Peak finding for crystallography

Po-Nan Li
Department of Electrical Engineering
Stanford University

Chun Hong Yoon
Linac Coherent Light Source
SLAC National Accelerator Laboratory

Motivation

- Peak finding is an essential part for crystallography to determine the crystal parameters and the structure of the sample
- Bragg peaks are pixel-like features that are sometimes very faint and difficult to be identified even by a human expert
- Current peak finders require a number of user–defined parameters that vary from experiment to experiment

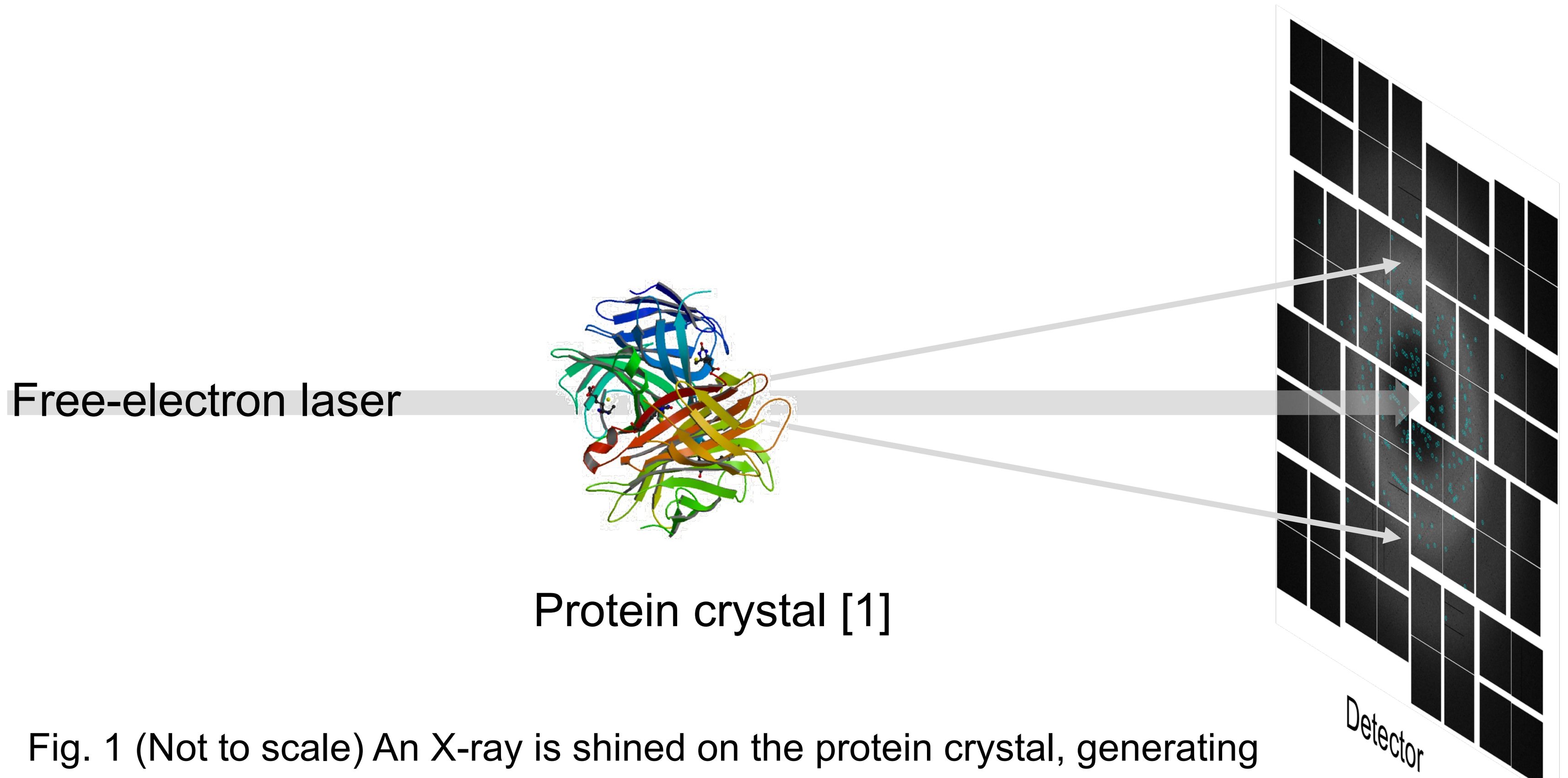

Fig. 1 (Not to scale) An X-ray is shined on the protein crystal, generating discrete and strong Fourier patterns, namely Bragg peaks

Data

- 200K+ diffraction images from CXIC0415 [2], a serial femtosecond crystallography (SFX) experiment conducted at LCLS's CXI beamline
- 0.5% for validation, 0.5% for test; rest for training

