MoE-DiffIR: Task-customized Diffusion Priors for Universal Compressed Image Restoration

Yulin Ren[1], Xin Li[1] (✉), Bingchen Li[1], Xingrui Wang[1], Mengxi Guo[2], Shijie Zhao[2], Li Zhang[2], and Zhibo Chen[1] (✉)

[1] University of Science and Technology of China, Hefei, Anhui, China
[2] Bytedance Inc., Beijing, China

ECCV 2024

Presenter: Hao Wang

Advisor: Prof. Chia-Wen Lin

Outline

- Introduction
- Framework
- Method
- Experiment
- Conclusion

Outline

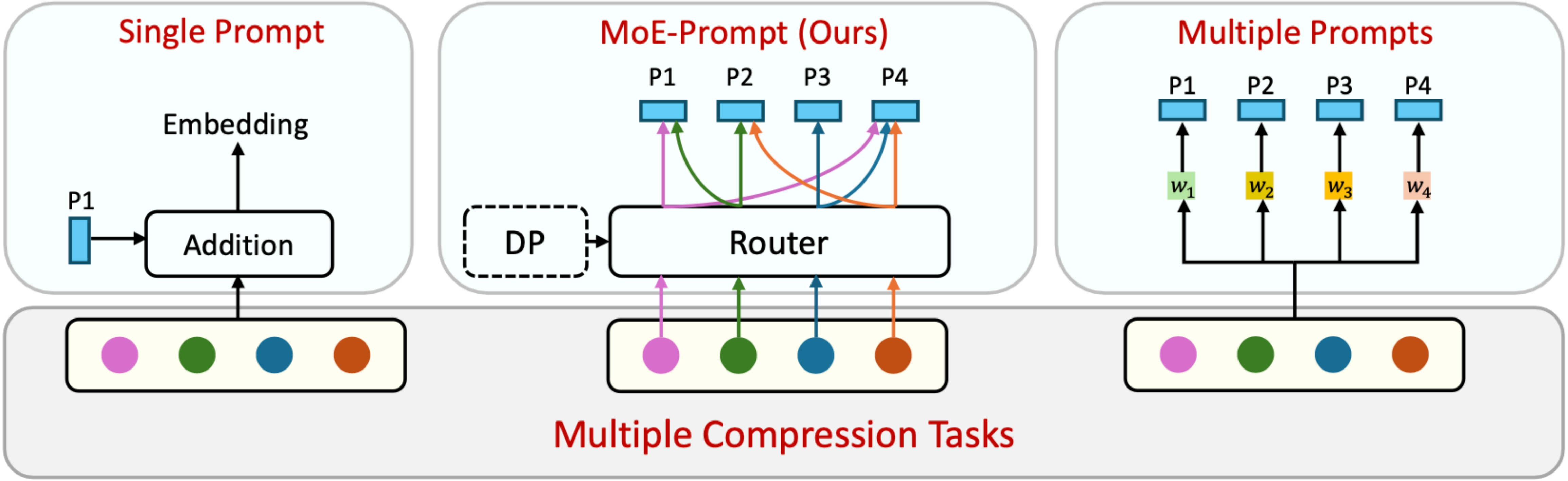
- Introduction
- Framework
- Method
- Experiment
- Conclusion

Introduction

- Challenges
  - lacking adaptability and universality for different image codecs
  - poor texture generation capability, particularly at low bitrates

- Method
  - develops the powerful **mixture-of-experts (MoE) prompt** module, the degradation-aware routing mechanism enable the flexible assignment of basic prompts
  - design the **visual-to-text adapter**, adapting the embedding of low-quality images from the visual domain to the textual domain as the textual guidance for SD, enabling more consistent and reasonable texture generation

