Electronic Supplementary Material (ESI) for Digital Discovery.

Supplementary Information for

Accelerating the Evaluation of Crucial Descriptors for Catalyst Screening via Message Passing Neural Network

Hieu A. Doan*[,a] Chenyang Li,[a] Logan Ward,[b] Mingxia Zhou,[a,c] Larry A. Curtiss[a] and Rajeev S. Assary *[a]

[a.] *Materials Science Division, Argonne National Laboratory, Lemont, Illinois 60439, United States.*

[b.] *Data Science and Learning Division, Argonne National Laboratory, Lemont, Illinois 60439, United States.*

[c.] *Present address: State Key Laboratory of Heavy Oil Processing, China University of Petroleum Beijing, Beijing 102249, China.*

*Corresponding authors: HAD(hadoan@anl.gov), RSA (assary@anl.gov), 630-252-3536

All relevant Python codes and instructions for running our LCG-MPNN models are provided on Github at https://github.com/MolecularMaterials/nfp. Data containing ~ 20,000 geometry optimizations using VASP were saved as Atomic Simulation Environment databases and can be downloaded from the Materials Data Facility using the following link: https://acdc.alcf.anl.gov/mdf/detail/doan_datasets_accelerating_representations_v1.1/

Table of Content

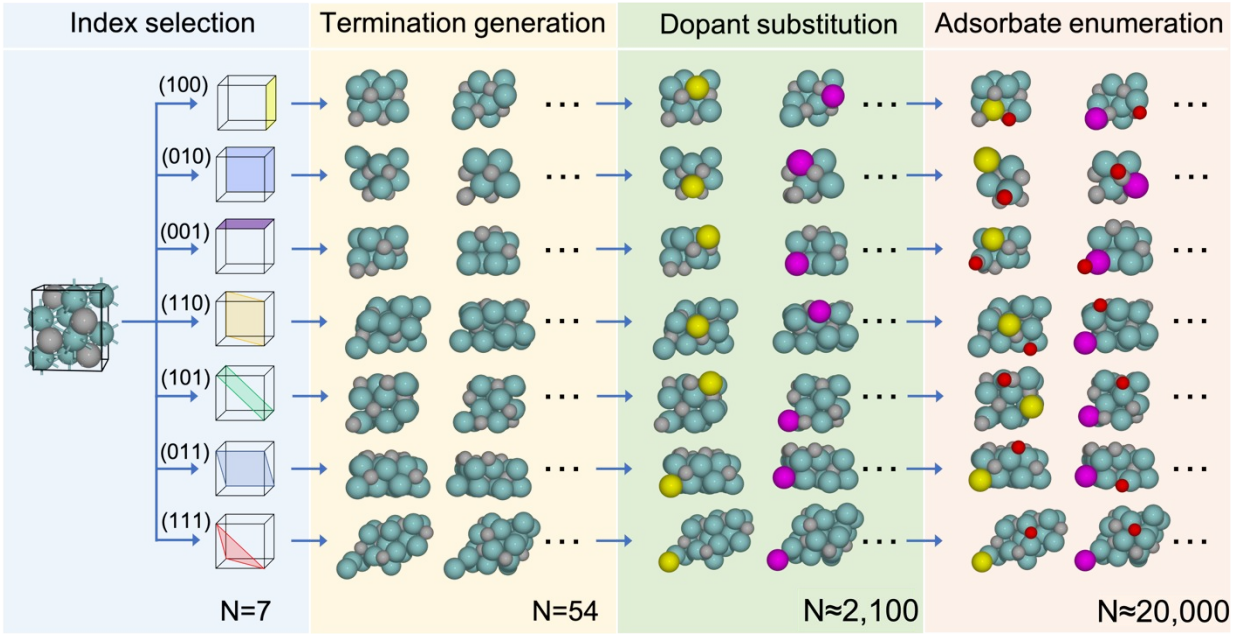

Figure S1 Scheme for generating oxygen binding geometries on Mo_2C surfaces. Top view of example surfaces is shown. N indicates the number of possible configurations (e.g., 7 low Miller-index surfaces were selected and cut from the bulk unit cell of orthorhombic Mo_2C).The atom color code is green for Mo, grey for C, yellow/magenta for a dopant atom, and red for O.

