第 3 章

CHAPTER 3

基于深度学习的图像检索

本章根据不同的深度学习体系结构，介绍深度学习在图像检索领域的研究进展，主要包括基于卷积神经网络的图像检索、基于生成对抗网络的图像检索、基于注意力机制的图像检索、基于循环神经网络的图像检索和基于强化学习的图像检索。

3.1 基于卷积神经网络的图像检索

基于特征学习的卷积神经网络(CNN)已经被广泛应用于图像检索。用于图像检索的典型 CNN 结构如图 3-1 所示。CNN 由不同的层组成，包括卷积层、非线性层、归一化层以及全连接层等。通常，全连接层学习到的抽象特征被用于生成哈希码和描述符。

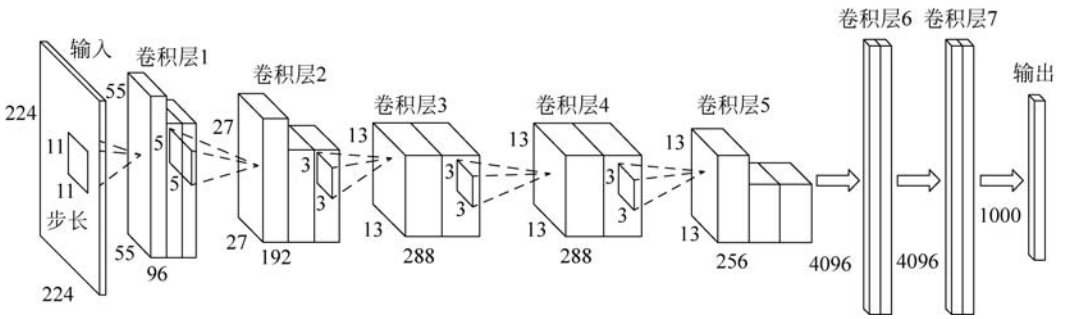

图 3-1　用于图像检索的卷积神经网络结构

2014 年，瑞典皇家理工学院的 Ali Sharif Razavian 使用从 OverFeat 网络提取的特征作为通用图像表示来处理图像分类、场景识别、细粒度识别、属性检测和图像检索等任务，通过与各种数据集上视觉分类任务中的最先进算法相比，CNN 获得了几乎一致的优异效果，实验结果表明，使用 CNN 学习获得特征是大多数视觉识别任务的主要选择。

同年，莫斯科物理技术大学的 Artem Babenko 在几个标准的图像检索基准数据集上评估深度神经网络在图像检索应用中的性能表现，并得出如下结论：①即使使用为分类任务

训练的 CNN 来检索图像，并且当训练数据集和检索数据集彼此差异很大时，CNN 也表现良好。当 CNN 在与检索数据集更相关的图像上重新训练时，这种性能会进一步提高。②图像检索的最佳性能不是在网络的最顶层获得的，而是在输出层的前两层获得的。即使使用 CNN 对相关的图像进行重新训练，这种效果仍然存在。

虽然复杂的图像外观变化对可靠检索构成巨大挑战，但鉴于 CNN 在各种视觉任务上学习的鲁棒性，2016 年，中国科学院大学的 Liu Haomiao 使用深度监督哈希（Deep Supervised Hashing，DSH）方法学习紧凑的二进制代码，在大规模数据集上进行高效的图像检索。DSH 的网络结构如图 3-2 所示。DSH 采用一种 CNN 架构，输入为图像对（无论两幅图像是否相似），输出为二进制编码。DSH 模型学习图像特征的二进制编码，其主要优点包括：①相似的图像在汉明空间上编码相似；②二进制编码计算效率更高。

