

分类号_____

学校代码 10487

学号 M201872600

密级 公开

硕士学位论文

华中科技大学

基于生成对抗网络的 红外图像生成方法研究

学位申请人：化彦伶

学科专业：控制工程

指导教师：左峥嵘 副教授

答辩日期：2020.5.16

**Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering**

**Research on Infrared Image Generation Method
based on Generative Adversarial Network**

Candidate : Hua Yanling
Major : Control Engineering
Supervisor : Assoc. Prof. Zuo Zhengrong

Huazhong University of Science & Technology
Wuhan, 430074, P. R. China
May, 2020

独创性声明

本人声明所呈交的学位论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除文中已经标明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对本文的研究做出贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签名：仇嘉俊

日期：2020 年 5 月 17 日

学位论文版权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，即：学校有权保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权 XXX 大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

本论文属于 ☐ 保密，在_____年解密后适用本授权书。
☒ 不保密。

（请在以上方框内打“√”）

学位论文作者签名：仇嘉俊

日期：2020 年 5 月 17 日

指导教师签名：李均峰

日期：2020 年 5 月 17 日

摘要

红外成像具有抗干扰能力强、远距离成像、夜间成像等优点，在民用和军事领域都有重要应用。在红外相关设备的研制过程中，需要大量多种条件下的红外图像作为验证测试数据。而外场试验实拍红外图像需要耗费巨大的人力、物力，同时很难获得全时段红外图像。针对红外图像样本不足的问题，论文将生成对抗网络引入红外图像生成任务，研究了基于可见光图像的红外图像生成方法和红外图像时段拓展技术，主要工作如下：

（1）针对雾霾对可见光图像的降质作用导致的异源红外图像生成性能下降的问题，论文提出了基于图像雾化程度的去雾算法。考虑到雾图对应的透射率图反映了图像的雾化程度，算法搭建了与去雾模块并行的透射率图预测模块；设计阈值自适应学习模块计算分割阈值，生成不同等级雾化程度的掩码图，指导分离去雾卷积模块复原出无雾图像。实验表明本算法在公开数据集和自然图像上都实现了很好的去雾效果。

（2）为了根据易获得的可见光图像生成对应场景的红外图像，论文尝试将 Pix2pix 网络应用到配对的可见光红外图像数据集。针对 Pix2pix 网络生成的红外图像细节信息缺失的问题，论文提出了多感受野特征融合 Pix2pix 网络，构建了基于 Unet++ 结构的多感受野特征提取器；提出了逐级嵌套像素级的多感受野特征融合机制。实验表明多感受野特征融合 Pix2pix 网络实现了更精细的红外纹理生成。

（3）针对配对的多时段红外图像数据的缺失问题，论文首先提出了基于 StarGAN 网络的红外图像时段拓展算法。针对 StarGAN 网络对同一材质图像区域拓展结果不一致的问题，论文提出了语义约束 StarGAN 网络，在 StarGAN 网络基础上添加了语义分割分支，加入场景语义特征，设计了语义编码一致损失。针对 StarGAN 网络对部分材质图像区域拓展结果错误的问题，论文改进了语义约束 StarGAN 网络的时段编码方式，提出了基于语义分割图和红外辐射时变曲线的时段信息编码方式。实验表明语义约束 StarGAN 网络更适用于红外图像时段拓展任务。

关键词：图像去雾，生成对抗网络，异源红外图像生成，红外图像时段拓展

ABSTRACT

Infrared imaging is widely used in civil and military fields, which has strong anti-jamming ability and can obtain clear images from a distance and in night scenes. In the process of developing infrared-related equipment, a large number of infrared images under various conditions are needed as verification data. However, it takes a lot of manpower and material resources to acquire real-time infrared images in the field. At the same time, it is difficult to obtain infrared images at all times. In order to solve the problem of insufficient infrared image datasets, this paper introduces the generative adversarial networks into infrared image generation task, and studies infrared image generation method based on visible image and infrared image period extension technology. The main contributions of this thesis are as follows:

(1) Aiming at performance degradation of generating infrared images caused by haze on visible images, this paper proposes a dehazing algorithm based on image fogging degree. Considering that the transmission map reflects the fogging degree of hazy images, we build a transmission map prediction module parallel to the dehazing module. The threshold adaptive learning module is designed to calculate the segmentation threshold. According to the segmentation threshold, we can get mask maps of different fogging degree, which can guide the separation defogging convolution module to recover clear images. Experiments show that our algorithm achieves good defogging effect on both public datasets and natural images.

(2) In order to generate infrared images of corresponding scenes based on easily available visible images, this paper tries to apply the Pix2pix network to the paired visible-infrared image datasets. To solve the problem of missing detail information of infrared image generated by Pix2pix network, this paper proposes the multi-receptive field feature fusion Pix2pix network. A multi-receptive field features extractor based on the Unet ++ structure is constructed and a multi-receptive field feature fusion mechanism is proposed. Experiments show that the multi-receptive field feature fusion Pix2pix network achieves finer infrared texture generation.

(3) Aiming at the lack of paired multi-period infrared image datasets, this paper firstly

proposes an infrared image period extension algorithm based on StarGAN network. In order to solve the problem that StarGAN network gets different results in expanding the same material image region, this paper proposes the semantic constrained StarGAN network. On the basis of the StarGAN network, a semantic segmentation branch is added. We inject the scene semantic features into net and design the semantic consistent loss. Aiming at the problem that the result of the expansion of some material image regions by StarGAN network is wrong, this paper improves the period-coding method of the semantic constrained StarGAN network. We propose a period-coding method based on semantic segmentation maps and infrared radiation time-varying curves. Experiments show that the semantic constrained StarGAN network is more suitable for the task of infrared image period extension.

Keywords: Image Dehazing, Generative Adversarial Networks, Heterogeneous Infrared Image Generation, Infrared Image Period Expansion

目录

摘 要	I
ABSTRACT	II
1 绪论	1
1.1 研究背景及意义	1
1.2 红外图像生成研究现状	2
1.3 图像去雾算法研究现状	6
1.4 本文的主要研究内容	7
2 基于图像雾化程度的图像去雾算法	9
2.1 图像去雾相关知识	9
2.2 基于图像雾化程度预测的图像去雾模型	10
2.3 本章小结	21
3 基于生成对抗网络的异源红外图像生成算法	22
3.1 异源红外图像生成相关理论介绍	22
3.2 基于 Pix2pix 的异源红外图像生成方法	25
3.3 基于多感受野特征融合 Pix2pix 的异源红外图像生成方法	34
3.4 本章小结	46
4 基于生成对抗网络的红外图像时段拓展	47
4.1 基于 StarGAN 的红外图像时段拓展	47
4.2 基于语义约束 StarGAN 的红外图像时段拓展	63
4.3 有雾图像多时段红外图像生成	79
4.4 本章小结	80
5 全文总结与展望	81
5.1 论文主要工作	81

华中科技大学硕士学位论文

5.2 未来工作展望	82
致谢	83
参考文献	84
附录 1 攻读硕士学位期间的科研成果	89

1 绪论

1.1 研究背景及意义

红外成像是将红外辐射信号转换为图像信号的信号转换过程。在红外成像过程中，红外探测设备获取到的红外辐射不仅包括太阳辐射，还包括自身热辐射，故可实现夜间成像，同时红外辐射对雾、霾有一定的抗干扰能力，在远距离条件和恶劣天气环境下也能得到清晰的成像结果。随着对红外辐射和红外光学的深入研究，红外成像技术快速发展，在民用和军事领域都有广泛应用。在民用领域，资源探测、红外遥感、红外救援等应用以及各种红外安防设备的出现，有力地保障了社会安全；在军事领域，红外目标识别、红外跟踪、红外成像制导都离不开红外成像技术的发展^[1]。

在红外相关设备的研制和红外成像技术的研究过程中，需要大量多种条件下的红外图像数据作为验证测试的素材。同时基于深度网络的算法也需要大量的红外图像用于算法的训练、测试与验证。然而，在大多数场景中，采用实拍红外图像的获取方式是十分困难的，例如卫星红外图像或者是军事目标红外图像的获取需要耗费大量的人力、物力^[2]，同时采集各种材质在所有背景、所有时段的红外图像是不现实的。因此利用红外仿真技术生成红外图像有重要研究意义。

目前主流的红外图像仿真方式是采用红外仿真软件平台进行场景红外仿真，通过对目标场景进行分析建模得到场景三维模型，然后根据红外辐射理论计算场景不同材质的红外辐射分布，之后利用大气传输模型计算红外辐射到探测器的辐射衰减，最后模拟成像器成像特性，为红外辐射添加传感器成像效应后进行灰度化得到红外仿真图像。相比采用外场实拍红外图像的方式，红外仿真软件大大减少了实测成本，同时通过调整红外辐射分布模型参数、传感器参数能够实现不同时段和不同波段的红外图像仿真。但是该方法存在着诸如目标温度模型仿真程度低、中间参数庞大、各系统耦合度高、处理过程繁琐等问题^[3]，不适用于快速生成大量红外图像。

为了实现快速、高效的大量红外图像数据的生成，本文针对红外仿真中的异源红外图像生成和红外时段拓展展开研究。可见光成像器技术发展成熟，并且数据获取方便、成本低，如果能够利用易获得的可见光数据去生成红外图像，便能很好的缓解红外图像获取困难的问题。同时，值得关注的是，由于可见光成像容易受到雾、霾天气的影响，会影响可见光图像到红外图像的转换性能，所以如何复原被雾霾污染的可见光图像也是本文要解决的一个问题。红外图像的时段拓展旨在根据某一时段的红外图像，通过时段拓展技术获得其他任意时段红外图像，进一步在时间维度上对红外图像样本空间进行扩充，从而为高级视觉任务提供更可靠有力的数据支撑。

人工智能、大数据挖掘、深度学习是近几年的研究热点，并且已经获得大量创新性成果，其中深度学习技术已经广泛应用于目标检测、目标识别、图像分割等领域。生成对抗网络是深度学习的一个新兴分支，不但满足了图像生成领域的研究和应用需求，同时也为无监督学习模型提供了强有力的算法框架^[4]。目前为止，国内外将深度学习应用到图像生成中的方法很少，在红外图像的时段拓展方法上几乎为零。如果能够将深度学习这一领域的技术应用到红外图像生成的工作中，不仅能够实现快速、高效的大量数据的生成，而且能够进一步推动红外图像生成技术的发展，为异源红外图像生成和红外图像的时段拓展提供新思路。

1.2 红外图像生成研究现状

从上世纪初叶开始，国内外陆续有很多研究机构开始了对红外图像仿真的研究，包括目标及背景的红外辐射分布计算、建模，红外图像仿真全流程研究，多时段红外图像生成，红外波段拓展等，目前基于红外仿真模型的仿真方法已经很成熟了。相比之下，深度学习发展历程相对短暂，而深度网络在红外图像生成任务上的应用更是寥寥无几。本文将针对以上两个方向对红外图像生成的国内外研究现状进行介绍。

1.2.1 基于仿真模型的红外图像生成方法

基于仿真模型的红外图像生成方法是被广泛采用的红外图像仿真方法。国外红

外仿真技术的研究始于二十世纪三四十年代,研究初期,主要是拟合出地物的表面温度分布,从而对地物建立热模型。1980年,Jeaoobs实现了基于简单一维热模型的地物表面温度建模^[5]。1985年,Ben-Yosef等人对8-12 μm 波段的红外图像进行辐射统计来建立经验模型^[6]。1987年,Biesel等人通过模拟技术实现了一种用于实时生成高准确度的红外图像的仿真方法^[7],创建了红外场景数据库。1994年,Gambotto等人提出了一种基于分割算法和区域分析相结合的红外仿真算法^[8],用于模拟地面表面温度并计算对应的红外热图像,同时实现了全天时红外温度模型计算。二十世纪九十年代,逐渐展开了红外图像仿真方面的研究。1992年,Cathcart J M根据场景几何模型、红外辐射模型和计算机图形学技术的结合对图像进行渲染^[9],生成了高分辨率合成图像。1996年,Hyun-Ki等人提出了一种根据结合目标内部热模型和背景环境温度分布的红外辐射模型计算方式^[10]。20世纪末期,Poglio T等人根据目标3D模型,搭建了一套能够实现高空间分辨率的3-14 μm 的红外图像仿真流程^[11]。

随着计算机计算能力的逐步强大,到了21世纪,出现了许多成熟的红外仿真平台,极大的简化了红外图像仿真的过程,缩短了仿真周期。2003年,英国的Lockheed Martin公司开发了CAMEO-SIM仿真软件^[12],能够根据场景几何模型、红外辐射模型、大气模型合成多种环境下的红外图像。目前受众最广泛的红外仿真软件是MuitiGen Paradigm公司开发的Vega系列软件包,该软件集成了大量精确的材料数据样本、辐射模型和大气模型,能够实现多环境、多时段的红外图像仿真。2007年,法国OKTAL-SE公司开发了SE-WORKBENCH^{[13][14]},该软件可以用于光电、激光雷达等多种传感器设备的环境建模仿真,SE-WORKBENCH中的专用红外模块可完成高可信度的红外图像仿真。往后几年,国外研究人员开始将具有强大并行能力的GPU引入红外仿真系统中,Mielikainen等人在2011年开发了一套基于GPU的高速、高性能的红外大气探测干涉仪IASI^[15],能够在安装多台GPU的计算机上运行。

相比国外红外图像仿真技术,我国红外仿真相关研究起步较晚,整体水平与国外相差较大,随着国家在红外仿真领域的加大投资,目前我国也取得了较大的进展。最初,我国的研究人员主要研究方向为模拟自然场景红外辐射分布。1997年,张建奇等人根据地表的辐射特性,结合热平衡方程,建立了地表红外辐射特性相关模型^[16]。

2000 年,王章野等人根据气象学和传热学理论,计算了建筑物屋顶、墙体、玻璃、沥青路等地物不同时段红外辐射强度^[17]。2002 年,宣益民等人实现了一种基于可见光图像的红外热像仿真方法,建立基于马尔可夫随机场理论的分割模型,模拟计算不同材质热辐射特性,将不同材质辐射度灰度化得到仿真红外热像^[18]。2010 年,陈珊等人提出了一种背景场景红外纹理仿真方法,利用可见光图像,根据场景温度模型、红外辐射模型、灰度映射模型生成不同时段红外图像^[19]。2015 年,周强等人根据可见光图像和近红外图像的辐射定标结果和场景发射率建立了可见光图像到近红外图像的灰度映射关系^[20],该方法可以实现较快的仿真速度,但是存在红外纹理真实感低的问题。2017 年,杨壹斌等人提出了一种基于红外灰度插值调制的红外图像仿真方法,利用传热学理论,根据灰度差值分布计算其他时段红外图像灰度分布^[21],适用于简单场景的红外图像仿真。从二十一世纪,我国逐渐开始了对完整红外仿真系统的研究。2000 年,吴亚平和张天序等人分析了在红外图像仿真过程中场景三维模型、大气传输模型和材质物理热性间的相互作用,并利用 Vega 软件完成了多波段红外图像仿真^[22]。2007 年,笪邦友和桑农等人采用 Vega 软件计算出不同材质在多种时段、天气、季节条件下的仿真图像,通过求解仿真结果与红外辐射量之间的关系,得到不同材质的红外辐射量数据作为对应材质的辐射量均值^[23]。2016 年,钟鸣在 SE-Workbench 软件的基础上进行二次开发,自动进行材质映射过程,减少了人工的过度干预^[24]。

以上基于仿真模型的红外图像生成方法都是基于红外成像的物理模型的,因此可解释性强,有一定的物理可信度,但是存在中间物理参数庞大、需要大量人工干预、目标背景间红外辐射计算耦合较少的问题,并不能实现快速大量红外图像的生成。

1.2.2 基于深度学习的红外图像生成方法

异源红外图像生成和红外图像时段拓展实质上是不同图像域间的相互映射过程,本文将图像生成问题建模为图像翻译问题。深度学习是近几年的研究热点,国外研究人员使用深度网络在可见光图像域上进行了一些图像翻译工作。2016 年,Gatys 等

人采用卷积神经网络实现了可见光图像的艺术风格转换^[25], 但是该算法存在运行效率底下、转换失真等问题。同年, Li Feifei 等人提出了基于感知损失函数的实时图像风格转换模型, 通过采用 Vgg-16 网络的中间特征层提取图像特征计算感知损失^[26]。

生成对抗网络 (GAN) 是深度学习中的热门话题, 最近 GAN 在图像到图像的翻译中取得了令人印象深刻的结果。到目前为止, 生成图像保持着低质量和低分辨率, 这种情况随着 Pix2pix 的出现改变, 是图像翻译领域一次彻底的革命。2016 年, Isola 等人提出了一种新的图像生成网络框架 Pix2pix, 通过采用生成器和判别器的对抗损失和 L1 损失实现了成对可见光图像间的图像转换问题^[27]。之后针对可见光图像生成对应的语义分割图问题, 为了进一步提高图像分辨率, Wang 等人提出了 Pix2pixHD 网络, 通过设计多级生成器和多尺度判别器, 在高分辨率图像翻译任务上取得了较好的结果^[28]。针对非配对图像域间的图像翻译问题, 2017 年, Zhu 等人提出了 CycleGAN 网络^[29], 通过采用对偶式的生成器网络和循环一致损失, 实现了非配对图像域间的相互转换。同年, Kim 等人提出的 DiscoGAN^[30]和 Yi 等人提出的 DualGAN^[31]采用了与 CycleGAN 相似的网络结构, 在图片上色和人脸内容编辑等方面得到了较好的结果。2018 年, Liu 等人提出了 UNIT 网络^[32], 将变分自动编码器与生成器结合, 每个图像域都有各自的编码网络和生成器网络, 并假定两个图像域编码网络的输出服从相同分布, 以学习跨域图像的联合分布。2019 年, Huang 等人基于特征解耦的思想提出 MUNIT 网络^[33], 将图像编码为与图像域无关的内容编码和图像域特征编码, 通过将图像内容编码与不同的域特征编码组合生成不同域图像。上述框架一次只能学习两个不同图像域间的映射关系, 在处理多个图像域时具有有限的可扩展性。2018 年, Choi 等人提出了 StarGAN 网络^[34], 将图像域信息进行编码与图像结合作为生成器输入, 采用一个生成器网络一次可进行多个图像域间的相互转换, 算法在人脸属性编辑任务上实现了较好的结果。

以上深度网络结构均是在可见光图像上展开的研究, 同时采用的数据集大多是经过对齐处理的人脸数据集, 数据集整体分布平缓。而本文红外数据集场景内容变化较大, 与可见光图像数据集数据分布差距较大, 所以实际应用仍有较大的性能差异。

近几年逐渐出现了一些基于生成对抗网络的红外图像生成算法。2019 年, 谢江荣等人提出了采用 DCGAN 网络实现一些简单红外目标图像的随机生成^[35]。同年, 冯云飞对红外波段拓展任务进行研究, 采用 CycleGAN 网络实现了中波红外图像到长波红外图像的转换^[3]。

1.3 图像去雾算法研究现状

雾霾导致的图像退化问题给人们的日常生活和科学研究造成了诸多不利影响, 因此研究图像去雾算法具有重要的应用价值和现实意义。图像去雾算法根据原理不同可以分为基于先验知识的算法和基于深度学习的算法。

基于先验知识的图像去雾算法。基于先验的去雾算法根据大气散射模型进行复原, 主要任务是估计透射率图。何开明等提出了暗通道先验 (DCP) 算法, 作者发现清晰图局部区域的暗通道值几乎接近零^[36]。根据该先验, 通过计算雾图的暗通道图获得透射率图。Zhu. 等人提出了一种颜色衰减先验 (CAP) 算法来恢复深度信息^[37]。在这种方法中, 他们通过观察深度信息与图像饱和度和亮度之差成正比来计算透射率图。Berman 等人提出了非局部图像去雾, 无雾图像的颜色可以近似为 RGB 空间中的数百个不同的颜色簇^[38]。作者假设由于雾的存在, 颜色簇呈线条形式, 基于此先验恢复了深度信息来计算透射率图。

基于深度学习的去雾算法。基于深度学习的去雾算法主要使用卷积神经网络来恢复清晰图像。Ren 等人提出了多尺度去雾网络 (MSCNN), MSCNN 包含粗规模和细规模的网络以获得透射率图^[39]。细尺度网络完善了从粗尺度网络获得的透射率图。Li 等人提出了一种端到端的去雾网络 (AOD-net), 通过变换大气散射模型计算方式, 在一个网络中恢复清晰图^[40]。Kangfu Mei 等人提出了一种基于编码器和解码器网络的去雾网络, 即渐进特征融合网络 (PFF-net), 直接学习从有雾图到无雾图的映射关系^[41]。Bharath Raj N. 等人提出了一种基于条件生成的网络去雾算法 (Dehaze-GAN)。该算法使用编码器-解码器结构作为生成器, 学习从有雾图到无雾图的映射关系, 并使用 PatchGAN 作为鉴别器^[42]。Malav 等人提出了一种端到端的去烟、去雾网络

(DHSGAN), 该模型是在生成的对抗网络框架下训练的, 将透射率图与雾图在通道上进行合并作为生成器输入复原无雾图像^[43]。

从上述方法中我们可以看到, 图像去雾算法在传统图像算法和深度学习领域都有较多的成果。但是已有算法在处理雾气分布不均的雾图、不同场景真实图像时, 泛化能力较弱。所以, 针对图像中雾化程度不一致这一现象, 设计更有针对性的网络结构是去雾算法性能提高的关键。

1.4 本文的主要研究内容

本文分为五章, 其中第一章为绪论, 第二章到第四章主要研究基于图像雾化程度的可见光图像去雾, 基于生成对抗网络的异源红外图像生成以及基于生成对抗网络的红外图像时段拓展, 最后一章为总结和展望。每章具体研究工作如下:

第一章为绪论, 主要介绍论文的研究目的和意义、红外图像生成的研究现状、图像去雾的研究现状以及本文的结构安排。

第二章研究可见光图像的去雾算法, 首先介绍了图像去雾的相关知识, 进一步地为了提高算法去雾性能, 提出了基于图像雾化程度的图像去雾算法, 分析了网络的构建方式, 最后通过实验证明去雾算法性能的优越性。

第三章研究基于可见光图像的红外图像生成, 首先利用 Pix2pix 网络完成异源红外图像任务。进一步地, 针对生成的红外图像细节信息缺失的问题, 提出了基于多感受野特征融合 Pix2pix 的异源红外图像生成算法, 介绍了基于 Unet++ 结构的多感受野特征提取器的构建方式, 分析了逐级嵌套像素级的多感受野特征融合机制, 详细介绍了网络的实现细节。最后通过实验证明多感受野特征融合 Pix2pix 在红外图像细节恢复上的仿真能力。

第四章研究基于生成对抗网络的红外图像时段拓展, 首先针对配对的红外图像时段数据缺失的问题, 提出了基于 StarGAN 网络的红外图像时段拓展算法。针对 StarGAN 网络拓展结果中出现的图像局部区域时段拓展失效的问题, 提出了语义约束 StarGAN 网络, 介绍了语义约束 StarGAN 网络基本思路和实现细节, 分析了基于

语义分割图和红外辐射时变曲线的时段信息编码方式，详细介绍了网络的实现细节。最后通过实验证明语义约束 StarGAN 网络在红外图像时段拓展任务上的算法性能，同时展示了从有雾可见光图像到多时段红外图像的仿真结果。

第五章对全文的工作进行总结，并对下一步研究工作给出建议。

2 基于图像雾化程度的图像去雾算法

实际成像环境中的雾霾将严重降低可见光图像的成像质量，导致地物信息获取能力下降，易劣化可见光图像到红外波段图像的转换性能，需要研究可见光图像的预处理方法，提升可见光图像质量，为后续异源红外图像生成提供高质量的输出数据，本章将进行可见光图像的去雾算法的研究，以恢复清晰图像。

2.1 图像去雾相关知识

为了有效地处理雾天图像，MCCARTNRY 根据雾天图像的成像机制，提出了大气散射模型^[44]，如下式所示：

$$I(x) = J(x) + A(x)(1 - t(x)) \quad (2-1)$$

$$t(x) = e^{-\beta d(x)} \quad (2-2)$$

其中 $I(x)$ 表示有雾图像； $J(x)$ 是待恢复的原始未退化的图像； $t(x)$ 为雾天图像对应的透射率图； $A(x)$ 为大气光值； $d(x)$ 为场景深度； β 为雾浓度参数。

目前针对单图像去雾的算法分为：基于先验知识的去雾算法、基于深度学习的去雾算法。基于先验的去雾算法需手工设计图像特征，算法的性能取决于先验知识的准确性，因此泛化能力较差。基于学习的去雾算法根据是否采用大气散射模型设计网络结构分为两类，依赖大气散射模型的网络分别估计透射率图和大气光值，然后根据式 2-1 恢复清晰图。该方法存在两个局限性，一是透射率图及大气光值估计不准确会影响雾图的去雾程度，二是线性模型不足以模拟复杂的成雾机制。另一类基于学习的去雾算法直接从雾图恢复清晰图，通过网络学习从雾图到清晰图的复杂变换函数。但是由于缺乏物理模型的限制，再加上卷积神经网络参数共享的性质，使得雾图中不同雾浓度的区域采用相同的参数，导致去雾后图像局部偏暗或亮度过高，去雾不彻底。

雾图的透射率图反映了场景反射光经过粒子衰减能够达到探测系统的比例，与图像的雾化程度成负相关，所以本文以透射率图为依据，设计了一种能够并行处理雾图中不同雾化程度区域的去雾卷积网络。

2.2 基于图像雾化程度预测的图像去雾模型

2.2.1 网络结构介绍

本文搭建的基于图像雾化程度的去雾模型如图 2-1 所示，包含四个模块：特征编码模块、透射率图估计模块、去雾模块和分离去雾模块。特征编码模块提取不同尺度的图像特征，分别传递给透射率图估计模块和去雾模块对应尺度的卷积块。透射率图估计模块得到的透射率图和去雾模块提取的图像特征输入分离去雾模块，由分离去雾模块恢复清晰图。

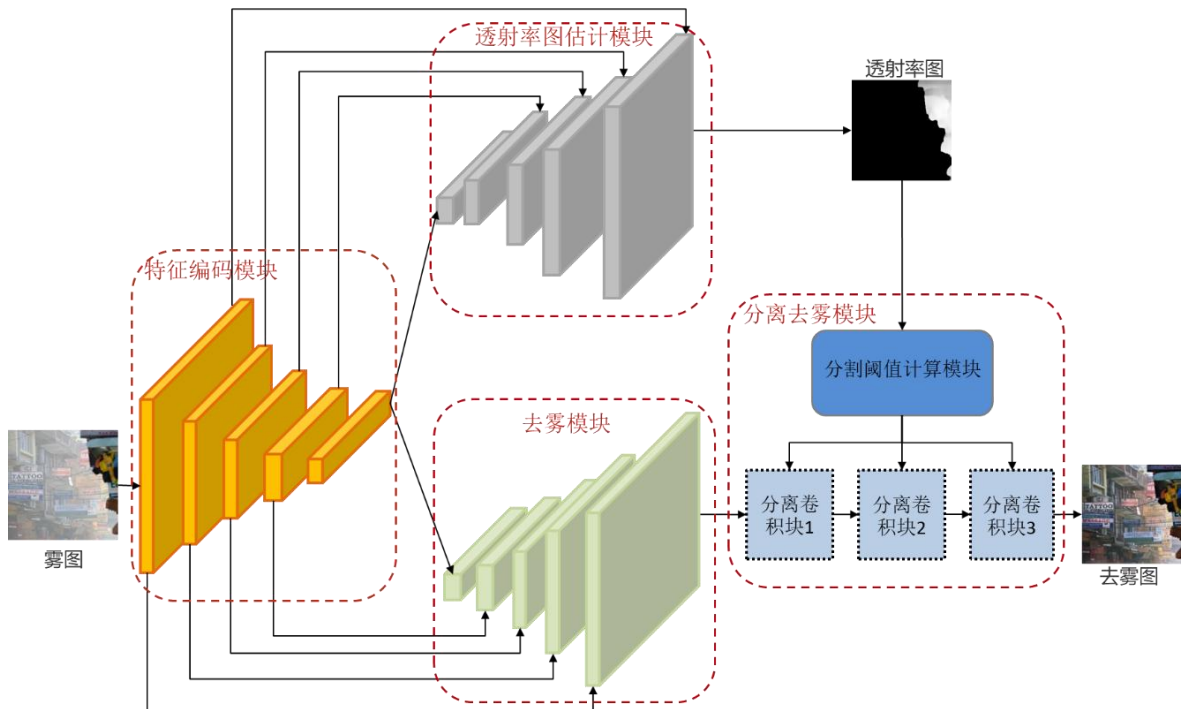


图 2-1 基于图像雾化程度的去雾模型

雾图中不同区域的雾化程度有所差异，一般图像近景区域的雾化程度低于远景区域。对图像整体去雾，可能会导致近景处过度增强，远景处去雾不彻底。基于此，本文构建了分离去雾模块。分离卷积模块的基本思路是：根据透射率图估计模块得

