神经网络校准

Ruslan Vasilev[*] **Alexander D'yakonov**[*]

artnitolog@yandex.com djakonov@mail.ru

摘要

解决现实问题的神经网络通常不仅需要做出准确的预测，还需要给出预测的置信度。模型的校准程度体现了估计置信度与真实概率之间的接近程度。本文综述了神经网络领域的置信度校准问题，并对校准方法进行了实证比较。我们分析了问题陈述、校准定义以及不同的评估方法：用于估计模型是否校准良好的可视化和标量度量。我们回顾了现代校准技术，包括基于后处理的方法，以及需要改变训练过程的方法。实证实验涵盖了各种数据集和模型，并根据不同标准比较了校准方法。

arXiv:2303.10761v1 [cs.NE] 19 Mar 2023

[*]这项工作于2020–2021在莫斯科国立罗蒙诺索夫大学完成，是Ruslan Vasilev的三年级课程作业，指导教师为Alexander D'yakonov。

目录

1 引言.2 2 问题陈述 .2 3 校准评估 .33.1 可视化.3

3.2 校准指标.4**4 校准方法 .6**

4.1 后处理.6 4.1.1 直方图分箱.6 4.1.2 保序回归.7 4.1.3 Platt 校准的推广.7 4.2 训练期间校准.8 4.2.1 标签平滑.8 4.2.2 焦点损失.9

5 实证研究　实验. **9**

5.1 实验设计.9 5.2 实验结果.10

6 结论 .12

参考文献 .14

附录 .16

A 分类准确率.16

B 基于分箱的指标 .18 C 适当评分规则 .24

1 引言

深度学习在各个领域的应用日益广泛。神经网络已积极应用于临床实践 [1], 自动驾驶汽车 [2], 机器翻译 [3], 以及其他各种应用。

在许多情况下，现实世界的问题要求模型不仅能给出正确的预测，还能提供可靠的置信度度量。置信度 是指预测正确的概率估计。例如，如果某个算法以 0.9 的置信度预测一组患者样本健康，我们预期其中 90% 确实健康。具有可靠置信度估计的模型称为校准后的模型。除了解读神经网络的预测结果外，当概率估计被输入后续算法步骤时（例如语言模型的解码策略 [4]），置信度校准也很重要。

现代神经网络往往校准得并不好 [5]。然而，许多其他机器学习算法也会产生有偏的置信度估计 [6, 7]。人们已针对“经典”机器学习提出了各种校准技术，其中一些也已应用于深度学习。

2 问题陈述

考虑一个针对集合 $\mathcal{X}$ 中对象的分类问题，其类别为 $\mathcal{Y} = \{1, \ldots, K\}$。假设我们训练了一个模型 ——一种为每个 $x \in \mathcal{X}$生成分数（置信度）向量 *(confidences)*$\mathbf{a}(x) = (a_1(x), \ldots, a_K(x))$, $\sum_{j=1}^{K} a_j(x) = 1$的算法。接下来，将置信度最高的类别分配给该对象：

$$\hat{y}(x) := \operatorname*{argmax}_{j \in \mathcal{Y}} a_j, \quad \hat{p}(x) := a_{\hat{y}}.$$

我们希望将估计器 $\hat{p}$理解为真实标签与预测标签一致的概率。如果估计足够准确，那么该模型就称为校准后的。形式上，校准（ [5]中的完美校准）的定义可写为：

$$\mathbb{P}\left(y = \hat{y} \mid \hat{p} = p\right) = p \quad \forall p \in [0, 1]\,. \tag{1}$$

模型校准还有比[1] （1）更严格的定义。例如，根据[8], , 如果校准良好（原文如此），则称该分类器为校准后的

$$\mathbb{P}(y = j \mid a_j = p) = p \quad \forall j \in \mathcal{Y}, \quad \forall p \in [0, 1] \tag{2}$$

这意味着 对每个类别（而不仅仅是预测类别）给出的置信度都经过校准 .

对于真实世界数据和模型，我们无法直接检查（1）和（2），因此各种校准指标和可视化也很有用，相关内容将在第 3 节中介绍。

[1]需要注意的是，术语校准也常指能够提高模型置信度准确性的方法 。

第 4 节介绍校准方法——让置信度更加可靠的技术。首先，可以事后校准置信度，也就是找到一种将有偏估计映射为校准后的估计的变换。存在多种算法可用于找到最优变换。其次，也可以在训练过程中应用特殊技术，例如修改损失函数。

第 5 节对现代神经网络架构的校准方法进行了实证比较，并展示了损失函数的选择如何影响校准。

3 校准评估

3.1 可视化

在定义校准度量之前，我们先将问题简化为二元分类：$\mathcal{Y} = \{0,1\}$ ——设模型生成置信度 $\hat{p}$ ，表示对象属于正类 $(y = 1)$。二元分类常见于各种应用以及“经典”机器学习模型中，例如逻辑回归、支持向量机、梯度提升树等——相关模型的校准问题已在 [6, 7] 中进行研究。

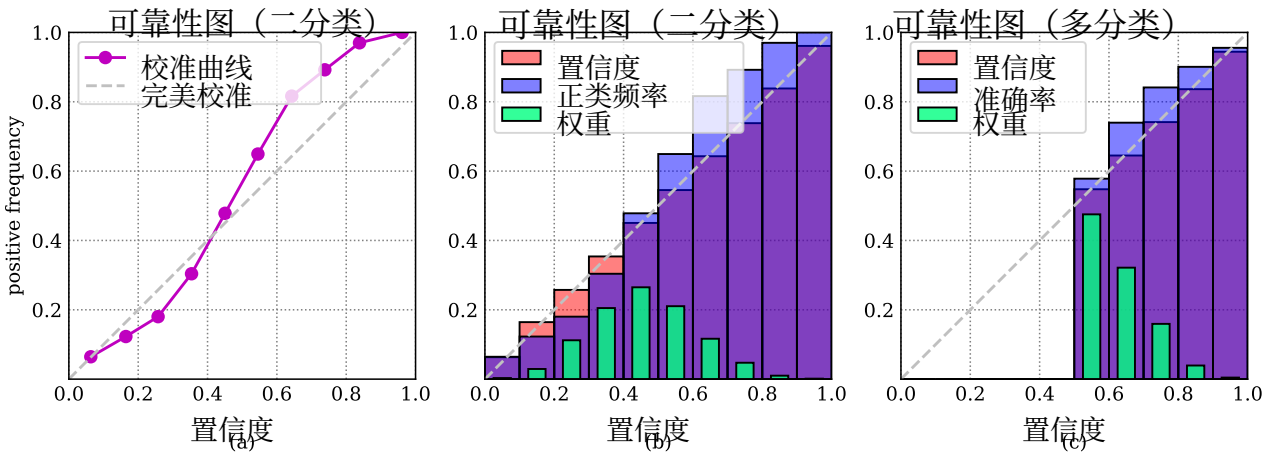


图*1*：可视化置信度校准的选项。为清晰起见，生成了合成数据。使用支持向量机作为模型（到分离超平面的距离经过缩放，映射到 $[0, 1]$）。

考虑一组对象（通常是另一个验证集），已知每个对象的真实类别，且模型会为其生成置信度分数。将所有可能的置信度值组成的区间 $[0, 1]$ 划分为 M 个等宽区间 I_m：

$$I_1 = \left[0, \frac{1}{M}\right), \; I_2 = \left[\frac{1}{M}, \frac{2}{M}\right), \; \ldots, \; I_{M-1} = \left[\frac{M-2}{M}, \frac{M-1}{M}\right), \; I_M = \left[\frac{M-1}{M}, 1\right]. \tag{3}$$

因此，每个置信度估计都落入其中一个区间——用 B_m 表示置信度位于 I_m中的对象的索引集合。 B_m 和 I_m 都称为分箱。

对于每个分箱 B_m，我们计算正类频率 A_m^1 和平均置信度 C_m^1：

$$A_m^1 = \frac{1}{|B_m|} \sum_{i \in B_m} \mathbb{1}(y_i = 1), \quad C_m^1 = \frac{1}{|B_m|} \sum_{i \in B_m} \hat{p}_i. \tag{4}$$

最后，我们可以绘制一张图 $(C_m^1,\ A_m^1)_{m=1}^M$，称为可靠性图图1（a）。此外，得到的曲线有时也称为校准曲线。如果模型的校准曲线接近对角线，则认为模型已校准良好（在小节 *3.2*中，我们将介绍用于衡量这种接近程度的标量指标）。

同样，$(C_m^1,\ A_m^1)_{m=1}^M$ 可以用直方图表示，这种图称为可靠性图。在图1(b) 中，平均置信度以红色表示，正类频率以蓝色表示。如果红色柱高于蓝色柱，则说明算法低估了置信度，即置信不足。相反的情况称为过度自信。为了便于解读，图中还可以显示落入该区间的对象所占比例（区间权重）。

当类别数量 $n > 2$时，可靠性图的绘制方式有所不同。最常见的方法对应于定义（1）。对于每个区间 B_m，我们估计“准确率”A_m 和平均置信度 C_m：

$$A_m = \frac{1}{|B_m|} \sum_{i \in B_m} \mathbb{1}(y_i = \hat{y}_i), \quad C_m = \frac{1}{|B_m|} \sum_{i \in B_m} \hat{p}_i. \tag{5}$$

（4）与（5）的区别在于，在多分类情形中，$\hat{}y_i$ 和$\hat{}p_i$ 分别对应于估计的类别及其置信度；而在二分类情形中，统计量仅针对正类计算。

注意，A_m 和 C_m 分别估计（1）的左侧和右侧。它们可以绘制在可靠性图上。对于两个类别，这种方法如图1 (c)所示。0.5 左侧的分箱最终为空，因为二元分类算法会将对象分配给置信度为 > 0.5 的类别。

还可以考虑逐类别可靠性图 [9]。绘制这类图时，应将每个类别分别指定为正类，并将所有其他类别视为负类，从而 n 可以绘制二分类情形下的可靠性图。尽管逐类别方法提供的信息更多（2），但当类别数量很大时（例如，ImageNet 中有 1’000 或 22’000 个类别 [10]），由于难以解读，通常不太实用。因此，仅考虑预测置信度的可靠性图（5）更为常见。

3.2 校准指标

除了可视化之外，还可以使用各种指标[2] 来评估模型的校准效果。最常用的指标之一是期望校准误差（ECE） [11]。它估算置信度与相应准确率之间绝对差值的期望：

$$\mathbb{E}_{\hat{p}} \left| \mathbb{P}\left(y = \hat{y} \mid \hat{p}\right) - \hat{p} \right|. \tag{6}$$

[2]这里，指标指用于评估算法的度量

(6) 可通过将置信度划分为多个分箱来近似（l 是所考虑集合中的对象总数）：

$$\begin{aligned}\mathrm{ECE} &= \sum_{m=1}^{M} \frac{|B_m|}{n} |A_m - C_m| \\ &= \sum_{m=1}^{M} \frac{|B_m|}{n} \left| \frac{1}{|B_m|} \sum_{i \in B_m} \mathbb{1}(y_i = \hat{y}_i) - \frac{1}{|B_m|} \sum_{i \in B_m} \hat{p}_i \right| \\ &= \frac{1}{n} \sum_{m=1}^{M} \left| \sum_{i \in B_m} \mathbb{1}(y_i = \hat{y}_i) - \sum_{i \in B_m} \hat{p}_i \right|.\end{aligned} \tag{7}$$

比较多分类问题的 (7) 和可靠性图，我们发现，ECE 恰好等于红色柱与蓝色柱之间差距的加权平均值图1。

还有一些基于将置信度划分为分箱的其他指标，不过较少使用。例如，可以计算置信度与准确率之间的最大差距[11]:

$$\mathrm{MCE} = \max_m |A_m - C_m|, \tag{8}$$

也可以不仅考虑预测类别的置信度，还考虑其他所有类别的置信度（逐类别期望校准误差）[9]:

$$\mathrm{cwECE} = \frac{1}{K} \sum_{j=1}^{K} \sum_{m=1}^{M} \frac{|B_m^j|}{n} |A_m^j - C_m^j|, \tag{9}$$

其中，B_m^j，A_m^j，C_m^j 分别表示第 m 个分箱的准确率和置信度；若将第 j 个类别视为正类，其余类别合并为负类。该指标对应逐类别可靠性图。

除了等宽分箱 (3)，还可以使用样本数量相同的分箱——有时置信度图就是这样构建的。在 [12], 中，作者提出使用等频分箱来计算上述指标。下文将考虑等宽方案。此外，除了 l_1-范数（即对模块取平均）之外，我们还可以使用 l_2 [13]。

分箱指标的问题在于其表现取决于分箱的数量。另一种方法是使用适当评分规则。我们考虑负对数似然 (NLL):

$$\mathrm{NLL} = -\frac{1}{l} \sum_{i=1}^{n} \log a_{i,y_i} \tag{10}$$

其中 y_i 是真实类别标签，$a_{i,\ y_i}$ 是算法对其的置信度，n 是对象总数。

另一种可用于评估模型校准的适当评分规则是布里尔分数:

$$\mathrm{BS} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{K} \left(a_{ij} - \mathbb{1}(y_i = j)\right)^2, \tag{11}$$

