

# 第一章 绪论

## 1.1 研究背景及意义

为了解决任意单一源图像无法提供全面可信度高的信息的问题，获取更加丰富的信息以便于加强对场景的描述，图像融合技术得到了广泛的应用<sup>[1]</sup>。从 20 世纪 80 年代开始，由于前端传感器技术的发展导致图像数据量猛增，而不同图像传感器所采集到的信息均具有各自的特征，成像机理及环境等条件的不一致导致了不同图像所包含信息的特殊性，单一源图像无法全面反映场景的所有特征，存在片面性，难以满足后续研究需求。面临着庞大的数据和繁杂的应用，迫切需要获取关于特定场景的全面信息的图像融合技术<sup>[2]</sup>。于是利用多个传感器对同一场景中的同一时刻的数据进行采集，或对同一场景中的不同时刻的数据进行采集，再进行融合操作。这样操作者就无须单独检索每幅源图像，而只需分析一幅经过融合后的图像便可以全面掌握特定场景的信息，从而减少了人工时间成本。另外，信号处理理论的进一步发展完善为融合算法设计奠定了理论基石，从而导致有关图像融合技术的基础理论研究成果更加广泛，图像融合技术也在现实应用中发挥着更加关键的作用。

图像融合<sup>[3]</sup>(Image Fusion)是用特定的算法将两幅或多幅图像综合成一幅新的图像。融合结果由于能利用两幅(或多幅)图像在时空上的相关性及信息上的互补性，使融合后得到的图像对场景有更全面、清晰的描述。图像

融合是图像处理中重要部分，能够协同利用同一场景的多种传感器图像信息，输出一幅更适合于人类视觉感知或计算机进一步处理与分析的融合图像。它可明显的改善单一传感器的不足，提高结果图像的清晰度及信息含量，有利于更为准确、更为可靠、更为全面地获取目标或场景的信息<sup>[4]</sup>。按照输入源图像类别的差异，图像融合也可以有很多研究领域，如红外与可见光图像融合<sup>[5][6]</sup>、多聚焦图像融合<sup>[7][8][9]</sup>、低秩表示图像融合及基于深度学习的图像融合等，相关学者也都在这些领域进行了深入的研究。图像融合技术在国内外得到了越来越多的运用，人体感知<sup>[10]</sup>、图像视觉领域与机器人视觉<sup>[11]</sup>、卫星图像和遥感绘图<sup>[12][13][14]</sup>、安防<sup>[15][16]</sup>、数码摄影<sup>[17][18]</sup>等领域。

## 1.2 图像融合的研究现状

### 1.2.1 多聚焦图像融合的研究现状

近年来，对多聚焦图像进行传统邻域研究的主要是像素级方法。其又分为空间域方法和变换域方法。基于变换域的方法，首先提取源图像对其进行分解到指定的变换域，经过人工设计融合规则，最后是将变换后的图像系数来进行融合，根据融合规则进行逆变换，通过融合图像系数对融合图像进行重建，得到最终的融合图像。可以看出变换域的方法最重要的是两点，变换域系数的选择和融合规则的设计。

目前，基于变换域方法的主要有双数复小波变换(DTCWT)<sup>[19]</sup>、离散小波变换(DWT)<sup>[20]</sup>、非下采样轮廓波变换(NSCT)<sup>[21]</sup>、拉普拉斯金字塔(LP)<sup>[22]</sup>、

曲波变换(CVT)<sup>[23]</sup>、稀疏表示(SR)<sup>[24]</sup>。虽然以上方法有不错的效果,但是在融合过程中忽略了空间一致性,上述方法丢失空间信息较为严重。但是,变换域的方法是容易出现方向不确定性和数据繁琐以及冗余数据量大,所以经常会出现被融合后的图像有伪影和模糊的边界。

