



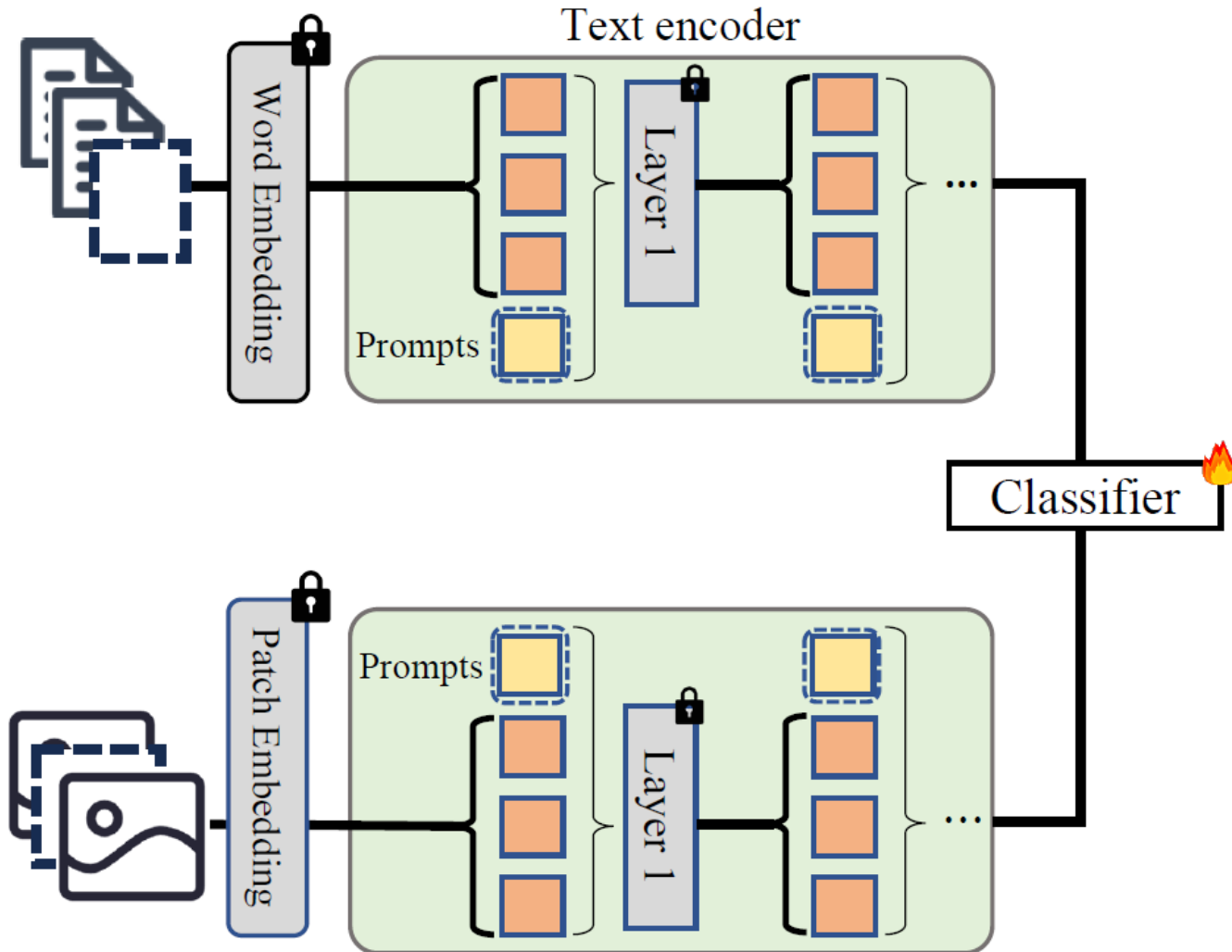
Deep Correlated Prompting for Visual Recognition with Missing Modalities

