高技术通讯 2018 年 第28 卷 第9 – 10 期:843 ~ 851

doi:10.3772/j.issn.1002-0470.2018.09 – 10.009

基于空间连续生成对抗网络的视频帧间图像生成①

张 涛②* 张 猛*** 蒋培培*

(*燕山大学信息科学与工程学院 秦皇岛 066004)

(**河北东软软件有限公司 秦皇岛 066004)

摘 要 针对低帧率视频播放不流畅以及使用传统方法提高视频帧率造成的边缘模糊问题,本文提出一种基于空间连续生成对抗网络(SC-GAN)的视频帧间图像生成方法。首先本文使用自编码器作为判别器,引入 Wasserstein 距离表示真实样本与生成样本损失分布的差异,替代传统生成对抗网络直接匹配数据分布的方式,其次利用生成器与判别器之间的平衡参数稳定训练过程,有效避免了模型崩溃的问题,最后利用连续视频帧图像在空间上的连续性,通过 Adam 在相邻两帧之间找到一个最优值,将其映射到图像空间,得到生成的帧间图像。为了说明生成的帧间图像的真实性,本文采用 PSNR 和 SSIM 对帧间图像进行了评估,评估结果证明生成的帧间图像具有较高的真实度,同时验证了本文提出的基于 SC-GAN 的视频帧间图像生成方法的可行性和有效性。

关键词 生成式对抗网络(GAN), 对抗式训练, 空间连续性, Adam, 帧间图像生成

0 引 言

提高视频帧率的技术一般称为插帧倍频或运动补偿算法(motion estimate/motion compensation, ME/MC),其基本思想是在视频序列中连续的 2 帧图像之间插入一个生成的中间图像帧,达到提高帧率的目的。

提高帧率的方法一般有非运动补偿类算法和运动补偿类算法[1]。非运动补偿类算法一般有帧重复和帧平均2 种形式[2,3],其优点是复杂度较低,容易集成到产品中,但是对于非静止的视频源视觉效果并不好。ME/MC 技术[4]在运动估计中采用块匹配算法获得运动矢量(motion vector, MV),经过计算得到较好的“运动补偿帧”,将其插入到 2 帧之间,以此来提高视频的帧率。优点是消除了视频抖动和拖尾现象,清晰度得到增强,但也会造成运动中的图像边缘不清晰。

帧间图像生成方面在深度学习领域的研究已经取得很大进展。虢齐[5]提出了图像生成模型可根据视频中过去帧和当前帧预测未来帧。Park 等人[6]提出了一种基于神经网络生成最优低动态范围图像。侯敬轩等人[7]提出了一种基于卷积神经网络的自学习帧率提升方法。龙古灿等人[8]提出一种用于视频图像帧间运动补偿的深度卷积神经网络,使用卷积网络学习视频中对象的运动规律从而生成视频的中间帧。

生成式对抗网络(generative adversarial networks,GAN)作为一种新的图像生成算法,在帧间图像生成方面的研究还比较少,其中最典型的是 Mathieu 等人[9]提出的一种超越均方误差的深度多尺度视频预测的方式,可以对视频的下一帧进行预测。这种方式是对 GAN 中对抗式思想在视频研究领域的首次应用。但是利用 GAN 模型在视频帧间

① 国家自然科学基金(61603327)和河北省自然科学基金(F2015203013)资助项目。

② 男,1979 年生,博士,副教授;研究方向:人工智能、认知计算、模式识别与数字图像处理;E-mail: zhtao@ysu.edu.cn

(收稿日期:2017-12-04)

图像生成方面的研究在文献中仍未见相关报道。

目前,GAN 已经可以生成人脸图像[10]、数字图像以及其他物体的图像,完成文字到图像的翻译、语义标注以及从低分辨率图像生成高分辨率图像等[11]。本文借鉴文献[8]中的深度卷积网络的思想,提出的基于空间连续生成对抗网络(spatial continuity generative adversarial networks, SC-GAN)的帧间图像生成方法是一种新的提高帧率的方法,其使用神经网络[12]训练的生成模型,并引入平衡参数 γ 控制生成器与判别器的训练平衡。利用连续视频帧图像的特征在空间上的连续性,通过 Adam 在相邻 2 帧之间找到一个最优值,将其映射到图像空间,得到生成的帧间图像,并采用 PSNR 和 SSIM 对帧间图像进行了评估。评估结果证明帧间图像具有较高的真实度,也验证了本文提出的基于 SC-GAN 的视频帧间图像生成方法的可行性和有效性。

1 GAN 的基本原理

GAN 的基本思想源自博弈论中的“零和博弈”,由一个生成器和一个判别器组成[13]。

图 1 显示了 GAN 的输入是一个随机噪声信号 z, 该噪声信号输入到生成器可以得出生成样本数据 $G(z)$。判别器以真实样本 x 与生成样本数据 $G(z)$ 作为输入,其输出表示判别器把生成样本 $G(z)$ 当作真实样本的概率,这个概率用来评判生成模型的质量。当判别器无法区分生成样本 $G(z)$ 和真实样本 x 时,则认为生成器和判别器达到最优状态。