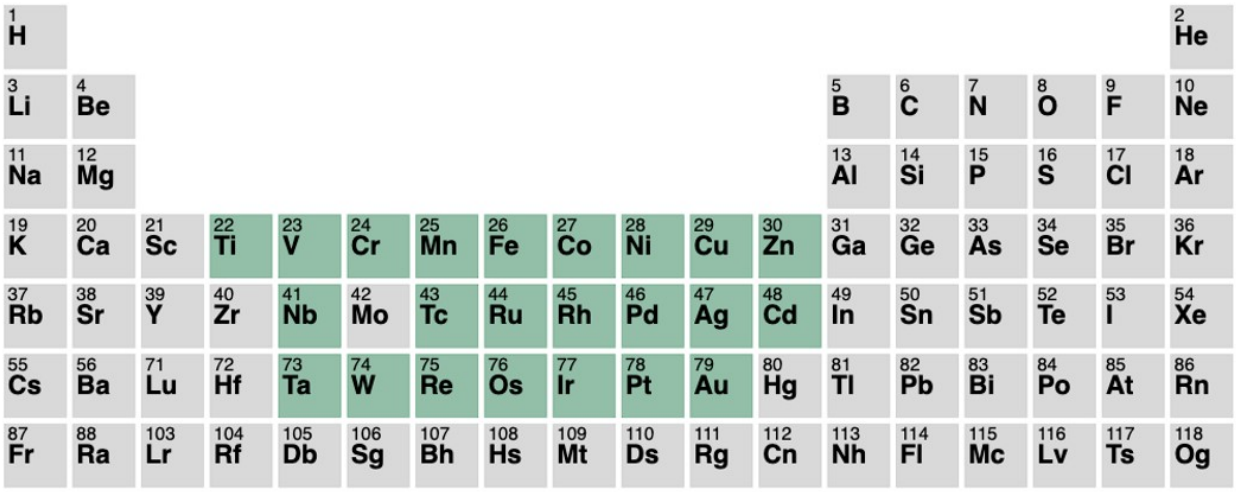

Figure S2 List of dopant elements (highlighted in green) used in this work

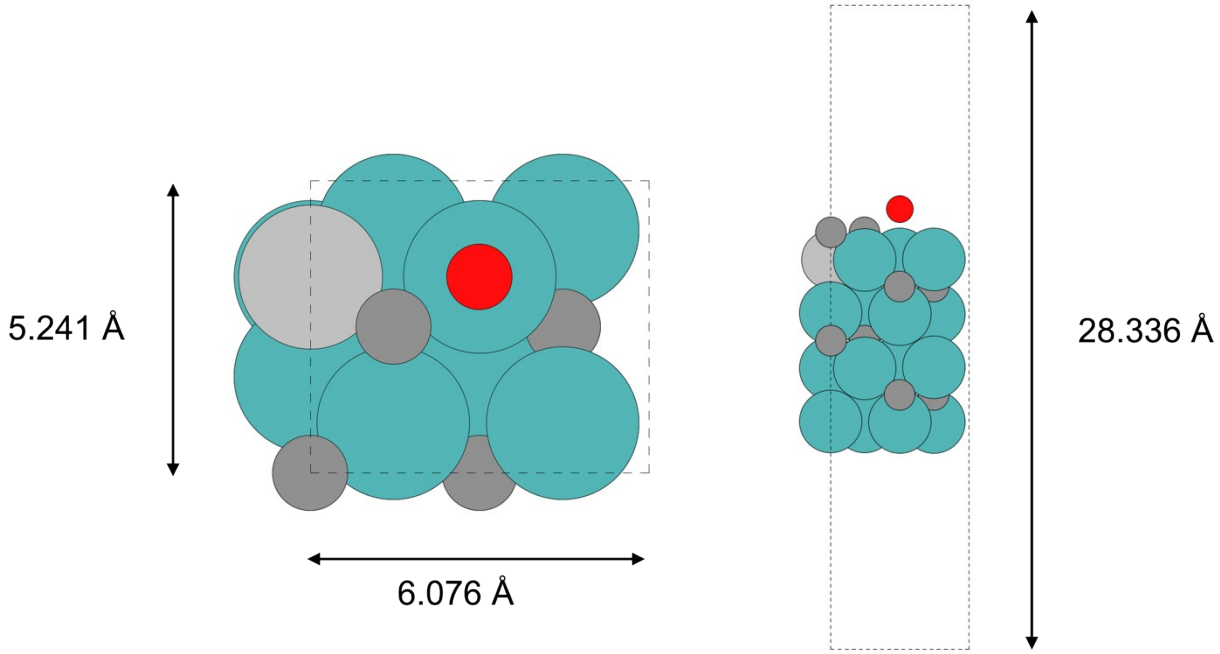

Figure S3 The top (left) and side (right) view of a typical (1x1) surface unit cell used in this work. In this particular example, the $Mo_2C(100)$ facet is doped with a Ag atom. The Mo, Ag, and C atoms are shown in green, light gray, and dark gray, respectively.

