量化权值激活的生成对抗网络

郑哲[1,2,3] 胡庆浩[2] 刘青山[1,3] 冷聪[2]

（南京信息工程大学 自动化学院，南京 210044）[1]

（中国科学院自动化研究所南京人工智能芯片创新研究院，南京 211100）[2]

（江苏省大数据分析技术重点实验室，南京 210044）[3]

摘要 近年来，生成对抗网络（Generative Adversarial Networks，GAN）在图像超分辨率、图像生成等许多计算机视觉任务中展现出优异的性能。借助于 GPU 强大的计算力，人们可以设计计算复杂度更高的 GAN 网络。然而，对于资源受限的移动端设备，高功耗、计算需求大的 GAN 将很难直接部署到实际应用中。得益于神经网络压缩技术取得的巨大进展，GAN 部署到移动端设备才成为可能。为此，文中提出一种同时对网络权值和激活进行量化的方案来压缩 GAN 网络，通过量化敏感性分析，发现与量化分类网络不同，GAN 中量化权重比量化激活更敏感，因此在量化时给予权重更多量化比特。文中比较两种评价 GAN 生成图像的指标 Inception Score(IS)和 Fréchet Inception Distance（FID），发现 FID 更适合评估量化后 GAN 性能。文中基于敏感性分析在 Mnist 和 Celeb-A 数据集上进行量化实验，用 FID 指标来评估量化 GAN 性能。实验结果表明：在生成图像质量不下降的情况下，依然可以取得 4 倍以上的压缩率，从而有效地解决了 GAN 的压缩问题。

关键词 生成对抗网络，资源受限，移动端设备，神经网络压缩，量化

中图法分类号 TP391.41/TP183 **文献标识码** A **DOI** （投稿时不提供 DOI 号）

Quantizing both weights and activations in Generative Adversarial Networks

Zhe zheng[1,2,3] Qinghao hu[2] Qingshan liu[1,3] Cong leng[2]

(School of Automation, Nanjing University of Information Science &Technology, Nanjing 210044)[1]

(CASIA-AIRIA, Nanjing 211100)[2]

(Jiangsu Key Lab of Big Data Analysis & Technology, Nanjing 2110044)[3]

AbstractIn recent years, Generative adversarial network has shown excellent performance in many computer vision tasks such as image super resolution, image generation and so on. GANs can be designed to be much more greedy in computation complexity because of huge quantity use of GPU application. For mobile devices that are resource-limited, however, on which it is intractable for GAN to be deployed due to high consumption both in energy and computation. Thanks for Great success in neural network compression, it presents methods to make it possible. In this paper, we propose a method to simultaneously quantize weights and activations in GANs.Sensitivity analysis shows that weights are more sensitive than activation in quantization process. We use Fréchet Inception Distance score to evaluate generated images of quantized GANs for Inception score is less applicable than FID . Motivated by sensitivity analysis, extensive experiments are conducted on Mnist and Celeb-A datasets. Results show that we can compress GANs by up to 4x and still achieve even higher performance than the original GANs. Thus, it effectively resolves the problem of compressing GANs.

Key words Generative adversarial networks; Source-limited; Mobile devices; Neural network compression; Quantization

到稿日期：返修日期：

郑哲（1995-），男，硕士在读，主要研究方向为深度网络压缩，E-mail：qsliu@nuist.edu.cn(刘青山)，CCF 会员。

1. 引言

近些年，生成对抗网络[1]（Generative Adversarial Networks，GAN）在计算机视觉领域取得了巨大成功，被广泛应用于图像生成、图像超分辨率、视频生成等任务。早期的生成对抗网络由全连接层组成，其网络较小，性能也差强人意。得益于深度学习[2]的快速发展，GAN 将卷积神经网络作为主干网络并在图像超分辨率、图像生成等计算机视觉任务上，取得了优异的性能。在深度学习中，全连接层相对于卷积层具有计算量小参数多的特点，而卷积层由于卷积操作和参数共享的特点，其参数较少但计算量较大。由于卷积层数的增多，先进主流的卷积网络模型往往具有占用空间大，计算消耗大的特点。因此，深度学习所具有的如上特点在 GAN 中也有体现。由于大量卷积层的使用，GAN 在前向推理时需要耗费更多的时间。例如，由谷歌开发的 BigGAN[3]，在图像生成任务中取得了最好的结果，但其模型占用空间达 1.9GB，前向推理需每秒上亿次浮点运算，模型的参数量和计算量都十分巨大。因此，将 GAN 应用到资源受限的移动端将变得非常困难。