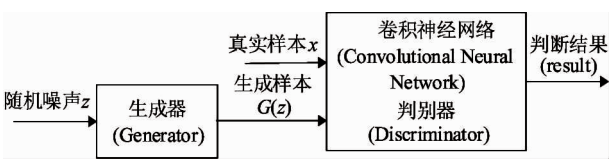

图 1 GAN 基本结构

判别器为了能够区分真实样本和生成样本,其目的是使 $D(x)$ 与 $D(G(z))$ 之间的差距最大化,即使 $D(x)$ 的值最大、$D(G(z))$ 的值最小[14]。生成器的目标是使自己产生的数据 $G(z)$ 在判别器上的表现 $D(G(z))$ 与真实样本 x 在判别器上的表现 $D(x)$ 尽可能一致,让判别器不能区分生成样本和真实样本。因此,这 2 个模块之间是一个相互竞争和对抗的过程,生成器与判别器的性能也在训练迭代过程中不断地提高。

$$\min_G \max_D V(D,G) = E_{x\sim P_{\text{data}}(x)}[\log D(x)] + E_{z\sim p_z(z)}[\log(1 - D(G(z)))] \tag{1}$$

GAN 模型的优化是一个“极小极大博弈”的过程,式(1)表示其目标函数[15]。在训练过程中,判别器尽可能地辨别出样本的真伪,即最大化 $\log D(x)$, 而生成器尽可能地骗过判别器,即最小化 $\log(1 - D(G(z)))$, 也就是最大化判别器的损失。生成器中隐式地定义了一个数据分布 P_g, 真实的数据分布定义为 P_{rd}, 经过判别器和生成器的对抗训练后,最优的结果是判别器无法辨别样本是真实样本还是生成样本,即 $P_g = P_{rd} = 0.5$。此时可以认为生成器已经学习到了真实样本的数据分布。

2 视频帧间生成算法

本文使用的是基于传统 GAN 模型的改进模型,其中使用自编码器作为判别器,并且使用由 Wasserstein 距离[16]衍生而来的损失函数对训练过程做出评估。引入平衡参数 γ 来控制生成器和判别器之间的平衡。本文使用的数据集是连续的视频帧图像,并且这些图像的特征在空间上具有连续性,利用训练好的生成模型并输入连续的 2 帧图像就可以生成帧间的图像。

2.1 模型框架

图 2 显示的是基于传统 GAN 模型的改进模型,使用自编码器作为判别器并且引入一个平衡参数 γ 控制生成器与判别器的平衡。自编码器是一种可以尽可能复现输入信号的神经网络,为了完成这种复现,自编码器就必须捕捉到可以代表输入信号最重要的特征。真实样本 x 输入到判别器,得到对真实样本重构后的损失分布 $\Gamma(x)$, 随机噪声 z 输入到生成器得到生成的数据 $G(z)$, 然后将 $G(z)$ 输入到判别器中会得到对生成样本重构后的损失分布 $\Gamma(G(z))$。通过 $\Gamma(x)$ 与 $\Gamma(G(z))$ 2 种分布之间的

Wasserstein 距离 $W(\Gamma(x),\ \Gamma(G(z)))$ 对模型的参数做出对应的调整使 D_w 的值最小。

$$D_w = W(\Gamma(x),\ \Gamma(G(z))) \tag{2}$$

同时,为了保证生成器与判别器训练过程的稳定性以及生成样本的多样性,在训练过程中引入平衡参数 γ 来平衡生成器和判别器,从而稳定训练过程。

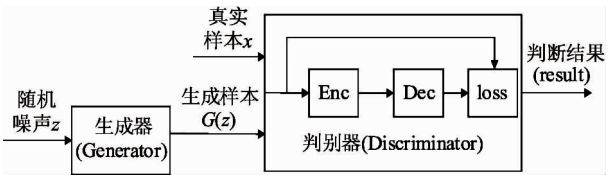

图 2　本文模型框架

2.2　Wasserstein 距离

Wassertein 距离又称作 Earth-Mover 距离(EM 距离),用于衡量 2 个分布之间的距离,定义为:

$$W(P_1,\ P_2) = \inf_{\lambda \sim \prod(P_1,\ P_2)} E_{(x,\ y)}[\ \|x - y\|\ ] \tag{3}$$

其中,$\prod(P_1,\ P_2)$ 是 P_1 和 P_2 2 个分布所有可能的联合分布的集合,对于每一种可能的联合分布 λ,从中采样 $(x,\ y) \sim \lambda$ 得到一对样本 x 和 y,这对样本之间的距离为 $\|x - y\|$。在 λ 的联合分布下,样本 x 和 y 对距离的期望值为 $E_{(x,\ y)}[\ \|x - y\|\ ]$。在所有可能的联合分布中,样本对距离期望值的下界,就是 Wasserstein 距离。