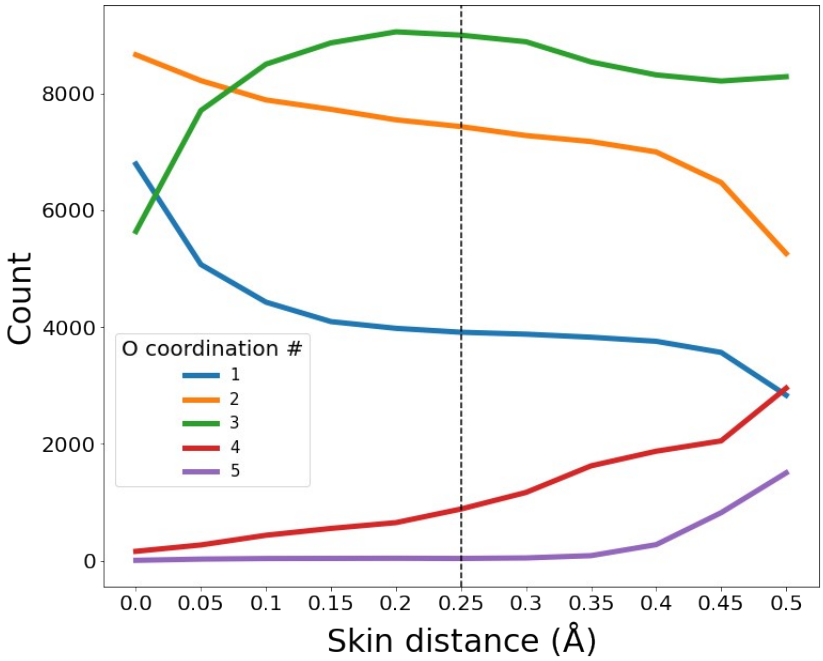

Figure S4 Skin distance analysis to determine the appropriate cut-off value for bonded interaction among atoms

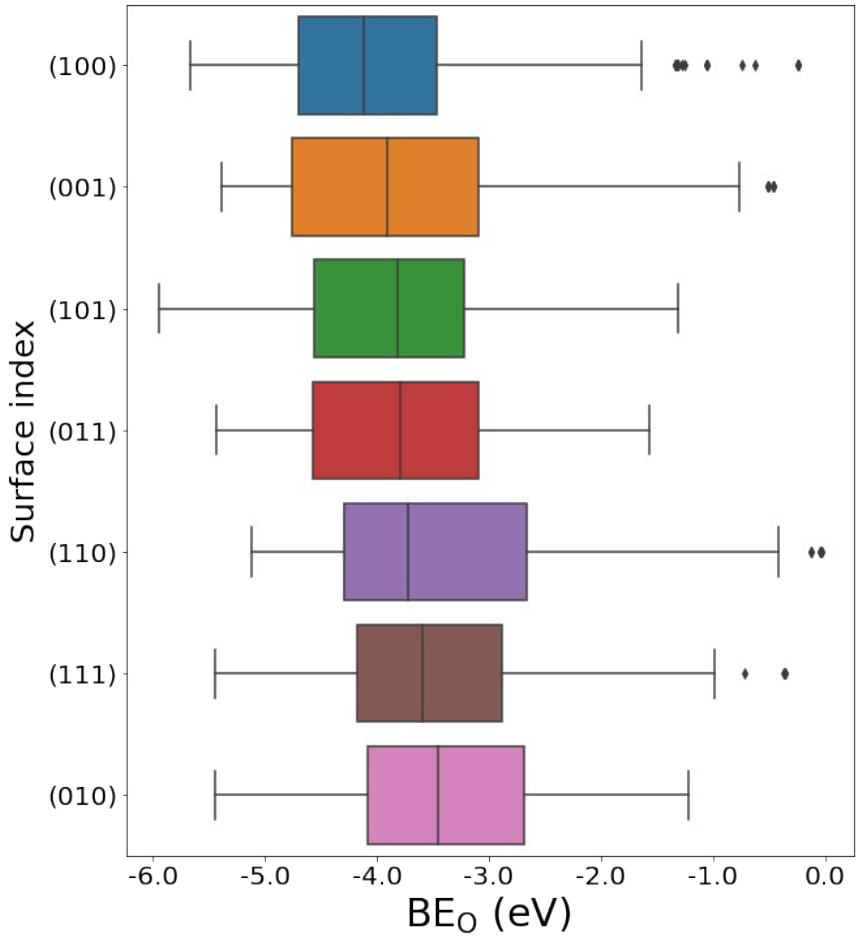

Figure S5 DFT computed oxygen binding energy (BE_O) distribution with respect to seven low Miller-index surfaces of pristine and doped Mo_2C

