EXOTIC: External Vision Guided Incomplete Multi-view Classification

Shilin Xu[1], Dezhong Peng[1, 3], Zhenwen Ren[2], Yuan Sun[1*]
[1]Sichuan University, Chengdu, China 610044
[2]Southwest University of Science and Technology, Mianyang, China 621010
[3]Tianfu Jincheng Laboratory, Chengdu, China 610093
xushilin990@gmail.com, pengdz@scu.edu.cn, rzw@njust.edu.cn, sunyuan_work@163.com

1. Appendix / Supplemental Material

This supplementary material provides additional details about EXOIIC to facilitate a deeper understanding of the proposed framework. Section 2 introduces the details of the dataset used in this paper. Section 3 presents the experimental results on the complete dataset. Section 4 describes the training process of EXOTIC. Section 5 introduces the visualization analysis on the HW and Fashion datasets. Section 6 presents the ablation studies and analysis on all datasets with 0.5 missing rate. Section 7 provides further parameter analysis. Section 8 provides more recent sota baselines. Section 9 analyze the impact of varying the size of the external knowledge base. Section 10 validates the imputation mechanism.

2. Dataset Details

The multi-view datasets used in this paper, The detailed statistics in this paper are provided in Tab.1. Detailed descriptions of each dataset are as follows

- **Caltech101**[3] consists of 8,677 images across 101 object categories, with our experiments utilizing the first 20 categories. Six distinct visual representations are extracted using Gabor filters, Wavelet Moments, CENTRIST, HOG, GIST, and LBP descriptors.
- **Scene15**[14] comprises 4,485 images spanning 15 scene categories, processed with three feature types: pixel intensity, PHOG, and LBP.
- **LandUse21**[15] provides 2,100 satellite images across 21 land-use classes, each represented by triple-view features (intensity, PHOG, and LBP).
- **HW**[2] contains 2,000 handwritten digit samples (0-9) with uniform class distribution (200 instances per digit), each characterized by six distinct feature representations.
- **Fashion**[13] dataset comprises images of various products. We interpret the three different styles of each product as three separate views representing the same entity.
- **NUSWIDE**[16] is a real-world web image subset, which contains 30000 images and 31 classes. Each smaple comprises 5 views (Color Moments, Color Hisorgram, Color Correlation, Edge Distribution and Wavelet Texture).
- **CUB**[11] contains 200 different bird categories with 11788 images and text descriptions. Features of 10 categories are extracted by GoogLeNet and Doc2Vec in Gensim5. In this paper,we select the first 14 categories and 600 samples.

3. Results Analysis on Complete Data

We evaluate the proposed EXOTIC with ten advanced multi-view classification methods. The experimental results on complete datasets are shown in Tab.1. We use classification accuracy as the evaluation metric. According to Tab.1, we can draw the following conclusions

- In the complete case, we can observe that EXOTIC achieves the best performance on four out of seven datasets, including Scene15, LandUse21, HW, and NUSWIDE.
- Our method maintains strong performance under complete case, demonstrating its effective modeling capability for complete multi-view data.

Table 1. The detailed statistics of all multi-view datasets.

Dataset	Size	Categories	View	Dimensionality
Caltech101	2,100	24	6	48;40;254;1,984;512;928
Scene15	4,485	19	3	20;59;40
LandUse21	2,100	25	3	20;59;40
HW	2,000	14	6	216;76;64;6;240;47
Fashion	10,000	14	3	784;784;784
NUSWIDE	30,000	35	5	65;226;145;74;129
CUB	600	14	2	1,024;300

4. Training Procedure

To clearly illustrate the step-by-step training process of our proposed EXOTIC framework, we provide the correspond-

*Corresponding authors.

Table 2. Classification accuracy (%) of our EXOTIC and ten compared methods on complete datasets.

