

## 深度学习的二维动画视觉领域修复综述

李雨航<sup>1</sup>, 谢良彬<sup>1,2</sup>, 董超<sup>1,3+</sup>

1. 中国科学院深圳先进技术研究院 机器视觉与模式识别重点实验室, 广东 深圳 518000

2. 澳门大学科技学院, 澳门 999078

3. 上海人工智能实验室, 上海 200000

+ 通信作者 E-mail: liyuhang\_98@163.com

**摘要:**传统二维动画是一种具有独特风格的影像作品,其制作方式和画面特征与现实场景的影像存在很大差异。传统二维动画通常需要逐帧绘制图片,并保存为位图。但是二维动画在存储、传输和播放过程中,可能会出现画质退化、分辨率不足、时序不连续等问题。随着深度学习技术的发展,基于深度学习的方法在动画修复领域得到了广泛应用。对基于深度学习的二维动画修复进行了全面总结。首先,调研了目前与动画相关的数据集,明确了深度学习在动画修复领域已有的数据支撑以及动画数据集建立的瓶颈。其次,通过调研和测试动画画质修复与动画视频插帧的深度学习算法,找出了目前动画修复的关键点与难点,介绍了一些针对动画画面特征设计的保证动画帧间一致性的方法,为动画视频的修复提供了思路。然后,通过分析现有图像质量评价方法对动画影像的有效性,找到能合理评价和指导动画修复结果的指标。最终,根据上述分析,阐明了动画相关修复任务中面临的挑战,并提出了未来深度学习在动画修复领域的发展方向。

**关键词:**动画修复;深度学习;超分辨率;动画插帧

**文献标志码:**A **中图分类号:**TP391.4

## Review of 2D Animation Restoration in Visual Domain Based on Deep Learning

LI Yuhang<sup>1</sup>, XIE Liangbin<sup>1,2</sup>, DONG Chao<sup>1,3+</sup>

1. Shenzhen Key Lab of Computer Vision and Pattern Recognition, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong 518000, China

2. Faculty of Science and Technology, University of Macau, Macau 999078, China

3. Shanghai Artificial Intelligence Laboratory, Shanghai 200000, China

**Abstract:** Traditional 2D animation is a distinct visual style with a production process and image characteristics that differ significantly from real-life scenes. It usually requires drawing pictures frame by frame and saving them as bitmaps. During the storage, transmission, and playback process, 2D animation may encounter problems such as picture quality degradation, insufficient resolution, and discontinuous timing. With the development of deep learning technology, it has been widely used in the field of animation restoration. This paper provides a comprehensive summary of 2D animation restoration based on deep learning. Firstly, exploring existing animation datasets can help identify the available data support and the bottleneck in establishing animation datasets. Secondly, investigating and testing deep learning-based algorithms for animation image quality restoration and animation interpolation can help identify key points and challenges in animation restoration. Additionally, introducing methods designed to ensure consistency between animation frames can provide insights for future animation video restoration. Analyzing the effectiveness of existing image quality assessment (IQA) methods for animation images can help identify practical

IQA methods to guide restoration results. Finally, based on the above analysis, this paper clarifies the challenges in animation restoration tasks and presents future development directions of deep learning in animation restoration field.

**Key words:** animation restoration; deep learning; super-resolution; animation interpolation

二维动画是一种风格独特的影像作品,其制作方式、画面特征和表现内容与记录真实场景的影像存在明显差异。传统二维动画在制作方式上具有独特的特点。在动画制作的前期阶段,通常设定具有超脱于真实世界的表达,具有高度的假设性和象征性。在动画制作的中期阶段,动画师根据设定绘制关键帧,并根据相邻关键帧的动作幅度、轨迹和形变状态合理补充流畅的中间帧。在这个过程中,动画师通常使用纸张或借助软件逐步绘制草图和清晰的线稿,最后进行上色。通常情况下,背景和角色是分开绘制的,因此可以在画面风格和播放速度等方面分别处理前景和背景。最后,将背景和角色合成,并添加后期效果,使用摄影设备或剪辑软件通常以每秒24帧的速度制作成动画影片,以确保观感上没有卡顿的现象。

制作完成后的二维动画的画面通常由清晰的边缘线条和平滑的色彩区域组成,并带有一些手绘笔触或艺术风格。二维动画中不存在真实的光源,对于光影的处理通常是模仿现实场景或根据画面风格进行艺术加工。画面背景与前景人物往往具有明显的虚实关系。因此,动画中常常出现明确的明暗交界线和独特的细节纹理,以及明显的虚实关系。在修复过程中,如何更好地保留动画的风格特点和虚实关系成为二维动画修复的难点和关键所在。

现有的动画影像受到多种因素的影响,从制作、保存到传输过程,可能会出现画质退化的现象,如噪声、模糊、压缩损失和色彩失真等。随着放映设备的更新换代,一些早期动画作品在高清屏幕上的效果并不理想。目前,日本的动画作品通常以720P的分辨率制作,而观众在视频网站上看到的1080P甚至更高分辨率的作品通常是通过算法进行插值拉伸得到的。这是因为目前动画制作仍未完全数字化,直接制作高分辨率动画成本高昂,并且原始数据过大不便于传输和保存。此外,动画播放过程中可能出现画面卡顿和不流畅的问题。因此,对于旧动画影像的修复工作以及对新动画影像的超分辨率处理是非常必要的,研究者需要从实际应用的角度进行分析和探索。

在传统工业界中针对动画画质问题(例如噪声、

模糊、压缩损失、色彩失真等)会使用一些传统算法。针对噪点问题,常使用滤波方法,包括中值滤波、均值滤波、高斯滤波等。然而,这些去噪方法只适用于特定类型的噪声,并且参数调整较为困难。对于画面模糊问题,常使用逆滤波方法、模拟模糊函数方法以及基于稀疏表示的去模糊方法。逆滤波方法通过傅里叶变换将模糊的图像转换到频域,再通过对频域图像进行滤波和反变换,恢复出清晰的图像。模拟模糊函数方法需要模拟多种模糊核及相关参数,以找到最佳的去模糊效果。而基于稀疏表示的去模糊方法则将模糊后的图像分解为稀疏表示和低通部分,从而得到清晰的图像。总体来说,传统的去模糊方法具有较强的局限性,无法应对复杂情况。针对压缩损失问题,工业界通常在动画传输之前进行预处理操作,如去噪、降低分辨率等,以减少图像数据量,从而降低传输过程中的压缩损失。另外,还可以选择更适合的压缩算法,如JPEG、H.264、H.265,或采用自适应压缩技术,在网络带宽等条件下自动选择合适的压缩比例和算法。在传输过程中,可以使用异常检测技术来检测传输错误、丢包、带宽不足等问题,并对已经损失的数据进行修复。对于已经受到压缩损失的动画数据恢复,常常使用降噪、去模糊、边缘增强等方法来恢复原始数据的细节和信息。然而,在数据损失较大的情况下,传统的降噪和去模糊方法可能会损失大量有效信息,从而影响观感。对于颜色失真问题,工业界也经常使用诸如色彩平衡、色调、饱和度、曲线等工具来调整图像或视频的颜色、亮度和对比度等参数。然而,传统的色彩校正方法通常只能对整个图像或视频进行调整,无法进行精细的局部调整。通过上述对传统工业界处理方法的分析可以看出,在面对不同的质量问题时,通常需要分成多个步骤来解决,并且需要调整和控制大量的参数,操作繁琐,解决问题的能力有限。

对于插帧任务,在传统工业界中有两种处理路线:第一种是根据前后帧图像进行线性插值;第二种是进行光流估计,即通过估计相邻帧之间的运动矢量生成新的中间帧。线性插值方法是通过对两个已知帧之间的像素值进行加权平均来计算新的像素

值,这种传统方法存在明显的模糊和失真。光流估计方法则需要进行像素间的匹配,通过对图像中某些特征的比较来匹配关键点,并计算出这些点之间的位移向量。相比传统的帧间线性插值,光流估计方法的效果更好。然而,在处理复杂场景时,目前仍存在一些挑战,效果不尽如人意。

近年来,深度学习方法在视频超分辨率、去噪、去模糊、光流估计以及视频插帧等领域快速发展,为动画影像修复工作提供了更多选择。

对于动画画质修复问题,既可以将画质问题拆分开来解决,也可以将动画影像修复理解作为一种盲修复任务。因为画质受损的动画往往具有复杂的退化情况,很难拆解成明确的子问题。在有监督的深度学习,根据动画作品真实的退化状况准备数据集,同时结合动画的特点进行模型构建,在模型学习的过程中要利用好蕴含在动画影像中的空间信息与时序信息。此外,尽可能端到端地解决动画画质问题,避免在传统工业界中多步骤处理影像。并且从真实的低质量动画场景中学习,能够增强深度学习网络的处理能力。在动画影像修复中,要在保留画面原有风格的基础上提高画面质量,达到更好、更自然的效果。

基于深度学习的动画画质修复方法,往往需要大量高质量和低质量的动画影像数据对。使用传统的下采样插值方法(例如双线性插值、双三次插值等)、模糊核(例如各向同性或各向异性高斯模糊)和随机噪声(例如椒盐噪声、高斯噪声、泊松噪音)等经典的退化方法,很难近似拟合出逼近真实场景的低质量动画,使得数据集的构建变得困难。

对于动画插帧任务,传统的光流估计算法通常依赖于人为设计的特征和约束,在复杂场景中处理困难。随着深度学习的发展,可以通过神经网络学习出更丰富、更准确的特征表示,从而改善光流估计的性能。基于深度学习的光流估计算法有更灵活的优化策略,可通过定义不同的损失函数来引入不同的约束条件。通常包括亮度约束项和平滑性约束项,用于保持光流的一致性和平滑性。这样可以提高算法的鲁棒性,更好地处理复杂场景,并获得更准确的光流估计结果。

但是二维动画数据缺乏精确的光流信息,在制作过程中很难进行矢量标定,从而阻碍了一些利用光流信息进行插帧的方法在动画领域的应用。

综上所述,二维动画影像具有独特的制作方法

与画面特征,因此其明显区别于现实场景的影像。传统工业界常使用一些传统算法来修复动画作品,但这些方法存在很大的局限性。随着深度学习的引入,动画影像修复方法得到了显著改进。然而,深度学习方法常受到数据集的限制,且针对动画自身特点的画质修复、插帧等算法以及有效的评价指标仍需要研究者们进行挖掘和探索。

