Supplementary Information for Voxel Image of Crystals for High-Throughput Materials Screening: Formation Energy Prediction by a Deep Convolutional Network

Ali Davariashtiyani[1] and Sara Kadkhodaei[1,*]

[1]*Department of Civil, Materials, and Environmental Engineering, University of Illinois Chicago, Chicago, IL, USA*[*]

(Dated: June 9, 2023)

[*] To whom correspondence should be addressed; Email: sarakad@uic.edu

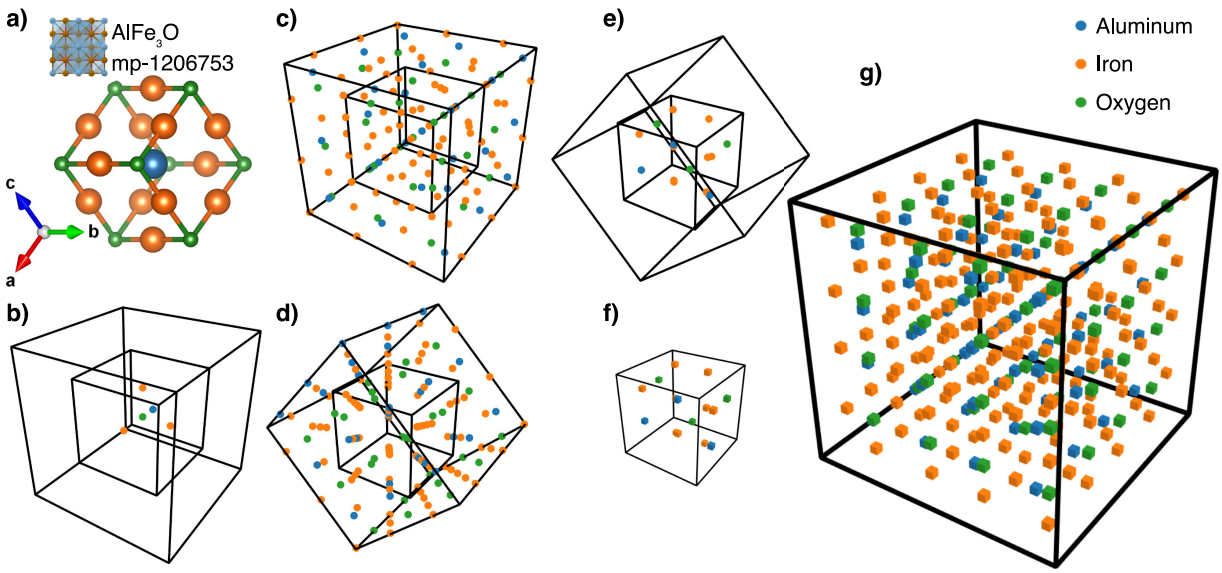

FIG. S1: The procedure for creating the sparse voxel images of crystals. (a) $AlFe_3O$ crystal structure. (b) Illustration of a sample unit cell converted to point cloud in an original box (6 $\mathring{A}$ side) centered at a large box (6 $\times\sqrt{3} = 10.4$ $\mathring{A}$ $\mathring{A}$ side). For better visualization, we depict a sample box of 6 $\mathring{A}$ instead of 17 $\mathring{A}$ that is used in our model. The crystal structure is placed in the center of the original box and is replicated by the lattice vectors to fill the larger box. (c) The point cloud is replicated along the lattice vectors to fill both boxes completely. (d) Application of a random rigid-body rotation (a rotation of 45, 0, and 0 degrees around the x, y, and z axes is shown) to the larger box and its containing crystal points while the original box is unchanged. (e) Removal of the crystal points out of the original box. (f) Voxelizing the original box into 16 $\times$ 16 $\times$ 16 and coloring the occupied voxels with RGB values of normalized atomic number, group number, and period number. For better visualization, we depict a sample grid of 16 $\times$ 16 $\times$ 16 instead of 32 $\times$ 32 $\times$ 32 that is used in our model (g) The final image of $AlFe_3O$ crystal rotated by (45°,0,0) with a box size of 17 $\mathring{A}$ and 32 $\times$ 32 $\times$ 32 voxel grid.

