

类别信息生成式对抗网络的单图超分辨重建

杨云¹, 张海宇¹, 朱宇², 张艳宁²

1. 陕西科技大学电气与信息工程学院, 西安 710021; 2. 西北工业大学计算机学院, 西安 710129

摘要: **目的** 基于生成式对抗网络的超分辨模型 (SRGAN) 以感知损失函数作为优化目标, 有效解决了传统基于均方误差 (MSE) 的损失函数导致重建图像模糊的问题。但是 SRGAN 的感知损失函数中并未添加明确指示模型生成对应特征的标志性信息, 使得其无法精准地将数据的具体维度与语义特征对应起来, 受此局限性影响, 模型对于生成图像的特征信息表示不足, 导致重建结果特征不明显, 给后续识别处理过程带来困难。针对上述问题, 在 SRGAN 方法的基础上, 提出一种类别信息生成式对抗网络的超分辨模型 (Class-info SRGAN), 通过附加的信息变量限制超分辨重建的解空间范围, 协助模型更加准确地参照数据的语义特征完成重建任务。**方法** 对 SRGAN 模型增设类别分类器, 并将类别损失项添加至生成网络损失中, 再利用反向传播训练更新网络参数权重, 以达到为模型提供特征类别信息的目的, 最终生成具有可识别特征的重建图像。创新及优势在于将特征类别信息引入损失函数, 改进了超分辨模型的优化目标, 使得重建结果的特征表示更加突出。**结果** 经 CelebA 数据集测试表明: 添加性别分类器的 Class-info SRGAN 的生成图像性别特征识别率整体偏高 (范围在 58% 至 97% 之间); 添加眼镜分类器的 Class-info SRGAN 的生成图像眼镜框架更加清晰。此外, 模型在 Fashion-mnist 与 Cifar-10 数据集上的结果同样表明其相较于 SRGAN 的重建质量更佳。**结论** 实验结果验证了本方法在超分辨重建任务中的优势和有效性, 同时结果显示: 虽然 Class-info SRGAN 更适用于具有简单、具体属性特征的图像, 但总体而言仍是一种效果显著的超分辨模型。

关键词: SRGAN; 感知损失函数; MSE; 类别信息; Class-info SRGAN

Class-info generative adversarial network for single image super-resolution

Yun Yang¹, Haiyu Zhang¹, Yu Zhu², Yanning Zhang²

1. College of Electrical & Information Engineering, Shaanxi University of Science and Technology, Xi'an, 710021; 2. School of Computer Science, Northwestern Polytechnical University, Xi'an, 710129

Abstract: **Objective** Image super-resolution reconstruction technology which means that by a set of low-quality and low-resolution images (or motion sequences) to produce a corresponding high-quality and high-resolution one. It has a wide range of applications in many fields, such as military, medical, public safety, and computer vision. In computer vision area, image super-resolution reconstruction may be able to make the image achieve a transformation from the detection level to recognition level, and even further to the identification level. In other words, image super-resolution reconstruction can

基金项目: 国家自然科学基金青年科学基金项目 (批准号: 61601271) (Project supported by the National Science Foundation for Young Scientists of China); 陕西省重点研发计划项目 (2017NY-124); 陕西省科学技术研究发展计划项目 (2014K15-03-06); 陕西省社会发展科技攻关项目(2016SF444, 2015SF277); 西安市科技计划项目(NC1319(1), NC1403, NC1403(2))

收稿日期: 2018-05-18 ; **修回日期:** 2018-06-25

Supported by: National Science Foundation for Young Scientists of China (Grant No. 61601271)

enhance the image recognition capability and identification accuracy. In addition, it can also realize a dedicated analysis for the target, so that a higher spatial resolution image of the interesting region can be obtained instead of employing the configuration of a high spatial resolution image with a large amount of data and calculation directly. The conventional approaches of super-resolution reconstruction, in general, includes example-based model, bicubic interpolation model and sparse coding methods, etc. In recent years, with the advent of the time of artificial intelligence (AI), deep learning (DL) has established a close connection with many associative subjects, and there are a lot of research achievements acquired by it, so does it in super-resolution reconstruction field. Convolutional neural networks (CNNs) and generative adversarial networks (GANs) have made numerous breakthroughs and achievements of the domain of image super-resolution reconstruction, such as using a convolutional neural network (SRCNN), using a very deep convolutional networks (VDSR) and using a generative adversarial network (SRGAN). In particular, the present of SRGAN model indicates that the single image super-resolution (SISR) technology has made a remarkable progress. Especially it uses the perceptual loss function as the optimization goal instead of traditional loss function based on mean square error (MSE), which effectively solved the problem of the model using the original loss function in order to obtaining a higher peak signal-to-noise ratio (PSNR) that caused the reconstruction results become fuzzy. Albeit this method makes image super-resolution reconstruction quality ameliorate significantly, how to better highlight the feature representation of reconstructed images, thereby comprehensively improving the reconstruction quality of generated images are still difficult issue which need to be resolved. Essentially, because of super-resolution reconstruction itself belongs to an ill-posed problem: images lose a certain amount of info during the process of down-sampling, therefore the reconstruction of the corresponding high-resolution image by the low-resolution image that lost part of the characteristics will inevitably bring about a generative deviation. In addition, out of the fact that SRGAN didn't add any auxiliary trademark info, which explicitly instruct the model to generate the corresponding features, into its loss function, made the model fail to accurately match specific dimensions and semantic features of data. The influence on controllability has confined its ability to sufficiently represent the features info of generated images, which caused a limitation to better improve the quality of reconstructed images and posed a difficulty to subsequent identification and process of the image. Aiming at above problems, based on SRGAN method, a super-resolution model based on class-info generative adversarial network (Class-info SRGAN) now is proposed, which be designed to utilize the additional info variables to restrict the solution space scope of super-resolution reconstruction and assist the model to fulfil the reconstruction task referring to the data semantic features in a more accurate way. **Method** For original SRGAN model, adding a class classifier and putting the class-loss item into the generative network loss, then making use of back-propagation during the training process to update parameters weights of the network to achieve the purpose of providing feature class-info for the model, and finally producing the reconstructed images which are possessed of the corresponding features. The innovation and advantage of proposed model lies in, to the original objective function, introducing feature class-info and improving the optimization objective of super-resolution model. Sequentially, it optimizes the network training process, meanwhile, makes the feature representation of reconstruction results become more prominent. **Result** The CelebA experiments indicated that the class-loss item helps the SRGAN model make minor changes towards better output. To be specific, compared with SRGAN model, respectively, with gender-class information, the differences were blurry, so it was harder to conclude that if the model does have a significant effect, although to some extent there were some slight improvements. The overall gender recognition rate of generated images by Class-info SRGAN model is, ranging from 58% to 97%, higher than the rate of SRGAN (from 8% to 98%). But with glasses-class information, the model more obviously learned how to form better shaped glasses. Besides, the results on the Fashion-mnist dataset and Cifar-10 dataset also indicated that the model is of significant effect, although the final results with Cifar-10 dataset were not very prominent as before experiments. In a word, the outcomes showed that the reconstruction quality of generated images from Class-info SRGAN model are better than original SRGAN. **Conclusion** Class-info does work in cases where there are clear-cut attributes that the model has learned as much as possible. The experimental results verify the superiority and effectiveness of the proposed model in the super-resolution reconstruction task. To sum up, based on some concrete and simple feature attributes, Class-info SRGAN seems to be a

