Scattering pattern classification in single-particle imaging experiments with a Siamese neural network

Forename Surname[a]* and Forename Surname[a]

[a] *SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA.*

E-mail: email

I will explore model generalizability in two directions — (1) similar size, different shapes; (2) different size, similar shapes (which might be harder to find those cases in PDB). Accordingly, the paper about this new networks can also diverge into two directions: (1) A paper about application of large computing facilities for enabling real-time SPI classification; (2) A paper about a generalized model for enabling real-time SPI classification.

1. Introduction

Single-particle imaging (SPI) with X-ray free-electron lasers (XFELs) is a promising method of determining the three-dimensional structure of noncrystalline biomolecules at room temperature. In SPI experiments, femtosecond coherent X-ray beams strike biomolecules sprayed into the beam path, causing scattering of radiation before destruction of the samples inflicted by the intense X-rays. This way of collecting scattering datasets is known as the 'diffraction before destruction' approach (Neutze *et al.*, 2000; Chapman *et al.*, 2006; Seibert *et al.*, 2011; Aquila *et al.*, 2015; Reddy *et al.*, 2017). Scattering patterns in SPI experiments largely fall into four categories, depending on how X-ray pulses hit sample particles delivered through jet streams. Astonishingly, 98% X-ray pulses miss their target particles (Shi *et al.*, 2019), leading to no scattering pattern identified as a 'no hit'. When X-ray photons collides with one and only one sample particle, the emerging scattering pattern is defined as a 'single hit'. Expectedly, a 'multi hit' happens when X-ray pulses and a cluster of sample particles intersect. In some cases, X-ray pulses might hit non-biological objects in delivery medium and those resulting scattering patterns are labeled as 'unintended hit'. Among the four categories, only 'single hit' images will contribute to reconstruction of electron density maps through downstream data processing, e.g. orientation recover and phase retrieval. Therefore, an efficient SPI scattering pattern classifier is long desired, but building one has remained to be an elusive task.

Some pioneering works that address the challenge of classifying SPI scattering pattern have been developed. Unsupervised methods (Yoon *et al.*, 2011; Giannakis *et al.*, 2012; Schwander *et al.*, 2012; Yoon, 2012; Andreasson *et al.*, 2014; Bobkov *et al.*, 2015) focus on extracting features from image samples and finding clusters of 'single hit' images. However, unsupervised solutions are highly problem-specific, that is to say, it's not realistic to expect 'single hit' images from different biological samples would form clusters that overlap in the feature space. Moreover, despite being unsupervised, those solutions also require prior knowledge of data distribution in feature space for classification, rendering it incapable of being automated. Supervised solutions based on neural network models learn from labelled images and provide predicted classification (Shi *et al.*, 2019; Ignatenko *et al.*, 2021). Convolution neural networks (CNN) were employed for feature extraction and a fully connected layer is directly attached to the convolutional layers for classification. Training such networks demands a large sample size for each class, which might not be available during the course of data collection. Specifically, the population of images vary substantially across different classes, resulting in imbalanced training data detrimental to model training.

In this work, we strive to approach the goal of classifying scattering patterns in nearly real-time SPI data collection with a figure to summarize what it is.

Describe the unqiue characteristics of Siamese network and triplet loss function. Our experiments show a high accuracy for diffraction pattern classification using this model.

2. Background (Related work)

Characterize related work in addressing the problem raised in the introduction. Meanwhile, what works have also inspised us to propose our method? (The FaceNet paper)

3. Our method

- An overview of our method – Siamese network.
- Characterize the classification proble as ”Are they different in nature?” problem.
- Quantitative aspects of the model:
  - Image embedding network
  - How to measure similarity?
  - Triplet loss function.
- Optimization.

4. Experiments

- Datasets.
- Data preprocessing (image, panel, mosaic)
- Data augmentation (for training, standardize, crop, random).
- Human baseline performance (game of hit)
- Accuracy-wise, our model has outperformed previous models.
- Less number of images to train the model (need to provide more results to justify).
- Limitations.

5. Conclusions

We introduced a model that does xxx. Our approach features xxx. We showed that this model provides state of the art performance on diffraction pattern classification. We evaluated our model on xxx experiments.

