

DOI:10.16136/j.joel.2022.04.0452

基于二值神经网络的大场景点云分类

章国道¹, 刘儒瑜^{2*}, 张志勇¹, 孔德伟¹, 邱飞岳³

(1. 浙江工业大学 计算机科学与技术学院, 浙江 杭州 310023; 2. 杭州师范大学 信息科学与技术学院, 浙江 杭州 311121; 3. 浙江工业大学 教育科学与技术学院, 浙江 杭州 310023)

摘要:近年随着3维数据采集技术不断发展,大场景点云数据的获取越来越方便。目前深度学习网络框架在2维图像处理领域越来越成熟,而大场景点云是一种3维无规则化的数据,3维卷积神经网络直接处理大场景3维数据会存在分类精度低和计算复杂等问题。因此为了有效解决基于深度学习的点云分类任务中存在的计算时间长和分类精度低的问题,本文提出基于二值神经网络的大场景点云分类方法,针对不规则的3维点云数据设计特征值计算方法,基于IR-Net二值神经网络处理输入的点云特征图像,进一步采用Dynamic ReLU激活函数,提高神经网络的计算效率,最后得出点云分类结果。实验结果表明,所提出的方法在Oakland数据集上分类精度达到97.6%,在GML数据集中取得了92.3%和97.2%的分类精度,实验结果证明Dy-ResNet能够有效提升了点云分类的精度,减少计算的复杂度,并提高了训练效率。

关键词:特征图; 特征融合; 二值神经网络; 大场景点云分类

中图分类号: TN24 文献标识码: A 文章编号: 1005-0086(2022)04-0364-09

Point cloud classification of large-scale scene based on binary neural network

ZHANG Guodao¹, LIU Ruyi^{2*}, ZHANG Zhiyong¹, KONG Dewei¹, QIU Feiyue³

(1. College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, Zhejiang 310023, China; 2. School of Information Science and Technology, Hangzhou Normal University, Hangzhou, Zhejiang 311121, China; 3. College of Education, Zhejiang University of Technology, Hangzhou, Zhejiang 310023, China)

Abstract: With the continuous development of three-dimensional (3D) data acquisition technology in recent years, the acquisition of three-dimensional large-scale scene point cloud data is becoming more and more convenient. At present, the deep learning network framework is becoming more and more mature in the field of two-dimensional (2D) image processing, while the large-scale scene point cloud is a kind of 3D irregular data. When using 3D convolutional neural networks in deep learning to directly process large-scale scene 3D data, there will be problems such as low classification accuracy and high computational complexity. Therefore, to effectively solve the problems of long computation time and low classification accuracy in point cloud classification based on deep learning, this paper proposes a binary neural network-based classification method for large-scale scene point cloud, designing the eigenvalue calculation method for irregular 3D point cloud data, processing the input point cloud feature images based on IR-Net binary neural network, further adopting Dynamic ReLU activation function to improve the computational efficiency of the neural network, and finally deriving the point cloud classification results. The experimental results show that the proposed method achieves 97.6% classification accuracy on the Oakland dataset and 92.3% and 97.2% in the GML dataset. Experimental results show that Dy-ResNet can effectively improve the accuracy of point cloud classification, reduce the complexity of calculation and im-

* E-mail: lry@hznu.edu.cn

收稿日期: 2021-06-22 修订日期: 2021-07-30

基金项目: 国家自然科学基金(71872131)、浙江省科技计划项目-重点研发(2018C01080)和浙江省自然科学基金探索青年(LQ22F030004)资助项目

prove training efficiency.

Key words: feature map; feature fusion; binary neural network; large-scale scene point cloud classification

1 引言

点云分类是3维视觉和机器学习等领域重要的研究课题。首先,点云数据的获取越来越方便,机载激光扫描(airborne laser scanning, ALS)和激光雷达(light detection and ranging, LiDAR)等设备可以获取大场景的空间距离信息生成3维点云。另一方面,基于大规模的室外场景点云数据,越来越多的应用,如智能驾驶、机器人导航、场景3维建模等迫切需要点云分类的结果,因此复杂大场景的点云分类工作是一项重要的任务。此外,本文还考虑了在一个小的局部邻域中对不同类型的点云对象进行准确分类,这显然是一项更具挑战性的工作。

随着深度学习的迅速发展,基于神经网络的3维点云分类方法效果显著,PointNet^[1]、PointNet++^[2]、RSNet^[3]、动态图卷积神经网络^[4]和点卷积神经网络^[5]集中于探索局部上下文和分层学习架构,在3维点云数据分类任务中都取得不错的效果。LIU等^[6]将空间点的3维邻域特征转化为2维图像,作为多尺度卷积神经网络的输入进行训练。现有的一些大场景点云分类方法通常将无规则的3维点云进行多视图转换或通过立体体素化转换成一种规则点云结构,但是在这些方法中,基于多视图的方法需要经过多个不同角度投影点云数据形成2维图像,导致点云的3维空间信息。基于体素化的方法^[7]将3维点云表示为一个规则的体素小方块集合,这就导致了这类方法的算法性能很大程度上取决于体素大小的划分,体素较大容易就会造成不能很好的表达点云的特征信息,体素量较小会造成计算量的增多,这种方法确实可以提高点云分类的性能,但由于网络参数计算量大,体素化网格分辨率低,需要占用大量内存,同时又伴随着局部信息丢失等问题,因此不适合大规模复杂场景下3维点云分类任务。

