



## 超分辨率重建技术及其在智能终端上的应用

祝谷乔<sup>1</sup>, 姜超<sup>2</sup>, 徐煜烨<sup>2</sup>

(1. 天翼数字生活科技有限公司, 上海 200085;

2. 中国信息通信科技集团有限公司, 湖北 武汉 430070)

**摘 要:** 简单介绍了超分辨率重建技术的发展历程和几种有代表性的超分方法及其实现原理, 提出了在智能终端上实现超分辨率重建技术的方案, 通过实验仿真, 给出了采用插值算法和深度学习算法实现的单图像超分辨率对终端处理性能、图像质量等方面的评估与分析, 提出了利用智能终端实现超分的适用场景的建议。并进一步探讨了该技术在娱乐视频与家庭监控业务中的典型应用场景, 展望了超分重建技术未来可能的研究方向以及与相关图像处理技术相融合的发展趋势。

**关键词:** 超分辨率; 插值方法; 深度学习算法; 视频业务

**中图分类号:** TP393

**文献标志码:** A

**doi:** 10.11959/j.issn.1000-0801.2023150

## Super-resolution reconstruction technology and its application on intelligent terminal device

ZHU Guqiao<sup>1</sup>, JIANG Chao<sup>2</sup>, XU Yuye<sup>2</sup>

1. E-surfing Digital Life Technology Co., Ltd., Shanghai 200085, China

2. China Information Communication Technologies Group Co., Ltd., Wuhan 430070, China

**Abstract:** The development history of super-resolution reconstruction technology and its typical approaches were briefly introduced. An implementation solution of super-resolution reconstruction was proposed on one kind of intelligent terminal device. The image super-resolution approaches implemented by interpolation algorithms and deep learning algorithms were experimented and simulated, and their results on terminal's processing performance and images' quality of experience were evaluated and analyzed. The suggestions on appropriate scenarios of super-resolution implemented by the intelligent terminal device were proposed. Furthermore, some typical cases of super-resolution reconstruction technology in the fields of entertainment video services and home surveillance services were discussed. The possible research direction and the trend of convergence of this technology and relevant image processing technologies were also prospected.

**Key words:** super resolution, interpolation approach, deep-learning algorithm, video service

## 0 引言

超分辨率重建 (super-resolution reconstruction, SR) 是计算机视觉领域的一类重要技术, 是一门具有较高科学研究价值且有广泛应用领域的现代图像处理技术。图像分辨率是图像质量的重要指标, 分辨率越高, 图像提供的信息越丰富, 细节越精细, 质量越佳。内容采集、存储等方面的限制以及传输过程中的干扰等因素, 导致图像质量出现不同程度的退化。超分辨率重建技术就是采用信号处理的方法提高图像分辨率, 它基于一张或多张低分辨率图像对缺失的图像细节进行恢复或补足<sup>[1-2]</sup>, 以实现图像或视频从低分辨率向高分辨率的转换, 在生物医学<sup>[3]</sup>、航空研究、视频监控、视频直播/点播<sup>[4]</sup>、虚拟现实等领域表现出巨大的应用潜力。

针对娱乐视频和家庭视频监控等业务领域, 超分辨率重建技术目前尚未得到规模化应用。本文通过在智能机顶盒终端上实现的几种插值算法和深度学习算法, 通过实验仿真, 给出了图像超分辨率重建对图像质量、终端性能等方面的评估与分析, 并基于当前芯片的处理能力及从用户视觉体验的角度, 提出了智能终端实现图像或视频超分辨率的适用场景, 为该技术在视频业务领域的应用推广提供参考。

## 1 主要算法

超分辨率技术诞生于 20 世纪 60 年代, 经历了单图像的超分辨率、与深度学习算法相结合、视频序列的超分辨率等不同阶段, 本文主要针对单图像超分辨率重建的几种典型算法进行简单介绍。

### 1.1 插值算法

插值算法<sup>[5]</sup>是超分辨率技术中最基本、简单与直接的一种方法, 其思路是通过不同的计算方法模拟像素之间的映射关系, 利用周围像素点的信息对目标像素点进行插值处理。它根据抽样函数或信号估计出连续位置的数值, 或者根据一系

列离散采样点重建原始的连续函数, 目标是尽可能地保留图像的细节。

#### 1.1.1 最近邻插值

最近邻插值 (nearest neighbor interpolation) 又称零阶插值, 其计算原理是取待插值点周围 4 个相邻像素点中距离最短的一个邻点的灰度值作为该点的灰度值。

#### 1.1.2 双线性插值

双线性插值 (bilinear interpolation) 又称一阶插值, 其插值原理是待插点像素值取原图像中与其相邻的 4 个点像素值的水平、垂直两个方向上的线性内插, 即根据待采样点与周围 4 个邻点的距离确定相应的权重, 从而计算出待采样点的像素值。

