Online Inference of Structure Factor Amplitudes:
Serial Merging for Serial Crystallography

Kevin M. Dalton [1]   Doeke R. Hekstra [1,2]

[1]Dept. of Molecular and Cellular Biology, Harvard University
[2]John A. Paulson School of Engineering and Applied Sciences

Introduction

With free-electron lasers reaching megahertz repetition rates [10], serial crystallography data sets are trending upward in size. For this reason, it is important to construct scalable algorithms for data reduction. Distributed computing is one way to accomplish this. National labs typically have ample compute resources to support large scale applications implemented with Message Passing Interface or other distributed computing protocols. However, a different approach is to construct algorithms which operate on small batches of data on a single computer. The extreme case, an online algorithm, learns to process data by only looking at a single example at a time. Here we recount the successful implementation of one such algorithm for scaling and merging reflection intensities. The algorithm uses deep learning to scale reflection intensities encouraging the merged structure factor estimates to follow a crystallographic prior distribution. The model is trained by gradient descent on a Bayesian objective function. We demonstrate that the model can estimate productive global parameter updates from single images. This approach has the advantage that it has relatively modest hardware requirements, can adapt on the fly as new data are acquired, and has the potential for transfer learning between data sets. We envision a future where convergence of this merging model can serve as the diagnostic criterion for terminating experiments at XFELs.

A Bayesian Merging Model for Diffraction

The goal of crystallographic data reduction is to produce estimates of structure factor amplitudes which, along with phases, determine the electron density of the sample. Ideally, these estimates should satisfy two criteria. Structure factor estimates should be:

- consistent with the observed data
- consistent with the expected distributions obtained from first principles [11]

It is possible to obtain such estimates using variational inference [1]. In the context of crystallography, this is accomplished by using gradient-based optimization to maximize the objective function, [4]

$$ELBO = \log P(\text{data}|\text{model}) - D_{KL}(\text{Structure Factors}||\text{Prior}) - D_{KL}(\text{Scale Factors}||\text{Prior}),$$

known as the Evidence Lower BOund. D_{KL} refers to the Kullback-Leibler divergence, a measure of distance between two probability distributions. The particular parameterization of this objective is a modeling choice.

Model Parameterization

Intensity Model:
$$F = \mathcal{T}runcated\mathcal{N}ormal\left(\boldsymbol{\mu_F}, \boldsymbol{\sigma_F}, 0, \infty\right)$$
$$\mu_\Sigma, \sigma_\Sigma = \boldsymbol{g_\theta}(F, I_{obs}, \sigma_{obs}, Refl.\ Metadata)$$
$$\Sigma = \mathcal{L}og\mathcal{N}ormal(\mu_\Sigma, \sigma_\Sigma)$$
$$\boldsymbol{k}I_{obs} \sim \mathcal{N}\left(\Sigma F^2, \boldsymbol{k}\sigma_{obs}\right)$$

Priors:
$$P_F = \mathcal{W}ilson$$
$$P_\Sigma = \mathcal{L}og\mathcal{N}ormal(0, 1)$$

Loss function:
$$\mathcal{L}(\boldsymbol{k}, \boldsymbol{\theta}, \boldsymbol{\mu_F}, \boldsymbol{\sigma_F}) = -ELBO$$
$$\approx -\sum_i^S \left\{ \log \mathcal{N}(\boldsymbol{k}I_{obs}|s_i f_i^2, \boldsymbol{k}\sigma_{obs}) + \right.$$
$$w_F \left[\log F(f_i) - \log P_F(f_i)\right] +$$
$$\left. w_\Sigma D_{KL}(\Sigma||P_\Sigma) \right\}$$

In this model we choose to parameterize structure factors by a truncated normal distribution with positive support as in previous work [4]. The optimization variables are indicated in bold. The scalar variable, $\boldsymbol{k}$, is a global scale parameter which is constrained to be positive. $\boldsymbol{k}$ is important to account for the fact that the prior on Σ has an arbitrary mean. $\boldsymbol{\mu_F}$ and $\boldsymbol{\sigma_F}$ are the location and scale parameters of the structure factor distributions. Both are constrained to be positive using the softplus function. We refer to the parameters of the scale model as θ. These consist exclusively of bias vectors and randomly initialized weight matrices in the linear layers of the model. In this implementation the first two terms of the $ELBO$ are estimated by drawing reparameterized samples [8] from the structure factor distribution and scale distribution. The Kullback-Leibler divergence between the scales and their prior is computed analytically. The weight terms, w_F and w_Σ, are hyperparameters which modulate the strength of the priors. During each training update, $\mathcal{L}$, is summed over the reflections in one image.