目前，神经网络压缩特别是在图像分类领域已经取得了巨大的进展。网络压缩方法主要有五种：网络剪枝、低秩、量化、知识蒸馏、轻量级网络设计。这些方法都可以保证原始网络精度和网络压缩率的折中，主要有两种情况：第一种是网络精度优先，在模型性能损失不大的情况下尽可能提高压缩率；第二种是压缩率优先以满足硬件需求的情况下，尽可能提高模型的精度。本文属于第一种，在性能无损的情况下，采用量化方法将 GAN 网络的比特数由浮点表示量化成低比特表示，直接在硬件上加速并减少模型占用的空间，以获得较高的压缩率。

即使图像分类任务上的网络压缩取得了巨大的进展，但是应用于 GAN 上的压缩工作依旧很少[4][5]，QGAN[4]基于 EM 算法提出新的量化函数对生成网络和判别网络进行权值的量化而保持激活为浮点，在低比特甚至是 1bit 的情况下依然取得可观的性能。Angeline[5]等人用知识蒸馏的方法，用一个大的教师 GAN 网络指导学习出一个小的 GAN 网络来获得较高的压缩率。与上述工作不同，对于图像生成任务而言，其最终目标是能够生成逼真的图片，也由于判别器的量化会导致生成器很难获得最优的梯度，从而使得生成网络很难收敛。因此，本文保持判别网络权重和激活为浮点，只对生成网络进行量化压缩。因此需要对权重和激活的量化进行敏感性的分析，而 GAN 中的权重和激活对量化的敏感性与量化分类网络相反（敏感性分析见 4.1 节）：量化权重比量化激活更敏感即量化权重对性能的影响比量化激活对性能的影响更大。本文主要工作如下：

- 通过敏感性分析，发现 GAN 中量化权重比量化激活更敏感，因此在量化时对生成网络的权重进行较高比特量化，对激活进行较低比特量化。
- Fréchet Inception Distance（以下简称 FID）[6]和 Inception Score[7]（以下简称 IS）是两种主要的评估 GAN 性能的方法。实验发现，FID 在评估量化 GAN 的性能比 IS 更合适。
- 本文是第一个同时将 GAN 模型的权重和激活同时量化的工作，实验表明，我们实现在性能无损的情况下依然能取得较高的压缩率。

2. 相关工作

2.1 生成对抗网络

生成对抗网络由两个网络组成：判别网络 D 和生成网络 G。D 判断 G 生成的图像是否满足真实图像分布即图像是否为真，G 从预设的分布 $z \sim N(0,1)$ 或 $z \sim U(-1,1)$ 中采样来尽可能生成逼真的图像来欺骗判别网络。通过二者的博弈，从而使得生成网络不断提升自身的能力直到能够欺骗判别网络。其训练的损失函数如下：

$$\min_G \max_D V(D,G) = \mathrm{E}_{\mathbf{x}\sim pdata(\mathbf{x})}[\log D(\mathbf{x})] + \mathrm{E}_{\mathbf{z}\sim p(\mathbf{z})}[\log(1-G(\mathbf{z}))] \quad (1)$$

通过最小化图像生成分布和真实分布的 JS 散度来训练生成网络。由于 GAN 训练不稳定易出现模式崩塌，早期的工作[8][9]探索更好的网络结构。目前先进的 GAN 网络结构大多采用 DCGAN[10]中所用的网络设置，为了进一步提高 GAN 的性能，有工作[11][12]提出了新的损失函数。目

前，BigGAN 生成的图像已经达到了以假乱真的效果。在此基础上，本文主要研究的是在性能无损的情况下，如何去降低 GAN 的模型存储，加快其推理速度。

2.2 量化

量化技术在网络压缩中直接有效，其将浮点型权重映射到对应的低比特定点表征空间，利用低比特来压缩模型，利用定点数运算来提高运算速度。例如，BNN[13]将网络中的权重和激活量化为 1bit，相较于原始网络可以带来 32 倍的压缩率，由于激活的二值化，这使得可以采用 xnor 操作快速的在硬件上实现前向推理，同时降低大量的内存占用和计算消耗。

