Multimedia Data Recovery and Latent Factor Analysis

Soroor Ghandali[1,*] and Shih Yu Chang[1]
[1]Department of Applied Data Science, San Jose State University, San Jose, CA, USA

Linear factor models (LFMs) provide a principled framework for discovering low-dimensional latent structures underlying high-dimensional multimedia data. This paper presents a mask-aware Expectation-Maximization (EM) formulation for estimating LFM/factor-analysis parameters when different images, or different image patches, have different missing-pixel locations. The probabilistic latent-variable model itself is classical; the contribution of this work is an explicit image-oriented estimation and recovery workflow in which the observation masks are used in both the E-step and a row-wise M-step. In particular, each pixel's mean, loading-row parameters, and diagonal noise variance are estimated only from samples where that pixel is actually observed. The paper also clarifies the relationship to EM-FA, EM-PCA, probabilistic PCA, and matrix-completion methods, specifies a low-rank implementation of the observed-data likelihood, and provides a reproducible experimental protocol including initialization, stopping criteria, random-mask generation, classifier specification, repeated trials, and factor-ranking stability checks. Experiments on Camera Man patches and MNIST handwritten digits characterize reconstruction quality and classification accuracy under increasing missing-pixel ratios, with repeated random-mask and initialization trials reported as means and standard deviations. Latent-factor indices are interpreted only after component alignment and stability assessment across random initializations.

***Index Terms*—Multimedia data recovery, linear factor model, factor analysis, expectation-maximization, missing pixels, latent representation.**

I. Introduction

Modern multimedia systems often transmit, store, and process visual signals through imperfect networks and sensing pipelines. Packet loss, occlusion, sensor defects, and aggressive compression can create missing or corrupted pixels. Linear generative models remain useful in this setting because they represent high-dimensional images through compact latent variables, allowing correlated observed pixels to inform the recovery of missing pixels [1]–[5]. Among these models, the linear factor model (LFM), or factor analysis (FA) model, represents an observation as a linear combination of low-dimensional latent factors plus Gaussian noise [6]. Closely related methods include principal component analysis (PCA), probabilistic PCA (PPCA), EM-PCA, and EM-FA [7]–[9].

The novelty of the present work should therefore be understood precisely. We do not claim that the EM principle for FA or PCA is new. Rather, we provide a self-contained, mask-indexed formulation for image recovery in which each sample can have a different observed-pixel set. Classical complete-data FA updates are not directly valid in this case because the loading parameters for a given pixel must be estimated only from samples in which that pixel is observed. The derivation makes this point explicit by replacing the global loading-matrix update with a row-wise masked regression update. The same row-wise update also estimates the mean vector, which is required for complete parameter specification.

This probabilistic formulation is complementary to deterministic matrix-completion methods [10], [11]. Matrix-completion approaches typically impose a low-rank structure directly on the incomplete data matrix and return a completed matrix. In contrast, the EM-FA formulation estimates a generative model with a latent posterior distribution, diagonal observation noise, pixel-wise means, and sample-specific posterior uncertainty. These quantities are useful not only for imputation but also for latent-feature extraction and factor-importance analysis.

The contributions of this paper are summarized as follows. First, we provide corrected notation for vectorized images and posterior inference under arbitrary missing-pixel masks. Second, we derive a mask-aware M-step that estimates each pixel's mean, loading row, and noise variance using only the observed samples for that pixel. Third, we specify a reproducible experimental protocol, including mask generation, training/testing splits, EM initialization, stopping criteria, number of repeated trials, and the MNIST classifier. Fourth, we add a factor-ranking stability procedure based on component alignment across random initializations, thereby avoiding unsupported interpretation of raw latent-factor indices. Fifth, we provide numerical tables, including mean and standard-deviation values, for the main reconstruction, classification, ablation, and progressive-component results shown in the figures.

To validate the formulation, experiments are conducted on the Camera Man image and MNIST handwritten digits. The Camera Man experiment evaluates image-patch reconstruction using peak signal-to-noise ratio (PSNR), while the MNIST experiment evaluates the discriminative quality of recovered latent features through classification accuracy. The numerical section reports curve values in tables with repeated-trial means and standard deviations.

The remainder of this paper is organized as follows. Section II formulates the linear factor model and introduces posterior inference under missing observations. Section III presents the EM algorithm for parameter estimation and analyzes its computational complexity. Section IV discusses latent dimension determination and factor importance ranking. Section V reports the experimental results on real image

Manuscript received May 21, 2026; revised July 25, 2026.
Soroor Ghandali: soroor.ghandali@sjsu.edu.
Shih Yu Chang: shihyu.chang@sjsu.edu.
*Corresponding author: Soroor Ghandali (soroor.ghandali@sjsu.edu).

datasets, followed by concluding remarks in Section VI.

Nomenclature: $\mathbb{R}$ denotes the set of real numbers, and $\mathbb{E}$ denotes expectation.