随着人们对小波分析的深入,发现它有新特性并且它具有良好的方向性和正交性。由于该算法在同一时间获得了图像的频率信息,因而在图像融合中的应用优于其它常规的融合算法。因此比很多传统方法都更适用于图像融合。Li 等人<sup>[25]</sup>首先引入了离散小波变换(DWT)进入图像融合领域。主要是通过对变换域提出的,主要包括有三个步骤。首先将原始图像通过活动水平的测量得到初始图,然后通过融合规则来得到新的结果图,最后通过后处理方法,就可以得到全聚焦的融合图像。该方法主要是通过局部区域求解的方法来得到活动水平,并且采用了最大选择的融合规则来计算,最后通过对系数的重构来得到高清图像。Rockinger<sup>[26]</sup>等人提出了基于位移不变 DWT 算法与其他方法相比也能得到不错的结果。希尔等人引入了基于双树复小波变换(DTCWT)的多聚焦图像融合方法,该方法可以提供近似的位移不变性和定向选择性。De 和 Chanda<sup>[27]</sup>提出了利用形态小波变换(MWT)的多聚焦图像融合方法,它本质上是由扩展和腐蚀等形态操作构造的 DWT 的非线性扩展。Lewis<sup>[28]</sup>等通过应用图像分割算法,将基于 DTCWT 的融合中基于像素的策略扩展到基于区域的策略。Tian 等<sup>[29]</sup>提出了利用小波系数分布的扩散进行基于小波的多聚焦图像融合的统计锐度测量方法。Aymaz 和 Kose<sup>[30]</sup>引入了基于主成分分析(PCA)的基于平稳小波变换(SWT)

的多聚焦图像融合方案。

基于空间域的图像融合方法在最近十年得到飞速的发展，同时也受到了很多研究者们的关注。与上述方法相比，基于空间域方法是在近聚焦图像和远聚焦图像上进行操作，因此更加简单，融合速度快，可以满足一些要求较低的场景。在空间域中可以分成基于像素方法，基于区域方法，基于块的方法。基于块的方法将大小一致的图像块对源图像进行分解，随后根据设计规则对其进行位置清晰度高的图像块对其进行计算得到拥有聚焦特性的图像块，将得到的图像块进行计算得到融合图像。该方法有一定的局限性，主要是块的尺寸切割不均匀导致带来了块效应。因此，Aslantas 等人<sup>[31]</sup>通过优化的方法随机选取合适的图像块尺寸，该算法用于解决图像块尺寸的融合问题，相对于常规的固定块，该方法能够较好地提高静态图像的可视化同时提高了自适应块静态图像融合的视觉质量，缺点是计算量大。随着基于区域的图像融合算法兴起<sup>[32]</sup>，该方法利用图像网络提取聚焦区域和非聚焦区域，然后根据后处理技术再聚焦到区域中，融合成最终的聚焦图像。由于硬件技术不成熟，使得该方法在精度上受到了很大的影响。随着各国对图像处理的大力支持下，出现了很多有代表性的方法，比如基于引导滤波(Guided Filter)<sup>[33]</sup>的算法，主要是通过像素提取，使其对聚焦区域的源图像像素进行计算，通过计算得出一个初始的决策图，最终的目的使其得到精确的分割出来的决策图，才能得到效果很好的融合图。两种像素级多聚焦融合传统算法相同之处是，几乎所有聚焦检测都依赖于高频信息，如梯度信息和边缘信息。2001 年李等人<sup>[34]</sup>最初对源图像进行分割使其

的尺寸大小不一的自适应块，然后采用空间频率技术作为对活动水平的计算，使其在融合规则下得到合适的块图像，最后使用后处理技术得到重建后的图像。并在活动水平测量、融合规则、块分割策略等方面进行了各种改进。Huang 和 Jing<sup>[35]</sup>对多聚焦图像融合中经常使用的聚焦度量进行了评价，包括方差<sup>[36]</sup>、强度势能<sup>[37]</sup>、LP 势能<sup>[38]</sup>、SML 势能<sup>[39]</sup>、空间频率(SF)等。研究者们对基于块的融合方法都是基于用户经验设置的固定块大小，仅仅是块的不同大小对最终的融合结果有至关重要的影响，基于固定尺寸大小的常用的方法在融合的图像中引入不良的阻塞或者残差效应，进一步提高源图像分割的灵活性。提出了基于区域的融合算法，基于区域的方法就不会以块为基础了而是区域进行计算。主要的最大的差别是活动水平实际测量是在每个不规则大小的分段区域进行，而不是块。Min 等<sup>[40]</sup>人引入了基于区域的多聚焦图像融合方法，基于 PCNN 的分割方法是为融合设计了一种共同使用显著性和可见性的活动测量方法。Li 和 Yang<sup>[41]</sup>提出了基于归一化分割和 SF 聚焦度量的多聚焦图像融合方法。Hao<sup>[42]</sup>等提出基于平均位移分割方法的多聚焦彩色图像融合方法，并采用 SML 作为活动度量。Yang 和 Guo<sup>[43]</sup>应用分水岭算法将源图像分割成多个像素，进行多聚焦图像融合。目前，大部分的像素算法都采用了像素最大权重框架设计，对融合图与原始图进行加权计算，其核心问题是获得每个源图像的权重图像。主流方法都是利用活动水平测量，聚焦测量在多聚焦图像融合中来应用去测试源图像的像素稳定性和鲁棒性。许多研究者将梯度图或者权重图来获取决策图的将其归为一个分类问题。然而，为了获得更准确的权值或分类结果。如 QWT<sup>[44][45]</sup>、ND 滤波<sup>[46]</sup>、ICA<sup>[47]</sup>、SR<sup>[48]</sup>、NSCT<sup>[49]</sup>、结构张量<sup>[50]</sup>等

