引用格式：MA Decao, XIAN Yong, SU Juan, et al. Visible-to-infrared Image Translation Based on an Improved Conditional Generative Adversarial Nets[J]. Acta Photonica Sinica, 2023, 52(4):0410003
马得草,鲜勇,苏娟,等. 基于改进的条件生成对抗网络的可见光红外图像转换算法[J].光子学报,2023,52(4):0410003

基于改进的条件生成对抗网络的可见光红外图像转换算法

马得草[1],鲜勇[1],苏娟[2],李少朋[1],李冰[1]
(1 火箭军工程大学 作战保障学院,西安 710025)
(2 火箭军工程大学 核工程学院,西安 710025)

摘　要:针对当前基于条件生成对抗网络的红外图像生成算法中红外生成图像纹理细节信息差和结构信息差的问题,提出了一种基于改进的条件生成对抗网络的红外图像生成算法。首先,基于ConvNext改进了生成网络,在生成网络解码部分通过添加残差连接增强了解码部分对编码部分提取的图像深层特征的利用;其次,生成网络采用UNet网络架构,增强了对图像底层特征的利用;最后,对抗网络通过对生成图像特征的一阶统计量(均值)和二阶统计量(标准差)的损失计算,进一步改善了红外生成图像的灰度信息和纹理细节信息。与现有典型红外图像生成算法的对比实验结果表明,该方法能够生成质量更高的红外图像,在主观视觉描述和客观指标评价上都取得了更好表现。匹配应用实验表明,该算法在可见光图像与红外图像异源匹配任务中体现了较好的应用价值。

关键词:红外图像;可见光图像;生成对抗网络;生成网络;对抗网络;匹配应用评价
中图分类号:TN219;TP391.41　　**文献标识码**:A　　**doi**:10.3788/gzxb20235204.0410003

0　引言

利用可见光图像得到对应红外图像的方法,能够有效解决红外图像在红外制导、红外对抗和红外目标识别任务中数据缺乏的问题。目前,采用红外特性建模得到红外仿真图的方法,能够有效仿真得到目标的红外辐射特性。但是,该方法的仿真过程需要进行目标材质分类、分割等繁琐的操作,并且仿真得到的红外图像缺乏纹理信息。因此,需要探索一种高效准确的将可见光图像转换为对应红外图像的新范式。

基于生成对抗网络(Generative Adversarial Nets, GAN)[1]的图像生成技术为红外图像的生成提供了新的思路。特别地,Pix2Pix[2]和CycleGAN[3]为可见光图像转换为红外图像提供了通用网络框架,在文献[4-8]中均采用条件生成对抗网络的方式生成了红外图像。其中,ThermalGAN[4]和LayerGAN[5]通过多模态数据生成红外图像。ThermalGAN分两步生成红外图像,首先用可见光图像和温度矢量生成目标的平均温度红外图像,然后用目标的平均温度红外图像和可见光图像生成更加精细的红外图像。LayerGAN包括两种方式生成红外图像:一种方法是使用温度矢量、语义分割图像和热分割图像生成红外图像;另一种方法是使用可见光图像、语义分割图像和热分割图像生成红外图像。这两种方法对数据要求高,需要多种模态的数据,对于基于红外图像的行人重识别任务有着重要意义。文献[6-8]算法仅将可见图像作为输入,输出等效的红外图像。I-GANs[6]和Pix2pix-MRFFF[7]通过改进生成网络来改善红外生成图像的质量,而InfraGAN[8]基于UNetGAN[9]的思路是通过改进对抗网络来提高红外生成图像的质量。

本文针对现有方法生成的红外图像在不同程度上存在纹理不清晰、结构缺失的问题,将可见图像作为输入,输出等效的红外图像,通过改善生成网络和对抗网络来提高红外生成图像的质量,基于ConvNext[10]设

基金项目:国家自然科学基金(No. 62103432)
第一作者:马得草, madecaoedu@163.com
通讯作者:鲜勇, xy603xy@163.com
收稿日期:2022-11-08;**录用日期**:2022-12-24

计生成网络有效利用提取的可见光图像的深层特征和浅层特征，对抗网络通过对红外生成图像进行特征统计，对红外生成图像的灰度和结构信息加以引导，减弱对生成网络的约束，从而释放生成网络更大的潜能。

1 网络结构

本文提出的网络是基于条件生成对抗网络[11](Conditional Generative Adversarial Nets，CGAN)改进而来的，由基于ConvNext编码解码结构的生成网络和特征统计对抗网络组成。其结构如图1所示，生成网络的解码部分通过对编码部分深层特征和底层特征的利用改善了红外生成图像的质量，对抗网络通过对图像特征的一阶特征统计量和二阶特征统计量进行损失计算得到GAN损失，与L1损失结合作为总的损失，最终生成了更为清晰的红外图像。

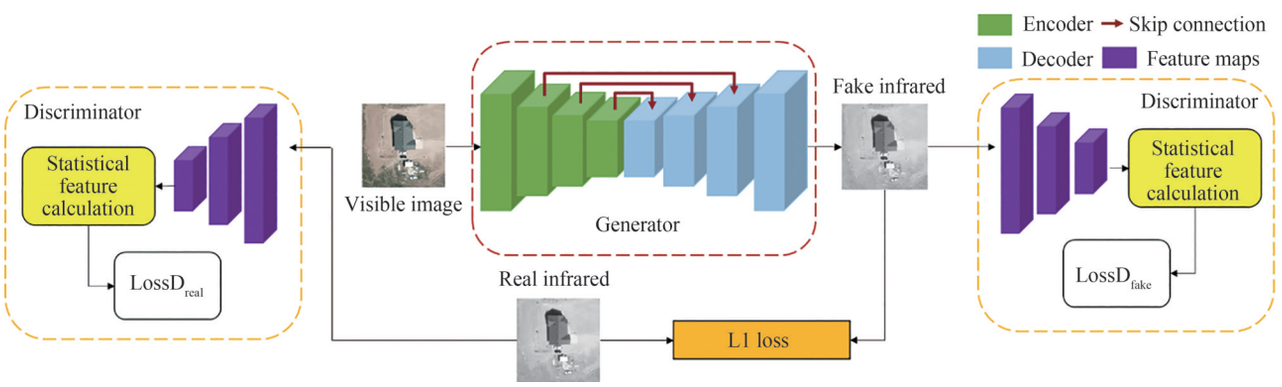


图1 算法结构

Fig. 1 Algorithm structure

1.1 生成网络

生成网络以256×256大小的可见光图像作为输入，将其转换为对应的红外图像，结构如图2所示。具体来讲，生成网络在编码部分通过步长为4和2的卷积对可见光图像进行下采样，通过卷积模块提取可见光图像的特征，如图2(b)所示，最终使用在ImageNet数据集上训练好的ConvNext_tiny作为网络的编码器。在解码部分，通过跳跃连接和残差连接分别加强了对编码部分提取的图像底层特征和深层特征的利用。在上采样过程中，通过上采样模块，如图2(c)和(d)所示，重建出256×256大小的红外图像。生成网络结构参数如表1所示。

