Лекція 2. КЛАСИФІКАЦІЯ ЗВУКУ

Класифікація звуку охоплює різні аспекти звукових хвиль та може бути розглянута з кількох перспектив, залежно від наукової або практичної області. Основні категорії класифікації звуку:

1. За частотою (сприйняття людським вухом)

- **Інфразвук** — частоти нижче 20 Гц. Людина не здатна сприймати ці звуки, але деякі тварини (наприклад, слони) їх чують.
- **Чутний звук** — діапазон від 20 Гц до 20 кГц. Це частоти, які може сприймати людське вухо.
- **Ультразвук** — частоти понад 20 кГц. Людина не чує ці звуки, але вони використовуються в медичних приладах, сонарі та для інших технічних цілей.

2. За природою джерела

- **Природні звуки** — звуки, що створюються природними процесами, такими як вітер, дощ, грім, шум океану, спів птахів.
- **Штучні звуки** — звуки, створені людиною або технічними пристроями, наприклад, музичні інструменти, двигуни, розмови.

3. За тембром

Тембр звуку визначає його унікальне звучання та дозволяє відрізнити один інструмент від іншого навіть при однаковій висоті. Тембр залежить від гармонічних складових та обертонів, які супроводжують основний тон.

4. За гучністю (амплітудою)

- **Тихі звуки** — мають низьку інтенсивність або амплітуду, наприклад, шепіт, шелест листя.
- **Гучні звуки** — мають велику інтенсивність, такі як звук двигуна, музичні концерти або грім.

5. За тривалістю

- **Короткотривалі** — миттєві або короткі звуки, як клацання або хлопок.
- **Довготривалі** — звуки з тривалим існуванням, як шум двигуна або постійний гул.

6. За музичними характеристиками

- **Музичні звуки** — звуки з визначеною висотою, які можна використовувати для створення музики, наприклад, ноти на музичних інструментах.
- **Шумові звуки** — мають невпорядковану частотну структуру та використовуються у створенні шуму (наприклад, аплодисменти, шурхіт).

7. За фізичною природою звукової хвилі

- **Поздовжні хвилі** — звукові хвилі, де коливання частинок середовища відбуваються в напрямку поширення хвилі (характерні для газів, рідин і твердих тіл).
- **Поперечні хвилі** — коливання частинок відбуваються перпендикулярно до напрямку поширення хвилі (можуть виникати лише в твердих тілах).

8. За акустичним середовищем

- **Акустично прозорі середовища** — середовища, в яких звук легко проходить, наприклад, повітря або вода.
- **Акустично непрозорі середовища** — середовища, що поглинають або відбивають звук, наприклад, щільні тверді матеріали.

9. За напрямком поширення

- **Монофонічний** — звук поширюється з одного джерела в одне місце.
- **Стерео** — звук поширюється з декількох джерел, створюючи ефект об'ємного звучання.

Ця класифікація допомагає краще зрозуміти природу та характеристики звуку, як у технічних, так і в музичних або природних контекстах.

Алгоритми класифікації звуку використовуються для аналізу та обробки аудіосигналів з метою їх категоризації за певними ознаками.

Ці алгоритми мають широкий спектр застосувань, зокрема в розпізнаванні мовлення, виявленні шуму, музичній індексації, біометрії та інших областях.

Основні алгоритми класифікації звуку включають методи машинного навчання, нейронні мережі, а також класичні статистичні методи.

Основні типи алгоритмів класифікації та їх особливості:.

1. Машинне навчання (ML)

Ці алгоритми використовують статистичні моделі для аналізу ознак звуку, таких як спектральні, тембральні, або часові характеристики, з метою їх класифікації.

1.1 Алгоритм K найближчих сусідів (KNN)

- **Принцип**: Для нового звукового зразка шукаються найближчі сусіди в навчальній вибірці за певними ознаками (наприклад, MFCC — коефіцієнти мел-частотного кепстра). Клас визначається більшістю серед сусідів.
- **Особливості**: Простий у реалізації та добре працює для невеликих обсягів даних, але неефективний для великих вибірок, оскільки вимагає багато обчислювальних ресурсів для пошуку сусідів.

1.2 Логістична регресія

- **Принцип**: Логістична регресія використовує імовірнісну модель для прогнозування ймовірності того, що даний звук належить до певної категорії.
- **Особливості**: Підходить для бінарних класифікаційних завдань, але може бути розширена для мультикласових завдань.

