分类号 C8/364　　密级 公 开
U D C 0005655　　编号 10741

兰州财经大学
LANZHOU UNIVERSITY OF FINANCE AND ECONOMICS

硕 士 学 位 论 文

论文题目 基于 A-HRNet 对抗自编码器的矩阵填充算法研究

研 究 生 姓 名：黄梓玉
指导教师姓名、职称：黄恒君 教授
学 科、专 业 名 称：应用经济学 统计学
研 究 方 向：调查技术与统计分析

提 交 日 期：2024 年 6 月 5 日

独创性声明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名：贾梓玉　　签字日期：2024.6.3

导师签名：黄恒君　　签字日期：2024.6.3

关于论文使用授权的说明

本人完全了解学校关于保留、使用学位论文的各项规定，同意（选择“同意”/“不同意”）以下事项：

1.学校有权保留本论文的复印件和磁盘，允许论文被查阅和借阅，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文；

2.学校有权将本人的学位论文提交至清华大学“中国学术期刊（光盘版）电子杂志社”用于出版和编入 CNKI《中国知识资源总库》或其他同类数据库，传播本学位论文的全部或部分内容。

学位论文作者签名：贾梓玉　　签字日期：2024.6.3

导师签名：黄恒君　　签字日期：2024.6.3

Research on matrix filling algorithm based on A-HRNet Adversarial Auto-encoder

Candidate : Huang Zi Yu

Supervisor: Huang Heng Jun

摘 要

矩阵填充技术提供了一种高效的方式来弥补数据的缺失。数据缺失是数据分析、机器学习、图像处理等诸多领域中的共性问题。利用矩阵填充，可发现隐藏在数据中的规律及变化规律，从而对其进行更深层次的认识和运用。矩阵填充还可以改善数据的完整性和质量，提高预测和决策的准确性。因此，在数据处理和实际应用中，解决矩阵填充的问题显得尤为关键。矩阵填充涉及多种数据类型，图像类填充是一个关键问题，因此构建可修复图像类数据的矩阵填充模型就显得尤为重要。通过填充缺失的像素，可使图像恢复完整，提高图像的可视化效果。图像填充还能优化图像的后续处理流程，例如图像的分类、对象的识别以及图像的重建工作。因此，在图像处理和分析领域，解决图像填充的问题具有不可忽视的重要性。

在此基础上，提出一种融合 HRNet 与注意力机制的对抗式自编码模型 AH-AAE。AH-AAE 模型从两个方面对基本对抗自编码器进行了改进。第一个改进的点是将生成器中的编码器部分进行重构为 HRNet，使其能够更好地表达和保持图像的细节，进而改善修复结果的质量和逼真度。第二个改进的地方是将注意力机制引入生成器中，注意力机制通过图像细节加强了对局部填充的控制，提高图像填充质量。当图像部分受损时，可以保证填充的结果只覆盖需要的部分，而不会对整幅图像做多余的改动。本文提出的 AH-AAE 模型通过引入通道注意力在跳跃连接处构建通道相似性融合模块来丰富通道之间的特征关系；在解码器网络中，结合空间注意力和位置编码的位置融合模块用于增强边界位置信息的表达。

为了验证模型效果，在 MS-COCO 以及 KITTI 数据集上进行了多项实验，证明了 AH-AAE 模型的填充能力、去噪能力以及环境感知能力，在多个指标上均达到了最优。本文所提的模型性能良好，未来可为敦煌壁画修复、计算机视觉以及视频填充等图像类数据修复领域发展提供参考。

关键词：深度学习 矩阵填充 注意力机制 HRNet 深度特征

Abstract

Matrix filling technology provides an efficient way to make up for the lack of data. Data missing is a common problem in many fields such as data analysis, machine learning and image processing. By using matrix filling, we can find the laws hidden in the data and the changing laws, so as to understand and apply them in a deeper level. Matrix filling can also improve the integrity and quality of data and improve the accuracy of prediction and decision-making. Therefore, in data processing and practical application, it is particularly critical to solve the problem of matrix filling. Matrix filling involves many data types, and image class filling is a key problem, so it is particularly important to build a matrix filling model that can repair image class data. By filling the missing pixels, the image can be restored completely and the visualization effect of the image can be improved. Image filling can also optimize the subsequent processing flow of images, such as image classification, object recognition and image reconstruction. Therefore, in the field of image processing and analysis, it is of great importance to solve the problem of image filling.

On this basis, an adversarial auto-encoding model AH-AAE integrating HRNet and attention mechanism was proposed. The AH-AAE model improves the basic adversarial auto-encoder from two aspects. The first aspect of the transformation is to reconstruct the encoder part of the generator into HRNet, which allows it to better express and preserve the

details of the image, thereby improving the quality and fidelity of the restoration results. The second aspect of the transformation is the introduction of attention mechanisms into the generator, which enhances the control of local fill through image details, improving the quality of image fill. When part of an image is damaged, you can be sure that the result of the fill will cover only the part you need, without making unnecessary changes to the entire image. The AH-AAE model proposed in this paper enriches the feature relationship between channels by introducing channel attention and constructing a channel similarity fusion module at the hopping connection. In the decoder network, the position fusion module combining spatial attention and position coding is used to enhance the expression of boundary position information.

In order to verify the effect of the model, a number of experiments were carried out on MS-COCO and KITTI datasets, which proved that the filling ability, denoising ability and environment perception ability of the AH-AAE model reached the optimal in multiple indicators. The model proposed in this paper has good performance and can provide a reference for the development of image data restoration fields such as Dunhuang mural restoration, computer vision, and video filling in the future.

Keywords： Deep learning; Matrix filling; Attention mechanism; HRNet; Depth characteristics

目 录

1 引言

1.1 研究背景及研究意义

缺失数据是指由于客观或主观原因而包含不完整数值的数据，它对各个领域的实际应用都有重大影响。由于某些客观因素，如隐私保护、不能被直接观察到的特性等，往往会造成数据缺失。此外，由于一些人为的原因，如疾病或事故，也可能造成资料的缺失。在数据科学、信息处理、社会调查、医学、生命科学、计算机视觉等许多领域，都存在大量的缺失数据。统计学和机器学习等领域面临着缺失数据带来的巨大不确定性，如何有效解决这些问题是大数据研究的目标。缺失数据补全的目标是使用适当的方法预测缺失数据，以减少不完整数据对后续分析的影响。

在处理缺失数据时，一般采用删除缺失值、内插法和回归模型等方法。删除缺失数据会减少样本中的数据量，从而影响结果的准确性。插值法是使用计算或模型来预测数据，如平均插值法、回归插值法和多重插值法。回归模型主要用于根据现有数据预测数据；最常用的方法是线性回归、逻辑回归和随机森林。然而，缺失数据的填充必须充分考虑数据本身的特点和缺失数据的分布情况，并最终考虑到后续分析。采用适当的填充措施可以增加数据的完整性，提高模型的准确性，减少对计算结果的影响。缺失数据是当今各领域普遍存在的一个问题，也是一个非常棘手的问题。建立合适的填充模型将有助于最大限度地发挥数据在信息处理中的作用，在确保准确性的前提下，为各领域的数据分析和决策提供便利。

矩阵填充是一种新的补全方法，能有效解决信息缺失和高维特征的现代问题。Netflix 的低秩矩阵填充问题就是这样一个例子，它根据电影评分预测用户的喜好，然后进行推荐。将这一过程转换为矩阵时，用户以行表示，已观看过的视频以列表示，评分以矩阵元素表示。填充矩阵的目的是将已知的电影评分与周围的环境因素相关联，从而推导出其他矩阵元素的可能值。矩阵本质上是一个低秩矩阵，用精确算法重构后会得到一个完整的矩阵。矩阵填充也可应用于其他数据结构，如文字处理、推荐等。

矩阵填充技术在计算机视觉、推荐系统、图像和视频处理以及协同过滤等多

个领域都有重要应用。其中，矩阵填充技术是计算机视觉领域的一个重要研究方向，已被提出作为一种图像分割方法。在实际应用中，推荐系统在提高用户满意度方面发挥着至关重要的作用。这种方法有效地解决了用户不了解自己偏好的问题。矩阵填充技术可以准确预测用户对产品的兴趣，并提供个性化的推荐。图像和视频填充是矩阵填充应用中最常用的领域，其中缺失区域的位置和内容是根据已知内容推导出来的，例如物体去除、污点去除和划痕修复。这项技术可以将受损图像恢复到原始状态，从而提高图像质量和可视化效果。协同过滤是推荐系统和信息检索的一种有效方法。通过分析海量用户的交互记录，创建一个预测模型，为用户提供推荐。这种方法通过矩阵填充提高了推荐系统的性能。矩阵填充方法的研究和应用使实际问题的解决更加高效和准确。

缺失数据的填充，常见的是数值类型和图像类型的填充方法。就数值类型而言，通常使用数值之间的相关性进行插值。在分析可用信息时，一般方法是找出与缺失值高度相关的其他变量，并从这些数据中推导出缺失值。能够有效地填补缺失数据，同时保持数据的完整性。图像类别数据的填补方法主要基于像素之间的相关性。图像的像素具有一定的空间相关性，即像素之间的关系一般是有规律的。因此，可以根据现有的像素数据来估计缺失的像素。矩阵填充法用于图像处理中，可以更好地保持图像的完整性。此外，近年来关于将矩阵填充方法与神经网络相结合的研究也有了长足的发展。基于神经网络的数据处理技术，可以对数据进行复杂的非线性建模，从而更准确地预测缺失值。基于神经网络的矩阵填充技术在许多学科中发挥了重要作用，为研究人员解决了大量实际问题。总之，矩阵填充是解决多种类型数据缺失问题较好的方法之一。通过挖掘数据之间的关联关系，并结合建模算法准确填充缺失数据，可以提高数据的完整性和可用性。这些研究方法具有重要的理论和实践价值。

1.2 国内外研究现状

随着计算机和通信技术、传感器技术及嵌入式技术的迅速发展，传感器设备被大量运用于遥感(Liu, P. et al., 2019)、医疗(Jifara et al., 2019)、日常生活(Park et al., 2019)等多个领域。传感器节点采集的数据能帮助人们更好地了解监测区域，然而由于设备软硬件故障、网络传输问题及工作环境恶劣等不可控因素，传感器

网络中的图像数据存在不同程度的缺失，降低了数据的可用性(谭丹丹 等, 2007)。直接丢弃缺失数据简单易行，但当缺失数据比例较大时，此方法会造成原始数据失真，损害推理能力。因此，对图像缺失数据进行填充是一种更合理且非常必要的方法(张网娟等, 2019)。图像填充的目的是在一个给定的 mask 情况下，填充缺失区域的像素，使其整体达到纹理和结构一致性，或者语义和视觉可信。在对图像数据进行填充的方法上已有较多研究，根据算法使用的硬件设备与自身内在逻辑，主要可以分为传统算法与深度学习算法。

（1）矩阵填充与传统算法

k 近邻（k-Nearest Neighbor, KNN）填充算法是通过计算缺失数据样本与完整数据样本之间的欧式距离，选出距离最小的 k 个样本作为缺失样本的最近邻，再通过距离的反比加权平均而得到缺少数据的填充值(Pan et al., 2015; Zhang, 2012)，该算法的核心在于 k 值的选择，不同的 k 值会造成显著的结果差异。均值填充是典型的基于统计的方法，将确实的数据用其他所有对象的取值的平均值进行填充(金勇进 等, 2000)，方法简单易行，但填充精度相对较低。大多研究采用针对传感器网络数据的特点来进行算法的选择，如曲志坚等通过属性约简划分属性的重要性，再根据属性的重要程度选择相应的填充算法(曲志坚 等, 2015)，王小平和曹立明考虑数据的时间相关性，使用局部时间索引策略对时间序列中存在的缺失数据进行填充(王小平 等 tan, 2002)，SRINIVAS 和 PATNAIK 考虑数据时空相关性，使用概率主元分析和混合概率主元分析论证了同时使用时间和空间信息可以提高填充的准确性(Srinivas et al., 1994)。

（2）矩阵填充与深度学习算法

对图像缺失数据填充本质上属于从低质量图像中恢复高质量图像，是计算机视觉中的图像去噪、图像超分辨率重建领域目前普遍研究的问题之一。1988 年就已经提出一种利用神经计算网络恢复由已知的平移不变模糊函数和加性噪声退化的灰度图像的方法(Zhou et al., 1988)，之后的 1990-2000 之间，也产生了基于多状态自适应线性神经元的边缘检测算法、类比细胞神经网络通用机算法以及改进的 Hopfield 网络，这就允许了一个神经元有一个有界的时间延迟可与其他神经元进行通信(Paik et al., 1990, 1992; Venetianer et al., 1995)。

卷积神经网络（Convolutional Neural Network, CNN）由于其独特的卷积计算

方式，以分层的方式收集局部特征作为图像表示，在计算机视觉中发挥出极为优秀的作用(Krizhevsky et al., 2017; Simonyan et al., 2014; Szegedy et al., 2015)。CNN 在视频空间和时间维度进行训练，可有效增强其空间分辨率以及生成器网络的结构足以在任何学习之前捕获大量的低级别图像统计信息(Kappeler et al., 2016; Ulyanov et al., 2018)。随着神经网络的普遍使用，将非局部操作引入递归神经网络获得非局部递归网络，并在图像恢复中获得第一次尝试，关键之处在于非本地模块可以灵活地集成到现有的深度网络的端到端的训练，捕捉每个位置和它的邻居之间的深层特征相关性(Liu et al., 2018)。随后也有人开始着眼用于图像复原任务的深度神经网络设计，大多数深 CNN 基于 IR 模型不充分利用原始低质量图像的层次特征，从而导致相对较低的性能(Liu, X. et al., 2019)。而有效的残差密集网络(RDN)来解决 IR 中的这个问题，通过利用所有卷积网络的层次特征，在效率和有效性之间进行更好的权衡分层问题(Zhang et al., 2020)。Dong 等首先将深度神经网络应用于图像超分辨率重建的研究上(Dong et al., 2014)，随后，Lim 等 Lim et al. (2017)与 Zhang 等(Zhang, Y. et al., 2018)也在图像超分辨率上基于残差网络与注意力机制对其进行深层次神经网络模型构建取得了较为显著的成果。与此同时，一些深度学习模型的提出也被应用于图像去噪，如 DnCNN(Zhang et al., 2017)、RPCNN(Xia et al., 2020)和 BRDNet(Tian et al., 2020)等。尽管 CNN 在获取图像局部特征上具有显著的优势，但在捕捉全局要素时则存在一定的问题，虽然可以通过扩大感受野的方式来进行直观的解决，但带来的一系列危害性后果也是无可避免的(Peng et al., 2021)。以 transformer 架构为主的注意力机制的出现(Vaswani et al., 2017)，很好的解决了 CNN 无法顾及长距离计算机视觉任务(Dosovitskiy et al., 2020; Wu et al., 2020)。ViT（Vision Transformer）方法通过将每个图像分割成具有位置嵌入的小块来构建一系列标记，并应用级联变换器块来提取参数化矢量作为视觉表示(Dosovitskiy et al., 2020)。由于注意力机制和多层感知器(Multilayer Perceptron, MLP)结构的存在，transformer 反映了复杂的空间转换和长距离特征依赖，实现了全局表示。但注意力机制忽略了局部特征细节，导致其在局部信息填充上的效果差强人意。

（3）矩阵填充与注意力机制

注意力机制在图像分类、目标检测、语义分割、视频理解、图像生成、三维

视觉、多通道任务和自监督学习等许多视觉任务中取得了巨大的成功。将计算机视觉中的各种注意力机制进行了全面的回顾和分类，可大体分为通道注意、空间注意、时间注意和分支注意等(Guo et al., 2022)。针对注意力机制这一部分，早期已有学者对其进行研究。2004 年拉维研究了选择性注意和认知控制的负荷理论。这一理论解决了长期存在的早期与晚期选择的争论，并阐明了认知控制在选择性注意中的作用(Lavie et al., 2004)。之后，Sao 等人介绍了双向注意流(BIDAF)网络，这是一个多级分层过程，代表不同粒度级别的上下文，并使用双向注意流机制来获得查询感知的上下文表示，在无需早期摘要的同时提出替代方法，扩展了自我关注机制以有效地考虑序列元素之间的相对位置(Seo et al., 2016)。随后有学者提出了一种替代方法，扩展了自我关注机制以有效地考虑序列元素之间的相对位置或距离的表示(Shaw et al., 2018)。在 CNN 的扩展方面，Bello 考虑将自我注意用于辨别性视觉任务作为卷积的替代(Bello et al., 2019)。为了对注意力机制有一个更好的一般理解,Zhu 提出了一项实证研究，在一个广义的注意力公式中消融了各种空间注意力元素，该公式包括主导的 Transformer 注意力以及流行的可变形卷积和动态卷积模块。为了使模型更易于访问，引入一个开源工具，在多个尺度上可视化注意力，每个尺度都提供了关于注意力机制的独特视角(Vig, 2019; Zhu et al., 2019)。更有学者提出应更专注于视觉解释的注意图，并以此来代表图像识别中注意位置的高响应值，同时针对注意力机制中存在的基于时间注意力机制方法而导致的识别错误和细节丢失得到进一步改善(Fukui et al., 2019)。为了解决上述问题，在自编码器神经网络中引入注意力机制来解决视频弹幕问题(Yan et al., 2019)。

在近两年关于注意力机制的研究方面，更多的是与其他网络进行了融合，研究自我注意在学习鲁棒表征中的作用(Zhou et al., 2022)、结合 Transformer 学习 sparse Boolean 函数的样本复杂性(Edelman et al., 2022)。而注意力机制在识别特征影响的极性方面存在弱点，尤其是结合线性化 Transformer 不受特定的电感偏差影响，那么忽略高层次的抽象就会对学习文本情感特征产生负面影响，进一步降低情感分类性能(Liu et al., 2022; Wu et al., 2022)。为了解决此问题，可通过整合情感智能 (El)和注意力机制来改进 LSTM 网络来提高分类精度(Huang et al., 2021)。注意力机制中导致精度偏低的原因也包括其内部的独立计算会导致噪声

和模糊的注意权重变化来压制性能，设计一个关注模块，在所有相关向量之间寻求共识来增强适当的相关性并抑制错误的相关性(Gao et al., 2022)；其内部依赖位置感知的感应偏差以及位置不可知无编码的部分添加到机制中可会导致精度降低(Ma et al., 2022)。

图像填充作为矩阵填充的实际应用领域之一，对多个领域都有着很强的影响。统计方法大多是利用图片的相关性，转换为灰度图，获取边缘，闭运算填充图像空间，再寻找以及绘制轮廓坐标点来实现图像填充。引入时间以及空间的概念后，图片的深度增加，精度略有提升。随着方法的迭代，顺延神经网络的发展，图像填充也与之结合，意在通过其内部相关关系来扩充图像。由此，引入注意力机制可在扩大感受野的同时尽量减少其他信息的损失，并使用多头注意力机制构造结构感知模型，扩大目标范围。细化损失函数，防止模型因拟合度欠佳产生噪声等影响，增强图片特征提取的能力。

以此为线索，构建矩阵填充模型处理图像类数据产生一系列值得讨论的问题：1、图片质量的选取。对于图像填充来说，不同形式的图片修复的完整度也不一样。那么在图片的选择上，就要既包含灰度图片、彩色图片、正常分辨率是图片以及部分有深度的图片。在大部分图像修复的算法中，能进行修复的都是带有显著特征的图像。如果缺失部分与旁边的依存关系不强，那么得到的结果也是不和谐的。所以，如何选取合适的图片是首要任务；2、填充算法的构建。如何提取更全面的特征。卷积神经网络提取特征能力稍弱，尤其是边界特征，如何扩大感受野就成了关键问题，有学者通过感受野分为密集和稀疏两部分分开操作确保采样的完整性。还有人结合了 transformer，在卷积层以及池化层添加了注意力机制来解决这类问题。那么由于处理的是图片数据，结合 transformer 更适合应用于视频数据，图片也可进行操作，但效果并不那么好。但注意力机制是目前获取特征的较好的方式，如何结合，以及在哪里结合可以避开之前的缺陷获得更优良的结果是重中之重。

1.3 研究内容与技术路线

1.3.1 研究内容

本文在原有矩阵填充的方法上进行改进，基于对抗编码器，融合注意力机制与 HR-Net 对原有模型进行创新，将提取目标更精准、算法更优良、结果更清晰，并可用于多种应用是本文研究的最终目的。其中研究主要内容如下：

1、引入注意力机制，将通道注意力与空间注意力进行改进。

（1）针对图像特征提取，创建通道相似性融合模块。通过将每个高层特征图看作一个区域特定的响应，不同区域响应也是相关联的，通过各个通道以及区域之间的关联性，可捕获不同的、更全面的区域响应。关键在于不仅可获得近处的信息，也可以捕捉到远距离区域的相对深度信息。这种捕捉到信息响应的结构感知模型可通过注意力机制来实现特征的具体计算值，每个通道的最终特征是所有通道的特征与原始特征的加权和。通过捕获特征图之间的远程依赖关系，可获得了编码场景结构丰富上下文信息的聚合特征。

