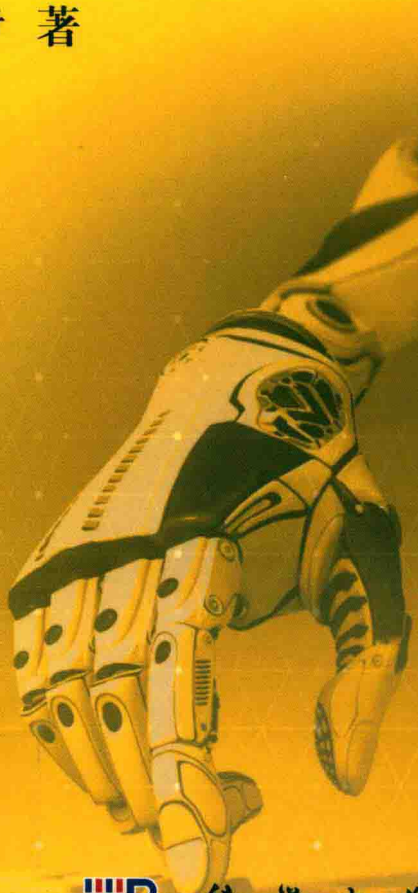
正则化深度学习及其在机器人环境感知中的应用

Regularied Deep Learning and Its Application in Robotics Perception

刘 勇 廖依伊 著



科学出版社

正则化深度学习及其在机器人环境感知中的应用

Regularized Deep Learning and Its Application in Robotics Perception

刘　勇　廖依伊　著

科学出版社

北京

内 容 简 介

近年来，随着人工智能技术的飞速发展，深度神经网络技术在图像分析、语音识别、自然语言理解等难点问题中都取得了十分显著的应用成果。本书系统地介绍了深度学习应用于机器人环境感知面临的难点与挑战，针对性地提出基于正则化深度学习的机器人环境感知方法，并结合机器人作业场景分类、多任务协同环境感知、机器人导航避障环境深度恢复、感知目标三维重建等应用案例对正则化深度学习方法应用进行介绍。本书紧紧围绕面向机器人环境感知的深度学习问题，深入分析相关概念，建立相关模型，并设计相关方法，为正则化深度学习机器人环境感知应用提出了较为系统的解决方案。

本书可供人工智能、机器人、计算机等专业的研究生、教师和科研人员参考。

图书在版编目(CIP)数据

正则化深度学习及其在机器人环境感知中的应用/刘勇，廖依伊著. —北京：科学出版社，2018.12

ISBN 978-7-03-059426-6

Ⅰ. ①正… Ⅱ. ①刘… ②廖… Ⅲ.①正则化-机器学习-应用-机器人-传感器-研究 Ⅳ. ①O177 ②TP242

中国版本图书馆 CIP 数据核字 (2018) 第 254119 号

责任编辑：胡庆家 郭学雯 / 责任校对：彭珍珍
责任印制：吴兆东 / 封面设计：陈 敬

科学出版社 出版
北京东黄城根北街 16 号
邮政编码：100717
http://www.sciencep.com

北京虎彩文化传播有限公司 印刷

科学出版社发行 各地新华书店经销
*
2018 年 12 月第 一 版 开本：720×1000 B5
2018 年 12 月第一次印刷 印张：9 插页：2
字数：170 000

定价：68.00 元

(如有印装质量问题，我社负责调换)

前　言

近年来，随着人工智能技术的飞速发展，深度学习技术在图像分析、语音识别、自然语言理解等难点问题中都取得了十分显著的应用成果。然而该技术在机器人感知领域的应用相对而言仍然不够成熟，主要源于深度学习往往需要大量的训练样本来避免过拟合，提升泛化能力，从而降低其在测试样本上的泛化误差，而机器人环境感知中涉及的任务与环境具有多样化特性，且严重依赖于机器人硬件平台，因而难以针对机器人各感知任务提供大量标注样本；其次，对于解不唯一的病态问题，即使提供大量的训练数据，深度学习方法也难以在测试数据上提供理想的估计，而机器人感知任务中所涉及的距离估计、模型重构等问题就是典型的病态问题，其输入中没有包含对应到唯一输出的足够信息。针对上述问题，本书以提升深度学习泛化能力为目标、以嵌入先验知识的正则化方法为手段、以机器人环境感知为应用背景展开研究，具体取得了以下四个方面的研究成果。