本文将从数据集、动画画质修复、动画时序修复、各算法结果以及评价指标的有效性这几个方面进行介绍和分析,全面总结深度学习方法在动画影像修复领域的应用情况,并最终指出动画影像修复领域的现存问题和发展前景。

## 1 二维动画相关数据集

根据动画数据的制作方式,本文将动画数据集划分为真实二维动画影像数据集和利用三维软件构建的二维动画数据集。真实二维动画影像通常由画师逐帧绘制,不包含光流、深度等矢量信息。真实二维动画影像数据集主要由从真实动画影片中截取的动画图片和从插画网站收集的插画图片组成。而利用三维软件构建的二维动画数据集可以利用三维软件进行批量渲染生成,能够准确标定每一帧画面中各个物体的矢量信息。

### 1.1 真实二维动画影像数据集

在动画画质修复任务中,通常需要使用从真实动画影片中收集的数据进行学习。有监督的深度学习需要大量成对的数据集,为了生成高质量的动画数据,可以通过特定的退化方式生成相应的低质量动画数据。其中AVC(animation video clip)<sup>[1]</sup>是一个大规模高质量的真实动画影片剪辑数据集,每个剪辑片段包含100张连续的动画图片。使用AVC-RealLQ(从网络上下载的44个真实低质量动画片段),通过视频超分辨率网络选择出高质量修复图片,从而形成数据对,再利用经典的退化模块和卷积层组成的退化网络进行训练。这种可学习的降质策略有助于缩小合成数据与真实数据之间的分布差距。

在动画影片的插帧任务中,ATD-12K<sup>[2]</sup>提供了一个在网络上收集的大型动画数据集,从30部不同风格的动画电影中挑选出了12 000个三联组。这些三联组的中间帧可作为算法插帧结果的监督图像。此外,ATD-12K的测试集根据运动幅度和遮挡面积划分了插帧难度级别,并提供了感兴趣区域(region of interest, RoI)和运动类别的标签。

除了针对动画超分和插帧的真实动画影片数据集,还有一些在插画网站上整理得到的插画数据集。Danbooru<sup>[3]</sup>是一个持续更新动画人物插画的网站,拥有百万级的插画数据和大量粗糙标签,但可以通过标签筛选来获取所需的相关数据,进而整理出可用的数据集。DanbooRegion<sup>[4]</sup>包含大量的动画插图与其对应的区域分割图,可以应用于动画区域追踪、线稿上色、动画图像内容分解等方向。nico-illst<sup>[5]</sup>包含50万张插图,注释标签包括作者、主题和图像的风格(水彩、铅笔、丙烯酸等)等信息。

对于动画角色的脸部识别,DAF:re<sup>[6]</sup>基于DanbooruAnimeFaces(DAF)动画人脸数据集,保留角色名称的标签,并标注出头部边界框,主要用于动画角色识别的相关研究。iCartoonFace<sup>[7]</sup>是目前规模最大的动画人物识别数据集,由40万张以上高质量图片组成,包含5013个动画人物。LSASRD<sup>[8]</sup>是一个大规模的动画风格识别数据集,包含了190部动画作品,每个作品包含10个以上的角色。不同人物的头像会被裁剪出来,并标记好年份、性别、地区、种族等信息。

## 1.2 利用三维软件构建的二维动画数据集

Blender<sup>[9]</sup>作为一款开源的三维图形图像软件,被广泛应用于动画制作和数据生成。它可以解决传统二维动画逐帧绘制繁琐及无法进行矢量标定的问题。通过“三渲二”技术来建立动画数据集,即模拟二维动画的风格和光源将三维模型渲染成具有二维动画风格的图片。但目前“三渲二”技术尚不成熟,与逐帧绘制的动画影片具有一定差距。因此利用三维软件构建的二维动画数据集不常用于动画画质修复任务,而更常用于动画插帧这类需要矢量信息辅助的任务。

Creative Flow+Dataset<sup>[10]</sup>使用Blender软件,在三维动画场景中获取每个物体在每个时刻的真实准确位置坐标,并创建了一个多样化的多风格艺术视频数据集。该数据集包含每个像素的光流、遮挡、分割标签、法线和深度的准确信息。数据集包括3000个动画序列,使用了40个纹理线样式和38个阴影样式随机选择进行渲染。然而,每个样本都是由简单的预定义运动驱动单一模型生成的,因此无法模拟真实的二维动画所需的运动和模型的复杂度。

AnimeRun<sup>[11]</sup>选择了三部Blender开源的三维动画影片,将其渲染成二维动画风格,同时生成相应的光流及深度信息。二维动画风格画面由明晰的边缘线条和纯色色块组成,尚未添加阴影部分以及复杂光源和特效,与真实二维动画仍存在差距。但相较

于Creative Flow+Dataset,AnimeRun包含能够移动的背景、多个模型间丰富的交互以及更复杂的运动。

AnimeCeleb<sup>[12]</sup>提供了240万张图片,并标注了相应的详细姿态信息。该数据集首先在网站上收集三维动画模型,通过改变模型的三维属性获得更多的表情与姿态,然后在Blender软件上渲染成二维动画风格的人物图像,再使用注释系统进行标注。

## 2 动画修复的相关算法

### 2.1 动画画质修复相关算法

在单张动画图片的修复方面,Anime4K<sup>[13]</sup>是一个开源的高质量实时动画超分去噪方法,早期的v1.0版本没有使用深度学习方法,而是考虑清晰的边缘对动画视觉效果的重要性。该方法首先使用双三次插值将图像放大到目标大小,然后根据周围像素的亮度和梯度信息对每个像素点进行优化。后面更新的版本中尝试使用深度学习方法,其官方提供了一个基于Tensorflow框架,参数量仅为2548的卷积神经网络模型。Anime4K的重心倾向于实时优化动画画质,更注重模型大小和运行时间,并有多种模式可以相互组合使用,以达到更好的结果。

目前有许多方法通过学习大量动画数据来提高修复效果。Waifu2x<sup>[14]</sup>最初基于SRCNN(super-resolution convolutional network)<sup>[15]</sup>网络,并应用于动画图像的超分和去噪工作。其提供了一个7层卷积神经网络模型,并通过反卷积的方法对图像进行上采样;随后更新了一个由两个U-Net结构<sup>[16]</sup>组成的UpCunet2X模型。目前最新版本的Waifu2x尝试使用Transformer结构,引入了5个SwinTransformer-Block<sup>[17]</sup>,构建了UpSwinUNet网络。以上Waifu2x的所有方法均提供了无降噪版与降噪版。

Real-CUGAN(real cascade U-Nets generative adversarial network)<sup>[18]</sup>在Waifu2x的UpCunet2X模型版本的基础上,利用生成对抗方法在百万级未公开的动画数据上进行训练。

Real-ESRGAN(Anime6B)<sup>[19]</sup>是Real-ESRGAN(real-world enhanced super-resolution generative adversarial network)的动画修复版本,是在原模型的基础上将残差稠密块(residual in residual dense block, RRDB)的数量减少至6个,并输入大量动画数据集进行训练,使其更好地应用于动画图像的修复。单张动画图片修复网络的发展路线如图1所示。

在动画视频超分辨率方面,Real-ESRGAN提供



图1 单张动画图片修复网络发展路线

Fig.1 Single animation image restoration method development line

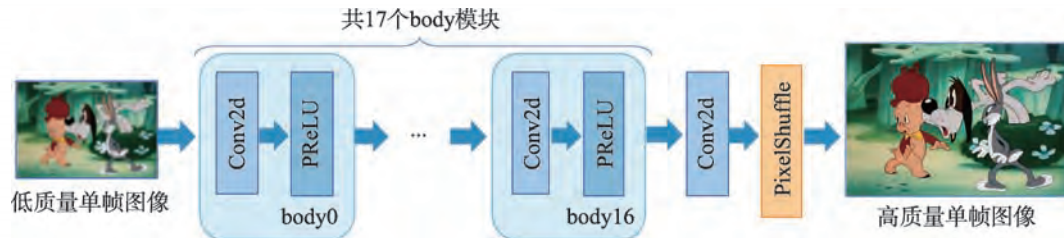


图2 AnimeVideo-v3 模型结构

Fig.2 AnimeVideo-v3 model structure

了一个名为 AnimeVideo-v3<sup>[20]</sup>的动画视频模型,该模型结构如图2所示。该方法以单帧作为输入,由17个串联的卷积模块组成,未利用帧间的时序信息。而 AnimeSR<sup>[1]</sup>是专为动画设计的深度学习网络,考虑了动画影像的特征。该方法使用单向循环网络和基于滑动窗口的结构,可以将多帧图像作为输入,并充分利用连续帧中蕴藏的时序信息。循环块内部采用多尺度设计,充分融合不同尺度下的特征,帮助网络理解画面内容,优化修复,该模型结构如图3所示。在第3.2节中,通过测试和比较这两种动画视频超分辨率网络,来验证 AnimeSR 结构设计的有效性。

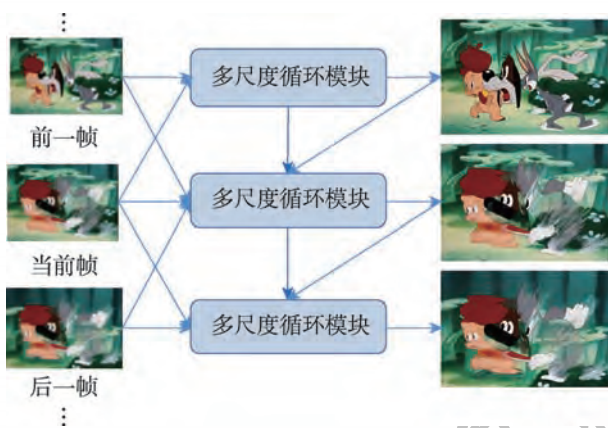


图3 AnimeSR 模型结构

Fig.3 AnimeSR model structure

## 2.2 动画时序修复相关算法

### 2.2.1 动画插帧算法

AnimeInterp<sup>[2]</sup>首次正式地定义和研究了专门针

对动画的插帧问题,其具体结构如图4所示。针对经典的二维动画具有清晰线条和单一颜色填充的封闭区域的特点,该方法首先利用相邻帧中色块对应匹配生成粗糙的光流(segment-guided matching);然后使用经过训练的循环结构光流细化模块(recurrent flow refinement)进一步优化粗糙的光流,以弥补二维动画中光流信息的缺失;最后采用多尺度的卷积神经网络进行扭曲变换合成中间帧。当时,AnimeInterp没有可用的二维动画光流数据进行学习,因此可以巧妙地利用二维动画中色块平坦和线条清晰的特点来模拟光流信息以进行插帧。然而,在第3.3节的测试中,发现这种不精确的光流会导致插帧结果出现画面模糊和边界不清的问题。

