方法：基于图像的表示方法、用深度网络、建模时间音调关系

摘要：

1. 深度网络挑战：难以把现有知识用于训练同时不改变训练流程
2. 音乐转化为2D表示，生动展现不同音高间的关系
   1. 用于CNN，学习有效的抽象编码方式表现音高关系
   2. 用于LSTM，捕捉音乐序列中的时序依赖性
3. Tonnetz启发式表示，对比pianoroll常用表示方法

介绍：

1. 近期工作：
   1. 深度神经网络：Magenta，音乐生成（Google Brain team）
   2. Eck and Schmidhuber（2002）RNN with LSTM cells
      1. Improvisation（即兴创作）12小节蓝调
      2. 25键相互独立，每个键阈值0.5
   3. Boulanger-Lewandowski, Bengio,andVincent(2012)
      1. a pianoroll representation of 88 keys（全健表示方法）
   4. Huang, Duvenaud, and Gajos (2016)
      1. 用符号序列表示和弦序列
      2. 和弦嵌入采用Word2vec使用skip-gram模型
   5. Herremans and Chuan (2017)
      1. Word2vec生成复调音乐
   6. Herremans, Chuan, and Chew (2017)
      1. 音乐生成更完整综述
2. 1. 广泛使用的高准确率图形识别算法使数据可用性提高，端到端方法省去手动提取特征
   2. 音乐标注数据稀缺、但理论历史悠久，如何应用到深度学习需要探索
   3. 乐理不止分类
      1. 学习复杂声调系统（西方音乐）分析作曲
      2. 分层结构(Kirlin and Jensen 2015)
3. 本工程
   1. 目标：将乐理融合进神经网络的输入表示，以便学习特定音乐特征
   2. 结构：时序能力LSTM+复调音乐自动编码CNN（2d表示法）
      1. Tonnetz表示法用于表示复调音乐时间片（尚未应用过）
      2. 多层CNN提取声调空间内不同调的音乐性关系
   3. 数据集：MuseData(Boulanger-Lewandowski, Bengio, and Vincent 2012)
   4. 结果：不同网络配置和不同池化策略下自动编码和原始tonnetz矩阵的sigmoid交叉熵
   5. 对比：pianoroll表示法
4. 复调音乐中的Tonnetz矩阵
   1. 音乐学家和乐理学家用于研究音调和音调空间的图像化表示方法
      1. 提出1739
      2. 三元结构1926？
      3. 多种变形、表达更复杂的音乐关系
      4. parsimonious voice-leading (Cohn 1997)
      5. 为音乐信息检索(MIR)提供理论基础
         1. 和弦识别器(Harte,Sandler,andGasser2006)
         2. 结构可视化(Bergstrom, Karahalios, and Hart 2007)
   2. 本文用具体音调代替音调类
      1. 网格四边形化
      2. 半音最近距离拓展
      3. 遍布C0到C8#
   3. 每个beat是一个片段
      1. 片段中出现的在矩阵中是1
      2. 否则是0
5. 网络结构

两层CNN编码，一层全连接解码

3层LSTM将训练好的编码训练到下一个时间的编码

Sigmoid交叉熵

二、arbitrary transcription algorithm + advice of an expert trained on symbolic sequences

任意转录算法+符号序列上的专家建议

1. Restricted Boltzmann machines
   1. Gibbs采样？
   2. NADE（The neural autoregressive distribution estimator）
      1. 条件分布vj表示为vk的非线性函数(k < j)
2. RT RBM(循环时序RBM)
   1. 约束：隐藏单元必须描述条件分布并传达时序信息
3. RNN-RBM
   1. 去掉限制，加入h(t)⑧ 平均场值？？
4. 初始化方法