Probabilistic Conformal Distillation for Enhancing Missing Modality Robustness

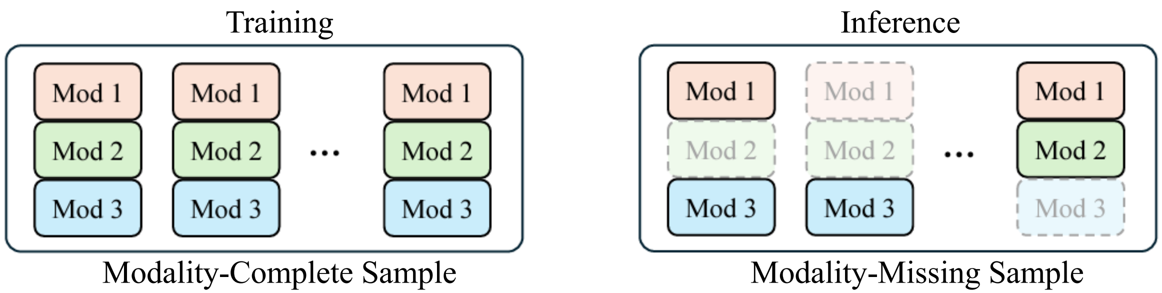
Mengxi Chen, Fei Zhang, Zihua Zhao, Jiangchao Yao, Ya Zhang, Yanfeng Wang

2024.12

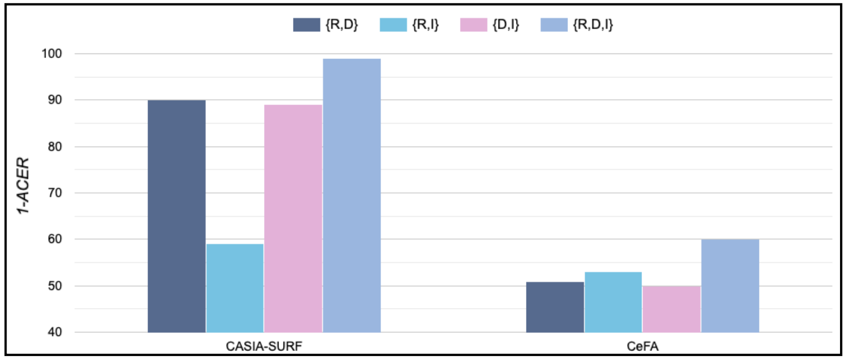
上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

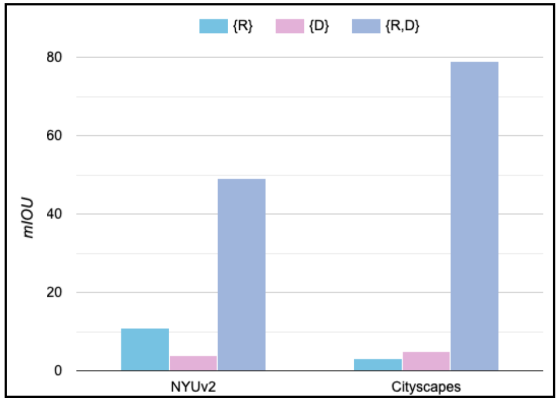
Background

What is Missing Modality Inference?



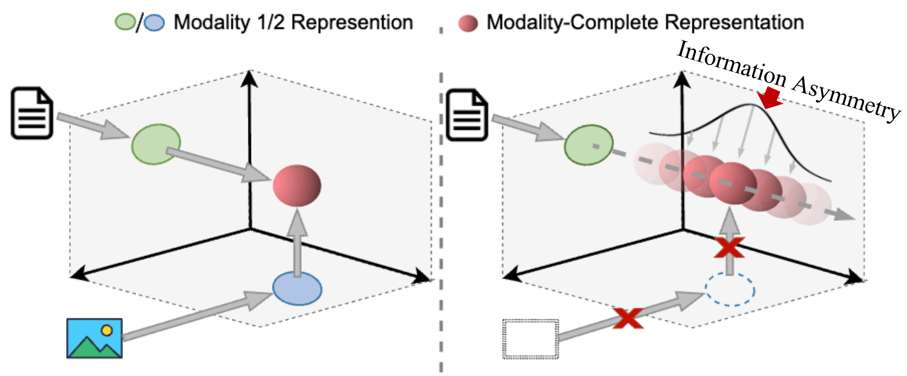
Multimodal models **trained on modality-complete samples** but **tested on modality-missing samples.**





