JAZZNET：用于音乐音频机器学习研究的基础钢琴模式数据集

Tosiron Adegbija

美国亚利桑那州图森市亚利桑那大学电气与计算机工程系 tosiron@arizona.edu

arXiv:2302.08632v1 [cs.SD] 17 Feb 2023

摘要

本文介绍了*jazznet* 数据集，这是一个包含基础爵士钢琴音乐模式的数据集，可用于开发音乐信息检索（MIR）机器学习（ML）算法。该数据集包含 162520 个带标签的钢琴模式，包括和弦、琶音、音阶以及带有转位的和弦进行，总计超过 26k 小时的音频，数据集大小为 95GB。本文介绍了数据集的组成、创建与生成过程，并提供了一个开源模式生成器，它采用名为基于距离的模式结构（*DBPS*）的方法，研究人员只需定义音乐模式中各音高之间的距离，即可轻松生成新的钢琴模式。我们通过使用卷积循环神经网络（CRNN）和深度卷积神经网络进行实验，展示了该数据集如何帮助研究人员针对具有挑战性的音乐信息检索任务对新模型进行基准测试。数据集和代码可通过以下链接获取：

https://github.com/tosiron/jazznet.

索引词— 音乐信息研究、机器学习、爵士钢琴数据集、大数据

1. 引言

音乐领域机器学习（ML）研究最重要、最基本的需求之一，是拥有可免费使用的数据集。遗憾的是，在高质量数据集的获取方面，音乐领域落后于其他领域（如图像和语音识别）。本文通过介绍 *jazznet* 数据集，为不断增长的可用音乐音频数据集作出贡献。该数据集可扩展，包含基础钢琴模式。

目前已有不少致力于创建音乐信息检索（MIR）数据集的工作，而 jazznet 与这些工作相辅相成。现有的一些音乐数据集包括 GTZAN [1], MSD [2], AudioSet [3], FMA 数据集 [4], MusicNet [5], 和 RWC[6]。jazznet 数据集与大多数现有数据集的主要区别在于其创建数据集的方法：侧重于基础模式，而非完整的音乐作品。这种方法的灵感来自人类有效学习钢琴音乐的方式。假设你想学习演奏爵士钢琴；如果先理解构成完整音乐作品基础的基本模式——和弦、琶音、音阶，以及和弦进行——你就能学得更快、更有效 [7]。我们专注于爵士钢琴模式，使数据集创建切实可行，同时也因为爵士乐具有多样性——爵士乐中的模式涵盖了其他几种音乐流派（例如布鲁斯、乡村音乐、流行音乐等）。鉴于表现形式多样是爵士乐的重要特征，数据集还包含了这些模式可以演奏的不同形式——即转位。

该数据集的多类别特性也旨在模拟人类学习爵士乐时所经历的难度层级。例如，人们很容易区分不同类型的模式（如和弦、琶音和音阶）。而在每个类别内部，要识别不同模式的模式则困难得多（如大调与小调，或增和弦与减和弦）。最后，只有天赋异禀且具备绝对音高 [8] 的音乐家，才能识别具体的模式（如 A 大调与 C 大调）。

总之，我们做出了两项主要贡献。首先，我们介绍了 jazznet 数据集[1]，。该数据集是基于对爵士乐教育资源和爵士标准曲的广泛调查整理而成。数据集包含 162520 个自动生成并标注的钢琴音乐模式，总计 95GB，音频时长超过 2.6 万小时。jazznet 数据集的主要目标是促进用于应对具有挑战性的音乐信息检索任务的机器学习模型开发。其次，我们开发了一个模式生成器，它使用基于距离的模式结构（*DBPS*），根据模式内音高之间的距离生成钢琴模式。这种方法可以轻松生成用于机器学习研究的新钢琴模式。为方便用户，我们已将实现该方法的 Python 脚本[2] 开源。

2. JAZZNET 数据集

尽管与现有音乐数据集互为补充，jazznet仍在多个方面独具特色。从结构上看，与jazznet最为相关的数据集是NSynth [9] 和MAPS[10] 数据集，但两者也存在一些显著差异。NSynth旨在探索不同音高、音色和包络的音符的神经音频合成。然而，它不包含任何音乐模式，只有单个音符；尽管它涵盖的乐器和力度比jazznet更多，但就其中音乐模式的多样性而言，两者并不可比。另一方面，MAPS数据集常用于自动音乐转录 [11]，其中包含独立的

[1]https://doi.org/10.5281/zenodo.7192653[2] https://github.com/tosiron/jazznet/tree/main/ PatternGenerator