在 AnimeInterp 之后,Chen 等人<sup>[2]</sup>为进一步提升动画插帧的画面质量,提出了 SSL 模块(Softsplat-Lite)。该模块采用前向扭曲结构进行特征提取、光流预测和图像融合。为了更好地利用二维动画的线条信息以获得更清晰的色块区域,作者设计了 DTM 模块(distance transform module)。该模块首先通过边缘检测提取图像的二值化线稿;然后利用欧几里德距离变换(normalized Euclidean distance transform, NEDT)处理前后参考帧与 SSL 输出结果的线稿,并通过训练来估计中间输出帧的 NEDT;最后将估计得到的中间输出帧的 NEDT 与 SSL 模块的结果进行融合,通过训练得到优化后的中间输出帧。以上训练涉及的两部分分别进行处理,整体流程图如图5所示。

经过对这两种插帧网络进行测试,可以发现插

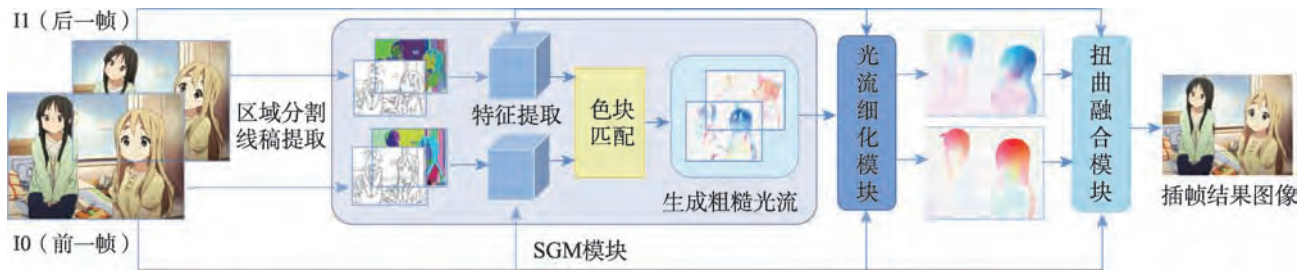


图4 AnimeInterp流程图

Fig.4 AnimeInterp pipeline

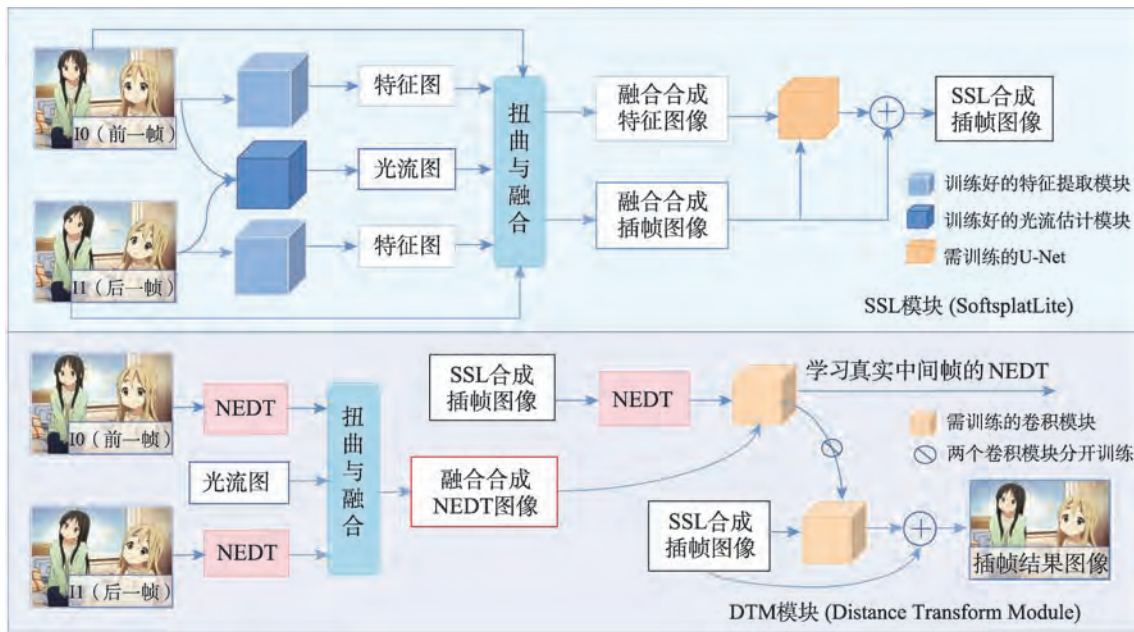


图5 Chen方法流程图

Fig.5 Chen's method pipeline

帧结果存在画面模糊、重影等问题,因此需要更好的动画插帧算法。这种算法不仅要能够解决视频插帧中常见的物体突然出现或遮挡消失的问题。还要能够准确估计二维动画中的光流信息,并有效处理动画画面中出现的特殊插帧问题,例如大面积纯色色块的运动,同色不同物体帧间对应等。

### 2.2.2 与帧间一致性有关的动画算法

为了解决动画影片超分辨率和动画插帧任务中出现的色块晕染和边缘线条模糊问题,通常需要引入确保画面帧间一致性的模块。在这方面,可以借鉴一些动画视频上色和动画追踪算法。这些算法通常涉及人物抠像、运动估计、边缘线条提取、色块分割、区域匹配、光流估计等领域。本小节重点介绍这些算法中为保证帧间一致性所采取的方法。

Sykora 等人<sup>[22]</sup>提出一种无监督的黑白动画视频上色算法。首先屏蔽背景层,只保留运动的人物,再

利用基于特征点匹配的KLT(Kanade-Lucas-Tomasi)<sup>[23]</sup>追踪算法,在第一帧中提取出高曲率点与交界点,然后在第二帧中寻找这些特征点的对应点,并计算它们之间的位移向量来得到光流场。根据这些匹配好的关键点,将每一帧的人物拆解成图像块,使用归一化差异平方和(L2-norm)估计参考图像与目标图像块间的结构相似程度。经过旋转和运动估计的处理后,对匹配好的图像块进行粘贴后再进行颜色统计并上色。如果出现相似结构但颜色不同导致匹配错误,可以通过建立一种图像块间归属关系图(attributed relational graphs, ARG)来表示共用边界线的图像切片。最后通过概率松弛方法排列该目标图像切片各种颜色的可能性,并选择可能性最大的颜色作为最终结果。

Zhu 等人<sup>[24]</sup>提出一种识别手绘动画中连续帧间的物体或区域的对应关系的动画追踪算法,使用一种漫

画的区域提取方法(stereoscopizing cel animations)<sup>[25]</sup>对序列中的每一帧进行区域标记。对于遮挡和后续区域发生分割的问题,该追踪算法考虑了序列中所有帧各区域间的关系,并且建立一个无环连通图,将区域间的外观匹配程度作为路径的损失,从而将跨帧区域匹配问题转化成一个最小路径问题。

目前许多动画视频上色算法使用深度学习方法,通常依靠相邻帧作为上色参考。Thasarathan等人<sup>[26]</sup>使用生成对抗网络进行训练,将线稿、灰度图以及参考帧的彩色图作为输入,结合条件损失、内容损失、风格损失共同训练。Shi等人<sup>[27]</sup>设计出一个由颜色转换网络和细化网络组成的模型。颜色转换网络的输入包括目标图像与参考图像的线稿、参考图像的彩色原图,以及目标图像与彩色图像线稿经过二值化处理后的欧几里德距离变换图。为了学习参考图像与目标线稿图像相似的空间特征,作者提出了一个距离注意力机制,利用非局部相似性将颜色信息从参考图像转换到目标图像。为保证上色的帧间一致性和统一的画面风格,颜色转换网络中还包含8个以自适应实例标准化(adaptive instance normalization, AdaIN)<sup>[28]</sup>作为归一化层的残差块。Li等人<sup>[29]</sup>提出了一种算法,用于完成关键帧之间的多帧上色。该算法使用两个关键帧和一张中间帧线稿草图作为指引。在第一步中,通过输入的两个关键帧获取颜色信息,学习中间帧草图简化后的线稿与彩色参考图之间的联系。然后,使用PWC-Net(pyramid, warping, and cost volume net)<sup>[30]</sup>生成光流估计结果,并进行光流一致性检查,得到遮挡处理的掩膜。最后,隐式地结合掩膜信息、线稿信息和光流信息,生成上色后的中间帧。为了减少由于线稿过细导致的画面模糊,还提出了一种基于倒角匹配的线条损失参与训练。在第二步中,假定根据光流估计结果按任意帧率生成其他中间帧的光流。然后,通过光流细化模块和光流一致性检查,混合生成上色后的结果。最后,将所有生成的中间帧作为输入,使用可变形卷积细化结果。Zhang等人<sup>[31]</sup>提出了一个相关匹配特征转移模块(correlation matching feature transfer model, CMFT)来保证时间一致性。作者认为相邻帧的线稿是风格相似但内容不同的图片,而同一帧的线稿与彩色图是内容相同但风格不同的图片。因此,可以先将参考图像的线稿和目标图像的线稿输入一个编码器,并将两个结果输入CMFT模块进行特征匹配,然后将彩色参考图经过编码器后的特征图输入

CMFT模块,进行特征转移。Yu等人<sup>[32]</sup>提出了一个两阶段光流着色网络来解决动画序列帧的着色问题。第一阶段,网络从给定的参考帧和下一帧线条之间的纹理变化中找到颜色像素流向,然后完成初始着色。第二阶段,网络对第一阶段的输出进行颜色校正,以此来增强帧间一致性。

综上所述,研究者在图像块或色块层面进行匹配;引入注意力机制、AdaIN、对相邻帧进行特征提取等深度学习方法;利用线稿或光流信息等方式来保证动画的帧间一致性。

### 2.3 工业界常用的二维动画修复软件

对于二维动画修复问题,学术界与工业界有不同的侧重目标和解决方案。学术界更重视算法的效果和创新性,通常致力于将前沿模型应用于动画领域,或设计出更符合动画特点的创新模型。而工业界更注重算法的实现、效率、可扩展性以及用户友好性,旨在实时、综合地解决动画修复问题。

