

用于超分辨率重建的深度网络递进学习方法

张毅锋^{1,2,3} 刘袁¹ 蒋程¹ 程旭⁴

摘要 本文针对深度学习在单幅图像超分辨率方面难以恢复高频纹理细节的问题,提出了一种基于递进学习的超分辨率算法.该算法首先采用灰度共生矩阵提取图像纹理特征,然后利用基于密度峰值的聚类方法实现对整个训练集的分类,其中每个训练子集具有相似的纹理复杂度.针对传统的递进学习方法会出现对已掌握知识“遗忘”的问题,本文根据网络模型在各个训练子集上的拟合情况,实时调整当前训练样本在各个子集上的概率分布,从而实现快速收敛,并获得更好的纹理细节复原效果.将本文提出的递进学习用于 DRCN、VDSR、SRCNN 等超分辨率网络的训练,实验结果表明超分辨率网络收敛速度得到提升,同时网络对复杂纹理等细节较多的图像也获得了较好的视觉恢复效果,峰值信噪比则平均获得 0.158 dB、0.18 dB、0.092 dB 的提升.

关键词 超分辨率,递进学习,共生矩阵,密度峰值

引用格式 张毅锋,刘袁,蒋程,程旭.用于超分辨率重建的深度网络递进学习方法.自动化学报,2020,46(2):274–282

DOI 10.16383/j.aas.2018.c180158

A Curriculum Learning Approach for Single Image Super Resolution

ZHANG Yi-Feng^{1,2,3} LIU Yuan¹ JIANG Cheng¹ CHENG Xu⁴

Abstract The main challenge of single image super resolution (SISR) is the recovery of high frequency details such as tiny textures. In order to solve this problem, a curriculum learning-based approach is proposed. In this paper, firstly, gray-level co-occurrence matrix is applied to extract texture features of images. Then, with the clustering algorithm based on density peaks, the training dataset is divided into several subsets on the basis of texture features. As for traditional curriculum learning, the performance is easy to get worse due to the “forgotten” phenomenon of knowledge that has been learned. Different from this, training examples are sampled from all the subsets based on the slope of the learning curve of each subsets. It is helpful for the speed of convergence and the recovery of high frequency details. Experiments show that when SISR networks such as DRCN, VDSR, SRCNN are trained with curriculum learning, training time is shortened and the visual effect is improved. The PSNR (Peak signal to noise ratio) values are increased by 0.158 dB, 0.18 dB and 0.092 dB, respectively.

Key words Super resolution, curriculum learning, co-occurrence matrix, density peaks

Citation Zhang Yi-Feng, Liu Yuan, Jiang Cheng, Cheng Xu. A curriculum learning approach for single image super resolution. *Acta Automatica Sinica*, 2020, 46(2): 274–282

收稿日期 2018-03-20 录用日期 2018-09-10
Manuscript received March 20, 2018; accepted September 10, 2018

国家自然科学基金(61673108, 61802058),江苏省自然科学基金(BK20151102),北京大学机器感知与智能教育部重点实验室开放课题(K-2016-03),东南大学水声信号处理教育部重点实验室开放项目(UASP1502)资助

Supported by National Natural Science Foundation of China (61673108, 61802058), National Natural Science Foundation of Jiangsu Province (BK20151102), Opening Project of Key Laboratory of Machine Perception, Peking University (K-2016-03), Opening Project of Key Laboratory of underwater acoustic signal processing, Southeast University (UASP1502)

本文责任编辑 王亮

Recommended by Associate Editor WANG Liang

1. 东南大学信息科学与工程学院 南京 210096 2. 东南大学南京通信技术研究院 南京 211100 3. 计算机软件新技术国家重点实验室(南京大学) 南京 210093 4. 南京信息工程大学计算机与软件学院 南京 210044

1. School of Information Science and Engineering, Southeast University, Nanjing 210096 2. Nanjing Institute of Communications Technologies, Southeast University, Nanjing 211100 3. State Key Laboratory for Novel Software Technology, Nan-

单幅图像超分辨率 (Single image super resolution, SISR) 重建的任务是由一幅低分辨率图像 (Low resolution image, LR) 去推断最为可能的高分辨率图像 (High resolution image, HR). 由于 LR 丢失了大量高频信息,理论上恢复出的高分辨率图像存在多种可能.算法的目标是尽可能多地恢复纹理细节等高频信息,找寻最为可能的 HR.

基于深度学习的算法在 SISR 上获得了巨大成功.其中 SRCNN^[1] 通过训练一个浅层网络来实现纹理细节的恢复.胡长胜等^[2] 则针对信息过压缩和网络收敛速度等问题对 SRCNN 做出改进,获得视觉效果上的改善.文献 [3] 和 [4] 则分别在深度学习和噪声图像的超分辨率上做了进一步探索.浅层网络的缺点之一是感受野很小,因而也限制了网络的

jing University, Nanjing 210093 4. School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044

表达和泛化能力. VDSR^[5]、SRResNet^[6]、EDSR^[7]等则采用深层网络来获取高分辨率图像. 损失函数的设计也获得越来越多的重视. 传统的均方误差函数 (Mean square error, MSE) 虽然有利于获得很高的峰值信噪比 (Peak signal to noise ratio, PSNR), 但恢复出的图像容易丢失纹理等细节信息, 视觉上趋于模糊. EnhanceNet^[8] 通过感知损失来恢复 LR 中丢失的高频信息, 获得了较好的视觉效果.

这些方法都是尝试在“学”这方面进行改进, 而深度学习的训练是一件较为困难的事情, 通常并不能收敛到较好的局部极小值. 本文尝试在“教”这个领域做出改进, “老师”通过递进学习 (Curriculum learning, CL)^[9] 为“学生” (待训练的神经网络) 合理安排课程 (读取训练样本的先后顺序), 先拟合简单的训练样本, 再拟合困难样本, 不仅可以加速收敛, 网络模型也有望收敛到接近全局最优解的局部极小值. 递进学习目前在语言模型生成^[10]、长短时记忆网络^[11] 的训练等方面都取得了不错的效果, 但对于传统的递进学习方法, 已拟合的简单样本在训练后期由于很少参与到模型训练中, 网络模型往往会出现对已学知识“遗忘”的现象.

