arXiv:2303.02310v1 [cs.LG] 4 Mar 2023

IKD+：用于视网膜病变分类的可靠、低复杂度深度模型

Shreyas Bhat Brahmavar[1], *Rohit Rajesh*[1],*Tirtharaj Dash*[2],*Lovekesh Vig*[3],*Tanmay Tulsidas Verlekar*[1],*Md Mahmudul Hasan*[4], *Tariq Khan*[4],*Erik Meijering*[4],*Ashwin Srinivasan*[1]

[1] APPCAIR，BITS Pilani，K.K. Birla Goa Campus，印度[2] Boolean Lab，加利福尼亚大学，圣迭戈，美国[3] TCS Research，印度[4] 计算机科学与工程学院，新南威尔士大学，悉尼，澳大利亚

摘要

用于视网膜病变的深度神经网络（DNN）模型，其预测准确率估计处于 90% 中高区间。然而，以下方面仍未得到解决：最先进的模型复杂，训练和部署需要大量计算基础设施；预测的可靠性可能存在很大差异。本文聚焦于这些问题，并提出一种称为 IKD+ 的迭代知识蒸馏（IKD）方法，在模型规模、准确率和可靠性之间进行权衡。我们使用两种广泛采用的模型校准估计技术（Platt 缩放和温度缩放），研究 IKD+ 的运行机制；所用模型是目前性能最佳的模型，即由 EfficientNets 集成而成、参数约为 1 亿的模型。我们证明，采用温度缩放的 IKD+ 模型，其参数数量最多可比原始集成减少约 500 倍，同时准确率没有显著损失。此外，IKD+ 模型的校准分数（可靠性）与基线模型相当或更优。

索引词— 视网膜病变、知识蒸馏、模型校准、EfficientNets、医学图像处理

1. 引言

视网膜病变是一种影响视网膜的疾病。视网膜是眼球后部对光敏感的组织层，预计病例数将达到约 1.91 亿例 2030 [1]。通常，它是糖尿病的常见并发症，也可能由青光眼和年龄相关性黄斑变性等其他疾病引起。如果未能及时诊断或治疗，视网膜病变可能导致视力丧失，甚至失明[2, 3, 4]。人工智能（AI），尤其是深度神经网络（DNN）的快速发展，以及其在医疗保健领域的成功应用，使其在视网膜病变诊断方面具有巨大的应用潜力。此外，视网膜病变深度学习模型已取得较高的预测准确率，有些模型的表现达到 90% 中高段 [5, 6, 7]。然而，这些模型通常非常复杂，需要大量计算基础设施和专业知识才能在真实场景中训练和部署 [7]。此外，其预测的可靠性可能存在很大差异，因此难以在临床实践中放心使用这些模型。再者，这些模型往往是黑箱，对其预测几乎不给出或完全不给出有意义的解释；当模型复杂度较高时，这一问题尤其突出。

在本文中，我们通过探索迭代式知识蒸馏来应对这些挑战，[8]该过程能够为视网膜病变逐步生成更简单的深度学习模型。我们的工作建立在先进的模型校准 [9] 技术之上，这些技术表明，在不牺牲预测准确率的情况下，可以提高高概率预测的可靠性。我们使用两种用于多疾病视网膜病变分类的最先进方法来验证我们的方法。结果表明，显著更简单的模型能够达到与原始模型相近的预测准确率。此外，简单模型的校准分数与复杂模型相当，甚至更优。

我们的方法有望通过提供更可靠、低复杂度且易于训练和部署于临床环境中的深度学习模型，显著改善视网膜病变的诊断（和治疗）。它还可以提高这些模型的透明度和可解释性，使医疗专业人员能够更好地理解结果并信任其预测。通过应对这些挑战，我们的方法有望显著改善全球数百万遭受可预防或未被发现的失明和视力损伤人群的医疗护理质量。除了促进视网膜病变的诊断和治疗外，我们还将校准技术扩展到多标签分类模型可靠性的估计。最先进的校准技术主要关注二分类或多分类问题。相比之下，多标签

