

融合智能

人工智能时代的图像融合技术

版本号：v25.04.28

张星辰 著

英国埃克塞特大学

2025 年 4 月

献给我的家人

前 言

每个人都喜欢完美。然而，世间万物不可能有完美的。一种重要的弥补方式，就是融合。作为资深金庸迷，笔者发现，融合的思想在金庸先生的笔下体现得淋漓尽致。不论是《神雕侠侣》中杨过和小龙女的双剑合璧，还是《侠客行》中石破天和白万剑的双剑合璧，还是《倚天屠龙记》中华山派和昆仑派的正反两仪剑法的融合，还是《雪山飞狐》中胡斐的一对双胞胎徒弟的互补剑法，均体现了融合的思想。

融合，可以拓宽人类感知的局限，可以克服相机原理的限制，也可以在实际应用中直接帮助人们。在实践中，由于各种成像传感器的限制，单一图像往往不能充分全面地反映一个场景的信息。图像融合是指将多幅图像中的有用信息提取出来生成一幅融合图像或者进行更好的决策的过程。根据源图像类别的不同，图像融合主要包含可见光红外图像融合、多聚焦图像融合、多曝光图像融合、医学图像融合和遥感图像融合。本质上，这些图像融合任务都是为了实现“兼听则明”，即利用不同图像中的互补信息以获得更高质量的图像或者进行更好的决策。

图像融合技术多年来一直是研究热点，并且在许多领域有着重要应用，例如目标跟踪、目标检测、生物信息识别、医学图像识别、深度估计、图像美化程序等。大约从 2017 年开始，深度学习技术被引入到图像融合领域，并进一步促进了图像融合领域的发展。截止到 2023 年 9 月，已有大量基于各种深度学习模型和算法的图像融合文章在学术期刊和会议上发表，引起了广泛关注，也吸引着越来越多的研究人员开始着手相关研究。

然而，现有的图像融合著作基本上都是多年以前出版的，其中并未涉及到深度学习相关的内容。此外，现有的图像融合著作主要关注理论方法层面，对图像融合的实践着墨不多。有鉴于此，笔者深感一部关注基于深度学习的图像融合方法及实践的著作是非常有必要的。本书的写作目的，即在于此。本书取名为《融合智能》，旨在强调本书关注的是基于以深度学习为代表的人工智能的图像融合方法。

本书主要包括三个部分。第一部分是背景与概念，包括第一章到第二章。其中，第一章是绪论，主要介绍图像融合方面的一些基础知识，如图像融合的基本概念、图像融合的分类、本书的写作目的等。第二章介绍深度学习基础，包括深度学习发展情况简介、深度学习基础知识和常用的深度学习框架等内容。第二部分是方法与技术，包括第三章到十章。其中，第三章总体介绍基于深度学习的图像融合，包括其必要性和发展状况，以及常用于图像融合的深度学习模型。第四章介绍图像融合算法的性能评价方法，包括定性评价方法和定量评价指标。第五章至第七章各介绍一个图像融合类别，分别是可见光与红外图像融合、多聚焦图像融合、多曝光图像融合。在介绍每一个图像融合类别时，均首先给出问题的定义，然后简单介绍传统方法，再重点介绍近年来迅速发展的基于深度学习的方法。第九章介绍通用图像融合方法，即可以同时应用于几种图像融合任务的融合方法。第十章介绍应用驱动的图像融合方法。本书第三部分是应用、实践与展望，主要包括第十二章到第十六章。其中，第十二章介绍可见光与红外图像融合的应用，第十三章介绍多聚焦图像融合的应用，第十四章介绍多曝光图像融合的应用。然后，第十一章介绍图像融合实践，包括编程语言的选择和评价基准的使用。第十五章总结介绍图像融合领域的前言进展。最后，第十六章对全书进行总结，并对图像融合领域的发展进行展望。

除了正文以外，笔者在附录A中还列出了图像融合相关的学术期刊和会议，便于读者朋友在进行论文投稿时参考。在附录B中，笔者给出了笔者的 Github 链接。读者朋友们可以从该链接中找到许多开源图像融合算法的下载链接，以便于开展研究工作。在附录C中，笔者结合自己多年的论文写作和审稿经验，简单介绍了一下论文的写作经验，以供相关读者参考。

需要说明的是，由于近年来发表的图像融合论文很多，本书只能挑选少数主要的方法来进行介绍。笔者在选取参考文献时，主要选择那些发表在顶会或者知名学术期刊上的文章。

此外，笔者认为图像融合领域需要取得更好的发展，有两个问题必须要解决。一是找到合适的应用，二是开发出合理的评价方法和评价基准。本书也介绍了笔者在这两个方面做的一些探索。

本书可作为图像融合领域的学生、科研人员和相关从业人员的参考书。本书也适合对图像融合领域感兴趣的学生和科研人员参考。

由于基于深度学习的图像融合发展非常迅速，因此笔者在写作的过程中需要经常更新内容，甚至将之前已写好的内容进行重新写作。笔者已尽了自己最大

的努力将本书写好。然而，由于图像融合领域发展迅速，加之笔者水平有限，书中难免存在不足和错误之处，敬请读者不吝批评指正。这也是笔者不断前进的动力。本书相关内容的更新和勘误会发表在微信公众号“笑书神侠读博学”上。欢迎读者朋友们关注并以各种方式与笔者交流。我们一起努力把下一版变得更好。

季羡林先生说：“出书必定要有用”。希望本书是一本有用的书。

常用缩略词

CNN	Convolutional Neural Network
CVPR	The IEEE/CVF Computer Vision and Pattern Recognition Conference
DDPM	Denoising Diffusion Probabilistic Model
ECCV	European Conference on Computer Vision
GAN	Generative Adversarial Network
GPU	Graphics Processing Unit
ICCV	International Conference on Computer Vision
MEFB	Multi-exposure image fusion benchmark
MFIFB	Multi-focus image fusion benchmark
MSE	Mean-squared error
NIR	Near-infrared image
NLP	Natural Language Processing
OTB	Online tracking benchmark
RGB	Red-Green-Blue
SOTA	State-of-the-art
VIF	Visible-infrared image fusion
VIFB	Visible-infrared image fusion benchmark
VOT	Visual online tracking

目 录

前 言	II
第一部分 图像融合背景与概念	1
第一章 绪论	1
1.1 引言	1
1.2 图像融合的基本概念	2
1.3 图像融合的一个特点和两个目的	3
1.4 图像融合的分类	4
1.5 图像融合中的配准	9
1.6 基于深度学习的图像融合	10
1.7 本书的写作目的	11
1.8 本书主要内容与特色	12
第二章 人工智能基础知识	15
2.1 什么是深度学习?	15
2.1.1 优化问题	17
2.1.2 基于梯度的优化	17
2.2 深度学习三要素	17
2.2.1 数据	17
2.2.2 算力	18
2.2.3 算法	19
2.3 深度学习的分类	19
2.3.1 监督学习	19
2.3.2 无监督学习	20

2.3.3 强化学习	20
2.3.4 分类问题和回归问题	21
2.3.5 判别式模型和生成式模型	22
2.4 深度学习算法的常规设计流程	22
2.5 图灵测试	23
2.6 常用深度学习框架简介	24
2.7 小结	27
第二部分 图像融合方法与技术	28
第三章 基于人工智能的图像融合概述	29
3.1 传统图像融合方法简介	29
3.1.1 图像融合方法的三个步骤	30
3.1.2 传统图像融合方法的类别	30
3.1.3 传统图像融合方法的缺点	31
3.2 基于深度学习的图像融合发展状况概述	31
3.2.1 基于深度学习的目的	31
3.2.2 有监督方法和无监督方法	31
3.3 常用于图像融合的深度学习模型	32
3.3.1 卷积神经网络	32
3.3.2 Transformer	32
3.3.3 生成对抗网络	33
3.3.4 扩散模型	33
3.4 常用于图像融合的重要深度学习技术	34
3.4.1 注意力机制	34
3.4.2 残差连接	34
3.4.3 稠密连接	34
3.4.4 自动网络架构搜索	35
3.4.5 其他重要技术	35
3.5 与多模态机器学习的关系	35
3.6 基于深度学习的图像融合发展趋势	36
3.6.1 多种深度学习模型被用于图像融合	36

3.6.2 从非端到端的方法到端到端的方法	37
3.6.3 从特定图像融合方法到通用图像融合方法	37
3.6.4 从生成融合图像到改进下游应用	37
3.6.5 新型图像融合类型开始出现	38
3.7 小结	38
第四章 图像融合算法性能评价	39
4.1 图像融合算法评价的特殊性	39
4.2 当前的主要图像融合评价方法	40
4.2.1 图像融合定性评价方法	41
4.2.2 图像融合定量评价方法	42
4.2.3 图像融合评价方法现状	44
4.3 其他评价方法	45
4.4 近年来的发展特点	47
4.5 图像融合评价方法的发展趋势	54
4.5.1 设计更好评价基准	54
4.5.2 基于具体应用的性能评价	55
4.6 小结	56
第五章 可见光与红外图像融合	57
5.1 红外图像：从另一个视角感知世界	57
5.2 可见光与红外图像融合概述	59
5.3 传统融合方法概述	61
5.4 使用深度学习做图像融合的动机	62
5.5 基于深度学习的融合方法发展历程概述	62
5.6 基于深度学习的可见光与红外图像融合方法分类	63
5.7 基于深度学习的可见光与红外图像融合方法介绍	63
5.7.1 单分支模型和双分支模型	64
5.7.2 基于卷积神经网络的图像融合方法	64
5.7.3 基于自编码器的图像融合方法	69
5.7.4 基于生成式对抗网络的图像融合方法	70
5.7.5 基于 Transformer 的图像融合方法	74
5.7.6 基于扩散模型的图像融合方法	75

5.8 可见光与红外图像融合的发展特点	76
5.9 未来发展趋势	82
5.10 小结	86
第六章 多聚焦图像融合	87
6.1 多聚焦图像融合概述	87
6.2 传统融合方法概述	88
6.3 基于深度学习的融合方法	89
6.3.1 基于深度学习的多聚焦图像融合方法的分类	89
6.3.2 基于有监督学习的融合方法	90
6.3.3 基于无监督学习的融合方法	94
6.4 训练数据的获取	95
6.4.1 有监督多聚焦图像融合方法	95
6.4.2 无监督多聚焦图像融合方法	96
6.5 多聚焦图像融合的发展趋势	96
6.6 小结	98
第七章 多曝光图像融合	99
7.1 多曝光图像融合概述	99
7.2 传统融合方法概述	100
7.3 多曝光图像融合的特点	100
7.4 多曝光图像融合方法的分类	100
7.4.1 融合两张源图像的方法和融合多张源图像的方法	100
7.4.2 静态图像融合和动态图像融合	101
7.4.3 传统方法和深度学习方法	102
7.5 基于深度学习的融合方法	102
7.5.1 有监督学习方法	102
7.5.2 无监督学习方法	103
7.6 多曝光图像融合的发展趋势	106
7.7 小结	107
第八章 医学图像融合	108
8.1 概述	108
8.2 问题定义	109

8.2.1 医学图像融合的类别	109
8.2.2 医学图像融合的关键点	109
8.3 传统融合方法概述	109
8.4 基于深度学习的融合方法	109
8.4.1 发展历程概述	109
8.4.2 基于有监督学习的融合方法	109
8.4.3 基于无监督学习的融合方法	109
8.4.4 训练数据的获取	109
8.4.5 3D 医学图像融合	109
8.5 小结	109
 第九章 通用图像融合方法	 110
9.1 传统通用图像融合方法	110
9.2 基于深度学习的通用图像融合方法	111
9.2.1 概述与分类	111
9.2.2 实现方式	111
9.3 通用图像融合方法的优缺点	113
9.3.1 优点	113
9.3.2 缺点	113
9.4 小结	114
 第十章 应用驱动的图像融合方法	 115
10.1 应用驱动的图像融合方法的优势	116
10.2 应用驱动的可见光与红外图像融合方法	117
10.2.1 行人检测驱动的图像融合方法	117
10.2.2 语义分割驱动的图像融合方法	118
10.2.3 通用目标检测驱动的图像融合方法	119
10.3 应用驱动的其他图像融合方法	119
10.4 小结	120
 第三部分 图像融合的实践与展望	 122
第十一章 图像融合实践	123

11.1 编程语言及深度学习框架选择	123
11.1.1 编程语言选择	123
11.1.2 深度学习框架选择	124
11.2 使用 VIFB 进行可见光红外图像融合	125
11.3 代表性图像融合方法的使用	128
11.3.1 FusonGAN	128
11.3.2 DenseFuse	128
11.3.3 Dif-Fusion	128
第十二章 可见光与红外图像融合的应用	129
12.1 红外图像的常见应用总结	129
12.2 红外图像的缺点	132
12.3 像素级可见光和红外图像融合的应用	133
12.4 其他层级的可见光与红外图像融合的应用	136
12.4.1 基于特征级融合的应用	136
12.4.2 基于决策级融合的应用	139
12.4.3 基于多个层级融合的应用	139
12.5 可见光与红外图像融合的应用小结	140
12.6 展望	140
12.7 小结	144
第十三章 多聚焦图像融合的应用	145
13.1 多聚焦图像融合的应用概述	145
13.2 基于多聚焦图像融合的远距离人脸检测	147
13.3 基于多聚焦图像融合的光学显微图像融合	148
13.4 基于多聚焦图像融合的深度估计	149
13.5 展望	150
13.6 小结	152
第十四章 多曝光图像融合的应用	153
14.1 多曝光图像融合的应用概述	153
14.2 基于多曝光图像融合的语义分割	153
14.3 提升显微图像质量	155
14.4 小结	157

第十五章 图像融合的前沿进展	158
15.1 与其他任务相结合	158
15.2 通用图像融合方法	158
15.3 关于评价基准的研究	159
15.4 基于具体应用的融合方法性能评价	159
15.5 将图像配准和图像融合进行结合	160
15.6 其他类型的图像融合	161
15.6.1 可见光图像与近红外图像融合	161
15.6.2 偏振图像融合	162
15.6.3 RGB 图像融合	162
15.6.4 RGB 图像和 Optimal Waveband 图像融合	163
15.6.5 可见光图像与深度图像融合	163
15.6.6 可见光图像与事件相机数据融合	164
15.6.7 多视角图像融合	165
15.7 其他应用	165
15.8 小结	165
第十六章 总结与展望	166
16.1 总结	166
16.2 待解决的问题	167
16.3 展望	169
附录 A 图像融合相关的学术期刊和学术会议	175
附录 B 图像融合相关开源代码下载链接	177
附录 C 图像融合论文写作经验	178
索 引	182
参考文献	183

第一部分

图像融合背景与概念

第1章

绪论

互补思想是图像融合的核心。

——笔者

本章为全书的绪论，主要对图像融合的基本概念和本书的主要内容做概括性的介绍，以便读者掌握相关背景知识和对全书有概括性的了解。

1.1 引言

杨过也乘机接回长剑，适才这一下当真是死里逃生，但人当危急之际心智特别灵敏，猛地里想起：“我和姑姑二人同使玉女剑法，难以抵挡。但我使全真剑法，她使玉女剑法，却均化险为夷。难道心经的最后一章，竟是如此行使不成？”当下大叫：“姑姑，‘浪迹天涯’！”说着斜剑刺出。小龙女未及多想，依言使出心经中所载的“浪迹天涯”，挥剑直劈。两招名称相同，招式却是大异，一招是全真剑法的厉害剑招，一招是玉女剑法的险恶家数，**双剑合璧，威力立时大得惊人**，金轮法王无法齐挡双剑击刺，向后急退，嗤嗤两声，身上两剑齐中。

...

金轮法王见二人剑招越来越怪，可是**相互呼应配合，所有破绽全为旁边一人补去**，厉害杀着却是层出不穷。

——金庸《神雕侠侣》第十四回 礼教大防

石破天内力强劲之极，所学武功也十分精妙，只是少了习练，更无临敌应变的经历，眼见敌招之来，不知该出哪一招去应付才是。他所学的金乌刀法，除了

最后一招之外，每一招都是针对雪山剑法而施，史婆婆传授之时，总也是和每招雪山剑法合并指点。此刻他心中慌乱，无暇细思，但见白万剑使什么招数，他便跟着使出那一招相应的招数，是以白万剑使“老枝横斜”，他便使“长者折枝”，白万剑使“双驼西来”，他便使“千钧压驼”。**那知这金乌刀法虽说是雪山剑法的克星，但正因为相克，一到联手并使之时，竟将双方招数中的空隙尽数弥合，变成了威力无穷的一套武功。**

——金庸《侠客行》第十回 太阳出来了

上面这两段话，分别出自金庸先生的著名小说《神雕侠侣》和《侠客行》。前一段话写的是金庸先生笔下一个非常著名的双剑合璧的例子，即杨过和小龙女的双剑合璧。在该例子中，杨过和小龙女单打独斗均打不过金轮法王，但是他们双剑合璧以后即可打败金轮法王了。类似地，第二个例子写的是《侠客行》中的双剑合璧。

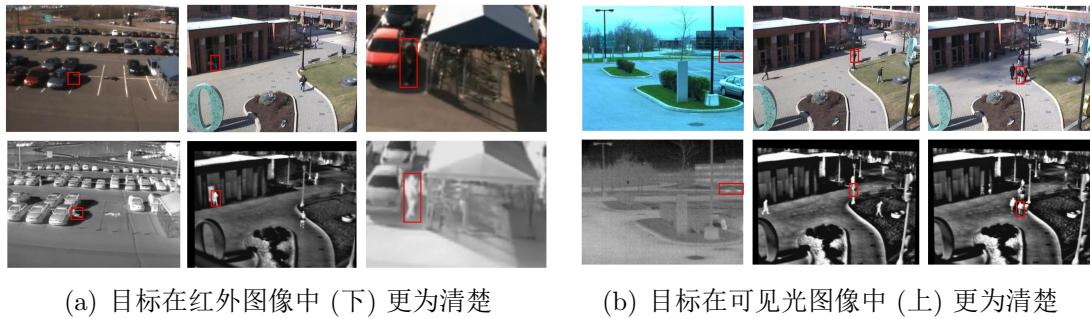
上面这两个例子均体现了互补的思想，即两种武功均各有缺陷，但是具有互补的特性。他们结合到一起以后（双剑合璧）互相弥补了对方的缺陷，从而威力大增。

互补是一种非常重要的思想。本书讨论的主题——图像融合，即是互补思想在图像处理领域的非常重要而且直观的体现。事实上，当双方优缺点都很明显，但彼此互补时，融合的作用最大。

1.2 图像融合的基本概念

受成像原理的限制，单一成像传感器是有缺陷的，正如引言中所述的全真剑法、雪山剑法等是有缺陷的。这种缺陷，使得由单一成像传感器在单次拍摄中获取的图像无法获取场景的全部信息，从而导致人或者计算机在基于该图像进行后续的处理时会出现问题。例如，我们生活中常见的照相机是可见光相机，这种相机在光线良好时可以捕获场景的主要信息。然而，当光线条件不好时，这种相机捕获场景信息的能力会急剧下降，从而使得后续的检测、识别、跟踪等操作无法很好地进行。

图1.1(a)给出了几个例子，即图像中的目标（用红色矩形框标记）由于处在阴影之中，导致在可见光相机捕获的图片中看不清楚这些目标。相反，在由红外相



(a) 目标在红外图像中(下) 更为清楚 (b) 目标在可见光图像中(上) 更为清楚

图 1.1: 可见光图像与红外图像性能互补的示例。图像来源于 [1]。

机¹捕获的图片中，则可以很清楚地看到这些目标。然而，红外相机虽然可以捕获热辐射信息，但是这种相机缺乏捕获图像细节的能力，也无法区分颜色，因此很多情况下在红外图像中也无法清楚地区分目标。例如，在图1.1(b)中，由于环境的干扰，由红色矩形框标记出来的目标在红外图像中不易区分，而在相应的可见光图像中则可以清楚区分。

上述例子表明了在不同工作环境下，可见光图像和红外图像的互补特性。这种互补特性还体现在其他类型的图像中，例如由于图像传感器景深有限导致的多聚焦图像、由于曝光程度不同导致的多曝光图像。如何充分利用这些互补的图像，以便获得质量更高的图像或者更好地完成检测、跟踪、识别等后续任务，是图像融合所要解决的问题。

如果要给图像融合下一个定义的话，笔者认为图像融合是指从两幅或多幅图像中获取互补的有用信息，形成质量更高的融合图像或者数据以便于人或者计算机可以更好的处理的过程。在实际应用中，不同类型的图像可以反映目标场景不同侧面的特性。通过对这些图像进行融合，可以获得同一目标场景较为准确全面的图像描述。

1.3 图像融合的一个特点和两个目的

笔者认为，图像融合具有一个特点和两个目的，如图1.2所示。

¹这里指远红外相机，即热像仪。这种相机基于物体的热辐射信息来成像。在本书中，如无特别说明，红外相机均指远红外相机

1. 图像融合的一个特点

图像融合领域有一个很重要的特点，即没有标准融合图像 (**ground truth**)。这一点与很多其他的计算机视觉任务（如目标跟踪、检测、图像分割）等是不一样的。没有标准融合图像，给图像融合任务带来了不少麻烦。首先，因为没有标准融合图像，所以在对深度学习模型进行训练的时候，缺乏有效的训练标签。虽然有一些科研人员通过一些方法来构造了伪标准融合图像（例如，利用多种现有方法得到融合图像，然后从中选择得到伪标准融合图像），但是这种伪标准融合图像毕竟不像真正的标签一样有用。其次，缺乏标准融合图像，也给图像融合算法的性能评价带来了很多困难。关于性能评价方面的困难，我们会在**第四章**进行详细讨论。

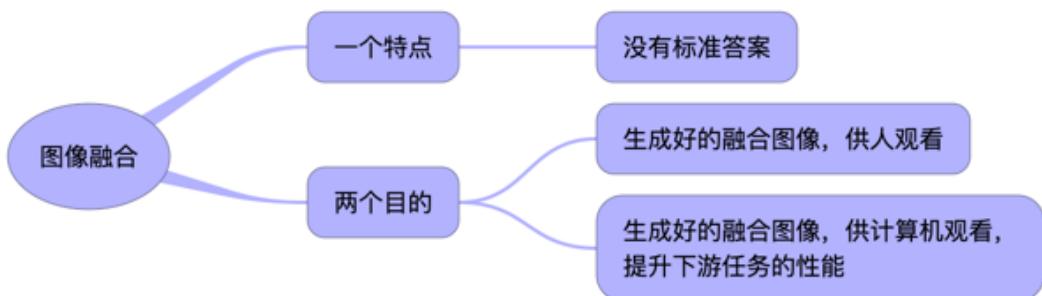


图 1.2: 图像融合的一个特点与两个目的。

2. 图像融合的两个目的

一般来说，图像融合有两个目的：一是生成好的融合图像供人观看，二是通过生成融合图像，提升下游任务的性能。第二个目的实际上也可以理解为给计算机“观看”。

图像融合在许多领域有着重要应用，例如目标跟踪、目标检测、生物信息识别、医学图像识别、深度估计、图像美化程序。因此，图像融合多年来一直是图像处理和计算机视觉中的研究热点。

1.4 图像融合的分类

图像融合有不同的分类方法，例如可以按照源图像类型分类，也可以按照融合的层次和融合方法进行分类。本节对图像融合的分类情况进行简单介绍。

1. 按源图像类型分类

根据源图像的不同，常见的图像融合类型主要包括可见光与红外图像融合、多聚焦图像融合、多曝光图像融合、医学图像融合和遥感图像融合。其中多聚焦图像融合和多曝光图像融合为单模态图像融合，而其他三种为多模态图像融合。本书仅关注可见光与红外图像融合、多聚焦图像融合和多曝光图像融合。

1) 可见光与红外图像融合

可见光与红外图像融合主要是指对由可见光相机和红外相机捕获的图像进行融合的过程。可见光与红外图像融合的主要目的，是为了将可见光与红外图像的互补特征提取出来并保留在融合图像中，从而可以使得融合图像免受光照条件等因素的影响，并且使得融合图像中的目标更具有区分度。可见光与红外图像融合广泛应用于目标检测、目标跟踪和生物信息识别等应用中。图1.3展示了一个可见光与红外图像融合的示例。从图中可以看出，融合图像保留了可见光图像和红外图像的特点。特别地，处于汽车阴影中的人可以看得非常清楚。



图 1.3: 可见光与红外图像融合示例。源图像来源于图像融合评价基准 VIFB[2]，融合图像由笔者使用 MGFF 算法 [3] 生成。

2) 多聚焦图像融合

通常我们通过相机获取的图像只有处于景深范围内的部分是清晰的，在景深范围之外的部分则是模糊的。在图1.4(a)中，由于前景中的小孩在景深范围之内，故图像中小孩是清晰的，而背景中的娃娃和大人是模糊的。而在图1.4(b)中，由于背景在景深范围之内，故图像中背景里的大人和娃娃是清晰的，而前景中的小孩是模糊的。



图 1.4: 多聚焦图像融合示例。源图像来源于 Lytro 图像融合数据集 [4]，融合图像由笔者使用 CBF 算法 [5] 生成。

在实际应用中，无论是出于视觉欣赏角度，还是出于后续应用的角度，通常我们希望能够获得清晰的图像²。然而，由于相机景深的限制，人们通常很难做到在同一张图像中使所有物体均处于景深范围内。因此，这个任务一般无法在单次拍摄时直接完成。为了解决这个问题，多聚焦图像融合应运而生。多聚焦图像是指通过对两幅或者多幅各只有部分清晰的图像进行融合，从而获得完全清晰的图像（如图1.4(c)所示）的过程。由于多聚焦图像融合可以通过算法克服相机景深范围有限的限制，生成清晰的图像，因此受到了广泛关注。除此之外，因为景深可以体现物体和相机之间的距离，因此包含着距离信息。因此，基于多聚焦图像还有可能对场景进行深度估计。

3) 多曝光图像融合

相较于大自然的万千色彩，通常使用的相机的动态范围是较窄的，因此在单次拍摄中无法完全捕获场景的信息。动态范围越高，图像表现的层次越丰富。此外，通常拍摄的图像还经常会出现曝光不足和曝光过度的现象，如图1.5所示。图1.5(a)中的图像由于曝光不足，导致图像中很多重要信息显示不清楚，而图1.5(b)中同一场景的图像则由于曝光过度，同样导致很多重要信息显示不清楚。

由于在日常生活和实际应用中对高质量图像的强烈需求，多曝光图像融合应运而生。多曝光图像融合是指将在两种或多种曝光条件下捕获的图像进行融合，以获得高动态范围、高质量图像（如图1.5(c)所示）的过程。经多曝光图像融合生成的图像，不仅有助于人眼对场景的辨识，也有助于计算机进行边缘检测、图像分割等图像处理操作。这项技术弥补了普通数码摄像及显示器材的动态范围

²特例外，如刻意使背景虚化的人像照片

窄于现实场景的局限性，也有助于处理曝光不足和曝光过度的图像。



图 1.5: 多曝光图像融合示例。图像由笔者拍摄，融合图像由笔者使用 MGFF 算法 [3] 生成。

4) 小结

从上述几种类别的图像融合任务中，读者可以看出，在每种图像融合任务中都体现了互补的思想，即源图像具有互补的特点。通过互补，单一图像的缺点得到了弥补，多种图像的有效信息得以在融合图像中保留，从而图像质量更好、后续任务更容易进行。可以说，互补思想是图像融合的核心。

2. 按融合层次分类

按照融合进行的层次来分，图像融合算法主要分为像素级、特征级和决策级融合。以目标跟踪为例，图1.6中的三张图分别是在像素级、特征级和决策级进行融合。像素级图像融合是指直接对源图像进行融合的过程，并且后续应用一般是在融合图像的基础上进行。像素级图像融合准确性最高，但同时计算量也最大，并且对于源图像有严格的配准要求。特征级图像融合是指先分别提取源图像的特征，然后对特征进行融合，并基于融合特征来进行后续任务的过程。特征级融合属于中间层次的图像融合，其既保留了足够的信息，又减小了计算量，并且对源图像的配准要求低于像素级图像融合。决策级图像融合是指先基于每幅源图像进行目标跟踪等实际任务，得到结果，然后对这些结果进行融合以得到最终结果的过程。决策级图像融合是最高级的图像融合，具有计算量小且对源图像配准要求更低等特点。然而，决策级图像融合会丢失源图像的许多信息。

值得说明的是，本书主要关注像素级图像融合方法。

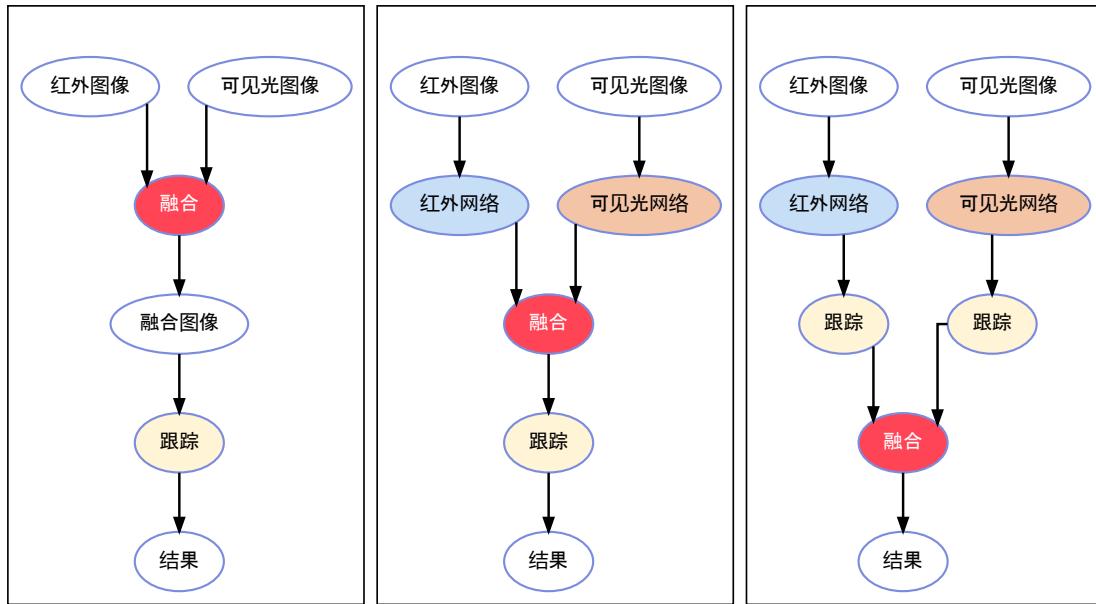


图 1.6: 不同层级的图像融合 (以目标跟踪为例)。

3. 按融合域分类

图像融合算法既可以在空间域完成，也可以在变换域完成。前者直接在空间域对图像进行融合，并且一般可以进一步分为基于像素的、基于块的和基于区域的融合方法。基于像素的图像融合方法是指直接对源图像的像素进行融合。例如，在多聚焦图像融合中，先判断源图像中每个像素的聚焦度（清晰度），然后根据聚焦度生成权值，最后对源图像的每个像素进行加权融合从而得到融合图像。基于块的融合方法是指首先将源图像划分为大小固定的图像块，然后以图像块为单位来进行融合。基于区域的图像融合方法是指首先使用图像分割算法将图像划分为大小不等或者不均匀的图像区域，然后以这些区域为单位进行融合。其中，基于块和基于区域的融合算法的效果非常依赖于图像块和图像区域的划分方式，并且容易在融合图像中的边界处（例如聚焦和不聚焦的分界处）产生瑕疵。

基于变换域的图像融合方法通常首先使用某种变换将源图像变换到另外一个域。在变换域中，源图像一般以系数的形式存在。然后，在变换域中，根据一定的算法或者规则，对源图像的系数进行融合。最后，对融合后的系数进行逆变换以得到融合图像。常见的变换有多尺度变换（包括小波变换、金字塔变换等）和稀疏表达等。

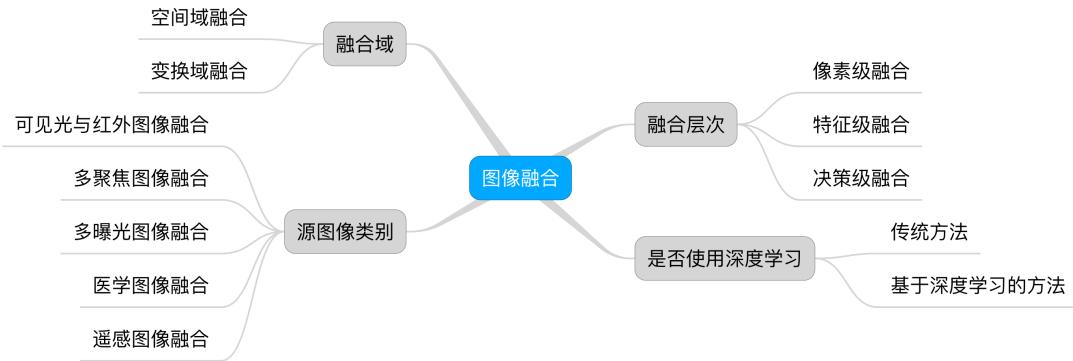


图 1.7: 图像融合主要分类方法总结。

4. 按是否使用深度学习分类

按照图像融合方法中是否使用了深度学习技术，图像融合方法可以被分为传统方法和基于深度学习的方法。具体地，在方法中使用了深度学习技术的图像融合方法，统称为基于深度学习的图像融合方法。其他方法则统称为传统方法。深度学习是在 2017 年左右被引入图像融合领域的，因此在 2017 年以前的图像融合方法，几乎均为传统方法。

5. 小结

上述几个小节介绍了图像融合的主要分类方法。图1.7对图像融合的主要分类方法进行了总结。需要注意的是，图像融合方法可能还有其他的分类方式。此外，不同的分类方式之间会存在交集。

1.5 图像融合中的配准

图像配准是指将图像对齐的过程，如图1.8所示。如前文所述，图像融合可以在像素级、特征级和决策级进行。这三种融合对于图像配准的要求依次递减：像素级融合要求最高，特征级融合次之，决策级再次之。



图 1.8: 图像配准示意图。

值得指出的是，目前用英文关键词“image fusion”搜索出的论文，绝大多数是像素级图像融合，并且目的是生成融合图像，一般并未考虑后续的应用。这一般包括红外可见光图像融合、多曝光图像融合、多聚焦图像融合、医学图像融合和遥感图像融合。因此，这类型的图像融合算法对于源图像的配准要求很高。

目前大多数基于可见光和红外图像进行行人检测和目标跟踪等任务的工作，是在特征级或者决策级进行的³，因此他们对于源图像配准的要求没有那么高，图像之间有些许偏差是可以接受的。当然，必须指出的是，图像之间的偏差会对跟踪和检测的结果有些许影响。在 2019 年的国际计算机视觉会议 (ICCV) 上，张等人 [6] 对从不同源图像提取的特征进行了对齐处理，并证明了这样可以提高目标检测的性能。

事实上，绝大多数开源的红外可见光图像数据集中的相当一部分图像并未严格配准。比如目标跟踪里的 GTOT[7] 和 RGBT234 数据集 [8]，还有著名的 Multi-spectral KAIST 数据集 [9]，其中都有很多图像没有严格配准。这主要是由于多模态图像之间的严格配准非常困难。

综上所述，在像素级图像融合中需要注意源图像是否严格配准了。在特征级或决策级的融合中，配准要求可以适当降低。

1.6 基于深度学习的图像融合

尽管深度学习技术从 2012 年开始就在计算机视觉领域大放异彩并得到了迅速发展，然而，一直到 2017 年左右深度学习技术才被引入到图像融合领域。自从深度学习技术被引入到图像融合领域以来，基于深度学习的图像融合方法引起了研究人员的广泛关注。据笔者统计，目前已有数百篇基于深度学习的图像融合相关的论文在国际期刊和学术会议上发表，且相当一部分是发表在顶级期刊和学

³像素级融合跟踪笔者也做过

术会议。如图1.9所示，仅是多聚焦图像融合这一个方向，截止到2020年5月份，就已经有超过50篇基于深度学习的文章发表，并且在逐年迅速增加。图1.10中也显示了截止到2022年9月份基于深度学习的可见光与红外图像融合相关论文的发表情况。

深度学习之所以在图像融合领域如此受关注，主要是因为深度学习可以解决图像融合中的几个重要问题。首先，深度学习可以提供相较于人工特征更好的特征，而特征在图像融合方法中起着至关重要的作用。其次，深度学习可以自动地学习融合规则，而不是像传统方法那样由研究人员来设计融合规则。例如，深度学习可以提供自适应的权值，这些权值在图像融合算法中用于对像素、图像块或者图像区域进行融合以获得融合图像，因此可以提高图像融合算法的鲁棒性。此外，在变换域图像融合方法中，深度学习方法可以提供由学习方法获得的变换，而不是像传统方法那样人为指定变换方法。

到目前为止，已有许多深度学习方法和模型被用于图像融合中。从学习方法的角度来讲，监督学习、无监督学习、自监督学习和迁移学习均已被用于图像融合。从模型的角度来讲，卷积神经网络(CNN)、生成对抗网络(GAN)、自编码器(autoencoder)、深度置信网络、Transformer和扩散模型等多种模型均已在图像融合中发挥了作用。虽然深度学习被引入图像融合领域才短短几年时间，但是基于深度学习的方法基本上已经成为了图像融合领域的主流方法了。

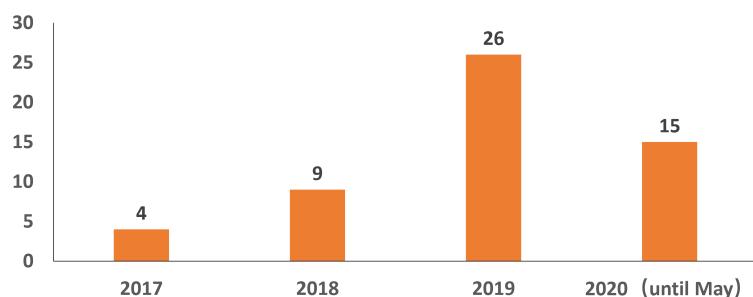


图 1.9: 2017 年至 2020 年 5 月份期间基于深度学习的多聚焦图像融合论文发表情况。图像来源于 [10]。

1.7 本书的写作目的

由于深度学习是2017年左右才被引入图像融合领域，所以以往的图像融合著作绝大多数没有涉及这方面的内容，或者只是非常简略地提到了一些。尽管已有不少相关研究成果发表，但是基于深度学习的图像融合方法相较于传统方法的

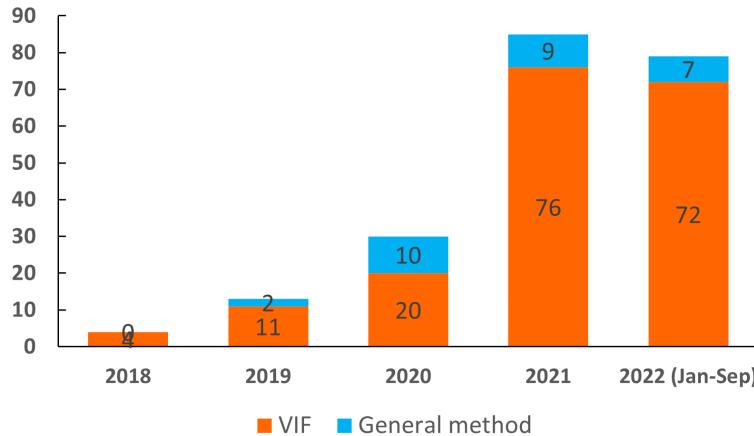


图 1.10: 2018 年至 2022 年 9 月份期间基于深度学习的可见光与红外图像融合论文发表情况。图像来源于 [11]。

优势和劣势还有不少未知的地方。例如，在图像融合领域是否像在目标检测、图像识别等领域一样，深度学习方法“完胜”传统方法，目前还是一个未知数。目前每年仍然有不少传统图像融合方法被发表，即是一个例证。此外，对于如何开展基于深度学习的图像融合，例如选用哪个深度学习模型、用哪个深度学习框架进行算法开发、如何获取大规模训练数据，目前仍缺乏理论和实践方面的著作来进行系统讨论。最后，在引入深度学习以后，图像融合有哪些新的应用、如何应用，目前也缺乏系统性的介绍材料。

基于上述考虑，笔者认为写作一部专门讨论基于深度学习的图像融合方法、应用和实践的著作是非常有必要的。这就是本书的写作目的。

1.8 本书主要内容与特色

1. 主要内容

本书主要包括三个部分。

第一部分是背景与概念，包括第一章和第二章。其中，第一章是绪论，主要介绍图像融合方面的一些基础知识，如图像融合的基本概念、图像融合的分类、本书的写作目的等。第二章介绍深度学习基础，包括深度学习发展情况简介、深度学习基础知识和常用的深度学习框架等内容。

第二部分是方法与技术，包括第三章到第十章。其中，第三章总体介绍基于深度学习的图像融合，包括其必要性和发展状况，以及常用于图像融合的深度学习模型。第四章介绍图像融合算法的性能评价方法，包括定性评价方法和定量评

价指标。第五章至第七章各介绍一个图像融合类别，分别是可见光与红外图像融合、多聚焦图像融合、多曝光图像融合。在介绍每一个图像融合类别时，均首先给出问题的定义，然后简单介绍传统方法，再重点介绍近年来迅速发展的基于深度学习的方法。第九章介绍通用图像融合方法，即可以同时应用于几种图像融合任务的融合方法。第十章介绍最近流行的应用驱动的图像融合方法。

第三部分是应用、实践与展望，主要包括第十二章到第十六章。其中，第十二章介绍可见光和红外图像融合的应用，第十三章介绍多聚焦图像融合的应用，第十四章介绍多曝光图像融合的应用。第十一章介绍图像融合实践，包括编程语言的选择和评价基准的使用。第十五章介绍图像融合领域的前沿进展。最后，第十六章对全书进行总结，并对图像融合领域的发展进行展望。

除了正文以外，笔者在附录A中还列出了图像融合相关的学术期刊和会议，便于读者朋友在进行论文投稿时参考。在附录B中，笔者给出了笔者的 GitHub 链接。读者朋友们可以从该链接中找到许多开源图像融合算法的下载链接，以便于开展研究工作。在附录C中，笔者结合自己多年的论文写作和审稿经验，简单介绍了一下论文的写作经验，以供相关读者参考。

2. 主要特色

本书主要具有以下特色：

- 前沿。本书主要关注近年来飞速发展的基于深度学习的图像融合方法。据笔者所知，本书是第一部关注基于深度学习的图像融合方法的著作。具体地，本书总结讨论了自 2017 年以来的基于深度学习的图像融合研究的主要进展，并介绍了笔者在相关领域的最新科研成果。本书绝大多数参考文献为 2017 年以后发表。因此，本书可以体现图像融合领域的最新发展现状和趋势。
- 理论与应用结合。本书不仅关注基于深度学习的图像融合理论，也十分关注基于深度学习的图像融合技术的实际应用。本书在第三部分系统总结了图像融合的应用。
- 重视实践。本书第十一章重点介绍了图像融合的实践知识，包含编程语言以及深度学习框架的选择，和图像融合评价基准的实验。此外，在附录中，笔者还介绍了图像融合相关的学术期刊和会议，给出了一些开源代码的下

载链接，并分享了图像融合论文的写作经验。与以往的图像融合著作相比，读者在阅读本书以后能尽快上手实践，更具有可操作性。

- 通俗易懂。虽然本书是关于图像融合的专业著作，但是笔者尽量使用通俗易懂的语言来进行叙述，力争使得即便没有图像融合研究经验的读者也能轻松看懂，并尽量不使用数学公式。此外，笔者从金庸小说中精心思考选取了一些例子作为对相关理论的补充说明，希望可以更好地帮助读者进行理解。

第 2 章

人工智能基础知识

我们认为人工智能是最值得学习的学科。

——《人工智能：一种现代方法》

深度学习技术在许多应用上已经取得了非常好的效果，并且已经渗透到了我们日常生活的方方面面，如人脸识别、推荐算法、智能监控等。在 2017 年左右，深度学习技术被引入到图像融合领域并取得了很好的效果。此后，基于深度学习技术的图像融合取得了迅速的发展。为了便于读者更好地理解本书内容，本章对深度学习的知识做简要介绍。

2.1 什么是深度学习？

我们经常在不同的地方看到深度学习、机器学习和人工智能（Artificial Intelligence, AI）这些名词。很多人无法区分这几个概念。实际上，机器学习是人工智能的子领域，而深度学习又是机器学习的子领域。他们三者之间的关系如图2.1所示。在人工智能经典名著《人工智能：一种现代方法》[12] 中，作者将机器学习列为人工智能的六大子领域之一。著名深度学习框架 Keras 的作者在他的《Python Deep Learning》一书中说“深度学习是机器学习发展最快且最重要的子领域”。

机器学习和深度学习都是从数据中学习规律。不过，在机器学习中，输入数据的特征需要人为设计，即需要进行特征工程。而在深度学习中，特征不再需要人为设计，而是从大量的数据中自动学习到的。相比于人为设计的特征而言，深度学习模型学习到的特征更好、更能适应复杂的情况。因此，在许多应用中，深

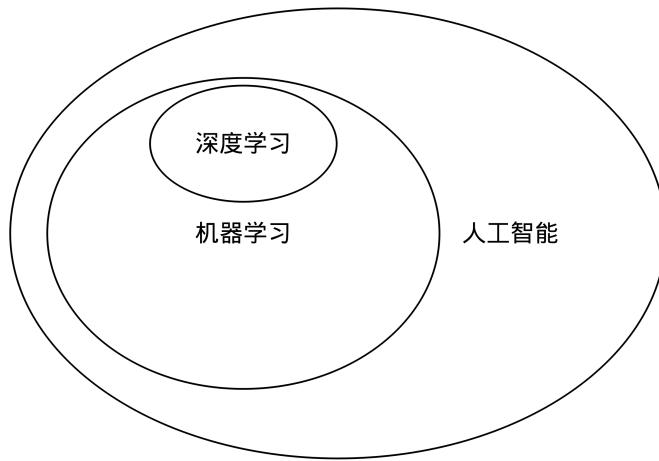


图 2.1: 人工智能、机器学习和深度学习的关系示意图。

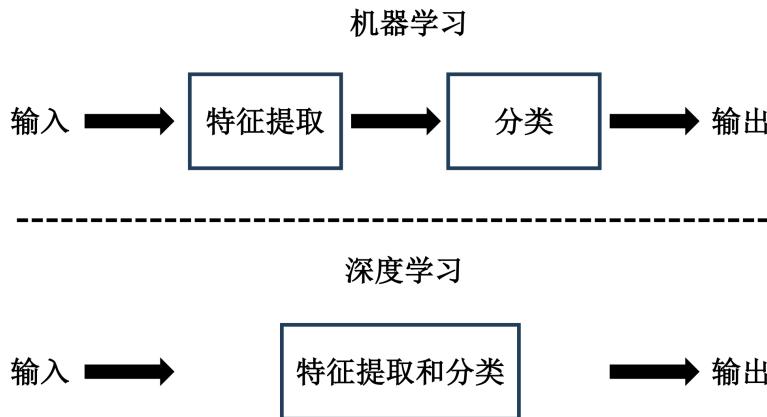


图 2.2: 机器学习和深度学习的区别。

度学习算法可以得到比机器学习算法更好的性能。机器学习和深度学习的对比如图2.2所示。此外，深度学习基本上特指深度神经网络。“深”是指这个神经网络可以有很多层，比如 100 多层乃至上千层。层数越多，这个深度神经网络的表达能力越强。

关于深度学习的发展，需要指出的一点是，尽管深度学习已经被研究了许多 年，但是一直到 2012 年，深度学习才开始腾飞。笔者于 2012 年初做本科毕业设计课题“基于视觉的目标跟踪与分析”时，还尚未听说过深度学习这个概念。那时也几乎没有人 在计算机视觉任务中使用深度学习技术。深度学习腾飞的标志性事件是，辛顿等人在 2012 年使用了基于 AlexNet 的方法，在 ImageNet 图像分类比赛中以绝对优势战胜了传统方法。

现在，深度学习技术已经成为了人工智能的代表技术。斯坦福大学吴恩达教授把人工智能比作为电，认为人工智能会像电一样无处不在，给人类生活带来巨大的改善。这也主要是得益于深度学习技术的迅速发展。

2.1.1 优化问题

深度学习问题，本质上是一个优化问题。我们一般首先根据我们的任务需求，建立一个优化问题并定义一个损失函数(loss function)¹，然后使用优化算法来优化这个损失函数，从而得到我们的模型。

2.1.2 基于梯度的优化

在深度学习中，人们使用基于梯度的优化算法。更具体地，常常使用的是梯度下降法。这种算法要求先求出损失函数的梯度，然后沿着梯度方向去优化损失函数。

起初，人们会用手工的方法去求梯度，但这样不仅费时费力，而且容易出错。后来，人们开始使用自动微分技术通过反向传播来求梯度。例如，笔者在读博士时，还曾使用过一种叫做 Tapenade [13] 的针对 Fortran 的自动微分软件。现在，自动微分技术已经被很好地集成在深度学习框架，如 Tensorflow 和 PyTorch 中。我们在开发深度学习模型的过程中，只需要调用代码去求梯度即可，不用再花费更多精力去研究其原理。例如，在 PyTorch 中，Autograd 可以为张量(tensor)求导数。在将张量封装为 Variable 之后，即可以调用它的.backward() 方法实现自动微分来求梯度。

2.2 深度学习三要素

如图2.3所示，深度学习的三要素为：数据，算力，算法。事实上，深度学习在近十年来的迅速发展，与这三要素息息相关。

2.2.1 数据

数据对于深度学习至关重要。在深度学习中，数据集一般分为三个部分，即训练集、验证集和测试集，如图2.4所示。训练集是用来训练模型的，一般包含数据和标签。验证集用来在训练的过程中检验模型的性能和选择模型。测试集则是在模型训练好以后，用来测试模型的性能的。需要指出的是，在训练阶段，往往我们只知道训练集和验证集，而不知道测试集。打个比方，训练集就像是老师给我们讲的例题，验证集就是模拟考试题，而测试集就是正式考试的题目。正如高

¹就是优化问题中的目标函数

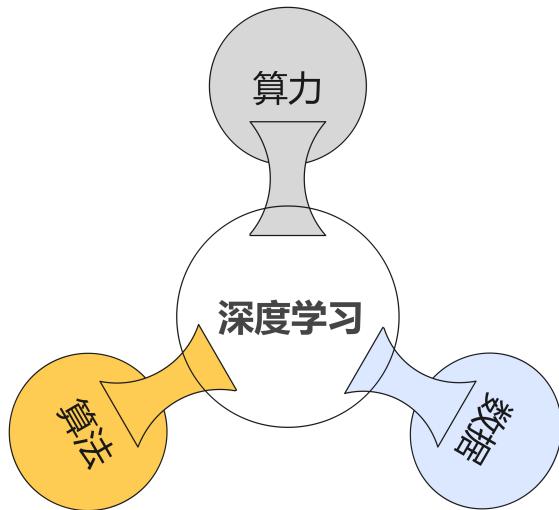


图 2.3: 深度学习三要素: 数据, 算力, 算法。

中的学生希望在高考中取得好成绩一样, 模型训练的目的是在测试集上取得好的效果。此外, 测试集在模型训练过程中是未知的, 正如高考题是未知的一样。

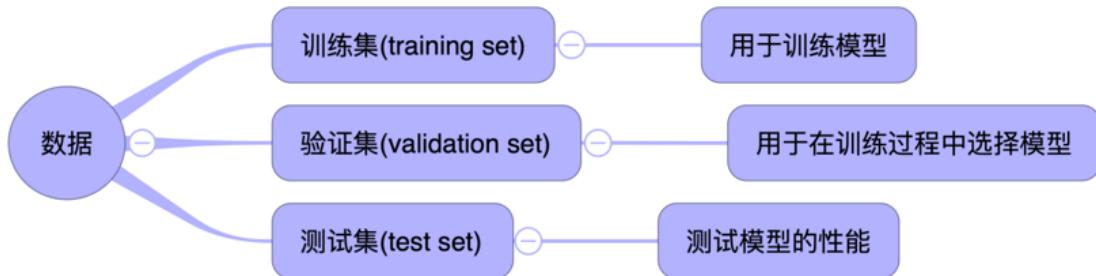


图 2.4: 深度学习数据集的划分。

2.2.2 算力

深度学习在近年来可以得到迅猛的发展, 与算力的发展是分不开的。具体来说, 以前即使有很好的模型和大量的数据, 但是因为算力不够, 人们也无法完成模型的训练。GPU (graphics processing unit)的应用, 使得人们可以更快地训练深度学习模型。正因为如此, 人工智能 GPU 的主要厂家英伟达在近年来得到了发展非常迅速, 其市值在 2023 年 5 月超过了万亿美元, 成为全球首家市值突破万亿美元的芯片制造商。

2.2.3 算法

深度学习算法的进步也是促进深度学习迅速发展的重要原因之一。例如，反向传播在深度学习里的使用，使得训练多层网络成为可能。此外，残差网络、注意力机制和 transformer 等算法及模型的提出，也使得深度学习得到了进一步的发展。

2.3 深度学习的分类

一般来说，深度学习可以分为监督学习、无监督学习和强化学习三种类型，如图2.5所示。

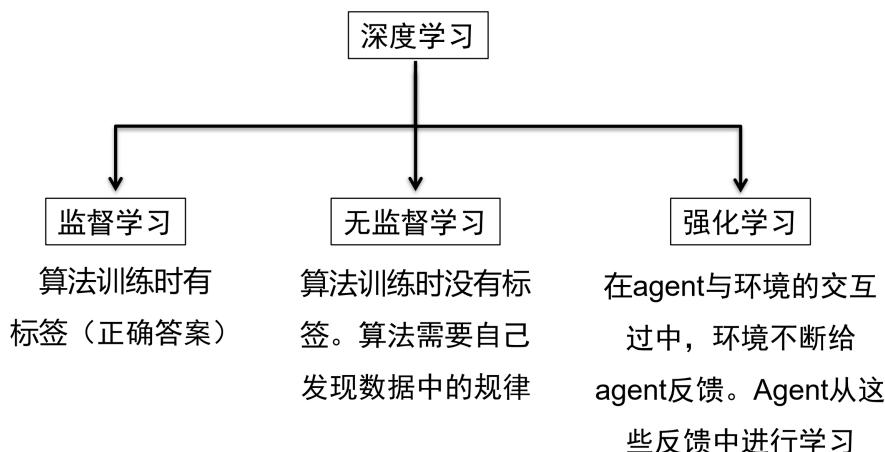


图 2.5: 深度学习的分类。

2.3.1 监督学习

监督学习是指，在进行算法训练时，我们需要使用数据的标签来指导模型的学习。为了获得标签，我们需要对数据进行标注，不过标注是一个很费时费力的过程。

监督学习可以分为判别式模型和生成式模型。判别式模型是指，在模型学到数据的规律以后，给新的数据，模型可以给出相应的结论，比如预测房价。而生成式模型指的话，在模型学到数据的规律以后，可以根据该规律生成新的数据。现在非常火爆的文图转换以及 ChatGPT，都是生成式模型。

2.3.2 无监督学习

无监督学习指的是在算法训练时不需要数据的标签，模型可以自行学习到数据的规律。典型的无监督学习方法如自编码器。

一般来说，无监督学习的效果没有监督学习好。但是，由于不需要标签，省略了耗时耗力的标注过程，无监督学习具有非常好的前景。例如，国际著名期刊 Nature Machine Intelligence 上曾发表过一篇文章，指出我们从婴儿的学习中受到启发，去开发新的无监督学习方法 [14]。

另外，值得注意的是，在无监督学习中有一种特殊的方法，称为自监督学习。自监督学习的主要思想是从数据本身来挖掘或者构造标签信息用于指导模型训练，因此也不需要人为进行标注。在本书中，笔者不对无监督学习和自监督学习进行区分。

2.3.3 强化学习

说到强化学习，让我们先来看看金庸先生的《侠客行》中的一个例子。在金庸先生的《侠客行》中，有这样一段话：

十余招后，石破天信心渐增，拆解快了许多。闵柔心中暗喜，**每当他一剑使得不错，便点头嘉许**。石破天早看出她在指点自己使剑，**倘若闵柔不点头，那便重使一招**，闵柔如认为他拆解不善，仍会第三次以同样招式进击。总要让他拆解无误方罢。

——金庸《侠客行》

这段话是对《侠客行》的主人公石破天与闵柔（可能是他母亲，但双方都不知道）过招的一段描述。在这段话里，有几个关键词，即“点头嘉许”、“不点头”。这两个词可以理解成正反馈与负反馈，其实就是强化学习里的“奖赏（reward）”²，是强化学习的关键。

强化学习强调智能体 (agent) 与环境的互动。在当前状态下，智能体做出动作以后，会从环境那里得到一个奖赏。基于这个奖赏信息，智能体决定下一步的行动。智能体的目标是获得最大化的预期利益。

对照强化学习和《侠客行》中的描述可以看出，石破天就是智能体，动作就是他使剑招，奖赏就是闵柔“点头嘉许”或者“不点头”。这样一看，金庸先生

²注意，奖赏可以是正的，也可以是负的

在《侠客行》中描述了一个很直观的强化学习过程。

事实上，强化学习的思想，与人的学习过程也非常类似。笔者记得在我儿子几个月大的时候，有一天我太太给我发来一段话，说的是和儿子玩的情况：

今天我们和他玩的时候，如果叫他名字他会转头回应，我们就笑并且亲亲他，他就知道回应我们会让我们开心。如果他不回头没回应我们，我们就不亲他。然后试了几次以后，每次叫他他都会回头回应我们了。

上面这段话，和金庸先生在《侠客行》里描述的石破天和“母亲”闵柔拆招的情况几乎一模一样，也是一个典型的强化学习过程。在这个过程中，我儿子是智能体。他妈妈亲他或者不亲他，就是对他动作的奖赏。

强化学习是一种非常有用的技术。英国的 DeepMind 公司使用强化学习技术开发出了很多著名的模型。例如，打败围棋世界冠军的 AlphaGo [15] 和用于控制核聚变的模型 [16]，都是基于强化学习算法开发的。

2.3.4 分类问题和回归问题

深度学习模型根据输出数据的类型，又可以分为分类问题和回归问题。在分类问题中，输出数据是离散的。例如，图2.6中展示的二分类和多分类的例子，就是典型的分类问题。



图 2.6: 深度学习中的分类问题。

在回归问题中，输出数据是连续的。例如，在房价预测问题中，因为房价可以是连续的数据，因此这是一个典型的回归问题。

2.3.5 判别式模型和生成式模型

给定一张图片，让一个模型来判断图片里是哪种动物，这种模型是判别式模型。如果给定一批图片，让模型学习这些图片的规律，然后用该模型来生成一张类似的图片。这样的模型，是生成式模型。

在 2022 年底以前，判别式模型是更主流的模型。2022 年 1 月，笔者在帝国理工学院电子与电气工程系担任《深度学习》课程讲师，负责主讲的内容之一就是生成式模型。然而，那个时候 OpenAI 还没有发布 ChatGPT，生成式模型还远远没有达到如今这种火爆程度。

在 ChatGPT 被发布以后，生成式模型变得异常火热，成为了科技领域最热门的话题之一。如今，生成式模型已经被用于许多的应用。

2024 年 7 月份，笔者有幸在伦敦科学博物馆听了著名华裔数学家陶哲轩的讲座。在讲座中，陶哲轩肯定了生成式 AI 的巨大价值。他特别提到，对于那些错误答案可能产生的后果并不严重的领域，例如用 AI 设计 PPT 背景图片和证明数学题，生成式模型会特别有用。

2.4 深度学习算法的常规设计流程

深度学习算法的常规设计流程如图2.7所示。一般来说，我们需要先分析我们的问题，尤其是我们需要处理的数据的类型（例如，语音，图像，文本，视频，或者图），以及我们需要达到的目标（分类问题，回归问题，或者生成新数据）。然后，我们可以选择数据集，设计模型和定义损失函数，并设置参数。此后，我们可以进行模型训练和测试。如果发现模型的效果不好，我们可以回过头再进行前面的步骤，对模型、损失函数和参数进行调整，然后重新训练。

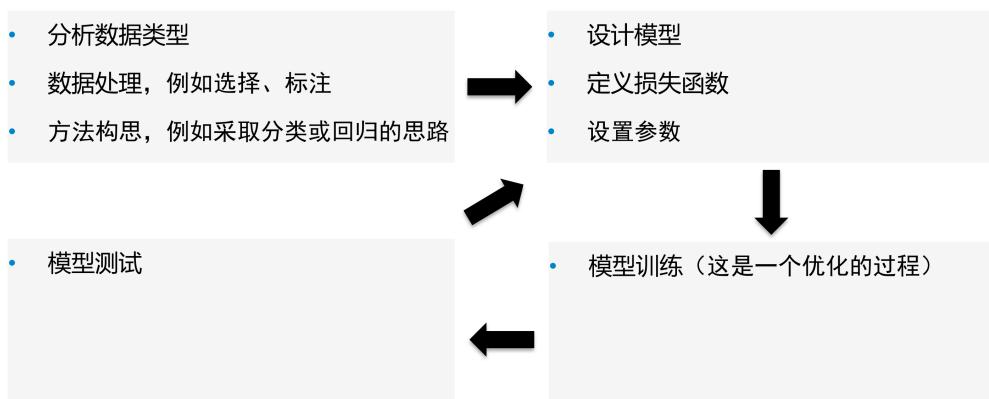


图 2.7: 深度学习算法的常规设计流程。

2.5 图灵测试

在讨论人工智能的时候，不可绕开的一个话题就是怎么判定什么是智能。1950年，英国数学家图灵在他的经典论文中提出了图灵测试。图灵测试的基本内容是说，如果人和一个模型对话时，无法区分对方是人还是计算机，那么可以认为这个模型具有智能。

截止到2022年11月以前，人们认为图灵测试在短期内是难以通过的。然而，在2022年11月份美国OpenAI公司发布了ChatGPT以后，有相当多专家认为ChatGPT已经可以通过图灵测试了。例如，笔者于2023年4月份在帝国理工学院听过一次英国皇家工程院院士、香港科技大学首席副校长郭毅可院士的讲座。在那次讲座中，郭院士就提到他认为ChatGPT已经通过了图灵测试。

2023年7月，国际顶级期刊Science上也发文^[17]讨论了如何判断AI系统是否智能的问题。文章中提出了几点质疑：

- 数据污染。即GPT-4表现出来的智能，可能是他们的训练数据中包含了后来用作测试的数据，或者非常类似的数据。
- 鲁棒性。作为人类，我们如果会回答一个问题，那么我们很可能回答一个类似的问题。但是对于AI系统来说就不一定了。比如，众所周知，GPT-4对于问题的描述非常敏感。对于同一个概念的考察，有可能换一种描述方式以后，GPT就不知道该如何处理了。
- 评价基准（Benchmark）的不可靠。

因此，该文章对于现在³的AI系统是否具备了人类的智能，持怀疑态度。文章还认为，设计合理的评估AI系统智能性的方法是急需要做的事情。2023年8月，Science继续刊文指出图灵测试已不太适合，建议使用Weizenbaum测试^[18]。

不管图灵测试是否可以继续判定智能的存在，图灵测试在人工智能的发展历史上的影响力是巨大的。值得说明的是，笔者十分崇拜图灵。2022年，笔者去英国曼彻斯特游玩时，还专程去看了曼彻斯特市区的图灵雕像（见图2.8）、瞻仰先贤。

