Computer Science and Application 计算机科学与应用, 2025, 15(5), 606-614
Published Online May 2025 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2025.155133

Hans汉斯

基于全局通道数剪枝的Wav2Lip模型轻量化的研究

徐康杰[1]，陈云翔[1,2]，张　龙[1,2]，唐　帅[1]，周庆华[1]

[1]长沙理工大学物理与电子科学学院，湖南 长沙
[2]深圳媲美科技有限公司，广东 深圳

收稿日期：2025年4月12日；录用日期：2025年5月14日；发布日期：2025年5月23日

摘　要

针对Wav2Lip模型计算量大，推理速度慢，在一些对实时性要求较高或算力较为有限的应用场景中可能难以满足预期效果等问题，论文提出了基于全局通道数剪枝的方法，选用了三种不同剪枝比例，对Wav2Lip模型进行了全局通道数剪枝并对比。实验结果表明，论文提出的全局通道数剪枝方案成功地：1) 提升了推理速度；2) 减小了模型体积；3) 保持或提升了所生成图像的效果。该方案在降低计算成本的同时，能够实现高效且稳定的推理性能。

关键词

Wav2Lip，深度学习，模型轻量化，全局通道数剪枝

Research on Lightweight Wav2Lip Model Based on Global Channel Number Pruning

Kangjie Xu[1], Yunxiang Chen[1,2], Long Zhang[1,2], Shuai Tang[1], Qinghua Zhou[1]

[1]School of Physics and Electronic Science, Changsha University of Science and Technology, Changsha Hunan
[2]Pimei Technology Co., Ltd., Shenzhen Guangdong

Received: Apr. 12th, 2025; accepted: May 14th, 2025; published: May 23rd, 2025

Abstract

In response to the issues of high computational complexity, slow inference speed, and potential difficulty in achieving expected results in some application scenarios that require high real-time performance or limited computing power for the Wav2Lip model, the paper proposes a method based on global channel pruning, using three different pruning ratios to perform global channel pruning

文章引用: 徐康杰, 陈云翔, 张龙, 唐帅, 周庆华. 基于全局通道数剪枝的 Wav2Lip 模型轻量化的研究[J]. 计算机科学与应用, 2025, 15(5): 606-614. DOI: 10.12677/csa.2025.155133

on the Wav2Lip model and compare them, the experimental results show that the global channel pruning scheme proposed in the paper successfully: 1) improves inference speed; 2) Reduced the size of the model; 3) Maintained or improved the effect of the generated image. This solution can achieve efficient and stable inference performance while reducing computational costs.

Keywords

Wav2Lip, Deep Learning, Model Lightweighting, Global Channel Pruning

1. 引言

语音驱动视频口型技术是一种通过音频信号来驱动和生成视频中的口型动画的方法。此类技术在数字人视频内容制作等方面得到了广泛应用。其中，基于生成式人工智能的 Wav2Lip 模型可以用来生成与音频内容高度同步的口型视频[1]。然而，Wav2Lip 模型巨大，可能存在大量的冗余参数。我们分析了 Wav2Lip 模型的结构，发现 Wav2Lip 的生成器层数过深，导致生成器部分聚集了大量冗余参数。根据 Kim B K 等的分析，Wav2Lip 生成器中的每一层的大量冗余信息和参数穿梭在整个网络当中[2]。这些参数对计算效率和推理速度造成了负担，特别是在一些资源有限的设备上(如移动设备和嵌入式系统)运行时，可能无法达到实时或高效的表现[3]。因此，如何在保证生成效果的前提下减少冗余参数、优化网络结构、加快推理速度是我们需要解决的问题。

深度学习发展越来越快，深度卷积神经网络在许多现实应用中的部署中很大程度上受到了极高的计算成本的阻碍，神经网络的轻量化在当下得到了更广泛的关注[4]，其已成为一个至关重要的研究领域[5]。剪枝是轻量化网络的其中一种较为有效的方法[6]，Mishra R 等表明通过模型剪枝，可以精剪模型[7]，从而提高推理速度，降低模型大小。剪枝最早可以追朔到 1988 年，Hanson S 等[8]从过度参数化的模型中删除不重要的权重。直到 2015 年[9]，研究界才逐渐意识到剪枝应用在消除深度神经网络中显著冗余方面的潜力是巨大的，大量关于更深度的模型剪枝方法的文献涌现了出来，其中比较典型的剪枝文献[10]成功地将剪枝方法应用于判别模型，然而，将它们应用在生成模型的研究相对较少，Shu H 等[11]在 2019 年首次提出针对 GAN 中生成网络的剪枝算法，他们使用生成模型参数冗余建模，借鉴传统的剪枝算法，直接最小化压缩生成模型前后的重建误差来获得压缩后的模型。