其中 K 是类别数。

4 校准方法

校准技术主要有两种。第一种是对模型输出进行后处理。特殊的变换，即校准映射，会将有偏的概率估计映射为校准后的概率估计。第二种是将校准融入模型训练本身。

4.1 后处理

这种变换通常在留出集 （校准集） $(x_i,\ y_i)_{i=1}^{n}$ 上确定。它可以使用与超参数调优相同的数据集，但不能使用训练集，因为模型在训练数据上的输出分布与未见数据上的不同。

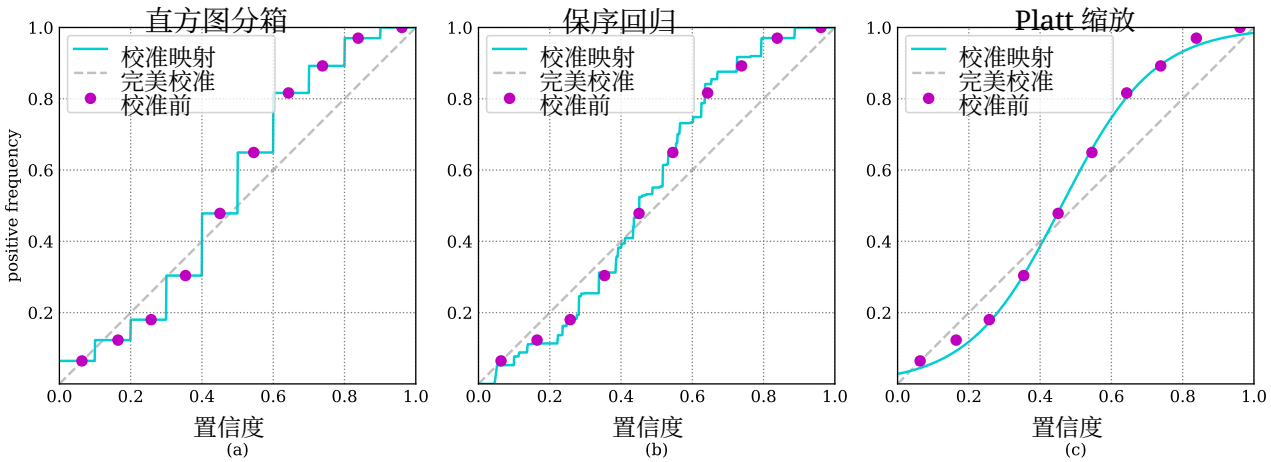


图2：二元分类的不同校准映射（数据和模型与图1相同）。

4.1.1 直方图分箱

该方法最初在 [14] 中提出，用于校准决策树和朴素贝叶斯分类器。直方图分箱中的校准映射是分段常数函数。考虑二分类情形：将输出置信度值集合划分为若干分箱 $B_1,\ \ldots\ldots,\ B_M$ （通常采用等宽(3) 或等频分箱），落入 B_m 的分数将替换为相同的 θ_m。为求得$\theta_1,\ \ldots\ldots,\ \theta_M$，需要求解以下优化问题：

$$\sum_{m=1}^{M}\sum_{i\in B_m}(\theta_m - y_i)^2 \ \rightarrow\ \min_{\theta_1,\ldots,\theta_M}. \tag{12}$$

在此类表达式中，最优的 θ_m 等于落入B_m的正类样本所占比例。校准映射如图2（a）所示。

该方法采用一对其余策略推广到多分类情形：将每个类别分别视为正类，并构造 K 分段常数函数。在推理阶段，还会对校准向量进行归一化。

4.1.2 保序回归

该方法由 [8]提出。对于二分类情形，映射仍为分段常数函数，但区间数量 M 及其边界均经过优化。约束条件是映射应为非递减函数。因此，求解以下问题：

$$\sum_{m=1}^{M}\sum_{i\in\tilde{B}_m}(\theta_m - y_i)^2 \;\to\; \min_{\substack{M\\ \theta_1\leqslant\cdots\leqslant\theta_M\\ 0=\alpha_0\leqslant\alpha_1\leqslant\cdots\leqslant\alpha_{M-1}\leqslant\alpha_M=1}}, \tag{13}$$

其中 $\tilde{B}_1=\{i:\ \alpha_0\leqslant\hat{}p_i<\alpha_1\}$，……，$\tilde{B}_m=\{i:\ \alpha_{m-1}\leqslant\hat{}p_i\leqslant\alpha_m\}$。函数的形状如图2所示。

保序回归以与直方图分箱相同的方式推广到多分类情形。

4.1.3 Platt 校准的推广

最初，该方法被提出用于 [15] 支持向量机的校准。如图1和图2所示，若将对象到分离超平面的距离 $r(x)$ 缩放为 $[0,1]$，并将其视为正类的置信度，则可靠性图将呈现 S 形函数：

$$\hat{p}(x)=\frac{1}{1+e^{-(\alpha\cdot r(x)+\beta)}}. \tag{14}$$

尺度参数 α 和位置参数 β 在校准集上通过最大似然估计进行优化。在此方法中，变换是连续的，并且支持多种推广到多分类情形的方式。

物体 x 的神经网络最后一个线性层输出逻辑值向量：$\mathbf{z}=(z_1,\ldots,z_K)$。为估计类别概率，可使用 Softmax $\sigma(\cdot)$ 对该向量进行变换：

$$\sigma(\mathbf{z})=\frac{1}{\sum_{j=1}^{K}\exp(z_j)}\left(\exp(z_1),\ldots,\exp(z_K)\right),$$

因此，可以通过为逻辑值引入缩放参数和位置参数，将 Platt 校准推广为：

$$a(x)=\sigma(\mathbf{W}\cdot\mathbf{z}+\mathbf{b}). \tag{15}$$

参数 $\mathbf{W}$ 和 $\mathbf{b}$ 也通过在校准集上进行最大似然估计来优化，这等价于最小化负对数似然（10）。根据 $\mathbf{W}$ 和 $\mathbf{b}$ 的形状，可以定义不同的推广形式：

1. 温度缩放：

$$\mathbf{W}=\frac{1}{T}\in\mathbb{R},\ T>0,\ \mathbf{b}=\mathbf{0}.$$

使用单个标量参数对 Platt 校准进行泛化。这是最常用的方法之一。温度升高 T 会增加不确定性，即输出分布的熵增大。相反，降低 T，

会提高对预测类别的置信度。同时，预测类别保持不变。

2. 向量缩放：

$$\mathbf{W} = \mathrm{diag}(\mathbf{v}) \in \mathbb{R}^{K\times K} \text{ — diagonal matrix}, \mathbf{v} \in \mathbb{R}^{K}.$$

在这种方法中，会为每个类别分别优化一个缩放因子（如果偏置 $\mathbf{b} \neq \mathbf{0}$也参与优化，则还会优化偏置）。

3. 矩阵缩放：

$$\mathbf{W} \in \mathbb{R}^{K\times K}, \mathbf{b} \in \mathbb{R}^{K}.$$

矩阵缩放是这组方法中最通用的参数化方式，相当于在 logit 空间中进行逻辑回归。不过，当类别数量较多时，这种方法的参数过多，可能导致过拟合，因为校准集通常不大。

请注意，要实现这些方法中的任意一种，只需在冻结的神经网络上添加一个线性层（维度符合要求）即可。

4.2 训练期间的校准

神经网络的性能很大程度上取决于损失函数。通常使用负对数似然（10）。给定对象 x，它等于真实分类独热分布$\mathbf{y}$ 与预测分布之间的交叉熵：

$$\mathrm{CE}(\mathbf{y}, \mathbf{a}) = -\sum_{j=1}^{K} y_j \log a_j. \tag{16}$$

为了提高模型的校准效果，可以直接修改损失函数。

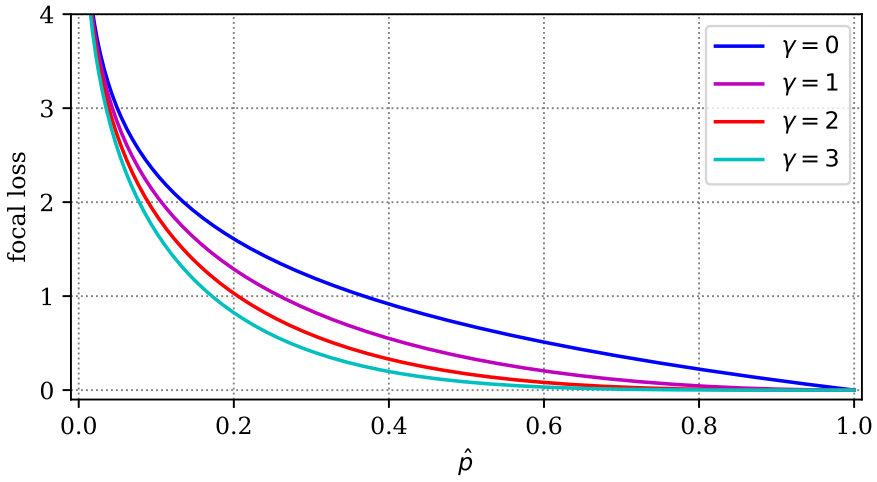
4.2.1 标签平滑

在这种方法中，目标的退化分布会被平滑后的分布取代。可以使用参数 $\alpha \in [0, 1]$ 调节平滑程度：

$$\mathbf{y} = (y_1, \ldots, y_K) \mapsto \left((1-\alpha)y_1 + \frac{\alpha}{K}, \ldots, (1-\alpha)y_K + \frac{\alpha}{K}\right) = \mathbf{y}'. \tag{17}$$

随着参数 α 增大，$\mathbf{y}'$ 的分布会变得更加均匀。经过这一变换后，交叉熵($\mathbf{y}'$，$\mathbf{a}$) 即平滑后的分类向量与预测分布之间的交叉熵会被最小化。

虽然使用平滑标签训练分类器并非新想法，但这种方法最初是为校准而提出的 [16]。



*Fi*图 *3*：单个对象的焦点损失。$\hat{}p$— 对真实类别的概率估计

4.2.2 焦点损失

焦点损失最初是为解决类别不平衡问题而提出的 [17]。在置信度校准方面，这一想法最早应用于 [18]。对于属于j类的对象，焦点损失的定义如下：

$$\mathrm{FL} = -(1 - a_j)^{\gamma} \cdot \log a_j, \quad \gamma \geqslant 0. \tag{18}$$

请注意，当 $\gamma = 0$ 时，损失函数就变为交叉熵。增大参数γ，如图 3所示，会降低对那些对真实类别已经具有较高置信度的对象的惩罚。交叉熵是真实$\mathbf{y}$ 分布与预测$\mathbf{a}$ 分布之间 Kullback–Leibler 散度的上界，而焦点误差则从估计值中减去预测分布的熵 $H(\mathbf{a})$[18]：

$$\mathrm{CE}(\mathbf{y}, \mathbf{a}) \geqslant \mathrm{KL}(\mathbf{y}||\mathbf{a}), \qquad \mathrm{FL}(\mathbf{y}, \mathbf{a}) \geqslant \mathrm{KL}(\mathbf{y}||\mathbf{a}) - \gamma \cdot \mathrm{H}(\mathbf{a}).$$

因此，优化焦点误差还会增加预测分布的熵，并有助于解决过度自信问题。

5 实证实验

5.1 实验设计

实验中使用了以下数据集：

- **CIFAR-10** [19]：该数据集由 60 000 张彩色图像组成 32×32。训练 / 验证集/ 测试集 划分分别为 50 000 / 5 000 / 5 000。
- **CIFAR-100** [19]：60 000 张彩色图像 32×32，100 个类别。训练/验证集/测试集：50 000/ 5 000/ 5 000。

- **ImageNet 2012** [10]：大型数据集，包含按 1000 个类别组织的彩色图像。训练/验证集/测试集：1.2 m/ 25 000/ 25 000。
- **Tiny ImageNet** [10]：110 000 张彩色图像 64×64，划分为 200 个类别，是前一个数据集的子集。训练 / 验证集 / 测试集：100 000 / 5 000 / 5 000。

计算使用了来自开放仓库的各种架构的预训练神经网络。在实验中，模型和数据集分为两个主要组：

1. 第一组包括在 CIFAR-10、CIFAR-100 和 ImageNet 上训练的神经网络。模型的权重分别取自仓库 [20, 21, 22]。该组中的模型用于比较基于后处理的校准方法。

2. 第二组包括来自仓库 [23] 的预训练神经网络。此处使用的数据集为 CIFAR-10、CIFAR-100 和 Tiny ImageNet。这些神经网络是在 [18] 中训练的——所研究的模型采用了焦点误差和标签平滑。

模型在训练 数据上（或其子集）进行训练，在验证集上进行校准——所有图表和指标均对应测试集 。

所有实验使用的代码已发布在 [24]：温度缩放、向量缩放和矩阵缩放使用 PyTorch 实现，其他方法和评估使用 SciPY 和 scikit-learn 实现。直方图分箱使用 20 个分箱；交叉熵、类别加权期望校准误差和最大校准误差基于 15 个分箱的划分计算；可靠性图使用 10 个分箱。

5.2 实验结果

包含测量结果的完整表格见附录；本研究所考虑的所有模型的其他可靠性图可在配套代码库中找到 [24]。

考察 ShuffleNetV2（CIFAR-100，图 4）的可靠性图：我们看到现代神经网络校准的“典型”情况——过度自信。校准方法有助于纠正这种情况：在本例中，温度缩放在所有指标上效果最佳。然而，当类别数量较多时，直方图分箱会过于激进地改变概率，这可以从权重的变化中看出。

