RMIT UNIVERSITY

Deep Learning
COSC 2779/2972 | Semester 2 2024
Assignment 1: Introduction to Deep Convolutional Neural Networks
(Human Action Recognition)

Assessment Type	Individual assignment. Submit online via Canvas → Assignments → Assignment 1. Marks awarded for meeting requirements as closely as possible. Clarifications/updates may be made via announcements/relevant discussion forums.
Due Date	5.00pm, Friday 30 August 2024 (Week 6)
Marks	30%

1 Overview

In this assignment you will explore a real dataset to practice the typical deep learning process. The assignment is designed to help you become more confident in applying deep learning approaches. In this assignment you will:

- Develop a deep learning system to solve a real-world problem.
- Analyse the output of the algorithm(s).
- Research how to extend the DL techniques that are taught in class.
- Provide an ultimate judgement of the final trained model that you would use in a real-world setting.

To complete this assignment, you will require skills and knowledge from lecture and lab material for Weeks 1 to 6 (inclusive). You may find that you will be unable to complete some of the activities until you have completed the relevant lab work. However, you will be able to commence work on some sections. Thus, do the work you can initially, and continue to build in new features as you learn the relevant skills. *A deep learning model cannot be developed within a day or two. Therefore, start early.*

2 Learning Outcomes

This assessment relates to all of the learning outcomes of the course which are:

- Discuss and critically analyse a variety of neural network architectures; Evaluate and Compare approaches and algorithms on the basis of the nature of the problem/task being addressed.
- Synthesise suitable solutions to address particular machine learning problems based on analysis of the problem and characteristics of the data involved.

- Communicate effectively with a variety of audiences through a range of modes and media, in particular to: interpret abstract theoretical propositions, choose methodologies, justify conclusions and defend professional decisions to both IT and non-IT personnel via technical reports of professional standard and technical presentations.

- Develop skills for further self-directed learning in the general context of neural networks and machine learning; Research, Discuss, and Use new and novel algorithms for solving problems; Adapt experience and knowledge to and from other computer sciences contexts such as artificial intelligence, machine learning, and software design.

3 Assessment details

3.1 Task

Human action recognition (HAR) has a wide range of applications, such as intelligent video surveillance and environmental home monitoring, video storage and retrieval, intelligent human–machine interfaces, and identity recognition. Usually, HAR is performed with videos. However, processing videos on low compute devices such as IoT edge devices can be challenging and therefore it is important to develop HAR systems that can work with still images.

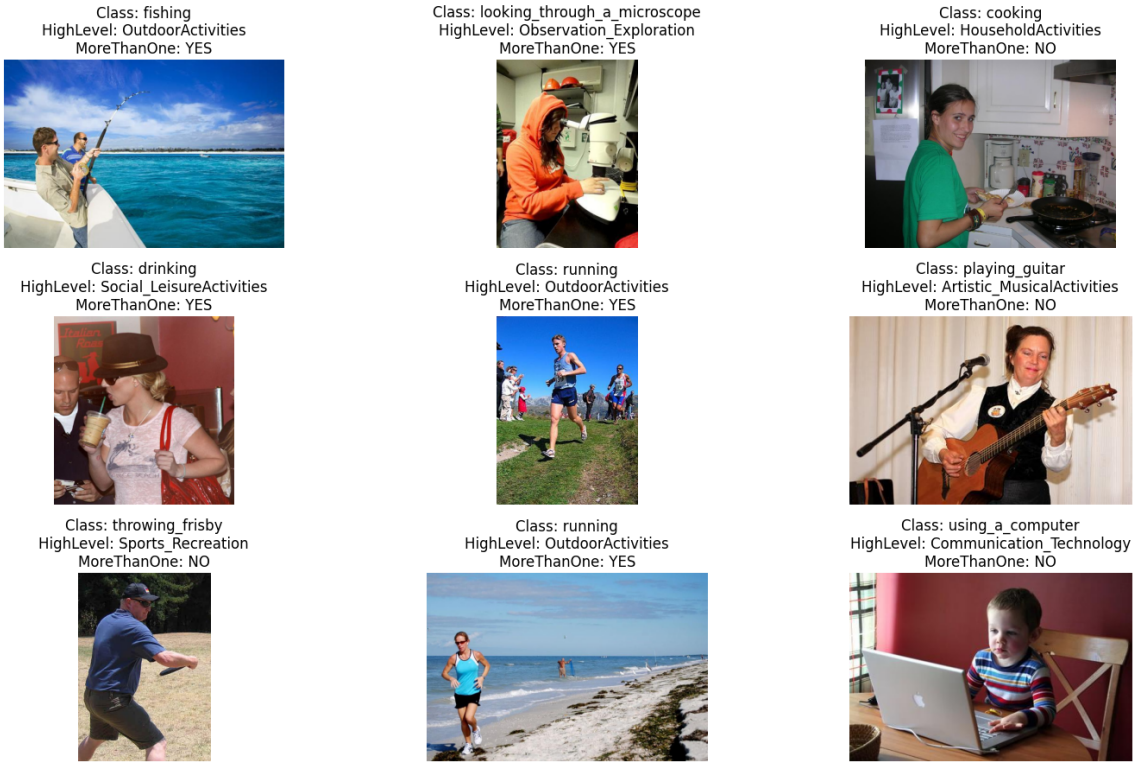

Figure 1: Example Images and labels from the dataset.