³截止到2023年7月



图 2.8: 笔者与英国曼彻斯特的图灵雕像。

2.6 常用深度学习框架简介

深度学习网络是高度模块化的，即由常见的一些深度学习模块组成。在实现深度学习算法的过程中，如果我自己来编写这些模块的程序，不仅费时费力，而且容易出错⁴。因此，为了便于开展深度学习的研究，研究员们开发了一些深度学习框架。通过使用这些深度学习框架，我们可以非常方便的搭建我们的深度学习模型。深度学习框架的出现，对于深度学习的快速发展也起到了很重要的促进作用。本节对一些深度学习框架进行简要介绍。

1. Theano

Theano 是第一个 Python 深度学习框架，由蒙特利尔大学的 MILA 实验室开发。它为许多后来的框架奠定了基础。Theano 具有以下一些特点：

符号计算：Theano 采用了符号计算的方法来定义计算图。用户首先构建符号图，然后再将具体数值输入进行计算，这种方式在一些情况下可以提高计算效率。

高度优化：Theano 在计算优化方面非常出色，它可以自动进行计算图优化，包括常数折叠、梯度计算的优化等，从而加速计算过程。

GPU 支持：类似于后来的深度学习框架，Theano 也支持在 GPU 上进行计算，以加速深度学习模型的训练和推理。

⁴例如，在深度学习里面求导数是非常重要的。笔者读博士期间还曾研究和使用过自动微分的软件。然而，那时候的自动微分软件很不稳定，而且不能处理很复杂的程序。

面向研究：Theano 的设计初衷是为了支持深度学习研究和实验，因此它提供了灵活的符号计算和模型定义方式，适合用于探索新的深度学习算法和模型结构。

在图像融合领域，有一个基于多曝光图像融合的方法 DeepFuse [19] 是使用 Theano 进行开发的。然而，虽然 Theano 在过去曾经是深度学习领域的重要框架之一，但由于其开发和支持的逐渐减少，以及一些现代框架（如 TensorFlow 和 PyTorch）的崛起，许多用户和开发者已经转向更现代化、功能更全面的框架。不过，Theano 对于深度学习的发展做出了重要贡献，并且其设计思想和技术在后来的框架中得到了继承和发展。

2. MatConvNet

当涉及到深度学习框架时，MatConvNet 也是一个选择。MatConvNet 是由牛津大学一个实验室开发的深度学习框架，专门用于处理计算机视觉任务。它是基于 MATLAB 的，因此特别适合那些熟悉 MATLAB 编程环境的用户。

以下是 MatConvNet 的一些特点：

专注于计算机视觉：MatConvNet 在设计上专注于处理计算机视觉任务，如图像分类、目标检测、语义分割等。它提供了丰富的预训练模型和网络结构，使得在这些任务上的实验和应用变得更加便捷。

MATLAB 集成：由于 MatConvNet 是基于 MATLAB 的，它与 MATLAB 的各种图像处理和计算工具集成良好，使得用户可以方便地在 MATLAB 环境下进行数据处理、可视化和实验调试。

性能优化：MatConvNet 在优化计算性能方面做了一些努力，特别是在 GPU 上的加速。它通过利用 GPU 计算能力，使得深度学习模型的训练和推理更加高效。

易用性：虽然 MatConvNet 相对于一些主流框架如 TensorFlow 和 PyTorch 在用户基数上较小，但它的设计目标之一是易用性，使得那些在 MATLAB 中进行计算机视觉研究的用户可以更快速地上手。

有一些知名的目标跟踪算法（如 SiamFC[20] 和 CFNet[21]）和图像融合算法 [22, 23] 是基于 MatConvNet 开发的。不过，由于主要针对的是计算机视觉，以及受 Matlab 的运算速度的限制，MatConvNet 的用户群体并不是很大。

3. Caffe

Caffe 是一个受欢迎的深度学习框架，由 Berkeley Vision and Learning Center 开发。Caffe 主要用于计算机视觉任务，特别擅长处理卷积神经网络。它以配置文件的形式定义网络结构，使得模型的构建和修改非常简单。Caffe 还提供了预训练的模型库，方便用户在特定任务上进行迁移学习。然而，相比于其他框架，Caffe 在灵活性和扩展性方面稍显不足，但在计算速度上表现优秀。

4. TensorFlow

TensorFlow 是由谷歌开发的开源深度学习框架，广泛应用于学术界和工业界。它以数据流图的形式表示计算任务，允许用户在 CPU 和 GPU 上高效地运行深度学习模型。TensorFlow 提供了丰富的 API 和工具，支持构建各种类型的神经网络，包括卷积神经网络、循环神经网络等。它还具备自动求导和分布式训练的能力，能够应对大规模的数据和模型。

在图像融合领域，有大量的方法是基于 TensorFlow 开发的，如 DenseFuse[24] 和 FusionGAN[25]。然而，由于 TensorFlow 版本之间的兼容性不太好，现在研究人员开始倾向于使用 PyTorch 了 [11]。

5. Keras

Keras 是一个高级神经网络 API，最初是建立在 Theano 之上，后来也支持 TensorFlow 作为后端。Keras 的设计目标是简单易用，使得用户能够快速搭建深度学习模型而不需要过多关注底层细节。它提供了丰富的预定义层和模型，也支持自定义层和模型，使得网络的组合和迁移变得十分便捷。虽然 Keras 本身的发展已经融入到 TensorFlow 中，但它作为一个高级 API 仍然被广泛使用。

6. PyTorch

PyTorch 是由 Facebook⁵开发的深度学习框架，也是一个开源项目。相比于 TensorFlow，PyTorch 更注重动态图计算，使得模型的构建和调试更加直观和灵活。此外，PyTorch 的不同版本之间的兼容性相对较好。这些特点使得 PyTorch 在学术界广受欢迎。许多研究人员喜欢使用 PyTorch 来探索新的深度学习思路。同时，PyTorch 也提供了丰富的工具和 API，支持分布式训练和混合精度计算，

⁵现已改名叫 Meta

能够在不同硬件平台上高效运行。

7. MXNet

MXNet 是一个开源深度学习框架，最初由亚马逊 (Amazon) 公司开发，现在由 Apache 孵化器托管。MXNet 提供了动态图和静态图两种方式来定义计算图，用户可以根据任务的需求选择适合的方式。MXNet 支持多种编程语言，如 Python、C++、R 等，具有良好的跨平台性。此外，MXNet 还在分布式训练和模型部署方面具有一定优势，适用于大规模和高效的深度学习计算。

综上所述，现有的深度学习框架主要分为四个阵营：Google, Facebook (Meta), Amazon, Matlab。在学术界，Tensorflow 和 Pytorch 占了绝对主流。值得注意的是，这里只是介绍了一些常用的深度学习框架，实际上市场上还有许多其他优秀的框架，每个框架都有其独特的优势和适用场景。在选择框架时，需要根据具体的任务和需求，综合考虑各个框架的特点，选择最适合自己的工具。随着深度学习领域的不断发展，这些框架也在不断更新和完善，为用户提供更好的深度学习体验。

2.7 小结

本章对深度学习的基础知识进行了简要介绍。我们介绍了深度学习本质上是一个优化问题，并且通常使用的是基于梯度的优化算法。我们也介绍了深度学习的三要素、分类以及常见的开发流程。此外，本章还介绍了图灵测试以及常用的深度学习框架。

第二部分

图像融合方法与技术

第 3 章

基于人工智能的图像融合概述

深度学习的引入，使得图像融合的研究掀起了新一轮高潮。

——笔者

近年来，深度学习技术被应用于许多领域并取得了非常好的效果。自 2016 年英国 DeepMind 公司的阿尔法狗战胜了围棋世界冠军李世石以后，深度学习得到了人们的高度的关注。此后，美国 OpenAI 公司于 2022 年 11 月推出 ChatGPT，使得深度学习进一步得到了全民的关注。

大约在 2017 年，深度学习被研究人员引入了图像融合领域。之后，基于深度学习的图像融合得到了蓬勃发展。每年均有大量基于深度学习的图像融合相关论文发表，并且数量呈现出迅速增长的趋势。深度学习的引入，使得图像融合的实现有了更多的可能性，各种新的想法和创意层出不穷，使得图像融合的研究掀起了新一轮高潮。

本章将对基于深度学习的图像融合进行概述。

3.1 传统图像融合方法简介

在深度学习技术被引入图像融合领域之前，图像融合已经被研究了 30 多年。在这 30 多年的研究中，国内外的研究人员提出了大量的图像融合方法。在本书中，这些方法统称为传统方法。

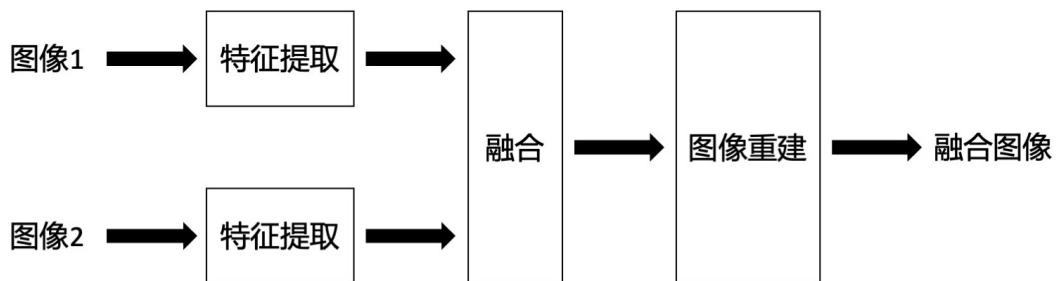


图 3.1: 图像融合的三个步骤。

3.1.1 图像融合方法的三个步骤

图像融合一般包含三个步骤，即特征提取、特征融合（规则）、图像重建，如图3.1所示。特征提取是指从源图像中提取出有用信息。这些有用信息会根据融合规则进行融合。融合以后的信息会再根据图像重建算法进行重建，从而得到融合图像。

3.1.2 传统图像融合方法的类别

如本书第一章所述，传统图像融合方法按照融合域进行分类，可以大致分为空间域和变换域的方法。前者直接在空间域对图像进行融合，并且一般可以进一步分为基于像素的、基于块的和基于区域的融合方法。基于像素的图像融合方法是指直接对源图像的像素进行融合。例如，在多聚焦图像融合中，先判断源图像中每个像素的聚焦度（清晰度），然后根据聚焦度生成权值，最后对源图像的每个像素进行加权融合从而得到融合图像。基于块的融合方法是指首先将源图像划分为大小固定的图像块，然后以图像块为单位来进行融合。基于区域的图像融合方法是指首先使用图像分割算法将图像划分为大小不等或者不均匀的图像区域，然后以这些区域为单位进行融合。其中，基于块和基于区域的融合算法的效果非常依赖于图像块和图像区域的划分方式，并且容易在融合图像中的边界处（例如聚焦和不聚焦的分界处）产生瑕疵。

基于变换域的图像融合方法通常首先使用某种变换将源图像变换到另外一个域。在变换域中，源图像一般以系数的形式存在。然后，在变换域中，根据一定的算法或者规则，对源图像的系数进行融合。最后，对融合后的系数进行逆变换以得到融合图像。常见的变换有多尺度变换（包括小波变换、金字塔变换等）和稀疏表达等。

3.1.3 传统图像融合方法的缺点

在传统图像融合方法中，这三个过程均由人为设计的方法完成。然而，人为设计的方法在完成这三个过程中不一定效果很好，尤其是不一定适用于大量源图像。例如，在基于块的空间域图像融合方法中，图像的分块方法对最终的融合效果影响比较大。但是人为设计的分块方法，不一定是最好的，也不一定对各个融合图像都有比较好的效果。再例如，在基于变换域的图像融合中，如何选择有效的变换域和融合规则，使得融合算法具有好的效果和鲁棒性，也是非常具有挑战性的。因此，在传统图像融合方法中，处处可见的人为因素对算法的有效性和鲁棒性会产生不可知的影响。

3.2 基于深度学习的图像融合发展状况概述

3.2.1 基于深度学习的目的

在3.1.3小节中，笔者介绍了传统图像融合方法的缺点。深度学习技术，可以帮助解决那些缺点。具体地，深度学习的引入，使得这三个过程可以更好地、更加自动化地完成。尤其是，深度学习模型具有很强的特征提取能力，可以获得比人为设计的特征提取方法更好的特征。此外，深度学习模型在实现融合规则和图像重建方面也具有更好的优势。最后，深度学习模型可以实现端到端的图像融合，即将三个步骤一起完成。

基于上述原因，基于深度学习的图像融合方法引起了研究人员的广泛兴趣，并在近几年得到了蓬勃发展。从2017年以来，每年均有大量的基于深度学习的图像融合方法发表，并且呈逐年上升的趋势。本书第一章的图1.9和1.10分别显示了基于深度学习的多聚焦图像融合和可见光与红外图像融合的相关文章的发表情况。从这些图中可以看出，基于深度学习的图像融合在近几年确实得到了迅速发展和大量关注。

3.2.2 有监督方法和无监督方法

根据训练过程是否需要标准答案(ground truth)，基于深度学习的图像融合方法可以分为有监督方法和无监督方法。在有监督方法中，标准答案被用于指导模型训练，而无监督方法不需要标准答案。

值得说明的是，图像融合任务并不存在标准答案。因此，绝大多数基于深度

学习的图像融合方法是无监督方法。在少数的有监督方法中，研究人员使用的标准答案也并不是真正的标准答案。例如，在多聚焦图像融合中，有研究人员基于清晰图像来生成多聚焦图像，然后在训练过程中将清晰图像用作标准答案。在多曝光图像融合中，有研究人员使用由其他融合方法生成的融合图像作为标准答案来对模型进行有监督训练。然而，这些构造标准答案的方法，都存在一些弊端。

3.3 常用于图像融合的深度学习模型

在近几年的发展中，已经有多种深度学习模型被用于图像融合中。笔者在这里对常用于图像融合的深度学习模型进行简要介绍。

3.3.1 卷积神经网络

基于卷积神经网络的方法，是基于深度学习的图像融合方法中的一个主要类别。在生成对抗网络被用于图像融合任务以前，几乎所有基于深度学习的图像融合方法都是采用的卷积神经网络。

在基于卷积神经网络的方法中，很重要的一个进步是全卷积网络的使用，因为全卷积网络可以接收任何尺寸的源图像作为输入，而不是像非全卷积网络那样只能接收特定尺寸的输入。

3.3.2 Transformer

自从谷歌的研究人员于 2017 年提出变换器 (Transformer) 以来，它就给深度学习领域带来了巨大的变化。许多著名的模型或者产品，例如 DeepMind 公司的 AlphaFold 和 OpenAI 的 ChatGPT，均是基于 Transformer 开发的。2020 年，谷歌的研究人员又提出了视觉 Transformer。后者成为了自 2012 年以后计算机视觉领域最大的创新。2021 年 ICCV 的最佳论文奖即授予了一个叫 Swin Transformer[26] 的模型。

相比于传统的卷积神经网络，Transformer 最大的优点是可以学习到全局的信息，而不再局限于卷积神经网络的局部感受野。这个特性对于图像融合任务也非常有益处，因此，研究人员也将 Transformer 引入到了图像融合领域。

3.3.3 生成对抗网络

著名物理学家费曼曾说，“我不理解我不能创造的东西”。这句话被广泛用于介绍生成式网络的文章中。与判别式网络不同，生成式网络学习数据的规律并按照该规律生成新的数据。作为生成式模型主要代表之一的生成对抗网络（GAN），由武汉大学马佳义教授团队于 2019 年率先引入到了图像融合领域。自此以后，基于生成式对抗网络的图像融合方法，成为了基于深度学习的图像融合研究里的主要方法之一。近几年有许多基于生成式对抗网络的图像融合方法被提出。

3.3.4 扩散模型

扩散模型（Diffusion model）是近两年非常热门的一种新型生成式模型。其在文图转换中的惊人效果使得扩散模型得到了广泛关注并被应用于许多任务之中。2023 年，武汉大学马佳义教授团队率先将扩散模型用于图像融合任务 [27]。之后，西安交通大学赵子祥博士提出了基于扩散模型的多模态图像融合模型 [28]。目前，基于扩散模型的图像融合方法，已成为图像融合领域一个新的热门方向，并在受到越来越多研究人员的关注。

各种深度学习模型被应用于图像融合任务的时间线如图 3.2 所示。由于篇幅原因，也由于在图像融合领域使用得很少，本书不对深度玻尔兹曼机和深度置信网络进行介绍。



图 3.2: 深度学习模型被应用于图像融合任务的时间线。

3.4 常用于图像融合的重要深度学习技术

3.4.1 注意力机制

李笑来在《财富自由之路》中说到，注意力是我们最重要的财富。吴军博士在《元智慧》一书中也说，“要专注于自己的事情，不要操不该操的心”。这主要是因为注意力机制可以使我们更加关注对我们重要的东西。深度学习里的注意力机制也是如此。因为注意力机制的重要性，它成为了在图像融合领域使用得最多的深度学习技术之一。

3.4.2 残差连接

残差连接 [29] 是著名计算机学者何恺明提出的一种技术。该技术使得深度学习网络可以被做得很深，是深度学习领域最为重要的基础技术之一。如图3.3所示，其基本思想是使虚线中的部分拟合出残差映射 $f(\mathbf{x}) - \mathbf{x}$ ，而这个残差映射在实际中更为容易优化。在基于深度学习的图像融合方法中，有许多方法在其深度学习模型中使用了残差连接。残差连接已成为基于深度学习的图像融合方法中的重要可选组件之一。

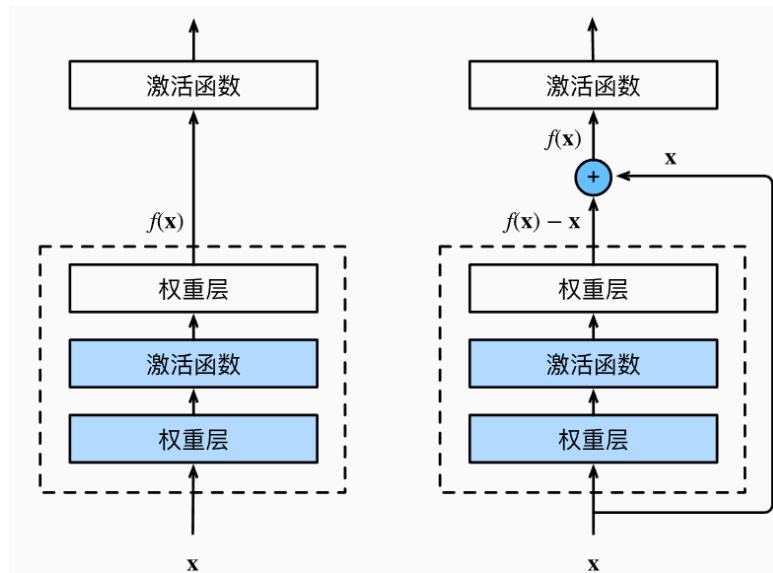


图 3.3: 正常网络 (左图) 和残差网络 (右图)。图片来源于 [30]。

3.4.3 稠密连接

稠密连接 [31] 是残差连接的逻辑扩展 [30]。具体地，稠密连接使用连接操作来取代残差连接中的相加操作，即在通道维上对输入和输出进行连接，如图3.4所

示。一个具体的稠密连接如图3.5所示。

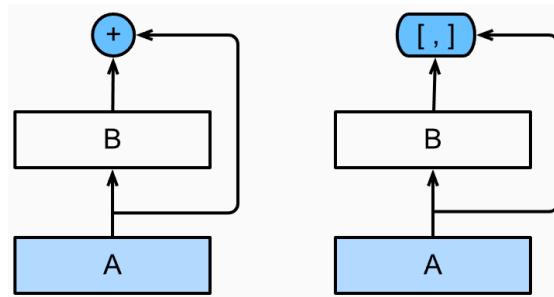


图 3.4: 残差连接 (左图) 和稠密连接 (右图)。图片来源于 [30]。

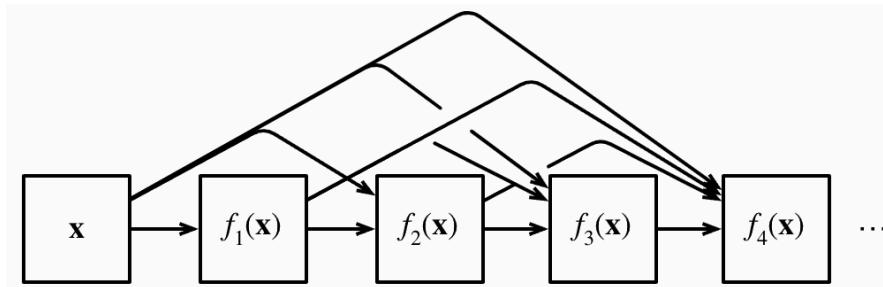


图 3.5: 稠密连接示意图。图片来源于 [30]。

3.4.4 自动网络架构搜索

在绝大多数基于深度学习的图像融合方法中，研究人员均需认真设计模型架构以便得到好的融合效果。然而，对每一个方法进行人为的架构设计，是一个很繁琐的过程，不仅费时而且费力。基于此种原因，有学者着手研发了基于自动网络架构搜索的图像融合方法。例如，大连理工大学樊鑫教授团队提出了基于自动架构搜索的可见光与红外图像融合方法（SMoA）。

3.4.5 其他重要技术

除了上述技术以外，其他一些技术，例如多尺度特征融合、对比学习、特征分解等，也常常被用于基于深度学习的图像融合方法中。由于篇幅原因，这里不再进行介绍。

3.5 与多模态机器学习的关系

从某种角度来说，基于深度学习的图像融合与多模态机器学习具有一定的相似性。例如，多模态图像融合（如可见光与红外图像融合、医学图像融合）和多

模态机器学习一样，都是基于多个模态的信息来进行学习，从而得到更好的效果。然而，基于深度学习的图像融合与多模态机器学习也有很大的区别。

首先，不是所有图像融合任务都是多模态的。例如多聚焦图像融合和多曝光图像融合，实际上是单模态的。

其次，基于深度学习的多模态图像融合，旨在通过融合源图像来生成高质量的融合图像。尽管现在有一些基于应用的图像融合方法出现（见本书第十章），这些方法仍把生成高质量的融合图像作为重要目标。然而，在多模态机器学习中，人们主要关注如何充分利用多模态信息以帮助其他下游任务取得更好的效果。例如，人们会关注如何利用可见光图像和激光雷达的信息，使得自动驾驶汽车可以在不同光照条件下都能感知外界环境，但是并不太关注把可见光图像和激光雷达的信息融合到一张图像中。换句话说，在多模态机器学习中，人们更关注如何利用多模态信息来提升下游应用的效果而并不追求把信息融合成单一的形式（如图像）。

此外，在基于深度学习的多模态图像融合中，一般只涉及两个模态，很少有两个以上的模态一起进行融合来生成融合图像。而在多模态机器学习中，可能会有多个模态一起进行融合。例如，Meta 公司开发的 ImageBind 模型 [32] 可以融合多达六个模态的信息。

因此，我们可以说图像融合与多模态机器学习既相关又有很大区别。在研究过程中，我们可以参考多模态机器学习的方法来进行多模态图像融合的研究，但同时也需要注意区分二者的不同之处。

3.6 基于深度学习的图像融合发展趋势

根据笔者的研究，基于深度学习的图像融合还大致呈现出以下发展趋势。

3.6.1 多种深度学习模型被用于图像融合

起初，只有少数基于深度神经网络和卷积神经网络的图像融合方法发表。近年来，越来越多的深度学习模型被用于图像融合，例如自编码器、变分自编码器、生成对抗网络、Transformer 和扩散模型。这些模型的引入，使得图像融合领域的研究呈现出百花齐放的效果，各种创新性的想法被提出，图像融合的效果也越来越好。各种深度学习模型被应用于图像融合任务的时间线如图3.2所示。

3.6.2 从非端到端的方法到端到端的方法

在深度学习刚刚被引入图像融合领域的时候，许多方法仍在是非端到端的方法。例如，一些方法先通过深度学习方法生成融合权值图（weight map），然后对该融合权值图进行一些后处理，最后再根据该权值图对源图像进行融合。另外，也有一些方法仅将深度学习应用于图像融合三个步骤中的一部分而不是全部步骤。

在最近两年，越来越多的端到端的方法被研究人员提出，即通过一个深度学习模型直接从源图像得到融合图像。

3.6.3 从特定图像融合方法到通用图像融合方法

起初，深度学习模型仅被用于开发用于特定图像融合任务的图像融合方法，例如可见光与红外图像融合、多聚焦图像融合。后来，研究人员通过研究，发现各种图像融合任务中存在一些共性，因此有可能用同一个深度学习方法来完成不同的图像融合任务。因此，一些研究人员提出了通用图像融合方法。关于通用图像融合方法的详细内容，请参见本书第九章。

3.6.4 从生成融合图像到改进下游应用

截止到本书写作之时，绝大多数基于深度学习的图像融合方法的目的是从源图像生成高质量的融合图像，如图10.2(a)所示。然而，图像融合的目的不仅仅是生成高质量的图像。图像融合更为重要的目的，是通过生成高质量的融合图像来给下游应用，例如目标跟踪、目标检测、场景分割，提供更好的数据源，从而提升下游应用的性能，如图10.2(b)所示。因此，生成高质量的融合图像是手段，而不是目的¹。这就像是金庸先生在《神雕侠侣》中提到的杨过和小龙女的双剑合璧。双剑合璧的目的应该是实用（打败对手），而不是为了好看。本书第十章将详细讨论应用驱动的图像融合方法。

¹当然，在一些图像融合任务中，生成高质量的融合图像也是目的，例如生成给安保人员查看的监控录像，在多聚焦和多曝光图像融合任务中生成高质量的摄影作品

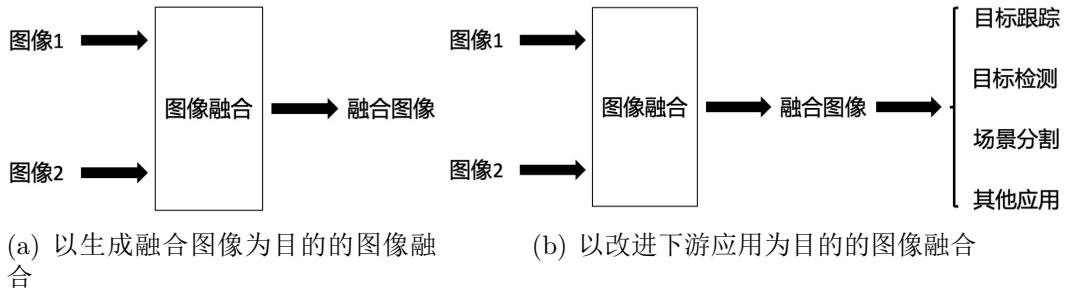


图 3.6: 以生成融合图像为目的的图像融合和以改进下游应用为目的的图像融合。前者的目的是生成高质量融合图像，而后者的目的通过生成融合图像来改进下游应用的性能。

3.6.5 新型图像融合类型开始出现

在深度学习被引入到图像融合领域之前，主要的图像融合任务包括可见光与红外图像融合、多聚焦图像融合、多曝光图像融合、医学图像融合和遥感图像融合。在近几年，一些新型的图像融合任务开始出现，例如可见光与近红外图像融合、可见光与深度图像融合、可见光与事件相机图像融合。此外，几乎所有的这些新型图像融合任务均是基于深度学习方法来实现的。由此也可以看出，深度学习对于图像融合领域的重要影响。本书第十五章将详细讨论近年来出现的新型图像融合类型。

3.7 小结

本章对基于深度学习的图像融合方法进行了概述，主要介绍了使用深度学习进行图像融合的必要性和基于深度学习的图像融合的基本发展状况。希望本章可以使读者朋友们对基于深度学习的图像融合有一个大概的了解，以便我们在后续章节中更加深入地介绍基于深度学习的图像融合方法。需要注意的是，本书第二章和本章仅对深度学习知识做了非常简单的介绍。如读者朋友希望了解更多关于深度学习的知识，笔者强烈推荐读者朋友们阅读相关书籍 [30, 33]。

第 4 章

图像融合算法性能评价

正如华山论剑可以确定武林高手的武功高低，图像融合评价基准可以用于确定图像融合算法的好坏

——笔者

图像融合算法的性能评价是图像融合研究中的一个重要部分。要确定一个图像融合算法的优劣，必须采用一定方法来对该算法进行评价。因此，研究人员们每开发出一种新的图像融合算法并想以论文形式发表时，都需要在论文中将该算法与现有算法进行性能对比，以证明该算法的优越性。本章介绍图像融合算法的性能评价方法，包括其特殊性、常用的图像融合评价方法、图像融合算法评价的特殊性、当前评价方法存在的问题，最新发展趋势和未来展望。

4.1 图像融合算法评价的特殊性

图像融合算法性能评价，是指通过一些评价方法来确定图像融合方法的好坏。性能评价方法，不管是对于我们了解一个算法的好坏，还是对于研究人员发表学术论文，都是必不可少的。然而，图像融合算法的性能评价具有一定的特殊性。

我们知道，在计算机视觉的很多研究领域如目标跟踪和目标检测中，存在标准答案（即 ground truth）。例如，在目标跟踪领域，OTB、VOT 等性能测试平台会提供标准答案（通常是以 bounding box 的形式存在），如图4.1所示。一个算法是否成功跟踪到目标，以及跟踪到的程度如何，是非常明确的¹。在目标检

¹当然，这些任务中也会有不同的评价指标，但都是与标准答案进行对比

测任务中，也会有目标类别、所在位置和大小等标准答案可供对比。

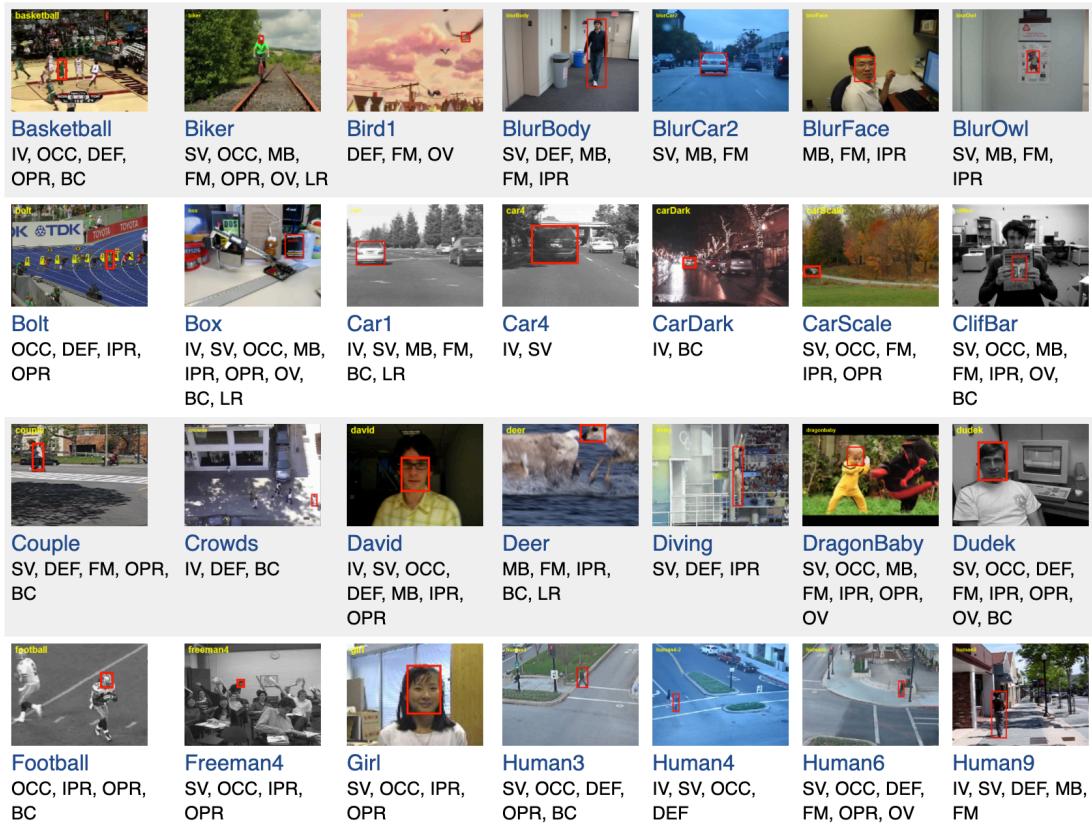


图 4.1: 目标跟踪评价基准 OTB 中的部分图像及标准答案。图中红色的矩形框即为该帧图像中相应目标的所在位置, 也就是标准答案。图片来源于 OTB 官网。

然而, 与目标跟踪、目标检测等领域不同, 在图像融合任务中, 一般不存在这种标准答案。因此, 图像融合算法的好坏程度不是那么直观。这就好比在《射雕英雄传》中, 想弄清楚郭靖和欧阳锋的武功高低很容易², 但要想知道郭靖和欧阳锋谁写诗写得好就不容易了, 因为前者有客观标准而后者没有。因此, 与很多其他的计算机视觉任务相比, 图像融合算法的性能评价具有其特殊性, 即没有标准答案。

4.2 当前的主要图像融合评价方法

目前, 图像融合算法的性能评价主要包含定性评价方法和定量评价方法。

²华山论剑

4.2.1 图像融合定性评价方法

定性评价方法，也叫主观评价方法，是指通过人眼来观察融合图像，以判断融合图像质量好坏的方法。这种方法非常重要，因为高质量的融合图像需要具备好的视觉效果。此外，一般来说，定性评价方法也比较直观。因此，在几乎所有的图像融合论文中，研究人员都会展示一些融合图像以便进行定性比较，如图4.2所示。



图 4.2: 笔者在论文中展示的可见光红外图像融合算法的定性评价。从图中可以看出，定性评价主要通过比较不同融合图像的视觉效果来对图像融合方法的性能进行对比。图像来源于 [11]。

然而，定性评价方法具有几个重要缺点。

首先，定性评价方法具有很强的主观性。所谓“萝卜白菜，各有所爱”，不同的人对于图像视觉效果的好坏可能存在不同的标准。因此，在进行定性评价的过程中很可能出现不同研究人员对于同一个算法的性能评价结果不同的情况，尤

其是在有些算法性能比较接近的情况下。

其次，定性评价需要观察者一张图一张图来观察，无法自动化操作。所以，当被评价的算法数量多、测试图像数量多的时候，定性评价方法非常耗时耗力。

最后，因为学术论文篇幅有限，因此研究人员不可能在论文中展示所有的定性结果。因此，通常的做法是在论文中放几个定性比较的例子来说明算法的性能。然而，这种做法对于图像融合算法性能的展示并不全面，会出现采样偏差[34]。比如，有可能出现某算法只在论文中展示的几个例子中表现好而在其他图像上表现不好的极端情况。

基于上述原因，仅仅依靠定性的方法对图像融合算法的性能进行评价是不够的。

4.2.2 图像融合定量评价方法

定量评价方法，又叫客观评价方法，是指通过一些客观的图像融合评价指标来对图像融合的结果进行评价的方法，如图4.3所示。

TABLE III
QUANTITATIVE PERFORMANCE COMPARISON ON VIFB. THE BEST THREE VALUES OF EACH METRIC ARE MARKED IN RED, GREEN AND BLUE, RESPECTIVELY. THE THREE NUMBERS AFTER METHOD NAMES DENOTE THE NUMBER OF BEST VALUE, SECOND BEST VALUE, AND THIRD BEST VALUE, RESPECTIVELY. BEST VIEWED IN COLOR

Method	Information theory-based				Information feature-based				Structural similarity-based		Human perception-inspired		
	CE (↓)	EN (↑)	MI (↑)	PSNR (↑)	AG (↑)	EI (↑)	Q^{ABP} (↑)	SD (↑)	SF (↑)	RMSSE (↓)	SSIM (↑)	Q_{CB} (↑)	Q_{CV} (↓)
ADF (0,0,0)	1.464	6.788	1.921	58.405	4.582	46.529	0.519	35.185	14.132	0.104	1.400	0.474	777.817
CBF (0,0,2)	0.994	7.324	2.161	57.595	7.154	74.590	0.578	48.544	20.380	0.126	1.171	0.526	1575.148
FPDE (0,0,0)	1.366	6.766	1.924	58.402	4.538	46.022	0.484	34.931	13.468	0.104	1.387	0.460	780.114
Gfce (0,3,0)	1.931	7.266	1.844	55.939	7.498	77.465	0.471	51.563	22.463	0.173	1.134	0.535	898.946
GFF (0,0,0)	1.189	7.210	2.638	58.100	5.326	55.198	0.624	50.059	17.272	0.112	1.398	0.619	881.625
GTF (0,0,0)	1.285	6.508	1.991	57.861	4.303	43.664	0.439	35.130	14.743	0.118	1.371	0.414	2138.369
HMSD_GF (0,1,0)	1.164	7.274	2.472	57.940	6.246	65.034	0.623	57.617	19.904	0.116	1.394	0.604	532.958
Hybrid_MSD (0,1,0)	1.257	7.304	2.619	58.173	6.126	63.491	0.636	54.922	19.659	0.110	1.405	0.623	510.866
IFEVIP (0,0,0)	1.339	6.936	2.248	57.174	4.984	51.782	0.486	48.491	15.846	0.138	1.391	0.462	573.767
LatLRR (3,0,0)	1.684	6.909	1.653	56.180	8.962	92.813	0.438	57.133	29.537	0.169	1.184	0.497	697.286
LP_SR (2,2,2)	0.957	7.339	2.809	57.951	5.851	60.781	0.661	57.314	18.807	0.117	1.390	0.645	522.687
MGFF (0,0,0)	1.295	7.114	1.768	58.212	5.839	60.607	0.573	44.290	17.916	0.109	1.406	0.542	676.887
MSVD (0,0,2)	1.462	6.705	1.955	58.415	3.545	36.202	0.331	34.372	12.525	0.104	1.425	0.426	808.993
NSCT_SR (3,0,1)	0.900	7.396	2.988	57.435	4.692	67.956	0.646	52.475	19.389	0.131	1.277	0.617	1447.340
RP_SR (0,1,2)	0.994	7.353	2.336	57.777	6.364	65.220	0.566	55.808	21.171	0.122	1.332	0.606	888.848
TIF (0,0,0)	1.371	7.075	1.767	58.225	5.558	57.839	0.584	42.643	17.739	0.109	1.399	0.545	613.004
VSMWLS (0,0,0)	1.409	7.028	2.035	58.194	5.612	57.252	0.554	46.253	17.662	0.109	1.417	0.496	754.704
CNN (1,1,2)	1.030	7.320	2.653	57.932	5.808	60.241	0.658	60.075	18.813	0.118	1.391	0.621	512.569
DLF (3,0,0)	1.413	6.724	2.030	58.444	3.825	38.569	0.434	34.717	12.491	0.103	1.461	0.445	759.814
IFCNN (0,0,1)	1.419	7.122	2.068	58.246	6.228	64.645	0.589	48.521	19.359	0.108	1.403	0.531	495.289
ResNet (0,3,0)	1.364	6.734	1.988	58.441	3.674	37.255	0.407	34.940	11.736	0.104	1.460	0.445	724.831
SeAFusion (0,1,0)	1.543	6.967	2.120	57.301	5.655	58.877	0.561	49.628	17.733	0.134	1.393	0.460	416.935
SwinFusion (1,0,0)	1.338	6.938	2.282	57.321	5.605	57.992	0.575	52.855	18.045	0.135	1.406	0.489	399.224
U2Fusion (0,0,0)	1.316	7.200	1.946	57.966	6.241	65.831	0.532	50.058	18.288	0.114	1.331	0.540	719.791
YDTR (0,0,1)	1.568	6.828	2.124	58.015	4.333	44.591	0.452	44.980	15.082	0.116	1.435	0.436	679.953

图 4.3: 笔者在论文中展示的可见光红外图像融合算法的定量评价。图像来源于[11]。

在图像融合领域，已经有数十种定量评价指标被提出和用来对图像融合算法进行评价。图像融合定量评价指标可以从不同的角度进行分类，如图4.4所示。例如，根据是否需要参考图像，可以分为基于参考图像的评价指标和不基于参考图像的评价指标。另外，根据评价指标的定义方法，图像融合评价指标大致可以

被分为四种类型 [35]:

- 基于信息论的评价指标
- 基于图像特征的评价指标
- 基于结构相似性的评价指标
- 基于感知启发的指标



图 4.4: 图像融合定量评价指标的分类方法。从图中可以看出，评价指标有不同的分类方法。

此外，近年来，还有部分学者提出了基于深度学习的评价指标。在本书中，笔者采用第二种分类方法，并将基于深度学习的评价指标作为第五类评价指标。

定量评价对于图像融合任务很重要，因此在几乎每篇图像融合论文中研究人们都会给出定量指标的结果。需要指出的，一般来说，这些评价指标可以用于多种图像融合任务的性能评价中。然而，也有部分评价指标是针对具体的图像融合任务而设计，因此只用于某种图像融合任务。例如，香港城市大学马柯德老师针对多曝光图像融合任务设计了 MEF-SSIM 这个评价指标 [36]。

使用定量指标进行算法的性能评价也存在一些问题。

首先，单个评价指标具有片面性。荷兰计量经济学家桑内·布劳（Sanne Blauw）在他的著作《数据如何误导了我们》中写道：

随着经济社会的发展，人们不可避免地创造出了越来越多的抽象概念，来解

释这个世界，比如经济增长、能力、智力、信用、幸福等等。但作者提醒我们，如果我们忘记了，这些概念只是人为创造出来的，而是把它们当做客观存在的事物，那就很危险了。因为这样的话，我们会很容易忽视了“抽象概念”跟“真实世界”之间的那道屏障；忽视了所有概念背后都隐含着人们的价值判断；忽视了并不是所有概念都能够被量化，量化的方式也不只有一种。所以，当我们看到一个数据，衡量的是一个人为创造出来的概念的时候，最多只能把它理解为，是真相的一个切面，而不能把它当作全部的真相。

——桑内·布劳《数据如何误导了我们》

当前用到的图像融合评价指标，都是人为创造出来的概念，因此都只能反映图像融合效果的一个切面。尤其是在图像融合任务中不存在标准答案，因此这些评价指标均不能直接与标准答案进行对比。这进一步导致了在这些客观指标中也存在着片面性。

其次，容易出现评价不全面，即“盲人摸象”的情况。由前述分类可知，评价指标只能从某一个或一些方面去评价图像的特性。这就导致了单个评价指标不能全面评价图像的性能。在论文中，研究员一般会选择多个评价指标来对算法进行评价。然而，如果评价指标选择不当，很可能也无法对融合算法进行全面的评价。

最后，结果不具备可比性。因为评价指标多种多样并且不存在最好的指标，所以研究员一般是根据自己的喜好或者需求来选择一些评价指标呈现在论文中。这就导致了不同的论文使用不同评价指标组合的情况。这种情况在几乎每一种图像融合任务中都存在。这种现象使得很多论文提出的算法不好进行直接对比，从而不容易知道算法性能的优劣情况。

基于上述这些原因，目前图像融合的定量评价方法还有很大的研究空间。

4.2.3 图像融合评价方法现状

基于图像融合算法评价的特殊性，以及定性方法和定量方法的优缺点，目前在研究中通常使用定性评价和定量评价相结合的方法来对图像融合算法进行评价。在目前已发表的绝大多数图像融合论文中，论文作者们均给出了定性评价和定量评价的结果。如果论文作者在投稿时只给了定性结果或者定量结果，那么审稿人基本上都会要求作者增加另一种结果。

定性评价和定量评价，可以在一定程度上对图像融合算法的性能进行评价。

然而，需要指出的是，目前在定性评价的图像选择以及定量评价的指标选择上，都带有研究人员很强的主观性。

4.3 其他评价方法

目前，在图像融合相关的学术论文中，常规的评价方法是上述的定性评价和定量评价相结合的方法。不过，论文中也存在一些其他的评价方法。本节将对这些方法进行简要介绍。

1. 采用人力对所有融合图像进行打分、定性分析

香港城市大学马柯德老师曾经进行过多曝光图像融合的主观评价研究。他们构建了一个多曝光图像融合的图像库。该图像库中包含了 17 组源图像，每组源图像包含至少 3 张图像，即欠曝光图像、过曝光图像和中间图像，如图4.5所示。他们选用了 8 种多曝光图像融合算法来对这 17 组源图像进行融合，并一共生成了 136 张融合图像。然后，他们找了 25 位观察者来观察这些融合图像，并就融合图像的好坏给出一个 1 到 10 的分数。其中 1 代表融合图像的质量最差而 10 代表融合图像的质量最好。

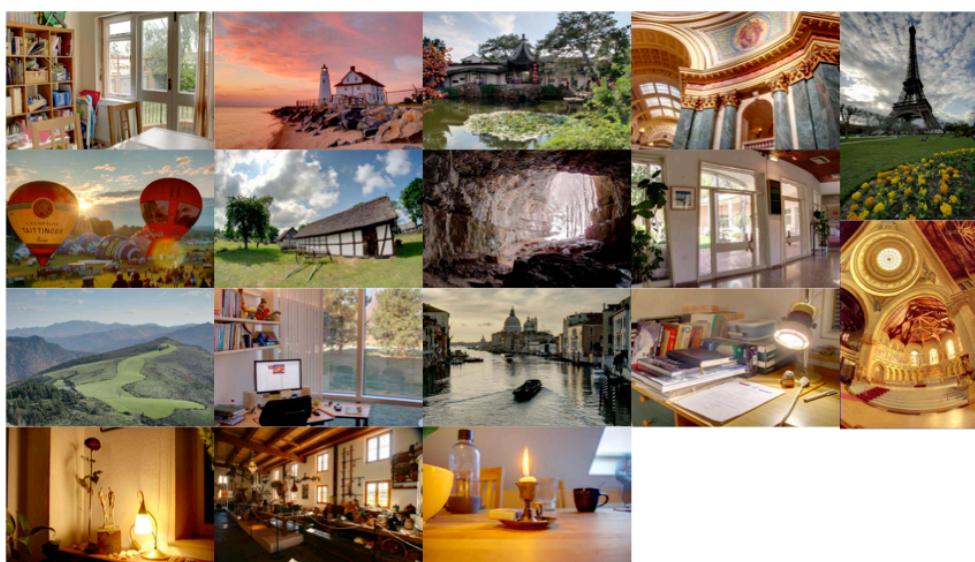


图 4.5：香港城市大学马柯德老师使用的多曝光图像融合源图像。图片来源于 [36]。

马柯德老师对得到的数据进行了详细的分析。他们得出了几下几点结论。首先，观察者对于这 8 种多曝光图像融合的结果的看法比较一致。其次，没有一种

算法可以在所有 17 组图像上得到最好的融合结果。第三，现有的³定量评价指标无法很好地评价多曝光图像融合算法的效果。基于这些分析，他们提出了一种针对多曝光图像融合任务的新的评价指标，即 MEF-SSIM。

类似的，研究人员们 [37] 最近对可见光与红外图像融合算法也做了类似的分析。他们找了 48 位观察者来对 200 组源图像的融合图像（由 20 种融合方法得到）进行评价。观察者每次比较两张图像，并将融合图像的质量分为 5 个等级。

这种找一批观察者对所有融合图像进行比较、打分的性能评价方法，可以对融合算法的性能进行比较好的评价，结果总体上比较可靠。然而，这种方法费时费力，也很难扩展到更大规模，例如更多源图像、更多对比方法的情况。因此，研究人员在论文中使用这种评价方法的情况并不多见。

2. 统计分析

除了单纯在论文中列出评价指标的值以外，有些学者在论文中会对图像融合的定量结果进行统计分析，以研究各方法的性能是否具有统计上的显著差异。例如，在医学图像融合中已有研究人员使用过这种方法 [38]。笔者在多曝光图像融合的研究中也使用过这种方法，如图4.6所示。总体来说，目前在图像融合论文中使用统计分析的情况还比较少见。

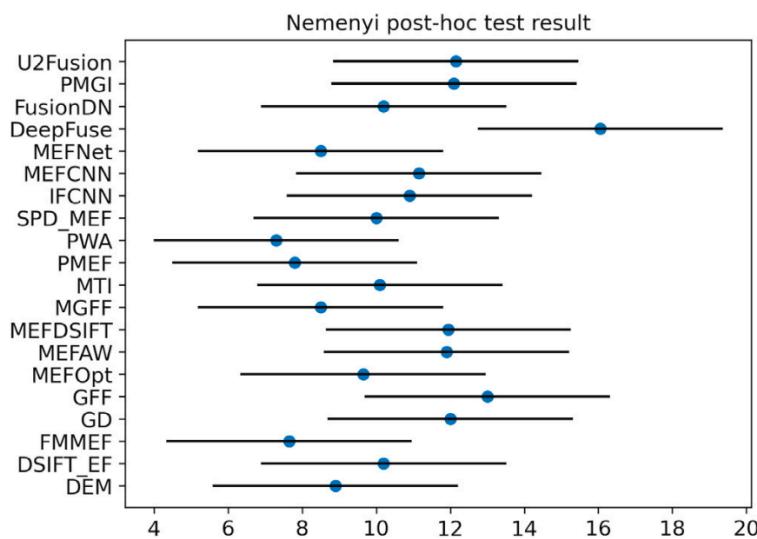


图 4.6: 笔者对多曝光图像融合算法进行的统计分析。图片来源于 [39]。

³该论文发表于 2015 年，所以这里“现有的”是指截止到 2015 年

4.4 近年来的发展特点

如前所述，不管是定性评价还是定量评价，都还存在着诸多问题。近年来，有一些学者（包括笔者），对如何更好地评价图像融合的性能进行了探索。本节简单讨论已有的一些探索。

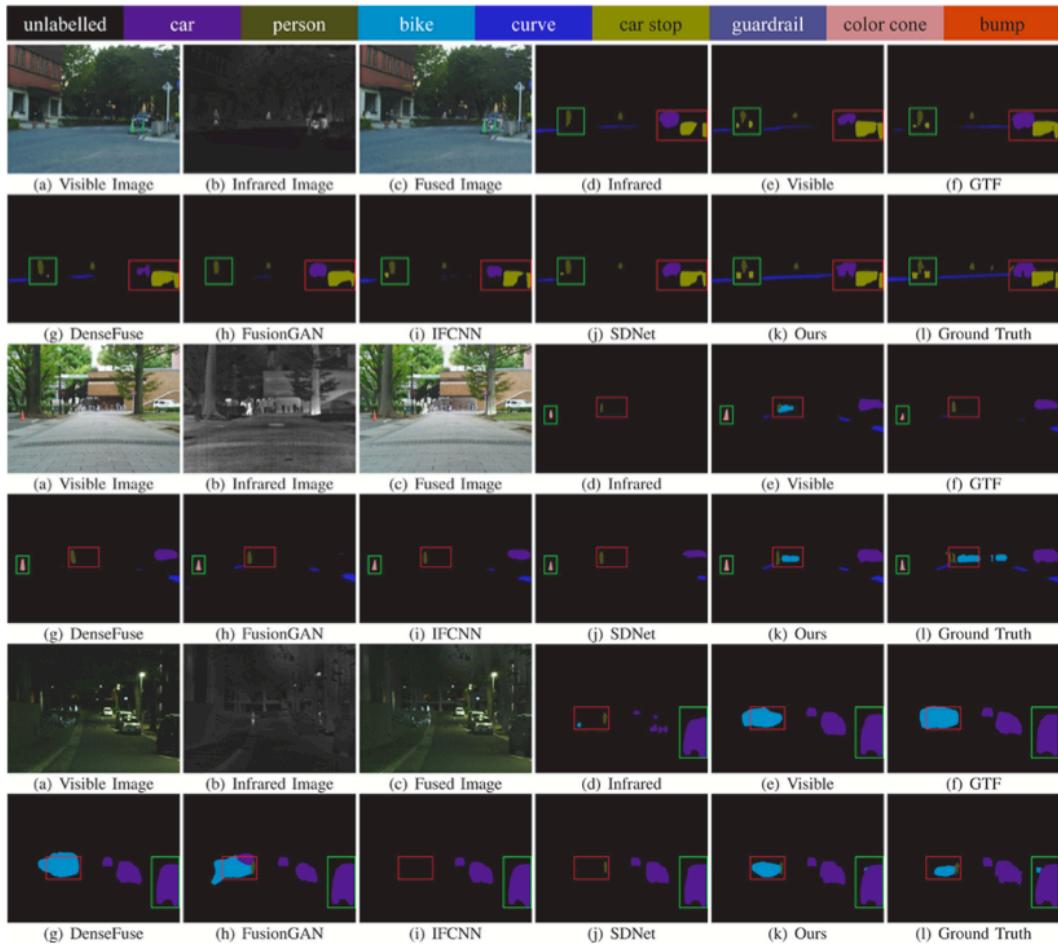


图 4.7: 武汉大学马佳义教授团队在论文中基于场景分割效果对可见光与红外图像融合算法进行性能评价。图片来源于 [40]。

1. 基于下游应用的性能评价

在 2022 年以前，几乎所有的可见光与红外图像融合论文在做性能评价时，均只考虑基于图像融合评价指标的定量评价和基于人眼观察的定性评价。2021 年，笔者在笔者的微信公众号“笑书神侠读博学”发表了《双剑合璧，为了好看还是实战》一文，指出在图像融合领域不能只考虑融合图像的质量，还得考虑融合图像对于下游应用的提升作用。因此，用下游应用来对图像融合算法的质量进行评价是很重要的。

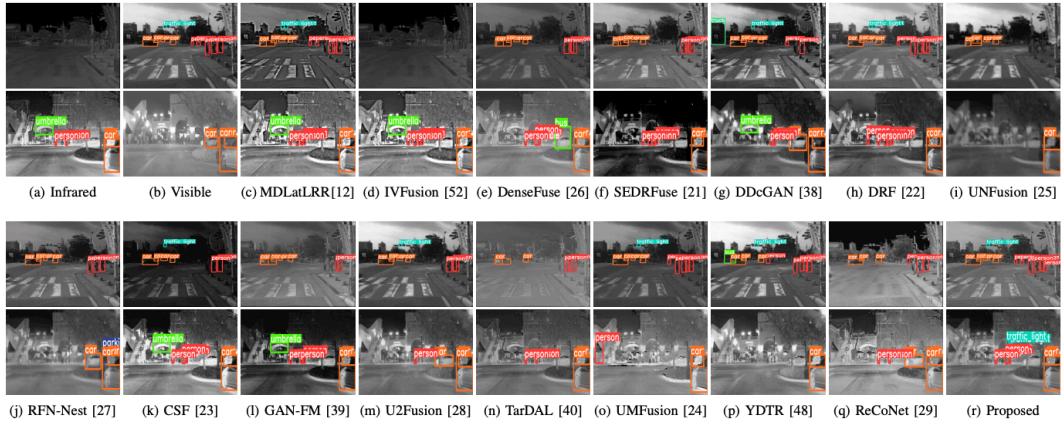


图 4.8: 基于目标检测效果对可见光与红外图像融合算法进行性能评价。图片来源于 [41]。

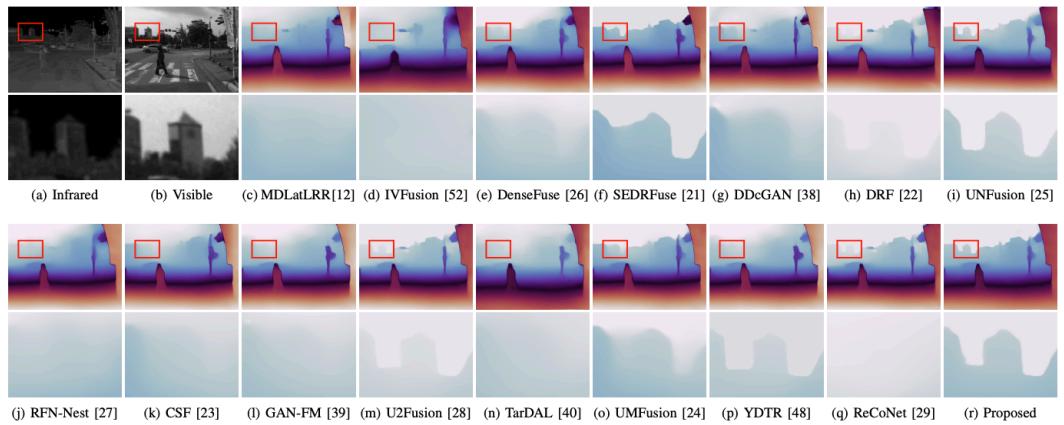


图 4.9: 基于深度估计效果对可见光与红外图像融合算法进行性能评价。图片来源于 [41]。

笔者的呼吁引起了一些图像融合研究人员的注意。在笔者的文章发出以后，一些学者在设计图像融合算法的阶段会将下游目标任务考虑在内。除此以外，一些学者在论文中除了进行前述的定性和定量评价以外，开始通过下游应用来对图像融合的效果进行评价。例如，武汉大学马佳义教授团队在他们的数篇论文中使用下游应用对图像融合算法的性能进行了评价。图4.7中给出了一个基于场景分割的效果对图像融合算法进行性能评价的示例。此外，Park 等人 [41] 在他们最新发表的论文中进行了基于目标检测和深度估计的性能评价，如图4.7和图4.9所示。

2. 建立评价基准

在计算机视觉的很多研究方向中，都有广泛使用的评价基准（benchmark），例如目标跟踪领域的 OTB[42, 43] 和 VOT[44]。这些评价基准为相关领域的方法

进行性能对比提供了很好的平台，为相关领域的发展做出了很大贡献。然而，在图像融合领域却一直没有评价基准。基于此种情况，笔者在近几年投入巨大的精力开发了图像融合领域的三个评价基准⁴，分别为可见光与红外图像融合评价基准 VIFB (Visible and infrared image fusion benchmark)、多曝光图像融合评价基准 MEFB (Multi-exposure image fusion benchmark) 和多聚焦图像融合评价基准 MFIFB (Multi-focus image fusion benchmark)。

本节简单介绍图像融合领域评价基准的必要性和笔者研发的可见光与红外图像融合评价基准。

1) 图像融合评价基准的必要性

在 2013 年以前，想要发表一篇目标跟踪相关的论文非常简单，只需要随便找几个图像序列做实验，和几个算法比较一下证明自己算法更优即可达到发表要求。这些序列和算法都可以自己随机找，因此在文献中出现了一种现象就是，大家说都自己的算法好，也都有实验证明。可是，由于在实验中使用的序列和进行对比的方法都不相同，因此谁也没办法证明自己的算法比其他算法都好。

这种现象就像《射雕英雄传》里第一次华山论剑以前的东邪、西毒、南帝、北丐、中神通一样，每个人都觉得自己是天下第一，但是又没办法证明自己是天下第一，也没法让对手服气。后来，他们组织了华山论剑，进行了一场公平的比赛才决出王重阳为天下第一，并且其他人都心服口服。

2013 年，加州大学美熹德分校的吴毅研究员（现为南京信息工程大学教授）发起了单目标跟踪领域里的“华山论剑”——OTB (Object tracking benchmark)^[42]。OTB 中包含了 50 个图像序列和 29 种目标跟踪算法，并提供了相关软件平台用于进行算法运行、结果绘制和对比。OTB 最大的意义在于提供了一个算法性能对比的平台，结束了大家各自随机找序列和对比算法的历史。从此以后，想要发表单目标视觉跟踪相关的论文，基本上必须提供算法在 OTB 平台下的运行结果。

近些年来，各个领域的评价基准被不断提出。例如，行人轨迹预测领域的 ETH 和 UCY，行人意图预测领域的 JAAD 和 PIE，人体姿态估计领域的 COCO，等等。这些评价基准已经成为了研究人员在论文中证明自己算法性能必不可少的一部分。

⁴笔者在上海交通大学工作期间，曾在上海交通大学人工智能研究院参加过一次著名学者 Ming-Hsuan Yang 的学术报告会。在该报告中，他介绍了他的团队当年制作目标跟踪领域首个评价基准——OTB 的经历，给了笔者一些启发

然而，长期以来，图像融合领域面临着和 2013 年以前的目标跟踪领域同样的问题。尽管关于图像融合的文献非常多，但是存在一个重大问题，那就是缺乏评价基准。无论是可见光与红外图像融合，多聚焦图像融合，多曝光图像融合还是医学图像融合，都是如此。这主要体现在以下几个方面：

- 没有统一的数据集。尽管有一些数据集相对常用，但并没有形成标准或者共识。因此，在文献里普遍存在测试图片不一致的情况。很多时候，是张三用 A 图像对进行性能测试，而李四用 B 图像对进行性能测试，导致张三和李四的算法无法进行直接对比。
- 没有统一的评价指标。和目标跟踪等有标准答案（ground truth）的领域不同，图像融合领域一般没有标准，从而导致在对融合结果的评价时没有标准答案。如前所述，研究人员从不同的角度出来，提出了各种各样的图像融合评价指标。事实上，到目前为止，研究人员大约设计了不下 30 种各种各样的评价指标用于评价融合图像的质量。然而，这些指标中并不存在最好的指标，目前也没有一些评价指标是研究人员公认必须使用的。实际上，指标的评价效果非常取决于测试图像，并且指标之间还存在互相矛盾的现象。因此，在文献中经常可以看到不同的论文中使用不同评价指标的现象。

Reference (year)	Venue	No. of test image pairs	Objective evaluation metrics
FusionGAN [32] (18) VIF-Net [59] (19) U2Fusion [63] (19) GANMcC [101] (20) Liu et al. [92] (20) SDNet [93] (20) Xu et al. [190] (22) MHTNet [114] (21)	INFUS TIP TPAMI TIM TIP IJCV PR TIM	7 (TNO) + 31 (INO) 9 (TNO and INO) 20 (TNO) + 45 (RoadScene) 16 (TNO) + 30 (RoadScene) 20 (TNO) 10 (TNO) 20 (TNO) + 20 (RoadScene) 20 (TNO) + 20 (KAIST) + 20 (BEMP)	5 (EN, SD, SSIM, SF, VIF) 5 (MI, $Q^{AB/F}$, PC, Q^{NCIE} , UIQI) 4 (SSIM, PSNR, CC, SCD) 6 (SSIM, CC, SCD, EN, SD, MI) 5 (VIF, AG, SF, SCD, $Q^{AB/F}$) 4 (EN, FMI _{dct} , PSNR, MG) 6 (Q_{abf} , SCD, VIF, SF, SSIM, PSNR) 6 (SCD, VIFF, EN, SD, MI, Q_{CV})
DDcGAN [49] (18) DIDFuse [54] (19) FusionDN [61] (19) PMGI [62] (19) Liu et al. [91] (20)	IJCAI IJCAI AAAI AAAI MM	20 (TNO) 40 (TNO) + 52 (NIR) + 40 (FLIR) 44 (RoadScene) 17 (TNO) 37 (TNO) + 26 (RoadScene)	4 (EN, SD, SF, PSNR) 6 (EN, MI, SD, SF, VIF, AG) 4 (SD, EN, VIF, SCD) 6 (SSIM, $Q^{AB/F}$, EN, FMI, SCD, CC) 4 (SD, VIF, CC, SCD)

图 4.10: 在可见光和红外图像融合领域，很多研究论文使用不同的测试集和评价指标来进行算法性能评价。图片来源于 [11]。

图 4.10 中展示了可见光与红外图像融合领域的研究论文使用不同测试图像和不同评价指标的现象。这种现象在深度学习被引入图像融合领域以前就存在，在深度学习被引入以后，依然存在而且似乎更严重了。基于这两个原因，在图像融合领域广泛存在着“王婆卖瓜，自卖自夸”的现象。这严重影响了研究人员对图像融合算法性能的正确认识，也严重阻碍了图像融合算法从学术界走向工业

