项目编号：201810006006



国家大学生创新创业训练计划

结题报告

项 目 名 称：MusiCoder:基于深度神经网络的

轻量化音频质感转换系统

立 项 人： 彭煦潭

所 在 单 位：北京航空航天大学计算机学院

手 机： 13121212519

电 子 信 箱： [pengxt@act.buaa.edu.cn](mailto:pengxt@act.buaa.edu.cn)

资 助 金 额： 7500 元

指 导 教 师： 李建欣

训 练 场 地： 北航计算机新技术研究所

研 究 期 间： 2018.04 - 2019.03

北京航空航天大学教务处

项 目 摘 要

音乐质感指包括风格流派、音质音色等在内的音乐高阶特征，它对于人类音乐审美感受和情感表现均具有极其重要的作用，是学术领域关注的重点问题。对给定的音乐进行质感转换，不仅可以助力相关的科学研究，而且在个人娱乐领域也对应着非常丰富的应用场景，并能激发专业创作者的灵感，具有巨大的市场价值。

然而，由于研究对象的抽象性和问题结构的复杂性，据我们所知，尚不存在满足实际应用要求的解决音乐质感转换问题的方法或模型。本文首先提出了一种全新的频谱化描述方法，成功实现了对音乐质感特征的可视化刻画；在对图像风格转换等其它领域模型所用网络进行迁移的基础上，搭建并训练了一款包含残差层和卷积层的深度神经网络，从而仅通过单次前向传播即可快速转换输入的音乐质感频谱；基于连续差分和并行架构优化，借助Griffin-Lim算法和音频幅值正则化方法，还实现了一款高性能的逆向编码器，可用于利用本文所提出的频谱化表示，快速重建目标音乐；最终成功将这一端到端的音乐质感转换网络部署在云端，并面向多种操作系统平台，开发了支持高并发的客户端。作为音乐质感转换领域内第一款取得完整产品形态的产品，MusiCoder系统具备算法先进、转换效果突出、可用性高、集成度强、可解释、轻量化等诸多优势，具有广阔的应用前景。

Abstract

Music texture is a collective concept of many high-level musical features, e.g. style, rhythm, melody, etc. As it plays a significant role in the artistic taste and emotional expression, it has attracted much attention from academia. Texture transfer on given music pieces can not only booster related researches, but also produce various application scenarios and inspire professional composers, which implies great marketing potential.

However, due to the abstractness of the subject and the complexity of the problem per se, to the most of our knowledge, there are no such method of model which can be put into real use. We propose a novel spectral representation aiming to visualize the texture feature of music. A deep neural network with convolutional layers and residual blocks is built and trained based on the model of image style transfer, which can transfer music texture with a single forward-propagation. Based on differential method, parallel optimization, Griffin-Lim Algorithmand magnitude normalization of audio signal, we implement a high-speed encoder which can reconstruct output audio via our proposed spectral representation. We successfully deploy this end-to-end music texture transfer network on cloud servers and develop highly concurrent clients for various operating systems. As the first consummate system for music texture transfer, MusiCoder has obvious advantages in progressiveness, output quality, serviceability, integration, interpretability and computational complexity.

关键词（3～5个）：

音乐质感转换Music texture transfer，频谱表示Spectral representation，深度神经网络Deep neural network，轻量化系统Light-weighted system，云应用Cloud application

