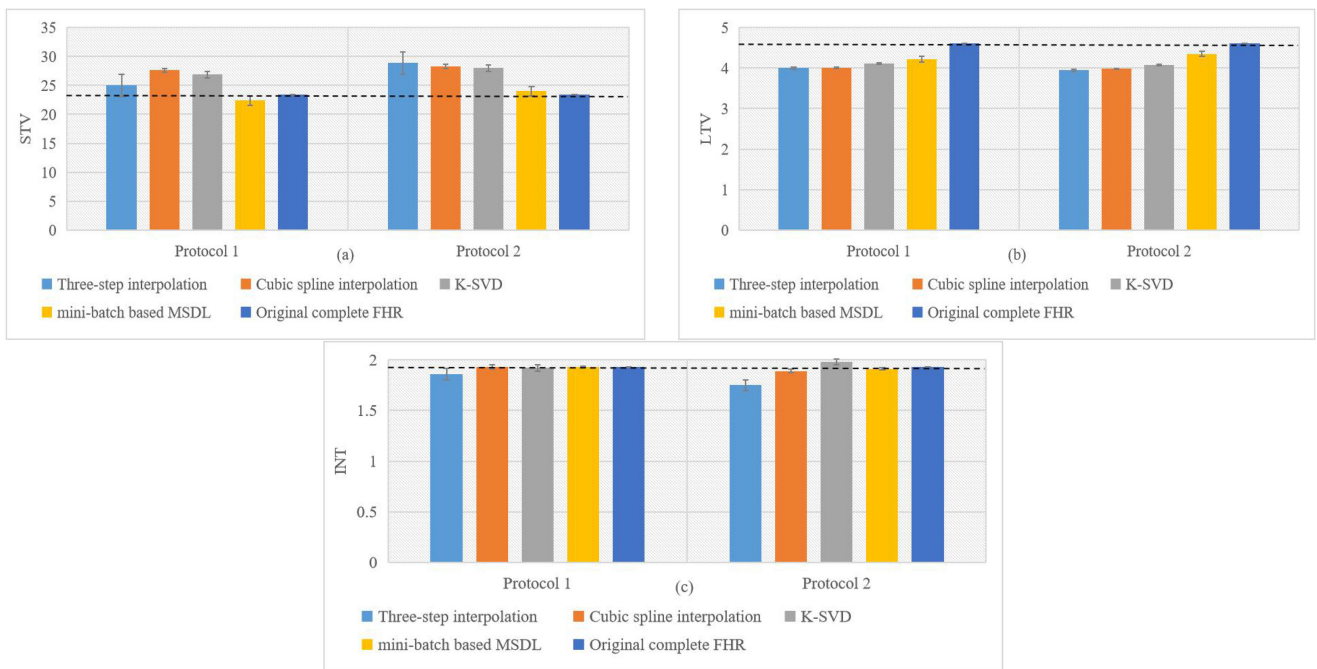

Fig. 10. Comparison of four classical approaches: The best performance of the secondary indicators in Protocols 1 (left) and 2 (right).

TABLE V
COMPARATIVE TABULATION OF THE EXPERIMENTAL RESULTS FOR VARIOUS ALGORITHMS

Authors	Algorithm type	Specific steps/algorithm	Performance
G Feng er al. (2017) [7]	Mathematics / Statistics	Gaussian process	Performed well when q was small, but the artifacts increased during feature extraction and time-frequency analysis
SK Lee et al. (2019) [9]		ARCH (1)	
VP Oikonomou et al. (2013) [18]	Machine learning	K-SVD	Dictionary had difficulty learning adaptively when q was large
F. Barzideh et al. (2018) [19]	Machine learning	SDL + Shift-Invariant Dictionary	Large computational burden
Proposed	**Machine learning**	**miniMSDL**	-

learning from the signal class, the dictionary introduces fewer artifacts during feature extraction and time-frequency analysis. It should be noted that in contrast to the methods used by V.P. Oikonomou et al. [18] and F. Barzideh et al. [19], which generated dictionaries based on pre-specified dictionary and a standard sparse coding solution based on L_0-normrespectively, we provide an L_p-norm-based MSDL approach to generating and updating the dictionary and sparse coefficient matrix. Also, we considered a conjugate gradient and mini-batch version to reduce the computational burden.