图 1。C 大调和弦及其三种形式的示意图：根音、第一转位和第二转位

音符、和弦和音乐作品。然而，它包含的和弦比 jazznet少，也不包括音阶、琶音或和弦进行。此外，MAPS数据集明显小于jazznet，数据量为31GB，而 jazznet为95GB。最重要的是，与以往的数据集不同，jazznet还附带钢琴模式生成器，使研究人员能够扩展数据集，加入新的模式。

虽然该数据集在设计上易于扩展，但我们通过广泛调查多种爵士乐演奏与教学资源，为当前版本整理了这些模式，例如《爵士钢琴书》 [7],《爵士乐理书》 [12]，以及 50 多首爵士标准曲目 [13, 14]。我们还与多位爵士乐音乐家交流，就一些常见的基础模式，尤其是和弦进行，达成共识。不过，我们也承认，鉴于爵士乐的复杂性与多样性，该数据集并未涵盖所有爵士钢琴模式。尽管如此，所提供的模式生成器（第 2.5 节）使用户能够生成数据集中没有的新模式。在以下各节中，我们将介绍理解该数据集创建过程所需的乐理背景，描述数据生成方法，随后展示 jazz net 数据集及其特征、模式生成器和潜在应用。

2.1. 简要背景：基础乐理

该数据集基于标准八个八度（88 键）钢琴的音符。钢琴（以及大多数其他弦乐器）的主要音符通常用 7 个字母表示： A，B，C，D，E，F，G（对应钢琴上的白键）。这些音符之间的音符（即黑键）称为紧接在前一个音符的升号（♯），或紧接在后一个音符的降号（♭）。例如，C 和 D 之间的音符根据语境称为 C♯ 或 D♭ 。八度是两个音高之间的音程，其频率相差一半或一倍。例如，钢琴上的中央 C（也称 C4）频率为 261.63 Hz，下一个 C（C5）频率为 523.25 Hz，而前一个 C（C3）频率为 130.81 Hz。每个音符与其八度之间有 12 个“半音”（也称半音）或 6 个“全音”（也称全音）。

该数据集由四种钢琴模式组成：和弦、琶音、音阶以及和弦进行。和弦是由多个同时演奏或听到的音符组成的音高或频率的和声音集。根据所包含的音符数量，可以将和弦分为：二音和弦（2个音符）、三和弦 （3个音符）、七和弦 （4个音符）等。琶音是音符依次演奏的和弦。音阶是按照音符频率排列的音符集合。可以重新排列这些模式中的音符，以创造不同的音乐“色彩”，这种排列称为转位 （图1以C大调三和弦为例进行了说明）。此外，新手爵士钢琴家最先且最重要的学习内容之一

算法 1： 基于距离的模式生成
模式结构（*DBPS*）

输入： 距离 = [d_0,d_1,d_2,..., d_n]；类型=“和弦”或“琶音”或“音阶”
输出： 所有调中的模式

对每个*MIDI* 音高，在 *range(24, 109)* **执行**
  $note_0 \leftarrow$MIDI 音高
  $time \leftarrow time$ #当前时间
  对于 *i* **在** *range(0, len(distance))***中执行**
    $note_{i+1} \leftarrow note_i + distance[i]$
    如果 类型 == “琶音”或“音阶”**则**
      $time \leftarrow time + 1$
    end
  end
结束

是常见的和弦进行。和弦进行是构成音乐基础的和弦连续组合 [12]。它们通常使用罗马数字表示，小写字母表示小和弦，大写字母表示大调和弦。这种表示方式使和弦进行不受乐曲所演奏的具体调的影响。

2.2. 通过基于距离的模式结构（*DBPS*）进行数据集创建

我们采用一种称为基于距离的模式结构（*DBPS*）的方法，以灵活直观的方式自动生成钢琴模式。由于大多数钢琴音乐模式都具有公式化的结构，因此可以根据模式内各音高之间的间距生成模式。我们使用 MIDI 音高编号表示音高，其中距离 1 表示两个音符之间最短的距离（即一个半音）。为说明这一方法，我们在算法 1 中提供了简化的伪代码。该算法以此前音符之间的音高间距为输入，从而生成任意模式。例如，对于包含三个音符的任意大三和弦 $note_0$，$note_1$，$note_2$，输入为 $distance = [4, 3]$。也就是说，以根音为 $note_0$ 时， $note_1$ 与其相距 4 个音高（或 4 个半音），而 $note_2$ 与其相距 3 个音高。同理，对于大七和弦 $distance = [4, 3, 4]$；依此类推。

