



重构迁移学习的红外目标分类

毛远宏, 贺占庄, 马 钟*

(西安微电子技术研究所 西安 710065)

【摘要】 红外图像目标分类在目标识别等领域有重要的应用价值, 目前卷积神经网络在可见光图像分类方面达到了优异的性能。但对于红外图像来说, 由于有标记样本数量少和图像成像差异大, 直接使用现有的网络模型来处理红外图像无法取得理想效果。该文将可见光图像作为源域, 将红外图像作为目标域, 在深度网络中使用迁移学习方法来解决此问题。在迁移学习中, 目标域网络提取的特征越能体现出本域数据的真实分布, 那么在此基础上进行两个域的特征分布适配就更加有效, 迁移后的目标域网络性能和泛化能力越好。该文首先利用大量无监督的红外样本训练了红外图像深度卷积自编码器, 增强了红外图像域网络的特征表达能力。其次, 通过减小源域和目标域的特征分布距离, 使得两个图像域特征分布相似, 从而将源域中深度网络的学习能力迁移到目标域。经过上述改进, 相比于可见光图像预训练微调的方法, 分类准确率提升了 11.27%。

关键词 卷积自编码器; 卷积神经网络; 红外图像; 目标分类; 迁移学习

中图分类号 TP391 **文献标志码** A **doi**:10.12178/1001-0548.2019162

Infrared Target Classification with Reconstruction Transfer Learning

MAO Yuan-hong, HE Zhan-zhuang, and MA Zhong*

(Xi'an Microelectronics Technology Institute Xi'an 710065)

Abstract Infrared target classification has important values in target recognition. At present, convolutional neural network has achieved excellent performance in visible image classification. However, for infrared images, the available networks can't achieve satisfying results due to the small number of annotated samples and large imaging differences. In this paper, visible images are used as source domain, infrared images as target domain. Transfer learning is used to address the challenges in the deep learning framework. In the transfer learning, if the target domain network can represent the distribution of its domain well, the performance and generalization of the target domain network should be more effective. Therefore, the convolutional autoencoder is trained with a large number of unannotated infrared samples, which greatly enhances the feature representation in the infrared image domain. By reducing the feature distribution distance between the two domains, the feature distributions become similar. The classification performance in the source domain is transferred to the target domain. With the changes above, the accuracy rate is improved by 11.27% compared with the method based on the visible images fine-tuning.

Key words convolutional autoencoder; convolutional neural networks; infrared images; target classification; transfer learning

红外目标分类在计算机视觉应用中有着重要的应用价值。例如, 在自动目标识别中, 能够实现对于检测目标在红外影像下的识别分类, 可以有效提高识别性能^[1-2]。红外目标分类相对于传统可见光图像分类, 也存在着更多困难。由于红外传感器的成像特点, 红外图像中虽然包含目标的外部轮廓和区域特征, 但是目标的颜色、纹理等细节信息非常有

限。其次, 对于同一类目标, 红外目标存在不同视角和形变等因素的影响, 增加了分类的困难。最后, 红外传感器图像的样本采集成本高, 很难获取大量有标注的数据集来进行训练。因此, 基于大规模监督数据的深度学习方法难以直接应用于红外目标分类。

传统红外图像目标分类的方法主要是基于人工

收稿日期: 2019-07-10; 修回日期: 2020-03-01

基金项目: 国家自然科学基金青年科学基金(61702413)

作者简介: 毛远宏(1980-), 男, 博士, 主要从事计算机视觉方面的研究。

通信作者: 马钟, E-mail: mazhong@mail.com

设计特征的分类。基于手工特征的匹配方法,主要通过 SIFT^[3], SURF^[4], ORB^[5] 等描述子,从红外图像中提取图像的特征点,然后输入分类器进行分类。文献 [6] 提出在红外图像上使用小波变换来提取特征,再进行分类识别。文献 [7] 使用 PLB 直方图和灰度直方图来生成红外图像特征,通过 SVM 进行分类。文献 [8] 提出红外目标的边缘特征,再使用 SOM 进行分类。上述方法,采用手工提取的特征无法和后续分类器进行端到端的整体优化。同时,由于获取的样本相对有限,也难以保证其在红外目标分类上的泛化性能。