(1) 提出约束隐层特征表示的图正则自编码器，以流形假设为先验知识约束隐层特征保留输入空间中的局部近邻特性，并通过理论分析论证了图正则项有助于学习对于输入的小量干扰具有鲁棒性的特征表示，从而提升自编码器网络的泛化能力。在此之上，本书将图正则自编码器应用于 2D 激光观测的场景分类问题，利用广义图正则项约束样本采集位置相邻的 2D 激光观测学习相似特征表示，说明图正则项可用于嵌入移动机器人空间位置等特定任务下的先验知识。

(2) 提出约束深度神经网络结构的语义正则网络，以机器人感知多任务之间的相关性为先验知识构造单输入多输出的正则化网络结构，其中像素级的语义分割任务作为图像级的场景分类任务的正则分支，约束网络在理解物体语义信息的基础上理解场景类别，从而在大幅减少所需训练样本数目的同时提升网络在图像场景分类任务上的泛化能力。

(3) 提出约束深度神经网络结构的嵌套残差网络，针对单目图像深度估计的病态特性，引入移动机器人感知中常见的稀疏深度观测并从中生成稠密参考深度，再利用稠密参考深度与真实深度的差值具有物理意义的先验知识构造正则化的嵌套残差网络结构，约束网络直接估计残差深度，从而在仅引入十分稀疏的深度观测 (如 2D 激光点云) 时即可显著降低单目图像估计深度的不确定性。

(4) 提出约束网络输出的深度移动立方体网络，针对从部分观测重构物体三维模型问题的病态特性，提出端到端的估计可表示任意拓扑结构的三维网格模型，使得直接对重构的三维网格模型进行正则成为可能，再以三维模型几何特性为先验

知识直接约束三维网格模型的平滑性以及复杂度，使得网络可直接从不完整且有噪声的观测给出一个理想的三维网格模型估计，对于机器人抓取操作的感知等实际应用具有重要意义。

对于上述提出的关键技术，本书在多种机器人环境感知任务上设计了定量与定性实验，在多个数据集上检验了算法的性能，充分验证了在正则化的统一框架下提升深度学习泛化能力的有效性。

作　者

2018 年 5 月于浙江大学

目　　录

第1章 绪 论

1.1 背景和意义

近年来，人工智能以飞速的发展引起了全世界的高度重视。在图像处理领域知名的 ImageNet 大规模视觉识别挑战赛上，微软研究院在 2015 年提出了误分类率低至 4.95%的分类算法[1]，首次在该挑战赛上成功超越人类 5.1%的误分类率。2016 年，AlphaGo 以 4:1 击败韩国围棋职业棋手李世石[2]，引起了大众对于人工智能的关注热潮。2018 年 1 月，亚马逊的无人零售商店 Amazon Go 开始向公众开放，使用人工智能机器人代替了传统的超市收银员。全球众多知名企业如谷歌、百度等纷纷投入大量物力和财力研究辅助驾驶以及无人驾驶，试图抢占无人驾驶领域的先机。为抓住人工智能发展的机遇，推动建设创新型国家和世界科技强国，我国也相继出台了一系列政策支持人工智能的发展。2017 年 7 月，国务院印发《新一代人工智能发展规划》，对我国新一代人工智能发展的总体思路、战略目标和主要任务、保障措施进行了系统的规划和部署。