Lianyu Hu, Tongkai Shi, Wei Feng, Fanhua Shang, Liang Wan

College of Intelligence and Computing, Tianjin University

Code: https://github.com/hulianyuyy/Deep_Correlated_Prompting

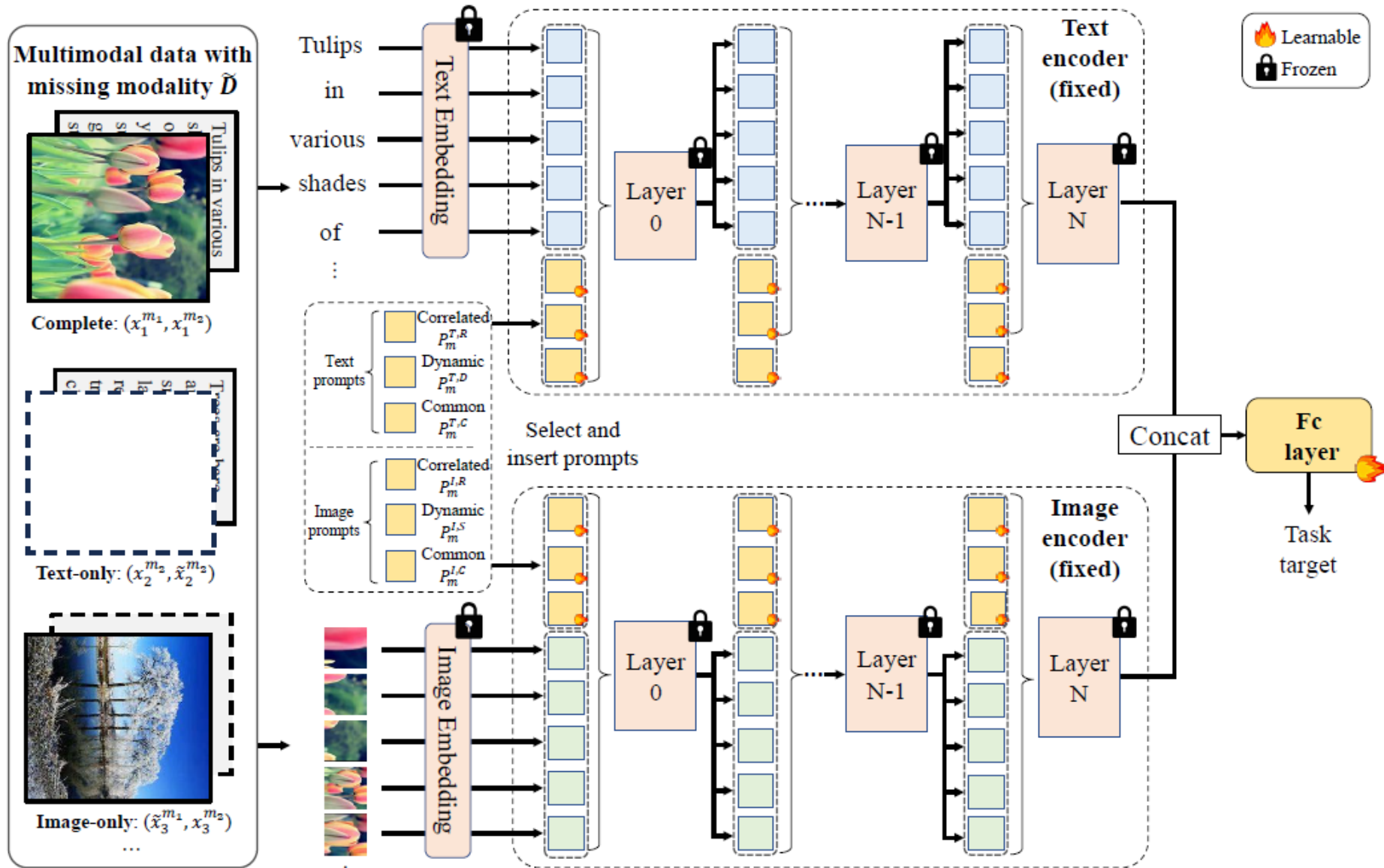
Recap



A previous method (MMP[1]) proposes to use prompt learning to handle missing-modality cases. Only the parameters of the prompts are updated, and the backbone is fixed.

