

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

CDDFuse: 用于多模态图像融合的相关性驱动双分支特征分解

作者: Zixiang Zhao, Haowen Bai, Jiangshe Zhang, Yulun Zhang, Shuang Xu, Zudi Lin, Radu Timofte, Luc Van Gool

简介：2023 年 4 月发表于**CVPR-2023**的多模态图像融合 SOTA 论文《CDDFuse: Correlation-Driven Dual-Branch Feature Decomposition for Multi-Modality Image Fusion》，作者来自西安交通大学、ETH Zurich、哈佛大学等机构，代码开源于<https://github.com/Zhaozixiang1228/MMIF-CDDFuse>。

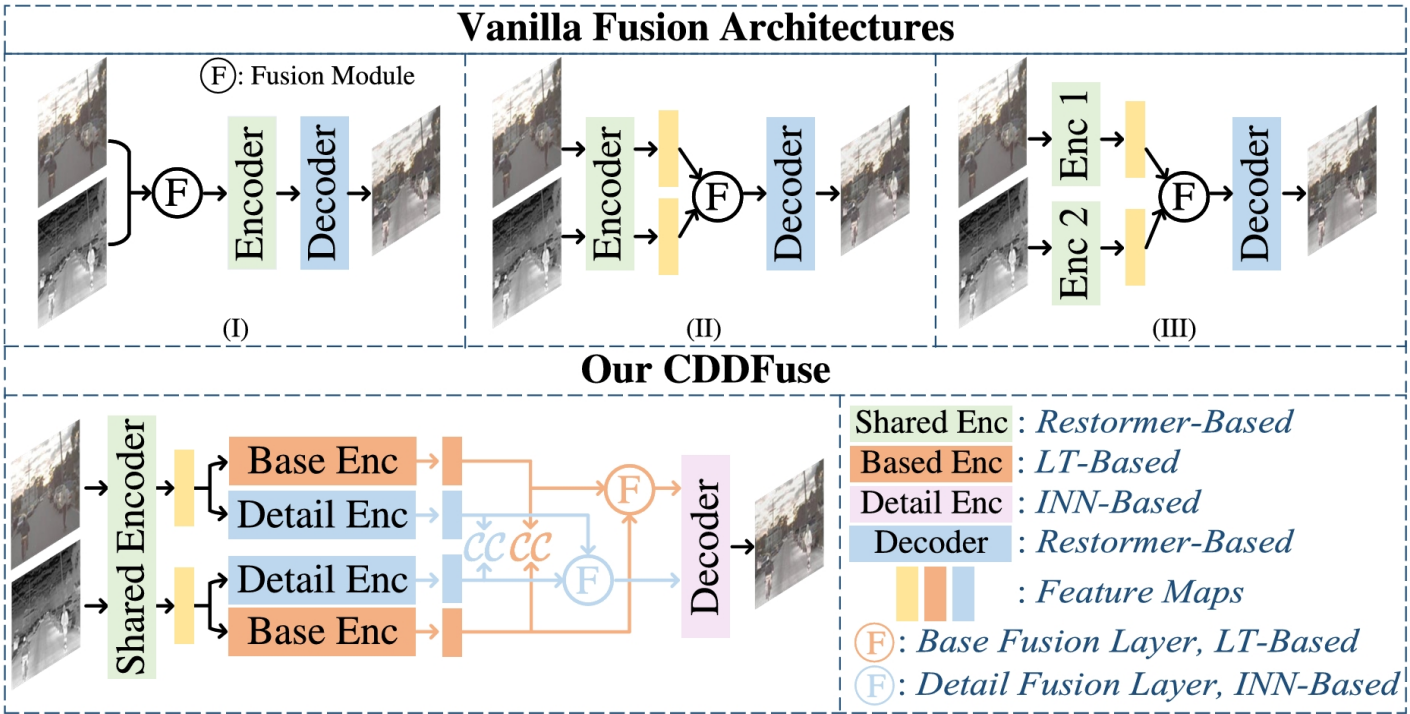
CVPR (IEEE/CVF Conference on Computer Vision and Pattern Recognition) 是**计算机视觉领域三大顶会之首**，也是全球 AI 领域影响力最大的学术会议之一。

一、研究背景

多模态图像融合 (MMIF) 旨在整合不同传感器的互补信息，生成兼具多模态优势的融合图像，核心任务包括**红外-可见光融合**和**医学图像融合**。现有主流方法多采用 CNN 自编码器架构，但存在三大固有缺陷：

- 1. CNN 的黑箱统计拟合特性导致无法有效解耦**模态共享特征**与**模态特有特征**；
- 2. CNN 的局部感受野限制了长距离依赖建模，难以提取全局结构信息；
- 3. CNN 前向传播过程中高频细节信息易丢失。

论文提出核心假设：不同模态的**低频特征**对应**模态共享信息**，具有高相关性；**高频特征**对应**模态特有信息**，具有**低相关性**，以此为基础设计了可控可解释的特征分解融合框架。



(a) Existing MMIF methods vs. CDDFuse. The *base* and *detail* encoders are responsible for extracting global and local features, respectively.

