

轻量化多模态医学图像融合

MACTFusion: Lightweight Cross Transformer for Adaptive Multimodal Medical Image Fusion

Xinyu Xie , Xiaozhi Zhang , Xinglong Tang , Jiaxi Zhao , Dongping Xiong , Lijun Ouyang ,
Bin Yang , Hong Zhou , Bingo Wing-Kuen Ling , and Kok Lay Teo

<https://github.com/millieXie/MACTFusion>

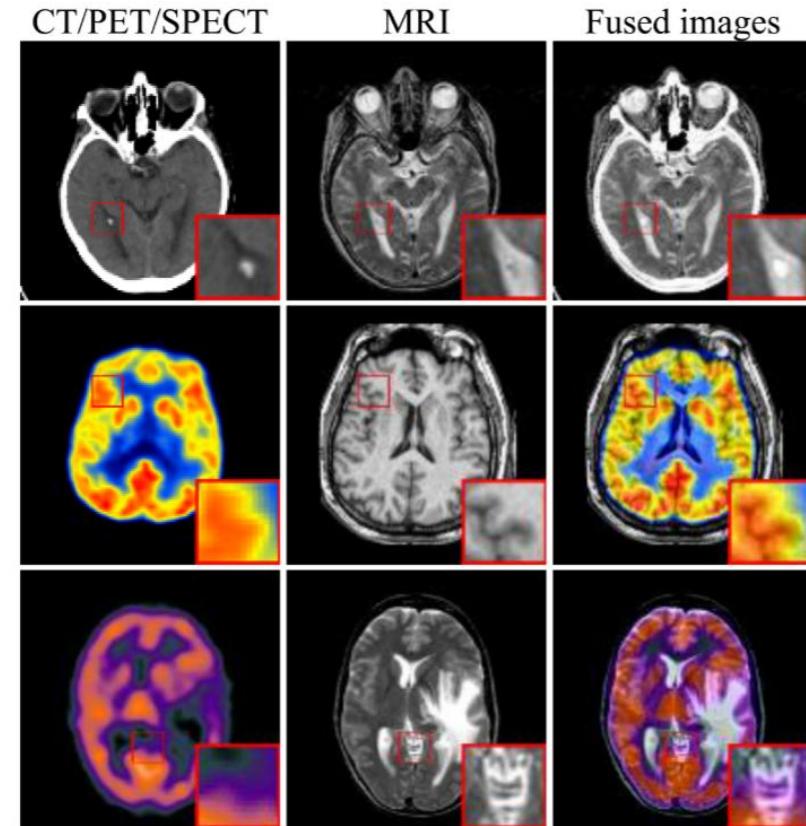


Background

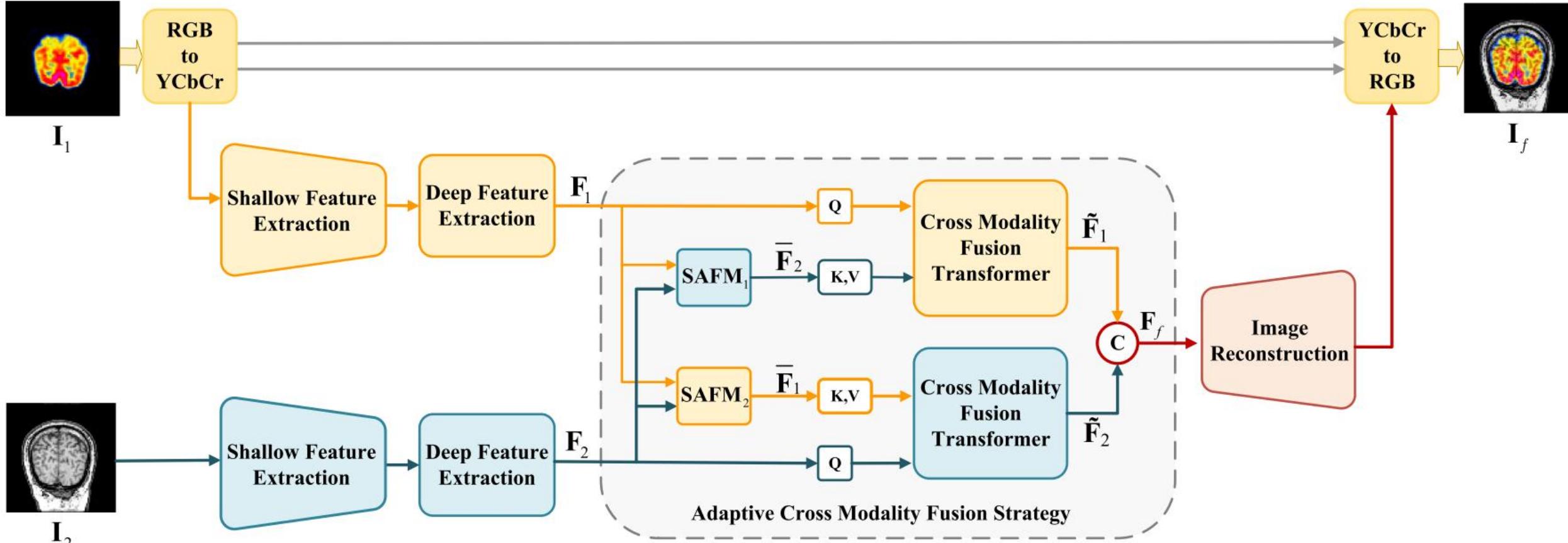
- 多模态医学影像在疾病诊断中至关重要，但不同模态影像各有优缺点。
- 经典的方法存在不可逆数据丢失、细节失真、融合规则设计复杂
- 深度学习的方法会忽略长程依赖交互，导致全局上下文特征丢失
- Transformer的方法存在局部特征提取和跨模态特征交互不足、计算成本高

