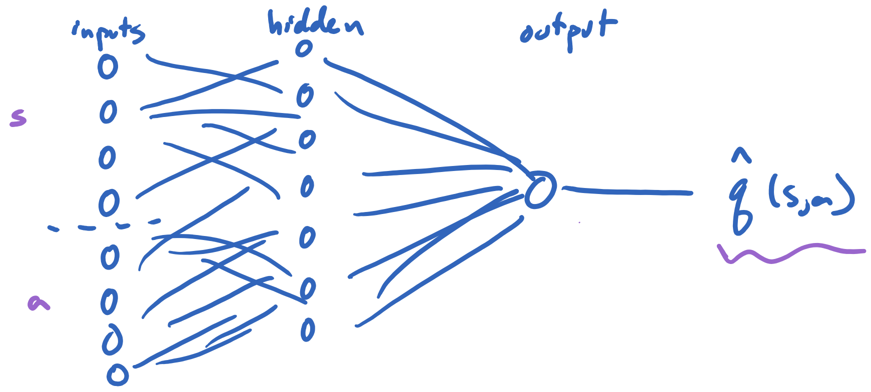
Deep Q Learning



$\hat{q}((80,4,10,24),0) = .38$

$\hat{q}((80,4,10,24),1) = .54$

$\hat{q}((80,4,10,24),2) = .21$

$\hat{q}((80,4,10,24),3) = .04$

reward 0   r

new state (64,4,10,22)   s'

