文章编号：1672-8785(2025)05-0001-10

单目深度估计研究综述

王　诚　李梦媛　李春领

（华北光电技术研究所，北京　100015）

摘　要：单目深度估计在三维重建、目标跟踪、场景理解等众多应用中起到非常重要的作用。由于单目摄像头具有成本低、设备较为普及、图像获取方便等特点，从单目图像中获取深度信息成为热门研究。首先概述了用于单目深度估计的常见深度学习模型，主要包括卷积神经网络(Convolutional Neural Network，CNN)、循环神经网络(Recurrent Neural Network，RNN)和生成对抗网络(Generative Adversarial Network，GAN)。然后从训练方法的角度归纳了用于单目深度估计的深度学习方法，并对单目深度估计的发展趋势进行了总结。

关键词：单目深度估计；计算机视觉；深度学习

中图分类号：TP391.41　　**文献标志码**：A

DOI：11.3969/j.issn.1672-8785.2025.05.001

A Review of Monocular Depth Estimation Research

WANG Cheng，LI Meng-yuan，LI Chun-ling

(*North China Research Institute of Electro-Optics，Beijing 100015，China*)

Abstract：Monocular depth estimation plays a very important role in many applications such as 3D reconstruction，target tracking，and scene understanding. Since monocular cameras have the characteristics of low cost，widespread equipment，and convenient image acquisition，obtaining depth information from monocular images has become a hot research topic. First，the common deep learning models used for monocular depth estimation are summarized，mainly including convolutional neural network (CNN)，recurrent neural network (RNN)，and generative adversarial network (GAN). Then，the deep learning methods for monocular depth estimation are summarized from the perspective of training methods，and the development trend of monocular depth estimation is summarized.

Key words：monocular depth estimation；computer vision；deep learning

收稿日期：2024-12-05

作者简介：王诚(1986-)，男，江苏扬州人，高级工程师，主要从事光电系统方面的研究。

E-mail：kim2005wang@163.com

0 引言

场景深度估计在计算机视觉领域占据着举足轻重的地位，不仅增强了人们对真实三维场景的感知和理解，而且还推动了众多应用的发展，比如机器人导航、自动驾驶和虚拟现实等。传统主动深度估计方法(如利用激光、结构光等的反射在物体表面获取深度点云)虽然精度较高，但往往伴随着高昂的人力和计算成本。

相比之下，基于图像的深度估计方法以其较低的成本和广泛的应用前景，逐渐成为研究的主流。目前，该领域的发展已经经历了从依赖深度线索(如消失点、对焦与散焦、阴影等)的初级阶段，到引入手工设计的特征(如尺度不变特征转换(Scale-Invariant Feature Transform，SIFT)、加速稳健特征(Speeded-Up Robust Features，SURF))以及概率图模型(如条件随机场(Conditional Random Field，CRF)、马尔科夫随机场(Markov Random Field，MRF))的中期阶段，再到通过深度学习技术为图像处理和深度估计带来革命性变革的现阶段，如图 1 所示。

传统上，基于双目相机的深度估计方法通过计算两幅二维图像(由双目相机拍摄)的视差，并结合立体匹配和三角测量技术来获取深度图。然而，这种方法至少需要两台固定的相机，且当场景缺乏纹理时，很难捕捉到足够多的图像特征来进行匹配。

因此，研究者们开始关注单目深度估计方法。该方法仅使用单个相机来获取图像或视频序列，无需额外的复杂设备或专业技术。由于大多数应用场景中只需要单个相机，单目深度估计方法具有广泛的应用需求。

单目图像以二维形式反映了三维世界，然而由于缺乏可靠的立体视觉关系，我们无法判断物体的大小和距离，也无法判断一个物体是否被另一个物体遮挡。因此，我们需要恢复单目图像的深度，并基于深度图判断物体的大小

