引用格式：HAO Yongping, CAO Zhaorui, BAI Fan, et al. Research on Infrared Visible Image Fusion and Target Recognition Algorithm Based on Region of Interest Mask Convolution Neural Network[J]. Acta Photonica Sinica, 2021, 50(2):0210002
郝永平,曹昭睿,白帆,等.基于兴趣区域掩码卷积神经网络的红外-可见光图像融合与目标识别算法研究[J].光子学报,2021,50(2):0210002

基于兴趣区域掩码卷积神经网络的红外-可见光图像融合与目标识别算法研究

郝永平[1],曹昭睿[1],白帆[1],孙颢洋[1],王兴[2],秦洁[1]
(1 沈阳理工大学 装备工程学院,沈阳 110159)
(2 沈阳理工大学 机械工程学院,沈阳 110159)

摘　要:建立权重独立的双通道残差卷积神经网络,对可见光与红外频段下的目标图像进行特征提取,生成多尺度复合频段特征图组。基于像点间的欧式距离计算双频段特征图显著性,根据目标在不同成像频段下的特征贡献值进行自适应融合。通过热源能量池化核与视觉注意力机制,分别生成目标在双频段下的兴趣区域逻辑掩码并叠加在融合图像上,凸显目标特征并抑制非目标区域信息。以端到端识别网络作为基础,利用交叉损失计算策略,对含有注意力掩码的多尺度双频段融合特征图进行目标识别。结果表明,所设计的识别网络能够有效地融合目标红外热源物理特征和可见光图像纹理特征,提高了信息融合深度,保留目标热辐射与纹理特征的同时降低了背景信息干扰,对全天候复杂环境下的多尺度热源目标具有良好的识别精度与鲁棒性。

关键词:卷积神经网络;注意力机制;图像融合;目标识别;多光谱成像
中图分类号:TP391.41　　**文献标识码**:A　　**doi**:10.3788/gzxb20215002.0210002

Research on Infrared Visible Image Fusion and Target Recognition Algorithm Based on Region of Interest Mask Convolution Neural Network

HAO Yongping[1], CAO Zhaorui[1], BAI Fan[1], SUN Haoyang[1], WANG Xing[2], QIN Jie[1]
(1 *College of Equipment Engineering, Shenyang Ligong University, Shenyang* 110159, *China*)
(2 *College of Mechanical Engineering, Shenyang Ligong University, Shenyang* 110159, *China*)

Abstract: A dual channel residual convolution neural network with independent weight is established. The features of target in visible and infrared images is extracted. The multi-scale composite frequency band feature maps are generated. Based on the Euclidean distance between image points, the saliency of each image point in the dual band feature map is calculated. The adaptive fusion is carried out according to the characteristic contribution value of the target in different imaging frequency bands. Through the thermal radiation pooling kernel and visual attention mechanism, the logical mask of the target region of interest under dual frequency band is generated and superimposed on the fusion image to highlight the target features and suppress the non target area. Based on end-to-end identification network and using the cross

基金项目:国防科技基础加强计划技术领域基金(No.2020-JCJQ-JJ-422),“十三五”装备预研兵器工业联合基金(No.6141B012841)
第一作者:郝永平(1960—),男,教授,博士,主要研究方向为光电装备智能探测。Email:yphsit@126.com
通讯作者:曹昭睿(1993—),男,博士研究生,主要研究方向为机器视觉与目标识别。Email: caozhaorui@163.com
收稿日期:2020-08-27;**录用日期**:2020-12-03

loss calculation strategy. The target recognition of multi-scale dual band fusion feature map with attention mask is carried out. The results show that the designed recognition network can effectively integrate the physical characteristics of infrared heat source and the line features of visible image. The depth of information fusion is improved. The thermal radiation and texture features of the target is retained. The interference of background information is reduced. It has good recognition accuracy and robustness for multi-size heat source targets in all-weather and complex environment.

Key words: Convolution neural network; Attention mechanism; Image fusion; Target recognition; Multispectral imaging

OCIS Codes: 100.4996; 150.0150; 110.3080; 100.3008

0 引言

基于图像信息的目标识别技术是机器视觉探测领域的重点研究方向之一。但随着探测环境的不断多元化与复杂化,检测任务对识别算法的抗干扰能力和目标深层特征信息挖掘能力的需求不断提高。现阶段多数识别算法以目标在可见光波段下的成像作为输入数据源,结合特征判据进行图像的全局检测与识别。这种依托可见光作为单一成像波段的识别方法,在低照度、有云雾等恶劣环境下具有较大的局限性,无法有效地获取目标的图像特征信息。同时全局检测方法需要对图像进行逐像素或逐区域的分析,无目标存在的背景图像将影响到识别计算效率与精度。例如在森林火点检测、海上舰艇探测、空中无人机标定、战场环境下装甲目标识别等应用中,待检测目标在可见光波段下的特征极易淹没于背景中,而在夜间和遮挡环境下,可见光几乎无法有效捕捉到目标图像。同时上述应用均有大视场拍摄需求,目标的成像区域在全画幅图像中占比较小,而任务对识别算法的精确性和实时性却有较高的要求,若对全局图像的所有区域进行无差别检测,必将使算法产生较大延迟与误差,进而影响识别效果。所以,设计一种能够提取目标深层特征,并具备高抗干扰能力的高效识别算法,既是当前机器视觉领域的核心需求,也是目标识别领域中极富挑战的研究方向。

为解决上述问题,国内外学者将研究重心转到了多频段图像融合与识别技术上,其中针对红外和可见光图像的融合与识别成为了该领域的重点研究内容。由于大多数待检测物为热源,目标自身的热辐射特征是将其所处环境进行区分的最佳判据之一。红外图像与可见光图像的成像频段覆盖区域较广,能够同时表征出物体多维度的光学特征,且因成像机理不同,位于各自成像域上的图像表征方式具备互补性[1]。对红外图像与可见光图像进行融合识别,可在过滤环境噪声的同时,提高算法针对复杂环境的抗干扰能力,实现全天候探测与识别。这种可以提取目标深层光学特征、并具备良好鲁棒性的图像融合方法,成为了当前多频段图像融合识别技术的首选策略。

目前,国内外研究人员对红外与可见光图像融合识别技术已经开展了部分研究。文献[2]提出了一种双树复小波域内结合区域分割的红外与可见光图像融合方法,将提升了全局对比度的红外图像与可见光图像融合,保持了场景细节信息的同时实现了算法在机载光电平台上的嵌入。文献[3]利用鲁棒主成分分析法(Robust Principle Component Analysis,RPCA)分解出原图像的稀疏分量和低秩分量,随后采用非下采样轮廓波变换(Non-subsampled Contourlet Transform,NSCT)进一步对低秩分量进行处理,并将其与稀疏分量叠加成新的红外-可见光融合图像。文献[4]利用非下采样剪切波变换(Non-subsampled Shearlet Transform,NSST)对红外与可见光图像进行多尺度、多方向的分解,将生成的高频子带和低频子带图像进行融合与逆变换,并在此基础上利用全卷积神经网络(Fully Convolutional Networks,FCN)对融合图像进行识别。上述文献所提出方法的主要思路均是采用不同的过滤器对红外和可见光图像进行分解,并将含有共同特性的特征分量进行像素级或区域融合。文献[5]根据视网膜大脑皮层的图像映射原理,利用红外图像补充可见光暗区图像,以降低融合图像噪声。文献[6]以导向滤波为基础,提出了红外图像辅助可见光降噪,同时增添红外目标纹理的融合方法。这些方法虽然能够对红外和可见光图像进行有效的融合和互补降噪,但在本质结构上均由高低频信号剥离、图像噪声滤波、像素加权融合构成,要求红外和可见光原图像在分辨率、像素匹配、像质、噪声影响程度、位置偏移量上保持高度一致,否则将会产生较大的融合误差。且上述方法对目标的红外成像特征仅在像素值上进行应用,特征利用和挖掘效率不高,在真实应用中局限性较大。

