音乐创作与深度学习:综述

**Carlos Hernandez-Olivan∗carloshero@unizar.es**

**何塞·r·贝尔特兰jrbelbla@unizar.es**

萨拉戈萨大学工程与通信系，María de Luna Calle

arXiv:2108.12290v2 [cs.]2021年9月7日

**摘要**

创造一件复杂的艺术作品(如音乐作品)需要展现出真正的创造力，这取决于与音乐语言等级相关的各种因素。音乐生成一直面临着算法方法，最近，深度学习模型正被用于其他领域，如计算机视觉。在这篇文章中，我们想要把基于人工智能的音乐作曲模型与人类音乐作曲和创作过程之间的现有关系置于背景之下。我们对最近的深度学习音乐创作模型进行了概述，并从理论的角度将这些模型与音乐创作过程进行了比较。我们试图通过分析当前的深度学习模型产生创造性音乐的能力，或者人工智能和人类作曲过程之间的相似性来回答这一任务中一些最相关的开放问题。

*关键词音乐生成深度学习机器学习神经网络*

**1简介**

音乐通常被定义为一系列的音调或节奏，或两者兼有，以某些特定的模式。音乐创作(或生成)是创作或写一首新的音乐作品的过程。音乐作曲术语也可以指原创作品或音乐作品。音乐创作需要创造力，这是一种人类独特的能力，可以理解和产生无限多的句子，其中大多数句子在[2]之前从未遇到或说过。这是在设计或提出基于ai的作曲算法时需要考虑的一个非常重要的方面。

具体地说，音乐创作是音乐信息检索领域的一个重要课题。它包括旋律生成、多音轨或多乐器生成、风格转换或和声等子任务。本文将从近年来在人工智能和DL基础上蓬勃发展的众多技术的角度来讨论这些方面。

**1.1从算法合成到深度学习**

从20世纪80年代开始，人们对计算机作曲的兴趣从未停止过增长。一些实验在20世纪80年代早期出现，如大卫·柯普在1983 - 1989年的《音乐智力实验》或伊安尼斯·谢纳基斯的《类比法A和B》。2000年代后期，也有David Cope提出将Markov链与语法相结合用于自动作曲，Koening的Project1 (PR1)[5]等相关作品相继诞生。这些技术可以归为算法音乐作曲，算法音乐作曲是一种通过形式化的方法进行作曲的方法。这种类型的组合由一个受控过程组成，该过程是基于必须按照固定顺序执行的数学指令。在算法组合中有几种方法，如马尔可夫模型、生成语法、元胞自动机、遗传算法、转移网络或曹氏理论[8]。有时，这些技术和其他概率方法与深度神经网络相结合，以调整它们或帮助它们更好地模拟音乐，这是DeepBach[9]的情况。这些模型可以生成和协调不同风格的旋律，但缺乏通用性

∗

https://carlosholivan.github.io

这些模型的能力，以及必须手工完成的基于规则的定义，使得这些方法与基于Depp学习的模型相比，没有那么强大和一般化。

从20世纪80年代到21世纪初，第一批尝试用网络音乐来模仿音乐的作品诞生了。近年来，随着深度学习(Deep Neural Networks, DL)技术的发展，许多研究都试图利用深度神经网络(Deep Neural Networks, NN)对音乐进行建模。用于音乐生成的DL模型通常使用已被证明在其他领域表现良好的神经网络架构，如计算机视觉或自然语言处理(NLP)。在这些领域也可以使用预先训练的模型，可以用于音乐生成。这被称为转移学习[14]。本文稍后将介绍一些神经网络技术和体系结构。今天的音乐创作采用了大规模的NLP应用程序的输入表示和神经网络体系结构，例如基于变压器的模型，在这项任务中显示了非常好的性能。这是因为音乐可以被理解为一种语言，在这种语言中，每一种风格或音乐流派都有自己的规则。

**1.2基于深度学习的音乐作曲神经网络架构**

首先，我们将概述目前为止在音乐作曲任务中提供最佳结果的最广泛使用的神经网络架构。在音乐作曲任务中，最常用的神经网络结构是生成模型(Variational AutoEencoders, VAEs)或生成对抗网络(Generative Adversarial Networks, GANs)，以及基于nlp的模型(Long - Short-Term Memory, LSTM)或transformer。下面是这些模型的概述。

**1.2.1变分自编码器(VAEs)**

原始VAE模型[15]采用编码器-解码器架构，通过重构输入产生一个潜在空间(见图1a)。潜在空间是压缩数据的多维空间，其中最相似的元素位于彼此最近的位置。在VAE中，编码器近似后验，解码器参数化似然。对编码器和解码器分别采用λ和θ参数的神经网络参数化后验近似和似然近似。后验推理是通过最小化编码器或近似后验之间的Kullback-Leiber (KL)发散，通过最大化证据下界(ELBO)来实现的。梯度是用所谓的重参数化技巧计算的。原始的VAE模型也有一些变化，如β-VAE[16]，它在重构损失中增加了一个惩罚项β，以改善潜在空间分布。在图1a中，我们展示了一般的VAE架构。一个基于VAE的音乐作曲DL模型的例子是MusicVAE[17]，我们将在本文的进一步章节中描述它。

**1.2.2生成对抗网络(GANs)**

GANs[18]是由两个nn组成的生成模型:生成器G和鉴别器d。生成器学习分布pg 进行训练是为了让鉴别器最大限度地为训练样本和生成器生成的样本分配正确标签的概率。这种训练思想可以被理解为D和G遵循Goodfellow等人[18]描述的双人极大极小博弈。在图1b中，我们展示了GAN的总体结构。

生成器和鉴别器可以由不同的神经网络层构成，如多层感知器(MLP)[19]、LSTM[20]或卷积神经网络(CNN)[21][22]。

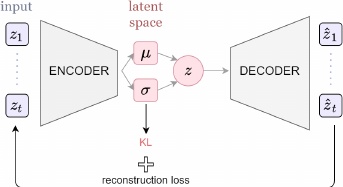
**1.2.3变压器**

变压器[23]不仅在自然语言处理中表现良好，而且在计算机视觉模型中也有很好的应用。transformer可以被用作自回归模型，就像lstm，它允许它们被用于生成任务。《变形金刚》背后的基本理念是注意力机制。Vaswani et al.[23]提出的原始注意机制在音乐作曲任务[24]中有多种变体。注意层与前馈层的结合形成了Transformer的编码器和解码器，这与同样由编码器和解码器组成的纯AutoEncoder模型不同。变形器是用符号来训练的，这些符号是输入的结构化表示。在图1c中，我们展示了通用Transformer架构。

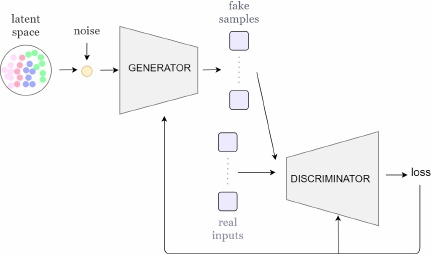
**1.3深度学习在音乐创作中的挑战**

有不同的观点，从挑战的角度来看，DL音乐创作，让我们问一些问题，与输入表示和DL模型有关，这些模型在这个领域已经被使用，实际的最先进的方法的输出质量或研究人员测量产生的音乐质量的方法。在这篇综述中，我们问自己以下问题，涉及到合成过程和输出

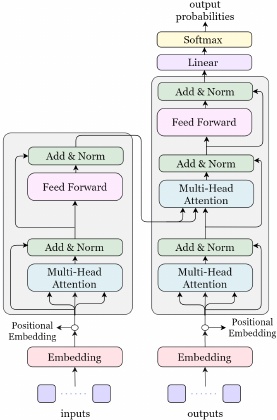
2



(一)



(b)



(c)

图1:a) VAE [15] b) GAN[18]和c)变压器总体架构摘自[23]。

能够产生音乐并具有一定创造力的模型?用DL来演奏音乐的最佳神经网络架构是什么?端到端方法能生成完整的结构化音乐片段吗?用DL创作的乐曲仅仅是对输入的模仿，还是神经网络能够以训练数据中不存在的风格生成新的音乐?神经网络是否应该按照人类的逻辑和过程来作曲?音乐生成的DL模型需要多少数据?当前的评价方法是否足以比较和衡量音乐创作的创造力?

