arXiv:2302.00986v1 [cs.AI] 2 Feb 2023

Eloss：智能驾驶中的一种敏感输入质量指标

Haobo Yang、Shiyan Zhang、Zhuoyi Yang、Xinyu Zhang*

s1911593@ed.ac.uk,

zshiyan@bupt.edu.cn,zhuoyiyang03241811@tju.edu.cn,

xyzhang@tsinghua.edu.cn,汽车安全与节能国家重点实验室，以及清华大学车辆与运载学院

摘要

随着交通环境日益复杂，安全感知在智能驾驶中的重要性不断提升。智能驾驶鲁棒感知的传统方法侧重于使用异常数据训练模型，让深度神经网络自行决定如何处理异常。然而，这些模型无法平稳适应多样且复杂的真实世界环境。本文提出一种名为 Eloss 的新型指标，并从异常检测的角度提出一种新颖的训练策略，以增强感知模型。Eloss 基于对感知模型信息压缩层的解释而设计。具体而言，受通信系统设计的启发，信息压缩网络的信息传输过程有两项预期：信息量平稳变化，信息熵持续下降。随后，根据上述预期可得出 Eloss，用于指导相关网络参数的更新，并生成一种敏感指标，在维持模型性能的同时识别异常。我们的实验表明，对于异常数据，Eloss 与标准数值的偏差可超过 100 倍；对于相似但类型不同的异常，Eloss 也会产生有明显区别的数值，这证明了所提方法的有效性。我们的代码见：（论文录用后提供代码）。

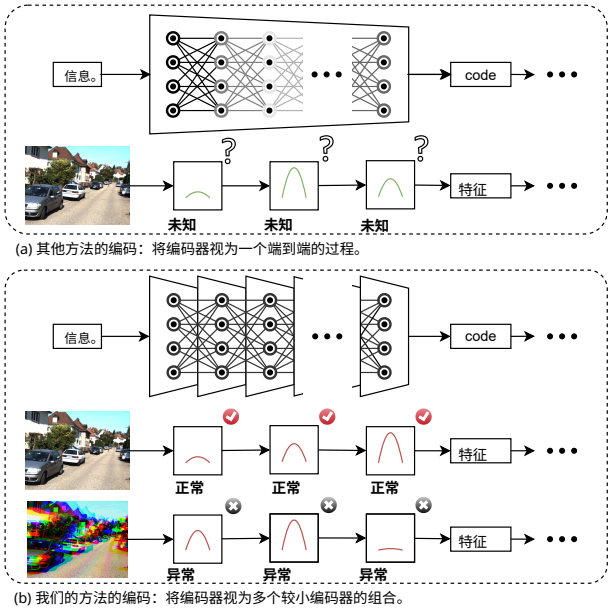


图 1：待更新

引言

智能驾驶是未来城市交通发展的必然趋势（Freudendal-Pedersen、Kesselring 和 Servou，2019），其中三维目标检测是实现智能驾驶的关键任务之一（Zhi 等人，2017）。近年来，随着计算设备的快速迭代以及高质量标注数据集的涌现，数据驱动的深度学习方法推动了三维目标检测任务的快速发展（Wu 等人，2021）。然而，与应用于其他领域的目标检测任务相比，由于驾驶员和道路安全方面，应用于智能驾驶的三维目标检测任务对准确率和速度有更高要求。此外，与其他应用场景相比，智能驾驶车辆中的《传感器》无法保证由于天气等条件不可控，无法稳定地采集高质量数据。因此，可能会出现一些异常数据。如果模型无法判断这些异常数据并进行特殊处理，就会导致智能驾驶系统的决策出错，甚至引发严重的交通事故。

面对智能驾驶对三维目标检测提出的更高要求，许多先进算法都提出了解决方案，以实现更高的准确率和更快的速度。然而，许多相关工作都基于高质量数据建模，着重于在干净数据集上取得更高性能。此外，即使数据集中存在少量低质量数据，模型也能通过大规模训练学到若干异常模式，从而实现较高的准确率。然而，在真实世界中采集的数据种类繁多，模型在数据集中学到的模式不足以应对复杂的驾驶环境。因此，即使在各种数据集上取得了较高的准确率，感知模型仍很可能做出错误判断和决策，造成无法挽回的后果。

*通讯作者 版权所有 © 2023，人工智能促进协会 (www.aaai.org)。保留所有权利。

无法挽回的后果。

异常检测在许多领域都有解决方案，例如 CCD 平整度检测和纺织品缺陷检测。然而，这类工作大多广泛依赖数据来训练“黑箱”模型，不适用于智能驾驶。此外，正如前文所述，模型只能通过训练数据学习有限数量的异常模式，而这些有限的模式不足以应对复杂多变的道路环境。因此，要解决异常问题，我们应从数据本身寻找正常与异常之间的差异。网络的最终目的是提取与目标任务相关的有用传感器信息；对于异常数据，网络无法从中提取足够的信息，甚至可能过度解读这些数据。基于这一思路，我们可以根据模型或模型各层提取的信息量实时检测异常数据，见图 1。

本文从将感知系统视为通信模型的角度出发（Zou 等人，2022），构建了一个理论模型，用以描述智能驾驶感知任务中神经网络底层的信息传输过程。本文的贡献主要包括以下几个方面。首先，基于通信系统中的信源编码理论（Jones 和 Mary Jones，2000），构建了模型信息压缩部分各层信息变化期望值的理论：信息压缩网络各层的信息变化是稳定的。其次，通过引入连续随机变量 X，为神经网络中的数据建立概率模型，以估计信息熵的变化。第三，根据信息变化的期望值，构建了即插即用的 Eloss 函数模块，可将其用作损失函数，引导神经网络参数的更新过程，并赋予网络检测异常的能力。