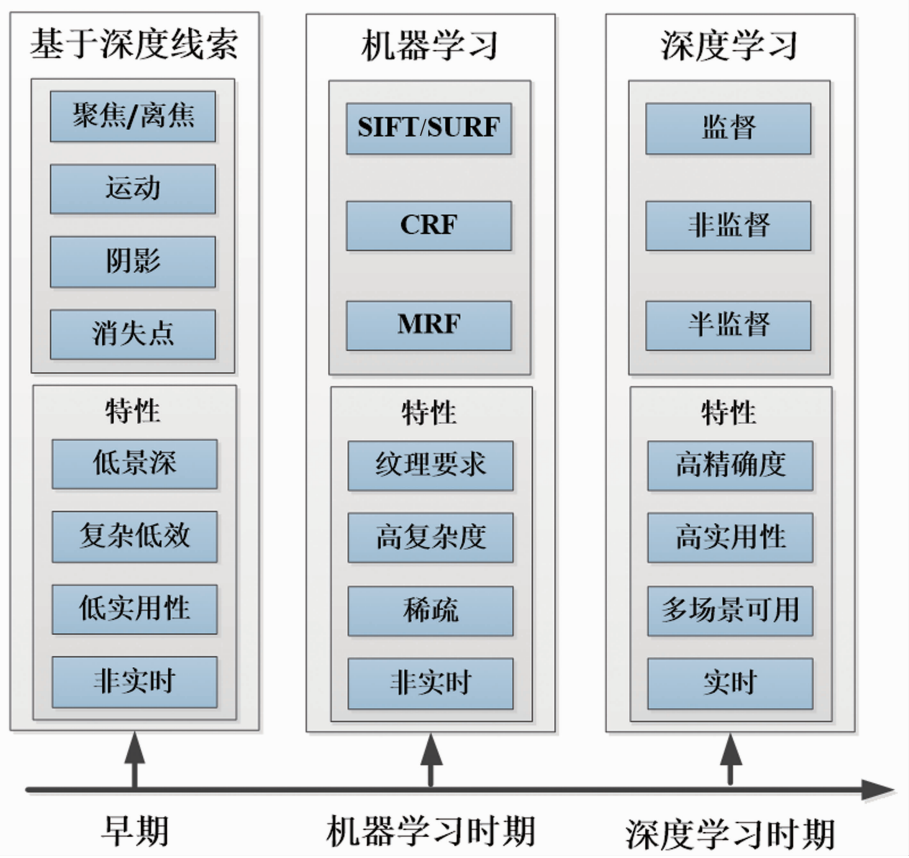

图 1 深度估计研究的发展

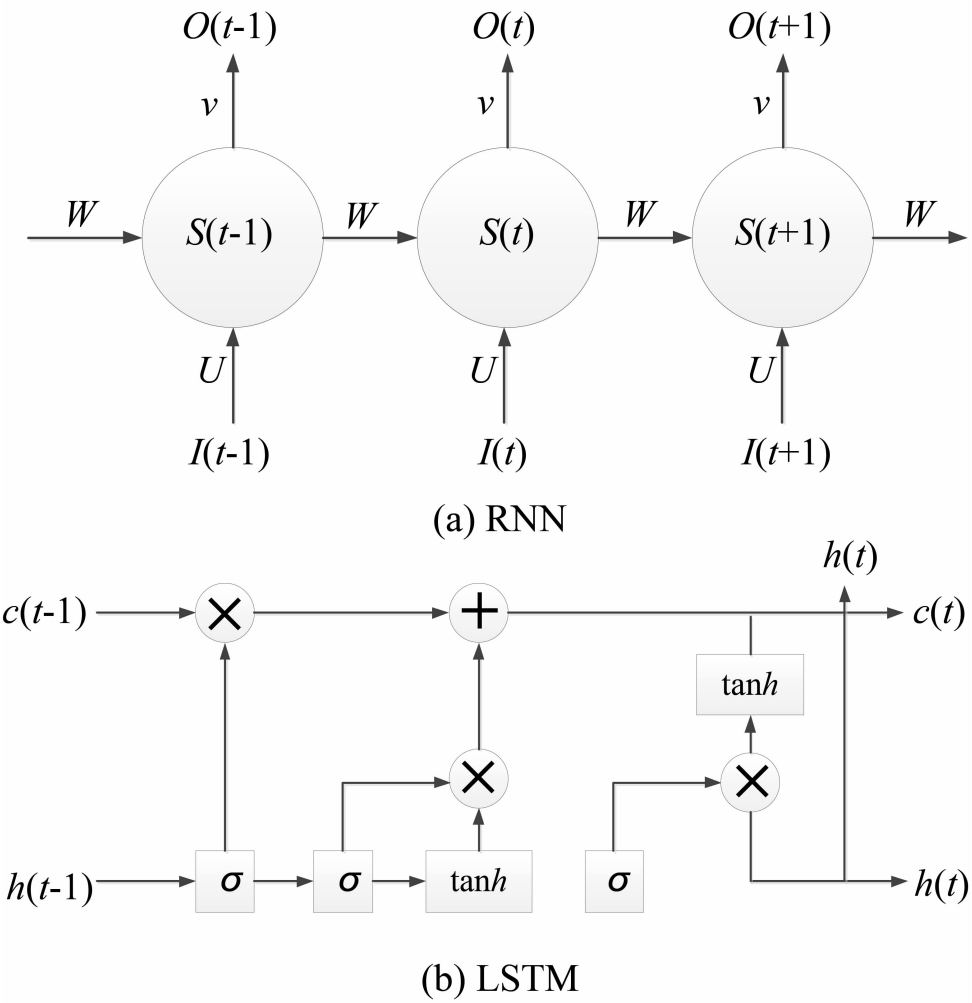

图 2　RNN 和 LSTM 示意图

和距离，以满足场景理解的需求。

本文主要聚焦于单目深度估计研究，对近年来基于深度学习的方法进行了综述。此外，本文还描述了现有方法的局限性，并展望了未来的发展趋势。

1　深度学习模型

1.1　CNN

CNN 非常擅长自动提取图像中的空间特征，包括场景的深度信息。它通过卷积层、池化层、全连接层和激活函数等结构，能够学习输入图像的二维空间特征。卷积层将输入数据转换为深度特征；池化层通过混合池化或平均池化的方式减小输入特征图的大小；全连接层通常位于 CNN 的末端，用于输出结果；激活函数则通常是一个连续可微的非线性函数，以避免纯线性组合。与传统方法相比，CNN 需要的参数更少，能够同时提取深度特征和重构深度图(卷积层在深度学习中的应用主要体现在特征提取上，它能够自动学习输入数据的局部特征，如边缘、纹理等)。