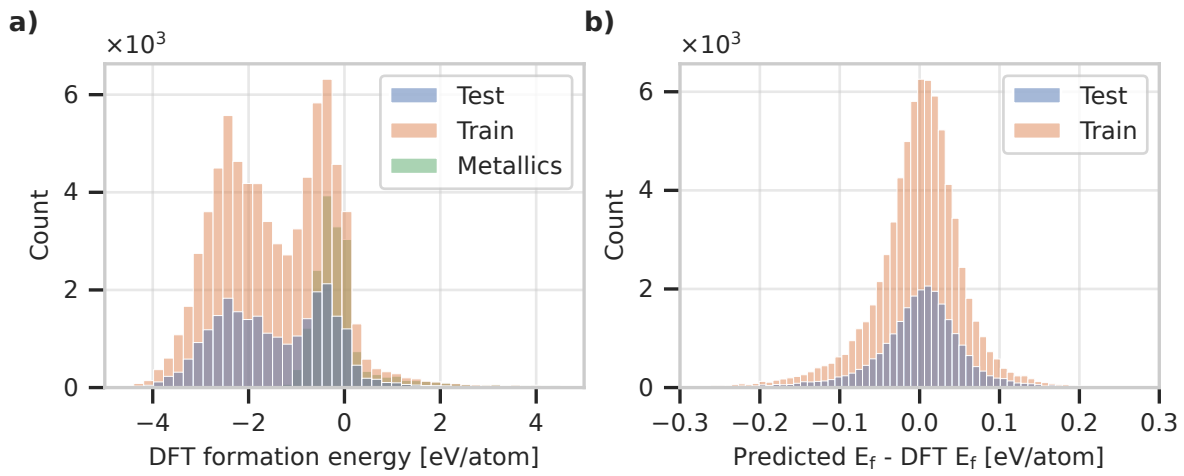

FIG. S2: a) Distribution of the DFT formation energy of crystal samples in the train and test data sets. The formation energy distribution of metallic samples in the train and test set is highlighted. b) Distribution of the formation energy error (predicted formation energy - DFT formation energy) over the train and test sets.

	Hull Count	Binary pairs MAE [eV/atom]	Avg.±Std. MAE [eV/atom]	Median Shape Error%	Avg. Hull Acc.%
Ceramic	785	0.069	0.113 ± 0.043	-5	43
Semiconductor	14	0.080	0.095 ± 0.031	-32	59
Metallic	1,867	0.045	0.053 ± 0.029	2	42
Semimetallic	394	0.045	0.079 ± 0.029	-0	56
Nonmetallic	55	0.094	0.153 ± 0.042	-22	21
All	3,115	0.054	0.073 ± 0.043	-2	44

TABLE S1: Statistics of various error metrics across the 3,115 binary convex hull predictions in this study. The second column, binary pairs MAE, reports the mean absolute error (MAE) of formation energy for each class of materials, which is computed over all the binary systems of the corresponding class. The third, fourth and fifth columns report the average formation energy MAE, median of the shape error of the convex hull, and average of the hull accuracy, respectively. MAE, shape error, and hull accuracy values are computed for individual binary systems and are subsequently averaged over binary systems belonging to each material class.