（2）针对图像数据边界缺失，引入权重向量以及位置编码增强融合表达。针对图像填充结果的清晰度。本文使用的方法是融合不同尺度的特征并强调重要细节以实现更清晰的深度估计。若编码器无经操作直接经过跳跃连接获得的信息更多为缺乏对局部细节的进一步处理，这种简单操作忽略不同层次之间的差距，导致预测深度图像中边界位置模糊。适当的处理边界缺失并不损失其他部位的信息是关键，如果能够清晰的知道边界特征的类别以及位置，就可以更精确的恢复图像的原有位置。通过计算卷积层所需的相应权重来进行融合，权重与自身相加的计算可保证某项特征的计算不会缩减至 0。

2、引入 HR-Net 网络，保持图像数据输出形式的分辨率。针对不同图像的分辨率像素点问题，不仅要模型可以处理基础图像，还可以处理具有深度的、分辨率高的图像。本文是通过引进 HR-Net 来实现，HR-Net 的优势在于通过并行多个分辨率的分支，加上不断进行不同分支之间的信息交互，同时达到保持图像高分辨率、增强语义信息以及精准位置信息的目的。

1.3.2 技术路线

对本文整体框架与思路进行整理，获得文章技术路线图如下图 1.1 所示。

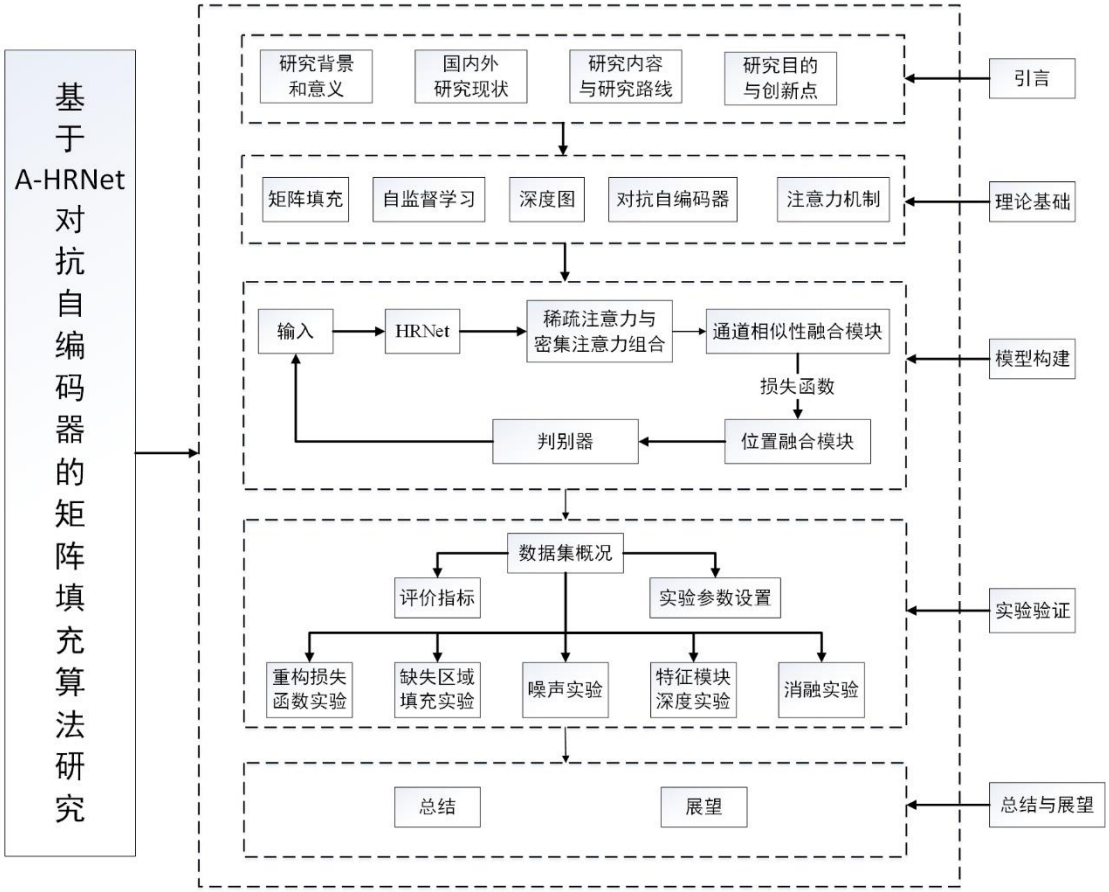


图 1.1 研究思路与技术路线图

通过图 1.1 技术路线图可以看出，本文通过“实际问题-理论基础-方法创新-模型构建-验证模型”的逻辑进行展开具体内容。图像在计算机中的表现形式是像素点的结合，从人眼中所看到的图像作为出发点，如何提高图像分辨率、清晰度，使图像获得更直观的表达是本文的研究重点。在原有矩阵填充的方法上进行改进，将提取目标更精准、算法更优良、结果更清晰，并可用于多种应用是本文研究的最终目的。

1.4 研究目的与创新点

1.4.1 研究目的

本研究旨在构建更高效的矩阵填充模型，以便为真实场景中得到更广泛的应用。图像填充作为矩阵填充的具体应用，对多个领域都有着很强的影响。

近年来，图像修复取得了重大进展，然而如何恢复纹理清晰、结构合理的破损图像仍然具有挑战性。由于卷积神经网络有限的感受野，一些特定的方法只处理规则纹理而丢失整体结构。另一方面，基于注意力机制的模型可以学习到更好的长期依赖关系用于结构恢复，但受限于大图像尺寸的推理计算量大。

卷积神经网络的各项优异推动了卷积自编码器的产生，严格上说，卷积自编码器属于自编码器的一个特例，它使用卷积层和池化层代替了原有的全连接层。经过卷积，可使二维图像的信息获得更全面的表达，但这种表达也不是完整的，会存在信息损失的情况。而经卷积自编码器延伸得到的多种自编码器模型处理图像的效果各有不同。有学者通过傅里叶卷积度对全局感受野的频率域特征进行编码，用于分辨率鲁棒的修复。但是它们不能保证图像的整体结构，对于弱纹理的图像效果较差。进一步地，利用具有长程依赖关系的基于 transformer 的方法，先对低分辨率的图像进行填充，再利用卷积神经网络对其进行上采样。但这种方式使变压器对于大图像需要巨大的内存占用。同时，生成式对抗网络种在图像生成和改进方面取得了显著的成果，例如生成高质量的图像、改进低质量的图像、图像风格转换等。但其中也存在一些缺陷，例如“纳什均衡”不稳定、模式崩溃问题以及计算资源需求高等问题。

矩阵填充处理图像类数据的算法繁多，不仅每种算法都有优缺点，针对的重点也比较局限。故本文在对抗自编码器的基础上进行改进，构建新的矩阵填充模型，可进行图像修复。

1.4.2 创新点

本文从生活中实际问题出发，构建一种更高效、输出结果更优质的矩阵填充模型，将模型应用于提高修复优质图像方向，并将模型经过测试至可应用于实践。

1、在原有对抗自编码器的基础上，将注意力机制融入其中，获得更强的特征提取能力。通过引入注意力机制，就不再要求原有的编码器结构将输入信息都编码进一个固定长度的向量中，而是编码器需要将输入编码成一个向量的序列，在解码的时候，每一步都会选择性的从向量序列中挑选一个子集进行进一步处理。这样，在产生每一个输出的时候，都能够做到充分利用输入序列携带的信息。那么这就提高了特征提取的能力。在提升特征提取能力更精准的同时保证信息的全面表达。

2、融合 HR-Net，使模型能够处理高分辨率图像数据，融合成功的模型可用于处理海量图像数据。HRNet 的设计核心是在整个网络结构中维持高分辨率的特征图，并在保持高分辨率的同时，通过融合不同分辨率的信息来增强最终的特征表示，达到多尺度融合的作用。HRNet 的模块化设计使得其易于与其他模型或模块结合，扩展性强。这意味着可以根据具体任务的需求，灵活地调整 HRNet 的结构和参数，以进一步优化分辨率和性能，这也为 HRNet 再其他领域的应用奠定了基础。本文将构建好的模型进行拓展应用，图像类数据均可进行测试并投入使用，图像增强、图像去噪以及视频修复等诸多领域。

1.5 文章组织结构

本文在对抗自编码器的基础模型中，引入深度学习网络中的高分辨率网络（High-Resolution Net，HRNet）和注意力机制，构建全新的矩阵填充模型，并将其应用于图像修复等相关问题。本文所做的工作如下：

第一章，引言。介绍了论文的研究背景和意义，简要论述了矩阵填充问题及创新点，并对矩阵填充研究现状进行了综述。

第二章，理论基础。介绍完成本论文所需的一些理论知识，其中包含矩阵填充、自监督学习、深度图、对抗自编码器和注意力机制，为后续的网络构建和实验研究提供了基本的方法说明。

第三章，AH-AAE 模型构建。详细介绍了模型的构建方法，并深入分析了各模块的构成、设计思路、推导分析和实现方法。在此基础上，进一步探讨了 AH-AAE 的损失函数，并进一步完善了模型的具体细节。

第四章，实验验证。首先解释了本论文中使用的数据集、指标要求和实验环

境。在此基础上，进行多种重构损失函数对比试验、多种模型对比实验、噪声抑制实验、环境感知实验以及消融实验，全面衡量各模型模块的有效性和模型的综合性能，并对实验结果逐个进行分析。

第五章，总结与展望。本章总结了本论文所开展的工作，指出了本论文为解决的问题，并提出了未来的研究方向。

2 理论基础

本章将对矩阵填充经典算法、自监督学习、深度图、对抗自编码器以及注意力机制的相关知识进行介绍。在 2.4 与 2.5 节中，着重介绍对抗自编码器的优劣势以及注意力机制引入的原因，同时在 2.1 中给出矩阵填充经典算法中的计算模型，可作为后续章节噪声实验中的对比算法模型。

2.1 矩阵填充经典算法

矩阵填充是在多个样本点缺失的情况下，利用已有的样本信息对原始数据矩阵进行精确还原的过程。当 Netflix 公司提出在稀疏数据的基础上确定用户最喜欢的电影时，矩阵填充法开始崭露头角。此后，许多研究人员在各个领域引入了矩阵填充方法，并受到越来越多的关注。标准矩阵填充问题是一个秩最小化优化模型，低秩条件约束下的目标函数为：

$$\begin{aligned} &\min_{X\in\mathbb{R}^{m\times n}} \ Rank(\boldsymbol{X}) \\ &\text{s.t. } P_{\Omega}(\boldsymbol{X}) = P_{\Omega}(\boldsymbol{M}) \end{aligned} \tag{1}$$

其中，$\boldsymbol{M}\in\mathbb{R}^{m\times n}$ 表示待补全的 m 行 n 列数据矩阵，$Rank(\boldsymbol{X})$ 表示矩阵 $\boldsymbol{X}$ 的秩，Ω 表示矩阵 $\boldsymbol{X}$ 中已知元素位置的集合，$P_{\Omega}:\mathbb{R}^{m\times n}\to\mathbb{R}^{m\times n}$ 表示矩阵 X 向矩阵子空间的正交投影。$P_{\Omega}(\boldsymbol{M})$ 定义为式(2)的形式：

$$P_{\Omega}(\boldsymbol{M}) = \begin{cases} \boldsymbol{M}_{ij}, (i,j)\in\Omega \\ 0, \quad \text{其他} \end{cases} \tag{2}$$

式（1）是一类特殊的仿射矩阵秩极小化问题，这里的秩是是非凸、非光滑且间断的，是一类 NP-hard 非凸优化问题。目前，现有的矩阵填充算法大多基于低秩性约束，包括三种类型的填充模型，即基于核范数放松的矩阵填充模型，基于矩阵分解的矩阵填充模型，基于非凸松弛的矩阵填充模型。从研究方法区分主要有两种：核范数最小化与矩阵分解。

2010 年，Candès 等人(Candes et al., 2012) 根据矩阵核范数是单位球内秩函数的最佳凸逼近(Candes and Recht, 2012)这一结论对式(1)进行调整，用核范数取代了秩范数，并提出基于核范数松弛的矩阵填充模型。该模型的优点在于它是一

种凸优化模型，存在全局最优解，核范数的邻近算子具有闭式解析，但该模型的解释涉及复杂的奇异值分解，求解效率有限，核范数不能近似地对目标矩阵的真实秩进行排序。自 Candès 之后，许多研究人员针对该模型展开研究并开发出更多优秀的求解算法，其中包括奇异值阈值算法（Singular Value Thresholding , SVT）、加速邻近梯度算法（Accelarated Proximal Gradient , APG）以及压缩感知重构算法（Fixed Point Continuation , FPC）。但因核范式的松弛方法存在求解效率和可扩展性有限的缺点，利用非凸函数松弛与矩阵分解方法进行的改进，提出了新的矩阵填充模型。

基于非凸函数松弛的建模方法是基于核范式数松弛建模方法的另一种选择。研究者认为，尽管核范数是单位球内秩函数的最佳凸近似，但两者之间仍存在很大差异。用一个非凸函数来逼近矩阵的秩函数，可以避免核范数估计矩阵秩时出现的估计偏差。Nie 等人于 2015 年提出在模型中引入 Schatten p-范数，作为秩函数的替代建模方法，并发现 Schatten p-范数松弛秩函数比核范数松弛秩函数得到的结果更准确(Nie et al., 2015)。Gu 首先提出了加权核范数近似矩阵秩函数，同时针对加权核范式必须人工设置权重的问题进行了改进，并在加权核范数的基础上构建加权 Schatten p-范数，有效解决了权重的问题(Gu et al., 2017; Xie et al., 2016)。凸松弛往往导致稀疏建模结果不准确，敏感的正则化参数会导致结果不稳定，Wen 等人提出了基于截断 Schatten p-范数的矩阵完备性模型和低秩稀疏分解模型，将 Schatten p-范数和截断核范数相结合，使模型更加灵活。应用也更加广泛。

矩阵分解是是求解低秩矩阵填充的另一种方法，是指一个秩为 r 的矩阵 $\boldsymbol{M} \in \mathbb{R}^{m \times n}$ 分解为两个较小的矩阵乘积形式，如: $\boldsymbol{M} = \boldsymbol{M}_L \boldsymbol{M}_R$。其中，$\boldsymbol{M}_L \in \mathbb{R}^{m \times r}$，$\boldsymbol{M}_R \in \mathbb{R}^{r \times n}$。分解出的矩阵数目由原始矩阵的秩决定，矩阵分解可有效降低模型求解过程中因奇异值分解而产生的计算成本，加速算法迭代执行速度。

基于矩阵分解的矩阵填充算法还包含：交替下降法、OptSpace、低秩矩阵拟合算法（Low-rank Matrix Fitting , LMaFit）和奇异值投影算法（Singular Value Projection , SVP）。采用矩阵分解法构建的填充模型在计算效率方面较好，可以避免对矩阵进行复杂的奇异值分解，但分解过程需提前估计目标矩阵的秩信息，同时模型存在非全局最优驻点解的可能性对填充效果有一定影响。

目前常用的经典填充方法有 SVT、SVP、$S_P - l_p$（*Schatten p* 范数和 l_p 范数）、TNNR-ADMM（截断核范数正则化），各算法的基本介绍如下：

（1）奇异值阈值算法（Singular Value Thresholding , SVT）是一种无监督的机器学习算法，用于对给定的输入(数据集)进行分类。用户输入训练数据，然后使用奇异值阈值算法来产生一个分类模型，称为阙值模型。通过阈值模型，用户可以有效地区分不同的分类，并为未知的分类结构提供有效的分类预测。

SVT 的基本原理是根据样本的特征和类别之间的统计学熵差异，找到最佳的分类模型。重点是计算各种统计熵，并将其优选为某些启发式算法，使其能够最大程度地反映类别之间的差异。熵信息用于评估不同的预测模型，并根据所选的最佳模型确定每个类别的临界值。

SVT 算法的主要思想是利用奇异值分解，选取一个合适的阈值，将矩阵的奇异值向零压缩，生成新的矩阵，再进行多次迭代，最终实现对原矩阵的无奇异性的有效处理。该算法利用式(3)来近似下式(1)中的矩阵核范数最小化问题

$$\begin{aligned} &\frac{\min}{\boldsymbol{X}} \| \boldsymbol{X} \|_* \\ &s.t. \boldsymbol{X}_{ij} = \boldsymbol{M}_{ij}, (i,j) \in \Omega \end{aligned} \tag{3}$$

$$\begin{aligned} &\frac{\min}{\boldsymbol{X}} \tau \| \boldsymbol{X} \|_* + \frac{1}{2} \| \boldsymbol{X} \|_F^2 \\ &s.t. \boldsymbol{X}_{ij} = \boldsymbol{M}_{ij}, (i,j) \in \Omega \end{aligned} \tag{4}$$

当 $\tau \to \infty$ 时，式(4)收敛于式(3)，该问题的迭代序列变成下式(5)的形式：

$$\begin{cases} \boldsymbol{X}^k \leftarrow D_\tau(\boldsymbol{Y}^{k-1}) \\ \boldsymbol{Y}^k \leftarrow \boldsymbol{Y}^{k-1} + \delta_k P_\Omega(\boldsymbol{M} - \boldsymbol{X}^k) \end{cases} \tag{5}$$

其中，$\{\delta_k\}_{k \geq 1}$ 为步长，D_τ 为奇异值收缩算子，SVT 算法的迭代步骤有两个关键的性质：稀疏性以及低秩性。

（2）奇异值投影算法（Singular Value Projection , SVP）是一种使用梯度投影法解决 ARMP 问题的投影梯度下降算法。如果矩阵中的元素不相关且分布均匀，该方法就能精确还原整个矩阵，并接近最佳样本数。在 SVP 算法中，引入了一个正交投影算子 P_r，其中 $P_r : \mathbb{R}^{m \times n} \to \mathbb{R}^{m \times n}$ 是向空间 $C(r) = \{\boldsymbol{X} : Rank(\boldsymbol{X}) \leq r\}$ 的正交投影，则：$P_r(\boldsymbol{X}) = \underset{\boldsymbol{Y}}{arg\,min} \| \boldsymbol{Y} - \boldsymbol{X} \|_F$，$s.t. \boldsymbol{Y} \in C(r)$。低秩矩阵的最佳近似

值是与其主成分分析相对应的前 r 个主成分，通过 SVD 方法求解，即找到 X 矩阵的前 r 个奇异值和奇异向量。这一点与前文提到的核范数最小化算法中的奇异值值算子 D_τ 类似。D_τ 算子使用的是软截断法（Soft-thresholding），即通过设置阈值将小的奇异值降至 0 来获得低秩解法。P_r 算子采用硬截断（Hard-thresholding）的方法，即直接截断原始变量的前 r 个奇异值。在每次迭代中，SVP 应用低秩正交投影算子 P_r 将矩阵映射到低秩空间，然后在以下迭代步骤中用投影梯度更新其值，如下式(6):

$$\begin{cases} \boldsymbol{X}^{k+1} \leftarrow P_\tau(\boldsymbol{Y}^{k+1}) \\ \boldsymbol{Y}^{k+1} \leftarrow \boldsymbol{Y}^k - \eta_k P_\Omega(\boldsymbol{X}^k - \boldsymbol{M}) \end{cases} \tag{6}$$

其中，η_k 为步长。另外，为了改善 SVP 算法的收敛性，还有了一种牛顿法 (Newton)，但在矩阵的秩较大时，其时间复杂度较高，计算量很大。后来，一些学者给出了 Stagewise-SVP 算法，并在 l_∞ 范数下证明了该算法的收敛性。

（3）$S_P - l_p$ 是指 *Schatten p* 范数和 l_p 范数联合最小化算法，$S_P - l_p$ 是一种用于矩阵鲁棒填充的特殊算法。联合 *Schatten p* 范数和 l_p 范数最小化的目标是找到与被破坏矩阵的观测值相匹配的低秩矩阵。该算法将 *Schatten p* 范数和 l_p 范数正则化项相结合，来达到稳健的目的。*Schatten p* 范数是度量矩阵奇异值之和的矩阵核范数的一种推广。这种方法可以得到低秩的解集，从而更好地理解矩阵的内部结构，而 l_p 范数的正则化项则有效地改善了矩阵的稀疏性。它允许许多矩阵条目精确为 0，从而更容易处理缺失或破损的条目。这种算法保持了矩阵的低秩结构，并通过将 *Schatten p* 范数和 l_p 范数联合极小化，有效地对矩阵中缺失或破损的条目进行复原。