为了生成和弦进行，我们采用一种类似的方法，根据进行中每个和弦内音符之间的距离进行构建。每个和弦的基础音符（*note0*）由其与参考音（该进行所处的调）之间的距离决定，具体依据罗马数字表示法。例如， ii 和弦的基础音符与参考音相距两个音高，而 IV 和弦的基础音符则相距五个音高。确定基础音符后，即可按照算法 1 所述生成进行中的和弦。例如，广为人知的爵士和弦进行 *ii-V-I* 可用以下模式结构表示：

和弦1：音符0=参考音+2；音符1=音符0+3；音符2=音符1+4
和弦2：音符0=参考音+7；音符1=音符0+4；音符2=音符1+3
和弦3：音符0=参考音；音符1=音符0+4；音符2=音符1+3

DBPS方法还可用于生成爵士乐中常见的变化或扩展和弦。

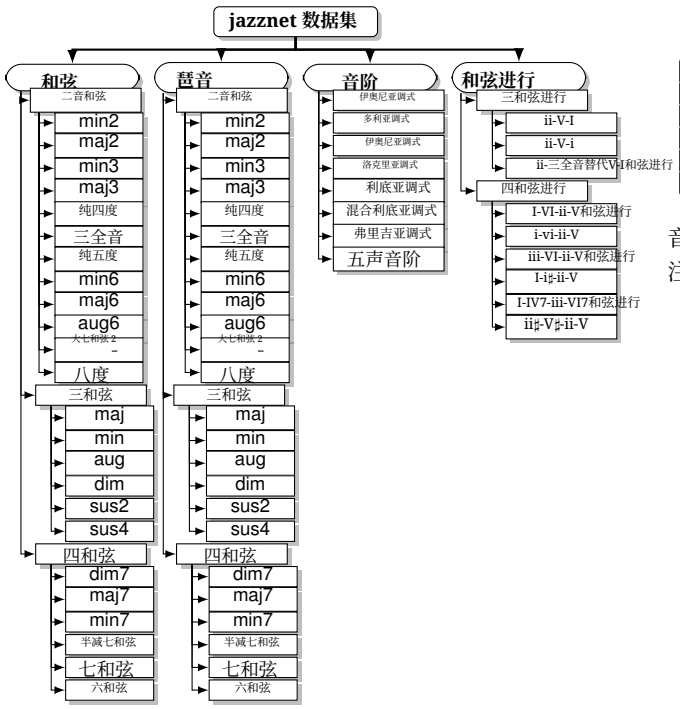


图2。jazznet 数据集的高层分类体系。

该方法通过施加变化或扩展所需的音程距离来生成新的和弦。例如，在包含音符$note0$、$note1$和 $note2$的和弦上进行属和弦 $7th$ 扩展时，会添加一个新音符 $note3$，其音高比 $note2$高三个音程。添加大和弦 $7th$ 会增加一个比 $note2$高四个音程的音符。同样，也可以通过修改和弦中某个音符的音高来生成变化音。例如，将 $note2$ 的音高降低一个音程即可生成 $\flat 5$ 。

我们为调查中选出的所有钢琴模式创建了模式结构，然后使用 Python 开发了一个模式生成器（第 2.5 节），利用这些结构为所有调和转位生成 MIDI 格式文件[3] 。随后，我们人工检查了这些 MIDI 文件的准确性，再使用 Timidity 合成器[4], 将其转换为 WAV 格式，采样率为 16KHz，位深为 24 位。为确保生成的模式质量，我们随机抽取 WAV 文件进行核验，由两位爵士钢琴家在实体钢琴上演奏这些模式，以确认声音和标签均正确。

2.3. 数据集描述与统计

图2展示了该数据集的高层分类体系。数据集包含三种和弦及琶音：二音和弦、三和弦和四和弦。音阶以全音阶（6个音符）为主，并包含一种五声音阶（5个音符）。数据集以和弦进行为主，分为两种：三和弦进行和四和弦进行（七和弦）进行。所有模式均包含其全部转位，并涵盖88键钢琴的所有调。

表1总结了数据集的统计信息。该数据集以 WAV 格式发布（也提供 MIDI 文件），并按以下类型： 和弦、琶音分别存放在不同的数据目录中,

[3]https://github.com/MarkCWirt/MIDIUtil[4] http://timidity.sourceforge.net/

表1。jazznet 数据集的统计信息

Type	#模式	#总计	时间（秒）	小时数	大小（GB）
和弦	24	5,525	3	259	1
琶音	24	5,525	4; 5; 6	433	1.7
音阶	8	4,590	9; 7	674	6.3
和弦进行	9	146,880	7; 10	25,568	85.6
总计	65	162,520		26,934	95