相关工作

不确定性量化

在真实场景中，自动驾驶感知系统面临着诸多挑战，例如传感器观测中的遮挡和噪声 [1]，而传感器观测往往只能有限地反映环境。因此，对不同环境中感知不确定性的量化引起了广泛关注。

深度学习中的不确定性分为两类：偶然不确定性和认知不确定性 [2]。偶然不确定性源于对环境的不完全了解，例如部分可观测性和测量噪声；即使获取更多乃至无限的数据，也无法降低这种不确定性，但可以通过显式建模来降低。通常使用异方差损失函数来学习偶然不确定性[3]。认知不确定性源于数据集不足以及模型尚未掌握的某些方法，只要有足够的训练数据，就可以消除这种不确定性。通常有两种常用方法来估计认知不确定性：蒙特卡洛（MC）丢弃法 [4]和集成方法 [5]。

以往研究中的不确定性量化缺少不确定性估计的真实值 [6]以及统一的定量评价指标 [7]。具体而言，在不同的机器学习任务中，不确定性的定义各不相同，例如分类、分割和回归。本文提出的方法可以为常规数据输入生成真实值 0，而我们的 Eloss 是一种统一的方法，适用于所有具有重复结构的任务或网络。我们将在下一节讨论这一点。

通信模型中的 Shannon 信源编码

我们可以利用通信模型构建神经网络，并运用信息论中的先验知识来解释和指导神经网络优化。

信息论主要由 Shannon 创立（Jones 和 Mary Jones，2000），它运用熵的概念，基于信息量化研究信息过程。信息熵通过量化信号中的不确定性，揭示了信号处理和通信操作的局限性。由于通信网络具有便于量化的特性，许多研究将通信领域的信息论知识引入深度神经网络。

MacKay 等人提出了由神经元或网络构建的信道模型（Mackay 2003）。Sharma 等人将基准编码引入变分自编码器（Sharma 等，2021）。利用香农第一定理，将编码的平均长度与熵的大小进行比较，以确保变分分布是无失真的编码过程。

2015 年，Tishby 和 Zaslavsky 使用信息瓶颈（tis 2000）进一步揭示了深度学习模型的机制。他们使用多层感知机进行的实验表明，网络倾向于先捕获相关信息，然后再将其组合起来。受 DAGSurv（Sharma 等，2021）工作的启发，我们将单一模态的特征提取融合前处理过程建模为信源编码，并通过将该部分与后续的融合网络和检测网络进行联合训练，实现了信息量的压缩并提高了模型效率，同时选择性地捕获有效特征并减少无效特征。此外，我们引入信息论中的熵来量化信源编码每一层输出中的信息量，通过限制熵的变化来减少网络可能的优化方向，从而加速模型优化。

神经网络优化器

机器学习中的优化器优化过程，是寻找能够显著降低损失函数的神经网络参数。损失函数通常包括在整个训练集上评估的性能指标，以及额外的正则化项。为了使模型输出接近或达到最优数值，我们需要使用各种优化策略和算法，更新并计算影响模型训练和模型输出的神经网络参数。目前，解决这一问题的主要方法

大致分为三类：梯度下降法、动量优化法和自适应学习法。

在梯度下降法中，小批量和中批量梯度方法结合了批量梯度下降（Hinton、Srivastava 和 Swersky，2012）与随机梯度下降（Bottou，2012）的优点，在样本选择时选取少于训练样本总数的小批量样本，既保证了训练速度，也保证了最终收敛的准确率（Ruder，2016）。

动量优化方法引入了物理学中的动量概念，以加速梯度下降；常见算法有动量法和 NAG。采用动量优化方法时，在梯度方向不变的维度上，参数更新会变快；而当梯度发生变化时，参数更新会变慢，从而加快收敛并减少震荡（Dozat，2016）。其中，NAG 是对动量法的改进：它在更新参数时保留上一次更新的方向，并利用当前梯度微调最终的更新方向，同时在梯度更新时引入修正。这种方法非常擅长加速收敛并抑制振荡。

在深度学习中，学习率是一个非常重要的超参数，通常难以确定，往往需要经过一定次数的训练才能找到最佳学习率（Smith 2017）。自适应学习率优化算法能够根据某些策略自适应地调整学习率大小，从而提高训练速度。目前，自适应学习率优化算法主要包括：AdaGrad算法、RMSProp算法、Adam算法和AdaDelta算法（Le et al. 2011；Zaheer and Shaziya 2019）。Adam中的动量法直接融入了一阶梯度矩的估计中。同样，Adam还包含偏差校正，用于校正从原点初始化的一阶和二阶矩估计。这使参数相对平滑，适用于大多数非凸优化问题，以及大型数据集和高维空间。

理论

为了高效融合各模态的信息，并去除对后续任务无益的信息，必须对信息进行压缩，使输入后续网络的特征能够为最终任务作出贡献。通信模型中的这一信息压缩过程可以表示为失真受限编码。为提高信息传输效率，在采用失真受限编码进行压缩的过程中，需要去除原始信源中出现频率较低的信源符号。