随着深度学习和卷积神经网络技术的日趋成熟,研究人员开始将这种具备强大自学习能力的算法应用在红外与可见光图像融合识别技术中。文献[7]提出了一种基于两级卷积神经网络的融合方法,利用近红外图像恢复去噪后可见光图像的颜色与纹理,弥补了弱光条件下可见光图像的缺失信息。但真实低照度环境下可见光图像所包含的噪声更为复杂,信息缺失程度更大,该方法在恶劣环境下的融合效果不佳。文献[8]提出了结合多光谱尺度不变特征变换(Scale Invariant Feature Transform,SIFT)算子匹配与核分类的双频光谱拓展融合方法,实现可见光与红外图像的匹配,但算法匹配精度无法满足训练需求。文献[9]将近红外图像的高频部分、低频部分与可见光图像进行加权分配后,在矩阵维度上进行连接,传入卷积神经网络进行通道压缩与整合,但该方法仅是对红外与可见光图像在通道上进行堆叠,双频融合信息在空间和语义上的高维特征关系并未得到体现。文献[10]基于卷积神经网络和离散余弦波的方法融合可见光与红外图像,并配合视觉注意力机制(Visual Attention Mechanism,VAM)强化融合图像的纹理细节与清晰度,但VAM注意力机制对非兴趣区域的抑制效果并不明显。文献[11]以VGG网络(Visual Geometry Group Network,VGG-Net)为特征提取网络构架,将像素加权处理和通道叠加后的融合图像作为输入,训练网络内参并进行融合图像的目标分类,但该方法易使内容丰富且维度更高的可见光特征图被低信息量的红外特征图稀释,难以在复杂环境下的目标检测任务中使用。文献[12]利用生成对抗网络将可见光图像向红外频段进行迁移,实现了单频低容量数据样本向双频高容量的扩充,由于该方法未建立热源目标在红外和可见光频段间的真实成像特征关系,可能会生成不符合实际情况的目标红外或可见光图像,在真实目标检测上存在一定风险。上述方法为红外与可见光图像的融合与识别提供了丰富的技术经验,但在目标深度特征提取、复杂环境下高维图像的目标检测与多频段数据融合方面,仍存在一定的缺陷。

针对当前红外可见光图像融合与目标识别技术存在的局限性与不足,以及为满足复杂环境下含有热辐射目标的精确识别需求,本文对目标位于红外和可见光频段下的图像特征进行深入研究,设计了一种含有注意力机制的红外-可见光融合与目标识别双通道卷积神经网络(Infrared Visible Image Fusion Neural Network,IVFNN)。该方法的主要创新点为:1)利用权重分离的双通道深度残差卷积网络分别提取目标在红外和可见光频段下的特征图,生成目标在不同频段下的高维抽象特征;2)以特征语义点间欧氏距离为计算准则,获得目标在各频段特征图中对热源信息与纹理特征的贡献值,依照贡献值大小进行自适应融合;3)利用热辐射池化核与视觉注意力机制,分别在红外与可见光频段下生成对应的目标兴趣区域逻辑掩码,整合后覆盖在多尺度复合频段融合特征图上,生成含有目标位置逻辑语义信息的待识别特征图。基于上述创新点,IVFNN减小了全局图像中非兴趣区域的计算干扰,充分利用并融合了目标在红外频段下的低频轮廓特征与可见光频段下的高频纹理特征,对目标在不同频段下的表现特征进行了最佳选取与保留,提升了复杂环境下的识别计算精度与鲁棒性。

1 算法结构

本文所设计的含有注意力机制的红外-可见光图像融合与目标识别双通道卷积神经网络(IVFNN)由红外可见光图像的双通道特征提取、自适应融合及兴趣逻辑掩码生成、多尺度复合频段目标识别三个核心计算部分构成,IVFNN网络结构如图1所示。首先,同一场景下等目标占比的红外与可见光图像将分别输入到两个权值独立、结构相同的特征提取网络中,获得目标在不同频段下的多尺度高维特征图。随后计算红外与可见光特征图中各像素的全局显著性,获得同一像点在不同频段下对自身识别特征的贡献程度,依照特征贡献值进行自适应图像融合,生成含有最佳高频纹理信息和低频热源信息的融合特征图。在不同尺度的双频特征图组上,利用热辐射池化核与视觉注意力机制生成红外与可见光下的目标兴趣区域逻辑掩码,整合各频段视觉注意力逻辑码并将其叠加在融合图像上,凸显兴趣目标并抑制非目标区域的图像语义信息。最后通过以YOLO V3为构架的目标识别网络进行多尺度融合特征图识别,完成红外与可见光频段下的目标位置与类别确定。

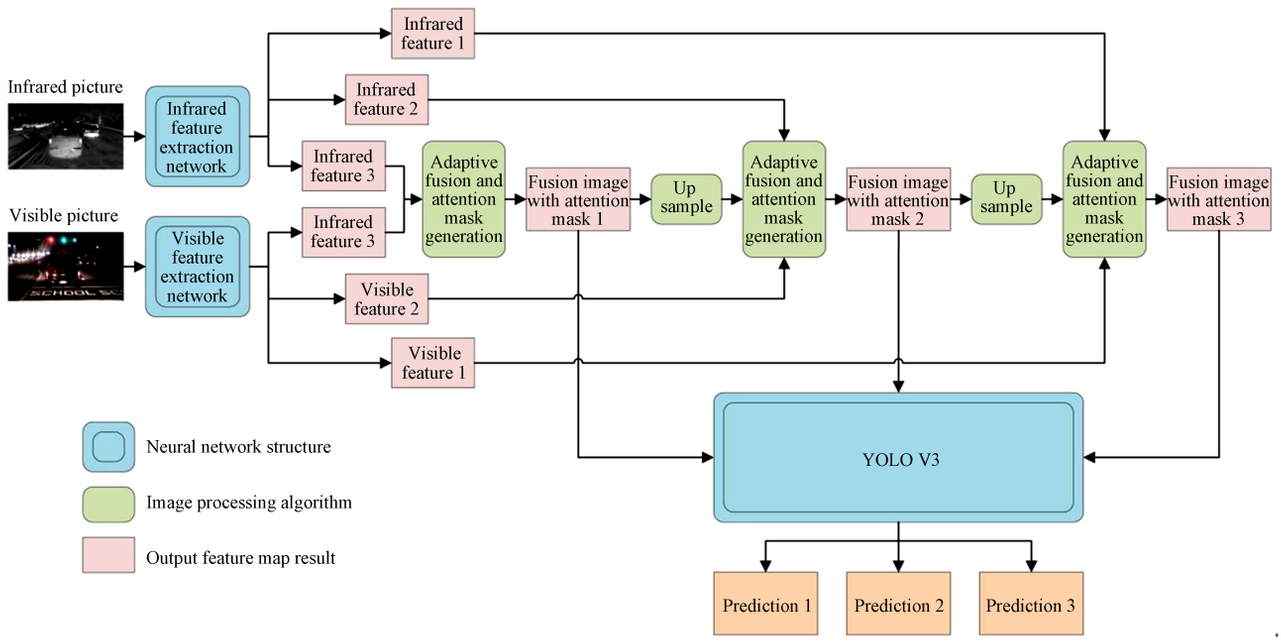

图 1 IVFNN 网络结构示意

Fig.1 Network structure of IVFNN

2 算法实现

2.1 红外–可见光双通道特征提取网络

在进行目标红外与可见光信息融合之前，需要对目标于双频段下的原始图像进行特征提取，获得图像语义信息更为抽象、目标热辐射与纹理信息表征能力更强的特征图。本文以如图 2 所示的深度残差特征提取网络作为 IVFNN 的特征提取网络基础构架。

