文章编号：1672-8785(2023)06-0019-08

基于无监督学习的单样本红外图像生成方法

易　星[1,2,3]　潘　昊[1*]　赵怀慈[2,3]　杨　斌[1]

(1.沈阳化工大学信息工程学院，辽宁 沈阳　110142；
2.中国科学院光电信息技术处理重点实验室，辽宁 沈阳　110169；
3.中国科学院沈阳自动化研究所，辽宁 沈阳　110169)

摘　要：针对当前可见光-红外图像数据集匮乏导致的模型特征学习能力不够以及生成图像质量低下等问题，提出了单样本的无监督学习方法来训练红外图像生成模型。首先，在数据集难以获取、匮乏的情况下，仅采用一对可见光-红外图像作为模型训练的数据，降低了数据获取的难度，解决了数据匮乏的问题。其次，为了在训练模型时充分提取图像特征，改进了网络结构。实验数据表明，本文方法能够在单样本图像生成中取得较好的效果。在艾睿光电数据集中，本文方法的峰值信噪比(Peak Signal-to-Noise Ratio，PSNR)与结构相似性(Structural Similarity，SSIM)指标分别达到了 26.5588 dB 和 0.8846；在俄亥俄州立大学(Ohio State University，OSU)数据集上的 PSNR 和 SSIM 分别达到了 30.3528 dB 和 0.9182。与基于风格的生成对抗网络(Style-based Generative Adversarial Network，StyleGAN)方法相比，本文方法在艾睿光电数据集上的 PSNR 和 SSIM 指标分别提高了 16.07%和 23.78%；在 OSU 数据集上的 PSNR 和 SSIM 指标分别提高了 31.8%和 40.4%。结果表明，本文方法在当前图像质量评价指标方面有较为明显的提高，生成的红外图像纹理细节丰富且接近于真实红外图像。该研究对于今后的红外图像生成技术优化具有一定的参考意义。

关键词：无监督学习；红外图像生成；AdaIN 归一化模块；少样本数据

中图分类号：TP391　　**文献标志码：**A　　**DOI：**10.3969/j.issn.1672-8785.2023.06.004

Single-Sample Infrared Image Generation Method Based on Unsupervised Learning

YI Xing[1,2,3], PAN Hao[1*], ZHAO Huai-ci[2,3], YANG Bin[1]

(*1. School of Information Engineering, Shenyang University of Chemical Technology, Shenyang 110142, China;*
2. Key Laboratory of Opto-Electronic Information Technology Processing, Chinese Academy of Sciences, Shenyang 110169, China;
3. Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110169, China)

收稿日期：2022-12-07

基金项目：装备预研重点项目(41401040105)

作者简介：易星(1995-)，男，湖南攸县人，硕士研究生，主要从事深度学习与图像生成方面的研究。

*通讯作者：E-mail：panhao@syuct.edu.cn

Abstract: Aiming at the problems such as insufficient learning ability of model features and low quality of generated image caused by the current scarcity of visible-infrared image datasets, a single-sample unsupervised learning method to train infrared image generation model is proposed in this paper. First of all, when the dataset is difficult to obtain, only a pair of visible-infrared images are used as the data for model training, which reduces the difficulty of data acquisition and solves the problem of data scarcity. Secondly, in order to fully extract image features during the training of the model, the network structure is improved. Experimental data show that good results can be achieved in single-sample image generation by the proposed method. In the InfiRay OE dataset, peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) of the proposed method reach 26.5588 dB and 0.8846, respectively. PSNR and SSIM of the Ohio State University (OSU) dataset reach 30.3528 dB and 0.9182, respectively. Compared with the style-based generative adversarial network (StyleGAN) method, PSNR and SSIM of the proposed method in the InfiRay OE dataset are increased by 16.07% and 23.78%, respectively. PSNR and SSIM of OSU dataset are increased by 31.8% and 40.4%, respectively. The results show that the image quality evaluation index of the proposed method is improved significantly, and the texture details of the generated infrared image are rich and close to the real infrared image. The research has a certain reference significance for the optimization of infrared image generation technology in the future.

