

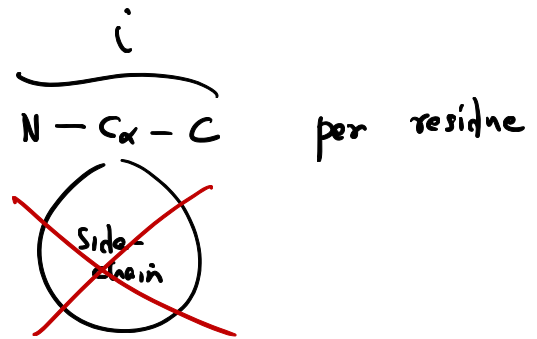
AlphaFold 1

Sidechainnet → proteinnet → PDB + CASP

CASP7
 ↪ blind test set
 PDB snapshot before CASP7

training - 30 valid - 10

only has x, y, z co-ords of



↪ includes C $_{\beta}$ as well as all sidechain atoms

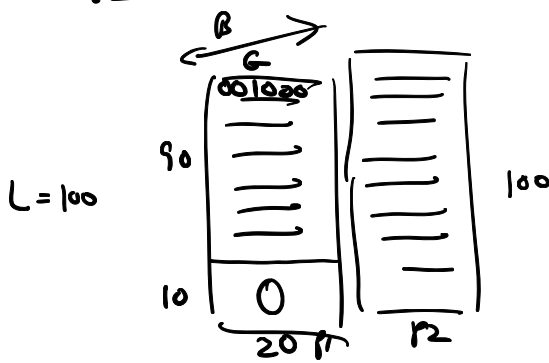
Sidechainnet Data

B: batch of proteins

B = 1
 dynamic = False

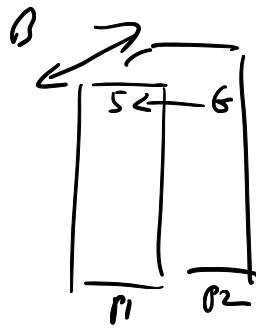
Seqs → one-hot sequence encoding

$B \times L \times 20$ or $B \times 20 \times L$



B = 2

int_seqs
B x L



We int_seqs as inputs
for creating embeddings for
each amino acid

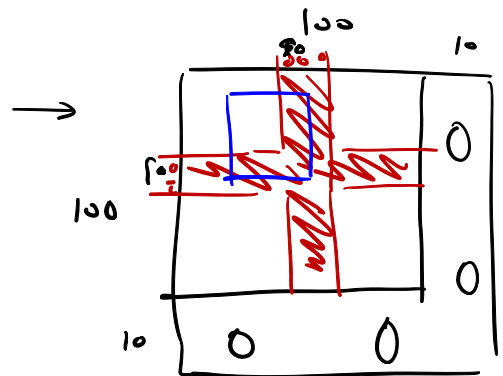
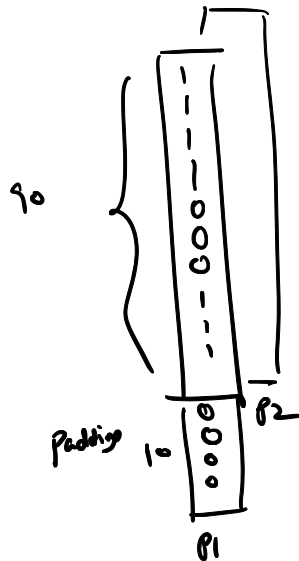
str_seqs

