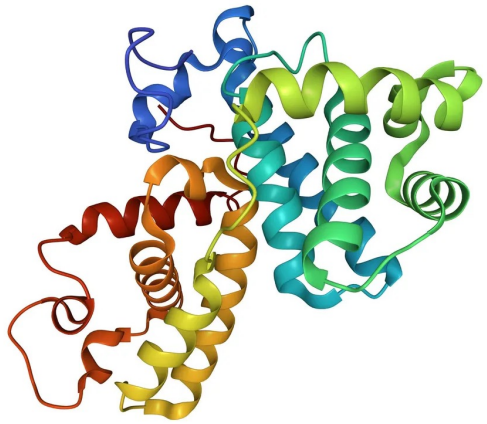
POINTLISTNET: DEEP LEARNING ON 3D POINT LISTS

Hehe Fan[1,2], Linchao Zhu[1], Yi Yang[1]
Mohan Kankanhalli[2]

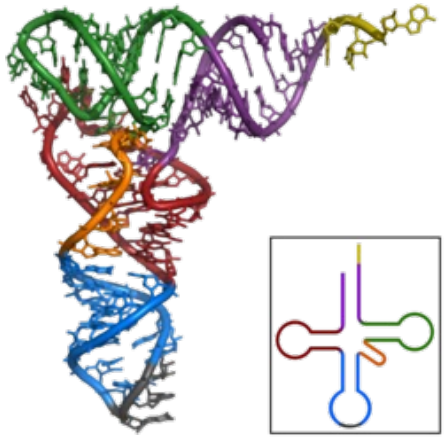
浙江大學
ZHEJIANG UNIVERSITY

NUS
National University
of Singapore

PointListNet: Deep Learning on 3D Point Lists

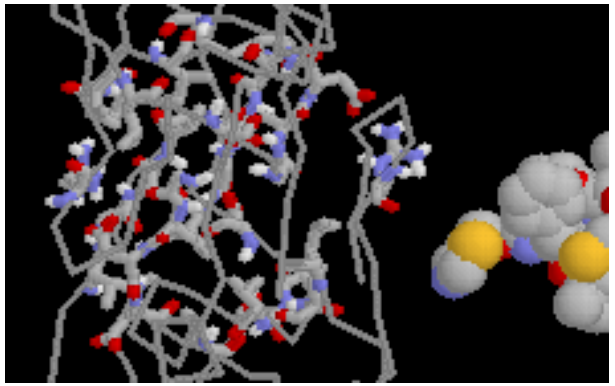


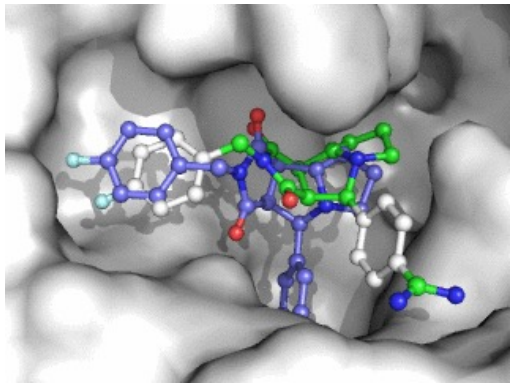
Protein

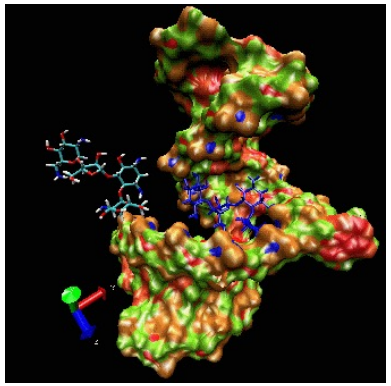


Non-coding RNA

1. *Proteins are made up of amino acids, which are linked by peptide bonds and form a chain, i.e., a list.*
2. *A protein functions only when it folds into a 3D shape.*
3. *Once we recognize its structure, we can know its function, attributes …*