例如,Topaz Labs<sup>[33]</sup>的“Topaz Photo AI”和“Topaz Video AI”利用深度学习方法,可以对图像和视频进行超分辨率、画质修复、运动插帧以及稳定抖动等多种处理。用户可以通过选择不同的模型和参数来修复低质量视频。

为解决各类影像问题,Topaz Labs使用了多种模型和训练策略。例如,在画质修复方面,使用Dione模型处理隔行扫描视频,并针对不同类型的数字视频格式和分辨率进行数据收集和训练,以得到相应的模型变体。他们还使用Proteus Fine Tune和Artemis模型作为通用模型,根据不同质量的数据进行训练,并获得各自的变体。此外,使用Gaia模型来优化高质量影像和计算机生成影像,以及使用Theia模型进行锐化和增强细节。对于运动插帧问题,使用Apollo模型处理非线性运动和稍微模糊的输入,以及使用Chronos模型处理帧之间变化较大的快速运动。

目前,Topaz Labs并未公开上述模型的具体结构和所使用的训练集。然而,Topaz Labs需要使用用户本地的硬件进行图片和视频处理,因此对用户的硬件有一定要求。例如要求用户的显卡至少为Nvidia GTX 900或更高版本,且具有4 GB以上的显存。显卡要求通常与模型的大小和复杂度相关联。可见,工业界为了确保用户能够正常使用,需要对模型的大小和复杂度进行限制。为了更好地处理复合问题,他们采用多种训练模型和训练策略,并联合使用

多个模型来解决复杂的问题。

除了使用深度学习的 Topaz Labs 之外, Adobe After Effects<sup>[34]</sup>、TVPaint Animation<sup>[35]</sup>等视频编辑和二维动画制作的专业软件也可用于动画影像的修复和增强。

### 3 结果分析与方法评估

#### 3.1 动画单张图片画质修复测试

本文对第 2.1 节中提到的基于深度学习网络的动画单张图片的超分方法进行了测试。方法包括 Anime4K 官方的 Tensorflow 版本、Waifu2x\_7Conv(无降噪版、降噪版)、Waifu2x\_CUNet(无降噪版、降噪版)、Waifu2x\_Swin(无降噪版、降噪版)、Real-CUGAN(无降噪版、降噪版)、Real-ESRGAN(Anime6B\_4X)。所有方法均使用官方提供的权重将原图像放大到两倍,并与其他方法进行对比。其中 Real-ESRGAN(Anime6B\_4X)版本上采样至四倍后再使用双三次插值方法下采样至两倍。

为了进行测试,从网络上选择了 169 张真实的低质量动画图片。后续会公开这 169 张真实低质量图片,它们包含多个国家不同风格、不同内容的动画作品,并且所有低质量图片存在不同程度的模糊、压缩、噪点等未知的复合退化情况。因此,这些低质量动画修复问题属于盲复原问题。

目前的网络均是利用拟合出的数据对进行训练,并没有学习过真实的低质量动画图片。同时,这些测试图片没有对应的高质量动画图片,因此需要使用无参考的评价方法来评估修复结果。在图像质量评价方面,选择了 NIQE(natural image quality evaluator)<sup>[36]</sup>、BRISQUE(blind/referenceless image spatial quality evaluator)<sup>[37]</sup>、NRQM(non-reference quality metric)<sup>[38]</sup>、感知指数(perceptual index, PI)<sup>[39]</sup>这些传统方法,它们不使用机器学习方法。另外,还采用了使用深度学习的方法,包括 GIQA(generated image quality)<sup>[40]</sup>、HyperIQA<sup>[41]</sup>、MANIQA(multi-dimension attention network image quality assessment)<sup>[42]</sup>。其中 GIQA 可以分为 KNN( $K$  nearest neighbor)版本、GMM(Gaussian mixture model)版本。然而,以上无参考的评价指标是否能有效评价动画图片仍需要进一步分析。

本文中所有实验均是在 Nvidia GeForce GTX TITAN X 上进行测试。

##### 3.1.1 动画单张图片修复结果

首先,对网络的无降噪版与降噪版进行了比较,具

体包括 Waifu2x\_7Conv、Waifu2x\_CUNet、Waifu2x\_Swin、Real-CUGAN。其中 Waifu2x 的三种网络参数量逐渐递增。人们通常认为深度学习网络参数量越多,修复效果会更好,但三种 Waifu2x 网络面对真实的低质量图片的处理结果并未展示出明显区别。Real-CUGAN 的表现明显优于三种 Waifu2x 网络。

真实的动画测试图片通常会存在噪声与压缩。如表 1 所示 Waifu2x 的三种无降噪版并没有学习到这类退化,因此无法处理这些问题,并且会产生伪影。而降噪版虽然只接触过拟合的低质量动画数据,但仍能在一定程度上去除噪声和压缩,从而取得更好的效果。并且在面对较为严重复合的噪声与压缩时,处理效果并不理想。Waifu2x 降噪版能使图像更平滑,但无法生成自然的纹理细节。在高模糊图片中 Waifu2x 的降噪版抹平了原本动画人物眼睛的高光,影响观感。相比之下,Real-CUGAN 的修复效果明显优于 Waifu2x 的三种网络,修复结果的边缘更加清晰,线条更加锐化。另外,Real-CUGAN 的无降噪版对图片中的噪声与压缩也有明显的效果,而 Real-CUGAN 降噪版的平滑效果更明显。

总体来说,与各自的无降噪版相比,各网络的降噪版修复结果的色块更加平坦,颜色过渡更加平滑,但也会损失部分细节。

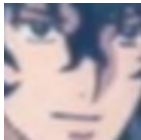


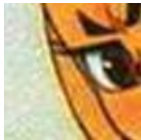

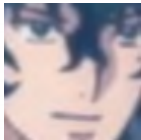


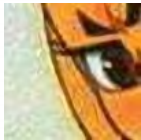

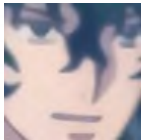


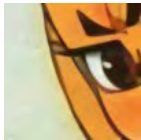

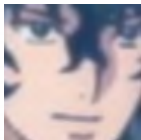




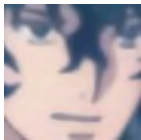
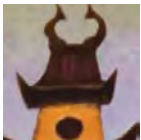

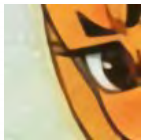

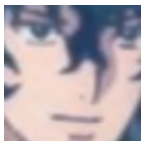




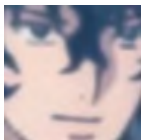


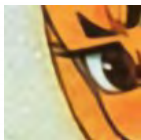


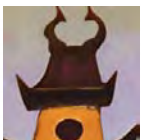

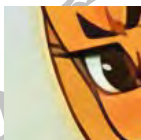


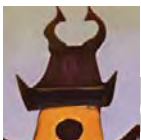

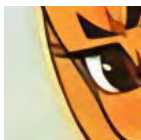
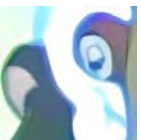
在二维动画的修复中,针对动画人物的处理效果对观众的感受至关重要。为了比较分析各方法对人物细节的修复效果,本文选择了 Anime4K、上一组实验中表现较好的 Real-CUGAN 无降噪版和降噪版,以及 Real-ESRGAN(Anime6B\_4X)方法。如表 2 所示,Anime4K 对动画图像的修复效果不太理想。而 Real-CUGAN 与 Real-ESRGAN 不仅能处理画面中模糊、噪点与压缩,还可以生成符合二维动画人物的眼睛、毛发、手部细节。特别是 Real-ESRGAN 生成的纹理最精细和清晰,对人物的处理效果最佳。

二维动画经常对画面中的某些部分进行虚化处理,通常需要通过网络保持画面原始的虚实关系,并在前景进行锐化,在背景进行虚化。为了实现这一目标,选择了在降噪去压缩和人脸修复方面表现最佳的方法 Real-CUGAN 和 Real-ESRGAN。如表 3 所示,Real-ESRGAN 缺乏对画面内容的理解,难以准确区分图片的前景和背景,这可能导致背景过度锐化,产生伪影,影响观感。相比之下,Real-CUGAN 相较于 Real-ESRGAN 对虚化区域保留更好。此外,Real-ESRGAN 在处理模糊的粗边缘线时容易产生中空现



表1 网络的无降噪版与降噪版修复结果对比

Table 1 Comparison of restoration results between no denoise models and denoise models

修复方法	模糊图片	噪声图片	压缩图片	压缩+噪声	严重退化
Bicubic					
Waifu2x_7Conv(无降噪版)					
Waifu2x_7Conv(降噪版)					
Waifu2x_CUNet(无降噪版)					
Waifu2x_CUNet(降噪版)					
Waifu2x_Swin(无降噪版)					
Waifu2x_Swin(降噪版)					
Real-CUGAN(无降噪版)					
Real-CUGAN(降噪版)					

象或高亮白边,这是其过度锐化的特点。然而,Real-CUGAN相对而言在这些情况下表现不太明显。在二维动画中常常具有一些特有的笔触和风格,但目前的网络对于笔触处理和风格的保留均不理想,经常出现涂抹感强的现象。

总体来说,Real-CUGAN与Real-ESRGAN的处理效果明显优于其他方法。这两种方法均倾向于锐

化线条边缘并使色块更加平坦均匀,这符合二维动画的特征。其中Real-CUGAN的修复结果对虚化区域的保留更好,而Real-ESRGAN更加善于构造纹理,尤其是人物的细节部分,但过锐的特点可能会产生边缘线中空、白边或生成伪影。因此希望未来的方法能够进一步理解画面内容,从而更好地保留画面中的虚实关系、笔触与风格。

表2 对动画人物细节的修复结果对比

Table 2 Comparison of restoration results for details of animation characters

修复方法	眼部细节	眼部细节	毛发细节	手部细节	面部整体	面部整体
Bicubic						
Anime4k						
Real-CUGAN(无降噪版)						
Real-CUGAN(降噪版)						
Real-ESRGAN						

表3 动画修复中的特殊情况

Table 3 Special cases in animation restoration

修复方法	前后背景 区别处理	虚化背景 人物处理	镜面反射 人物处理	过锐导致 边缘中空	过锐导致 边缘白边	笔触 处理	水彩风格 画面处理
Bicubic							
Real-CUGAN(无降噪版)							
Real-CUGAN(降噪版)							
Real-ESRGAN							