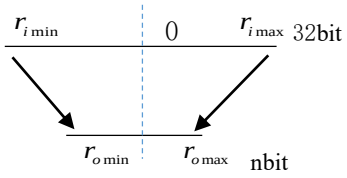
目前，有许多应用于图像分类网络的量化方法。权重量化方面：早期的工作 BinaryConnect[14]将模型的权重二值化（-1，1）并在小数据集上取得了近乎无损的效果。BWN[15]引入浮点型因子来提高网络表征能力，提升了二值化网络在大数据集上的性能，但与原始网络在准确率上仍有较大差距。BWNH[16]从哈希保持内积的角度递进求解每层最优参数。ABCNet[17]将卷积核近似成若干二值基卷积核的线性组合。INQ[18]提出增量量化的策略。High-order[18]通过不断近似量化的残差来减少量化误差。HBNN[20]提出在神经元层次上学到各自的量化比特。HAQ[21]将不同量化比特对硬件的影响作为输入，通过强化学习来自动学得网络每层的量化比特数。权重激活量化方面：BNN 在 BinaryConnect 基础上将权重和激活同时二值化。Dorefa[22]通过对浮点型权重、激活、梯度选择不同的量化 bit 数来取得模型在性能、大小上的折中。LQ-Net[23]通过让网络学到最优的量化器来减少性能损失。GroupNet[24]从网络结构近似的角度学得原始网络对应的二值网络。而目前只有 QGAN 对 GAN 网络进行量化，其基于最大期望算法提出新的量化函数对生成网络和判别网络进行权值的量化，在低比特甚至是 1bit 的情况下依然取得可观的性能。本文与 QGAN 不同的地方在于：本文仅对生成网络进行量化，同时量化生成网络的权重和激活，保持判别网络为浮点，QGAN 则是保持激活不变，对生成网络和判别网络的权值进行量化；采用 FID 来有效评估量化后模型生成图像的质量。

3. 具体方法

本节分别介绍权重量化和激活量化及其对应的前向传播和反向传播，对权重进行多比特量化，对激活二值量化。

3.1 量化方法

量化浮点权重和浮点激活为低比特，即将高比特空间的值映射到低比特空间。如下图一所示，



图一 低比特量化示意图

Fig.1 low bit quantization

图一中，n 为量化比特数，r_i 和 r_o 分别表示原始浮点参数和量化后的定点数，$r_{i\min}$ 和 $r_{i\max}$ 分别表示 r_i 的最小值和最大值。通过量化函数 Q 将输入 r_i 映射到 r_o 上，即 $r_o = Q(r_i)$ 。显然，量化比特数越小，量化误差越大，由此带来的模型性能损失也越大，因此在量化时我们需要在性能和压缩之间进行折中。在量化之前，我们有必要对 GAN 网络中权重和激活进行敏感性分析，确定量化对权重和激活的具体影响来指导量化。与 HWGQ[25]一致，我们采用二值量化来进行敏感性的验证，此时 $r_{o\min} = -1$，$r_{o\max} = 1$ 敏感性分析见 4.1 节。由于量化 GAN 权重比量化激活对模型性能的影响更大，因此本文对权重给予更多的量化比特数，给予激活较少的量化比特数。

3.1.1 权重量化

BWN[15]提出了一个有效的二值量化方法，其将浮点权重近似为一个常数与二值矩阵点乘。假设要量化的卷积核 $W \in \Re^{chw}$，c、h和w分别为卷积核的通道数、高和宽，则卷积核可近似表示为 $\mathbf{W} \approx \alpha \mathbf{B}$，其与输入 $\mathbf{I}$ 的卷积表示为

$$\mathbf{I} * \mathbf{W} \approx \alpha(\mathbf{I} \oplus \mathbf{B}) \tag{2}$$

其中，$\oplus$ 表示无乘法的卷积运算，$\mathbf{B}=\text{sign}(\mathbf{W})$，$\alpha = \frac{1}{chw}\|\mathbf{W}\|_1$。