Deteriorate remarkably !!!

Motivation



When partial modalities are missing, the retaining information is merely correlated to that of modality-complete input in a probabilistic sense.

Objective: Transfer privileged information of modality-complete representation by considering the indeterminacy in the mapping from incompleteness to completeness.

$$z_i^{\star} = \arg\max_{z_i \in Z} p(z_i|\mathrm{x}_i),$$

x_i: modality-missing sample z_i^*: modality-complete representation

Method——Probabilistic Conformal Distillation

Modeling a distribution to learn the PDF by satisfying two key characteristics:

Probability Extremum

- Points **closer to** the modality-complete representation have **high probabilities.**
- Points **farther away** the modality-complete representation have **low probabilities.**

$$q(z_p^\star \in Z_p|\mathrm{x}_i) \gg q(z_n^\star \in Z_n|\mathrm{x}_i) \approx 0.$$

Z_p: Positive set of modality-complete representations
Z_n: Negative set of modality-complete representations

Geometric Conformality

- The relation of peak points of modeled distributions $\xleftrightarrow{\textbf{conformal}}$ The relation of modality-complete representations:

$$s(g_p^\star \in G_p, g_i) \gg s(g_n^\star \in G_n, g_i),$$

G_p: Positive set of modality-complete geometric vectors
G_n: Negative set of modality-complete geometric vectors

Objective Function:

$$\max \frac{\prod_{g_p^\star \in G_p} s(g_p^\star, g_i) \prod_{z_p^\star \in Z_p} q(z_p^\star|\mathrm{x}_i)}{\prod_{z_n^\star \in Z_n} q(z_n^\star|\mathrm{x}_i)}. \Longrightarrow \max \underbrace{\left(\sum_{z_p^\star \in Z_p} \log q(z_p^\star|\mathrm{x}_i) - \sum_{z_n^\star \in Z_n} \log q(z_p^\star|\mathrm{x}_i)\right)}_{\text{Probability Extremum}} + \underbrace{\sum_{g_p^\star \in G_p} \log s(g_p^\star, g_i)}_{\text{Geometric Consistency}}.$$

Method——Probabilistic Conformal Distillation

■ Multimodal Probabilistic Modeling

$$q(z_i|\mathrm{x}_i) \sim \mathcal{N}\left(z_i; \mu_i, \sigma_i^2\right), \text{ where } \mu_i = f\left(x_i\right), \sigma_i = h\left(\mu_i\right).$$

