Supplementary Material

In this document, we give the detailed proof process of Theorem 2 in our work. Additionally, we also give the detailed parameter analysis convergence analysis, and a comparative study of pre-trained and re-trained models for the proposed LRNDP algorithm.

A. Convergence Analysis

In this subsection, we establish detailed convergence analysis for the proposed algorithm in the framework of PAM algorithm. For convenience, we define the following functions:

$$\begin{aligned} L(\boldsymbol{\mathcal{R}}, \boldsymbol{\mathcal{Z}}, \boldsymbol{\mathcal{M}}, \boldsymbol{\mathcal{X}}) &:= \frac{1}{2}\|\boldsymbol{\mathcal{A}}(\boldsymbol{\mathcal{X}}) - \boldsymbol{\mathcal{Y}}\|_F^2 \\ &+ \sum_{p=1}^{P}(\sum_{i=1}^{2}\alpha\|\nabla_i \boldsymbol{\mathcal{R}}_p\|_{\boldsymbol{w}_{\boldsymbol{\mathcal{R}}_p},\star} + \frac{\lambda}{2}\|\boldsymbol{\mathcal{M}}_p - \boldsymbol{\mathcal{R}}_p\|_F^2) \\ &+ \beta\Phi(\boldsymbol{\mathcal{Z}}) + \frac{\rho}{2}\|\boldsymbol{\mathcal{M}} - \boldsymbol{\mathcal{Z}}\|_F^2 + \frac{\tau}{2}\|\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{M}}_{\times 3}\boldsymbol{E}\|_F^2, \end{aligned}$$

$$\begin{aligned} f(\boldsymbol{\mathcal{R}}, \boldsymbol{\mathcal{Z}}, \boldsymbol{\mathcal{M}}, \boldsymbol{\mathcal{X}}) &:= \frac{1}{2}\|\boldsymbol{\mathcal{A}}(\boldsymbol{\mathcal{X}}) - \boldsymbol{\mathcal{Y}}\|_F^2 + \sum_{p=1}^{P}\frac{\lambda}{2}\|\boldsymbol{\mathcal{M}}_p - \boldsymbol{\mathcal{R}}_p\|_F^2 \\ &+ \frac{\rho}{2}\|\boldsymbol{\mathcal{M}} - \boldsymbol{\mathcal{Z}}\|_F^2 + \frac{\tau}{2}\|\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{M}}_{\times 3}\boldsymbol{E}\|_F^2, \end{aligned}$$

$$\begin{aligned} f_1(\boldsymbol{\mathcal{R}}) &:= \sum_{p=1}^{P}\sum_{i=1}^{2}\alpha\|\nabla_i \boldsymbol{\mathcal{R}}_p\|_{\boldsymbol{w}_{\boldsymbol{\mathcal{R}}_p},\star}, \\ f_2(\boldsymbol{\mathcal{Z}}) &:= \beta\Phi(\boldsymbol{\mathcal{Z}}). \end{aligned}$$

Next, we give the global convergence theorem of the proposed LRNDP algorithm as follows.

Theorem 2 (Global convergence)**.** *Assuming that the sequence $\{\boldsymbol{\mathcal{R}}^t, \boldsymbol{\mathcal{Z}}^t, \boldsymbol{\mathcal{M}}^t, \boldsymbol{\mathcal{X}}^t\}$ generated by Algorithm 1 is bounded and $\Phi(\boldsymbol{\mathcal{Z}})$ is a KŁ function, then the whole sequence $\{\boldsymbol{\mathcal{R}}^t, \boldsymbol{\mathcal{Z}}^t, \boldsymbol{\mathcal{M}}^t, \boldsymbol{\mathcal{X}}^t\}$ converges to a critical point of $L(\boldsymbol{\mathcal{R}}, \boldsymbol{\mathcal{Z}}, \boldsymbol{\mathcal{M}}, \boldsymbol{\mathcal{X}})$.*

The proof of Theorem 2 is inspired by Theorem 6.2 in [1]. If the following three key conditions are met, we can use the established results to prove that the proposed LRNDP algorithm is convergent.

1) $L(\boldsymbol{\mathcal{R}}^t, \boldsymbol{\mathcal{Z}}^t, \boldsymbol{\mathcal{M}}^t, \boldsymbol{\mathcal{X}}^t)$ satisfies the KŁ property at each point;
2) the sufficient decrease condition;
3) the relative error condition.

In other words, the proof of theorem 2 is by verifying that the above three key conditions are established. Before verifying these conditions, we first introduce some related definitions.