（4）TNNR-ADMM（截断核范数正则化）是一种将 TNNR 算法与 ADMM 算法相结合的改进算法。其基本思想是将张量重建问题简化为一类非凸优化问题，并使用交替方向乘法（ADMM）来解决这些问题。ADMM 是一种比较流行的优化方法，尤其适用于具有线性约束的凸优化问题。该方法将原始问题分解为一系列子问题，通过迭代求解这些子问题来逐步优化原始问题。结合这两种方法，在

TNNR 算法的基础上引入了 ADMM 算法的思想，通过迭代求解每个迭代点的子问题来优化原始问题。在适当的情况下，TNNR-ADMM 可实现全局收敛，并且收敛速度很快。

2.2 自监督学习

自监督学习是无监督学习的一种，它主要利用辅助任务从大量无监督数据中提取自身的监督信息，然后用于训练网络并从中提取有用的特征。监督学习需要大量标注的有标签数据，且强化学习需要与环境进行多次交互。而自监督学习的特点是无需对数据打标签，其主要优势是可以以极低的扩展成本轻松扩展到大型数据集。Tonioni 等人认为，很难从每个目标领域获得足够数量的样本来进行高效训练，这限制了它在许多实际应用中的实用性；而无监督学习可执行深度网络和持续在线适应，保证了算法在任何情况下的准确性(Tonioni et al., 2019)。

自监督学习被认为是机器学习的 "理想状态"，其重点在于如何从包含自监督的信号中提取辨别特征。在语义分割、物体识别、图像分类和人类行为识别等方面具有重要的理论意义和很高的应用价值。同时，它也逐渐扩展到不同的学习场景，如领域适应、小样本或无样本学习、分布式识别、生成对抗网络、卷积图网络等。Yuan 在深度相关框架下提出了一种有效的基于自监督学习的跟踪器，经实验结果表明在标准评估基准上，与最先进的监督和无监督跟踪方法相比，所提出的自监督深度相关跟踪器获得了有竞争力的跟踪性能(Yuan et al., 2020)。自监督学习在多领域的发展对人们生活发挥着至关重要的作用。Buzau 和 Xu 利用混合深度神经网络和多传感器特征融合提出了新的集成模型，可用于检测智能电表中异常以及欺诈的特性(Buzau et al., 2019; Xu et al., 2020)。 Zhang 通过深度学习与自监督学习结合提高垃圾分类的准确性，使终端显示的垃圾分类提供了一种可能(Zhang et al., 2021)。Chang 与 Jung 在样本不足的情况下，结合自监督学习，分别针对工业物联网智能故障定量识别与结构健康监测系统进行重新建模，提出了新的深度双重强化学习模型，有效实现了自我监控(Chang et al., 2022)。

2.3 深度图

深度图是三维场景中每个像素点的距离信息。它用来表示图像中每个像素点

相对于摄像机的距离，以及像素点之间的距离信息。深度图在立体视觉、三维重建和现实融合等计算机视觉任务中非常有用。深度图通常有两种类型：灰度图和彩色图，较远的物体通常用较暗的像素值表示，较近的物体显示为较亮的像素值。深度图可用于从图像中提取场景的三维信息，在建模、距离测量等方面有广泛的应用。

深度图是利用深度学习中卷积神经网络（CNN）学习的高级特征表征。这种方法是对图像等输入数据进行多重卷积和池化操作融合后创建的二维矩阵，能更好地表示抽象的图像特征。深度图是一种直观的图像处理方法，具有强大的可视化特征。该方法能有效处理不同比例和类别的图像，从而实现更准确的分类和标记。在目标检测中，深度图可以帮助定位图像中的目标区域，确定物体的位置，并创建更精确的边界。在图像分割中，对图像进行深度分析并细分为有意义的区域，从而更好地理解和处理图像内容。

在图像处理过程中，可以通过详细分析相关模块来验证深度图在图像处理中的性能。由于深度图的特性与网络的结构和参数密切相关，可以利用深度图来研究不同的网络结构和运算方法对图像特征提取效果的影响。其次，进行深度图实验有助于研究人员更好地理解深度图的特性，从而进一步优化网络结构，提高图像处理效率。

深度图可以提供目标在场景中的位置信息，对障碍物检测、距离测量等非常有用。该方法可以获得目标的位置、形状、大小等信息，从而完成对目标的精确检测、分割和分类，还可以估计目标的距离和位置，因此进行深度图的实验可以验证该模型推导空间位置的能力。

2.4 对抗自编码器

对抗式自编码器（Adversarial Auto-encoder，AAE）是一种生成式模型，由两部分组成，分别是生成器和判别器，该算法由 Makhzan 在 2015 年提出的，利用自编码器隐含码矢量的聚集后验，来实现变分推理(Makhzani et al., 2015)。与变分自编码器相比，AAE 具有可调整的分子特征提取、无需监督预训练，以及较高的计算效率(Kadurin et al., 2017)等优势。

编码器把图像等输入资料转化成潜在空间中的表达形式，也叫编码。解码器

则将潜在表示还原为原始输入数据的重建，旨在保留原始数据的特征和结构。在该算法中，除了编码器解码器，还有一种判别器。判别器的主要任务是判定重构出的样本值是否来自原始数据分布。生成器的目的是通过学习对抗损失，欺骗判别器，使重构的样本更贴近真实数据分布。通过对编码器和解码器进行对抗训练，可以有效地提取出更具体、更精确的隐藏表示，从而使重建样本值更接近真实分布。

AAE 是将自编码程序转换为生成模型的常用方法。该算法基于重构错误准则和对抗准则的自编码训练，在自编码过程中引入新的误差准则，使自编码器隐含表达式的聚合后验概率分布与任意概率分布匹配。可用于数据生成、图像修复和特征学习等任务。它的优势在于，通过引入对抗训练，可以提高重建样本的质量，增加对数据分布的理解，生成更丰富、更真实的数据样本。

2.4.1 自动编码器

自动编码器（Auto-Encoder, AE）主要用于对数据进行有效表达，并对数据进行压缩与解压。Mei 等提出一种基于渐进式特征融合的 U-Net 编码器-解码器深层网络，并将深层神经网络机理视为“黑箱”，从而产生具有适应性和可训练性的端到端模型(Mei et al., 2019)。Chen 等针对无标签图像，提出了一种基于上下文修复的自主学习方法(Chen et al., 2020)。这些研究对深度学习网络中的自编码器模型结构进行了初步设计，使其能够应用于简单图像处理任务，但在复杂的图像处理领域仍然有很大的局限性。

因此，这类算法的核心难点在于如何求解其高复杂性。Deng 在这方面做了一些初步的探索，并在此基础上提出了一种基于完全参数化梯度下降的深层卷积神经网络，在降低模型复杂性的前提下，有效地解决多模态图像的复原与融合问题(Deng et al., 2020)。目前，虽然研究表明使用自动编码器的网络结构可以简化模型，但在映射的过程中，会丢失一些细节和高频信息，导致图像重构的质量下降。如果在训练过程中出现噪声或损坏特别严重的部分，其内部噪声会被自动编码器吸收，进而影响图像质量。

2.4.2 生成对抗网络

生成对抗网络（Generative Adversarial Networks，GAN）是无监督学习的神经网络，训练一个生成的对抗性网络需要对一个困难的程序进行详细的优化(Li et al., 2015)。 Springenberg 等人在没有标签或者部分标签的情况下，提出了一个判别分类器。该方法是基于一个目标函数，权衡观察到的例子和他们的预测分类分布之间的互信息对生成模型的分类器进行综合评价(Springenberg, 2015)。

2017 年，Choi 等人首次报道了一种基于 medGAN 的算法，用于生成与医学统计领域相关的人工人类病例数据(Choi et al., 2017)。在此基础上，Pathak 等人提出了一种基于自编码和对抗训练网络的图像重建算法，利用生成网络学习图像的空间分布规律，重建出更真实的图像，并在融合全局上下文信息的同时有效保留图像的结构完整性和连续性(Pathak et al., 2016) ，但该方法受限于有限的图像分辨率。为解决这一难题，Zhang 等人提出了堆叠生成对抗网络 (StackGANs) ，以获得高精度、高分辨率的逼真图像。其次，他们开发了一种新的多级生成对抗网络（StackGAN-V2），并将其应用于特定场景(Zhang, H. et al., 2018)。

GAN 可以通过调整生成和判别网络的结构和参数，以及选择训练样本和预处理训练模型来控制修复结果的质量和形状。然而，GAN 的学习算法通常非常复杂，对设备的要求较大。在修复图像时，噪声或损坏的影响可能会转移到修复后的图像上，这往往会导致一些伪影和失真。此外，GAN 虽然具有很强的提取细节的能力，但其性能仍然受到很多因素的限制，如网络结构、训练数据的质量和数量以及训练过程的稳定性等。

2.5 注意力机制

近年来，注意力机制被广泛应用于图像分类、物体识别、语义分割、视频理解、三维视觉、多通道任务和自主学习等领域(Niu et al., 2021)。注意力机制可以模拟人类的注意力机制，从而提高模型性能，因此在机器学习、深度学习等领域有着重要的应用。2004 年，Lavie (Lavie et al., 2004) 等人针对“早选择”和“晚选择”的争论提出了改进的对策，并证明了选择控制在认知调节中的作用。Xu (Xu et al., 2015) 《Show, attend and tell: Neural image caption generation with visual

attention》可以说是关于注意力机制的第一部著作，它引入了软注意力（soft attention）和硬注意力（hard attention）。在每个时间点，给定图像的每个区域都分配了注意力权重，从而提高了语句与特定模块的相关性。

注意力机制是一个逐步发展的过程，它们都是在以往研究的基础上不断完善和扩展。为了降低计算量，人们最初提出并改进了简化注意力机制(Shazeer et al., 2017)。在此基础上，研究了可变注意力机制，提高了算法的计算速度和稳定性，并融合了多重注意的多目标信息，获得了更丰富的特征。Bello (Bello et al., 2019) 在卷积神经网络（Convolutional Neural Network ,CNN）的基础上，引入了一种新的自注意力学习方法，并利用视觉分辨任务替代卷积，将自注意力引入识别不同视觉信息的新学习方法。同时，Vig (Vig, 2019) 开发了一款开源软件，可以实现多层次注意力的可视化，并在此基础上进一步探索注意力机制。Fukui (Fukui et al., 2019) 提出了基于注意力机制的视觉解释模型，该模型具有很高的响应值。注意力机制可以提高图像的细节质量，实现更精确的局部检索；通过调整权重，可以实现检索时模型重点区域的视觉呈现，并通过调整注意力机制参数和各种注意模式，提高模型的可解释性和可控性。

2.6 本章小结

本章主要介绍了一些与矩阵填充相关的理论知识和算法，其中涉及后面章节进行实验对比的方法。也介绍了本文网络搭建以及后续实验所需的基础知识，包括自监督学习、深度图、对抗自编码器以及注意力机制的相关内容。

3 AH-AAE 模型构建

基于前两章的实际问题背景与理论基础，本章将介绍具体的创新点与模型构建方法，对抗自编码器相比原始 GAN 和 AE 而言，具有了很多让生成结果更可控的特性，但很难把 AAE 扩展到高分辨率图像数据上，同时产生的细节也可能会过度随机导致边缘模糊。故本章引入 HRNet 以及注意力机制进行建模以重点解决的问题包括：1、针对图像特征提取的问题；2、针对提升图像分辨率的问题；3、针对边界的恢复问题。以下为具体建模过程。

3.1 AH-AAE 网络结构

本文模型是基于传统 AAE 引入注意力机制以及 HRNet 进行改进的网络，通过引入注意力机制可有效提升模型填充效率以及准确度，以下将基于注意力机制与 HRNet 的矩阵填充模型简称为 AH-AAE，模型构建如下图 3.1 所示。

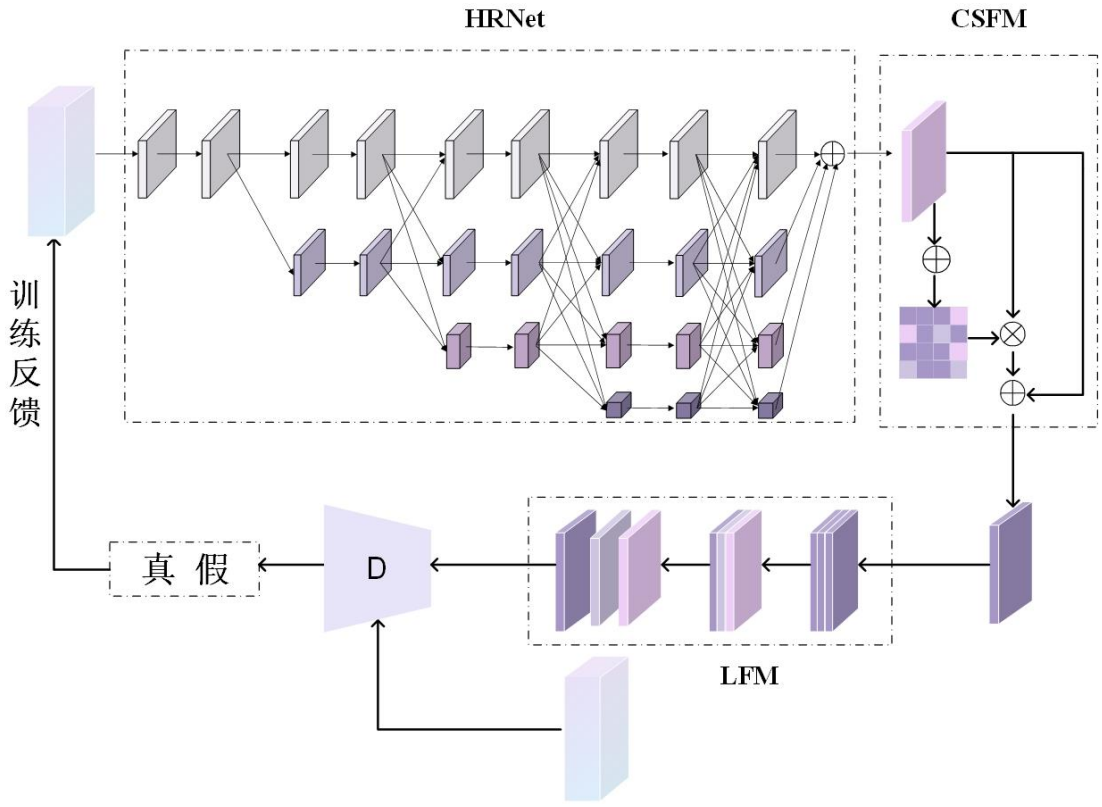


图 3.1 AH-AAE 网络构架图

从图 3.1 的网络结构图中可以看出，编码器部分使用预先训练好的 HRNet 进行特征提取。创建一个高分辨率子网络，再添加多个分辨率子网络，将多个子网络依序相加形成更多阶段。多个高分辨率子网络并行连接，在处理过程中反复进行信息交互，以实现多个特征的融合。经过编码器处理后，引入注意力机制，具体分为通道注意力以及空间注意力进行顺序操作。分辨率最高的输出特征图被输入通道相似性融合模块（Channel Similarity Fusion Module , CSFM）进行强化学习，用于提高通道之间的语义关联，然后由位置融合模型（Location Fusion Model ,LFM）将掩码的位置信息和周围相关联的信息融合在一起。使用位置融合模型进行解码恢复掩码区域，以提高待修复位置预测的准确性。最后，将其与原始数据一起送入判别器进行验证，直到掩码位置信息得到充分表达。通过这种方式，重要的特征会得到改进，少量的特征则会被抑制。

3.2 HRNet 网络

高分辨率网络（High-Resolution Net，HRNet）是由中科大和微软研究院联合开发的人体姿态识别网络模型。与特征提取中常用的下采样和上采样方法不同，HRNet 能够在整个计算过程中保持高分辨率表征。如下图 3.2 所示，从一组高分辨率卷积开始，逐步添加低分辨率卷积并将其并行连接起来。HRNet 由多个阶段组成，其中第 n 个阶段包含 n 个卷积分支和 n 种不同的分辨率。在此过程中，不同的并行操作组合之间通过多分辨率融合不断交换信息。通过这种方法获得的高分辨率表征不仅具有完整的语义特征，还具有很高的空间精度。首先，由于多分辨率的分支是并行而非串行的，因此整体高分辨率特征保持不变。其次，传统方法将高阶低分辨率特征与高阶放大的低分辨率特征融合在一起，而 HRNet 始终保持高阶和低分辨率特征，不断将它们融合在一起并相互促进。

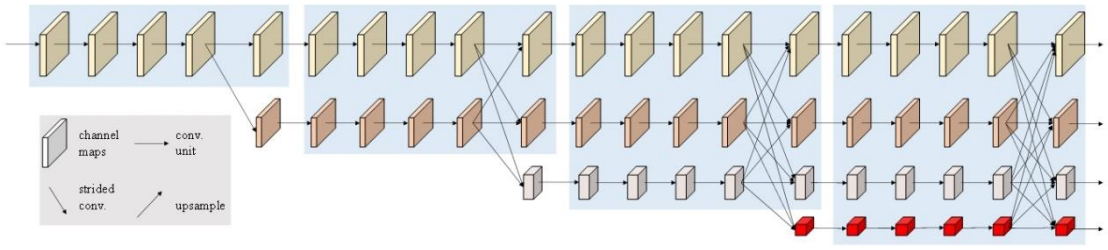


图 3.2 HRNet 网络结构图

HRNet 网络的结构如图 3.2 所示。网络的主体部分由四个阶段和四个并行卷积分支组成。分辨率分别为 1/4、1/8、1/16 和 1/32。第一阶段包含四个宽度为 64 的残差单元，每个单元后都有一个 3 × 3 的卷积，将特征图的数量变为 C。第二、第三和第四级分别包含 1、4 和 3 个模块。每个模块包含 4 个残差单元。每个单元对每个分辨率进行两次 3×3 的卷积，然后通过 ReLU 函数的非线性激活进行 BatchNorm 处理。四个分辨率下的卷积通道数依次为 C、2C、4C 和 8C，每个阶段的末尾都有一个多分辨率融合模块。

如图 3.3 展示的 HRNet 模块图可观察到，每个模块可以分为两部分：(a) 多分辨率并行卷积，(b)多分辨率融合。

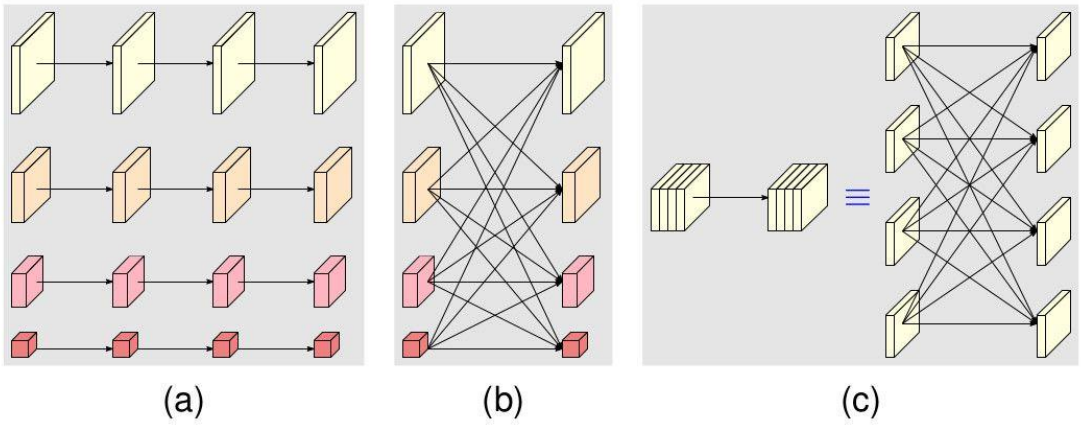


图 3.3 HRNet 并行模块图

多分辨率并行卷积，类似群卷积（group convolution）。每个分辨率的卷积操作相互独立，保持相同的分辨率。多分辨率卷积使用传统多分支卷积的全连接模式（图 3.3（c））。传统卷积可以分解为多个较小的卷积。输入通道被分为若干子集，而输出通道也同样被分为若干子集。输入子集与输出子集紧密相连，每个连接都使用传统的卷积运算。每个输出通道的子集是每个输入通道子集的卷积输出之和。唯一不同的是，HRNet 必须考虑分辨率的差异问题。

HRNet 是在低分辨率的帮助下，多次融合高分辨率，从图 3.2 及图 3.3 中可以看出 HRNet 网络的 3 个关键特点：1）以并行而非串行的方式连接不同分辨率的分支；2）整个运算过程都保留了高分辨率表征； 高低分辨率图之间不断地交换信息；3）不断地融合不同分辨率的表征，得到对位置敏感的高分辨率表征。