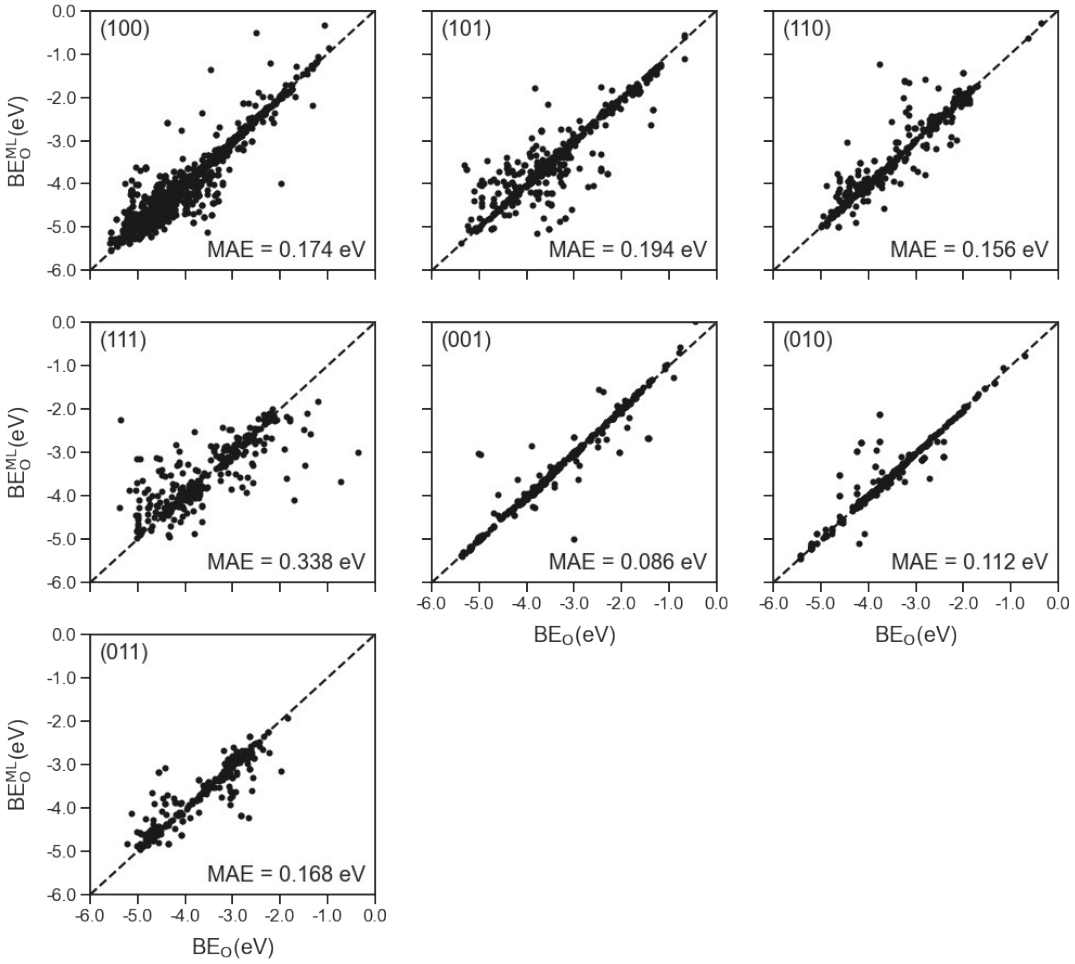

Figure S6 Prediction accuracy, measured as MAE between BE_O^{ML} and BE_O, of the LCG-MPNN model with respect to seven low Miller-index surfaces of pristine and doped Mo_2C in the test set