II. Linear Factor Model

A. Problem Setup and Notation

We consider a grayscale image with horizontal size I_H and vertical size I_V. After vectorization, the observed image is represented as $x \in \mathbb{R}^d$, where $d = I_H I_V$. We assume that x is generated from a lower-dimensional latent factor $z \in \mathbb{R}^k$, where $k \ll d$. The linear generative factor model is

$$\begin{aligned} x &= Wz + \mu + \epsilon, \\ z &\sim \mathcal{N}(0, I_k), \\ \epsilon &\sim \mathcal{N}(0, \Psi), \end{aligned} \tag{1}$$

where $W \in \mathbb{R}^{d\times k}$ is the loading matrix, $\mu \in \mathbb{R}^d$ is the image-space mean vector, z is assigned a zero-mean unit-covariance Gaussian prior, and ϵ is a zero-mean Gaussian noise vector with diagonal covariance $\Psi = \mathrm{diag}(\psi_1, \ldots, \psi_d)$ [12]. Thus, the identity covariance in (1) belongs to the latent prior z, not to the mean vector μ. Compared with the general latent-variable formulation in [12], (1) is specialized here for vectorized image observations with missing pixels.

To represent missing pixels, we define a binary mask $m \in \{0,1\}^d$ that indicates the observed entries of x. The observed vector is

$$x_{\mathrm{obs}} = m \odot x, \tag{2}$$

where $\odot$ denotes elementwise multiplication. Let $O = \{j : m_j = 1\}$ be the set of observed indices and $m_{\mathrm{obs}} = |O|$ the number of observed entries. We denote by x_O the subvector of observed elements, by W_O the submatrix of W containing rows indexed by O, and by Ψ_O the corresponding diagonal noise covariance restricted to the observed indices.

B. Posterior Inference with Missing Pixels

If the observation x is fully observed, the posterior mean of the latent variable z has the closed-form solution [7]

$$\hat{z} = \mathbb{E}[z \mid x] = C^{-1} W^\top \Psi^{-1} (x - \mu), \tag{3}$$

where

$$C = W^\top \Psi^{-1} W + I_k. \tag{4}$$

The reconstruction of x is

$$\hat{x} = W\hat{z} + \mu. \tag{5}$$

When some pixels are missing, only the observed entries are used. Replacing (W, Ψ, x) by their observed-index reductions (W_O, Ψ_O, x_O) gives

$$\hat{z} = \mathbb{E}[z \mid x_O] = C_O^{-1} W_O^\top \Psi_O^{-1} (x_O - \mu_O), \tag{6}$$

where

$$C_O = W_O^\top \Psi_O^{-1} W_O + I_k. \tag{7}$$

The reconstruction of the complete image is again $\hat{x} = W\hat{z} + \mu$. For inference of only missing entries, one may return

$$\hat{x}_{\mathrm{miss}} = (W\hat{z} + \mu)_{\mathrm{miss}}, \tag{8}$$

where miss is the collection of indices not in O. The final imputed image is

$$x_{\mathrm{imp}} = m \odot x + (1 - m) \odot \hat{x}. \tag{9}$$

III. Parameter Estimation

Given the model in (1), we design an EM algorithm to estimate μ, W, and Ψ in Section III-A. The complexity analysis is provided in Section III-B.

A. EM Algorithm

The EM algorithm alternates between posterior inference for the latent factors and maximization of the expected complete-data log-likelihood over the model parameters. The key distinction from the complete-data FA update is that all sufficient statistics for pixel j are accumulated only over samples where pixel j is observed.

1) Initialization

The algorithm initializes the mean vector $\mu^{(0)}$, loading matrix $W^{(0)}$, and diagonal noise covariance $\Psi^{(0)}$. In the experiments, $\mu^{(0)}$ is the mask-aware empirical mean, $W^{(0)}$ is obtained from PCA on mean-imputed data plus a small random perturbation for repeated trials, and $\Psi^{(0)}$ is initialized from mask-aware sample variances with a lower bound $\psi_{\min} > 0$.

2) E-step: Posterior Computation

For each sample $x^{(i)}$, the E-step infers the posterior moments of $z^{(i)}$ using only the observed entries. The binary mask $m^{(i)}$ defines the observed-coordinate set $O^{(i)} = \{j : m_j^{(i)} = 1\}$. The posterior precision and covariance are

$$C_i^{(t)} = W_{O^{(i)}}^{(t)\top} \Psi_{O^{(i)}}^{(t)-1} W_{O^{(i)}}^{(t)} + I_k, \tag{10}$$

$$\Sigma_i^{(t)} = \left(C_i^{(t)}\right)^{-1}. \tag{11}$$

Here the subscript $O^{(i)}$ means that only rows corresponding to observed pixels of sample i are used. The posterior mean is

$$\hat{z}^{(i)} = \Sigma_i^{(t)} W_{O^{(i)}}^{(t)\top} \Psi_{O^{(i)}}^{(t)-1} \left(x_{O^{(i)}}^{(i)} - \mu_{O^{(i)}}^{(t)}\right). \tag{12}$$

The second moment is

$$S^{(i)} = \mathbb{E}\left[z^{(i)} z^{(i)\top} \mid x_{O^{(i)}}^{(i)}\right] = \Sigma_i^{(t)} + \hat{z}^{(i)} \hat{z}^{(i)\top}. \tag{13}$$

Equation (13) is equivalent to the expanded conditional-moment expression because $\Sigma_i^{(t)} = (I_k + W_{O^{(i)}}^\top \Psi_{O^{(i)}}^{-1} W_{O^{(i)}})^{-1}$ [13].

