

УДК 004.932

ПОРІВНЯННЯ КЛАСИФІКАЦІЇ ЗОБРАЖЕНЬ ХРОМАГРАМ І СПЕКТРОГРАМ ДЛЯ РОЗПІЗНАВАННЯ АКОРДІВ

Пономаренко А.О.

Науковий керівник – к.т.н., доц. Турута О.П.

Харківський національний університет радіоелектроніки, каф. ПІ
м. Харків, Україна

тел.: +38(068) 567-88-08, e-mail: artur.ponomarenko@nure.ua

This work is devoted to a comparison of a usage of a chroma and a spectrogram for a chord recognition. Image classification is one of fields of neural networks usage. After a model is learned by annotated images it can be used to recognize a class of another one. This approach is applicable also for sounds. And the problem is how to represent a sound as an image. There are different ways. The choice depends on the domain. In the chord recognition a tone frequency is the most important part. There are two representation of the sound tones – spectrogram and chroma. The spectrogram shows the dependence of the signal power spectral density on time. A chromogram is a spectrogram, but the signal is divided into 12 logarithmically distributed parts that corresponds to 12 notes.

Класифікація зображень за допомогою нейронної мережі є одним з найбільш поширених завдань у галузі обробки зображень та машинного навчання. Для розв'язання цього завдання, використовують нейронні мережі, які навчаються розпізнавати певні ознаки та шаблони на зображеннях. Цей самий підхід може бути використаний для аудіо даних, так як аудіо сигнал може бути перетворений у зображення. Для цього існують різні способи: waveform, спектрограма і хромограма, та інші. Вибір залежить від конкретного завдання.

В рамках цієї роботи була поставлена задача розпізнавання акордів. Тому для цього важливо знати частоту тонів в аудіозаписі. Для цього підходять спектрограма і хромограма. Спектрограма – це графік залежності частоти від часу. По осі Y вказана частота, по осі X час. Хромограма – це також спектрограма, але сигнал поділений логарифмічно на 12 частин, що відповідає 12 нотам європейської музики. Тобто технічно хромограма має менше деталей. У межах цього завдання було порівняно, яке з представлень аудіо є більш ефективним для розпізнавання акордів.

В першу чергу було проведено збір даних для научного дослідження. А саме було записано аудіовиконання акордів на гітарі. Запис виконувався за допомогою звичайного диктофону на мобільному телефоні. Було записано акорди різних всі тонів (за виключенням дієзних) і різних видів (мінор, мажор, септакорди та інші). Після дані були опрацьовані, щоб на кожному окремому примірнику аудіо звучав лише один окремо зіграний

акорд, та очищені від сміттєвих даних, яких важко уникнути під час запису.

Після того, отримані дані було доповнено. А саме, для скорочення часу запису, було створено пропущені дієзні тони акордів шляхом програмного підвищення тону оригінальних записів. Таким чином для навчання нейронної мережі додалися класи тонів акордів: C# (до дієз), D# (ре дієз), F# (фа дієз), G# (соль дієз) та A# (ля дієз). І вже після цього аудіо було перетворено на зображення: спектрограми і хромограми. На рисунку 1 наведено приклад спектрограми (зліва) і хромограми (справа) для акорду ля мажор.

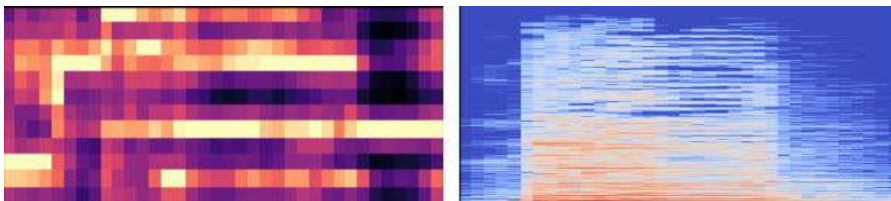


Рисунок 1 – Хромаграма (зліва) та спектрограма (справа) акорд ля мажор

Код для навчання моделі був написаний на python за допомогою фреймворку tensorflow. Для навчання була створена базова keras модель. Кожний з типів акордів мав по 9 зображень, з яких 7 було використано для навчання і 2 – для валідації. Навчання складається з 10 епох. Кожне зображення перемасштабовується і переводиться в чорнобіле для полегшення процесу навчання.

Для порівняння подивимось точність і втрати навчання та валідації протягом 10 епох навчання. На рисунку 2 зображені результати.

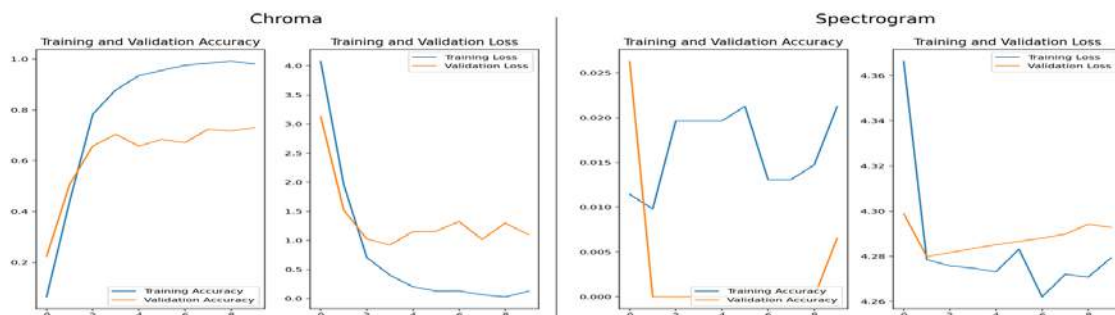


Рисунок 2 – Показники accuracy та loss навчання і валідації моделі для хромаграми (зліва) і спектрограми (справа)

Як ми можемо побачити результати навчання по зображенням хромаграми дуже точні – якість зростає стабільно кожної епохи і на останніх точність майже близиться до 100%. В той самий час, графік для спектрограми дуже нестабільний, і точність не переважає 2%. З цього можемо зробити висновок, що хромаграма значно краще підходить для вирішення поставленої задачі розпізнавання акордів.