Key words: unsupervised learning; infrared image generation; adaptive instance normalization module; few sample data

0 引言

红外图像在军民两用领域应用广泛。尤其在军用领域，通过深度学习方法获取红外图像，可以更好地辨别目标，从而实现精准目标打击。目前，随着相关研究的快速发展，图像生成技术已经广泛应用于军事及电影场景制作领域。

少样本以及单样本图像生成是指基于已知种类的样本，通过模型训练得到给定种类图像的过程，而样本数据集通常只有一张或者几张图片。在图像生成领域，无监督学习指的是源域图像与目标域图像在模型训练时无需成对匹配及一一对应。

传统的少样本学习方法基于数据扩充技术，利用随机平移、旋转和翻转、添加高斯噪声等操作来实现类之间的转换。这种方法适用于数据量较小的数据集。当前，随着深度学习的快速发展，研究者们利用神经网络实现了少样本学习并将其广泛应用于图像生成领域。Clouâtre L 等人最早基于优化方法提出采用爬虫的小样本图像生成(Few-shot Image Generation using Reptile, FIGR)算法来研究少样本图像生成[1]。该算法实现了少量样本场景下的图像生成，但是生成的图像质量非常差，没有体现优化方法的优势。Antoniou A 等人在 2017 年提出了较为有名的数据增强生成对抗网络(Data Augmentation Generative Adversarial Network, DAGAN)算法[2]。该算法通过输入单张图像并对图像进行变换，得到属于同一类别的图像。通过这样的思想实现了少样本图像生成，但是用该方法生成的图像多样性不够，仍有很大的提升空间。Hong Y 等人在 2020 年提出了德尔塔生成对抗网络(Delta Generative Adversarial Network, DeltaGAN)算法[3]。他们通过建立随机向量与同类两张图片的差异，设计了生成子网络与重构子网络，提出了样本相关匹配损失，实现了少样本图像生成。用该算法生成的图像质量相较于 DAGAN 算法有了明显的提升，在复杂数据集上也能生成较好的样本图像。

2020 年，Hong Y 等人基于融合思想提出了基于匹配的生成对抗网络(Matching-based Generative Adversarial Network, Matching-GAN)算法[4]。他们将条件图像与随机向量输入到公共的匹配区域来学习较为合理的插值系数，同时将条件图像的深层特征图进行线性融

合来生成混合特征图像。但由于深层特征线性融合度不够，得到的图像效果不佳。针对MatchingGAN算法存在的问题，Hong Y等人同年又提出了融合/填充生成对抗网络(Fusing-and-Filling Generative Adversarial Network，F2GAN)方法[5]。该方法采用融合填充的思想获取到较好效果的图像，但在训练模型时效率低。Saito K等人2020年提出的少样本、无监督的图像到图像转换(Few-shot，UNsupervised Image-to-image Translation，FUNIT)算法[6]将基于融合的方法与基于转换的方法相结合，对中间表征进行解耦，然后针对每一个解耦的表征图再进行单独操作，从而实现图像生成。2021年，Gu Z等人又提出了局部融合生成对抗网络(Local-Fusion Generative Adversarial Network，LoFGAN)方法[7]。他们将整体数据图像随机分成基本图像与参考图像，并通过对两者的特征语义匹配以及对局部特征的临近替换，在细粒度水平上产生了更逼真的多样化图像。同时，他们还提出一种局部重建损失，为模型训练提供了较好的稳定性和生成质量。Li T等人2021年提出的记忆-概念-注意(Memory-Concept-Attention，MoCA)方法[8]通过对记忆原型进行聚类来生成图像。该方法提高了少样本图像生成的质量，且在模型鲁棒性方面也有不错的优势。