Dorefa[22]中提出多比特量化函数来量化卷积网络，可以将浮点参数量化至用户设定的比特数。其量化函数为

$$r_o = \frac{1}{2^k - 1} round((2^k - 1) r_i) \tag{3}$$

其中，k 为要量化的比特数, round 为取整操作。

训练时，由量化后的权重进行前向传播计算梯度，由于权重在量化后，损失函数将对其不可导，为了解决影响参数更新的不可导问题，我们采用 STE[26]来解决这个问题，即在反向传播时，损失函数 c 对 r_i 的导数近似为

$$\frac{\partial c}{\partial r_i} = \frac{\partial c}{\partial r_o} \tag{4}$$

虽然，量化权重可以极大的压缩模型的物理占用空间，但是其在计算复杂度上并无太大的提高，因此对激活进行量化显得十分必要。同时量化权重和激活，可以采用更快速更易实现的逻辑和位计数运算[13][15]，来极大的减少计算量。本文采用 BWN 来进行 1bit 量化，采用 Dorefa 来进行多 bit 量化。

3.1.2 激活二值量化

激活二值化中，一个直接有效的方法是通过 sign 函数来量化。在 XnorNet 中，对权重二值化而引入的浮点因子在激活二值化中影响可忽略，因此，我们直接对激活进行 c 量化，即前向的量化函数为

$$r_o = sign(r_i) \tag{5}$$

由于 sign 函数不可导，因此采用 STE，近似损失函数对输入的导数为

$$\frac{\partial c}{\partial r_i} = \frac{\partial c}{\partial r_o} \mathrm{I}_{|r_i \le 1|} \tag{6}$$

其中 $\mathrm{I}_{|r_i \le 1|}$ 为输入 r_i 绝对值小于 1 的参数，大于 1 的参数梯度截断为 0。

4. 实验

我们在 DCGAN、WGAN[11]上进行实验，采用 pytorch 框架，在 Celeb-A 数据集上评估 IS 和 FID 能否反映生成图像质量的好坏，并对 GAN 中权重和激活的量化进行敏感性分析，最后在 Mnist 和 Celeb-A 给出不同比特数量化下的结果。

4.1 生成图像评价

IS 和 FID 被广泛应用来评估生成网络生成图像的质量和多样性。**IS 越大，FID 越小**，则代表生成图像质量越高，模型性能越好。然而，[27]指出用 IS 来评价生成图像质量并不理想，较大的 IS 并不能代表生成的图片就越真实。因此，对于量化后 GAN 模型性能的评估，我们需要确定 IS 和 FID 是否能够反映生成图像质量的高低，以此来评估量化方法是否有效。我们共采样 5 万张量化后模型生成的图像，分别计算其与原始数据集图像的 IS 和 FID，具体结果如表一和表四所示。

表一中给出了 DCGAN 和 WGAN 在 Celeb-A 数据集上的 IS 得分。表中 W/A 中 W 代表权重量化的比特数，A 代表激活量化的比特数。我们分别对两个网络进行不同比特数的量化。可见 DCGAN 在 1/32 取得了最高 IS，WGAN 在 4/1 取得了最高 IS。

表一 在 CelebA 上 DCGAN/WGAN 不同比特数 IS

Table 1 IS score of DCGAN and WGAN on dataset CelebA with different quantization bits

W/A	DCGAN	WGAN
32/32	2.47	2.40
1/32	**2.87**	-
32/1	2.51	2.31
8/1	2.70	2.314
4/1	2.43	**2.637**

为了验证该指标是否能反映图像质量，我们生成了对应的可视化图像。然而从可视化生成的图像来看，IS 指标并不能反映图像质量的好坏。可视化生成图像如图二示。

（a）32bit/32bit

（b） 1bit/32bit

（c）4bit/1bit

图二 DCGAN:（a）IS=2.47，（b）IS=2.87；WGAN:(c) IS=2.637

Fig.2 IS score of DCGAN and WGAN by W/A bits

从图二中可见，（b）和（c）的图像质量明显差于（a）

中的图像，（c）中出现模式崩塌却依然可以得到很高的 IS，可见采用 IS 来评估模型性能并不合适[27]。结合表五和图六、图七的结果来看，生成图像的质量与 FID 有较好的吻合度。因此，与 QGAN[4]中采用 IS 来评估量化后模型和原始模型所生成图像质量不同，我们采用 FID 来评估模型性能。

4.2 量化敏感性分析

在量化图像分类模型中，激活量化比权重量化更敏感[25]。如下表二所示，

表二 AlexNet 在 ILSVRC12 数据集量化准确率[25]

