Missing Data Imputation by Reducing Mutual Information with Rectified Flows

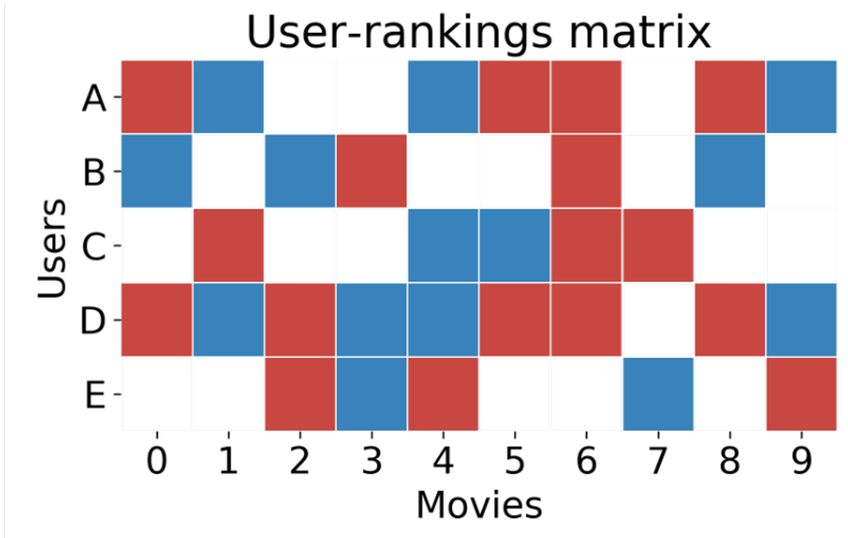
Jiahao Yu[1], *Qizhen Ying*[2], *Leyang Wang*[3], *Ziyue Jiang*[4], *Song Liu*[4]

[1] University of Cambridge [2] University of Oxford [3] UCL [4] University of Bristol

[1,2,3] This work was partially done at University of Bristol.

Neural Information Processing Systems

High-Dimensional Missing Data



The Netflix Price

64×64 CelebA

Missing Data Imputation

We observe pairs of data and missing mask:

$$D = \{(x, m)\}, m \in \{0,1\}^d, x_j = \begin{cases} x_j^*, & m_j = 1 \\ \text{NaN} & m_j = 0 \end{cases}.$$

Goal: Given D, guess$\{x_j^* \mid m_j = 0\}$.

Missingness Assumptions [1]

- Missing Completely at Random (**MCAR**): $x^* \perp\!\!\!\perp m$
- Missing at Random (**MAR**): $x_{1-m}^* \perp\!\!\!\perp m \mid x_m^*$
- Missing not at Random (**MNAR**): Otherwise

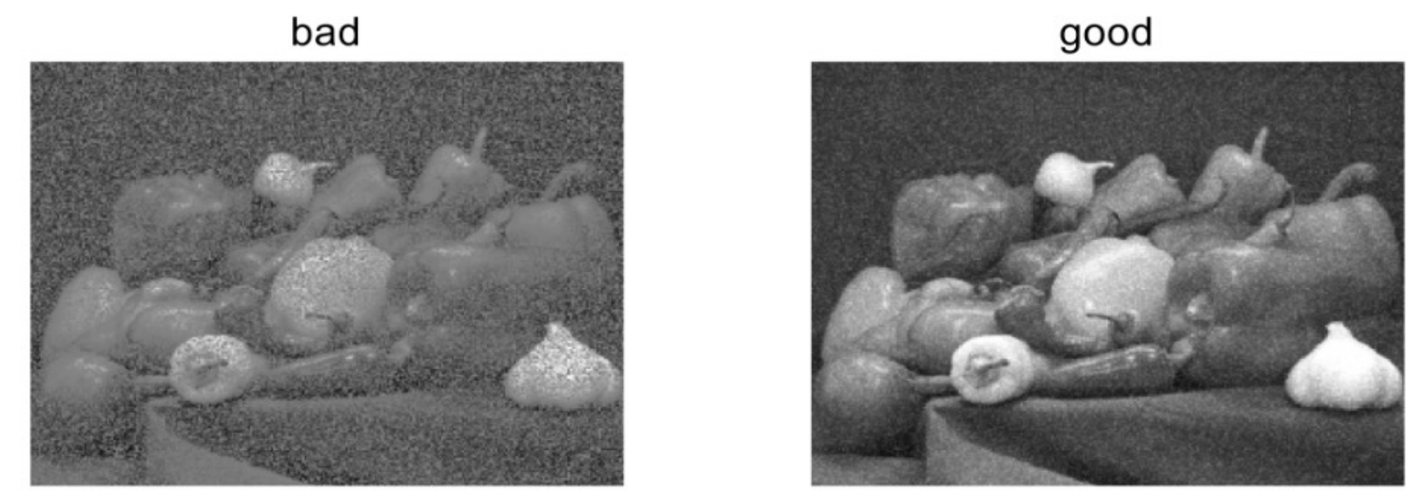
Existing Methods

- **One-shot** (e.g., Mean, Median):
  Not using feature-wise relationships.
- **Round-robin** (e.g., MICE, HyperImpute):
  Fail to scale effectively to high-dimensional data.
- **Others** using generative modelling, e.g.,
  GAIN [2]: Alternate adversarial steps
  DiffPuter: EM + Diffusion iterations
  OTImpute: Match incomplete with complete samples

⇒ *Can we design a joint imputation method?*

Criterion for Good Imputation [2]

You cannot tell which pixel is imputed and which is not.



