第 45 卷 第 5 期
2024 年 5 月

仪 器 仪 表 学 报
Chinese Journal of Scientific Instrument

Vol. 45 No. 5
May 2024

DOI: 10.19650/j.cnki.cjsi.J2312185

基于知识蒸馏自适应 DenseNet 的无人机对地目标可见光与红外图像融合*

童小钟,赵宗庆,苏绍璟,左 震,孙 备
(国防科技大学智能科学学院 长沙 410072)

摘 要:可见光与红外图像融合旨在利用两种不同传感器之间有效的信息,通过互补的图像特征实现图像增强。然而,当前基于深度学习的融合方法倾向于优先考虑评价指标。模型的复杂性较高,权重参数较大,推理性能低,泛化性较差,不易部署到无人机载边缘计算端。为了应对这些挑战,本文提出了一种新颖的可见光与红外图像融合方法,即知识蒸馏的自适应 DenseNet 来学习预先存在的融合模型,通过使用超参数(例如宽度和深度)来实现融合效果和模型轻量化。本文提出的方法在典型地面目标数据集进行了评估,实验结果表明,该模型参数仅为 77 KB,推理时间为 0.95 ms,具有超轻量的网络结构,良好的图像融合效果和复杂场景下较强的泛化能力。

关键词: 可见光与红外图像;图像融合;知识蒸馏;自适应;无人机
中图分类号: TP391.4 TH701 **文献标识码**: A **国家标准学科分类代码**: 510.4050

Fusion of visible and infrared images of ground targets by unmanned aerial vehicles based on knowledge distillation adaptive DenseNet

Tong Xiaozhong, Zhao Zongqing, Su Shaojing, Zuo Zhen, Sun Bei
(*College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410072, China*)

Abstract: Visible and infrared image fusion aims to exploit the effective information between two different sensors to achieve image enhancement through complementary image features. However, current deep learning-based fusion methods tend to priorities evaluation metrics, and the models have high complexity, large weight parameters, low inference performance, poor generalization, and are not easy to deploy on the UAV edge computing platform. To address these challenges, this paper proposes a novel approach for visible and infrared image fusion, i.e., adaptive DenseNet with knowledge distillation to learn a pre-existing fusion model, which achieves fusion effectiveness and model lightweighting through the use of hyperparameters (e.g., width and depth). The proposed method is evaluated on a typical ground target dataset, and the experimental results show that the model parameter is only 77 KB and the inference time is 0.95 ms, which has an ultra-light network structure, excellent image fusion effect and strong generalization ability in complex scenes.
Keywords: visible and infrared images; image fusion; knowledge distillation; adaptive; unmanned aerial vehicles

0 引 言

无人机(unmanned aerial vehicles, UAV)具有成本低廉、灵活性高、操作简单、部署方便和体积小等优点,在遥感测绘、农业植保、电力巡检和安防监视等民用领域[1]以及战场监视、侦察打击和毁伤评估等军事领域具有重要的应用价值。人工智能和计算机视觉技术的发展极大的拓展了无人机的环境感知能力和自主能力[2]。随着无人机携带载荷的多源化、作业任务的多样化以及任务区域环境复杂化的影响,提升无人机感知能力已成为热门的研究课题[3]。

无人机通常搭载可见光与红外载荷,可见光传感器在理想光照条件下能够捕获物体的纹理结构细节和颜色信息,但容易受光照条件的影响。红外传感器能够检测物体的红外辐射和温度变化,不易受光照变化的影响,可

收稿日期:2023-11-23 Received Date: 2023-11-23
*基金项目:国家自然科学基金(52101377,62201598)、湖南省研究生科研创新项目(CX20220015)资助

全天候使用,但对比度较低,纹理细节缺乏。可见光与红外图像融合旨在利用可见光和红外两种不同传感器之间有效的信息,提取并结合互补的特征实现图像增强。因此,图像融合在低照度和浓烟场景具有较广泛的应用前景,通过融合可见光和红外图像,无人机可以全天候执行侦察感知任务,无人车可以增强弱光和浓雾天气条件的自主导航和目标识别性能。学术上已有不少研究人员针对可见光与红外图像融合开展相关的课题研究[4-7],但这些研究未从无人机边缘计算平台的角度对这两种传感器高效的融合进行有效的探讨。本文主要解决复杂场景下无人机对地目标检测时存在的两个关键问题:1)如何在无人机平台实际应用中充分发挥两种传感器的特点,形成优势互补;2)如何设计轻量化的网络结构,使其容易部署到无人机实现实时图像融合。

随着图像融合任务的拓展,研究学者提出了许多方法来解决可见光与红外图像融合面临的挑战。根据特征提取和融合策略的不同,这些方法可以分为传统方法和深度学习方法。根据手工设计的特征分解和生成规则,传统的融合方法主要分为基于多尺度变换[8]、基于稀疏表示[9-10]、基于显著性[11-12]、基于模糊集[13]和基于混合[14]的方法。传统方法主要包括 3 个阶段。首先是特征提取阶段,利用特定的变换从源图像中提取特征;然后在特征融合阶段应用融合策略来组合这些特征;最后,在特征重建阶段应用相应的逆变换得到最终的融合图像。基于深度学习的方法主要包括基于自动编码器[14]、卷积神经网络[15]、生成对抗网络[16-17]和 Transformer 网络[18-19],这些方法通过精心设计的网络结构和损失函数实现高效的特征提取和特征重建,从而得到良好的融合结果。需要注意的是,与目标检测、目标跟踪和语义分割等有真值标签的高级视觉任务不同,可见光与红外图像融合缺少真值标签,无法提供有监督模型训练所需的标签数据。

虽然现有的融合方法具备一定的融合效果,但大多数方法忽略了无人机边缘设备计算资源稀缺以及算法实时性的要求,通过加深网络结构来优化融合图像的视觉质量和评价指标。因此,模型变得越来越复杂,模型的推理效率降低。知识蒸馏是一种常用的模型压缩技术,有助于将知识从预先训练的大型教师模型转移到小型且功能强大的学生模型[20]。本文受知识蒸馏的启发,设计了一种用于无人机平台的实时可见光与红外图像融合的自适应知识蒸馏方法。具体来说,该网络首先利用教师模型进行训练,学生模型通过知识蒸馏实现与教师模型相当的性能。然后通过自适应机制调整网络的超参数以实现更轻量化的网络结构。最后,将算法部署到无人机 NVIDIA Orin 边缘计算平台实现典型场景对地目标的可见光与红外图像融合。实验结果表明,该算法具有超轻量的网络结构,高效的推理性能和良好的图像融合效果,不同复杂场景的融合结果表明其具有良好的泛化性能。

