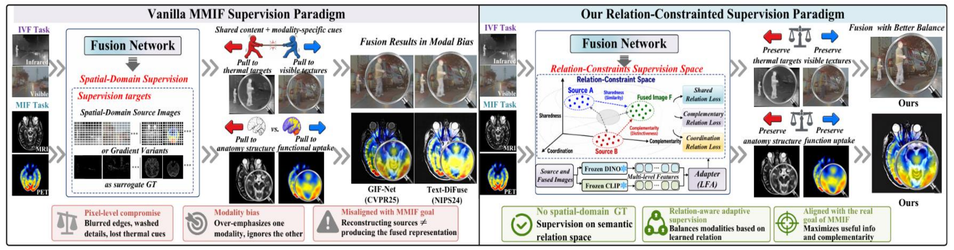
Beyond Spatial-Domain Supervision: A Relation Constrained Space for Multi-Modal Image Fusion

Zeyu Wang[1], Mingyu Ge[1], Haiyu Song[1,*], Haoran Duan[2,*]

[1] Dalian Minzu University [2] Tsinghua University

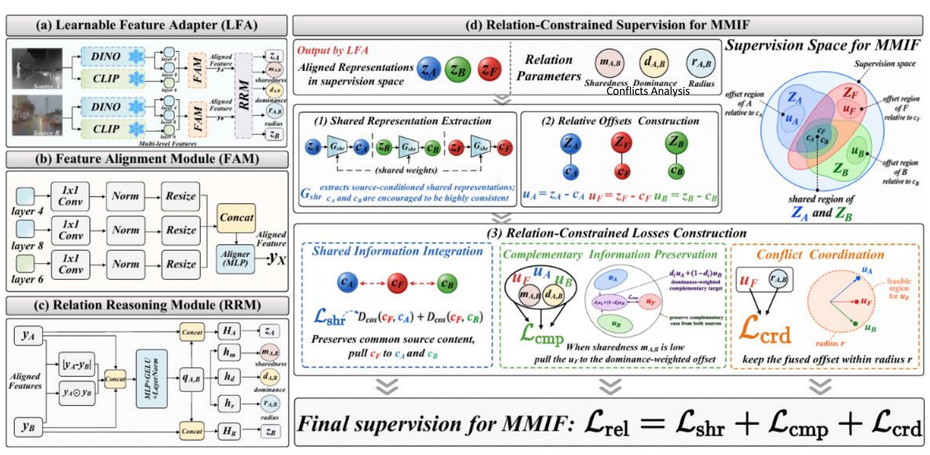
NEURAL INFORMATION PROCESSING SYSTEMS

Motivation

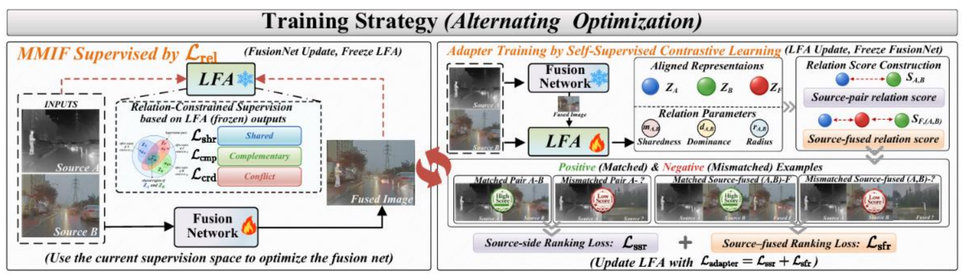


Vanilla MMIF Supervision Paradigm vs. OurParadigm.

Method



Overview of the Learnable Feature Adapter and Relation-Constrained Supervision.

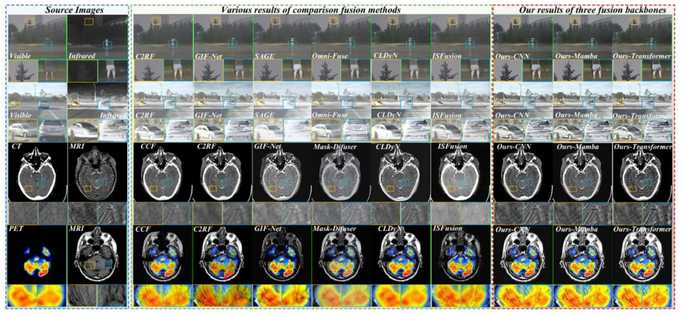


Overview of the Alternating Optimization Strategy.