一、研究背景

创新点

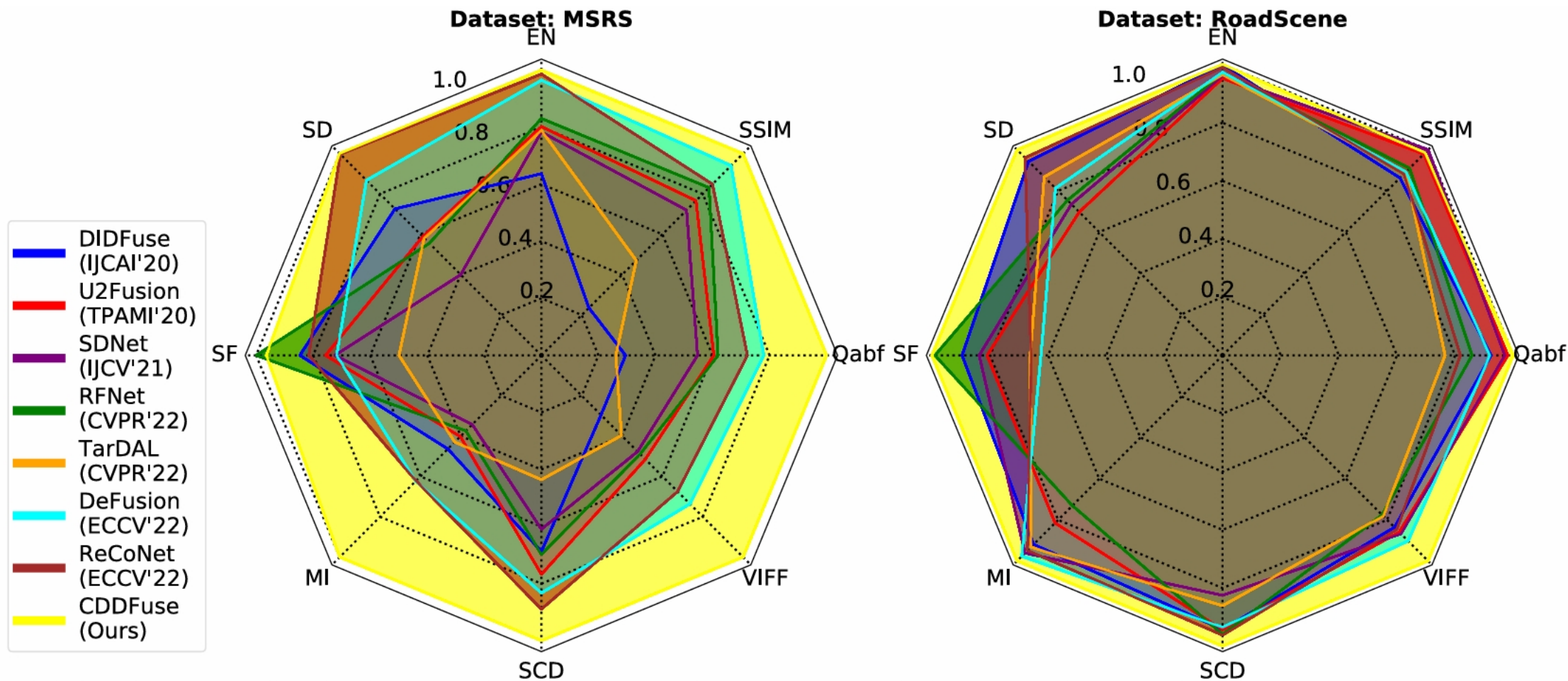
- 1、双分支Transformer-CNN框架：**提出了一个结合了Transformer和CNN的双分支框架，用于提取和融合全局和局部特征，更好地反映了不同的模态特定和模态共享特征。
- 2、特征提取和分解：**通过Restormer块、Lite Transformer (LT) 块和可逆神经网络 (INN) 块，实现了模态特定和模态共享特征的提取和分解。
- 3、相关性驱动的分解损失函数：**提出了一种相关性驱动的分解损失函数，强制模态共享/特定特征分解，使得跨模态基特征相关联，同时不同模态中的详细高频特征去相关。
- 4、领先的图像融合性能：**在红外-可见光图像融合和医学图像融合任务中取得了领先的性能，并提出了统一的测量基准来证明融合图像如何促进下游多模态目标检测和语义分割任务。

不足点

特殊环境下的增加训练的复杂性和时间成本。

一、研究背景

性能雷达图对比



(b) CDDFuse (highlighted in yellow) achieves the state-of-the-art performance on MSRS [57] and RoadScene [71] in eight metrics.