#### 1.1.3 双三次插值

双三次插值 (bicubic interpolation) 又称立方卷积插值<sup>[6]</sup>, 是一种比较复杂的插值方法, 其原理是利用以采样点为中心的  $4 \times 4$  邻域内的 16 个已知像素点的灰度值估计采样点的灰度值。双三次插值的示意图如图 1 所示。

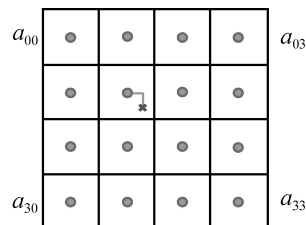


图 1 双三次插值的示意图

#### 1.1.4 兰索斯插值

兰索斯插值 (Lanczos interpolation) 是选取待采样点在水平、垂直方向上相邻的 8 个像素点的值并进行加权求和的计算, 以生成目标像素点的像素值。与双线性插值算法不同的是, 插值点的权重由高阶函数计算得到。

## 1.2 深度学习算法

深度学习算法是对特征类型的图片进行训练, 并利用专门设计的神经网络对低分辨率图像进行纹理补充、细节补充、降噪、去隔行等处理,



从而重建具有丰富细节内容的高分辨率图像。该方法通过学习算法获得高分辨与低分辨率图像之间关系的先验知识,以取代传统的重建方法所用的正则化算法。

基于深度学习的超分辨率图像重建方法主要采用以下几种网络结构。

### 1.2.1 基于卷积神经网络

基于卷积神经网络<sup>[7]</sup> (convolutional neural network, CNN) 的深度学习算法是直接对低分辨率图像和高分辨率图像进行端到端映射的学习,弥补传统重建算法对高频细节信息丢失的缺陷。

#### (1) 线性网络

线性网络有最为简单的结构,仅由单一的信号流动路径组成,多个卷积层相互叠加,网络的输入从初始层依次传递向后层。根据上采样方式的不同,可分为预上采样模式和后上采样模式两种类型。

在使用预上采样模式的线性网络中,首先会将输入的低分辨率图像通过上采样操作放大至目标高分辨率图像的尺寸,然后送入网络学习分层,经特征表示后,得到最终的超分辨率重建结果,预上采样模式中的上采样操作通常使用双三次插值算法。该方法学习难度较低,时间和空间复杂度较高。

后上采样模式的网络首先会对低分辨率的输入进行学习,然后对网络输出的特征进行上采样,

完成重建任务。这种方法让特征提取过程在低维空间上进行,训练速度和推理速度较快,但它不能适应不同的放大倍数。

#### (2) 残差网络

与线性网络相比,基于残差学习的网络在设计中采用跳跃连接避免梯度消失,支持设计更深的网络结构。网络模型用于学习生成一张目标高分辨率图像与经过插值后的低分辨率图像之间的残差图像,通过将残差图像与输入低分辨率图像的插值结果逐像素相加得到重建结果。残差网络对图像中高频信息的学习更为充分。采用了残差网络的 VDSR (accurate image super-resolution using very deep convolutional networks) 流程如图 2 所示。

#### (3) 递归网络

递归网络<sup>[8]</sup>采用的是递归连接的卷积层,即通过在网络的不同深度复用同一个卷积,达到降低网络参数的目的。它将一个困难的任务分解成简单任务,可降低学习难度,但存在模型复杂、训练难度大的问题。

### 1.2.2 基于生成对抗网络

基于生成对抗网络<sup>[9]</sup> (generative adversarial network, GAN) 的深度学习算法是利用“对抗博弈”的思想,将模型训练无监督化,借助反向传播不断优化模型缩减原始图像与重建图像之间的