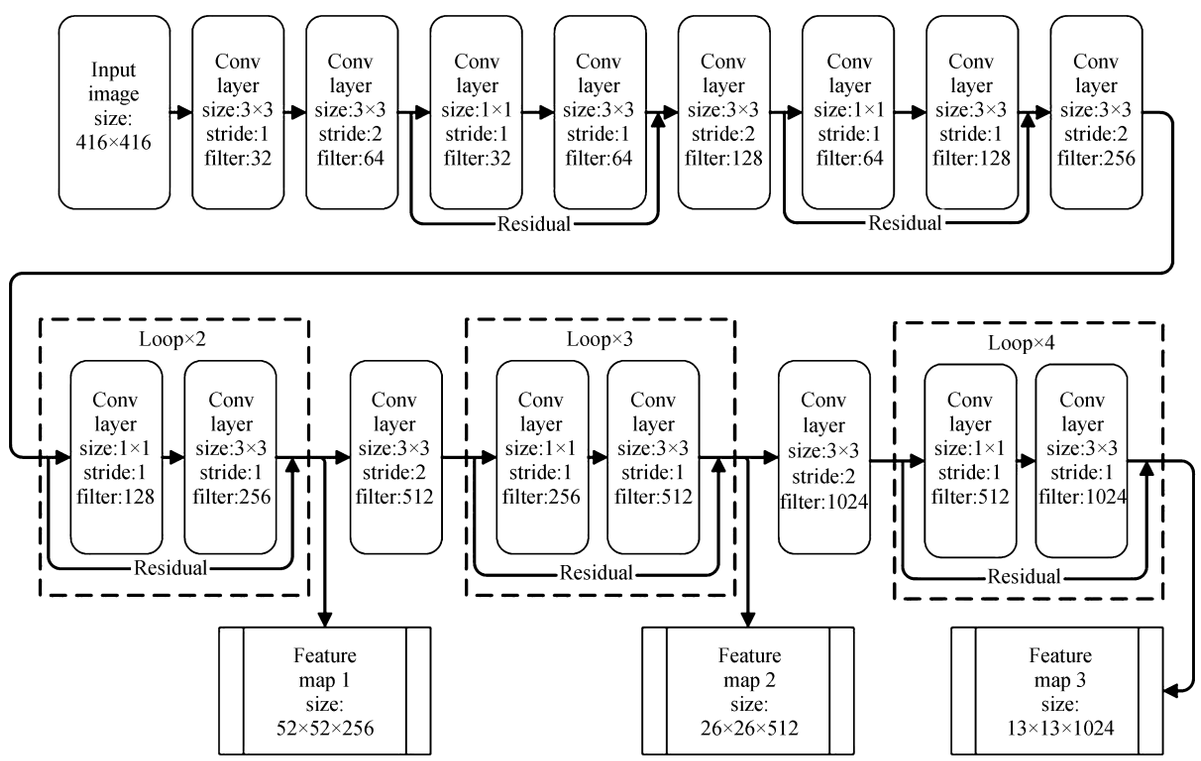

图 2 IVFNN 特征提取网络结构

Fig.2 Feature extraction network structure of IVFNN

特征提取网络由10个独立卷积层与3个残差模块构成，其中每个残差模块内包含两个卷积层。被整形成416×416尺寸的输入图像将分别经过各自的特征提取网络进行下采样计算，并利用loop循环模块与residual残差模块将浅层图像信息与深层语义信息不断堆叠，防止目标低维度纹理与几何特征丢失。每个循环模块完成残差特征叠加后，分别输出尺寸为13×13、26×26与52×52的特征图，该特征图组将作为后续图像融合与兴趣逻辑掩码生成的输入源。其中尺寸较小的特征图主要用于提供语义信息、尺度较大的特征图主要用于提供目标位置信息[13]。

由于红外图像与可见光图像在组成矩阵维度、色彩分布、光学信息响应敏感范围等参数上存在差异，同一目标在不同频段下，表征纹理、轮廓、颜色、热辐射等信息的方式也不同。若将两种不同频段的图像直接融合，再经过同一网络进行特征提取，生成红外-可见光共享识别权值，则会稀释并破坏目标在不同频段上的固有特征。故本文在特征提取网络上采用双通道构架，使红外和可见光图像分别经过两个结构相同、权值和参数不共享的特征提取网络进行训练和识别。

在这一过程中，所生成的红外与可见光特征图分别表示为

$$\begin{cases} F_{\text{inf}} = \{F_{\text{inf}}^{52}, F_{\text{inf}}^{26}, F_{\text{inf}}^{13}\} \\ F_{\text{vis}} = \{F_{\text{vis}}^{52}, F_{\text{vis}}^{26}, F_{\text{vis}}^{13}\} \end{cases} \tag{1}$$

式中，F^{n}_{channel}为某一尺度与频段下，经由特征提取网络生成的特征图组，channel表示该特征图组源自红外(inf)或可见光(vis)频段，n表示该特征图组的尺度。红外与可见光频段将分别生成三组尺度与维度相对应的特征图组，并作为后续图像处理的输入信息源，进行双频自适应融合、兴趣区域逻辑码生成与识别计算。

2.2 红外-可见光高维特征图融合与处理

完成红外与可见光图像的特征提取后，生成的多尺度双频特征图组将作为输入数据，分别进行自适应融合与兴趣区域逻辑码生成。其中自适应融合将以红外和可见光特征图上同一像点在不同频段下的特征贡献值作为融合参考，兴趣区域逻辑码由基于热辐射池化核计算的红外逻辑码和基于视觉注意力机制的可见光逻辑码构成。IVFNN网络的双频段自适应融合与视觉注意力逻辑码计算过程中，图像数据变化与计算流程如图3。

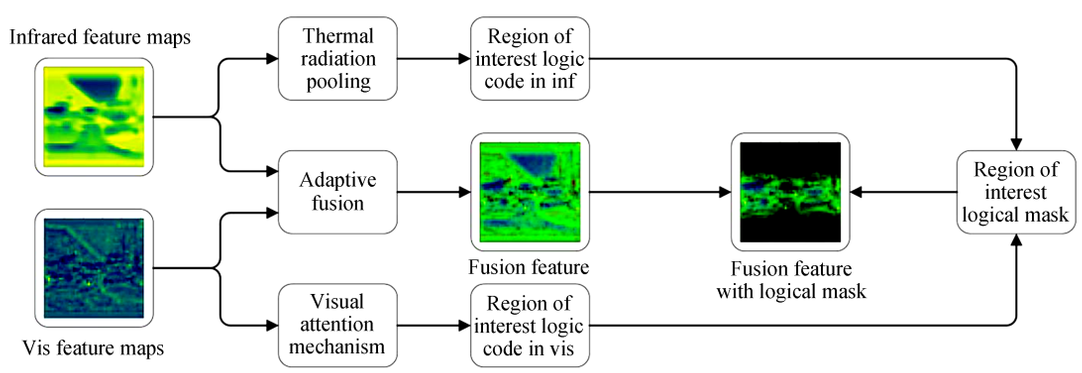

图3 输入图像在双频自适应融合与注意力逻辑码计算过程中变化状态

Fig.3 Input image changes in the process of dual channel adaptive fusion and attention logic code calculation

2.2.1 红外-可见光特征图的自适应融合

在红外和可见光的成像过程中，感光器件像元单位、成像色域分布、目标图像特征表达方式上存在一定差异，即使能够满足双频成像分辨率与图像尺寸的统一，红外与可见光对于同一目标在不同环境影响下的深层图像语义特征表现方式和突出程度也存在不同。当处于如图4(a)所示的光照充足的日间时，目标在可见光频段成像的纹理特征表达量相比红外频段下的温度与轮廓特征表达量要更为清晰与丰富，红外频段的目标热源特征还会因外界环境温度较高而被淡化；相反在如图4(b)所示的光照不充足且热源目标与环境目标相差较大的夜间时，目标的红外频段表征信息相比可见光更为明显。

常规的双频融合算法一般采用加权融合法或范数融合法。加权融合法一般采用人工定义的方式分配不同频段下的融合权重比，因不同环境下同种目标在红外与可见光下的表征清晰程度并不恒定，使得这种权值固定的融合方式无法针对每个像素点生成最佳的融合比，存在全天候环境下双频图像融合特征模糊的

(a) Images in visible and infrared during the daytime

(b) Images in visible and infrared during the nighttime

图4 不同外界环境下红外与可见光的目标特征成像贡献对比

Fig.4 Comparison of imaging contribution of infrared and visible light targets in different environments

缺陷。范数融合法一般直接用于双频原始图像的像素级融合,融合重心为像点的像素值,当融合像点与周围像点存在一定的梯度关系时,这一特征将会被范数融合策略所破坏[14-15]。针对上述问题,本文将分别针对红外和可见光特征图在不同外界环境下的语义特点及其对目标特征表达的贡献值进行判断,采用自适应融合策略,依照目标高维图像信息的表征程度,有选择地从不同频段中提取像素点并分配融合权值,获得当前状态下能够最丰富地表达目标信息的融合特征图。