在 SISR 中, 纹理细节多的图像由于不易恢复属于较为困难的样本, 整体平滑的图像则属于简单样本. 在课程设置方面, 本文提出通过灰度共生矩阵^[12] 提取图像纹理特征, 随后采用基于密度峰值的聚类方法^[13] 将训练集分成多个子集, 每一个子集的图片纹理复杂度比较接近, 这些子集就是相应的课程. 针对模型对已学课程“遗忘”的问题, 本文依据模型在不同训练子集上的学习曲线的斜率来选取相应的训练数据. 斜率越大, 在相应的子集上抽取训练样本的概率就越高, 但每个训练子集都有一定概率参与到网络模型训练中. 这样避免了人为设定训练次序造成的不稳定, 同时也抑制了对已掌握知识的遗忘, 从而保证训练过程始终朝着负梯度方向优化. 本文采用改进的递进学习训练 DRCN^[14]、VDSR^[5]、SRCNN^[1] 等 SISR 网络, 实验表明本文提出的改进的递进学习方法在加速网络模型收敛的同时, 模型的超分辨率效果也获得进一步提升.

1 基于递进学习的网络训练

1.1 递进学习的重要性

基于深度学习的超分辨率算法通常是随机选择训练样本来训练模型. 本小节将论述合理地选择训练样本有利于网络收敛. 设网络模型在时刻 t 的参数为 M_t , 网络全局最优解对应的参数为 M^* . 当前网络的训练样本对记为 (x_t, y_t) , $t + 1$ 时刻的参数

M_{t+1} 如式 (1) 所示.

$$M_{t+1} = M_t - \eta \frac{\partial \Phi(\langle M_t, x_t \rangle, y_t)}{\partial M_t} \quad (1)$$

其中, η 是学习率, 函数 $\Phi(\cdot)$ 是损失函数. 相应地, M_{t+1} 与 M^* 的距离定义为

$$\begin{aligned} \|M_{t+1} - M^*\|_2^2 &= \\ \left\| M_t - \eta \frac{\partial \Phi(\langle M_t, x_t \rangle, y_t)}{\partial M_t} - M^* \right\|_2^2 &= \\ \|M_t - M^*\|_2^2 + \eta^2 Q_1 - 2\eta Q_2 & \end{aligned} \quad (2)$$

其中,

$$\begin{aligned} Q_1 &= \left\| \frac{\partial \Phi(\langle M_t, x_t \rangle, y_t)}{\partial M_t} \right\|_2^2 \\ Q_2 &= \left\langle M_t - M^*, \frac{\partial \Phi(\langle M_t, x_t \rangle, y_t)}{\partial M_t} \right\rangle \end{aligned} \quad (3)$$

神经网络迭代的优化目标是

$$\min \|M_{t+1} - M^*\|_2^2 \quad (4)$$

即寻找合适的训练样本 (x_t, y_t) , 满足

$$\arg \min_{(x_t, y_t)} (\eta^2 - 2\eta Q_2) \quad (5)$$

这样可以使得模型参数沿着负的梯度方向逼近 M^* . Q_1 值反映当前损失函数下梯度的大小. 如果所选择的训练样本 (x_t, y_t) 已被网络模型较好地拟合, 那么 Q_1 值会较小. Q_2 值反映梯度与 $M_t - M^*$ 的匹配程度, 与训练样本所含信息量有关. 通常称网络模型难以拟合的样本为困难样本, 容易拟合的样本为简单样本. 网络模型对样本的拟合能力可反映出样本所含信息的多少, 网络模型难以拟合的样本通常包含的信息比较多, 反之亦然. 由此可知困难样本含有比较丰富的信息, 基于其计算所得的梯度与 $M_t - M^*$ 的匹配程度较高.

在训练的不同阶段,

$$\frac{\partial \Phi(\langle M_t, x_t \rangle, y_t)}{\partial M_t} \quad (6)$$

式 (6) 取值不同, 在训练初期该值比较大, 在训练后期该值比较小. 所以为了模型可以快速收敛, 在训练的不同阶段, 优化目标有所不同. 在训练初期, 最小化 Q_1 值是主要任务, 即选择简单的训练样本进行训练. 在训练后期, 最大化 Q_2 值是主要任务, 此时应选取较为困难的训练样本进行训练. 因而神经网络通过递进学习的方法可以更快地收敛, 并且更有可能收敛到更优解.

1.2 图像纹理特征提取

将训练集划分成不同的子集是递进学习的前提。SISR 用于恢复纹理信息, 因而本文将依据图像纹理的分布来进行分类。纹理特征提取方法大致分为四类: 统计方法、模型方法、信号处理方法和结构方法。统计方法中的灰度共生矩阵法^[12] 可较好地描述各像素值的空间分布特性以及纹理特征, 有较强的适应性和鲁棒性。相比较而言, 基于模型的方法主要通过模型参数来衡量图像纹理特征, 在参数求解上有较大困难; 信号处理的方法除了计算量大之外, 小波包分析对非规则纹理不能很好地提取特征, 正交小波变换的多分辨率分解则往往忽略高频部分从而导致信息丢失; 结构方法则通常适用于有规则的人造纹理, 对不规则的自然纹理效果较差。因而本文采用灰度共生矩阵来提取图像纹理信息。

灰度共生矩阵定义如下: 设尺寸为 $L \times M$ 的图像的灰度值范围为 $(0, W)$, 图像上坐标偏差为 $(\Delta x, \Delta y)$ 的两个点灰度值分别为 i, j 的概率记为 $P(i, j)$, 该概率由图像灰度值的整体分布决定。由此获得大小为 $W \times W$ 的矩阵 G , 其中 $G(i, j) = P(i, j)$ 。矩阵 G 通常被称为灰度共生矩阵, 该矩阵反应了图像纹理的整体情况。

