KMD: Koopman Multi-modality Decomposition for Generalized Brain Tumor Segmentation under Incomplete Modalities

Supplementary Material

7. Additional Experiments

7.1. Segmentation Performance

To highlight the effectiveness of KMD in balanced missing rate scenarios compared to Passion [12], the segmentation results for the WT, TC, and ET classes are presented in Table 5, Table 6 and Table 7 respectively. The results indicate that incorporating Passion into the balanced missing rate dataset adversely impacts the performance of the baseline models, demonstrating that Passion is effective only under imbalanced missing rate conditions. In contrast, KMD consistently performs well across both balanced and imbalanced missing rate settings.

7.2. Computational Comparison

We have evaluated the KMD module and Passion [12] by the time taken per training iteration, and GPU memory usage. Both methods are trained with the baseline model mmFormer [20] using a batch size of 1. To ensure a fair comparison, we measure the average time consumption over 30 iterations on a single Nvidia 3090Ti GPU. Passion [12] constantly consumes 9942 MiB of GPU memory with training iterations 1.502s while our KMD constantly consumes 6424 MiB of GPU memory with training iterations 0.863s. In Section 5.3, we have highlighted that KMD outperforms Passion in terms of effectiveness. From the data above, it is evident that KMD also requires less memory and achieves faster computation speeds compared to Passion. This demonstrates that KMD is more efficient under imbalanced settings.

7.3. Ablation Studies

Hyper-parameter Hyper-parameter sensitivities are presented in Table 9. Empirically, $[\alpha, \beta, \gamma]$ should be set as [$1/10$, $1/100$, 1] to make respective losses in the same magnitude. Predictions may decline by setting them as [$1/10$, $1/10$, 1], [$1/10$, $1/10$, $1/10$], and [1, $1/10$, 10]. To further study the impact of hyper-parameters, the values of [α, β, γ] are fine-tuned within the given range, with adjustments of $1/3$ and $1/6$ of the range, respectively. From the results, we found predictions do not decline drastically.

Similarity Measurement The cosine similarity (D_{cos}) is a suitable measurement in the proposed KMD because it measures similarities between two vectors of an inner product space, rather than an absolute difference between two samples. In line with this, common features will be parallel to each other, while those specific will be orthogonal. Differently, absolute measurement such as MSE emphasizes the actual difference between corresponding elements of the vectors. It is sensitive to vector magnitude but has nothing to do with inner product space measurement. Constructing a modality relationship with MSE would bring a negative impact since it heavily penalizes small differences in magnitude even though respective inner products are insignificant. Results in Table 8 verify that MSE-based measures that evaluate similarity with Euclidean distance do not bring prediction improvement, while cosine similarity does.

Table 5. Comparison of WT Dice scores when different modalities are missing with balanced missing rates on BraTS2018. • represents present modalities. Dice scores of baselines, baselines with Passion [12] and baselines with KMD (Δ) are presented. Average means the average Dice score of all the scenarios.

Modalities				WT								
Flair	T1	T1ce	T2	[5]	[12]	Δ	[20]	[12]	Δ	[11]	[12]	Δ
			•	80.52	76.04	76.94	84.09	81.79	84.07	79.47	77.68	85.09
		•		67.06	63.14	78.33	72.85	67.95	81.64	83.06	64.69	87.37
	•			68.42	61.88	79.49	73.37	64.45	84.14	85.97	66.07	85.92
•				82.96	78.09	80.69	85.60	83.58	83.73	84.37	80.52	86.69
		•	•	82.57	78.61	80.63	85.97	82.81	85.54	83.76	79.66	86.60
	•	•		71.97	68.26	79.46	76.93	72.84	84.96	85.98	70.99	86.42
•	•			85.82	81.62	81.81	87.09	85.96	84.25	86.04	84.07	86.52
	•		•	83.25	77.81	81.42	86.09	83.30	84.39	85.73	79.91	85.61
•			•	86.00	82.35	80.59	87.55	86.51	84.08	84.58	84.05	87.27
•		•		84.94	81.12	81.14	87.94	85.99	84.83	84.32	83.49	87.13
•	•	•		86.06	82.64	81.78	88.36	86.65	85.01	86.10	84.31	86.79
•		•	•	86.53	83.00	81.86	88.16	87.11	84.39	85.95	85.06	87.10
	•	•	•	86.34	83.36	81.22	88.74	87.38	84.95	84.30	84.59	87.61
•	•		•	83.61	78.80	81.73	85.96	83.31	85.22	85.94	80.36	86.88
Average				81.15	76.91	80.51	84.19	81.40	84.37	84.68	78.96	86.64

Table 6. Comparison of TC Dice scores when different modalities are missing with balanced missing rates on BraTS2018. • represents present modalities. Dice scores of baselines, baselines with Passion [12] and baselines with KMD (Δ) are presented. Average means the average Dice score of all the scenarios.