因此为了确保点云3维信息的完整性,本文提出一种基于二值神经网络的大场景点云分类方法。将3维点云投影到 x 、 y 、 z 3个不同的平面计算其特征值,再将得到的特征值进行组合计算形成2维点云特征图像来融合多个不同视角的点云特征。同时为了减少网络参数计算复杂度,本文将二值神经网络引入到点云分类任务中,实验证明二值神经网络可以提高点云处理的效率。在这

2方面的结合下,本方法在3维点云处理任务上取得很好的效果,具有一定的研究意义。

2 相关工作

目前大部分点云分类工作都集中如何修改点云的网络框架,一般是通过加深网络框架,或者采用一些比较有效的局部特征提取方法来提取更多有效的点云特征,例如基于局部邻域的点云特征计算^[8],以体素、块和柱的形式^[9]或以空间块、平面段和局部区域的形式^[10],特征本身通常包括来自3维(three-dimensional, 3D)结构张量的几何特征^[11,12]。此外,高度和局部平面特征^[13]、局部邻域内点分布的2维(two-dimensional, 2D)和3D几何特性^[14]、高度特征^[8]等也已经在分类任务中大量地使用。

传统神经网络通常通过不断增加网络层的数量达到提升精度的目的,但是也会带来网络模型参数多,计算复杂度高的问题。网络层越深的深度卷积神经网络模型就需要越多的存储空间和昂贵的计算资源,同时网络层越深也会导致一些梯度爆炸和过拟合的现象。近年来,许多学者研究二值神经网络在图像分类方面的工作,其在降低模型的存储和计算量等方面具有很大的优势,与全精度卷积神经网络相比,二值神经网络的图像分类准确率比较低。二值神经网络是通过+1与-1两个值来进行位运算完成信息的传递,这样可以避免传统卷积乘除运算,节省大量的计算。然而也是因为只进行加法与位运算,所以会导致网络传递中大量的有效信息丢失的问题。

二值神经网络前向传播的激活值和权值被局限在+1和-1两个值,导致整体网络精度的下降。但是目前许多学者提出了多种改进方法,其中包括2种主流的方法用于增加神经网络多样性,分别是通过增加神经元的数量和增加特征的多样性。Bi-RealNet^[15]网络在激活层中额外增加一个网络层达到改进网络层特征多样性的目的同时提升了网络的分类精度,然而随着网络不断加深,网络层浮点加减法位运算不断地增加,Bi-RealNet的效果比普通二进制神经网络要更差。MARTINEZ等^[16]通过模型蒸馏和量化2种策略,进一步挖掘二值神经网络模型的学习能力,并基于该方法搭建ResNet18网络结构,在著名的ImageNet图像分类数据集上达到了65.4%的分类精度,只是牺牲了1%计算量,在准确率方面比传统全精度卷积网络提高了3.9%。RAS-

TEGAR 等^[17]通过使用新的位运算量进一步将二值神经网络框架的输入二值化,取得了与卷积神经网络不相上下的效果。

3 基于二值卷积神经网络的大场景点云分类方法

传统神经网络为了保证高准确率往往会不断地增加卷积的网络层深度,因此对硬件的内存计算空间要求比较高并且存在计算复杂度高的局限性,不利于对实时性有要求的任务。二值化神经网络具有自身模型压缩率高和计算速度快等优点,成为了近些年的研究热点。其中,二值神经网络 IR-Net^[18]从全新的角度思考二值神经网络的前向传播和反向传播过程,解决神经网络层中的信息流的传递和保留问题。IR-Net 按照传统二值神经网络的设计将每一

个网络层的权值和隐藏层的激活值二值化为+1 或者-1,使得整个网络模型占用更少的运算空间。IR-Net 基于传统二值神经网络,提出 Libra-PB 参数二值化方法和误差衰减估计器(error decay estimator, EDE)梯度计算方法,能够保留更多的神经网络传递的特征信息,并且很大程度上降低运算时间,提高运行效率。因此本文基于 IR-Net 进一步提出适用于点云分类的网络框架,包括整个算法框架的搭建、点云特征提取模块和网络模块的设计。二值神经网络大场景点云分类网络框架如图 1 所示。针对点云分类任务,本文的方法通过将原始的 3 维点云和在垂直轴的点云进行平面投影,进行手动特征值计算点云的特征值,然后将包含点云特征信息的特征值进行组合形成点云的 2 维特征矩阵,将 3 维点云的计算降低到 2 维的计算,能够有效的降低计算量。为了进一步减少点云分类任务的网络参数和计算量,采用 IR-Net 的二值神经网络和 Dynamic Re-LU 激活函