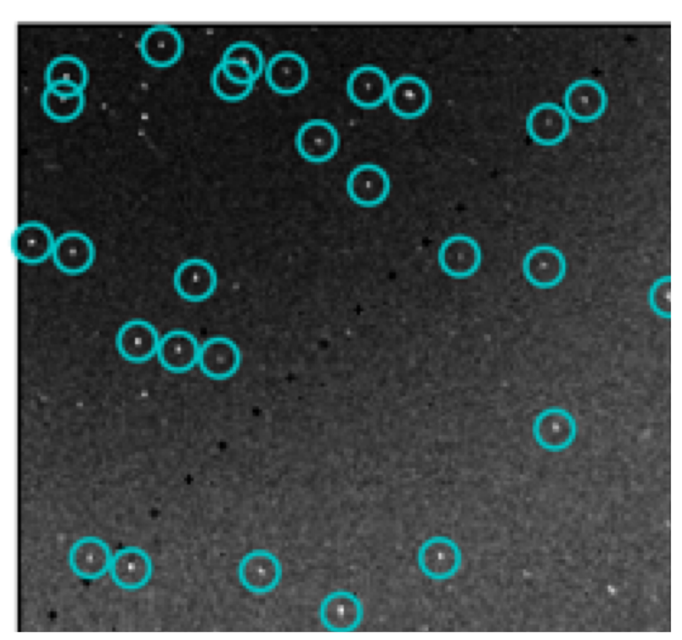

Fig. 2 Bragg peaks (bright pixels) and labels (cyan circles)

Subset	Train	Val	Test	Total
# Diffraction images	199,675	1,000	1,000	201,675
# Bragg peaks	12,498,810	63,543	63,917	12,626,270

YOLO

YOLO is a fast object detecting system that uses CNN as feature extractor and divides the image into grid cells for detections [3-5]. Each cell is responsible for detecting the object whose center falls within it and has maximum probability. YOLOv2 has 19 conv. Layers [4]. The latest YOLOv3 has 53 conv. layers and shortcuts [5].

Architecture

The original YOLOv2 backbone cannot detect Bragg peaks so we modified it by reducing the number of convolutional and maxpooling layers to preserve the features of Bragg peaks. We experimented with several custom architectures (PeakNet-X) and YOLOv3, which was released just two months ago [5]. PeakNet-Fat is finally chosen for its >0.9 precision at 0.5 sensitivity.

Type	Filters	Size	Output
conv	32	3x3/1	1552x1480x32
max		2x2/2	776x740x32
conv	64	3x3/1	776x740x64
max		2x2/2	388x370x64
conv	128	3x3/1	388x370x128
conv	64	1x1/1	388x370x64
conv	128	3x3/1	388x370x128
			388x370x192
conv	96	3x3/1	388x370x96
conv	30	1x1/1	388x370x30
detection			

conv
reorg

Fig. 3 Architecture of PeakNet-Fat

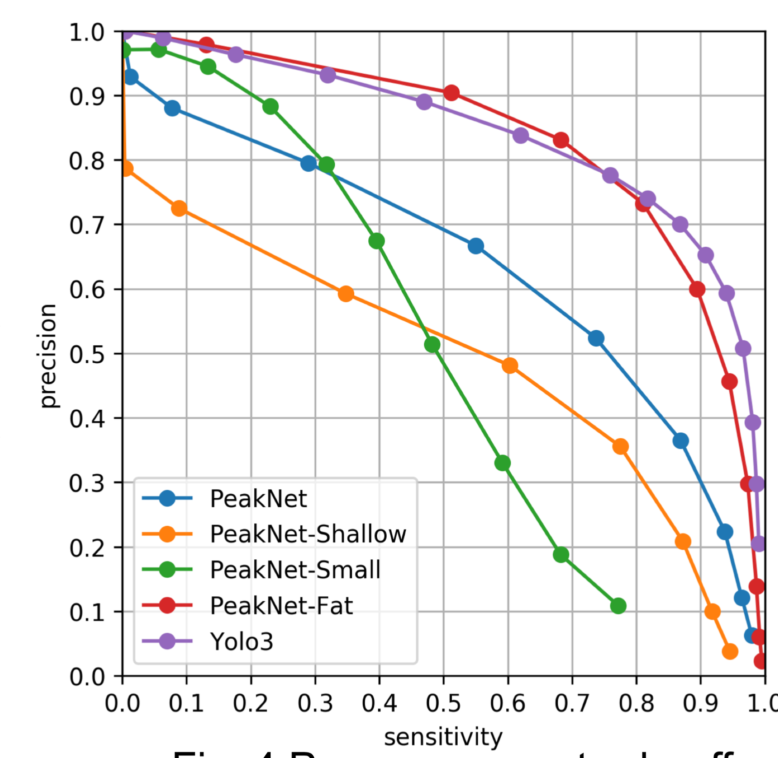

Fig. 4 Prec. vs. sens. trade-off

Arch	CONVs	BFLOPs	Time	Sens.	Prec.	*Dist.	*IOU
PeakNet	16	71.19	465	0.08	0.88	0.82	0.15
PNet-shallow	12	59.43	557	<0.01	0.79	1.08	0.18
PNet-small	16	9.83	347	0.13	0.95	0.75	0.23
PNet-Fat	8	119.46	760	0.68	0.83	0.82	0.30
YOLOv3	53	213.25	845	0.82	0.74	0.85	0.35