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 项目名称 | | | | | MusiCoder:基于深度神经网络的轻量化音频质感转换系统 | | | | | | | | | | |
| 项目来源 | | | | | 科学研究 | | | | | | | | | | |
| 审批经费（元） | | | | | 7500元 | | | | 起止时间 | 2018年4月 至 2019年3月 | | | | | |
| 项目组成员(第一人为立项人) | 姓名 | 学号 | | 专业/年级 | | | 学院 | | 项目分工 | | E-mail | | | 联系电话 | |
| 彭煦潭 | 15061204 | | 计算机科学与技术/2015级 | | | 计算机学院 | | 算法设计，整体架构，实现与设计，生产与展示 | | pengxt@act.buaa.edu.cn | | | 13121212519 | |
| 蔡帜 | 15061077 | | 计算机科学与技术/2015级 | | | 计算机学院 | | 算法设计，实现与调试 | | felixfelicis@buaa.edu.cn | | | 13261428544 | |
| 金银株 | 14231028 | | 计算机科学与技术/2015级 | | | 计算机学院 | | 实现与调试（前端） | | jinyinzhu@buaa.edu.cn | | | 13552565917 | |
| 张嘉 | 15061150 | | 计算机科学与技术/2015级 | | | 计算机学院 | | 界面设计、产品调试 | | facade1801@163.com | | | 18109866992 | |
| 石发强 | 14061115 | | 计算机科学与技术/2014级 | | | 计算机学院 | | 服务器运维 | | 14061115@buaa.edu.cn | | | 18293106300 | |
| 导 师 | 姓名 | | 李建欣 | | | 单位 | | 北京航空航天大学计算机学院 | | 职务 | | 博导 | 职称 | | 正教授 |
| 电话（手机） | | 010-82339274 | | | E-mail | | lijx@buaa.edu.cn | | 性别 | | 男 | 年龄 | | 40 |

|  |
| --- |
| **一、**项目背景  你是否曾经羡慕过阿卡贝拉艺术家们用嗓音模仿各种音效的神奇技能？你是否曾经幻想过自己也能像专业的混音师一样，将乐器演奏的曲目与其它的声音素材搭配融合，成就一段自然优美的旋律？你是否憧憬过拥有这样一款工具，像给照片选取滤镜一样，自由地改变自己喜欢的音乐的风格与流派？  在讨论这些问题时，我们实际谈到了音乐质感（music texture）。它是包含风格、音色、曲式、织体在内的一系列音乐要素共同组成的概念[1]。这些要素通过不同的结合方式，能借助音响效果带给人们迥异的感受。不同的音乐质感往往表达了不同的情感或主题。  相关商业应用  特征转换一直是音乐与音频处理的热点问题。音乐质感转换可以帮助普通用户改变自己歌唱或者演奏的音乐，从而为之赋予更加丰富的内涵，还可以为专业创作者提供新的思路与灵感，因而具备很高的商业价值，备受关注。当前被广泛投入应用的软件都是基于滤波器对信号进行的处理来完成各类音乐转换任务。其中，Adobe Audition作为一个专业音频编辑和混合环境，可以提供多种信号处理效果；Steinberg Cubase是一款面向创意的数字音频工作站；Abelton Live则是一款支持实时效果的专业音序器。但是，这类软件的安装包动辄数个GB甚至超过数十个GB，对运行环境有着较高的要求，且功能复杂，难以入门[3]。同时，要完成音频质感转换这样的复杂任务，往往严重依赖用户的经验与直觉，耗费大量人力物力，极难操作。更重要的是，由于它们使用的滤波器均基于固定的规则，而现实的音乐种类繁多，加之不同个体心中的理想生成效果通常不同且不可知的，因此在实际操作中，这类软件往往表现出扩展性与普适性差的特点，甚至可能掣肘音乐人的创作。  相关机器学习方法  作为基于统计的方法，机器学习在天然属性上相对于上述基于规则的商业软件具备较大优势。事实上，在音频之外的此类转换问题上，机器学习已经取得了令人印象深刻的效果。借助卷积神经网络，Gatys等人成功完成了图像风格转换[4]，而Johnson等人提出了与之效果近似的前馈网络方案，几乎可以实时地完成任务[5]；在自然语言处理领域，Fu等人提出了用非平行的数据学习风格转换，并在实验中取得了可以与人类相媲美的高正确率[6]。  在向音频领域扩展的过程中，也出现了一些探索性方法。Ulyanov用不同的频率代替了图像问题中不同的颜色频道，得到了一个沿时序表达不同频率上音乐强度的一维频谱。他搭建了一个权重随机的有4096个卷积核的单层网络进行音乐质感与风格的转换[7]；类似地，Wyse用一个类似于Gatys用于图像风格转换的网络[4]进行音乐强度波谱的直接转换，并使用Griffin-Lim算法[2]进行重建，但是效果并不理想[8]； Barry等人在这个方法的基础上，引入梅尔频谱和常数Q频谱等新的谱图进行实验。但是他们的方法仍然没有在效果上取得明显的突破[9]。这些方法的时间和空间性能都很差。其它尝试包括在时域上直接进行变换[10]和分离音乐内容和风格后分别变换[11]等。在生成式方法的研究中，来自谷歌DeepMind实验室的基于原始波形与卷积神经网络的Wavenet[12]和索尼CSL实验室基于循环神经网络设计的DeepBach[13]均为音乐质感转换提供了新思路。 |