Простий і швидкий алгоритм, однак може бути менш точним для складних звукових сигналів.

1.3 Support Vector Machine (SVM)

- **Принцип**: Алгоритм будує гіперплощину, що максимально розділяє зразки різних класів на основі їх ознак (таких як спектрограми).
- **Особливості**: Ефективний для класифікації звуків, особливо при обмеженій кількості даних, однак може вимагати багато часу для навчання на великих вибірках.

2. Нейронні мережі (Deep Learning)

Нейронні мережі, особливо глибокі архітектури, широко застосовуються для класифікації звуку завдяки їх здатності виявляти складні патерни та особливості в аудіоданих.

2.1 Рекурентні нейронні мережі (RNN)

- **Принцип**: RNN використовують зворотні зв’язки для обробки послідовностей даних, що дозволяє їм враховувати часові залежності в звуці. Популярні варіації, такі як LSTM (Long Short-Term Memory), особливо ефективні для обробки довготривалих залежностей.
- **Особливості**: Дуже ефективні для класифікації аудіо з часовою залежністю (наприклад, розпізнавання мовлення). Вимагають великих обчислювальних ресурсів і великих наборів даних для навчання.

2.2 Згорткові нейронні мережі (CNN)

- **Принцип**: CNN обробляють двовимірні дані, такі як спектрограми звукових сигналів, для виявлення просторових залежностей між частотними компонентами звуку.
- **Особливості**: Чудово підходять для обробки акустичних ознак і звукових образів, часто використовуються для класифікації звуків у таких завданнях, як розпізнавання музичних жанрів або видів шуму.

2.3 Гібридні моделі CNN+RNN

- **Принцип**: Поєднують переваги CNN для обробки просторових ознак (спектрограми) з RNN для аналізу часових залежностей.
- **Особливості**: Ефективні в багатьох задачах класифікації звуку, оскільки дозволяють одночасно враховувати як частотну, так і часову структуру сигналу.

3. Функції ознак (Feature Engineering)

Важливою частиною класифікації звуку є вилучення ознак, які потім подаються на вхід класифікатору.

Найпопулярніші з методів вилучення ознак.

3.1 MFCC (Mel-Frequency Cepstral Coefficients)

- **Принцип**: Коефіцієнти мел-частотного кепстрал аналізують частотні характеристики звуку відповідно до людського слуху.
- **Особливості**: Один з найпоширеніших методів вилучення ознак для мовлення та музики. Використовується в багатьох алгоритмах класифікації звуку.

3.2 Chroma Features

- **Принцип**: Хрома-фічі представляють силу звуку в 12 тональних класах (відповідних нотах) і використовуються для аналізу гармонії музичних композицій.
- **Особливості**: Часто застосовуються в музичних додатках для класифікації жанрів або інструментів.

3.3 Zero-Crossing Rate

- **Принцип**: Визначає кількість разів, коли сигнал перетинає осьову лінію (нульове значення амплітуди).
- **Особливості**: Використовується для класифікації шумових або негармонічних звуків.

4. Традиційні статистичні методи

До появи машинного навчання та глибокого навчання, звукові сигнали класифікувалися з використанням статистичних методів, які і зараз можуть застосовуватися в простіших випадках.

4.1 Лінійний дискримінантний аналіз (LDA)

- **Принцип**: Цей метод шукає лінійні комбінації ознак, що найкраще відділяють класи.
- **Особливості**: Ефективний для задач із невеликою кількістю ознак і класів, але може бути недостатньо гнучким для складних сигналів.

5. Баєсові моделі

- **Принцип**: Баєсові моделі (наприклад, наївний баєсовий класифікатор) базуються на теоремі Байєса, яка дозволяє визначити ймовірність належності звуку до певного класу.
- **Особливості**: Прості для реалізації, добре працюють для початкових задач класифікації з обмеженими обсягами даних.

Алгоритми класифікації звуку використовують різні підходи залежно від складності задачі, типу звукових даних та обсягу доступних навчальних даних.

Нейронні мережі, особливо CNN і RNN, є найефективнішими для класифікації складних звуків, тоді як традиційні методи машинного навчання можуть бути корисними для більш простих задач.

Алгоритм класифікації на базі згорткових мереж

Алгоритм класифікації звуку на основі згорткових нейронних мереж (CNN) є потужним інструментом для аналізу та обробки аудіоданих.

