

基于生成对抗网络的图像清晰度提升方法^①



范晓烨, 王 敏

(河海大学 计算机与信息学院, 南京 210024)
通讯作者: 王 敏, E-mail: mwang@hhu.edu.cn

摘 要: 视频监控、军事目标识别以及消费型摄影等众多领域对图像清晰度有很高的要求. 近年来, 深度学习在视觉和定量评估的应用研究中取得较大进展, 但是其结果一般缺乏图像纹理的细节, 边缘过度平滑, 给人一种模糊的视觉体验. 本文提出了一种基于生成对抗网络的图像清晰度提升方法. 为了更好的传递图像的细节信息, 采用改进的残差块和跳跃连接作为生成网络的主体架构, 生成器损失函数除了对抗损失, 还包括内容损失、感知损失和纹理损失. 在 DIV2K 数据集上的实验表明, 该方法在提升图像清晰度方面有较好的视觉体验和定量评估.

关键词: 图像清晰度; 生成对抗网络; 曼哈顿距离

引用格式: 范晓烨, 王敏. 基于生成对抗网络的图像清晰度提升方法. 计算机系统应用, 2021, 30(2): 176-181. <http://www.c-s-a.org.cn/1003-3254/7775.html>

Improved Image Sharpness Method Based on Generative Adversarial Network

FAN Xiao-Ye, WANG Min

(College of Computer and Information, Hohai University, Nanjing 210024, China)

Abstract: Video surveillance, military object recognition, consumer photography, and many other fields have high requirements for image sharpness. In recent years, deep neural networks have made great progress in the applied research on visual and quantitative evaluation, but the results generally lack the details of image textures, and the edges are too smooth, providing blurry visual experience. For this reason, we propose a method of improving image sharpness based on the generative adversarial network in this study. In order to better delivery the image details, this method adopts the improved residual block and skip connection as the main structure of the generative network, and the generator loss function consists of content loss, perception loss, and texture loss in addition to adversarial loss. Finally, the experiments on the DIV2K dataset prove that the proposed method exhibits good visual experience and quantitative evaluation in terms of improving image sharpness.

Key words: image sharpness; Generative Adversarial Network (GAN); Manhattan distance

提升图像清晰度是计算机视觉领域中的常见问题. 它已经成为视频监控、军事目标识别以及消费型摄影、摄像等众多应用领域的一个重要研究课题. 模糊的图像严重影响目标识别、语义理解等高级视觉任务, 所以, 对模糊图像的处理, 提高图像清晰度具有重要意义, 因此受到了计算机视觉研究领域的高度关注.

最近, 在图像修复^[1,2], 超分辨率^[3-7]和去模糊^[8-10]相关研究领域中, 通过应用生成对抗网络 (GAN)^[11]取得了重大进展. 它是一种基于真实图像与生成图像比较来进行训练的模型, 以博弈的思想使得输出尽可能的接近输入的真实分布, 该训练方式可以很好地从样本数据中学习特征, 更好地解决分类和回归问题. GAN

^① 收稿时间: 2020-06-10; 修改时间: 2020-07-10, 2020-07-17; 采用时间: 2020-07-21; csa 在线出版时间: 2021-01-27

包括两部分:生成器和判别器.生成器生成图像,判别器判断图像是来自真实分布还是生成网络.在训练期间,判别网络计算对抗损失对生成网络进行评分.GAN可以保留图像的纹理细节,创造出接近真实图像的生成图像,令人无法分辨是否为真实图像.因此受到生成对抗网络对图像处理的启发,通过使用GAN也可将提高图像清晰度视为图像到图像的转换.

本文采用GAN的思想,生成网络中采用改进的残差块(residual block)作为主体,生成器损失函数在原有基础上加入内容损失、感知损失、纹理损失,在视觉和定量评估方面取得较大进展.

1 国内外相关工作

卷积神经网络在计算机视觉任务中普遍应用,从识别,语义分割到降噪,超分辨率,去模糊,修复,都表现出了卓越的性能.深度网络以端到端的多层方式将低级、中级、高级特征和分类器集成在一起,并且可以通过增加网络深度来丰富特征.