二、核心方法

1、整体架构

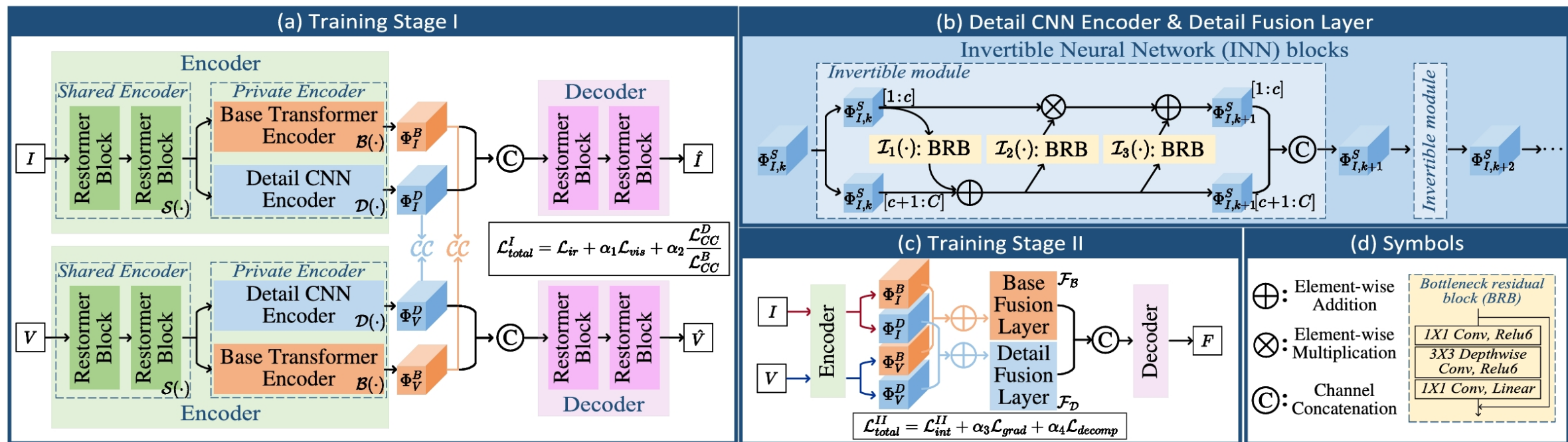


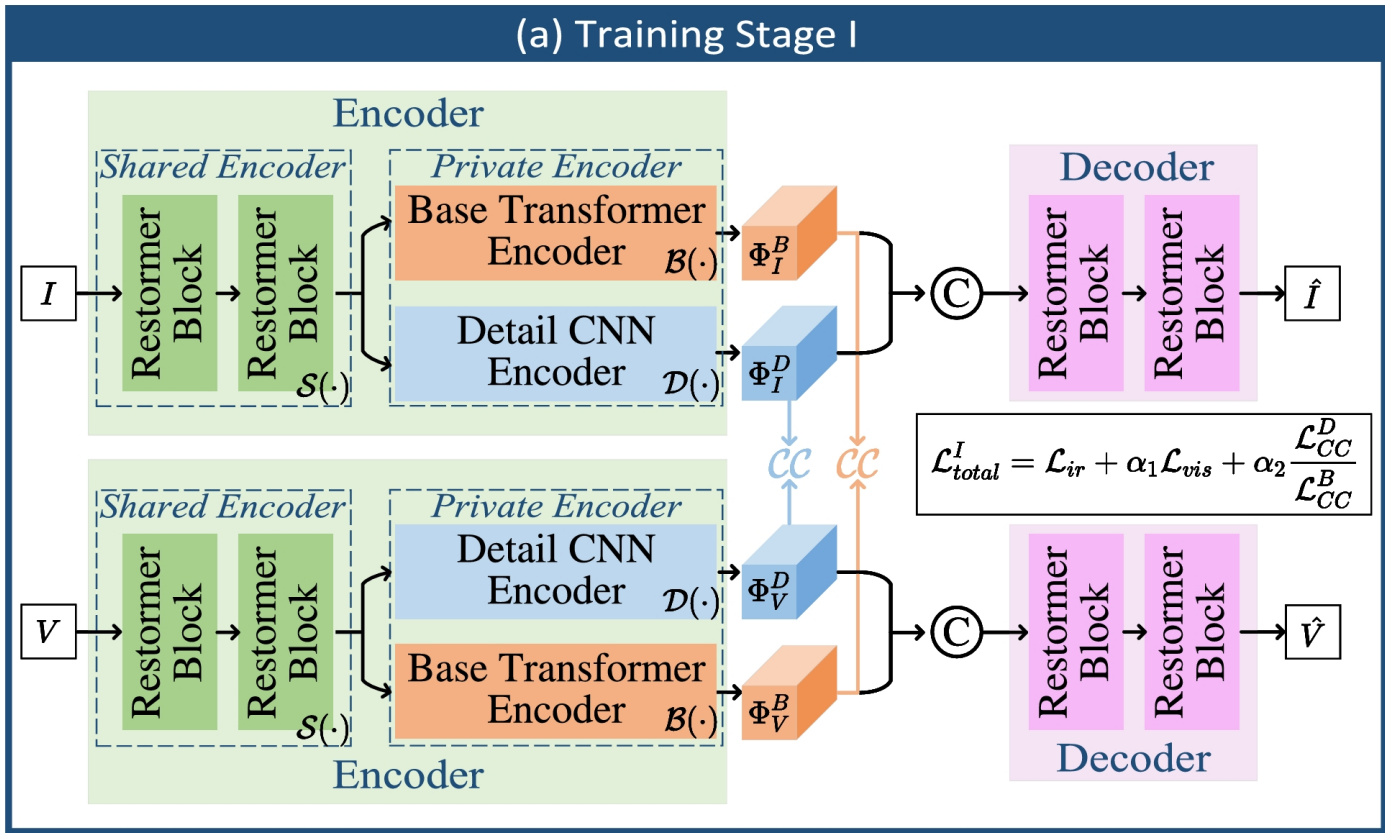
Figure 2. The architecture of our CDDFuse method (IVF as an example). **(a)** The pipeline in training stage I, which aims to train an AE structure for base/detail feature decomposition and reconstructing source images. **(b)** The INN block-based DCE and detail fusion layer, and the BRB block in the affine coupling layers of INN. **(c)** The pipeline in training stage II, which aims to obtain fusion images.