Згорткові мережі відомі своєю здатністю ефективно розпізнавати просторові закономірності в даних (у випадку аудіо — в спектрограмах).

1. Попередня обробка звуку

Перед тим, як використовувати CNN, аудіодані потребують перетворення в таку форму, яка дозволить нейронній мережі працювати з ними.

1.1 Оцифровка звуку

Звук зазвичай представлений у вигляді дискретного сигналу — набору відліків амплітуди у часі. Його можна записати у вигляді вектора значень амплітуд за часовими точками (одновимірний сигнал).

1.2 Витяг ознак

- Для застосування CNN звук перетворюють у **двовимірну форму** (2D-дані), з якими згорткові мережі працюють найефективніше. Це можна зробити за допомогою таких методів:

  - **Спектрограма**: являє собою зображення, де одна вісь відповідає часу, інша — частоті, а кольорова шкала відображає амплітуду частот у певний момент часу.

  - **Мел-спектрограма** або **MFCC** (Mel-Frequency Cepstral Coefficients) — це перетворена спектрограма з використанням мел-шкали, яка більше відповідає людському сприйняттю звуку.

2. Архітектура CNN для класифікації звуку

2.1 Вхідний шар

Вхідні дані мають розмірність, наприклад, M×N, де M — кількість частотних смуг, а N — кількість часових фреймів (у випадку спектрограми або MFCC).

Вхідний шар передає ці дані в згорткові шари.

2.2 Згорткові шари (Convolutional Layers)

У цих шарах виконується операція згортки над вхідними даними з використанням фільтрів (ядра згортки). Кожен фільтр аналізує певну локальну частину вхідного сигналу, знаходячи характерні ознаки (патерни) у часових і частотних залежностях.

- **Згортка (Convolution)**: Для кожного фільтра відбувається ковзання ядра по входу і обчислюється згортка — результат накладення ядра на кожну ділянку сигналу.
- **Активація (ReLU)**: Зазвичай після згортки застосовується функція активації ReLU (Rectified Linear Unit), яка залишає тільки додатні значення, що прискорює навчання і робить модель більш стабільною.

2.3 Шар підвибірки (Pooling Layer)

Після згорткових шарів застосовують шар підвибірки, зазвичай **максимальне підвибіркування (Max Pooling)**.

Цей шар зменшує розмірність даних (знижує роздільну здатність), залишаючи найбільш важливі ознаки.

Це допомагає зменшити обчислювальні витрати та уникнути перенавчання.

Наприклад, шар Max Pooling з вікном 2x2 вибирає максимальне значення з кожної області розміром 2x2 і зменшує розмір вхідного зображення удвічі.

2.4 Повторення згорткових і підвибіркових шарів

Мережа зазвичай має кілька шарів згортки і підвибірки, що дозволяє поступово витягувати більш складні і абстрактні ознаки із звукового сигналу. Нижчі шари знаходять простіші закономірності, такі як зміни частоти або амплітуди, а вищі — більш складні зв'язки.

3. Шари повного з'єднання (Fully Connected Layers)

Після кількох згорткових і підвибіркових шарів дані передаються на шар повного з'єднання. Цей шар функціонує як звичайна нейронна мережа, яка приймає на вхід згорнуті та зменшені ознаки і виконує класифікацію.

Мета цього шару — обробити витягнуті ознаки і визначити, до якого класу належить звук.

4. Фінальний шар класифікації

Фінальний шар зазвичай використовує **Softmax** або **Sigmoid** функції для визначення ймовірності належності зразка до кожного класу.

- **Softmax** використовується для мультикласової класифікації (коли звук може належати до одного з кількох класів).
- **Sigmoid** використовується для бінарної класифікації.

5. Процес навчання

Мережу навчають на великому наборі даних звуків із відомими класами, використовуючи **метод зворотного поширення помилки (backpropagation)** і оптимізатори (наприклад, **Adam**, **SGD**).

Під час навчання мережа налаштовує свої ваги (параметри згорткових фільтрів і шарів повного з'єднання), щоб мінімізувати різницю між прогнозованими та справжніми класами.

6. Тестування та оцінка

Після навчання модель тестують на нових зразках, щоб оцінити її точність, використовуючи метрики, такі як **точність (accuracy)**, **повнота (recall)**, **F1-score** тощо.