深度学习 (deep learning) 是推动当前人工智能热潮的一个关键因素，在 AlphaGo、ImageNet 挑战赛、Amazon Go 以及无人驾驶中均发挥着重要作用，它被广泛应用于图像处理、语音处理、自然语言处理以及决策问题等[3−6]，深刻地影响着各行各业以及人们的日常生活。具体而言，深度学习代表了一类机器学习算法，其特点是利用多层级联的非线性处理单元学习输入样本的特征表示，并在这些特征表示上构造可导的代价函数，然后通过反向梯度传播算法最小化代价函数并同时实现特征表示的学习。相比传统机器学习方法，深度学习通过这种端到端 (end-to-end) 的训练方式直接从原始输入数据学习特征表示并进行分类回归等决策，而传统机器学习方法首先根据人类自身的经验知识设计特征提取方法，例如，图像处理领域知名的手工特征提取方法 SIFT 和 SURF[7, 8]，再独立设计分类或回归等决策器。深度学习通过特征表示的学习获得了两点优势：第一，传统机器学习方法中可训练的部分仅包含分类器决策部分，而深度学习模型从特征提取到分类决策都是可训练的，增强了模型对于问题的拟合能力；第二，深度学习可通过多层级联的

非线性处理单元由简单到复杂、由细节到局部地提取特征，降低了直接设计复杂特征的难度。

尽管深度学习在诸多问题上体现了优异的性能，但它仍然存在一定的局限。首先，有监督的深度学习往往需要大量输入数据及其对应标注来实现理想的泛化 (generalization) 能力，所谓泛化能力是指算法对未曾见过的样本给出理想估计的能力。定义用于模型训练的样本为训练数据，以及一组与训练数据无交集的样本为测试数据，则带标注的训练数据的规模大小是深度学习能否在测试数据上实现优秀性能的决定性因素之一。举例来说，Zhou 等[9] 通过采用两百多万张有标注的场景图像训练深度学习模型，在一系列场景分类数据集上获得了大幅度领先的分类结果，Johnson-Roberson 等[10] 则通过采集大量仿真图像作为训练数据，提升了深度学习在车辆检测上的性能。然而，获取大量标注数据的代价十分高昂，通过大量标注训练样本提升深度学习泛化能力的方法仅适用于应用对象非常广泛的任务。在训练样本不足的情况下，由于深度学习模型包含大量可学习参数，很容易产生过拟合 (over-fitting) 现象，即模型在训练样本上表现优异，而无法正确理解训练时未曾见过的测试样本。此外，对于病态的、解不唯一的问题 (如图像去噪[11, 12]、图像超分辨率重构[13, 14]、物体三维模型重构[15–17])，即使提供了充足的训练样本，在测试数据上作出理想估计也是对深度学习的一大挑战。

相比之下，人类可以通过少量的样本迅速获取知识，并将知识应用于新的环境，其中一个关键的原因是人类具有长期知识的积累，因此在理解和认识事物时已具备一定的先验知识。类似地，本书的研究思路是通过正则化 (regularization) 将先验知识引入到深度学习模型中，狭义的正则化一般理解为约束算法中参数的数量或模值，例如，最小化参数的绝对值之和 (1 范数) 或均方和 (2 范数)。而本书所考虑的是更为广义的正则化，定义为 “任意试图减小模型的泛化误差而非降低训练误差的方法”[4]。综上所述，本书拟从正则化约束出发，将多方面先验知识引入深度学习模型中，加强深度学习算法在面对小样本、欠定解等问题下的鲁棒性和泛化性，并在机器人环境感知的一系列任务上验证算法的有效性。