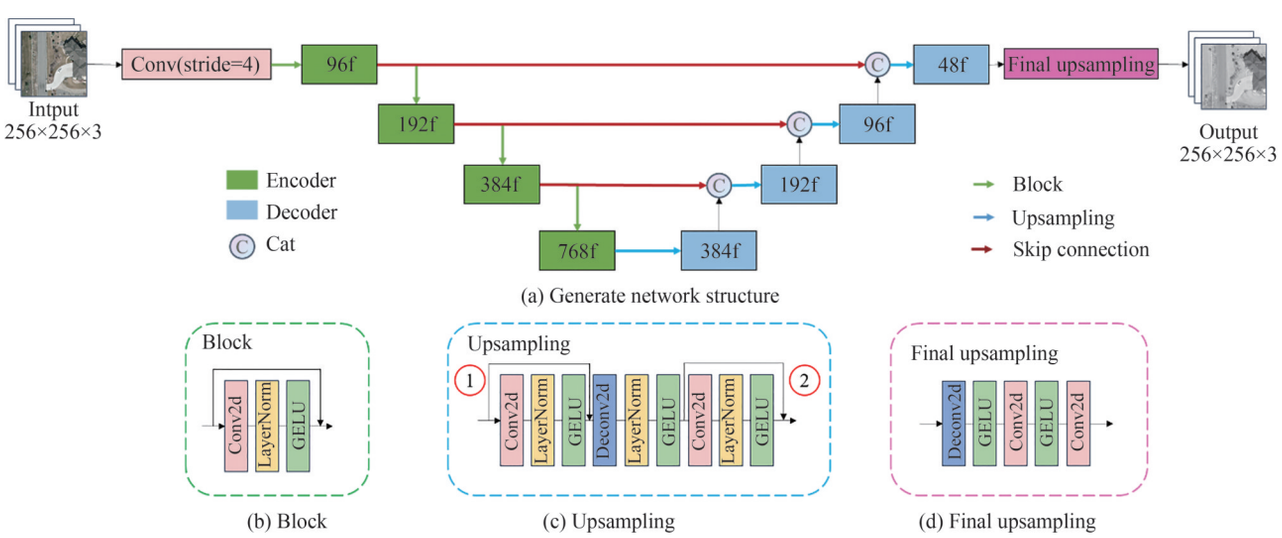


图2 生成网络

Fig.2 Generative Networks

表1　生成网络结构参数

Table 1　Generative networks structure parameter

Layers name	Output size	Networks layers
Down1	64×64	4×4Conv2d,96,Stride=4
Stages1	64×64	$\begin{bmatrix}7\times7,96\\7\times7,96\end{bmatrix}\times3$
Down2	32×32	2×2Conv2d,192,Stride=2
Stages2	32×32	$\begin{bmatrix}7\times7,192\\7\times7,192\end{bmatrix}\times3$
Down3	16×16	2×2Conv2d,384,Stride=2
Stages3	16×16	$\begin{bmatrix}7\times7,384\\7\times7,384\end{bmatrix}\times9$
Down4	8×8	2×2Conv2d,768,Stride=2
Stages4	8×8	$\begin{bmatrix}7\times7,768\\7\times7,768\end{bmatrix}\times3$
Dconv1	16×16	$\begin{bmatrix}1\times1,768,\text{Conv2d}\\3\times3,384,\text{ConvTranspose2d}\\1\times1,384,\text{Conv2d}\end{bmatrix}$
Dconv2	32×32	$\begin{bmatrix}1\times1,768,\text{Conv2d}\\3\times3,192,\text{ConvTranspose2d}\\1\times1,192,\text{Conv2d}\end{bmatrix}$
Dconv3	64×64	$\begin{bmatrix}1\times1,384,\text{Conv2d}\\3\times3,96,\text{ConvTranspose2d}\\1\times1,96,\text{Conv2d}\end{bmatrix}$
Dconv4	128×128	$\begin{bmatrix}1\times1,192,\text{Conv2d}\\3\times3,48,\text{ConvTranspose2d}\\1\times1,48,\text{Conv2d}\end{bmatrix}$
Final Dconv	256×256	$\begin{bmatrix}4\times4,24,\text{ConvTranspose2d}\\3\times3,12,\text{Conv2d}\\3\times3,3,\text{Conv2d}\end{bmatrix}$

该生成网络在编码部分主要采用的卷积是7×7的大卷积核,步长为1,填充为3,每个Stages的卷积通道数借鉴Tranformer的经验,如ConvNext_tiny的通道数分别设为96、192、384和768,且每个Stages中堆叠的卷积模块之比为3∶3∶9∶3。在解码部分,主要采用具有相同结构的上采用模块,包括两个1×1卷积核的卷积和一个3×3卷积核的反卷积。其中卷积主要用于对反卷积结果进行微调,反卷积主要用于扩充图像的尺寸,两个残差连接分别加强了编码部分提取的特征和反卷积后结果的利用。

图2(b)所示的卷积模块采用了ConvNext中的卷积模块设计,文献[10]证明了这种模块的设计方法在分类、检测和分割方面的优异表现。此外,文献[12]表明ConvNext可以像当前最先进的Transformer一样鲁棒和可靠,甚至更加可靠。UNet网络及其变体主要关注于编码部分和解码部分之间的跳跃连接,注重对图像底层特征的利用而缺少对图像深层特征的关注。图2(c)所示的上采样模块通过残差连接加强了对编码部分图像深层特征的利用。具体过程如式(1)所示。此外,上采样模块融合了卷积模块中一些有益的设计,如采用更少的归一化层,层归一化(Layer Normalization, LN)替代批量归一化(Batch Normalization, BN)等技巧。

$$U(F)=C(D(C(F_{\text{deep}})+F_{\text{deep}}))+D(C(F_{\text{deep}})+F_{\text{deep}})=C(F'_{\text{deep}})+F'_{\text{deep}} \tag{1}$$

式中,$U(\cdot)$表示上采样模块操作,$C(\cdot)$表示卷积模块操作,$D(\cdot)$表示采用转置卷积进行上采样,F_{deep}表示深层特征。最终上采样的结果与编码部分的底层特征进行拼接,进行下一步的上采样。生成网络一共进行了5次上采样,最后一次上采样采用图2(d)所示的上采样模块,逐渐降低卷积的通道数,平滑地生成图像。

1.2 对抗网络

提出的对抗网络称为特征统计网络(Statistical Feature Discriminator, SPatchGAN)是将一些常见的对抗网络如PatchGAN[13]专注于图像感受野特征的研究转化为对图像特征统计信息的研究,减小对生成网络的约束,从而释放生成网络更大的潜能。统计特征网络由特征提取层、特征计算部分和线性层组成,最后将图像的特征统计量作为损失进行监督。结构如图3所示。表2中展示了对抗网络的结构参数。具体地,对抗网络通过平均池化对图像进行降采样,在图像的三个不同尺度上通过相同结构的卷积层进行特征提取(表2中Conv1~Conv5),然后进行统计特征计算,最后通过线性层得出的损失相加得到总的损失。文献[8]表明,L1损失和基于结构相似性(Structural Similarity Index Measure, SSIM)的损失结合可以有效地提高红外生成图像的质量,其中评价指标SSIM由均值、标准差和协方差组成,即通过图像一阶信息和二阶信息的结合构建的。因此,选取图像特征的一阶统计量(均值)和二阶统计量(标准差)作为损失监督。

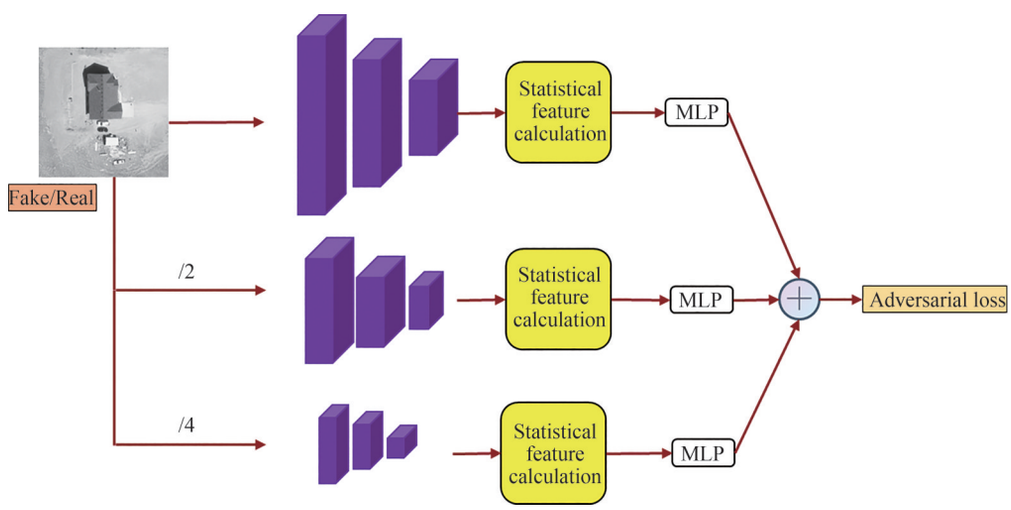


图3 对抗网络

Fig.3 Adversarial networks

表2 对抗网络结构参数

Table 2 Adversarial networks structure parameter

Layers name	Output size	Networks layers
Down	256×256, 128×128, 64×64	2×2AvgPool, Stride=2
Conv1	128×128, 64×64, 32×32	4×4Conv2d, 64, Stride=2
Conv2	64×64, 32×32, 16×16	4×4Conv2d, 128, Stride=2
Conv3	32×32, 16×16, 8×8	4×4Conv2d, 256, Stride=2
Conv4	32×32, 16×16, 8×8	1×1Conv2d, 512, Stride=1
Conv5	32×32, 16×16, 8×8	1×1Conv2d, 1 024, Stride=1