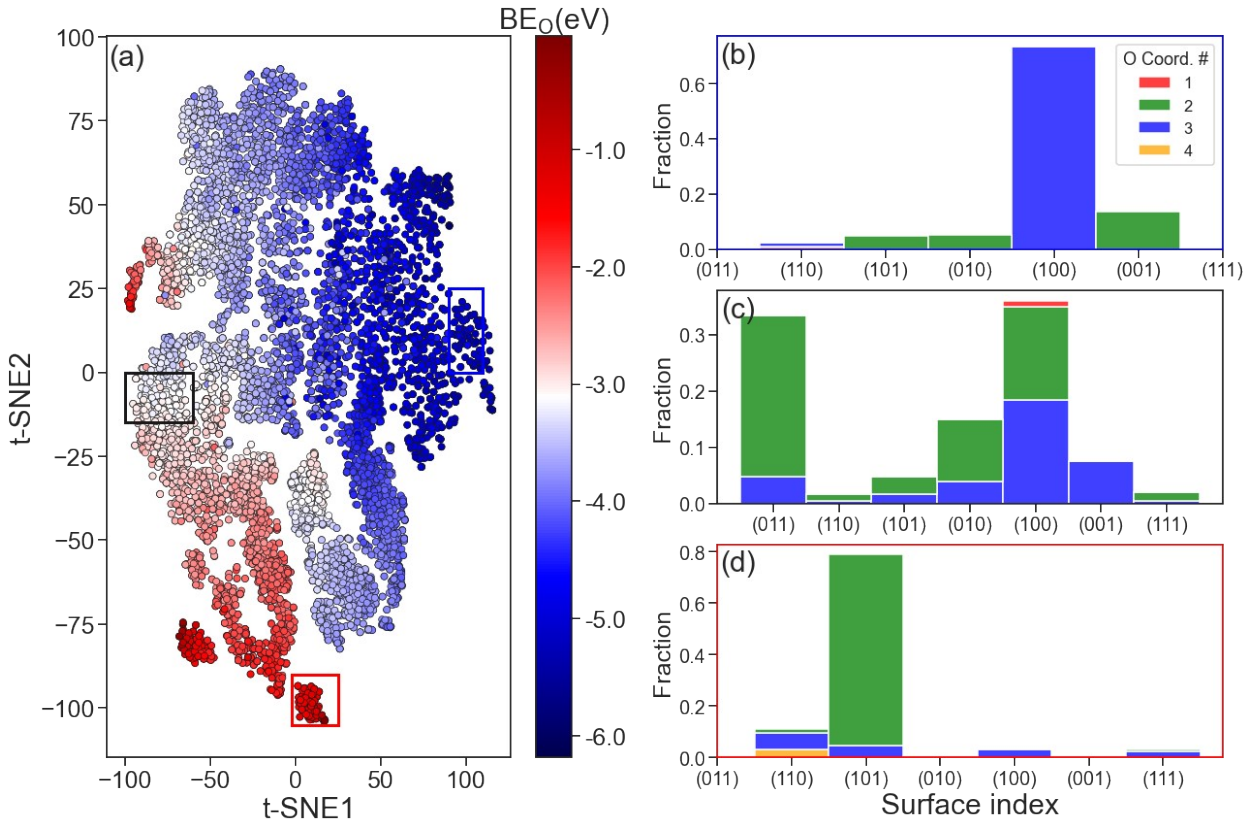

Figure S7 a) 2D t-distributed Stochastic Neighbor Embedding (t-SNE) plot of graph-level features from the readout layer. Each point represents an adsorption geometry from the training set, of which color was mapped to the computed oxygen binding energy (BE_O). The oxygen coordination number distribution of adsorption geometries in the regions enclosed by blue, black, and red rectangles are shown in (b), (c), and (d), respectively.