经过双通道特征提取网络计算后,原始输入红外与可见光图像将转化为某一尺寸的红外与可见光特征图组为 F^{n}_{channel},红外与可见光特征图组自适应融合流程如图5。

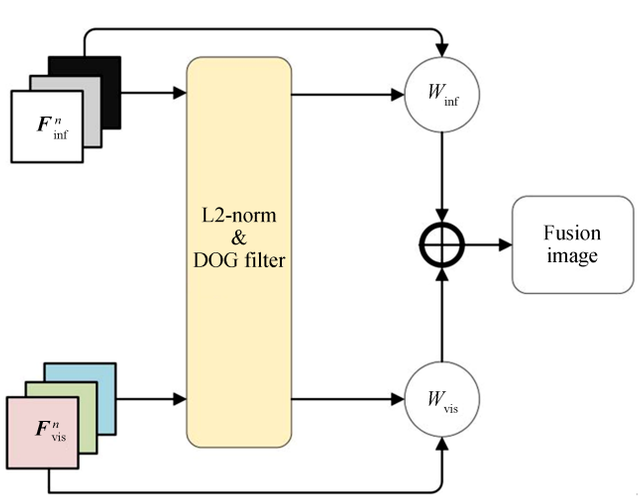

图5 红外与可见光特征图组自适应融合流程

Fig.5 Adaptive fusion process of infrared and visible feature maps group

对 F^{n}_{channel} 内某一维度特征图上各像素点进行全局显著性计算。首先获得该像素点所在维度特征图内的显著值 $S^{n}_{\text{channel}}(x,y)$,表示为

$$S^{n}_{\text{channel}}(x,y)=\left\|I_{\mu}-I_{\text{whc}}\left(F^{n}_{\text{channel}}(x,y)\right)\right\|_{2} \tag{2}$$

式中,I_{μ} 为输入特征图像内像素均值,I_{whc} 为高斯差分滤波后像素值,$\|\ \|_{2}$ 为L2范数,(x,y) 为该特征图内像素

点位置。由于$F^{n}_{\text{channel}}(x,y)$为高维特征图像点，该点与$I_{\mu}$的欧式距离仍保留了深层语义信息与梯度信息，优化了原始图像进行范数融合时导致的非色度特征淡化问题。

为实现双频特征图的自适应融合，根据目标在不同频段上的全局图像显著值，计算目标在当前环境下对于高频纹理信息和低频热辐射信息的贡献值。对于某一尺度下的特征图组，位于不同维度的特征图将使用不同的贡献值进行融合。则该尺度下，同一像点在红外与可见光频段下的特征贡献值W^{n}_{channel}为

$$W^{n}_{\text{channel}}(x,y)=\frac{S^{n}_{\text{channel}}(x,y)}{S^{n}_{\text{vis}}(x,y)+S^{n}_{\text{inf}}(x,y)} \tag{3}$$

该尺度下红外与可见光融合特征图F_{f}为

$$F_{\text{f}}=\sum(W^{n}_{\text{vis}}\times F^{n}_{\text{vis}}+W^{n}_{\text{inf}}\times F^{n}_{\text{inf}}) \tag{4}$$

其中同尺度同维度下的红外与可见光特征图采用add叠加，同尺度不同维度的融合特征图采用concatenate叠加。融合前后各频段语义信息热力特征图如图6。

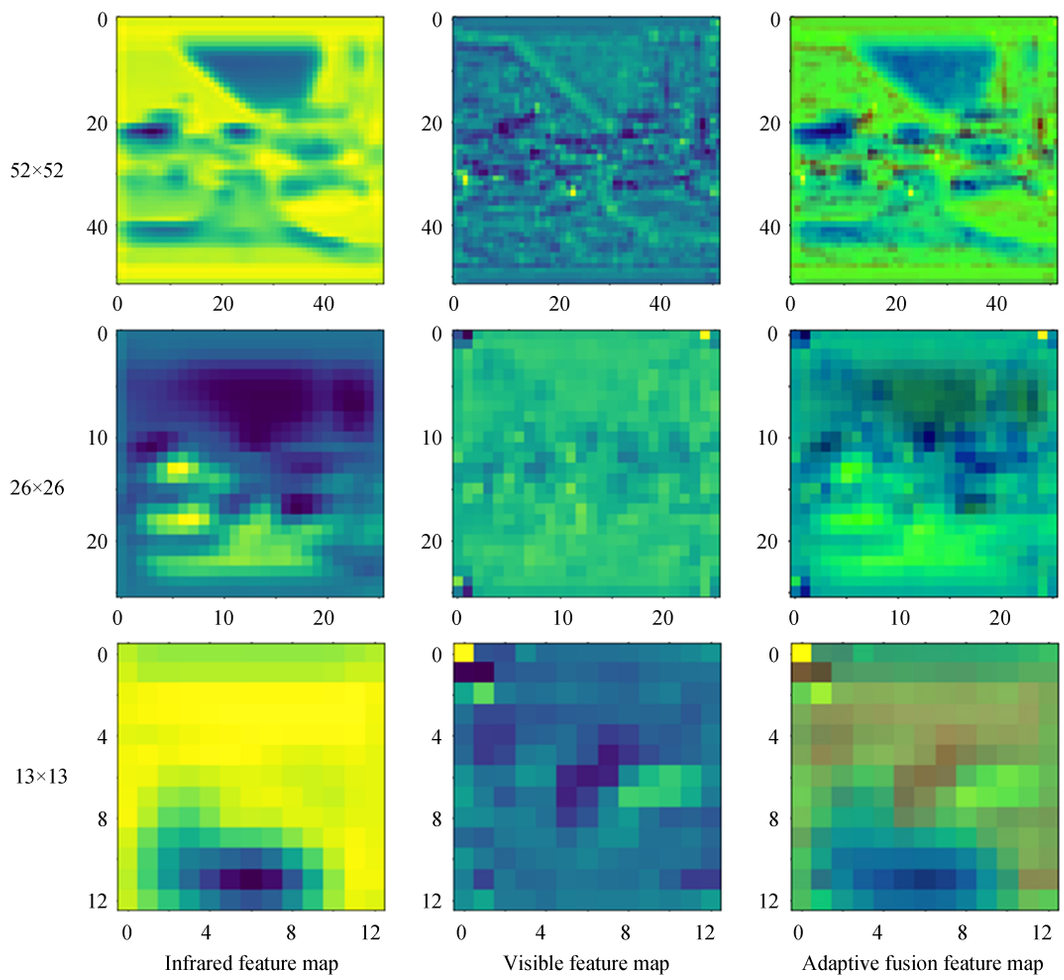

图6　融合前后各频段特征图

Fig.6　Feature maps of each channel before and after fusion

由图6可以看出，融合后的F_{f}相比融合前各频段的特征图，目标表征的高维语义信息更为丰富，对于不同环境下的目标特征信息凸显能力更强。

2.2.2　基于热辐射池化的红外兴趣掩码生成

目标在可见光和红外频段下，会将自身的轮廓特征分别以高频和低频的方式进行投射。而环境背景因温度或纹理信息量较低，导致其轮廓特征在不同频段下投射不明显。由于红外图像主要含有目标的低频轮廓信息，且这一信息由目标上各热源点的热辐射值表现[16-17]。故设计了一种热辐射池化核，通过计算不同像

点的热源辐射情况，选择高热区域作为目标的兴趣识别范围，以过滤环境和非热源区域，进而提升目标识别精度并降低冗余信息计算量。

在红外图像中，各像点根据自身成像色度与其对应目标点的真实热辐射情况，可分为核心能量点（e_{core}）、主动能量点（$e_{\text{initiative}}$）和被动能量点（e_{passive}）。其中核心能量点主动发热且自身热量大于周围的能量点，集中于待识别目标轮廓内；主动能量点主动发热但受到周围温度大于自身的核心能量点影响，分布于待识别目标内部部分低温区；被动能量点自身不发热且热能来自周围其他能量点的辐射贡献，集中于背景中。三种能量点表示方式如式5。

$$\begin{cases} e_{\text{core}} = E_{\text{self}} \\ e_{\text{initiative}} = E_{\text{self}} + E_{\text{stimulaed}} \\ e_{\text{passive}} = E_{\text{stimulaed}} \end{cases} \tag{5}$$

