CDDFuse: Correlation-Driven Dual-Branch Feature Decomposition for Multi-Modality Image Fusion

Zixiang Zhao[1,2], Haowen Bai[1], Jiangshe Zhang[1], Yulun Zhang[2],
Shuang Xu[3], Zudi Lin[4], Radu Timofte[2,5], Luc Van Gool[2]

[1]Xi’an Jiaotong University [2]Computer Vision Lab, ETH Zurich
[3]Northwestern Polytechnical University [4]Harvard University [5]University of Wurzburg

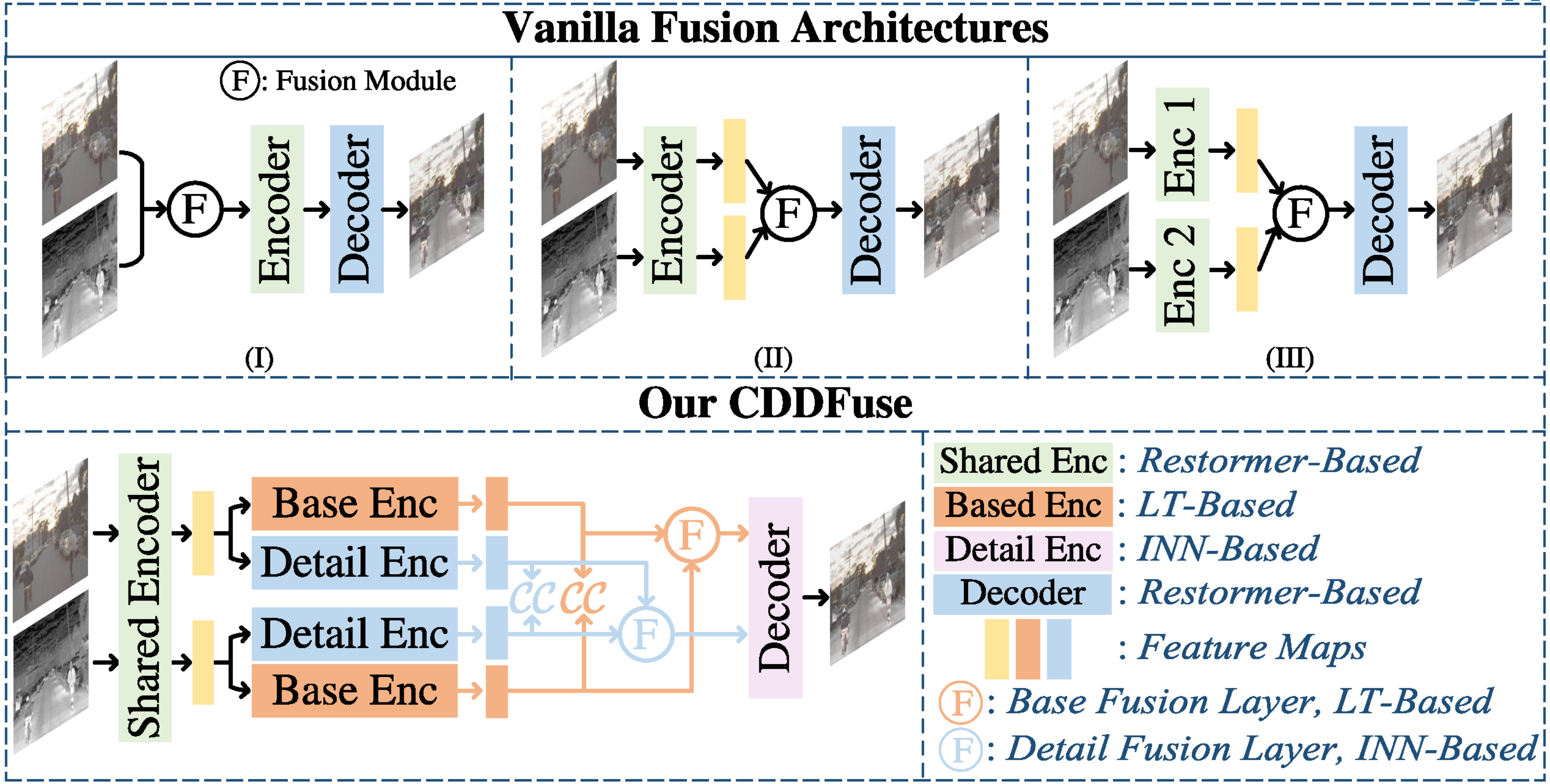
Infrared and visible image fusion

- Infrared and visible images:
  - ✓ infrared images: discriminative thermal radiations & ignoring illumination.
  - ✓ visible images: textural details & high spatial resolution.