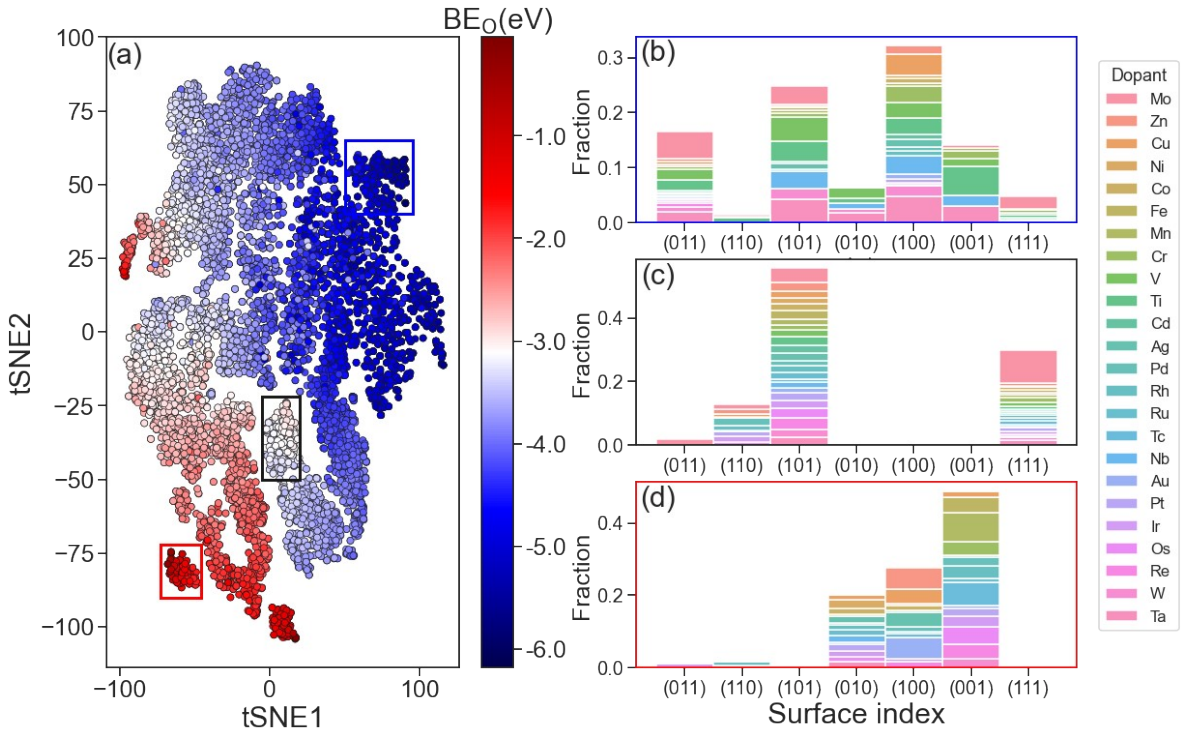

Figure S8 a) 2D t-distributed Stochastic Neighbor Embedding (t-SNE) plot of graph-level features from the readout layer. Each point represents an adsorption geometry from the training set, of which color was mapped to the computed oxygen binding energy (BE_O). The dopant element distribution of adsorption geometries in blue, black, and red regions are shown in (b), (c), and (d), respectively. The selected regions in this plot are identical to the ones in Fig. 5 of the main text.

Table S1 Hyperpameter optimization using 200 trials of Tree Parzen Estimator method.[1] The best set of hyperparameters was found in trial #118.

Trial #	Batch size	Dense 1	Dense 2	Dense 3	Learning rate	No. of message passing steps	Read-out	Validation MAE (eV)
1	50	128	64	32	0.0010	3	sum	0.198
2	50	32	32	32	0.0003	9	max	0.207
3	50	256	64	256	0.0004	9	mean	0.193
4	80	128	256	128	0.0005	7	mean	0.180
5	60	512	256	128	0.0004	5	set2set	0.181
6	50	256	256	128	0.0041	7	mean	0.748
7	50	128	256	512	0.0008	5	set2set	0.185
8	90	64	512	128	0.0009	5	max	0.182
9	60	512	128	128	0.0080	3	sum	0.484
10	70	64	512	512	0.0300	7	mean	21.519
11	80	128	512	256	0.0026	5	max	0.194
12	70	128	64	256	0.0010	9	set2set	0.170
13	80	32	32	64	0.0042	9	max	0.700
14	70	64	256	32	0.0004	7	set2set	0.183
15	70	64	256	32	0.0125	7	max	0.627
16	60	64	32	32	0.0041	5	set2set	0.542
17	50	64	512	256	0.0001	7	mean	0.200
18	80	128	512	512	0.0001	7	set2set	0.219
19	100	64	256	128	0.0478	3	max	0.411
20	60	32	64	32	0.0019	9	mean	0.186
21	90	128	64	512	0.0002	3	sum	0.193
22	100	256	128	64	0.0950	3	sum	0.430
23	60	32	64	32	0.0017	9	sum	0.182
24	90	128	64	512	0.0002	3	sum	0.193
25	100	256	128	64	0.0951	3	sum	0.651
26	60	32	64	32	0.0015	9	sum	0.182
27	90	128	64	512	0.0002	3	sum	0.195
28	100	256	128	64	0.0098	3	sum	0.529
29	50	32	64	32	0.0015	9	sum	0.182
30	90	128	64	512	0.0001	3	sum	0.205
31	100	256	128	64	0.0103	3	sum	0.519
32	50	32	32	32	0.0006	9	sum	0.182
33	90	512	64	512	0.0001	3	sum	0.205
34	100	256	128	64	0.0221	3	sum	0.500
35	50	32	32	32	0.0006	9	sum	0.182

