**TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TP HỒ CHÍ MINH**

**KHOA CÔNG NGHỆ THÔNG TIN**

****

**BÁO CÁO CUỐI KÌ**

**HỌC SÂU**

**PHÂN LOẠI THỂ LOẠI ÂM NHẠC BẰNG CONVOLUTIONAL NEURAL NETWORK (CNN)**

*Người thực hiện*: **PHẠM QUỐC HUY– 20042071**

Lớp **: 420300411601**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2024**

**TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TP HỒ CHÍ MINH**

**KHOA CÔNG NGHỆ THÔNG TIN**

****

**BÁO CÁO CUỐI KÌ**

**HỌC SÂU**

**PHÂN LOẠI THỂ LOẠI ÂM NHẠC BẰNG CONVOLUTIONAL NEURAL NETWORK (CNN)**

*Người thực hiện*: **PHẠM QUỐC HUY– 20042071**

Lớp **: 420300411601**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2024**

LỜI CẢM ƠN

Đầu tiên, em xin gửi lời cảm ơn đến Trường Đại học Công nghiệp Thành phố Hồ Chí Minh đã đưa bộ môn học sâu vào chương trình giảng dạy để chúng em có cơ hội tiếp thu kiến thức quý giá. Đặc biệt, em xin gửi lời cảm ơn chân thành nhất đến cô Đặng Thị Phúc (Giảng viên bộ môn) đã truyền đạt cho chúng em kiến thức bằng cả tất cả tâm huyết. Thời gian học bộ môn của cô là khoảng thời gian tuyệt vời vì em không chỉ được học lý thuyết mà còn nắm bắt được những kinh nghiệm thực tế hữu ích. Đây sẽ là hành trang để em có thể vững bước trên con đường đã lựa chọn ban đầu.

Bộ môn học sâu là môn vô cùng quan trọng và cơ bản trong quá trình học tập cũng như là môn bổ trợ kiến thức cho quá trình học và làm việc của em sau này. Tuy nhiên, do vốn kiến thức còn nhiều hạn chế và khả năng tiếp thu thực tế còn nhiều bỡ ngỡ. Mặc dù em đã cố gắng hết sức nhưng chắc chắn bài tiểu luận khó có thể tránh khỏi những thiếu sót và nhiều chỗ còn chưa chính xác, kính mong cô xem xét và góp ý để bài tiểu luận của em được hoàn thiện hơn. Em xin chân thành cảm ơn!

PHẦN ĐÁNH GIÁ CỦA GIẢNG VIÊN

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Tp. Hồ Chí Minh, ngày tháng năm

(kí và ghi họ tên)

TÓM TẮT

Phân loại thể loại bài hát bằng mô hình CNN (Convolutional Neural Network) là một bài toán trong học sâu (deep learning), nơi mục tiêu là phân loại một bài hát vào một thể loại cụ thể (như pop, rock, jazz, v.v.). Quy trình thực hiện thường bao gồm các bước chính sau:

* Chuẩn bị dữ liệu:
* **Thu thập dữ liệu âm thanh**: Tải về hoặc thu thập các bài hát và phân loại chúng theo thể loại.
* **Tiền xử lý dữ liệu âm thanh**: Chuyển đổi dữ liệu âm thanh (thường ở định dạng WAV hoặc MP3) thành các đặc trưng có thể sử dụng cho CNN.
* Xây dựng mô hình CNN
* Huấn luyện mô hình
* Đánh giá và tối ưu hóa

MỤC LỤC

[LỜI CẢM ƠN i](#_Toc184197725)

[TÓM TẮT iii](#_Toc184197726)

[1.1 Giới thiệu về bài toán 2](#_Toc184197727)

[1.2 Phân tích yêu cầu của bài toán 2](#_Toc184197728)

[1.3 Phương pháp giải quyết bài toán 2](#_Toc184197729)

[1.3.1 Mô hình tổng quát 2](#_Toc184197730)

[1.3.2 Đặc trưng của mô hình đề xuất 2](#_Toc184197731)

[1.4 Thực nghiệm 7](#_Toc184197732)

[1.4.1 Dữ liệu 7](#_Toc184197733)

[1.4.2 Xử lý dữ liệu 8](#_Toc184197734)

[1.4.3 Công nghệ sử dụng 8](#_Toc184197735)

[1.4.4 Cách đánh giá 8](#_Toc184197736)

[1.5 Kết quả đạt được 8](#_Toc184197737)

[1.6 Kết luận 13](#_Toc184197738)

# 1.1 Giới thiệu về bài toán

Phân loại thể loại bài hát là một bài toán phổ biến trong lĩnh vực xử lý tín hiệu âm thanh và học sâu. Mục tiêu của bài toán là dựa trên dữ liệu âm thanh của một bài hát, xác định thể loại của bài hát đó (như Pop, Rock, Jazz, Classical, Hip-hop, v.v.).

