tekhnicheskoy konferentsii [Problems of efficiency and security of functioning complex technical and information systems : proceedings of the 32nd All-Russian scientific technical conference]. Serpukhov, Branch of the Military Academy of Strategic Missile Forces named after Peter The Great Publ., 2013, part 5, pp. 242–247.

21. Chistoprudov D. A., Kuzmin A. A. Primenenie apparata nechetkogo logucheskogo vyvoda v predskazyvauchikh *sistemakh* [The use of the apparatus of fuzzy inference in prediction systems]. *Sbornik trudov 7-y Mezhdunarodnoy nauchno-prakticheskoy konferentsii uchashchikhsya i studentov* [Proceedings of the 7th International Scientific and Practical Conference of Pupils and Students]. Protvino, Department of Education and Science Publ., 2014, part 2, pp. 87–90.

22. Chistoprudov D. A., Kovalkov D. A., Tebyakina M. A. Effektivnost primeneniya nechetkikh modelei po otnosheniu k parametricheskim v otdelnykh voprosakh iz radiotekhniki [Efficiency of application of fuzzy models in relation to parametric in separate questions from radio engineering]. *Informatsionnye tekhnologii v obrazovanii, nauke i proizvodstve : sbornik trudov 4-y Mezhdunarodnoy nauchno-prakticheskoy konferentsii* [Information Technologies in Education, Science and Production : Proceedings of the 4th International Scientific and Practical Conference]. Serpukhov, 2010, part 2, pp. 277–279.

23. Chistoprudov D. A., Kovalkov D. A., Ustinov I. A. *Primenenie neirosetevykh tekhnologiy dlya resheniya zadachi optimizatsii na setyakh svyazi* [Application of neural network technologies to solve the optimization problem on communication networks]. *Informatsionnye tekhnologii v proektirovanii i proizvodstve* [Information Technologies in Design and Production]. Moscow, Scientific and Technical Center of the Defense Complex “Compass”, 2008, no. 1, pp. 111–113.

24. Bellman R. E., Giertz M. On the analytical of fuzzy sets. *Sciences*, 1975, vol. 5, pp. 149–156.

25. Zadeh L. J. Fuzzy Sets. *Information and Control*, 1965, vol. 8, pp. 338–353.

УДК 004.81

ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ЗАДАЧИ ГЕНЕРАЦИИ МУЗЫКАЛЬНЫХ КОМПОЗИЦИЙ[1]

Статья поступила в редакцию 08.11.2018, в окончательном варианте – 17.11.2018.

Никитин Никита Андреевич, Волгоградский государственный технический университет, 400005, Российская Федерация, г. Волгоград, пр. им. Ленина, 28,
аспирант, e-mail: set.enter@mail.ru

Розалиев Владимир Леонидович, Волгоградский государственный технический университет, 400005, Российская Федерация, г. Волгоград, пр. им. Ленина, 28,
кандидат технических наук, доцент, ORCID 0000-0002-7372-8364, e-mail: vladimir.rozaliev@gmail.com

Орлова Юлия Александровна, Волгоградский государственный технический университет, 400005, Российская Федерация, г. Волгоград, пр. им. Ленина, 28,
кандидат педагогических наук, доктор технических наук, доцент, ORCID 0000-0003-4854-7462, e-mail: yulia.orlova@gmail.com

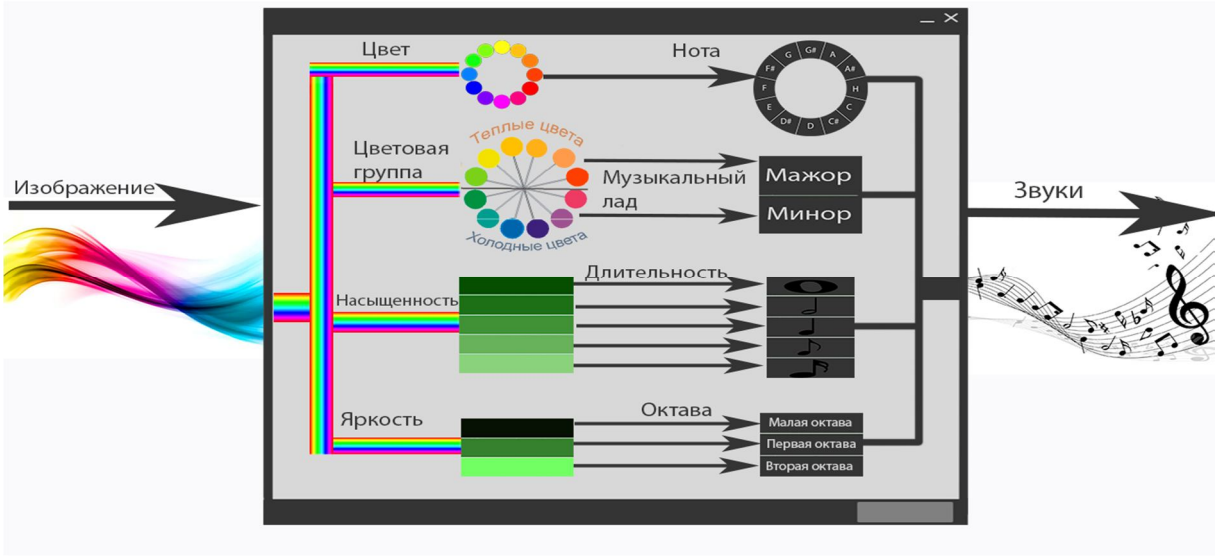
Заболеева-Зотова Алла Викторовна, Волгоградский государственный технический университет, 400005, Российская Федерация, г. Волгоград, пр. им. Ленина, 28,
доктор технических наук, профессор, ORCID 0000-0003-1918-667X, e-mail:zabzot@gmail.com

Цель исследования, описанного в данной статье, – увеличение гармоничности и мелодичности генерации звуков по изображению. Проведена разработка метода, реализующего комбинированный подход в генерации звуковых последовательностей. Предложенный метод использует рекуррентную нейронную сеть для генерации музыкального материала и цветомузыкальную теорию, которая используется для определения параметров композиции по изображению. Описана разработанная программа, предназначенная для генерации звуков по изображению. В ее основе лежит разработанный авторами метод, а также использование языков Python и библиотека Keras. Приведены результаты экспериментов, которые показывают высокую эффективность комплексного использования методов машинного обучения и светозвуковой теории для задачи генерации звуков по изображению.