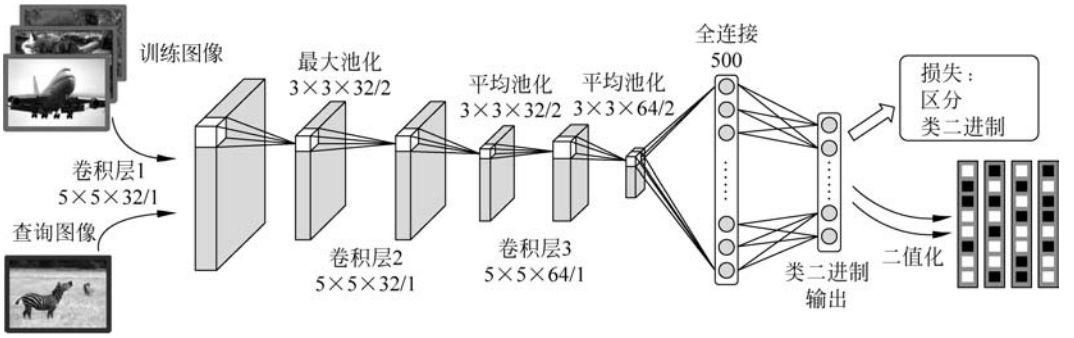

图 3-2　DSH 的网络结构

通过设计不同的损失函数，CNN 模型被大量用于生成哈希编码，从而进行有效的图像检索，例如，2016 年施乐欧洲研究中心的 Albert Gordo 使用三流连体网络，利用三重排序损失优化卷积区域中最大激活表示的权重；清华大学的 Cao Yue 使用配对损失构建相似度学习，利用量化损失控制哈希质量；2017 年清华大学的 Cao Zhangjie 提出加权成对交叉熵损失函数，用于从不平衡的相似性关系中进行相似性保留学习。

学习到的 CNN 抽象特征可用于不同模式下的图像检索，例如无监督图像检索模型 DBD-MQ（Deep Binary Descriptor with Multi-Quantization，多量化深度二进制描述器）、SADH（Similarity-Adaptive Deep Hashing，相似性自适应深度哈希）、DeepBit，有监督图像检索模型 FusionNet，以及半监督图像检索模型 SSDH（Semi-Supervised Deep Hashing，半监督深度哈希）等。

3.2　基于生成对抗网络的图像检索

2017 年，电子科技大学的 Wang Bokun 基于对抗学习机制在不同模态之间互相作用获得有效的共享子空间，提出一种对抗性的跨模态检索方法（Adversarial Cross-Modal

Retrieval，ACMR)，如图 3-3 所示。跨模态检索任务的核心是特征映射器和模态分类器之间的相互作用。特征映射器为公共子空间中的不同模态的项目生成模态不变表示，其目的是混淆充当对手的模态分类器；模态分类器试图根据其模态区分项目，并以这种方式控制特征映射器的学习。

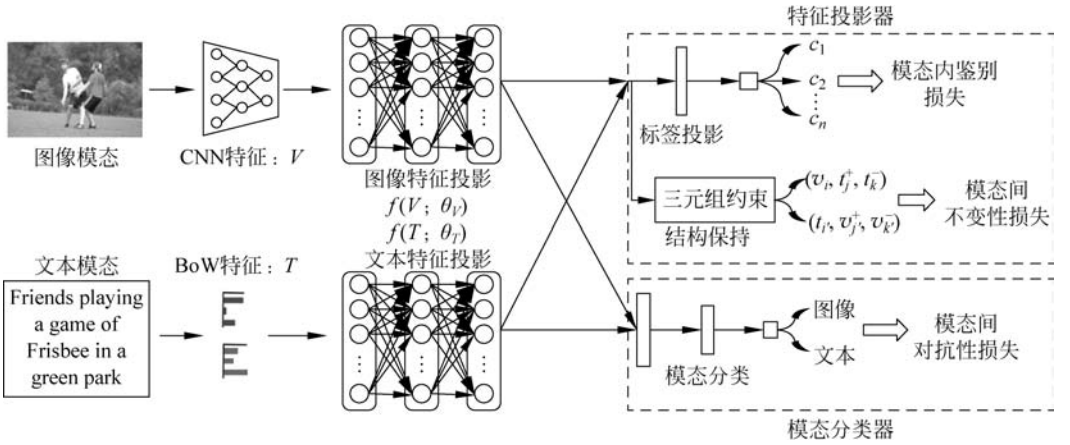

图 3-3　ACMR 的网络结构

在另一项工作中，中山大学的 Zhang Xi 提出一种具有注意力机制的对抗性哈希网络，通过选择性地关注多模态数据的信息部分来增强内容相似性的测量。

2018 年，电子科技大学的 Song Jingkuan 提出二进制生成对抗网络(Binary Generative Adversarial Network，BGAN)，利用无监督的方式实现图像检索，主要通过设计新的激活函数和目标函数解决两个问题：