Definition 1 (Kurdyka-Lojasiewicz property [1])**.** *A proper lower semi-continuous function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is said to have the KŁ property at point $\boldsymbol{x}^* \in \mathrm{dom}(\partial f)$ if there exists $\eta \in \mathbb{R}^+$, a neighborhood U of $\boldsymbol{x}^*$ and a continuous concave function φ such that for each $\boldsymbol{x} \in U \cap \{\boldsymbol{x}|f(\boldsymbol{x}^*) < f(\boldsymbol{x}) < f(\boldsymbol{x}^*) + \eta\}$, the KŁ inequality holds:*

$$\varphi'(f(\boldsymbol{x}) - f(\boldsymbol{x}^*))\,\mathrm{dist}(0, \partial f(\boldsymbol{x})) \geq 1,$$

where $\varphi : [0, \eta) \to \mathbb{R}^+$ satisfies: 1) $\varphi(0) = 0$; 2) φ is C^1-type on $[0, \eta)$; and 3) $\varphi'(s) > 0$ for any $s \in (0, \eta)$.

A function f that satisfies the KŁ property at every point of $dom(\partial f)$ is called a KŁ function.

Definition 2 (Semi-algebraic set and function [1])**.** *If there exists a finite number of real polynomial functions g_{ij} and n_{ij} satisfy $S = \bigcap_j \bigcup_i \{\boldsymbol{x} \in \mathbb{R}^n : g_{ij}(\boldsymbol{x}) = 0, n_{ij} < 0\}$, then the subset $S \in \mathbb{R}^n$ is called the semi-algebraic set. If the graph $\{(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^n \times \mathbb{R}, f(\boldsymbol{x}) = \boldsymbol{y}\}$ of the function f is a semi-algebraic set, then f is called the semi-algebraic function.*

Proposition 1. *A semi-algebraic real valued function f is a KŁ function, i.e., f satisfies KŁ property at every $\boldsymbol{x} \in dom(f)$ [2].*

Next, we first verify the KŁ property of function $L(\boldsymbol{\mathcal{R}}, \boldsymbol{\mathcal{Z}}, \boldsymbol{\mathcal{M}}, \boldsymbol{\mathcal{X}})$. Secondly, we prove that the bounded sequence $\{\boldsymbol{\mathcal{R}}^t, \boldsymbol{\mathcal{Z}}^t, \boldsymbol{\mathcal{M}}^t, \boldsymbol{\mathcal{X}}^t\}$ satisfies the sufficient decrease and relative error conditions. Finally, we have completed the proof of Theorem 2.

Lemma A1 (KŁ Lemma)**.** *The function $L(\boldsymbol{\mathcal{R}}, \boldsymbol{\mathcal{Z}}, \boldsymbol{\mathcal{M}}, \boldsymbol{\mathcal{X}})$ satisfies the KŁ property at each point.*

Proof: We prove Lemma A1 by verifying that each part of the function $L(\boldsymbol{\mathcal{R}}, \boldsymbol{\mathcal{Z}}, \boldsymbol{\mathcal{M}}, \boldsymbol{\mathcal{X}})$ satisfies the KŁ property. As the nuclear norm and Frobenius norm are semi-algebraic functions [2], and the proposition 1 indicates that the semi-algebraic function satisfies the KŁ property, therefore, $\|\boldsymbol{\mathcal{A}}(\boldsymbol{\mathcal{X}}) - \boldsymbol{\mathcal{Y}}\|_F^2$, $\sum_{p=1}^{P}\sum_{i=1}^{2}\|\nabla_i \boldsymbol{\mathcal{R}}_p\|_{\boldsymbol{w}_{\boldsymbol{\mathcal{R}}_p},\star}$, $\sum_{p=1}^{P}\|\boldsymbol{\mathcal{M}}_p - \boldsymbol{\mathcal{R}}_p\|_F^2$, $\|\boldsymbol{\mathcal{M}} - \boldsymbol{\mathcal{Z}}\|_F^2$, $\|\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{M}}_{\times 3}\boldsymbol{E}\|_F^2$ are KŁ functions. Moreover, according to the assumption of Theorem 2, $\Phi(\boldsymbol{\mathcal{Z}})$ also satisfy the KŁ property. Hence, the function $L(\boldsymbol{\mathcal{R}}, \boldsymbol{\mathcal{Z}}, \boldsymbol{\mathcal{M}}, \boldsymbol{\mathcal{X}})$ is a KŁ function. ■

Lemma A2 (Sufficient decrease lemma)**.** *For $\eta > 0$, the sequence $\{\boldsymbol{\mathcal{R}}^t, \boldsymbol{\mathcal{Z}}^t, \boldsymbol{\mathcal{M}}^t, \boldsymbol{\mathcal{X}}^t\}$ generated by Algorithm 1 satisfies*

the following inequalities:

$$\begin{cases} L(\boldsymbol{\mathcal{R}}^{t+1},\boldsymbol{\mathcal{Z}}^{t},\boldsymbol{\mathcal{M}}^{t},\boldsymbol{\mathcal{X}}^{t})+\frac{\eta}{2}\|\boldsymbol{\mathcal{R}}^{t+1}-\boldsymbol{\mathcal{R}}^{t}\|_F^2 \\ \quad \le L(\boldsymbol{\mathcal{R}}^{t},\boldsymbol{\mathcal{Z}}^{t},\boldsymbol{\mathcal{M}}^{t},\boldsymbol{\mathcal{X}}^{t}), \\ L(\boldsymbol{\mathcal{R}}^{t+1},\boldsymbol{\mathcal{Z}}^{t+1},\boldsymbol{\mathcal{M}}^{t},\boldsymbol{\mathcal{X}}^{t})+\frac{\eta}{2}\|\boldsymbol{\mathcal{Z}}^{t+1}-\boldsymbol{\mathcal{Z}}^{t}\|_F^2 \\ \quad \le L(\boldsymbol{\mathcal{R}}^{t+1},\boldsymbol{\mathcal{Z}}^{t},\boldsymbol{\mathcal{M}}^{t},\boldsymbol{\mathcal{X}}^{t}), \\ L(\boldsymbol{\mathcal{R}}^{t+1},\boldsymbol{\mathcal{Z}}^{t+1},\boldsymbol{\mathcal{M}}^{t+1},\boldsymbol{\mathcal{X}}^{t})+\frac{\eta}{2}\|\boldsymbol{\mathcal{M}}^{t+1}-\boldsymbol{\mathcal{M}}^{t}\|_F^2 \\ \quad \le L(\boldsymbol{\mathcal{R}}^{t+1},\boldsymbol{\mathcal{Z}}^{t+1},\boldsymbol{\mathcal{M}}^{t},\boldsymbol{\mathcal{X}}^{t}), \\ L(\boldsymbol{\mathcal{R}}^{t+1},\boldsymbol{\mathcal{Z}}^{t+1},\boldsymbol{\mathcal{M}}^{t+1},\boldsymbol{\mathcal{X}}^{t+1})+\frac{\eta}{2}\|\boldsymbol{\mathcal{X}}^{t+1}-\boldsymbol{\mathcal{X}}^{t}\|_F^2 \\ \quad \le L(\boldsymbol{\mathcal{R}}^{t+1},\boldsymbol{\mathcal{Z}}^{t+1},\boldsymbol{\mathcal{M}}^{t+1},\boldsymbol{\mathcal{X}}^{t}). \end{cases}$$

Proof: Suppose $\boldsymbol{\mathcal{R}}^{t+1},\boldsymbol{\mathcal{Z}}^{t+1},\boldsymbol{\mathcal{M}}^{t+1}$ and $\boldsymbol{\mathcal{X}}^{t+1}$ are the optimal solutions of F_1,F_2,F_3, and F_4 in the optimization problem (8), then we have

$$\begin{aligned} &L(\boldsymbol{\mathcal{R}}^{t+1},\boldsymbol{\mathcal{Z}}^{t},\boldsymbol{\mathcal{M}}^{t},\boldsymbol{\mathcal{X}}^{t})+\frac{\eta}{2}\|\boldsymbol{\mathcal{R}}^{t+1}-\boldsymbol{\mathcal{R}}^{t}\|_F^2 \\ &=F_1(\boldsymbol{\mathcal{R}}^{t+1}|\boldsymbol{\mathcal{R}}^{t})\le F_1(\boldsymbol{\mathcal{R}}^{t}|\boldsymbol{\mathcal{R}}^{t})=L(\boldsymbol{\mathcal{R}}^{t},\boldsymbol{\mathcal{Z}}^{t},\boldsymbol{\mathcal{M}}^{t},\boldsymbol{\mathcal{X}}^{t}), \\ &L(\boldsymbol{\mathcal{R}}^{t+1},\boldsymbol{\mathcal{Z}}^{t+1},\boldsymbol{\mathcal{M}}^{t},\boldsymbol{\mathcal{X}}^{t})+\frac{\eta}{2}\|\boldsymbol{\mathcal{Z}}^{t+1}-\boldsymbol{\mathcal{Z}}^{t}\|_F^2 \\ &=F_2(\boldsymbol{\mathcal{Z}}^{t+1}|\boldsymbol{\mathcal{Z}}^{t})\le F_2(\boldsymbol{\mathcal{Z}}^{t}|\boldsymbol{\mathcal{Z}}^{t})=L(\boldsymbol{\mathcal{R}}^{t+1},\boldsymbol{\mathcal{Z}}^{t},\boldsymbol{\mathcal{M}}^{t},\boldsymbol{\mathcal{X}}^{t}), \\ &L(\boldsymbol{\mathcal{R}}^{t+1},\boldsymbol{\mathcal{Z}}^{t+1},\boldsymbol{\mathcal{M}}^{t+1},\boldsymbol{\mathcal{X}}^{t})+\frac{\eta}{2}\|\boldsymbol{\mathcal{M}}^{t+1}-\boldsymbol{\mathcal{M}}^{t}\|_F^2 \\ &=F_3(\boldsymbol{\mathcal{M}}^{t+1}|\boldsymbol{\mathcal{M}}^{t})\le F_3(\boldsymbol{\mathcal{M}}^{t}|\boldsymbol{\mathcal{M}}^{t})=L(\boldsymbol{\mathcal{R}}^{t+1},\boldsymbol{\mathcal{Z}}^{t+1},\boldsymbol{\mathcal{M}}^{t},\boldsymbol{\mathcal{X}}^{t}), \\ &L(\boldsymbol{\mathcal{R}}^{t+1},\boldsymbol{\mathcal{Z}}^{t+1},\boldsymbol{\mathcal{M}}^{t+1},\boldsymbol{\mathcal{X}}^{t+1})+\frac{\eta}{2}\|\boldsymbol{\mathcal{X}}^{t+1}-\boldsymbol{\mathcal{X}}^{t}\|_F^2 \\ &=F_4(\boldsymbol{\mathcal{X}}^{t+1}|\boldsymbol{\mathcal{X}}^{t})\le F_4(\boldsymbol{\mathcal{X}}^{t}|\boldsymbol{\mathcal{X}}^{t})=L(\boldsymbol{\mathcal{R}}^{t+1},\boldsymbol{\mathcal{Z}}^{t+1},\boldsymbol{\mathcal{M}}^{t+1},\boldsymbol{\mathcal{X}}^{t}). \end{aligned}$$