However, it fails to consider the relationships between different prompts, and the correlations between prompts and inputs are ignored.

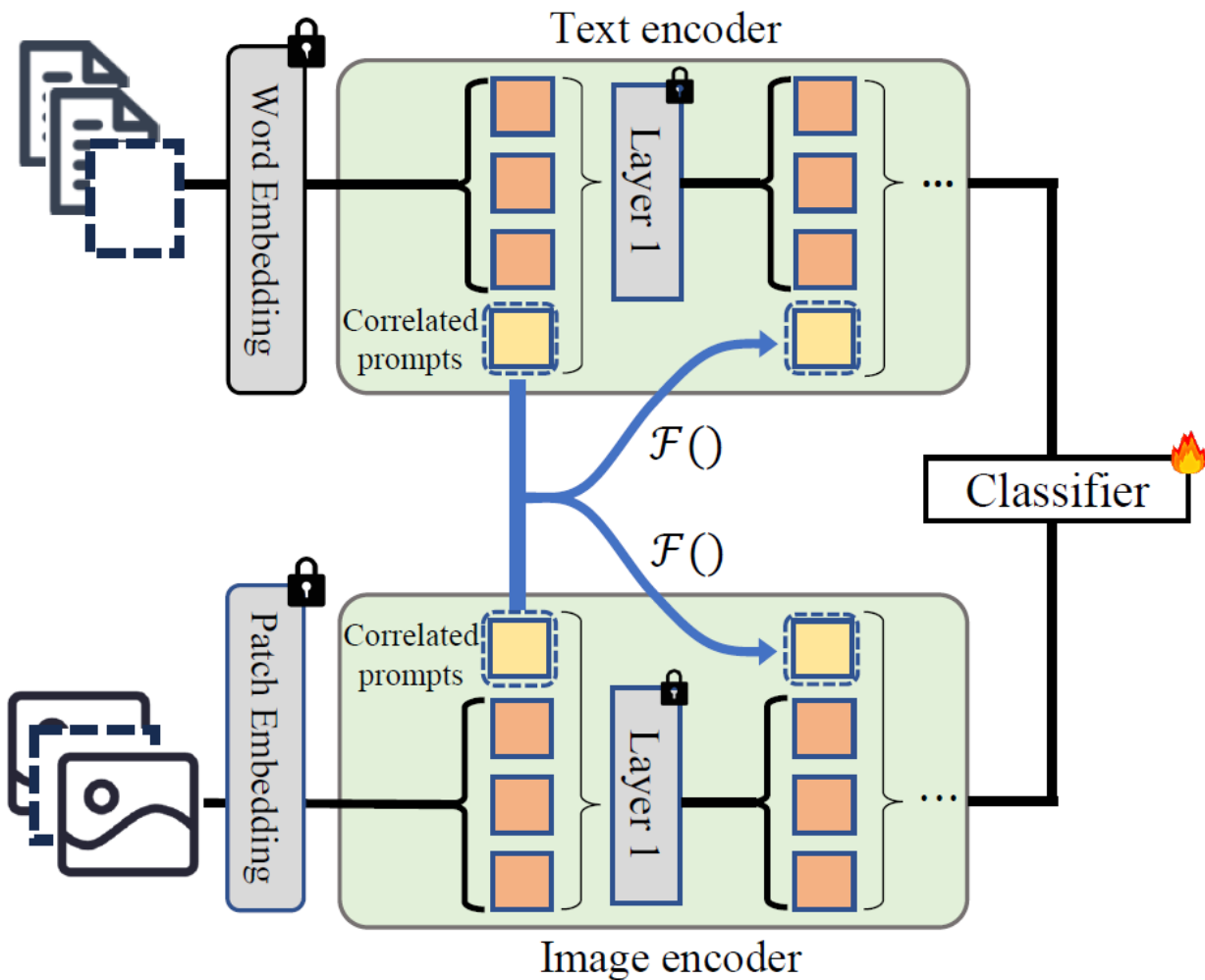
Framework overview



We propose

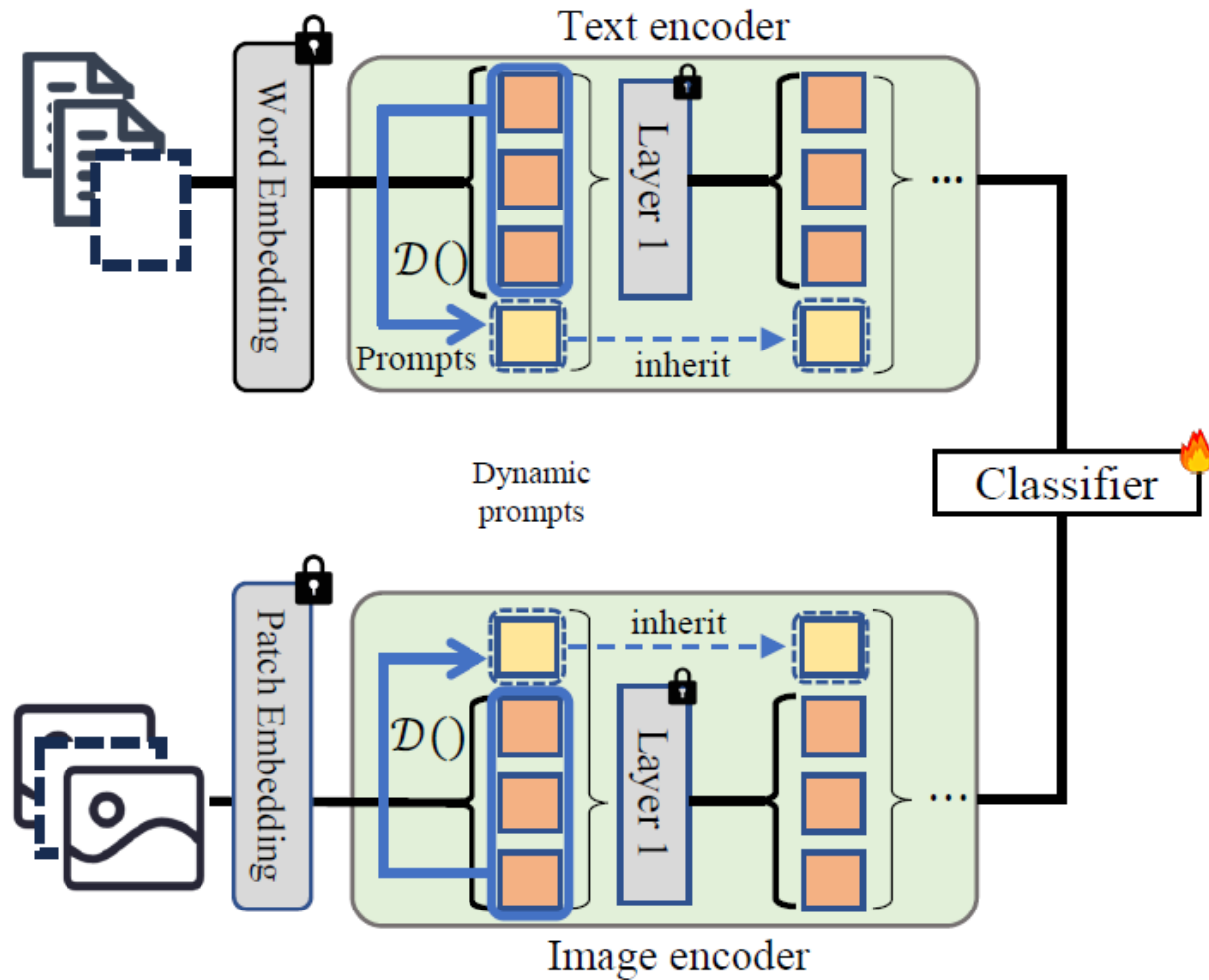
- (1) correlated prompts, which capture relations between different layers;
- (2) dynamic prompts, which dynamically compute the prompts based on inputs;
- (3) modal-common prompts, which store the shared information across different modalities.