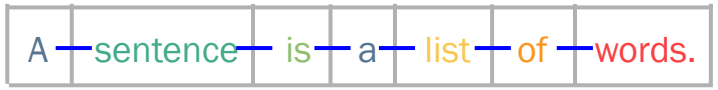
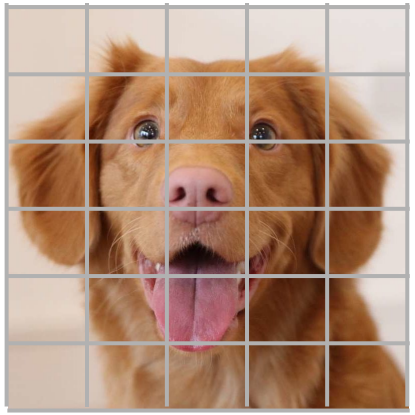
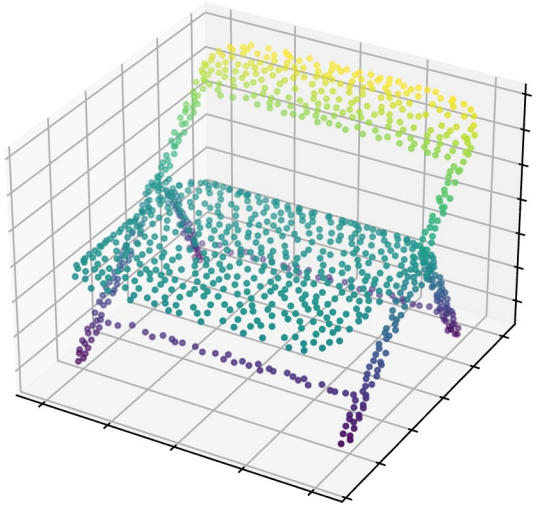
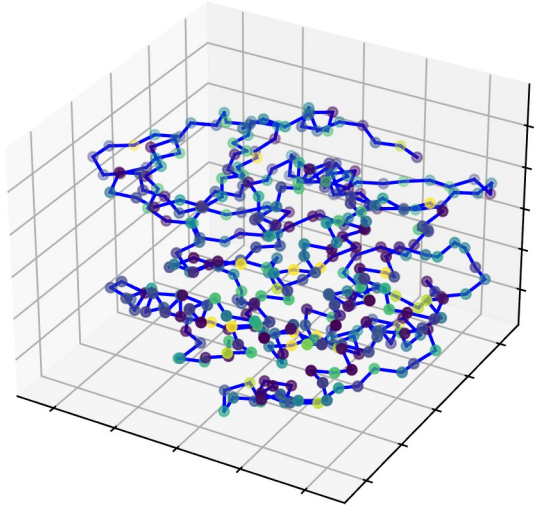






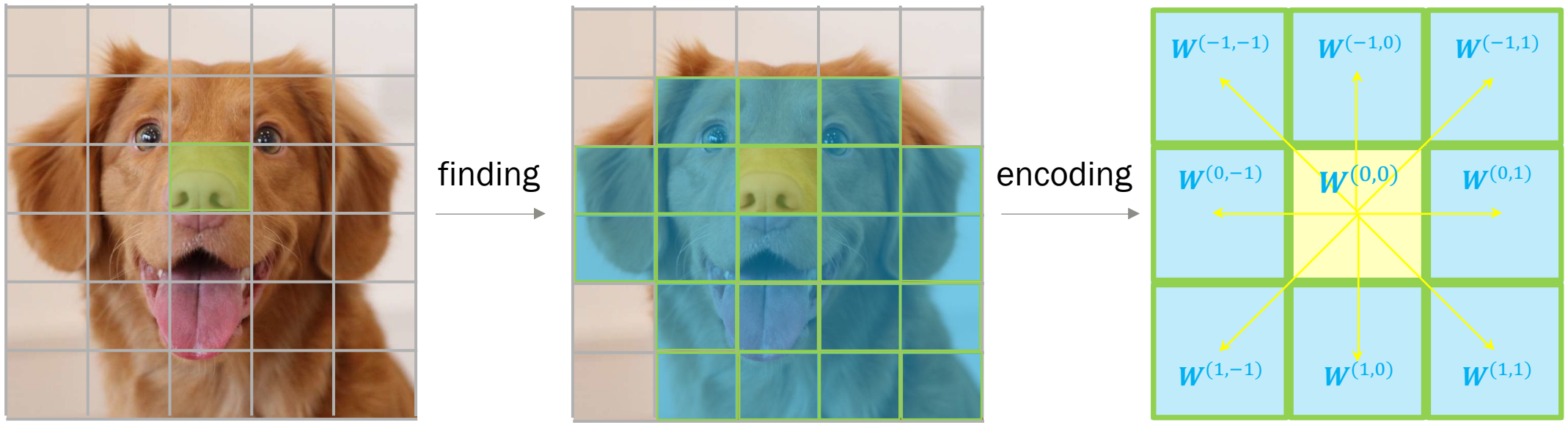
1. *We investigate a range of point cloud methods for protein modeling.*
2. *We propose a Transformer-style network, i.e., PointListNet, for 3D point list modeling.*
3. *Experiments on two protein classification tasks show the effectiveness PointListNet.*