Element	Convex Hull Count	Binary pairs MAE [eV/atom]	Avg. MAE [eV/atom]	Median Shape Error%	Element	Convex Hull Count	Binary pairs MAE [eV/atom]	Avg. MAE [eV/atom]	Median Shape Error%	Element	Convex Hull Count	Binary pairs MAE [eV/atom]	Avg. MAE [eV/atom]	Median Shape Error%
H	78	0.061	0.132	-2.8	As	80	0.049	0.074	4.4	Sm	77	0.039	0.054	-3.5
Li	77	0.041	0.047	-6.2	Se	82	0.059	0.073	-3.1	Eu	74	0.049	0.051	1.7
Be	73	0.054	0.077	-25.2	Br	80	0.057	0.088	-4.7	Gd	65	0.115	0.069	-5.0
B	81	0.081	0.112	-15.9	Kr	1	0.042	0.205	-100.0	Tb	76	0.037	0.054	-2.6
C	76	0.101	0.178	16.5	Rb	76	0.058	0.044	-3.8	Dy	75	0.03	0.053	-0.1
N	83	0.096	0.118	-7.6	Sr	73	0.04	0.061	-1.1	Ho	75	0.034	0.056	1.1
O	84	0.069	0.106	-4.3	Y	75	0.039	0.05	-1.9	Er	72	0.036	0.056	1.2
F	84	0.073	0.178	-14.8	Zr	78	0.057	0.06	-2.1	Tm	74	0.04	0.062	0.4
Na	79	0.051	0.047	2.8	Nb	65	0.071	0.075	10.7	Yb	83	0.061	0.059	1.7
Mg	83	0.028	0.048	-5.3	Mo	58	0.074	0.088	34.9	Lu	73	0.044	0.067	0.4
Al	81	0.041	0.047	4.2	Tc	60	0.08	0.073	-6.6	Hf	70	0.059	0.072	3.2
Si	82	0.046	0.063	4.7	Ru	69	0.068	0.071	7.3	Ta	67	0.08	0.107	11.6
P	79	0.068	0.079	-2.5	Rh	79	0.056	0.087	0.9	W	54	0.084	0.119	5.4
S	82	0.067	0.132	-6.5	Pd	80	0.049	0.059	-2.0	Re	72	0.093	0.065	-25.3
Cl	83	0.055	0.12	-12.8	Ag	75	0.046	0.051	-6.9	Os	66	0.072	0.079	14.4
K	79	0.045	0.048	-3.3	Cd	72	0.037	0.052	-3.3	Ir	78	0.065	0.065	2.6
Ca	63	0.042	0.068	0.6	In	76	0.036	0.043	1.2	Pt	79	0.055	0.071	4.3
Sc	68	0.043	0.077	-3.1	Sn	80	0.046	0.045	-0.7	Au	78	0.046	0.063	2.7
Ti	70	0.054	0.063	2.7	Sb	78	0.046	0.114	-3.5	Hg	73	0.039	0.034	-5.0
V	67	0.053	0.064	-4.3	Te	78	0.059	0.09	-7.5	Tl	76	0.044	0.049	8.5
Cr	65	0.064	0.093	37.2	I	81	0.057	0.095	-3.8	Pb	76	0.047	0.059	7.3
Mn	72	0.057	0.062	1.7	Xe	2	0.27	0.151	-76.1	Bi	75	0.047	0.056	4.1
Fe	72	0.055	0.072	-5.2	Cs	51	0.051	0.042	-6.0	Ac	57	0.068	0.078	-2.8
Co	74	0.065	0.06	3.5	Ba	64	0.057	0.084	-1.3	Th	81	0.057	0.073	3.8
Ni	78	0.051	0.073	1.4	La	75	0.037	0.06	-1.8	Pa	61	0.081	0.063	0.6
Cu	75	0.043	0.058	-10.2	Ce	71	0.071	0.075	16.0	U	70	0.076	0.102	-1.4
Zn	82	0.036	0.045	-2.5	Pr	78	0.034	0.046	-2.2	Np	70	0.084	0.093	4.2
Ga	83	0.039	0.043	-2.8	Nd	78	0.036	0.058	-2.2	Pu	66	0.117	0.133	0.2
Ge	83	0.052	0.055	7.6	Pm	66	0.035	0.053	-1.3					

TABLE S2: Element-wise analysis of various error metrics across 3,115 binary convex hull predictions in this study. The binary pairs mean absolute error (MAE) reports the formation energy error over all the binary pairs associated with each element. The average formation energy MAE, on the other hand, reports the formation energy MAE averaged over different binary systems that the corresponsing element can make. Similarly, the median convex hull shape error is computed over individual binary sytems associated with each element.

