Classification of diffraction patterns in Single Particle Imaging (SPI) experiments performed at X-ray free-electron lasers using a CNN

Alexandr Ignatenko

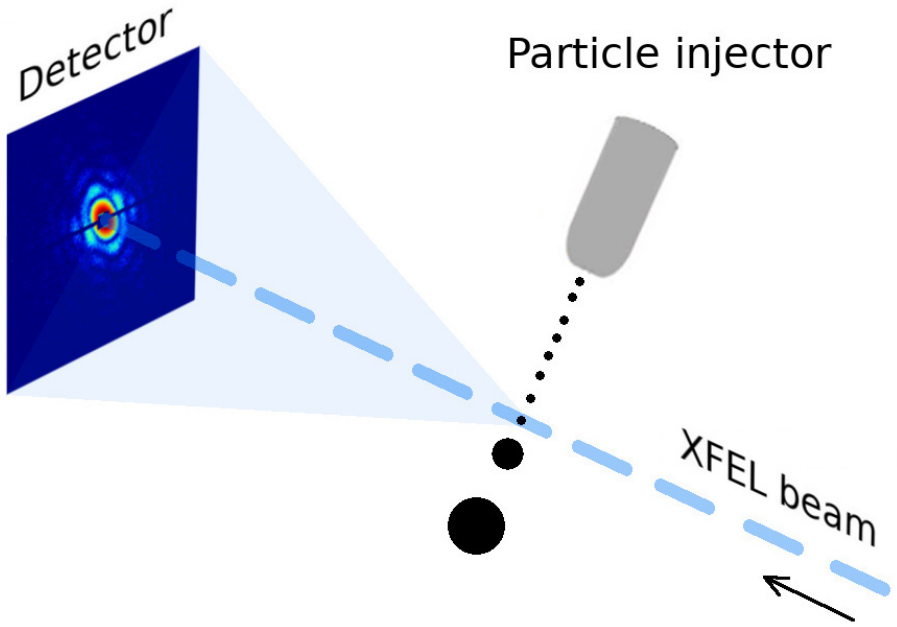
Hamburg, 17.12.2020

HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

RSF Russian Science Foundation

DESY.