1.2　问题与挑战

1.2.1　深度学习问题描述

为方便后面解释，本节首先通过符号和公式定义本书主要关注的监督学习问题以及深度学习模型。给定输入 x 以及目标输出 y，深度学习要解决的问题是拟

合一个 $y = f(x)$ 的函数。举例来说，图像分类问题中 x 对应输入图像，y 对应图像类别，$f(\cdot)$ 则是从输入图像到其类别的映射函数。$f(\cdot)$ 通常是一个十分复杂的映射，例如，对于相同的类别 y，其输入图像 x 会受到个例、光照以及环境等多种变量的影响。为解决这一问题，深度学习的思路是构造一组简单的映射 $f_1, f_2, \cdots, f_L$ 来拟合一个 $f(\cdot)$ 的复杂映射。令 f^* 为 $f_1, f_2, \cdots, f_L$ 组合在一起的函数，θ 为其中所有参数，则深度学习的目标是学习一组参数 θ 使得 $y^* = f^*(x;\theta)$ 尽可能接近 y。

深度学习中最常用的人工神经网络结构是深度前馈网络 (deep feedforward networks)，称之为 "**网络**" 是由于它是多个函数 $f_1, f_2, \cdots, f_L$ 的组合，称之为 "**前馈**" 是因为它通过级联的方式组合这些函数，例如，$L = 3$ 时网络的输出为 $y^* = f_3(f_2(f_1(x)))$，f_i 的输出只影响 $f_{i+1}, f_{i+2}, \cdots$ 而不影响自身，因此不构成反馈连接。由于这种级联的结构，f_i 称为网络的一个层，且 x 称为输入层 (input layer)，f_1, $f_2, \cdots, f_{L-1}$ 为隐层 (hidden layer)，y^* 为输出层 (output layer)，L 则为网络的层数。"**深度**" 一词则是由 L 的数值而来，现有的深度学习算法已经可以有效地训练 $L > 100$ 的网络[18]。需要说明的是，万能近似定理 (universal approximation theorem) 指出，即使 $L = 2$ 时 (隐层数量等于 1) 网络也可以无限逼近任意函数[19]，然而现有的理论研究表明通过增加网络的层数可获得指数倍增长的表达能力，从而大幅度降低所需要的参数数量[20–22]。

给定深度学习待解决的问题及其网络结构，参数 θ 的学习是通过最小化网络在所有训练样本上的估计误差而实现的。定义多组输入 x 及其目标输出 y 的集合为训练数据，标记为 $\mathscr{D} = \{(x_1, y_1), \cdots, (x_N, y_N)\}$，则深度学习的代价函数可以定义如下：

$$\mathcal{L} = \sum_i E(f^*(x_i;\theta), y_i) \tag{1.1}$$

其中 $E(y^*, y)$ 是衡量 y^* 与 y 之间的距离，例如，回归问题中 $E(y^*, y)$ 可以是 ℓ_1 或 ℓ_2 距离，分类问题中 $E(y^*, y)$ 可以是两个概率分布之间的互信息熵。深度学习的训练常采用梯度下降法 (gradient descent) 来最小化代价函数 $\mathcal{L}$，即根据梯度 $\delta\mathcal{L}/\delta\theta$ 来调整 θ。

1.2.2 深度学习的挑战

1.2.1 节提到给定训练数据集 $\mathscr{D}$，深度学习需要更新网络参数 θ 来拟合一个复杂映射。一般来说，深度学习模型中 θ 包含的未知变量个数庞大，$\mathscr{D}$ 中样本量的规

模远小于 θ 中未知变量的个数且存在噪声，因此可以使式 (1.1) 中代价函数最小的参数 θ 取值不唯一。由于网络参数自由度高，网络有可能产生过拟合现象，即深度学习模型完全拟合了训练数据对的分布特性，却不一定能在训练数据之外的测试数据上给出合理估计，也就是说网络的泛化能力差。事实上，从有限训练集 $\mathscr{D}$ 估计参数 θ 的问题是病态 (ill-posed) 的[23]，而导致这个病态问题的根本原因是训练数据集 $\mathscr{D}$ 缺少完整覆盖真实样本空间分布的信息，这也解释了为什么扩充 $\mathscr{D}$ 中样本的数量有助于避免过拟合，提高深度学习算法的性能。