一味地增加网络层数带来的问题就是梯度消失,不仅难以训练,而且收敛速度变慢,为解决上述问题,He等提出了ResNet^[12],它可以在增加特征图通道数的同时增加网络深度,而不增加网络的复杂度,并通过实验证明残差网络收敛速度更快,当网络深度显著增加时,其准确率也会随之上升.这是选择ResNet作为GAN模型中生成网络主体部分的原因之一.

之所以选择GAN作为网络框架,是因为我们想要的是和真实图像无法区分的输出,这正是最近提出的GAN所研究的^[13-15].Isola等^[16]提出图像到图像的转换问题通常是逐像素的分类或回归问题,在本文中,通过逐像素的方式计算输出和真实图像之间的内容损失.

原始ResNet中的快捷连接仅仅是简单地执行恒等映射,并将其输出添加至几个网络层之后的输出,恒等快捷连接既不增加额外的参数也不增加计算复杂度.文献[14]为了解决图像分类等高级的计算机视觉问题,因此将ResNet体系结构直接应用于提高图像清晰度这类的低级计算机视觉问题是达不到最优效果的.

传统残差网络中有批归一化操作,但是在提高图像清晰度的应用中,残差块中BN是没有必要的,因为残差块中快捷连接已经在不增加额外的参数和计算复杂度的前提下,加速了训练过程,因此本文的生成网络部分中的残差网络也将去除BN层.同样重要的是He

和Srivastava等^[17]说明跳跃连接可以将梯度反向传播到底层,并将图像细节传递到顶层,有助于减轻网络中携带相同信息的负担并获得高质量的局部最优化.在提高图像清晰度的过程中要考虑到图像的局部多样性,因此残差结构非常适用于提高图像清晰度.

Pathak等^[18]发现将GAN与传统的损失函数混合起来对于训练效果是有益的,例如 L_2 损失函数.但是如果我们要求GAN最小化预测像素和真实像素之间的 L_2 距离,它将倾向于产生模糊的结果.Zhao等^[19]实验使用 L_1 损失函数训练的网络比使用 L_2 可以获得更好的性能表现,收敛也更快.

综上所述,为了生成图像尽可能逼真,所以使用GAN框架,并和不同损失函数结合.在提高图像清晰度方向上,要考虑到图像特征的范围灵活性和局部多样性,所以生成器采用去除BN的残差块作为主体结构.

2 基于生成对抗网络的图像清晰度提升方法

本文遵循文献[20]的思想,生成网络由多个残差块组成.GAN中的生成网络和判别网络均使用SAME模式进行卷积,该操作会保持原特征图的维数不变,从而实现无损特征提取.

2.1 生成对抗网络

受到文献[11]的启发,GAN是一个最大最小的博弈问题,生成网络希望生成图像尽可能地接近真实分布,致使判别网络认为是真的;而判别网络则是尽可能地分辨出图像的真假,如式(1)所示:

$$\max_{\theta_G} \max_{\theta_D} E_{x \sim p_{\text{data}}} [\log D(x)] + E_{x \sim p_G} [\log(1 - D(x))] \quad (1)$$

其中, p_{data} 为真实分布, p_G 为生成器分布, D 和 G 分别表示判别器和生成器.将传统的最大最小问题应用于本文中便有式(2):

$$\max_{\theta_G} \max_{\theta_D} E_{I^T \sim p_T(I^T)} [\log D_{\theta_D}(I^T)] + E_{I^B \sim p_G(I^B)} [\log(1 - D_{\theta_D}(G_{\theta_G}(I^B)))] \quad (2)$$

其中, I^B 为模糊图像, I^T 为真实清晰图像.

2.2 网络整体架构

提高图像清晰度对图像纹理生成的要求极高,因此GAN中的生成网络至关重要.它需要对通过网络层的信息进行整合,为了改善反向传播中的梯度流,快捷连接可以增强残差网络的梯度流.