promising super-resolution model. But when it comes to the advancement, the answer must be definite. For example, how to develop a general Class-info SRGAN that can be used for a variety of super-resolution reconstruction tasks, how to successfully carry out a Class-info SRGAN with multiple attributes simultaneously, and how to insert the auxiliary class-info into Class-info SRGANs' architectures more efficiently and conveniently, so on and so forth. These assumptions provide references and conditions for acquiring a better performance of super-resolution reconstruction in the future.

Key words: super-resolution based on generative adversarial network (SRGAN); perceptual loss function; Mean Square Error (MSE); class-info; Class-info SRGAN

0 引言

单图超分辨 (Single Image Super-resolution, SISR) 的主要任务是通过单幅低分辨率 (Low Resolution, LR) 图像重建出高分辨率 (High Resolution, HR) 图像。由于单图超分辨的输入相比序列超分辨更易获得, 因此其应用前景十分广阔^[1-2]。

近年来, 将深度学习应用于超分辨重建任务成为了发展的趋势, 随之衍生出了许多著名的超分辨模型。其中, SRCNN^[3] (Super-resolution based on Convolutional Neural Network) 是首个基于深度学习的超分辨模型。尽管 SRCNN 效果明显优于传统方法, 但仍存在像是对超参数变化敏感, 以及训练期间网络不稳定导致训练困难等问题有待改善。

随着 Goodfellow^[4]等人于 2014 年 10 月提出生成式对抗网络^[4] (Generative Adversarial Networks, GANs) 之后, Ledig^[5]等人将其应用于超分辨重建领域并提出了 SRGAN^[5] (Super-resolution based on Generative Adversarial Network) 模型。该方法效果明显, 特别是在单图超分辨的 2 至 4 倍放大倍数方面作用显著, 超越了之前 SRCNN^[3]模型的重建结果。

其中, 值得注意的是 SRGAN^[5]模型改进了原始损失函数, 将感知损失函数 (Perceptual Loss Function) 作为其优化目标, 从而有效地解决了传统的基于均方误差 (Mean Square Error, MSE) 损失函数的模型为了获得较高的峰值信噪比 (Peak Signal to Noise Ratio, PSNR) 而导致重建图像模糊的问题。

但是, 超分辨重建本身属于一种病态问题 (Ill-posed Problem), 该问题成立的基础是构建以先验知识将高分辨率图像映射成为低分辨率图像的观测模型 (Observation Model)。由于图像降采样过程中一定量信息的损失, 因而再由采样获得的已丢失部分特征的低分辨率图像重建出对应的高分辨率图像必然会存在一定的生成偏差, 这种偏差随着图像

内容、类别、纹理等复杂程度的提高而提高。

另一方面, SRGAN^[5]模型的感知损失函数中并未添加明确指示模型生成对应特征的标志性信息, 造成超分辨重建的解不具有确定性, 从而使得模型无法精准地将数据的具体维度与语义特征对应起来。这种在可控性上受到的限制造成模型对于生成图像的特征信息无法进行充分的表示, 导致重建结果的局部特征不明显, 进而无法全面提升重建质量。

为了解决以上问题, 本文提出一种类别信息生成式对抗网络的超分辨模型 (Class-info SRGAN), 核心是改进了以 SRGAN^[5]为基础的网络结构, 引入特征类别信息以突出重建结果的具体属性特征, 使其在目前较为成功的超分辨放大倍数上实现更为理想的重建效果。方法的理论基础是通过 SRGAN 模型设法增添类别分类器, 引入类别损失项, 利用反向传播 (Back Propagation) 训练的方式更新网络参数权重以提供图像特征类别信息, 协助模型训练并生成具有可识别特征的重建图像。

Class-info SRGAN 模型的创新及优势之处在于, 将特征类别信息引入损失函数, 改进了原始 SRGAN^[5]模型的目标函数, 优化了网络训练过程, 使得重建结果的特征表示更加突出, 并为之后针对超分辨重建图像做进一步的检测、识别和分析等工作提供了重要的前提与保障。

1 相关工作

多年来, 有关超分辨重建的研究一直吸引着人们的目光, 其方法种类较多, 现大致将超分辨模型分为两类: 非深度学习方法 (Non-deep Learning Approaches) 和深度学习方法 (Deep Learning Approaches)。

