Supporting Information:

A Multitask Approach to Learn Molecular Properties

Zheng Tan[a,*], Yan Li[b], Weimei Shi[a], Shiqing Yang[a]

[a] Chengdu Polytechnic, 83 Tianyi Street, Chengdu, Sichuan 610000, P. R. China

[b] Xiyuan Quantitative Technology, 388 Yizhou Road, Chengdu, Sichuan 610000, P. R. China

[*] Corresponding author

E-mail address: zhengtan1983@yahoo.com (Zheng Tan)

BASE MODEL HYPERPARAMETERS

Table S1: Hyperparameters employed by each base regressor in QM9. Note that n_estimators and max depth denote the number of trees and the depth of the tree in xgboost respectively.

	xgboost			**one-layer net**	**two-layer net**	**three-layer net**	**GCN**
	n_estimators	max depth	learning rate	epoch	epoch	epoch	epoch
mu	350	10	0.2	180	100	80	8
alpha	100	8	0.1	350	180	360	25
homo	390	10	0.2	250	130	160	10
lumo	390	10	0.2	340	280	250	16
gap	390	10	0.15	330	170	170	4
r2	200	3	0.1	370	200	280	27
zpve	390	3	0.3	380	210	230	26
cv	60	3	0.1	340	180	340	24
u0	120	3	0.1	330	180	330	28
u298	120	3	0.1	320	240	390	28
h298	120	3	0.1	340	210	390	29
g298	110	3	0.1	370	180	320	29

Table S2: Hyperparameters employed by each base regressor in Alchemy.

	xgboost			**one-layer net**	**two-layer net**	**three-layer net**	**GCN**
	n_estimators	max depth	learning rate	epoch	epoch	epoch	epoch
mu	390	12	0.20	190	240	240	30
alpha	390	12	0.20	390	220	190	12
homo	390	12	0.15	280	210	350	11
lumo	390	12	0.20	270	350	370	30
gap	390	12	0.20	360	160	240	10
r2	390	12	0.20	290	280	320	24
zpve	390	12	0.20	400	220	260	29
cv	390	12	0.30	400	290	340	24
U0	390	12	0.20	370	340	370	6
U	390	12	0.20	400	190	150	18
H	390	12	0.15	380	90	120	10
G	390	12	0.20	390	190	210	21

Table S3: Hyperparameters employed by each base classifier in Tox21. Note that C and gamma denote the penalty parameter and RBF kernel coefficient in SVM

respectively.

	xgboost			SVM		RF		two-layer net
	n_estimators	max depth	learning rate	C	gamma	n_estimators	max depth	epoch
NR-AR	50	3	0.1	2^{3}	2^{-2}	10	3	210
NR-AR-LBD	120	3	0.01	2^{-4}	2^{-1}	330	3	150
NR-AhR	280	6	0.1	2^{-1}	2^{-5}	130	12	60
NR-Aromatase	50	10	0.1	2^{-5}	2^{-2}	360	5	300
NR-ER	70	3	0.1	2^{3}	2^{-4}	150	6	70
NR-ER-LBD	300	7	0.4	2^{2}	2^{-3}	40	8	130
NR-PPAR-gamma	50	1	0.2	2^{-4}	2	110	2	240
SR-ARE	120	3	0.1	2^{-3}	2^{-4}	390	11	70
SR-ATAD5	270	2	0.1	2^{-4}	2^{-5}	140	3	130
SR-HSE	50	5	0.1	2^{-2}	2^{-4}	10	3	360
SR-MMP	220	3	0.1	2^{-3}	2^{-4}	130	12	50
SR-p53	280	3	0.1	2^{-1}	2^{-4}	120	6	80

ADAPTED MULTITASK MODEL HYPERPARAMETERS

Table S4: Hyperparameters employed by the adapted multitask models in different datasets.