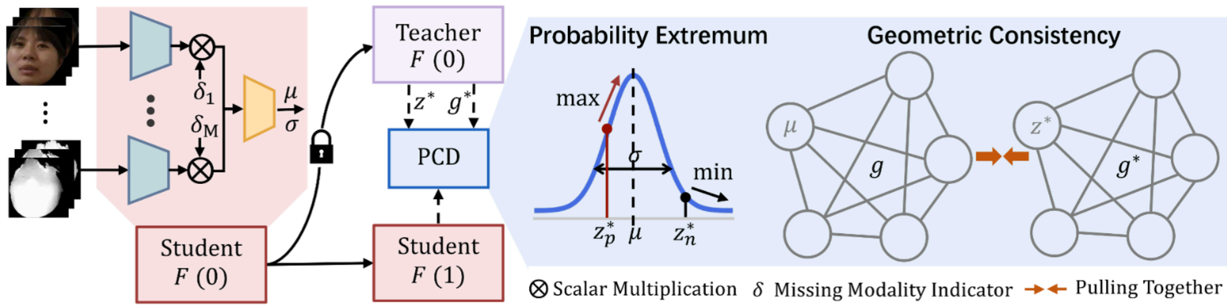
■ Probability Extremum

$$\underbrace{\left(\sum_{z_p^\star \in Z_p} \log q(z_p^\star|\mathrm{x}_i) - \sum_{z_n^\star \in Z_n} \log q(z_p^\star|\mathrm{x}_i)\right)}_{\text{Probability Extremum}} \Rightarrow \mathcal{L}_u = \sum_{\{p|y_p=y_i\}} \sum_d \left(\frac{(z_{p,d}^\star - \mu_{i,d})^2}{2(\sigma_{i,d})^2} + \log \sigma_{i,d}\right) - \sum_{\{n|y_n \neq y_i\}} \sum_d \left(\frac{(z_{n,d}^\star - \mu_{i,d})^2}{2(\sigma_{i,d})^2} + \log \sigma_{i,d}\right)$$

■ Geometric Conformality

$$s(g_p^\star, g_i) = \frac{\exp(\beta(g_p^\star, g_i)/\tau)}{\exp(\beta(g_p^\star, g_i)/\tau) + \sum_{\{n|y_n \neq y_i\}} \exp(\beta(g_n^\star, g_i)/\tau)}, \quad g_i^\star(b) = \alpha(z_i^\star, z_b^\star), \; g_i(b) = \alpha(\mu_i, \mu_b),$$

$$\underbrace{\sum_{g_p^\star \in G_p} \log s(g_p^\star, g_i)}_{\text{Geometric Consistency}} \Rightarrow \mathcal{L}_g = -\sum_{\{p|y_p=y_i\}} \log s(g_p^\star, g_i),$$



■ Overall Loss

$$\mathcal{L} = \mathcal{L}_t + \lambda(\mathcal{L}_u + \mathcal{L}_g),$$

Experiments

Table 1: Performance under different modality-missing inference condition on two classification datasets and two segmentation datasets.

Method	CASIA-SURF (ACER ↓) {R}	{D}	{I}	{R,D}	{R,I}	{D,I}	{R,D,I}	Average
Traditional [49]	23.03	17.10	49.53	10.40	41.02	11.26	1.40	22.11
Separate Model [49]	10.01	4.45	11.65	3.41	6.32	3.54	1.23	5.80
Augmentation [1]	11.75	5.87	16.62	4.61	6.68	4.95	2.21	7.52
HeMIS [15]	14.36	4.70	16.21	3.23	6.27	3.68	1.97	7.18
MMFormer [50]	11.15	4.67	13.99	1.93	4.77	3.10	1.94	5.93
MMANET [46]	8.57	2.27	10.04	1.61	3.01	1.18	0.87	3.94
MD [12]	10.84	6.65	19.43	12.64	7.84	3.99	0.96	7.30
ETMC [14]	7.91	4.73	7.54	1.39	4.56	1.46	0.76	4.05
RAML [6]	11.26	3.10	11.65	1.92	5.35	1.76	1.09	5.16
PCD	**7.23**	**2.20**	**5.66**	**0.99**	**2.86**	**0.89**	**0.74**	**2.93**
Δ	0.74%↓	0.07%↓	1.88%↓	0.40%↓	0.15%↓	0.29%↓	0.02%↓	1.01%↓

