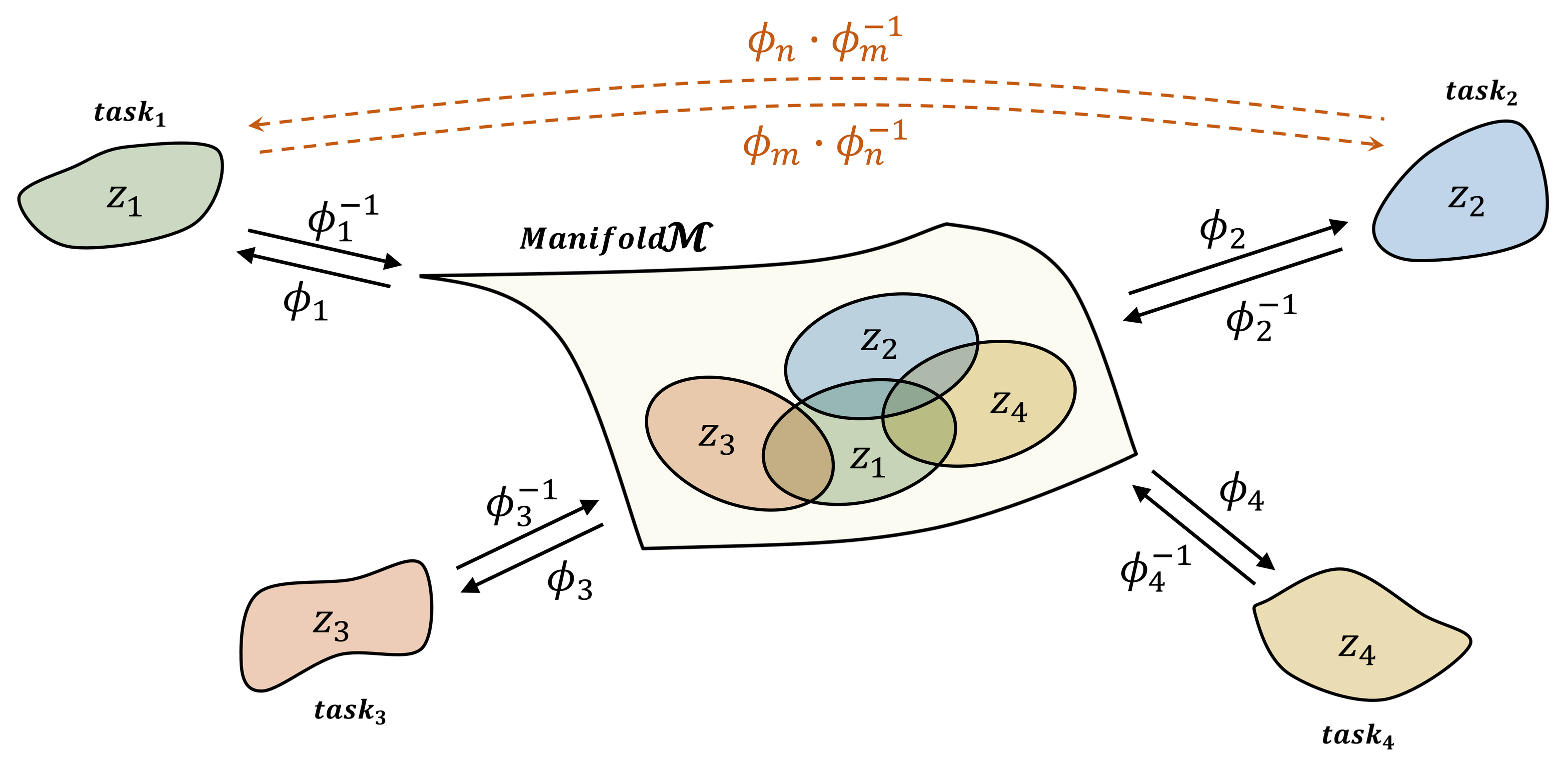
LG AI Research

Multi-task Extension of Geometrically Aligned Transfer Encoder

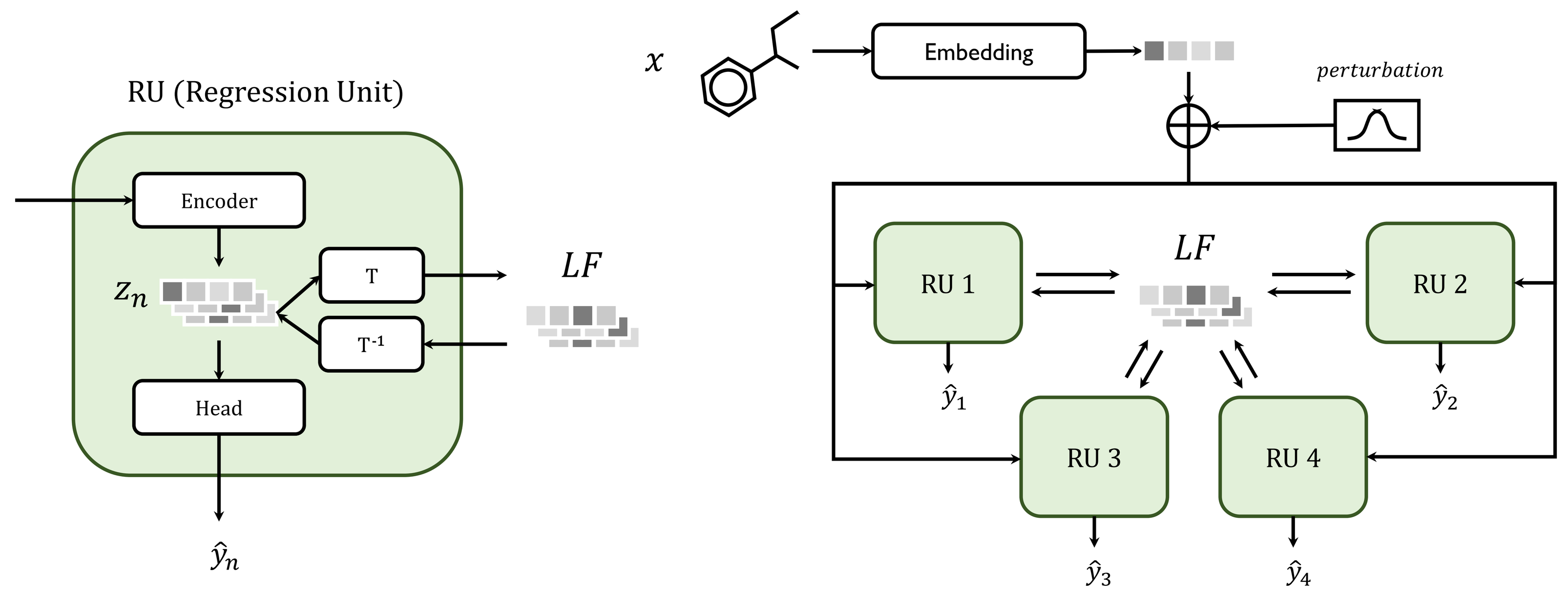
Sung Moon Ko*, Sumin Lee*, Dae-Woong Jeong*, Hyunseung Kim, Chanhui Lee, Soorin Yim, Sehui Han
LG AI research

INTRODUCTION

- Scientific experiments or simulations often require significant amounts of time and effort, making it challenging to amass abundant data in the field.
- One of the recent advances for mitigating data shortage issue, Geometrically Aligned Transfer Encoder (GATE) algorithm utilizes the concept of curved geometry in a Riemannian scheme to align the geometrical shapes of the underlying latent spaces of data abundant source task and target task.
- GATE is proven to work in a two-task setting, with one target and one source task. Yet, theoretically, it is not restricted to two tasks. Therefore, we extend the concept of GATE to multiple sources.
- Our main contribution of the article is as follows.
  - ✓ We extend the GATE to encode multiple source taskssetup.
  - ✓ Extension to multiple tasks provides a positive leveraging effect.
  - ✓ Proposed model outperforms conventional method in multi-task molecular property setup.



