深度学习中的图像融合

西北工业大学计算机学院
School of Computer Science, Northwestern Polytechnical University

图像融合

由于硬件设备理论和技术上的限制，**单一传感器或单一拍摄设置所拍摄的图像无法有效、全面地描述成像场景。**
而图像融合能够将不同源图像中的有意义信息结合起来，生成单一图像，该图像包含更丰富的信息，更有利于后续应用。
可以作为下游任务,包括目标检测,遥感图像监测,医学图像诊断以及RGB-T图像物体追踪任务等的一种图像增强技术。

不同拍摄设备

Infrared image

visible image

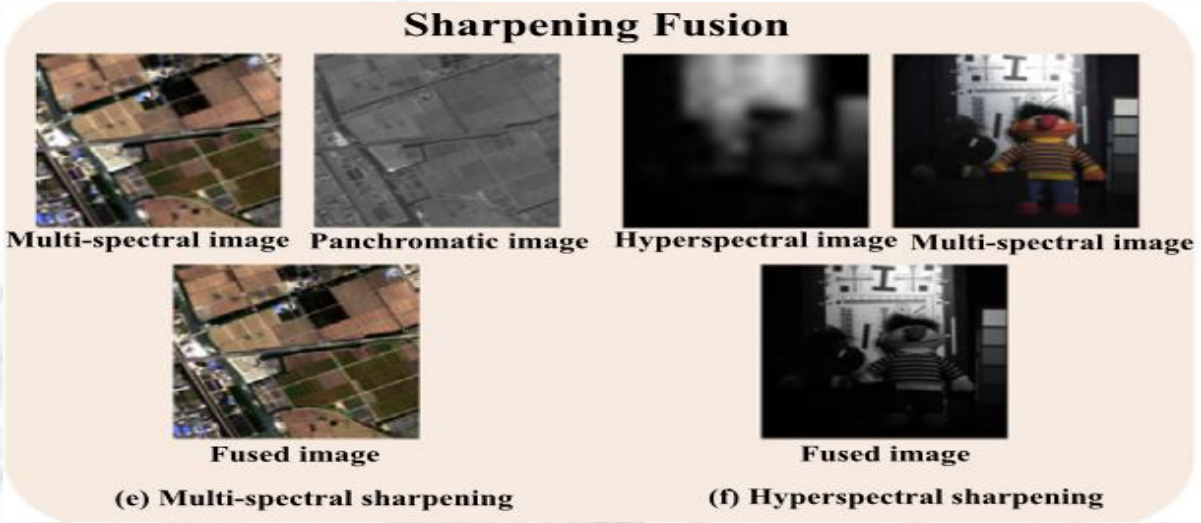
常见图像融合场景

- **数码摄影图像融合**(Digital photography image fusion)

光学镜头受到景深、曝光程度影响,可能出现无法聚焦所有图物体或者过曝等情况。
可以利用图像融合将在不同拍摄设置下拍摄的多幅图像结合起来,生成具有高动态范围的全清晰图像。比如多曝光图像与多聚焦图像融合。

- 锐化融合(Sharpening fusion)

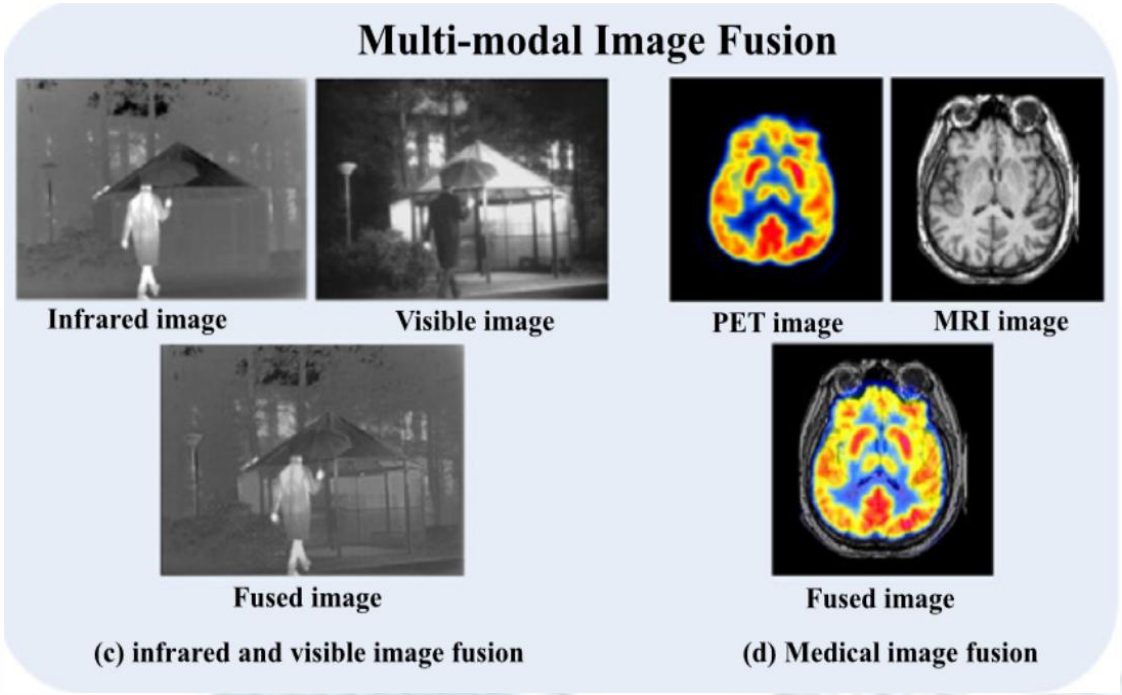
保证信噪比下,图像光谱分辨率和空间分辨率存在矛盾。
锐化融合是克服光谱分辨率和空间分辨率之间矛盾的有效技术。通常包括多光谱锐化和超广谱锐化。



常见图像融合场景

多模态图像融合(Multi-modal image fusion)

不同传感器的成像原理各不相同,其拍摄的多模态图像在描述场景时的侧重点也大相径庭。通过融合不同模态图像中互补和有益的信息，可以更全面地描述成像场景或目标。最具代表性的两种多模态图像任务包括红外和可见光图像融合以及医学图像(PET/MRI)融合。



- **红外-可见光图像融合**(Infrared and visible image fusion)

红外图像具有明显的对比度，即使在恶劣天气下也能从背景中有效地突出目标。
可见光图像包含丰富的纹理细节，更符合人类的视觉感知。红外和可见光图像融合就是要将这两种特性结合起来，生成对比度高、纹理丰富的图像。

- 医学图像融合(Medical image fusion)

医学影像按所代表的信息可分为结构影像和功能影像。比如PET(正电子发射断层扫描)可以描述人体代谢强度,CT反应组织结构。

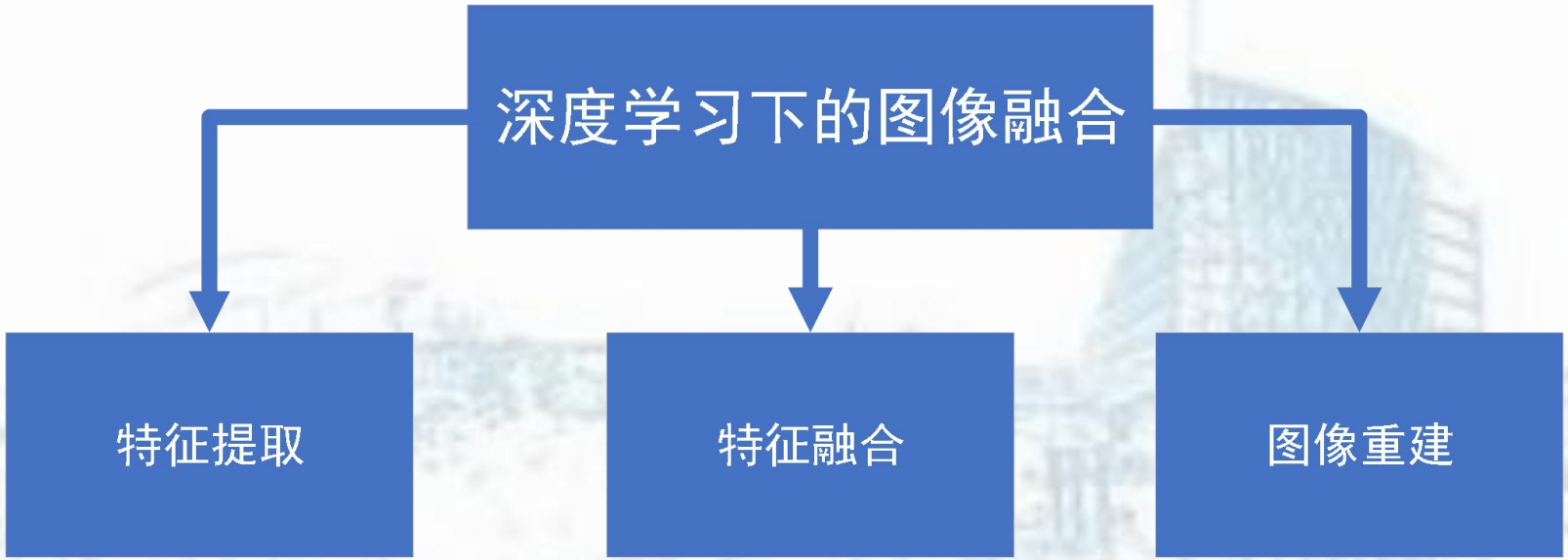
医学影像融合将两种不同类型的医学影像结合在一起，生成具有更丰富信息的单一图像,有利于更准确地诊断疾病。

基于深度学习的图像融合方法

传统的图像融合需要**人工设计融合策略,融合性能有限**；而且常常对不同的源图像采用相同的变换来提取特征，**没有考虑源图像特征差异,可能导致所提取的特征表达能力差。**

而基于深度学习的方法可以**设计不同分支或模块实现差异化特征提取**从而获得更有针对性的特征，此外利用设计好的**损失函数**也能学习到更合理的特征融合策略。

图像融合可以简单地分为三个子问题,不同的深度学习方法可以单独或者同时解决这些问题。

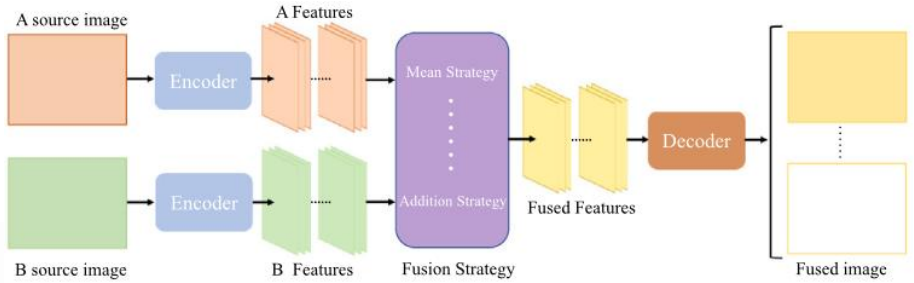


基于深度学习的图像融合方法

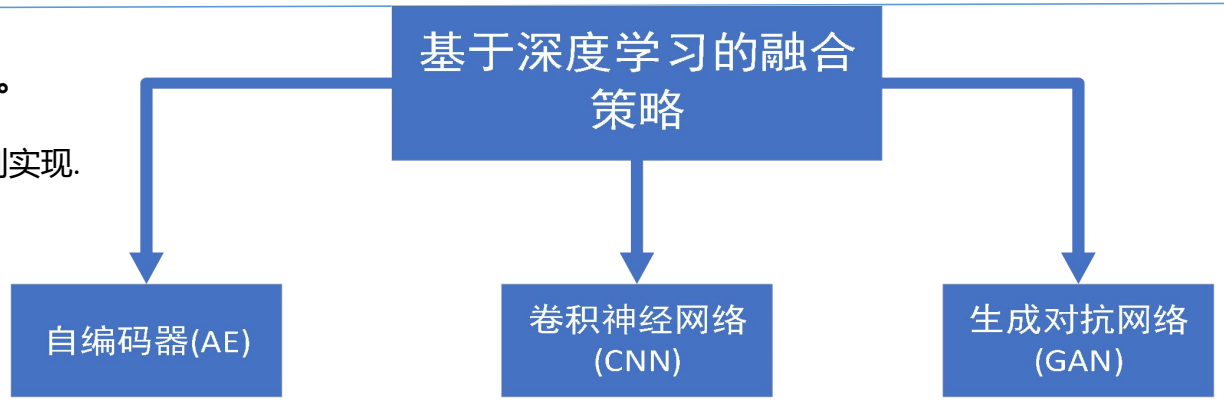
基于深度学习的图像融合方法包括基于AE,CNN以及GAN的方法。

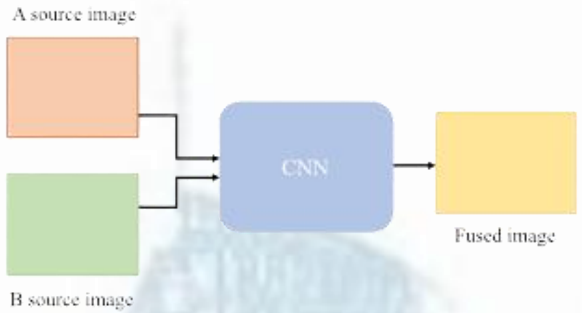
- AE-based

使用预训练的autoencoder实现特征提取和重建,中间的融合特征利用传统的融合规则实现.