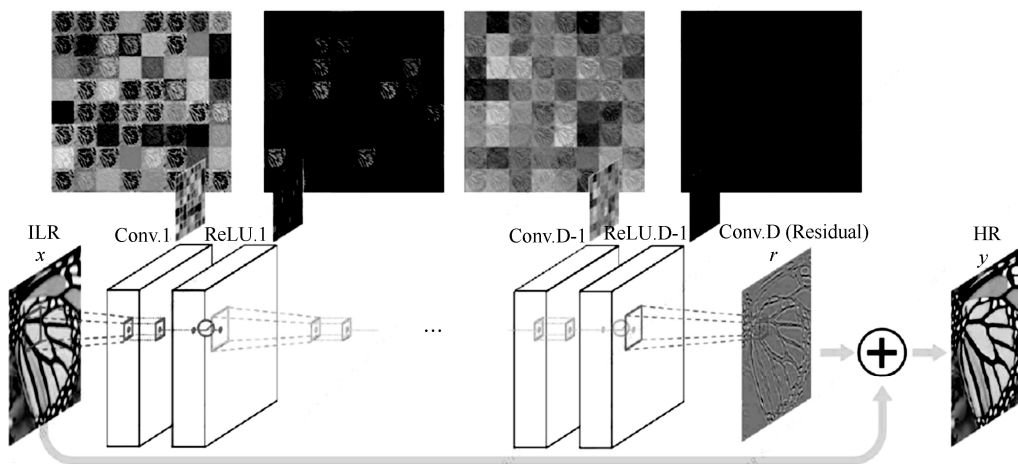


图2 采用了残差网络的 VDSR 流程

差距。对抗网络引入生成器和判别器的两个神经网络,通过迭代训练使之达到纳什均衡状态。通常情况下,进行对抗损失训练的模型具有更好的感知质量,但训练过程存在困难与不稳定性<sup>[10]</sup>。

对抗网络典型结构示意图如图 3 所示。

## 2 实现方案

本文在一款 8K 智能机顶盒上实现了超分辨率重建技术,通过将 OpenCV 开发工具库和部分 OpenCV\_contrib 移植进机顶盒系统版本中,形成软件开发工具包 (software development kit, SDK) 供上层应用调用,支持图像或视频超分辨率的处理能力。将多种典型的插值算法、深度学习算法集成进 OpenCV 的 resize 函数和 DnnSuperResImpl 函数,同时也实现了均方误差<sup>[7]</sup> (mean square error, MSE)、峰值信噪比<sup>[7]</sup> (peak signal-to-noise ratio, PSNR)、结构相似性<sup>[7]</sup> (structural similarity, SSIM) 等图像质量评价的算法,支持对超分后图像的质量实时提供量化的评估结果。

从逻辑功能的角度,智能机顶盒主要由图像超分处理、图像全屏显示和图像处理计算等功能模块组成。其中图像超分处理模块主要是调用 OpenCV 的方法实现对图像的插值或 AI 超分辨率处理 (C 语言实现),图像全屏显示模块主要实现对超分图像的全屏显示功能 (Java 语言实现),图像处理计算模块主要用于对超分后图像的 PSNR、SSIM 指

标进行计算,从而给出图像质量的客观评价数据。

智能机顶盒的图像超分辨实现流程如图 4 所示。

## 3 实验结果

### 3.1 质量评价方法

图像质量评价方法通常分为客观评价和主观评价两类。客观评价<sup>[11]</sup>是模拟人对图像质量的感知过程,构建与主观评价结果相一致的评价算法;而主观评价是以人作为观测者,对图像进行主观评价,力求能够真实地反映人的视觉感知。

本文实验所采用的质量评价方法主要是全参考客观评价 PSNR 与 SSIM 指标以及主观绝对评价平均意见得分 (mean opinion score, MOS) 指标。MOS 是衡量图像质量最具代表性的主观评价方法,它通过对观察者的评价归一判断图像质量,实验主要采用双刺激连续质量分级法的方式,即将待评价图像和原始图像按一定规则交替播放持续一定时间给观察者,观察者根据自己的知识和理解对图像进行打分,将所有分数取平均得到主观评价值。客观评估 PSNR 为该领域应用最早和最广的参数之一,它用于衡量图像噪声水平或图像有损变换 (图像压缩、图像修复) 重建质量的客观评价;通过图像之间的最大可能像素值 (通常为 255) 和 MSE 进行定义,正常情况下 PSNR 的范围为 20~40 dB,越接近 40 dB,表示图像重建的质量越好。SSIM 为一种衡量两幅图像相似度的客观指标,它与主观体