1.3 损失函数

仅使用L1损失和GAN损失用于生成高质量的红外图像,本文算法的损失函数表示为

$$G^{*}=\arg\min_{G}\max_{D}\ell_{\mathrm{CGAN}}(G,D)+\lambda\ell_{\mathrm{L1}} \tag{2}$$

$$\ell_{\mathrm{CGAN}}(G,D)=E_{x,y:p_{\mathrm{data}}}\left[\log D(x,y)\right]+E_{x:p_{\mathrm{data}}}\left[\log(1-D(x,G(x,z)))\right] \tag{3}$$

$$\ell_{\mathrm{L1}}=E\left[\left\|y-G(x,z)\right\|_{1}\right] \tag{4}$$

式中,G^{*}表示总的损失,ℓ_{CGAN}表示条件生成对抗网络的损失,ℓ_{L1}表示L1损失;$E(\cdot)$表示期望值,下标$x\sim P_{\mathrm{data}}$表示x取自可见光图像的数据,下标$x,y\sim P_{\mathrm{data}}$表示$x$取自可见光图像及$y$取自$x$对应的真实红外图像的数据;$\lambda$表示L1损失的权重,$y$表示标签信息(真实红外图像),$D(x,y)$表示判别器判断真实数据是否真实的概率;$G(x,z)$表示生成器根据源域图像$x$(可见光图像)生成的目标域图像$z$(红外图像),$D(x,G(x,z))$表示判别器判断生成数据$G(x,z)$是否真实的概率。

2 实验验证

2.1 数据集

实验涉及3个不同的数据集（VEDAI Dataset[14]、OSU Color-Thermal Dataset[15]和KAIST Dataset[16]），均由配准好的可见光图像和红外图像对组成。VEDAI数据集采集的是2012年美国犹他州AGRC卫星的春季图像，包含1 024×1 024（其中每个像素代表了12.5 cm×12.5 cm的区域）和512×512（其中每个像素代表了25 cm×25 cm的区域）两种尺寸的图像，红外图像为近红外图像。将VEDAI数据集中512×512尺寸的图像裁剪为包含目标的256×256尺寸的图像以适应网络输入。其中，1 046对用于训练，200对用于测试。OSU数据集常用于可见光图像和红外图像的融合，其中红外相机采用Raytheon PalmIR 250D，25 mm的镜头，光学相机采用Sony TRV87 Handycam，采样频率为30 Hz，得到的图像分辨率为320×240，最终拍摄得到的图像采用单应性矩阵和人工选点的方式进行配准。OSU数据集的拍摄场景是美国俄亥俄州立大学校园内繁忙的道路交叉口。OSU数据集图像被裁剪并放大为256×256尺寸以适应网络输入。因为该数据集得到的图像场景单一，对该数据进行了抽样处理。最终，683对图像用于训练，170对图像用于测试。KAIST数据集包含学校、街道和乡村的各种日常交通场景，一般用于行人检测，其中彩色相机是PointGrey Flea3，彩色图像尺寸为640×480，红外相机是FLIR-A35，红外图像尺寸为320×256，经过相机标定，最终得到640×512尺寸的图像对。实验中选择KAIST数据集中白天拍摄的图像，并被缩放为256×256尺寸以适应网络输入。经过抽样处理，减少大量重复的图像对，最终，5 008对用于训练，1 239对用于测试。

2.2 评价指标

6个客观评价指标用于评估生成图像质量，包括峰值信噪比（Peak Signal-to-Noise Ratio，PSNR）[17]、结构相似性（SSIM）、多尺度结构相似性（Multi-scale Structural Similarity Index Measure，MS-SSIM）[18]，学习感知图像块相似度（Learned Perceptual Image Patch Similarity，LPIPS）[19]、Fréhet Inception距离（Fréchet Inception Distance，FID）[20]和灰度相似性。PSNR和SSIM通常一起用于评估图像质量。多尺度结构相似性弥补了结构相似性在图像多尺度评价上的不足。LPIPS测量两个图像特征向量之间的欧几里得距离。为了计算指标，比较特征是从在ImageNet上预训练的基于卷积神经网络（Convolutional Neural Network，CNN）的主干中获得的（实验中使用了AlexNet网络模型）。FID是两个图像数据集之间相似性的度量，在本文被用于评估生成红外图像集和真实红外图像集的相似性，它被证明与人类对视觉质量的判断有很好的相关性，并且常用于评估生成对抗网络样本的质量。灰度相关性是采用灰度相关匹配中常用的归一化积相关算法中归一化相关性（Normalized Cross Correlation，NCC）作为评价指标。文献[21]中证明了红外图像的灰度特征与温度分布有着密切的关系，因而灰度相关性评价指标可以在一定程度上表示温度分布相关性。

2.3 训练细节

所有实验均在Intel（R）Core（TM）i9-10980XE CPU 3GHz和一块NVIDIA RTX 3090 GPU上运行，采用的深度学习框架是Pytorch。提出的基于改进的条件生成对抗网络的红外生成算法，使用了Adam优化器，其中β_1和β_2分别设置为0.5和0.999。网络训练共包含200轮训练，以确保模型收敛。其中，生成网络前100轮训练的学习率固定在0.000 2，然后在剩余的100轮训练中线性下降到0，对抗网络前100轮训练的学习率固定在0.000 002，然后在剩余的100轮训练中线性下降到0。

在本文算法中，L1损失和GAN损失对于网络的训练起着重要作用，设置λ作为L1损失的权重系数用于协调L1损失和GAN损失。L1损失设置过大将导致对抗网络不能有效工作，从而导致生成图像缺乏纹理信息，不添加L1损失或者L1损失设置过小将导致网络难以训练。图4表明当λ取100时，SSIM、MS-SSIM、PSNR、LPIPS、FID和NCC评价指标均取得了最好值，因而将λ设置为100。