1.1 非深度学习方法

传统的超分辨模型之一是基于实例的超分辨^[6] (Example-Based Super-resolution), 其通过最近邻搜索^[7]在三次样条插值^[8]中锐化边缘和图像细节, 该方法可以获得丰富的高频信息, 但缺点也较为明

显,比如对于训练样本的要求较高,以及对噪声极为敏感。

另一种模型是双立方插值超分辨^[9](Bicubic Interpolation Super-resolution),其突出特点是在 2 倍放大倍数上效果理想,并且成功改进了之前先进的双线性插值模型^[10](Bilinear Interpolation Model)。基于插值的方法通常算法简单、快速,但却容易产生模糊或者锯齿边缘。

此外,还有利用稀疏编码方法^[11](Sparse Coding Methods)学习与优化低分辨率图像块映射函数,并聚合相应的高分辨率图像块进行超分辨重建。但该方法的缺陷在于过完备词典^[12](Over-complete Dictionary)的选择,随机选取只能实现特定领域图像的超分辨,而对于通用图像的超分辨重建效果较差。

一般来说,非深度学习的超分辨重建模型由于受到抗噪能力、重建质量,以及不同类型图像之间泛化程度的限制,因此效果并不理想。

1.2 深度学习方法

SRCNN^[3]是最早将深度学习技术应用于超分辨重建的神经网络模型,其主要可分为三个步骤:图像块提取与特征表示,特征非线性映射以及最终的重建。SRCNN 在 2 倍放大倍数方面效果显著,并启发了其他一些基于卷积神经网络(Convolutional Neural Network, CNN)的方法,主要包括像素递归超分辨^[13](Pixel Recursive Super-resolution, PRSR)和感知损失^[14](Perceptual Loss)。

另一种基于卷积神经网络的方法是使用更深层的 CNN 模型,即 VDSR^[15](Super-resolution Using Very Deep Convolutional Networks)。VDSR 重复地级联一对卷积和非线性层,深度范围从 5 至 20,并且随着网络深度的增加,模型性能也有所提高。VDSR 与 SRCNN 相比有了明显的改进,特别是在更高放大倍数的情况下。

另外,香港中文大学的 Dong 和 Tang^[16]等人在 SRCNN^[3]之后又提出了 FSRCNN^[16](Fast Super-resolution Convolutional Neural Networks)。其主要从以下三个方面对 SRCNN 进行改进:首先,FSRCNN 不再将双立方插值作为输入,而是采用原始低分辨率图像,并在网络末端引入去卷积层以执行上采样。其次, SRCNN 中的非线性映射被 FSRCNN 中的收缩、映射和扩展所替代。最后, FSRCNN 采用更小的滤波器尺寸和更深的网络结构。这些改进在为

FSRCNN 提供了更好性能的同时也降低了计算成本。

上述基于 CNN 的超分辨重建模型虽然效果与传统方法相比有了大幅提升,但是由于模型对超参数的变化较为敏感,以及在训练期间网络不稳定导致训练困难等问题的存在,给重建过程的顺利完成造成了阻碍。

在过去几年中,使用 GANs^[4]实现超分辨重建已经取得了显著的进展。Radford 和 Metz^[17]等人于 2015 年提出深度卷积生成式对抗网络^[17](Deep Convolutional Generative Adversarial Networks, DCGAN)以提升训练 GANs 执行超分辨重建任务时的稳定性。还有一种重要的基于深度学习的超分辨模型即 SRGAN^[5]。Ledig^[5]提出使用一种带有残差块(Residual Blocks)的前馈 CNN 网络作为其生成函数,该函数使用感知损失(Perceptual Loss)作为优化目标。从本质上来说,SRGAN 的主要特点是利用深度残差网络^[18](Deep Residual Network, DRN)使该模型能够实现 4 倍的放大倍数,而且改进了基于原始目标函数的重建模型为获得较高的 PSNR 造成图像模糊的缺陷。目前来说,SRGAN 模型被广泛认为能够产生出当前最先进的重建效果。

但是,SRGAN^[5]在超分辨重建过程中缺少显式添加的指示模型生成对应特征类别信息,这会导致重建结果的特征类别表征不明,从而造成重建图像与原始图像存在一定的语义特征偏差,甚至可能会重建出具有错误特征信息的图像。这对超分辨重建问题本身来说是至关重要的,因为即使成功恢复出了具有高质量清晰度的图像,但是由于图像内容的错误会使得整个重建过程及结果的价值被大幅削减。本文提出的 Class-info SRGAN 模型,针对这一问题,通过将特征类别信息引入损失函数,从而改进了超分辨模型的优化目标,使得最终重建结果的特征表示更加突出,为后续的分类处理等相关工作奠定了良好的基础。

2 方法

2.1 GANs

生成式对抗网络^[4](GANs)是一种通过对抗过程估计生成模型的网络框架,其对生成模型(Generative Model)的发展产生了深远的影响。

GANs 由生成网络 G 和判别网络 D 构成, D 可以判断出样本是来自于原始数据还是由生成网络产生,而 G 试图伪造原始数据来使判别模型做出错误判断,

直至生成真假难辨的样本。整个过程中 G 与 D 同时运行，可近似将其看作是一种二元博弈游戏 (Two-player Minmax Game)。GANs 的原理示意图如图 1 所示，目标函数表示为公式 (1)，其中 z 是潜在表征，初始化为随机噪声， x 是真实样本， p 代表概率分布， \mathbb{E} 代表期望。