In this assignment, you will develop a deep convolutional neural network (CNN) to identify the actions of a person from still images. The actions of interest are categorized into 40 types (Classes). Predicting the action of the subject becomes challenging when multiple people are visible in the image. Therefore, we are also interested in detecting if there is more than one person in the image, as this will provide an indication of confidence in the prediction. The task is to design a thorough experiment and determine how well the following two attributes can be predicted:

- Predict the action performed by a person in an input RGB image. Actions are categorised into 40 categories (column name - '`Class`').
- Predict whether more than one person is present in the image (column name - '`MoreThanOnePerson`').
- An additional label '`HighLevelCategory`' is present in the data. You may use this in any way you like during the model development/analysis phase. However, predicting this label is not required.

Please read the following **requirements and guidelines** carefully.

- **Must** use Tensorflow 2 with Keras. Zero marks otherwise.
- You need to design a network that takes in an image as input and predicts the action '`Class`' and if '`MoreThanOnePerson`' in Image. For higher grades (CR/DI/HD) you should develop a **single network** that makes both predictions (NOT two separate networks).
- You *may* use pre-trained networks as part of your solution. However, there needs to be **"clearly identifiable" network segment(s) that is designed and trained by you**. You should show how this segment is developed (tuned) in you code.
- **Only neural network** based techniques can be used in the assignment. Other ML techniques such as SVM, and RF cannot be used.
- (Notebook Markdown) You need to come up with a deep learning system, where **each element of the system is *justified*** using data analysis, performance analysis and/or knowledge from relevant literature. Clearly document at the start of the Notebook Section using Markdown.
- (Notebook Markdown) You should **clearly explain your evaluation framework**, including how you selected appropriate performance measures, and how you determined the data splits.
- (Notebook Markdown) Finally you need to analyse the results from your model using appropriate techniques and establish how adequate your model is to perform the task in the real world and discuss limitations if there are any (**ultimate judgment**).
- (Notebook Markdown) While there are many practical utilities of HAR technology, there are many **ethically challenging scenarios** raised by such technology as well. discuss such challenges in general and biases in the dataset. Don't write lengthy paragraphs. A point-form answer is acceptable.

- Predict the '`Class`' and '`MoreThanOnePerson`'. For the future data provided with in the `Future_data_2024.csv` file.

3.2 Dataset

The data set for this assignment is available on Canvas. There are the following files:

- "`README.md`": Description of dataset.
- "`Images.zip`": Contain all the images (test and train set).
- "`train_data_2024.csv`": This data is to be used in developing the models. Use this for your own exploration and evaluation of which approach you think is "best" for this prediction task.
- "`future_data_2024.csv`": You need to predict the action & action-class for this data and submit the prediction via canvas. The teaching team will use this data to evaluate the performance of the model you have developed. This data should not be considered as a part of the model development data.
- "`s1234567_predictions.csv`": Shows the expected format for your predictions on the unseen test data. You should organize your predictions in this format. Any deviation from this format will result on zero marks for the results part (this include the order of images). Change the number to your student ID. **Note:** Your predictions should be in text format and match category labels in train data file - integer predictions will not be accepted.

The original data is from Stanford 40 Actions Dataset published with *B. Yao, X. Jiang, A. Khosla, A.L. Lin, L.J. Guibas, and L. Fei-Fei. Human Action Recognition by Learning Bases of Action Attributes and Parts. Internation Conference on Computer Vision (ICCV), Barcelona, Spain. November 6-13, 2011.*

Licence agreement: The dataset can only be used for the purpose of this assignment. Sharing or distributing this data or using this data for any other commercial or non-commercial purposes is prohibited (including uploading the data to a public GitHub repository).

4 Suggested Schedule

We expect that you will start the assignment immediately and follow a schedule similar to the one shown below. Do not fall behind, *A deep learning model cannot be developed within a day or two.*

- **Week 2-3**: Read the specification and familiarize yourself with the problem. Explore the data set and the task. Identify biases and ethical issues and start writing the report.
- **Week4**: Design the experiments. Develop the data loading mechanism. Search relevant literature and read. Start writing the report.
- **Week5**: Develop the model - design, and train. Do model analysis. Update the report.
- **Week6**: Do model analysis. Update the report. Submission!