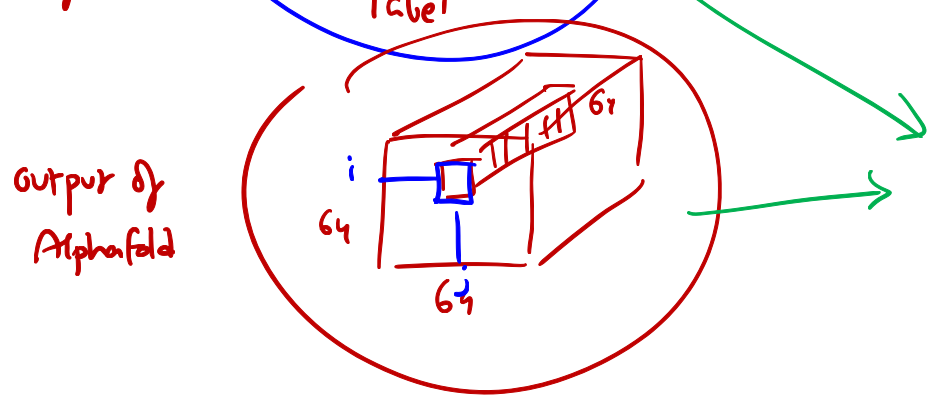
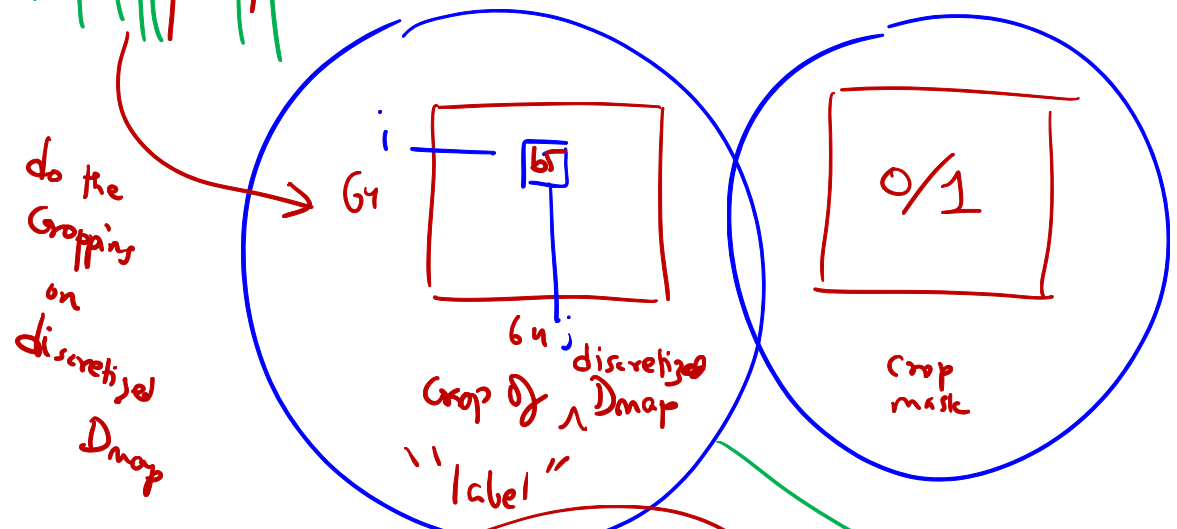
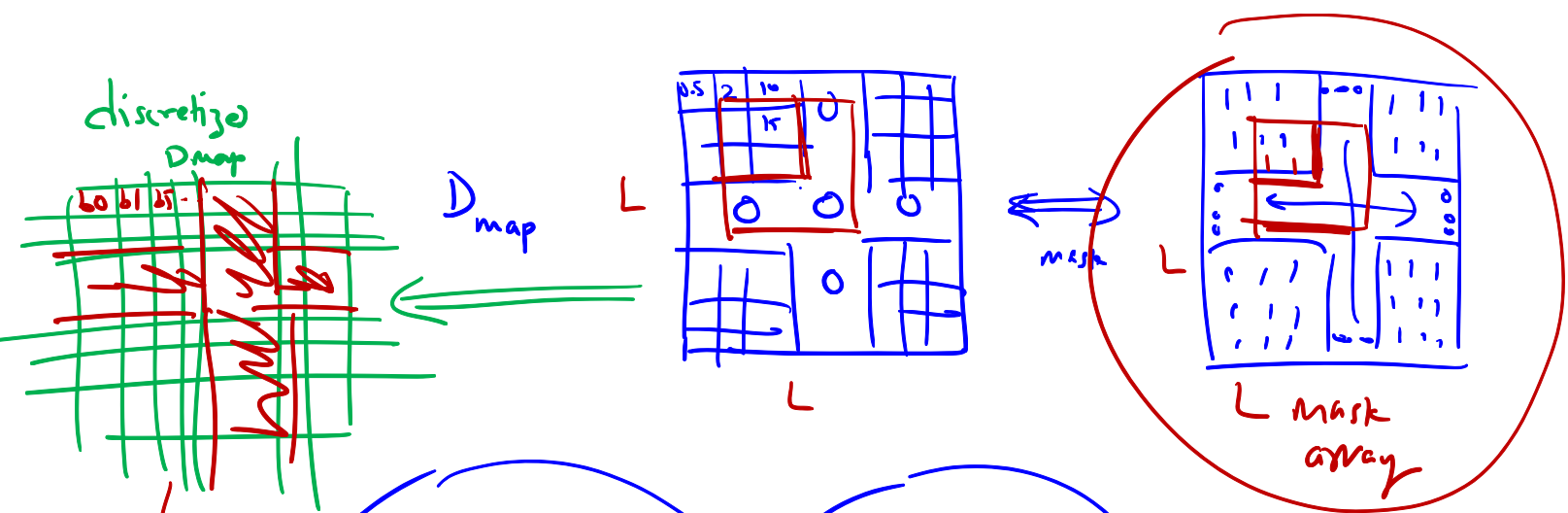
B x L