为了回答这些问题，我们从音乐作曲或生成(Music Composition or Generation)所遵循的过程的角度出发，来获取最终的作曲和DL模型的输出，即人类作曲过程与基于深度学习的音乐生成过程的比较，以及生成的音乐所表现出的艺术性和创造性特征。我们还分析了最近使用深度学习的最先进的音乐作曲模型，以展示这些模型所提供的结果(主题、完整的作曲……)。分析的另一个重要方面是这些模型用来生成音乐的输入表示，以了解这些表示是否正确

3.

适合作曲。如果这些神经网络架构足够强大，能够创作出具有一定创造力的新音乐，那么这些模型将如何改进，以及深度学习音乐作曲的研究方向和未来工作。

**1.4论文结构**

在这篇综述中，我们从作曲过程和生成输出的类型来分析音乐作曲任务，而不包括演奏或合成任务。本文的结构如下:第二节概述了音乐的创作过程和音乐的基本原理。在第3节中，我们从旋律创作的角度概述了最先进的方法，并描述了用于创作结构化音乐的DL模型。在第4节中，我们描述了产生多音轨或多乐器音乐的DL模型，也就是说，为一种以上乐器制作的音乐。在第5节中，我们展示了不同的方法和度量，通常用于评估音乐生成模型的输出。在第6节中，我们通过分析我们在第3节和第4节中描述的模型来描述音乐生成中仍然存在的开放问题。最后，在第7节中，我们揭示了未来的工作和研究中仍存在的挑战。

**2音乐创作过程**

与书面语言一样，音乐创作过程是一个复杂的过程，它依赖于大量的决策。在音乐领域，这个过程[26]取决于我们所使用的音乐风格。例如，在古典音乐中，很常见的是由一个或两个小节组成的小单元，称为主题，然后发展成旋律或乐句，而在流行音乐或爵士乐中，更常见的是采用一个和弦进行，并在其之前创作或即兴创作旋律。不管我们所创作的音乐风格如何，当一个作曲家开始创作一段音乐时，在它背后总会有一些基本的旋律或和声理念。从古典音乐的角度来看，这一理念(或母题)是由作曲家发展起来的，以构建旋律或短语，遵循一定的和声进程，然后这些短语被组成部分。每个小节都有自己的目的，所以它可以用不同的键来写，它的乐句通常遵循不同的和声进行，而不是其他小节。通常，作品有旋律部分和伴奏部分。一段音乐的旋律部分可以由不同的乐器演奏，它们的频率范围可能相似，也可能不相似，而和声部分给乐曲一种深刻和有组织的感觉。乐器，不一定在相同的频率范围，结合仪表和管弦乐技术(见3.2节)。这些元素在音乐创作中是至关重要的，也是定义音乐作品的风格或体裁的重要关键。音乐有两个维度，时间维度和和声维度。时间维度由音符的持续时间或节奏来表示，它是这个轴上的最低水平。在这个维度中，音符可以被分组或以一种叫做条的单位来度量，条是一组有序的音符。另一个维度，和声，与音符值或音高有关。如果我们想象一个图像，时间维度是横轴，和谐维度是纵轴。和声也有一个短暂的演变，但这并没有体现在乐谱上。有一种非常常见的基于软件的音乐表现形式，叫做钢琴卷，它遵循这种逻辑。

音乐的时间维度是由低层次的单位构成的，这些单位是音符，这些音符被组合在构成(主题)的小节中。在时间的高级维度中，我们可以找到由持续8小节或更多小节的短语组成的部分(这取决于风格和作曲家)。和声的最低级是音符级，不同乐器演奏的音符叠加在一起就形成了和弦。和弦的序列被称为和弦进行，与乐曲有关，它们在时间维度上也有相关性。话虽如此，我们可以把音乐看作是一种复杂的语言模型，它由短期和长期的关系组成。这些关系延伸到两个维度，时间维度与音乐结构有关，和声维度与音符或音高和和弦有关，即和声。

从符号音乐的产生和分析的角度来看，基于Walton[27]的思想，基本的音乐原则或元素是(见图2):

•和谐。它是构成和弦的音符的叠加，从而构成和弦进行。音级可以被认为是和声中最低的音级，其后是和弦级。最高一级可以认为是递进级，通常属于某个键。

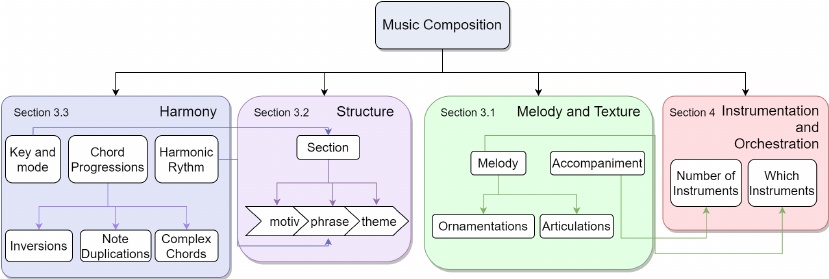
•音乐形式或结构。它是音乐呈现的高层次结构，与时间维度相关。乐曲中最小的部分是母题，母题在乐句中展开，乐句的组合构成乐段。音乐部分的顺序取决于音乐风格，如一些流行歌曲(也表示为ABCBA)的内调-合唱-外调或奏鸣曲的叙述-展开-再现或ABA。可以在不同的尺度和模式的部分串联给了我们整个组成。

4

•旋律和纹理。在音乐术语中，织体是指乐曲中必须结合的旋律、节奏和和声内容。音乐可以是单音或复音取决于同时演奏的音符，主音或支声取决于旋律，如果它有或没有伴奏。

•乐器和管弦乐。这些是考虑到音乐作品中乐器或音轨数量的音乐技巧。管弦乐指的是将不同乐器组合在一起，共同谱写一首乐曲。管弦乐指的是将旋律和伴奏分配给不同的乐器，共同谱写一首乐曲。在录音或基于软件的音乐表现中，“乐器”被组织成音轨。每一个音轨都包含了在一个乐器[28]上演奏的音符集合。因此，我们可以把一个有多个乐器的作品称为多音轨，多音轨是指包含两个或多个音轨的信息，其中每个音轨由一个乐器演奏。每个音轨可以包含一个或多个音符，同时发声，导致单音轨道和复音轨道分别。

音乐类别之间是有联系的。和声与结构有关，因为一个部分通常以相同的音阶和模式演奏。小节之间有节奏，也可以有改变音阶的调式。织体和乐器与音色特征有关，它们之间的关系是基于一个事实，即不是所有的乐器都能演奏相同的旋律。一个例子是，当我们有一个旋律，有很多装饰元素，不能用特定的乐器家族演奏(因为每个乐器的技术可能性或设计师的原因)。



(一)



(b)

图2:a)一般的音乐创作方案，b)贝多芬第五交响曲开头的一个音乐层次或类别的例子。

另一个重要的音乐属性是动态，但它们与演奏有关，而不是与作曲本身有关，所以我们不会在本文中讨论它们。在图2中，我们展示了我们在这篇综述中所涉及的音乐作曲过程的各个方面，并且描述了各个主题所讨论的类别和论文各部分之间的关系。

5

**3旋律生成**

旋律是一组具有一定节奏、以美学的方式排列的音符。旋律可以是单音或复音。单音是指在一个时间步中只演奏一个音符的旋律，而在复音旋律中，在同一时间步中有多个音符被演奏。旋律生成是音乐创作的重要组成部分，它已经被尝试与算法合成和几个神经网络架构，包括生成模型，如VAEs或GANs，循环神经网络(RNNs)用于自回归任务，如LSTM，神经自回归分布估计器(NADEs)[29]或当前用于自然语言处理的模型，如transformer[23]。在图3中，我们展示了一个旋律生成模型的输出式乐谱的音乐基本原理的方案。