Feed Forward Layer
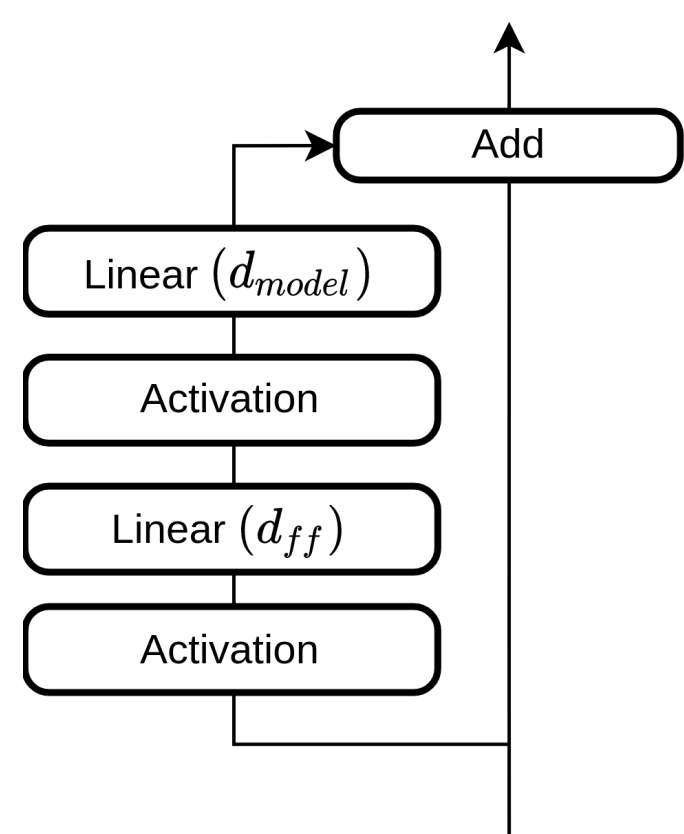
The fundamental nonlinear building block of the scale model is a feed forward neural network. This component of the model takes a batch of vectors with d_{model} dimensions and outputs a modified batch of vectors in the same dimensional space. In this work, we use a ResNet version 2 block [5]. The main advantage of this type of layer is that it contains a residual connection between the input and output. The output layer takes the elementwise sum of the input and a nonlinear projection thereof. This ensures that gradients of the loss function do not vanish in earlier layers leading to stalled learning.