由于这些信源符号出现的频率较低，即使它们在编码过程中丢失，在信道另一端仍能实现较高的数据恢复率，并大幅提高传输效率。同样，在信息压缩网络中，被移除的信息通常对后续任务没有贡献，移除这些信息可以大幅提高后续网络的效率（Jones 和 Mary Jones，2000）。

在通信模型和信息论中，熵常用来衡量信息量。熵越低，信息量越大。因此，我们提出了一种熵计算方法，用于描述失真受限编码过程中信息的变化，稍后将对此进行详细说明。同时，为了确保信息压缩过程平稳，我们使信息通过神经网络各层时的熵变化保持恒定，防止信息发生突变失真。

神经网络层的熵期望

神经网络训练过程可以看作是不断寻找输入与期望输出之间映射关系，以实现学习效果的过程（Liu等人，2017）。可以认为，在神经网络训练之前，所学到的映射关系较弱；而在使用数据训练神经网络时，所学到的映射关系会不断增强，网络各层的特征提取能力，也就是信息压缩能力，也会持续提升。然而，对于这样的黑盒模型（Buhrmester，M¨unch和Arens，2021），目前尚无对其信息压缩过程的良好解释，更无法有效指导信息压缩模型的参数优化过程。

为解释信息压缩网络的底层机制，参照通信模型，特征压缩可视为一种信源编码过程。为确保网络能够以更完整的方式提取对后续任务有用的数据，信源编码应采用失真受限编码，如上文所述。

现在，我们引入信息熵的概念，以反映网络输出的信息量。在带宽恒定的通信系统中，如果传输数据的信息熵持续降低，信息传输效率便会逐步提高（Zou 等人，2022）。

在失真受限编码器中，信息熵会随着编码程度的提高而降低，特征压缩网络也是如此。由于特征压缩网络通常具有连续的重复网络结构，例如 SECOND 网络中的重复线性层（Yan、Mao 和 Li，2018年），且每个重复的网络层都应具有相似的特征压缩能力，因此可以认为网络中的数据传输带宽是恒定的。总而言之，对于特征压缩网络而言，逐层信息量变化的预期是：网络每个重复层输出结果的信息熵持续降低。

根据期望，我们可以优化网络参数，使其朝着保持稳定变化的方向调整

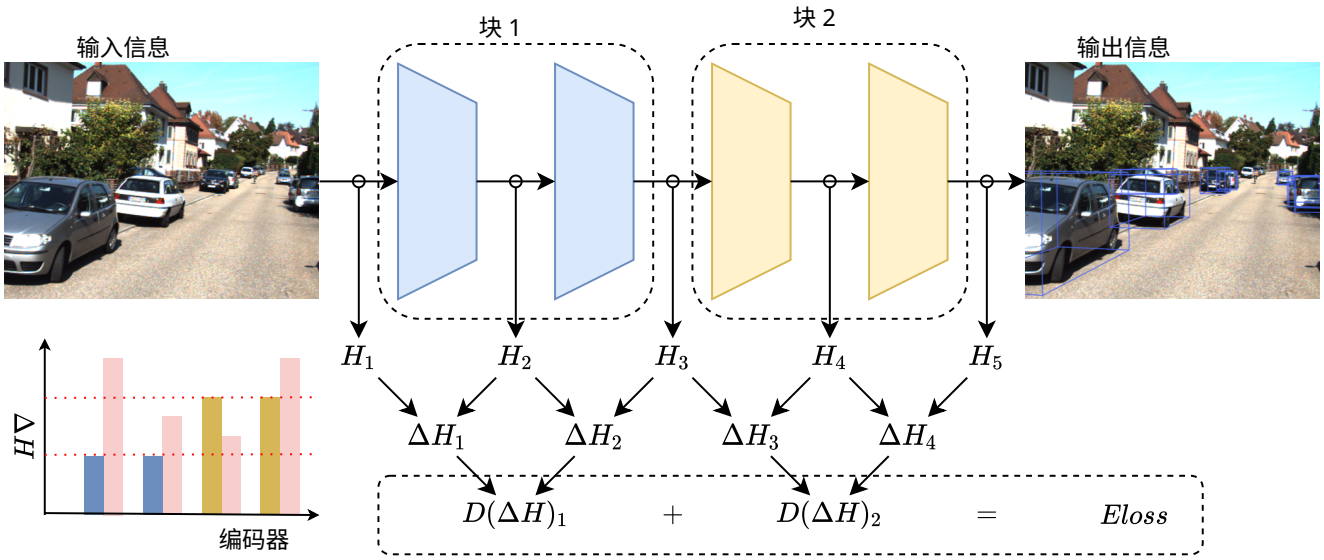


图2：我们提出的 Eloss 方法的详细示意图。

通过构建熵损失函数，逐层优化信息熵。因此，我们能够深入黑盒模型，找到进一步优化训练过程的方法。

异常信息输入的不确定性量化

对于具有特定功能的特征提取网络，原始输入数据经过各个特征提取层后会形成新的特征图。在这一过程中，信息量会发生变化，也就是信息熵会发生变化。然而，由于特征提取过程中不会引入新的信息，且会滤除大量与目标任务无关的信息，只保留与目标任务相关的信息，因此通常预期这一过程会使信息熵降低。

在同一个特征提取网络中，网络各层的特征提取能力与该层参数的规模呈正相关。因此，当特征图依次经过若干个参数量相同且结构相似的特征提取层时，每次特征提取后都应保持相近的信息削减量，以确保网络中的信息压缩过程平稳进行。