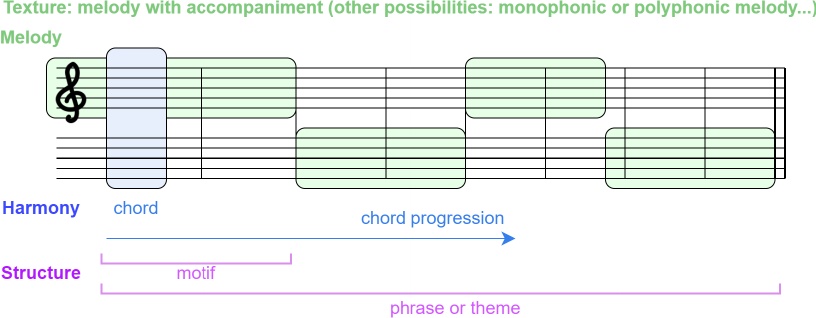


图3:类似于输出的旋律生成模型。

**3.1旋律生成的深度学习模型:从主题到旋律短语**

根据我们所处领域的音乐类型，人类的作曲过程通常从一个主题或和弦开始，然后扩展为一个乐句或旋律。当涉及到DL方法的音乐生成，几个模型可以生成短期的音符序列。2016年，第一个DL模型尝试使用循环神经网络(rnn)和语义模型(如单元选择[30])生成短旋律。这些模型适用于短序列，所以创造完整旋律的兴趣随着新的神经网络的诞生而增长。从这些最初的作品中衍生出来，并以创造更长的序列(或旋律)为目标，其他结合了神经网络和概率方法的模型出现了。这方面的一个例子是谷歌的品红旋律RNN模型[31]发布于2016年，预期RNN[32]或DeepBach[9]都发布于2017年。DeepBach被认为是目前最先进的音乐生成模式之一，因为它能够生成巴赫风格的4声赞美诗。

然而，这些方法不能产生新的旋律与高水平的创造力从零开始。为了改进生成任务，研究人员选择生成模型进行音乐创作。事实上，如今，在2到16小节中生成主题或短旋律的最佳模型之一就是MusicVAE2 [17]出版于2018年。MusicVAE是一个基于VAE[15]的音乐生成模型。有了这个模型，音乐可以通过插入一个潜在的空间来产生。该模型使用来自LMD 3[33]的大约150万首歌曲进行训练，它可以为几乎3种乐器生成复调旋律:旋律、贝斯和鼓。在创建了MusicVAE模型之后，随着其他领域新的神经网络架构的诞生，新的基于dlm的模型可以创建更长的旋律的必要性和可用性增加了。2018年推出的“Music Transformer[24]”等以《变形金刚》为基础的新机型，以及OpenAI[34]提出的“MuseNet 2019”等使用“GPT-2”等预先训练过的《变形金刚》的机型。这些基于Transformer的模型，如Music Transformer，可以生成较长的旋律并继续给定的序列，但在几个小节或几秒钟后，旋律就会变得有点随机，也就是说，有一些音符和和声不符合乐曲的乐感。

为了克服这个问题，并开发出能够生成更长的序列，同时又不会失去在之前的酒吧或主要主题中产生的音乐感的模型，新的模型在2020年和2021年诞生，它是VAEs、变形金刚或其他神经网络或机器学习算法的组合。这些模型的一些例子是

https://magenta.tensorflow.org/music-vae，于2021年8月访问

2

3.

https://colinraffel.com/projects/lmd/，于2021年8月访问

6

TransformerVAE[35]和PianoTree[36]。这些模型甚至在复调音乐中也表现良好，它们可以生成音乐短语。最新发布的生成完整短语的模型之一是Mittal et al.[37]在2021年提出的模型，它基于去噪扩散概率模型(去噪扩散概率模型ddpm)[38]，这是一种新的生成模型，通过学习将扩散过程从数据转化为高斯噪声来生成高质量的样本。该模型采用MusicVAE 2杆模型训练扩散模型，以捕捉VAE潜伏期z之间的时间关系k k = 32是32个潜在变量，可以产生64个bar (2 bar per latent)。虽然可以产生较长的复调旋律，但它们没有遵循中心母题，因此往往会失去一定的方向感。

**3.2结构感知**

正如我们在第一节中提到的，音乐是一种结构化语言。一旦旋律被创造出来，它们必须被分成更大的部分(见图2)，这些部分在乐曲中起着重要的作用。这些部分的名称根据音乐风格而有所不同。流行音乐或trap音乐的介绍、合唱、韵文，古典奏鸣曲的阐述、发展、再现等都有不同的名称。部分也可以用大写字母命名，歌曲结构可以表示为ABAB，例如。在DL音乐创作中，生成有结构的音乐是最困难的任务之一，因为结构意味着对节奏、和弦进行和旋律的美感，这些美感是由桥梁和节奏串联起来的。

在DL中，有一些模型试图通过强加自相似性约束的高级结构来生成结构化的音乐。其中的一个例子是Lattner等人在2018年提出的模型[39]，该模型使用卷积受限玻尔兹曼机(C-RBM)生成音乐，并使用自相似矩阵[40]来施加乐曲的结构，就像它是一个模板一样。这种强加结构模板的方法类似于作曲家在作曲时所遵循的作曲过程，而最终的音乐作品也遵循强加的结构模板。虽然新的DL模型有端到端的趋势，关于用结构对音乐建模的新研究也在[41]中发布，但还没有一种DL模型能够自己生成结构化的音乐，即不需要模板的帮助，也不需要高层结构信息传递给NN。

**3.3和声和旋律调理**

在DL的音乐创作中，有一项任务是对给定旋律进行和声，这与从零开始创作复调旋律的任务是不同的。一方面，如果我们用DL模型从头开始分析创作出来的旋律的和声，我们会发现用DL模型生成的音乐结构并不好，因为它还没有以端到端的方式组成不同的部分，也没有写出美学的节奏或部分之间的桥梁。尽管如此，基于变形金刚的模型所产生的和声在生成的乐曲[24]的第一个小节中是连贯的，因为它遵循一个特定的键。我们必须强调的是，这些旋律是为钢琴而写的，这与多乐器音乐不同，后者带来了额外的挑战，比如为每种乐器生成合适的旋律或伴奏，或者决定哪些乐器组成合奏(见第4节)。另一方面，旋律协调的任务包括生成一个特定旋律的和声。伴奏可以是和弦伴奏，不管乐器是什么，也不管和弦所在的轨道是什么，多轨道伴奏，每个和弦中的音符都属于一个特定的乐器。首先，协调模型使用HMM，但这些模型被RNNs改进。一些模型预测和弦功能[42]和其他模型匹配和弦伴奏的特定旋律[43]。对于不同轨迹的伴奏生成，已经提出了基于gan的模型来实现lead sheet安排。2018年，提出了一种名为MICA[44]的多乐器协同排列模型，并在2020年提出了其改进MSMICA[45]，以生成多声道伴奏。还有一个模型叫做巴赫Doodle[46]，它使用椰子[47]为巴赫风格的特定旋律生成伴奏。这些模型的和声质量提高了从头开始创建复调旋律的模型所产生的和声，因为模型专注于旋律内容来执行和声，这比从头开始生成一个完整的和谐作品的挑战更小。

在使用条件反射的DL生成音乐的过程中还有更多的任务，比如生成一个给定和弦进程的旋律，这是人类遵循的一种作曲方式。这些任务已经通过变分自动编码器(VAEs)[48]、生成式对抗网络或基于gan的模型[49]、[50]得到了解决4 端到端模型[44]。其他模型执行完整的组合过程，如ChordAL[51]。这个模型生成和弦，然后得到的和弦进程被发送到旋律生成器，最终的输出被发送到音乐风格处理器。

https://shunithaviv.github.io/bebopnet/，于2021年8月访问

4

7

像BebopNet[50]这样的模型从爵士乐和弦中生成旋律，因为这种风格在和声环境中提出了额外的挑战。

**3.4体裁转换与体裁转移**

在音乐中，风格或流派被定义为从音乐理论到声音设计的复杂混合。这些特征包括音色、作曲过程和音乐作品中使用的乐器或音乐合成的效果。由于存在许多音乐类型，并且缺少一些类型的数据集，所以我们通常会使用风格转换技术，通过改变现有音符的音高或添加适合我们想要转换音乐的风格的新乐器，将一种特定风格的音乐转换成其他风格。