到透射率图，可以得到对应雾图的雾化程度；通过设定一个合适的阈值，可以计算出弱雾掩码和强雾掩码，采用弱雾掩码和强雾掩码来指导分离卷积块的卷积过程。以上分析均是基于全卷积神经网络的空间不变性。对于阈值的设定，本文设计了专门的阈值估计组件，采用卷积块来自动学习针对每张雾图的合适的阈值。设定的阈值用来界定弱雾区域和强雾区域，应与透射率图全局信息相关，所以预测阈值卷积层的卷积核大小应与透射率图大小相同。为了减少参数量，首先将透射率图进行 32 倍均值下采样，

$$trans = Avgpool(trans, 32) \quad (2-3)$$

之后采用卷积核大小为原图像大小 1/32 倍的卷积层计算阈值，最后将其映射到 (0,1) 范围内，计算方式如下：

$$Threh = \frac{1}{1 + e^{-f(trans)}} \quad (2-4)$$

其中 $trans$ 表示透射率图， f 表示针对透射率图设计的卷积模块。

通过比较透射率图中的像素值与估计的阈值的大小，来确定弱雾区域掩码图 ML 和强雾区域掩码图 MD ，计算方式分别如式 2-5、2-6 所示：

$$ML(x, y) = \begin{cases} 1, trans(x, y) \geq Threh \\ 0, trans(x, y) < Threh \end{cases} \quad (2-5)$$

$$MD(x, y) = \begin{cases} 1, trans(x, y) < Threh \\ 0, trans(x, y) \geq Threh \end{cases} \quad (2-6)$$

分离卷积块基于全卷积神经网络的空间不变性，根据弱雾区域掩码图 ML 和强雾区域掩码图 MD 将特征 f 进行分离，得到弱雾区域特征 f_{ML} 和强雾区域特征 f_{MD} ，计算方式分别如式 2-7、2-8 所示：

$$f_{ML}(x, y) = f(x, y) \times ML(x, y) \quad (2-7)$$

$$f_{MD}(x, y) = f(x, y) \times MD(x, y) \quad (2-8)$$

采用 2 个并行的卷积块分别处理弱雾区域特征 f_{ML} 和强雾区域特征 f_{MD} ，将两个

卷积块 F_1 、 F_2 的输出进行像素相加得到完整特征，计算方式如下式所示：

$$f_r(x, y) = F_1(f_{ML}(x, y)) + F_2(f_{MD}(x, y)) \quad (2-9)$$

最后采用卷积块对并行分离去雾得到的特征进一步优化，综上所述，分离卷积块的结构如图 2-2 所示：

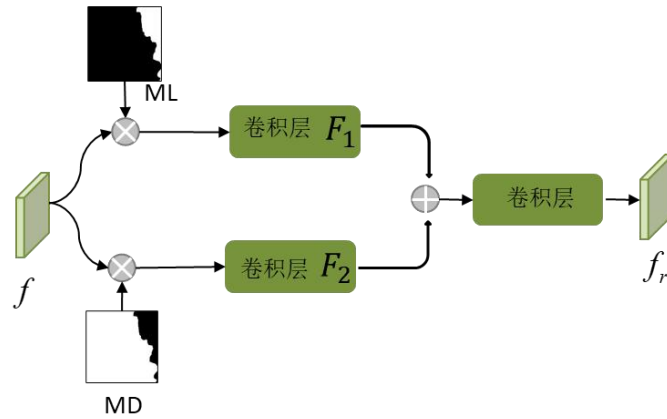


图 2-2 分离卷积块示意图

2.2.2 模型构建分析

基于图像雾化程度的图像去雾算法模型如图 2-3 所示，其中编码模块由 5 个下采样卷积块组成，透射率图估计模块和去雾模块由传输卷积层和 5 个上采样卷积块组成，分离卷积模块由 2 个阈值预测卷积块和 3 个分离卷积块组成。

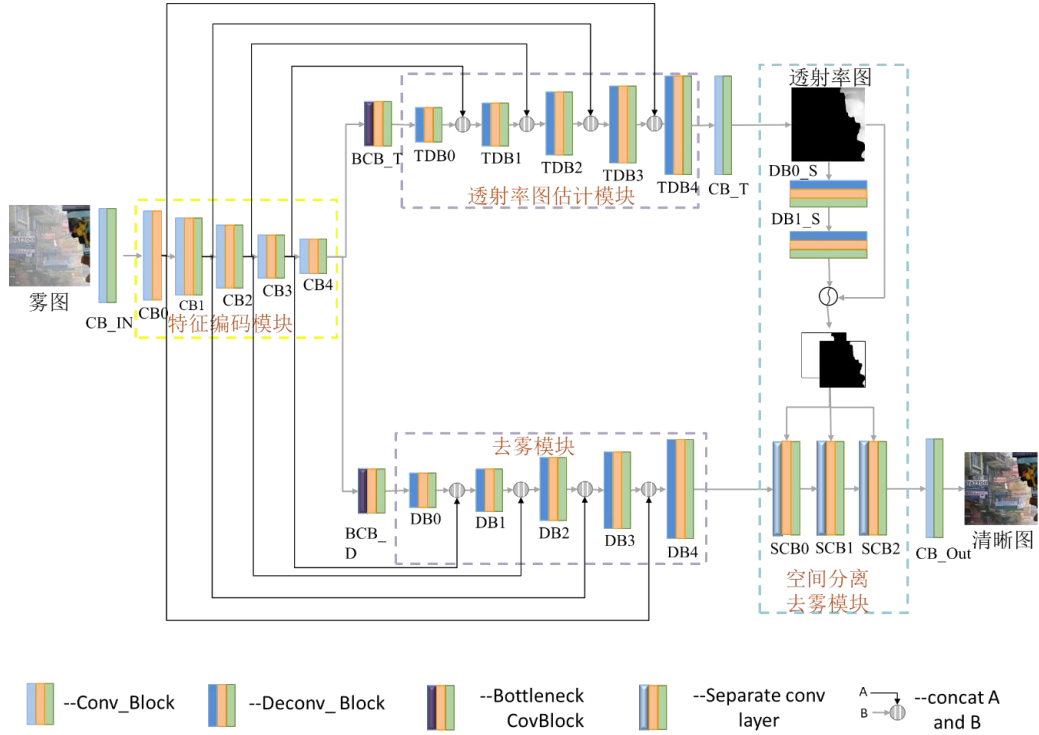


图 2-3 去雾模型网络结构

Densenet^[45]提出对卷积层的输入和输出添加短连接,可以有效地减少因为网络深度加深导致的梯度弥散,更易于网络训练。借鉴 Densenet 的思想,对于卷积块 Conv_Block、传输层 Bottleneck_Conv_Block 和上采样块 Decov_Block,我们将密集连接层 Dense_Layer 作为网络基础卷积块。Dense_Layer 包括 3 个子卷积块,每个子卷积块的输入输出进行通道上的合并,Conv_Block、Bottleneck_Conv_Block 和 Decov_Block 模块结构如图 2-4 所示,其中的基本模块如图 2-5 所示:

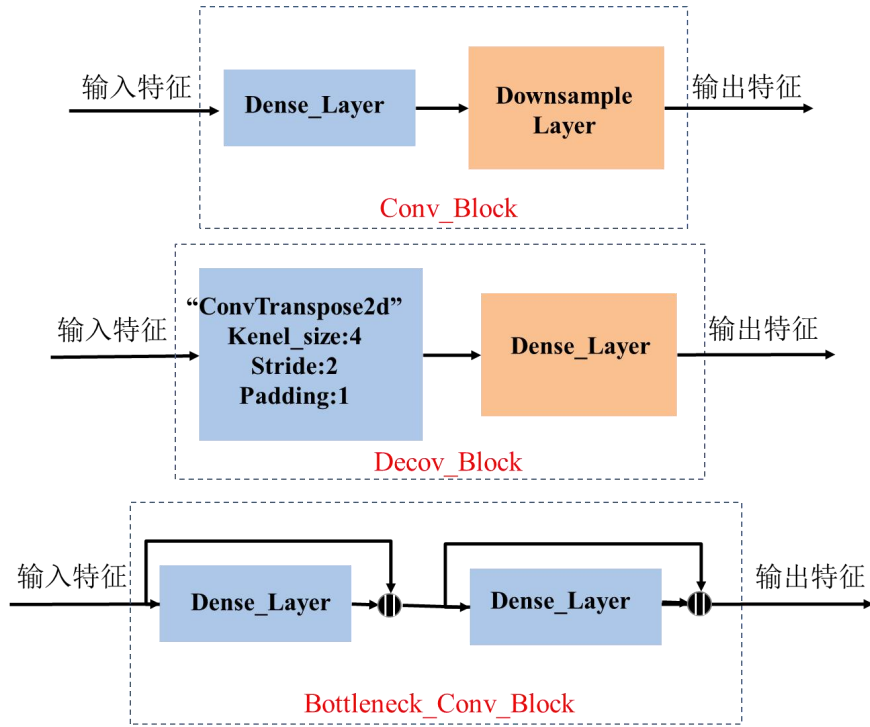


图 2-4 去雾网络子模块网络结构

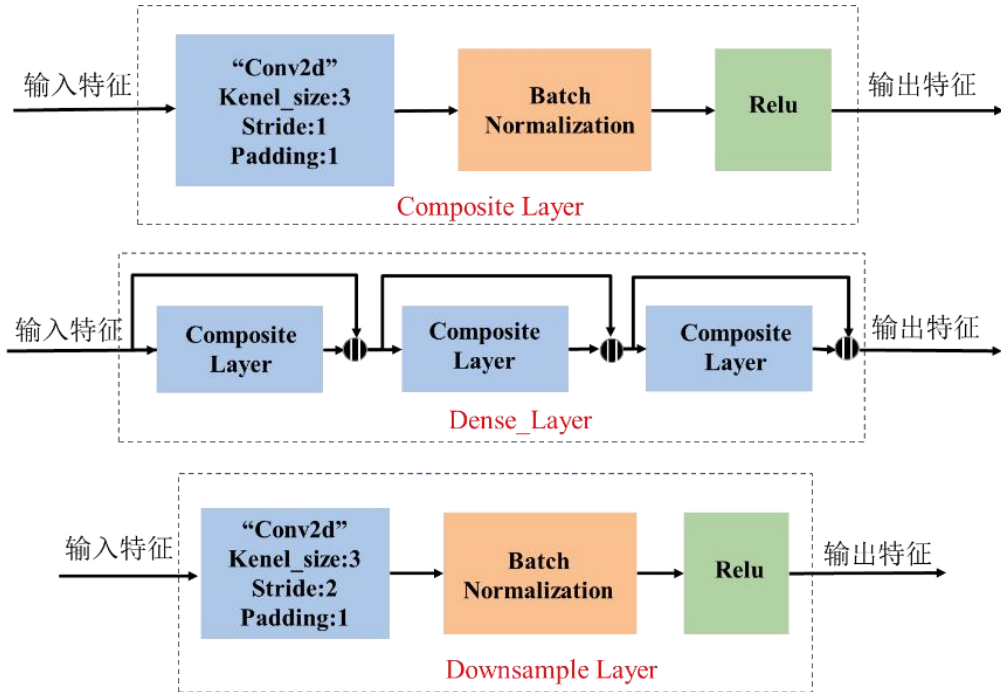


图 2-5 基本模块网络结构

阈值自适应计算模块如图 2-6 所示, 首先将透射率图进行 32 倍下采样, 之后采

用卷积核为 $256/32=8$ 大小的卷积层学习全局特征，最后采用 Sigmoid 映射函数将阈值映射到 $(0, 1)$ 。

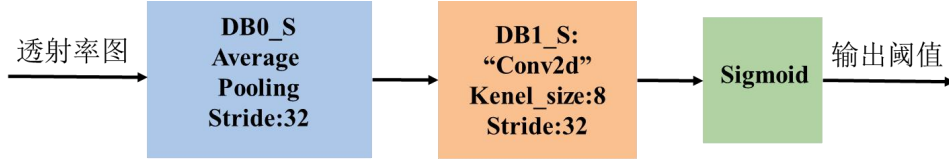


图 2-6 阈值自适应计算模块网络结构

对于分离卷积模块，根据得到的掩码图将输入特征进行分离，之后采用两个卷积核分别学习两部分特征，同样的我们采用 Densenet 的思想，将每个子模块的输入输出进行通道上的合并，具体模块结构如图 2-7 所示：

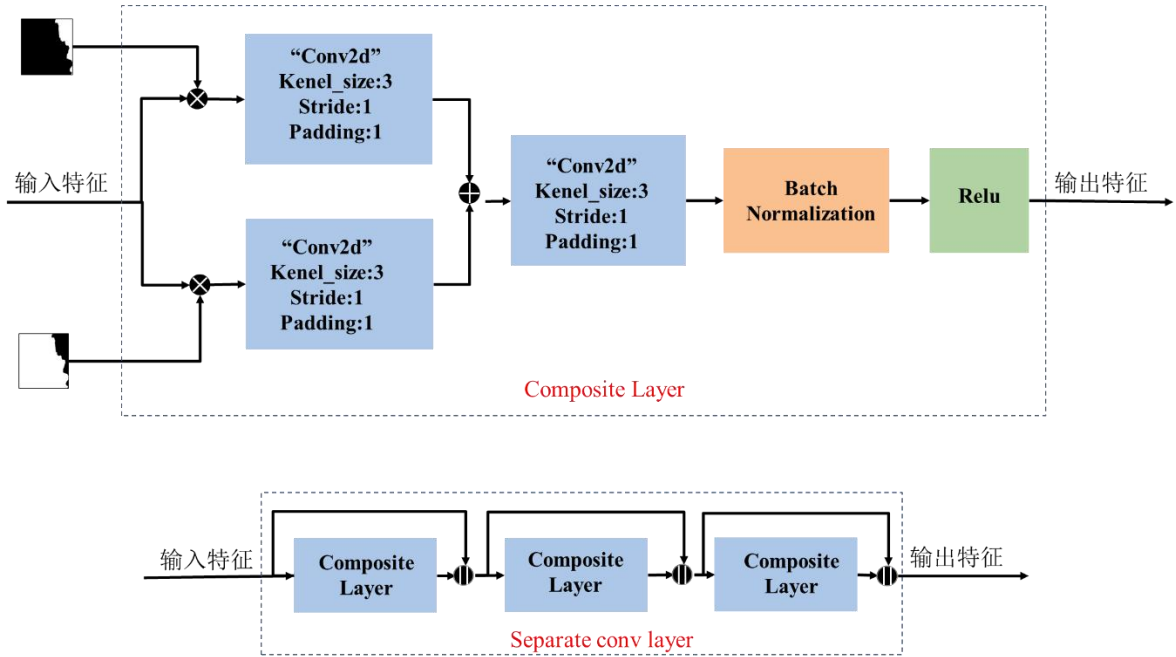


图 2-7 分离卷积模块网络结构

2.2.3 损失函数

本文的去雾网络主要完成两个任务，一个是透射率图的估计，一个是去雾，所以损失函数主要分为两部分：估计透射率图的损失函数、去雾损失函数。

估计透射率图的损失函数。本文设计的透射率图估计损失函数 L_{trans} 包含两部分：基于像素值的误差函数 L1 损失函数 L_{l1_t} 和梯度损失函数 L_{grad} 。 L_{l1_t} 损失函数的定义如下式所示

$$L_{l1_t} = \sum_{w,h,c} \|T(I) - t\|_1 \quad (2-10)$$

其中 T 表示透射率图估计网络， I 表示输入雾图， t 为目标透射率图， w 、 h 、 c 分别为图像的长、宽和通道数。

为了对透射率图的边缘信息进行优化，本文在 L1 损失函数的基础上引入梯度损失函数 L_{grad} ， 计算方式如式2-11所示：

$$L_{grad} = \sum_{w,h} \|G_x(T(I)) - G_x(t)\|_2 + \|G_y(T(I)) - G_y(t)\|_2 \quad (2-11)$$

其中 G_x 表示水平梯度计算函数， G_y 表示垂直梯度计算函数。

图像去雾的损失函数。去雾的损失函数为基于像素值的 L1 损失函数，定义如下：

$$L_{dehaze} = \sum_{w,h,c} \|S - C\|_1 \quad (2-12)$$

其中 S 为去雾后的清晰图， C 为目标清晰图。

综上所述，本文基于图像雾化程度的图像去雾算法的整体损失函数为：

$$L_{sum} = L_{dehaze} + L_{trans} \quad (2-13)$$

2.2.4 评价指标

峰值信噪比：峰值信噪比 PSNR 通常用来评价图像重建失真度，分贝值越高表示图像重建质量越好，本文采用 PSNR 计算雾图复原后的清晰图相比真实清晰图的失真度，其计算公式如下：

$$PSNR = 10 \log_{10} \left(\frac{P_{peak}}{MSE} \right) \quad (2.14)$$

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (p(i, j) - g(i, j))^2 \quad (2.15)$$

其中 p_{peak} 为图像灰度值最大值 255，MSE 为均方差值， $p(i, j)$ 和 $p(i, j)$ 表示雾图复原后的清晰图和真实清晰图的像素值。

结构相似度：结构相似度 SSIM 是用来评价两张图像的结构信息，是对两张图像的亮度、对比度和协方差进行综合计算的评价指标，数值越大，代表两张图像的结构越具有一致性，本文采用 SSIM 计算真实清晰图与雾图复原后的清晰图间的结构相似度，其计算公式如下：

$$SSIM(f, g) = \frac{(2\mu_f\mu_g + c_1)(2\sigma_{fg} + c_2)}{(\mu_f^2 + \mu_g^2 + c_1)(\sigma_f^2 + \sigma_g^2 + c_2)} \quad (2.16)$$

其中 f 、 g 分别为参与评价的两张图像， μ_f 、 σ_f^2 分别为图像 f 的均值和方差， μ_g 、 σ_g^2 为图像 g 的均值和方差， σ_{fg} 为图像 f 、 g 的协方差。

2.2.5 模型训练与实验结果分析

(1) 数据集介绍

由于没有实拍的真实的成对雾图、清晰图、透射率图数据库，所以目前被广泛采用的是利用 NYU_depth 数据库合成的雾图数据库。但是 NYU_depth 数据库的场景都是室内图像，而去雾算法的主要应用场景是室外，所以本文基于 ReDWeb_V1 数据集合成了一个场景丰富的图像去雾数据库。ReDWeb_V1 数据库^[46]是一个包含室内、室外图像及图像对应深度信息的数据库。给定清晰图 J 和深度图 d ，生成随机大气光值 A ($A \in (0.8, 1)$)、大气散射系数 β ($\beta \in (0.4, 1.6)$)，根据大气散射模型合成有雾图像。最后将清晰图和合成的雾图的大小规范为 256×256 。

本文采用公开数据集 RESIDE 中的 SOTS 数据集作为算法性能评价的测试数据集，该数据集中有 500 对户外的雾图、清晰图图像对，与训练数据库没有重叠。

(2) 估计传输图的损失函数。

为了说明估计透射率图损失函数的有效性，本文针对相同的网络结构尝试了不同的损失函数：(i) L1 损失，(ii) L1 损失+梯度损失。实验结果如图 2-8 所示。我们可以看到，使用 L1 损失函数和梯度损失函数获得的透射图边缘信息更加平滑。



图 2-8 损失函数对透射率图估计的影响

(3) 训练过程与实验结果展示与分析

根据 2.2.2 节搭建的去雾网络，采用基于 ReDWeb_V1 的合成数据集开展实验，经过 6000 次迭代后网络收敛，网络整体损失函数 L_{sum} 、去雾损失函数 L_{dehaze} 和估计透射率图损失函数 L_{trans} 变化曲线如图 2-9 所示，整体上看，震荡幅度变小，并且保持下降趋势，在 4000 步之后损失函数下降逐渐变缓，最终模型收敛。

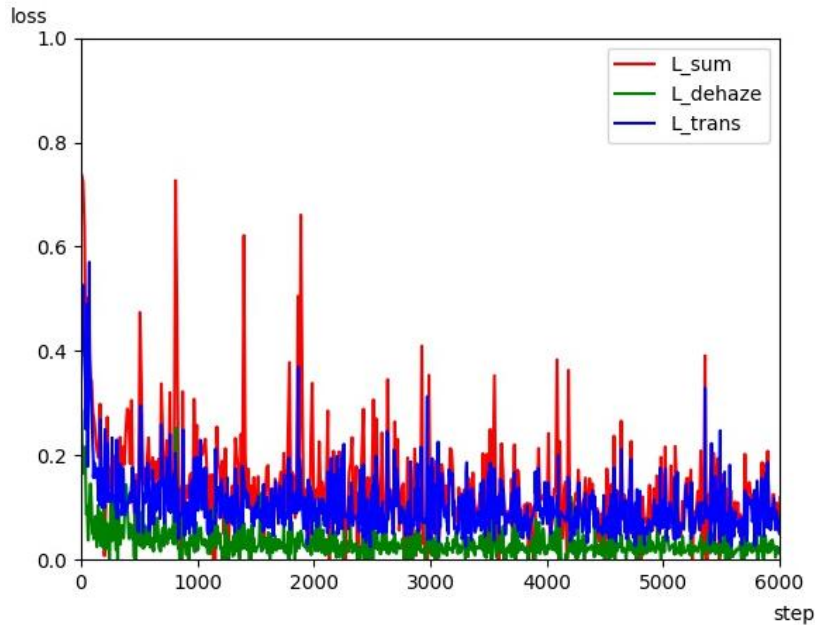


图 2-9 去雾损失曲线

本文将阈值估计组件计算得到的弱雾掩码和强雾掩码图进行了可视化，如图

2-10 所示，第一列为真实场景中的雾图、第二列为弱雾掩码图、第三列为强雾掩码图、第四列为网络估计的透射率图。弱雾掩码在雾化程度较小区域像素值为 1，其他区域为 0，强雾掩码与弱雾掩码相反。从图中我们可以看出，透射率图反映了图像的雾化程度，雾化程度较大的图像区域对应的透射率图灰度值较小。同时根据阈值估计组件计算得到阈值将图像区域进行分割，分割后得到的掩码图很好地反映了雾图中不同雾化程度的图像区域，证明了分离去雾模块对不同雾化程度区域分离并行去雾的有效性。

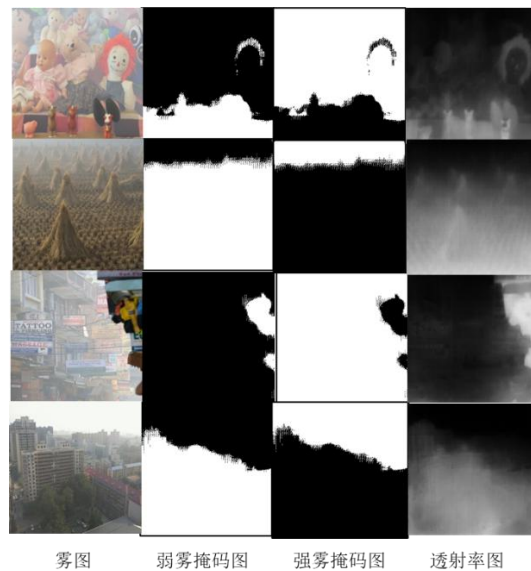


图 2-10 雾图对应的弱雾掩码图、强雾掩码图

在公开数据集 SOTS 的户外数据集上，将本文算法与其他性能较好的基于先验知识的去雾算法（DCP^[36]、CAP^[37]）和基于深度学习的去雾算法（AOD-net^[40]、PFF-net^[41]、Dehaze-GAN^[42]、DHSGAN^[43]）进行比较，对比结果如表 2-1 所示。本文提出的算法虽然在 PSNR 上略差于 DHSGAN 算法，但是实现了最优的 SSIM 值。PSNR 计算两张图像每个像素点的绝对误差，是误差敏感型的评价指标。而 SSIM 从两张图像的亮度、对比度、结构上进行综合考虑，在图像复原领域相比 PSNR 有更高的准确度。实验表明该本文提出的算法有更高的复原准确度。

表 2-1 不同去雾算法去雾效果

Model	PSNR	SSIM
-------	------	------

DCP ^[36]	19.155	0.899
CAP ^[37]	18.481	0.773
AOD-net ^[40]	20.292	0.875
PFF-net ^[41]	21.252	0.839
Dehaze-GAN ^[42]	23.595	0.909
DHSGAN ^[43]	26.612	0.910
本文算法	24.827	0.925

图 2-11 为 6 个真实场景的有雾图像及几种算法生成的相应的去雾结果。从第一、二、五行可以看出 AOD、Dehaze-GAN 和 DHSGAN 保留了较多的雾霾，同时 DCP 和 DHSGAN 使得天空区域过度增强。观察第四行，DCP、AOD、Dehaze-GAN 的去雾结果都出现了局部区域过暗、信息丢失的现象（中下树木部分），而 DHSGAN 整去雾不彻底。在第三行中，DCP、AOD、Dehaze-GAN 去雾结果的楼房（左角）过暗，与本文算法得到的结果形成鲜明对比。在最后一行可以发现本文提出的去雾网络较好的恢复了真实的色彩信息，天空区域色彩信息没有发生畸变。

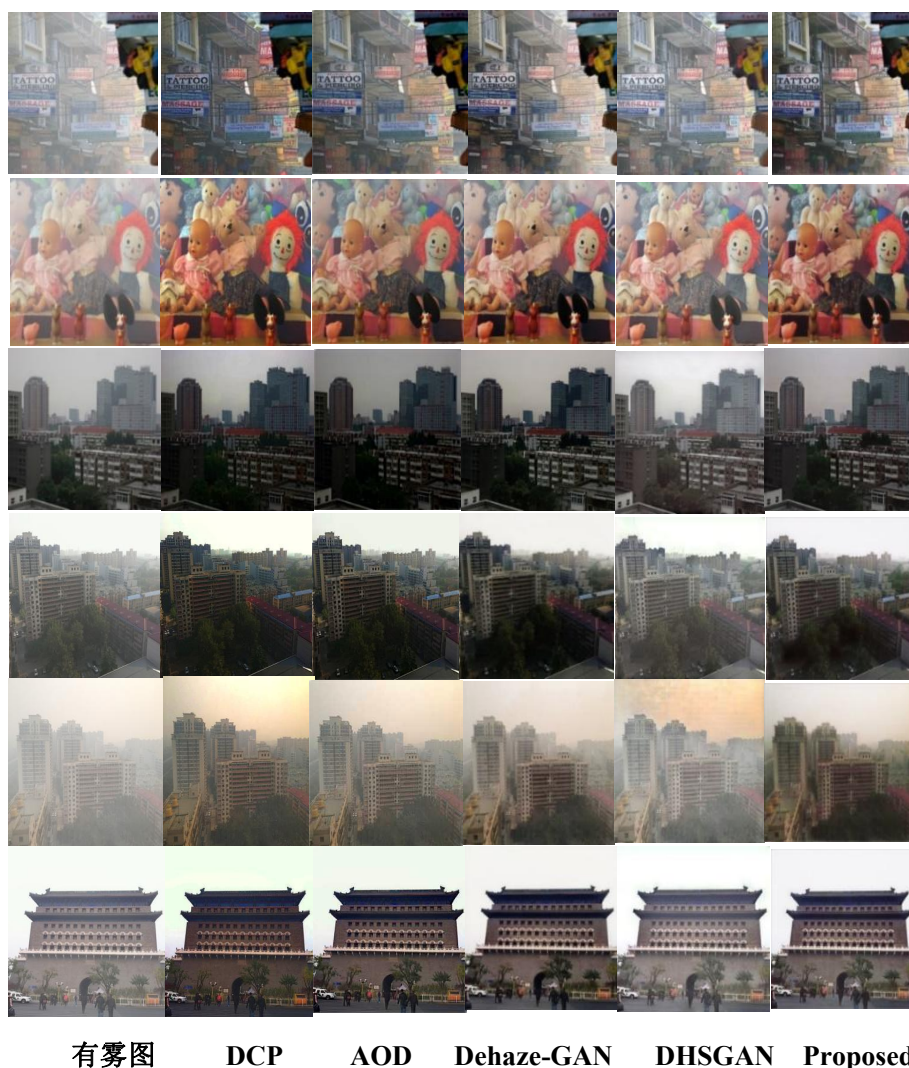


图 2-11 真实图像去雾方法的比较

2.3 本章小结

本章主要针对雾图局部区域雾化程度不同的问题，提出了一种基于图像雾化程度的图像去雾算法。考虑到雾图对应的透射率图反映了图像的雾化程度，算法添加了与去雾模块并行的透射率图预测模块；根据得到的透射率图，设计阈值自适应学习模块计算分割阈值，生成不同等级雾化程度的掩码图，指导分离去雾卷积模块恢复无雾图。通过在公开数据集上的性能评估和对真实图像的去雾效果进行比较，本文提出的去雾网络相比其他去雾算法实现了更高精度的去雾能力。

3 基于生成对抗网络的异源红外图像生成算法

基于可见光图像的红外图像生成方法能够解决红外图像获取难、采集成本高的问题，本章研究了基于生成对抗网络的红外图像生成方法，采用 Pix2pix 网络来自动学习可见光图像到红外图像的映射关系。针对 Pix2pix 网络生成的红外图像细节信息缺失的问题，本文提出了多感受野特征融合 Pix2pix 网络，进一步优化生成的红外图像的细节信息，提高仿真图像的真实度。

3.1 异源红外图像生成相关理论介绍

3.1.1 红外成像影响因素

红外辐射存在于世界各个角落，自然界中一切温度高于绝对零度的物体都会发出红外辐射，并且表面温度越高，产生的红外辐射越强。红外线是一种电磁波，波长范围介于 $0.75\mu m \sim 1000\mu m$ ，根据其波长范围，我们可以将红外光谱分为三个波段： $0.7\mu m \sim 2.5\mu m$ 间的红外光谱成为近红外、 $2.5\mu m \sim 25\mu m$ 间的红外光谱成为中波红外、 $25\mu m \sim 1000\mu m$ 间的红外光谱成为远红外。红外图像的成像效果主要与场景温度、红外成像设备波长范围、大气传输介质相关。接下来主要分析温度和波长对红外成像效果的影响。

普朗克定律是描述红外热辐射的基本定律之一，该定律定量地分析了黑体的光谱辐出度与温度和波长的关系，公式如下：

$$M_{b\lambda} = \frac{c_1}{\lambda^5} \cdot \frac{1}{e^{c_2/\lambda T} - 1} \quad (3-1)$$