1　方法与理论分析

1.1　知识蒸馏自适应 DenseNet 的可见光红外图像融合方法

1)方法概述

本文将可见光与红外图像的融合建模为神经网络的拟合问题,方法的整体框架如图 1 所示。

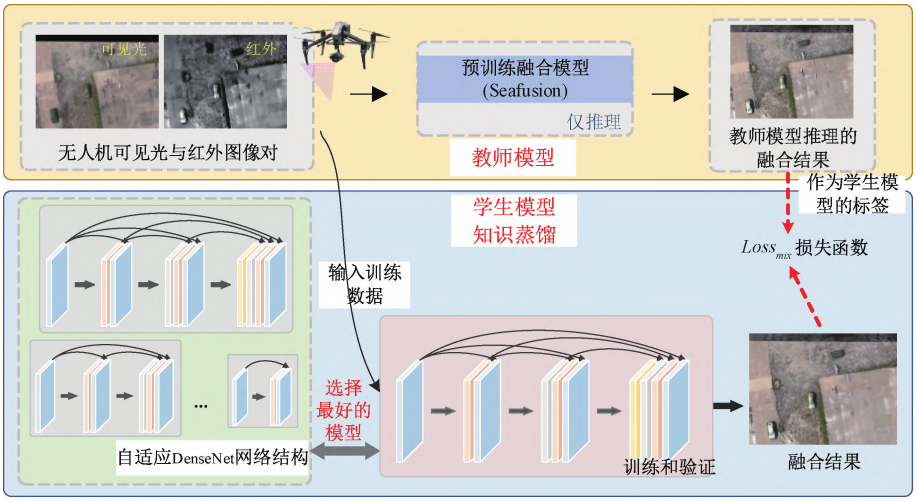


图 1　拟议的基于知识蒸馏自适应 DenseNet 框架

Fig. 1　The proposed adaptive DenseNet framework based on knowledge distillation

具体来说,首先,将公开的可见光与红外图像对[21]作为训练数据,该数据场景与实验场景存在明显的区别。然后通过教师模型推理得到的融合结果作为学生模型训练的标签。之后,可见光与红外图像对输入 DenseNet 模

型获得融合后的图像,计算其与教师模型推理获得的融合图像标签的损失函数 $Loss_{mix}$,使其收敛至最小值即可获得理想的融合模型。考虑无人机载边缘部署算力和功耗的需求,设计了两个调节模型复杂度的变量(n,m)为拟合任务选择最佳 DenseNet 网络结构。值得注意的是,与现有复杂的融合模型相比,本文提出的方法具有更低的复杂性、更少的参数量和更高的推理效率,同时提供良好的图像融合效果。当模型的复杂度参数设置为(n = 4,m = 8),只包含约 77 KB 参数,输入图像分辨率大小为 640×480 时,模型计算量为 5.66 G FLOPs。具体步骤包括以下 5 个方面:

(1)构建可见光和红外图像配对数据集(I_{vis},I_{ir})。

(2)选择教师模型(SeAFusion[22]),使用预训练权重推理数据集的可见光和红外图像(I_{vis},I_{ir}),获得融合图像结果。

(3)将步骤(2)获得的融合图像作为标签 $Label$,与原始可见光和红外图像配对数据集(I_{vis},I_{ir})相结合,组成一组新的带标签的数据集$\{I_{vis},I_{ir},Label\}$。

(4)使用步骤(3)数据集$\{I_{vis},I_{ir},Label\}$训练学生模型,即 DenseNet 网络结构。此步骤是实际的知识蒸馏过程,学生模型训练时使用教师模型推理得到的融合图像作为软标签进行有监督的学习。

(5)训练获得学生模型自适应 DenseNet 权重。学生模型与教师模型 SeAFusion 具有类似的融合效果,但在推理速度和模型复杂度方面远远优于教师模型。轻量化的网络结构使其适合在无人机载端部署。

2)自适应 DenseNet 的总体框架

本文提出的可见光红外图像融合的自适应 DenseNet 网络结构如图 2 所示。DenseNet 集成了两个调节模型复杂度的变量(n,m),n 表示密集层(DenseLayer)的数量,m 表示每个密集层输出的通道数。具体来说,自适应 DenseNet 模型的训练过程主要包括以下几个方面,首先将配准后的可见光红外图像对作为模型的输入;然后经过 ConvBlock 模块和 DenseLayer 模块处理;最后通过卷积和 Tanh 激活函数获得可见光红外融合图像输出结果。DenseLayer 模块有助于浅层的特征传递到网络的深层,从而增强可见光红外图像融合的效果。本文构建的模型没有上采样或下采样层,确保了每个密集层的输入和输出仅在通道数上有区别,特征图尺寸保持不变。本文提出的知识蒸馏 DenseNet 模型仅由极少数的卷积层组成,当模型的复杂度参数设置为(n = 4,m = 8),只包含约 77 KB 参数,NVIDIA Orin 边缘计算平台对 640×480 分辨率大小的图像推理时间为每秒 28 帧。本文设计的轻量化网络结构仅需要消耗非常少的计算资源,具有高效的推理性能,能够作为无人机载边缘计算部署较为理想的解决方案。

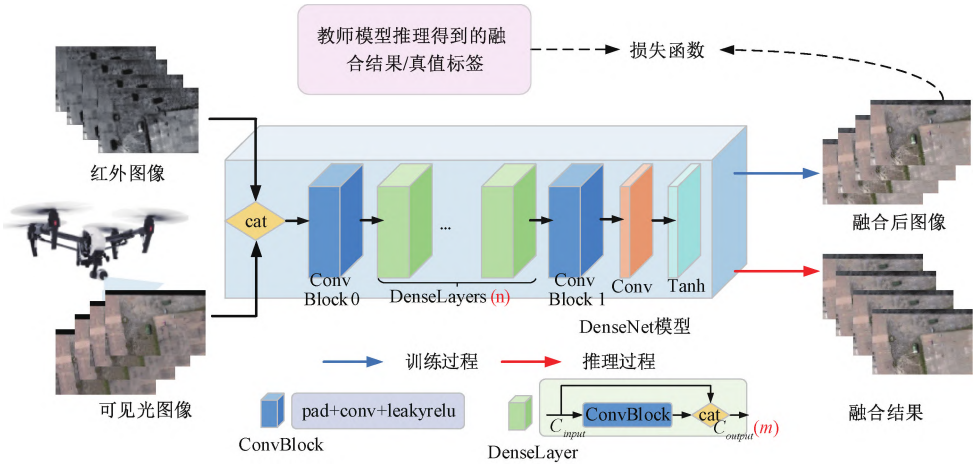


图 2 本文提出的 DenseNet 网络结构

Fig. 2 The proposed network structure of DenseNet

为了进一步说明构建的网络结构,表 1 详细描述了网络层,以及每层的输入和输出通道数,并提供了(n = 2,m = 8)和(n = 5,m = 16)示例。自适应调整设计知识蒸馏 DenseNet 模型性能的超参数包括密集层的数量 n 和每个密集层输出的通道数 m。

3)损失函数设计

本文提出的方法旨在更好的适应其他模型融合的结果。因此,将平均绝对误差(mean absolute error, MAE)和均方误差(mean square error, MSE)作为损失函数。平均绝对误差损失函数的定义如下:

$$loss_{mae} = | x_i - y_i | \tag{1}$$

其中,x_i 是自适应模型输出的融合图像序列,y_i 是教师模型输出的融合图像序列并将其作为标签。均方误差损失函数的公式如下:

$$loss_{mse} = (x_i - y_i)^2 \tag{2}$$

表 1　不同参数对应的网络结构

Table 1　Network structure corresponding to different parameters

网络层	$n=2,m=8$		$n=5,\ m=16$	
	输入通道	输出通道	输入通道	输出通道
ConvBlock0	4	8	4	16
DenseLayer1	8	16	16	32
DenseLayer2	16	24	32	48
DenseLayer3	–	–	48	64
DenseLayer4	–	–	64	80
DenseLayer5	–	–	80	96
ConvBlock1	24	32	96	32
Conv	32	3	32	3
激活函数	Tanh			

平均绝对误差损失对异常值表现出更大的弹性,但容易在全局最小值附近振荡,较难收敛到最优值。均方误差损失具有更快的收敛速度,但对异常值较为敏感。本文综合利用平均绝对误差和均方误差损失的优势,使用混合损失函数 $Loss_{mix}$ 对模型训练,计算方式如下:

$$Loss_{mix}=\begin{cases}\dfrac{1}{2}loss_{mse}^{\ 2}, & |x_i-y_i|\leqslant\delta\\ \delta\cdot loss_{mae}-\dfrac{1}{2}\delta^2, & |x_i-y_i|>\delta\end{cases}\tag{3}$$