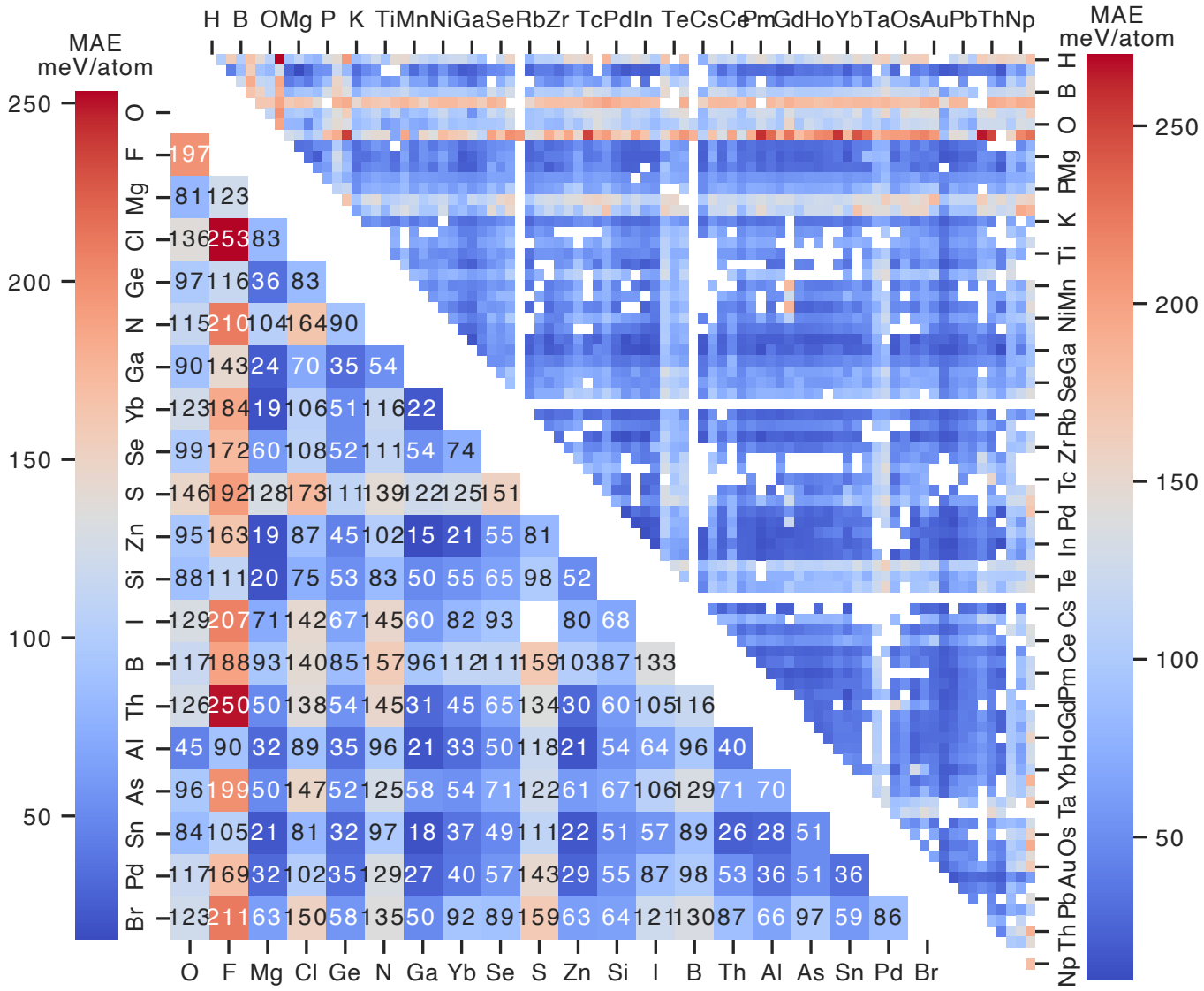

FIG. S3: Top right: The formation energy MAE for different pairs of all the binary compounds of this study. For labels of different rows refer to Table S2 rows which have the same order as this plot. Bottom left: The formation energy MAE for the most frequently appearing elements in the binary compounds of this study.

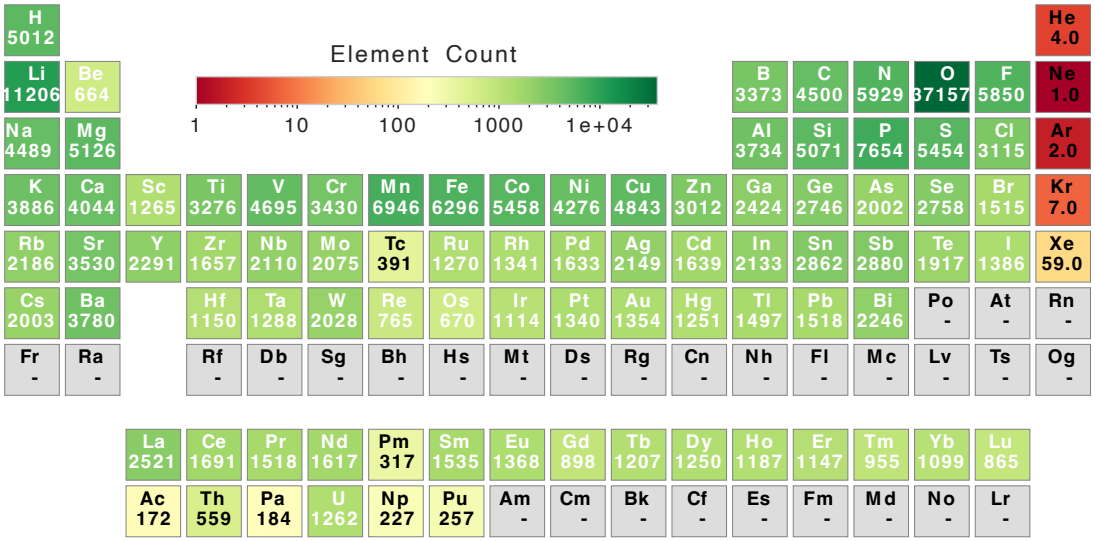

FIG. S4: Number of the elements in our training set from the Materials Project Database (v2021.05.13). Each appearance of an element in a crystal structure, regardless of its stoichiometry, counts as one.