除了参数学习是一个病态问题之外，深度学习中还可能涉及另外一类病态问题。对于一些特定任务而言，任务本身就是一个病态映射。举例来说，对于图像去噪、图像升采样、从单目图像估计深度、三维重构等一系列问题，输入 x 本身就不包含足够确定唯一输出 y 的信息，因此 y 的取值是有二义性 (ambiguity) 的。尽管在充分的训练下，网络可以记忆训练数据中 x 与 y 的对应关系，但是对于测试数据来说，在 x 信息不足的情况下重构出理想的 y 更具挑战性。

为了区分这两类病态问题，本书分别称之为参数学习病态性以及任务映射病态性，值得一提的是，无论任务映射病态性是否存在，参数学习都具有病态性。这两种病态问题也反映了深度学习中存在着的两点挑战。

- 为了克服参数学习病态性，深度学习需要提供大量的标注样本，尽可能地覆盖真实样本空间的分布，然而大量标注样本的获取代价高昂；
- 即使具备大规模训练样本，拟合具有二义性的病态映射，且在测试数据上作出准确估计也是对深度学习的挑战。

由上面的分析可见，造成病态问题的一个关键原因是信息的缺失，即参数学习病态问题中训练数据集的信息不足以覆盖真实样本空间分布，而任务映射病态问题中输入的信息不足以确定准确输出。在信息缺失的前提下，引入有效的先验信息是应对这些挑战的一个重要思路，正如数学家 Lanczos 所言[24]：

> A lack of information cannot be remedied by any mathematical trickery.

因此，本书的思路就是在深度学习模型中引入先验信息来提升算法的泛化能力，而先验知识则通过各类正则化方法体现在深度学习模型中。

1.2.3　机器人环境感知

在深度学习的实际应用中，机器人的环境感知具有重要的现实意义，同时其应用场合差异大、标注样本少、包含病态映射等特性又对深度学习的泛化能力提出了

更高的要求。因此，本书以机器人环境感知的一系列任务为应用场景验证正则化深度学习的有效性。机器人环境感知的特性可以总结如下。

• 首先，机器人环境感知对于机器人的自主性和智能性而言具有重要意义。对于人类来说，环境感知也是日常生活中无时无刻不在进行的，在认知环境的基础上人类才能执行许多高等任务。机器人也需要从传感器的信息中提炼出语义信息、结构信息，才能进一步执行具体任务。

• 其次，机器人应用中往往需要面对多样化环境与多类感知问题，而针对不同的问题，在不同场景中采集大量数据并进行标注的做法，代价十分高昂且难以实现，尤其是对于一些特殊的应用场合如救援、恶劣天气等，即使收集无标注数据也并不容易，样本数量的不足凸显了深度学习的参数学习病态问题。

• 再次，机器人应用中不仅需要对语义信息进行定性感知，也包括对距离、结构等几何信息的定量感知，而在许多定量感知问题中输入信息不足以确定唯一的输出，对应了前文提到的任务映射病态问题。

• 最后，机器人具有运动连续性、多传感器、需同时处理多个任务等特性，这些特性为机器人环境感知提供了丰富的先验知识。

由此可见，机器人环境感知反映了深度学习中的两类病态问题，本书希望在探索通用先验知识的基础上，利用机器人领域特有的先验知识构造正则化深度学习方法，从而提升深度学习在机器人环境感知问题上的泛化能力。由于机器人环境感知所涉及的范围十分广泛，本书主要探讨其中两类具有代表性的问题。

• 定性语义感知：理解语义知识有助于机器人与人类的交互，使得机器人可以在语义层面上理解待执行的任务，在这类问题中本书所关注的具体内容是对于场景类别以及物体类别的理解。

• 定量距离结构感知：在理解场景语义的基础上，机器人执行任务时还需要知道目标的距离，对于抓取等操作还需要知道物体的三维模型，在这类问题中本书所关注的是针对场景的距离感知以及针对物体的三维模型重构。