其中,δ 设置为 1。混合损失函数 $Loss_{mix}$ 结合了 $loss_{mae}$ 和 $loss_{mse}$ 的优点,帮助模型更快收敛的同时,减少异常值的影响。对于样本集,通过平均损失表示,定义如下:

$$L(x,y)=mean\{l_1,l_2,\cdots,l_i\}\tag{4}$$

1.2　DenseNet 自适应优化策略

本文的目的是通过构建最优的 DenseNet 结构以适应现有的融合算法实现知识蒸馏。因此,将问题简化为离散条件下的自适应优化问题,通过寻找最优的 (n,m) 值来实现良好的拟合性能,同时考虑无人机边缘计算平台的功耗和推理性能,建立自适应优化的数学模型,定义如下:

变量:(n,m)

目标:$\min\{$给定每对图像的推理时间$|(n,m)\}$

约束:$|f(n,m)-f(n_{best},m_{best})|<\sigma,\ (n,m)\in N$

其中,m 为密集层输出通道数,n 为密集层数,$f(n,m)$ 表示拟合性能,$f(n_{best},m_{best})$ 表示理论上最优拟合性能,σ 表示阈值,N 是自然数。DenseNet 的自适应优化策略包括以下 3 个步骤:

1)(预处理阶段):首先,对于给定的输出通道数、密集层的数量和可见光红外图像对,计算每对图像执行推理所需的时间,以选择推理性能最好的模型。如图 3 所示显示了不同的 (n,m) 值自适应 DenseNet 可见光与红外图像融合的推理时间。横轴表示每个密集层的输出通道数(即 m 值,$m\in[4,16]$),4 条曲线表示密集层的数量(即 n 值,$n\in[2,5]$),纵轴表示模型的推理时间。从图 3 可以看出,密集层数量越多,则推理时间越长,但推理时间不一定随着 m 的增加而增加。此外,当 m 为 2 的指数幂时,在不增加推理时间的情况下,模型复杂度增加,具有更高的成本效益。因此,预处理阶段淘汰了部分 (n,m) 值构建的低成本 DenseNet 结构。例如,$(n=2,m=11)$ DenseNet 结构的实际复杂度明显低于 $(n=2,m=12)$ 的结构,但其推理时间明显更高。表 2 展示了去除低成本效益的 (n,m) 值后按推理时间从低到高排序的可选自适应 DenseNet 网络结构,其中每个 (n,m) 对对应一个序号 s。

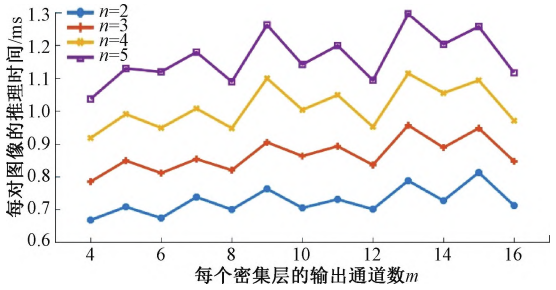


图 3　使用不同数量的密集层和每个密集层的输出通道对每对图像进行推理所需的时间

Fig. 3　Time required for inference for each pair of images using different number of dense layers and output channels of each dense layer

表 2　自适应 DenseNet 网络结构

Table 2　Adaptive DenseNet network architecture

序号	1	2	3	4	5	6	7	8
n	2	2	2	2	3	3	3	3
m	4	8	12	16	4	8	12	16
推理时间/ms	0.668	0.700	0.701	0.712	0.785	0.82	0.836	0.847
序号	9	10	11	12	13	14	15	16
n	4	4	4	4	5	5	5	5
m	4	8	12	16	4	8	12	16
推理时间/ms	0.918	0.948	0.952	0.971	1.037	1.09	1.094	1.117

2)(假设阶段):本文提出的方法基于机器学习知识执行自适应搜索最优解,利用以下假设:

A:删除低成本效益的 (n,m),模型与推理时间正相关,即序号越大,模型越复杂。

B:模型的复杂度与泛化拟合能力正相关。具体来说,给定两个模型 A 和 B,如果 A 比 B 更复杂,并且两个模型在具有足够训练和验证数据的拟合期间收敛,则 A 的拟合性能将与 B 一样好。相反,如果训练和验证数据不足,模型 A 和模型 B 都无法收敛。

3)(求解阶段):将 $f(n,m)$ 定义为 $1-Loss_{mix}(n,m)$,其中 $Loss_{mix}(n,m)$ 表示不同 (n,m) 值模型训练的损失函数。很明显,随着 $Loss_{mix}(n,m)$ 变小,模型生成的图像更接近预期图像,表明拟合效果更好。自适应最优搜索策略的伪代码如算法 1 所示。为了获得最优的自适应网络结构和可见光与红外图像融合效果,采用二分搜索策略自适应地获得最佳超参数 (n,m)。超参数 $(n,m)_{s_{low}}$、$(n,m)_{smid}$ 和 $(n,m)_{s_{high}}$ 分别表示 $s=1$、8 和 16 时的网络结构。初始化模型的超参数设置为 $(n_0,m_0)=(5,16)$ 和 $(n_1,m_1)=(2,4)$,初始阈值设置为 $\sigma=0.01\times[f(n_0,m_0)-f(n_1,m_1)]$。当 s 值最小时则认为达到最优情况。根据解获得的 (n,m) 值构建的自适应 DenseNet 即为最优网络。

算法 1:DenseNet 的自适应优化策略

输入:训练集 T $\{visible_i, infrared_i, label_i\}$,
验证集 V $\{visible_j, infrared_j, label_j\}$
输出:$s \rightarrow (n,m)_s$
开始
1. 计算 $f(n_0,m_0)$ 和 σ
2. $s_{low}=1 \rightarrow (n,m)_{s_{low}}$
3. $s_{high}=16 \rightarrow (n,m)_{s_{high}}$
4. $s_{mid}=(s_{high}+s_{low})//2=8 \rightarrow (n,m)_{s_{mid}}$
5. **While** $s_{low} \neq s_{high}$ **do**
6. **if** $|f(n,m)_{s_{mid}}-f(n_0,m_0)|<\sigma$ **do**
7. $s_{high}=s_{mid}$
8. $s_{mid}=(s_{high}+s_{low})//2$
9. **else do**
10. $s_{low}=s_{mid}+1$
11. $s_{mid}=(s_{high}+s_{low})//2$
12. return $s_{mid} \rightarrow (n,m)_{s_{mid}}$
结束

2 实 验

2.1 无人机对地目标图像融合硬件平台设计

根据无人机对地目标图像融合的任务需求,设计了图像融合的总体架构,如图 4 所示。无人机对地目标图像融合系统在无人机载端搭载光电吊舱(包括可见光相机、红外热像仪和激光测距仪)、GPS 模块、加速度计、边缘计算模块以及开源飞控和图数传一体的传输模块。机载端首先通过光电吊舱获取可见光与红外图像数据,然后边缘计算模块读取光电吊舱数据,在线实时处理可见光与红外图像,输出融合结果。地面端通过通信电台远程控制的方式对 NVIDIA Orin 进行操作,并通过显控平台实时接收机载数据,查看无人机载端实时操作和处理结果。