Method	Caltech101	Scene15	LandUse21	HW	Fashion	NUSWIDE	CUB
CPM-NET [17]	86.03±0.52	37.53±1.15	25.48±0.62	38.70±2.48	79.34±0.11	31.93±0.38	85.00±0.00
DCP [4]	87.78±1.13	76.32±1.19	71.38±2.29	97.50±0.45	97.46±0.55	32.40±1.32	89.00±1.78
UIMC [14]	**94.88±1.25**	76.74±0.35	54.67±2.92	98.62±0.38	98.11±0.20	40.72±0.10	**92.50±1.77**
DIMC [12]	88.64±1.54	74.38±0.92	65.81±1.65	98.45±0.19	98.32±0.32	46.75±0.44	80.42±6.25
DICNET [5]	88.72±1.54	74.18±1.24	66.10±1.93	98.40±0.20	**98.59±0.33**	46.91±0.43	86.67±2.53
LMVCAT [6]	93.06±0.10	79.55±1.08	71.24±1.07	98.75±0.25	87.35±0.40	44.11±0.32	89.17±1.67
MTD [8]	44.00±0.72	62.53±1.53	62.06±2.80	95.88±0.88	62.65±2.45	20.54±0.18	80.00±2.50
AIMNET [7]	94.10±0.31	73.21±0.93	53.90±2.26	97.75±0.25	92.38±0.47	42.94±0.68	91.67±0.00
SIP [9]	64.60±0.21	13.41±0.44	37.86±1.90	89.25±3.50	47.20±1.10	6.31±1.59	85.42±0.42
RANK [10]	56.94±3.31	42.29±1.83	55.36±0.36	90.75±0.25	54.45±0.15	16.85±2.91	88.33±0.83
EXOTIC	94.29±0.76	**79.53±1.57**	**78.67±2.43**	**98.95±0.48**	97.57±0.07	**50.96±0.45**	89.67±2.01

Algorithm 1 Training algorithm of the proposed EXOTIC

Require: Multi-view data $\left\{X^v \in \mathbb{R}^{d_v}\right\}_{v=1}^{V}$ with labels $\{y_i \in \{0,1\}^C\}_{i=1}^{N}$; External knowledges M; Hyper-parameters α, β, γ; Train epochs T; External knowledge match number k;

Ensure: Parameters of the trained model and calibrated label.

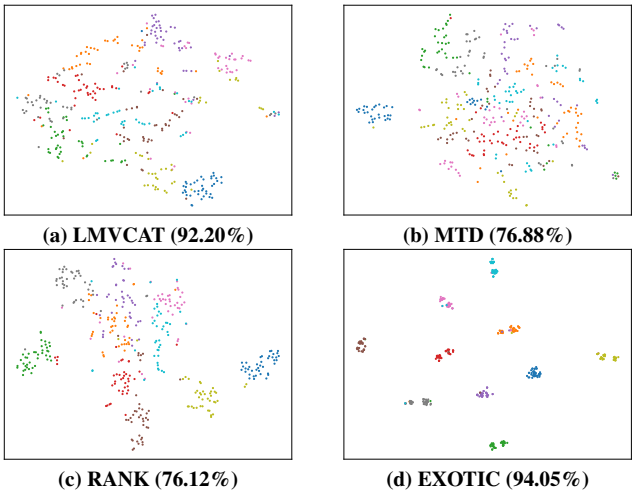
1: **for** each $i \in [1, T]$ **do**
2: Extract the first channel embedding feature $\left\{\mathbf{C}^{(v)}\right\}_{v=1}^{m}$, the second channel embedding feature $\left\{\mathbf{P}^{(v)}\right\}_{v=1}^{m}$ and external knowledge representation U.
3: Compute the fusion representation Z of the first channel according to Eq.1.
4: Match k external knowledge representations for each sample according to Eq.2.
5: Compute the weighted external knowledge according to Eq.3.
6: Compute the classification result of weighted external knowledge.
7: Impute the missing view by Eq.7 and get imputed fusion representation.
8: Compute the classification result of the imputed fusion representation.
9: Compute knowledge-specific contrastive loss $\mathcal{L}_{vc}$, category-wise contrastive loss $\mathcal{L}_{kc}$, and cross-entropy loss $\mathcal{L}_c$ according to Eq.4, Eq.5, and Eq.7, respectively.
10: Compute total loss $\mathcal{L}_{all}$ by Eq.8
11: Update network parameters.
12: **end for**

ing pseudocode in Algorithm 1. This algorithm details the key operations and data flow.

5. Visualization Analysis

To further validate the effectiveness of EXOTIC in learning discriminative representations under extreme data incompleteness, we perform t-SNE visualization on the HW dataset with a missing rate of 0.9, comparing it with several state-of-the-art methods, including LMVCAT, MTD, and RANK, as shown in Fig. 1. The visualized distributions in Fig. 1 (a)–(c) reveal severe class overlaps and scattered samples, suggesting that these baselines fail to capture robust feature structures when most views are absent. In contrast, our EXOTIC method (Fig. 1 (d)) produces compact, well-separated clusters, where samples from the same class are tightly grouped and distinct from other classes. This clear class separability demonstrates that external vision knowledge effectively guides the model toward semantically consistent and class-aligned feature representations, thereby enhancing the robustness of model and discriminative ability even under high missing rates.