|  |
| --- |
| **二、**研究内容    对音乐质感转换问题，我们给出如下定义：给定音频片段*A*和音乐片段*M*，当且仅当*A*和*M*共享可判定的相同质感特征时，我们有关系为真。类似地，当且仅当*M*和另一音乐片段*N*被判定为同一段音乐内容的不同版本时，关系为真。那么，音乐质感转换即指给定音频-音乐二元组*(Mt,At)*，生成音乐片段*Mc*，并满足为真。  本项目即是利用机器学习算法与自研的信号处理算法，在开发一款通用的轻量化音乐质感转化系统，并部署在云端。 |
| **三、**技术路线及解决方案  我们的系统主要包含两个模块：部署在服务器后端的音乐质感转换网络和供不同用户进行并发交互的客户端程序。我们通过网络连接实现二者之间的数据通信。系统的整体架构见图1。  图片包含 屏幕截图  已生成高可信度的说明  **图1 Musicoder整体架构方案**  音乐质感转换网络  我们将在这一节介绍并讨论本系统的核心算法。我们的基本思路是通过实现将音频转换为图像，从而可以通过问题域的迁移，借助图形图像相关方向的方法与模型实现音乐质感转换。同时，我们还需要确保可以从转换得到的频谱化表示出发，重建得到输出音频，且重建过程的损失是有界的。在这个思路的指导下，我们设计了一款端到端网络，主要组成包括由音频生成自研彩色频谱的转换器、支持彩色频谱转换的深度神经网络、通过分析频谱从而重建输出音频的编码器。整个网络的设计见图2。其中，虚线部分表示仅在训练过程中会出现的数据流。  图片包含 屏幕截图  已生成极高可信度的说明  **图2 音乐质感转换网络**  1、音频-频谱转换器  音频信号图形化的经典做法是通过短时傅里叶变换(STFT)，确定时变信号其局部区域正弦波的频率与相位，继而将一个时间段内的相关数据输出到相互正交的两个矩阵。接下来，将表示频率与相位平面上信号强度的矩阵保留，全过程如式1所示，  （式1）  其中*w(n)*表示以0为中心的高斯窗口，*x(n)*代表音频信号。我们选取其中的强度部分（实数），记为*Xi*。我们可以将它按相对灰度投影为二维图像，如图3(a)所示。这样的变换与投影很难捕捉和突出与音频质感相关的高阶信息。  声学与心理学方面的研究表明，人类对声音的感知并不是按强度线性变化的[14]。因此，我们从分贝这一更好地刻画人类听觉的量度标准出发，对我们在式2中得到的矩阵数值通过式1进行重整，  （式2）  其中*r*代表矩阵*Xi*的最小元。  我们得到的新矩阵*XdB*的图像如图3(b)所示。我们注意到，由于有效像素的增加，这一新频谱已经包含了音频质感信息。但是，我们还发现，由于环境噪音等问题，有部分无效信息被引入，它们会严重降低之后神经网络训练与转换的效果。我们借助了近年来在深度学习模型里被广泛采用的注意力机制，通过我们设计的式3进行滤波，  （式3）  其中是取值在[0,1]范围的超参，\*表示Hadamard乘积，返回矩阵*X*的最小元。当取黄金分割比0.618时，滤波得到的信号*Xa*和被滤去的强度图像分别如图3(c)和图3(d)所示。通过与图3(b)进行对比，我们发现多数的噪音被有效过滤，而滤除部分几乎不包含任何语义信息。  通过对图像风格转换网络的考察，我们发现色彩频道和网络的感知能力具有非常强的相关性，而Ulyanov的研究也对这一结论进行了佐证[7]。因此，在进行图像化处理时，我们将频谱*Xa*中的单通道数值元素映射到RGB色彩空间。为此，我们采用了magma色彩方案,将其转换得到*Xrgb*，如图3(e)。  图片包含 户外  已生成高可信度的说明  **(a)*Xi* (b)*XdB* (c)*Xa***  图片包含 户外  已生成高可信度的说明  **(d)*XdB - Xa*  (e)*Xrgb* (f)*Yrgb***  **图3 音频质感转换各阶段频谱图**  其中，magma色带是由256个离散的RGB三元组按序排列的，代表从蓝色到黄色的均匀色彩分布。由于所有的三个通道上的数值均严格满足全序关系，因此在实际操作中，我们通过三通道的和来建立RGB矩阵和单通道矩阵之间的联系。  2、深度神经网络  图片包含 屏幕截图  已生成极高可信度的说明  **图4 用于频谱质感转换的深度神经网络**  我们借鉴了取得杰出实验效果的快速图像风格转换网络，搭建了用于实现频谱上的质感转换的深度网络。如图4所示，这个网络首先包含3个搭配了ReLu非线性激活函数的卷积层，接下来是5个残差层，3个通过ReLu激活的卷积层和一个生成输出的tanh激活层。我们还使用了Ulyanov等人提出的实例正则化方法（Instance Normalization）[15]以提高转换质量。