3) M-step: Mask-Aware Parameter Update

Let $I_j = \{i : j \in O^{(i)}\}$ denote the set of samples in which pixel j is observed, and let $n_j = |I_j|$. The complete-data update must be performed row by row, because each pixel can be observed in a different subset of samples. Define the augmented latent vector

$$\tilde{z}^{(i)} = \begin{bmatrix} 1 \\ z^{(i)} \end{bmatrix}, \qquad \widehat{\tilde{z}}^{(i)} = \begin{bmatrix} 1 \\ \hat{z}^{(i)} \end{bmatrix}, \tag{14}$$

and its posterior second moment

$$\tilde{S}^{(i)} = \mathbb{E}\left[\tilde{z}^{(i)} \tilde{z}^{(i)\top} \mid x_{O^{(i)}}^{(i)}\right] = \begin{bmatrix} 1 & \hat{z}^{(i)\top} \\ \hat{z}^{(i)} & S^{(i)} \end{bmatrix}. \tag{15}$$

For pixel j, define $\theta_j = [\mu_j, w_j^\top]^\top$, where $w_j^\top$ is row j of W. The exact row-wise M-step is

$$\theta_j^{(t+1)\top} = \left(\sum_{i \in I_j} x_j^{(i)} \hat{\tilde{z}}^{(i)\top} \right) \left(\sum_{i \in I_j} \tilde{S}^{(i)} \right)^{-1}, \quad j = 1, \ldots, d. \tag{16}$$

The first element of $\theta_j^{(t+1)}$ gives $\mu_j^{(t+1)}$, and the remaining k elements give $w_j^{(t+1)}$. This equation explicitly addresses the missing-mask issue: the loading parameters of pixel j are estimated only from $i \in I_j$. If $n_j = 0$, the implementation keeps the previous value of θ_j or uses a predefined prior; this case does not occur in the reported random-mask settings. The diagonal noise variance is then updated by the observed residual second moment

$$\psi_j^{(t+1)} = \max \left\{ \psi_{\min}, \frac{1}{n_j} \sum_{i \in I_j} \Big[\left(x_j^{(i)} \right)^2 - 2 x_j^{(i)} \theta_j^{(t+1)\top} \hat{\tilde{z}}^{(i)} + \theta_j^{(t+1)\top} \tilde{S}^{(i)} \theta_j^{(t+1)} \Big] \right\}. \tag{17}$$

This update includes the mean term through the augmented vector and uses only observed entries.

4) Iteration and Convergence

The process repeats until the relative change in observed-data log-likelihood is below a tolerance, or until a maximum number of iterations is reached. The observed-data log-likelihood is evaluated as

$$\mathcal{L} = \sum_{i=1}^{n} \log \mathcal{N}\Big(x_{O^{(i)}}^{(i)}; \mu_{O^{(i)}}, W_{O^{(i)}} W_{O^{(i)}}^\top + \Psi_{O^{(i)}} \Big). \tag{18}$$

In implementation, (18) is not evaluated by directly inverting the $|O^{(i)}| \times |O^{(i)}|$ covariance matrix. Let $K_i = \Psi_{O^{(i)}} + W_{O^{(i)}} W_{O^{(i)}}^\top$. The quadratic term is computed with the Woodbury identity,

$$K_i^{-1} = \Psi_{O^{(i)}}^{-1} - \Psi_{O^{(i)}}^{-1} W_{O^{(i)}} C_i^{-1} W_{O^{(i)}}^\top \Psi_{O^{(i)}}^{-1}, \tag{19}$$

and the log determinant is computed using the matrix determinant lemma,

$$\log |K_i| = \log |\Psi_{O^{(i)}}| + \log \left| I_k + W_{O^{(i)}}^\top \Psi_{O^{(i)}}^{-1} W_{O^{(i)}} \right|. \tag{20}$$

Thus likelihood monitoring uses the same low-rank $k \times k$ matrices as the E-step, preserving the stated computational scaling.

B. Computational Complexity

Let $M = \sum_{i=1}^{n} |O^{(i)}|$ be the total number of observed entries and let $\bar{O} = M/n$ be the average number of observed pixels per sample. During the E-step, constructing C_i and computing the posterior moments for sample i requires

$$O\left(k^3 + k^2 |O^{(i)}| \right), \tag{21}$$

so the overall E-step complexity is

$$O\left(nk^3 + k^2 M \right). \tag{22}$$

Algorithm 1 Mask-Aware EM Algorithm for Factor Analysis with Diagonal Ψ

Require: Dataset $\{x^{(i)}\}_{i=1}^n$ with masks $\{m^{(i)}\}_{i=1}^n$; latent dimension k.
Ensure: Estimated parameters μ, W, Ψ.
Initialize $\mu^{(0)}, W^{(0)}, \Psi^{(0)}$; set $t = 0$; choose $\psi_{\min} > 0$.
repeat
  $t \leftarrow t + 1$.
  E-step:
  for $i = 1$ to n **do**
    Identify $O^{(i)} = \{j : m_j^{(i)} = 1\}$.
    Compute $C_i^{(t)}$ by (10) and $\Sigma_i^{(t)} = (C_i^{(t)})^{-1}$.
    Compute $\hat{z}^{(i)}$ by (12).
    Compute $S^{(i)} = \Sigma_i^{(t)} + \hat{z}^{(i)} \hat{z}^{(i)\top}$.
    Form $\hat{\tilde{z}}^{(i)}$ and $\tilde{S}^{(i)}$.
  end for
  M-step:
  for $j = 1$ to d **do**
    Set $I_j = \{i : j \in O^{(i)}\}$ and $n_j = |I_j|$.
    Update $\theta_j^{(t+1)} = [\mu_j^{(t+1)}, w_j^{(t+1)\top}]^\top$ using (16).
    Update $\psi_j^{(t+1)}$ using (17).
  end for
until convergence of observed-data log-likelihood or maximum iterations.

