DOI 10.37882/2223–2966.2023.03.09

РАЗРАБОТКА КЛАССИФИКАТОРА ДЛЯ ОПТИЧЕСКОГО РАСПОЗНАВАНИЯ МУЗЫКАЛЬНОЙ НОТАЦИИ

DEVELOPMENT OF A CLASSIFIER FOR OPTICAL RECOGNITION OF MUSICAL NOTATION

N. Gubenko
S. Molodyakov
T. Kolikova

Summary. The representations that are used for two tasks related to sheet music are considered: identification of scores (musical notes) and obtaining the corresponding spectrograms and audio performances, taking into account the score as a search query. The scheme of the classifier is presented, the distinctive feature of which is the sequential use of several neural networks. The classifier is trained on a new data set of musical scores, which are collected from the DeepScores data set for segmentation, detection and classification of tiny objects, from the HOMUS data set — a pen-based musical notation and a data set of labeled GUITARPRO songs for sequence models. The combined data will become publicly available in the future. The results of extraction experiments using scans of real notes of high complexity are presented.

Keywords: recognition, musical notation, dataset, classifier, deep learning, spectrogram.

Губенко Надежда Олеговна
Магистрант, Санкт-Петербургский политехнический университет Петра Великого
gubenko.no@edu.spbstu.ru
Молодяков Сергей Александрович
Д.т.н., профессор, Санкт-Петербургский политехнический университет Петра Великого
molodyakov_sa@spbstu.ru
Коликова Татьяна Всеволодовна
Старший преподаватель, Санкт-Петербургский политехнический университет Петра Великого
kolikova_tv@spbstu.ru

Аннотация. Рассматриваются представления, которые используются для двух задач, связанных с нотами: идентификация партитур (нотных записей) и получение соответствующих спектрограмм и аудио выступлений с учетом партитуры в качестве поискового запроса. Представлена схема классификатора, отличительной особенностью которого является последовательное использование нескольких нейронных сетей. Классификатор обучается на новом наборе данных нотных партитур, которые собраны из набора данных DeepScores для сегментации, обнаружения и классификации крошечных объектов, из набора данных HOMUS — нотной записи на основе пера и набора данных маркированных песен GUITARPRO для моделей последовательности. Объединенные данные в дальнейшем станут общедоступными. Приводятся результаты экспериментов по извлечению, в которых используются сканы реальных нот высокой сложности.

Ключевые слова: распознавание, музыкальная нотация, набор данных, классификатор, глубокое обучение, спектрограмма.

Введение

Многие важные приложения в области поиска музыкальной информации — от сценариев поиска до записи в реальном времени — требуют согласования между различными представлениями произведения, чаще всего между печатной партитурой и записанным звуковым исполнением. Область распознавания и анализа нотных записей обычно называют оптическим распознаванием музыки (OPM) [1] .

Классификация музыкальных символов или нотации произведений — это подзадача системы OPM, в которой изолированным символам присваиваются метки классов. В представляемой работе разрабатывается универсальный классификатор музыкальных символов и партитур. Для обучения классификатора создается новый большой наборов данных, состоящий из нескольких доступных наборов данных. Создаваемый универсальный классификатор способен классифицировать музыкальные символы независимо от того, хорошо ли они напечатаны или просто написаны от руки, а также получить соответствующие спектрограммы для дальнейшего преобразования в звуковой формат MIDI.

1. Рассмотрение известных методов и систем распознавания музыкальной нотации

Традиционно автоматические методы связывания аудио и нот опирались на некоторые общие представления, которые позволяют сравнивать и сопоставлять моменты времени в аудио и позиции нот. Известны работы, которые демонстрируют примеры среднего уровня представления символических описаний событий, которые связаны с ошибками шагов автоматической

