

基于向量量化的深度学习图像重建方法的研究

杜成名, 王根, 李永春

巢湖学院, 安徽 合肥 238024

DOI: 10.61369/TACS.2025100035

摘 要 : 图像重建是从退化质量的观测中复原出清晰图像的过程, 其本质上是个信息不全的问题。传统深度学习方法有效但往往为了像素精确而损失了纹理细节, 所得到的图像是光滑无质感的。向量量化技术为解决此问题带来了新思路, 其实质就是把一张图像表示成一系列离散、有代表性的特征符号以便构造一个包含自然图像基本规律的符号库, 在这个库指引下做语义推理和细节生成即可完成图像重建工作。本文主要研究该方法背后的理论依据, 首先提出向量量化是如何对重建解空间施加强先验知识的, 然后分析基于向量量化自编码器的生成式重建模型的原理机制, 并讨论在离散符号空间中进行概率建模的优势。

关 键 词 : 向量量化; 图像重建; 生成先验; 离散表征; 深度学习

Research on Deep Learning Image Reconstruction Method Based on Vector Quantization

Du Chengming, Wang Gen, Li Yongchun

Chaohu University, Hefei, Anhui 238024

Abstract : Image reconstruction is the process of restoring a clear image from observations with degraded quality, which is essentially a problem of incomplete information. Traditional deep learning methods are effective but often lose texture details in pursuit of pixel accuracy, resulting in smooth and textureless images. Vector quantization technology brings new ideas to solve this problem. Its essence is to represent an image as a series of discrete and representative feature symbols to construct a symbol library containing the basic laws of natural images. Under the guidance of this library, semantic reasoning and detail generation can be carried out to complete the image reconstruction work. This paper mainly studies the theoretical basis behind this method. Firstly, it puts forward how vector quantization applies strong prior knowledge to the reconstruction solution space. Then, it analyzes the principle and mechanism of the generative reconstruction model based on the vector quantization autoencoder, and discusses the advantages of probabilistic modeling in the discrete symbol space.

Keywords : vector quantization; image reconstruction; generative prior; discrete representation; deep learning

引言

图像重建是计算机视觉领域的一个经典难题, 它旨在从模糊、有噪或低分辨率的退化图像中恢复出清晰且高质量的原始图像。从数学的角度来看, 这属于典型的“逆问题”, 我们已知结果以及部分过程, 要反推出原因。由于在退化过程中损失了大量信息, 因此这个反推的过程具有天然的不确定性, 即存在无穷多张可能的清晰图像都可以解释同一个退化图像, 这就是所谓的“病态问题”。早期的重建方法往往依靠人工设计的一些图像规律作为约束条件来寻找一个看似合理的答案, 比如平滑性或者边缘连续等等。而深度学习的优点在于可以从大量的数据集中自动学习到这些复杂规律。但是很多模型只是简单地以“像素级差异最小”为目标来进行训练, 这种做法虽然可以获得数值上接近的答案, 但从数学意义上来说却是所有可能解的平均值, 所以最终得到的重建图像就会显得非常粗糙, 并缺乏那种锐利的感觉。为了摆脱这个局限, 研究者们开始把目光转向生成式模型。它的思路不再是“猜”一个最接近的像素答案, 而是“创造”一张符合自然图像规律、且与观测信息一致的同时又视觉上逼真的图片。向量量化技术在这种情况下就有着独特的理论价值: 它为图像特征构建了一套自己的“字典”或者“密码本”, 使得重建的过程变成了“查字典造句”的过程—先根据退化图像推断出大概的“字符”, 再利用强大的语言模型把这些字符组合成通顺而生动的“句子”。

项目信息: 巢湖学院高层次人才科研启动经费项目 (KYQD202211; KYQD202114); 安徽省高校杰出青年科研项目 (2022AH020093); 安徽省自然科学基金项目 (2408085MD102)

作者简介: 杜成名 (1985-), 男, 安徽濉溪人, 博士, 高级工程师, 研究方向: 闪烁体探测器研制测试、数字信号处理以及图像重建工作。

一、向量量化与图像重建的理论关联

（一）向量量化：从连续特征到离散符号的本质提炼

向量量化就是一种数据压缩技术，它的工作叫做“归纳摘要”，我们设想一下，要用有限的词汇去描述世界上的一切事物，向量量化也是如此，它想要创建一个容量有限的“基础词汇库”，这个库里的每个“词”都是一个典型的特征模式，当我们碰到一个复杂的图像特征的时候，并不是把它直接记录下来，而是从词汇库当中找出那个最接近它的词来替代。这个过程我们就叫作“量化”。在深度学习里加上一层向量量化层，其理论目的可绝不仅仅是压缩这么简单。它实际上在神经网络内部设立了一个“信息瓶颈”，这个瓶颈迫使网络传递信息的时候，只能舍弃那些细枝末节、非本质的信息，只留下最核心、最有代表性的特征。这些被留下来的不同符号，就是对图像内容的一种非常抽象的“概念性”描述，这个基础词汇库就是在数据驱动下学到的一个关于“什么是自然图像”的强先验知识库。