分类问题的相关研究相对较少，具体内容将在以下章节中介绍。

1.1. 相关工作

Lyu 等人 [10]提出了一种基于疾病对视网膜眼底图像进行分类的多专家深度学习网络。该网络由多个分类器组成，这些分类器的对数几率会进行堆叠，以形成集成结果并完成最终预测。在 M¨uller 等人 [7], 的研究中，作者提出了一种多分类器方法，用于将视网膜图像分为正常和异常类别，并在 28 个类别中识别具体的疾病标签。该方法利用集成学习，结合多个异构深度卷积神经网络模型的预测能力。作者展示了其流程的高准确率，以及其与其他视网膜疾病预测领域最先进流程的可比性。此外，还有关于深度学习在视网膜病变诊断中应用的综合性综述 [11, 12]。

许多基于深度学习的技术采用复杂的深度网络结构，通常需要大量计算资源进行训练和部署，例如使用 DenseNet201、EfficientNetB4 构建的模型。一些技术还采用机器学习模型集成，例如堆叠逻辑回归，这会进一步增加计算需求。当前最先进方法的另一项局限在于其预测的可靠性。大多数模型采用堆叠设置，即针对每个类别分别应用二元逻辑回归算法，从而产生 29 个不同的逻辑回归模型（针对本文所解决的问题）。然而，这些模型的校准程度以及其能否生成可靠的预测尚不明确。校准是指预测概率与真实结果的匹配程度。在医疗领域，确保模型能够以更高的可靠性进行预测至关重要。因此，我们使用期望校准误差评估所提出模型的校准情况 [9, 13]。

1.2. 主要贡献

我们对构建可靠的视网膜病变诊断深度学习模型所作的贡献如下：

- 我们提出了一种称为 IKD+ 的新方法，以迭代方式将大型复杂模型（即教师模型）中的知识准确地迁移到更小、更高效的模型（即学生模型）中；
- 我们采用 Platt 缩放和温度缩放，以增强学生模型预测结果的可靠性；
- 通过大量实验，我们证明了所提出的方法在提高视网膜病变诊断的准确率和可靠性方面的有效性。

本文其余部分安排如下：第 2 节详述我们提出的方法，第 3 节介绍评估细节，并在第 4 节讨论我们的结果。

2. 可靠的 IKD（IKD+）

本节介绍可靠迭代知识蒸馏（IKD+）的概念，其目标是找出一个（大型、复杂）模型的较简单变体，该变体（a）具有与较复杂模型相当的准确率，且（b）能够做出可靠的预测（大致而言，即高置信度预测的预测误差较低）。后一个要求源于人们如今已认识到，现代深度网络模型有时会以极高的置信度做出非常糟糕的预测 [9] [14]。如果此类预测被用于提供医疗辅助，这一问题尤其严重。为实现这两个目标，我们采用 IKD 方法（见 [8]所述），并使用一种在迭代流程的每一步都考虑模型校准的损失函数。为完整起见，下面给出了 IKD 流程：随后将介绍损失函数和校准误差的具体细节。IKD 将模型 M 视为一对（σ， $\boldsymbol{\pi}$），其中σ 表示模型的结构，即神经元、层和激活函数的组织方式； $\boldsymbol{\pi}$ 表示模型的参数（突触权重）。知识蒸馏可以视为一种函数 KD，它寻找满足条件的 $\boldsymbol{\pi}_j$。

$$KD((\sigma_i, \boldsymbol{\pi}_i), \cdot) = (\sigma_j, \boldsymbol{\pi}_j). \tag{1}$$

与 [8] 类似，我们假设存在一个精化算子 ρ，将其应用于深度网络的结构，从而得到一个新的、通常更简单的结构。更简单的模型宽度和/或深度更小，且参数更少。受 Yalburgi 等人 [8] 的启发，我们提出了一种精化（ρ）算子，可实现块缩减，即：将结构中每个块内的特征数量按一定比例减少 p，并将每个块中的卷积滤波器数量减少 p。遵循 [8] 中的研究，我们的工作也将 p 的值设为 0.5。

我们工作的主要贡献是将模型的预测可靠性这一概念纳入 IKD 流程。为此，我们修改了 [8]中采用的损失函数，将[9]中提出的期望校准误差（ECE）纳入其中。下一小节将详细介绍这些内容。