深度学习的方法兴起之后,在可见光图像分类方面,取得了前所未有的性能进步^[9-12]。深度学习通过 CNN(convolutional neural network) 分支来提取待匹配图像的特征,使用 Softmax 函数对于样本进行分类。由于红外图像采集成本过高,目前没有公开的大规模有监督红外目标分类数据集。在深度学习中,如果仅使用少量的红外图像数据来进行学习,网络很快会发生过拟合,影响模型后续的泛化能力。

为了避免卷积神经网络在小样本训练中出现过拟合,通常在 ImageNet 数据集的预训练网络模型基础上,使用少量的红外标注样本进行再次训练,也就是业界常用的网络微调^[13]。虽然此种方法在一定程度上可以缓解模型过拟合的影响,但是由于传感器存在成像原理的差异,可见光和红外图像样本之间不是同分布的关系。因此,仅仅采用参数微调的方法,在红外目标分类上很难取得好的效果。

近年来,随着深度学习的广泛应用,与迁移学习方法之间的结合也不断加深^[14-15]。通过迁移学习,源域和目标域能够在特征空间中实现域间分布适配,解决了红外和可见光图像由于不同成像机理造成的样本分布差异。但是在迁移学习中,目标域中用于训练的样本通常非常有限,并且这些训练样本的分布可能和整个目标域的总体分布有比较大的偏差。在这种情况下,即使域适配的方法本身没有问题,迁移后的目标域模型在性能和泛化能力上也可能达不到很好的效果。因此,迁移学习时,要尽量提高目标域网络的特征表示能力,使得用于域适配的高层特征尽可能和真实分布接近,从而保证迁移学习效果。

另一方面,与可见光图像分类不同,红外目标分类领域中缺乏公开的大规模标注数据库,但是在

数据采集过程中,大量无标注的红外图像却相对容易获得。这些无标签红外样本实际上也潜在包含了红外图像域的特征分布。深度自编码网络^[16]可以用于红外图像进行编码和解码,充分利用这些无监督样本进行学习。通过重构,保证目标的红外特征在目标域中不丢失,从而提高网络中红外图像的特征表达能力。

本文基于 VGG16 网络构造了卷积自编码器,对大量无标记的红外图像进行无监督学习,红外域分支的自编码器通过无监督学习提高了红外域特征提取能力。同时,使用迁移学习方法,将红外域分支和可见光域分支的高层特征分布进行域适配,使其特征分布相似,从而将可见光图像(源域)网络的学习能力迁移给红外图像域(目标域)。实验证明,基于重构的迁移学习,有效提高了网络对红外特征的表达能,也提升了红外目标的分类效果。

1 基于重构的迁移学习网络架构

本文在训练过程中采用了双分支的神经网络架构,主要由红外卷积自编码器和可见光分类网络组成。为了保证两个网络的特征维度保持一致,两个分支都使用了 VGG16 作为主干。VGG16 网络中包含多个 Block。每个 Block 由 2 个或者 3 个卷积层、一个激活层和一个池化层组成。红外和可见光输入图像都采用了 224×224×3 像素的尺寸,红外图像由单通道图像被复制为 3 通道图像。

其网络架构如图 1 所示。可见光图像输入网络使用传统的 VGG16 网络来进行可见光的分类任务。在全连接(full connection, FC)层,可见光域和红外域通过 MMD(maximum mean discrepancy) 距离来评估两个域的图像特征分布,减少特征分布差异。