传统的生成对抗网络模型在训练过程中,生成的数据分布直接匹配真实样本的分布,不易学习到样本数据的所有关键性特征,并且收敛速度比较慢。在本文的改进模型中,使用自编码器作为判别器,通过 Encoder 和 Decoder 对真实样本进行重构,得到重构后的损失分布 $\Gamma(x)$。2 种完全相同的数据分布,那么其损失分布也应该是相同的。$\Gamma(x)$ 与生成样本数据损失分布 $\Gamma(G(z))$ 的 Wasserstein 距离可以反映出 2 种分布之间存在的差异性,从而可以进一步优化生成模型。

2.3　平衡参数 γ

本文定义 2 个正态分布 $\eta_1 = N(m_1,\ c_1)$ 和 $\eta_2 = N(m_2,\ c_2)$,其中均值 m_1、$m_2 \in \mathbb{R}^p$,协方差 c_1、$c_2 \in \mathbb{R}^{p\times p}$。根据 Wasserstein 公式,两个正态分布之间的 Wasserstein 距离的平方如式(4)所示。

$$W(\eta_1,\eta_2)^2 = \|m_1 - m_2\|_2^2 + trace(c_1 + c_2 - 2(c_2^{1/2}c_1c_2^{1/2})^{1/2}) \tag{4}$$

其中,$trace(\ )$ 表示求迹的操作,当 $p = 1$ 时,式(4)可简化为式(5)。

$$W(\eta_1,\eta_2)^2 = \|m_1 - m_2\|_2^2 + c_1 + c_2 - 2(c_1c_2)^{1/2} \tag{5}$$

由于本文采用的是优化判别器重构损失分布之间的 Wasserstein 距离达到优化模型的目的,因此,只要式(5)满足单调性,即 $\dfrac{c_1 + c_2 - 2(c_1c_2)^{1/2}}{\|m_1 - m_2\|_2^2}$ 为一个常数或者单调递增函数,就可以对 2 个分布之间的 Wasserstein 距离的平方进行优化,即满足式(6)。

$$W(\eta_1,\eta_2)^2 \propto \|m_1 - m_2\|_2^2 \tag{6}$$

生成样本经过判别器的损失分布为 $\Gamma(G(z))$,真实样本数据经过判别器的损失分布为 $\Gamma(x)$,在训练过程中,当 2 个分布满足 $m_1 = m_2$ 时,则认为生成器和判别器保持平衡稳定训练的状态如式(7)所示。

$$E[\Gamma(x)] = E[\Gamma(G(z))] \tag{7}$$

但是从 2 个分布之间的 Wasserstein 距离优化方面考虑,当 $m_1 = m_2$ 时,$\dfrac{c_1 + c_2 - 2(c_1c_2)^{1/2}}{\|m_1 - m_2\|_2^2}$ 便趋近于无穷大,此时模型无法进行优化,甚至会造成模型崩溃。为了解决模型不稳定的问题,本文引入一个平衡参数 $\gamma \in [0,1]$,用于平衡生成器与判别器,避免 $m_1 = m_2$ 的条件发生,使其中一方不会赢过另一方,从而使模型的训练过程更稳定。

$$\gamma E[\Gamma(x)] = E[\Gamma(G(z))] \tag{8}$$

式(8)显示了参数在训练过程中用于平衡生成器与判别器。本模型中,判别器有 2 个目的,即自动编码真实的图像和能够辨别真实的图像和生成的样本图像。参数 γ 可以保证判别器和生成器训练过程的稳定性,γ 值偏小时,判别器侧重于对真实图像进行自编码,所以生成的图像多样性会减少。本模型的目的在于生成帧间的图像,更侧重于生成图像的质量,因此在一定范围内,γ 的值越大,生成模型的

质量越好。

2.4 空间连续性

本文中的数据集由一段视频的每一帧图像组成,可以认为这些图像的特征在空间上具有连续性。使用 Adam 优化器在相邻两帧图像之间找到一个最优值 z_r 满足式(9)使 e_r 值最小。

$$e_r = \| x1_r - G(z_r) | - | x2_r - G(z_r) \| \quad (9)$$

在式(9)中,$x1_r$ 和 $x2_r$ 分别对应前一帧和后一帧图像,将 z_r 映射到图像空间,得到生成的帧间图像。这种方式能够实现视频帧间的图像生成,同时证明了经过训练的生成模型并不是对图像进行简单记忆,而是在训练过程中真正学习到了图像的特征和内容。在真实图像之间进行图像生成为提高视频分辨率提供了一种新的有效的方式。