(1) 如何在不松弛的情况下生成图像的哈希(二进制)表示；

(2) 如何利用哈希实现准确的图像检索。

BGAN 的网络结构如图 3-4 所示。

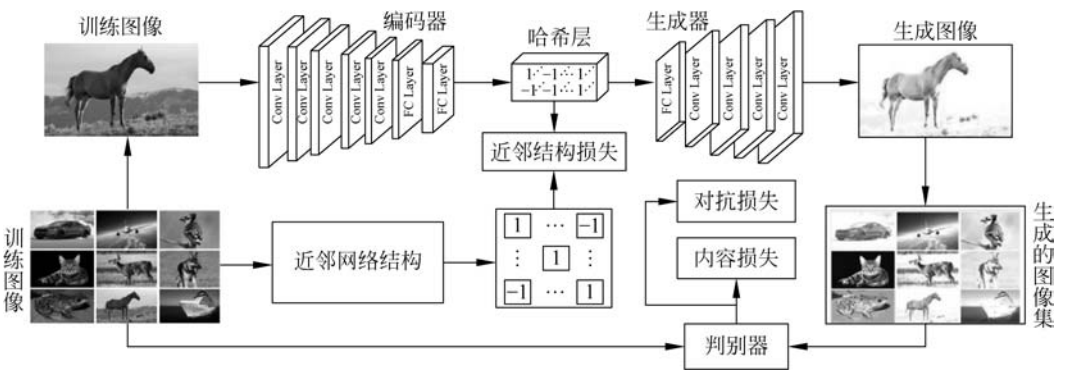

图 3-4　BGAN 的网络结构

同年，匹兹堡大学的 Kamran Ghasedi Dizaji 基于对抗网络提出一种新的深度无监督哈希网络 HashGAN，无须任何监督训练即可有效地输入图像的二进制表示。HashGAN 由生成器、判别器和编码器三部分组成，通过共享编码器和判别器的参数训练深度哈希网络。

与此同时，清华大学的 Cao Yue 研究另外一个 HashGAN，通过使用成对的条件 Wasserstein GAN 生成更多样本，进一步学习图像检索的哈希码。

西安电子科技大学的 Li Chao 提出一种自监督对抗性哈希（Self-Supervised Adversarial Hashing，SSAH）方法，它是早期尝试将对抗学习以自监督方式纳入跨模态哈希的方法之一。这项工作的主要贡献是利用两个对立的网络最大限度地提高不同模态之间表示的语义相关性和表示一致性，并利用自监督语义网络以多标签标注的形式发现高层次的语义信息。这些信息指导了特征学习过程，并在共同语义空间和汉明空间中保持了模态关系。SSAH 的网络结构如图 3-5 所示。

2019 年，中国科学院大学的 Gu Wendell 提出一种用于跨模态检索的对抗引导非对称哈希（Adversary Guided Asymmetric Hashing，AGAH）方法。如图 3-6 所示，该方法联合学习端到端架构中每种模态的特征表示和哈希码。为了增强特征学习部分，采用一个对抗性学习引导的多标签注意模块。在该模块中，首先利用对抗策略学习特征，达到跨模态表示的分布一致性，然后通过多标签注意力关注每个特征的标签信息，从而学习有区别的特征表示。此外，生成的二进制代码应保留每个项目的多标签语义。为了实现这一目标，采用多标签二进制码映射，可以为哈希码配备多标签语义信息，从而保证多标签语义的保持。为了保证所有相似对与不相似对的相似度更高，采用三元组边界约束和余弦量化技术进行汉明空间相似度保持，从而保留所有项目对之间的较高等级相关性。

西安电子科技大学的 Deng Cheng 通过将语义相似性保留目标与对抗性哈希学习框架相结合，提出一种无监督对抗性哈希（Unsupervised ADversarial Hashing，UADH）方法。如图 3-7 所示，UADH 包括三个神经网络：①编码器网络，用于从真实图像生成哈希码；②生成器网络，用于从哈希码生成合成图像；③判别器网络，旨在分别区分来自编码器网络和生成器网络的哈希码和图像对。

2020 年，印度理工学院马德拉斯分校的 Anubha Pandey 为基于零样本的草图检索提出一种多阶段生成模型。该模型的灵感来自 StackGAN 架构，多级模型的输出被馈送到孪生网络以学习更好地嵌入并减少枢纽点问题。使用多阶段生成模型，可以生成更接近原始图像特征空间的细化特征。此外，孪生网络使用对比损失函数区分投影空间中给定的生成和真实图像特征对，这种方法有助于将基于草图的图像检索问题简化为多个子问题。如图 3-8 所示，第 1 阶段，将草图特征投影到图像域；第 2 阶段，生成图像特征的细节信息；第 3 阶段，使用孪生网络生成更鲜明的特征。

