Latent Bayesian Optimization via Autoregressive Normalizing Flows

Seunghun Lee[1], Jinyoung Park[1], Jaewon Chu[1], Minseo Yoon[1], Hyunwoo J. Kim[2]

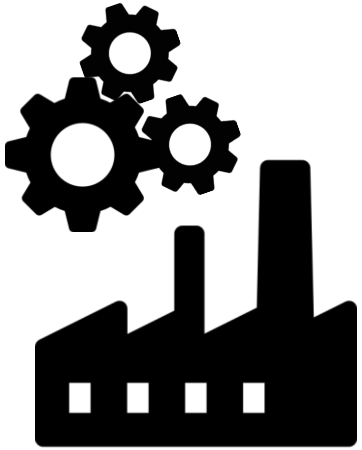
MLV Lab, Korea University[1], KAIST[2]

Bayesian Optimization

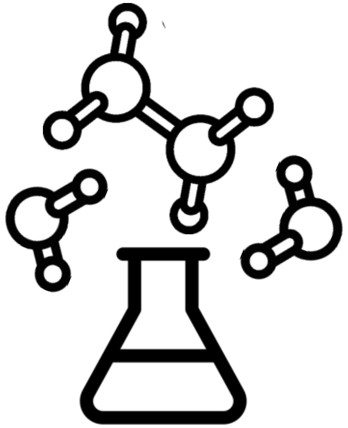
Goal: Optimize a Black-box Function Efficiently

$$\arg\max_{x \in \mathcal{X}} f(x)$$

Applications



Automated Design



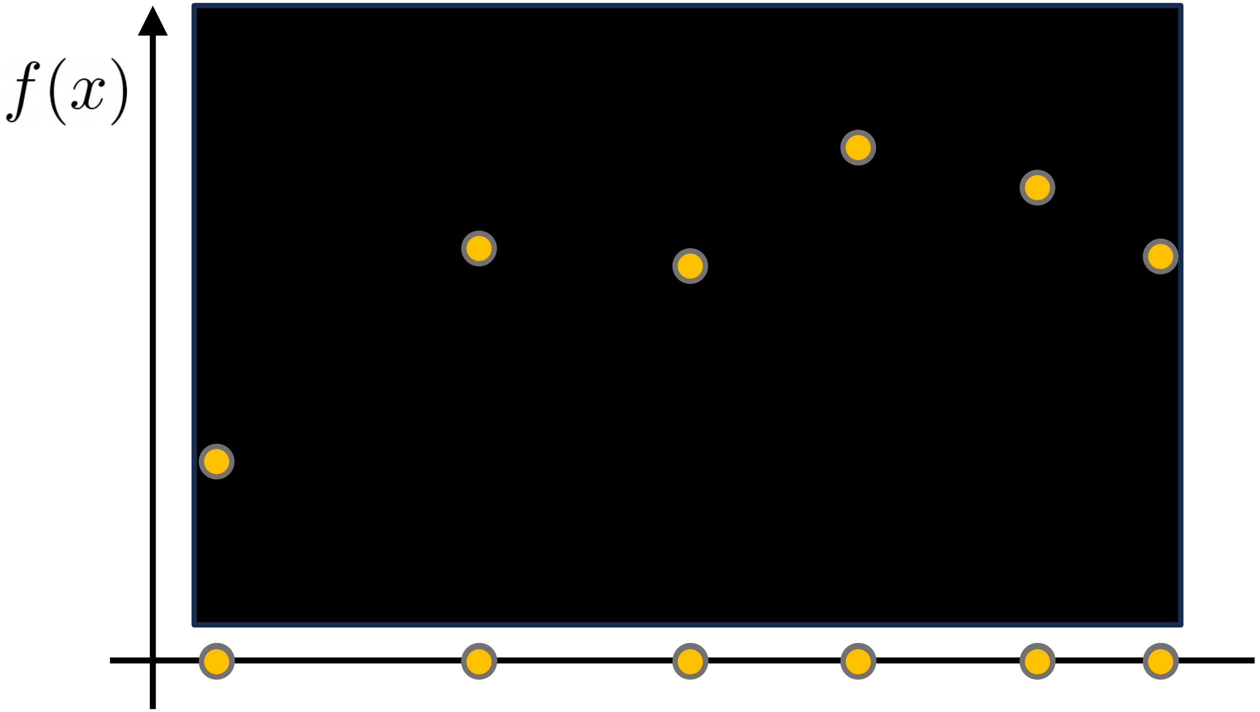
Drug Discovery

...