Correlated prompts



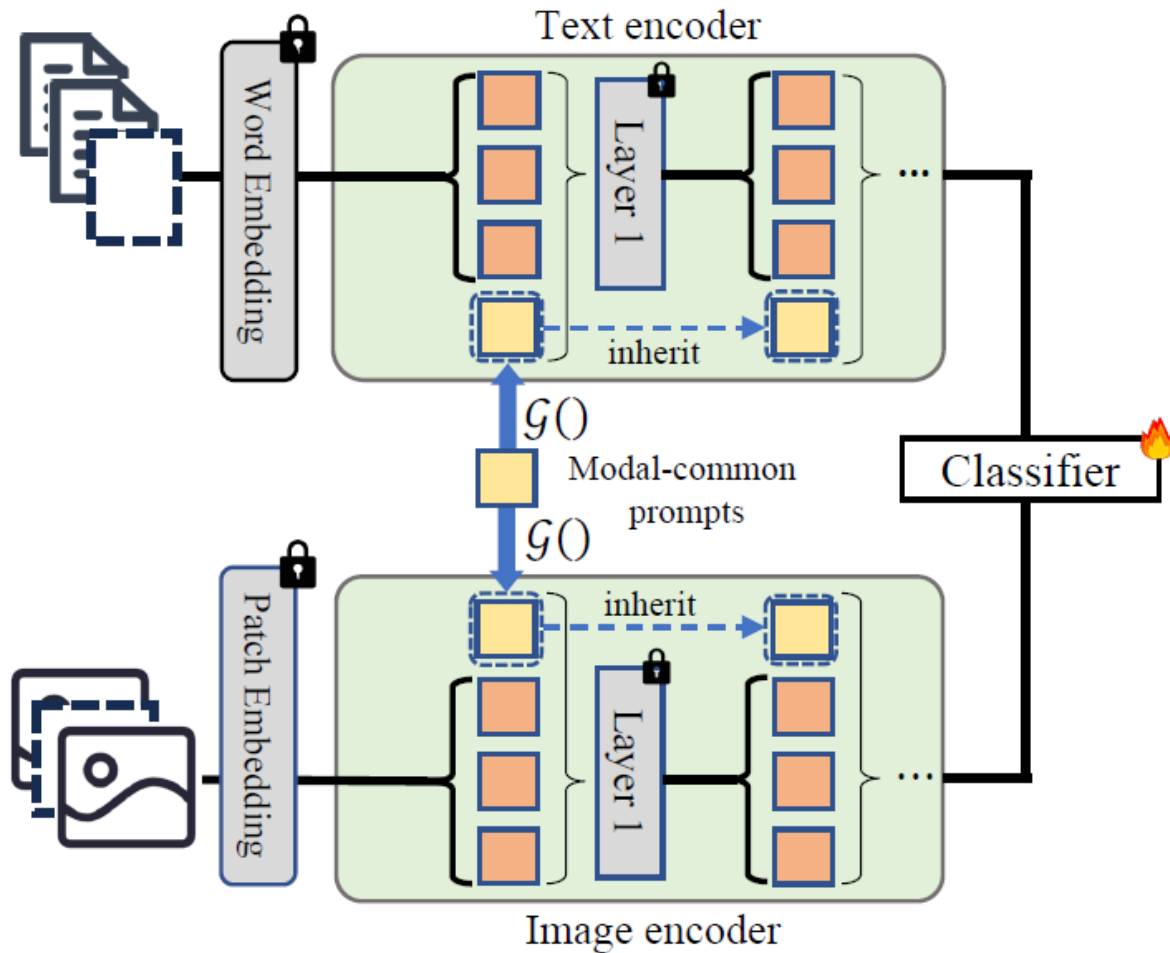
It generates the prompts of the next layer based on the prompts of both modalities in the current layer to enable cooperation of prompts from both modalities.