3.3 融合注意力机制的特征模块

3.3.1 稀疏注意力与密集注意力组合

在传统的注意力机制中，注意权重的计算往往需要被注意对象的参与，而在编码过程中，权重的计算既要考虑编码过程的隐含状态，也要考虑解码过程的隐含状态。自注意力机制不是输入指令和输出指令之间的注意力机制，而是发生在输入指令内部元素之间或输出指令内部元素之间的注意力机制。例如，在 transformer 中计算权重参数时，文本向量被转换成相应的 KQV，只对序列进行相应的矩阵运算，而不使用目标序列的信息。神经网络接收到的输入包含许多大小不同的向量，不同向量之间存在一定的关系，但实际训练中无法充分利用这些输入之间的关系，导致模型训练效果不佳。例如，机器翻译问题（序列-序列问题，机器决定多少个标签）、词性标注问题（一个向量对应一个标签）、语义分析问题（多个向量对应一个标签）以及其他文本处理问题。

全连接神经网络无法对多个相关输入建立相关性，这一问题可通过自注意力机制来解决，自注意力机制试图让机器识别整个输入不同部分之间的相关性。但在实践中存在两个问题：第一，算法的复杂性（时间、计算量）是输入数据序列长度的二次方。第二，先验知识的结构性。自注意力机制并不预设输入的结构偏差。换句话说，transformer 必须从训练数据中学习数据序列的信息。因此，transformer 很容易在中小型数据上过拟合。在此基础上，从稀疏自注意、线性自注意、原型和存储压缩以及低阶自注意等方面对自注意力机制可进行优化。

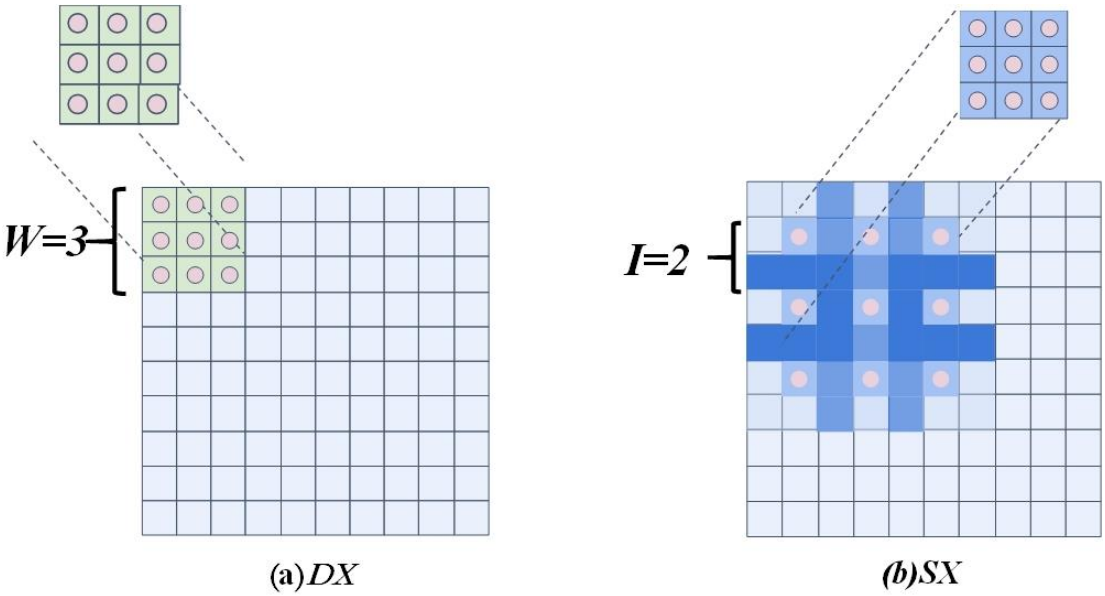


图 3.4 自注意力机制

自注意力机制最大的问题就是高昂的计算成本，用多头注意力机制进行计算有：$\Omega(\text{MSA})=4hwC^2+2(hw)^2C$，其中，以 P 为步长进行投影，$h=\frac{H}{P}$、$w=\frac{D}{P}$ 为尺寸 $H\times D$ 的投影特征图 $\boldsymbol{X}$（$\boldsymbol{X}\in\mathbb{R}^{h\times w\times C}$）。而密集注意力如图 3.4（a）所示，密集注意力可使每个单元可以从一个不重叠的 $W\times W$ 窗口的邻域位置与较少数量的单元进行交互。将单元进行分组，每组均有 $W\times W$ 个单元，再将其用于计算 $\frac{h}{W}\times\frac{w}{W}$ 次的自注意力模块，生成的密集注意力 $\boldsymbol{DX}$ 的计算成本为下式(7)：

$$\Omega(\boldsymbol{DX})=(4W^2C^2+2W^4C)\times\frac{h}{W}\times\frac{w}{W}=4hwC^2+2W^2hwC \tag{7}$$

稀疏注意力如图 3.4（b）所示，允许每个单元与其余数量的单元产生交互，这些单元来自间隔大小为 I 的稀疏位置。之后，所有单元的更新也被分成若干组，每个组有 $\frac{h}{I}\times\frac{w}{I}$ 个单元。进一步利用这些组计算 $I\times I$ 次的自注意力，稀疏注意力 $\boldsymbol{SX}$ 的计算成本为式(8)：

$$\Omega(\boldsymbol{SX})=(4\frac{h}{I}\times\frac{w}{I}C^2+2(\frac{h}{I}\times\frac{w}{I})^2C)\times I\times I=4hwC^2+2\frac{h}{I}\frac{w}{I}hwC \tag{8}$$

将密集注意力与稀疏注意力如图 3.5 所示进行结合来提取深度特征。

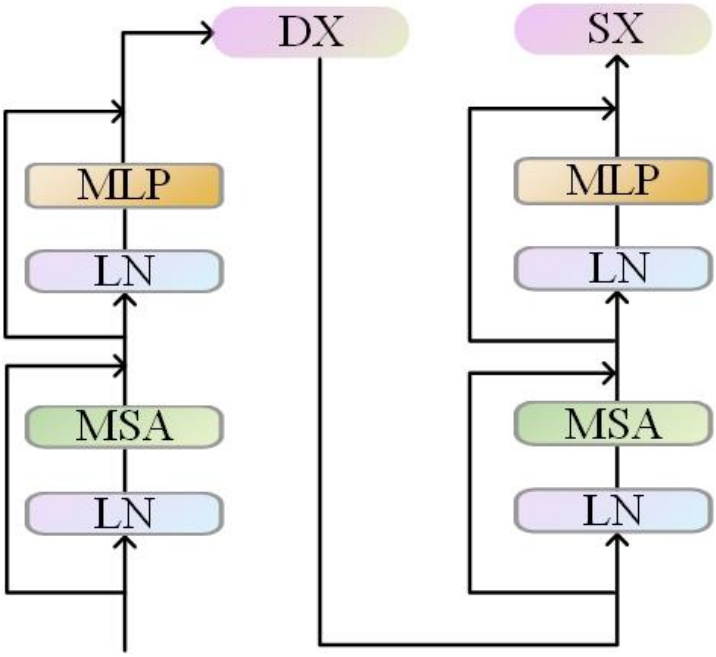


图 3.5 密集注意力与稀疏注意力组合

输入特征 x 经过层归一化(layer normalization, LN)和多头自注意力(multi-head self-attention, MSA)。输出 x' 被送入多层感知(multi-layer perception , MLP)。dx 是经密集注意力的最终输出。该过程可描述为式(9)：

$$\begin{aligned} x' &= \mathrm{MSA}(\mathrm{LN}(x)) + x \\ dx &= \mathrm{MLP}(\mathrm{LN}(x')) + x' \end{aligned} \tag{9}$$

获得密集注意力矩阵后 $\boldsymbol{DX}$ ，再经同样操作获得稀疏注意力矩阵 $\boldsymbol{SX}$ 。

3.3.2 通道相似性融合模块

在深度估算中，每个特征图都可视为一个区域的响应，响应区域之间存在一定的相关性。在此基础上，充分利用各通道之间的相关性，创建出更完整的响应区域。在处理 RGB 数据时，非常重要的一点是，每个通道的信息都能有机地整合在一起。

通道注意力主要检查输入数据的通道信息，并计算每个通道的注意力权重。根据每个通道的特征，对这些通道进行加权，并突出重要的通道。传统的通道注意力方法不是将通道作为一个整体进行平均，而是学习每个通道的权重，以达到对通道特征响应增强或抑制的目的。该方法采用非线性运算，旨在提高模型性能

的同时不牺牲通道权重。然而，传统的通道注意力方法未能很好地解决这一问题。

结合通道注意力的特点，并在关于稀疏注意力与密集注意力(Zhang et al., 2022)的启发下，，构建了如图 4 的通道相似性融合模块（Channel Similarity Fusion Module , CSFM），旨在减少噪声或冗余的特征，提高模型的准确性与泛化能力。

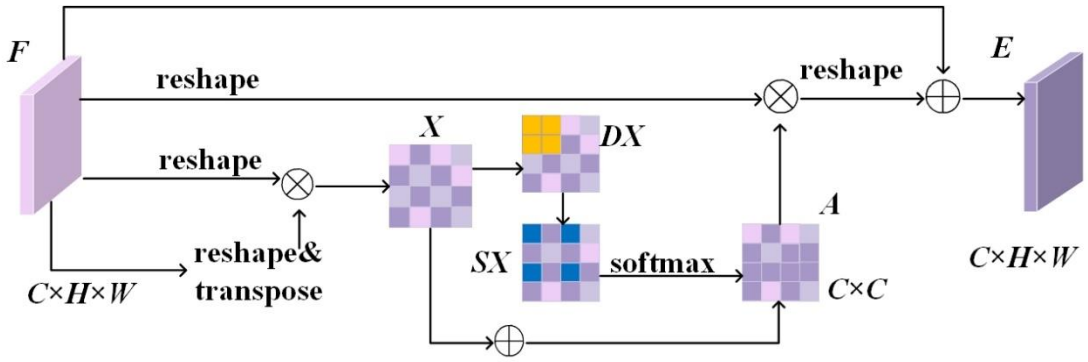


图 3.4 CSFM 网络

第一步，生成一个注意力矩阵，用来模拟任何两个通道图之间的关联。将由编码器生成的特征图 $\boldsymbol{F}$（$\boldsymbol{F} \in \mathbb{R}^{C\times H\times W}$），重塑为 $\mathbb{R}^{C\times N}$，其中 $N = H \times W$ 是像素数，再通过矩阵相乘来计算每个通道的特征相似性，得到 $\boldsymbol{S} \in R^{C\times C}$。

第二步，在传统的通道注意力中，是将两个特征图相乘再通过归一化获得通道注意力图。在这里，结合注意力深度特征提取策略(Zhang et al., 2022)，将相似度矩阵的输出为浅层特征，设计密集与稀疏两种类型。将密集特征层经过归一化和多头注意力后输入多层感知机中，再与其稀疏特征映射进行交互，从而形成更深层次的特征图，最后利用元素相加获得最终特征图。在此基础上，按照常规的处理方式，对浅层特征进行规范化处理，获得含有深度特征的注意力图 $\boldsymbol{A} \in \mathbb{R}^{C\times C}$，$\boldsymbol{A}$ 中的每个元素可表示为：$a_{ij} = \dfrac{exp(\boldsymbol{a}_i \cdot \boldsymbol{a}_j)}{\sum_{j=1}^{C} exp(\boldsymbol{a}_i \cdot \boldsymbol{a}_j)}$，其中，$\boldsymbol{a}_i$、$\boldsymbol{a}_j$ 表示 $\boldsymbol{A}$ 中的向量，a_{ij} 表示第 j 通道对第 i 通道的影响。

第三步，在 $\boldsymbol{A}$ 和 $\boldsymbol{F}$ 做乘积并乘尺度系数 β，将结果重塑为 $\mathbb{R}^{C\times H\times W}$，最后将 $\boldsymbol{F}$ 与结果相加获得最终的输出 $\boldsymbol{E} \in \mathbb{R}^{C\times H\times W}$：

$$\boldsymbol{E}_i = \sum_{j=1}^{C} (\beta a_{ij} \boldsymbol{F}_j) + \boldsymbol{F}_i \tag{10}$$

由式(10)可知，每个通道的最终特征是各通道的特征与原特征的加权和，通过扩大感受野获得更多的特征信息，再利用特征图之间的相似性，构造的通道相似模块，可有效地丰富各通道间的特征关系，可减少各通道的特征损失。

3.3.3 位置融合模块

深度神经网络的解码器通过跳跃连接与编码器的特征相连，从而获得更详细的信息。然而，像这样简单的求和与连接操作过于单调，会使缺失数据的边界变得极其模糊。数据缺失部分通常有明确的边界点作为特征参考，由此引入空间注意力和位置编码，构建位置融合模块（LFM）。

在分析空间结构数据时，空间注意力是重点，其核心思想是将注意力加权与输入数据的空间位置信息相结合。在空间注意力中，位置信息指的是节点与相邻节点的连接方式。充分利用节点之间的位置信息，空间注意力可以更好地捕捉节点之间的联系，实现信息传递。而位置编码则是描述图中节点具体位置的附加向量或矩阵。位置编码有多种生成方式，如正弦编码、余弦编码等(Vaswani et al., 2017)。通过将位置编码与空间注意力相结合，节点的位置信息一并纳入考虑，并对其进行加权融合，从而使空间注意力机制能够更好地模拟节点之间的关系。

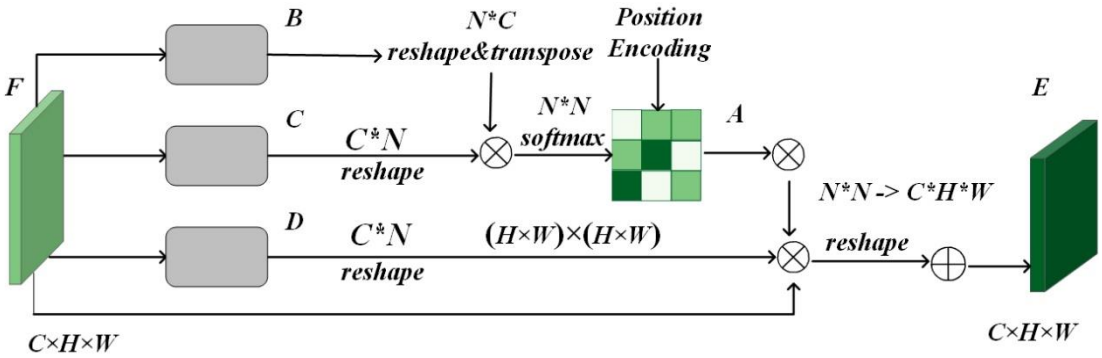


图 3.5 LFM 网络

第一步，特征图 $\boldsymbol{F}$（$\boldsymbol{F} \in \mathbb{R}^{C \times H \times W}$）通过三个卷积层得到三个子特征图 $\boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$

($\boldsymbol{B},\boldsymbol{C},\boldsymbol{D}\in\mathbb{R}^{C\times H\times W}$) 重塑为 $\mathbb{R}^{C\times N}$，分别记作 $\boldsymbol{G},\boldsymbol{P},\boldsymbol{Q},\boldsymbol{U}\in R^{C\times N}$，$C$、$H$、$W$ 分别为特征图的通道数、高度和宽度，重塑后的特征图中 $N=H\times W$ 为像素个数。利用特征图之间的相关性计算注意力权重，矩阵 $\boldsymbol{P}$ 的转置 $\boldsymbol{P}^T$ 和 $\boldsymbol{Q}$ 相乘，再通过归一化得到注意力权重矩阵 $\boldsymbol{S}$ ($\boldsymbol{S}\in\mathbb{R}^{N\times N}$)。矩阵 $\boldsymbol{S}$ 中元素 s_{ij} 计算方式为：

$$s_{ij}=f(\boldsymbol{p}_i,\boldsymbol{q}_j)=\frac{exp(\boldsymbol{p}_i^T\cdot\boldsymbol{q}_j)}{\sum_{i=1}^{N}\sum_{j=1}^{N}exp(\boldsymbol{p}_i^T\cdot\boldsymbol{q}_j)} \tag{11}$$

其中，$f(\cdot,\cdot)$ 为注意力计算函数，$\boldsymbol{p}_i^T$ 为 $\boldsymbol{P}^T$ 中第 i 行向量，$\boldsymbol{q}_j$ 为 $\boldsymbol{Q}$ 中第 j 行向量，$N=H\times W$ 为像素数，$\boldsymbol{S}$ 中的每个元素 s_{ij} 表示在空间上位置 i 与位置 j 之间的注意力权重。

第二步，引入位置编码函数 $\mathrm{PE}(i)$ 和 $\mathrm{PE}(j)$，将位置编码与注意力权重结合经编码函数计算获得融合位置的注意力矩阵 $\boldsymbol{A}$ ($\boldsymbol{A}\in\mathbb{R}^{N\times N}$)。

$$a_{ij}=s_{ij}*g(PE(i),PE(j)) \tag{12}$$

其中 $g(\cdot,\cdot)$ 表示用于融合位置编码的函数，位置编码的计算公式(Vaswani et al., 2017)：

$$\begin{aligned}PE(\mathrm{pos},2\mathrm{i})&=\sin(\mathrm{pos}/10000^{2i/d_{\mathrm{model}}})\\PE(\mathrm{pos},2\mathrm{i}+1)&=\cos(\mathrm{pos}/10000^{2i/d_{\mathrm{model}}})\end{aligned} \tag{13}$$

第三步，将 $\boldsymbol{U}$ ($\boldsymbol{U}\in\mathbb{R}^{C\times N}$) 再和 $\boldsymbol{A}$ 相乘，得到加权后的矩阵最后和 $\boldsymbol{G}$ 进行相加并将 $\mathbb{R}^{C\times N}$ 重塑为 $\mathbb{R}^{C\times H\times W}$ 得到最终输出 $\boldsymbol{E}$（$\boldsymbol{E}\in\mathbb{R}^{C\times H\times W}$），计算方式如下所示：

$$\boldsymbol{E}_j=\alpha\sum_{i=1}^{N}(a_{ij}\odot\boldsymbol{u}_i)+\boldsymbol{G}_j \tag{14}$$

其中，α 为尺度因子，$\boldsymbol{E}_j$ 为第 j 个通道的最终特征，a_{ij} 为矩阵 $\boldsymbol{A}$ 中的元素，$\boldsymbol{u}_i$ 为 $\boldsymbol{U}$ 中的第 i 行向量，两者哈达玛积运算结果表示 $\boldsymbol{u}_i$ 中所有像素按照注意力矩阵中对应位置的权重进行加权求和后的结果，$\boldsymbol{G}_j$ 为重塑后的特征图 $\boldsymbol{G}$ 中第 j 个通道。

将位置编码与空间注意力融合的空间注意力机制，能够同时考虑特征与位置信息，从而增强模型的泛化能力与鲁棒性，提高图像边界部分的修复效果。

3.4 损失函数

为了训练所提出的网络，遵循将损失函数 L_{total} 定义为以下四个损失函数之和(Dong et al., 2022; Suvorov et al., 2022)，表示为：$L_{\text{total}} = L_{\text{rec}} + L_{\text{adv}} + L_{\text{fm}} + L_{\text{hrf}}$

L_{rec} 为重构损失(Zhao et al., 2016)，L_1 损失函数能保持较好的颜色及亮度，但没有考虑人类视觉感知，可能会陷入局部最优解，以及会发生过拟合；MS_SSIM 损失函数(Zhao et al., 2016)考虑到了分辨率的问题，可以保留高频信息，但同时也会导致亮度的改变和颜色的偏差，故此处的重构损失函数采用两者结合的 MS-SSIM-L_1，获得：

$$L_{\text{rec}} = \alpha L^{\text{MS\_SSIM}} + (1-\alpha)\,\mathrm{G}_{\sigma_G^M}\cdot \mathrm{L}^{\mathrm{l}_1} = \alpha(1-\text{MS\_SSIM}(\tilde{p})) + (1-\alpha)\,\mathrm{G}_{\sigma_G^M}\cdot \mathrm{L}^{\mathrm{l}_1} \tag{15}$$

L_{adv} 作为对抗损失，由生成器损失和判别器损失组成，采用中的对抗损失表示为：

$$\begin{aligned} L_{\mathrm{D}} &= -E_{\boldsymbol{I}}[\log \mathrm{D}(\mathrm{I})] - E_{\boldsymbol{I},\mathrm{M}}[\log \mathrm{D}(\hat{\mathrm{I}}) \odot (1-\mathrm{M})] - E_{\boldsymbol{I},\mathrm{M}}[\log(1-\mathrm{D}(\boldsymbol{I})) \odot \mathrm{M}] \\ L_{\mathrm{G}} &= -E_{\boldsymbol{I}}[\log \mathrm{D}(\boldsymbol{I})] \\ L_{\text{adv}} &= L_{\mathrm{D}} + L_{\mathrm{G}} + \lambda_{\mathrm{GP}} L_{\mathrm{GP}} \end{aligned} \tag{16}$$