### 3.1.2 无参考图像质量评价结果

传统无参考图像质量评价指标不需要进行训练,通常是比较目标图片与真实的自然场景图片在

空间域上分布的差异。如NIQE、BRISQUE引入归一化亮度系数MSCN(mean-subtracted contrast normalization)比较目标图像MSCN分布与广义高斯分布的

偏离情况;NRQM提取了自然图像在多个色彩空间的特征;PI则结合了NRQM与NIQE的结果。然而二维动画图像与自然图像在空间域上存在明显差异,因此传统的无参考图像质量评价指标不能很好地评估动画图像。

深度学习的图像质量评价指标通常通过网络提取图片特征。如GIQA原本是为生成图片设计的评价指标,具有参数化模型(GMM-GIQA)和非参数化模型(KNN-GIQA)两种。GIQA首先对真实数据在特征层面建模,从AVC数据集中选择了909张1920×1080的真实高质量动画数据,并对其进行多尺度处理,采用0.75、0.50、0.375作为下采样参数,从而获得四种不同尺寸的真实图片进行特征提取。然后对待测图片进行特征提取,比较真实数据与待测图像分布,通过得分的排名来判断图片质量的高低。而HyperIQA通过多尺度模块加强对图片语义特征的理解,MANIQA使用多尺度注意力机制,通过ViT(vision transformer)<sup>[43]</sup>提取特征,加强全局和局部图像不同区域之间的交互作用。HyperIQA与MANIQA都通过增强对图片的语义理解来提高评估的准确性,并且两种IQA网络目前没有使用动画数据进行训练。

首先,需要验证各图像质量评价指标是否符合基本的视觉规律。根据第3.1.1小节中169张真实低质量动画图片修复结果,可以发现Bicubic修复效果明显不如Waifu2x\_CUNet(降噪版),而Real-CUGAN(无降噪版)与Real-ESRGAN的结果明显优于其他所有方法,更符合人类视觉感知。此外,两种方法各有特点,Real-CUGAN(无降噪版)对虚化区域保留得更好,而Real-ESRGAN对人物细节处理得更好,画面更锐利清晰。

因此一个成功的动画图片质量评价指标必须要正确地区分各方法的结果,需要使Waifu2x\_CUNet

(降噪版)的得分明显优于Bicubic,而Real-CUGAN(无降噪版)与Real-ESRGAN的打分可以相近,但得分要明显优于其他方法。

由表4所示,从169张低质量动画数据修复后的平均得分来看,NIQE、BRISQUE、NRQM、PI、KNN-GIQA的结果与视觉感知不符;GMM-GIQA对四种方法得分的排序基本正确,但Waifu2x\_CUNet(降噪版)与Real-CUGAN(无降噪版)区分并不明显,只算是大致符合;HyperIQA与MANIQA在未经过动画数据集微调的情况下已经取得了基本符合人类视觉感知的结果。从得分平均值来看GIQA、HyperIQA与MANIQA均更偏爱Real-ESRGAN的修复。

然而,从表5中两个具体的单张图像结果可见,第一幅“民族服饰女孩”图片中Real-CUGAN(无降噪版)清晰且无明显伪影,Real-ESRGAN在装饰处有伪影,因此对于这张图像,评价指标Real-CUGAN(无降噪版)的修复结果应该获得最高分。对于下面的动画群像图像,Real-ESRGAN清晰且对人物面部处理最好,因此对于这张图像,评价指标应该是Real-ESRGAN的处理结果得分更高。然而,GMM-GIQA、HyperIQA、MANIQA在Real-CUGAN(无降噪版)与Real-ESRGAN结果有明显的优劣时,打分情况没有完全符合人类感知。

除了满足基本的视觉规律,本文也验证了评价指标对动画图片中的压缩与噪声的敏感度。从169张真实动画图片中筛选出了80张有明显压缩、噪声的图片,使用Waifu2x\_7Conv、Waifu2x\_CUNet、Real-CUGAN这三种网络的无降噪版与降噪版进行比较。在处理噪声和压缩图片方面,Waifu2x两种网络的降噪版在噪声、压缩图片处理上均明显优于无降噪版;Real-CUGAN的两种版本都能有效地降噪去压缩,只是降噪版的结果更趋于平滑。总体而言,Real-

表4 低质量数据修复结果的各图像质量评价指标平均得分


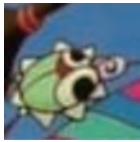
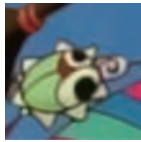
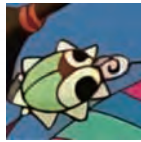
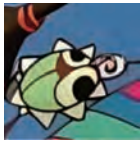





Table 4 Average score of image quality assessment methods for low-quality data restoration results

图像质量评价指标	Bicubic	Waifu2x_CUNet(降噪版)	Real-CUGAN(无降噪版)	Real-ESRGAN
NIQE ↓ (不符合)	6.933 3(1)	8.406 8(4)	7.592 4(3)	7.419 3(2)
BRISQUE ↓ (不符合)	56.420 3(2)	63.429 5(4)	42.752 7(1)	59.761 2(3)
NRQM ↑ (不符合)	6.517 8(4)	7.633 6(1)	7.375 4(2)	7.336 7(3)
PI ↓ (不符合)	6.601 3(3)	7.214 3(4)	5.727 2(1)	5.824 9(2)
KNN-GIQA ↑ (不符合)	0.005 391(4)	0.005 440(1)	0.005 420(2)	0.005 397(3)
GMM-GIQA ↑ (大致符合)	-5 833.31(4)	-5 555.17(3)	-5 531.35(2)	-5 483.46(1)
HyperIQA ↑ (符合)	48.389 4(4)	56.606 3(3)	70.013 7(2)	71.396 4(1)
MANIQA ↑ (符合)	0.608 1(4)	0.641 6(3)	0.671 6(2)	0.705 4(1)

注:“↑”表示该指标认为分数越高的图片越好,“↓”反之;数字右侧“( )”内部分表示IQA得分排名。

表5 单张图片修复结果在各图像质量评价指标下得分

Table 5 Score of single image after restoration evaluated by image quality assessment methods

图像质量评价指标	Bicubic	Waifu2x_CUNet(降噪版)	Real-CUGAN(无降噪版)	Real-ESRGAN
				
GMM-GIQA ↑ (不符合)	-3 742.62(2)	-4 870.67(4)	-3 553.59(1)	-4 041.78(3)
HyperIQA ↑ (不符合)	43.17(4)	49.31(3)	78.35(2)	79.67(1)
MANIQA ↑ (不符合)	0.415 3(3)	0.297 9(4)	0.617 0(1)	0.595 7(2)
				
GMM-GIQA ↑ (不符合)	-4 627.28(1)	-5 126.67(4)	-5 114.93(2)	-5 116.44(3)
HyperIQA ↑ (不符合)	46.26(4)	72.33(2)	69.71(3)	72.17(1)
MANIQA ↑ (符合)	0.622 8(4)	0.683 0(3)	0.720 7(2)	0.747 3(1)

CUGAN网络的结果比Waifu2x的两种网络的结果更加清晰。

因此,在评价各指标对噪声和压缩敏感度的时候需要明显区分Waifu2x网络的降噪版和无降噪版。Real-CUGAN的降噪版和无降噪版的得分可以接近,但总体而言,Real-CUGAN的结果应优于Waifu2x的两种网络的结果。

如表6所示,NIQE、BRIQUE、NRQM、PI无法区分噪声压缩图像与正常图像;KNN-GIQA可以正确区分两种Waifu2x网络的降噪版与非降噪版,但是Waifu2x的两种网络与Real-CUGAN的分数接近;GMM-IQA大致符合,但Waifu2x\_CUNet(降噪版)与Real-CUGAN两种版本的得分较为接近;HyperIQA的打分完全符合;MANIQA的评分与预期不符。具体单张图像结果可见于表7和表8,MANIQA对图片中的噪声敏感,给出了正确的评分,但对压缩不敏

感,导致处理压缩图片失败的结果得到较高的分数。这可能与MANIQA的训练数据有关。

另外,本文还测试了目前表现最好的三种评价指标,即GMM-GIQA、HyperIQA、MANIQA来评估它们对于画面中特殊情况的准确性,例如虚化区域的保留程度和边缘问题的出现等。

如表9所示,对于背景虚化图片,Real-ESRGAN过度锐化了虚化区域,导致产生伪影。然而,这三种评价指标仍将其评为最佳,这与视觉规律不符。当经过Real-ESRGAN处理后,图片边缘出现白边时,HyperIQA和MANIQA仍将其选为最佳,也不符合视觉规律。

综上所述,目前传统的无监督图像质量评价方法不适用于动画图像。而利用深度学习网络的这三种评价指标,即GMM-GIQA、HyperIQA和MANIQA,对于动画图像的评价结果基本符合视觉规律。目前,

表6 压缩与噪声明显数据的各图像质量评价指标平均得分

Table 6 Average score of image quality assessment methods for significant compression and noise data

图像质量评价指标	Waifu2x_7Conv (无降噪版)	Waifu2x_7Conv (降噪版)	Waifu2x_CUNet (无降噪版)	Waifu2x_CUNet (降噪版)	Real-CUGAN (无降噪版)	Real-CUGAN (降噪版)
NIQE ↓ (不符合)	5.755 0	7.687 4	6.058 2	8.063 0	7.255 1	7.171 7
BRISQUE ↓ (不符合)	50.578 3	55.640 1	50.593 4	59.521 7	40.653 4	47.017 2
NRQM ↑ (不符合)	4.539 2	4.546 1	4.537 5	4.300 4	6.132 7	6.210 9
PI ↓ (不符合)	5.607 9	6.570 6	5.760 3	6.881 3	5.561 1	5.480 3
KNN-GIQA ↑ (不符合)	0.005 332	0.005 446	0.005 335	0.005 479	0.005 469	0.005 477
GMM-GIQA ↑ (大致符合)	-5 936.18	-5 511.66	-5 941.97	-5 474.68	-5 496.61	-5 404.25
HyperIQA ↑ (符合)	48.495 6	56.532 0	48.296 8	56.425 4	71.043 3	71.687 9
MANIQA ↑ (不符合)	0.435 7	0.405 4	0.443 4	0.400 5	0.561 1	0.570 3