式中，E代表该点的能量辐射类型。红外频段下的兴趣逻辑码将基于各像点的热辐射贡献程度对其进行类型区分，以增强核心能量点信息、保留主动能量点信息以及抑制被动能量点为目标。对于红外特征图$F_{\text{inf}}^{\text{size}\times\text{size}}$上某一像点$i$，利用能量池化核$\omega$对其所在的邻域$X$内像点$m_k$进行能量辐射量计算，并获得$i$点的能量类型，该过程如图7所示。

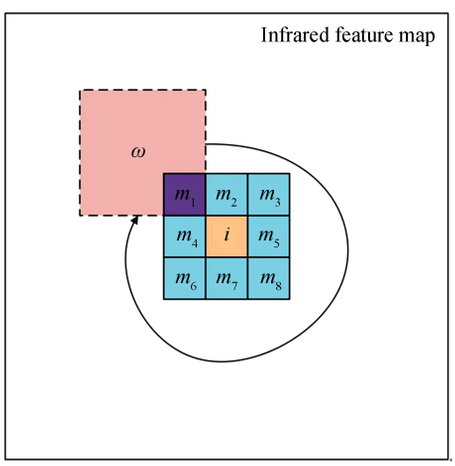

图7 计算像点热辐射类型过程

Fig.7 The process of calculating the thermal radiation type type of image point

ω是尺度为3×3的最大池化核，ω将遍历X区域内像点，并计算每次的最大池化像素值，像点i在X邻域内的能量辐射度E_i为

$$E_i = \sigma\left[\sum_{k=1}^{8}(I_i - \omega X_{m_k})\right] \tag{6}$$

式中，X_{mk}为如图7所示的以m_k像点为基准的池化区域，I_i为X区域中心i点的像素值，σ为sigmoid函数。根据不同像点的E_i值分析像点能量类型，并生成红外频段下低频兴趣逻辑掩码M_{inf}，则掩码值标记规则为

$$M_{\text{inf}}(x,y) = \begin{cases} 1 & m = e_{\text{core}} \quad 0.4 \leqslant E_i \leqslant 0.7 \\ E_i & m = e_{\text{initiative}} \quad 0.7 < E_i < 1 \\ 0 & m = e_{\text{passive}} \quad 0 \leqslant E_i < 0.4 \end{cases} \tag{7}$$

2.2.3 基于视觉注意力机制的可见光兴趣掩码生成

可见光图像主要含有目标高频轮廓信息，这一信息由目标上各点像素与周围像点间的差异所表现[18]。为保留可见光频段上更为丰富的轮廓信息，精确目标注意力区域的划分，采用如图8所示的可见光注意力机制网络生成该频段下的兴趣逻辑码。

视觉注意力机制网络将对可见光特征图$F_{\text{vis}}^{\text{size}\times\text{size}}$分别进行全局最大池化、全局平均池化以及下、上采样计算。其中下采样为kernel size=3、stride=2的卷积计算，该过程循环两次；上采样为对应的反卷积计算，将完成卷积计算的特征图还原为输入尺寸。经上述三个过程计算后生成的特征图进行叠加，合成初级高维兴

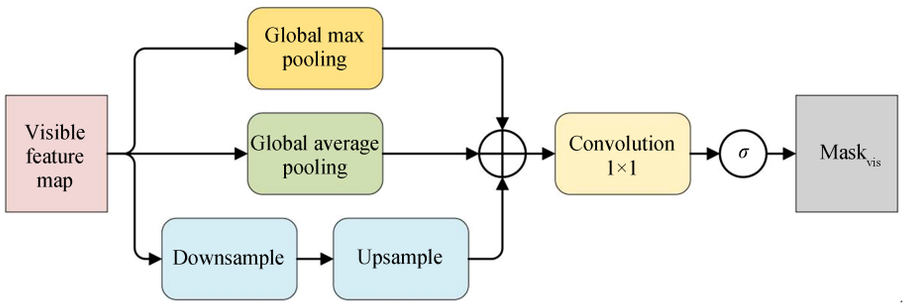

图8　可见光频段下的注意力网络

Fig.8　Attention network in visible channel

趣逻辑码$M_{\text{vis-p}}$，表示为

$$M_{\text{vis}-\text{p}}=\text{gmp}(F)\oplus\text{gap}(F)\oplus\text{conv}(F) \tag{8}$$

式中，F为可见光特征图$F_{\text{vis}}^{\text{size}\times\text{size}}$，gmp为全局最大池化计算，gap为全局平均池化计算，conv为采用卷积操作的下、上采样计算，$\oplus$采用concatenate叠加。

$M_{\text{vis-p}}$经由kernel size=3、stride=2的低通道数卷积层进行降维，获得低维度初级兴趣逻辑码$M_{\text{vis}-\text{p}}^{*}$。最后利用sigmoid函数对$M_{\text{vis}-\text{p}}^{*}$内部各点逻辑掩码值进行归一化处理，获得可见光频段下高频兴趣逻辑掩码M_{vis}，则掩码值标记规则为

$$M_{\text{vis}}(x,y)=\begin{cases}1 & 0.7\leqslant T_i\leqslant 1\\ T_i & 0.4<T_i<0.7\\ 0 & 0\leqslant T_i<0.4\end{cases} \tag{9}$$

式中，T_i为M_{vis}上像点(x,y)兴趣逻辑掩码值。将同尺寸下双频特征图的兴趣逻辑码M_{inf}与M_{vis}进行整合，获得该特征图的目标兴趣区域逻辑掩码M为

$$M(x,y)=\begin{cases}1 & 0.7\leqslant T_i\leqslant 1\quad\text{or}\quad 0.4\leqslant E_i\leqslant 0.7\\ \dfrac{(T_i+E_i)}{2} & 0.4\leqslant T_i<0.7\quad\text{and}\quad 0.7<E_i\leqslant 1\\ 0 & 0\leqslant T_i<0.4\quad\text{and}\quad 0\leqslant E_i<0.4\end{cases}$$

$$M(x,y)=\begin{cases}1 & 0.7\leqslant T_i\leqslant 1\vee 0.4\leqslant E_i\leqslant 0.7\\ \dfrac{(T_i+E_i)}{2} & 0.4\leqslant T_i<0.7\wedge 0.7\leqslant E_i\leqslant 1\\ 0 & 0\leqslant T_i<0.4\wedge 0\leqslant E_i<0.4\end{cases} \tag{10}$$

则附有兴趣逻辑码的双频融合图像F_{mask}为

$$F_{\text{mask}}(x,y)=F_f(x,y)\otimes M(x,y) \tag{11}$$

由于用于检测的$F_{\text{f}-\text{a}}^{\text{size}\times\text{size}}$中含有目标的信息区域得到了保留，而非目标区域信息被抑制，所以相对于目标区域所在的网格，仅有非目标区域信息的网格置信值更低。不同频段融合前与叠加兴趣区域逻辑码后的

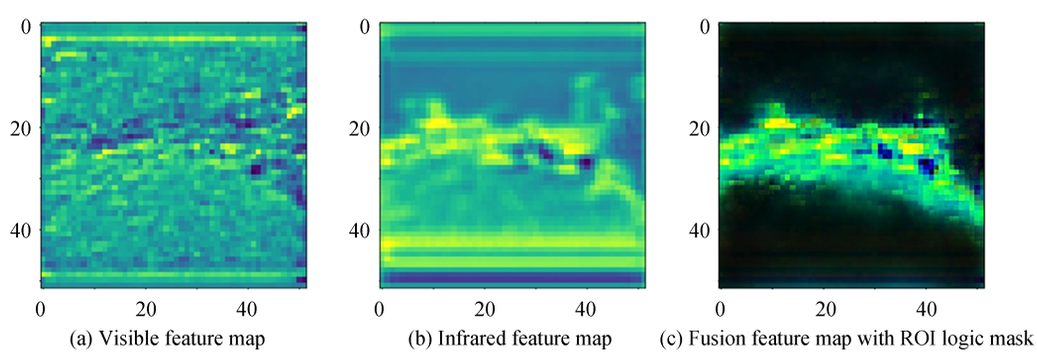

图9　叠加兴趣区域逻辑掩码前后各频段特征图

Fig.9　Feature maps of each channel before and after adding the logical mask of ROI