相反，对于类别数量较少的情况，直方图分箱在预测置信度方面效果最佳（表5，表6 ——在 CIFAR-10 上几乎适用于所有模型）。请注意，此时神经网络已经能够以非常高的准确率解决分类问题。预测类别的几乎所有概率都接近 1，例如图 5所示。从最大校准误差来看（表9，表10）——该指标不考虑分箱的权重——直方图分箱会导致校准效果较差。

对于矩阵缩放，我们不提供可靠性图：当类别数量较多时，该方法会过度拟合。因此，除了类别较少的 CIFAR-10 外，矩阵缩放都会显著降低所有数据集的分类质量（表 3，表 4）。

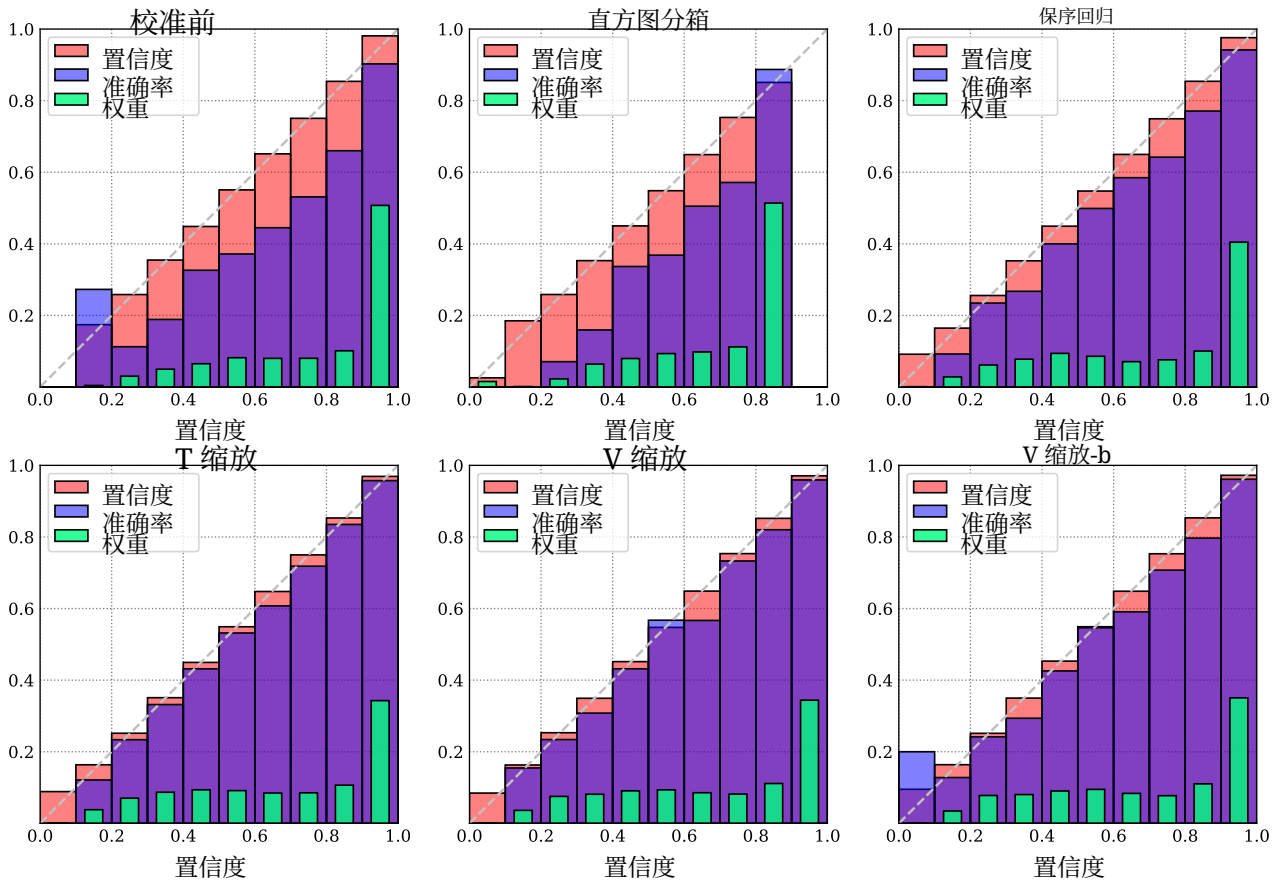


图 *4*：*CIFAR-100*，*ShuffleNetV2_x0_5*

神经网络校准最常用的方法之一是温度缩放。该方法不会影响分类预测，而其他校准选项几乎总会降低准确率（表 3，表 4）。

就负对数似然而言，温度缩放和向量缩放的表现最好，这与预期一致，因为校准过程中最小化的就是这一损失： 表 11、 表 12。就布里尔分数而言，许多情况下表现最好的校准方法是保序回归： 表 13、表 14。

数据集	模型	CE	FL 1	FL 2	FL 3	LS 0.05
CIFAR-10	DenseNet121	4.53	3.47	2.02	1.68	**1.65**
CIFAR-10	ResNet110	4.73	3.70	2.78	**1.61**	2.20
CIFAR-10	ResNet50	4.26	3.88	2.55	**1.58**	3.07
CIFAR-10	Wide-ResNet-26-10	3.25	2.66	**1.57**	1.98	4.33
CIFAR-100	DenseNet121	20.90	14.54	8.40	**4.49**	13.27
CIFAR-100	ResNet110	19.76	15.35	12.10	**9.22**	11.44
CIFAR-100	ResNet50	18.14	13.36	8.60	**4.99**	8.15
CIFAR-100	Wide-ResNet-26-10	16.28	9.12	4.22	**2.20**	5.27
TinyImageNet	ResNet50	15.98	7.87	3.32	**1.93**	15.73

表 *1*：*ECE*，*%* – 期望校准误差误差（越低越好），未进行后处理时的 *15*个分箱，各列对应不同的损失函数

数据集	模型	CE	FL 1	FL 2	FL 3	LS 0.05
CIFAR-10	DenseNet121	0.948	0.755	**0.514**	0.524	0.576
CIFAR-10	ResNet110	0.990	0.804	0.660	**0.505**	0.673
CIFAR-10	ResNet50	0.941	0.836	0.625	**0.524**	0.766
CIFAR-10	Wide-ResNet-26-10	0.699	0.611	**0.479**	0.523	0.869
CIFAR-100	DenseNet121	0.458	0.364	0.280	**0.254**	0.315
CIFAR-100	ResNet110	0.433	0.372	0.321	**0.281**	0.299
CIFAR-100	ResNet50	0.412	0.337	0.282	**0.256**	0.271
CIFAR-100	Wide-ResNet-26-10	0.372	0.264	**0.218**	0.226	0.239
TinyImageNet	ResNet50	0.250	0.218	0.205	**0.203**	0.231

表 *2*：类别加权期望校准误差（*cwECE*，*%*）——类别加权期望校准误差（越低越好），未进行后处理的 *15*个分箱，各列对应不同的损失函数

焦点损失和标签平滑比优化交叉熵的模型能得到校准效果更好的模型——无论是从预测类别的校准情况（表 1）还是逐类别

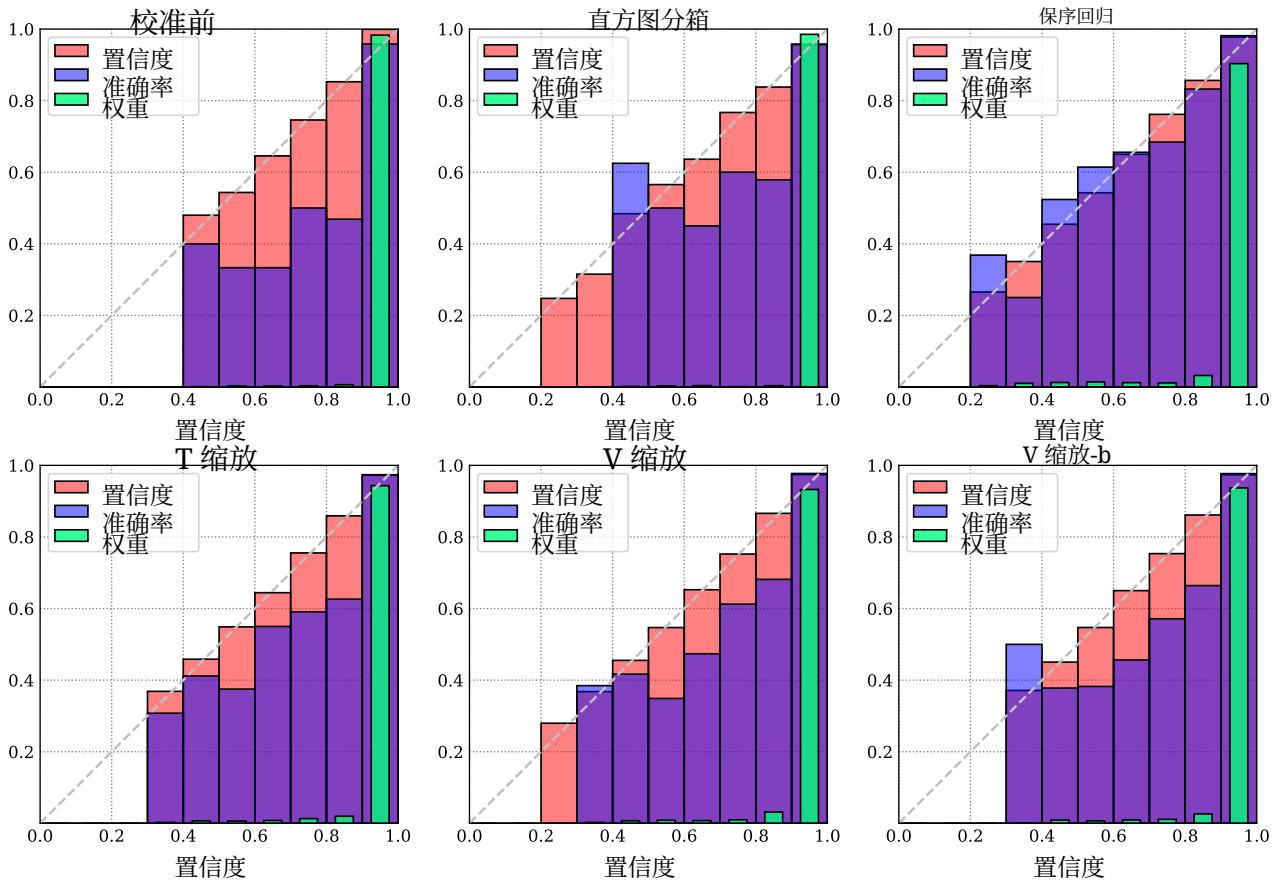


图 *5*：*CIFAR-10*，*DenseNet121*（交叉熵）

估计（表 2）来看。同时，这类模型还可以通过后处理进一步校准。在原始论文中，[18]，使用温度缩放来校准通过焦点损失训练的模型。尽管这种方法在期望校准误差（ECE）方面效果不错（表 6），但向量缩放在逐类别 cwECE 方面表现更好（表 8）。对于第一组模型（表 7），向量缩放基本上也能将 cwECE 降至最低。这些结果并不意外，因为向量缩放会为每个类别分别求得系数。

另请看 EfficientNet 的可靠性图（图 6）。在所有使用的模型中，这种神经网络的置信不足现象最为明显：大多数预测结果落在 [0。9, 1], 而非 [0。8，0。9）。这种现象可能是由训练设置造成的：该模型使用标签平滑（$\alpha = 0。1$）进行训练。[22]所有校准方法都使答案的置信度显著提高。

6 结论

本研究比较了不同神经网络架构、数据集和标准下的主要置信度校准方法。

某种方法是否适用，很大程度上取决于数据量和所选的评估标准。对于每个类别分别求得校准映射的算法，只有在校准集中的数据量充足时才能发挥良好效果（通常在类别数量较少时才能满足这一条件）。基于线性