然而，对于异常的输入特征图，例如其中某些信息含有噪声时，特征图信息量在经过特征提取器的各层时，其变化将变得不规则，而不是平滑递减，熵值甚至可能增加。因此，我们可以观察输入数据经过各个特征提取器时的熵，以判断输入信息是否异常。

信息概率建模

基于熵的期望计算的损失函数，需要一种方法来估计每个信息压缩网络中层输出的熵。此外，估计熵需要对输出数据的分布进行概率建模。

在信息压缩中，通常会将卷积网络视为一种选择（Zou 等人，2022）。为了将卷积神经网络建模为概率模型，我们作如下设定：由卷积核生成的特征通道 $X=\{x_1, x_2, ..., x_i\}$ 被视为多维连续随机变量 X的样本，其中 i 是通道数，而每个通道中的数值数量是 X 的维度 d。

由于任何网络各层的输出都可以视为连续随机变量 X，并且任意一层的输出都可以选作X的一组样本 x_i，因此，除了卷积神经网络之外，概率建模也可应用于其他神经网络结构，以计算数据分布的熵。因此，所提出的针对卷积神经网络的概率建模方法可以扩展到其他神经网络结构。

熵计算

现在，估计神经网络各层熵的问题转化为根据未知连续随机变量 X的概率分布来估计熵。

解决这个问题有很多方法，其本质在于计算连续随机变量的微分熵。微分熵也称为连续熵，是信息论中的一个概念，源于 Shannon 尝试将其熵概念扩展到连续概率分布（Jones 和 Mary Jones，2000）。设随机变量 X，其概率密度函数 f 的定义域为集合 X。该微分熵 $h(X)$ 或 $h(f)$ 定义如下：

$$h(X) = -\int f(x)\log f(x)dx \tag{1}$$

由于这个问题中随机变量的概率分布事先未知，概率密度函数也未知，而该概率分布中只有有限个样本值可用。下面介绍可用于计算其信息熵的 K近邻熵估计法（van de Water 和 Schram，1988）：

连续变量经过采样后成为离散变量。为了用 n 个样本近似整个样本空间，将每个样本点扩展为一个 d 维超球体，球体半径为该样本点与最近样本点之间的距离。当变量在样本空间中均匀分布时，每个样本点的概率可近似为 $1/n$。

由于随机变量在样本空间中的分布未知，其分布可能与均匀分布存在较大差异，因此利用样本在样本空间中的分布对空间内随机变量的分布进行修正。样本在样本空间中的密度和稀疏程度会直接影响各样本点附近的概率密度。每个样本点的离散概率估计如下：

$$p(x_i) = [(n-1)\cdot r_d(x_i)^d]\cdot V_d]^{-1} \tag{2}$$

其中，n 为样本数，r_d（x_i）是样本 x_i 与其最近样本点之间的 d 维欧几里得距离，V_d 是 d 维空间中单位球的体积。

随机变量 X 的熵估计值为：

$$H(X) = \frac{1}{n}\sum_{i=1}^{n}[-\log p(x_i)] + \gamma \tag{3}$$

其中，γ 为欧拉-马歇罗尼常数，约等于 0.5772。

K近邻熵估计法将每个样本点与其最近样本点之间的距离扩展为该样本点与第 k 个最近样本点之间的距离，随机变量 X 的熵估计值变为：

$$H(X,k) = -\psi(k) + \psi(n) + \log V_d + \frac{d}{n}\sum_{i=1}^{n}\log r_{d,k}(x_i) \tag{4}$$

其中，ψ 是双伽马函数，$\psi(1) = -\gamma$,$\psi(n) \sim \log(n-1)$。$r_{d,k}(x_i)$是样本 x_i 与其第 k 个最近样本点之间的 d 维欧几里得距离。可以证明，当 $k = 1$时，$H(X)$等价于$H(X,k)$。我们将 $H(X)$作为网络各层输出的熵值，进而获得网络各层的熵变量 ΔH，$\Delta H_n = H_{n+1} - H_n$，其中 n 是网络层的索引。

信息压缩网络损失函数

根据对信息压缩网络的信息传输预期，可以建立两个损失函数：函数 L_1 关注信息的稳定变化，函数 L_2 关注信息变化的方向。

损失函数 L_1 根据熵变化 ΔH的方差构造，且 L_1 的期望值为 0，其中 L_1为方差。

$$L_1 = \frac{\sum_{n=1}^{N}(\Delta H_n - \widehat{\Delta H})^2}{N} \tag{5}$$

其中，N、n 为重复网络$\widehat{\ \ }$层的层索引，ΔH 为所有重复网络层的平均熵变化。

根据熵递减预期，损失函数 L_2 的形式如下：

$$L_2 = -\sum_{n=1}^{N}\Delta H_n^2 \tag{6}$$

我们将 L_1 和 L_2 合称为Eloss。

当网络使用Eloss进行训练时，Eloss的影响可以被描述为一种放大器，增强网络某些部分的可解释性，此时该部分可以被描述为特征压缩网络。但这也是Eloss方法的局限性，因为 L_1 和 L_2能够产生影响的网络层，仅限于那些重复多次且执行与通信网络密切相关的特定任务的层。我们将在以下章节中评估这一局限性。因此，Eloss是对损失函数 L的补充，而后者与最终目标完全对应。

