

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN



NHẬP MÔN CÔNG NGHỆ THÔNG TIN

Đồ án

PHÂN LOẠI CHỮ SỐ

Giảng viên thực hành: Lê Đức Khoan

Thành phố Hồ Chí Minh, 11/2025

Mục lục

1	Chuẩn đầu ra	1
2	Mô tả đồ án	1
2.1	Nội dung chính	1
2.2	Mục tiêu đồ án	1
2.3	Nội dung chi tiết	1
3	Các quy định và yêu cầu	2
4	Yêu cầu mỗi tuần	2
5	Hướng dẫn thực hành	2
5.1	Cấu trúc thư mục	3
5.2	Cài đặt môi trường	4
5.2.1	Cài đặt quản lí môi trường	4
5.2.2	Cài đặt thư viện	4
5.2.3	Tập dữ liệu	4



1 Chuẩn đầu ra

Đồ án này nhằm mục tiêu đạt được các chuẩn đầu ra sau:

- Tham gia thảo luận, tranh luận theo nhóm trên chủ đề môn học.
- Phân tích, tổng hợp và viết tài liệu kỹ thuật theo mẫu cho trước theo cá nhân hoặc cộng tác nhóm.
- Biết, hiểu thuật ngữ tiếng Anh thuộc các ngành và chuyên ngành.
- Xây dựng một ứng dụng đơn giản, ứng dụng kiến thức ngành Thị giác máy tính.

2 Mô tả đồ án

2.1 Nội dung chính

Tự huấn luyện một mô hình AI cho bài toán phân loại chữ số.

2.2 Mục tiêu đồ án

Sinh viên thực hiện đồ án sẽ nắm được các kiến thức:

- Cấu trúc của game đơn giản (CNPM).
- Các bước xây dựng phần mềm (CNPM).
- Một số công cụ hỗ trợ làm việc trên môi trường Internet (MMT).

Sinh viên thực hiện đồ án sẽ rèn luyện các kỹ năng:

- Lập kế hoạch, làm việc nhóm, tổ chức và quản lý nhóm.
- Tìm hiểu và viết báo cáo.
- Đọc hiểu tiếng Anh và các tài liệu chuyên ngành.

2.3 Nội dung chi tiết

Sinh viên được cung cấp đoạn code xử lý dữ liệu đơn giản. Hãy thực hiện những yêu cầu sau:

1. Cài đặt thành công môi trường huấn luyện mô hình AI.



2. Chuẩn bị dữ liệu bao gồm: hình ảnh, âm thanh,...
3. Cách tổ chức mã nguồn với các dự án liên quan đến trí tuệ nhân tạo.
4. Thiết kế mô hình, huấn luyện, đánh giá mô hình.
5. Xử lý và các kỹ thuật làm giàu dữ liệu.

3 Các quy định và yêu cầu

Các kết quả cần đạt được:

- Báo cáo mô tả về đề án: Mô tả về cách làm, công cụ sử dụng, khó khăn, thuận lợi, các chức năng của chương trình, kết hoạch thực hiện, phân công công việc và đánh giá từng thành viên.
- Github chứa mã nguồn của đề án. Đảm bảo có phân chia công việc và sử dụng các chức năng của Git như branch, pull request, merge (kiểm tra mỗi tuần về mã nguồn).
- Video báo cáo tiến độ từng tuần.
- Sản phẩm hoàn thiện.

4 Yêu cầu mỗi tuần

- **Tuần 1:** Cài đặt thành công môi trường lập trình, thư viện, chuẩn bị dữ liệu.
- **Tuần 2:** Xây dựng dataloader.
- **Tuần 3:** Xây dựng kiến trúc mô hình.
- **Tuần 4:** Xây dựng file huấn luyện mô hình.
- **Tuần 5:** Đánh giá và chạy thử nghiệm mô hình.

5 Hướng dẫn thực hành

Đề án này sẽ sử dụng thư viện Pytorch của Python để thiết kế:

5.1 Cấu trúc thư mục

```
digits_classification/  
├── configs/  
│   └── config.yaml  
├── src/  
│   ├── losses/  
│   │   └── loss.py  
│   ├── models/  
│   │   └── model.py  
│   └── data/  
│       └── dataloader.py  
├── trainer.py  
├── requirements.txt  
└── README.md
```

Giải thích chức năng từng file:

- **configs/config.yaml:** Chứa các thông số huấn luyện như số epoch, batch size, learning rate, và đường dẫn dữ liệu. Giúp dễ dàng thay đổi cấu hình mà không cần sửa code.
- **src/losses/loss.py:** Định nghĩa hàm mất mát (*loss function*) — ví dụ: CrossEntropyLoss — dùng để đo sự khác biệt giữa dự đoán của mô hình và nhãn thật.
- **src/models/model.py:** Xây dựng kiến trúc mạng nơ-ron, ví dụ MLP hoặc CNN, để nhận dạng chữ số từ ảnh (dataset như MNIST).
- **src/data/dataloader.py:** Chuẩn bị và tải dữ liệu huấn luyện, chia thành tập train/validation/test, đồng thời chuẩn hóa dữ liệu (0–1).
- **trainer.py:** Chứa vòng lặp huấn luyện chính. Gọi model, loss, optimizer; tính toán độ chính xác, lưu mô hình tốt nhất.
- **requirements.txt:** Liệt kê các thư viện cần thiết (ví dụ: torch, torchvision, numpy, tqdm) để chạy dự án.
- **README.md:** Mô tả tổng quan dự án — mục tiêu, cách cài đặt môi trường, cách chạy huấn luyện và đánh giá mô hình.



5.2 Cài đặt môi trường

5.2.1 Cài đặt quản lí môi trường

Với đồ án này, sinh viên sử dụng các công cụ như conda, miniconda, micromamba để tạo môi trường và cài đặt thư viện. Tham khảo tại: [Micromamba](#).

5.2.2 Cài đặt thư viện

Cài đặt thư viện pytorch. Sau khi đã tạo được môi trường với micromamba.

```
micromamba activate digits_classification  
pip install -r requirements.txt
```

5.2.3 Tập dữ liệu

Tham khảo tại: [MNIST](#).