二、核心方法

2、共享特征编码器（SFE）

$$\Phi_I^S = \mathcal{S}(I), \Phi_V^S = \mathcal{S}(V).$$

SFE的目标是从红外和可见光输入图像中提取浅层特征。



使用ResTormer块是因为其能够在高分辨率输入图像中通过跨特征维度的自注意力机制提取全局特征，从而在不显著增加计算量的情况下提取跨模态的浅层特征。

具体地，SFE从红外图像I和可见光图像V中提取浅层特征 Φ_I^S 和 Φ_V^S

二、核心方法

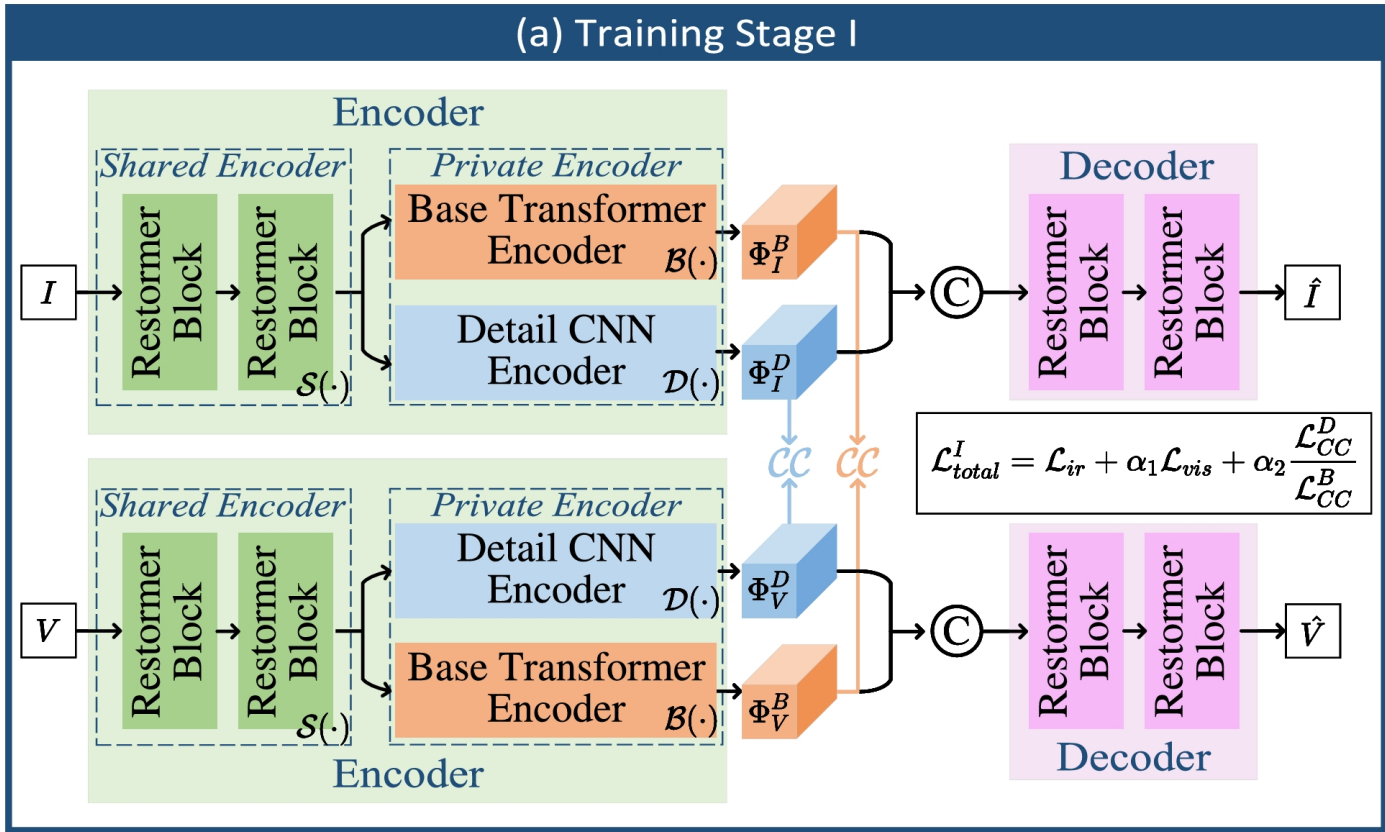
3、基础变换编码器（BTE）

$$\Phi_I^B = \mathcal{B}(\Phi_I^S), \Phi_V^B = \mathcal{B}(\Phi_V^S).$$

BTE负责从共享特征中提取低频基特征。

为了提取长距离依赖特征，使用了空间自注意力机制的Transformer，考虑到性能和计算效率的平衡，选择了LT块作为BTE的基本单元。

通过Flattened feed-forward网络结构，LT块在减少参数数量的同时保持了相同的性能。



二、核心方法

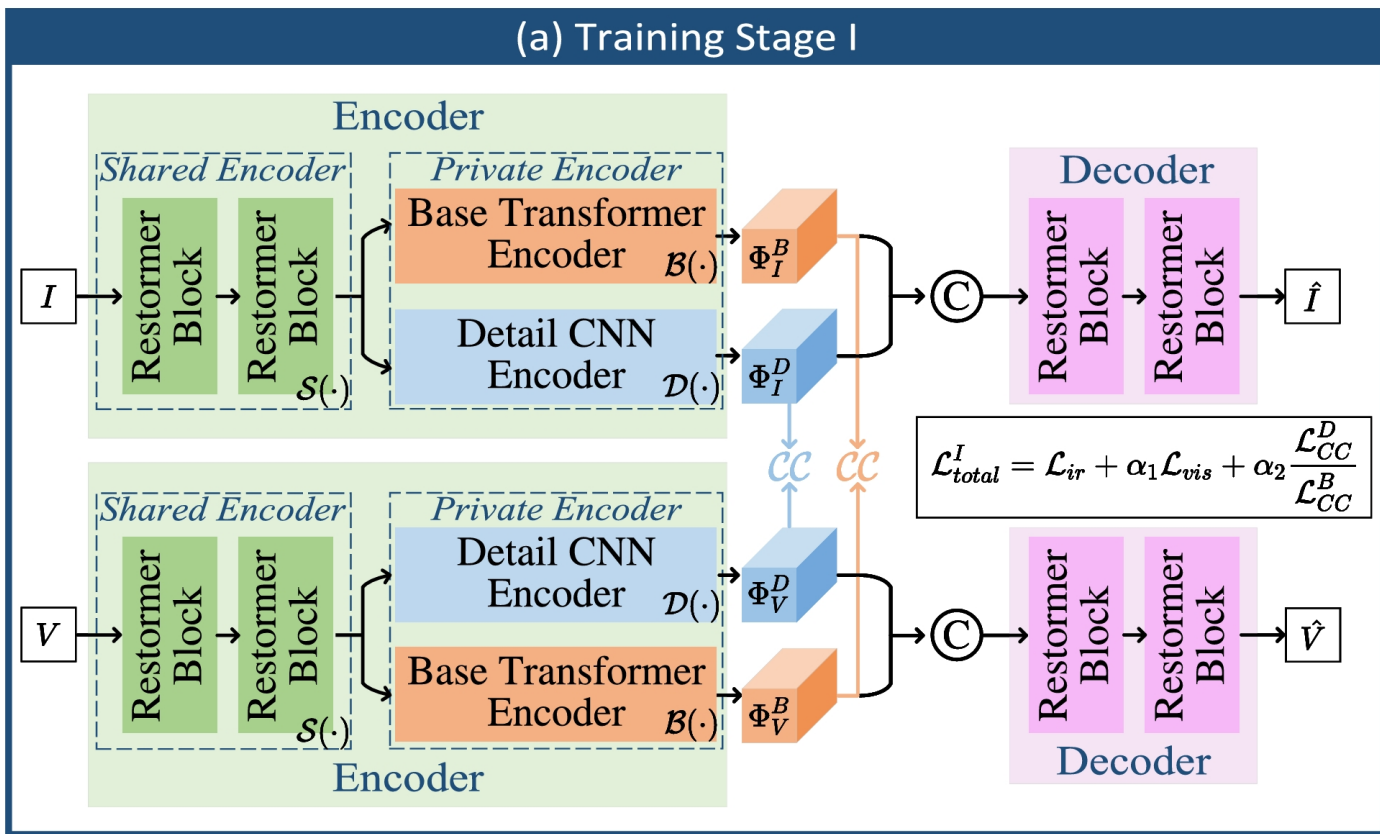
4、细节CNN编码器（DCE）