实验

数据集。我们在 KITTI 数据集和 nuScenes 数据集上进行了实验。KITTI 数据集由德国卡尔斯鲁厄理工学院和美国丰田技术学院共同创建（Geiger、Lenz 和 Urtasun，2012；Geiger 等，2013）。此外，该数据集是智能驾驶场景中使用最广泛的计算机视觉算法评测数据集之一。数据集包含从城市、村庄和高速公路等场景中采集的真实世界图像数据。KITTI 数据集包含多模态数据，例如激光雷达点云、GPS数据、右侧彩色摄像头数据和灰度相机图像。KITTI 数据集分为训练集和测试集，其中训练集包含 7481 个样本，测试集包含 7518 个样本。nuScenes（NuTonomy Scenes）数据集是一个用于自动驾驶车辆行驶的多模态数据集。它是首个为自动驾驶车辆提供全套传感器数据的大规模数据集，其中包括六个摄像头、一台激光雷达、五个毫米波雷达、GPS和惯性测量单元。其目标标注数量是 KITTI 数据集的七倍以上，包括 140 万张相机图像、39 万份激光雷达扫描数据，以及人工标注的 140 万个三维边界框，涵盖

23个物体类别。nuScenes 数据集由 1000 个场景组成，每个场景相当于 20 秒的视频，涵盖多种场景。每个场景有 40 个关键帧，即每秒两个关键帧，其余为扫描帧。关键帧均由人工标注，每帧都包含一些以边界框形式呈现的标注，例如尺寸、围栏、类别和可见性等。现有自动驾驶数据集缺乏用于构建自动驾驶感知系统的全套多模态数据，而 nuScenes 弥补了这一不足。

实现细节。 实验使用 Nvidia RTX 3090 设备进行，模型基于 MMDetection3D（Contributors 2020）框架，使用 PyTorch 构建。PyTorch 广泛应用于深度学习，并提供大量 Python接口，使框架能够轻松调用和使用 Python 自带的函数包。PyTorch 内置的自动微分功能使其成为一种非常流行的动态图框架。MMDetection3D 是一个基于 PyTorch 的开源三维目标检测工具箱，本文使用的模型均基于 PyTorch，在 MMDetection3D 框架上开发。

评估 Eloss 对异常输入的敏感性分别使用经过 Eloss 训练和未经 Eloss 训练的 PointPillars模型处理测试数据集中的 KITTI 数据集和 nuScenes 数据集，并使用经过 Eloss 训练和未经 Eloss 训练的 VoxelNet模型分别处理测试数据集中的 KITTI 数据集，然后比较所得的 Eloss。接着计算未添加 Eloss 的模型对数据集进行处理后所得结果的置信度。将实验结果中的 Eloss 与实验结果的置信度进行比较，并进一步探究经过 Eloss 训练的模型所得结果与未经 Eloss 训练的模型所得结果之间的 Eloss 差异，探究哪些指标能够更好地反映异常数据，并说明对异常数据的敏感性差异。结果如表1所示。

将未添加 eloss 训练的模型所获得的置信度，与处理不同测试数据集后得到的 eloss 数值进行比较。研究发现，在不同模型下处理添加了不同噪声的数据集时，添加噪声后的置信度低于未添加噪声时的置信度。由此可以识别出此时的数据异常。但是，当噪声添加比例不同时，结果的置信度变化很小，对数据异常的敏感性较低，缺乏区分度，也无法判断噪声添加比例。据推测，置信度存在对添加噪声过度解读的问题。在这种情况下，eloss 数值同样能够识别数据异常，发挥相同的作用，尽管其数值呈现出混乱状态。不过，它对某些具有特定噪声比例的数据尤其敏感，能够检测出不同大小的噪声。也就是说，在使用同一模型获得相同检测结果时，以 eloss 作为指标能够更好地判断数据是否异常，并能检测出具有区分度的异常数据。

比较未添加 eloss 训练的模型处理不同测试数据集后得到的 eloss，与添加 eloss 训练的模型处理不同测试数据集后得到的 eloss 数值，可以发现，后者的 eloss 数值对数据异常更加敏感，波动也更加明显，尽管结果的准确率有所降低。也就是说，使用相同模型时，经过 eloss 训练的模型会牺牲部分结果准确率；但以牺牲一定检测准确率为代价，可以进一步提升对异常数据的检测能力。

比较这两种情况下的 eloss 数值与置信度可以发现，尽管置信度能够反映数据是否异常，但在添加不同比例的噪声时，其变化幅度极小，无法进一步区分；而将 eloss 数值作为判断数据异常程度的指标，尤其是使用经过 eloss 训练的模型所得到的 eloss 数值，尽管检测准确率有所下降，却不仅能够同样判断数据异常，而且 eloss 波动更大，进一步提升了对异常数据的检测能力。

与 Eloss 在训练过程中的比较

为衡量 Eloss 对模型训练过程的影响，我们首先在无噪声的 KITTI数据集和 Nusenes数据集上，针对同一模型分别开展有 Eloss 和无 Eloss 的对照实验。我们将部分实验结果绘制在图3中，以更直观地展示 Eloss 对训练过程波动性的影响。