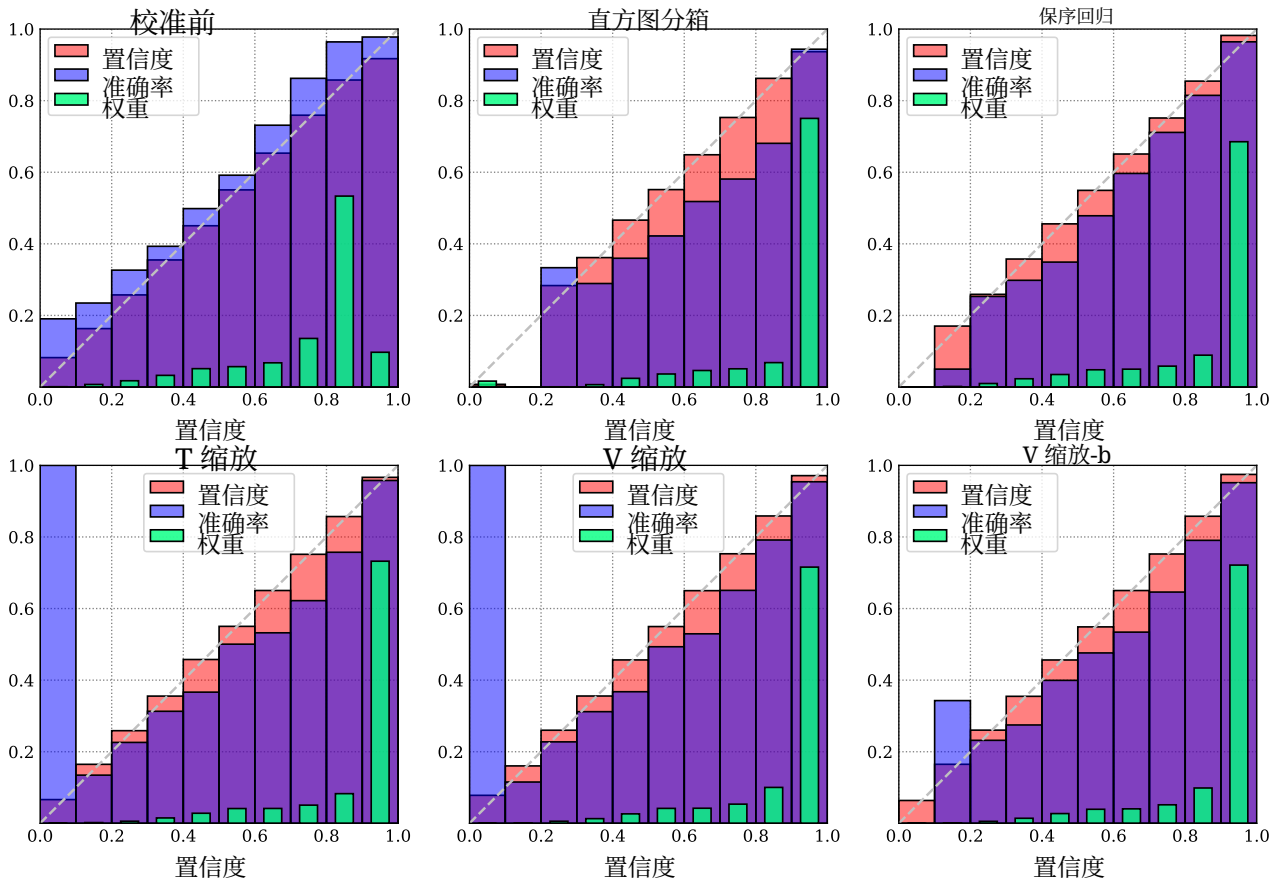


图 6：*ImageNet*，*EfficientNet_b8*

逻辑值变换的策略（例如温度缩放）在类别数量较多的问题中表现出色，但参数化过度时容易出现过拟合（矩阵缩放）。

置信度校准在机器学习领域仍有待进一步研究：正如本文所示，不仅校准方法的选择会导致不同结果，评估标准的选择也会如此。

参考文献

1. Anwar S. M. [等]. 使用卷积神经网络进行医学图像分析：综述 // 医学系统杂志. — 2018年. — Vol. 42, no. 11. — P. 1–13. 2. Grigorescu S., Trasnea B., Cocias T., Macesanu G. 自动驾驶中的深度学习技术综述 // 野外机器人学杂志. — 2020年. — Vol. 37,no. 3. — P. 362–386. 3. Wu Y. [等]. 谷歌神经机器翻译系统：弥合人工翻译与机器翻译之间的差距 // arXiv 预印本：1609.08144. — 2016年. 4. Holtzman A. [等]. 神经文本退化的奇特案例 // arXiv 预印本 arXiv:1904.09751. — 2019年. 5. Guo C., Pleiss G., Sun Y., Weinberger K. Q. 现代神经网络的校准 // 国际机器学习大会. — PMLR. 2017年. — P. 1321–1330. 6. Niculescu-Mizil A., Caruana R. 使用监督学习预测良好的概率 // 第22届国际机器学习大会论文集. — 2005年. — P. 625–632. 7. Caruana R., Niculescu-Mizil A. 监督学习算法的实证比较 // 第23届国际机器学习大会论文集. — 2006年. — P. 161–168. 8. Zadrozny B., Elkan C. 将分类器得分转换为准确的多分类概率估计 // 第八届 ACM 知识发现与数据挖掘国际会议论文集. — 2002年. — P. 694–699. 9. Kull M. [等]. 超越温度缩放：通过狄利克雷校准获得校准良好的多分类概率 // 神经信息处理系统进展. Vol. 32 / H. Wallach [等]编. — Curran Associates, Inc., 2019年. 10. Deng J. [等]. ImageNet：大型分层图像数据库 // 2009年 IEEE 计算机视觉与模式识别会议. — IEEE. 2009年. — P. 248–255. 11. Naeini M. P., Cooper G., Hauskrecht M. 使用贝叶斯分箱获得校准良好的概率 // AAAI 人工智能大会论文集. Vol. 29. — 2015年. 12. Nixon J. [等]. 衡量深度学习中的校准 // arXiv 预印本：1904.01685. — 2020年. 13. Kumar A., Liang P., Ma T. 经验证的不确定性校准 // 神经信息处理系统进展（NeurIPS）. — 2019年. 14. Zadrozny B., Elkan C. 从决策树和朴素贝叶斯分类器获得校准后的概率估计 // ICML. Vol. 1. — Citeseer. 2001年. — P. 609–616. 15. Platt J. [等]. 支持向量机的概率输出及其与正则化似然方法的比较 //. Vol. 10. — Cambridge, MA, 1999年. — P. 61–74. 16. Müller R., Kornblith S., Hinton G. 标签平滑何时有帮助？ //. — 2019年.

17. Lin T.-Y. [等。用于密集目标检测的焦点损失 // IEEE计算机视觉国际会议论文集。— 2017年。— 第2980–2988页。18. Mukhoti J. [等。使用焦点损失校准深度神经网络。— 2020年。19. Krizhevsky A., Hinton G. 从微小图像中学习多层特征 // 硕士论文，计算机科学系，多伦多大学。— 2009年。20. Phan H. huyvnphan/PyTorch_CIFAR10。— 版本v3.0.1。— 2021年1月。— URL：https: //doi.org/10.5281/zenodo.4431043。21. chenyaofo。PyTorch CIFAR模型。— 2021年。— URL：https://github.com/chenyaofo/pytorch-cifar-models。22. Wightman R. PyTorch图像模型。— 2019年。— URL：https://github.com/rwightman/pytorch-image-models。23. Mukhoti J., Kulharia V. [18]的代码。— URL：https : // github . com / torrvision / focal_calibration。24. Vasilev R. 校准（配套代码库）。— URL：https : / / github . com / artnitolog/calibration。

附录

A 分类准确率

数据集	模型	校准前	直方图分箱	保序回归	T 缩放	V 缩放	V 缩放-b	M 缩放-b
CIFAR-10	DenseNet121	**93.92**	93.52	93.72	**93.92**	93.84	93.80	93.80
CIFAR-10	DenseNet161	93.70	**93.74**	93.58	93.70	93.72	**93.74**	93.52
CIFAR-10	DenseNet169	**94.08**	93.44	93.84	**94.08**	94.02	93.94	93.78
CIFAR-10	GoogleNet	**92.92**	92.58	92.68	**92.92**	92.90	92.84	92.76
CIFAR-10	InceptionV3	**93.32**	93.12	93.28	**93.32**	93.22	**93.32**	93.26
CIFAR-10	MobileNetV2	**93.42**	93.40	**93.42**	**93.42**	93.26	93.36	93.36
CIFAR-10	ResNet18	**92.54**	92.14	92.34	**92.54**	92.48	92.40	92.18
CIFAR-10	ResNet34	**93.24**	92.74	92.92	**93.24**	93.14	93.16	92.94
CIFAR-10	ResNet50	**93.44**	93.10	93.22	**93.44**	93.40	93.38	93.24
CIFAR-10	VGG11 _bn	91.96	91.66	91.84	91.96	91.86	91.80	**91.98**
CIFAR-10	VGG13 _bn	93.86	93.40	93.74	93.86	**93.92**	93.82	93.68
CIFAR-10	VGG16 _bn	**93.52**	93.38	93.42	**93.52**	93.36	93.36	93.36
CIFAR-10	VGG19 _bn	93.76	93.36	93.62	93.76	93.70	**93.84**	93.56
CIFAR-100	MobileNetV2 _x0_5	**70.32**	67.56	69.86	**70.32**	70.26	69.94	55.38
CIFAR-100	MobileNetV2 _x1_0	73.34	70.74	72.98	73.34	73.20	**73.50**	59.24
CIFAR-100	MobileNetV2 _x1_4	**76.22**	72.96	75.58	**76.22**	75.72	75.80	63.10
CIFAR-100	ResNet20	**67.80**	63.80	66.72	**67.80**	67.36	67.54	49.94
CIFAR-100	ResNet32	**69.10**	65.90	68.64	**69.10**	68.90	68.70	52.20
CIFAR-100	ResNet44	**70.82**	67.86	70.22	**70.82**	70.42	70.56	55.16
CIFAR-100	ResNet56	**72.04**	69.54	71.62	**72.04**	71.72	71.58	56.34
CIFAR-100	ShuffleNetV2 _x0_5	66.86	64.22	66.68	66.86	**66.98**	66.58	50.02
CIFAR-100	ShuffleNetV2 _x1_0	**71.58**	69.16	71.24	**71.58**	71.26	71.30	57.50
CIFAR-100	ShuffleNetV2 _x1_5	73.60	70.90	73.34	73.60	73.72	**73.80**	61.30
CIFAR-100	ShuffleNetV2 _x2_0	**74.92**	72.30	74.44	**74.92**	74.84	74.82	63.04
CIFAR-100	VGG11 _bn	**69.96**	67.76	69.34	**69.96**	69.56	69.58	59.90
CIFAR-100	VGG13 _bn	**73.90**	71.84	73.02	**73.90**	73.56	73.36	62.58
CIFAR-100	VGG16 _bn	**73.30**	71.34	72.78	**73.30**	72.96	72.94	64.28
CIFAR-100	VGG19 _bn	**73.14**	71.78	72.76	**73.14**	73.00	72.82	63.78
ImageNet	EfficientNet _b8	85.52	83.43	84.92	85.52	**85.54**	85.47	79.12
ImageNet	MobileNetV2 _120d	**77.68**	73.57	76.79	**77.68**	77.36	77.16	59.84
ImageNet	RepVGG_b3	**80.57**	77.68	79.89	**80.57**	80.22	80.17	66.38
ImageNet	VGG19 _bn	**74.35**	70.47	73.72	**74.35**	73.92	73.62	54.06