Train a **classifier** and an **imputer** alternately, so that the classifier cannot tell which coordinate is imputed.

Our Interpretation

If **MCAR**, the perfect imputation $\hat{x} = x^*$ implies $\hat{x} \perp\!\!\!\perp m$.

- Minimizing $\text{MI}[\hat{x}, m] := \text{KL}[p_{\hat{x},m} \| p_m p_{\hat{x}}]$.

Mutual Information Reducing Iterations (MIRI)

- Set initial imputation $\hat{x}^{(0)}$, $t = 1$.
- Repeat:

$$\hat{x}^{(t)} = \arg\min_{\hat{x}} \text{KL}\left[p_{\hat{x},m} \| p_m p_{\hat{x}^{(t-1)}}\right], \ t \leftarrow t+1.$$

Prop. *MIRI is proven to reduce the mutual information!*

How to obtain $\hat{x}^{(t)}$?

Prop. *$\hat{x}^{(t)}$ optimal iff*

$$p_{\hat{x}^{(t)}}(x_{1-m} \mid x_m, m) = p_{\hat{x}^{(t-1)}}(x_{1-m} \mid x_m).$$

- Finding the optimal imputer is the same as *matching two conditional distributions*!
- Construct a rectified flow ODE [3] that *"transports" $\hat{x}^{(t-1)}$ to $\hat{x}^{(t)}$.*

Train a (Conditional) Rectified Flow [3]

Learn velocity field v by minimizing MSE loss:

$$v^* = \arg\min_v \int_\tau \mathbb{E}\left\| \tilde{x}^{(t-1)} - \hat{x}^{(t-1)} - v_\tau\left[x_{1-m}(\tau), \hat{x}_m^{(t-1)}, \hat{x}_m^{(t)}\right]\right\|^2$$

where $x(\tau) = \tau\tilde{x}^{(t)} + (1-\tau)\hat{x}^{(t-1)}$ and $\tilde{x} = \texttt{shuffle}(\hat{x})$.