（二）图像重建的病态性与先验知识的必要性

为什么先验知识这么重要呢？我们打个比方，假如我看到了一个模糊的影子，这个影子可能是只猫也有可能是个花瓶。如果没有先验的知识，我们就完全无法分辨出来。但是如果我们脑子里面存了清晰的记忆中的猫和花瓶的形象，就可以根据这个影子的轮廓，再调动这些脑补出来的先验知识拼凑出最合理的那个图像是什么样的。图像重建也是如此，退化的图像就像是一个“模糊的影子”，重建算法要从无数种可能中作出选择。向量量化学到的这些“基础词汇库”就给了我们一种重要的先验知识，它告诉我们自然图像是由这些基本、合理的“词汇”组成的，那些不在词汇库里的奇怪像素组合就是假的。这样就把重建过程限制在一个合理且有高视觉质量的空间里。^[1-5]

二、基于向量量化的深度学习图像重建理论模型

（一）基础框架：向量量化自编码器的两阶段建模

基于向量量化的重建方法一般遵循一个两阶段的理论框架，这就是它的“解耦”思想。第一阶段，学习“世界的字典”，也就是训练一个向量量化自编码器。编码器将一张高清图像分析成一系列连续的特征。量化层把这些连续特征映射到预设好的“基础词汇库”里去寻找对应离散符号。最后通过解码器只用那些离散符号就能尽可能完美地重新构建出原图来——如果成功了就意味着这个离散词汇库已经足够丰富能够捕捉自然图像的主要模式，并且解码器也掌握了如何把这些符号“翻译”回逼真画面的能力此时它本身就是一个非常厉害的图像生成工具。第二阶段，解决“逆向推理”问题。我们现在面对的不是高清图像，而是一张退化后的图像，我们需要一个专门的“推理网络”，学习如何从一张退化的图片中推测出第一阶段得到的离散符号序列。只要我们得到了预测出来的符号序列，就可以调用之前训练好的强大的解码器来生成最终的高清重建结果。这样的框架具有理论上的优势：将原本非常复杂、难以处理的从图到图的问题，分解成了两个相对简单一些的小问题——

一个是学习怎么生成一幅图像，另一个是学习怎么从退化的观测反推出来原来的图像符号。这样就使得任务变得容易许多，并且还借助了很强的生成先验保证输出的质量^[6-10]。

（二）模型演进：融入对抗训练以增强生成逼真度

在基础框架当中，如果重建的目标只是像素级相似的话，解码器就会倾向于产生平滑的结果。于是就有了向量量化生成对抗网络模型，它把生成对抗网络的思想引入进来。“生成器”与“判别器”的博弈是核心所在，在这个模型里，解码器充当着生成器的角色，而且多了一个判别器的任务就是分辨出这张图片到底是“真实的高清图”，还是“由解码器所生成的”。解码器不仅需要去重建图像，还要做到以假乱真，让判别器无法辨别出来，这种对抗性的训练方式，从理论上讲，是在使生成出来的图像分布和真实图像分布间的差异最小化，迫使解码器产生的结果无论在整体统计特性上诸如纹理、光泽、结构等等方面都不差于真实图片，并且在视觉感知层面达到极至程度的状态。

这种对抗训练机制促使解码器不断优化自身，在训练过程中，解码器会依据判别器的反馈不断调整生成策略。判别器每一次准确分辨出真实图片与生成图片，都会给解码器传递一个信号，告知其当前生成结果存在的不足。解码器接收到这些信号后，会针对性地改进，努力提升生成图像的质量，使其在纹理细节上更加细腻真实，光泽表现上更加自然逼真，结构布局上更加合理准确。经过多轮这样的对抗训练，解码器生成图像的能力会得到显著提升，最终生成出在视觉上几乎与真实图片难以区分的图像。^[7-15]