在本文中，我们提出了一种对 Wav2Lip 模型有效网络剪枝训练方案——全局通道数剪枝。我们研究在 Wav2Lip 生成器上进行剪枝对 Wav2Lip 模型推理速度和效果的影响。针对 Wav2Lip 的生成器，我们分别对剪枝率为 20%、50%和 75%三种不同剪枝方案进行实验比较，寻找在模型大小、推理速度和推理效果方面展现更为全面的剪枝比例方案。实验结果表明，该剪枝策略有效解决了在有限的设备、资源下，部署 Wav2Lip 时面临的种种挑战。

2. Wav2Lip 生成器剪枝方法

本文将专注于对 Wav2Lip 模型的生成器进行剪枝。图 1 是由编码器–解码器结构构成的 Wav2Lip 生成器模型。Wav2Lip 的生成器包含一个由人脸编码器和音频编码器组成的双路径编码器(图 1 虚线框中的部分)，其中人脸编码器以参考帧与姿态先验帧串联，音频编码器以一段语音作为输入。Wav2Lip 生成器

两个编码器的通道数(即滤波器的数量)均随着网络层数的增加而增加，其中，人脸编码器的滤波器个数从16个增加到512个，音频编码器的滤波器个数从32个增加到512个。人脸编码器和解码器之间有跳跃连接，这些连接会将人脸编码器的某一层的输出直接传递到解码器的对应层，解码器的滤波器个数与编码器的滤波器个数呈现反向对称的关系，其滤波器个数逐渐减小，从512个减少到64个。这个模型的深度较大，其信息在每一层的卷积核中得以存储，层与层之间传递着大量的参数和信息，部分参数对于高保真的图像生成可能并非必需[12]。减少生成器各卷积核的通道数有可能简化整个模型，降低计算消耗，加快模型的推理速度[13]。

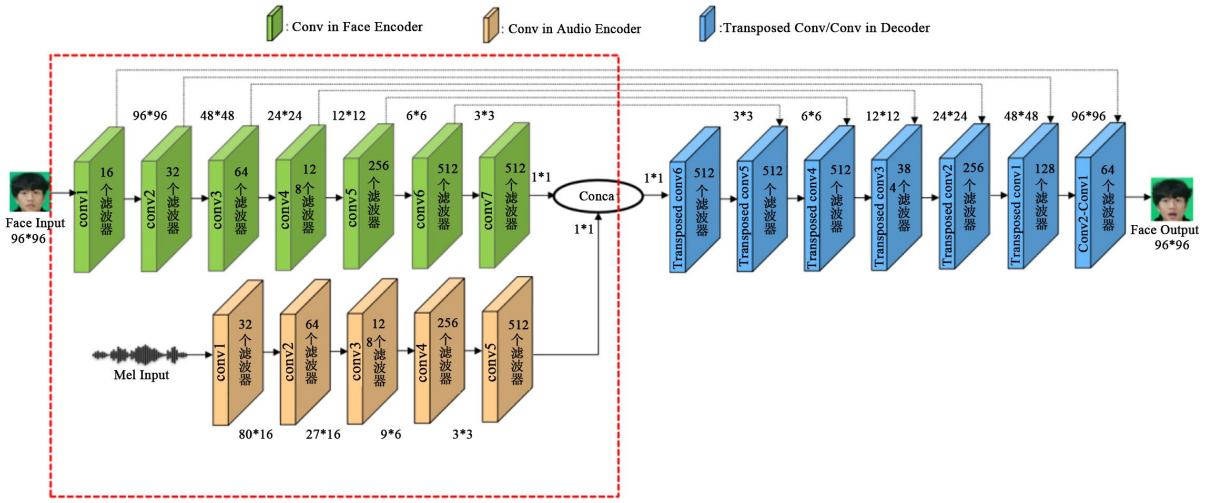


Figure 1. The generator structure of Wav2Lip model
图 1. Wav2Lip 模型的生成器结构

2.1. 编码器的滤波器重要性排序

在神经网络训练过程中，随着各层参数的不断更新，每一层的输入数据分布会不断发生变化，这被称为“内部协变量偏移”。这种变化会使得网络的训练变得困难，因为网络需要不断适应新的数据分布。批标准化BN (Batch Normalization)层的基本思想是对每一层的输入进行标准化处理，使得其均值为0，方差为1，从而把一个小批量(mini-batch)内的所有数据，从不规范的分布拉到正态分布。批标准化的好处是使得后续的激活函数作用在一个更加均匀、稳定的数据分布上，可以在一定程度上避免梯度爆炸或梯度消失的问题。具体来说，对于一个小批量的数据，计算其输入数据的均值和方差，然后使用式(1)对该批数据进行标准化：

