智能系统学报
CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

基于二进制生成对抗网络的视觉回环检测研究

杨慧，张婷，金晟，陈良，孙荣川，孙立宁

引用本文:
杨慧, 张婷, 金晟, 等. 基于二进制生成对抗网络的视觉回环检测研究[J]. 智能系统学报, 2021, 16(4): 673–682.
YANG Hui, ZHANG Ting, JIN Sheng, et al. Visual loop closure detection based on binary generative adversarial network[J]. *CAAI Transactions on Intelligent Systems*, 2021, 16(4): 673–682.

在线阅读 View online: https://dx.doi.org/10.11992/tis.202007007

DOI: 10.11992/tis.202007007
网络出版地址: https://kns.cnki.net/kcms/detail/23.1538.TP.20210412.1346.010.html

基于二进制生成对抗网络的视觉回环检测研究

杨慧，张婷，金晟，陈良，孙荣川，孙立宁
（苏州大学 机电工程学院，江苏 苏州 215021）

摘　要：针对现有的回环检测模型大多基于有监督学习进行训练，需要大量标注数据的问题，提出一种视觉回环检测新方法，利用生成对抗思想设计一个深度网络，以无监督学习的方式训练该网络并提取高区分度和低维度的二进制特征。将距离传播损失函数和二值化表示熵损失函数引入神经网络，将高维特征空间的海明距离关系传播到低维特征空间并增加低维特征表示的多样性，进而利用 BoVW 模型将提取的局部特征融合为全局特征用于回环检测。实验结果表明：相比 SIFT 和 ORB 等特征提取方法，所述方法在具有强烈视角变化和外观变化的复杂场景下具有更好的性能，可以与 AlexNet 和 AMOSNet 等有监督深度网络相媲美。但采用无监督学习，从根本上避免了费时费力的数据标注过程，特别适用于大规模开放场景的回环检测，同时二进制特征描述符极大地节约了存储空间和计算资源。

关键词：回环检测；无监督学习；二进制描述符；BoVW；视觉 SLAM；生成对抗；特征提取；深度学习
中图分类号：TP181　　**文献标志码**：A　　**文章编号**：1673−4785(2021)04−0673−10

中文引用格式：杨慧，张婷，金晟，等. 基于二进制生成对抗网络的视觉回环检测研究 [J]. 智能系统学报，2021，16(4)：673–682.
英文引用格式：YANG Hui, ZHANG Ting, JIN Sheng, et al. Visual loop closure detection based on binary generative adversarial network[J]. CAAI transactions on intelligent systems, 2021, 16(4): 673–682.

Visual loop closure detection based on binary generative adversarial network

YANG Hui, ZHANG Ting, JIN Sheng, CHEN Liang, SUN Rongchuan, SUN Lining
(School of Mechanical and Electric Engineering, Soochow University, Suzhou 215021, China)

Abstract: In view of the problem that the existing loop closure detection models are mostly trained based on supervised learning and require a large amount of labeled data, this paper proposes a new method for visual loop closure detection. The idea of the generative adversarial network is adopted, and thus, a deep neural network is designed and trained through unsupervised learning methods to extract more discriminative binary feature descriptors with low dimensions. The distance propagation loss function and a binarized representation entropy loss function are introduced into the neural network. The first loss function can help spread the Hamming distance relationship of the high-dimensional feature space to the low-dimensional feature space, and the second one increases the diversity of the low-dimensional feature representation. The extracted local features are fused into global features by using the BoVW model for further loop closure detection. Experimental results show that the proposed method has better performance than feature extraction algorithms such as SIFT and ORB in complex scenes that have a strong viewpoint and appearance changes, and its performance is comparable with that of supervised deep networks such as AlexNet and AMOSNet. It is especially suitable for loop closure detection in large-scale open scenes because the time-consuming and tedious process of supervised data annotation is completely avoided with the use of unsupervised learning. Moreover, the binary feature descriptors can greatly save storage space and computing resources.

Keywords: loop closure detection; unsupervised learning; binary descriptor; BoVW; visual SLAM; generative adversarial; feature extraction; deep learning

收稿日期：2020−07−08.　**网络出版日期**：2021−04−12.
基金项目：国家自然科学基金面上项目 (61673288).
通信作者：陈良. E-mail: chenl@suda.edu.cn.

利用三维空间中的信息进行避障、定位以及和三维空间中的物体进行交互对于移动机器人等

自主无人系统来说是必不可少的能力。通常，三维感知能力由定位和建图两部分组成。当前主流的方法支持同步定位与建图，即 SLAM(simultaneous localization and mapping)。在 SLAM 系统中，机器人需要对自身所处的环境进行建图并同时估计自己的位姿[1]。视觉 SLAM 系统主要包括 3 个部分：前端视觉里程计、后端优化、回环检测[2]。其中，回环检测的目的在于判断机器人所在区域是否处于以前访问过的区域，以便消除机器人在长时间导航与定位中产生的累计误差，对于机器人进行准确定位以及地图构建起着至关重要的作用[3]。但是，机器人在利用视觉 SLAM 进行导航时不可避免地会面临光照变化、季节更替、视角改变、动态场景等情况，这些因素都会导致回环检测的性能大大降低，从而影响机器人定位的准确性以及地图构建的可靠性，因此需要更加鲁棒以及稳定的回环检测方法。

针对视觉回环检测问题，目前主流的方法主要分为传统方法以及基于深度学习的方法[4]。SIFT[5](scale invariant feature transform) 及 SURF[6](speeded up robust feature) 等是目前使用较为广泛的传统特征提取方法。前者对尺度及光照都具有一定的鲁棒性，但在提取特征时十分耗时，运行效率较为低下。SURF 相比于 SIFT 计算效率有所提高，但对旋转以及尺度变换的鲁棒性却远远低于 SIFT。SURF 和 SIFT 描述符都属于局部描述符，为了让基于局部描述符的方法应用于视觉 SLAM 系统，应用于自然语言处理及检索领域的词袋模型被引入视觉领域，形成了视觉词袋模型 BoVW[7](bag of visual word)。该方法主要分为提取视觉词汇、构建视觉词典、计算相似度 3 个部分。提取视觉词汇即利用 SURF 或者 SIFT 提取图片的局部特征，形成不同的视觉单词向量。将所有特征向量进行聚类，构建包含若干视觉词汇的词典。测试时，将输入图片与视觉词典进行对比得到该图片在视觉词典中的直方图，计算两张图片直方图之间的距离即可完成相似度计算。BoVW 模型对于环境变化，例如尺度变化、旋转以及视角变化具有鲁棒性，但研究表明该方法在光照变化严重的情况下表现不佳。