表 *3*：准确率 %（越高越好）– 正确预测的比例，第*1*组

数据集	模型	校准前	直方图分箱	保序回归	T 缩放	V 缩放	V 缩放-b	M 缩放-b
CIFAR-10	DenseNet121（交叉熵）	94.92	**94.98**	94.84	94.92	94.94	94.96	94.76
CIFAR-10	DenseNet121 (FL 1)	94.84	94.68	94.82	94.84	95.00	**95.02**	94.52
CIFAR-10	DenseNet121 (FL 2)	94.90	94.60	**94.96**	94.90	94.84	94.80	94.80
CIFAR-10	DenseNet121 (FL 3)	94.24	94.24	94.20	94.24	94.34	94.24	**94.36**
CIFAR-10	DenseNet121 (LS 0.05)	94.52	94.62	**94.66**	94.52	94.56	94.58	94.62
CIFAR-10	ResNet110 (交叉熵)	94.82	94.70	94.76	94.82	94.84	94.86	**94.90**
CIFAR-10	ResNet110 (FL 1)	94.80	94.90	94.84	94.80	94.86	**95.00**	94.98
CIFAR-10	ResNet110 (FL 2)	94.82	94.82	**94.96**	94.82	**94.96**	94.90	94.84
CIFAR-10	ResNet110 (FL 3)	94.70	94.64	94.84	94.70	94.84	**94.86**	94.70
CIFAR-10	ResNet110 (LS 0.05)	94.36	94.36	**94.48**	94.36	94.32	94.32	94.40
CIFAR-10	ResNet50（交叉熵）	**95.02**	94.62	94.66	**95.02**	95.00	94.96	94.82
CIFAR-10	ResNet50（FL 1）	94.58	94.34	94.52	94.58	**94.60**	94.52	94.58
CIFAR-10	ResNet50（FL 2）	94.80	94.62	94.82	94.80	**94.90**	**94.90**	94.88
CIFAR-10	ResNet50（FL 3）	94.52	94.48	**94.66**	94.52	94.58	94.60	94.44
CIFAR-10	ResNet50（LS 0.05）	94.26	94.02	94.18	94.26	94.26	94.16	**94.30**
CIFAR-10	Wide-ResNet-26-10（交叉熵）	96.12	96.02	96.08	96.12	**96.20**	**96.20**	96.10
CIFAR-10	Wide-ResNet-26-10（FL 1）	**95.70**	95.26	95.56	**95.70**	95.64	95.66	**95.70**
CIFAR-10	Wide-ResNet-26-10（FL 2）	**95.56**	95.06	95.50	**95.56**	95.40	95.46	95.48
CIFAR-10	Wide-ResNet-26-10（FL 3）	95.64	95.62	**95.80**	95.64	95.74	95.74	95.64
CIFAR-10	Wide-ResNet-26-10（LS 0.05）	95.66	95.50	95.40	95.66	95.64	**95.68**	95.42
CIFAR-100	DenseNet121（交叉熵）	75.32	74.52	75.26	75.32	**75.76**	75.70	65.60
CIFAR-100	DenseNet121（FL 1）	75.66	74.12	75.90	75.66	**76.02**	76.00	65.24
CIFAR-100	DenseNet121（FL 2）	75.90	73.50	**76.22**	75.90	75.90	76.02	65.32
CIFAR-100	DenseNet121（FL 3）	76.10	72.90	75.86	76.10	76.16	**76.20**	65.06
CIFAR-100	DenseNet121（LS 0.05）	75.54	74.56	75.66	75.54	75.66	**75.74**	66.60
CIFAR-100	ResNet110（交叉熵）	76.44	76.02	76.50	76.44	76.80	**76.88**	66.20
CIFAR-100	ResNet110（FL 1）	76.84	75.28	76.48	76.84	**77.34**	77.28	66.52
CIFAR-100	ResNet110（FL 2）	76.74	74.48	76.68	76.74	**77.04**	76.90	67.00
CIFAR-100	ResNet110 (FL 3)	76.30	74.80	76.56	76.30	76.66	**76.82**	66.46
CIFAR-100	ResNet110 (LS 0.05)	75.96	74.48	75.68	75.96	76.10	**76.18**	66.66
CIFAR-100	ResNet50 (交叉熵)	76.00	74.48	75.62	76.00	76.12	**76.14**	65.86
CIFAR-100	ResNet50（FL 1）	76.38	74.22	76.36	76.38	**76.66**	76.46	66.32
CIFAR-100	ResNet50（FL 2）	**76.58**	73.90	76.10	**76.58**	76.52	76.46	65.40
CIFAR-100	ResNet50（FL 3）	**77.08**	74.44	76.42	**77.08**	76.86	76.94	64.62
CIFAR-100	ResNet50（LS 0.05）	76.18	74.38	76.24	76.18	76.42	**76.48**	67.98
CIFAR-100	Wide-ResNet-26-10（交叉熵）	78.26	77.22	77.98	78.26	**78.38**	78.22	67.72
CIFAR-100	Wide-ResNet-26-10（FL 1）	**79.84**	78.10	79.80	**79.84**	79.76	79.76	68.82
CIFAR-100	Wide-ResNet-26-10（FL 2）	79.64	77.54	79.64	79.64	79.96	**80.04**	69.44
CIFAR-100	Wide-ResNet-26-10（FL 3）	79.70	76.96	79.62	79.70	**79.86**	79.80	70.18
CIFAR-100	Wide-ResNet-26-10（LS 0.05）	78.14	75.98	77.90	78.14	**78.26**	78.24	69.62
TinyImageNet ResNet50 (CE)		**49.98**	44.92	49.10	**49.98**	49.44	49.24	32.66
TinyImageNet	ResNet50（FL 1）	**50.10**	44.96	49.48	**50.10**	50.04	49.94	31.32
TinyImageNet	ResNet50（FL 2）	**51.86**	46.56	51.14	**51.86**	51.82	51.62	31.96
TinyImageNet	ResNet50（FL 3）	**51.06**	44.56	50.18	**51.06**	50.82	50.94	32.08
TinyImageNet	ResNet50（LS 0.05）	**53.62**	47.18	51.94	**53.62**	52.84	52.78	36.82

表 *4*：准确率 %（越高越好）– 正确预测所占比例，组 *2*

B 基于分箱的指标

数据集	模型	校准前	直方图分箱	保序回归	T 缩放	V 缩放	V 缩放-b	M 缩放-b
CIFAR-10	DenseNet121	1.86	**1.08**	2.10	1.64	1.67	1.58	1.49
CIFAR-10	DenseNet161	2.11	**1.08**	1.64	1.68	1.87	1.75	1.75
CIFAR-10	DenseNet169	2.04	**1.42**	1.87	1.72	1.64	1.62	1.64
CIFAR-10	GoogleNet	1.65	**0.89**	1.36	1.21	1.46	1.37	1.34
CIFAR-10	InceptionV3	2.06	**1.35**	1.91	1.88	1.99	1.91	1.86
CIFAR-10	MobileNetV2	2.92	**1.45**	1.74	1.87	2.09	1.95	1.78
CIFAR-10	ResNet18	2.51	**1.60**	2.03	2.13	1.98	1.93	2.16
CIFAR-10	ResNet34	2.67	**1.44**	2.09	1.96	1.70	1.88	1.75
CIFAR-10	ResNet50	2.50	**1.06**	1.70	1.67	1.72	1.59	1.62
CIFAR-10	VGG11_bn	1.87	2.21	**1.65**	1.83	1.88	1.90	1.76
CIFAR-10	VGG13_bn	**1.41**	1.42	1.54	1.44	1.44	1.43	1.72
CIFAR-10	VGG16_bn	1.86	**1.08**	1.61	1.71	1.93	1.81	1.74
CIFAR-10	VGG19_bn	2.15	**1.02**	1.34	1.87	2.00	1.98	2.14
CIFAR-100	MobileNetV2_x0_5	12.69	9.39	5.86	**3.28**	3.31	3.57	44.57
CIFAR-100	MobileNetV2_x1_0	11.77	10.13	6.02	3.78	**3.46**	3.69	40.68
CIFAR-100	MobileNetV2_x1_4	9.66	10.13	4.29	2.90	**2.69**	2.86	36.79
CIFAR-100	ResNet20	11.21	9.01	5.79	**2.48**	3.11	3.13	50.05
CIFAR-100	ResNet32	13.95	10.34	5.47	2.61	**2.28**	2.77	47.79
CIFAR-100	ResNet44	14.99	8.42	6.42	**3.18**	3.49	3.48	44.84
CIFAR-100	ResNet56	14.72	7.86	5.68	3.13	**3.10**	3.40	43.65
CIFAR-100	ShuffleNetV2_x0_5	12.93	9.62	5.38	**2.22**	2.38	2.80	49.77
CIFAR-100	ShuffleNetV2_x1_0	11.68	8.15	5.90	**3.68**	4.24	4.18	42.32
CIFAR-100	ShuffleNetV2_x1_5	9.86	8.94	5.44	**4.56**	4.88	4.64	38.44
CIFAR-100	ShuffleNetV2_x2_0	7.68	10.14	4.91	**4.26**	4.78	4.78	36.80
CIFAR-100	VGG11_bn	15.86	10.61	7.78	**5.17**	5.80	6.08	29.38
CIFAR-100	VGG13_bn	14.06	8.12	7.19	5.84	**5.80**	5.84	27.53
CIFAR-100	VGG16_bn	19.33	8.87	5.74	4.18	**3.93**	4.03	28.60
CIFAR-100	VGG19_bn	20.17	8.79	5.22	4.53	3.74	**3.61**	31.98
ImageNet	EfficientNet_b8	8.98	4.88	**2.96**	3.17	3.51	3.95	20.15
ImageNet	MobileNetV2_120d	6.85	7.02	1.96	**1.61**	2.11	2.90	38.18
ImageNet	RepVGG_b3	**3.23**	6.17	3.61	3.95	4.08	4.76	32.42
ImageNet	VGG19_bn	3.69	9.26	3.93	1.92	**1.53**	2.16	44.68

表*5*：*ECE %* – 期望校准误差（越低越好），*15*个分箱，第*1*组

数据集	模型	校准前	直方图分箱	保序回归	T 缩放	V 缩放	V 缩放-b	M 缩放-b
CIFAR-10	DenseNet121（交叉熵）	4.53	**0.38**	1.24	1.64	1.19	1.29	1.45
CIFAR-10	DenseNet121 (FL 1)	3.47	0.87	**0.65**	1.26	0.88	1.05	1.44
CIFAR-10	DenseNet121 (FL 2)	2.02	0.97	**0.73**	0.95	1.08	1.01	1.18
CIFAR-10	DenseNet121 (FL 3)	1.68	**1.17**	1.32	1.49	1.32	1.38	1.84
CIFAR-10	DenseNet121 (LS 0.05)	1.65	**0.93**	1.55	1.22	1.28	1.24	1.29
CIFAR-10	ResNet110 (交叉熵)	4.73	**1.01**	1.11	1.23	1.36	1.39	1.25
CIFAR-10	ResNet110 (FL 1)	3.70	**0.71**	0.99	1.11	1.20	1.06	1.05
CIFAR-10	ResNet110（FL 2）	2.78	**0.55**	1.20	1.03	0.94	0.96	1.07
CIFAR-10	ResNet110（FL 3）	1.61	**0.84**	0.96	1.24	0.94	0.84	1.01
CIFAR-10	ResNet110（LS 0.05）	2.20	1.77	0.90	1.56	1.07	0.88	**0.72**
CIFAR-10	ResNet50（交叉熵）	4.26	**0.70**	0.96	1.41	1.17	1.10	1.23
CIFAR-10	ResNet50（FL 1）	3.88	1.88	**1.46**	1.58	1.65	1.67	1.85
CIFAR-10	ResNet50（FL 2）	2.55	**1.08**	1.09	1.17	1.52	1.52	1.14
CIFAR-10	ResNet50（FL 3）	1.58	**0.78**	0.98	1.11	1.16	1.01	1.56
CIFAR-10	ResNet50（LS 0.05）	3.07	1.23	**1.14**	1.35	1.35	1.43	1.40
CIFAR-10	Wide-ResNet-26-10（交叉熵）	3.25	0.60	**0.50**	1.07	0.84	0.91	0.90
CIFAR-10	Wide-ResNet-26-10（FL 1）	2.66	0.95	1.03	**0.87**	1.08	1.17	1.05
CIFAR-10	Wide-ResNet-26-10（FL 2）	1.57	1.40	**1.12**	1.18	1.37	1.31	1.42
CIFAR-10	Wide-ResNet-26-10（FL 3）	1.98	**0.82**	0.89	1.06	1.08	0.86	0.89
CIFAR-10	Wide-ResNet-26-10（LS 0.05）	4.33	**0.72**	0.94	0.99	1.17	1.17	1.23
CIFAR-100	DenseNet121 (交叉熵)	20.90	7.00	6.18	4.82	4.74	**4.72**	34.38
CIFAR-100	DenseNet121 (FL 1)	14.54	6.03	7.31	**5.18**	5.22	5.38	34.71
CIFAR-100	DenseNet121 (FL 2)	8.40	10.42	4.47	**4.16**	4.37	4.42	34.63
CIFAR-100	DenseNet121 (FL 3)	4.49	9.10	4.16	**3.98**	4.55	4.45	34.84
CIFAR-100	DenseNet121 (LS 0.05)	13.27	6.98	7.03	7.44	**3.32**	3.49	33.31
CIFAR-100	ResNet110（交叉熵）	19.76	**3.99**	6.09	5.63	5.37	5.44	33.74
CIFAR-100	ResNet110（FL 1）	15.35	6.16	6.56	4.87	**4.83**	4.84	33.41
CIFAR-100	ResNet110（FL 2）	12.10	8.50	5.92	**5.48**	5.51	5.61	32.91
CIFAR-100	ResNet110 (FL 3)	9.22	8.68	**4.67**	5.36	5.41	5.39	33.53
CIFAR-100	ResNet110 (LS 0.05)	11.44	6.95	5.86	**4.41**	4.50	4.66	33.25
CIFAR-100	ResNet50 (CE)	18.14	**4.75**	7.24	6.02	6.52	6.53	34.07
CIFAR-100	ResNet50（FL 1）	13.36	5.81	7.06	**5.29**	5.44	5.61	33.62
CIFAR-100	ResNet50（FL 2）	8.60	9.41	4.90	**4.72**	4.91	4.90	34.52
CIFAR-100	ResNet50（FL 3）	4.99	7.73	4.07	**3.39**	3.83	3.98	35.30
CIFAR-100	ResNet50 (LS 0.05)	8.15	7.68	6.61	**5.27**	5.32	5.32	31.94
CIFAR-100	Wide-ResNet-26-10 (交叉熵)	16.28	4.91	6.43	**4.79**	5.21	5.00	32.23
CIFAR-100	Wide-ResNet-26-10（FL 1）	9.12	8.12	5.40	4.31	4.35	**4.24**	31.16
CIFAR-100	Wide-ResNet-26-10（FL 2）	4.22	7.61	**3.47**	3.87	4.01	3.75	30.53
CIFAR-100	Wide-ResNet-26-10（FL 3）	**2.20**	8.05	3.33	3.62	3.62	3.79	29.80
CIFAR-100	Wide-ResNet-26-10（LS 0.05）	5.27	7.52	5.36	**4.78**	4.85	4.83	30.37
TinyImageNet	ResNet50（交叉熵）	15.98	9.96	10.06	**6.67**	7.26	7.83	65.77
TinyImageNet	ResNet50（FL 1）	7.87	9.96	6.12	**4.17**	4.76	5.54	67.23
TinyImageNet	ResNet50（FL 2）	3.32	9.49	4.94	**2.98**	3.63	3.82	66.81
TinyImageNet	ResNet50 (FL 3)	1.93	9.88	3.13	**1.77**	2.61	2.70	66.65
TinyImageNet	ResNet50 (LS 0.05)	15.73	6.29	**3.02**	7.02	7.18	7.56	60.64

