Supplementary Material for

Fine-grained Image Quality Assessment for Perceptual Image Restoration

This supplementary document provides comprehensive experimental evidence, detailed analyses, and extensive visual demonstrations that substantiate the findings presented in our main paper. The material herein offers deeper insights into the fine-grained challenges of image restoration quality assessment and validates the effectiveness of our proposed approach across diverse scenarios.

This document encompasses the following components:

• Detailed Ablation Study Results: Systematic investigation of our degradation-aware feature learning (DFL) module's contribution, demonstrating consistent improvements across multiple evaluation metrics (SRCC, PLCC, ACC) and providing empirical justification for our architectural design choices.

• Comprehensive Performance Analysis Across Quality Ranges: Exhaustive evaluation of state-of-the-art methods across 13 datasets, revealing consistent performance degradation of existing approaches within narrow MOS score intervals and highlighting the fundamental challenges in fine-grained quality assessment.

• Exemplars of Fine-Grained Quality Differences: Carefully curated image pairs demonstrating subtle quality variations across six restoration tasks - deblurring, denoising, deraining, super-resolution, dehazing, and mixture restoration. These examples illustrate the perceptual challenges that make fine-grained assessment particularly demanding for image quality assessment.

• Thorough Qualitative Analysis: Side-by-side visual comparisons showcasing how different quality assessment methods respond to restoration results with nuanced quality differences. These comparisons reveal the limitations of traditional metrics and the inconsistent behavior of deep learning-based approaches in fine-grained evaluation scenarios, while demonstrating our proposed FGResQ method in accurately identifying quality preferences that align with human perceptual judgments.

All source code and implementation details are provided in the supplementary material.

Ablation Study

As shown in Table 3 in main paper, , unlike task-specific methods that excel in particular scenarios, FGResQ maintains robust performance across all six IR tasks, suggesting that incorporating degradation knowledge enables more generalizable quality assessment capabilities.To further validate the effectiveness of our degradation-aware feature learning (DFL) module, we conduct ablation experiments. Table S1 presents the average performance across all IR tasks, comparing our full FGResQ model with a variant that removes the degradation-aware learning component (w/o DFL).

Table S1: Ablation study.

Method	SRCC	PLCC	ACC
w/o DFL	<u>0.698</u>	<u>0.711</u>	<u>0.743</u>
FGResQ	**0.703**	**0.717**	**0.752**

Comprehensive Performance Analysis Across Quality Ranges

Following our methodology in the main paper, we partition quality score ranges into discrete intervals, and systematically evaluate existing IQA methods across 13 benchmark datasets. This comprehensive analysis encompasses both full-reference (FR) methods including PSNR, SSIM, LPIPS, and DISTS, as well as no-reference (NR) methods ranging from traditional approaches (NIQE, IL-NIQE, BRISQUE) to state-of-the-art learning-based methods (DBCNN, HyperIQA, MetaIQA, LIQE, CLIP-IQA, Q-Align, DeQA-Score). Tables 2 through 14 present comprehensive results across two categories of datasets. General IQA datasets include TID2013, CSIQ, LIVE, KADID-10k and VCL@FER for synthetic distortions. Image restoration quality assessment datasets encompass PIPAL for super-resolution, denoise and mixture restoration, MDD13 for deblurring evaluation, exBeDDE and IVCDehazing for dehazing assessment, IVIPC-DQA for deraining quality evaluation, and super-resolution datasets QADS, RealSRQ, and SISRSet covering different scenarios. The results consistently demonstrate the same pattern observed in our main analysis: while most methods achieve reasonable overall correlation coefficients when evaluated on complete datasets, their performance dramatically deteriorates within narrow quality ranges. This reveals the inability of traditional methods to effectively conduct fine-grained image quality assessment.

Fine-grained Image Pairs Visualization

This section presents visual examples demonstrating the perceptual challenges in fine-grained quality assessment across six restoration tasks. These examples represent the subtle quality variations commonly encountered in practical image restoration applications, where algorithmic comparisons and parameter optimization require sensitive discrimination between restoration results with marginal quality differences.

Figures S1 through S6 showcase image pairs with subtle quality differences in deblurring, denoising, deraining, super-resolution, dehazing, and mixture restoration tasks. The image pairs demonstrate cases where slight variations in restoration effectiveness can significantly impact perceptual quality despite producing similar overall visual appearance.

Table S2: Performance analysis across different MOS score ranges on PIPAL dataset. Methods are evaluated on subsets with MOS scores in specific ranges.

Type	Method	[0.0,0.2)		[0.2,0.4)		[0.4,0.6)		[0.6,0.8)		[0.8,1.0]		[0.0,1.0]	
		SRCC	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**
FR	PSNR	0.323	0.266	0.082	0.107	0.209	0.224	0.161	0.135	0.072	0.021	0.422	0.420
	SSIM	0.293	0.243	0.108	0.101	0.258	0.273	0.254	0.195	0.049	0.076	0.530	0.516
	LPIPS	-0.034	0.025	0.077	0.080	0.325	0.323	0.287	0.267	0.124	0.096	0.612	0.595
	DISTS	0.168	0.134	0.159	0.165	0.310	0.304	0.242	0.227	0.165	0.128	0.585	0.584
NR	NIQE	-0.126	-0.101	-0.002	0.036	0.107	0.114	0.001	-0.096	0.080	0.038	0.153	0.018
	IL-NIQE	-0.235	-0.144	-0.098	-0.093	0.126	0.128	0.128	0.141	0.054	0.053	0.289	0.274
	BRISQUE	-0.142	-0.105	0.025	0.014	0.125	0.100	0.035	-0.048	0.131	0.096	0.185	0.078
	DB-CNN	-0.157	-0.072	0.321	0.316	0.353	0.362	0.330	0.354	-0.016	0.060	0.636	0.669
	HyperIQA	0.100	0.174	0.274	0.259	0.314	0.307	0.292	0.298	0.032	-0.045	0.584	0.611
	MetaIQA	0.037	0.074	0.160	0.163	0.204	0.196	0.174	0.178	-0.101	-0.060	0.423	0.432
	LIQE	-0.232	-0.035	0.053	0.059	0.175	0.084	0.299	0.240	0.107	0.182	0.479	0.318
	CLIP-IQA	-0.152	-0.130	0.211	0.190	0.238	0.236	0.293	0.299	0.071	0.049	0.530	0.542
	Q-Align	0.230	0.271	0.301	0.333	0.337	0.324	0.213	0.220	0.178	0.154	0.418	0.410
	DeQA-Score	0.568	0.652	0.676	0.672	0.623	0.636	0.516	0.562	0.350	0.339	0.747	0.777

Table S3: Performance comparison across different MOS ranges on TID2013 dataset. FR and NR represent full-reference and no-reference methods, respectively.