Metrics on validation data at threshold 0.25. Architecture; number of convolutional layers; billions of floating point operations per image; processing time per image in milliseconds; sensitivity TP/(TP+FN); precision TP/(TP+FP); average coordinate error; average intersection-over-union. *conditioned on IOU>0.0

Results

Fig. 5 Representative image from the test set. The diffraction image is inverted for visualization, so Bragg peaks are dark; bright indicates lower intensity. Cyan circles mark the labels; magenta boxes mark the detection results. Threshold = 0.25

Test stats
Sensitivity: 0.68
Precision: 0.83
Avg. error: 0.82
Avg. IOU: 0.30

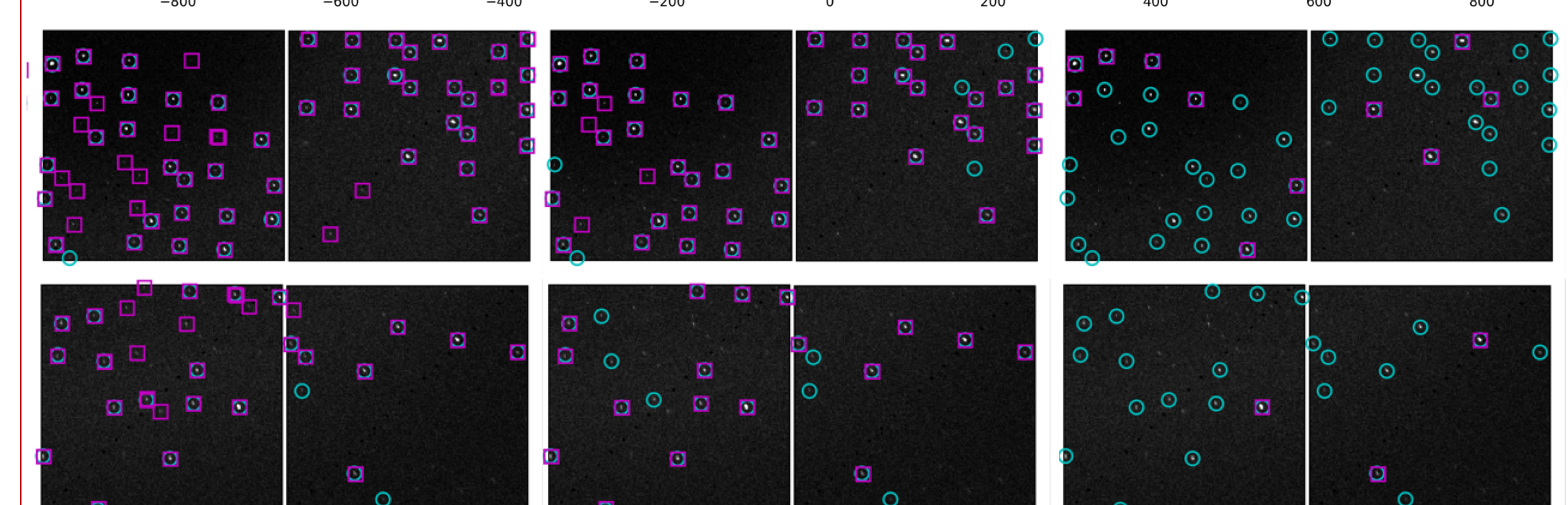

Fig. 6 Zoom-in of the yellow box in Fig. 5. Higher detection threshold gives better precision but lower sensitivity. Note that many "false positives" are actually Bragg peaks that weren't labelled.

[1] PDB ID: 5JD2
[2] M. S. Hunter et al., "Selenium single-wavelength anomalous diffraction de novo phasing using an X- ray-free electron laser," Nat. Comm. 7, 13388 (2016).
[3] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, "You only look once: Unified, real-time object detection," *arXiv:1506.02640* (2015).
[4] J. Redmon, A. Farhadi, "YOLO9000: Better, faster, stronger," *arXiv:1612.08242*, (2016).
[5] J. Redmon, A.Farhadi,,"YOLOv3: An Incremental Improvement" arXiv:1804.02767 (2018).
GitHub: https://github.com/leeneil/peaknet