$$\Phi_I^D = \mathcal{D}(\Phi_I^S), \Phi_V^D = \mathcal{D}(\Phi_V^S).$$

与BTE不同，DCE负责从共享特征中提取高频细节信息。

考虑到边缘和纹理信息在细节特征中的重要性，DCE中的CNN架构需要尽可能多地保留细节信息。

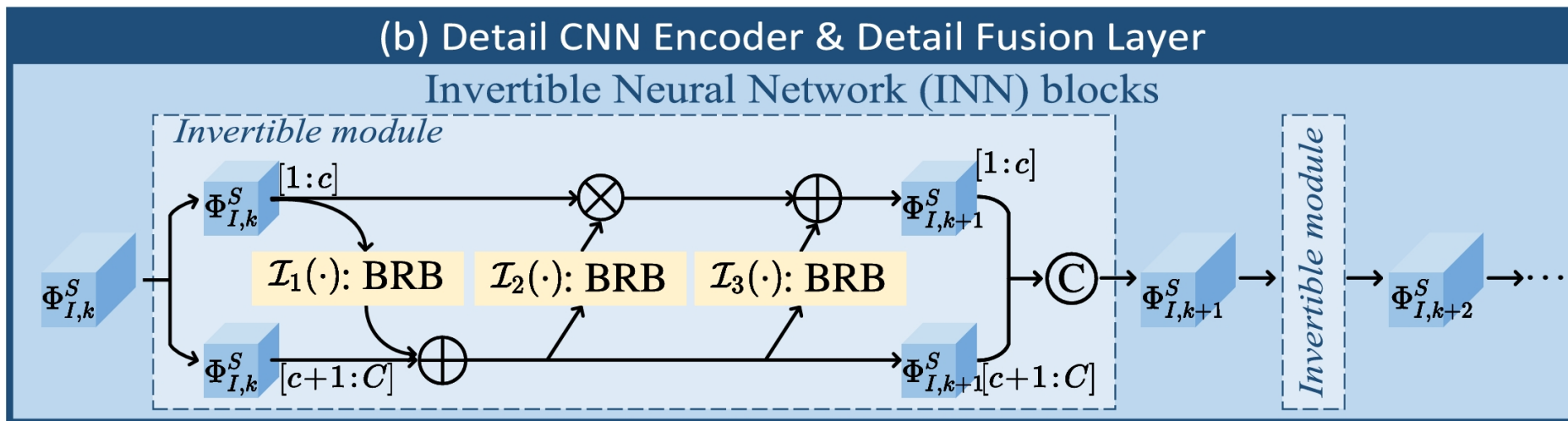
INN模块通过使输入和输出特征相互生成来更好地保留输入信息，因此可以被视为无损特征提取模块，非常适合用于图像融合任务。



二、核心方法

4、细节CNN编码器（DCE）

$$\Phi_I^D = \mathcal{D}(\Phi_I^S), \quad \Phi_V^D = \mathcal{D}(\Phi_V^S).$$



$$\Phi_{I,k+1}^S [c+1:C] = \Phi_{I,k}^S [c+1:C] + \mathcal{I}_1(\Phi_{I,k}^S [1:c]),$$

$$\begin{aligned} \Phi_{I,k+1}^S [1:c] &= \Phi_{I,k}^S [1:c] \odot \exp(\mathcal{I}_2(\Phi_{I,k+1}^S [c+1:C])) \\ &\quad + \mathcal{I}_3(\Phi_{I,k+1}^S [c+1:C]), \end{aligned}$$