Type	Method	[0.0,0.2)		[0.2,0.4)		[0.4,0.6)		[0.6,0.8)		[0.8,1.0]		[0.0,1.0]	
		SRCC	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**
FR	PSNR	0.064	-0.275	0.164	0.213	0.303	0.330	0.329	0.327	-0.033	-0.126	0.670	0.651
	SSIM	-0.018	-0.197	0.192	0.262	0.194	0.178	0.380	0.371	0.175	0.113	0.596	0.633
	LPIPS	-0.318	-0.010	0.308	0.371	0.269	0.252	0.407	0.356	-0.090	-0.045	0.724	0.731
	DISTS	-0.136	0.052	0.345	0.419	0.268	0.268	0.343	0.350	0.026	0.060	0.688	0.736
NR	NIQE	-0.191	-0.119	0.221	0.176	0.054	0.076	0.100	0.151	0.014	-0.003	0.313	0.360
	IL-NIQE	-0.027	0.004	0.236	0.272	0.158	0.157	0.305	0.288	-0.141	-0.068	0.520	0.597
	BRISQUE	-0.227	-0.246	0.267	0.210	0.091	0.106	0.195	0.196	0.136	0.160	0.370	0.473
	DB-CNN	0.636	0.654	0.621	0.591	0.649	0.636	0.496	0.507	0.293	0.303	0.889	0.905
	HyperIQA	0.391	0.453	0.587	0.605	0.391	0.347	0.339	0.333	0.000	0.027	0.810	0.839
	MetaIQA	0.400	0.502	0.595	0.618	0.563	0.546	0.560	0.526	0.529	0.631	0.885	0.875
	LIQE	-0.018	-0.140	0.484	0.321	0.468	0.335	0.513	0.508	0.322	0.375	0.869	0.797
	CLIP-IQA	0.582	0.567	0.438	0.446	0.467	0.495	0.403	0.413	0.087	0.040	0.813	0.835
	Q-Align	0.346	0.464	0.489	0.497	0.510	0.451	0.469	0.472	0.303	0.349	0.867	0.870
	DeQA-Score	0.382	0.448	0.566	0.560	0.572	0.566	0.514	0.525	0.467	0.429	0.891	0.898

Table S4: Performance comparison across different MOS ranges on CSIQ dataset. FR and NR represent full-reference and no-reference methods, respectively.

Type	Method	[0.0,0.2)		[0.2,0.4)		[0.4,0.6)		[0.6,0.8)		[0.8,1.0]		[0.0,1.0]	
		SRCC	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**
FR	PSNR	0.006	0.236	0.536	0.564	0.010	-0.003	0.305	0.411	0.534	0.620	0.808	0.801
	SSIM	0.103	-0.035	0.409	0.455	-0.040	0.012	0.378	0.355	0.646	0.475	0.850	0.793
	LPIPS	0.333	0.250	0.441	0.461	0.343	0.331	0.620	0.582	0.657	0.568	0.925	0.903
	DISTS	0.055	0.037	0.664	0.661	0.288	0.334	0.508	0.509	0.673	0.593	<u>0.933</u>	<u>0.929</u>
NR	NIQE	0.661	0.501	0.424	0.351	-0.320	-0.161	0.316	0.346	0.296	0.260	0.619	0.700
	IL-NIQE	0.346	0.086	0.446	0.395	0.367	0.318	0.524	0.428	0.216	0.193	0.809	0.787
	BRISQUE	0.200	0.215	0.385	0.278	0.076	0.064	0.302	0.301	0.146	0.172	0.559	0.688
	DB-CNN	0.624	0.605	0.696	0.657	0.693	0.638	0.526	0.550	0.448	0.483	0.929	0.942
	HyperIQA	0.730	0.629	0.472	0.408	0.340	0.419	0.566	0.575	0.316	0.315	0.858	0.891
	MetaIQA	0.640	0.605	0.624	0.575	0.552	0.493	0.477	0.417	0.469	0.450	0.888	0.882
	LIQE	0.566	0.476	0.579	0.433	0.469	0.489	0.686	0.687	0.675	0.664	0.947	0.888
	CLIP-IQA	0.697	0.591	0.306	0.372	0.165	0.203	0.488	0.498	0.352	0.367	0.842	0.868
	Q-Align	0.649	0.549	0.576	0.560	0.409	0.444	0.637	0.607	0.341	0.370	0.877	0.904
	DeQA-Score	**0.830**	0.586	0.521	0.491	0.528	0.569	0.546	0.560	0.552	0.551	**0.938**	**0.952**

Table S5: Performance comparison across different MOS ranges on LIVE dataset. FR and NR represent full-reference and no-reference methods, respectively.

Type	Method	[0.0,0.2)		[0.2,0.4)		[0.4,0.6)		[0.6,0.8)		[0.8,1.0]		[0.0,1.0]	
		SRCC	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**
FR	PSNR	0.687	0.627	0.296	0.233	0.525	0.494	0.618	0.628	0.882	0.922	0.952	0.806
	SSIM	0.685	0.654	0.298	0.326	0.624	0.637	0.706	0.682	0.894	0.858	0.971	0.885
	LPIPS	0.710	0.655	0.263	0.225	0.625	0.641	0.669	0.639	0.891	0.847	0.960	0.845
	DISTS	0.710	0.635	0.286	0.215	0.728	0.702	0.703	0.719	0.880	0.879	0.979	0.936
NR	NIQE	0.591	0.610	0.151	0.100	0.442	0.449	0.677	0.581	0.474	0.535	0.931	0.645
	IL-NIQE	0.677	0.612	0.220	0.113	0.610	0.619	0.501	0.540	0.297	0.196	0.921	0.690
	BRISQUE	0.576	0.546	0.506	0.543	0.765	0.768	0.864	0.808	0.541	0.515	0.971	0.976
	DB-CNN	0.476	0.486	0.486	0.615	0.797	0.770	0.816	0.780	0.666	0.676	0.977	0.982
	HyperIQA	0.676	0.694	0.605	0.715	0.813	0.803	0.856	0.834	0.625	0.655	0.977	0.986
	MetaIQA	0.786	0.604	0.581	0.589	0.494	0.462	0.284	0.157	0.266	0.131	0.874	0.837
	LIQE	-0.185	-0.169	0.461	0.397	0.791	0.769	0.693	0.691	0.365	0.287	0.959	0.943
	CLIP-IQA	0.428	0.410	0.452	0.518	0.629	0.615	0.736	0.700	0.500	0.541	0.961	0.968
	Q-Align	0.889	0.994	0.907	0.953	0.803	0.716	0.891	0.926	0.554	0.972	0.923	0.928
	DeQA-Score	0.845	0.987	0.895	0.922	**0.773**	**0.842**	**0.895**	**0.933**	**0.877**	**0.991**	**0.974**	**0.976**

Table S6: Performance analysis across different MOS score ranges on KADID-10K dataset. Methods are evaluated on subsets with MOS scores in specific ranges.