$$\min_G \max_D (D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (1)$$

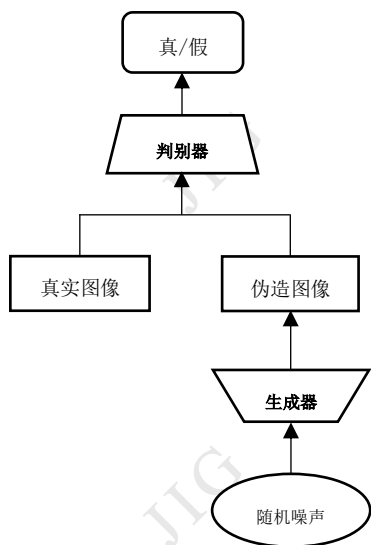


图 1 生成式对抗网络原理示意图
Figure 1. GANs schematic diagram

2.2 SRGAN

随机噪声的输入使得 GANs^[4] 在训练过程中显得过于自由，这种在可控性方面的缺陷最终导致生成模型效果不佳。对此，SRGAN^[5] 不再使用随机噪声作为初始化变量，而是将来自于原始数据集的高分辨率图像 I^{HR} 经高斯滤波降采样得到的低分辨率图像 I^{LR} 作为训练样本。SRGAN 的原理示意图如图 2 所示，其目标函数表示为公式 (2)。

$$\min_{\theta_G} \max_{\theta_D} \mathbb{E}_{I^{HR} \sim p_{train}(I^{HR})} [\log D_{\theta_D}(I^{HR})] + \mathbb{E}_{I^{LR} \sim p_G(I^{LR})} [\log(1 - D_{\theta_D}(G_{\theta_G}(I^{LR})))] \quad (2)$$

其中，由 θ_G 参数化的前馈 CNN 生成网络记作 G_{θ_G} ，这里， $\theta_G = \{W_{1:L}; b_{1:L}\}$ ，表示 L 层深度网络的权重和偏差，并通过优化超分辨重建的特定损失函数 l^{SR} 获得，同理有参数化判别网络 D_{θ_D} 。

SRGAN 以感知损失作为优化目标，该损失由内容损失 (Content Loss) 和对抗损失 (Adversarial Loss) 两部分加权组合构成，如公式 (3) 所示。

$$l^{SR} = l_x^{SR} + 10^{-3} l_{Gen}^{SR} \quad (3)$$

其中， l_x^{SR} 是内容损失函数，使用 VGG^[19] (Very Deep Convolutional Networks) 网络特征图谱损失替代了之前基于 MSE 的内容损失， l_{Gen}^{SR} 是原始对抗损失函数。

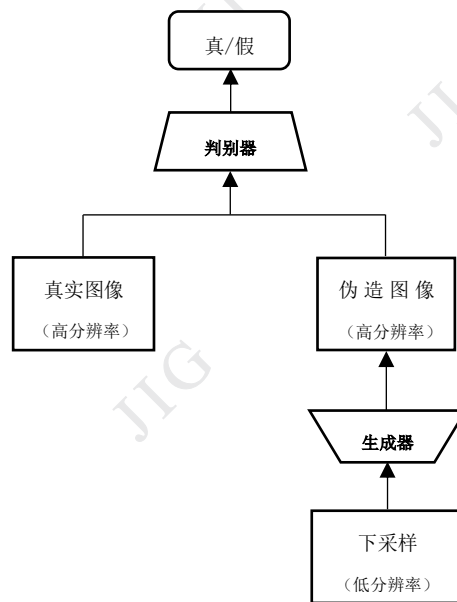


图 2 生成式对抗网络的超分辨原理示意图
Figure 2. SRGAN schematic diagram

2.3 Class-info SRGAN

如前所述，SRGAN^[5] 的感知损失函数中缺少特征类别信息，使得模型对于生成图像的特征信息表示不足，导致重建结果的特征表示不明显甚至出现错误，给后续相关工作造成障碍。

本文提出一种用于单图超分辨的 Class-info SRGAN 模型。在该模型下，设 $C(*)$ 为计算类别损失的附加项，通过对目标函数显式地添加并最小化该类别损失项从而对超分辨模型引入特征类别信息，这样一方面改进了超分辨模型的目标函数，优化了训练过程，另一方面更突出了重建结果的特征表示。Class-info SRGAN 的原理示意图如图 3 所示，其目标函数表示为公式 (4)。

$$\min_G \max_D \mathbb{E}_{I^{HR} \sim p_{train}(I^{HR})} [\log D_{\theta_D}(I^{HR})] + \mathbb{E}_{I^{LR} \sim p_G(I^{LR})} [\log(1 - D_{\theta_D}(G_{\theta_G}(I^{LR}) + C(*)))] \quad (4)$$

为了实现该结构，首先根据一组预先确定类别的属性 (比如性别或眼镜等) 在原始数据集上训练分类器，具体参照 3.2.1 节中所述。然后固定分类器权重并将网络结构导入 SRGAN 模型，因此在模型训练期间不会再对分类器参数的权重进行更新。最后替换分类器中的节点，将现有的 SRGAN 结构图与分类器结构图连接起来，以便在 SRGAN 模型中使用

附加的类别信息,从而实现Class-info SRGAN 模型。

整个过程概括来说,就是使用与 SRGAN 连接后的分类器模型首先得到原始高分辨率图像(64×64)的分数,同理可得到生成器生成图像(64×64)的分数,再将原始高分辨率图像分数视为标签,生成器分数作为 Softmax 交叉熵损失的对数值,产生的类别损失项然后被添加到生成器损失中并且使用 Adam 函数优化,该方法类似于原始 GANs 模型中的损失函数最小化。