Bayesian Optimization

Goal: Optimize a Black-box Function Efficiently

$$\arg\max_{x \in \mathcal{X}} f(x)$$



Objective function $f(x)$

→ Expensive and complex

→ Black-box function

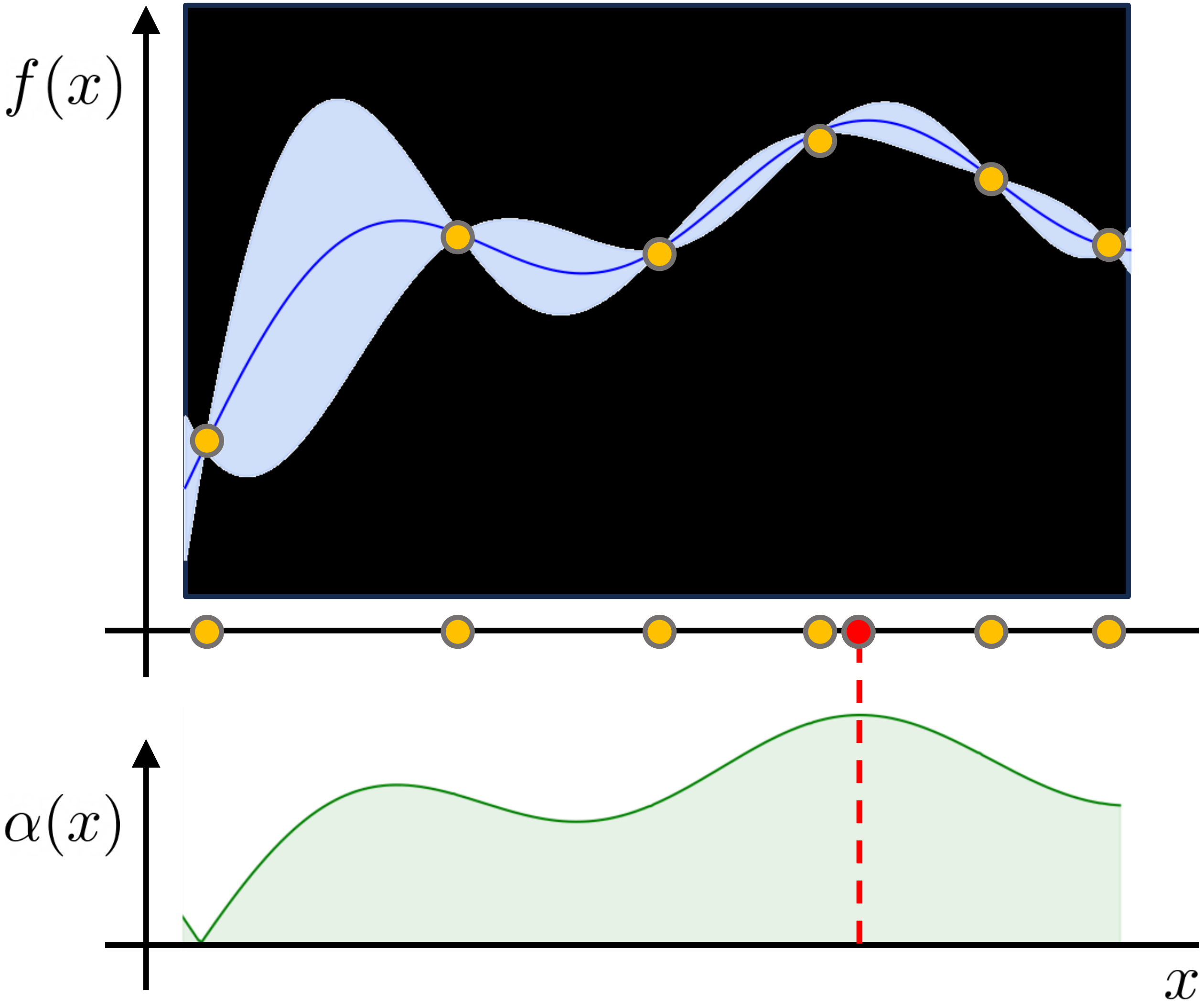
Function evaluation (O)