界。因此，建立图像融合领域的评价基准，以便全面、公平地测试和评价图像融合算法的性能，是非常有必要的。笔者始终认为，只有明确和正视基于深度学习的图像融合算法的真实性能，才有可能实现进一步的发展。否则的话，沉浸在“自我 SOTA (state-of-the-art) ”的假象之下，领域难以真正进步。

需要指出的是，缺乏合适的评价基准的问题也出现在人工智能的其他一些领域。2020 年 5 月 27 日，国际顶级期刊 Science 上发表了一篇题为 “Eye-catching advances in some AI fields are not real” 的文章。文章指出目前一些 AI 领域的发展是虚假繁荣，并引用了一篇麻省理工学院的研究人员针对模型剪枝算法所做的实验，表明很多宣称为 SOTA 的算法，其实性能并不怎样。此外，该文章认为目前很多算法之间根本不存在可比性（因为模型不同，数据集不同等等原因），并提供了一个比较模型剪枝算法的平台。

2) 可见光与图像融合评价基准: VIFB

本节介绍笔者开发的可见光与红外图像融合 benchmark，即 visible and infrared image fusion benchmark (VIFB)。VIFB 是可见光与红外图像融合领域的首个评价基准，也是整个图像融合领域的首个评价基准，发表以后引起了研究人员的广泛关注。截止到 2024 年 8 月，VIFB 的论文被引用了 201 次。VIFB 被包括英国帝国理工学院、美国波士顿大学、加拿大国家研究委员会、加拿大英属哥伦比亚大学、加拿大卡尔加里大学、挪威科技大学、纽约大学阿布扎比分校、土耳其伊斯坦布尔科技大学、土耳其伊兹密尔经济大学、马来西亚多媒体大学、上海交通大学、中国科学院、中国科学院大学、国防科技大学、上海科技大学、西北工业大学、厦门大学、东南大学、清华大学深圳研究生院、南京航空航天大学在内的数十家国内外单位的研究人员在论文中使用过。

在 VIFB 中，笔者收集制作了一个含有 21 对可见光与红外图像对的测试集、一个含有 20 种可见光与红外图像融合算法的代码库和 13 种评价指标。此外，我们还制作了软件平台，在该平台上我们统一了这 20 种算法的接口，可以一键运行并获得融合结果（420 张融合图像）和评价指标的计算结果。

另外，在 VIFB 中还可以非常方便地添加和运行基于 Matlab 的算法（使用我们设计的函数接口）并计算评价指标。或者，将在其他环境中（例如 Tensorflow, PyTorch）运行得到的融合图像添加到 VIFB 中进行评价指标的计算。例如，在笔者最新发表的论文中 [11]，笔者在 VIFB 中加入了最新发表的 5 种图像融合算法的结果。此外，在 VIFB 中加入新的源图像也非常方便。

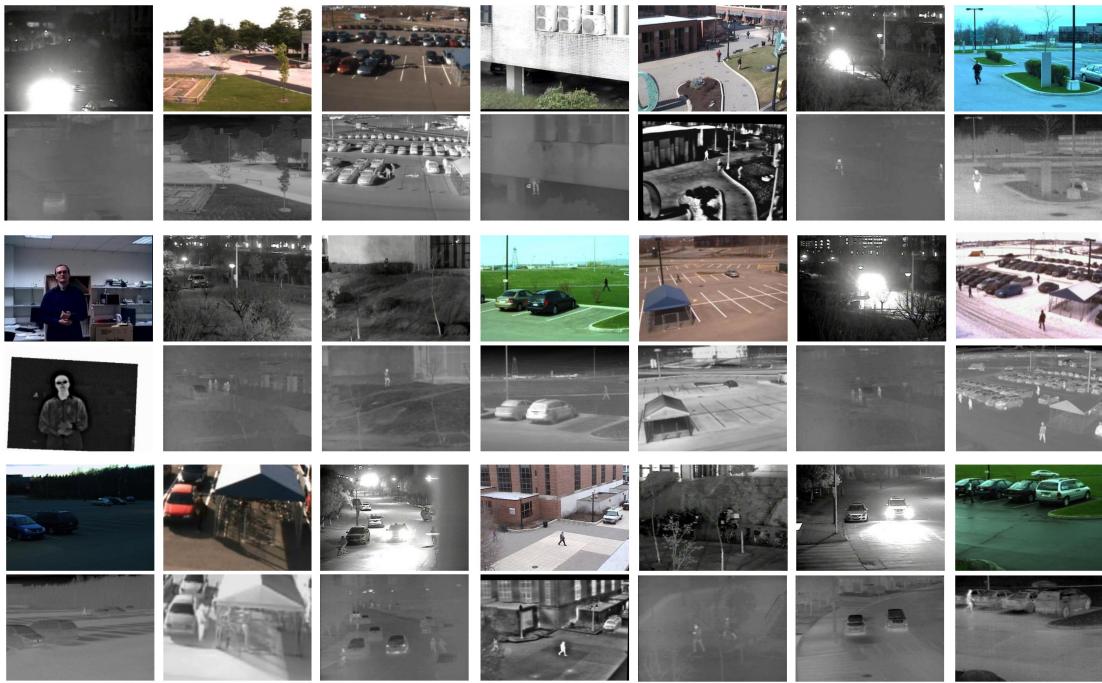


图 4.11: VIFB 中的可见光与红外图像对。图中第 1、3、5 行是可见光图像，第 2、4、6 行是对应的红外图像。图像来源于 [2]。

VIFB 对于可见光与红外图像融合领域的主要贡献为：

(1) 数据集

VIFB 包含 21 对可见光和红外图像对。这些图像是由笔者从互联网和融合跟踪数据集 [8] 中搜集整理得到的。这些图像对涵盖了许多不同的工作环境和条件，例如室内、室外、低光照强度、过曝光等。因此，这个数据集可以用来测试算法的鲁棒性和泛化性能。每一个图像对都进行了严格配准，以保证图像融合可以顺利进行。这些图像对同时涵盖了多种分辨率，例如 320*240, 630*460, 512*184 和 452*332。图4.11展示了这些图像对。

(2) 代码库

如前所述，近年来研究人员提出了很多种可见光与红外图像融合算法。然而，仅有一部分研究论文提供了源代码，而且这些源代码有不同的输入和输出接口（格式）。这使得使用这些代码来生成结果和进行性能分析比较困难。为了解决这个问题，在 VIFB 中，笔者集成了 20 种近年来发表的可见光与红外图像融合方法，包括 MSVD [45], GFF [46], MST_SR [47], RP_SR [47], NSCT_SR [47], CBF [5], ADF [48], GFCE [49], HMSD_GF [49], Hybrid-MSD [50], TIF [51], GTF [52], FPDE [53], IFEVIP [54], VSM_WLS [55], DLF [56], LatLRR [57], CNN [22], MGFF [3] 和 ResNet [58] 法。这 20 种算法的出处、发表年份、方法类别如表4.1所

示。

表 4.1: VIFB 中已集成的可见光与红外图像融合算法。

方法名称	发表年份	期刊/会议	分类
ADF [48]	2016	IEEE Sensors Journal	多尺度
CBF [5]	2015	Signal, image and video processing	多尺度
CNN [22]	2018	IJWMIP	深度学习
DLF [56]	2018	International Conference on Pattern Recognition	深度学习
FPDE [53]	2017	International Conference on Information Fusion	子空间
GFCE [49]	2016	Applied Optics	多尺度
GFF [46]	2013	IEEE Transactions on Image Processing	多尺度
GTF [52]	2016	Information Fusion	其他
HMSD_GF[49]	2016	Applied Optics	多尺度
Hybrid_MSD [50]	2016	Information Fusion	多尺度
IFEVIP [54]	2017	Infrared Physics & Technology	其他
LatLRR [57]	2018	arXiv	显著性检测
MGFF [3]	2019	Circuits, Systems, and Signal Processing	多尺度
MST_SR [47]	2015	Information Fusion	混合
MSVD [45]	2011	Defense Science Journal	多尺度
NSCT_SR [47]	2015	Information Fusion	混合
ResNet [58]	2019	Infrared Physics & Technology	深度学习
RP_SR [47]	2015	Information Fusion	混合
TIF [51]	2016	Infrared Physics & Technology	显著性检测
VSMWLS [55]	2017	Infrared Physics & Technology	混合

这些算法涵盖了许多不同类型的可见光与红外图像融合算法因此基本可以代表可见光与红外图像融合领域的最新发展水平和趋势。2023 年，笔者又往 VIFB 中添加了 5 种最新发表的图像融合算法，即 SeAFusion [40], SwinFusion[59], U2Fusion[60], YDTR[61], IFCNN[62]。

为了将算法集成到图像融合框架中，并且为了便于用户使用，笔者在 VIFB 中设计了统一的代码接口。通过使用这个接口，任何基于 Matlab 开发的图像融合算法均可以很方便地被集成到 VIFB 中。需要说明的是，对于那些不是基于 Matlab 开发的代码，我们也设计了接口以便将这些算法的融合结果接入到 VIFB 中进行性能对比分析。

(3) 全面的性能评价

在可见光与图像融合领域，已经有多种评价指标被提出来，例如互信息 (mutual information)，空间频率 (spatial frequency) 和交叉熵 (cross entropy)。然而，在这些评价指标中，没有任何一种比其他所有指标都好。为了全面地评价融合算法的性能，在 VIFB 中，我们集成了 13 种评价指标以便可以全面地评价图像融合算法的性能。在 VIFB 中可以非常方便地计算这些评价指标，从而可以非常方便地进行性能比较。在 VIFB 中实现的 13 种评价指标已在表 4.2 中列出。表中的“+”表示当该指标值越大时，图像融合算法性能越好；“-”表示当该指标值越小时，图像融合算法性能越好。从表中可以看出，VIFB 中的评价指标包含了 Liu 等人 [35] 提出的全部四种类型的评价指标。此外，笔者基于 VIFB 的数据集运

表 4.2: VIFB 中已集成的图像融合算法评价指标。‘+’表示指标值越大，性能越好；‘-’表示指标值越小，性能越好。

种类	名称	含义	+/-
基于信息论的评价指标	CE [63]	交叉熵	-
	EN [64]	熵	+
	MI [65]	互信息	+
	PSNR [66]	峰值信噪比	+
基于结构相似性的指标	SSIM [67]	结构相似性度量	+
	RMSE [66]	平均均方误差	-
基于图像特征的指标	AG [68]	平均梯度	+
	EI [69]	边缘强度	+
	SD [70]	标准差	+
	SF [71]	空间频率	+
	$Q^{AB/F}$ [72]	基于梯度的评价指标	+
基于人类感知启发的指标	Q_{CB} [73]	Chen-Blum metric	+
	Q_{CV} [74]	Chen-Varshney metric	-

行了所有 20 种图像融合算法，计算了所有融合图像的 13 种评价指标，并进行了大量的性能对比分析，以便理解这些图像融合算法。

(4) 小结

作为初步的尝试，我们希望 VIFB 的出现，可以给可见光与红外图像融合领域的研究者们提供一个运行算法和评价算法的平台，可以缓解可见光与红外图像融合领域文献中缺乏统一评价平台而导致的“王婆卖瓜、自卖自夸”现象，可以促进代码开源，为领域的发展助一分力。

4.5 图像融合评价方法的发展趋势

本节简单介绍笔者对图像融合评价方法的发展趋势的看法。

4.5.1 设计更好评价基准

笔者认为图像融合算法性能评价的一个重要发展趋势，是设计更好的评价基准。尽管笔者近年来针对 3 个图像融合任务提出了 3 个评价基准，对图像融合

算法的性能评价进行了一些有益的探索，但是目前的评价基准还有很大的改进空间。笔者认为，图像融合评价基准可以在以下方面进行改进：

1. 使用大规模数据集

笔者研发的 VIFB 中使用了 21 组源图像，MEFB 中使用了 100 组源图像，规模都不大，并且只包含了测试集。考虑到场景的多样性需求，笔者认为未来的图像融合评价基准中需要包含更大规模的数据集，并且需要包含训练集、验证集和测试集。这样可以使得算法的比价更加公平一些。

2. 选用更加科学的评价指标组合

笔者在 VIFB 中选择了 13 种定量评价指标，在 MEFB 中选用了 20 种评价指标。笔者认为未来的图像融合评价基准可以考虑包含更加科学的评价指标组合，以便使得图像融合的定量评价更加科学。然而，这是一项非常有挑战性的任务。

3. 找一批观察者对所有融合图像进行评价和打分

香港科技大学马柯德老师在多曝光图像融合的研究过程中曾经找过 25 位观察者对 136 张融合图像进行过打分。笔者认为在未来的图像融合评价基准中可以包含类似的由一批观察者打分得到的定性评价结果。然而，考虑到大规模的数据集和大量的对比算法，这项工作很费时费力。

4.5.2 基于具体应用的性能评价

著名人工智能专家、硅谷知名投资人吴军博士在他的著作《大学之路》中谈到大学排名的时候提到，对大学进行排名的时候是一件非常困难的事情，并且往往会出现同一所大学在不同排行榜上差别非常大的情况。因此，他建议家长和学生在选择大学时，与其看大学的排名，不如根据自己的需要从几个角度去综合地“感受”那些大学。

笔者认为，在对图像融合算法进行性能评价也是如此。选择不同的指标和数据集，可能得到大相径庭的图像融合方法排名。因此，笔者认为在对图像融合方法进行排名的时候，不妨也根据自己的需要来进行排名。具体来说，图像融合任务主要的目的是提供更好的源图像以便提升下游应用的性能。因此，在对图像融合方法进行排名的时候，不妨基于应用来进行排名。例如，将各种图像融合方法

应用于某个基于可见光和红外图像的目标跟踪数据集，然后看看这些方法生成的融合图像对于目标跟踪任务的提升作用有多大。

这样做有两个好处。首先，一般来说，目标跟踪、目标检测等下游任务是有标准答案的，因此，这种做法可以将没有标准答案的图像融合任务通过有标准答案的下游任务来进行评价。其次，通过以应用为导向的排名，我们可以筛选出真正对自己感兴趣的应用有利的融合方法，以便在实际中进行应用。这也有利于促进图像融合算法的实际应用。

值得说明的是，目前基本上只对可见光与红外图像融合算法有基于应用的性能评价方案，而在其他图像融合任务中鲜有基于应用的性能评价方法。

4.6 小结

图像融合方法的性能评价是图像融合研究中的一个重要领域，也是一个非常困难的领域。在本章中，笔者对主要的图像融合性能评价方法进行了介绍，并介绍了近年来图像融合领域的评价基准的发展情况。最后，笔者对图像融合算法性能评价的发展趋势进行了展望。

第5章

可见光与红外图像融合

红外相机给我们提供了从另一个视角感知世界的方法。

——笔者

本章主要讨论基于深度学习的可见光与红外图像融合。在几种图像融合任务中，可见光与红外图像融合是最受关注、发表论文最多的任务之一。这主要是因为可见光与红外图像融合具有很好的应用价值。

5.1 红外图像：从另一个视角感知世界

请读者朋友们看一下下面这两段文字：

他又倒跃一步。其时天色将明，但天明之前一刻最是黑暗，除了刀光闪闪之外，**睁眼不见一物**。他所学的独孤九剑，要旨是看到敌人招数的破绽所在，乘虚而入，此时敌人的身法招式全然无法见到，**剑法便使不出来**。

——金庸《笑傲江湖》二十四 蒙冤

“独孤九剑”的要旨，在于一眼见到对方招式中的破绽，便即乘虚而入，后发先至，一招制胜，**但在这漆黑一团的山洞之中，连敌人也见不到，何况他的招式，更何况他招式中的破绽？**处此情景，“独孤九剑”便全无用处。

——金庸《笑傲江湖》三十八 聚歼

上面这两段文字出自金庸先生的著名武侠小说《笑傲江湖》。小说中的文字充分描述了黑暗对于人的视觉的影响。可以看到，在光线很差的时候，人的视觉

会受到很大的影响，难以正常工作。

斯坦福大学李飞飞教授在《我看的世界》一书中介绍动物学家安德鲁·帕克的观点时说，引发寒武纪生命大爆发的导火线是一种能力的出现：光敏感性，这也是现代眼睛形成的基础。

相机的成像原理与人眼类似。现实生活中最常见的相机是可见光相机，例如手机上搭载的相机和人们通常使用的数码相机。这种相机通过相机中的感光元件接收可见光来生成图像，因此这种相机生成的图像是可见光图像，包含灰度图像和彩色图像¹。这里的彩色图像，也就是我们常说的 RGB (Red-Green-Blue) 图像。这些图像之所以被称为 RGB 图像，是因为每张图像包含 3 个通道，分别为 R (红色)、G (绿色)、B (蓝色) 通道。图5.1中展示了一个例子。注意，彩色图像也可转换为单通道的灰度图像。

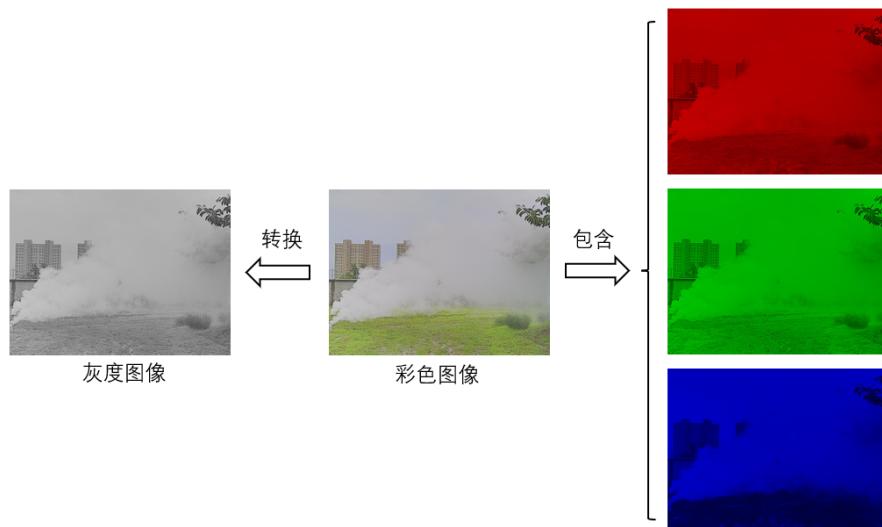


图 5.1: 一张彩色图像可以分为 R (红色)、G (绿色)、B (蓝色) 三个通道，也可以转换成单通道的灰度图像。图像来源于 M³FD 数据集 [75]。

长期以来，人们习惯了通过可见光图像来描述和感知世界。到目前为止，绝大多数计算机视觉应用都是基于可见光图像来开展的。这是因为可见光图像包含了丰富的细节和纹理信息，为许多计算机视觉应用提供了好的图像源。然而，可见光图像有一个明显的缺点，即容易受到光照条件和烟、霾等的影响。这和前文中《笑傲江湖》中人的视觉受光线影响是类似的。例如，在图5.1中，烟的后面是有两个人的，但是我们在可见光图像中完全看不见他们。

实际上，除了可见光相机以外，还有一些其他种类的相机，其中一种称为热

¹现在灰度相机已经很少了。在本书中，除非特别说明，笔者并不区分可见光图像和彩色图像

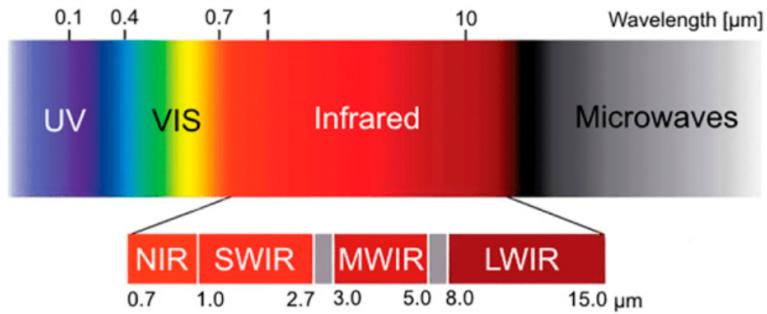


图 5.2: 可见光和红外图像的波长范围。其中“VIS”是指可见光图像，而“Infrared”是指红外图像，“LWIR”是指热红外图像。图像来源于 [76].



图 5.3: 与图5.1中可见光图像对应的热红外图像。图像来源于 M³FD 数据集 [75].

红外相机²。热红外相机通过感知物体的热辐射来进行成像，因此热红外相机探测的是物体的温度信息。因此，热红外相机的成像不受光照条件和烟、霾等的影响。图5.2中显示了红外线和可见光的波长范围。

图5.3中显示了与图5.1中的可见光图像对应的热红外图像。从图中可以看出，尽管现场有浓烟，我们依然可以在红外图像中非常清楚地观察到两个人。

从以上分析中可以看出，红外相机可以给我们提供从 RGB 相机中看不到的热量信息。也就是说，红外相机给我们提供了从另一个视角感知世界的方法。

5.2 可见光与红外图像融合概述

尽管红外图像给人们提供了一个新的感知世界的视角，并且这个视角在一些时候非常有效，但是仅仅依靠这个视角来感知世界很多时候也是不够的。这主要是因为红外图像中缺乏一些在可见光图像中存在的纹理和细节信息。例如，在图5.4中，在第一列图像中，因为浓烟的存在，我们无法从可见光图像中看见人，

²除非特别说明，在本书中提到红外的时候，均是指热红外（即远红外）。



图 5.4: 可见光与红外图像融合示意图。从上到下分别为可见光图像、红外图像和融合图像。在第一列图像中，因为浓烟的存在，我们无法从可见光图像中看见人，但是可以从红外图像中看到。在第二列和第三列图像中，我们较难从红外图像中将人和周围的人或者背景分开，但是在可见光图像中可以轻易做到。这体现出了可见光和红外图像的互补性。这里的融合图像由笔者使用 IFCNN 方法 [62] 获得。

但是可以从红外图像中看到。在第二列和第三列图像中，我们较难从红外图像中将人和周围的人或者背景分开，但是在可见光图像中可以轻易做到。

由上述例子可知，可见光和红外在一定程度上具有互补的特性。如果能将可见光图像和红外图像进行融合，那么人们将可以更好地感知世界。可见光与红外图像融合（Visible and infrared image fusion, VIF）因此应运而生。

可见光与红外图像融合是图像融合领域的一个重要分支。如前文所述，可见光与红外图像融合主要是为了将可见光与红外图像中的互补信息进行融合，以便得到更好的信息便于人们观察或者便于下游任务取得更好的效果，如图5.4所示。

多年以来，可见光与红外图像融合一直是比较活跃的研究领域。西安理工大学和西京大学的科研团队是最早将深度学习引入到可见光和红外图像融合领域的科研团队之一。他们于 2017 年将深度学习引入了可见光与红外图像融合领域 [77]。从那以后，许多基于深度学习的可见光与红外图像融合方法被提出来。图5.5中展示了笔者统计的从 2018 年到 2022 年，每年发表的基于深度学习的可见光与红外图像融合方法的论文数量的情况³。可以看出，基于深度学习的可见

³不完全统计

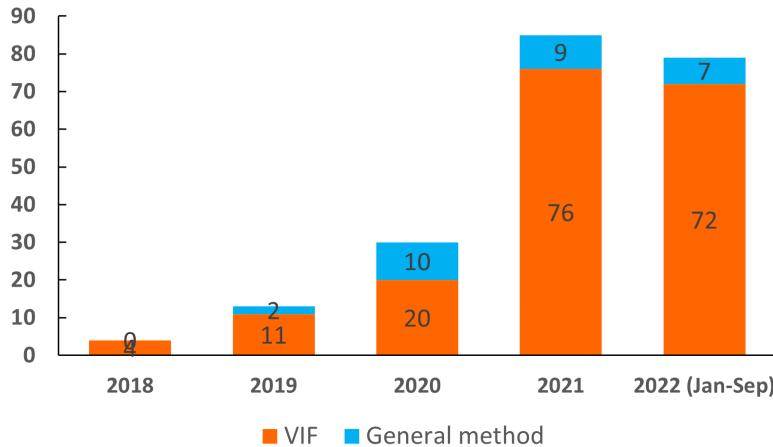


图 5.5：基于深度学习的可见光与红外图像融合方法的论文数量迅速增加。注意，基于深度学习的通用图像融合方法（可以应用于包括可见光与红外图像融合在内的多种图像融合任务）也被统计在内。

光和红外图像融合的论文数量增加得很快。

目前，可见光与红外图像融合已是图像融合领域被研究得最多的任务之一，也成为了图像处理的重要研究领域之一。可见光与红外图像融合已被应用于许多应用中，例如目标跟踪、目标检测、场景分割、生物特征识别、人群计数和机器人导航。

可以预见，在未来，图像融合领域发展的一个重要趋势是基于深度学习的方法，并且将有一大批研究成果出现。

5.3 传统融合方法概述

从方法上来讲，可见光与红外图像融合方法主要分为两类，即传统方法和基于深度学习的方法。在深度学习方法被引入图像融合领域来讲，根据采用的理论来进行划分，主要的图像融合方法包括基于多尺度变换的方法，基于稀疏表达的方法，基于子空间的方法，基于显著性检测的方法，混合模型和其他方法。这是武汉大学马佳义教授团队的分类方法 [78]。

总体来说，这些方法包含两种。一种是基于空间域的方法，一种是基于变换域的方法。基于空间域的方法是指直接在空间域对源图像进行操作从而得到融合图像的方法，主要包含基于像素的、基于块的和基于区域的方法。基于变换域的方法是指首先将源图像变换到某个变换域，然后在该变换域内进行图像融合（一般以系数的形式），最后再用逆变换得到融合图像的过程。常用的变换包括多尺度变换（例如小波变换）、压缩感知、稀疏表达等。

然而，由于传统方法具有一些缺点，近年来，研究人员开始研究基于深度学习的可见光红外图像融合方法。

5.4 使用深度学习做图像融合的动机

一个典型的可见光与红外图像融合方法通常包括三个阶段：特征提取、特征融合和图像重建。首先，提取可见光和红外图像的特征。然后，将这些特征进行融合。最后，执行图像重建以获得融合图像。

对于可见光与红外图像融合方法，所有三个阶段都对性能至关重要。首先，应该提取可见光和红外图像的互补特征。具体来说，需要提取可见光图像中的细节和纹理信息以及红外图像中的显著信息。其次，需要有效地融合这些互补特征。最后，需要有效地重建融合图像。然而，在传统可见光与红外图像融合方法中，这些阶段都是手工设计的，在处理各种工作条件时并不十分有效。因此，传统方法在提供高质量融合图像方面可能能力有限。

深度学习在各种图像处理领域表现出巨大成功。在过去几年中，研究人员开始关注使用深度学习进行可见光与红外图像融合，旨在实现更好的融合性能。深度学习可以解决图像融合中的几个重要问题。首先，深度学习可以提供相较于人工特征更好的特征。其次，深度学习可以提供自适应的权值，这些权值在图像融合中非常重要。值得注意的是，深度学习可能只应用于图像融合过程的一部分。

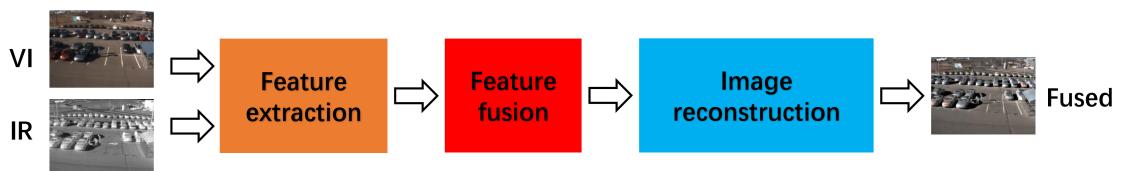


图 5.6: 可见光与红外图像融合方法的三个阶段，即特征提取、特征融合和图像重建。

5.5 基于深度学习的融合方法发展历程概述

最近几年，有许多基于深度学习的图像融合方法被提出来。关于深度学习方法，目前卷积神经网络 (CNN)、生成对抗网络 (GAN)、自编码器 (Autoencoder)、变换器 (Transformer) 和扩散模型 (Diffusion model) 均已被用于进行图像融合。除此之外，深度学习方法也被用于图像融合方法性能评价。此外，有监督学习方

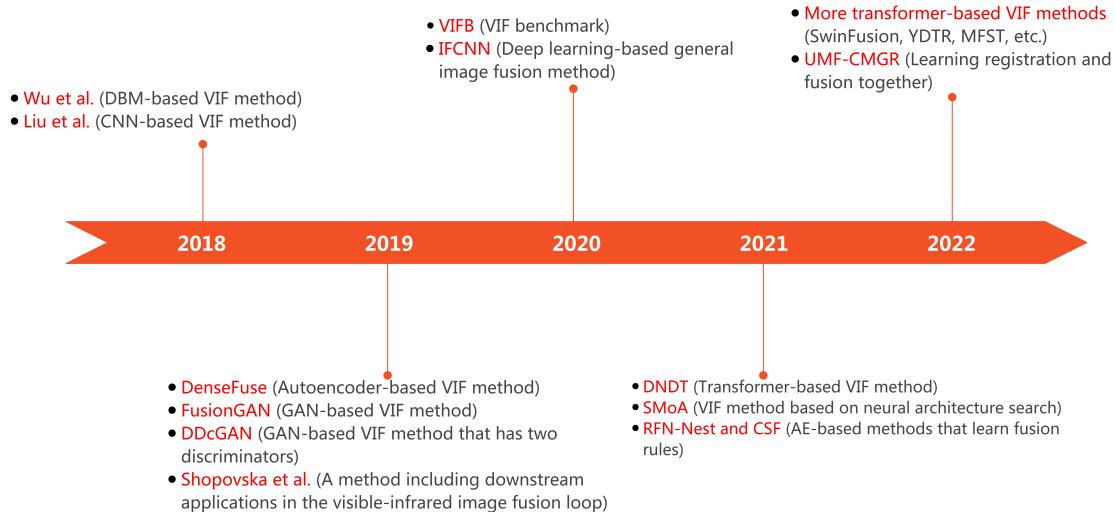


图 5.7: 基于深度学习的可见光与红外图像融合方法发展时间轴及里程碑方法。

法和无监督学习方法也均已被用于进行图像融合。图5.7展示了基于深度学习的可见光与红外图像融合方法发展时间轴及关键里程碑。

5.6 基于深度学习的可见光与红外图像融合方法分类

基于深度学习的可见光与红外图像融合方法可以根据不同的方式进行分类。根据训练过程中是否使用了标签，可以分为监督方法和无监督方法。注意，这里的监督方法是指在训练过程中需要某种形式的标注（不一定是融合图像的真实标签）。根据采用的模型类型，可以分为基于卷积神经网络的方法、基于自编码器的方法、基于生成式对抗网络的方法、基于变换器的方法等。根据方法是否是端到端的，可以分为端到端方法和非端到端方法。在端到端方法中，融合图像是直接从源图像生成的，而不使用手工设计的步骤（例如，图像分解、独立训练、基于权重图的加权求和）。相反，非端到端方法至少需要一个手工设计的步骤。此外，可见光与红外图像融合方法可以分为全卷积方法和非全卷积方法。

5.7 基于深度学习的可见光与红外图像融合方法介绍

本节介绍一些具有代表性的基于深度学习的可见光与红外图像融合方法。

5.7.1 单分支模型和双分支模型

在图像融合中，源图像（以两幅源图像的情况为例）的处理一般有两种架构，即单分支和双分支架构，如图5.8所示。在单分支架构中，可见光和红外图像在通道方向上连接在一起，形成一个4通道输入，然后输入到深度学习模型中。相反，在双分支架构中，可见光和红外图像分别通过两个支路进行处理。在某些情况下，这两个支路是相同的（共享权重）。然而，由于可见光和红外图像是不同的模态，在许多情况下，采用两个不同的支路（不共享权重）来分别处理它们会更加合适一些。

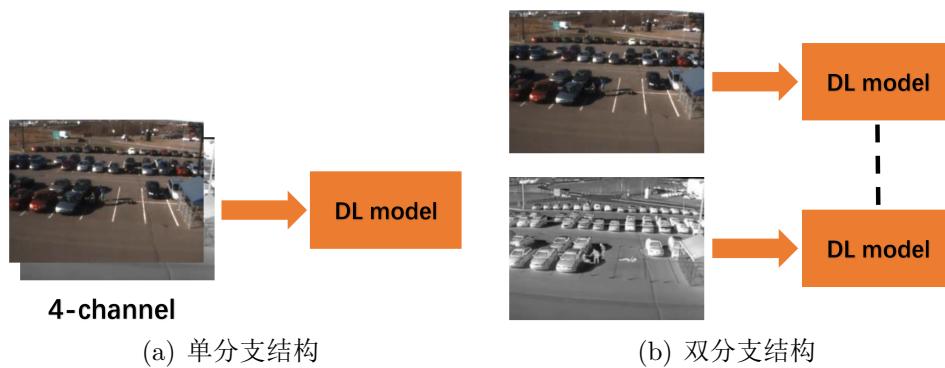


图 5.8: 基于深度学习的图像融合里的单分支模型和双分支模型。在某些双分支模型中，两个深度模型会共享权重，如图 (b) 中虚线所示。

5.7.2 基于卷积神经网络的图像融合方法

合肥工业大学刘羽老师是最早提出基于卷积神经网络的可见光与红外图像融合方法的研究人员之一。2018年，刘老师团队使用孪生卷积神经网络从源图像生成权重图 [22]。在该方法中，他们使用拉普拉斯金字塔对源图像进行分解，使用高斯金字塔对权重图进行分解。然后，他们在多尺度下进行融合。该方法使用高质量图像及其通过多尺度高斯滤波和随机采样生成的模糊版本进行训练。在刘羽老师发表该方法以来，研究人员提出了许多基于卷积神经网络的可见光与红外图像融合方法。

基于卷积神经网络的可见光与红外图像融合方法的主要架构如图5.9所示。无监督和监督方法之间的主要区别在于是否在训练过程中使用了标签。这常常体现在损失函数的构建方式上。值得说明的是，在一些研究中，卷积神经网络仅应用于图像融合的部分阶段。在这种情况下，图5.9所示的基于卷积神经网络的模型可能包含其他手动步骤，例如手动融合规则。

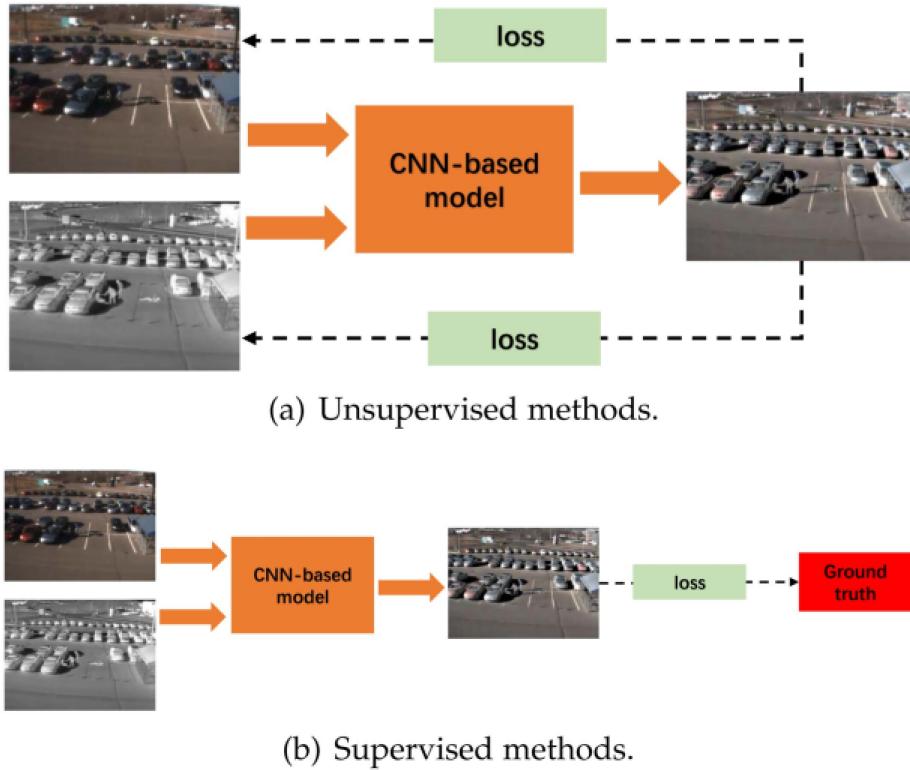


图 5.9: 基于卷积神经网络的可见光与红外图像融合方法示意图。

1. 无监督方法

基于卷积神经网络的无监督融合方法的整体架构如图5.9(a) 所示。由于缺乏标签，损失函数通常使用融合图像和源图像来进行计算。一般来说，损失函数包含基于图像融合评价指标构建的项。

在许多无监督方法中，卷积神经网络仅应用于可见光与红外图像融合过程的一部分。例如，Liu 等人 [79] 将源图像分解为基底部分和细节部分。然后，他们使用加权求和方法融合基底部分，并使用 CNN 和多层特征融合策略融合细节部分。Hou 等人 [80] 提出了 VIF-Net，在特征提取和图像重建中使用卷积神经网络。与此相对，在一些无监督方法中，卷积神经网络则应用于整个可见光与红外图像融合过程。例如，Xu 等人 [81] 和 Mustafa 等人 [82] 在可见光与红外图像融合的所有三个阶段都应用了卷积神经网络。

目前已经有许多基于卷积神经网络的无监督图像融合方法被提出来。在图5.9(a) 所示的基础架构上，研究员们使用了许多不同的措施来提升融合性能。具体措施总结如下：

- 残差连接：江南大学李辉等人 [58] 在 2019 年将残差连接引入可见光与红外图像融合领域。之后，许多可见光与红外图像融合方法利用了残差连接

并表现出良好的性能。例如，Li 等人 [83] 在编码器和解码器中均使用了残差连接。

- **密集连接：**密集连接 [31] 在许多应用中都表现出了良好的性能。Li 等人 [24] 在 2019 年将其引入可见光与红外图像融合领域。他们在编码器中使用了密集连接来提高模型的表示能力。从那时起，密集连接被广泛应用于可见光与红外图像融合方法中以提高性能。其中，大多数方法仅在特征提取阶段应用密集连接，而一些方法 [84] 在特征提取和图像重建阶段都应用了密集连接。
- **注意力机制：**引入注意力机制是提高可见光与红外图像融合性能的最常用技术之一。注意力机制可以应用于可见光与红外图像融合方法的不同阶段。在一些方法中，注意力机制被用于细化特征。例如，Zhu 等人 [85] 利用多通道注意力机制来改进特征提取，Li 等人 [86] 使用预训练的 CNN 从源图像提取特征，然后采用注意力机制以实现更高效的特征提取，Mustafa 等人 [82] 利用自注意机制来细化可见光和红外特征。在其他一些方法中，注意力机制被用作融合策略 [87]。此外，在另外一些方法中，注意力机制也用于特征细化和特征融合。例如，Wang 等人 [84] 利用 L_p 归一化的注意力策略来细化和结合深度特征。
- **多尺度特征：**多尺度特征在基于卷积神经网络的可见光与红外图像融合方法中也被频繁使用。实现多尺度特征的一种方法是使用不同大小的卷积核，因为较小的卷积核可以有效提取低频信息，而较大的卷积核可以捕捉大特征 [85]。例如，Wang 等人 [84] 和 Liu 等人 [88] 使用两种不同大小的卷积核，即 1×1 和 3×3 卷积核，来提取多尺度特征。
- **多层次特征：**融合多层次特征也是提高融合性能的一种常用技术。获取多层次特征的一种方法是使用图像分解方法。例如，Yan 等人 [89] 首先将源图像分解为基底层和细节层。基底层使用视觉显著性权重图方法进行融合，而细节层首先通过多分辨率奇异值分解生成多层次特征，然后通过预训练的 VGG19 进行融合。另一种实现方法是使用 CNN 的不同层的特征，首次由 Li 等人 [56] 提出。
- **对比学习：**在可见光与红外图像融合中，通常在融合图像中保留红外图像中的显著目标和可见光图像中的背景细节。因此，融合图像通常在显著目

标上与红外图像相似，在背景细节上与可见光图像相似。基于这一思想，可以将对比学习应用于可见光与红外图像融合。例如，Zhu 等人 [?] 通过设计对比学习框架和对比损失，将对比学习应用于可见光与红外图像融合。此外，Luo 等人 [90] 提出了一种对比差异损失，用于进行特征解耦，从而最大化源图像共同特征和私有特征之间的区别。这种方法可以应用于多种图像融合任务。

- **神经架构搜索：**大多数研究人员专注于设计不同的架构来进行可见光与红外图像融合，这在很大程度上依赖于研究人员的经验，可能需要大量时间。因此，有必要自动学习架构。大连理工大学樊鑫教授团队在这方面做了一些有益的探索。2021 年，他们提出了 SMoA[91]。该方法利用神经架构搜索（NAS）来发现可见光和红外图像的模态导向特征表示。这种方法可以缓解手动设计架构的问题。此外，他们在另一种可见光与红外图像融合方法中提出了一种架构搜索方案 [92]。具体来说，他们首先构建了一个分层聚合融合架构，然后建立了一个可以搜索潜在架构的搜索空间。此外，他们提出了一种基于 NAS 的轻量级架构 [?]。笔者认为，这是一个有前途的方向，因为它可以节省人工设计架构的工作，并有可能找到更好的架构。
- **特征分解：**除了图像分解外，一些方法执行特征分解以提高融合性能。例如，Xu 等人 [93] 提出了为可见光和红外图像学习解耦表示的方法，具体来说，为可见光和红外图像学习场景相关和属性相关的表示。然后，分别对场景相关和属性相关特征应用不同的融合策略。最终的融合图像通过基于残差块和反卷积层的 CNN 获得。与 [93] 不同，Xu 等人 [81] 进行特征图分解以获得共同部分和独特部分。然后，他们对共同部分和独特部分应用不同的融合规则。
- **考虑光照的模块：**光照条件显著影响可见光和红外图像的可靠性。因此，在图像融合过程中考虑光照条件非常重要。马佳义教授团队 [94] 提出了一个考虑光照的模块来评估光照条件，然后通过光照感知损失引导图像融合过程。
- **其他类型的卷积：**一些研究人员提出在可见光与红外图像融合方法中使用其他类型的卷积，而不是常规卷积，因为这些卷积具有特殊的性质。例如，Mustafa 等人 [82] 利用膨胀卷积来增加感受野。

上述措施是基于卷积神经网络的可见光与红外图像融合方法中的重要构建模块。许多可见光与红外图像融合方法结合了其中的一些措施来提高性能。例如，Li 等人 [86] 提出了一种基于 DenseNet 和注意力机制的可见光与红外图像融合方法，Mustafa 等人 [82] 利用多尺度特征和残差连接提高性能，Long 等人 [95] 提出了一种基于密集残差网络的可见光与红外图像融合方法，将 ResNet 和 DenseNet 结合起来，Zou 等人 [96] 在他们的方法中采用了注意力机制和多尺度特征，Shen 等人 [97] 在他们的方法中使用了注意力机制、残差连接和 DenseNet。

2. 有监督方法

虽然大多数基于卷积神经网络的方法是无监督的，但研究人员也提出了一些有监督方法，其整体架构如图 1.9(b) 所示。这些监督方法中使用了几种类型的“标签”。

第一种类型的标签是由其他方法生成的融合图像。例如，An 等人 [98] 使用 Zhang 等人 [54] 提出的方法来生成融合图像作为训练标签。

第二种类型的标签是在其他图像融合任务使用的合成标签。例如，Liu 等人 [22] 使用高质量图像及其通过多尺度高斯滤波和随机采样生成的模糊版本训练他们的模型。此外，Feng 等人 [99] 提出了一种基于全卷积网络的可见光与红外图像融合方法，其训练数据同样是清晰的 RGB 图像及其模糊版本。此外，Wang 等人 [100] 使用清晰的 RGB 图像及其模糊版本微调预训练的 VGG 和 ResNet 模型。Zhu 等人 [101] 使用可见光和红外图像生成合成训练数据。

第三种类型的标签是目标掩码。在可见光与红外图像融合中，突出红外图像中更可见的目标非常重要。因此，一些方法尝试通过目标掩码提高融合性能。例如，马佳义教授团队 [102] 将可见光与红外图像融合过程定义为可见光图像中的纹理信息和红外图像中的显著目标的融合。因此，他们在源图像中标注目标掩码，并使用这些掩码帮助网络进行显著目标检测和信息融合。

第四种类型的标签是最近使用的下游应用的地面真相。例如，马佳义教授团队 [40] 在训练中使用了场景分割的标签。

总之，监督方法可以具有与无监督方法相似的网络架构。然而，由于在可见光与红外图像融合任务中没有真实的标签，不同类型的“标签”被用来弥补这一点并计算损失函数。此外，在前文中介绍的用于提高无监督可见光与红外图像融合方法性能的措施也经常被用于有监督方法中。

3. 基于迁移学习的方法

有一些基于卷积神经网络的方法是基于迁移学习的。这些方法使用了在大规模数据集（如 ImageNet）上预训练的现成模型来提取特征。例如，Li 等人 [58] 和 Zhang 等人 [103] 使用 ResNet50 从源图像的高频部分提取特征，Li 等人 [56] 和 Ren 等人 [104] 使用 VGG19 提取特征，Yang 等人 [105] 采用了 VGG16，Li 等人 [86] 则使用在 ImageNet 上预训练的 DenseNet-201 提取深度特征。基于迁移学习的方法通常使用手动设计的融合规则来对提取的特征进行融合。值得说明的是，在一些方法中，研究人员会在进行融合之前对这些特征进行一些处理。例如，江南大学李辉等人 [58] 先应用零相位成分分析和 l_1 范数对提取的特征进行归一化，然后再进行融合。

5.7.3 基于自编码器的图像融合方法

基于自编码器的图像融合方法主要分为两步。第一步是对一个自编码器进行预训练，如图5.10的 (a) 图所示。在第一步中，会通过训练自编码器来重建输入图像的方式对模型进行训练。通过这个步骤，我们可以得到一个训练好的编码器和一个训练好的解码器。第二步，如图 7(b) 所示，使用训练好的编码器进行特征提取，并使用训练好的解码器进行图像重建。编码器和解码器之间的融合通常根据手动的融合规则进行，或者通过使用可见光-红外图像对进行第二次训练来学习。

基于自编码器的可见光与红外图像融合方法的一个早期且典型的例子是江南大学吴小俊教授团队提出的 DenseFuse 方法。该方法使用 MS-COCO 来预训练自编码器，并使用不同的融合策略（加法和 l_1 范数）来进行特征融合。

之后，有许多基于自编码器的方法被提出来。例如，Jian 等人 [106] 提出了 SEDRFuse 方法，首先使用可见光和红外图像训练自动编码器，以获得能够提取特征的编码器和能够重建融合图像的解码器。他们设计了一种基于注意力的特征融合策略来融合中间特征，并使用选择最大值策略来融合补偿特征。类似地，Li 等人 [107] 提出了一种基于空间注意力和通道注意力的融合策略，用于融合多尺度特征。

然而，上述 AE 基础方法的特征融合步骤是根据手动融合规则进行的，可能不太有效。为了解决这个问题，Li 等人 [108] 提出了 RFN-Nest，使用通过可见光和红外图像对训练的残差融合网络来进行特征融合。此外，他们在编码器中

利用了多尺度特征，在解码器中利用了嵌套连接以提高性能。类似地，Xu 等人 [109] 首先使用可见光和红外图像训练一个自动编码器，然后训练一个分类器以获得特征的分类显著图，并使用该显著图以像素级权重方式融合特征。最终使用解码器基于融合特征重建融合图像。

需要说明的是，许多 AE 基础方法仅使用 RGB 图像来训练自动编码器，并直接将训练好的自动编码器应用于红外图像。因此，由于 RGB 和红外图像之间的差异，性能可能有限。为了解决这个问题，一些方法轮流使用 RGB 和红外图像来训练自动编码器，而一些方法使用 RGB-红外图像对来训练编码器。此外，一些方法训练一个可见光自动编码器和一个红外自动编码器。另一种方法是使用可见光-红外图像对来训练融合模块。

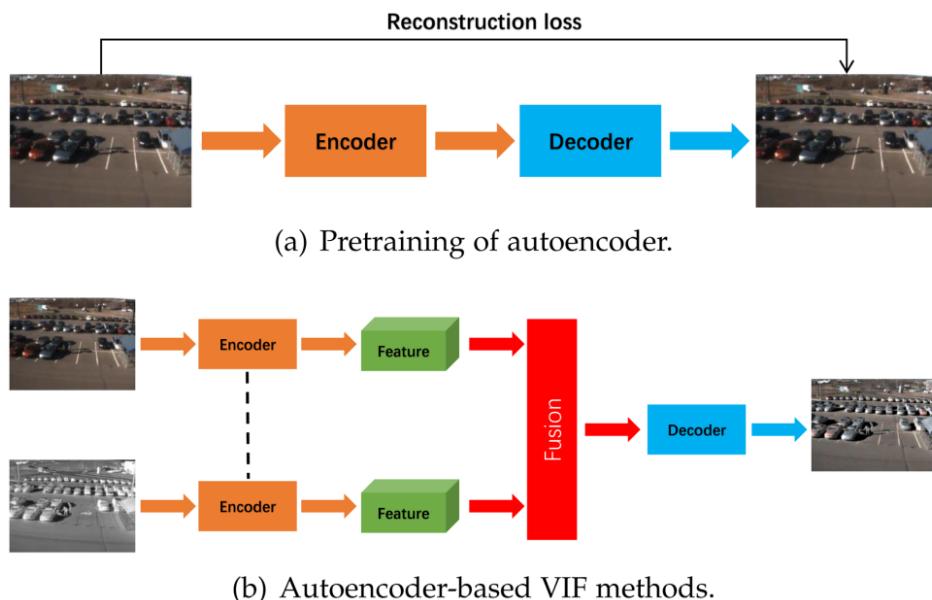
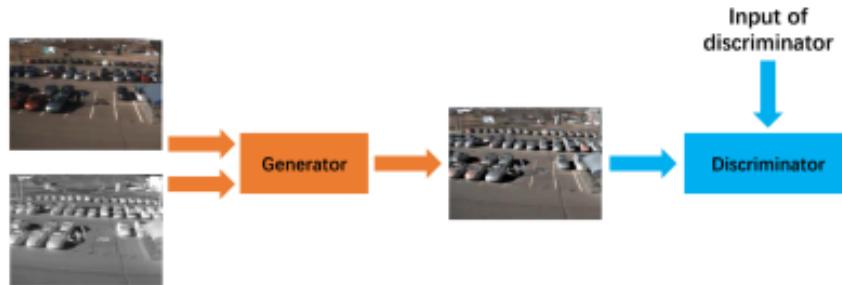


图 5.10: 基于自编码器的可见光与红外图像融合方法示意图。

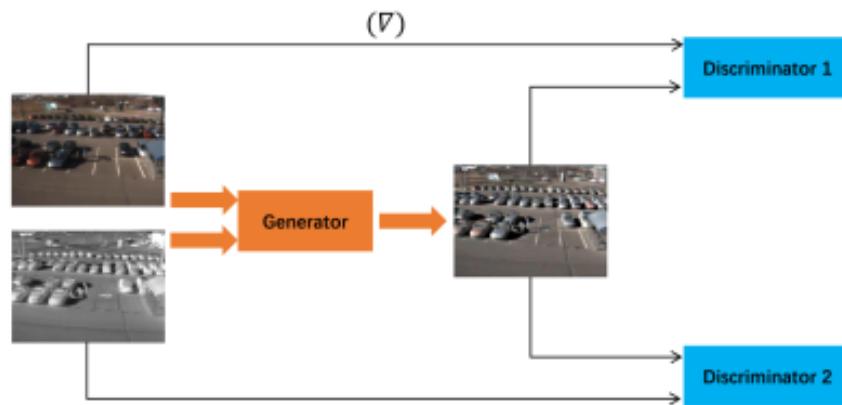
5.7.4 基于生成式对抗网络的图像融合方法

2019 年，武汉大学马佳义教授团队提出了 FusionGAN 方法 [25]，首次将生成式对抗网络应用于可见光和红外图像融合任务。之后，有许多基于生成式对抗网络的图像融合方法被提出，使得基于生成式对抗网络的方法成为了图像融合领域的主要方法之一。

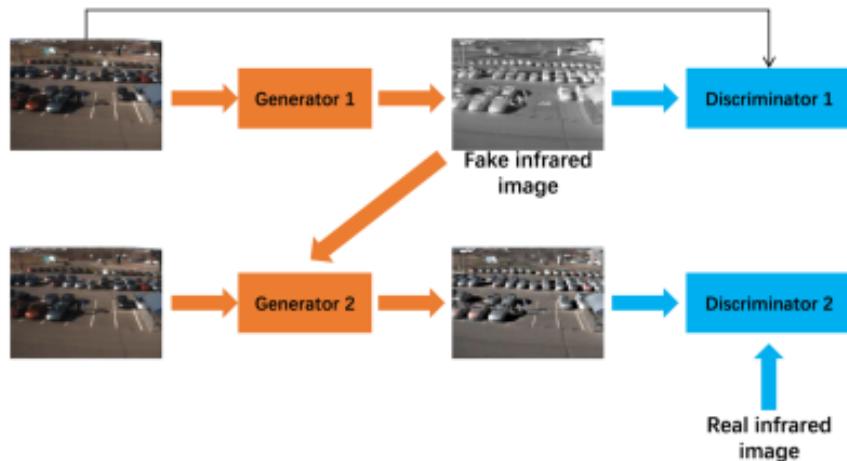
基于生成式对抗网络的可见光和红外图像融合方法的主要思路是使用生成器来生成融合图像，然后使用判别器来判断生成的融合图像是否包含了源图像的重要信息。按照训练过程是否需要标签，基于生成式对抗网络的图像融合方法可分为无监督方法和有监督方法。



(a) One generator and one discriminator [6], [33], [66], [69], [70]. The generator is used to generate fused images, and the discriminator is used to make the fused images similar to either the visible or the infrared image.



(b) One generator and two discriminators [50], [51], [67], [68], [72], [99]. The two discriminators are used to make the fused image contain features from both visible and infrared images.



(c) Two generators and two discriminators [71]. Generator 1 is used to generate a fake infrared image, and Generator 2 is used to produce fused image based on the visible image and the fake infrared image.