表 *6*：*ECE %* – 期望校准误差（越低越好），*15*个分箱，组 *2*

数据集	模型	校准前	直方图分箱	保序回归	T 缩放	V 缩放	V 缩放-b	M 缩放-b
CIFAR-10	DenseNet121	0.513	0.509	0.566	0.514	0.548	0.491	**0.483**
CIFAR-10	DenseNet161	0.646	0.536	0.534	0.629	0.531	**0.527**	0.536
CIFAR-10	DenseNet169	0.551	0.566	0.527	0.536	**0.485**	0.490	0.518
CIFAR-10	GoogleNet	0.641	0.579	0.543	0.559	0.544	**0.520**	0.545
CIFAR-10	InceptionV3	0.596	0.573	**0.543**	0.598	0.543	0.573	0.553
CIFAR-10	MobileNetV2	0.638	**0.500**	0.502	0.543	0.572	0.574	0.546
CIFAR-10	ResNet18	0.649	0.578	0.570	0.634	0.590	**0.538**	0.581
CIFAR-10	ResNet34	0.714	**0.549**	0.549	0.671	0.584	0.566	0.570
CIFAR-10	ResNet50	0.596	0.523	0.522	0.555	0.556	**0.511**	0.530
CIFAR-10	VGG11_bn	0.633	0.579	0.517	0.635	0.620	**0.506**	0.523
CIFAR-10	VGG13_bn	0.563	0.536	**0.443**	0.572	0.510	0.488	0.461
CIFAR-10	VGG16_bn	0.542	0.617	0.478	0.548	0.550	**0.477**	0.518
CIFAR-10	VGG19_bn	0.578	0.526	**0.443**	0.514	0.531	0.460	0.495
CIFAR-100	MobileNetV2_x0_5	0.364	0.325	0.289	**0.260**	0.270	0.266	0.893
CIFAR-100	MobileNetV2_x1_0	0.337	0.295	0.266	0.264	0.262	**0.254**	0.815
CIFAR-100	MobileNetV2_x1_4	0.305	0.297	0.254	0.246	0.251	**0.245**	0.737
CIFAR-100	ResNet20	0.357	0.338	0.287	**0.270**	0.274	0.276	1.001
CIFAR-100	ResNet32	0.383	0.309	0.285	**0.262**	0.267	0.274	0.956
CIFAR-100	ResNet44	0.398	0.303	0.283	**0.263**	0.274	0.270	0.897
CIFAR-100	ResNet56	0.391	0.285	0.275	**0.260**	0.272	0.262	0.873
CIFAR-100	ShuffleNetV2_x0_5	0.379	0.312	0.288	**0.261**	0.272	0.276	0.999
CIFAR-100	ShuffleNetV2_x1_0	0.343	0.315	0.281	**0.253**	0.269	0.266	0.849
CIFAR-100	ShuffleNetV2_x1_5	0.302	0.290	0.259	0.259	0.265	**0.255**	0.773
CIFAR-100	ShuffleNetV2_x2_0	0.274	0.291	0.256	0.252	0.261	**0.245**	0.738
CIFAR-100	VGG11_bn	0.399	0.286	0.272	0.271	0.285	**0.262**	0.678
CIFAR-100	VGG13_bn	0.359	0.262	0.272	0.264	0.267	**0.254**	0.639
CIFAR-100	VGG16_bn	0.448	0.244	0.249	0.252	0.259	**0.244**	0.645
CIFAR-100	VGG19_bn	0.462	**0.237**	0.244	0.254	0.266	0.246	0.696
ImageNet	EfficientNet_b8	0.035	0.025	**0.022**	0.023	0.024	0.023	0.042
ImageNet	MobileNetV2_120d	0.036	0.034	0.030	**0.030**	0.030	0.030	0.080
ImageNet	RepVGG_b3	0.028	0.030	**0.027**	0.028	0.028	0.028	0.067
ImageNet	VGG19_bn	0.032	0.036	**0.031**	0.032	0.032	0.032	0.091

表 7：类别加权期望误差 交叉熵 % – 分类期望校准误差（越低越好），*15*个分箱，第*1*组

数据集	模型	校准前	直方图分箱	保序回归	T 缩放	V 缩放	V 缩放-b	M 缩放-b
CIFAR-10	DenseNet121（交叉熵）	0.948	0.478	0.530	0.548	0.485	**0.455**	0.461
CIFAR-10	DenseNet121（FL 1）	0.755	0.487	**0.376**	0.462	0.430	0.416	0.540
CIFAR-10	DenseNet121（FL 2）	0.514	0.488	0.424	0.462	**0.419**	0.431	0.428
CIFAR-10	DenseNet121（FL 3）	0.524	0.489	**0.474**	0.532	0.508	0.483	0.517
CIFAR-10	DenseNet121 (LS 0.05)	0.576	0.376	0.446	0.514	0.410	**0.317**	0.392
CIFAR-10	ResNet110 (交叉熵)	0.990	0.491	0.547	0.550	0.495	**0.449**	0.476
CIFAR-10	ResNet110 (FL 1)	0.804	0.483	**0.404**	0.517	0.511	0.449	0.458
CIFAR-10	ResNet110 (FL 2)	0.660	0.470	0.411	0.486	0.453	**0.408**	0.438
CIFAR-10	ResNet110 (FL 3)	0.505	0.436	**0.400**	0.500	0.417	0.418	0.425
CIFAR-10	ResNet110 (LS 0.05)	0.673	0.497	0.488	0.617	0.505	**0.420**	0.431
CIFAR-10	ResNet50（交叉熵）	0.941	0.469	0.493	0.524	0.454	0.435	**0.428**
CIFAR-10	ResNet50（FL 1）	0.836	0.520	0.465	0.519	0.465	**0.464**	0.470
CIFAR-10	ResNet50（FL 2）	0.625	0.506	0.457	0.511	0.485	0.465	**0.455**
CIFAR-10	ResNet50（FL 3）	0.524	0.525	**0.429**	0.531	0.528	0.464	0.513
CIFAR-10	ResNet50（LS 0.05）	0.766	0.520	0.455	0.634	0.525	0.459	**0.440**
CIFAR-10	Wide-ResNet-26-10（交叉熵）	0.699	0.399	0.388	0.440	0.404	**0.371**	0.382
CIFAR-10	Wide-ResNet-26-10 (FL 1)	0.611	0.442	0.399	0.412	0.402	0.400	**0.375**
CIFAR-10	Wide-ResNet-26-10（FL 2）	0.479	0.443	0.436	0.462	**0.428**	0.437	0.429
CIFAR-10	Wide-ResNet-26-10（FL 3）	0.523	0.401	0.377	0.452	0.400	**0.373**	0.407
CIFAR-10	Wide-ResNet-26-10（LS 0.05）	0.869	0.417	0.424	0.502	0.458	**0.400**	0.415
CIFAR-100	DenseNet121（交叉熵）	0.458	**0.210**	0.239	0.258	0.252	0.226	0.688
CIFAR-100	DenseNet121（FL 1）	0.364	**0.231**	0.249	0.267	0.261	0.237	0.695
CIFAR-100	DenseNet121（FL 2）	0.280	0.278	0.241	0.256	0.252	**0.239**	0.694
CIFAR-100	DenseNet121（FL 3）	0.254	0.283	0.248	0.252	0.258	**0.244**	0.698
CIFAR-100	DenseNet121（LS 0.05）	0.315	0.223	0.238	0.252	0.223	**0.206**	0.668
CIFAR-100	ResNet110（交叉熵）	0.433	**0.183**	0.233	0.252	0.254	0.233	0.676
CIFAR-100	ResNet110（FL 1）	0.372	0.231	0.243	0.262	0.257	**0.229**	0.669
CIFAR-100	ResNet110（FL 2）	0.321	0.253	0.246	0.254	0.254	**0.243**	0.660
CIFAR-100	ResNet110 (FL 3)	0.281	0.257	0.245	0.258	0.258	**0.241**	0.671
CIFAR-100	ResNet110 (LS 0.05)	0.299	**0.217**	0.245	0.253	0.250	0.221	0.667
CIFAR-100	ResNet50 (交叉熵)	0.412	**0.215**	0.250	0.259	0.259	0.240	0.683
CIFAR-100	ResNet50（FL 1）	0.337	0.251	0.250	0.250	0.253	**0.243**	0.674
CIFAR-100	ResNet50（FL 2）	0.282	0.272	0.251	0.250	0.244	**0.234**	0.692
CIFAR-100	ResNet50（FL 3）	0.256	0.274	0.250	0.252	0.245	**0.238**	0.707
CIFAR-100	ResNet50（LS 0.05）	0.271	**0.222**	0.239	0.258	0.243	0.225	0.640
CIFAR-100	Wide-ResNet-26-10（交叉熵）	0.372	**0.198**	0.229	0.230	0.241	0.219	0.645
CIFAR-100	Wide-ResNet-26-10（FL 1）	0.264	0.238	0.229	0.223	0.219	**0.210**	0.624
CIFAR-100	Wide-ResNet-26-10（FL 2）	0.218	0.245	0.211	0.216	0.215	**0.208**	0.611
CIFAR-100	Wide-ResNet-26-10（FL 3）	0.226	0.260	0.216	0.228	0.220	**0.215**	0.596
CIFAR-100	Wide-ResNet-26-10（LS 0.05）	0.239	0.222	0.229	0.240	0.238	**0.219**	0.608
TinyImageNet	ResNet50（交叉熵）	0.250	0.218	0.216	**0.201**	0.213	0.208	0.671
TinyImageNet	ResNet50（FL 1）	0.218	0.226	**0.199**	0.205	0.216	0.212	0.684
TinyImageNet	ResNet50（FL 2）	0.205	0.228	0.208	**0.203**	0.214	0.204	0.678
TinyImageNet	ResNet50（FL 3）	0.203	0.227	**0.202**	0.204	0.217	0.209	0.677
TinyImageNet	ResNet50（LS 0.05）	0.231	0.210	0.203	**0.200**	0.206	0.203	0.628

表 *8*：*cwECE %* – 分类期望校准误差（越低越好），*15*个分箱，组 *2*

数据集	模型	校准前	直方图分箱	保序回归	T 缩放	V 缩放	V 缩放-b	M 缩放-b
CIFAR-10	DenseNet121	42.22	32.98	80.26	41.59	**25.41**	29.82	36.27
CIFAR-10	DenseNet161	25.30	38.11	**16.86**	27.21	24.37	26.92	34.27
CIFAR-10	DenseNet169	20.25	34.25	**16.18**	22.95	41.38	36.91	26.38
CIFAR-10	GoogleNet	19.85	24.06	24.77	**13.08**	16.88	25.35	42.19
CIFAR-10	InceptionV3	31.42	80.60	**17.31**	31.74	20.06	33.00	28.17
CIFAR-10	MobileNetV2	33.59	68.23	25.32	**24.74**	32.21	74.59	27.27
CIFAR-10	ResNet18	34.79	28.21	**17.80**	28.45	76.02	24.15	24.30
CIFAR-10	ResNet34	23.94	42.79	**18.42**	24.47	31.98	25.98	21.92
CIFAR-10	ResNet50	26.73	41.10	19.27	24.68	24.24	**16.52**	31.25
CIFAR-10	VGG11 _bn	23.38	16.46	16.86	23.40	18.33	**16.37**	25.19
CIFAR-10	VGG13 _bn	33.53	25.74	**14.47**	36.17	40.39	25.12	23.99
CIFAR-10	VGG16 _bn	43.19	34.06	**19.88**	43.57	26.82	22.98	20.03
CIFAR-10	VGG19 _bn	35.49	33.63	**18.37**	22.27	36.65	24.40	19.50
CIFAR-100	MobileNetV2 _x0_5	39.98	22.95	14.67	43.49	**9.70**	11.54	89.36
CIFAR-100	MobileNetV2 _x1_0	28.71	25.03	17.36	9.74	**7.37**	9.69	77.00
CIFAR-100	MobileNetV2 _x1_4	25.60	25.06	12.64	**6.32**	8.72	10.38	90.57
CIFAR-100	ResNet20	38.15	23.61	13.66	12.52	**8.89**	14.32	76.15
CIFAR-100	ResNet32	32.21	22.11	13.88	**9.12**	9.37	9.23	84.30
CIFAR-100	ResNet44	32.68	28.18	15.00	**9.09**	10.82	12.42	92.21
CIFAR-100	ResNet56	30.81	61.69	14.19	10.67	**7.78**	9.63	82.18
CIFAR-100	ShuffleNetV2 _x0_5	23.82	21.63	12.26	8.51	**7.89**	8.38	76.58
CIFAR-100	ShuffleNetV2 _x1_0	24.66	24.33	18.31	13.27	14.08	**10.62**	82.94
CIFAR-100	ShuffleNetV2 _x1_5	24.47	24.70	14.58	17.98	13.47	**12.31**	71.59
CIFAR-100	ShuffleNetV2 _x2_0	93.58	24.84	13.84	27.37	27.23	**12.18**	83.28
CIFAR-100	VGG11 _bn	40.53	29.29	20.59	15.36	**13.28**	14.15	47.91
CIFAR-100	VGG13 _bn	34.00	33.38	18.11	16.34	13.75	**13.11**	51.88
CIFAR-100	VGG16 _bn	48.47	45.52	16.72	**12.39**	13.18	15.35	76.81
CIFAR-100	VGG19 _bn	51.11	38.46	16.52	17.08	14.98	**13.73**	54.69
ImageNet	EfficientNet _b8	44.80	22.83	**12.03**	93.37	22.61	17.02	55.00
ImageNet	MobileNetV2 _120d	12.31	18.22	6.16	**3.75**	5.92	5.65	63.64
ImageNet	RepVGG_b3	11.15	19.16	**7.36**	10.09	10.68	14.01	61.53
ImageNet	VGG19 _bn	8.33	20.24	8.72	6.40	5.48	**4.71**	72.88