This completes the proof. ■

Lemma A3 (Relative error lemma)**.** *For $\eta>0$, the sequence $\{\boldsymbol{\mathcal{R}}^{t},\boldsymbol{\mathcal{Z}}^{t},\boldsymbol{\mathcal{M}}^{t},\boldsymbol{\mathcal{X}}^{t}\}$ is generated by Algorithm 1. Then, there exist $\boldsymbol{\mathcal{V}}_1^{t+1}$, $\boldsymbol{\mathcal{V}}_2^{t+1}$, $\boldsymbol{\mathcal{V}}_3^{t+1}$ and $\boldsymbol{\mathcal{V}}_4^{t+1}$ that satisfy the following inequalities:*

$$\begin{cases} \|\boldsymbol{\mathcal{V}}_1^{t+1}+\nabla_{\boldsymbol{\mathcal{R}}}f(\boldsymbol{\mathcal{R}}^{t+1},\boldsymbol{\mathcal{Z}}^{t},\boldsymbol{\mathcal{M}}^{t},\boldsymbol{\mathcal{X}}^{t})\|_F \\ \quad \le \eta\|\boldsymbol{\mathcal{R}}^{t+1}-\boldsymbol{\mathcal{R}}^{t}\|_F, \\ \|\boldsymbol{\mathcal{V}}_2^{t+1}+\nabla_{\boldsymbol{\mathcal{Z}}}f(\boldsymbol{\mathcal{R}}^{t+1},\boldsymbol{\mathcal{Z}}^{t+1},\boldsymbol{\mathcal{M}}^{t},\boldsymbol{\mathcal{X}}^{t})\|_F \\ \quad \le \eta\|\boldsymbol{\mathcal{Z}}^{t+1}-\boldsymbol{\mathcal{Z}}^{t}\|_F, \\ \|\boldsymbol{\mathcal{V}}_3^{t+1}+\nabla_{\boldsymbol{\mathcal{M}}}f(\boldsymbol{\mathcal{R}}^{t+1},\boldsymbol{\mathcal{Z}}^{t+1},\boldsymbol{\mathcal{M}}^{t+1},\boldsymbol{\mathcal{X}}^{t})\|_F \\ \quad \le \eta\|\boldsymbol{\mathcal{M}}^{t+1}-\boldsymbol{\mathcal{M}}^{t}\|_F, \\ \|\boldsymbol{\mathcal{V}}_4^{t+1}+\nabla_{\boldsymbol{\mathcal{X}}}f(\boldsymbol{\mathcal{R}}^{t+1},\boldsymbol{\mathcal{Z}}^{t+1},\boldsymbol{\mathcal{M}}^{t+1},\boldsymbol{\mathcal{X}}^{t+1})\|_F \\ \quad \le \eta\|\boldsymbol{\mathcal{X}}^{t+1}-\boldsymbol{\mathcal{X}}^{t}\|_F. \end{cases}$$

Proof: Suppose $\boldsymbol{\mathcal{R}}^{t+1},\boldsymbol{\mathcal{Z}}^{t+1},\boldsymbol{\mathcal{M}}^{t+1}$ and $\boldsymbol{\mathcal{X}}^{t+1}$ are the optimal solutions of F_1,F_2,F_3, and F_4 in the optimization problem (8), then we have

$$\begin{aligned} &0\in\partial f_1(\boldsymbol{\mathcal{R}}^{t+1})+\nabla_{\boldsymbol{\mathcal{R}}}f(\boldsymbol{\mathcal{R}}^{t+1},\boldsymbol{\mathcal{Z}}^{t},\boldsymbol{\mathcal{M}}^{t},\boldsymbol{\mathcal{X}}^{t}) \\ &+\eta(\boldsymbol{\mathcal{R}}^{t+1}-\boldsymbol{\mathcal{R}}^{t}), \\ &0\in\partial f_2(\boldsymbol{\mathcal{Z}}^{t+1})+\nabla_{\boldsymbol{\mathcal{Z}}}f(\boldsymbol{\mathcal{R}}^{t+1},\boldsymbol{\mathcal{Z}}^{t+1},\boldsymbol{\mathcal{M}}^{t},\boldsymbol{\mathcal{X}}^{t}) \\ &+\eta(\boldsymbol{\mathcal{Z}}^{t+1}-\boldsymbol{\mathcal{Z}}^{t}). \end{aligned}$$