近年来，随着深度学习的迅速发展，越来越多基于深度学习的特征提取方法被提出。Chen 等[8]率先利用 ImageNet 的预训练卷积神经网络 (convolutional neural network, CNN) 模型提取图片的深度特征并与空间和序列滤波器相结合应用于场景识别，实验表明该方法在场景识别中精度较高。文献 [9] 第一次提出了基于卷积神经网络的场景识别系统，通过将 CNN 中高层和中层提取的特征相结合，实现了较为鲁棒的大规模场景识别。

上述特征提取方法都存在一定的局限性。SURF、SIFT 等人工特征描述符无法自动提取图片深层特征，需要人为设计特征描述符，随着大规模开放场景下数据集规模的不断增加，手工设计全面且准确的特征描述符越来越困难。而基于 CNN 等深度学习的方法虽然可以自动提取图片的深度特征，但在模型训练时大多使用有监督学习，需要大量的有标签数据，而数据的标注过程费时费力。

因此，研究基于无监督学习的特征表达，是当前机器视觉领域的研究热点和难点。Gao 等[10] 使用堆栈去噪自编码器 (stacked denoising auto-encoder, SDA) 模型进行无监督回环检测。然而，该方法需要离线训练，且训练集和测试集相同，因此实用性不强。最近，生成对抗网络 (generative adversarial network, GAN)[11] 作为一种新的无监督学习方法受到越来越多的关注，成为新的研究热点。GAN 作为一种优秀的生成模型，与其他生成模型，如自编码器 (auto-encoder, AE)[12]、受限玻尔兹曼机 (restricted Boltzmann machine, RBM)[13] 相比，无需大量的先验知识，也无需显式地对生成数据的分布进行建模。由于 GAN 独特的对抗式训练方法，在训练过程中可以从大量的无标签数据中无监督地学习数据的特征表达，同时生成高质量的样本，相比于传统机器学习算法具有更强大的特征学习以及特征表达能力。因此，GAN 被广泛应用于机器视觉等领域。也有学者将 GAN 应用于回环检测任务中[14]。该方法从鉴别器的高维特征空间中提取特征描述子。但是，该方法提取的特征描述子维度较高，会占用大量的存储空间以及计算资源。

受 Shin 等[14] 的启发，本文以无监督学习的方式训练 GAN 来进行回环检测。考虑到低维二进制描述子能够降低存储资源的消耗，同时加速回环检测的决策过程。因此，本文在鉴别器中加入激活函数，将传统的非二进制描述子转换成二进制描述子。同时为了弥补低维特征所带来的信息损失，提高二进制特征描述符的区分度，使其在复杂场景外观变化下具有鲁棒性，本文将距离传播损失函数 L_{DP}(distance propagating) 和二值化表示熵损失函数 L_{BRE}(binarized representation entropy) 引入鉴别器中，将高维特征空间的海明距离关系传播到低维特征空间中，并利用 BoVW 模型将提取的局部特征融合为全局特征用于回环检测。实验结果表明，该描述符可以解决复杂场景

下的回环检测问题，对于视角及环境变化具有较强的鲁棒性，用生成对抗的方式开展无监督回环检测不但是可行的，而且以该方法生成的二进制特征描述符具有较高的区分度，减少了低维特征的信息损失。

综上所述，本文创新点总结如下：1) 提出一种视觉回环检测新方法，该方法利用生成对抗的思想设计一个深度网络以无监督的方式训练该网络，并利用该网络提取高区分度和低维度的二进制特征；2) 将距离传播损失函数引入神经网络，将高维空间之间的海明距离关系传播到低维空间，使高维空间特征与低维空间特征具有相似的距离关系；3) 将二值化表示熵损失函数引入神经网络，提高了低维特征空间二进制描述符的多样性，进一步弥补低维特征所带来的信息损失；4) 利用 BoVW 模型将提取的局部特征融合为全局特征，有助于大规模开放场景下的回环检测。

1　无监督二进制描述符的提出

1.1　生成对抗思想

GAN 由生成器 G(Generator) 和鉴别器 D(Discriminator) 组成，二者在训练时相互对抗，相互进化。在训练时，生成器 G 的主要目标是学习潜在样本的数据分布，并生成尽可能真实的新样本以骗过鉴别器 D，而鉴别器 D 则要判断出输入数据的真实性，即输入数据是来自真实数据还是来自由生成器 G 生成的虚假数据。根据上述思想，Goodfellow 等[13] 给出了 GAN 的损失函数：

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

式中：x 表示真实数据；$D(x)$ 为鉴别器判断 x 为真实数据的可能性；z 代表输入生成器的随机变量；$G(z)$ 为生成器 G 生成的尽量服从真实数据分布的虚假样本；$D(G(z))$ 表示鉴别器 D 判断 $G(z)$ 为虚假数据的概率。鉴别器 D 的目标是对输入数据进行正确的二分类，而生成器 G 的目标则是让其生成的虚假数据 $G(z)$ 在鉴别器 D 上的表现 $D(G(z))$ 和真实数据 x 在鉴别器 D 上的表现 $D(x)$ 尽可能一致。

1.2　无监督二进制描述符的定义

GAN 不仅具有强大的生成能力，而且研究表明可将 GAN 的鉴别器 D 作为特征提取器，其表现同样令人满意[15-16]。原因在于 GAN 在进行对抗训练的过程中，生成器 G 会生成质量不断提高的虚假图像，而鉴别器为了提高判断准确性，不断提升自身的特征表达能力以提取更有区分度的特征。因此，本文利用 GAN 的鉴别器 D 作为视觉回环检测任务的特征提取器，其优势在于可以充分利用生成对抗的思想进行特征的无监督学习，不需要额外的标签数据，也不需要人工干预，就可以自动获得区分度高的特征描述符。