	QM9				Alchemy				Tox21
	one-layer net	two-layer net	three-layer net	GCN	one-layer net	two-layer net	three-layer net	GCN	two-layer net
epoch	390	350	460	5	390	390	390	13	320

QM9 INTER-TASK CORRELATIONS AND TARGET IMPORTANCE DISTRIBUTIONS

Table S5: Correlation matrix of quantum chemical properties in QM9.

	mu	**alpha**	**homo**	**lumo**	**gap**	**r2**	**zpve**	**cv**	**u0**	**u298**	**h298**	**g298**
mu	1.0000	-0.2392	-0.1012	-0.3856	-0.3336	0.0108	-0.3734	-0.1623	-0.2711	-0.2711	-0.2711	-0.2711
alpha	-0.2392	1.0000	0.2416	0.3043	0.1879	0.5576	0.7313	0.6986	0.2522	0.2522	0.2522	0.2521
homo	-0.1012	0.2416	1.0000	0.2081	-0.2602	-0.0984	0.1317	0.0095	0.1034	0.1033	0.1033	0.1034
lumo	-0.3856	0.3043	0.2081	1.0000	0.8903	0.0165	0.6546	0.3155	0.3340	0.3340	0.3340	0.3340
gap	-0.3336	0.1879	-0.2602	0.8903	1.0000	0.0621	0.5849	0.3071	0.2816	0.2816	0.2816	0.2816
r2	0.0108	0.5576	-0.0984	0.0165	0.0621	1.0000	0.3792	0.6938	-0.2019	-0.2019	-0.2019	-0.2020
zpve	-0.3734	0.7313	0.1317	0.6546	0.5849	0.3792	1.0000	0.7405	0.2958	0.2959	0.2959	0.2958
cv	-0.1623	0.6986	0.0095	0.3155	0.3071	0.6938	0.7405	1.0000	-0.0858	-0.0858	-0.0858	-0.0859
u0	-0.2711	0.2522	0.1034	0.3340	0.2816	-0.2019	0.2958	-0.0858	1.0000	1.0000	1.0000	1.0000
u298	-0.2711	0.2522	0.1033	0.3340	0.2816	-0.2019	0.2959	-0.0858	1.0000	1.0000	1.0000	1.0000
h298	-0.2711	0.2522	0.1033	0.3340	0.2816	-0.2019	0.2959	-0.0858	1.0000	1.0000	1.0000	1.0000
g298	-0.2711	0.2521	0.1034	0.3340	0.2816	-0.2020	0.2958	-0.0859	1.0000	1.0000	1.0000	1.0000

Table S6: Target importance in the SST-xgboost model, which is computed from the relative influence of each variable in trees growing. Each row in the table represents a SST fitting for a QM9 task. Weights in the expanded feature space for those target estimates are extracted.

	mu	**alpha**	**homo**	**lumo**	**gap**	**r2**	**zpve**	**cv**	**u0**	**u298**	**h298**	**g298**
mu	0.1695	0.0008	0.0008	0.0011	0.0009	0.0007	0.0008	0.0008	0.0010	0.0022	0.0000	0.0016
alpha	0.0014	0.1370	0.0019	0.0016	0.0020	0.0013	0.0018	0.0020	0.0028	0.0046	0.0000	0.0030
homo	0.0010	0.0007	0.3468	0.0008	0.0010	0.0007	0.0007	0.0007	0.0008	0.0017	0.0000	0.0015
lumo	0.0004	0.0003	0.0004	0.6830	0.0004	0.0003	0.0004	0.0003	0.0003	0.0002	0.0000	0.0006

gap	0.0005	0.0005	0.0018	0.0016	0.5918	0.0005	0.0006	0.0004	0.0005	0.0004	0.0000	0.0007
r2	0.0053	0.0113	0.0029	0.0008	0.0029	0.3723	0.0030	0.0037	0.0110	0.0163	0.0000	0.0085
zpve	0.0012	0.0036	0.0039	0.0038	0.0013	0.0011	0.5417	0.0014	0.0025	0.0030	0.0000	0.0018
cv	0.0000	0.0253	0.0083	0.0091	0.0026	0.0374	0.0465	0.5104	0.0178	0.0112	0.0000	0.0073
u0	0.0155	0.0092	0.0045	0.0040	0.0043	0.0025	0.0011	0.0029	0.3121	0.2339	0.0000	0.0368
u298	0.0165	0.0095	0.0045	0.0039	0.0043	0.0026	0.0012	0.0029	0.3096	0.2337	0.0000	0.0368
h298	0.0165	0.0095	0.0045	0.0039	0.0043	0.0026	0.0012	0.0029	0.3096	0.2337	0.0000	0.0368
g298	0.0172	0.0095	0.0051	0.0037	0.0048	0.0027	0.0011	0.0029	0.3065	0.2265	0.0000	0.0425