1.2　RNN

RNN 是一种序列到序列的模型，具有记忆能力，非常适合处理视频序列等时间序列数据，如图 2(a)所示。RNN 包括三个部分：输入单元、隐藏单元和输出单元。其中，隐藏单元的输入由当前输入单元的输出和前一隐藏单元的输出共同组成。此外，Hochreiter S 等人[1]提出了图 2(b)所示的长短期记忆(Long Short-Term Memory，LSTM)单元。它利用一个包含输入门层、遗忘门层和输出门层的三门结构来学习长期依赖关系。通过引入 LSTM 等单元，RNN 能够学习长期依赖关系。这非

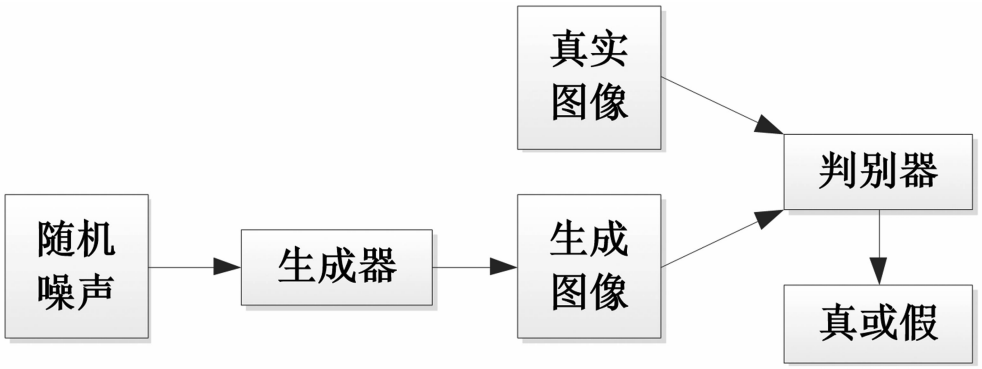

图 3 GAN 的基本框架

常有助于从视频序列中学习时间特征，进而实现单目深度估计。

1.3 GAN

GAN 在深度估计领域的应用确实为这一任务带来了许多创新和提升。GAN 通过独特的对抗性训练机制，使得生成的深度图更加清晰和逼真。这对缺乏真实场景时的深度图尤为重要。