我们通过使用预训练好的损失网络，通过式4和式5，来分别计算转换生成的频谱与内容输入和质感输入频谱之间的损失，  （式4）  （式5）  其中，和分别表示内容输入频谱和生成音频频谱的特征图，和分别表示内容输入频谱和生成音频频谱在层的Gram矩阵。Gram矩阵通过他们的特征图集合*X*按式6进行了定义，  （式6）  进而，如式7，我们可以写出完整的目标函数，  （式7）  ，和分别代表内容损失，质感损失和用于改善输出频谱局部连续性的总变量正则项。  在训练过程中，我们选定一个目标质感频谱，将大量的内容频谱作为训练集。我们通过损失网络，在每例通过转换网络生成得到频谱输出时进行反向传播，更新转换网络的权重，迭代地降低目标函数值，提升整体转换效果。在完成训练的转换网络上，输入的新内容频谱仅需经过单次前向传播，就可以快速获得目标质感特征，实现设计规格。  3、频谱-音频编码器  为了通过给定的频谱图重建网络，我们需要将频谱图还原为原始状态下的强度矩阵，进而设法获得匹配的相位矩阵，获得输出音频。首先，我们要从RGB色彩空间的彩色频谱中取得各个位置对应的实际数值。为此，我们基于连续求差分的方法，设计了快速利用magma色谱降维到单值的算法，   |  | | --- | | **输入： 数值递增的RGB三通道magma色带列表,RGB三通道频谱,**  **输出： 单通道数值频谱,** | | ***初始化：;***  **forto**  **end for**  **return** |   接下来，我们通过反解式2中的函数关系，获得了复原到线性量度的近似映射，如式8，  （式8）  此处的*r*与音频-频谱转换器中的*r*取值相同。  由于magma色带上的RGB值是有界且离散分布的，因而矩阵每个元素的误差是有界的。由此，我们确保了式8中的变换引入的误差是有界的。  获得转换生成的强度矩阵后，我们利用Griffin-Lim算法，对相位矩阵进行重建。该算法通过迭代地使用自编码器，根据强度信息不断去拟合对应的相位值。随着轮数的增加，获得的相位值与强度值合成的音频越接近实际，且对应误差随迭代数收敛至0。相关证明在此不进行赘述。  为了提升本编码器的效率，我们对它进行了并行优化，即在进行相位估值时，我们将输出频谱进行切分，在获得相位矩阵后进行串列，从而最大限度利用了资源，整体速度得到了很大提升。  客户端及通信单元  为了验证我们提出的模型在实际场景下的可用性，我们设计并实现了运行在个人电脑上的客户端软件（界面如图5所示），并通过相应的通信单元进行会话管理，确保客户端和服务器端的数据交换。  图片包含 屏幕截图  已生成极高可信度的说明  **图5 Musicoder客户端界面**  客户端主要实现了与用户的个性化交互及部分信号的预处理。用户可以选择或拖拽待转换的音乐进入客户端，音频的开头部分将会被截取并上传服务器，服务器端程序会发起一个新会话，将得到的歌曲转换为对应频谱并发还，以确认服务和通信状态正常。用户可以选取最喜欢的音乐片段并在客户端进行截取，并能几乎实时地获得对应的频谱。选定目标歌曲后，用户可以试听待选质感并进行选择。我们还提供了三种不同的转换质量供用户选择，质量越高，耗时越长。完成全部选择并点击转换按钮后，服务器端会分配资源给对应的新任务，并在转换完成后将输出频谱和对应的音频发送至客户端。用户可以将其保存在本地任意路径，也可以改变之前做过的任何一项选择，以开启一项新的转换任务。值得一提的是，为了提升用户体验，我们在客户端的所有音频播放单元都设置了滤波器，用于音频强度归一化，从而稳定音量，便于用户进行比较和调整。  我们的客户端及通信单元在架构上与服务器端具有极低的耦合度，因而扩展性极强。同时，经过测试，我们的会话管理与任务分配模式确保本系统可以高效应对包含多用户的并发场景。  实验与验证  （一）神经网络训练细节  我们选取Free Music Archive(FMA)数据集[16]作为内容来源，即本模型的训练集。它包含106574种不同风格的30秒歌曲片段，我们将每段曲子均匀分成三份后，输入网络进行训练。我们基于TensorFlow框架搭建并训练了我们的神经网络，使用LibRosa库完成了与音频信号处理相关的工作。我们选取的频谱尺寸为1025\*862像素。在进行傅里叶变换时，我们选取窗口尺寸为2048。在训练前馈转换网络时，我们的批尺寸(batch size)设置为16，共训练10轮以得到满意的结果。我们选取1e-3作为学习率。在损失网络中，我们选取了经有ImageNet[17]训练得到的VGG-19网络[18]，并分别设置，和为7.5，5e2和2e2。