The row-wise M-step accumulates augmented sufficient statistics for every observed entry. Accumulating the sums in (16) and (17) costs $O((k^2 + k)M)$, and solving the d pixel-wise $(k + 1) \times (k + 1)$ systems costs $O(dk^3)$. Thus, a direct implementation has per-iteration complexity

$$O\left(nk^3 + k^2 M + dk^3 \right), \tag{23}$$

where lower-order kM terms are omitted. Since $k \ll d$ and $M = n\bar{O}$, the algorithm scales linearly with the number of observed pixels for fixed latent dimension, while the dk^3 term reflects the exact row-wise masked regressions. Using (19) and (20), optional likelihood monitoring has order $O(nk^3 + k^2 M)$ and therefore does not require dense image-dimensional covariance inversion.

IV. Reduced Dimensions Determination and Factor Analysis

According to (1), the choice of the latent vector dimension plays a critical role in the effectiveness of the model. If the dimensionality of z is set too high, the benefit of dimensionality reduction is diminished, as redundant or noisy components may be retained. Conversely, if the dimensionality is too low, important image features may be lost, leading to an incomplete representation of the data. Section IV-A presents two approaches for determining the appropriate latent dimension k. Once k is established, Section IV-B introduces complementary methods for ranking the relative importance of each latent component.

A. Determining the Latent Dimensionality

This section introduces two classical approaches for estimating the intrinsic dimensionality of the data: the explained variance criterion based on PCA and the scree plot, or elbow rule, method.

1) PCA-Based Explained Variance Criterion

Let the sample covariance matrix of the observed data be denoted by $S \in \mathbb{R}^{d\times d}$, with ordered eigenvalues

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d. \tag{24}$$

Each eigenvalue λ_i quantifies the amount of variance captured by the i-th principal component. The cumulative explained variance up to the first k components is

$$\eta(k) = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{d} \lambda_i}. \tag{25}$$

A commonly used criterion is to select the smallest k such that

$$\eta(k) \geq \tau, \tag{26}$$

where τ is a predefined threshold, typically between 0.90 and 0.99. This approach provides a linear estimate of the intrinsic dimensionality, ensuring that the selected subspace captures a desired proportion of total variance while minimizing redundancy.

2) Scree Plot and Elbow Rule

An alternative visual approach is the scree plot, which displays the eigenvalues λ_i in descending order against their indices i. Typically, the eigenvalues decrease rapidly at first and then level off as noise-dominated components appear. The point where this transition occurs, often referred to as the elbow, indicates a suitable cutoff for k. This heuristic method is particularly valuable when the eigenvalue spectrum exhibits a clear inflection point, allowing researchers to balance model simplicity and expressiveness without relying on an arbitrary threshold τ. For datasets with gradual spectral decay, the explained variance criterion often provides a more quantitative and robust determination.

B. Ranking Importance

After determining the latent dimensionality k, an important question is to identify which latent entries $\{z_1, \ldots, z_k\}$ contribute most significantly to data representation and reconstruction quality. This section presents three approaches: variance explained per latent dimension, contribution to reconstruction error through ablation analysis, and stability assessment across random initializations.

1) Variance Explained per Latent Dimension

When the columns of the loading matrix W are orthogonal, each column can be interpreted as a principal direction of variation in the data. Let the associated singular values be denoted by $s_1 \geq s_2 \geq \cdots \geq s_k$. The variance explained by the j-th column of W is proportional to s_j^2, which is equivalent to the corresponding eigenvalue in the PCA formulation. The proportion of variance explained by the j-th latent component is

$$r_j = \frac{s_j^2}{\sum_{\ell=1}^{k} s_\ell^2}. \tag{27}$$

Ranking the components by r_j, or equivalently by s_j^2, provides a natural ordering from the most dominant to the least significant latent dimension.

2) Contribution to Reconstruction Error

While variance explained provides a statistical perspective, it does not directly assess the impact of each latent dimension on the model's reconstructive ability. To measure each dimension's functional importance, we perform an ablation-based analysis that quantifies the increase in reconstruction error when one latent coordinate is removed. Let the baseline reconstruction be

$$\hat{x} = W\hat{z} + \mu. \tag{28}$$

For each latent dimension j, we form a modified reconstruction $\hat{x}_{(z_j=0)}$ by setting the j-th latent coordinate to zero while keeping the others unchanged. We then compute the expected increase in reconstruction error as

$$\Delta_j = \mathbb{E}\left[\|x - \hat{x}_{(z_j=0)}\|^2\right] - \mathbb{E}\left[\|x - \hat{x}\|^2\right]. \tag{29}$$