在基于计算机的音乐创作中，音乐中最常见的表现风格转移技术是获得风格的嵌入，并利用这种嵌入或特征向量生成新的音乐。神经网络中的风格传递[52]是由Gatys等人在2016年引入的，其思想是将风格特征从另一幅图像应用到一幅图像上。2018年的MIDI-VAE[53]是最早将风格转移用于符号音乐生成的研究之一。MIDI-VAE将潜伏空间中的风格编码为音高、力度和乐器特征的结合，生成复调音乐。风格迁移也可以通过迁移学习[14]实现。第一个使用迁移学习来进行风格迁移的作品是Hung et al.[54]在2019年提出的一个循环的jazz VAE模型。迁移学习是通过在源数据集上训练模型来完成的，然后在目标数据集上微调产生的模型参数，目标数据集可以与源数据集具有不同的风格。这个模型表明，使用转移学习将一个确定风格的音乐作品转换为另一个是一个很好的解决方案，因为它不仅可以用于转换新类型的现有作品，而且还可以用于在正在使用的音乐作曲数据集中不存在的类型中从头开始创作音乐现在。这方面的一个例子是使用一个用大的流行数据集(如LMD[33])训练的神经网络，并使用这个预先训练的模型通过迁移学习生成城市音乐。

其他音乐特征如和声和纹理(见图2)也被用作风格转换特征[55]，[36]，[56]。也有人研究了融合体裁模型，将不同的风格混合在一起，以一种未知的风格[57]产生音乐。

**4乐器和管弦乐**

正如我们在第二节中提到的，器乐和管弦乐是所创作的音乐体例中的基本元素，通过使用特定的乐器或他们的作品被编排的方式，可以代表每个作曲家的特征签名。一个例子就是贝多芬在他的交响曲中使用的管弦乐，它改变了音乐创作的方式。*配器是研究如何将相似或不同的乐器组合在不同的数字中以创建一个整体配器是选择和组合类似或不同的得分部分[59]。*由此，我们可以将器乐作为乐曲的颜色，管弦乐与乐曲的审美联系起来。器乐和管弦乐对我们感知音乐的方式产生了巨大的影响，因此，对于音乐的情感部分来说，尽管它们代表了音乐的一个基本部分，但情感超出了这部作品的范围。

**4.1从复音到多乐器音乐的生成**

在基于计算机的音乐创作中，我们可以将多乐器或多声道音乐中的器乐和管弦乐概念分组。然而，基于dlc的多仪器生成模型并不完全适用于这些概念。基于dl的多乐器模型为多个乐器生成多音阶音乐，但是，生成的音乐是否遵循一个连贯的和声进程?在器乐和管弦乐方面，最终的安排是一致的吗?还是基于cd的模型只是生成了多乐器音乐，而没有考虑到每个乐器的颜色或安排?在第3节中，我们展示了复音音乐的生成可以创作出具有一定和声感的音乐，但是在面对多乐器音乐时，最需要考虑的一个方面就是乐器的颜色和合奏。如何确定合奏中有多少种乐器、哪些乐器，以及它们之间的旋律和伴奏如何划分，是目前DL音乐生成中尚未解决的问题。近年来，我们面临的挑战是构建一个从头开始生成音乐的DL模型，它可以是一个交互模型，在这个模型中，人类可以选择合奏乐队[28]的乐器。还有一些模型允许在工具或酒吧上着色。我们将描述这些模型并回答4.2节中暴露的问题。在图4中，我们展示了多乐器生成模型的类似输出的乐谱的音乐基本原理的方案。

8

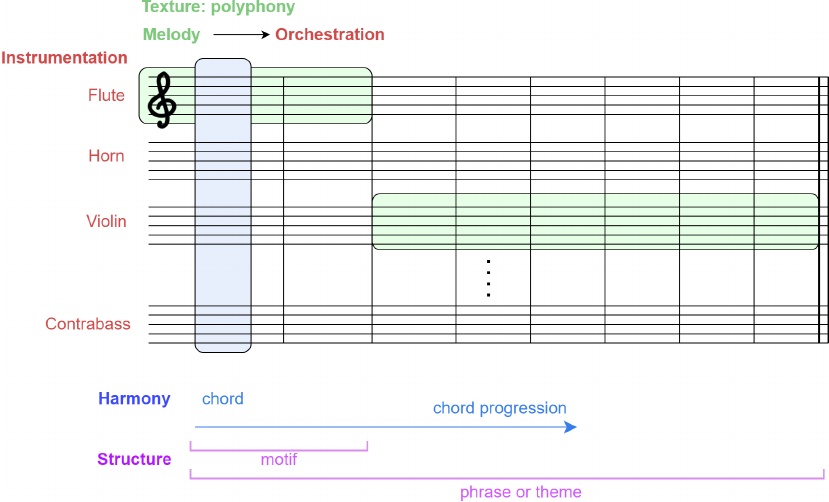


图4:多仪器生成模型类输出评分方案。

**4.2多仪器从零开始生成**

第一个可以产生多音轨音乐的模型最近被提出。在多音轨音乐生成之前，一些模型为给定的旋律或和弦生成鼓声音轨。这些模型的一个例子是Kang等人在2012年提出的模型。[60]用自动鼓发生器在给定的音阶上为旋律伴奏。随后，2017年Chu等人[61]使用分层RNN生成有鼓的流行音乐。

在音乐生成中最常用的架构之一是生成模型，如GANs和VAEs。第一个被考虑的、也是最著名的多声道音乐生成模型是MuseGAN[62]，于2017年推出。随后，更多的模型遵循多仪器生成任务[63]，[64]，并在2020年晚些时候发布了其他基于AutoEncoders的模型，如MusAE[65]。最近被用来生成音乐的另一大群神经网络架构是变形金刚。Transformer最著名的音乐生成模型是用于钢琴复调音乐生成的music Transformer[24]。2019年，Donahue等人[66]提出了LakhNES用于多声道音乐生成，2020年，Ens等人提出了基于LakhNES的条件多声道音乐生成模型(MMM)[28]，通过将多声道连接成单个序列，改进了该模型的令牌表示。该模型使用多仪器和BarFill表示，如图5所示。在图5中，我们展示了MultiInstrument表示，它包含了MMM模型用来生成音乐的符号，以及用于上漆的BarFill表示，即生成一个或几个小节，但保留了其余的组成部分。

从作曲过程的角度来看，这些模型不进行管弦或乐器演奏，但他们从零开始或通过内画创作音乐。这意味着这些模型不选择乐器的数量，也不产生高质量的旋律或伴奏内容相关的乐器被选择。例如，MMM模型为预定义的乐器生成旋律内容，它遵循该乐器的音质特征，但当在保持其他音轨的情况下对单个乐器进行绘制或重新创建时，有时很难跟上其他乐器的琴键。这使我们得出结论，多乐器生成音乐的模型专注于端到端生成，但在器乐或管弦乐方面仍然不能很好地工作，因为它们仍然不能决定生成的音乐作品中乐器的数量。他们为他们所训练的合奏产生音乐，如LakhNES[66]，或者他们采用预定义的音轨来生成每个音轨[28]的内容。更近期的模型，如MMM，在多乐器生成方面开放了人类和人工智能之间的互动，这将允许更好地跟踪人类作曲过程，从而改善由多乐器生成的音乐。