为量化 Eloss 对训练过程的影响，我们使用平均绝对值斜率（Mean Absolute Value Slope，MAVP）来衡量 Eloss 对模型训练曲线波动性的影响，并使用最大精度指标来衡量 Eloss 对模型训练准确率的影响。MAVP 的公式如下，其中 N 为滑动窗格的数量，(k, k+1) 表示两个相邻的时间窗口。

$$\text{MAVP} = \frac{1}{N}\sum_{k=1}^{N}(|x_{k+1}| - |x_k|)) \tag{7}$$

我们将 PointPillars 和 VoxelNet模型应用于 KITTI数据集，以开展上述对照实验。实验结果如表2所示。

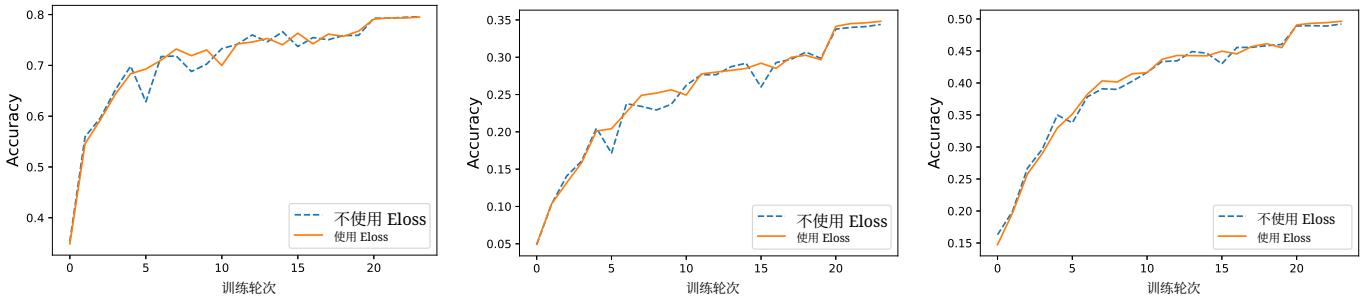
数据透视表中的实验结果表明，在两个模型中添加 eloss 后，最大训练准确率均有所下降。就 MAVP 而言，添加 eloss 后 MAVP 降低，这意味着加入 eloss 使上述训练过程更加平稳。

在 Nuscence数据集上，我们使用三种不同的指标：Car $AP_dist1.0$、平均精度均值和 NDS，对 PointPillars模型进行了上述对照实验。实验结果如表3所示。

表中显示，添加 Eloss 后，Car类别在训练过程中的最大准确率有所下降。不过，MAVP 的下降表明，添加 Eloss 减缓了上述训练过程中的波动。对于多类别的平均精度均值，

设置			置信度			Eloss（指标）			Eloss（指标和损失函数）		
模型	数据集	数值	干净	噪声1	噪声2	干净	噪声1	噪声2	干净	噪声1	噪声2
VoxelNet	KITTI	Mean	0.495	0.248	0.248	0.015	0.008	0.009	1.584E-03	9.085E-03	8.697E-03
		变化百分比	0.0%	-49.9%	-49.9%	0.0%	-48.5%	-39.1%	0.0%	**473.5%**	**449.0%**
PointPillars	KITTI	Mean	0.487	0.344	0.344	0.012	2.086	0.008	1.091E-01	3.836E+00	-
		变化百分比	0.0%	-29.3%	-29.3%	0.0%	**17475.6%**	-36.1%	0.0%	**3416.0%**	-
PointPillars	NuScenes	Mean	0.168	0.128	0.128	0.034	1.918	0.016	2.556E-04	1.478E-01	4.273E-03
		变化百分比	0.0%	-23.7%	-23.7%	0.0%	**5494.7%**	-54.5%	0.0%	**57746.3%**	**1571.9%**

表1：比较置信度、作为指标的 Eloss，以及作为指标和损失函数的 Eloss。在不同噪声设置下的观察结果。



(a) 汽车 $AP_{dist1.0}$　(b) 平均精度均值　(c) NDS

图3：采用或不采用 Eloss 时，PointPillars 模型在 NuScenes 验证集上的准确率收敛曲线。(a) 距离阈值为 1.0 米时的汽车检测平均精度；(b) 在 10 个物体类别上计算的平均精度均值；(c) nuScenes 检测得分。

模型	方法	最大值（%）	MAVP（%）
PointPillars	不使用 Eloss	90.694	11.946
	使用 Eloss	88.916	11.932
	增量	-1.778	**-0.014**
VoxelNet	不使用 Eloss	94.873	10.959
	使用 Eloss	94.586	10.937
	增量	-0.287	**-0.022**

表2：训练过程中模型在 KITTI 验证集上的汽车 AP_{R40} 最大值和 MAVP。

指标	方法	最大值（%）	MAVP（%）
汽车 $AP_{dist1.0}$	不使用 Eloss	79.580	2.945
	使用 Eloss	79.520	2.811
	增量	-0.060	**-0.135**
mAP	不使用 Eloss	34.393	6.036
	使用 Eloss	34.815	4.883
	增量	**0.422**	**-1.153**
NDS	不使用 Eloss	49.217	4.902
	使用 Eloss	49.637	3.902
	增量	**0.420**	**-1.000**

表3：PointPillars 在 NuScenes 验证集训练过程中的最大值和 MAVP。

以及 NuScenes 检测评分 NDS，也有类似观察结果，这表明在模型中添加 Eloss 能使训练过程更加平稳。

以上是关于 eloss 对模型训练影响的实验，未受到噪声干扰。为了进一步了解 eloss 在实验中的作用，我们将在网络的不同部分加入 Eloss，或使用异常数据进行对照实验。