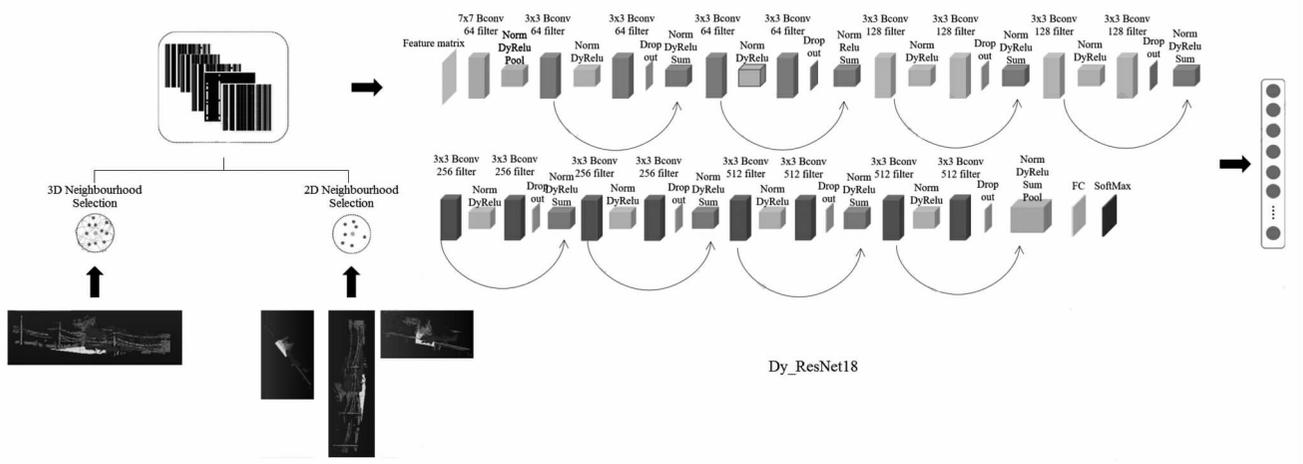


图 1 二值神经网络大场景点云分类框架

Fig. 1 binary neural network framework for large-scale scene point cloud classification

数结合的方法,进一步提高网络的计算效率。

3.1 点云特征提取模块

传统的点云处理方法,是直接将 3 维点云输入到神经网络中进行处理,处理这些 3 维的点云信息需要消耗巨大的计算和存储资源,这对硬件配置和性能具有一定的要求。如果可以将 3 维的点云信息进行降维到 2 维,在通过神经网络提取特征计算,这样可以节省很大的计算量和运行时间,所以本文将点云进行手动的特征计算,得出多个能够体现点云中各个物体特征的特征值。传统的方法都是通过支持向量机(support vector machine, SVM)、ada-boost、random forest 和 decision tree 等方法进行处理,然而这些方法在处理过程中会丢失较多的信息。

针对此方面,本文将得到的点云特征值进行 2 维矩阵组合,组合成一个 2 维矩阵。因为这些特征值能够反应点云的特征信息,所以将这些 2 维矩阵看作一系列黑白点云特征图像,然后将这些黑白点云特征图像输入到卷积神经网络框架中进一步处理,得到更多的特征信息,从而完成对 3 维点云的特征提取。

3.2 二值卷积神经网络

二值神经网络具有内存消耗少和计算速度快等特点。在图像分类任务中,二值神经网络可以从损失函数的设计和训练策略的方法来提高整体网络的训练能力。

本文采用的二值神经网络是将权重和激活值都

进行二值化操作,用1位数表示浮点权值或激活值。二值神经网络计算过程可以表示为:

$$Q_x(x) = \alpha B_x, \quad (1)$$

式中, α 表示二进制值的标量,包括权重的 α_w 和激活值的 α_a , $B_x \in \{-1, +1\}^n$ 表示二进制数值,包括二值化权重 B_w 和激活 B_a 。符号函数 B_x 可以表示为:

$$B_x = \text{sign}(x) = \begin{cases} +1, & \text{if } x \geq 0 \\ -1, & \text{otherwise} \end{cases}, \quad (2)$$

二值化的权值和激活值用前向传播中的向量乘法表示为:

$$z = Q_w(w)^\top Q_a(a) = \alpha_w \alpha_a (B_w \diamond B_a), \quad (3)$$

式中, \diamond 表示按位操作向量的内积。

二值神经网络的反向传播过程中,由于 sign 函数的梯度连续不可导,因此采用直通估计器(straight through estimator, STE)的方法来对梯度进行近似化操作,但是由于这两个之间存在着明显的梯度不匹配,很容易导致反向传播存在错误信息,使传播的值出现偏差,因此在后续方法中引入了 Libra-PB 方法,可以传播更多有效的信息。

3.3 基于参数二值化(Libra-PB)的信息保留网络

传统的二值神经网络直接将激活函数和权重量化为+1和-1,这种方法虽然简单,同时也会丢失传递过程中的特征和权重信息,极大降低网络的表达能力,因此 QIN 等^[18]采用基于参数二值化(Libra-PB)的信息保留网络,它可以在二值网络传播过程中尽可能地保留有效信息。

在正向传播过程中,量化运算会造成信息部分损失,因此使用最小量化误差来寻找二值神经网络模型的最小误差值,计算式如下所示:

$$\min_j(Q_x(x)) = \|x - Q_x(x)\|^2, \quad (4)$$

式中, x 为全精度参数, $Q_x(x)$ 为量化参数, $j(Q_x(x))$ 为全精度参数与二值化参数之间的误差。

二值神经网络的参数值被限制在+1和-1两个值内,这使得网络层中神经元所携带的信息很容易丢失。只经过减少参数之间的误差来训练出一个有效的二值神经网络模型具有一定难度。Libra-PB 方法^[10]可以保证信息在前向传播过程中损失最小化,并综合考虑在前向传播和反向传播过程中的量化误差和信息损失。随机变量 $b \in \{-1, +1\}$ 服从 Bernoulli 分布,其概率质量函数表示为:

$$f(b) = \begin{cases} p, & \text{if } b = +1 \\ 1 - p, & \text{if } b = -1 \end{cases}, \quad (5)$$