9

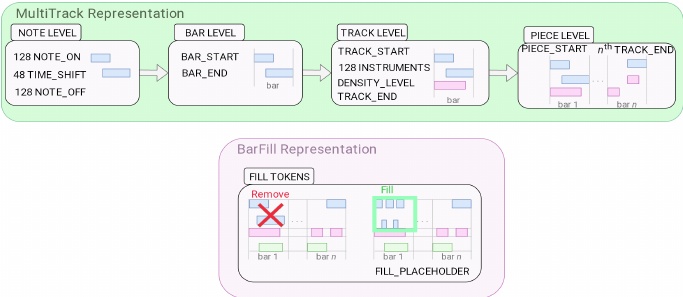


图5:从[28]复制的MMM令牌表示

**5评价与指标**

评价在音乐生成可以划分的方式，其中输出的DL模式是测量。Ji等[67]对客观评价和主观评价进行了区分。在音乐中，有必要从主观的角度来衡量结果，因为这是一种评估类型，告诉我们与人类的创造力相比，该模型带来了多少创造力。有时，计算模型结果的指标的客观评价可以让我们了解这些结果的质量，但很难找到一种方法将其与创造力的概念联系起来。在本节中，我们将展示最先进的模型如何从客观和主观的角度来衡量其结果的质量。

**5.1客观评价**

客观评价使用一些数值指标衡量模型的性能及其输出的质量。在音乐生成中，有一个问题是比较为不同目的训练的模型和用不同数据集训练的模型，所以我们给出了最常用的最先进模型的描述。Ji等[67]区分了模型度量和音乐度量或描述性统计，以及其他方法，如模式重复[68]或剽窃检测[9]。

当你想要测量一个模型的性能，最常用的指标取决于DL模型用于生成音乐是:损失，困惑，BLEU分数，精度(P)，召回(R)或F分数(F1)。通常，这些指标用于比较为相同目的而构建的不同DL模型。

Loss通常从数学的角度来表示模型的输入和输出的差异，perplexity则告诉我们一个模型所具有的泛化能力，这与模型如何生成新的音乐有更多的关系。例如，Music Transformer[24]使用损耗和困惑来比较不同Transformer架构之间的输出，以验证模型，TonicNet[69]仅使用损耗用于相同的目的，而MusicVAE[17]仅使用一个度量，表明模型具有的重构质量，但没有使用任何指标来比较其他DL音乐生成模型。

关于与音乐相关的参数，也就是那些考虑到音乐描述符的参数，我们可以发现这些参数有助于衡量音乐作品的质量。Ji等[67]将这些度量标准分为四类:音高相关、节奏相关、和声相关和风格转移相关。与音高相关的指标[67]，如规模一致性、音频垃圾、空条的比例或使用的音高类的数量，都是衡量一般音高属性的指标。与节奏相关的量度考虑到音符的持续时间或模式，例如，节奏变化，同时出现的三个或四个音符的数量，或重复音调的持续时间。和声相关指标衡量的是和弦的熵，距离或覆盖范围。这三个度量类别被MuseGAN[62]、C-RNN-GAN[70]或JazzGAN[49]等模型所使用。最后，与样式传递相关的技术有助于理解生成的样式与所需的样式有多接近或多远。其中包括风格契合、内容保存或迁移力度[71]。

**5.2主观评价**

主观的观点决定了所生成的音乐在创造性和新颖性方面是怎样的，也就是说，所生成的音乐在多大程度上可以被认为是艺术。虽然艺术包含着创造性和创造性，但是我们无法给艺术下一个定义

10

美学Sternberg和Kaufman[72]将创造力定义为对任务做出新颖和合适的贡献的能力，通常还带有质量、惊喜或有用等附加成分。创造力需要对音乐知识的本质和运用有更深的理解。Ji等人[67]认为，音乐质量的定量评价与人的判断缺乏相关性，这意味着音乐生成模型也必须从主观的角度进行评价，这将使我们对模型的创造性有一个深刻的认识。在主观评价中最常用的方法是听力测试，它通常由人类试图区分机器生成的音乐和人类创造的音乐组成。这种方法被称为图灵测试，用于测试DeepBach[9]。在这个模型中，属于不同音乐体验组的1.272人参加了测试。实验表明，模型越复杂，输出越好。MusicVAE[17]还进行了听力测试和Kruskal Wallis H-Test来验证模型的质量，结果表明采用分层解码器后，模型性能更好。MuseGAN[62]也进行了一项听力测试，将144名用户分为不同音乐体验的组，但有一个预先设定的问题，用户必须在1到5的范围内投票:和声的顺从度、节奏、结构、连贯性和总体评分。

其他的听力方法需要对生成的音乐进行评分，这被称为并行评分[67]。根据模型的生成目标，还可以向听众询问一些关于模型的创造性或生成片段的自然性的问题。在听力测试中，有一件重要的事情要记住，那就是测试人群的可变性(如果听者是有基本音乐理论知识的音乐系学生，如果他们是业余爱好者，所以他们没有任何音乐知识，或者如果他们是专业的音乐家)。听者必须有相同的刺激，也听相同的片段，并有作为参考(如果适用)相同的人类创造的片段。听觉疲劳也必须考虑到，因为如果听者长时间听类似的样本，可能会产生一种诱发的偏见。

话虽如此，我们可以得出结论，当涉及到音乐生成时，听力测试是必不可少的，因为它提供了模型质量的反馈，它们也可以找到正在研究的更好的神经网络架构或DL模型。

**6讨论**

我们已经证明，音乐是一种结构化的语言模型，具有时间和和声的短期和长期关系，这需要对其所有的深刻理解才能被建模。这一点，再加上音乐中存在的各种体例和子体例，以及在创作音乐作品时可以遵循的大量作曲策略，使得DL音乐生成领域成为一个不断发展和具有挑战性的领域。在描述了音乐创作过程和DL在音乐生成方面的最新工作之后，我们现在将讨论1.3节中提出的问题。

**目前DL的模型是否能够创作出具有一定创造力的音乐?**

第一批用DL生成音乐的模型使用的是lstm等rnn。这些模型可以生成笔记，但当它们生成长期序列时就失败了。这是因为这些网络并不处理音乐生成所需的长期序列。为了解决这个问题，并能够通过插入两个现有的主题或从分布中取样来生成短主题，我们创建了MusicVAE。但由此产生了一些问题:现有主题之间的插值是否能够在相同的音乐作品中产生高质量的有意义的内容?如果我们使用MusicVAE来创造一个短的主题，我们可以得到非常好的结果，但如果我们使用这种模型来产生更长的短语或主题，类似于输入，这些插值可以输出具有美感的主题，但有时它们并不遵循输入所具有的任何节奏或音符方向(上升或下降)模式。因此，这些插值通常不能产生高质量的主题，因为模型不了解节奏模式和音符方向。此外，和弦进行通常有倒转，在古典音乐或流行音乐、爵士乐或城市音乐的风格限制中有规则，这些规则决定了每个和弦如何被另一个和弦跟随。如果我们分析DL方法生成的多音旋律，就和声内容而言，存在质量不足的问题，因为训练生成音乐的神经网络无法理解音乐语言中出现的所有这些复杂之处，或者因为这些信息应该作为输入的一部分传递给神经网络，例如作为令牌。

**用DL来演奏音乐的最佳神经网络架构是什么?**

Transformer体系结构与不同的注意力机制一起使用，这些机制允许对较长的序列进行建模。一个成功的例子就是MMM模型[28]，它使用GPT-2生成多音轨音乐。尽管该模型使用了一个经过训练的Transformer来生成文本，但它还是生成了和声和节奏方面的连贯音乐。其他体系结构使用生成网络，如GANs或VAEs，也使用组合

11

这些架构的变形金刚。带来这些模型的力量是提取高级音乐属性(如风格)的可能性，以及组织在潜在空间中的低级特征。然后，这个潜在的空间被用来在这些特征和属性之间插入，以根据现有的作品和音乐风格生成新的音乐。