根据图 4 设计的总体架构搭建的无人机实物硬件平台如图 5 所示,该六旋翼无人机具有强大的负载能力和较长的续航能力,最大支持 10 kg 任务载荷和 50 min 标准续航时间。实验中将 NVIDIA Orin 边缘计算模块和 Q30T 光电吊舱挂载到无人机载端,地面端显控平台通过通信电台远程查看和控制 NVIDIA Orin 边缘计算模块,实现机载端图像融合实时处理。

2.2 典型数据集构建

数据集构建包括可见光图像和红外图像。融合模型要求输入图像空间上严格对齐,光电吊舱的可见光与红外图像分辨率和视场角不一致,无法直接用于模型的推理。因此,在数据采集和无人机载算法部署验证时,通过计算单应性矩阵裁剪可见光与红外图像空间不一致区域,使可见光图像和红外图像在空间上始终保持对齐。此时光电吊舱的可见光传感器为 2 倍变焦,红外传感器为 1 倍焦距。接下来证明可见光与红外图像对齐时单应性矩阵的不变性。

对于可见光图像的任一点 Q,通过投影变换模型映射到红外图像的点 q,该投影变换表示如下:

$$q=s\boldsymbol{M}\boldsymbol{T}Q \tag{5}$$

其中,$\boldsymbol{M}$ 为相机的内参数矩阵,s 为非零常数因子,$\boldsymbol{T}=[\boldsymbol{R}\ \boldsymbol{t}]$ 为物理变换矩阵,$\boldsymbol{R}$ 表示旋转矩阵,$\boldsymbol{t}$ 表示 3 维的列向量。$Q=(X,Y,1)^T$ 表示平面点的齐次坐标,式(5)可以写为:

$$\begin{bmatrix}x\\y\\1\end{bmatrix}=s\boldsymbol{M}[r_1\ r_2\ t]\begin{bmatrix}X\\Y\\1\end{bmatrix} \tag{6}$$

记 $H=s\boldsymbol{M}[r_1\ r_2\ t]$,则式(6)可以表示为:

$$q=\boldsymbol{H}Q,\ \boldsymbol{H}=\begin{bmatrix}h_1&h_2&h_3\\h_4&h_5&h_6\\h_7&h_8&1\end{bmatrix} \tag{7}$$

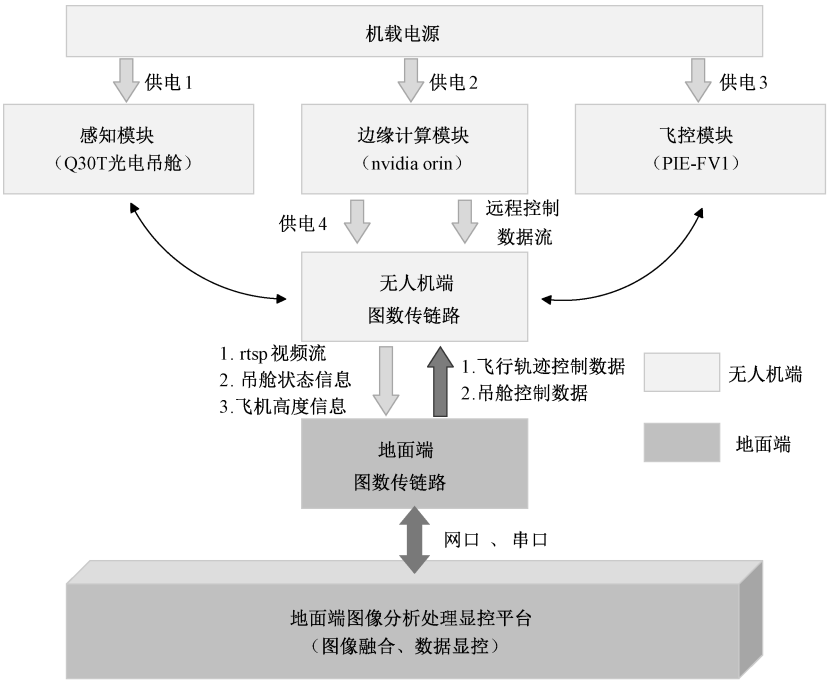


图 4　无人机对地目标图像融合总体架构

Fig. 4　Overall architecture of UAV to ground target image fusion

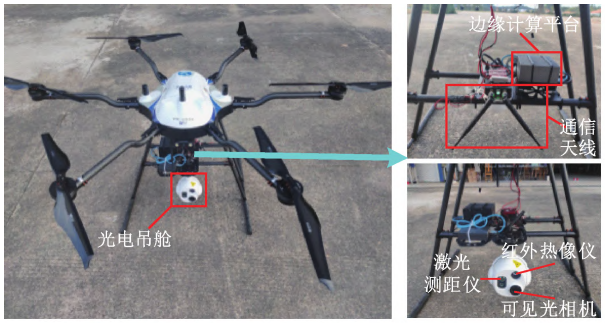


图 5　搭建的开源无人机硬件平台

Fig. 5　UAV hardware platform

其中,$\boldsymbol{H}$ 称为单应矩阵。单应矩阵 $\boldsymbol{H}$ 有 8 个自由度,至少需要 4 对不共线的匹配点才能计算得到。通过 $\boldsymbol{H}$ 的计算可以得出,两个图像像素点之间的对应关系与相机内参数、图像旋转和平移参数相关。因此,实验过程保持光电吊舱的内外参不变能够保证单应矩阵的不变性。

之后,利用无人机搭载的光电吊舱对随机摆放的地面目标从不同高度和角度进行拍摄,构建了配准的可见光与红外图像地面目标数据集。值得注意的是,为了验证算法对浓烟场景可见光与红外图像的融合效果,本实验通过发烟罐遮挡地面目标以模拟典型浓烟场景。通过人工去重、删除低质量图像等步骤建立地面典型目标数据集共 32 对可见光与红外图片。如图 6 所示,为构建的地面目标可见光与红外图像对示例,包含阴天、晴天等多种场景,覆盖伪装车辆、伪装帐篷和小型固定翼无人机等多种类型目标。

图 6　可见光与红外图像典型场景数据集

Fig. 6　Typical scene dataset of visible and infrared images

2.3　实验结果与分析

为了证明本文提出的方法用于无人机对地目标可见光与红外图像融合的有效性,在地面典型场景目标数据集与其他最先进的方法进行全面的比较。之后将模型部署到无人机平台,验证提出的算法在典型场景对地目标可见光与红外图像融合的效果。

1)实验设置

实验的硬件环境为配备 AMD Ryzen Threadripper PRO 5995WX CPU,Tesla A100 GPU 和 128 G 内存;软件环境为 Ubuntu 20. 04 系统,Pytorch 1. 8. 0 深度学习框架

和 CUDA 版本为 11.7 的小型深度学习工作站。用于训练和测试的图像被裁剪为 640×480 的尺寸,并在空间上严格配准。使用 Adam 优化器[23],批量大小设置为 16,初始学习率设为 0.001,网络总共训练迭代 20 轮。为了保证对比实验的一致性,所有对比方法的网络均未加载预训练模型。