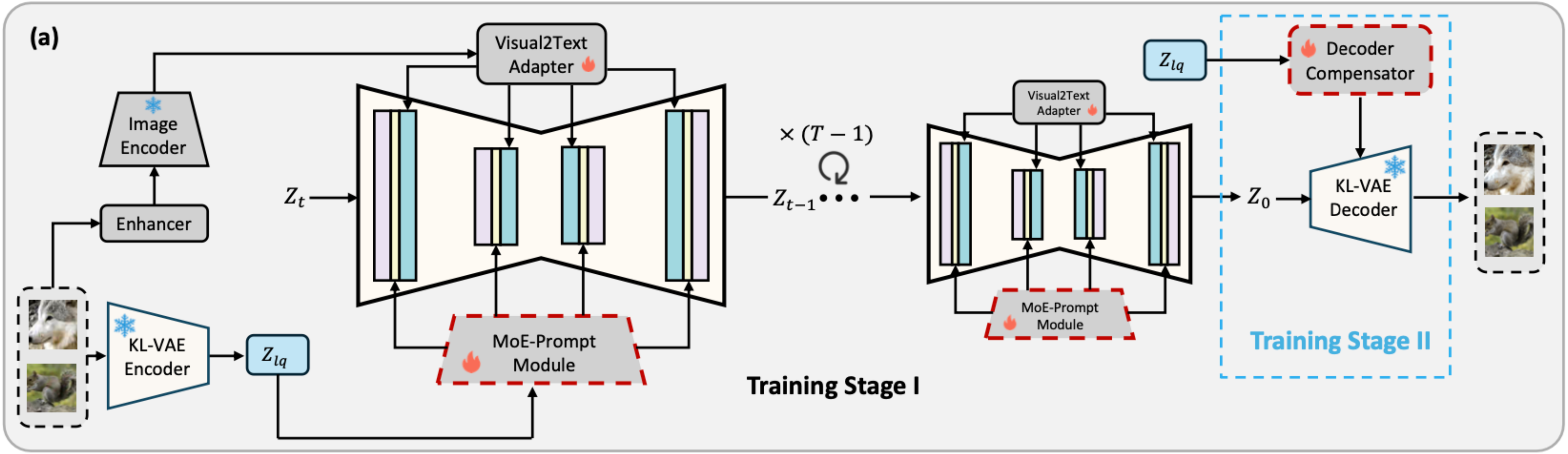
Introduction



Outline

- Introduction
- Framework
- Method
- Experiment
- Conclusion

Framework



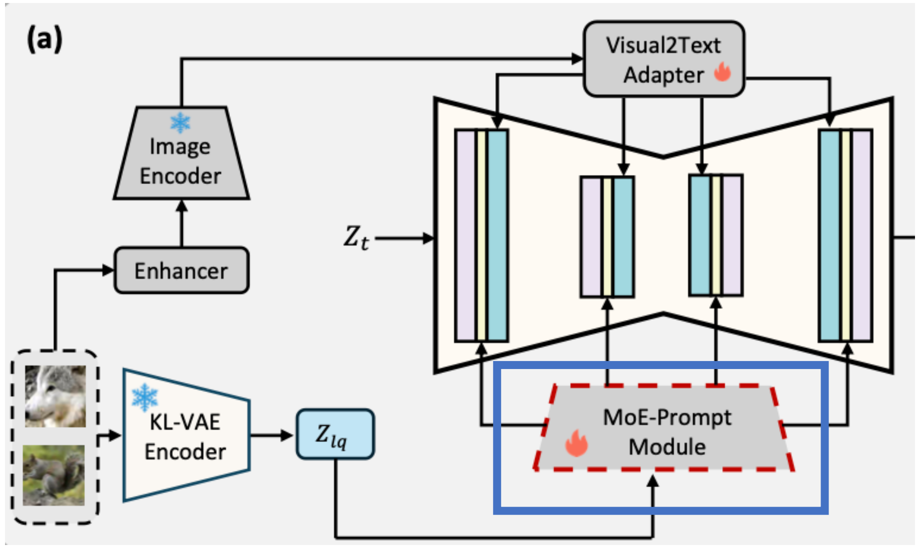
(a)
Visual2Text Adapter
Image Encoder
Enhancer
KL-VAE Encoder
Z_{lq}
Z_t
MoE-Prompt Module
$\times (T-1)$
Z_{t-1}
Visual2Text Adapter
MoE-Prompt Module
Training Stage I
Z_{lq}
Decoder Compensator
Z_0
KL-VAE Decoder
Training Stage II