AE-based





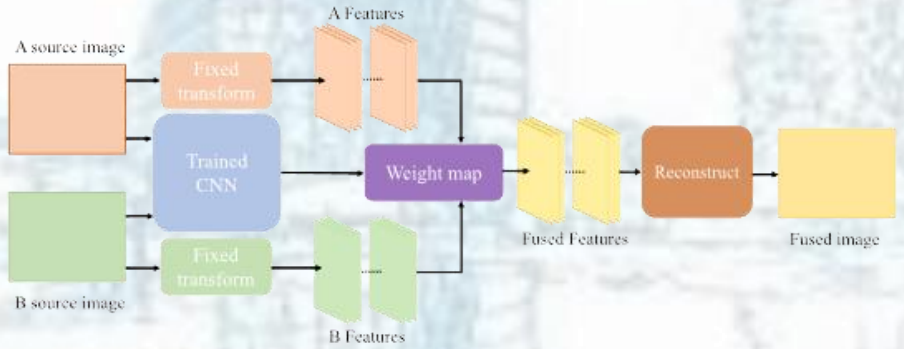
CNN-based

- CNN-based

通常有两种实现方式

1. 通过使用精心设计的损失函数和网络结构,实现端到端的特征提取、特征融合和图像重建。

2. 另一种方法是采用 CNN 来拟合融合规则,而特征提取和图像重建则采用传统方法进行。比如使用预先训练好的网络（如 VGGNet）从源图像中提取特征，并根据这些特征生成融合权重图。

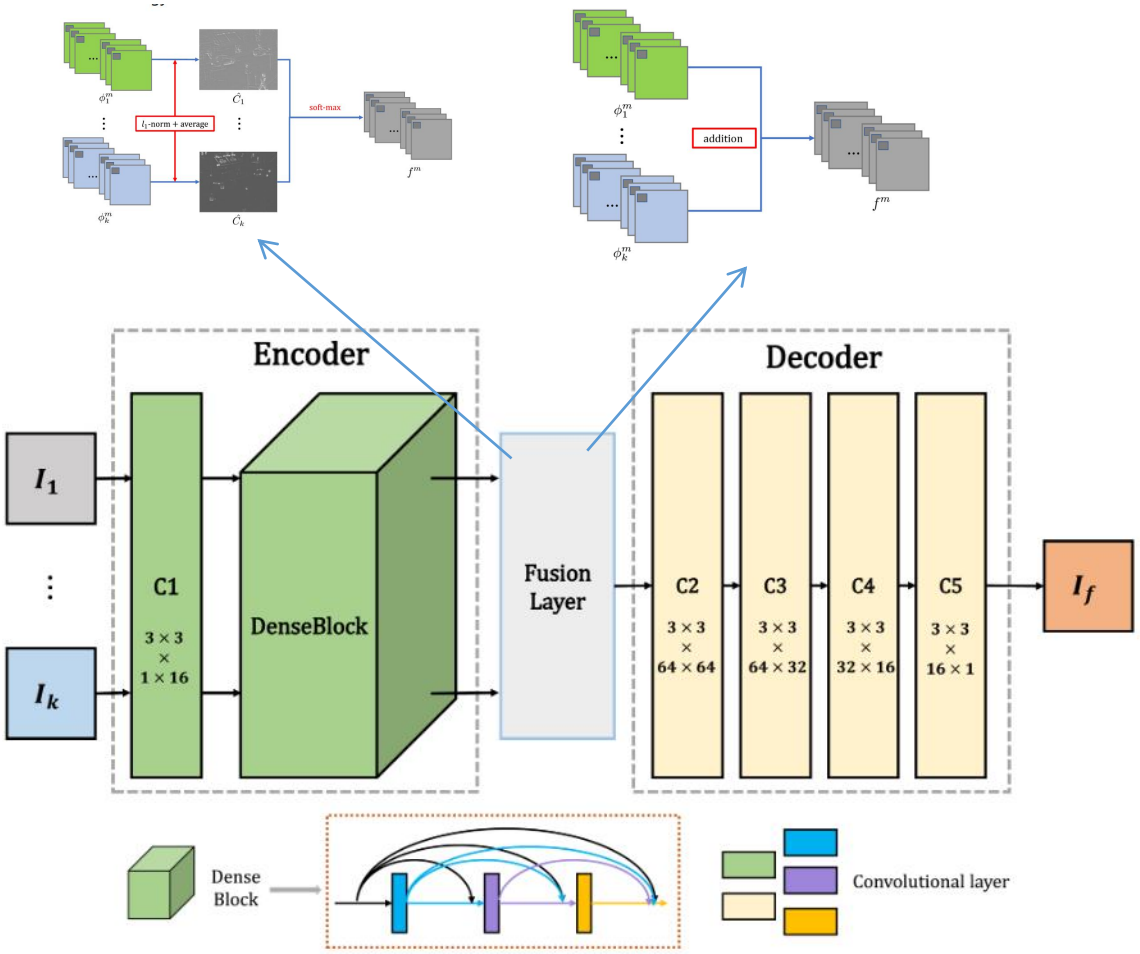


基于深度学习的图像融合方法

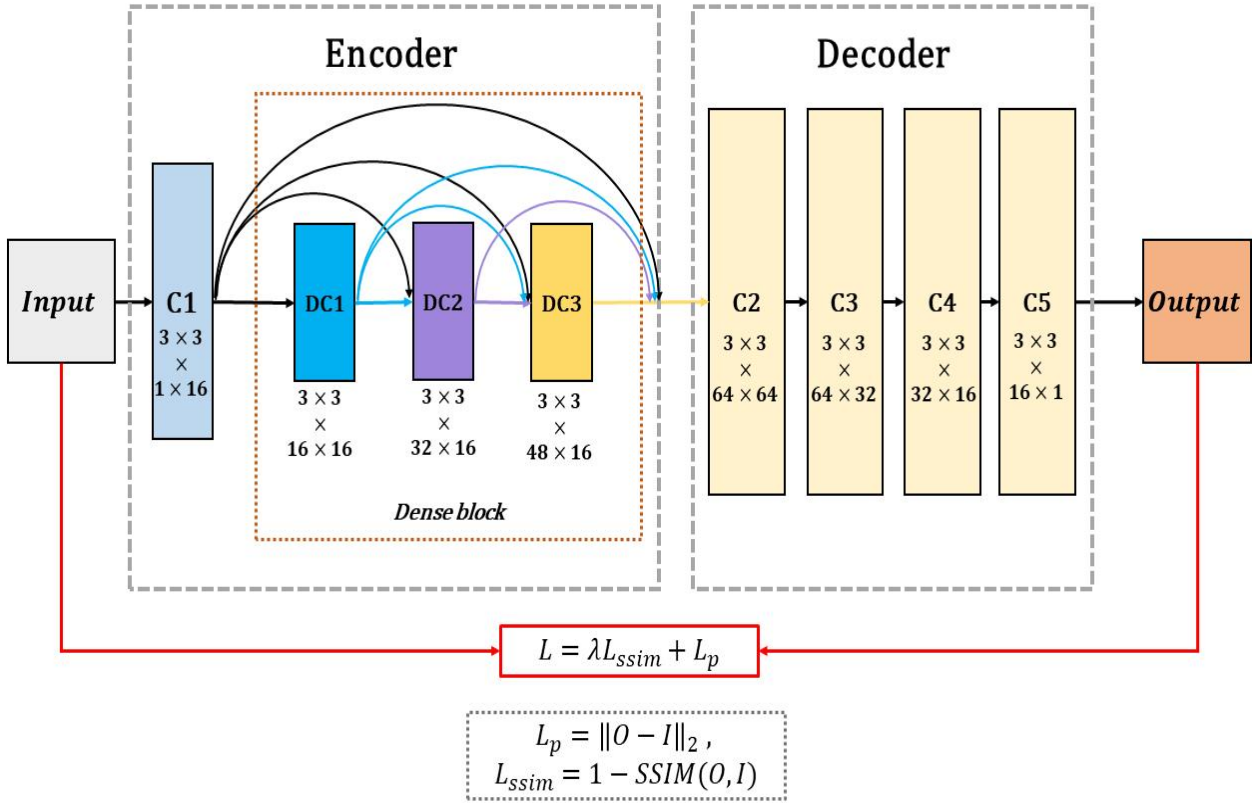
西北工业大学计算机学院
School of Computer Science, Northwestern Polytechnical University

Densefuse: A Fusion Approach to Infrared and Visible Images

2019 IEEE Transactions on Image Processing



during generation

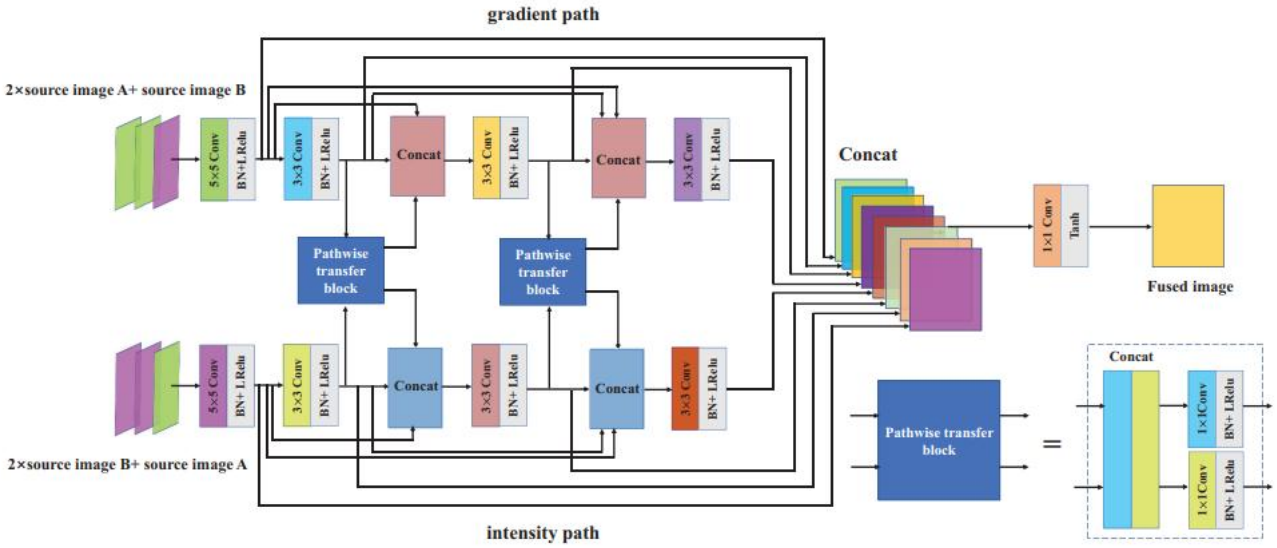


during training

基于深度学习的图像融合方法

Rethinking the Image Fusion: A Fast Unified Image Fusion Network based on Proportional Maintenance of Gradient and Intensity

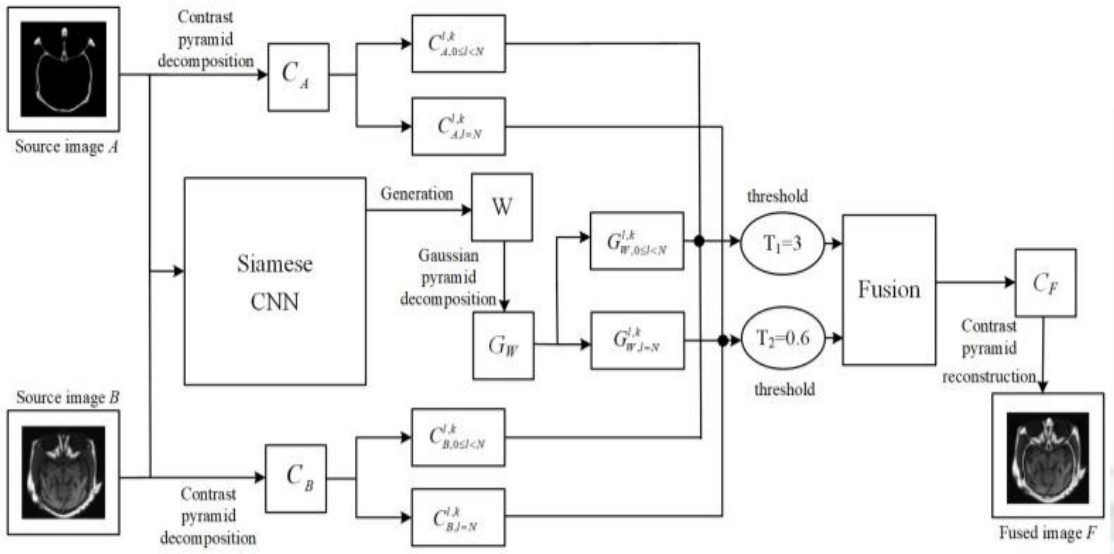
2020 AAAI



基于端到端 CNN 方法的一个代表作是 PMGI,它提出了梯度和强度的比例维护损失，引导网络直接生成融合图像。

Multi-Modality Medical Image Fusion Using Convolutional Neural Network and Contrast Pyramid

2020 Sensors



另一种方法是采用训练好的的 CNN 来制定融合规则,而特征提取和图像重建则采用传统方法进行。

基于深度学习的图像融合方法

- **GAN-based**

GAN 方法依靠生成器和判别器之间的对抗博弈来估计目标的概率分布，从而以隐含的方式共同完成特征提取、特征融合和图像重构

- 该模型由两个（至少）具有博弈学习功能的模块实现：生成模型（G）和判别模型（D）。一般来说，模块越多，训练难度越复杂。具体来说，在生成模型中，通过构建从先验分布 Pz(z) 到数据空间的映射函数 G(z;θg)，可以在真实数据集 x 上学习到生成分布 P_g，同时根据 $D(x;\theta_d)$ 可以得到判断输入是真还是假的概率值。优化过程可以看作是一个最小-最大的双人博弈，其目标函数定义为：

$$\min_G \max_D \mathbb{E}[\log D(I_{\text{real}})] + \mathbb{E}[\log(1 - D(I_{\text{fake}}))].$$

- 在训练过程中，生成模型的目标是尽可能生成真实图像来欺骗判别模型。判别模型的目标是尽可能地将生成模型生成的图像（即假数据）与真实图像（即真数据）分离开来。这样，生成模型和判别模型就构成了一个动态的 "博弈过程"。

优点:GAN 方法能充分利用源图像中的信息（如曝光条件、场景结构等）来建立无监督对抗模型可能是实现高质量多曝光融合的一个不错选择。

缺点:这种对抗模型可以使融合图像包含尽可能多的源图像信息，但假设融合图像中的信息总是源图像的某种累加可能并不准确。
此外针对GAN模型本身,也有训练不稳定等问题。

WGAN

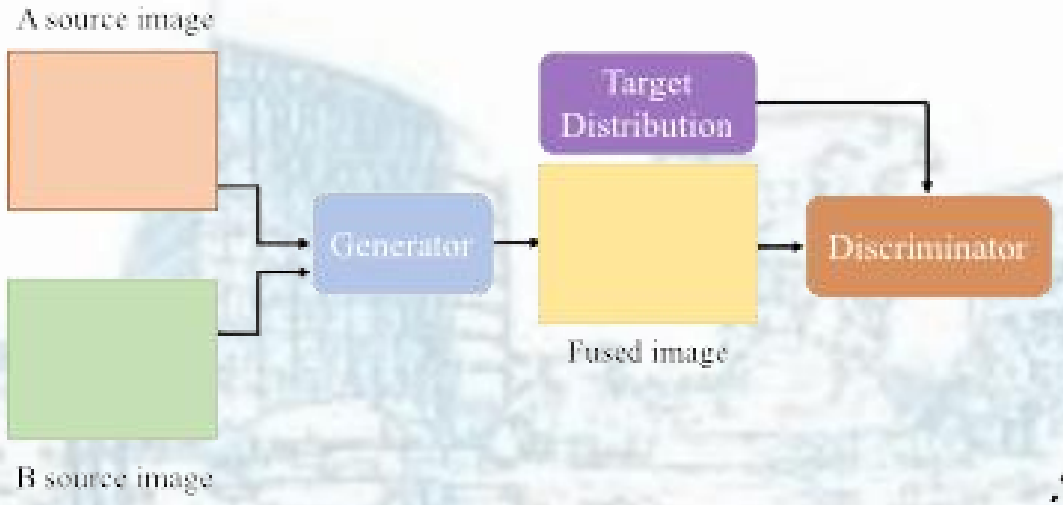
$$L = E_{x\sim p_{data}}[f_\omega(x)] - E_{x\sim p_g}[f_\omega(x)]$$

s.t. $|f(x_1) - f(x_2)| \le K|x_1 - x_2|$

WGAN-GP

$$L(D) = -E_{x\sim p_{data}}[D(x)] + E_{x\sim p_g}[D(x)] + \lambda E_{x\sim P_{\hat{x}}}[(\|\nabla_x D(x)\|_p - 1]^2$$