2022年，Yang M等人通过基于融合策略的方式提出了WaveGAN方法[9]。这是一种融合了频率感知的少样本图像生成模型，具体如下：通过特征编码将其分解为多个频带分量，然后利用高频分量与低频分量的特征信息处理能力的强弱，将丰富信息提供给高频分量进行融合处理，从而使图像生成的质量大幅提高。2022年，Li Z等人提出了FakeCLR算法[10]。他们利用对比学习的方法缓解数据增强的不稳定性问题。与其他对比学习的策略相比，该方法只对扰动的假样本应用对比学习，可生成较好的图像。

为了解决图像生成质量差以及图像数据集匮乏的问题，本文借鉴了StyleGAN生成器与辨别器思想[11]，提出了单样本图像生成的方法。与StyleGAN方法不同的是，我们不再需要大量的数据集，数据集源域与目标域样本只有一张图像。其次，通过对生成器以及辨别器网络深度进行加深操作来充分提取图像特征，使得模型具有更好的生成能力和辨别能力，进而生成质量较高的红外图像。

1 相关理论及方法

1.1 网络整体框架

本文采用的生成器网络[12]主要分为两部分。第一部分为映射网络(Mapping Network)，其作用主要是生成所需的风格参数。映射网络将随机采样的数据投影到特征空间，由8个相同的FC全连接层构成。通过对随机向量的特征解缠，输出不同的特征变量来控制不同的视觉特征(见图1)。第二部分采用了合成网络(Synthesis Network)。它主要将通过映射网络得到的不同视觉特征合成目标域中的红外图像。网络的整体流程如图1所示。将可见光图像变成特征向量，并将其输入到由映射网络以及添加随机噪声的合成网络组成的生成器中，生成假的红外图像。与此同时，将假的红外图像与真的红外图像一起输入到鉴别器中进行鉴别，再将得到的鉴别损失值反馈给生成器和鉴别器，以便更好地调整生成器网络的生成能力。

映射网络仅使用输入向量进行特征控制的能力有限，影响图像生成质量，且图像容易产生伪影。为了解决这些问题，我们在合成网络中引入了自适应实例归一化(Adaptive Instance Normalization，AdaIN)模块[13]。该模块是2019年Karras T等人[13]为了对源随机向量进行控制而提出的。在自然风格迁移的任务中，AdaIN模块应用在归一化层。相关研究表明，它能够解耦低水平的风格特征以及高水平的内容特征。通过归一化统计可以将风格与内容特征结合起来。因此，在生成器中引入AdaIN模块能够解决生成的假红外图像出现的伪影问