文献 [16] 表明，从鉴别器 D 的高维中间层中提取的特征具有更高的区分度，但是高维特征需要更多的存储空间以及消耗更多的计算资源。因此，大多数研究中都会将高维特征进行降维以减少其对存储空间的占用，提高回环检测的运行速度。但是降维操作会不可避免地导致特征描述符损失信息。因此，本文将距离传播损失函数 L_{DP} 和二值化表示熵损失函数 L_{BRE} 引入生成对抗网络的无监督学习过程，将高维特征空间的海明距离关系传播到低维特征空间中并增加低维特征表示的多样性，获得更紧凑的二进制特征描述符。

综上所述，本文将改进后的生成对抗网络称为二进制生成对抗网络，基于无监督学习从二进制生成对抗网络的鉴别器 D 中提取的二进制特征向量称为无监督二进制描述符。

2　无监督视觉回环检测方法

2.1　方法总体框架

本文基于所提出的基于二进制生成对抗网络进行视觉回环检测的新方法的总体框架如图 1 所示。

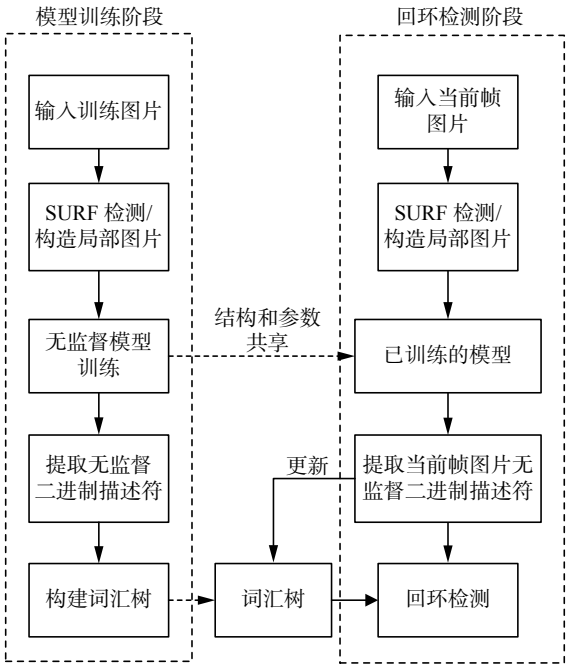

图 1　无监督视觉回环检测总体框架

Fig. 1　Overall framework of unsupervised visual loop closure detection

在模型训练阶段，首先利用 SURF 进行关键点检测并构造局部图片，基于下文所述的距离传播损失函数以及二值化表示熵损失函数交替训练鉴别器 D 及生成器 G，利用训练好的二进制生成对抗网络的鉴别器 D 提取无监督二进制描述符，并基于 BoVW 方法构建词汇树。在回环检测阶段，将实时获取的图像帧进行同样的关键点检测并构造局部图片，利用已训练好的模型提取当前帧图片的无监督二进制描述符，与现有词汇树进行比较以判断是否存在回环；当系统在大规模开放场景下运行，可以根据需要更新词汇树，以提高所述方法的适应性。

2.2 构造局部图片

本研究属于基于局部特征的回环检测方法。为获取图像的局部特征，首先将数据集中的全局图片进行分割以获取所需的局部图片。对于数据集中的每一张图片，本文利用 SURF 描述符检测关键点，将接近图片边缘的关键点丢弃后，以剩余每个关键点为中心构建尺寸为 32×32 的局部图片。图 2 为 SURF 关键点的检测和构造局部图片的示意图。下文将介绍如何利用这些局部图片对模型进行无监督训练。

图 2 局部图片的构造

Fig. 2 Local image patch construction

2.3 距离传播损失函数

为了获得低维且区分度高的无监督二进制描述符，本文在 GAN 的鉴别器中加入了距离传播损失函数 L_{DP}。该损失函数的作用在于将高维特征空间中的关系映射到低维空间，也就是说，在鉴别器 D 的高维特征空间和低维特征空间之间传播海明距离，使这两层之间具有相似的距离关系。为了达到这个目标，需要迫使鉴别器 D 的高维特征空间和低维特征空间的输出具有相似的归一化点积结果。

假设 $L(x)$ 表示鉴别器 D 中神经元个数为 K 的低维中间层，$H(x)$ 表示神经元个数为 M 的高维中间层。为了将特征空间中连续的特征向量转化为相应的二进制特征向量 $\boldsymbol{b}_L$、$\boldsymbol{b}_H$，本文使用以下激活函数[17]：

$$\mathrm{BAF}(x)=\varepsilon(s(x)-0.5) \tag{2}$$

式中：$\varepsilon(\cdot)$ 为阶跃函数，$s(x)$ 为 sigmoid 函数。利用该激活函数可将处于 [0,1] 的连续特征向量转换为二进制特征向量。

两个二进制向量之间的海明距离可以用下式进行计算：

$$d_H(\boldsymbol{b}_i,\boldsymbol{b}_j)=A-\left(\boldsymbol{b}_i^{\mathrm{T}}\boldsymbol{b}_j+(\boldsymbol{b}_i-1)^{\mathrm{T}}(\boldsymbol{b}_j-1)\right) \tag{3}$$

式中：A 是二进制特征向量的维度，因此可以用点积反映两个二进制特征向量之间的距离关系，令：

$$\mathrm{Dot}_{\boldsymbol{b}_i,\boldsymbol{b}_j}=\boldsymbol{b}_i^{\mathrm{T}}\boldsymbol{b}_j+(\boldsymbol{b}_i-1)^{\mathrm{T}}(\boldsymbol{b}_j-1) \tag{4}$$

$\mathrm{Dot}_{\boldsymbol{b}_i,\boldsymbol{b}_j}$ 越大，则二进制向量 $\boldsymbol{b}_i$、$\boldsymbol{b}_j$ 之间距离越相近，反之亦然。因此本文将提出的用于回环检测问题的距离传播损失函数定义为

$$L_{\mathrm{DP}}=\frac{1}{N(N-1)}\sum_{i,j=1,i\neq j}^{N}\left|\frac{\mathrm{Dot}_{i,j}^{H}}{M}-\frac{\mathrm{Dot}_{i,j}^{L}}{K}\right| \tag{5}$$