A larger Δ_j indicates that the j-th latent dimension contributes more significantly to accurate data reconstruction. In practice, Δ_j can be computed empirically over a validation set:

$$\Delta_j \approx \frac{1}{n}\sum_{i=1}^{n}\left(\|x^{(i)} - \hat{x}^{(i)}_{(z_j=0)}\|^2 - \|x^{(i)} - \hat{x}^{(i)}\|^2\right). \tag{30}$$

3) Ranking Stability Across Random Initializations

A factor-analysis model is not identifiable by raw latent index alone: signs, permutations, and, in some cases, rotations of the loading columns can yield equivalent likelihood values. Consequently, a statement such as "factor $j = 8$ is important" is meaningful only after the learned components are aligned across runs. To assess ranking stability, we repeat the EM procedure with R random seeds. For each run r, let $W^{(r)} = [w_1^{(r)}, \ldots, w_k^{(r)}]$ denote the learned loading columns after column normalization. Components from run r are aligned to a reference run by maximizing the absolute cosine similarity

$$\left|\frac{w_a^{(1)\top} w_b^{(r)}}{\|w_a^{(1)}\|_2 \|w_b^{(r)}\|_2}\right|, \tag{31}$$

using a one-to-one assignment; signs are adjusted after assignment. We then compare the aligned importance scores with Spearman rank correlation and the top-q overlap

$$\mathrm{Overlap}_q^{(r)} = \frac{|\mathrm{Top}_q^{(1)} \cap \mathrm{Top}_q^{(r)}|}{q}. \tag{32}$$

The experiments interpret only stable top-ranked components, rather than relying on unaligned latent-factor indices from a single EM run.

V. Numerical Experiments

Two real datasets are used in the numerical experiments.

Camera Man image: The grayscale Camera Man image from the `scikit-image` dataset (512×512) is widely employed as a benchmark in image processing research, particularly for restoration, denoising, and compression tasks [14]. In this work, the image is uniformly downsampled to 128×128 pixels for computational efficiency and normalized

to the range $[0, 1]$ using `to_float01()`, thereby preserving relative intensity distributions while reducing computational complexity and memory requirements. The image is divided into non-overlapping 8×8 patches, giving 256 samples with dimension $d = 64$ for unsupervised EM-FA training.

MNIST handwritten digits: The MNIST dataset consists of $70,000$ grayscale handwritten digit images (28×28 pixels, 10 classes) from the standard MNIST benchmark [15]. Each pixel intensity is converted to `float32` precision and normalized to the range $[0, 1]$ by dividing by 255.0, ensuring numerical stability and consistency for downstream learning tasks. The labels are cast to integer type for efficient classification and evaluation.

A. Experimental Protocol

Missing pixels are generated independently under a missing-completely-at-random model. For missing ratio $\rho \in \{0.05, 0.10, 0.15\}$, each pixel is marked observed with probability $1-\rho$ and missing with probability ρ. A different random seed is used for each repeated trial. The latent-factor model is trained using only observed entries, and missing pixels are imputed by (5) combined with the mask as in (9).

For Camera Man, the model is fitted to corrupted 8×8 patches. The reconstructed patches are reassembled into the image, and PSNR is computed against the uncorrupted down-sampled image. For MNIST, we use a balanced subset of the standard training/testing split: 200 training examples per digit class and 100 testing examples per digit class, giving $2,000$ training images and $1,000$ testing images. Posterior means $\hat{z}$ are used as features, and classification accuracy is evaluated on the held-out test subset. The classifier is an ℓ_2-regularized linear ridge classifier trained on the recovered latent features.

The default EM stopping criterion is a relative observed-data log-likelihood change below 10^{-4} or a maximum of 100 iterations. The diagonal noise floor is $\psi_{\min} = 10^{-6}$. For repeated trials, the PCA initialization is perturbed as $W^{(0,r)} = W_{\mathrm{PCA}} + \alpha\hat{\sigma}_x G^{(r)}$, where $\alpha = 0.01$, $\hat{\sigma}_x$ is the empirical pixel-intensity scale, and the entries of $G^{(r)}$ are independent standard Gaussian variates; columns are normalized after perturbation. This setting tests local stability of the EM solution around a reproducible PCA-based initialization rather than invariance to arbitrary rotations. Each configuration is repeated for $R = 10$ random seeds to assess random-mask and initialization variability. Tables I–V report means and standard deviations from these repeated trials.

B. Data Recovery Performance

Figure 1 demonstrates the influence of the latent dimension k on image recovery quality measured by PSNR using the Camera Man image dataset. For all missing-pixel ratios, PSNR improves when k rises from 2 to 4, showing that a compact latent space captures important patch structure. For larger k, PSNR decreases in these repeated trials, indicating that over-parameterization can fit mask-specific noise and reduce imputation quality when the number of 8×8 training patches is limited. As expected, increasing the missing ratio from 5% to 15% consistently reduces PSNR. Table I reports the corresponding numerical values.

