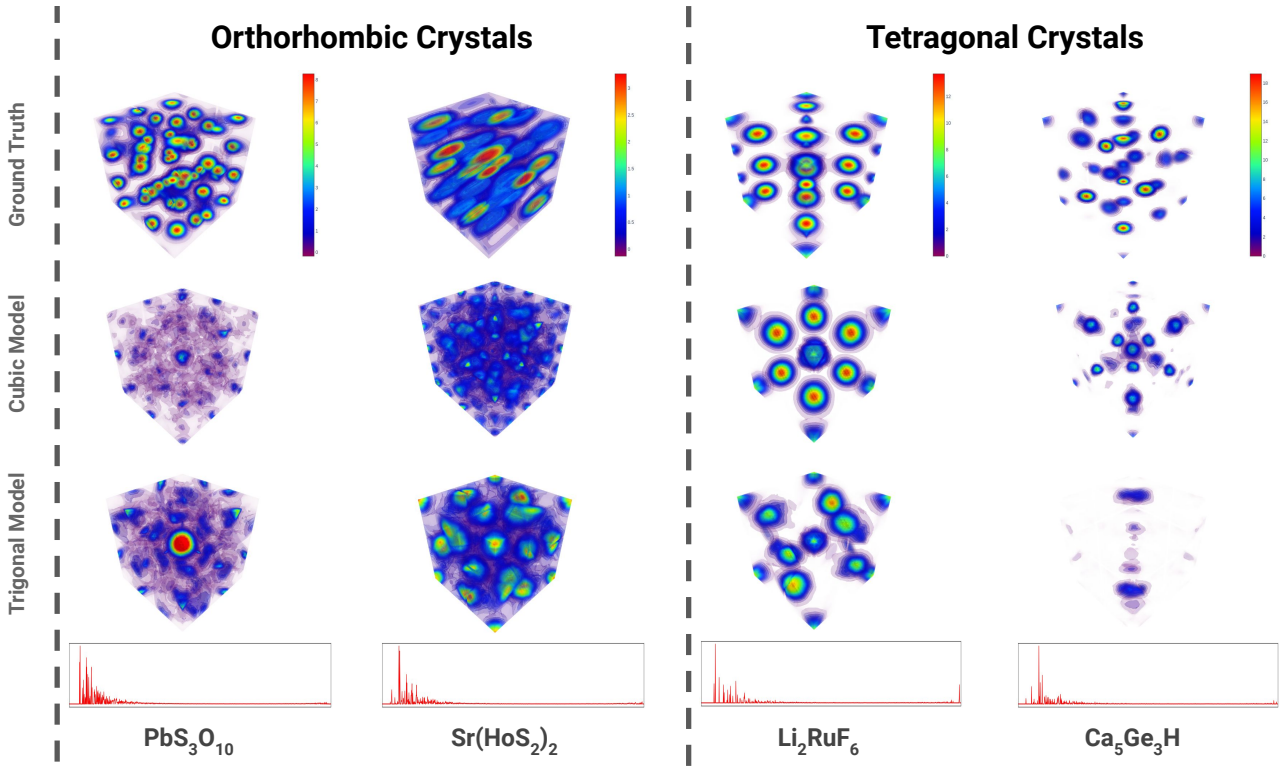
Supplementary Information for

Towards End-to-End Structure Determination from X-Ray Diffraction Data Using Deep Learning

Gabe Guo *et al.*



Supplementary Figure 1. Zero-Shot Adaptation to Unseen Systems: This qualitatively shows the reconstruction ability of our models (which were trained on the cubic and trigonal systems) on a few samples of crystals from systems (*i.e.*, orthorhombic, tetragonal) that were never seen during train time.

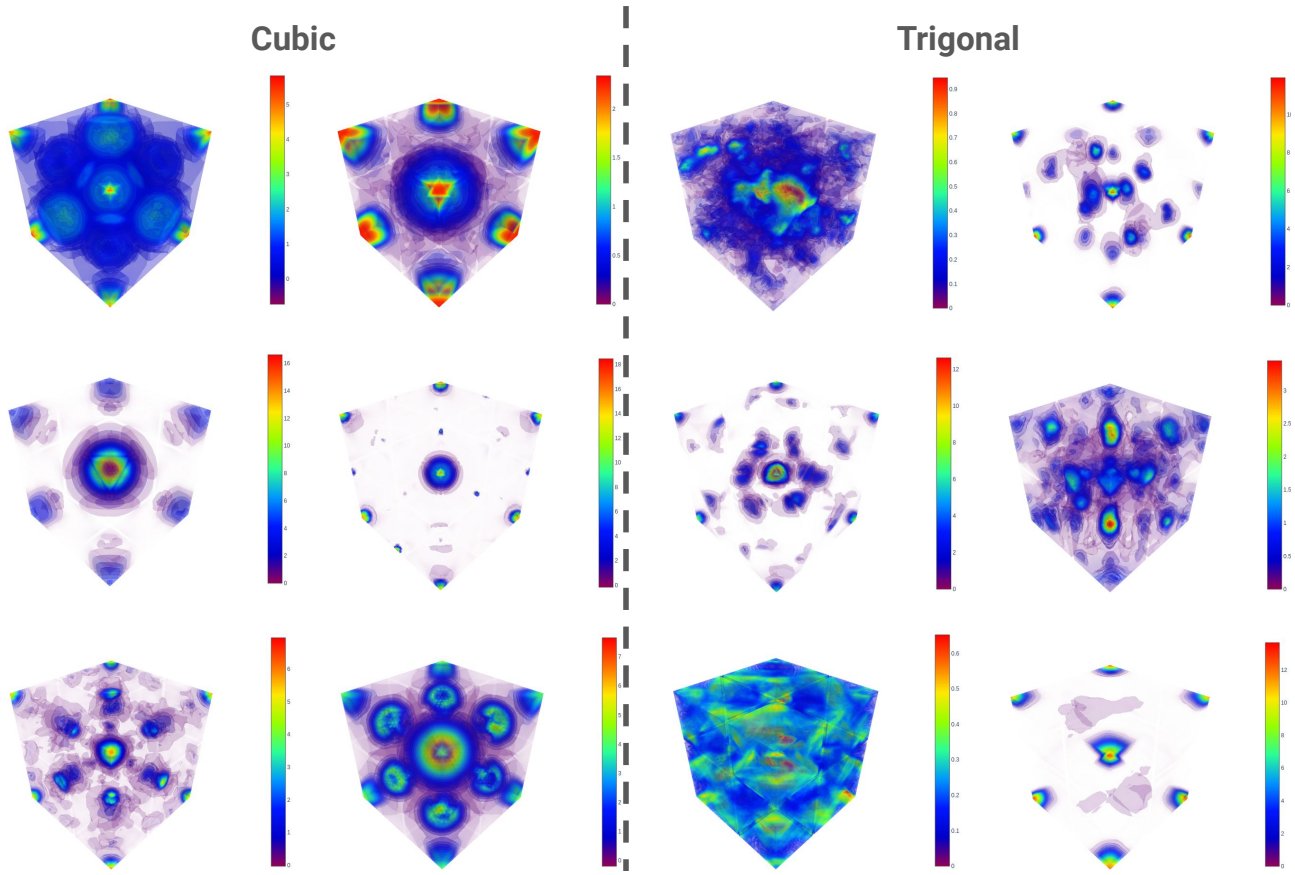
Supplementary Note 1: Zero-Shot Adaptation to Unseen Systems