式中：N 是一个 batch 的大小；$\mathrm{Dot}_{i,j}^{H}$ 为高维特征空间中二进制特征表示 $\boldsymbol{b}_i$ 与 $\boldsymbol{b}_j$ 之间的点积值，同理 $\mathrm{Dot}_{i,j}^{L}$ 则表示低维特征空间二进制特征表示之间的点积值。同时，为了使高维特征空间与低维特征空间中二进制特征表示之间的海明距离具有可比性，需要对点积值进行归一化处理。

在利用深度学习进行特征提取时，为了获得好的特征表达，一般会提取高维空间的特征描述子，虽然这样得到的特征向量表现较好，但是其维度过大，会占用过多的存储空间及计算资源。通过使用距离传播损失函数 L_{DP}，可以得到低维且区分度高的二进制特征向量，就可以在好的特征表达和高效的计算效率之间求取平衡。

2.4 二值化表示熵损失函数

相比于高维特征描述子，低维特征描述子不可避免地会面临信息的损失，因此为了进一步提高低维特征空间中二进制特征表示的信息多样性，本文利用了二值化表示熵损失函数 L_{BRE}，这一损失函数在文献 [18] 中被提出，它由边缘熵 L_{ME}(marginal entropy) 及激活相关 L_{AC}(activation correlation) 两部分组成：

$$L_{\mathrm{BRE}}=L_{\mathrm{ME}}+L_{\mathrm{AC}} \tag{6}$$

L_{BRE} 通过最大化联合熵降低低维特征空间中特征向量之间的联系，以增加其多样性。利用二值化表示熵损失函数 L_{BRE} 可以提高特征描述符的区分度，从而增强鉴别器对于真实数据以及虚假数据的区分能力。如此一来，利用连接鉴别器与生成器的损失函数则可以提高生成器对于潜在样本分布的估计能力。对视觉回环检测而言，使用二值化表示熵损失函数 L_{BRE} 不仅可以使得鉴别器输出高区分度的二进制描述符提高模型在回环检测阶段的性能，而且可以加快无监督学习进程使得模型收敛更快。

2.5　网络设计

所设计的用于视觉回环检测的二进制生成对抗网络模型如图 3 所示。鉴别器 D 包含 7 个卷积层，其中卷积核大小为 3×3，通道数分别为{96, 96, 96, 128, 128, 128, 128}，stride 为{1, 1, 2, 1, 1, 2, 1}，两个 NIN(network-in-network) 结构 (神经元个数分别为 256、128) 以及一个全连接层。本文将最后一个卷积层 CONV7 作为高维特征空间，从该层提取高维特征描述子，将包含 256 个神经元的 NIN 层作为低维特征空间，提取低维特征描述子。生成器 G 包含一个全连接层及 3 个反卷积层，其中卷积核大小为 5×5，通道数分别为{256, 128, 3}。生成器的输入为维度 100 的随机噪声，输出为尺寸为 32×32 的虚假图像，并将该虚假图像作为输入与真实图像同时输入鉴别器中，而鉴别器的输出则为输入图像为真的概率。

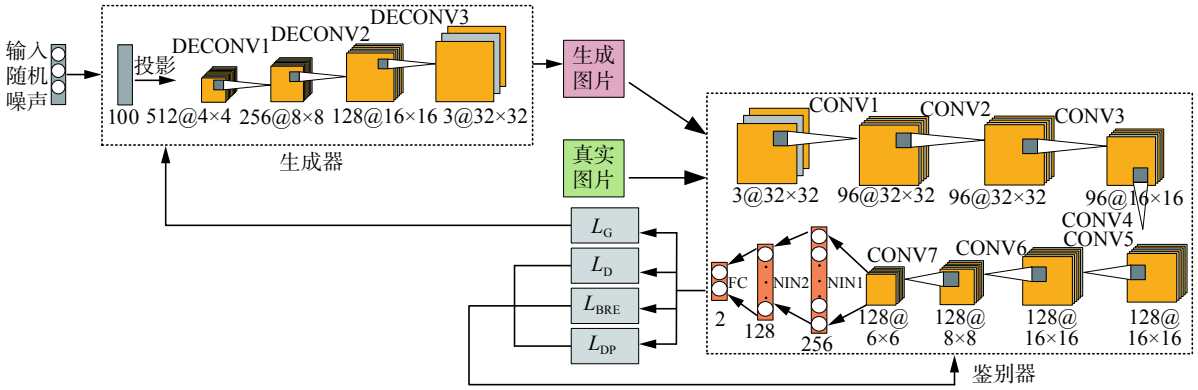

图 3　用于视觉回环检测的网络模型

Fig. 3　Network model for visual loop closure detection

2.6　模型训练

本文使用无监督的方法对模型进行训练，交替训练鉴别器 D 及生成器 G。GAN 训练的总目标函数、生成器 G 的损失函数与文献 [11] 相同。根据前文所述，鉴别器 D 训练时的损失函数可以表示为

$$L = L_{\mathrm{D}} + \lambda_{\mathrm{DP}} \cdot L_{\mathrm{DP}} + \lambda_{\mathrm{BRE}} \cdot L_{\mathrm{BRE}} \tag{7}$$

其中 L_{D} 是 Goodfellow 等[11] 给出原始损失函数，即

$$L_{\mathrm{D}} = -E_{x\sim p_{\mathrm{data}}(x)}[\log(D(x))] - E_{z\sim p_z(z)}[\log(1-D(G(z)))] \tag{8}$$

λ_{DP} 与 λ_{BRE} 为超参数，加入这两个超参数的目的在于调节距离传播损失函数以及二值化表示熵损失函数对于鉴别器目标函数的影响。在实验部分将通过改变 λ_{DP} 与 λ_{BRE} 的值验证距离传播损失函数以及二值化表示熵损失函数对整个模型性能以及训练过程的影响。

2.7　参数设置

本文所述模型和训练算法共有 8 个超参数，实验中设置的具体值如表 1 所示。所述参数值并非唯一值，可以根据具体情况进行调整以加速训练过程。图像分割后的局部图片大小为 32×32，为默认值。众所周知，GAN 的训练相对困难，λ_{DP} 与 λ_{BRE} 与特征提取能力相关，同时，合适的数值可以加快模型的训练过程，使得模型收敛速度更快，表中数值为优选值。