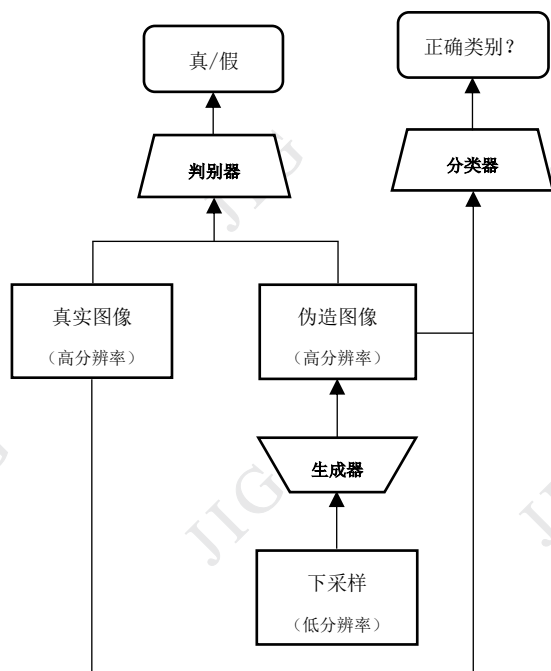


图 3 类别信息生成式对抗网络的超分辨原理示意图

Figure 3. Class-info SRGAN schematic diagram

通过向生成器损失中添加分类器类别损失项,使生成器针对分类器上的正确预测进行了优化,从而显式地表明结果接近于原始图像,而类别损失反推至生成器的参数权重,使得分类器呈现生成器学到的附加信息,例如一些性别特征或眼镜的样子等。因此,类别损失项提供的信息对超分辨模型的训练过程和结果而言都是十分重要的。

3 实验

一般来说,特别是当针对一些背景复杂,内容丰富的数据进行重建时,类别信息可能不会对超分辨模型产生较为显著的影响。因为这些图像内包含的特征数量庞大,细节信息繁多,导致类别信息的添加与使用会变得复杂而困难。相比之下,图像像素和边缘信息通常被认为对重建结果更为重要。

本文所述方法旨在从具体特征层面来增强单图

超分辨的重建效果,因此,此处首先选择 CelebA^[20] 作为训练与测试数据集,理由是因为相较于其他数据集,CelebA 的内容和背景较为单一,故推测类别信息对其超分辨重建结果产生影响的可能性较大,而且实验效果也会更加显著。

3.1 CelebA 数据集

名人面孔特征数据集 (Large-scale CelebFaces Attributes, CelebA) 是一个大规模人脸特征数据集,每个图像具有 40 个属性标签(例如“男”,“眼镜”,“刘海”,“胡子”等)。CelebA 内容丰富,总共包括 10,177 种身份,202,599 位名人人脸图像和 5 个地标位置。

CelebA 提供数据集的预处理对齐和剪切版本,这些对上半身进行了对齐和裁剪后的图像大小为 178×218 (例如图 4(a))。此处,在原始图像脸部周围随机裁剪生成 128×128 大小的图像并缩小至 64×64 (例如图 4(b)),最后使用经过 4 倍降采样的 16×16 的图像作为模型输入。



(a) 对齐和裁剪图像(178×218) (b) 预处理和裁剪图像(64×64)

图 4 CelebA 数据集样本图像

Figure 4. CelebA dataset sample image ((a) aligned and

cropped image (178×218); (b) preprocessed and cropped image (64×64))

3.2 类别信息

此处旨在以原始 SRGAN 模型重建质量为基础的前提下提高人眼对特定属性的辨别度。比如下文中通过添加性别信息,希望在生成图像中使该特征更加突出,从而提高整体重建质量。

3.2.1 性别分类器

首先,在 CelebA 上训练一个 CNN 性别分类器,分类器结构与训练方法的详细信息如附件 new_srgan 下 classifier 中的 cnn.ipynb 文件所示。分类器组成如下:

Conv Layer 1: 包含 32 个滤波器,每个滤波器尺寸为 3×3,最后通过 2×2 大小的最大池化对图像进行下采样处理使尺寸缩减为原来的一半(64×64)。

Conv Layer 2 和 3: 第 2 卷积层(3×3×32)和第 3 卷积层(3×3×64)分别以第 1 和第 2 卷积层的输出作为输入,输入通道的数量对应于先前卷积层中的滤波器个数。

Flatten Layer: 由于卷积层输出的是 4 维张

量, 现欲将其用作全连接网络的输入, 因此需要将输出重新塑造为 2 维张量。

Fully-connected Layer 1: 输入来自之前的 Flatten 层, 全连接层中的神经元数量为 128, 使用修正线性单元 (Rectified Linear Unit, ReLU) 作为激活函数。

Fully-connected Layer 2: 输入向量长度为 128, 输出向量长度为 2, 以确定输入图像属于哪种类别 (“男” 或 “女”), 这里不使用 ReLU 函数, 而是通过 Softmax 函数对网络输出进行归一化处理。

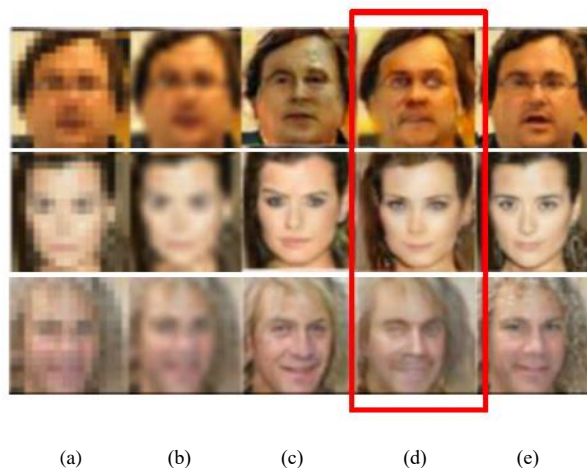
TensorFlow 中内置有可用于计算交叉熵 (Cross-entropy) 的函数, 其主要用于对分类效果的性能进行度量评价, 然后通过对神经网络层中的参数变量进行适当调整, 使模型更好地对输入图像进行分类。交叉熵是一个连续的正函数, 若模型预测输出与期望输出完全一致则交叉熵为零, 因此网络的优化目标是最小化交叉熵。在计算出每个图像的分类交叉熵之后, 为了使用交叉熵来指导模型参数变量的优化, 因而需要一个标量值, 此处选择所有图像分类交叉熵的平均值。