For a more comprehensive study, we employ t-SNE to visually illustrate the learned representations on the Fashion dataset under missing rates of 0.5 and 0.9, respectively. As shown in Fig.2, it can be observed that EXOTIC learns more compact and discriminative representations under both 0.5 and 0.9 missing rates. This further demonstrates the superiority of our method.



(a) LMVCAT (92.20%) **(b) MTD (76.88%)**

(c) RANK (76.12%) **(d) EXOTIC (94.05%)**

Figure 1. t-SNE visualization on the HW dataset with 0.9 missing rate, where classification accuracy is reported in brackets.

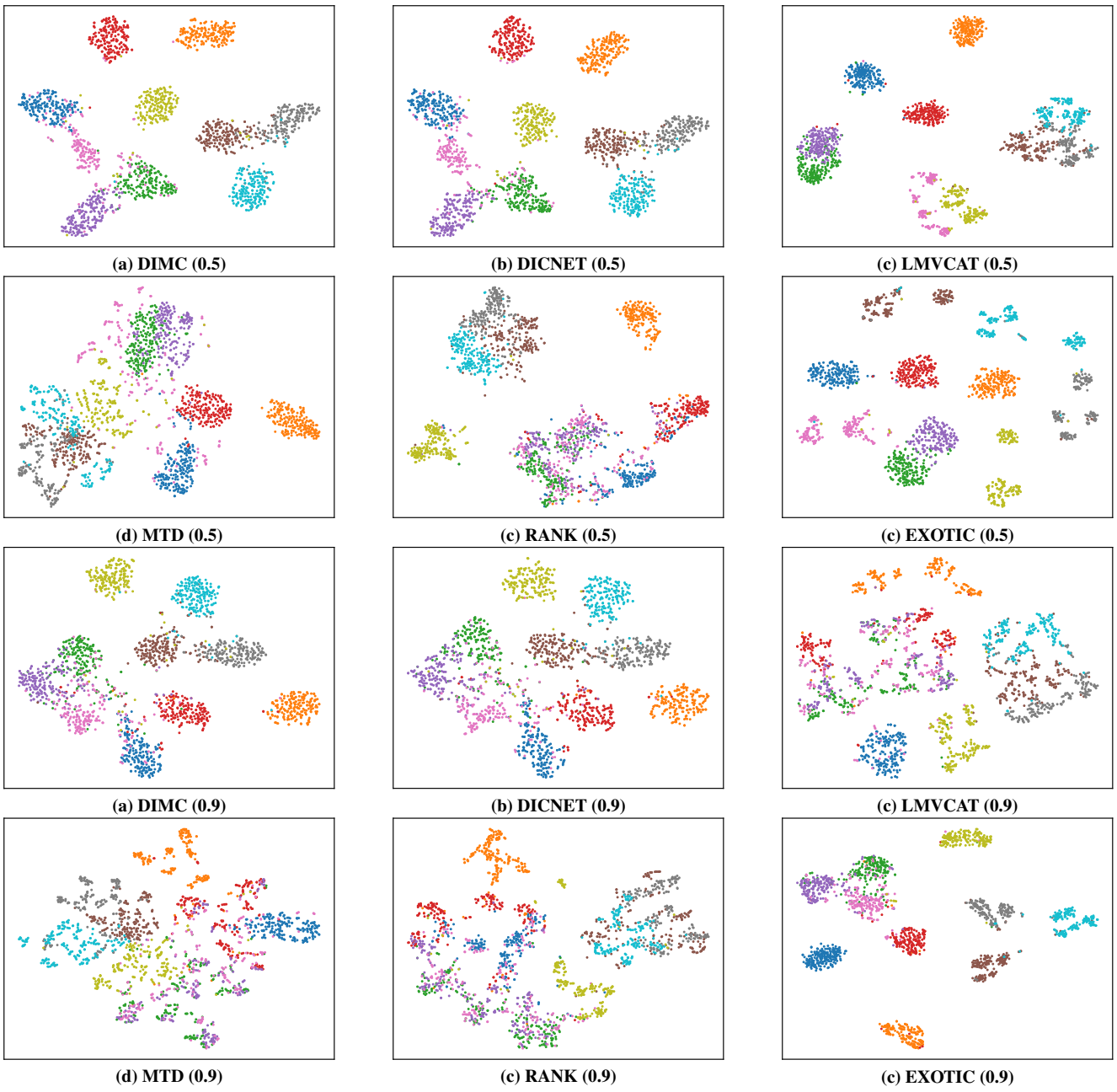


Figure 2. t-SNE visualization under 0.5 and 0.9 missing rate on the Fashion dataset.