Ключевые слова: рекуррентная нейронная сеть, цветомузыкальная теория, Keras, автоматизированная генерация музыки, схемы соотнесения цветов и нот, анализ изображений

[1] Работа выполнена при поддержке РФФИ (грант 17-07-01601, 17-29-07021, 18-07-00220).

Графическая аннотация (Graphical annotation)



THE APPLICATION OF MACHINE LEARNING METHODS FOR THE TASK OF GENERATING MUSICAL COMPOSITIONS

The article was received 08.11.2018, in the final version – 17.11.2018.

Nikitin Nikita A., Volgograd State Technical University, 28 Lenin Ave., Volgograd, 400005, Russian Federation,
post-graduate student, e-mail: set.enter@mail.ru
Rozaliev Vladimir L., Volgograd State Technical University, 28 Lenin Ave., Volgograd, 400005, Russian Federation,
Cand. Sci. (Engineering), Assistant Professor, ORCID 0000-0002-7372-8364, e-mail: vladimir.rozaliev@gmail.com
Orlova Yulia A., Volgograd State Technical University, 28 Lenin Ave., Volgograd, 400005, Russian Federation,
Cand. Sci. (Pedagogy), Doct. Sci. (Engineering), Assistant Professor, ORCID 0000-0003-4854-7462, e-mail: yulia.orlova@gmail.com
Zaboleeva-Zotova Alla V., Volgograd State Technical University, 28 Lenin Ave., Volgograd, 400005, Russian Federation,
Doct. Sci. (Engineering), Professor, ORCID 0000-0003-1918-667X, e-mail: zabzot@gmail.com

The purpose of the study described in this article is to increase the harmony and melody of the sound generation based on images. Development of methods that implement a combined approach to the generation of sound sequences. The developed method uses a recurrent neural network to generate musical material and a color music theory, which is used to determine the parameters of a composition by an image. Describes the developed program for generating sounds by the image, based on the developed method, as well as the Python language and the Keras library. The results of experiments that show the high efficiency of the integrated use of methods of machine learning and color music theory for the problem of generating sounds by image are presented.

Keywords: recurrent neural network, color music theory, Keras, automated music generation, color and note matching schemes, image analysis

Введение. Несмотря на все достижения в понимании творческих процессов, создание музыки не может проходить автоматически. Роль пользователя-композитора очень высока, и можно лишь говорить об автоматизации этого процесса. Передаваемая музыкой и картинами эмоциональность сложно распознаваема [2]. Сам процесс создания музыки на данный момент не поддаётся чёткой формализации, хотя и основывается на чётко определённых музыкальных правилах. Наибольших успехов автоматизация процесса написания и создания музыки достигла сравнительно недавно (в последние десятилетия). Однако по большей части связана с изучением и повторением различных музыкальных стилей [5].

В рамках данной работы предполагается автоматизация процесса создания музыки путём автоматизированной генерации звуков по изображению. Другими словами, генерация звуков по изображению есть процесс преобразования изображения в одну или несколько последовательностей нот, с определённым основным тоном и длительностью [12].

Гипотеза исследования заключается в том, что использование изображений для получения характеристик композиции позволит снизить роль пользователя. Кроме того, использование нейронных сетей позволит повысить качество результирующей музыкальной композиции.

Целью работы является увеличение гармоничности и мелодичности генерации звуков по изображению.

Для достижения поставленной цели были поставлены следующие задачи:

- проанализировать программы для генерации звуковых последовательностей;
- найти и изучить методы и подходы автоматизированной генерации звуковых последовательностей;
- изучить методы синтеза звуков и анализа изображений;
- разработать модули анализа изображения, генерации звуковых последовательностей и синтеза звуков;
- протестировать разработанную программу на предмет мелодичности и гармоничности выходной музыкальной композиции (с привлечение музыкальных экспертов).

Метод соотнесения цветовых и музыкальных характеристик. Для снижения роли пользователя-композитора в генерации звуков, часть характеристик музыкального произведения получается путём анализа цветовой гаммы изображения. Таким образом, характер полученной музыкальной композиции будет соответствовать входному изображению. Данная особенность делает возможным применение указанного подхода для создания фоновых музыкальных произведений в компьютерных играх, рекламе и фильмах.

Ключевыми характеристиками музыкального произведения является его тональность и темп. Именно эти параметры определяются путём анализа цветовой гаммы изображения. Для начала определим соотношение цветовых и музыкальных характеристик [4] (табл. 1).

Таблица 1 – Соотношение цветовых и музыкальных характеристик

Цветовые характеристики	Музыкальные характеристики
Оттенок (красный, синий, жёлтый…)	Нота (до, до-диез, ре, ре-диез, ми, ми, фа, фа-диез, соль, соль-диез, ля, ля-диез, си)
Цветовая группа (тёплый/холодный)	Музыкальный лад (мажор/минор)
Яркость	Октава ноты
Насыщенность	Длительность ноты

Согласно цветомузыкальной теории В.В. Афанасьева, невозможно навсегда привязать какую-либо ноту к определенному цвету в силу их различной природы. Поэтому необходимо связать отношения звуков и цветов в зависимости от того, в какой плоскости они представлены: мелодия, гармония или тональность. Иначе говоря, одна и та же нота в произведении может быть окрашена разными цветами [1].

1. Анализируя в целом изображение, определяем преимущественный цвет и последовательность цветов с характеристиками. Далее, согласно преимущественному цвету и одной из пяти схем соотнесения цветов и нот, определяется тональность будущего произведения.

Максимальное родство цветов определяется между двумя соседними хроматическими цветами, а степень тонального родства – между звуками, находящимися на расстоянии семи полутонов друг от друга. Это отражается в квинтовом круге тональностей.

Таким образом, совмещая хроматический цветовой круг и квинтовый круг тональностей, получим необходимую последовательность для модуляций.

Например, пусть в качестве преимущественного цвета был выбран красный. Согласно схеме Ньютона, это соответствует ноте до (рис. 1).