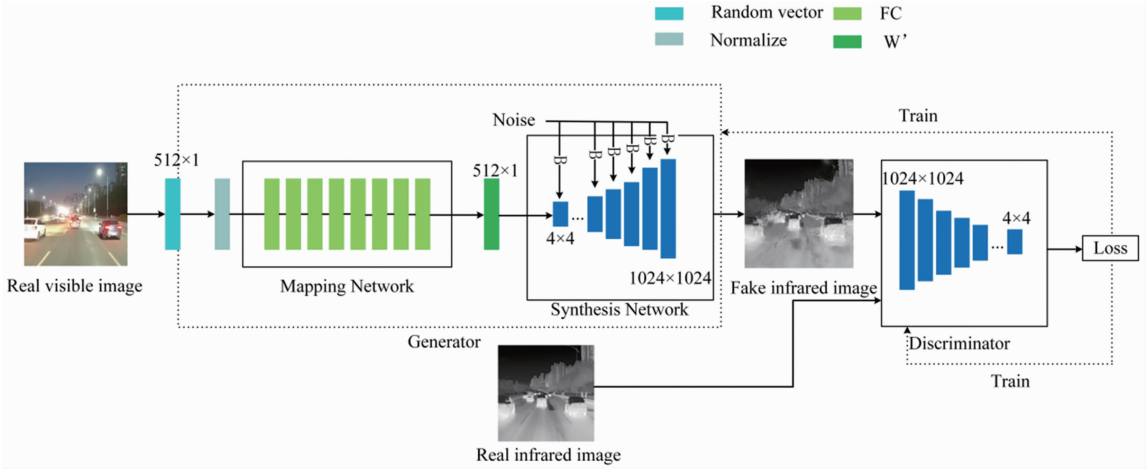

图 1 整体网络框架结构示意图

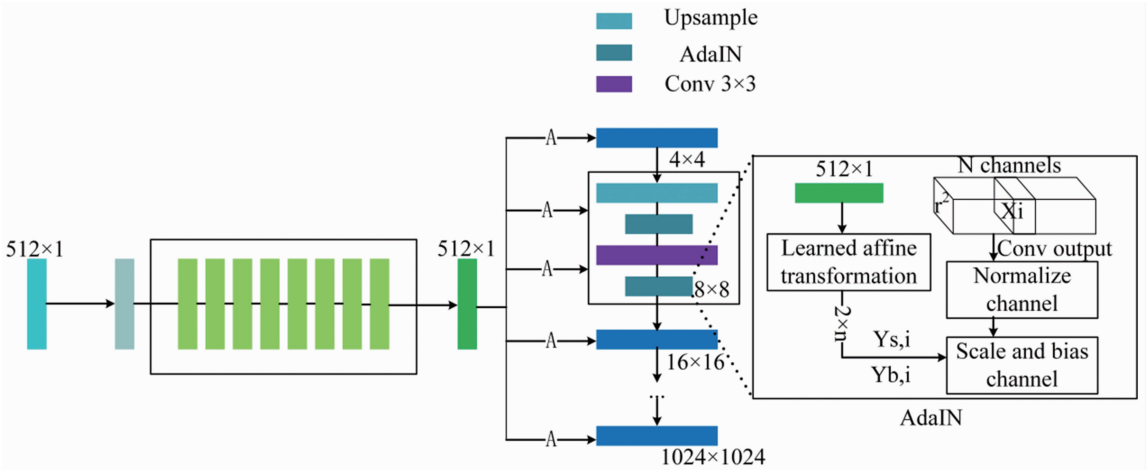

图 2 AdaIN 模块的结构示意图

题。它与同类的归一化层类似，主要对网络中间层的输出结果进行 scale 和 shift 操作。利用可学习的 scale 和 shift 参数实现特征图中不同位置的对齐，提高网络的学习效果，避免梯度消失。图 2 为 AdaIN 模块的结构示意图。

1.2 损失函数

在图像生成领域内，图像的输入域为 $x\in R^{H\times W\times C}$，输出图像域为 $y\in R^{H\times W\times 3}$，数据集为 $X=\{X\,|X\in x\}$，$Y=\{Y\,|Y\in y\}$。生成器主要分为编码(Encoder)和解码(Decoder)两个部分，因此输出图像的过程即为 $\hat{y}=G(z)=G_{dec}(G_enc(x))$。在生成器部分，需要保证生成的图像与目标域图像接近相似，因此采用对抗损失。该损失函数为

$$L(G,D,X,Y)=E_{y\sim Y}\log D(y)+E_{x\sim X}\log(1-D(G(x))) \quad (1)$$

在传统的无监督学习中，可能需要通过预先设计好的损失函数来衡量预测的表现情况。本文使用噪声对比估计来学习编码，将最大化形成互信息，通过关联信号信息与数据集中的样本图像来形成对比。噪声对比估计需要对查询信号与正样本信号进行关联后再将其与负样本形成对比。首先将查询样本与正样本及负样本分别映射 K 维向量 V，$V^+\in R^K$，$V^-\in R^{N\times K}$。其中，V^- 属于负样本。将样本归一化到单位球中，从而防止空间坍塌或者空间扩张。通过这样的方式形成一个 $N+1$ 分类问题。那么正样本被选中的概率为

$$\gamma(v,v^+,v^-)=-\log\left[\frac{\exp(v\times v^+/\tau)}{\exp(v\times\frac{v^+}{\tau})+\sum_{n=1}^{N}\exp(v\times\frac{v^-}{\tau})}\right] \quad (2)$$