式中, p 表示为取值+1的概率, $p \in (0, 1)$,熵 $Q_x(x)$ 可以通过式(6)计算得出:

$$H(Q_x(x)) = H(B_x) = -p \ln(p) - (1-p) \ln(1-p), \quad (6)$$

式中, p 取值为0.5时,信息熵 $Q_x(x)$ 将取到最大值。

为了使网络模型更加稳定,不会因为权重大小和梯度变化而产生影响,可以进一步将权重进行归一化处理,标准化权重 \hat{w}_{std} 可表示为:

$$\hat{w}_{\text{std}} = \frac{\hat{w}}{\sigma(\hat{w})}, \hat{w} = w - \bar{w}, \quad (7)$$

式中, \hat{w} 为权重, $\sigma(\cdot)$ 为标准差。 \hat{w}_{std} 有2个作用:1)使得二值权值的信息熵最大化;2)二值化权重值加。使用标准化的权重表现得更加稳定,二值化权重 $Q_w(\hat{w}_{\text{std}})$ 在训练过程中维持很好的鲁棒性。

针对二值化方法和量化误差中存在着某些计算代价比较高的浮点计算,Libra-PB 进一步引入整数位移标量来提高权值的性能。最优的位移标量计算式表示为:

$$s^* = \text{round}(\log_2(\|\hat{w}_{\text{std}}\|_1/n)). \quad (8)$$

在二值神经网络中,对权重进行平衡和标准化能够实现有效信息的传递和保留,通过引入有效的损失函数可以在不增加网络运行时间的同时进一步提升网络的整体效率。例如 Libra-PB 方法可以直接插入到现有的一些神经网络模型中,降低网络的计算成本,并提高效率。

3.4 Dynamic ReLU 激活函数

ReLU 激活函数是最常用的激活函数之一,因为能够提高神经网络的整体性能,所以被广泛地应用于各种神经网络结构中。传统的 ReLU、LeakyReLU、PReLU 等静态激活函数不会根据输入的变化而变化,因此具有一定的局限性,Dynamic ReLU^[19]是一种新的动态激活函数,它可以根据输入参数的变化改变自身的激活函数,在不造成网络深度或宽度增加的情况下,能有效提高网络模型性能。Dynamic ReLU 函数可以很方便地嵌入到神经网络框架中,提高网络的计算效率,得出更好的结果。

传统的静态 ReLU 可以表示为:

$$y = \max\{x, 0\}, \quad (9)$$

式中, x 是输入向量。对于输入的 x_c 的 c th通道,激活值 $y_c = \max\{x_c, 0\}$ 。

Dynamic ReLU 激活函数表示为:

$$y_c = f_{\theta(x)}(x_c) = \max_{1 \leq k \leq K} \{a_c^k(x) x_c + b_c^k(x)\}, \quad (10)$$

式中,以 (a_c^k, b_c^k) 为系数的超函数 $\theta(x)$ 的输出表示为: $[a_1^1, \dots, a_1^K, \dots, a_c^1, \dots, a_c^K, b_1^1, \dots, b_1^K, \dots, b_c^1, \dots, b_c^K]^\top = \theta(x)$,

式中,函数数量用 K 来表示,通道数量记为 C 。激活参数 (a_c^k, b_c^k) 一方面与对应的输入元素 x_c 有关,另一方面还与其他输入元素 $x_{j \neq c}$ 有关。

Dynamic ReLU 在处理维数为 $C \times H \times W$ 的输入张量 x 时,使用池化层对输入特征空间维度信息进行压缩。然后输入到两个全连接层(中间包含一个 ReLU 层),最后输出 $2KC$ 个元素,可以用 $a_1^1:K$ 和 $b_1^1:K$ 来表示,记为 $\Delta a_1^1:K$ 和 $\Delta b_1^1:K$ 。通过 $2\sigma(x) - 1$ 函数来标准化 $+1$ 到 -1 之间的误差范围,其中 sigmoid 函数记为 $\sigma(x)$,其计算表达式为:

$$a_c^k(x) = \alpha^k + \lambda_a \Delta a_c^k(x), b_c^k = \beta^k + \lambda_b \Delta b_c^k(x), \quad (12)$$

式中, α^k 和 β^k 分别是 a_c^k 和 b_c^k 的初始值。 λ_a 和 λ_b 表示误差范围取值。

从上面的方法可以看出,Dynamic ReLU 激活函数可以随着输入参数的变化而动态的改变激活函数 (K 值),能在一定程度上提高整体网络框架效率,因此本文将其与二值神经网络结合,进一步提高点云特征图像的分类准确率。

3.5 大场景点云分类网络总体框架设计

基于二值神经网络的大场景点云分类框架具体的设计如下:首先是从 3 维和 2 维 2 个不同维度上对

大场景点云进行特征提取,得到点云特征图像,再将大小为 [BatchSize, Channel, Feature_Width, Feature_Height] 的特征图像输入二值神经网络框架。其设计表如表 1 所示, DyResNet18 为 IR-Net 中采用 Dynamic ReLU 的二值神经网络, ResNet18_1w1a、ResNet20_1w1a、ResNet20_1w32a 分别为 IR-Net 中三个对比网络,其中 1w1a 指的是当正向传播 1 比特权重,反馈激活输出是 1 比特。1w32a 指的是当正向传播 1 比特权重,反馈激活输出是 32 比特。DyResNet18 总共有四个 block,每一个 block 的大小为 [2, 2, 2, 2], 在 block_1 中 Conv 参数为 3×3 的卷积核和通道大小为 64 的卷积层, block_2 的 Conv 参数为 3×3 和通道为 128 的卷积层, block_3 中 Conv 的参数为 3×3 的卷积核和通道为 256 的卷积层, block_4 的 Conv 参数为 3×3 和通道为 512 的卷积层,通过这些卷积层可以提取点云特征图像的特征信息,得到一个 [常量, K] 的一维向量,得出分数进而输出分类结果,所提出方法采用 Dynamic ReLU 激活函数,进一步缩短了网络的训练时间。