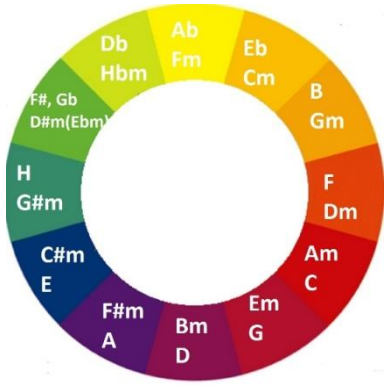


Рисунок 1 – Наложение цветового круга и квинтового круга тональностей

2. Затем, сопоставляя хроматическую гамму нашей тоники с цветовым кругом, получим соответствие нот и цветов для мелодической части будущего произведения (рис. 2).

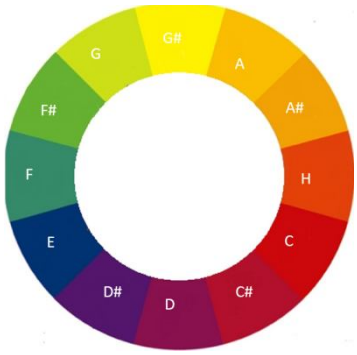


Рисунок 2 – Наложение хроматической гаммы и цветового круга

3. После этого необходимо определить соответствие цветов и ступеней для построения гармоний произведения.

В музыке существуют функциональные отношения (тоника – субдоминанта – доминанта), а в цветоведении этому соответствует принцип дополнительности цветов (красный – синий – желтый). То есть главными цветами при принятом основном (например, красном) являются дополнительные – желтый и синий; главными трезвучиями в музыке – тоническое, субдоминантовое и доминантовое (рис. 3).

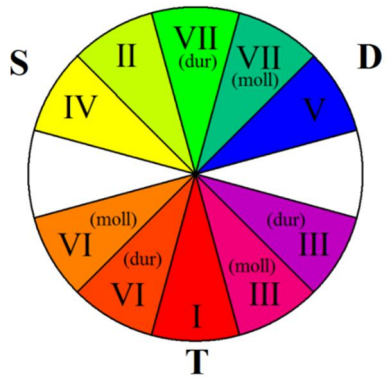


Рисунок 3 – Гармоническое соответствие

В предыдущих пунктах (с 1-го по 3-й) получили тональность будущего произведения, определили необходимую последовательность для модуляций, выявили соответствие цветов и нот применительно для мелодического сопровождения, а также получили гармонию.

4. Затем, согласно определённым ранее соответствиям цветов и нот для мелодического сопровождения, необходимо преобразовать полученный набор цветов в набор нот.

Выбор нейронной сети для генерации музыкальных композиций. Важной особенностью нейронных сетей прямого распространения (feedforward neural networks) является то, что у них есть общее ограничение: входные и выходные данные имеют фиксированный, заранее обозначенный размер. Например – картинка 100×100 пикселей или последовательность из 256 бит. Нейросеть с математической точки зрения ведет себя как обычная функция, хоть и очень сложно устроенная: у нее есть заранее обозначенное число аргументов, а также обозначенный формат, в котором она выдает ответ.

Вышеперечисленные особенности не представляют больших трудностей, если речь идет о тех же картинках или заранее определенных последовательностях символов. Однако для обработки любой условно бесконечной последовательности, в которой важно не только содержание, но и порядок, в котором следует информация (например, текст или музыка) необходимо использовать нейронные сети с обратными связями – рекуррентные нейронные сети (RNN). В рекуррентных нейросетях нейроны обмениваются информацией между собой: например, вдобавок к новому «кусочку» входящих данных нейрон также получает некоторую информацию о предыдущем состоянии сети. Таким образом, в сети реализуется «память», что принципиально меняет характер ее работы и позволяет анализировать любые последовательности данных, в которых важно, в каком порядке идут значения [11].

Однако большой сложностью сетей RNN является проблема исчезающего (или взрывного) градиента – она заключается в быстрой потере информации с течением времени. Конечно, это влияет лишь на веса для нейронов, а не на их состояния. Однако ведь именно в весах нейронов накапливается информация. Сети с долгой краткосрочной памятью (long short term memory, LSTM) стараются решить вышеупо-

мянутую проблему потери информации, используя фильтры и явно заданную клетку памяти. У каждого нейрона есть клетка памяти и три фильтра: входной, выходной и забывающий. Целью этих фильтров является защита информации. Входной фильтр определяет, сколько информации из предыдущего слоя будет храниться в клетке. Выходной фильтр определяет, сколько информации получат следующие слои. Такие сети способны научиться создавать сложные структуры, например, сочинять тексты в стиле определённого автора или сочинять простую музыку. Однако при этом потребляется большое количество ресурсов [6].

Таким образом, для реализации программы автоматизированной генерации музыкальных композиций необходимо использовать именно рекуррентные нейронные сети с долгой краткосрочной памятью – RNN LSTM (долгая краткосрочная память – разновидность архитектуры рекуррентных нейронных сетей). Именно данный вид нейронных сетей используется для генерации музыкальных композиций в программе Magenta – это музыкальный проект с открытым исходным кодом от Google; RNN LSTM используется в программе сочинения композиций в стиле И.С. Баха – BachBot; DeepJaz – система позволяет генерировать джазовые композиции на основе анализа midi-файлов [3].

Описание используемой искусственной нейронной сети. Рекуррентная нейронная сеть (RNN) имеет циклические или повторяющиеся соединения, которые позволяют сети хранить информацию по входам. Эти связи можно считать похожими на память. RNN особенно полезны для изучения последовательных данных, таких как музыка.

В TensorFlow повторяющиеся соединения на графе разворачиваются в эквивалентную нейронную сеть прямого распространения (feedforward neural network) (рис. 4). Затем эту сеть обучают с использованием техники градиентного спуска, называемой back propagation through time (BPT).