表7 单张压缩图片修复结果在各图像质量评价指标下得分

Table 7 Score of single compressed image after restoration evaluated by image quality assessment methods






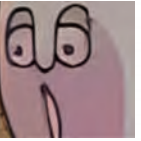
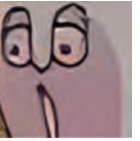
图像质量评价指标	Waifu2x_7Conv (无降噪版)	Waifu2x_7Conv (降噪版)	Waifu2x_CUNet (无降噪版)	Waifu2x_CUNet (降噪版)	Real-CUGAN (无降噪版)	Real-CUGAN (降噪版)
						
GMM-GIQA ↑ (大致符合)	-4 956.13	-3 865.02	-5 002.84	-3 768.61	-3 938.04	-4 166.11
HyperIQA ↑ (符合)	50.59	60.70	51.55	63.29	70.77	71.41
MANIQA ↑ (不符合)	0.478 1	0.420 7	0.508 9	0.444 4	0.650 9	0.623 9

表8 单张噪声图片修复结果在各图像质量评价指标下得分

Table 8 Score of single noise image after restoration evaluated by image quality assessment methods



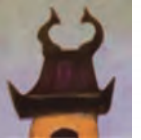

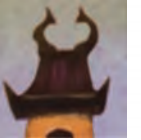
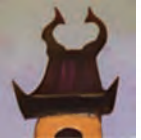







图像质量评价指标	Waifu2x_7Conv (无降噪版)	Waifu2x_7Conv (降噪版)	Waifu2x_CUNet (无降噪版)	Waifu2x_CUNet (降噪版)	Real-CUGAN (无降噪版)	Real-CUGAN (降噪版)
						
GMM-GIQA ↑ (大致符合)	-9 288.59	-8 057.73	-9 357.14	-8 069.28	-8 701.52	-8 553.18
HyperIQA ↑ (符合)	60.64	62.71	59.45	64.11	66.41	67.46
MANIQA ↑ (符合)	0.317 7	0.326 8	0.332 7	0.368 3	0.448 2	0.491 6

表9 修复图片特殊情况在各图像质量评价指标下得分

Table 9 Score of special image case after restoration evaluated by image quality assessment methods

背景虚化图片	Real-CUGAN(降噪版)	Real-ESRGAN	处理后产生边缘问题图片	Real-CUGAN(降噪版)	Real-ESRGAN
					
GMM-GIQA ↑ (不符合)	-4 299.03	-4 191.14	GMM-GIQA ↑ (符合)	-5 831.83	-6 213.98
HyperIQA ↑ (不符合)	74.83	75.72	HyperIQA ↑ (不符合)	76.35	78.01
MANIQA ↑ (不符合)	0.613 6	0.624 6	MANIQA ↑ (不符合)	0.480 8	0.528 8

MANIQA 对于图像中的压缩问题不敏感。这三种方法在背景虚化问题上失效,而 HyperIQA 和 MANIQA 在过度锐化的边缘问题上的评价也失效。如果 HyperIQA 和 MANIQA 能够学习更多种类的退化动画图片,也许它们的表现会更好。

### 3.1.3 动画单图修复算法效率比较

本小节对上述基于 Pytorch 框架<sup>[44]</sup>的深度学习网络的二倍超分辨率模型进行了算法效率的比较(其中 Real-ESRGAN 针对动画训练的权重是四倍超分辨率模型)。本文针对几个关键指标进行分析,分别是模型参数量、每秒浮点运算数(FLOPs)、激活次数(activation)、平均运行时间。

首先,比较了模型复杂度方面的指标。模型参数量是衡量模型大小的指标,它表示了模型中需要学习的参数的数量。每秒浮点运算数是衡量模型计算量的指标,它表示模型在每秒内执行的浮点运算的数量。这两个指标都对模型的效率和计算资源的需求有重要影响。其次,考虑了激活次数这一指标。激活次数表示模型中激活函数的执行次数,它直接关系到模型的计算速度和内存使用情况。最后,对平均运行时间进行了分析。平均运行时间是指在给定输入条件下,模型完成一次前向传播的平均时间。这个指标反映了模型的实时性能和计算效率。

表 10 所示的所有关键指标结果均是使用 256×

256 的图片在 Nvidia GeForce GTX TITAN X 上进行测试得到的。需要注意的是,网络参数量越大复杂度越高,则运行时间越长。其中 Waifu2x\_Swin 仅有 4 层卷积层,每秒浮点运算数和激活次数较小,但它仍然需要处理大量的自注意力操作和窗口划分操作,反而增加了计算时间。而 Waifu2x\_CUNet 和 Real-CUGAN 网络模型参数量相同,但用户可以设置不同的参数。此外,两个官方网络的代码实现细节、计算精度和编译器优化也存在差异。

表 10 动画单图修复算法效率比较

Table 10 Comparison of effectiveness of single image animation restoration algorithms

模型名称	参数量/ 10 <sup>6</sup>	每秒浮点 运算数/10 <sup>9</sup>	激活次 数/10 <sup>7</sup>	平均运行 时间/s
Waifu2x_7Conv	0.552 5	35.404 9	3.862 69	0.032 726
Waifu2x_CUNet	1.284 6	47.197 2	9.970 26	0.042 718
Waifu2x_Swin	3.758 3	4.319 7	1.264 92	0.099 735
Real-ESRGAN	4.467 8	374.278 7	38.902 17	0.164 037
Real-CUGAN	1.284 6	63.104 5	13.297 70	0.052 398

通过对以上指标的比较和分析,可以评估模型的算法效率,并选择最适合实际应用场景的模型。同时,也可以根据这些指标来进一步优化模型,以提高算法效率并减少资源消耗。

### 3.2 动画视频画质修复测试结果

接下来对第 2.1 节中提到的动画视频超分方法 AnimeVideo-v3 与 AnimeSR 进行了测试。测试数据选取了真实的低质量动画序列,包括一些 AVC-RealLQ 序列与网络下载的动画视频剪辑。为了评估它们的性能,使用了第 3.1.2 小节介绍的评价指标,即 GMM-GIQA、HyperIQA 和 MANIQA。

由表 11 可见,在“pig\_13”画面中,动画的笔触明显,形变剧烈,并带有马赛克状色块断层。AnimeVideo-v3 与 AnimeSR 在平滑画面的同时,也导致了严重的涂抹感。马赛克断层现象得到了平滑,但尚未完全解决。总体上,AnimeSR 的处理结果在画面边缘清晰,色块平滑方面表现更好,相比之下,AnimeSR 的结果更加锐利。

由表 12 可见,GMM-GIQA 在 9 组动画序列中认为有 5 组中 AnimeSR 的表现更佳,而 HyperIQA 和 MANIQA 则基本认为 AnimeSR 的结果更好。因为它采用了多尺度结构,增强了对画面内容的理解,并且输入了多帧图片,从而能够挖掘蕴藏在动画序列中的时序信息,更好地保证帧间的一致性。然而,目前

表 11 pig\_13 部分片段处理结果

Table 11 Restoration results of pig\_13 sequence



的评价指标只能对单张图像的结果进行评估,无法评估视频帧间的一致性。

### 3.3 动画插帧算法测试

本文对第 3.2 节中提到的动画插帧算法进行了测试,方法包括 AnimeInterp<sup>[21]</sup>与 Chen 方法<sup>[21]</sup>。测试数据选取了 ATD-12K 中 2 000 组测试集三联组。由于三联组的中间帧是插帧的真实值,评价指标可以选择全参考的峰值信噪比(peak signal-to-noise ratio, PSNR)、结构相似性指数(structural similarity, SSIM)<sup>[45]</sup>、学习感知图像切片相似度(learned perceptual image patch similarity, LPIPS)<sup>[46]</sup>。

从表 13 中可以看出,针对三种评价指标,Chen 方法表现更好。然而,目前的图像质量评价方法无法从时序角度分析插帧的合理性和一致性。

复杂运动与物体间遮挡是当前插帧算法的难点。此外,动画影像通常具有大面积平滑色块,缺乏纹理特征,并且经常出现非现实的夸张运动,使得动画插帧变得更加困难。AnimeInterp 与 Chen 的插帧方法均依赖光流,在结构中使用了能进行光流预测的 RAFT/RFR<sup>[47]</sup>模块。

如表 14 所示,当新物体突然在画面中出现时,如

表 12 视频超分算法结果的各图像质量评价指标平均得分

Table 12 Average score of image quality assessment methods for restoration results of animation video algorithm

动画剪辑名	AnimeVideo-v3			AnimeSR		
	GMM-GIQA	HyperIQA	MANIQA	GMM-GIQA	HyperIQA	MANIQA
Anime_dance_level_0	-5 119.31 (1)	68.71	0.533 7	-5 231.06	70.16 (1)	0.542 7 (1)
daria	-5 613.20	70.65	0.511 7	-5 593.60 (1)	73.04 (1)	0.532 4 (1)
digimon	-8 332.36	67.53	0.542 9 (1)	-8 190.67 (1)	67.93 (1)	0.529 9
Little_carp	-6 311.58 (1)	73.73	0.368 8	-6 432.67	74.99 (1)	0.385 6 (1)
pokemon	-5 890.85	72.61	0.520 9	-5 633.29 (1)	74.67 (1)	0.557 2 (1)
The_monkey_king	-6 304.64 (1)	73.00	0.453 0	-6 354.99	73.73 (1)	0.462 9 (1)
Timon_pumbaa	-8 567.88	72.80 (1)	0.571 2	-8 334.70 (1)	72.73	0.576 5 (1)
Tom_and_jerry	-4 405.63	74.38	0.508 4 (1)	-4 277.52 (1)	75.24 (1)	0.488 1
pig_13	-4 646.91 (1)	69.87	0.373 1 (1)	-4 929.52	71.18 (1)	0.372 0

注: 分数值内“(1)”表示在该评价指标中的最优分数。

表 13 ATD-12KTest 插帧后平均得分

Table 13 Average score of ATD-12KTest after interpolation

插帧方法	PSNR ↑	SSIM ↑	LPIPS ↓
AnimeInterp <sup>[2]</sup>	27.02	0.920 6	0.065
Chen 方法 <sup>[21]</sup>	29.22 (1)	0.930 7 (1)	0.036 (1)