References

Andreasson, J., Martin, A. V., Liang, M., Timneanu, N., Aquila, A., Wang, F., Iwan, B., Svenda, M., Ekeberg, T., Hantke, M., Bielecki, J., Rolles, D., Rudenko, A., Foucar, L., Hartmann, R., Erk, B., Rudek, B., Chapman, H. N., Hajdu, J. & Barty, A. (2014). *Optics Express*, **22**(3), 2497.

Aquila, A., Barty, A., Bostedt, C., Boutet, S., Carini, G., dePonte, D., Drell, P., Doniach, S., Downing, K. H., Earnest, T., Elmlund, H., Elser, V., Gühr, M., Hajdu, J., Hastings, J., Hau-Riege, S. P., Huang, Z., Lattman, E. E., Maia, F. R. N. C., Marchesini, S., Ourmazd, A., Pellegrini, C., Santra, R., Schlichting, I., Schroer, C., Spence, J. C. H., Vartanyants, I. A., Wakatsuki, S., Weis, W. I. & Williams, G. J. (2015). *Structural Dynamics*, **2**(4), 041701.

Bobkov, S. A., Teslyuk, A. B., Kurta, R. P., Gorobtsov, O. Y., Yefanov, O. M., Ilyin, V. A., Senin, R. A. & Vartanyants, I. A. (2015). *Journal of Synchrotron Radiation*, **22**(6), 1345–1352.

Bromley, J., Guyon, I., LeCun, Y., Säckinger, E. & Shah, R. (1994). *Advances in Neural Information Processing Systems*, **6**, 737–744.

Chapman, H. N., Barty, A., Bogan, M. J., Boutet, S., Frank, M., Hau-Riege, S. P., Marchesini, S., Woods, B. W., Bajt, S., Benner, W. H., London, R. A., Plönjes, E., Kuhlmann, M., Treusch, R., Düsterer, S., Tschentscher, T., Schneider, J. R., Spiller, E., Möller, T., Bostedt, C., Hoener, M., Shapiro, D. A., Hodgson, K. O., van der Spoel, D., Burmeister, F., Bergh, M., Caleman, C., Huldt, G., Seibert, M. M., Maia, F. R. N. C., Lee, R. W., Szöke, A., Timneanu, N. & Hajdu, J. (2006). *Nature Physics*, **2**(12), 839–843.

Chopra, S., Hadsell, R. & LeCun, Y. (2005). In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pp. 539–546. San Diego, CA, USA: IEEE.

Giannakis, D., Schwander, P. & Ourmazd, A. (2012). *Optics Express*, **20**(12), 12799.

Ignatenko, A., Assalauova, D., Bobkov, S. A., Gelisio, L., Teslyuk, A. B., Ilyin, V. A. & Vartanyants, I. A. (2021). *Machine Learning: Science and Technology*, **2**(2), 025014.

Koch, G., Zemel, R. & Salakhutdinov, R. (2015). In *ICML 2015 Deep Learning Workshop*.

Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. (2000). *Nature*, **406**(6797), 752–757.

Reddy, H. K., Yoon, C. H., Aquila, A., Awel, S., Ayyer, K., Barty, A., Berntsen, P., Bielecki, J., Bobkov, S., Bucher, M., Carini, G. A., Carron, S., Chapman, H., Daurer, B., DeMirci, H., Ekeberg, T., Fromme, P., Hajdu, J., Hanke, M. F., Hart, P., Hogue, B. G., Hosseinizadeh, A., Kim, Y., Kirian, R. A., Kurta, R. P., Larsson, D. S., Duane Loh, N., Maia, F. R., Mancuso, A. P., Mühlig, K., Munke, A., Nam, D., Nettelblad, C., Ourmazd, A., Rose, M., Schwander, P., Seibert, M., Sellberg, J. A., Song, C., Spence, J. C., Svenda, M., Van der Schot, G., Vartanyants, I. A., Williams, G. J. & Xavier, P. L. (2017). *Scientific Data*, **4**(1), 170079.