GAN 包含两个主要模块：生成器和判别器。其中，生成器负责预测深度图，用作深度估计网络的核心部分；而判别器则负责判断输入的深度图是真实的还是由生成器生成的。这种对抗性训练的方式促使生成器不断优化其生成的深度图，以欺骗判别器，从而生成更加接近真实深度图的结果，如图 3 所示。

有几种具有代表性的 GAN 在深度估计中得到应用，包括条件 GAN、深度卷积 GAN、WGAN、堆叠 GAN、SimGAN 和 Cycle-GAN。其中，条件 GAN 可以引入额外的条件信息来指导生成过程，深度卷积 GAN 通过引入卷积层来改进 GAN 的架构，WGAN 通过改进损失函数来解决 GAN 训练中的不稳定问题，堆叠 GAN 通过堆叠多个 GAN 来提升生成质量，SimGAN 通过模拟和无监督学习来生成逼真的图像，而 Cycle-GAN 则通过引入循环一致性损失来实现无监督的图像到图像的转换。

在深度估计中，使用 GAN 模型可以提供对抗性约束，帮助模型更好地学习三维映射和尺度信息，从而生成更加准确和逼真的深度图。这种方法不仅可以缓解真实场景下深度图获取困难的问题，而且还可以提升深度估计模型的泛化能力和鲁棒性。

2 用于单目深度估计的深度学习方法

深度学习方法在单目深度估计中的应用是一个非常重要的研究领域。它旨在通过深度神经网络从单个二维彩色图像中学习出深度图，这一任务最初由 Eigen D 等人[2]在 2014 年提出。他们采用了一个由粗到细的框架，其中粗网络学习整个图像的全局深度以获得粗略的深度图，而细网络则学习局部特征以优化深度图。从那以后，许多研究者都提出了用于单目深度估计的深度学习方法。

基于深度学习的单目深度估计框架通常是一个编码器–解码器网络，输入 RGB 图像，输出深度图。编码器网络由卷积层和池化层组成，用于捕获深度特征；解码器网络则包含反卷积层，用于回归出与输入图像大小相同的像素级深度图。此外，为了保留各尺度的特征，编码器和解码器对应的层通过跳跃连接进行连接。整个网络通过深度损失函数进行约束和训练，当生成所需的深度图时，网络收敛。

在训练深度神经网络时，深度学习方法通常采用梯度下降法来获得局部最优解。这个最优解的质量取决于初始化和特定的参数设置。在初始化过程中，通常需要调整图像大小以满足网络学习的需求。此外，还需要设置初始学习率、优化器参数、批量大小和迷你批量大小，以学习和保存图像特征。常用的学习方法是随机梯度下降法；通常采用 Adam 优化器。

当梯度不再变化且损失函数趋于稳定时，网络收敛。

单目深度估计中的深度学习方法相较于传统方法，通过构建多层神经网络来学习深度特征，确实展现了更高的准确性。面对单目图像中的小遮挡或部分真实深度缺失的情况，深度学习方法依然能够估计出场景的深度，并且误差较小。而在场景中存在大遮挡或者没有真实深度数据的情况下，通过增加网络约束，深度学习方法也能够学习到场景的深度。这种强大的鲁棒性正是深度学习在单目深度估计中的显著优势。

在深度学习中，单目深度估计网络通过从真实深度图中学习场景结构信息来估计深度图。然而，获取真实深度图的成本非常高。因此，一些单目深度估计网络需要在较少或没有真实深度图的情况下进行训练，这些方法被称为半监督或无监督学习方法。接下来，将从训练方式的角度来介绍深度学习中的单目深度估计方法：监督学习、无监督学习和半监督学习模型。

2.1　监督学习方法

基于监督学习的单目深度估计网络使用真实深度图作为训练数据。监督学习方法依赖于大量的标注数据，因此其性能通常较高，但标注数据的获取成本高、耗时长。这种方法能够为其他方法(如无监督或半监督学习方法)提供一个基准或起点。

2.1.1　CNN 在单目深度估计中的应用

在深度学习中，研究者们设计了基于 CNN 的单目深度估计网络，无需进行建模，直接从单目标图像中获得深度信息。下面介绍从单目图像中学习到的绝对深度和相对深度。

对于绝对深度学习，Li J 等人[3]提出的 VGG-16 双流框架结合深度回归和深度梯度，通过深度梯度融合模块获得了一致的深度图。这种方法不仅提高了模型的泛化能力，而且还丰富了三维投影的信息。此外，还有许多基于更复杂 CNN 的单目深度估计方法用来学习像素级深度，比如基于 VGG 的模型、基于 ResNet 的模型和基于 DenseNet 的模型。