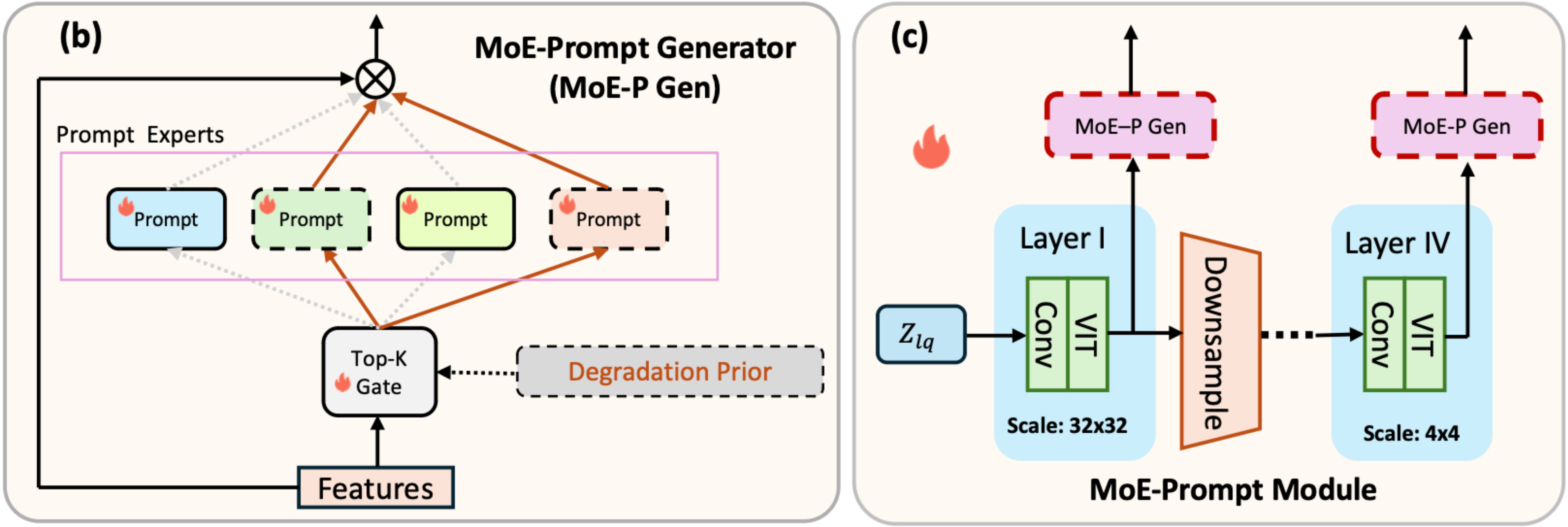
Outline

- Introduction
- Framework
- Method
- Experiment
- Conclusion

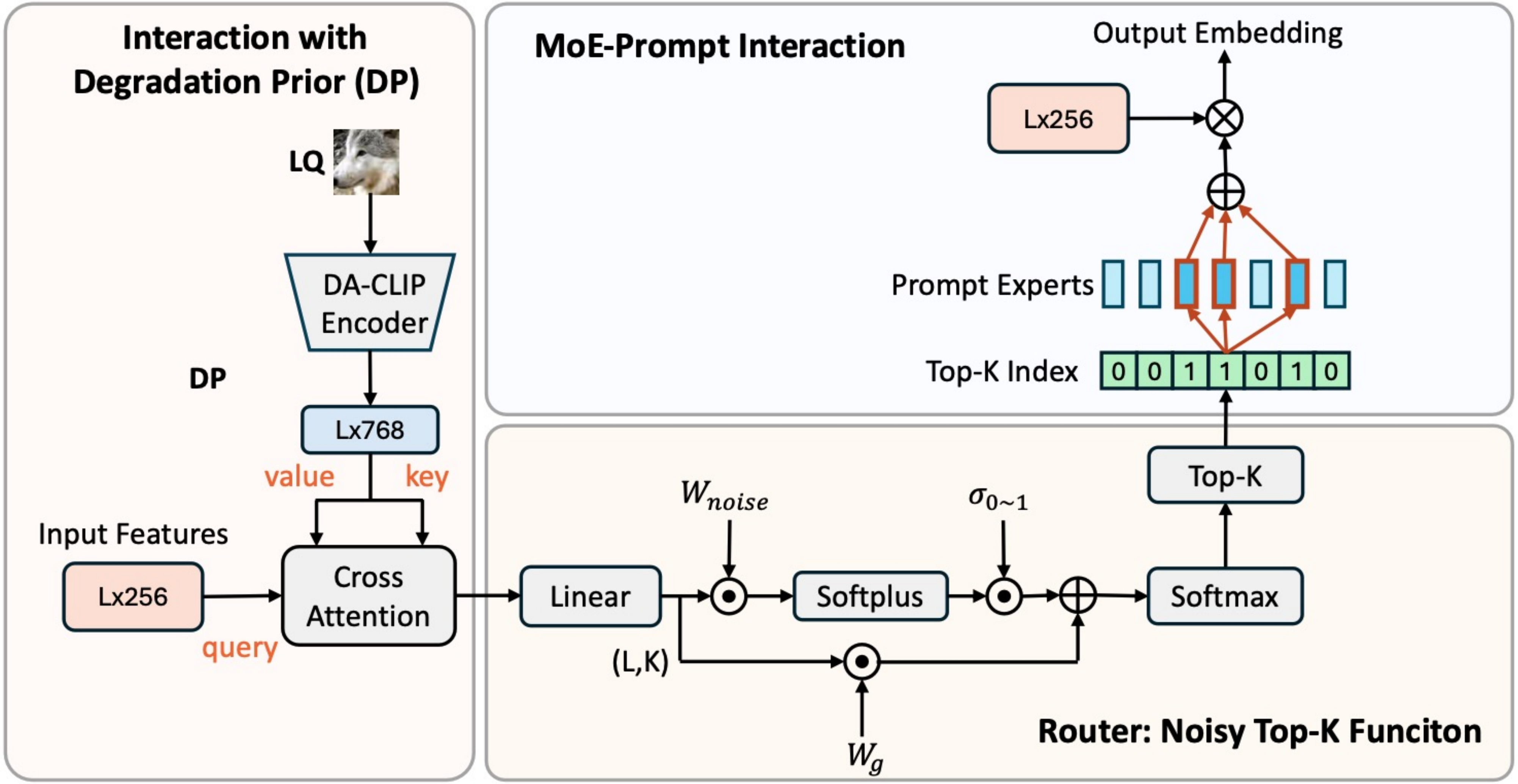
MoE-Prompt Module

$$y = \sum_{i=1}^{n} G(x)_i E_i(x) \qquad (1)$$



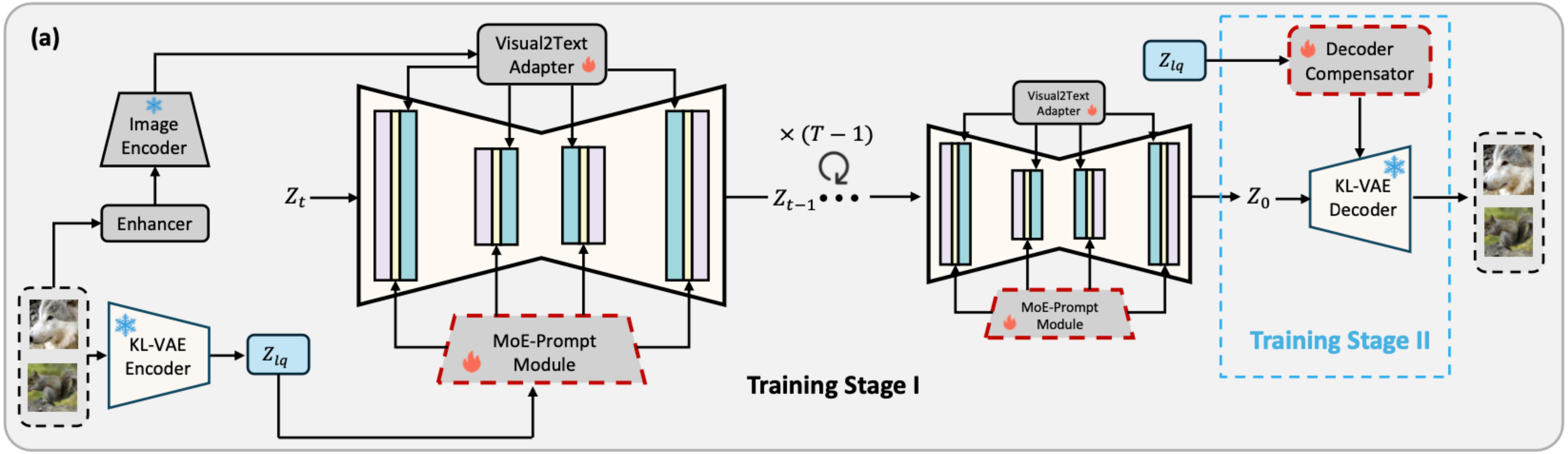


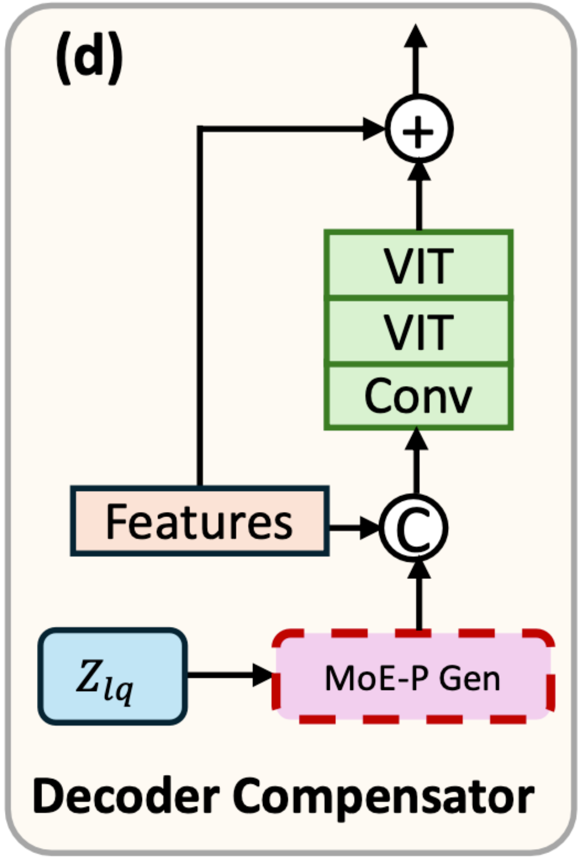
MoE-Prompt Generator



$$G(x) = \text{Top-K}(\text{Softmax}(xW_g + \mathcal{N}(0,1)\text{Softplus}(xW_{\text{noise}}))) \tag{3}$$