Type	Method	[0.0,0.2)		[0.2,0.4)		[0.4,0.6)		[0.6,0.8)		[0.8,1.0]		[0.0,1.0]	
		SRCC	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**
FR	PSNR	0.409	0.370	0.212	0.219	0.147	0.126	0.163	0.168	0.162	0.153	0.675	0.554
	SSIM	0.442	0.371	0.210	0.176	0.137	0.176	0.180	0.152	0.288	0.293	0.609	0.581
	LPIPS	0.416	0.323	0.355	0.298	0.224	0.231	0.276	0.233	0.412	0.399	0.825	0.752
	DISTS	0.365	0.315	0.312	0.285	0.244	0.241	0.311	0.301	0.473	0.458	0.807	0.799
NR	NIQE	0.036	0.043	0.091	0.076	0.043	0.058	0.061	0.051	0.366	0.337	0.386	0.410
	IL-NIQE	-0.059	-0.048	0.095	0.111	0.057	0.089	0.160	0.192	0.290	0.302	0.526	0.543
	BRISQUE	0.082	0.039	0.045	0.020	0.048	0.037	0.041	0.047	0.256	0.272	0.323	0.387
	DB-CNN	0.004	0.009	0.255	0.248	0.311	0.298	0.382	0.386	0.337	0.327	0.822	0.809
	HyperIQA	0.047	0.022	0.275	0.256	0.310	0.306	0.373	0.362	0.519	0.460	0.803	0.806
	MetaIQA	0.023	0.006	0.159	0.178	0.228	0.208	0.367	0.362	0.487	0.474	0.793	0.782
	LIQE	0.194	0.145	0.318	0.284	0.343	0.335	0.359	0.270	0.285	0.195	0.798	0.758
	CLIP-IQA	0.125	0.110	0.228	0.235	0.217	0.215	0.313	0.317	0.374	0.346	0.790	0.790
	Q-Align	0.436	0.426	0.488	0.482	0.534	0.502	0.534	0.436	0.604	0.445	0.958	0.962
	DeQA-Score	0.468	0.532	0.638	0.622	0.673	0.742	0.689	0.762	0.726	0.753	0.926	0.923

Table S7: Performance analysis across different MOS score ranges on VCL@FER. Methods are evaluated on subsets with MOS scores in specific ranges.

Type	Method	[0.0,0.2)		[0.2,0.4)		[0.4,0.6)		[0.6,0.8)		[0.8,1.0]		[0.0,1.0]	
		SRCC	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**
FR	PSNR	0.114	0.111	0.159	0.178	0.316	0.287	0.422	0.429	0.413	0.488	0.846	0.857
	SSIM	0.414	0.480	0.483	0.484	0.398	0.432	0.488	0.381	0.063	0.242	0.919	0.812
	LPIPS	0.569	0.606	0.575	0.569	0.449	0.423	0.488	0.358	0.301	0.289	0.896	0.808
	DISTS	0.546	0.650	0.630	0.650	0.532	0.562	0.545	0.530	0.462	0.329	0.942	0.913
NR	NIQE	0.185	0.089	0.461	0.414	0.413	0.373	0.515	0.439	-0.119	0.174	0.835	0.549
	IL-NIQE	0.072	0.100	0.516	0.414	0.306	0.257	0.403	0.385	0.301	0.110	0.796	0.761
	BRISQUE	0.274	0.325	0.525	0.409	0.607	0.600	0.448	0.451	0.601	0.571	0.917	0.904
	DB-CNN	0.654	0.628	0.733	0.689	0.577	0.627	0.695	0.713	0.273	0.239	0.960	0.959
	HyperIQA	0.725	0.711	0.779	0.740	0.558	0.608	0.647	0.612	-0.112	-0.134	0.946	0.945
	MetaIQA	0.336	0.384	0.277	0.293	0.233	0.110	0.216	0.169	0.123	-0.115	0.687	0.635
	LIQE	0.728	0.400	0.745	0.742	0.630	0.677	0.654	0.612	0.406	0.493	0.963	0.954
	CLIP-IQA	0.345	0.291	0.547	0.542	0.356	0.372	0.708	0.696	0.371	0.389	0.952	0.958
	Q-Align	0.718	0.275	0.576	0.629	0.435	0.380	0.633	0.692	0.594	0.607	0.890	0.858
	DeQA-Score	0.631	0.631	0.720	0.782	0.678	0.717	0.812	0.809	0.559	0.661	0.957	0.956

Table S8: Performance analysis across different MOS score ranges on exBeDDE. Methods are evaluated on subsets with MOS scores in specific ranges.

Type	Method	[0.0,0.2)		[0.2,0.4)		[0.4,0.6)		[0.6,0.8)		[0.8,1.0]		[0.0,1.0]	
		SRCC	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**
FR	PSNR	-0.235	-0.234	0.077	0.049	-0.290	-0.225	-0.304	-0.322	-0.233	-0.309	-0.705	-0.701
	SSIM	-0.176	-0.177	0.002	0.002	-0.027	0.017	-0.311	-0.320	-0.455	-0.491	-0.730	-0.720
	LPIPS	-0.075	-0.055	-0.117	-0.189	-0.421	-0.450	-0.245	-0.246	-0.321	-0.403	-0.747	-0.752
	DISTS	0.001	-0.060	-0.190	-0.236	-0.322	-0.285	-0.404	-0.341	-0.263	-0.227	-0.783	-0.755
NR	NIQE	0.123	0.091	-0.082	-0.065	0.198	0.133	-0.128	-0.145	-0.050	-0.055	0.029	0.031
	IL-NIQE	0.108	0.082	-0.036	-0.039	0.032	0.061	0.014	-0.032	-0.402	-0.335	0.107	0.074
	BRISQUE	-0.086	-0.091	0.116	0.140	0.180	0.195	-0.223	-0.237	0.241	0.220	0.118	0.068
	DB-CNN	0.155	0.197	0.344	0.371	0.200	0.264	0.396	0.384	0.731	0.721	0.899	0.905
	HyperIQA	0.125	0.191	0.324	0.298	0.226	0.206	0.159	0.143	0.594	0.610	0.847	0.863
	MetaIQA	0.135	0.239	0.297	0.298	0.262	0.263	0.223	0.221	0.662	0.617	0.860	0.881
	LIQE	0.017	0.143	0.433	0.407	0.383	0.356	0.385	0.376	0.734	0.492	0.895	0.893
	CLIP-IQA	0.168	0.179	0.397	0.406	0.328	0.342	0.215	0.217	0.524	0.491	0.879	0.903
	Q-Align	0.013	0.044	0.153	0.170	0.218	0.260	0.286	0.289	0.472	0.432	0.814	0.815
	DeQA-Score	0.105	0.142	0.201	0.186	0.172	0.275	0.258	0.265	0.707	0.727	0.863	0.879

Table S9: Performance analysis across different MOS score ranges on IVCDehazing. Methods are evaluated on subsets with MOS scores in specific ranges.