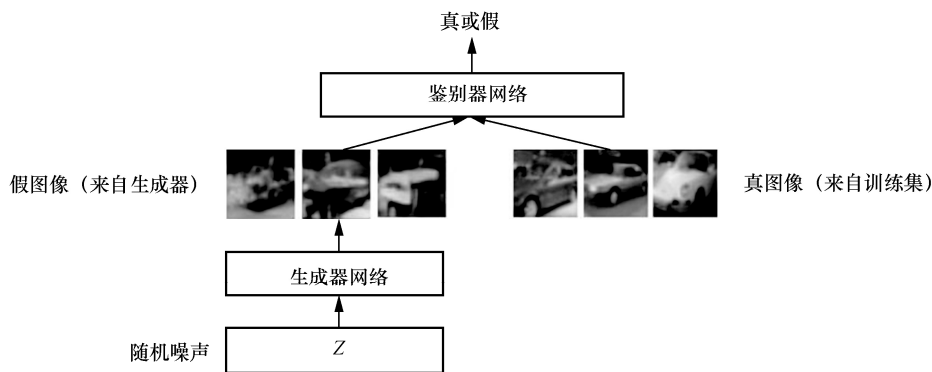


图 3 对抗网络典型结构示意图

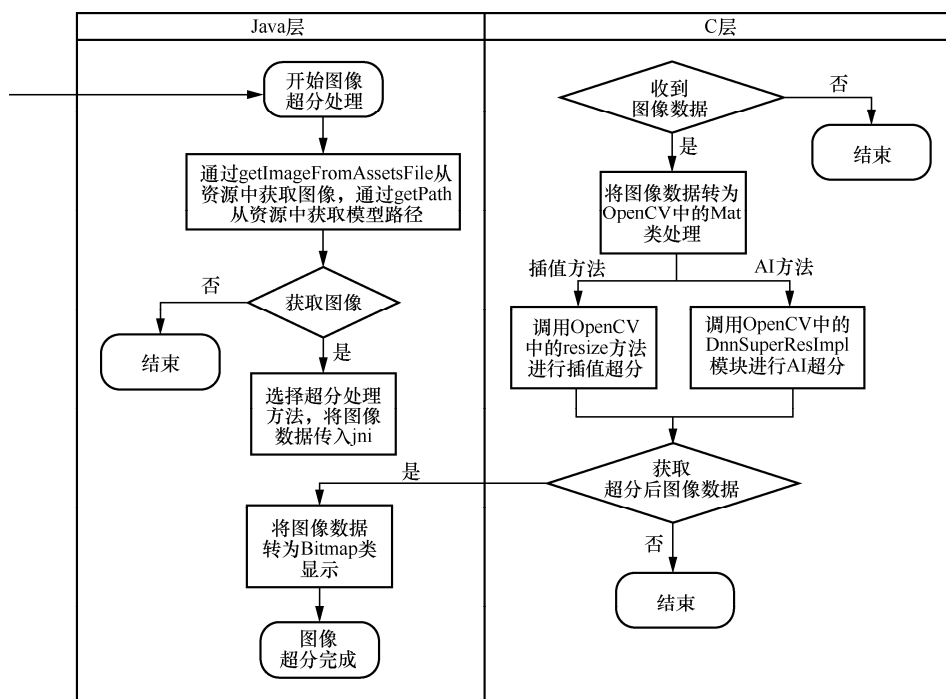


图4 智能机顶盒的图像超分辨率实现流程

验的一致性最高, 已成为当前全参考图像质量评价的基准算法; 它从人类视觉系统 (human visual system) 的角度出发, 从图像中提取结构信息, 将亮度 (均值)、对比度 (方差) 和结构这 3 个相对独立的层次进行比较, SSIM 的取值范围为 0~1, 数值越大, 表示结构相似性越高, 重建效果越好。

### 3.2 实验方法说明

实验采用一款 8K 智能机顶盒终端, 按照第 2 节的方案进行了超分辨率重建的功能实现, 集成了以下对比实验所使用的各种插值算法、深度学习算法以及客观质量评价算法。机顶盒的运行环境为 CPU (8 核 A73)、GPU (6 核 G52)、NPU (int4 9 T 算力)、6 GB (LPDDR4); 机顶盒的软件版本为 OpenCV 4.5.0 + OpenCV\_contrib4.5.0。

实验选取了分辨率为 576P、720P、1080P 的多组图像数据集 (每项实验采用的图片数量均大于 30 张), 运行不同算法对图集进行了超分辨率重建, 针对不同超分辨率算法和不同超分比例的多种情况, 给出图像超分辨处理对终端性能消耗、

图像质量等方面平均表现的评估。实验采用以下几种典型的插值算法和深度学习算法进行比较: 插值算法包括最邻近插值、双线性插值、双三次插值、兰索斯插值, 深度学习算法包括快速超分辨率卷积神经网络 (fast super resolution convolutional neural network, FSRCNN)、小型快速超分辨率卷积神经网络 (small FSRCNN, FSRCNN-s)、高效子像素卷积神经网络 (efficient sub-pixel convolutional neural network, ESPCN) 和拉普拉斯金字塔超分辨率网络 (Laplacian pyramid network for fast and accurate super-resolution, LAPSRN)。

### 3.3 实验数据

#### 3.3.1 插值算法的比较

各种插值算法的特性比较见表 1。

本实验选取了分辨率为 576P 和 1080P 的数据集, 分别超分成 1080P 和 4K 的高清图像, 720 pixel×576 pixel 超分至 1920 pixel×1080 pixel 和 1920 pixel×1080 pixel 超分至 4096 pixel×2160 pixel 插值算法的性能与质量结果分别见