Infrared & Visible images   Fusion image

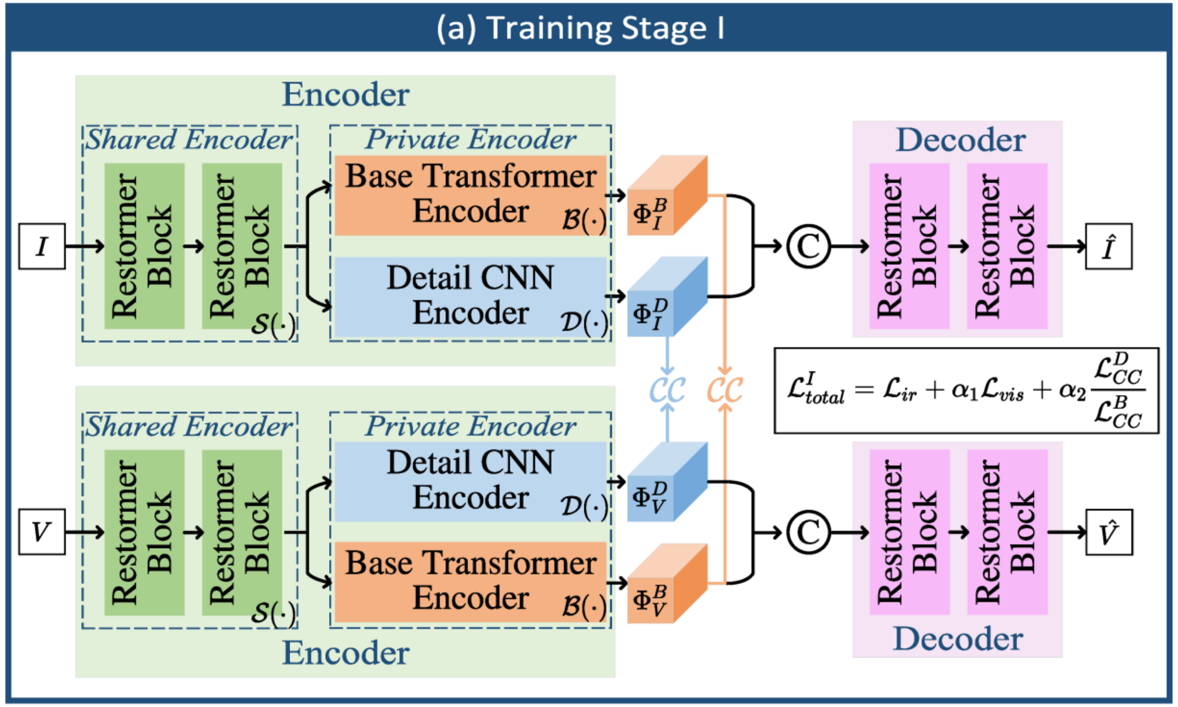
- Fusion images:
  - ✓ highlight radiation information (brightness and contrast)
  - ✓ detailed texture information (gradients and edges)
  - ✓ a clear, complete and accurate description of the targets

Challenge & Solution



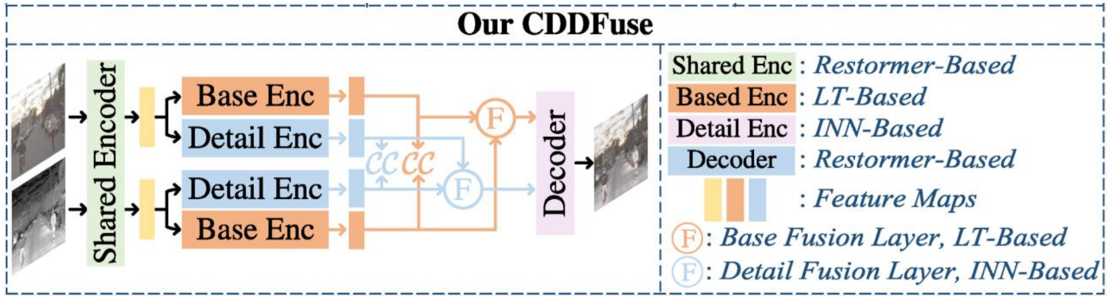
- Challenges :
  - ✓ interpreting the working mechanism
  - ✓ extracting cross-modal features
  - ✓ loss of high-frequency information
- CDDFuse:
  - ✓ adding correlation restrictions
  - ✓ dual-branch Transformer-CNN exactor
  - ✓ INN block in detail encoder

CDDFuse: Workflow



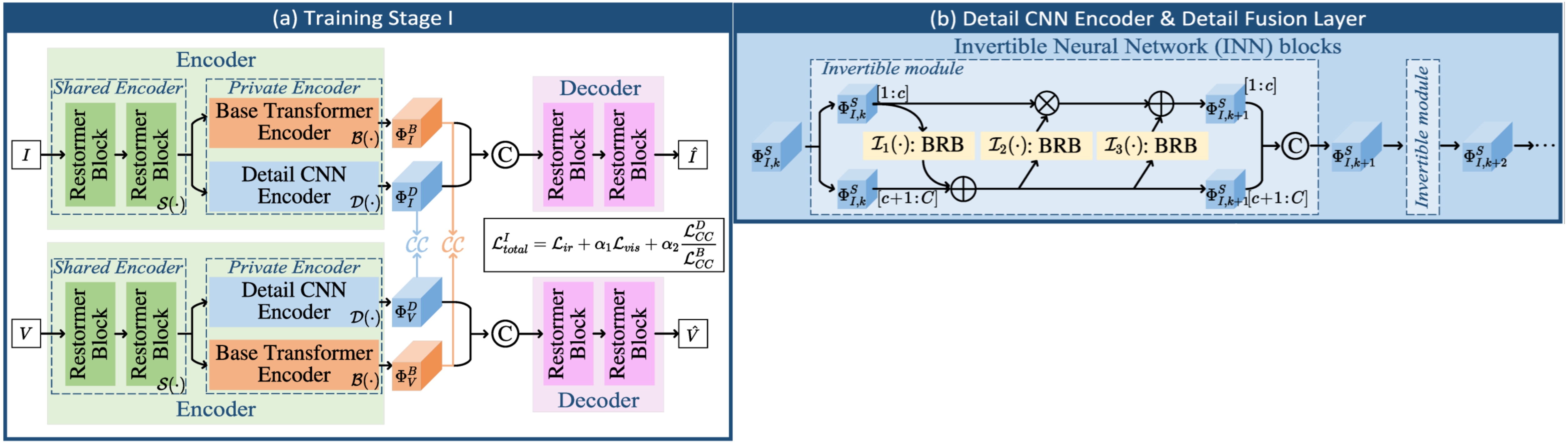
$$\mathcal{L}_{total}^{I} = \mathcal{L}_{ir} + \alpha_1 \mathcal{L}_{vis} + \alpha_2 \mathcal{L}_{decomp}, \quad \mathcal{L}_{decomp} = \frac{\left(\mathcal{L}_{CC}^{D}\right)^2}{\mathcal{L}_{CC}^{B}} = \frac{\left(\mathcal{CC}\left(\Phi_I^D, \Phi_V^D\right)\right)^2}{\mathcal{CC}\left(\Phi_I^B, \Phi_V^B\right) + \epsilon}$$