1.3 研究现状

本节针对与深度学习相关的国内外研究现状进行综述和小结。从探讨深度学习的起源、发展及其体系结构出发，再介绍深度学习中常用的正则化方法，最后说

明目前深度学习在机器人环境感知中的研究现状。

1.3.1　深度学习发展

深度学习的起源可以追溯到 1958 年感知器的提出[25]，该文受生物神经细胞启发实现了一种最简单的单层前馈神经网络 $y^* = H(wx+b)$，其中 $H(\cdot)$ 为单位阶跃函数。虽然最初感知器被认为是一种很有潜力的模型，1969 年 Minsky 和 Papert[26] 在 *Perceptrons* 书中指出感知器无法解决以异或 (XOR) 为代表的线性不可分问题。尽管，1970 年 Linnainmaa[27] 在其硕士论文中指出可以通过多层感知器 (multilayer perceptrons) 解决非线性问题，然而主流学者的批判态度仍然大幅度降低了人们对感知器的研究热情。

1986 年，Rumelhart 等[28] 在 *Nature* 发文，说明了多层感知器可解决复杂非线性问题，重新引发了人们对于人工神经网络的关注。多层感知器是对单层感知器的推广，通过多个非线性层的组合克服了单层感知器无法处理线性不可分问题的局限。具体来说，多层感知器的一层可以表示为 $f_i(x) = s(W_i x + b_i)$，其中 $s(\cdot)$ 可以是任意的非线性函数，一般常用的是 Sigmoid 函数[29]。Rumelhart 等的主要贡献是介绍了用于训练多层传感器的反向传播 (back-propagation) 算法，其核心思想是通过沿代价函数 $\mathcal{L}$ 下降的方向来调整权值，即根据梯度 $\delta\mathcal{L}/\delta W$ 来调整 W，其中 $\mathcal{L}$ 是网络输出 y^* 与目标输出 y 的距离。现在反向传播算法仍是训练人工神经网络的基础算法。然而，由于网络的隐层个数大于 2 时反向传播算法存在梯度消失的问题，层数太深的网络难以训练，加深网络层数反而难以获得更优的性能。再加上当时计算资源的局限和训练数据的不足，学界在 20 世纪 90 年代后开始将关注重点更多地放在其他机器学习算法上，人工神经网络逐渐淡出人们的视线。

2006 年，Hinton 和 Salakhutdinov[30] 在 *Science* 上提出了贪婪逐层预训练方法，解决了多层网络难以训练的问题，重新引发了人们对于神经网络的关注，也普遍被认为是现代 “深度学习” 的开端。具体而言，该文提出了一种由多个受限玻尔兹曼机 (restricted Boltzmann machine，RBM) 组成的深度置信网络 (deep belief networks，DBN)。RBM 是一种随机概率模型，可以从输入 x 学习特征 h 并从 h 重构 x。因此文中首先无监督地从低往高训练每一层 RBM 的参数，达到初始化权值的效果，低层的 RBM 训练完成后则固定其参数，将其输出作为后一层 RBM 的输入，这种训练方式称为无监督预训练 (unsupervised pre-training)。最后可以将所有

RBM 连接起来同时训练并进行微调 (fine-tuning)。对于分类问题，可以在最末一层的特征上接入分类器进行微调。DBN 提出后，基于同样的无监督预训练思路的栈式自编码器 (stacked auto-encoders，AE) 也在许多问题上取得了不错的性能[31−33]。栈式自编码器的思路与 DBN 类似，两者最大的区别在于 DBN 是随机概率生成模型，而栈式自编码器属于确定性模型。