其中，D 为判别器；$\boldsymbol{M}$ 为图像掩码区域，$\boldsymbol{I}$ 和 $\hat{\boldsymbol{I}}$ 分别为真实图像以及预测后的图像；L_{D} 为判别器损失；L_{G} 为生成器损失；$\lambda_{\mathrm{GP}} = 0.001$，$L_{\mathrm{GP}} = E_{\boldsymbol{I}} \left\| \nabla_{\boldsymbol{I}} D(\boldsymbol{I}) \right\|^2$ 为梯度惩罚。

同时添加 L_{fm} 和 L_{hrf}。L_{fm} (Wang et al., 2018)为特征匹配损失，重在改善 GAN 损失的均衡，通过比较生成图像和真实图像的特征表示，可更好的学习到特征分布而提高图像生成质量。L_{hrf} (Suvorov et al., 2022)为感知损失，添加此项损失改良了 GAN 在训练过程中出现的模式崩溃以及梯度消失等问题，同时可引导生成器学习到更真实、更清晰的图像特征，提高图像输出质量。最终的损失函数记作：

$$L_{\text{total}} = \lambda_{L_{\text{rec}}} L_{\text{rec}} + \lambda_{L_{\text{adv}}} L_{\text{adv}} + \lambda_{L_{\text{fm}}} L_{\text{fm}} + \lambda_{L_{\text{hrf}}} L_{\text{hrf}} \tag{17}$$

3.5 本章小结

本章首先介绍了 AH-AAE 模型的网络结构，并对网络的构建方式、模型中的各个组成部分做了简要说明。在此基础上，详细讨论了以 HRNet 为中心的编码器结构和以注意力机制为中心的特征模块结构。在特征模块方面，重点讨论了稀疏和密集注意力联合方法的构建、将通道注意力机制转换为通道相似性融合模块的过程，以及合并空间注意力机制和位置编码形成的位置融合模块。最后，讨论了模型中使用的损失函数，并得到了损失函数的最终表达式。

4 实验验证

本章主要将第三章的模型进行验证，本文的模型为矩阵填充模型，实验对象为图像类数据，模型填充效果验证包括填充实验、噪声实验以及消融实验。另外设置重构函数实验以及特征模块深度实验测试本文模型在目标检测方面的效果。

4.1 数据集概述

矩阵填充是一个算法问题，涉及数字、图像、文本等多种数据类型。本文的研究对象是图像类数据。传统的图像填充方法可分为三大类：基于特征的图像填充、基于数据的图像填充和基于内容的图像填充。这些算法在很大程度上依赖于数据本身的特征，并利用这些特征来确定缺失区域的内容。然而，在现实生活中，由于光线、背景等因素的影响，物体自身的特征会发生变化，使其在颜色、纹理等方面与周围环境会有所不同。因此，如何有效填充这种差异是当前图像处理中亟待解决的关键问题。本文使用两个数据集分别对其进行测试。基本介绍如下：

（1）实验数据选自微软于 2014 年资助的 MS-COCO （Microsoft Common Objects in Context, MS-COCO）数据集，其中包含大量用于目标识别、分割、关键点识别和标注的图像。本实验使用的版本为 2017 版，其中包含：训练集 118,287 张图像、测试集 40,670 张图像。此数据集用于训练模型。本章的具体应用的实验情况为 4.4.1、4.4.2、4.4.3 和 4.4.5。

（2）KITTI 数据集是一个广泛应用于计算机视觉领域的数据集，用于研究和发展自动驾驶、场景理解等相关算法和模型。该数据集提供了多种传感器数据，包括高分辨率图像、激光雷达点云和惯性测量单元数据等，并提供了地面真实标注数据。KITTI 数据集涵盖了道路检测、目标检测与跟踪、语义分割、立体视觉、激光点云处理和车辆自我定位等多个任务领域。数据集中的数据采集环境多样，覆盖了城市、乡村和高速公路等不同场景，并考虑了各种挑战因素，如摄像机姿态变化、遮挡和低光照等。KITTI 数据集规模庞大，提供了丰富的数据和详细的标注，可用于训练、验证和评估计算机视觉模型的性能。它的存在为算法改进、性能评估和实验比较提供了有力的支持，KITTI 数据集中的数据具有代表性，广泛应用于深度网络中检验模块的特征提取能力。本文使用该数据集中的自动驾驶

数据测试拟合后模型中的特征模块作用是否达到预期。具体使用的实验情形为本章 4.4.4。

4.2 评价指标

本次实验所选用的评价指标为四项，分别为峰值信噪比（PSNR）、图像通用质量（UQI）、FID、不同阈值（thr）下准确率精度δ（$thr = 1.25, 1.25^2, 1.25^3$）。

（1） PSNR 表示破坏性噪声的功率与信号表示精度之间的比率。由于许多信号的动态范围很宽，PSNR 通常用分贝的对数来表示。在图像处理领域，图像质量已成为一项重要指标。PSNR 的表达式如下：

$$PSNR = 10 \cdot \log_{10}(\frac{MAX_I^2}{MSE}) = 20 \cdot \log_{10}(\frac{MAX_I}{\sqrt{MSE}}) \tag{18}$$

（2）UQI 是进行图像评价的常用指标之一，用于衡量图像的相似度和偏差度。UQI 是一种基于亮度、对比度和图像结构的测量方法，UQI 的表达式如下：

$$UQI = \frac{4\mu_O\mu_F\sigma_{OF}}{(\mu_O^2 + \mu_F^2)(\sigma_O^2 + \sigma_F^2)} \tag{19}$$

（3）FID 全称为 Fréchet Inception Distance，用于衡量生成的图像与真实图像的相似程度。FID 是根据图像与真实图像之间的分布差异来评估图像的一种度量，通过比较最终图像与原始图像来进行计算，FID 的表达式如下：

$$FID = \| \mu_r - \mu_g \|^2 + Tr(\Sigma_r + \Sigma_g - 2(\Sigma_r\Sigma_g)^{1/2}) \tag{20}$$

（4）阈值准确率是用来度量分类模型性能的优劣的一种方法，是指二分类问题中，分类器在不同的阈值条件下分类器的正确率。将各分类器输出的概率或得分与某个阈值相比较，高于该阈值的判为正类，低于该阈值的判定为负类。其中，阈值精度是衡量深度计算精度的一个重要指标，是指在一定的阈值范围内，预测深度和真实深度之间的最大比值在某个阈值范围内的像素占比。该方法采用三种不同的阈值，分别对三种阈值进行了比较，阈值准确率越大，表示预测深度越接近真实深度，性能越好，像素点越全面。δ的表达式如下：

$$\delta = \max(\frac{p_i}{g_i}, \frac{g_i}{p_i}) < thr, thr = 1.25, 1.25^2, 1.25^3 \tag{21}$$

在上述指标中，PSNR、UQI、δ的值均为越大越好，FID 越小越好。其中，

FID 为两个图像之间的相似度，数值越低表示生成的图像与原始图像越相似；PSNR、UQI 数值越高表示生成的图像质量越好：不同阈值时的准确率可用于检验预测深度与真实深度之间的差距，表示小于 thr 的像素点占总像素点的百分比，数值越接近 1 效果越好,像素点越全面。

AH-AAE 模型的损失函数中涉及SSIM 损失函数，若使用此指标将会对评判结果产生较大误差，故本次实验的评价指标不包含SSIM 。

4.3 实验参数设置

本文模型是在 PyTorch 框架的基础上实现的，并在深度学习工作站上运行，该工作站配备 IntelCore i7-13700KF@2.7GHz（16 核 24 线程）、NVIDIA GeForce RTX4080（64 GB 内存和 16 GB 显存）、Win11 操作系统、Python 版本 3.7.12、PyTorch 版本 1.13.1、cuda 版本 11.7，torchvision 版本为 0.14.1。 模型在 MS-COCO 公共数据集上进行训练，使用 Adam 优化器进行优化，初始学习率为 0.0001 且不进行调整，共训练了 40 次。模型使用训练集来进行训练，比例划分为： 80% 训练数据、 20%测试数据。

4.4 实验分析

本章对 AH-AAE 模型的性能进行多项测试，并通过检测结果判断模型处理不同类型问题的能力。

4.4.1 不同损失函数下 A-ACE 模型区域填充比较

本实验重在根据已有的研究成果，对三种损失函数（MSE、MS-SSIM 以及 MS-SSIM-L_1）进行了分析(Zhao et al., 2016)。

（1）最基本的损失函数是均方误差（Mean Squared Error，MSE），MSE 可以用 $l_{\mathrm{MSE}}=\frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{x}_i)^2$ 表示，其中 x_i 是输入数据的第 i 个元素，$\hat{x}_i$ 是相应的重建结果，n 是元素总数。因此，总损失函数可以表示为所有样本的均方误差。在训练过程中，首先用反向传播算法计算梯度，然后用梯度下降等优化方法求解，并

更新模型参数。算法使用重构结果时，首先对输入信号进行编码，再将其传递给解码器进行解码。然后计算重构结果与原始输入之间的均方误差，再通过反向传播算法计算梯度并更新模型参数。

（2）另一项为MS-SSIM 损失函数， MS-SSIM 为多尺度结构相似函数，MS-SSIM 是SSIM 的多尺度形式，误差函数是感知驱动的像素为 p 的SSIM 函数定义有：$\mathrm{SSIM(p)}=\frac{2\mu_x\mu_y+C_1}{\mu_x^{\ 2}+\mu_y^{\ 2}+C_1}\cdot\frac{2\sigma_{xy}+C_2}{\sigma_x^{\ 2}+\sigma_y^{\ 2}+C_2}=l(\mathrm{p})\cdot\mathrm{cs(p)}$。其中，均值和标准差由一个带标准差 σ_G 的高斯滤波器 G_{σ_G} 计算。σ_G 的选择对使用SSIM 训练的网络的处理结果质量有影响。具体来说，对于较小的 σ_G 值，网络失去了保持局部结构的能力，并且在平坦区域重新引入了斑点伪影。对于较大的 σ_G，可观察到网络倾向于在边缘附近保留噪声。与微调 σ_G 不同，提出使用多尺度版本的SSIM，即MS-SSIM 。给定一个由 M 个尺度组成的二进金字塔，MS-SSIM 定义为：$\mathrm{MS\_SSIM}(P)=l_M^{\alpha}(p)\cdot\prod_{j=1}^{M}cs_j^{\beta_j}(p)$，相应的损失函数可写作：$\mathcal{L}^{\text{MS-SSIM}}(p)=1-\mathrm{MS-SSIM}(\tilde{p})$

（3）最后进行对比的损失函数为MS-SSIM－$\mathrm{L_1}$ 损失函数，MS-SSIM 和SSIM 对均匀偏差都不是特别敏感。这就导致了亮度的改变，或者是色彩上的改变，一般看起来比较暗。然而，MS-SSIM 比其他损失函数更好地保持了高频区域的对比度。另一方面，$\mathrm{L_1}$ 保留了颜色和亮度，无论局部结构如何，误差都被同等地权衡，但不会产生与MS-SSIM 相当的对比度。为了捕获两种误差函数的最佳特性，将它们结合起来有：

$$\begin{aligned}\mathcal{L}^{\text{MS-SSIM-}L_1}&=\alpha L^{\mathrm{MS\_SSIM}}+(1-\alpha)\mathrm{G}_{\sigma_G^M}\cdot L^{l_1}\\&=\alpha(1-\mathrm{MS\_SSIM}(\tilde{p}))+(1-\alpha)\mathrm{G}_{\sigma_G^M}\cdot L^{l_1}\end{aligned}\tag{18}$$

式（18）中，省略了对所有损失函数对 p 的依赖，设定 $\alpha=0.84$。式（18）的导数只是它的两项导数的加权和。$G_{\sigma_G^M}$ 和 L^{l_1} 之间添加了一个逐点相乘：这是因

为MS-SSIM 根据其对中心像素 MS-SSIM 的贡献 $\tilde{p}$ 来传播像素 q 处的误差，由高斯权重决定。

训练三个损失函数，并将运行结果进行整理，得到的数值和图像对比如下表 1 以及图 4.1 所示。实验设置， gt 为原始图像；mask 为掩膜图像。

表 1 损失函数对比结果

模型	PSNR	UQI	FID	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
MSE	27.8304	0.9815	2.3760	**0.9039**	0.9439	0.9596
MS-SSIM	26.5735	0.9694	3.9783	0.8713	0.9254	0.9489
MS-SSIM-L_1	**27.8351**	**0.9821**	**2.1236**	0.9036	**0.9448**	**0.9606**

从表 1 可以看出，在三种损失函数中， MS-SSIM $-$ L_1 的综合性能优于其它两种损失函数。但是，$\delta < 1.25$项的 MS-SSIM $-$ L_1 值要低于 MSE 的值，这是因为在不同的图像类型下，掩膜位置的随机性会对训练效果造成一定的影响，从而使预测深度出现偏差，最直接的反映在了阈值为 1.25 上的准确率上。

(a) gt (b) mask (c) MSE

(d) MS-SSIM (e)MS-SSIM-L_1

图 4.1 不同损失函数对比图

图 4.1 显示了使用 ACAE 模型的不同损失函数对损坏区域的填充情况。比较了三种不同损失函数对修复损坏图像的影响，发现使用 MSE 和MS-SSIM 训练的神经网络在传统性能和感知损失方面都略低于MS-SSIM$-L_1$算法。从图4.2（d）（e）中可以看出，（d）中彩色图块的中心融合在一起，而（e）中由于上方彩色图块偏移的边缘内容比（d）略强。在各种图像质量评估指标中，MS-SSIM和L_1组合的损失函数结果最好，因此MS-SSIM$-L_1$损失函数被用作重建损失函数的一部分。

4.4.2 随机区域缺失值填充方法对比实验

为了说明方法的有效性，将本文模型与基准模型进行对比实验。本文模型为AH-AAE。将对比实验设置如下：

（1）基础卷积自编码器模型(Mao et al., 2016)，记作 CAE，采用卷积层代替全连接层，原理与自编码器类似，下采样输入以提供更低维的隐表示，并强制自编码器学习象征的压缩版本。

（2）融合注意力的卷积自编码器(Bodapati, 2022)，记作 ACAE，注意力机制在深度学习中有着巨大潜力，该网络是一种基于端到端训练空间注意力的卷积神经网络体系结构来进行识别的模型。在分类模块中引入的注意力机制，确保识别区域得到更高的注意力得分；

（3）基于 PGGAN 的生成对抗模型(Demir et al., 2018)，记作 PGGAN，PGGAN 方法包括一个鉴别器网络，它结合了一个全球性的 GAN 架构与补丁 GAN 的方法 PGGAN 首先共享 G-GAN 和 patchGAN 之间的网络层，然后将路径进行拆分以产生两个对抗性的变量，这些变量为生成器网络提供数据，以便捕获图像纹理的局部连续性和图像中普遍的全局特征。

（4）Context-Encoder 模型(Pathak et al., 2016)，记作 CE，是第一个基于 GAN 的修正算法，也是将上下文编码器与 GAN 进行融合的算法，它强调在修改过程中理解整体图像的背景是非常重要的，并且使用完全连接层来完成这个功能。

实验设置，gt 为原始图像；mask 为掩膜图像；CAE、PGGAN、CE 表示同上述定义；其中 CAE 为基准模型，ACAE 模型使用重构损失函数 MS-SSIM-L_1，

未加入 HRNet 的网络记作。A-AAE 具体参数值如表 2 所示，图像对比结果如图 4.2 所示。

表 2 多模型对比结果

模型	PSNR	UQI	FID	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
CAE	26.56	0.9753	4.9308	0.8672	0.9336	0.9568
ACAE	27.8351	0.9821	2.1236	0.9036	0.9448	0.9606
PGGAN	27.6873	0.9809	6.5873	0.9555	0.9682	0.9752
CE	27.7481	0.9813	5.5743	0.9569	0.9694	0.9763
A-AAE	31.0375	0.9908	1.3868	0.9602	0.9704	0.9808
AH-AAE	**33.7825**	**0.9928**	**1.1492**	**0.9748**	**0.9806**	**0.9931**

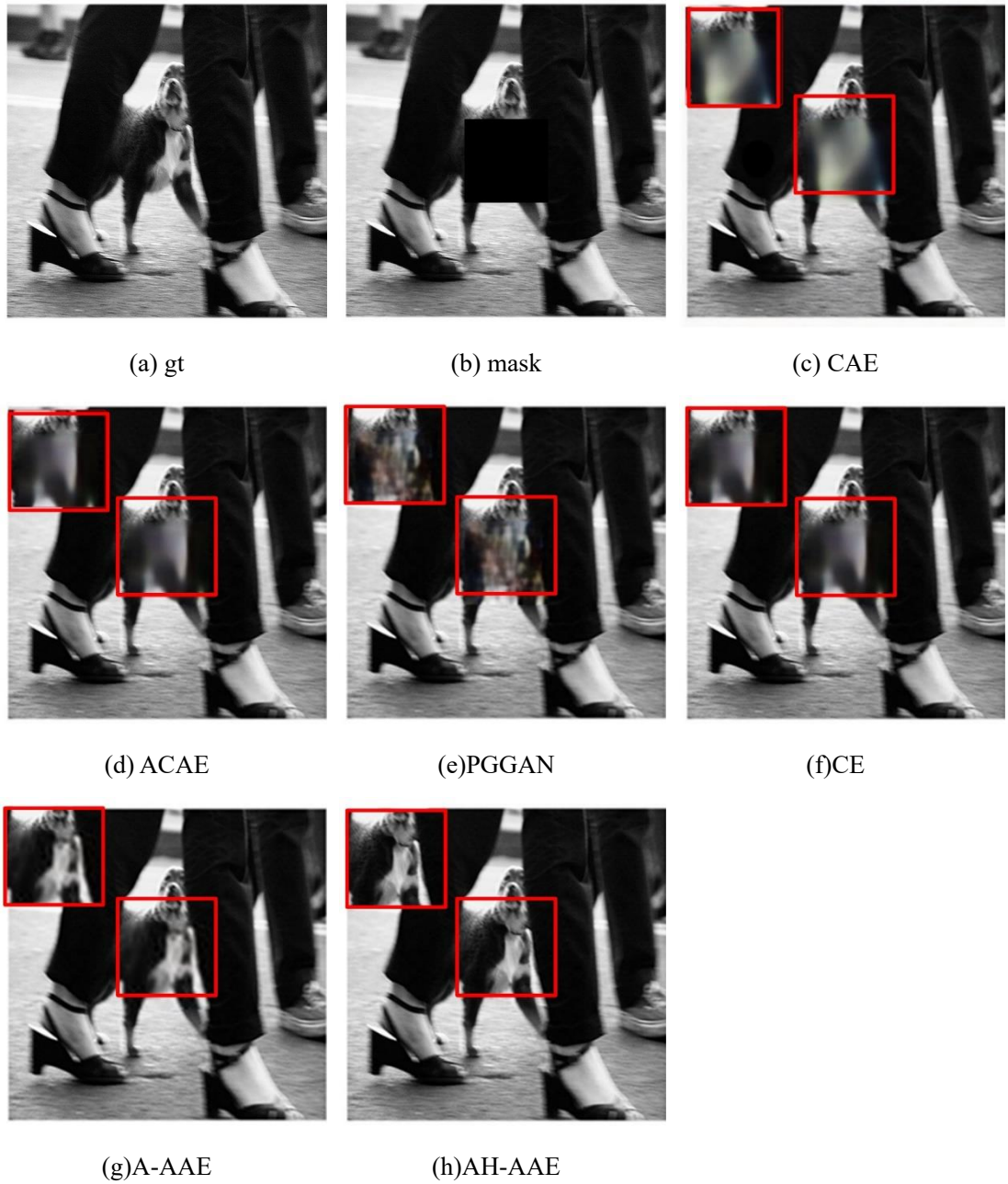
由表 2 中的指标结果可知，引入注意力机制和 HRNet 的 AH-AAE 网络的填充效果要好得多。在进行模型对比的指标方面，可以进行以下分析：

（1）PSNR 根据原始图像与重建图像之间均方误差的对数标度来衡量图像的差异，即 PSNR 值越大，重建图像与原始图像的差异就越小，从而可以更好地还原图像。如果 PSNR 值大于 30 dB，通常被认为重建质量较好，而小于或等于 20dB 则表示图像质量较差。从表 2 可以看出，四种方法的信噪比峰值都小于 28dB，而 AH-AAE 模型的信噪比峰值都大于 30 dB，说明在相同条件下，注意力机制和 HRNet 的加入能更好地重建图像。

（2）与 PSNR 不同，UQI 是一种基于亮度协方差、亮度均值和亮度方差等图像属性的相似度量。UQI 值范围从 0 到 1，0 代表图像质量最差，1 代表质量最好。表 2 中列出的六种算法的 UQI 值都在 0.98 左右，这意味着这些方法对人眼的恢复效果基本上可以忽略不计。由此可以得出结论，加入注意力机制和 HRNet 可提高成像质量、相似度以及保真度。