表 1 各种插值算法的特性比较

| 插值算法  | 灰度连续情况 | 相邻像素的影响 | 锯齿现象         | 运算速度 |
|-------|--------|---------|--------------|------|
| 最邻近插值 | 不连续    | 未考虑     | 不可避免, 有块状效应  | 最快   |
| 双线性插值 | 连续     | 未考虑     | 可避免, 轻微马赛克效应 | 较快   |
| 双三次插值 | 连续     | 考虑      | 可避免          | 一般   |
| 兰索斯插值 | 连续     | 考虑      | 可避免          | 较慢   |

表 2 插值算法的性能与质量结果 (720 pixel×576 pixel 超分至 1 920 pixel×1 080 pixel)

| 插值算法        | 评价指标      |        |          |             |     |
|-------------|-----------|--------|----------|-------------|-----|
|             | 处理用时/s    | CPU 资源 | PSNR     | SSIM        | MOS |
| 最邻近插值       | 0.007 121 | 22%    | 30.261 5 | 0.845 27    | 1   |
| 双线性插值       | 0.011 045 | 24%    | 32.238 5 | 0.887 91    | 2   |
| 双三次插值 (4×4) | 0.011 437 | 37%    | 32.883 6 | 0.900 796   | 3   |
| 兰索斯插值 (8×8) | 0.011 511 | 35%    | 32.965 5 | 0.902 504 1 | 4   |

表 3 插值算法的性能与质量结果 (1 920 pixel×1 080 pixel 超分至 4 096 pixel×2 160 pixel)

| 插值算法        | 评价指标      |        |          |           |     |
|-------------|-----------|--------|----------|-----------|-----|
|             | 处理用时/s    | CPU 资源 | PSNR     | SSIM      | MOS |
| 最邻近插值       | 0.016 045 | 28%    | 35.622   | 0.959 338 | 1   |
| 双线性插值       | 0.024 829 | 29%    | 36.175 8 | 0.963 938 | 2   |
| 双三次插值 (4×4) | 0.029 973 | 36%    | 37.268 5 | 0.966 419 | 3   |
| 兰索斯插值 (8×8) | 0.031 027 | 40%    | 37.353   | 0.966 267 | 4   |

表 2 和表 3, 为采用不同插值算法实现单图像超分辨率的处理性能和图像质量等方面的平均表现。

从表 2 和表 3 可以得到以下内容。

(1) 在本实验中, 超分辨率的处理时长与原图分辨率、目标分辨率成正相关, 分辨率越大耗时越长。

(2) 4 种插值算法中, 最近邻插值和双线性插值的图像质量较差, 仅适用于实时性要求特别高但处理能力非常有限的场景。使用双三次插值和兰索斯插值的图像质量较好, 这两者相比, 双三次插值的运算速度快, 而兰索斯插值的质量更胜一筹。

(3) 实验所用的智能机顶盒和插值算法 (尤其是兰索斯算法), 将标清图像 (720 pixel×576 pixel) 超分至高清图像 (1 920 pixel×1 080 pixel) 所用的处理时长均在 12 ms 以内 (小于 16 ms), 适用于 60 f/s (f/s 为画面每秒传输帧数) 及以下的视频进行实时超分处理的场景, 如高清直播业务。而对于从高清图

像 (1 920 pixel×1 080 pixel) 超分到超高清图像 (4 096 pixel×2 160 pixel) 的情况, 由于处理用时超过 32 ms (大于 16 ms 且小于 33 ms), 对高于 30 f/s 的视频可能引起内容播放过程中的卡顿、丢帧等现象, 不适用于此类高帧率实时视频处理的场景, 但可用于普通帧率 (30 f/s 及以下) 的实时超分处理, 也可用于高帧率视频内容的超分预处理的场景, 如点播业务、本地播放业务等。

### 3.3.2 深度学习算法的比较

本实验选取分辨率为 576P 和 1 080P 的数据集, 分别超分成 1 080P 和 4K 的高清图像, 720 pixel×576 pixel 超分至 1 920 pixel×1 080 pixel 和 1 920 pixel×1 080 pixel 超分至 4 096 pixel×2 160 pixel 深度学习的性能与质量结果分别见表 4 和表 5, 为采用不同深度学习算法实现单图像超分辨率的处理性能和图像质量等方面的平均表现。