果前一帧没有像素点可以映射到新物体的区域,使用前向扭曲(forward warping)方法会导致空洞的产生。而使用后向扭曲(backward warping)方法时,新物体的部分区域在前一帧中没有可以映射到的位置,因此插帧效果差。反之,若原本画面中的物体在下一帧中被遮挡或消失,插帧效果也不理想,正如表 14 中的前两组示例所示。

对动画由大量纯色色块组成的特点进行分析,表 14 第 3 组示例中的挥刀转动属于纯色物体的非直线运动,该实例中“刀”由纯色组成,缺乏纹理特征,造成光流方向估计错误。同理,第 4 组中的“红桃队列”具有相同颜色,但每个“红桃”均是独立的物体,它们在动画中的运动方向不同,插帧算法很难分辨出不同帧间每个“红桃”的对应关系,因此两种插帧方法在保证帧间一致性上表现不佳。

在 AnimeInterp 与 Chen 方法的对比中,发现 AnimeInterp 的插帧结果边缘模糊,涂抹感强,插帧不佳的部分呈现模糊态。Chen 方法使用了含有边缘信息的欧几里德变换,因此边缘特征更明显,插帧不佳的部分呈现重影态而非模糊。由于 Chen 更重视边缘的做法更符合动画作品的特点,其插帧结果也常有优于 AnimeIntrep 的表现。如表 14 中最后两组示例所示,Chen 插帧结果在对人物毛发、五官等细节部分

边缘清晰,更接近真实中间帧结果,而 AnimeInterp 模糊不清。

综上所述,动画插帧面临的难点包括新物体进入画面与原有物体被遮挡或消失的情况。除此以外,大面积纯色缺乏纹理、不同物体相同颜色等动画特点也会增大光流估计的难度。AnimeInterp 插帧常会造成画面模糊,边缘不清;而 Chen 方法插帧常会造成重影,但边缘相比 AnimeInterp 结果更加清晰。

## 4 动画修复问题的瓶颈与方向


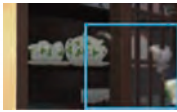



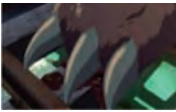
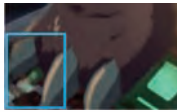
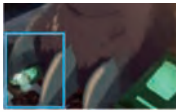
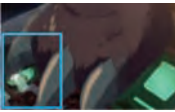

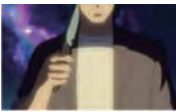
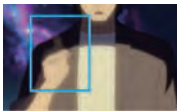
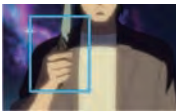
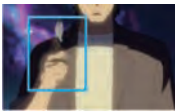


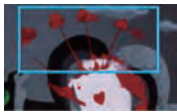
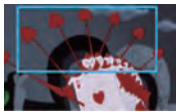
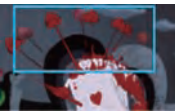

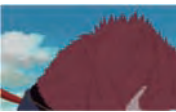
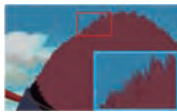




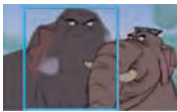
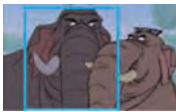
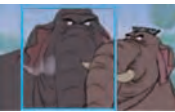
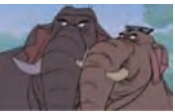
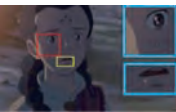
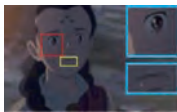
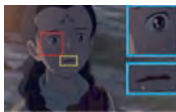

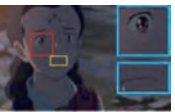


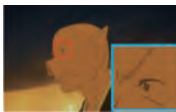
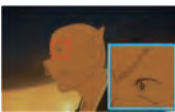

### 4.1 动画数据集

在画质修复方面,缺乏成对的真实动画数据。因此只能使用降质方法生成对应的低质量图像,或者利用画质修复方法得到高质量图像。然而,这种拟合的数据必须尽可能接近真实数据集的分布,这是收集动画画质相关数据时的难点。虽然 Real-ESRGAN 与 AnimeSR 均有自己的降质策略,但表 11 “pig\_13”动画中的马赛克颜色断层问题仍未得到完全解决,因此需要找到更适合动画的降质策略。

在时序方面,二维动画数据缺乏矢量标记。在 AnimeInterp 算法中,通过色块匹配的方式来弥补光流数据的缺失。Chen 方法注重线条的提取,取得了更好的效果。然而,当面临大面积纯色或不同物体具有相同颜色的情况时,这两种方法都可能由于光流预测错误而导致插帧失败。

AnimeRun 与 AnimeCeleb 利用三维软件建模,模拟二维动画特征进行渲染,得到具有矢量标记的二维动画数据。但是三维软件建模的成本较高,模拟二维动画的光效较为单一,难以得到逼真的二维动

表 14 插帧方法修复情况  
Table 14 Animation interpolation restoration results

情况描述	前一帧	AnimeInterp 插帧结果	真实中间帧	Chen 方法插帧结果	后一帧
新物体突然出现					
原有物体遮挡或消失					
纯色块非直线运动(挥刀转动)					
同色不同物体相邻帧间对应困难					
两种插帧算法画面特点区别					
					
Chen 方法优于 AnimeInterp 案例					
					

画风格。因此,结合二维动画画面特征与三维软件辅助是动画矢量标记的重要方向。

## 4.2 动画画质修复相关方法

在单张动画图片画质修复方面,目前Real-CUGAN与Real-ESRGAN的处理效果最优。然而,这两种方法并没有针对动画画面特点进行模型的改进。其中Real-CUGAN对人物处理粗糙,细节不佳。Real-ESRGAN过锐的特点会产生边缘线中空、白边和伪影,尤其在虚化的背景区域。两种方法对笔触和动画风格的保留均不足,涂抹感强。

要解决以上问题,今后的网络设计需要结合动画画面特点,更好地理解画面内容,从而更好地保留画面中的虚实关系、笔触与风格。在网络结构上可以加入对边缘特征的提取,通过在数据集上进行降质方法改进、蒙版处理、特殊情况筛选等步骤,来进一步保留画面原本风格。

多帧输入的AnimeSR在表现上优于单张输入的AnimeVideo-v3,这提示了动画序列间的时序信息的重要性。AnimeSR采用了多尺度的循环模块,增强了网络对画面内容的理解和对时序信息的利用,为未来模型的探索提供了参考。

## 4.3 动画插帧算法

AnimeInterp和Chen的插帧方法都没有利用具有矢量标注的数据集进行学习。除了视频插帧中常见的问题,如物体突然出现或原有物体遮挡消失,还存在纯色或相似色块光流估计不准的问题。例如,人物的两只眼睛通常具有相似的形状和颜色,如果在相邻帧之间无法正确对应眼睛,对画面的影响将非常明显。

而AnimeRun则具有真实的矢量标注,尽管其模拟的二维动画与真实二维动画之间存在差距,但光流信息是准确的,并且同色但不同物体之间的帧间



匹配也是准确的。因此,利用类似 AnimeRun 这样使用三维软件制作的光流数据集参与动画插帧网络训练是一个方向。同时 AnimeInterp 与 Chen 方法对动画特征的利用也值得学习和借鉴。

#### 4.4 评价指标

目前,传统的无参考图像质量评价方法,如 NIQE、BRISQUE、NRQM 和 PI,其设计初衷是针对自然图像,因此在动画评估中全部失效。然而,它们在自然图像中的特征提取方法可以提供借鉴的思路。在利用深度学习的无参考图像质量评价方法中,GIQA 使用大量真实的动画图像作为参考,通过筛选和收集这些真实图像可以提高 GIQA 的效果。HyperIQA 和 MANIQA 表现最佳,但它们偏向于过锐化的效果。建立一个动画质量评价数据集,并对 HyperIQA 和 MANIQA 进行重新训练,将使这两种方法在动画中更加适用。

对于全参考的图像质量评价方法,如 PSNR、SSIM 和 LPIPS,它们可以用于对单张图像进行评价,但在动画作品中,“观感”比“精确度”更为重要。与超分辨率或插帧后的图像与参考图像之间的误差相比,画面中线条、人物等细节的纹理对观感的影响更为明显。例如人物面部占画面的比例很小,但往往能最直观地体现修复效果的好坏。因此,在动画修复相关工作中,应更加注重观感和细节。除了单张画面的细节和画质外,动画中帧间的一致性也非常重要。色块匹配错误会导致边缘模糊、色块晕染和画面不流畅等问题。然而,目前尚缺乏用于评估帧间一致性的指标。

#### 4.5 对工业界发展的展望

Topaz Labs 的成功商业化验证了深度学习方法在实际应用中的可行性。然而,利用用户本地资源进行计算需要对模型的复杂度进行限制。因此,工业界需要采用模型压缩技术,如剪枝、量化和分解等,以减少模型的存储需求和计算复杂度。同时,优化训练策略和数据集,提升网络的能力,使网络更高效。此外,网络设计需要尽可能地端到端,以优化用户的使用体验。

随着计算能力的提升和云计算的普及,实时动画修复技术和云服务将成为重要的发展方向。云服务能够提供高性能的计算资源,使动画修复工作能够利用更复杂的模型,从而获得更好的修复效果。

为了进一步提升用户体验,便捷的交互方式和多模态技术具有广阔的发展前景。用户可以通过文

字或其他交互方法智能而优雅地修改和完善动画画面,从而进一步提升其质量和可操作性。

综上所述,工业界的发展方向包括使用更高效的网络模型为用户提供优质的效果,以及探索深度学习和云平台的结合,将科研界前沿的方法实际应用于工业领域。这将推动深度学习技术在动画修复领域的应用,并为用户提供更优质的服务体验。

#### 5 结束语

目前动画的修复任务已取得一定的进展,但仍存在一些亟待解决的关键问题。本文展示了这些方法的实际效果,对关键问题与难点进行了总结归纳,并提出了一些潜在的研究方向。未来希望能够出现更多的动画相关数据集、修复算法以及评价指标。同时,这些方法需要应用到真实的动画场景中,例如动画相关的超分算法可以使过去的动画作品在 2K、4K 的屏幕上重现光彩。动画插帧算法可以极大地减少动画工作者的劳动时间。希望研究者能更加注重观感与实际,并进一步优化和改进动画相关修复工作。