Contribution

- 提出一种用于**多模态医学图像融合**框架，有效地挖掘和整合有意义的信息，同时提高融合性能，降低计算成本
- 设计一种具有**交叉多轴注意力机制**的轻量化交叉Transformer，用于整合多模态特征的局部和全局交互
- 引入了一种新的**空间自适应融合模块**来自适应重新映射特征分布，同时设计了一个特殊的**特征提取模块**用于获取局部特征和全局特征。



Overview



CT
MRI → PET
→ SPECT

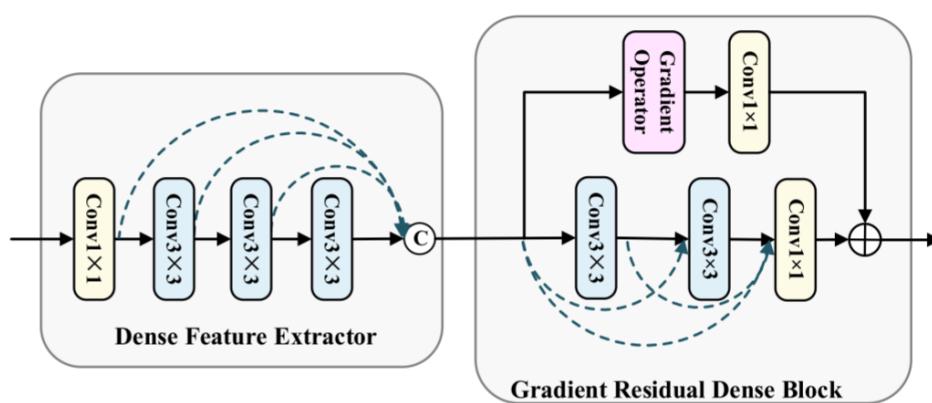
CT可直接与MRI融合

PET和SPECT图像首先需要从RGB转换
为YCbCr, 然后与MRI融合



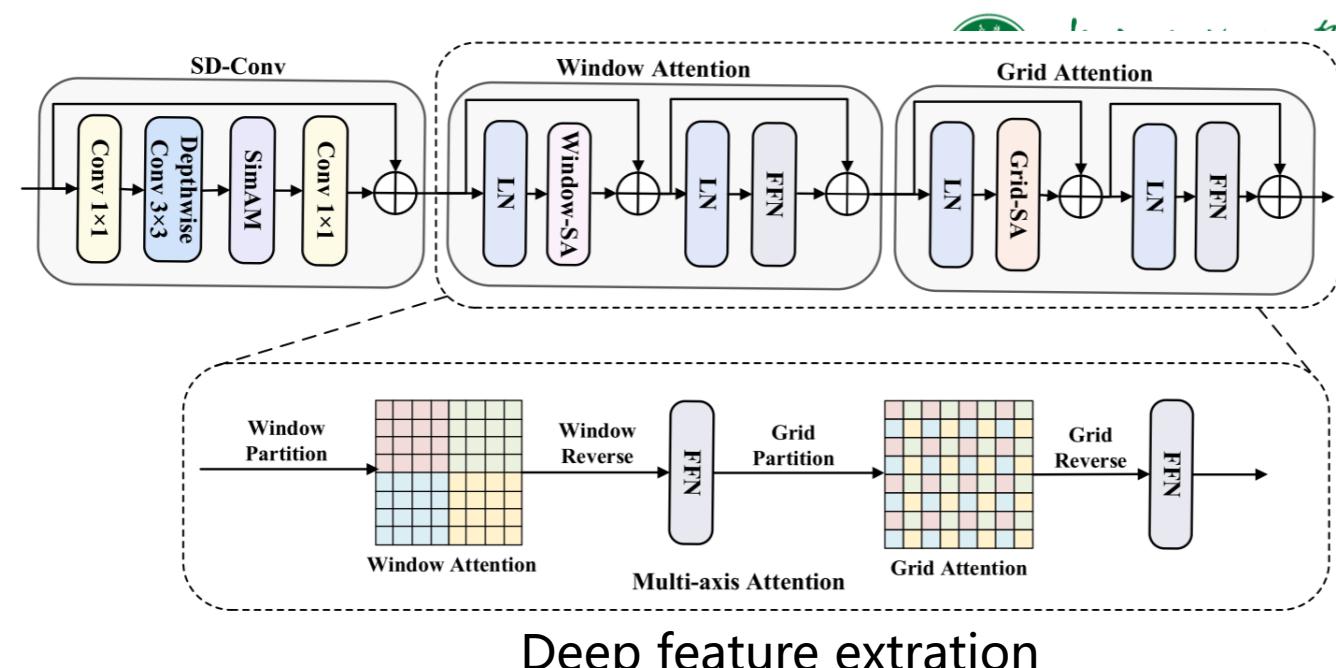
丰富的纹理和适当的代谢信息

Feature Extraction



Shallow feature extraction

DFE通过 1×1 卷积降低通道维度，再用三个级联的 3×3 卷积层提取局部特征并拼接输出，GRDB则通过梯度操作增强对细节的学习



Deep feature extraction

由深度可分离卷积(SD-Conv)和多轴注意力构成。深度可分离卷积可作为位置编码生成器，使模型无需显式的位置编码层；多轴注意力由Window Attention和Grid Attention组成，可降低计算复杂度，同时获取局部和全局交互信息