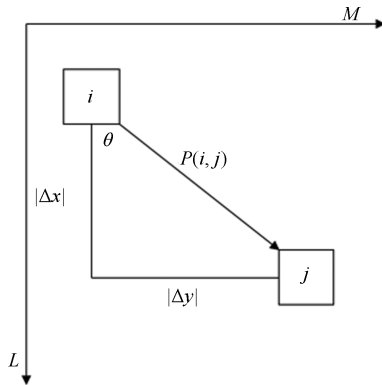


图 1 灰度共生矩阵示意图

Fig. 1 The schematic diagram of gray-level co-occurrence matrix

灰度共生矩阵维度较大, 这不利于后续处理。通常情况下, 可以通过提取特征量来实现降维, 本文提取以下 5 种特征: 纹理能量 Ene 、纹理均匀性 $Homo$ 、纹理对比度 Con 、纹理的熵 Ent 以及纹理相关性 Cor , 以此来降维。这 5 个特征量反映了图像的纹理复杂度、一致性、均匀程度等信息。本文以 5 维数据 $(Ene, Homo, Con, Ent, Cor)$ 来表征一幅图像的纹理信息。

1.3 训练集的划分

本文采用基于密度峰值的聚类方法^[13] (Density peaks, DP) 将训练集按照纹理信息分成多个子集。

本文首先基于余弦相似度来定义图像之间的纹理相似度。设图像 I_1, I_2 的纹理特征分别为

$$\begin{cases} Fea1 = (Ene1, HOMO1, Con1, Ent1, Cor1) \\ Fea2 = (Ene2, HOMO2, Con2, Ent2, Cor2) \end{cases} \quad (7)$$

则它们的余弦相似度为:

$$Sim(I_1, I_2) = \frac{Fea1 Fea2^T}{\|Fea1\| \cdot \|Fea2\|} \quad (8)$$

对于图像 I_i , 其密度定义为:

$$\rho_i = \sum_{I_i \in I_s \setminus I_i} \psi(Sim(i_i, I_i) - \gamma) \quad (9)$$

其中,

$$\psi(s) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} \quad (10)$$

γ 根据实际情况设定的阈值。所以图像 I_i 的密度 ρ_i 反映了训练集 I_s 中与 I_i 余弦相似度大于 γ 的图像数目。

DP 聚类算法的核心思想是每个子集的聚类中心密度最大, 每个中心都连接其周围密度相对低的点, 同时不同子集的聚类中心彼此相似度较小。主要步骤如下:

1) 筛选出密度大同时彼此相似度较小的多幅图像作为多个聚类中心。

2) I_s 中的所有图像连接到与其最相似且密度比它大的图像样本, 直至连接到聚类中心。

所有共享一个聚类中心的图像属于同一个训练子集, 这些图像在纹理复杂度、均匀程度、一致性等方面均基本保持一致。许多基于欧氏距离或曼哈顿距离的聚类算法, 通常只能发现相似尺寸和密度的球状簇或凸型类簇, 本文的 DP 算法采用余弦相似度定义图像之间的距离, 同时基于相对密度和相对相似度来连接点, 所以其在聚类过程中可以发现任意形状的簇, 并且对不同密度的训练子集也能很好地聚类。另一方面, DP 算法在有噪声的数据中仍然可以基于纹理特征对图像较好地分类, 为下一步的递进学习奠定了基础。

1.4 递进学习方法

递进学习方法在网络的训练中发挥着重要作用, 主要需要考虑以下三个问题:

1) 按照网络拟合难度将训练集分成不同的子集。

2) 设定阈值来衡量模型对每个子集的拟合程度。

3) 不断地将困难子集与简易子集混合来防止模型对已掌握知识的遗忘。

本文提出的基于灰度共生矩阵和密度峰值的聚类方法可以实现对训练集合理地划分. 在此基础上, 选择合适的递进学习方法对网络最终的训练结果有着重要影响. 通常的递进学习方法分为如下三类:

1) 递进学习 V1: 按照课程由易到难的顺序, 先拟合简单的数据集, 达到一定效果后转到下一个训练子集继续训练.

2) 递进学习 V2: 按照由易到难的顺序, 与递进学习 V1 的不同在于会不断复习已经掌握的课程, 即简单的数据集仍会用于网络参数的训练, 有利于避免网络模型对已掌握课程的遗忘.

3) 递进学习 V3: 将由易到难与在所有训练子集中随机抽取训练样本结合起来. 网络快速收敛的同时, 随机抽取的训练样本造成的波动有利于模型参数逼近全局最优解.

以上三种方法都是建立在训练子集的难度能够合理地排序这一基础之上. 递进学习对简单训练子集的定义是网络模型能够较容易地拟合其分布, 从而迅速提升在该数据集上的性能效果. 但很多情况下, 人为对训练子集难易程度所作的排序未必符合网络模型内在的规律. 不同的任务, 不同的网络架构对训练集难易的定义也会不同. 为避免课程设置造成的训练难以收敛等问题, 本文提出了新的递进学习方法, 具体步骤如下所示:

1) 训练初始化, 对 N 个训练子集以相等概率抽取样本来训练网络模型.

2) 每迭代 M 次, 分别测试网络模型在 N 个训练子集上的性能效果. 迭代 L 次时 ($L = K \times M$), 在第 i 个训练子集上最近的 K 次测试结果记为 $S_i = S_i^j, j = 1, \dots, K$. 为避免噪声的影响, 剔除最大值和最小值, 即:

$$S_i^* = \{S_i^j, j = 1, \dots, K\} \setminus \{\max(S_i), \min(S_i)\} \quad (11)$$

3) 最近 L 次迭代, 网络模型在训练集 i 上的学习曲线斜率定义为:

$$C_i = \frac{|\max(S_i^*) - \min(S_i^*)|}{K} \quad (12)$$

4) 重新分配在每个训练子集上的样本抽取概率:

$$P_i = \frac{C_i}{\sum_{k=1}^N C_k} \quad (13)$$

重复执行步骤 2) ~ 4) 至模型训练结束.

在训练集已经依据纹理特征分成 N 个训练子集的前提下, 本文提出的递进学习方法抛弃了人为对训练子集设置先后学习顺序, 而是依据训练时网络模型在各个子集上的学习曲线斜率来实时调整训

练样本在不同子集上的概率分布. 模型在特定训练子集上的超分辨率测试效果进步越快, 可以认为该子集简单容易拟合. 相应地, 模型会更多地从中提取样本迭代训练. 训练的中后期, 由于网络模型已拟合简单子集, 因而在这类子集上的测试效果提升较慢, 训练样本更多地转移到从难度较大的训练子集中提取. 通过依据网络模型的训练情况动态调整在不同子集上的样本抽取概率, 保证了训练过程始终朝着负梯度方向优化. 另一方面, 在整个训练周期, 所有的训练子集都对网络训练发挥作用, 一定程度上避免了对已学知识的“遗忘”.