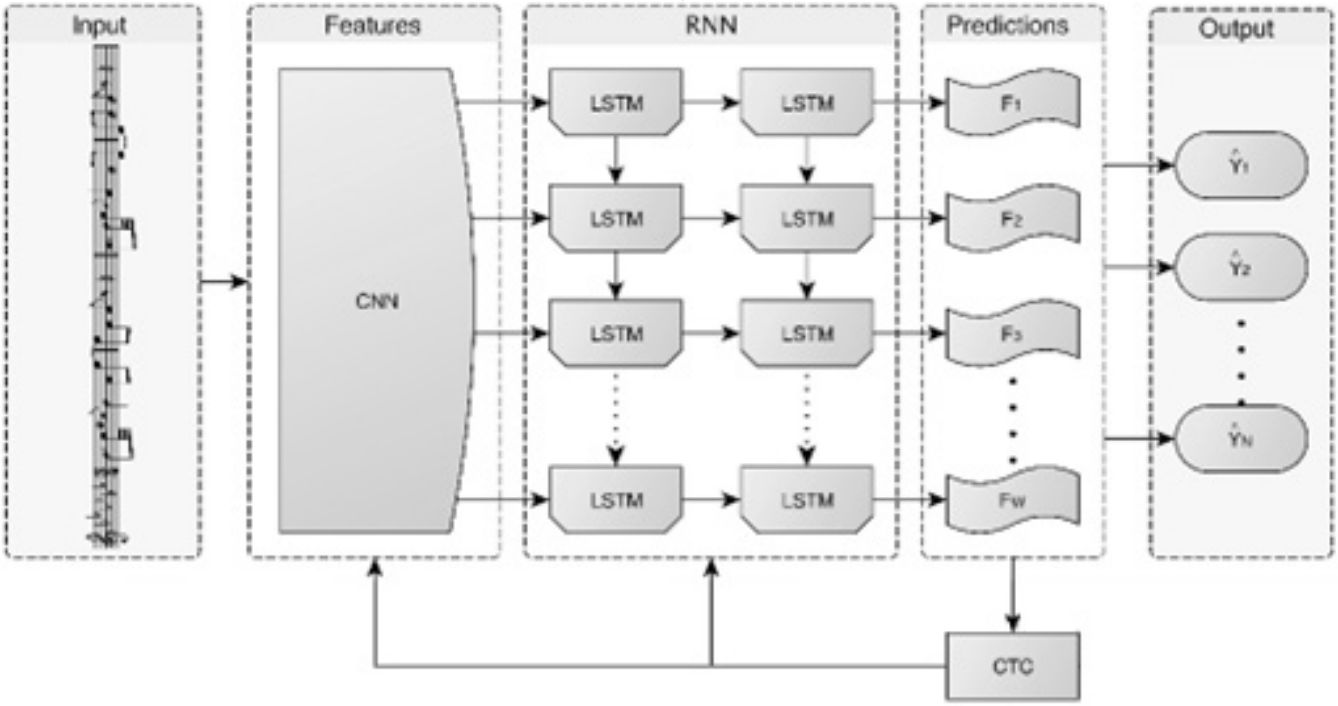
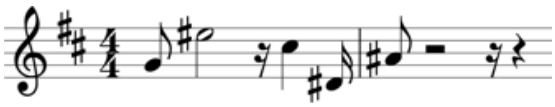


Рис. 1. Схема классификатора распознавания партитур



MRRN; RK+3R"QR<L+6

Рис. 2. Нотная партитура и соответствующие лэйблы в программе

записи музыки на аудио стороне [2, 3], или демонстрируют методы оптического распознавания музыки [4, 5]; или спектральные характеристики, которые избегают явного шага транскрипции звука [6] .

В статье [7] представлена система, основанная на OPM, которая использует аннотированные данные для контролируемого обучения. Этими данными является объединение нескольких наборов данных. Существуют системы, распознающие ноты с помощью глубокого обучения. В статье [8] рассматриваются алгоритмы, которые представляют собой конвейерные сверточные модели, имеющие комбинированные функции затрат для обнаружения и классификации объектов партитур. Внимание проблеме распознавания нотных партитур уделяется в работе [9]. В ней используются нейронная модель, сочетающая в себе возможности сверточных нейронных сетей, которые работают с входным изображением, и рекуррентных нейронных сетей, которые имеют дело с последовательным характером преобразования.