Deep Learning on Different Kinds of Data

text: 1D word list	image: 2D pixel grid	3D point cloud/set	protein: 3D point list
			
Texts are regular 1D lists of words. The position is the word's sequential order in the text and the feature is the word itself.	Images are regular 2D grids of pixels. The position is the row and column where the pixel is located and the feature is the color.	Point clouds are irregular 3D point sets. The position is the 3D coordinate and the feature is the point attributes.	Proteins can be seen as 3D point lists. The position of an amino acid involves a regular 1D sequential order and an irregular 3D coordinate. The feature is the amino acid (residue) type.

How to Model Structure?

- When a deep neural network tries to model the struture of a certain kind of data, it includes two setps:
  - *Finding related regions.*
  - *Encoding related regions.*



PointListNet

- Finding related points.
  - *When employing Transformer for protein modeling, we find that it may be the 1D and 3D distances, instead of amino acids' features, that mainly determine amino acids' correlations.*
  - *We replace the self-attention in the vanilla Transformer with a non-parametric distanced-based attention mechanism.*

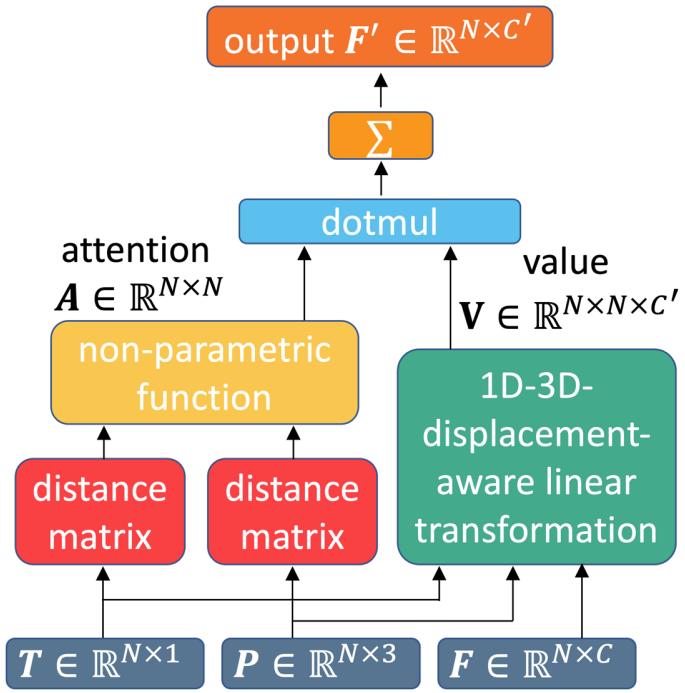
$$\boldsymbol{A} = g\left(\frac{\boldsymbol{D}^{1d}}{L}, \frac{\mathbf{D}^{3d}}{R}\right)$$