Table S6 exhibits the target importance distribution derived from the SST-xgboost model in QM9 task fitting. The 0 value for the h298 column in all the task fitting possibly demonstrates this target variable is not contributing to the trees splitting (due to the strong similarity between h298 and u298). No direct relationship between the target importance and the target correlation (shown in Table S5) can be traced, implying that the enhanced prediction accuracy in SST may not result from the correlated tasks.

COMPARISON WITH THE BYPASS MULTITASK MODEL FOR QM9

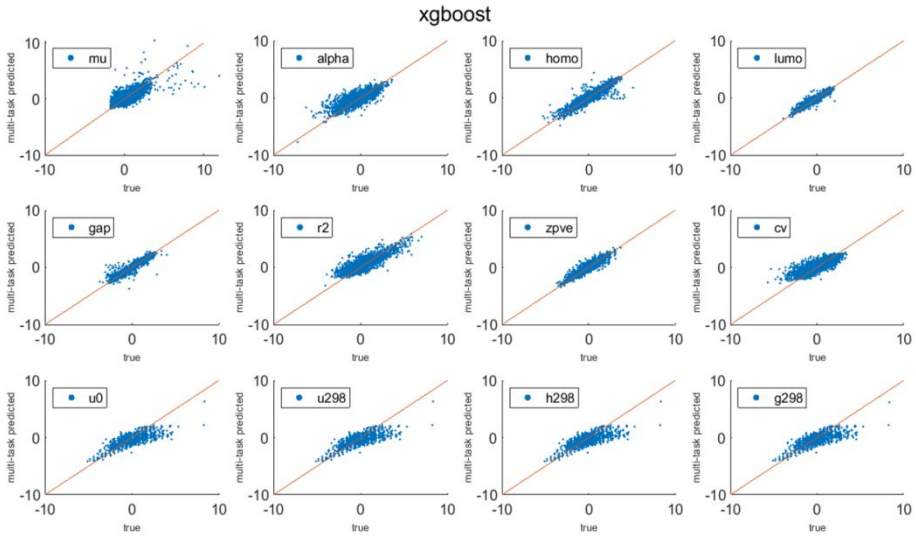
Table S7: Test set prediction *RRMSE*, *aRRMSE* and *RP* for the bypass network model compared to the ST, SST and the adapted counterpart in QM9.

RRMSE	**ST DNN 2L**	**SST DNN 2L**	**adapted DNN 2L**	**bypass DNN 2L**
mu	0.5550	0.5596	0.6021	0.5736
alpha	0.4425	0.4109	0.4592	0.5965
homo	0.3757	0.3607	0.4194	0.3713
lumo	0.2144	0.2033	0.2606	0.2101
gap	0.2807	0.2495	0.3028	0.2643
r2	0.4002	0.3849	0.4356	0.5102
zpve	0.2828	0.2609	0.3005	0.3287
cv	0.3996	0.3765	0.4162	0.4983
u0	0.4469	0.4048	0.4412	0.5419
u298	0.4416	0.4106	0.4406	0.5428
h298	0.4367	0.4045	0.4408	0.5417
g298	0.4425	0.4063	0.4418	0.5437
aRRMSE	0.3932	0.3694	0.4134	0.4602
RP		1.0646	0.9512	0.8544