在本模型中,放弃了传统 GAN 中直接匹配数据分布的方式,而是采用自编码器作为判别器,通过 Encoder 和 Decoder 对输入样本进行重构,得到对样本重构后的损失分布 $\Gamma(x)$,通过引入 Wasserstein 距离来衡量 $\Gamma(x)$ 与 $\Gamma(G(z))$ 两种分布的差异,相比于传统 GAN 直接匹配样本数据分布的方式,收敛速度更快,生成效果更好。同时,引入平衡参数 γ 平衡生成器与判别器,有效解决了传统 GAN 中模型过于自由不可控的缺陷,使训练模型更稳定,极大可能地避免了模型崩溃的问题。连续的帧图像组成的数据集,在图像的特征空间上具有一定的连续性,使用 Adam 在连续的两帧图像之间找到一个最优值并将其映射到图像空间,得到生成的帧间图像。

3 实验分析

3.1 生成模型质量评估

为了说明本文 SC-GAN 模型对图像的生成能力,本文分别采用数据量为 200k 的 CelebA 的人脸数据集和 50k 的动漫头像数据集 CartoonFaces 对生成模型的质量进行测试,两个数据集中均有不同角度,不同表情和不同亮度的图像。采用 PSNR 和 SSIM 指标对本文生成模型的质量进行评估。

本文采用学习率 $lr \in [5 \times 10^{-5}, 10^{-4}]$ 的 Adam 训练模型。输入模型图像的分辨率均为 64 × 64,batch _ size = 16,epoch = 300,在此条件下,本文对不同生成模型的结果做了以下的对比实验。

图 3 和图 4 显示了 CelebA 数据集和 CartoonFaces 数据集在不同生成模型下生成的随机样本,其中,每一行分别对应 DCGAN、EBGAN 和本文的模型。在相同条件下,本文使用的模型相比于 DCGAN 和 EBGAN,生成图像的清晰度和多样性均有一定的优势,视觉效果也更加逼真和自然。

表 1 显示了使用 PSNR 和 SSIM 对 DCGAN 模型、EBGAN 模型和本文模型 SC-GAN 的质量评估结

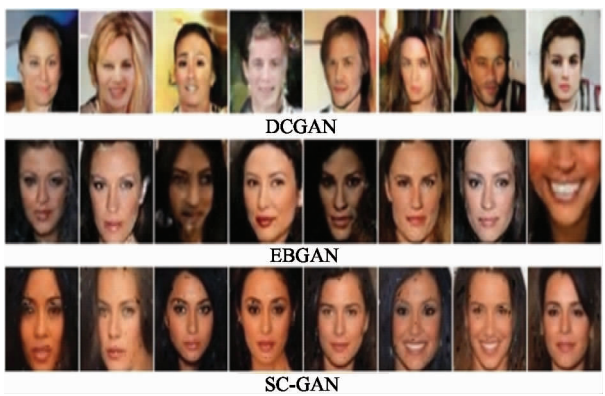

图 3 CelebA 数据集随机样本生成结果

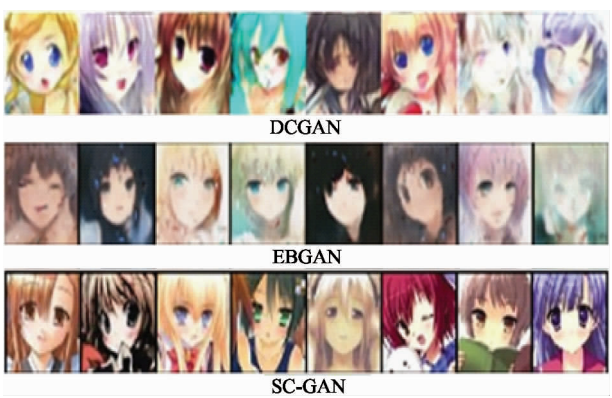

图 4 CartoonFaces 数据集随机样本生成结果对比

表 1 PSNR 和 SSIM 评估结果

Dataset	Method	PSNR	SSIM
CelebA	DCGAN[14]	30.0103	0.8924
	EBGAN[15]	31.1094	0.9007
	SC-GAN	32.5128	0.9218
CartoonFaces	DCGAN[14]	31.0845	0.9102
	EBGAN[15]	30.0026	0.8821
	SC-GAN	33.0249	0.9363

果。该结果显示本文模型使用不同的数据集生成的图像要更优于 DCGAN 和 EBGAN 模型,也证明了本文模型的质量也更优于其他的模型。

3.2　帧间图像生成

本文生成器和判别器的优化器均采用基于梯度的优化算法 Adam,为了使梯度下降法有更好的性能,将学习率 lr 设定在$[5\times10^{-5},\ 10^{-4}]$范围内。

为了说明平衡参数对模型在训练过程中的影响,本文选用场景较为简单的 M_{64} 数据集进行实验。在其他参数不变的情况下,分别对参数 γ 选取 0.3、0.5、0.7、0.9 进行 4 组实验,并对生成结果做出对比。

图 5 显示了 M_{64} 训练集在不同的 γ 值时的生成样本的结果,其中,每一列图像对应实验中参数 γ 的值分别为0.3、0.5、0.7和0.9。γ 较低时,生成的样本单一,多样性明显不足。γ 逐渐增大,生成的图像多样性更高,图像更清晰,γ 接近于临界值时,生成的样本质量变差,模型开始变得不稳定。当 γ 越接近 1,说明 m_1 的值就越接近 m_2,模型优化效果就会越来越差。