Dynamic prompts



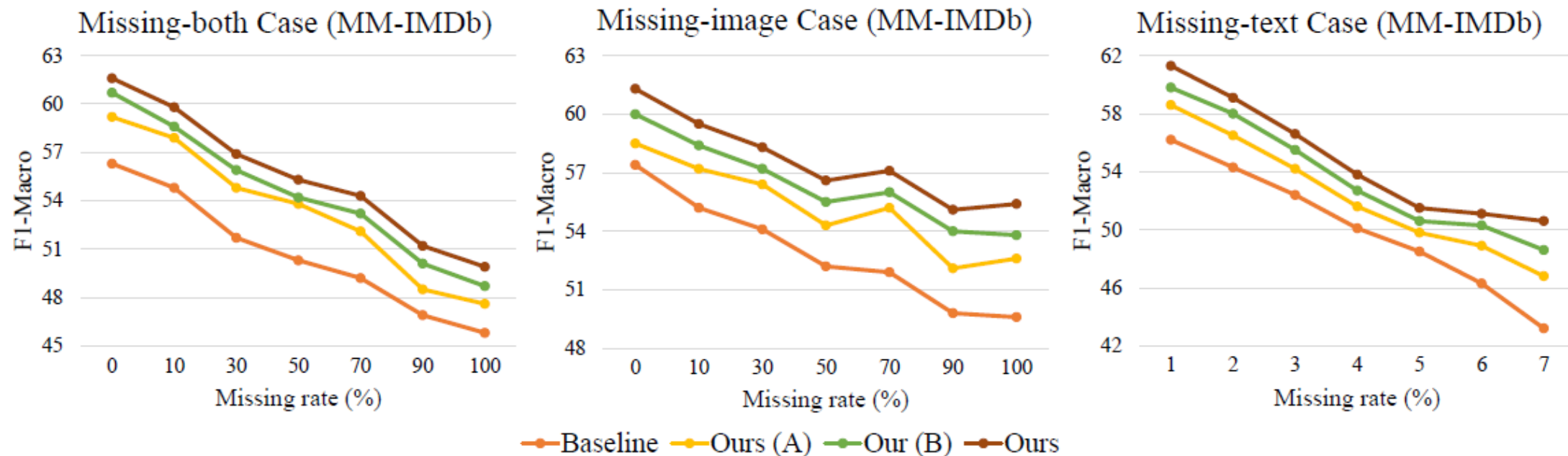
It dynamically computes the prompts based on different input features to better guide the behavior of the model, avoiding using fixed prompts for different inputs.

Modal-common prompts



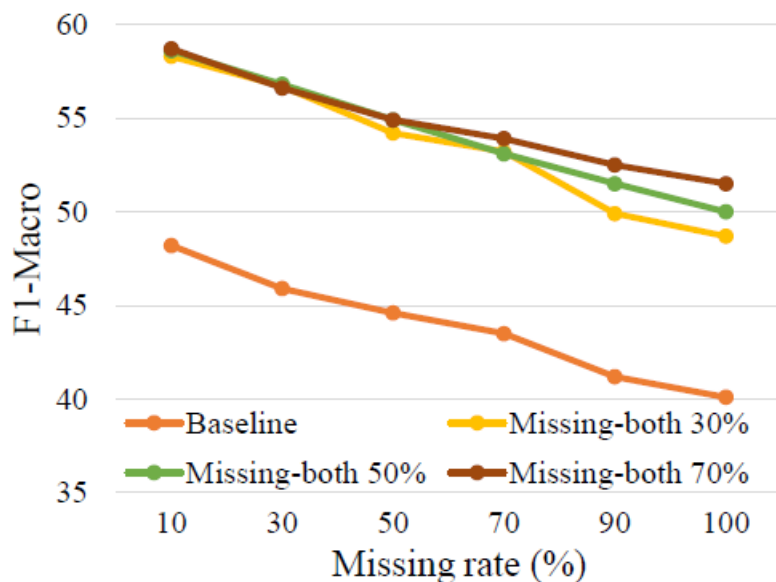
It stores the shared information across different modalities and facilitate the model to encode modal-specific information to better handle the missing scenarios in each modality.