Method	CeFA (ACER ↓) {R}	{D}	{I}	{R,D}	{R,I}	{D,I}	{R,D,I}	Average
Traditional [49]	50.00	50.00	49.96	49.25	47.28	48.95	39.62	47.86
Separate Model [49]	27.44	33.75	36.17	35.62	31.62	36.62	24.15	32.20
Augmentation [1]	27.93	36.90	36.14	32.10	28.47	35.12	31.87	32.65
HeMIS [15]	34.14	37.97	36.94	36.02	33.94	31.92	40.66	35.94
MMFormer [50]	28.51	33.58	39.56	29.47	27.66	32.17	30.72	31.52
MMANET [46]	27.15	32.50	35.62	22.87	23.27	30.45	23.68	27.94
MD [12]	27.13	35.81	37.99	26.25	31.29	34.69	30.49	31.95
ETMC [14]	24.74	34.28	37.62	22.52	24.25	30.63	21.59	27.95
RAML [6]	28.54	33.88	40.01	23.82	28.81	28.85	22.11	29.43
PCD	**21.38**	**28.01**	**34.79**	**17.19**	**20.92**	**21.68**	**14.39**	**22.63**
Δ	3.36%↓	4.49%↓	0.83%↓	5.33%↓	2.35%↓	5.75%↓	7.20%↓	5.31%↓

Method	NYUv2 (mIOU ↑) {R}	{D}	{R,D}	Average	Cityscapes (mIOU ↑) {R}	{D}	{R,D}	Average
Traditional [36]	11.15	4.18	48.78	21.41	3.17	4.87	78.73	28.89
Separate Model [36]	44.22	40.55	48.89	44.55	77.60	59.11	78.62	71.77
Augmentation [1]	41.34	39.76	47.23	42.77	76.89	57.42	78.13	70.81
MMFormer [50]	43.22	41.12	48.45	44.26	76.62	58.53	78.01	71.05
MMANET [46]	44.93	42.75	**49.62**	45.58	77.61	60.12	78.89	72.20
PCD	**45.68**	**44.34**	49.44	**46.49**	**78.26**	**61.30**	**79.53**	**73.03**
Δ	0.75%↑	1.59%↑	0.18%↓	0.91%↑	0.65%↑	1.18%↑	0.64%↑	0.83%↑

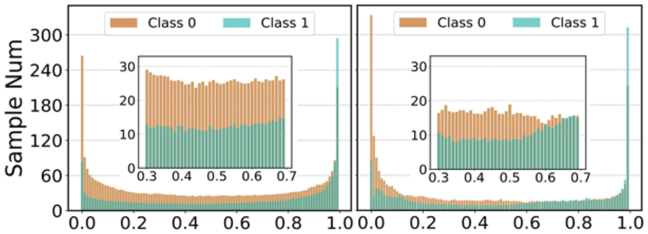
Table 2: Ablation Study

$\mathcal{L}_c$	$\mathcal{L}_u$	$\mathcal{L}_g$	CASIA-SURF {R}	{D}	{I}	{R,D}	{R,I}	{D,I}	{R,D,I}	Average
✓	×	×	12.31	2.89	19.24	1.31	8.16	2.19	1.35	6.78
✓	×	✓	13.55	**2.01**	18.02	**0.86**	5.81	2.53	0.85	6.24
✓	✓	×	7.59	4.10	7.97	1.83	3.86	2.04	0.97	4.05
✓	✓	✓	**7.23**	2.20	**5.66**	0.99	**2.86**	**0.89**	**0.74**	**2.93**

$\mathcal{L}_c$	$\mathcal{L}_u$	$\mathcal{L}_g$	CeFA {R}	{D}	{I}	{R,D}	{R,I}	{D,I}	{R,D,I}	Average
✓	×	×	26.95	38.06	37.06	24.18	24.75	32.82	25.38	29.89
✓	✓	×	21.14	33.76	37.22	21.28	23.61	27.56	21.19	26.53
✓	×	✓	**20.62**	34.43	35.23	18.18	21.86	32.63	21.72	26.38
✓	✓	✓	21.38	**28.01**	**34.79**	**17.19**	**20.92**	**21.68**	**14.39**	**22.63**