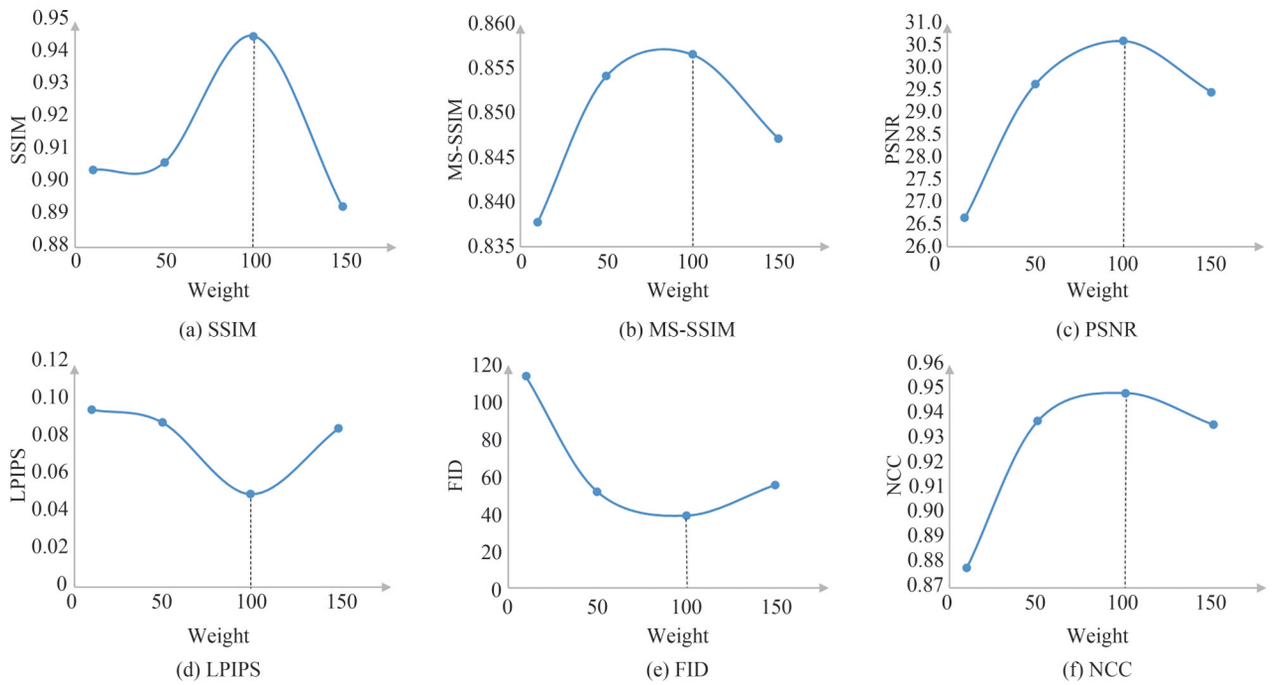


图 4 评价指标随权重系数 λ 的变化

Fig.4 Change of evaluation metrics with the weight coefficient λ

2.4 消融实验

为了验证本文方法的有效性，在VEDAI数据集上进行了消融实验，Pix2Pix作为基线模型分别验证了改进的生成网络和对抗网络的效果，为了公平起见消融实验中生成网络的编码部分没有加载预训练参数，将特征统计对抗网络（SPatchGAN）、无两种残差连接的上采样模块（Upsampling1）、仅有第2种残差连接的上采样模块（Upsampling2）、仅有第1种残差连接的上采样模块（Upsampling3）、有两种残差连接的上采样模块（Upsampling4）和Pix2Pix模型生成结果进行对比。主观评价结果如图5所示，客观评价结果如表3所示。

图 5 消融实验主观对比结果

Fig.5 The subjective comparison results of ablation experiment

表3　消融实验客观对比结果

Table 3　The objective comparison results of ablation experiment

Evaluation indexes	SSIM ↑	MS-SSIM ↑	PSNR ↑	LPIPS ↓	FID ↓	NCC ↑
Pix2Pix	0.910 3	0.799 4	26.979 9	0.085 6	51.231 1	0.931 1
Baseline＋SPatchGAN	0.918 9	0.834 6	27.247 8	0.077 0	48.790 9	0.937 0
Baseline＋Upsampling1	0.920 2	0.853 4	30.086 3	0.057 9	43.034 6	0.941 4
Baseline＋Upsampling2	0.928 0	0.855 0	29.937 5	0.055 3	42.220 2	0.942 9
Baseline＋Upsampling3	0.943 9	0.859 5	30.580 4	0.050 7	38.969 9	0.948 9
Baseline＋Upsampling4	0.945 4	0.856 6	30.620 7	0.050 3	38.895 5	0.948 7

观察图5的生成结果可以发现，采用本文提出的统计特征对抗网络生成的红外图像在图像的灰度信息上保持较好（图5(d)红色框所示），原因在于本文对抗网络中采用均值和标准差作为统计量，均值有助于引导图像生成过程中灰度信息的保持，标准差有助于图像生成过程中结构的保持。这在文献[22]中也得到了证明。图5(e)～(h)中红色框结果显示了本文提出的上采样模块对于红外图像生成过程中的灰度真实性有较大的改善。

从表3中的数据对比可以发现，SPatchGAN对MS-SSIM的提升较大，这主要在于其相比PatchGAN提取了生成图像的多尺度特征，从不同尺度的角度分别约束了图像的生成。从表3中的第4行至第7行数据观察可以发现，提出的几种上采样模块对5种指标的提升均有帮助，从4种上采样模块的对比中可以看出上采样模块中第一种残差连接的效果更加显著（Upsampling3的结果相比Upsampling2的结果指标提升更明显），第一种残差连接主要将编码部分的深层特征传入解码部分，说明加强对生成网络编码部分提取图像深层特征的利用有助于改善生成图像的质量。

2.5　对比实验

为了验证本文方法的优越性，在3个数据集上进行了实验验证，将本文方法与基于条件生成对抗网络的红外图像生成算法Pix2Pix、ThermalGAN、I-GANs、InfraGAN进行对比实验。ThermalGAN使用了可见光图像和温度矢量作为输入，为了对比的公平，实验中的ThermalGAN仅使用了可见光图像作为输入。InfraGAN的输入是512×512大小的图像，对InfraGAN进行了修改以适应256×256大小图像的输入。图6展示了实验的主观对比结果，表4展示了客观评价结果。

图6中3个数据集典型红外图像的生成结果表示5种模型在VEDAI数据集上的生成效果均较好，5种模型在VEDAI数据集上的生成结果主要体现在图像的灰度信息差别上，从最终结果可以看出本文算法生成的红外图像灰度信息更加准确。OSU数据集场景较为单一，本文算法在OSU数据集上对于微小物体有着更好的生成效果，细节更加分明。本文方法在KAIST数据集上相比其他4个模型，红外生成图像中的物体纹理更清晰，细节信息更丰富。

表4数据表明本文方法在3个不同数据集上有着最好的表现，尤其在高分辨率的VEDAI数据集上。从对比结果来看，随着图像分辨率的增加，算法的性能出现了下降，这将在下一步研究中进行解决。综合主观和客观对比结果，本文算法的红外生成图像质量优于其他算法的生成结果。

表5中展示了5种算法的不同结构，本文算法仅使用L1损失和GAN损失对生成网络进行指导，便取得了较好的红外图像生成效果。算法设计时没有考虑网络规模和训练耗时的因素，下一步将在网络轻量化和减少网络训练耗时方面进行改善。