而在相对深度学习中，研究者们则利用图像中点对之间的相对关系来推断深度信息。根据输出点对之间的相对关系，利用数值优化方法获得密集的深度图。Chen W[4]和 Lee J[5]均设计了一种 CNN 来估计单目图像在不同尺度下的相对深度。结果表明，该方法经优化重建后的深度图像均方根误差可与绝对深度估计方法相当，同时其鲁棒性更强，不受数据同态性的影响。

总的来说，这两种方法各有优缺点，适用于不同的应用场景。如果需要更高的深度估计精度，那么绝对深度学习方法可能更适合；如果希望模型在复杂环境中也能保持较好的性能，那么相对深度学习方法可能是一个更好的选择。

2.1.2　RNN 在单目深度估计中的应用

基于 RNN 的网络编码器通常全部由 LSTM 层或 ConvLSTM 层组成，或者由卷积层和 LSTM 层混合组成，以提取和保留用于单目深度估计的空间-时序特征。例如，Kumar A C 等人[6]提出的深度网络使用了 8 个 ConvLSTM 层来预测单目深度图，可使网络充分利用序列中的时序信息，而卷积操作则有助于保持单元格之间的空间几何关系。另一方面，Mancini M 等人[7]则采用 LSTM 单元来按顺序利用输入流并预测场景深度。在他们的方法中，LSTM 层位于编码器网络中的卷积层之后。这些方法展示了 RNN 及其变体在处理具有时序信息的单目图像序列时的有效性，特别是在需要捕捉动态场景中的深度变化时。

2.1.3　GAN 在单目深度估计中的应用

GAN 通过生成器和判别器的对抗训练，能够生成与真实数据非常接近的深度图。在 Jung H 等人[8]的工作中，他们结合了 GlobalNet 和 RefinementNet，分别用于提取全局特征和估计局部结构。这样的设计使得模型能够更

全面地捕捉场景中的信息。

通过以上分析可知，CNN、RNN 和 GAN 各有优势。CNN 擅长学习场景的空间信息，RNN 擅长学习视频序列中的时间信息，而 GAN 则擅长生成和辨别深度图。这些方法在有足够真实深度图作为监督的情况下，能够实现很高的准确率。但是，真实深度图的获取确实是一个挑战。

2.2 无监督学习方法

监督学习方法需要大量的带有真实深度图的训练图像，这些图像的获取通常需要昂贵的设备和大量的人工标注工作，而无监督单目深度估计方法则通常利用图像之间的几何关系(如视差、光流等)来估计深度信息。这种方法不需要真实深度图作为监督，因此大大降低了对数据集的要求。例如，对于立体图像对，可以通过计算左右视图之间的视差来估计深度；对于单目图像序列，则可通过相邻帧之间的光流信息来推断场景的深度结构。虽然无监督学习方法的准确率可能略低于监督学习方法，但它们在实际应用中具有更大的灵活性和潜力。比如，在自动驾驶、机器人导航等领域，我们可以利用车载摄像头或机器人携带的相机所拍摄的视频序列进行无监督训练，从而实现对场景深度的实时估计。随着技术的不断进步和算法的不断优化，无监督学习方法在单目深度估计领域的应用前景将会越来越广阔。

2.2.1 无监督学习的立体匹配应用

立体匹配方法通常利用左右两张图像来计算深度值，而无监督学习方法则是受此启发，通过训练立体图像对来预测单张图像的深度信息。该方法不需要大量的标注数据，这在现实中是非常有用的。

Garg R 等人[9]最早在 2016 年提出了无监督单目深度估计模型。随后，越来越多的研究者开始利用左右视图来训练网络。这些网络基于二维或三维 CNN，进一步推动了无监督立体匹配方法的发展。