$$\mathcal{L}_{total}^{II} = \mathcal{L}_{int}^{II} + \alpha_3 \mathcal{L}_{grad} + \alpha_4 \mathcal{L}_{decomp},$$



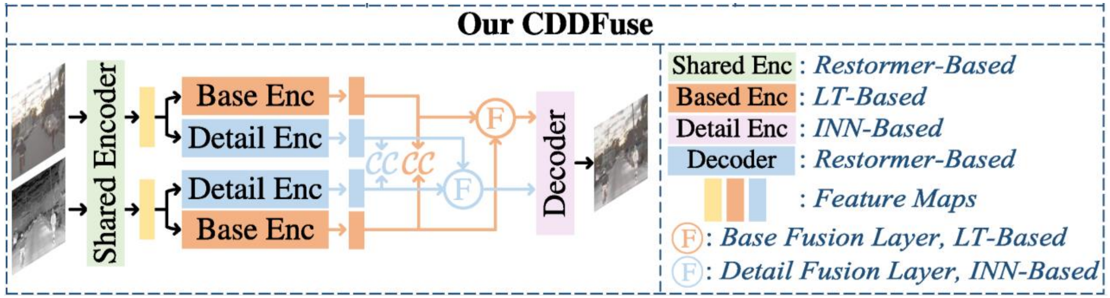
Zamir et al. Restormer: Efficient transformer for high-resolution image restoration. CVPR 2022.
Wu et al. Lite transformer with long-short range attention. ICLR 2020.
Dinh et al. Density estimation using real NVP. ICLR 2017

CDDFuse: Workflow



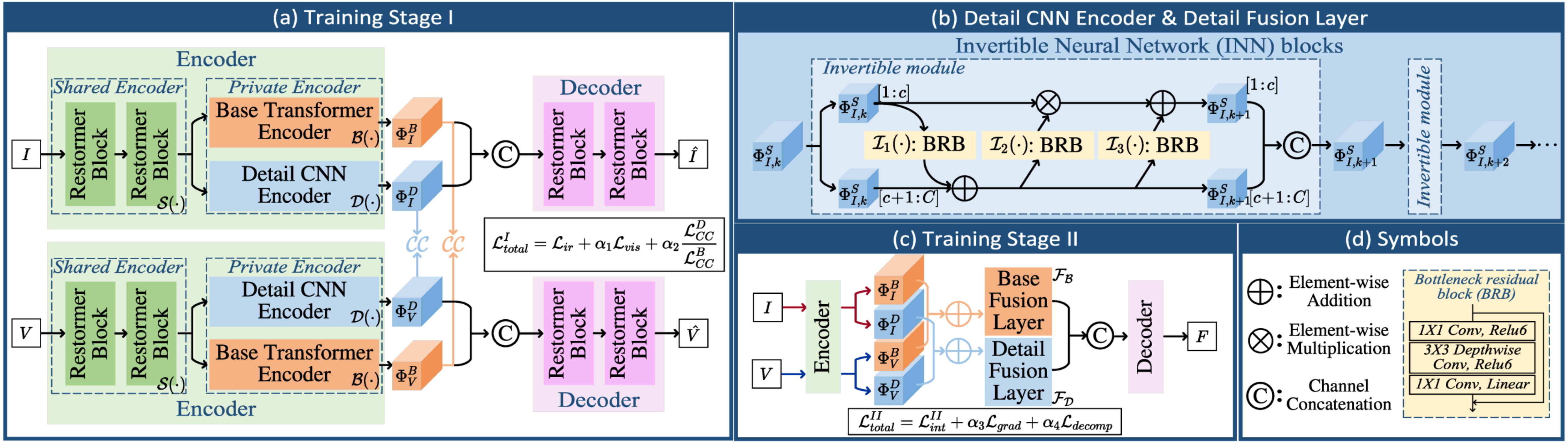
$$\mathcal{L}^{I}_{total} = \mathcal{L}_{ir} + \alpha_1 \mathcal{L}_{vis} + \alpha_2 \mathcal{L}_{decomp}, \quad \mathcal{L}_{decomp} = \frac{\left(\mathcal{L}^D_{CC}\right)^2}{\mathcal{L}^B_{CC}} = \frac{\left(CC\left(\Phi^D_I, \Phi^D_V\right)\right)^2}{CC\left(\Phi^B_I, \Phi^B_V\right) + \epsilon}$$

$$\mathcal{L}^{II}_{total} = \mathcal{L}^{II}_{int} + \alpha_3 \mathcal{L}_{grad} + \alpha_4 \mathcal{L}_{decomp},$$