$$\hat{x}^{(k)} = \frac{x^{(k)} - \mu}{\sqrt{\sigma^2 + \varepsilon}} \tag{1}$$

其中，$\hat{x}^{(k)}$是第k个样本对应的卷积层的输出值，作为BN的输入。μ是该批数据的均值，σ^2是其方差，ε则是一个很小的常数，用于防止分母为0。在标准化后，为了保证模型具有足够的表达能力，使用式(2)引入两个可学习的参数γ (缩放参数)和β (偏移参数)，对标准化后的数据进行线性变换。两个参数γ和β是模型在训练的时候所得到的，每进行一次BN层的操作，都会得到一个γ和β值。

$$y^{(k)} = \gamma^{(k)}\hat{x}^{(k)} + \beta^{(k)} \tag{2}$$

Li H等发现，较小权重的滤波器对输入图像的响应较弱，导致卷积操作后的输出值较小，经过激活

函数处理后，这些输出值可能会变得非常小，甚至为零，从而意味着网络没有充分学习到有效的特征[14]。去除这些滤波器对模型性能的影响较小，甚至几乎不会下降。因此，他们提出可以基于权重大小进行滤波器剪枝。我们参考该方法，把 BN 层中线性变换的缩放因子视作每一个人脸编码器和音频编码器的滤波器的重要性因子，并基于此重要性因子对 Wav2Lip 进行剪枝。图 2 是剪枝之前的 Wav2Lip 模型人脸编码器和音频编码器卷积层与 BN 层，其中 $C \in R^{n_{in} \times n_{out} \times k \times k}$ 表示卷积层的核矩阵，n_{in}、n_{out}、k 分别是输入、输出通道数和空间核大小。对于第 i 层卷积层的第 m 个滤波器 C_{im}，可求出其权重值 γ_{im}。我们仿照 Liu 等[15]的做法，对 γ 的值进行从小到大排序：γ 越小，则说明其对应的滤波器越不重要。为了使 γ 值的排序呈现较大的差异性，拉大排序分值差距，我们还使用了 L1 正则化对其进行稀疏化，以便更好地筛选不重要的通道数。我们把 L1 正则化项加入训练的总损失，训练模型时的总损失函数为

$$L_{total} = L_{original} + \lambda \sum_i |\gamma_{im}| \tag{3}$$

其中，$L_{original}$ 是原始 Wav2Lip 网络的损失函数，γ_{im} 表示第 m 个滤波器对应的权重值，λ 是 L1 正则化的超参数，用于控制稀疏化的程度。我们对训练了 51000 步的模型进行了 L1 正则化前后的取样分析对比。图 3 分别是稀疏化前后训练到 51000 步时，模型的 γ 值分布。图中的横轴代表了 γ 的值，纵轴代表了数量。从图 3(a)可以看出，在稀疏化前，即 $\lambda = 0$ 时，γ 的值大部分分布在 1 附近，其值较为集中。我们取 $\lambda = 0.001$ 进行稀疏化，从图 3(b)可以看出稀疏化使 γ 的值大部分分布在了 0 附近，且 0~1 之间的值分布较广泛，比稀疏化前更平均。为了更直观地比较稀疏化前后的 γ 分布，图 4 显示了从开始训练到训练了 51000 步的过程中，γ 值的分布。从图 4(a)可以看出，在稀疏化前随着训练步数的增加，γ 值有所分散，但大致集中在 1 附近，其中位数约为 1。在取 $\lambda = 0.001$ 进行稀疏化后，随着训练步数的增加，γ 值明显分散，且中位数趋于 0，见图 4(b)。超参数 λ 值的增加会导致模型的精度下降，从图 3 与图 4 的 γ 值分布来看，在取 $\lambda = 0.001$ 时，已有较好的稀疏效果，所以我们在后面的剪枝过程中，把 λ 的值设为 0.001，无需取太大的值。