SPI experiment

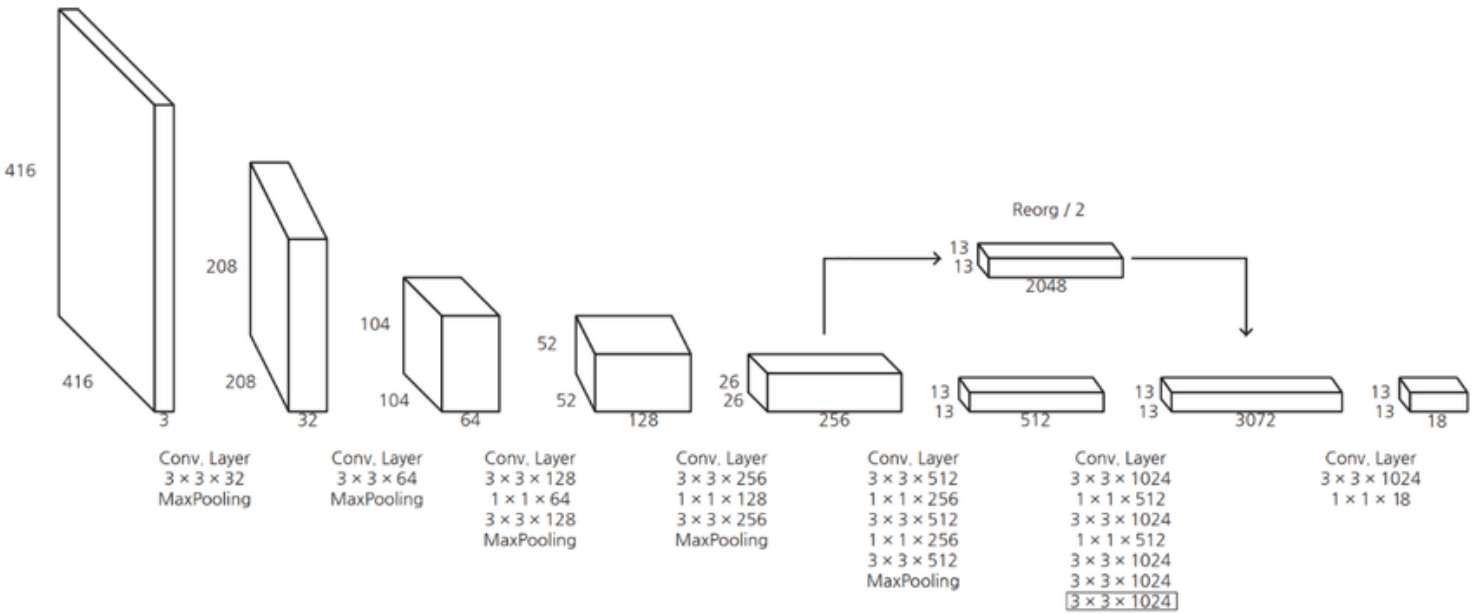


Data processing pipeline:

Pre-processing → Classification → Orientation determination & background subtraction → Phase retrieval and structure reconstruction

Classification: hit finding, size filtering, single-hit classification

YOLOv2



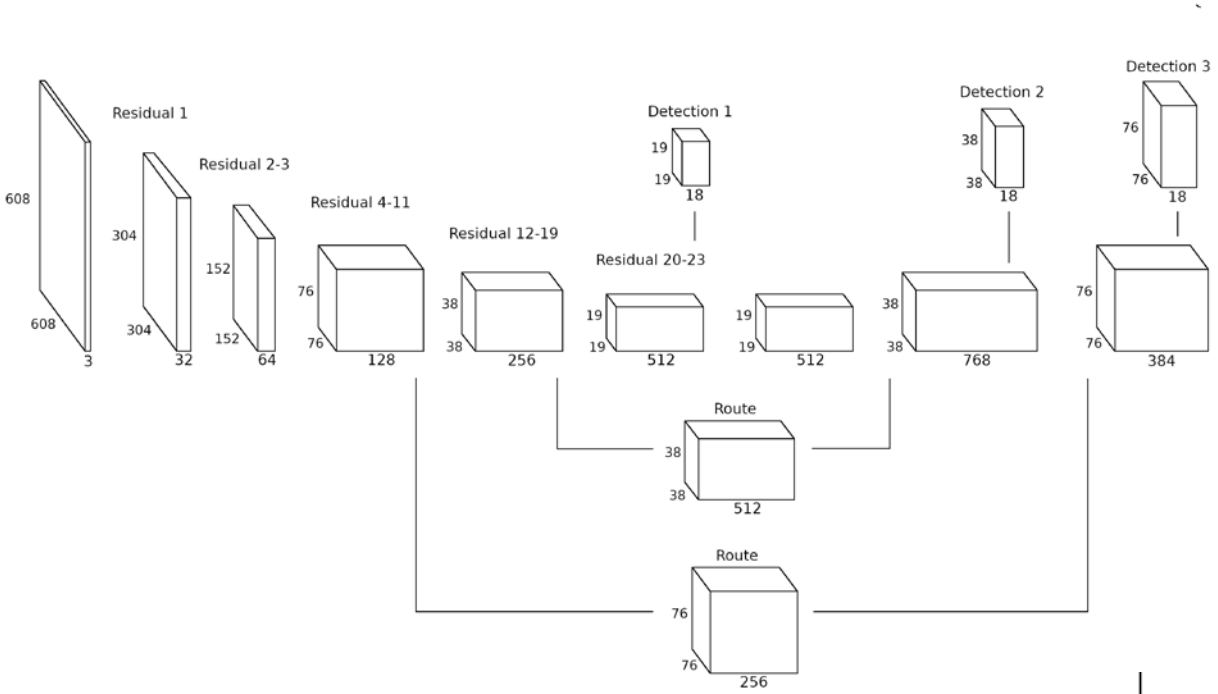
S. Seong et al., Determination of Vehicle Trajectory through Optimization of Vehicle Bounding Boxes using a Convolutional Neural Network. *Sensors* **2019**, *19*, 4263