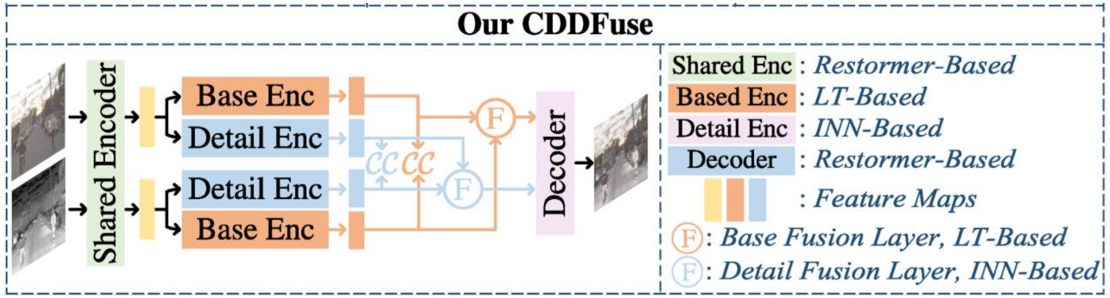
Zamir et al. Restormer: Efficient transformer for high-resolution image restoration. CVPR 2022.
Wu et al. Lite transformer with long-short range attention. ICLR 2020.
Dinh et al. Density estimation using real NVP. ICLR 2017

CDDFuse: Workflow



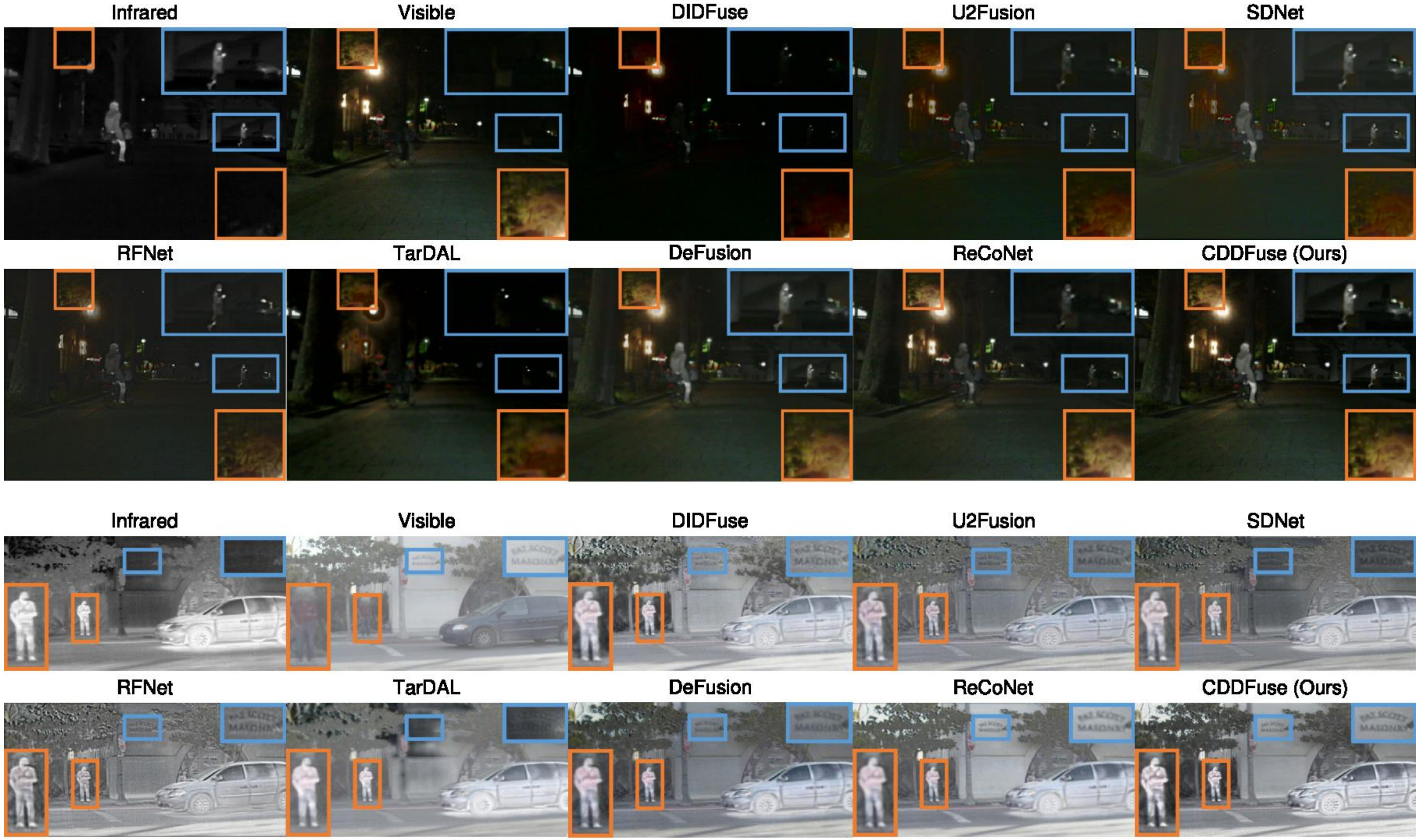
$$\mathcal{L}^{I}_{total} = \mathcal{L}_{ir} + \alpha_1\mathcal{L}_{vis} + \alpha_2\mathcal{L}_{decomp}, \quad \mathcal{L}_{decomp} = \frac{\left(\mathcal{L}^D_{CC}\right)^2}{\mathcal{L}^B_{CC}} = \frac{\left(\mathcal{CC}\left(\Phi^D_I, \Phi^D_V\right)\right)^2}{\mathcal{CC}\left(\Phi^B_I, \Phi^B_V\right) + \epsilon}$$

$$\mathcal{L}^{II}_{total} = \mathcal{L}^{II}_{int} + \alpha_3\mathcal{L}_{grad} + \alpha_4\mathcal{L}_{decomp},$$