在频谱-音频编码器部分，我们设置Griffin-Lim算法的迭代次数为1e2，即最终系统原型的标准音质对应数值。  （二）系统部署与测试  我们的服务器端部署在了一台配置为双核CPU与4G内存的云服务器上。借助pyQt框架与ffmpeg工具包，我们开发了面向Windows和Mac系统的客户端。这一搭配代表了一种相当典型的应用场景。  经测试，在转换10s的音乐，并选择输出标准音质时，本系统在除了转换之外的环节均做到几乎实时。在质感转换环节，平均用时仅需30.84秒，甚至短于相同时长的音频在本地运行的Adobe Audition上获得近似音质输出的时间，更显著强于Gatys等人的图像风格转换同等条件下超过十分钟的耗时[4]。在计算资源占用方面，全流程服务器端内存占用平均峰值仅为16.7%(1.33G)，堪称极其轻量的深度神经网络。同时，我们也进行了多用户连续并发转换的压力测试，整个系统运行正常。  由上可见，作为此领域第一款面向实际应用的产品，本系统在易用性、交互性、轻量化、稳定性等诸多方面都具备明显的优势。  （三）转换效果测试  在音乐质感转换方面，我们选取水流音效、脉冲音效和未来感音效作为目标质感，训练得到了对应网络。我们选取了不同流派、不同年代、不同演唱或演奏形式的许多歌曲进行了转换。  同时，我们进行了音频质感合成实验，并选取了相关研究中被反复借鉴与讨论的Ulyanov等的方法[7]作为基准方法。音频质感合成对于学术届的相关研究具有重要意义。我们设置水流声作为质感输入，用基于国际万维网联盟的Webaudio接口[19]采集获得粉噪声（强度和频率成倒数分布的噪声）作为内容输入，并设置目标函数中的内容权重*Lcontent*为0，从而最大程度排除内容对质感合成的影响。涉及的四段音频音频各自的频谱表示见图6。  图片包含 户外  已生成高可信度的说明 图片包含 家具, 窗帘, 夜晚, 深色  已生成高可信度的说明  **(a)粉噪声 (b)质感输入（水流音效）**  图片包含 家具, 窗帘  已生成高可信度的说明  **(c)基准方法输出 (d)Musicoder输出**  **图6 音频质感合成实验频谱图**  我们注意到，与图6(c)相比，图6(d)中本系统的输出不但与图6(b)中的质感输入共享了更多特征，而且来自图6(a)中粉噪声的内容信息也少得多。同时，我们的输出音频更为洁净，质量更高。  为了定量地进行对比，我们分别选取了两种方法的目标函数值作为评判指标，结果如图8所示。  **图7 音频质感合成中两种方法目标函数值对比图**  图7中，基准方法经计算得到的损失值在两种不同的评判指标下，均是相同实验条件下Musicoder的数倍。由此可以看出，我们的模型在捕捉质感特点与合成对应输出音频方面，均具备明显的优势。 |
| 四**、**项目特色与创新点   1. 通过集成新型可重建音频信号频谱表征方法和转换网络，提出了一种端到端的音频质感转换算法 2. 率先实现并在云端部署了满足实际生产要求的深度音乐质感转换系统，并可同时用于音乐质感抽取等其它下游任务 3. 提出了一种新型评判方法，可用于全面地评估类似系统的转换质量。相关指标下，本系统达到了业余人类转换者的水平 |
| 五、项目总结（不少于500字）  本项目提出了一种端到端的音乐质感转换方法，并将它部署在云端，实现了具有高可用性的跨平台系统。我们首先设计了一种可逆的频谱表示，将对音乐质感特征的研究转化为了对相应图像的考察；进而，架构并训练了一款深度神经网络，完成了音乐质感的转换，并实现了用于重建输出音频的编码器；最后，通过对算法的优化，成功开发出了可以投入实际应用的MusiCoder系统。  实验表明，作为第一款面向应用的音乐质感转换模型，MusiCoder在转化效果、计算性能、扩展性与稳定性等各个方面均展示出了显著的优势，在消费娱乐级市场和专业级音乐处理软件市场均具有广阔的前景。同时，它还在音频质感合成等任务上取得了突出的效果，对相关的学术研究亦能起到推动作用。 |
| 六、项目进度安排 （从项目提出、论证、开展工作、到答辩验收全部过程）   * 项目提出：2018.01 * 前期调研：2018.02 * 正式立项：2018.04 * 算法设计：2018.04 * 架构设计：2018.04 * 后端实现：2018.05 * 客户端与网络通信模块实现：2018.05 * 调试与demo发布：2018.06 * 论文写作与投稿：2018.08 * 其它材料与物料准备：2018.12 * 答辩验收：2019.03 |