本文GAN中生成网络和判别网络的总体结构如

图 1, 图 2 所示. 生成网络主体部分由改进的残差块 (去除 BN 层) 和跳跃连接嵌套而成, 不使用池化层和反池化层, 因为池化层会丢失图像细节, 然而这些细节信息对提高图像清晰度至关重要. 中间由 5 个改进的残差块组成, 利用跳跃连接进行像素级相加操作, 保证了将图像细节传递到特征图顶层, 有利于降低网络的复杂度, 获得高质量的局部最优化, 将梯度反向传播到底层, 使得训练深层网络更加容易.

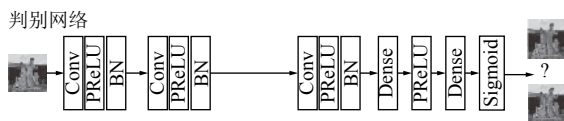


图 1 判别网络架构

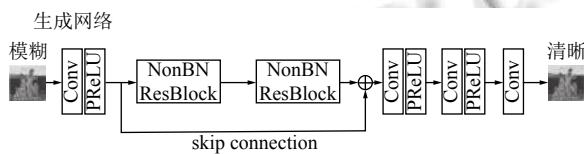


图 2 生成网络架构

判别网络部分主要由 8 个卷积-激活-归一操作组成, 用两个全连接层在尾部进行重新拟合, 减少特征信息的损失, 根据特征的组合进行分类, 大大减少特征位置对分类带来的影响. 最后用 Sigmoid 对全连接层的输出进行二分类, 判断输入图像真假与否.

2.3 改进的残差块

如图 3 所示图 3(a) 为原始的残差块, 其适用于图像分类等高级计算机视觉问题, 均达到较高水平, 在提高图像清晰度这类低级计算机视觉问题中, 原始的残差块并没有表现出满意的效果, 其中 BN 层是不必要的模块, 将其去除可简化网络结构, 在稳定 GAN 训练过程的同时, 性能也有所提升, 图 3(b) 为改进的残差块.

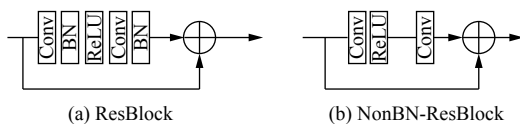


图 3 残差块

另外将网络中的 ReLU 激活函数改为了 PReLU, 整流激活函数对神经网络性能的巨大提升功不可没, 由文献 [21] 提出的 PReLU 激活函数可以自适应地学习整流器的参数, 在 ImageNet 分类上, PReLU 是超越

人类分类水平的关键所在, 在可忽略的计算成本下提高精度.

实验证明, 选用 LReLU 和 ReLU 作为整流器对最终的实验结果几乎没有影响, 因此采用可以不断寻求最优参数的 PReLU 作为整流器.

2.4 损失函数

本文 GAN 模型的损失函数主要分为两部分: 生成器损失函数和判别器损失函数, 其中生成器损失函数由内容损失、感知损失、纹理损失和对抗损失构成, 下面将进行展开介绍.

2.4.1 内容损失

本文 GAN 模型的损失函数在原来最大最小问题的基础上加入内容损失函数, 文献 [4, 18] 使用的是均方误差 (L_2) 作为损失函数, 这可以在定量评估方面取得很高的分数, 但是它会导致图像的细节纹理太过平滑, 给人视觉上有模糊的感觉. 文献 [19] 通过实验发现使用曼哈顿距离 (L_1) 训练出来的模型性能更好, 训练过程的收敛速度也比 L_2 快, 因此使用 L_1 作为本文模型中的内容损失函数:

$$l_{con} = \frac{1}{WH} \sum_{i=1}^W \sum_{j=1}^H |I_{i,j}^T - G_{\theta_G}(I^B)_{i,j}| \quad (3)$$

其中, W 和 H 分别表示图像的宽和高.