图 5.11: 无监督的基于生成式对抗网络的可见光与红外图像融合方法示意图。

1. 无监督方法

大多数基于生成式对抗网络的可见光与红外图像融合方法是无监督方法。这些方法的训练通常由一个比较融合图像与源图像之间的差异的损失函数驱动。此损失函数通常包含反映不同角度差异的多个项。如图5.11所示，不同的基于生成式对抗网络的图像融合方法可能包含不同数量的生成器和判别器。本节将基于生成器和判别器的数量讨论来简要这些方法。

1) 一个生成器和一个判别器

FusionGAN 是第一个基于生成式对抗网络的可见光与红外图像融合方法。FusionGAN 含有一个生成器和一个判别器。其中，生成器用于生成具有突出目标的融合图像，判别器用于强制融合图像包含可见光图像中的纹理细节。

FusionGAN 引起了图像融合研究人员的注意。许多科研人员开始研究改进 FusionGAN 的方法。例如，马佳义教授团队在随后的一篇论文中 [110] 通过引入细节损失和目标边缘增强损失，使得融合图像具有更多的纹理细节。

其他科研人员也采取了一些措施来提升基于生成式对抗网络的可见光和红外图像融合方法。例如，Xu 等人 [111] 在生成器中使用了残差块和跳跃连接。Fu 等人 [112] 提出利用密集块让生成器学习更多信息。他们不仅将浅层和深层特征连接起来，还在生成器的每一层插入可见光图像以帮助网络学习可见光信息。此外，注意力机制 [113] 也得到了应用。

上述这些基于生成式对抗网络的图像融合方法只采用一个判别器来强制生成的融合图像类似于可见光图像 [25, 114, 110, 112, 111, 115] 或红外图像 [116]。然而，无论哪种方式，随着对抗训练的进行，融合图像都会丢失源图像的一些细节。为了解决这个问题，一些研究人员提出使用多个判别器。

2) 一个生成器和多个判别器

为了处理判别器仅考虑单一源图像的问题，一些研究人员将基于生成式对抗网络的方法扩展为使用两个或更多的判别器。使用更多判别器的主要优点是可以保留两个源图像中的特征。例如，马佳义教授团队 [117] 提出了一种具有两个判别器的基于 GAN 的可见光与红外图像融合方法，可用于保留两个源图像中的特征。随后，马佳义教授团队 [118] 对该方法进行了扩展，主要改进包括：首先，使用密集连接的 CNN 替换生成器中的 U-Net；其次，判别器以图像本身而不是图像的梯度作为输入；最后，使用反卷积层替换两个上采样层对红外图像进行上采样，以生成生成器的输入。

其他研究人员也注意到采用两个判别器的益处。例如，Li 等人 [119, 120, 121,

[122] 设计了一系列使用一个生成器和两个判别器的可见光和红外图像融合方法。此外，马佳义教授团队 [123] 提出使用一个基于全尺度跳跃连接的生成器和两个 Markovian 判别器来保留可见光和红外源图像中的有用信息。

除了使用两个判别器的方法外，Song 等人 [124] 最近在他们的可见光与红外图像融合方法中使用了一个生成器和三个判别器。除了可见光和红外判别器外，他们还设计了一个差异图像判别器来处理可见光和红外图像之间的差异，也取得了良好的效果。

3) 两个生成器和两个判别器

Zhao 等人 [125] 提出了一种使用两个生成器和两个判别器的可见光与红外图像融合方法。他们首先使用第一个生成器从可见光图像生成一个伪红外图像。然后，他们使用第二个生成器来融合可见光图像和伪红外图像以获得融合图像。第一个判别器用于比较融合图像和可见光图像，第二个判别器用于比较融合图像、真实红外图像和伪红外图像。该方法是第一个在测试阶段仅使用可见光图像作为输入的可见光与红外图像融合方法。

2. 有监督方法

有少数基于生成式对抗网络的图像融合方法是有监督方法。这些方法大致使用了三种不同类型的标签。

第一种类型使用其他方法生成的融合图像作为标签。例如，Lebedev 等人 [126] 使用拉普拉斯金字塔算法和多尺度 Retinex[127] 生成的融合图像作为标签。随后，Li 等人 [128] 提出了基于耦合 GAN[129] 的 RCGAN 方法，并采用由 GFF[46] 生成的预融合图像来优化耦合生成器。然而，RCGAN 的性能会受到所选择的生成预融合图像的方法的影响。

第二种类型使用目标掩码作为监督信号。例如，Gu 等人 [130] 使用标记的掩码用于帮助融合图像包含来自红外图像的显著热目标。然而，该方法中判别器的输入只包含显著热目标，因此融合图像可能会丢失可见光图像的纹理细节。最近，Zhou 等人 [131] 提出了一种使用一个生成器和两个判别器的语义监督可见光与红外图像融合方法。该工作的一个创新之处在于设计了一个信息量判别器块来生成融合权重，然后用于保留可见光和红外图像中的语义信息。此外，这两个判别器有助于使融合图像包含可见光图像的纹理细节和红外图像的热辐射。该模型的训练需要用到标记好的语义掩码。

第三种类型使用 RGB-D 数据集中的 RGB 图像在 YCbCr 空间中的 Y

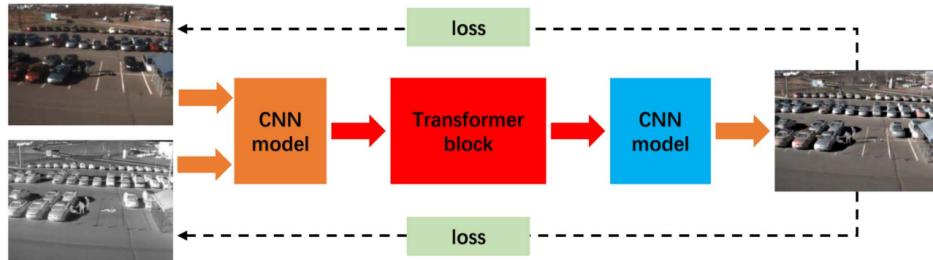


图 5.12: 基于变压器的可见光与红外图像融合方法的示例架构。该架构使用卷积神经网络来提取特征，这些特征通过变压器模块进行融合。融合图像通过另一个卷积神经网络模型获得。

通道作为标签 [132]。这种方法基于该标签来生成合成红外和可见光图像。然后，使用合成数据集进行训练。

5.7.5 基于 Transformer 的图像融合方法

变换器能够处理长距离依赖关系，并已被应用于各种自然语言处理和视觉任务中。在 2021 年，变换器被引入图像融合领域，并被应用于可见光与红外图像融合、其他图像融合任务和通用图像融合。

基于变换器的图像融合方法有不同的思路。在一些方法中，变换器仅用于特征融合。例如，武汉大学马佳义教授团队提出了 SwinFusion [59]，这是一种基于 Swin 变换器的通用图像融合方法。该方法使用卷积神经网络来做特征提取和图像重建。特征融合部分则是通过变换器来实现的。

在其他一些方法中，变换器也被应用于可见光与红外图像融合方法的其他阶段。例如，王等人在一个基于自编码器的框架中设计了 SwinFuse [133]，并使用基于变换器的编码器来提取全局特征。此外，合肥工业大学刘羽教授团队 [61] 提出的 YDTR 方法将卷积神经网络和变换器结合在编码分支和解码分支中。该团队在另外一个方法中 [134] 还设计了一个基于变换器的全局特征提取分支，与他们的基于卷积神经网络的局部特征提取分支平行。

总之，基于变换器的方法的架构因方法而异。值得一提的是，现有的基于变换器的可见光和红外图像融合方法很少是完全基于变换器的。如图5.12所示，在图像融合过程中，现有方法通常将卷积神经网络与变换器结合使用。此外，目前所有的基于变换器的可见光与红外图像融合方法都是无监督的方法。在这种情况下，损失函数是通过融合图像和源图像来进行计算的。

5.7.6 基于扩散模型的图像融合方法

近一两年来，扩散模型是一种非常流行的生成式模型。扩散模型通过前向过程和反向过程进行数据建模。在前向过程中，输入数据在多个时间步中逐渐添加高斯噪声，直到数据接近纯噪声。在反向过程中，模型通过多个逆向时间步逐步去除添加的噪声，恢复原始输入数据。

扩散模型作为一种强大的生成式模型，在图像生成、图像修复、图像超分辨率、图像到图像翻译等领域取得了突破性进展，并被广泛应用于许多任务中。

2023 年，扩散模型也被引入到了可见光和红外图像融合领域，并展现了卓越的性能。一个典型的例子 [27] 如图5.13所示。该方法将可见光图像和红外图像的多通道数据作为输入，并使用扩散模型来构建多通道数据的联合分布，并提取多通道互补信息，从而实现高保真的融合结果。

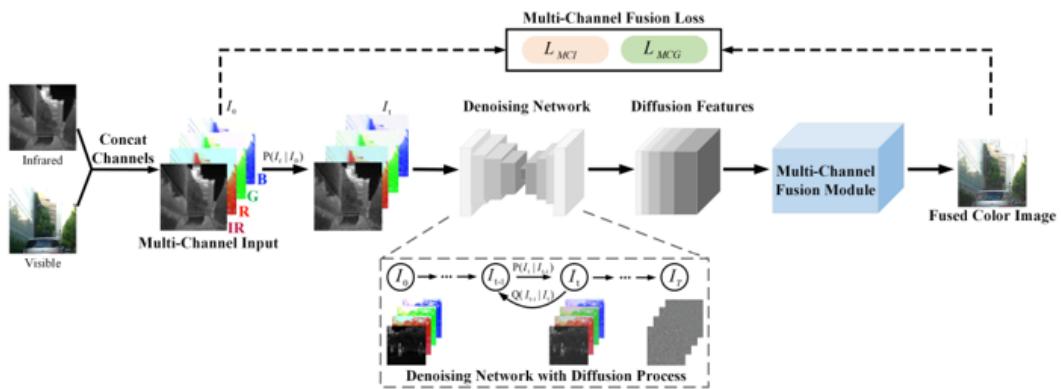


图 5.13: 一种基于扩散模型的可见光与红外图像融合方法。图片来源于 [27]。

具体来说，该方法的实现过程包括以下几个步骤：

- 构建多通道输入。该方法采用了本章5.7.1部分所说的单分支结构，将红外图像和可见光图像在通道维度上拼接形成多通道输入。
- 多通道扩散特征提取：采用去噪扩散概率模型（Denoising Diffusion Probabilistic Model, DDPM）构建多通道数据分布。在正向扩散过程中，高斯噪声被逐步添加到多通道输入数据上。经过多个时间步后，数据接近于纯噪声。在反向扩散过程中，模型通过去噪网络逐步去除噪声，重构原始输入数据。通过训练去噪网络，学习可见光和红外图像的联合潜在结构。
- 多通道融合模块：从去噪网络中提取多通道扩散特征，这些特征包含可见光和红外图像的信息。将这些多通道扩散特征输入到多通道融合模块中，直接生成三通道的融合图像。

- 损失函数设计：为了保留丰富的纹理信息，采用多通道梯度损失来保持梯度的真实性。为了使融合图像的强度分布与红外图像和可见光图像相似，采用多通道强度损失来保持强度的真实性。最终的损失函数结合了多通道梯度损失和多通道强度损失，用于指导融合网络的训练。

实验结果表明，基于扩散模型的方法在可见光和红外图像融合任务中表现出色，特别是在色彩保真度和多源信息聚合能力方面，显著优于现有的其他方法。

5.8 可见光与红外图像融合的发展特点

1. 越来越多的深度学习模型被应用于可见光与红外图像融合

当深度学习被引入可见光与红外图像融合领域时，只有深度玻尔兹曼机 [77] 和卷积神经网络 [22, 56] 被应用于可见光与红外图像融合。在稍后的时间（2019年），生成式对抗网络 [25] 和自编码器 [24] 被引入可见光与红外图像融合并成为非常重要的可见光与红外图像融合方法类型。2021年，变分自动编码器 [135] 和变换器 [136, 137, 138] 也被引入到该领域。2023年，扩散模型被引入图像融合领域。此外，一些重要的网络架构，如稠密网络（DenseNet）和残差网络（ResNet），也已被引入到可见光与红外图像融合并成为重要的构建模块。

2. 大多数方法是无监督方法

因为在可见光与红外图像融合中没有真实的标准融合图像，因此大多数基于深度学习的可见光与红外图像融合方法是无监督方法。这与多聚集图像融合（见第六章）和多曝光图像融合（见第七章）不同。在这两个任务中，可以分别使用完全清晰和曝光良好的图像作为合成数据集的标准融合图像。因此，在可见光与红外图像融合中，研究人员非常重视设计基于融合图像和源图像的各种损失函数。

3. 深度学习与传统图像处理技术的结合

深度学习和传统图像处理技术一起应用于可见光与红外图像融合的方法并不少见。例如，Raza 等人 [139] 使用了密集多尺度网络、四叉树分解以及贝塞尔插值来提取不同的特征。也有一些方法将生成式对抗网络与传统图像处理技术结合起来。例如，Wang 等人 [140] 首先将源图像分解为基底层和细节层，然后

使用拉普拉斯金字塔方法融合基底层，而细节层则使用生成式对抗网络进行融合。此外，Yang 等人 [141] 将生成式对抗网络和自适应引导滤波器相结合。具体来说，生成器用于生成一个组合纹理图，该图被用作自适应引导滤波器的引导图像。通过结合深度学习和传统图像处理技术，可以在可见光与红外图像融合方法中保留它们各自的优势。

4. 可见光与红外图像融合与其他任务的结合

以前，在几乎所有的可见光与红外图像融合研究中，可见光与红外图像融合是唯一的目标。近年来，一些研究人员开始研究将可见光与红外图像融合与其他任务一起执行。例如，Li 等人 [142] 提出了一种可以同时执行可见光与红外图像融合和图像超分辨率的可见光与红外图像融合方法。Xiao 等人 [143] 提出了一种用于同时执行可见光与红外图像融合和超分辨率的知识蒸馏方法。通过将可见光与红外图像融合与其他任务结合使用，模型可以更高效地使用，因为一个模型可以执行更多任务。

5. 图像融合与配准联合学习

由于可见光和红外图像的成像机制不同以及可见光和红外摄像机的参数不同，精确对齐可见光-红外图像对是很困难的。许多方法已经被提出用于执行可见光-红外图像配准 [144, 145, 146]。然而，这些方法几乎都不考虑图像融合任务。为了解决这个问题，一些研究人员开始将图像融合和配准联合学习 [147, 148]。

例如，大连理工大学樊鑫教授团队 [147] 首先使用跨模态感知风格迁移网络生成伪红外图像。然后，他们学习真实红外图像和伪红外图像之间的位移向量场，这是一个更容易的单模态配准问题。学习到的位移向量场随后被用来重建已配准的真实红外图像。最后，他们通过一个双路径交互融合网络，使用可见光图像和已配准的红外图像进行图像融合。一个包含风格迁移损失、交叉正则化损失、配准损失和图像融合损失的损失函数被设计用于指导模型训练。此外，武汉大学马佳义教授团队 [148] 提出了 RFNet。这是一种互相增强的框架，可以同时学习融合和配准。具体来说，RFNet 利用图像融合为图像配准提供反馈。然而，尽管这个想法很有启发性，RFNet 被设计用于可见光图像和近红外图像的配准，而不是可见光图像和热红外图像的配准。

6. 不同分辨率图像的可见光与红外图像融合方法

现有的大多数可见光与红外图像融合方法旨在融合相同分辨率的可见光和红外图像。然而，在实际中，由于传感器的成本原因，通常会出现高分辨率的可见光图像和低分辨率的红外图像。最近，一些研究人员开始开发融合不同分辨率图像的方法。例如，武汉大学马佳义教授团队 [117] 提出了一种方法，可以融合高分辨率的可见光图像和低分辨率的红外图像。这种方法首先通过两个上采样层对红外图像进行上采样，然后采用判别器来区分原始红外图像和下采样后的融合图像。该方法随后被进一步扩展 [118]。具体来说，他们使用了反卷积层来替代两个上采样层，从而实现从低分辨率特征到高分辨率特征的映射。这样一来，参数是通过训练而不是预定义获得的。然而，这两种方法对源图像的输入尺寸有要求，即可见光图像和红外图像分辨率之间的比例应为 4。

此外，刘羽老师团队 [142] 提出了一种基于元学习的可见光与红外图像融合方法，可以将不同分辨率的源图像融合成任意分辨率的融合图像。

7. 关于评价基准的研究

与计算机视觉中的许多任务不同，图像融合长期以来一直缺乏评价基准。直到 2020 年，笔者团队开始在图像融合领域创建基准测试 [149, 39]。关于可见光与红外图像融合，笔者提出了第一个可见光-红外图像融合基准（VIFB）[2]。该基准包括 21 对可见光-红外图像对的测试集、20 种可见光与红外图像融合方法的代码库和 13 个评估指标。VIFB 已被超过 10 多个国家的许多研究人员采用。

8. 面向应用的融合方法

大多数现有的可见光与红外图像融合方法在图像融合过程中不考虑下游应用，如图5.14(a)所示。因此，在图像融合过程中学习和融合的特征是通用特征，这可能导致视觉上令人满意的融合图像，但未必对下游应用最优。笔者称这种方法为面向视觉质量的可见光与红外图像融合方法。

在过去三年中，可见光和红外图像融合领域一个非常重要的特点是开发面向应用的可见光与红外图像融合方法。这与大多数现有的可见光与红外图像融合研究不同。后者在图像融合过程中不考虑下游应用。据笔者所知，Shopovska 等人 [150] 是最早考虑在可见光与红外图像融合中引入下游应用的研究之一。具体来说，他们在损失函数中采用了辅助行人检测误差，以帮助定义人类外观的相关

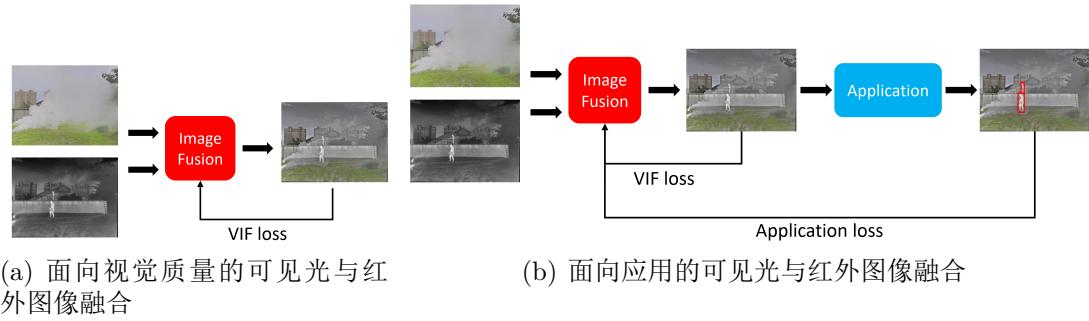


图 5.14: 面向视觉质量的可见光与红外图像融合与面向应用的可见光与红外图像融合。以行人检测为例。可见光和红外图像来自 M³FD 数据集 [75]。融合图像由作者使用 MGFF [3] 图像融合算法生成。

特征。这项工作的主要重点是增强融合图像中行人在视觉上的可见性。之后，马佳义教授团队提出了 SeAFusion[40]，在图像融合过程中考虑了场景分割。樊鑫教授团队 [75] 在双层优化公式中将图像融合和目标检测结合起来，并提出了联合训练方法。与面向视觉质量的可见光与红外图像融合方法相比，面向应用的可见光与红外图像融合方法通过在损失函数中加入基于应用的损失项，在融合过程中直接考虑了下游应用的性能，如图5.14(b)所示。因此，这些方法可以提供更适合特定应用的融合图像。

9. 在损失函数中使用不同项

基于深度学习的可见光与红外图像融合方法的一个共同特点是，通常在损失函数中包含不同的项。一个主要原因是这个任务没有真实的标签，因此融合图像的获得主要是靠损失函数的设计，从而融合质量很大程度上依赖于损失函数。在这种情况下，研究人员必须设计自己的损失函数来指导模型训练。

在设计损失函数时，一个直接的方法是根据图像融合的评价指标来设计损失函数。实际上，大多数基于深度学习的可见光与红外图像融合方法都包含根据图像融合评价指标设计的损失项。值得注意的是，大多数基于深度学习的可见光与红外图像融合方法的损失函数只考虑图像融合性能。因此，我们称这种损失函数为可见光与红外图像融合损失函数项，如图5.14(a)所示。可见光与红外图像融合损失函数项通常是通过计算融合图像和源图像或伪真实标签之间的差异得出的。

然而，正如笔者曾在论文中 [2] 指出的那样，一个可见光与红外图像融合方法在不同类型的图像融合评价指标（如基于结构的指标和基于信息论的指标）上的表现可能非常不同。因此，基于单一指标的可见光与红外图像融合损失函数不足以训练出一个好的可见光与红外图像融合方法。因此，研究人员开始在可见光

与红外图像融合损失函数中加入不同的项。通常，这些项对应于不同类型的评价指标。例如，陈勋教授团队 [151] 使用了强度损失项和结构相似性指数（SSIM）损失项，刘羽老师团队 [61] 使用了空间频率（SF）损失项和 SSIM 损失项，吴小俊教授团队 [152] 使用了 SSIM 损失项、基于梯度的损失项和均方误差（MSE）损失项。

此外，正如在本书5.8节中提到的，近年来已有研究人员开始在损失函数中加入基于应用的项，除了可见光与红外图像融合损失外。我们称这些项为应用损失函数项，如图5.14(b)所示。例如，Shopovska 等人 [150] 和樊鑫教授团队 [75] 在损失函数中加入了目标检测损失项，马佳义教授团队 [40] 在损失函数中加入了语义分割损失项。应用损失函数项通常是通过计算网络输出与应用的真实标签之间的差异得出的。而

总的来说，大多数现有的可见光与红外图像融合方法只使用可见光与红外图像融合损失函数项。然而，一个更有前景的方法是同时使用可见光与红外图像融合损失函数项和应用损失函数项。

10. 可以直接融合彩色图像的方法

大多数可见光与红外图像融合方法只能融合灰度图像。为了融合彩色图像，这些方法通常先将 RGB 图像转换到 YCbCr 颜色空间，然后将 Y 通道与红外图像进行融合 [92, 59, 61, 153]。之后，再进行逆颜色空间转换以获得彩色融合图像。然而，这个过程较为复杂。此外，大多数方法仅使用深度学习方法融合 Y 通道，而 Cr 和 Cb 通道则采用传统方法（例如手动方法）进行融合。这可能会导致信息丢失，因为 Cb 和 Cr 通道也包含重要信息。

最近，研究人员提出了一些可以直接融合可见光彩色图像和红外图像的可见光与红外图像融合方法 [154]，我们认为这是一个重要的发展特点，也是未来的一个有前途的研究方向。

11. 编程语言和深度学习框架

我们回顾了图5.5中展示的所有基于深度学习的可见光与红外图像融合方法，以检查所使用的编程框架。图5.15 显示了各年使用的编程框架。可以看到，使用 Tensorflow 的方法数量在 2021 年之前迅速增加，但在 2022 年开始减少。相反，使用 Pytorch 的方法数量迅速增加，这表明 Pytorch 已成为基于深度学习的可见

光与红外图像融合方法中最受欢迎的编程框架。使用 Matlab 的方法数量从 2018 年到 2022 年略有增加。

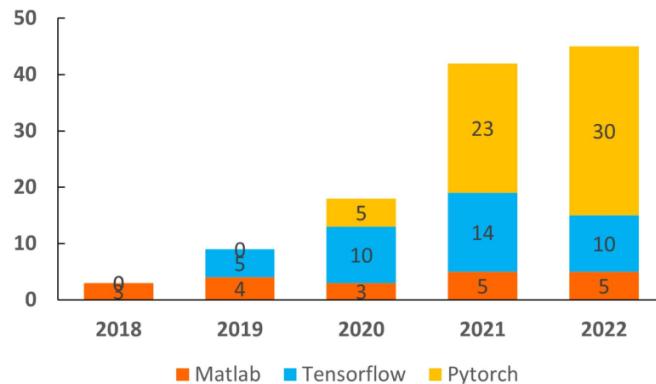


图 5.15: 基于深度学习的可见光与红外图像融合方法（包括可以应用于可见光与红外图像融合的通用图像融合方法）使用不同编程框架的论文数量。我们只统计提供开源代码或明确提及所用编程框架的论文。

12. 研究人员分布

笔者研究了自 2018 年到 2022 年 12 月期间发表的 212 篇基于深度学习的可见光红外图像融合的论文，并统计了作者的分布情况。如图5.16所示，超过 90% 的论文有中国机构参与。具体来说，77.7% 的论文仅由中国机构发表，13.3% 的论文由中国和其他国家的机构共同发表，只有 9.0% 的论文仅由其他国家的机构发表。这些数据表明，中国机构在基于深度学习的可见光与红外图像融合方法研究中非常活跃。在涉及其他国家机构的 47 篇论文中，不同国家研究人员的具体分布如图??所示。可以看出，英国是最活跃的国家（24%），其次是印度（12%）、韩国（12%）、美国（10%）和加拿大（10%）。

总之，从事基于深度学习的可见光与红外图像融合方法研究的大多数研究人员来自中国机构。其他国家的研究人员对这一研究方向关注较少。应该采取更多措施使这一研究方向更加流行。

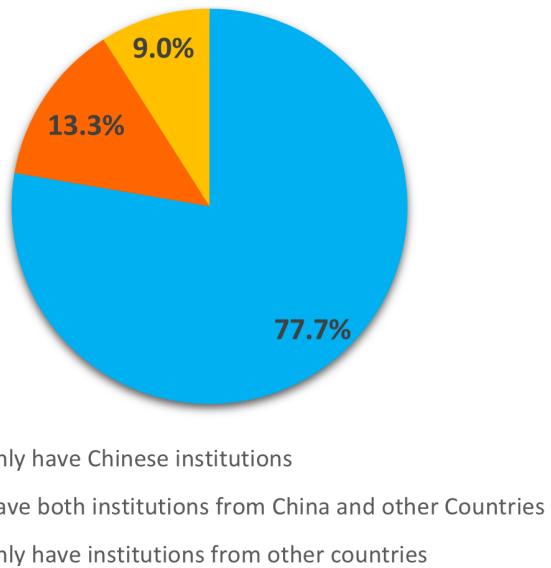


图 5.16: 中国和其他国家机构发表的论文数量。

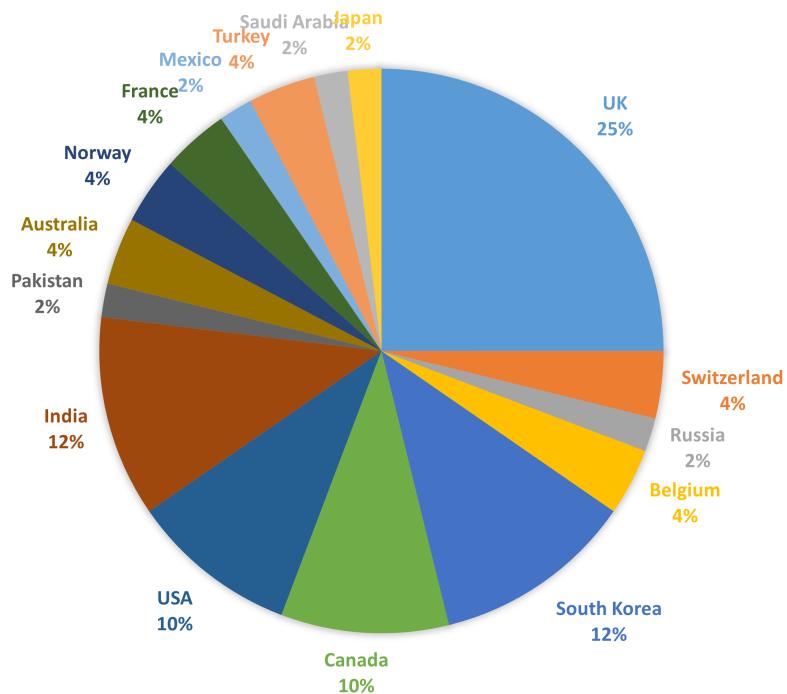


图 5.17: 除了中国之外的研究人员分布。

5.9 未来发展趋势

1. 更好的评价指标

可见光与红外图像融合方法在现有论文中通过定性比较（使用视觉性能）和定量比较（使用图像融合评价指标）进行评估。然而，如表5.1所示，不同论文中

表 5.1: 一些可见光与红外图像融合方法发表在顶级期刊和会议上。可以看出，由于使用了不同的测试集和评价指标，很难了解可见光与红外图像融合方法的真实性能比较。

Method (year)	Venue	No. of test image pairs	Objective evaluation metrics
FusionGAN [25] (19)	INFUS	7 (TNO) + 31 (INO)	5 (EN, SD, SSIM, SF, VIF)
VIF-Net [80] (20)	TIP	9 (TNO and INO)	5 (MI, $Q^{AB/F}$, PC, Q^{NCIE} , UIQI)
U2Fusion [156] (20)	TPAMI	20 (TNO) + 45 (RoadScene)	4 (SSIM, PSNR, CC, SCD)
GANMcC [116] (21)	TIM	16 (TNO) + 30 (RoadScene)	6 (SSIM, CC, SCD, EN, SD, MI)
Liu et al. [157] (21)	TIP	20 (TNO)	5 (VIF, AG, SF, SCD, $Q^{AB/F}$)
SDNet [158] (21)	IJCV	10 (TNO)	4 (EN, FMI _{dct} , PSNR, MG)
[159] (22)	PR	20 (TNO) + 20 (RoadScene)	6 (Q_{abf} , SCD, VIF, SF, SSIM, PSNR)
MHTNet [83] (22)	TIM	20 (TNO) + 20 (KAIST) + 20 (BEMP)	6 (SCD, VIFF, EN, SD, MI, Q_{CV})
DDcGAN [117] (19)	IJCAI	20 (TNO)	4 (EN, SD, SF, PSNR)
DIDFuse [160] (20)	IJCAI	40 (TNO) + 52 (NIR) + 40 (FLIR)	6 (EN, MI, SD, SF, VIF, AG)
FusionDN [161] (20)	AAAI	44 (RoadScene)	4 (SD, EN, VIF, SCD)
PMGI [162] (20)	AAAI	17 (TNO)	6 (SSIM, $Q^{AB/F}$, EN, FMI, SCD, CC)
[92] (21)	MM	37 (TNO) + 26 (RoadScene)	4 (SD, VIF, CC, SCD)

选择的指标各不相同，导致公平的性能比较变得困难。此外，一个指标通常只能从某些方面评估融合结果 [2]。此外，在图像融合领域，定性结果通常与定量结果不一致 [155, 2]。因此，需要更好的指标。理想的指标应该与视觉性能一致，并能够全面反映融合性能。

2. 更好的评价基准

尽管笔者开发了一个可见光与红外图像融合基准 [2]，但可见光与红外图像融合基准研究仍处于早期阶段。我们选择了五种最近的基于深度学习的方法 (IFCNN [62]、SeAFusion [40]、SwinFusion [59]、U2Fusion [156] 和 YDTR [61])，这些方法可以代表可见光与红外图像融合领域的最新发展，并在 VIFB [2] 上测试了它们的性能。在这些方法中，IFCNN 和 U2Fusion 是基于 CNN 的通用图像融合方法，SeAFusion 是应用驱动的可见光与红外图像融合方法，SwinFusion 是基于 Transformer 的通用图像融合方法，YDTR 是基于 Transformer 的可见光与红外图像融合方法。定量结果可以在表5.2中找到。由于页面限制，我们将定性结果放在补充材料中。我们从 VIFB 的结果中有以下几点观察。

首先，最新的基于深度学习的方法在 VIFB 上的表现并不比旧的基于深度学习的方法更好。例如，2022 年发表的三种方法，即 SeAFusion、SwinFusion 和 YDTR，在 VIFB 上的定量表现比 CNN 和 DLF 更差。其次，一些传统的可见光与红外图像融合方法，即 LatLRR、LP_SR 和 NSCT_SR，表现出与基于深度学习的方法非常竞争的定量性能，这表明基于深度学习的方法在 VIFB 上并没有表现出明显的优势。第三，从定性比较中很难得出哪种类型的方法更好。具体来说，最新的基于深度学习的方法在某些情况下表现出良好的融合性能，但在

表 5.2: 在 VIFB 上的定量性能比较。每个指标的最佳三个值分别用红色、绿色和蓝色表示。方法名称后面的三个数字分别表示最佳值、次佳值和第三佳值的数量。最好在彩色显示器上查看。

Method	Information theory-based				Information feature-based				Structural similarity-based		Human perception-inspired		
	CE (↓)	EN (↑)	MI (↑)	PSNR (↑)	AG (↑)	EI (↑)	$Q^{AB/F}$ (↑)	SD (↑)	SF (↑)	RMSE (↓)	SSIM (↑)	Q_{CB} (↑)	Q_{CV} (↓)
ADF (0,0,0)	1.464	6.788	1.921	58.405	4.582	46.529	0.519	35.185	14.132	0.104	1.400	0.474	777.817
CBF (0,0,2)	0.994	7.324	2.161	57.595	7.154	74.590	0.578	48.544	20.380	0.126	1.171	0.526	1575.148
FPDE (0,0,0)	1.366	6.766	1.924	58.402	4.538	46.022	0.484	34.931	13.468	0.104	1.387	0.460	780.114
GFCE (0,3,0)	1.931	7.266	1.844	55.939	7.498	77.465	0.471	51.563	22.463	0.173	1.134	0.535	898.946
GFF (0,0,0)	1.189	7.210	2.638	58.100	5.326	55.198	0.624	50.059	17.272	0.112	1.398	0.619	881.625
GTF (0,0,0)	1.285	6.508	1.991	57.861	4.303	43.664	0.439	35.130	14.743	0.118	1.371	0.414	2138.369
HMSD_GF (0,1,0)	1.164	7.274	2.472	57.940	6.246	65.034	0.623	57.617	19.904	0.116	1.394	0.604	532.958
Hybrid_MSD (0,1,0)	1.257	7.304	2.619	58.173	6.126	63.491	0.636	54.922	19.659	0.110	1.405	0.623	510.866
IEFVIP (0,0,0)	1.339	6.936	2.248	57.174	4.984	51.782	0.486	48.491	15.846	0.138	1.391	0.462	573.767
LatLRR (3,0,0)	1.684	6.909	1.653	56.180	8.962	92.813	0.438	57.133	29.537	0.169	1.184	0.497	697.286
LP_SR (2,2,2)	0.957	7.339	2.809	57.951	5.851	60.781	0.661	57.314	18.807	0.117	1.390	0.645	522.687
MGFF (0,0,0)	1.295	7.114	1.768	58.212	5.839	60.607	0.573	44.290	17.916	0.109	1.406	0.542	676.887
MSVD (0,0,2)	1.462	6.705	1.955	58.415	3.545	36.202	0.331	34.372	12.525	0.104	1.425	0.426	808.993
NSCT_SR (3,0,1)	0.900	7.396	2.988	57.435	6.492	67.956	0.646	52.475	19.389	0.131	1.277	0.617	1447.340
RP_SR (0,1,2)	0.994	7.353	2.336	57.777	6.364	65.220	0.566	55.808	21.171	0.122	1.332	0.606	888.848
TIF (0,0,0)	1.371	7.075	1.767	58.225	5.556	57.839	0.584	42.643	17.739	0.109	1.399	0.545	613.004
VSMWLS (0,0,0)	1.409	7.028	2.035	58.194	5.612	57.252	0.554	46.253	17.662	0.109	1.417	0.496	754.704
CNN (1,1,2)	1.030	7.320	2.653	57.932	5.808	60.241	0.658	60.075	18.813	0.118	1.391	0.621	512.569
DLF (3,0,0)	1.413	6.724	2.030	58.444	3.825	38.569	0.434	34.717	12.491	0.103	1.461	0.445	759.814
IFCNN (0,0,1)	1.419	7.122	2.068	58.246	6.228	64.645	0.589	48.521	19.359	0.108	1.403	0.531	495.289
ResNet (0,3,0)	1.364	6.734	1.988	58.441	3.674	37.255	0.407	34.940	11.736	0.104	1.460	0.445	724.831
SeAFusion (0,1,0)	1.543	6.967	2.120	57.301	5.655	58.877	0.561	49.628	17.733	0.134	1.393	0.460	416.935
SwinFusion (1,0,0)	1.338	6.938	2.282	57.321	5.605	57.992	0.575	52.855	18.045	0.135	1.406	0.489	399.224
U2Fusion (0,0,0)	1.316	7.200	1.946	57.966	6.241	65.831	0.532	50.058	18.288	0.114	1.331	0.540	719.791
YDTR (0,0,1)	1.568	6.828	2.124	58.015	4.333	44.591	0.452	44.980	15.082	0.116	1.435	0.436	679.953

其他情况下表现不佳。一些传统的可见光与红外图像融合方法也表现出非常竞争的定性性能。最后，定量性能与定性性能不一致，这是图像融合领域的一个常见问题。需要注意的是，这些观察结果是基于 VIFB [2] 得出的。如果使用不同的测试图像集和评价指标，可能会得出不同的观察结果。

VIFB 是开发可见光与红外图像融合基准的初步尝试。它有一些局限性，例如测试图像的数量少且分辨率低。需要更多的努力来开发更好的基准，以更好地比较可见光与红外图像融合方法。例如，未来的基准可能包含更多的测试图像、更丰富的场景以及更高质量评价指标的更好组合。

3. 更多基于变换器的方法

变换器在许多计算机视觉任务中取得了优异的性能。然而，变换器在可见光与红外图像融合中的应用仍处于非常早期的阶段。我们预计，在未来几年中，将会有许多基于变换器的可见光与红外图像融合方法。特别是，开发纯粹基于变换器的可见光与红外图像融合方法是一个有趣的研究方向。此外，阐明在可见光与红外图像融合背景下的全局信息是什么也至关重要，而这一点在现有的基于变换器的可见光与红外图像融合方法中很少被解释。

4. 面向应用的图像融合方法

使用可见光与红外图像融合的动机之一是提高下游应用的性能。然而，从我们的综述中可以看出，大多数现有的可见光与红外图像融合方法并没有考虑下

游应用。这可以从性能评估方法中看出，即通过视觉性能的定性比较和图像融合指标的定量比较来评估性能。然而，这种设计方式的可见光与红外图像融合方法学习的是通用特征和融合规则，可能并未针对下游应用进行优化。因此，在设计可见光与红外图像融合方法时，考虑下游应用会更好。

一种实现这一目标的可能框架如图5.14(b)所示，在该框架中同时使用可见光与红外图像融合损失和应用损失来指导训练。我们预计，面向应用的图像融合方法将成为这一领域的主流方法。

5. 更多应用

可见光与红外图像融合有潜力提高许多应用的性能，尤其是那些需要在各种光照条件下工作的应用。然而，可见光与红外图像融合主要应用于目标跟踪 [163, 164]、目标检测 [165, 75]、显著目标检测 [?, 102] 和场景分割 [166, 40]。许多其他应用，如人员救援 [167] 和机器人 [168]，具有很大的价值，但很少被研究。我们认为未来应该探索更多的可见光与红外图像融合应用。

6. 处理未对齐问题

可见光和红外图像的未对齐可能会降低应用的性能，例如行人检测 [6]。因此，处理未对齐问题对于图像融合非常重要，这也有助于促进可见光与红外图像融合方法的应用。然而，尽管许多研究已经致力于处理可见光和红外图像的未对齐问题，对齐仍然是一个未解决的问题，完美对齐可见光和红外图像非常具有挑战性。实际上，几乎所有现有的 RGB-红外数据集，如 LasHeR [169]、RGBT234 [8] 和 M³FD [75]，都有一些未对齐的问题。深度学习可能为这一问题提供潜在的解决方案，如最近的两项研究 [147, 148] 所探讨的那样，这些研究将图像融合和配准一起学习。

值得进一步研究的是，将图像融合、配准和下游应用结合起来，如 Tang 等人 [170] 所做的那样。我们期望未来会有更多基于深度学习的配准方法被提出，以解决未对齐问题。

7. 结合其他任务

在大多数 VIF 研究中，通常只考虑了可见光和红外图像融合。最近，一些研究人员开始将 VIF 与其他任务结合起来，这可能更有效且更高效。例如，Li

等人 [142] 和 Gu 等人 [171] 将 VIF 与图像超分辨率结合起来。然而，关于 VIF 与其他任务结合的研究仍然非常有限。我们期望在未来几年中，沿着这一方向的更多研究将被提出，以进一步探索 VIF 与其他任务的相互收益。

8. 提高融合效率

随着基于深度学习的可见光与红外图像融合方法的发展，研究人员设计了更大更深的模型来执行可见光与红外图像融合。然而，更大的模型使得可见光与红外图像融合方法的效率不足，阻碍了可见光与红外图像融合方法在实际应用中的价值，如目标跟踪和检测。一些研究人员 [172] 注意到了这一点，并尝试设计高效的深度可见光与红外图像融合方法。然而，这些研究仍然非常有限。设计高效的可见光与红外图像融合方法将是未来可见光与红外图像融合发展的一个重要趋势。

5.10 小结

本章介绍了图像融合中最重要的任务之一——可见光与红外图像融合。我们首先介绍了红外图像，指出红外图像给我们提供了一种从另外一个角度，即温度，来感知世界的方法。然后，我们介绍了可见光与红外图像融合这个任务，并简单介绍了传统可见光与红外图像融合方法。之后，我们重点介绍了近几年发展起来的基于深度学习的可见光与红外图像融合方法。

第 6 章

多聚焦图像融合

本章主要介绍基于深度学习的多聚焦图像融合技术。

6.1 多聚焦图像融合概述

在现实生活中，我们拍照时一般希望获得清晰的图像。在计算机视觉中，清晰的输入图像有利于各种算法取得好的效果。然而，在相机成像时，由于相机景深的限制，我们往往只能聚焦于场景中的某一部分，因此难以获得处处都聚焦的清晰图像。通常，我们获得的图像其中一部分是清晰的，而另一部分是不清晰的。图6.1中展示了一个这样的示例。从图中可以清楚看出，这幅图像中成像清晰的地方是不一致的。

为了获得处处清晰的图像，我们需要将这些聚焦点不同的图像进行处理。多聚焦图像融合，就是指通过融合两幅或多幅聚焦点不同的图像来生成清晰图像



(a) 图像前景不清晰而背景清晰 (b) 图像前景不清晰而背景清晰

图 6.1：聚焦点不同的图像示例。图像来源于 Lytro 数据集 [4]。



图 6.2: 多聚焦图像融合的例子。图像来源于 [10]。

的过程。图6.2中展示了一些多聚焦图像融合的例子。从图中可以看出，通过融合两幅聚焦点不同的源图像，我们获得了处处清晰的融合图像。

6.2 传统融合方法概述

多聚焦图像融合这个任务已经被研究了 30 多年。在深度学习被引入这个领域以前提出的方法，我们统称为传统融合方法。传统多聚焦图像融合方法主要分为基于空间域的方法和基于变换域的方法。具体来说，基于空间域的方法直接在空间域中操作，并且可以进一步分为三类：基于像素的方法、基于块的方法和基于区域的方法。相比之下，基于变换域的方法首先将图像转换到另一个域，然后使用转换后的系数进行融合。最终通过相应的逆变换得到融合后的图像。研究人员已经提出了许多基于变换域的方法，例如稀疏表示方法、多尺度方法、基于梯度域的方法和混合方法。

传统多聚焦图像融合方法存在一些缺点。在多聚焦图像融合中，焦点测量 (**focus measure, FM**) 或活动水平测量 (**activity measure, AM**) 以及融合规则是两个关键任务。传统多聚焦图像融合方法在处理这些任务时存在一些问题。首先，焦点测量和融合规则都是手动设计的。然而，真实图像的情况非常复杂，因此手动设计的焦点测量和融合规则可能会限制解决方案的搜索空间。其次，传统多聚焦图像融合方法将焦点测量和融合规则分开，这进一步限制了融合性能，因为人们几乎不可能通过组合所有必要因素来实现理想的设计。最后，基于变换域的方法利用手工设计的变换，无法充分表示源图像，并且对各种输入不

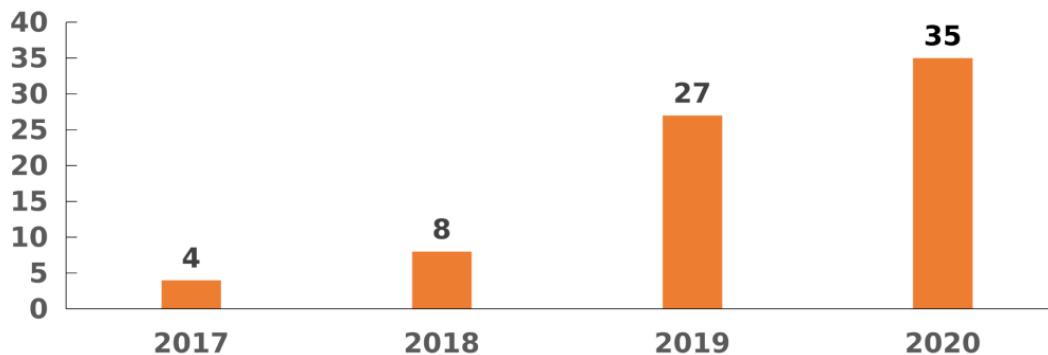


图 6.3: 2017 年到 2020 年间，基于深度学习的多聚焦图像融合论述数量。图像来源于 [10]。

够鲁棒。

深度学习可以帮助解决这些问题。具体来说，深度学习可以基于比手工设计特征更强大的深度特征来学习焦点测量。深度学习方法还可以在训练过程中自动学习融合规则。此外，它们可以通过学习同时处理活动水平测量和融合规则，这比传统算法更合理。最后，一些深度学习模型可以看作是自适应变换，比基于变换域的方法更鲁棒。因此，深度学习技术有潜力在多聚焦图像融合任务中提供更好的融合性能。然而，应该注意的是，并非所有基于深度学习的多聚焦图像融合方法都能通过深度学习解决所有这些问题。在一些多聚焦图像融合方法中，深度学习仅作为该方法的一部分使用。

6.3 基于深度学习的融合方法

基于深度学习的多聚焦图像融合方法在近年来得到了迅速的发展。图6.3显示了 2017 年到 2020 年间发表的基于深度学习的多聚焦图像融合论文的数量。从图中可以看出，每年发表的相关论文的数量在迅速增加，这表明基于深度学习的多聚焦图像融合得到了迅速的发展。

6.3.1 基于深度学习的多聚焦图像融合方法的分类

基于深度学习的多焦点图像融合方法可以通过不同的方式分类。按照是否需要标签进行训练，基于深度学习的多聚焦图像融合方法可以分为有监督方法和无监督方法。基于深度学习的多焦点图像融合方法还可以分为基于决策图的方法和端到端的方法。在基于决策图的方法中，首先生成一个指示焦点水平（或活动水平）的决策图。然后根据这个决策图进行图像融合。在这些方法中，深度

学习通常用于生成决策图，随后可能会有后处理步骤以获得更好的决策图。相比之下，端到端的多焦点图像融合算法通过将源图像输入到网络中直接生成融合图像。

6.3.2 基于有监督学习的融合方法

1. 基于卷积神经网络的方法

合肥工业大学刘羽老师 [173] 提出了第一个基于卷积神经网络的多聚焦图像融合方法。该方法利用卷积神经网络学习从源图像到焦点图的映射。通过这种方式，传统方法中分别处理的活动水平测量和融合规则可以联合学习。经过一系列后处理步骤，得到最终的决策图 (decision map)，然后通过像素级加权平均方法生成融合图像。从那时起，各种基于卷积神经网络的方法相继被提出。这些方法大致可以分为基于决策图的方法和端到端的方法。

基于决策图的方法是指首先生成一个决策图，然后对该决策图进行一些后处理步骤，再基于处理后的决策图对源图像进行加权融合的过程。而端到端的方法是指从源图像直接生成得到融合图像，因此中间并没有后处理的步骤。

1) 基于决策图的方法

在刘羽老师提出基于卷积神经网络的多聚焦图像融合方法以后，基于决策图的方法主要是在朝着下面一些方向在发展。

- 改进的焦点测量。有不少方法致力于改进多聚焦图像融合中的焦点测量。

例如，Tang 等人 [174] 提出了一种像素级卷积神经网络 (p-CNN)。该方法可以比传统的手工设计的焦点测量方法更好地识别源图像中的聚焦和散焦像素。Wang 等人 [175] 提出了一种类似的方法，利用孪生卷积神经网络模型来识别聚焦和散焦像素。在这些方法中，焦点测量在训练过程中自动学习，因此比手工设计的焦点更具鲁棒性。

- 多层次特征。多层次特征可以用于使特征表示更强大。例如，Yang 等人 [176] 提出了一种用于多聚焦图像融合的多层次特征卷积神经网络 (MLFCNN) 架构。其主要创新点在于使用了来自不同层的多层次特征，并采用 1×1 卷积来降低特征空间的维度。

- 多尺度特征。在一些研究中，多尺度特征被用来提高性能。例如，Du 等人 [177] 将刘羽老师的方法 [173] 扩展到多尺度框架 (MCNN)。MCNN 使用

多尺度输入进行训练以获得焦点图，然后通过形态学和分水岭操作处理以获得理想的决策图。Lai 等人 [178] 提出的多尺度视觉注意深度卷积神经网络（MADCNN）也使用来自不同空间尺度的互补特征。此外，Wang 等人 [179] 提出了一种基于离散小波变换域的 CNN 新颖多聚焦图像融合算法。具体来说，小波被应用于源图像以获得高频和低频图像。然后将高频和低频图像输入到相应的 CNN 中以获得相应的决策图，这些决策图随后被用于生成融合的高频和低频图像。最终的融合图像通过逆小波变换得到。通过这种方式，所提出的方法结合了基于 CNN 和变换域方法的优点。

- 注意力机制。在 Lai 等人 [178] 提出的 MADCNN 中，使用了一个视觉注意力单元来帮助网络更准确地定位聚焦区域。在 Guo 等人 [180] 提出的 SSAN 中，也使用了注意力机制，试图缓解卷积算子局部感受野的限制。
- 处理聚焦和散焦边界。聚焦和散焦边界（FDB）是一个重要区域，许多算法在 FDB 附近的表现不佳。为了解决这个问题，Ma 等人 [181] 提出使用两个不同的网络，即边界细化网和普通细化网，来处理从初始网络获得的初始评分图。边界细化网专门设计用于处理 FDB 附近的图像块，而普通细化网则应用于距离 FDB 较远的图像块。此外，在 MMFNet[182] 中，设计了一种 a-matte 边界散焦模型，通过精确模拟 FDB 附近的散焦扩展效应（DSE）来生成逼真的训练数据。据我们所知，这是第一个在多聚焦图像融合训练数据生成中考虑 DSE 的工作。该方法设计了两个级联子网，即初始融合子网和边界融合子网。具体来说，初始融合子网首先生成指导图，然后边界融合子网细化 FDB 附近的融合结果。
- 全卷积算法。在上述基于决策图的方法中，除了 MADCNN[178] 和 MMFNet[182]，网络中都使用了全连接层，因此测试图像的输入大小受到训练图像大小的限制。解决这个问题的一种方法是使用全卷积网络，这样对输入图像的大小没有要求，因此不需要将输入图像划分为小块。Guo 等人 [183] 提出了第一个基于全卷积决策图的方法。后来，一些其他全卷积方法被开发出来。例如，Li 等人 [184] 提出了 DRPL，该方法直接将整个图像转换为二值掩码（加权图），而不进行任何图像块操作。然后根据这些加权图生成融合图像。该方法通过对源图像施加互补约束，假设两个源图像是互补的。然而，正如 [182] 中指出的，前景和背景区域之间的清晰边界并不总是存在。因此，该方法在处理 FDB 时可能会遇到问题。

2) 端到端的方法

端到端方法的主要优势在于通过训练直接学习源图像与融合图像之间的映射，因此不需要后处理步骤。端到端的多聚焦图像融合方法通常采用编码器-解码器架构。其中，有些方法采用的是双分支结构，而另外一些方法采用的是单分支结构。

采用双分支结构的一个例子是 Xu 等人 [185] 提出的方法。该方法的每个分支处理一个源图像。采用单分支结构的一个例子是 Li 等人 [186] 提出的 U-net。在这种方法中，源图像首先转换为 YCbCr 颜色空间，然后使用 U 形网络融合亮度分量 (Y)¹。Cb 和 Cr 分量使用加权融合方法进行融合。

2. 基于生成式对抗网络的方法

生成对抗网络也被应用于有监督多聚焦图像融合。据笔者所知，Guo 等人 [187] 提出的 FuseGAN 是第一个基于 GAN 的多聚焦图像融合算法。FuseGAN 将多聚焦图像融合表述为图像到图像的翻译问题，并利用最小二乘 GAN 目标来增强 FuseGAN 的训练稳定性。GAN 生成一个置信图，然后通过高级卷积条件随机场进行细化。后来，Huang 等人 [188] 提出了 ACGAN，这是一种基于 GAN 的端到多聚焦图像融合方法。ACGAN 和 FuseGAN 之间有一些区别。首先，ACGAN 是一种端到端方法，可以直接输出融合图像，而 FuseGAN 需要后处理步骤。其次，ACGAN 使用由强度损失、梯度损失和结构相似性指数 (SSIM) 损失组成的生成器损失，而 FuseGAN 采用最小二乘 GAN 目标。第三，ACGAN 被训练用于融合灰度图像，因此在融合彩色图像时需要进一步处理，而 FuseGAN 可以直接融合彩色图像。

3. 基于变换器的方法

因为变换器可以捕捉全局的信息，因此在图像融合领域也得到了较多的应用。例如，山东工商学院的科研团队提出了一种有监督的基于变换器的多聚焦图像融合方法 [189]。他们提出了一种端到端的基于变换器和注意力机制的方法。

¹将源图像转换到 YCbCr 颜色空间再对 Y 分量进行操作，是很多图像融合算法使用的一种操作方式

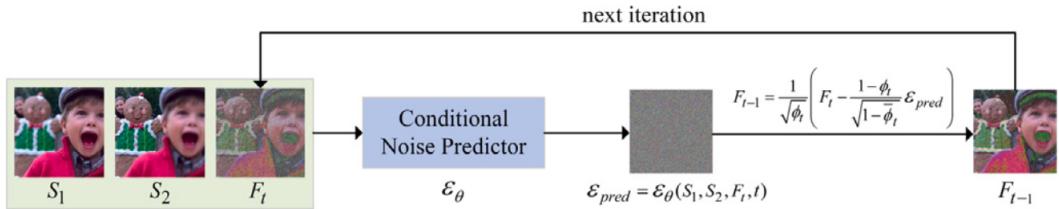


图 6.4: 基于扩散模型的多聚焦图像融合方法。图像来源于 [190]。

4. 基于扩散模型的方法

中国科技大学和中国科学院的研究人员于 2024 年发表了第一个基于扩散模型的多聚焦图像融合方法 FusionDiff [190]。该方法是基于去噪扩散概率模型 [191] 开发的。其示意图见图6.4。

FusionDiff 分为前向过程和反向扩散过程。

1) 前向扩散过程:

在前向扩散过程中, FusionDiff 使用一系列高斯噪声将清晰图像转换为纯噪声图像。在该方法的实现过程中, 这个清晰图像就是训练时候的标签图像。此外, 两张清晰区域不同的源图像, 也作为输入一起用于训练过程中。

2) 反向扩散过程:

在反向扩散过程中, 两张源图像以及潜变量被一起输入到条件噪声预测器中。条件噪声预测器的目标是预测在每一步前向扩散中添加的噪声, 然后生成比图像 F_t 更清晰的图像 F_{t+1} 。也就是说, 使用两张源图像作为指导信息, 对初始随机噪声图像进行多次迭代去噪, 以获得融合图像。

FusionDiff 有几个特点。首先, FusionDiff 在融合原理上与现有的多聚焦图像融合方法不同, 因为 FusionDiff 不依赖于任何手动活动水平测量、融合规则或复杂的特征提取网络。这意味着设计和训练 FusionDiff 更简单。其次, 前文介绍的基于生成式对抗网络是使用对抗游戏来强迫融合图像和源图像是相似的。与基于生成式对抗网络的方法不同, 基于扩散模型的方法不比较融合图像和源图像的相似性。此外, 为了使得每次运行该算法得到的融合图像差不多, FusionDiff 的作者在反向扩散过程中没有加上随机噪声。

需要指出的是, 由于扩散模型往往需要进行很多次去噪的步骤, 因此基于扩散模型的多聚焦图像融合方法的运行速度通常较慢。

5. 其他方法

除了上述提到的有监督方法外，还有其他一些有监督的方法被提出。例如，Zhai 等人 [192] 提出了一种基于去噪自动编码器 (DAE) 和深度神经网络 (DNN) 的多聚焦图像融合方法。Deshmukh 等人 [193] 提出了一种算法，利用深度置信网络 (DBN) 计算指示输入图像清晰区域的权重。此外，Lahoud 等人 [194] 提出了使用预训练神经网络提取特征，从而减轻训练阶段的负担。此外，还有基于集成学习的方法。与上述仅使用一个模型的方法不同，Naji 等人提出了基于三种卷积神经网络 (CNNs) 集成的融合算法 (ECNN) [195]。其主要思想是使用多种模型和数据集，而不仅仅是一个，以减少在训练数据集上过拟合的问题。Naji 等人还提出了一种类似的方法，称为 HCNN[196]。基于集成学习的方法的主要创新在于它们可以结合多个模型或数据集的优势，从而对各种输入更加鲁棒。

6.3.3 基于无监督学习的融合方法

有监督方法需要大量的带标签训练数据。然而，在多聚焦图像融合中通常没有可用的真实标签，因此几乎所有有监督的方法都使用实际上不真实的合成数据集。因此，开发无监督的多聚焦图像融合方法是非常理想的。

1. 基于卷积神经网络的方法

无监督的基于卷积神经网络的多聚焦图像融合方法也可以分为基于决策图的方法和端到端的方法。

1) 基于决策图的方法

除了端到端方法之外，还提出了一些基于决策图的无监督多聚焦图像融合方法。例如，Ma 等人 [197] 提出了一种基于编码器-解码器网络 (SESF) 的无监督多聚焦图像融合算法，该算法使用 SSIM 作为损失函数的一部分。在 SESF 中，编码器在推理阶段用于获取输入图像的深度特征。然后利用这些特征和空间频率 [197] 获得焦点图。接着使用一致性验证方法来获得最终的决策图。SESF 并没有使用多聚焦图像对来进行训练。该方法仅使用 MS COCO 数据集对编码器-解码器网络进行训练。

2) 端到端的方法

据我们所知，Yan 等人 [198] 提出了第一个基于卷积神经网络的无监督多聚焦图像融合算法 (MFNet)。MFNet 以一对多焦点图像作为输入，可以直接输

出全焦点图像。在 MFNet 中实现无监督训练的关键在于使用了一种基于无参考 SSIM 的损失函数 [67]。SSIM 是一种广泛使用的图像融合评估指标，用于测量源图像与融合图像之间的结构相似性。通过训练深度学习模型，优化参数以减少 SSIM 的值，从而增强源图像与融合图像之间的结构相似性。通过这种方式，不需要真实标签。

除了使用多尺度特征和稠密连接等方式来改进性能以外，一个重要的改变性能的方式是往损失函数里添加不同的项。例如，Mustafa et al. 等人在他们的一系列方法中 [199] 都在 SSIM 损失函数项的基础上添加了一个像素损失函数项。

2. 基于生成式对抗网络的方法

生成式对抗网络也已经被用于无监督多聚焦图像融合方法。据笔者所知，MFF-GAN[200] 是第一个基于生成式对抗网络的无监督多聚焦图像融合方法。该方法首先使用自适应决策块根据重复模糊原理评估源图像每个像素的清晰度。具体来说，如果像素的清晰度较高，那么在添加模糊后像素值变化更多。然后，设计了一种内容损失，专门强制生成器生成与聚焦源图像具有相同分布的融合结果。最后，使用判别器与生成器建立对抗游戏，使得融合图像的梯度图与基于源图像构建的联合梯度图相似，从而进一步增强纹理细节。

3. 基于变换器的方法

2022 年，变换器被引入到了多聚焦图像融合领域。例如，云南大学的科研团队于 2022 年提出了一个将 U 型网络和变换器结合的多聚焦图像融合方法。在该方法中，编码器部分使用的是卷积层和变换器结合的形式，而解码器部分使用的是卷积层。该方法的训练过程不需要标签。他们通过将融合图像与源图像进行对比来指导模型进行训练。

6.4 训练数据的获取

6.4.1 有监督多聚焦图像融合方法

一般来说，多聚焦图像融合任务没有标签。因此，绝大多数有监督的多聚焦图像融合方法使用伪标准图像来指导训练。伪标准图像是自然获取的图像。然后研究人员通过在这些图像像添加不同的模糊程度和区域，来制造源图像。例如，Guo 等人 [187] 基于 PASCAL VOC 2012 [201] 合成了一个大型多焦点图像数据

集，通过利用归一化的圆盘点扩散函数来模拟散焦，并分离每张图像的背景和前景。Ma 等人 [182] 通过利用一种 a-matte 边界散焦模型来表示散焦模糊效应 (DSE)，生成了训练数据。对有监督算法来说，另外一种获取训练数据的方法是使用光场相机。这种相机获取的图像可以在后期处理时再调整聚焦区域，从而形成不同模糊区域的数据集。Zhang 等人 [202] 提出了一个名为 Real-MFF 的多焦点图像数据集，该数据集包含 710 对多焦点图像及其对应的真实标签图像。该数据集就是通过对光场数据应用重新聚焦生成的。

6.4.2 无监督多聚焦图像融合方法

对于无监督多聚焦图像融合算法来说，有三种选择来生成训练数据。第一种方法是基于现有的多聚焦图像融合数据集创建训练数据。例如，Xu 等人 [161, 60] 从 Lytro 数据集中选择了 10 对图像，并使用裁剪和翻转进行数据增强。然而，由于公开的多聚焦图像融合数据集非常小，这种方法生成的训练数据可能数量不足。此外，Lytro 数据集中的散焦模糊效应 (DSE) 不多。第二种选择是使用大规模的 RGB 数据集作为训练数据并且不添加模糊。这种选择被一些无监督方法采用，以训练一个能够有效提取图像特征的网络。例如，Ma 等人 [197] 使用 MS COCO[203] 作为训练集来训练他们的网络。第三种方法是使用合成数据，如 Zhang 等人在 [200] 中所做的那样。需要注意的是，Zhang 等人在 [200] 中没有使用人工创建的真实标签。

6.5 多聚焦图像融合的发展趋势

1. 开发真实数据集

如前文所述，在多聚焦图像融合领域，主要用于训练的数据集大多是仿真数据集，即从清晰图像中添加噪声来得到一批模糊区域不同的图像。然而，这种数据与真实的多聚焦图像数据存在一些差异。在未来，开发真实的多聚焦图像融合数据集很重要。实现这一目标的一个可能方法是从光场数据中生成具有真实标签的训练数据。

2. 开发更多应用

本书第十三章将讨论多聚焦图像融合的应用。读者朋友们在那一章将看到，目前多聚焦图像融合的应用还相对有限。笔者认为，多聚焦图像融合未来的一个发展趋势是开发更多相关应用，让这项技术能够落地。

3. 高效融合多张源图像的多聚焦图像融合算法

目前大多数多聚焦融合算法的输入图像是 2 张。然而，实际应用中会经常遇到要融合多张源图像的情况，例如显微图像融合。目前常见的做法是先融合 2 张图像，然后将融合图像与第 3 张图像进行融合来得到新的融合图像。然后以此类推，直到融合所有源图像得到最终的融合图像。

然而，这种融合方法有一些缺点。首先，这种融合方法不是太高效，因为要融合 N 张源图像，我们需要将融合算法运行 $(N-1)$ 次。其次，这种融合方法无法同时充分利用所有源图像的信息，可能会使得最终的融合效果不是很好。因此，未来有必要设计开发更加高效的算法来融合多张源图像。

4. 更好的多聚焦图像融合评价基准

笔者在可见光与红外图像融合评价基准 VIFB 的基础上开发了多聚焦图像融合的评价基准，即 multi-focus image fusion benchmark (MFIFB)[149]，并且在另外一篇论文中对一些多聚焦图像融合算法进行了详细的评价。然而，在未来，开发更好的多聚焦图像融合评价基准是很有必要的。更好的评价基准可能包含更多的数据、更合适的评价指标组合以及更加恰当的性能评价策略。

5. 动态多聚焦图像融合

截止到本书写作时为止，多聚焦图像融合的研究主要聚焦于静态多聚焦图像融合，即认为在拍摄多张源图像时场景和场景中的物体是静止的。然而，在现实生活中，当我们拍摄同一个场景的不同聚焦点的图像时，场景往往会发生一些变化，例如行人走动、物体移动或者相机位置发生了变化。在这种情况下，如果使用现有的多聚焦图像融合方法来进行融合，在生成的融合图像中可能会产生幻影和模糊。因此，在这种情况下需要进行动态多聚焦图像融合。需要指出的是，动态图像融合在多曝光图像融合（见第七章）中研究得较多，而到目前为止在多聚

焦图像融合中研究得非常少。考虑到动态多聚焦图像融合的实际性，笔者认为动态多聚焦图像融合是未来的一个发展趋势。

6. 应用驱动的多聚焦图像融合方法

截止到本书写作为止，几乎所有的基于深度学习的图像融合方法的目的都是生成高质量的清晰图像。然而，这些生成的清晰图像对于下游应用的促进作用，很少被谈及。这些下游应用也并没有在图像融合过程中被考虑进去。这种设计方法，有可能会限制生成的图像对于下游任务的促进作用。在未来，笔者认为有必要设计应用驱动的多聚焦图像融合方法，即在图像融合过程中直接考虑下游应用，这样可以保证生成的融合图像对于下游应用有切实的促进作用。

6.6 小结

本章主要讨论了多聚焦图像融合，包括定义、传统多聚焦图像融合方法、基于深度学习的多聚焦图像融合方法、多聚集图像融合数据集，以及多聚集图像融合的发展趋势。

从本章的讨论我们可以看出，尽管深度学习的引入是多聚焦图像融合领域得到了很大的发展，但仍有很多问题需要解决。

第 7 章

多曝光图像融合

多曝光图像融合在于捕捉和重现丰富的视觉细节，超越单一曝光的局限。

——笔者

多曝光图像融合 (multi-exposure image fusion, MEF) 是图像融合领域的重要子任务之一，近年来吸引了越来越多的关注。除了传统方法以外，深度学习技术也越来越多地被应用于多曝光图像融合任务中。本章主要对基于深度学习的多曝光图像融合进行介绍。

7.1 多曝光图像融合概述

由于常见成像传感器的捕捉范围有限，单一图像通常因为曝光不足或过度曝光而无法揭示所有细节。多曝光图像融合技术通过融合多张曝光不同的图像来解决这一问题。如图7.1所示，多曝光图像融合是将多张不同曝光的图像融合成一张视觉上令人满意且高质量的融合图像的过程。可以看出，与其他图像融合任务类似，多曝光图像融合也是从多张源图像中结合重要信息生成高质量融合图像。大多数多曝光图像融合算法的输入是两张或多张曝光程度不同的源图像，输出是一张曝光情况良好、高质量的融合图像。

对多曝光图像融合的研究已经进行了大约 30 年。据笔者所知，早在 1993 年，Burt 等人 [204] 就进行了多曝光图像融合研究。他们提出了一种基于金字塔变换的方法来执行多种图像融合任务，包括可见光-红外图像融合、多聚焦图像融合和多曝光图像融合。在那之后，已有许多的多曝光图像融合算法被提出来。



图 7.1: 多曝光图像融合示意图。图中第一行为欠曝光图像，第二行为过曝光图像，第三行为使用 MTI 算法 [205] 进行融合得到的融合图像。从图中可以看出，跟源图像相比，融合图像的质量大大提升了。图像来源于 [39]。

7.2 传统融合方法概述

7.3 多曝光图像融合的特点

多曝光图像融合这个任务具有几个特点。首先，和其他图像融合任务一样，多曝光图像融合任务通常没有标准答案（标签），即没有标准的融合图像。其次，在多曝光图像融合里，经常会出现多张源图像的情况。这一点和可见光与红外图像融合不太一样，因为可见光与红外图像融合任务中通常只有两张源图像。此外，在多曝光图像融合里经常要考虑动态图像融合的情况，即在拍摄源图像时，相机或者物体出现了运动的情况（参见第7.4.2小节）。

7.4 多曝光图像融合方法的分类

多曝光图像融合方法可以按照不同的方式进行分类。例如，按照训练过程是否需要标签信息，可以分为有监督学习方法和无监督学习方法。此外，多曝光图像融合方法还有其他的分类方式，如图7.2所示。

7.4.1 融合两张源图像的方法和融合多张源图像的方法

根据源图像的数量，多曝光图像融合方法可以分为融合两张源图像的方法和融合多张源图像的方法。很多多曝光图像融合方法只能对两张源图像进行融合。然而，在多曝光图像融合这个任务中，经常会出现有多张源图像的情况。如果我们用一个只能融合两张源图像的多曝光图像融合方法来融合多张源图像，那么需

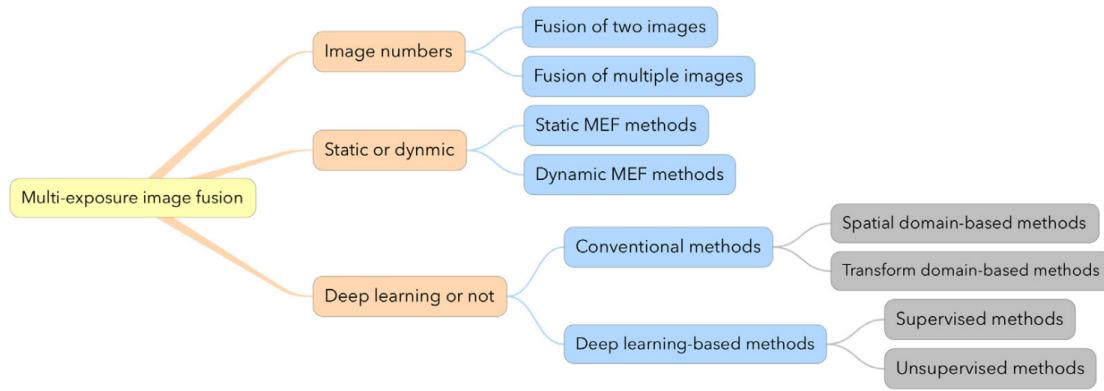


图 7.2: 多曝光图像融合方法分类示意图。图像来源于 [39]。

要反复使用该算法来对源图像进行融合。比如，先融合源图像 1 和源图像 2 得到一张融合图像，然后将该融合图像与源图像 3 进行融合，以得到一张新的融合图像，以此类推。然而，这种方法需要将多曝光图像融合算法运行多次，效率不高。因此，有些研究人员使用多分支结构来同时对多张源图像进行融合。

7.4.2 静态图像融合和动态图像融合

按照源图像中场景或者物体是否发生了变化，可以分为静态多曝光图像融合和动态多曝光图像融合。在静态多曝光图像融合中，不考虑相机运动和物体运动。换句话说，这些运动足够小，可以忽略不计。例如，使用安装在三脚架上的相机捕捉静态物体的图像是典型的静态场景。也就是说，在静态多曝光图像融合中，源图像之间只有曝光程度的不同，其他部分，如背景、前景，都是一致的。

相反，在动态多曝光图像融合中，需要考虑相机运动或物体运动，否则融合图像会出现模糊或重影效果。一个典型的动态场景例子是用手机在没有三脚架的情况下拍摄图像，这时手部运动会导致动态图像。也就是说，在动态多曝光图像融合中，源图像之间除了曝光程度不一样以外，背景和前景都可能不一样。例如，场景中的物体可能移动了，从而在不同的源图像中位置不一样。

动态多曝光图像融合方法比静态多曝光图像融合方法更难开发，因为动态多曝光图像融合算法需要处理由于相机运动导致的物体运动或不对齐。处理动态场景的一种典型方法是首先对源图像进行配准，将动态场景转换为静态场景，然后应用静态多曝光图像融合方法获取融合图像。例如，Chen 等人 [206] 提出了一种基于卷积神经网络的多曝光图像融合方法。该方法包含一个单应性网络，用于估计源图像之间的单应性矩阵，以进行图像配准。需要注意的是，动态多曝光

图像融合方法通常也适用于静态场景。

7.4.3 传统方法和深度学习方法

按照是否使用了深度学习，多曝光图像融合方法可以分为传统多曝光图像融合方法和基于深度学习的多曝光图像融合方法。其中，传统方法可以分为基于空间域的方法和基于变换域的方法。基于空间域的方法直接在空间域中操作，可以大致分为三类：基于像素的方法、基于块的方法和基于优化的方法。相比之下，基于变换域的方法首先将图像转换到另一个域，然后在该变换域中进行融合。最后，通过逆变换得到融合后的图像。