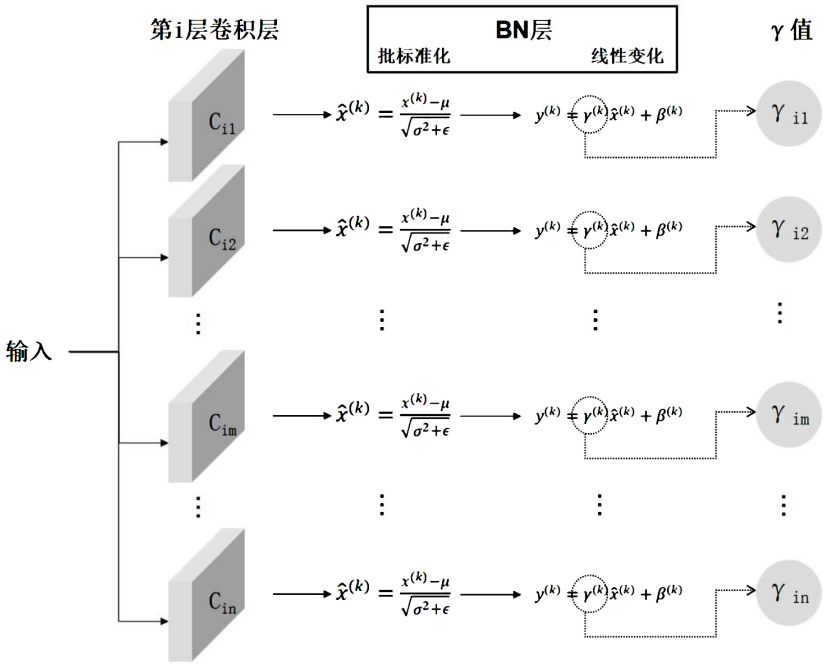
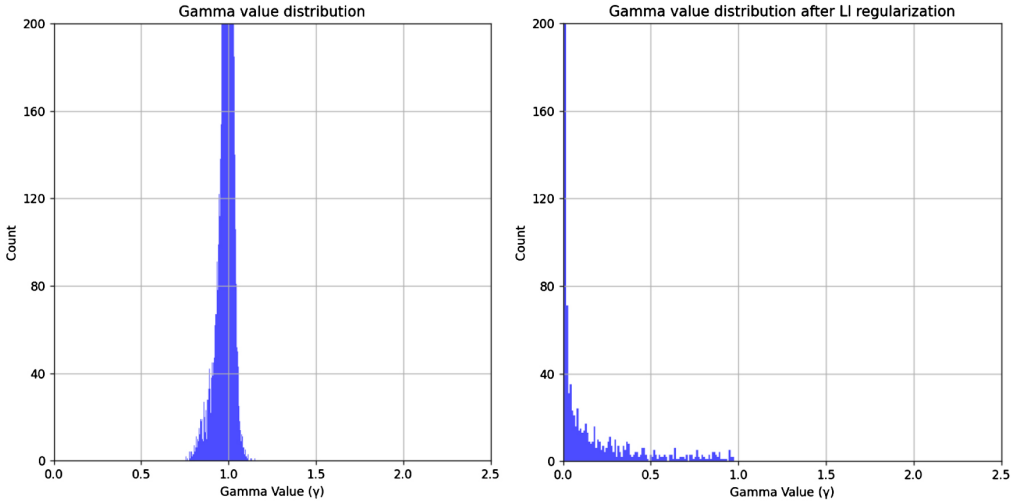
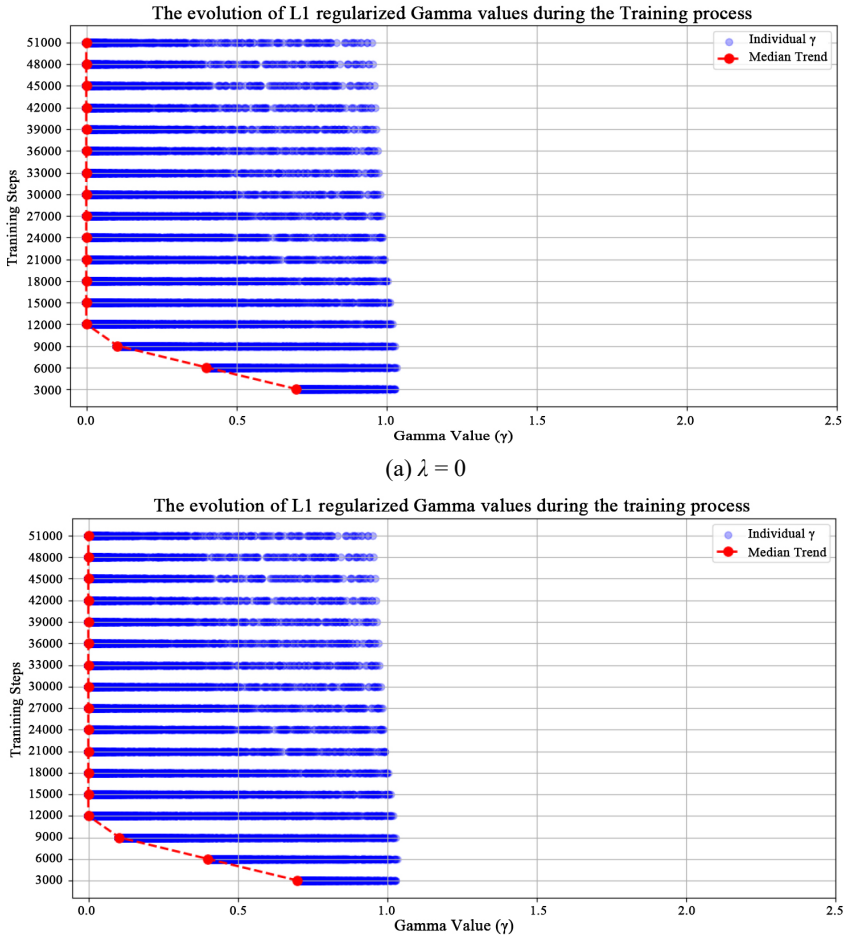