对于二维 CNN，Godard C 等人[10]提出左右一致性约束，通过同时重建左右视图来训练无监督网络。这大大提高了预测深度图的准确性。他们的工作为后来的研究者提供了很多有价值的思路和方法。除此之外，还有很多其他的研究者在这个领域作出了贡献。Goldman M 等人[11]构建一个孪生网络来学习立体图像，Andraghetti L 等人[12]通过传统的视觉里程计增强了深度估计，Watson J 等人[13]通过深度提示增强了立体匹配，Rehman S U 等人[14]应用了无监督预训练滤波器方法。这些研究都不断地推动着无监督立体匹配方法的进步和发展。

对于三维 CNN，一些研究者在三维卷积块中采用上下文信息来约束无监督网络，以进行单目深度估计。Chang J R 等人[15]提出了 PSMNet，该网络采用自上而下/自下而上的方式进行训练，以执行无监督的单目深度估计。其中，使用空间金字塔池化模块作为匹配成本体积；通过聚合半全局环境信息，并使用三维卷积模块结合多个堆叠的沙漏型三维 CNN 与中间监督来调整匹配成本体积。基于立体匹配的无监督学习模型主要受左右成对图像之间的投影和映射关系的约束，这仍然需要包含立体图像的数据集。因此，如何在训练阶段仅使用单个摄像头进行无监督的单目深度估计已经引起了研究人员的关注。

2.2.2 无监督学习在单目序列深度估计中的应用

无监督学习模型在训练时会同时考虑场景结构和相机运动。相机姿态估计在这里类似于图像变换估计，并对单目深度估计产生了积极影响。最近，研究人员将视觉里程计引入到基于单目序列的深度估计中，通过预测摄像机运动来学习场景深度。

基于单目序列的无监督学习深度估计的一般模型，包括深度网络和姿态网络两个子网络。这两个网络在训练阶段会联合训练，整个模型受到与立体匹配方法类似的图像重建损失的约束。但不同的是，这里的图像扭曲是建立

在单目序列的相邻帧之上的。除了重建损失函数之外，还采用了立体匹配方法中的平滑度损失函数和光度一致性损失函数。这些损失函数的结合使得模型能够更准确地估计深度和相机姿态[16]。

Zhou T 等人[17]设计了两个网络来分别估计单目视频中的深度图和相机运动，而且这些网络可以联合训练或单独训练。这一工作为后续的研究提供了很多有用的参考，比如使用三维几何约束训练的模型、具有不确定性或置信度图的估计、设计具有自注意力的网络等。

与监督学习方法相比，无监督学习方法不需要真实的深度图作为标签。这降低了构建深度标签的成本，但代价是可能会面临较低的准确性。简而言之，无监督单目深度估计通过图像间的几何关系来学习，无需真实深度数据，但精度可能不如监督学习方法。

2.3　半监督学习方法

半监督学习方法是指利用大量未标记数据作为半监督学习的辅助手段，以减少模型对真实深度图像的依赖。该方法既增强了深度图的尺度一致性，又提高了其估计准确性。

2.3.1　结合合成数据的单目深度估计方法

合成数据由图形引擎生成，为收集大量深度数据提供了一种可能的解决方案。因此，研究者们引入带有深度标签的合成数据集来进行单目深度估计。然而，在训练过程中，如何克服合成数据与现实数据之间的领域差距是一个挑战。

随着图像风格迁移与领域自适应之间的联系变得日益紧密，研究者们采用风格迁移和对抗训练来估计现实场景中的深度图。该方法依赖于大量合成数据训练的模型。Zheng C 等人[18]提出了一种双模块领域自适应网络。其中一个模块训练合成和真实图像，并通过重建损失和生成对抗损失相互重建，然后再将其输入到另一个模块以预测真实深度图。此外，还有其他一些模型采用自注意力、循环一致性、跨领域等方法来进行领域自适应，以预测单目深度图。

2.3.2　结合雷达辅助的单目深度估计技术

研究者们采用雷达这样的辅助深度传感器来捕捉真实世界中的深度信息。不过，雷达采集的数据可能会有一些噪声，而且通常会比真实的深度图稀疏。所以，在利用这些数据时，需要特别小心。

Kuznietsov Y 等人[19]提出的半监督学习网络就是一个很好的例子。他们输入左右图像，并构建一个立体对齐的几何约束形式，实现了对稀疏数据的有效利用。其中的深度一致性损失主要是左右深度估计图与稀疏数据之间的误差。试验结果表明，这种方法确实比传统的监督和无监督方法表现更好。