36	60	512	64	32	0.0024	9	mean	0.178
37	90	128	64	512	0.0002	3	sum	0.193
38	100	64	256	128	0.0674	3	max	7.438
39	100	256	128	128	0.0530	3	max	0.389
40	100	256	256	128	0.0299	5	max	0.490
41	90	64	256	128	0.0176	3	max	0.465
42	90	512	256	128	0.0452	3	max	668.718
43	80	64	256	256	0.0041	5	max	0.240
44	90	128	64	512	0.0005	3	sum	0.187
45	90	128	64	512	0.0002	3	set2set	0.197
46	90	128	64	512	0.0003	5	mean	0.197
47	90	128	512	512	0.0010	7	sum	0.174
48	70	128	32	512	0.0003	3	set2set	0.200
49	90	64	512	256	0.0064	5	max	0.590
50	80	64	512	128	0.0012	5	max	0.187
51	70	512	512	128	0.0003	5	mean	0.191
52	90	64	512	256	0.0008	5	max	0.179
53	90	64	512	128	0.0063	5	set2set	0.595
54	60	512	256	128	0.0008	5	set2set	0.173
55	60	512	512	128	0.0031	5	set2set	0.227
56	60	512	256	128	0.0023	5	set2set	0.184
57	80	512	256	128	0.0005	7	mean	0.186
58	80	512	256	128	0.0001	7	mean	0.210
59	80	512	256	128	0.0006	7	mean	0.176
60	80	128	256	256	0.0004	7	mean	0.178
61	60	512	256	64	0.0012	7	set2set	0.198
62	80	32	32	128	0.0001	7	mean	0.222
63	60	128	256	128	0.0019	7	set2set	0.183
64	70	512	256	64	0.0002	7	mean	0.199
65	70	128	32	256	0.0051	9	set2set	0.667
66	70	128	256	256	0.0009	9	set2set	0.181
67	70	128	64	32	0.0024	9	mean	0.189
68	70	128	64	32	0.0032	9	mean	0.725
69	50	128	64	256	0.0033	9	mean	0.716
70	60	128	64	32	0.0012	9	mean	0.177
71	70	32	64	32	0.0007	9	mean	0.182
72	80	256	64	256	0.0018	9	mean	0.192
73	60	128	128	32	0.0142	9	mean	0.634
74	50	512	64	256	0.0081	9	mean	0.661

75	80	128	64	32	0.0006	9	set2set	0.176
76	70	128	64	256	0.0015	9	mean	0.176
77	60	512	128	32	0.0022	9	mean	0.189
78	80	256	32	256	0.0010	7	set2set	0.174
79	70	32	64	64	0.0003	9	mean	0.189
80	50	128	64	32	0.0048	9	set2set	0.674
81	60	128	128	256	0.0005	7	mean	0.190
82	80	512	64	32	0.0027	9	mean	0.183
83	70	256	32	256	0.0013	7	set2set	0.180
84	100	128	64	64	0.0036	9	mean	0.715
85	90	128	512	512	0.0011	7	sum	0.181
86	90	512	512	512	0.0058	9	sum	0.659
87	60	32	512	512	0.0076	7	sum	1.425
88	90	128	512	512	0.0007	9	sum	0.180
89	60	512	512	512	0.0113	7	sum	0.799
90	70	128	512	512	0.0021	9	sum	0.178
91	50	64	512	512	0.0009	7	set2set	0.181
92	90	256	64	32	0.0027	9	sum	0.189
93	100	512	256	128	0.0004	5	set2set	0.187
94	60	32	512	256	0.0003	5	set2set	0.183
95	70	128	128	512	0.0008	5	set2set	0.178
96	90	512	256	128	0.0002	5	set2set	0.206
97	60	128	32	64	0.0014	5	set2set	0.196
98	70	64	512	256	0.0005	5	set2set	0.178
99	50	512	256	128	0.0006	5	set2set	0.202
100	100	128	256	512	0.0017	5	set2set	0.185
101	90	256	512	256	0.0004	7	max	0.195
102	60	32	128	128	0.0003	5	set2set	0.183
103	70	128	32	512	0.0008	7	sum	0.188
104	90	512	256	64	0.0011	5	set2set	0.171
105	60	128	512	128	0.0001	7	set2set	0.208
106	70	64	256	256	0.0041	5	sum	0.328
107	50	512	128	512	0.0007	3	max	0.180
108	100	128	32	128	0.0002	7	set2set	0.198
109	90	256	512	256	0.0019	5	sum	0.176
110	80	256	32	256	0.0009	7	set2set	0.188
111	80	256	32	256	0.0002	7	set2set	0.297
112	80	256	32	256	0.0037	3	set2set	0.205
113	80	256	32	256	0.0029	5	set2set	0.189