Darknet-19 feature extractor trained on ImageNet
+ 1 convolutional layer: 20 convolutional layers

Loss function with classification and localization terms

Stochastic gradient descent (SGD) used for minimization
Batch size: 64

YOLOv3

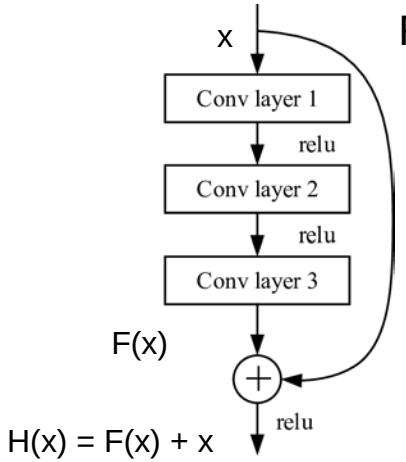


Darknet-53 feature extractor trained on ImageNet
53 convolutional layer in feature extraction

Residual architecture

Loss function with classification and localization terms

Stochastic gradient descent (SGD) used for minimization
Batch size: 24



Residual block

Metrics

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

TP – True positives

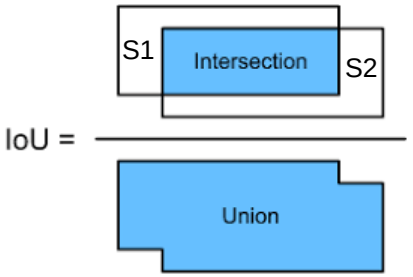
FP – False positives

FN – False negatives

$$F_\beta = (1 + \beta^2) \cdot \frac{\text{precision} \cdot \text{recall}}{(\beta^2 \cdot \text{precision}) + \text{recall}}$$

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

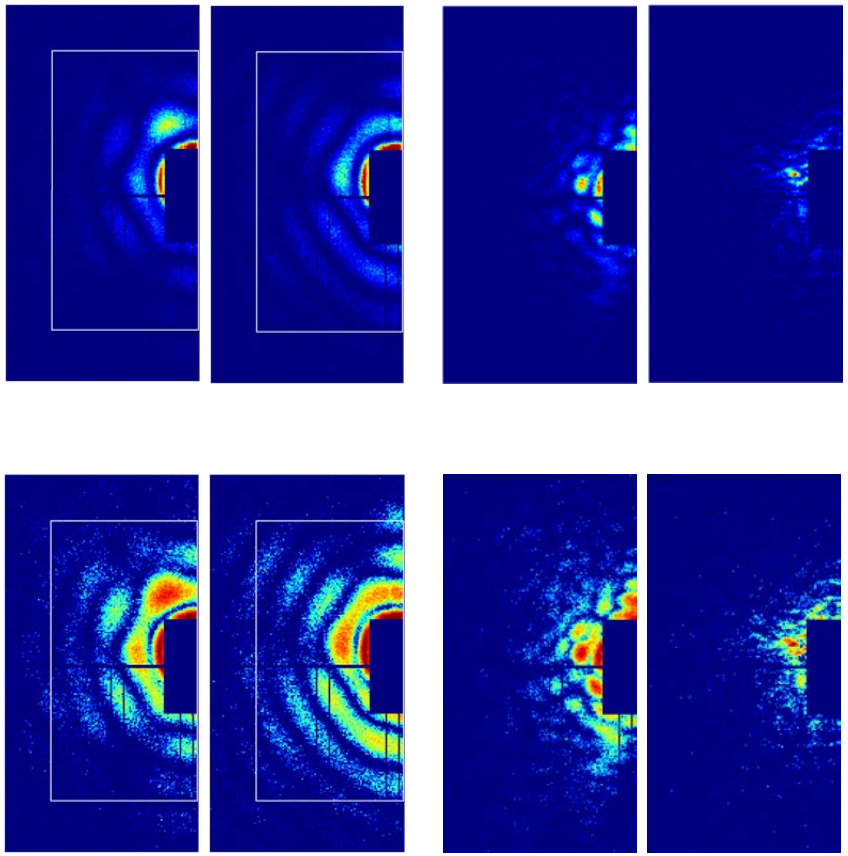
$$\text{IoU} = \frac{Intersection}{Union} = \frac{Intersection}{S_1 + S_2 - Intersection}$$