其中， $M_{b\lambda}$ 表示辐出度， c_1 表示第一辐射常量， c_2 表示第二辐射常量， λ 表示波长， T 表示温度， k 表示玻尔兹曼常数。 $c_1 = 3.7413 \times 10^{-16} W \cdot m^2$ ， $c_2 = 1.43879 \times 10^{-2} W \cdot m^2$ ， $k = 1.380662 \times 10^{-23} J \cdot K^{-2}$ 。

图 3-1 绘制了黑体的光谱辐出度与温度和波长的关系曲线，从图中我们可以看

到不同温度下黑体辐出度随着波长变大变化趋势大致相同，先大幅升高，到达峰值逐步减小，随着温度升高，波长峰值在逐步减小。根据维恩位移定律，黑体辐出度峰值对应的波长与黑体温度成反比。对公式计算波长 0 到无穷的积分，可以得到黑体辐出度与温度的关系，也就是玻尔兹曼定律，如式所示：

$$M_0(T) = \sigma T^4 \quad (3-2)$$

其中 $\sigma = 5.6694 \times 10^{-8} W / (m^2 \cdot K^4)$ 。玻尔兹曼定律表明黑体辐出度与温度的四次方成正比。

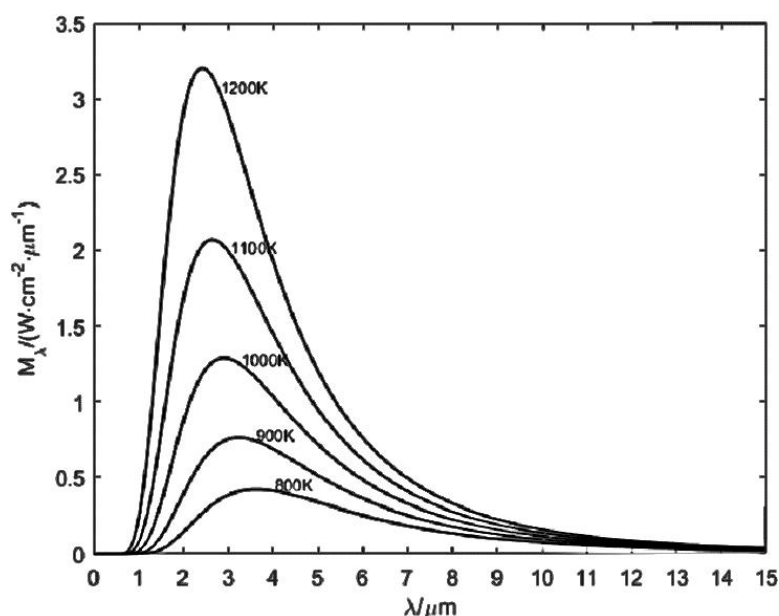


图 3-1 黑体辐出度与温度和波长的关系

以上分析均是针对黑体的讨论，而黑体是一种理论模型。为了将上述物理分析与实际场景相适用，便需要量化黑体辐射与实际场景的关系。发射率是一种量化目标物体辐射量与黑体辐射量之比的物理量，比值越接近 0 则目标物辐出量与黑体辐出量相差愈大。这样我们只要知道任意目标物发射率，便可以利用黑体辐射理论进行目标物辐射分析。不同物体由于物理属性不同，发射率也不同。一般地，金属的发射率与温度成正比，非金属的发射率与温度成反比。物体的反射率与波长也有关系，通常随着红外波长增大，对应的发射率逐渐减小。

3.1.2 生成对抗网络基础

生成对抗网络（GAN）是一种基于零和博弈思想的生成模型。与传统的网络模型不同，GAN 网络由生成网络和判别网络两个子网络组成，其模型结构如图 3-2 所示。生成网络的目标是生成符合真实数据分布的样本，并将学习到的映射函数应用到给定的噪声 z 中。判别器的目标是能够准确地鉴别来自真实数据分布的真实样本和由生成器生成的生成样本。在训练过程中，生成器从判别器的决策中得到反馈，进一步优化模型参数，学习下次如何更好地欺骗判别器。假设真实样本 x 符合真实数据分布 $P_r(x)$ ，隐向量 z 符合定义的隐式向量分布 $P_z(z)$ ，例如均匀分布或球形高斯分布。对 $P_z(z)$ 分布采样得到向量 z ，然后输入生成网络生成 $x' = G(z)$ ，将生成的数据 x' 和真实数据 x 分别输入判别器，我们期望判别器能够具有准确的判别能力输出对应的 0（Fake）和 1（Real）。实质上，判别器实现了一个二分类模型的功能，可以采用交叉熵损失函数进行模型训练。同时，生成网络则希望生成数据逐渐符合真实数据分布，能够被判别网络判别为 1，基于此，生成对抗网络的损失函数的定义为：

$$\min_G \max_D V(G, D) = E_{x \sim P_r} [\log D(x)] + E_{z \sim P_z} [\log(1 - D(G(z)))] \quad (3-3)$$

其中 $V(G, D)$ 为生成对抗损失函数，通过极小化 $V(G, D)$ 训练生成网络，极大化 $V(G, D)$ 训练判别网络。

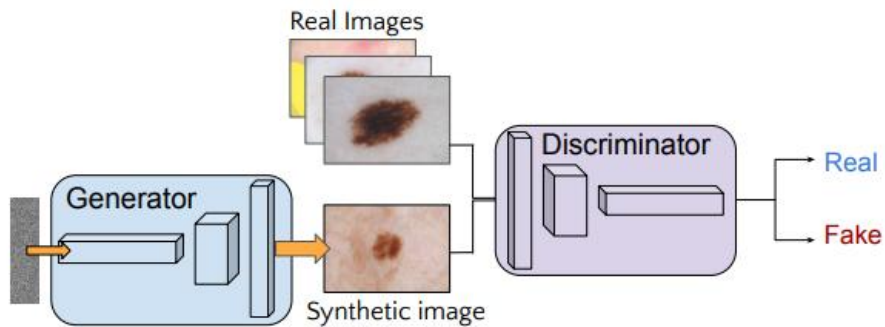


图 3-2 生成对抗网络模型结构

3.1.3 异源红外图像生成可行性分析

本文主要针对可见光图像对应同一波段和同一时段的红外图像间的变换关系，

拟定波长范围为 $7.5\sim 13\ \mu\text{m}$ 、时段为 14:00 的红外图像, 图像均在同一时段采用统一的拍摄条件采集。由 3.1.1 节的分析可知, 不同目标物体的发射率不同, 对应的红外辐射量不同。在传统的红外仿真领域, 需要将可见光图像进行材质分割, 分别计算每种材质在对应波段和大气条件下的红外辐射, 可见不同材质与对应红外图像亮度分布有物理相关性。可见光图像一般都具有较好的纹理信息和对比度, 根据可见光图像获得的场景信息, 如果能够获得不同材质间可见光图像到红外图像的映射关系, 那么就可以利用该映射关系来实现可见光图像到对应场景的红外图像的生成。

生成对抗网络在图像生成任务中有广泛的应用, 可以学习两个图像域间的映射关系。通过搭建生成器可以拟合不同材质可见光图像到红外图像的复杂的变化函数, 并能够充分考虑到不同材质红外辐射间的相互影响导致的红外成像效果的改变。因此可以搭建生成对抗网络来实现异源红外图像生成, 通过生成器与判别器的博弈学习得到可见光图像到红外图像的映射函数。本文采用生成器网络学习可见光图像域与中波红外图像域间的映射关系, 实验表明生成对抗网络适用于异源红外图像生成任务。

3.2 基于 Pix2pix 的异源红外图像生成方法

3.2.1 Pix2pix 网络

本文将由可见光图像生成红外图像的异源红外图像生成问题建模为两个图像域间的图像翻译问题。在本文已知的范围内, 采用 GAN 网络进行红外图像生成的算法很少。近几年的 GAN 网络的衍生算法中, 算法主要解决的是生成图像的多样性、非配对图像翻译以及高分辨率的图像生成问题, 而本文专注于由可见光图像生成红外图像的准确性, 基于此, 本文选择了由 Isola 在 2017 年 CVPR 会议上提出的 Pix2pix 网络^[27]。Pix2pix 网络是采用条件生成对抗网络 (CGAN) 的结构来解决图像翻译任务的通用网络架构。该网络不仅能够学习图像到图像间的映射关系, 同时能够根据样本数据特征学习针对特定翻译任务的损失函数, 使得条件生成对抗网络适用于图像翻译任务。

条件生成对抗网络是首次将条件标签引入生成对抗网络的结构，一定程度上改善了原始生成对抗网络由于任何不加约束导致生成模式过于自由的缺点。与基于离散标签、文本的条件对抗网络不同，Pix2pix 主要针对图像到图像的翻译任务，目的是将高分辨率的输入映射到高分辨率的输出结果，故 Pix2pix 将图像作为条件标签。考虑到参与映射的图像对结构大致对齐，生成器采用 U-net 结构^[47]，其结构类似于编码解码网络，输入图像首先经过网络的编码部分来降低特征图的分辨率，之后采用解码网络恢复到原始图像分辨率，同时利用跨层通道连接使得高层特征图能够感知低层特征图的细节信息。对于判别器结构，原始 GAN 的判别器网络对于要求高分辨率、高细节保真度的图像翻译任务并不适用，Pix2pix 采用 PatchGAN 判别器结构。传统 GAN 网络的输出为 ‘True’ 或者 ‘False’ 的布尔值，表示对于判别器输入的整张图像的判定，而 PatchGAN 判别器的输出为 $N \times N$ 的矩阵，矩阵中的每个像素代表对应于输入图像感受野 $M \times M$ 大小的图像块的判定，从而使得判别器更多的关注于局部细节信息的恢复。

本节尝试将 Pix2pix 网络应用到异源红外图像生成任务中，利用该网络学习可见光图像到红外图像的映射关系，搭建的模型框架如下图 3-3 所示：

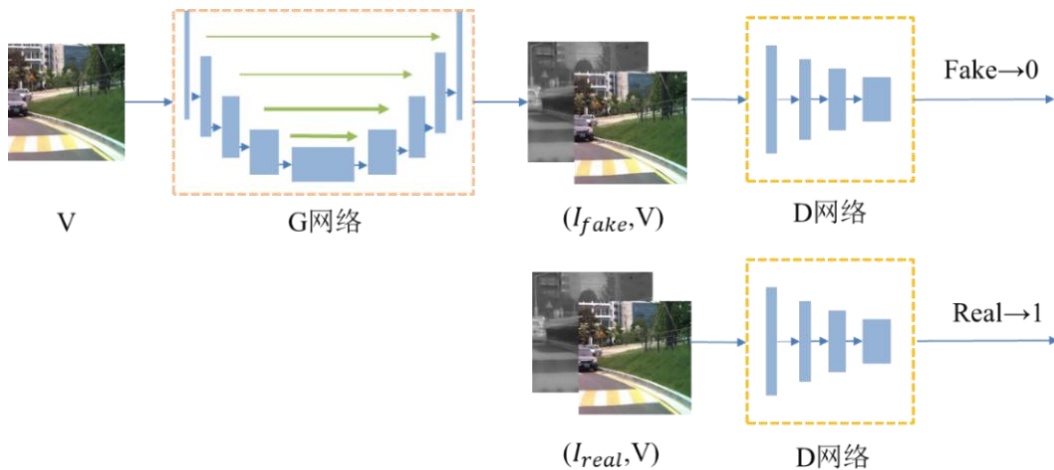


图 3-3 Pix2pix 模型结构

图中 V 表示样本对中的可见光图像， I 表示样本对中的红外图像， I' 表示生成器生成的红外图像；训练样本对描述为 (V, I) 。Pix2pix 网络由生成器和判别器组成，生

成器表示为 G 网络，判别器表示为 D 网络。

3.2.2 Pix2pix 模型构建分析

生成器结构。生成器为 U-net 结构，类似于编码解码网络。对于编码器和解码器卷积层数的确定，本文针对不同层数的网络结构进行了实验，最终采用的结构为：编码部分由 5 层卷积层组成，用来提取图像特征，解码部分由 5 层反卷积层组成，利用提取到的图像特征生成红外图像，本文搭建的生成器结构如图 3-4 所示：

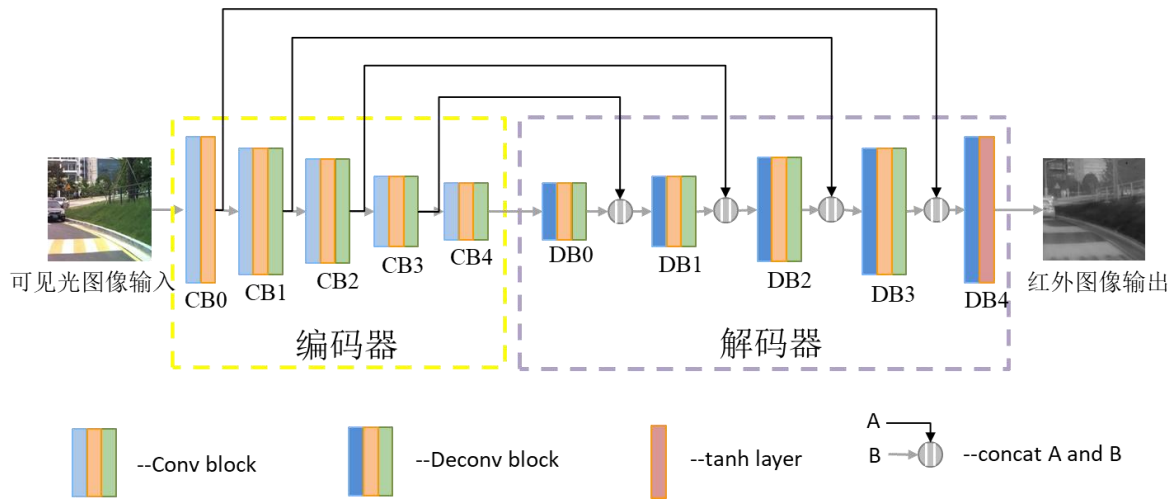


图 3-4 Pix2pix 生成器网络结构

图中黄色虚线部分为编码器，紫色虚线部分为解码器，黑色实线为编码器和解码器间的跨层连接。其中编码模块和解码模块示意图如图 3-5、图 3-6 所示：

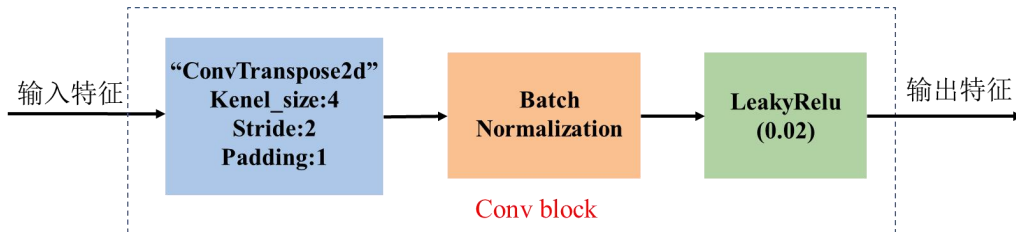


图 3-5 Pix2pix 编码器子模块网络结构

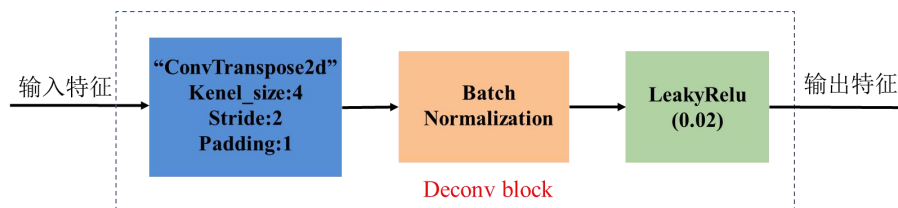


图 3-6 Pix2pix 解码器子模块网络结构

G 网络的参数见表 3-1，表中卷积层参数(f,k,s,p)分别表示卷积核数量、大小、步长、填充尺寸，输入参数和输出参数(b,c,h,w)分别表示训练批大小、通道数、特征图高、特征图宽。由表可知，编码部分通过卷积层设置步长为 2 来进行下采样，解码层每层特征图都与编码层相同分辨率的特征图进行通道上的连接，并通过反卷积逐步恢复图像的分辨率，保证了输出和输入图像的分辨率相同。

表 3-1 Pix2pix 生成器网络参数表

模块	卷积块名称	卷积层参数 (f,k,s,p)	输入特征 (b,c,h,w)	输出特征 (b,c,h,w)
输入	-	-	(1, 3,256,256)	-
编码器	CB0	(64,4,1,2)	(1, 3,256,256)	(1,64,128,128)
	CB1	(128,4,1,2)	(1,64,128,128)	(1,128,64,64)
	CB2	(256,4,1,2)	(1,128,64,64)	(1,256,32,32)
	CB3	(512,4,1,2)	(1,256,32,32)	(1,512,16,16)
	CB4	(512,4,1,2)	(1,512,16,16)	(1,512,8,8)
解码器	DB0	(512,4,1,2)	(1,512*2,8,8)	(1,512,16,16)
	DB1	(256,4,1,2)	(1,512*2,16,16)	(1,256,32,32)
	DB2	(128,4,1,2)	(1,256*2,32,32)	(1,128,64,64)
	DB3	(64,4,1,2)	(1,128*2,64,64)	(1,64,128,128)
	DB4	(3,4,1,2)	(1,64*2,128,128)	(1,3,256,256)

判别器结构。D 网络用来辨别输入的“真假”，实质上是一种可训练的损失函数。与原始 GAN 网络不同，本文搭建的 D 网络采用了 PatchGAN 的思想，网络的输出为 30×30 的矩阵，从而输出的每个像素值对应于输入 70×70 大小的图像块，对于图像翻译任务来说，能够实现重现更多的细节信息。本文搭建的 D 网络的结构如图 3-7 所示，其中卷积模块与 G 网络相同。

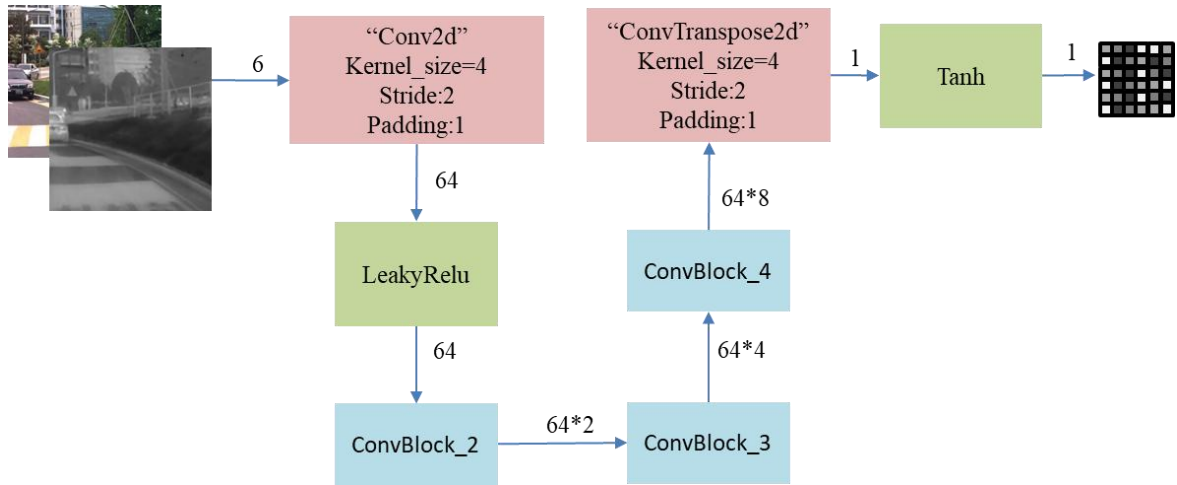


图 3-7 Pix2pix 判别器网络结构

Pix2pix 的判别器 D 网络的输入为图像对，将可见光图像分别与真实红外图像、生成的红外图像在第三通道维度进行拼接，作为判别器的真样本对和假样本对。D 网络的参数见表 3-2，表中卷积层参数(f,k,s,p)分别表示卷积核数量、大小、步长、填充尺寸，输入参数和输出参数(b,c,h,w)分别表示训练批大小、通道数、特征图高、宽。

表 3-2 Pix2pix 判别器网络参数表

模块	卷积层参数 (f,k,s,p)	输入特征 (b,c,h,w)	输出特征 (b,c,h,w)
输入	-	(1, 6,256,256)	-
下采样	(64,4,1,2)	(1, 6,256,256)	(1,64,128,128)
	(128,4,1,2)	(1,64,128,128)	(1,128,64,64)
	(256,4,1,2)	(1,128,64,64)	(1,256,32,32)
	(512,4,1,1)	(1,256,32,32)	(1,512,31,31)
	(3,4,1,1)	(1,512,31,31)	(1,3,30,30)

由表可知，D 网络输入为(batchsize, 6 , 256, 256)样本对，经过三次下采样，根据表中数据我们可以计算 D 网络输出矩阵每个值多对应原图的感受野。在卷积神经网络

络中，感受野用来表示每层卷积层输出的特征层中的像素点相对于输入图像的映射范围大小。感受野计算公式为：

$$RF^N = \begin{cases} 1, N = 0 \\ k_1, N = 1 \\ RF^{N-1} \times k_n - (k_n - 1) \times (RF^{N-1} - \prod_{i=1}^{n-1} s_i), N \geq 2 \end{cases} \quad (3-4)$$

其中 RF^N 表示第 N 层卷积层的感受野，同理 RF^{N-1} 表示第 $N-1$ 层特征层的感受野， k_n 表示第 n 层卷积核大小， s_i 表示第 i 层卷积层的步长。根据公式，D 网络每层卷积层的感受野为：

Conv1: $k_1=4, s_1=2, RF^1=4$;

Conv2: $k_2=4, s_2=2, RF^2=10$

Conv3: $k_3=4, s_3=2, RF^3=22$

Conv4: $k_4=4, s_4=2, RF^4=46$

Conv5: $k_5=4, s_5=2, RF^5=70$

由上述分析可知，判别器输出 30×30 的矩阵中，每个像素对应原图像 70×70 像素的图像块，体现了对图像局部真假信息的判断。

3.2.3 Pix2pix 损失函数

Pix2pix 的损失函数由 G 网络损失函数和 D 网络损失函数组成，网络的训练过程 G 网络和 D 网络的博弈过程，G 网络的目标为最小化 L_{pixGAN} ，D 网络的目标为最大化 $L_{pixGAN}(G, D)$ 如式 3-5 所示：

$$L_{pixGAN}(G, D) = E_{V, I}[\log D(V, I)] + E_V[\log(1 - D(V, G(V)))] \quad (3-5)$$

同时考虑到 G 网络的任务不仅是“欺骗”判别器，同时还要实现图像翻译任务，使得生成的图像无限接近样本对中真实的红外图像。因此，G 网络的损失函数加入

了 L1 损失函数，最终 G 网络和 D 网络的损失函数如式 3-6、3-7 所示：

$$L_{G_net}(G) = L_{pixGAN}(G) + \lambda L_{l1}(G) = E_V[\log(1 - D(V, G(V)))] + \lambda L_{l1}(I, G(V)) \quad (3-6)$$

$$L_{D_net}(D) = -L_{pixGAN}(D) = -(E_{V,I}[\log D(V, I)] + E_V[\log(1 - D(V, G(V)))] \quad (3-7)$$

其中 λ 为损失项 $L_{l1}(I, G(V))$ 的权重。

G 网络的损失函数在 $L_{pixGAN}(G)$ 损失的约束下使得生成的红外图像分布符合目标图像分布，在 $L_{l1}(G)$ 损失的约束下使得生成的红外图像与目标图像在像素级上保持灰度一致性。

3.2.4 模型训练与实验结果分析

实验中采用的数据集 Kaist 实验室发布车载道路场景数据集。数据集中红外图像为 FLIR 红外摄像头拍摄，范围为 7.5~13um，选定的时段为 14:00，同时可见光图像与对应的红外图像结构相同、一一配对。整个数据集的场景具有一定的相似性，包括道路、树木、草地、车辆等景物及少量的行人干扰。图像的分辨率为 256×256。训练数据集包含 2500 对可见光-红外图像样本对，测试数据集共有 500 对。

对于训练参数，本文将 batchsize 设置为 1，总共进行 140000 次迭代，前 20000 次迭代学习率保持 0.0001 不变，之后每 1000 次迭代学习率下降 0.0001/120 直到学习率为 0。本文采用 Adam 优化器， β_1 、 β_2 分别设置为 0.5、0.999。

为了针对可见光图像生成红外图像这一任务设计更合适的网络结构，在生成器拓补结构为 U-net 网络的约束下，本文针对生成器的网络的多种深度结构进行实验，实验结果如表 3-3 所示，所以最终采用深度为 5 的 U-net 结构。

表 3-3 不同深度 U-net 网络算法性能

网络结构	PSNR	SSIM	模型大小 (M)
U-net4	26.562016	0.784729	6.171
U-net5	26.645653	0.793285	16.659
U-net6	26.613107	0.792967	29.244

U-net7	26.596768	0.789918	41.829
--------	-----------	----------	--------

根据可见光图像、红外图像图像对和设置的损失函数对网络进行训练，经过 140000 次迭代之后，模型基本收敛。训练过程中的生成器和判别器的变化趋势分别如图 3-8、图 3-9 所示。训练开始时，生成器网络相对判别器拟合能力较弱，训练过程比较不稳定。随着训练的进行，生成器逐渐学到了图像域间的映射关系，生成器生成对抗损失函数逐步下降，同时判别器函数的震荡减弱。L1 损失函数从整体上看一直保持逐步减小的趋势。

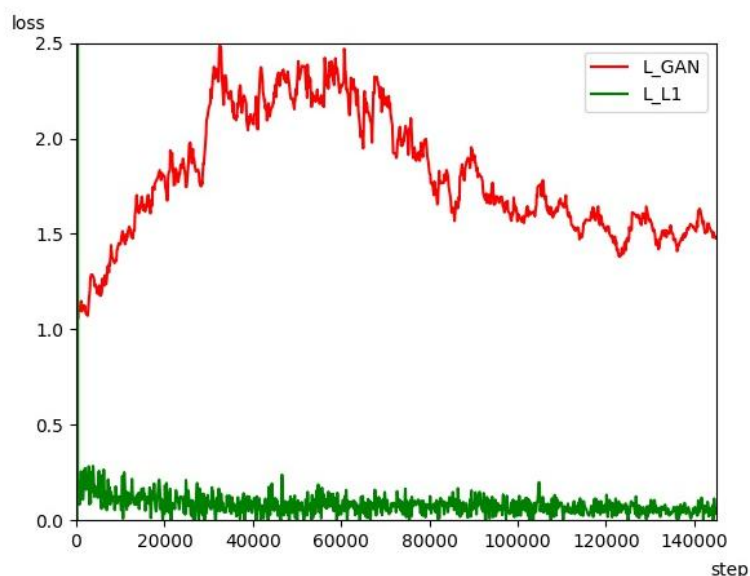


图 3-8 Pix2pix 生成器损失曲线

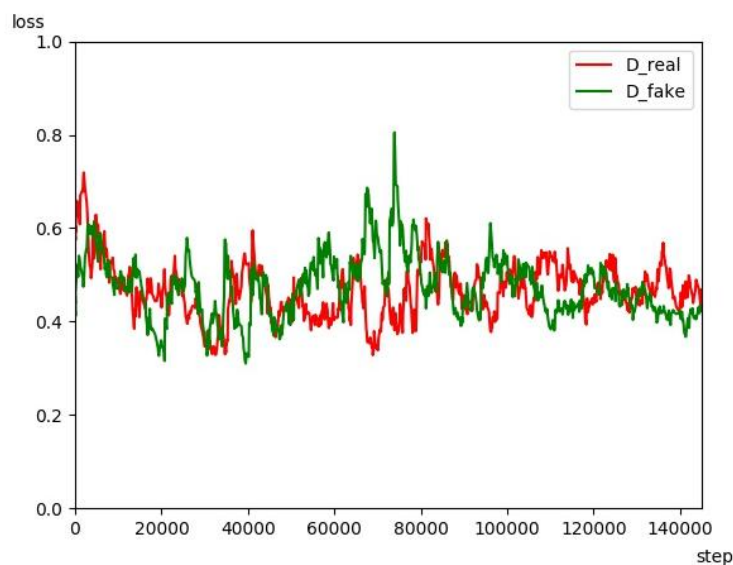


图 3-9 Pix2pix 判别器损失曲线

采用训练好的网络对测试集中的数据进行测试，得到的结果如图 3-10 所示，其中第一列为可见光图像、第二列为可见光图像对应的真实红外图像、第三列为 Pix2pix 网络生成的红外图像，红色标注部分为失真局部区域。从图中我们可以看出，生成图像效果方面，部分简单场景和道路两边的环境实现的较好的生成效果，整体上基本实现红外生成任务。我们观察到红色方框标注到的局部区域失真现象，第一行车主的轮廓信息没有生成，后排小轿车边缘信息模糊；第二行小轿车的车顶边缘生成不平缓；第三行，场景中有两个行人但是生成后我们只能观察到一个行人信息。虽然我们对 Pix2pix 网络针对异源红外图像生成任务进行了改进，但是仍然存在局部信息缺失的现象。