Table 2 Top1/Top5 accuracy of quantized AlexNet on dataset ILSVRC12

准确率	1/32	32/1
Top1	53.9	46.7
Top5	77.3	71

可以看出保持激活为浮点，只对权重量化的模型性能明显好于仅量化激活的模型性能。因此，我们需要验证：GAN 的量化是否也有相同的现象。

本文在 Celeb-A 数据集上做敏感性分析，并在 DCGAN 和 WGAN 上分别对比权重二值激活浮点和权重浮点激活二值的 FID 得分。

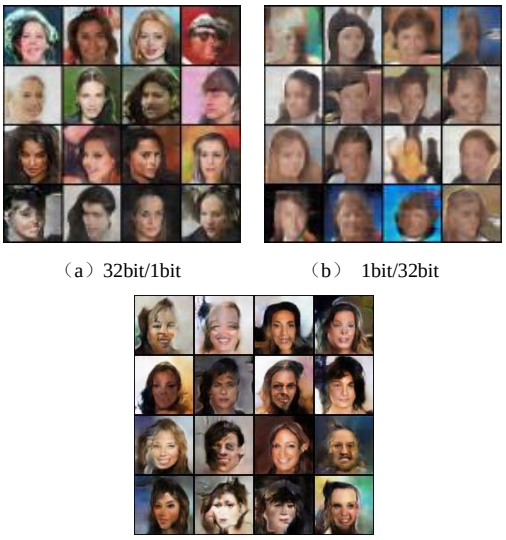
表三 DCGAN/WGAN 敏感性分析

Table 3 Sensitivity analysis of DCGAN and WGAN

W/A	DCGAN	WGAN
32/1	**81.6**	**92.3**
1/32	220.1	不收敛

从表三可以看出，DCGAN 中 1/32 的 FID 得分为 220.06，32/1 的 FID 得分为 81.56。WGAN 中 32/1 的 FID 得分为 92.30，1/32 不收敛。所以我们可以得到：在 GAN 的量化中，**权重量化比激活量化更加敏感**。

由于 WGAN 在权重二值激活浮点的情况下不能收敛，因此我们对剩余三种情况下进行可视化，从图三中可以看出在超参数设置和量化函数相同的情况下，只二值化激活生成的图片明显好于只二值化权重所生成的图片，这一现象与量化图像分类模型相反即 GAN 中量化权重比量化激活更敏感。因此，相较于激活，我们在对上述 GAN 进行量化时给予权重更多的量化比特数。



（a）32bit/1bit　　（b） 1bit/32bit

（c）32bit/1bit

图三 DCGAN：（a）FID=81.6（b）FID=220.1

WGAN:（c）FID=92.3

Fig.3 FID score of DCGAN and WGAN by W/A bits

4.3 Mnist 数据集

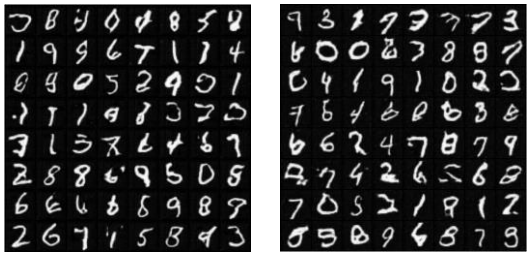
Mnist 数据集是一个黑白手写数字数据库，包含从 0 到 9 的 55000 张训练集图像，10000 张测试集图像，每张图片通道数为 1，大小为 28x28 像素。表四给出了 DCGAN、WGAN 在不同比特数量化下的 FID 得分。

表四 Mnist 数据集 FID 得分

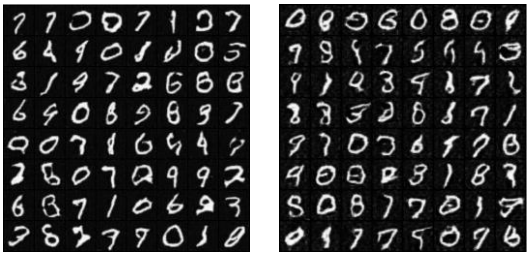
Table 4 FID score on dataset Mnist

W/A	DCGAN	WGAN
32/1	4.66	64.02
8/1	14.98	**56.08**
4/1	102.75	/

进一步地，我们将不同比特数量化后 GAN 所生成的图像可视化。从图四和图五中可见，量化后模型依然取得了较好的效果，为了进一步验证量化的有效性，我们在较大 Celeb-A 数据集上进行实验。