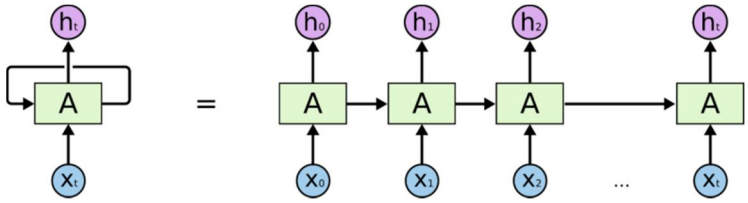


Рисунок 4 – Циклическое соединение RNN, развёрнутое во времени

Есть большое количество способов, с помощью которых RNN может соединяться с собой с циклическими соединениями. Наиболее распространенными являются сети с долгой краткосрочной памятью (long short term memory, LSTM) и управляемые рекуррентные нейроны (gated recurrent units, GRU). В обоих случаях сети имеют мультипликативные нейроны, которые защищают их внутреннюю память от перезаписи, позволяя нейронным сетям обрабатывать более длинные последовательности. В данной работе предполагается использование LSTM [9].

Все рекуррентные нейронные сети имеют форму цепочки повторяющихся модулей нейронной сети. В стандартных RNN этот повторяющийся модуль будет иметь очень простую структуру, например, один слой tanh (рис. 5) [10].

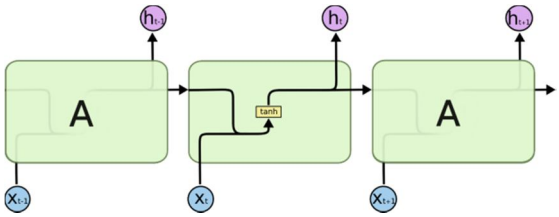


Рисунок 5 – Повторяющийся модуль в стандартной однослойной рекуррентной нейронной сети

LSTM также имеют эту цепочку, но повторяющийся модуль имеет более сложную структуру. Вместо того чтобы иметь один слой нейронной сети, существует четыре, взаимодействующих между собой особым образом (рис. 6) [10].

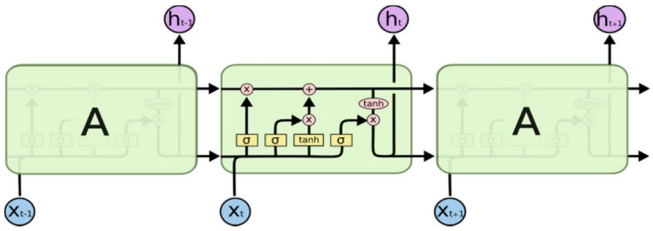
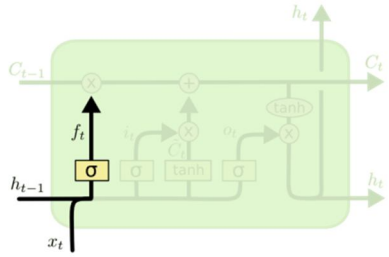


Рисунок 6 – Повторяющийся модуль в стандартной однослойной рекуррентной нейронной сети

На приведенной выше диаграмме каждая строка несет целый вектор, начиная с выхода одного узла и входа другого. Розовые круги представляют поточечные операции, такие как сложение векторов, в то время как в желтых ячейках представлены уровни нейронной сети. Соединение линий означает объединение, в то время как разделение линии обозначает, что ее содержимое копируется в различные входы.

Первым шагом в LSTM является решение, какую информацию мы собираемся выбросить из состояния ячейки. Это решение принимается сигмоидным слоем (sigmoid layer). Данный слой «смотрит» на значение выхода ht-1 и входа xt, рассчитывает значение в диапазон от 0 до 1 для каждого состояния ячейки Ct-1. Если слой вернул значение «1», это означает, что данное значение необходимо оставить (запомнить), если «0» – удалить из состояния ячейки (рис. 7). Например, в состоянии ячейки могут храниться характеристики текущего такта – если такт ещё не закончен, в таком случае необходимо оставить характеристики в памяти. Если идёт работа уже с новым тактом, то необходимо запомнить новые параметры.

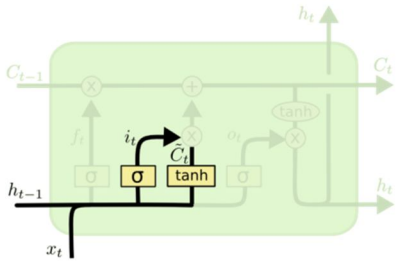


$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$

Рисунок 7 – Использование сигмоидного слоя для удаления информации из состояния

Следующим шагом является принятие решения о том, какую новую информацию мы собираемся хранить в состоянии ячейки. Для этого, во-первых, сигмоидный слой принимает решение, какие значения мы будем обновлять. Далее слой tanh создает вектор новых значений кандидата Ct, которые могут быть добавлены в состояние.

Следующим шагом является обновление старого состояния ячейки Ct-1 в новом состоянии ячейки Ct. Для этого необходимо умножить старое состояние ft – таким образом произведём удаление информации из состояния. Затем необходимо сложить полученное значение и it * Ct. Таким образом, получим новые значения кандидатов, масштабируемые значением коэффициента обновления каждого значения состояния (рис. 9).



$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$

Рисунок 9 – Обновление старого и нового состояния ячейки

На последнем шаге необходимо решить, что будет выводить данный слой. Этот вывод будет основан на состоянии ячейки. Сначала пропускаем входное значение через сигмоидный слой, который решает, какие части состояния ячейки необходимо вывести. Затем обрабатываем состояние ячейки с использованием tanh (чтобы сдвинуть значения между «–1» и «1») и умножаем его на выход сигмоидного слоя.

Поведение нейронной сети определяется набором весов и смещений, которые имеет каждый из узлов. Поэтому для корректной работы нейронной сети необходимо настроить их на некоторое правильное значение.

Во-первых, необходимо определить, насколько хорошим или плохим является какой-либо выход согласно входному значению. Это значение называется стоимостью. Как только стоимость получена, необходимо использовать метод обратного распространения ошибки (back propagation). По сути, он сводится к вычислению градиента стоимости по отношению к весам (т.е. производной стоимости по каждому весу для каждого узла в каждом слое). Затем необходимо использовать метод оптимизации для корректировки весов с целью снижения стоимости. В данной работе мы будем использовать метод градиентного спуска.

Для обучения нейронной сети предполагается подавать на вход вектор, который содержит следующие части [8]:

- название ноты: MIDI-обозначение текущей ноты. Используется для представления высоты ноты;
- время начала включения ноты;

- время выключения ноты;
- сила (громкость) воспроизведения ноты (velocity).