校准敏感损失：

我们考虑两种不同的损失函数，分别称为温度缩放和 Platt 缩放。两种情况下，损失均表示为：

$$L\left(M_s \mid D\right) = \frac{1}{m} \sum_{j=1}^{m} E_j, \tag{2}$$

其中， E_j 表示温度缩放， M_s 是学生模型，D 是训练数据，而 Platt 缩放分别可表示为公式（3）和（4）。

$$2T^2 \alpha D_{KL}\left(\frac{P^{(j)}}{T}, \frac{Q^{(j)}}{T}\right) + (1-\alpha) \cdot \mathrm{CE}(\mathrm{Y}, \hat{\mathrm{Y}}) \tag{3}$$

$$\alpha D_{KL}\left(WP^{(j)}, WQ^{(j)}\right) + (1-\alpha)\cdot \mathrm{CE}(\mathrm{Y}, \hat{\mathrm{Y}}) \quad (4)$$

在公式（3）和（4）中，CE 表示交叉熵损失D_{KL} 表示 KL 散度 [15], ，而 T、 W 均在验证集上进行优化；m 表示批大小，c 表示类别数量。 α 表示教师指导相对于来自数据的硬目标的重要性。 $P^{(j)}$和 $Q^{(j)}$ 分别表示学生模型和教师模型输出的类别概率分布，其中 j 表示数据实例；$Y = [y_1, \ldots, y_c]$ 表示独热编码真实标签，而$\hat{}$ $Y = [\hat{}y_1$、.....、$\hat{}y_c]$ 表示学生模型的预测结果。

过程 1： 所提出的 IKD+ 过程

输入： 预训练教师模型 M_0；数量
迭代次数 k； σ_1 第一个学生模型；
精化算子 ρ；

输出： 一个压缩模型

初始化 i ： $= 1$;
当 $i \leq k$ **时**
  $M_i = \text{IKD+}(M_{i-1}, \sigma_i, \cdot)$;
  其中 IKD+使用校准敏感损失函数;
  $\sigma_{i+1} = \rho(\sigma_i)$;
  $increment\ i$;
结束
返回 M_k;

3. 实证评估

我们的评估旨在研究使用 IKD+ 时预测准确率、可靠性与压缩之间的权衡。具体而言，我们针对糖尿病视网膜病变的预测研究以下假设：

- 配备了合适的期望校准误差（ECE）估计器的 IKD+ 流程能够构建出比 IKD 更可靠的模型；
- 使用 IKD+ 不会显著降低预测性能，相较于规模相当的最先进模型或通过 IKD 获得的模型皆是如此。

3.1. 数据集和机器规格

我们在视网膜眼底多疾病图像数据集上进行评估 [16]。该数据集由 3200 张视网膜图像组成，其中 1920 张用于训练。与该数据集相关的一项挑战是疾病的分布。尽管糖尿病视网膜病变（DR）包含376 个样本（$\approx$ 10%），分流（ST）仅包含 5 个（$\approx$ 0.1%）。数据集的完整分布见图 1。所有图像均采用 RGB 颜色格式，且 768×768 像素。

我们提出的方法，即1，是对 [8]中所介绍算法的改进。该方法使用 Python 实现，运行于配备 NVIDIA Tesla V100-SXM2 GPU（32GB）的系统上

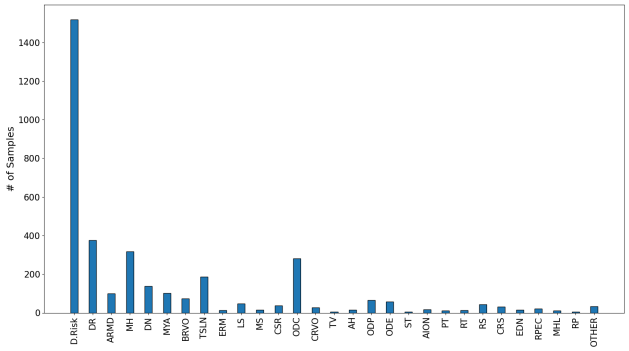

图 1。数据集的分布。

内存。代码基于 PyTorch 深度学习框架 [17] 开发。期望校准误差（ECE）使用 net:cal 库 [18] 计算。

3.2. 方法

给定一个数据集：
将数据划分为训练集 D_{train} 和测试集 D_{test}；使用 D_{train} 作为训练数据，将 M 设为准确率为 A、ECE 为 E 且大小为 S的教师模型；
设 M_0 为从 M 获得的 IKD 基线模型，其准确率为 A_0，期望校准误差为 E_0，大小为 S_0；
现在对于 $k = 1$、2、.....、n

