**Project Notes**

Diss sections

1. Dolphin Setup
2. Memory accessing

Literature

Hierarchical reinforcement learning for self‐driving decision‐making without reliance on labelled driving data

* Independent Manoeuvre Selection and Motion Control
* APRL – Asynchronous Parallel RL – Each learner has own policy network, average is taken at each iteration, learners synchronise form shared network before making new decisions

Hands-On Reinforcement Learning with Python

* Reinforcement Learning Basics – Policy Function, Value Function, Agent-Environment Interface, Q Function, MDP, Bellman Equation, Monte Carlo Methods
* Policy Optimisation
* Feature exploitation vs Feature exploration

Deep Learning, Reinforcement Learning and World Models

* Generalised Policy Updates
* World Models

Lightweight 3D hand pose estimation by cascading CNNs with reinforcement learning

* CNN and feature extraction
* Feature exploitation vs Feature exploration

Source - <https://github.com/JackWBoynton/mariokart-rl/README.md>

Dolphin Config : GFX.ini -> Added `DumpFramesAsImages = True` in `[Settings]`

: Added hotkeys for frame advancing, learning can run independent of game framerate. Can be done not in real time

Installed Tesseract OCR Engine <https://github.com/tesseract-ocr/tesseract>

Funky Flame Runner: MT = 113 ,Wheelie = 97 ,Normal = 84

If I choose to advance the emulator frame by frame to allow more time for processing, then I will need to take into account the fact that dolphin only recognises inputs during a frame change. So the input would need to be held, frame advanced, then released. This approach could be very interesting to show the process of what happens each frame.

Charged MT = 270

Game version – ntsc-u

## Gecko Codes

To display relevant in-game information, a gecko code is used. This is a type of cheat code that is often used by people performing TASs (Tool-Assisted Speedruns) when they need to know the exact state of the game, which will also be very useful for me. These codes are supported by the emulator and can be activated/disactivated using the GUI.

The format for the Values I will be using is as follows

* XZ Velocity – A decimal number always to 2 DP, surprisingly this stays positive even when moving backwards
* Race% - A decimal number always to 5dp
  + Number before decimal is important too
* MT – A whole number between 0 and 270

To process the information displayed by the ‘TAS Debug Panel’ created by [Melg, stebler, Seeky, SwareJonge, kierio04, Gaberboo] I need to extract the text from each frame. To do this I used Google’s Tesseract Optical Character Recognition engine. Running the game at native (640x528) resolution gives uncertain results, so to improve the accuracy I increased the resolution of the game in the emulators settings. This would give the framedumps a higher resolution as the emulator dumps frames at the resolution it is running at. When deciding the crop regions, I made sure to leave a small border around the character. This along with the configuration settings helps tesseract to provide accurate results. These configurations include setting the psm (page segmentation mode) level which tells tesseract the high level format of the text within the image, such as paragraphs, single lines or single characters. For my use case this is set to a single line (level 6). Additionally I enforced a whitelist of the numerical characters. This is mainly to avoid confusion between similar looking pairs of characters such as 0 and O and I and 1. I decided to not include the decimal point since I can divide by the appropriate amount to get the numbers into the expected formats.

## Dolphin wrapper

* Init
  + Open dolphin
  + Assert config options – framedumps, input recording
  + Load default savestate
  + Pause emulation
* Step
  + Read frame
  + Calculate estimated reward for each input
  + Update Q function?
  + Return input with highest estimated reward
* Reset
  + Load default savestate
  + Reset controller
* Terminate
  + Return sum of rewards
  + Stop emulation
  + Reset controller
  + Wipe framedumps folder

## Controller

* Reset
* Press button
  + Button name
* Angle stick
  + Stick
  + Anlge (-1 -to 1)

## Frame Reader

* Open file
* Greyscale
* Downsample
* Return information

## Dolphin Hotkeys

Toggle Pause – P

Frame Advance - ]

Reset – R

## Dolphin Configs info

**To use custom Dolphin User Directory use -u option in command line**

C:\Path\To\Dolphin\Build\Dolphin.exe -u E:\Path\To\User\Directory\

* Dolphin.ini
  + [Display]
    - Fullscreen Resolution – set to 720p
    - Render Window X/Y Pos -
    - Render Window Width/Height
  + [Interface]
    - ShowLogWindow = True
    - ShowLogConfigWindow = True
* GFX.ini
  + [Settings]
    - DumpFramesAsImages = True

## Problems

* Frame Processing
  + Tesseract config
    - Character whitelist
    - Specify single line
  + Clarity
    - Low resolution and small text makes blurry characters
      * Lambda to make b/w giving clearer borders
    - Font that resembled 7-segment display lead to a lot of errors
      * Change gecko code so that numbers are normal font
* Controller inputs
  + Dolphin controller state would not change unless I call get\_gc\_buttons() after set\_gc\_buttons()
* Savestate loading
  + Dolphin API has known bug where loading a savestate using the API (https:\\github.com/TASLabz/dolphin/issues/123)
* Reward function – valued different components too much
  + Agent would charge MT into a wall – not even release
    - Give -ve reward if agent ended with charged mt
    - Give reward for surviving longer
* When given an empty q-table, if exploiting would always return 0 (first action in list)
  + Add condition to check if its empty, if so return random action

## Decisions

* Reward function – iterative design process
  + Variables to use
  + Thresholds
  + Reward amounts
  + Decision to add CP - bonus
* Processing every 8 frames – effects of changing action can be seen
  + Deciding this value - iterative
* State Space – rounding values – unnecessary precision in race% and vel – will change significantly after T frames
  + Rounding too much meant values that were far apart horizontally were seen as same state – stuck in optima
* Termination check min speed – soon into offroad, byt not too soon so that wheelie+direction terminates

## Investigation

Reward stays low for over 7000 episodes – why?

Investigate how much of the reward each part is contributing

Race% was increasing by ~0.0005 every frame and as I was rounding to 3.dp, this was giving a reward of 0.1 every other frame. -> round to 4dp to ensure different value every frame (if speed is low enough to give a smaller increase then episode will end anyway)

Additionally I want to balance reward from speed and reward from completion. Speed is between 0.5-0.7. -> give race% reward a weight of 6

Race% and Velocity are now balanced and give a similarly scaled value each frame

* Going fast is as valuable as going round the track

No single state fully explored after 350 episodes…

Too complex – every 10 frames significant changes – 0.dp is enough to represent

Race% Investigation – driving across track gives very small increase

Goes round corners – horizontal position in road is measured

Action space – remove hopping from action space – if B is held with no direction in one step and with direction in the next step then a drift will not be performed

Problem – random in unexplored state is always first action in list

Termination check – get road type

Q-Values converging

Race% to 3.dp, investigation seeing how value changes driving across track

**2 MAJOR OVERSIGHTS**

**State could not tell which direction a drift was started in**

**-**mt part now is {0,1,2} where 0 is no mt, 1 is left, 2 is right

**State could not tell whether a wheelie was performed or not when speed below 84**

* Caused steering to be weird

**Training started – 12:10,**

**5hrs + 16 20 hrs**

**6000 eps + 16000**

**NEW: (Boolean mt charge): 6681(23/3-16:55) +**

q creatd 22 march 10:18, 9hrs stopped at 10:46 ~20 hours

**14692 (23/3 – 23:45), 7729 23/3-16:55 ~7000 episode ~20 hours**

**Xzpos Run – ep3257, turn oil spill is defeated. Started speedup at 31/3/24, approximately 250%**

**FULL\_LAP\_Q\_TABLE = 43**

Agent B – 2491 episodes, 09:55-14:36, Time: 37.703