	Number of structures	MAE [eV/atom]	Positional distance	Adjacency distance	Hull point match
Pt	1	0.108	0	0	Agree
$TmPt_3$	1	0.043	0	0	Agree
$TmPt_2$	1	0.033	0	0	Agree
Tm_3Pt_4	1	0.018	0	0	Agree
TmPt	1	0.062	0	0	Agree
Tm_5Pt_4	1	0.005	0	0	Agree
Tm_5Pt_3	1	0.022	0	0	Disagree
Tm_2Pt	1	0.022	0	0	Agree
Tm_3Pt	1	0.021	0	0	Agree
Tm_4Pt	1	0.030	0	0	Agree
Tm	4	0.054	4	2	Agree

TABLE S3: The formation energy MAE, positional and adjacency distances, and hull match for different compositions of Pt-Tm.

	Number of structures	MAE [eV/atom]	Positional distance	Adjacency distance	Hull point match
Cl_2	3	0.294	4	2	Disagree
$NaCl_7$	1	0.002	0	0	Disagree
$NaCl_3$	2	0.043	0	0	Agree
NaCl	3	0.031	0	0	Agree
Na_3Cl_2	2	0.038	2	1	Agree
Na_2Cl	4	0.037	2	2	Disagree
Na_3Cl	1	0.026	0	0	Agree
Na	13	0.029	22	9	Disagree

TABLE S4: The formation energy MAE, positional and adjacency distances, and hull match for different compositions of Na-Cl

	Number of structures	MAE [eV/atom]	Positional distance	Adjacency distance	Hull point match
N_2	14	0.148	8	10	Disagree
Tb_2N_3	1	0.024	0	0	Agree
TbN	1	0.076	0	0	Agree
Tb	4	0.035	6	3	Disagree

TABLE S5: The formation energy MAE, positional and adjacency distances, and hull match for different compositions of Tb-N

	Number of structures	MAE [eV/atom]	Positional distance	Adjacency distance	Hull point match
Se	10	0.073	10	6	Disagree
V_2Se_9	1	0.007	0	0	Agree
VSe_2	1	0.076	0	0	Agree
V_5Se_8	1	0.051	0	0	Agree
V_3Se_4	1	0.006	0	0	Disagree
VSe	3	0.083	0	0	Agree
V_5Se_4	1	0.018	0	0	Agree
V	2	0.053	0	0	Agree

TABLE S6: The formation energy MAE, positional and adjacency distances, and hull match for different compositions of V-Se

	Number of structures	MAE [eV/atom]	Positional distance	Adjacency distance	Hull point match
Y	3	0.011	2	2	Agree
NdY_3	3	0.024	4	1	Disagree
Nd_3Y	2	0.015	2	1	Disagree
Nd	3	0.052	4	2	Disagree

TABLE S7: The formation energy MAE, positional and adjacency distances, and hull match for different compositions of Nd-Y

	Number of structures	MAE [eV/atom]	Positional distance	Adjacency distance	Hull point match
Se	10	0.073	10	6	Disagree
$MoSe_2$	4	0.032	6	3	Disagree
Mo_3Se_4	2	0.164	0	0	Agree
Mo_9Se_{11}	2	0.078	0	0	Disagree
MoSe	1	0.200	0	0	Agree
Mo_3Se	1	0.055	0	0	Agree
Mo	7	0.104	8	4	Agree

TABLE S8: The formation energy MAE, positional and adjacency distances, and hull match for different compositions of MoSe