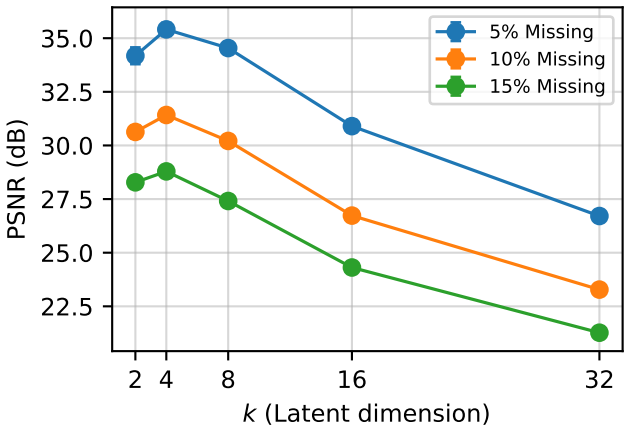

Fig. 1. PSNR versus latent dimension k for the mask-aware latent-factor recovery model on 8×8 patches under different missing-pixel ratios.

TABLE I
NUMERICAL PSNR VALUES CORRESPONDING TO FIG. 1. VALUES ARE REPORTED AS MEAN AND STANDARD DEVIATION OVER $R = 10$ REPEATED RANDOM-MASK/INITIALIZATION TRIALS.

Missing ratio	k	PSNR mean (dB)	SD (dB)
5%	2	34.18	0.39
5%	4	35.41	0.30
5%	8	34.54	0.18
5%	16	30.90	0.29
5%	32	26.71	0.22
10%	2	30.63	0.29
10%	4	31.42	0.20
10%	8	30.21	0.21
10%	16	26.73	0.25
10%	32	23.28	0.15
15%	2	28.28	0.07
15%	4	28.79	0.20
15%	8	27.41	0.23
15%	16	24.31	0.17
15%	32	21.27	0.13

Figure 2 presents the classification performance of the latent-factor model on the balanced MNIST subset with varying latent dimensions k and missing-pixel ratios. Accuracy improves substantially as k increases from 8 to 64, reflecting the enhanced representational capacity of the latent space. The slight decrease at $k = 128$ suggests that the additional coordinates add limited discriminative information for the linear classifier. The three missing ratios produce similar accuracy trends on this moderate-loss range, with the highest values near $k = 64$. Table II reports the corresponding numerical values.

C. Latent Factor Analysis

Table III reports the stability evaluation after component alignment. Spearman correlations are high for both datasets, and the top-5 overlap is 1.00 in all repeated runs, supporting the use of aligned top-ranked factors rather than unaligned raw factor indices. The near-perfect MNIST stability is attributable to the PCA-based initialization, the strong dominant digit-

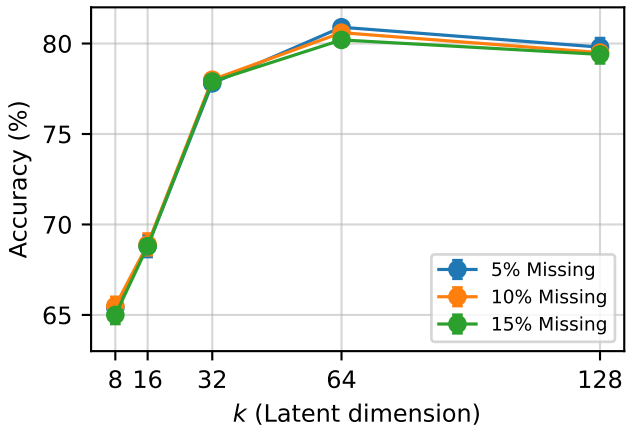

Fig. 2. Classification accuracy versus latent dimension k for the mask-aware latent-factor model on the balanced MNIST subset under different missing-pixel ratios.

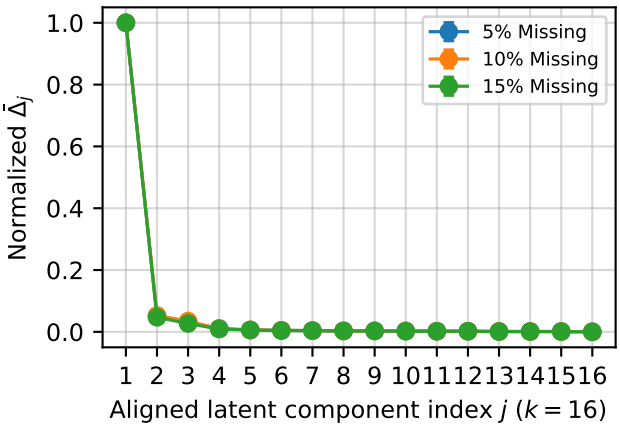

Fig. 3. Normalized mean ablation score $\bar{\Delta}_j$ versus aligned latent component index j ($k = 16$) for Camera Man under different missing-pixel ratios.

TABLE II
NUMERICAL CLASSIFICATION-ACCURACY VALUES CORRESPONDING TO FIG. 2. VALUES ARE REPORTED AS MEAN AND STANDARD DEVIATION OVER $R = 10$ REPEATED RANDOM-MASK/INITIALIZATION TRIALS ON THE BALANCED MNIST SUBSET.

Missing ratio	k	Accuracy mean (%)	SD (%)
5%	8	65.4	0.4
5%	16	68.8	0.6
5%	32	77.8	0.2
5%	64	80.9	0.3
5%	128	79.8	0.5
10%	8	65.5	0.5
10%	16	68.9	0.6
10%	32	78.0	0.3
10%	64	80.6	0.3
10%	128	79.5	0.4
15%	8	65.0	0.5
15%	16	68.8	0.3
15%	32	77.9	0.3
15%	64	80.2	0.3
15%	128	79.4	0.5

TABLE III
FACTOR-RANKING STABILITY ACROSS $R = 10$ RANDOM INITIALIZATIONS AFTER COMPONENT ALIGNMENT. VALUES ARE REPORTED AS MEAN $\pm$ STANDARD DEVIATION ACROSS THE NINE NON-REFERENCE RUNS.