Для определения правильного выхода согласно входу предполагается преобразовать вектор следующим образом: пусть имеется вектор нот {c, d, e, f, g, a, h}, тогда обучающий вектор будет {{c, d}, {d, e}, {e, f}, {f, g}, {g, a}, {a, h}}. Такой способ обучения нейронной сети используется, например, для прогнозирования временных рядов [7].

Метод получения композиции по изображению. В данной работе предполагается следующий алгоритм получения композиции по изображению:

1) согласно методу соотнесения цветовых и музыкальных характеристик получаем тональность произведения и последовательность первых 20 % нот, считанных с изображения;

2) далее по полученной последовательности нот предсказываем продолжение произведения с помощью обученной модели и нейронной сети;

3) по итоговой последовательности нот и тональности, согласно методу соотнесения цветовых и музыкальных характеристик, строим гармоническую часть произведения.

Описание разработанной программы для генерации звуков по изображению. Для подтверждения эффективности предложенных алгоритмов и методов была разработана программа для генерации звуков по цветовой гамме изображений. Данная программа представляет собой веб-сайт, реализованный на языке Python. На вход программа получает изображение, которое пользователь загружает вручную. После получения пути к изображению, программа загружает изображение в память, используя библиотеку OpenCV. Затем происходит конвертация изображения в цветовое пространство HSV. Затем, в процессе анализа изображения, программа определяет преимущественный цвет изображения – по нему определяется тональность и темп композиции. После происходит предсказание (достраивание) композиции согласно полученным характеристикам с помощью нейронной сети.

Архитектура программы представлена на рисунке 10.

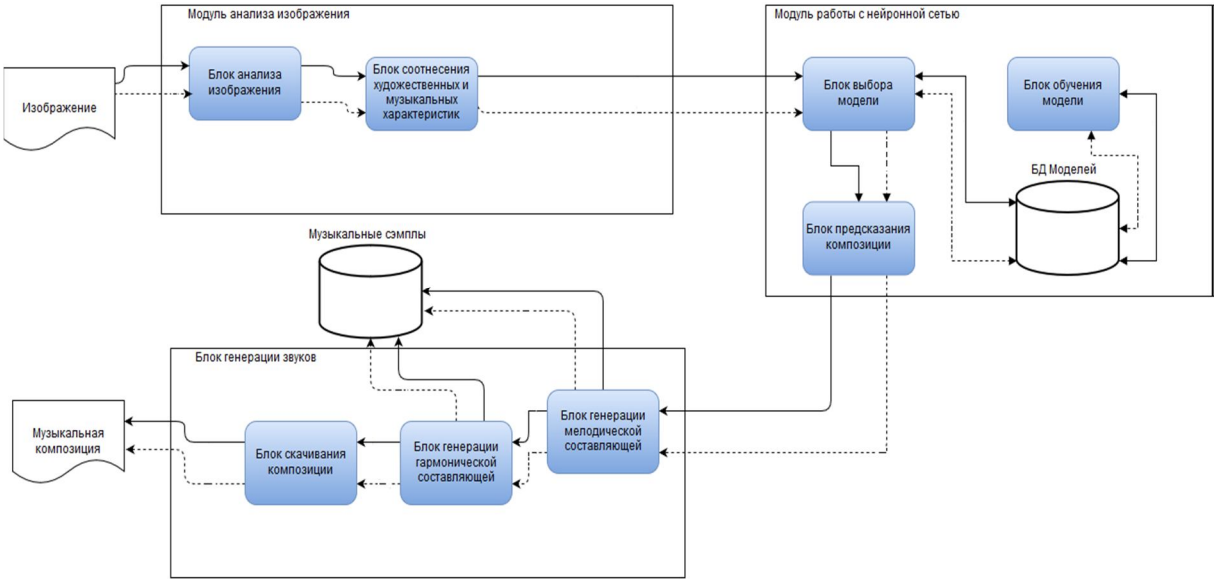


Рисунок 10 – Архитектура программы

Проведение экспериментов. Для подтверждения эффективности предложенных методов были проведены три эксперимента:

- эксперимент по анализу зависимости качества генерируемых звуков от размера обучающей выборки;
- эксперимент по анализу качества генерируемых звуков;
- эксперимент по анализу зависимости качества генерируемых звуков с использованием нейронных сетей от подхода без использования нейронных сетей.

Для оценки композиций была использована экспертная группа, состоящая из 10 независимых специалистов-экспертов, имеющих среднее или высшее музыкальное образование, которые оценивали композиции по следующим критериям:

- соответствие характеру изображения (по пятибалльной шкале);
- реалистичность звучания инструмента (фортепьяно или гитара);
- мелодичность композиции;

- качество гармонии (аккомпанемента);
- приятность мелодии для восприятия;
- цельность композиции;
- реалистичность/искусственность композиции.

Эксперимент по анализу зависимости качества генерируемых звуков от размера обучающей выборки. Входные данные: обучающие выборки различного размера (120 классических произведений, 500, 1000, 2000, 4295). Выходные данные: зависимость качества генерируемы звуков от размера обучающей выборки.

Для проведения данного эксперимента была взята обучающая выборка, состоящая из 4295 музыкальных произведений в формате .midi. Были обучены 5 моделей на различном размере выборки. Для первой модели были взяты 120 классических произведений, для второй – 500, для третьей – 1000 классических произведений. Для четвёртой и пятой модели были взяты 2000 и 4295 (максимальное количество музыкальных произведений в формате .midi, которое доступно для обучения) произведений различного жанра.

По всем моделям были сгенерированы 3 композиции и отправлены экспертам для анализа.

Результаты проведения эксперимента представлены в таблице 2.

Таблица 2 – Средние оценки экспертов для эксперимента по анализу зависимости качества генерируемых звуков от размера обучающей выборки

Размер выборки	**X1**	**X2**	**X3**	**X4**	**X5**	**X6**	**X7**
120	3,5	4	3	5	3	3	3
500	4	4	3,5	5	3,5	3,5	3
1000	4	4	4	5	4	4	3,5
2000	5	4	4	5	5	4,5	4
4295	4	4	4	5	4	4	4

Примечание. X**1** – соответствие характеру изображения (по пятибалльной шкале); X**2** –реалистичность звучания инструмента (фортепьяно или гитара); **X3** – мелодичность композиции; **X4** – качество гармонии (аккомпанемента); **X5** – приятность мелодии для восприятия; **X6** – цельность композиции; **X7** –реалистичность/искусственность композиции.