Известны работы, в которых предлагается обратное преобразование — из аудио формируются ноты и даже аккорды [10, 11]. В работе [11] для преобразования коротких отрывков аудио в нотные изображения предлагается использовать мультимодальные сверточные нейронные сети.

2. Объединение наборов данных

Для обучения моделей нейронных сетей предлагается объединять следующие общедоступные наборы данных:

- Набор данных рукописных онлайн-музыкальных символов (HOMUS) [6] содержит 15200 образцов изолированных музыкальных символов.
- Набор данных MUSCIMA++ [12] является крупнейшим доступным набором данных, который содержит 55000 полных символов для базового набора данных рукописных музыкальных партитур.

Таблица 1. Точность обучения набора музыкальных символов

Тип	Средняя точность
Ключ Соль 	0.98
Ключ Фа 	0.98
Диез 	0.96
Бемоль 	0.97
Бекар 	0.92
Тактовая черта |	0.89
Размер такта	0.83
Ноты	0.91
Паузы	0.90

- Набор данных печатных музыкальных символов DeepScores [8] — это высококачественный набор данных, состоящий из страниц записанной музыки. В нем 300000 полных страниц в виде изображений, содержащих десятки миллионов объектов.
- Набор данных песен с меткой GUITAR PRO для моделей последовательности — содержит около 12000 полных музыкальных произведений в виде изображений раздельных музыкальных партий

Результирующий набор данных содержит более 70000 рукописных и более 350000 печатных партитур с существенным количеством межклассовых различий. Набор состоит из 63 символов: музыкальных нот от C3 до B6, 4 длительности (половина, четверть, восьмая, шестнадцатая), 4 паузы тех же длительностей, символы размеров (3/4, 4/4, 6/8), знаки альтерации (диез, бемоль, бекар), скрипичный ключ, тактовая черта. Полные печатные страницы в дальнейшем можно разделить на отдельные сегменты в самом коде.

3. Создание универсального музыкального классификатора

Универсальный музыкальный классификатор должен быть способен распознавать большинство видов музыкальных символов, независимо от того, написаны они от руки или напечатаны. Глубокие нейронные сети, особенно сверточные и рекуррентные нейронные сети, предлагают удобный и мощный способ решения задач компьютерного зрения. Поэтому основная задача — построить такой классификатор путем обучения сверточной и рекуррентной нейронной сети на представленном расширенном наборе данных.

На рис. 1 представлена схема классификатора распознавания нотных партитур. На вход подаются изображения нот, на выходе формируются соответствующие нотам лэйблы и интервалы, которые конвертируются в MIDI-файлы с помощью библиотеки mido языка Python. Сверточная нейронную сеть (CNN) отвечает за обучение и обработку входного изображения. RNN отвечает за создание последовательности музыкальных символов. Далее идет предсказание и выходной результат в виде лэйблов (знаковых обозначений для каждого символа). Функция потери коннекционистской временной классификации (CTC) используется только для обучения. CTC выполняет локальную оптимизацию с использованием алгоритма максимизации ожиданий, чтобы с большой вероятностью выдавать правильную последовательность.

В процессе считывания и преобразования в цифровой массив изображение из этого набора данных подвергается различным искажениям и зашумлению, что требует эффективных средств его улучшения. Фильтрация — это устранение шумов в изображении, таких как невысокое качество отдельных элементов,

возникновение погрешностей сканирования, неверно выбранное разрешение, неоднородность контура, изолированные пустоты внутри объекта, разрывы, слияния. Для фильтрации шумов полутоновых изображений наиболее часто используются усредняющий, пороговый и медианный фильтры, которые включены на этапе обучения.