$\mathcal{L}_c$	$\mathcal{L}_u$	$\mathcal{L}_g$	NYUv2 {R}	{T}	{R,T}	Average	Cityscapes {R}	{T}	{R,T}	Average
✓	×	×	44.24	41.17	47.89	44.43	77.54	59.64	78.46	71.89
✓	×	✓	45.96	42.95	48.54	45.82	78.11	60.62	79.07	72.60
✓	✓	×	44.48	42.02	48.86	45.12	77.52	59.94	78.91	72.17
✓	✓	✓	**45.68**	**44.34**	**49.44**	**46.49**	**78.26**	**61.30**	**79.53**	**73.03**

Experiments



(a) Warm-up (b) MPCD

Figure 3: The prediction distributions of both the teacher and the distilled student of PCD under all multimodal combinations on CeFA. The X-axis represents the normalized logit output and the Y-axis is the number of samples after taking the square root.

Table 3: Performance under different modality-missing inference condition with **modality-missing training data**.

Missing	Method	CASIA-SURF (ACER ↓)							
		{R}	{D}	{I}	{R,D}	{R,I}	{D,I}	{R,D,I}	Average
30%	MMANET [46]	13.50	3.38	**6.57**	6.57	3.72	1.83	1.31	4.67
	ETMC [14]	**7.63**	3.62	10.18	**1.12**	5.21	**1.43**	0.96	4.31
	PCD	8.28	**2.13**	6.66	1.24	**2.66**	2.66	**0.60**	**3.18**
	Δ	0.65%↑	1.25%↓	0.09%↑	0.12%↑	1.06%↓	1.23%↑	0.36%↓	1.13%↓
40%	MMANET [46]	14.96	5.22	9.03	3.24	5.14	2.31	2.10	6.00
	ETMC [14]	9.38	7.42	**7.44**	1.41	3.98	3.16	**0.58**	4.77
	PCD	**7.14**	**1.77**	10.88	**1.08**	**3.70**	**1.10**	0.88	**3.79**
	Δ	2.24%↓	3.45%↓	3.44%↑	0.33%↓	0.28%↓	1.21%↓	0.30%↑	0.98%↓

Missing	Method	CeFA (ACER ↓)							
		{R}	{D}	{I}	{R,D}	{R,I}	{D,I}	{R,D,I}	Average
30%	MMANET [46]	28.39	39.61	**34.12**	34.19	23.39	34.12	27.11	31.56
	ETMC [14]	25.96	34.69	38.60	24.15	24.58	31.83	24.03	29.12
	PCD	**23.42**	**30.23**	34.60	**18.34**	**21.98**	**24.50**	**15.07**	**23.73**
	Δ	2.54%↓	4.46%↓	0.48%↑	5.81%↓	1.41%↓	7.33%↓	8.96%↓	5.39%↓
40%	MMANET [46]	29.94	43.40	37.29	31.60	28.62	44.97	31.80	35.38
	ETMC [14]	**24.38**	37.82	38.33	25.04	24.39	36.96	24.03	30.13
	PCD	24.91	**31.23**	**34.40**	**21.09**	**23.98**	**23.31**	**16.30**	**25.03**
	Δ	0.53%↑	6.58%↓	2.89%↓	3.95%↓	0.40%↓	13.65%↓	7.73%↓	5.10%↓

Conclusion

- We propose a Probabilistic Conformal Distillation (PCD) method to handle the missing modality problem, which transfers privileged information of modality-complete representation by considering the indeterminacy in the mapping from incompleteness to completeness.
- We parameterize different modality-missing representations as distinct distributions to fit their unknown PDFs in the modality-complete space. This is specially realized by considering the probabilities of extreme points and ensuring the geometric consistency between peak points of different PDFs and modeled distributions.
- We conduct comprehensive experiments to demonstrate the effectiveness of PCD across a range of modality-missing scenarios. Extensive comparison on multimodal classification and segmentation tasks consistently validate the superior performance of our method compared to the state-of-the-art approaches.