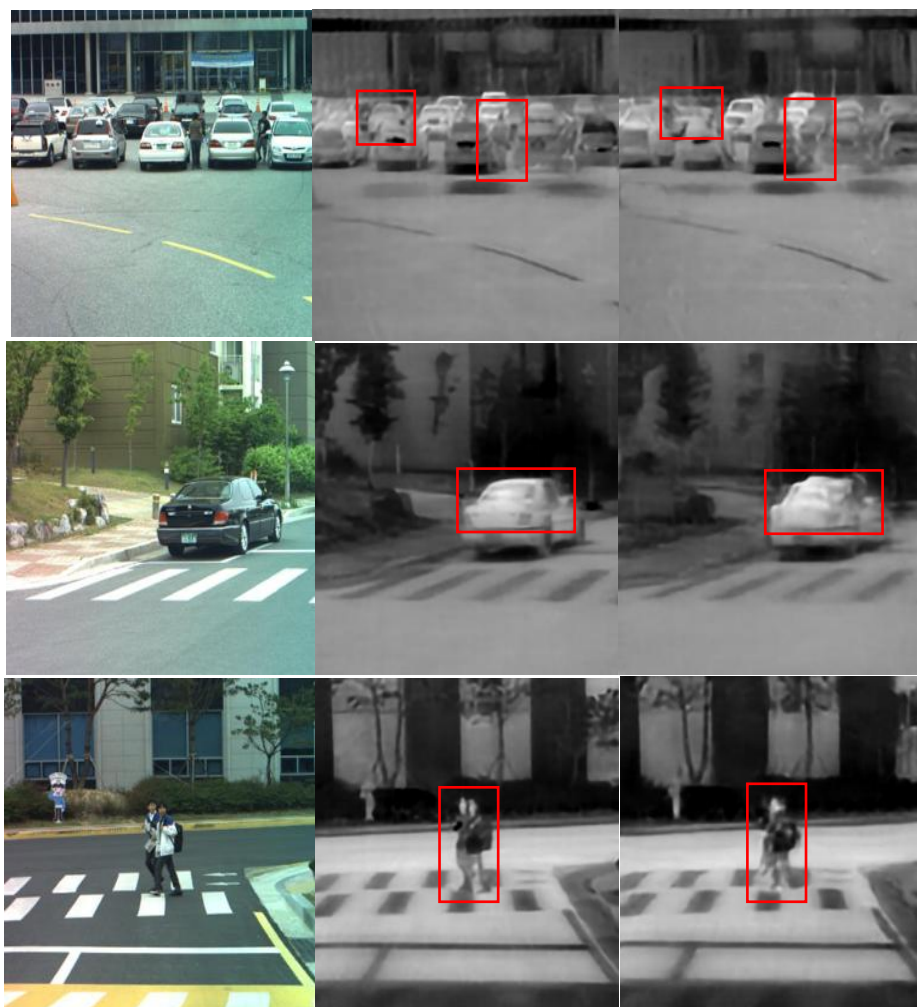


图 3-10 Pix2pix 实验结果

由以上展示的实验结果可知，Pix2pix 基本完成了由可见光图像生成红外图像的任务，生成的红外图像分布与真实的红外图像有较高的一致性，但是仍然存在转换失败的样本。由于成像机制不同，同一场景可见光图像与红外图像在分布上有很大差异，可见光图像有更分明的色彩特征和更清晰的纹理信息，而红外图像经过灰度量化后只有一个通道，并且大部分场景较可见光图像边缘信息更加模糊。一般我们认为浅层网络提取图像的浅层特征，例如边缘特征、颜色特征等，深层网络抽取高级语义特征，U-net 通过跨层连接，将浅层网络的特征与深层网络特征相连，一定程度上增强了浅层特征的信息。但是同时浅层特征与深层特征具有较大的特征表达级别的差别，而这种较大的特征间的语义差别可能会导致网络更侧重深层更抽象的特征，从而加大网络的训练难度。为了解决 Pix2pix 网络生成的红外图像细节缺失的问题，本文提出了基于多感受野特征融合 Pix2pix 的异源红外图像生成方法。

3.3 基于多感受野特征融合 Pix2pix 的异源红外图像生成方法

3.3.1 多感受野特征融合 Pix2pix 网络

在异源红外图像生成任务中，算法要保证由可见光图像生成的红外图像的准确度，尽可能生成更精细的红外纹理信息。本节针对 Pix2pix 网络生成的红外图像细节信息缺失的问题，进行了以下两点改进：

（1）多感受野特征提取器。为了进一步增强较小感受野特征的特征利用率，本文借鉴 Unet++ 网络的模型结构，在 3.2 节 Unet 网络的基础上搭建了基于 Unet++ 结构的多感受野特征提取器，提取到分辨率相同的不同感受野的图像特征

（2）多感受野特征融合机制。与广泛采用的在通道上进行合并的特征融合方式不同，本文通过逐级嵌套多个注意力机制模块，学习不同感受野特征像素级的权重，实现多感受野特征融合。

U-net ++^[48]是一种用于医学图像分割的功能更强大的网络结构。该架构本质上是一个深度监督的编码器/解码器网络，其中，编码器和解码器子网络通过一系列嵌套

的密集跨层连接路径进行整合。重新设计的跨层连接路径旨在减少编码器和解码器子网络的特征图之间的语义差异。Unet++架构背后的基本假设是，在与来自解码器网络的相应语义特征图融合之前，逐渐丰富来自编码器网络的高分辨率特征图，从而更有效地捕获前景对象的细粒度细节。当来自解码器和编码器网络的特征映射在语义上相似时，网络将更轻松地处理学习任务。这与 U-Net 中常用的普通跳过连接相反，后者直接从编码器向解码器网络提供对应分辨率特征图，从而导致语义差异较大的特征融合。

U-net ++由通过一系列嵌套的密集卷积块连接的编码器和解码器组成，网络结构如图 3-11 (a) 所示，其中黑色标注路径表示原始的 U-Net 结构，绿色标注路径和蓝色标注路径表示跨层连接路径上的密集连接方式，而红色标注路径表示添加的监督学习方式。由图 3-11 可以看到，Unet++网络是一种将“空心”的 U-net “填满”的特征提取器，将 4 种不同层次的特征通过 4 条不同的解码路径还原到原始高分辨率。不同感受野间特征的连接方式有 2 种：邻层短连接和跨层长连接，图 3-11 (b) 展示了 Unet++第一条跨层连接路径。邻层短连接保证了 Unet++网络实心区域（图 3-11 红色三角标注区域）在反向传播时的连续性，跨层长连接提高了不同层次特征的利用率。对于不同感受野特征的融合，U-net 网络在不同解码器输出路径后连接了卷积核大小为 1×1 的卷积层，之后计算所有路径输出的平均值。Unet++网络的改进主要是充分利用不同感受野的特征，感受野较大的特征能够充分识别图像中较大尺寸物体，而感受野较小的特征能保留图像中边缘信息，使得尺度较小物体的特征不至于丢失。

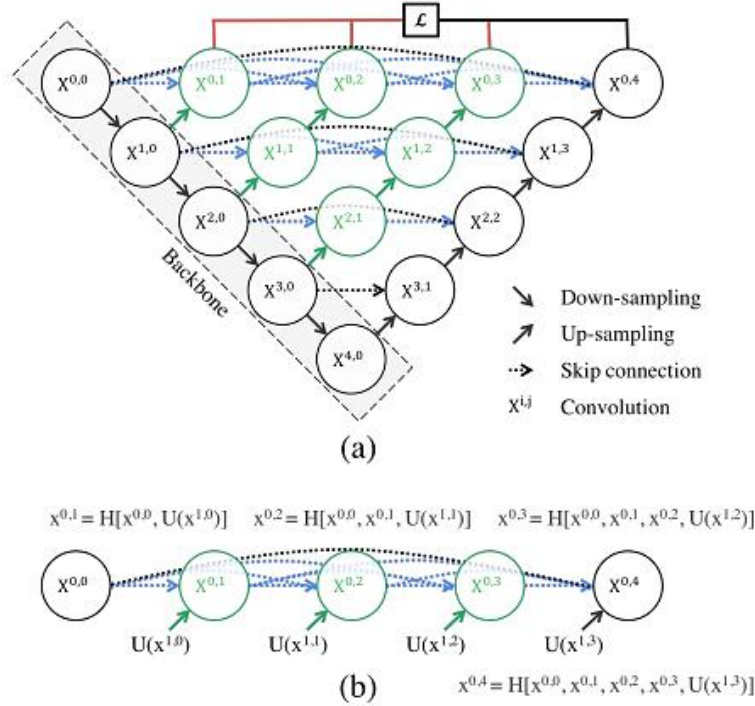


图 3-11 Unet++模型结构

图像翻译任务中同样存在深层网络较强的特征提取能力和图像边缘信息、局部信息丢失的矛盾，所以对于 Pix2pix 网络，本文采用 Unet++结构作为生成网络的特征提取器，在 3.3 节 U-net 结构的基础上添加 Unet++网络的实心部分。对于提取到的不同感受野的特征，本文设计了多感受野特征融合机制，使得网络能够自动学习不同感受野特征的重要程度。

最近，注意力机制已在许多图像处理任务中得到利用，包括视觉问答、图像分类和图像生成。SAGAN^[49]首次通过在生成器和判别器中添加自注意力模块来提高 GAN 网络对全局结构建模的能力。注意力机制通过采用一系列的权重系数，使得网络能够自动学习特征的重要程度，强调重要信息，同时抑制无关信息。常见的注意力机制有两种，一个是空间注意力机制，另一个是通道注意力机制。对于空间注意力模块，网络通过引入了注意机制来捕获特征图的任意两个位置之间的空间依赖性。对于某个位置上的特征，可通过使用在所有位置上特征的加权求和得到的聚合特征来进行更新，其中权重由相应两个位置之间的特征相似性决定。即，具有相似特征

的任何两个位置处的特征都可以相互促进，而无论它们在空间维度上的距离如何，从而能够实现长距离特征间的相互约束。对于通道注意力模块，网络使用类似的自我关注机制来捕获任意两个通道映射之间的通道约束关系，并使用所有通道映射的加权和更新每个通道映射权重。

空间注意力机制和通道注意力机制分别针对空间、通道两个维度进行特征汇总，并不适用于多感受野特征的特征融合。借鉴注意力机制的思想，本文提出了多感受野特征融合模块，其结构如图 3-12 所示。本文采用嵌套式的多感受野特征融合方式，每次设计一个注意力融合结构来融合两个感受野差异较小的特征，得到的新的特征再与较大的特征进行融合。对于注意力融合层，如图 3-12 右下方所示，指定两个特征 f_i 和 f_{i+1} 的特征图，我们引入一个像素级的注意力机制模块来学习这两个特征的像素级的权重。对于权重的学习，首先将输入融合块的两个特征进行通道上的合并 $[f_i, f_{i+1}]$ ，之后采用卷积层学习像素级的注意力图，同时映射函数将注意力图中的权重限制到 $[0,1]$ ，权重学习如下式所示

$$\alpha_i = g(H([f_i, f_{i+1}])) = \frac{1}{1 + e^{-H([f_i, f_{i+1}]})} \quad (3-8)$$

其中 g 表示映射函数， H 表示卷积函数。

根据得到注意力图 α_i ，特征 f_i 和 f_{i+1} 具体的融合方式如式 3-9 所示：

$$FeatureFuse(f_i, f_{i+1}) = f_i \odot \alpha_i + f_{i+1} \odot (1 - \alpha_i) \quad (3-9)$$

其中 f_i 和 f_{i+1} 两个特征的注意力权重大小负相关

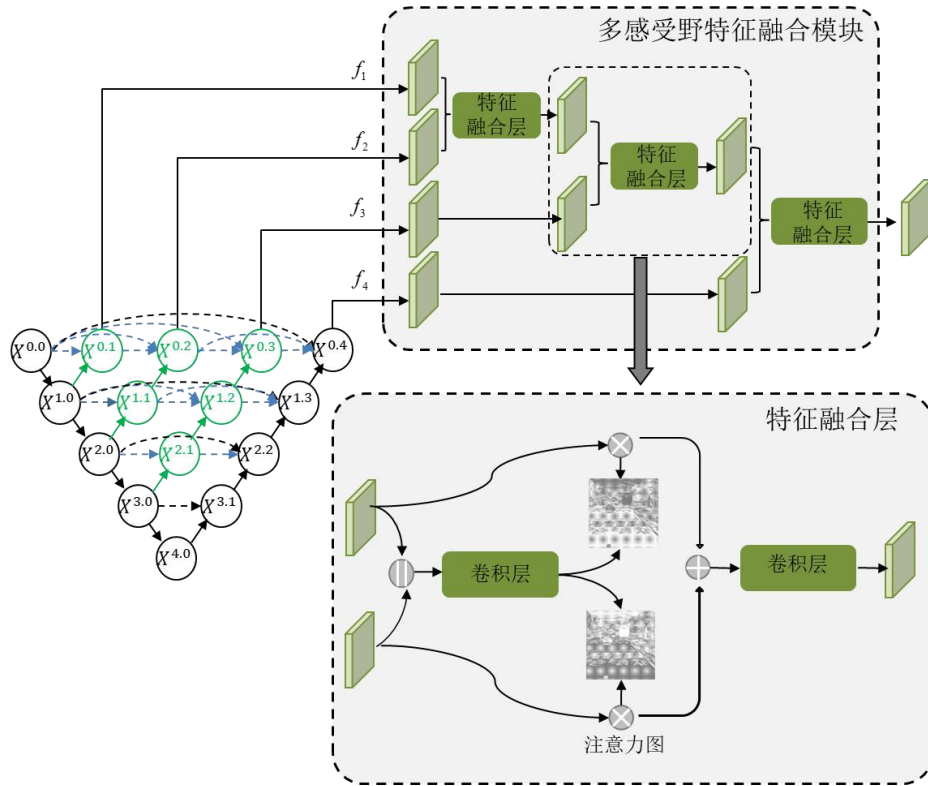


图 3-12 多感受野注意力融合模块

本文设计的多感受野特征融合 Pix2pix 网络的总体网络模型如图 3-13 所示。生成器网络首先通过采用 Unet++提取多感受野特征，然后利用多感受野特征融合模块自动地学习不同感受野特征的像素级的权重，并且得到融合后的特征，生成对应的红外图像。判别器网络与原始 Pix2pix 网络判别器结构相同，采用 PatchGAN 判别器结构。

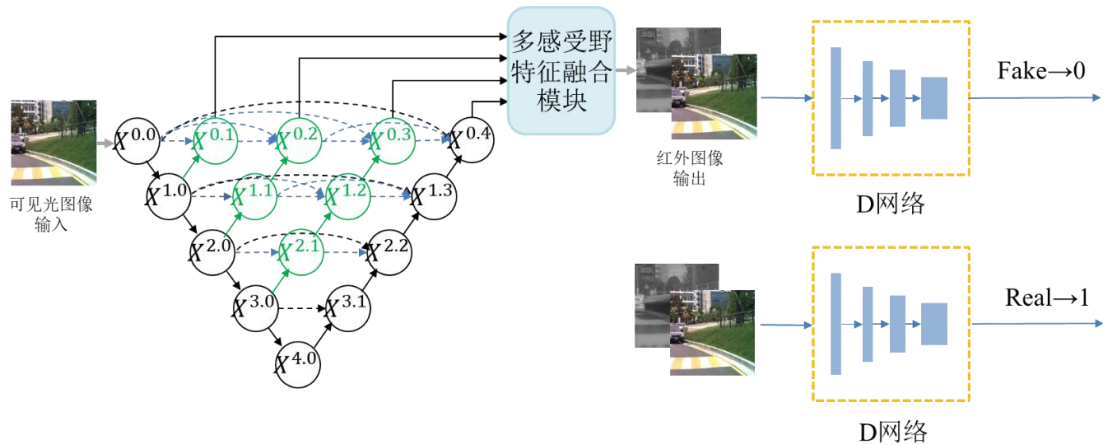


图 3-13 多感受野特征融合 Pix2pix 模型结构

3.3.2 多感受野特征融合 Pix2pix 模型构建分析

本文针对可见光图像生成红外图像的任务，在 Pix2pix 网络的基础上，提出了一种多感受野特征融合 Pix2pix 网络，首先是利用 Unet++ 网络提取多感受野特征，之后采用多感受野特征融合模块得到融合后的特征，生成对应的红外图像，最后通过生成器和判别器的博弈学习得到训练后的变换函数。判别器网络结构与 3.2.2 节介绍的判别器网络结构相同，所以本节重点介绍生成器网络结构。

本文采用 Unet++ 网络作为多感受野特征提取器，由于 Unet++ 原始结构提出的初衷是用于图像语义分割，相当于像素级的语义分类，但是本文异源红外图像生成在像素级上完成红外图像灰度预测的任务。Unet++ 网络应用到图像变换任务上在结构上与图像分割相同，主要区别在于：

- (1) 下采样方式：本文采用步长为 2 的卷积层代替原始 Unet++ 下采样层；
- (2) 多感受野特征处理方式：本文采用 Unet++ 网络作为多感受野特征提取器，采用多感受野特征融合模块来融合提取到的特征，代替原始 Unet++ 网络中对多条上采样路径输出取均值的方式；
- (3) 损失函数：本文采用生成对抗损失和 L1 损失函数训练网络，代替原始 Unet++ 网络的交叉熵损失函数。

生成器结构主要分为 3 部分：基于 Unet++ 网络的特征提取器、多感受野特征融合模块、输出卷积层。基于 Unet++ 网络搭建的基础网络为 3.2.4 节经过试验验证的 U-net 网络，如图 3-14 阴影框标注部分，图 3-14 无阴影部分为填充的跨层连接路径上的密集连接卷积层。

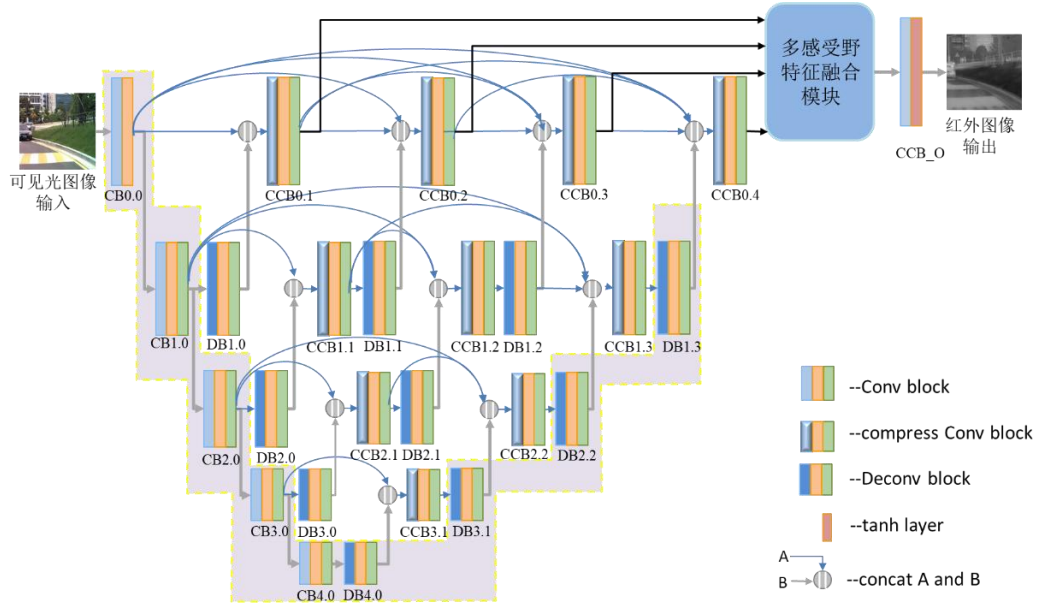


图 3-14 多感受野特征融合 Pix2pix 生成器网络结构

基于 Unet++ 网络的特征提取器。图 3-14 描述了基于 Unet++ 网络的特征提取器详细网络结构，除了阴影框标注的 U-net 网络外，Unet++ 网络的实心部分由成对的上采样模块和特征压缩模块组成。上采样模块保证了网络长连接的可行性，同时为了减小网络大小和加快推理过程，添加了特征压缩模块，减小特征通道数，网络中卷积层的详细参数将会在下文介绍。

多感受野特征融合模块。多感受野特征融合模块通过采用像素级的注意力机制融合不同感受野的特征，其中每 2 个特征采用一个融合块，逐层嵌套，其结构如图 3-15 所示，每个融合块参数不共享。给定四个相同尺寸的特征 f_1 、 f_2 、 f_3 、 f_4 ，我们首先采样四个卷积层对特征分别进行降维，之后采用两两嵌套的方式实现不同感受野特征的融合。本文对其中一个融合块进行详细说明：对于 f_1 、 f_2 ，首先分别进行步长为 2 的下采样，再进行通道上的合并，之后采用两个卷积核大小为 3 的卷积层 Conv12、Conv12_a 学习注意力图 Map；根据得到的注意力图 Map，进行 f_1 、 f_2 特征融合得到 $f_{12} = f_1 \times Map + f_2 \times (1 - Map)$ ；最后采用卷积核大小为 3 的卷积核 f_{12} 进一步优化得到 f_1 、 f_2 融合后的最终结果 f_{12} 。 f_1 、 f_2 融合后的特征 f_{12} 再与 f_3 进行特征融合，以此类推，逐级嵌套。

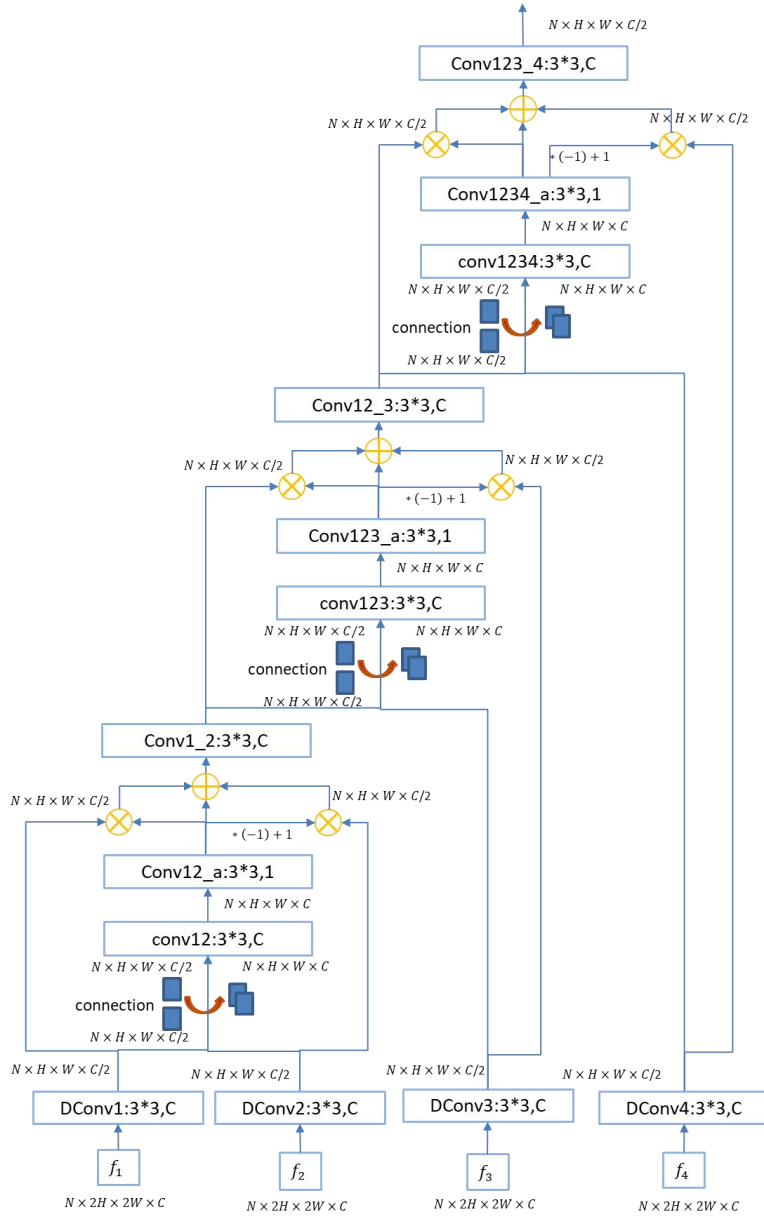


图 3-15 多感受野特征融合模块网络结构

输出卷积层。输出卷积层为卷积核大小为 3 的卷积层，输出 3 通道图像。

G 网络的参数见表 3-4，表中卷积层参数(f,k,s,p)分别表示卷积核数量、大小、步长、填充尺寸，输入参数和输出参数(b,c,h,w)分别表示训练批大小、通道数、特征图高、特征图宽，多感受野特征融合模块的卷积层参数在图中已经标注，故表格中不再进行详细描述。

华中科技大学硕士学位论文

表 3-4 多感受野特征融合 Pix2pix 生成器网络参数表

模块	卷积块名称	卷积层参数 (f,k,s,p)	输入特征 (b,c,h,w)	输出特征 (b,c,h,w)
输入	-	-	(1, 3,256,256)	-
Unet++特 征提取器	CB0.0	(64,4,1,2)	(1, 3,256,256)	(1,64,128,128)
	CCB0.1	(64,4,1,1)	(1,128,128,128)	(1,64,128,128)
	CCB0.2	(64,4,1,1)	(1,192,128,12,8)	(1,64,128,128)
	CCB0.3	(64,4,1,1)	(1,256,128,128)	(1,64,128,128)
	CCB0.4	(64,4,1,1)	(1,320,128,128)	(1,64,128,128)
	CB1.0	(128,4,1,2)	(1,64,128,128)	(1,128,64,64)
	CCB1.1	(128,4,1,1)	(1,256,64,64)	(1,128,64,64)
	CCB1.2	(128,4,1,1)	(1,384,64,64)	(1,128,64,64)
	CCB1.3	(128,4,1,1)	(1,512,64,64)	(1,128,64,64)
	DB1.0、1.1、1.2、1.3	(64,4,1,2)	(1,128,64,64)	(1,64,128,128)
	CB2.0	(256,4,1,2)	(1,128,64,64)	(1,256,32,32)
	CCB2.1	(256,4,1,1)	(1,512,32,32)	(1,256,32,32)
	CCB2.2	(256,4,1,1)	(1,768,32,32)	(1,256,32,32)
	DB2.0、2.1、2.2	(128,4,1,1)	(1,256,32,32)	(1,128,64,64)
	CB3.0	(512,4,1,2)	(1,256,32,32)	(1,512,16,16)
	CCB3.1	(512,4,1,1)	(1,1024,16,16)	(1,512,16,16)
	DB3.0、DB3.1	(512,4,1,2)	(1,512,16,16)	(1,256,32,32)
	CB4.0	(1024,4,1,2)	(1,512,16,16)	(1,1024,8,8)
	DB4.0	(512,4,1,2)	(1,1024,8,8)	(1,512,16,16)
输出	CCB_O	(3,4,1,1)	(1,32,32,32)	(1,3,256,256)

3.3.3 模型训练与实验结果分析

根据 3.3.2 节搭建的多感受野特征融合 Pix2pix 网络，采用与 3.2.4 节实验相同的

数据集开展实验，经过 60000 次迭代后网络收敛，相比 Pix2pix 网络 140000 步收敛，多感受野特征融合 Pix2pix 网络训练速度明显加快。生成器和判别器的损失函数分别如图 3-16、图 3-17 所示，其中生成器损失函数和 L1 损失函数中一直保持下降，训练过程比较稳定。判别器损失开始时由于多感受野特征融合 Pix2pix 网络的生成器结构相对判别器网络较复杂，相对判别器网络收敛较慢，开始时判别器网络损失很小，随着生成器网络拟合能力的逐渐增强，判别器网络损失慢慢变大，最终收敛到了理论平衡点 $\log(1-0.5) \approx 0.7$ 左右。

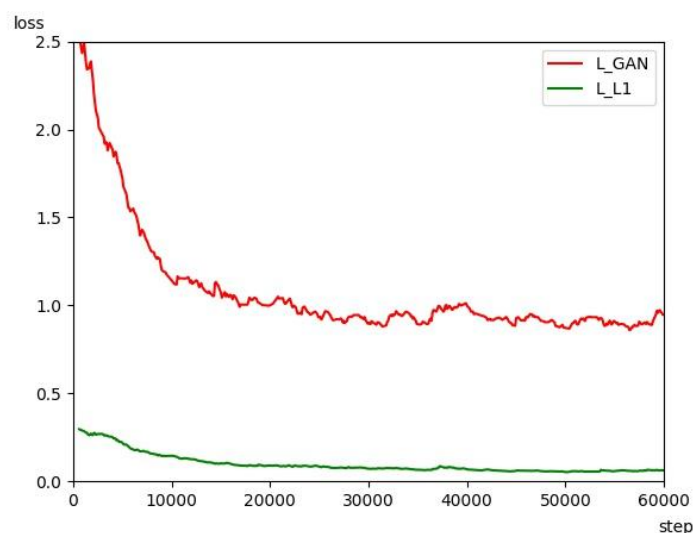


图 3-16 多感受野特征融合 Pix2pix 生成器损失曲线

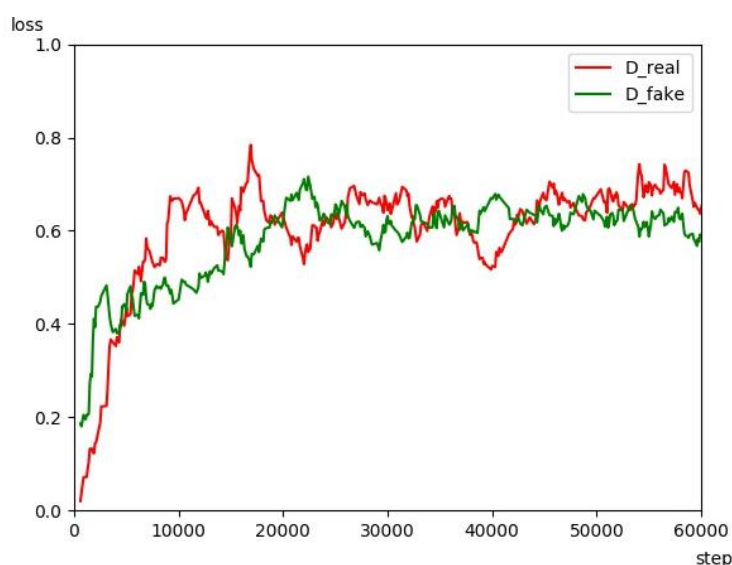


图 3-17 多感受野特征融合 Pix2pix 判别器损失曲线

采用训练好的网络对测试集中的数据进行测试，得到的结果如图 3-18 所示，其中第一列为可见光图像、第二列为可见光图像对应的真实红外图像、第三列为 Pix2pix 网络生成的红外图像、第四列为多感受野特征融合网络生成的红外图像。从图中我们可以看出，生成图像效果方面，多感受野特征融合网络实现了更准确的细节恢复