从表 4 和表 5 可以得到以下内容。



表 4 深度学习的性能与质量结果 (720 pixel×576 pixel 超分至 1 920 pixel×1 080 pixel)

| 深度学习算法   | 评价指标      |        |          |           |     |
|----------|-----------|--------|----------|-----------|-----|
|          | 处理用时/s    | CPU 资源 | PSNR     | SSIM      | MOS |
| FSRCNN-s | 0.120 766 | 45%    | 34.237 4 | 0.948 478 | 1   |
| FSRCNN   | 0.310 031 | 49%    | 35.025 7 | 0.959 921 | 4   |
| ESPCN    | 0.432 786 | 52%    | 34.565 7 | 0.960 559 | 3   |
| LAPSRN   | 5.858 17  | 62%    | 34.217 6 | 0.953 263 | 2   |

表 5 深度学习的性能与质量结果 (1 920 pixel×1 080 pixel 超分至 4 096 pixel×2 160 pixel)

| 深度学习算法   | 评价指标     |        |          |           |     |
|----------|----------|--------|----------|-----------|-----|
|          | 处理用时/s   | CPU 资源 | PSNR     | SSIM      | MOS |
| FSRCNN-s | 5.430 97 | 60%    | 36.079 5 | 0.966 591 | 1   |
| FSRCNN   | 14.416 7 | 80%    | 36.932 9 | 0.971 221 | 4   |
| ESPCN    | 21.194 9 | 85%    | 36.441 1 | 0.970 014 | 3   |
| LAPSRN   | 29.024 0 | 88%    | 36.127 1 | 0.969 024 | 2   |

(1) 在本实验中, 采用不同深度学习算法实现的图像超分辨率, 在客观与主观质量方面存在较大的差异, 造成这种差异的原因在于训练集、训练网络等因素的不同。

(2) 4 种深度学习算法中, FSRCNN、ESPCN<sup>[4]</sup>算法的图像质量较好, 但两者相比, ESPCN 算法对于终端的性能消耗更大, 图像质量略微优于 FSRCNN 算法。

(3) 实验所用的智能机顶盒和深度学习算法, 不论将标清图像超分至高清图像还是高清图像超分至超高清图像的情况, 终端所用处理时长基本超过 300 ms, 因此, 不适用于实时超分处理的场景, 但可用于对视频内容进行超分预处理后再进行播放的场景。

### 3.3.3 插值算法和深度学习算法的比较

本实验选取分辨率为 576P、720P、1 080P 的数据集, 采用兰索斯插值、FSRCNN 与 ESPCN 算法将

原始图像超分至 2 倍或 3 倍大小, 插值和深度学习算法的性能结果见表 6, 为采用 3 种算法实现单图像超分辨率的处理性能和图像质量等方面的平均表现。但由于图集中有部分原始图像缺失, 因此表 6 没有给出图像质量均值的统计, 仅做了相关的分析描述。

从表 6 可以得到以下内容。

(1) 在本实验中, 超分辨率的处理时长与图像原始分辨率、目标分辨率及算法的设计均表现有相关性, 但非线性相关。横向比较下, 插值算法与目标分辨率的相关性权重更大; 而基于深度学习的算法, 其性能与训练数据集、训练网络等算法设计因素相关的权重最大, 而与图像原始分辨率、目标分辨率相关性权重次之。

(2) 插值算法和深度学习算法相比而言, 兰索斯插值算法的运算速度具有明显优势, 深度学习算法相比插值算法速度慢 2 个数量级以上。

(3) 针对部分有原始图像的样本, 实验进行

表 6 插值和深度学习算法的性能结果

| 分辨率                                             | 平均处理用时/s |        |        |
|-------------------------------------------------|----------|--------|--------|
|                                                 | 兰索斯插值    | FSRCNN | ESPCN  |
| 720 pixel×576 pixel→1 440 pixel×1 152 pixel     | 0.011    | 7.621  | 10.868 |
| 1 280 pixel×720 pixel→3 840 pixel×2 160 pixel   | 0.052    | 4.001  | 5.611  |
| 1 920 pixel×1 280 pixel→3 840 pixel×2 560 pixel | 0.051    | 8.592  | 11.950 |

了 PSNR/SSIM 指标计算和主观评分,结果显示,采用兰索斯插值算法的图像客观质量优于深度学习算法,如针对图像放大倍数为 3 的实验,实际测得兰索斯插值算法的 PSNR/SSIM 约为 35/0.9,深度学习算法的 PSNR/SSIM 约为 30/0.75;同样针对这种放大倍数超过 2 的情况,采用兰索斯插值算法的主观评分下降明显,如实际测得 MOS 为 3,而采用深度学习算法的主观体验优于兰索斯插值算法,如实际测得 MOS 为 3.5~4。

(4) 相比传统方法,采用深度学习算法处理的图像的主观体验更优,尤其对于人眼比较敏感的图像信息,如人脸、字幕、亮度、景深区域等,表现出更强的细节还原能力,使得人为感知提升效果明显。