1. 设 M_k 为 [8] 中介绍的 IKD 流程构建的模型，设 M_k^+ 为所提出的 IKD+流程构建的模型（参见流程 1）。2. 将 M_k 的准确率、期望校准误差和大小分别设为 A_k、 E_k 和 S_k。3. 同样地，将 M_k^+ 的准确率、期望校准误差和大小分别设为 A_k^+、 E_k^+ 和 S_k^+。4. 最后进行比较：将 (A_k、 E_k、 S_k) 与 (A、 E、 S) 比较，将 (A^+、 E_k^+、 S_k^+) 与 (A、 E、 S) 比较，并将 (A_k^+、 E_k^+、 S_k^+) 与 (A_k、 E_k、 S_k) 比较。

对于我们的评估，以下细节与之相关：
– 我们将 RI- ADD Grand Challenge 的获胜者视为教师模型 [19]。不过，它是由多个 EfficientNets 组成的集成 [20]。教师模型 M_0 对于步骤3.2 ，指定为集成中准确率最高的模型，即 EfficientNet-B6；
– 评估时将 k 设为 5；
由于数据集不平衡，我们进行数据增强以平衡类别分布。增强操作包括旋转、翻转，以及调整亮度、色调和对比度；
– IKD+ 过程的 α 值设为 0。7；
– IKD+采用两种不同的技术：温度缩放和 Platt 缩放；
– 对于每个 k的取值，IKD 和 IKD+ 采用相同的精化算子 $S_k = S_k^+$；

– 所有情况下，我们均按相对于 M_0的比例报告模型规模；– 模型使用 Adam 优化器进行训练 [21], , 动量因子为 β（0.9，0.999）；– 通过实验确定超参数，批大小设为 16，学习率设为 10^{-4}。

4. 结果

我们的实验结果见表1和表2。在所考察的两种校准敏感损失中，温度缩放的表现优于Platt缩放。不过，在这两种情况下，对于大多数 $k > 0$ 值，IKD+ 模型的准确率都优于 IKD 模型。IKD+的 ECE 值也始终低于 IKD 模型。这表明，与对应的 IKD 模型相比，IKD+模型具有更高的可靠性，且预测置信度更高——见图2。还应注意，与基线模型M_0相比，IKD+得到的模型准确率相同，但体积缩小了八倍。这一结果与另一项实验一致；在该实验中，我们将集成中的次优模型 EfficientNet-B5 作为基线模型 M_0——见表4。因此可以得出结论，我们的方法通过八倍× 压缩即可达到与最先进方法相当的准确率。即使经过 128× 压缩，准确率也仅下降 5%——见表2。与最先进的集成模型相比，这一结果意义重大，例如 [7], , 其包含 20 个深度卷积网络。表2中的M_5 学生模型相较于此类模型实现了 500×压缩，同时仍保持 90% 的准确率。表3和表4报告了使用集成中的次优模型 EfficientNet-B5 作为基线模型进行 IKD+的更多结果。其准确率下降幅度略高于使用 EfficientNet-B6 进行 IKD+的情况。这种下降是意料之中的，因为 EfficientNet-B5 是 EfficientNet-B6 的 29×压缩版本。结果表明，与从 EfficientNet-B6 换用 EfficientNet-B5 相比，IKD+ 更适合获得小尺寸模型，并能在准确率、可靠性和模型大小之间实现更好的权衡。

表 1. IKD 的比较+ 以及在 EfficientNet-B6 上使用 Platt缩放校准的 IKDIKD IKD

	学生模型	教师模型	压缩			+	
				ECE	Acc	ECE	Acc
Base	M0	-	1x	1.95	0.912	0.483	0.942
Platt缩放	M1	M0	8x	1.77	0.917	0.472	0.896
	M2	M1	16x	1.899	0.892	0.459	0.931
	M3	M2	32x	1.43	0.88	0.428	0.931
	M4	M3	64x	1.61	0.801	0.412	0.889
	M5	M4	128x	1.823	0.797	0.412	0.844

5. 结论

本文提出了一种新的模型压缩算法，称为 IKD+ ， 可针对传统方法获得的预测可靠性进行精准提升

表2。使用温度缩放校准对 EfficientNet-B6 上的 IKD+ 与 IKD 进行比较

	学生教师压缩			IKD		IKD+	
				ECE	Acc	ECE	Acc
Base	M0	-	1x	1.81	0.927	0.577	0.955
温度缩放	M1	M0	8x	1.762	0.919	0.48	0.955
	M2	M1	16x	1.856	0.9	0.44	0.93
	M3	M2	32x	1.88	0.893	0.397	0.912
	M4	M3	64x	1.75	0.845	0.397	0.879
	M5	M4	128x	1.897	0.861	0.438	0.90