本文中构建的二值卷积神经网络的结构参数如表 1 所示。

表 1 二值神经网络框架设计表

Tab. 1 Binary neural network framework design table

layer_name	Dy-ResNet18_1w1a	ResNet18_1w1a	ResNet20_1w1a	ResNet20_1w32a
conv1_x			3×3, 64, stride 1	
conv2_x	[1w1a_Conv, 3×3, 64 1w1a_Conv, 3×3, 64]×2	[1w1a_Conv, 3×3, 64]×2	[1w1a_Conv, 3×3, 16]×3	[1w32a_Conv, 3×3, 16]×3
conv3_x	[1w1a_Conv, 3×3, 128 1w1a_Conv, 3×3, 128]×2	[1w1a_Conv, 3×3, 128]×2	[1w1a_Conv, 3×3, 32]×3	[1w32a_Conv, 3×3, 32]×3
conv4_x	[1w1a_Conv, 3×3, 256 1w1a_Conv, 3×3, 256]×2	[1w1a_Conv, 3×3, 256]×2	[1w1a_Conv, 3×3, 64]×3	[1w32a_Conv, 3×3, 64]×3
conv5_x	[1w1a_Conv, 3×3, 512 1w1a_Conv, 3×3, 512]×2	[1w1a_Conv, 3×3, 512]×2		

分析表 1 可知:在 Dy-ResNet18_1w1a 的网络框架中输入大小为 [50, 32, 32, 1] 的特征图像,经过 conv1_x 卷积层后,输出为 [1, 64, 32, 32] 的特征图,经过 conv2_x 之后输出的特征图为 [1, 64, 32, 32], conv3_x 之后输出的特征图为 [1, 128, 16, 16], conv4_x 之后输出的特征图为 [1, 256, 8, 8], conv5_x 之后输出的特征图为 [1, 512, 4, 4], 经过 DyReLU 激活函数后输出结果。ResNet18_1w1a 的卷积方法与 Dy-ResNet18_1w1a 一样,只是在激活函数上使用 hard-tanh 激活函数。ResNet20_1w1a 与 ResNet18_1w1a 的不同之处在于特征图像经过 conv1_x 卷积层后,

输出通道数为 16 的特征图,经过 conv2_x 之后输出通道数为 16 的特征图, conv3_x 之后输出通道数为 32 的特征图, conv4_x 之后输出通道数为 64 的特征图, ResNet20_1w32a 相对于 ResNet20_1w1a 的不同之处在于卷积方法选择不同。

4 实验结果分析

4.1 实验环境

实验中 epoch 设置为 120, 采用随机梯度下降 (Stochastic Gradient Descent, SGD) 的优化方法训练网络, 初始学习率 lr 设置为 0.001, weight_decay

设置为 1×10^{-5} , betas 为 (0.9, 0.999), 并采用余弦退火的学习率衰减策略。在提取特征过程中, 以固定的 50 个点为单位计算特征值。通过将所提出的 Dy-ResNet18 网络模型与不同的二值神经网络模型实验对比, 通过实验结果可以发现, 在训练的总时间上本文方法要花费更少的时间, 但在准确率上稍许降低。

4.2 Oakland 数据集分类任务

实验部分采用 Oakland 3D 数据集作为实验数据集, Oakland 3D 数据集是使用 Navlab11 和侧视的 LMS 激光扫描仪收集的。数据集中每个点都有对应的类别信息, 总共包括电线(wire)、电线杆(pole)、建筑物(facade)、地面(ground)和植被(scatter_misc)等 5 类标签信息。

4.2.1 不同二值神经网络框架对比

在 IR-Net 中, IR-Net 二值神经网络与 GEVAERT 等^[10]提出的 3 种不同网络模型 ResNet18_1w1a、ResNet20_1w1a、ResNet20_1w32a 进行对比证明网络的有效性。在本文的点云分类任务中, 采用初始学习率为 1×10^{-3} , adam 优化器, 总共训练了 120 轮, 实验也进一步将所提出的方法与 3 个网络模型进行对比分析。在 Oakland 3D 数据集下得到的点云分类结果对比如表 2 所示, 实验结果证明本文设计的网络是有效的, 其中 Dynamic ReLU 动态激活函数进一步提高了神经网络效果, 使 Dy-ResNet18 取得了最高的分类精度 97.6%。

表 2 不同二值神经网络框架实验结果对比

Tab. 2 Comparison of experimental results of different network frameworks

Network model	Classification accuracy
Dy-ResNet18_1w1a	97.6%
ResNet18_1w1a	97.2%
ResNet20_1w1a	97.2%
ResNet20_1w32a	96.5%