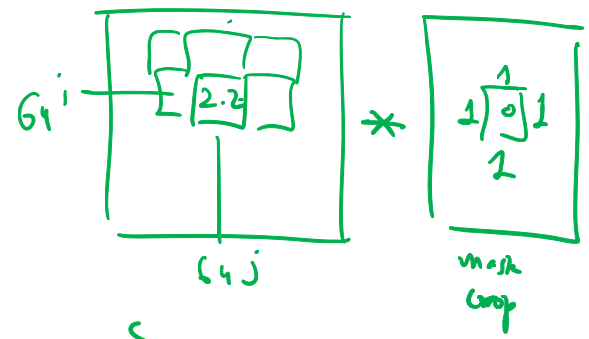
actual symbol

masks → mask per position in each seq in the batch





CE with logits
Reduction = 'none'



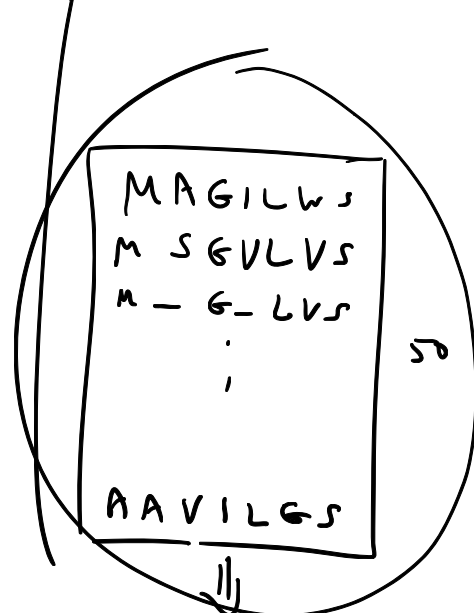
loss = Sum after $*$

ev(s)
 ↙
 evolutionary
 PSSM

$B \times L \times (20+1)$

Real tensor

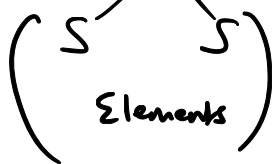
20 diff AA
 1 info content



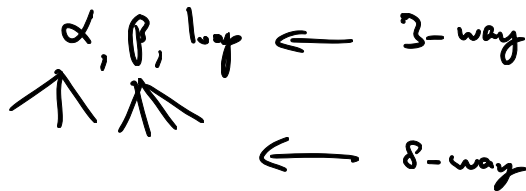
M	0.95		
A	0.05		

L

~~secondary structure (secs):~~



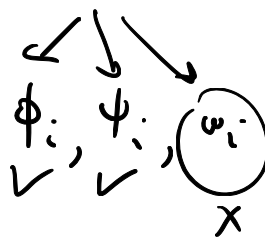
8-way labeling of SSE info



torsion

Angles (angs)

$B \times L \times 3$



Real (radians)