(5) 相比表 4 和表 5 的实验数据,深度学习算法在超分处理时长、图像质量等方面表现出一定的差异和不稳定性,造成这种差异的原因主要在于训练集的不同,深度学习算法比较依赖于数据集的选取,对于不同的图像特征,如图像类型(如自然图像、计算机绘制图像等)、表达内容(如自然景观、人物、字幕等)、图像格式等,超分效果呈现出较大的差异。

720 pixel × 576 pixel 超分至 4 096 pixel × 2 160 pixel 的图像效果如图 5 所示,展示了 720P 图像分别采用插值算法和深度学习算法产生的 4K 分辨率图像效果。

### 3.4 实验小结

基于第 4.3 节实验数据的比较分析发现,通过 8K 智能机顶盒实现的几种超分辨率方法中,兰索斯插值算法相对轻量、图像质量稳定较好,采用兰索斯插值算法的智能终端具备将 60 f/s 及以下

的视频内容实时超分为 1 080P 分辨率的能力,对于 4K 目标分辨率,智能终端只能支持对 30 f/s 及以下的视频内容的实时超分处理。相对而言,深度学习算法的超分处理用时长、受训练数据集或训练网络的影响较大,在算力资源有限的情况下,采用深度学习算法的智能终端不具备对视频内容进行实时超分处理的能力,仅能应用在视频内容离线预处理的业务场景。

## 4 场景与应用

### 4.1 视频业务典型应用场景

利用超分辨率技术能够在不升级现有采集设备的情况下提高图像的分辨率,也能在不增加传输带宽的情况下改善图像或视频的画面质量,在遥感影像、医学成像、自动驾驶、安防监控<sup>[12]</sup>等领域得到广泛应用。本文主要给出在娱乐视频、家庭视频监控中的几种典型应用场景。

#### (1) 影视作品老片新看

广播电视、IPTV 业务都经历了从标准、高清到超高清的发展过程,早期一些经典的影视作品拍摄于 20 世纪 80 年代到 90 年代(如央视的四大名著电视剧),当时内容采集所用设备是模拟标清摄像机,采用磁带作为存储介质进行保存,尽管内容制作精良,但如今通过 4K/8K 电视机进行播放,必然出现清晰度差、脱磁、边缘锯齿、色彩失真等问题,使作品的艺术表现力大打折扣。因此,在业务平台侧对存量的电视节目内容进行超分辨率修复处理,可以把原本模糊、灰暗的视频内容变得清晰、明亮,同时消除掉人



图5 720 pixel×576 pixel 超分至 4 096 pixel×2 160 pixel 的图像效果





物伪影、文字毛刺等瑕疵,使好的影视作品成为永恒的经典。

## (2) 图像/视频的实时超分

随着终端硬件性能与应用处理能力的提升,目前一些高端配置的终端,如手机、智能电视等,支持端侧实时超分的技术<sup>[13]</sup>,能够在不增加网络开销的情况下解决内容分辨率低、网络传输不稳定等问题。例如,当前流量占比较高的用户自制的短视频或者网络直播,内容的采集较多使用的是手机终端(竖屏),分辨率以 720P 为主,当这些内容在大屏幕终端上进行播放,清晰度明显不够;随着 8K 智能电视机的面世,除了 8K 视频内容的稀缺,更主要的是绝大部分用户界面的分辨率都在 4K 以下,导致节目海报、文字等内容在大尺寸电视屏上的显示效果不理想;用户使用多屏互动功能将一些老照片投屏至电视屏上进行回放,或者用户想检查家庭监控视频中的某个内容片段,由于录制内容的分辨率低而无法满足用户辨识图像细节的需求。在这些场景中应用实时超分辨率技术,可有效提高画面清晰度,提升用户体验。

## 4.2 与相关技术的融合

将超分辨率技术与图像增强<sup>[14]</sup>、视频插帧、HDR<sup>[15]</sup>等技术相结合,能进一步地提高视频画质。其中,图像增强技术是以提高图像质量和辨识度为目的,有选择地突出图像中的某些特征而抑制其他一些特征,如对边缘信息、轮廓信息、对比度做突出处理。视频插帧技术则以提高视频画面的连贯性和流畅度为目标,它通过计算视频的帧间运动关系,生成可在两帧图像之间插入的符合运动关系的中间帧,在不增加内容制作成本的同时有效提高视频帧率,特别适用于快速变化的视频画面或放慢镜头时的视频播放等场景,视频插帧技术可以看作一种在时间序列上的超分辨率复原。HDR 成像技术的目的是还原图像的亮度层次范围以达到与真实世界中太阳光照射相一致的亮