混沌矩阵分析可以更好的对实验结果进行可视化分析, 对应的各个类别信息的混淆矩阵如表 3 所示。

表 3 Dy-ResNet 网络框架混淆矩阵

Tab. 3 Confusion matrix of Dy-Resnet network framework

	Wire	Scatter_misc	Pole	Ground	Facade
Wire	22.5%	0.01%	60.3%	1.08%	15.2%
Scatter_misc	0.28%	71.8%	19.6%	0.16%	8.16%
Pole	5.02%	0.0%	93.7%	0.01%	1.0%
Ground	0.01%	0.01%	0.00%	99.7%	0.01%
Facade	2.7%	0.01%	0.88%	0.01%	96.4%

由表 3 中的实验数据可知, Dy-ResNet18 的电线分类准确率为 22.5%, 大多数(60.3%)的电线被错误地分类成了电线杆, 5.02% 电线杆也为错误地分类为电线。另外, 有 15.2% 电线被分类为建筑物, 2.7% 的建筑物被区分为电线, 虽然这 2 个物体在形状上有所不同。电线被误分类的情况最为严重, 一方面是因为电线和其他类别物体的重叠遮挡多, 另一方面是电线相较于其他类别物对于局部特征提取的要求更加苛刻, 所以导致了严重的误分类问题, 未来的工作中要加强这一问题的优化。

各个二值神经网络框架不同类别准确率, 如表 4 所示。在表 4 中 Dy-ResNet 18_1w1a 的整体性能最好, Dy-ResNet 18_1w1a 和 ResNet 18_1w1a, ResNet 20_1w32a 的实验对比分析下, 可以得出在同样残差网络的情况下, Dynamic ReLU 动态激活函数能有效提升电线这种依赖局部特征的类别。ResNet20_1w32a 和 ResNet18_1w1a, ResNet20_1w1a 的实验对比分析, 可以得出激活输出的比特更大时, 能更均匀的对各个类别进行识别, 但是会牺牲一定的计算量。未来的工作可以继续探索搭载 Dynamic ReLU 动态激活函数的网络设计。

表 4 各个二值神经网络框架不同类别准确率

Tab. 4 Accuracy of different categories of each binary neural network framework

	Wire	Scatter_misc	Pole	Ground	Facade
Dy-ResNet18_1w1a	22.5%	71.8%	93.7%	99.7%	96.4%
ResNet18_1w1a	0.0%	92.1%	94.8%	98.9%	99.2%
ResNet20_1w1a	10.0%	78.6%	68.8%	99.0%	92.9%
ResNet20_1w32a	34.2%	93.0%	23.9%	99.2%	93.7%

4.2.2 训练时间对比

表5展示不同点云分类网络模型训练所花费的时间。

从表5可知,通过在不同的分类网络框架中进行多次实验,Dy-ResNet18的训练时间消耗与ResNet18接近,说明Dynamic ReLU动态激活函数不会影响计算时间,网络层的深度和激活输出的比特量与计算时间成正比。

表5 不同网络模型训练时间对比

Tab.5 Time efficiency comparison of different network model

Network model	Training time
Dy-ResNet18	6 h±0.5 h
ResNet18_1w1a	6 h±0.5 h
ResNet20_1w1	6.5 h±0.5 h
ResNet20_1w32a	7 h±0.5 h

4.2.3 与其他方法比较

与其他方法在Oakland 3D数据集上实验对比精度如表6所示,在电线杆、建筑物以及整体的分类

上,本文所提方法相对于其他的分类方法,有着最高的平均分类精度,达到了97.6%。

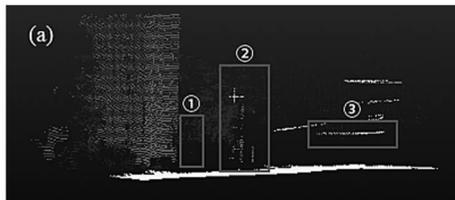
4种不同网络框架分类结果可视化,如图2所示。

从实验结果图2可以得知,图2(d)相较于图2(a),在电线杆(图中紫色部分)的分类上更优。相较于图2(b),在电线杆和电线(图中黄色部分)分类上更优。相较于图2(c),在电线的分类上更优。所以,综合考虑Dy-ResNet18在Oakland 3D数据集上相较于其他的二值神经网络分类效果最好。

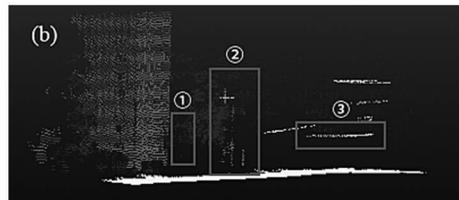
表6 不同方法的对比精度

Tab.6 Comparison accuracy of different methods

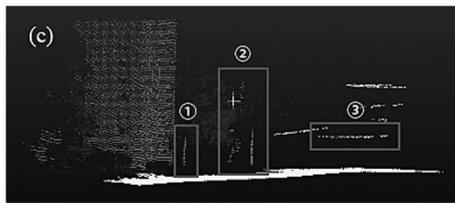
Category	Ref. [19]	Ref. [20]	Ref. [21]	Ref. [22]	Ours
Scatter_misc	90.7%	80.68%	80.55%	92.0%	71.8%
Wire	5.3%	92.93%	93.08%	10.7%	22.5%
Pole	22.3%	68.42%	70.11%	59.7%	93.7%
Ground	99.6%	98.37%	98.22%	99.9%	99.7%
Facade	87.6%	71.13%	70.95%	94.6%	96.4%
All	93.53%	94.75%	94.68%	95.5%	97.6%