Image Encoder Model
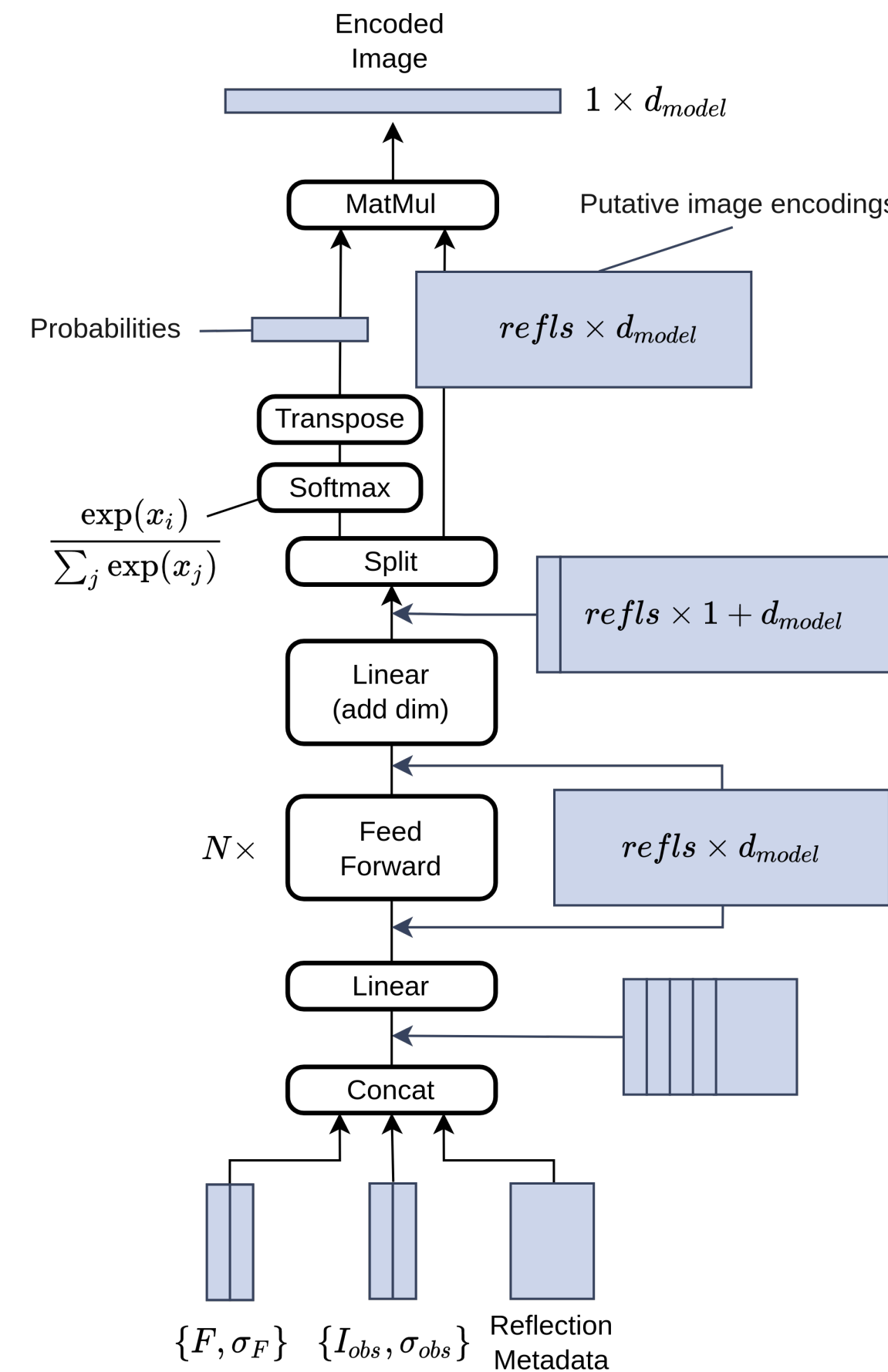
The idea behind the image encoder is to learn a function which looks at the current structure factor estimates, the context of each reflection (reflection metadata), and the observed intensities and suggests a vector that describes the image. Hopefully, this vector includes information needed to scale the reflections on this image. For instance, the vector should encode information about the orientation of the crystal, its size, shape, and mosaic properties, the brightness and deflection of the X-ray beam, etc. The way this representation comes about is that a neural network considers each reflection