Experiments

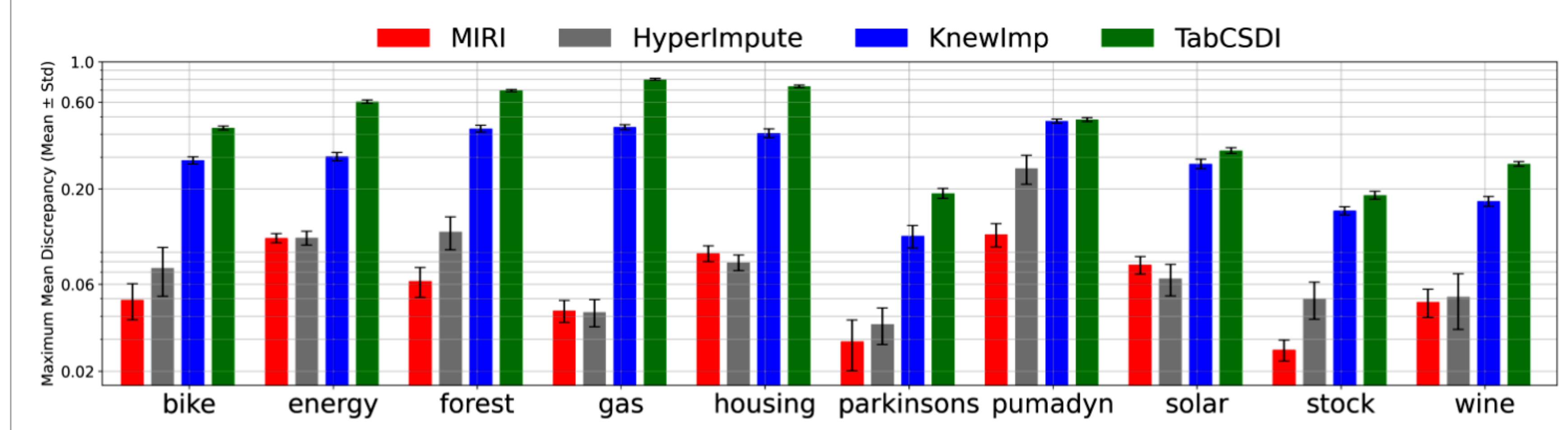
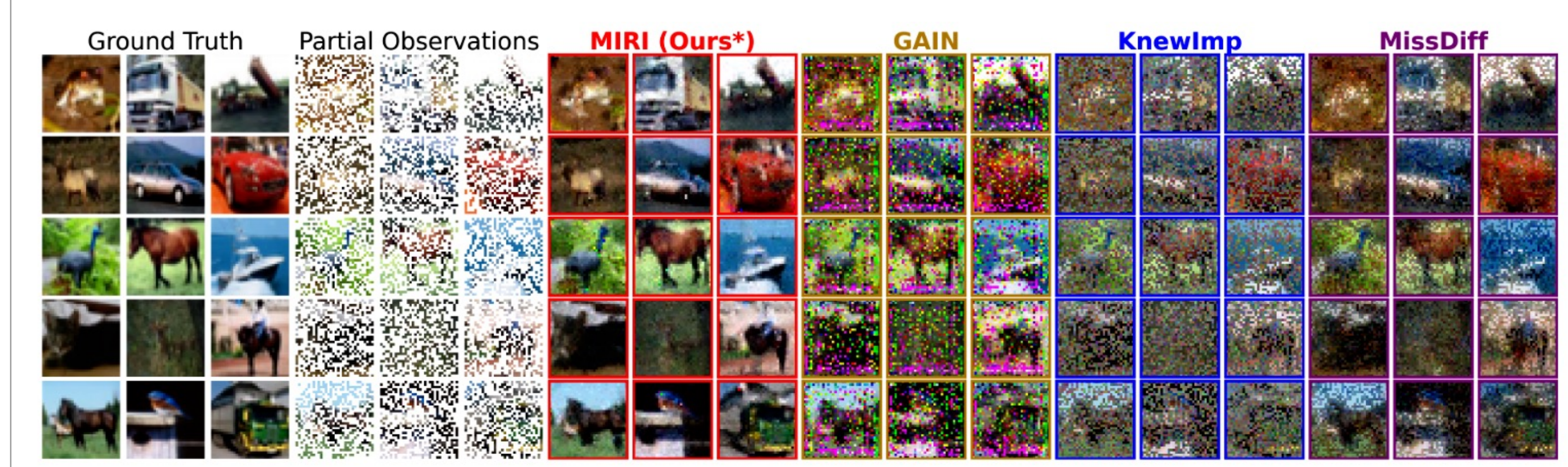
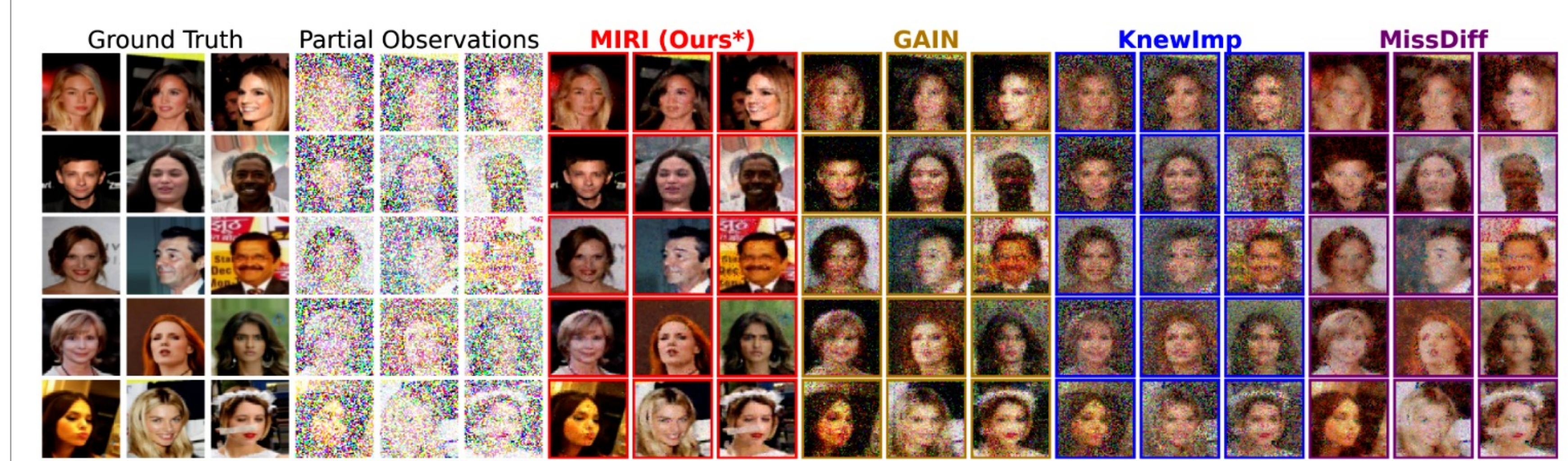


Figure 2: MMD on UCI datasets with 60% data missing. The lower the better.



(a) 15 uncurated 32×32 CIFAR-10 images and their imputations. Pixels are removed from *all RGB channels*.



(b) 15 uncurated 64×64 CelebA images and their imputations. Pixels are removed from *each RGB channel independently*.

References

[1] D. B. Rubin. Inference and missing data. Biometrika, 63(3), 581-592, 1976.
[2] J. Yoon, J. Jordon, and M. van der Schaar. GAIN: Missing data imputation using generative adversarial nets. ICML 2018.
[3] X. Liu, C. Gong, and Q. Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. ICLR 2023.