Derivative/gradient (X)

Bayesian Optimization

Goal: Optimize a Black-box Function Efficiently

$$\arg\max_{x \in \mathcal{X}} f(x)$$



Probabilistic surrogate model $\hat{f}(x)$

→ predicts objective with uncertainty

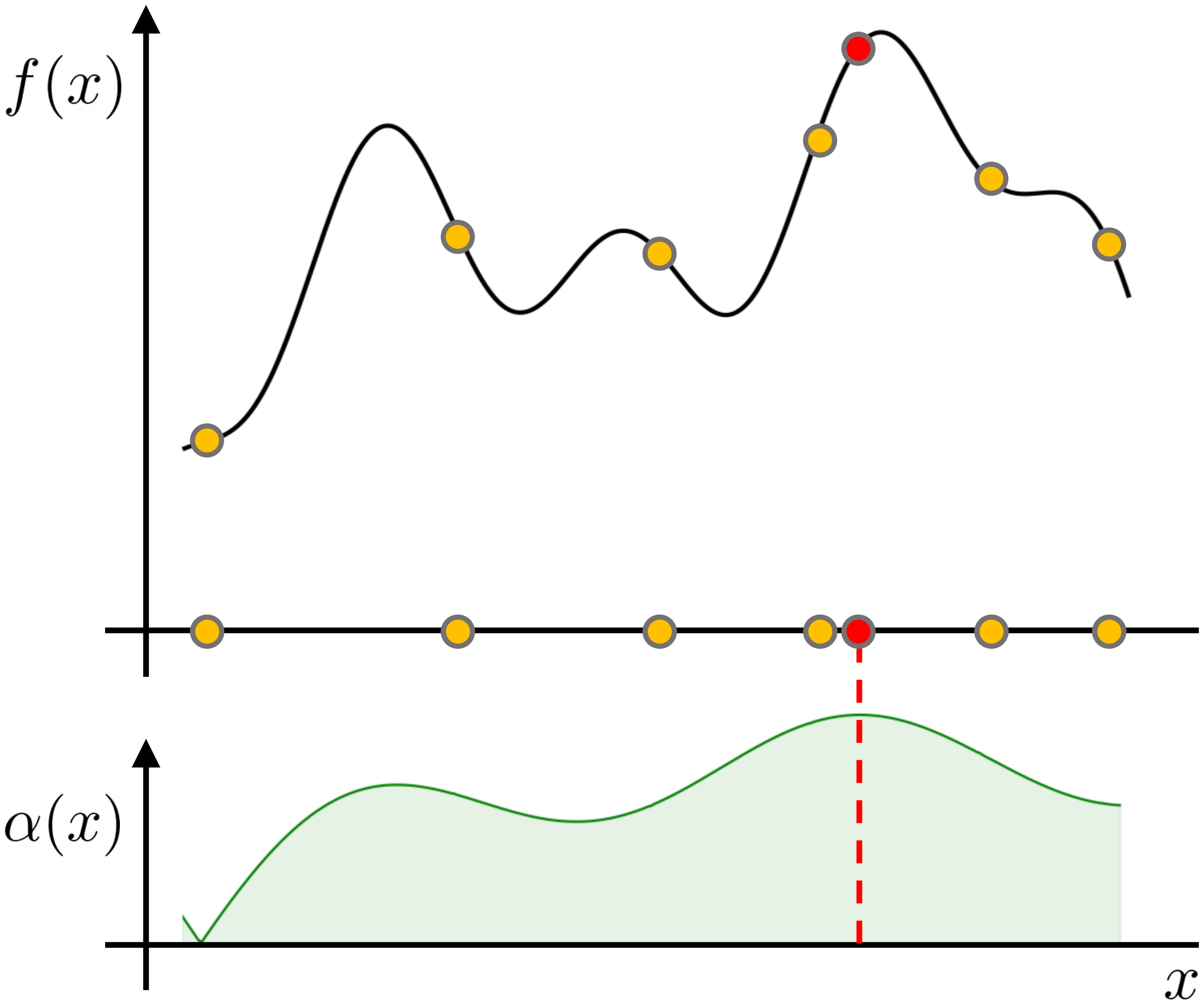
Acquisition function $\alpha(x)$

→ scores where to evaluate next

Bayesian Optimization

Goal: Optimize a Black-box Function Efficiently