7.5 基于深度学习的融合方法

传统多曝光图像融合方法使用手工设计的特征或变换来融合多曝光图像，因此可能对不同的输入条件不够鲁棒，融合性能也受到限制。随着深度学习技术在许多领域的巨大成功，Prabhakar 等人 [207] 在 2017 年将深度学习引入多曝光图像融合领域。此后，基于深度学习的多曝光图像融合方法得到了迅速发展，许多不同的深度学习模型也纷纷被应用于这个任务中。截止到本书交稿之时，卷积神经网络、生成式对抗网络、变换器等，均已被用于多曝光图像融合任务中。

7.5.1 有监督学习方法

由于多曝光图像融合任务没有真正的标签，因此有监督方法使用的标签通常是人为“制造”的。2018 年，Wang 等人提出了一种有监督的基于卷积神经网络的多曝光图像融合方法 [208]。在该方法中，他们基于 ILSVRC 2012 数据集 [209] 生成了人工的数据集。具体地，他们将 ILSVRC 2012 数据集中的图像作为标签，然后调整那些图像的像素来得到不同曝光程度的图像作为源图像。

另一种制造标签的方式是从已有的一些方法中选择好的融合结果，并将它们用作标签。例如，在多曝光图像融合领域用得比较多的 SICE 数据集 [210] 的标签是从十三种代表性算法的融合结果中选择的。该数据集提供了 500 多个具有参考图像的静态场景数据。

2020 年，Chen 等人 [206] 提出了一种基于生成式对抗网络的方法来执行动态多曝光图像融合。据笔者所知，该方法是最早的基于生成式对抗网络的多曝光图像融合方法之一。该方法包含三个主要组件：用于补偿相机运动的单应性估计

网络、用于生成融合图像的生成器和用于对抗学习的判别器网络。该方法的主要创新点有两个。首先，该方法使用单应性网络来估计单应性以补偿相机运动，因此可以应用于动态场景。其次，该方法采用了基于生成式对抗网络的模型，通过对抗学习可以减轻伪影。该方式是使用 SICE 数据集进行训练的，是一种有监督学习的方法。

此外，武汉大学马佳义教授团队也于 2020 年提出了一种基于生成式对抗网络的多曝光图像融合方法——MEF-GAN [155]，并用 SICE 数据集进行了训练。MEF-GAN 与 Chen 等人 [206] 的方法的不同之处主要有三点。首先，Chen 等人的方法包含一个单应性估计网络，因此可以应用于动态场景，而 MEF-GAN 设计用于静态场景。其次，MEF-GAN 中采用了自注意力机制，以实现注意力驱动和长程依赖。第三，MEF-GAN 的损失函数包含内容损失，包括均方误差 (MSE) 项和梯度项，而 Chen 等人的方法则没有。然而，MEF-GAN 对源图像的尺寸有要求（图像的宽度和高度都应能被 8 整除），因此不能应用于任意空间分辨率，这在一定程度上限制了其应用。

值得说明的是，由于许多多曝光图像融合方法中的标签是从其他算法的融合结果中选择的，而不是通过光学镜头拍摄的，因此它们可能不够精确或不适用。因此，这种有监督的多曝光图像融合方法的结果受到其他算法的限制。为了解决这些问题，研究人员提出了一系列无监督的多曝光图像融合方法。

7.5.2 无监督学习方法

1) 基于卷积神经网络的方法

2017 年，Prabhakar 等人提出了一个基于卷积神经网络的多曝光图像融合算法——DeepFuse [207]。据笔者所知，DeepFuse 不仅是第一个基于深度学习的多曝光图像融合方法，也是第一个无监督方法。DeepFuse 的方法示意图如图 7.3 所示。可以看到，它包括几个步骤。首先，将输入图像转换为 YCbCr 颜色通道数据。其次，利用包含特征层、融合层和重建层的 CNN 来融合 Y 通道。色度通道 (Cb 和 Cr) 通过加权融合进行融合。这是因为图像的结构细节主要体现在亮度通道，而亮度通道的亮度变化比色度通道更明显。第三，将融合后的 YCbCr 数据转换到 RGB 空间，以获得融合图像。为了实现无监督学习，DeepFuse 使用了一种无参考图像质量度量 MEF-SSIM [36] 作为损失函数。然而，它需要在不同的颜色空间之间转换图像，这相比直接融合 RGB 图像的方法来说并不直观。此

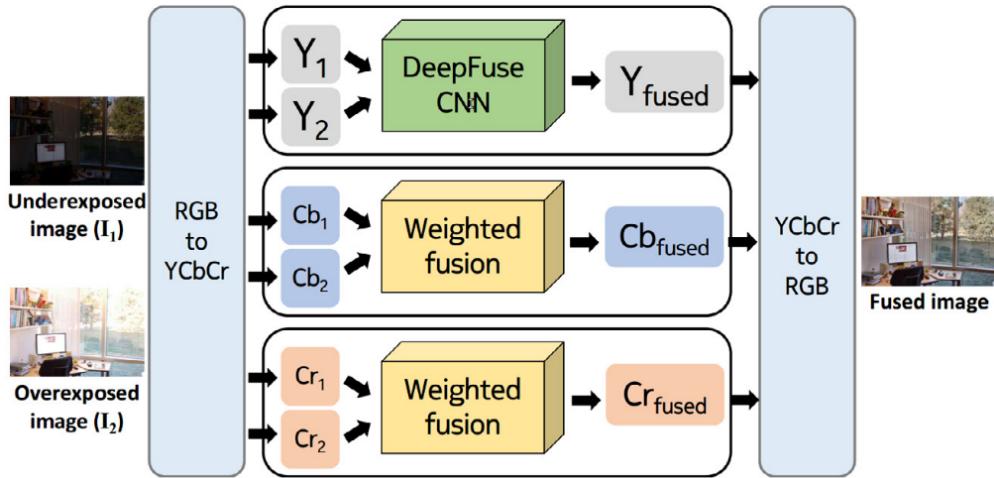


图 7.3: DeepFuse 的方法示意图。图像来源于 [207]。

外，它仅使用 MEF-SSIM 作为损失函数，因此无法学习和保留 MEF-SSIM 未涵盖的其他特征。

在 DeepFuse 之后，研究员们提出了一些列基于卷积神经网络的方法。这些方法从不同的方面对 DeepFuse 进行改进，例如使用稠密连接网络代替传统的卷积神经网络、使用多分支网络来支持多张源图像、对下采样的图像进行操作以提升融合效率、增加损失函数项以提升融合效果、使用 MEF-SSIMc [211] 代替 MEF-SSIM 以便直接在 RGB 颜色空间内对图像进行融合。

2) 基于生成式对抗网络的方法

Yang 等人提出了一种基于生成式对抗网络的多曝光图像融合方法，即 GANFuse[212]。GANFuse 与本书 7.5.1 小节中提到的另外两种基于生成式对抗网络的多聚焦图像融合方法有两个主要区别。首先，GANFuse 是一种无监督方法，使用无监督损失函数来训练。具体而言，该损失函数旨在衡量融合图像与源图像之间的相似性，而不是与真实图像的相似性。其次，GANFuse 由一个生成器和两个判别器组成。每个判别器用于区分融合图像和一个源图像。

3) 基于变换器的方法

变换器最主要的特点是可以捕获全局信息，因此对于多曝光图像融合任务也很有好处。2022 年，复旦大学的科研人员提出了第一个基于变换器的多曝光图像融合方法——TransMEF [213]。如图 7.4 所示，TransMEF 是基于编码器-解码器结构设计的。

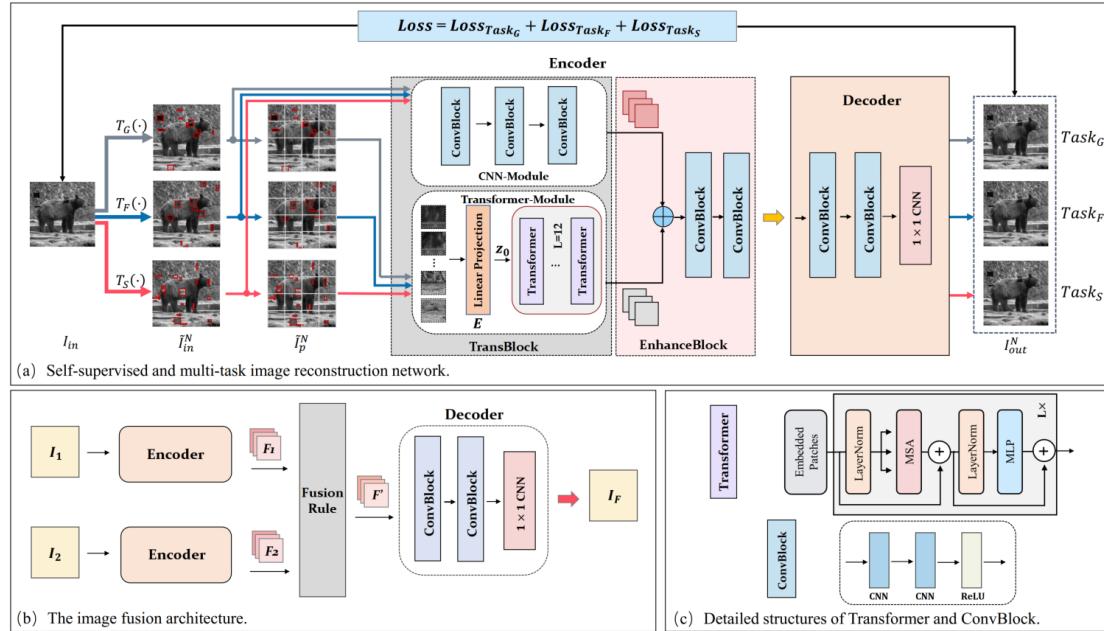


图 7.4: TransMEF 的方法示意图。图像来源于 [213]。

TransMEF 的编码器中包含了特征提取模块和特征提升模块。其中，特征提取模块里包含了卷积神经网络模块和变换器模块。因此，该编码器既可以利用局部信息，也可以利用全局信息。TransMEF 的解码器一些卷积模块。

TransMEF 是使用自然图像（MS COCO 数据集）并且利用自监督多任务学习的方式来进行训练的，因此不需要使用不同曝光程度的源图像来进行训练。具体地，TransMEF 的作者们设计了三种变化方式。在训练的时候，将变换后的图像分别输入给编码器进行特征提取和增强，然后输入给解码器进行图像重建。这种基于图像重建训练编码器-解码器结构的思路，和笔者在本书第五章介绍过的 DenseFuse 有点类似。值得说明的是，由于使用了三种变化方式，TransMEF 的训练过程是一种多任务学习的方式。

在训练完成以后做融合时，TransMEF 先使用训练好的编码器从两张源图像中提取特征。然后，TransMEF 将来自两张源图像的特性求平均以得到融合特征。最后，使用训练好的解码器将融合特征重建为融合图像。与其他很多图像融合方法类似，在对 RGB 图像进行融合时，TransMEF 先将源图像转换到 YCbCr 色彩空间，然后对 Y 通道使用上述方法进行融合。Cb 和 Cr 通道则采用传统的加权融合方法进行融合。然后，将融合图像从 YCbCr 色彩空间转换到 RGB 色彩空间。

在 TransMEF 之后，还有一些其他的基于变换器的方法可以进行多曝光图

像融合。典型的例子包括合肥工业大学刘羽副教授于 2023 年提出的方法 [214]。该方法与 TransMEF 有不少共性，如都是基于编码器-解码器结构，都结合了卷积神经网络和变换器，都使用自然图像进行训练。然而，该方法使用了多尺度特征并且使用了一个特征交互模块加强特征之间的交互。此外，该方法在解码器中使用了嵌套连接架构。

值得说明的是，还有一些基于变换器的通用图像融合方法可以进行多曝光图像融合，如 SwinFusion [59]。通用图像融合方法将在本书第九章中讨论。

4) 其他方法

值得指出的是，目前绝大多数多曝光图像融合方法仅仅是用来完成多曝光图像融合这一个任务。然而，也有一些研究人员尝试了将多曝光图像融合和其他任务进行结合。例如，北京航空航天大学邓欣副教授曾提出了一个将多曝光图像融合和图像超分辨率两个任务结合起来的方法 [215]。

7.6 多曝光图像融合的发展趋势

1. 高效地融合多张源图像

如前文所述，目前很多多曝光图像融合算法只能对固定数量的源图像（例如两张源图像）进行融合。这和《射雕英雄传》里全真教的“天罡北斗阵”一样，必须要七个人一起使用才行。在这种情况下，多曝光图像融合方法的使用可能不是太方便。因此，笔者认为，未来有必要开发可以融合不固定数量源图像多曝光图像融合算法。理想的多曝光图像融合算法是像《倚天屠龙记》里面张三丰发明的“真武七截阵”那样，七个人可以用，六个人、五个人都可以用，自由度很大。

2. 更好的评价基准

笔者于 2021 年发表了多曝光图像融合领域的第一个评价基准，即 multi-exposure image fusion benchmark (MEFB) [39]。与本书 4.4 部分介绍过的 VIFB 类似，MEFB 中也包含测试集、代码库、评价指标以及相关接口，并且也是基于 Matlab 进行开发的。MEFB 中包含 100 对多曝光图像，其中每对图像包含一张过曝光图像和一张欠曝光图像。

笔者研发的多曝光图像融合评价基准 (MEFB) 是对多曝光图像融合评价基

准的初步探索，因此还有许多不足之处。例如，MEFB 中的每对图像仅包含两张图像，而实际中可能会有多张不同曝光程度的图像需要进行融合。此外，MEFB 中包含的测试集主要是静态图像，而实际中很多时候场景是动态的，需要进行动态曝光图像融合。

笔者认为，未来需要开发更好的评价基准来解决这些问题。此外，更好的评价基准还需要从更大的数据集和更合理的评价指标组合等方面去考虑。

3. 应用驱动的多曝光图像融合

截止到本书完稿为止，几乎所有的多曝光图像融合的目的都是生成高质量的融合图像，而缺乏对下游任务的详细考虑。尤其是，已有的多曝光图像融合方法没有在融合的过程中考虑下游任务的需求，因此，这些方法生成的融合图像对下游任务有多大的促进效果是一个未知数。笔者认为，应用驱动的多曝光图像融合算法应当是多曝光图像融合任务里一个重要的未来发展趋势。

此外，笔者认为需要进一步挖掘多曝光图像融合的应用价值。关于多曝光图像融合的应用，参见本书第十四章。

7.7 小结

本章主要对多曝光图像融合技术进行了介绍。我们介绍了多曝光图像融合的分类以及一些典型的基于深度学习的多曝光图像融合方法。我们还讨论了多曝光图像融合领域未来的发展趋势。

第8章

医学图像融合

8.1 概述

根据笔者的总结分析，目前绝大多数图像融合的研究人员来自中国。然而有一个例外，那就是医学图像融合。

8.2 问题定义

8.2.1 医学图像融合的类别

8.2.2 医学图像融合的关键点

8.3 传统融合方法概述

8.4 基于深度学习的融合方法

8.4.1 发展历程概述

8.4.2 基于有监督学习的融合方法

8.4.3 基于无监督学习的融合方法

8.4.4 训练数据的获取

8.4.5 3D 医学图像融合

[216, 217]

8.5 小结

第 9 章

通用图像融合方法

如前文所述，图像融合按照任务的不同可以分为不同的类型，例如可见光红外图像融合、多聚焦图像融合、多曝光图像融合。到目前为止，绝大多数图像融合算法是针对某一类图像融合任务进行设计的。例如笔者在本书前几章中介绍的那些算法。近年来，也有一些学者开始研究通用图像融合方法，即可以应用于多种图像融合任务的方法。本章简要讨论一下通用图像融合方法。

9.1 传统通用图像融合方法

在深度学习被引入图像融合领域之前，已经有一些学者研究过通用图像融合方法，称为传统通用图像融合方法。其中的代表性的方法包括湖南大学李树涛教授提出的基于引导滤波 [218] 的图像融合算法 (GFF)[46]。在 GFF 论文中，其作者将该算法应用于多曝光图像融合、多聚焦图像融合、红外可见光图像融合和医学图像融合，取得了较好的效果。GFF 算法的原理如图9.1所示。

GFF 算法自从被提出以来，引起了图像融合领域研究人员的重点关注，并累计被引用超过 1400 多次¹，成为了最具代表性的传统通用图像融合算法之一。此外，有一些通用图像融合算法是在 GFF 算法的基础上进行改进而得到的，如上海交通大学和上海电力大学的研究团队提出的基于多尺度引导滤波的图像融合算法（MGFF）[3]。

然而，传统通用图像融合方法仍然具有传统图像融合方法的缺点，即图像融合的三个阶段均是人为设计的，因此难以在各种工况下取得良好的融合性能。因此，近年来，研究人员开始研究基于深度学习的通用图像融合方法。

¹ 截止到 2023 年 4 月份

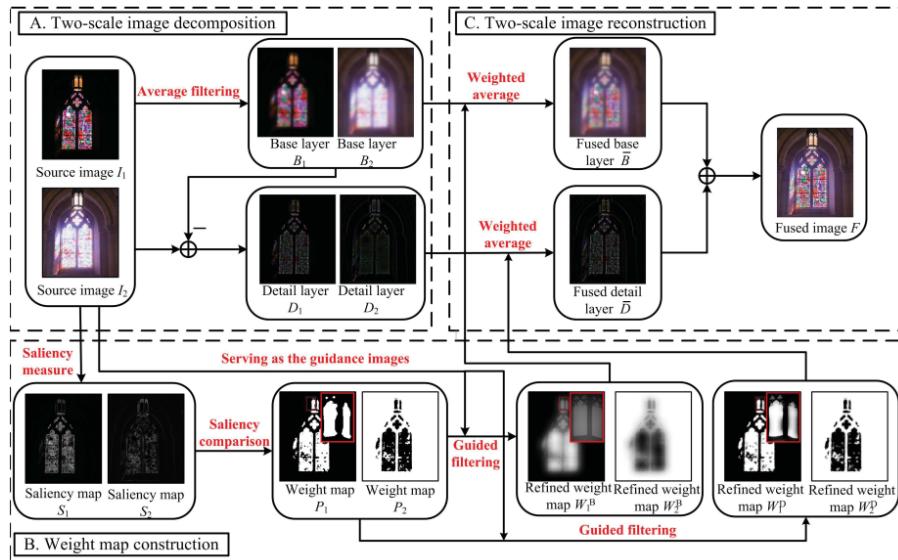


图 9.1: GFF 算法原理示意图。图片来源于 [46]。

9.2 基于深度学习的通用图像融合方法

9.2.1 概述与分类

基于深度学习的通用图像融合方法是指基于深度学习的、可以用于多种图像融合任务的图像融合方法。在基于深度学习的通用图像融合方法中，采用的基本思路很接近，即收集训练数据、训练模型、将模型用于融合不同类型的图像融合任务。

基于深度学习的通用图像融合方法也可以分为基于卷积神经网络的方法、基于生成对抗网络的方法、基于自编码器的方法、基于 transformer 的方法和其他方法。比较有代表性的方法包括 IFCNN [62]、U2Fusion [60] 和 SDNet [158] 等。这些方法均可以被用于多种图像融合任务。需要注意的是，由于遥感图像融合任务与可见光红外图像融合、多聚焦图像融合、多曝光图像融合等差别较大，因此，通用图像融合方法一般不会用来同时处理遥感图像融合任务和其他类型的图像融合任务。

9.2.2 实现方式

一般来说，通用图像融合方法仅含有一个网络模型。为了实现将同一个模型用于多种不同的图像融合任务，通常采用的方法有以下几种。

第一，针对不同的图像融合任务，选取不同的损失函数权值。例如，在武汉

表 9.1: 代表性的通用图像融合方法。VIF 为可见光与红外图像融合, VNIF 为可见光与近红外图像融合, MFIF 为多聚焦图像融合, MEF 为多曝光图像融合, MEDIF 为医学图像融合。

方法	发表地方	年份	类别	实现方法	图像融合任务
[194]	arXiv	2019	基于 CNN	预训练模型	VIF, MEDIF, MFIF
IFCNN [62]	INFUS	2020	基于 CNN	不同融合策略	VIF, MEDIF, MFIF, MEF
FusionDN [161]	AAAI	2020	基于 CNN	弹性权重固化、不同训练数据	VIF, MEDIF, MFIF, MEF
PMGI [162]	AAAI	2020	基于 CNN	不同损失函数权值、不同训练数据	VIF, MEDIF, MFIF, MEF
SDNet[158]	IJCV	2021	基于 CNN	不同损失函数权值、不同训练数据	VIF, MEDIF, MFIF, MEF
U2Fusion [60]	TPAMI	2022	基于 CNN	弹性权重固化、不同训练数据	VIF, MEDIF, MFIF, MEF
SwinFusion[59]	JAS	2022	基于 Transformer	不同训练数据	VIF, VNIF, MEDIF, MFIF, MEF

大学马佳义教授团队提出的 PMGI 方法 [162] 中, 他们将不同的图像融合任务统一为梯度信息和强度信息的比例维持问题, 并据此设计了统一的损失函数形式。在实践中, 他们针对不同图像融合任务的特点, 手动调整损失函数中各项的权值。

第二, 针对不同的图像融合任务, 采用不同的特征融合方法。这类方法使用同一个模型来完成不同的图像融合任务, 但是考虑了不同类型的源图像的不同特点。因此, 在进行融合的时候, 会针对不同的图像融合任务采取不同的融合策略。这类方法的典型代表是 IFCNN [62]。在 IFCNN 中, 进行多曝光图像融合时采取的融合方法时是逐元素均值策略, 而在进行其他类型的图像融合(可见光与红外图像融合、多聚焦图像融合、医学图像融合)时采取的融合方法是逐元素最大值策略。

第三, 使用针对其他视觉任务(例如图像分类)的预训练模型来提取特征, 从而进行图像融合。这方面的典型例子是 Lahoud 等人开发的图像融合算法 [194]。他们使用在 ImageNet 上预训练的模型来提取源图像特征, 并将源图像特征进行融合以完成不同的图像融合任务。

表9.1中总结了一些代表性的基于深度学习的通用图像融合方法, 供读者朋友参考。值得说明的是, 通用图像融合已经吸引了不少研究人员的关注, 并有一些其他的实验通用图像融合的方式被提出, 例如 FusionDN 和 U2Fusion 中使用的弹性权重固化方法。由于篇幅限制, 本书在此不做详细介绍。此外, 值得说明的是, 一些通用图像融合方法 [194, 62] 使用某个任务的数据来进行模型训练, 然后直接将训练好的模型用于其他图像融合任务。也有很多通用融合方法 [60, 158, 59] 是针对每种图像融合任务使用不同的训练数据, 以便获得不同图像融合任务上的优异性能。

9.3 通用图像融合方法的优缺点

9.3.1 优点

通用图像融合方法的优点显而易见，即使用方便并且实用——一个方法即可完成多种图像融合任务。此外，一些通用图像融合方法还可以利用不同图像融合任务之间的共性特点和内在联系 [219]。在深度学习被引入图像融合领域以后，鉴于深度学习强大的特征学习能力，不断有学者提出基于深度学习的通用图像融合方法，取得了较好的进展。

9.3.2 缺点

虽然通用图像融合方法有着很好的优点，但这类算法也存在着显著的缺点，即性能可能不是特别好。从笔者在前文中对几种不同图像融合任务的介绍中可以知道，各种图像融合任务有着不同的源图像、目的和特点，因此，一个通用的模型很难同时捕捉到这些特点。例如，一个能够很好地提取并融合可见光图像和红外图像信息的模型，很可能无法同时很好地提取不同聚焦点的图像信息。笔者曾在人工智能顶级期刊 TPAMI 上发文 [?] 对一些通用图像融合算法在多聚焦图像融合这个问题上的性能进行过实验分析，发现通用图像融合算法在多聚焦图像融合这个任务上的性能往往比不上专门的多聚焦图像融合算法。

中国台湾大学的李宏毅老师在他的课程中介绍 Transformer 时曾指出，虽然各种自然语言处理（NLP）的任务都可以用序列到序列（seq2seq）的模型去解决，但是对多数自然语言处理任务，往往对各个任务定制化模型效果会更好。他还举例说，瑞士军刀虽然可以用来做很多事情，但是它不见得是一个最好用的。另外一个例子是 2022 年底火起来的 ChatGPT。虽然它可以针对很多问题给出很好的答案，但是对于一些很专业的问题，它就无法给出好的答案了。同样的道理，虽然通用图像融合算法可以被用于不同的图像融合任务，但是针对每种图像融合任务的目的和特点来单独设计定制化的融合算法，往往能得到更好的融合效果。

有趣的是，笔者就通用图像融合方法和特定图像融合方法的对比询问过 ChatGPT 的意见²，它给出的答案如图9.2所示。应该来说，笔者非常赞同 ChatGPT 的看法。

²2023 年 4 月 25 日



通用图像融合方法和特定的图像融合方法，哪个更好一些？



通用图像融合方法和特定的图像融合方法各有其优缺点，哪个更好取决于具体的应用场景和需求。

通用图像融合方法是指能够适用于多种不同类型图像的融合方法，通常是基于机器学习或深度学习的技术实现。这种方法具有较强的通用性和灵活性，能够在多种应用场景下实现图像融合。但是，由于其通用性，其融合效果可能不够优秀，在某些特定场景下可能无法满足需求。

特定的图像融合方法是指针对特定场景或任务而设计的图像融合方法，通常是基于领域知识和专业技术实现。这种方法具有较强的针对性和专业性，能够在特定场景下实现高质量的图像融合。但是，由于其专业性，其适用范围可能相对较窄，难以在其他场景下使用。

综上所述，对于不同的应用场景和需求，需要综合考虑通用图像融合方法和特定的图像融合方法的优缺点，并选择最适合的方法来实现图像融合。

图 9.2: ChatGPT 对于通用图像融合方法和特定图像融合方法的看法。

9.4 小结

本节简单地介绍了通用图像融合算法，包括传统通用图像融合算法和基于深度学习的通用图像融合算法。此外，本节还对通用图像融合算法的优缺点进行了简要分析。

尽管通用图像融合算法目前还没有取得特别好的融合性能，但是由于它们可以被应用于多种不同的图像融合任务，因此仍然有不少研究人员在研究通用图像融合算法。或许，在不久的将来，会有很多更好的通用图像融合算法被提出。让我们拭目以待吧！

第 10 章

应用驱动的图像融合方法

在前文中，我们介绍了多种图像融合方法¹。在那些图像融合方法中，输入是源图像，输出是融合图像。也就是说，研究人员设计那些算法的目的是生成高质量的、视觉效果好的融合图像。然而，正如本书第一章所介绍过的，图像融合的目的除了生成高质量的融合图像以外，还有通过图像融合来提升后续应用的性能。实际上，在很多情况下，图像融合更重要的目的是通过提供更好的源图像来提升后续应用的性能。尽管笔者在前文中介绍的那些图像融合方法可能可以提升后续应用的性能，但在那些图像融合算法的设计过程中并未考虑后续应用的情况。因此，目前绝大多数图像融合算法的设计是与后续应用脱节的。

笔者于 2021 年在笔者的公众号“笑书神侠读博学”发表了《双剑合璧，为了观赏还是为了实战?》一文（图10.1），指出了图像融合领域重视生成高质量图像而忽略后续应用的问题，引起了研究人员的重视。近一两年来，已经有研究人员着手做了一些初步研究。然而，截止本书写作为止，学术界关于应用驱动的图像融合方法的研究还非常少。在本章中，笔者将介绍近年来应用驱动的图像融合方法的发展情况，并简单介绍几个代表性的工作。

¹如无特别说明，则本章所说的图像融合方法均是指像素级图像融合方法



图 10.1: 笔者的文章

10.1 应用驱动的图像融合方法的优势

如前所述，在绝大多数图像融合研究中，输入是源图像，输出是融合图像。在整个图像融合过程中，未考虑后续的应用问题，如图10.2(a)所示。在这种情况下，融合图像包含了一些源图像里的通用特征，因此生成的融合图像对于后续应用可能会有一定的提升作用。例如，笔者曾经用实验证明这种类型的图像融合方法可以在一定程度上提升目标跟踪的性能 [163]。然而，由于在图像融合过程中未考虑后续应用的情况，因此生成的融合图像对于后续应用不一定是非常合适的。在这种情况下，融合图像对于后续应用的提升效果可能很有限。与此相反，应用驱动的图像融合算法在设计图像融合算法的时候就将后续应用考虑在内，如图10.2(b)所示。因此，应用驱动的图像融合方法所生成的融合图像对于后续应用有更好的性能提升作用。事实上，在一些其他的图像融合领域，应用的性能才是更加重视的方面。典型例子是可见光与深度图像的融合 (RGBD)。

此外，由于在图像融合任务中没有 ground truth 存在，因此在做性能评价的

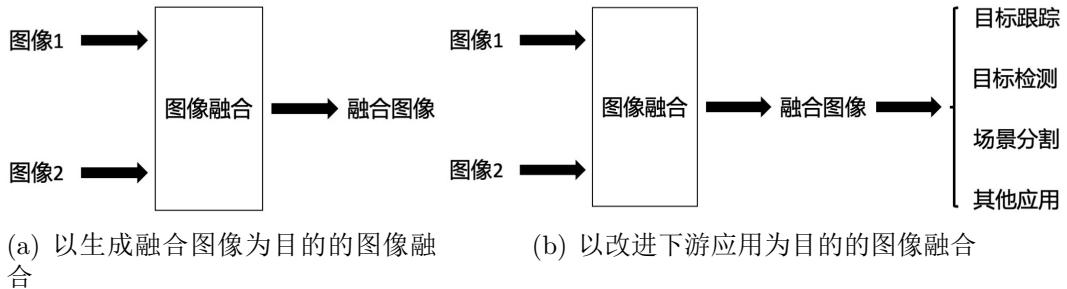


图 10.2: 以生成融合图像为目的的图像融合和以改进下游应用为目的的图像融合。前者的目的是生成高质量融合图像，而后者的目的则是通过生成融合图像来改进下游应用的性能。

时候不是很容易。本书第四章介绍过一些图像融合性能评价方法和当前图像融合领域性能评价方面存在的问题。另一方面，一般来说，后续的应用，例如目标跟踪和目标检测，是有 ground truth 的。因此，以应用为导向的图像融合方法的另一个好处，就是可以通过后续应用的 **ground truth** 来评价图像融合方法的性能好坏。

基于上述原因，近年来，包括笔者在内的一些学者开始研究应用驱动的图像融合方法。

10.2 应用驱动的可见光与红外图像融合方法

10.2.1 行人检测驱动的图像融合方法

比利时根特大学的科研人员于 2019 年发表了一项研究工作，用行人检测来辅助可见光与红外图像融合 [150]。该研究的目的是为了在汽车辅助驾驶系统 (ADAS) 中给人类司机提供更便于观察的融合图像。具体地，他们希望在融合图像和可见光图像尽可能接近，但是行人区域在各种环境条件下均能看清。为了实现这个目的，他们将融合图像输入给一个行人检测器，然后用行人检测的结果和图像融合的结果一起构成目标函数来训练图像融合模型。因为该方法在训练过程中同时考虑了图像融合的性能和目标检测的性能，因此用该方法训练得到的图像融合算法相较于一般的图像融合方法，更能提升行人检测的性能。此外，相较于特征级的图像融合方法，该方法可以输出便于人类观察的融合图像，以供 ADAS 系统使用。

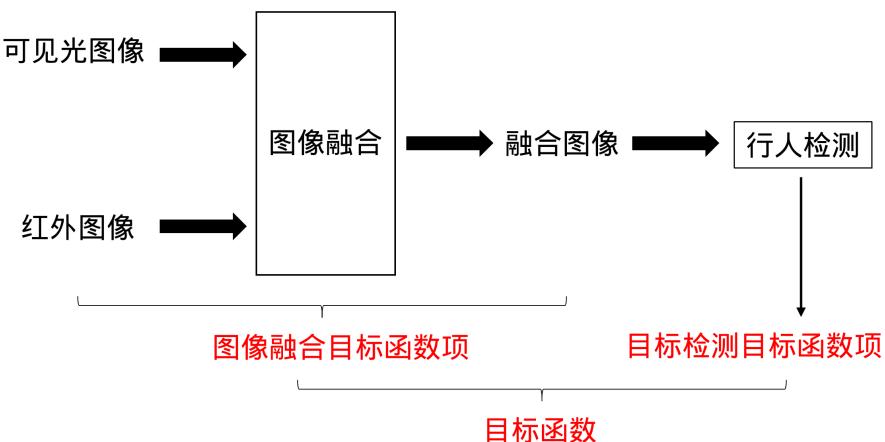


图 10.3: 行人检测驱动的可见光与红外图像融合方法

这个方法的关键是他们的目标函数中不仅仅包含了基于图像融合评价指标设计出来的目标项，也包括基于行人检测性能设计出来的目标项，如图10.3所示。事实上，这也是各种应用驱动的图像融合方法的关键。值得注意的是，在该方法中，只有图像融合方法参与训练，而行人检测器不参与训练。行人检测器的主要目的是提供行人检测目标函数项。

10.2.2 语义分割驱动的图像融合方法

2022 年，武汉大学马佳义教授团队提出了一种同时考虑可见光与红外图像融合和语义分割的新方法 [40]，称为 SeAFusion。该方法将可见光与红外图像融合和语义分割模型串联在同一个网络模型中，并且使用一个包含了内容目标函数项和语义目标函数项的目标函数来训练模型，如图10.4所示。由于使用了从语义分割的结果构造的语义目标函数项，SeAFusion 既能提供好的融合图像，也能较好地提升语义分割应用的性能。此外，为了提升算法的实时性，马教授团队在该方法中设计了一个轻量型的梯度残差稠密模块来进行图像融合。值得注意的是，在 SeAFusion 中，图像融合算法和语义分割算法均参与训练，这是与前文所述比利时根特大学的那项研究工作的重要区别。

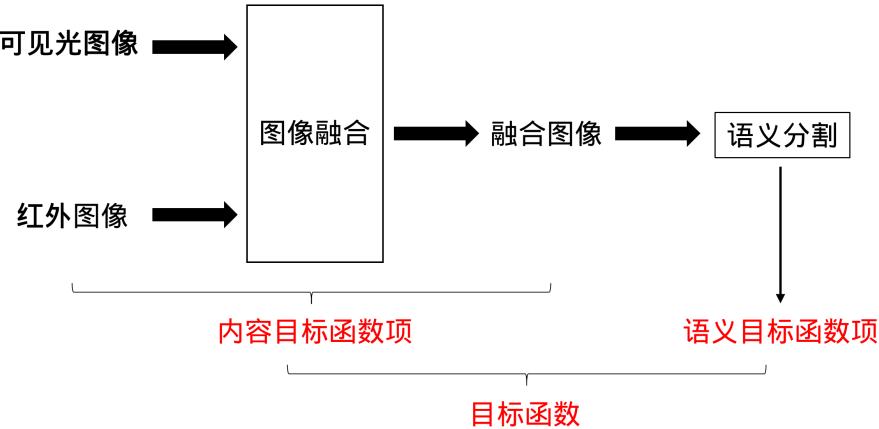


图 10.4: 语义分割驱动的可见光与红外图像融合方法

10.2.3 通用目标检测驱动的图像融合方法

2022 年，大连理工大学樊鑫教授团队在计算机视觉顶级会议 CVPR 上发表了一项研究工作，称为 TarDAL^[75]。TarDAL 包含一个图像融合模型和一个目标检测模型。其目的在于生成可以提升目标检测性能的融合图像。前文所述比利时根特大学的那项研究工作主要关注行人检测，而 TarDAL 关注不同类别的目标检测。

为了实现生成高质量的融合图像并能提升目标检测的性能，樊鑫教授团队的研究人员设计了一个双层优化问题，并提出了一个联合训练策略用以同时对图像融合模型和目标检测模型的参数进行训练。TarDAL 的图像融合算法以生成对抗网络为基础，包含一个生成器和两个判别器。TarDAL 的目标检测算法是以著名的 YOLOV5 为基础。该团队用实验证明，他们的设计不仅可以生成视觉效果良好的融合图像，并且可以很好地提升后续目标检测任务的性能。

10.3 应用驱动的其他图像融合方法

据笔者所知，截止到本书写作时为止，尚未有应用驱动的多聚焦图像融合方法和多曝光图像融合方法被发表。此外，应用驱动的医学图像融合方法也非常非常少。2022 年，合肥工业大学的刘羽副教授团队提出了一项相关工作^[220]。他们在设计多模态核磁共振图像的融合算法的时候，将神经胶细胞瘤的分割任务也考虑在内。具体地，他们的方法是基于生成对抗网络设计的，其中生成器是图像融合算法，而 2 个判别器是神经胶细胞瘤分割网络。这项工作很有意义，因为以往的绝大多数医学图像融合方法未考虑实际应用，而只是用一些图像融合的

评价指标去评价融合性能。笔者在英国帝国理工学院的实验室里有一位同事是伦敦某医院重症监护室的全职医生²。笔者曾就医学图像融合的实际应用需求咨询过他。当时他给我的回复说医学图像融合是有实际价值的，只是通过图像融合评价指标来评价，与实际应用有些脱节。

10.4 小结

本章主要介绍近年来开始发展的应用驱动的图像融合算法研究。笔者首先分析了现有图像融合算法研究存在的问题，然后指出了应用驱动的图像融合研究的益处。接着，笔者介绍了近年来比较有代表性的几个应用驱动的图像融合研究工作。值得注意的是，目前基于应用驱动的图像融合方法主要出现在可见光和红外图像融合，以及医学图像融合上。在多聚焦图像融合和多曝光图像融合中，尚未见到应用驱动的方法。

附：《双剑合璧，为了观赏还是为了实战？》³

图像融合是一个被研究了很多年的领域。

所谓图像融合，是指把具有互补信息的两幅或多幅图像进行融合以得到包含信息更丰富或者质量更高的图像的过程。

常见的图像融合类型包括多曝光图像融合、红外与可见光图像融合、多聚焦图像融合、医学图像融合和遥感图像融合。

图像融合和金庸先生笔下杨过和小龙女的双剑合璧很类似。双剑合璧是为了融合全真剑法和玉女剑法，取长补短以增强威力，而图像融合是为了将不同图像进行取长补短，为后续应用（如目标跟踪）提供更好的输入图像以提高性能。

最近几年，随着深度学习的发展，越来越多的研究人员开始关注图像融合领域。每年发表的图像融合论文数量，尤其是基于深度学习的图像融合论文数量，在迅速增加。

其中，红外可见光图像融合是最受关注的领域之一。

然而，当前红外可见光图像融合研究存在一个重要问题，即重视生成图像而忽略了后续应用。具体来说，就是绝大多数图像融合算法的目的都是生成融合图像，使其在某些评价指标上取得好的结果，而忽略了其在后续应用中起的作用。

²这位同事是个奇人，也是位大牛。他是一位全职医生，平时只能用业余时间或者假期攻读博士学位、做科研，并且懂很多东西

³本文于 2021 年发表于笔者的微信公众号“笑书神侠读博学”

类比到双剑合璧中，就是重观赏性而轻实战性。打比方说，就是杨过和小龙女进行双剑合璧是为了观赏性强，而不是为了打赢别人。

但是实际上，除了少数情况（比如双剑合璧的发明人林朝英，她发明双剑合璧是为了有朝一日能和王重阳调情）以外，大多数时候双剑合璧的目的是打败对手。比如，杨过和小龙女双剑合璧是为了打败金轮法王。也就是说，实战性比观赏性更重要。

同样地，在红外可见光图像融合里，在绝大多数情况下融合的首要目的应该是改善后续应用（如目标跟踪）的性能，而不是为了生成好看的图像。

遗憾的是，到目前为止，绝大多数红外可见光图像融合相关论文的关注点均是生成好的融合图像，而忽视了生成的图像对后续应用是否有帮助。而能在SD、SSMI等评价指标上取得好的结果的融合图像不一定有助于提高后续应用的性能。

欣喜的是，已有部分研究人员开始对此进行反思。例如最近我就收到几个相关的提问。

希望“重观赏轻实战”的问题可以引起越来越多的红外可见光图像融合研究人员的重视，并一起来进行改善！

第三部分

图像融合的实践与展望

第 11 章

图像融合实践

纸上得来终觉浅，绝知此事要躬行

——陆游

11.1 编程语言及深度学习框架选择

11.1.1 编程语言选择

根据笔者的经验，在深度学习被引入图像融合领域以前，绝大多数图像融合算法的代码是用 Matlab 写的。典型例子如湖南大学李树涛教授团队的 GFF 图像融合方法 [46]。在深度学习刚刚被引入图像融合时，Matlab 也依然被用于编写图像融合算法，典型例子如合肥工业大学刘羽老师团队开发的基于卷积神经网络的图像融合算法 [22] 和江南大学吴小俊教授团队开发的基于 ResNet 的图像融合算法 [58]。这两个算法的深度学习部分都是用 MatConvNet 实现的。然而，随着深度学习方法渐渐成为图像融合领域的主流方法，Python 成为了当前基于深度学习的图像融合方法首选的语言。这主要是因为当前主要的深度学习框架，例如 Tensorflow 和 Pytorch，均是基于 Python 语言的。除此以外，使用 Python 进行图像融合算法开发还可以方便地调用 OpenCV。

基于上述原因，笔者建议，如果要开发基于传统方法的图像融合方法，那么 **Matlab** 是不错的选择；如果要开发基于深度学习的图像融合方法，那么首选的编程语言是 **Python**。

表 11.1: 在图像融合算法中使用过的深度学习框架

序号	深度学习框架	算法示例
1	Theano	DeepFuse [207]
2	MatConvNet	[22, 58]
3	TensorFlow	U2Fusion [60], FusionDN [161]
4	Pytorch	IFCNN[62], RFN-Nest [108], SEDRFuse [106]

11.1.2 深度学习框架选择

在第二章中，笔者已对现有的一些深度学习框架进行过简单介绍。在此处，笔者将根据自己的实践经验和对文献的分析，讨论一下在开发基于深度学习的图像融合算法时该如何选择深度学习框架。

在阅读和整理文献的过程中，笔者对在图像融合中使用过的深度学习框架进行了总结。在图像融合中使用过的深度学习框架如表11.1所示，主要包括如下几种：

在上述框架中，Theano 在图像融合中使用得很少 (仅在 DeepFuse 算法中使用过)，并且 DeepFuse 算法还需要安装 Lasagne 和 Matlab，因此使用起来不太方便。MatConvNet 主要在深度学习刚刚被引入图像融合时在少数几个算法中使用过。随着深度学习框架的发展，MatConvNet 已基本上退出了图像融合领域。目前来说，TensorFlow 和 Pytorch 是基于深度学习的图像融合算法使用的主要深度学习框架。

在图11.1中，笔者统计了在 2018 年到 2022 年 10 月份间发表的基于深度学习的可见光与红外图像融合方法所使用的深度学习框架的情况¹。从图中我们可以总结出以下几点：

- Pytorch 是目前在可见光与红外图像融合中使用得最多的深度学习框架。
- Tensorflow 在 2019 年左右被引入可见光与红外图像融合领域。其使用率迅速超过了 Matmab 的使用率。之后，Tensorflow 的快速提高，并在 2022 年开始出现下滑。
- Pytorch 在 2020 年左右被引入可见光与红外图像融合领域。其使用率迅速增加，并很快超过了 Tensorflow 的使用率。目前，Pytorch 在基于深度学习的可见光与红外算法中占据绝对优势。

¹需要指出的是，有许多图像融合论文没有开源源代码，并且有一些论文并没有在论文中提到其使用的深度学习框架，因此图11.1并不是完全统计，故仅供读者参考。

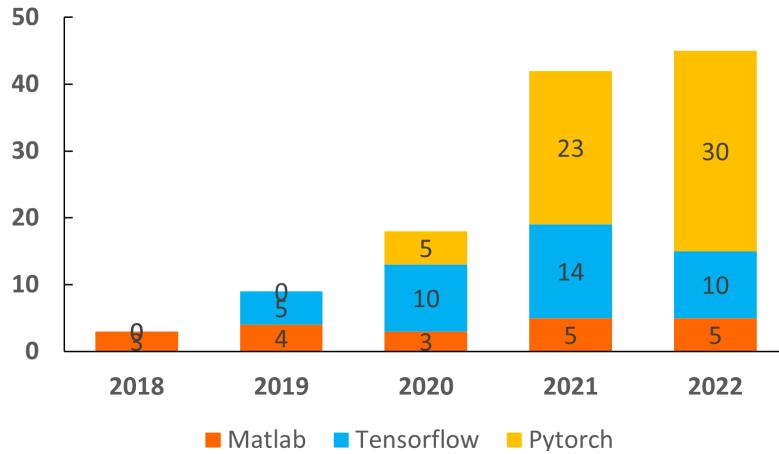


图 11.1: 2018 年至 2022 年 10 月份期间，基于深度学习的可见光与红外图像融合算法使用的深度学习框架情况统计。

基于上述分析，并考虑到 Pytorch 各版本之间的兼容性优于 TensorFlow 各版本之间的兼容性，笔者建议读者朋友们在开发图像融合算法时优先考虑使用 Pytorch。

11.2 使用 VIFB 进行可见光红外图像融合

本节演示一下如何使用笔者开发的可见光与红外图像融合领域的首个可见光红外图像融合评价基准，即 Visible-infrared image fusion benchmark (VIFB)，进行图像融合。关于 VIFB 的细节，可以参考本书第4.4小节和文献 [2]。

VIFB 中主要包含几个部分：测试集、融合算法和评价指标。因此，在使用 VIFB 时需要对这三个方面进行设置。

(1) 选择测试图像

测试图像的选择需要通过 *util* 文件夹下的 *configImgsvI.m* 和 *configImgsvIR.m* 两个文件来完成，如图11.2所示。VIFB 中一共包含了 21 对可见光与红外图像。用户可以选择其中任意一对或者多对图像来进行图像融合。

```

14 %function imgs=configImgVI
15
16 %path = 'Your own path\VIFB\input\' ;
17 - path = 'D:\Dropbox\Github\VIFB\input\' ;
18
19 - img_VI = {struct('name','carLight','path',strcat(path,'VIN'),'ext','jpg'),...
20 struct('name','carShadow','path',strcat(path,'VIN'),'ext','jpg'),...
21 struct('name','carWhite','path',strcat(path,'VIN'),'ext','jpg'),...
22 struct('name','elecbike','path',strcat(path,'VIN'),'ext','jpg'),...
23 struct('name','fight','path',strcat(path,'VIN'),'ext','jpg'),...
24 struct('name','kettle','path',strcat(path,'VIN'),'ext','jpg'),...
25 struct('name','labMan','path',strcat(path,'VIN'),'ext','jpg'),...
26 struct('name','man','path',strcat(path,'VIN'),'ext','jpg'),...
27 struct('name','manCall','path',strcat(path,'VIN'),'ext','jpg'),...
28 struct('name','manCar','path',strcat(path,'VIN'),'ext','jpg'),...

```

图 11.2: 在 VIFB 中设置测试图像。

(2) 选择融合算法

图像融合算法的选择需要通过 *util* 文件夹下的 *configMethods.m* 文件来进行, 如图11.3所示。VIFB 中集成了 20 种图像融合算法, 读者可以选择其中一种或者多种来进行图像融合。

```

7 %function methods=configMethods
8
9 - methodVIFB={struct('name','ADP'),...
10 struct('name','CBF'),...
11 struct('name','CNN'),...
12 struct('name','DLF'),...
13 struct('name','FPDE'),...
14 struct('name','GRCE'),...
15 struct('name','GFF'),...
16 struct('name','GTF'),...
17 struct('name','HMSD_GF'),...
18 struct('name','Hybrid_MSD'),...
19 struct('name','IFEVIP'),...
20 struct('name','LatLRR'),...
21 struct('name','MGFF'),...

```

图 11.3: 在 VIFB 中设置图像融合算法。

(3) 选择评价指标

图像融合指标的选择需要通过 *util* 文件夹下的 *configMetrics.m* 文件来进行, 如图11.4所示。VIFB 中集成了 13 种图像融合评价指标, 读者可以选择其中一种或者多种评价指标来对图像融合的结果进行评价。

```
configMethods.m x +  
7 function methods=configMethods  
8  
9 - methodvIEFB={struct('name','ADF'),...  
10     struct('name','CBF'),...  
11     struct('name','CNN'),...  
12     struct('name','DLF'),...  
13     struct('name','FPDB'),...  
14     struct('name','GFCB'),...  
15     struct('name','GFF'),...  
16     struct('name','GTF'),...  
17     struct('name','HMSD_GF'),...  
18     struct('name','Hybrid_MSD'),...  
19     struct('name','IFEVIP'),...  
20     struct('name','LatLRR'),...  
21     struct('name','MGFF'),...  
     ...};  
     % (length(vIEFB))
```

图 11.4: 在 VIFB 中设置图像融合算法。

在设置好测试图像、融合算法和评价指标以后，用户可以运行 *main_running.m* 程序² 来开始进行图像融合。所选择的图像融合方法将对每一对用户所选择的可见光与红外图像融合对进行融合操作。所得到的融合图像将保存在 *output/fused_images* 文件夹中。然后，用户可以运行 *compute_metrics.m* 来计算评价指标的结果。每一幅融合图像都会使用用户所选择的评价指标进行评价。评价指标的计算结果会保存在 *output/evaluation_metrics* 和 *output/evaluation_metrics_single* 文件夹下。图11.5展示了使用 VIFB 中的 20 种图像融合算法得到的关于 *fight* 这对可见光与红外图像的 20 幅融合图像。

²务必设置正确的路径



图 11.5: 在 VIFB 中得到的关于 fight 图像对的 20 种融合结果。

从上述过程可以看出，使用 VIFB 来进行可见光与红外图像融合是很容易的。除了上述使用方法以外，用户还可以将新的测试图像添加到 VIFB 中以便使用 VIFB 中的融合方法进行融合。用户也可以将在别处生成的融合图像（如使用基于 Python 的融合方法得到的融合图像）添加到 VIFB 中以便使用 VIFB 中的评价指标进行性能评价。

当然，VIFB 最重要的价值是提供一个统一的标准以对可见光图像融合算法进行性能评价，一改以往文献中“王婆卖瓜，自卖自夸”的局面，并将这种思想传递给研究社区。

11.3 代表性图像融合方法的使用

11.3.1 FusionGAN

11.3.2 DenseFuse

11.3.3 Dif-Fusion

第 12 章

可见光与红外图像融合的应用

本章主要介绍可见光与红外图像融合的应用。本质上，像素级图像融合对于下游任务的促进作用，都是通过产生高质量的融合图像而实现的。不管是可见光红外图像融合，多聚焦图像融合还是多曝光图像融合，都是如此。本章首先对红外图像的常见应用进行介绍，然后对像素级可见光与红外图像融合的应用进行介绍。最后，本章将对其他层级的可见光与红外图像融合的应用进行简单介绍。

12.1 红外图像的常见应用总结

作为一种与 RGB 图像很不一样的图像类型，红外图像有其独特的优势。红外图像由于不受光照条件、雾、烟等的影响，并且常常可以显著地区分行人、动物和背景，因此，研究人员已经将红外图像用于各种不同的应用中。本节对红外图像的常见应用进行简单总结。

1. 计算机视觉领域

红外图像在计算机视觉领域的多个任务里有应用。典型应用如下：

目标跟踪。红外图像已被一些研究人员用于目标跟踪。例如，哈尔滨工业大学深圳校区的研究人员基于红外图像开展了一系列的目标跟踪相关的工作 [221, 222, 223]。该团队研发的 LSOTB-TIR 是基于红外图像的目标跟踪领域里非常重要的数据集和评价基准。

人体感知及相关应用。由于人体一般与环境中的物体具有不同的温度，因此红外图像非常适合用于感知人体。基于此，研究员们已将红外图像应用于一系



图 12.1: 笔者在英国演示基于红外图像的人体姿态估计。图片由笔者拍摄。

列与人体感知有关的任务中。例如，南京南空航天大学的研究人员使用红外相机对夜晚对停泊在机场的飞机周围进行人体运动识别 [224]，另外有研究人员基于红外相机进行行人横穿马路的行为识别 [225, 226]、行人检测 [227, 228]、行人跟踪 [229, 230]。此外，在 2023 年的 CVPR 上，研究人员开发了一种基于红外图像来推断人体过去一段时间所做运动的算法 [231]。具体地，该算法利用人物接触过程中人体留在物体上的热量信息来推断人在 3 秒钟以前的行为。除了上述应用以外，笔者发现，基于红外图像来做人体姿态估计，效果相当不错。图 12.1 中展示的是笔者在 2023 年英国 Great Exhibition Road Festival（大展览路科技艺术节）上演示基于红外图像的人体姿态估计时候的场景。可以看出，该演示引起了大量观众的兴趣。

2. 机器人领域

红外图像的感知目标物体表面温度信息的能力，使得其在黑暗中和有烟雾的情况下也可以正常工作，因此在机器人领域有广泛的应用。以下是几个红外图像在机器人领域的应用示例：

目标检测和跟踪：机器人可以使用红外图像来检测和跟踪热源，例如人体、动物和机器设备等。这对于机器人在不同环境中实现自主导航和避障非常有用。

导航和避障：红外图像可以帮助机器人导航和避障，因为它们可以识别环境中的热源。例如，机器人可以使用红外图像来找到人类或动物的位置，以避免与他们碰撞。此外，红外图像也可用于机器人建图和导航。例如，有研究人员给机

器人装备了红外相机，以便其可以在有浓烟的情况下进行建图和导航 [232]。

工业检测：红外图像可以用于检测机器和设备中的故障。例如，它们可以检测机器零件的温度是否超过了正常范围，从而帮助机器人识别可能需要修理或更换的部件。美国波士顿动力公司开发的 Spot 机器人，在装备红外相机以后，已经被用于对工厂设备进行检测¹。

工业自动化：红外图像可用于检测生产线中的设备是否正常运行。机器人可以使用红外图像来监测电气设备或机器部件的温度，以便及时发现问题并采取措施避免故障。

消防和安全：机器人可以使用红外图像来检测火灾和其他潜在的危险情况。机器人可以快速扫描建筑物并识别温度异常的区域，以便消防员或安全人员进行更有效的响应。

受害者检测与识别：机器人领域顶级期刊《科学机器人》(Science Robotics)于 2021 年 6 月报道了在无人机上搭载红外传感器，以对森林中的受害者进行检测和识别，从而便于营救的文章 [167]。该研究展现出了红外相机在受害者检测营救方面的良好潜力。

总之，红外图像在机器人领域具有广泛的应用，可以帮助机器人更好地感知环境，实现自主导航和控制，并为人们提供更好的安全和健康保障。

3. 其他应用

1) 人体生理指标分析

机器人可以使用红外图像来监测病人的体温，并检测是否存在发热等症状。例如，在疫情期间，美国麻省理工学院的科研人员在机器狗（Spot）上放置了红外相机，并基于开红外相机开展了人体生理指标的分析 [233]。图12.2中展示该研究中机器人的示例。

2) 野生动物检测

因为野生动物通常也具有和环境物体不同的温度，因此红外相机也十分适合用于检测野生动物。澳大利亚的一些研究人员在无人机上搭载了红外相机，以便对野生动物进行检测 [234]。

¹<https://www.bostondynamics.com/solutions/inspection/thermal>

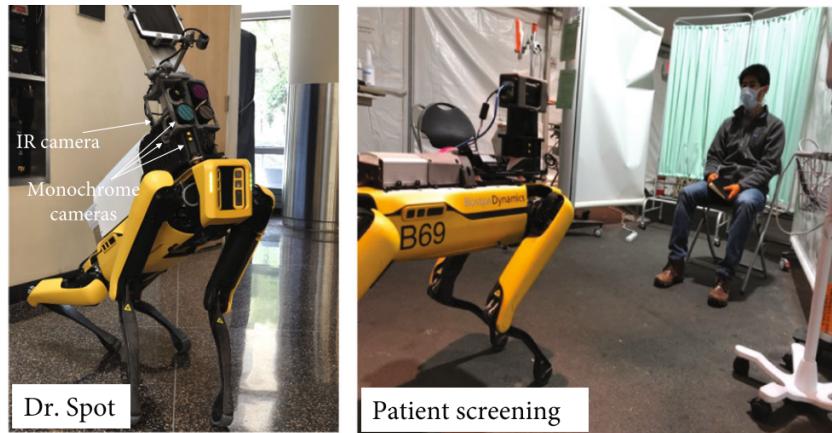


图 12.2: 在 Spot 机器狗上放置相机以进行人体生理指标测量。图片来源于 [233]。

3) 其他

红外图像还有许多其他用处，笔者在此不再详细介绍。

12.2 红外图像的缺点

尽管红外图像有很多优点，但是红外图像也有缺点。在本节中，我们将探讨热红外图像技术的一些缺点。

- 分辨率有限，且无法提供像 RGB 图像那样丰富的纹理和细节信息：相比于传统的可见光图像，热红外图像的分辨率通常较低。由于红外辐射的波长较长，所以在相同的像素尺寸下，其细节捕捉能力相对较差。这可能导致在某些应用场景下无法准确地辨别目标物体或识别细微的特征。
- 红外相机价格高昂。：热红外成像设备通常价格较高，特别是高性能的工业级和军用级设备。笔者实验室曾购买过一个 FLIR DUO PRO R 相机，花费高达 5000 英镑。
- 受环境影响：热红外图像的质量和可靠性很大程度上受环境条件的影响。例如，雨、雪、雾等恶劣天气会导致红外辐射的传播受阻，从而影响图像的清晰度和可见范围。此外，大气湿度和温度的变化也可能干扰热红外图像的获取和解释。
- 不能穿透遮挡物：热红外图像只能探测目标表面的红外辐射，而无法穿透实体物体或遮挡物，这限制了其在一些特定应用中的使用。例如，在寻找隐藏在建筑物内部的目标或者地下的物体时，热红外图像无能为力。

- 不能提供颜色信息：由于热红外图像是基于目标物体辐射的温度差异来生成的，它不能提供物体的真实颜色信息。这在某些情况下可能导致对目标的准确定位和识别产生困难。
- 受到大气吸收影响：大气中的某些气体对红外辐射具有吸收作用，特别是在长波红外范围内。这会导致热红外图像在一些特定波长段的信号衰减，从而影响图像质量和可靠性。

虽然热红外图像技术存在一些缺点，但随着科技的不断发展，相信这些问题将逐渐得到解决或缓解。例如，国际顶级学术期刊《自然》杂志于 2023 年 9 月 7 日发表了封面文章 [235]，介绍了美国科研团队对于热成像的改进，展现出了非常好的潜力。也有一些研究人员在使用一些深度学习技术，例如 transformer，来对红外图像进行后处理以改善质量 [236]。同时，热红外图像技术在许多应用领域仍然具有不可替代的优势，尤其是在夜间观测、隐蔽目标探测和高温环境下的工作等方面。因此，我们期待未来热红外图像技术在更广泛领域取得更多突破和应用。

12.3 像素级可见光和红外图像融合的应用

本节介绍像素级可见光和红外图像融合的应用。像素级融合是指先生成融合图像，然后基于融合图像进行下游任务，如图12.3所示。需要指出的是，在笔者于 2021 年呼吁研究人员关注基于图像融合的应用的以后，基于像素级可见光和红外融合的应用多了起来。

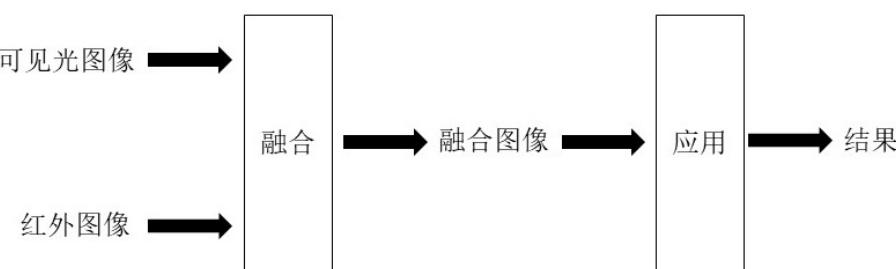


图 12.3: 基于像素级图像融合的应用的基本思路。

1. 目标跟踪

基于视觉的运动目标跟踪是指在视频中对指定目标进行定位跟踪的过程。一般来说，需要指定目标在视频第一帧中的位置，然后要求跟踪算法在后续帧中找

到目标的位置来形成跟踪轨迹。运动目标跟踪在民用及军事领域均有重大意义和迫切需求。近年来，随着深度学习和相关滤波技术的迅速发展，目标跟踪领域得到了巨大的发展，目标跟踪效果取得了很大的进步。然而，目前绝大多数目标跟踪方法是基于可见光图像的，而这类方法的性能在可见光图像不可靠时会大大降低。例如，当光线条件不好，或者在雨、雾等天气条件下，基于可见光的目标跟踪算法性能会大大降低。红外图像检测物体的热信息，因此不受光线和雨、雾等因素的影响。因此，将可见光和红外图像的互补信息在目标跟踪中合理利用，有望大大提高目标跟踪算法的性能、适用范围和鲁棒性。

笔者于 2019 年研发了基于孪生网络的可见光与红外图像融合目标跟踪方法，对于像素级可见光和红外图像融合在目标跟踪中的应用进行了研究成 [163]。该算法的基本思路是，首先使用图像融合算法对可见光和红外图像进行融合，得到融合图像。然后从融合图像中裁剪出模版图像和搜索图像，然后使用基于孪生网络的目标跟踪算法 (SiamFC) 来进行目标跟踪。笔者的工作证明了像素级可见光和红外图像融合可以提升目标跟踪的性能。

2. 目标检测

2022 年，大连理工大学樊鑫教授团队提出了一个结合像素级目标跟踪和目标检测的新算法 (TarDAL)，并提出了一个新的可见光和红外图像目标跟踪数据集 (M³FD) [75]。该算法的基本思路是在目标函数中同时使用图像融合的目标项和目标检测的目标项来对模型进行训练。因此，通过该算法得到的融合图像，不仅具有好的质量，而且对于下游的目标检测任务也具有很好的提升效果。该算法的示意图和在 M³FD 数据集上的测试效果分别如图 12.4 和 ?? 所示。

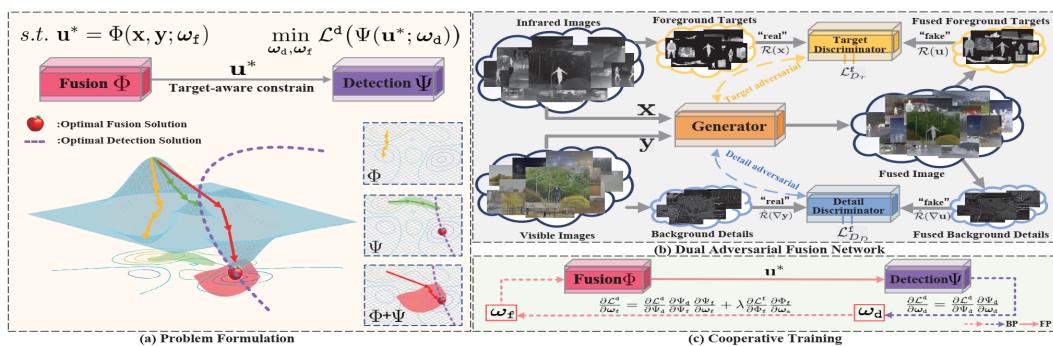
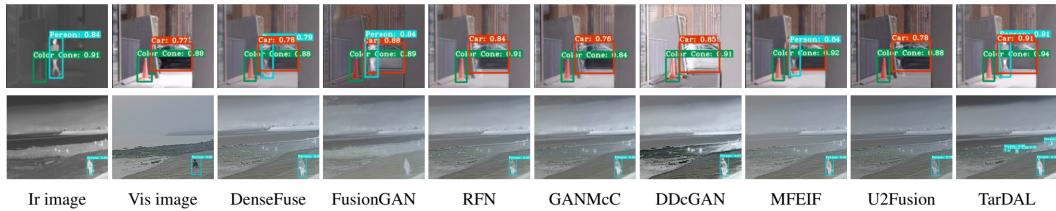