音阶和和弦进行；子目录则按不同的模式划分。共有65种已标注的模式，具体如下:

•和弦/琶音 （每类 24 个，分别加上“- chord”或“- arpeggio”后缀）：*12* 种二音和弦（每种 1 个转位）：小二度（min2）、大二度（maj2）、小三度（min3）、大三度（maj3）、纯四度（perf4）、三全音、纯五度（perf4）、小六度（min6）、大六和弦（maj6）、小七度或增六度（aug6）、大七和弦（maj7 2）、八度；*6* 种三和弦（每种 2 个转位）：大调（maj）、小调（min）、增和弦（aug）、减和弦（dim）、挂二和弦（sus2）、挂四和弦（sus4）；以及 *6* 种四和弦（每种 3 个转位）：dim7、maj7、min7、min7♭5、七和弦、六和弦。和弦和琶音总计各有 5,525 个。

• **音阶** (8)：*5*种自然音阶（每种有6种转位）：伊奥尼亚调式、多利亚调式、伊奥尼亚调式、洛克里亚调式、利底亚调式、混合利底亚调式、弗里吉亚调式；*1*种五声音阶 （4种转位）。总计有4590个音阶。

• **和弦进行**（9种）：三和弦进行 （64种组合）：ii-V-I、ii-V-i和ii-三全音替代V-I和弦进行（triV表示对V进行三全音替代）；四和弦进行 （256种组合）：I-VI-ii-V、i-vi-ii-V、iii-VI-ii-V、I-i♯-ii-V、I-IV7-iii-VI7和ii♯-V♯-ii-V。总计有146,880个和弦进行。

每个调式子目录都包含该调式中的模式，并根据特定的音符/音高、八度、调式和转位进行标记。和弦进行中的转位根据和弦组合进行标记。

所有模式均以每分钟 60 拍（bpm）的速度录制，并在结尾保留两拍的衰减。和弦录音时长均为 3 秒。琶音和音阶的录制方式为每拍演奏一个音符。琶音时长为 4 至 6 秒，音阶时长为 7 至 9 秒，具体取决于音符数量。和弦进行的录制方式为每小节演奏两个和弦，时长为 7 至 10 秒。

2.4. 建议的子集

鉴于和弦进行数量导致数据集存在不平衡，我们建议使用数据集的一些子集，以满足对数据集平衡性要求较高的场景，或在下载/使用完整数据集不切实际时（例如，初步测试模型）。所有子集均包含全部和弦、音阶和琶音（15640 个音频片段）。小型、中型和大型子集分别增加 5876、14688 和 36720 个和弦进行，总计分别包含 21516、30328 和 52360 个音频片段。所有子集均以伪随机方式生成，以确保涵盖所有模式。元数据文件采用 CSV格式，并建议按 80/10/10% 的比例划分训练集、验证集和测试集。验证集和测试集从不同模式中随机选取，且不与训练集重叠。

2.5. 模式生成器

我们使用 Python 开发了钢琴模式生成器，用户可以借此轻松创建新模式。由于篇幅所限，我们省略了生成器函数的完整细节（详情请见 Github 页面[5]），这里只作简要说明：按照第 2.2 节所述指定音高间距，即可生成和弦、琶音和音阶。同样，指定和弦的罗马数字以及支持的扩展音/变化音（例如，大七和弦 ♭5 ♯ 等），即可生成和弦进行。关键在于，这些模式会在88键钢琴的所有调和所有转位中生成。随后，用户可以使用自己选择的工具，将生成的 MIDI 文件转换为所需格式，例如 WAV。所提供的工具可用于创建大量新模式，而且脚本是开源的，因此可以进一步扩展该工具，以支持其他模式。

2.6. 应用

Jazznet 可用于评估针对各种机器学习音乐信息检索任务开发的模型。最直接、最基础的任务涉及基于数据集类别层级的**自动音乐理解**：类型识别（判断输入是和弦、琶音、音阶还是和弦进行）或调式识别，例如判断输入是增和弦（增和弦）还是大和弦（大和弦），是伊奥尼亚调式，还是 ii-V-I进行。还可以预测音乐模式的具体细节，例如音高内容、具体调（如 C-大和弦与 E♭-大和弦）、八度和转位。