$(-\pi, \pi)$

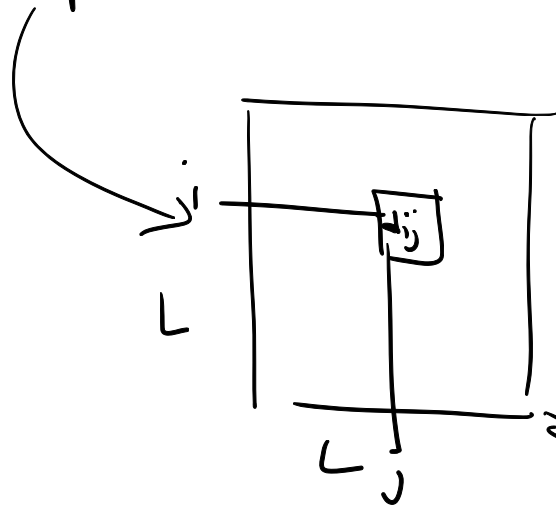
1296 bins!

Coords:

$B \times L \times 13 \times 3$

x, y, z coords of all atoms

$C_{\beta}^i - x^i, y^i, z^i$ for all except G which uses C_{α}



$L \times L$ Dmap of $(d_{ij})_{i,j=1}^L$

discretize

\downarrow
 One-hot seq $B \times L \times 20$
 ever $B \times L \times 21$
 mask (seq) $B \times L \times 1$
 Dmap ($B \times L \times L$)
 Dmask ($B \times L \times L$)

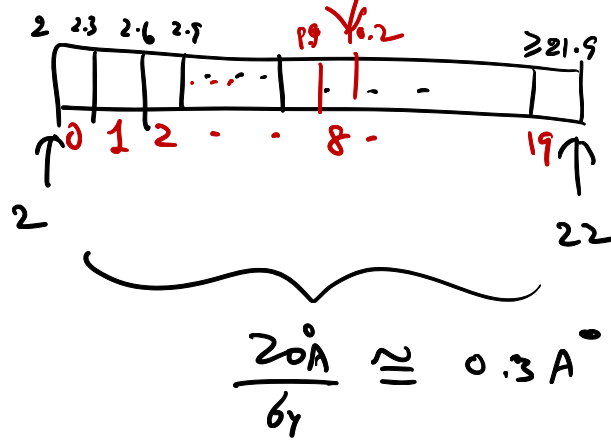
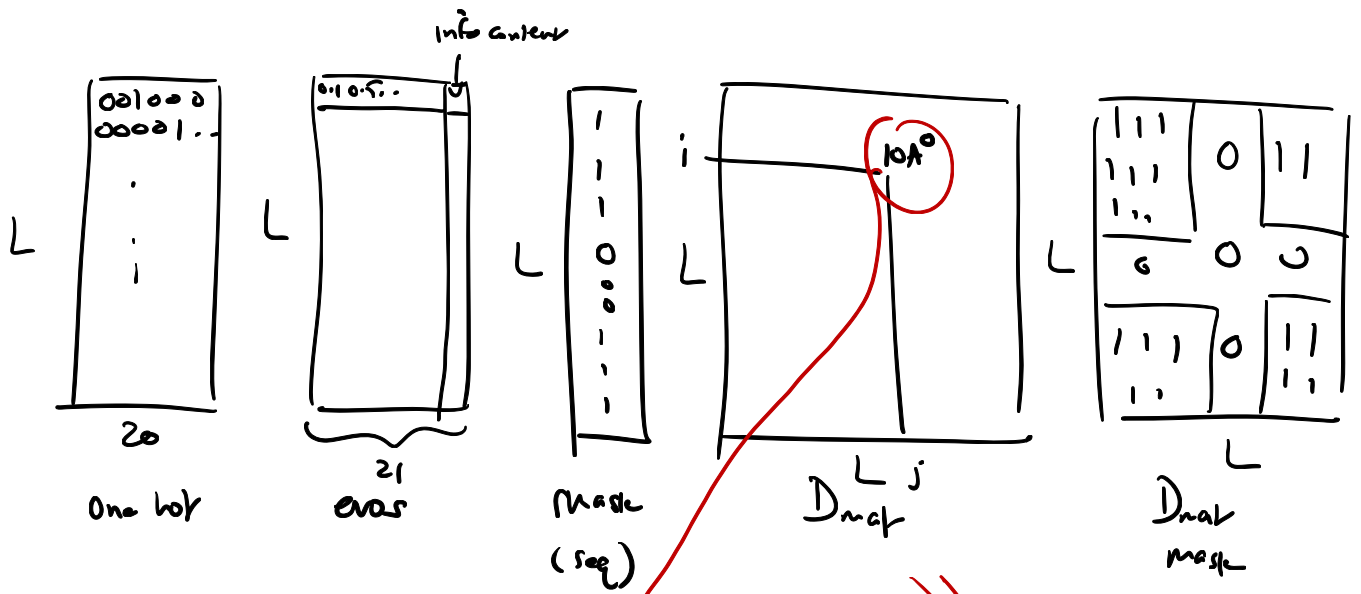
\downarrow
 0 1 2 3 4 5 6 7 10 1
 M A G I U V S P . . . G
 M V G L . . .