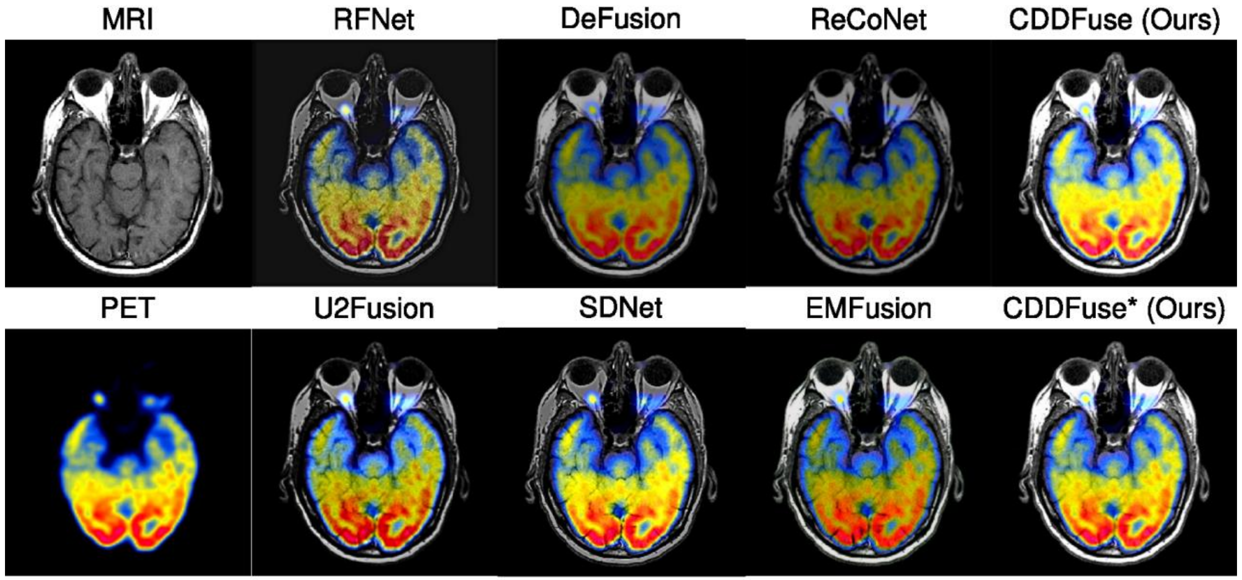
Zamir et al. Restormer: Efficient transformer for high-resolution image restoration. CVPR 2022.
Wu et al. Lite transformer with long-short range attention. ICLR 2020.
Dinh et al. Density estimation using real NVP. ICLR 2017

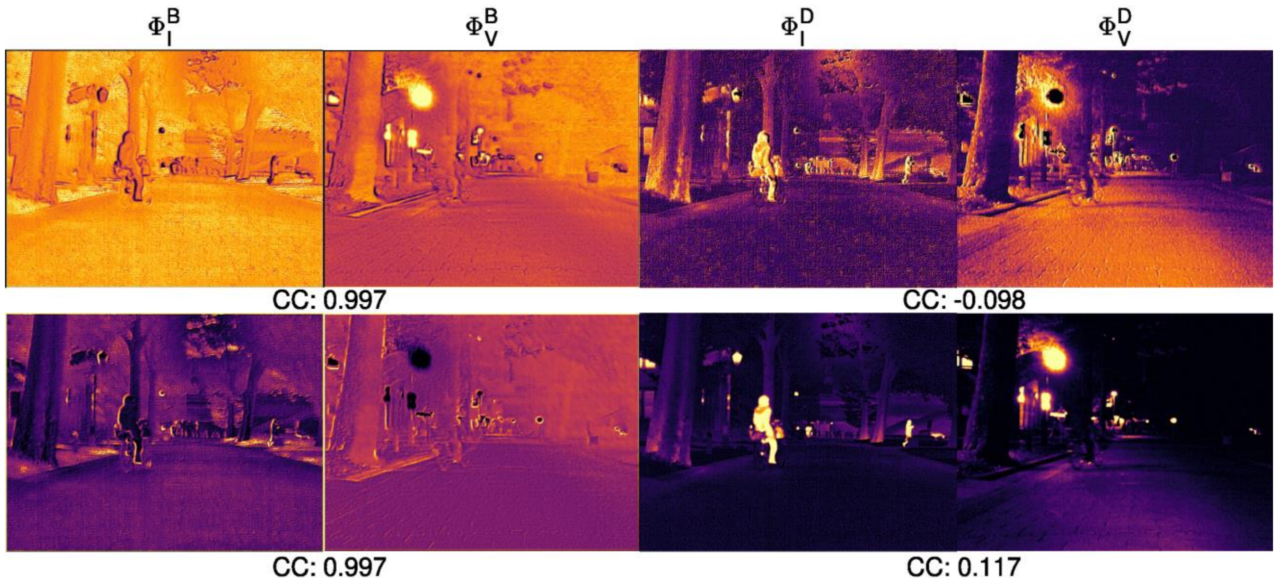
Qualitative comparison



Visual comparison for “FLIR 04602” and “00706N” in infrared-visible image fusion.