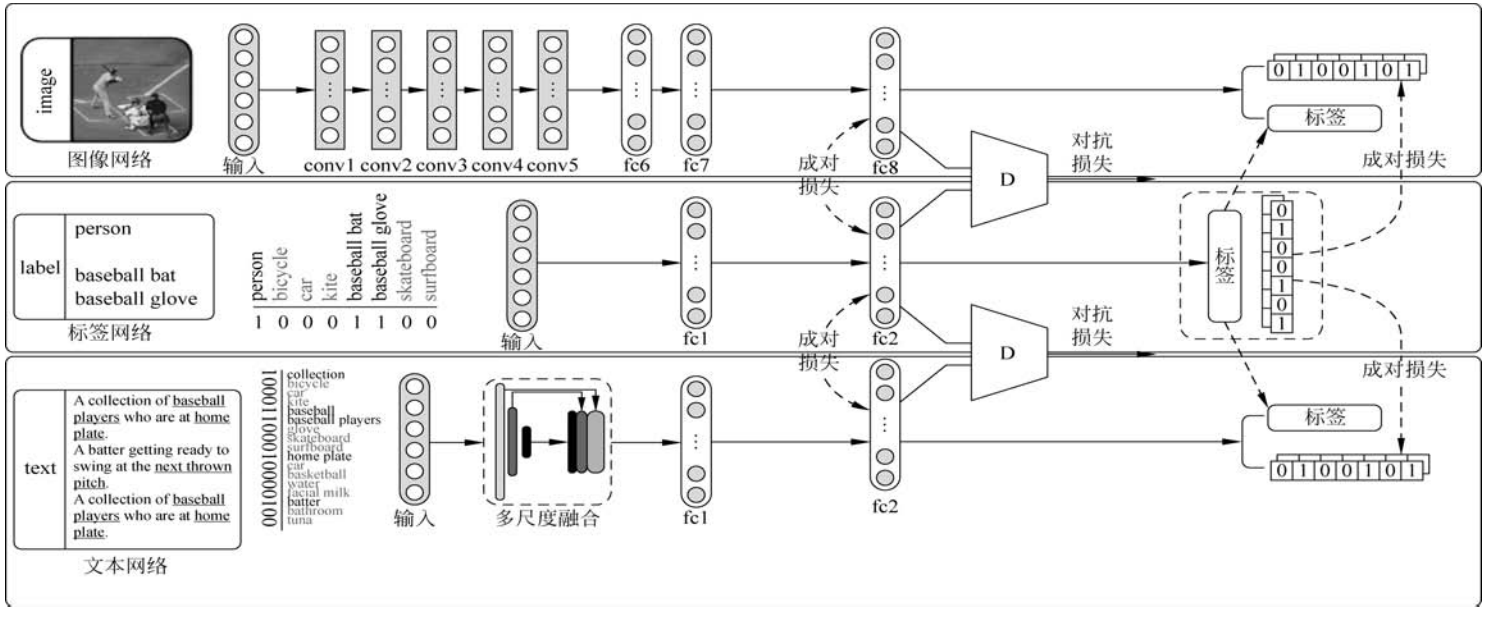

图 3-5　SSAH 的网络结构

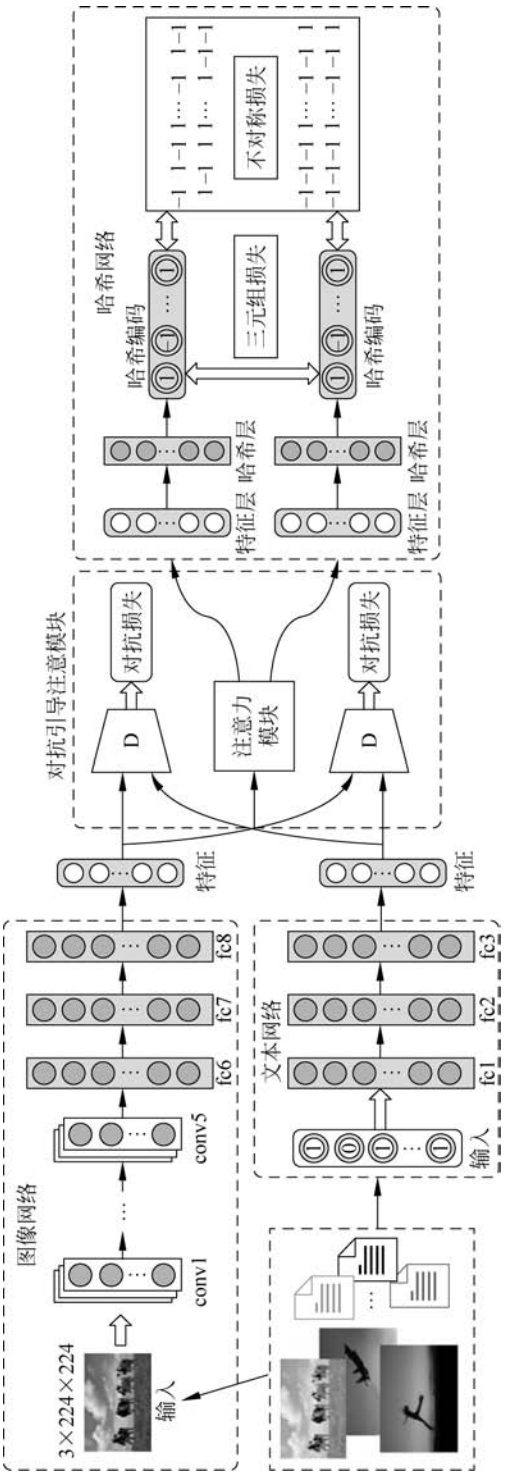

图 3-6　AGAH 的网络结构

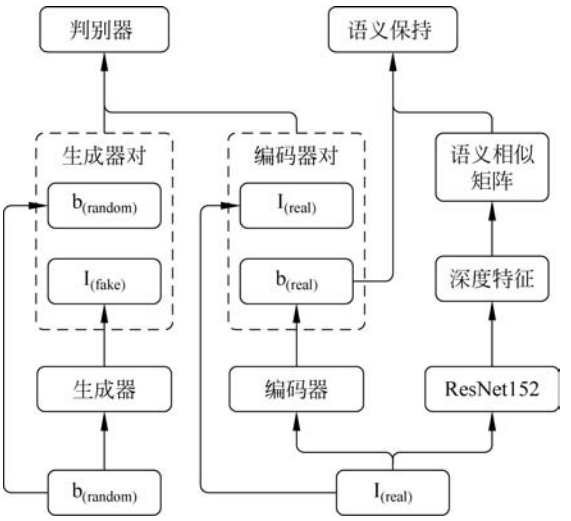

图 3-7 UADH 的网络结构

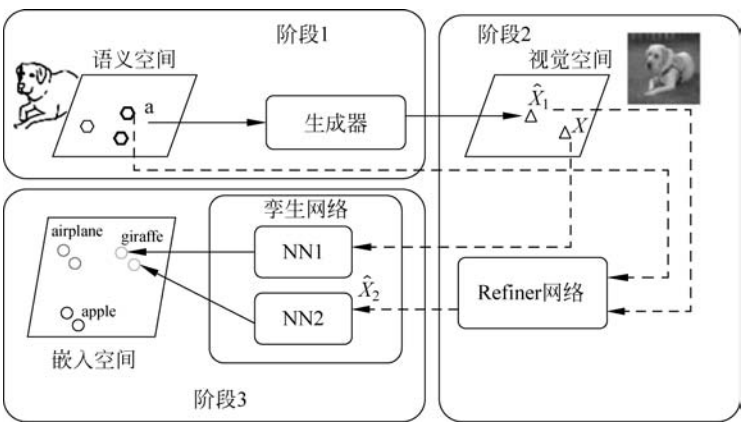

图 3-8 基于零样本草图检索的多阶段生成模型

3.3 基于注意力机制的图像检索

注意力机制是将显著性信息建模到特征空间中以避免背景噪声影响的一种非常有效的方式。

2017 年，韩国浦项科技大学的 Hyeonwoo Noh 提出一种适合于大规模图像检索的局部特征，称为深度局部特征(Deep Local Feature，DELF)。新的特征利用卷积神经网络，基于图像级别标注的地标图像数据进行训练。为了确保该局部特征对图像检索任务的有效性，引入了一个选取关键点的注意力机制，该机制与局部特征共享大部分的网络参数，如图 3-9