技术都可以获取决策图。

## 1.2.2 红外与可见光图像融合研究现状与动态分析

由于图像融合技术和夜视技术的迅速发展，红外和可见光图像的融合技术已被广泛应用到各类现实情景中。在 2005 年，加拿大 Octec 集团成功地将红外线视频图像与自然彩色可见光图像融合在一起，并研发了一个给警用直升飞机使用的图像融合系统，其结果主要是图像具有可见光和自然彩色特征<sup>[51]</sup>。2009 年，美国军队在极低照度状态下，融合通过新一代增强型夜视镜头所获取的红外与可见光图像，大大提高了对目标的感知性能<sup>[52]</sup>。不仅如此，英美还开发了“追踪者”战术侦察车系统，为了快速准确地识别目标，判断目标位置及相距距离，融合红外传感器、激光测距传感器以及电视摄像传感器所获得的数据<sup>[53]</sup>。之后，TNO 研究所研发了一种融合系统，可以生成彩色的融合图像<sup>[54]</sup>。

对于红外与可见光图像融合技术的研究，我国经过长时间的发展也取得了一定成果。在卫星遥感领域，中国首先采用了在卫星上加装的 CCD 摄像机和红外线多光谱扫描仪收集图像，然后再加以融合处理<sup>[55]</sup>。同时，红外与可见光图像融合的研究也在国内外许多科研机构展开，并在实践中取得了良好的效果。北京理工大学在 2001 年开发了一套双声道彩色影像融合系统<sup>[56]</sup>。华中理工大学于 2005 年完成了一套用于红外双频带图像的融合系统。在这之后，南京理工大学也进行了尝试，改善了融合图像的特性，开发了一个融合系统，对红外线和弱可见光图像进行处理<sup>[57]</sup>。近年来，一直参与图像融合系统研究工作的单位还有中科院，图像融合性能已明显提高，

也在国内各高校得到了广泛应用。由此可见，研究红外与可见光图像融合技术十分有意义。

当前，随着对红外和可见光图像融合的研究日益增多，如何选择合适的融合算法是影响图像融合效果的关键因素。红外和可见光融合的不同方法，可以将其归类为基于多尺度变换<sup>[58][59]</sup>、低秩表示<sup>[60]</sup>、神经网络<sup>[61][62]</sup>、生成对抗网络<sup>[63][64]</sup>等其他方法。多尺度变换能有效实现图像不同尺度空-时域信息分离，可以将原始图像分解为不同尺度的子图，每个子图像有效体现真实图像不同强度结构、纹理信息，在图像融合中表现出较好的性能；基于低秩表示的融合方法主要将图像分为低秩部分和细节部分再进行后续处理，可以提取源图像的全局结构信息；基于神经网络的方法经过预训练，然后进行特征提取，容错能力强；基于生成对抗网络的融合方法无需进行预训练，由生成器与鉴别器进行对峙学习得到融合结果，该方法获得的融合图像性能更优。

### 1.2.3 低秩表示在图像融合中的研究现状与动态分析

低秩表示是以多维分析、压缩和可见性为基础的计算机可视化和图形分析的一个研究对象。当前所采用的低秩表示模型的关键在于：在稀疏性和低秩优化的条件下，矩阵的特征可以体现为稀疏和低秩的交迭。尤其是，如果矩阵信息取自于图像，稀疏成分和低秩成分将会分别表现出明显的特征和背景<sup>[65]</sup>。

在基于表示学习的红外与可见光图像融合算法中经常运用稀疏表示，

虽然可以获得较好的融合图像，但是也容易存在细节信息不足及获取全局结构的能力不足等问题。因此，低秩表示受到了人们的关注。低秩表示主要侧重于图像的主要特征表示，近年来在显著性检测、视频前景背景分离、图像去噪等一些相关领域进行了丰富的研究，特别是在图像处理领域也取得了一定的研究成果。