[4] scalar
 200000100
 . . . 0

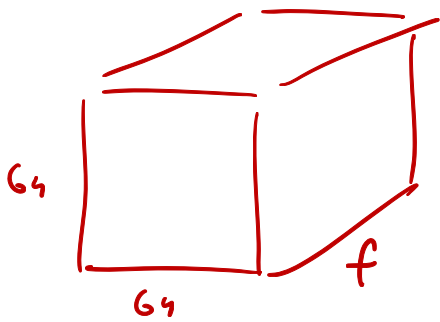
pos ending for 4th position

4 0 0 1 0 0
 - - - - -

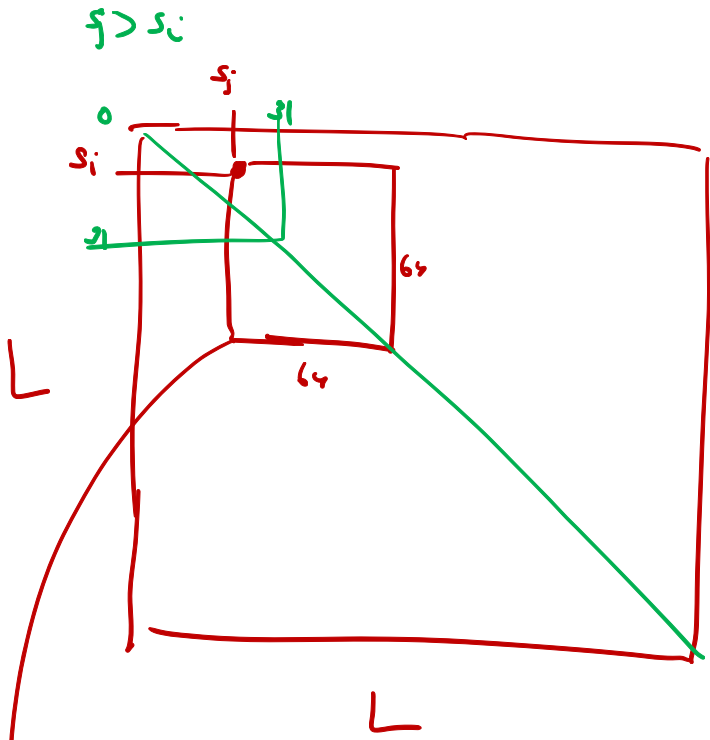
101
 64 22 16 8 4 2 1
 |



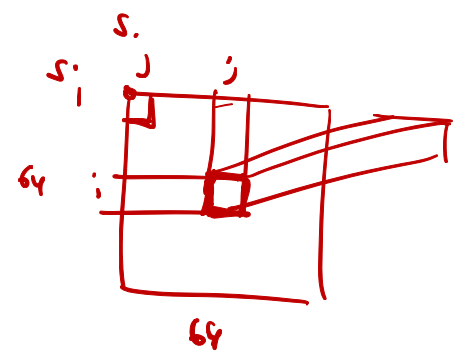