#### 参考文献:

- [1] WU Y, WANG X, LI G, et al. AnimeSR: learning real-world super-resolution models for animation videos[J]. arXiv: 2206.07038, 2022.
- [2] SIYAO L, ZHAO S, YU W, et al. Deep animation video interpolation in the wild[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Jun 20-25, 2021. Piscataway: IEEE, 2021: 6587-6595.
- [3] GWERN B. Danbooru2021: a large-scale crowdsourced and tagged anime illustration dataset[EB/OL]. (2022-01-23)[2023-03-22]. <https://www.gwern.net/Danbooru2021>.
- [4] ZHANG L, JI Y, LIU C. DanbooRegion: an illustration region dataset[C]//LNCS 12358: Proceedings of the 16th European Conference on Computer Vision, Glasgow, Aug 23-28, 2020. Cham: Springer, 2020: 137-154.
- [5] IKUTA H, OGAKI K, ODAGIRI Y. Blending texture features from multiple reference images for style transfer[C]//Proceedings of the SIGGRAPH Asia 2016 Technical Briefs, Macao, China, Dec 5-8, 2016. New York: ACM, 2016: 1-4.
- [6] RIOS E A, CHENG W H, LAI B C. DAF:re: a challenging, crowd-sourced, large-scale, long-tailed dataset for anime character recognition[J]. arXiv:2101.08674, 2021.
- [7] ZHENG Y, ZHAO Y, REN M, et al. Cartoon face recognition: a benchmark dataset[C]//Proceedings of the 28th

- ACM International Conference on Multimedia, Seattle, Oct 12-16, 2020. New York: ACM, 2020: 2264-2272.
- [8] LI H, GUO S, LYU K, et al. A challenging benchmark of anime style recognition[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, Jun 19-20, 2022. Piscataway: IEEE, 2022: 4720-4729.
- [9] BLENDER. Blender open source 3D creation suite[CP/OL]. Netherlands: Blender Conference (1994) [2023-01-20]. <https://www.blender.org/>.
- [10] SHUGRINA M, LIANG Z, KAR A, et al. Creative flow + dataset[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, Jun 16-20, 2019. Piscataway: IEEE, 2019: 5384-5393.
- [11] LI S Y, LI Y H, LI B, et al. AnimeRun: 2D animation visual correspondence from open source 3D movies[J]. arXiv: 2211.05709, 2022.
- [12] KIM K, PARK S, LEE J, et al. AnimeCeleb: large-scale animation CelebHeads dataset for head reenactment[C]//LNCS 13668: Proceedings of the 17th European Conference on Computer Vision, Tel Aviv, Oct 23-27, 2022. Cham: Springer, 2022: 414-430.
- [13] bloc97/Anime4K: a high-quality real time upscale for anime video[EB/OL]. (2018)[2023-01-22]. <https://github.com/bloc97/Anime4K>.
- [14] NAGADOMI. Waifu2x: image and video super-resolution using deep convolutional neural networks[EB/OL]. (2015) [2023-01-22]. <https://github.com/nagadomi/waifu2x>.
- [15] DONG C, LOY C C, HE K, et al. Image super-resolution using deep convolutional networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 38(2): 295-307.
- [16] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[C]//LNCS 9351: Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Oct 5-9, 2015. Cham: Springer, 2015: 234-241.
- [17] LIU Z, LIN Y, CAO Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision, Montreal, Oct 10-17, 2021. Piscataway: IEEE, 2021: 9992-10002.
- [18] BILIBILI. RealCUGAN: improved training of generative adversarial networks using realistic unpaired data synthesis [EB/OL]. (2021)[2023-01-22]. <https://github.com/bilibili/ailab/tree/main/RealCUGAN>.
- [19] WANG X T. Anime model[EB/OL]. (2021)[2023-01-22]. [https://github.com/xinntao/Real-ESRGAN/blob/master/docs/anime\\_model.md](https://github.com/xinntao/Real-ESRGAN/blob/master/docs/anime_model.md).
- [20] WANG X T. Anime video model[EB/OL]. (2021)[2023-01-22]. [https://github.com/xinntao/Real-ESRGAN/blob/master/docs/anime\\_video\\_model.md](https://github.com/xinntao/Real-ESRGAN/blob/master/docs/anime_video_model.md).
- [21] CHEN S, ZWICKER M. Improving the perceptual quality of 2D animation interpolation[C]//LNCS 13677: Proceedings of the 17th European Conference on Computer Vision, Tel Aviv, Oct 23-27, 2022. Cham: Springer, 2022: 271-287.
- [22] SÝKORA D, BURIÁNEK J, ŽÁRA J. Unsupervised colorization of black-and-white cartoons[C]//Proceedings of the 3rd International Symposium on Non-photorealistic Animation and Rendering, Annecy, Jun 7-9, 2004. New York: ACM, 2004: 121-127.
- [23] BRUCE D L, Takeo K. An iterative image registration technique with an application to stereo vision[C]//Proceedings of the 7th International Joint Conference on Artificial Intelligence, Vancouver, Aug 24-28, 1981. San Francisco: Morgan Kaufmann Publishers Inc, 1981: 674-679.
- [24] ZHU H, LIU X, WONG T T, et al. Globally optimal toon tracking[J]. ACM Transactions on Graphics, 2016, 35(4): 75.
- [25] LIU X, MAO X, YANG X, et al. Stereoscopizing cel animations[J]. ACM Transactions on Graphics, 2013, 32(6): 223.
- [26] THASARATHAN H, EBRAHIMI M. Artist-guided semi-automatic animation colorization[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshops, Seoul, Oct 27-Nov 2, 2019. Piscataway: IEEE, 2019: 3157-3160.
- [27] SHI M, ZHANG J Q, CHEN S Y, et al. Reference-based deep line art video colorization with a few references[J]. IEEE Transactions on Visualization and Computer Graphics, 2023, 29(6): 2965-2979.
- [28] HUANG X, BELONGIE S. Arbitrary style transfer in real-time with adaptive instance normalization[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision, Venice, Oct 22-29, 2017. Piscataway: IEEE, 2017:1501-1510.
- [29] LI X, ZHANG B, LIAO J, et al. Deep sketch-guided cartoon video inbetweening[J]. IEEE Transactions on Visualization and Computer Graphics, 2021, 28(8): 2938-2952.
- [30] SUN D, YANG X, LIU M Y, et al. PWC-net: CNNs for optical flow using pyramid, warping, and cost volume[C]//Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Jun 18-23, 2018. Piscataway: IEEE, 2018: 8934-8943.
- [31] ZHANG Q, WANG B, WEN W, et al. Line art correlation matching feature transfer network for automatic animation

- colorization[C]//Proceedings of the 2021 IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, Jan 3-8, 2021, Piscataway: IEEE, 2021: 3872-3881.
- [32] YU Y, QIAN J, WANG C, et al. Animation line art colorization based on optical flow method[J]. SSRN Electronic Journal, 2022. DOI:10.2139/ssrn.4202289.
- [33] Topaz Labs. Topaz Labs: AI image quality software[EB/OL]. [2023-01-28]. <https://topazlabs.com>.
- [34] Adobe after effects: create motion graphics and visual effects for film, TV, video, and web[EB/OL]. [2023-01-28]. <https://www.adobe.com/products/aftereffects.html>.
- [35] TVPaint 2D animation software[CP/OL]. (1991) [2023-01-28]. <http://www.tvpaint.com/>.
- [36] MITTAL A, SOUNDARARAJAN R, BOVIK A C. Making a “completely blind” image quality analyzer[J]. IEEE Signal Processing Letters, 2012, 20(3): 209-212.
- [37] MITTAL A, MOORTHY A K, BOVIK A C. No-reference image quality assessment in the spatial domain[J]. IEEE Transactions on Image Processing, 2012, 21(12): 4695-4708.
- [38] MA C, YANG C Y, YANG X, et al. Learning a no-reference quality metric for single-image super-resolution[J]. Computer Vision and Image Understanding, 2017, 158: 1-16.
- [39] BLAU Y, MICHAELI T. The perception-distortion tradeoff [C]//Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Jun 18-23, 2018. Piscataway: IEEE, 2018: 6228-6237.
- [40] GU S, BAO J, CHEN D, et al. GIQA: generated image quality assessment[C]//LNCS 12356: Proceedings of the 16th European Conference on Computer Vision, Glasgow, Aug 23-28, 2020. Cham: Springer, 2020: 369-385.
- [41] SU S, YAN Q, ZHU Y, et al. Blindly assess image quality in the wild guided by a self-adaptive hyper network[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, Jun 13-19, 2020, Piscataway: IEEE, 2020: 3667-3676.
- [42] YANG S, WU T, SHI S, et al. MANIQA: multi-dimension attention network for no-reference image quality assessment [C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, Jun 18-24, 2022. Piscataway: IEEE, 2022: 1191-1200.
- [43] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: transformers for image recognition at scale[J]. arXiv:2010.11929, 2020.
- [44] PASZKE A, GROSS S, MASSA F, et al. Pytorch: an imperative style, high-performance deep learning library [C]//Advances in Neural Information Processing Systems 32, Vancouver, Dec 8-14, 2019: 8026-8037.
- [45] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE transactions on image processing, 2004, 13(4): 600-612.
- [46] ZHANG R, ISOLA P, EFROS A A, et al. The unreasonable effectiveness of deep features as a perceptual metric[C]//Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Jun 18-23, 2018. Piscataway: IEEE, 2018: 586-595.
- [47] TEED Z, DENG J. RAFT: recurrent all-pairs field transforms for optical flow[C]//LNCS 12347: Proceedings of the 16th European Conference on Computer Vision, Glasgow, Aug 23-28, 2020. Cham: Springer, 2020: 402-419.



李雨航(1998—),女,吉林人,硕士研究生,主要研究方向为图像超分辨率、动画修复。

**LI Yuhang**, born in 1998, M.S. candidate. Her research interests include image super-resolution and animation restoration.



谢良彬(1997—),男,江西人,博士研究生,主要研究方向为视频/图像超分辨率、图像生成等。

**XIE Liangbin**, born in 1997, Ph.D. candidate. His research interests include video/image super-resolution, image generation, etc.



董超(1989—),男,山东人,博士,研究员,博士生导师,主要研究方向为图像/视频超分辨率、去噪和画质增强等。

**DONG Chao**, born in 1989, Ph.D., professor, Ph.D. supervisor. His research interests include image/video super-resolution, denoising and image quality enhancement, etc.