2)评价指标

评价指标包括图像融合质量的评价和算法推理效率的评价。图像融合质量评价包括结构相似性指数测量(structure similarity index measure, SSIM)、信息熵(entropy, EN)、空间频率(spatial frequency, SF)和视觉信息保真度(visual information fidelity for fusion, VIF)等 4 个常用的图像融合评价指标[24]对算法的性能进行评估。这些评价指标从结构相似度,信息理论,图像特征和人眼视觉感知等角度全面的分析融合图像的质量。具体来说,SSIM 是感知模型,更符合人眼的直观感受,可以衡量图片的失真程度,定义如下:

$$SSIM(x,y)=\frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)} \tag{8}$$

其中,C_1 和 C_2 是常数,μ_x 和 μ_y 表示平均灰度衡量的亮度。σ_x 和 σ_y 表示通过灰度标准差衡量的对比度。

信息熵从信息论的角度反映图像信息的丰富程度,是度量图像包含信息量多少的客观评价指标。通常情况下,图像信息熵越大,信息量就越丰富,质量越好,计算公式为:

$$H(A)=-\sum_a P_A(a)\log p_A(a) \tag{9}$$

其中,a 表示灰度值,$P(a)$ 表示灰度概率分布。

空间频率反映图像灰度的变化率,空间频率越大表示图像越清晰,融合图像质量越好,计算公式如下:

$$\begin{aligned} SF&=\sqrt{RF^2+CF^2}\\ RF&=\sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}|H(i,j)-H(i,j-1)|^2}\\ CF&=\sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}|H(i,j)-H(i-1,j)|^2} \end{aligned} \tag{10}$$

其中,M,N 为图片的宽高;$H(i,j)$ 表示图像在(i,j)位置的像素值。

VIF 是基于视觉信息保真度提出的衡量融合图像质量的指标,VIF 与主观视觉具有较高的一致性,数值越大则图像质量越好,计算公式如下:

$$\mathrm{VIF}=\frac{\sum_{j\in \text{subbands}} I(\vec{C}N,j;\vec{F}N,j\,|\,s^{N,j})}{\sum_{j\in \text{subbands}} I(\vec{C}N,j;\vec{E}N,j\,|\,s^{N,j})} \tag{11}$$

VIF 的取值范围是[0,1],E 和 F 分别代表视觉失真通道和图像失真通道获取的最终信息。

3)消融实验

(1)非线性激活函数层的有效性。为了证明 DenseNet 非线性激活函数层的重要性和效果,开展了非线性激活函数层的消融实验,实验结果如表 3 所示。可以观察到,去除非线性激活函数层后,模型的评价指标出现明显的下降,去除非线性激活函数层前后的可见光与红外图像融合效果可视化对比如图 7 所示。第一行和第二行分别表示没有和有非线性激活函数层模型的结果。可以看出,去除非线性激活函数层后,融合图像的亮度明显偏暗,红外特征不够明显,说明非线性激活函数层对模型融合具有重要贡献。

表 3 非线性激活函数层的消融结果

Table 3 Ablation results of the nonlinear activation layers

方法	SSIM	EN	SF	VIF
没有非线性激活函数层	0.883	6.933	0.039	0.765
本文方法	**0.884**	**7.027**	**0.043**	**0.782**

图 7 非线性激活函数层的消融结果可视化

Fig. 7 Visualization of the ablation results of nonlinear activation function layers

(2)知识蒸馏 DenseNet 结构自适应能力。为了证明本文设计的自适应优化策略能够挑选出最优的密集层 n 和输出通道数 m,进行了网络结构自适应优化策略的消融实验,实验结果如表 4 所示。表中为不同的(n,m)值对应的图像融合效果和推理性能,当$(n=4,m=8)$时获得了最好的可见光与红外融合质量和推理效率的平衡。

表 4 网络自适应优化策略的消融结果

Table 4 Ablation results for network adaptive optimization strategies

方法(n,m)	SSIM	EN	SF	VIF	推理时间/ms
(2,4)	0.883	7.008	0.043	0.777	**0.668**
(5,16)	**0.892**	7.025	0.043	0.768	1.117
(4,8)本文方法	0.884	**7.027**	**0.043**	**0.782**	0.948

上述实验充分说明本文针对 DenseNet 自适应优化策略的有效性,能够获得可见光与红外图像良好的融合效果和高效的推理效率。

4)与其他最先进方法的对比实验

(1)融合结果定量分析。为了测试知识蒸馏自适应 DenseNet 模型的有效性和可见光与红外图像融合的效果,与其他 4 种流行的融合方法及相应的蒸馏模型在构建的测试数据集上进行了性能评估,这些方法包括 DenseFuse[25]、SeAFusion[22]、SwinFusion[26] 和 (Y-shape dynamic transformer, YDTR[27])。不同方法在测试集的定量分析结果如图 8 所示。从图中可以看出,本文使用知识蒸馏获得的融合结果分布与原始模型的分布相似。具体来说,对于 VIF 评价指标,拟合 SeAFusion 融合算法在大多数图像对实现了最佳性能,这表明提出的知识蒸馏方法在拟合 SeAFusion 算法时能够保持与主观视觉高度的一致性。这是因为本文提出的方法不仅能够拟合 SeAFusion 算法的结果,而且还可以从输入的可见光与红外图像对获取信息。

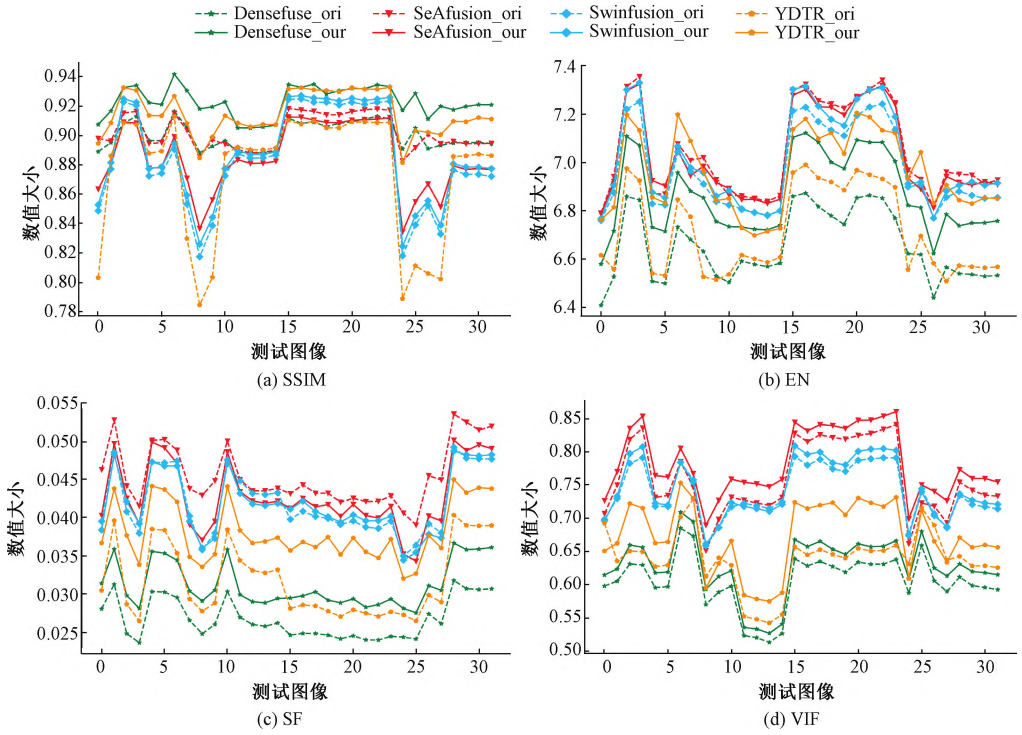


图 8　定量分析结果,原始模型以后缀“_ori”表示,知识蒸馏模型以后缀“_our”表示

Fig. 8　Quantitative analysis results, indicated by the suffix “_ori” for the original model and the suffix “_our” for the knowledge distillation model