Additionally, using the subproblem (23) and (25), we get that

$$\begin{aligned} 0&=\lambda\sum_{p=1}^{P}ST_p^T(ST_p(\boldsymbol{\mathcal{M}}^{t+1})-ST_p(\tilde{\boldsymbol{\mathcal{R}}}^{t+1})) \\ &+\rho(\boldsymbol{\mathcal{M}}^{t+1}-\boldsymbol{\mathcal{Z}}^{t+1})-\tau(\boldsymbol{\mathcal{X}}^{t}_{\times 3}\boldsymbol{E}^T-\boldsymbol{\mathcal{M}}^{t+1}) \\ &+\eta(\boldsymbol{\mathcal{M}}^{t+1}-\boldsymbol{\mathcal{M}}^{t}), \\ 0&=\boldsymbol{\mathcal{A}}^T((\boldsymbol{\mathcal{A}}(\boldsymbol{\mathcal{X}})-\boldsymbol{\mathcal{Y}})+\tau(\boldsymbol{\mathcal{X}}^{t+1}-\boldsymbol{\mathcal{M}}^{t+1}_{\times 3}\boldsymbol{E}) \\ &+\eta(\boldsymbol{\mathcal{X}}^{t+1}-\boldsymbol{\mathcal{X}}^{t}). \end{aligned}$$

Then, the $\boldsymbol{\mathcal{V}}_1,\boldsymbol{\mathcal{V}}_2,\boldsymbol{\mathcal{V}}_3$, and $\boldsymbol{\mathcal{V}}_4$ are defined as follows

$$\begin{aligned} \boldsymbol{\mathcal{V}}_1^{t+1}&=-\nabla_{\boldsymbol{\mathcal{R}}}f(\boldsymbol{\mathcal{R}}^{t+1},\boldsymbol{\mathcal{Z}}^{t},\boldsymbol{\mathcal{M}}^{t},\boldsymbol{\mathcal{X}}^{t}) \\ &\quad-\eta\|\boldsymbol{\mathcal{R}}^{t+1}-\boldsymbol{\mathcal{R}}^{t}\|_F, \\ \boldsymbol{\mathcal{V}}_2^{t+1}&=-\nabla_{\boldsymbol{\mathcal{Z}}}f(\boldsymbol{\mathcal{R}}^{t+1},\boldsymbol{\mathcal{Z}}^{t+1},\boldsymbol{\mathcal{M}}^{t},\boldsymbol{\mathcal{X}}^{t})\|_F \\ &\quad-\eta\|\boldsymbol{\mathcal{Z}}^{t+1}-\boldsymbol{\mathcal{Z}}^{t}\|_F, \\ \boldsymbol{\mathcal{V}}_3^{t+1}&=0, \\ \boldsymbol{\mathcal{V}}_4^{t+1}&=0. \end{aligned}$$

It's easy to know that $\boldsymbol{\mathcal{V}}_1^{t+1}\in\partial f_1(\boldsymbol{\mathcal{R}}^{t+1})$ and $\boldsymbol{\mathcal{V}}_2^{t+1}\in\partial f_2(\boldsymbol{\mathcal{Z}}^{t+1})$. Thus, we have

$$\begin{aligned} &\|\boldsymbol{\mathcal{V}}_1^{t+1}+\nabla_{\boldsymbol{\mathcal{R}}}f(\boldsymbol{\mathcal{R}}^{t+1},\boldsymbol{\mathcal{Z}}^{t},\boldsymbol{\mathcal{M}}^{t},\boldsymbol{\mathcal{X}}^{t})\|_F \\ &\quad=\eta\|\boldsymbol{\mathcal{R}}^{t+1}-\boldsymbol{\mathcal{R}}^{t}\|_F\le\eta\|\boldsymbol{\mathcal{R}}^{t+1}-\boldsymbol{\mathcal{R}}^{t}\|_F, \\ &\|\boldsymbol{\mathcal{V}}_2^{t+1}+\nabla_{\boldsymbol{\mathcal{Z}}}f(\boldsymbol{\mathcal{R}}^{t+1},\boldsymbol{\mathcal{Z}}^{t+1},\boldsymbol{\mathcal{M}}^{t},\boldsymbol{\mathcal{X}}^{t})\|_F \\ &\quad=\eta\|\boldsymbol{\mathcal{Z}}^{t+1}-\boldsymbol{\mathcal{Z}}^{t}\|_F\le\eta\|\boldsymbol{\mathcal{Z}}^{t+1}-\boldsymbol{\mathcal{Z}}^{t}\|_F, \\ &\|\boldsymbol{\mathcal{V}}_3^{t+1}+\nabla_{\boldsymbol{\mathcal{M}}}f(\boldsymbol{\mathcal{R}}^{t+1},\boldsymbol{\mathcal{Z}}^{t+1},\boldsymbol{\mathcal{M}}^{t+1},\boldsymbol{\mathcal{X}}^{t})\|_F \\ &\quad\le\eta\|\boldsymbol{\mathcal{M}}^{t+1}-\boldsymbol{\mathcal{M}}^{t}\|_F, \\ &\|\boldsymbol{\mathcal{V}}_4^{t+1}+\nabla_{\boldsymbol{\mathcal{X}}}f(\boldsymbol{\mathcal{R}}^{t+1},\boldsymbol{\mathcal{Z}}^{t+1},\boldsymbol{\mathcal{M}}^{t+1},\boldsymbol{\mathcal{X}}^{t+1})\|_F \\ &\quad\le\eta\|\boldsymbol{\mathcal{X}}^{t+1}-\boldsymbol{\mathcal{X}}^{t}\|_F. \end{aligned}$$