红外分支的卷积自编码器,先单独使用大量无标签的红外图像进行无监督训练。训练完成之后,再将整个自编码器嵌入迁移学习网络,作为红外图像分支的特征提取网络。自编码器的编码网络结构和可见光分类网络结构相同。这样设计是为了便于在相同的网络层次,可见光域和红外域都能够获取到相同维度的特征来计算 MMD 损失。在红外和可见光域进行域适配的同时,红外自编码器的解码网络能够从红外高层特征重构出原始的红外图像,避免了目标的红外特征信息丢失。

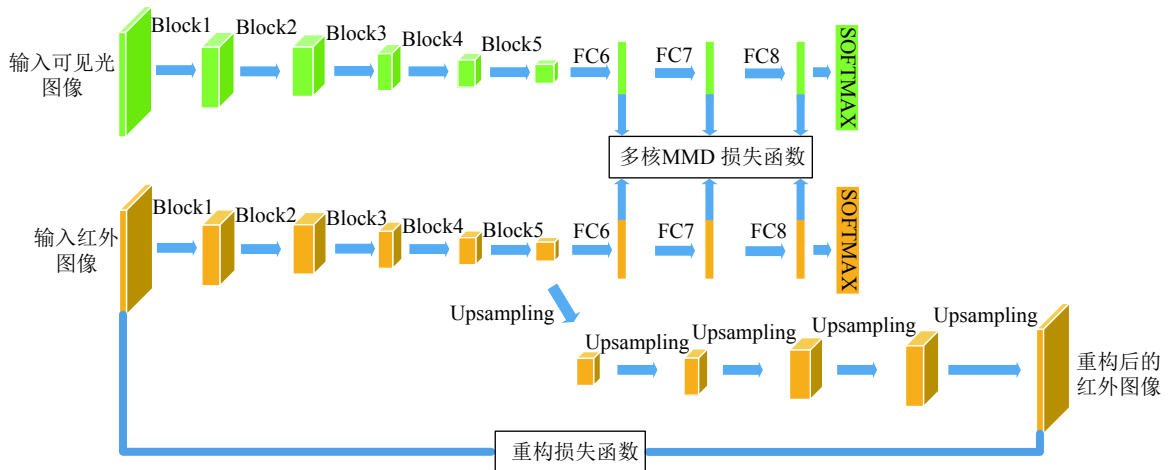


图 1 基于重构的迁移学习训练网络架构图

在整体网络训练的过程中, 待优化的损失函数如下:

$$\sum_{i=1}^L \lambda_i \text{MMDLoss}(D_s, D_t) + \sum_{i=1}^2 \text{SoftmaxLoss}_{br_i} + \lambda_r \text{Recon Loss} \quad (1)$$

式中, 第一项损失函数为红外分支和可见光图像特征的特征的 MMD 距离。为了取得更好的效果, 本文采用了多层特征 MMD 距离进行衡量。L 表示卷积神经网络的层数。在本文网络中, 分别在 FC6、FC7 和 FC8 层计算两个域的 MMD 距离; D_s 表示源域在相应网络层的特征; D_t 表示目标域在网络层的特征。通过优化两个域的 MMD 距离, 实现了源域和目标域在特征空间分布相似。第二项函数表示用于红外和可见光网络分类损失总和。第三项为红外图像分支卷积自编码器的重构损失函数。损失函数为输入红外图像像素和重构图像的欧式距离。训练的过程中, 保证了目标的红外特征不丢失。

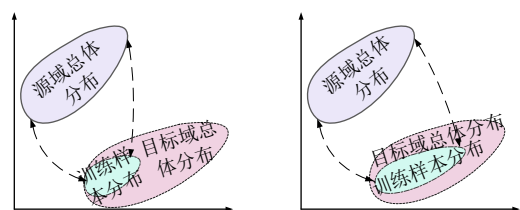
通过多次试验效果确定了超参数的取值, 将 MMD 损失的超参数 λ_i 分别设置为 0.10。ReconLoss 按照原图尺寸所有像素值来计算欧式差距, 考虑到数量比较大, 其 λ_r 设置为 0.001。