2 实验与结果分析

2.1 数据集

本文使用 DIV2K^[15] 数据集来训练. 该数据集包含 1000 张不同的 2K 分辨率 RGB 图像. 其中 800 张用于训练, 100 张用于验证, 剩下的 100 张用于测试. 训练好的模型在四个公开的数据集上验证效果, 分别是 Set5^[16], Set14^[17], BSD100^[18] 和 Urban100^[19]. 对于数据集 Set5 和 Set14, 他们分别包含 5 张和 14 张图像. BSD100 是伯克利语义分割数据集 BSD300^[18] 的测试集. Urban100 则包含了 100 张富含多种实体建筑物结构元素的图片. 对于超分辨率的放大尺寸, 本文的实验分别进行了 2 倍、3 倍和 4 倍放大. 算法效果的评价准则采用了峰值信噪比 (Peak signal to noise ratio, PSNR) 和结构相似性 (Structural similarity index, SSIM) 两种指标.

2.2 训练细节以及参数设置

训练过程中, 神经网络的输入尺寸是 48×48 , 输出是 $48r \times 48r$, r 是超分辨率放大比例. 测试时, 输入的图像则可以是任意尺寸. 为了避免过拟合, 本文通过随机的水平、垂直翻转以及旋转来增加训练样本数量. 所有的训练图像都来自 DIV2K 数据集. 为了加速网络收敛, 像素值被归一化到 $[0, 1]$. 算法流程基于 Tensorflow^[20] 搭建.

2.3 递进学习训练效果对比

本文在经典的 SISR 神经网络上比较了传统训练方法 (Baseline)、递进学习 V1、递进学习 V2、递进学习 V3 以及本文算法下的训练效果. 图 2、3 分别是 DRCN^[14] 在 Set14 数据集, VDSR^[5] 在 Set5 数据集上的测试结果. 除了训练样本学习顺序上的不同, 模型结构, 学习率等超参数均保持一致. 可以看出, 采用递进学习 V1 的训练方法在训练后期会出现效果变差的现象. 由于专注于困难数据集的拟合, 对简单数据集出现了“遗忘现象”, 甚至最终在

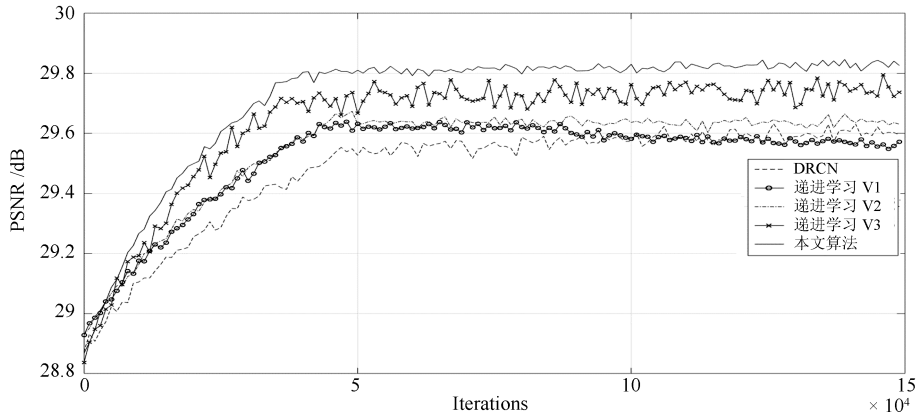


图 2 DRCN 在不同学习方法下的训练收敛情况 (测试集是 Set14, 放大倍数为 3 倍)

Fig. 2 The training performance of DRCN under different learning strategies on $\times 3$ SR (The test set is Set14)

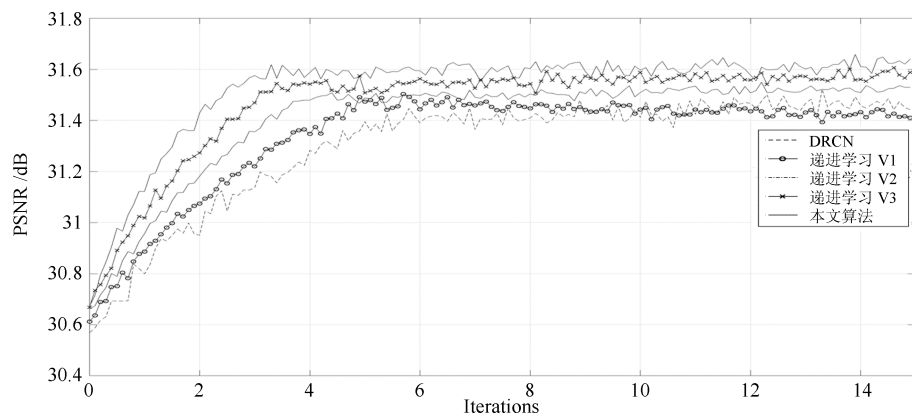


图 3 VDSR 在不同学习方法下的训练收敛情况 (测试集是 Set5, 放大倍数为 4 倍)

Fig. 3 The training performance of VDSR under different learning strategies on $\times 4$ SR (The test set is Set5)

PSNR 指标上的结果比常规的训练方法还要低. 采用递进学习 V2、递进学习 V3 训练网络模型时, 在遵循由易到难的同时, 不断“复习”已掌握的“课程”, 在测试数据集上获得了更高的 PSNR 值. 采用本文提出的递进学习方法, 相较于 Baseline, 平均获得了 0.32 dB 的 PSNR 值提升, 不仅收敛速度更快, 性能指标上也比递进学习 V2、递进学习 V3 要好. 证明了本文递进学习方法的有效性.