6. Ablation Studies

To further evaluate the contribution of each component, ablation experiments are conducted under four variants with 0.5 missing rate. From the Tab. 3, we can conclude that: (1) When we removed $\mathcal{L}_{kc}$, the performance significantly dropped, indicating its importance in balancing internal and external information and preventing semantic conflicts. (2) When the loss term $\mathcal{L}_{vc}$ is removed, noticeable performance drops occur across almost all datasets, especially on the NUSWIDE dataset, showing its role in enhancing the task relevance of external knowledge. (3) When both losses are removed, accuracy declines further on most datasets, confirming their complementary effect to enhance classification accuracy. 4) When external knowledge completion is removed, performance on all datasets shows a significant decline, which demonstrates the powerful role of knowledge completion in handling missing issues.

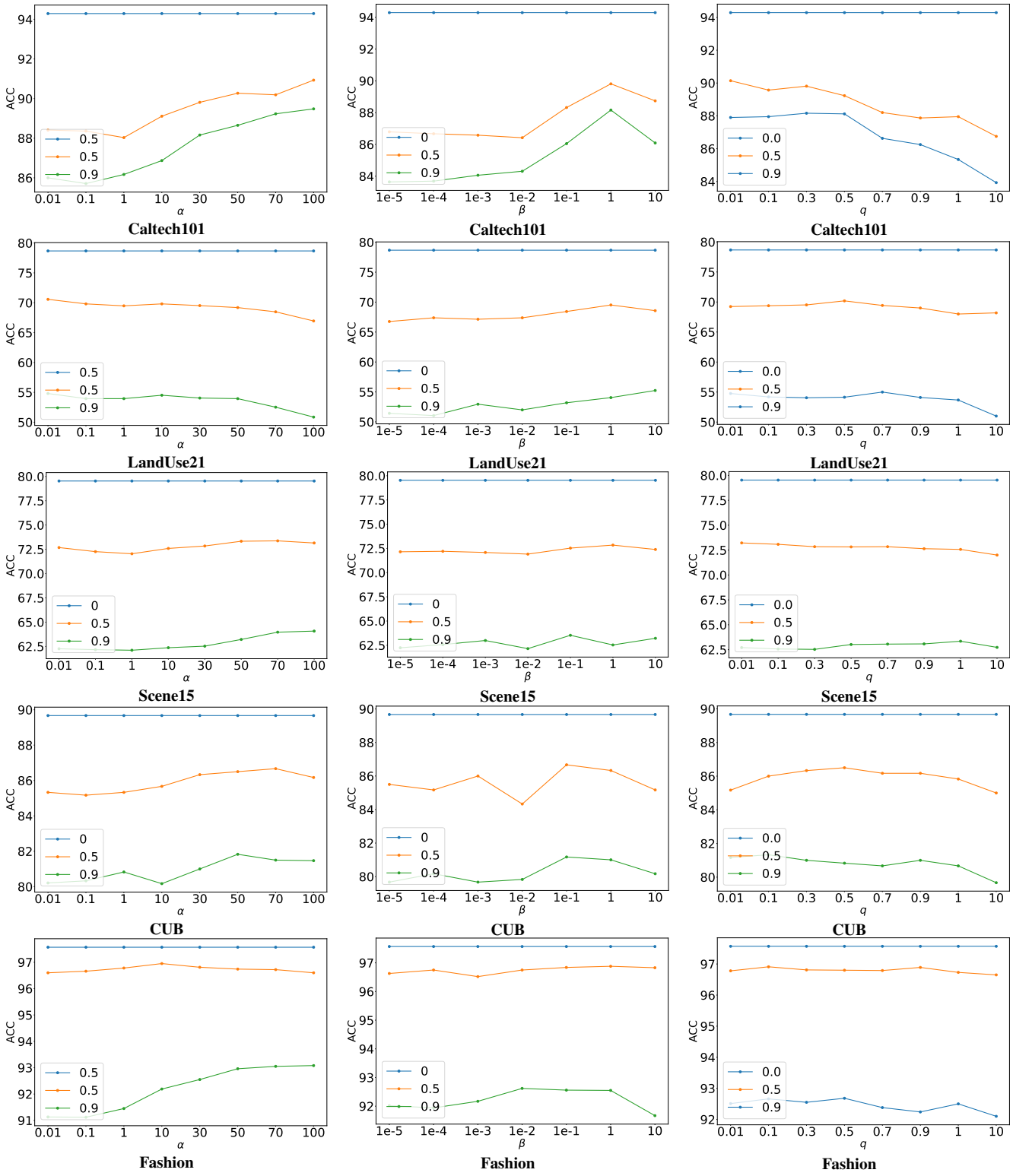


Figure 3. Parameter analysis for α, β, and q on five datasets with different missing rates.

7. Parameter Analysis

To further evaluate the impact of each hyperparameter α, β, and q, we conduct a sensitivity analysis on other four

Table 3. Ablation study on different datasets with 0.5 missing rate.

Dataset	EXOTIC-1	EXOTIC-2	EXOTIC-3	EXOTIC-4	EXOTIC
Caltech101	85.92	87.87	86.63	87.49	**89.81**
Scene15	72.08	72.51	71.75	71.64	**72.84**
LandUse21	68.76	**69.95**	67.76	66.62	69.52
HW	95.35	**96.95**	96.40	96.35	96.60
Fashion	96.29	96.64	96.73	94.62	**96.81**
NUSWIDE	**52.40**	50.69	51.03	45.67	52.32
CUB	84.33	85.17	84.83	82.50	**86.33**

datasets under different missing rates, e.g., 0, 0.5, and 0.9. From the Fig. 3, we can observe that: (1) The model exhibits different sensitivities to different parameters. Specifically, it is slightly more sensitive to alpha and beta than to q. (2) Our method maintains stable performance over a relatively wide range of parameter settings. (3) Overall, compared with the low missing rate, parameter variations have a greater impact on performance under