*Ta*表 *9*：*MCE %* – 最大校准误差（越低越好），*15*个分箱，第*1*组

数据集	模型	校准前	直方图分箱	保序回归	T 缩放	V 缩放	V 缩放-b	M 缩放-b
CIFAR-10	DenseNet121（交叉熵）	46.58	53.58	23.62	**22.26**	29.52	23.21	67.72
CIFAR-10	DenseNet121 (FL 1)	36.54	37.23	26.98	29.38	31.21	70.77	**22.20**
CIFAR-10	DenseNet121 (FL 2)	27.80	35.64	**10.22**	66.85	29.08	28.61	29.06
CIFAR-10	DenseNet121 (FL 3)	77.08	38.19	38.07	76.59	74.49	**19.89**	23.57
CIFAR-10	DenseNet121 (LS 0.05)	55.74	56.99	80.80	70.26	57.18	66.81	**40.70**
CIFAR-10	ResNet110 (交叉熵)	47.15	50.61	35.38	**25.12**	25.90	27.26	31.14
CIFAR-10	ResNet110 (FL 1)	73.35	61.93	26.03	49.35	**19.64**	42.96	29.91
CIFAR-10	ResNet110（FL 2）	26.15	30.54	24.24	**23.77**	25.80	32.68	28.42
CIFAR-10	ResNet110（FL 3）	**25.15**	30.68	29.37	38.39	30.89	69.55	30.64
CIFAR-10	ResNet110（LS 0.05）	55.53	83.03	**20.67**	59.80	58.54	59.94	48.80
CIFAR-10	ResNet50（交叉熵）	42.08	55.72	22.08	20.49	32.80	**18.27**	31.52
CIFAR-10	ResNet50（FL 1）	37.23	73.04	**22.43**	31.58	73.41	26.62	34.54
CIFAR-10	ResNet50（FL 2）	**25.00**	41.04	74.60	67.95	76.43	29.61	32.46
CIFAR-10	ResNet50（FL 3）	28.99	32.00	**11.89**	19.15	24.35	31.04	30.76
CIFAR-10	ResNet50（LS 0.05）	50.21	45.06	**21.05**	74.74	51.06	76.52	43.41
CIFAR-10	Wide-ResNet-26-10（交叉熵）	48.03	63.54	**25.41**	30.04	30.41	31.41	32.78
CIFAR-10	Wide-ResNet-26-10（FL 1）	26.49	48.07	**18.55**	75.41	24.17	31.02	21.96
CIFAR-10	Wide-ResNet-26-10（FL 2）	33.49	74.05	**23.92**	31.05	29.04	24.23	37.53
CIFAR-10	Wide-ResNet-26-10（FL 3）	28.18	19.14	18.52	69.02	31.10	**17.94**	29.90
CIFAR-10	Wide-ResNet-26-10（LS 0.05）	39.10	36.02	**19.03**	31.45	71.21	36.22	42.35
CIFAR-100	DenseNet121（交叉熵）	54.14	29.97	21.75	28.70	**12.36**	12.84	77.88
CIFAR-100	DenseNet121（FL 1）	35.37	36.84	17.62	16.51	14.52	**14.35**	86.25
CIFAR-100	DenseNet121 (FL 2)	26.13	24.50	15.21	**12.86**	14.74	15.88	82.62
CIFAR-100	DenseNet121 (FL 3)	14.76	23.33	11.81	12.16	13.74	**9.81**	76.96
CIFAR-100	DenseNet121 (LS 0.05)	46.99	30.43	22.79	60.42	12.61	**11.28**	75.69
CIFAR-100	ResNet110（交叉熵）	58.32	43.75	19.58	24.80	19.10	**17.70**	85.79
CIFAR-100	ResNet110（FL 1）	38.23	41.78	20.14	18.67	20.87	**17.51**	70.51
CIFAR-100	ResNet110（FL 2）	30.27	30.83	15.95	19.63	18.86	**14.13**	89.28
CIFAR-100	ResNet110（FL 3）	25.44	23.25	16.91	21.06	16.78	**15.19**	69.21
CIFAR-100	ResNet110（LS 0.05）	40.00	30.48	**15.85**	30.16	20.27	20.47	89.53
CIFAR-100	ResNet50（交叉熵）	47.19	47.37	20.20	**16.26**	20.44	22.76	83.57
CIFAR-100	ResNet50（FL 1）	34.78	32.35	21.41	16.26	**13.78**	15.34	53.71
CIFAR-100	ResNet50（FL 2）	23.83	25.69	14.62	**14.00**	15.24	16.00	91.02
CIFAR-100	ResNet50（FL 3）	14.00	20.36	**9.20**	12.66	12.36	10.70	89.46
CIFAR-100	ResNet50（LS 0.05）	31.28	29.37	94.39	23.55	21.02	**18.71**	69.93
CIFAR-100	Wide-ResNet-26-10（交叉熵）	55.83	46.57	21.84	15.99	14.89	**14.20**	83.06
CIFAR-100	Wide-ResNet-26-10（FL 1）	27.40	30.07	16.22	**12.96**	14.48	13.07	72.31
CIFAR-100	Wide-ResNet-26-10（FL 2）	16.16	27.35	**11.15**	12.62	13.02	26.94	88.77
CIFAR-100	Wide-ResNet-26-10（FL 3）	**9.90**	24.61	13.33	12.65	16.52	44.60	80.98
CIFAR-100	Wide-ResNet-26-10（LS 0.05）	26.84	32.39	20.92	27.45	24.19	**18.73**	75.74
TinyImageNet	ResNet50（交叉熵）	33.75	21.46	26.91	17.26	**15.41**	17.65	80.79
TinyImageNet	ResNet50（FL 1）	18.63	14.14	17.73	**9.08**	12.43	17.18	78.34
TinyImageNet	ResNet50（FL 2）	9.53	14.57	9.68	**7.69**	9.86	10.83	71.59
TinyImageNet	ResNet50（FL 3）	6.39	14.29	10.00	**6.21**	7.30	11.51	77.58
TinyImageNet	ResNet50（LS 0.05）	26.59	15.09	**7.08**	13.56	17.30	18.21	77.05

Table 10：*MCE %* – 最大校准误差（越低越好），*15*个分箱，组 *2*

C 适当评分规则

数据集	模型	校准前	直方图分箱	保序回归	T 缩放	V 缩放	V 缩放-b	M 缩放-b
CIFAR-10	DenseNet121	0.257	0.427	0.354	0.257	0.258	**0.255**	0.256
CIFAR-10	DenseNet161	0.261	0.495	0.285	0.262	0.255	**0.253**	0.281
CIFAR-10	DenseNet169	0.245	0.559	0.309	0.245	**0.242**	0.242	0.258
CIFAR-10	GoogleNet	0.246	0.290	0.300	0.241	0.241	0.239	**0.235**
CIFAR-10	InceptionV3	0.265	0.550	0.349	0.266	0.264	**0.264**	0.264
CIFAR-10	MobileNetV2	0.241	0.494	0.309	0.239	0.239	**0.237**	0.239
CIFAR-10	ResNet18	0.269	0.616	0.327	0.269	0.265	**0.263**	0.269
CIFAR-10	ResNet34	0.269	0.508	0.317	0.265	0.261	**0.260**	0.265
CIFAR-10	ResNet50	0.255	0.589	0.289	0.253	0.252	**0.251**	0.258
CIFAR-10	VGG11_bn	0.257	0.410	0.342	0.257	0.255	**0.252**	0.254
CIFAR-10	VGG13_bn	0.207	0.487	0.249	0.207	0.205	**0.202**	0.206
CIFAR-10	VGG16_bn	0.233	0.522	0.276	0.232	0.230	**0.229**	0.237
CIFAR-10	VGG19_bn	0.248	0.553	0.291	0.245	0.244	**0.242**	0.250
CIFAR-100	MobileNetV2_x0_5	1.210	4.128	1.738	**1.066**	1.073	1.083	15.290
CIFAR-100	MobileNetV2_x1_0	1.102	4.019	1.480	**0.979**	0.986	0.986	13.881
CIFAR-100	MobileNetV2_x1_4	0.997	3.823	1.499	**0.908**	0.911	0.909	12.531
CIFAR-100	ResNet20	1.267	4.445	1.771	**1.155**	1.161	1.164	17.232
CIFAR-100	ResNet32	1.370	4.573	1.762	**1.148**	1.155	1.162	16.467
CIFAR-100	ResNet44	1.358	4.350	1.656	**1.097**	1.104	1.110	15.459
CIFAR-100	ResNet56	1.340	3.933	1.711	**1.069**	1.076	1.075	15.020
CIFAR-100	ShuffleNetV2_x0_5	1.342	4.479	1.795	**1.196**	1.199	1.206	16.835
CIFAR-100	ShuffleNetV2_x1_0	1.200	4.205	1.760	**1.089**	1.098	1.103	14.203
CIFAR-100	ShuffleNetV2_x1_5	1.109	3.727	1.594	**1.052**	1.053	1.054	12.794
CIFAR-100	ShuffleNetV2_x2_0	1.018	3.610	1.613	**0.992**	0.999	0.997	12.425
CIFAR-100	VGG11_bn	1.541	3.978	1.840	**1.259**	1.275	1.270	4.707
CIFAR-100	VGG13_bn	1.320	3.361	1.642	**1.106**	1.118	1.117	4.230
CIFAR-100	VGG16_bn	1.677	3.162	1.584	**1.142**	1.148	1.144	5.757
CIFAR-100	VGG19_bn	1.836	3.034	1.621	**1.158**	1.162	1.158	7.711
ImageNet	EfficientNet_b8	0.656	2.778	1.432	**0.572**	0.581	0.664	5.554
ImageNet	MobileNetV2_120d	0.946	4.282	1.917	**0.894**	0.900	0.931	10.508
ImageNet	RepVGG_b3	0.823	3.751	1.795	0.815	**0.811**	0.871	9.225
ImageNet	VGG19_bn	1.034	4.892	2.100	**1.017**	1.021	1.040	13.665