Data representation (SPI @ LCLS 2018)

Color (3 RGB layers, “jet” color scheme)

Grayscale (3 identical layers)



Linear scale

Logarithmic scale

Positive

Negative

Positive

Negative

ROI
123×240 detector px

$I(q) \sim q^{-3} \div q^{-4}$

Training, validation and test sets

Training set

165 positive examples (single hits)

390 negative examples

Validation set

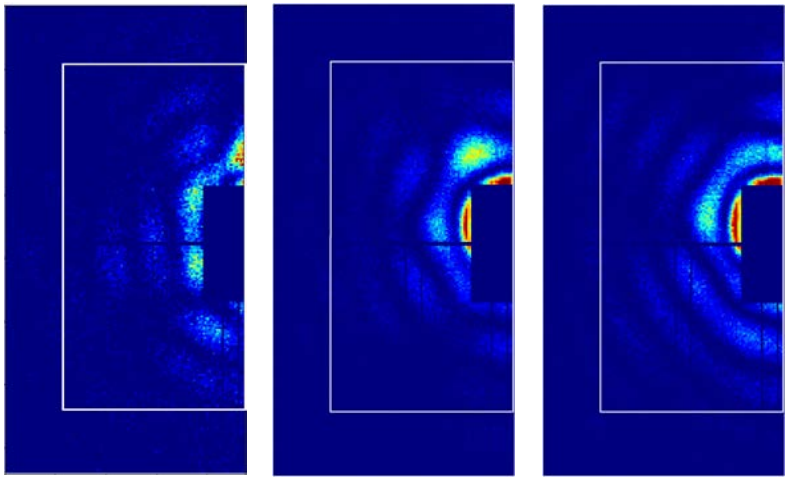
53 positive examples (single hits)

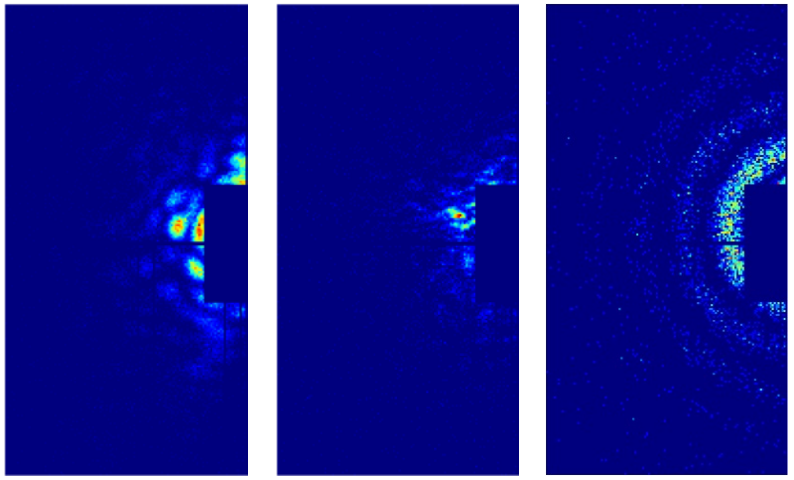
283 negative examples

Test set
(55-84 nm)