度效果,它通过将多幅不同曝光范围的图像信息进行融合与处理,获得更广泛的色彩范围,提升最亮和最暗画面的对比度,HDR 技术可以看作是一种在像素亮度范围上的超分辨率复原。

基于压缩域的超分辨率技术<sup>[16]</sup>是另外一种融合应用。对于一些因传输或存储严重受限但图像质量要求较高的场景,图像压缩方法与超分辨率方法有机结合,能有效缓解由压缩算法引起的图像/视频视觉质量的下降,它通过在压缩过程中增加一些用于超分辨率处理的有用先验信息,解决对于压缩图像或视频的超分辨率复原问题。

## 5 结束语

随着用户对高清晰度的不断追求,超分辨率技术因其成本低、兼容性强等技术优势在越来越多的领域得到应用。理论上超分辨率可实现的方法种类繁多,对于传统方法而言,定量指标的提升越来越困难;结合深度学习算法的超分辨率技术,受制于不同应用场景下的训练数据集、运行终端的性能和算力能力等因素的影响,在超分处理性能、图像质量效果等方面表现出较大的差异,因此,进一步探索优越和高效的深度学习算法仍是图像处理领域的主要研究方向。

## 参考文献:

- [1] AGUSTSSON E, TIMOFTE R. NTIRE 2017 challenge on single image super-resolution: dataset and study[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE Press, 2017: 1122-1131.
- [2] TIMOFTE R, AGUSTSSON E, GOOL L V, et al. NTIRE 2017 challenge on single image super-resolution: methods and results[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE Press, 2017: 1110-1121.
- [3] CHEN C, XIONG Z W, TIAN X M, et al. Camera lens super-resolution[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 1652-1660.
- [4] SHI W Z, CABALLERO J, HUSZÁR F, et al. Real-time single

- image and video super-resolution using an efficient sub-pixel convolutional neural network[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2016: 1874-1883.
- [5] 黄健, 赵元元, 郭苹, 等. 深度学习的单幅图像超分辨率重建方法综述[J]. 计算机工程与应用, 2021, 57(18): 13-23.  
HUANG J, ZHAO Y Y, GUO P, et al. Survey of single image super-resolution based on deep learning[J]. Computer Engineering and Applications, 2021, 57(18): 13-23.
- [6] SCHULTZ R R, STEVENSON R L. A Bayesian approach to image expansion for improved definition[J]. IEEE Transactions on Image Processing, 1994, 3(3): 233-242.
- [7] 刘文星. 基于深度学习的单图像及视频超分辨率重建算法研究[D]. 重庆: 重庆理工大学, 2022.  
LIU W X. Research on super-resolution reconstruction algorithm of single image and video based on deep learning[D]. Chongqing: Chongqing University of Technology, 2022.
- [8] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all You need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM Press, 2017: 6000-6010.
- [9] WANG X T, YU K, DONG C, et al. Recovering realistic texture in image super-resolution by deep spatial feature transform[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 606-615.
- [10] SAJJADI M S M, SCHÖLKOPF B, HIRSCH M. EnhanceNet: single image super-resolution through automated texture synthesis[C]//Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2017: 4501-4510.
- [11] 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: a Publication of the IEEE Signal Processing Society, 2004, 13(4): 600-612.
- [12] YANG F Z, YANG H, FU J L, et al. Learning texture transformer network for image super-resolution[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 5790-5799.
- [13] HE Z L, HUANG H X, JIANG M, et al. FPGA-based real-time super-resolution system for ultra high definition videos[C]//Proceedings of 2018 IEEE 26th Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM). Piscataway: IEEE Press, 2018: 181-188.
- [14] WEI P X, XIE Z W, LU H N, et al. Component divide-and-conquer for real-world image super-resolution[C]//European Conference on Computer Vision. Cham: Springer, 2020: 101-117.
- [15] ZENG H M, ZHANG X L, YU Z B, et al. SR-ITM-GAN: learning 4k UHD HDR with a generative adversarial network[J]. IEEE Access, 2020, 8: 182815-182827.
- [16] LIU Z, CUI C. A new low bit-rate coding scheme for ultra high definition video based on super-resolution[C]//2018 IEEE International Conference on Computer and Communication Engineering Technology (CCET). Piscataway: IEEE Press, 2018: 325-329.

#### [作者简介]



祝谷乔 (1976—), 女, 天翼数字生活科技有限公司高级工程师, 主要研究方向为 IPTV 关键技术与标准化。



姜超 (1985—), 男, 中国信息通信科技集团有限公司正高级工程师, 主要研究方向为多媒体通信、光纤通信与智能终端等。



徐煜桦 (1992—), 女, 中国信息通信科技集团有限公司工程师, 主要研究方向为流媒体传输协议、音/视频编解码等。