GAN-based



图像融合的评价指标

图像融合的评价与具体融合任务相关,**针对红外图像与可见光图像的融合任务。**
红外图像可以根据热辐射的差异将目标与背景区分开来,具有**明显的对比度**,这**在任何昼夜和天气条件下都很有效。**
相比之下,可见光图像可以提供**丰富的纹理细节,空间分辨率高**,清晰度高,符合人类视觉系统。

由于融合后的图像没有ground truth,一些常见的图像质量评价指标可以用于评估融合图像与可见光图像、红外图像的相似性。具体包括平均结构相似性SSIM,峰值信噪比PSNR,互信息MI,熵EN,标准差SD,空间频率SF等等。

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2$$

$$PSNR = 10 \cdot log_{10}\frac{(MAX_I)^2}{MSE} = 20 \cdot log_{10}\frac{MAX_I}{\sqrt{MSE}}$$

$$SF = \sqrt{RF^2 + CF^2}$$

$$RF = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}|H(i,j) - H(i,j-1)|^2}$$

$$CF = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}|H(i,j) - H(i-1,j)|^2}$$

$$SSIM(x,y) = [l(x,y)^{\alpha} \cdot c(x,y)^{\beta} \cdot s(x,y)^{\gamma}]$$

$$l(x,y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1}$$

$$\mu_x = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W}x(i,j)$$

$$\mu_y = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W}y(i,j)$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2}s(x,y)$$

$$\sigma_x = [\frac{1}{H \times W - 1}\sum_{i=1}^{H}\sum_{j=1}^{W}(x(i,j) - \mu_x)]^{1/2}$$

$$\sigma_y = [\frac{1}{H \times W - 1}\sum_{i=1}^{H}\sum_{j=1}^{W}(x(i,j) - \mu_y)]^{1/2}$$

$$s(x,y) = \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3}$$

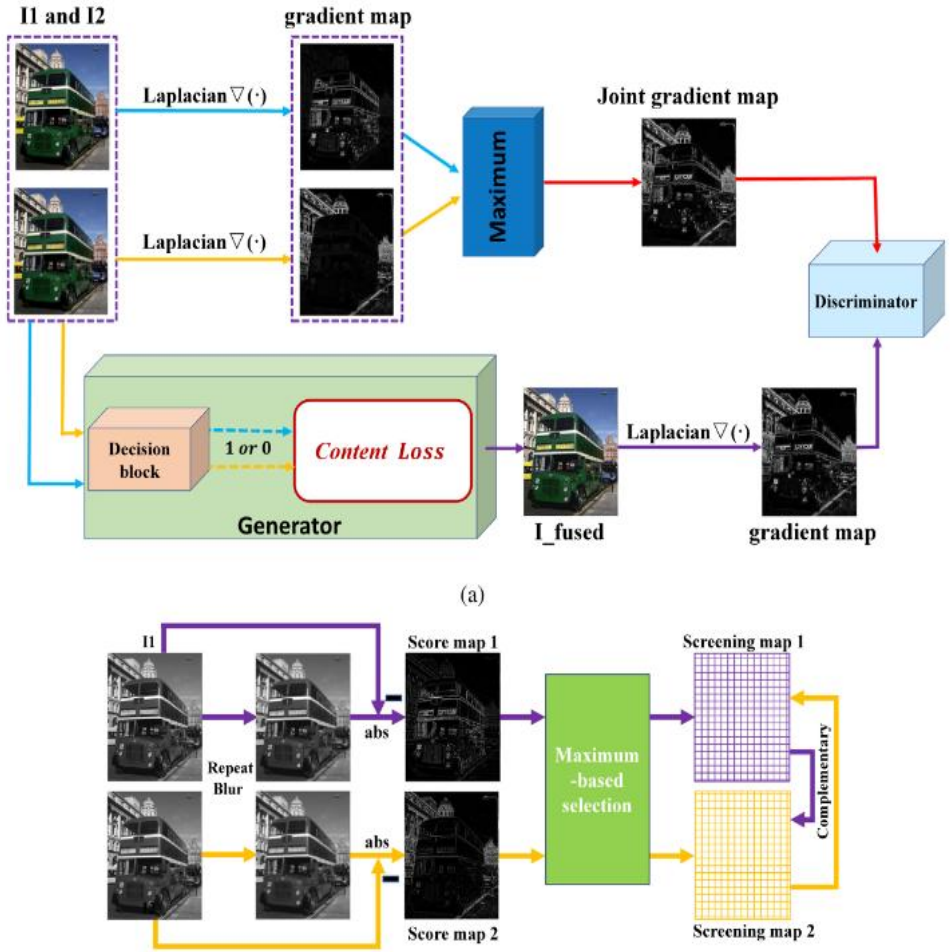
多焦距图像融合

西北工业大学计算机学院
School of Computer Science, Northwestern Polytechnical University

MFF-GAN: An unsupervised generative adversarial network with adaptive and gradient joint constraints for multi-focus image fusion

2021 Information Fusion

动机:**由于光学镜头的局限性，很难在一张图像中将所有不同景深的物体全部聚焦。**在此背景下,多焦点图像融合作为一种图像增强方法，可以将不同聚焦区域的图像融合在一起，获得单个全清晰图像，在各个领域具有良好的应用前景。



对于多焦点图像融合,最有意义的信息是源图像中的锐利区域(图像梯度大的区域),这些区域反映图像的强度分布和纹理细节。

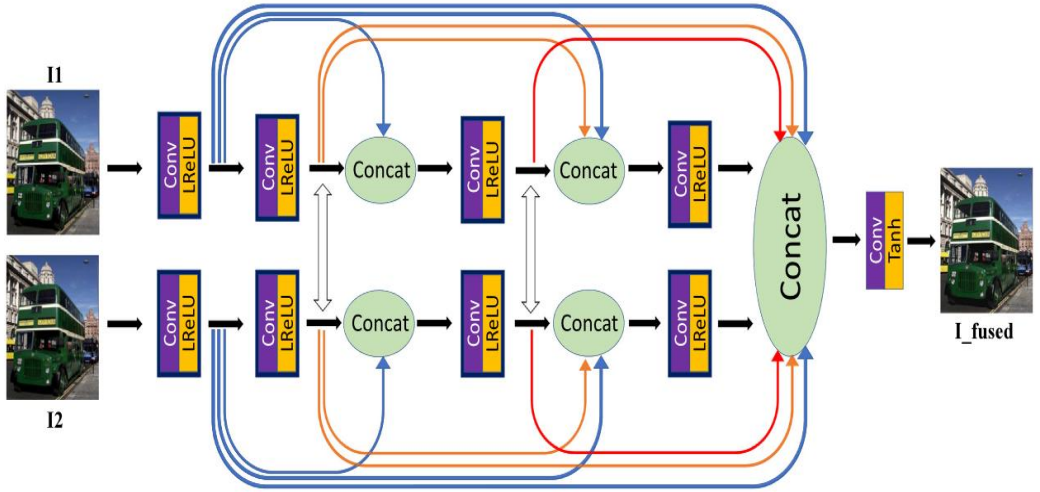
在信息提取的过程中,应该保留锐利区域的这些信息,而模糊区域的这些信息应该被丢弃。因此,有必要在优化过程中调整损失函数的机制,从而约束网络有选择地提取和重构信息。

此外,为加强融合结果的细节,以减少神经网络在图像生成任务中常见的平滑效应。基于这些考虑,设计了一个具有自适应和梯度联合约束的生成对抗网络,

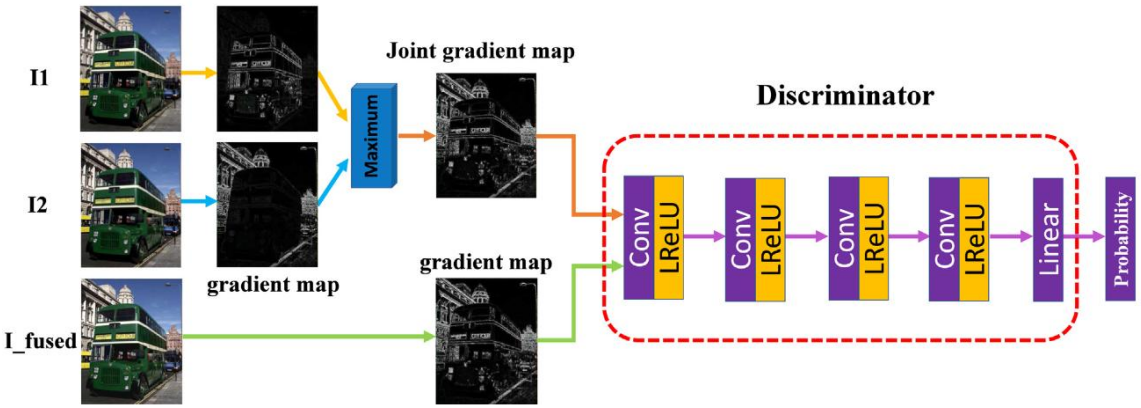
主要关注图像的强度以及纹理,而聚焦的图像与没有聚焦的图像可以根据图像梯度的大小分为前景和后景,这在许多多焦距图像融合论文中体现.

MFF-GAN: An unsupervised generative adversarial network with adaptive and gradient joint constraints for multi-focus image fusion

西北工业大学计算机学院
School of Computer Science, Northwestern Polytechnical University



生成器



辨别器

生成器损失函数

$$\mathcal{L}_G = \mathcal{L}_{G_{\text{adv}}} + \alpha L_{G_{\text{con}}}$$

$$\mathcal{L}_{G_{\text{adv}}} = \frac{1}{N}\sum_{n=1}^{N}(D(\nabla(I^n_{\text{fused}})) - a)^2$$

$$L_{G_{\text{con}}} = \beta_1 \mathcal{L}_{\text{int}} + \beta_2 \mathcal{L}_{\text{grad}}$$

$$\mathcal{L}_{\text{int}} = \frac{1}{HW}\sum_{i}\sum_{j} S_{1_{i,j}} \cdot (I_{\text{fused}_{i,j}} - I_{1_{i,j}})^2 + S_{2_{i,j}} \cdot (I_{\text{fused}_{i,j}} - I_{2_{i,j}})^2$$

$$S_{1_{i,j}} = \text{sign}(RB(I_{1_{i,j}}) - \min(RB(I_{1_{i,j}}), RB(I_{2_{i,j}}))),$$

$$S_{2_{i,j}} = 1 - S_{1_{i,j}},$$

$$RB(\cdot) = abs(I_{i,j} - LP(I_{i,j}))$$

LP指低通滤波

$$\mathcal{L}_{\text{grad}} = \frac{1}{HW}\sum_{i}\sum_{j} S_{1,j} \cdot (\nabla I_{\text{fused}_{i,j}} - \nabla I_{1_{i,j}})^2 + S_{2_{i,j}} \cdot (\nabla I_{\text{fused}_{i,j}} - \nabla I_{2_{i,j}})^2.$$

辨别器损失函数

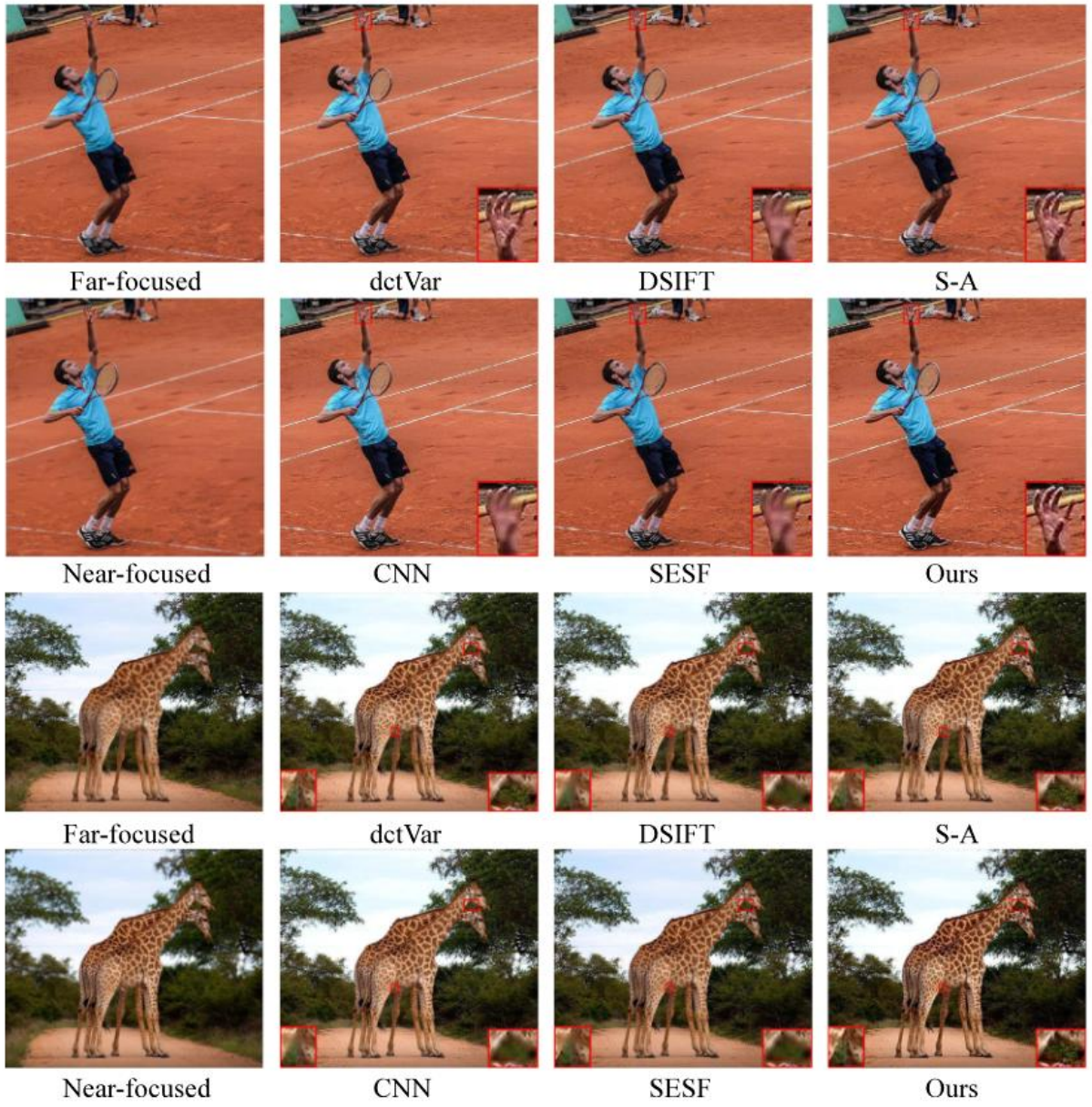
$$Grad_{\text{fused}} = \text{abs}(\nabla I_{\text{fused}})$$

$$Grad_{\text{joint}} = \max(\text{abs}(\nabla I_1), \text{abs}(\nabla I_2)),$$

$$\mathcal{L}_{D_{\text{adv}}} = \frac{1}{N}\sum_{n=1}^{N}[D(Grad^n_{\text{fused}}) - b]^2 + [D(Grad^n_{\text{joint}}) - c]^2$$