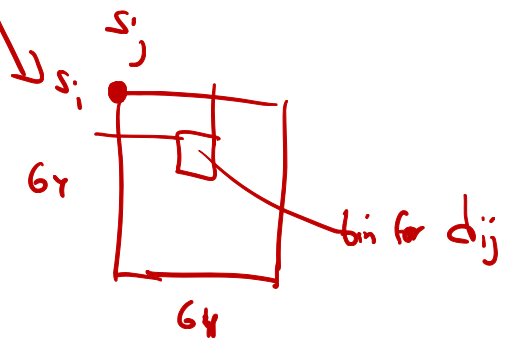
D_{mat} -discretized
'symbolic' bin labels



tensor input for CNN



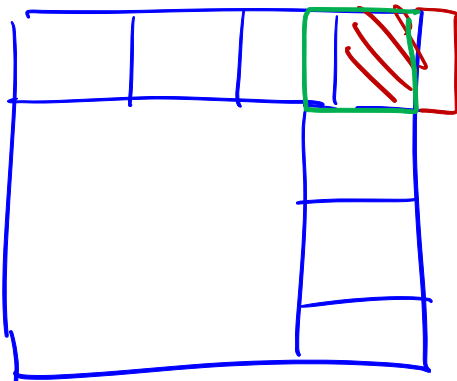
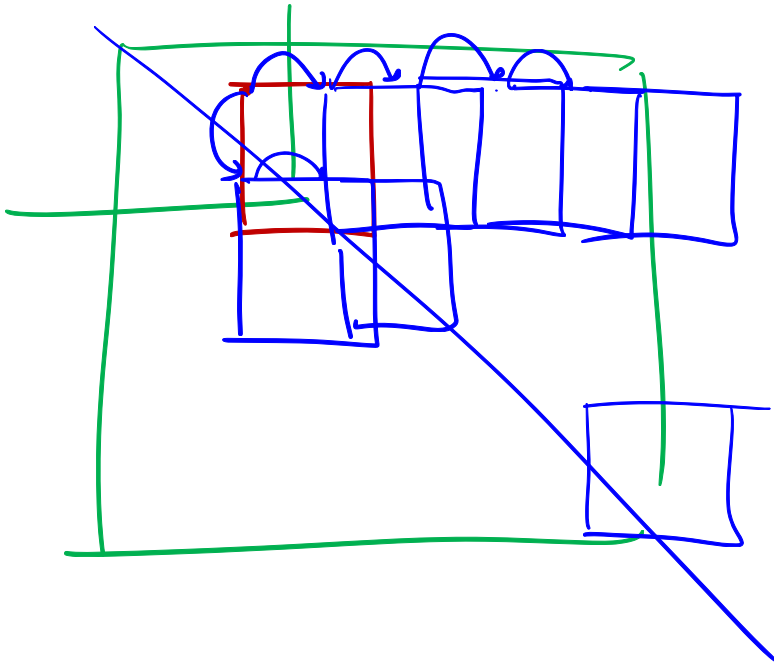
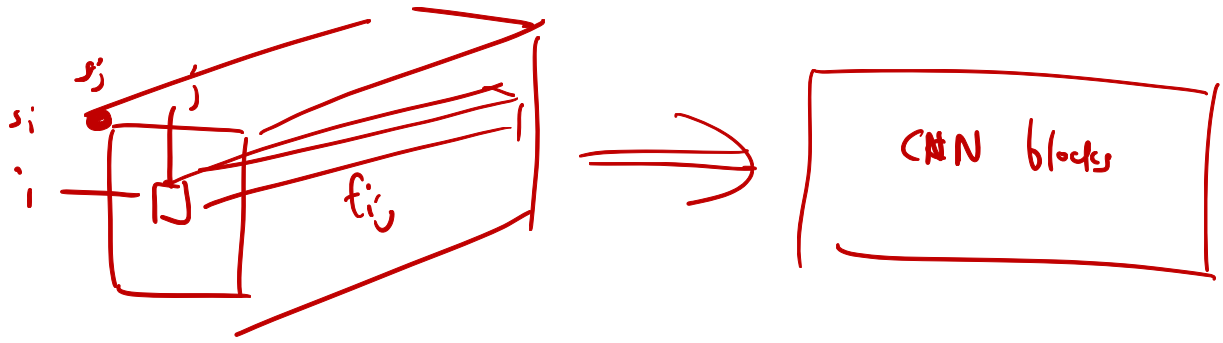
Dmcr-discrete