This completes the proof. ■

Proof of Theorem 2: Firstly, the Lemma A1 shows that the function $L(\boldsymbol{\mathcal{R}},\boldsymbol{\mathcal{Z}},\boldsymbol{\mathcal{M}},\boldsymbol{\mathcal{X}})$ satisfies the KŁ property at each point. Secondly, according to Lemmas A2 and A3, we can know that the bounded sequence $\{\boldsymbol{\mathcal{R}}^{t},\boldsymbol{\mathcal{Z}}^{t},\boldsymbol{\mathcal{M}}^{t},\boldsymbol{\mathcal{X}}^{t}\}$ satisfies the sufficient decrease condition and relative error condition. By verifying that the above three key conditions are established, it shows that the proposed LRNDP algorithm satisfies Theorem 6.2 in [1]. Therefore, the bounded sequence $\{\boldsymbol{\mathcal{R}}^{t},\boldsymbol{\mathcal{Z}}^{t},\boldsymbol{\mathcal{M}}^{t},\boldsymbol{\mathcal{X}}^{t}\}$ globally converges to a critical point of $L(\boldsymbol{\mathcal{R}},\boldsymbol{\mathcal{Z}},\boldsymbol{\mathcal{M}},\boldsymbol{\mathcal{X}})$. This completes the proof of Theorem 2. ■

B. Parameter Analysis

To illustrate the impact of each parameter on the quality of the reconstructed image, we plot the PSNR and SSIM curves under different parameters for 1×10^6 incident photons in Fig. 1. As can be seen from Figs. 1 (a)-(d), the parameters α and β have a significant impact on the quality of the reconstructed image, and the PSNR and SSIM obtain satisfactory results when $\alpha=1\times10^{-5}$ and $\beta=5\times10^{-4}$. Figs. 1 (e)-(n) show how the balance factors λ,ρ,ρ_1,τ and η affect the changing trends of PSNR and SSIM. Through parameter adjustment, it is determined that the best reconstruction results can be obtained when $\lambda=1\times10^{-3},\rho=5\times10^{-5},\rho_1=5\times10^{-4},\tau=1\times10^{-2}$ and $\eta=1\times10^{-4}$. In Figs. 1 (o) and