Средние оценки для данного эксперимента приведены в таблице 3.

Таблица 3 – Средние оценки по всем критериям для эксперимента по анализу зависимости качества генерируемых звуков от размера обучающей выборки

Размер выборки	Средняя оценка для всех критериев
120	3,5
500	3,7
1000	4
2000	4,5

Таким образом, можно сделать вывод, что наилучшей моделью с точки зрения качества генерируемой музыкальной композиции является модель, обученная на 2000 классических произведениях.

Из таблицы видно, что последняя модель, обученная на 4295 композиций, генерирует произведения хуже, чем предыдущая модель. Это связано с тем, что для обучения последней модели были взяты произведения различного жанра. В этом случае на выходе модель – произведение, состоящее из смеси различных жанров. Такие произведения были оценены экспертами хуже, чем в случае использования для обучения только классических музыкальных произведений.

Эксперимент по анализу качества генерируемых звуков. Для тестирования были привлечены 10 человек, имеющих музыкальное образование для того, чтобы они смогли оценить работу программы без временных затрат на установку программного обеспечения, был составлен набор из 10 папок, имеющих название Тест № 1, Тест № 2 и так далее до Тест № 10. В каждой папке находилось исходное изображение и сгенерированная композиция в .mp3 формате. Данный набор тестов был отправлен 10 экспертам в виде архива по электронной почте.

Средние оценки экспертов по каждому из критериев по пятибалльной шкале представлены в таблице 4.

Проанализировав оценки всех экспертов и высчитав средние по каждому критерию, можно сделать вывод о том, что фортепьяно на слух экспертов звучит реалистичнее, чем гитара. Также можно сделать вывод о том, что зачастую композиция, сгенерированная по абстрактным изображениям, более приятна на слух, чем генерация по пейзажам. Однако тест № 10 является городским пейзажем и при этом получил довольно неплохие экспертные оценки.

Таблица 4 – Средние оценки экспертов для эксперимента по анализу качества генерируемых звуков

	X1	**X2**	**X3**	**X4**	**X5**	**X6**	**X7**
Тест № 1	5	4	4	5	4	4	4
Тест № 2	5	4	5	5	5	5	5
Тест № 3	5	4	5	4	5	4	4
Тест № 4	5	4	5	5	5	5	5
Тест № 5	5	4	5	5	5	5	4
Тест № 6	5	3	4	5	5	5	5
Тест № 7	5	4	4	5	4	4	4
Тест № 8	4	4	4	5	4	4	4
Тест № 9	5	4	3	5	3,5	3,5	3
Тест № 10	5	4	5	5	5	5	5

Примечание. **X1** – соответствие характеру изображения (по пятибалльной шкале); **X2** – реалистичность звучания инструмента (фортепьяно или гитара); **X3** – мелодичность композиции; **X4** – качество гармонии (аккомпанемента); **X5** – приятность мелодии для восприятия; **X6** – цельность композиции; **X7** – реалистичность/искусственность композиции.

В целом общее впечатление от сгенерированных звуков у экспертов положительное. Среди минусов некоторые эксперты выделяют однотипность гармонии, иногда рваность и недостаточную реалистичность произведения; недостаточную реалистичность звучания гитары.

Делая вывод по каждому критерию, можно сказать, что все эксперты оценили на высоком уровне соответствие произведения характеру изображения (кроме 8-го теста). По второму критерию – инструмент фортепьяно звучит довольно реалистично. При оценке мелодичности композиций мнения экспертов разделились пополам, т.е. половину композиций эксперты оценили на высший бал, другую половину – на «4» (в целом – неплохой результат). Качество гармонии также было оценено экспертами на высший бал (кроме 3-го теста). Приятность мелодий для восприятия получила 60 % высших балов и 40 % «четвёрок». Это говорит о том, что некоторые произведения (тесты 7, 8, 9) звучат не вполне реалистично. Реалистичность и цельность композиций в среднем были оценены на «4», что является естественным результатом для компьютерной генерации звуков.

Для окончательного подведения итогов тестирования можно высчитать средние значения для каждого критерия по всем тестам. Результаты такого расчёта представлены в таблице 5.

Таблица 5 – Средние значения критериев по всем тестам для эксперимента по анализу качества генерируемых звуков

Критерий	**Среднее значение для всех тестов**
Соответствие характеру изображения	4,9
Реалистичность звучания инструмента	3,9
Мелодичность композиции	4,4
Качество гармонии	4,9
Приятность для восприятия	4,6
Цельность композиции	4,5
Реалистичность композиции	4,3

Таким образом, все сгенерированные произведения получили в отношении отдельных характеристик оценки выше «4» (кроме реалистичности звучания инструмента, что связано с некоторой искусственностью звучания гитары).

Эксперимент по анализу зависимости качества генерируемых звуков с использованием нейронных сетей от подхода, без использования нейронных сетей. Входные данные: модель, обученная на 2000 классических произведений. Выходные данные: зависимости качества генерируемых звуков с использованием нейронных сетей от подхода, без использования нейронных сетей.

В данном эксперименте были сгенерированы 5 музыкальных композиций с использованием искусственных нейронных сетей, и 5 – без их использования. Эти композиции были отправлены специалистам-экспертам для анализа. Эксперты оценивали каждую пару произведений с помощью одной из трёх оценок:

- 1 – произведение, созданное без использования нейронных сетей, лучше, чем с их использованием;
- 0 – произведение, полученное без использования нейронных сетей, похоже (по качеству) на произведение с их использованием;
- 1 – произведение, полученное с использованием нейронных сетей, лучше, чем без их использования.

Результаты оценки экспертами пар произведений представлены в таблице 6.