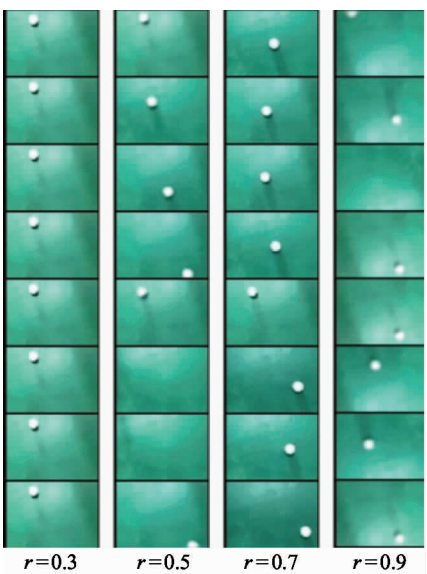

图 5　生成结果对比 $\gamma\in\{0.3,0.5,0.7,0.9\}$

为了实验中参数 γ 的选取更具有说服力,本文使用基于太极教学视频的 Taiji 数据集在相同的条件下,分别选取 $\gamma\in[0,1]$,以0.1为步长进行了 11 组实验,并以 PSNR 和 SSIM 评估的结果得到对不同 γ 值的评估结果折线图,如图 6 所示。

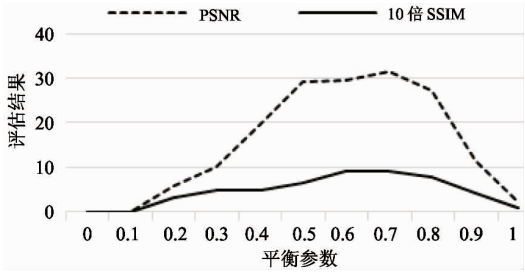

图 6　不同 γ 值评估结果折线图

为了更直观地显示评估结果的变化趋势,图中采用了 10 倍的 SSIM 值进行比较。根据对不同的 γ 值进行多次测试,权衡 PSNR 和 SSIM 的评估结果,在本文的实验中设定 γ 值为0.7。

在帧间图像生成实验中,使用基于太极教学视频的 Taiji 数据集和基于美国动画视频弹力球的 Ball 数据集。每个数据集均包含50 k幅图像,分辨率均为 64×64。

为了说明模型的生成能力,在帧间图像生成实验中,首先从数据集中选取不同场景的 6 组任意连续的 2 帧图像,然后分别将其输入到生成模型中,得到 6 组不同的帧间图像,并将其编号为 InFNo.1 ~ InFNo.6。

数据集 Taiji 和 Ball 中输入的第 1 帧图像和第 2 帧图像如图 7、图 8 所示。

图 7　Taiji 数据集模型输入图像

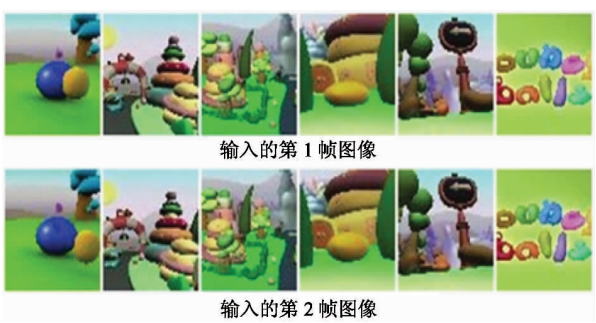

图 8　Ball 数据集模型输入图像

生成的帧间图像如图 9、图 10 所示。

图 9　Taiji 帧间图像生成结果

图 10　Ball 帧间图像生成结果

表 2 和表 3 分别为 Taiji 数据集和 Ball 数据集实验的评估结果。

表 2　Taiji 帧间图像评估结果对比

图像组编号	PSNR	SSIM
InFNo. 1	31.6415	0.9556
InFNo. 2	32.0383	0.9550
InFNo. 3	40.7233	0.9711
InFNo. 4	41.3777	0.9760
InFNo. 5	28.8292	0.9663
InFNo. 6	41.0436	0.9773

表 3　Ball 帧间图像评估结果对比

图像组编号	PSNR	SSIM
InFNo. 1	29.7372	0.8789
InFNo. 2	30.0038	0.8810
InFNo. 3	30.2259	0.8829
InFNo. 4	30.1032	0.8967
InFNo. 5	31.5499	0.8848
InFNo. 6	29.6556	0.8489

为了能够说明生成的帧间图像与真实图像之间相似度，本文进行帧间生成质量验证实验。首先从数据集中选取 6 组不同角度，不同场景及不同色调的任意连续 3 帧图像，然后将第 1 帧和第 3 帧图像输入到生成模型中，得到对应的 6 组帧间图像，分别将其编号为 InCNo. 1 ~ InCNo. 6。在本实验中增加了与真实视频帧对比的实验过程，将生成的帧间图像分别与 6 组中的第 2 帧真实图像进行对比，并使用 PSNR 和 SSIM 对其进行评估。