Adaptive Cross-Modality Fusion Strategy

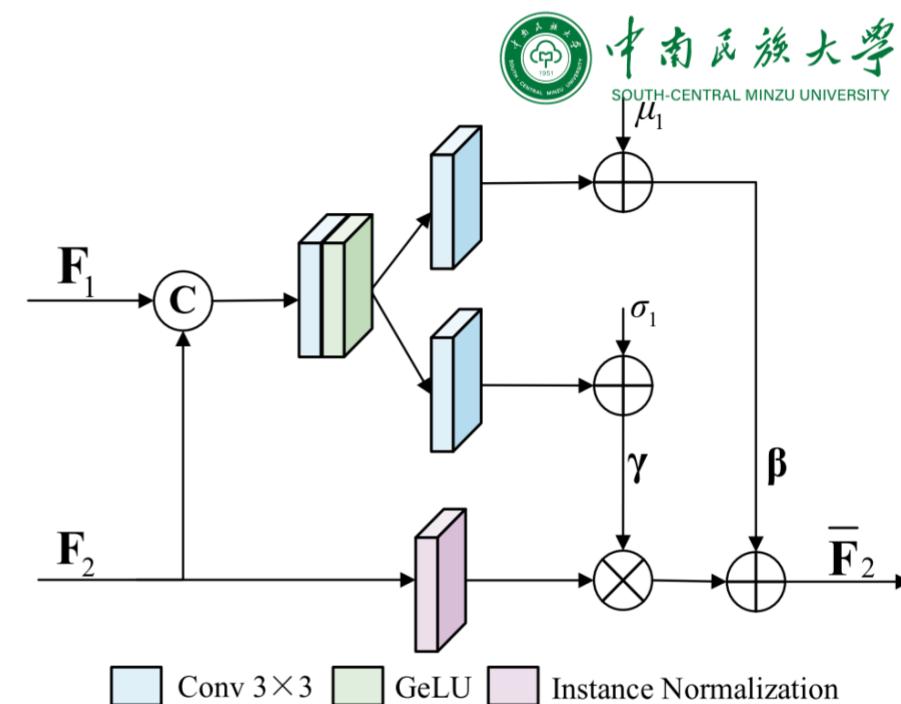
- 由两个平行的对称分支组成，每个分支包括空间自适应融合模块和跨模态融合Transformer。它将特征处理为查询、键和值，并在两个分支之间交换信息，整合互补信息。
- F_2 通过 β 和 γ 的调整，将其分布重新映射到 F_1 的分布。 μ_1 表示 F_1 的通道均值， σ_1 表示标准差。

$$\beta \leftarrow \beta + \mu_1 \quad \gamma \leftarrow \gamma + \sigma_1$$

$$F_2 \leftarrow \frac{F_2 - \mu_2}{\sigma_2}$$

$$\bar{F}_2 = F_2 \odot \gamma + \beta$$

参数会在每一次的迭代时更新



空间自适应融合模块($SAFM_1$)

因此， $SAFM_1$ 帮助 F_2 近似到 F_1 的分布，反之， $SAFM_2$ 帮助 F_1 近似 F_2 的分布

Adaptive Cross-Modality Fusion Strategy

- 首先分别通过两个SD-Conv进行处理，其输出分别作为查询(Q)，键(K)和值(V)特征

$$F_c = \text{CrossWindow Attention} (\text{LN}\{Q, K, V\}) + Q,$$

$$\tilde{Q} = \text{FFN}(\text{LN}(F_c)) + F_c,$$

$$\bar{F}_c = \text{CrossGridAttention} (\text{LN}\{\tilde{Q}, K, V\}) + \tilde{Q},$$