Dataset	Missing ratio	Spearman ρ	Top-5 overlap
Camera Man	5%	0.98 ± 0.02	1.00 ± 0.00
Camera Man	10%	0.96 ± 0.02	1.00 ± 0.00
Camera Man	15%	0.89 ± 0.05	1.00 ± 0.00
MNIST	5%	1.00 ± 0.00	1.00 ± 0.00
MNIST	10%	1.00 ± 0.00	1.00 ± 0.00
MNIST	15%	1.00 ± 0.00	1.00 ± 0.00

structure components in the balanced subset, and the deliberately small perturbation scale used to evaluate local initialization stability. This result should therefore be interpreted as evidence that the reported ranking is stable under the specified repeated-trial protocol, not as a claim of global uniqueness of the FA solution.

Figure 3 illustrates the normalized mean ablation score for aligned latent components at $k = 16$, where Δ_j quantifies the increase in mean-squared reconstruction error when the j-th aligned latent factor is removed. The values fluctuate across component indices, indicating that some aligned factors contribute more strongly to image reconstruction than others. Because raw factor indices are not identifiable before alignment, the interpretation is restricted to the aligned and stability-checked rankings described in Section IV-B. Table IV reports the numerical mean and standard deviation values corresponding to Fig. 3.

TABLE IV
NUMERICAL NORMALIZED ABLATION SCORES CORRESPONDING TO FIG. 3. VALUES ARE REPORTED AS MEAN $\pm$ STANDARD DEVIATION OVER $R = 10$ REPEATED RANDOM-MASK/INITIALIZATION TRIALS AFTER COMPONENT ALIGNMENT.

Aligned component j	5% missing	10% missing	15% missing
1	1.000 ± 0.000	1.000 ± 0.000	1.000 ± 0.000
2	0.052 ± 0.006	0.053 ± 0.007	0.047 ± 0.006
3	0.031 ± 0.005	0.035 ± 0.006	0.027 ± 0.005
4	0.011 ± 0.003	0.010 ± 0.003	0.009 ± 0.003
5	0.006 ± 0.002	0.006 ± 0.002	0.006 ± 0.002
6	0.005 ± 0.002	0.005 ± 0.002	0.004 ± 0.001
7	0.004 ± 0.001	0.004 ± 0.001	0.004 ± 0.001
8	0.003 ± 0.001	0.003 ± 0.001	0.003 ± 0.001
9	0.003 ± 0.001	0.003 ± 0.001	0.003 ± 0.001
10	0.003 ± 0.001	0.003 ± 0.001	0.002 ± 0.001
11	0.002 ± 0.001	0.002 ± 0.001	0.002 ± 0.001
12	0.002 ± 0.001	0.002 ± 0.001	0.002 ± 0.001
13	0.001 ± 0.000	0.001 ± 0.000	0.001 ± 0.000
14	0.001 ± 0.000	0.001 ± 0.000	0.001 ± 0.000
15	0.001 ± 0.000	0.001 ± 0.000	0.001 ± 0.000
16	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000

Figure 4 presents the classification effect of progressively including the top-ranked aligned latent components on the MNIST subset. Accuracy generally improves when additional high-ranked components are included, then stabilizes as lower-ranked components add less discriminative information. The curves confirm that the most important aligned components carry much of the classification-relevant structure. Table V gives the numerical mean and standard deviation values corresponding to Fig. 4.

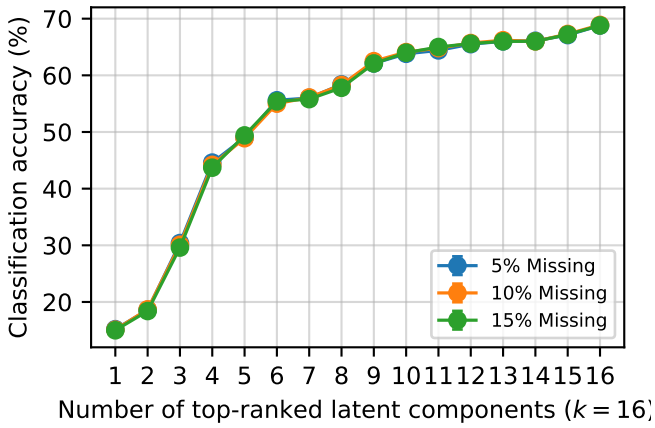

Fig. 4. Classification accuracy versus the number of top-ranked latent components included ($k = 16$) for the MNIST subset under different missing-pixel ratios.

TABLE V
NUMERICAL PROGRESSIVE-COMPONENT CLASSIFICATION ACCURACIES CORRESPONDING TO FIG. 4. VALUES ARE REPORTED AS MEAN ACCURACY (%) $\pm$ STANDARD DEVIATION OVER $R = 10$ REPEATED RANDOM-MASK/INITIALIZATION TRIALS AFTER COMPONENT ALIGNMENT AND RANKING.