Особливості CNN для класифікації звуку:

1. **Автоматичне витягування ознак**: На відміну від класичних методів, де потрібне ручне вилучення ознак (наприклад, MFCC), CNN автоматично виявляє важливі характеристики аудіосигналу.

2. **Ефективність для великих наборів даних**: CNN добре працюють з великими і складними даними, що робить їх ідеальними для класифікації аудіо в реальних додатках, як-от розпізнавання мовлення або музичних жанрів.

3. **Стійкість до шуму**: Згорткові мережі стійкі до шуму і здатні ігнорувати незначні зміни у вхідних даних.

4. **Велика кількість параметрів**: CNN можуть мати велику кількість параметрів, тому для їхнього навчання потрібні великі обсяги даних і потужні обчислювальні ресурси (GPU).

Алгоритм класифікації звуку на базі згорткових мереж полягає в перетворенні звукових сигналів у двовимірні форми (спектрограми) та виявленні ключових патернів через багато рівнів згорткових та підвибіркових шарів.

Це дозволяє мережі ідентифікувати і класифікувати звуки на основі складних ознак, що забезпечує високу точність у різних завданнях.

ДАТАСЕТ URBAN SOUND 8K

Urban Sound 8K — це широко використовуваний датасет для класифікації та розпізнавання звуків міського середовища. Він складається з коротких аудіозаписів різних типів міських шумів, зібраних з різних джерел, і є популярним вибором для досліджень у галузі звукової класифікації.

Основні характеристики датасету:

1. **Кількість аудіофайлів**: Датасет містить **8732 аудіозаписи** у форматі WAV. Тривалість кожного запису варіюється від кількох секунд до максимальної тривалості у 4 секунди.

2. **Класи звуків**: Записи в Urban Sound 8K розподілені на **10 класів** звуків, що часто зустрічаються в міському середовищі:

- Автомобільні гудки (car horn)
- Собачий гавкіт (dog bark)
- Дрель (drilling)
- Двигуни (engine idling)
- Постріли (gun shot)
- Дитячі ігри на майданчику (children playing)
- Сирени (siren)
- Повітряні кондиціонери (air conditioner)
- Автобуси (street music)
- Сміттєві баки (jackhammer)

3. **Аудіоформат**:

- Формат файлів: WAV
- Частота дискретизації: 44.1 кГц
- Канали: моно

4. **Анотації**: Кожен аудіофайл має анотації, що включають:

- Назву класу, до якого належить записаний звук.
- Метадані, такі як тривалість запису, файлова структура, а також геолокаційна інформація (місця, де були зроблені записи, якщо це можливо).

5. **Структура датасету**: Датасет розділений на **10 папок**, кожна з яких містить близько 1000 файлів. Така структура дозволяє легко використовувати його для крос-валідації, де кожна з папок може слугувати як навчальна або тестова вибірка.

Посилання:

https://www.kaggle.com/datasets/chrisfilo/urbansound8k

https://www.kaggle.com/datasets/smaildurcan/urban-sound-8k-image-png-dataset

6. **Застосування**: Urban Sound 8K призначений для навчання моделей машинного навчання з метою автоматичної класифікації міських звуків. Основні застосування:

- Розпізнавання шумів у міських умовах (наприклад, сирени, гудки).
- Розпізнавання звукових подій у реальному часі.
- Розробка алгоритмів для смарт-міст або систем моніторингу звуку.

7. **Джерела звуків**: Звуки були взяті з відкритих баз даних і є реальними записами міського середовища.

Приклад використання:

Urban Sound 8K широко використовується в задачах звукової класифікації, де моделі (як-от згорткові нейронні мережі) навчаються класифікувати звуки міста.

Зазвичай для роботи з цим датасетом звукові сигнали перетворюють на спектрограми або використовують коефіцієнти мел-частотного кепстра (MFCC) як ознаки для моделей машинного навчання.

Чому Urban Sound 8K важливий:

- **Великий набір даних**: Велика кількість аудіофайлів дає можливість створювати точні та узагальнені моделі.
- **Різноманітність звуків**: Звуки, включені до датасету, покривають широкий спектр міських шумів, що робить його корисним для задач реального часу.
- **Зручність використання**: Датасет структурований таким чином, що дозволяє легко проводити експерименти з крос-валідацією і навчанням моделей.

Датасет Urban Sound 8K доступний для завантаження з відкритих джерел, що робить його чудовим ресурсом для дослідників та інженерів у сфері аудіоаналітики.