$$\tilde{F}_c = \text{FFN}(\text{LN}(\bar{F}_c)) + \bar{F}_c$$

$$\tilde{F}_1 = \tilde{F}_c \otimes T_1 + F_1$$

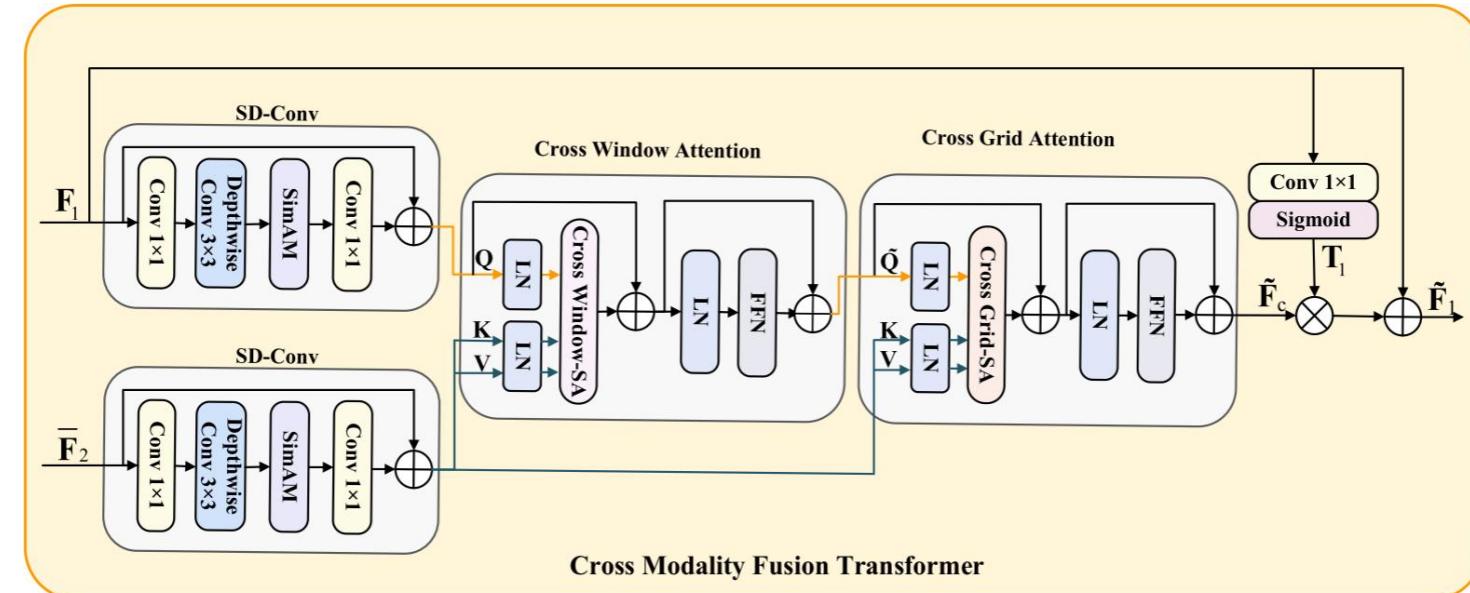
$$T_1 = \text{sigmoid}(F_1)$$

我们将两个输出特征 \tilde{F}_1 和 \tilde{F}_2 连接起来，得到最终的融合特征 F_f

Image Reconstruction

$$I_f = H_R(F_f)$$

$$1 \times \text{Conv}_1 + 2 \times \text{Conv}_3$$



跨模态融合Transformer

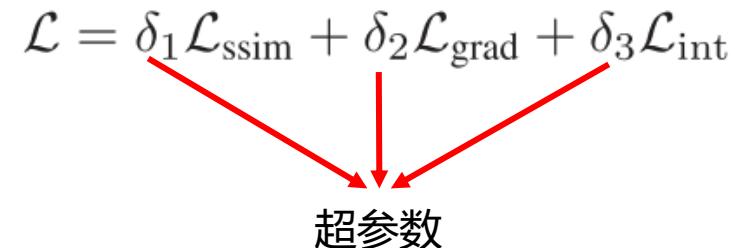
- 结构相似系数(SSIM)损失, 梯度损失和强度损失

$$\begin{aligned}\mathcal{L}_{\text{ssim}} = & \eta \cdot (1 - \text{SSIM}(\mathbf{I}_f, \mathbf{I}_1)) \\ & + (1 - \eta) \cdot (1 - \text{SSIM}(\mathbf{I}_f, \mathbf{I}_2))\end{aligned}$$

$$\mathcal{L}_{\text{grad}} = \frac{1}{HW} \|\|\nabla \mathbf{I}_f| - \max(|\nabla \mathbf{I}_1|, |\nabla \mathbf{I}_2|)\|_1$$

$$\mathcal{L}_{\text{int}} = \frac{1}{HW} \|\mathbf{I}_f - \max(\mathbf{I}_1, \mathbf{I}_2)\|_1$$

- 总损失计算

$$\mathcal{L} = \delta_1 \mathcal{L}_{\text{ssim}} + \delta_2 \mathcal{L}_{\text{grad}} + \delta_3 \mathcal{L}_{\text{int}}$$


Experiment setup

传统方法 基于CNN 基于Transformer

- 对比的方: CSMCA, EMFusion, FusionGAN, IFCNN, U2Fusion , IFT 和SwinFusion.
- 评估指标: 差异相关性总和(SCD)、多尺度结构相似性度量(MS-SSIM), 边缘信息度量($Q^{AB/F}$)、标准差(SD)、加权融合质量度量(Q_w)、视觉信息保真度(VIF)

$$\begin{pmatrix} Y \\ Cb \\ Cr \end{pmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & 0.500 \\ 0.500 & -0.419 & -0.081 \end{bmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} + \begin{pmatrix} 0 \\ 128 \\ 128 \end{pmatrix}$$

➤ CT-MRI融合

Methods	SCD↑	MS-SSIM↑	Q ^{AB/F} ↑	SD↑	Q _w ↑	VIF↑
CSMCA [39]	1.060	0.944	0.599	9.285	0.795	0.317
EMFusion [40]	1.387	0.914	0.512	9.615	0.645	0.416
FusionGAN [20]	0.132	0.723	0.207	8.534	0.414	0.198
IFCNN [15]	1.297	0.868	0.538	9.631	0.563	0.349
U2Fusion [16]	0.840	0.836	0.287	9.097	0.575	0.312
IFT [22]	1.119	0.893	0.457	9.322	0.684	0.378
SwinFusion [1]	1.566	0.954	0.631	9.651	0.798	0.497
MACTFusion (Ours)	1.715	0.959	0.660	9.685	0.809	0.507

➤ PET-MRI融合

Methods	SCD↑	MS-SSIM↑	Q ^{AB/F} ↑	SD↑	Q _w ↑	VIF↑
CSMCA [39]	1.060	0.932	0.507	9.423	0.853	0.507
EMFusion [40]	0.994	0.903	0.736	9.815	0.914	0.455
FusionGAN [20]	0.322	0.628	0.200	7.769	0.458	0.214
IFCNN [15]	1.238	0.932	0.689	9.534	0.900	0.520
U2Fusion [16]	1.004	0.892	0.505	8.847	0.733	0.485
IFT [22]	1.300	0.910	0.523	9.278	0.720	0.529
SwinFusion [1]	1.562	0.923	0.743	9.877	0.922	0.553
MACTFusion (Ours)	1.691	0.931	0.771	9.900	0.931	0.571