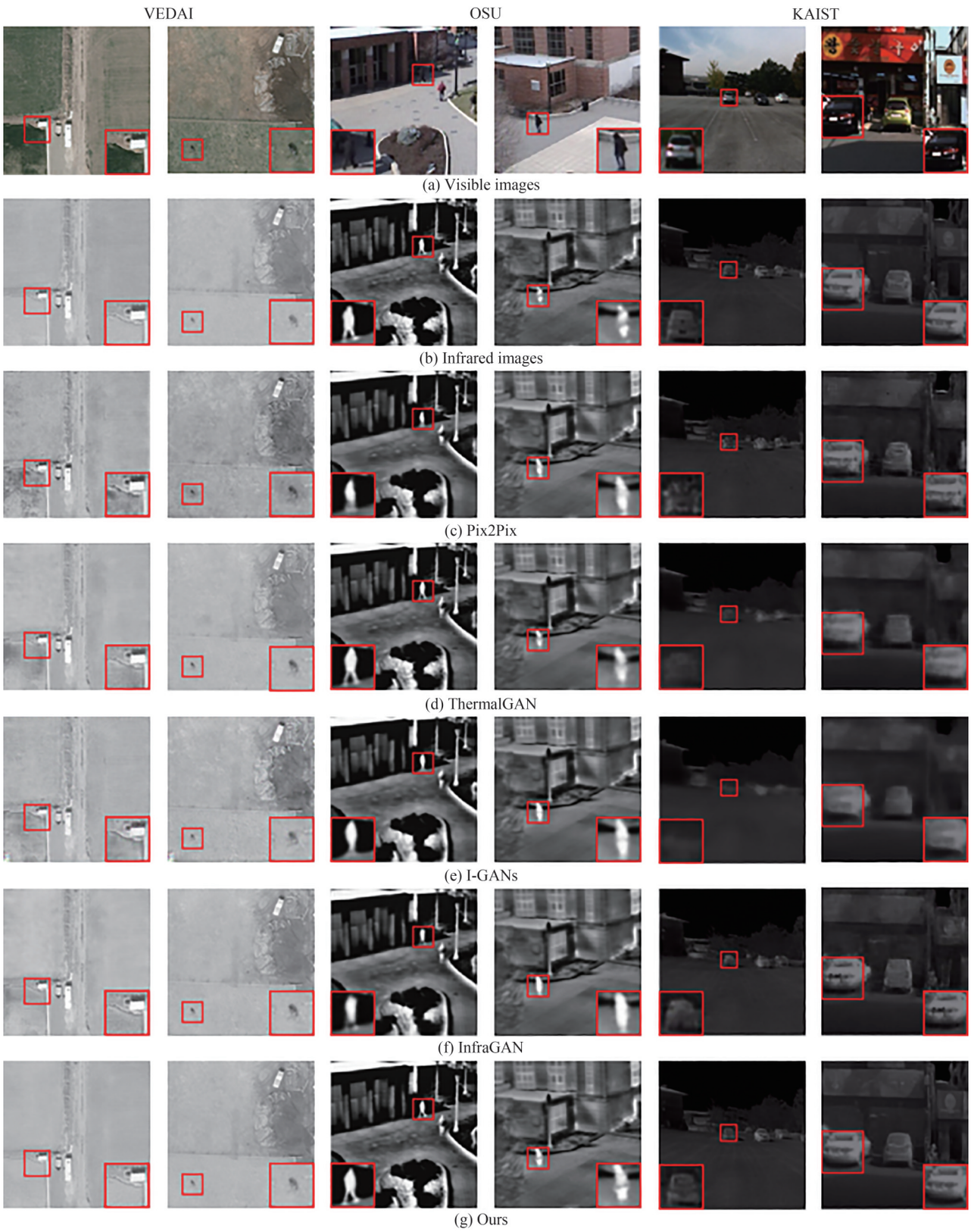


图 6　不同算法的主观实验结果对比

Fig.6　The subjective experiment results comparison of different algorithms

表4　不同算法的客观实验结果对比

Table 4　The objective experiment results comparison of different algorithms

VEDAI	SSIM ↑	MS-SSIM ↑	PSNR ↑	LPIPS ↓	FID ↓	NCC ↑
Pix2Pix	0.910 3	0.799 4	26.979 9	0.085 6	51.231 1	0.931 1
ThermalGAN	0.946 6	0.853 9	30.048 9	0.052 5	39.967 8	0.947 3
I-GANs	0.814 6	0.744 2	27.287 7	0.158 1	119.674 3	0.893 5
InfraGAN	0.948 7	0.858 5	29.987 7	0.051 1	39.796 1	0.948 0
Ours	**0.955 1**	**0.881 6**	**31.329 4**	**0.042 3**	**33.954 0**	**0.960 4**
Ours(512×512)	0.950 8	0.854 7	31.223 7	0.059 7	37.853 0	0.960 1
Ours(1024×1024)	0.955 5	0.848 1	31.864 9	0.085 6	53.615 7	0.951 0
OSU	SSIM ↑	MS-SSIM ↑	PSNR ↑	LPIPS ↓	FID ↓	NCC ↑
Pix2Pix	0.901 5	0.885 4	24.187 6	0.118 8	84.363 2	0.975 6
ThermalGAN	0.904 3	0.906 1	29.239 8	0.132 2	78.079 8	0.974 5
I-GANs	0.881 2	0.867 7	27.767 0	0.160 3	107.761 4	0.964 3
InfraGAN	0.905 1	0.904 6	29.025 5	0.137 4	89.047 5	0.973 1
Ours	**0.923 4**	**0.936 8**	**31.331 2**	**0.064 5**	**57.973 5**	**0.983 9**
KAIST	SSIM ↑	MS-SSIM ↑	PSNR ↑	LPIPS ↓	FID ↓	NCC ↑
Pix2Pix	0.827 7	0.593 1	22.732 8	0.201 7	77.878 9	0.922 6
ThermalGAN	0.854 1	0.613 8	23.550 4	0.283 1	112.654 8	0.937 1
I-GANs	0.844 1	0.572 0	22.950 3	0.361 8	156.585 2	0.927 9
InfraGAN	0.844 5	0.608 2	23.021 4	0.181 0	63.930 8	0.929 5
Ours	**0.869 2**	**0.700 8**	**24.452 4**	**0.112 3**	**27.533 1**	**0.948 3**

表5　不同算法的网络结构

Table 5　Network structure of different algorithms

Method	Generator	Discriminator	Loss	
			L1 loss	SSIM loss
Pix2Pix	ResNet9block	PatchGAN	√	
ThermalGAN	UNet	PatchGAN	√	
I-GANs	D-Linket34	PatchGAN	√	
InfraGAN	UNet	UNetGAN	√	√
Ours	UConvNext	SPatchGAN	√	

3　匹配应用实验

为了验证本文算法在可见光图像与红外图像异源匹配方面的应用价值，采用3种经典的传统匹配算法和3种较为先进的基于深度学习的匹配算法进行实验验证。传统匹配算法分别是尺度不变特征变换(Scale-Invariant Feature Transform，SIFT)算法、SURF(Speeded Up Robust Features)算法和ORB(Oriented FAST and Rotated BRIEF)算法，基于深度学习的匹配算法分别是D2-Net算法[23]、SuperGlue算法[24]和LoFTR算法[25]。匹配实验结果采用匹配终点误差(Matching end Point Error，EPE)进行评价，表示为

$$\mathrm{EPE}=\frac{1}{n}\sum_{i=1}^{n}\sqrt{\left(x_i-\hat{x}_i\right)^2-\left(y_i-\hat{y}_i\right)^2} \tag{5}$$

式中，n代表匹配成功的点数，(x_i,y_i)代表特征点在目标图像上的位置，$(\hat{x}_i,\hat{y}_i)$代表相应特征点经过真实单应性矩阵转换后的位置。所使用的数据集均是配准好的图像对，因此单应性矩阵实际上是一个单位矩阵。

3.1　传统匹配算法

SIFT算法是通过求一幅图中的特征点及其有关尺度和方向的描述子得到特征，并进行图像特征点匹配。SIFT所查找到的关键点是一些十分突出，不会因光照、仿射变换和噪音等因素而变化的点。SURF算法是一种稳健的局部特征点检测和描述算法。SURF的出现很大程度是对SIFT算法的改进，用一种更为高效的方式改进了特征的提取和描述方式。ORB算法是一种快速特征点提取和描述的算法，其特征检测是

将FAST特征点的检测方法与BRIEF特征描述子结合起来，并在它们原来的基础上做了改进与优化。采用不同的传统匹配算法（SIFT算法、SURF算法和ORB算法）进行图像特征提取，之后用K近邻算法匹配特征点，并用随机采样一致性算法剔除错误的匹配点得到最终的匹配结果。特别地，由于算法限制，其中部分图像无法匹配，因此在评价生成红外图像的传统匹配算法结果时，只计算了成功匹配图像对的匹配终点误差。