2.4 不同聚类方法效果比较

衡量聚类算法的性能通常包括紧密性指数 (Compactness, CP), 间隔性指数 (Separation, SP), DBI 指数 (Davies-bouldin index, DBI) 和 $Dunn$ 指数 (Dunn validity index, DVI). 聚类结果的子集划分不妨设为 $C = \{C_1, C_2, \dots, C_k\}$, 首先定义以下变量:

$$avg(C_i) = \frac{2}{|C_i|(|C_i| - 1)} \sum_{1 \leq i < j \leq |C_i|} dist(x_i, x_j) \quad (14)$$

$$diam(C_i) = \max_{1 \leq i < j \leq |C_i|} dist(x_i, x_j) \quad (15)$$

$$d_{\min}(C_i, C_j) = \min_{x_i \in C_i, x_j \in C_j} dist(x_i, x_j) \quad (16)$$

$$d_{cen}(C_i, C_j) = dist(u_i, u_j) \quad (17)$$

其中, 四个变量 $avg(C_i)$, $diam(C_i)$, $d_{\min}(C_i, C_j)$, $d_{cen}(C_i, C_j)$ 分别对应划分的训练子集内样本间平均距离、子集内样本间最远距离、两个子集 C_i, C_j 间最近样本的距离以及两个子集 C_i, C_j 中心点间的距离. $dist(\cdot)$ 用于计算样本间距离, $u = \frac{1}{|C|} \sum_{1 \leq i \leq |C|} x_i$ 代表子集的中心点. 相应地, CP, SP, DBI 以及 DVI 定义如下:

$$CP = \frac{1}{k} \sum_{i=1}^k \frac{1}{|C_i|} \sum_{x_i \in C_i} dist(x_i, u_i) \quad (18)$$

$$SP = \frac{2}{k^2 - k} \sum_{i=1}^k \sum_{j=i+1}^k dist(u_i, u_j) \quad (19)$$

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \frac{avg(C_i) + avg(C_j)}{d_{cen}(u_i, u_j)} \quad (20)$$

$$DVI = \min_{1 \leq i \leq k} \left\{ \min_{j \neq i} \frac{d_{\min}(c_i, c_j)}{\max_{1 \leq l \leq k} \text{diam}(C_l)} \right\} \quad (21)$$

其中, CP 越低, 反映类内距离越近, SP 越高意味着类间距离越远, DBI 小反映类内距离小同时类间距离大, DVI 大则表明类内距离小同时类间距离大。

针对 DIV2K 数据集, 在基于灰度共生矩阵提取纹理特征的基础上, 采用不同的聚类方法, 包括 FCM^[21]、BIRCH^[22]、MCLUST^[23]、STING^[24] 以及本文采用的基于密度峰值的聚类方法 (DP)。四种指标如表 1 所示。DP 算法在四种指标中都优于其他的聚类算法, 聚类效果是类内距离小同时类间距离大。

表 1 不同聚类算法在 CP、SP、DBI、DVI 上的性能指标

Table 1 The performance of different clustering algorithms in CP, SP, DBI and DVI

	FCM	BIRCH	MCLUST	STING	DP (ours)
CP	3.04	2.19	2.86	2.32	1.78
SP	2.67	2.96	3.08	3.12	3.89
DBI	7.23	6.91	8.23	6.58	6.01
DVI	0.52	0.57	0.49	0.55	0.63

采用本文提出的递进学习方法训练 VDSR 网络, 在课程设置方面, 在基于灰度共生矩阵的基础上, 采用包括 FCM、BIRCH、MCLUST、STING 以及 DP 在内的聚类算法对训练集进行分类, 其他参数保持一致。在测试集 Set5、Set14、BSD 100 和 Urban100 上的表现如表 2 所示。其中评价指标是 PSNR 和 SSIM, 基于 5 种聚类方法在测试集上的表现分别标记为 VDSR + FCM、VDSR +

BI、VDSR + MC、VDSR + ST、VDSR + DP。其中, PSNR、SSIM 较高的用黑体表示。相较于其他聚类算法下 VDSR 的性能指标, 基于 DP 聚类方法的 VDSR 在 PSNR、SSIM 上基本都获得了最高值。

2.5 定量分析

本文采用 PSNR 和 SSIM 两种评价指标来定量分析递进学习对 SISR 网络训练效果的影响。分别比较了 SRCNN、VDSR、DRCN 三种网络采用传统方法和本文提出的递进学习两种训练方法下的性能指标, 测试集是 Set5、Set14、BSD 100 和 Urban100。采用递进学习训练的网络分别标记为 SRCNN + CL、VDSR + CL 和 DRCN + CL。具体数值见表 3, 其中 PSNR、SSIM 较高的在表 3 中用粗体表示。相较于传统训练下的网络模型, 采用本文训练方式的 SRCNN、VDSR、DRCN 在 PSNR 上分别获得了平均 0.158 dB、0.18 dB、0.092 dB 的提升, 在 SSIM 上则分别获得了平均 0.0056、0.0061、0.0053 的提升。实验表明采用改进的递进学习算法训练 SISR 网络模型可以获得更好的超分辨率效果。

2.6 定性分析

DRCN 采用深度递归神经网络实现超分辨率, 可以较好地恢复图像的纹理细节。本节采用递进学习训练 DRCN 网络, 测试结果与包括双三次差值 (Bicubic interpolation)、A + ^[25]、VDSR 以及传统方法训练的 DRCN 对比。具体的视觉效果对比情况如图 4 所示。图 4 展示了 3 幅图像的 3 倍超分辨率放大效果图, 测试图像来自 Set14、BSD100 数据集。可以看出, 相比 A + 算法, VDSR、DRCN 较好地恢复出图像细节, 采用递进学习训练的 DRCN