语义信息热力特征图如图9。

本文通过上述方法实现了对目标区域信息的凸显与非目标区域的抑制，提升算法针对复杂环境下目标的检测精度与抗干扰能力。

2.3 含有注意力逻辑码的双频融合识别算法

2.3.1 双频融合图像目标识别

IVFNN的目标识别网络以YOLO V3为基础构架。经IVFNN进行特征提取、双频图像融合与视觉注意力分析后，网络将产生尺度分别为13×13、26×26与52×52、利用兴趣区域逻辑码进行目标增强的三个待识别融合特征图$F_{\mathrm{f-a}}^{\mathrm{size}\times\mathrm{size}}$。不同尺度的$F_{\mathrm{f-a}}^{\mathrm{size}\times\mathrm{size}}$将被分割成size×size个网格，每个网格内的特征图信息将分别传送到输入层并向后传递计算。若目标物的中心落在某个网格中，则该网格就将对这个区域内的目标物进行检测。置信值表示当前网格中是否包含目标物以及目标物位置的准确性，置信值为检测目标物的概率与IOU的乘积，表达式为

$$C = P(o)\times \mathrm{IOU}_{\mathrm{pred}}^{\mathrm{truth}} \tag{12}$$

$P(o)$表示目标是否存在于当前网格内，若存在则值为1，不存在值为0。IOU为交并比(Intersection Over Union)，即目标物产生的目标框与该目标物的范围框的交并比，其表达式为

$$\mathrm{IOU} = \frac{\mathrm{DR}\cap\mathrm{GT}}{\mathrm{DR}\cup\mathrm{GT}} \tag{13}$$

式中，DR为检测目标框范围(Detection Result)，GT为真实目标覆盖范围(Ground Truth)。由于IVFNN的网络训练数据集由可见光与红外双频标记图像及其对应标签构成，故GT范围由红外真实目标覆盖范围($\mathrm{GT_{inf}}$)与可见光真实目标覆盖范围($\mathrm{GT_{vis}}$)构成。则IVFNN网络中IOU′表达式为

$$\mathrm{IOU}' = \frac{\mathrm{DR}\cap\left(\mathrm{GT_{inf}}\cup\mathrm{GT_{vis}}\right)}{\mathrm{DR}\cup\left(\mathrm{GT_{inf}}\cup\mathrm{GT_{vis}}\right)} \tag{14}$$

受到双频人工标记误差与红外-可见光特征图融合影响，以及兴趣区域逻辑码对目标特征图边缘信息抑制作用，部分检测结果的DR范围会小于GT甚至位于GT内部。由于IOU对重叠框之间的距离不敏感，所以本文引入双频道CIOU(Complete IOU)作为衡量DR与GT间关系的依据，并针对红外与可见光频段下的合成真实目标覆盖范围$\mathrm{GT_f}$进行调整。IVFNN网络中CIOU表达方式为

$$L_{\mathrm{CIOU}} = 1-\mathrm{IOU}'+\frac{\rho^2(b,b^*)}{c^{*2}}+\alpha v \tag{15}$$

$$\alpha = \frac{v}{(1-\mathrm{IOU}')+v} \tag{16}$$

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^*}{h^*}-\arctan\frac{w}{h}\right)^2 \tag{17}$$

式中，α为权重函数，v为DR与GT长宽比相似性的度量，ρ为两点间欧式距离，w、h、b分别为DR框的宽、高与中心坐标，w^*、h^*、b^*分别为$\mathrm{GT_f}$框的宽、高与中心坐标，c^*为b与b^*间距离。IVFNN网络中CIOU如图10。

每个网格同时还将预测C个目标物类别概率，即第i类目标物中心落在该网格内的概率，记为$P(C_i|O)$。C为目标物的类别数，与目标物个数B无关。将目标物类别概率与置信值相乘，得到的乘积作为置信值评分，表示为

$$P(C_i|O)\times P(O)\times \mathrm{CIOU} = P(C_i)\times \mathrm{CIOU} \tag{18}$$

最终，根据每个网格所得出的置信值评分，确定网格中是否存在目标物，从而完成目标物的确定与识别。

2.3.2 双频融合图像目标训练

IVFNN使用均方和误差作为损失函数来优化模型参数，即网络输出的size×size×(B×5+c)维双频向量与真实红外和可见光图像的对应size×size×(B×5 + c)维向量的均方和误差。对于目标检测网络，需要计算的损失由以下三部分构成：

$$E_{\mathrm{v\text{-}i}}^{\mathrm{coord}} = \lambda E_{\mathrm{f\text{-}v}}^{\mathrm{coord}}+(1-\lambda)E_{\mathrm{f\text{-}i}}^{\mathrm{coord}} \tag{19}$$

$$E_{\mathrm{v\text{-}i}}^{\mathrm{ciou}} = \lambda E_{\mathrm{f\text{-}v}}^{\mathrm{ciou}}+(1-\lambda)E_{\mathrm{f\text{-}i}}^{\mathrm{ciou}} \tag{20}$$

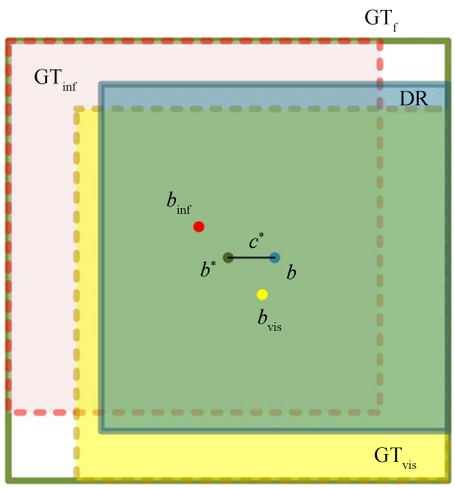

图 10　IVFNN 中 CIOU 示意图

Fig.10　Schematic diagram of CIOU in IVFNN

$$E_{\text{v-i}}^{\text{class}}=\lambda E_{\text{f-v}}^{\text{class}}+(1-\lambda)E_{\text{f-i}}^{\text{class}} \tag{21}$$

$E_{\text{v-i}}^{\text{coord}}$、$E_{\text{v-i}}^{\text{ciou}}$、$E_{\text{v-i}}^{\text{class}}$ 为双频融合图像分别关于红外和可见光频段图像在 DR 和 GT_f、CIOU 以及目标分类的之间的误差。由于 IVFNN 引入了红外与可见光双频图像作为原始输入和训练数据源，故本文采用交叉损失计算的方式，分别获得融合-红外(f-i)与融合-可见光(f-v)下的各类损失值，然后利用 $\lambda=0.5$ 调整结构误差权值，交叉损失计算方法如图 11。

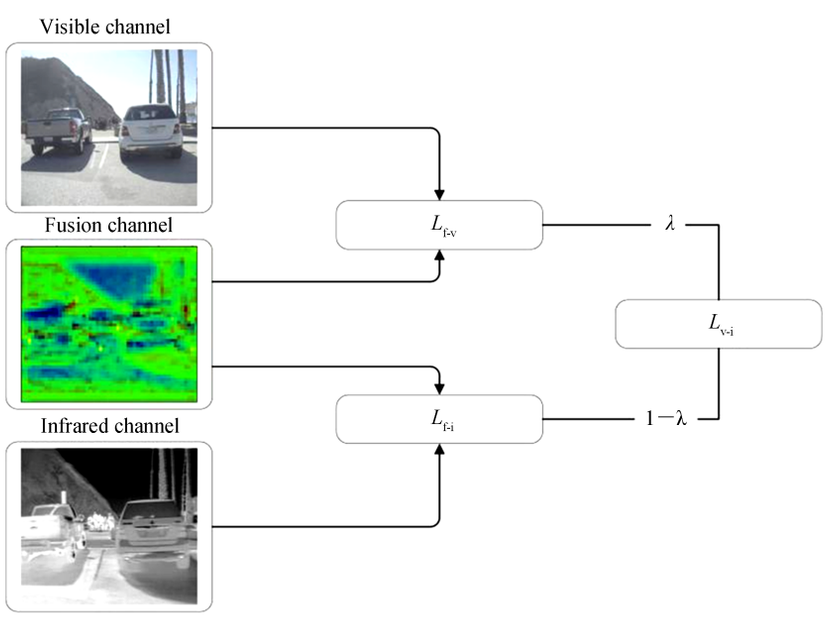

图 11　双频与融合图像交叉损失计算

Fig.11　Cross loss calculation of dual channels and fused images

融合-红外(f-i)与融合-可见光(f-v)下的各类损失值计算方式为

$$\text{coordError}=\lambda_{\text{coord}}\sum_{j=0}^{B}\prod_{ij}^{\text{obj}}[(x_i-x_i^{\,*})^2+(y_i-y_i^{\,*})^2]+\lambda_{\text{coord}}\sum_{j=0}^{B}\prod_{ij}^{\text{obj}}[(\sqrt{w_i}-\sqrt{w_i^{\,*}})^2+(\sqrt{h_i}-\sqrt{h_i^{\,*}})^2] \tag{22}$$