in isolation and proposes a putative image encoding based on that reflection. Alongside the putative encoding, the network also outputs a score representing the probability that the reflection has something meaningful to say about the image overall. The consensus encoding is taken to be the sum over candidate encodings weighted by this probability. Pooling the candidate encodings into a single vector of length d_{model} prevents this model from passing reflections' intensities directly to the scale model. This is essential to prevent the overall model from overfitting during training.
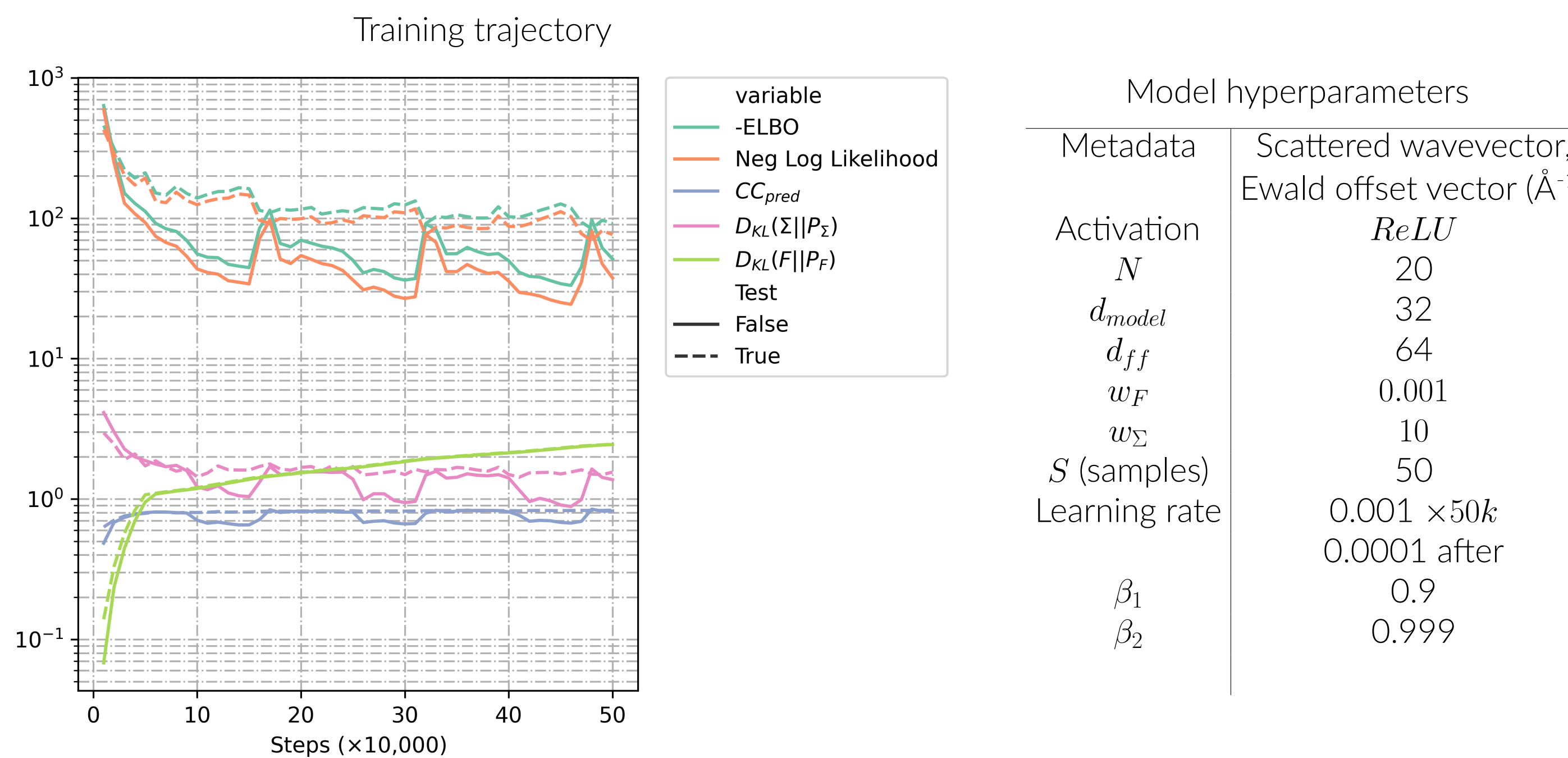
Reflection Scaling Model