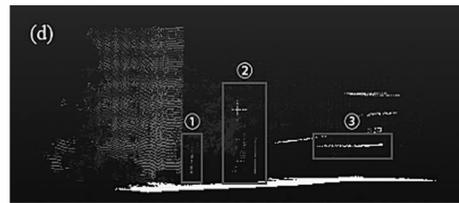
(a) ResNet20_1w32a



(b) ResNet20_1w1a



(c) ResNet18_1w1a



(d) Dy-ResNet18

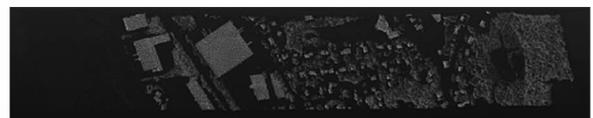
图2 4种不同网络框架分类结果可视化

Fig.2 Visualization of classification results of four different network frameworks

4.3 GML数据集分类任务

GML数据集包含2077万个标记坐标点,有地面、建筑、汽车、树木和低矮植被等7个数据类别,本文从GML_A和GML_B数据集中通过子采样的方法来进行点云分类任务,采样后的数据集可视化如

图3所示。



(a) GML_A



(b) GML_B

图 3 GML_A 和 GML_B 训练数据集可视化

Fig. 3 Visualisation of GML_A and GML_B training datasets

4.3.1 训练消耗时间对比

本方法在 GML 数据集上采用初始学习率为 1×10^{-3} , 学习率(当前学习率 = $1 \times 10^{-3} / 40 \wedge \text{int}(\text{轮数} / 5)$)随着训练轮数而逐渐变化, 总共训练 60 轮, 与 ResNet18_1w1a, ResNet20_1w1 和 ResNet20_1w32a 这 3 种不同网络框架的训练时间对比结果如表 7 所示。

表 7 不同网络模型训练时间对比

Tab. 7 Time efficiency comparison of different network model

Network model	Training time
Dy-ResNet18	4h+0.1h
ResNet18_1w1a	4 h±0.1 h
ResNet20_1w1	5 h±0.1 h
ResNet20_1w32a	5 h±0.5 h

4.3.2 与其他方法进行分类准确率对比

本文将自身的实验结果其他方法进行分类准确率对比, 通过对比实验数据表明, 本方法在 GML_A 数据集上的分类准确率达到 92.3%, 在 GML_B 数据集上达到了 97.2%, 相对于其他的分类方法对比具有一定的优势, 其结果如表 8 所示。

表 8 不同网络模型准确率对比

Tab. 8 Accuracy comparison of different network

Network	GML_A DataSet accuracy	GML_B DataSet accuracy
Our method	92.3%	97.2%
Ref. [23]	90.5%	—
Ref. [24]	78.58%	94.16%
Ref. [25]	—	94.3%

从上表中可以得出 Dy-ResNet 网络框架在 GML 数据集中能够保持很好的稳定性, 不会因为数据集的变化在准确率上变化很大, 具有一定的借鉴意义。

4.3.3 GML_A 和 GML_B 的实验对比结果

图 4 表示 Dy-ResNet 在 GML_A 和 GML_B 的准确率变化, 在 epoch = 30 时, 准确率已经趋向于稳定。

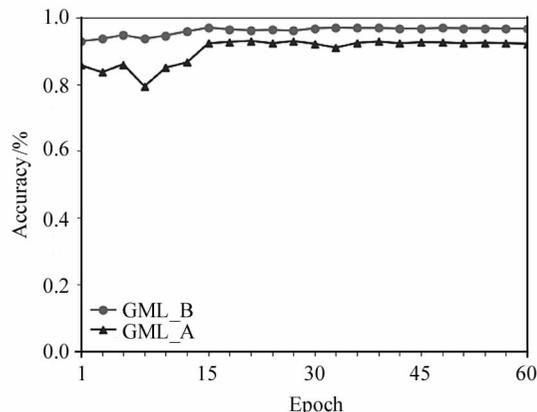


图 4 Dy-ResNet 在 GML_A 和 GML_B 的准确率变化

Fig. 4 The accuracy variation of Dy-ResNet in GML_A and GML_B

5 结 论

本文提出了一种基于二值神经网络的大场景 3 维点云分类框架, 在 Oakland 和 GML 数据集上对比了 Dy-ResNet18, ResNet18_1w1a, ResNet20_1w1a 和 ResNet20_1w32a 等 4 个网络框架。实验结果表明, Dy-ResNet18 二值神经网络在 Oakland 数据集上总训练时间维持在 6.5 小时左右, 总体分类精度上达到了 97.6%, 在单类别物体上的分类精度也具备一定优势, 在 GML 数据集上的训练时间为 4 h 左右, 在 GML_A 和 GML_B 数据集上的分类精度分别达到了 92.3% 和 97.2%, 取得了较好点云分类效果。本文未来的工作主要是在保证准确率的前提下, 降低点云分类的工作对高性能硬件配置的依赖。

参考文献:

[1] QI C R, SU H, MO K, et al. Pointnet: deep learning on point sets for 3d classification and segmentation[C]//IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, Hawaii, USA. New York: IEEE, 2017: 652-660.

[2] QI C R, YI L, SU H, et al. PointNet++: deep hierarchical feature learning on point sets in a metric space[J]. Advances in Neural Information Processing Systems, 2017, 30(3): 231-237.