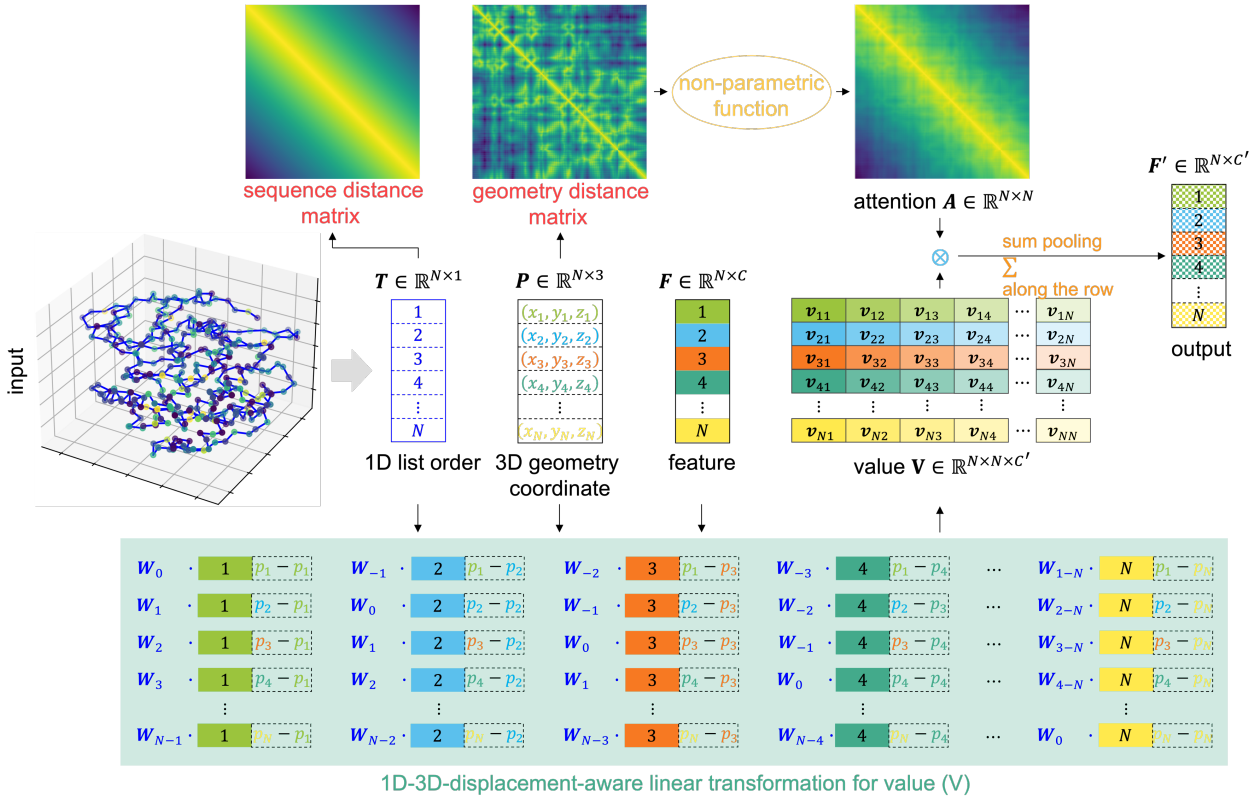
$\boldsymbol{D}^{1d}$ *is the distance matrix of 1D point orders in the peptide chain.*

$\boldsymbol{D}^{3d}$ *is the distance matrix of 3D point cooridinates in Euclidean space.*

L *and* R *are the longest sequence and geometry distances from the center to the farthest points, respectively.*

PointListNet





- Encoding related points.
  - *We propose a 1D-3D-Displacement-Aware Value method to integrate relative structure modeling into Transformer.*
  - *The 1D-3D-Displacement-Aware Value method employs regular and irregular methods to capture the sequence and geometry structures, respectively.*
  - *Rotationally invariant.*

Experiments

- We evaluate the proposed method on two recognition tasks: protein fold classification and enzyme reaction classification.
- Comparison with point cloud methods.