另外, 选取 Adam 优化器对模型训练过程进行优化。Adam 是一种可以替代传统随机梯度下降 (Stochastic Gradient Descent, SGD) 过程的一阶优化算法, 它能基于训练数据迭代地更新神经网络权重。Adam 算法与传统 SGD 算法的不同之处在于, SGD 保持单一的学习率更新所有权重, 学习率在整个训练过程中不会发生改变。而 Adam 通过计算梯度的一阶和二阶矩估计来为不同的参数设计独立的自适应学习率, 使得整体训练过程尽可能接近最优。

需要注意的是, 最初性别分类器在训练集上的准确率为 0.938, 经过 10 个 epoch 后分类器在测试集上的准确率为 0.856。如 3.1 节中所述, 由于该实验对样本图像 (178×218) 做了大幅裁减 (64×64), 几乎只保留了原图中的人脸面部, 因而分类器基本不能依据颈部或头发等特征来进行区分。鉴于上述情况, 分类器仅通过面部形状及特征完成分类, 故取得当前准确率已属较为理想。

之后, 如前文 2.3 中节所述, 固定性别分类器的权重并将网络导入 SRGAN 模型, 使用分类器来计算类别损失, 然后将类别损失项添加至生成器损失中。在尝试过不同大小的类别损失权重后最终确定数值为 0.5, 因为从结果上来看此时的重建质量最佳。细节过程可参照附件代码 new_srgan 中的 model.py 文件。

如图 5 所示为随机选取的生成结果, 观察可知, SRGAN 和 Class-info SRGAN 输出结果几乎相同, 无法显式地表现出两者之间的真实性别差异。图 6 展示了 SRGAN 产生错误输出的女性图像, 值得注意的是, Class-info SRGAN 模型会产生更加女性化的面孔 (图 6 中 (d))。



(a) 16×16 降采样 (b) 双立方插值 (c) SRGAN (d) 本文提出 (e) 64×64 原图

图 5 添加性别分类器的类别信息生成式对抗网络的超分辨率输出结果

Figure 5. Output results of Class-info SRGAN by adding gender classifier ((a) 16×16 down-sampling; (b) bicubic interpolation; (c) SRGAN; (d) proposed; (e) 64×64 original)



(a) 16×16 降采样 (b) 双立方插值 (c) SRGAN (d) 本文提出 (e) 64×64 原图

图 6 包含错误信息的输出结果

Figure 6. Output results including error message

((a) 16×16 down-sampling; (b) bicubic interpolation; (c) SRGAN; (d) proposed; (e) 64×64 original)

表 1 展示了对应于图 5 和图 6 中两种重建模型示例样本的分类准确率。由表 1 中数据可知, 第 1、3 和 4 栏中 Class-info SRGAN 的生成结果分类准确率均高于 SRGAN, 第 2 栏中两者结果也十分接近, 因而从整体上来看 Class-info SRGAN 模型重建结果的性别特征识别率更高, 说明其重建质量更佳。

表 1 SRGAN 模型与 Class-info SRGAN 模型示例样本在性别分类器上的分类准确率比较

Table 1 Comparison of classification accuracy between examples of SRGAN model and Class-info SRGAN model

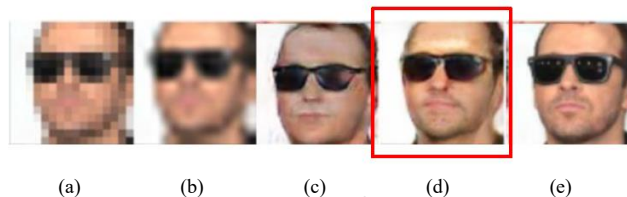
on gender classifier

序号	SRGAN	Class-info SRGAN
1	69%	97%
2	98%	93%
3	24%	59%
4	8%	58%

综合以上结果，分类器类别损失项在超分辨率模型中确实发挥了作用，但由于性别属于一种较为宽泛的特征类别范畴，包含内容众多，却未涉及具体的细节信息，以至于使模型无法全面有效地学习该类别特征(具体分析见 3.2.2 节)。为了检验该假设，需要选择一种更为具体的类别信息，比如眼镜。

3.2.2 眼镜分类器

此处设置相同结构的 CNN 分类器模型，并将输入中的性别标签更换为眼镜标签，与通过训练性别分类器向模型提供性别特征信息最终实现超分辨率重建的方法类似，添加眼镜分类器的 Class-info SRGAN 模型的重建结果如图 7 所示。由图可知，相较于 SRGAN，Class-info SRGAN 生成的眼镜看起来更加清晰并准确覆盖了眼部周围。



16×16 降采样 双立方插值 SRGAN 本文提出 64×64 原图
图 7 添加眼镜分类器的类别信息生成式对抗网络的超分辨率输出结果

Figure 7. Output results of Class-info SRGAN by adding eyeglasses classifier ((a) 16×16 down-sampling; (b) bicubic interpolation; (c) SRGAN; (d) proposed; (e) 64×64 original)

分析产生该现象的原因是因为，相比于性别特征，眼镜本身属于一种易于被察觉且十分具体的类别特征，故分类器中相应类别损失项的阈值范围会受到更加明确的限定，这样更有利于网络的训练与收敛，因而模型也就更容易学习并生成具有对应特征的重建图像。

为了表明实验中上述结果的产生并非偶然，现再次从 Class-info SRGAN 模型在 CelebA 数据集的重建结果中随机选取如图 8 所示的 8 幅图像，左右两列分别对应着添加性别分类器与眼镜分类器的 Class-info SRGAN 模型生成图像以及原始高分辨率