通过分析过去几年使用DL生成音乐的神经网络模型和体系结构，我们发现没有一种特定的神经网络体系结构能够更好地实现这一目的，因为用于构建音乐生成模型的最佳神经网络体系结构将取决于我们想要获得的输出。尽管如此，变形金刚和生成模型正在成为最好的替代品，在这个领域的最新作品证明了这一点。这两个模型的组合也是执行音乐生成[35]的一个很好的选择，尽管它取决于我们想要生成的输出，有时最好的解决方案来自DL和概率方法的组合。另一个需要考虑的方面是，通常情况下，音乐生成需要具有大量参数和数据的模型。我们可以通过将一个预训练的模型作为我们在前几节中描述的一些最先进的模型来解决这个问题，然后对另一个NN架构进行微调。另一种选择是使用MusicVAE建议的巨大数据集训练一个大模型，生成一个预训练的潜在空间，然后利用预训练的潜在空间训练一个数据较少的较小的神经网络，以便像midi建议的那样调整音乐作品的风格[73]。

**端到端方法能生成完整的结构化音乐片段吗?**

正如我们在3.2节中所描述的，现在已经有基于结构模板的模型可以生成结构化的音乐[39]，但是还没有一种端到端的方法可以构建结构化的音乐片段。人类作曲家遵循的音乐作曲过程与这种基于模板的方法类似。在不久的将来，人工智能可能会从头开始创作结构化的音乐，但这里的问题是，用于音乐生成的人工智能模型是否会用于从头开始创作完整的音乐作品，或者这些模型是否会作为作曲家的辅助，从而作为人类和人工智能之间的互动更有用。

**用DL创作的乐曲仅仅是对输入的模仿，还是神经网络能够以训练数据中不存在的风格生成新的音乐?**

当训练DL模型时，一些在输入中传递给NN的信息可以在输出中不做任何修改。即使如此，“MusicVAE[17]”等DL音乐生成模型也表明，新的音乐可以在不模仿现有音乐或不抄袭的情况下创作。模拟输入可能是一种过拟合的情况，而这从来不是DL模型的目标。还应该考虑到，在音乐的生成过程中，由于可能出现在一段音乐中的乐器、音调、节奏或和弦的多样性，很难进行剽窃。

**神经网络是否应该按照人类的逻辑和过程来作曲?**

我们发现，研究人员开始建立能够产生复调旋律的模型，但这些旋律在几个小节之后并不遵循任何方向。当《MusicVAE》问世时，我们有可能生成高质量的主题，这鼓励了新的研究，以过去的时间步骤为信息来生成旋律。像扩散模型[37]这样的新模型正在使用这种预先训练的模型来生成更长的序列，让旋律遵循模式或方向。我们还表明，有一些模型可以通过和弦进行调节来生成旋律，这是在流行音乐风格中作曲的方式。将人类的作曲方式与DL架构进行比较，我们可以看到这两种过程的一些相似之处，特别是在自回归模型中。自回归(AR)是指从过去的事件中预测未来的值。有些DL方法是自回归的，新模型试图通过提取过去的时间步长的信息来生成更长的序列，类似于人类创作古典音乐的过程。

**音乐生成的DL模型需要多少数据?**

如果我们看看最先进的模型，这个问题就可以部分地得到回答。MusicVAE使用LMD[33]，有370万个旋律，460万个鼓模式和16万6千个三重奏。相反，《音乐变形金刚》只使用1.100首来自钢琴电子比赛的钢琴曲来训练模型。其他的模型，如MMM采用GPT-2，这是一个预先训练的具有大量文本数据的Transformer。这让我们确认用于音乐生成的DL模型确实需要大量的数据，特别是在训练生成模型或变形器时，但采用预先训练的模型并执行迁移学习也是一个很好的解决方案，特别是对于那些在符号音乐生成的实际数据集中没有表示出来的音乐类型和子类型。

12

**当前的评价方法是否足以比较和衡量音乐创作的创造力?**

正如我们在第5节中所描述的，有两种评价类别:客观评价和主观评价。现有方法的客观评价指标相似，但缺乏一种通用的主观评价方法。听力测试是最常用的主观评价方法，但有时图灵测试只要求区分计算机作曲和人工作曲，不足以了解由神经网络创建的作曲的所有特征。解决这一问题的方法是，按照MuseGAN的建议，就图2所示的音乐特征的质量提出一般性的问题，并在DL模型中使用相同的问题和相同的评分方法，设定一个一般性的主观评价方法。

**7结论与未来工作**

在本文中，我们通过概述用于DL音乐生成的神经网络架构来描述DL音乐生成的最新技术，并讨论了在使用深度神经网络进行音乐生成时仍然存在的挑战。

使用DL架构和技术来生成音乐(以及其他艺术内容)是一个日益增长的研究领域。然而，这里也存在一些开放的挑战，如基于结构生成音乐，分析生成音乐的创造性以及构建能够帮助作曲家的互动模型。未来的工作应该专注于更好地建模长期关系(在时间和和声轴上)，以生成结构良好和和谐的音乐，不会在几个小节之后变得松散，并在绘画或人类ai互动，这是近年来越来越感兴趣的任务。还有一个悬而未决的挑战，与迁移学习或风格生成的条件反射有关，这允许不局限于出现在公共数据集中的相同的作者和类型，如JSB Chorales数据集或laklakmidi数据集，这使得大多数最先进的作品只专注于相同的音乐风格。当涉及到多仪器世代时，这一任务并不遵循人类的作曲过程，看到新的DL模型可能会很有趣，它们首先创作出高质量的旋律内容，然后自己或在人类的帮助下决定音乐作品的乐器数量，并能够根据每个乐器的音色特征编写高质量的音乐。进一步的问题与DL音乐生成应该关注的方向有关，也就是说，构建端到端模型，可以从无到有地生成高创造性的音乐，或者作曲家可以与人工智能互动的交互模型，这是未来将解决的任务，尽管人类与人工智能互动的趋势每天都在加快发展。

在DL音乐创作中还有更多的开放性问题不在本文的研究范围之内。比如，如果网络用户接受的是受版权保护的音乐，那么使用DL生成的音乐的知识产权归谁所有。我们认为这将是商业应用中的一个重要关键。这里的主要关键是定义是什么使一个组合不同于其他，有几个特征发挥了重要的作用。正如我们在第一节中提到的，这些特征不仅包括乐曲本身，还包括乐器发声时所使用的音色和效果。从作曲的角度来看，这是我们研究的范围，我们可以声明，当用DL生成音乐时，它总是可能生成与输入相似的音乐，有时生成的音乐有直接从输入中获取的模式，因此，从音乐理论、知识产权和科学的角度来看，这一领域还需要做进一步的研究，以定义是什么让一个作品不同于其他作品，以及用DL生成的音乐如何注册。

我们希望本文的分析将有助于更好地理解问题和可能的解决方案，从而对基于深度学习的音乐生成的整体研究议程做出贡献。

**致谢**

该研究得到了西班牙科学、创新和大学部RTI2018- 096986-B-C31合同和阿拉贡政府AffectiveLab-T60-20R项目的部分支持。

我们要感谢Jürgen Schmidhuber的建议。

**参考文献**

[1] https://www.copyright.gov/prereg/music.html, 2019。于2021年7月访问。

Noam Chomsky。*句法结构。*De Gruyter Mouton, 2009。

[3]大卫·科普。音乐智能实验:基于非线性语言的作曲。*新音乐研究，18(1):117-139,1989。*

13

Iannis Xenakis。新原则音乐形成了音乐创作。1981.