2.3.3　结合表面法线信息的单目深度估计方法

结合表面法线信息的单目深度估计方法利用表面法线与深度之间的强相关性，通过从输入的 RGB 图像中提取额外信息（如表面法线）来增强深度估计的准确性。

表面法线作为三维空间中某点局部切平面的法向量，与深度信息紧密相连。一方面，表面法线可以从深度信息中估计出来；另一方面，深度信息也受到由表面法线确定的局部切平面的约束。这种相互依赖的关系使得结合表面法线进行深度估计成为了一个有效的策略。

在 Qi X 等人[20]提出的 GeoNet 模型中，就体现了这种策略的应用。GeoNet 由两个主要部分组成：一个是从深度到表面法线的网络，利用最小二乘法从深度信息中求解表面法线；另一个是从表面法线到深度的网络，通过核回归模块对初始深度图进行细化。这种方法利用表面法线在局部平面上不发生较大变化的特性来提高单目深度估计的准确度。此外，还有其他一些模型也采用了类似的策略：通过引入深度-法线一致性、表面正则化约束或深度补全等方法，进一步提升结合表面法线进行单目深度估计的效果。

总的来说，结合辅助信息（如虚拟数据、

稀疏深度和表面法线)的半监督学习方法，在单目深度估计中表现出了更高的准确性。虽然辅助信息需要增加输入的数据量和内容复杂性，但是可以显著提高深度估计的精度和可靠性。

2.4 总结

本节主要从训练方式的角度对单目深度估计中的深度学习方法进行了分类，包括监督学习、无监督学习和半监督学习方法。

对于单目深度估计，监督学习方法具有最高的准确性，但它强烈依赖于真实的深度图。这意味着为了训练模型，需要大量的带有真实深度信息的图像数据，而这在现实应用中很难实现。

无监督学习方法通过在输入图像上构建几何约束来预测深度图，而不需要真实的深度图作为监督。这种方法虽然避免了数据标注的难题，但其准确性通常略低于监督学习和半监督学习方法。

半监督学习方法综合了监督学习和无监督学习的优点，它依赖于一些辅助信息，这些信息比真实的深度图更容易获取。这种方法在保持一定准确性的同时，也减少了对大量标注数据的依赖。

3 展望

近几年来，基于深度学习的单目深度估计技术得到了广泛的研究和发展，然而仍有一些局限性需要克服：(1)为了提高准确性，研究人员加深了深度神经网络的层数，这增加了内存使用量和空间复杂度。(2)单目深度估计网络通常是编码-解码网络，深度特征经过多层信息处理后严重丢失，导致估计的深度图准确性较低，无法满足实际应用的要求。下面对单目深度估计未来面临的关键挑战以及研究方向进行了展望。

3.1 网络框架的整合与优化

在许多监督学习模型中，语义分割将与深度估计结合，但它仍然是一个处理独立任务的独立模块。在无监督学习方法中，通常存在多个子网络，分别能够学习深度估计、视觉里程计和流估计。然而，这些网络并没有很好地连接在一起，导致参数数量庞大，增加了内存需求和计算量。如何更好地整合网络是一个研究方向，值得未来探索。或许可以通过使用相同的深度学习网络同时获得不同的特征，如语义信息、光流特征和深度特征。在编码阶段，同时提取和匹配不同类型的特征；在解码阶段，它们被分别解码以满足应用需求。

3.2 轻量级网络设计

对于深度网络带来的高内存和计算复杂度问题，往往神经网络要求更多的时间进行计算，故设计更加轻量级的网络结构也是一个发展方向。例如，使用深度可分离卷积、瓶颈层等技术来减少参数数量和计算量。同时，可以探索网络剪枝、量化技术等来减少模型的复杂度，同时尽量保持其性能。另外，也可以考虑结合硬件加速技术，比如使用专用的人工智能芯片或图形处理器来加速计算过程[21]。

3.3 数据集构建

数据集的质量对深度学习模型的泛化能力和鲁棒性至关重要。为了提升深度估计的结果，需要更多质量更好、场景类型更丰富的数据。然而，构建新数据集不仅耗时还昂贵。利用计算机生成大量图像虽然可行，但质量参差不齐。未来可以探索如何结合真实数据和生成数据，以及如何通过自动化和半自动化的方法来加速数据集的构建过程。