Дані в датасеті **Urban Sound 8K** представлені у вигляді аудіофайлів та метаданих, що описують ці файли.

1. Аудіофайли

- Формат: **WAV**
- Частота дискретизації: **44.1 кГц**
- Канали: **моно** (одноканальний запис)
- Тривалість аудіозаписів: від кількох мілісекунд до **4 секунд**.

Кожен аудіофайл містить запис звукової події з міського середовища, який належить до одного з 10 класів. Ці файли організовані у папках для зручності.

2. Метадані

Метадані представлені у вигляді **CSV-файлу** з інформацією про кожен аудіофайл. Вони включають такі поля:

- **slice_file_name**: Назва аудіофайлу.
- **fsID**: Ідентифікатор користувача, який завантажив файл (джерело).
- **start**: Початкова точка аудіо (секунди), з якої було вирізано звуковий фрагмент.
- **end**: Кінцева точка аудіо (секунди).
- **salience**: Важливість звукової події (вказує на ступінь фокусування на звуці — від менш важливих до домінуючих звуків).
- **fold**: Номер папки (від 1 до 10), у якій зберігається файл. Це поле використовується для крос-валідації.

- **classID**: Ідентифікатор класу звукової події (від 0 до 9).
- **class**: Назва класу (наприклад, "dog bark", "car horn").

Ці метадані дозволяють дослідникам та розробникам легко аналізувати дані, здійснювати крос-валідацію та розподіляти вибірки для навчання, тестування та валідації моделей.

ДАТАСЕТ URBAN-SOUND-8K-IMAGE

Датасет **Urban-Sound-8K-Image** — це модифікація оригінального датасету **Urban Sound 8K**, де аудіофайли були перетворені у візуальні форми, зазвичай у вигляді **спектрограм** або **мел-спектрограм**.

Ця модифікація призначена для використання з методами комп'ютерного зору, зокрема зі згортковими нейронними мережами (CNN), які ефективно працюють з двовимірними зображеннями.

Основні характеристики:

1. **Формат даних**:

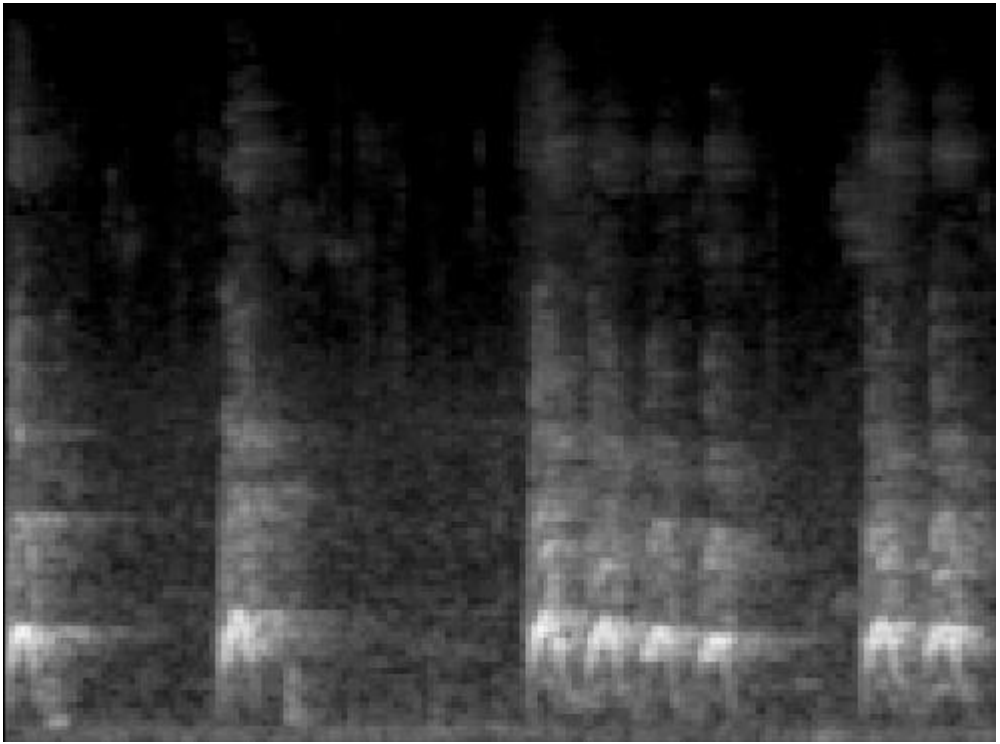
- Кожен аудіозапис у форматі WAV був перетворений у зображення.
- Тип зображень: спектрограми або мел-спектрограми (частотні зображення), які візуалізують звукові хвилі.
- Формат зображень: зазвичай **PNG** або **JPG**.