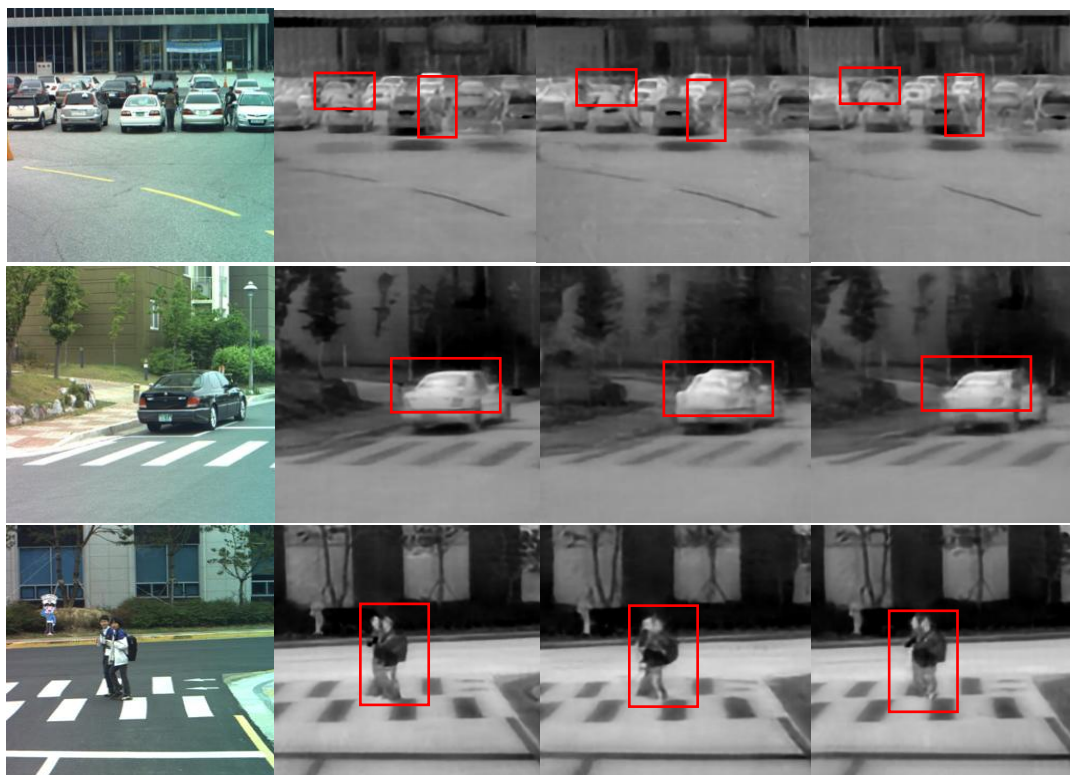


图 3-18 Pix2pix 与多感受野特征融合 Pix2pix 实验结果

本文采用 SSIM 和 PSNR 作为异源红外图像生成算法的评价指标，计算生成的红外图像相对于真实红外图像的结构相似度和失真度，SSIM 和 PSNR 的计算方式在 2.2.4 节有具体介绍。本文对 Pix2pix 网络实验结果和多感受野特征融合 Pix2pix 网络实验结果进行定量分析，得到的结果如表 3-5 所示，本文提出的算法有更高的性能指标。

表 3-5 Pix2pix 和的多感受野特征融合 Pix2pix 网络性能定量分析

模型	SSIM	PSNR
Pix2pix	0.793	26.646
多感受野特征融合 Pix2pix 网络	0.861	28.709

本文对多感受野特征融合模块学习到的不同感受野特征的注意力图进行了可视化，如图 3-19、图 3-20 所示，(a)、(b) 分别为可见光图像和对应的红外图像，(c)、(d)、(e)、(f) 分别为感受野从小到大的特征注意力图。从图中我们可以看出，感受野较小的特征对应的注意图更关注于图像的纹理、局部边缘，而感受野较大的特征对应的注意力图关注于图像较大的轮廓信息，与红外图像有较大的结构相似性。

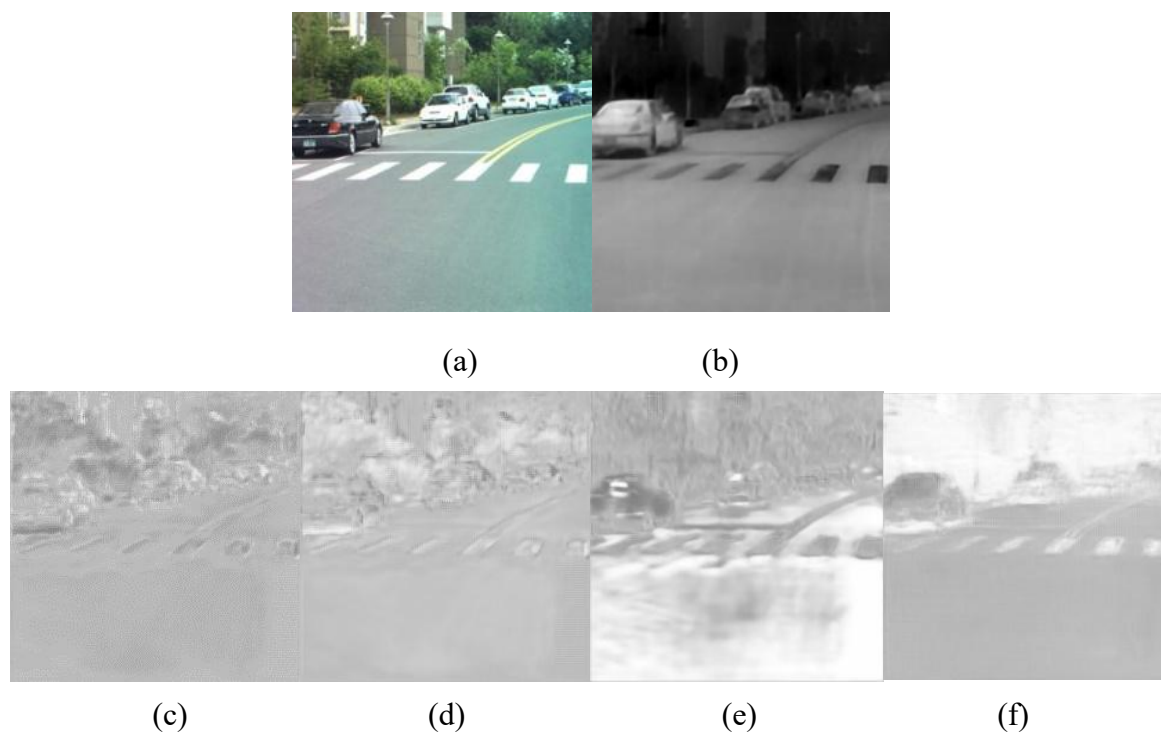


图 3-19 不同感受野特征融合注意力图 1

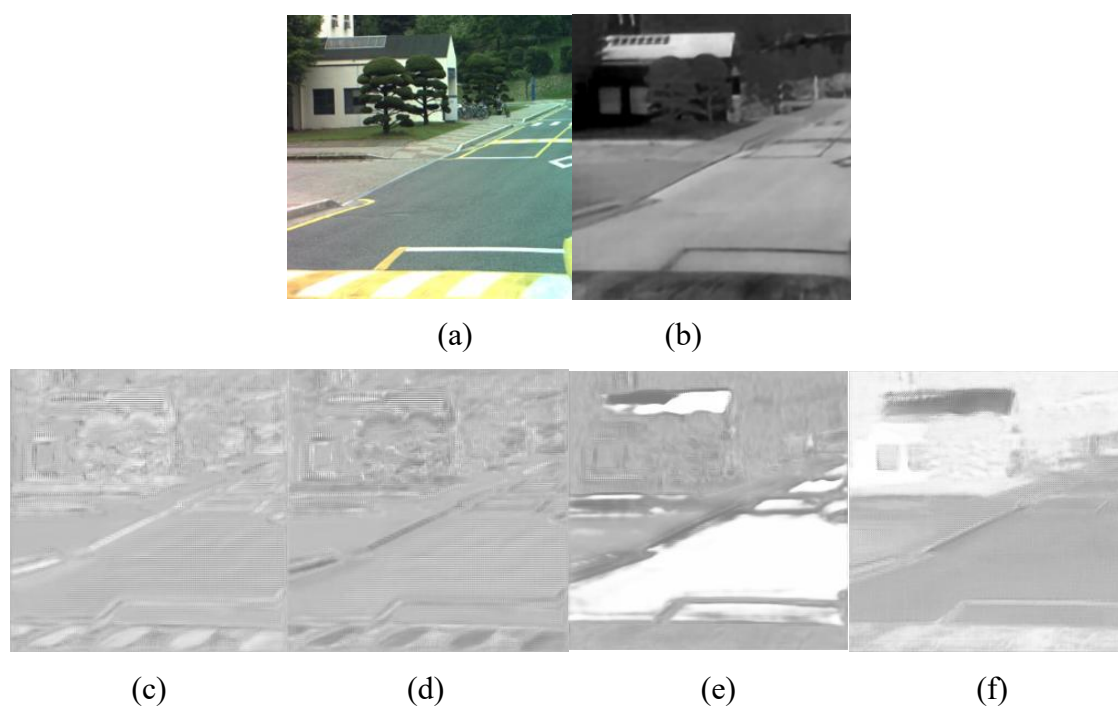


图 3-20 不同感受野特征融合注意力图 2

3.4 本章小结

本章针对异源红外图像生成任务展开研究,首先分析了使用 Pix2pix 网络完成红外图像生成任务的原因,并且根据学习任务对网络结构进行了调整。之后针对 Pix2pix 网络生成的红外图像局部信息缺失的问题,提出了多感受野特征融合 Pix2pix 网络,搭建基于 Unet++网络的多感受野特征提取,设计了一种多感受野特征融合机制,实验表明多感受野特征融合网络有更高的性能指标。

4 基于生成对抗网络的红外图像时段拓展

由第三章的研究可知,利用基于多感受野特征融合 Pix2pix 的异源红外图像生成算法能够实现可见光图像到某一时段的红外图像的转换。为了高效地完成大量不同时段的红外图像的生成,本章研究了基于生成对抗网络的红外图像时段拓展。由于对于时段拓展任务来说,很难获得多时段的配对红外图像,所以本章的研究基于一种能够进行非配对图像间相互转换的网络—StarGAN 网络。

4.1 基于 StarGAN 的红外图像时段拓展

4.1.1 StarGAN 网络

红外成像主要和温度、波长、传输介质有关,不同时段红外图像的变换关系不仅与温度有关,而且与波长和大气传输环节相关,式 4-1 和 4-2 分别表示了两个时段红外图像的成像因素:

$$I_{T_1} = f(T_1, \lambda_1, \varepsilon_1, \theta_1) \quad (4-1)$$

$$I_{T_2} = f(T_2, \lambda_2, \varepsilon_2, \theta_2) \quad (4-2)$$

I_{T_1} 表示时段 1 红外图像, I_{T_2} 表示时段 2 红外图像, λ_1 、 T_1 、 ε_1 表示时段 1 的影响因素, λ_2 、 T_2 、 ε_2 表示时段 2 的影响因素, λ 表示波长, T 表示温度, ε 表示大气传输影响, θ 表示其他影响因素。

本文选取三个时段红外图像分别代表了各自时段的红外辐射特性:早上 5 点红外图像、下午 2 点红外图像、晚上 7 点红外图像。红外图像数据集采用相同的成像设备集中拍摄,所以波长和大气环境大致不变。根据以上分析,如果能够获得映射函数满足式(4-3),那么就可以利用 F 来进行红外图像的时段拓展,所以时段拓展任务的可行性取决于能否拟合出时段 1 图像和时段 2 图像之间的映射函数。

$$I_{T_1} = F(I_{T_2}) \quad (4-3)$$

本文将红外图像时段拓展问题建模为不同时段红外图像域间的图像翻译问题。为了解决需要配对数据集的问题，已经有许多学者提出了针对非配对数据集的图像翻译模型：CycleGAN^[29]、DiscoGAN^[30]和 DualGAN^[31]采用对偶式的生成器结构，同时采用循环一致损失函数保留源图像域的关键特征，减少不合理的映射关系；UNIT^[32]和 MUNIT^[33]基于图像信息解码的思想，采用 VAE 与 GAN 网络的结合，将图像信息解码为内容特征和图像域特有的风格特征，实现了一对一和多对多的图像生成。但是上述模型一个生成器一次只能学习两种图像域间的映射关系，在进行多种图像域间的图像翻译时鲁棒性和扩展性非常有限。如图 4-1 (a) 所示，上述模型为了学习 k 个图像域之间的所有变换映射，必须训练 $k(k-1)$ 个生成器，同时每次训练仅采用两个图像域数据集，无法学习整个数据集的全局特征。为了提高算法效率，本文的算法实现了仅使用一个模型就可以对多个图像域进行相互转换。如图 4-1 (b) 所示，该框架在输入端添加目标域图像时段信息，使用一个生成器网络学习多个时段红外图像间的映射关系。这种统一模型架构允许在训练单个网络的同时得到多种图像域间的变换关系，能够将输入图像灵活转换为任何所需目标域图像。

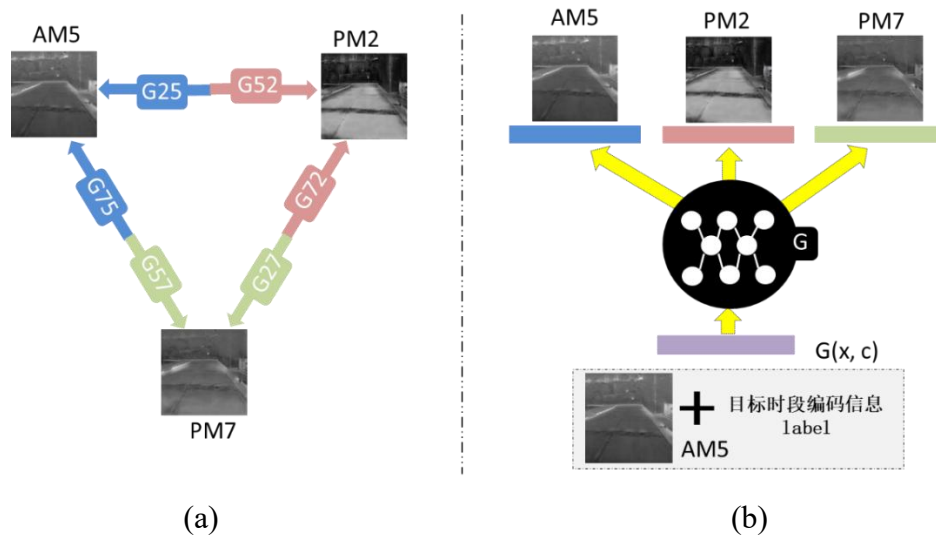


图 4-1 模型结构对比

StarGAN 网络^[34]是 Choi 等人提出的一种实现人脸属性编辑的网络，该网络通过对输入端添加目标人脸属性，实现了基于一个生成器的多属性的人脸图像编辑。StarGAN 模型的生成器没有学习固定的两个图像域间的变换关系，而是将图像和域

信息合并作为网络输入，将图像灵活地转换到对应的图像域。对于域信息的编码方式，StarGAN 使用独热编码来表示域信息。根据之前的分析，StarGAN 网络基本适应于红外时段拓展的任务，故本文采用 StarGAN 网络作为算法实现的基本框架。基于 StarGAN 网络的模型结构如图 4-2 所示，网络由生成器网络 G 和判别器网络 D 组成。生成器网络输入为时段编码信息（绿色系方块）和输入红外图像的合并，同时判别器网络添加了时段信息分类分支。

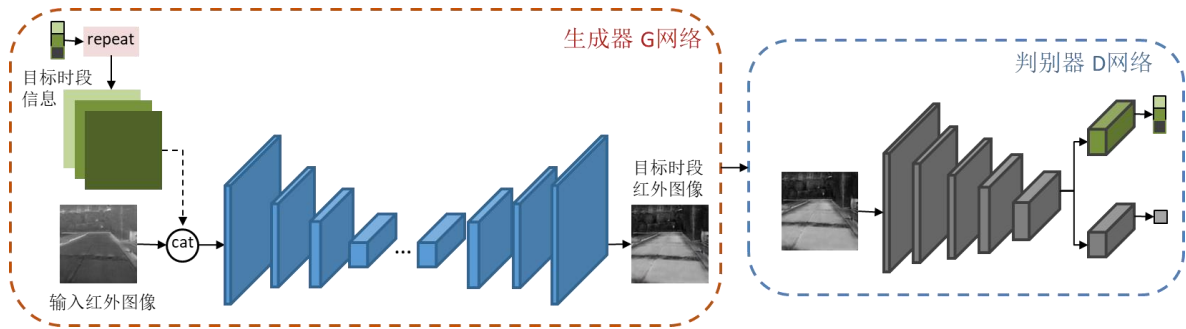


图 4-2 StarGAN 网络模型结构

在训练过程中，StarGAN 随机生成目标域时段标签，并训练模型将输入红外图像灵活地转换为目标时段图像域，从而在测试阶段，可以通过控制目标时段标签，将图像转换为任何所需的目标时段红外图像。基于 StarGAN 网络的红外时段拓展算法训练示意图如图 4-3 所示，图像顶部标注了时段信息及对应的编码方式：

(a) 训练 D 网络：将真实的红外图像 x 和生成器生成的红外图像 x' 分别输入 D 网络，D 网络要学会区分真实红外图像和假红外图像，并将真实红外图像分类到对应的时段域（100）。

(b) 训练 G 网络：源域到目标域的映射。随机生成目标域时段编码信息 $label_r$ （010），目标域标签 $label_r$ （010）在空间上复制并与输入图像 x 结合输入生成器网络生成对应时段的红外图像 x' 。

(c) 训练 G 网络：目标域到源域的映射。G 网络尝试在给定原始红外图像时段信息编码标签的情况下，从生成的图像中重建原始图像，将生成图像 x' 与原始图像时段编码信息 $label_s$ 结合输入 G 网络，生成重建后的原始红外图像 x'^{rec} 。

(d) 训练 G 网络：“欺骗”判别器，生成与真实红外图像无法区分的目标时段红外图像。G 网络生成的红外图像 x' 输入 D 网络，判别器判别图像真假，同时输出对应的时段编码信息，此时，生成器期望判别器判别结果为“真”，同时正确输出目标域时段编码 $label_r$ (010)。

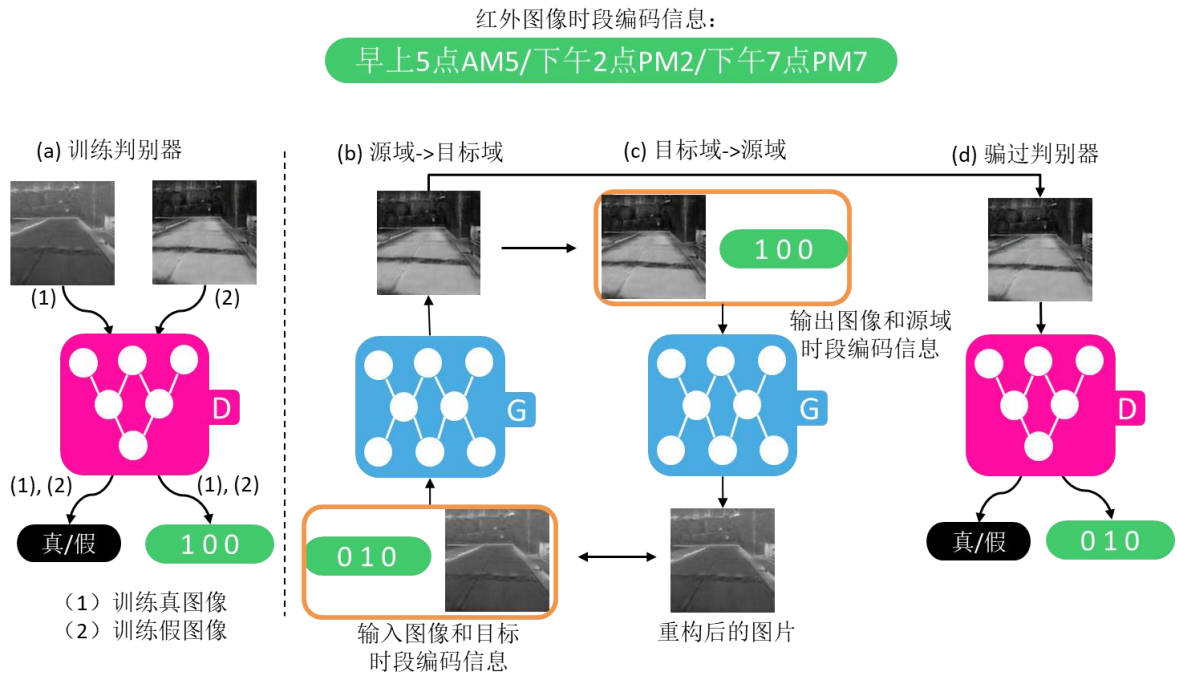


图 4-3 StarGAN 训练示意图

4.1.2 StarGAN 模型构建分析

生成器结构。本文搭建的生成器的网络结构如图 4-4 所示，图中黄色虚线部分为编码器，紫色虚线部分为解码器，黑色虚线为传输层。编码部分由 3 个下采样卷积块组成，降低特征层的分辨率，传输层由 6 个残差块组成，解码部分经过 3 个上采样卷积块恢复原始分辨率。为了实现多时段红外图像的生成，网络的输入端添加时段编码，原始时段编码方式为独热编码方式，将原始时段编码在第 3、4 维度延伸并且与输入图像在第 2 维度上连接输入生成器，通过变换原始时段编码信息，控制网络输出结果。图 4-4 为生成器网络结构，展示了早上 5 点的红外图像通过连接 2 种不同的时段编码信息得到的下午 2 点和晚上 7 点的红外图像。

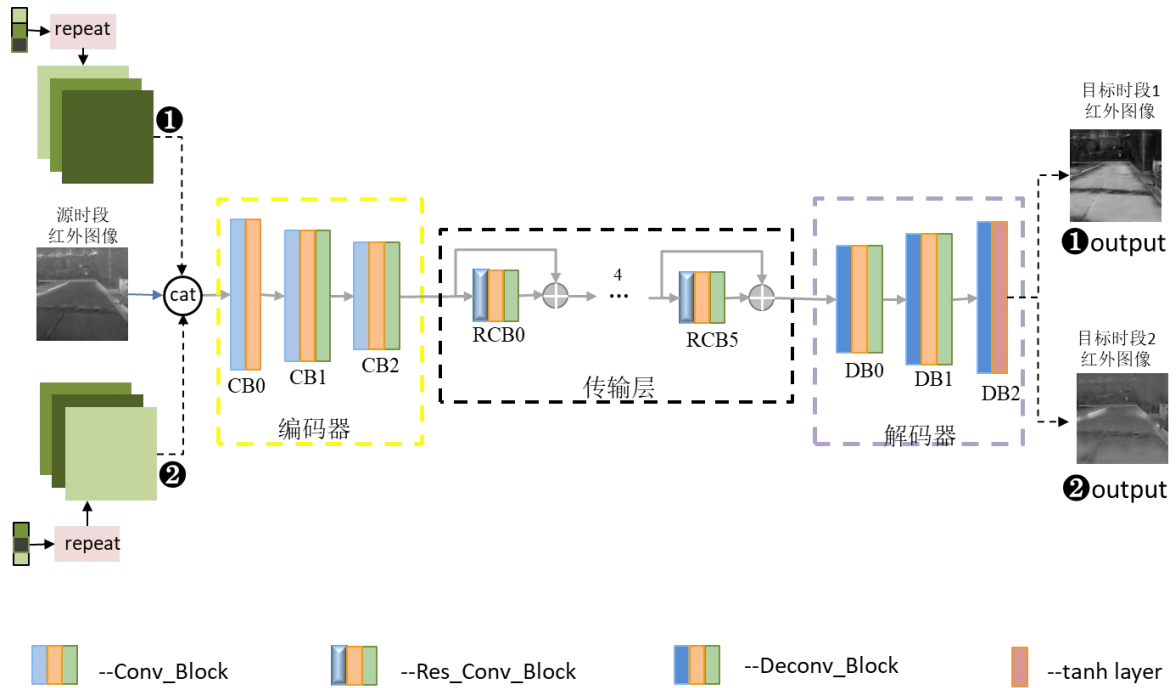


图 4-4 StarGAN 生成器网络结构

图 4-4 中下采样卷积块 Conv_Block、残差卷积块 Res_Conc_Block、上采样卷积 Deconv_Block 块的结构示意图分别如图 4-5、图 4-6、图 4-7 所示。对于归一化层的选择，生成器网络采用了在图像风格转换任务中实验效果较好的实例归一化层 Instance Normalization (IN 层)。BN 是目前卷积神经网络广泛采用的归一化层，通过对训练过程中每个批次的数据进行归一化，保证了数据分布的一致性，更适应于学习数据的整体分布。而在无配对数据集的图像翻译任务中，图像生成的结果主要依赖对应的图像实例，与同一批次中其他图像实例的数据分布关联较小。所以，本文采用 IN 层来对每个图像实例进行归一化处理。同时本文在传输层中使用了残差卷积神经网络的结构，将卷积块的输入与输出相加，有助于加快网络的学习速度，改善了反向传播时的梯度消失问题。

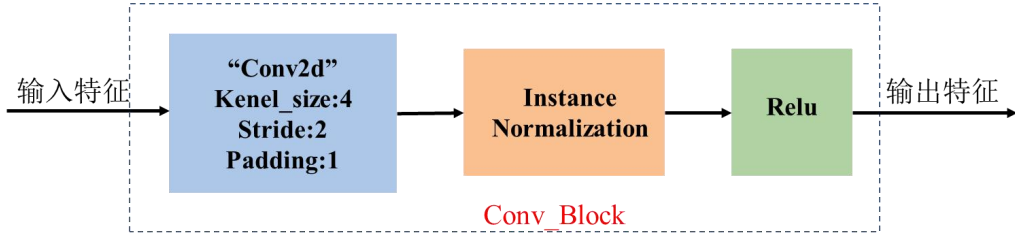


图 4-5 StarGAN 编码器子模块网络结构

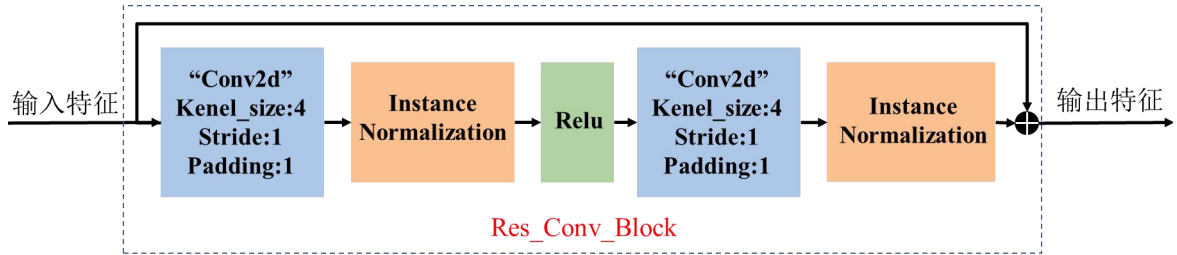


图 4-6 StarGAN 传输层子模块网络结构

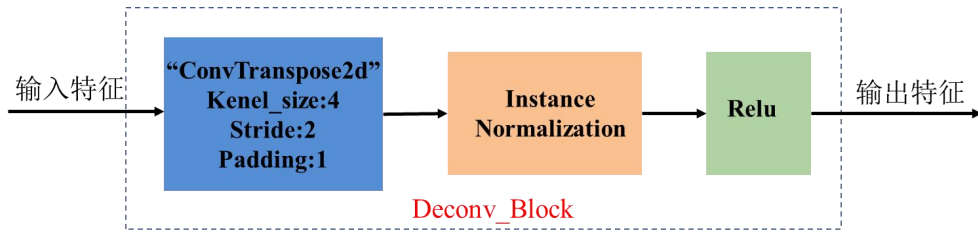


图 4-7 StarGAN 解码器子模块网络结构

G 网络的参数见表 4-1，表中卷积层参数(f,k,s,p)分别表示卷积核数量、大小、步长、填充尺寸，输入特征和输出特征参数(b,c,h,w)分别表示训练批大小、通道数、特征图高、特征图宽。由表可知，编码器通过卷积层设置步长为 2 来进行下采样，传输层每个子模块由两个卷积层组成，特征分辨率不变，并通过解码器的反卷积层逐步恢复图像的分辨率，保证了输出和输入图像的分辨率相同。