表 1　参数设置表

Table 1　Parameter setting

参数	大小	参数	大小
batch-size	25	局部图片长	32
epoch	100	局部图片宽	32
learning-rate	0.0003	λ_{DP}	0.5
momentum	0.5	λ_{BRE}	0.1

3　实验

3.1　实验数据集

本文选择的训练集为 Places365-Standard[19]，该数据集包含 365 个互不相关的场景类别，且无任何的标签数据。在本实验中，为了加快模型训练速度，减少训练时间，只选取了该数据集前 2 000 张图片作为训练集 (也可以增加训练样本)，并将训练集中的图片进行分割后，最终获得 140 000 张局部图片。

本文选取 3 个数据集作为测试集进行验证，分别是 NC(new college) 数据集、CC(city centre) 数据集以及 KAIST(korea advanced institute of science and technology) 数据集。NC 数据集和 CC 数据集是由英国牛津大学移动机器人小组发布的数据集[20]。其中 CC 数据集由左右两边搭载相机的移动设备沿着 2 km 的城市路段所收集，包含行人、移动的汽车等动态物体，而且视角及外观变化较

为强烈。NC 数据集同样是由左右两边搭载相机的移动设备所拍摄的，和 CC 数据集不同的是，NC 数据集的拍摄环境为校园，且含有较多的重复元素，例如墙壁等。KAIST[21] 数据集是由韩国科学技术院发布的公开数据集，该数据集是通过配备在车辆上的摄像头以及传感器于一天中不同时段在同一条街道所拍摄的。KAIST数据集中又包括 3 个子数据集：North、West、East。

以上 3 个数据集都有不同程度的视角及外观变化，具体可见表 2。对于传统手工提取特征的方法来说，强烈的视角及外观变化对回环检测是一个巨大的挑战，因此使用以上数据集可以有效验证本文所提出的方法相对于传统方法的优势，以及在大规模开放场景下的适应性。

表 2　数据集描述

Table 2　Dataset description

数据集	拍摄环境	视角变化	外观变化
NC	校园	强烈	中等
CC	市中心	强烈	强烈
KAIST	街道	中等	强烈

3.2　实验结果

作为对比，本文选取 ORB、BRIEF 和 SURF 3 个手工提取的特征描述符方法，以及基于有监督学习的 AlexNet[22]、AMOSNet 和 HybridNet[23] 深度学习方法，在 3 个测试集上进行对比。除此之外，为了验证二进制描述符相对于非二进制描述符的优势，本文还将对比二进制描述符与非二进制描述符之间的性能差异。

为了对比各类方法的性能，本文绘制了不同方法的准确率−回召率曲线，即 PR(precision-recall) 曲线[24]，并按照学术研究的常规做法，将 PR 曲线与横纵坐标围成的面积，即 AUC 作为评判标准[19]。AUC 的计算公式为

$$\mathrm{AUC}=\sum_{i=1}^{M-1}\frac{(p_i+p_{i+1})}{2}\times(r_{i+1}-r_i) \tag{9}$$

式中：M 为图片序列的数量；p_i 代表在点 i 时的准确率；而 r_i 则为回召率。AUC 越大则表明该方法的性能越好。

为了对比不同参数对于模型性能的影响，调整 λ_{DP} 与 λ_{BRE} 的数值，并计算不同数值下的各个数据集的 AUC。同时绘制了在模型中加入与不加入距离传播损失函数的情况下高维与低维空间特征海明距离之间的距离关系图。

3.2.1　不同方法的结果对比

图 4~8 绘制了各方法在 3 个测试数据集上的 PR 曲线，为方便量化对比，AUC 值列于表 3。下面将分析比较不同方法的性能和差异。

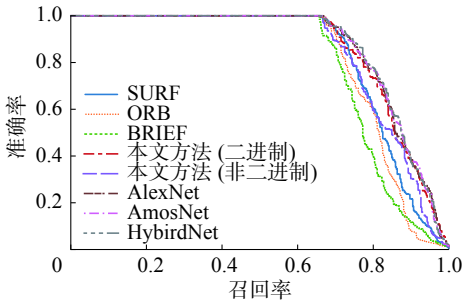

图 4　CC 数据集下各方法的 PR 曲线

Fig. 4　AUC under PR curves on the CC dataset

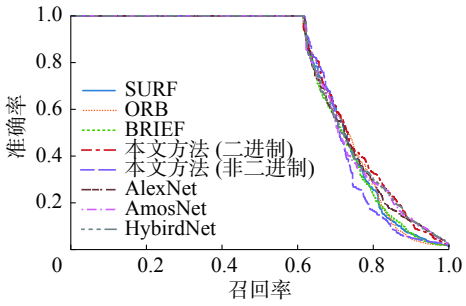

图 5　NC 数据集下各方法的 PR 曲线

Fig. 5　AUC under PR curves on the NC dataset

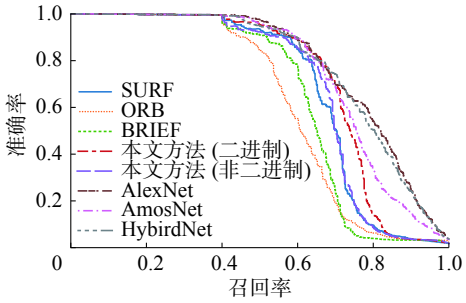

图 6　Kaist(East) 数据集下各方法的 PR 曲线

Fig. 6　AUC under PR curves on the Kaist(East) dataset

从图 4~8 及表 3 中可以得出如下结论：

1) 相比于人工提取特征的传统方法，基于深度学习的方法性能有较大的提升。无论是基于有监督学习的 AlexNet、AMOSNet 和 HybridNet，还是本文所提出的基于二进制生成对抗网络的方法都要比传统 SIFT、ORB、BRIEF 等人工特征描述符有更突出的表现，主要原因在于深度学习的方法可以在复杂的环境下自动且精准地提取图像的深层特征。

2) 相比于有监督方法，本文所提出的无监督回环检测方法在性能上略有下降，AlexNet 和 HybridNet 相对最优，本文所述方法与 AMOSNet 性能相近。由于有监督学习方法利用了大量的有标签数据，可以通过已知的训练样本训练出最优模