Hui 等<sup>[66]</sup>通过运用低秩表示方法可以提取源图像全局结构的特征，将字典学习与低秩表示相结合，设计了一种新的融合方法。这种方法既可以获得源图像的全局结构信息，又可以获得局部结构信息。但是低秩表示侧重于图像的主要特征表示，虽然可以获取源图像的全局结构特征，但对细节特性考虑不足，融合性能有待进一步提高。随后，Li 等<sup>[67]</sup>在红外和可见光图像的融合中使用了潜在低秩表示(Latent Low-Rank Representation, LatLRR)。首先通过 LatLRR 算法对源图像进行低秩部分和细节部分的分解，再分别利用不同的融合策略进行处理和重建。Cheng 等<sup>[68]</sup>人结合多尺度分解和低秩表示理论，将输入图像分解成高通子带和低通子带图像。然后在高、低通两个通道中分别提取高频分量以及低频分量；最后通过计算每个像素值与其对应的高维系数之间的相关系数矩阵，确定最优权值向量。并用 LatLRR 方法对该图像进行了分解，得到了一个有意义的显式图像。采用显式图的权值对低频段图像进行加权合并，并采用平均梯度法进行高通子频段的融合。文献<sup>[69]</sup>提出了一种基于潜在低秩表示和脉冲式耦合神经网络的多尺度融合模型，通过全新的融合规则，可将包含在红外图像内部的显著性特征与可见光图像的梯度特征完美结合，从而获取具有优良视觉效



果的融合图像。文献<sup>[70]</sup>提出了一种基于潜在低秩表示与复合滤波的红外与弱可见光增强图像融合方法，该方法得到的融合图像细节信息丰富，清晰度高，具有良好的可视性。潜在低秩表示技术在实现图像的主要特征表示的同时考虑图像细节信息的分解，若将潜在低秩表示应用于红外与可见光图像的融合中，势必进一步提高图像融合的性能。

### 1.2.4 深度学习在图像融合中的研究现状与动态分析

最近几年，深度学习逐步发展成现在最盛行的热门研究课题之一，也成为了人工智能领域中的一个关键课题。它是人工智能在自然语言处理和电脑视觉等应用领域中实现突破的关键。传统的融合方法实现起来比较复杂，往往需要人工设计融合方法和规则<sup>[71]</sup>。随着计算机视觉技术快速发展以及人工智能算法的不断完善，“机器换人”已经逐渐被广泛应用于各个领域。深度学习利用深度神经网络结构，模拟大脑的视觉机制对文字、音频或图形等信息做出理解。深层神经网络又可以叫做深度分层学习网络，利用逐层获得底层特性并整合得到更抽象的高层特性，学习一些较高层级的信息表达<sup>[72]</sup>。这种深度学习算法，能够实现对输入信号的识别。在此基础上，该网络的整体集成不依赖于人力，能够从网络中抽取出更多的深层次的特点，并将其最终的结果直接呈现出来，智能化程度更高，应用范围逐步扩大。它能用于图像分析、人脸识别、语音处理、自动驾驶、文字翻译、智能监控等场合中，呈现出卓越特性。

深度学习技术在红外与可见光图像融合任务中发挥着越来越重要的作用，融合图像的性能也越来越高，这都得益于网络模型的不断优化提升。

目前, 在处理图像融合任务中, 深度学习主要有三种网络结构, 分别为卷积神经网络、卷积稀疏表示及堆栈自编码<sup>[73]</sup>。2015年, Huang 等人<sup>[74]</sup>通过构建堆栈自编码网络使得全色图像的细节特征及多光谱图像的光谱特性得以保留与综合。Yu 等人<sup>[75]</sup>提出了一个基于卷积稀疏表示的融合框架, 该方法可有效利用图像的稀疏性特征, 有助于融合图像的代表。但最常用的还是卷积神经网络及其相应的改进模型, 网络模型的不断优化, 带来网络性能的逐步提升。2016年, Liu 等人<sup>[76]</sup>使用深度卷积神经网络结构处理多聚焦图像融合任务, 同时第一次在红外和可见光图像融合任务中使用 CNN。然而, 红外和可见光图像的特征及像素值分布不一样, 使用这种空间域的方法融合结果性能较差。文献<sup>[77]</sup>提出了一种基于卷积神经网络的红外与可见光图像融合算法, 该方法可以较好的将特定尺度信息进行保存, 并减小滤波对边缘细节带来的光晕影响, 融合后图像噪声较少, 细节呈现的更加自然。由于合理提高卷积网络深度可以提高网络信息提取能力, Hui 等人<sup>[78]</sup>使用 VGG-19<sup>[79]</sup>处理红外和可见光图像融合任务, 将中间层的信息利用到后面层次的信息提取中, 可以减少信息丢失, 获得深度特征。文献<sup>[80]</sup>提出了一种有效的图像融合方法, 将多尺度变换与 VGG 网络相结合, 生成了包含更多红外与可见光图像的有效信息的融合图像。不过 VGG 网络仍存在多层结构过于简单, 无法合理利用所有信息的问题。基于密集连接块图像特征丢失较少并且减少过拟合现象等的优点, Li 等人<sup>[81]</sup>设计了一种密集融合结构用于红外与可见光图像融合。该结构主要分为编码网络、融合层和解码网络三个部分, 三者各自承担不同的工作协同完成融合任务。文献<sup>[82]</sup>提出了一种基于 ResNet152 深度学习模型的红外与可见光图像融合算法, 该