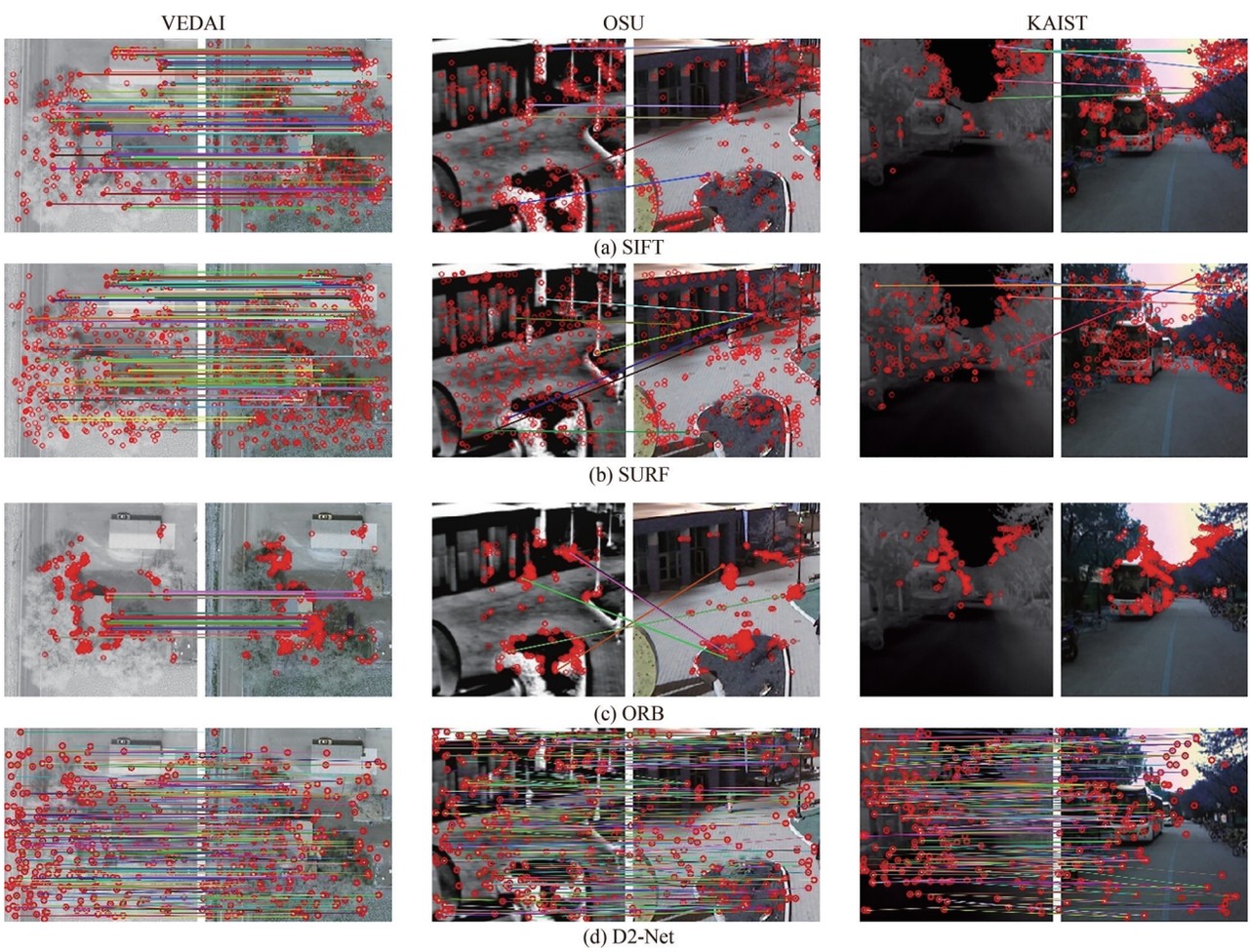
3.2 基于深度学习的匹配算法

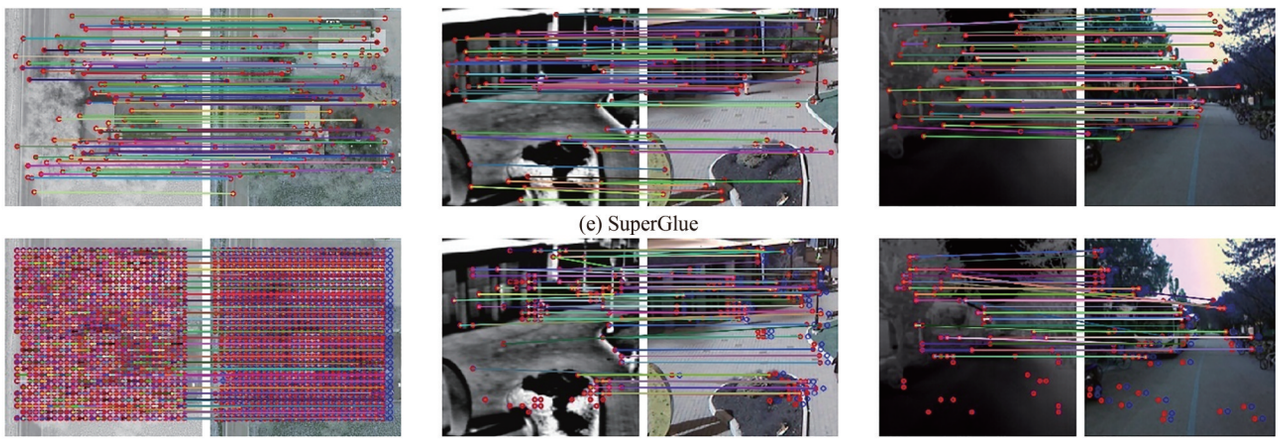
D2-Net[23]直接从特征描述子进行关键特征检测，将检测放在处理的后期阶段，从而获得更稳定的关键点，解决了在困难的成像条件下找到可靠像素级匹配的问题。SuperGlue[24]提出了一种基于图卷积神经网络的特征匹配算法，采用SuperPoint提取特征点及描述符，通过求解可微分最优化转移问题实现特征匹配。LoFTR[25]是一种不依赖关键点检测的端到端特征匹配方法，利用卷积神经网络初步提取特征，再利用Transformer的全局注意力加强特征，可以较好地对低纹理图片和相似区域进行匹配。使用这3种匹配算法时，均采用它们提供的预训练权重进行匹配实验。

3.3 匹配实验验证

为了验证本文算法在可见光图像和红外图像异源匹配任务中的应用价值，在3个数据集上对生成的结果进行了匹配实验，分别对比了Pix2Pix、ThermalGAN、I-GANs、InfraGAN和本文算法生成的红外图像与真实红外图像之间的匹配结果。此外，还添加了可见光图像和红外图像的匹配实验作为对照。图7展示了6种匹配方法在可见光图像与红外图像之间的匹配结果，图8展示了6种匹配方法在相同红外图像之间的匹配结果。表6展示了6种匹配方法的匹配终点误差。

图7和图8每组匹配图像对中左侧图像为真实红外图像作为基准图像，右侧为待匹配图像。对比发现，6种算法在同源图像的匹配上相比异源图像提取的特征点更多且错误匹配的情况更少。从图7可以发现基于深度学习的匹配算法相比传统匹配算法在异源图像的匹配上提取的特征点更多且误匹配现象更少。特别地，传统图像匹配算法在异源图像匹配过程中可能会出现失败的现象，如图7(c)KAIST数据集上的结果所示。