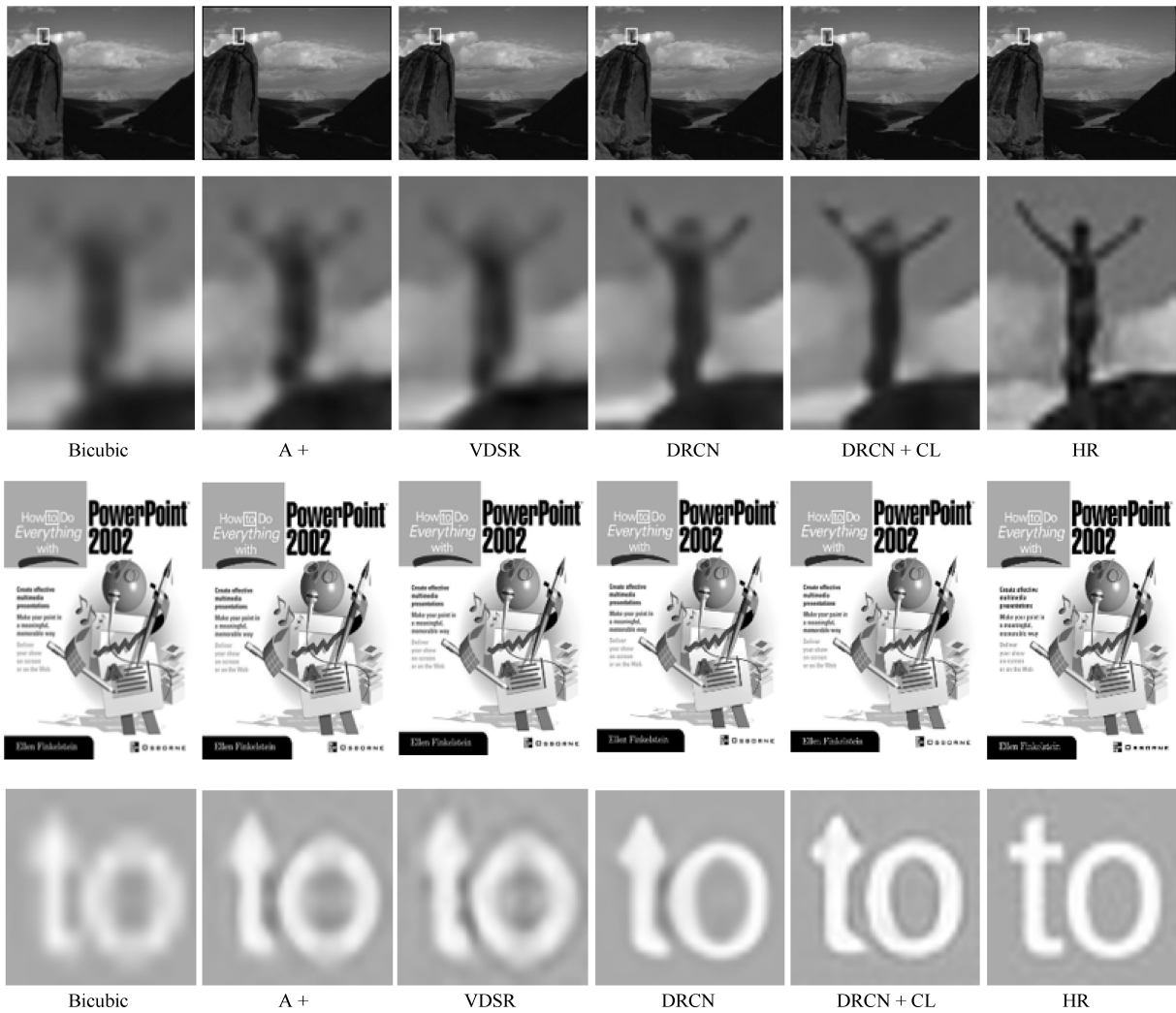
表 2 基于不同聚类算法的 VDSR 在数据集 Set5、Set14、BSD 100、Urban100 上的性能指标

Table 2 The performance of VDSR based on different clustering algorithms in Set5, Set14, BSD 100, and Urban100

数据集	放大比例	VDSR (PSNR/SSIM)	VDSR + FCM (PSNR/SSIM)	VDSR + BI (PSNR/SSIM)	VDSR + MC (PSNR/SSIM)	VDSR + ST (PSNR/SSIM)	VDSR + DP (PSNR/SSIM)
Set5	× 2	37.53/0.9587	37.56/0.9462	37.68/0.9538	37.56/0.9589	37.65/0.9581	37.74/0.9592
	× 3	33.66/0.9213	33.67/0.9241	33.70/0.9258	33.71/0.9232	33.63/0.9222	33.79/0.9264
	× 4	31.35/0.8838	31.38/0.8799	31.53/0.8812	31.41/0.8861	31.47/0.8846	31.49/0.8897
Set14	× 2	33.03/0.9124	33.06/0.9126	33.01/0.9112	33.07/0.9129	33.09/0.9125	33.11/0.9122
	× 3	29.77/0.8314	29.80/0.8352	29.86/0.8356	29.79/0.8329	29.81/0.8329	29.91/0.8402
	× 4	28.01/0.7674	28.12/0.7650	28.29/0.7710	28.13/0.7703	28.26/0.7717	28.32/0.7738
BSD 100	× 2	31.90/0.8960	31.99/0.8978	31.93/0.9010	32.09/0.8992	32.05/0.8993	32.13/0.9071
	× 3	28.82/0.7976	28.84/0.7977	28.92/0.7954	28.89/0.7988	29.04/0.8004	29.11/0.8011
	× 4	27.29/0.7251	27.41/0.7196	27.32/0.7260	27.35/0.7273	27.32/0.7278	27.28/0.7310
Urban100	× 2	30.76/0.9140	30.77/0.9139	30.74/0.9123	30.91/0.9169	30.84/0.9156	30.81/0.9193
	× 3	27.14/0.8279	27.22/0.8264	27.22/0.8282	27.29/0.8277	27.16/0.8288	27.35/0.8291
	× 4	25.18/0.7524	25.33/0.7569	25.21/0.7554	25.29/0.7551	25.32/0.7542	25.41/0.7567

表 3 不同算法在数据集 Set5、Set14、BSD 100、Urban100 上的性能指标
Table 3 The performance of different algorithms in Set5, Set14, BSD 100, and Urban100