995 positive examples (single hits)

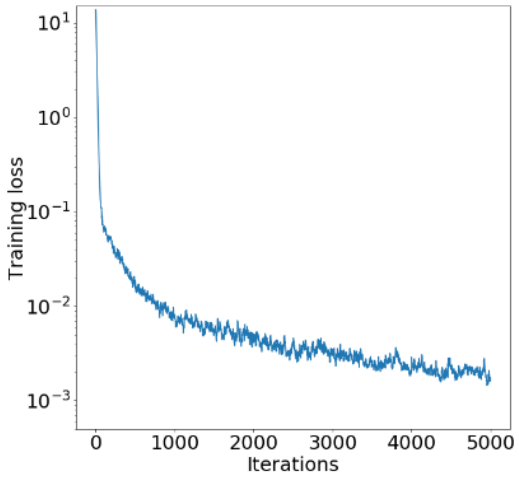
16 937 negative examples

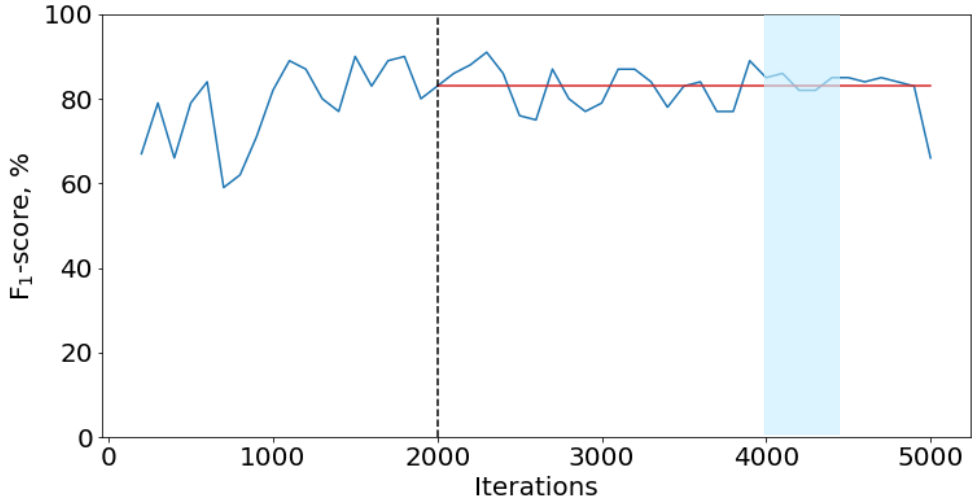




Training, validation and test

Observe training loss every iteration and save model every 100 iterations





Validation procedure to find an optimal training stage (not to optimize the hyper-parameters, *i.e.* size and number of filters, learning rate, batch size etc)

Calculate F_1-score on validation data set (every 100 iterations) to control overfitting

Intersection selection (4000, 4100, 4200, 4300 and 4400 iterations) for test

Results

Intersection selection

Model	Number of selected single hits	Intersection with manual selection	IoU for manual selection, %	Accuracy, %	Precision, %	Recall, %
YOLOv2, color, linear	1185	597	38	95	50	60
YOLOv2, color, log	1368	614	35	94	45	62
YOLOv2, grayscale, linear	1756	622	29	92	35	63
YOLOv2, grayscale, log	904	487	34	95	54	50
YOLOv3, color, linear	1041	505	33	94	49	51
YOLOv3, grayscale, linear	2316	465	16	87	20	47

Results

Comparison with a reference method

Model	Number of selected single hits	Intersection with manual selection	IoU for manual selection, %	Accuracy, %	Precision, %	Recall, %
YOLOv2, color, linear	1185	597	38	95	50	60
EM based	1085	574	34	94	53	48