$$\text{ciouError}=\sum_{j=0}^{B}\prod_{ij}^{\text{obj}}(C_i-C_i^{\,*})^2+\lambda_{\text{noobj}}\sum_{j=0}^{B}\prod_{ij}^{\text{obj}}(C_i-C_i^{\,*})^2 \tag{23}$$

$$\text{classError}=\prod_{i}^{\text{obj}}\sum_{c\in\text{class}}(p_i(c)-p_i^{\,*}(c))^2 \tag{24}$$

式中，p 为 DR 下的预测值，p^* 为 GT_f 下的标注值。Π_i^{obj} 表示物体落入网格 i 中的情况，Π_{ij}^{obj} 和 Π_{ij}^{noobj} 表示物体

落入与未落入网格i的第j个目标框内的情况。综上，IVFNN的整体损失函数为

$$\mathrm{Loss}=\sum_{i=0}^{S^2}\mathrm{coordError}_{v-i}+\mathrm{iouError}_{v-i}+\mathrm{classError}_{v-i} \tag{25}$$

3 实验结果

本文为检验IVFNN的计算效果，以FLIR红外-可见光图像数据集作为训练样本。其中红外图像包含7 153张，可见光图像包含6 936张，检测类别分为人、车、自行车三类。训练过程中batch size为32，训练学习率自调。仿真测试平台为DELL Z840，CPU配置为Intel Xeon E5-2643 V3，主频3.4 GHz，GPU为Quadro P5000，运行内存32 GB，计算环境为Ubuntu 18.04。测试时IVFNN编写语言为Python3.7，配合Tensorflow 2.0与Opencv 3.2作为辅助高级API。IVFNN完成上述训练后损失函数收敛情况如图12。

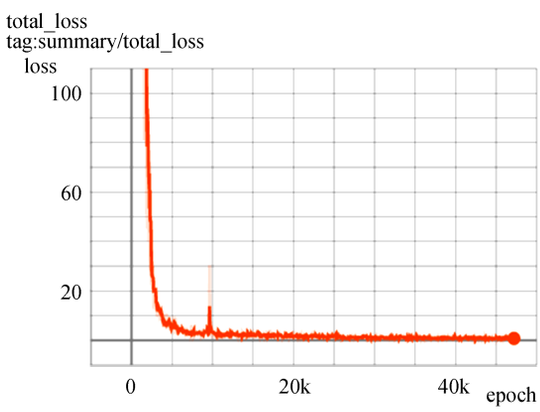

图12 IVFNN损失函数曲线

Fig.12 Loss function curve of IVFNN

通过图12可以看出，IVFNN在上述训练过程中，损失函数有较好的收敛现象，未出现梯度爆炸或梯度丢失等问题，证明所设计的网络结构具备合理性与可行性。为对比IVFNN的双频自适应融合与兴趣区域凸显功能对目标识别精度的提升，在完成融合图像数据的训练后，分别将原始红外与可见光频段下的数据集输入至IVFNN中进行单频段检测测试。此时IVFNN中关于图像融合、交叉损失函数计算、双通道特征

图13 IVFNN双频融合与各频段独立识别效果

Fig.13 Recognition effect of IVFNN and each frequency band

提取网络等双频信息交互部分全部关闭，保留单频图像数据计算环节，还原网络至单通道训练与识别状态。完成红外、可见光与双频融合的训练后，各频段识别效果如图13。

从图13的测试结果可以看出，在使用独立通道进行识别时，图13(a)所示的可见光频段能够将前方的车辆进行识别，但无法识别纹理信息较少的行人。图13(b)所示的红外频段能够识别出车辆与行人，但由于特征信息量较少且维度较低，语义信息提取程度不高，车辆的识别精度相对可见光较差。图13(c)所示的融合识别策略因扩展了双频语义信息，对目标与行人的识别精度有所提升，但受到环境信息干扰，识别框与目标真实轮廓依然未达到精确吻合，且部分非目标特征在融合后出现了特征增幅现象，即某目标在一个频段内信息特征不明显，而在另一频段内信息特征明显，融合后特征发生了中和，使得融合特征图中语义信息达到阈值进而被错误识别。图13(d)所示的含有兴趣区域逻辑掩码的自适应融合识别保留了双频下的高维语义信息，精确地将目标所在区域与背景进行分割和标记，减少了识别过程中预选区域判定计算量，在提高识别精度的同时降低了环境信息与融合后非目标特征增幅对识别结果的影响。

部分待检测目标因辐射效果较弱，于红外频段下信息反馈量较低。此时应要求IVFNN的自适应融合算法对该类型目标的红外特征与可见光特征融合比例进行调整，选取最佳的特征融合权重分配模式。为了验证IVFNN对弱热辐射目标成像的自适应融合效果，进行了自行车目标的识别测试，以体现IVFNN对低热辐射目标的识别效果。IVFNN对低辐射目标识别效果如图14。

(a) Visible light recognition　(b) Infrared recognition　(c) Fusion recognition with ROI mask

图14　低热辐射目标的识别效果

Fig.14　Recognition effect of low-heat radiation target

通过图14可知，针对低热辐射的自行车目标，IVFNN的识别精度高于单频段可见光或红外的识别精度，在面对热辐射信息量较低的目标时，IVFNN能够有效地进行自适应融合调整与权值分配，防止强特征贡献频段的信息被弱特征贡献频段的信息所淡化。

为验证本文算法对多尺度目标的识别效果，测量IVFNN的最小可识别目标像素占比及其识别精度，本文基于FLIR数据集中运动车辆视频流数据作为识别样本，对同一目标在不同距离与像素占比的图像，利用红外、可见光与双频自适应融合三种方式进行识别测试。测试将从视频中目标车辆完整进入视场后开始，随着目标车辆的不断远离，IVFNN将持续对车辆进行双频融合自适应识别，并记录识别精度及车辆像素占比。当目标车辆出现连续3帧识别错误、连续10帧识别波动或目标识别精度达到最低识别阈值(60%)时，则认为已达IVFNN的最小目标识别极限。不同成像频段下目标多尺度识别测试效果如表1。

表1　不同成像频段下目标多尺度识别测试效果

Table 1　Multi-scale target recognition ability of IVFNN and each frequency band

Imaging frequency band	Minimum pixel of recognizable	Minimum identifiable target pixel ratio	Limit accuracy of minimum target recognition
Infrared	370×232 pixels	6.98%	65.3%
Visible light	452×292 pixels	10.7%	69.5%
Dual channels adaptive fusion of IVFNN	248×179 pixels	3.61%	60.7%

由于IVFNN融合了红外与可见光的复合特征，当目标像素信息量减少时，IVFNN的自适应融合方法将调整不同频段的特征表达权重，选取当前目标占比下特征贡献值最大的方式进行语义信息配比，能有效减缓因目标尺度减小导致的分类波动和识别精度骤降。通过表1可知，IVFNN在最小目标识别像素量与最小目标识别极限精度稳定性上均优于仅使用单探测频段的识别方法，证明IVFNN在多尺度目标识别上具有良好表现。