Eloss 的不同用量

在本实验中，我们对已训练 80 个训练轮次的模型继续进行训练，并观察将 Eloss 添加到网络的不同部分后，对模型在 KITTI 测试集上的准确率有何影响。

Eloss 是一种即插即用模块，这意味着可以将 Eloss 添加到具有重复网络结构的任意位置并使其发挥作用。因此，我们可以通过考虑网络的不同部分来评估 Eloss。为方便起见，本实验仅考虑 SECOND 骨干网络中按块分组的重复结构。

整理最佳模型的实验结果后，我们得到表4。

在这些表格中，80 个训练轮次时的模型准确率与微调模型的准确率之间存在差距。因此，可以合理地预期模型在完成 80 个训练轮次后继续训练。在验证集中，应用 Eloss 后准确率如预期般提升。然而，在测试集中，我们得到了相互矛盾的结果。

对于 PointPillars，在表 4 中，随着 Eloss 影响更多 SECOND 块，汽车 AP_R40 数值呈下降趋势，这与之前的实验结果一致。原因是 Eloss 对模型参数施加了更多约束，使模型的训练过程更加困难。

模型	训练轮次	Eloss	汽车 AP_{R40}			时间(ms)
			简单（%）	中等（%）	困难（%）	
PointPillars	80	0 个块	84.20	74.82	68.02	4.50
	85	1 个块	82.65	71.30	64.66	7.41
	85	2 个块	80.90	69.79	64.34	16.74
	85	3 个块	80.23	68.92	63.49	34.88
	best	0 个块	-	77.60	-	-
VoxelNet	80	0 个块	82.35	73.35	68.59	6.94
	85	1 个块	81.34	70.46	65.29	14.72
	85	2 个块	85.34	74.44	67.51	33.75
	best	0 块	-	79.07	-	-

表 4：在 KITTI数据集上连续训练后，采用不同 Eloss 设置的 VoxelNet 和 PointPillars 的汽车三维检测准确率。

表 4 中 voxel-net 的结果与之前的观察不同。当 Eloss 影响一个 SECOND 块时，测试集上的准确率略有下降；但当 Eloss 影响两个 SECOND 块时，准确率则有所提高。

针对 VoxelNet 使用 Eloss 时出现的异常现象，我们进行了更多实验，以作进一步评估。

不同模型之间的比较

在本实验中，我们使用与 VoxelNet 相同的体素编码器，并对 SECOND 中可能受到 Eloss 影响的块进行预训练（使用或不使用 Eloss），然后冻结其参数。接着，我们改变整个模型的规模，增加模态数量和复杂程度。结果见表 5。

在 SECOND+ResNet（He 等人（2016））实验中，我们使用了两种模态的信息：点云和图像。骑行者和行人检测的准确率均有所提高，增幅均超过 3%，但汽车检测的准确率有所下降。

在最后一个模型 SEC- OND+RESNET+相关性（Yan、Mao和 Li（2018年）；He 等人（2016年）；Zheng 等人（2022年））+图神经网络（Scarselli 等人（2008年））+特征金字塔网络（Lin 等人，2017 年）中，情况变得更糟，只有行人检测的准确率有所提高。

这些观察结果表明，对于采用体素编码器的网络，Eloss 的影响越显著，该网络在目标检测任务上的表现就越好。

在这个阶段，我们推测出现上述现象的原因是：受 Eloss 影响的块有利于压缩，因为信息会在这些块中传递；但压缩结果很脆弱，在训练过程中很容易受到后续推理过程的干扰而失真。这表明，如果使用 Eloss，我们需要考虑降低 Eloss 覆盖范围以外各层的学习率。未来我们将投入更多精力研究这一问题。

结论

本文提出了 Eloss，这是一种基于通信系统背后理念、用于增强特征压缩网络可解释性的工具。Eloss 通过将网络层输出与根据信源编码生成的信息变化预期进行比较而构建。我们利用 Eloss 优化网络，使其朝着收敛于预期的方向发展。

通过从三个不同方面开展实验，我们的研究表明，使用 Eloss 训练三维目标检测模型有利于提高训练速度和模型可解释性。尽管取得了这些结果，我们仍然发现研究存在局限性。对于只有很小一部分网络可受 Eloss 影响的模型，加入 Eloss 可能会对模型训练过程造成意外干扰。未来，我们将详细研究这一局限性，并继续开展智能驾驶领域的模型可解释性研究。

致谢

本研究得到国家高技术研究发展计划（项目编号：2018YFE0204300）和国家自然科学基金委员会（项目编号：U1964203）资助，并由美团和清华大学-滴滴未来出行联合研究中心赞助。