图 12.4: TarDAL 算法的基本思路。图像来源于 [75]。

图 12.5: TarDAL 算法在 M³FD 数据集上的测试效果。图像来源于 [75]。

3. 语义分割

本书第十章中在介绍应用驱动的图像融合方法时，曾介绍过语义分割驱动的可见光和红外图像融合方法。该方法是像素级可见光和红外图像融合在语义分割这个任务上的典型应用。图10.4中介绍了该方法的基本原理。图12.6中展示了该方法的结果。

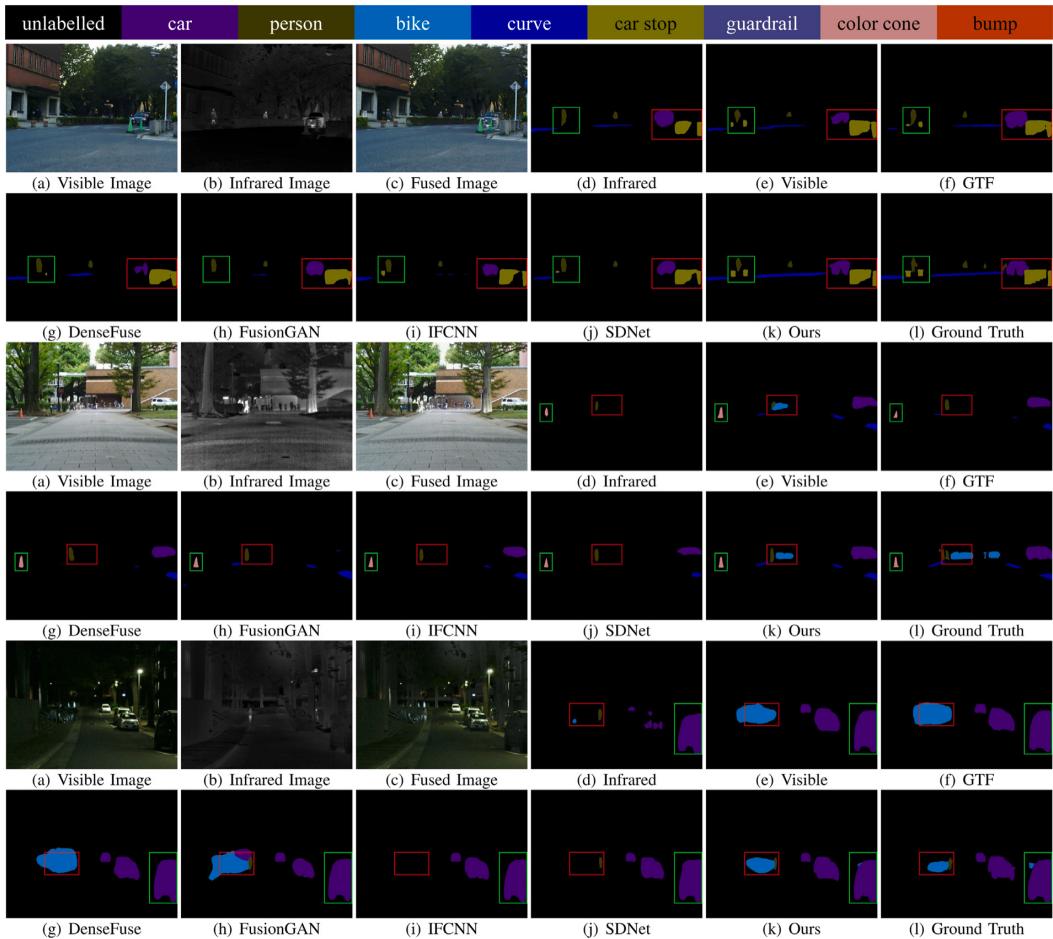


图 12.6: 像素级可见光和红外图像融合在语义分割中的应用效果示意图。图像来源于 [40]。

4. 三维重建

谢等人 [237] 于 2023 年提出了一种基于语义信息的可见光与红外图像配准算法。在该论文中，他们将像素级可见光和红外图像融合应用到了三维重建中，即先对可见光和红外图像进行融合，然后基于融合图像进行三维重建。图12.7中展示了他们的三维重建实验结果。他们的实验结果表明了像素级可见光和红外图像融合的确对三维重建任务有帮助。

5. 深度估计

谢等人 [237] 也将像素级可见光和红外图像融合应用到了深度估计任务中，即首先对可见光和红外图像进行融合，然后基于融合图像进行深度估计。

图12.7中展示了他们的三维重建实验结果。他们的实验结果表明了像素级可见光和红外图像融合的确对三维重建任务有帮助。

6. 其他应用

除了上述应用以外，像素级可见光和红外图像融合还在不少其他任务上得到了应用，如混凝土结构的缺陷分割 [238, 239]、飞行视觉系统 [240]、野火检测 [241]。笔者相信，未来会有越来越多的基于像素级可见光和红外图像融合的应用出现。

12.4 其他层级的可见光与红外图像融合的应用

12.4.1 基于特征级融合的应用

除了在像素级对可见光和红外图像进行融合以外，研究人员还对可见光和红外图像进行特征层级的融合并应用于许多不同的应用中，如图12.9所示。实际上，大多数基于可见光和红外图像的应用都是基于特征级融合进行的。比较典型的算法有安徽大学李成龙教授团队提出的 DMCNet 跟踪算法 [242] 和大连理工大学卢湖川教授团队提出的 ADRNet 跟踪算法 [243]。

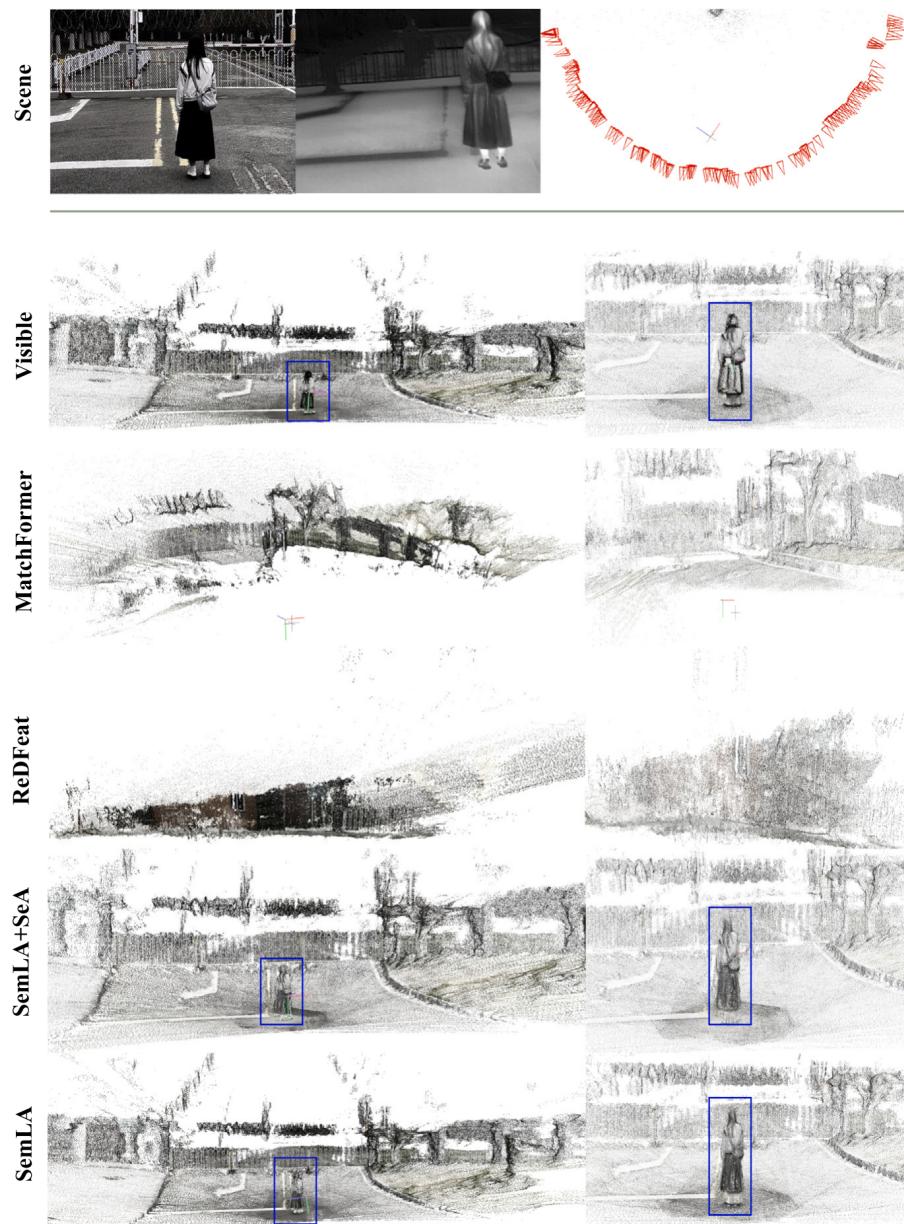


图 12.7: 像素级可见光和红外图像融合在三维重建中的应用效果示意图。图像来源于 [237]。

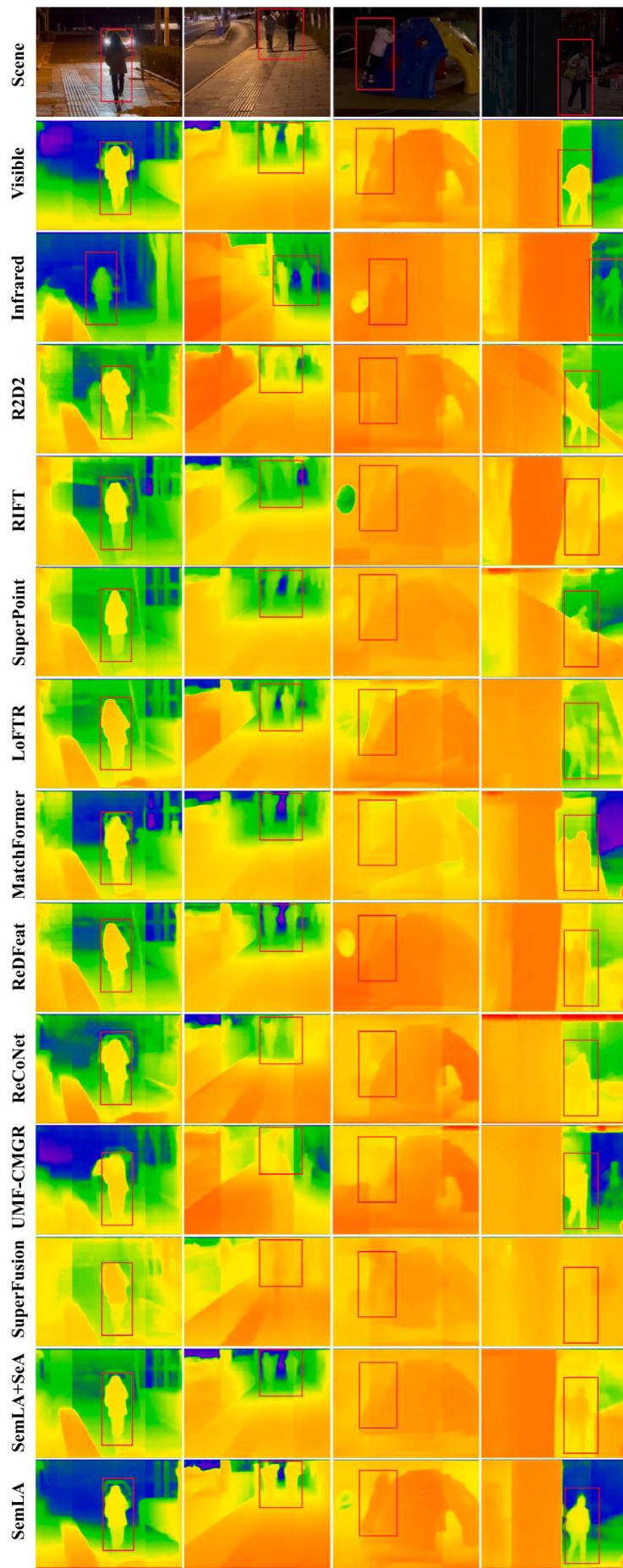


图 12.8: 像素级可见光和红外图像融合在深度估计中的应用效果示意图。图像来源于 [237]。

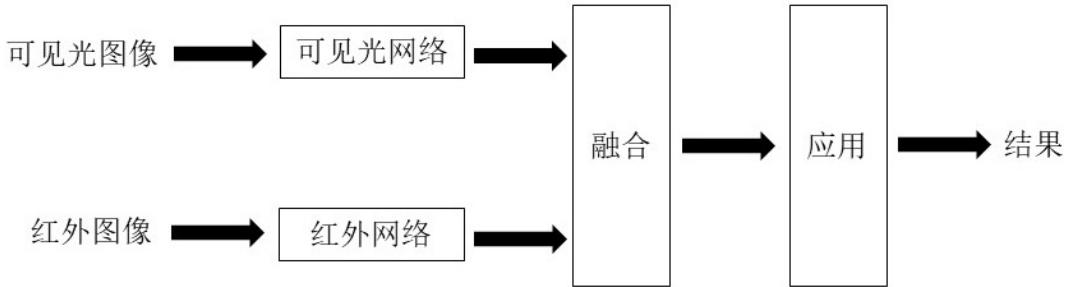


图 12.9: 基于特征级图像融合的应用的基本思路。

12.4.2 基于决策级融合的应用

笔者曾荣获 2019 年第九届中国信息融合论文最佳论文提名奖。当时的获奖论文，就是基于决策级融合的可见光与红外图像目标跟踪 [244]。在该论文中，笔者提出了一种基于孪生网络的决策级融合算法，基本思路如图 12.10 所示。首先，笔者使用两个孪生网络分别基于可见光视频和红外视频进行目标跟踪，然后将两种跟踪结果在决策级进行融合。在融合之前，笔者设计了一种方法对可见光和红外图像的可靠性进行了计算，并基于该可靠性进行两种跟踪结果的决策级融合，取得了不错的效果。

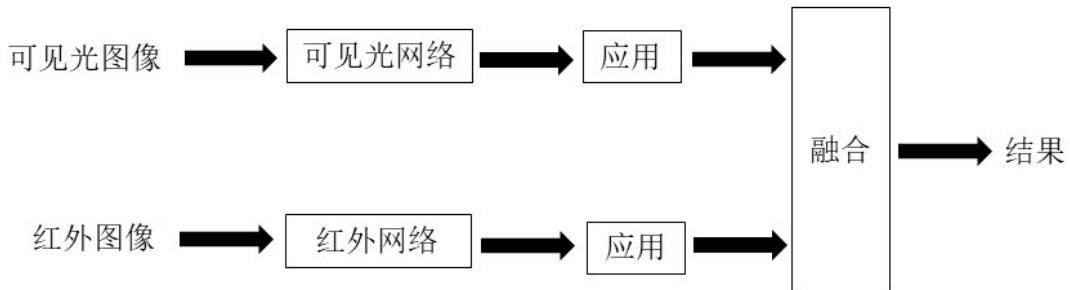


图 12.10: 基于决策级图像融合的应用的基本思路。

12.4.3 基于多个层级融合的应用

也有一些学者尝试将多个层级的融合进行结合，以取得更好的效果。例如，大连理工大学卢湖川教授团队于 2022 年在计算机视觉顶级会议 CVPR 上提出了一个叫 HMFT 的可见光-红外融合跟踪算法。在该算法中，研究人员对可见光和红外图像的信息在不同层次进行融合，即先在特征级进行融合，然后在决策级进行融合，取得了非常好的跟踪效果。



图 12.11: 可见光与红外图像融合的应用。

12.5 可见光与红外图像融合的应用小结

总体来说，可见光与红外图像融合已经被应用于许多领域中。除了非常常见的目标跟踪和行人检测以外，可见光与红外图像融合的应用还包括场景分割 [245, 246, 247]、深度估计 [248]、目标检测 [249, 250, 251, 63]、人群计数 [252, 253, 254]、视频监控 [255, 256]、生理指标监测 [257, 258]、人脸识别 [259, 260]、明火探测 [261]、即时定位与地图构建 [168]、电力设备检测 [262, 263, 264]、显著性检测 [265]、自动驾驶 [250]、玻璃分割 [266]、交通违规检测 [267]、建筑物裂纹检测 [268] 等。图12.11对可见光与红外图像融合的一些应用进行了总结。笔者认为，随着技术的进步和越来越多的研究人员的加入，可见光与红外图像融合在未来将被应用于更多的应用中。

12.6 展望

在研究过程中，笔者发现尽管近年来有大量的基于可见光和红外图像融合的应用被提出，但是该领域仍有许多问题尚未解决，有待进一步的研究。在此，笔

者就可见光和红外图像融合的应用进行一下展望。

1. 基于未配准的可见光与红外图像对的应用

据笔者所知，到目前为止，几乎所有的可见光与红外融合跟踪算法均是基于配准好的可见光与红外视频对的。然而，可见光与红外视频的严格配准在实际应用中比较难实现。例如，图12.12中展示了上海交通大学先进航空电子和智能信息处理实验室（AAII）拍摄的未严格配准的视频对。除此之外，尽管号称是严格配准的，但是目前的融合跟踪数据集中有大量视频并未进行严格配准，例如GTOT[7] 数据集和 RGBT234 数据集 [8]。图12.13展示了一些未配准的例子。因此，设计能够基于未配准的可见光与红外视频对进行目标跟踪的融合跟踪算法具有很重要的实际价值。



图 12.12: 未配准的可见光与红外视频对



图 12.13: 常见的可见光与红外数据集 (GTOT [7] and RGBT234 [8]) 中的未配准现象。

在各个层级融合中，笔者认为决策级融合跟踪更适合用于处理未配准的问题，因为决策级融合跟踪算法对于图像配准的要求最低。一个可以用来处理未配准问题的方法是用图像预处理技术将图像进行配准，然而这样很难实现精确配准。另一个更加实际可行的方法是求出可见光图像和红外图像之间的仿射变换，

然后把可见光图像的目标框变换到红外图像（或者把红外图像的目标框变换到可见光图像）以得到最终的目标跟踪结果。此时，需要考虑可见光和红外图像这两个模态中，哪一个是主要模态。

2. 模态缺失的问题

大多数可见光与红外图像融合应用中假设可见光图像和红外图像可以被同时获得。然而，在现实中，由于隐私问题或者成本问题或者故障问题，常常会有模态缺失的情况出现。例如，在模型运行时只有可见光图像或者红外图像。在这种情况下，我们的模型需要有处理模态缺失的能力。

笔者的一种思路是通过模态转换来处理模态缺失的问题。具体思路如图12.14所示。假设我们只有可见光图像，我们可以通过图像到图像转换的方法，例如著名的 CycleGAN 算法 [269] 来实现从可见光图像到红外图像的转换。然后，我们可以将可见光和生成的伪红外图像进行融合，得到融合图像，然后运行下游应用。笔者通过实验证明了这种方法的确可以得到比单独使用可见光图像更好的效果。



图 12.14: 基于图像转换的方法来处理模态缺失的思路。

3. 提高应用的实时性

在目标跟踪中，跟踪精度和速度是两个最重要的性能指标。然而，当前绝大多数基于深度学习的融合跟踪方法，仅仅追求好的跟踪精度而忽略了速度方面的要求，因此将网络结构设计得非常复杂，导致算法跟踪速度很慢，缺乏实用性。

实时性不强的问题不仅仅存在于基于可见光和红外图像融合的目标跟踪中，也存在于基于可见光和红外图像融合的其他应用中。因此，如果在保持应用性能的前提下提升算法的实时性，是未来基于可见光和红外图像融合的应用领域需要重点解决的问题之一。

4. 基础模型的开发

目前的面向应用的可见光与红外图像融合算法基本都是针对特定应用开发的。例如，有专门应用于目标跟踪的算法，专门应用于行人检测的算法，也有专门应用于目标检测的算法，等等。然而，据笔者所知，当前很少有研究人员研究能够用于多种任务的可见光与红外图像融合基础模型。尽管目前已有一些多模态模型，但很少有针对可见光和红外图像的。在这种情况下，每针对一个新任务都要开发新模型，不仅效率低，而且造成浪费，并且对每一种任务都需要大量的训练数据。因此，笔者认为，未来的一个重要发展趋势，是开发可见光与红外图像融合的基础模型。这样一来，在面向新的任务的时候，仅需要在基础模型后面加上适当的层，并使用少数新的数据对模型进行训练，即可使用。此外，自监督的训练方法可以用来训练这些基础模型。

5. 面向应用的融合

本书第??章介绍过基于面向应用的可见光和红外图像融合。在此，笔者希望再次重复一下，开发面向应用的可见光和红外图像融合算法，是未来最重要的发展趋势之一，也是将可见光与红外图像融合技术从论文层面带到实际应用中的关键步骤。

6. 大规模应用数据集和标注

如前文所述，基于深度学习的可见光与红外图像融合跟踪方法是近年来，也是未来的发展趋势。然而，在基于深度学习的可见光与红外图像融合跟踪研究中，大规模融合跟踪数据集非常重要。这主要有两方面的原因。首先，深度学习模型的训练需要大规模的数据集，然而目前几乎所有的融合跟踪算法，均仍只采用可见光视频作为训练数据。其主要原因就是因为大规模融合跟踪数据集的缺乏。其次，为了合理地评价融合跟踪算法的性能，需要使用覆盖多个使用场景、多种挑战因素并严格配准的数据集。尽管目前已经有研究人员提出了相关数据集，如 GTOT, RGBT210 和 RGBT234，但不管在规模还是配准程度上，均还不够。

因此，本文作者认为，在可见光与红外融合跟踪领域的一项非常紧迫也非常重要的工作，就是建立大规模的融合跟踪数据集。具体地，这个数据集应当满足如下几点要求：

- 规模足够大。这个数据集应该具有足夠数量的可见光与红外图像对，并且

应当划分成训练集、验证集和测试集。训练集应当包含足够的图片数量，以便可以从头开始训练融合跟踪算法。

- 具有多样性。这个数据集应当包含多种多样的挑战因素，即那些在实际应用中可能遇到的各种挑战因素，例如遮挡、光照条件变化、低光照强度等。
- 具有标注信息。这个数据集应该提供标注信息，包括目标位置的标注信息和视频属性的标注信息。

7. 多目标融合跟踪

据本文作者所知，目前文献中所有通用目标融合跟踪算法均是单目标跟踪算法。这里的通用目标跟踪算法是指算法可以跟踪任意指定物体，而不是只跟踪某一个类别的物体。然而，在实际应用中，多目标跟踪相比较于单目标跟踪更具有实用价值。因此，本文作者认为，融合跟踪的一个重要发展方向将是多目标融合跟踪。然而，开展多目标融合跟踪较单目标融合跟踪更为困难。目前主要面临着的难点有数据集的缺乏，以及多目标融合跟踪中多个运动目标的特征的融合机制。

8. 更加智能和自动化的自适应特征融合方法

在前文介绍过的基于孪生网络的可见光红外融合跟踪算法中，笔者提出了自适应特征融合网络，并考虑了模态可靠性。尽管这个方法能够在很多情况下表示可见光和红外图像的可靠性，但是在这个方法中需要人为选定阈值，使用起来不是很方便。在未来，应当继续研究更加智能和自动化的自适应特征融合方法。例如，可以设计一个网络结构直接对模态权值进行学习。

12.7 小结

本章首先介绍了红外图像的一些应用，然后对可见光和红外图像融合的应用进行了介绍。具体地，笔者介绍了多个不同层次的可见光和红外图像融合的应用。最后，笔者对于可见光和红外图像融合的应用的未来发展趋势进行了展望。

第 13 章

多聚焦图像融合的应用

双剑合璧，为了好看，更为了有用。

——笔者

如本书第六章所述，目前几乎所有的关于多聚焦图像融合的研究都聚焦于生成处处都清晰的融合图像。研究人员们在发表相应的科研成果时，一方面是通过人眼来观察对比融合图像来对融合结果进行定性评价，一方面又通过本书第四章中介绍过的一些图像融合评价指标来对融合结果进行定量评价。然而，在这些关于多聚焦图像融合的研究中，多聚焦图像融合的应用极少被关注到。本章中，笔者将讨论多聚焦图像融合的应用。

13.1 多聚焦图像融合的应用概述

多聚焦图像融合的主要目的是生成一幅在各处都清晰的图像。一般认为，多聚焦图像融合可以用于克服相机成像时景深的限制，从而生成高质量的图像。因此，多聚焦图像融合被广泛用于提高摄影图像的质量中。这是目前绝大多数多聚焦图像融合算法在做的事情，如图13.1(a)所示。

多聚焦图像融合可以将只有部分清晰的多张图像融合成一张处处清晰的融合图像。基于融合图像来开展其他应用，将比基于原始图像来开展其他应用有更好的效果。因此，理论上，多聚焦图像融合技术对于很多应用都有促进作用，如图13.1(b)所示。因此，近年来，研究人员们也开始将多聚焦图像融合技术应用于不同的任务之中。总体来说，多聚焦图像融合的应用主要是通过给下游应用提供高质量的源图像来实现的。

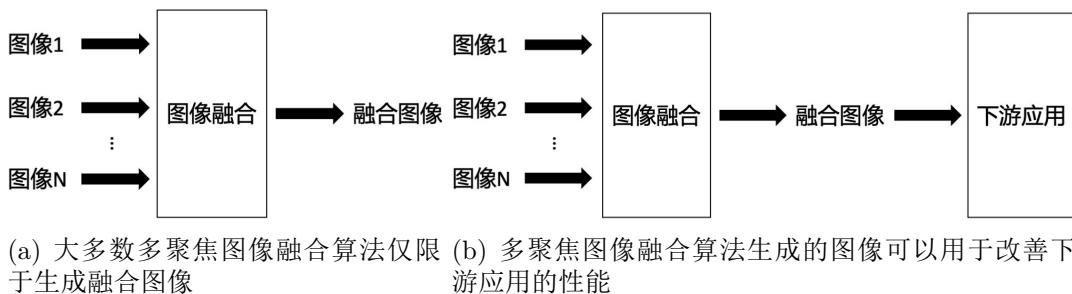


图 13.1: 以生成融合图像为目的的多聚焦图像融合和以改进下游应用为目的的多聚焦图像融合。前者的目的是生成高质量融合图像，后者的目的是通过生成融合图像来改进下游应用的性能。

当前，研究人员已经将多聚焦图像融合技术应用于帕金森症图像的分类 [270]、高倍血膜显微镜图像的融合 [271]、生物医学显微镜图像的融合 [272]、显微镜图像融合 [273]、视觉功率巡检 [274]、深度估计 [59]、人脸识别 [275] 等。图13.2中概括了几种主要的应用。

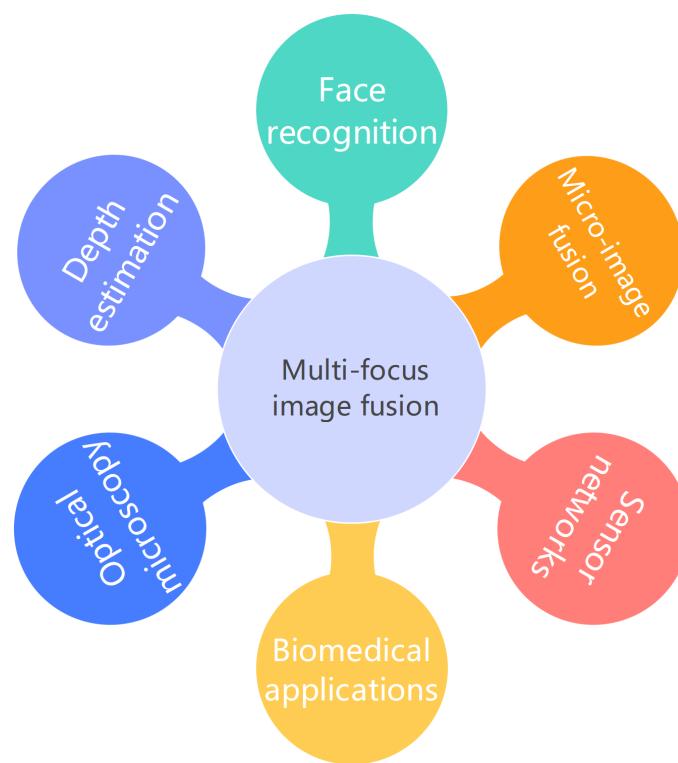


图 13.2: 多聚焦图像融合的主要应用。从图中可以看出，到目前为止，多聚焦图像融合已在不同的任务中得到了应用。

13.2 基于多聚焦图像融合的远距离人脸检测

远距离（10米到20米远）多张人脸检测在很多场景下都有需求，但是一个有挑战性的任务。这主要是因为传统的相机受景深的限制，难以产生处处清晰的图像（如图13.3所示），因为使得在不同距离的人脸的检测效果参差不齐。Raghavendra等人[276]证明了使用光场相机比使用传统相机可以获得更好的远距离人脸检测的效果。他们进一步研究了如何基于多聚焦图像融合来开展多张人脸的远距离检测。具体地，他们[275]将多张包含多张人脸的图像融合成一张处处清晰的图像，然后开展人脸识别。他们在论文中处理的是远距离（10米到20米远）人脸检测问题。

在Raghavendra等人的研究中，他们使用了光场相机。光场相机是一种特殊的相机，他们可以在一次拍摄中得到不同聚焦点（不同深度）的一批图像，如图13.3所示。他们的算法流程如图13.4所示。首先，他们使用光场相机获得同一场景的一系列不同聚焦点的图像，然后将这些图像融合得到一张处处清晰的图像。然后，他们基于这张清晰的图像来做人脸识别。如图??所示，他们的实验证明使用这种方法得到的远距离多张人脸的检测效果远好于使用传统相机得到的结果。



图 13.3: 使用 Lytro 光场相机拍摄得到的不同聚焦深度的图像（图 (a) 到图 (d)）。图 (e) 是融合 4 张源图像而得到的处处清晰的图像。图像来源于 [275]。

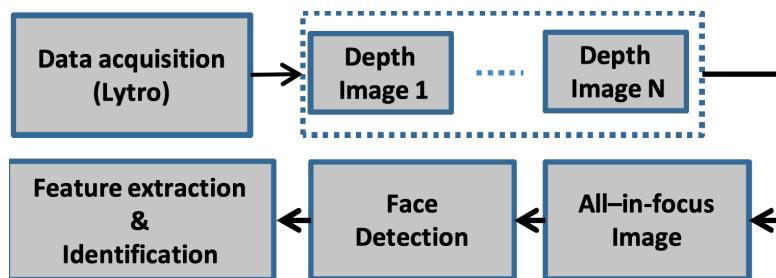


图 13.4: 基于多聚焦图像的人脸检测流程图。图像来源于 [275]。

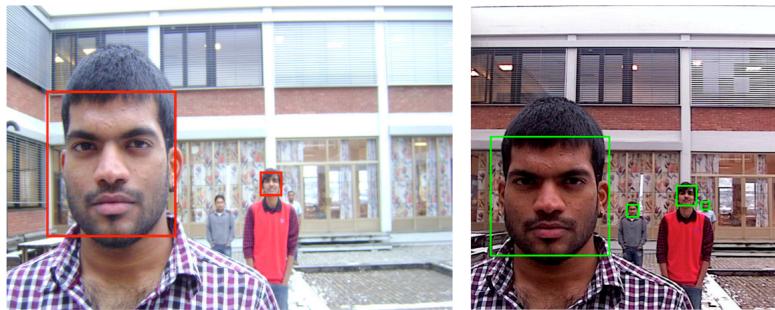


图 13.5: 基于多聚焦图像融合的远距离多人脸检测的效果比使用传统相机的人脸检测效果要好很多。左图: 基于传统相机拍摄的图像得到的人脸检测效果。右图: 基于多聚焦图像融合进行的人脸检测的效果。图像来源于 [275]。

13.3 基于多聚焦图像融合的光学显微图像融合

多聚焦图像融合的另外一个很重要也很常见的应用领域, 是光学显微图像融合。这是因为, 光学显微镜的放大倍数和景深是相互限制的关系。也就是说, 光学显微镜为了获取高分辨图像, 通常具有非常有限的景深。为了获取厚度大于景深的样本(例如血液样本)图像, 常常需要获取一些不同聚焦点的图像。为了便于医学检验人员观察或者便于人工智能系统进行自动处理, 需要将这一些不同聚集点的图像进行融合, 得到处处清晰的图像。这实际上类似于实现了扩展景深。

例如, 2022 年, 英国伦敦大学学院等单位的研究人员 [271] 对高放大倍数的血涂片图像进行了图像融合。这些图像聚焦点不一样, 因此这是一个多聚焦图像融合的任务。他们设计了一个基于卷积神经网络的方法来实现 7 张或者 3 张源图像的融合, 如图 13.6 所示。他们的实验表明, 他们得到的融合图像比基于小波变换的方法得到的融合图像更适合进行疟原虫检测。类似地, 中国科技大学和中国科学院的研究人员 [272] 提出了一种基于图像分割网络的多聚焦图像融合方法, 用于对 5 张¹不同聚焦点的病理学显微图像进行融合, 以得到处处清晰的图像。他们的方法示意图和融合结果分别如图 13.7 和图 13.8 所示。

需要指出的是, 光学显微图像融合的源图像一般有多张。这样很多方法主要针对两张源图像的融合是不一样的。另外, 这类显微图像融合的应用对图像融合算法的实时性一般要求比较高。

¹该论文作者通过实验发现, 对于癌细胞病理学图像来说, 5 张不同聚焦点的图像足以用来得到处处清晰的图像

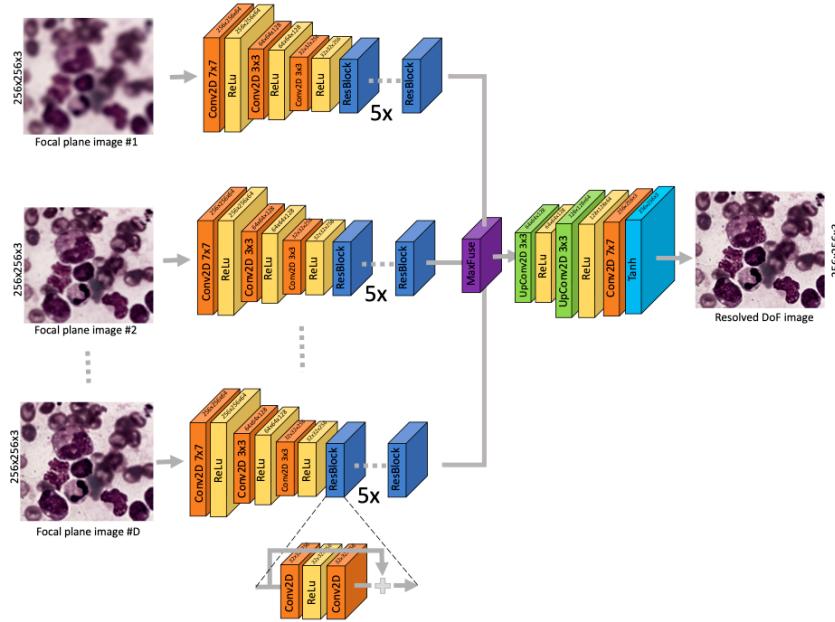


图 13.6: 多聚焦图像融合在血涂片显微图像融合上的应用。图像来源于 [271]。

13.4 基于多聚焦图像融合的深度估计

现在有很多基于深度估计算法可以基于单张 RGB 图像估计出深度信息。多聚焦图像融合算法由于可以从多张不同焦点的图像中融合产生清晰图像，因此对于这些深度估计算法的效果是有提升作用的。

2022 年，武汉大学马佳义教授团队在论文中用实验证明了多聚焦图像融合算法对于基于 RGB 图像的深度估计的性能有较大的提升作用 [59]，如图 13.9 所示。然而，值得指出的是，该研究中使用的数据集是人为产生的模拟数据集，并非真实世界中由相机拍摄产生的多聚焦源图像。要使得多聚焦图像融合对于深度估计真的有用，我们需要找到真实的应用场景，即的确需要产生多聚焦图像的地方。

此外，图像中的聚焦区域和非聚焦区域反应了物体和相机之间的距离信息，因此，可以从多聚焦图像中推测出深度信息。因此，可以基于多聚焦图像进行三维重建和深度估计。这二者分别被称为 shape from focus [277] 和 depth from focus [278]。

然而，需要指出的是，在基于多聚焦图像进行三维重建和深度估计的时候，并非一定要先生成融合图像，然后基于融合图像进行三维重建或者深度估计。相反，在许多基于多聚焦图像进行三维重建和深度估计的工作中，并不需要先生成融合图像。

2021 年，中国台湾清华大学的研究人员 [278] 提出了一种可以同时生成融

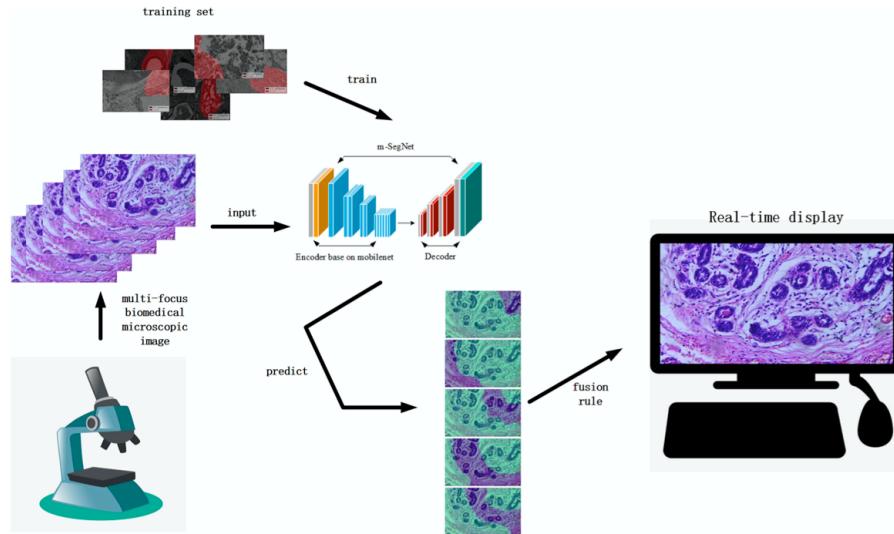


图 13.7: 多聚焦图像融合在病理学显微图像融合上的应用示意图。图像来源于 [272]。

合图像和深度信息的方法, 如图13.10所示。该方法既可以使用深度图的 ground truth 来做有监督训练, 也可以使用处处清晰的融合图像来做无监督的深度估计训练。另外, 因为使用了三维卷积, 所以该方法可以处理不同数量的源图像。

13.5 展望

本节对于多聚焦图像融合的应用研究进行一些展望。

(1) 开展更多应用

高质量的图像对于绝大多数计算机视觉任务都会有好处。因此, 理论上, 多聚焦图像融合对于许多下游任务都会有帮助。然而, 如前所述, 目前多聚焦图像融合的应用还局限于少数几个任务上。笔者认为, 未来有必要将多聚焦图像融合算法应用于更多的计算机视觉任务中。

(2) 开发面向应用的多聚焦图像融合数据集

为了将多聚焦图像融合应用于更多的任务中, 有必要开发面向应用的多聚焦图像融合数据集。然而, 需要注意的是, 当前很多的多聚焦图像融合数据集是通过人为添加模糊产生的模拟数据集, 而并不是在原始图像中存在模糊区域。我们不能为了融合而融合, 必须是在确有将多张不同焦点的图像进行融合的需求的

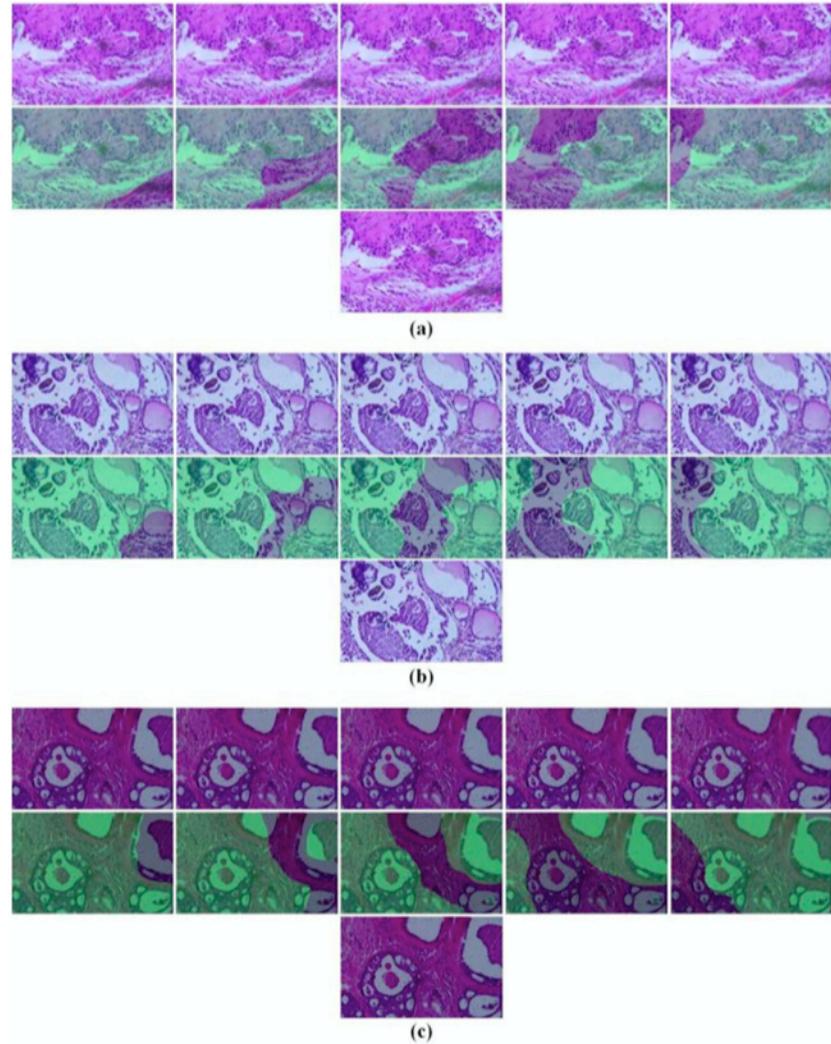


图 13.8: 多聚焦图像融合在病理学显微图像融合上的应用结果。(a) 肺癌的病理学显微图像 (b) 甲状腺乳头状癌的病理学显微图像 (c) 腺样囊性癌的病理学显微图像。每一个子图的第一行是不同聚焦点的源图像，第二行是 decision map，第三行是融合图像。图像来源于 [272]。

情况下（如本章前面介绍过的远距离多人脸检测的例子），才应用多聚焦图像融合算法，并制作相关数据集。

(3) 开发应用驱动的多聚焦图像融合算法

本书第十章介绍过应用驱动的图像融合算法。然而，目前基本上只有应用驱动的可见光与红外图像融合方法和医学图像融合方法。截止到本书写作为止，似乎还没有应用驱动的多聚集图像融合算法。在考虑多聚焦图像融合算法的应用时，应用驱动的多聚焦图像融合方法应当是一个重要的考虑方向。例如，在多聚焦显微图像融合中，可以在融合过程中将下游任务考虑在内。

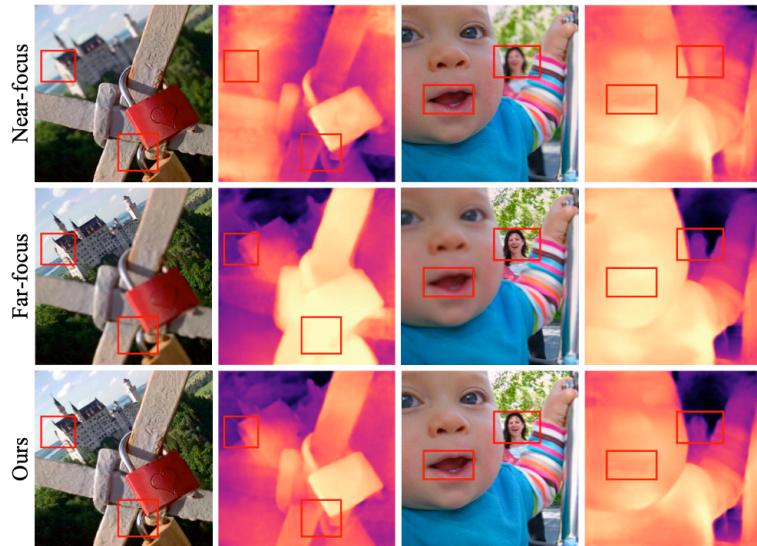


图 13.9: 多聚焦图像融合在基于 RGB 图像的深度估计中的作用。图中前两行为源图像和基于源图像的深度估计结果。第三行为融合图像和基于融合图像的融合结果。从图中可以看出，进行多聚焦图像融合以后，深度估计的效果更好了。图像来源于 [59]。

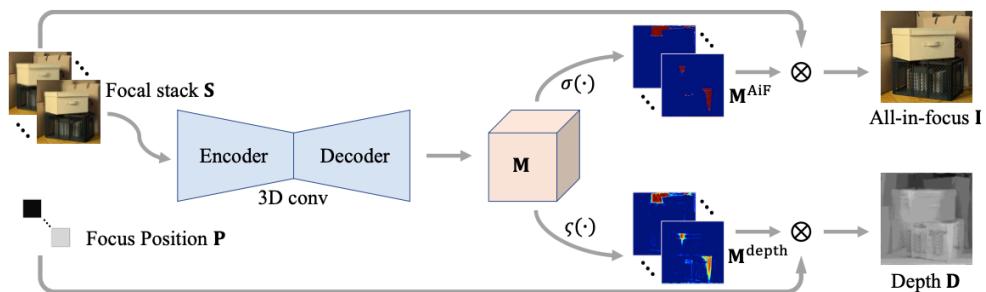


图 13.10: 可以同时生成深度图和融合图像的方法。该方法既可以使用深度图的 ground truth 来做有监督深度估计训练，也可以使用清晰的图像来做无监督深度估计训练。图像来源于 [278]。

13.6 小结

本章主要讨论多聚焦图像融合的应用。首先，我们对多聚焦图像融合的应用进行了总结，然后介绍了几个典型的例子。最后，我们的多聚焦图像融合的应用研究进行了展望。

第 14 章

多曝光图像融合的应用

多曝光图像融合旨在将两张或者多张不同曝光程度的图像融合成一张曝光程度良好的图像。本章主要介绍多曝光图像融合的应用。

14.1 多曝光图像融合的应用概述

总体来说，多曝光图像融合的主要应用在于提升图像的视觉效果，也就是将过曝光和欠曝光的图像进行融合以得到高质量的图像。这部分内容在本书第七章已详细介绍过，故本章不再赘述。除此以外，多曝光图像融合在其他方面的应用不是很多。这与可见光和红外图像融合以及多聚焦图像融合有些区别。

14.2 基于多曝光图像融合的语义分割

本节简要多曝光图像融合在自动驾驶语义分割中的一个应用。

在自动驾驶中，语义分割是很重要的一项基本技术，因为它可以帮助自动驾驶汽车更好地理解场景，例如障碍物和可行驶区域。目前研究人员已经开发了大量的基于视觉的语义分割方法。然而，这些方法通常容易受到光照条件变化的影响。为了克服这个问题，2022 年，厦门大学和英国埃塞克斯大学的研究人员使用了多曝光图像融合技术来提升自动驾驶中语义分割的性能 [279]。他们首先使用一个配准模块来对多曝光图像进行配准，以便得到配准以后的输入图像序列。然后，他们使用了一个神经网络模型来预测每张输入图像的权值图。最后，他们使用权值图和引导滤波技术得到融合图像。融合图像之后被用于进行语义分割。该方法的流程示意图如图 14.1 所示。

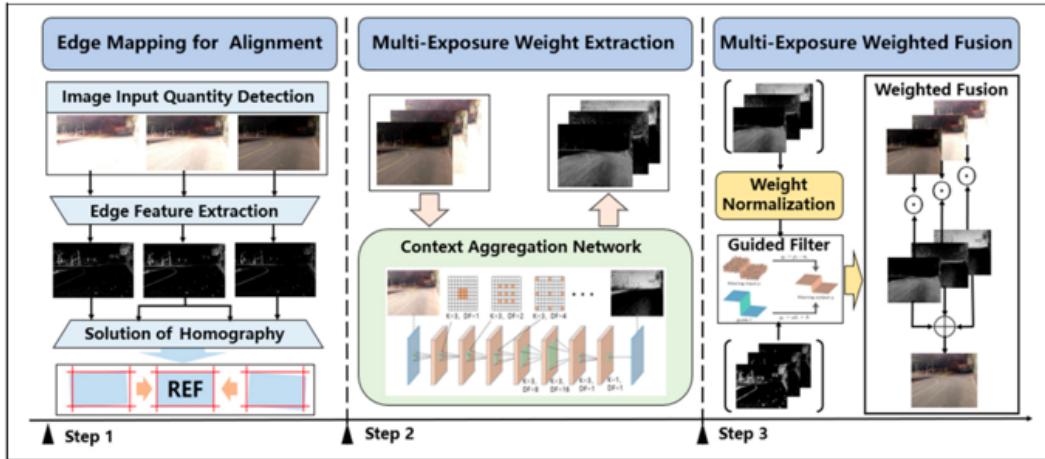


图 14.1: 基于多曝光图像融合的语义分割方法示意图。图像来源于 [279]。

需要说明的是，该团队不仅在已有数据集中测试了该方法的有效性，他们还收集了真实的实验数据来测试该方法。图14.2中展示了他们在厦门大学收集数据时所用的车辆和相机。在收集数据时，3个相机的曝光参数被设置成各不相同。该团队共收集了 1800 张图像。

该方法在实际收集到的图像上的效果如图14.3所示。从图中可以看出，基于该方法生成的融合图像的分割结果比基于 MEF-Net 和 GGIF 方法生成的融合图像的分割结果要好。

厦门大学和埃塞克斯大学的这项工作，是最早将多曝光图像融合应用于自动驾驶语义分割任务的工作。该工作证明了多曝光图像融合技术对语义分割任务有帮助作用。笔者认为，该工作会对后续工作产生很好的指导意义。



图 14.2: 实验设备示意图。图像来源于 [279]。

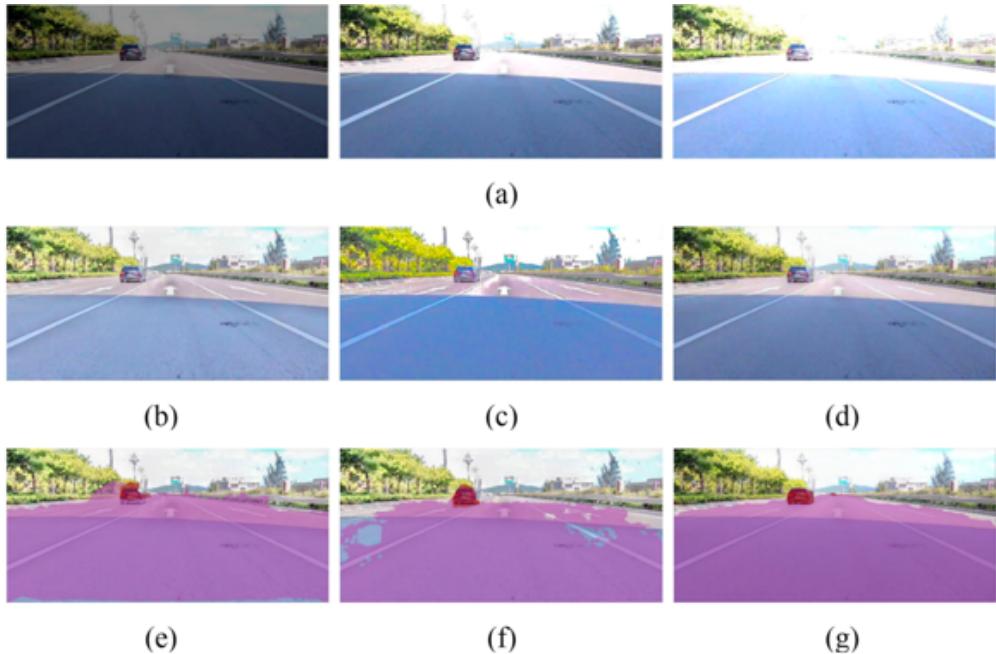


图 14.3: 结果示意图。(a) 输入图像。(b) 使用 MEF-Net 方法 [280] 得到的融合图像。(c) 使用 GGIF 方法 [281] 得到的融合图像。(d) 使用该团队开发的方法得到的融合图像。(e) 到 (f) 为基于 (b)-(d) 的融合图像产生的分割结果。图像来源于 [279]。

14.3 提升显微图像质量

多曝光图像融合已有的另外一个应用是提升显微图像的质量。本节将介绍一个典型例子。

硅藻是一类重要的微小藻类，它们在生态系统中扮演着关键角色。硅藻通过光合作用吸收二氧化碳并释放氧气，同时它们的硅基细胞壁在生物学和环境科学研究中具有重要意义。硅藻的多样性使其成为水质监测、环境变化指示以及石油勘探等领域的重要研究对象。通过显微图像分析，科学家们能够更精确地分类和识别不同种类的硅藻，从而更好地理解它们在生态系统中的作用和应用价值。

在显微图像分析中，传统的显微成像技术难以捕获具有复杂硅藻壳体和多尺度结构的硅藻种类的完整动态范围。这些硅藻壳体由硅基细胞壁构成，具有独特的几何图案和纳米级的细微结构。由于硅藻呈现出多种多样的形态和结构，明场和暗场显微镜可以捕获多张具有重要和有价值信息的图像。为了提取更多的细节，研究人员通常需要在不同曝光设置下拍摄一系列照片，如图14.4所示。这是因为单一曝光图像往往难以同时捕获明亮区域和暗弱区域的细节，这会导致一些区域过度曝光或曝光不足。在这种情况下，图像融合的概念就派上了用场。

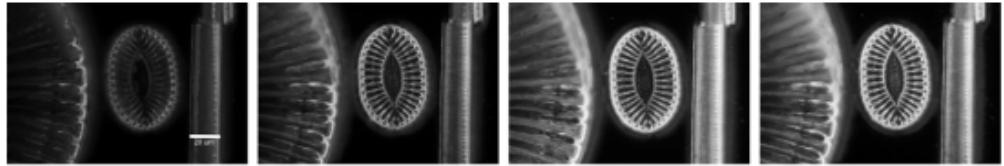


图 14.4: 曝光时间不同的新巧弯盘藻 (*Campylodiscus neofastuosus*) 源图像示意图。从左往右，曝光时间分别为 50 毫秒，150 毫秒，250 毫秒和 300 毫秒。图像来源于 [282]。

多曝光图像融合技术的动机在于克服标准数码相机在捕捉高动态范围场景或标本细节时的限制。通过结合不同曝光设置下的图像，多曝光图像融合能够生成一幅融合图像，既保留了亮区的细节，又展现了暗区的特征。具体到硅藻的研究中，硅藻的透明细胞壁在显微镜下的不同曝光图像中可能表现出过度或不足的曝光。使用多曝光图像融合技术，可以在传统显示设备上直接呈现细节丰富且曝光均匀的融合图像，从而更好地观察和分析这些微观结构。

哈宾德·辛格等在其 2022 年的论文《基于多曝光显微图像融合的细节增强算法》[282] 中提出了一种新的细胞区域敏感曝光融合 (CS-EF) 方法。该方法利用局部信息测量来选择输入曝光中的良好曝光区域，并引入了改进的直方图均衡化技术，以在融合之前提高输入多曝光图像的均匀性。

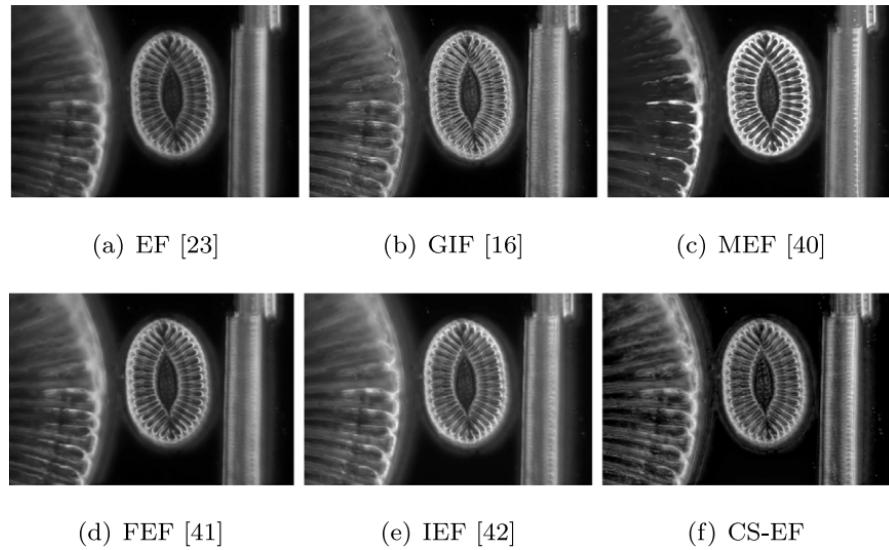


图 14.5: 使用新巧弯盘藻暗场显微镜图像对比五种现有图像融合算法和 CS-EF 方法。(a-f) 分别为：原始图像、EF（曝光融合）方法、GIF（引导图像滤波）方法、MEF（基于抠图的曝光融合）方法、FEF（快速曝光融合）方法、IEF（基于照明估计的曝光融合）方法和 CS-EF（细胞区域敏感曝光融合）方法。图像来源于 [282]。

该方法在 MATLAB 上实现，并在不同的显微图像数据集上进行了测试，如

图14.5所示。该方法也可以应用于彩色图像，如图14.6所示。结果表明，CS-EF方法能够在融合图像中显著提升细节，并在保持颜色信息的同时有效减少背景噪声。这一研究不仅验证了CS-EF在细节保留和噪声抑制方面的优越性，还为多曝光显微图像的融合提供了新的思路和方法。

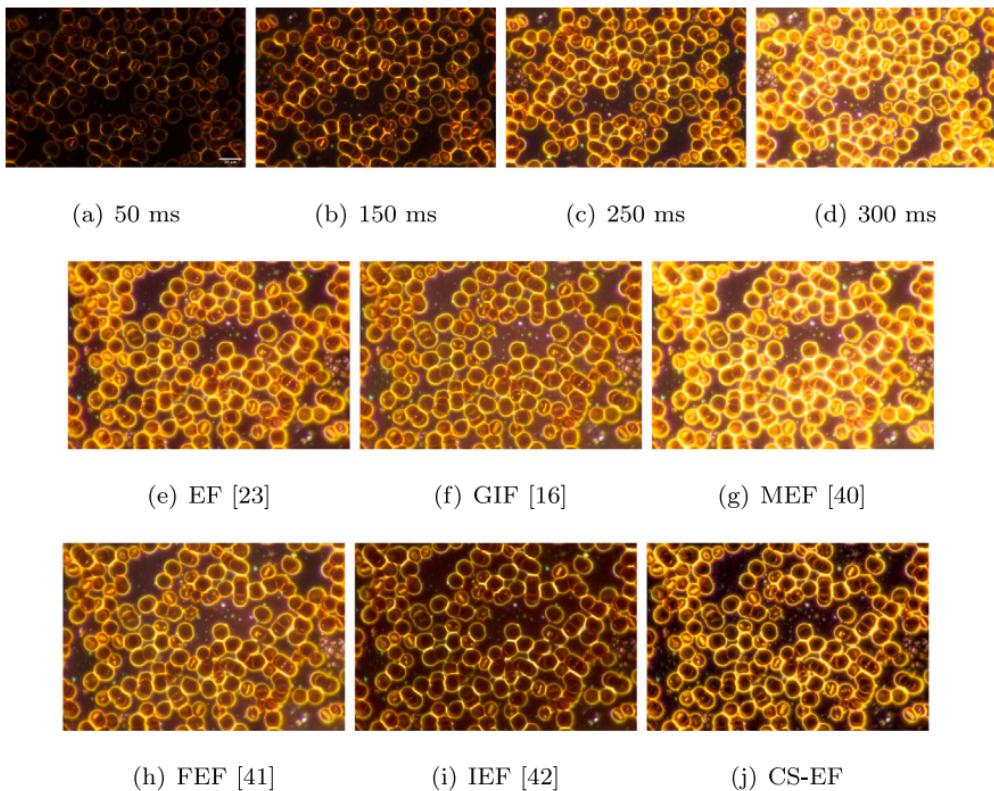


图 14.6: 扩展到彩色图像数据集。(a-d) 四张多曝光输入图像; (e-j) 使用暗场显微镜拍摄的 RAT-blood 图像, 对比五种现有图像融合算法和 CS-EF 方法。(e-i) 分别为: EF (曝光融合) 方法、GIF (引导图像滤波) 方法、MEF (基于抠图的曝光融合) 方法、FEF (快速曝光融合) 方法、IEF (基于照明估计的曝光融合) 方法; (j) CS-EF (细胞区域敏感曝光融合) 方法。图像来源于 [282]。

此外,值得注意的是,该团队还曾同时使用多聚焦图像融合技术和多曝光图像融合技术来提升显微图像的性能 [283]。

14.4 小结

本章主要介绍了多曝光图像融合的应用。总体来说,多曝光图像融合目前主要应用于提升图像的质量,以后进行后续分析和处理,例如语义分割。多曝光图像融合这项技术是很有价值的,其更多应用还有待研究人员挖掘。笔者相信,在未来一定会出现更多的基于多曝光图像融合的应用。

第 15 章

图像融合的前沿进展

本章将对图像融合领域的一些前沿进展进行总结，以供读者朋友参考。

15.1 与其他任务相结合

自从图像融合这个研究领域出现以来，绝大部分的像素级图像融合任务与其他任务是相互独立的。也就是说，这些图像融合方法的结果仅仅是生成融合图像，并不涉及其他任务。近年来，有些研究人员开始将图像融合任务与其他任务进行结合，取得了不错的效果。例如，昆明大学李华锋副教授与其他研究人员合作，将医学图像融合、去噪和图像质量提高进行了结合 [284]。李华锋副教授还将多种图像融合任务与图像超分辨率进行过结合 [285]。另外，北京航空航天大学邓欣副教授将多曝光图像融合和图像超分辨率相结合 [215]，北京航空航天大学 Li 等人将图像融合与去噪相结合 [286]，均取得了不错的成绩。

将图像融合与其他任务相结合，既可以建立图像融合与其他研究领域的联系、扩大图像融合的研究范围，也可以使图像融合技术得到更广泛的应用。此外，由于一个模型可以同时实现两个或多个不同的任务，因此模型的利用效率比较高。基于这些原因，与其他任务相结合的图像融合方法，成为了图像融合领域的一个热门的研究方向。