$$\Phi_{I,k+1}^S = \text{CAT}\{\Phi_{I,k+1}^S [1:c], \Phi_{I,k+1}^S [c+1:C]\}$$

因此，文章采用了带有仿射耦合层的INN块，并且在每个可逆层中，变换如下：

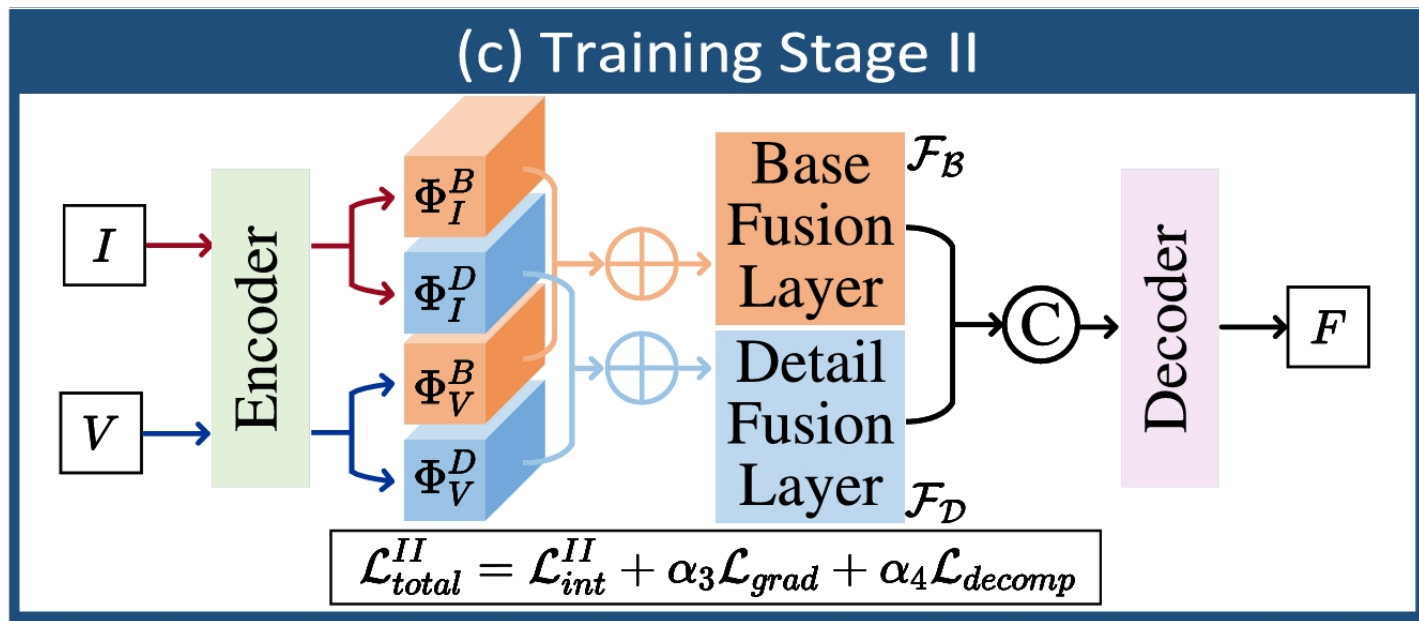
二、核心方法

5、融合层

基特征融合层 (FB) 负责将来自不同模态的基特征进行融合。

细节特征融合 (FD) 负责将来自不同模态的细节特征进行融合。

$$\Phi^B = \mathcal{F}_B (\Phi_I^B, \Phi_I^B), \Phi^D = \mathcal{F}_D (\Phi_I^D, \Phi_V^D),$$



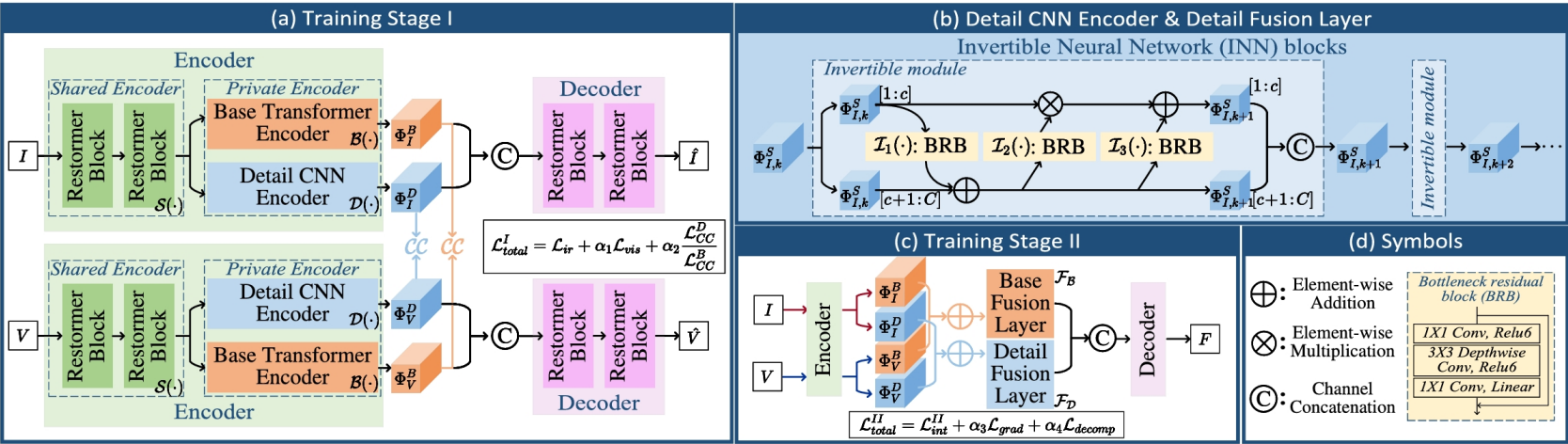
二、核心方法

6、解码器

训练阶段I:
解码器接收来自红外和可见光图像的基特征 (Φ^B_I 和 Φ^B_V) 和细节特征 (Φ^D_I 和 Φ^D_V) ， 并重建对应的原始红外和可见光图像

训练阶段II:
解码器接收融合后的基特征 (Φ^B) 和细节特征 (Φ^D) ， 并输出最终的融合图像。

Stage I: $\hat{I} = \mathcal{DC} \left(\Phi^B_I, \Phi^D_I \right), \hat{V} = \mathcal{DC} \left(\Phi^B_V, \Phi^D_V \right);$
Stage II: $F = \mathcal{DC} \left(\Phi^B, \Phi^D \right).$



