

XPSI: X-ray Free Electron Laser-based Protein Structure Identifier Paula Olaya*, Michael R. Wyatt II*, Silvina Caino-Lores*, Florence Tama⁺, Osamu Miyashita⁺, Piotr Luszczek*, Michela Taufer*



*University of Tennessee, Knoxville, USA ⁺RIKEN, Japan

orkflow		Platform		Autoencode			ler Tr	er Training	
for analysis of XFEL and conformation r protein diffraction		Executed on a single node of the Tellico cluster • 32-core Power9		Test	Prediction	Data size	Time [s] (over 20 trials)	Test 1: It takes 45 mins to train the autoencoder with 39692 samples Test 2: It takes 90	
-	1			Orientation [φ,Θ]	39,692	2709.8			
	RAM) • 2 x Nvidia V100		2	Orientation + conformation $[\phi,\Theta, conf]$	79,384	5415.6	mins to train the autoencoder with 79384 samples		
0	kNN Training								
	Test Prediction		Data	Over 20 trials			$\mathbf{T}_{\mathbf{a},\mathbf{c},\mathbf{t}} 1_{\mathbf{b}} 0 0 7_{\mathbf{c},\mathbf{t},\mathbf{c}} 1_{\mathbf{c}}$		
ining			size	Training time [s] (90%)	Validati [s] (10%	ion time 6)	train on 90% of the data, and 0.10 s to		
Prediction	ons: ion iation	1	Orientation $[\phi, \Theta]$	39692	0.07	0.10		validate on 10% Test 2: 0.34 s to train on 90% of the data, and 0.66 s to validate on 10%	
 Orientation Conformat 		2	Orientation + conformation $[\phi,\Theta, conf]$	79384	0.34	0.66			
Fror, Res	sults	5, al	nd Less	ons	Learne	d			
		Validation Errors (20 trials)					Me	etrics and Results	
Intensity High Intensity				Low Intensity			Frror [Frror Degree: The distance	
Selected K=5	99% of data: 99% of data: error < 2.2° • • • • • • • • • • • • • • • • • • •			8 - 8 - 9 6 - 9 6 - 9 6 - 9 6 - 9 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	99% of da error < 2	ata: 2.2°	in degination of the second se	in degrees between two points on a sphere (ϕ , Θ) Conformation accuracy: Th proportion [%] of correct predictions among the tota number of cases examined	
- Median 9 10 11 12 13 14 15 16 17 18 19 20 K		Image: Constraint of the second state of the seco		Image: Contract of the second state			For Hi Test 1	For High and Low intensity Test 1: 99% of the data	
Selected K=3 999.5 990.5 990.5 990.5 990.5 990.5 990.5 990.5 990.5 990.5 990.5 990.5 990.5 990.5 990.5 990.5 900.5 1.2.3.4		99% of data: error < 2.2°		 100 00 00	$\begin{array}{c} 100 \\ 99.5 \\ 99 \\ 99 \\ 99 \\ 99 \\ 99 \\ 99 \\ $			 have an error degree with 2.2° Test 2: 99% of the data have a error degree within 2. 100% of the data have conformation accurace of 100% 	
								Acknowledgem	



he

nin

In collaboration with the JLESC Consortium (https://jlesc.github.io/) Supported by JDRD Program at UTK and IBM Shared University (SUR) Award