Type	Method	[0.0,0.2)		[0.2,0.4)		[0.4,0.6)		[0.6,0.8)		[0.8,1.0]		[0.0,1.0]	
		SRCC	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**
FR	PSNR	-	-	-	-	-	-	-	-	-	-	-	-
	SSIM	-	-	-	-	-	-	-	-	-	-	-	-
	LPIPS	-	-	-	-	-	-	-	-	-	-	-	-
	DISTS	-	-	-	-	-	-	-	-	-	-	-	-
NR	NIQE	-1.000	-1.000	-0.113	-0.056	0.127	0.059	0.193	0.150	0.600	0.920	0.357	0.455
	IL-NIQE	1.000	1.000	-0.042	-0.162	0.582	0.440	0.132	0.105	-0.200	-0.576	0.381	0.509
	BRISQUE	1.000	1.000	0.459	0.407	-0.291	-0.022	0.062	-0.003	0.200	0.582	0.122	0.124
	DB-CNN	1.000	1.000	-0.127	-0.175	0.255	0.268	0.324	0.397	-0.200	0.310	0.753	0.654
	HyperIQA	1.000	1.000	-0.177	-0.101	0.800	0.715	0.130	0.002	-0.400	-0.667	0.404	0.592
	MetaIQA	1.000	1.000	0.127	0.150	-0.346	-0.422	0.171	0.270	0.000	0.085	0.683	0.618
	LIQE	1.000	1.000	-0.438	-0.513	0.255	0.455	0.443	0.380	0.200	0.377	0.630	0.540
	CLIP-IQA	1.000	1.000	0.223	0.132	-0.255	-0.332	0.107	0.179	0.000	0.154	0.530	0.559
	Q-Align	1.000	1.000	0.329	0.278	0.143	0.072	0.600	0.634	-1.000	-0.978	0.328	0.303
	DeQA-Score	1.000	1.000	0.410	0.532	0.643	0.778	-0.150	-0.323	-0.400	-0.460	0.103	0.177

Table S10: Performance analysis across different MOS score ranges on IVIPC-DQA dataset. Methods are evaluated on subsets with MOS scores in specific ranges.

Type	Method	[0.0,0.2)		[0.2,0.4)		[0.4,0.6)		[0.6,0.8)		[0.8,1.0]		[0.0,1.0]	
		SRCC	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**
FR	PSNR	-	-	-	-	-	-	-	-	-	-	-	-
	SSIM	-	-	-	-	-	-	-	-	-	-	-	-
	LPIPS	-	-	-	-	-	-	-	-	-	-	-	-
	DISTS	-	-	-	-	-	-	-	-	-	-	-	-
NR	NIQE	0.255	0.240	0.077	0.073	0.121	0.077	-	-	-	-	0.032	0.008
	IL-NIQE	-0.094	-0.073	0.038	0.045	0.184	0.126	-	-	-	-	0.002	0.033
	BRISQUE	-0.003	-0.115	-0.018	-0.028	0.113	0.172	-	-	-	-	0.090	0.094
	DB-CNN	0.164	0.165	0.338	0.369	0.238	0.264	-	-	-	-	0.491	0.525
	HyperIQA	0.011	0.038	0.369	0.417	0.050	0.111	-	-	-	-	0.475	0.503
	MetaIQA	-0.022	-0.040	0.393	0.353	0.319	0.364	-	-	-	-	0.550	0.546
	LIQE	-0.098	0.044	0.329	0.332	0.136	0.152	-	-	-	-	0.483	0.479
	CLIP-IQA	-0.043	-0.041	0.159	0.189	0.215	0.240	-	-	-	-	0.400	0.450
	Q-Align	0.015	-0.105	0.272	0.349	0.245	0.324	-	-	-	-	0.255	0.302
	DeQA-Score	0.571	0.544	0.512	0.602	0.506	0.502	-	-	-	-	0.533	0.581

Table S11: Performance analysis across different MOS score ranges on MDD13 dataset. Methods are evaluated on subsets with MOS scores in specific ranges.

Type	Method	[0.0,0.2)		[0.2,0.4)		[0.4,0.6)		[0.6,0.8)		[0.8,1.0]		[0.0,1.0]	
		SRCC	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**
FR	PSNR	1.000	0.656	0.182	0.154	0.321	0.317	0.048	0.066	0.014	-0.068	0.186	0.168
	SSIM	1.000	0.657	0.015	-0.008	0.261	0.294	0.289	0.167	0.125	0.071	0.445	0.353
	LPIPS	0.800	0.799	0.232	0.093	0.252	0.091	0.340	0.326	0.472	0.451	0.777	0.702
	DISTS	1.000	0.922	0.400	0.336	0.295	0.292	0.600	0.591	0.790	0.820	0.907	0.901
NR	NIQE	-0.200	-0.702	0.296	0.302	0.128	0.097	0.253	0.287	0.372	0.475	0.386	0.413
	IL-NIQE	0.000	-0.287	0.331	0.127	0.176	0.150	0.258	0.287	0.394	0.382	0.591	0.515
	BRISQUE	0.200	0.029	-0.130	-0.238	0.091	0.075	0.304	0.304	0.206	0.227	0.356	0.356
	DB-CNN	0.800	0.698	0.540	0.671	0.437	0.447	0.538	0.551	0.476	0.515	0.868	0.888
	HyperIQA	0.000	0.318	0.577	0.603	0.370	0.388	0.551	0.551	0.377	0.387	0.862	0.879
	MetaIQA	1.000	0.843	0.455	0.445	0.310	0.279	0.462	0.391	0.282	0.218	0.806	0.826
	LIQE	0.949	0.656	0.536	0.559	0.497	0.442	0.594	0.582	0.370	0.400	0.899	0.814
	CLIP-IQA	0.316	0.649	0.449	0.610	0.499	0.524	0.604	0.628	0.556	0.622	0.908	0.910
	Q-Align	-0.400	-0.086	0.380	0.488	0.296	0.301	0.589	0.588	0.672	0.663	0.884	0.880
	DeQA-Score	0.800	0.916	0.432	0.608	0.429	0.445	0.529	0.532	0.493	0.539	0.887	0.905

Table S12: Performance analysis across different MOS score ranges on QADS dataset. Methods are evaluated on subsets with MOS scores in specific ranges.

Type	Method	[0.0,0.2)		[0.2,0.4)		[0.4,0.6)		[0.6,0.8)		[0.8,1.0]		[0.0,1.0]	
		SRCC	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**
FR	PSNR	1.000	0.656	0.182	0.154	0.321	0.317	0.048	0.066	0.014	-0.068	0.186	0.168
	SSIM	1.000	0.657	0.015	-0.008	0.261	0.294	0.289	0.167	0.125	0.071	0.445	0.353
	LPIPS	0.800	0.799	0.232	0.093	0.252	0.091	0.340	0.326	0.472	0.451	0.777	0.702
	DISTS	1.000	0.922	0.400	0.336	0.295	0.292	0.600	0.591	0.790	0.820	0.907	0.901
NR	NIQE	-0.200	-0.702	0.296	0.302	0.128	0.097	0.253	0.287	0.372	0.475	0.386	0.413
	IL-NIQE	0.000	-0.287	0.331	0.127	0.176	0.150	0.258	0.287	0.394	0.382	0.591	0.515
	BRISQUE	0.200	0.029	-0.130	-0.238	0.091	0.075	0.304	0.304	0.206	0.227	0.356	0.356
	DB-CNN	0.800	0.698	0.540	0.671	0.437	0.447	0.538	0.551	0.476	0.515	0.868	0.888
	HyperIQA	0.000	0.318	0.577	0.603	0.370	0.388	0.551	0.551	0.377	0.387	0.862	0.879
	MetaIQA	1.000	0.843	0.455	0.445	0.310	0.279	0.462	0.391	0.282	0.218	0.806	0.826
	LIQE	0.949	0.656	0.536	0.559	0.497	0.442	0.594	0.582	0.370	0.400	0.899	0.814
	CLIP-IQA	0.316	0.649	0.449	0.610	0.499	0.524	0.604	0.628	0.556	0.622	0.908	0.910
	Q-Align	-0.400	-0.086	0.380	0.488	0.296	0.301	0.589	0.588	0.672	0.663	0.884	0.880
	DeQA-Score	0.800	0.916	0.432	0.608	0.429	0.445	0.529	0.532	0.493	0.539	0.887	0.905

Table S13: Performance analysis across different MOS score ranges on RealSRQ dataset. Methods are evaluated on subsets with MOS scores in specific ranges.