型，因此性能表现更为出色。但是，有监督学习方法需要大量标签数据且训练时间更长。而无监督学习方法由于不需要标签，则更适用于大规模场景、复杂场景和开放场景下的回环检测问题。除此之外，AlexNet、HybridNet 及 AMOSNet 在训练时都需要大量有标签数据，其中，Krizhevsky 等[21] 在训练 AlexNet 时采用 120 万张图片作为训练集，AMOSNet 和 HybridNet 在训练时的数据集更是包含了 250 万张图片[23]，而本文所述方法仅仅需要 2 000 个无标签数据对模型进行训练即可获得较为出色的结果。而且值得注意的是，在 NC 数据集上无监督回环检测的表现甚至优于有监督方法。NC 的拍摄环境为校园，且含有较多重复元素和强烈的视角变化。这证明了本文所述方法在复杂场景下，特别是强烈的视角变化具有鲁棒性。所以综上所述，本文的方法与有监督方法之间的性能差异是完全可以接受的。

3) 对比二进制特征描述符和非二进制特征描述符，可以发现，在无监督回环检测框架下，在本文所提出的 3 个测试集上二进制特征描述符的性能更优。本文利用距离传播损失函数使得高维特征空间与低维特征空间之间具有相似的海明距离关系，利用二值化表示熵损失函数能进一步增强低维二进制特征描述子的表征能力，弥补其信息损失，提高其可靠性。在性能接近的情况下，二进制特征描述符对于回环检测应用非常有吸引力，因为使用二进制特征描述符可以节省更多的存储空间以及计算资源，加快回环检测速度[25]。

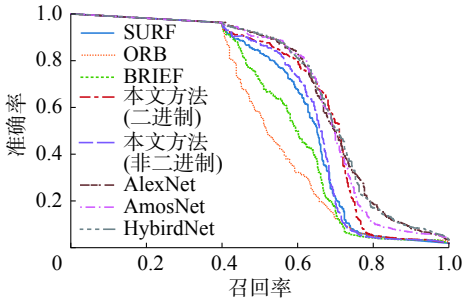

图 7　Kaist(North) 数据集下各方法的 PR 曲线

Fig. 7　AUC under PR curves on the Kaist(North) dataset

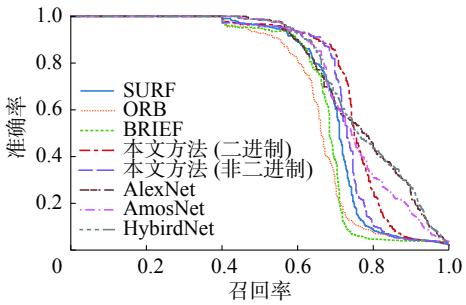

图 8　Kaist(West) 数据集下各方法的 PR 曲线

Fig. 8　AUC under PR curves on the Kaist(West) dataset

表 3　AUC 汇总表

Table 3　AUC summary

数据集	SURF	ORB	BRIEF	Ous(binary)	Ous(nonbinary)	AlexNet	AmosNet	HybridNet
CC	0.827	0.806	0.786	0.858	0.834	0.864	0.865	**0.867**
NC	0.734	0.735	0.730	**0.752**	0.724	0.742	0.742	0.745
KAIAT(East)	0.677	0.611	0.639	0.719	0.686	**0.787**	0.749	0.778
KAIAT(North)	0.610	0.536	0.574	0.656	0.622	0.682	0.671	**0.684**
KAIAT(West)	0.702	0.662	0.675	0.750	0.726	0.771	0.755	**0.778**

3.2.2　不同参数的结果对比

为了进一步研究距离传播损失函数 L_{DP} 和二值化表示熵损失函数 L_{BRE} 对无监督回环检测性能的影响，本文改变参数 λ_{DP} 以及 λ_{BRE} 的值，并计算了不同参数值在各个数据集下相对应的 AUC，结果如表 4 所示。从表中可以看出，只有在同时加入距离传播损失函数 L_{DP} 和二值化表示熵损失函数 L_{BRE} 后，视觉回环检测的性能才会有实质的提升。因此，在无监督回环检测中，距离传播损失函数 L_{DP} 和二值化表示熵损失函数 L_{BRE} 缺一不可，前者实现高维特征到低维特征的映射，获得维度更低，更为紧凑且区分度高的二进制特征描述子，后者在熵损失最小的情况下进一步提高低维二进制描述符的多样性和表征能力。在本实验中，优选的参数是 λ_{DP}=0.5，λ_{BRE}=0.1。

除此之外，为了验证距离传播损失函数的有效性，测试其是否将高维空间特征的距离关系映射至低维空间，我们以 KAIST(North) 数据集为例，分别提取其在 λ_{DP}=0.5，λ_{BRE}=0.1 和 λ_{DP}=0，λ_{BRE}=0.1 时高维空间特征以及低维空间特征，对不同维度的特征进行归一化操作后，利用式 (4) 计算相同参数下高维空间与低维空间相对应特征之间的相似性。

实验结果如图 9 所示。

表 4　不同参数下的 AUC

Table 4　AUC under different parameters

参数	CC	NC	KAIST(East)	KAIST(North)	KAIST(West)
λ_{DP}=0.5, λ_{BRE}=0.1	0.858	0.752	0.719	0.656	0.750
λ_{DP}=0, λ_{BRE}=0.1	0.752	0.690	0.522	0.511	0.574
λ_{DP}=0.5, λ_{BRE}=0	0.768	0.692	0.512	0.493	0.541

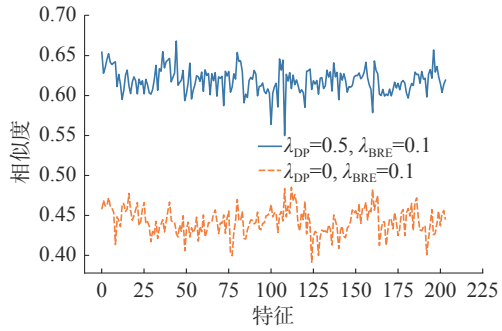

图 9　高维空间与低维空间距离关系图

Fig. 9　Distance diagram between two layers

图 9 中横坐标表示 KAIST(North) 数据集中图片的特征描述子，纵坐标则为不同维度特征之间的相似性。从图中可以清楚地看出，在两组不同参数下，高维特征空间与低维特征空间之间距离关系的相似性具有明显的差异。在 λ_{DP}=0，λ_{BRE}=0.1 时，高维特征空间与低维特征空间的距离关系相似性位于 0.39~0.49，而当 λ_{DP}=0.5，λ_{BRE}=0.1 时，其相似性则位于 0.55~0.67。由此可得，距离传播损失函数的加入有助于将高维特征空间的海明距离关系映射到低维空间，获得更加紧凑，区分度更高的特征。