Modalities				TC								
Flair	T1	T1ce	T2	[5]	[12]	Δ	[20]	[12]	Δ	[11]	[12]	Δ
			•	64.03	63.01	68.95	67.80	63.26	76.90	72.42	62.73	78.02
		•		74.53	69.48	74.07	77.32	72.92	76.66	75.70	70.78	79.44
	•			58.63	57.76	74.50	64.56	56.46	78.38	79.11	59.91	77.95
•				61.95	60.25	74.35	64.08	60.60	75.98	76.72	61.56	78.67
		•	•	79.20	76.61	75.19	81.51	77.97	78.87	76.39	30.33	79.60
	•	•		77.45	74.37	74.68	79.43	76.25	79.07	79.37	74.77	79.25
•	•			69.25	66.79	75.97	69.14	66.28	78.00	79.64	68.23	79.04
	•		•	67.48	66.01	75.59	70.63	66.41	78.23	78.91	66.85	79.07
•			•	67.98	66.39	74.53	68.60	66.86	77.08	78.10	66.05	79.58
•		•		78.85	77.23	75.75	80.75	77.94	78.27	77.72	77.73	79.43
•	•		•	80.15	78.15	75.88	81.75	79.18	79.31	80.14	78.60	79.28
•		•	•	70.75	68.64	75.88	70.92	68.88	78.09	79.71	68.73	79.66
	•	•	•	79.40	78.21	75.70	81.74	78.29	79.39	77.26	77.88	79.92
•	•		•	80.15	77.11	77.02	81.55	78.43	79.39	79.69	76.81	79.74
Average				72.13	70.00	74.86	74.27	70.70	78.12	77.92	67.21	79.19

8. Find linear operator in Multi-modal Data

In this section, following the theorems proved in [13], we can expand the theory to multi-modality data settings.

Table 7. Comparison of ET Dice scores when different modalities are missing with balanced missing rates on BraTS2018. • represents present modalities. Dice scores of baselines, baselines with Passion [12] and baselines with KMD (Δ) are presented. Average means the average Dice score of all the scenarios.

Modalities				ET								
Flair	T1	T1ce	T2	[5]	[12]	Δ	[20]	[12]	Δ	[11]	[12]	Δ
			•	38.69	37.74	60.41	40.08	39.39	72.14	61.06	35.53	69.01
		•		69.22	66.28	68.40	72.19	67.71	70.69	63.78	64.64	70.08
	•			30.89	25.89	68.50	38.89	27.98	71.74	64.69	31.84	70.59
•				33.56	28.96	66.02	37.23	32.79	70.91	60.56	30.33	70.25
		•	•	71.40	69.21	68.26	73.11	71.29	74.97	64.53	68.36	70.53
	•	•		70.90	68.03	69.18	73.06	71.00	71.27	67.25	67.04	70.76
•	•			38.53	35.09	69.38	40.64	36.67	71.27	66.93	36.91	70.85
	•		•	41.91	39.22	68.11	42.27	40.71	72.43	68.76	39.69	70.61
•			•	40.90	39.01	66.08	43.65	41.64	72.03	63.15	39.24	71.06
•		•		69.51	67.07	68.95	75.56	70.35	72.98	64.13	67.61	70.51
•	•	•		71.61	69.46	69.35	43.34	41.82	73.00	67.45	67.73	70.89
•		•	•	43.37	40.98	68.86	81.74	71.04	72.28	72.45	39.89	71.36
	•	•	•	71.17	68.74	68.88	73.36	72.83	73.14	64.81	68.13	71.26
•	•		•	74.20	70.21	69.58	75.31	73.55	73.22	67.29	68.75	71.20
Average				54.70	51.85	67.85	57.89	54.20	72.29	65.49	51.84	70.64

Assumption 4. *Consider a multi-modal dataset containing N samples, where each sample consists of M modalities. Each modality m can be represented by a function $g_m : \mathcal{M} \to \mathbb{R}$ (or $\mathbb{C}$), where $m = 1, 2, \ldots, M$. We assume that, as the number of observations m increases, the time-average and space-average of the data for each modality converge for almost all $x \in \mathcal{M}$. Specifically, for each modality m, we have:*

$$\lim_{m\to\infty} \frac{1}{m} \sum_{j=0}^{m-1} g_m(\boldsymbol{x}_j) = \int_{\mathcal{M}} g_m(\boldsymbol{x})\, d\mu(\boldsymbol{x}), \tag{3}$$

Assumption 4 holds under the condition that different modalities from the same dataset, obtained from the same patient, vary primarily in their signal representation. In this context, each spatial slice within a volumetric dataset can be interpreted as a component of spatial information, while the sequence of these slices can be regarded as temporal information.