| 七、具体成果   * 客户端安装包：<https://sourceforge.net/projects/musicoder/files/v0.9.9-beta.1/> * 开源项目（75+ star，25+ fork）：<https://github.com/Pzoom522/MusiCoder> * 发表的论文：X. Peng, C. Li, Z. Cai, F. Shi, Y. Liu, and J. Li. [*A Lightweight Music Texture Transfer System*](https://arxiv.org/abs/1810.01248). CoRR **abs/1810.01248** (2018) * 学术海报：<https://www.doi.org/10.13140/RG.2.2.35251.89120> * 软件著作权：**MusiCoder音乐质感转换系统v1.0**. 著作权人：彭煦潭;蔡帜;金银株;张嘉. 原始取得. 首次发表日期：2018年04月02日. 登记时间：2018年5月14日. 登记号：2018SR336023 * 取得的部分奖项： * “华为杯”全国智能设计大赛一等奖 * 全国大学生人工智能创新大赛第四名 * 国际大学生类脑计算大赛优秀奖 * 全球AI艺术大赛创新应用奖 * 北航“冯如杯”科技竞赛一等奖 |
| 八、项目经费明细（不少于200字）   |  |  |  | | --- | --- | --- | | **日期** | **花费项目（用途+数量）** | **金额（元）** | | 2018.07.24 | 三星存储卡\*1 | 199 | | 2018.07.05 | 阅读器\*1 | 958 | | 2018.10.14 | 打印费\*1 | 350 | | 2018.07.11 | surface键盘\*1 | 678 | | 2018.10.04 | 纸质文具及用品\*1 | 131 | | 2018.08.20 | 广告宣传印刷品 | 80 | | 2018.06.28 | 声卡板\*1 | 350 | | 2018.06.05 | 图书\*1 | 59.8 | | 2018.11.09 | 打印费\*1 | 133.9 | | 2018.06.29 | 图书\*1 | 177.7 | | 2018.06.27 | 胶水\*1 | 34.82 | | 2018.07.21 | 图书\*1 | 51.4 | | 2018.09.08 | 触控笔\*1 | 668 | | 2018.11.04 | 数码产品及配件\*1，配送费\*1 | 16.9 | | 2018.09.08 | 通讯器材及配件\*1 | 139.9 | | 2018.12.06 | 收派服务费\*1 | 19 | | 2018.12.07 | 收派服务费\*1 | 20 | | 2018.12.08 | 收派服务费\*1 | 22 | | 2018.06.27 | 尤里克克\*1，调音器\*1 | 97.41 | | 2018.07.24 | 三星存储卡\*1 | 171.55 | | 2018.07.05 | 阅读器\*1 | 825.86 | | 2018.10.14 | 打印费\*1 | 339.81 | | 2018.08.09 | MAVIC桨叶保护罩\*1 | 85.34 | | 2018.07.11 | surface键盘\*1 | 584.48 | | 2018.10.04 | 纸质文具及用品\*1 | 127.18 | | 2018.08.20 | 广告宣传印刷品 | 77.67 | | 2018.06.28 | 声卡板\*1 | 333.33 | | 2018.06.05 | 图书\*1 | 59.8 | | 2018.11.09 | 打印费\*1 | 130 | | 2018.07.12 | 鼠标\*1 | 132.76 | | 2018.06.29 | 图书\*1 | 177.7 | | 2018.09.08 | 键盘盖\*1 | 679.31 | | 2018.06.27 | 胶水\*1 | 30.02 | | 2018.07.24 | 回车键\*1 | 44.66 | | 2018.07.21 | 图书\*1 | 51.4 | | 2018.09.08 | 触控笔\*1 | 575.86 | | 2018.11.04 | 数码产品及配件\*1，配送费\*1 | 14.57 | | 2018.09.08 | 通讯器材及配件\*1 | 120.6 | | 2018.07.20 | 参赛报名费\*1 | 188.68 | | 2018.07.21 | 参赛报名费\*1 | 754.72 | |