Type	Method	[0.0,0.2)		[0.2,0.4)		[0.4,0.6)		[0.6,0.8)		[0.8,1.0]		[0.0,1.0]	
		SRCC	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**
FR	PSNR	1.000	1.000	-0.149	-0.151	-0.069	0.010	-	-	-	-	-0.035	0.013
	SSIM	1.000	1.000	-0.173	-0.116	0.021	0.118	-	-	-	-	-0.029	0.050
	LPIPS	1.000	1.000	-0.148	-0.106	0.271	0.202	-	-	-	-	-0.003	0.071
	DISTS	1.000	1.000	-0.147	-0.132	0.311	0.309	-	-	-	-	-0.034	0.057
NR	NIQE	-1.000	-1.000	-0.017	-0.005	0.359	0.362	-	-	-	-	0.013	0.050
	IL-NIQE	-1.000	-1.000	-0.031	-0.014	0.218	0.205	-	-	-	-	0.006	0.032
	BRISQUE	-1.000	-1.000	0.059	0.117	0.095	0.156	-	-	-	-	0.045	-0.018
	DB-CNN	-1.000	-1.000	0.610	0.630	0.786	0.801	-	-	-	-	0.739	0.838
	HyperIQA	-1.000	-1.000	0.474	0.490	0.705	0.743	-	-	-	-	0.602	0.609
	MetaIQA	1.000	1.000	0.106	-0.003	-0.061	0.029	-	-	-	-	0.171	0.134
	LIQE	-1.000	-1.000	0.412	0.347	0.561	0.421	-	-	-	-	0.537	0.525
	CLIP-IQA	-1.000	-1.000	0.395	0.439	0.636	0.660	-	-	-	-	0.555	0.659
	Q-Align	-1.000	-1.000	-0.058	0.005	0.201	0.303	-	-	-	-	0.100	0.227
	DeQA-Score	-1.000	-1.000	0.267	0.269	0.685	0.679	-	-	-	-	0.445	0.609

Table S14: Performance analysis across different MOS score ranges on SISRSet dataset. Methods are evaluated on subsets with MOS scores in specific ranges.

Type	Method	[0.0,0.2)		[0.2,0.4)		[0.4,0.6)		[0.6,0.8)		[0.8,1.0]		[0.0,1.0]	
		SRCC	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**	**SRCC**	**PLCC**
FR	PSNR	-0.314	-0.168	-0.136	0.077	0.261	0.077	0.762	0.741	-0.367	-0.231	0.589	0.613
	SSIM	-0.074	-0.072	-0.009	-0.068	0.029	-0.240	0.833	0.648	-0.346	-0.196	0.371	0.406
	LPIPS	0.279	0.165	-0.600	-0.511	-0.540	-0.592	-0.167	0.129	-0.217	-0.364	0.517	0.560
	DISTS	0.220	0.122	-0.409	-0.427	-0.503	-0.576	-0.048	0.242	-0.045	-0.415	0.665	0.653
NR	NIQE	0.288	0.260	0.446	0.525	0.753	0.809	-0.191	-0.238	0.351	0.430	0.680	0.683
	IL-NIQE	0.712	0.687	0.236	0.371	0.562	0.594	-0.071	-0.120	0.338	0.449	0.639	0.616
	BRISQUE	0.664	0.680	0.055	-0.064	0.213	0.444	-0.310	-0.194	0.512	0.458	0.888	0.876
	DB-CNN	0.664	0.655	0.791	0.675	0.852	0.804	0.476	0.040	0.772	0.683	0.915	0.902
	HyperIQA	0.801	0.810	0.636	0.642	0.617	0.623	0.262	0.053	0.452	0.491	0.959	0.946
	MetaIQA	0.444	0.385	0.091	0.180	0.050	0.435	-0.357	-0.477	0.479	0.222	0.638	0.511
	LIQE	0.191	0.266	0.364	0.100	0.631	0.596	-0.333	-0.476	0.397	0.122	0.307	0.436
	CLIP-IQA	0.606	0.653	0.164	0.225	0.125	0.043	0.071	0.032	0.436	0.371	0.945	0.941
	Q-Align	0.805	0.799	0.564	0.528	0.208	0.197	-0.119	-0.175	-0.204	-0.081	0.679	0.665
	DeQA-Score	0.553	0.528	0.427	0.360	0.356	0.336	-0.381	-0.282	0.591	0.191	0.496	0.417

These carefully selected examples illustrate the inherent complexity of fine-grained quality assessment in image restoration. The visual evidence demonstrates that distinguishing subtle quality differences requires sophisticated understanding of restoration-specific artifacts and perceptual quality factors. These examples highlight the necessity for FGResQ designed to handle the nuanced quality differences encountered in image restoration applications, where traditional evaluation paradigms may not capture the full spectrum of perceptual quality variations.

Qualitative Analysis

This section provides visual evidence of how different quality assessment methods respond to fine-grained quality differences in restored images. Figures S7 through S12 present qualitative comparisons on representative fine-grained image pairs across different restoration tasks. The results demonstrate systematic limitations of existing IQA methods in fine-grained scenarios.

Traditional metrics like PSNR and SSIM often produce nearly identical scores for images with subtle quality differences, failing to provide meaningful quality discrimination. Advanced learning-based methods such as CLIP-IQA and DeQA-Score show improved sensitivity compared to traditional metrics but still exhibit inconsistent behavior in fine-grained evaluation. These methods frequently produce incorrect rankings when quality differences become subtle, particularly in challenging scenarios like mixture restoration where multiple degradation types interact.

In contrast, our proposed FGResQ demonstrates superior performance across all tested scenarios. The method consistently identifies the superior image in fine-grained pairs where existing approaches fail, particularly excelling in challenging mixture restoration and dehazing scenarios. FGResQ's degradation-aware feature learning enables more nuanced understanding of restoration quality, while the dual-branch architecture effectively captures both coarse-grained and fine-grained quality relationships. These visual comparisons complement our quantitative analysis and provide intuitive evidence demonstrating FGResQ's effectiveness in addressing the fundamental limitations of existing approaches for fine-grained quality assessment in image restoration applications.