Contributions

- *We **revisit MMIF supervision** from the objective of fusion and **propose a relation-constrained paradigm** beyond spatial domain source images as surrogate ground truths.*
- *We define the supervision space through **sharedness, complementarity, and coordination,** aligning supervision with the requirements of fusion.*
- *We **design a learnable feature adapter** that aligns heterogeneous pretrained features and predicts relation parameters to automatically construct the supervision space.*
- *We **propose a self-supervised relation ranking strategy and an alternating optimization scheme** to jointly train the adapter and fusion network, forming a bidirectional optimization mechanism.*

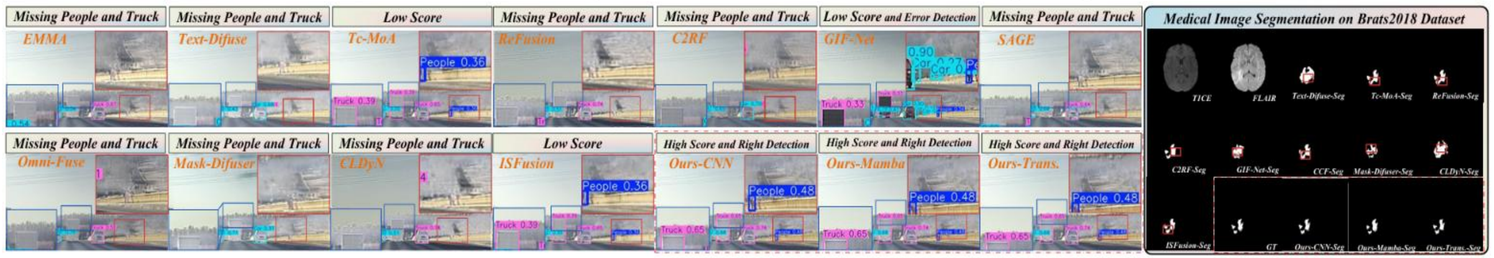
Qualiative and Quantitative Results



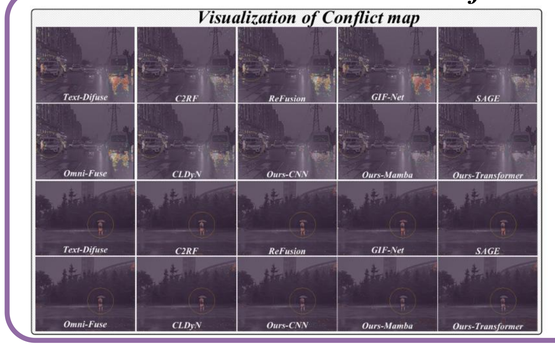
IVF Task		RoadScene							M³FD							MSRS						
Methods	Pub/Year	AG ↑	EN ↑	SD ↑	MI ↑	$Q_{AB/F}$ ↑	Q_M ↑	Q_P ↑	AG ↑	EN ↑	SD ↑	MI ↑	$Q_{AB/F}$ ↑	Q_M ↑	Q_P ↑	AG ↑	EN ↑	SD ↑	MI ↑	$Q_{AB/F}$ ↑	Q_M ↑	Q_P ↑
EMMA	CVPR 24	4.779	49.127	46.777	2.809	0.436	0.432	0.355	5.871	57.569	42.377	3.771	0.589	0.512	0.460	3.788	40.539	40.590	4.164	0.541	0.760	0.496
Text-DiFuse	NIPS 24	3.962	42.663	48.966	1.961	0.140	0.288	0.052	1.861	20.580	40.568	2.054	0.066	0.240	0.332	2.792	30.723	39.679	1.310	0.118	0.288	0.038
Tc-MoA	CVPR 24	4.192	44.883	42.610	1.043	0.413	0.556	0.403	5.180	54.156	39.030	3.333	0.651	0.891	0.513	3.618	38.414	42.335	3.493	0.630	1.090	0.483
ReFusion	IJCV 24	5.357	59.336	48.229	2.726	0.437	0.566	0.397	5.338	57.331	40.803	3.548	0.637	0.731	0.489	3.813	40.641	42.890	4.102	0.442	1.085	0.576
C2RF	IJCV 25	4.989	52.270	45.576	2.720	0.519	0.565	0.388	4.779	50.329	39.348	2.498	0.223	0.287	0.399	3.474	37.153	41.348	1.729	0.185	0.330	0.092
GIF-Net	CVPR 25	5.705	59.855	45.000	2.236	0.352	0.365	0.252	5.435	52.096	42.714	2.529	0.471	0.319	0.319	3.379	36.267	32.910	1.926	0.367	0.396	0.277
SAGE	CVPR 25	3.715	39.057	46.384	1.035	0.362	0.451	0.332	4.885	50.959	40.930	2.851	0.592	0.599	0.434	3.342	35.157	37.485	2.990	0.539	0.714	0.435
Omni-Fuse	TPAMI 25	3.206	35.212	44.830	2.187	0.303	0.396	0.267	4.514	49.067	41.029	3.208	0.430	0.362	0.277	3.091	34.123	40.043	2.664	0.361	0.412	0.217
CLDyN	CVPR 26	4.918	51.612	40.644	1.827	0.417	0.475	0.329	5.824	57.312	35.254	3.235	0.689	0.756	0.520	3.568	37.882	39.377	2.111	0.661	0.867	0.484
ISFusion	CVPR 26	5.320	54.393	35.946	2.345	0.522	0.484	0.306	5.080	52.927	36.718	2.188	0.511	0.422	0.343	4.124	42.850	34.100	3.303	0.491	0.559	0.357
Ours(CNN)	- -	5.701	59.876	49.673	2.962	0.541	0.604	0.411	5.478	57.921	42.315	4.238	0.664	1.041	0.529	3.842	40.982	43.517	4.301	0.651	1.337	0.509
Ours(Mamba)	- -	5.858	60.742	50.214	3.084	0.552	0.621	0.417	5.602	58.736	42.982	4.612	0.671	1.102	0.533	3.902	41.427	43.918	4.493	0.664	1.462	0.518
Ours(Transformer)	- -	5.861	61.069	50.581	3.179	0.564	0.643	0.425	5.666	59.158	43.407	4.837	0.679	1.157	0.540	3.953	41.854	44.231	4.651	0.678	1.576	0.526