（a）32bit/32bit　　（b）32bit/1bit



（c）8bit/1bit　　（d）4bit/1bit

图四 DCGAN 不同比特数量化结果

Fig.4 Results of quantized DCGAN by different bits

WGAN 生成图像如图五所示，

（a）32bit/32bit　　（b）32bit/1bit

（c）8bit/1bit

图五 WGAN 不同比特数量化结果

Fig.5 Results of quantized WGAN by different bits

4.4 CelebA 数据集

CelebA 是一个大规模的名人人脸属性数据集，其中有 10,177 个名人身份的 202,599 张人脸图片，CelebA 由香港中文大学开放提供，广泛用于人脸相关的计算机视觉训练任务，可用于人脸属性标识训练、人脸检测训练以及 landmark 标记。我们设置原始模型和量化模型的训练超参数相同，生成图片共 5 万张来计算 FID。量化结果如表五所示，

表五 DCGAN/WGAN 不同比特数 FID 得分

Table 5 FID score of quantized DCGAN and WGAN with different bits

W/A	DCGAN	WGAN
32/32	103.86	91.07
32/1	81.56	92.30
8/1	85.37	**88.39**
4/1	**81.45**	189.64

从表五中我们可以看出，量化后的 DCGAN 和 WGAN 在 FID 指标上明显小于原始模型。生成图像分别如图六和图七所示。

（a）32bit/32bit　　（b）32bit/1bit

（c）8bit/1bit　　（d）4bit/1bit

图六 DCGAN 不同比特数量化结果

Fig. 6 Results of quantized DCGAN by different bits

从图六中可以看出，DCGAN 即使在权重为 4bit 激活为 1bit 的情况下，其生成的图像质量仍媲美原始模型生成的图像。

（a）32bit/32bit　　（b）32bit/1bit

（c）8bit/1bit　　（d）4bit/1bit

图七 WGAN 不同比特数量化结果

Fog. 7 Results of quantized WGAN by different bits

图七中，量化权重为 8bit 激活为 1bit 的 WGAN，其性能较于原始模型并没有损失，而当权重量化为 4bit 时，由于过低比特量化导致的模型表征能力受限，模型性能显著下降。因此，在一定比特数下对 GAN 进行量化，其生成的图像

质量上甚至优于原始模型，这是由于量化的正则化[14]作用。

5. 结论

本文同时对 GAN 的权重和激活量化，发现 FID 比 IS 更能对量化模型进行准确的评估；通过权重和激活的量化敏感性分析，发现与量化分类网络不同，在量化 GAN 中，量化权重比量化激活更加敏感；与分类网络量化相似，当权重和激活都量化为较低比特时，模型性能显著下降，这是由于低比特表示的参数使得模型相较于浮点型缺少强大的表征能力，未来的工作将着重解决低比特量化下模型性能显著下降的问题。

参 考 文 献

[1]. Goodfellow Ian J.,Generative adversarial nets.//Proceedings of the Neural Information Processing Systems.Montreal,Canada,2014.2672—2680

[2]. Krizhevsky Alex., ImageNet classification with deep convolutional neural networks.//Proceedings of the Neural Information Processing Systems. Montreal,Lake Tahoe,US,2012,1106-1114.

[3]. Brock, A., Donahue, J., & Simonyan, K. Large Scale GAN Training for High Fidelity Natural Image Synthesis. *ArXiv, abs/1809.11096*. 2019.

[4]. Wang, P., Wang, D., Ji, Y., Xie, X., Song, H.F., Liu, X., Lyu, Y., & Xie, Y. QGAN: Quantized Generative Adversarial Networks.arXiv:1901.08263.2019

[5]. Angeline A., Chiang P,. Alex Gain, Ameya Patil, Kolten Pearson,Soheil Feizi.Compressing gans using knowledge distillation.arXiv:1902.00159,2019.

[6]. Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, Sepp Hochreiter, GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium.arXiv:1706.08500.2017

[7]. Salimans, Tim and Goodfellow, Ian and Zaremba, Wojciech and Cheung, Vicki and Radford, Alec and Chen, Xi and Chen, Xi., Improved Techniques for Training GANs.//Proceedings of the Neural Information Processing Systems.2016.2234—2242

[8]. Mirza, M. and Osindero, S. Conditional generative adversarial nets.arXiv:1411.1784, 2014

[9]. Karras, T., Aila, T., Laine, S., and Lehtinen, J. Progressive growing of gans for improved quality, stability, and variation.arXiv:1710.10196, 2017

[10]. Radford, A., Metz, L., and Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv:1511.06434, 2015

[11]. Arjovsky, M., Chintala, S., and Bottou, L Wasserstein gan.arXiv:1701.07875, 2017.