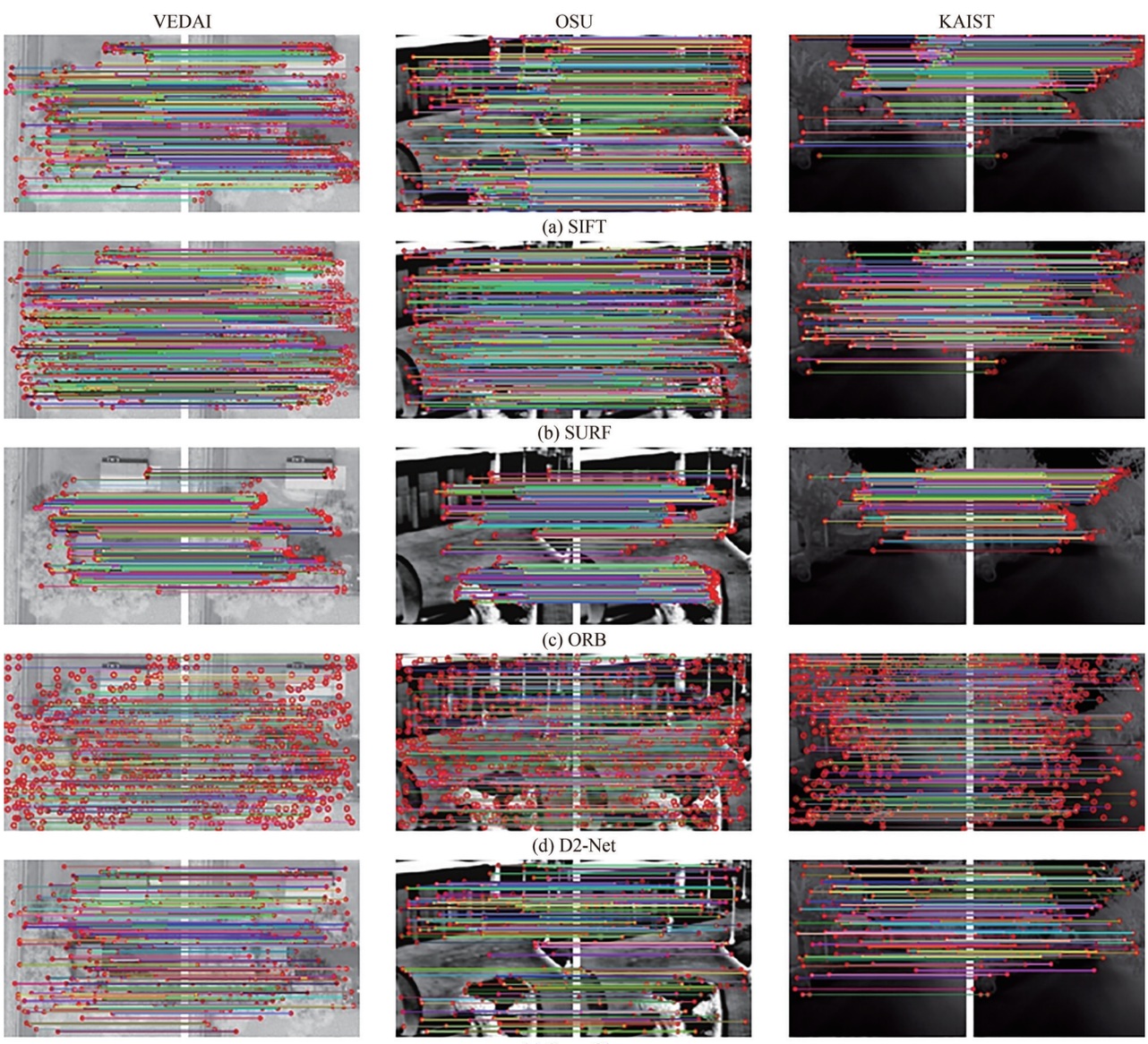


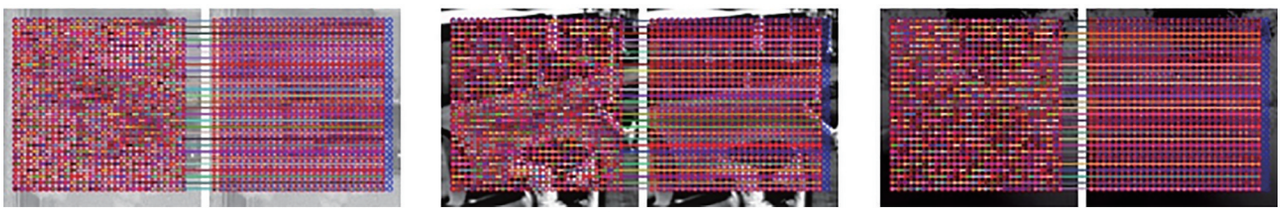
(e) SuperGlue

(f) LoFTR

图7　6种匹配算法在异源匹配中的表现

Fig.7　Performance of six matching algorithms in heterogeneous matching





(f) LoFTR

图8 6种匹配算法在红外图像匹配上的表现

Fig.8 Performance of six matching algorithms in infrared image matching

表6 6种匹配算法的匹配终点误差

Table 6 EPE of six matching algorithms

EPE VEDAI	SIFT	SURF	ORB	D2-Net	SuperGlue	LoFTR
RGB	0.131 7	2.334 4	8.151 8	0.767 3	0.334 8	0.136 5
Pix2Pix	0.311 6	0.245 3	0.660 1	0.387 1	0.477 0	0.052 6
ThermalGAN	0.118 0	0.170 6	0.196 7	0.284 3	0.302 3	0.027 2
I-GANs	0.631 0	0.795 4	1.029 8	0.542 0	0.433 1	0.070 4
InfraGAN	0.109 6	0.166 2	0.287 3	0.265 9	0.314 5	0.021 6
Ours	**0.081 8**	**0.154 7**	**0.106 5**	**0.238 3**	**0.257 1**	**0.017 8**
EPE OSU	SIFT	SURF	ORB	D2-Net	SuperGlue	LoFTR
RGB	52.614 8	28.740 6	38.851 5	1.695 6	0.877 6	2.533 4
Pix2Pix	0.105 3	0.099 9	0.125 5	0.218 2	0.178 0	0.043 3
ThermalGAN	0.126 1	0.105 0	0.138 0	0.233 4	0.173 5	0.042 7
I-GANs	0.147 1	0.117 4	0.154 0	0.259 3	0.212 5	0.058 7
InfraGAN	0.122 7	0.105 6	0.140 0	0.224 6	0.163 8	0.040 0
Ours	**0.080 8**	**0.078 9**	**0.088 8**	**0.167 1**	**0.145 6**	**0.034 9**
EPE KAIST	SIFT	SURF	ORB	D2-Net	SuperGlue	LoFTR
RGB	40.716 0	36.108 2	33.319 1	2.765 0	4.387 7	10.397 7
Pix2Pix	2.885 9	1.744 7	2.430 1	0.855 9	1.045 2	0.546 8
ThermalGAN	1.498 3	1.533 5	1.055 8	0.970 2	0.944 1	0.807 6
I-GANs	2.089 5	2.399 4	1.814 8	1.201 2	1.523 7	2.574 3
InfraGAN	2.024 1	1.438 6	2.184 2	0.726 6	0.803 6	0.385 4
Ours	**0.630 3**	**0.552 1**	**0.622 2**	**0.477 9**	**0.495 5**	**0.181 3**

从表6的结果可以发现，LoFTR匹配算法在3个数据集上表现最好。相比采用可见光图像进行匹配，利用图像转换算法将可见光图像转换的对应红外图像进行匹配，可以有效降低匹配终点误差。对比表4和表6的数据可以发现，匹配终点误差和6种客观评价指标并不成严格的正比关系，但是具有正相关关系，即一般客观评价指标表现越好，相应的匹配终点误差也越小。从表4和表6的结果中可以发现，本文算法不仅主客观上表现较好，而且在匹配任务中也有着较好的表现。

4 结论

本文提出了一种基于改进的条件生成对抗网络的可见光红外图像转换算法，用于解决当前典型红外生成算法中生成图像纹理细节信息差和结构信息差的问题。该算法提出的生成网络不仅注重对图像底层特征的利用，而且加强了对图像深层特征的利用。对抗网络通过图像特征的一阶统计量（均值）来引导红外图像在生成过程中产生更加真实的灰度信息，通过图像特征的二阶统计量（标准差）来引导红外图像在生成过程中保持结构信息。实验结果表明，所提算法生成了纹理细腻、结构清晰的红外图像，并且生成的红外图像

在可见光图像与红外图像匹配方面有较好的应用价值,适用于室外温度相对固定情况下的可见光图像和对应红外图像的转换。

参考文献

[1] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.

[2] ISOLA P, ZHU Junyan, ZHOU Tinghui, et al. Image-to-image translation with conditional adversarial networks[C]. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017:5967-5976.

[3] ZHU Junyan, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]. Proceedings of the 2017 IEEE Conference on Computer Vision (ICCV), 2017:2223-2232.