多层感知器、深度置信网络以及栈式自编码器都属于全连接 (fully connected) 结构。对于两个相邻的表示层 $\boldsymbol{h}_i$ 和 $\boldsymbol{h}_j$ 来说，$\boldsymbol{h}_i$ 中的任意一个神经元 $h_{ik}, k = 1, \cdots, K$ 都和 $\boldsymbol{h}_j$ 中的所有神经元 $h_{jl}, l = 1, \cdots, L$ 相连，因此两层之间需要权值矩阵 $\boldsymbol{W} \in \mathbb{R}^{K \times L}$，在数据维度高时权值矩阵中包含的参数数目庞大。1989 年，由 LeCun 等[34, 35] 提出的卷积神经网络 (convolutional neural networks，CNN) 则通过局部连接以及权值共享的方式，大幅度降低了所需参数的数量，最终它被成功地应用到了图像数字识别的任务中。2012 年，Hinton 及其同事 Krizhevsky，Sutskevery[36] 在 ImageNet 大规模视觉分类挑战赛上首次通过 CNN 获得冠军，由此之后 CNN 在图像处理等多个领域备受青睐[37−41]。

从网络中信息的流向来说，前面介绍的网络结构都属于深度前馈神经网络 (deep feedforward networks)[4] 即网络中信息的流向是单向且不存在循环结构的。除此之外，递归神经网络 (recurrent neural networks)[42−44] 是另一类包含反馈结构的网络，往往用于序列化数据的特征提取与决策，如自然语言处理[42] 和视频处理[44] 等。由于本书研究内容并非局限于序列化数据，因此本书的研究针对前馈神经网络进行展开。

1.3.2 深度学习与正则化

在 1.2.2 节中提到，引入先验知识是解决病态问题的一个关键，而先验知识通常以正则化的方式嵌入到模型中。深度学习中最常用的三种正则化方式分别为：① 关于数据 $\mathscr{D} = \{(x_1, y_1), \cdots, (x_N, y_N)\}$ 的正则化；②关于结构 $f^*(\cdot)$ 的正则化；③加入正则项 $R(\cdots)$ 的正则化，此时网络代价函数由式 (1.1) 扩展如下：

$$\mathcal{L} = \sum_i E(f^*(x_i; \theta), y_i) + \alpha R(\cdot) \tag{1.2}$$

下面将分别对这三类最常用的正则化方法进行分析。

1. 关于数据的正则化

关于数据的正则化方法中主流的一类方法是数据增强，一般指通过添加噪声和干扰的方式扩充训练数据集 $\mathscr{D}$[45-48]。理想情况下，如果给定一个覆盖了所有样本空间可能性的训练数据集，那么模型在测试数据上也可以达到优秀性能，所以增大训练数据集也是一个提升模型泛化能力的有效方式。对于某些特定的问题来说，数据增强是很容易实现的。例如，图像识别问题中，输入图像 x 的旋转、平移、翻转、尺度缩放等多种变化都不会改变其标签 y，因此每个标注样本对 (x,y) 可扩充为多组样本对 $(x',y),(x'',y),\cdots$。此外，直接在输入数据 x 上加入少量噪声也是一种关于数据的正则化方式[49, 50]。

深度学习中通用的技巧之一 Dropout[51]，也可以被认为是一种关于数据的正则化方法。Dropout 指在网络的节点上加入 0/1 的随机掩码，所以网络训练时被激活的节点是随机的，这可以看成是在数据上乘上对应的随机二值掩码。

此外，批标准化 (batch normalization, BN)[52] 也可以看成是一种数据正则化方法，它对每个批 (batch) 的特征层分布进行标准化操作，也就是说每个训练样本不再作为一个独立样本，Ioffe 和 Szegedy[52] 认为这有助于提高网络的泛化能力。

2. 关于结构的正则化

深度神经网络结构 $f^*(\cdot)$ 的设计本身也可以看成是一种正则化方法，这意味着将关于待拟合问题的先验知识表示在了结构设计中。举例来说，如果 $f^*(\cdot)$ 由一个层数较浅、参数较少的网络来构成，则说明先验知识表明待拟合问题本身的复杂度较低，反之，如果 $f^*(\cdot)$ 采用一个层数较深、参数较多的网络结构，则说明先验知识认为待拟合问题本身是复杂的，且这个复杂问题可以通过逐层抽象的方式来拟合。