The goal of the scaling model is to produce a distribution of probable scale factors appropriate for each reflection in an image. As in previous work [4], it uses metadata associated with each reflection in order to infer scales. Unlike the previously published model, it has access to the image representation produced by the upstream encoder model. The input layers of this model simply concatenate the image vector onto each reflection's metadata. A neural network uses this enriched representation of the reflection to predict appropriate scale factors. In the parameterization presented here, the neural network predicts a two-vector interpreted as the location and scale parameters of a log-normal distribution. The LogNormal layer converts the scale parameter to a strictly positive number by applying the softplus function, $f(x) = \log(1 + exp(x))$.
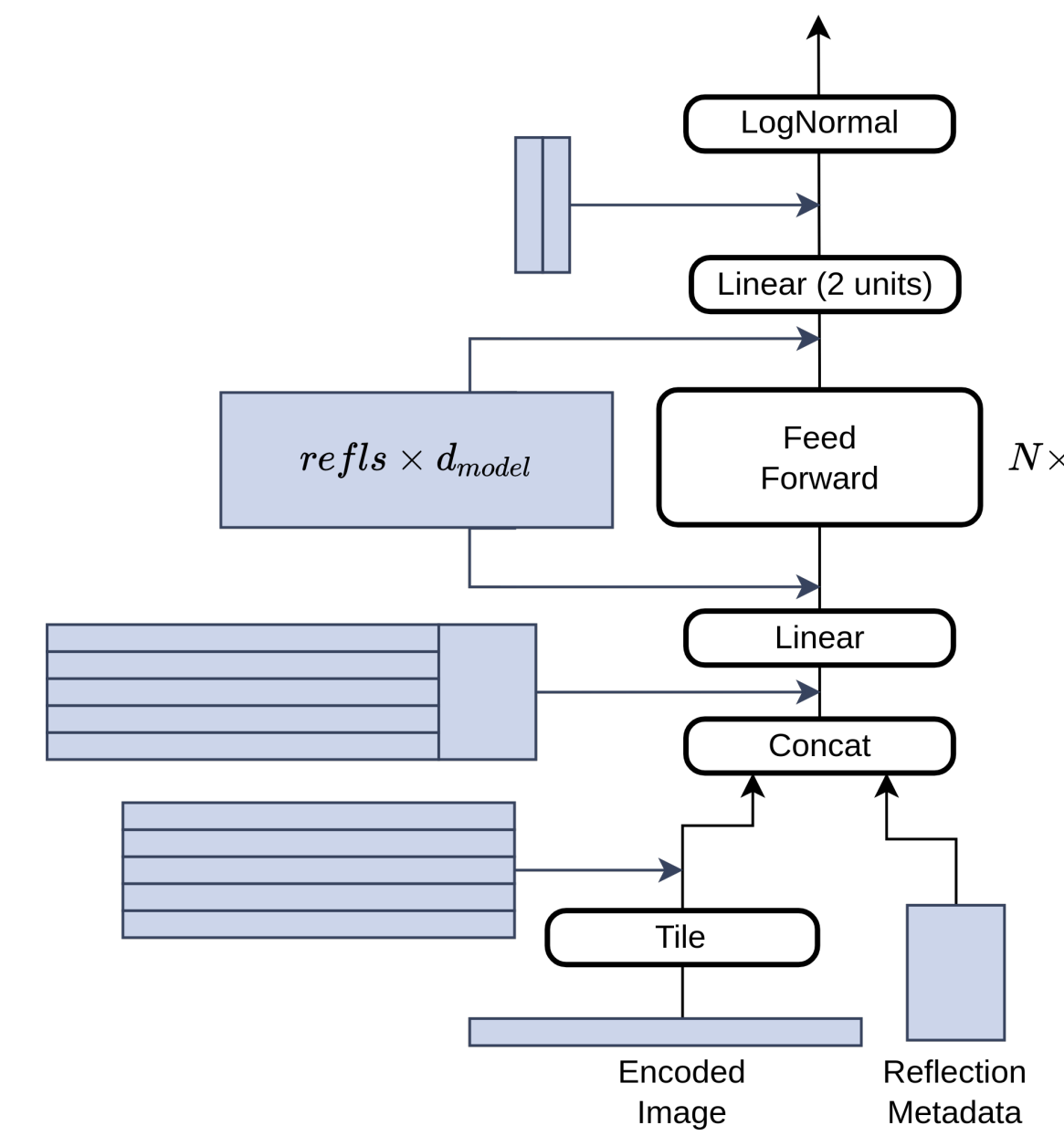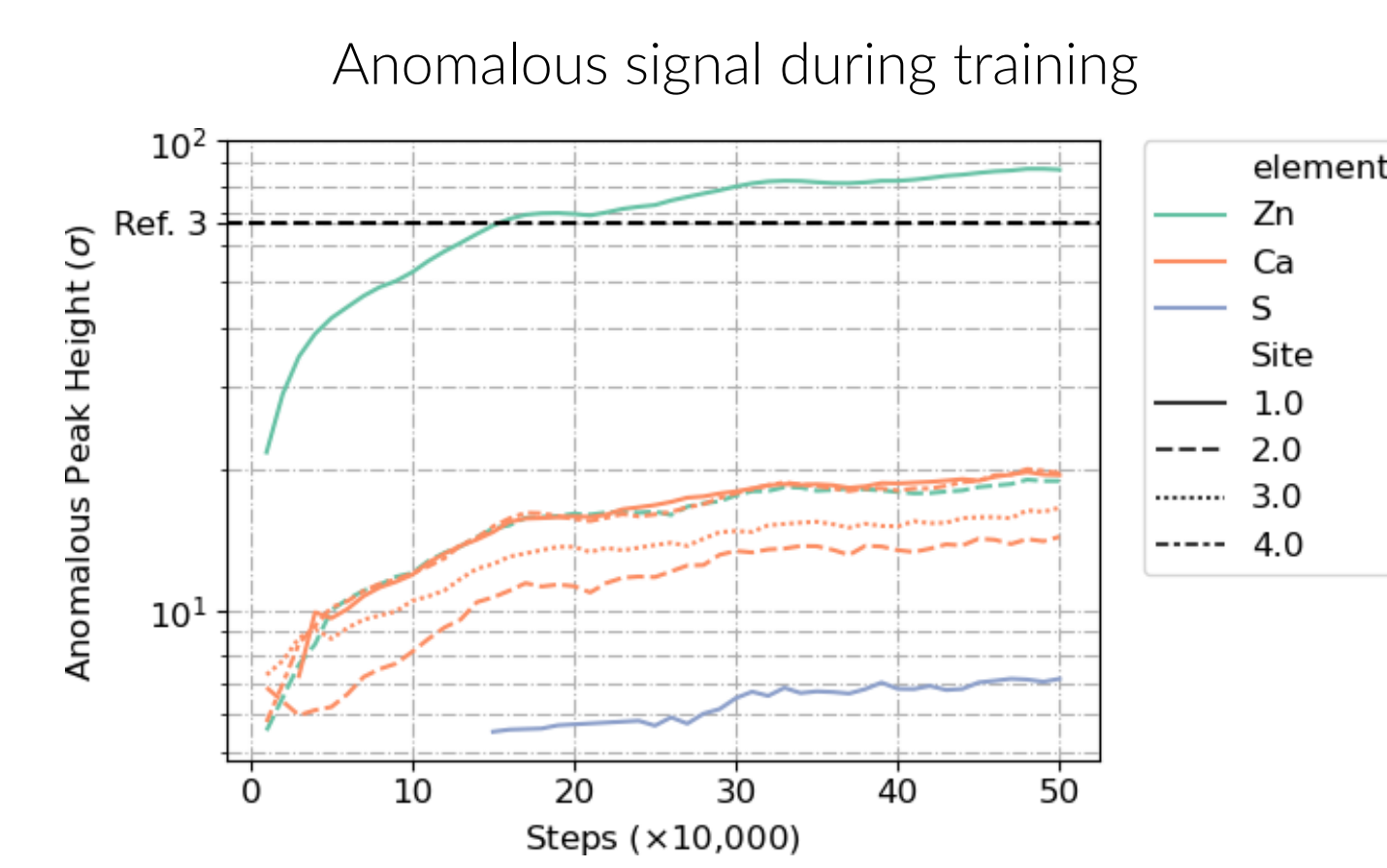
Example Application to Single Wavelength Anomalous Diffraction Data