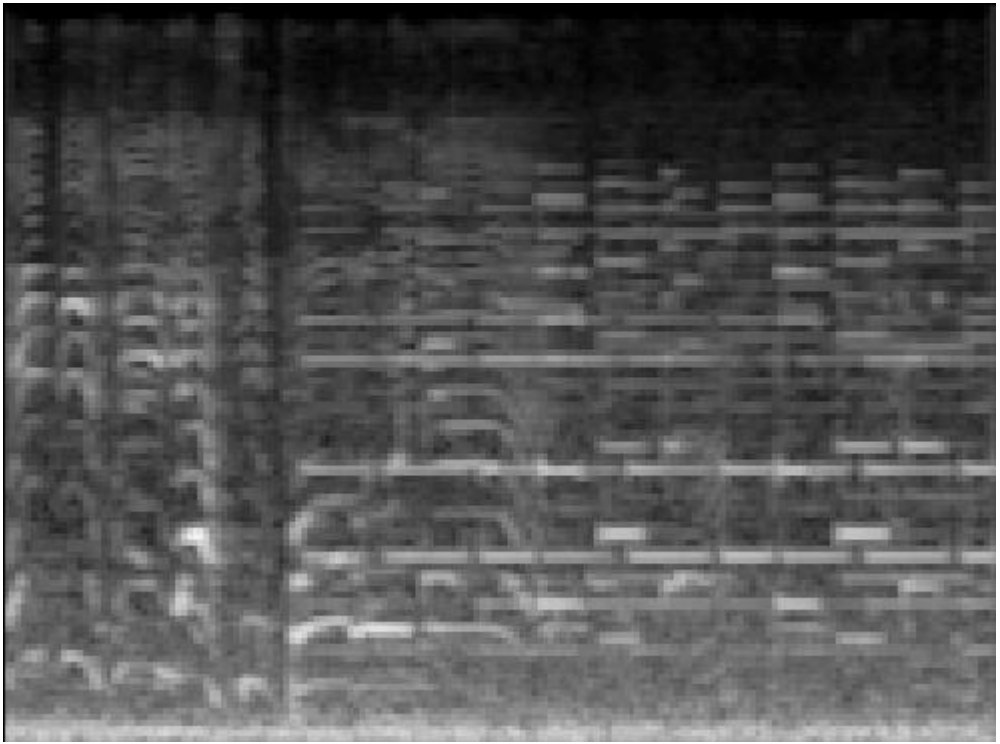
2. **Спектрограми**:

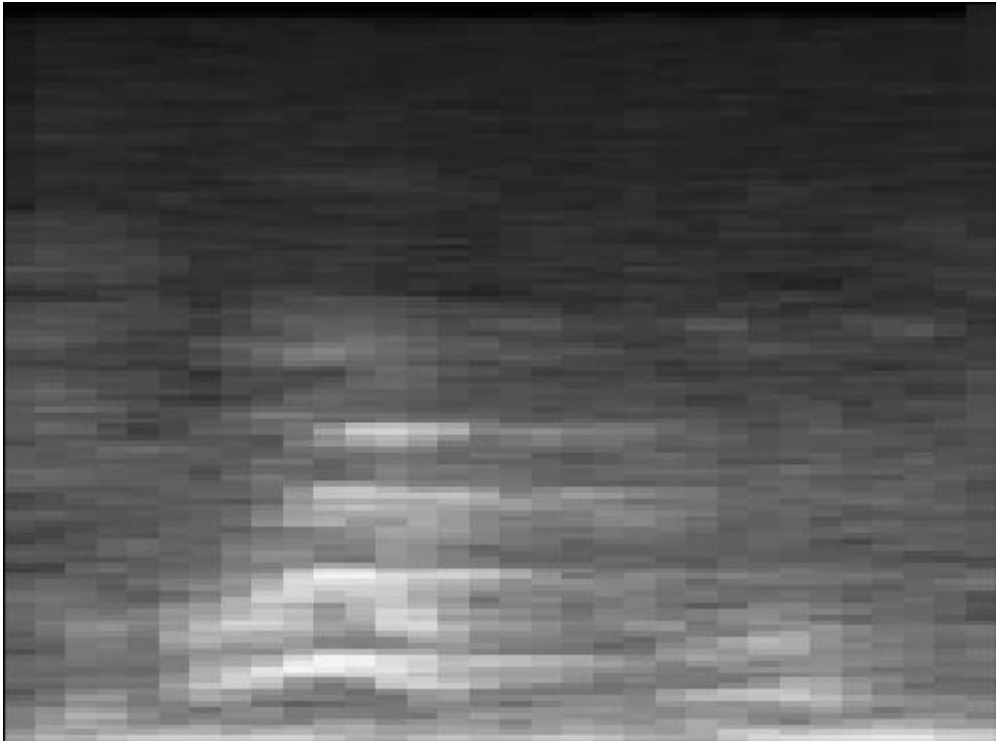
- Спектрограма — це візуалізація зміни частотного складу звукового сигналу за часом. Одна вісь спектрограми відповідає часу, інша — частотам, а кольори або інтенсивність відображають амплітуду частот на певний момент часу.
- Мел-спектрограма — це особлива версія спектрограми, де частоти відображаються за мел-шкалою, яка більше відповідає сприйняттю звуку людським вухом.