We conduct preliminary qualtitative explorations on our models' reconstruction ability on a few crystals from the orthorhombic and tetragonal systems; neither system was seen during the train period. We use our models that were trained on the cubic and trigonal systems, and feed in 25 testing samples from the orthorhombic system and 25 testing samples from the tetragonal system to each. We only feed in 25 samples for each of these previously unseen systems because the goal of this exploration is simply to get a sense of how the model may behave on these other systems. See Supplementary Figure 1 and Supplementary Table 1.

As expected in this zero-shot setting, reconstruction is not very successful, judging from quantitative metrics and visual inspection. We see examples of reconstruction attempts that are barely coherent, such as the trigonal model on Ca_5Ge_3H. Even for the coherent reconstructions, both the cubic-trained and trigonal-trained models seem not to capture the symmetries of the orthorhombic and tetragonal systems. Instead, these models seem to predict crystals with symmetries from the systems they were trained on. That being said, the models do seem to have some notion of the frequency and intensity of the electron density signal, looking at examples like the cubic model's reconstruction of Li_2RuF_6. Furthermore, the SSIM gap between the performance on these never-before-seen systems ($0.528 - 0.598$ SSIM) and the performance of the trigonal model on the

System	**Model**			
	SSIM $(\mu \pm \sigma)$		**PSNR** $(\mu \pm \sigma)$	
	Cubic	**Trigonal**	**Cubic**	**Trigonal**
Orthorhombic $(n = 25)$	0.528 ± 0.174	0.571 ± 0.184	22.4 ± 3.8	22.7 ± 3.7
Tetragonal $(n = 25)$	0.586 ± 0.175	0.598 ± 0.170	21.8 ± 2.8	21.9 ± 2.5

Supplementary Table 1. Preliminary Zero-Shot Adaptation Results: The table displays the mean and standard deviation of SSIM and PSNR scores for our models trained on the cubic and trigonal systems, when evaluated on 25 examples each from the orthorhombic and tetragonal systems.



Supplementary Figure 2. Latent Space Exploration: This figure shows the unconditional generation capabilities of our models, where we randomly sample XRD and chemical formula vectors from the latent space, (*i.e.*, $\mathbf{E}_d, \mathbf{E}_f \sim \mathcal{N}(0,1)$) and feed them through the rest of the model. Results displayed are for the cubic and trigonal models.

trigonal system (0.741 SSIM) is similar to the SSIM gap between the performance of the trigonal model on the trigonal system and the cubic model on the cubic system (0.934 SSIM).

Supplementary Note 2: Generative Modeling via Latent Space Exploration

Inspired by the VAE's decoder-only unconditional generative capabilities[38], we attempt to use the latter parts of *CrystalNet* as a generative model. Specifically, we bypass XRD and chemical formula inputs (and also bypass their encoders). Instead, we randomly sample XRD and chemical formula representation vectors from the latent space, *i.e.*, $\mathbf{E}_d, \mathbf{E}_f \sim \mathcal{N}(0,1)$. We then feed these randomly sampled vectors through the feature fusion block, and combine them with the usual queried coordinates to pass through the charge regressor. This gives us unconditionally generated crystals, *i.e.*, crystals generated without any XRD or chemical information.

See Supplementary Figure 2. Firstly, we see that in both the cubic and trigonal systems, the latent space is generally well-behaved, in that sampling randomly from the latent space (under the assumption of a multivariate Gaussian distribution) gives us coherent outputs. We note that the outputs for the cubic system are generally sharper than those for the trigonal system. We also see the existence of "prototype" crystals, *i.e.*, many of the crystals that were unconditionally generated actually look very similar to the crystals that were predicted from XRD and chemical information (except for the absolute charge density). With this in mind, one interpretation of the XRD and chemical encoder branches is that they guide the sampling towards promising regions in the latent space.