所示。提出的框架可以替代图像检索领域的其他关键点特征提取方式，带来更高精度的特征匹配和几何验证。

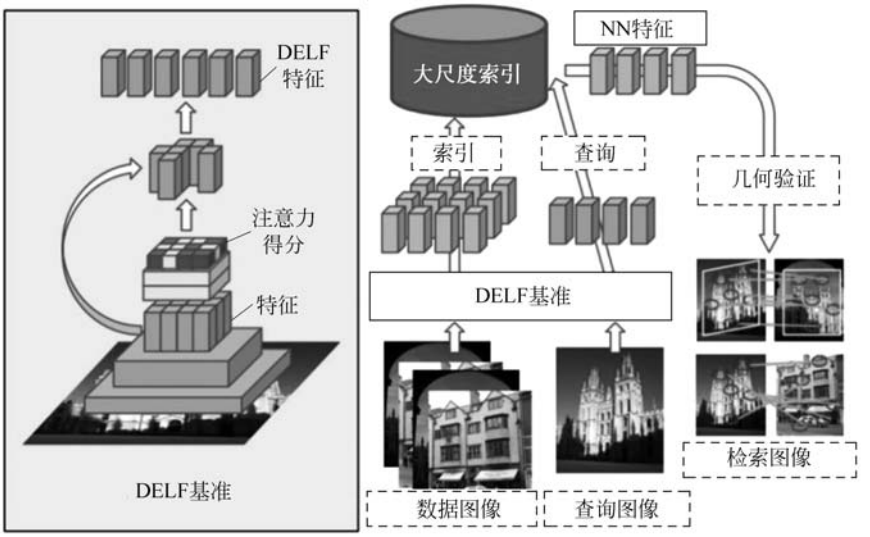

图 3-9　基于 DELF 和注意力的图像检索系统结构图

2019 年，新加坡南洋理工大学的 Huang Longkai 发现在深度哈希模型中使用基于梯度下降的算法可能会导致一对训练实例的哈希码在优化过程中同时朝着彼此方向更新，因此提出一种梯度注意力机制，通过神经网络为每对训练实例哈希码的梯度生成注意力，如图 3-10 所示。通过梯度注意力机制，可以加快学习过程。

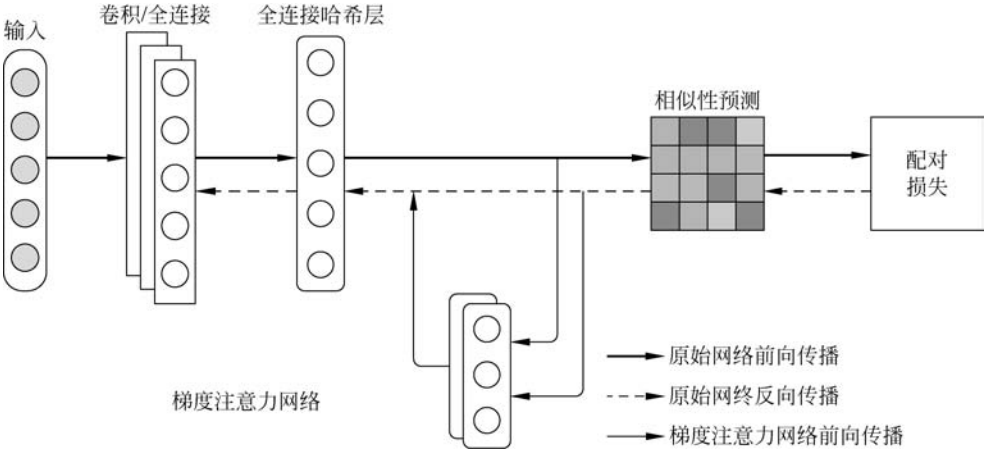

图 3-10　基于梯度注意力机制的深度哈希算法

2020 年，伦敦帝国理工学院的 Tony Ng 利用不同空间位置特征之间的二阶关系，结合二阶描述符的相似性，提出用于图像检索的二阶损失和注意力描述符（Second-Order Loss

and Attention for Image Retrieval，SOLAR）。如图 3-11 所示，左图表示在空间上学习最佳的相对特征贡献，右图表示在描述符空间中使用二阶相似度使集群之间的距离保持一致。该方法在图像检索和图像匹配两个不同任务上都带来了显著的性能改进。