Ground truth: manual selection

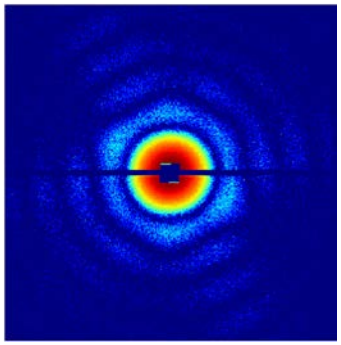
A. Ignatenko et al. (2020) "Classification of diffraction patterns in single particle imaging experiments performed at X-ray free-electron lasers using a convolutional neural network", *Machine Learning: Science and Technology* (accepted). Preprint available http://arxiv.org/abs/2008.07288

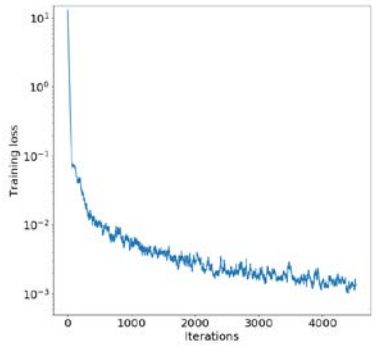
Results (SPI @LCLS 2015)

Training

165 positive examples

390 negative examples

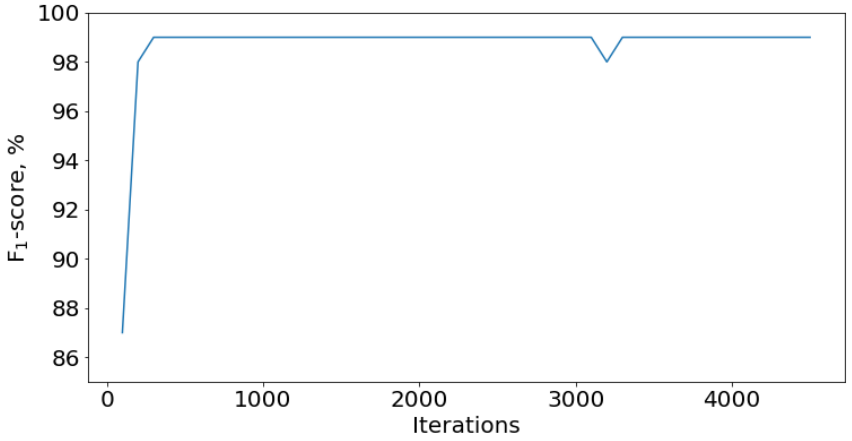




Validation

80 positive examples

180 negative examples



Test

~44k patterns

	Accuracy, %	Precision, %	Recall, %
Iteration 2000	83	53	99

Ground truth – PCA selection

Conclusions & Outlook

- Classification of single hits with CNNS is promising
- Moderate depth of the CNN is enough for the task
- Difficulties: large variation in experimental data, uncertainty in ground truth

- Understand the nature of variation in experimental data
- Learn to construct a more reliable ground truth
- Generalize the classification model for different particles

Thank you for your attention!

Coherent X-ray Scattering and Imaging Group at DESY

Prof. Dr. I. A. Vartanyants
N. Mukharamova
D. Assalauova
D. Lapkin

Dr. J. Carnis
Dr. Y-Y. Kim
Dr. A. Ignatenko
R. Khubbutdinov

Prof. Dr. E. Weckert (DESY)

F. Isensee (DKFZ)

S. A. Bobkov (National research Center “Kurchatov Institute”)

A. B. Teslyuk (National research Center “Kurchatov Institute”, Moscow Institute of Physics and Technology)

SPI consortium @ LCLS

A. Aquila (LCLS)

Helmholtz Associations Initiative and Networking Fund and the Russian Science Foundation grant HRSF-0002

RSF Russian Science Foundation