Jazznet 还可用于更复杂的音乐信息检索任务。例如，**自动音乐转录** [11, 15] 被视为音乐信息检索中的一项经典任务，即提取音频录音中的音乐组成部分（如和弦）。目前大多数音乐转录模型都聚焦于音乐中出现的孤立音符或和弦。但利用在 jazznet 上训练的模型，还可以预测音阶和和弦进行等其他音乐模式。该数据集还可用于开发**音乐推荐系统** [16, 17]。推荐系统可以基于音乐模式构建：如果听众喜欢某首歌曲中的某些和弦或音阶（例如一首以小调音阶或小和弦为主的歌曲），那么他们可能也会喜欢包含类似和弦调式的其他歌曲。与音乐理解相关的是，jazznet 可用于**自动音乐生成** [18, 19]。许多爵士乐作品都包含爵士和弦进行的重复，并在其上演奏音阶或琶音。例如，几乎所有爵士标准曲中都会出现的*ii-V-I* 进行，可以使用每个和弦的琶音或匹配的音阶进行即兴演奏 [7]，（例如，在*ii* 和弦上使用多利亚调式）。J azznet 可用于学习音乐作品中的基本模式及其出现方式（例如和弦进行的频率、与和弦进行配合使用的音阶等）。

3. 实验

实验设置： 我们使用了两个简单模型，以展示 jazz net 数据集在音乐音频识别的两项任务——类型识别和调式识别——中的潜力。第一个模型是卷积循环神经网络（CRNN），如 [20] 中所述。该模型以梅尔频谱图作为输入，并使用循环神经网络（RNN）对音乐结构中的长期依赖关系进行建模。第二个模型是 [21] 中介绍的 M5 模型：一种以时间序列波形为输入、无需预处理的深度卷积神经网络。由于数据预处理通常是音频机器学习的瓶颈，该模型能够反映在完全不进行预处理的情况下使用数据集所能达到的性能。两个模型均包含 4 个带批归一化的卷积层，并使用 ReLU 激活函数。我们使用 Adam 优化器和分类交叉熵损失函数训练模型，并按照先前在该数据集的中等规模子集（30328 个音频片段）上的研究 [21],，使用平均精度均值（mAP）和曲线下面积（AUC）分数评估模型性能。

表2. 调式预测的平均精确率和平均精度均值。

	和弦	琶音	音阶	和弦进行	mAP
CRNN	0.28	0.16	0.10	0.67	0.63
M5	0.36	0.09	0.06	0.48	0.30

结果： 对模型而言，识别和弦、琶音、音阶和和弦进行是一项简单的任务。CRNN 和 M5 模型得分较高，平均曲线下面积为 0.984。它们始终能够识别全部四种类型，平均精度均值分别为 0.99 和 0.97。相比之下，两位爵士乐钢琴家曾非正式地尝试识别随机抽取的样本，二人均达到了 100% 的准确率。

另一方面，如表2所示，模型要识别 65 种模式则困难得多。CRNN 和 M5 模型的平均精度均值分别为 0.63 和 0.30，平均曲线下面积为 0.04。两个模型都发现琶音和音阶最具挑战性：CRNN 对二者的平均精度分别为 0.16 和 0.10，M5 则分别为 0.09 和 0.06。不过，模型在和弦进行上的表现最好，CRNN 和 M5 的平均精度分别为 0.67 和 0.48。人类也发现调式预测比类型识别更困难，准确率在 57% 至 73% 之间。总而言之，该数据集旨在体现识别音乐模式的难度层级，并为各种具有挑战性的音乐信息检索任务提供模型基准测试的机会。

4. 结论

本文介绍了*jazznet*，这是一个收录了爵士钢琴核心音乐模式的数据集，同时也是一个开源模式生成器，可供研究人员对用于复杂音乐信息检索（MIR）任务的机器学习模型进行基准测试。我们的目标是让该数据集助力音乐信息检索领域的机器学习研究。未来，我们计划扩充数据集，纳入更多音乐听觉属性，例如力度和节奏变化，以及更复杂的模式，如五音和弦和更长的和弦进行。此外，我们还打算探索更先进的模型，以提升使用该数据集完成相关任务的表现，并进一步研究DBPS方法自动生成不同类型数据的潜力。