表 4-1 StarGAN 生成器网络参数表

模块	卷积块名称	卷积层参数 (f,k,s,p)	输入特征 (b,c,h,w)	输出特征 (b,c,h,w)
输入	-	-	(1,3,256,256)	(1,6,256, 56)
	-	-	(1,3,256,256)	
编码器	CB0	(64,7,3,1)	(1,6,256,256)	(1,64,256,256)
	CB1	(128,4,1,2)	(1,64,256,256)	(1,128,128,128)
	CB2	(256,4,1,2)	(1,128,128,128)	(1,256,64,64)
传输层	RCB0	(256,3,1,1)	(1,256,64,64)	(1,256,64,64)
		(256,3,1,1)	(1,256,64,64)	(1,256,64,64)
	RCB1	(256,3,1,1)	(1,256,64,64)	(1,256,64,64)
		(256,3,1,1)	(1,256,64,64)	(1,256,64,64)
	RCB2	(256,3,1,1)	(1,256,64,64)	(1,256,64,64)
		(256,3,1,1)	(1,256,64,64)	(1,256,64,64)
	RCB3	(256,3,1,1)	(1,256,64,64)	(1,256,64,64)
		(256,3,1,1)	(1,256,64,64)	(1,256,64,64)
	RCB4	(256,3,1,1)	(1,256,64,64)	(1,256,64,64)
		(256,3,1,1)	(1,256,64,64)	(1,256,64,64)
	RCB5	(256,3,1,1)	(1,256,64,64)	(1,256,64,64)
		(256,3,1,1)	(1,256,64,64)	(1,256,64,64)
解码器	DB0	(128,4,1,2)	(1,256,64,64)	(1,128,128,128)
	DB1	(64,4,1,2)	(1,128,128,128)	(1,64,256,256)
	DB2	(3,7,1,1)	(1,64,256,256)	(1,3,256,256)

判别器结构。对于一个给定的输入图像 x 和一个目标时段标签 c ，时段拓展的目的是将 x 转化为具有目标时段 c 图像域特征的输出图像 y 。为了实现生成的图像满足

对应时段的图像域条件，D 网络添加了一个辅助分类器，对判别器输入图像进行时段信息分类，并且在训练 G 网络和 D 网络时增加时段分类损失函数，具体实现细节见 4.1.3 节。同时，D 网络保留用来辨别输入的“真假”的输出路径，本文搭建的 D 网络的结构如图 4-8 所示，ConvBlock 子模块如图 4-9 所示：

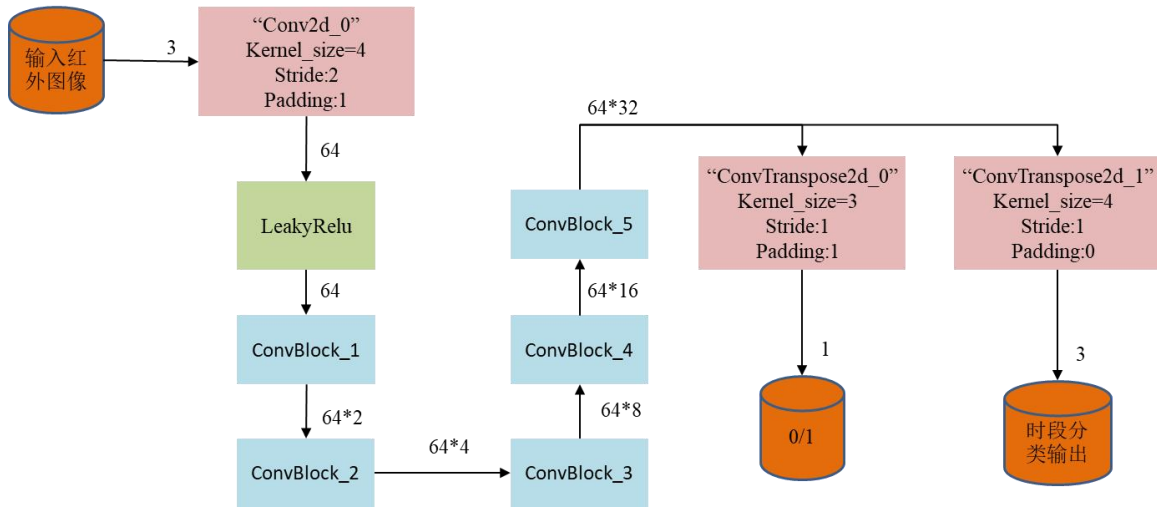


图 4-8 StarGAN 判别器网络结构

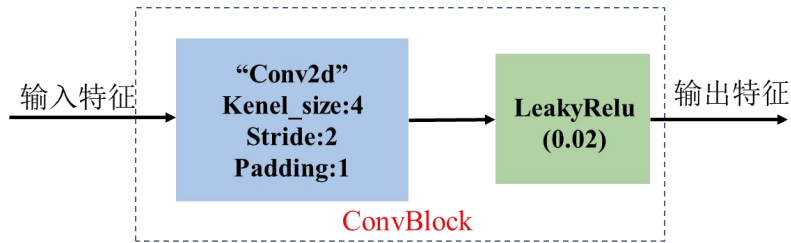


图 4-9 ConvBlock 子模块网络结构

D 网络的参数见表 4-2，表中卷积层参数(f,k,s,p)分别表示卷积核数量、大小、步长、填充尺寸，输入特征和输出特征参数(b,c,h,w)分别表示训练批大小、通道数、特征图高、特征图宽，输出模块包含两个分支，时段信息分类分支和真假样本判断分支。

表 4-2 StarGAN 判别器网络参数表

模块	卷积块名称	卷积层参数 (f,k,s,p)	输入特征 (b,c,h,w)	输出特征 (b,c,h,w)
输入	-	-	(1, 3,256,256)	-
下采样	Conv2d_0	(64,4,1,2)	(1, 3,256,256)	(1,64,128,128)
	ConvBlock_1	(128,4,1,2)	(1,64,128,128)	(1,128,64,64)
	ConvBlock_2	(256,4,1,2)	(1,128,64,64)	(1,256,32,32)
	ConvBlock_3	(512,4,1,2)	(1,256,32,32)	(1,512,16,16)
	ConvBlock_4	(1024,4,1,2)	(1,512,16,16)	(1,1024,8,8)
	ConvBlock_5	(2048,4,1,2)	(1,1024,8,8)	(1,2048,4,4)
输出模块	Contranpose2d_0	(1,3,1,1)	(1,2048,4,4)	(1,1,4,4)
	Contranpose2d_1	(3,4,0,1)	(1,2048,4,4)	(1,3,1,1)

4.1.3 StarGAN 损失函数

生成对抗网络是强大的生成模型，但存在训练不稳定、梯度消失、梯度爆炸的问题，经过分析，主要原因是在最优判别器的优化条件下，生成器的最终优化目标为最小化真实样本分布和生成样本分布间的 JS 距离。JS 距离可以反映两个数据分布的相似度，但是当两个数据分布的支撑面完全不重叠时，JS 距离得到的距离测度将失效，而在图像生成任务中高维数据间的相互映射过程中， P_1 、 P_2 分布不重叠或者重叠测度几乎为 0 的可能性很大。

Wasserstein 距离的提出一定程度上解决了原始 GAN 训练不稳定的问题，其优越性在于即使两个数据分布间没有重叠，仍然可以反映两个分布间的距离远近。在 WGAN 中^[50]，作者采用 Wasserstein 距离作为生成对抗损失函数，计算公式如下：

$$W(P_1, P_2) = \frac{1}{k} \sup_{\|f\|_L \leq k} E_{x \sim P_1}[f(x)] - E_{x \sim P_2}[f(x)] \quad (4-4)$$

其中 P_1 、 P_2 分别为真实数据和生成数据分布，函数 f 必须满足 Lipschitz 连续，函数 f 导数的绝对值小于 k 。WGAN 采用权重裁剪对判别器施加 Lipschitz 约束，限制判别器的所有参数 $w_i \in [-0.01, 0.01]$ 。但是在实验过程中，WGAN 仍然存在收敛速度慢、调参困难等问题，原因在于对判别器强制进行参数截断，在拉大真假样本的判别结果的训练过程中，判别器参数更倾向于取最大或最小值，减弱了判别器的拟合能力。同时，对于参数截断的阈值选择，过高或过低会导致梯度爆炸或梯度消失，加大调参的难度。

WGAN-GP^[51]采用梯度惩罚的策略代替参数截断，通过设置额外的损失函数 $(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2$ 来满足 Lipschitz 限制，将梯度惩罚损失项与原始 WGAN 损失函数合并，计算公式如下：

$$W(P_1, P_2) = E_{x \sim P_1}[f(x)] - E_{x \sim P_2}[f(x)] - \lambda E_{\hat{x} \sim P_{\hat{x}}}[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2] \quad (4-5)$$

$$\hat{x} = \varepsilon x_1 + (1 - \varepsilon)x_2, \quad x_1 \in P_1, \quad x_2 \in P_2, \quad \varepsilon \in \text{Uniform}[0, 1] \quad (4-6)$$

其中 $\hat{x}$ 表示真实样本与假样本间的随机插值采样。WGAN-GP 采用额外的梯度惩罚损失项代替对判别器网络的参数截断，避免了训练困难和梯度消失、梯度爆炸的问题，同时实验表明 WGAN-GP 保证了模型的训练过程的稳定性和收敛速度。所以本文采用 WGAN-GP 损失函数作为非配对数据形式图像生成的目标函数。根据 4.1.1 节训练过程的分析，红外图像时段拓展的损失函数如下所示：

对抗损失。根据上述分析，本文选用 WGAN-GP 对抗损失 L_{adv} ，G 网络的目标为最小化 L_{adv} ，D 网络的目标为最大化 L_{adv} ，对抗损失的定义如下：

$$L_{adv} = E_x[D_{src}(x)] - E_{x,c}[D_{src}(G(x,c))] - \lambda_{gp} E_{\hat{x}}[(\|\nabla_{\hat{x}} D_{src}(\hat{x})\|_2 - 1)^2] \quad (4-7)$$

其中 D_{src} 为判别器真假样本判断分支， λ_{gp} 为梯度惩罚损失项超参数。

时段分类损失。对于给定的输入红外图像 x 和目标域时段标签 c ，时段拓展是将 x 转换为具有目标域时段特征的输出图像 y 。为了达到这一条件，D 网络添加了一

个辅助分类器，并在优化 D 网络和 G 网络时添加了时段分类损失。也就是说，我们将时段分类损失函数分解为两个损失项：用于优化 D 网络的真实红外图像的时段分类损失，如式 4-8 所示；以及用于优化 G 网络的拓展后红外图像的时段分类损失，如式 4-9 所示：

$$L'_{cls} = E_{x,c'}[-c' \log(D_{cls}(x))] \quad (4-8)$$

$$L^f_{cls} = E_{x,c}[-c \log(D_{cls}(G(x,c)))] \quad (4-9)$$

其中 c 为目标域时段标签， c' 为原始红外图像时段标签， D_{cls} 为判别器时段信息分类分支。

重建损失。通过最小化对抗损失和时段分类损失，训练 G 网络生成逼真的红外图像并将其分类到正确的目标时段域。但是，将生成对抗损失最小化并不能保证时段拓展后的图像保留输入图像的场景内容信息。为了解决这个问题，StarGAN 网络将循环一致性损失应用于生成器的优化过程，将生成器生成的目标时段红外图像与原始红外图像的时段信息编码相连，输入生成器网络，重新映射到原始时段图像域。循环一致损失函数要求重新生成的原始时段红外图像与真实红外图像无限接近，其定义如下：

$$L_{rec} = E_{x,c,c'}[\|x - G(G(x,c),c')\|_1] \quad (4-10)$$

综上所述，生成器和判别器的优化损失函数分别如式 4-11、4-12 所示：

$$L_D = -L_{adv} + \lambda_{cls} L'_{cls} \quad (4-11)$$

$$L_G = L_{adv} + \lambda_{cls} L^f_{cls} + \lambda_{rec} L_{rec} \quad (4-12)$$

其中 λ_{cls} 和 λ_{rec} 是分别控制时段分类损失和重建损失相对于对抗损失的相对重要性的超参数。

综上所述，基于 StarGAN 网络的红外图像时段拓展的损失函数示意图如图 4-10 所示：

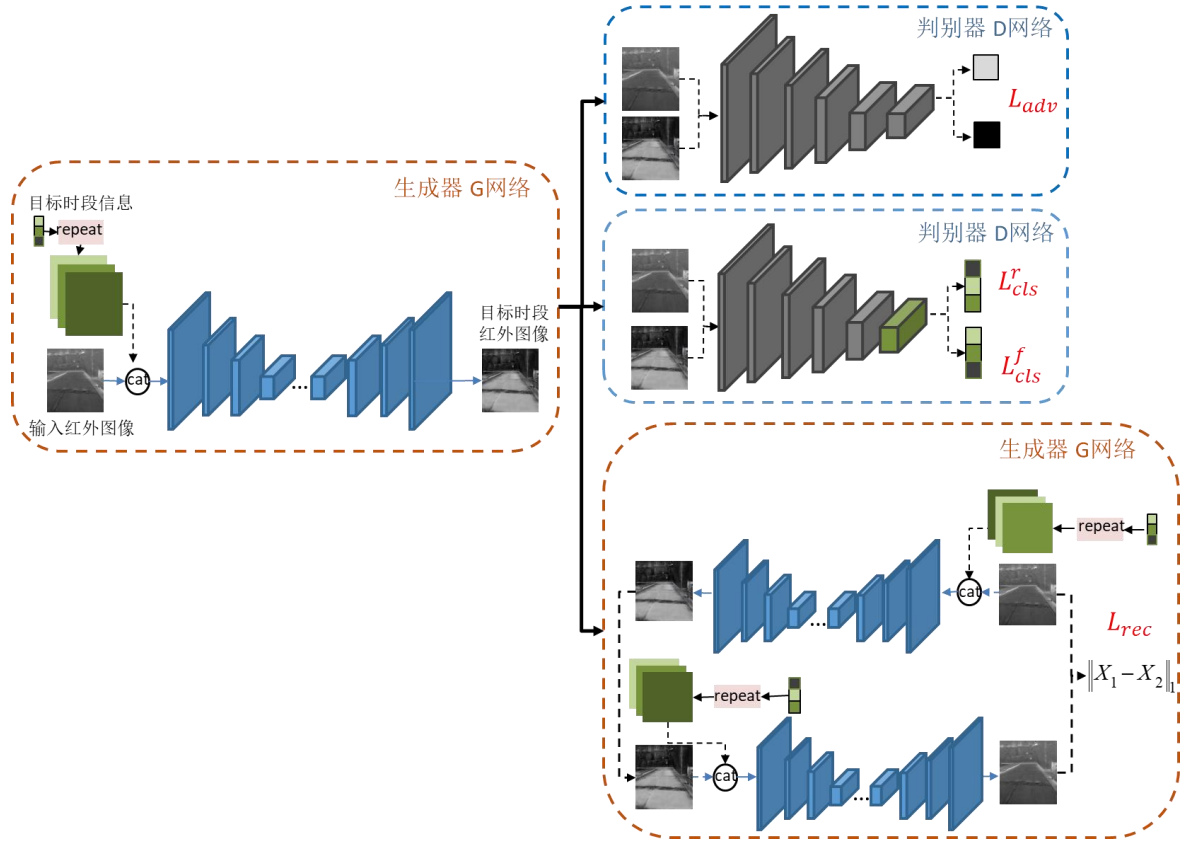


图 4-10 StarGAN 网络损失函数示意图

4.1.4 模型训练与实验结果分析

(1) 数据集介绍

实验中采用的数据集为 Kaist 实验室发布的车载道路场景数据集。数据集中红外图像为 FLIR 红外摄像头拍摄，波段范围为 7.5~13um。本文选定了 3 个时段红外图像展开研究分别为早上 5 点、下午 2 点、晚上 7 点。每个时间段红外图像训练数据集数量为 2000 张，测试数据集红外图像数量为 200 张，同时每个时段红外图像数据不是一一配对的，但是数据集中的场景相似，主要包括道路、树木、草地、车辆等景物及少量的行人干扰。

(2) 训练细节及参数设置

对于数据集的处理，本文首先将全部图像数据进行切割、调整尺寸的处理使得图像大小为 (256, 256)，之后对每张红外图像对应的时段信息标号，并将标号与红外

图像相连作为一个数据样本{0: 红外图像, 1: 标号}, 将所有时段红外图像进行相同的处理, 最终得到 6000 个训练数据样本。训练过程中, 算法将红外图像对应的标号进行独热编码作为时段编码信息, 将独热编码信息复制扩展到 (256, 256), 扩展后的编码结果与对应的红外图像在第二维度相连输入生成器网络。

对于训练参数, 本文将 `batchsize` 设置为 8, 总共进行 20000 次迭代, 前 10000 次迭代学习率保持 0.0001 不变, 之后每 1000 次迭代学习率下降 0.0001/10 直到学习率为 0。本文采用 Adam 优化器, β_1 、 β_2 分别设置为 0.5、0.999。

(3) 训练过程及实验结果展示与分析

根据红外时段数据集和设置的损失函数对网络进行训练, 经过 100 个 epoch 之后, 训练过程中的生成器和判别器的变化趋势如图 4-11、图 4-12 所示。总体来看, 循环一致损失保持下降趋势, 并且送 17500 步之后逐渐收敛; 判别器网络震荡趋势逐渐减小, 同时真样本判别损失和假样本判别损失相当, 没有出现模式崩塌。

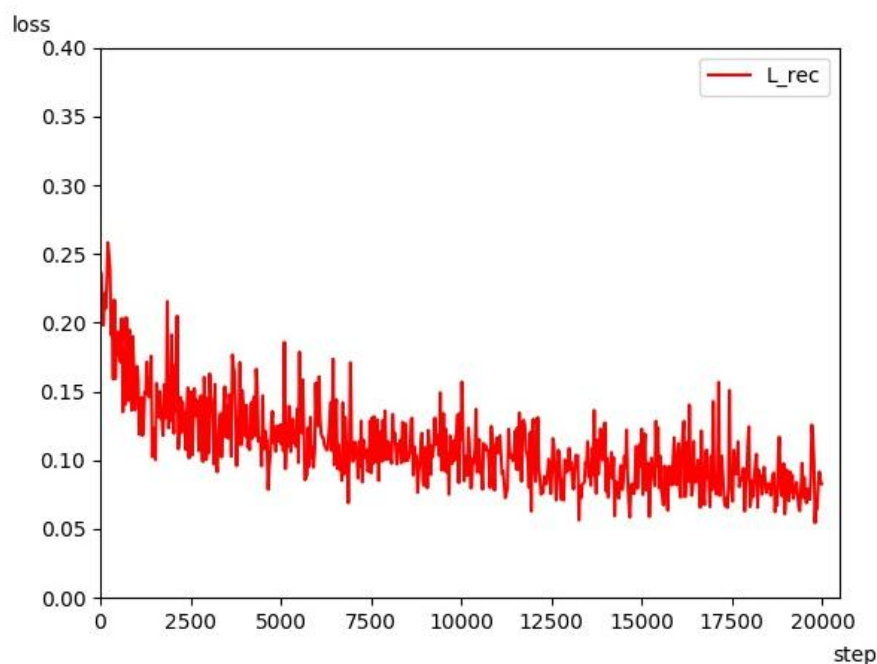


图 4-11 StarGAN 生成器循环一致损失曲线

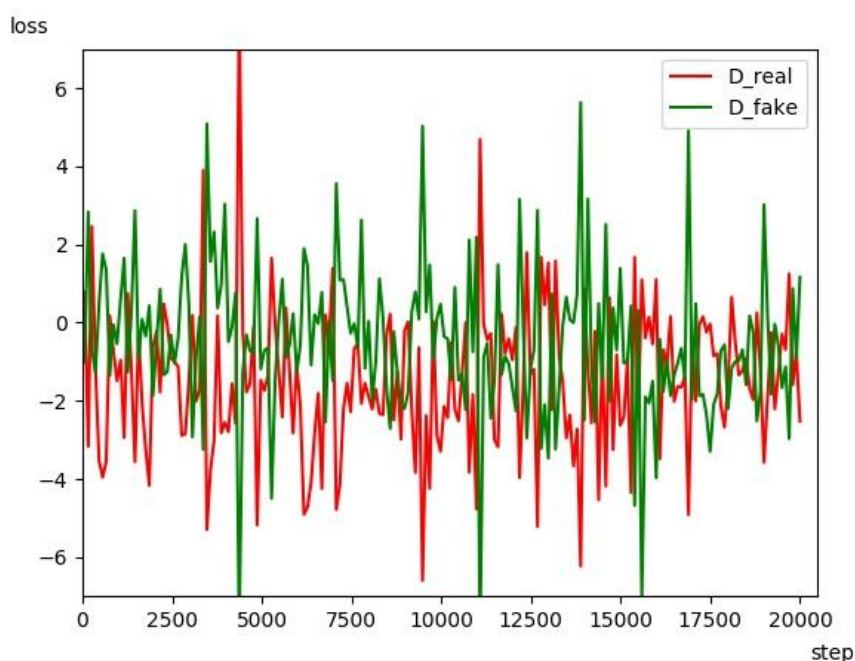
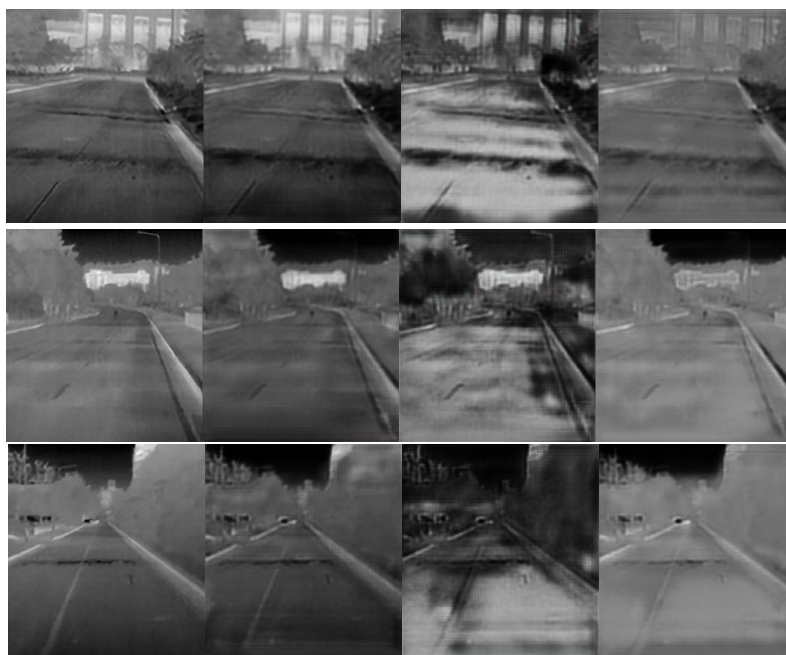


图 4-12 StarGAN 判别器损失曲线

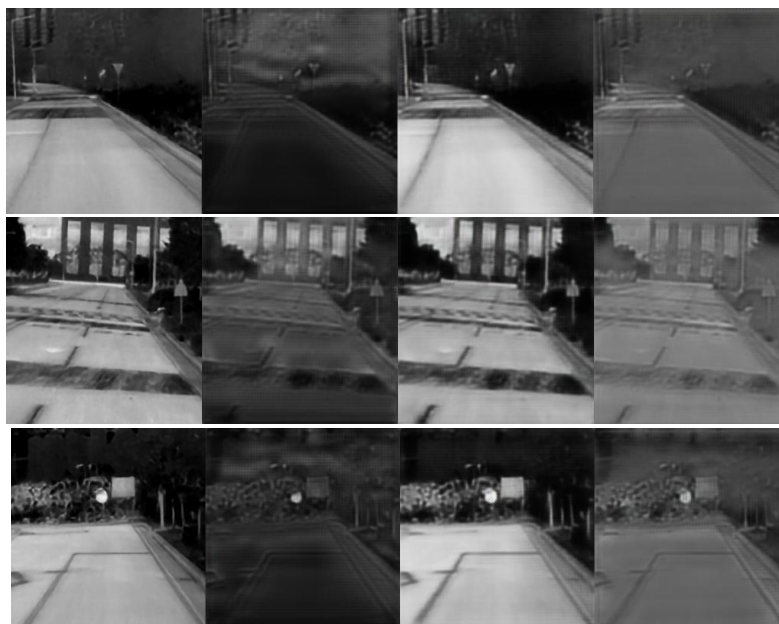
利用训练完成的生成器网络进行测试集数据的时段拓展,本节首先采用主观判断的方法衡量拓展结果的图像真实感和时段信息的有效性,图 4-13 展示了早上 5 点的红外图像拓展到下午 2 点和晚上 7 点的实验结果。其中第一列为输入的早上 5 点的红外图像,第二、三、四列分别为经过时段拓展得到的早上 5 点、下午 2 点、晚上 7 点红外图像。图中第一列的 3 个场景均为道路、建筑物、树木构成的实拍红外图像,可以看到早上 5 点、晚上 7 点拓展结果较好,每种场景语义范围内均没有出现亮度突变,但是下午 2 点的拓展结果中出现不同程度的失真。我们可以观察道路这一语义场景部分,第一行拓展结果较好,但是第二、三行较远处的道路信息拓展失败,仍保持早上 5 点实拍红外图像的灰度特性,亮度较暗。分析原因在于早上 5 点实拍红外图像红外成像效果总体上灰度均值较小、对比度低,晚上 7 时段红外图像总体上亮度变大,成像效果与早上 5 点相对较近,所以拓展结果较好。而下午 2 点红外图像相比早上 5 点、晚上 7 点时段,红外图像整体上不仅亮度变大,同时图像不同语义场景灰度信息变化明显,对比度大,导致生成效果较差。



实拍早上 5 点 生成早上 5 点 生成下午 2 点 生成晚上 7 点

图 4-13 早上 5 点红外图像时段拓展结果

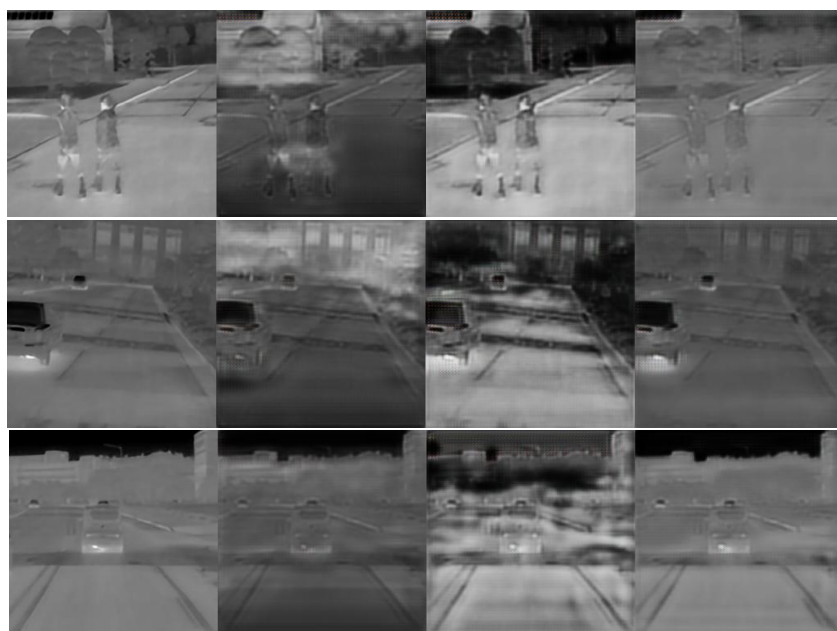
图 4-14 为下午 2 点红外图像拓展结果，相比早上 5 点，成像效果明显提升，分析原因主要在于下午 2 点时段图像对比度大，不同材质间的轮廓明显。



实拍下午 2 点 生成早上 5 点 生成下午 2 点 生成晚上 7 点

图 4-14 下午 2 点红外图像时段拓展结果

图 4-15 为晚上 7 点红外图像时段拓展结果，同样存在由于对比度较弱导致的拓展后图像质量较差的问题。同时我们可以观察到，晚上 7 点到早上 5 点拓展结果在树木这一语义场景出现了拓展错误的现象，树木由于比热容小，红外辐射变化不大，同时相对于道路场景发射率较小，所以整体上灰度会较暗，而图中第一行和第二行拓展后的早上 5 点红外图像树木部分灰度明显变亮。



实拍晚上 7 点

生成早上 5 点

生成下午 2 点

生成晚上 7 点

图 4-15 晚上 7 点红外图像时段拓展结果

根据上述分析，可以看到 StarGAN 网络初步实现了时段拓展的任务，在一些场景情况下实现了较好的拓展结果。但是也出现了拓展后图像质量较差、红外纹理时段拓展错误的现象。考虑原因一部分在于红外图像灰度较暗、对比度较低的成像特性导致，另一部分由于本文采用的数据集存在较大的场景信息变化，导致 StarGAN 网络在一部分生成结果中出现了比较明显的场景语义信息生成错误的情况。为了保留原图像场景信息的同时保证网络有较强的泛化能力，本文提出了语义约束 StarGAN 网络。

4.2 基于语义约束 StarGAN 的红外图像时段拓展

4.2.1 语义约束 StarGAN 网络介绍

在红外图像拓展的任务中，算法一方面要保证生成的红外图像保持原红外图像场景语义信息，另一方面实现生成的红外图像符合对应时段的红外纹理分布特征。本节提出的语义约束 StarGAN 为满足语义信息不变性和时段信息有效性进行了以下两点改进：

(1) 语义信息不变性：在原始 StarGAN 网络的结构上，添加语义信息编码分支，同时在训练生成器的过程中添加语义编码一致损失函数，使得生成后的红外图像与原始红外图像的语义信息趋向一致；

(2) 时段信息有效性：不同语义场景的红外辐射量时变趋势不同，比热容较大的场景昼夜温差大，辐射量变化大，比热容较小的场景温度变化平缓，辐射量变化小，基于此，本节提出了一种基于语义分割图和红外辐射时变曲线的时段信息编码方式。

基于语义约束 StarGAN 的红外图像时段拓展模型结构如图 4-16 所示。生成器网络的编码器由时段信息编码分支、语义编码分支和图像特征编码分支组成，将时段信息编码、语义编码分支最后一层特征、图像特征编码分支编码后特征相连输入解码器，生成对应时段编码信息的红外图像。判别器网络与 StarGAN 判别器网络相同。