MFF-GAN: An unsupervised generative adversarial network with adaptive and gradient joint constraints for multi-focus image fusion

西北工业大学计算机学院
School of Computer Science, Northwestern Polytechnical University



论文中的定性分析

作者认为,所提出的方法能更好地保持规则纹理,而 dctVar 和 S-A 则不能。例如，它们模糊了长颈鹿的轮廓，而 MFF-GAN 结果中这些轮廓被锐化了。

同时还能够准确保留聚焦和散焦区域边界线附近的细节，而 CNN、SESF 和 DSIFT 在第一组结果中运动员的手模糊了。

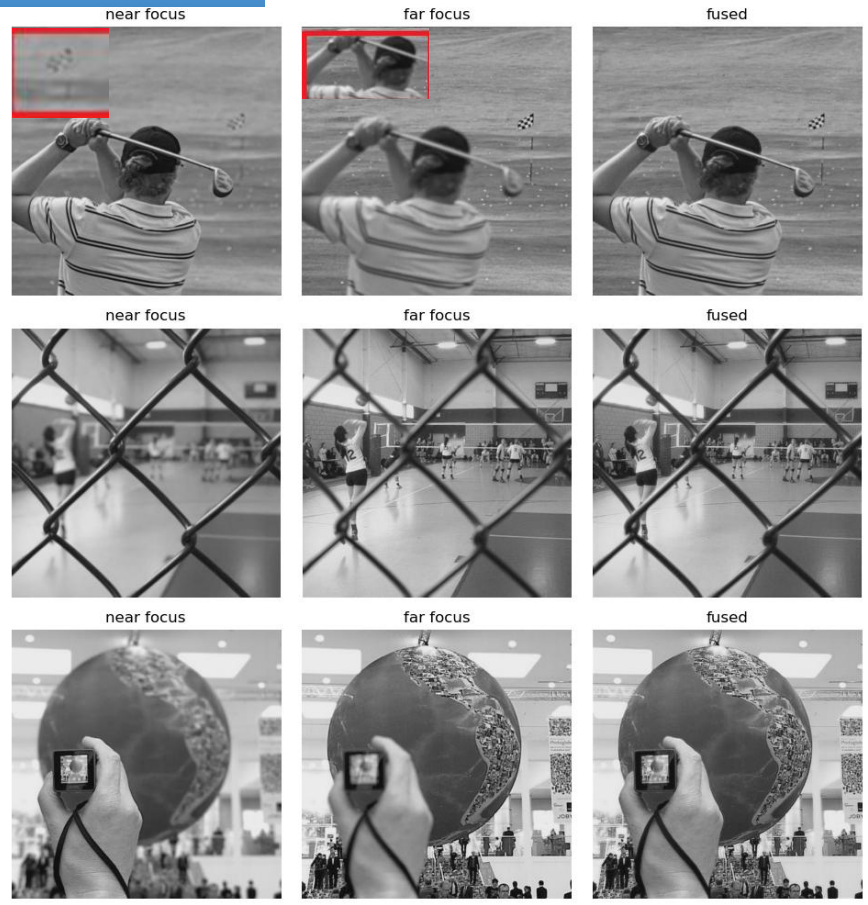
总的来说，所提出方法不仅具有良好的整体清晰度，而且还能保持局部细节，尤其是在聚焦和散焦区域的交界处

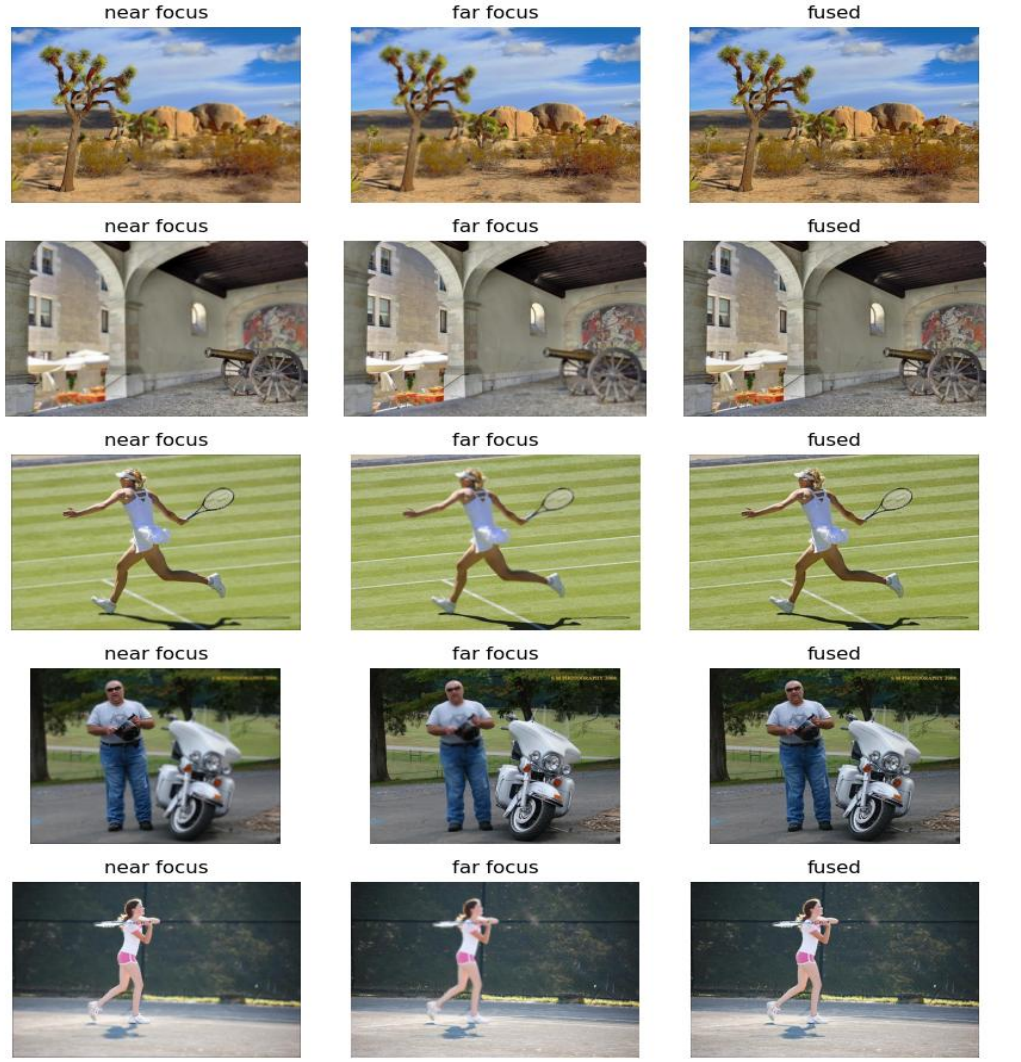
MFF-GAN: An unsupervised generative adversarial network with adaptive and gradient joint constraints for multi-focus image fusion

西北工业大学计算机学院
School of Computer Science, Northwestern Polytechnical University

在训练时,常常会把图像分为多个patch然后对于patch进行处理,一方面因为一张图像上不同部分可能在纹理细节等方面有较大差异,这样处理考虑到了图像不同部分,同时也提供了更多训练数据。论文中具体处理三通道RGB图时,会将RGB图片转为YCrCb格式然后只融合Y通道,其余两个通道由传统方式融合.
另外也有论文使用目标检测数据集Pascal VOC,利用segmentation mask制作多焦点数据集用于训练(***MFIF-GAN: A new generative adversarial network for multi-focus image fusion***)

定性分析

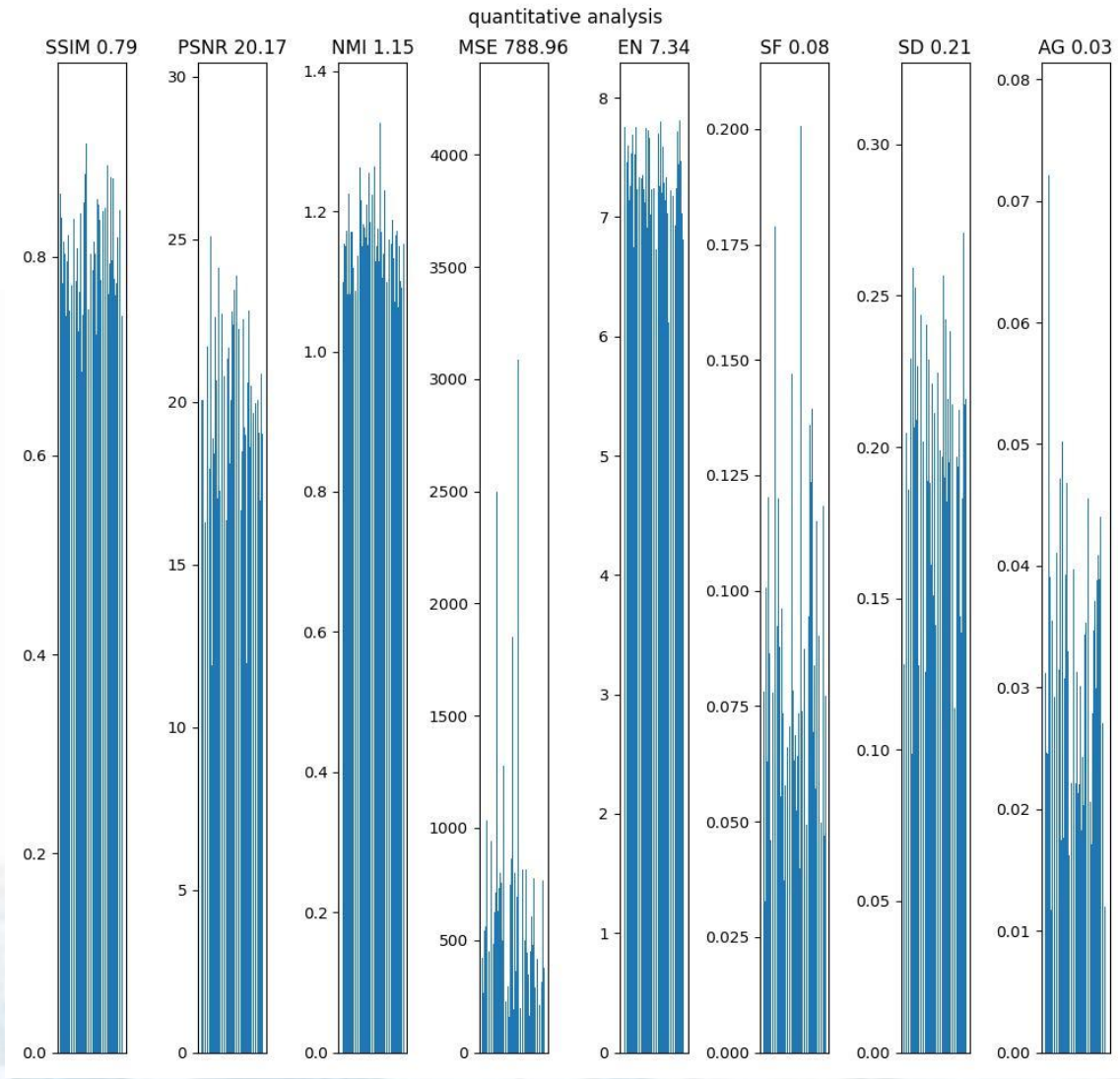


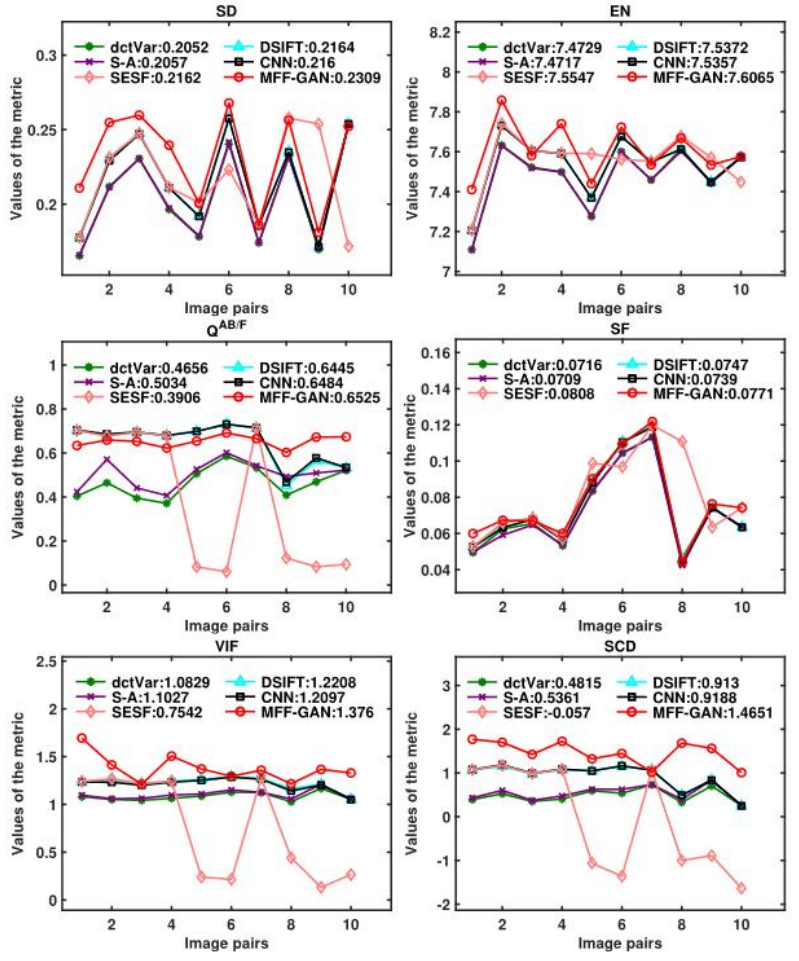


MFF-GAN: An unsupervised generative adversarial network with adaptive and gradient joint constraints for multi-focus image fusion