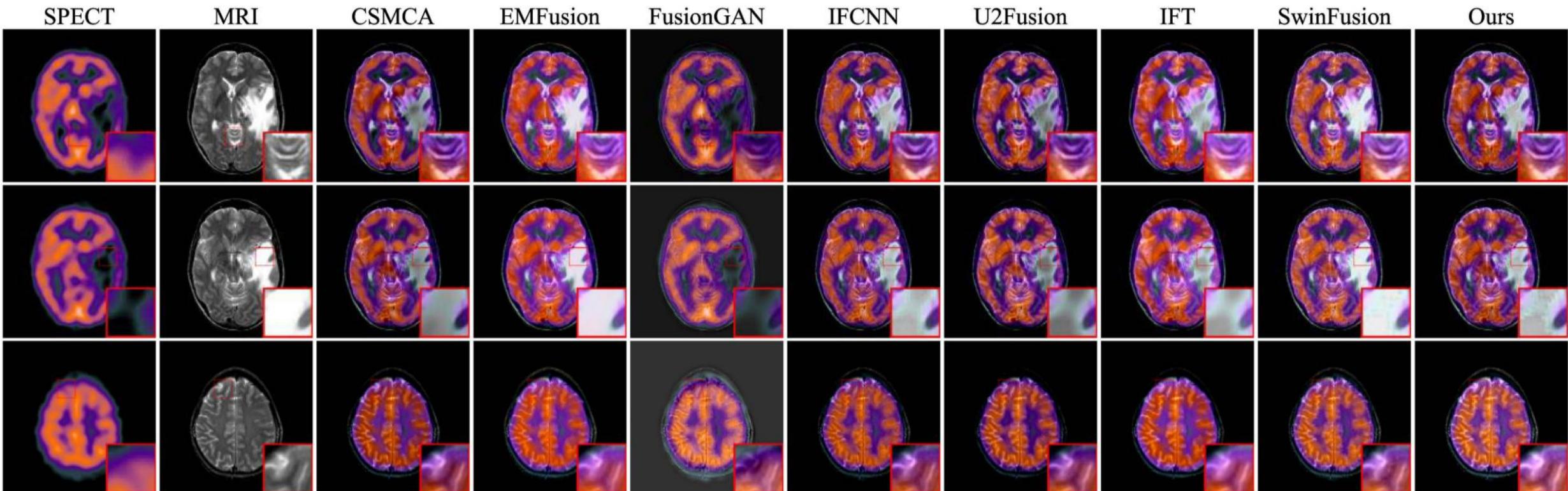
Experiment Results

➤ SPECT-MRI融合

Methods	SCD↑	MS-SSIM↑	Q ^{AB/F} ↑	SD↑	Q _w ↑	VIF↑
CSMCA [39]	0.890	0.931	0.697	8.124	0.841	0.445
EMFusion [40]	1.082	0.947	0.682	8.287	0.857	0.490
FusionGAN [20]	0.190	0.810	0.317	7.431	0.469	0.288
IFCNN [15]	1.220	0.966	0.642	8.257	0.865	0.549
U2Fusion [16]	0.826	0.937	0.532	7.636	0.798	0.544
IFT [22]	1.217	0.929	0.658	8.274	0.811	0.510
SwinFusion [1]	1.631	0.960	0.680	8.393	0.8511	0.644
MACTFusion (Ours)	1.810	0.955	0.699	8.415	0.839	0.842