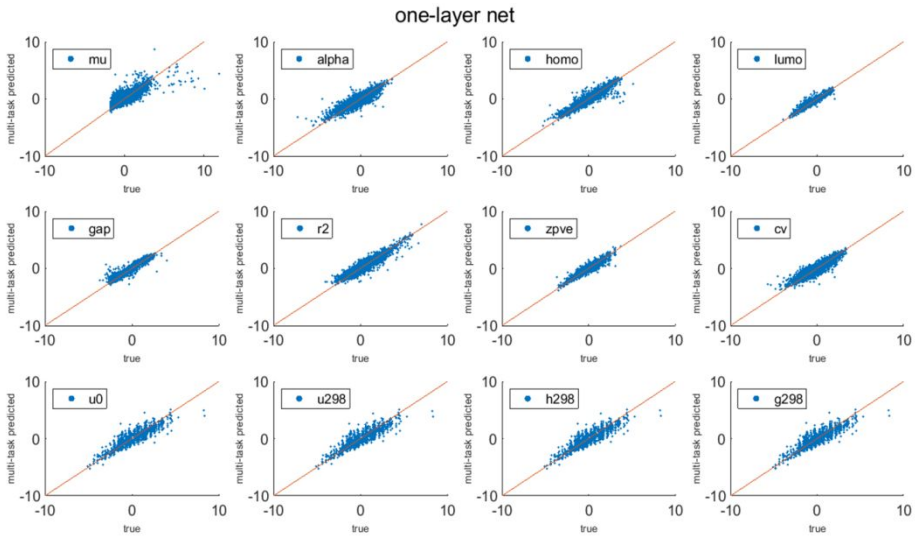
We implement the bypass network model mentioned in “Is Multitask Deep Learning Practical for Pharma” for QM9, and compare the target prediction quality to the ST, SST and the adapted counterpart. It is observed that the bypass model underperforms ST, SST and the conventional multitask DNN (adapted DNN), which is consistent with the corresponding reference where moderate performance of bypass is found.

Note that the bypass architecture has two shared hidden layers of size [1024, 512], while having two per-task bypass layers of size [1024, 512]. A batch size of 128 is used, with the epoch number optimized to be 25. All other settings are identical to the adapted DNN regressor.