需要说明, 在 MMD 损失函数和红外分支的重构损失函数仅在训练过程中参与运算。在测试阶段, 仅使用图 1 中的红外图像分类分支, 通过 Softmax 函数输出相应的红外目标分类概率。因此在测试过程中, 网络前向推理的计算量相比之前 VGG16 网络, 没有增加。

2 红外图像域的深度卷积自编码器

迁移学习的实质是要求红外图像和可见光图像

的特征分布接近, 本文用 ImageNet 预训练模型可以近似代表可见光图像总体分布。对于红外域的网络, 只能通过已知的训练样本来估计本域的总体分布。图 2 分析了目标域的训练样本大小和分布对于迁移学习的影响。图中带箭头的虚线代表了域适配的过程。如图 2a 所示, 如果目标域已知的训练样本分布不能代表目标域整体的分布, 并且有比较大的偏差, 那么在域适配后, 目标域分类器泛化能力也同样受到影响。在图 2b 中, 如果目标域训练样本的分布能大致体现目标域的真实分布, 则迁移后的模型在分类性能和泛化性上都有很大的提升。所以, 在迁移学习中, 目标域网络的特征表达能力越强, 生成的高层特征越能体现目标域的真实分布, 迁移效果越好。当然, 目前目标域训练样本规模还不能完全反映目标域的真实分布, 否则就没有必要进行迁移学习。



a. 训练样本与总体分布差异大 b. 训练样本与总体分布差异小

图 2 目标域训练样本对于迁移学习影响分析

从样本采集上看, 用于迁移学习的分类样本选择非常严格。要求待分类目标在红外和可见光图像要成对出现。图像目标特征明显, 且有角度、尺度和背景的变化。很多红外图像因为达不到上述要求, 而没有得到利用。但是, 这些无监督数据同样包含了红外域的样本分布信息。为了充分使用这些

无监督数据, 本文采用了卷积自编码网络对于红外特征进行学习。如图 1 所示, 红外目标网络的编码部分和可见光图像特征提取网络保持相同的网络结构。因此, 使用无监督学习的过程, 也是提高红外目标特征表示的过程。

在式 (2) 中, 无监督的重构损失函数使用了自编码器中常用的欧式距离损失, 从而最小化输入红外图像和重构后的图像之间差异。

$$\text{Recon Loss} = \|x - f(x')\|_2 \quad (2)$$

式中, x 为输入图像; $f(x')$ 为重构后图像; 重构损失为归一化后像素值之间的欧式距离, 其维度为 $224 \times 224 \times 3$ 。

3 红外图像目标分类的迁移学习

虽然红外和可见光图像成像原理不同, 但是同一目标的边缘轮廓在红外和可见光图像具有相似之处。因此, 在红外图像目标分类中使用迁移学习具有可行性。另一方面, 由于红外图像样本数量少, 容易引起过拟合。基于可见光分类模型的参数微调在一定程度上可以降低过拟合, 然而红外和可见光的图像样本之间不是独立同分布的关系, 无法保证两个域在高维空间的特征分布相似。因此, 在红外目标分类任务中需要基于域适配的迁移学习。

在进行红外和可见光图像迁移学习的过程中, 要抽象出两个域的共性。一般认为高层特征能更关注类别差异, 而忽略底层成像差异。因此本文在 FC6、FC7 和 FC8 层比较了高层特征的 MMD 距离来衡量分布差异, 保证两个域的高层特征的分布相似。而对于底层特征来说, 由于两个域成像机理存在差异, 应该保持两个域在底层网络权值的独立, 通过深度学习中的反向传播算法, 让底层卷积网络去自主地选择各自域中最适合的滤波器, 更好的提取各个域所特有的特征信息。

本文中使用的 MMD 损失函数为:

$$D(D_s, D_t)^2 = \sum_{l=1}^L \left\| \frac{1}{m} \sum_{x_s \in D_s} \phi(x_s^l) - \frac{1}{n} \sum_{x_t \in D_t} \phi(x_t^l) \right\|_H^2 = \sum_{l=1}^L \left(\frac{1}{m^2} \sum_{i,j=1}^m k(x_s^i, x_s^j) - \frac{2}{mn} \sum_{i,j=1}^{m,n} k(x_s^i, x_t^j) + \frac{1}{n^2} \sum_{i,j=1}^n k(x_t^i, x_t^j) \right) \quad (3)$$