2.4.2 感知损失

当两幅图像内容完全一样, 但相对位置却偏差 1 个像素时, 求得的内容损失却不是 0, 其关注的是图像中每一个对应位置像素之间的距离, 因此鲁棒性不强. 感知损失允许模型的输出可能与真实图像不完全匹配, 但鼓励网络生成具有类似特征表示的图像, 如式 (4) 所示:

$$l_{pec} = \|\varphi(G_{\theta_G}(I^B)) - \varphi(I^T)\|_2^2 \quad (4)$$

本文使用预训练的 VGG-19 网络来获取特征图 φ , 为了同时考虑到图像中的多尺度信息捕获高级和低级特征, 因此将第 2 个和第 5 个池化层结果融合. 避免了要求网络输出图像与真实清晰图像在像素级别上的一致, 而是鼓励两幅图具有相似的特征.

2.4.3 纹理损失

Ulyanov 等^[22] 在风格迁移中将全局纹理合成到其他图像, 与风格迁移任务的区别在于, 本文任务中图像中的纹理特征多样化, 不仅限于一种风格, 因此在训练过程中通过计算基于块的纹理损失, 使得网络生成局

部相似的纹理,可以促进产生的图像具有更丰富的纹理信息.如式(5)所示:

$$l_{\text{tex}} = \left\| Gr(\varphi(G_{\theta_G}(I^B))) - Gr(\varphi(I^T)) \right\|_2^2 \quad (5)$$

其中, $Gr(F) = FF^T$.

2.4.4 对抗损失

生成器损失函数还有原始的对抗损失,其鼓励生成网络能够生成更加自然的图像.对抗损失如式(6):

$$l_{\text{adv}} = \sum_{n=1}^N -\log D_{\theta_D}(G_{\theta_G}(I^B)) \quad (6)$$

我们希望生成图像真实到可以愚弄判别网络,即 $D_{\theta_D}(G_{\theta_G}(I^B))D_{\theta_D}(G_{\theta_G}(I^B))$ 趋向于 1,为了同判别器损失函数保持一致,我们将式(6)改写为 $-D_{\theta_D}(G_{\theta_G}(I^B))$,对其求最小化,不影响其收敛性.

2.4.5 判别器损失

GAN网络的损失函数中除了生成器损失函数,还包括判别器损失函数,判别网络同时对生成图像和真实图像进行判断,判别器损失函数如式(7):

$$l_{\text{dis}} = \sum_{n=1}^N -(\log(1 - D_{\theta_D}(G_{\theta_G}(I^B))) + \log D_{\theta_D}(I^T)) \quad (7)$$

式中,对于真实样本 I^T 而言 $D_{\theta_D}(I^T)$ 越大越好;对于假样本 $G_{\theta_G}(I^B)$ 而言 $D_{\theta_D}(G_{\theta_G}(I^B))$ 越小越好,为了这两项的期望趋势相同,我们将 $D_{\theta_D}(G_{\theta_G}(I^B))$ 改写为 $1 - D_{\theta_D}(G_{\theta_G}(I^B))$.

3 实验与分析

3.1 DIV2K 数据集

DIV2K数据集^[22]是新发布的用于图像复原任务的高质量图像数据集,包括800张训练图像,100张验证图像和100张测试图像,图像种类包罗万象应有尽有,所以本文采用该数据集进行实验,尽可能增加模型的泛化能力.

3.2 实验设计

由于数据集图像过大,因此训练过程中对图像进行随机裁剪 24×24 并翻转,使用Adam优化算法进行500000次迭代,训练25万次后进行学习速率衰减,衰减率设置为0.1,训练生成网络每更新一次,判别网络更新两次,防止GAN训练的不平衡.

3.3 实验结果与分析

本小结通过消融实验证明改进的残差块和损失函数对于提高图像清晰度的有效性,如图4所示,图4(a)和图4(b)分别为生成模型的输入和真实清晰图像,图4(c)至图4(f)为分别进行实验得出的效果图.具体

来说,图4(a)为输入的模糊图像,图4(b)为真实清晰图像,图4(c)为没有使用残差块的生成效果,定义该生成模型为StackedG;图4(d)为使用传统残差块的生成效果,定义该生成模型为ResG;图4(e)为使用改进残差块的生成效果,定义该生成模型为NonBN-ResG;图4(f)为使用改进残差块和损失函数的生成效果,定义该生成模型为NonBN-ResG-L.