该方法采用的网络既有图像层次，又有图像块层次，整个输入、输出图像应该有同样的结构，对应的图像块之间也应该有相应的多层次图像块的学习目标。在 Encoder G 的特征编码层中，不同的层、不同的空间位置代表不同的图像块。层数越深，图像块越大。假设选择感兴趣的共 L 层的特征图，将其通过 2 层 MLP 网络 H_1 产生的特征为

$$\{Z_1\}_L = \{H_1(G^1_{enc}(x))\}L \tag{3}$$

式中，G^1_{enc} 表示第 1 层输出特征；Z_1 表示第 1 层特征；L 表示层数($L\in\{1,2,3,\cdots,L\}$)；S 表示每一层的图像块数($S\in\{1,2,3,\cdots,S_1\}$)，S_1 表示第 L 层有 S_1 个空间位置；$Z_l\in R^{C_l}$ 表示第 l 层里第 S 个图像块对于特征向量的维度为 C_l。式(4)表示第 l 层所有的图像块 S 中除去 s 的特征：

$$Z_l^{S/s} \in R^{(S_L-1)\times C_1} \tag{4}$$

输出图像 $\hat{y}$ 则表示为

$$\{\hat{Z}_l\}L = \{H_l(G^1_{enc}(x))\}L \tag{5}$$

对输入输出对应位置的图像块进行匹配，并将同一张图像其他位置的图像块作为负样本，则损失记为 patch loss。该损失函数如下：

$$L_{PatchNCE}(G,H,X) = E_{x\sim X}\sum_{l=1}^{L}\sum_{s=1}^{S_L} l(\hat{Z}_l^S, Z_l^S, z_l^{S/s}) \tag{6}$$

因此，最终的目标函数添加一致性损失函数 $L_{PatchNCE}(G,H,X)$，使得 $E_{y\sim Y}(G(y)-y_l)$ 尽量最小，避免生成器生成的图像与源域图像相差太大。总损失函数为

$$\begin{aligned} L_{all} &= L_{PatchNCE}(G,H,X,Y) \\ &+\lambda_x L_{PatchNCE}(G,H,X) \\ &+\lambda_y L_{PatchNCE}(G,H,Y) \end{aligned} \tag{7}$$

式中，λ_x 与 λ_y 为权重系数，均设置为 1。

2 分析与讨论

2.1 数据集与实验过程

本文实验训练图像生成模型采用的平台配置如下：在 Ubuntu 系统中配置 Python3.7 的环境；采用 Pytorch 框架；显卡为 GeForce RTX 2080Ti；运行内存为 11G。测试后的结果和消融实验结果指标都是在 Windows 10 操作系统中运行 Matlab 2019a 后得出的。

在训练期间，所用数据集是 OSU 色热数据集以及艾睿光电数据库的红外图像与可见光图像序列对[15]。由于本文研究的是单样本图像生成，因此在数据集中分别采用三对 256×256 像素的可见光-红外图像作为训练图像，另外取三对数据集作为测试图像。本实验相关参数如下：训练 epoch 数设置为 16 个；前 8 个 epoch 的学习率设置为 0.0002，后 8 个 epoch 设置为线性梯度下降；每个 epoch 迭代 10000 次图像数据；batch_size 设置为 4。

2.2 实验结果

为了验证本文实验结果的有效性，从而正确评价使用该方法得到的红外图像生成质量，本文从客观上采用 PSNR 和 SSIM 两项指标进行图像评价。评价结果和可视化结果分别如表 1 和图 3 所示。与传统方法相比，本文提出的算法在艾睿光电数据集序列 1 上的 PSNR 与 SSIM[16-20]值分别提高了 16.07%和 23.78%，在艾睿光电数据集序列 2 上的 PSNR 与 SSIM 值分别提高了 11.5%和 12.8%，在 OSU 数据集上的 PSNR 与 SSIM[20]值分别提高了 31.8%和 40.4%。由表 1 可知，相较于 CycleGAN 方法与 Recycle-GAN 方法，本文方法的数据指标有很大提升。从图 3 中可以看出，采用本文方法生成的图像细节纹理接近于真实的红外图像，图像质量高；而采用 StyleGAN 方法(Original)生成的红外图像细节丢失严重，图像模糊不清。因此，本文提出的算法在单样本图像生成任务中取得了显著效果，客观上验证了其有效性。