15.2 通用图像融合方法

本书第九章对通用图像融合方法进行过介绍。尽管目前的通用图像融合方法尚存在一些缺点，但是从近几年的文献中可以发现，基于深度学习的通用图像

融合方法的性能越来越好，并且越来越多的研究人员在着手开发基于深度学习的通用图像融合方法。例如，武汉大学马佳义教授团队在近一两年就提出了数个通用图像融合方法，如基于 Transformer 的 SwinFusion [59]、基于连续学习和生成对抗网络的 UIFGAN[287]。江南大学吴小俊教授团队也提出了 UNIFusion[152]。通用图像融合方法实际上已成为当前图像融合领域的一个热点研究方向。

15.3 关于评价基准的研究

图像融合领域近年来的一个热点研究方向是关于评价基准的研究。这主要是因为在图像融合任务中没有真正的标准答案（请参考本书第四章），因此性能评价是一个困难的问题。因此，多年以来，图像融合领域没有评价基准。这个计算机视觉的很多其他领域，例如目标跟踪、目标检测、图像分割，有着很大的不同。

近年来，笔者意识到了图像融合领域对评价基准的急切需求，从 2020 年开始先后开发了可见光和红外图像融合领域的第一个评价基准 VIFB、多曝光图像融合领域的第一个评价基准 MEFB，以及多聚焦图像融合领域的第一个评价基准 MFIFB。这些评价基准，引起了图像融合领域众多研究人员的关注，并分别被多个研究人员在论文中使用。关于图像融合评价基准的研究，已然成为图像融合领域的一个热点研究方向。

15.4 基于具体应用的融合方法性能评价

在前文中，笔者对图像融合方法的性能评价方法进行过介绍。从前文可知，在过去的研究中，对图像融合方法的定量评价是通过一系列图像融合评价指标进行的。然而，因为没有标准答案的存在，这些评价指标很难全面、客观地反应图像融合算法的性能。有鉴于此，部分学者对如何更好地对图像融合方法进行性能评价展开了研究。

这两年此方面的一个前沿进展是基于具体应用来对图像融合方法进行评价。也就是说，在评价图像融合方法的性能时，不仅仅看生成的融合图像如何，还要看生成的图像对于下游任务具有多大的促进作用，如图 15.1 所示。这样做的一个很重要的好处是将图像融合这个没有标准答案的任务的性能评价，转换为具有标准答案的下游任务的性能评价。这是因为，目标跟踪、检测、场景分割等下游

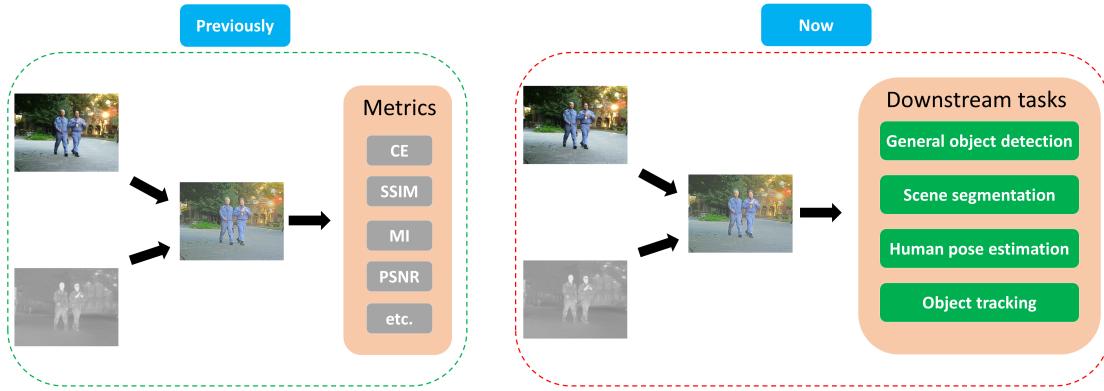


图 15.1: 图像融合性能评价方式的转换。以前, 图像融合的定量评价是通过一系列的图像融合评价指标来完成的。当前, 有学者提出应当通过下游任务的性能来评价图像融合方法的性能。

任务通常具有标准答案, 因而很容易进行客观的性能评价。

例如, 武汉大学马佳义教授团队在最近的一些研究中 [40, 59] 使用场景分割和目标检测任务对一些图像融合算法进行了对比。大连理工大学樊鑫教授团队提出了新的基于可见光和红外图像的目标检测数据集, 并使用目标检测任务来比较了一些图像融合算法 [75]。

15.5 将图像配准和图像融合进行结合

图像配准是一个被研究了很多年的领域。不过近年来出现的一个新趋势, 是将图像配准和图像融合进行结合, 一起研究。这个趋势主要体现在可见光与红外图像融合方面。

由于可见光和红外图像的成像机制以及可见光和红外相机的不同参数, 准确对齐可见光和红外图像对是非常困难的。研究人员已经提出了许多方法来进行可见光和红外图像的配准 [144, 145, 146]。然而, 几乎所有这些研究都没有考虑图像融合任务。为了解决这个问题, 一些研究人员开始同时学习图像融合和配准 [147, 148]。例如, 王等人 [147] 首先使用跨模态感知风格转移网络生成伪红外图像。然后, 他们学习真实红外图像与伪红外图像之间的位移矢量场, 这是一个较简单的单模态配准问题。学习得到的位移矢量场被用于重建注册后的真实红外图像。最后, 他们使用可见光图像和配准后的红外图像通过双路径交互融合网络进行图像融合。他们设计了包括风格转移损失、交叉正则化损失、配准损失和图像融合损失的损失函数来指导模型训练。此外, 武汉大学徐涵等人 [148] 提出了



图 15.2: 可见光图像 (左)、远红外图像 (中)、近红外图像 (右)。图像来源于 [249]。

RFNet。这是一个相互增强的框架，可以同时学习图像融合和配准。具体而言，RFNet 利用图像融合为图像配准提供反馈。然而，尽管这个想法很有启发性，但 RFNet 是为可见光图像和近红外图像的配准而设计的，而不是热红外图像。

15.6 其他类型的图像融合

除了本书中介绍的可见光与红外图像融合、多聚焦图像融合和多曝光图像融合以外，另外两种常见的图像融合任务是医学图像融合和遥感图像融合¹。到目前为止，这 5 种图像融合任务是图像融合领域里最常见的，也是最受研究人员关注的。除了这 5 种图像融合任务以外，近年来还出现了一些新型的图像融合任务。本节将对这些新出现的图像融合任务进行简单介绍。

15.6.1 可见光图像与近红外图像融合

笔者在前文中介绍的可见光与红外图像融合中的红外图像，指的是由波长范围在 8 到 15 微米之内²的远红外光成的像，又称热红外图像。除了热红外图像以外，研究人员也研究过另外一种红外图像和可见光图像的融合问题，即近红外图像(Near-infrared image, NIR)。近红外图像的由波长范围在 0.8 到 2.0 微米之内的近红外光成的像。

由于波长不同，近红外图像与远红外图像具备不同的特点。近红外图像的主要特点是可以提供一些更清楚的细节信息。图15.2中展示了可见光图像、远红外和近红外图像对比的一个例子。

从上述分析可知，近红外图像也具备一些和可见光图像互补的特点。因此，将可见光图像与近红外图像进行融合，也可以得到更好的图像，并在一些领域中具有应用价值。到目前为止，国内外已经有不少研究人员开展了可见光与近红外

¹由于篇幅限制和笔者研究领域的原因，本书未对医学图像融合和遥感图像融合进行介绍

²关于远红外和近红外的波长范围，不同文献的说法略有差异

图像融合的研究 [288, 289, 290, 291, 292, 293, 294, 295]，并在许多领域有了应用，例如图像去噪 [296]、远距离人脸识别 [297]、视频监控 [298, 299]、人体姿态估计 [300, 301]、人脸识别 [302]、图像去雾 [303]、水果检测 [304]。由此可见，可见光与近红外图像融合也是一个很有应用前景的研究方向，在未来或许会吸引更多的研究人员参与进来。

15.6.2 偏振图像融合

另外一种在近年来较受关注的新型图像融合任务，是偏振图像融合 [305, 306, 307]。偏振图像融合旨在将强度图像和线偏振度图像融合成一个具有更多细节的图像。这个有更多细节的图像有助于提高在复杂背景下目标检测的能力。张等人提出的偏振图像融合算法如图15.3所示。偏振图像融合也是一种像素级图像融合。

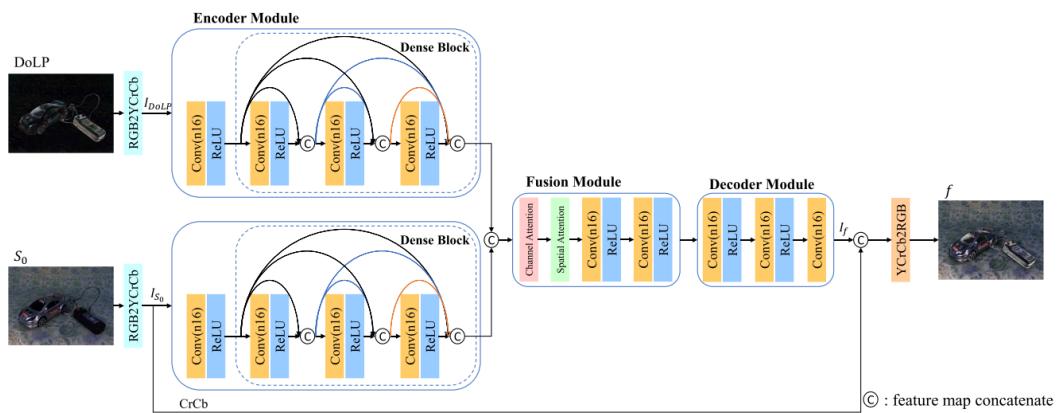


图 15.3：武汉大学马佳义教授团队提出的偏振图像融合算法 PAPIF。图像来源于 [306]。

15.6.3 RGB 图像融合

2022 年，有研究人员 [308] 在国际顶级学术会议 ECCV 上发表了 FusionVAE 算法，用于对多张 RGB 图像进行融合。该算法的输入是少于等于 3 张的 RGB 图像，其中每一张 RGB 图像都被遮挡了不同部分。该算法通过一个变分自编码器来获得一幅融合的图像。值得指出的是，该融合图像是在从训练集获得的先验知识的基础上产生的。该算法的示意图如图15.4所示。从图中可以看出，这种类型的 RGB 图像融合旨在生成高质量的融合图像，属于像素级融合。

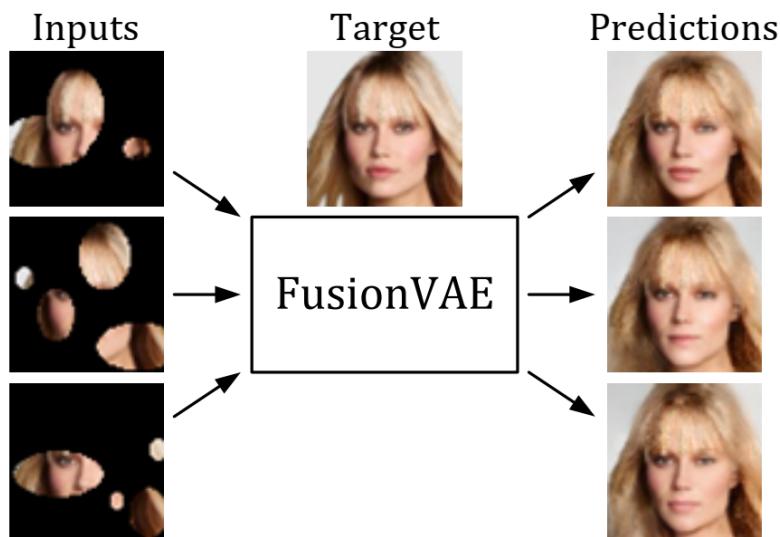


图 15.4: FusionVAE 算法示意图。图像来源于 [308]。

15.6.4 RGB 图像和 Optimal Waveband 图像融合

孙等人 [309] 于 2023 年提出了一种将 RGB 图像和 Optimal Waveband 图像进行融合，以得到高质量水下图像的方法。这是另外一种新型的图像融合任务。

15.6.5 可见光图像与深度图像融合

近年来，另外一种受到了很多关注的图像融合是可见光与深度图的融合 (RGBD)。我们的世界是 3D 的。深度信息的引入，使得很多 3D 应用有了可能，也为很多应用中的遮挡处理提供了方便。这个研究方向的发展也受到了深度相机的促进，例如 Realsense L515 和 Microsoft Kinect。图 15.5 展示了一个可见光图像和深度图像对比的例子。

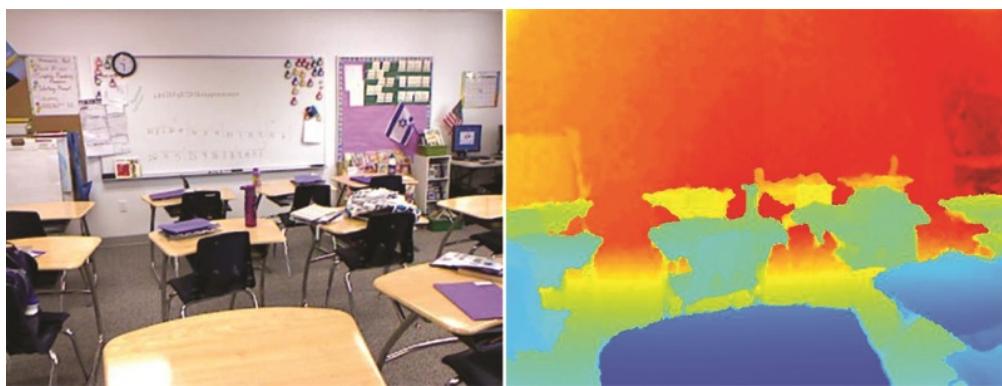


图 15.5: RGB 图像和深度图像示例。图像来源于 NYU Depth Dataset V2[310]。

目前已经有了不少基于可见光和深度图像融合的研究，涉及到了 3D 目标检

测 [311]、三维重建 [312]、SLAM [313]、显著性检测 [314, 315, 265]、目标跟踪 [316, 317]、场景分割 [318, 319]、人体姿态估计 [320, 321]、机器人 [322, 323] 等诸多应用场景。可见，可见光和深度图像的融合已经成为很活跃的一个研究方向。

值得注意的是，目前大多数 **RGBD** 图像融合方法是特征级或者决策级的方法，而不是像素级的方法。这与可见光与红外图像融合很不相同。实际上，在 **RGBD** 图像融合的研究中，研究人员更加关注的是应用的性能，而不注重生成的融合图像的视觉效果。

15.6.6 可见光图像与事件相机数据融合

事件相机是一种新型的相机，其工作原理与传统相机截然不同并且具有传统相机所不具备的一些优点，例如低延时和高时间分辨率。因此，将传统相机和事件相机结合起来使用以结合二者的优点，是一个很有价值的方向。

值得注意的是，可见光相机产生的图像与事件相机产生的“图像”截然不同（如图15.6所示）。因此，可见光图像和事件相机图像的结合主要是在特征级进行操作的，而不是在像素级。当然，目前也有一些研究人员在尝试首先将事件相机的图像转换为传统图像的格式，然后再与可见光图像进行融合使用。

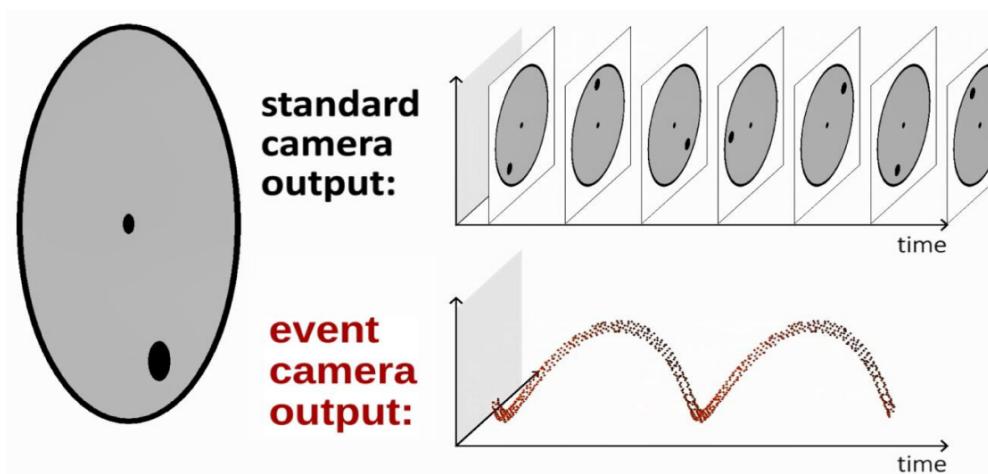


图 15.6: 可见光图像和事件相机“图像”示例。图像来源于苏黎世联邦理工学院研究员 Davide Scaramuzza 的事件相机讲义。

目前，在这个方向上已经有了一些初步的工作。例如，研究人员将可见光图像和事件相机的“图像”结合起来做目标跟踪，大幅提升了在目标快速运动情况下的跟踪性能 [324]。也有研究人员将可见光图像和事件相机图像结合起来做高动态状态下的深度估计 [325]。还有研究人员将可见光图像和事件相机图像进行

融合进行恶劣天气条件下的目标检测 [326]。此外，可见光与事件相机融合也被用于语义分割 [327, 328]、机器人抓取 [329] 和其他一些应用中 [330]。

15.6.7 多视角图像融合

多视角图像融合是指将从多个不同视角捕获的图像进行融合的过程。一般来说，多视角图像融合的目的不是生成一幅新的视觉效果更好的图像，而是为了得到更好的特征从而提升下游应用的性能。因此，多视角图像融合一般也是在特征级进行的。

到目前为止，多视角图像融合已经被用于多个不同的应用中。例如，Pilikos 等人通过对从两个不同视觉得到的超声图像进行融合，得到了更好的分割结果 [331]。Riera 等人 [332] 将多视角图像融合应用于大豆的产量估计中。他们使用机器人在田间从不同视角拍摄图像，然后将这些图像输入一个深度学习模型进行大豆荚果计数。王等人 [333] 将从不同视角拍摄的深度图像进行特征级融合，从而实现更好的 3D 动作识别。

15.7 其他应用

Xingchen To do: 完成这个

建筑物表面损伤检测：我审稿的文章

15.8 小结

本章对图像融合的其他应用和前沿进展进行了介绍。具体地，本章介绍了可见光与红外图像融合、多聚焦图像融合和多曝光图像融合的一些应用。此外，本章还介绍了目前图像融合领域的一些前沿进展，包括与其他任务相结合、基于具体应用的融合方法性能评价以及一些新型的图像融合任务。从本章的内容可以看出，图像融合可以被用于各种应用中。各种新型应用和新型图像融合任务还待广大研究人员来挖掘研究。

第 16 章

总结与展望

路漫漫其修远兮，吾将上下而求索。

——屈原

16.1 总结

作为第一本专门讨论基于深度学习的图像融合方法的著作，本书重点介绍了近年来蓬勃发展的基于深度学习的图像融合方法及相关应用。基于笔者的研究经历和科研实践，本书重点介绍了可见光与红外图像融合、多聚焦图像融合和多曝光图像融合，并介绍了它们的一些相关应用。此外，本书还简单介绍了一些近年来新出现的图像融合任务。值得指出的是，本书所介绍的绝大部分成果是近 5 年发表的，因此具有很好的前沿性。

基于本书介绍的相关内容，我们可以发现，图像融合作为一个已经存在了多年的研究领域，正在吸引着国内外越来越多的研究人员参与其中，并且研究范围在不断扩展。这其中的一个重要原因是图像融合是一个很有前景的研究领域，有着很多实际应用。此外，一个非常重要的促进因素是深度学习方法的引入。深度学习方法的引入，不仅使得图像融合任务本身的方法变得多种多样，同时也使得基于图像融合的各种应用的实现方法变得多种多样。

当然，虽然已有大量的基于深度学习的图像融合论文被发表，并且基于深度学习的图像融合方法已经成为了主流，但是基于深度学习的图像融合仍处于蓬勃发展时期，仍有许多问题有待探索。笔者相信，在今后几年，基于深度学习的图像融合必将引起更多的关注，也会有更多优秀的科研成果涌现出来。

16.2 待解决的问题

虽然图像融合领域已经发展了几十年，并且取得了长足的进步，但是仍然存在一些问题有待解决。本节对图像融合领域存在的一些问题进行一些总结。

1. 缺乏更好的评价指标

笔者在前文中已经介绍过，图像融合领域目前缺乏公认的很好的评价指标。这给图像算法的定量评价带来了很大的困难。这也是目前研究人员在图像融合的论文中常常选用不同评价指标的重要原因之一。然而，为了使得图像融合领域能够得到更好的发展，设计出更好的性能评价指标是必须要做的事情。

笔者认为，好的评价指标需要具备以下几个特点：

- 和基于视觉效果的定性评价基本一致。目前很多评价指标与融合图像的视觉效果是不一致甚至是矛盾的。这给图像融合算法的进行性能造成了很大的困难。
- 泛化性好。现在的图像融合算法常常要在几个数据集上进行测试。因此，图像融合评价指标的泛化性要好。也就是说，该指标要能在多个数据集上对图像融合算法进行客观评价。
- 易于计算。现在的图像融合算法常常要在大量的测试图像上进行测试。尤其是在现在的论文中，研究人员往往要对十几种甚至几十种图像融合算法进行比较。如果评价指标的计算量太大，那么开展图像融合的定量评价就是一件较为耗时的事情。

2. 缺乏合适且客观的性能评价方法

笔者在前文中介绍不同类型的图像融合任务时已经指出过，由于在图像融合中一般没有标准融合图像（标签）存在，因此想像在目标检测和目标跟踪任务中那样客观地评价图像融合算法的性能非常困难。遗憾地是，这种现象并没有随着基于深度学习的图像融合技术的发展而得到缓解。缺乏合适且客观的性能评价方法导致的一个直接结果就是我们很难知道一个算法的真实性能，这对于图像融合领域的发展和图像融合算法的实际应用都是不利的。

3. 脱离实际应用

图像融合是一个实用性很强的领域。实际上，在图像融合的两个目的（见第1.3小节）中，第二个目的更为重要，因为图像融合的初衷就是为了克服传感器的局限，以提高各种实际应用（如目标跟踪、检测）等的性能。然而，目前相当一部分的图像融合研究并没有考虑到实际的应用。这种现象在像素级图像融合中表现得尤为明显。以可见光红外图像融合为例，绝大部分图像融合算法的目的仅仅是生成质量更好的融合图像。然而，一方面，这些算法在设计过程中并没有考虑到后续应用的需求；另一方面，绝大多数图像融合算法对实际应用的益处也没有通过实验进行验证。因此，这些算法对于实际应用是否真的有帮助以及有多大程度的帮助，是一个未知数。

另外一个典型例子是医学图像融合。据笔者所知，目前绝大部分的医学图像融合论文都没有考虑实际应用。大多数论文都是基于现有的少量公开数据集进行一些实验，然后通过一些定性分析（通过人眼观察）和定量分析（通过相关评价指标来进行）来判断融合图像的质量。然而，这些融合图像是否对医生有帮助并不清楚，因为整个算法的评价过程中没有医生参与进来，也没有考虑医生的临床需求。

综上所述，目前绝大多数图像融合的研究还停留在学术研究阶段，在实际中的应用较少，并且在算法研发过程中对于后续应用的考虑也很少。当然，近年来也有一些研究人员开始关注图像融合对实际应用的帮助，但是还远远不够。

4. 在国际上受重视不够

到本书写作为止，图像融合领域在国际上受到的重视程度还不是很够。以可见光红外图像融合为例，图16.1对于从2018年到2022年9月底之间发表的基于深度学习的可见光与红外图像融合相关的论文的发表机构进行了统计分析。从图中可以看出，在91%的论文中有国内研究机构的参与。其中77.7%的论文完全由国内的研究机构完成。仅有9%的论文是仅由国外研究机构完成的。从上述分析可以看出，目前可见光与红外图像融合的研究在国际上受到的重视程度不太够，参与进来的国外科研机构不多。此外，以笔者所在的英国埃克塞特大学和帝国理工学院为例，除了笔者之外，从事图像融合研究的科研人员非常少。

图16.2进一步分析了参与可见光与红外图像融合的国外科研机构的分布情况。从图中可以看出，国外对于图像融合的研究主要集中于英国、印度、韩国、美国和加拿大。其他国家的科研机构参与得非常少。这也可以在一定程度上说

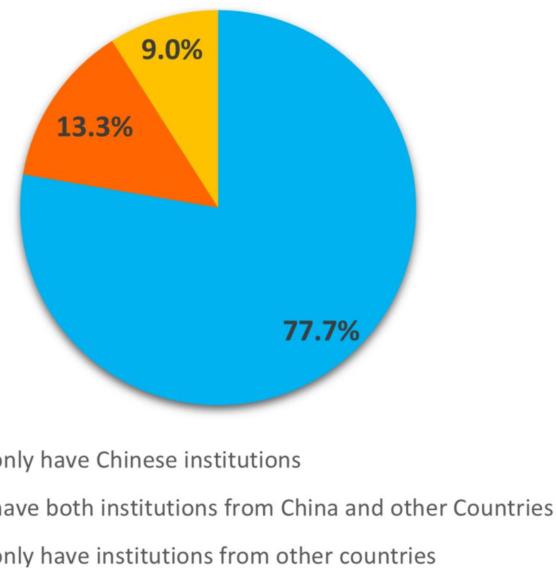


图 16.1: 基于深度学习的可见光与红外图像融合论文发表机构统计分析 (2018 年至 2022 年 9 月底)。

明，图像融合的研究在国际上受到的重视程度不是太够。

图像融合领域在国际上受到的重视程度不够有多方面的原因。要想提高图像融合领域在国际上的知名度，需要广大研究人员多产生高质量的科研成果，尤其是有实用价值的科研成果，来证明这个领域真正有研究价值。目前来看，仍然任重而道远。

笔者正在尽自己的绵薄之力在国际上推广宣传图像融合这个领域。2024 年 11 月，笔者作为主要组织者在国际知名学术会议——英国机器视觉会议上组织了一个基于深度学习的图像融合研讨会。一方面，希望可以给国内外的相关研究人员一个讨论的平台。另一方面，笔者也希望该研讨会可以使得国际上的研究人员对图像融合这个领域有更多的关注。

16.3 展望

如前所述，在图像融合领域仍有很多问题有待研究人员去探索。根据笔者的研究与总结，本节对图像融合领域未来的发展进行一些展望，供读者朋友们参考。

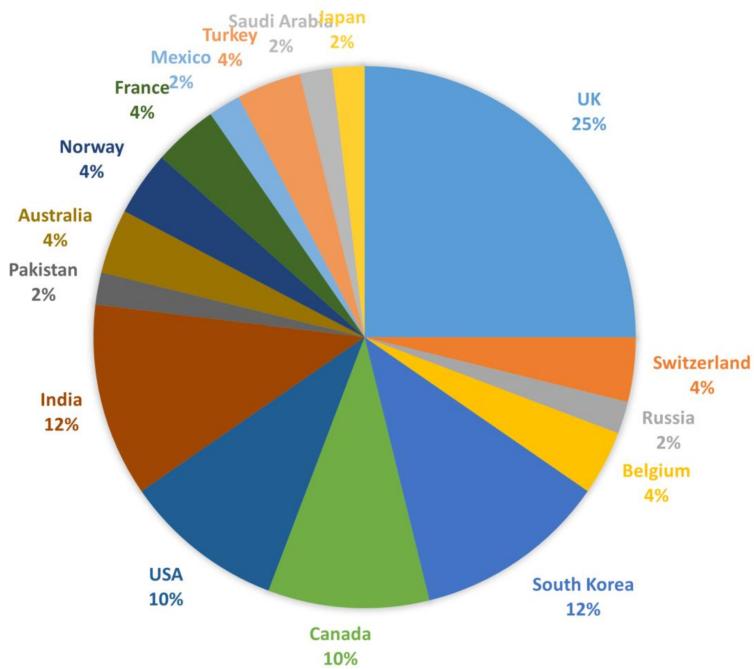


图 16.2: 基于深度学习的可见光与红外图像融合论文发表机构统计分析 (2018 年至 2022 年 9 月底)。

1. 会有更多的基于深度学习的工作发表

近几年来，基于深度学习的图像融合方法已经成为图像融合领域的主流方法。毫无疑问，未来会有更多的基于深度学习的图像融合方法发表。特别地，一些新的深度学习方法和模型，会被逐渐引入图像融合领域。例如，2019 年，武汉大学马佳义教授团队将生成式对抗网络引入了图像融合领域。2021 年，研究人员将变换器引入了图像融合领域。2023 年，马佳义教授团队将扩散模型引入了图像融合领域。

2. 与多模态机器学习建立更加紧密的联系

图像融合领域的可见光与红外图像融合和医学图像融合，属于多模态图像融合。基于深度学习的多模态图像融合，与当前非常热门的多模态机器学习是很相关的。多模态机器学习吸引了很多研究人员参与，取得了很好的研究成果。笔者认为，多模态机器学习的一些研究成果，对于图像融合领域的发展是有很好的促进作用和借鉴价值的。多模态机器学习最权威的研究团队之一是美国卡耐基梅隆大学 Louis-Philippe Morency 教授的研究团队。关于多模态机器学习的内容，可以参考卡耐基梅隆大学的《多模态机器学习》课程。

3. 联系实际任务，重视实际应用

如前文所述，绝大多数图像融合（像素级）方法在设计的过程中并没有考虑实际应用的需求。此外，绝大多数图像融合方法也没有通过实验证明是其是否能改善应用的性能。因此，笔者认为，在未来的图像融合研究中，研究人员们需要在这方面进行考虑，联系实际任务，重视实际应用，以便图像融合技术可以真正地得到应用、创造实际价值。

值得说明的是，目前已有一些学者意识到了这个问题并进行了一些初步研究。例如，江南大学吴小俊教授团队的研究人员在他们的一些论文中研究了该团队研发的可见光红外图像融合对于目标跟踪任务的促进作用。笔者也初步做了一些相关的工作来探索图像融合算法对于目标跟踪的促进作用和影响。

此外，笔者认为，未来会有越来越多的图像融合应用出现。

4. 应用驱动的图像融合方法会成为主流

笔者在前文介绍了应用驱动的图像融合方法。考虑到图像融合研究的更重要的目的是提升下游应用的效果，笔者认为应用驱动的图像融合方法会越来越得到重视，甚至成为主流的图像融合方法。

5. 重视评价指标和评价方法的研究

尽管目前已有许多种图像融合算法的评价指标被提出，但是这些指标均不能充分反映图像融合算法的好坏。在实际研究工作中，选择哪种或哪些评价指标对图像融合算法进行评价具有非常大的随机性，因此大大阻碍了算法性能的评价和领域的发展。因此，笔者认为，在未来的图像融合研究中，研究更有效的算法评价指标会是一个非常有前景的研究方向。

此外，鉴于在图像融合领域没有标准答案存在，笔者认为可以将图像融合与实际应用（例如目标跟踪、检测）相结合，并利用这些有标准答案存在的应用来对图像融合算法进行评价。具体地，当某种图像融合算法更能提高实际应用的性能时，证明该算法更加有效。

除了研究具体的评价指标和方法以外，图像融合评价基准的研发也将是未来图像融合领域的研究重点。笔者在工作中已对图像融合领域的评价基准进行过初步探索，然而，要研发更好的图像融合评价基准，仍需要大量研究人员的持续投入。

6. 解决图像配准问题

本书所介绍的绝大多数图像融合算法使用时有一个前提条件：待融合图像是严格配准的。然而，由于成像机理或者成像条件的差异，这个条件在实际中往往很难满足。图16.3中展示了常用的可见光与红外数据集 GTOT[7] 和 RGBT234[8] 中的未配准现象。从图中可以看出，可见光和红外图像之间是有比较明显的差异的。这种差异，会导致图像融合的效果较差，出现重影、鬼影等现象。



图 16.3: GTOT 和 RGBT234 数据集中可见光和红外图像未配准的情况。

图像配准是一个研究了很久的问题，不过一直未能得到很好的解决。近年来，图像配准问题又得到了一些图像融合研究人员的注意。一个比较重要的趋势是，研究人员将图像配准和图像融合放在同一个深度学习模型中去学习。例如，武汉大学马佳义教授团队提出的 RFNet 算法 [148] 和大连理工大学樊鑫教授团队提出的 PIAFusion 算法 [147]，均是同时处理图像配准和图像融合的算法。当然，值得指出的是，RFNet 算法是针对可见光与近红外图像融合所设计。

不过，尽管有这些方法的存在，图像的严格配准仍然是一个难题。笔者认为，未来应当有一批研究人员将精力投入到图像配准上。此外，笔者认为，图像融合和图像配准的发展会是相辅相成的。

7. 配准、融合和应用一体化

在图像融合领域，图像融合以前的配准、图像融合算法和图像融合以后的应用，是三个关键的阶段。在绝大多数研究中，这三个步骤是分开进行的，并且很少有研究人员会同时研究这三个问题。

不过，目前已有部分学者开始研究配准、融合和应用一体化的问题。例如，武汉大学马佳义教授团队于 2022 年发表了 SuperFusion 方法 [170]，将配准、融

合和应用在一个模型中一并完成，取得了很好的效果。

8. 提升图像融合效率

到目前为止，基于深度学习的总体发展是模型越来越大。然而，吴军博士在他的《全球科技通史》中指出，人类文明发展的本质就是用更少的能量获得更多信息的过程。根据这个观点，笔者认为，基于深度学习的图像融合方法，也应当朝着用更少的能量去融合更多信息这个方向发展。换句话说，我们需要努力去提升图像融合的效率，以便将图像融合技术应用于更多的任务中。

9. 与其他任务相结合

目前觉得图像融合算法只能处理图像融合任务。然而，如前文所述，近年来也有一些算法可以同时处理图像融合和其他任务，即将图像融合和其他任务结合在一个模型之中。一种实现的思路是使用多任务学习。这样做好处是模型的利用效率更高，因为一个模型可以完成 2 种或者多种任务。因此，笔者认为图像融合领域另外一个未来的发展趋势，就是与其他任务相结合。

10. 多幅源图像的融合

目前绝大多数图像融合方法是针对两幅源图像的情况设计的。然而，现实中也可能出现多幅源图像的融合情况，尤其是在多聚焦图像融合和多曝光图像融合中。尽管有研究人员提出先融合两幅图像得到一幅融合图像，然后将该融合图像和第三幅图像融合得到新的融合图像（以此类推）的方法。但是是否存在效率更高、效果更好的融合方法，是一个值得研究的问题。

11. 研究其他类型传感器的图像融合

笔者在前文中详细介绍了被研究得比较多的几种图像融合任务，并简单介绍了一些近年来新出现的图像融合任务，如可见光图像和深度图像的融合、可见光图像和近红外图像的融合、多视角图像融合。然而，尽管这些图像融合任务都有非常好的应用前景，但是相关的研究还非常少。因此，笔者认为，未来的图像融合研究不会再局限于可见光红外图像融合、多聚焦图像融合和多曝光图像融合等几种传统的的图像任务。未来会有越来越多的研究人员开展其他类型的图像融合任务的研究与应用。

最后,尽管本书主要介绍图像融合,融合这种思想却广泛存在于许多领域。笔者认为,智能融合技术在未来会受到越来越多的重视,得到越来越多的应用。麻省理工学院计算机科学与人工智能实验室主任丹尼尔那·露丝在她的新书《The Heart and the Chip》一书中描述了很多人和机器人协作的例子。笔者认为,人机融合也是未来的重要发展方向。

附录 A

图像融合相关的学术期刊和学术会议

图像融合是一个蓬勃发展的研究领域。最近几年，越来越多的研究人员开始在顶级学术期刊和会议上发表图像融合相关的文章，不断地提出新的图像融合方法，也不断地开拓图像融合的新应用领域。

为了方便读者朋友们寻找图像融合相关的论文进行阅读，也为了方便读者们在进行论文投稿时更容易地找到合适的学术期刊和会议，笔者对主要的图像融合相关的国际期刊和会议进行了小结。这些期刊和会议可以在表A.1中找到。读者朋友们在投图像融合相关的论文时，如果选择这些期刊和会议，则不用担心方向不对口的问题。

在表中，笔者还提供了在每个期刊或会议上发表的图像融合论文样例一篇或几篇，以便读者朋友们在寻找期刊和会议时进行对比。值得说明的是，笔者在表A.1中列出的仅是比较有代表性、质量也较高的一些期刊和会议。还有很多其他学术期刊和会议也会接收图像融合相关的论文。

表 A.1: 图像融合相关的学术期刊和会议

序号	简称	全称	样例
学术期刊			
1	TPAMI	IEEE Transactions on Pattern Analysis and Machine Intelligence	[35, 60, 10]
2	IJCV	International Journal of Computer Vision	[158]
3	TIP	IEEE Transactions on Image Processing	[46, 184, 155]
4	TNNLS	IEEE Transactions on Neural Networks and Learning Systems	[334, 335]
5	TMM	IEEE Transactions on Multimedia	[187, 122, 336]
6	TIM	IEEE Transactions on Instrumentation and Measurement	[106, 121, 116]
7	TCSVT	IEEE Transactions on Circuits and Systems for Video Technology	[337, 338, 289]
8	TCI	IEEE Transactions on Computational Imaging	[176, 339, 80]
9	TGRS	IEEE Transactions on Geoscience and Remote Sensing	[340, 341]
10	INFUS	Information Fusion	[39, 1, 342]
11	PR	Pattern Recognition	[149, 343]
12	PRL	Pattern Recognition Letters	[344, 202]
13	IPT	Infrared Physics & Technology	[345, 58]
14		Chinese Journal of Information Fusion (中国信息融合学报)	
学术会议			
1	CVPR	IEEE/CVF Conference on Computer Vision and Pattern Recognition	[148]
2	ICCV	IEEE International Conference on Computer Vision	[207, 346]
3	ECCV	European Conference on Computer Vision	[347, 348]
4	AAAI	AAAI conference on Artificial Intelligence	[162, 161]
5	IJCAI	International Joint Conference on Artificial Intelligence	[117, 160]
6	ACM MM	ACM Multimedia	[349]
7	ICASSP	IEEE International Conference on Acoustics, Speech and Signal Processing	[206, 350]
8	ICME	IEEE International Conference on Multimedia and Expo	[351, 352]
9	ICPR	International Conference on Pattern Recognition	[56]
10	ICIP	International Conference on Image Processing	[353, 354]
11	Fusion	International Conference on Information Fusion	[355, 356]
12		中国信息融合大会	[244]

附录 B

图像融合相关开源代码下载链接

与计算机视觉中的其他很多研究领域相比，图像融合领域的开源氛围并不活跃，因为文献中很多图像融合算法的代码并未开源。幸运的是，仍有一批研究人员无私地开源了他们的代码，大大方便了图像融合领域的研究，为图像融合领域的发展做出了很大贡献。为了方便读者，笔者对截止到本书交稿为止时的图像融合开源代码的下载链接进行了整理。具体的下载链接，以及更多关于图像融合的资源，可前往笔者的 GitHub 主页进行查看。

值得指出的是，近年来图像融合领域的开源氛围有了很大的改善。近几年发表的许多算法都有对应的开源代码。笔者认为开源氛围的改善可以促进图像融合领域进一步蓬勃发展，并将吸引更多的研究人员加入这个领域。

附录 C

图像融合论文写作经验

笔者在知名学术期刊和学术会议上发表过多篇图像融合相关的论文，也经常替 IEEE TPAMI、IJCV 等知名学术期刊和 CVPR、ICCV、ECCV、AAAI 等学术会议审阅图像融合相关的稿件，因此在图像融合论文的写作方面积累了一些心得。在本节中，笔者简要介绍一些图像融合论文的写作经验，供读者朋友参考。值得指出的是，这些经验都是笔者的个人经验，其他专家和审稿人可能会有不同要求。因此，这些经验仅供读者朋友们参考。在实际写作中，还需要根据具体的研究工作质量以及目标期刊、会议，进行调整。

1. 研究动机要写清楚

研究动机是一篇论文的核心要素之一，是必须要交代清楚的。然而，笔者在替学术期刊和会议审稿时，经常发些的一个问题是论文的研究动机写得不是特别清楚。有相当一部分的论文只是笼统地指出已有方法不能很好地在融合图像中保留源图像的细节信息，但是并未进行具体说明。还有不少论文只是笼统地说已有算法效果不好，但是哪些算法效果不好，怎么不好，为什么不好，并没有具体指出。这种含糊的动机，会很容易让审稿人觉得论文作者对于研究领域的把握不充分、文献调研不到位。事实上，如果研究动机没有写清楚，那么算法的创新性也很难写清楚，因此会影响审稿人对于论文的评价，甚至会质疑论文的创新性。

因此，笔者建议读者朋友们再写论文的时候，一定要清楚地指出已有算法的不足之处。比较典型的是方法是选取最具代表性的算法，先扬后抑地客观描述其优点与不足之处。

2. 文章要有新意

在审了和读了图像融合领域里的许多论文以后，笔者觉得很多文章没有新意。有大量的文章是把注意力机制、稠密网络等模块翻来覆去地用，进行各种排列组合，但却不明确指出这样做的原因。这样的文章显得没有新意，无法让审稿人眼前一亮，因此有较大的被拒稿的可能性。

季羡林先生曾经写过一篇文章，叫做《没有新意，不要写文章》。笔者觉得这篇文章写得非常好，说得非常有道理¹。文中说：

单篇论文的核心是讲自己的看法、自己异于前人的新意，要发前人未发之覆。有这样的文章，学术才能一步步、一代代向前发展。如果写一部专著，其中可能有自己的新意，也可能没有。因为大多数的专著是综合的、全面的叙述。即使不是自己的新意，也必须写进去，否则就不算全面。论文则没有这种负担，它的目的不是全面，而是深入，而是有新意，它与专著的关系可以说是相辅相成的。

——季羡林《没有新意，不要写文章》

至于如何找到新意，季羡林先生在文章中也提出了他自己的见解：

在大多数情况下，只有到杂志缝里才能找到新意。在大部头的专著中，在字里行间，也能找到新意的，旧日所谓“读书得间”，指的就是这种情况。因为，一般说来，杂志上发表的文章往往只谈一个问题、一个新问题，里面是有新意的。你读过以后，受到启发，举一反三，自己也产生了新意，然后写成文章，让别的学人也受到启发，再举一反三。如此往复循环，学术的进步就寓于其中了。

——季羡林《没有新意，不要写文章》

现在图像融合领域里的大多数论文是发表在学术期刊上，也有一部分文章是发表的学术会议上的。因此，笔者将季羡林先生的观点稍作修改，即只有到杂志和学术会议缝里才能找到新意。

当然，文章有了新意，还得把新意写清楚。季羡林先生说：“有了创见写论文，也不要下笔千言，离题万里。空洞的废话少说不说微妙”。

¹季先生虽然是搞人文社科的，但是他的很多见解对于其他学科也很有借鉴作用

3. 实验要做充分

笔者在审稿时经常发现一些图像融合论文的实验不是很充分，容易让审稿人产生质疑。这一般主要体现在如下几个方面：

- 对比的方法太少。在一些论文中，作者仅选取少数几个图像融合算法进行性能对比，而这些算法无法覆盖较多类型的图像融合算法，因此使得实验不是很有说服性。
- 对比的方法太老。图像融合领域发展非常迅速，每年都有许多新的方法被提出。因此，在论文中需要尽可能地选取一些较新的方法进行实验对比，这样才更具有说服力。有部分作者会选取数年前乃至十几年前的算法进行性能对比，这样不是很有说服力，会很容易让审稿人对方法的真实性能产生质疑。
- 使用的评价指标太少。在图像融合领域中，一个评价指标往往只能从某个方面对图像融合的结果进行评价。在这种情况下，如果只选取很少的评价指标对算法进行评价的话，很容易出现“盲人摸象”的情况，从而导致性能评价不是特别客观。这也会影响审稿人对于论文的评价。尤其需要指出的是，在进行图像融合论文写作时，作者最好对于自己所选的评价指标进行一定的解释，说清楚自己为什么要选择这些评价指标而不是其他的一些评价指标。
- 只有定性结果或者只有定量结果。在图像融合论文中，只有定性结果或者定量结果不足以表明算法的优越性，因为定性结果和定量结果并不完全是对应和匹配的。因此，在论文中要同时展示定性结果和定量结果，以便审稿人可以对算法的性能进行全面评价。

4. 算法的细节要描述到位

在审稿过程中，笔者经常发现一些论文中缺乏某些细节信息。有些时候，缺乏的甚至是非常重要的信息。例如，在某些基于深度学习的图像融合算法的论文中，没有讨论训练集，导致审稿人在读完论文以后都不知道该算法是用什么数据进行训练的。还有些论文作者会忘记给出一些很重要的超参数，因此使得论文不具备可复现性。缺乏这些信息，都很有可能会影响审稿人对论文的印象，从而影

响审稿结果。因此，在论文中，一定要把细节，尤其是重要细节，描述清楚、到位。

5. 对结果进行充分分析

在很多图像融合论文中，作者只列举实验结果并且指出所提出的算法性能优于已有算法。然而，对于为什么所提算法的性能比现有方法好，则解释甚少。需要指出的是，在审稿时，审稿人会很关注所提算法为什么好，希望可以看到作者对于所设计的算法有充分深入的了解、实验与解释。因此，论文作者一定要对实验结果进行充分的分析和讨论，这样会使得审稿人对于作者所提的算法更有信心，会大大提高论文的命中率。

6. 适当展示所提方法的应用效果

在对所提方法进行定性和定量描述以外，作者们还可以适当地对所提方法的应用效果进行展示。例如，在可见光和红外图像融合的论文中，可以适当展示该方法对于下游的目标跟踪、行人检测等任务的促进作用。这样可以更有说服力地证明所提方法的好处。

索引

- Diffusion model, 33
GFF, 110
GPU, 18
IFCNN, 112
PMGI, 112
RGB, 58
RGBD, 116, 163
tensor, 17
事件相机, 164
优化问题, 17
像素级图像融合, 7
光场相机, 147
决策级图像融合, 7
判别式模型, 19
图像融合, 3
图像融合相关的期刊和会议, 175
图像配准, 9
多曝光图像融合, 6
多聚焦图像融合, 6
定性评价, 40
定量评价, 40, 42
张量, 17
强化学习, 19
扩散模型, 33
损失函数, 17
无监督学习, 19
梯度下降法, 17
深度图, 163
特征级图像融合, 7
生成式对抗网络, 33
生成式模型, 19
监督学习, 19
论文写作经验, 178
近红外图像, 161
通用图像融合方法, 110

参考文献

- [1] Xingchen Zhang, Ping Ye, Henry Leung, Ke Gong, and Gang Xiao. Object fusion tracking based on visible and infrared images: A comprehensive review. *Information Fusion*, 63:166–187, 2020.
- [2] Xingchen Zhang, Ping Ye, and Gang Xiao. VIFB: a visible and infrared image fusion benchmark. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 104–105, 2020.
- [3] Durga Prasad Bavirisetti, Gang Xiao, Junhao Zhao, Ravindra Dhuli, and Gang Liu. Multi-scale guided image and video fusion: A fast and efficient approach. *Circuits, Systems, and Signal Processing*, 38(12):5576–5605, Dec 2019.
- [4] Mansour Nejati, Shadrokh Samavi, and Shahram Shirani. Multi-focus image fusion using dictionary-based sparse representation. *Information Fusion*, 25:72–84, 2015.
- [5] B. K. Shreyamsha Kumar. Image fusion based on pixel significance using cross bilateral filter. *Signal, Image and Video Processing*, 9(5):1193–1204, Jul 2015.
- [6] Lu Zhang, Xiangyu Zhu, Xiangyu Chen, Xu Yang, Zhen Lei, and Zhiyong Liu. Weakly aligned cross-modal learning for multispectral pedestrian detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5127–5137, 2019.
- [7] Chenglong Li, Hui Cheng, Shiyi Hu, Xiaobai Liu, Jin Tang, and Liang Lin. Learning collaborative sparse representation for grayscale-thermal tracking. *IEEE Transactions on Image Processing*, 25(12):5743–5756, 2016.

- [8] Chenglong Li, Xinyan Liang, Yijuan Lu, Nan Zhao, and Jin Tang. RGB-T object tracking: benchmark and baseline. *Pattern Recognition*, page 106977, 2019.
- [9] Soonmin Hwang, Jaesik Park, Namil Kim, Yukyung Choi, and In So Kweon. Multispectral pedestrian detection: Benchmark dataset and baseline. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1037–1045, 2015.
- [10] Xingchen Zhang. Deep learning-based multi-focus image fusion: A survey and a comparative study. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9):4819–4838, 2021.
- [11] Xingchen Zhang and Yiannis Demiris. Visible and infrared image fusion using deep learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.
- [12] Stuart Russell and Peter Norvig. Artificial intelligence: a modern approach, global edition 4th. *Foundations*, 19:23, 2021.
- [13] Laurent Hascoet and Valérie Pascual. The tapenade automatic differentiation tool: principles, model, and specification. *ACM Transactions on Mathematical Software (TOMS)*, 39(3):1–43, 2013.
- [14] Lorijn Zaadnoordijk, Tarek R Besold, and Rhodri Cusack. Lessons from infant learning for unsupervised machine learning. *Nature Machine Intelligence*, 4(6):510–520, 2022.
- [15] David Silver, Julian Schrittwieser, Karen Simonyan, et al. Mastering the game of go without human knowledge. *Nature*, 550:354–359, October 2017.
- [16] Jonas Degrave, Federico Felici, Jonas Buchli, Michael Neunert, Brendan Tracey, Francesco Carpanese, Timo Ewalds, Roland Hafner, Abbas Abdolmaleki, Diego de Las Casas, et al. Magnetic control of tokamak plasmas through deep reinforcement learning. *Nature*, 602(7897):414–419, 2022.
- [17] Melanie Mitchell. How do we know how smart ai systems are?, 2023.

- [18] Jack Stilgoe. We need a weizenbaum test for ai, 2023.
- [19] K Ram Prabhakar, V Sai Srikanth, and R Venkatesh Babu. Deepfuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs. In *Proceedings of the IEEE international conference on computer vision*, pages 4714–4722, 2017.
- [20] Luca Bertinetto, Jack Valmadre, Joao F Henriques, Andrea Vedaldi, and Philip HS Torr. Fully-convolutional siamese networks for object tracking. In *European conference on computer vision*, pages 850–865. Springer, 2016.
- [21] Jack Valmadre, Luca Bertinetto, João Henriques, Andrea Vedaldi, and Philip HS Torr. End-to-end representation learning for correlation filter based tracking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2805–2813, 2017.
- [22] Yu Liu, Xun Chen, Juan Cheng, Hu Peng, and Zengfu Wang. Infrared and visible image fusion with convolutional neural networks. *International Journal of Wavelets, Multiresolution and Information Processing*, 16(03):1850018, 2018.
- [23] Hui Li, Xiao-Jun Wu, and Josef Kittler. Infrared and visible image fusion using a deep learning framework. In *2018 24th international conference on pattern recognition (ICPR)*, pages 2705–2710. IEEE, 2018.
- [24] Hui Li and Xiaojun Wu. DenseFuse: A Fusion Approach to Infrared and Visible Images. *IEEE Transactions on Image Processing*, 28(5):2614–2623, 2019.
- [25] Jiayi Ma, Wei Yu, Pengwei Liang, Chang Li, and Junjun Jiang. FusionGAN: A generative adversarial network for infrared and visible image fusion. *Information Fusion*, 48(June 2018):11–26, 2019.
- [26] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10012–10022, 2021.

- [27] Jun Yue, Leyuan Fang, Shaobo Xia, Yue Deng, and Jiayi Ma. Dif-fusion: Towards high color fidelity in infrared and visible image fusion with diffusion models. *arXiv preprint arXiv:2301.08072*, 2023.
- [28] Zixiang Zhao, Haowen Bai, Yuanzhi Zhu, Jiangshe Zhang, Shuang Xu, Yulun Zhang, Kai Zhang, Deyu Meng, Radu Timofte, and Luc Van Gool. Ddfm: denoising diffusion model for multi-modality image fusion. *arXiv preprint arXiv:2303.06840*, 2023.
- [29] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [30] Aston Zhang, Zachary C Lipton, Mu Li, and Alexander J Smola. Dive into deep learning. *arXiv preprint arXiv:2106.11342*, 2021.
- [31] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.
- [32] Rohit Girdhar, Alaaeldin El-Nouby, Zhuang Liu, Mannat Singh, Kalyan Vasudev Alwala, Armand Joulin, and Ishan Misra. Imagebind: One embedding space to bind them all. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15180–15190, 2023.
- [33] Francois Chollet. *Deep learning with Python*. Simon and Schuster, 2021.
- [34] Peibei Cao, Zhangyang Wang, and Kede Ma. Debiased subjective assessment of real-world image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 711–721, 2021.
- [35] Zheng Liu, Erik Blasch, Zhiyun Xue, Jiying Zhao, Robert Laganiere, and Wei Wu. Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: A comparative study. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34:94–109, 2012.

- [36] Kede Ma, Kai Zeng, and Zhou Wang. Perceptual quality assessment for multi-exposure image fusion. *IEEE Transactions on Image Processing*, 24(11):3345–3356, 2015.
- [37] Zhihao Chang, Shuyuan Yang, Zhixi Feng, Quanwei Gao, Shengzhe Wang, and Yuyong Cui. Semantic-relation transformer for visible and infrared fused image quality assessment. *Information Fusion*, 95:454–470, 2023.
- [38] Jiao Du, Meie Fang, Yufeng Yu, and Gang Lu. An adaptive two-scale biomedical image fusion method with statistical comparisons. *Computer Methods and Programs in Biomedicine*, 196:105603, 2020.
- [39] Xingchen Zhang. Benchmarking and comparing multi-exposure image fusion algorithms. *Information Fusion*, 74:111–131, 2021.
- [40] Linfeng Tang, Jiteng Yuan, and Jiayi Ma. Image fusion in the loop of high-level vision tasks: A semantic-aware real-time infrared and visible image fusion network. *Information Fusion*, 2022.
- [41] Seonghyun Park, An Gia Vien, and Chul Lee. Cross-modal transformers for infrared and visible image fusion. *IEEE Transactions on Circuits and Systems for Video Technology*, pages 1–1, 2023.
- [42] Yi Wu, Jongwoo Lim, and Ming-Hsuan Yang. Online object tracking: A benchmark. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2411–2418, 2013.
- [43] Yi Wu, Jongwoo Lim, and Ming-Hsuan Yang. Object tracking benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(9):1834–1848, 2015.
- [44] Matej Kristan, Jiri Matas, Aleš Leonardis, et al. A novel performance evaluation methodology for single-target trackers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(11):2137–2155, Nov 2016.
- [45] VPS Naidu. Image fusion technique using multi-resolution singular value decomposition. *Defence Science Journal*, 61(5):479–484, 2011.

- [46] Shutao Li, Xudong Kang, and Jianwen Hu. Image fusion with guided filtering. *IEEE Transactions on Image Processing*, 22(7):2864–2875, 2013.
- [47] Yu Liu, Shuping Liu, and Zengfu Wang. A general framework for image fusion based on multi-scale transform and sparse representation. *Information Fusion*, 24:147–164, 2015.
- [48] Durga Prasad Bavirisetti and Ravindra Dhuli. Fusion of infrared and visible sensor images based on anisotropic diffusion and karhunen-loeve transform. *IEEE Sensors Journal*, 16(1):203–209, 2016.
- [49] Zhiqiang Zhou, Mingjie Dong, Xiaozhu Xie, and Zhifeng Gao. Fusion of infrared and visible images for night-vision context enhancement. *Applied optics*, 55(23):6480–6490, 2016.
- [50] Zhiqiang Zhou, Bo Wang, Sun Li, and Mingjie Dong. Perceptual fusion of infrared and visible images through a hybrid multi-scale decomposition with gaussian and bilateral filters. *Information Fusion*, 30:15–26, 2016.
- [51] Durga Prasad Bavirisetti and Ravindra Dhuli. Two-scale image fusion of visible and infrared images using saliency detection. *Infrared Physics & Technology*, 76:52–64, 2016.
- [52] Jiayi Ma, Chen Chen, Chang Li, and Jun Huang. Infrared and visible image fusion via gradient transfer and total variation minimization. *Information Fusion*, 31:100–109, 2016.
- [53] Durga Prasad Bavirisetti, Gang Xiao, and Gang Liu. Multi-sensor image fusion based on fourth order partial differential equations. In *2017 20th International Conference on Information Fusion*, pages 1–9. IEEE, 2017.
- [54] Yu Zhang, Lijia Zhang, Xiangzhi Bai, and Li Zhang. Infrared and visual image fusion through infrared feature extraction and visual information preservation. *Infrared Physics & Technology*, 83:227 – 237, 2017.
- [55] Jinlei Ma, Zhiqiang Zhou, Bo Wang, and Hua Zong. Infrared and visible image fusion based on visual saliency map and weighted least square optimization. *Infrared Physics & Technology*, 82:8–17, 2017.

- [56] Hui Li, Xiao-Jun Wu, and Josef Kittler. Infrared and visible image fusion using a deep learning framework. *24th International Conference on Pattern Recognition*, 2018.
- [57] Hui Li and Xiaojun Wu. Infrared and visible image fusion using latent low-rank representation. *arXiv:1804.08992*, 2018.
- [58] Hui Li, Xiao-Jun Wu, and Tariq S Durrani. Infrared and Visible Image Fusion with ResNet and zero-phase component analysis. *Infrared Physics & Technology*, 102:103039, 2019.
- [59] Jiayi Ma, Linfeng Tang, Fan Fan, Jun Huang, Xiaoguang Mei, and Yong Ma. SwinFusion: Cross-domain Long-range Learning for General Image Fusion via Swin Transformer. *IEEE/CAA Journal of Automatica Sinica*, 9(7):1200–1217, 2022.
- [60] Han Xu, Jiayi Ma, Junjun Jiang, Xiaojie Guo, and Haibin Ling. U2fusion: A unified unsupervised image fusion network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.
- [61] Wei Tang, Fazhi He, and Yu Liu. YDTR: infrared and visible image fusion via y-shape dynamic transformer. *IEEE Transactions on Multimedia*, 2022.
- [62] Yu Zhang, Yu Liu, Peng Sun, Han Yan, Xiaolin Zhao, and Li Zhang. IFCNN: A general image fusion framework based on convolutional neural network. *Information Fusion*, 54(August 2018):99–118, 2020.
- [63] D. M. Bulanon, T.F. Burks, and V. Alchanatis. Image fusion of visible and thermal images for fruit detection. *Biosystems Engineering*, 103(1):12–22, 2009.
- [64] Van Aardt and Jan. Assessment of image fusion procedures using entropy, image quality, and multispectral classification. *Journal of Applied Remote Sensing*, 2(1):023522, 2008.
- [65] Guihong Qu, Dali Zhang, and Pingfan Yan. Information measure for performance of image fusion. *Electronics letters*, 38(7):313–315, 2002.

- [66] P Jagalingam and Arkal Vittal Hegde. A review of quality metrics for fused image. *Aquatic Procedia*, 4:133–142, 2015.
- [67] Zhou Wang, Alan C Bovik, Hamid R Sheikh, Eero P Simoncelli, et al. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.
- [68] Guangmang Cui, Huajun Feng, Zhihai Xu, Qi Li, and Yueling Chen. Detail preserved fusion of visible and infrared images using regional saliency extraction and multi-scale image decomposition. *Optics Communications*, 341:199 – 209, 2015.
- [69] B Rajalingam and R Priya. Hybrid multimodality medical image fusion technique for feature enhancement in medical diagnosis. *International Journal of Engineering Science Invention*, 2(Special issue):52–60, 2018.
- [70] Yun-Jiang Rao. In-fibre bragg grating sensors. *Measurement science and technology*, 8(4):355, 1997.
- [71] Ahmet M Eskicioglu and Paul S Fisher. Image quality measures and their performance. *IEEE Transactions on Communications*, 43(12):2959–2965, 1995.
- [72] C. S. Xydeas and Petrovic V V. Objective image fusion performance measure. *Military Technical Courier*, 36(4):308–309, 2000.
- [73] Yin Chen and Rick S Blum. A new automated quality assessment algorithm for image fusion. *Image and vision computing*, 27(10):1421–1432, 2009.
- [74] Hao Chen and Pramod K Varshney. A human perception inspired quality metric for image fusion based on regional information. *Information fusion*, 8(2):193–207, 2007.
- [75] Jinyuan Liu, Xin Fan, Zhanbo Huang, Guanyao Wu, Risheng Liu, Wei Zhong, and Zhongxuan Luo. Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

- [76] Shailesh Nirgudkar, Michael DeFilippo, Michael Sacarny, Michael Benjamin, and Paul Robinette. Massmind: Massachusetts maritime infrared dataset. *The International Journal of Robotics Research*, 42(1-2):21–32, 2023.
- [77] Wei Wu, Zongming Qiu, Min Zhao, QiuHong Huang, and Yang Lei. Visible and infrared image fusion using NSST and deep Boltzmann machine. *Optik*, 157:334–342, 2018.
- [78] Jiayi Ma, Yong Ma, and Chang Li. Infrared and visible image fusion methods and applications: A survey. *Information Fusion*, 45:153–178, 2019.
- [79] Yaochen Liu, Lili Dong, Yuanyuan Ji, et al. Infrared and visible image fusion through details preservation. *Sensors*, 19(20):4556, 2019.
- [80] Ruichao Hou, Dongming Zhou, Rencan Nie, Dong Liu, Lei Xiong, Yanbu Guo, and Chuanbo Yu. VIF-Net: An Unsupervised Framework for Infrared and Visible Image Fusion. *IEEE Transactions on Computational Imaging*, 6:640–651, 2020.
- [81] Han Xu, Meiqi Gong, et al. CUFD: An encoder–decoder network for visible and infrared image fusion based on common and unique feature decomposition. *Computer Vision and Image Understanding*, 218:103407, 2022.
- [82] Hafiz Tayyab Mustafa, Jie Yang, Hamza Mustafa, et al. Infrared and visible image fusion based on dilated residual attention network. *Optik*, 224:165409, 2020.
- [83] Qingqing Li, Guangliang Han, Peixun Liu, Hang Yang, Dianbing Chen, Xinglong Sun, Jiajia Wu, and Dongxu Liu. A multilevel hybrid transmission network for infrared and visible image fusion. *IEEE Transactions on Instrumentation and Measurement*, 71:1–14, 2022.
- [84] Zhishe Wang, Junyao Wang, Yuanyuan Wu, Jiawei Xu, and Xiaoqin Zhang. Unfusion: A unified multi-scale densely connected network for infrared and visible image fusion. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021.