Method	Modeling	Protein Fold Classification			Enzyme Reaction
		Fold	Superfamily	Family	Classification
PointNet++	$f'_t = \underset{\|p_{t'}-p_t\|\le r}{\text{MAX}} \text{MLP}([f_{t'}, p_{t'}-p_t])$	26.0	37.7	93.8	78.4
DGCNN	$f'_t = \underset{f_{t'}\in \text{TopK}(f_t)}{\text{MAX}} \text{MLP}([f_{t'}, f_{t'}-f_t])$	25.6	39.2	94.4	80.1
Point Transformer	$f'_t = \sum_{\|p_{t'}-p_t\|\le r} \alpha_{tt'} \times (W_3 \cdot f_{t'} + \delta_{tt'})$ $\delta_{tt'} = \text{MLP}(p_t - p_{t'})$ $\alpha_{tt'} = \text{MLP}(W_1 \cdot f_t - W_2 \cdot f_{t'} + \delta_{tt'})$ $\{\alpha_{tt'}\} = \text{softmax}(\{\alpha_{tt'}\}_{\|p_{t'}-p_t\|\le r})$	26.4	40.1	92.0	81.3
PointMLP	$f'_t = \text{MLP}(\underset{f_{t'}\in \text{TopK}(f_t)}{\text{MAX}} \text{MLP}(f_{tt'}))$ $f_{tt'} = \alpha \odot \frac{f_{t'}-f_t}{\delta+\epsilon} + \beta$ $\delta = \sqrt{\frac{1}{K\times N\times C}\sum_{i=1}^{N}\sum_{j=1}^{K}(f_{t'}-f_t)^2}$	26.8	38.8	94.2	79.7
PointListNet (ours)	3D Coordinate	36.8	55.3	97.4	84.5
	1D Order & 3D Coordinate	**55.2**	**76.4**	**99.5**	**88.0**

- Because PointListNet models both 1D and 3D strucutre, it significantly outperforms those point cloud methods.
- Even only with 3D geometry structure, PointListNet still significantly outperforms the point cloud methods.

Experiments

- Comparison with the SOTA methods.

Level	Input	Method	Protein Fold Classification			Enzyme Reaction
			Fold	Superfamily	Family	Classification
Atom Level (Molecule)	3D Coordinate	GCN	16.8	21.3	82.8	67.3
		EdgePool (GNN)	12.9	16.3	72.5	57.9
		3D CNN	31.6	45.4	92.5	72.2
	3D Coordinate & Bond	IEConv	45.0	69.7	98.9	87.2
Residue Level (Protein)	1D Order	1D ResNet	10.1	7.21	23.5	24.1
		DeepFS (1D ResNet)	17.0	31.0	77.0	70.9
		DeepFS (1D CNN)	40.9	50.7	76.2	-
		LSTM	6.41	4.33	18.1	11.0
		Transformer	9.22	8.81	40.4	26.6
	3D Coordinate	GAT	12.4	16.5	72.7	55.6
	1D Order & 3D Coordinate	GraphQA	23.7	32.5	84.4	60.8
		GVP	16.0	22.5	83.8	65.5
		IEConv	47.6	70.2	99.2	87.2
		GearNet	28.4	42.6	95.3	79.4
		GearNet-IEConv	42.3	64.1	99.1	83.7
		GearNet-Edge	44.0	66.7	99.1	86.6
		GearNet-Edge-IEConv	48.3	70.3	**99.5**	85.3
		PointListNet (ours)	**55.2**	**76.4**	**99.5**	**88.0**

- PointListNetsignificantly outperforms those SOTA methods

Summary

- Among the early efforts, we investigate a range of point cloud methods for protein modeling.
- We propose a Transformer-style network, i.e., PointListNet, for 3D point list modeling.
- We replace self-attention with non-parametric distance-based attention, which is more efficient and effective to achieve the correlation among microparticles in some cases.
- We integrate relative structure modeling into Transformer and employ regular and irregular methods to capture the sequence and geometry structures, respectively.
- We conduct extensive experiments on two protein tasks and the proposed method significantly outperforms existing methods.