2.3 消融实验

本文的消融实验主要采用艾睿光电数据集，通过对网络生成器与辨别器结构进行不同程度的加深来验证网络生成图像的稳定性是否有效。实验主要包含以下几个部分：(1)生成器深度设置为 4，辨别器深度设置为 4；(2)生成器深度设置为 8，辨别器深度设置为 8；(3)生成器深

图 3 单样本图像在不同数据集上生成的可视化结果

表 1 单样本图像生成在不同数据集中的评价指标数据

数据集	StyleGAN（Original）		CycleGAN		RecycleGAN[15]		本文方法	
	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
艾睿光电序列 1	21.1119	0.6601	14.87	0.54	16.84	0.56	24.5053	0.8171
艾睿光电序列 2	23.8190	0.7842	15.78	0.48	16.34	0.43	26.5588	0.8846
OSU 序列 3	23.0149	0.6540	13.68	0.25	17.03	0.68	30.3528	0.9182

度设置为 10，辨别器深度设置为 8；（4）生成器深度设置为 16，辨别器深度设置为 16；（5）生成器深度设置为 32，辨别器深度设置为 32；（6）生成器深度设置为 64，辨别器深度设置为 64。表 2 列出了单样本图像生成在不同深度参数下的消融实验评价指标数据。可以看出，在生成器网络深度与辨别器网络深度都等于 4 时，生成图像的 PSNR 与 SSIM 评价指标较低。随着生成器和辨别器网络深度的增加，生成图像的 PSNR 和 SSIM 值越来越高，说明生成图像的质量也越来越好。

将消融实验可视化训练结果（见图 4）与测试结果（见图 5）相结合并比较后可以明显看出，当生成器与辨别器深度值较小时，网络训练模型图像与测试模型图像生成的结果相差较大，主要表现如下：训练时生成的图像纹理细节丰富，图像质量好；测试时生成的图像纹理细节缺失严重，图像质量低下。以 $ngf=4$、$ndf=4$ 为例，训练图像的生成质量较好，但是模型在测试时生成的图像纹理细节混乱，质量低下。随着网络深度的加深，训练模型生成的图像与测试模型生成的图像越来越接近。以 $ndf=64$、$ngf=64$ 为例，训练阶段与测试阶段生成的图像差距越来越小，图像的 PSNR 值和 SSIM 值较高，纹理生成信息较为丰富，没有失真现象，模型的泛化能力强。相反，当

表 2 单样本图像生成在不同深度参数下的消融实验评价指标数据

深度参数	评价指标	
	SSIM	PSNR/dB
$ngf=4$，$ndf=4$	21.1119	0.6601
$ngf=8$，$ndf=8$	22.6182	0.7262
$ngf=10$，$ndf=8$	23.4527	0.7257
$ngf=16$，$ndf=16$	24.1758	0.7700
$ngf=32$，$ndf=32$	24.6880	0.8023
$ngf=64$，$ndf=64$	24.5053	0.8171