Qualitative comparison





Visual comparison for "MRI-PET-16" in medical image fusion.

Visualization of the decomposed features.

Quantitative comparison

Quantitative results of the IVF task.

Dataset: MSRS Infrared-Visible Fusion Dataset [57]	EN	SD	SF	MI	SCD	VIF	Qbaf	SSIM
DID [88]	4.27	31.49	10.15	1.61	1.11	0.31	0.20	0.24
U2F [70]	5.37	25.52	9.07	1.40	1.24	0.54	0.42	0.77
SDN [82]	5.25	17.35	8.67	1.19	0.99	0.50	0.38	0.72
RFN [72]	5.56	24.09	**11.98**	1.30	1.13	0.51	0.43	0.83
TarD [35]	5.28	25.22	5.98	1.49	0.71	0.42	0.18	0.47
DeF [32]	6.46	37.63	8.60	2.16	1.35	0.77	0.54	0.94
ReC [19]	6.61	43.24	9.77	2.16	1.44	0.71	0.50	0.85
CDDFuse	**6.70**	**43.38**	11.56	**3.47**	**1.62**	**1.05**	**0.69**	**1.00**

Dataset: TNO Infrared-Visible Fusion Dataset [59]	EN	SD	SF	MI	SCD	VIF	Qbaf	SSIM
DID [88]	6.97	45.12	12.59	1.70	1.71	0.60	0.40	0.81
U2F [70]	6.83	34.55	11.52	1.37	1.71	0.58	0.44	0.99
SDN [82]	6.64	32.66	12.05	1.52	1.49	0.56	0.44	1.00
RFN [72]	6.83	34.50	**15.71**	1.20	1.67	0.51	0.39	0.92
TarD [35]	6.84	45.63	8.68	1.86	1.52	0.53	0.32	0.88
DeF [32]	6.95	38.41	8.21	1.78	1.64	0.60	0.41	0.96
ReC [19]	7.10	44.85	8.73	1.78	1.70	0.57	0.39	0.88
CDDFuse	**7.12**	**46.00**	13.15	**2.19**	**1.76**	**0.77**	**0.54**	**1.03**

Dataset: RoadScene Infrared-Visible Fusion Dataset [71]	EN	SD	SF	MI	SCD	VIF	Qbaf	SSIM
DID [88]	7.43	51.58	14.66	2.11	1.70	0.58	0.48	0.86
U2F [70]	7.09	38.12	13.25	1.87	1.70	0.60	0.51	0.97
SDN [82]	7.14	40.20	13.70	2.21	1.49	0.60	0.51	**0.99**
RFN [72]	7.21	41.25	16.19	1.68	1.73	0.54	0.45	0.90
TarD [35]	7.17	47.44	10.83	2.14	1.55	0.54	0.40	0.88
DeF [32]	7.23	44.44	10.22	2.25	1.69	0.63	0.48	0.89
ReC [19]	7.36	52.54	10.78	2.18	1.74	0.59	0.43	0.88
CDDFuse	**7.44**	**54.67**	**16.36**	**2.30**	**1.81**	**0.69**	**0.52**	0.98

Quantitative results of the MIF task.

Dataset: MRI-CT Medical Image Fusion	EN	SD	SF	MI	SCD	VIF	$Q^{AB/F}$	SSIM
TarD [35]	4.75	61.14	28.38	1.94	0.81	0.32	0.35	0.61
RFN [72]	**5.30**	52.95	33.42	1.98	0.58	0.33	0.52	0.49
DeF [32]	4.63	66.38	21.56	2.20	1.12	0.47	0.44	1.29
ReC [19]	4.41	66.96	20.16	2.03	1.24	0.40	0.42	1.29
CDDFuse	4.83	**88.59**	**33.83**	**2.24**	**1.74**	**0.50**	**0.59**	**1.31**
U2F [70]	4.88	52.98	22.54	2.08	0.75	0.37	0.46	0.49
SDN [82]	**5.02**	60.07	29.41	2.14	0.97	0.38	0.47	0.51
EMF [69]	4.76	72.76	22.56	2.34	1.32	0.56	0.49	1.31
CDDFuse*	4.88	**79.17**	**38.14**	**2.61**	**1.41**	**0.61**	**0.68**	**1.34**

Dataset: MRI-PET Medical Image Fusion	EN	SD	SF	MI	SCD	VIF	$Q^{AB/F}$	SSIM
TarD [35]	3.81	57.65	23.65	1.36	1.46	0.57	0.58	0.68
RFN [72]	**4.77**	50.57	**29.11**	1.53	0.96	0.39	0.52	0.42
DeF [32]	4.17	64.65	22.35	1.74	1.48	0.58	0.56	1.45
ReC [19]	3.66	65.25	21.72	1.51	1.49	0.44	0.51	1.40
CDDFuse	4.24	**81.72**	28.04	**1.87**	**1.82**	**0.66**	**0.65**	**1.46**
U2F [70]	3.73	57.07	23.27	1.69	1.27	0.40	0.49	1.39
SDN [82]	3.83	61.40	**31.97**	1.71	1.40	0.47	0.57	1.46
EMF [69]	4.21	56.80	26.01	1.82	1.31	0.62	0.67	1.47
CDDFuse*	**4.23**	**70.73**	29.57	**2.03**	**1.69**	**0.71**	**0.71**	**1.49**