数据集 Taiji 和 Ball 的模型输入图像如图 11、图 12所示，第 1 行和第 2 行分别为输入的第一帧图像和输入的第 3 帧图像。

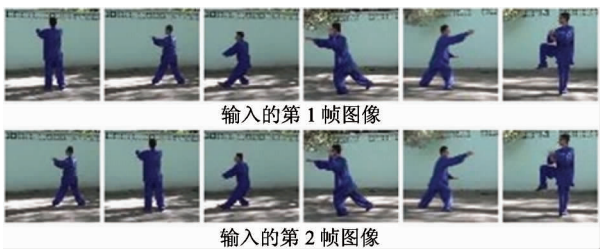

图 11　Taiji 数据集模型输入图像

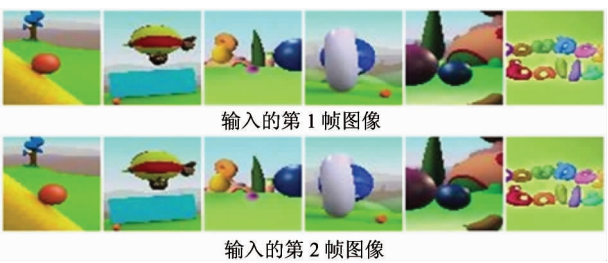

图 12　Ball 数据集模型输入图像

Taiji 和 Ball 数据集的帧间图像生成实验结果如图 13、图 14 所示，第 1 行和第 2 行分别为真实的第 2 帧图像和生成的第 2 帧图像。

图 13　Taiji 帧间图像生成对比结果

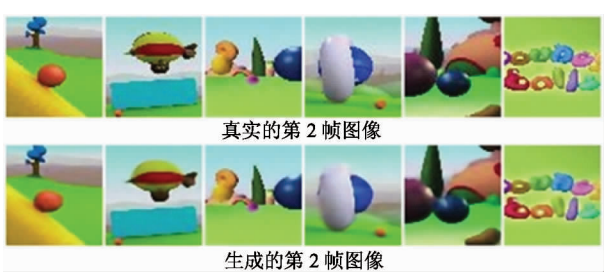

图 14　Ball 数据集帧间图像生成对比结果

Taiji 数据集和 Ball 数据集的评估结果分别如表 4 和表 5 所示。

表 4　Taiji 帧间图像评估结果对比

图像组编号	PSNR	SSIM
InCNo. 1	33.0463	0.9737
InCNo. 2	33.4009	0.9591
InCNo. 3	34.5261	0.9629
InCNo. 4	36.8102	0.9732
InCNo. 5	31.4005	0.9735
InCNo. 6	35.3619	0.9690

表 5　Ball 帧间图像评估结果对比

图像组编号	PSNR	SSIM
InCNo. 1	29.8137	0.8781
InCNo. 2	30.5376	0.9171
InCNo. 3	30.2366	0.9090
InCNo. 4	29.8696	0.8611
InCNo. 5	27.2198	0.6738
InCNo. 6	29.9387	0.8885

在实验过程中,模型的收敛情况如图 15 所示。

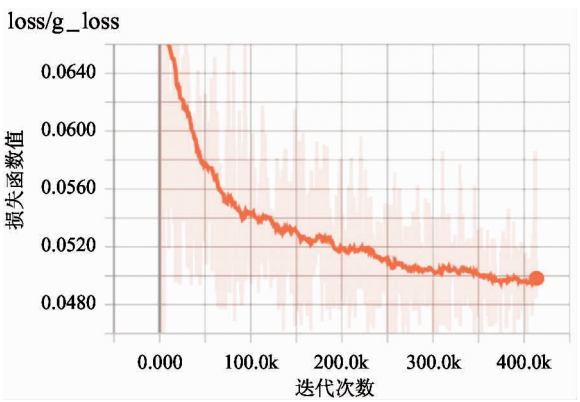

图 15　训练过程中模型收敛走势

在模型的质量验证实验中,相较于其他生成模型,GAN 模型生成的图像质量更高,因此,本文的对比实验采用相同的输入图像,基于 GAN 模型进行帧间图像生成实验,对比实验结果如图 16 和图 17 所示。图中第 1 行为输入的第 1 帧图像,第 2 行为真实的第 2 帧图像,第 3 行为生成的帧间图像,第 4 行为输入的第 3 帧图像。