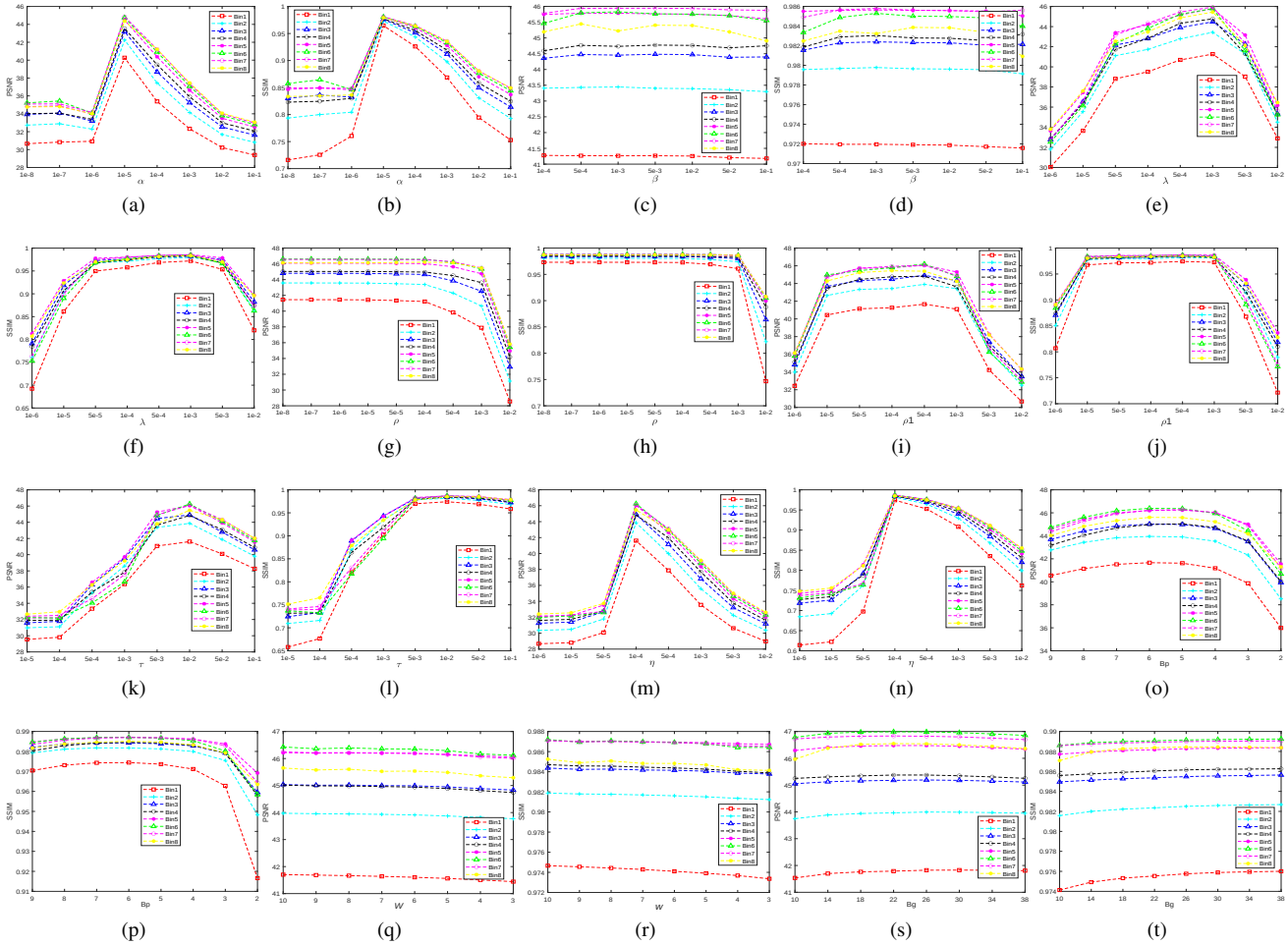


Fig. 1: Parameter analysis of the proposed LRNDP algorithm. (a) and (b), (c) and (d), (e) and (f), (g) and (h), (i) and (j), (k) and (l), (m) and (n), (o) and (p), (q) and (r), (s) and (t) plot the PSNR and SSIM for different setting of the parameter $\alpha, \beta, \lambda, \rho, \rho_1, \tau\ \eta, B_p, W$ and B_g, respectively.

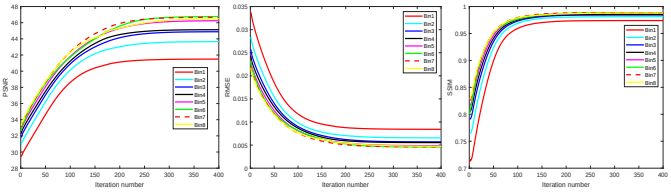
(p), we compare the PSNR and SSIM values for patch sizes ranging from 3×3 to 9×9. The experimental results show that when the patch size B_p is set to 6×6, the best reconstruction effect can be achieved. Further, Figs. 1 (q)-(t) reveal the impact of the search window size and the number of similar patches on the reconstruction performance. After careful analysis, we find that when the search window size W is 7 and the number of similar patches B_g is 30, the minimum PSNR value and the maximum SSIM value can be obtained in all energy channels. At the same time, we also set the step size for extracting similar patches S to 4. Based on the above detailed analysis and experimental results, this paper sets the parameters to the following optimal values: $\alpha = 1\times10^{-5}, \beta = 5\times10^{-4}, \lambda = 1\times10^{-3}, \rho = 5\times10^{-5}, \rho_1 = 5\times10^{-4}, \tau = 1\times10^{-2}$ $\eta = 1\times10^{-4}, B_p = 36, W = 7$ and $B_g = 30$. For 1×10^5 incident photons, we set $\beta = 5\times10^{-4}, \rho_1 = 5\times10^{-4}, \tau = 5\times10^{-3}$ and $\eta = 1\times10^{-6}$ for thoracic images and abdominal images, and $\beta = 5\times10^{-2}, \rho_1 = 1\times10^{-3}, \tau = 5\times10^{-3}$ and $\eta = 1\times10^{-3}$ for mouse thorax phantom. At the same time, we set $\rho = 5\times10^{-4}$, $\rho = 1\times10^{-3}$, and $\rho = 5\times10^{-3}$ for thoracic images, abdominal images, and mouse thorax phantom, respectively. The settings of other parameters are the same as those for the case of 1×10^6 incident photons.