the high missing rate. To verify the role of q in controlling the sharpness of the penalty, we conduct a direct comparison with the InfoNCE loss on different datasets with 90% missing data. As shown in Tab. 4, our method performance consistently outperforms InfoNCE, demonstrating that q effectively controls the sharpness of the penalty and leads to improved performance over standard cross-entropy loss such as InfoNCE.

Table 4. Our Loss vs. Standard InfoNCE (%) on six datasets with 0.9 missing rate..

Loss	HW	Caltech101	Fashion	CUB	Scene15	LandUse21
InfoNCE	94.15	87.49	92.15	80.00	**63.41**	54.29
Ours	**94.20**	**88.16**	**92.55**	**81.00**	62.53	**55.24**

8. More Recent Sota Baselines

We include additional comparison method, and the corresponding experimental results are reported in Tab. 5. Due to the page limitation, these experiments are conducted on the HW and Caltech101 datasets under different missing rates. The results show that our method consistently outperforms recent baselines, providing strong evidence of its effectiveness.

Table 5. Classification accuracy (%) on the HW and Scene15 datasets with 0.5 missing rate

Data	Method	0.1	0.2	0.3	0.4	0.5
HW	RIML[1]	96.50	95.00	94.30	93.80	92.90
	Ours	**99.00**	**98.90**	**98.30**	**98.10**	**96.40**
Scene15	RIML[1]	68.19	65.65	64.25	62.41	59.54
	Ours	**80.84**	**80.27**	**76.25**	**72.48**	**72.84**

9. Impact of Knowledge Size

We conduct additional experiments on the HW and Scene15 datasets under a 90% missing rate, where the size of the external knowledge base is gradually increased. As shown in Tab. 6, increasing the size of the external knowledge base consistently improves performance at first, but the gains gradually saturate once a moderate scale is reached. This suggests that EXOTIC does not rely on massive external datasets to be effective. In practice, a reasonably sized knowledge base is sufficient to capture useful semantic information, which can provide a clear trade-off for real-world deployment.

Table 6. The impact of knowledge size (%) on the HW and Scene15 datasets with 0.9 missing rate.