3. **Класи**: Як і в оригінальному Urban Sound 8K, у цьому датасеті є **10 класів звукових подій**:

- Автомобільні гудки
- Собачий гавкіт
- Дрель
- Двигуни
- Постріли
- Дитячі ігри на майданчику
- Сирени
- Кондиціонери
- Автобуси
- Сміттєві баки (відбійні молотки)







4. **Метадані**: Оскільки цей датасет є варіацією Urban Sound 8K, метадані можуть залишатися тими ж, що й в оригіналі:

- Ім'я файлу
- Ідентифікатор класу
- Назва класу
- Папка, до якої належить файл (для крос-валідації)
- Інші атрибути, що дозволяють аналізувати або розподіляти дані для навчання та тестування.

5. **Застосування**:

- Датасет призначений для задач класифікації зображень, використовуючи аудіо, представлене у візуальній формі.
- Його використовують у дослідженнях для навчання згорткових нейронних мереж (CNN) та інших моделей комп'ютерного зору для класифікації звукових подій.
- Може бути корисним у системах розпізнавання звуків міського середовища, таких як інтелектуальні міста або автоматичні системи моніторингу шуму.

Особливості:

- **Простота використання з CNN**: Замість роботи з аудіо, модель аналізує зображення, що дозволяє використовувати методи комп'ютерного зору.
- **Перетворення звуку в зображення**: Спектрограми та мел-спектрограми є стандартним способом перетворення звукових даних у форму, зручну для CNN.
- **Візуалізація звукових ознак**: У спектрограмах легко видно зміни частоти, інтенсивності звуку, що допомагає моделям виявляти важливі патерни.

Переваги використання Urban-Sound-8K-Image:

- **Ефективне використання згорткових мереж**: CNN показують високу продуктивність при роботі з візуальними даними, і цей датасет дозволяє адаптувати їх до задач класифікації звуку.
- **Можливість переднавчання на інших зображеннях**: Моделі, натреновані на великих наборах зображень (наприклад, ImageNet), можуть бути адаптовані для класифікації звуків через техніку transfer learning.

Таким чином, Urban-Sound-8K-Image є корисним для експериментів, де аудіо подається у вигляді зображень, що дозволяє використовувати сучасні методи комп'ютерного зору для аналізу звукових сигналів.

Алгоритм побудови згорткової нейронной мережі

1. **Завантаження даних**: використовуємо метадані з датасету Urban Sound 8K, щоб отримати шляхи до зображень (спектрограм) та відповідні класи звуків.

2. Генеруємо зображення з папок і нормалізуємо їх шляхом перетворення значень пікселів у діапазон [0, 1].

3. **Архітектура моделі**: Створюємо згорткову нейронну мережу з кількома шарами згортки та підвибірки, щоб витягувати ознаки з зображень, а потім використовуємо щільні шари для класифікації звуків.

4. **Навчання та оцінка**: Модель тренується на навчальній вибірці, і потім ми оцінюємо її продуктивність на тестових даних. Матриця плутанини та класифікаційний звіт показують детальну інформацію про точність для кожного класу.

5. **Візуалізація результатів**: Графіки відображають динаміку точності та втрат під час навчання.