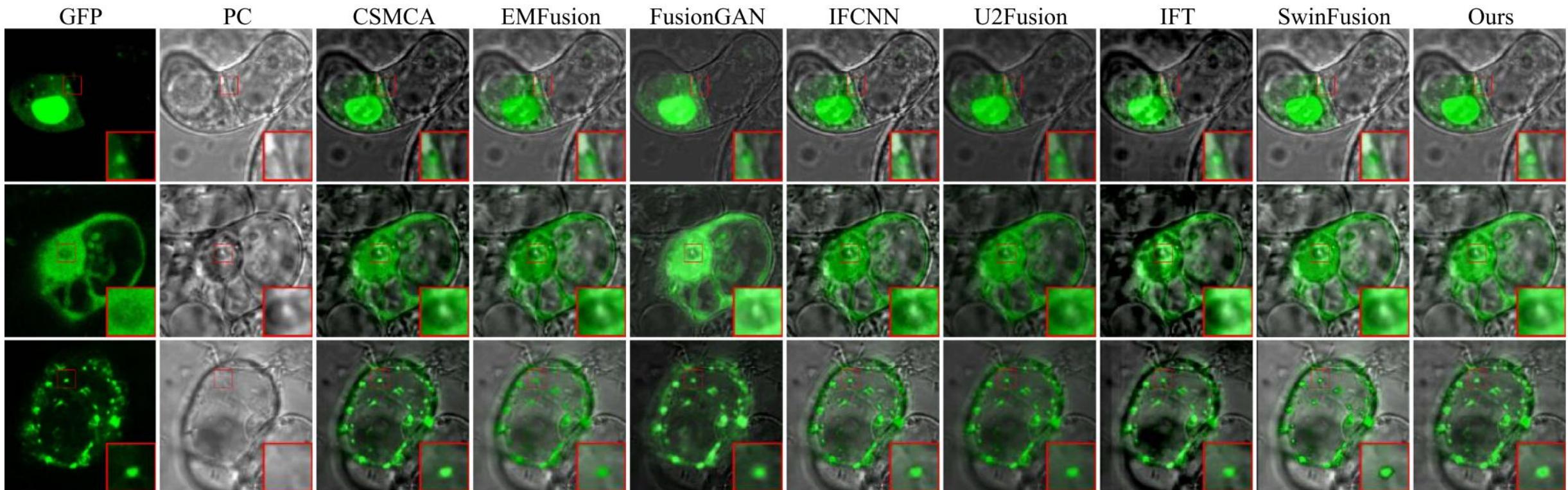
Experiment Results

➤ 可视化



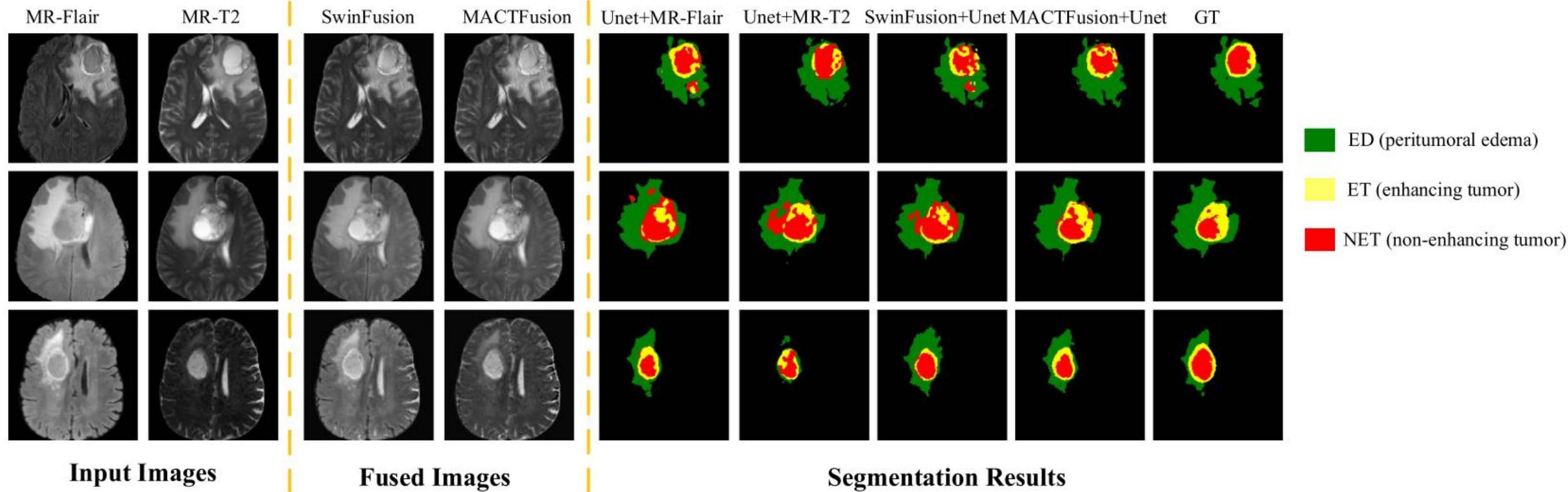
Experiment Results

➤ 绿色荧光蛋白(GFP)与相位对比图像(PC)图像融合



Experiment Results

➤ 下游任务



➤ 网络结构的分析

Shallow Feature Extraction(SFE)

Spatial Adaptation Fusion Module (SAFM)

Cross-Modality Fusion Transformer (CMFT)