Size	100	1000	5000	10,000	50,000
HW	84.0	94.65	94.45	94.55	94.20
Caltech101	75.40	86.87	87.78	88.20	88.16

10. Validation of the Imputation Mechanism

To validate our imputation mechanism, we compare it with a standard KNN imputation mechanism on the HW dataset under 90% missing rate. As shown in Tab. 7, our method consistently outperforms KNN, demonstrating the clear advantage of our mechanism over simple heuristics.

Table 7. Prototype-based vs. KNN Imputation (%) on six datasets with 0.9 missing rate.

Mechanism	HW	Caltech101	Fashion	CUB	Scene15	LandUse21
KNN	93.05	85.71	90.21	80.52	62.24	52.76
Ours	**94.20**	**88.16**	**92.55**	**81.00**	**62.53**	**55.24**

References

[1] Haishun Chen, Cai Xu, Ziyu Guan, Wei Zhao, and Jinlong Liu. Biased incomplete multi-view learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 15767–15775, 2025. 5

[2] Yuhang Lan, Shilin Xu, Chao Su, Run Ye, Dezhong Peng, and Yuan Sun. Multi-view hashing classification. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pages 2122–2130, 2025. 1

[3] Yeqing Li, Feiping Nie, Heng Huang, and Junzhou Huang. Large-scale multi-view spectral clustering via bipartite graph. In *Proceedings of the AAAI conference on artificial intelligence*, 2015. 1

[4] Yijie Lin, Yuanbiao Gou, Xiaotian Liu, Jinfeng Bai, Jiancheng Lv, and Xi Peng. Dual contrastive prediction for incomplete multi-view representation learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4): 4447–4461, 2022. 2

[5] Chengliang Liu, Jie Wen, Xiaoling Luo, Chao Huang, Zhihao Wu, and Yong Xu. Dicnet: Deep instance-level contrastive network for double incomplete multi-view multi-label classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 8807–8815, 2023. 2

[6] Chengliang Liu, Jie Wen, Xiaoling Luo, and Yong Xu. Incomplete multi-view multi-label learning via label-guided masked view-and category-aware transformers. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 8816–8824, 2023. 2

[7] Chengliang Liu, Jinlong Jia, Jie Wen, Yabo Liu, Xiaoling Luo, Chao Huang, and Yong Xu. Attention-induced embedding imputation for incomplete multi-view partial multi-label classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 13864–13872, 2024. 2

[8] Chengliang Liu, Jie Wen, Yabo Liu, Chao Huang, Zhihao Wu, Xiaoling Luo, and Yong Xu. Masked two-channel decoupling framework for incomplete multi-view weak multi-label learning. *Advances in Neural Information Processing Systems*, 36, 2024. 2

[9] Chengliang Liu, Gehui Xu, Jie Wen, Yabo Liu, Chao Huang, and Yong Xu. Partial multi-view multi-label classification via semantic invariance learning and prototype modeling. In *Forty-first International Conference on Machine Learning*, 2024. 2

[10] Chengliang Liu, Jie Wen, Yong Xu, Bob Zhang, Liqiang Nie, and Min Zhang. Reliable representation learning for incomplete multi-view missing multi-label classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(6):4940 – 4956, 2025. 2

[11] Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The caltech-ucsd birds-200-2011 dataset. 2011. 1

[12] Jie Wen, Chengliang Liu, Shijie Deng, Yicheng Liu, Lunke Fei, Ke Yan, and Yong Xu. Deep double incomplete multi-view multi-label learning with incomplete labels and missing views. *IEEE Transactions on Neural Networks and Learning Systems*, 2023. 2

[13] Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017. 1

[14] Mengyao Xie, Zongbo Han, Changqing Zhang, Yichen Bai, and Qinghua Hu. Exploring and exploiting uncertainty for incomplete multi-view classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19873–19882, 2023. 1, 2

[15] Yi Yang and Shawn Newsam. Bag-of-visual-words and spatial extensions for land-use classification. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 270–279, 2010. 1

[16] Shengju Yu, Siwei Wang, Pei Zhang, Miao Wang, Ziming Wang, Zhe Liu, Liming Fang, En Zhu, and Xinwang Liu. Dvsai: Diverse view-shared anchors based incomplete multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 16568–16577, 2024. 1

[17] Changqing Zhang, Zongbo Han, Huazhu Fu, Joey Tianyi Zhou, Qinghua Hu, et al. Cpm-nets: Cross partial multi-view networks. *Advances in Neural Information Processing Systems*, 32, 2019. 2