C. Convergence Analysis

In Theorem 2, we establish the convergence guarantee of the proposed algorithm from a theoretical perspective. Here, we utilize the stability of experimental results to further verify the proposed algorithm's convergence. Taking the thoracic images as example, Fig. 2 plots the curves of the PSNR, RMSE, and SSIM versus the iterations under 65 projection views and 1×10^6 photons. We observe that the PSNR, RMSE and SSIM curves of images with different energy channels eventually converged to a stable solution after 300 iterations as the iteration number increasing. Hence, this paper sets the number of iterations to 300.

D. Comparative Study of Pre-trained and Re-trained models

In this subsection, we utilize thoracic and abdominal images as examples to contrast the reconstruction capabilities of our proposed method when applied under both a pre-trained model and a re-trained model. To facilitate this comparison, we gather 5794 medical images specifically for re-training FFDNet.



(a) (b) (c)

Fig. 2: The figure plots PSNR, RMSE and SSIM versus the iterations for reconstructed images of various energy channels under 65 views and noisy projection data. (a) PSNR, (b) RMSE, (c) SSIM.

Subsequently, we train a set of denoisers across a noise level range of 0 to 75, incorporating these into our method, designated as LRNDP1. For incident photons of 1×10^6 and 1×10^5, Table A1 presents quantitative results for the LRNDP and LRNDP1 methods, obtained from 65 projection views. The results indicate that the FFDNet model, when re-trained for specific applications and noise characteristics, can capture image details more accurately while effectively reducing noise and artifacts. Relatively speaking, the LRNDP1 method can capture complex image features more effectively than the LRNDP method, thereby achieving better image reconstruction effects.

TABLE A1: PSNR, RMSE and SSIM values by the proposed LRNDP and LRNDP1 methods.

Datasets	Noise levels	Channel Number	LRNDP			LRNDP1		
			PSNR	RMSE	SSIM	PSNR	RMSE	SSIM
Thoracic images	Noise level I 1×10^6 photons	Bin1	41.80	0.0081	0.9758	41.83	0.0081	0.9759
		Bin2	43.94	0.0064	0.9825	43.99	0.0063	0.9826
		Bin3	45.12	0.0055	0.9855	45.18	0.0055	0.9856
		Bin4	45.29	0.0054	0.9862	45.36	0.0054	0.9862
		Bin5	46.37	0.0048	0.9883	46.46	0.0048	0.9883
		Bin6	46.89	0.0045	0.9891	46.97	0.0045	0.9892
		Bin7	46.74	0.0046	0.9890	46.81	0.0046	0.9890
		Bin8	46.42	0.0048	0.9884	46.51	0.0047	0.9884
	Noise level II 1×10^5 photons	Bin1	37.39	0.0135	0.9545	37.69	0.0130	0.9553
		Bin2	39.66	0.0104	0.9664	40.01	0.0100	0.9674
		Bin3	41.12	0.0088	0.9723	41.46	0.0085	0.9732
		Bin4	41.91	0.0080	0.9754	42.21	0.0078	0.9761
		Bin5	42.29	0.0077	0.9764	42.55	0.0075	0.9769
		Bin6	41.52	0.0084	0.9731	41.94	0.0080	0.9738
		Bin7	43.32	0.0068	0.9795	43.53	0.0067	0.9799
		Bin8	43.32	0.0068	0.9794	43.54	0.0067	0.9798
Abdominal images	Noise level I 1×10^6 photons	Bin1	44.81	0.0057	0.9821	44.86	0.0057	0.9823
		Bin2	46.99	0.0045	0.9868	47.08	0.0044	0.9871
		Bin3	47.56	0.0042	0.9877	47.61	0.0042	0.9878
		Bin4	48.30	0.0038	0.9892	48.36	0.0038	0.9893
		Bin5	49.28	0.0034	0.9910	49.37	0.0034	0.9911
		Bin6	49.56	0.0033	0.9914	49.64	0.0033	0.9915
		Bin7	49.56	0.0033	0.9914	49.61	0.0033	0.9915
		Bin8	49.07	0.0035	0.9907	49.08	0.0035	0.9907
	Noise level II 1×10^5 photons	Bin1	39.86	0.0102	0.9589	39.89	0.0101	0.9590
		Bin2	42.40	0.0076	0.9709	42.46	0.0075	0.9710
		Bin3	43.93	0.0064	0.9765	44.05	0.0063	0.9768
		Bin4	44.70	0.0058	0.9793	44.87	0.0057	0.9796
		Bin5	45.11	0.0056	0.9803	45.26	0.0055	0.9806
		Bin6	45.23	0.0055	0.9804	45.49	0.0053	0.9810
		Bin7	45.81	0.0051	0.9825	45.94	0.0050	0.9827
		Bin8	46.46	0.0048	0.9843	46.60	0.0047	0.9846

REFERENCES

[1] H. Attouch, J. Bolte, and B. F. Svaiter, "Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized gauss–seidel methods," *Mathematical Programming*, vol. 137, no. 1, pp. 91–129, 2013.

[2] J. Bolte, S. Sabach, and M. Teboulle, "Proximal alternating linearized minimization for nonconvex and nonsmooth problems," *Mathematical Programming*, vol. 146, no. 1, pp. 459–494, 2014.