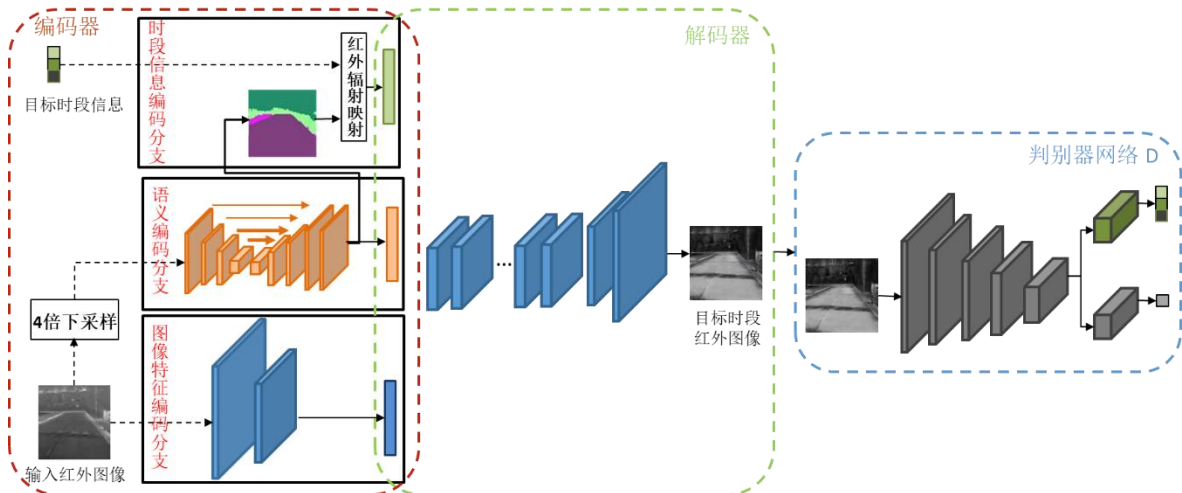


图 4-16 语义约束 StarGAN 模型结构

G 网络的编码器在 StarGAN 网络的基础上，添加了时段编码分支和语义编码分支。时段信息编码分支根据语义编码分支得到的语义分割结果和红外辐射时变曲线，计算时段编码，其实现方式将在 4.2.2 节介绍。语义编码分支为一个基于 U-net 的语义分割网络，对输入红外图像进行像素级的分类。在网络训练的过程中，算法对语义分割分支采用有标签的监督式的学习方式，如果语义分割分支可以准确地反映语义信息，则该语义分割网络的倒数第二层的特征层便可以捕捉该图像的语义信息。基于此，我们将分割网络的倒数第二层特征与编码器输出相连作为时段无关特征，一方面保留原红外图像特征信息，另一方面约束网络保证拓展后语义信息的不变性。同时，算法在训练中添加了语义编码一致性损失，约束拓展后的红外图像与原时段红外图像有一致的语义信息，具体实现方式将在 4.2.3 节进行介绍。

以上便是语义约束 StarGAN 的具体实现细节。基于语义约束 StarGAN 网络的红外时段拓展的算法流程图如下：

(a) 训练 D 网络：将真实的红外图像 x 和生成器生成的红外图像 x' 分别输入 D 网络，D 网络要学会区分真实红外图像和假红外图像，并将红外图像分类到对应的时段域。

(b) 训练 G 网络：源域到目标域的映射。将红外图像 x 输入语义编码分支和图像特征编码分支，随机生成目标域时段标号 r ，根据时段标号 r 与语义编码分支分割结果，计算时段编码 $label_r$ ；将图像特征编码分支输出、语义编码分支倒数第一层特征层与时段编码 $label_r$ 相连输入解码器，生成对应时段的红外图像 x' 。

(c) 训练 G 网络：目标域到源域的映射。G 网络尝试在给定原始时段信息编码标签的情况下从生成的图像中重建原始时段红外图像，将生成图像 x' 输入语义编码分支和图像特征编码分支，根据原始红外图像时段标号 s 与语义编码分支分割结果，计算时段编码 $label_s$ ；将图像特征编码分支输出、语义编码分支倒数第一层特征层与时段编码 $label_s$ 相连输入解码器，生成重建后的原始红外图像 x^{rec} 。

(d) 训练 D 网络：“欺骗”判别器。G 网络生成的红外图像 x' 输入 D 网络，判别器判别图像真假，同时输出对应的时段编码信息，此时，生成器期望判别器判别结果为“真”同时正确分类到对应的时段域。

4.2.2 语义约束 StarGAN 时段编码方式

在 StarGAN 网络中，全幅红外图像不同材质间的时段编码信息相同。然而红外辐射与材质特性密切相关，不同材质红外辐射分布符合对应的日变化特性。基于此，结合不同材质红外辐射的日变化特性的先验知识，本文提出了一种基于语义分割图和红外辐射时变曲线的时段编码方式，

不同材质辐射变化量与温度变化密切相关，我们可以构建不同材质红外辐射日变化数据库，作为红外图像时段拓展的先验知识。为了统计不同材质红外辐射时变曲线，本文采用 SE-WORKBENCH 仿真软件对几种材质的红外辐射量进行统计。SE-WORKBENCH 仿真软件是一种被广泛采用的红外仿真软件，通过对场景进行材质分割结合计算好的大气参数，能够仿真任意场景在不同波段不同时段红外图像。

在本文采用的数据集中，数据集中的场景相似，主要包括道路、树木、草地、车辆等景物及少量的行人干扰。本文利用 SE-WORKBENCH 软件，对数据集中大量出现的天空、草地、地面、树四种场景的辐射量进行红外辐射量时变数据统计，如表 4-3 所示。由表可知，天空辐射变化量在全天时段变化趋于平缓；草地和树木辐射量变化趋势和幅度几乎一致，由于植被的比热容较小，温度变化范围小，红外辐射变化范围小；地面比热容较大，红外辐射变化范围相对较大。

表 4-3 中波不同材质时段辐射量变化统计（单位： $W/(m^2 \cdot sr)$ ）

时间	天空	树	草地	地面
05 点	1.48125	1.53412	1.52892	1.57349
14 点	1.55423	2.39011	2.53156	3.28729
19 点	1.46777	1.57837	1.59181	1.63271

对于红外图像的材质分割,本文利用 4.2.1 节的语义分割分支输出的语义分割结果,作为对应的材质分割结果。该等效关系成立的前提为,本文仅对天空、草地、地面、树四种场景进行材质到红外辐射量的映射,这四个场景内材质统一,不存在歧义信息。根据构建的不同材质红外辐射日变化数据库和语义分割结果,本文提出了基于语义分割图和红外纹理时变曲线的时段信息编码方式,实现方式如下:

(1) 首先将统计到的中波红外辐射变化量进行归一化,归一化后结果如表 4-4 所示;

(2) 根据原红外图像语义分割后语义图,将场景的语义编码信息与目标时段信息结合在表 4-4 找到对应归一化后的辐射值,替换语义图上的像素值;

(3) 对语义图上的每个像素点重复上述操作,对于除了天空、草地、地面、树之外的语义编码对应的像素点,设置为 0,得到材质相关时段编码信息;

(4) 为了保证除天空、草地、地面、树之外材质的时段拓展可行性,将材质相关时段编码信息与 4.1.1 节提出的时段编码信息进行通道上的合并,作为语义约束 StarGAN 的时段编码。

表 4-4 不同材质时段编码信息映射表

时间	天空	树	草地	地面
05 点	0.007	0.036	0.034	0.058
14 点	0.047	0.507	0.585	1
19 点	0	0.061	0.068	0.091

4.2.3 语义约束 StarGAN 模型构建分析

本文针对红外图像拓展的任务,在 StarGAN 网络的基础上,提出了语义约束 StarGAN,增加了语义编码分支,将语义编码分支的倒数第一层的特征作为红外图像的语义编码,同时提出了一种基于不同材质红外辐射时变先验的时段编码方式。判别器网络结构与 4.1.2 节介绍的判别器网络结构相同,本小节重点介绍生成器网络结

构。

本文搭建的生成器的网络结构如图 4-17 所示,语义约束 StarGAN 网络包括编码器和解码器两大组件,其中编码器由三个编码分支组成:时段编码分支、语义编码分支、图像特征编码分支。图中红色虚线部分为编码器,绿色虚线部分为解码器。图像特征编码分支与 4.1.2 节 StarGAN 网络中的编码器相同,解码器为 StarGAN 网络传输层个解码器的组合,语义编码分支基础网络为 U-net 结构,时段编码分支为语义分割结果到红外辐射量的映射过程。

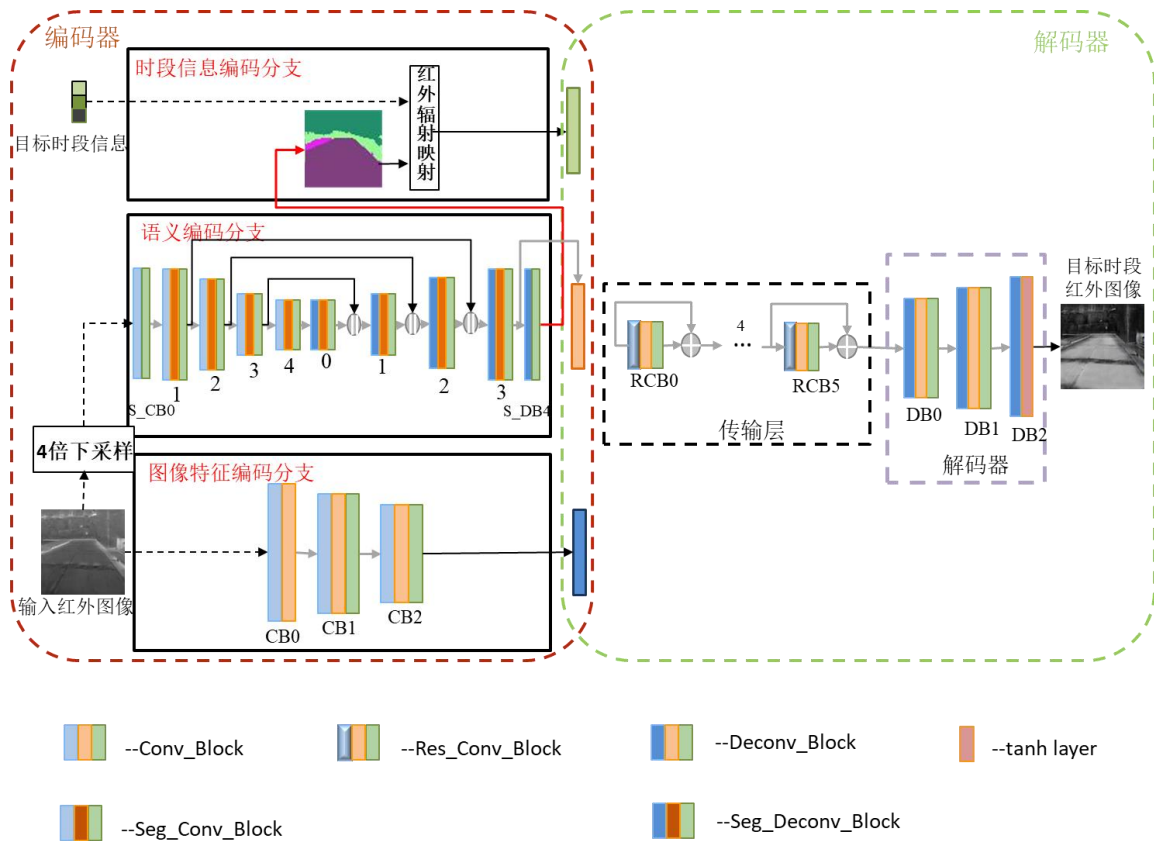


图 4-17 语义约束 StarGAN 生成器网络结构

生成器网络除了语义编码分支外,其他卷积块结构与 4.1.2 节相同,在此不再分析。语义编码模块主要由输入卷积层、4 个编码器子模块、4 个解码器子模块和输出卷积层组成,编码器子模块和解码器子模块结构分别如图 4-18、图 4-19 所示:

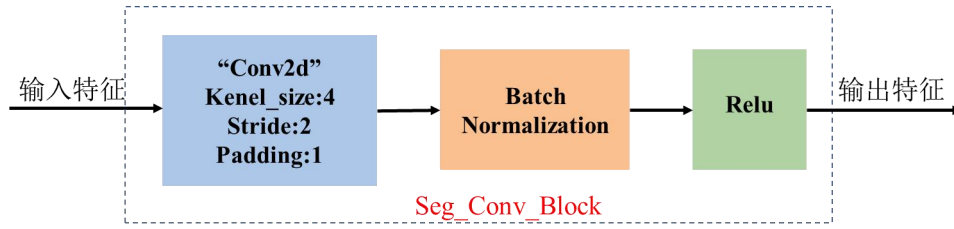


图 4-18 语义编码分支编码器子模块网络结构

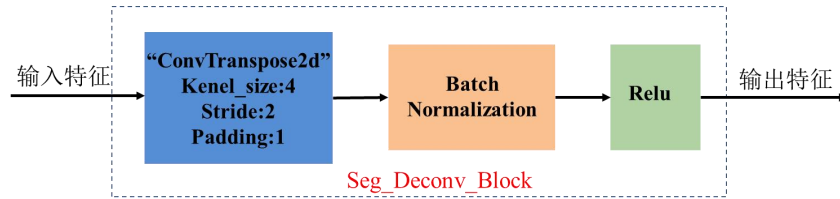


图 4-19 语义编码分支解码器子模块网络结构

G 网络的参数见表 4-5，表中卷积层参数(f,k,s,p)分别表示卷积核数量、大小、步长、填充尺寸，输入和输出参数(b,c,h,w)分别表示训练批大小、通道数、特征图高、宽。图像特征编码分支参数与 StarGAN 编码器参数相同。解码器结构与 StarGAN 传输层和解码器相同，但是参数设置不同。表 4-5 中解码器模块的 RCBs 表示传输层中的 6 个子模块，每个子模块参数相同。

表 4-5 语义约束 StarGAN 生成器网络参数表

模块	卷积块名称	卷积层参数 (f,k,s,p)	输入特征 (b,c,h,w)	输出特征 (b,c,h,w)
语义编码模块	S_CB0	(64,4,1,1)	(1, 3,64,64)	(1,64,64,64)
	S_CB1	(128,4,1,2)	(1, 64,64,64)	(1,128,32,32)
	S_CB2	(256,4,1,2)	(1,128,32,32)	(1,256,16,16)
	S_CB3	(512,4,1,2)	(1,256,16,16)	(1,512,8,8)
	S_CB4	(1024,4,1,2)	(1,512,8,8)	(1,1024,4,4)
	S_DB0	(512,4,1,2)	(1,1024,4,4)	(1,512,8,8)
	S_DB1	(256,4,1,2)	(1,512*2,8,8)	(1,256,16,16)
	S_DB2	(128,4,1,2)	(1,256*2,16,16)	(1,128,32,32)
	S_DB3	(64,4,1,2)	(1,128*2,32,32)	(1,64,64,64)

模块	卷积块名称	卷积层参数 (f,k,s,p)	输入特征 (b,c,h,w)	输出特征 (b,c,h,w)
	S_DB4	(20,4,1,1)	(1,64,64,64)	(1,20,64,64)
解码器	RCBs	(336,3,1,1)	(1,336,64,64)	(1,336,64,64)
	DB0	(128,4,1,2)	(1,336,64,64)	(1,128,128,128)
	DB1	(64,4,1,2)	(1,128,128,128)	(1,64,256,256)
	DB2	(3,7,1,1)	(1,64,256,256)	(1,3,256,256)

4.2.4 语义约束 StarGAN 损失函数

在 4.1.3 节介绍的 StarGAN 损失函数的基础上, 语义约束 StarGAN 网络添加了两
种损失函数: 语义分割损失函数、语义编码一致损失函数。

语义分割损失函数。 对于语义分割分支的训练, 本文采用有标签的监督式学习
方式, 采用交叉熵损失函数训练网络, 计算方式如下

$$L'_{seg} = E_{x,c^s} [-c^s \log(S(x))] \quad (4-13)$$

其中 c^s 为场景类别, S 表示语义编码分支。

语义编码一致损失函数。 为了保证生成的红外图像的语义信息与输入红外图像
保持一致, 本文提出了语义编码一致性损失, 要求生成的红外图像经过语义编码分
支后输出的语义分割结果与原始的红外图像语义信息一致, 如下式所示:

$$L_{rec_seg} = E_{x,c,c'} [-c^s \log(S(G(x,c)))] \quad (4-14)$$

综上所述, 生成器和判别器的优化损失函数分别如式 4-15、4-16 所示:

$$L_D = -L_{adv} + \lambda_{cls} L'_{cls} \quad (4-15)$$

$$L_G = L_{adv} + \lambda_{cls} L'_{cls} + \lambda_{rec} L_{rec} + \lambda_{seg} L'_{seg} + \lambda_{rec_seg} L_{rec_seg} \quad (4-16)$$

其中 λ_{cls} 、 λ_{rec} 、 λ_{seg} 、 λ_{rec_seg} 是分别控制时段分类损失、重建损失、语义分割损
失函数、语义编码一致损失函数相对于对抗损失的相对重要性的超参数。至此, 基
于语义约束 StarGAN 的红外图像时段拓展损失函数示意图如图 4-20 所示:

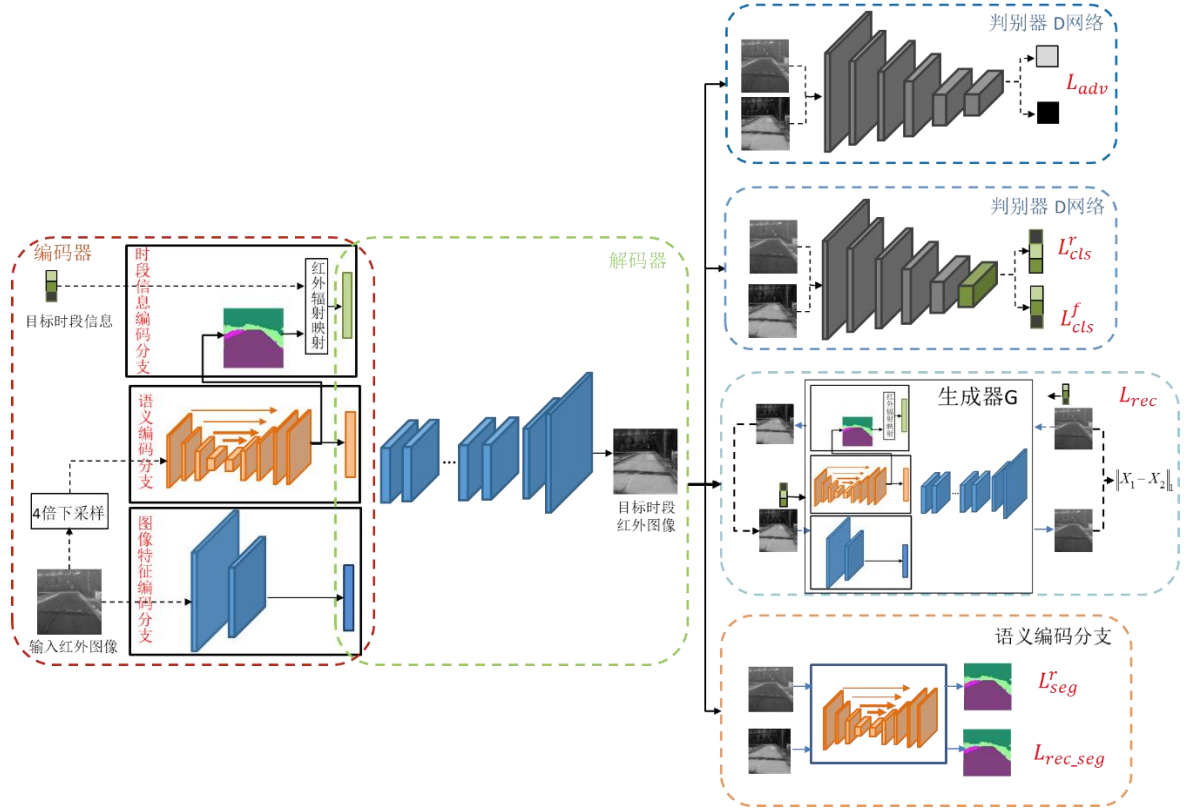


图 4-20 语义约束 StarGAN 损失函数示意图

4.2.5 红外图像时段拓展评价指标

灰度分布直方图：本文采用灰度分布直方图来统计图像的灰度分布情况，灰度分布直方图的计算公式如下：

$$p_h = \frac{n_h}{\sum n_h} \quad (4-17)$$

其中 h 表示灰度值(0-255)， p_h 表示灰度值 h 在图像中出现的概率， n_h 表示具有灰度值 h 的所有像素的个数。

巴氏系数：本文采用巴氏系数来衡量直方图分布的一致性，其计算公式如下：

$$BC(p, q) = \sum_{i=1}^n \sqrt{p_i q_i} \quad (4-18)$$

其中， p 和 q 分别是真实红外图像和拓展红外图像的直方图， n 是直方图分区数

结构相似度：由于本文采用的数据集没有图像对应的真实标签，为了衡量时段拓展后，图像结构及对应的语义结构的不变性，我们采用 SSIM 衡量时段拓展前后两张红外图像的结构相似度。结构相似度 SSIM 是用来评价两张图像的结构信息，数值越大，代表两张图像的结构越具有一致性，其计算公式已在 2.2.4 节说明。

4.2.6 模型算法与实验结果分析

根据 4.2.3 节搭建的语义约束 StarGAN 网络，采用与 4.1.4 节实验相同的数据集开展实验，经过 20000 次迭代后网络收敛，生成器循环一致损失和判别器的损失函数分别如图 4-21、图 4-22 所示，其中生成器损失函数一直保持下降，判别器对真假样本的判别能力大致一致，没有出现梯度崩塌现象。相比 StarGAN 网络的训练过程，判别器损失函数的震荡明显减小。

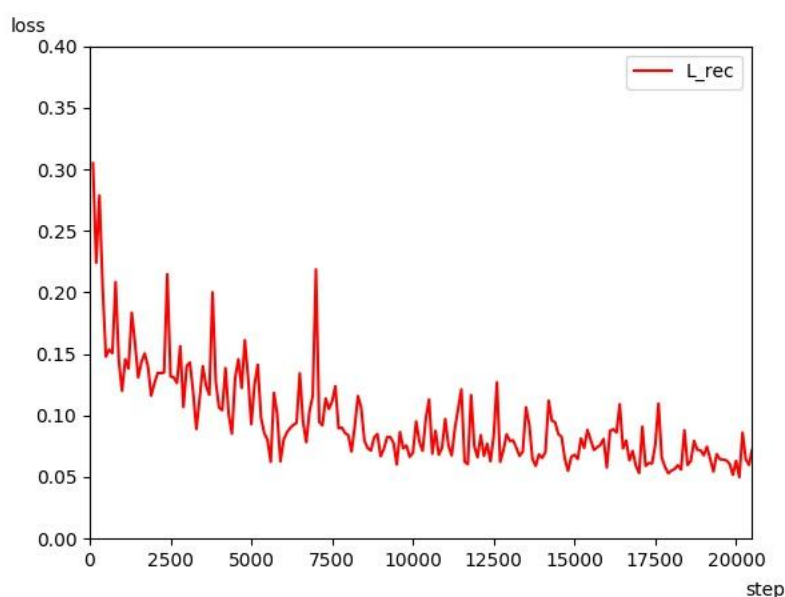


图 4-21 语义约束 StarGAN 生成器循环一致损失曲线

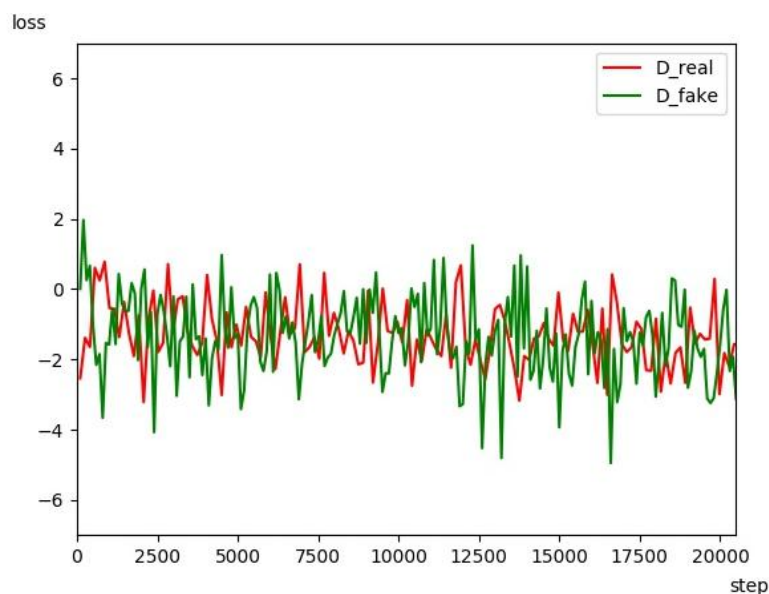
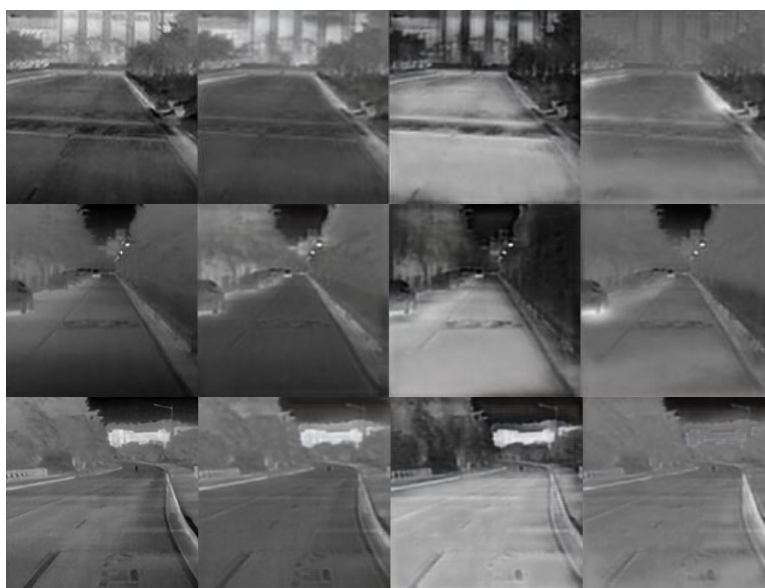


图 4-22 语义约束 StarGAN 判别器损失曲线

利用训练完成的生成器网络进行测试集数据的时段拓展，本节首先采用主观判断的方法衡量拓展结果的图像真实感和时段信息的有效性，图 4-23、图 4-24、图 4-25 分别展示了早上 5 点、下午 2 点、晚上 7 点的时段拓展结果。其中第一列为输入实拍红外图像，与 4.3.3 节采用的场景相同，第二、三、四列分别为经过时段拓展得到的早上 5 点、下午 2 点、晚上 7 点红外图像。



实拍早上 5 点 生成早上 5 点 生成下午 2 点 生成晚上 7 点

图 4-23 早上 5 点红外图像拓展后结果



实拍下午 2 点 生成早上 5 点 生成下午 2 点 生成晚上 7 点

图 4-24 下午 2 点红外图像拓展后结果



实拍晚上 7 点 生成早上 5 点 生成下午 2 点 生成晚上 7 点

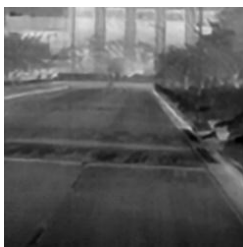
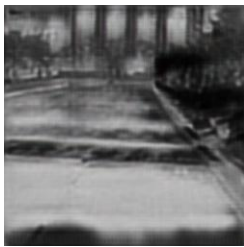
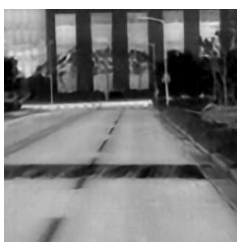
图 4-25 晚上 7 点红外图像拓展后结果

从图 4-23、图 4-24、图 4-25 中我们可以看出，语义约束 StarGAN 网络每个

时段拓展得到的红外图像都有较逼真的红外纹理，保留了原始时段红外图像的场景信息，同时也较好的拓展到了对应时段的红外图像域。相比 4.1.4 节采用相同场景得到的实验结果，语义约束 StarGAN 实现了明显的生成性能提升。

由于本文采用的数据集为非配对数据集，所以无法采用 SSIM、PSNR 等相似度量指标进行网络性能评估。本文采用相似场景的对应时段红外图像作为参考图像对拓展性能进行主观判断。从上述分析我们可以知道针对实拍图像为下午 2 点的红外图像，StarGAN 网络和语义约束 StarGAN 网络都得到了较好的拓展结果，所以论文重点关注两种网络对实拍早上 5 点红外图像的拓展结果和实拍晚上 7 点红外图像的拓展结果，分别如表 4-6、表 4-7 所示。

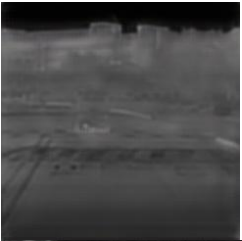
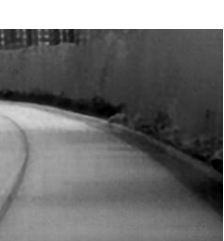
表 4-6 实拍早上 5 点红外图像的拓展结果

拓展时段	早上 5 点实拍图像	StarGAN 网络生成图像	语义约束 StarGAN 网络生成图像	相似场景实拍图像
拓展到下午 2 点				
拓展到晚上 7 点				