式中, L 为代表了在深度网络上需要计算 MMD 距离的网络层数; x_s 代表了源域的样本特征; x_t 代表

了目标域中的样本特征; m 是源域中训练样本数量; n 是目标域中的训练样本数量; $\phi(x)$ 表示样本特征在再生希尔伯特空间的映射; 为了计算简便, 再生希尔伯特空间的内积将转化为核函数 $k(x_i, x_j)$ 。在 MMD 损失函数计算中, 为了核函数具有更强的映射能力, 会采用多个高斯核线性组合进行表示^[17]。在训练过程中, 使用一次训练中输入网络样本计算出方差 σ^2 , 分别使用 $(0.5\sigma^2, \sigma^2, 2\sigma^2)$ 构造 3 个高斯核。在 MMD 损失函数中使用多个高斯核的线性组合, 能够更加有效地评估源域和目标域的样本分布相似性。

4 实验

4.1 实验数据

对于可见光和红外图像目标分类的迁移学习, 目前还没有公开的数据集, 本文构建了数据集进行实验。通过在无人机上搭载红外和可见光摄像头, 获取了不同目标在不同角度、光照、尺度下的图像。本文共标注了 2 000 幅原始图像 (包含可见光和红外图像), 共 25 种类别, 其中包含各种建筑、汽车、桥梁、厂区、天然气站等。可见光和红外图像的数量比为 1 : 1。80% 的样本用来进行训练, 剩余样本用来进行测试。图 3 列出相应目标类别的红外和可见光图像样本, 同一目标的红外和可见光图像除了成像差异之外, 对于不同尺度、角度和背景等因素也进行考虑。

样本标注完成后, 剩余大量无标注的红外图像作为无监督数据集。本文共收集整理 50 000 张红外图像, 用于训练红外图像的深度卷积自编码器。将 90% 的图像作为训练集, 剩余部分作为测试集。无监督的红外图像样本, 对于图像内容没有太多的要求, 需要说明, 在红外无监督样本中不包含用于迁移学习的红外目标图像。图 4 列举了红外无标注样本。



图 3 红外和可见光迁移标注的学习样本示例



图 4 红外图像无标注样本示例

4.2 实验方法

在进行域适配训练之前, 首先需要对红外分支的自编码网络进行训练。训练卷积自编码器时, 本文借鉴了 SegNet^[18] 网络结构。在编码阶段下采样时, 池化层会记录每个池化窗口最大值的索引。在解码阶段中, 上采样层使用这些索引值来恢复相应的特征图。相比于传统的反卷积操作, 此方法能够更大程度减少重构误差。

红外分支的深度卷积自编码器训练完成之后, 本文将和可见光的分类网络进行并联, 开始进行迁移学习训练。可见光分类网络采用 VGG16 作为主干, 使用 ImageNet 预训练的权值进行初始化。在迁移学习训练中, 可见光图像输入可见光分支网络, 红外图像输入自编码器网络。在 FC6, FC7 和 FC8 层进行 MMD 损失优化。梯度下降采用 Momentum 算法。训练参数设置如下: 迭代次数设置为 50 000 次; 网络基础学习率按照迭代次数从 10^{-3} 匀速下降到 10^{-5} ; 硬件平台为 Nvidia TITAN XP GPU; 软件平台使用 Caffe。

测试阶段, 只保留红外图像分支的编码器部分, Softmax 层输出红外图像分类预测。

4.3 实验结果

为了验证本文方法的有效性, 进行了不同模型和方法的对比实验:

1) SIFT 方法。在红外图像上, 使用传统的 SIFT 方法在红外图像上提取特征点, 然后通过 SVM 分类器进行分类。

2) 基于 VGG16 分类网络, 直接使用带标签的红外图像样本从随机初始值开始训练。

3) 基于 VGG16 分类网络, 使用 VGG16 在 ImageNet 数据集上预训练模型来初始化网络, 然后使用有标注的红外图像样本进行再训练, 对网络参数进行微调。由于小样本场景下, 业界通常采用此种训练方法。因此, 在实验结果评估中, 将其设置为性能基线。

4) 基于孪生神经网络进行迁移学习。红外和可见光特征提取均采用 VGG16 作为分支网络, 两个分支都采用 ImageNet 可见光预训练模型的权值进行初始化。两个分支在 FC6, FC7 和 FC8 层优化 MMD 距离, 拟合两个域的特征分布。

表 1 对比了上述方法在红外目标分类的平均准确度。在使用迁移学习的方法中, 括号中的数值为迁移学习中可见光图像 (源域) 分类的准确率。基

于实验结果, 可以看出, 相比于业界通常采用的参数微调方法, 本文方法分类的准确率增加了 11.27%。

表 1 分类准确率结果对比表

测试网络	准确率/%
SIFT+SVM	67.21
VGG16 train from scratch	65.75
VGG16+fine-tuning	78.63
transfer learning(two VGG16 branches)	86.74(92.29)
transfer learning(VGG16+autoencoder branches)	89.90(92.31)

通过对比实验, 可以得出以下结论:

1) 使用传统的 SIFT 描述子, 在红外图像中提取的特征点, 分类效果不理想。

2) 由于红外有标记样本数量有限, 直接在分类网络上进行训练, 网络容易过拟合。如果基于可见光预训练模型进行微调, 由于分布存在差异, 分类效果不理想。

3) 使用迁移学习后, 能够实现目标域和源域在高层特征的分布对齐, 分类准确率有明显提升。另一方面, 通过加入红外图像的重构环节, 准确率提升了 3.16%。这说明通过红外图像的无监督学习, 对于红外域的特征表示更加有效, 能更真实反映红外域的特征分布, 提高迁移学习效果。

图 5 列出了在迁移学习样本集中原始红外图像和使用红外自编码器重构出的图像。第一行为原始红外图像, 第二行为重构后的红外图像。可以看出, 重构出的红外图像基本保留了原始图像的信息, 这说明红外图像自编码器在重构过程中没有丢失红外域图像特征。

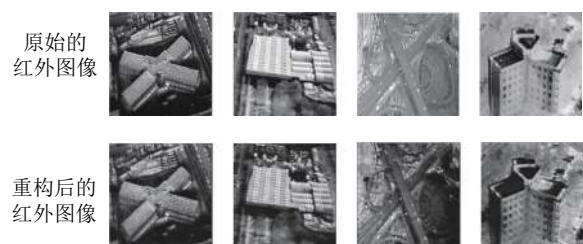
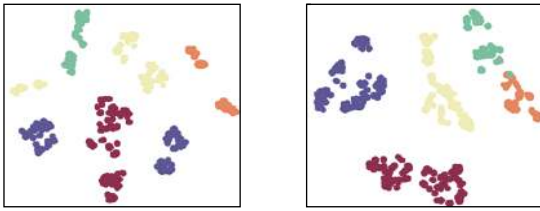


图 5 红外输入图像和重构图像示例

为了进一步分析本文所提出的方法效果优于传统方法的原因, 列举了红外分支和可见光分支的 5 种类别在 FC8 层的特征, 通过 t-SNE^[19] 进行了可视化。不同的颜色代表了不同的类别特征。从图 6 可以看出, 基于参数微调的方法并不能保证两个域中相同类别特征分布对齐。本文方法实现了红外和