Based on the above analysis, the proposed miniMSDL approach provides optimal performance in all the conditions considered and guarantees reduced distortion of the original signal features.

VI. CONCLUSION

The study and recovery of missing FHR recordings has very important practical value for the development of modern intelligent auxiliary diagnosis technology to detect fetal distress in late pregnancy, and is an important research direction in the field of intelligent health applications. The major concerns for normal inpainting methods appear to be long missing length and suitability for the online mode with timely data recovery.

In this study, we developed a new inpainting method for the recovery of missing FHR data, optimizing the model by combining mini-batches and minimized sparse dictionary learning, which addresses the problem of inpainting missing samples which can be used in both antepartum and intrapartum FHR measurements. We designed two protocols and used 100 clinical FHR recordings to simulate two missing data scenarios.

Then 7 evaluation indicators (4 metrics typical in data analysis and 3 clinical indicators) were applied to adjust and determine 4 major parameters (block length $N = 400$, batch size $M = 125$, L_p-norm, $p = 0.6$, and genetic factor $\lambda = 0.6$). These are the only parameters that need to be adjusted in the algorithm. Therefore, it is not difficult to solve the algorithm. Our study identified the following key points:

1) The proposed mini-batch-based minimized sparse dictionary learning algorithm relies exclusively on a reduced set of samples and has low computational burden, making it compliant with online implementation.
2) Compared with the traditional methods such as K-SVD, the computational load is greatly reduced due to the mini-batch solution, and memory learning and training for dictionaries.
3) For dictionary learning-based methods, missing length $q$and total missing percentage Q are important parameters when attempting to recover the missing data in 1D signal.

Of course, the main limitation of this paper is that in order to obtain an optimal dictionary, we went through all the possible options to choose the best one. A possible direction for future work is to find a more efficient way to identify the best dictionary.