[5] https://koenigproject.nl/project-1/， 2019。于2021年7月访问。

我是彼得·m·托德。算法合成的连接主义方法。*计算机音乐，13(4):27 - 43,1989。*

格哈德·尼尔豪斯。*算法作曲:自动音乐生成的范例。*施普林格科学与商业传媒，2009。

Lejaren A Hiller Jr和Leonard M Isaacson。用高速数字计算机作曲。在音频工程学会会议9。音响工程学会，1957。

[9] Gaëtan Hadjeres, François Pachet，和Frank Nielsen。Deepbach:巴赫赞美诗一代的可操纵模型。在Doina Precup和Yee Whye Teh，编辑，第34届机器学习国际会议论文集，ICML 2017，悉尼，新南威尔士州，澳大利亚，2017年8月6-11日，机器学习研究论文集第70卷，第1362-1371页。Pmlr, 2017。

[10]迈克尔·c·莫泽。神经网络音乐预测作曲:探索心理声学约束和多尺度加工的好处。*联系。科学。*， 6(2-3):247 - 280,1994。

我是道格拉斯·艾克。在节奏中找到重拍的松弛振荡器网络。在Georg Dorffner, Horst Bischof和Kurt Hornik，编辑，人工神经网络- ICANN 2001，维也纳国际会议，奥地利，2001年8月21-25日，会议录，计算机科学讲座讲稿2130卷，1239-1247页。施普林格，2001。

[12] Douglas Eck和Jürgen Schmidhuber。学习蓝调的长期结构。在José， R. Dorronsoro，编辑，人工神经网络- ICANN 2002，国际会议，马德里，西班牙，2002年8月28-30日，会议录，计算机科学讲座讲稿2415卷，284-289页。施普林格，2002。

Jamshed J Bharucha和Peter M Todd。用神经网络建模音调结构的感知。*计算机音乐，13(4):44-53,1989。*

庄福珍，祁志远，段克玉，奚东波，朱永春，朱恒树，熊辉，何清。迁移学习研究综述。*acta physica sinica, 2018, 37(1): 1 - 7, 2 - 9。*

Diederik P. Kingma和Max Welling。自编码变分贝叶斯。在Yoshua Bengio和Yann LeCun，编辑，第二届学习表征国际会议，ICLR 2014, Banff, AB，加拿大，2014年4月14-16日，会议跟踪学报，2014。

[16] Irina Higgins, Loïc Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed，和Alexander Lerchner。vae:用约束变分框架学习基本的视觉概念。第五届学习表征国际会议，ICLR 2017，法国土伦，2017年4月24-26日，会议纪要。openreview。net, 2017。

亚当·罗伯茨，杰西·h·恩格尔，科林·拉斐尔，柯蒂斯·霍桑和道格拉斯·艾克。音乐长期结构学习的层次潜向量模型。Jennifer G. Dy和Andreas Krause，编辑，第35届机器学习国际会议论文集，ICML 2018, Stockholmsmässan，瑞典斯德哥尔摩，2018年7月10-15日，机器学习研究论文集80卷，4361-4370页。Pmlr, 2018。

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David ward - farley, Sherjil Ozair, Aaron Courville和Yoshua Bengio。生成式对抗网。NIPS ' 14，第2672-2680页，剑桥，马萨诸塞州，美国，2014。麻省理工学院出版社。

弗兰克·罗森布拉特。感知器:大脑中信息存储和组织的概率模型。*心理学报，65(6):386,1958。*

[20] Sepp Hochreiter和Jürgen Schmidhuber。长期短期记忆。*关键词:神经网络，神经网络，神经网络引用本文:*

福岛邦彦和三宅清。Neocognitron:一种新的模式识别算法，可容忍变形和位置偏移。*模式识别。*， 15(6): 455-469, 1982。

[22] Yann LeCun, Patrick Haffner, Léon Bottou, Yoshua Bengio。基于梯度学习的目标识别。在David A. Forsyth, Joseph L. Mundy, Vito Di Gesù和Roberto Cipolla，编辑，计算机视觉的形状，轮廓和组织，计算机科学讲座讲稿1681卷，319页。施普林格，1999。

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser和Illia Polosukhin。你所需要的就是关注。在Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett，编辑，神经信息处理系统进展30:2017年神经信息处理系统年会，2017年12月4-9日，美国，CA, Long Beach, 2017年，页5998-6008。

14

Anna Huang, Ashish Vaswani, Jakob Uszkoreit, Noam Shazeer, Curtis Hawthorne, Andrew M Dai, Matthew D Hoffman和Douglas Eck。音乐转换器:生成具有长期结构的音乐。*arXiv预印本:1809.04281,2018。*

雷蒙德·格伦·利瓦伊。*二年级儿童原创语言与音乐作品创作过程之实地调查。*博士论文，凯斯西储大学，1991年。

[26]大卫·柯林斯。音乐创作中创造性思维的综合过程模型。*音乐心理学，33(2):193-216,2005。*

查尔斯·沃尔顿。*《音乐的基本形式》*阿尔弗雷德音乐，2005年。

[28] Jeff Ens和Philippe Pasquier。探索用变压器生成条件多轨音乐。*arXiv预印本:2008.06048,2020。*

雨果·拉罗谢尔和伊恩·穆雷。神经自回归分布估计。Geoffrey J. Gordon, David B. Dunson和Miroslav Dudík，编辑，《第十四届人工智能与统计国际会议论文集》，AISTATS 2011, Fort Lauderdale, USA, 2011年4月11-13日，JMLR论文集第15卷，第29-37页。JMLR.org, 2011。

[30]梅森·布雷坦，吉尔·温伯格和拉里·p·赫克。一种基于深度神经网络的音乐单元选择方法。Ashok K. Goel, Anna Jordanous和Alison Pease，编辑，《第八届计算创造力国际会议论文集》，ICCC 2017，亚特兰大，乔治亚，美国，2017年6月19-23日，第72-79页。计算创造力协会(ACC)， 2017。

[31] Elliot Waite等人。在歌曲和故事中生成长期结构。*网络博客文章。紫红，15(4)，2016。*

[32] Gaëtan哈杰斯和弗兰克·尼尔森。使用预期神经网络生成具有位置约束的交互式音乐。*2017年。*

[33]科林·拉斐尔。*基于学习的方法比较序列，并应用于音频到midi的对齐和匹配。*博士论文，哥伦比亚大学，2016。

我是克里斯汀·佩恩。Musenet, 2019。*网址:https://openai。Com/blog/musenet, 2019。*

姜君彦，夏格斯，戴夫·b·卡尔顿，克里斯·n·安德森，瑞安·H. Miyakawa。Transformer VAE:一种结构感知和可解释音乐表现学习的层次模型。2020年IEEE声学、语音和信号处理国际会议，ICASSP 2020，巴塞罗那，西班牙，2020年5月4-8日，第516-520页。Ieee, 2020。

王子宇，张依依，张一晓，姜俊艳，杨瑞涵，赵俊波，夏格斯。钢琴树VAE:多声部音乐的结构化表示学习。*中图分类号:tq126.22文献综述*

高塔姆·米塔尔，杰西·h·恩格尔，柯蒂斯·霍桑和伊恩·西蒙。符号音乐生成与扩散模型。*CoRR, abs/2103.16091, 2021。*

Jonathan Ho, Ajay Jain和Pieter Abbeel。去噪扩散概率模型。《神经信息处理系统进展:2020年神经信息处理系统年会》，neuroips 2020, 2020年12月6-12日，虚拟，2020年。

[39] Stefan latner, Maarten Grachten和Gerhard Widmer。利用卷积受限玻尔兹曼机和约束在多声部音乐生成中强加更高层次的结构。*2016年4月，中国科学院研究生院。*

[40] Meinard Müller。*音乐处理基础-音频，分析，算法，应用。*施普林格，2015。

陈可，张伟林，Shlomo Dubnov，夏格斯，李伟。深层神经网络显式结构编码对符号音乐生成的影响。2019国际多层音乐表现与处理研讨会(MMRP)，第77-84页。Ieee, 2019。

叶寅政、萧文义、山satoru Fukayama、Tetsuro Kitahara、Benjamin Genchel、Liu Hao-Min - min、Dong Hao-Wen - wen、Chen Yian、Terence Leong、Yang yi - xuan。三和弦自动旋律和声的比较研究。*中国科学(d辑)，2001.02360,2020。*

杨伟，孙平，张毅，张颖。Clstms:两个lstm模型的组合，为象征性旋律生成和弦伴奏。2019高性能大数据与智能系统国际会议(HPBD&IS)，第176-180页Ieee, 2019。

[44]朱宏远，刘琦，袁静，秦川，李佳伟，张坤，广州，魏福如，徐元春，陈恩宏。小冰乐队:流行音乐的旋律和编曲生成框架。《第24届ACM SIGKDD知识发现与数据挖掘国际会议论文集》，KDD 2018，伦敦，英国，2018年8月19-23日，2837-2846页。Acm, 2018。