Schroff, F., Kalenichenko, D. & Philbin, J. (2015). In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 815–823. Boston, MA, USA: IEEE.

Schwander, P., Giannakis, D., Yoon, C. H. & Ourmazd, A. (2012). *Optics Express*, **20**(12), 12827.

Seibert, M. M., Ekeberg, T., Maia, F. R. N. C., Svenda, M., Andreasson, J., Jönsson, O., Odić, D., Iwan, B., Rocker, A., Westphal, D., Hantke, M., DePonte, D. P., Barty, A., Schulz, J., Gumprecht, L., Coppola, N., Aquila, A., Liang, M., White, T. A., Martin, A., Caleman, C., Stern, S., Abergel, C., Seltzer, V., Claverie, J.-M., Bostedt, C., Bozek, J. D., Boutet, S., Miahnahri, A. A., Messerschmidt, M., Krzywinski, J., Williams, G., Hodgson, K. O., Bogan, M. J., Hampton, C. Y., Sierra, R. G., Starodub, D., Andersson, I., Bajt, S., Barthelmess, M., Spence, J. C. H., Fromme, P., Weierstall, U., Kirian, R., Hunter, M., Doak, R. B., Marchesini, S., Hau-Riege, S. P., Frank, M., Shoeman, R. L., Lomb, L., Epp, S. W., Hartmann, R., Rolles, D., Rudenko, A., Schmidt, C., Foucar, L., Kimmel, N., Holl, P., Rudek, B., Erk, B., Hömke, A., Reich, C., Pietschner, D., Weidenspointner, G., Strüder, L., Hauser, G., Gorke, H., Ullrich, J., Schlichting, I., Herrmann, S., Schaller, G., Schopper, F., Soltau, H., Kühnel, K.-U., Andritschke, R., Schröter, C.-D., Krasniqi, F., Bott, M., Schorb, S., Rupp, D., Adolph, M., Gorkhover, T., Hirsemann, H., Potdevin, G., Graafsma, H., Nilsson, B., Chapman, H. N. & Hajdu, J. (2011). *Nature*, **470**(7332), 78–81.

Shi, Y., Yin, K., Tai, X., DeMirci, H., Hosseinizadeh, A., Hogue, B. G., Li, H., Ourmazd, A., Schwander, P., Vartanyants, I. A., Yoon, C. H., Aquila, A. & Liu, H. (2019). *IUCrJ*, **6**(2), 331–340.

Taigman, Y., Yang, M., Ranzato, M. & Wolf, L. (2014). In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1701–1708. Columbus, OH, USA: IEEE.

Yin, K. & Tai, X.-C. (2018). *Journal of Scientific Computing*, **74**(1), 175–196.

Yoon, C. (2012). In *SPIE Optical Engineering + Applications*, edited by P. J. Bones, M. A. Fiddy & R. P. Millane, p. 85000H. San Diego, California, USA.

Yoon, C. H., Schwander, P., Abergel, C., Andersson, I., Andreasson, J., Aquila, A., Bajt, S., Barthelmess, M., Barty, A., Bogan, M. J., Bostedt, C., Bozek, J., Chapman, H. N., Claverie, J.-M., Coppola, N., DePonte, D. P., Ekeberg, T., Epp, S. W., Erk, B., Fleckenstein, H., Foucar, L., Graafsma, H., Gumprecht, L., Hajdu, J., Hampton, C. Y., Hartmann, A., Hartmann, E., Hartmann, R., Hauser, G., Hirsemann, H., Holl, P., Kassemeyer, S., Kimmel, N., Kiskinova, M., Liang, M., Loh, N.-T. D., Lomb, L., Maia, F. R. N. C., Martin, A. V., Nass, K., Pedersoli, E., Reich, C., Rolles, D., Rudek, B., Rudenko, A., Schlichting, I., Schulz, J., Seibert, M., Seltzer, V., Shoeman, R. L., Sierra, R. G., Soltau, H., Starodub, D., Steinbrener, J., Stier, G., Strüder, L., Svenda, M., Ullrich, J., Weidenspointner, G., White, T. A., Wunderer, C. & Ourmazd, A. (2011). *Optics Express*, **19**(17), 16542.