Effectiveness

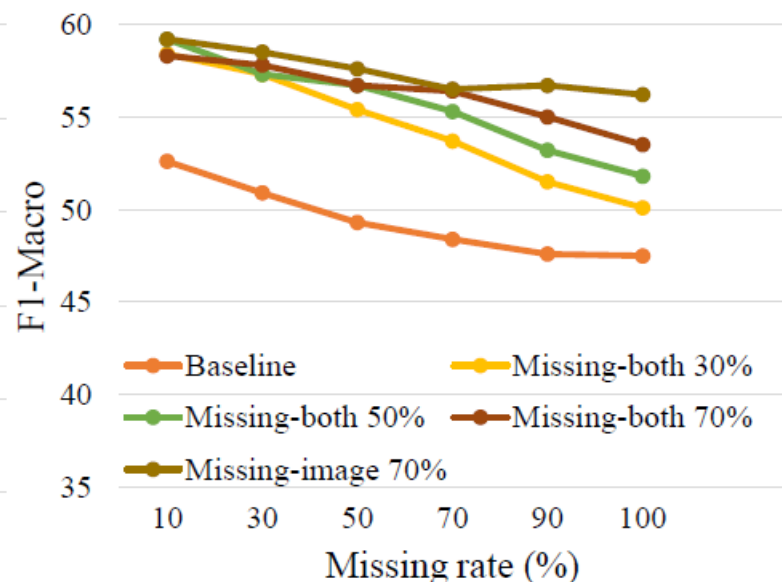


(1) baseline, which directly drops the features when a modality is missing; (2) Ours (A), which only equips the correlated prompts; (3) Ours (B), which equips both the correlated prompts and the dynamic prompts. The experiments are conducted on the val set of MM-IMDb dataset across different missing rates (0–100%) upon three different missing-modality scenarios (missing-both, missing-image and missing-text).

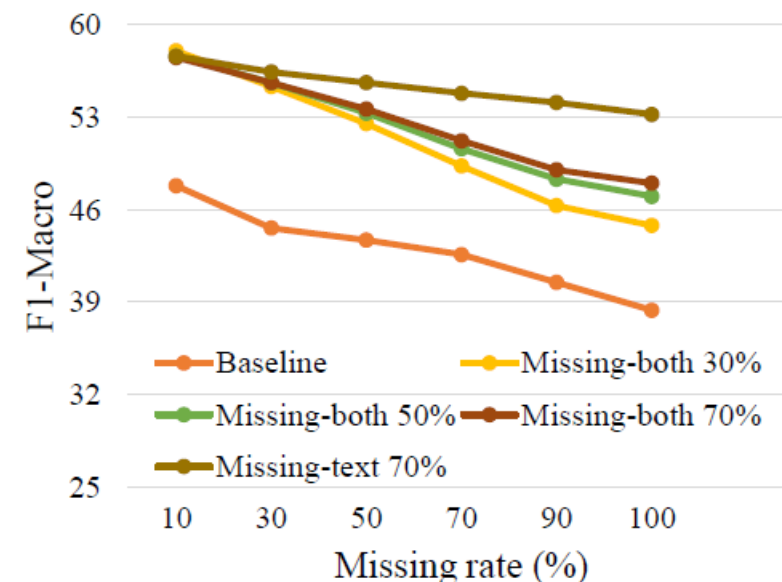
Robustness



(a) Test with missing-both



(b) Test with missing-image



(c) Test with missing-text

- (a) All models are trained on missing-both cases, and evaluated on missing-both cases with different missing rates.
- (b) Models are trained on missing-both or missing-image cases, and evaluated on missing-image cases with different missing rates.
- (c) Models are trained on missing-both or missing-text cases, and evaluated on missing-text cases with different missing rates.