Figure 2. Convolutional and BN layers of facial and audio encoders before pruning
图 2. 剪枝前人脸、音频编码器的卷积层和 BN 层



(a) $\lambda = 0$ (b) $\lambda = 0.001$

Figure 3. Distribution of BN layer gamma values before and after sparsity during 51000 training steps
图 3. 训练 51000 步时稀疏化前后的 BN 层 γ 值分布



(a) $\lambda = 0$

(b) $\lambda = 0.001$

Figure 4. Distribution of BN layer gamma values before and after sparsity under different training steps
图 4. 不同训练步数下稀疏化前后的 BN 层 γ 值分布

2.2. 生成器的剪枝与模型再训练

我们根据需要保留的通道数比例，确定了编码器的权重阈值 γ_T。我们剪枝编码器中所有权重低于该阈值的滤波器，同时，人脸解码器中同样比例的滤波器也被相对应地去除，以保证编码器与解码器之间输入输出的匹配。由于每一层滤波器的数量，决定这一层输出的通道数，因此去除一个滤波器，实质上相当于删除一个通道的所有输入和输出连接，于是通过剪枝，我们可以获得一个新的通道数较少的网络。

我们以 50%剪枝为例，给出剪枝后的网络结构。由于 Wav2Lip 主干网络有人脸编码器、音频编码器和解码器三层结构，我们仅展示剪枝后主干网络的人脸编码器的结构。剪枝后的人脸编码器网络结构如表 1 所示。剪枝后的模型保持了与原模型相同的层数和卷积核大小，但每一层的通道数较原模型有所减少，表里的红色括号代表了剪枝前原模型通道数结构，例如，第一层从 16 个输出通道减少到了 8 个。

Table 1. 50% pruned facial encoder network structure
表 1. 50%剪枝后的人脸编码器网络结构

Layer	Input Size	Operator (Kernel, In Channels, Out Channels, Residual)	Stride	Padding
1	- (depends on input)	Conv2d (7 × 7, 6,8)	1	3
2	96 × 96 × 8 (96 × 96 × 16)	Conv2d (3 × 3, 8, 16)	2	1
2	48 × 48 × 16 (96 × 96 × 32)	Conv2d (3 × 3, 16, 16, True)	1	1
2	48 × 48 × 16 (96 × 96 × 32)	Conv2d (3 × 3, 16, 16, True)	1	1
3	48 × 48 × 32 (96 × 96 × 64)	Conv2d (3 × 3, 16, 32)	2	1
…	……	……	…	…
7	1 × 1 × 256 (96 × 96 × 512)	Conv2d (1 × 1, 256, 256)	1	0

剪枝操作改变了神经网络的结构，显著减少了模型的复杂度，但是会导致其性能有所下降，因此有必要对剪枝后的神经网络进行再训练。这一过程旨在通过调整剩余的权重，提升剪枝后模型的精度。