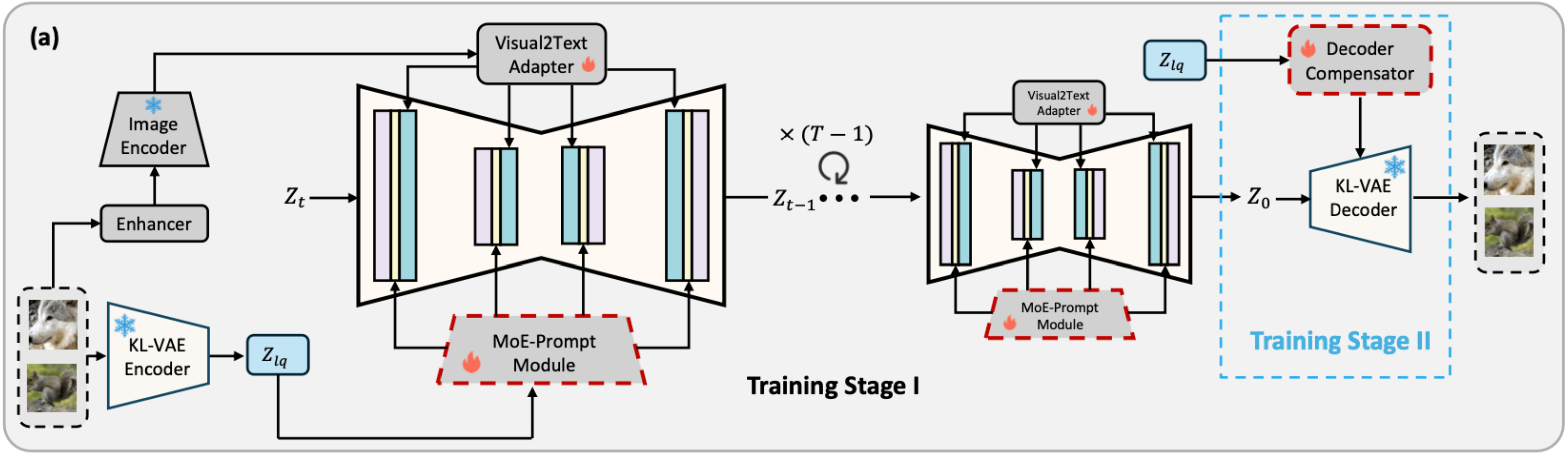
Visual2Text Adapter and Decoder Compensator





- Stable Diffusion, trained on large-scale datasets, stores an **abundance of text-to-image priors**
- employ several MLP layers to translate visual information into the textual domain of SD

Fine-tuning Procedure



$$\mathcal{L}_{\mathrm{SD}} = \mathbb{E}_{\epsilon\sim\mathcal{N}(0,1)}\left[\|\epsilon - \epsilon(z_t, t)\|_2^2\right] \quad (2)$$