对于 SSIM 指标,SwinFusion 知识蒸馏获得的结果与原始模型的结果相似度较高,并且整体表现良好。该实验结果证明了本文设计方法的有效性,能够高度拟合其他模型得到的融合结果。与原始模型不同的是,SwinFusion 的融合策略是常规策略,局限于特定场景应用。相比之下,本文提出的知识蒸馏模型应用场景限制少,具有很强的场景泛化能力。此外,对于信息论的评价指标 EN 和图像特征的评价指标 SF,原始融合结果和知识蒸馏融合结果整体分布一致性较高。对于 Densefuse、SwinFusion 和 YDTR 模型,知识蒸馏得到的融合结果比原始结果更好,本文设计的模型更加轻量化,具有更强的泛化性和更加广阔的应用前景。上述实验结果表明本文提出的自适应蒸馏模型用于可见光与红外融合的有效性。

表 5 为不同融合算法原始融合结果和知识蒸馏结果不同评价指标的平均分数,红色加粗表示排名第一,蓝色粗体数字表示排名第二。从表中可以看出,本文提出的知识蒸馏方法有效拟合了不同模型的融合结果,与原始模型相比,不同评价指标具有相似甚至更好的性能。具体来说,Densefuse 蒸馏模型的 SSIM 值为 0.923,超过原始模型的 0.9,这是因为本文提出的方法可以最大限度地保留不同卷积层从可见光与红外图像提取的特征。SeAFusion 算法获得的原始模型和蒸馏模型在 EN、SF 和 VIF 评价指标取得了排名前二的结果,这是因为原始模型能够实现良好的融合效果,本文构建的知识蒸馏模型能够有效的拟合原始模型的融合结果。尽管蒸馏模型的 EN 和 SF 评价指标不如原始模型,但差异非常小,本文提

出的蒸馏模型在训练时间、推理效率和模型复杂度方面性能远好于原始模型。综合上述实验可以预见,若提供更好的融合结果(用于监督训练的软标签)进行拟合,知识蒸馏模型将获得更优异的性能。此外,对于其他 3 种方法,蒸馏模型比原始模型获得更好的评价指标,这是因为自适应 DenseNet 模型可以更好地反映融合图像的浅层图像特征和原始像素特征,从而在图像的对比度和纹理特征方面具有更好的性能。

表 5 不同模型对应的测试结果

Table 5 Test results corresponding to different models

方法	SSIM	EN	SF	VIF
Densefuse_ori	0.900	6.651	0.027	0.605
SeAFusion_ori	0.902	**7.041**	**0.046**	**0.759**
SwinFusion_ori	0.884	6.970	0.042	0.740
YDTR_ori	0.875	6.712	0.032	0.639
Densefuse_our	**0.923**	6.862	0.031	0.628
SeAFusion_our	0.884	**7.027**	**0.043**	**0.782**
SwinFusion_our	0.887	7.009	0.042	0.745
YDTR_our	**0.915**	6.959	0.038	0.673

(2)模型大小和推理性能分析。测试集的定量分析结果可以得出,本文提出的方法能够很好的拟合给定的融合图像。为了进一步说明构建的知识蒸馏模型的参数量和推理性能远优于原始算法,表 6 对比了 4 种不同融合模型获得的相应自适应 DenseNet 的模型权重大小和推理时间。从表中可以看出,通过知识蒸馏生成的自适应 DenseNet 显著减少了推理时间和模型参数大小。具体来说,SwinFusion 实现了最大程度的推理效率提升和网络复杂度降低,知识蒸馏模型推理时间是原始算法的 0.000 57 倍,模型参数变为原来的 0.002 倍,取得了比原始算法更佳的融合效果。Densefuse 和 YDTR 的知识蒸馏模型推理时间和模型大小也发生了显著的变化,它们同样取得了比原始算法更好的融合效果。SeAFusion 的知识蒸馏模型推理时间仅为原始算法的 0.238 倍,模型仅为原始算法的 0.12 倍,具有多场景应用的泛化性,更轻量化的网络结构和更快的推理性能有助于部署到无人机平台实现机载边缘端的实时处理。此外,与原始模型相比,本文提出的方法收敛速度更快,通过知识蒸馏搜索模型最优结构的方法可以在 20 轮内收敛到损失函数的稳定值。此外,当 $s \in [1,16]$ 时,最多需要 5 次搜索即可找到每个最佳 (n,m) 组合,即相应的最佳自适应 DenseNet 结构。

表 6 原始融合算法及相应的自适应知识蒸馏 DenseNet 的模型大小和推理时间

Table 6 Model size and inference time for the original fusion algorithm and the corresponding adaptive knowledge distillation DenseNet

模型	原始算法		本文算法(自适应 DenseNet)				性能比值大小	
	推理时间/ms	模型大小/KB	n	m	推理时间/ms	模型大小/KB	推理时间比值	模型大小比值
Densefuse	3	296	2	8	0.7	42	0.233	0.14
YDTR	63	873	2	8	0.7	42	0.011	0.048
SwinFusion	7 920	54 025	5	8	1.1	98	0.000 57	0.002
SeAFusion	4	667	4	8	0.95	77	0.238	0.12

(3)定性分析。为了更直观的展示本文提出的算法对可见光与红外图像融合效果的有效性,比较了 4 种不同原始融合模型方法与相应的知识蒸馏模型获得的可见光与红外融合图像的对比度和纹理细节。如图 9 所示,为不同模型对典型场景地面目标可见光与红外图像融合的测试结果。观察 SeAFusion 原始模型和蒸馏模型获得的结果可以得出,蒸馏模型在图像细节和红外产生的伪影处理方面具有更好的效果,显示出更高的整体对比度,更符合人眼视觉感知。本文设计的蒸馏模型能够最大限度地保留不同卷积层从可见光与红外图像提取的特征。

(4)局限性分析。尽管将 SeAFusion 作为教师模型能够获得良好的可见光与红外图像融合效果,但本文设计的自适应知识蒸馏模型需要教师模型推理得到的融合图像作为软标签对模型进行训练。软标签质量的好坏关系到自适应知识蒸馏模型融合的效果。如图 9 所示,对于 Densefuse、SwinFusion 和 YDTR 方法,由于训练集和实验场景差异较大,原始模型不具有不同场景融合的泛化性,它们无法得到理想的融合结果作为知识蒸馏自适应 DenseNet 的软标签,因此本文提出的知识蒸馏模型无法得到理想的可见光与红外图像融合结果。然而,可见光与红外图像融合往往无法得到真值标签,如何得到高质量的软标签是本文算法融合效果好坏的关键。

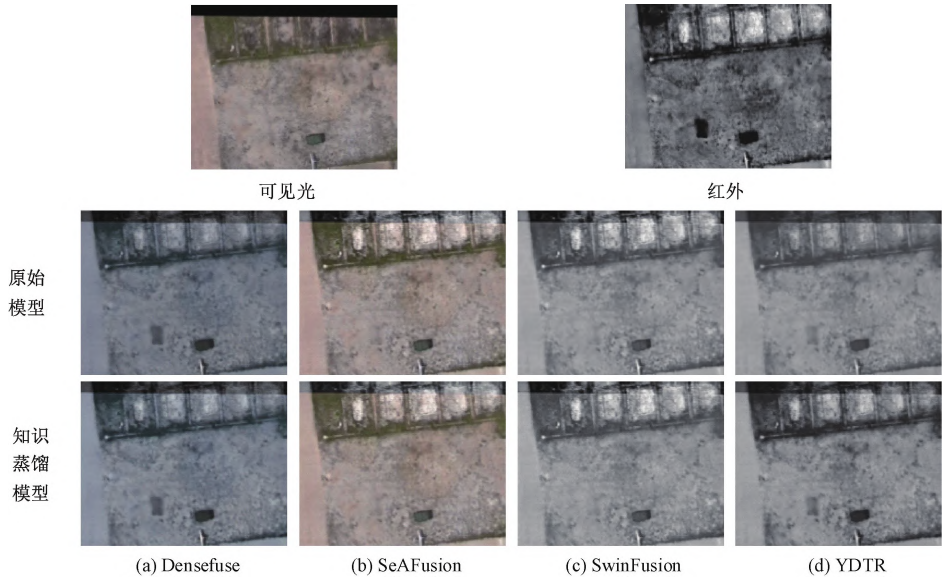


图 9　不同融合算法原始模型与知识蒸馏模型得到的融合结果

Fig. 9　Fusion results obtained from the original model of different fusion algorithms and the knowledge distillation model