114	80	256	32	256	0.0159	7	set2set	0.622
115	80	256	32	256	0.0218	5	set2set	0.709
116	80	256	32	256	0.0048	3	set2set	0.243
117	60	256	32	256	0.0014	7	set2set	0.189
118	**70**	**256**	**32**	**256**	**0.0010**	**9**	**set2set**	**0.168**
119	60	32	256	64	0.0001	5	set2set	0.220
120	80	512	256	128	0.0075	7	max	0.672
121	70	64	32	256	0.0003	9	set2set	0.218
122	50	256	64	256	0.0005	3	set2set	0.185
123	60	512	128	128	0.0003	5	set2set	0.181
124	100	32	256	256	0.0002	7	max	0.199
125	80	256	64	64	0.0088	9	set2set	0.578
126	70	512	32	128	0.0020	5	set2set	0.177
127	60	64	64	256	0.0364	9	set2set	0.857
128	70	512	256	128	0.0004	7	set2set	0.181
129	80	256	128	256	0.0007	3	max	0.186
130	100	32	32	64	0.0015	9	set2set	0.180
131	50	512	64	256	0.0001	5	set2set	0.217
132	60	256	256	128	0.0055	7	set2set	0.632
133	70	64	64	256	0.0024	9	set2set	0.180
134	80	512	256	128	0.0012	5	max	0.173
135	60	256	32	256	0.0782	7	set2set	0.813
136	70	32	128	64	0.0006	3	set2set	0.186
137	90	512	256	64	0.0004	5	set2set	0.180
138	90	512	256	64	0.0011	5	set2set	0.173
139	90	512	256	64	0.0017	5	max	0.175
140	90	512	256	64	0.0010	5	set2set	0.194
141	90	512	256	64	0.0008	5	set2set	0.178
142	90	512	256	64	0.0030	5	set2set	0.185
143	100	512	256	64	0.0070	5	set2set	0.628
144	50	512	256	64	0.0002	5	set2set	0.191
145	60	512	256	64	0.0037	5	max	0.237
146	80	512	256	64	0.0010	5	set2set	0.171
147	90	512	64	32	0.0005	5	set2set	0.197
148	70	512	256	128	0.0021	5	set2set	0.169
149	80	128	32	256	0.0006	9	set2set	0.179
150	60	64	256	64	0.0014	5	set2set	0.185
151	70	256	64	256	0.0009	7	set2set	0.176
152	90	512	32	128	0.0002	9	set2set	0.287

153	50	128	128	256	0.0017	5	set2set	0.195
154	100	256	256	32	0.0012	7	set2set	0.240
155	80	32	64	128	0.0006	5	set2set	0.192
156	60	512	32	256	0.0003	9	set2set	0.375
157	70	128	256	64	0.0004	3	set2set	0.180
158	90	64	64	256	0.0024	7	set2set	0.176
159	80	256	256	128	0.0007	5	set2set	0.181
160	60	512	32	256	0.0044	9	set2set	0.722
161	70	128	128	64	0.0010	9	set2set	0.187
162	70	256	64	32	0.0016	9	max	0.184
163	70	512	256	128	0.0028	9	set2set	0.207
164	70	32	32	256	0.0003	9	set2set	0.189
165	70	128	256	256	0.0008	9	sum	0.178
166	70	256	64	64	0.0010	9	set2set	0.198
167	70	512	32	128	0.0034	9	set2set	0.744
168	70	64	256	256	0.0005	9	mean	0.183
169	70	512	128	32	0.0019	9	set2set	0.180
170	70	128	64	64	0.0097	9	max	0.654
171	70	256	32	256	0.0013	9	set2set	0.174
172	90	512	256	64	0.0011	9	sum	0.170
173	90	512	256	64	0.0022	5	set2set	0.184
174	90	32	64	64	0.0053	3	set2set	0.240
175	90	128	256	64	0.0062	9	set2set	0.575
176	90	256	32	64	0.0005	5	mean	0.182
177	90	64	512	64	0.0011	9	max	0.176
178	90	512	128	64	0.0006	5	set2set	0.181
179	90	128	256	64	0.0041	5	sum	0.256
180	90	512	64	64	0.0004	9	set2set	0.195
181	90	256	32	64	0.0015	3	set2set	0.177
182	80	32	256	64	0.0002	5	set2set	0.210
183	50	512	64	512	0.0025	9	set2set	0.653
184	100	128	256	64	0.0002	5	mean	0.200
185	70	256	512	128	0.0122	9	set2set	0.717
186	70	512	32	128	0.0019	5	max	0.195
187	70	64	128	256	0.0022	9	set2set	0.194
188	70	128	256	256	0.0032	3	sum	0.192
189	70	512	64	128	0.0222	5	set2set	102.544
190	70	256	32	512	0.0044	5	set2set	0.484
191	70	32	256	32	0.0067	9	set2set	0.701

192	70	512	256	256	0.0017	5	mean	0.185
193	70	128	64	128	0.0037	9	set2set	0.682
194	80	256	512	256	0.0009	5	max	0.195
195	70	512	32	256	0.0088	3	set2set	0.479
196	70	64	128	128	0.0028	9	sum	0.225
197	80	128	256	32	0.0021	5	set2set	0.174
198	70	512	64	512	0.0008	9	set2set	0.173
199	50	256	32	256	0.0013	5	set2set	0.177
200	100	32	256	128	0.0006	9	mean	0.191

References

(1) Bergstra, J.; Yamins, D.; Cox, D. D. Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures. In *ICML*; 2013; Vol. 28, pp 115–123. https://doi.org/10.1080/01459740.2015.1058375.