可见光中同类目标的分布相似,使得红外目标的不同类之间具有更好的区分性。



a. 未采用迁移学习的特征分布 b. 采用重构迁移学习的特征分布

图6 迁移学习前后的特征分布可视化

5 结束语

在红外图像分类中,针对红外图像中样本过少和样本不同分布的问题,本文提出了一种基于重构迁移学习的方法,充分使用了大量的无监督红外样本,提高了红外域特征表示能力,通过迁移学习实现了红外域图像和可见光图像特征分布相似。通过上述改进,相比于目前广泛使用的参数微调,本文方法的分类准确率提升了 11.27%。

参 考 文 献

- [1] CHENG Kai-sheng, LIN Hwei-yung. Automatic target recognition by infrared and visible image matching[C]// 2015 14th IAPR International Conference on Machine Vision Applications. Piscataway, NJ, USA: IEEE, 2015: 312-315.
- [2] 张迪飞, 张金锁, 姚克明, 等. 基于 SVM 分类的红外舰船目标识别[J]. *红外与激光工程*, 2016, 45(1): 104004.
ZHANG Di-fei, ZHANG Jin-suo, YAO Ke-ming, et al. Infrared ship-target recognition based on SVM classification[J]. *Infrared and Laser Engineering*, 2016, 45(1): 104004.
- [3] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.
- [4] BAY H, TUYTELAARS T, GOOL L J V. SURF: Speeded up robust features[C]//9th European Conference on Computer Vision. Graz, Austria: Springer-Verlag, 2006: 404-417.
- [5] RUBLEE E, RABAUD V, KONOLIGE K, et al. ORB: An efficient alternative to SIFT or SURF[C]//2011 International Conference on Computer Vision. Piscataway, NJ, USA: IEEE, 2011: 2564-2571.
- [6] 李炯, 雷虎民. 一种基于红外图像的目标自动识别算法[J]. *航空计算技术*, 2005, 35(4): 26-28.
LI Jiong, LEI Hu-min. A method of automated recognition and classification based on infrared images[J]. *Aeronautical Computer Technique*, 2005, 35(4): 26-28.
- [7] 李瑞东, 孙协昌, 李勤. 空间目标红外特征提取与识别技术[J]. *红外技术*, 2017, 39(5): 427-435.
LI Rui-dong, SUN Xie-chang, LI Meng. Infrared feature extraction and recognition technology of space target[J]. *Infrared Technology*, 2017, 39(5): 427-435.
- [8] SHAIK J S, IFTEKHARUDDIN K M. Automated tracking and classification of infrared images[C]// Proceedings of the International Joint Conference on Neural Networks. Piscataway, NJ, USA: IEEE, 2003: 1201-1206.
- [9] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of The ACM*, 2012, 60(6): 84-90.
- [10] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition. [EB/OL]. [2019-05-15]. <https://arxiv.org/pdf/1409.1556.pdf>.
- [11] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2016: 770-778.
- [12] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 7132-7141.
- [13] YOSINSKI J, CLUNE J, BENGIO Y, et al. How transferable are features in deep neural networks[C]// Advances in Neural Information Processing Systems. Montreal, Canada: Curran Associates Inc, 2014: 3320-3328.
- [14] LONG M, CAO Y, WANG J, et al. Learning transferable features with deep adaptation networks[C]//International Conference on Machine Learning. Lille, France: IMLS, 2015: 97-105.
- [15] LONG M, WANG J, DING G, et al. Transfer feature learning with joint distribution adaptation[C]//IEEE International Conference on Computer Vision. Washington DC, USA: IEEE, 2013: 2200-2207.
- [16] HINTON G E. Reducing the dimensionality of data with neural networks[J]. *Science*, 2006, 313(5786): 504-507.
- [17] GRETTON A, SEJDINOVIE D, STRATHMANN H, et al. Optimal kernel choice for large-scale two-sample tests[C]//Advances in Neural Information Processing Systems. Nevada, USA: Curran Associates Inc, 2012: 1205-1213.
- [18] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(12): 2481-2495.
- [19] MAATEN L VAN DER, HINTON G E. Visualizing data using t-SNE[J]. *Journal of Machine Learning Research*, 2008, 9: 2579-2605.

编辑 刘飞阳