Згорткова нейронна мережа (CNN) складається з кількох основних шарів

```
model = Sequential([
    Conv2D(32, (3, 3), activation='relu', input_shape=(64, 64, 3)),
    MaxPooling2D(pool_size=(2, 2)),
    Conv2D(64, (3, 3), activation='relu'),
    MaxPooling2D(pool_size=(2, 2)),
    Conv2D(128, (3, 3), activation='relu'),
    MaxPooling2D(pool_size=(2, 2)),
    Flatten(),
    Dense(128, activation='relu'),
    Dropout(0.5),
    Dense(10, activation='softmax')  # 10 класів       ])
```

1. Вхідний шар

Conv2D(32, (3, 3), activation='relu', input_shape=(64, 64, 3))

Це згортковий шар, який має 32 фільтри розміром 3x3. Фільтри використовуються для виявлення локальних ознак у зображеннях (наприклад, контурів або текстур).

Активатор: Функція активації ReLU (Rectified Linear Unit), яка допомагає моделі вчитися нелінійним залежностям.

Вхідна форма: input_shape=(64, 64, 3) означає, що мережа очікує зображення розміром 64x64 пікселів із трьома каналами (RGB).

2. Шар підвибірки (Pooling)

MaxPooling2D(pool_size=(2, 2))

Цей шар зменшує просторовий розмір вхідного зображення. Він обирає максимальне значення в кожному 2x2 регіоні, що дозволяє видаляти зайві або менш важливі деталі, зберігаючи основні ознаки.

3. Другий згортковий шар

Conv2D(64, (3, 3), activation='relu')

Цей шар має 64 фільтри розміром 3x3 і також використовує функцію активації ReLU.

Другий згортковий шар аналізує вихід попереднього шару та виділяє ще більш складні ознаки з зображення.

4. Другий шар підвибірки

MaxPooling2D(pool_size=(2, 2))

Як і попередній шар підвибірки, цей шар зменшує розмір даних, обираючи максимальні значення в 2x2 регіонах, що далі скорочує розмір зображення.

5. Третій згортковий шар

Conv2D(128, (3, 3), activation='relu')

Цей шар має 128 фільтрів і функцію активації ReLU. Третій згортковий шар дозволяє моделі знаходити ще більш складні патерни у зображенні.

6. Третій шар підвибірки

MaxPooling2D(pool_size=(2, 2))

Знову ж таки, шар підвибірки зменшує розмір даних, щоб зменшити обчислювальну складність і залишити тільки найбільш важливі ознаки.

7. Шар розгортання (Flattening)

Flatten()

Цей шар розгортає багатовимірний вхід у плоский вектор. Після кількох шарів згортки та підвибірки ми маємо багатовимірний тензор, який потрібно перетворити у вектор для подальшої обробки у повнозв'язних шарах.

8. Повнозв'язний (Dense) шар

Dense(128, activation='relu')

Це повнозв'язний шар, який має 128 нейронів. Його завдання — обробляти плоский вектор ознак, отриманий з попередніх шарів, та знаходити зв'язки між ознаками.

ReLU: Функція активації ReLU, яка дозволяє моделі навчатися нелінійним залежностям.

9. Шар Dropout

Dropout(0.5)

Цей шар випадково "відключає" 50% нейронів під час навчання, щоб запобігти перенавчанню. Це допомагає зробити модель більш узагальненою і стійкою до надмірного підлаштування до тренувальних даних.

10. Вихідний шар

Dense(10, activation='softmax')

Це повнозв'язний вихідний шар із 10 нейронами, де кожен нейрон відповідає одному з 10 класів звуків у датасеті.

Функція активації softmax, яка перетворює виходи нейронів у ймовірності для кожного класу. Це стандартна функція для багатокласової класифікації, оскільки вона дозволяє моделі вибрати найімовірніший клас.

Основні особливості:

1. **Три згорткові шари** для виділення ознак з зображень.
2. **Три шари підвибірки (Pooling)** для зменшення просторового розміру ознак та зменшення обчислювальної складності.
3. **Повнозв'язні шари** для кінцевої класифікації на 10 класів.
4. **Dropout** для уникнення перенавчання.
5. **Softmax** для отримання ймовірнісної оцінки кожного класу.

Ця структура є типовою для задач класифікації зображень і добре підходить для аналізу спектрограм у задачах класифікації звуку.