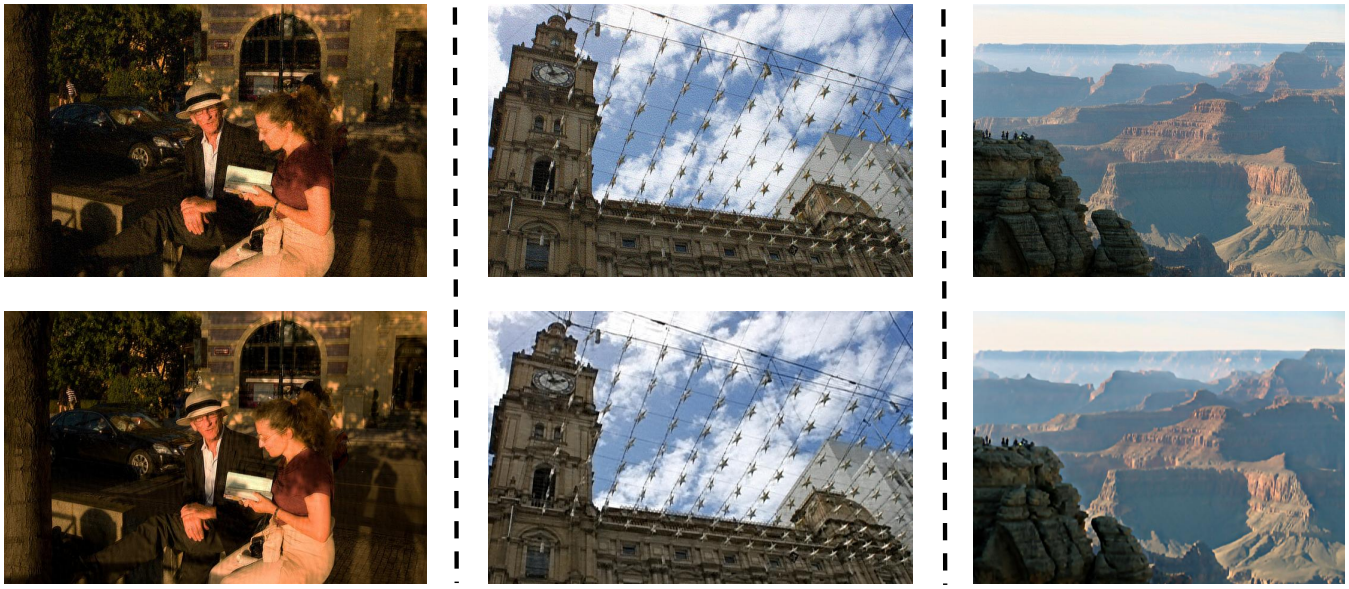


Figure S1: Fine-grained image pairs for Deblurring

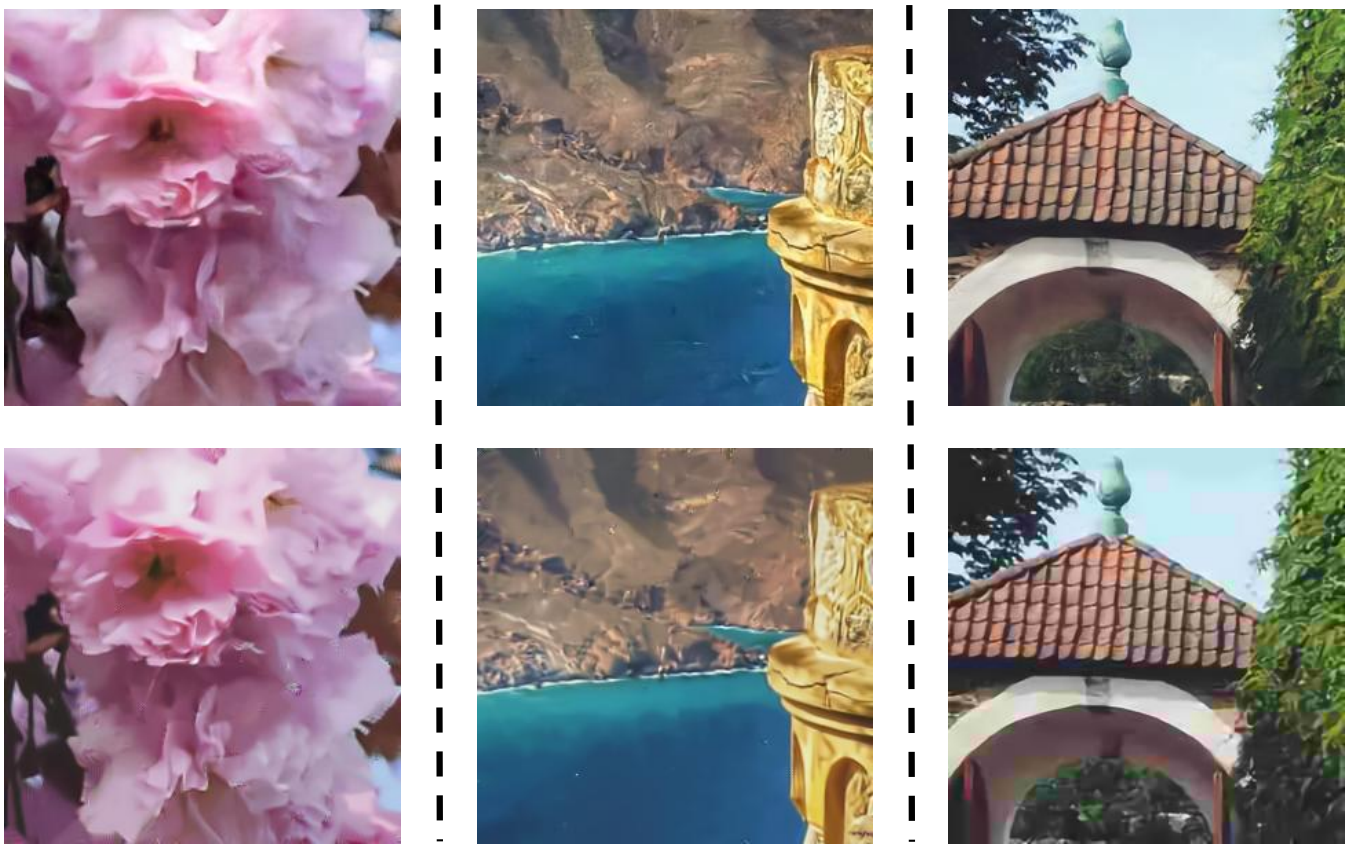


Figure S2: Fine-grained image pairs for Denoising

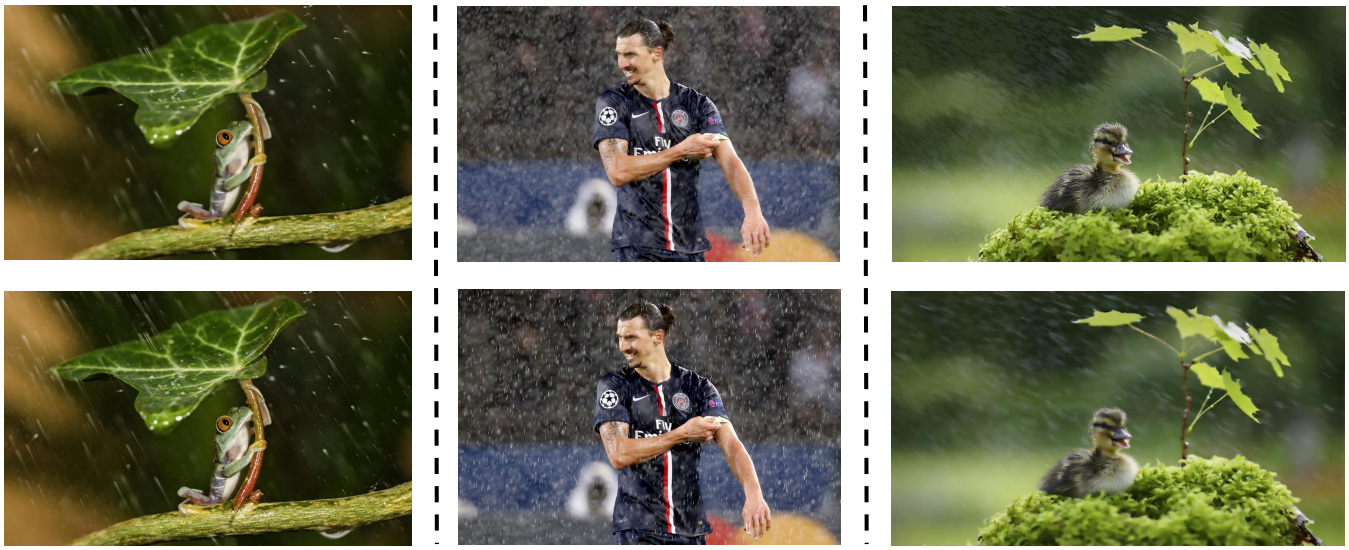


Figure S3: Fine-grained image pairs for Deraining

Figure S4: Fine-grained image pairs for Super-Resolution

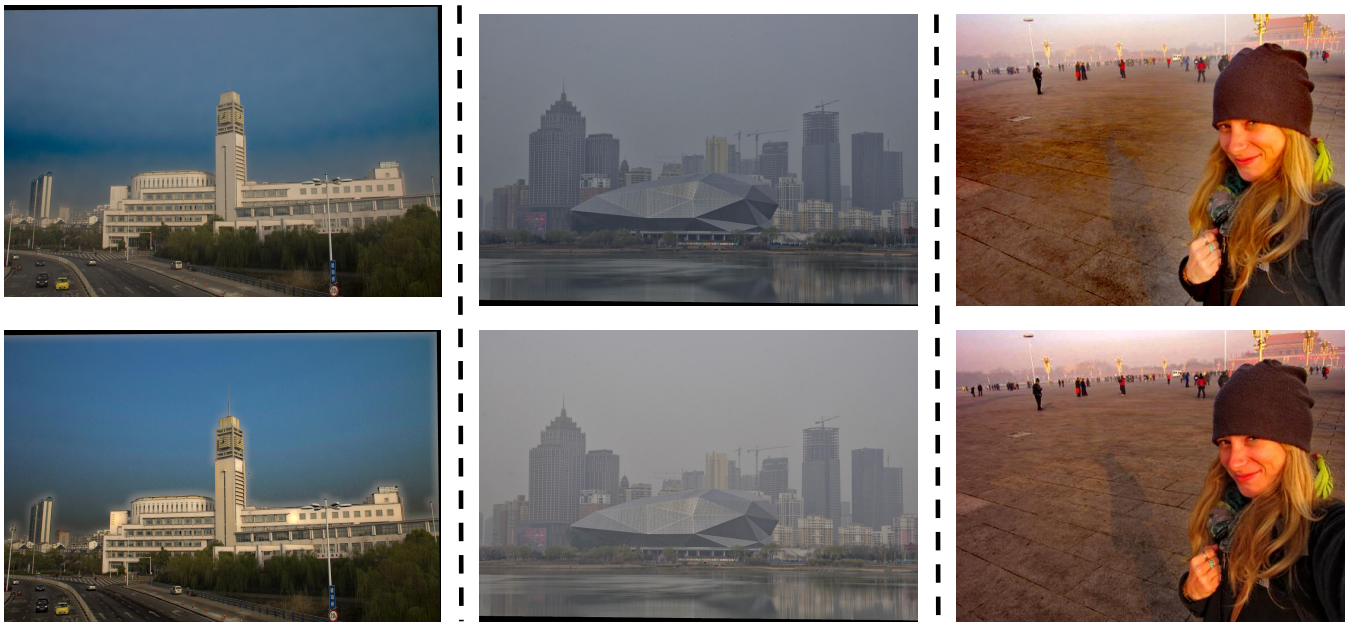


Figure S5: Fine-grained image pairs for Dehazing

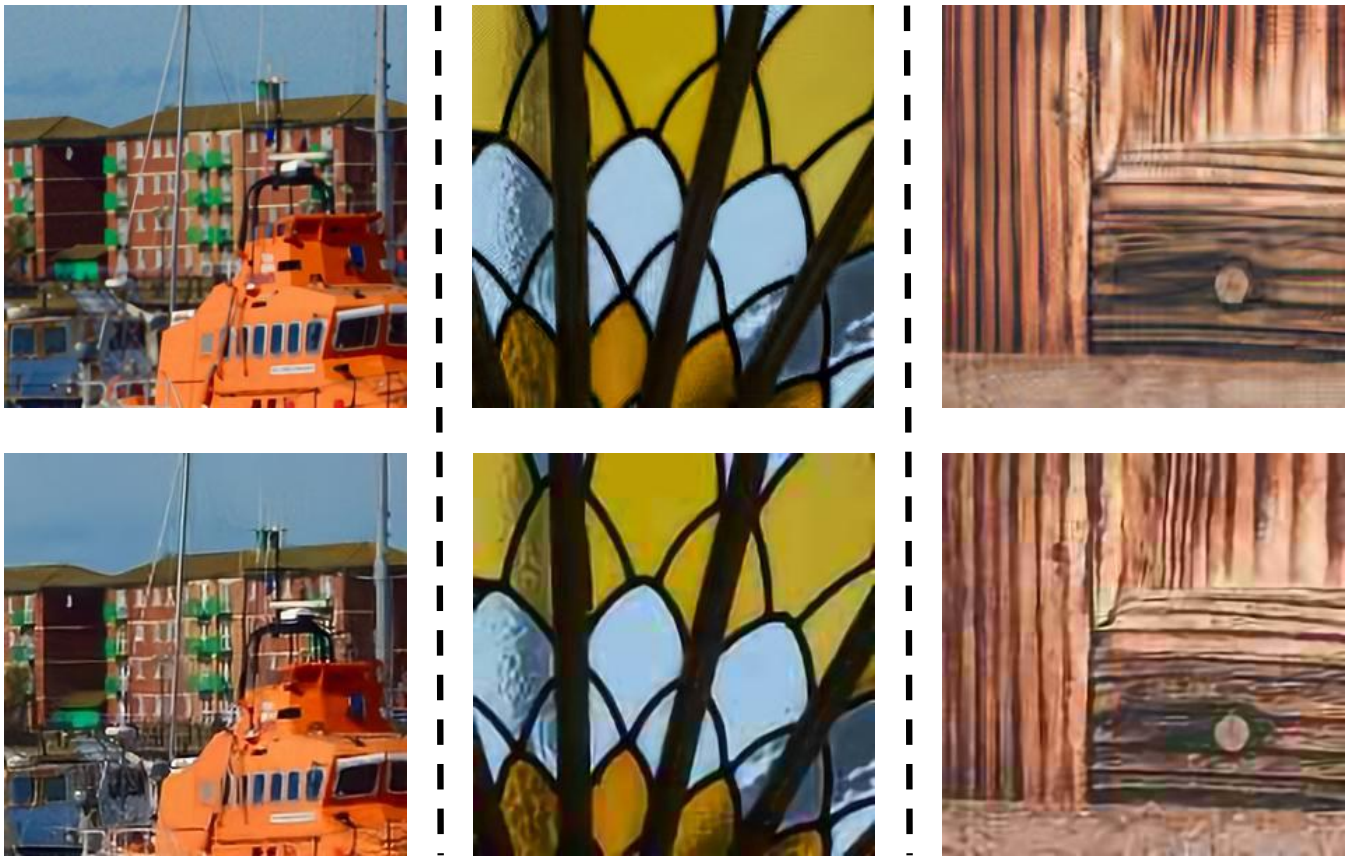


Figure S6: Fine-grained image pairs for Mixture Restoration

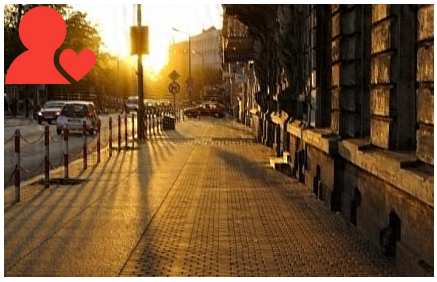
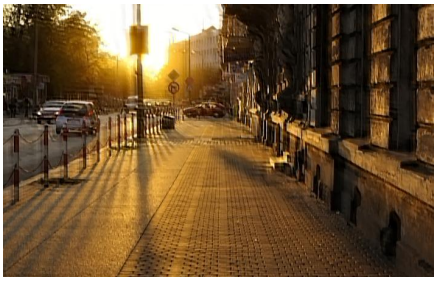
Image A	Image B	Metric	$Score_A$	RANK	$Score_B$
		PSNR ↑	16.00	✓	15.81
		SSIM ↑	0.576	✗	0.585
		DISTS ↓	0.164	✓	0.198
		CLIP-IQA ↑	0.998	✗	0.999
		DeQA-Score ↑	3.572	✓	3.114
		FGResQ	Image A is better ✓		
		PSNR ↑	15.57	✗	15.76
		SSIM ↑	0.249	✗	0.254
		DISTS ↓	0.195	✓	0.204
		CLIP-IQA ↑	0.757	✓	0.219
		DeQA-Score ↑	3.750	✗	3.882
		FGResQ	Image A is better ✓		

Figure S7: Qualitative comparison for Deblurring