3. 实验结果分析

3.1. 数据集和评价指标

在 Wav2Lip 模型的训练中，我们选择了 LRS3 数据集作为基础数据集[16]，该数据集包含来自约 5000 个不同 TED 演讲者的约 187000 个视频片段。为了进一步丰富训练数据，我们还录制了一部分高质量的中文视听数据集。该数据集包含来自 278 名不同年龄、性别和背景的中文说话人的约 7300 个视频片段。我们将中文数据集与 LRS3 数据集混合使用，从而增强了模型的泛化能力。在数据预处理方面，我们对每个视频进行了切割，并将其转化为单帧图像，以便用于模型的训练和评估。此外，为了使模型更好地适应中文语音特征，我们采用了分片的方式对数据进行处理。这些数据片段不仅能提供丰富的语音与视觉信息，还帮助模型更精准地学习中文语音和面部动作之间的关系，从而提高了模型在中文语音生成的表现。

为了评估生成效果，我们采用了 Frechet Inception Distance (FID)、置信度(LSE-C)和唇形同步误差距离(LSE-D)作为评价指标。FID 能够量化生成数据与真实数据之间的相似度，从而为我们提供了一个衡量生成质量的标准[17]。LSE-C 和 LSE-D 是 Wav2Lip 中利用 syncnet 网络提出的新指标，其可以评估语音和生成的人脸样本之间的唇动同步质量[18]。通过这些指标，我们能够准确评估模型在语音同步中的表现，并不断优化模型的生成效果。

3.2. 对比实验

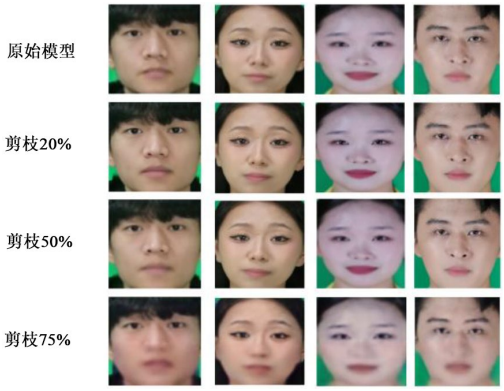
我们为 Wav2Lip 模型的生成器设置了三个不同的剪枝比例，使图 1 中的视觉编码器和音频编码器中的通道数在剪枝后分别减少了 20%、50%和 75%，并通过对比实验来确定合理的剪枝比例。表 2 给出了不同剪枝比例下的定量结果，↑和↓分别表示较高和较低的值更好。从表 1 可以看出，剪枝比例越大，模型的数据量和推理时间越少。剪枝比例为 20%、50%和 75%时，模型的大小分别下降到原始模型的 94%、24%和 6%；其对应的推理时间也分别下降到原始模型的 96%、55%和 45%。

在 FID 得分方面，当剪枝比例为 50%以内时，其 FID 得分逐渐略微下降，即生成图像质量略优于原模型，这可能与剪枝减轻过拟合的能力和删除有缺陷的权重有关[19]；但是在剪枝比例为 75%时，其 FID 分数迅速上升，导致图像质量变差，这是由于剪枝程度过重导致的。在 LSE-C 得分方面，剪枝比例为 50%时，其得分优于原始模型，但略低于 20%剪枝(差距为 8%)，这一差距在可接受范围内；在 LSE-D 得分方面，50%的剪枝比例得分优于原模型，但比 20%剪枝略逊色(差距为 7%)，这一差距同样在可接受范围内。然而，在 75%的剪枝比例下，LSE-C 的得分显著变小、LSE-D 的得分显著变大，说明推理效果有可能明显变差。

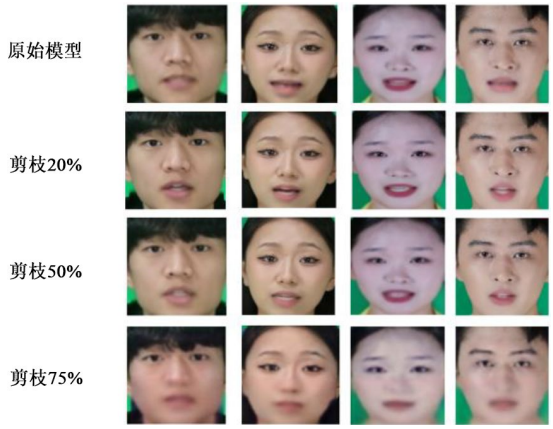
Table 2. Comparison of model size, inference time, FID, LSE-C, LSE-D performance under different pruning ratios
表 2. 不同剪枝比例下的模型大小、推理时间、FID、LSE-C、LSE-D 性能比较