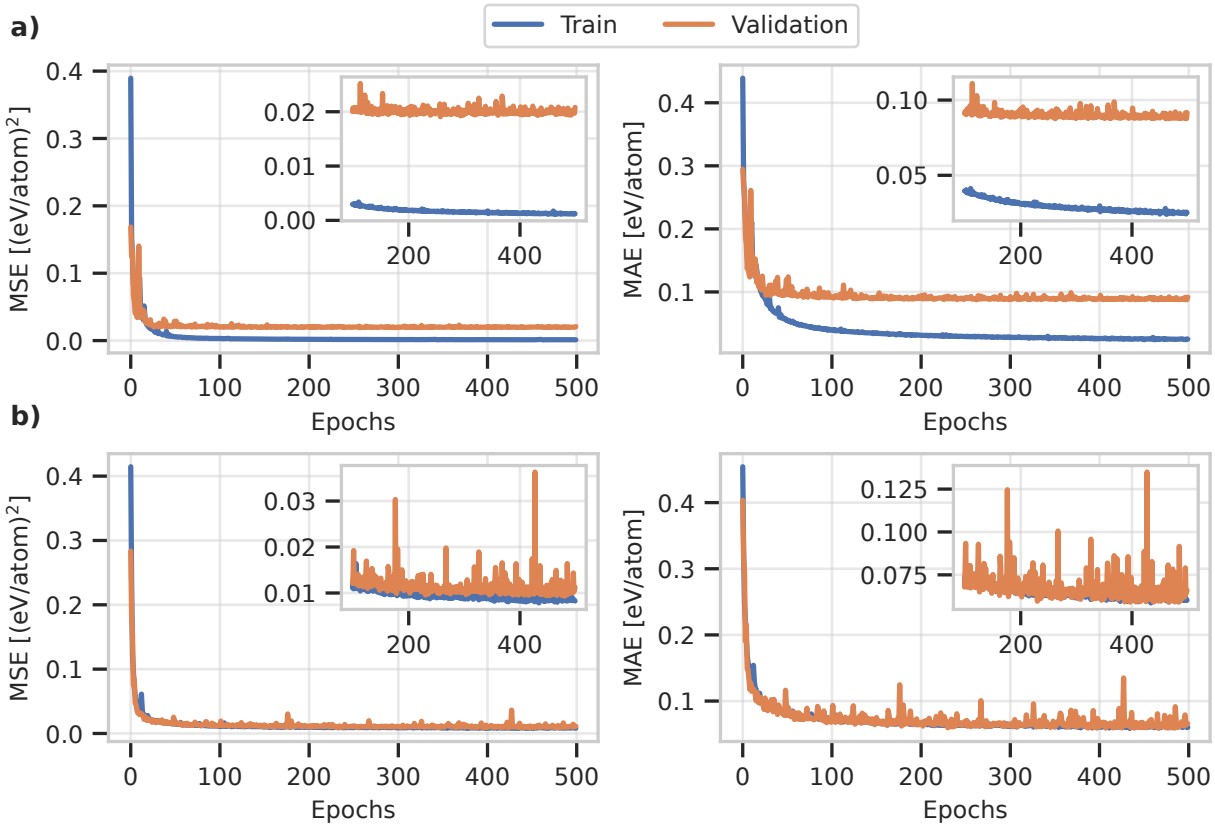

FIG. S5: The mean squared error (MSE) - the loss function - and the mean absolute error (MAE) as a function of the epoch number during training of the model shown for both the training and validation data a) without data augmentation and b) with data augmentation.