（3）FID 是评估生成模型质量的一个重要指标；FID 值没有严格的上下限，但一般为正值；FID 值越小，真实图像与重建图像之间的差异越小，生成图像的质量越好。UQI 因自身的限制，不能很好的反映人的主观感受，因此，从主观感知的角度来看，FID 对 UQI 结果的主观感知有一定的辅助作用。从表 2 可以看出，对比模型的 FID 值有高有低，而加入注意力机制和 HRNet 的 FID 值均小于 1.5，说明引入注意力机制和 HRNet 的生成模型在生成重建图像方面具有较高的水平。

（4）其定义是在一个模型中，有正确类别的样品数目与样本总数的比率。高精度代表了该模型在识别任务中具有更高的准确度。该方法是用来衡量深度和实际深度的差距，准确率的值是指低于阈值 thr 的像素点所占总体像素点的百分比，越接近于 1 效果越好，像素点越全面。从表 2 可以看出，引入注意机制的模型在三种不同的阈值下都有较好的性能，其中基于注意机制和 HRNet 的 AH-AAE 模型的预测准确率高达 0.9748。



(a) gt (b) mask (c) CAE

(d) ACAE (e)PGGAN (f)CE

(g)A-AAE (h)AH-AAE

图 4.2 图像对比结果

从图 4.2 中可以看出，加入注意力机制后，图（c）和（d）中的图像细节更加丰富。然而，图（e）和图（f）是包含 GAN 的网络结构，当 GAN 填充图像时，不同图像的纹理和结构会出现差异，出现模糊和失真等严重问题。从图（e）和图（f）可以看出，虽然填充后图像的颜色分布相对一致，但仍有部分像素模糊。GAN 的不足之处在于要使用大量的训练模型，并结合其他注意机制来提高生成算法的表现力和理解力。(g)为融入注意力机制的填充效果，可以清楚地看到不同的输入图像会根据自身的特点调整注意力的分配，使原来被破坏的部分得到更多的关注，从而减少不一致、失真、模糊等现象的发生。（h）作为注意力机制和 HRNet 的填充结果的整合，采用 HRNet 进行多层次特征提取，以提高对填充区域细节的表达能力；同时，注意力机制可以让生成器更多地关注填充区域。这两种方法的结合可以很好地解决填充图像时出现的纹理一致性、细节保留和真实感等问题。将（h）与其他图像进行比较，可以明显看出（h）的图像更加清晰，分辨率更高。

4.4.3 噪声实验

噪声是图像中出现的一种不必要的随机变化，主要表现为图像中像素值的连续或无规则的移动。灰度图像每个像素只有一个亮度值。相比之下，彩色图像包含多个颜色通道；每个像素包含多个颜色分量，当噪声出现在某个特定颜色通道时，彩色图像中噪声的影响可能会被颜色变化所掩盖。为了验证这次试验的精度，本节噪声实验在 MS-COCO 数据集中加入了不同比例的噪声进行测试性能。由于篇幅所限，此部分所提到的表格资料和图像资料均为 10 张图像的试验结果；待检测的图像如图 4.3 所示，其中包含灰度图像、彩色图像各 5 张。

将 A-AAE 与传统矩阵填充算法进行对比，检验模型的去噪能力。其中涉及的基础算法包含：SVT、SVP、$S_P - l_p$（*Schatten p* 范数和 l_p 范数）、TNNR-ADMM。

在此基础上，通过去除 HRNet 后只添加注意力机制的模型（记作 A-AAE)，来验证 HRNet 是否可增强模型的去噪能力以及改善图像分辨率。

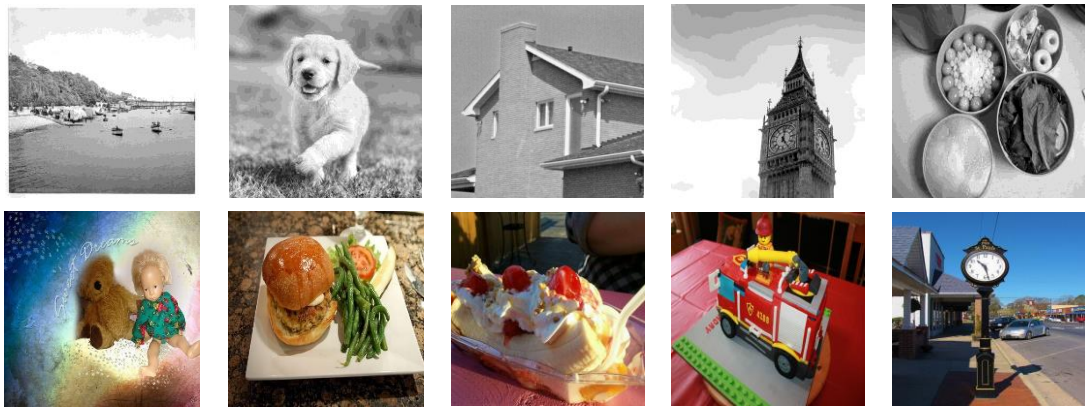


图 4.3 待检测图像

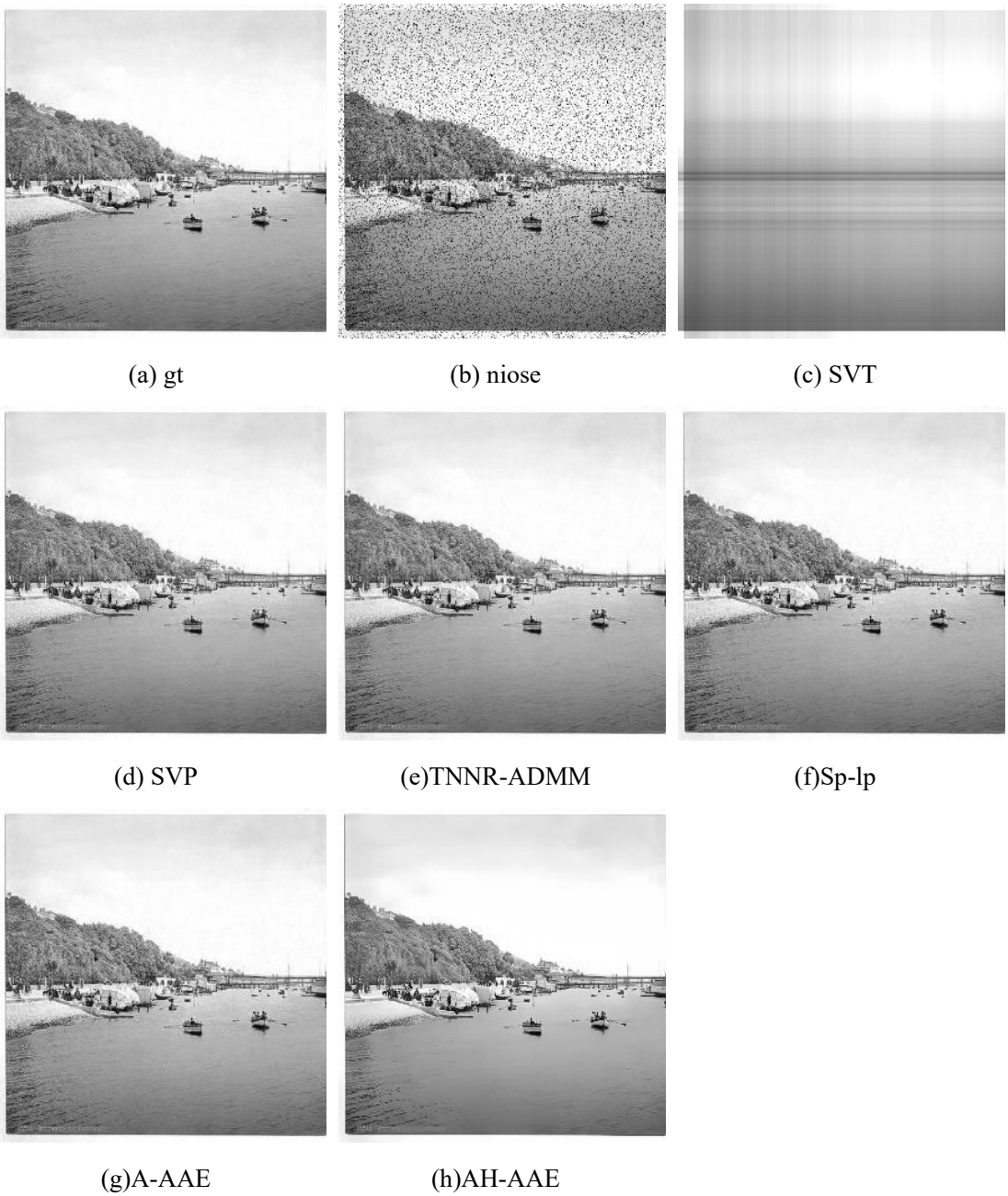
在复原效果方面，以 PSNR 值作为客观评价标准。在此基础上，本文也将重构出的图像进行了直观的对比。在表 3、4、5 中给出了噪声测试指标结果；图 4.4-4.7 展示了具体的噪声图像恢复结果。

表 3 随机添加 10%噪声像素修复 PSNR 指标对比

图像	SVT	SVP	$S_P - l_p$	TNNR-ADMM	A-AAE	**AH-AAE**
beach	28.9852	37.6556	37.3420	39.3190	37.6806	**39.6871**
dog	28.3926	35.2700	36.9115	36.8768	37.0487	**37.6461**
house	28.3380	35.5403	37.1251	38.5108	37.3642	**38.5343**
loft	36.1849	38.9590	38.0100	37.0939	38.2220	**38.2977**
lunch	33.0964	35.4930	40.4203	40.9818	41.4459	**42.9971**
toy	34.5365	37.9884	38.5340	41.3978	38.9156	**42.5101**
food	34.6526	35.6308	41.3720	43.0175	41.8579	**43.8505**
cake	34.7889	37.1905	44.9467	46.2685	45.4523	**46.2786**
car	36.2918	36.5718	47.6295	47.4520	48.0653	**48.0823**
lamp	35.4411	37.0327	39.6574	41.8474	40.0299	**41.8656**
Average	33.0708	36.7332	40.1949	41.2766	40.6082	**41.9749**

表 3 是随机加入 10%噪音的像素修补的结果，从表 3 可以看出，不管用哪一种算法进行向上比较，后 5 个彩色级图像的指标值都比前面 5 个灰度级图像的指标值稍微高一点。采用 SVT 和 SVP 方法获得的图像平均 PSNR 值分别为 33.0708 db 和 36.7332 db，产生这种情况的可能是这两种算法对彩色图像均采取矩阵化处理方式，忽视了图像各色组通道间的结构信息；或是图像中的噪声并不是高斯白噪声而是具有特性的噪声，图像内容受到奇异值算法的影响，导致此类算法的去

噪效果不佳。AH-AAE 算法对彩色和灰度两种图像都有较好的去噪效果，特别是与未加入 HRNet 的方法相比，峰值信噪比平均提升 1.3667，表明引入 HRNet 不仅可以改善图像的分辨率，而且可以有效地抑制噪声。



(a) gt (b) niose (c) SVT

(d) SVP (e)TNNR-ADMM (f)Sp-lp

(g)A-AAE (h)AH-AAE

图 4.4 图像 beach 随机添加 10%噪声灰度图像对比

图 4.4 显示了各种不同的算法在图像 Beach 随机加入 10%噪音之后的去噪的效果比较。从图像中可以清楚地看到，虽然加入了一些噪音，但是仍然可以看

到目标的轮廓。SVT 方法是根据信号和噪声在奇异值分解过程中的能量差，进行去噪处理。如果信号和噪声的能量相差较小时，很难对噪声进行准确的辨识和分离，从去噪的效果来看，SVP、TNNR-ADMM、Sp-lp 三者的修补效果相似，从表 3 中还可以看出，这三种方法的 PSNR 也比较接近，这意味着在低噪音情况下，模型修复效果很好。如图 4.4 所示，（g）和（h）的修补结果色块更加显眼，两幅图像色彩的亮度和暗度都有显著的变化，呈现出更加丰富和层次丰富的效果，尤其是在（h）中，虽然是灰度图像，但是仍可以清楚地看到内在细节，并没有由于灰度等级的类似而产生的模糊和混乱，可以更加精确地分辨出不同的对象或特征。实验结果显示，AH-AAE 算法在保持图像高分辨率的输出的情况下，具有较好的降噪效果。

表 4 随机添加 30%噪声像素修复 PSNR 指标对比

图像	SVT	SVP	$S_P - l_p$	TNNR-ADMM	A-AAE	**AH-AAE**
beach	28.3914	34.6679	30.5059	35.1927	33.6604	**35.6871**
dog	28.2785	32.8993	34.7142	35.1506	34.8635	**36.6461**
house	33.4389	32.9095	33.4242	35.3269	33.6531	**35.5342**
loft	35.2522	36.1442	35.1636	36.7830	35.2048	**36.7977**
lunch	33.3806	32.3472	33.7062	36.1585	34.0116	**36.4203**
toy	33.2056	33.7741	33.0036	35.1942	33.1623	**35.5101**
food	33.2962	32.0728	33.8101	35.4083	34.0177	**35.8502**
cake	35.1495	33.4200	36.2026	38.0747	36.4586	**38.2086**
car	35.7065	33.1963	37.7000	38.6177	37.9192	**39.0523**
lamp	34.4257	33.5712	34.0972	35.8925	34.3291	**36.5656**
Average	33.0525	33.5002	34.2328	36.1799	34.7280	**36.6272**

表 4 是随机加入 30%噪声的像素修复结果，其中有 30%为低噪声，虽然会对影像造成一定程度的干扰，但仍然保有较高比例的原始图像信息，属于一种可控制的情况。从表 4 可以看出，在各种类型的图像上，随机加入 30%的噪音后，大多数算法的数值相差不大，只有 SVT 方法存在低于 30db 情况。另外，从比较表 3 的结果中还可以看出一个有意思的现象，即每一种算法的平均峰信噪比较器（PSNR）都会随著噪声比率的增大而下降。这说明，当图像被破坏的程度越大，样本图像间的相关性就越小，修复的困难也就越大。随着图像中噪声比重的增大，图像中缺失的信息也越来越多，因此，修复算法必须对缺失的信息进行更加精确

的评估与还原，这给修复带来了更大的挑战。因此， PSNR 值的下降反应了修复的困难程度增加以及修补质量的降低。

(a) gt (b) niose (c) SVT

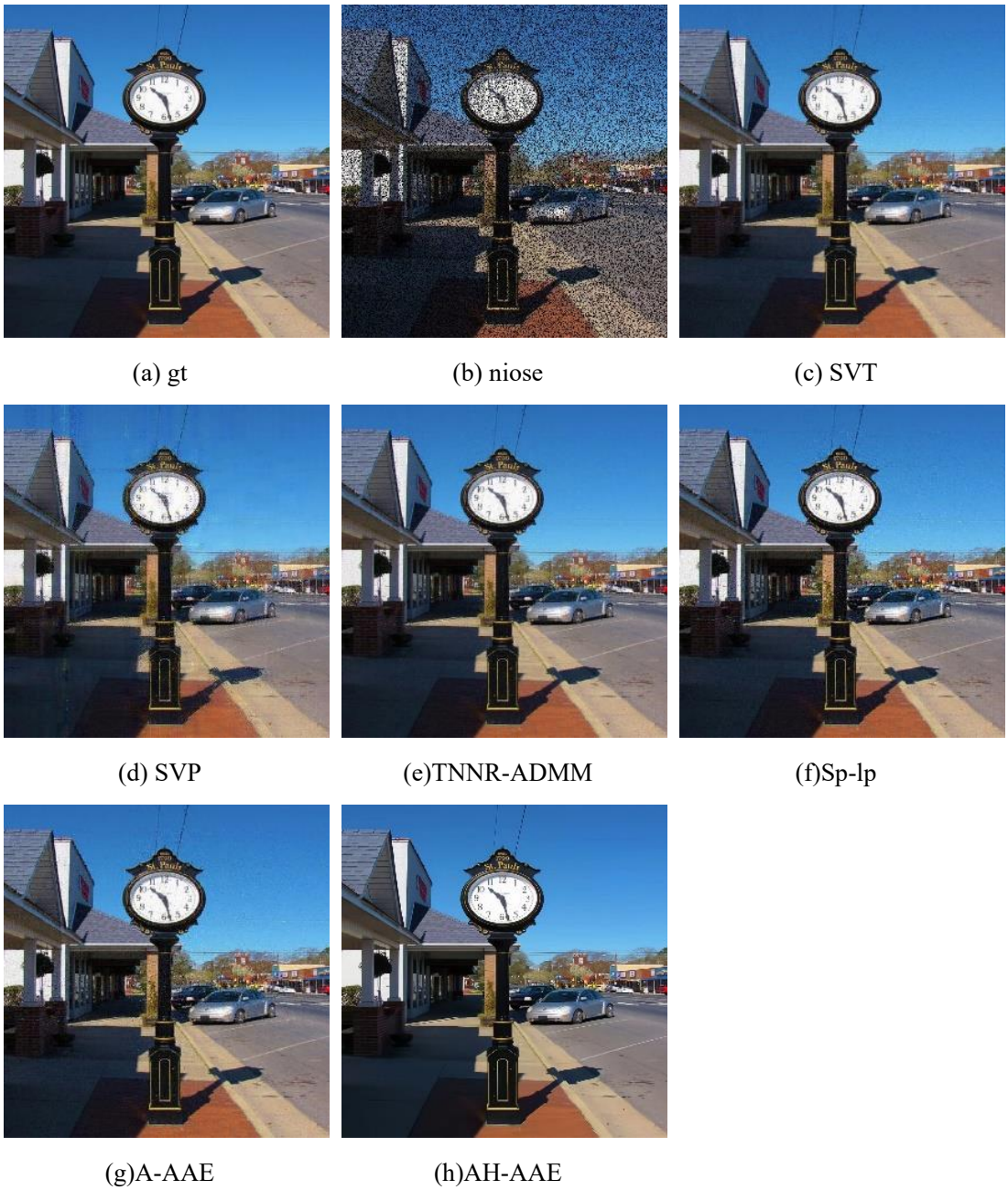
(d) SVP (e)TNNR-ADMM (f)Sp-lp

(g)A-AAE (h)AH-AAE

图 4.5 图像 dog 随机添加 30%噪声灰度图像对比

图 4.5 是一个灰度级的图像，在这组图像中，（c）出现与图 4.4（c）相同的问题，大多数算法都能很好地复原，虽然细节上有些缺陷，但仍可以对整个图像进行有效地去噪。在此基础上，将图像放大进行观察，可注意到包含注意力机制

的模型可提升图像的对比度，提高图像的分辨率，增强图像的颜色，减少噪声带来的伪影，从而提高图像的真实性和清晰度。



(a) gt (b) niose (c) SVT

(d) SVP (e)TNNR-ADMM (f)Sp-lp

(g)A-AAE (h)AH-AAE

图 4.6 图像 lamp 随机添加 30%噪声彩色图像对比

图 4.6 为彩色图像修复，虽然图像（c）、（d）、（e）、（f）、（g）均实现了有效的降噪，但（e）、（f）中的天空部分出现了一些棋盘形状错乱。（c）、（d）的天空存在伪影，且标识牌轮廓线不够清楚；（g）通过引入注意力机制，能够更好的保

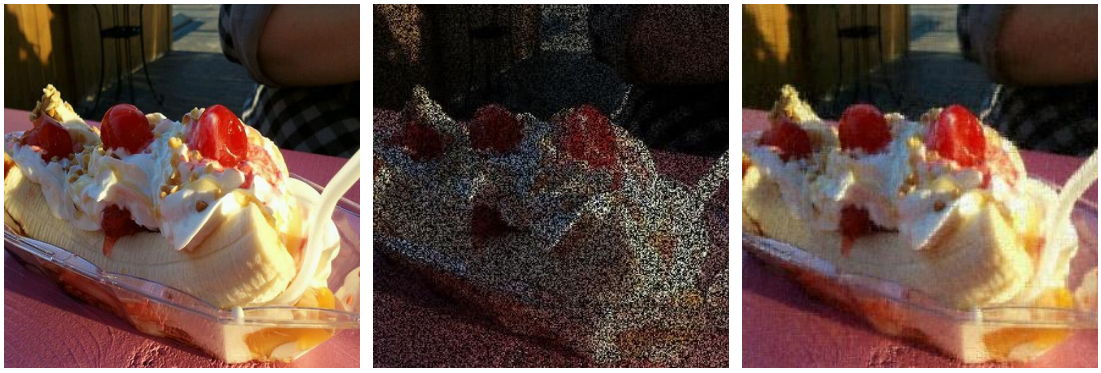
留图像的细节，同时房屋轮廓更加完整，周边的人工痕迹也更加明显；（h）由 AH-AAE 模型生成，提高了（g）的分辨率，更好地还原了图像的色彩信息，使图像更加光滑、更加细致、更加细致。