[12]. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. C Improved training of wasserstein gans.//Proceedings of the Neural Information Processing Systems.2017.5767–5777

[13]. Hubara, I., Courbariaux, M., Soudry, D., El-Yaniv, R., Bengio, Y.: Binarized neural networks. //Proceedings of Neural Information Processing Systems . 2016.4107–4115

[14]. M. Courbariaux, Y. Bengio, and J. David. Binaryconnect: Training deep neural networks with binary weights during propagations.//Proceedings of the Neural Information Processing.Montreal,Canada,2015..3123-3131

[15]. M.Rastegari, XNOR-Net: ImageNet Classification Using Binary Convolutional Neural Networks.//Proceedings of the European Conference on Computer Vision.Amsterdam,Holland,2016,525-542.

[16]. Qinghao Hu, Peisong Wang and Jian Cheng, From Hashing to CNNs: Training BinaryWeight Networks via Hashing.//Proceedings of the AAAI Conferenceon Artificial Intelligence. New Orleans, Louisiana.USA.2018,3247-3254

[17]. Lin, X., Zhao, C., Pan, W.: Towards accurate binary convolutional neural network..//Proceedings of the Neural Information Processing Systems.Montreal,2017.345–353

[18]. Aojun Zhou, Anbang Yao, Yiwen Guo, Lin Xu, Yurong Chen. Incremental Network Quantization: Towards Lossless CNNs with Low-Precision WeightsarXiv preprint arXiv:1702.03044(2017)

[19]. Li, Z., Ni, B., Zhang, W., Yang, X., Gao, W.: Performance guaranteed network acceleration via high-order residual quantization.//Proceedings of the International Conference on Computer Vision 2017.2584–2592

[20]. Josh Fromm, Shwetak Patel, Matthai Philipose.Heterogeneous Bitwidth Binarization in Convolutional Neural Networks.//Proceedings of the Neural Information Processing Systems,Montréal,Canada,2018.4006—4015

[21]. Kuan Wang, Zhijian Liu, Yujun Lin, Ji Lin, Song Han.HAQ: Hardware-Aware Automated Quantization.arXiv:1811.08886.2018

[22]. Shuchang Zhou, Yuxin Wu, Zekun Ni, Xinyu Zhou,He Wen, Yuheng Zou., DoReFa-Net: Training Low Bitwidth Convolutional Neural Networks with Low Bitwidth Gradients.arXiv:1606.06160.2016

[23]. Zhang, Dongqing and Yang, Jiaolong and Ye, Dongqiangzi and Hua, Gang, LQ-Nets: Learned Quantization for Highly Accurate and Compact Deep Neural Networks.//Proceedings of European Conference on Computer Vision.2018

[24]. Bohan Zhuang, Chunhua Shen, Mingkui Tan, Lingqiao Liu, Ian Reid, Structured Binary Neural Networks for Accurate Image Classification and Semantic Segmentation.arXiv:1811.10413.2018

[25]. Cai, Z., He, X., Sun, J., Vasconcelos, N.: Deep learning with low precision by halfwave gaussian quantization. //Proceedings of IEEE Conference on Computer Vision and Pattern Recognitio

n. New York ,IEEE Press.2017. 5918–5926.

[26]. Bengio, Yoshua, L´eonard, Nicholas, and Courville, Aaron. Estimating or propagating gradients through stochastic neurons for conditional computation.arXiv:1308.3432, 2013

[27]. Shane Barratt, Rishi Sharma.A Note on the Inception Score. arXiv:1801.01973.2018

作者简介：

郑哲，男，目前研究生二年级，就读于南京信息工程大学自动化学院控制科学与工程系。 z.zheng@nuist.edu.cn

胡庆浩，男，博士，2019 年 6 月博士毕业于中科院自动化研究所，在 AAAI,ECCV，ACM-MM 发表多篇文章。 huqinghao@airia.cn

刘青山，男，南京信息工程大学自动化学院院长，教授，博士生导师，2003 年毕业于中科学院自动化研究所，IEEE 高级会员。 qsliu@nuist.edu.cn

冷聪，男，博士，中国科学院自动化研究所南京人工智能芯片创新研究院副院长，2016 年博士毕业于中科院自动化研究所。lengcong@airia.cn