Comparison with other methods

Datasets	Missing rate η	Train/Test		Validation set					Testing set				
		Image	Text	CoOp	MMP	MaPLe	DePT	Ours	CoOp	MMP	MaPLe	DePT	Ours
MM-IMDb (F1-Macro)	50%	100%	50%	51.23	52.07	52.76	53.87	55.23	48.06	48.88	49.58	50.64	52.13
		50%	100%	53.04	54.52	55.26	56.04	57.32	49.89	51.46	52.32	52.78	54.32
		75%	75%	51.46	52.12	52.87	54.02	55.45	48.37	49.32	49.56	50.87	52.32
	70%	100%	30%	47.26	48.23	48.75	49.87	51.35	44.13	45.64	45.52	46.38	48.52
		30%	100%	52.32	53.21	53.98	55.04	56.21	48.82	50.52	50.64	52.13	53.14
		65%	65%	50.22	51.34	52.31	53.17	54.24	46.84	48.12	49.16	50.32	51.42
	90%	100%	10%	47.86	48.84	50.12	50.98	52.36	44.76	45.32	46.84	47.56	49.26
		10%	100%	51.65	52.36	53.14	54.12	55.42	48.32	49.12	50.13	50.88	52.22
		55%	55%	47.44	48.04	48.82	49.98	51.26	44.12	44.87	45.12	46.54	48.04
Food101 (Accuracy)	50%	100%	50%	77.36	78.24	79.87	80.24	82.33	77.45	77.89	79.64	80.16	82.11
		50%	100%	86.98	87.12	87.48	87.85	89.23	87.02	87.16	87.35	82.14	89.12
		75%	75%	81.76	81.98	82.58	83.26	85.25	81.24	81.72	82.34	83.12	85.24
	70%	100%	30%	76.65	76.74	76.87	76.87	79.18	76.34	76.52	77.02	77.34	78.87
		30%	100%	85.21	86.12	86.36	86.52	87.53	84.78	85.64	85.89	86.12	87.32
		65%	65%	79.14	79.56	80.06	81.85	82.38	78.87	79.12	79.84	81.46	81.87
	90%	100%	10%	72.65	73.74	73.25	74.22	75.54	71.87	73.14	73.46	74.12	75.26
		10%	100%	82.16	82.78	83.42	84.02	86.26	81.67	82.14	83.12	83.56	85.78
		55%	55%	77.36	77.78	78.26	78.66	80.39	76.46	76.58	77.85	78.12	79.87
Hateful Memes (AUROC)	50%	100%	50%	58.32	58.56	58.78	59.31	60.24	60.56	60.31	60.87	61.87	62.32
		50%	100%	60.34	61.12	61.34	61.78	62.34	62.41	62.35	63.13	63.88	64.46
		75%	75%	62.34	62.87	63.14	63.24	63.78	64.87	65.84	65.46	65.86	66.02
	70%	100%	30%	58.54	59.02	59.36	60.02	60.56	60.74	61.12	61.26	61.56	62.82
		30%	100%	60.12	60.78	61.32	61.54	62.32	62.74	63.24	63.14	63.48	64.12
		65%	65%	62.34	62.56	63.12	63.32	63.78	64.82	65.04	65.23	65.48	66.08
	90%	100%	10%	58.02	57.34	58.32	59.02	60.34	60.03	57.21	60.74	61.14	62.08
		10%	100%	59.02	59.32	60.21	60.56	61.34	61.46	61.52	61.87	62.42	63.87
		55%	55%	62.32	62.56	63.24	63.78	64.34	64.32	63.34	64.85	65.37	66.78