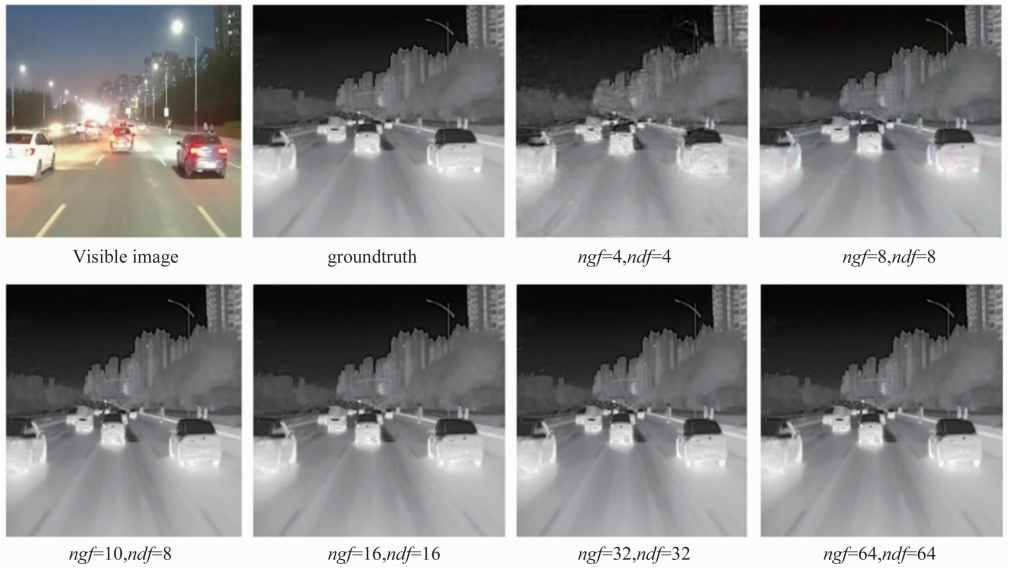

图 4 单样本图像在艾睿光电数据集上训练时生成的可视化结果

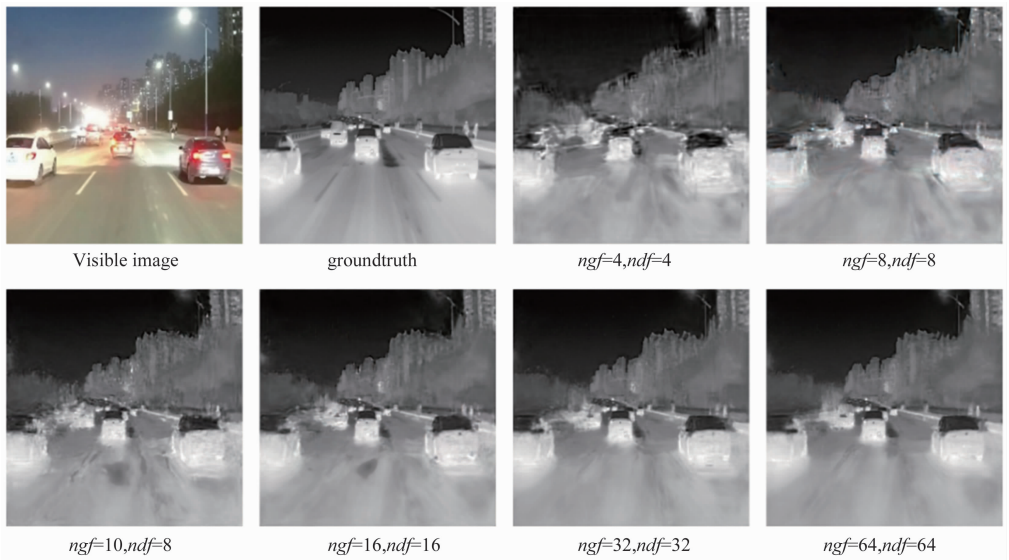

图 5 单样本图像在艾睿光电数据集上进行模型测试时生成的可视化结果

ndf 为 4、8、10，ngf 为 4、8 时，在生成器和辨别器深度不够的条件下所测试的图像模糊不清，失真严重。由此可见，本文增加生成器与辨别器深度后能够有效提高红外图像生成的质量。

3 结束语

针对当前红外图像数据匮乏、难以获取的

问题，本文提出了单样本图像的生成方法。此外，针对生成的红外图像质量低下的问题，增加了生成器网络以及辨别器网络的深度，对模型生成图像时有更好的特征提取能力，可生成高质量的红外图像。本文方法在当前图像质量评价指标中都有较为明显的提高，生成图像的纹理细节丰富，质量与真实红外图像非常接近。因此，本文提出的红外图像生成方法是有效的。但由于在训练期间网络深度增加，该模型所需的计算量也非常巨大。模型训练时间较长，所需的硬件设备资源较大。未来将聚焦于改善模型的结构，用更轻量化的模型代替生成器网络，在保证图像质量的前提下优化网络结构，从而生成更高质量的红外图像。

参考文献

[1] Clouâtre L, Demers M. Figr: Few-shot image generation with reptile [J]. *arXiv*: 1901.02199, 2019.