西北工业大学计算机学院
School of Computer Science, Northwestern Polytechnical University

定量分析





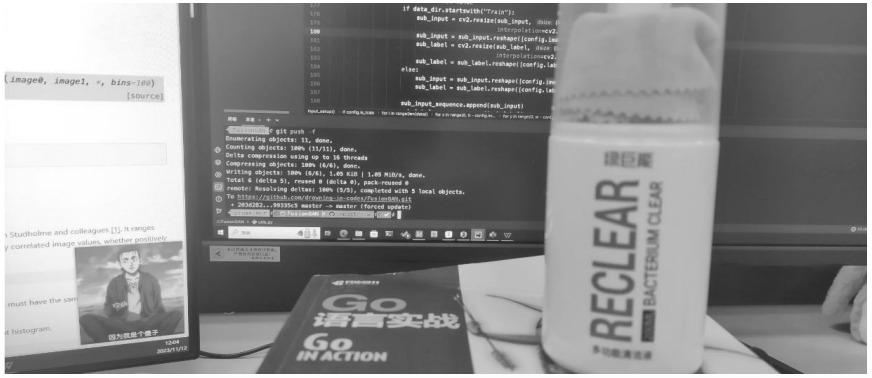
from MFF-GAN

MFF-GAN: An unsupervised generative adversarial network with adaptive and gradient joint constraints for multi-focus image fusion

西北工业大学计算机学院
School of Computer Science, Northwestern Polytechnical University

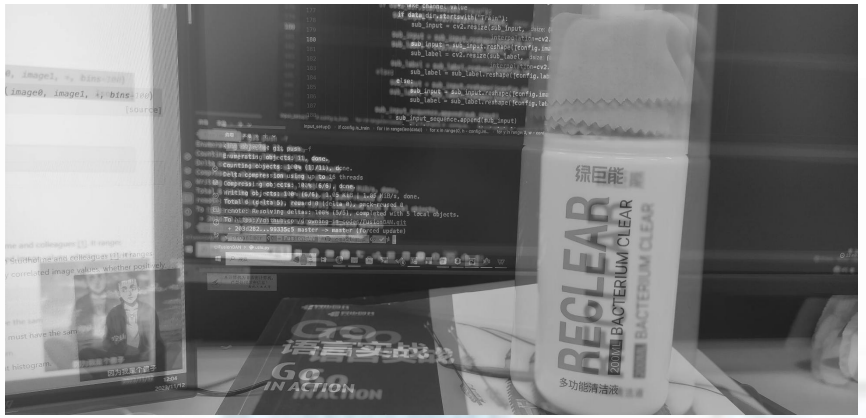
遇到的问题

near focus



far focus

fused

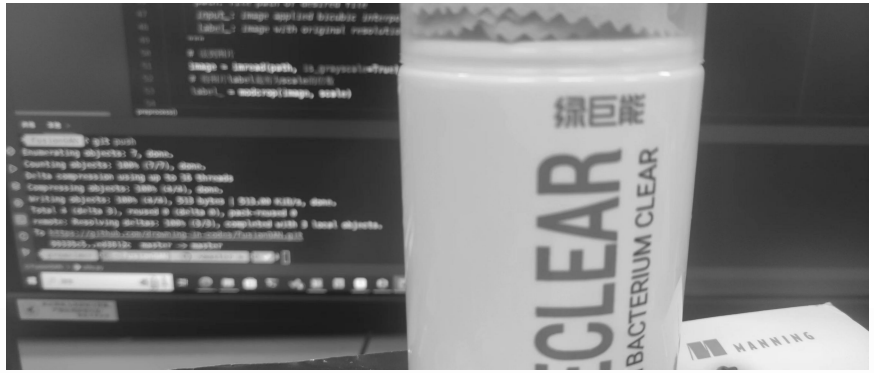


MFF-GAN: An unsupervised generative adversarial network with adaptive and gradient joint constraints for multi-focus image fusion

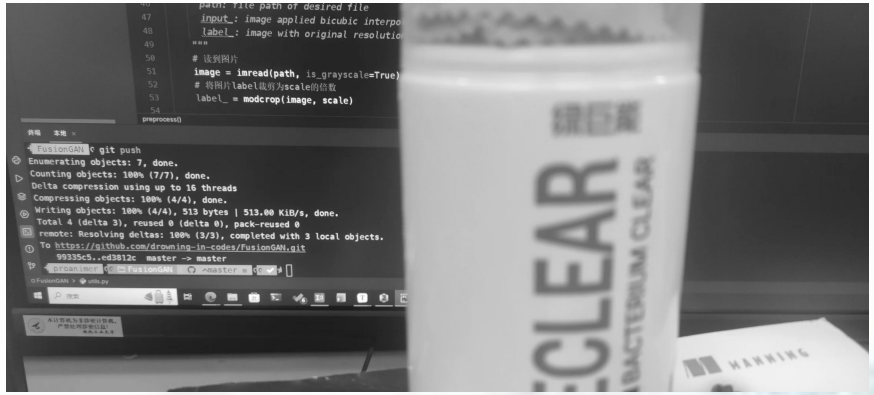
西北工业大学计算机学院
School of Computer Science, Northwestern Polytechnical University

遇到的问题

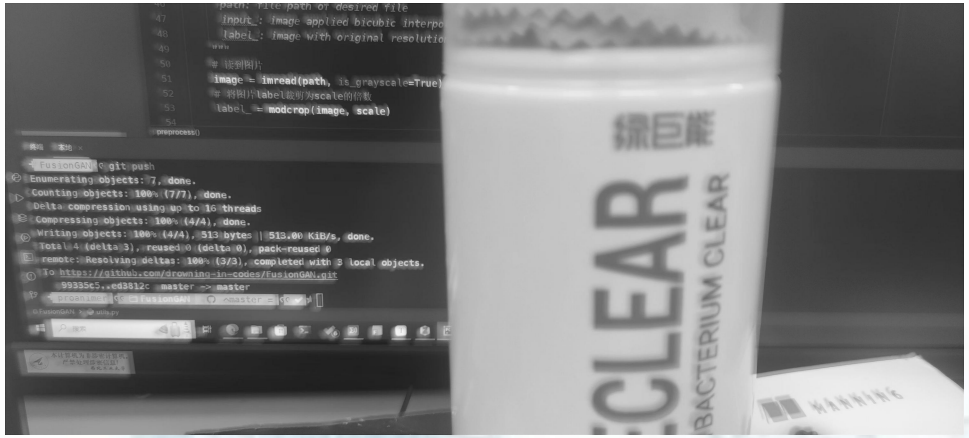
near focus



far focus



fused



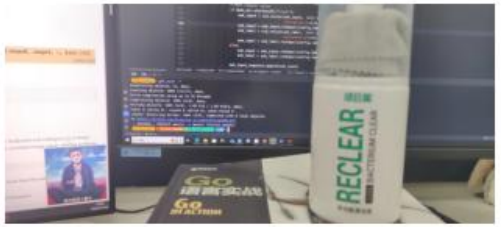
融合图像出现了重影,**推测原因是手机拍摄照片时,改变焦距实际上会进行裁剪。这样使得拍摄的图像会有些许偏移,不是对齐的**

MFF-GAN: An unsupervised generative adversarial network with adaptive and gradient joint constraints for multi-focus image fusion

西北工业大学计算机学院
School of Computer Science, Northwestern Polytechnical University

near focus

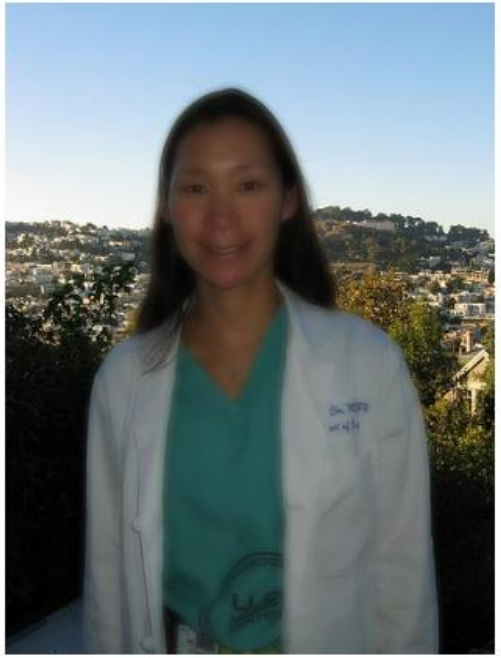
far focus



fused

near focus

far focus



fused

original

使用了Pascal VOC数据集中的图片,利用了分割的mask制作近焦和远焦的数据集

红外与可见光图像融合

西北工业大学计算机学院
School of Computer Science, Northwestern Polytechnical University

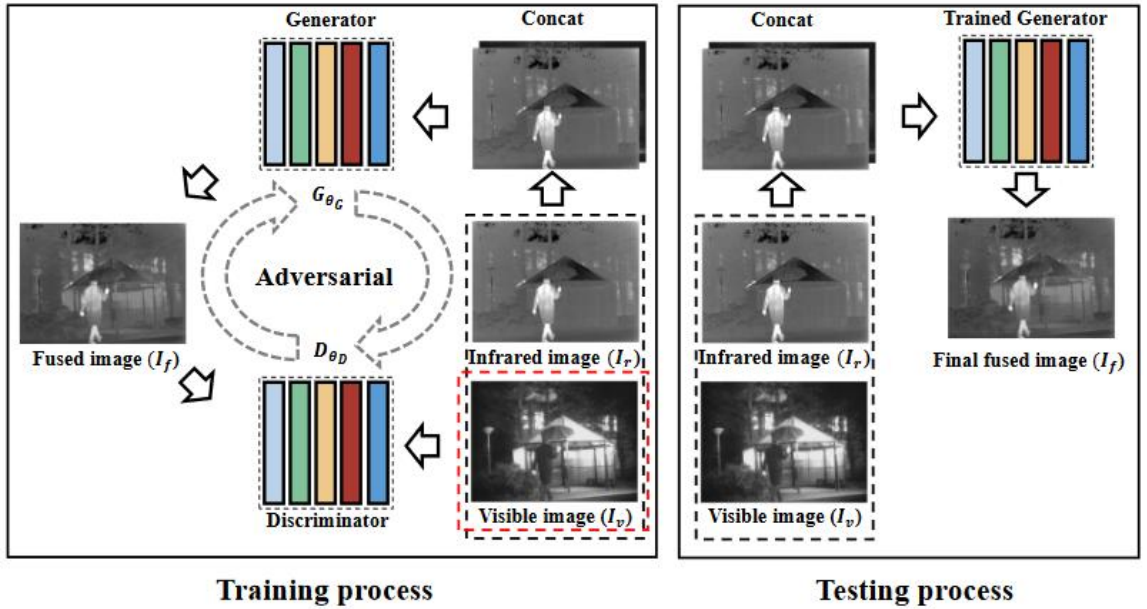
FusionGAN: A generative adversarial network for infrared and visible image fusion

2019 Information Fusion

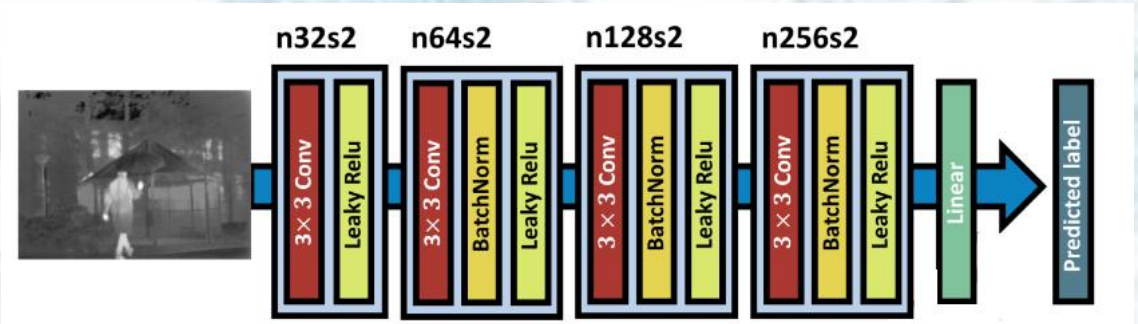
动机:红外图像可以根据热辐射的差异将目标与背景区分开来,这在白天/黑夜和所有天气条件下效果都很好。相比之下,可见光图像可以以与人类视觉系统一致的方式提供具有高空间分辨率和清晰度的纹理细节。因此可以**充分利用红外图像和可见光图像的互补信息来融合具有视觉吸引力的图像或支持更高层次的视觉任务**，如分割、跟踪和检测等任务。

第一个提出了基于GAN模型融合红外图像与可见光图像的方法,设计了相应的损失函数较好地同时保留了红外图像中的热辐射相关的对比度以及可见光图像中的纹理。

端到端的模型，避免了传统方法中手动设计复杂的活动水平测量和融合规则的情况。



生成器结构 卷积中没有填充操作



辨别器结构 第一层加入padding

FusionGAN: A generative adversarial network for infrared and visible image fusion

生成器损失函数

$$\mathcal{L}_G = V_{\text{FusionGAN}}(G) + \lambda \mathcal{L}_{\text{content}},$$

$$V_{\text{FusionGAN}}(G) = \frac{1}{N}\sum_{n=1}^{N}\left(D_{\theta_D}(I_f^n) - c\right)^2,$$

$$\mathcal{L}_{\text{content}} = \frac{1}{HW}\left(\|I_f - I_r\|_F^2 + \xi\|\nabla I_f - \nabla I_v\|_F^2\right)$$

生成器损失,包括对抗性损失以及内容损失。

由于红外图像的热辐射信息由其像素强度表征，而可见光图像的纹理细节信息可部分由其梯度表征,因此强制要求融合后的图像 具有与红外图像相似的图像强度和与可见图像相似的梯度。

而内容损失包括图像强度值差异以及图像梯度差异,分别使用图像灰度以及图像梯度表示。而由于仅使用梯度信息并不能完全代表可见光图像中的纹理细节,另外还引入了对抗损失,根据可见光图像调整融合图像。