评估结果分别如表 6 和表 7 所示。

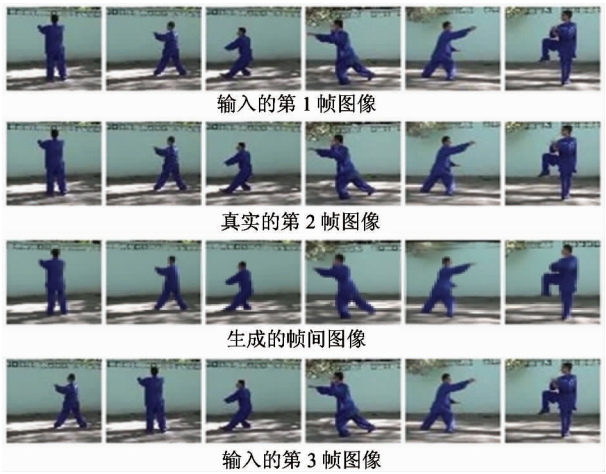

图 16　Taiji 数据集基于 GAN 生成结果

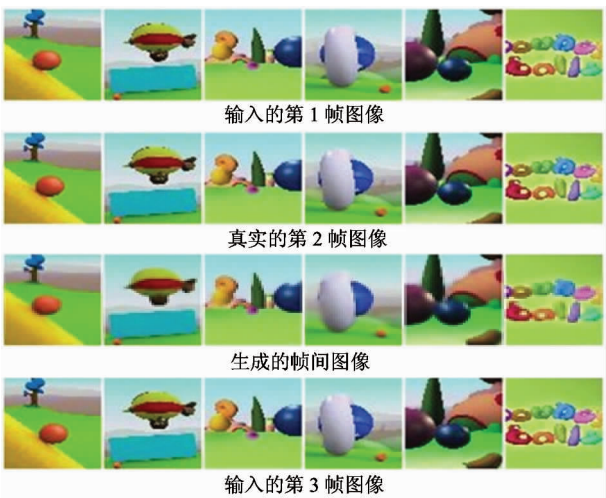

图 17　Ball 数据集基于 GAN 生成结果

表 6　Taiji 帧间图像 GAN 模型评估结果对比

图像组编号	PSNR	SSIM
InCNo. 1	24.4376	0.7039
InCNo. 2	26.6432	0.7934
InCNo. 3	25.8363	0.8301
InCNo. 4	25.4636	0.7294
InCNo. 5	23.7490	0.5829
InCNo. 6	26.3718	0.7493

表 7　Ball 帧间图像 GAN 模型评估结果对比

图像组编号	PSNR	SSIM
InCNo. 1	22.7739	0.6342
InCNo. 2	24.7438	0.6837
InCNo. 3	25.2936	0.7353
InCNo. 4	24.7493	0.6538
InCNo. 5	23.8469	0.6258
InCNo. 6	25.9736	0.7653

通过模型质量评估实验与 2 个阶段性的帧间图像生成实验的结果以及 GAN 模型的对比实验结果可知,本文构造的 SC-GAN 模型在基于视频的帧间图像生成中比传统方式能生成质量更好的图像,不具有传统方式带来的边缘模糊等问题,并且收敛速度快,生成图像更高效。相比于 GAN 模型,SC-GAN 模型从生成帧间图像的视觉效果和评估结果上都证明了其具有良好的生成能力。

在每一个阶段的实验中均进行多组不重复采样实验,并通过 PSNR 和 SSIM 两种方式对生成结果进行质量评估。在视觉上,本文构造的 SC-GAN 模型相比于其他生成模型,能生成质量更好的图像,在场景较为简单的视频中,更是无法区分出生成图像与真实图像,并且不会产生图像轮廓扭曲、图像模糊的问题。在量化结果上,SC-GAN 模型生成的帧间图像具有较高的真实性,并且与真实的视频帧之间具有很高的结构相似性。

在本实验中,针对分辨率为 64 × 64 的视频序列,在完成模型训练的情况下生成单帧图像的平均时间为 6.7 ms。将视频分辨率扩大为 512 × 512 进行相同测试,在完成模型训练的情况下生成单帧图像的平均时间为 24.3 ms,满足 25 Hz 到 50 Hz 进行帧间图像生成的实时性要求。考虑到本实验采用编程语言为 python 且计算设备为 GTX1060 显卡,这些因素限制了计算速度的提高。若改为C++ 语言编写并将计算设备速度提高,运行时间有望大幅缩短。

4 结 论

本文通过重构损失分布之间的 Wasserstein 距离对 GAN 模型进行优化,同时引入一个平衡参数 γ 维持模型稳定训练的状态,重点讨论了如何利用 GAN 的生成模型实现视频帧间的图像生成。平衡参数 γ 偏低时,生成的样本单一,多样性明显不足;γ 逐渐增大,生成的图像多样性更高,图像更清晰,γ 接近于临界值时,生成的样本质量变差,模型开始变得不稳定。实验中使用 CelebA 人脸头像数据集和 CartoonFaces 动漫头像数据集对不同的生成模型的质量进行了评估,首先从生成图像的质量以及评估的结果 2 个方面证明本文使用的生成模型更优于其他的生成模型。然后使用 Ball 和 Taiji 2 个数据集进行了帧间图像生成的实验。利用 Adam 选择 2 帧之间的最优数据分布并映射到图像空间,模型能够生成连续 2 帧之间的图像。为了进一步说明生成图像的可靠性和真实性,本文实验随机选取连续 3 帧图像中的第 1 帧和最后 1 帧图像输入到模型中,生成的帧间图像与中间的真实图像进行对比并采用 PSNR 和 SSIM 2 种形式进行了量化评估。通过生成结果和量化评估 2 种形式对模型的生成能力进行了说明。