结合上述测试结果可知，IVFNN通过自适应融合的方式，在特征图层面上深化了目标在红外与可见光双频段上的语义信息，使识别的抗干扰能力更强，进一步提升了目标识别精确性。同时，具有目标兴趣区域标记能力的IVFNN能够有效地针对目标特征与所在位置进行精确识别，提高了算法对目标所在位置的锚定能力，降低了非目标区域对识别精度的影响。

在上述测试的基础上，将所提及的部分红外-可见光融合算法与IVFNN进行对比，以评判本文方法在双频融合领域的提升效果。由于其他参考算法仅涉及了融合策略与特征提取策略，尚未构成完整的端到端学习、训练与目标识别能力，故将该类算法的核心部分迁移至IVFNN的等效计算环节中，构建相同的数据计算流程与结构。各算法与IVFNN的测试结果对比如表2。

表2 各算法测试结果对比

Table 2 Comparison of test results of each algorithm

Algorithm used	Processes	Average recognition rate	Average calculation speed
Ref.[2]	Feature matching and fusion	70.6%	35 fps
Ref.[3]	Filtering fusion	77.9%	37 fps
Ref.[6]	Filtering fusion	76.1%	32 fps
Ref.[7]	Neural network fusion	80.5%	22 fps
Ref.[10]	Neural network fusion	81.3%	23 fps
Proposed algorithm	Global IVFNN	83.2%	25 fps

通过表2可以看出，IVFNN在整体结构上具有较好的计算精度与稳定性，由于采用了双通道特征提取网络构架，IVFNN的平均计算速度相比其他算法较低，但仍能达到实时解算速度，若需要进一步提升计算效果，可选用深度可分离卷积层作为特征提取基础模块，或对网络进行剪枝优化。对于复杂环境下热源目标的识别任务，IVFNN可以通过自适应融合策略获取目标在双频成像下的高维特征，并利用兴趣逻辑码减少环境背景的识别干扰，有效地提升了多尺度目标的识别效率。

4 结论

本文提出了一种双通道红外-可见光双频融合识别神经网络，采用自适应策略的方式在双频特征图上进行高维语义特征融合，保留了多维度目标特征信息。配合能量池化计算与视觉注意力机制生成红外与可见光频段下的目标兴趣区域逻辑掩码，增强目标特征信息并抑制环境干扰。实验结果表明，所设计的算法能够在复杂环境中对多尺度热源目标进行红外与可见光频段下的精确识别，并具备良好的抗干扰和自适应能力。

参考文献

[1] MA Jiayi, MA Yong, LI Chang. Infrared and visible image fusion methods and applications: A survey[J]. Information Fusion, 2019: 153-178.

[2] WANG Ting. Research and application of infrared image and visible image fusion[D]. Xi′an: Xi′an University of Technology, 2019.
汪廷. 红外图像与可见光图像融合研究与应用[D]. 西安: 西安理工大学, 2019.

[3] FENG Yufang, YIN Hong, LU Houqing, et al. Infrared and visible image fusion based on improved convolutional neural network[J]. Computer Engineering, 2020, 46(8): 243-249+257.
冯玉芳, 殷宏, 卢厚清, 等. 基于改进卷积神经网络的红外与可见光图像融合[J]. 计算机工程, 2020, 46(8): 243-249+257.

[4] ZUO Yujia. Research on key technology of infrared and visible image fusion system based on airborne photoelectric platform[D]. Changchun: Changchun Institute of Optics, Precision Machinery And Physics, Chinese Academy of Sciences, 2017.

左羽佳.机载光电平台红外与可见光图像融合系统关键技术研究[D].长春:中国科学院长春光学精密机械与物理研究所,2017.

[5] JUNG T Y, SON D M, LEE S H. Visible and NIR image blending for night vision[C]. International Conference on Image Processing, Computer Vision and Pattern Recognition(IPCV), 2019. Las Vegas: CSREA Press, 2019:110-113.

[6] JEE S, KANG M G. Sensitivity improvement of extremely low light scenes with rgb-nir multispectral filter array sensor [J]. Sensors, 2019, 19(5):1256.

[7] JUNG C K, ZHOU K L, FENG J W. FusionNet: Multispectral fusion of RGB and NIR images using two stage convolutional neural networks[J]. IEEE Access, 2020, 8: 23912-23919.

[8] MATTHEW B, SABINE S. Multi-spectral SIFT for scene category recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2011, Providence, RI, USA, New York: IEEE, 12218780.

[9] TANG Chaoying, PU Shiliang, YE Pengzhao, et al. Fusion of low-light visible and NIR images based on convolutional neural networks[J]. Acta Optica Sinica, 2020, 40(16):37-45.
唐超影,浦世亮,叶鹏钊,等.基于卷积神经网络的低照度可见光与近红外融合[J].光学学报,2020,40(16):37-45.

[10] LIU Yaochen. Research on fusion method of visible light image and infrared image [D]. Dalian: Dalian Maritime University, 2019.
刘耀晨.可见光图像与红外图像的融合方法研究[D].大连:大连海事大学,2019.

[11] JIANG Zetao, LIU Xiaoyan, HU Shuo. Scene recognition of infrared and visible light fusion images based on CNN[J]. Computer Engineering and Design, 2019, 40(8):2289-2294.
江泽涛,刘小艳,胡硕.基于CNN的红外与可见光融合图像的场景识别[J].计算机工程与设计,2019,40(8):2289-2294.

[12] WANG Lan, Research on action recognition methods based on infrared and visible spectrum[D]. Chongqing: Chongqing University of Posts and Telecommunications, 2019.
汪澜.基于红外与可见光双光谱视频的行为识别算法研究[D].重庆:重庆邮电大学,2019.

[13] PENG Yuqing, ZHAO Xiaosong, TAO Huifang, et al. Hand gesture recognition against complex background based on deep learning[J]. Robot, 2019, 41(4):534-542.
彭玉青,赵晓松,陶慧芳,等.复杂背景下基于深度学习的手势识别[J].机器人,2019,41(4):534-542.

[14] LONG Zhiyong. Research on infrared and visible image registration and fusion algorithm[D]. Chengdu: University of Electronic Science and Technology of China, 2020.
龙勇志.红外与可见光图像配准与融合算法研究[D].成都:电子科技大学,2020.

[15] SHEN Yu, CHEN Xiaopeng, YUAN Yubin, et al. Infrared and visible image fusion based on significant matrix and neural network [J/OL]. Laser & Optoelectronics Progress: 1-16 [2020-08-13]. http://kns.cnki.net/kcms/detail/31.1690.TN.20200711.1640.002.html.
沈瑜,陈小朋,苑玉彬,等.基于显著矩阵与神经网络的红外与可见光图像融合[J/OL].激光与光电子学进展:1-16[2020-08-13].http://kns.cnki.net/kcms/detail/31.1690.TN.20200711.1640.002.html.

[16] AN Haonan, ZHAO Ming, PAN Shengda, et al. Infrared target detection algorithm based on pseudo multimodal images [J]. Acta Photonica Sinica, 2020, 49(8):176-188.
安浩南,赵明,潘胜达,等.基于伪模态转换的红外目标融合检测算法[J].光子学报,2020,49(8):176-188.

[17] JI Lu, LIU Shijian, WANG Xiao, et al. Infrared aircraft classification method with small samples based on improved relation network[J]. Acta Optica Sinica, 2020, 40(8):87-96.
金璐,刘士建,王霄,等.基于改进关系网络的小样本红外空中目标分类方法[J].光学学报,2020,40(8):87-96.

[18] SHEN Yu, CHEN Xiaopeng, LIU Cheng, et al. Infrared and visible image fusion based on hybrid model driving[J/OL]. Control and Decision:1-8[2020-08-13].https://doi.org/10.13195/j.kzyjc.2019.1749.
沈瑜,陈小朋,刘成,等.基于混合模型驱动的红外与可见光图像融合[J/OL].控制与决策:1-8[2020-08-13].https://doi.org/10.13195/j.kzyjc.2019.1749.

Foundation item: National Defense Science and Technology Foundation Enhancement Plan Technology Field Fund(No.2020-JCJQ-JJ-422), Joint Foundation of Equipment Pre-research and Weapons Industry (No.6141B012841)