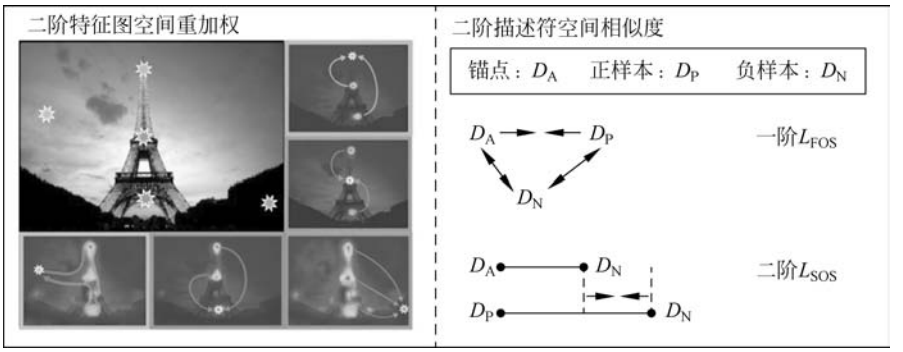

图 3-11 SOLAR 示例

3.4 基于循环神经网络的图像检索

近年来，利用循环神经网络（Recurrent Neural Network，RNN）和长短期记忆网络（Long Short-Term Memory，LSTM）学习图像描述进行图像检索的研究不断涌现出来。

2017 年，英国东英吉利大学的 Shen Yuming 使用基于区域的卷积神经网络和 LSTM 模块进行文本-视觉交叉检索。中国科学院西安光学精密机械研究所的 Lu Xiaoqiang 利用分层 RNN 生成用于图像检索的有效哈希码。

2019 年，新加坡南洋理工大学的 Chen Zhuo 使用 LSTM 模块作为孪生网络框架中卷积和全连接块之间的基准来学习图像检索描述符。

3.5 基于强化学习的图像检索

2018 年，清华大学的 Yuan Xin 利用强化学习进行图像检索，提出一种通过策略梯度进行可扩展图像检索的无松弛深度哈希方法，如图 3-12 所示。

2020 年，复旦大学的 Yang Juexu 提出一种具有冗余消除的深度强化哈希模型，称为深度强化去冗余哈希（Deep Reinforcement De-Redundancy Hashing，DRDH），它可以充分利用大规模相似性信息并通过深度强化学习消除冗余哈希位，减少图像检索相似度计算中的歧义，如图 3-13 所示。