从表 4-6 可以看出，StarGAN 网络和语义约束 StarGAN 网络能够较好地将早上 5 点的红外图像拓展到晚上 7 点的红外图像，与对应的相似场景晚上 7 点的实拍图像有较高的相似度。而对于从早上 5 点红外图像拓展到下午 2 点的拓展结果可以看出，语义约束 StarGAN 网络相比 StarGAN 网络生成的红外图像，在道路上的拓展结果与实拍场景更相似。实验表明语义约束 GAN 通过在训练过程中对拓展结果进行语义约

束，避免了相同材质拓展结果出现灰度突变等拓展结果错误的问题。

表 4-7 实拍晚上 7 点红外图像的拓展结果

拓展时段	晚上 7 点实拍图像	StarGAN 网络生成图像	语义约束 StarGAN 网络生成图像	相似场景实拍图像
拓展到早上 5 点				
拓展到下午 2 点				

从表 4-7 可以看出，在早上 5 点的拓展结果中，语义约束 StarGAN 网络相比 StarGAN 网络各种材质的拓展结果更准确，树木和建筑物的拓展结果与实拍图像更一致，进一步证明了语义编码分支和语义约束时段编码信息的有效性。在下午两点的拓展结果中可以看出，语义约束 StarGAN 网络在道路场景的拓展结果与实拍图像相似度更高。

为了更加客观的判断语义约束 StarGAN 网络的优越性，本文采用了实拍图像、StarGAN 网络时段拓展结果和语义约束 StarGAN 网络的拓展结果的灰度直方图进行对比。我们针对道路和树木两种材质进行分析，表 4-8 展示道路直方图的统计结果，表 4-9 展示树木直方图的统计结果。根据直方图统计结果，对于道路场景，我们可以看到语义约束 StarGAN 网络与实拍红外图像更相近，StarGAN 网络在下午 2 点时段与实拍直方图分布偏差较大，实拍红外图像像素集中分布于灰度值大于 150 以上，而 StarGAN 网络集中于 150 以下。对于树木场景，StarGAN 灰度分布在下午 2 点出

现了较大偏差，实拍红外图像像素值集中分布在灰度值小于 100 以下区域，而 StarGAN 网络仍有较多像素处于 100 以上；在晚上 7 点时段，实拍红外图像像素值集中分布在灰度值大于 100 以下区域，而 StarGAN 网络仍有较多像素处于 100 以下。通过对道路和树木两种材质的拓展结果进行分析，证明了语义约束 StarGAN 网络生成图像相比 StarGAN 网络有更好的红外纹理分布仿真能力。

表 4-8 道路灰度直方图对比

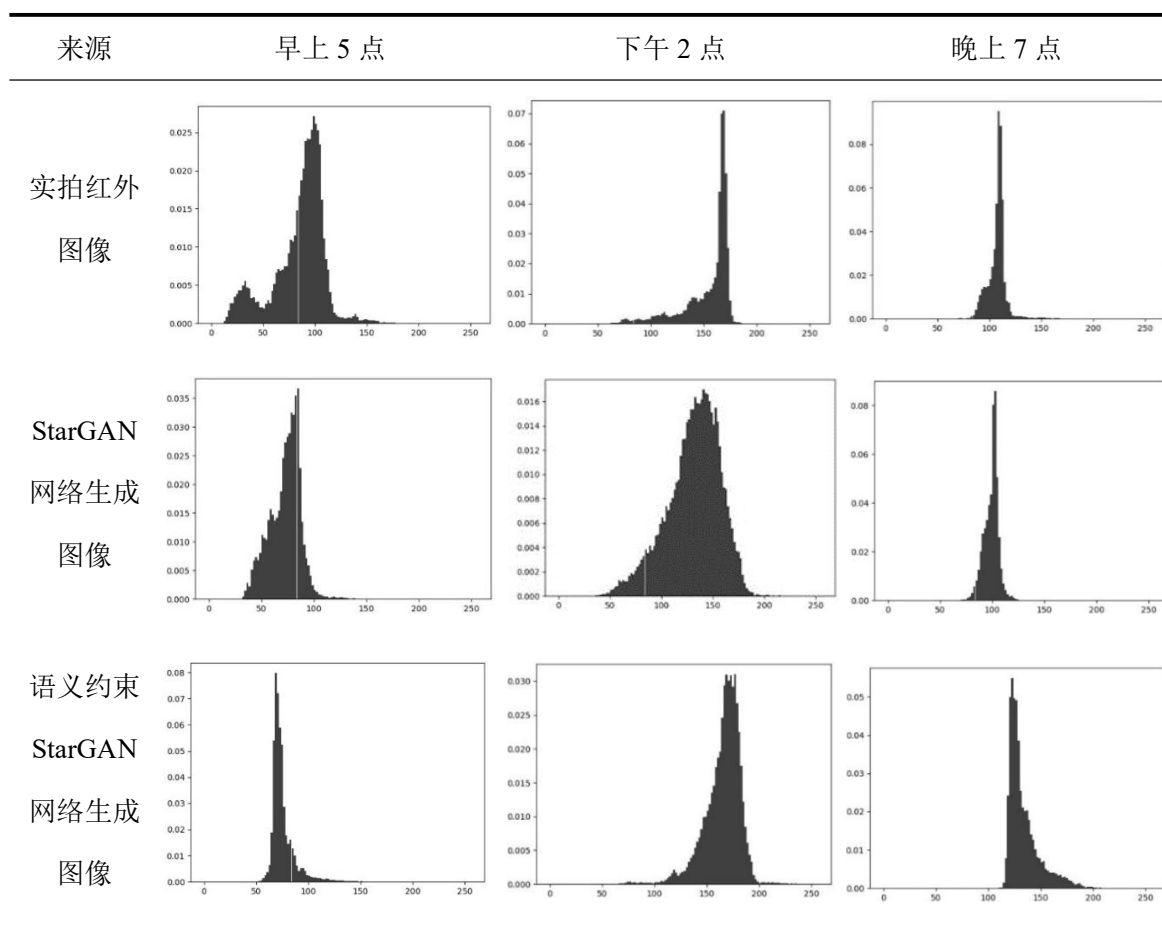
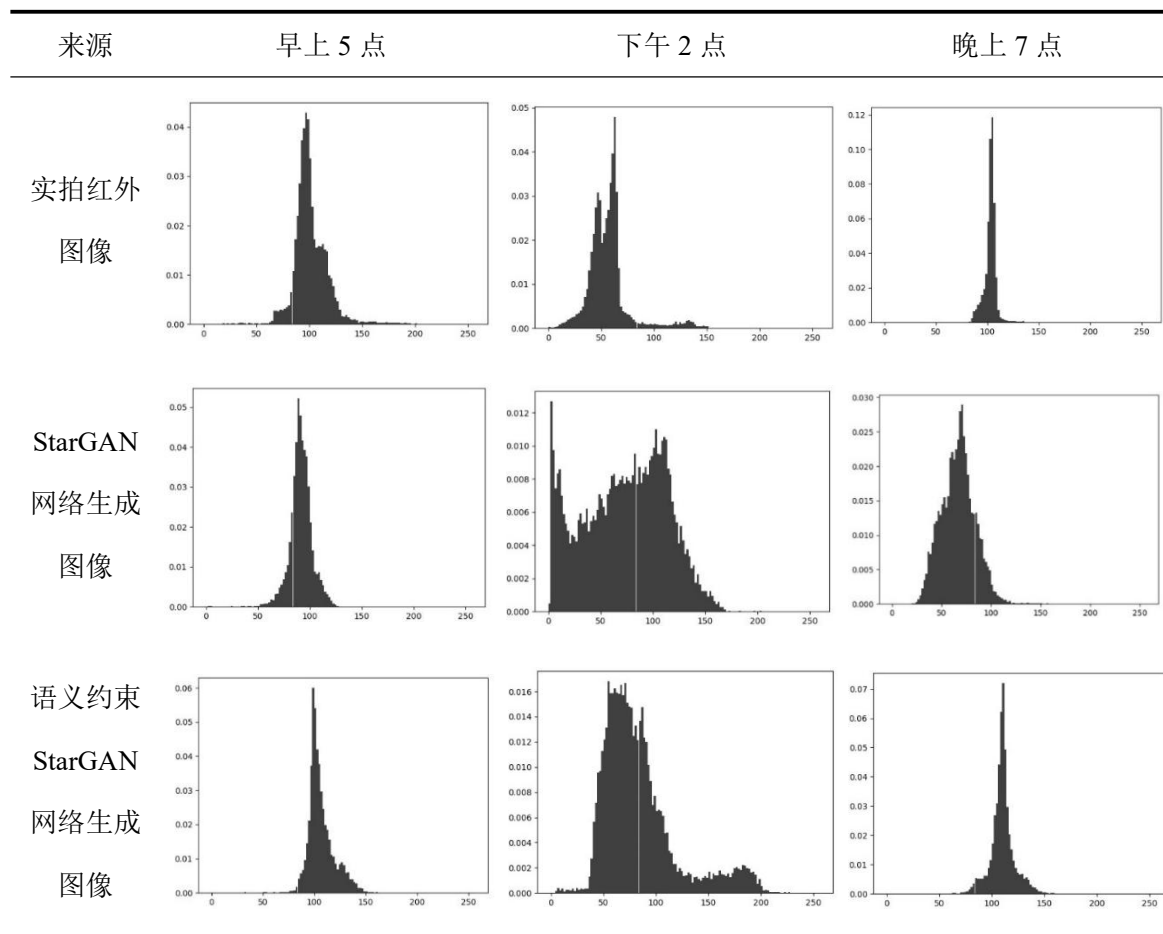


表 4-9 树木灰度直方图对比



同时在对比实验中，本文分别采用 StarGAN 网络和语义约束 StarGAN 网络，生成了早上 5 点、下午 2 点和晚上 7 点的仿真图像，并于同时刻、同场景的真实图像进行了对比，评价指标采用了巴氏系数和 SSIM，分别如表 4-10 和表 4-11。

从表 4-10 中可以看出，不管是道路还是树木场景，语义约束 StarGAN 网络生成结果相比 StarGAN 有较大巴氏系数值，说明语义约束 StarGAN 网络拓展得到的红外图像灰度分布更符合实拍红外图像灰度分布。

表 4-10 语义约束 StarGAN 与 StarGAN 性能指标

模型	拓展后时段	巴氏系数（道路）	巴氏系数（树木）
StarGAN	早上 5 点	0.765	0.800
	下午 2 点	0.891	0.948
	晚上 7 点	0.877	0.911
语义约束 StarGAN	早上 5 点	0.882	0.893
	下午 2 点	0.894	0.978
	晚上 7 点	0.902	0.935

表 4-11 统计了每个时段的红外图像时段拓展前后图像对的结构相似度。从表中可以看出，语义约束 StarGAN 网络在原图像时段为早上 5 点的拓展结果，相比 StarGAN 网络有显著的性能提升，与之前的分析相符，早上 5 点的红外图像由于灰度均值较小、对比度较低，StarGAN 网络拓展结果相比其他时段较差，而语义约束 StarGAN 拓展结果仍实现了一定的结构相似度。下午 2 点和晚上 7 点的 SSIM 值显示，语义约束 StarGAN 网络相比 StarGAN 仍保留了较多的结构信息。

表 4-11 时段拓展前后 SSIM 值

模型	原图像时段	拓展后图像时段		
		早上 5 点	下午 2 点	晚上 7 点
StarGAN	早上 5 点	0.361	0.242	0.352
	下午 2 点	0.552	0.804	0.708
	晚上 7 点	0.723	0.625	0.847
语义约束 StarGAN	早上 5 点	0.883	0.673	0.772
	下午 2 点	0.620	0.841	0.652
	晚上 7 点	0.808	0.748	0.843

4.3 有雾图像多时段红外图像生成

根据第二章搭建的去雾网络、第三章搭建的异源红外图像生成网络和本章搭建的红外图像时段拓展网络，本节将 4 张有雾图像分别进行去雾处理、可见光图像生成红外图像处理、时段拓展处理，分别得到了对应的早上 5 点、下午 2 点、晚上 7 点时段红外图像，如图 4-26 所示，其中第一列为雾图，第二列为去雾后的清晰图，第三列为生成的早上 5 点红外图像，第四列为生成的下午 2 点红外图像，第四列为生成的晚上 7 点红外图像。从图中我们可以看到，由一张雾图生成了三张不同时段的红外图像，实验表明本文算法能够根据易获得的可见光图像进行快速大量的红外多时段红外图像生成。

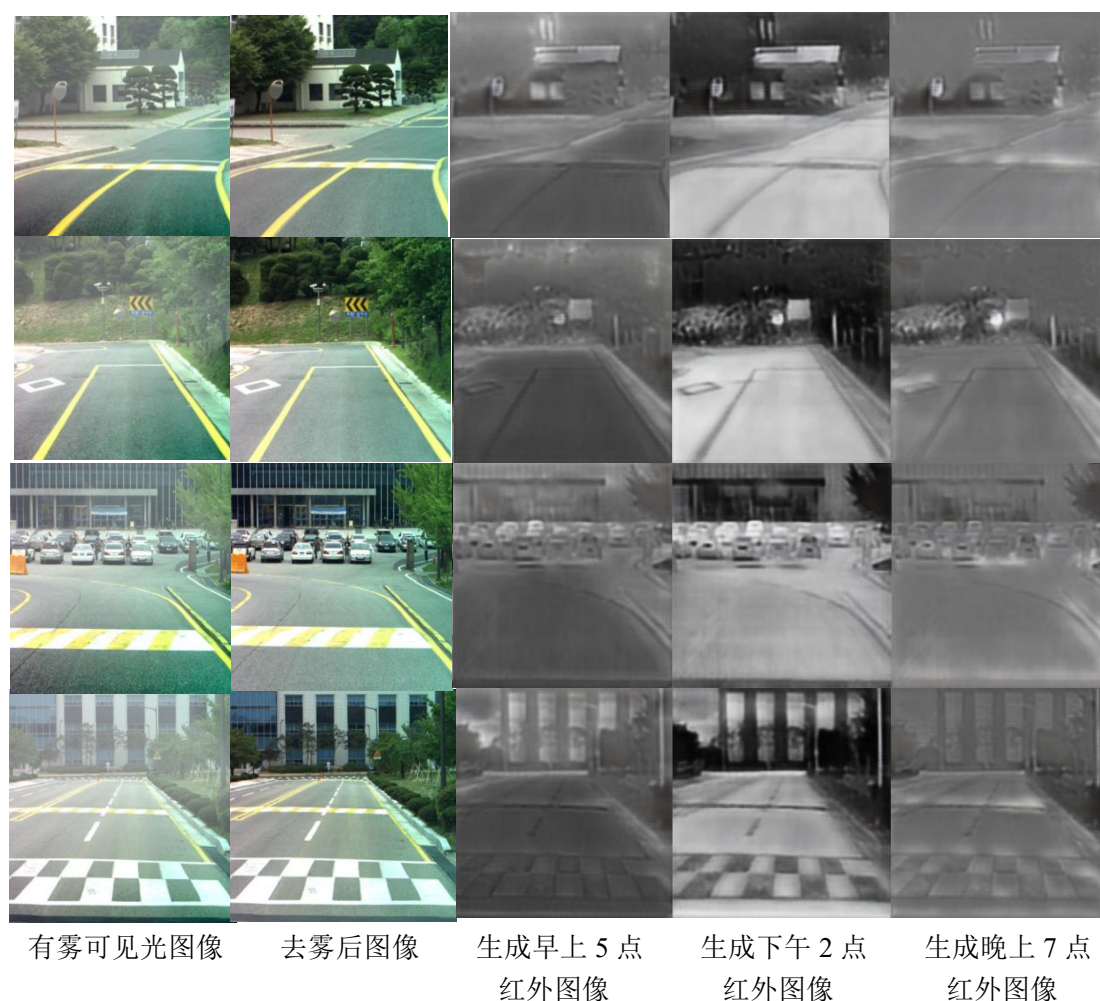


图 4-26 雾图生成对应三个时段的红外图像结果图

4.4 本章小结

本章针对非匹配形式的红外图像时段拓展任务，首先分析了使用 StarGAN 网络进行红外图像时段拓展的原因，并对实验结果进行分析。之后针对 StarGAN 网络拓展后图像出现同一材质拓展结果不同和时段拓展结果错误的问题进行分析，提出了基于语义约束 StarGAN 网络的红外图像时段拓展方法，一方面对网络训练进行语义约束，加入场景语义分割信息和语义编码一致损失；另一方面，提出了基于语义分割图和红外纹理时变曲线的场景时段信息编码方式。实验表明语义约束 StarGAN 网络更适用于红外图像拓展任务。

5 全文总结与展望

5.1 论文主要工作

针对红外图像样本不足的问题，论文将生成对抗网络引入红外图像生成任务，研究了基于可见光图像的红外图像生成方法和红外图像时段拓展技术，来实现快速、高效的红外图像生成。具体工作主要包含以下三个部分：

(1) 对可见光图像的去雾算法进行研究。针对可见光图像与红外图像的转换任务中可见光图像易受雾霾影响的问题，研究基于卷积神经网络的图像去雾算法。考虑到雾图中局部区域雾化程度不同，提出了一种基于图像雾化程度的图像去雾算法，通过根据透射率自适应的图学习分割阈值，对图像特征进行分割，实现了对不同雾化程度图像区域的并行处理。实验表明本算法在公开数据集和自然图像上都实现了很好的去雾效果。

(2) 对基于可见光图像的红外图像生成研究进行研究。针对 Pix2pix 网络生成的红外图像细节信息缺失的问题，本文提出基于多感受野特征融合 Pix2pix 的异源红外图像生成算法，构建了多感受野特征提取器，提出了逐级嵌套像素级的多感受野特征融合机制，实现了具有细节仿真能力的红外图像快速生成方法。实验表明本文提出的异源图像生成的算法具有更细节的红外图像纹理仿真能力。

(3) 对基于生成对抗网络的红外图像时段拓展进行研究。针对配对的红外图像时段数据缺失的问题，提出了基于语义约束的 StarGAN 网络的红外图像时段拓展算法。针对 StarGAN 网络对同一材质图像区域拓展结果不同的问题，论文提出了语义约束 StarGAN 网络，在 StarGAN 网络基础上添加了语义分割分支，加入场景语义特征，设计了语义编码一致损失。针对 StarGAN 网络对部分材质图像区域拓展结果错误的问题，论文改进了语义约束 StarGAN 网络的时段编码方式，提出了基于语义分割图和红外辐射时变曲线的时段信息编码方式。实验表明语义约束 StarGAN 网络更适用于红外图像时段拓展任务。

5.2 未来工作展望

本文在图像去雾和红外图像生成任务上取得了一定的成果，但是受研究时间、本人研究水平的限制，论文中的工作仍然需要进一步地分析与研究：

（1）在本文提出的基于图像雾化程度的图像去雾算法中，由于阈值模块的卷积层需要学习全局特征并且得到一维阈值，所以卷积核大小需根据雾图图像大小进行计算，故不能满足任意大小的图像生成，后续研究可以进一步优化网络结构设计，实现任意大小图像去雾。

（2）本文在进行时段拓展任务时，采用的数据集大小为 6000 张，样本空间有限，后续需制备更大规模的样本集，以提升算法的性能，验证算法的泛化能力。

（3）红外辐射与材质特性密切相关，后续研究可以进一步考虑对红外场景的材质进行预测，制备相应的材质数据库，结合材质信息进一步提升生成图像的真实感。

（4）红外辐射不仅与材质特性密切相关，而且受环境影响，如雨雪天气、沙尘暴天气等，论文仅对有雾条件下的红外图像生成算法进行了研究，后续实验可进一步完成不同环境下红外图像的生成实验。

致谢

时光荏苒，转眼间两年的研究生生活也步入尾声，我也将结束我的大学生活，开启人生的另一个征程。借此机会，我要向在求学期间一直给予我帮助的老师和朋友表示感谢。

感谢我的导师左峥嵘老师的悉心教导。从大四保研结束，左峥嵘老师便提前开始了对我的指导，帮助我尽早熟悉实验室相关课题。在研一初期，左峥嵘老师根据我平时的文献积累和研究兴趣，帮我确定了生成对抗网络这一研究方向，并且针对红外图像生成这一课题，在前人工作量较少的前提下，鼓励我打开思路，当我遇到瓶颈的时候，帮助我解疑答惑。左峥嵘老师对待科研的严谨认真和对学生的耐心负责，在我心中留下了不可磨灭的印象，同时也让我有了新的学习榜样。由于新冠疫情的影响，我们都开始了在家办公的方式，每周例会左峥嵘老师都会针对论文的完成情况进行讨论，尽力解决我们的问题。尽管不能对论文内容进行面对面交谈，但是左峥嵘老师对我们的论文进行了仔细地批注、修改。学生两年来让您费心了，衷心地希望您身体健康，工作顺利！

感谢实验室的吴双忱师兄，感谢吴师兄在我遇到困难时给我提出建议，帮助我走出困境；感谢湛炳汉师兄、刘嘉文师姐、韩金成师兄在科研生活中给我的帮助；感谢张维、侯云龙帮我出谋划策；感谢袁定兴、沈凡姝、马阳阳、李凌岳师弟师妹的支持与帮助；感谢我的好朋友陈曦的两年的陪伴，给予我温暖和依赖。

最后，我要感谢我的家人，感谢我的父母和弟弟一直以来对我无条件的信任和支持，给予我不断向前的动力，谢谢你们！

参考文献

- [1] 李晨阳. 基于三维场景的红外成像仿真系统及实现[D].中国科学院大学(中国科学院国家空间科学中心), 2019.
- [2] 罗晓春, 孙继银, 刘婧, 夏晶. 利用反演可见光图像的方法实现红外图像的获取[J]. 红外与激光工程, 2008(05): 773-776.
- [3] 冯云飞. 基于实拍图像的红外图像波段拓展方法研究[D]. 西安电子科技大学, 2019.
- [4] 姜枫. 基于生成对抗网络的图像渐进式生成算法研究[D]. 浙江大学, 2019.
- [5] Jacobs P A M. Simulation of the thermal behavior of an object and its nearby surroundings[J]. 1980.
- [6] Ben-Yosef N, Wilner K, Simhony S, et al. Measurement and analysis of 2-D infrared natural background[J]. Applied optics, 1985, 24(14): 2109-2113.
- [7] Biesel H, Rohlfing T. Real-time simulated forward looking infrared (FLIR) imagery for training[C]//Infrared Image Processing and Enhancement. International Society for Optics and Photonics, 1987, 781: 71-81.
- [8] Gambotto J P. Combining image analysis and thermal models for infrared scene simulations[C]//Proceedings of 1st International Conference on Image Processing. IEEE, 1994, 1: 710-714.
- [9] Cathcart, Michael J. Generation and application of high-resolution infrared computer imagery[J]. Optical Engineering, 1991, 30(11):1745.
- [10] Hong H K, Han S H, Hong G P, et al. Simulation of reticle seekers using the generated thermal images[C]//Proceedings of APCCAS'96-Asia Pacific Conference on Circuits and Systems. IEEE, 1996: 183-186.
- [11] Poglio T, Savaria E, Wald L. Specifications and conceptual architecture of a thermal infrared simulator of landscapes[J]. Proceedings of SPIE - The International Society for Optical Engineering, 2001, 4540:488-497.

- [12]Haynes A W, Gilmore M A, Filbee D R, et al. Accurate scene modeling using synthetic imagery[C]//Targets and Backgrounds IX: Characterization and Representation. International Society for Optics and Photonics, 2003, 5075: 85-96.
- [13]Latger J, Cathala T, Douchin N, et al. Simulation of active and passive infrared images using the SE-WORKBENCH[C]//Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XVIII. International Society for Optics and Photonics, 2007, 6543: 654302.
- [14]Cathala T, Douchin N, Joly A, et al. The use of SE-WORKBENCH for aircraft infrared signature, taken into account body, engine, and plume contributions[C]//Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXI. International Society for Optics and Photonics, 2010, 7662: 76620U.
- [15]Mielikainen J, Huang B, Huang H L A. GPU-accelerated multi-profile radiative transfer model for the infrared atmospheric sounding interferometer[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2011, 4(3): 691-700
- [16]张建奇, 方小平, 张海兴, 杨威, 朱长纯. 自然地表红外辐射特性统计分析[J]. 红外与毫米波学报, 1997(04): 27-33.
- [17]王章野, 陆艳青, 彭群生, 江照意, 朱华. 基于气象学和传热学的城市建筑物红外成像模型[J]. 系统仿真学报, 2000(05): 517-523.
- [18]宣益民, 李德沧, 韩玉阁. 复杂地面背景的红外热像合成[J]. 红外与毫米波学报, 2002(02): 133-136.
- [19]陈珊, 孙继银, 郭文普, 李琳琳. 一种新的红外纹理生成方法[J]. 大连海事大学学报, 2010, 36(04): 103-106.
- [20]周强, 白廷柱, 刘明奇, 邱纯. 基于可见光图像的近红外场景仿真[J]. 红外技术, 2015, 37(01): 11-15.
- [21]杨壹斌, 李敏, 杨敏, 王亚楠. 基于图像差值调制模型的红外纹理生成方法[J]. 红外技术, 2017, 39(03): 214-220.

- [22]吴亚平, 张天序. 红外图像的计算机仿真及仿真软件[J]. 红外与激光工程, 2000(04): 1-3.
- [23]笪邦友, 桑农, 张天序. 一种利用 Vega 软件的长波红外图像仿真方法[J]. 红外与激光工程, 2007(03): 333-337.
- [24]钟鸣. 基于 SE-Workbench 的红外场景图像自动生成方法研究[D]. 华中科技大学, 2016.
- [25]Gatys L A, Ecker A S, Bethge M. Image style transfer using convolutional neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition.2016: 2414-2423.
- [26]Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfe and super-resolution[C]//European conference on computer vision. Springer, Cham, 2016: 694-711.
- [27]Isola P, Zhu J Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1125-1134.
- [28]Wang T C, Liu M Y, Zhu J Y, et al. High-resolution image synthesis and semantic manipulation with conditional gans[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8798-8807.
- [29]Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2223-2232.
- [30]Kim T, Cha M, Kim H, et al. Learning to discover cross-domain relations with generative adversarial networks[C]//International Conference on Machine Learning.2017: 1857-1865.
- [31]Yi Z, Zhang H, Tan P, et al. DualGAN: Unsupervised duel learning for image-to-image translation[C]//International Conference on Computer Vision. 2017: 2849-2857.

- [32]Liu M Y, Breuel T, Kautz J. Unsupervised image-to-image translation networks[C]//Advances in Neural Information Processing Systems. 2017: 700-708.
- [33]Huang X, Liu M Y, Belongie S, et al. Multimodal unsupervised image-to-image translation[C]//European Conference on Computer Vision. 2018: 172-189.
- [34]Choi Y, Ghoi M, Kim M, et al. StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation[C]//International Conference on Computer Vision and Recognition. 2018: 8789-8797.
- [35]谢江荣, 李范鸣, 卫红, 李冰. 基于生成对抗式神经网络的红外目标仿真方法[J]. 光学学报, 2019, 39(03): 150-156.
- [36]He K, Sun J, Tang X. Single image haze removal using dark channel prior[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 33(12): 2341-2353.
- [37]Zhu Q, Mai J, Shao L. A fast single image haze removal algorithm using color attenuation prior[J]. IEEE Transactions on Image Processing, 2015, 24(11): 3522-3533.
- [38]Berman D, Avidan S. Non-local image dehazing[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 1674-1682.
- [39]Ren W, Liu S, Zhang H, et al. Single image dehazing via multi-scale convolutional neural networks[C]//European Conference on Computer Vision. Springer, Cham, 2016: 154-169.
- [40]Li Boyi, Peng Xiuliam, et al. An all-in-one network for dehazing and beyond[C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 4770-4778.
- [41]Mei K, Jiang A, Li J, et al. Progressive feature fusion network for realistic image dehazing[C]//Asian Conference on Computer Vision. Springer, Cham, 2018: 203-215.
- [42]Bharath Raj N., Venkateswaran N. Single Image Haze Removal using a Generative Adversarial Network. arXiv:1810.09479, 2018
- [43]Malav R, Kim A, Sahoo S R, et al. DHSGAN: An End to End Dehazing Network for

- Fog and Smoke[C]//Asian Conference on Computer Vision. Springer, Cham, 2018: 593-608.
- [44]McCartney E J. Optics of the atmosphere:scattering by mole-cules and particles[M]. New York:John Wiley and Sons,1976:421-421.
- [45]Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 4700-4708.
- [46]Xian K, Shen C, Cao Z, et al. Monocular relative depth perception with web stereo data supervision[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 311-320.
- [47]Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation[J]. 2015, 9351:234-241.
- [48]Zhou Z, Siddiquee M M R, Tajbakhsh N, et al. Unet++: A nested u-net architecture for medical image segmentation[M]//Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer, Cham, 2018: 3-11.
- [49]Zhang H, Goodfellow I, Metaxas D, et al. Self-Attention Generative Adversarial Networks[C]//International Conference on Machine Learning. 2019: 7354-7363.
- [50]Arjovsky M, Chintala S, Bottou L. Wasserstein Generative Adversarial Networks[C]//International Conference on Machine Learning. 2017: 214-223.
- [51]Gulrajani I, Ahmed F, Arjovsky M, et al. Improved training of wasserstein gans[C]//Advances in Neural Information Processing Systems. 2017: 5767-5777.

附录 1 攻读硕士学位期间的科研成果

（一）发表学术论文

- [1] Yanling Hua, Zhengrong Zuo. Dehazing network based on haze density[C]//MIPPR 2019: Pattern Recognition and Computer Vision. International Society for Optics and Photonics, 2020, 11430: 114300V.

（二）申请（授权）专利

- [1] 左峥嵘, 化彦伶, 吴双忱, 申请号: 201911016783.1 《一种基于卷积神经网络的图像去雾方法》。