表3。IKD+ 与 IKD 在 EfficientNet-B5 上使用 Platt缩放校准的比较IKD IKD

	学生教师压缩					+	
				ECE	Acc	ECE	Acc
Base	M0	-	1x	1.89	0.891	0.513	0.92
Platt缩放	M1	M0	8x	1.771	0.917	0.452	0.915
	M2	M1	16x	1.625	0.918	0.409	0.913
	M3	M2	32x	1.722	0.896	0.399	0.896
	M4	M3	64x	1.781	0.822	0.397	0.83
	M5	M4	128x	1.78	0.797	0.382	0.811

表 4。使用 EfficientNet-B5 上的温度缩放校准比较 IKD+ 和 IKD

	学生教师压缩			IKD		IKD+	
				ECE	Acc	ECE	Acc
Base	M0	-	1x	1.918	0.91	0.493	0.923
温度缩放	M1	M0	8x	1.922	0.91	0.487	0.925
	M2	M1	16x	1.895	0.879	0.491	0.886
	M3	M2	32x	1.863	0.882	0.471	0.845
	M4	M3	64x	1.821	0.853	0.421	0.856
	M5	M4	128x	1.79	0.826	0.359	0.821

IKD [8]。我们在损失函数中引入温度缩放和 Platt 缩放，从而带来这些改进。模型的准确率也因此得到提升。与对应的 IKD 模型相比，IKD+还通过给出置信度更高的预测结果，提升了模型的可靠性。我们应用 IKD+ 来解决影响视网膜的疾病分类问题。目前，视网膜病变领域的最先进方法是集成模型，由多个（某些情况下多达 20 个）深度卷积神经网络组成。这些模型需要大量计算资源，不适合在现实场景中部署。我们提出的方法将集成模型的规模缩小了 500 倍，同时保持了与最先进方法相当的准确率。此外，我们的方法也更加可靠；与在不同版本的 EfficientNets 之间切换相比，它在准确率、期望校准误差和模型规模之间实现了更好的权衡。

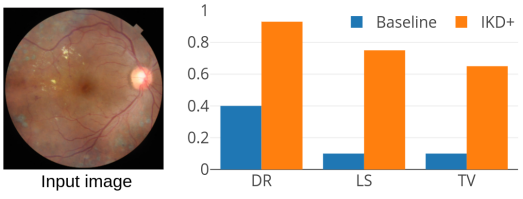

图 2。柱状图展示了使用基线和 IKD+模型将图像分类为糖尿病视网膜病变（DR）、激光瘢痕（LS）和血管迂曲（TV）的概率。

6. 参考文献

[1] Yingfeng Zheng、Mingguang He 和 Nathan Congdon，“糖尿病视网膜病变的全球流行病”，印度眼科学杂志，第 60 卷，第 5 期，第 428 页，2012 年。[2] Tien Y. Wong、Chui Ming Gemmy Cheung、Michael˜Larsen、Sanjay Sharma 和 Rafael SimA3，“糖尿病视网膜病变”，*Nature ReviewsDisease Primers*，第 2 卷，第 1 期，16012 页，2016 年 3 月。[3] Janet M. Torpy、Tiffany J. Glass 和 Richard M. Glass，“视网膜病变”，*JAMA*，第 298 卷，第 8 期，第 944–944 页，2007 年 8 月。[4] Quresh Mohamed、Mark C Gillies 和 Tien Y Wong，“糖尿病视网膜病变管理：系统综述”，*JAMA*，第 298 卷，第 8 期，第 902–916 页，2007 年。[5] Rishab Gargeya 和 Theodore Leng，“利用深度学习自动识别糖尿病视网膜病变”，*Ophthalmology*，第 124 卷，第 962–969 页，2017 年 7 月。[6] R.S. Latha、G.R. Sreekanth、R.C. Suganthe、B. Bizu、K. Suvalakshmi 和 K. Venkatachalam，“利用深度卷积神经网络根据视网膜图像诊断糖尿病视网膜病变和青光眼”，发表于*2022* 年国际计算机通信与信息学会议 *(ICCCI)*，2022 年，第 01–06 页。[7] Dominik M¨uller、I˜naki Soto-Rey 和 Frank Kramer，“基于异构深度学习模型集成的视网膜成像多疾病检测”，发表于德国医学数据科学 *2021*：*Digital Medicine: Recognize–Understand–Heal*，第 23–31 页。IOS Press，2021 年。[8] Sharan Yalburgi、Tirtharaj Dash、Ramya Hebbalaguppe、Srinidhi Hegde 和 Ashwin Srinivasan，“神经网络压缩中迭代知识蒸馏的实证研究”，发表于*ESANN*，2020 年，第 217–222 页。[9] Chuan Guo、Geoff Pleiss、Yu Sun 和 Kilian Q Weinberger，“现代神经网络的校准”，发表于*International conference on machinelearning*。PMLR，2017 年，第 1321–1330 页。