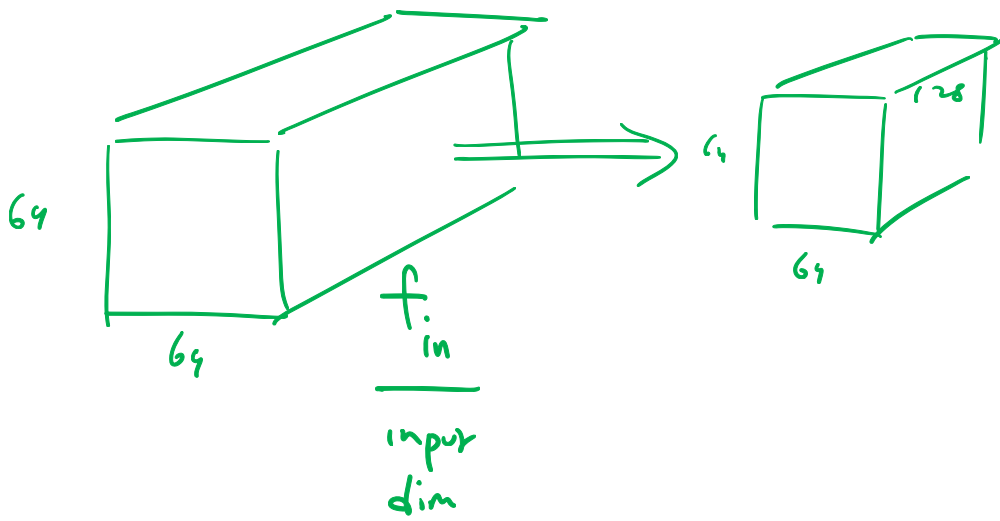


$$f_{ij} = \text{One-hot}_i \parallel \text{evos}_i \parallel \text{one-hot}_j \parallel \text{evos}_j \parallel$$

$$|\text{evos}_i - \text{evos}_j| \parallel \text{evos}_i \odot \text{evos}_j \parallel$$

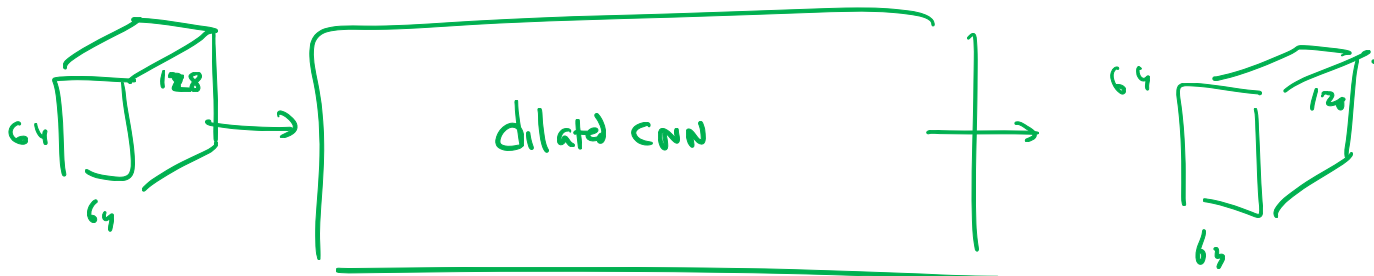
\parallel : Concat
 \odot : element wise product





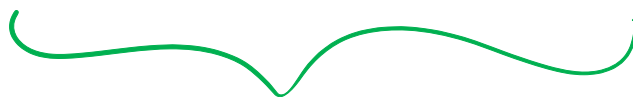
$$2d\text{-conv} (1 \times 1, \frac{f_{in}}{\text{in-channels}}, \frac{f_{out}}{128})$$

...



$d=1$
 $d=2$
 $d=4$
 $d=8$

in succession



dilata = d
 padding
 3x3 kernel