Number of top-ranked components	5% missing	10% missing	15% missing
1	15.2 ± 0.2	15.1 ± 0.2	15.0 ± 0.2
2	18.6 ± 0.3	18.7 ± 0.3	18.4 ± 0.3
3	30.4 ± 0.4	30.1 ± 0.4	29.6 ± 0.5
4	44.6 ± 0.5	44.2 ± 0.5	43.7 ± 0.6
5	49.0 ± 0.6	48.9 ± 0.6	49.4 ± 0.5
6	55.6 ± 0.5	55.0 ± 0.5	55.4 ± 0.5
7	56.0 ± 0.5	56.1 ± 0.5	55.8 ± 0.4
8	58.4 ± 0.4	58.3 ± 0.4	57.8 ± 0.4
9	62.1 ± 0.4	62.5 ± 0.4	62.1 ± 0.4
10	63.8 ± 0.4	64.1 ± 0.4	64.0 ± 0.4
11	64.4 ± 0.3	64.8 ± 0.3	65.0 ± 0.3
12	65.5 ± 0.3	65.7 ± 0.3	65.6 ± 0.3
13	66.0 ± 0.3	66.2 ± 0.3	66.0 ± 0.3
14	66.1 ± 0.4	66.0 ± 0.4	66.0 ± 0.3
15	67.1 ± 0.4	67.3 ± 0.4	67.2 ± 0.4
16	68.8 ± 0.6	68.9 ± 0.6	68.8 ± 0.5

VI. CONCLUSIONS

This paper presented a mask-aware EM formulation for linear factor analysis under missing-pixel conditions. The derivation explicitly incorporates sample-dependent observation masks in both posterior inference and parameter estimation. In particular, the loading row, mean entry, and diagonal noise variance for each pixel are estimated only from samples where that pixel is observed. This resolves the inconsistency of applying complete-data updates to incomplete image vectors. The paper also clarifies that EM-FA, EM-PCA, and PPCA are classical foundations; the contribution here is the explicit image-oriented, mask-indexed recovery workflow and its accompanying factor-ranking analysis.

The experimental section specifies the missing-pixel process, initialization, stopping criterion, repeated-trial protocol, MNIST subset, and classifier. Numerical tables report mean and standard-deviation results alongside the plots. The factor-interpretation section includes component alignment and ranking-stability evaluation, preventing unsupported conclusions from a single arbitrary latent-factor index. Future work will extend the framework to nonlinear generative models and multimodal networked data streams.

ACKNOWLEDGMENTS

The authors thank Prof. Hsiao-Chun Wu of National Sun Yat-Sen University, Taiwan, for valuable technical discussions.

REFERENCES

[1] Y. Wu, X. Wang, P. Angueira, J. Montalban, N. Hur, S. Ahn, and S.-I. Park, "Local content insertion in the SFN environment for geographically segmented localcasting," in *Proceedings of IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, Jun. 2025, pp. 1–6.

[2] S. Ahn, B.-m. Lim, J. Kim, C. E. C. Ribeiro, L. Fausto, and S.-I. Park, "Geographic segmented localcasting based on LDM: Design and plans in Brazil," in *Proceedings of IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, Jun. 2025, pp. 1–4.

[3] L. N. D. S. Junior, R. S. Rabaca, G. H. M. G. de Oliveira, G. L. Montalvao, G. A. Chiesa, C. Akamine, B.-M. Lim, S.-I. Park, and Y. Wu, "TV 3.0 geographical segmentation by over-the-air physical layer," in *Proceedings of IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, Jun. 2025, pp. 1–4.

[4] G. C. Livanos and V. Anishchenko, "Development of an ultra-long-range wireless backhaul solution using ATSC 3.0," *IEEE Transactions on Broadcasting*, vol. 69, no. 2, pp. 629–634, May 2023.

[5] Q. Gao, Z. Li, J. Li, J. Xue, and Y. Xing, "Mobile computing force network architecture and quality of service (QoS) mechanism towards network-computing integration," in *Proceedings of IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, Jun. 2025, pp. 1–6.

[6] D. J. Bartholomew, M. Knott, and I. Moustaki, *Latent Variable Models and Factor Analysis: A Unified Approach*. John Wiley & Sons, 2011.

[7] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 61, no. 3, pp. 611–622, Sep. 1999.

[8] Z. Ghahramani and G. E. Hinton, "The EM algorithm for factor analysis," University of Toronto Technical Report CRG-TR-96-1, 1996.

[9] S. T. Roweis, "EM algorithms for PCA and SPCA," in *Advances in Neural Information Processing Systems*, vol. 10, 1998, pp. 626–632.

[10] E. J. Candes and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717–772, 2009.

[11] R. Mazumder, T. Hastie, and R. Tibshirani, "Spectral regularization algorithms for learning large incomplete matrices," *Journal of Machine Learning Research*, vol. 11, pp. 2287–2322, 2010.

[12] C. M. Bishop, "Latent variable models," in *Learning in Graphical Models*. Springer, 1998, pp. 371–403.

[13] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY: Springer, 2006.

[14] S. van der Walt *et al.*, "scikit-image: Image processing in Python," 2014.

[15] Y. LeCun, C. Cortes, and C. Burges, "The MNIST database of handwritten digits," http://yann.lecun.com/exdb/mnist/, 1998, accessed: 2025-10-31.