[2] Antoniou A, Storkey A, Edwards H. Data augmentation generative adversarial networks [J]. *arXiv*: 1711.04340, 2017.

[3] Hong Y, Niu L, Zhang J, et al. Deltagan: Towards diverse few-shot image generation with sample-specific delta[C]. Tel Aviv: European Conference on Computer Vision, 2022.

[4] Hong Y, Niu L, Zhang J, et al. Matchinggan: Matching-based few-shot image generation [C]. London: 2020 IEEE International Conference on Multimedia and Expo (ICME), 2020.

[5] Hong Y, Niu L, Zhang J, et al. F2gan: Fusing-and-filling gan for few-shot image generation [C]. Seattle: 28th ACM International Conference on Multimedia, 2020.

[6] Saito K, Saenko K, Liu M Y. Coco-funit: Few-shot unsupervised image translation with a content conditioned style encoder [C]. Glasgow: European Conference on Computer Vision, 2020.

[7] Gu Z, Li W, Huo J, et al. Lofgan: Fusing local representations for few-shot image generation [C]. Montreal: IEEE/CVF International Conference on Computer Vision, 2021.

[8] Li T, Li Z, Luo A, et al. Prototype memory and attention mechanisms for few shot image generation [C]. Vienna: International Conference on Learning Representations, 2021.

[9] Yang M, Wang Z, Chi Z, et al. WaveGAN: Frequency-Aware GAN for High-Fidelity Few-Shot Image Generation[C]. Tel Aviv: European Conference on Computer Vision, 2022.

[10] Li Z, Wang C, Zheng H, et al. FakeCLR: Exploring Contrastive Learning for Solving Latent Discontinuity in Data-Efficient GANs [C]. Tel Aviv: European Conference on Computer Vision, 2022.

[11] Huang J, Liao J, Kwong S. Unsupervised image-to-image translation via pre-trained stylegan2 network [J]. *IEEE Transactions on Multimedia*, 2021, **24**: 1435-1448.

[12] Park T, Efros A, Zhang R, et al. Contrastive learning for unpaired image-to-image translation [C]. Glasgow: European Conference on Computer Vision,2020.

[13] Karras T, Lainel S, Aila T. A style-based generator architecture for generative adversarial networks [C]. Long Beach: IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019.

[14] Li S, Han B, Yu Z, et al. I2v-gan: Unpaired infrared-to-visible video translation [C]. Chengdu: 29th ACM International Conference on Multimedia, 2021.

[15] Hore A, Ziou D. Image quality metrics: PSNR vs. SSIM [C]. Istanbul: 20th International Conference on Pattern Recognition, 2010.

[16] Winkler S, Mohandas P. The evolution of video quality measurement: From PSNR to hybrid metrics [J]. *IEEE Transactions on Broadcasting*, 2008, **54**(3): 660-668.

[17] Ssara U, Akter M, Uddin M S. Image quality assessment through FSIM, SSIM, MSE and PSNR — a comparative study [J]. *Journal of Computer and Communications*, 2019, **7**(3): 8-18.

[18] Setiadi D. PSNR vs SSIM: imperceptibility quality assessment for image steganography [J]. *Multimedia Tools and Applications*, 2021, **80**(6): 8423-8444.

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

[20] Qian X Y, Zhang M, Zhang F. Sparse GANs for thermal infrared image generation from optical image [J]. *IEEE Access*, 2020, **8**: 180124-180132.