1.2 Phân tích yêu cầu của bài toán

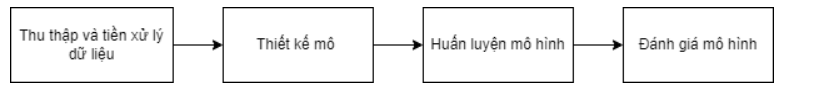
Đầu vào:

* Dữ liệu âm thanh : các file âm thanh ở dạng MP3 hoặc WAV, có độ dài là 30 giây mỗi file.
* Nhãn dán (label) : mỗi file âm thanh được đặt trong thư mục được đặt tên theo thể loại nhạc của file âm thanh đó.

Đầu ra:

* Nhãn dán duy nhất thể hiện thể loại nhạc của file âm thanh được đưa vào

1.3 Phương pháp giải quyết bài toán

 1.3.1 Mô hình tổng quát

1.3.2 Đặc trưng của mô hình đề xuất

* Đầu vào của mô hình :

Là Spectrogram hoặc Mel-spectrogram, chuyển đổi tín hiệu âm thanh thành biểu diễn hình ảnh, trong đó: trục X là thời gian, trục Y là tần số, cường đô âm thanh được biểu thị bằng màu sắc

Kích thước đầu vào đươc chuẩn hóa thành 150x150

* Kiến trúc của CNN

Lớp tích chập (Convolutional Layers): trích xuất đặc trưng từ hình ảnh âm thanh

Lớp pooling (max pooling): giảm kích thước và độ phức tạp, làm nổi bật các đặc trưng.

Lớp fully connected ( Dense): tổng hợp các đặc trưng đã học từ các lớp trước

Thu thập và tiền xử lý dữ liệu:

* Thu thập dữ liệu: Sử dụng các bộ dữ liệu tiêu chuẩn GTZAN Dataset
* Tiền xử lý dữ liệu:
* A screen shot of a graph

  Description automatically generatedChuyển dữ liệu từ dạng âm thanh sang dạng hình ảnh
* Chia dữ có độ dài là 30 giây thành các đoạn (chunk) có độ dài là 4 giây với độ chồng lập (overlap duration) trên mỗi đoạn là 2 giây

A screenshot of a computer screen

Description automatically generated

Thiết kế mô hình:

* Mô hình được xây dựng theo dạng tuần tự (Sequential)
* Gồm 4 lớp tích chập (Convolutional Layers):
* Lớp 1 :

Số lượng filter : 32

Kích thươc kernel : 3x3

Hàm kích hoạt : Relu

Padding : giữ nguyên

Kích thước đầu vào : 150x150x1

* Các lớp tiếp theo :

Số lượng filter tăng dần : 64, 128, 256, 512.

* Dropout layer :

Tắt ngẫu nhiên một số neurons trong qua trình huấn luyện

* Flatten layer :

Chuyển đầu ra 2D từ lớp tích chập thành 1D

* Fully Connected (Dense) Layer 1:

Có một lớp dense với 1200 units và hàm kích hoạt ReLU

* Fully Connected (Dense) Layer 2:
* A screenshot of a computer

  Description automatically generatedCó một lớp dense với số lượng units bằng và hàm kích hoạt softmax để kết thúc mô hình

A screenshot of a computer

Description automatically generated

Huấn luyện và đánh giá mô hình:

* Bộ dữ liệu huấn luyện: Chia bộ dữ liệu thành tập huấn luyện và tập kiểm tra.
* Tiêu chí đánh giá: Sử dụng các chỉ số đánh giá như độ chính xác (accuracy), hàm mất mát, độ chính xác trên tập test, hàm mất mát trên tập test, độ nhạy (recall)

1.4 Thực nghiệm

1.4.1 Dữ liệu

Nguồn dữ liệu:

* Nguồn: [*GTZAN Dataset - Music Genre Classification*](https://www.kaggle.com/datasets/andradaolteanu/gtzan-dataset-music-genre-classification/data)

Mô tả dữ liệu:

* A screenshot of a computer

  Description automatically generatedA screenshot of a computer

  Description automatically generatedBộ dữ liệu gồm 10 thể loại nhạc với 100 file âm thanh 30 giây cho mỗi thể loại

1.4.2 Xử lý dữ liệu

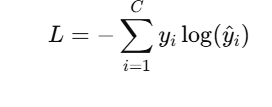
* Chuyển dữ liệu từ dạng âm thanh thành hình ảnh
* Chuyển kich thước ảnh về 150x150

1.4.3 Công nghệ sử dụng

* Ngôn ngữ lập trình: Python.
* Thư viện: TensorFlow, Keras, Numpy
* Công cụ: Jupyter Notebook.