[5]https://github.com/tosiron/jazznet/tree/main/ PatternGenerator

5. 参考文献

[1] George Tzanetakis 和 Perry Cook，《音频信号的音乐流派分类》，*IEEETransactionson speech andaudio processing*，第 10 卷，第 5 期，第 293–302 页，2002 年。[2]Thierry Bertin-Mahieux、Daniel PW Ellis、Brian Whitman 和 Paul Lamere，《百万歌曲数据集》，2011 年。[3] J ort F Gemmeke、Daniel PW Ellis、Dylan Freedman、Aren Jansen、Wade Lawrence、R Channing Moore、Manoj Plakal 和 Marvin Ritter，《Audio Set：用于音频事件的本体与人工标注数据集》，载于*2017IEEEInternational Conference on Acous-tics, Speechand Signal Processing (ICASSP)*。IEEE，2017 年，第 776–780 页。[4]Micha¨el Defferrard、Kirell Benzi、Pierre Van- dergheynst 和 Xavier Bresson，《FMA：用于音乐分析的数据集》，*arXiv preprintarXiv:1612.01840*，2016 年。[5] John Thickstun、Zaid Harchaoui 和 Sham Kakade，《从头学习音乐特征》，*arXivpreprintarXiv:1611.09827*，2016 年。[6] Masataka Goto、Hiroki Hashiguchi、Takuichi Nishimura 和 Ryuichi Oka，《RWC 音乐数据库：音乐流派数据库与乐器音色数据库》，2003 年。[7] Mark Levine，《爵士钢琴书》，“O’Reilly Media, Inc.”，2011 年。[8] Diana Deutsch，《绝对音高之谜》，*Acoustics Today*，第 2 卷，第 4 期，第 11 页，2006 年。[9] Jesse Engel、Cinjon Resnick、Adam Roberts、Sander Dieleman、Mohammad Norouzi、Douglas Eck 和 Karen Simonyan，《使用 WaveNet 自编码器合成音乐音符的神经音频合成》，载于*InternationalConferenceonMachineLearning*。P MLR，2017 年，第 1068–1077 页。

[10] Valentin Emiya、Nancy Bertin、Bertrand David 和 Roland Badeau，《MAPS：用于多音高估计和音乐自动转录的钢琴数据库》，研究报告，*Inria-00544155*，2010年。

[11] Emmanouil Benetos、Simon Dixon、Zhiyao Duan 和 Sebastian Ewert，《自动音乐转录：综述》，*IEEE* 信号处理杂志，第36卷第1期，第20–30页，2018年。

[12] Mark Levine，《爵士乐理书》，“O’Reilly Media, Inc.”，2011年。

[13] Ted Gioia，《爵士标准曲：曲目指南》，牛津大学出版社，2021年。

[14] Hal Leonard 公司，你应该在钢琴上演奏的前 *50* 首爵士标准曲，Hal Leonard，2017 年。

[15] Emmanouil Benetos、Simon Dixon、Dimitrios Gian- noulis、Holger Kirchhoff 和 Anssi Klapuri，“自动音乐转录：挑战与未来方向”，《智能信息系统杂志》，第 41 卷，第 3 期，第 407–434 页，2013 年。[16] Markus Schedl、Peter Knees、Brian McFee、Dmitry Bogdanov 和 Marius Kaminskas，“音乐推荐系统”，载于《推荐系统手册》，第 453–492 页。施普林格，2015 年。[17] Aaron Van den Oord、Sander Dieleman 和 Benjamin Schrauwen，“基于深度内容的音乐推荐”，《神经信息处理系统进展》，第 26 卷，2013 年。[18] Dorien Herremans、Ching-Hua Chuan 和 Elaine Chew，“音乐生成系统的功能分类法”，《*ACM*计算评论》（*CSUR*），第 50 卷，第 5 期，第 1–30 页，2017 年。[19] Filippo Carnovalini 和 Antonio Rod`a，“自动音乐生成与表现性演奏的多层方法”，载于*2019* 年多层音乐表征与处理国际研讨会（*MMRP*）。IEEE，2019 年，第 41–48 页。[20] Keunwoo Choi、Gy¨orgy Fazekas、Mark Sandler 和 Kyunghyun Cho，“用于音乐分类的卷积循环神经网络”，载于*2017*年*IEEE*声学、语音与信号处理国际会议（*ICASSP*）。IEEE，2017 年，第 2392–2396 页。[21] Wei Dai、Chia Dai、Shuhui Qu、Juncheng Li 和 Samarjit Das，“用于原始波形的超深卷积神经网络”，载于*2017*年*IEEE*国际声学、语音与信号处理会议（*ICASSP*）。IEEE，2017 年，第 421–425 页。