

引言

随着深度神经网络（DNNs）在各种应用中展现出巨大潜力，如何在模型规模较大的部署到边缘设备之前进行有效的压缩技术，如权值量化和解剪枝，已经成为近年来研究的热点。尽管量化技术能够有效地减少模型大小，从而提高部署效率，但同时也可能导致性能的下降。因此，如何在量化技术与性能之间找到一个平衡点，成为了许多研究共同探讨的问题。本文首篇论文探讨了深度神经网络的量化如何用于准确的边缘计算，实验结果表明，量化技术在三个广泛使用的应用上，可以提高准确度，并在一定程度上减少内存消耗。

另一篇论文主要针对自动传染病分类问题进行研究。自动从图像中分类传染性疾病可以帮助医疗工作者识别因资源限制而未被充分诊断的疾病，以及临床医生在诊断新颖或新兴疾病（如猴痘）方面缺乏经验或洞察力。为了避免漏诊或延迟诊断，从而防止进一步传播，并改善临床结果，分析学习到的表示是必要的。该工作提出了自动发现概念的方法，即人类可解释的属性，允许在医学图像分析任务中深入理解学习信息，beyond 训练标签或协议。同时，该研究评估了代表方法在结核病预测和猴痘预测任务上的表现，并提出了NMFx，一种通过概念发现的一般NMF公式，在无监督、弱监督和监督场景中以统一方式工作。这两篇论文都为深度神经网络在边缘计算和医学图像分析领域的应用提供了新的思路和方法。

正文

深度神经网络的量化用于准确的边缘计算

近年来，深度神经网络（DNNs）在许多应用中展现出了巨大潜力，甚至超越了人类专家的表现。然而，由于其规模较大，通常在将其部署到边缘之前，会应用压缩技术，如权重量化和解构。人们普遍认为量化会导致性能下降，因此，许多现有研究都探讨了量化策略，以最小化精度损失。在本文中，作者认为量化，本质上对权重表示施加规律化，有时可以帮助提高精度。我们在三个广泛使用的应用上进行了全面的实验：全连接网络（FCN）用于生物医学图像分割、卷积神经网络（CNN）用于在ImageNet上对图像进行分类，以及循环神经网络（RNN）用于自动语音识别。实验结果表明，在三个应用上，量化分别可以提高精度1%、1.95%、4.23%，并且在3.5倍到6.4倍内存减少的情况下表现优异。

自动传染病分类分析与概念发现

自动从图像中分类传染性疾病可以协助必要的医学诊断。这种方法可以识别诸如结核病等因资源限制而未被充分诊断的疾病，以及临床医生在诊断新颖或 emerging 疾病（如猴痘）方面缺乏经验或敏锐性的疾病。避免漏诊或延迟诊断可以防止进一步传播，并改善临床结果。为了理解并信任神经网络预测，分析学习到的表示是必要的。在这个工作中，我们认为自动发现概念，即人类可解释的属性，允许在医学图像分析任务中深入理解学习到的信息，超越训练标签或协议。我们概述了医学图像和计算机视觉领域中的现有概念发现方法，并在结核病预测和猴痘预测任务上评估了代表性方法。最后，我们提出了 NMFx，一种通过概念发现的一般 NMF 公式，它在未监督、弱监督和监督的场景中都按统一方式工作。

结论

综上所述，本文研究了深度神经网络在边缘计算和医学图像分析中的应用。实验结果表明，深度神经网络的量化技术可以提高边缘设备上的性能，同时减少内存占用。在两个应用场景中，深度神经网络的自动传染病分类分析与概念发现也取得了显著的效果。自动传染病分类分析可以提高疾病的诊断效果，避免误诊或延迟诊断，从而防止疾病的进一步传播。同时，通过概念发现，可以深入理解学习到的信息，提高医学图像分析任务的性能。未来，我们期待看到更多研究将深度神经网络的量化技术应用到更多的场景，以提高性能并实现更准确的预测。