$$L_{Decoder} = \mathcal{L}_{lpips}[z_{lq}, z_0, hr] \quad (4)$$

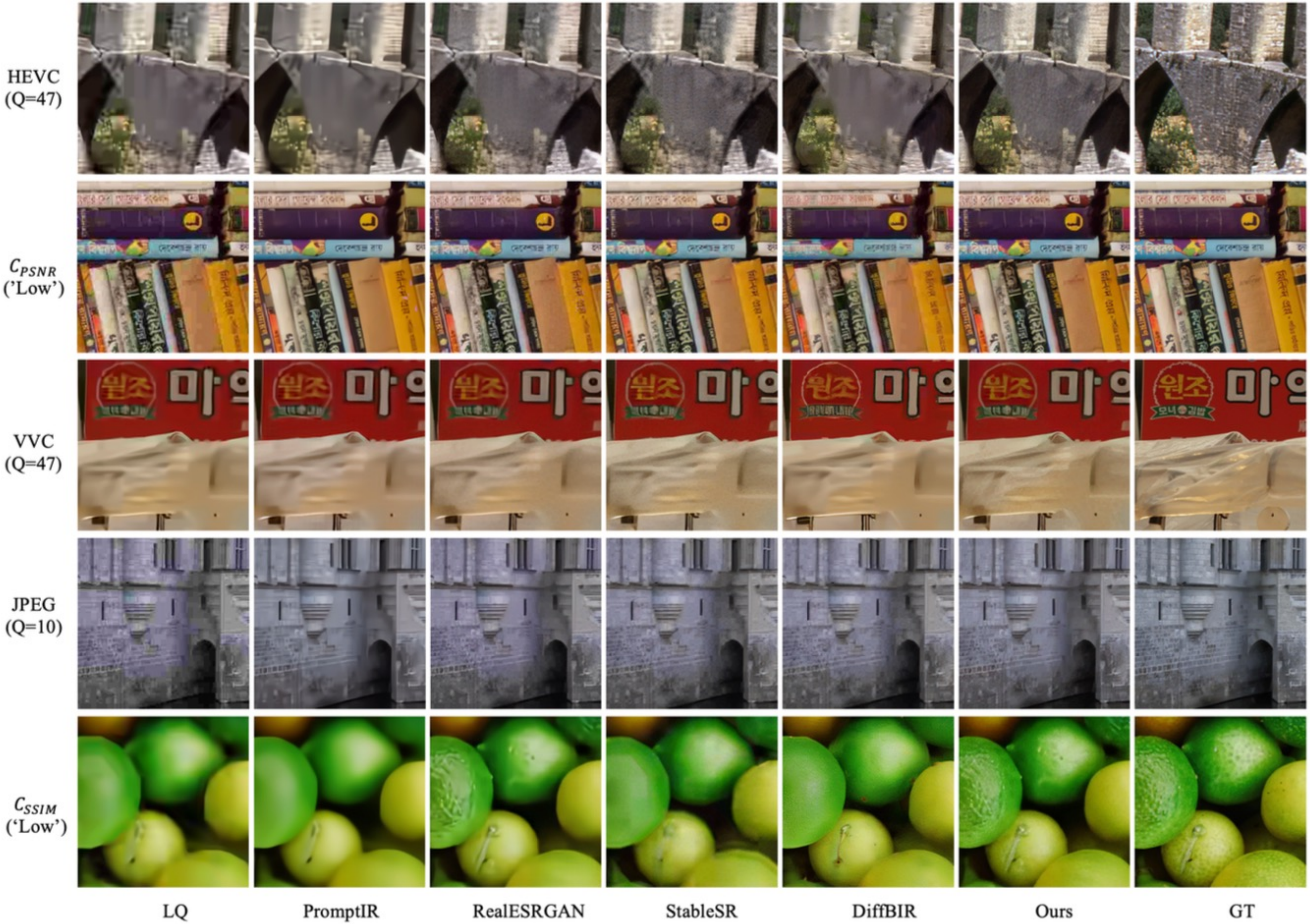
Outline

- Introduction
- Framework
- Method
- Experiment
- Conclusion

Results

Codecs	Methods	LIVE1 [60]				Classic5 [84]				BSDS500 [5]				DIV2K [3]				ICB [17]			
		PSNR	SSIM	LPIPS	FID	PSNR	SSIM	LPIPS	FID	PSNR	SSIM	LPIPS	FID	PSNR	SSIM	LPIPS	FID	PSNR	SSIM	LPIPS	FID
JPEG [64]	AirNet [33]	30.06	0.858	0.2241	113.59	35.84	0.955	0.1213	62.78	30.99	0.864	0.1975	96.56	29.91	0.872	0.2037	107.22	31.25	0.878	0.2565	179.67
	HAT [8]	30.18	0.860	0.1952	115.53	33.69	0.911	0.1185	57.80	31.07	0.865	0.1809	86.32	29.58	0.869	0.1980	107.22	30.98	0.881	0.2308	173.19
	PromptIR [54]	31.42	0.885	0.2131	111.15	35.66	0.949	0.1124	49.84	31.95	0.881	0.1791	87.50	31.14	0.896	0.1879	98.83	31.67	0.896	0.2113	167.68
	RealESRGAN [68]	30.26	0.860	0.1423	76.28	33.21	0.915	0.1106	56.49	30.24	0.848	0.1485	81.43	29.39	0.866	0.1357	74.46	30.22	0.876	0.1443	130.66
	DiffBIR [42]	28.42	0.812	0.0995	67.23	30.38	0.86	0.1026	56.95	28.67	0.810	0.1013	73.43	27.99	0.805	0.0898	65.84	29.72	0.837	0.1086	113.05
	PASD [79]	28.39	0.806	0.1055	70.46	29.97	0.839	0.0959	54.98	27.70	0.758	0.0934	70.93	28.04	0.784	0.0906	67.61	28.75	0.668	0.1736	115.67
	StableSR [66]	30.19	0.855	0.1069	68.19	31.08	0.875	0.1227	46.97	30.60	0.847	0.1127	67.94	29.44	0.863	0.1067	62.55	31.00	0.875	0.1213	125.56
	SUPIR [82]	27.41	0.747	0.1254	71.04	29.05	0.834	0.1110	60.60	27.89	0.713	0.1076	69.63	27.56	0.780	0.1362	70.23	28.78	0.673	0.1319	121.60
	Ours	30.50	0.857	0.0964	62.72	32.30	0.890	0.0902	40.83	30.95	0.852	0.1006	68.00	29.86	0.868	0.0906	55.75	31.48	0.881	0.1059	100.93
VVC [6]	AirNet [33]	29.08	0.803	0.3055	185.2	33.74	0.914	0.1715	165.37	29.71	0.796	0.3147	183.47	28.15	0.808	0.2863	155.61	28.17	0.797	0.3252	208.07
	HAT [8]	29.19	0.806	0.2801	182.68	32.27	0.874	0.1629	156.23	29.75	0.797	0.2874	170.61	28.51	0.812	0.2832	154.78	29.04	0.817	0.3069	213.37
	PromptIR [54]	30.00	0.825	0.3053	168.36	34.06	0.916	0.1662	148.57	30.32	0.811	0.2818	162.25	29.41	0.832	0.2628	145.96	29.35	0.816	0.2798	205.30
	RealESRGAN [68]	28.64	0.786	0.2082	131.83	32.56	0.884	0.1196	87.79	28.24	0.765	0.2185	130.61	27.77	0.792	0.1978	114.20	27.89	0.790	0.2118	188.82
	DiffBIR [42]	27.54	0.771	0.1687	95.49	29.64	0.811	0.1407	85.55	27.97	0.773	0.1775	111.42	27.06	0.763	0.1468	93.68	28.03	0.775	0.1746	154.91
	PASD [79]	27.48	0.766	0.1746	101.67	28.78	0.779	0.1381	86.43	26.84	0.749	0.1717	109.04	26.91	0.738	0.1355	96.20	26.80	0.708	0.1857	154.37
	StableSR [66]	28.49	0.771	0.1679	98.45	30.23	0.822	0.1318	84.59	28.75	0.756	0.1831	100.38	27.85	0.789	0.1473	92.08	28.46	0.789	0.1824	165.84
	SUPIR [82]	27.46	0.714	0.1468	99.73	28.45	0.774	0.1415	87.62	27.16	0.784	0.1759	105.71	26.49	0.733	0.1833	99.80	26.84	0.715	0.1990	159.98
	Ours	28.76	0.777	0.1444	88.83	31.01	0.845	0.1121	78.28	28.94	0.755	0.1577	84.92	28.05	0.786	0.1316	80.96	28.63	0.781	0.1540	144.51
HEVC [62]	AirNet [33]	28.70	0.792	0.3159	167.81	33.59	0.906	0.1790	172.80	29.31	0.784	0.3244	174.49	27.78	0.794	0.3010	158.35	27.35	0.827	0.3041	209.11