辨别器损失函数

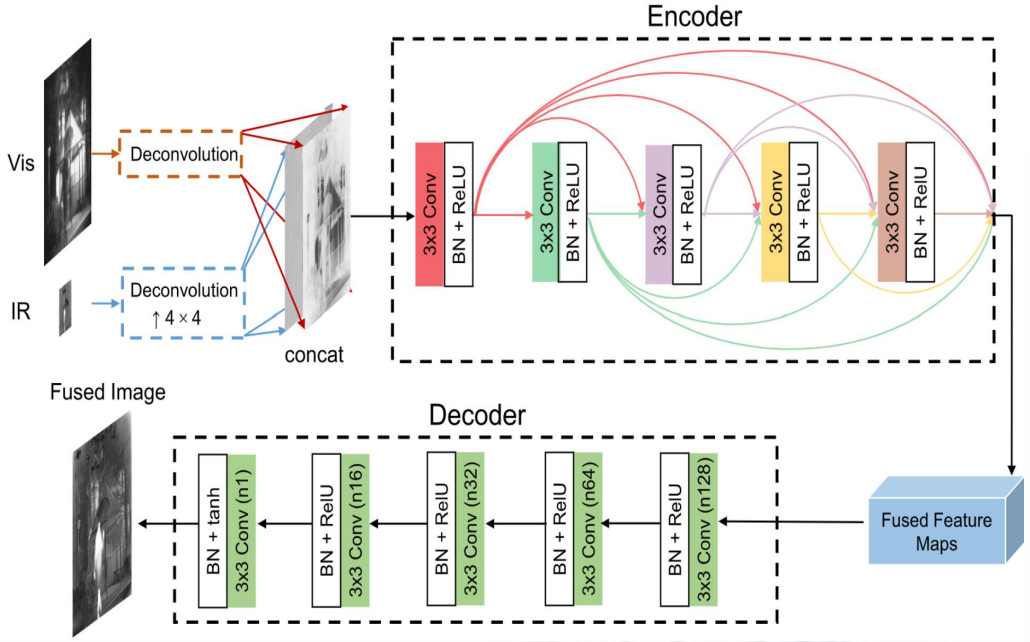
$$\mathcal{L}_D = \frac{1}{N}\sum_{n=1}^{N}\left(D_{\theta_D}(I_v) - b\right)^2 + \frac{1}{N}\sum_{n=1}^{N}\left(D_{\theta_D}(I_f) - a\right)^2$$

分辨器损失,包括对抗性损失。

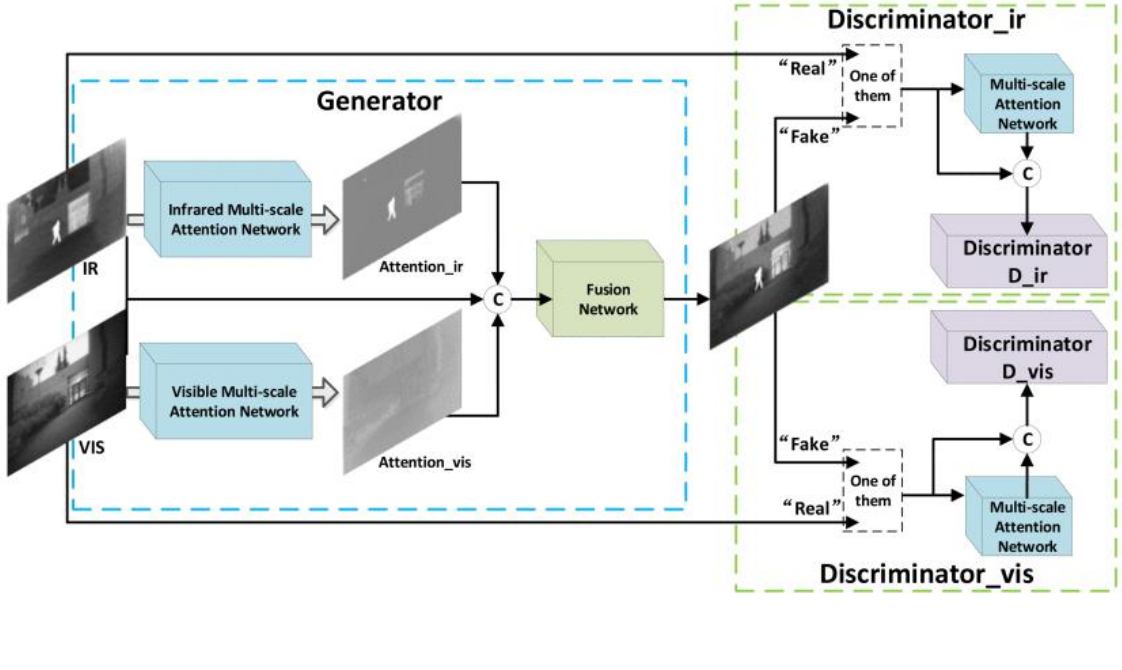
FusionGAN: A generative adversarial network for infrared and visible image fusion

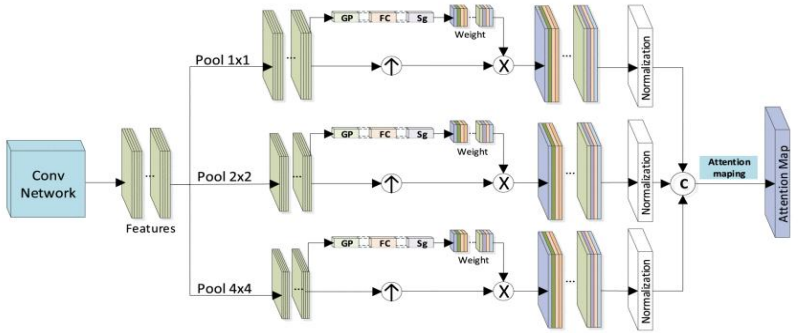
西北工业大学计算机学院
School of Computer Science, Northwestern Polytechnical University

Similar Work



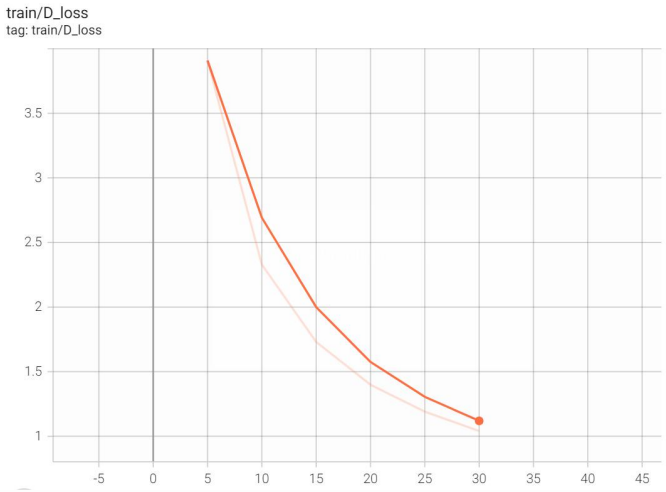
from DDcGAN(2020)
Information Fusion

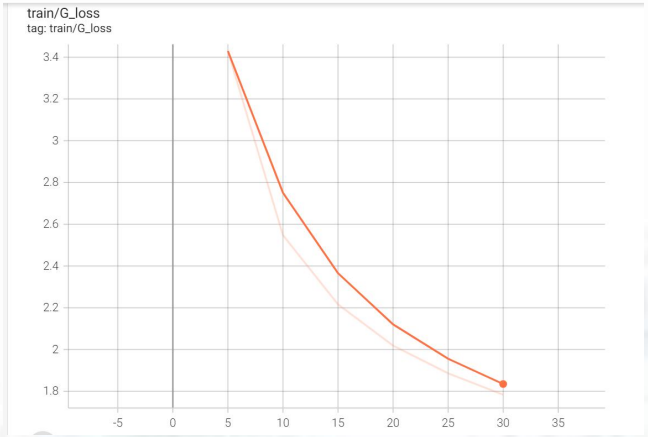




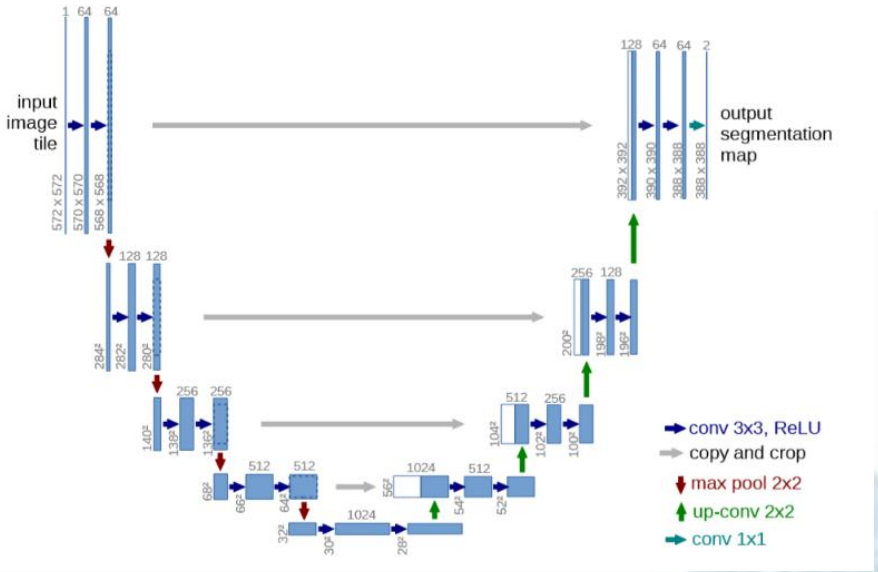
from AttentionFGAN(2020)
IEEE Transactions on Multimedia

可行的改进

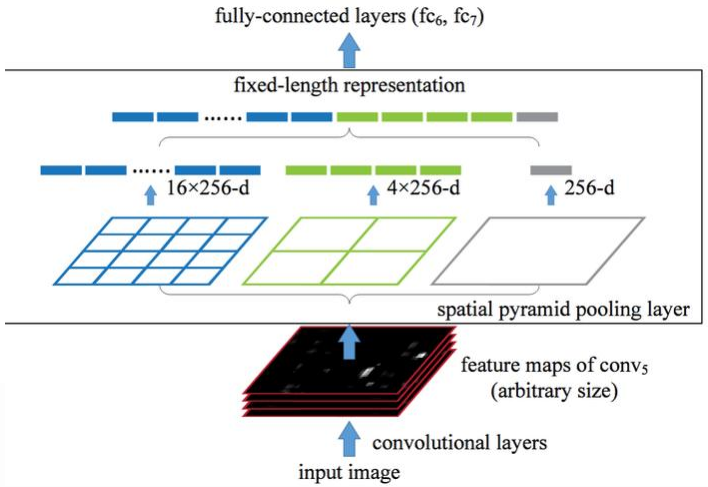




训练损失



U-Net



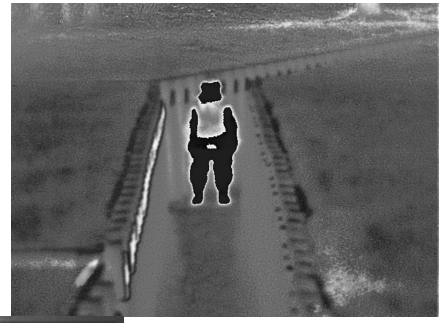
SSP-Net

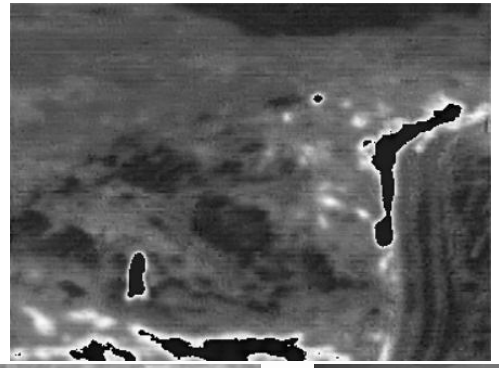
改进：模型U-Net;增加残差,增加损失函数,加入SSIM_loss;适应输入的GAP+FCN或者使用SSP-Net