参考文献

2000年。 信息瓶颈方法。arXiv:physics/0004057。

Bottou, L. 2012. 随机梯度下降技巧。载于神经网络：技巧与窍门，421–436。施普林格。Buhrmester, V.; Munch, D.; and Arens, M. 2021. 面向计算机视觉的黑盒深度神经网络解释器分析：一项综述。机器学习与知识提取，3(4): 966–989。Contributors, M. 2020. MMDetection3D：OpenMMLab 新一代通用三维目标检测平台。https://github.com/open-mmlab/mmdetection3d。Dozat, T. 2016. 将 Nesterov 动量融入 Adam。Freudendal-Pedersen, M.; Kesselring, S.; and Servou, E. 2019. 什么才是未来城市的智慧之选？出行与自动化。可持续发展，11(1): 221。Geiger, A.; Lenz, P.; Stiller, C.; and Urtasun, R. 2013. 视觉与机器人学相遇：KITTI数据集。《国际机器人研究杂志》，32(11): 1231–1237。Geiger, A.; Lenz, P.; and Urtasun, R. 2012. 我们准备好迎接自动驾驶了吗？KITTI视觉基准套件。载于*2012*年*IEEE*计算机视觉与模式识别会议，3354–3361。IEEE。He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. 用于图像识别的深度残差学习。载于*IEEE*计算机视觉与模式识别会议论文集，770–778。Hinton, G.; Srivastava, N.; and Swersky, K. 2012. 机器学习神经网络第6a讲：小批量梯度下降概述。引文，14(8): 2。Jones, G. A.; and Mary Jones, J. 2000. 信息与编码理论。施普林格伦敦。Le, Q. V.; Ngiam, J.; Coates, A.; Lahiri, A.; Prochnow, B.; and Ng, A. Y. 2011. 深度学习的优化方法。载于国际机器学习大会。

模型	方法	汽车 3D AP_{R40}			骑行者 3D AP_{R40}			行人 3D AP_{R40}		
		简单（%）	中等（%）	困难（%）	简单（%）	中等（%）	困难（%）	简单（%）	中等（%）	困难（%）
SECOND（Yan、Mao和Li（2018））	不使用 Eloss	82.35	73.35	68.59	70.89	56.72	50.68	50.75	40.76	36.96
	使用 Eloss	82.68	73.67	67.21	71.99	58.00	50.94	50.49	41.16	37.43
	增量	**+0.33**	**+0.32**	-1.38	**+1.1**	**+1.28**	**+0.26**	-0.26	**+0.4**	**+0.47**
+ResNet（He 等人（2016））	不使用 Eloss	80.29	67.37	60.94	75.70	52.37	46.10	39.64	31.13	28.95
	使用 Eloss	77.62	64.92	60.36	71.47	55.79	49.64	44.85	35.60	32.66
	增量	-2.67	-2.45	-0.58	-4.23	**+3.42**	**+3.54**	**+5.21**	**+4.47**	**+3.71**
+相关性（Zheng 等人，2022 年）	不含 Eloss	73.47	62.47	57.99	63.08	49.55	44.33	42.46	35.11	32.16
+图神经网络（Scarselli 等人，2008 年）	采用 Eloss	67.33	58.70	54.13	57.16	46.36	41.54	45.02	36.39	33.29
+特征金字塔网络（Lin 等人，2017 年）	增量	-6.14	-3.77	-3.86	-5.92	-3.19	-2.79	**+2.56**	**+1.28**	**+1.13**

表 5：比较 3 种不同模型在训练 40 轮次后，在 3 个类别（汽车、骑行者和行人）的测试集上使用或不使用 Eloss 时的准确率。

Lin, T.-Y.; Doll´ar, P.; Girshick, R.; He, K.; Hariharan, B.; and Belongie, S. 2017. 用于目标检测的特征金字塔网络。载于*IEEE*计算机视觉与模式识别会议论文集，2117–2125。Liu, W.; Wang, Z.; Liu, X.; Zeng, N.; Liu, Y.; and Alsaadi, F. E. 2017. 深度神经网络架构及其应用综述。神经计算，234: 11–26。Mackay, D. J. C. 2003. 信息论、推断与学习算法。剑桥大学出版社，（Imp. Ruder, S. 2016. 梯度下降优化算法概述。*arXiv*预印本 *arXiv:1609.04747*。Scarselli, F.; Gori, M.; Tsoi, A. C.; Hagenbuchner, M.; and Monfardini, G. 2008. 图神经网络模型。*IEEE*神经网络汇刊，20(1): 61–80。Sharma, A. K.; Kukreja, R.; Prasad, R.; and Rao, S. 2021. DAGSurv：基于有向无环图的深度神经网络生存分析。载于亚洲机器学习会议，1065–1080。PMLR。Smith, L. N. 2017. 用于训练神经网络的循环学习率。载于*2017*年*IEEE*冬季计算机视觉应用会议（*WACV*），464–472。IEEE。van de Water, W.; and Schram, P. 1988. 基于近邻信息的广义维数。《物理评论*A*》，37(8): 3118。Wu, Y.; Jiang, X.; Fang, Z.; Gao, Y.; and Fujita, H. 2021. 通过二维引导的精确锚框提议和多层融合实现多模态三维目标检测。应用软计算，108: 107405。Yan, Y.; Mao, Y.; and Li, B. 2018. SECOND：稀疏嵌入式卷积检测。《传感器》，18(10): 3337。Zaheer, R.; and Shaziya, H. 2019. 深度学习中优化算法的研究。载于*2019*年第三届国际创新系统与控制会议（*ICISC*），536–539。IEEE。Zheng, S.; Zhu, Z.; Liu, Z.; Guo, Z.; Liu, Y.; Yang, Y.; and Zhao, Y. 2022. 用于疾病预测的多模态图学习。*IEEE*医学影像汇刊。Zhi, S.; Liu, Y.; Li, X.; and Guo, Y. 2017. LightNet：一种用于实时三维目标识别的轻量级三维卷积神经网络。载于*3DOR@ Eurographics*。Zou, Z.; Zhang, X.; Liu, H.; Li, Z.; Hussain, A.; and Li, J. 2022. 一种基于联合编码模型的车道线分割新型多模态融合网络。信息融合，80: 167–178。