METHODS



$$l_{tot} = l_{reg} + \alpha l_{auto} + \beta l_{cons} + \gamma l_{map} + \delta l_{dis}$$

- We design a mapping function with an autoencoder model for each task. The encoder T indicates mapping from latent space to universal manifold, and the decoder T^{-1} indicates mapping the other way around.

$$z'_\alpha = \text{Transfer}_{\alpha\to LF}(z_\alpha) \qquad z'_t = \text{Transfer}_{t\to LF}(z_t) \qquad z'_\alpha = \text{Transfer}_{\alpha\to LF}(z_\alpha)$$
$$\hat{z}_\alpha = \text{Transfer}^{-1}{}_{LF\to\alpha}(z'_\alpha) \qquad \hat{z}_t = \text{Transfer}^{-1}{}_{LF\to t}(z'_t) \qquad \hat{z}_{\alpha\to t} = \text{Transfer}^{-1}{}_{LF\to t}(z'_\alpha)$$

$$l_{\text{auto}} = \sum_\alpha \text{MSE}(z_\alpha, \hat{z}_\alpha) \qquad l_{cons} = \sum_\alpha \text{MSE}(z'_\alpha, z'_t) \qquad l_{map} = \sum_\alpha \text{MSE}(y_t, \hat{y}_{\alpha\to t})$$

- Latent spaces for task pairs are aligned by matching not only transferred vectors in locally flat frame (consistency loss) but also transferred latent vectors corresponds to outputs of other tasks (mapping loss).

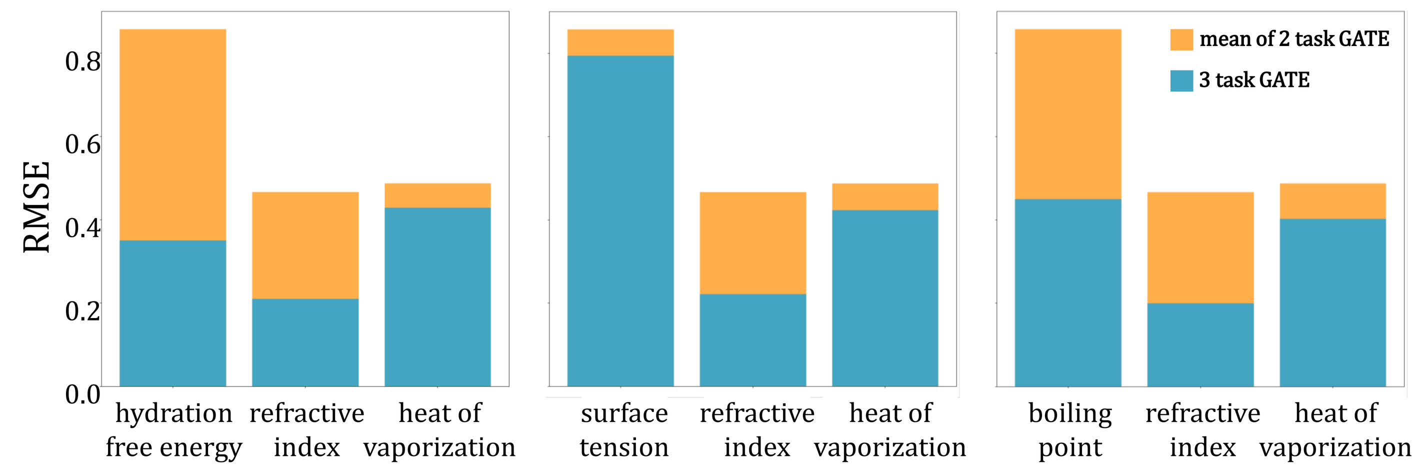
$$s^i_\alpha \equiv |(z'_\alpha) - (z'^i_\alpha)| \qquad s^i_t \equiv |(z'_t) - (z'^i_t)|$$
$$z'^i_\alpha = \text{Transfer}_{\alpha\to LF}(\text{Encoder}_\alpha(x^i)) \qquad l_{dis} = \frac{1}{M}\sum_\alpha C_\alpha \sum_i^M \text{MSE}(s^i_\alpha, s^i_t)$$
$$z'^i_t = \text{Transfer}_{t\to LF}(\text{Encoder}_t(x^i))$$