总的来说,本文提出的知识蒸馏自适应 DenseNet 方法能够有效拟合原始融合模型的性能,并以更快的推理效率、更轻量化的网络结构实现高质量的可见光与红外图像融合。

3　典型场景对地目标的可见光与红外融合应用

本文将算法搭载在无人机平台开展典型场景对地面伪装车辆、伪装指挥所、各种不同尺寸的小型无人机等目标进行可见光与红外图像融合的实验验证。算法部署和验证均在无人机载 NVIDIA Orin 计算平台进行。该边缘计算平台具有丰富的 GPU 和 CPU 计算资源,并且具有良好的散热性能和较低的功耗,已经广泛应用于高效实时的边缘数据处理和自主检测等计算机视觉任务。图 10 展示了本实验部署在无人机平台的 NVIDIA Orin 边缘计算平台的实物图。

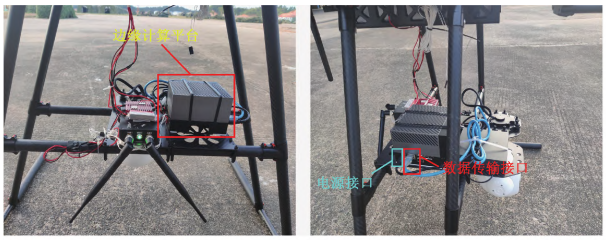
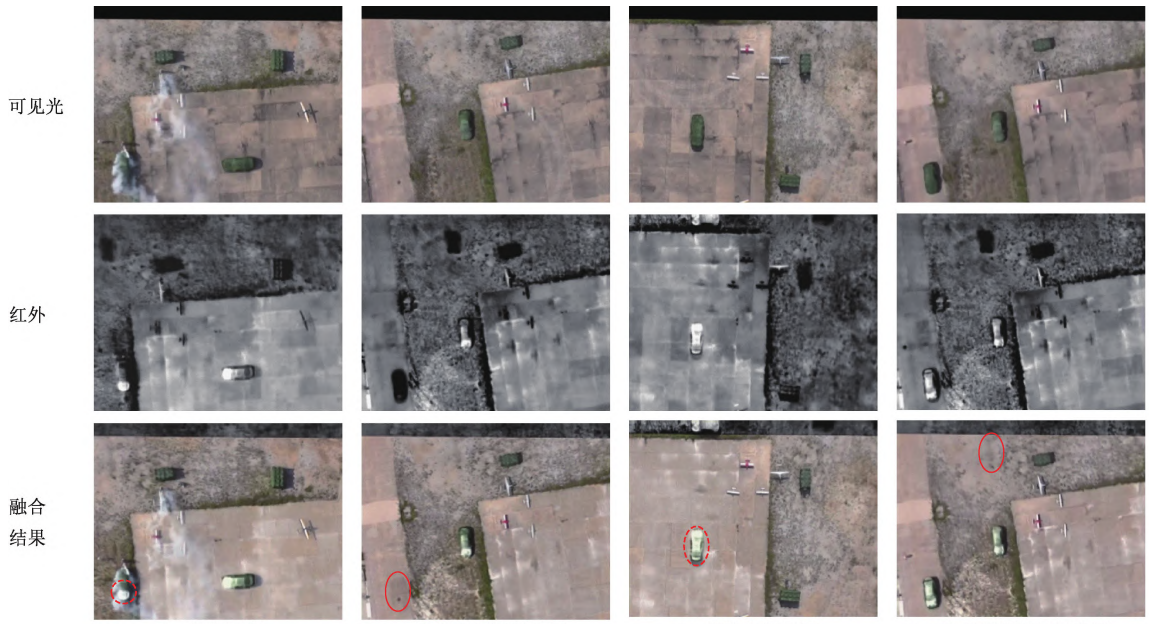


图 10　NVIDIA Orin 边缘计算平台实物

Fig. 10　NVIDIA Orin edge computing platform

NVIDIA Orin 通过网口读取可见光与红外载荷的图像数据,然后对获取的可见光与红外图像进行融合处理。得益于本文设计轻量化的自适应网络结构和边缘计算平台的优异性能,算法处理速度达到了 28 帧每秒,实现了实时的可见光与红外图像融合。如图 11 所示为无人机在典型场景对地面不同类型目标的可见光与红外图像融合结果。从图中可以看出,本文提出的算法能够在多种场景实现地面不同类型的目标高效融合。

图 11(a)可见光图像受到了浓烟干扰的影响,红外图像未受到干扰,从融合结果可以看出,本文提出的方法将伪装车辆的红外热信息融合到可见光图像,从而增强了浓烟场景下伪装车辆目标的可鉴别性(图中虚线圆圈标注区域)。由于伪装车辆和伪装指挥所长时间停放,图 11(b)和(d)中红外图像出现对应目标的伪影,但可见光图像能够准确观察地面目标。因此,本文提出的方法获得的融合结果能够去除红外伪影可能带来的虚警现象(图中实线圆圈标注区域)。图 11(c)为遮盖伪装网的车辆目标,算法将红外热信息融合到可见光图像,从融合结果可以看出,融合红外信息后的伪装车辆目标更具判别性,与伪装指挥所目标具有更加明显的可区分性(图中虚线圆圈标注区域)。



(a) 浓烟干扰
(a) Smoke interference

(b) 红外车辆目标伪影
(b) Infrared vehicle target artefacts

(c) 伪装车辆目标
(c) Camouflaged vehicle targets

(d) 红外指挥所目标伪影
(d) Infrared command post target artefacts

图 11 典型场景无人机对地目标可见光与红外图像融合结果

Fig. 11 Typical scenario UAV fusion results of visible and infrared images of ground targets

4 结 论

本文提出了一种可见光与红外图像融合的新方法，该方法将可见光与红外融合问题建模为神经网络拟合问题。首先，设计了知识蒸馏的自适应 DenseNet 从预先存在的融合模型中学习知识，通过使用模型结构的超参数（包括宽度和深度）实现可见光和红外图像的融合。其次，考虑无人机载边缘端算力和功耗的要求，设计了两个调节模型复杂度的变量（n，m）获得轻量化和高质量融合效果的 DenseNet 网络结构。当模型的复杂度参数设置为（$n=4$，$m=8$），只包含约 77 KB 参数。最后，将算法部署到 NVIDIA Orin 机载边缘计算平台，可对分辨率大小为 640×480 的图像实现每秒 28 帧典型场景的可见光与红外融合。本文构建了首个无人机视角对地目标典型场景的可见光与红外图像数据集，为无人机开展可见光与红外图像融合提供数据支撑。

实验结果表明，本文算法对典型场景的可见光与红外图像具有优异的融合效果和良好的泛化性能，轻量化的网络结构易于部署到各种类型的边缘计算平台，实现高效的推理速度。需要指出的是，制约该算法融合性能的主要原因是缺少理想的可见光与红外融合结果作为模型监督训练必需的软标签。本文提出的方法为可见光与红外图像融合提供了新的思路和潜在的途径，未来的研究仍需继续探索其它超参数对融合模型的知识蒸馏性能的影响，聚焦可见光与红外图像提高目标检测性能等高级视觉任务的融合方式，更好提升无人机平台环境感知能力和自主识别能力。

参考文献

[1] 彭继慎，孙礼鑫，王凯，等. 基于模型压缩的 ED-YOLO 电力巡检无人机避障目标检测算法[J]. 仪器仪表学报，2021，42(10)：161-170.
PENG J SH，SUN L X，WANG K，et al. ED-YOLO power inspection UAV obstacle avoidance target detection algorithm based on model compression [J]. Chinese Journal of Scientific Instrument，2021，42(10)：161-170.