3.2.3　可视化分析

在这部分，以 NC 数据集为例，通过可视化的方式来证明基于无监督二进制描述符的视觉回环检测方法的有效性。图 10 为根据图片的已有标签绘制的真实回环图，若第 i 帧图片与第 j 帧图片形成回环，则在图中对应坐标为 (i,j) 的点为白色。所以真实回环图根据对角线完全对称。图 11~13 为 ORB、BRIEF、SURF 以及本文所述方法给出的回环检测图，用相似度矩阵来表示，其中坐标为 (i,j) 的点表示第 i 帧图片与第 j 帧图片之间的相似度，坐标点的颜色根据对应帧之间的相似度的变化而变化，颜色越亮则相似度越高，两帧图片之间的相似度越高则二者成为回环的几率越大。

图 10　NC 数据集的真实回环图

Fig. 10　The ground truth of NC dataset

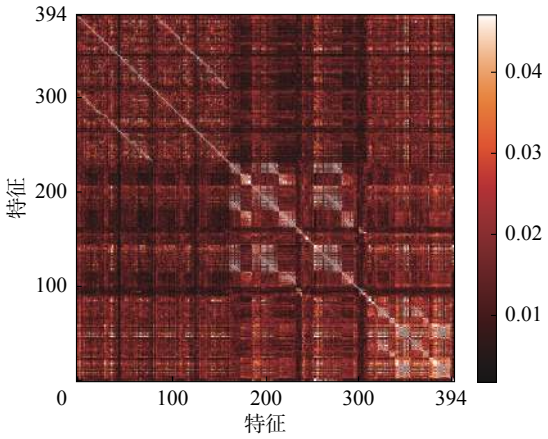

图 11　基于 BRIEF 的相似度矩阵

Fig. 11　Similarity matrix of BRIEF

通过对比真实回环与不同方法检测出的回环，不难发现，不论是传统的 ORB、BRIEF 以及 SURF 还是本文所述方法都可以检测出较为明显的回环，不同的是传统方法在面对不易检测的回环时会出现遗漏的情况，因此相比于图 11~13，图 14 会出现更多的明亮点以及色块，明暗对比较为明显，这充分说明本文所述方法会为回环检测提供更多的相似帧，减少遗漏情况的出现。因此在面对较强的视角及外观变化时本文所述方法可以检测出更多的回环，效果更加突出，这表明无监督二进制描述符更有区分度。

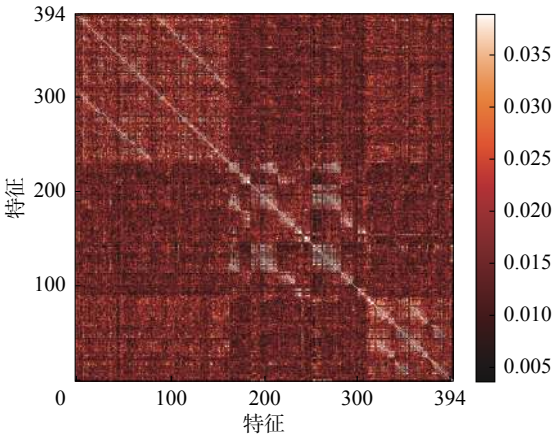

图 12　基于 ORB 的相似度矩阵
Fig. 12　Similarity matrix of ORB

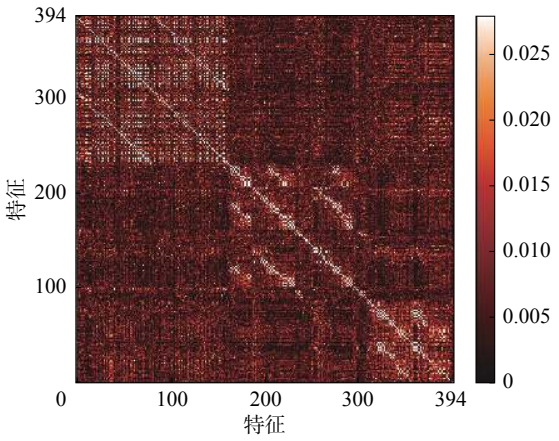

图 13　基于 SURF 的相似度矩阵
Fig. 13　Similarity matrix of SURF

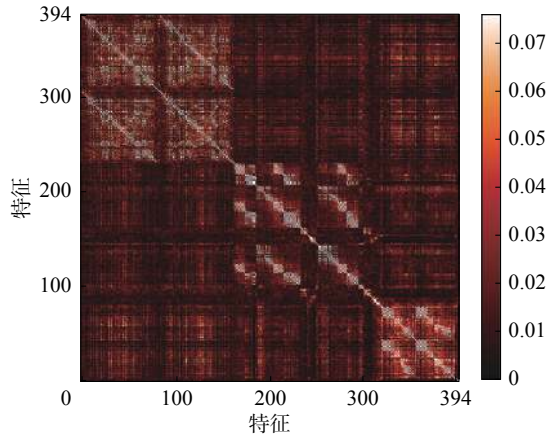

图 14　基于无监督二进制描述符的相似度矩阵
Fig. 14　Similarity matrix of unsupervised binary descriptor

4　结束语

针对现有的视觉回环检测方法大多依赖有监督学习且特征向量维度较高，占用较大存储空间的问题，本文受生成对抗网络的启发，提出了一种无监督二进制描述符，并将其与 BoVW 结合用于视觉回环检测。该方法在模型训练时采用无监督学习方式，训练集为互不相关的场景图片且无任何标签数据。为了获得高区分度及低维度的无监督二进制描述符，利用距离传播损失函数将高维特征空间中的关系映射到低维空间，并且利用二值化表示熵损失函数提高低维空间二进制特征表示的多样性，进一步改善低维特征所带来的信息损失问题。在 NC 数据集、CC 数据集以及 KAIST 数据集上对本文所提出的无监督二进制描述符的有效性进行了验证，并和 ORB、BRIEF、SURF 这 3 种人工特征描述符，以及 AlexNet、AMOSNet 和 HybridNet 3 种深度学习方法进行了比较。结果表明，无监督二进制描述符在具有强烈视角及外观变化的复杂场景下具有鲁棒性，性能可以与有监督深度网络媲美。但无监督方法从根本上避免了费时费力的有监督数据标注过程，同时极大地节约了存储空间和计算资源，加快回环检测的进程，在大规模开放场景的视觉 SLAM 中具有较大价值。

参考文献：

[1] KONOLIGE K, AGRAWAL M. FrameSLAM: from bundle adjustment to real-time visual mapping[J]. IEEE transactions on robotics, 2008, 24(5): 1066–1077.