	HAT [8]	28.82	0.795	0.2957	167.72	31.80	0.862	0.1717	165.78	29.42	0.786	0.3002	176.55	28.19	0.800	0.2994	153.22	29.45	0.831	0.3075	232.96
	PromptIR [54]	29.54	0.814	0.3138	172.58	33.87	0.909	0.1721	157.71	29.88	0.799	0.2901	156.25	29.04	0.822	0.2677	146.35	29.85	0.834	0.2847	223.55
	RealESRGAN [68]	28.31	0.776	0.2269	139.67	32.43	0.873	0.1266	104.17	28.02	0.754	0.2355	137.33	27.53	0.781	0.2148	117.01	28.33	0.807	0.2192	200.91
	DiffBIR [42]	27.53	0.762	0.1790	104.09	29.54	0.809	0.1317	94.85	27.79	0.773	0.1784	111.20	26.96	0.751	0.1595	96.19	28.35	0.780	0.1666	176.30
	PASD [79]	27.50	0.76	0.1801	108.89	29.02	0.78	0.1300	91.91	26.07	0.747	0.1853	106.77	26.15	0.704	0.1461	100.83	27.02	0.706	0.1801	177.79
	StableSR [66]	28.19	0.759	0.1845	106.00	30.39	0.821	0.1323	85.98	28.64	0.751	0.1902	108.52	27.67	0.780	0.1656	95.47	28.84	0.803	0.1895	184.51
	SUPIR [82]	26.80	0.679	0.1605	108.69	28.39	0.777	0.1339	97.12	26.40	0.743	0.1895	107.19	26.47	0.700	0.1721	103.54	27.05	0.710	0.1938	182.37
	Ours	28.50	0.768	0.1622	98.66	31.08	0.839	0.1157	82.82	28.71	0.745	0.1749	96.23	27.83	0.777	0.1470	84.63	28.87	0.791	0.1776	171.01
WebP [20]	AirNet [33]	29.40	0.822	0.2537	154.54	34.59	0.930	0.1453	106.65	29.55	0.803	0.2701	161.94	28.31	0.831	0.2253	145.74	27.02	0.831	0.2804	221.70
	HAT [8]	29.51	0.825	0.2384	153.36	32.31	0.882	0.1404	99.30	29.68	0.804	0.2582	152.92	28.84	0.832	0.2252	139.85	30.33	0.857	0.2358	211.06
	PromptIR [54]	30.49	0.856	0.2501	149.40	35.1	0.936	0.1383	94.66	30.35	0.832	0.2533	145.45	30.04	0.866	0.2137	135.30	29.61	0.866	0.2352	207.33
	RealESRGAN [68]	28.95	0.816	0.1697	100.49	33.66	0.913	0.0928	89.85	28.39	0.784	0.1890	117.83	28.16	0.824	0.1568	94.41	29.23	0.844	0.1735	173.90
	DiffBIR [42]	27.96	0.779	0.1472	91.40	30.01	0.875	0.1021	66.42	28.30	0.777	0.1705	103.80	27.73	0.778	0.1318	91.27	29.05	0.807	0.1407	136.89
	PASD [79]	27.88	0.77	0.1511	96.90	29.71	0.844	0.0940	63.11	27.68	0.810	0.1638	101.99	26.22	0.723	0.1218	97.58	27.70	0.707	0.1379	138.46
	StableSR [66]	28.88	0.805	0.1206	75.13	32.11	0.872	0.1074	69.10	29.00	0.780	0.1528	97.23	28.34	0.820	0.1235	84.54	29.86	0.839	0.1487	142.77
	SUPIR [82]	27.36	0.671	0.1418	95.55	30.14	0.837	0.1073	69.50	26.17	0.757	0.1791	100.12	26.84	0.716	0.1530	101.02	27.72	0.711	0.1389	141.48
	Ours	29.28	0.815	0.1098	70.64	32.82	0.895	0.0781	61.70	29.13	0.783	0.1300	80.32	28.57	0.825	0.1019	68.87	30.27	0.845	0.1255	136.27
C_{PSNR} [9]	AirNet [33]	30.33	0.847	0.2386	151.30	35.15	0.940	0.1502	165.52	30.35	0.825	0.2678	156.72	29.92	0.859	0.2110	145.42	28.83	0.863	0.2421	188.50
	HAT [8]	30.40	0.850	0.2070	153.25	33.33	0.898	0.1520	167.65	30.43	0.827	0.2460	155.76	29.79	0.861	0.2095	135.84	31.38	0.880	0.2384	178.90
	PromptIR [54]	30.84	0.865	0.2341	141.38	35.87	0.945	0.1461	153.46	30.91	0.838	0.2478	145.03	30.54	0.879	0.1857	129.72	31.49	0.880	0.2067	175.59
	RealESRGAN [68]	30.07	0.840	0.1467	85.38	33.76	0.917	0.1096	80.31	29.62	0.813	0.2235	142.41	29.12	0.845	0.1443	81.47	29.79	0.859	0.1580	151.03
	DiffBIR [42]	28.42	0.812	0.1064	85.38	30.13	0.841	0.1307	79.13	28.06	0.795	0.1534	103.09	28.06	0.804	0.1030	80.08	29.75	0.835	0.1096	109.86
