

# 引言

近年来，深度神经网络（DNNs）在众多领域已经展现出了巨大潜力，甚至超越了人类专家的表现。然而，由于其模型规模较大，通常需要在部署到边缘设备之前应用压缩技术，如权重量化和解构。虽然量化通常被认为会导致性能下降，但本文却提出量化（实质上对权重表示施加正则化）有时可以帮助提高准确度。在本文中，我们通过三个广泛使用的应用场景进行了全面实验，实验结果表明，量化在这些应用场景中分别可以提高准确度1%、1.95%、4.23%，并且在3.5倍到6.4倍的内存减少情况下表现良好。

同时，自动传染病分类分析与概念发现也是一个重要的研究方向。自动从图像中分类传染性疾病可以协助必要的医学诊断，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兴或猴痘等疾病方面缺乏经验和敏锐度。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，而提出自动发现概念的方法可以实现这一目标。在医学图像分析和计算机视觉领域中，已有多种概念发现方法，而本文则评估了代表方法，并提出了NMF<sub>x</sub>，一种通过概念发现的一般NMF形式，它可以在无监督、弱监督和监督场景中工作。

## 正文

### 深度神经网络的量化对边缘计算的准确性

近年来，深度神经网络（DNNs）在许多应用中已经展现出了巨大潜力，甚至超越了人类专家的表现。然而，由于其模型规模较大，通常需要在部署到边缘设备之前应用压缩技术，如权重量化和解构。虽然量化通常被认为会导致性能下降，但本文却提出量化（实质上对权重表示施加正则化）有时可以帮助提高准确度。

作者在三个方面进行了全面的实验，包括用于生物医学图像分割的完全连接网络（FCN）、用于ImageNet上图像分类的卷积神经网络（CNN）以及用于自动语音识别的循环神经网络（RNN）。实验结果显示，在三个应用上，量化分别可以提高精度1%、1.95%、4.23%，并且在与3.5x-6.4x内存减少的比较下，量化效果更为显著。

### 自动传染病分类分析与概念发现

自动从图像中进行传染病分类有助于进行必要的医学诊断，这可以识别出因资源限制而未被充分诊断的疾病，如结核病，以及临床医生在诊断中缺乏经验或敏锐度的创新或新兴疾病，如猴痘。避免或延迟诊断可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，需要分析学习到的表示。在这项工作中，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对学习到的信息有深刻的理解，beyond 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NMF<sub>x</sub>，一种通过概念发现获得的具有统一性的解释性的一般NMF公式，在无监督、弱监督和监督场景中都有效。

## 结论

结论：

1. 深度神经网络（DNNs）在许多应用中展现出巨大潜力，但由于其规模较大，通常需要在部署到边缘设备之前应用压缩技术，如权重量化和解构。
2. 量化，即对权重表示施加规律化的过程，有时可以帮助提高准确率。作者在三个广泛使用的应用上进行了实验，结果表明，量化在这三个应用中的准确率分别提高了1%、1.95%、4.23%，并且在3.5倍到6.4倍的内存减少之间。
3. 自动传染病分类分析与概念发现可以协助必要的医学诊断，识别因资源限制而未被充分诊断的疾病，也可以识别临床医生在诊断方面缺乏经验或智慧的崭新的疾病。
4. 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。自动发现概念（即人类可解释的特征）允许在医学图像分析任务中对学习到的信息有深刻的理解。
5. 最后，作者提出了NMF<sub>x</sub>，一种通过概念发现获得