图像。可以看到，生成图像的内容清晰，轮廓分明，而且针对所添加的特征类别信息表征均较为明显。



(a)本文提出 (b)原始图像 (c)本文提出 (d)原始图像
图 8 类别信息生成式对抗网络的超分辨率重建结果
在 CelebA 上的重建结果

Figure 8. Reconstruction results on CelebA of Class-info SRGAN ((a) proposed; (b) original; (c) proposed; (d) original)

3.3 其他数据集

为了进一步说明文中方法的有效性，本文还在 Fashion-mnist^[21]和 Cifar-10^[22]数据集上实现了该模型。

3.3.1 Fashion-mnist 数据集

Fashion-mnist 作为 Mnist 手写数字数据集的一种替代，涵盖了来自 10 种类别的共计 7 万个不同商品的正面图像。Fashion-mnist 的大小、格式和训练/测试集划分与原始的 Mnist 数据集完全一致，而原来 Mnist 中的数字 0-9 现在依次对应着 T 恤、裤子、套衫、裙子、外套、凉鞋、汗衫、运动鞋、包和踝靴这些物品。

此处，对原始 Fashion-mnist 数据集(28×28 的灰度图片)通过 4 倍采样因子降采样处理作为超分辨率模型的输入。由于该数据集中的商品可识别度高，而且分布相当均匀，因而有助于更好地了解不同类型商品的重建情况，同时也期望能够实现比原始 SRGAN 模型更好的性能效果。

最终结果如图 9 所示，为了便于比较，设置生成图像由两部分构成，其中前 5 行表示 Class-info SRGAN 的结果，后 5 行表示 SRGAN 的结果。从图中

可明显看出,前者在清晰度、纹理和轮廓等方面均优于后者,从而证实了附加的特征类别信息对超分辨重建结果的显著作用。

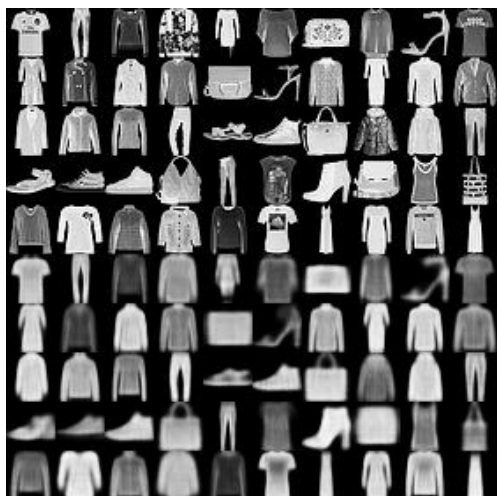


图 9 类别信息生成式对抗网络的超分辨与生成式对抗网络的超分辨生成结果比较

Figure 9. Comparison of generated results between Class-info SRGAN and SRGAN

3.3.2 Cifar-10 数据集

Cifar-10 由包含 10 个类别(飞机、汽车、鸟、猫、鹿、狗、蛙、马、船、卡车)的 60000 幅彩色图像(32×32)构成,每个类别有 6000 张,数据集总共被划分成 50000 张训练图像和 10000 张测试图像。

图 10 和 11 分别展示了 SRGAN 模型与 Class-info SRGAN 模型关于 Cifar-10 数据集各自的实验结果。从两幅图中可以看出,虽然两个模型产生的图像并不相同,但使用 Class-info SRGAN 生成的图像比使用 SRGAN 生成的图像看起来会显得更加干净清晰。

但同时也注意到,相比于 Class-info SRGAN 模型之前在 CelebA 与 Fashion-mnist 数据集上的重建结果,此时的图像清晰度并不理想,细节特征也并不突出,其中部分图像基本只能通过大致轮廓进行辨识,而 SRGAN 模型的重建效果甚至更差。

3.3.3 结果分析

如前文所述,对于背景复杂,内容丰富,细节信息数量较多的数据集,Class-info SRGAN 模型的效果可能并不会十分显著。3.3.2 节中的实验结果验证了该假设,重建图像质量相比于 SRGAN 模型并无明显提升,体现出了特征类别信息的适用范围,反映出 Class-info SRGAN 模型存在一定的局限性。现从网络结构与算法方面对此进行分析。

首先,使用本文方法在复杂数据集上实现重建

时必然需要添加多种类型的类别信息。同时添加这些类别信息的难度较高,而且如何对这些附加信息进行叠加组合从而使其能够发挥出较为理想的作用来指导模型最终生成对应的细节特征还有待研究。

其次,如前文所述,Class-info SRGAN 模型产生作用的关键在于分类器类别损失项的引入,在类别信息尚不明确或同时加入多种类别信息的情况下,会使得相应的类别损失项阈值范围得不到更加具体的限定,造成后期反向传播训练的参数更新不够准确及时,从而减弱了对超分辨模型的训练和收敛原有的指导作用。

最后,由于原始 SRGAN 模型为了获得较大的放大倍数而采用 ResNet 网络结构,因而在判别器与生成器中并未使用到全连接层,从而无法将关联特征合并至生成器与判别器当中。故与使用全连接层的网络结构相比,额外的复杂度会使得模型在整个训练过程中对于超参数的调节变得十分敏感。因而大量特征类别信息的加入对于当前的网络结构而言会影响到整体训练的稳定性,进而使类别信息的作用受到抑制,最终导致重建质量无显著提高。

4 结论

本文提出一种基于类别信息生成式对抗网络的超分辨模型(Class-info SRGAN),经标准数据集(CelebA、Fashion-mnist 和 Cifar-10)测试表明,通过对原始 SRGAN 模型的损失函数添加具体的类别损失项,从而引入特征类别信息,改进了超分辨模型的优化目标,使得重建结果的特征表示更加突出,有助于实现更好的重建结果,为后续相关工作的开展创造了条件。