a: SAFM+CMFT

b: Element Addition+CMFT

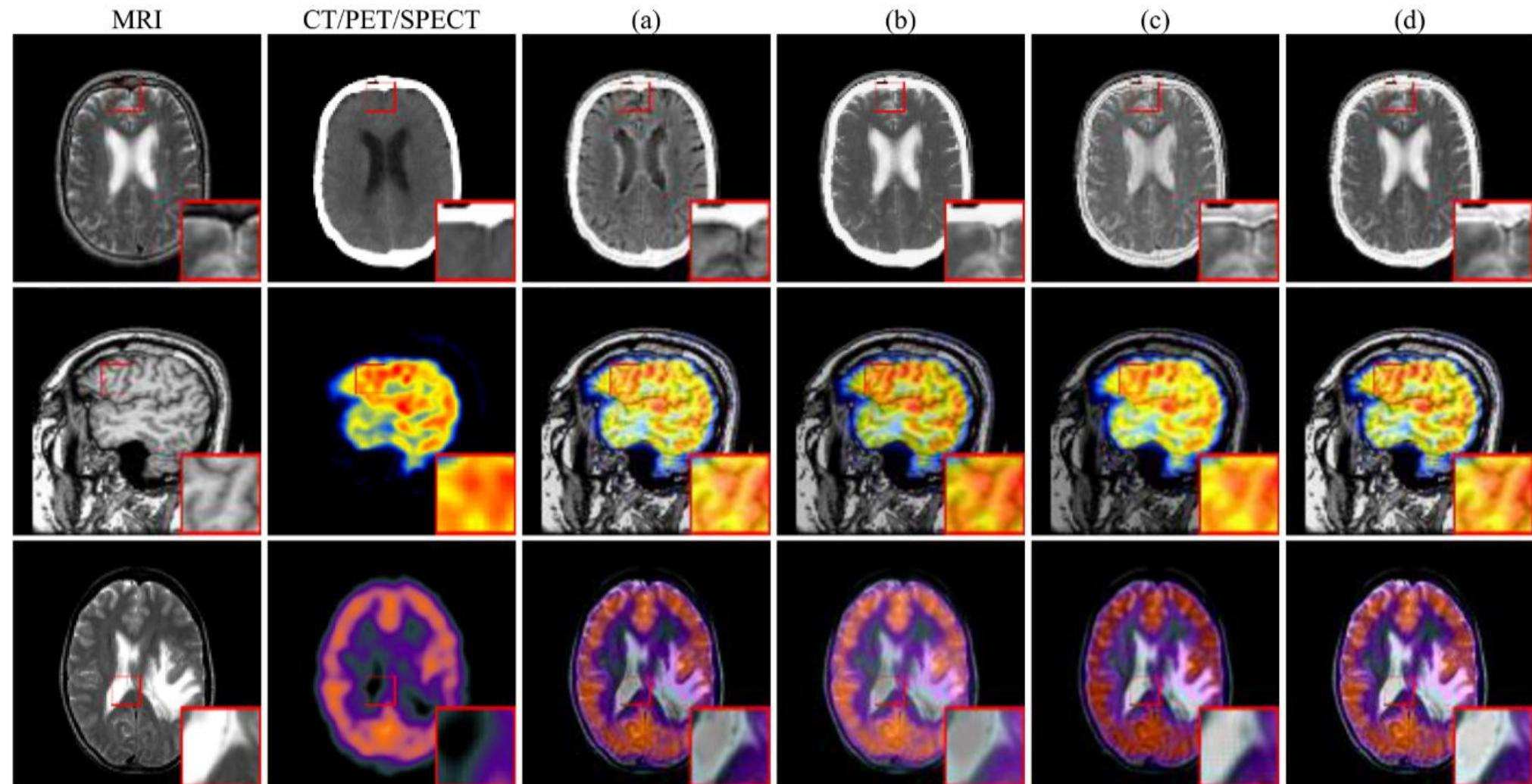
c: SAFM+Generic Fusion Approach

d: MACTFusion

Fusion Tasks		SCD↑	MS-SSIM↑	Q ^{AB/F} ↑	SD↑	Q _w ↑	VIF↑
CT-MRI	a	1.661	0.897	0.601	9.667	0.702	0.493
	b	1.691	0.900	0.637	9.648	0.795	0.497
	c	1.683	0.844	0.634	9.650	0.789	0.491
	d	1.715	0.959	0.661	9.685	0.809	0.507
PET-MRI	a	1.601	0.913	0.706	9.878	0.921	0.561
	b	1.680	0.829	0.760	9.844	0.929	0.501
	c	1.613	0.890	0.764	9.852	0.930	0.534
	d	1.691	0.931	0.771	9.900	0.931	0.571
SPECT-MRI	a	1.529	0.945	0.592	8.398	0.820	0.570
	b	1.687	0.937	0.614	8.295	0.830	0.609
	c	1.584	0.922	0.691	8.288	0.823	0.517
	d	1.810	0.955	0.699	8.415	0.839	0.842

Ablation Study

➤ 网络结构的分析



Ablation Study

➤ Loss分析

$$\mathcal{L} = \delta_1 \mathcal{L}_{\text{ssim}} + \delta_2 \mathcal{L}_{\text{grad}} + \delta_3 \mathcal{L}_{\text{int}}$$