Dataset: MRI-SPECT Medical Image Fusion	EN	SD	SF	MI	SCD	VIF	$Q^{AB/F}$	SSIM
TarD [35]	3.66	53.46	18.50	1.44	0.90	0.64	0.52	0.36
RFN [72]	**4.39**	44.01	**23.77**	1.60	0.72	0.45	0.58	0.37
DeF [32]	3.81	56.65	15.45	1.80	1.27	0.61	0.56	**1.46**
ReC [19]	3.22	60.07	17.40	1.50	1.47	0.46	0.54	1.40
CDDFuse	3.91	**71.82**	20.68	**1.89**	**1.92**	**0.66**	**0.69**	1.44
U2F [70]	3.47	52.97	19.58	1.68	1.28	0.48	0.57	1.41
SDN [82]	3.43	49.62	**22.20**	1.69	1.09	0.55	0.66	1.48
EMF [69]	3.74	51.93	17.14	1.88	1.12	0.71	0.74	**1.49**
CDDFuse*	**3.90**	**58.31**	20.87	**2.49**	**1.35**	**0.97**	**0.78**	1.48

Quantitative comparison

Results of the multi-modal detection.

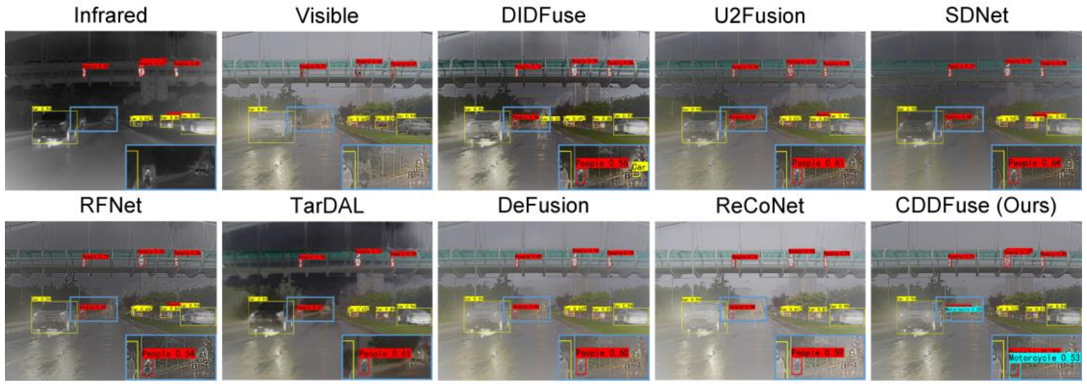
Table 3. AP@0.5(%) values for MM detection on M^3FD dataset.

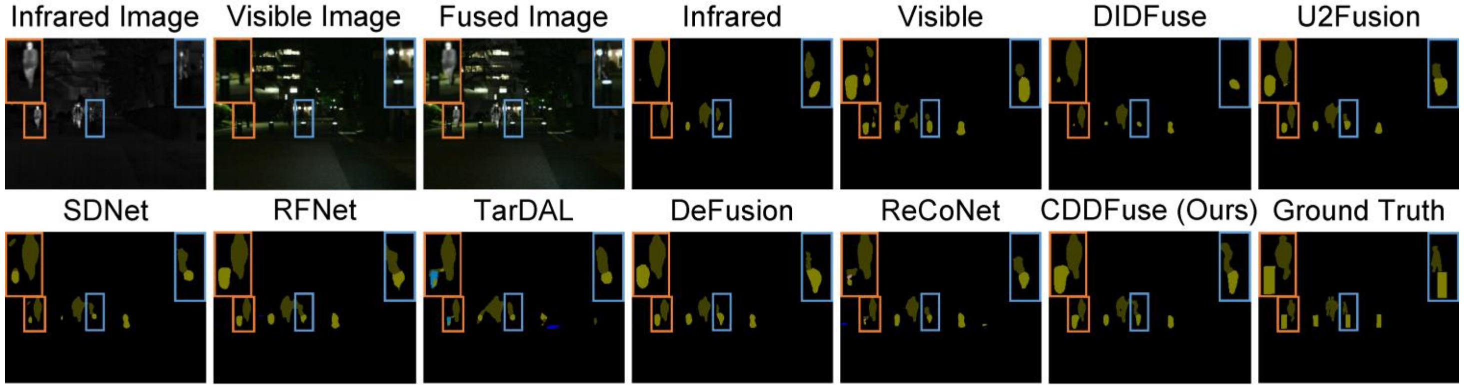
	Bus	Car	Lam	Mot	Peo	Tru	mAP@0.5
IR	78.75	88.69	70.17	63.42	80.91	65.77	74.62
VI	78.29	90.73	86.35	69.33	70.53	70.91	77.69
DID	79.65	92.51	84.70	68.72	79.61	68.78	78.99
U2F	79.15	92.29	87.61	66.75	80.67	71.37	79.64
SDN	81.44	92.33	84.14	67.37	79.35	69.29	78.99
RFN	78.15	91.94	84.95	**72.80**	79.41	69.04	79.38
TarD	81.33	**94.76**	87.13	69.34	81.52	68.65	80.45
DeF	**82.94**	92.49	**87.78**	69.45	80.82	71.44	80.82
ReC	78.92	91.79	87.41	69.34	79.41	69.98	79.48
Ours	82.60	92.54	86.88	71.62	**81.60**	**71.53**	**81.13**

Results of the multi-modal segmentation.