As a case study for the suitability of this model for merging serial femtosecond crystallography data, we applied it to a publicly available dataset consisting of 166,250 diffraction images [6] of the zinc metalloprotease, thermolysin, which were integrated using DIALS [2] (CXIDB: 81). The data were acquired at a wavelength of 1.27 Å leading to considerable anomalous signal from the zinc ions in the active site. In addition to zinc, thermolysin also has four ordered calcium ions and two methionine residues which may contribute anomalous signal. We trained our model for 500,000 gradient updates using the Adam optimizer for stochastic objective functions [7]. This equates to slightly more than 3 passes over the dataset. We trained the model on a consumer GPU (NVIDIA RTX 3060). With the reported hyperparameters, the model requires 797 megabytes of GPU memory and executes a gradient step in roughly 70ms. Total training time was less than 12 hours. By optimizing the input pipeline and tuning hyperparameters, we expect a 3 to 10-fold speedup is attainable.


Training trajectory

variable	
-ELBO	
Neg Log Likelihood	
CCpred	
$D_{KL}(\Sigma||P_\Sigma)$	
$D_{KL}(F||P_F)$	
Test	
False	
True	

Model hyperparameters	
Metadata	Scattered wavevector, Ewald offset vector (Å^{-1})
Activation	ReLU
N	20
d_{model}	32
d_{ff}	64
w_F	0.001
w_Σ	10
S (samples)	50
Learning rate	0.001 $\times 50k$ 0.0001 after
β_1	0.9
β_2	0.999

Thermolysin Phasing Results

Anomalous signal during training

	element
	Zn
	Ca
	S

	1.0
	2.0
	3.0
	4.0

Comparison of SAD phasing results

	Ours	Reference [3]
Zn^{2+} peak (σ)	83.2	67.0
Model-map CC	0.84	0.80±0.001
Sites found	7	6
Residues built	305	297.2±6.5
R_{work}	0.204	0.212±1.3
R_{free}	0.231	0.237±1.7

The table presents SAD Phasing Results from AutoSol [9] using the same parameters as [3]. The zinc peak height in the table is from `phenix.find_peaks_holes` using our structure factor estimates and the autobuilt model.

Thermolysin active site   Methionine-205
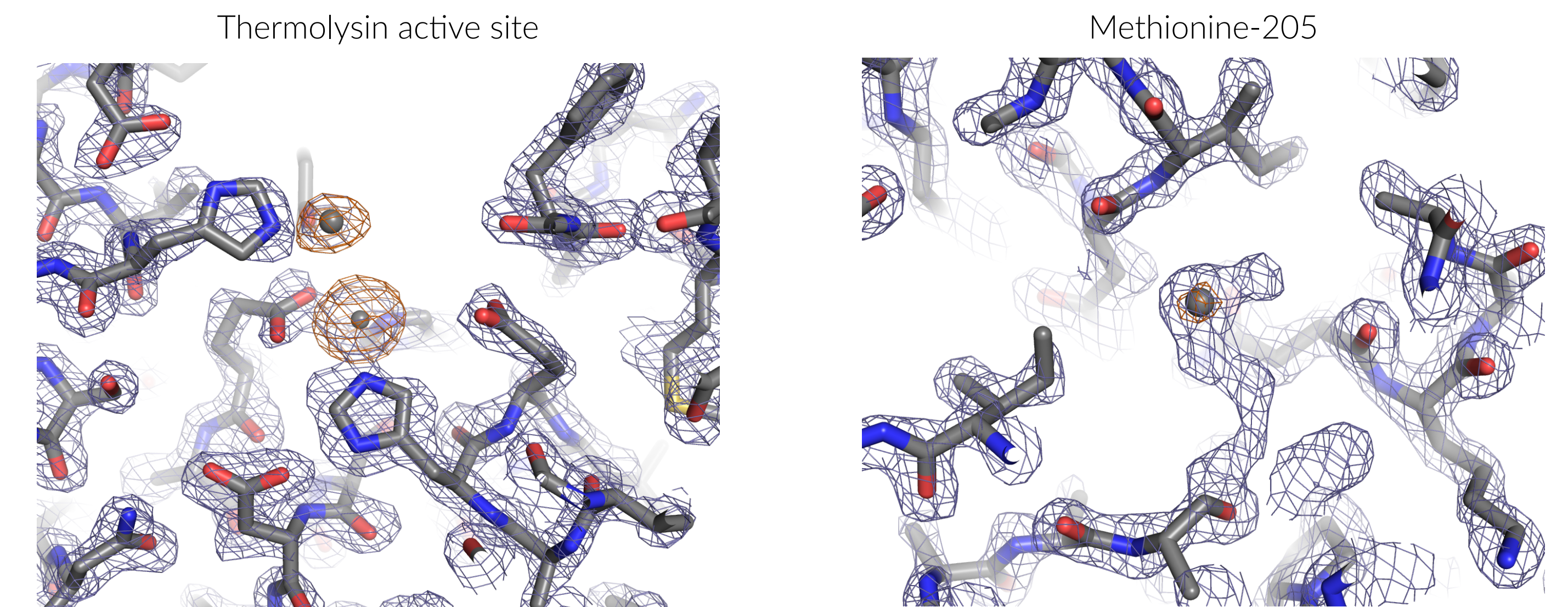
Thermolysin SAD phasing results featuring the autobuilt model from AutoSol [9]. The blue electron density, contoured at 2σ, is the experimental, density-modified map. The orange map is the anomalous difference map using the pictured model for phases and the structure factor estimates from our merging algorithm contoured at 5σ.