剪枝比例/%	模型大小/KB↓	推理时间/S↓	FID↓	LSE-C↑	LSE-D↓
0	425,594	13.20	56.85	5.2963295	10.929478
20	401,837	7.87	53.92	**7.9609256**	**7.93514**
50	106,811	4.59	**47.47**	7.220083	8.572339
75	**27,318**	**3.7**	111.63	0.24874306	18.790245

在给定语音的情况下，我们使用四张不同人脸图像进行推理。图 5(a)(b)分别展示了在闭唇和开唇两种状态下，采用不同剪枝比例进行推理的结果。闭唇状态下，在 20%和 50%的剪枝比例时，生成的唇语同步人脸图像仍然与原始模型保持了高度的视觉一致性，几乎看不到与原始未裁剪模型之间的视觉差异，而在 75%的剪枝比例下，生成图像明显模糊，质量变差；在开唇状态下，唇部的张开程度决定了生成图像中口腔的可见度，剪枝比例较低时，模型能够较好地保留唇形和口腔的自然开度，而随着剪枝比例增大，唇形的张开程度会略有失真，导致口腔的可见度略有变化，在 75%的剪枝比例时，唇形完全失真。进一步地，我们对开唇状态下的牙齿清晰度进行了分析。随着剪枝比例的增加，开唇状态下牙齿的细节逐渐模糊，尤其是 75%的剪枝比例时，牙齿的轮廓完全不可见。但值得注意的是，在 50%剪枝比例下，牙齿的清晰度基本保持，且与原始未裁剪模型相比差异不大。因此，50%以内的剪枝比例对牙齿细节的影响较小。



(a) 闭唇



(b) 开唇

Figure 5. Inference results under different pruning ratios
图 5. 不同剪枝比例下的推理结果

上述分析结果表明，在 50%的剪枝比例以内，我们的剪枝方法在成功地提高了模型推理的效率同时，还保持了图像的生成质量。在综合考虑各项指标后，我们认为将剪枝阈值设定为 50%是最为合理的方案。在这个剪枝比例下，模型的整体性能优于原始模型。尽管与 20%剪枝比例相比，50%的剪枝在精度和唇形同步性方面略有下降，但这一差距仍在可接受范围内。同时，相较于 20%剪枝，50%的剪枝显著降低了模型大小和推理时间，分别减少了 74%和 42%。

经过我们的全局通道数剪枝后，Wav2Lip 不仅在模型大小上取得了显著的压缩，而且在推理速度方面也表现出了明显的优势。尤其是在移动设备等计算资源受限的轻量级硬件上，这种压缩使得模型能够更高效地部署，并有效减少了推理过程中的延迟。这对于实际应用，特别是在需要实时唇语同步的视频生成或语音交互系统中，具有非常重要的意义。

4. 结语

本文提出了一种全局通道数剪枝的方法，并将其应用于 Wav2Lip 模型。该方法通过对批标准化层中的缩放因子进行稀疏诱导正则化，得到一系列可排序的值，从而实现对不重要通道的剪枝。通过全局通道数剪枝，对比分析不同的剪枝比例，并结合模型大小、推理速度、FID、LSE-C、LSE-D 等指标进行全面评估。我们发现，对于 Wav2Lip 模型而言，当剪枝比例不超过 50%时，剪枝后的模型的各项指标均优于原始模型，且精度不会下降。在剪枝比例为 50%时，模型的推理达到了最优的效能，在这一剪枝比例下，Wav2Lip 模型的大小缩小至原模型的 25%，推理耗时减少至原模型的一半以上，同时推理精度不仅未降低，反而有所提高。总体而言，通过全局通道数剪枝，能够有效减少模型的参数量，压缩模型体积，同时显著提升推理速度，而且在保持生成质量的同时，某些任务中的生成质量甚至有所提升。

本研究为 Wav2Lip 模型的优化提供了一种行之有效的剪枝方案，并为生成对抗网络以及其他深度学习模型的优化提供了新的思路和方法。我们期望这一成果能够为未来相关研究提供有价值的参考与启示。

基金项目

国家自然科学基金项目(42074198)。

参考文献

[1] Prajwal, K.R., Mukhopadhyay, R., Namboodiri, V.P. and Jawahar, C.V. (2020) A Lip Sync Expert Is All You Need for

Speech to Lip Generation in the Wild. *Proceedings of the* 28*th ACM International Conference on Multimedia*, Seattle, 12-16 October 2020, 484-492. https://doi.org/10.1145/3394171.3413532

[2] Kim, B.K., Choi, S. and Park, H. (2022) Cut Inner Layers: A Structured Pruning Strategy for Efficient U-Net Gans. arXiv:2206.14658.