- [85] Jiahui Zhu, Qingyu Dou, Lihua Jian, Kai Liu, Farhan Hussain, and Xiaomin Yang. Multiscale channel attention network for infrared and visible image fusion. *Concurrency and Computation: Practice and Experience*, page e6155, 2020.
- [86] Yang Li, Jixiao Wang, Zhuang Miao, and Jiabao Wang. Unsupervised densely attention network for infrared and visible image fusion. *Multimedia Tools and Applications*, 79(45):34685–34696, 2020.
- [87] Zhaisheng Ding, Haiyan Li, Dongming Zhou, et al. A robust infrared and visible image fusion framework via multi-receptive-field attention and color visual perception. *Applied Intelligence*, pages 1–19, 2022.
- [88] Yi Liu, Changyun Miao, Jianhua Ji, and Xianguo Li. Mmf: A multi-scale mobilenet based fusion method for infrared and visible image. *Infrared Physics & Technology*, 119:103894, 2021.
- [89] Lei Yan, Jie Cao, Saad Rizvi, Kaiyu Zhang, Qun Hao, and Xuemin Cheng. Improving the Performance of Image Fusion Based on Visual Saliency Weight Map Combined With CNN. *IEEE Access*, 8:59976–59986, 2020.
- [90] Xiaoqing Luo, Yuanhao Gao, Anqi Wang, Zhancheng Zhang, and Xiao-Jun Wu. IFSepR: A general framework for image fusion based on separate representation learning. *IEEE Transactions on Multimedia*, 2021.
- [91] Jinyuan Liu, Yuhui Wu, Zhanbo Huang, Risheng Liu, and Xin Fan. Smoa: Searching a modality-oriented architecture for infrared and visible image fusion. *IEEE Signal Processing Letters*, 28:1818–1822, 2021.
- [92] Risheng Liu, Zhu Liu, Jinyuan Liu, and Xin Fan. Searching a hierarchically aggregated fusion architecture for fast multi-modality image fusion. In *Proceedings of the 29th ACM International Conference on Multimedia*, pages 1600–1608, 2021.
- [93] Han Xu, Xinya Wang, and Jiayi Ma. DRF: Disentangled Representation for Visible and Infrared Image Fusion. *IEEE Transactions on Instrumentation and Measurement*, 70:1–13, 2021.

- [94] Linfeng Tang, Jiteng Yuan, Hao Zhang, Xingyu Jiang, and Jiayi Ma. PI-AFusion: A progressive infrared and visible image fusion network based on illumination aware. *Information Fusion*, 83:79–92, 2022.
- [95] Yongzhi Long, Haitao Jia, Yida Zhong, Yadong Jiang, and Yuming Jia. RXDNFuse: A aggregated residual dense network for infrared and visible image fusion. *Information Fusion*, 69:128–141, 2021.
- [96] Yan Zou, Linfei Zhang, Chengqian Liu, Bowen Wang, Yan Hu, and Qian Chen. Infrared visible color night vision image fusion based on deep learning. In *AI and Optical Data Sciences II*, volume 11703, page 117031S. International Society for Optics and Photonics, 2021.
- [97] Zhengwen Shen, Jun Wang, Zaiyu Pan, Yulian Li, and Jiangyu Wang. Cross attention-guided dense network for images fusion. *arXiv preprint arXiv:2109.11393*, 2021.
- [98] Wen-Bo An and Hong-Mei Wang. Infrared and visible image fusion with supervised convolutional neural network. *Optik*, 219:165120, 2020.
- [99] Yufang Feng, Houqing Lu, Jingbo Bai, Lin Cao, and Hong Yin. Fully convolutional network-based infrared and visible image fusion. *Multimedia Tools and Applications*, 79(21):15001–15014, 2020.
- [100] Meng Wang, Xingwang Liu, and Huaiping Jin. A generative image fusion approach based on supervised deep convolution network driven by weighted gradient flow. *Image and Vision Computing*, 86:1–16, 2019.
- [101] Depeng Zhu, Weida Zhan, Yichun Jiang, et al. IPLF: A novel image pair learning fusion network for infrared and visible image. *IEEE Sensors Journal*, 22(9):8808–8817, 2022.
- [102] Jiayi Ma, Linfeng Tang, Meilong Xu, Hao Zhang, and Guobao Xiao. STD-FusionNet: An Infrared and Visible Image Fusion Network Based on Salient Target Detection. *IEEE Transactions on Instrumentation and Measurement*, 2021.

- [103] Dawei Zhang, Kan Ren, Jing Zhou, Guohua Gu, and Qian Chen. An infrared and visible image fusion method based on deep learning. In *4th Optics Young Scientist Summit*, volume 11781, page 1178109. International Society for Optics and Photonics, 2021.
- [104] Xianyi Ren, Fanyang Meng, Tao Hu, Zhijun Liu, and Changwei Wang. Infrared-visible image fusion based on convolutional neural networks. In *International Conference on Intelligent Science and Big Data Engineering*, pages 301–307. Springer, 2018.
- [105] Yong Yang, Jia-Xiang Liu, Shu-Ying Huang, Hang-Yuan Lu, and Wen-Ying Wen. VMDM-fusion: a saliency feature representation method for infrared and visible image fusion. *Signal, Image and Video Processing*, pages 1–9, 2021.
- [106] Lihua Jian, Xiaomin Yang, Zheng Liu, Gwanggil Jeon, Mingliang Gao, and David Chisholm. SEDRFuse: A Symmetric Encoder–Decoder With Residual Block Network for Infrared and Visible Image Fusion. *IEEE Transactions on Instrumentation and Measurement*, 70:1–15, 2021.
- [107] Hui Li, Xiao-Jun Wu, and Tariq Durrani. NestFuse: An infrared and visible image fusion architecture based on nest connection and spatial/channel attention models. *IEEE Transactions on Instrumentation and Measurement*, 69(12):9645–9656, 2020.
- [108] Hui Li, Xiao-Jun Wu, and Josef Kittler. RFN-Nest: An end-to-end residual fusion network for infrared and visible images. *Information Fusion*, 2021.
- [109] Han Xu, Hao Zhang, and Jiayi Ma. Classification saliency-based rule for visible and infrared image fusion. *IEEE Transactions on Computational Imaging*, 7:824–836, 2021.
- [110] Jiayi Ma, Pengwei Liang, Wei Yu, Chen Chen, Xiaojie Guo, Jia Wu, and Junjun Jiang. Infrared and visible image fusion via detail preserving adversarial learning. *Information Fusion*, 54:85–98, 2020.

- [111] Dongdong Xu, Yongcheng Wang, Shuyan Xu, Kaiguang Zhu, Ning Zhang, and Xin Zhang. Infrared and visible image fusion with a generative adversarial network and a residual network. *Applied Sciences*, 10(2):554, 2020.
- [112] Yu Fu, Xiao-Jun Wu, and Tariq Durrani. Image fusion based on generative adversarial network consistent with perception. *Information Fusion*, 72:110–125, 2021.
- [113] Jixiao Wang, Yang Li, and Zhuang Miao. A new infrared and visible image fusion method based on generative adversarial networks and attention mechanism. In *2021 The 4th International Conference on Image and Graphics Processing*, pages 109–119, 2021.
- [114] C Yuan, CQ Sun, XY Tang, and RF Liu. FLGC-Fusion GAN: An Enhanced Fusion GAN Model by Importing Fully Learnable Group Convolution. *Mathematical Problems in Engineering*, 2020, 2020.
- [115] Snigdha Bhagat, Shiv Dutt Joshi, Brejesh Lall, and Smriti Gupta. Multimodal sensor fusion using symmetric skip autoencoder via an adversarial regulariser. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:1146–1157, 2021.
- [116] Jiayi Ma, Hao Zhang, Zhenfeng Shao, Pengwei Liang, and Han Xu. GAN-McC: A Generative Adversarial Network With Multiclassification Constraints for Infrared and Visible Image Fusion. *IEEE Transactions on Instrumentation and Measurement*, 70, 2021.
- [117] Han Xu, Pengwei Liang, Wei Yu, Junjun Jiang, and Jiayi Ma. Learning a generative model for fusing infrared and visible images via conditional generative adversarial network with dual discriminators. In *IJCAI*, pages 3954–3960, 2019.
- [118] Jiayi Ma, Han Xu, Junjun Jiang, Xiaoguang Mei, and Xiao-Ping Zhang. DDcGAN: A dual-discriminator conditional generative adversarial network for multi-resolution image fusion. *IEEE Transactions on Image Processing*, 29:4980–4995, 2020.

- [119] Jing Li, Hongtao Huo, Kejian Liu, Chang Li, Shuo Li, and Xin Yang. Infrared and visible image fusion via multi-discriminators wasserstein generative adversarial network. In *2019 18th IEEE International Conference On Machine Learning And Applications*, pages 2014–2019. IEEE, 2019.
- [120] Jing Li, Hongtao Huo, Kejian Liu, and Chang Li. Infrared and visible image fusion using dual discriminators generative adversarial networks with wasserstein distance. *Information Sciences*, 529:28–41, 2020.
- [121] Jing Li, Hongtao Huo, Chang Li, Renhua Wang, Chenhong Sui, and Zhao Liu. Multigrained attention network for infrared and visible image fusion. *IEEE Transactions on Instrumentation and Measurement*, 70, 2021.
- [122] Jing Li, Hongtao Huo, Chang Li, Renhua Wang, and Qi Feng. Attention-FGAN: Infrared and visible image fusion using attention-based generative adversarial networks. *IEEE Transactions on Multimedia*, 2020.
- [123] Hao Zhang, Jiteng Yuan, Xin Tian, and Jiayi Ma. GAN-FM: Infrared and Visible Image Fusion Using GAN With Full-Scale Skip Connection and Dual Markovian Discriminators. *IEEE Transactions on Computational Imaging*, 7:1134–1147, 2021.
- [124] Anyang Song, Huixian Duan, Haodong Pei, and Lei Ding. Triple-discriminator generative adversarial network for infrared and visible image fusion. *Neurocomputing*, 483:183–194, 2022.
- [125] Yuqing Zhao, Guangyuan Fu, Hongqiao Wang, and Shaolei Zhang. The fusion of unmatched infrared and visible images based on generative adversarial networks. *Mathematical Problems in Engineering*, 2020, 2020.
- [126] MA Lebedev, DV Komarov, OV Vygolov, and Yu V Vizilter. Multisensor image fusion based on generative adversarial networks. In *Image and Signal Processing for Remote Sensing XXV*, volume 11155, page 111551T. International Society for Optics and Photonics, 2019.
- [127] Ana Belén Petro, Catalina Sbert, and Jean-Michel Morel. Multiscale retinex. *Image Processing On Line*, pages 71–88, 2014.

- [128] Qilei Li, Lu Lu, Zhen Li, Wei Wu, Zheng Liu, Gwanggil Jeon, and Xiaomin Yang. Coupled GAN with relativistic discriminators for infrared and visible images fusion. *IEEE Sensors Journal*, 21(6), 2021.
- [129] Ming-Yu Liu and Oncel Tuzel. Coupled generative adversarial networks. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, pages 469–477, 2016.
- [130] Yansong Gu, Xinya Wang, Can Zhang, and Baiyang Li. Advanced driving assistance based on the fusion of infrared and visible images. *Entropy*, 23(2):239, 2021.
- [131] Huabing Zhou, Wei Wu, Yanduo Zhang, Jiayi Ma, and Haibin Ling. Semantic-supervised infrared and visible image fusion via a dual-discriminator generative adversarial network. *IEEE Transactions on Multimedia*, 2021.
- [132] Xiaoqing Luo, Anqi Wang, Zhancheng Zhang, Xinguang Xiang, and Xiao-Jun Wu. Latraivf: An infrared and visible image fusion method based on latent regression and adversarial training. *IEEE Transactions on Instrumentation and Measurement*, 70:1–16, 2021.
- [133] Zhishe Wang, Yanlin Chen, Wenyu Shao, Hui Li, and Lei Zhang. Swinfuse: A residual swin transformer fusion network for infrared and visible images. *IEEE Transactions on Instrumentation and Measurement*, 71:1–12, 2022.
- [134] Wei Tang, Fazhi He, and Yu Liu. Tccfusion: An infrared and visible image fusion method based on transformer and cross correlation. *Pattern Recognition*, 137:109295, 2023.
- [135] Long Ren, Zhibin Pan, Jianzhong Cao, and Jiawen Liao. Infrared and visible image fusion based on variational auto-encoder and infrared feature compensation. *Infrared Physics & Technology*, 117:103839, 2021.
- [136] Haibo Zhao and Rencan Nie. DNDT: Infrared and Visible Image Fusion Via DenseNet and Dual-Transformer. In *2021 International Conference on*

- Information Technology and Biomedical Engineering (ICITBE)*, pages 71–75. IEEE, 2021.
- [137] Yu Fu, TianYang Xu, XiaoJun Wu, and Josef Kittler. PPT Fusion: Pyramid Patch Transformerfor a Case Study in Image Fusion. *arXiv preprint arXiv:2107.13967*, 2021.
- [138] Dongyu Rao, Tianyang Xu, and Xiao-Jun Wu. Tgfuse: An infrared and visible image fusion approach based on transformer and generative adversarial network. *IEEE Transactions on Image Processing*, 2023.
- [139] A. Raza, J. Liu, Y. Liu, Z. Li, J. Liu, X. Chen, H. Huo, and T. Fang. IR-MSDNet: Infrared and Visible Image Fusion based on Infrared Features Multiscale Dense Network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, pages 1–1, 2021.
- [140] Juan Wang, Cong Ke, Minghu Wu, Min Liu, and Chunyan Zeng. Infrared and visible image fusion based on Laplacian pyramid and generative adversarial network. *KSII Transactions on Internet & Information Systems*, 15(5), 2021.
- [141] Yong Yang, Jiaxiang Liu, Shuying Huang, Weiguo Wan, Wenying Wen, and Juwei Guan. Infrared and visible image fusion via texture conditional generative adversarial network. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021.
- [142] Huafeng Li, Yueliang Cen, Yu Liu, Xun Chen, and Zhengtao Yu. Different input resolutions and arbitrary output resolution: A meta learning-based deep framework for infrared and visible image fusion. *IEEE Transactions on Image Processing*, 30:4070–4083, 2021.
- [143] Wanxin Xiao, Yafei Zhang, Hongbin Wang, Fan Li, and Hua Jin. Heterogeneous knowledge distillation for simultaneous infrared-visible image fusion and super-resolution. *IEEE Transactions on Instrumentation and Measurement*, 71:1–15, 2022.

- [144] Jae Hak Lee, Yong Sun Kim, Duhgoon Lee, Dong-Goo Kang, and Jong Beom Ra. Robust ccd and ir image registration using gradient-based statistical information. *IEEE Signal Processing Letters*, 17(4):347–350, 2010.
- [145] Jungong Han, Eric J Pauwels, and Paul De Zeeuw. Visible and infrared image registration in man-made environments employing hybrid visual features. *Pattern Recognition Letters*, 34(1):42–51, 2013.
- [146] Chaobo Min, Yan Gu, Yingjie Li, and Feng Yang. Non-rigid infrared and visible image registration by enhanced affine transformation. *Pattern Recognition*, 106:107377, 2020.
- [147] Di Wang, Jinyuan Liu, Xin Fan, and Risheng Liu. Unsupervised misaligned infrared and visible image fusion via cross-modality image generation and registration. In *IJCAI*, 2022.
- [148] Han Xu, Jiayi Ma, Jiteng Yuan, Zhuliang Le, and Wei Liu. RFNet: Unsupervised Network for Mutually Reinforcing Multi-modal Image Registration and Fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.
- [149] Pengyu Zhang, Dong Wang, and Huchuan Lu. Multi-modal visual tracking: Review and experimental comparison. *arXiv preprint arXiv:2012.04176*, 2020.
- [150] Ivana Shopovska, Ljubomir Jovanov, and Wilfried Philips. Deep visible and thermal image fusion for enhanced pedestrian visibility. *Sensors*, 19(17):3727, 2019.
- [151] Jing Li, Jianming Zhu, Chang Li, Xun Chen, and Bin Yang. CGTF: Convolution-Guided Transformer for Infrared and Visible Image Fusion. *IEEE Transactions on Instrumentation and Measurement*, 2022.
- [152] Chunyang Cheng, Xiao-Jun Wu, Tianyang Xu, and Guoyang Chen. Uni-fusion: A lightweight unified image fusion network. *IEEE Transactions on Instrumentation and Measurement*, 70:1–14, 2021.

- [153] Yuan Gao, Shiwei Ma, and Jingjing Liu. DCDR-GAN: A densely connected disentangled representation generative adversarial network for infrared and visible image fusion. *IEEE Trans. Circuits Syst. Video Technol.*, 2022.
- [154] Hyungjoo Jung, Youngjung Kim, Hyunsung Jang, Namkoo Ha, and Kwanghoon Sohn. Unsupervised Deep Image Fusion with Structure Tensor Representations. *IEEE Transactions on Image Processing*, 29:3845–3858, 2020.
- [155] Han Xu, Jiayi Ma, and Xiao-Ping Zhang. MEF-GAN: multi-exposure image fusion via generative adversarial networks. *IEEE Transactions on Image Processing*, 29:7203–7216, 2020.
- [156] Han Xu, Jiayi Ma, Junjun Jiang, et al. U2fusion: A unified unsupervised image fusion network. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(1):502–518, 2022.
- [157] Risheng Liu, Jinyuan Liu, Zhiying Jiang, Xin Fan, and Zhongxuan Luo. A bilevel integrated model with data-driven layer ensemble for multi-modality image fusion. *IEEE Transactions on Image Processing*, 30:1261–1274, 2021.
- [158] Hao Zhang and Jiayi Ma. SDNet: A Versatile Squeeze-and-Decomposition Network for Real-Time Image Fusion. *International Journal of Computer Vision*, pages 1–25, 2021.
- [159] Meilong Xu, Linfeng Tang, Hao Zhang, and Jiayi Ma. Infrared and visible image fusion via parallel scene and texture learning. *Pattern Recognition*, 132:108929, 2022.
- [160] Zixiang Zhao, Shuang Xu, Chunxia Zhang, Junmin Liu, Pengfei Li, and Jiangshe Zhang. DIDFuse: Deep Image Decomposition for Infrared and Visible Image Fusion. In *Proceedings of IJCAI*, 2020.
- [161] H Xu, J Ma, Z Le, J Jiang, and X Guo. FusionDN: A unified densely connected network for image fusion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020.

- [162] Hao Zhang, Han Xu, Yang Xiao, Xiaojie Guo, and Jiayi Ma. Rethinking the image fusion: A fast unified image fusion network based on proportional maintenance of gradient and intensity. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020.
- [163] Xingchen Zhang, Gang Xiao, Ping Ye, Dan Qiao, Junhao Zhao, and Shengyun Peng. Object fusion tracking based on visible and infrared images using fully convolutional siamese networks. In *Proceedings of the 22nd International Conference on Information Fusion*. IEEE, 2019.
- [164] Yabin Zhu, Chenglong Li, Jin Tang, and Bin Luo. Quality-aware feature aggregation network for robust rgbt tracking. *IEEE Transactions on Intelligent Vehicles*, 2020.
- [165] Haijiang Sun, Qiaoyuan Liu, Jiacheng Wang, Jinchang Ren, Yanfeng Wu, Huimin Zhao, and Huakang Li. Fusion of infrared and visible images for remote detection of low-altitude slow-speed small targets. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:2971–2983, 2021.
- [166] Wujie Zhou, Shaohua Dong, Caie Xu, and Yaguan Qian. Edge-Aware Guidance Fusion Network for RGB-Thermal Scene Parsing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 3571–3579, 2022.
- [167] David C Schedl, Indrajit Kurmi, and Oliver Bimber. An autonomous drone for search and rescue in forests using airborne optical sectioning. *Science Robotics*, 6(55):eabg1188, 2021.
- [168] Long Chen, Libo Sun, Teng Yang, Lei Fan, Kai Huang, and Zhe Xuanyuan. RGB-T SLAM: A flexible SLAM framework by combining appearance and thermal information. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5682–5687. IEEE, 2017.
- [169] Chenglong Li, Wanlin Xue, Yaqing Jia, Zhichen Qu, Bin Luo, and Jin Tang. LasHeR: A Large-scale High-diversity Benchmark for RGBT Tracking. *arXiv preprint arXiv:2104.13202*, 2021.

- [170] Linfeng Tang, Yuxin Deng, Yong Ma, et al. SuperFusion: A versatile image registration and fusion network with semantic awareness. *IEEE/CAA Journal of Automatica Sinica*, 9(12):2121–2137, 2022.
- [171] Yuanjie Gu, Zhibo Xiao, Hailun Wang, Cheng Liu, and Shouyu Wang. A dataset-free self-supervised disentangled learning method for adaptive infrared and visible images super-resolution fusion. *arXiv preprint arXiv:2112.02869*, 2021.
- [172] Sedat Özer, Mert Ege, and Mehmet Akif Özkanoglu. SiameseFuse: A computationally efficient and a not-so-deep network to fuse visible and infrared images. *Pattern Recognition*, 129:108712, 2022.
- [173] Yu Liu, Xun Chen, Hu Peng, and Zengfu Wang. Multi-focus image fusion with a deep convolutional neural network. *Information Fusion*, 36:191–207, 2017.
- [174] Han Tang, Bin Xiao, Weisheng Li, and Guoyin Wang. Pixel convolutional neural network for multi-focus image fusion. *Information Sciences*, 433:125–141, 2018.
- [175] Chang Wang, Zongya Zhao, Qiongqiong Ren, Yongtao Xu, and Yi Yu. A novel multi-focus image fusion by combining simplified very deep convolutional networks and patch-based sequential reconstruction strategy. *Applied Soft Computing*, page 106253, 2020.
- [176] Yong Yang, Zhipeng Nie, Shuying Huang, Pan Lin, and Jiahua Wu. Multilevel features convolutional neural network for multifocus image fusion. *IEEE Transactions on Computational Imaging*, 5(2):262–273, 2019.
- [177] Chaoben Du and Shesheng Gao. Image segmentation-based multi-focus image fusion through multi-scale convolutional neural network. *IEEE Access*, 5:15750–15761, 2017.
- [178] Rui Lai, Yongxue Li, Juntao Guan, and Ai Xiong. Multi-scale visual attention deep convolutional neural network for multi-focus image fusion. *IEEE Access*, 7:114385–114399, 2019.

- [179] Zeyu Wang, Xiongfei Li, Haoran Duan, Xiaoli Zhang, and Hancheng Wang. Multifocus image fusion using convolutional neural networks in the discrete wavelet transform domain. *Multimedia Tools and Applications*, 78(24):34483–34512, 2019.
- [180] Xiaopeng Guo, Lingyu Meng, Liye Mei, Yueyun Weng, and Hengqing Tong. Multi-focus image fusion with siamese self-attention network. *IET Image Processing*, 14(7):1339–1346, 2020.
- [181] Haoyu Ma, Juncheng Zhang, Shaojun Liu, and Qingmin Liao. Boundary aware multi-focus image fusion using deep neural network. In *IEEE International Conference on Multimedia and Expo*, pages 1150–1155. IEEE, 2019.
- [182] Haoyu Ma, Qingmin Liao, Juncheng Zhang, Shaojun Liu, and Jing-Hao Xue. An α -matte boundary defocus model-based cascaded network for multi-focus image fusion. *IEEE Transactions on Image Processing*, 29:8668–8679, 2020.
- [183] Xiaopeng Guo, Rencan Nie, Jinde Cao, Dongming Zhou, and Wenhua Qian. Fully convolutional network-based multifocus image fusion. *Neural computation*, 30(7):1775–1800, 2018.
- [184] Jinxing Li, Xiaobao Guo, Guangming Lu, Bob Zhang, Yong Xu, Feng Wu, and David Zhang. DRPL: Deep Regression Pair Learning for Multi-Focus Image Fusion. *IEEE Transactions on Image Processing*, 29:4816–4831, 2020.
- [185] Kaiping Xu, Zheng Qin, Guolong Wang, Huidi Zhang, Kai Huang, and Shuxiong Ye. Multi-focus Image Fusion using Fully Convolutional Two-stream Network for Visual Sensors. *KSII Transactions on Internet and Information Systems*, 12(5):2253–2272, 2018.
- [186] Huaguang Li, Rencan Nie, Jinde Cao, Xiaopeng Guo, Dongming Zhou, and Kangjian He. Multi-focus Image Fusion using U-shaped Networks with a Hybrid Objective. *IEEE Sensors Journal*, 1748(c):1–1, 2019.
- [187] Xiaopeng Guo, Rencan Nie, Jinde Cao, Dongming Zhou, Liye Mei, Kangjian He, Student Member, Rencan Nie, Jinde Cao, and Dongming Zhou.

- FuseGAN: Learning to fuse Multi-focus Image via Conditional Generative Adversarial Network. *IEEE Transactions on Multimedia*, 21(8):1–1, 2019.
- [188] Jun Huang, Zhuliang Le, Yong Ma, Xiaoguang Mei, and Fan Fan. A generative adversarial network with adaptive constraints for multi-focus image fusion. *Neural Computing and Applications*, 32(18):15119–15129, 2020.
- [189] Pan Wu, Limai Jiang, Zhen Hua, and Jinjiang Li. Multi-focus image fusion: Transformer and shallow feature attention matters. *Displays*, 76:102353, 2023.
- [190] Mining Li, Ronghao Pei, Tianyou Zheng, Yang Zhang, and Weiwei Fu. Fusiondiff: Multi-focus image fusion using denoising diffusion probabilistic models. *Expert Systems with Applications*, 238:121664, 2024.
- [191] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.
- [192] Hao Zhai and Yi Zhuang. Multi-focus image fusion method using energy of Laplacian and a deep neural network. *Applied Optics*, 59(6):1684–1694, 2020.
- [193] Vaidehi Deshmukh, Arti Khaparde, and Sana Shaikh. Multi-focus image fusion using deep belief network. In *International Conference on Information and Communication Technology for Intelligent Systems*, pages 233–241. Springer, 2018.
- [194] Fayed Lahoud and Sabine Süsstrunk. Fast and efficient zero-learning image fusion. *arXiv:1905.03590*, 2019.
- [195] Mostafa Amin-Naji, Ali Aghagolzadeh, and Mehdi Ezoji. Ensemble of CNN for Multi-Focus Image Fusion. *Information Fusion*, 51:201–214, 2019.
- [196] Mostafa Amin-Naji, Ali Aghagolzadeh, and Mehdi Ezoji. CNNs hard voting for multi-focus image fusion. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–21, 2019.

- [197] Boyuan Ma, Yu Zhu, Xiang Yin, Xiaojuan Ban, Haiyou Huang, and Michele Mukeshimana. SESF-Fuse: An unsupervised deep model for multi-focus image fusion. *Neural Computing and Applications*, pages 1–12, 2020.
- [198] Xiang Yan, Syed Zulqarnain Gilani, Hanlin Qin, and Ajmal Mian. Structural similarity loss for learning to fuse multi-focus images. *Sensors*, 20(22):6647, 2020.
- [199] Hafiz Tayyab Mustafa, Fanghui Liu, Jie Yang, Zubair Khan, and Qiao Huang. Dense multi-focus fusion net: A deep unsupervised convolutional network for multi-focus image fusion. In *International Conference on Artificial Intelligence and Soft Computing*, pages 153–163. Springer, 2019.
- [200] Hao Zhang, Zhuliang Le, Zhenfeng Shao, Han Xu, and Jiayi Ma. MFF-GAN: An unsupervised generative adversarial network with adaptive and gradient joint constraints for multi-focus image fusion. *Information Fusion*, 66:40–53, 2021.
- [201] Mark Everingham, SM Ali Eslami, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 111(1):98–136, 2015.
- [202] Juncheng Zhang, Qingmin Liao, Shaojun Liu, Haoyu Ma, Wenming Yang, and Jing-Hao Xue. Real-mff: A large realistic multi-focus image dataset with ground truth. *Pattern Recognition Letters*, 138:370–377, 2020.
- [203] Tsung-Yi Lin, Michael Maire, Serge Belongie, et al. Microsoft COCO: Common objects in context. In *European Conference on Computer Vision*, pages 740–755. Springer, 2014.
- [204] Peter J Burt and Raymond J Kolczynski. Enhanced image capture through fusion. In *1993 (4th) international Conference on Computer Vision*, pages 173–182. IEEE, 1993.

- [205] Yi Yang, Wei Cao, Shiqian Wu, and Zhengguo Li. Multi-scale fusion of two large-exposure-ratio images. *IEEE Signal Processing Letters*, 25(12):1885–1889, 2018.
- [206] Sheng-Yeh Chen and Yung-Yu Chuang. Deep exposure fusion with deghosting via homography estimation and attention learning. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1464–1468. IEEE, 2020.
- [207] K Ram Prabhakar, V Sai Srikar, and R Venkatesh Babu. Deepfuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs. In *2017 IEEE International Conference on Computer Vision (ICCV). IEEE*, pages 4724–4732, 2017.
- [208] Jinhua Wang, Weiqiang Wang, Guangmei Xu, and Hongzhe Liu. End-to-end exposure fusion using convolutional neural network. *IEICE TRANSACTIONS on Information and Systems*, 101(2):560–563, 2018.
- [209] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.
- [210] Jianrui Cai, Shuhang Gu, and Lei Zhang. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Transactions on Image Processing*, 27(4):2049–2062, 2018.
- [211] Kede Ma, Zhengfang Duanmu, Hojatollah Yeganeh, and Zhou Wang. Multi-exposure image fusion by optimizing a structural similarity index. *IEEE Transactions on Computational Imaging*, 4(1):60–72, 2017.
- [212] Zhiguang Yang, Youping Chen, Zhuliang Le, and Yong Ma. Ganfuse: a novel multi-exposure image fusion method based on generative adversarial networks. *Neural Computing and Applications*, pages 1–13, 2020.
- [213] Linhao Qu, Shaolei Liu, Manning Wang, and Zhijian Song. Transmef: A transformer-based multi-exposure image fusion framework using self-

- supervised multi-task learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 2126–2134, 2022.
- [214] Jinyuan Liu, Zhu Liu, Guanyao Wu, Long Ma, Risheng Liu, Wei Zhong, Zhongxuan Luo, and Xin Fan. Multi-interactive feature learning and a full-time multi-modality benchmark for image fusion and segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8115–8124, 2023.
- [215] Xin Deng, Yutong Zhang, Mai Xu, Shuhang Gu, and Yiping Duan. Deep coupled feedback network for joint exposure fusion and image super-resolution. *IEEE Transactions on Image Processing*, 30:3098–3112, 2021.
- [216] Nicholas J Swerdlow, Douglas W Jones, Alexander B Pothof, Thomas FX O’Donnell, Patric Liang, Chun Li, Mark C Wyers, and Marc L Schermerhorn. Three-dimensional image fusion is associated with lower radiation exposure and shorter time to carotid cannulation during carotid artery stenting. *Journal of vascular surgery*, 69(4):1111–1120, 2019.
- [217] Jochen von Spiczak, Manoj Mannil, Hanna Model, Chris Schwemmer, Sebastian Kozerke, Frank Ruschitzka, Hatem Alkadhi, and Robert Manka. Multimodal multiparametric three-dimensional image fusion in coronary artery disease: Combining the best of two worlds. *Radiology: Cardiothoracic Imaging*, 2(2):e190116, 2020.
- [218] Kaiming He, Jian Sun, and Xiaou Tang. Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(6):1397–1409, 2012.
- [219] 唐霖峰, 张浩, 徐涵, and 马佳义. 基于深度学习的图像融合方法综述. 中国图象图形学报, 2023.
- [220] Yu Liu, Yu Shi, Fuhao Mu, Juan Cheng, and Xun Chen. Glioma segmentation-oriented multi-modal mr image fusion with adversarial learning. *IEEE/CAA Journal of Automatica Sinica*, 9(8):1528–1531, 2022.

- [221] Qiao Liu, Di Yuan, Nana Fan, Peng Gao, Xin Li, and Zhenyu He. Learning dual-level deep representation for thermal infrared tracking. *IEEE Transactions on Multimedia*, 25:1269–1281, 2022.
- [222] Qiao Liu, Xin Li, Zhenyu He, Chenglong Li, Jun Li, Zikun Zhou, Di Yuan, Jing Li, Kai Yang, Nana Fan, et al. Lsotb-tir: A large-scale high-diversity thermal infrared object tracking benchmark. In *Proceedings of the 28th ACM international conference on multimedia*, pages 3847–3856, 2020.
- [223] Qiao Liu, Xin Li, Di Yuan, Chao Yang, Xiaojun Chang, and Zhenyu He. Lsotb-tir: A large-scale high-diversity thermal infrared single object tracking benchmark. *IEEE Transactions on Neural Networks and Learning Systems*, 2023.
- [224] Meng Ding, Yuanyuan Ding, Li Wei, Yiming Xu, and Yunfeng Cao. Individual surveillance around parked aircraft at nighttime: Thermal infrared vision-based human action recognition. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 53(2):1084–1094, 2022.
- [225] Raluca Didona Brehar, Mircea Paul Muresan, Tiberiu Mariță, Cristian-Cosmin Vancea, Mihai Negru, and Sergiu Nedevschi. Pedestrian street-cross action recognition in monocular far infrared sequences. *IEEE Access*, 9:74302–74324, 2021.
- [226] Raluca Didona Brehar, Cristian Cosmin Vancea, Mircea Paul Mureșan, Sergiu Nedevschi, and Radu Dănescu. Pose based pedestrian street cross action recognition in infrared images. In *2021 IEEE 17th International Conference on Intelligent Computer Communication and Processing (ICCP)*, pages 41–46. IEEE, 2021.
- [227] Fatih Altay and Senem Velipasalar. The use of thermal cameras for pedestrian detection. *IEEE Sensors Journal*, 22(12):11489–11498, 2022.
- [228] Zhewei Xu, Jiajun Zhuang, Qiong Liu, Jingkai Zhou, and Shaowu Peng. Benchmarking a large-scale fir dataset for on-road pedestrian detection. *Infrared Physics & Technology*, 96:199–208, 2019.

- [229] Meng Ding, Wen-Hua Chen, and Yun-Feng Cao. Thermal infrared single-pedestrian tracking for advanced driver assistance system. *IEEE Transactions on Intelligent Vehicles*, 8(1):814–824, 2022.
- [230] Qiao Liu, Zhenyu He, Xin Li, and Yuan Zheng. Ptbtir: A thermal infrared pedestrian tracking benchmark. *IEEE Transactions on Multimedia*, 2019.
- [231] Zitian Tang, Wenjie Ye, Wei-Chiu Ma, and Hang Zhao. What happened 3 seconds ago? inferring the past with thermal imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17111–17120, 2023.
- [232] Benjamin Ronald van Manen, Victor Sluiter, and Abeje Yenehun Mersha. Firebotslam: thermal slam to increase situational awareness in smoke-filled environments. *Sensors*, 23(17):7611, 2023.
- [233] Hen-Wei Huang, Jack Chen, Peter R Chai, Claas Ehmke, Philipp Rupp, Farah Z Dadabhoy, Annie Feng, Canchen Li, Akhil J Thomas, Marco da Silva, et al. Mobile robotic platform for contactless vital sign monitoring. *Cyborg and Bionic Systems*, 2022.
- [234] Sean Ward, Jordon Hensler, Bilal Alsalam, and Luis Felipe Gonzalez. Autonomous uavs wildlife detection using thermal imaging, predictive navigation and computer vision. In *2016 IEEE aerospace conference*, pages 1–8. IEEE, 2016.
- [235] Fanglin Bao, Xueji Wang, Shree Hari Sureshbabu, Gautam Sreekumar, Liping Yang, Vaneet Aggarwal, Vishnu N Boddeti, and Zubin Jacob. Heat-assisted detection and ranging. *Nature*, 619(7971):743–748, 2023.
- [236] Mohamed Amine Marnissi and Abir Fathallah. Gan-based vision transformer for high-quality thermal image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 817–825, 2023.
- [237] Housheng Xie, Yukuan Zhang, Junhui Qiu, Xiangshuai Zhai, Xuedong Liu, Yang Yang, Shan Zhao, Yongfang Luo, and Jianbo Zhong. Semantics lead

- all: Towards unified image registration and fusion from a semantic perspective. *Information Fusion*, page 101835, 2023.
- [238] Sandra Pozzer, Marcos Paulo Vieira de Souza, Bata Hena, Reza Khoshkbary Rezayiye, Setayesh Hesam, Fernando Lopez, and Xavier Maldague. Defect segmentation in concrete structures combining registered infrared and visible images: A comparative experimental study. *Engineering Proceedings*, 8(1):29, 2021.
- [239] Sandra Pozzer, Marcos Paulo Vieira De Souza, Bata Hena, Setayesh Hesam, Reza Khoshkbary Rezayiye, Ehsan Rezazadeh Azar, Fernando Lopez, and Xavier Maldague. Effect of different imaging modalities on the performance of a CNN: An experimental study on damage segmentation in infrared, visible, and fused images of concrete structures. *NDT & E International*, 132:102709, 2022.
- [240] Xuyang Gao, Yibing Shi, Qi Zhu, Qiang Fu, and Yuezhou Wu. Infrared and visible image fusion with deep neural network in enhanced flight vision system. *Remote Sensing*, 14(12):2789, 2022.
- [241] Tanmay Kacker, Adolfo Perrusquia, and Weisi Guo. Multi-spectral fusion using generative adversarial networks for uav detection of wild fires. In *2023 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pages 182–187. IEEE, 2023.
- [242] Andong Lu, Cun Qian, Chenglong Li, Jin Tang, and Liang Wang. Duality-gated mutual condition network for rgbt tracking. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.
- [243] Pengyu Zhang, Dong Wang, Huchuan Lu, and Xiaoyun Yang. Learning Adaptive Attribute-Driven Representation for Real-Time RGB-T Tracking. *International Journal of Computer Vision*, pages 1–16, 2021.
- [244] Xingchen Zhang, Ping Ye, Jun Liu, Ke Gonge, and Gang Xiao. Decision-level visible and infrared fusion tracking via siamese networks. In *Proceedings of the 9th Chinese Conference on Information Fusion*, 2019.

- [245] Johan Vertens, Jannik Zürn, and Wolfram Burgard. Heatnet: Bridging the day-night domain gap in semantic segmentation with thermal images. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 8461–8468. IEEE, 2020.
- [246] V John, A Boyali, S Thompson, and S Mita. Bvtnet: Multi-label multi-class fusion of visible and thermal camera for free space and pedestrian segmentation. 2021.
- [247] Qiang Zhang, Shenlu Zhao, Yongjiang Luo, Dingwen Zhang, Nianchang Huang, and Jungong Han. Abmdrnet: Adaptive-weighted bi-directional modality difference reduction network for rgb-t semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2633–2642, June 2021.
- [248] Yuqi Li, Haitao Zhao, Zhengwei Hu, Qianqian Wang, and Yuru Chen. IV-FuseNet: Fusion of infrared and visible light images for depth prediction. *Information Fusion*, 58:1–12, 2020.
- [249] Karasawa Takumi, Kohei Watanabe, Qishen Ha, Antonio Tejero-De-Pablos, Yoshitaka Ushiku, and Tatsuya Harada. Multispectral object detection for autonomous vehicles. In *Proceedings of the on Thematic Workshops of ACM Multimedia 2017*, pages 35–43, 2017.
- [250] Ravi Yadav, Ahmed Samir, Hazem Rashed, Senthil Yogamani, and Rozenn Dahyot. CNN based Color and Thermal Image Fusion for Object Detection in Automated Driving. (July), 2020.
- [251] Heng Zhang, Elisa Fromont, Sébastien Lefèvre, and Bruno Avignon. Multispectral fusion for object detection with cyclic fuse-and-refine blocks. In *2020 IEEE International Conference on Image Processing (ICIP)*, pages 276–280. IEEE, 2020.
- [252] Joaquín Filipic, Martín Biagini, Ignacio Mas, et al. People counting using visible and infrared images. *Neurocomputing*, 450:25–32, 2021.

- [253] Tao Peng, Qing Li, and Pengfei Zhu. RGB-T Crowd Counting from Drone: A Benchmark and MMCCN Network. In *Proceedings of the Asian Conference on Computer Vision*, 2020.
- [254] Lingbo Liu, Jiaqi Chen, Hefeng Wu, Guanbin Li, Chenglong Li, and Liang Lin. Cross-modal collaborative representation learning and a large-scale rgbt benchmark for crowd counting. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4823–4833, 2021.
- [255] Praveen Kumar, Ankush Mittal, and Padam Kumar. Fusion of thermal infrared and visible spectrum video for robust surveillance. In *Computer Vision, Graphics and Image Processing*, pages 528–539. Springer, 2006.
- [256] D Bhavana, K Kishore Kumar, and D Ravi Tej. Infrared and visible image fusion using latent low rank technique for surveillance applications. *International Journal of Speech Technology*, pages 1–10, 2021.
- [257] Toshiaki Negishi, Shigeto Abe, Takemi Matsui, He Liu, Masaki Kurosawa, Tetsuo Kirimoto, and Guanghao Sun. Contactless vital signs measurement system using rgb-thermal image sensors and its clinical screening test on patients with seasonal influenza. *Sensors*, 20(8):2171, 2020.
- [258] Juncun Wei, Jiancheng Zou, Jiaxin Li, Zhengzheng Li, and Xin Yang. Non-contact heart rate detection based on fusion method of visible images and infrared images. In *International Conference on Artificial Intelligence and Security*, pages 62–75. Springer, 2022.
- [259] Usman Cheema, Mobeen Ahmad, Dongil Han, and Seungbin Moon. Heterogeneous visible-thermal and visible-infrared face recognition using unit-class loss and cross-modality discriminator. *arXiv preprint arXiv:2111.14339*, 2021.
- [260] Xianglong Chen, Haipeng Wang, Yaohui Liang, Ying Meng, and Shifeng Wang. A novel infrared and visible image fusion approach based on adversarial neural network. *Sensors*, 22(1):304, 2022.

- [261] JF Ciprián-Sánchez, G Ochoa-Ruiz, M Gonzalez-Mendoza, and L Rossi. Assessing the applicability of deep learning-based visible-infrared fusion methods for fire imagery. *arXiv preprint arXiv:2101.11745*, 2021.
- [262] Xiang Xu, Gang Liu, Durga Prasad Bavirisetti, Xiangbo Zhang, Boyang Sun, and Gang Xiao. Fast detection fusion network (fdfnet): An end to end object detection framework based on heterogeneous image fusion for power facility inspection. *IEEE Transactions on Power Delivery*, 2022.
- [263] Xihong Zhou, Gang Liu, Xiangbo Zhang, Bavirisetti Durga Prasad, Xinjie Gu, and Yonghua Li. Re2fad: A differential image registration and robust image fusion method framework for power thermal anomaly detection. *Optik*, page 168817, 2022.
- [264] Jiale Ma, Kun Qian, Xiaobo Zhang, and Xudong Ma. Weakly supervised instance segmentation of electrical equipment based on RGB-T automatic annotation. *IEEE Transactions on Instrumentation and Measurement*, 69(12):9720–9731, 2020.
- [265] Wei Gao, Guibiao Liao, Siwei Ma, Ge Li, Yongsheng Liang, and Weisi Lin. Unified information fusion network for multi-modal rgb-d and rgb-t salient object detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021.
- [266] Dong Huo, Jian Wang, Yiming Qian, and Yee-Hong Yang. Glass segmentation with rgb-thermal image pairs. *arXiv preprint arXiv:2204.05453*, 2022.
- [267] Vijay John, Ali Boyali, Simon Thompson, Annamalai Lakshmanan, and Seiichi Mita. Visible and thermal camera-based jaywalking estimation using a hierarchical deep learning framework. In *Proceedings of the Asian Conference on Computer Vision*, 2020.
- [268] Quincy G Alexander, Vedhus Hoskere, Yasutaka Narazaki, Andrew Maxwell, and Billie F Spencer. Fusion of thermal and rgb images for automated deep learning based crack detection in civil infrastructure. *AI in Civil Engineering*, 1(1):1–10, 2022.

- [269] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.
- [270] Weichen Dai, Yu Zhang, Shenzhou Chen, Donglei Sun, and Da Kong. A multi-spectral dataset for evaluating motion estimation systems. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5560–5566. IEEE, 2021.
- [271] Petru Manescu, Michael Shaw, Lydia Neary-Zajiczek, Christopher Bendkowski, Remy Claveau, Muna Elmi, Biobele J Brown, and Delmiro Fernandez-Reyes. Content aware multi-focus image fusion for high-magnification blood film microscopy. *Biomedical optics express*, 13(2):1005–1016, 2022.
- [272] Ronghao Pei, Weiwei Fu, Kang Yao, Tianli Zheng, Shangshang Ding, Hetong Zhang, and Yang Zhang. Real-time multi-focus biomedical microscopic image fusion based on m-segnet. *IEEE Photonics Journal*, 2021.
- [273] R Hurtado-Pérez, C Toxqui-Quitl, A Padilla-Vivanco, and G Ortega-Mendoza. Extending the depth-of-field for microscopic imaging by means of multifocus color image fusion. In *Current Developments in Lens Design and Optical Engineering XVI*, volume 9578, pages 191–199. SPIE, 2015.
- [274] Zhiyu Chen, Dong Wang, Shaoyan Gong, and Feng Zhao. Application of multi-focus image fusion in visual power patrol inspection. In *2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference*, pages 1688–1692. IEEE, 2017.
- [275] Ramachandra Raghavendra, Kiran B Raja, Bian Yang, and Christoph Busch. Multi-face recognition at a distance using light-field camera. In *2013 Ninth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pages 346–349. IEEE, 2013.

- [276] Ramachandra Raghavendra, Bian Yang, Kiran B Raja, and Christoph Busch. A new perspective—face recognition with light-field camera. In *2013 International Conference on Biometrics (ICB)*, pages 1–8. IEEE, 2013.
- [277] Tao Yan, Zhiguo Hu, Yuhua Qian, Zhiwei Qiao, and Linyuan Zhang. 3D shape reconstruction from multifocus image fusion using a multidirectional modified Laplacian operator. *Pattern Recognition*, 98:107065, 2020.
- [278] Ning-Hsu Wang, Ren Wang, Yu-Lun Liu, Yu-Hao Huang, Yu-Lin Chang, Chia-Ping Chen, and Kevin Jou. Bridging unsupervised and supervised depth from focus via all-in-focus supervision. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 12621–12631, 2021.
- [279] Tengchao Huang, Shuang Song, Qianjie Liu, Wei He, Qingyuan Zhu, and Huosheng Hu. A novel multi-exposure fusion approach for enhancing visual semantic segmentation of autonomous driving. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, page 09544070221097851, 2022.
- [280] Kede Ma, Zhengfang Duanmu, Hanwei Zhu, Yuming Fang, and Zhou Wang. Deep guided learning for fast multi-exposure image fusion. *IEEE Transactions on Image Processing*, 29:2808–2819, 2019.
- [281] Peng Ke, Cheolkon Jung, and Ying Fang. Perceptual multi-exposure image fusion with overall image quality index and local saturation. *Multimedia Systems*, 23:239–250, 2017.
- [282] Harbinder Singh, Gabriel Cristobal, Gloria Bueno, Saul Blanco, Simran-deep Singh, PN Hrisheekesha, and Nitin Mittal. Multi-exposure microscopic image fusion-based detail enhancement algorithm. *Ultramicroscopy*, 236:113499, 2022.
- [283] Harbinder Singh, Gabriel Cristóbal, and Vinay Kumar. Multifocus and multiexposure techniques. *Modern trends in diatom identification: Fundamentals and applications*, pages 165–181, 2020.

- [284] Huafeng Li, Xiaoge He, Dapeng Tao, Yuanyan Tang, and Ruxin Wang. Joint medical image fusion, denoising and enhancement via discriminative low-rank sparse dictionaries learning. *Pattern Recognition*, 79:130–146, 2018.
- [285] Huafeng Li, Moyuan Yang, and Zhengtao Yu. Joint image fusion and super-resolution for enhanced visualization via semi-coupled discriminative dictionary learning and advantage embedding. *Neurocomputing*, 422:62–84, 2021.
- [286] Xaosong Li, Fuqiang Zhou, and Haishu Tan. Joint image fusion and denoising via three-layer decomposition and sparse representation. *Knowledge-Based Systems*, 224:107087, 2021.
- [287] Zhuliang Le, Jun Huang, Han Xu, Fan Fan, Yong Ma, Xiaoguang Mei, and Jiayi Ma. Uifgan: An unsupervised continual-learning generative adversarial network for unified image fusion. *Information Fusion*, 2022.
- [288] MA Herrera-Arellano, Hayde Peregrina-Barreto, and Iván Terol-Villalobos. Visible-nir image fusion based on top-hat transform. *IEEE Transactions on Image Processing*, 2021.
- [289] Jinyuan Liu, Xin Fan, Ji Jiang, Risheng Liu, and Zhongxuan Luo. Learning a deep multi-scale feature ensemble and an edge-attention guidance for image fusion. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021.
- [290] Zhuo Li, Hai-Miao Hu, Wei Zhang, Shiliang Pu, and Bo Li. Spectrum characteristics preserved visible and near-infrared image fusion algorithm. *IEEE Transactions on Multimedia*, 23:306–319, 2020.
- [291] Alex E Hayes, Graham D Finlayson, and Roberto Montagna. Rgb-nir image fusion: metric and psychophysical experiments. In *Image Quality and System Performance XII*, volume 9396, page 93960U. International Society for Optics and Photonics, 2015.
- [292] Takashi Shibata, Masayuki Tanaka, and Masatoshi Okutomi. Visible and near-infrared image fusion based on visually salient area selection. In *Digital*

- Photography XI*, volume 9404, page 94040G. International Society for Optics and Photonics, 2015.
- [293] Vivek Sharma, Jon Yngve Hardeberg, and Sony George. Rgb-nir image enhancement by fusing bilateral and weighted least squares filters. *Journal of Imaging Science and Technology*, 61(4):40409–1, 2017.
- [294] Ashish V Vanmali and Vikram M Gadre. Visible and nir image fusion using weight-map-guided laplacian–gaussian pyramid for improving scene visibility. *Sādhanā*, 42(7):1063–1082, 2017.
- [295] Cheolkon Jung, Kailong Zhou, and Jiawei Feng. Fusionnet: Multispectral fusion of rgb and nir images using two stage convolutional neural networks. *IEEE Access*, 8:23912–23919, 2020.
- [296] Lex Schaul, Clément Fredembach, and Sabine Süsstrunk. Color image dehazing using the near-infrared. In *2009 16th IEEE International Conference on Image Processing (ICIP)*, pages 1629–1632. IEEE, 2009.
- [297] Faten Omri, Sebti Foufou, and Mongi Abidi. Nir and visible image fusion for improving face recognition at long distance. In *International Conference on Image and Signal Processing*, pages 549–557. Springer, 2014.
- [298] Duncan L Hickman. Colour fusion of rgb and nir imagery for surveillance applications. In *Electro-Optical and Infrared Systems: Technology and Applications XVII*, volume 11537, page 115370H. International Society for Optics and Photonics, 2020.
- [299] Hyuk-Ju Kwon and Sung-Hak Lee. Visible and near-infrared image acquisition and fusion for night surveillance. *Chemosensors*, 9(4):75, 2021.
- [300] Viviana Crescitelli, Atsutake Kosuge, and Takashi Oshima. An RGB/infrared camera fusion approach for multi-person pose estimation in low light environments. In *IEEE Sensors Applications Symposium*, pages 1–6. IEEE, 2020.

- [301] Viviana Crescitelli, Atsutake Kosuge, and Takashi Oshima. Poison: Human pose estimation in insufficient lighting conditions using sensor fusion. *IEEE Transactions on Instrumentation and Measurement*, 70:1–8, 2021.
- [302] Zhongli Ma, Jie Wen, Quanyong Liu, and Guanjun Tuo. Near-infrared and visible light image fusion algorithm for face recognition. *Journal of Modern Optics*, 62(9):745–753, 2015.
- [303] Jian Liang, Wenfei Zhang, Liyong Ren, Haijuan Ju, and Enshi Qu. Polarimetric dehazing method for visibility improvement based on visible and infrared image fusion. *Applied optics*, 55(29):8221–8226, 2016.
- [304] Zhihao Liu, Jingzhu Wu, Longsheng Fu, Yaqoob Majeed, Yali Feng, Rui Li, and Yongjie Cui. Improved kiwifruit detection using pre-trained vgg16 with rgb and nir information fusion. *IEEE Access*, 8:2327–2336, 2019.
- [305] 王霞, 赵家碧, 孙晶, and 金伟其. 偏振图像融合技术综述. *航天返回与遥感*, 42(6):9–21, 2021.
- [306] Han Xu, Yucheng Sun, Xiaoguang Mei, Xin Tian, and Jiayi Ma. Attention-guided polarization image fusion using salient information distribution. *IEEE Transactions on Computational Imaging*, 8:1117–1130, 2022.
- [307] Junchao Zhang, Jianbo Shao, Jianlai Chen, Degui Yang, and Buge Liang. Polarization image fusion with self-learned fusion strategy. *Pattern Recognition*, page 108045, 2021.
- [308] Fabian Duffhauss, Ngo Anh Vien, Hanna Ziesche, and Gerhard Neumann. FusionVAE: A Deep Hierarchical Variational Autoencoder for RGB Image Fusion. In *ECCV*, 2022.
- [309] Xudong Sun, Yuan Zhu, and Xianping Fu. Rgb and optimal waveband image fusion for real-time underwater clear image acquisition. *IEEE Transactions on Instrumentation and Measurement*, 2023.
- [310] Pushmeet Kohli Nathan Silberman, Derek Hoiem and Rob Fergus. Indoor segmentation and support inference from rgbd images. In *ECCV*, 2012.

- [311] Charles R Qi, Wei Liu, Chenxia Wu, Hao Su, and Leonidas J Guibas. Frustum pointnets for 3d object detection from rgb-d data. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 918–927, 2018.
- [312] Michael Zollhöfer, Patrick Stotko, Andreas Görilitz, Christian Theobalt, Matthias Nießner, Reinhard Klein, and Andreas Kolb. State of the art on 3d reconstruction with rgb-d cameras. In *Computer graphics forum*, volume 37, pages 625–652. Wiley Online Library, 2018.
- [313] Yuxiang Sun, Ming Liu, and Max Q-H Meng. Motion removal for reliable rgb-d slam in dynamic environments. *Robotics and Autonomous Systems*, 108:115–128, 2018.
- [314] Gongyang Li, Zhi Liu, and Haibin Ling. Icnet: Information conversion network for rgb-d based salient object detection. *IEEE Transactions on Image Processing*, 29:4873–4884, 2020.
- [315] Deng-Ping Fan, Zheng Lin, Zhao Zhang, Menglong Zhu, and Ming-Ming Cheng. Rethinking rgb-d salient object detection: Models, data sets, and large-scale benchmarks. *IEEE Transactions on neural networks and learning systems*, 32(5):2075–2089, 2020.
- [316] Song Yan, Jinyu Yang, Jani Käpylä, Feng Zheng, Aleš Leonardis, and Joni-Kristian Kämäräinen. Depthtrack: Unveiling the power of rgbd tracking. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10725–10733, 2021.
- [317] Norman Muller, Yu-Shiang Wong, Niloy J Mitra, Angela Dai, and Matthias Nießner. Seeing behind objects for 3d multi-object tracking in rgbd sequences. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6071–6080, 2021.
- [318] Linhui Li, Bo Qian, Jing Lian, Weina Zheng, and Yafu Zhou. Traffic scene segmentation based on rgbd image and deep learning. *IEEE Transactions on Intelligent Transportation Systems*, 19(5):1664–1669, 2017.

- [319] Hao Zhou, Lu Qi, Hai Huang, Xu Yang, Zhaoliang Wan, and Xianglong Wen. Canet: Co-attention network for rgb-d semantic segmentation. *Pattern Recognition*, 124:108468, 2022.
- [320] Christian Zimmermann, Tim Welschehold, Christian Dornhege, Wolfram Burgard, and Thomas Brox. 3d human pose estimation in rgbd images for robotic task learning. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1986–1992. IEEE, 2018.
- [321] Renat Bashirov, Anastasia Ianina, Karim Iskakov, Yevgeniy Kononenko, Valeriya Strizhkova, Victor Lempitsky, and Alexander Vakhitov. Real-time rgbd-based extended body pose estimation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2807–2816, 2021.
- [322] Fan Zhang and Yiannis Demiris. Learning garment manipulation policies toward robot-assisted dressing. *Science robotics*, 7(65):eabm6010, 2022.
- [323] Minghao Gou, Hao-Shu Fang, Zhanda Zhu, Sheng Xu, Chenxi Wang, and Cewu Lu. Rgb matters: Learning 7-dof grasp poses on monocular rgbd images. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 13459–13466. IEEE, 2021.
- [324] Xiao Wang, Jianing Li, Lin Zhu, Zhipeng Zhang, Zhe Chen, Xin Li, Yaowei Wang, Yonghong Tian, and Feng Wu. Visevent: Reliable object tracking via collaboration of frame and event flows. *arXiv preprint arXiv:2108.05015*, 2021.
- [325] Yi-Fan Zuo, Li Cui, Xin Peng, Yanyu Xu, Shenghua Gao, Xia Wang, and Laurent Kneip. Accurate depth estimation from a hybrid event-rgb stereo setup. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 6833–6840. IEEE, 2021.
- [326] Abhishek Tomy, Anshul Paigwar, Khushdeep Singh Mann, Alessandro Renzaglia, and Christian Laugier. Fusing event-based and rgb camera for robust object detection in adverse conditions. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2022.

- [327] Huayao Liu, Jiaming Zhang, Kailun Yang, Xinxin Hu, and Rainer Stiefelhagen. Cmx: Cross-modal fusion for rgb-x semantic segmentation with transformers. *arXiv preprint arXiv:2203.04838*, 2022.
- [328] Jiaming Zhang, Kailun Yang, and Rainer Stiefelhagen. Exploring event-driven dynamic context for accident scene segmentation. *IEEE Transactions on Intelligent Transportation Systems*, 23(3):2606–2622, 2021.
- [329] Hu Cao, Guang Chen, Zhijun Li, Yingbai Hu, and Alois Knoll. Neurograsp: Multi-modal neural network with euler region regression for neuromorphic vision-based grasp pose estimation. *IEEE Transactions on Instrumentation and Measurement*, 2022.
- [330] Xiurong Jiang, Lin Zhu, and Hui Tian. Learning event guided network for salient object detection. *Pattern Recognition Letters*, 151:317–324, 2021.
- [331] Georgios Pilikos, Lars Horchens, Tristan van Leeuwen, and Felix Lucka. Deep learning for multi-view ultrasonic image fusion. *arXiv preprint arXiv:2109.03616*, 2021.
- [332] Luis G Riera, Matthew E Carroll, Zhisheng Zhang, Johnathon M Shook, Sambuddha Ghosal, Tianshuang Gao, Arti Singh, Sourabh Bhattacharya, Baskar Ganapathysubramanian, Asheesh K Singh, et al. Deep multiview image fusion for soybean yield estimation in breeding applications. *Plant Phenomics*, 2021.
- [333] Yancheng Wang, Yang Xiao, Junyi Lu, Bo Tan, Zhiguo Cao, Zhenjun Zhang, and Joey Tianyi Zhou. Discriminative Multi-View Dynamic Image Fusion for Cross-View 3-D Action Recognition. *IEEE Transactions on Neural Networks and Learning Systems*, 2021.
- [334] Wu Wang, Xueyang Fu, Weihong Zeng, Liyan Sun, Ronghui Zhan, Yue Huang, and Xinghao Ding. Enhanced deep blind hyperspectral image fusion. *IEEE transactions on neural networks and learning systems*, 2021.
- [335] Jia-Li Yin, Bo-Hao Chen, Yan-Tsung Peng, and Hau Hwang. Automatic intermediate generation with deep reinforcement learning for robust two-

- exposure image fusion. *IEEE Transactions on Neural Networks and Learning Systems*, 2021.
- [336] Fan Zhao et al. Learning specific and general realm feature representations for image fusion. *IEEE Transactions on Multimedia*, 2020.
- [337] Wenda Zhao, Dong Wang, and Huchuan Lu. Multi-focus image fusion with a natural enhancement via a joint multi-level deeply supervised convolutional neural network. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(4):1102–1115, 2019.
- [338] Zixiang Zhao, Shuang Xu, Jiangshe Zhang, Chengyang Liang, Chunxia Zhang, and Junmin Liu. Efficient and model-based infrared and visible image fusion via algorithm unrolling. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021.
- [339] Shuang Xu, Lizhen Ji, Zhe Wang, Pengfei Li, Kai Sun, Chunxia Zhang, and Jiangshe Zhang. Towards reducing severe defocus spread effects for multi-focus image fusion via an optimization based strategy. *IEEE Transactions on Computational Imaging*, 6:1561–1570, 2020.
- [340] Jingxiang Yang, Liang Xiao, Yong-Qiang Zhao, and Jonathan Cheung-Wai Chan. Variational regularization network with attentive deep prior for hyperspectral-multispectral image fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 2021.
- [341] Yuehong Chen, Kaixin Shi, Yong Ge, et al. Spatiotemporal remote sensing image fusion using multiscale two-stream convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 2021.
- [342] Fan Zhao, Wenda Zhao, Libo Yao, and Yu Liu. Self-supervised feature adaption for infrared and visible image fusion. *Information Fusion*, 2021.
- [343] Aiqing Fang, Xinbo Zhao, Jiaqi Yang, Yanning Zhang, and Xiang Zheng. Non-linear and selective fusion of cross-modal images. *Pattern Recognition*, 119:108042, 2021.

- [344] Heng Li, Liming Zhang, Meirong Jiang, and Yulong Li. Multi-focus image fusion algorithm based on supervised learning for fully convolutional neural network. *Pattern Recognition Letters*, 141:45–53, 2020.
- [345] Shi Yi, Junjie Li, and Xuesong Yuan. Dfpgan: Dual fusion path generative adversarial network for infrared and visible image fusion. *Infrared Physics & Technology*, 119:103947, 2021.
- [346] Bin Xiao, Haifeng Wu, and Xiuli Bi. Dtmnet: A discrete tchebichef moments-based deep neural network for multi-focus image fusion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 43–51, 2021.
- [347] Zhanbo Huang, Jinyuan Liu, Xin Fan, Risheng Liu, Wei Zhong, and Zhongxuan Luo. Reconet: Recurrent correction network for fast and efficient multi-modality image fusion. In *ECCV*, 2022.
- [348] Pengwei Liang, Junjun Jiang, Xianming Liu, and Jiayi Ma. Fusion from decomposition: A self-supervised decomposition approach for image fusion. In *ECCV*, 2022.
- [349] Sabine Süsstrunk, Clément Fredembach, and Daniel Tamburrino. Automatic skin enhancement with visible and near-infrared image fusion. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 1693–1696, 2010.
- [350] Zeeshan Ahmad, Anika Tabassum, Ling Guan, and Naimul Khan. Ecg heart-beat classification using multimodal image fusion. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1330–1334. IEEE, 2021.
- [351] Fei Kou, Zhengguo Li, Changyun Wen, and Weihai Chen. Multi-scale exposure fusion via gradient domain guided image filtering. In *2017 IEEE International Conference on Multimedia and Expo*, pages 1105–1110. IEEE, 2017.

- [352] Jinyuan Liu, Jingjie Shang, Risheng Liu, and Xin Fan. Halder: Hierarchical attention-guided learning with detail-refinement for multi-exposure image fusion. In *2021 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2021.
- [353] Kede Ma and Zhou Wang. Multi-exposure image fusion: A patch-wise approach. In *2015 IEEE International Conference on Image Processing*, pages 1717–1721. IEEE, 2015.
- [354] Sang-hoon Lee, Jae Sung Park, and Nam Ik Cho. A multi-exposure image fusion based on the adaptive weights reflecting the relative pixel intensity and global gradient. In *2018 25th IEEE International Conference on Image Processing*, pages 1737–1741. IEEE, 2018.
- [355] Fahimeh Farahnakian, Jussi Poikonen, Markus Laurinen, Dimitrios Makris, and Jukka Heikkonen. Visible and infrared image fusion framework based on retinanet for marine environment. In *2019 22th International Conference on Information Fusion*, pages 1–7. IEEE, 2019.
- [356] Durga Prasad Bavirisetti, Gang Xiao, Junhao Zhao, Xingchen Zhang, and Pengbo Wang. A new image and video fusion method based on cross bilateral filter. In *2018 21st International Conference on Information Fusion*, pages 1–8. IEEE, 2018.