但同时也发现,在模型没有尽可能多的学习到具有明确属性的特征信息,或者重建对象包含十分复杂的细节特征的情况下,特征类别信息的作用有一定局限。不过,总体来说,对于相对简单的具体属性特征,Class-info SRGAN 仍是一种较为理想的超分辨模型。

随着 GANs 的研究与发展,可实现同时添加多种类别信息的 Class-info SRGAN 模型以及引入类别信息后可应用于多种任务的泛化能力更强的 Class-info SRGAN 模型,都将可能会成为未来基于生成式对抗网络的超分辨重建研究的重点。



图 10 生成式对抗网络的超分辨率在 Cifar-10 上的重建结果

Figure 10. Reconstruction result of SRGAN on Cifar-10



图 11 类别信息生成式对抗网络的超分辨率在 Cifar-10 上的重建结果

Figure 11. Reconstruction result of Class-info SRGAN on Cifar-10

参考文献(References)

[1] Daniel G, Shai B, Michal I. Super-Resolution from a Single Image [C]// International Conference on Computer Vision. 2009:349-356. [DOI: 10.1109/ICCV.2009.5459271]
[2] Nasrollahi K, Moeslund T. Super-resolution: a comprehensive survey[J]. Machine Vision and Applications, 2014, 25(6): 1423-1468. [DOI: 10.1007/s00138-014-0623-4]
[3] Dong C, Loy CC, He K, et al. Learning a deep convolutional network for image super-resolution [C]//European Conference on Computer Vision. Springer, Cham, 2014:184-199. [DOI:10.1007/978-3-319-1059302_13]
[4] Ian J G, Yoshua B, et al. Generative Adversarial Nets [J].

arXiv preprint, 2014. [arXiv:1406.2661]
[5] Ledig C, Wang C, Shi W, Theis L, et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network [C]// IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Hawaii, USA, 2017:105-114. [DOI:10.1109/CVPR.2017.19]
[6] William TF, Thouis RJ, Egon CP. Example-based super-resolution [J]. IEEE Computer graphics and Applications, 2002, 22 (2): 56-65. [DOI: 10.1109/38.988747]
[7] Arya S, Mount D, Netanyahu N, Silverman R, Wu A. An Optimal Algorithm for Approximate Nearest Neighbor Searching in Fixed Dimensions [J]. Journal of the ACM, 1998, 45(6): 891-923. [DOI: 10.1145/293347.293348]
[8] Keys R. Cubic Convolution Interpolation for Digital Image Processing [J]. IEEE Trans. Acoustics, Speech, Signal Processing, 1981, 29(6):1153-1160. [DOI: 10.1109/TASSP.1981.1163711]
[9] Lukin A, Krylov A, Nasonov A. Image interpolation by super-resolution. Proceedings of GraphiCon, 2006: 239-242. [PSU:10.1.1.178.3520]
[10] Yuan S, Taguchi A, Kawamata M. High accuracy wadi image interpolation with local gradient features [C]//Intelligent Signal Processing and Communication Systems, 2005, 105: 49-54. [DOI: 10.1109/ISPACS.2005.1595352]
[11] Yang J, Wright J, Huang TS, Ma Y. Image Super-Resolution via Sparse Representation [J]. IEEE transactions on image processing, 2010, 19(11): 2861-2873. [DOI: 10.1109/TIP.2010.2050625]
[12] Zhang Z, Xu Y, Yang J, Li X, Zhang D. A Survey of Sparse Representation: Algorithms and Applications [J]. IEEE Access, 2015, 3: 490-530. [DOI: 10.1109/ACCESS.2015.2430359]
[13] Dahl R, Norouzi M, Jonathon S. Pixel recursive super resolution [C]//IEEE International Conference on Computer Vision, 2017: 5449-5458. [DOI: 10.1109/ICCV.2017.581]
[14] Johnson J, Alahi A, Li F. Perceptual losses for real-time style transfer and super-resolution [C]//European Conference on Computer Vision. Springer International Publishing, 2016. [DOI: 10.1007/978-3-319-46475-6_43]
[15] Kim J, Lee J, Lee K. Accurate image super-resolution using very deep convolutional networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016. [DOI:10.1109/cvpr.2016.182]
[16] Dong C, Tang X. Accelerating the Super-Resolution Convolutional Neural Network [C]//European Conference on Computer Vision, 2016. [DOI:10.1007/978-3-319-46475-6_25]

- [17] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. arXiv preprint, 2015. [arXiv: 1511. 06434]
- [18] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778. [DOI: 10. 1109/CVPR.2016.90]
- [19] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint, 2014. [arXiv:1409.1556, 2014]
- [20] Liu Z, Luo P, Wang X, Tang X. Deep Learning Face Attributes in the Wild [C]//Proceedings of International Conference on Computer Vision, 2015:3730-3738. [DOI: 10.1109/ICCV.2015.425]
- [21] Han X, Kashif R, Roland V. Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms[J]. arXiv preprint, 2017. [arXiv:1708.07747]
- [22] Alex K. Learning Multiple Layers of Features from Tiny Images[M]. California, 2009:54-60.

第一作者简介:



杨云 (1965 -), 女, 教授, 2009 年于陕西科技大学获计算机专业工学博士学位, 主要研究方向为数据挖掘与机器学习, E-mail: yangyun11@163.com。

通信作者:



张海宇, 男, 陕西科技大学硕士研究生在读, 主要研究方向: 计算机视觉与机器学习, E-mail: 18049276681@163.com。

其他作者简介:

朱宇, 男, 助理研究员, 主要研究方向: 图像处理、模式识别、计算机视觉与智能信息处理。

张艳宁, 女, 教授, 主要研究方向: 图像处理、模式识别、计算机视觉与智能信息处理。