	PASD [79]	28.34	0.81	0.1101	89.69	29.24	0.810	0.1269	66.64	26.78	0.782	0.1549	99.44	27.43	0.786	0.0933	82.53	28.56	0.785	0.1802	117.99
	StableSR [66]	29.85	0.833	0.1082	73.86	31.29	0.847	0.1199	70.77	29.64	0.806	0.1595	103.12	28.76	0.835	0.1003	68.97	30.40	0.853	0.1280	114.55
	SUPIR [82]	27.48	0.721	0.1074	90.06	29.71	0.803	0.1410	71.20	27.95	0.803	0.1719	108.30	27.00	0.779	0.1046	84.73	28.58	0.796	0.1252	120.47
	Ours	30.18	0.837	0.0996	72.23	31.75	0.866	0.1029	64.18	30.12	0.821	0.1623	101.35	29.46	0.848	0.0865	58.97	31.25	0.864	0.1041	109.05
C_{SSIM} [9]	AirNet [33]	27.54	0.816	0.3325	171.53	33.99	0.947	0.1490	131.64	28.32	0.806	0.3270	173.52	27.09	0.822	0.3183	165.87	26.64	0.824	0.3268	197.18
	HAT [8]	27.63	0.819	0.2780	167.27	32.79	0.911	0.1428	120.24	28.35	0.809	0.2864	161.61	26.88	0.813	0.3180	161.40	28.88	0.854	0.2651	193.93
	PromptIR [54]	27.95	0.826	0.3284	160.14	34.26	0.948	0.138	123.52	28.49	0.817	0.2939	158.59	27.23	0.822	0.2904	156.75	28.97	0.843	0.2543	185.23
	RealESRGAN [68]	27.08	0.802	0.1986	108.08	32.72	0.92	0.1009	61.25	27.47	0.791	0.2765	150.52	26.20	0.786	0.2131	116.11	27.38	0.815	0.1996	169.79
	DiffBIR [42]	26.50	0.766	0.1364	82.54	29.62	0.841	0.1354	69.74	26.67	0.757	0.1517	95.84	26.17	0.748	0.1351	89.88	28.44	0.806	0.1352	128.31
	PASD [79]	26.45	0.762	0.1377	86.59	28.64	0.819	0.1273	68.14	26.72	0.744	0.1483	95.09	25.56	0.725	0.1372	87.22	27.48	0.706	0.1758	127.86
	StableSR [66]	27.00	0.789	0.1535	90.74	30.53	0.844	0.1271	71.17	27.81	0.785	0.1897	109.78	26.31	0.786	0.1506	96.62	28.35	0.824	0.1509	129.93
	SUPIR [82]	25.78	0.661	0.1358	87.00	28.57	0.813	0.1414	70.98	25.54	0.695	0.1587	92.55	25.00	0.719	0.1277	90.01	27.50	0.709	0.1842	133.68
	Ours	26.85	0.781	0.1352	80.23	31.03	0.87	0.1154	66.02	28.09	0.798	0.1573	101.87	25.97	0.773	0.1294	84.58	28.20	0.819	0.1322	122.25
HIFIC [49]	AirNet [33]	29.22	0.853	0.1258	78.78	32.35	0.897	0.1061	60.69	27.63	0.841	0.1339	77.67	28.93	0.864	0.1227	82.71	29.04	0.880	0.1956	135.95
	HAT [8]	29.32	0.857	0.1254	79.54	32.48	0.909	0.1009	62.95	27.70	0.841	0.1368	75.22	28.10	0.857	0.1220	76.08	30.35	0.888	0.1723	126.98
	PromptIR [54]	29.89	0.876	0.1206	82.81	32.69	0.918	0.0958	64.17	28.12	0.858	0.1325	79.23	29.06	0.879	0.1153	79.58	30.14	0.895	0.1577	122.82
	RealESRGAN [68]	28.86	0.849	0.0766	53.47	30.86	0.918	0.084	47.83	26.76	0.818	0.1040	70.85	27.62	0.839	0.0982	66.55	29.40	0.875	0.1014	104.11
	DiffBIR [42]	28.03	0.805	0.0595	45.14	28.72	0.849	0.0777	41.64	26.46	0.795	0.0920	69.37	27.13	0.796	0.0781	59.07	29.15	0.838	0.0776	79.42
	PASD [79]	28.00	0.801	0.0652	50.70	28.40	0.818	0.0774	40.57	27.52	0.787	0.0846	65.70	27.59	0.734	0.0690	58.63	28.90	0.762	0.1297	78.86
	StableSR [66]	28.87	0.841	0.0706	46.69	29.78	0.877	0.0792	39.27	26.96	0.807	0.0905	62.49	27.49	0.837	0.0705	53.48	29.46	0.867	0.0876	85.59
	SUPIR [82]	27.81	0.748	0.0735	48.40	28.43	0.812	0.0799	44.23	27.37	0.756	0.1021	63.67	27.17	0.733	0.0834	62.03	28.93	0.769	0.0936	83.47
	Ours	29.18	0.847	0.0593	43.58	30.24	0.88	0.0767	35.89	27.54	0.824	0.0753	53.70	28.08	0.848	0.0628	48.57	30.15	0.875	0.0685	74.12