Table 4. IoU(%) values for MM segmentation on MSRS dataset.

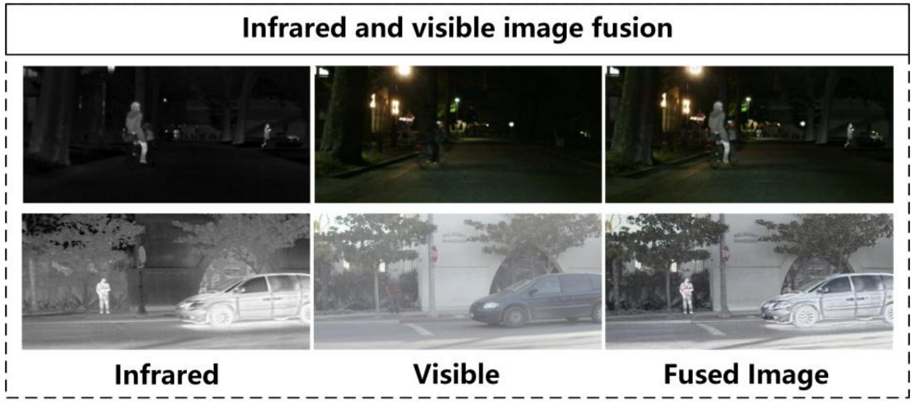
Models	Unl	Car	Per	Bik	Cur	CS	GD	CC	Bu	mIOU
VI	90.5	75.6	45.4	59.4	37.2	51.0	46.4	43.5	50.2	55.4
IR	84.7	67.8	56.4	51.8	34.6	39.3	42.2	40.2	48.4	51.7
DID [88]	97.2	78.3	58.7	60.9	36.2	52.9	**62.4**	44.0	55.7	60.7
U2F [70]	97.5	82.3	63.4	62.6	40.3	52.6	51.9	44.8	**59.5**	61.7
SDN [82]	97.3	78.4	62.5	61.7	35.7	49.3	52.4	42.2	52.9	59.2
RFN [72]	97.3	78.7	60.6	61.3	36.3	49.4	45.6	45.7	48.0	58.1
TarD [35]	97.1	79.1	55.4	59.0	33.6	49.4	54.9	42.6	53.5	58.3
DeF [32]	97.5	82.6	61.1	62.6	40.4	51.5	48.1	47.9	54.8	60.7
ReC [19]	97.4	81.0	59.9	61.4	41.0	51.3	54.4	47.4	55.9	61.1
Ours	**97.7**	**84.6**	**64.2**	**65.1**	**43.9**	**53.8**	61.7	**50.6**	57.3	**64.3**

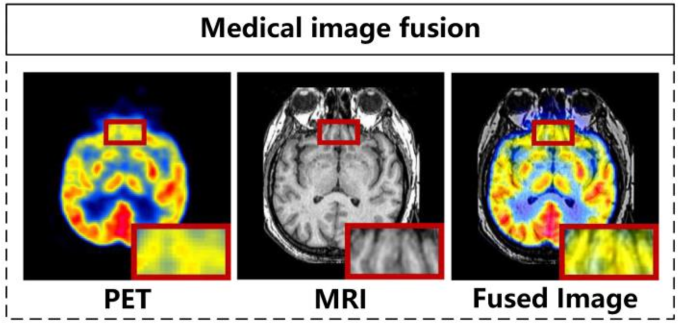


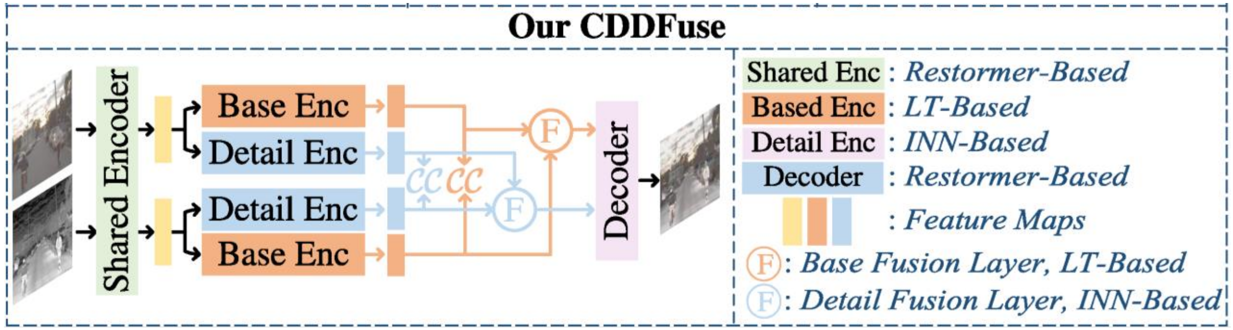


Take-home message

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA







- Image Fusion:
  - ✓ Highlight thermal radiation (infrared)
  - ✓ Detailed texture information (visible)
  - ✓ Clear and accurate representation (fused)
- Challenges :
  - ✓ Interpreting the working mechanism
  - ✓ Extracting cross-modal features
  - ✓ Loss of high-frequency information

- Cddfuse:
  - ✓ Adding correlation restrictions
  - ✓ Dual-branch Transformer-CNN exactor
  - ✓ INN block in detail encoder

Thanks For Listening!

https://github.com/Zhaozixiang1228/GDSR-DCTNet
zixiangzhao@stu.xjtu.edu.cn