REFERENCES

[1] R. H. Paul, "ACOG practice bulletin no. 145: Antepartum fetalSurveillance," *Obstet. Gynecol.*, vol. 124, no. 1, pp. 182–192, Jun. 2014.
[2] D. Ayres-De-Campos *et al.*, "FIGO consensus guidelines on intrapartum fetal monitoring: Cardiotocography," *Int. J. Gynecol. Obstet.*, vol. 131, no. 1, pp. 13–24, Jun. 2015.
[3] M. A. G. Santos *et al.*, "Online heart monitoring systems on the internet of health things environments: A survey, a reference model and an outlook," *Inf. Fusion*, vol. 53, pp. 222–239, Jan. 2020.
[4] W. Ding, M. Abdel-Basset, K. A. Eldrandaly, L. Abdel-Fatah, and V. H. C. de Albuquerque, "Smart supervision of cardiomyopathy based on fuzzy Harris Hawks optimizer and wearable sensing data optimization: A new model," *IEEE Trans. Cybern.*, to be published, doi: 10.1109/TCYB.2020.3000440.
[5] J. A. L. Marques *et al.*, "IoT-based smart health system for ambulatory maternal and fetal monitoring," *IEEE Internet Things J.*, to be published, doi: 10.1109/JIOT.2020.3037759.
[6] P. Bakker *et al.*, "The quality of intrapartum fetal heart rate monitoring," *Eur. J. Obstet. Gynecol. Reprod. Biol.*, vol. 116, no. 1, pp. 22–27, Oct. 2004.
[7] G. Feng, J. G. Quirk, P. M. Djuric., "Recovery of missing samples in fetal heart rate recordings with Gaussian processes," in *Proc. 25th Eur. Signal Process. Conf.*, Aug. 2017.
[8] Z. F. Gao *et al.*, "Learning physical properties in complex visual scenes: An intelligent machine for perceiving blood flow dynamics from static CT angiography imaging," *Neural Netw.*, vol. 123, pp. 82–93, Mar. 2020.
[9] S. K. Lee, Y. S. Park, K. J. Cha, "Recovery of signal loss adopting the residual bootstrap method in fetal heart rate dynamics," *Biomed. Tech., (Berl)*, vol. 64, no. 2, pp. 157–161, Apr. 2019.
[10] Y. Jia, J. Ma, "What can machine learning do for seismic data processing? An interpolation application," *Geophysics*, vol. 82, no. 3, pp. 163–177, May. 2017.
[11] Z. Tang, Y. Bao, H. Li, "Group sparsity-aware convolutional neural network for continuous missing data recovery of structural health monitoring," *Struct. Health Monit.*, Jul. 2020.
[12] X. Chai, G. Tang, S. Wang, K. Lin, and R. Peng, "Deep learning for irregularly and regularly missing 3-D data reconstruction," *IEEE Trans. Geoence Remote Sens.*, vol. 59, no. 7, pp. 6244–6265, Jul. 2021.
[13] A. Javaheri, H. Zayyani, F. Marvasti, "Recovery of missing samples using sparse approximation via a convex similarity measure," in *Proc. Int. Conf. SampTA*, Jul. 2017.
[14] A. Javaheri, H. Zayyani, F. Marvasti, "Sparse recovery of missing image samples using a convex similarity index," *Signal Process.*, vol. 152, pp. 90–103, Nov. 2018.
[15] Y. Deng *et al.*, "DeepFHR: Intelligent prediction of fetal Acidemia using fetal heart rate signals based on convolutional neural network," *BMC Med. Inf. Decis. Mak.*, vol. 19, no. 1, Dec. 2019.
[16] G. G. J. Spilka *et al.*, "Discriminating normal from 'Abnormal' pregnancy cases using an automated FHR evaluation method," in *Proc. SETN*, pp. 521–531, 2014.
[17] M. Cesarelli, M. Romano, P. Bifulco, F. Fedele, M. Bracale, "An algorithm for the recovery of fetal heart rate series from CTG data," *Comput. Biol. Med.*, vol. 37, no. 5, pp. 663–669, May 2007.
[18] V. P. Oikonomou, J. Spilka, C. Stylios, and L. Lhostka, "An adaptive method for the recovery of missing samples from FHR time series," in *Proc. IEEE Int. Symp. Comput.-Based Med. Syst.*, Jun. 2013.
[19] F. Barzideh *et al.*, "Estimation of missing data in fetal heart rate signals using shift-invariant dictionary," in *Proc. 26th Eur. Signal Process. Conf.*, Sep. 2018.
[20] G.S. Mallat, "Adaptive time-frequency decompositions," *Opt. Eng.*, vol. 33, no. 7, pp. 7–10, Jul. 1994.
[21] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
[22] P. Vincent, Y. Bengio, "Kernel matching pursuit," *Mach. Learn.*, vol. 48, pp. 165–187, 2002.
[23] M. Tan, I. W. Tsang, L. Wang, "Matching pursuit LASSO Part I: Sparse recovery over big dictionary," *IEEE Trans. Signal Process.*, vol. 63, no. 3, pp. 727–741, 2015.
[24] CTU-CHB Intrapartum Cardiotocography Database [EB/OL]. [Online]. Available: https://www.physionet.org/physiobank/database/ctu-uhb-ctgdb/
[25] A. L. Goldberger *et al.*, "PhysioBank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. E215–E220, 2000.
[26] V. Chudacek *et al.*, "Fetal heart rate data pre-processing and annotation," in *Proc. Int. Conf. Inf. Technol. Appl. Biomed.*, Nov. 2009.
[27] G. E. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, no. 2, Mar. 1978.
[28] G. Georgoulas, D. Stylios, and P. Groumpos., "Predicting the risk of metabolic acidosis for newborns based on fetal heart rate signal classification using support vector machines," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 5, pp. 875–884, May 2006.