（三）核心突破：在符号空间内进行上下文推理

上面的模型假定各个离散符号之间是独立的，但现实中这种假设很难成立——特别是像图像这样的高度结构化数据中，不同区域间常常有着很强的上下文依赖和复杂的相互作用关系。为了应对这个挑战，当前理论研究有了重大突破：研究人员开始使用自回归模型（如带有注意力机制的变换器）来建模这些离散符号序列里隐藏着的复杂依赖模式。这就好比在已经有了一个完整的“视觉词汇库”之后，又掌握了一套非常抽象的“语法规则系统”。在做图像重建的时候，模型不会把每个符号当成孤立单元来进行预测，而是用序列生成的方法，在预测每一个当前位置上的符号时，会全面地融合参考之前已经生成出来的所有符号所共同构成起来的那个丰富的上下文信息。借助这种机制，模型可以生成具备高度全局一致性视觉内容，拿人脸图像修复来讲，当模型执行这项任务时，它既会单独产出左眼，也会按照已有的右眼位置、尺寸、角度信息以及鼻子和其他面部组件的结构特点等要素来动态调节生成流程，进而保证新产生的局部与整体在几何及语义层面保持协调统一状态，依靠这样的基于符号空间概率建模和序列推断策略，在很大程度上提升了所生成图像的合理性及其结构性完整性，并且促进了有关视觉表示学习方面取得了重要进展。

三、结束语

综上所述，本文从理论角度较为全面地阐述了向量量化深度学习图像重建方法。这一技术的理论贡献在于，它创建起一种离

散符号系统，给病态图像重建问题增添了一种结构性强的先验知识，并且这种先验不是人为指定的，而是经由数据学习来的关于自然图像的本质规律。它的理论模型表现出了某种层次化的分解思路，依靠向量量化去提炼出基本特征单元，从而构筑“视觉词汇库”，然后借助对抗训练改善解码器，使其变成一个可以从词汇库生成近似真实图像的“大师画师”。

参考文献

[1] 陈渤. 基于向量量化的深度学习图像重建方法的研究 [D]. 电子科技大学, 2025.DOI: 10.27005/d.cnki.gdzku.2025.005183.

[2] 张涛, 王彬沅, 付莹, 等. 基于深度学习的光谱图像超分辨率综述 [J]. 中国图象图形学报, 2024, 29(08): 2113–2136.

[3] 毕海霞, 况祖正, 李凡, 等. 极化 SAR 图像分类深度学习算法综述 [J]. 科学通报, 2024, 69(35): 5108–5128.

[4] 陈万强, 陆永华, 朱赟, 等. 基于深度学习图像配准的燃调表面缺陷检测算法 [J]. 测试技术学报, 2024, 38(06): 586–592.

[5] 刘冠, 邵继中, 王宇琪, 等. 风景园林图像与图形在深度学习中的应用分析及未来展望 [J]. 南京师大学报 (自然科学版), 2024, 47(02): 44–53.

[6] 张天驰, 刘宇轩. 深度学习驱动的水下图像处理研究进展 [J]. 计算机科学, 2024, 51(S1): 283–294.

[7] 邹伟. 基于深度学习的脊柱多模态医学图像分割与配准方法研究 [D]. 苏州大学, 2024.DOI: 10.27351/d.cnki.gszhu.2024.002629.

[8] 陈峰伟. 基于深度学习的 CT 图像肝组织区域分割方法研究 [D]. 苏州大学, 2024.DOI: 10.27351/d.cnki.gszhu.2024.001265.

[9] 刘乐元, 孙见池, 高韵琪, 等. 基于深度学习的单图像三维人体重建研究综述 [J]. 华中科技大学学报 (自然科学版), 2024, 52(05): 98–122.DOI: 10.13245/j.hust.240614.

[10] 权静. 深度学习算法在新能源汽车图像识别故障检测方法中的应用 [J]. 汽车测试报告, 2024, (04): 152–154.

[11] 赵宇鹏. 深度学习在图像识别中的应用 [J]. 信息与电脑 (理论版), 2024, 36(04): 126–128.

[12] 潘美莲, 陈洁. 深度学习算法的图像识别技术在电子元件分拣中的应用 [J]. 电脑编程技巧与维护, 2024, (02): 140–142+169.DOI: 10.16184/j.cnki.comprg.2024.02.029.

[13] 杨米娜. 基于深度学习的图像识别系统研究 [J]. 山西电子技术, 2024, (01): 113–115.

[14] 李益. 深度学习在遥感图像分割算法中的应用 [J]. 集成电路应用, 2024, 41(02): 324–325.DOI: 10.19339/j.issn.1674–2583.2024.02.150.

[15] 苏远超, 许若晴, 高连如, 等. 基于深度学习的高光谱遥感图像混合像元分解研究综述 [J]. 遥感学报, 2024, 28(01): 1–19.