Таблица 6 – Результаты проведения эксперимента по анализу зависимости качества генерируемых звуков с использованием нейронных сетей от подхода, без использования нейронных сетей

	X1	**X2**	**X3**	**X4**	**X5**	**X6**	**X7**
Тест № 1	0	0	1	1	1	1	1
Тест № 2	0	0	1	1	1	1	1
Тест № 3	0	0	1	1	1	1	1
Тест № 4	0	0	1	1	1	1	0
Тест № 5	0	0	0	1	1	0	1

Примечание. **X1** – соответствие характеру изображения (по пятибалльной шкале); **X2** – реалистичность звучания инструмента (фортепьяно или гитара); **X3** – мелодичность композиции; **X4** – качество гармонии (аккомпанемента); **X5** – приятность мелодии для восприятия; X6 – цельность композиции; **X7** – реалистичность/искусственность композиции.

Для исследования результатов эксперимента по анализу зависимости качества генерируемых звуков с использованием нейронных сетей от подхода, без использования нейронных сетей, были просчитаны средние оценки для каждого критерия по всем тестам (табл. 7).

Таблица 7 – Средние оценки для каждого критерия по всем тестам для эксперимента по анализу зависимости качества генерируемых звуков с использованием нейронных сетей от подхода, без использования нейронных сетей

Критерий	Среднее значение для всех тестов
Соответствие характеру изображения	0
Реалистичность звучания инструмента	0
Мелодичность композиции	0,8
Качество гармонии	1
Приятность для восприятия	1
Цельность композиции	0,8
Реалистичность композиции	0,8

Таким образом, из таблицы 7 видно, что критерии «Соответствие характеру изображения» и «Реалистичность звучания инструмента» не изменились при изменении способа генерации звуков (с использованием искусственных нейронных сетей и без них). Это связано с тем, что параметр «Соответствие характеру изображения» определяется выбранной тональностью, которая не зависит от использования нейронных сетей. А параметр «Реалистичность звучания инструмента» зависит от модуля синтеза звуков, который тоже не был «затронут» при изменении подхода генерации композиций.

Наиболее важные параметры для оценки достижения цели исследования – «Мелодичность композиции», «Качество гармонии» и «Приятность для восприятия» – были отмечены экспертами как улучшившиеся. Это связано с тем, что без использования нейронных сетей был использован «наивный» подход к генерации композиций. В то же время нейронные сети позволяют выявить такие связи, которые не видны человеку – это сказывается на качестве генерируемых музыкальных композиций.

Параметры «Цельность композиции» и «Реалистичность композиции» немного улучшились при использовании нейронных сетей. Это связано с тем, что данные критерии зависят от предыдущих трёх критериев. Поскольку они улучшились, то использование нейронных сетей положительно повлияло и на данные критерии.

Заключение. В результате выполнения данной работы был предложен комбинированный подход к генерации звуковых последовательностей. Он использует рекуррентную нейронную сеть для генерации музыкального материала и цветомузыкальную теорию для определения параметров композиции по изображению. В процессе выбора нейронной сети для генерации музыкальных композиций было выявлено, что для реализации программы автоматизированной генерации музыкальных композиций необходимо использовать именно рекуррентные нейронные сети с долгой краткосрочной памятью – RNN LSTM (долгая краткосрочная память – разновидность архитектуры рекуррентных нейронных сетей).

В ходе проведения валидации моделей было выполнено три эксперимента:

1) по анализу зависимости качества генерируемых звуков от размера обучающей выборки;
2) по оценке качества генерируемых звуков;
3) по анализу зависимости качества генерируемых звуков с использованием нейронных сетей от подхода, без использования нейронных сетей.

По результатам первого эксперимента можно сделать вывод о том, что оптимальной моделью следует считать модель, обученную на 2000 классических произведениях, так как именно в этом случае были получены наивысшие экспертные оценки.

По результатам второго эксперимента можно сделать вывод о том, что в целом общее впечатление от сгенерированных звуков у экспертов было положительное. Среди минусов некоторые эксперты выделяют однотипность гармонии, иногда рваность и недостаточную реалистичность произведения; недостаточную реалистичность звучания гитары.

По результатам проведения третьего эксперимента можно сделать вывод о том, что использование нейронных сетей оправдано для задачи генерации музыкальных композиций по цветовой гамме изображений, поскольку для наиболее важных критериев, отвечающих за гармоничность и мелодичность звуков, было отмечено улучшение качества выходной музыкальной композиции.

Для оценки достижения цели исследования были выдвинуты следующие критерии:

- средняя оценка качества генерируемых звуков по изображению выше «4»;
- средняя оценка для эксперимента № 3 больше «0,5» (т.е. в процессе округления даст «1»).

Для проверки достижения «первого критерия» рассчитаем среднюю оценку по всем критериям для всех тестов. Для этого необходимо сложить все средние оценки из таблицы 5 и поделить на общее количество критериев. Таким образом, получим среднюю оценку «4,5». Поскольку «4,5» больше «4», то цель согласно первому критерию достигнута.

Для проверки достижения «второго критерия» необходимо рассчитать среднюю оценку по всем критериям, приведенным в таблице 7. Средняя оценка – «0,6», что больше «0,5». Следовательно, цель была достигнута и по второму критерию.

Список литературы

1. Афанасьев В. В. Светозвуковой музыкальный строй. Элементарная теория аудиовизуальных стимулов / В. В. Афанасьев. – Москва : Музыка, 2002. – 70 с.

2. Розалиев В. Л. Methods and Models for Identifying Human Emotions by Recognition Gestures and Motion / В. Л. Розалиев, А. В. Заболеева-Зотова // The 2013 2nd International Symposium on Computer, Communication, Control and Automation 3CA 2013, December 1–2, 2013, Singapore. – Amsterdam – Beijing – Paris : Atlantis Press, 2013. – P. 67–71.

3. Brinkkemper F. Analyzing Six Deep Learning Tools for Music Generation. – 2015. – Режим доступа: http://www.asimovinstitute.org/analyzing-deep-learning-tools-music/, свободный. – Заглавие с экрана. – Яз. англ. (дата обращения 03.07.2017).

4. Caivano J. L. Colour and sound: Physical and Psychophysical Relations / J. L. Caivano // Colour Research and Application. – 1994. – № 12 (2). – P. 126–132.

5. Cope D. Computer Models of Musical Creativity / D. Cope. – Cambridge, MA : MIT Press, 2005.

6. Doornbusch P. Gerhard Nierhaus: Algorithmic Composition: Paradigms of Automated Music Generation / P. Doornbusch // Computer Music Journal. – 2014. – Vol. 34, iss. 3.