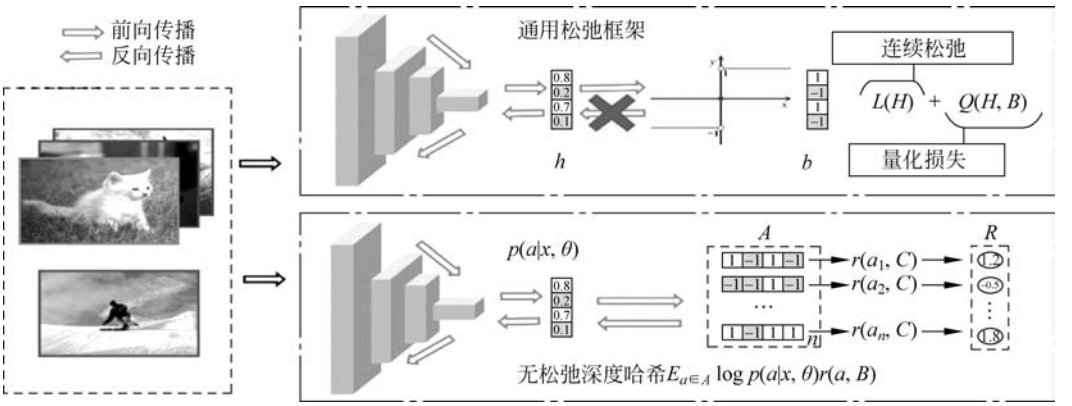

图 3-12　基于策略梯度的无松弛深度哈希方法

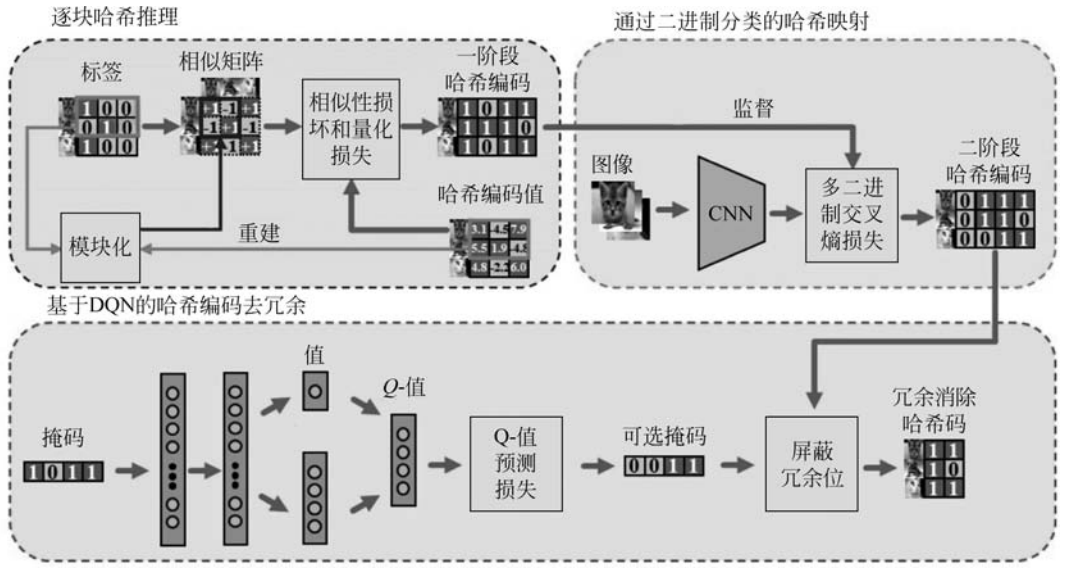

图 3-13　DRDH 模型结构图

3.6　本章小结

本章总结了近几年深度学习技术在图像检索领域的发展情况，包括基于卷积神经网络的图像检索、基于生成对抗网络的图像检索、基于注意力机制的图像检索、基于循环神经网络的图像检索以及基于强化学习的图像检索。

参考文献

[1] RAZAVIAN A S,AZIZPOUR H, SULLIVAN J,et al. CNN Features Off-the-Shelf: An Astounding Baseline for Recognition[C]. Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR),2014: 512-519.

[2] BABENKO A, SLESAREV A, CHIGORIN A, et al. Neural Codes for Image Retrieval[C]. Proceedings of the 2014 European Conference on Computer Vision(ECCV),2014: 584-599.

[3] LIU H,WANG R, SHAN S, et al. Deep Supervised Hashing for Fast Image Retrieval[C]. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2016: 2064-2072.

[4] CAO Y,LONG M, WANG J,et al. Deep Visual-semantic Quantization for Efficient Image Retrieval [C]. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2017: 916-925.

[5] CAO Z,LONG M, WANG J, et al. HashNet: Deep Learning to Hash by Continuation[C]. Proceedings of the 2017 IEEE International Conference on Computer Vision(ICCV),2017: 5609-5618.

[6] SHEN F,XU Y, LIU L, et al. Unsupervised Deep Hashing with Similarity-adaptive and Discrete Optimization[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(12): 3034-3044.

[7] LIN K,LU J,CHEN C S, et al. Unsupervised Deep Learning of Compact Binary Descriptors[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2018,41(6): 1501-1514.

[8] WANG B,YANG Y, XU X,et al. Adversarial Cross-modal Retrieval[C]. Proceedings of the 2017 ACM International Conference on Multimedia (ACMMM),2017: 154-162.

[9] ZHANG X,LAI H, FENG J. Attention-aware Deep Adversarial Hashing for Cross-modal Retrieval [C]. Proceedings of the 2018 European Conference on Computer Vision(ECCV),2018: 591-606.

[10] DIZAJI K G,ZHENG F, NOURABADI N S, et al. Unsupervised Deep Generative Adversarial Hashing Network[C]. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR),2018: 3664-3673.

[11] LI C,DENG C,LI N, et al. Self-supervised Adversarial Hashing Networks for Cross-modal Retrieval [C]. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),2018: 4242-4251.

[12] WEI S,LIAO L,LI J, ET AL. Saliency Inside: Learning Attentive CNNs for Content-based Image Retrieval[J]. IEEE Transactions on Image Processing,2019,28(9): 4580-4593.

[13] YANG J,ZHANG Y, FENG R,et al. Deep Reinforcement Hashing with Redundancy Elimination for Effective Image Retrieval[J]. Pattern Recognition,2020,100: 107-116.