[2] 张毅, 沙建松. 基于图优化的移动机器人视觉 SLAM[J]. 智能系统学报, 2018, 13(2): 290–295.
ZHANG Yi, SHA Jiansong. Visual-SLAM for mobile robot based on graph optimization[J]. CAAI transactions on intelligent systems, 2018, 13(2): 290–295.

[3] HO K L, NEWMAN P. Detecting loop closure with scene sequences[J]. International journal of computer vision, 2007, 74(3): 261–286.

[4] 刘强, 段富海, 桑勇, 等. 复杂环境下视觉 SLAM 闭环检测方法综述 [J]. 机器人, 2019, 41(1): 112–123, 136.
LIU Qiang, DUAN Fuhai, SANG Yong, et al. A survey of loop-closure detection method of visual SLAM in complex environments[J]. Robot, 2019, 41(1): 112–123, 136.

[5] LOWE D G. Object recognition from local scale-invariant features[C]//Proceedings of the 17th IEEE International Conference on Computer Vision. Kerkyra, Greece, 2002: 1150−1157.

[6] BAY H, TUYTELAARS T, VAN GOOL L. SURF: speeded up robust features[C]//Computer vision-ECCV 2006. Graz, Austria, 2006: 404−417.

[7] SIVIC J, ZISSERMAN A. Video Google: A text retrieval approach to object matching in videos[C]//Proceedings of the 9th IEEE International Conference on Computer Vision. Nice, France, 2003: 1470−1477.

[8] CHEN Zetao, LAM O, JACOBSON A, et al. Convolutional neural network-based place recognition[C]//Australasi-

an Conference on Robotics and Automati-on. Melbourne, Australasian, 2014: 8−14.

[9] SÜNDERHAUF N, SHIRAZI S, DAYOUB F, et al. On the performance of ConvNet features for place recognition[C]//Proceedings of 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems. Hamburg, Germany, 2015: 4297−4304.

[10] GAO Xiang, ZHANG Tao. Unsupervised learning to detect loops using deep neural networks for visual SLAM system[J]. Autonomous robots, 2017, 41(1): 1–18.

[11] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal, Canada, 2014: 2672−2680.

[12] HINTON G E, ZEME R S. Autoencoders, minimum description length and Helmholtz free energy[C]//Proceedings of the 6th International Conference on Neural Information Processing Systems. Denver, Colorado, USA, 1993: 3−10.

[13] SMOLENSKY P. Information processing in dynamical systems: foundations of harmony theory[M]//RUMELHART D E, MCCLELLAND J L. Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Cambridge: MIT Press, 1986.

[14] SHIN D W, HO Y S, KIM E S. Loop closure detection in simultaneous localization and mapping using descriptor from generative adversarial network[J]. Journal of electronic imaging, 2019, 28(1): 013014.

[15] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks[C]//Proceedings of the 4th International Conference on Learning Representations. San Juan, Puerto Rico, 2016: 97−108.

[16] SALIMANS T, GOODFELLOW I, ZAREMBA W, et al. Improved techniques for training GANs[C]//Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain, 2016: 2234−2242.

[17] DONG Haowen, YANG Y H. Training generative adversarial networks with binary neurons by end-to-end backpropagation[EB/OL]. (2018-12-12) [2020-01-01] https://arxiv.org/abs/1810.04714.

[18] CAO Yanshuai, DING G W, LUI K Y C, et al. Improving GAN training via binarized representation entropy (BRE) regularization[C]//Proceedings of the 6th International Conference on Learning Representations. Vancouver, Canada, 2018: 1−22.

[19] ZHOU Bolei, LAPEDRIZA A, KHOSLA A, et al. Places: A 10 million image database for scene recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2018, 40(6): 1452–1464.

[20] CUMMINS M, NEWMAN P. FAB-MAP: Probabilistic localization and mapping in the space of appearance[J]. The international journal of robotics research, 2008, 27(6): 647–665.

[21] CHOI Y, KIM N, PARK K, et al. All-day visual place recognition: benchmark dataset and baseline[C]//Proceedings of 2015 IEEE International Conference on Computer Vision and Pattern Recognition Workshops. Boston, USA, 2015: 8−13.

[22] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]//Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, USA, 2012: 1097−1105.

[23] CHEN Zetao, JACOBSON A, SÜNDERHAUF N, et al. Deep learning features at scale for visual place recognition[C]//Proceedings of 2017 IEEE International Conference on Robotics and Automation. Singapore, 2017: 3223−3230.

[24] ZAFFAR M, KHALIQ A, EHSAN S, et al. Levelling the playing field: A comprehensive comparison of visual place recognition approaches under changing conditions[EB/OL]. (2019-04-29) [2020-02-01] https://arxiv.org/abs/1903.09107?context=cs.CV.

[25] MEMON A R, WANG Hesheng, HUSSAIN A. Loop closure detection using supervised and unsupervised deep neural networks for monocular SLAM systems[J]. Robotics and autonomous systems, 2020, 126: 103470.

作者简介：

杨慧，硕士研究生，主要研究方向为视觉回环检测。

陈良，副教授，主要研究方向为基于深度学习的人工智能系统、新一代智能控制理论及应用。

孙立宁，教授，博士生导师，主要研究方向为先进机器人技术。主持“863”计划、973 计划、国家重大专项、国家自然科学基金等 20 多项。获国家技术发明/科技进步二等奖 2 项、教育部技术发明奖二等奖 1 项、省级技术发明/科技进步一等奖 3 项，二等奖 2 项。发表学术论文 400 多篇，获授权国家发明专利 40 余项。