本文算法通过对已有视频集进行分帧处理形成的数据集进行训练实现,并且本文的算法可以并行处理,因此,提高设备的计算能力,在并行资源足够的前提下,可以达到实时观看的效果,极大地改善了用户的体验。通过本文的实验也证明了针对已有的低帧率视频,通过利用图像特征在空间上的连续性进行帧间的图像生成是一种可行并且具有实际意义和应用价值的方法。

参考文献

[1] Ha T, Lee S, Kim J. Motion compensated frame interpolation by new block-based motion estimation algorithm [J]. *IEEE Transactions on Consumer Electronics*, 2004, 50(2):752-759

[2] Dikbas S, Altunbasak Y. Novel true-motion estimation algorithm and its application to motion-compensated temporal frame interoplation[J]. *IEEE Transactions on Image Processing*, 2013, 22(8):2931-2945

[3] Jeong S G, Lee C, Kim C S. Motion-compensated frame interpolation based on multi hypothesis motion estimation and texture optimization[J]. *IEEE Transactions on Image Processing*, 2013, 22(11):4497-4509

[4] 陈伟. 基于运动估计和运动补偿的帧率上转换算法研究 [D]. 哈尔滨:哈尔滨工业大学机电工程与自动化学院, 2016.6-17

[5] 虢齐. 基于深度学习的图像生成技术研究与应用 [D]. 成都:电子科技大学电子工程学院, 2017.25-40

[6] Park K, Yu S, Park S. An optimal low dynamic range image generation method using a neural network [J]. *IEEE Transactions on Consumer Electronics*, 2018, 64

(1):69-76

[7] 侯敬轩, 赵耀, 林春雨,等. 基于卷积网络的帧率提升算法研究[J]. 计算机应用研究,2018,35(2):611-614

[8] 龙古灿, 张小虎, 于起峰. 用于视频图像帧间运动补偿的深度卷积神经网络[J]. 国防科技大学学报, 2016, 38(5):143-148

[9] Mathieu M, Couprie C, Lecun Y. Deep multi-scale video prediction beyond mean square error[J]. *Electrical Engineering and Systems Science*, 2017,32(24):1091-1105

[10] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[J]. *Advances in Neural Information Processing Systems*, 2014,3(24):2672-2680

[11] 王坤峰, 苟超, 段艳杰,等. 生成式对抗网络 GAN 的研究进展与展望[J]. 自动化学报, 2017, 43(3):321-332

[12] 程学旗, 靳小龙, 王元卓,等. 大数据系统和分析技术综述[J]. 软件学报, 2014,25(9):1889-1908

[13] Berthelot D, Schumm T, Metz L. BEGAN: boundary equilibrium generative adversarial networks[J]. *Electrical Engineering and Systems Science*, 2017,21(15):2118-2123

[14] Mirza M, Osindero S. Conditional generative adversarial nets[J]. *Computer Science*, 2014, 10(21):2672-2680

[15] Poole B, Alemi A, Sohldickstein J, et al. Improved generator objectives for GANs[J]. *Computer Science*, 2016, 8(9):1060-1069

[16] Arjovsky M, Chintala S, Bottou L. Wasserstein GAN [J]. *Electrical Engineering and Systems Science*, 2017, 34(22):1201-1213

Inter-frame video image generation based on spatial continuity generative adversarial networks

Zhang Tao*, Zhang Meng*,**, Jiang Peipei*

(*School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004)

(**Neusoft Software Co., Ltd., Qinhuangdao 066004)

Abstract

A method for generating inter-frame video images based on spatial continuity generative adversarial networks (SC-GAN) is proposed to smooth the playing of low frame rate videos and to clarify blurry image edges caused by the use of traditional methods to improve the video frame rate. Firstly, this paper uses the auto-encoder as a discriminator. Introducing the Wasserstein distance represents the difference between the loss distribution of the real sample and the generated sample, instead of the traditional method of generative adversarial networks to directly match data distribution. Secondly, by using the balance parameter between generator and discriminator, the training process can be stabilized, which effectively prevents the model from collapsing. Finally, this paper uses the spatial continuity of the image features of continuous video frames and finds an optimal value between two adjacent frames by Adam. Then the value is mapped to the image space to generate inter-frame images. In order to illustrate the authenticity of the generated inter-frame images, this paper evaluates the inter-frame images by the use of PSNR and SSIM. The evaluation results show that the generated inter-frame images have a high degree of authenticity. The feasibility and validity of the proposed method based on SC-GAN are verified.

Key words: generative adversarial network (GAN), adversarial training, spatial continuity, Adam, inter-frame image generation