数据集	放大比例	SRCNN (PSNR/SSIM)	SRCNN + CL (PSNR/SSIM)	VDSR (PSNR/SSIM)	VDSR + CL (PSNR/SSIM)	DRCN (PSNR/SSIM)	DRCN + CL (PSNR/SSIM)
Set5	× 2	36.66/0.9542	36.92/0.9623	37.53/0.9587	37.74/0.9592	37.63/0.9588	37.71/0.9591
	× 3	32.75/0.9090	32.81/0.9136	33.66/0.9213	33.79/0.9264	33.82/0.9226	33.91/0.9239
	× 4	30.48/ 0.8628	30.56/0.8623	31.35/0.8838	31.49/0.8897	31.53/0.8854	31.61/0.8896
Set14	× 2	32.42/0.9063	32.63/0.9136	33.03/ 0.9124	33.11/0.9122	33.04/0.9118	33.11/0.9145
	× 3	29.28/0.8209	29.41/0.8261	29.77/0.8314	29.91/0.8402	29.76/0.8311	29.81/0.8423
	× 4	27.49/ 0.7503	27.62/0.7501	28.01/0.7674	28.32/0.7738	28.02/0.7670	28.13/0.7722
BSD 100	× 2	31.36/0.8879	31.52/0.8935	31.90/0.8960	32.13/0.9071	31.85/0.8942	31.91/0.9062
	× 3	28.41/0.7863	28.63/0.7912	28.82/0.7976	29.11/0.8011	28.80/0.7963	28.92/0.8037
	× 4	26.90/0.7101	26.99/0.7234	27.29/0.7251	27.28/0.7310	27.23/0.7233	27.35/0.7274
Urban100	× 2	29.50/0.8946	29.72/0.9064	30.76/0.9140	30.81/0.9193	30.75/0.9133	30.86/0.9201
	× 3	26.24/0.7989	26.41/0.8035	27.14/0.8279	27.35/0.8291	27.15/0.8276	27.23/0.8294
	× 4	24.52/0.7221	24.69/0.7316	25.18/0.7524	25.41/0.7567	25.14/0.7510	25.21/0.7572



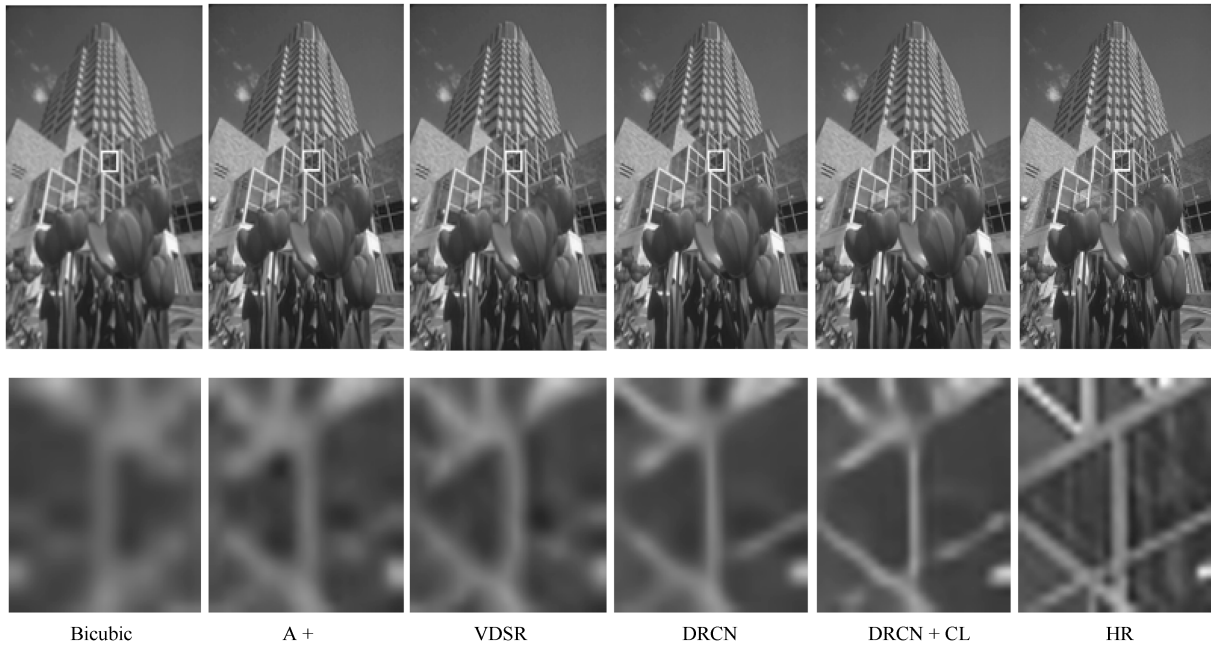


图 4 不同超分辨率算法视觉效果对比图 (放大倍数为 3 倍)

Fig. 4 Comparison of different algorithms in visual effects with upscaling factor 3

(DRCN + CL) 则具有更加丰富的纹理细节, 边缘也更为锐化, 视觉效果也更接近原始的高分辨率图像 (HR). 实验表明采用改进的递进学习算法有利于恢复纹理细节等高频信息, 从而获得更好的视觉效果.

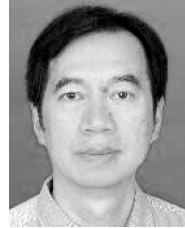
3 结论

本文提出用于超分辨率重建的深度网络递进学习方法. 递进学习有利于网络模型在“由易到难”的拟合过程中, 较快地收敛到接近全局最优解的局部极小值. 课程设置方面, 本文采用灰度共生矩阵提取图像纹理特征, 通过基于密度峰值的聚类方法实现对图像的分类. 在递进学习方法上, 本文针对传统递进学习方法的优缺点做出改进, 依据模型在各个子集上的学习曲线斜率来实时调整训练样本在不同子集上的概率分布, 获得比其他递进学习更好的训练效果. 实验表明, 本文提出的训练方法在加快网络训练的同时, 最终的超分辨率测试效果在 PSNR、SSIM 上都有相应的提升. 但本文的递进学习方法还需要人为对训练集进行分类, 分类方法直接影响到后续的训练. 在之后的研究中, 将尝试通过强化学习的方式, 让网络“学习”出选择训练样本的能力, 从而可以进一步减少人为设置课程对训练造成的不稳定.