[3] 林景栋, 吴欣怡, 柴毅, 等. 卷积神经网络结构优化综述[J]. 自动化学报, 2020, 46(1): 24-37.

[4] Mathew, M., Desappan, K., Swami, P.K. and Nagori, S. (2017) Sparse, Quantized, Full Frame CNN for Low Power Embedded Devices. 2017 *IEEE Conference on Computer Vision and Pattern Recognition Workshops* (*CVPRW*), Honolulu, 21-26 July 2017, 328-336. https://doi.org/10.1109/cvprw.2017.46

[5] 毕鹏程, 罗健欣, 陈卫卫. 轻量化卷积神经网络技术研究[J]. 计算机工程与应用, 2019, 55(16): 25-35.

[6] Reiners, M., Klamroth, K., Heldmann, F. and Stiglmayr, M. (2022) Efficient and Sparse Neural Networks by Pruning Weights in a Multiobjective Learning Approach. *Computers & Operations Research*, **141**, Article 105676. https://doi.org/10.1016/j.cor.2021.105676

[7] Mishra, R., Gupta, H.P. and Dutta, T. (2020) A Survey on Deep Neural Network Compression: Challenges, Overview, and Solutions. arXiv:2010.03954.

[8] Hanson, S. and Pratt, L. (1988) Comparing Biases for Minimal Network Construction with Back-Propagation. *Advances in Neural Information Processing Systems* 1, 1 January 1988, 177-185.

[9] Han, S., Mao, H. and Dally, W.J. (2015) Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding. arXiv:1510.00149.

[10] Frankle, J. and Carbin, M. (2018) The Lottery Ticket Hypothesis: Finding Sparse, Trainable Neural Networks. arXiv:1803.03635.

[11] Shu, H., Wang, Y., Jia, X., Han, K., Chen, H., Xu, C., *et al*. (2019) Co-Evolutionary Compression for Unpaired Image Translation. 2019 *IEEE*/*CVF International Conference on Computer Vision* (*ICCV*), Seoul, 27 October 2019-2 November 2019, 3234-3243. https://doi.org/10.1109/iccv.2019.00333

[12] 张良, 张增, 等. 基于 YOLOv3 的卷积层结构化剪枝[J]. 计算机工程与应用, 2021, 57(6): 131-137.

[13] 黄文斌, 陈仁文, 袁婷婷. 改进 YOLOv3-SPP 的无人机目标检测模型压缩方案[J]. 计算机工程与应用, 2021, 57(21): 165-173.

[14] Li, H., Kadav, A., Durdanovic, I., *et al*. (2016) Pruning Filters for Efficient Convnets. arXiv:1608.08710.

[15] Liu, Z., Li, J., Shen, Z., Huang, G., Yan, S. and Zhang, C. (2017) Learning Efficient Convolutional Networks through Network Slimming. 2017 *IEEE International Conference on Computer Vision* (*ICCV*), Venice, 22-29 October 2017, 2755-2763. https://doi.org/10.1109/iccv.2017.298

[16] Afouras, T., Chung, J.S. and Zisserman, A. (2018) LRS3-TED: A Large-Scale Dataset for Visual Speech Recognition. arXiv:1809.00496.

[17] Heusel, M., Ramsauer, H., Unterthiner, T., *et al*. (2017) Gans Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. *Advances in Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 6629-6640.

[18] Chung, J.S. and Zisserman, A. (2017) Out of Time: Automated Lip Sync in the Wild. In: Chen, C.S., Lu, J. and Ma, K.K., Eds., *Lecture Notes in Computer Science*, Springer International Publishing, 251-263. https://doi.org/10.1007/978-3-319-54427-4_19

[19] Tousi, A., Jeong, H., Han, J., Choi, H. and Choi, J. (2021) Automatic Correction of Internal Units in Generative Neural Networks. 2021 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Nashville, 20-25 June 2021, 7928-7936. https://doi.org/10.1109/cvpr46437.2021.00784