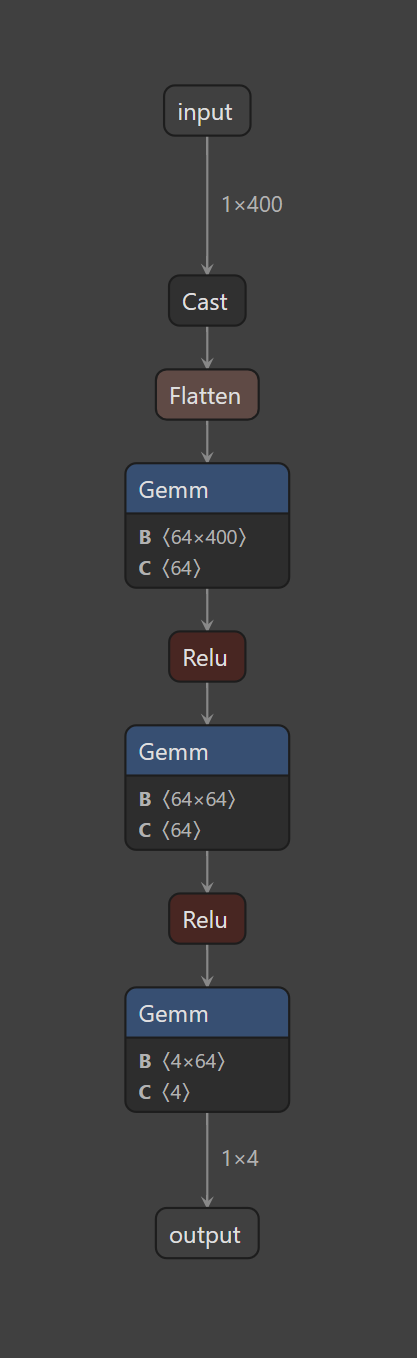
**Visualization of the DQN by Snake Game**

**performing RL on itself**



This network is the visualization of the model created by the game through RL by playing with itself close to 3200 episodes.

Instead of using a formal Q maximization and divergence minimization between Q predictions and the actual Q results in a game, I will avoid the hard-to-understand equations (as well as avoid needing to deal with Latex)

Using a rather simple DQN, it has produced this deep neural network architecture taking in 400 inputs (for the 20x20 grid Snake Environment) to approximate the calculation of all possible Q (rewards/outputs) for any action done on any given state on the 400 positions in the grid (Snake environment).

There are only 4 possible actions (snake moves up, down, left or right).

There are also 4 total possible states for any point in the grid:

1. Occupied by a part of the snake
2. Occupied by a trap/obstacle
3. Free
4. Occupied by an apple/fruit

*Side observation: The auto generated code of the snake game got all these possibilities but appears not to do the usual human programmer intuition (not a big deal) of ordering the states by reward points and for ease of code readability (meaning the most desirable is to eat the fruit for a value of 4, while least score is to go to a grid occupied by a snake as it will kill itself and end the game). See actual code generated for critic but for what it is, it is a perfectly good code that is generated in less than 2 minutes of combined reasoning and output.*

The resulting AI model avoids the need to create a 400x4x4 combinatoric system for maximizing Q by creating a 3-layer NN with 400 original inputs into 64 neurons and outputting eventually just 4 actions and corresponding Q values. All heuristics and activation logic for approximating the values are basic ML functions like ReLU.

Whether or not a training run of more episodes or other epsilon values or a more sophisticated DQN approach like double-DQN etc. would produce a significantly smarter game logic, is something to investigate but not very important at the moment.

Here below are print outs about the model as created, which are just some output of typical characteristics that describe the NN after loading the binary image of the model (created during the RL training session).

We are calling this model “DQN\_snake\_model1” as we anticipate producing other models with other RL sessions sometime for comparisons. This was also converted to .onnx format in the same python script to output the characteristics printed below. The above interpretation from Netron of the .onnx file appears to be a great visualization tool.

=== Model Policy Architecture ===

DQNPolicy(

(q\_net): QNetwork(

(features\_extractor): FlattenExtractor(

(flatten): Flatten(start\_dim=1, end\_dim=-1)

)

(q\_net): Sequential(

(0): Linear(in\_features=400, out\_features=64, bias=True)

(1): ReLU()

(2): Linear(in\_features=64, out\_features=64, bias=True)

(3): ReLU()

(4): Linear(in\_features=64, out\_features=4, bias=True)

)

)

(q\_net\_target): QNetwork(

(features\_extractor): FlattenExtractor(

(flatten): Flatten(start\_dim=1, end\_dim=-1)

)

(q\_net): Sequential(

(0): Linear(in\_features=400, out\_features=64, bias=True)

(1): ReLU()

(2): Linear(in\_features=64, out\_features=64, bias=True)

(3): ReLU()

(4): Linear(in\_features=64, out\_features=4, bias=True)

)

)

)

=== Q-Network Architecture ===

QNetwork(

(features\_extractor): FlattenExtractor(

(flatten): Flatten(start\_dim=1, end\_dim=-1)

)

(q\_net): Sequential(

(0): Linear(in\_features=400, out\_features=64, bias=True)

(1): ReLU()

(2): Linear(in\_features=64, out\_features=64, bias=True)

(3): ReLU()

(4): Linear(in\_features=64, out\_features=4, bias=True)

)

)

=== Layer Details and Weights ===

q\_net.0.weight: shape = (64, 400), requires\_grad = True

q\_net.0.bias: shape = (64,), requires\_grad = True

q\_net.2.weight: shape = (64, 64), requires\_grad = True

q\_net.2.bias: shape = (64,), requires\_grad = True

q\_net.4.weight: shape = (4, 64), requires\_grad = True

q\_net.4.bias: shape = (4,), requires\_grad = True

Using fallback dummy input of shape: (1, 400)

Running a forward pass through the Q-network to capture activations...

Output of Q-network forward pass:

tensor([[-34.8706, -38.9922, -36.1658, -44.3678]])

Captured Activations (layer: activation shape):

features\_extractor.flatten: (1, 400)

q\_net.0: (1, 64)

q\_net.1: (1, 64)

q\_net.2: (1, 64)

q\_net.3: (1, 64)

q\_net.4: (1, 4)

Exported graph: graph(%input : Float(1, 400, strides=[400, 1], requires\_grad=0, device=cpu),

%q\_net.0.weight : Float(64, 400, strides=[400, 1], requires\_grad=1, device=cpu),

%q\_net.0.bias : Float(64, strides=[1], requires\_grad=1, device=cpu),

%q\_net.2.weight : Float(64, 64, strides=[64, 1], requires\_grad=1, device=cpu),

%q\_net.2.bias : Float(64, strides=[1], requires\_grad=1, device=cpu),

%q\_net.4.weight : Float(4, 64, strides=[64, 1], requires\_grad=1, device=cpu),

%q\_net.4.bias : Float(4, strides=[1], requires\_grad=1, device=cpu)):