Acknowledgements

References

[1] David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. "Variational Inference: A Review for Statisticians". In: *Journal of the American Statistical Association* 112.518 (Apr. 3, 2017), pp. 859–877.

[2] A. S. Brewster et al. "Improving signal strength in serial crystallography with DIALS geometry refinement". In: *Acta Crystallographica Section D: Structural Biology* 74.9 (Sept. 1, 2018). Number: 9 Publisher: International Union of Crystallography, pp. 877–894.

[3] A. S. Brewster et al. "SAD phasing of XFEL data depends critically on the error model". In: *Acta Crystallographica Section D: Structural Biology* 75.11 (Nov. 1, 2019). Publisher: International Union of Crystallography, pp. 959–968.

[4] Kevin M. Dalton, Jack B. Greisman, and Doeke R. Hekstra. "Careless: A Variational Bayesian Model for Merging X-ray Diffraction Data". In: (Jan. 7, 2021). Publisher: Cold Spring Harbor Laboratory Section: New Results, p. 2021.01.05.425510.

[5] Kaiming He et al. *Identity Mappings in Deep Residual Networks*. Number: arXiv:1603.05027. July 25, 2016.

[6] Jan Kern et al. "Taking snapshots of photosynthetic water oxidation using femtosecond X-ray diffraction and spectroscopy". In: *Nature Communications* 5.1 (July 9, 2014). Number: 1 Publisher: Nature Publishing Group, p. 4371.

[7] Diederik P. Kingma and Jimmy Ba. "Adam: A Method for Stochastic Optimization". In: *arXiv:1412.6980 [cs]* (Jan. 29, 2017).

[8] Diederik P. Kingma and Max Welling. "Auto-Encoding Variational Bayes". In: *arXiv:1312.6114 [cs, stat]* (May 1, 2014).

[9] T. C. Terwilliger et al. "Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard". In: *Acta Crystallographica Section D: Biological Crystallography* 65.6 (June 1, 2009). Number: 6 Publisher: International Union of Crystallography, pp. 582–601.

[10] Max O. Wiedorn et al. "Megahertz serial crystallography". In: *Nature Communications* 9.1 (Oct. 2, 2018). Number: 1 Publisher: Nature Publishing Group, p. 4025.

[11] A. J. C. Wilson. "The probability distribution of X-ray intensities". In: *Acta Crystallographica* 2.5 (Oct. 2, 1949). Number: 5 Publisher: International Union of Crystallography, pp. 318–321.