7. Fernández J. D. AI Methods in Algorithmic Composition: A Comprehensive Survey / J. D. Fernández, F. Vico // Journal of Artificial Intelligence Research. – Spain, Málaga, 2013. – № 48. – P. 513–582.

8. Kim H. K. Transactions on Engineering Technologies: Special Issue of the World Congress on Engineering and Computer Science / Haeng Kon Kim, Sio-Iong Ao, A. Mahyar. – New York : Springer Publishing Company, 2013. – P. 796.

9. Mazurowski L. Computer models for algorithmic music composition / L. Mazurowski // Proceedings of the Federated Conference on Computer Science and Information Systems. – Poland : Szczecin, 2012. – P. 733–737.

10. Mikolov T. Recurrent neural network based language model / T. Mikolov, M. Karafiat, L. Burget, J. Cernocky, S. Khudanpur // Proceedings of INTERSPEECH International Speech Communication Association. – 2010. – № 9. – P. 1045–1048.

11. Sak H. Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition / H. Sak, A. Senior, F. Beaufays // ArXiv e-prints. – 2014.

12. Wu X. A study on image-based music generation / X. Wu. – Burnaby : Simon Fraser University, 2008.

References

1. Afanasiev V. V. *Svetozvukovon muzykalnyy stroy. Elementarnaya teoriya audiovizualnykh stimulov* [Light and sound musical system. Elementary theory of audiovisual incentives]. Moscow, Muzika Publ., 2002. 70 p.

2. Rozaliev V. L., Zaboleeva-Zotova, A. V. Methods and Models for Identifying Human Emotions by Recognition Gestures and Motion. *The 2013 2nd International Symposium on Computer, Communication, Control and Automation 3CA 2013, December 1–2, 2013, Singapore*. Amsterdam – Beijing – Paris, Atlantis Press, 2013, pp. 67–71.

3. Brinkkemper F. Analyzing Six Deep Learning Tools for Music Generation. 2015. Available at: http://www.asimovinstitute.org/analyzing-deep-learning-tools-music/ (accessed 03.07.2017).

4. Caivano J. L. Colour and sound: Physical and Psychophysical Relations. *Colour Research and Application*, 1994, no. 12 (2), pp. 126–132.

5. Cope D. *Computer Models of Musical Creativity*. Cambridge, MA, MIT Press, 2005.

6. Doornbusch P. Gerhard Nierhaus: Algorithmic Composition: Paradigms of Automated Music Generation. *Computer Music Journal*, 2014, vol. 34, iss. 3.

7. Fernández J. D., Vico F. AI Methods in Algorithmic Composition: A Comprehensive Survey. *Journal of Artificial Intelligence Research*, Spain, Málaga, 2013, no. 48, pp. 513–582.

8. Haeng Kon Kim, Sio-Iong Ao, Mahyar A. *Transactions on Engineering Technologies: Special Issue of the World Congress on Engineering and Computer Science.* New York, Springer Publishing Company, 2013, p. 796

9. Mazurowski L. Computer models for algorithmic music composition. *Proceedings of the Federated Conference on Computer Science and Information Systems*. Poland, Szczecin, 2012, p. 733–737.

10. Mikolov T. Recurrent neural network based language model. *Proceedings of INTERSPEECH International Speech Communication Association*, 2010, no. 9, pp. 1045–1048.

11. Sak H., Senior A., Beaufays F. Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition. *ArXiv e-prints*, 2014.

12. Xiaoying Wu. *A study on image-based music generation*. Burnaby, Simon Fraser University, 2008.

УДК 004.932

БЫСТРОДЕЙСТВУЮЩИЙ АЛГОРИТМ ОПТИМИЗАЦИИ ТОНОВОЙ АППРОКСИМАЦИИ МОНОХРОМНЫХ РАСТРОВЫХ ИЗОБРАЖЕНИЙ

Статья поступила в редакцию 07.11.2018, в окончательном варианте – 18.11.2018.

Агаджанян Альберт Грантович, Донской государственный технический университет, 344002, Российская Федерация, г. Ростов-на-Дону, пл. Гагарина, 1,

аспирант, e-mail: foralbert92@gmail.com

Обоснована целесообразность использования тоновой аппроксимации монохромных мультитоновых изображениий. Процедура тоновой аппроксимации заключается в понижении размерности цветовой (тоновой) палитры изображения и реализуется заменой оригинальных тонов на аппроксимирующие. Главной проблемой рассматриваемой задачи является подбор конкретных аппроксимирующих тонов, т.е. нахождение оптимальной структуры аппроксимирующей палитры. С целью оптимизации тоновой аппроксимации монохромных растровых изображений в ранее опубликованных коллективных работах автора разработан и экспериментально исследован гибридный алгоритм. Его особенностью является двухэтапная обработка изображения. Первый этап подразумевает использование модифицированного эволюционно-генетического алгоритма с целью сокращения области поиска оптимальной аппроксимирующей палитры. Второй этап гибридного алгоритма предполагает реализацию простого, но эффективного алгоритма перебора ближайшей окрестности субоптимального решения, которое, как предполагается, было найдено на первом этапе. Это позволяет гарантированно обеспечить экстремальную аппроксимирующую палитру согласно выбранному критерию оптимизации. Настоящая статья посвящена сравнению гибридного алгоритма с авторскими модификациями известного алгоритма кластеризация *k*-средних, который является популярным решением рассматриваемой задачи. Экспериментальное исследование продемонстрировало однозначное преимущество гибридного алгоритма в сравнении с модифицированным алгоритмом *k*-средних в отношении временных затрат и точности решения. Также в рамках исследования рассматривается новый метод, предложенный автором, который предполагает использование координатно сжатой матрицы изображения для ускорения вычислительного процесса. Анализ полученных результатов продемонстрировал высокую эффективность данного подхода применительно к решению задачи тоновой аппроксимации изображений.

Ключевые слова: монохромные изображения, оптимизация тоновой аппроксимации, гибридный алгоритм, алгоритм *k*-средних, модификация метода, эволюционно-генетический алгоритм, квантование

Графическая аннотация (Graphical annotation)