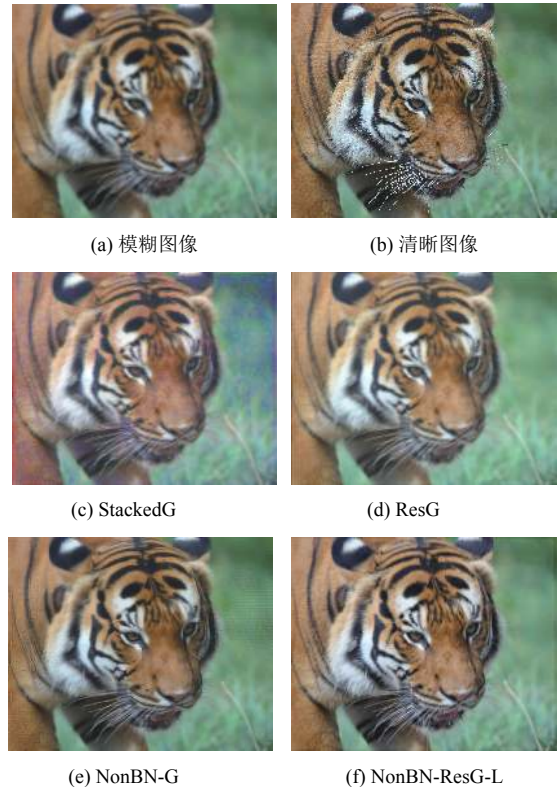


图4 消融实验

如图5所示,更加细节地展示实验效果,为了方便起见标号同图4.对该女人的人脸细节观察,图5(c)有明显的色彩失真,图5(d)则有振铃效应,图5(e)的细节重建不足,较为模糊,图5(f)的纹理特征,精细化程度均明显高于其他3种情况,包括头发、眼睛、牙齿,甚至身后的文字均清晰可见,与真实的高清晰度图像相差无几.表1为定量分析,可得出使用 L_1 作为内容损失函数训练的网络虽然在定量评估方面没有达到最优,但比 L_2 训练获得更好的性能表现.

4 结论与展望

本文提出了一种基于生成对抗网络的提升图像清晰度方法,生成网络主体部分由改进的残差块和跳

跃连接嵌套而成,保证了将图像细节传递到特征图顶层,有利于降低网络的复杂度,获得高质量的局部最优化,将梯度反向传播到底层,使得训练深层网络更加容易。

将来视频监控等众多领域对图像清晰度会有更高的要求,目前有多项工作研究重点放在超分辨率的工作上,因此未来我们可以将提高清晰度和超分辨率结合起来,在实现超分辨率的基础上并提高图像清晰度,这可以成为将来的研究重点。