5 Submission

You have to submit all the relevant material as listed below via Canvas.

1. **PDF version of your Notebook**: This needs to be the final version of your notebook that includes the analysis and explanations. *Will be used for plagiarism checking.*

2. **Your code/analysis in Notebook format**: Colab notebook(s) used to perform your modelling with justifications and critical analysis of your approach. Should be a ZIP file. The Jupyter notebook(s) should be clearly commented on in markdown format (see labs and lectorial exercises). **The final outputs should be visible**.

3. Predicted labels for the future data set. You should organize your predictions in the format given in `s1234567_predictions.csv`. Any deviation from this format will result in zero marks.

The submission portal on canvas consists of pages sub-pages. Page one for PDF-code submission, Page two for code-Notebook submission, and Page three for Prediction submission. More information is provided on canvas.

We strongly recommend you to attach a README file with instructions on how to run your application. Make sure that your assignment can run with the code included in your zip file and files provided in the assignment specifications! Include a PDF version of your report.

After the due date, you will have 5 business days to submit your assignment as a late submission. Late submissions will incur a penalty of 10% per day. After these five days, Canvas will be closed and you will lose ALL the assignment marks.

Assessment declaration:
When you submit work electronically, you agree to the assessment declaration - `https://www.rmit.edu.au/students/student-essentials/assessment-and-exams/assessment/assessment-declaration`

6 Teams

Not relevant. This is an individual assignment.

7 Academic integrity and plagiarism (standard warning)

Academic integrity is about honest presentation of your academic work. It means acknowledging the work of others while developing your own insights, knowledge and ideas. You should take extreme care that you have:

- Acknowledged words, data, diagrams, models, frameworks and/or ideas of others you have quoted (i.e. directly copied), summarised, paraphrased, discussed or mentioned in your assessment through the appropriate referencing methods

- Provided a reference list of the publication details so your reader can locate the source if necessary. This includes material taken from Internet sites. If you do not acknowledge the sources of your material, you may be accused of plagiarism because

you have passed off the work and ideas of another person without appropriate referencing, as if they were your own.

RMIT University treats plagiarism as a very serious offence constituting misconduct. Plagiarism covers a variety of inappropriate behaviours, including:

- Failure to properly document a source
- Copyright material from the internet or databases
- Collusion between students

For further information on our policies and procedures, please refer to the following: `https://www.rmit.edu.au/students/student-essentials/rights-and-responsibilities/academic-integrity`.

8 Marking guidelines

A detailed rubric is attached on canvas. In summary:

- Approach 40%;
- Ultimate Judgment, performance & Analysis 40%;
- Discussion on Ethical issues and biases 20%;

Approach: You are required to use a suitable approach to find a predictive model. Each element of the approach need to be *justified* using data analysis, performance analysis and/or published work in literature. *This assignment isn't just about your code or model, but the thought process behind your work.* The elements of your approach may include:

- Setting up the evaluation framework
- Selecting CNN architecture, loss function and optimization procedure.
- Hyper-parameter setting and tuning
- Identify problem specific issues/properties and solutions

Ultimate Judgement: You must make an *ultimate judgement* of the "best" model that you would use and recommend in a real-world setting for this problem. It is up to you to determine the criteria by which you evaluate your model and determine what is means to be "the best model". You need to provide evidence to support your ultimate judgement and discuss limitation of your approach/ultimate model if there are any.

Discussion on Ethical issues and biases: While there are many practical utilities of emotion recognition technology, there are many ethically challenging scenarios raised by such technology as well. discuss such challenges in general and biases in the dataset. **Don't need lengthy paragraphs. a point-form answer is acceptable**.

Critical Analysis: Finally, you must include markdown text in the notebook describing and analysing the approach that you have taken to find a suitable model and make your ultimate judgement.

In this analysis you should describe elements such as:

- Your final selected approach
- Why you selected this approach
- Parameter settings and other approaches you have tried.
- Limitation and improvements that are required for real-world implantation.

This will allow us to understand your rationale. We encourage you to explore this problem and not just focus on maximising a single performance metric. By the end of your notebook, we should be convinced that of your ultimate judgement and that you have considered all reasonable aspects in investigating this problem.

Remember that good analysis provides *factual statements, evidence and justifications for conclusions* that you draw. Statements such as:

> *"I did xyz because I felt that it was good"*

is not analysis. This is an unjustified opinion. Instead, you should aim for statements such as:

> *"I did xyz because it is more efficient. It is more efficient because . . . "*