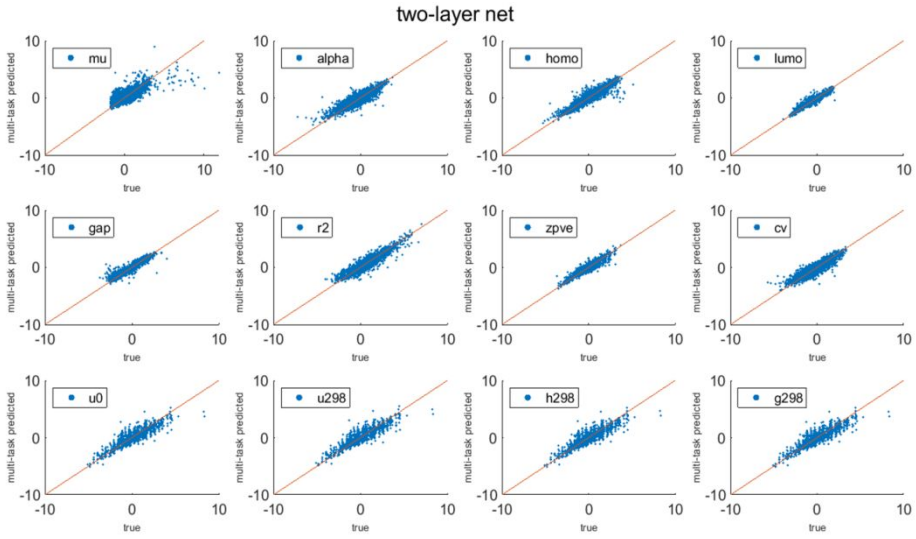
SST TEST SET PREDICTIONS IN QM9



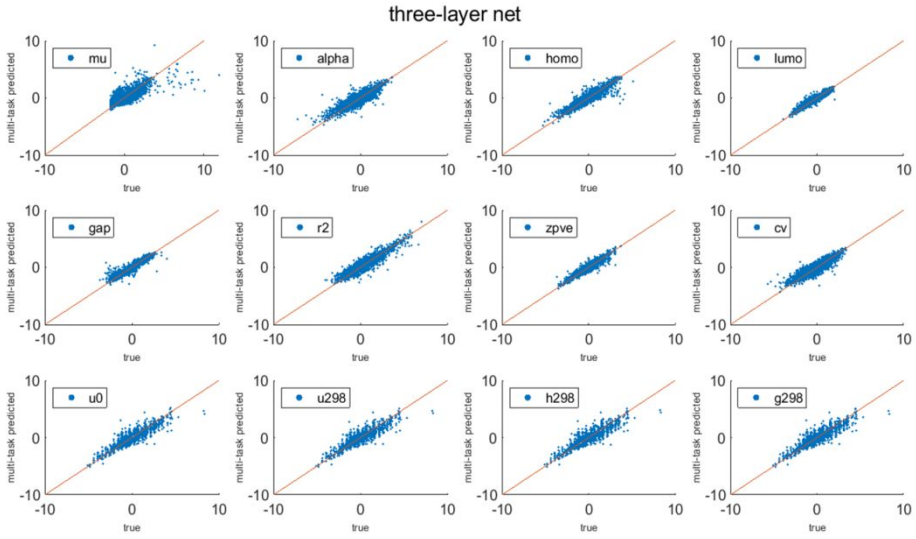
(a)


one-layer net
mu
alpha
homo
lumo
gap
r2
zpve
cv
u0
u298
h298
g298
multi-task predicted
true

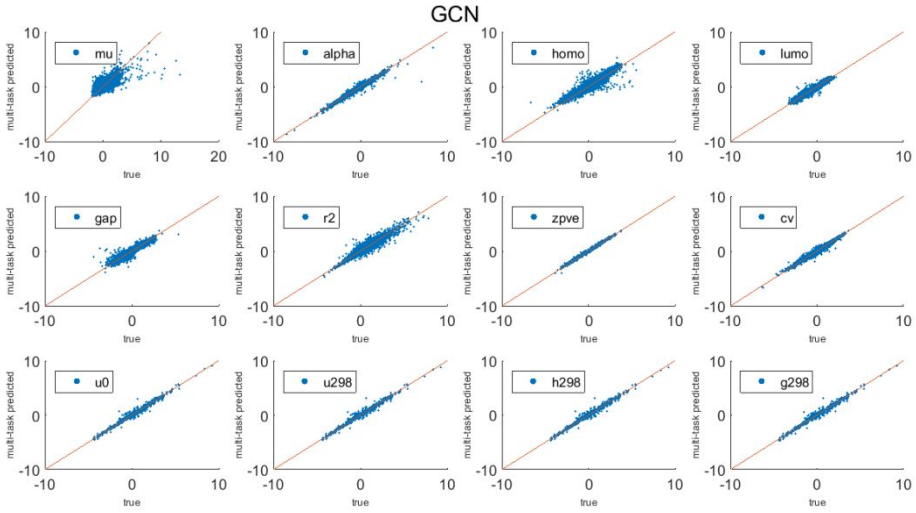
(b)


two-layer net
mu
alpha
homo
lumo
gap
r2
zpve
cv
u0
u298
h298
g298
multi-task predicted
true

(c)


three-layer net
mu
alpha
homo
lumo
gap
r2
zpve
cv
u0
u298
h298
g298
multi-task predicted
true

(d)



(e)

Figure S1: Multitask out-of-sample predictions on the test dataset in QM9 with various base regressors. (a) xgboost base regressor. (b) DNN base regressor with 1 hidden layer. (c) DNN base regressor with 2 hidden layers. (d) DNN base regressor with 3 hidden layers. (e) GCN base regressor. The red line indicates the identity mapping.

Figure S1 shows the SST performances on the QM9 test set with various base regressors. A more scattered profile for the dipole moment (mu) prediction can be visualized, corresponding to the notable *RRMSE* for mu. The prediction qualities of the SST-DNN models with different number of hidden layers are comparable with each other, which is consistent with the error analysis. The SST-GCN gives a more converged pattern for the predicted/true data map, especially for the vibrational (zpve) and thermodynamic (cv, u0, u298, h298 and g298) properties.