[4] KNIAZ V V, KNYAZ V A, HLADUVKA J, et al. Thermalgan: multimodal color-to-thermal image translation for person re-identification in multispectral dataset[C]. Proceedings of the European Conference on Computer Vision (ECCV) Workshops, 2018: 606-624.

[5] MIZGINOV V A, KNIAZ V V, FOMIN N A. A method for synthesizing thermal images using gan multi-layered approach[J]. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2021, 4421: 155-162.

[6] LI Bing, XIAN Yong, ZHANG Daqiao. Infrared image generation algorithm based on conditional generation adversarial networks[J]. Acta Photonica Sinica, 2021, 50(11): 1110004.
李冰,鲜勇,张大巧 . 基于条件生成对抗网络的红外图像生成算法[J]. 光子学报, 2021, 50(11):1110004.

[7] MA Yangyang, HUA Yanling, ZUO Zhengrong. Infrared image generation by pix2pix based on multi-receptive field feature fusion[C]. 2021 International Conference on Control, Automation and Information Sciences (ICCAIS), 2021: 1029-1036.

[8] ÖZKANOĞLU M A, OZER S. InfraGAN: A GAN architecture to transfer visible images to infrared domain[J]. Pattern Recognition Letters, 2022, 155: 69-76.

[9] SCHONFELD E, SCHIELE B, KHOREVA A. A u-net based discriminator for generative adversarial networks[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 8207-8216.

[10] LIU Zhuang, MAO Hanzi, WU Chaoyuan, et al. A convnet for the 2020s[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 11976-11986.

[11] MIRZA M, OSINDERO S. Conditional generative adversarial nets[DB/OL]. [2022-11-08]. https://arxiv.org/abs/1411.1784.

[12] PINTO F, TORR P H S, DOKANIA P K. An impartial take to the CNN vs transformer robustness contest[DB/OL]. [2022-11-08]. https://arxiv.org/abs/2207.11347.

[13] LI Chuan, WAND M. Precomputed real-time texture synthesis with markovian generative adversarial networks[C]. Proceedings of the 2016 European Conference on Computer Vision (ICCV), 2016:702-716.

[14] RAZAKARIVONY S, JURIE F. Vehicle detection in aerial imagery: a benchmark[J]. Journal of Visual Communication and Image Representation, 2015, 34:187-203.

[15] DAVIS J W, SHAARMA V. Background-subtraction using contour-based fusion of thermal and visible imagery[J]. Computer Vision and Image Understanding, 2007, 106(2-3): 162-182.

[16] HWANG S, PARK J, KIM N, et al. Multispectral pedestrian detection: benchmark dataset and baseline[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1037-1045.

[17] HORE A, ZIOU D. Image quality metrics: PSNR vs. SSIM[C]. 2010 20th International Conference on Pattern Recognition, 2010: 2366-2369.

[18] WANG Z, SIMONCELLI E P, BOVIK A C. Multiscale structural similarity for image quality assessment[C]. The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003, 2: 1398-1402.

[19] ZHANG R, ISOLA P, EFROS A A, et al. The unreasonable effectiveness of deep features as a perceptual metric[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 586-595.

[20] HEUSEL M, RAMSAUER H, UNTERTHINER T, et al. GANs trained by a two time-scale update rule converge to a local nash equilibrium[C]. Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017: 6629-6640.

[21] GONG Jiamin, GUO Tao, CAO Yi, et al. Correlativity analysis between image gray value and temperature based on infrared target[J]. Infrared and Laser Engineering, 2016, 45(3): 0304006.
巩稼民, 郭涛, 曹懿, 等 . 红外靶标的图像灰度与温度相关性剖析[J]. 红外与激光工程, 2016, 45(3): 0304006.

[22] ZHANG Yabin, LI Minghan, LI Ruihuang, et al. Exact feature distribution matching for arbitrary style transfer and domain generalization[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 8035-8045.

[23] DUSMANU M, ROCCO I, PAJDLA T, et al. D2-net: a trainable CNN for joint description and detection of local features[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 8092-8101.
[24] SARLIN P E, DETONE D, MALISIEWICZ T, et al. Superglue: learning feature matching with graph neural networks[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 4938-4947.
[25] SUN J, SHEN Z, WANG Y, et al. LoFTR: detector-free local feature matching with transformers[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 8922-8931.

Visible-to-infrared Image Translation Based on an Improved Conditional Generative Adversarial Nets

MA Decao[1], XIAN Yong[1], SU Juan[2], LI Shaopeng[1], LI Bing[1]
(1 *College of War Support, Rocket Force University of Engineering, Xi'an* 710025, *China*)
(2 *College of Nuclear Engineering, Rocket Force University of Engineering, Xi'an* 710025, *China*)

Abstract: Using visible images to obtain corresponding infrared images is an effective approach to address the lack of infrared images in infrared guidance, infrared countermeasures, and infrared object recognition tasks. At present, the infrared radiation properties of the target can be efficiently simulated by current methods that use modeling of infrared properties to obtain infrared simulation images. However, the simulation process of this method requires tedious operations such as the classification and segmentation of target materials, and the infrared images obtained by the simulation lack texture information. The infrared image generation algorithm based on Generative Adversarial Networks (GAN) can effectively alleviate the problems of cumbersome and labor-intensive infrared image generation. However, some current infrared image generation algorithms based on GANs are prone to the problems of lack of image detail information and lack of structural information. This paper proposes a visible-to-infrared image translation algorithm based on an improved Conditional Generative Adversarial Nets (CGAN). Different from the current UNet network and its variants which focus on the utilization of the underlying features of the image, the generative network not only focuses on the utilization of the underlying features of the image, but also strengthens the utilization of the underlying features of the image. In addition, some network tricks of the ConvNext network are incorporated. Techniques such as using fewer normalization layers, layer normalization instead of batch normalization, etc. The adversarial network improves the quality of the generated images by calculating the first-order feature statistics (mean) and second-order feature statistics (standard deviation) of the generated images. The mean value contributes to the generation of grayscale information of infrared images, and the standard deviation contributes to the generation of structural information of infrared images. The research of the adversarial network focusing on the image receptive field features is transformed into the research of the image feature statistical information, which reduces the constraints on the generative network and releases the greater potential of the generative network. In the experiment, three datasets were used, namely the VEDAI dataset, the OSU dataset and the KAIST dataset, and six objective evaluation metrics were used to evaluate the quality of the generated images, including peak signal-to-noise ratio, structural similarity, multi-scale structural similarity, learning perceptual image patch similarity, Fréchet inception distance and normalized cross correlation. Compared with existing typical infrared image generation algorithms, the experimental results show that the proposed method can generate higher quality infrared images and achieve better performance in both subjective visual description and objective metric evaluation. In the matching application experiment, this paper adopts three traditional matching algorithms: SIFT algorithm, SURF algorithm and ORB algorithm, and three matching algorithms based on deep learning: the D2-Net algorithm, SuperGlue algorithm and LoFTR algorithm. The experimental results show that compared with the use of the visible image for matching, the image conversion algorithm is used to match the corresponding infrared image converted by the visible image, which can effectively reduce the matching endpoint error. The experimental results show that matching end point error is not strictly proportional to the six objective evaluation indexes, but there is a positive correlation between matching end point error and objective evaluation indexes. In general, the better the performance of objective evaluation indicators, the smaller the corresponding matching end point

error. In summary, this paper proposes an improved conditional generative adversarial network for converting visible images to corresponding infrared images. The proposed method effectively alleviates the problems of the lack of texture detail information and the lack of structural information in the current infrared generation algorithm based on conditional generative adversarial network image generation in the process of image generation. The generated infrared image has good application value in the matching of the visible image and the infrared image. In addition, this paper provides new ideas for other image translation tasks.

Key words: Infrared image; Visible image; Generative adversarial networks; Generative networks; Adversarial networks; Matching application evaluation

OCIS Codes: 100.3008; 110.3080; 150.1135; 350.2660

Foundation item: National Natural Science Foundation of China (No. 62103432)