Image A	Image B	Metric	$Score_A$	RANK	$Score_B$
		PSNR ↑	26.30	✗	25.55
		SSIM ↑	0.844	✓	0.849
		DISTS ↓	0.173	✗	0.184
		CLIP-IQA ↑	0.852	✗	0.323
		DeQA-Score ↑	3.097	✗	2.918
		FGResQ	Image B is better ✓		
		PSNR ↑	24.47	✗	23.88
		SSIM ↑	0.783	✓	0.788
		DISTS ↓	0.173	✗	0.233
		CLIP-IQA ↑	0.080	✗	0.015
		DeQA-Score ↑	3.318	✓	3.336
		FGResQ	Image B is better ✓		

Figure S8: Qualitative comparison for Denoising

Image A	Image B	Metric	$Score_A$	RANK	$Score_B$
		PSNR ↑	23.23	✗	27.14
		SSIM ↑	0.897	✓	0.813
		DISTS ↓	0.086	✓	0.129
		CLIP-IQA ↑	0.309	✗	0.914
		DeQA-Score ↑	2.393	✓	2.225
		FGResQ	Image A is better ✓		
		PSNR ↑	34.44	✗	32.41
		SSIM ↑	0.915	✗	0.893
		DISTS ↓	0.085	✗	0.097
		CLIP-IQA ↑	0.139	✓	0.593
		DeQA-Score ↑	1.998	✓	2.219
		FGResQ	Image B is better ✓		

Figure S9: Qualitative comparison for Deraining

Image A	Image B	Metric	$Score_A$	RANK	$Score_B$
		PSNR ↑	19.18	✗	17.34
		SSIM ↑	0.625	✗	0.548
		DISTS ↓	0.174	✓	0.154
		CLIP-IQA ↑	0.153	✓	1.000
		DeQA-Score ↑	3.162	✓	3.283
		FGResQ	Image B is better ✓		
		PSNR ↑	22.55	✗	24.47