3.4 动态物体和遮挡问题

现有的大多数深度估计模型都是在理想条件下设计的，难以应对复杂的动态和遮挡场景。为了提升模型在实际应用中的表现，我们需要开发能够处理动态物体和遮挡问题的新模型或算法。这需要结合物体检测、跟踪和场景理解等多方面的技术。

3.5 高分辨率深度图的输出

对于增强现实和虚拟现实等实际应用，高分辨率深度图的输出至关重要。然而，为了提高计算效率，现有的大多数深度估计模型预测

的深度图分辨率通常较低。虽然研究人员已经尝试使用彩色图像超分辨率模型来细化深度图的超分辨率，但如何直接输出高分辨率深度图仍然是一个需要研究的方向。未来可以探索如何在保证计算效率的同时，提升深度图的分辨率和准确性。

参考文献

[1] Hochreiter S, Schmidhuber J. Long short-term memory [J]. *Neural Computation*, 1997, **9**(8): 1735-1780.

[2] Eigen D, Puhrsch C, Fergus R. Depth map prediction from a single image using a multi-scale deep network [C]. Montreal: 28th Annual Conference on Neural Information Processing Systems, 2014.

[3] Li J, Klein R, Yao A. A two-streamed network for estimating fine-scaled depth maps from single RGB images [C]. Venice: 2017 IEEE International Conference on Computer Vision, 2017.

[4] Chen W, Fu Z, Yang D, et al. Single-image depth perception in the wild [C]. Las Vegas: 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[5] Lee J, Kim C S. Monocular depth estimation using relative depth maps [C]. Long Beach: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019.

[6] Kumar A C, Bhandarkar S M, Prasad M. DepthNet: A recurrent neural network architecture for monocular depth prediction [C]. Salt Lake City: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018.

[7] Mancini M, Costante G, Valigi P, et al. Toward domain independence for learning-based monocular depth estimation [J]. *IEEE Robotics and Automation Letters*, 2017, **2**(3): 1778-1785.

[8] Jung H, Kim Y, Min D, et al. Depth prediction from a single image with conditional adversarial networks [C]. Beijing: 2017 IEEE International Conference on Image Processing, 2017.

[9] Garg R, BG V K, Carneiro G, et al. Unsupervised CNN for single view depth estimation: geometry to the rescue [C]. Amsterdam: 14th European Conference on Computer Vision, 2016.

[10] Godard C, Aodha O M, Brostow G J. Unsupervised monocular depth estimation with left-right consistency [C]. Honolulu: 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[11] Goldman M, Hassner T, Avidan S, et al. Learn stereo, infer mono: Siamese networks for self-supervised, monocular, depth estimation [C]. Long Beach: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019.

[12] Andraghetti L, Myriokefalitakis P, Dovesi P L, et al. Enhancing self-supervised monocular depth estimation with traditional visual odometry [C]. Quebec City: 2019 International Conference on 3D Vision, 2019.

[13] Watson J, Firman M, Brostow G, et al. Self-supervised monocular depth hints [C]. Seoul: 2019 IEEE/CVF International Conference on Computer Vision, 2019.

[14] Rehman S U, Tu S, Waqas M, et al. Unsupervised pre-trained filter learning approach for efficient convolution neural network [J]. *Neurocomputing*, 2019, **365**: 171-190.

[15] Chang J R, Chen Y S. Pyramid stereo matching network [C]. Salt Lake City: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018.

[16] 刘斌. 基于无监督学习的单目红外图像深度估计 [D]. 哈尔滨：哈尔滨工程大学, 2022.

[17] Zhou T, Brown M, Snavely N, et al. Unsupervised learning of depth and ego-motion from video [C]. Honolulu: 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[18] Zheng C, Cham T J, Cai J. T2net: synthetic-to-realistic translation for solving single-image depth estimation tasks [C]. Munich: 15th European Conference on Computer Vision, 2018.

[19] Kuznietsov Y, Stuckler J, Leibe B. Semi-supervised deep learning for monocular depth map pre-

diction [C]. Honolulu: 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[20] Qi X, Liao R, Liu Z, et al. Geonet: Geometric neural network for joint depth and surface normal estimation [C]. Salt Lake City: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018.

[21] 江俊君, 李震宇, 刘贤明. 基于深度学习的单目深度估计方法综述 [J]. **计算机学报**, 2022, **45** (6): 1276-1307.