- Finally, distances between the latent vector and its perturbations from each task are computed and restricted to be same (distance loss) for geometrical alignment.

RESULT & DISCUSSION

Effect of multi-task extension from two-task GATE to three-task GATE

- Across all three experiment sets, there is a consistent reduction in the RMSE of the three-task GATE compared to the two-task GATE, even when different additional tasks are included in the sets.
- This result indicates that synergy effect can be achieved through the proposed multi-task extension of the GATE.



Regression performance of many-task GATE

- Pearson correlation of GATE outperforms MTL and STL for 7 out of 10 tasks. Moreover, GATE's regression performance ranks within the top 2 for all tasks, demonstrating robust performance with an average rank of 1.3.

Tasks	GATE	MTL	STL
Parachor	0.9309±0.0073	**0.9358±0.0060**	0.9287±0.0086
Surface Tension	**0.8440±0.0073**	0.8195±0.0236	0.7171±0.0211
Dielectric Constant	**0.9228±0.0169**	0.9099±0.0176	0.9216±0.0070
Hydration Free Energy	**0.9504±0.0107**	0.9409±0.0160	0.9414±0.0097
Heat of Vaporization	0.8962±0.0057	**0.9018±0.0067**	0.8618±0.0160
Boiling Point	**0.9113±0.0066**	0.9076±0.0087	0.8847±0.0316
Refractive Index	**0.9793±0.0025**	0.9781±0.0030	0.9761±0.0009
Density	**0.8581±0.0115**	0.8512±0.0143	0.8237±0.0330
Melting Point	0.8739±0.0052	0.8714±0.0073	**0.8901±0.0019**
Viscosity	**0.9105±0.0072**	0.8952±0.0061	0.8967±0.0134
No. 1st	**7**	2	1
Avg. Rank	**1.3**	2.2	2.5

- As shown in the table, in many cases, multi-task setup enhances regression performance, but in some cases, it can actually reduce regression performance.
- This decline in performance can be attributed to the negative transfer of undesired interfering information among the tasks.
- GATE shows a reduction of performance in only one task, while classical MTL exhibits a performance decrease in four tasks out of ten tasks.

Tasks	GATE	MTL
Parachor	0.24	0.78
Surface Tension	17.69	14.28
Dielectric Constant	0.13	**-1.26**
Hydration Free Energy	0.96	**-0.06**
Heat of Vaporization	3.99	4.64
Boiling Point	3.01	2.59
Refractive Index	0.32	0.21
Density	4.18	3.33
Melting Point	**-1.83**	**-2.09**
Viscosity	1.54	**-0.16**

Discussion

- In this work, we designed the mathematical notion of the extended GATE with newly introduced hyperparameters and extended losses, and we have demonstrated the superior performance of the model using numerous open database datasets.
- While our model outperforms conventional setups, there are several areas for improvement:

1. the model's computational complexity grows significantly with the number of source tasks. Since the distance and mapping losses must be computed for every pair of source and target tasks, the complexity is on the order of $O(N^2)$. Therefore, compactifying the model architecture is one research direction to explore.
2. the distance loss can potentially be omitted if one can directly calculate the curvature of the space by finding the analytic form of the metric tensor. While this is normally impossible, by utilizing the notion of operator learning, it can be achieved.