[3] HUANG Q, WANG W, NEUMANN U. Recurrent slice networks for 3d segmentation of point clouds[C]//IEEE Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, Utah, USA. New York: IEEE, 2018: 2626-2635.

[4] WANG Y, SUN Y, LIU Z, et al. Dynamic graph CNN for learning on point clouds[J]. ACM Transactions on Graph-

- ics, 2019, 38(5):1-12.
- [5] LI Y, BU R, SUN M, et al. PointCNN: convolution on x-transformed points[J]. *Advances in Neural Information Processing Systems*, 2018, 31(2):820-830.
- [6] LIU X, CHEN Y, LI S, et al. Hierarchical classification of urban ALS data by using geometry and intensity information[J]. *Sensors*, 2019, 19(20):4583-4588.
- [7] NIKHIL G N, MERAZ M, JAVED M. Automatic on-road object detection in lidar-point cloud data using modified voxelnet architecture[C]//International Conference on Computer Vision and Image Processing, December 4-6, 2020, Prayagraj, India. Singapore: Springer, 2020:201-213.
- [8] MUNOZ D, BAGNELL J A, VANDAPEL N, et al. Contextual classification with functional Max-Margin Markov Networks[C]//Computer Vision and Pattern Recognition, June 20-25, 2009, Miami, Florida, USA. New York: IEEE, 2009:975-982.
- [9] HU H, MUNOZ D, BAGNELL J A, et al. Efficient 3D scene analysis from streaming data[C]//International Conference on Robotics and Automation, May 6-10, 2013, Karlsruhe, Germany. New York: IEEE, 2013:2297-2304.
- [10] GEVAERT C M, PERSELLO C, VOSSelman G. Optimizing multiple kernel learning for the classification of UAV data[J]. *Remote Sensing*, 2016, 8(2):1025-1039.
- [11] WEST K F, WEBB B N, LERSCH J R, et al. Context-driven automated target detection in 3D data[C]//Automatic Target Recognition XIV, April 13-15, 2004, Orlando, Florida, USA. Bellingham: SPIE, 2004:133-143.
- [12] WOLF D, PRANKL J, VINCZE M. Fast semantic segmentation of 3D point clouds using a dense CRF with learned parameters[C]//IEEE International Conference on Robotics and Automation, May 26-30, 2015, Washington, USA. New York: IEEE, 2015:4867-4873.
- [13] MALLET C, BRETAR F, ROUX M, et al. Relevance assessment of full-waveform lidar data for urban area classification[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2011, 66(6):71-84.
- [14] WEINMANN M, JUTZI B, HINZ S, et al. Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2015, 105(105):286-304.
- [15] LIU Z, WU B, LUO W, et al. Bi-real Net: enhancing the performance of 1-bit cnns with improved representational capability and advanced training algorithm[C]//European Conference on Computer Vision, September 8-14, 2018, Munich, Germany. Berlin: Springer, 2018:722-737.
- [16] MARTINEZ B, YANG J, BULAT A, et al. Training binary neural networks with real-to-binary convolutions[C]//International Conference on Learning Representations, April 26-May 1, 2020, Addis Ababa, Ethiopia. Berlin: Springer, 2020:11535-11541.
- [17] RASTEGARI M, ORDONEZ V, REDMON J, et al. XNOR-Net: imagenet classification using binary convolutional neural networks[C]//European Conference on Computer Vision, October 11-14, 2016, Amsterdam, The Netherlands. Berlin: Springer, 2016:525-542.
- [18] QIN H, GONG R, LIU X, et al. Forward and backward Information retention for accurate binary neural networks [C]//Computer Vision and Pattern Recognition, June 13-19, 2020, Seattle, Washington, USA. New York: IEEE, 2020:2250-2259.
- [19] CHEN Y, DAI X, LIU M, et al. Dynamic relu[C]//European Conference on Computer Vision, August 23-28, 2020, Glasgow, US. Berlin: Springer, 2020:351-367.
- [20] HE E, WANG H, CHEN Q, et al. An improved contextual classification method of point cloud[J]. *Acta Geodaetica et Cartographica Sinica*, 2017, 46(3):362-370.
何鄂龙, 王红平, 陈奇, 等. 一种改进的空间上下文点云分类方法[J]. *测绘学报*, 2017, 46(3):362-370.
- [21] WANG L, MENG W, XI R, et al. Large-scale 3D point cloud classification based on feature description matrix by CNN[C]//Computer Animation and Social Agents, May 21-23, 2018, Beijing, China. Berlin: Springer, 2018:43-47.
- [22] MUNOZ D, VANDAPEL N, HEBERT M. Directional associative markov network for 3D point cloud classification [J]. *International Symposium on 3-D Data Processing, Visualization, and Transmission*, 2008, 4(2):23-31.
- [23] BLOMLEY R, WEINMANN M. Using multi-scale features for the 3D semantic labeling of airborne laser scanning data [J]. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 2017, 4(3):15-21.
- [24] Li H, Meng W, Liu X, et al. Parameter optimization criteria guided 3D point cloud classification[J]. *Multimedia Tools and Applications*, 2019, 78(4):5081-5104.
- [25] CHEN Y, LIU G, XU Y et al. PointNet++ network architecture with individual point level and global features on centroid for ALS point cloud classification[J]. *Remote Sensing*, 2021, 13(3):472-476.

作者简介:

刘儒瑜 (1994—), 女, 博士, 讲师, 主要从事机器人定位与建图, 3 维视觉, 深度学习等方面的研究。