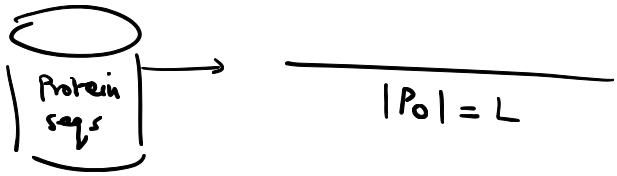
Alphafold 2

Training process



Create crops of size 256

(pick random start pos per sequence)

↳ a) use sliding window to cover the protein

↳ b) use a single random crop per epoch

Nres = 256 for training

1) Evoformer

↳ M tensor N_{clust} c_m Nres Nres

↳ Z tensor Nres Nres c_z

Multiple blocks

c_s Nres S matrix

2) Structure module (S, Z) ← bias

8 layers

IPA

Backbone Update

Angle prediction

L_{aux}^{l} = (FAPE + angle) loss

$L_{aux} = \underset{l=1..8}{\text{mean}} \{ L_{aux}^{l} \}$

∀ epochs

for each batch

1) Evoformer

2) structure

$L = L_{aux}$

L.backwards()

Optim. step()

Algo 2a

Ignore
- ↳ Compute all atoms
- ↳ rename symmetric
- ↳ 2e: FAPE not required for $\mathcal{L}_{aux}$ (Laux already has this)

Testing

$|P| = L$

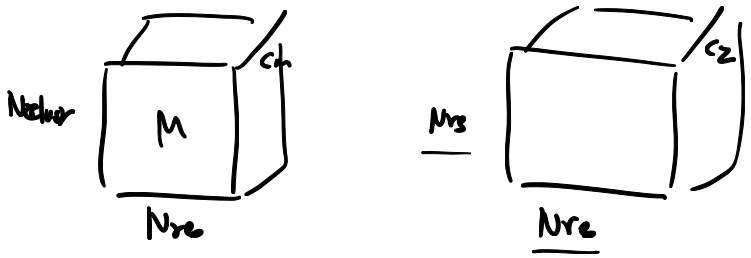
No cropping! (no restriction to 256 len crops)

$N_{res} = L$

↖ do not hard code leave it as N_{res}

Since all trainable parameters are linear projections of the last dim / feature dim

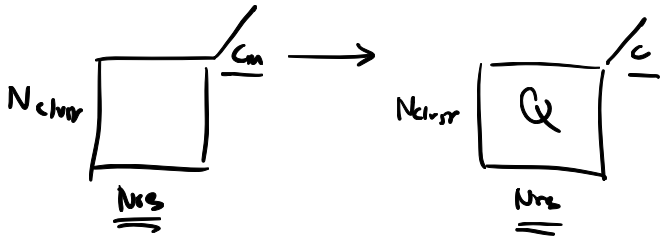
1) Row Attn:



attn: Q, K, V

also some gate

Q: linear ($\underline{M}$, bias = False)



$C_m \rightarrow C$ mapping

$A = QK^T$

$N_{res} \times N_{res}$

256x256 ← train

LxL ← test

testing is done on full protein in one shot

$\forall$ proteins in test set

$\forall\ i = 1 \ldots L$

$C_i^{\alpha} = (x, y, z)$ coordinates

print out the loss per test protein

(FAPE + torsion angle)

for one protein at a time

batch = 1

---

avg loss across proteins

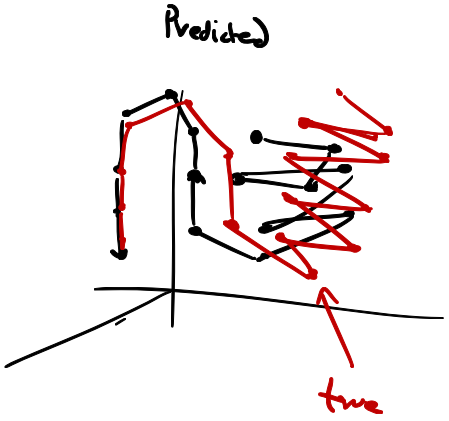
plot the predicted coords

$P_i \leftarrow$ test protein i

↳ pick one with low loss

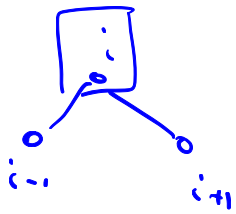
medium ~

high λ



Alphafold 2 : limitations

1) Using attn to model dependence between diff positions in a protein
2) however, each position (x, y, z) $(\alpha_\phi, \alpha_\psi \ldots)$ independently predicted



dependent prediction

Conditional prediction

$$\left((x,y,z)_i \mid (x,y,z)_{i-1}, (x,y,z)_{i+1} \right)$$

Systems biology

→ Network of metabolite, protein, gene

→ gene interaction networks

→ signalling

→ transcription

→ genomics

↪ proteomics

↪ drug discovery

→ etc.

Please complete the course eval

↪ leave comments

↪ 100% response