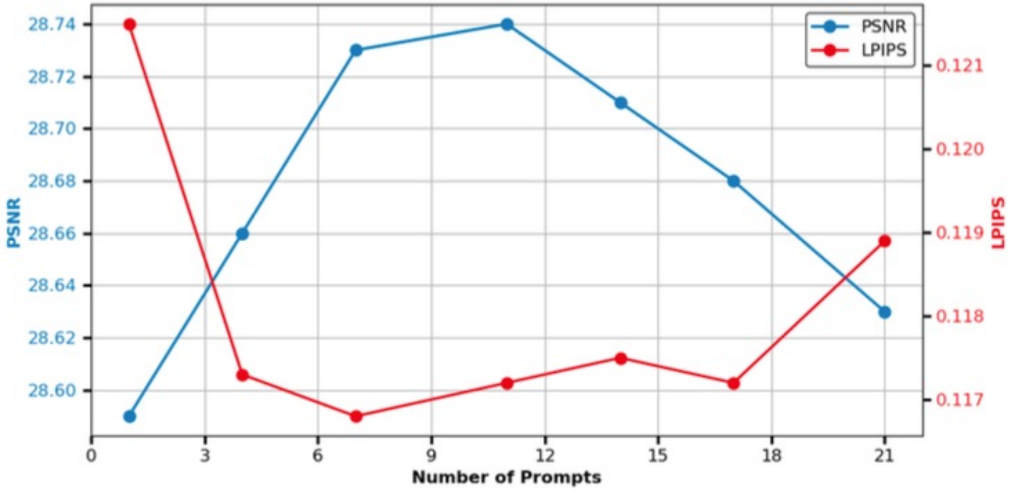
Results


HEVC (Q=47)
C_{PSNR} ('Low')
VVC (Q=47)
JPEG (Q=10)
C_{SSIM} ('Low')
LQ
PromptIR
RealESRGAN
StableSR
DiffBIR
Ours
GT

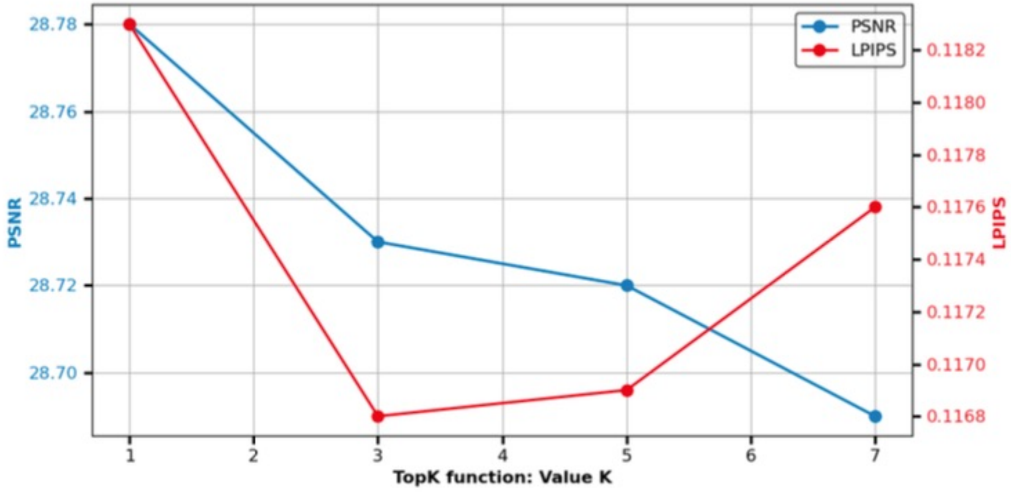
Ablation study

Methods	Seen tasks						Unseen Tasks			
	LIVE1		BSDS500		DIV2K		LIVE1 (Cross Degrees)		ICB (Cross Types)	
	PSNR/SSIM	LPIPS/FID	PSNR/SSIM	LPIPS/FID	PSNR/SSIM	LPIPS/FID	PSNR/SSIM	LPIPS/FID	PSNR/SSIM	LPIPS/FID
No Prompt	28.73/0.806	0.1343/85.87	28.56/0.770	0.1591/96.45	27.86/0.813	0.1295/79.5	31.79/0.893	0.064/37.33	28.61/0.787	0.1933/210.84
Single Prompt	28.86/0.806	0.1272/79.45	28.78/0.791	0.1530/89.62	28.02/0.816	0.1143/71.26	33.25/0.910	0.0457/28.60	28.88/0.793	0.179/187.29
Multiple Prompt	28.98/0.810	0.1212/77.09	28.93/0.794	0.1482/89.34	28.22/0.817	0.1124/71.65	33.32/0.913	0.0432/28.23	28.89/0.792	0.1756/187.89
MoE-Prompt (Ours)	**29.02/0.811**	**0.1179/75.86**	**28.97/0.794**	**0.1430/88.14**	**28.29/0.821**	**0.1071/68.91**	**33.45/0.916**	**0.0411/25.65**	**29.02/0.800**	**0.1690/176.87**

Methods	Datasets					
	LIVE1		BSDS500		ICB	
	PSNR/SSIM	LPIPS/FID	PSNR/SSIM	LPIPS/FID	PSNR/SSIM	LPIPS/FID
MoE-Prompt	29.02/0.810	0.1179/75.86	28.97/0.794	0.1430/88.14	29.83/0.839	0.1277/122.59
MoE-Prompt+V2T Adapter	29.03/0.812	0.1145/74.13	28.94/0.796	0.1367/**86.77**	29.83/0.840	0.1239/119.78
MoE-Prompt+DP	29.07/0.814	0.1154/76.60	**29.06**/0.795	0.1405/88.00	29.87/0.841	0.1269/122.32
MoE-Prompt+V2T Adapter+DP	**29.10/0.814**	**0.1136/73.60**	29.02/**0.797**	**0.1356**/86.81	**29.88/0.841**	**0.1235/119.29**



(a) Effects of Number of Prompts



(b) Effects of K value

Results


HEVC (Q=47)
VVC (Q=47)
C_{SSIM} (Q=1)
LQ
Single Prompt
Multi Prompts
MOE Promtps
MOE+DP
MOE+Visual

Outline

- Introduction
- Framework
- Method
- Experiment
- Conclusion

Conclusion

- Propose the **MoE-Prompt** to excavate task-customized diffusion priors, maximizing the utilization of different prompts, enabling them to collaboratively perceive different distortions.

- By utilizing a **Visual2Text** adapter, we integrate visual information into the **text inputs of the Stable Diffusion** model, thereby improving the perceptual restoration capabilities of the model at low bitrates.

- Our extensive experiments have demonstrated that MoE-DiffIR not only improves perceptual performance at low bitrates but also facilitates rapid transferability across various compression tasks.