4. Вычисление спектрограммы

Каждый отдельный символ представления в нотации с Lilypond [12] сопоставляется с соответствующим лэйблом, например на рис. 2 представлен фрагмент партитуры и соответствующих его символам лейблы, где, например, скрипичный ключ (первый элемент) — это буква «M» лэйбла.

После того, как партитуры разделились и распознались, им присваиваются соответствующие лэйблы, по которым в дальнейшем выстраивается длительность, частота, тональность и другое для представления спектрограмм.

Вычисляются спектрограммы логарифмической частоты с частотой дискретизации 22,05 кГц и размером окна FFT 2048 выборок. Для уменьшения размерности применяются нормализованный логарифмический банк фильтров с 16 полосами на октаву, допускающими только частоты от 30 Гц до 6 кГц. Это приводит к 92 частотным ячейкам. Частота кадров спектрограммы составляет 20 кадров в секунду.

5. Результаты

Для оценки точности обучения использовались значения точности определения длительности символа и определение класса символа. При корректном определении всех значений тест засчитывался за позитивный образец (PS), в противном случае засчитывался за негативный образец (NS). Точность рассчитывается по формуле: Accuracy — PS / (PS + NS). Полученная точность обучения представлена в табл. 1.

6. Выводы

Предоставленный классификатор имеет достаточную для многих задач точность распознавания музыкальных нотаций. Созданный новый наборов данных не идеален и в настоящее время страдает от некоторой несбалансированности: некоторые классы имеют менее 10 экземпляров, в то время как другие имеют более 1000. Это создает проблему для любого классификатора, который оптимизирует точность этого набора данных, поскольку он может просто изучить базовое распределение, как было сделано в распознавании нотаций с помощью нейронной сети, также и просто игнорировать классы с наименьшим количеством выборок. Поэтому необходимо собрать больше образцов из классов с недостаточным количеством экземпляров.

ЛИТЕРАТУРА

1. Klapuri A., Virtanen T. Automatic music transcription //Handbook of Signal Processing in Acoustics, Springer, 2011,—443 pp.
2. D.W. Aha, D. Kelz, and M.K. Albert, Instance-based learning algorithms, Mach. Learn., 1991, pp. 37–66.
3. Sigtia Wang, Pao-Chi Chang, Jian-Jiun Ding Spectral-Temporal Receptive Field-Based Descriptors and Hierarchical Cascade Deep Belief Network for Guitar Playing Technique Classification // IEEE Trans Cybern 2022 May pp. 66
4. Levenshtein V.I. Binary codes capable of correcting deletions, insertions, and reversals // Tech. Rep. 8, 1966, pp. 373–380
5. Luwei Yang, Akira Maezawa, Jordan B.L. Smith, Elaine Chew Probabilistic transcription of sung melody using a pitch dynamic model Speech and Signal Processing // Processing — March 2017, P. 373–380.
6. Jorge Calvo-Zaragoza et al. Pen-based Musical Notation Recognition: HOMUS Dataset, 2017, P. 154.
7. Alexander Pacha Towards a Universal Music Symbol Classifier, 2017
8. Jorge Calvo-Zaragoza, David Rizo End-to-End Neural Optical Music Recognition of Monophonic Scores 2018, P. 44–48.
9. Lukas Tuggener, Ismail Elezi DeepScores — A Dataset for Segmentation, Detection and Classification of Tiny Objects, 2018, P. 56
10. Voinov N.V., Ivanov D.A., Leontieva T.V., Molodyakov S.A. Implementation and Analysis of Algorithms for Pitch Estimation in Musical Fragments // Proceedings of 2021 24th International Conference on Soft Computing and Measurements, 2021, P. 113–116, doi: 10.1109/SCM52931.2021.9507134
11. Matthias Dorfer et al. Studying the correspondences of audio notes for cross-modal search and identification of works, 2018, P. 59–65.
12. Brooks F.P., Hopkins A.L. Jr., P.G. Neumann P.G., Wright W.V. An experiment in musical composition. IRE Transactions on Electronic Computers // Proc. Conf. Data Sci. Technol. Appl. — 2018. — P. 373–380.