| 九、应用前景评价 （不少于500字）  据我们所知，对于音乐质感转换问题，由于其研究对象的抽象性和问题结构的复杂性，尚不存在像MusiCoder这样基于深度学习算法构建的轻量化模型。我们的系统可以面向非常丰富的应用场景，具有巨大的价值：   1. 受欢迎的个人娱乐应用   具体的模式可以参考之前非常火爆的、利用神经网络完成照片风格转换的手机应用Prisma。用户可以上传自己唱的歌曲，然后通过MusiCoder改变其质感，得到有趣动听的输出，进而分享到社交网络；   1. 面向专业人士的生产力工具   传统模式下，对质感这一音乐基本属性的改变意味着对音乐制作全流程的重复，非常浪费人力物力。但是，配合MusiCoder,音乐制作人可以方便地利用已完成的乐曲，获得不同质感的预览版本，成倍加速创意迭代速度；   1. 相关领域科研的上游算法库   我们的相关论文涉及的音频信号表征方法对于音乐信息检索(MIR)等社区具备重要意义。它成功地沟通了音频与图像两个领域的信号，从而能够作为引入计算机图形图像领域成熟模型和方法的桥梁；更重要的是，它在突出质感特征同时，保留了音乐内容信息，并且给出了对应的复原算法，这使得它可以成为音乐特征提取、音乐类别划分等诸多下游任务的基础方法。如果论文被录用，除生产环境外，我们还计划开源训练模块的相关代码。   1. 其它相关应用的集成模块   本系统内核可以嵌入其它相关应用，比如当前的在进行音乐质感转换时效率低且效果差的专业软件，从而弥补相关功能短板。或者，对于需要实现音乐信号图像化的应用，MusiCoder的组件也可以单独进行集成。 |
| 十、存在的问题及建议  未来，我们将会在以下几个方面继续提升项目，以达到更好的效果：   1. 尝试其它的频谱表示方法。项目将引入梅尔频谱、常数Q频谱等其它变换，与当前的短时傅里叶变换进行集成，从而从更多角度对音乐质感进行刻画； 2. 推出面向更多平台的发行版。下一步，本系统客户端计划迁移到移动端，以便满足更多用户在不同场景下的使用需求； 3. 持续扩展程序可用性。当前的系统主要对短片段音乐进行质感转换。随着版本迭代，MusiCoder将支持转换时长更长且可变的片段。同时，系统将开放相关API，面向包含自行进行新质感训练等在内的一系列需求，向不同的用户群推出更具针对性的服务。 |
| 十一、参考资料（整个工作中参考和查阅的文件、书籍等，按照文献引用的顺序和标准规范列出）  [1]D. Griffin and Jae Lim. Signal estimation from modified short-time fourier transform. IEEE Transactions on Acoustics, Speech, and Signal Processing, 32(2), Apr 1984.  [2]E Goldstein. Sensation and perception. Wadsworth, Cengage Learning, Belmont, CA, 2014.  [3]Strähle, Jochen, and A. C. Kriegel. Fashion and Music: A Literature Review. Fashion & Music. 2018.  [4]L. A. Gatys, A. S. Ecker, and M. Bethge. Image style transfer using convolutional neural networks. In 2016 IEEE Conference on Computer Vision and Pattern Recognition, June 2016.  [5]Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super resolution. In European Conference on Computer Vision, 2016.  [6]Zhenxin Fu, Xiaoye Tan, Nanyun Peng, Dongyan Zhao, and Rui Yan. Style transfer in text: Exploration and evaluation. In Proc. of the 32nd AAAI Conference on Artificial Intelligence, 2018.  [7]Dmitry Ulyanov and Vadim Lebedev. Audio texture synthesis and style transfer. https://dmitryulyanov.github.io/audio-texture-synthesis-and-styletransfer/, 2016.  [8]L. Wyse. Audio Spectrogram Representations for Processing with Convolutional Neural Networks. In Proc. of the 1st International Workshop on Deep Learning and Music joint with IJCNN, volume 1, 2017.  [9]Shaun Barry and Youngmoo Kim. style transfer for musical audio using multiple time-frequency representations, 2018.  [10]Iman Malik and Carl Henrik Ek. Neural translation of musical style. In Proc. of the NIPS 2017 Workshop on Machine Learning for Audio Signal Processing, 2017.  [11]Gautham Santhosh. Audio style transfer. http://gauthamzz.com/2017/09/23/AudioStyleTransfer/, 2017.  [12]Aäron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew W. Senior, and Koray Kavukcuoglu. Wavenet: A generative model for raw audio. In SSW, 2016.  [13]Gaetan Hadjeres, Francois Pachet, and Frank Nielsen. DeepBach: a steerable model for Bach chorales generation. In Proc. of the 34th International Conference on Machine Learning, International Convention Centre, 2017.  [14]JoshH. McDermott and EeroP. Simoncelli. Sound texture perception via statistics of the auditory periphery: Evidence from sound synthesis. Neuron, 71(5), 2011.  [15]Dmitry Ulyanov, Andrea Vedaldi, and Victor S. Lempitsky. Instance normalization: The missing ingredient for fast stylization. CoRR, abs/1607.08022, 2016.  [16]Michael Defferrard, Kirell Benzi, Pierre Vandergheynst, and Xavier Bresson. Fma: A dataset for music analysis. In 18th International Society for Music Information Retrieval Conference, 2017.  [17]J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009.  [18]Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. CoRR, abs/1409.1556, 2014.  [19]Paul Adenot and Chris Rogers and Chris Wilson. Web Audio API. http://www.w3.org/TR/2015/WD-webaudio-20151208/, 2015. |
| 十二、指导教师意见  （学生工作的意义、项目的难度和工作量，本项目的特点以及学生对项目的理解与独创性等方面的评价）    签名：  年 月 日 |
| 十三、专家组评审意见      签名：  年 月 日 |
| 十四、学校意见  负责人（签名、公章）：  年 月 日 |