二、核心方法

7、两阶段训练策略

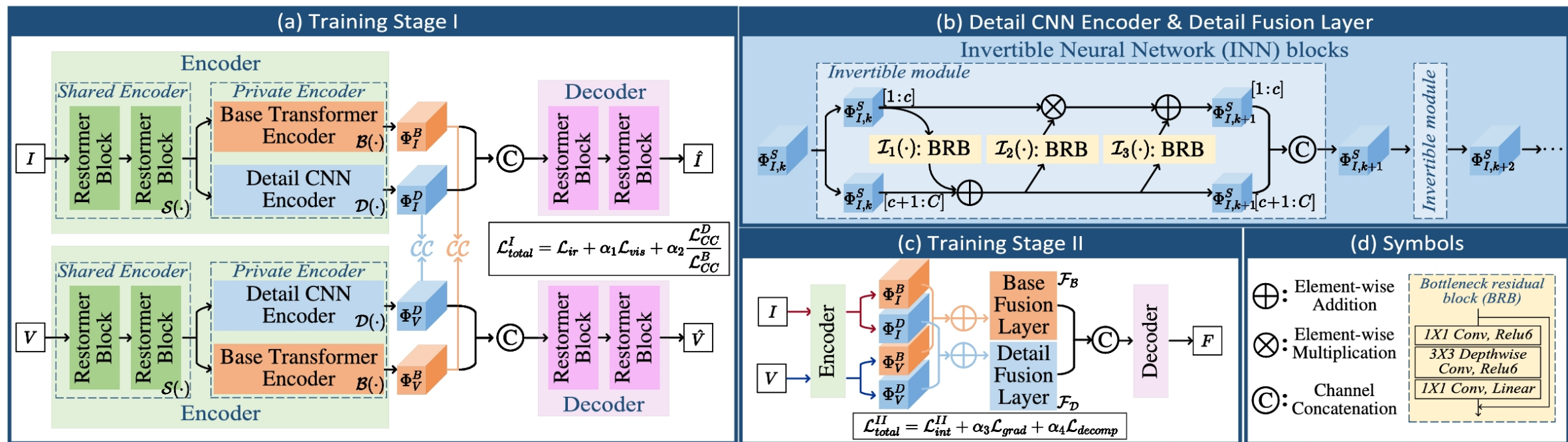


Figure 2. The architecture of our CDDFuse method (IVF as an example). **(a)** The pipeline in training stage I, which aims to train an AE structure for base/detail feature decomposition and reconstructing source images. **(b)** The INN block-based DCE and detail fusion layer, and the BRB block in the affine coupling layers of INN. **(c)** The pipeline in training stage II, which aims to obtain fusion images.

二、核心方法

8、损失

在训练阶段I，总损失 $\mathcal{L}_{total}^I$ 由红外和可见光图像的重建损失（ $\mathcal{L}_{ir}$ 和 $\mathcal{L}_{vis}$ ）、特征分解损失（ $\mathcal{L}_{decomp}$ ）以及调整参数（ α_1 和 α_2 ）组成：

$$\mathcal{L}_{total}^I = \mathcal{L}_{ir} + \alpha_1 \mathcal{L}_{vis} + \alpha_2 \mathcal{L}_{decomp},$$

在训练阶段II，总损失 $\mathcal{L}_{total}^{II}$ 包括内部损失（ $\mathcal{L}_{int}$ ）、梯度损失（ $\mathcal{L}_{grad}$ ）和特征分解损失（ $\mathcal{L}_{decomp}$ ），以及调整参数（ α_3 和 α_4 ）：

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

三、实验分析

1、红外 - 可见光融合

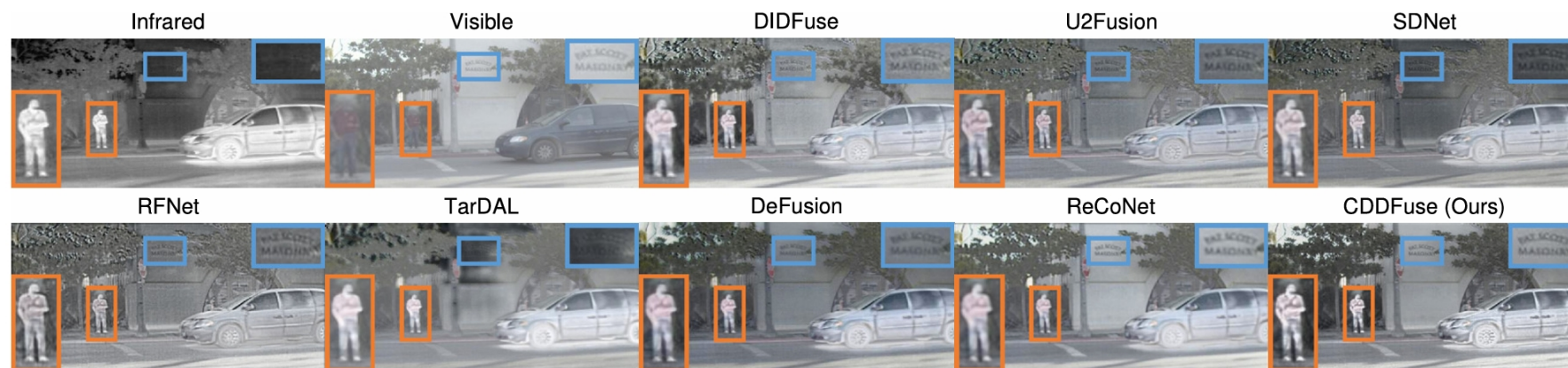


Figure 3. Visual comparison for “FLIR_04602” in RoadScene IVF dataset.



Figure 4. Visual comparison for “00706N” in MSRS IVF dataset.

三、实验分析

2、消融实验

消融实验有三部分。第一部分是分解损失，将base和detail损失由相除变为相减，另一个是直接去除分解损失。编码器中的LT和INN块也进行了消融实验，分别是将BTE中的LT换为INN、DCE中的INN换为LT、DCE中的INN换为CNN。最后一个是两阶段训练的实验，也就是直接采用一阶段训练法，用第二阶段的融合图像作为约束训练网络，可以看到效果差了很多。

Table 2. Ablation experiment results in the testset of MSRS. **Bold** indicates the best value.