[2] 方鑫，朱婧，黄大荣，等. 低 SNR 场景下微型无人机跟踪-检测融合方法[J]. 仪器仪表学报，2022，43(4)：79-88.
FANG X，ZHU J，HUANG D R，et al. Integrated tracking and detection of micro-UAV under low SNR environment [J]. Chinese Journal of Scientific Instrument，2022，43(4)：79-88.

[3] 吴立珍，李宏男，牛轶峰. 无人机小样本条件下遮挡和

混淆目标识别方法[J]. 国防科技大学学报, 2022, 44(4): 13-21.
WU L ZH, LI H N, NIU Y F. Occlusion and confusion targets recognition method for UAV under small sample conditions[J]. Journal of National University of Defense Technology,2022, 43(4): 13-21.

[4] 陈卓,方明,柴旭,等. 红外与可见光图像融合的 U-GAN 模型[J]. 西北工业大学学报, 2020, 38(4): 904-912.
CHEN ZH,FANG M,CHAI X,et al. U-GAN model for infrared and visible images fusion [J]. Journal of Northwestern Polytechnical University, 2020, 38(4): 904-912.

[5] 杨艳春,高晓宇,党建武,等. 基于 WEMD 和生成对抗网络重建的红外与可见光图像融合[J]. 光学 精密工程, 2022, 30(3): 320-330.
YANG Y CH,GAO X Y,DANG J W, et al. Infrared and visible image fusion based on WEMD and generative adversarial network reconstruction [J]. Optics and Precision Engineering,2022,30(3): 320-330.

[6] 闵莉,曹思健,赵怀慈,等. 改进生成对抗网络实现红外与可见光图像融合 [J]. 红外与激光工程, 2022, 51(4): 405-414.
MIN L,CAO S J,ZHAO H C, et al. Infrared and visible image fusion using improved generative adversarial networks [J]. Infrared and Laser Engineering, 2022, 51(4): 405-414.

[7] 胡建平,郝梦云,杜影,等. 结构和纹理感知的 Retinex 融合红外与可见光图像[J]. 光学 精密工程, 2022, 30(24): 3225-3238.
HU J P, HAO M Y, DU Y, et al. Fusion of infrared and visible images via structure and texture-aware Retinex[J]. Optics and Precision Engineering, 2022, 30(24): 3225-3238.

[8] LEWIS J J, O'CALLAGHAN R J, NIKOLOV S G, et al. Pixel-and region-based image fusion with complex wavelets [J]. Inf Fusion, 2007, 8: 119-30.

[9] ZHANG Q, LIU Y, BLUM R S, et al. Sparse representation based multi-sensor image fusion for multi-focus and multi-modality images: A review [J]. Inf Fusion, 2018, 40: 57-75.

[10] ZHANG Q, LI G, CAO Y, et al. Multi-focus image fusion based on non-negative sparse representation and patch-level consistency rectification [J]. Pattern Recognition, 2020, 104: 107325.

[11] CHEN J, WU K, CHENG Z, et al. A saliency-based multiscale approach for infrared and visible image fusion [J]. Signal Process, 2021, 182: 107936.

[12] MA J, TANG L, XU M, et al. STDFusionNet: An infrared and visible image fusion network based on salient target detection [J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1-13.

[13] CAI H, LIRAN Z, CHEN X, et al. Infrared and visible image fusion based on BEMSD and improved fuzzy set[J]. Infrared Physics & Technology, 2019, 98: 201-211.

[14] YIN W, HE K, XU D, et al. Adaptive enhanced infrared and visible image fusion using hybrid decomposition and coupled dictionary [J]. Neural Computing and Applications, 2022, 34: 20831-20849.

[15] AN W, WANG H. Infrared and visible image fusion with supervised convolutional neural network [J]. Optik, 2020, 219: 165120.

[16] LI J, HUO H, LI C, et al. AttentionFGAN: Infrared and visible image fusion using attention-based generative adversarial networks [J]. IEEE Transactions on Multimedia, 2021, 23: 1383-96.

[17] MA J, YU W, LIANG P, et al. FusionGAN: A generative adversarial network for infrared and visible image fusion [J]. Inf Fusion, 2019, 48: 11-26.

[18] WANG Z, CHEN Y, SHAO W, et al. SwinFuse: A residual swin transformer fusion network for infrared and visible images [J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 1-12.

[19] TANG W, HE F, LIU Y. TCCFusion: An infrared and visible image fusion method based on transformer and cross correlation[J]. Pattern Recognition, 2023, 137: 109295.

[20] MIRZADEH S I, FARAJTABAR M, LI A, et al. Improved knowledge distillation via teacher assistant[C]. Proceedings of the AAAI conference on artificial intelligence. 2020, 34(4): 5191-5198.

[21] ZHAO Z, SU S, WEI J, et al. Lightweight infrared and visible image fusion via adaptive DenseNet with knowledge distillation [J]. Electronics Newsweekly, 2023, 12(13): 2773.

[22] TANG L, YUAN J, MA J. Image fusion in the loop of high-level vision tasks: A semantic-aware real-time infrared and visible image fusion network [J]. Inf Fusion, 2022, 82: 28-42.

[23] KINGMA D P, BA J. Adam: A method for stochastic optimization [J]. Arxiv preprint Arxiv: 1412.6980, 2014.

[24] ZHANG X. Deep learning-based multi-focus image fusion: A survey and a comparative study [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44: 4819-38.

[25] LI H, WU X. DenseFuse: A fusion approach to infrared and visible images [J]. IEEE Transactions on Image Processing, 2018, 28: 2614-23.

[26] MA J, TANG L, FAN F, et al. SwinFusion: Cross-domain long-range learning for general image fusion via swin transformer[J]. IEEE/CAA Journal of Automatica Sinica, 2022, 9(7): 1200-1217.

[27] TANG W, HE F, LIU Y. YDTR: Infrared and visible image fusion via Y-shape dynamic transformer [J]. IEEE Transactions on Multimedia, 2023, 25: 5413-28.

作者简介

童小钟,分别于 2018 年和 2021 年于国防科技大学获得学士学位和硕士学位,正在国防科技大学攻读博士学位,主要研究方向为智能侦测和目标感知。

E-mail: tongxiaozhong@nudt.edu.cn

Tong Xiaozhong received his B. Sc. and M. Sc. degree in 2018 and 2021 both from National University of Defense Technology. Now, he is a Ph. D. candidate in National University of Defense Technology. His main research interests include intelligent detection and target perception.

赵宗庆(通信作者),分别于 2021 年和 2023 年于国防科技大学获得学士学位和硕士学位,正在国防科技大学攻读博士学位,主要研究方向为智能侦测和目标感知。

E-mail: zhaozongqing17@nudt.edu.cn

Zhao Zongqing (Corresponding author) received his B. Sc. and M. Sc. degree in 2021 and 2023 both from National University of Defense Technology. Now, he is a Ph. D. candidate in National University of Defense Technology. His main research interests include intelligent detection and target perception.