a: w/o ssim loss

b: w/o gradient loss

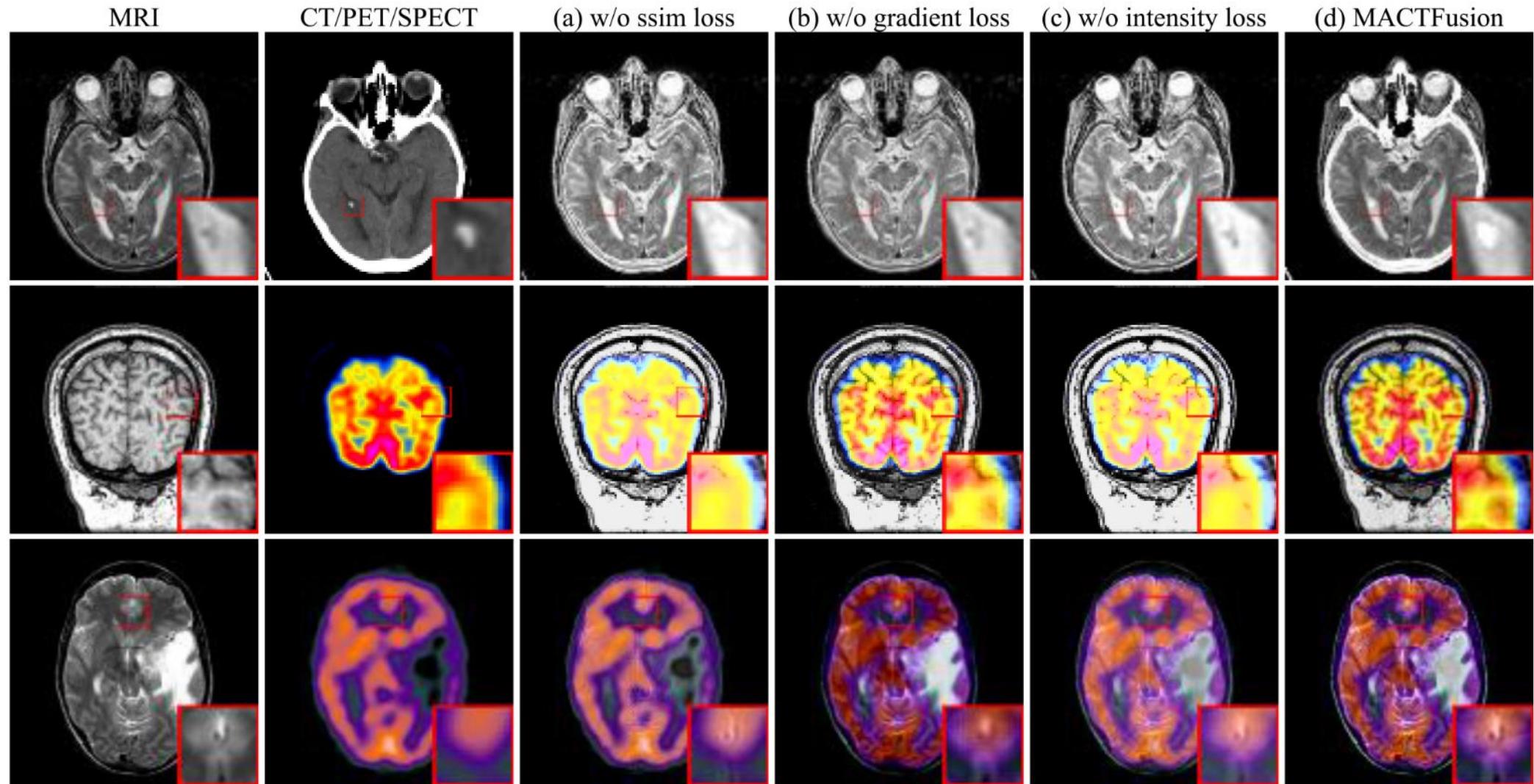
c: w/o intensity losss

d: MACTFusion

Fusion Tasks		SCD↑	MS-SSIM↑	Q ^{AB/F} ↑	SD↑	Q _w ↑	VIF↑
CT-MRI	a	1.479	0.778	0.640	9.574	0.719	0.496
	b	1.570	0.948	0.596	9.640	0.718	0.483
	c	1.686	0.943	0.649	9.556	0.612	0.482
	d	1.715	0.959	0.661	9.685	0.809	0.507
PET-MRI	a	1.530	0.758	0.579	9.382	0.673	0.564
	b	1.681	0.870	0.518	8.920	0.838	0.561
	c	1.535	0.835	0.587	9.374	0.702	0.551
	d	1.691	0.931	0.771	9.900	0.931	0.571
SPECT-MRI	a	1.365	0.760	0.626	8.266	0.809	0.526
	b	1.499	0.863	0.553	8.410	0.793	0.585
	c	1.513	0.918	0.630	8.288	0.619	0.617
	d	1.810	0.955	0.699	8.415	0.839	0.842

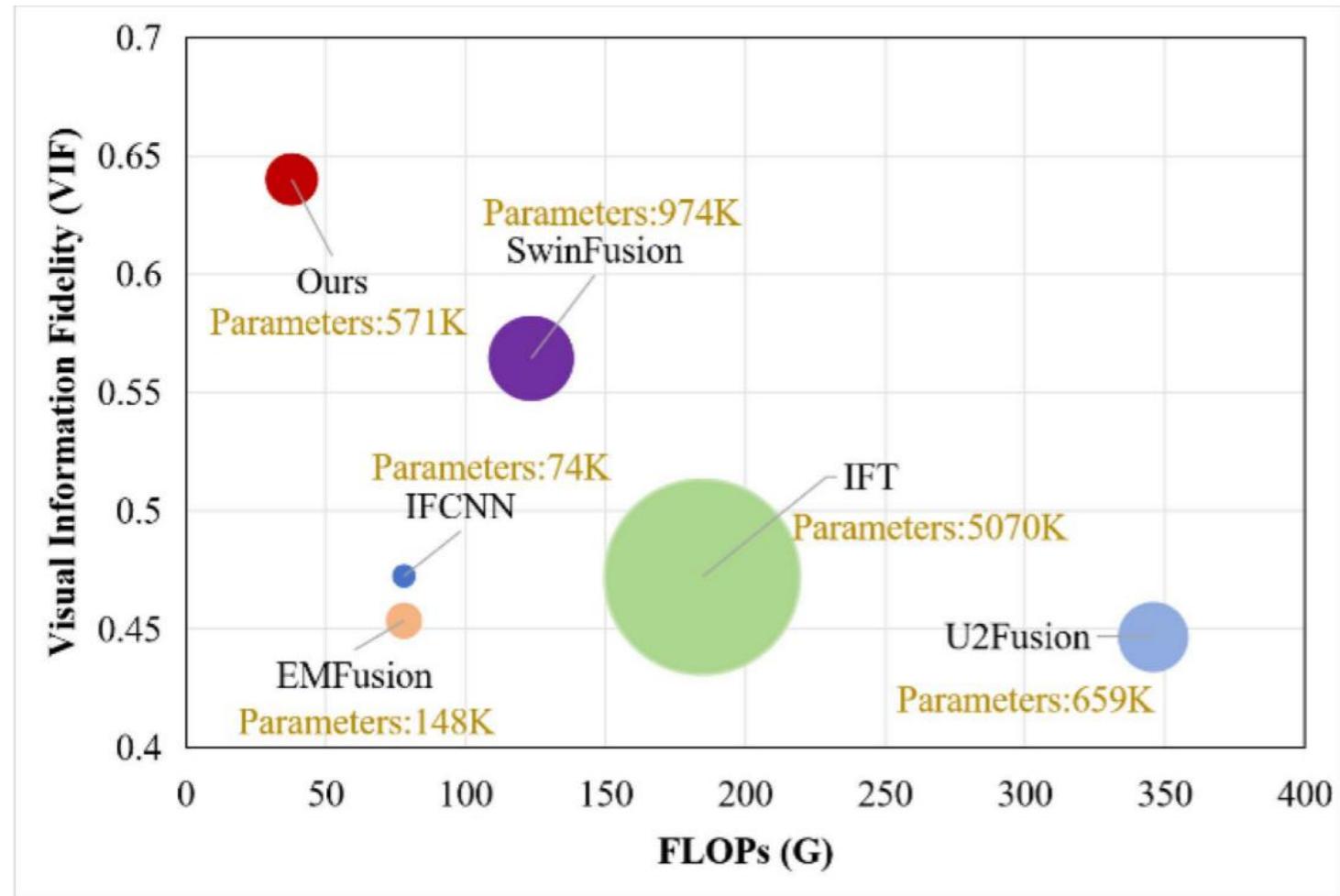
Ablation Study

➤ Loss分析



Computational Complexity Analysis

▶ 参数量分析



本文设计的**交叉多轴注意力机制**能整合不同模态的局部和全局特征，**空间自适应融合模块**引导**交叉Transformer** 聚焦关键信息，**特征提取模块**可有效获取局部和全局特征，MACTFusion在降低计算成本的同时，保持了较好的融合效果。



Future Work

将所提出的MACTFusion应用于临床医学图像融合，并且将MACTFusion扩展到三维图像融合。

Thank you