		SSIM ↑	0.795	✗	0.851
		DISTS ↓	0.184	✓	0.211
		CLIP-IQA ↑	0.001	✗	0.229
		DeQA-Score ↑	2.604	✗	2.642
		FGResQ	Image A is better ✓		

Figure S10: Qualitative comparison for Super-Resolution

Image A	Image B	Metric	$Score_A$	RANK	$Score_B$
		PSNR ↑	20.7267	✗	19.1576
		SSIM ↑	0.9202	✗	0.8777
		DISTS ↓	0.1659	✗	0.2339
		CLIP-IQA ↑	0.8345	✗	0.6686
		DeQA-Score ↑	2.2108	✓	2.3442
		FGResQ	Image B is better ✓		
		PSNR ↑	16.15	✗	14.15
		SSIM ↑	0.912	✗	0.833
		DISTS ↓	0.056	✗	0.074
		CLIP-IQA ↑	0.999	✓	1.000
		DeQA-Score ↑	3.752	✓	3.789
		FGResQ	Image A is better ✓		

Figure S11: Qualitative comparison for Dehazing

Image A	Image B	Metric	$Score_A$	RANK	$Score_B$
		PSNR ↑	27.34	✗	29.62
		SSIM ↑	0.833	✗	0.886
		DISTS ↓	0.134	✗	0.117
		CLIP-IQA ↑	0.860	✗	0.970
		DeQA-Score ↑	2.970	✗	3.085
		FGResQ	Image A is better ✓		
		PSNR ↑	21.57	✗	18.98
		SSIM ↑	0.829	✗	0.764
		DISTS ↓	0.177	✗	0.217
		CLIP-IQA ↑	0.207	✓	0.979
		DeQA-Score ↑	3.582	✓	3.627
		FGResQ	Image B is better ✓		

Figure S12: Qualitative comparison for Mixture Restoration