[10] Linquan Lyu、Imad Eddine Toubal 和 K. Palaniappan，《用于视网膜眼底图像多疾病检测的多专家深度网络》，载于*2022*年第*44*届*IEEE*医学与生物工程学会工程医学与生物学年会（*EMBC*），2022年，第1818–1822页。

[11] NorahAsiri、MuhammadHussain、FadwaAl Adel 和 Nazih Alzaidi，《基于深度学习的糖尿病视网膜病变计算机辅助诊断系统：综述》，医学人工智能，第 99 卷，第 101701 页，2019 年。

[12] Mohammad Z Atwany、Abdulwahab H Sahyoun 和 Mohammad Yaqub，《糖尿病视网膜病变分类的深度学习技术：综述》，*IEEE Ac-cess*，2022 年。

[13] Rishabh Patra、Ramya Hebbalaguppe、Tirtharaj Dash、Gautam Shroff 和 Lovekesh Vig，《通过显式正则化和动态数据剪枝校准深度神经网络》，载于*IEEE/CVF*计算机视觉应用冬季会议论文集，2023年，第1541–1549页。

[14] Xiaoqian Jiang、Melanie Osl、Jihoon Kim 和 Lucila Ohno-Machado，《校准预测模型估计以支持个性化医学》，美国医学信息学会杂志，第19卷，第263–274页，2012年3月。

[15] Solomon Kullback 和 Richard A Leibler，《论信息与充分性》，数学统计年鉴，第 22 卷，第 1 期，第 79–86 页，1951 年。

[16]Samiksha Pachade、Prasanna Porwal、Dhanshree Thulkar、Manesh Kokare、Girish Deshmukh、Vivek Sa-˜hasrabuddhe、Luca Giancardo、GwenolA© Quellec 和˜Fabrice MA©riaudeau，《视网膜眼底多疾病图像数据集（RFMID）：一个用于多疾病检测研究的数据集》，数据，第 6 卷，第 2 期，2021 年。[17]Adam Paszke、Sam Gross、Francisco Massa、Adam Lerer、James Bradbury、Gregory Chanan、Trevor Killeen、Zeming Lin、Natalia Gimelshein、Luca Antiga、Alban Desmaison、Andreas Kopf、Edward Yang、Zachary DeVito、Martin Raison、Alykhan Tejani、Sasank Chilamkurthy、Benoit Steiner、Lu Fang、Junjie Bai 和 Soumith Chintala，《PyTorch：一种命令式风格的高性能深度学习库》，载于神经信息处理系统进展第 *32* 卷，第 8024–8035 页。Curran Associates, Inc.，2019 年。[18]Fabian KA1⁄4ppers、Jan Kronenberger、Amirhossein Shantia 和 Anselm Haselhoff，《目标检测的多变量置信度校准》，载于*IEEE/CVF*计算机视觉与模式识别会议（*CVPR*）研讨会，2020 年 6 月。[19] 《RIADD（ISBI-2021）——Grand Challenge》，2021 年。[20] Mingxing Tan 和 Quoc Le，《EfficientNet：重新思考卷积神经网络的模型缩放》，载于国际机器学习会议。PMLR，2019 年，第 6105–6114 页。

[21] Diederik P. Kingma 和 Jimmy Ba，《Adam：随机优化的方法》，载于第三届国际学习表征会议，*ICLR 2015*，美国加利福尼亚州圣迭戈，*2015*年*5*月*7*日至*9*日，会议论文集，Yoshua Bengio 和 Yann LeCun 编，2015年。