References

- Dong C, Chen C L, He K M, Tang X O. Image super-resolution using using deep convolutional networks. *IEEE Transactions on Pattern & Machine Intelligence*, 2016, **38**(2): 295–307
- Hu Chang-Sheng, Zhan Shu, Wu Cong-Zhong. Image super-resolution reconstruction based on deep feature learning. *Acta Automatica Sinica*, 2017, **43**(5): 814–821 (胡长胜, 詹曙, 吴从中. 基于深度特征学习的图像超分辨率重建. 自动化学报, 2017, **43**(5): 814–821)
- Sun X, Li X G, Li J F, Zhuo L. Review on deep learning based image super-resolution restoration algorithms. *Acta Automatica Sinica*, 2017, **43**(5): 697–709
- Li Tao, He Xiao-Hai, Qing Lin-Bo, Teng Qi-Zhi. Noisy image super-resolution reconstruction with adaptive patch-group-cuts prior. *Acta Automatica Sinica*, 2017, **43**(5): 765–777 (李滔, 何小海, 卿麟波, 滕奇志. 基于自适应块组割先验的噪声图像超分辨率重建. 自动化学报, 2017, **43**(5): 765–777)
- Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks. In: *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, 2016, 1646–1654
- Ledig C, Theis L, Huszár F, Caballero J, Cunn A, Acosta A, et al. Photo-realistic single image superresolution using a generative adversarial network, arXiv preprint, arXiv: 1609.04802, 2016
- Lim B, Son S, Kim H, Nah S, MuLee K. Enhanced deep residual networks for single image super-resolution. In: *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, HI, USA: IEEE, July 2017, 1132–1140
- Sajjadi M S, Scholkopf B, Hirsch M. Enhancenet: single image super-resolution through automated texture synthesis, arXiv preprint, arXiv: 1612.07919, 2016

- 9 Bengio Y, Louradour J, Collobert R, Weston J. Curriculum learning. In: Proceedings of the 26th Annual International Conference on Machine Learning. Montreal Quebec, Canada: ACM, 2009
- 10 Graves A, Bellemare G, Menick J, Munos R, Kavukcuoglu K. Automated curriculum learning for neural networks. arXiv preprint arXiv: 1704.03003, 2017
- 11 Hochreiter, Sepp, Jürgen Schmidhuber. LSTM can solve hard long time lag problems. *Advances in Neural Information Processing Systems*. 1997
- 12 Haralick R M, Shanmugam K. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 1973, (6): 610–621
- 13 Rodriguez A, Laio A. Clustering by fast search and find of density peaks. *Science*, 2014, **344**(6191): 1492
- 14 Kim J, Lee J K, Lee K M. Deeply-recursive convolutional network for image super-resolution. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA: IEEE, 2016, 1637–1645
- 15 Timofte R, Agustsson E, Gool L V, Yang Ming-Hsuan, Zhang L, Lim B, et al. Ntire 2017 challenge on single image super-resolution: Methods and results. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA: IEEE, 2017
- 16 Bevilacqua M, Roumy A, Guillemot C. Low-complexity singleimage super-resolution based on nonnegative neighbor embedding. In: Proceedings of the 2012 British Machine Vision Conference, Surrey, UK: 2012, 135.1–135.10
- 17 Zeyde R, Elad M, Protter M. On single image scale-up using sparse-representations. In: Proceedings of the 2010 International conference on curves and surfaces. Avignon, France: Springer, 2010, 711–730
- 18 Martin D, Fowlkes C, Tal D. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proceedings of the 8th IEEE International Conference on Computer Vision, 2001, **2**(11): 416–423
- 19 Huang J B, Singh A, Ahuja N. Single image super-resolution from transformed self-exemplars. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, Massachusetts, USA: 2015, 5197–5206
- 20 Abadi M, Agarwal A, Barham P, Brevdo E, Chen Zhi-Feng, Citro C, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint, arXiv: 1603.04467, 2016
- 21 Bezdek J C, Ehrlich R, Full W. FCM: the fuzzy c-means clustering algorithm. *Computers & Geosciences*, 1984, **10**(2-3): 191–203
- 22 Zhang T, Raghu R, Miron L. BIRCH: an efficient data clustering method for very large databases. *ACM Sigmod Record*. **25**(2): ACM, 1996
- 23 Fraley C, Raftery A E. MCLUST: Software for model-based cluster analysis. *Journal of Classification*, 1999, **16**(2): 297–306
- 24 Wang W, Jiong Y, Richard M. STING: A statistical information grid approach to spatial data mining. *VLDB*. Athens, Greece, Vol. 97, 1997: 186–195
- 25 Timofte R, Smet V D, Gool L V. A +: Adjusted anchored neighborhood regression for fast super-resolution. In: Proceedings of the 2014 Asian Conference on Computer Vision (ACCV). Singapore: Springer, 2014, 111–126



张毅锋 博士, 东南大学信息科学与工程学院副教授, IEEE 高级会员. 主要研究方向为计算机视觉, 机器学习, 数字水印与信息隐藏, 混沌神经信息处理和无线通信. E-mail: yfz@seu.edu.cn
(**ZHANG Yi-Feng** Ph.D., associate professor at the School of Information Science and Engineering, Southeast University. Senior member of the IEEE. His research interest covers computer vision, machine learning, digital watermarking and information hiding, chaotic neural information processing, and wireless communication.)



刘袁 东南大学信息科学与工程学院硕士研究生. 主要研究方向为超分辨率, 视频理解, 语义分割. 本文通信作者. E-mail: liuyuan@seu.edu.cn
(**LIU Yuan** Master student at the School of Information Science and Engineering, Southeast University. His research interest covers super resolution, video understanding and semantic segmentation. Corresponding author of this paper.)



蒋程 东南大学信息科学与工程学院硕士研究生. 主要研究方向为机器学习, 人脸检测. E-mail: 220150747@seu.edu.cn
(**JIANG Cheng** Master student at the School of Information Science and Engineering, Southeast University. His research interest covers machine learning and face detection.)



程旭 南京信息工程大学计算机与软件学院副教授, 2015年在东南大学获得博士学位. 主要研究方向为计算机视觉和模式识别. E-mail: xcheng@nuist.edu.cn
(**CHENG Xu** Associate professor at the School of Computer and Software, Nanjing University of Information Science and Technology. He received his Ph.D. degree in Information and Communication Engineering from Southeast University (Nanjing), in 2015. His research interest covers computer vision and pattern recognition.)