表*11*：负对数似然（越低越好），第*1*组

数据集	模型	校准前	直方图分箱	保序回归	T 缩放	V 缩放	V 缩放-b	M 缩放-b
CIFAR-10	DenseNet121（交叉熵）	0.422	0.398	0.213	0.216	0.209	0.206	**0.198**
CIFAR-10	DenseNet121 (FL 1)	0.230	0.508	0.271	0.190	0.185	**0.182**	0.193
CIFAR-10	DenseNet121 (FL 2)	0.186	0.548	0.278	0.181	0.178	**0.178**	0.182
CIFAR-10	DenseNet121 (FL 3)	0.193	0.369	0.312	0.194	0.191	**0.191**	0.206
CIFAR-10	DenseNet121 (LS 0.05)	0.306	0.389	0.334	0.306	0.298	**0.293**	0.302
CIFAR-10	ResNet110 (交叉熵)	0.512	0.454	0.227	0.230	0.224	**0.222**	0.228
CIFAR-10	ResNet110 (FL 1)	0.234	0.547	0.285	0.181	0.181	**0.177**	0.182
CIFAR-10	ResNet110（FL 2）	0.195	0.477	0.261	0.179	0.177	**0.174**	0.176
CIFAR-10	ResNet110（FL 3）	0.185	0.478	0.252	0.183	0.178	**0.177**	0.185
CIFAR-10	ResNet110（LS 0.05）	0.301	0.434	0.302	0.301	0.294	**0.289**	0.290
CIFAR-10	ResNet50（交叉熵）	0.428	0.516	0.207	0.212	0.207	**0.203**	0.203
CIFAR-10	ResNet50（FL 1）	0.253	0.604	0.246	0.193	0.187	**0.184**	0.188
CIFAR-10	ResNet50（FL 2）	0.191	0.530	0.292	0.181	0.177	**0.176**	0.186
CIFAR-10	ResNet50（FL 3）	0.190	0.485	0.237	0.188	0.186	**0.185**	0.199
CIFAR-10	ResNet50（LS 0.05）	0.293	0.575	0.310	0.292	0.285	**0.280**	0.296
CIFAR-10	Wide-ResNet-26-10（交叉熵）	0.273	0.400	0.171	0.156	0.156	**0.153**	0.154
CIFAR-10	Wide-ResNet-26-10（FL 1）	0.181	0.454	0.209	0.158	0.158	**0.157**	0.162
CIFAR-10	Wide-ResNet-26-10 (FL 2)	0.161	0.447	0.244	0.159	0.157	**0.156**	0.160
CIFAR-10	Wide-ResNet-26-10（FL 3）	0.158	0.334	0.206	0.157	0.152	**0.151**	0.161
CIFAR-10	Wide-ResNet-26-10（LS 0.05）	0.224	0.389	0.279	0.217	0.211	**0.207**	0.230
CIFAR-100	DenseNet121（交叉熵）	2.063	2.325	1.564	1.200	1.183	**1.173**	11.759
CIFAR-100	DenseNet121（FL 1）	1.188	3.153	1.582	1.031	1.012	**1.006**	11.888
CIFAR-100	DenseNet121（FL 2）	0.948	3.645	1.437	0.928	0.917	**0.913**	11.761
CIFAR-100	DenseNet121（FL 3）	0.903	3.609	1.484	**0.902**	0.903	0.905	11.792
CIFAR-100	DenseNet121（LS 0.05）	1.435	2.492	1.899	1.375	1.356	**1.349**	11.259
CIFAR-100	ResNet110（交叉熵）	1.843	2.391	1.382	1.074	1.076	**1.070**	11.560
CIFAR-100	ResNet110（FL 1）	1.165	2.900	1.493	0.969	0.964	**0.962**	11.421
CIFAR-100	ResNet110（FL 2）	1.006	3.352	1.507	0.942	0.937	**0.935**	11.200
CIFAR-100	ResNet110 (FL 3)	0.941	3.254	1.507	0.922	0.918	**0.913**	11.475
CIFAR-100	ResNet110 (LS 0.05)	1.363	2.603	1.847	1.325	1.325	**1.313**	11.344
CIFAR-100	ResNet50 (交叉熵)	1.582	2.785	1.538	1.089	1.092	**1.084**	11.625
CIFAR-100	ResNet50（FL 1）	1.094	3.426	1.625	0.969	0.969	**0.965**	11.530
CIFAR-100	ResNet50（FL 2）	0.951	3.301	1.436	0.924	**0.920**	0.920	11.816
CIFAR-100	ResNet50（FL 3）	0.897	3.491	1.514	0.893	0.887	**0.885**	12.042
CIFAR-100	ResNet50（LS 0.05）	1.230	3.014	1.727	1.220	1.216	**1.205**	10.760
CIFAR-100	Wide-ResNet-26-10（交叉熵）	1.437	2.430	1.354	**0.979**	0.990	0.982	10.962
CIFAR-100	Wide-ResNet-26-10（FL 1）	0.898	2.725	1.293	0.846	0.845	**0.844**	10.631
CIFAR-100	Wide-ResNet-26-10（FL 2）	0.780	3.031	1.245	0.780	**0.779**	0.780	10.419
CIFAR-100	Wide-ResNet-26-10（FL 3）	0.759	3.206	1.327	0.755	**0.752**	0.755	10.240
CIFAR-100	Wide-ResNet-26-10（LS 0.05）	1.109	2.754	1.549	1.109	1.106	**1.095**	10.395
TinyImageNet	ResNet50（交叉熵）	2.325	6.905	3.478	**2.199**	2.213	2.214	21.369
TinyImageNet	ResNet50（FL 1）	2.180	6.833	3.411	**2.160**	2.170	2.167	22.083
TinyImageNet	ResNet50（FL 2）	2.033	6.626	3.158	**2.032**	2.048	2.042	21.973
TinyImageNet	ResNet50（FL 3）	2.038	6.768	3.004	**2.036**	2.057	2.047	21.756
TinyImageNet	ResNet50（LS 0.05）	2.334	5.121	3.108	**2.129**	2.142	2.141	18.609

表 *12*：负对数似然（越低越好），组 *2*

数据集	模型	校准前	直方图分箱	保序回归	T 缩放	V 缩放	V 缩放-b	M 缩放-b
CIFAR-10	DenseNet121	0.102	0.106	**0.099**	0.102	0.102	0.102	0.101
CIFAR-10	DenseNet161	0.102	0.107	**0.098**	0.103	0.101	0.101	0.104
CIFAR-10	DenseNet169	0.098	0.106	**0.095**	0.099	0.098	0.098	0.100
CIFAR-10	GoogleNet	**0.109**	0.112	0.109	0.109	0.110	0.109	0.110
CIFAR-10	InceptionV3	0.108	0.115	**0.105**	0.109	0.109	0.108	0.109
CIFAR-10	MobileNetV2	0.102	0.106	**0.099**	0.103	0.103	0.102	0.103
CIFAR-10	ResNet18	0.117	0.124	**0.113**	0.116	0.115	0.115	0.117
CIFAR-10	ResNet34	0.113	0.116	**0.106**	0.111	0.110	0.110	0.111
CIFAR-10	ResNet50	0.107	0.112	**0.102**	0.106	0.105	0.105	0.106
CIFAR-10	VGG11 _bn	0.118	0.122	**0.115**	0.118	0.118	0.117	0.118
CIFAR-10	VGG13 _bn	0.092	0.102	**0.091**	0.093	0.093	0.092	0.094
CIFAR-10	VGG16 _bn	0.100	0.108	**0.097**	0.100	0.099	0.099	0.101
CIFAR-10	VGG19 _bn	0.102	0.109	**0.098**	0.101	0.101	0.100	0.102
CIFAR-100	MobileNetV2 _x0_5	0.431	0.469	0.413	**0.404**	0.406	0.408	0.892
CIFAR-100	MobileNetV2 _x1_0	0.387	0.431	0.373	**0.365**	0.367	0.368	0.814
CIFAR-100	MobileNetV2 _x1_4	0.352	0.402	0.341	**0.336**	0.337	0.337	0.736
CIFAR-100	ResNet20	0.459	0.511	0.448	**0.438**	0.440	0.441	1.001
CIFAR-100	ResNet32	0.453	0.493	0.431	**0.420**	0.423	0.424	0.956
CIFAR-100	ResNet44	0.441	0.475	0.411	**0.403**	0.405	0.406	0.897
CIFAR-100	ResNet56	0.429	0.451	0.397	**0.392**	0.394	0.395	0.873
CIFAR-100	ShuffleNetV2 _x0_5	0.468	0.510	0.450	**0.442**	0.443	0.445	0.996
CIFAR-100	ShuffleNetV2 _x1_0	0.413	0.451	0.401	**0.392**	0.395	0.396	0.847
CIFAR-100	ShuffleNetV2 _x1_5	0.384	0.426	0.377	**0.373**	0.373	0.374	0.770
CIFAR-100	ShuffleNetV2 _x2_0	0.359	0.408	0.354	**0.352**	0.353	0.353	0.736
CIFAR-100	VGG11 _bn	0.455	0.473	0.424	**0.415**	0.419	0.419	0.673
CIFAR-100	VGG13 _bn	0.406	0.425	0.381	**0.373**	0.375	0.375	0.629
CIFAR-100	VGG16 _bn	0.450	0.439	0.386	**0.382**	0.383	0.382	0.633
CIFAR-100	VGG19 _bn	0.456	0.436	0.380	0.379	0.378	**0.377**	0.672
ImageNet	EfficientNet _b8	0.222	0.251	0.217	**0.215**	0.217	0.219	0.407
ImageNet	MobileNetV2 _120d	0.323	0.379	0.324	**0.314**	0.319	0.322	0.776
ImageNet	RepVGG_b3	0.283	0.332	0.286	**0.282**	0.285	0.289	0.655
ImageNet	VGG19 _bn	0.354	0.423	0.363	**0.353**	0.358	0.361	0.900

表 *13*：布里尔分数（越低越好），第*1*组

数据集	模型	校准前	直方图分箱	保序回归	T 缩放	V 缩放	V 缩放-b	M 缩放-b
CIFAR-10	DenseNet121（交叉熵）	0.094	0.094	**0.081**	0.084	0.082	0.082	0.082
CIFAR-10	DenseNet121 (FL 1)	0.086	0.092	**0.079**	0.081	0.080	0.080	0.084
CIFAR-10	DenseNet121 (FL 2)	0.081	0.089	**0.079**	0.080	0.079	0.079	0.080
CIFAR-10	DenseNet121 (FL 3)	0.089	0.091	**0.088**	0.089	0.088	0.088	0.090
CIFAR-10	DenseNet121 (LS 0.05)	0.102	0.102	**0.091**	0.103	0.102	0.101	0.100
CIFAR-10	ResNet110 (交叉熵)	0.097	0.098	0.086	0.087	0.086	0.086	**0.086**
CIFAR-10	ResNet110 (FL 1)	0.086	0.088	0.078	0.079	0.078	**0.078**	0.078
CIFAR-10	ResNet110（FL 2）	0.082	0.087	0.079	0.079	0.078	**0.078**	0.079
CIFAR-10	ResNet110（FL 3）	0.083	0.088	0.081	0.082	**0.079**	0.079	0.082
CIFAR-10	ResNet110（LS 0.05）	0.103	0.103	**0.092**	0.103	0.102	0.101	0.099
CIFAR-10	ResNet50（交叉熵）	0.093	0.098	0.084	0.085	0.084	**0.083**	0.084
CIFAR-10	ResNet50（FL 1）	0.091	0.098	0.083	0.084	0.083	**0.083**	0.083
CIFAR-10	ResNet50（FL 2）	0.083	0.091	0.080	0.081	0.079	**0.079**	0.081
CIFAR-10	ResNet50（FL 3）	0.085	0.092	0.084	0.084	0.084	**0.084**	0.088
CIFAR-10	ResNet50（LS 0.05）	0.102	0.106	**0.094**	0.102	0.101	0.101	0.099
CIFAR-10	Wide-ResNet-26-10（交叉熵）	0.069	0.070	**0.061**	0.061	0.061	0.061	0.061
CIFAR-10	Wide-ResNet-26-10（FL 1）	0.070	0.077	0.067	**0.067**	0.068	0.068	0.067
CIFAR-10	Wide-ResNet-26-10（FL 2）	0.071	0.078	**0.068**	0.070	0.071	0.070	0.071
CIFAR-10	Wide-ResNet-26-10（FL 3）	0.068	0.072	**0.067**	0.069	0.068	0.067	0.069
CIFAR-10	Wide-ResNet-26-10（LS 0.05）	0.076	0.078	**0.072**	0.077	0.076	0.076	0.077
CIFAR-100	DenseNet121（交叉熵）	0.446	0.413	0.368	0.372	0.364	**0.364**	0.687
CIFAR-100	DenseNet121（FL 1）	0.385	0.394	0.351	0.354	0.348	**0.347**	0.694
CIFAR-100	DenseNet121 (FL 2)	0.348	0.389	**0.336**	0.339	0.336	0.336	0.693
CIFAR-100	DenseNet121 (FL 3)	0.336	0.391	0.337	0.336	0.336	**0.336**	0.697
CIFAR-100	DenseNet121 (LS 0.05)	0.406	0.405	**0.378**	0.387	0.378	0.378	0.667
CIFAR-100	ResNet110（交叉熵）	0.421	0.393	0.347	0.349	0.346	**0.345**	0.675
CIFAR-100	ResNet110（FL 1）	0.375	0.384	0.333	0.336	0.333	**0.332**	0.669
CIFAR-100	ResNet110（FL 2）	0.358	0.387	**0.334**	0.339	0.337	0.336	0.658
CIFAR-100	ResNet110（FL 3）	0.346	0.379	**0.331**	0.337	0.334	0.333	0.670
CIFAR-100	ResNet110（LS 0.05）	0.386	0.401	**0.366**	0.372	0.372	0.371	0.665
CIFAR-100	ResNet50（交叉熵）	0.411	0.406	0.355	0.353	0.353	**0.353**	0.682
CIFAR-100	ResNet50（FL 1）	0.365	0.393	0.341	0.337	0.335	**0.334**	0.673
CIFAR-100	ResNet50（FL 2）	0.346	0.388	0.337	0.337	**0.335**	0.335	0.691
CIFAR-100	ResNet50（FL 3）	0.331	0.382	0.332	0.329	0.328	**0.328**	0.706
CIFAR-100	ResNet50 (LS 0.05)	0.365	0.390	**0.351**	0.359	0.357	0.356	0.639
CIFAR-100	Wide-ResNet-26-10 (交叉熵)	0.365	0.364	0.316	**0.310**	0.312	0.311	0.645
CIFAR-100	Wide-ResNet-26-10（FL 1）	0.304	0.337	0.294	0.291	**0.290**	0.290	0.623
CIFAR-100	Wide-ResNet-26-10（FL 2）	0.287	0.331	0.284	0.286	0.284	**0.284**	0.611
CIFAR-100	Wide-ResNet-26-10（FL 3）	0.281	0.336	**0.280**	0.283	0.281	0.281	0.596
CIFAR-100	Wide-ResNet-26-10（LS 0.05）	0.326	0.363	**0.317**	0.325	0.323	0.322	0.607
TinyImageNet	ResNet50（交叉熵）	0.677	0.720	0.661	**0.646**	0.652	0.654	1.324
TinyImageNet	ResNet50（FL 1）	0.648	0.714	0.650	**0.641**	0.645	0.646	1.352
TinyImageNet	ResNet50（FL 2）	0.620	0.688	0.630	**0.620**	0.627	0.627	1.343
TinyImageNet	ResNet50（FL 3）	0.624	0.707	0.634	**0.623**	0.631	0.631	1.340
TinyImageNet	ResNet50（LS 0.05）	0.645	0.682	0.617	**0.613**	0.620	0.621	1.229

表 *14*：布里尔分数（越低越好），组 *2*