Configurations		EN	SD	VIF	SSIM
I	Division $\rightarrow$ Subtraction in $\mathcal{L}_{decomp}$	6.55	42.20	0.98	1.00
II	w/o $\mathcal{L}_{decomp}$	6.19	36.49	0.96	0.97
III	LT block $\rightarrow$ INN in BTE	6.47	41.39	1.00	0.98
IV	INN $\rightarrow$ LT block in DCE	6.56	42.18	1.00	0.99
V	INN $\rightarrow$ CNN block in DCE	6.54	42.10	0.98	0.98
VI	w/o two-stage training	6.28	38.42	0.97	0.99
Ours		6.70	43.38	1.05	1.00

三、实验分析

3、医学图像融合

Table 5. Quantitative results of the MIF task. **Boldface** and underline show the best and second-best values, respectively. CDDFuse* represents the results after training on MIF datasets.

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	<u>33.42</u>	1.98	0.58	0.33	<u>0.52</u>	0.49
DeF [32]	4.63	66.38	21.56	<u>2.20</u>	1.12	<u>0.47</u>	0.44	<u>1.29</u>
ReC [19]	4.41	<u>66.96</u>	20.16	2.03	<u>1.24</u>	0.40	0.42	1.29
CDDFuse	<u>4.83</u>	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	<u>29.41</u>	2.14	0.97	0.38	0.47	0.51
EMF [69]	4.76	<u>72.76</u>	22.56	<u>2.34</u>	<u>1.32</u>	<u>0.56</u>	<u>0.49</u>	<u>1.31</u>
CDDFuse*	<u>4.88</u>	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	<u>0.58</u>	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	<u>1.74</u>	1.48	<u>0.58</u>	0.56	<u>1.45</u>
ReC [19]	3.66	<u>65.25</u>	21.72	1.51	<u>1.49</u>	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	<u>61.40</u>	31.97	1.71	<u>1.40</u>	0.47	0.57	1.46
EMF [69]	4.21	56.80	26.01	1.82	1.31	<u>0.62</u>	0.67	1.47
CDDFuse*	4.23	70.73	<u>29.57</u>	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	<u>0.64</u>	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	<u>1.80</u>	1.27	0.61	0.56	1.46
ReC [19]	3.22	<u>60.07</u>	17.40	1.50	<u>1.47</u>	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	<u>52.97</u>	19.58	1.68	<u>1.28</u>	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]	<u>3.74</u>	51.93	17.14	1.88	1.12	<u>0.71</u>	<u>0.74</u>	1.49
CDDFuse*	3.90	58.31	<u>20.87</u>	2.49	1.35	0.97	0.78	<u>1.48</u>

CDDFuse 即使不微调医学数据，也能超过大部分专门做医学融合的方法，泛化能力极强。

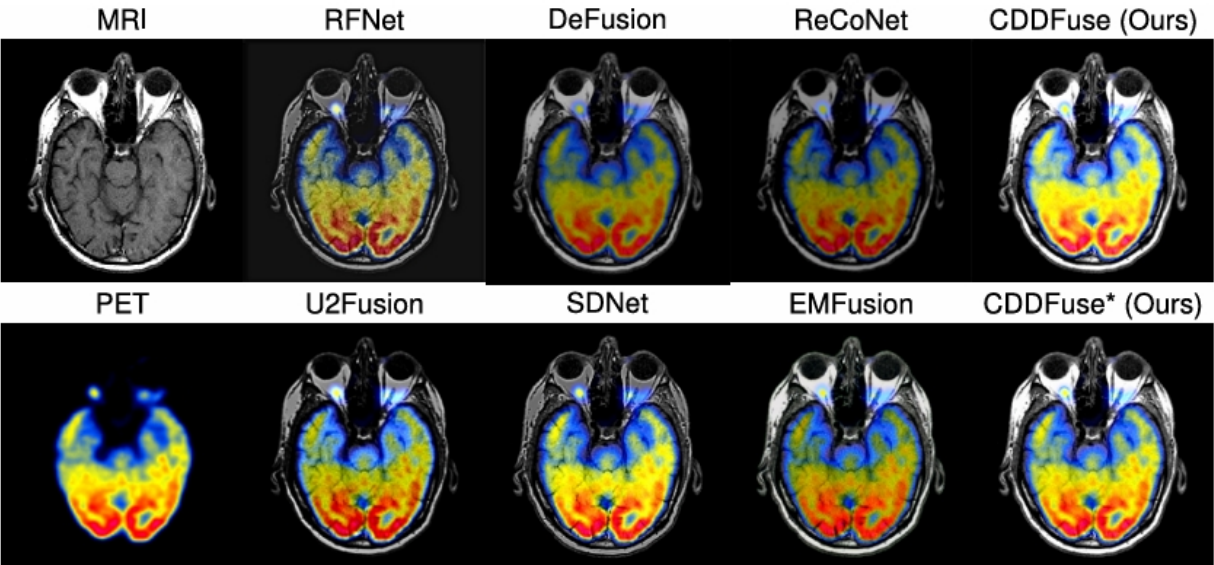


Figure 6. Visual comparison for “MRI-PET-16” in MRI-PET MIF.

四、结论

本文提出一种用于多模态图像融合的双分支 Transformer-CNN 架构。借助 Restormer、轻量级 Transformer 和可逆神经网络模块，模型能更好地提取模态专属特征与模态共享特征，且所提出的相关性驱动分解损失让这些特征的分解过程更直观、高效。实验验证了所提 CDDFuse 模型的融合效果，同时也能提升下游多模态模式识别任务的精度。

谢 谢