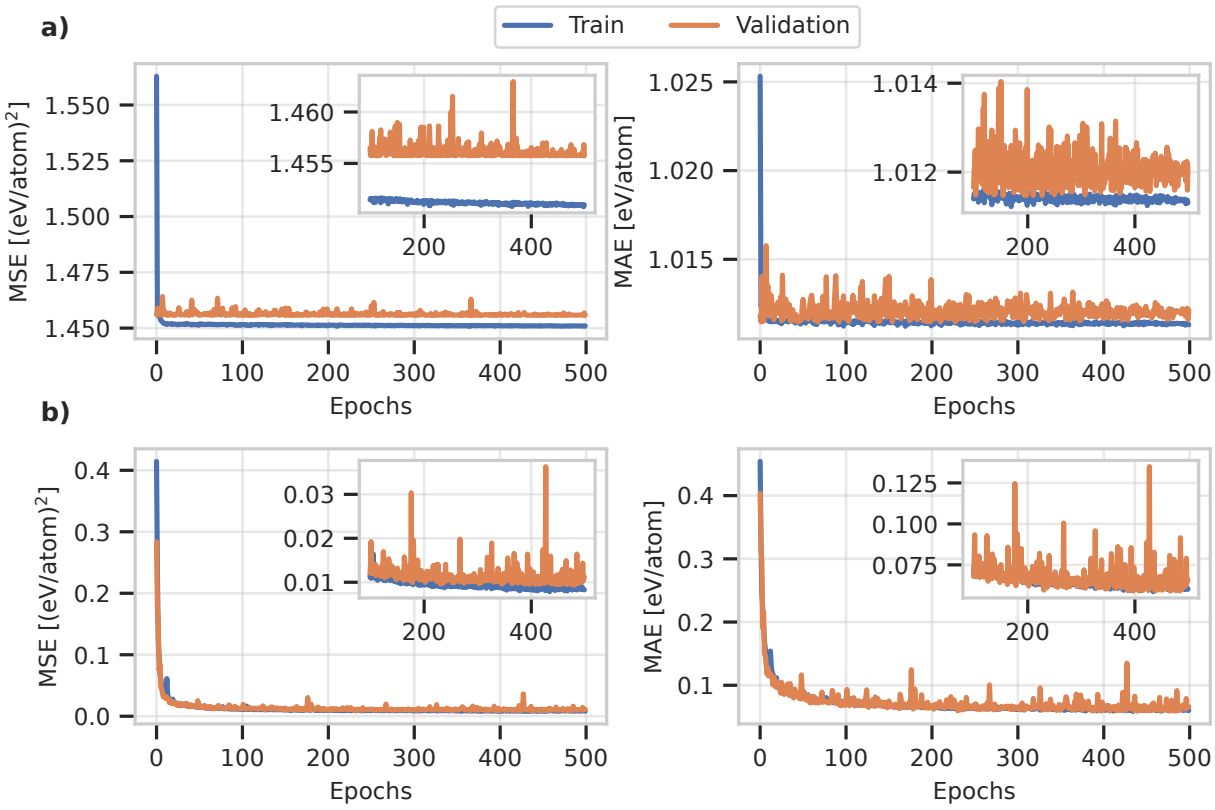

FIG. S6: The mean squared error (MSE) - the loss function - and the mean absolute error (MAE) as a function of the epoch number during training of the model shown for the training and validation data a) without skip connections and b) with skip connections incorporated into the CNN.