15

[45]朱宏远，刘琦，袁静，张坤，广州，陈恩宏。流行音乐的生成:从旋律到多风格编排。*ACM翻译。诺尔。发现。数据通报，14(5):54:1-54:31,2020。*

黄承志，柯蒂斯·霍桑，亚当·罗伯茨，莫妮卡·丁库莱斯库，詹姆斯·韦克斯勒，莱昂·洪，雅各布·豪克罗夫特。可接近的音乐作曲与机器学习的规模。Arthur Flexer, Geoffroy Peeters, Julián Urbano和Anja Volk，编辑，第20届国际音乐信息检索会议论文集，ISMIR 2019，荷兰，2019年11月4-8日，第793-800页，2019年。

黄成智，蒂姆·库伊曼斯，亚当·罗伯茨，亚伦·c·考维尔和道格拉斯·艾克。通过卷积对位。在Sally Jo Cunningham, Zhiyao Duan, Xiao Hu，和Douglas Turnbull，编辑，第18届国际音乐信息检索会议学报，ISMIR 2017，中国苏州，2017年10月23-27日，2017年11 - 218页。

[48]滕亦菲，赵安妮，和卡米尔·Goudeseune。为音乐服务生成不平凡的旋律。在Sally Jo Cunningham, Zhiyao Duan, Xiao Hu, Douglas Turnbull，编辑，第18届国际音乐信息检索会议论文集，ISMIR 2017，中国苏州，2017年10月23-27日，页657-663,2017。

[49] Nicholas Trieu和R Keller。Jazzgan:利用生成式对抗网络即兴创作。在MUME工作室，2018。

Shunit Haviv Hakimi, Nadav Bhonker和Ran El-Yaniv。Bebopnet:个性化爵士乐即兴创作的深层神经模型。第21届国际音乐信息检索会议论文集，ismir, 2020。

谭浩浩。和弦:一种基于和弦的使用双lstms生成音乐的方法。作者:Kazjon Grace, Michael Cook, Dan Ventura和Mary Lou Maher，编辑，第十届计算创造力国际会议论文集，ICCC 2019，夏洛特，北卡罗来纳州，美国，2019年6月17-21日，第364-365页。计算创造力(ACC)， 2019。

利昂·a·盖提斯，亚历山大·s·埃克和马蒂亚斯·贝茨。使用卷积神经网络的图像风格传输。2016年IEEE计算机视觉与模式识别会议，CVPR 2016，拉斯维加斯，美国，2016年6月27-30日，页2414-2423。计算机学报，2016。

Gino Brunner, Andres Konrad, Yuyi Wang和Roger Wattenhofer。MIDI-VAE:音乐的动力学建模和乐器演奏，以及在风格转换方面的应用。在Emilia Gómez, Xiao Hu, Eric Humphrey和Emmanouil Benetos，编辑，第19届国际音乐信息检索大会论文集，ISMIR 2018, 2018年9月23-27日，法国巴黎，2018年747-754页。

洪小志，王仲洋，杨义宣，王新民。利用迁移学习技术改进爵士乐旋律的自动生成。2019亚太信号与信息处理协会年会，APSIPA ASC 2019，中国，兰州，2019年11月18-21日，第339-346页。Ieee, 2019。

王子玉，王丁苏，张一晓，夏格斯。可控多音音乐生成的可解释表示学习。*国家自然科学基金资助项目(2008.012125,2008.0125,2008.0125,2009.07)。*

吴世伦和杨义宣。musemorphse:全歌曲和细粒度的音乐风格转移，只需一个变压器VAE。*CoRR, abs/2105.04090, 2021。*

陈志谦、吴志伟、卢彦成、Alexander Lerch、卢昌田。学习融合音乐流派与生成对抗双重学习。在Vijay Raghavan, Srinivas Aluru, George Karypis, Lucio Miele和Xindong Wu，编辑，2017 IEEE国际数据挖掘会议，ICDM 2017，新奥尔良，洛杉矶，美国，2017年11月18-21日，第817-822页。计算机学报，2017。

乔治·格罗夫。*贝多芬和他的九部交响曲，334卷。*快递公司(1962年)。

[59] Ertu˘grul Sevsay。*剑桥编曲指南。*剑桥大学出版社，2013。

[60] Soo-Yol Ok Semin Kang和Young-Min Kang。自动音乐生成和基于机器学习的评估。多媒体与信号处理，vol . 32, no . 4, 2012。

[61]朱恒，拉奎尔·乌尔塔桑，桑佳·菲德勒。来自PI的歌曲:一个音乐上可信的流行音乐一代的网络。在第五届学习表征国际会议上，ICLR 2017，土伦，法国，2017年4月24-26日，研讨会Track Proceedings。openreview。net, 2017。

[62]董浩文，杨丽嘉，杨义宣。Musegan:用于符号音乐生成和伴奏的多轨道顺序生成对抗网络。在Sheila A. McIlraith和Kilian Q. Weinberger，编辑，第三十二届AAAI人工智能会议论文集(AAAI-18)，第三十届人工智能创新应用(IAAI-18)和第八届AAAI教育研讨会

16

*人工智能研究进展(EAAI-18)， 2018年2月2-7日，美国路易斯安那州新奥尔良，2018年2月2-7日，第34-41页。*AAAI出版社，2018。

[63]于兰涛，张渭南，王军，于勇。Seqgan:具有策略梯度的序列生成对抗网。Satinder P. Singh和Shaul Markovitch编辑，《第三十一次AAAI人工智能会议论文集》，2017年2月4-9日，美国加利福尼亚州旧金山，2852-2858页。AAAI出版社，2017。

[64]董浩文，杨义宣。基于二元神经元的卷积生成对抗网络的多音音乐生成。在Emilia Gómez, Xiao Hu, Eric Humphrey和Emmanouil Benetos，编辑，第19届国际音乐信息检索会议记录，ISMIR 2018, 2018年9月23-27日，法国巴黎，2018年9月23-27日，190-196页。

[65] Andrea Valenti, Antonio Carta, Davide Bacciu。通过敌对的自动编码器学习潜在的风格意识符号音乐表征空间。*中国石油大学学报(自然科学版)，2020。*

[66]李义廷，李义廷。Lakhnes:通过跨领域的预训练来提高多乐器音乐的生成。Arthur Flexer, Geoffroy Peeters, Julián Urbano和Anja Volk，编辑，《第20届国际音乐信息检索会议论文集》，ISMIR 2019，荷兰代尔夫特，2019年11月4-8日，第685-692页，2019年。

[67]纪树雷，罗静，杨新宇。深度音乐生成的全面调查:多层次表现、算法、评估和未来方向。*国家自然科学基金资助:*

[68]王成义。用可变马尔可夫预言引导音乐合成。在Philippe Pasquier, Arne Eigenfeldt和Oliver Bown，编辑，音乐元创造，论文来自2014 AIIDE Workshop, 2014年10月4日，罗利，NC，美国，卷WS-14-18的AAAI Workshop。AAAI出版社，2014。

[69]奥马尔·珀拉查。用丰富的特征编码改进复调音乐模型。*CoRR, abs/1911.11775, 2019。*

[70]奥洛夫·莫根。C-RNN-GAN:具有对抗训练的连续循环神经网络。*2016年3月，中国科学院研究生院院士。*

[71]赵苏木。象征音乐流派转移与自行车。Lefteri H. Tsoukalas， Éric Grégoire，和Miltiadis Alamaniotis，编辑，IEEE第30届人工智能工具国际会议，ICTAI 2018, 2018年11月5-7日，Volos，希腊，页786-793。Ieee, 2018。

[72]李志强，李志强。*人类创造力的本质。*剑桥大学出版社，2018年。

[73]莫尼卡·丁库莱斯库、杰西·恩格尔和亚当·罗伯茨是编辑。*midi:使用用户数据个性化音乐vae模型，2019年。*

17