卷积神经网络[34, 53, 54] 是一种关于结构的正则化方法。相比全连接网络，卷积神经网络中卷积层 (convolutional layer) 所考虑的先验知识的特征往往是局部的，因此一个相同的卷积核可以应用于全图的特征提取，而池化层 (pooling layer) 所假设的先验知识是输入样本的响应在局部范围内是不变的。

除了规定网络中每一层的基本操作之外，还有一类关于结构的正则化方法是约束整个网络中的连接方式。相比于深度前馈网络中的基本连接方式，即 1.2.1 节中介绍的逐层级联的链式结构，跳跃连接 (skip connection) 属于一种正则化连接方法，例如，将网络中第 i 层的特征与第 j 层的特征进行并联并作为第 $j+1$ 层的输

入[46, 55]，其中 $j > i+1$。采用跳跃连接的思路而建立的残差网络 (ResNet)[18,56] 也是近年来被广泛应用的一种网络结构，它通过跳跃连接构造对于残差的学习，而非关于任务本身的学习，因此它的先验假设是残差的学习相比于直接学习任务本身难度更低，一系列基于 ResNet 的优秀成果也验证了这一假设的正确性[57, 58]。

多任务学习也可以看成是一种关于结构的正则化方法[59−61]，即网络由一个单输入单输出的结构改变成单输入多输出，甚至多输入多输出的结构。此时网络除拟合 $y = f(x)$ 还需要拟合多组 $y^{(i)} = f^{(i)}(x^{(i)})$，其中多个任务的输入可以是相同的。在这类网络结构中，一般部分参数属于多任务共享，而另一部分参数单独用于其中某个任务。在基于深度学习的多任务学习上，Wang 等[59] 提出同时从单目图像估计每个像素的语义信息以及深度信息；Cadena 等[60] 在此基础上进一步考虑了对原图像的重构，即同时考虑图像重构，深度估计以及语义分割；Cheng 等[61] 以松耦合形式同时估计语义分割以及光流。将无标注数据引入学习过程的半监督学习 (semi-supervised learning) 也属于多任务学习的结构正则[62, 63]，此时网络不仅需要拟合训练样本输入与输出之间的联合概率密度 $P(x, y)$，还需要拟合所有有标注数据与无标注数据的概率密度分布 $P(x)$。

3. 加入正则项的正则化

如式 (1.2) 所示，除了在数据或结构中嵌入先验知识之外，还有一类正则化方法是直接在代价函数中加入正则项 $R(\cdot)$。$R(\cdot)$ 的形式根据先验知识而定，按先验知识的不同，又可以分为关于参数的正则、关于隐层的正则以及关于输出的正则。

关于参数的正则是机器学习算法中最常用的正则方法，它直接约束网络中的参数 θ，其中最常见的是 2 范数参数正则 (ℓ_2 parameter regularization) 以及 1 范数参数正则 (ℓ_1 parameter regularization)。深度学习中的 2 范数正则通常被称为权值衰减 (weight decay)[64]，是深度学习中十分常用的正则方法[65−67]。相比于 2 范数正则，1 范数正则可以获得更为稀疏的参数分布，即一部分参数数值会趋向于 0，也在一些深度学习文章中有所应用[68]。

关于隐层的正则也是深度学习中一类常见的正则化方法[31, 32, 69]，它约束学习的特征表示 h，这实际上也是对参数的一个非显式约束。特征表示约束中，最常见的是特征稀疏性约束。Poultney 等[31] 提出在自编码器的编码器和解码器之间插入一个稀疏化模块，将自编码器输出的非稀疏特征表示转化为稀疏特征；Lee 等[32]