表 5 随机添加 60%噪声像素修复 PSNR 指标对比

图像	SVT	SVP	$S_P - l_p$	TNNR-ADMM	A-AAE	**AH-AAE**
beach	28.9968	32.7775	31.6033	32.8135	31.6723	**33.6871**
dog	28.3880	30.9736	31.7562	32.6423	31.8563	**36.6461**
house	31.3352	31.1828	30.6432	31.9654	30.7430	**34.5343**
loft	33.4235	34.1720	33.7961	34.6871	33.8109	**34.6877**
lunch	31.0524	30.5371	30.4725	31.7459	30.5926	**36.4203**
toy	30.9476	31.3404	30.4389	31.6133	30.5302	**34.5101**
food	30.5727	30.2333	30.5115	31.1265	30.5907	**35.8502**
cake	31.9567	31.2546	31.8383	32.8609	31.9575	**37.2086**
car	32.1964	31.1325	32.3394	33.0139	32.4735	**37.0523**
lamp	31.7744	31.5945	31.2383	32.3419	31.3350	**35.5656**
Average	31.0644	31.5198	31.4638	32.4811	31.5562	**35.6162**

表 5 为添加 60%高比例噪声的实验。具体来说，加入较高层次的噪声，会增加影像中的噪声比重，进而对影像品质及视觉效果造成较大的影响。从表 5 可以看出，不同算法之间的间隔在不断减小，这意味着遇到高噪声的情况下，强度会变得非常大，不同的算法都会遇到巨大的挑战。当噪声强度达到一定程度时，某些方法的降噪效果会有很大的局限性。将模型处于高噪音环境下获得的检验结果将为故障诊断技术的进一步完善与优化提供理论依据。

图 4.7 展示了在高噪音环境下的修复结果，从图 4.7 的不同修复算法可以看到，（c）、（d）、（e）中存在的伪影较多，图像的色彩也不平衡。通常情况下，噪点越多，图像的清晰度就越差。因此，如果将噪声增加到 60%时，图像就会出现更多的噪点，也会有色彩变化、颜色偏移以及失真等问题。在噪点较小的情况下这些问题并不明显。添加更高水平的噪声可以更直观地测试图像处理算法的性能，同时也会带来更多的挑战与优化空间。



(a) gt (b) niose (c) SVT

(d) SVP (e)TNNR-ADMM (f)Sp-lp

(g)A-AAE (h)AH-AAE

图 4.7 图像 cake 随机添加 60%噪声彩色图像对比

4.7 (f)为 SPLP 算法，可以清楚地看到整个图像的修复效果很好，但是图像中仍有众多彩色噪声，导致颜色不平衡。(g)（h）两种方法都引入了注意机制以改善图像的细节，引入注意力机制后，该模型在高噪声的环境中具有较强的去噪能力。注意机制允许模型更加关注图像中的重要信息，从而有效过滤掉噪声。通过对图像不同区域的注意力分配，模型能够优先处理局部噪声较强的区域，改善降噪效果。实验结果显示，在噪声强度为 60%的情况下，模型仍能有效地保留图

像中的细节及色彩，说明注意力机制对于去噪任务的性能改进起到了积极的作用。(h)在注意力机制的基础上，引入 HRNet，能够更准确地捕捉噪声并对其进行降噪处理，利用特征抽取与信息融合的能力，实现图像色彩信息的精确恢复。AH-AAE 在噪声强度为 60%的情况下，能够有效地抑制噪声，保留图像细节并精确地还原色彩，对噪声去除具有较强的优势。实验结果表明， HRNet 的应用对去噪效果的提高有很大的促进作用。

4.4.4 特征模块深度实验

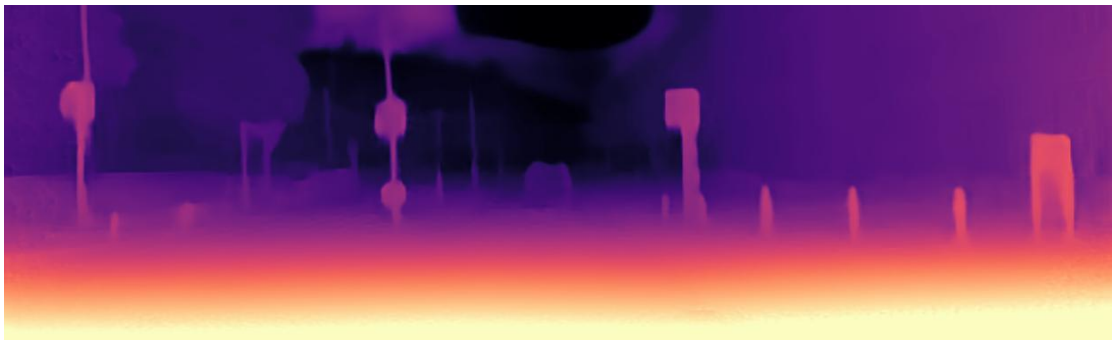
在这一部分中，将深入探讨引入注意力机制所产生的深度图的变化。

CSFM 特征模块能够通过生成特定区域或通道的权重向量来重新校准通道图，以增强网络对重要特征的关注程度，提升模型的表达能力和性能。Sigmoid 函数生成的权重分数在 0 到 1 之间，表示相应特征的重要性，即分数越高表示通道越重要，用于显示网络关注的特征。LFM 特征模块采用了空间注意力机制与位置编码结合的方法，可提升深度图分析任务模型的准确性和鲁棒性。这种机制能够更好地捕捉深度图中的位置信息和结构特征，并更精确地描述对象的边界。通过逐步恢复空间分辨率，模块能够利用细节信息，提高在物体识别、深度估计、场景重建和关键点检测等任务中的处理能力。这些功能帮助模型获得更有效的深度图，从而得到更准确和可靠的结果。引入注意力机制后，CSFM、LFM 可辅助 AH-AAE 学习权重，来使其更好地关注到图像中最重要、最有鉴别能力的区域。这有助于更好地理解图像的空间结构与之间的联系，进而提升图像处理任务的准确性和鲁棒性。

图 4.4 (a)和 4.5 （a）是待检测的图像，可以很清楚地看到，这两幅图都是马路图像，观察图像的内容来看，两幅图所包含的元素都比较完整。图 4.4 (a)显示了较少的车辆和更多的交通标志，而两侧则有明显的边界，例如房子和树；从图 4.5 (a)可以看到，这张照片属于交通较堵塞的高速公路，图像的上部分天空占比较大，两侧为较高的树木及高速公路护栏。

（a） 输入图像

（b） 深度图



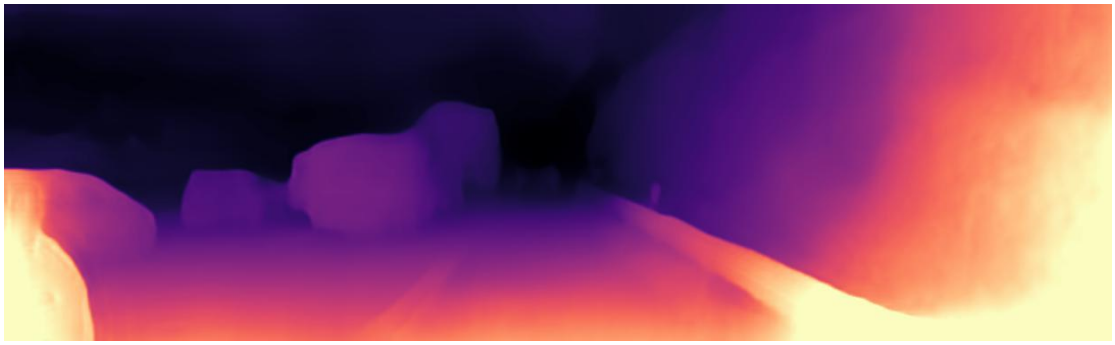
（c） 注意力模块环境感知输出

图 4.4 空旷马路深度图对比

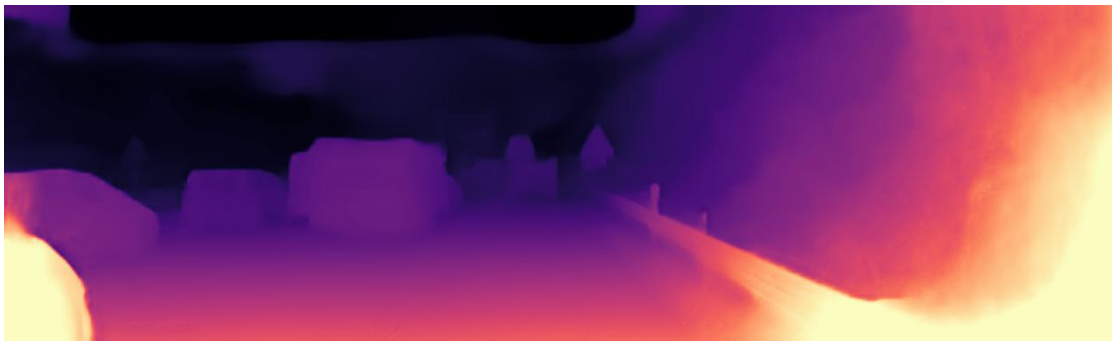
由图 4.4（a）（b）可知，马路在深度图中会识别为亮度较大的区域，周围的标识牌、结构柱等以及树木房屋均会被识别。对比图 4.4（b）（c），可以发现，加入注意力机制后各种标志的轮廓变得明显，一些在原深度图中与背景颜色近似但未识别的区域也识别到了轮廓，这表明注意力机制能够自动学习图像中哪些区域是关键的，如何着重于此区域使生成的深度图能在重要的目标或结构上得到加强。同时，注意力机制有助于减少或去除一些无关紧要的背景信息对于最终特征

图的干扰。模型更加专注于目标或感兴趣的区域，使得背景的影响减弱，从而使得生成的特征图更加干净和清晰。图 4.4（c）经过注意力机制的加持，更精确地描述了对象的边界，这表明解码器部分自适应地选择信息丰富的局部细节，并帮助网络处理和定位物体边缘，以获得更清晰的深度预测，可逐步恢复空间分辨率。

（a） 输入图像



（b） 深度图



（c） 注意力模块环境感知输出

图 4.5 拥堵车流深度图对比

图 4.5 中的（c）为通过注意力机制生成的特征图，能够很好地捕捉图像中的

细节和纹理。这是因为注意力机制可以聚焦于图像中具有高频信息的区域，使得特征图在细节方面更加丰富和明显。经注意力机制生成的深度图能够根据不同的输入图像来自适应地调整关注点，即使是相似的场景，特征图也可以根据图像的不同重点和需要进行调整，使得特征图能够更好地适应不同目标或环境。与图 4.4（c）不同，图 4.5（c）探测出了更远处的特征，充分说明了每个通道图都得到来自较远区域的更多注意。此外，它还特别强调了消失点区域，这可作为理解场景几何形状的有力线索。

将注意力机制引入到图像中，形成的深度图是一种动态的视觉表征。该方法利用注意力机制，能够根据不同的视觉特点，自适应地学习最感兴趣的区域，并对其进行重点研究。在此基础上，引入注意力机制所产生的特征图，使图像具有更加细致和丰富的信息表达能力。就像人的视觉注意力可以迅速将焦点集中到图像上的重要对象或区域一样，这样的特征图还可以对图像中的关键要素进行更多的关注。由注意力机制生成的特征图可视为一种加权表达，用权值来表征各像素的重要程度。该方法在去除不相关背景信息的情况下，更好地保留图像中的关键细节与结构，使其生成的深度图能够更加灵活地适应不同样本之间的差异，从而提升模型的泛化能力和鲁棒性。同时，注意力机制能够在不同的尺度和层级间进行信息传递和融合。生成的深度图可以看作是对不同层级特征的综合表示，能够有效地融合整体与局部信息，从而具有更加完整、更加丰富的特征表达能力。生成的深度图通常能够保留高分辨率的特征。这意味着在深度图中，能够更清晰地看到目标的细节和边缘等重要信息，有助于后续任务的精确性和准确性。此外，注意力机制还具有跨层次、跨规模、跨层次的信息转移与融合能力。生成的深度图可以看作是对不同层级特征的综合表示，能够有效地融合整体与局部信息，从而具有更加完整、更加丰富的特征表达能力。所得到的深度图像往往具有较高的分辨率。这种方法可以使图像中物体的细节、边界等重要信息更加清晰，从而提高了图像填充的精度。

通过本节的实验可以看出，加入注意力机制后生成的深度图在重要特征突出、信息融合、高分辨率特征保留以及动态自适应性等方面都表现出了优秀的特点，为后续的任务提供了更强大且有意义的特征表示。

4.4.5 消融实验

消融实验用来评价和了解深度网络中各关键要素对模型表现的作用。在评价过程中，通过选择性地去除特定的模块，以评价模型与那些模块之间的依赖关系，选取合适的评价指数。评价标准主要有分类精度、损失函数和收敛速率等。通过消融实验，可以定量地分析某个部件对模型性能的影响，从而深入了解模型的内在机理。通过对模型进行消融实验，可以揭示模型的关键驱动因素，为模式的设计与优化提供指导。

为了验证 AH-AAE 模型的各个功能模块的性能，设计了相应的消融实验。通过对每个组件的添加进行控制，并通过相关的数值和图像结果整体评估每个组件能否显著提高模型的填充能力。

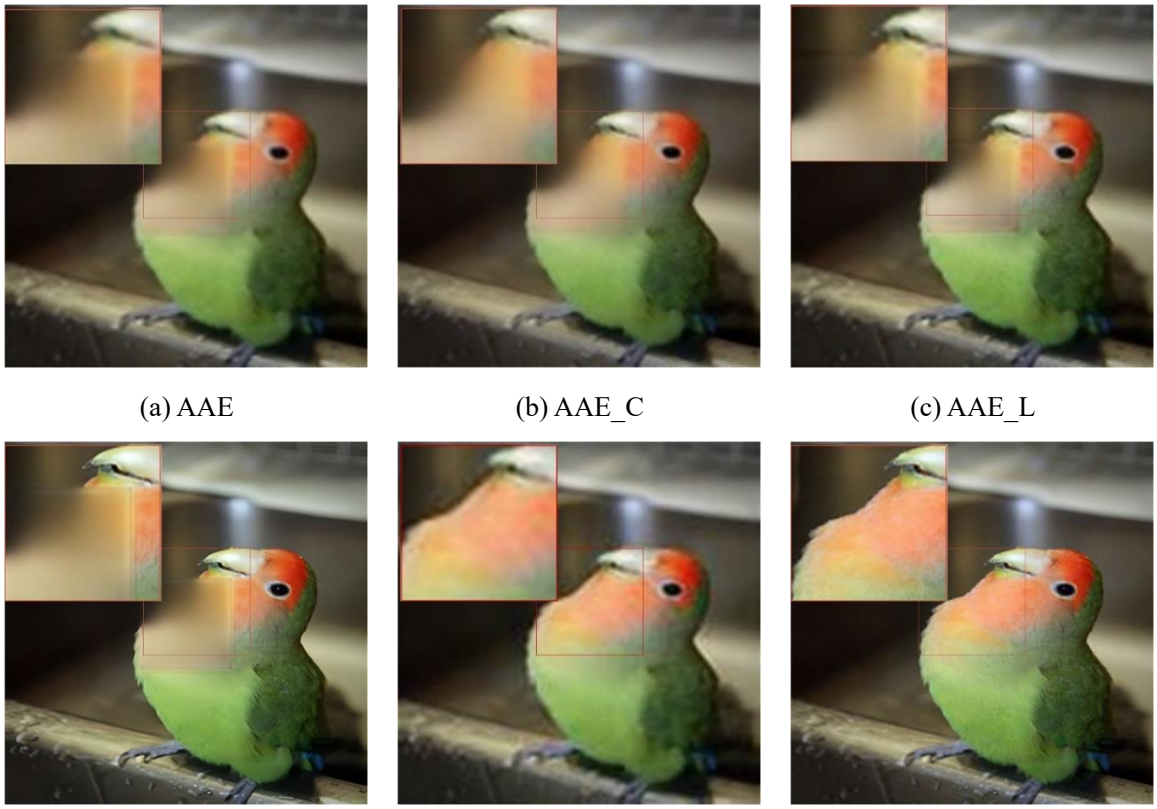
AH-AAE 模型的创新点在于融合了 HRNet 与注意力机制，其中注意力机制由 CSFM、LFM 两部分组成。为验证上述三个子模块对模型填充能力，将这些部分依次去除，进行消融实验。实验设定 AAE 作为基准模型，将只包含 CSFM 模块的模型记作 AAE_C，只包含 LFM 模块的模型记作 AAE_L，只包含 HRNet 模块的模型记作 AAE_H，同时包含 CSFM、LFM 的模型记作 AAE_CL,本文研究的模型记作 AH-AAE。具体的指标结果如表 6 所示，表 6 为以对抗型自编码为基线的各模块的消融实验。相关的图像修复结果如图 4.6 所示。

表 6 不同改进网络评价指标对比

模型	PSNR	UQI	FID	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
AAE	26.1804	0.9703	7.0524	0.8648	0.9228	0.9481
AAE_C	26.3890	0.9737	5.7244	0.8670	0.9333	0.9567
AAE_L	26.5524	0.9758	6.7586	0.8700	0.9300	0.9519
AAE_H	26.3475	0.9784	4.2384	0.8938	0.9482	0.9715
AAE_CL	31.0375	0.9908	1.3868	0.9602	0.9704	0.9808
AH-AAE	**33.7825**	**0.9928**	**1.1492**	**0.9748**	**0.9806**	**0.9931**

如表 6 所示，三个模块均可提高了基准模型的性能，但三者分别加入对基准

模型的提高效果并不明显。可以注意到，三模块组合的 AH-AAE 模型提升最多，说明引入注意力机制与 HRNet 后，图像复原能力得到了显著提高。同时 AH-AAE 在精度指标上得分显著提高也表明 AH-AAE 可以实现更准确和更真实的修复。



(a) AAE (b) AAE_C (c) AAE_L

(d) AAE_H (e) AAE_CL (f) AH-AAE

图 4.6 图像消融结果

通过对该模型进行消融实验后的具体修复效果如图 4.6 所示，可以看出各个模块的功能都很明显，其中包括：

（1）由（d）可知，AAE_C 能够捕捉到背景中的重要信息，并获取边缘信息和区别色彩的颜色的像素点，在左上方的部分，左上角放大部分中橙色羽毛区域出现轮廓，而在（C）中则是一个黑色的区域；

（2）AAE-L 在边缘提取方面略有优势，能较快地发现有明显特点的轮廓点；

（3）（f）为仅加入 HRNet 的深度学习网络，利用 HRNet 在图像特征抽取方面的优势，构建多尺度特征融合机制，在保证高空间分辨率的前提下，提高图像修复的鲁棒性与完备性。在图像修复方面， HRNet 也可以和解码器一起构成一个完整的图像修补框架。形成端到端的图像修复框架。这种端到端的结构可以

充分利用 HRNet 编码器提取的高质量特征，通过解码器对受损图像进行修复，从而得到更好的重建结果。(f)中羽毛的细节部分清晰可见；

（4）将 CSFM、LFM 进行结合，获得了（g）的模型修复图，与（d）、（e）结合观察，可以发现，尽管（g）的分辨率不高，但是二者相结合可以获得更加完善的填充结果，使掩模信息能够完整表达；

（5）在（g）的基础上，（h）展示了三个模块均保留的修复效果，可以发现（g）的填充效果较好，但就清晰度而言，（h）更清楚。HRNet 在保持高分辨率的同时，保留了图像的特征信息，减少了图像的损失。该算法能够更加细致地对图像的细节、纹理进行细致处理，从而最大限度地缩小与原图像的微小差别。

综上所述，当三个模块相互独立探测时，AH-AAE 模型的表现均会有所提高。但是，这三个模块组合在一起，可以发挥最优的效果。

4.5 本章小结

本章介绍了实验所用的数据集以及评价指标，并使用该数据进行的实验。进行的试验包括重构损失函数实验、多模型对比实验、噪声实验、特征实验以及消融实验。为了有效测试注意力机制模块对环境的感知能力，通过绘制输入注意力机制前后的深度图进行对比。这有助于验证是否通过添加注意力机制增强了环境感知能力。通过多模型对比证明了 AH-AAE 在图像修复方面的能力，并通过消融实验验证了各个模块的有效性。

5 总结与展望

5.1 总结

本文针对矩阵填充问题中的图像类数据填充问题展开研究，通过对图像问题深入剖析并借鉴深度学习在图像修补领域的发展趋势，设计了一种全新的矩阵填充模型应用于图像类数据填充，并通过仿真实验来验证该方法的有效性和优越性。本文主要研究成果总结为以下两点：

（1）本文以对抗自编码器为基础，将注意力机制与 HRNet 相结合，建立了 AH-AAE 矩阵填充模型。AH-AAE 主要针对图像类数据，对传统的神经网络填充模型进行了改进，有效克服了现有图像填充方法分辨率低、信息提取能力差、传输过程中信息丢失等问题。改进包括两个方面：1.使用 HRNet 作为生成器进行编码，提高了特征表达和图像信息提取能力。2.引入注意力机制，构造通道相似性融合模块和位置融合模块。通道相似性融合模块将相似性、稀疏注意力矩阵和密集注意力矩阵有机地结合在一起，扩大了感受野，提高了特征表示能力。位置融合模块则结合了位置编码和空间位置信息，不仅增强了模型的泛化能力，还显著提高了边界信息的获取能力。

（2）使用 MS-COCO 和 KITTI 数据集对提出的模型进行了实验，以测试模型的填充效果。结果表明，与其他常用算法相比，AH-AAE 模型在填充性能方面具有明显优势。通过消融实验检验各子模块对整体模型的提升效果，将各子模块整合到标准模型中，发现 AH-AAE 模型的每个子模块都能改善性能，但在三个模块组合的情况下，能达到更好的填充效果。除了填充性能，还进行了噪声实验来验证 AH-AAE 的去噪能力。从去噪实验可以看出，灰度图像的去噪难度要大于彩色图像，而 AH-AAE 模型对这两类图像的 PSNR 值都高于其他方法，说明 AH-AAE 虽然是填充模型，但在去噪领域也有很好的表现。在此基础上，对 AH-AAE 模型中注意力机制的两个特征模块进行了深入研究，大量实验表明，AH-AAE 模型的特征模块能够更好地反映场景的关键特征，提高去噪能力。

5.2 展望

本项目提出的 AH-AAE 模型能够有效地求解图像填充问题，但还存在以下不足之处：

（1）研究对象是图像，能否在保持基本结构不变的前提下，对其进行改进，使其适用于视频填充、声音填充和文字生成等领域；

（2）本文的重点是如何将 HRNet 与注意力机制相结合，使之更好地服务于该模型。在 2023 WAIC 年会的专题论坛上，合合信息关于 AI 影像内容安全的关键技术方案备受瞩目。其方案基于 HRNet 的“编码器-解码器”架构，融合图像自身的噪声、频谱等信息，实现对截图篡改轨迹的精确定位，并对生成的照片进行智能识别，从而有效地阻止非法获取照片中的信息，这种结构具有很高的分辨力。沿着此思路，思考对 AH-AAE 网络进行改进，能否提高模型的性能；

（3）当前 AH-AAE 模型结构仍有些烦琐，多层嵌套执行代码。HRNet 虽然具有很好的性能，但是其计算量很大，非常耗时。因此，对其进行改进以提高其计算效率也是一个需要进一步研究的问题。

参考文献

[1] Bello, I., Zoph, B., Vaswani, A., Shlens, J., Le, Q.V.. Attention augmented convolutional networks [C]. Proceedings of the IEEE/CVF international conference on computer vision, 2019: 3286-3295.