UFGAN:Stable Fusion with transfer loss and U-Net

本组方法

遇到的问题





修改模型参数

修改patch大小

生成图像对模型参数、结构比较敏感

UFGAN:Stable Fusion with transfer loss and U-Net

定性分析

from FusionGAN

U_GAN

DDcGAN

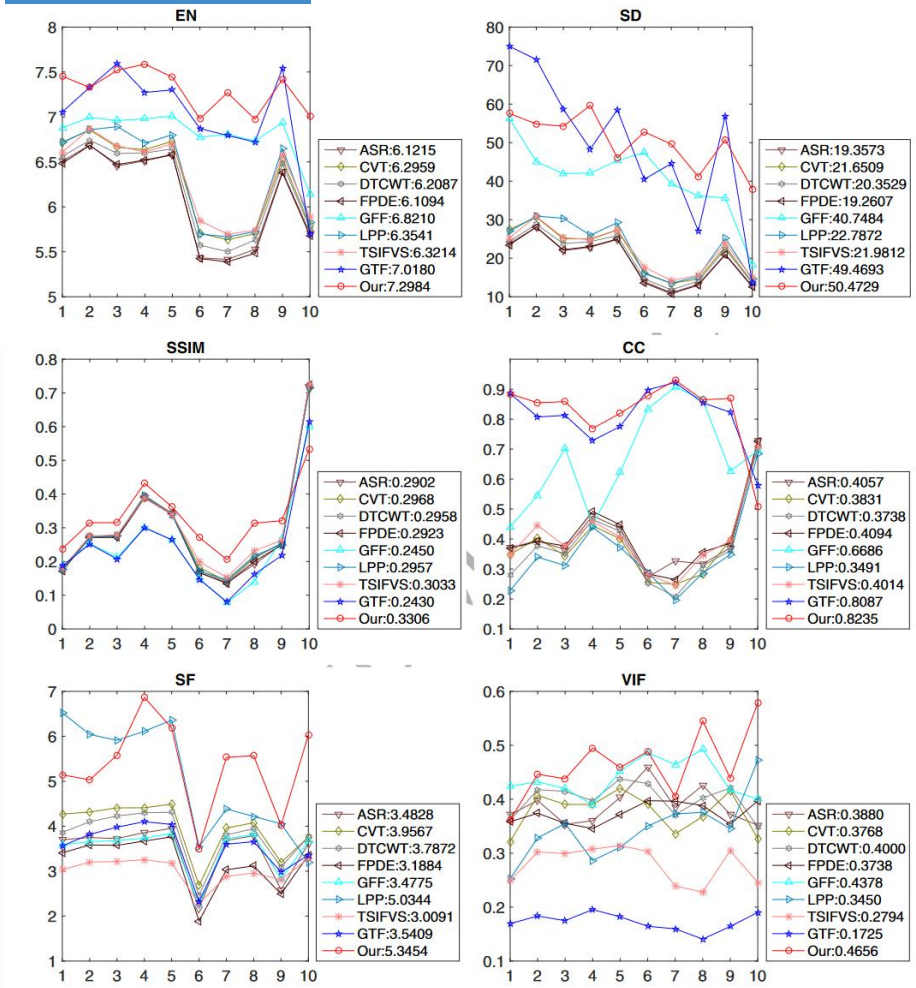
U_GAN

DDcGAN

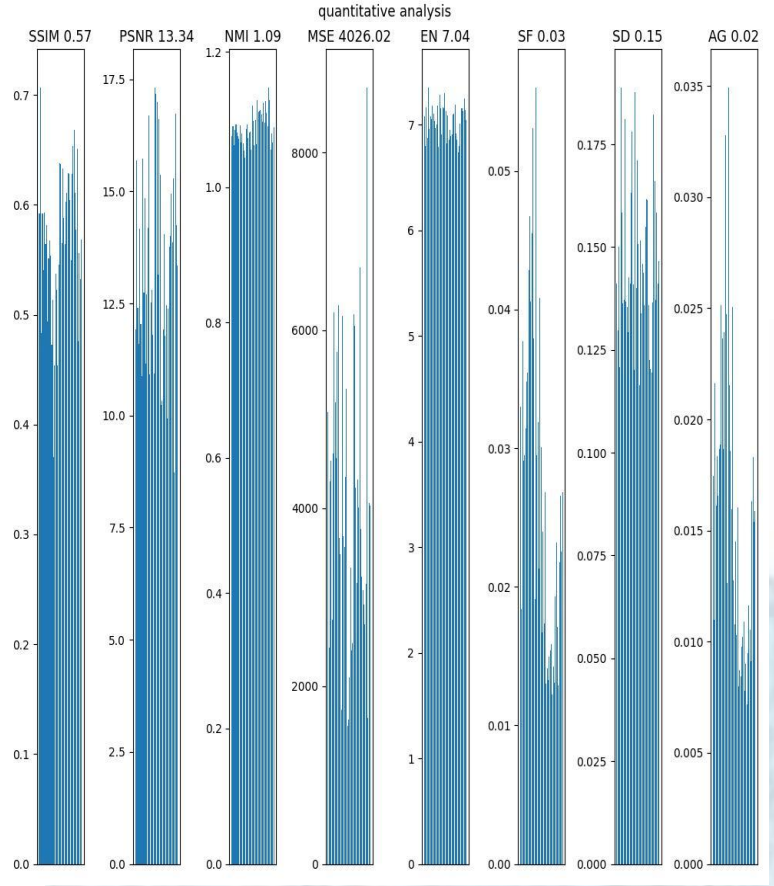
From top to bottom: infrared images, visible images, results of FusionGAN without adversarial loss and with adversarial loss

UFGAN:Stable Fusion with transfer loss and U-Net

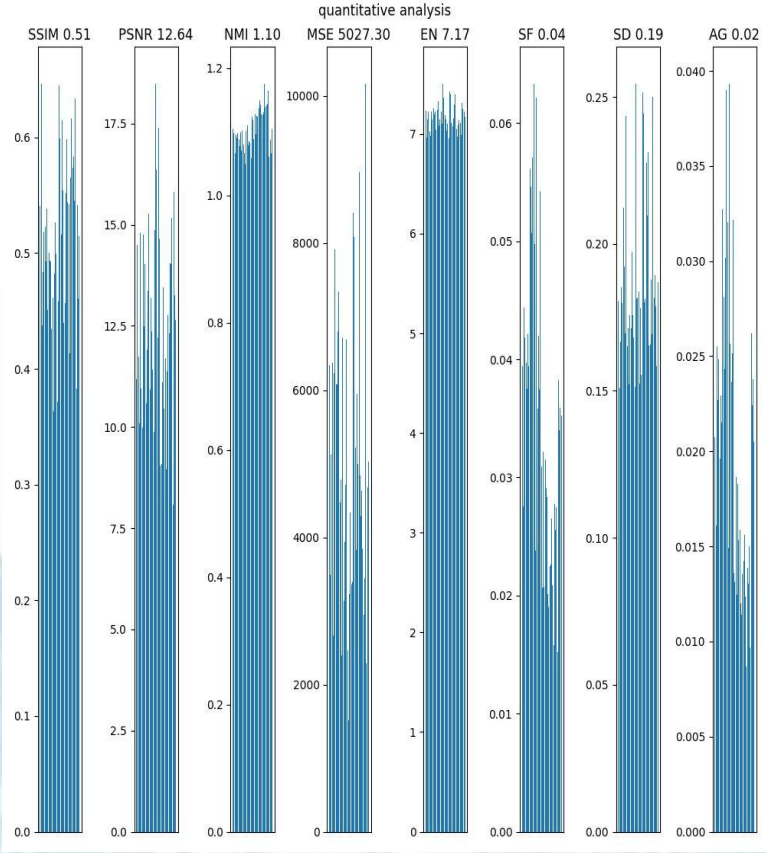
定量分析



from FusionGAN



U_GAN

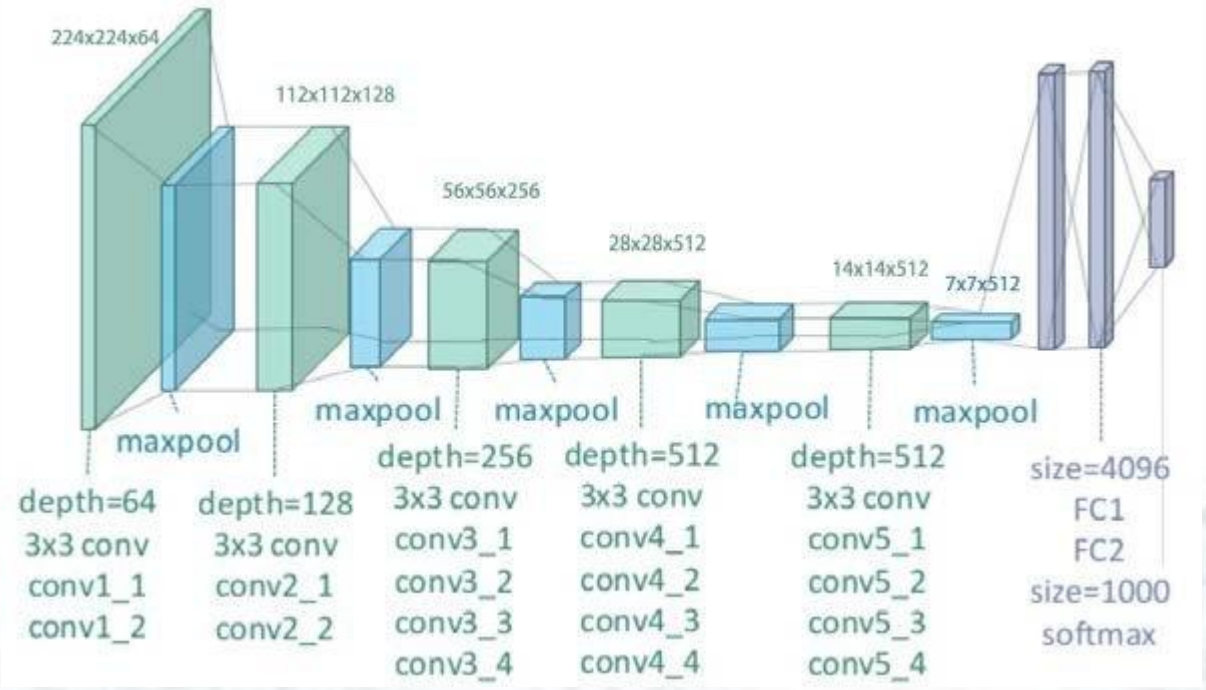


DDcGAN

UFGAN:Stable Fusion with transfer loss and U-Net

除了在模型上的改动,还可以?

引入新的损失,受transfer learning影响,使用一种损失函数同时控制融合图像内容与"风格"



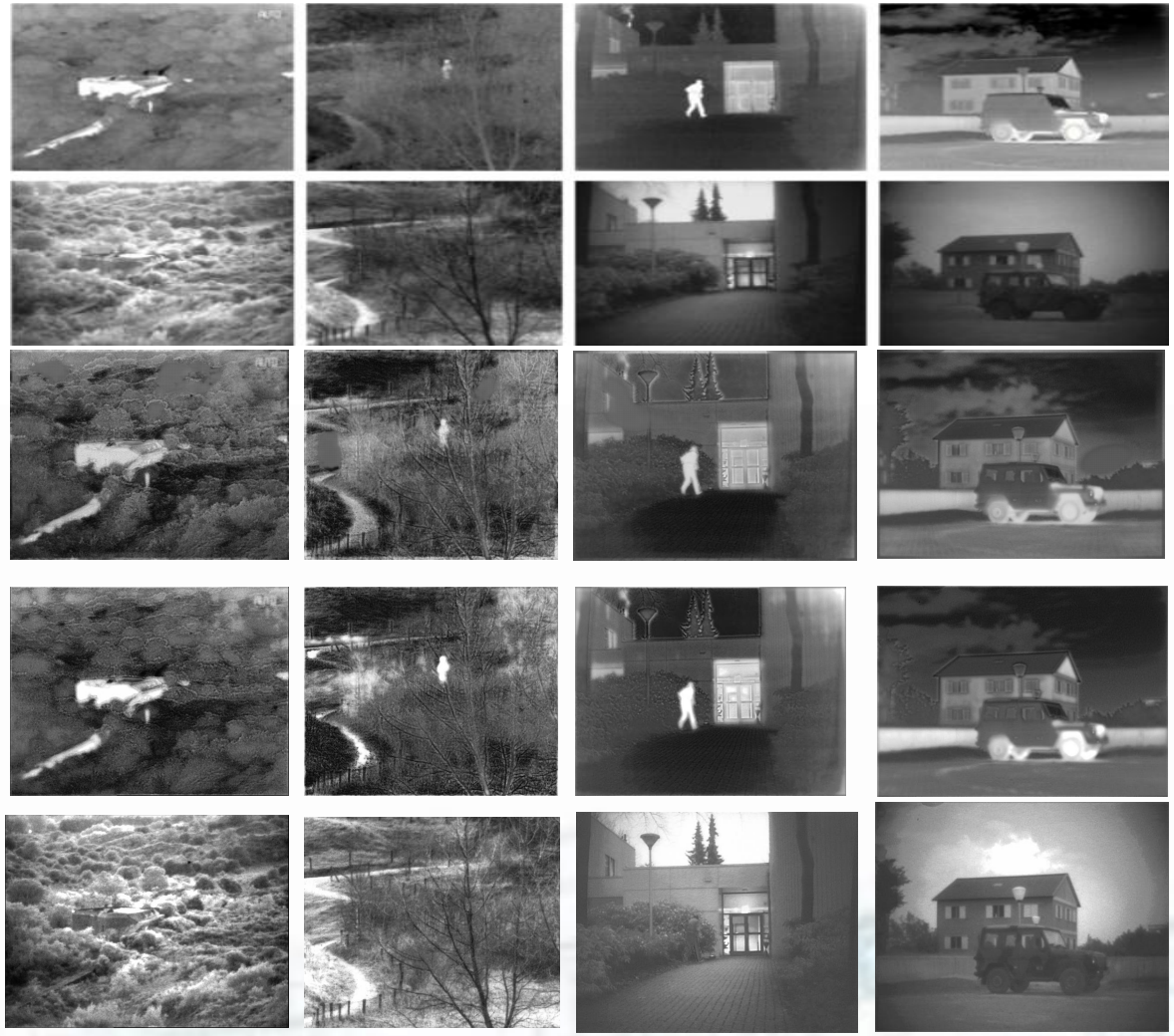
VGG19

$$L_{st} = L_{content} + k * L_{style} + \lambda * L_{tv}$$
$$L_{tv} = \sum_{i,j} |x_{i,j}x_{i+1,j}| + |x_{i,j}x_{i,j+1}|$$
$$L_{content} = MSE(O - I)$$
$$L_{style} = MSE(gram(O) - gram(I))$$
$$gram(I) = I * I^T$$

- 使用预训练网络的输出作为损失。一般来说，越靠近输入层，越容易抽取图像的细节信息；反之，则越容易抽取图像的全局信息。
- 其中包含内容损失,风格损失以及全变分损失。gram矩阵表示了特征不同通道的相关性,用以表示风格层输出的风格。
- 有时合成图像里面有大量高频噪点，即有特别亮或者特别暗的颗粒像素,这里使用全变分损失去噪。