Theorem 5. *(Linear Operator G in Koopman Invariant Subspaces [13]) Let $\boldsymbol{Y}_0 = [\boldsymbol{g}(\boldsymbol{x}_0) \cdots \boldsymbol{g}(\boldsymbol{x}_{m-1})]$ and $\boldsymbol{Y}_1 = [\boldsymbol{g}(\boldsymbol{f}(\boldsymbol{x}_0)) \cdots \boldsymbol{g}(\boldsymbol{f}(\boldsymbol{x}_{m-1}))]$, and suppose that Assumption 4 holds. If all modes are sufficiently excited in the data (i.e., $\operatorname{rank}(\boldsymbol{Y}_0) = n$), then the matrix $\boldsymbol{A} = \boldsymbol{Y}_1\boldsymbol{Y}_0^{\dagger}$ almost surely converges to the matrix representation of the linear operator G as $m \to \infty$.*

This theorem indicates that by sampling a sufficient number of system states and under specific conditions, such as sufficient excitation of the modes, we can approximate an infinite-dimensional Koopman operator $\mathcal{G}$ using a finite-dimensional matrix $\boldsymbol{A}$. Based on the Assumption 4 and Theorem 5, we have Theorem 6 for multi-modal images in this study.

Theorem 6. *For each modality $m \in [M]$, define $\boldsymbol{Y}_0^{(m)} = [\boldsymbol{g}_m(\boldsymbol{x}_0) \cdots \boldsymbol{g}_m(\boldsymbol{x}_{m-1})]$ and $\boldsymbol{Y}_1^{(m)} = [\boldsymbol{g}_m(\boldsymbol{f}(\boldsymbol{x}_0)) \cdots \boldsymbol{g}_m(\boldsymbol{f}(\boldsymbol{x}_{m-1}))]$, and suppose that Assumption 4 holds for each modality. If all modalities are sufficiently excited in the data (i.e., $\operatorname{rank}\left(\boldsymbol{Y}_0^{(m)}\right) = n$ for each $m \in [M]$), then for each modality m, the matrix $\boldsymbol{A}^{(m)} = \boldsymbol{Y}_1^{(m)}\left(\boldsymbol{Y}_0^{(m)}\right)^{\dagger}$ almost surely converges to the matrix form of the linear operator G_m as $m \to \infty$.*

Proof. For a single modality m, the given data $\boldsymbol{Y}_0^{(m)}$ and $\boldsymbol{Y}_1^{(m)}$ correspond to the setup in Theorem 5. The function $\boldsymbol{g}_m(\cdot)$ in Theorem 6 plays the role of $\boldsymbol{g}(\cdot)$ in Theorem 5, and the data matrices $\boldsymbol{Y}_0^{(m)}$ and $\boldsymbol{Y}_1^{(m)}$ in Theorem 6 are analogous to $\boldsymbol{Y}_0$ and $\boldsymbol{Y}_1$ in Theorem 5. Since Assumption 4 holds for each modality m, and the rank condition $\operatorname{rank}\left(\boldsymbol{Y}_0^{(m)}\right) = n$ is satisfied, Theorem 5 guarantees that the matrix $\boldsymbol{A}^{(m)} = \boldsymbol{Y}_1^{(m)}\left(\boldsymbol{Y}_0^{(m)}\right)^{\dagger}$ almost surely converges to the matrix form of the linear operator G_m as $m \to \infty$. Since Theorem 5 applies independently to each modality m, the same convergence result holds for each $\boldsymbol{A}^{(m)}$: almost surely converges to the matrix form of the linear operator G_m as $m \to \infty$. □

Theorem 6 is proven by directly applying Theorem 5 to each modality m. The convergence property for the matrix $\boldsymbol{A}^{(m)}$ in Theorem 6 follows from the established convergence in Theorem 5.

Table 8. Segmentation results of models with (•) different losses.

Loss	RFNet				mmFormer				M2FTrans			
MSE-based Loss	79.58	69.97	62.10	70.55	83.01	71.46	70.45	74.97	81.23	76.26	68.32	75.27
Cosine Loss	80.51	74.86	67.85	74.41	84.37	78.12	72.29	78.26	86.64	79.19	70.64	78.82

Table 9. Average segmentation results of models with different hyper-parameters under different missing rates (MR).

MR	Hyper			RFNet	mmFormer	M2FTrans	Hyper			RFNet	mmFormer	M2FTrans
	α	β	γ				α	β	γ			
(s,m,l)				71.29	70.76	73.10				70.95	71.96	74.11
(m,s,l)	1	1/10	1/10	70.61	73.18	74.96	1/10	1/10	1	71.00	69.90	73.70
(l,m,s)				70.48	71.89	72.59				71.39	71.27	72.71
(s,m,l)				73.89	72.76	74.89				74.69	73.37	75.11
(m,s,l)	1/10	1/10	1/10	71.28	74.44	74.45	1/10	1/100	1	73.44	74.10	74.48
(l,m,s)				70.11	68.17	72.11				72.85	74.50	73.17
(s,m,l)				71.74	74.68	74.82				71.78	76.59	74.57
(m,s,l)	1/10	1/300	1	73.76	75.23	74.35	1/10	1/600	1	71.95	76.13	74.75
(l,m,s)				72.33	72.78	72.98				73.13	74.36	73.57
(s,m,l)				72.21	75.49	74.27				72.88	74.27	73.04
(m,s,l)	1/10	1/100	1/3	73.12	74.99	74.89	1/30	1/600	1	71.28	75.44	75.58
(l,m,s)				72.62	73.24	72.94				72.80	74.93	73.56