$v(s') = \max_{a'} q(s',a')$

$\hat{v}(s') = \max_{a'} \hat{q}(s',a')$

$= 0.68$

from target network

training data

x   y

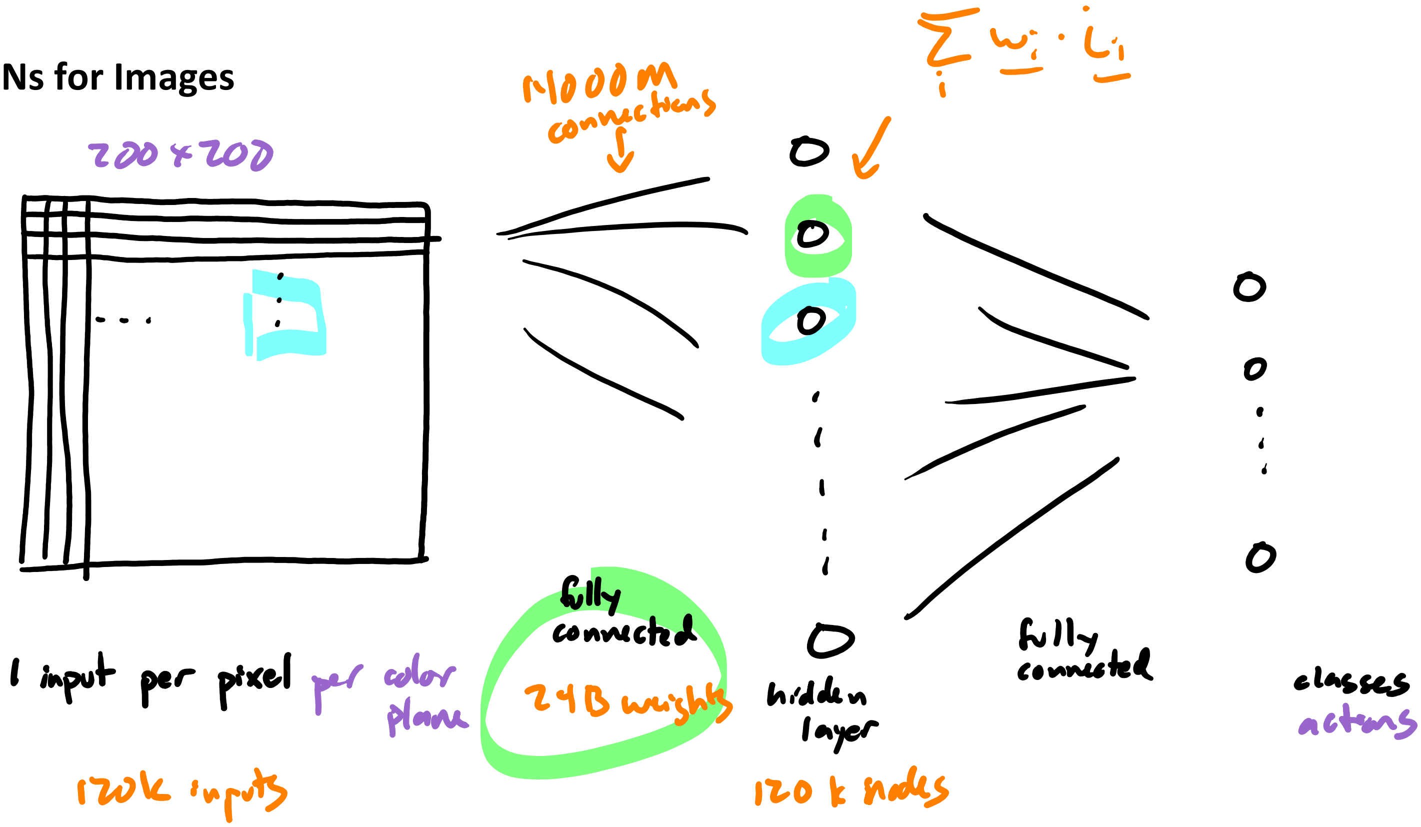
(80,4,10,24), 1   0.68

adjust weights on this

produces target output

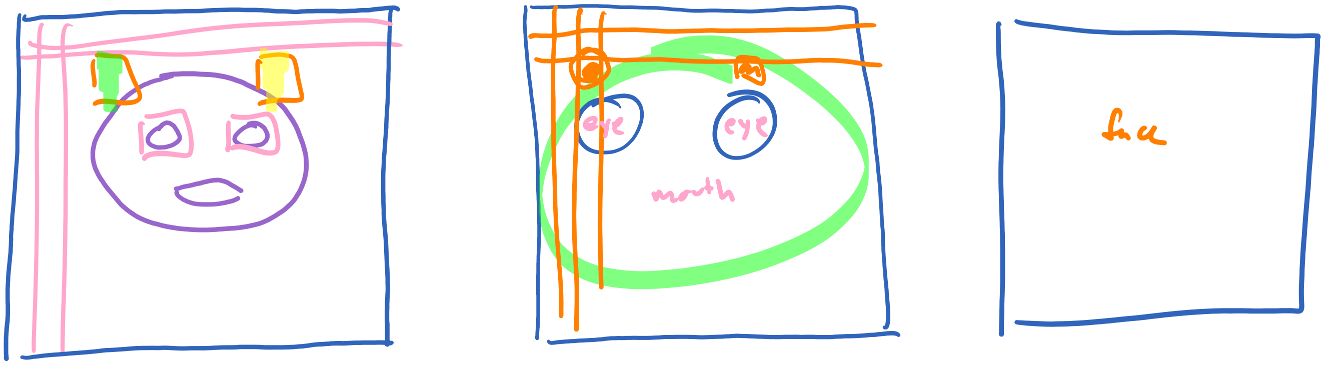
```
initialize learning, target networks
for each iteration
    for each of n episodes
        for each event
            add (s,a,s',r) to replay database
    sample replay database
    train learning network toward $r + \max_{a'} \hat{q}_{target}(s',a')$

    if enough time passed
        copy learning network to target network
```

ANNs for Images

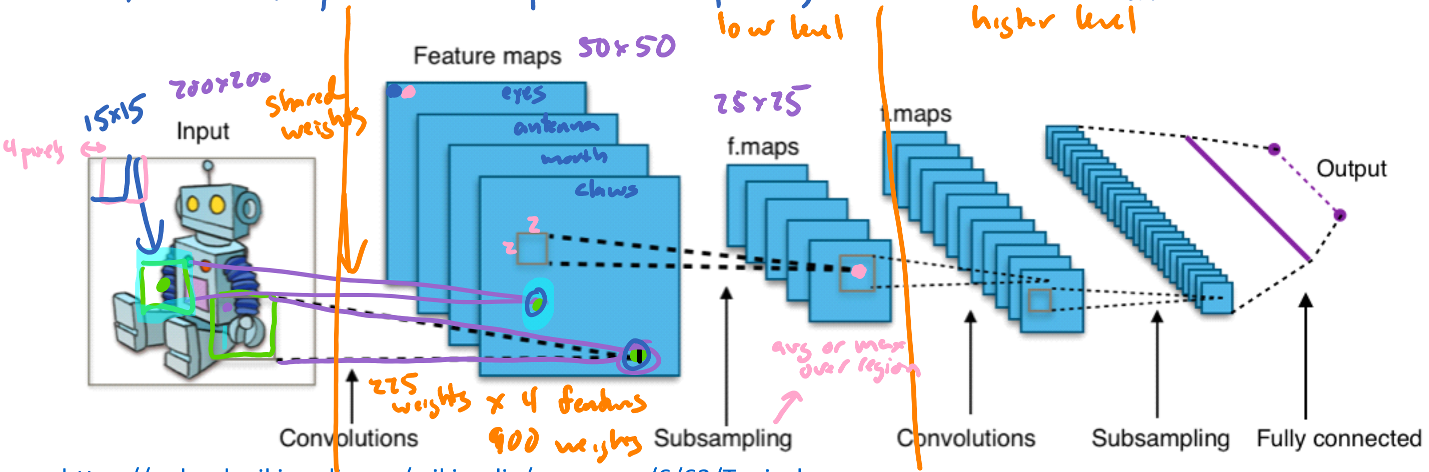


Convolutional Neural Networks

Deep Q network learning to play Pong



A much better picture from wikipedia user aphex 34, who does not endorse these notes.



https://upload.wikimedia.org/wikipedia/commons/6/63/Typical_cnn.png

SIMD
single instruction
multiple data

AlphaGo (2014-2017) DeepMind

Step 1: supervised learning for convolutional deep neural network

3 weeks

use database from games of expert players

13 layers

input: 19x19x48
locations features

- black
- white
- empty
- # opp captured
- # own captured

hand-coded features
- liberties
- ladder capture
- ladder escape

output: a move (19x19 + 1)

- matched 55% of time

+ smaller (faster) 75% of time

Step 2: reinforcement learning for convolutional deep neural network

1 day

beat SL network 80% of time

Step 3: reinforcement learning for valve network

+1 black win
⋮
-1 white win

Step 4: MCTS

Elo → →

2015 (Fan Hui) 2016 (Lee Sedol) 2017 (retired)

Δ Elo → higher rated player has chance of winning