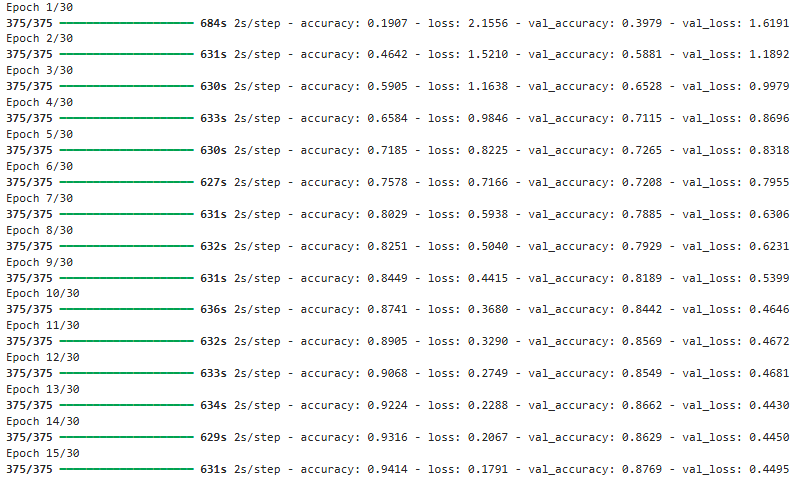
1.4.4 Cách đánh giá

* Độ đo đánh giá:
* A close up of a sign

  Description automatically generatedĐộ chính xác (accuracy)
* Hàm mất mát (loss)
* Độ chính xác trên tập test (Validation Accuracy)
* Hàm mất mát trên tập test (Validation Loss)
* Độ nhạy ( recall)

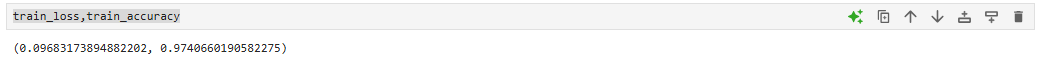
1.5 Kết quả đạt được

* Tham số thực nghiệm:
* Learning rate : 0.0001
* Hàm mất mát : cross entropy
* Số lượng epochs : 30
* Batch : 32A screenshot of a computer code

  Description automatically generated
* Quá trình huấn luyện

A screenshot of a computer

Description automatically generated

* Kết quả sau khi huấn luyện mô hình:
* accuracy: 0.9853
* loss: 0.0487
* val\_accuracy: 0.8773
* val\_loss: 0.5895
* Kết quả của mô hình trên tập Train:
* accuracy: 0.9737
* loss: 0.0941
* train\_loss: 0.0968
* train\_accuracy: 0.974

A graph of a graph showing the value of a performance

Description automatically generated with medium confidenceA graph of loss results

Description automatically generated

* A screenshot of a calendar

  Description automatically generatedKiểmA screenshot of a computer

  Description automatically generated tra mô hình với dữ liệu ngoài

A screenshot of a computer program

Description automatically generatedA white and black background

Description automatically generated with medium confidence

1.6 Kết luận

* Kết quả đánh giá mô hình:
* Độ chính xác: Nếu mô hình đạt độ chính xác cao trên tập kiểm tra (85%)
* Hàm mất mát: Giá trị loss thấp, không bị overfitting (quá khớp).
* Hạn chế và thách thức
* Dữ liệu hạn chế
* Độ phức tạp tính toán chưa cao

TỰ ĐÁNH GIÁ (Bài cá nhân)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| STT | Nội dung | Điểm chuẩn | Tự chấm | Ghi chú |
| 1  (9đ) | 1.1 Giới thiệu về bài toán | 0.5 | 0,5 |  |
| 1.2 Phân tích yêu cầu của bài toán | 1 | 1 |  |
| 1.3 Phương pháp giải quyết bài toán | 1.5 | 1 |  |
| 1.4 Thực nghiệm | 4.5 | 3 |  |
| **1.5 Kết quả đạt được** | 1 | 1 |  |
| 1.6 Kết luận | 0.5 | 0,5 |  |
| 2  (1đ) | Báo cáo (chú ý các chú ý 2,3,4,6 ở trang trước, nếu sai sẽ bị trừ điểm nặng) | 1đ | 1 |  |
| Tổng điểm | | |  |  |