UFGAN:Stable Fusion with transfer loss and U-Net

西北工业大学计算机学院
School of Computer Science, Northwestern Polytechnical University

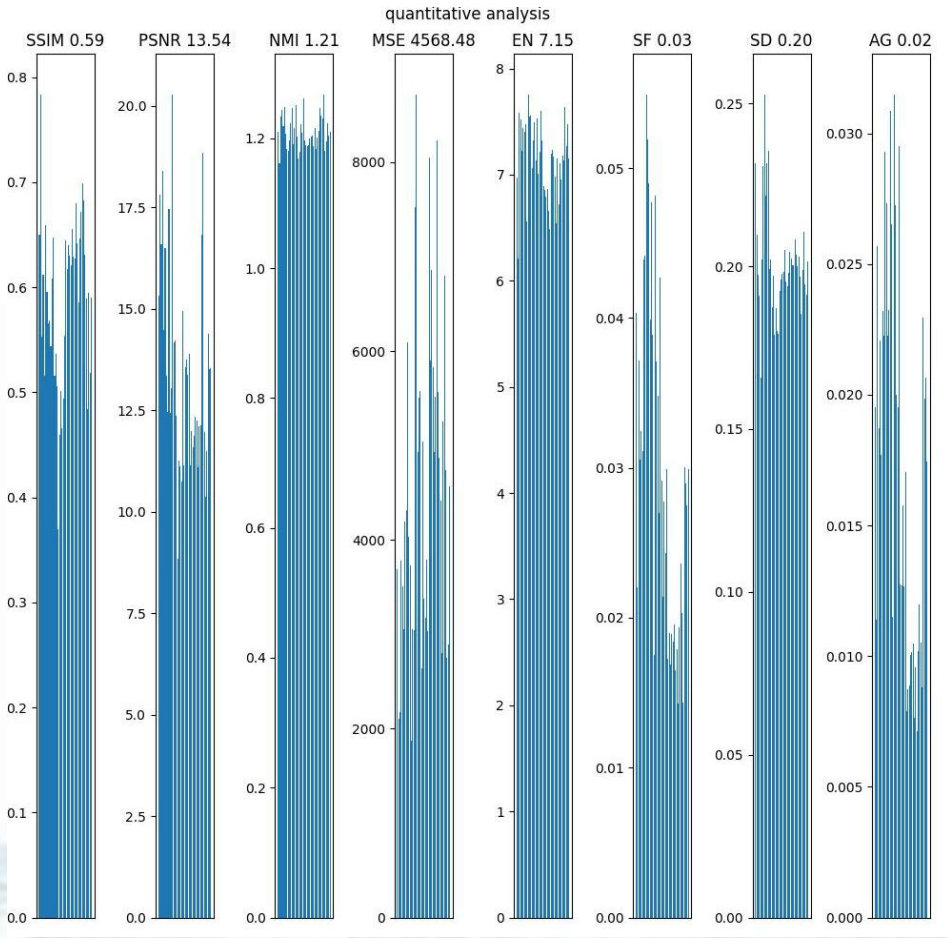


U_GAN

DDcGAN

U_GAN with style transfer loss

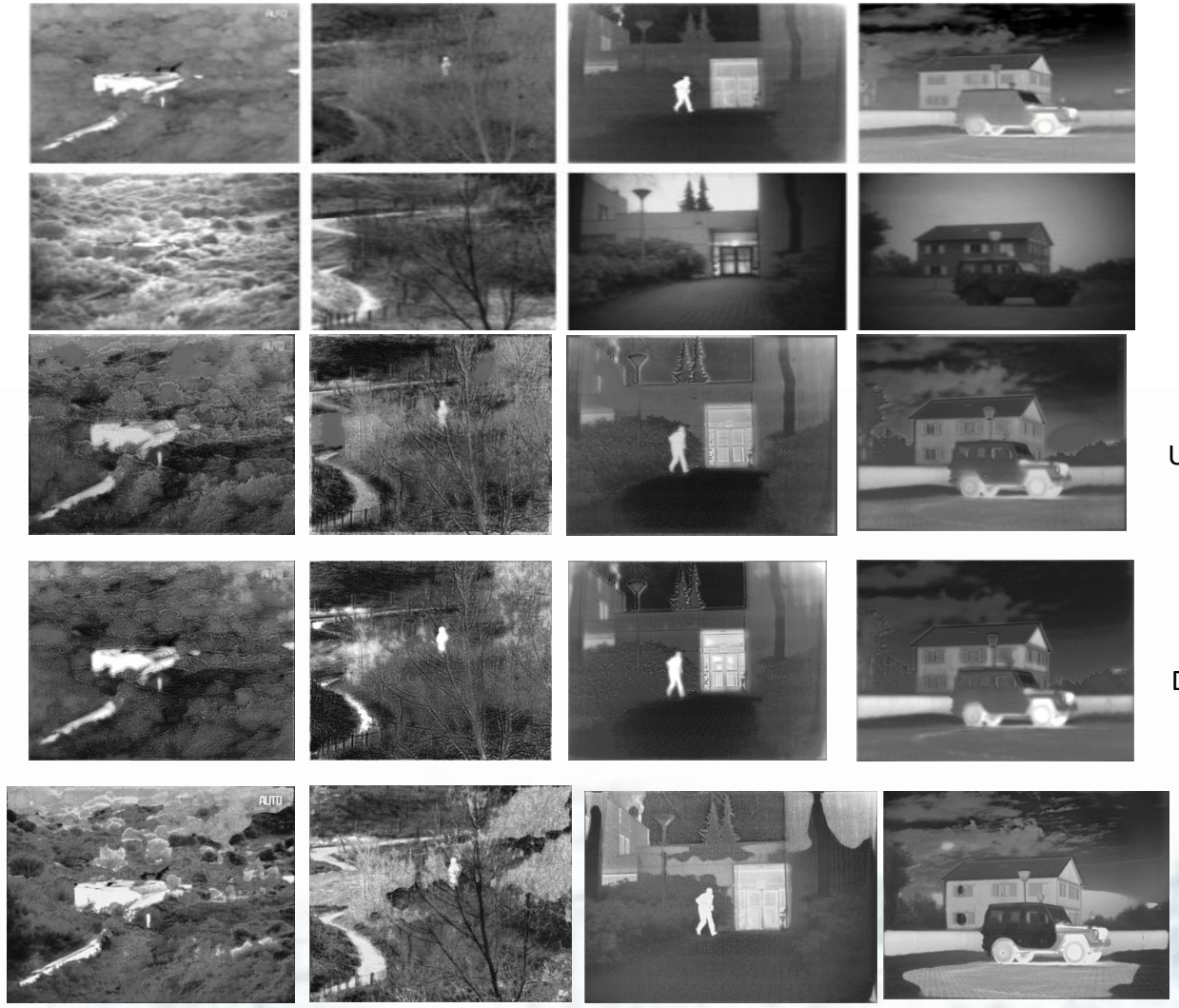
遇到的问题



融合的图像与可见光图像过于相似,即使在ssim上评价提升但没有太多实际意义

UFGAN:Stable Fusion with transfer loss and U-Net

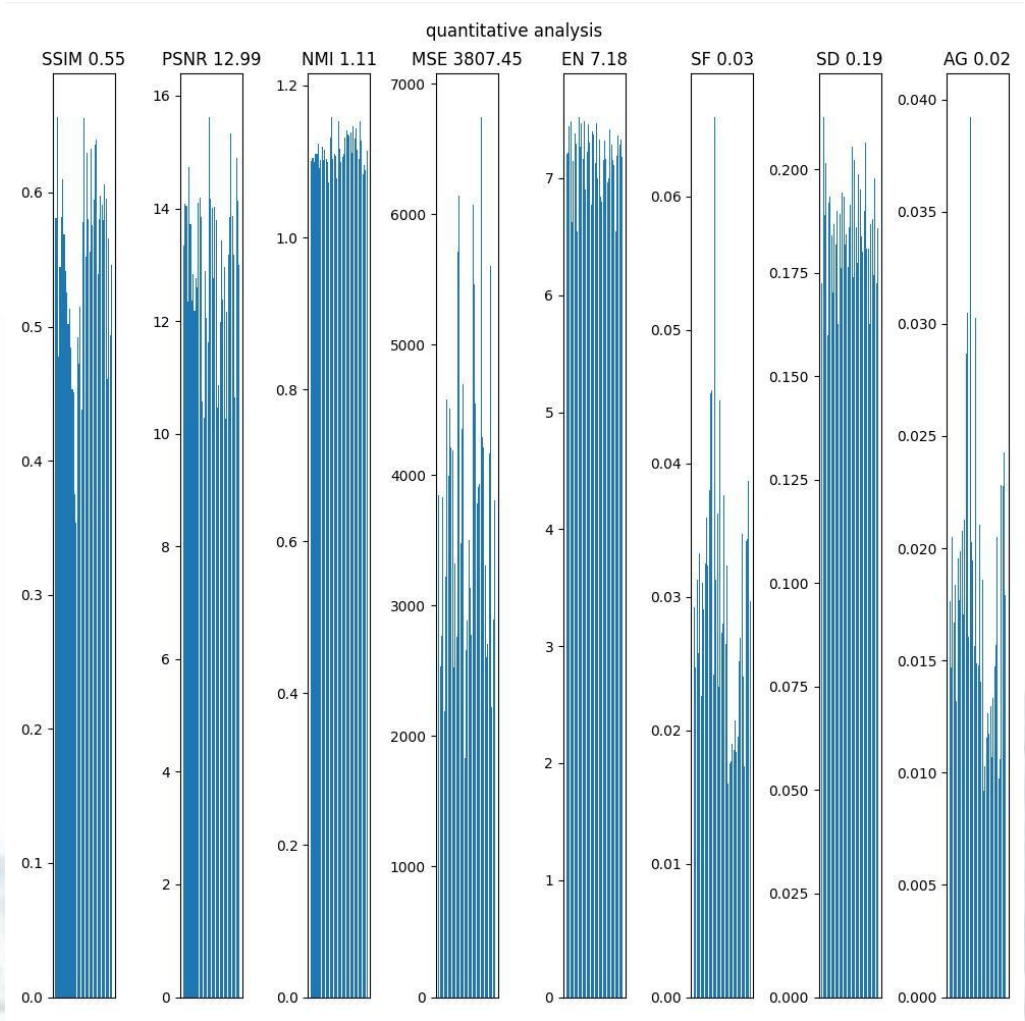
西北工业大学计算机学院
School of Computer Science, Northwestern Polytechnical University



U_GAN

DDcGAN

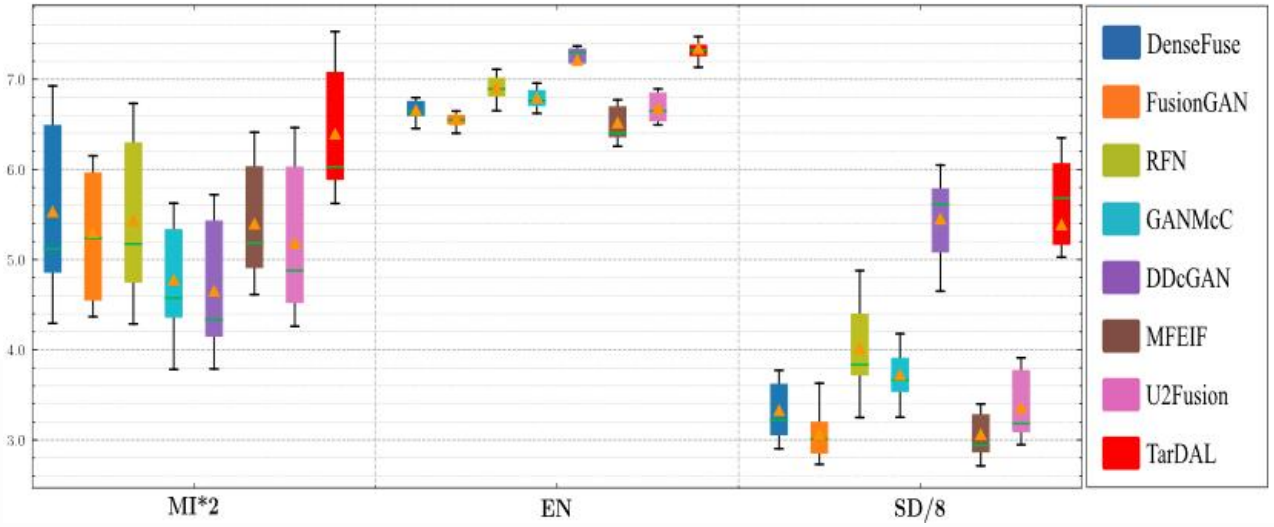
U_GAN with style transfer loss



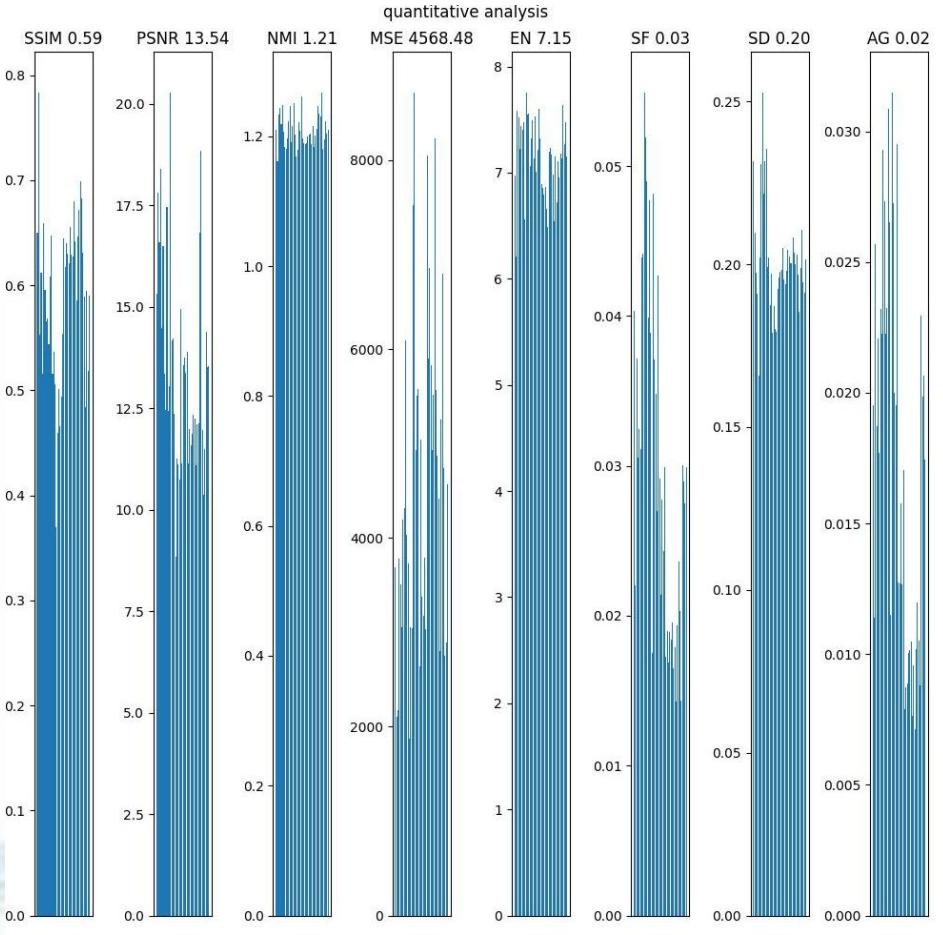
调参之后的结果

code link:https://github.com/drowning-in-codes/UFGAN.git
Blog:https://www.sekyoro.top

UFGAN:Stable Fusion with transfer loss and U-Net



from **Target-aware Dual Adversarial Learning and a Multi-scenario Multi-Modality Benchmark to Fuse Infrared and Visible for Object Detection (CVPR 2022)**



comparions between methods

	EN	SD/8
TarDAL	7.36	5.66
UFGAN(ours)	7.15	6.375

展望

1. 由于深度网络的学习效果与数据集中图像的数量和质量有关，因此有必要针对某些任务开发数据集。

2.目前**在几乎所有的图像融合任务中,基于深度学习的方法都假定源图像是预先配准好的。**然而,在真实场景中,由于视差、尺度差异和其他因素，多模态图像和数字摄影图像都无法进行配准。因此,现有深度学习方法中沿空间像素位置的操作无法用于现实源图像。虽然很多现成的方法可以用来对源图像进行预配准,但依赖于配准算法的预处理可能会导致某些局限性，如效率低和对注册精度的依赖性。因此，开发非配准融合算法，以隐式方式实现图像配准和融合是有必要的.

3. 由于大多数图像融合任务都没有真正的ground truth,因此评估融合结果的质量非常具有挑战性.一方面，所提出的指标可用于构建损失函数，以指导更高质量的融合。另一方面,新设计的指标还能公平地评价融合结果。即,设计更**合理更能解释的损失函数**,比如利用预训练目标检测网络的输出作为损失.

4.不同传感器采集的图像分辨率不同,可以考虑结合**超分**来克服分辨率差异问题,此外还有**在损失函数设计中引入后续任务(例如目标检测)的准确性**,从决策层面指导融合过程,**融合实时性**以及**统一的图像融合框架**等。
在模型设计上可以考虑:
1.利用transfer learning的思想,类似StyleGAN中的多尺度融合,将图像的不同模态作为一种style,设计网络保证content的不变而style改变。

2.利用cGAN的思想控制融合程度,人为添加标签作为不同的模态的标签。