[2] Bodapati, J.D.. Stacked convolutional auto-encoder representations with spatial attention for efficient diabetic retinopathy diagnosis [J].Multimedia Tools and Applications, 2022, 81(22): 32033-32056.

[3] Buzau, M.-M., Tejedor-Aguilera, J., Cruz-Romero, P., Gomez-Exposito, A.. Hybrid deep neural networks for detection of non-technical losses in electricity smart meters[J].IEEE Transactions on Power Systems, 2019,35(2):1254-1263.

[4] Candes, E., Recht, B.. Exact matrix completion via convex optimization[J].Comm unications of the ACM, 2012,55(6):111-119.

[5] Chang, Y., Chen, J., Wu, W., Pan, T., Zhou, Z., He, S.. Intelligent fault quantitative identification for industrial Internet of Things (IIoT) via a novel deep dual reinforcement learning model accompanied with insufficient samples[J]. IEEE Internet of Things Journal, 2022,9(20):19811-19822.

[6] Chen, X., You, S., Tezcan, K.C., Konukoglu, E.. Unsupervised lesion detection via image restoration with a normative prior[J]. Medical image analysis, 2020,64(1):101713.

[7] Choi, E., Biswal, S., Malin, B., Duke, J., Stewart, W.F., Sun, J.. Generating multi-label discrete patient records using generative adversarial networks[C]. Machine learning for healthcare conference. PMLR, 2017: 286-305.

[8] Demir, U., Unal, G..Patch-based image inpainting with generative adversarial networks[J]. arXiv preprint arXiv:1803.07422, 2018.

[9] Deng, X., Dragotti, P.L.. Deep convolutional neural network for multi-modal image restoration and fusion[J]. IEEE transactions on pattern analysis and machine intelligence, 2020,43(10):3333-3348.

[10] Dong, C., Loy, C.C., He, K., Tang, X.. Learning a deep convolutional network for image super-resolution[C]. Computer Vision–ECCV 2014: 13th European Confere

nce, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13. Springer, 2014:184-199.

[11] Dong, Q., Cao, C., Fu, Y.. Incremental transformer structure enhanced image inpainting with masking positional encoding[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022:11358-11368.

[12] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S.. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arxiv preprint arxiv:2010.11929, 2020.

[13] Edelman, B.L., Goel, S., Kakade, S., Zhang, C.. Inductive biases and variable creation in self-attention mechanisms [C]. International Conference on Machine Learning. PMLR, 2022: 5793-5831.

[14] Fukui, H., Hirakawa, T., Yamashita, T., Fujiyoshi, H.. Attention branch network: Learning of attention mechanism for visual explanation [C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019: 10705-10714.

[15] Gao, S., Zhou, C., Ma, C., Wang, X., Yuan, J.. Aiatrack: Attention in attention for transformer visual tracking [C]. Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXII. Springer, 2022: 146-164.

[16] Gu, S., Xie, Q., Meng, D., Zuo, W., Feng, X., Zhang, L.. Weighted nuclear norm minimization and its applications to low level vision [J]. International journal of computer vision, 2017, 121: 183-208.

[17] Guo, M.-H., Xu, T.-X., Liu, J.-J., Liu, Z.-N., Jiang, P.-T., Mu, T.-J., Zhang, S.-H., Martin, R.R., Cheng, M.-M., Hu, S.-M.. Attention mechanisms in computer vision: A survey[J]. Computational visual media, 2022, 8(3): 331-368.

[18] Huang, F., Li, X., Yuan, C., Zhang, S., Zhang, J., Qiao, S.. Attention-emotion-enhanced convolutional LSTM for sentiment analysis[J]. IEEE transactions on neural networks and learning systems, 2021,33(9): 4332-4345.

[19] Jifara, W., Jiang, F., Rho, S., Cheng, M., Liu, S.. Medical image denoising using

convolutional neural network: a residual learning approach[J]. The Journal of Supercomputing, 2019,75(1): 704-718.

[20] Kadurin, A., Nikolenko, S., Khrabrov, K., Aliper, A., Zhavoronkov, A.. druGAN: an advanced generative adversarial autoencoder model for de novo generation of new molecules with desired molecular properties in silico [J]. Molecular pharmace utics, 2017, 14(9): 3098-3104.

[21] Kappeler, A., Yoo, S., Dai, Q., Katsaggelos, A.K.. Video super-resolution with convolutional neural networks[J]. IEEE transactions on computational imaging 2016, 2(2): 109-122.

[22] Krizhevsky, A., Sutskever, I., Hinton, G.E.. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017,60(6): 84-90.

[23] Lavie, N., Hirst, A., De Fockert, J.W., Viding, E.. Load theory of selective attention and cognitive control[J]. Journal of experimental psychology: General, 2004,133 (3): 339-354.

[24] Li, Y., Swersky, K., Zemel, R.. Generative moment matching networks [C]. International conference on machine learning. PMLR, 2015: 1718-1727.

[25] Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K.. Enhanced deep residual networks for single image super-resolution [C]. Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2017:136-144.

[26] Liu, D., Wen, B., Fan, Y., Loy, C.C., Huang, T.S..Non-local recurrent network for image restoration [J]. Advances in neural information processing systems, 2018, 31.

[27] Liu, P., Wang, M., Wang, L., Han, W.. Remote-sensing image denoising with multi-sourced information[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2019,12(2): 660-674.

[28] Liu, X., Suganuma, M., Sun, Z., Okatani, T.. Dual residual networks leveraging the potential of paired operations for image restoration[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 7007-7016.

[29] Liu, Y., Li, H., Guo, Y., Kong, C., Li, J., Wang, S.. Rethinking attention-model explainability through faithfulness violation test[C]. International Conference on Machine Learning. PMLR, 2022: 13807-13824..

[30] Ma, X., Zhou, C., Kong, X., He, J., Gui, L., Neubig, G., May, J., Zettlemoyer, L., 2022. Mega: moving average equipped gated attention [J]. arxiv preprint arxiv:2209.10655, 2022.

[31] Makhzani, A., Shlens, J., Jaitly, N., Goodfellow, I., Frey, B., 2015. Adversarial autoencoders [J]. arXiv preprint arXiv:1511.05644, 2015.

[32] Mao, X.-J., Shen, C., Yang, Y.-B., 2016. Image restoration using convolutional auto-encoders with symmetric skip connections [J]. arXiv preprint arXiv:1606.08921, 2016.

[33] Mei, K., Jiang, A., Li, J., Wang, M.. Progressive feature fusion network for realistic image dehazing [C]. Computer Vision–ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers, Part I 14. Springer International Publishing, 2019: 203-215.

[34] Nie, F., Wang, H., Huang, H., Ding, C.. Joint Schatten p-norm and ℓ p-norm robust matrix completion for missing value recovery [J]. Knowledge and Information Systems, 2015, 42(3): 525-544.

[35] Niu, Z., Zhong, G., Yu, H.. A review on the attention mechanism of deep learning[J]. Neurocomputing, 2021, 452: 48-62.

[36] Paik, J.K., Katsaggelos, A.K.. Edge detection using a neural network [C]. International Conference on Acoustics, Speech, and Signal Processing. IEEE , 1990: 2145-2148.

[37] Paik, J.K., Katsaggelos, A.K.. Image restoration using a modified Hopfield network [J]. IEEE Transactions on image processing, 1992, 1(1): 49-63.

[38] Pan, R., Yang, T., Cao, J., Lu, K., Zhang, Z.. Missing data imputation by K nearest neighbours based on grey relational structure and mutual information [J]. Applied Intelligence, 2015, 43: 614-632.

[39] Park, B., Yu, S., Jeong, J.. Densely connected hierarchical network for image denoising[C]. Proceedings of the IEEE/CVF conference on computer vision and

pattern recognition workshops(CVPRW) , 2019:2104–2113.

[40] Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., Efros, A.A.. Context encoders: Feature learning by inpainting[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 2536-2544.

[41] Peng, Z., Huang, W., Gu, S., Xie, L., Wang, Y., Jiao, J., Ye, Q.. Conformer: Local features coupling global representations for visual recognition[C]. Proceedings of the IEEE/CVF international conference on computer vision, 2021:367-376.

[42] Seo, M., Kembhavi, A., Farhadi, A., Hajishirzi, H.. Bidirectional attention flow for machine comprehension[J]. arXiv preprint arXiv:1611.01603, 2016.

[43] Shaw, P., Uszkoreit, J., Vaswani, A.. Self-attention with relative position representations[J]. arXiv preprint arXiv:1803.02155, 2018.

[44] Shazeer, N., Mirhoseini, A., Maziarz, K., Davis, A., Le, Q., Hinton, G., Dean, J.. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer[J]. arXiv preprint arXiv:1701.06538, 2017.

[45] Simonyan, K., Zisserman, A.. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[46] Springenberg, J.T.. Unsupervised and semi-supervised learning with categorical generative adversarial networks[J]. arXiv preprint arXiv:1511.06390, 2015.

[47] Srinivas, M., Patnaik, L.M., 1994. Adaptive probabilities of crossover and mutation in genetic algorithms[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1994,24(4): 656-667.

[48] Suvorov, R., Logacheva, E., Mashikhin, A., Remizova, A., Ashukha, A., Silvestrov, A., Kong, N., Goka, H., Park, K., Lempitsky, V.. Resolution-robust large mask inpainting with fourier convolutions[C]. Proceedings of the IEEE/CVF winter conference on applications of computer vision, 2022: 2149-2159.

[49] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.. Going deeper with convolutions[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2015:1-9.

[50] Tian, C., Xu, Y., Zuo, W.. Image denoising using deep CNN with batch renormalization[J]. Neural Networks, 2020,121(1): 461-473.

[51] Tonioni, A., Tosi, F., Poggi, M., Mattoccia, S., Stefano, L.D.. Real-time self-adaptive deep stereo[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 195-204.

[52] Ulyanov, D., Vedaldi, A., Lempitsky, V.. Deep image prior[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018: 9446-9454.

[53] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017: 5998–6008.

[54] Venetianer, P.L., Werblin, F., Roska, T., Chua, L.O.. Analogic CNN algorithms for some image compression and restoration tasks[J]. IEEE transactions on circuits and systems I: Fundamental theory and applications, 1995,42(5): 278-284.

[55] Vig, J.. A multiscale visualization of attention in the transformer model[J]. arXiv preprint arXiv:1906.05714, 2019.

[56] Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Tao, A., Kautz, J., Catanzaro, B.. High-resolution image synthesis and semantic manipulation with conditional gans[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018: 8798-8807.

[57] Wu, B., Xu, C., Dai, X., Wan, A., Zhang, P., Yan, Z., Tomizuka, M., Gonzalez, J., Keutzer, K., Vajda, P.. Visual transformers: Token-based image representation and processing for computer vision [J]. arXiv preprint arXiv:2006.03677, 2020.

[58] Wu, H., Wu, J., Xu, J., Wang, J., Long, M.. Flowformer: Linearizing transformers with conservation flows[J]. arXiv preprint arXiv:2202.06258, 2022.

[59] Xia, Z., Chakrabarti, A.. Identifying recurring patterns with deep neural networks for natural image denoising [C]. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2020: 2426-2434.

[60] Xie, Y., Gu, S., Liu, Y., Zuo, W., Zhang, W., Zhang, L.. Weighted Schatten $ p $-norm minimization for image denoising and background subtraction [J]. IEEE transactions on image processing, 2016,25(10): 4842-4857.

[61] Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., Zemel, R., Bengio, Y. Show, attend and tell: Neural image caption generation with visual

attention, International conference on machine learning [C]. PMLR, 2015: 2048-2057.

[62] Xu, X., Tao, Z., Ming, W., An, Q., Chen, M.. Intelligent monitoring and diagnostics using a novel integrated model based on deep learning and multi-sensor feature fusion [J]. Measurement, 2020,165(1): 108086.

[63] Yan, C., Tu, Y., Wang, X., Zhang, Y., Hao, X., Zhang, Y., Dai, Q.. STAT: Spatial-temporal attention mechanism for video captioning[J]. IEEE transactions on multimedia, 2019,22(1): 229-241.

[64] Yuan, D., Chang, X., Huang, P.-Y., Liu, Q., He, Z.. Self-supervised deep correlation tracking [J]. IEEE Transactions on Image Processing, 2020, 30(1): 976-985.

[65] Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X., Metaxas, D.N.. Stackgan++: Realistic image synthesis with stacked generative adversarial networks [J]. IEEE transactions on pattern analysis and machine intelligence, 2018, 41(8): 1947-1962.

[66] Zhang, J., Zhang, Y., Gu, J., Zhang, Y., Kong, L., Yuan, X.. Accurate image restoration with attention retractable transformer [J]. arXiv preprint arXiv:2210.01427, 2022.

[67] Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising [J]. IEEE transactions on image processing, 2017,26(7): 3142-3155.

[68] Zhang, Q., Zhang, X., Mu, X., Wang, Z., Tian, R., Wang, X., Liu, X.. Recyclable waste image recognition based on deep learning [J]. Resources, Conservation and Recycling, 2021,171(1): 105636.

[69] Zhang, S.. Nearest neighbor selection for iteratively kNN imputation [J]. Journal of Systems and Software, 2012,85(11): 2541-2552.

[70] Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.. Image super-resolution using very deep residual channel attention networks [C], Proceedings of the European conference on computer vision (ECCV), 2018: 286-301.

[71] Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y.. Residual dense network for image restoration [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,

2020,43(7): 2480-2495.

[72] Zhao, H., Gallo, O., Frosio, I., Kautz, J. Loss functions for image restoration with neural networks [J]. IEEE Transactions on computational imaging, 2016, 3(01): 47-57.

[73] Zhou, D., Yu, Z., Xie, E., Xiao, C., Anandkumar, A., Feng, J., Alvarez, J.M.. Understanding the robustness in vision transformers [C], International Conference on Machine Learning. PMLR, 2022: 27378-27394.

[74] Zhou, Y.-T., Chellappa, R., Vaid, A., Jenkins, B.K.. Image restoration using a neural network [J]. IEEE transactions on acoustics, speech, and signal processing, 1988,36(7): 1141-1151.

[75] Zhu, X., Cheng, D., Zhang, Z., Lin, S., Dai, J., 2019. An empirical study of spatial attention mechanisms in deep networks [C], Proceedings of the IEEE/CVF international conference on computer vision. 2019: 6688-6697.

[76] 金勇进, 朱琳. 不同差补方法的比较 [J].数理统计与管理, 2000, (04): 50-54.

[77] 曲志坚, 张先伟, 曹雁锋, 刘晓红, 冯晓华. 基于自适应机制的遗传算法研究 [J]. 计算机应用研究, 2015, 32 (11): 3222-3225+3229.

[78] 谭丹丹, 曹斌, 刘俊, 姜风华, 王亚.一种基于配对的改进遗传算法[J]. 计算机应用, 2007, (S2): 170-171.

[79] 王小平, 曹立明. 遗传算法：理论、应用与软件实现[M].西安：西安交通大学出版社，2002: 235-251.

[80] 张网娟, 许国艳, 李敏佳, 朱帅. 基于卷积神经网络的缺失数据填充方法 [J]. 微电子学与计算机, 2019, 36 (03): 48-52+57.

攻读硕士学位期间承担的科研任务及主要成果

主要学术论文情况

[1] **黄梓玉**,钱崇辉,黄恒君 . 一种基于注意力机制的对抗型自编码器图像修复模型 [J]. 湖北民族大学学报(自然科学版), 2024, 42 (01): 81-85+91. DOI:10.13501/j.cnki.42-1908/n.2024.03.013

主要参与科研项目情况

[1] 面向城市计算的多领域数据融合方法研究，国家社会科学基金项目。

[2] 城市计算方法体系构建及甘肃智慧城市应用，中央引导地方科技发展项目。

[3] 基于数据融合的相对贫困测度及识别方法研究及应用，甘肃省软科学专项课题。

致 谢

花开花落万物道，聚散离别终有时。行文至此，百感交集，从一个怀揣理想的追梦少年，一路跌跌撞撞走到现在，些许遗憾，些许不舍，些许憧憬，对未来满怀热爱的我又要开始新的征程。回首在兰州财经大学的这三年，感恩每个从我身边出现的人，正是有你们的善意和陪伴，才拼凑出我对这段旅程的不舍与热爱。祝福你们的未来如鲜花般灿烂，回首过往，不忘曾全力以赴的自己。

桃李不言，下自成蹊。感谢我的指导老师黄恒君老师。从论文的选题到最终成文，感谢您陪我字斟句酌，倾尽所能的点播和指导我。三年的求学之路，不论。是传道授业、未来规划还是生活琐事，感谢您的体谅、包容与关爱。回首一封封邮件，一条条语音，这些点点滴滴记录着您的心血：也是您在学术研究上对我的辛勤栽培和耐心教导，让我端正了学习态度，养成了优秀的研学习惯，面对未来的考验，愈发严谨、稳重、坚韧。感谢学院的每一位呕心沥血的教授，给予我们追逐理想的勇气。

焉得谖草,言树之背。感谢我的父母，感谢你们见证了我的成长，教会我正直、真诚的对待别人，谢谢你们一路以来默默的陪伴，在我最困难的时候，给我肩膀依靠。你们是我前进路上最大的底气，有你们，我不怕输的一无所有。也谢谢你们，把最好的都给了我。

山水一程，三生有幸。感恩师门所有的兄弟姐妹们和研究生生涯中认识的朋友们。谢谢你们到过我的世界，是你们陪我成长，陪我走过这一程，让我在学习工作之余感受到温暖和快乐。感谢你们在研究过程中的合作、帮助和支持，我们相互鼓励、相互学习，共同度过了研究生生涯中的困难和挑战。美好的时光总是很短很短，真心祝福即将远行的你们前程似锦，取得更多的辉煌。

追风赶月末停留，平芜尽处是春山。在我 26 岁的这一年，希望我可以接受每一个时间段的自己，始终向上、始终热烈、始终勇敢的自己。关关难过关关过，前路漫漫亦灿烂，祝我永远向前看，始终努力向上，奔赴下一片山海。

行文至此，谨向每一位在我求学路上支持和鼓励过我的人表达我最真挚的感谢！