算法在保留图像显著特征的同时使融合图像获得了更多的细节纹理信息，且有效地降低了伪影和噪声。2018年，Ma等人<sup>[83]</sup>创新性地将GAN引入图像融合，从而更好地引导网络通过对抗性学习和特定内容丢失，在无监督的情况下保留显著的对比度和纹理细节。随后，Ma等人<sup>[84]</sup>基于FusionGAN使用多分类约束和专门设计的内容丢失，考虑源图像中包含的二次信息，生成具有良好视觉效果的融合图像。文献<sup>[85]</sup>提出了一种基于双通路生成对抗网络的红外与可见光图像融合方法，通过对抗方式训练生成网络和鉴别网络，并将训练的生成网络作为最终的图像融合模型，该方法得到的融合图像对比度更高，细节纹理更丰富，且能良好地契合人类的视觉感知。基于生成对抗网络的方法通过特定的内容丢失来提取源图像的信息，不需要进行模型预训练，算法效率高。但在网络结构的设计中对浅层网络的信息提取考虑不足，若能提高生成对抗网络在训练过程中对信息的利用程度，势必进一步提高图像融合的性能。

### 1.3 研究的主要内容

本文首先对多聚焦图像融合方法和红外与可见光图像融合方法进行了相关背景和意义的阐述，总结了两个算法在各自领域的应用和研究现状。

第二，对于多聚焦图像融合方法，我们进行了研究和分析，分析了多聚焦图像融合成像的机理与特征，并对多聚焦图像融合方法进行了对比，分析了两者的联系。本文对像素级方法和深度学习方法进行了概述和分析。还介绍和分析了目前多聚焦图像融合结果质量评价。对于红外与可见光图像融合方法，为了减弱低秩表示在图像融合任务中容易造成细节信

息模糊的问题，针对图像样本少的情况，利用潜在低秩表示技术兼顾考虑图像的主要特征与细节信息的优点，在红外与可见光图像融合中引入潜在低秩表示技术，建立一种高效的细节分解的潜在低秩表示红外与可见光图像融合方法。该方法具有良好的融合性能，可以更好地保留源图像中的全局结构和局部结构信息。

第三，在多聚焦图像融合算法中，研究了引导滤波器和图像抠图技术的理论，并分析了引导滤波器和图像抠图技术的理论。本文构建了一种利用引导滤波器与 KNN-Matting 相结合的多聚焦图像融合算法。该方法首先通过中值滤波器对源图像进行重复模糊处理，并使用差分算子获得图像的梯度信息。在此基础上，利用引导滤波器对梯度图像进行最优解计算，得到了初始的决策图。为了获得更准确的决策图，本文对初始决策图进行膨胀和腐蚀操作生成图像抠图中的三值图，并把三值图和原始图像送入 KNN-Matting 抠图算法中进行优化得到精确决策图。在公开的数据库上进行实验，采用了基于主观性和客观性的评估方法，并将该方法与其他主要的融合算法相比较。实验结果表明，所提出的方法具有较高的精度和效率，该算法能够获得清晰度高的图像和不错的融合效果。而在红外与可见光图像融合中，为了进一步实现红外和可见光图像信息的有效融合，改善目前融合算法存在的设计繁杂、实施困难与数据计算量大等问题，利用生成对抗网络技术兼顾考虑对比度信息和纹理细节的优点，建立一种深层次多分类的生成对抗网络红外与可见光图像融合方法。

第四，在多聚焦图像融合算法中构建了一种以无监督的学习方式来训

练并引入了多尺度残差网络的模型。该模型从输入的源图像获取深层特征，利用空间频率计算深层特征的活动水平，获得初始决策图。进而利用一致性验证方法优化初始决策图与源图像，计算出最终的融合结果图。

第五，对多聚焦图像融合算法和红外与可见光图像融合算法分别进行了软件设计，方便用户对所设计的算法进行调用。

第六，进行全文的总结与展望。对所做工作归纳说明并深入探讨。