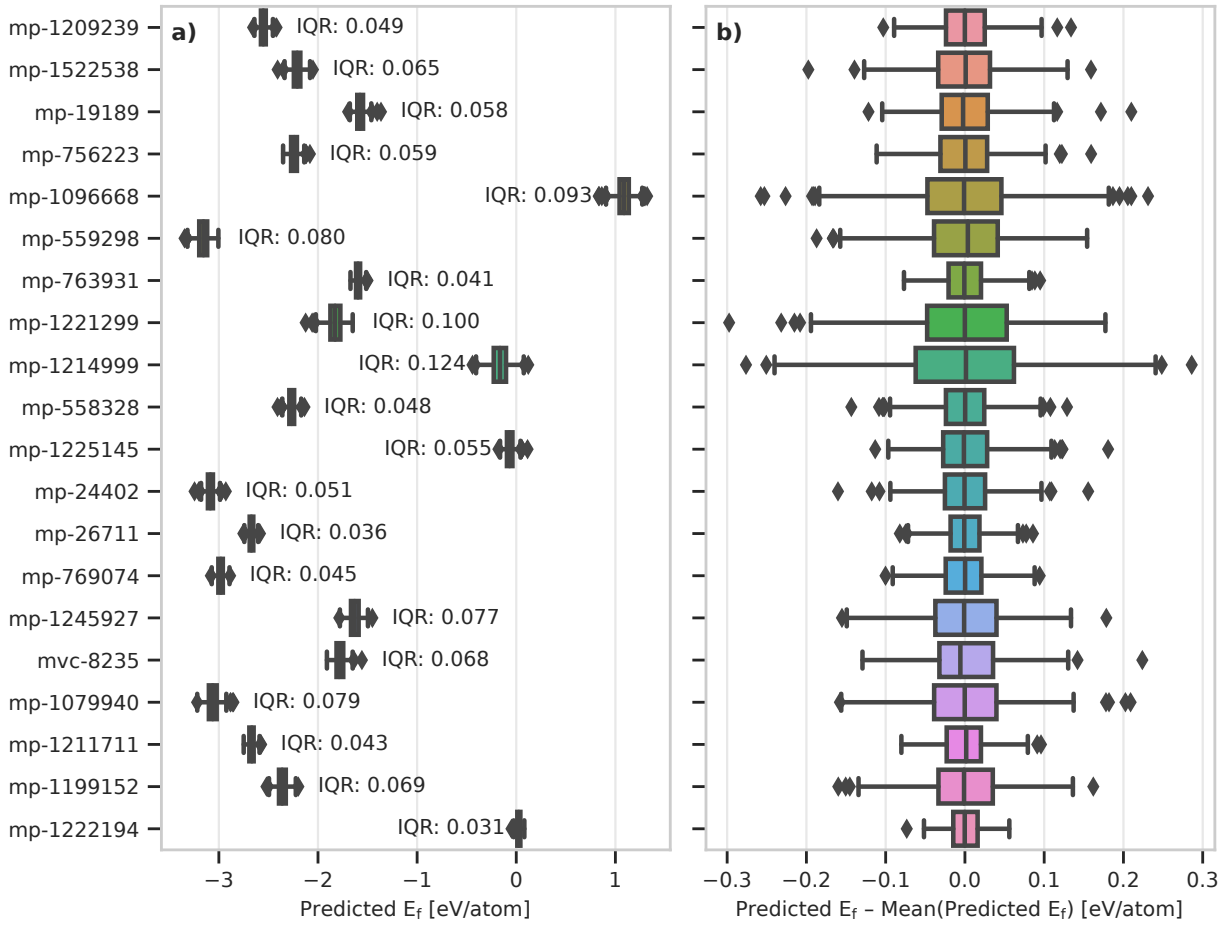

FIG. S7: Assessing the approximate rotation invariance of the formation energy prediction once no ensemble averaging is employed. a) The formation energy prediction span over 500 rotated instances of each crystal sample identified by its Materials Project's number. b) Box and whisker representation of the formation energy prediction spread (i.e., predicted E_f - mean (predicted E_f)) among 500 rotations of each crystal sample.

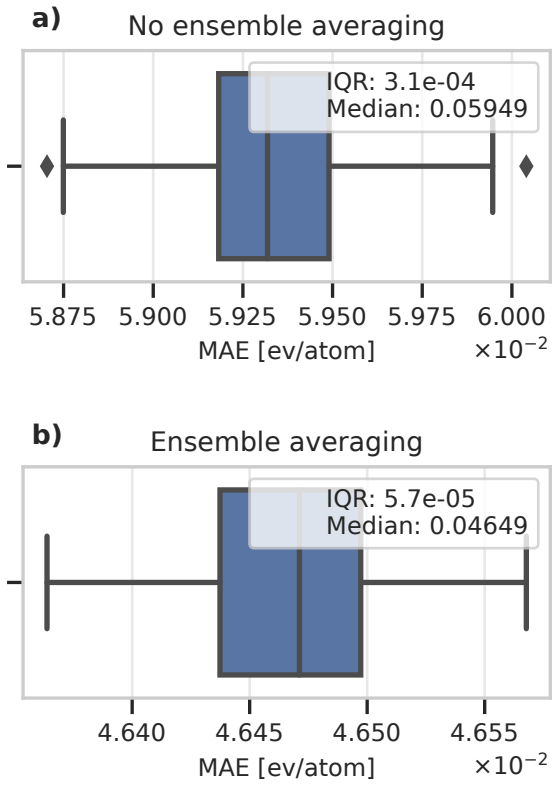

FIG. S8: The variation of the MAE between model's prediction and DFT values over 50 rotated instances of the test data a) when no ensemble averaging is employed and b) when ensemble-averaging is employed.

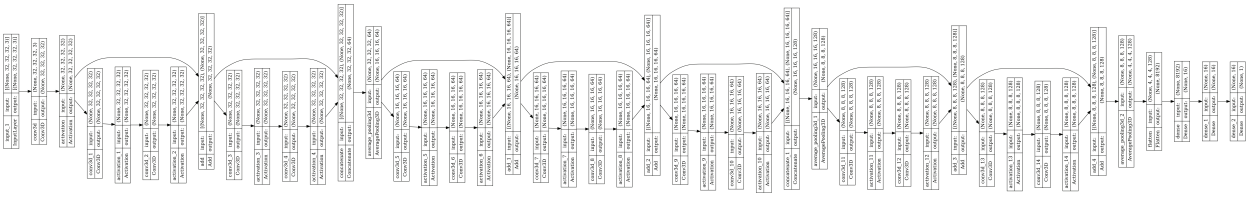

FIG. S9: The detailed architecture of the 15-convolutional-layer deep network input into a fully connected neural network used in this study. Each residual block consists of two convolution and activation layers with a skip connection (shown as a curved arrow) bypassing the output of the previous block to the next block.