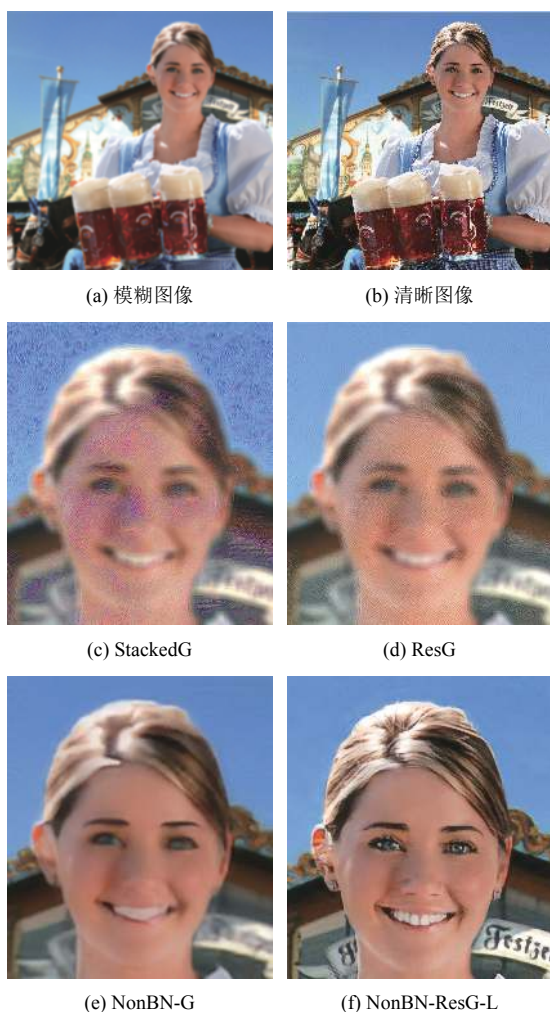


图5 实验细节对比

表1 不同模型的定量评估

模型	PSNR	SSIM
StackedG	25.11	0.804
ResG	25.82	0.890
NonBN-ResG	28.54	0.917
NonBN-ResG-L	28.39	0.933

参考文献

- Demir U, Unal G. Patch-based image inpainting with generative adversarial networks. arXiv: 1803.07422, 2018.
- Mao XJ, Shen CH, Yang YB. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain. 2016. 2810–2818.
- Zhang YL, Tian YP, Kong Y, *et al.* Residual dense network for image super-resolution. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 2472–2481.
- Ledig C, Theis L, Huszár F, *et al.* Photo-realistic single image super-resolution using a generative adversarial network. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 105–114.
- Wang XT, Yu K, Wu SX, *et al.* Esrgan: Enhanced super-resolution generative adversarial networks. Proceedings of European Conference on Computer Vision. Munich, Germany. 2018. 63–79.
- Shi WZ, Caballero J, Huszár F, *et al.* Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 1874–1883.
- Dong C, Loy CC, He KM, *et al.* Image super-resolution using deep convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(2): 295–307. [doi: 10.1109/TPAMI.2015.2439281]
- Kupyn O, Budzan V, Mykhailych M, *et al.* DeblurGAN: Blind motion deblurring using conditional adversarial networks. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 8183–8192.
- Kupyn O, Martyniuk T, Wu JR, *et al.* DeblurGAN-v2: Deblurring (orders-of-magnitude) faster and better. Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Republic of Korea. 2019. 8877–8886.
- Shao WZ, Liu YY, Ye LY, *et al.* DeblurGAN+: Revisiting blind motion deblurring using conditional adversarial networks. Signal Processing, 2020, 168: 107338. [doi: 10.1016/j.sigpro.2019.107338]
- Goodfellow IJ, Pouget-Abadie J, Mirza M, *et al.* Generative adversarial nets. Proceedings of the 27th International

- Conference on Neural Information Processing Systems. Cambridge, UK. 2014. 2672–2680.
- 12 He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 770–778.
- 13 Denton E, Chintala S, Szlam A, *et al.* Deep generative image models using a laplacian pyramid of adversarial networks. Proceedings of the 28th International Conference on Neural Information Processing Systems Montreal, QC, Canada 2015. 1486–1494.
- 14 Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv: 1511.06434, 2015.
- 15 Zhao JB, Mathieu M, LeCun Y. Energy-based generative adversarial network. arXiv: 1609.03126, 2016.
- 16 Isola P, Zhu JY, Zhou TH, *et al.* Image-to-image translation with conditional adversarial networks. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 5967–5976.
- 17 Srivastava RK, Greff K, Schmidhuber J. Training very deep networks. Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal, QC, Canada. 2015. 2377–2385.
- 18 Pathak D, Krähenbühl P, Donahue J, *et al.* Context encoders: Feature learning by inpainting. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 2536–2544.
- 19 Zhao H, Gallo O, Frosio I, *et al.* Loss functions for neural networks for image processing. arXiv: 1511.08861, 2015.
- 20 Johnson J, Alahi A, Li FF. Perceptual losses for real-time style transfer and super-resolution. Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherland. 2016. 694–711.
- 21 He KM, Zhang XY, Ren SQ, *et al.* Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile. 2015. 1026–1034.
- 22 Ulyanov D, Lebedev V, Vedaldi A, *et al.* Texture networks: Feed-forward synthesis of textures and stylized images. Proceedings of the 33rd International Conference on International Conference on Machine Learning. New York City, NY, USA. 2016. 1349–1357.