MIF Task		CT-MRI							MRI-PET							MRI-SPECT						
Methods	Pub/Year	AG ↑	EN ↑	SD ↑	MI ↑	$Q_{AB/F}$ ↑	Q_M ↑	Q_P ↑	AG ↑	EN ↑	SD ↑	MI ↑	$Q_{AB/F}$ ↑	Q_M ↑	Q_P ↑	AG ↑	EN ↑	SD ↑	MI ↑	$Q_{AB/F}$ ↑	Q_M ↑	Q_P ↑
EMMA	CVPR 24	7.317	76.268	72.599	3.279	0.551	0.153	0.316	7.116	75.114	76.489	2.448	0.541	0.143	0.318	4.689	48.655	62.720	2.361	0.516	0.223	0.337
Text-DiFuse	NIPS 24	6.831	70.390	66.701	3.159	0.499	0.155	0.264	8.581	88.178	66.353	2.331	0.489	0.144	0.312	5.935	60.168	58.726	2.225	0.539	0.226	0.365
Tc-MoA	CVPR 24	6.888	72.785	72.894	3.317	0.592	0.162	0.369	7.148	75.676	69.960	2.450	0.640	0.195	0.469	4.251	44.348	55.776	2.333	0.426	0.272	0.356
ReFusion	IJCV 24	8.731	88.753	76.113	3.176	0.538	0.223	0.356	7.800	80.686	72.422	2.486	0.539	0.212	0.458	5.490	55.098	61.274	2.310	0.580	0.557	0.372
C2RF	IJCV 25	7.978	81.643	76.192	2.953	0.332	0.171	0.131	7.933	81.310	70.626	2.183	0.458	0.216	0.284	5.369	54.366	59.497	2.257	0.547	0.400	0.309
GIF-Net	CVPR 25	9.544	94.839	70.216	3.067	0.466	0.128	0.236	6.489	67.343	61.177	2.299	0.348	0.105	0.245	4.742	48.710	51.468	2.244	0.432	0.176	0.304
CCF	NIPS 24	6.724	67.686	70.568	3.284	0.546	0.283	0.546	6.461	65.927	65.437	2.493	0.512	0.193	0.273	4.398	44.729	64.837	2.241	0.526	0.285	0.391
Mask-Difuser	TPAMI 25	5.984	60.253	69.600	3.299	0.484	0.166	0.213	6.126	61.783	62.930	2.706	0.572	0.200	0.440	4.164	41.390	58.787	2.127	0.548	0.303	0.302
CLDyN	CVPR 26	7.777	80.208	73.716	3.245	0.597	0.206	0.394	7.618	79.791	72.779	2.466	0.670	0.220	0.490	4.919	50.583	63.594	2.870	0.528	0.330	0.609
ISFusion	CVPR 26	8.944	88.753	72.945	2.953	0.556	0.171	0.315	8.322	81.579	66.199	2.171	0.579	0.177	0.349	5.531	54.280	58.691	2.169	0.570	0.267	0.393
Ours(CNN)	- -	9.558	94.876	75.642	3.352	0.611	0.249	0.397	9.127	96.512	76.938	2.514	0.604	0.219	0.463	6.145	63.876	66.873	2.344	0.581	0.423	0.396
Ours(Mamba)	- -	9.573	95.214	76.308	3.421	0.626	0.267	0.443	9.412	97.864	77.614	2.592	0.619	0.228	0.474	6.382	65.214	67.942	2.401	0.594	0.451	0.405
Ours(Transformer)	- -	9.608	95.842	76.527	3.548	0.637	0.290	0.411	9.690	99.031	78.072	2.667	0.627	0.235	0.481	6.571	66.683	68.496	2.448	0.608	0.464	0.413

Downstream Tasks: Object Detection and Medical Imaga Segmentation



Conflict Analysis



Method	CR (%) ↓	CS ↓	UI ↓
Text-DiFuse	27.25	0.806	0.4527
C2RF	16.04	0.715	0.3379
ReFusion	20.91	0.784	0.3865
GIF-Net	18.36	0.692	0.3514
SAGE	15.28	0.641	0.3196
Omni-Fuse	17.63	0.667	0.3421
CLDyN	13.72	0.584	0.2968
Ours (CNN)	7.42	0.315	0.218
Ours (Mamba)	6.37	0.287	0.198
Ours (Transformer)	4.77	0.265	0.176

Contact Us **GitHub**