$$\arg\max_{x \in \mathcal{X}} f(x)$$



Probabilistic surrogate model $\hat{f}(x)$

→ predicts objective with uncertainty

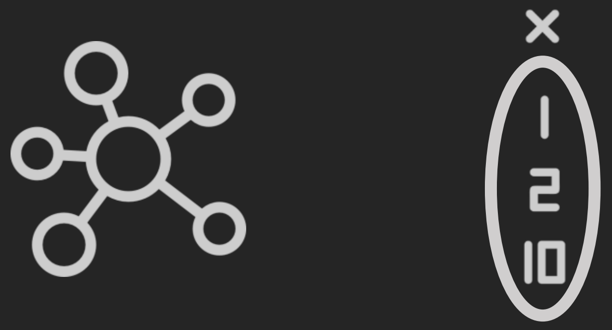
Acquisition function $\alpha(x)$

→ scores where to evaluate next

Bayesian Optimization

- Feasible set is often **Complex**

(structured/discrete)



Bayesian Optimization

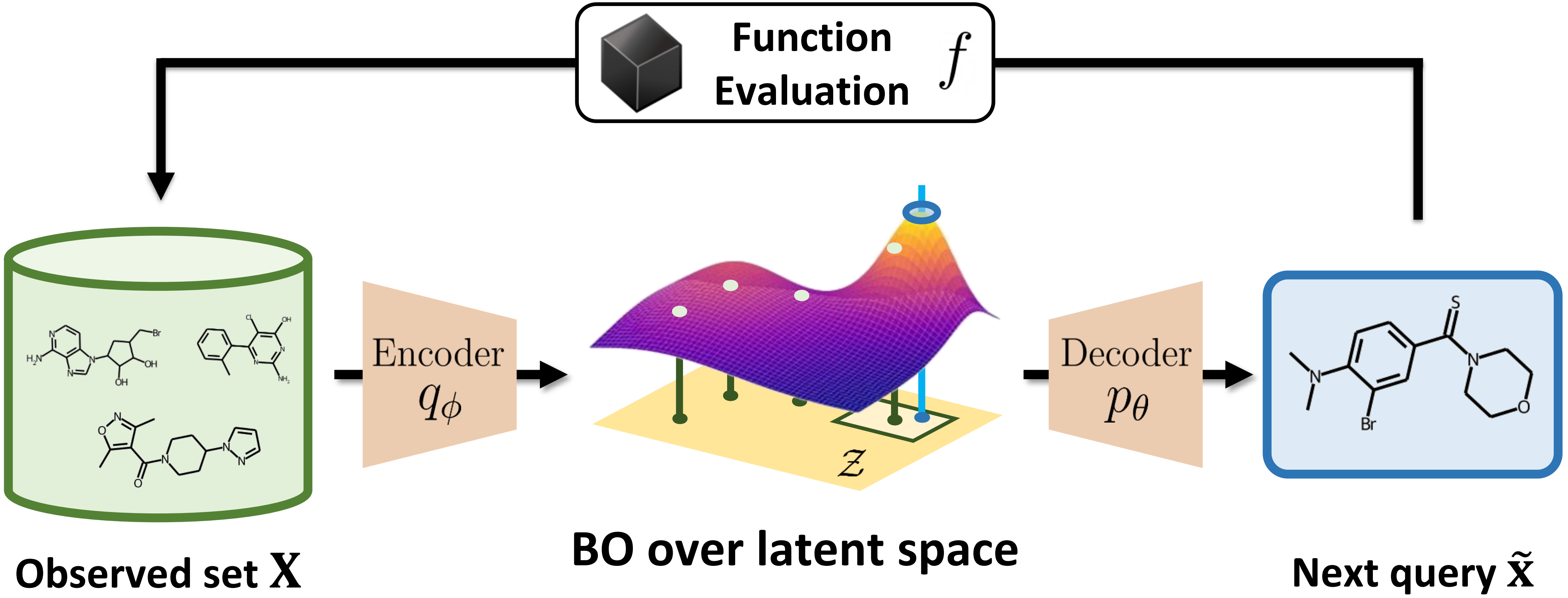
- Feasible set is often Complex

  (structured/discrete)

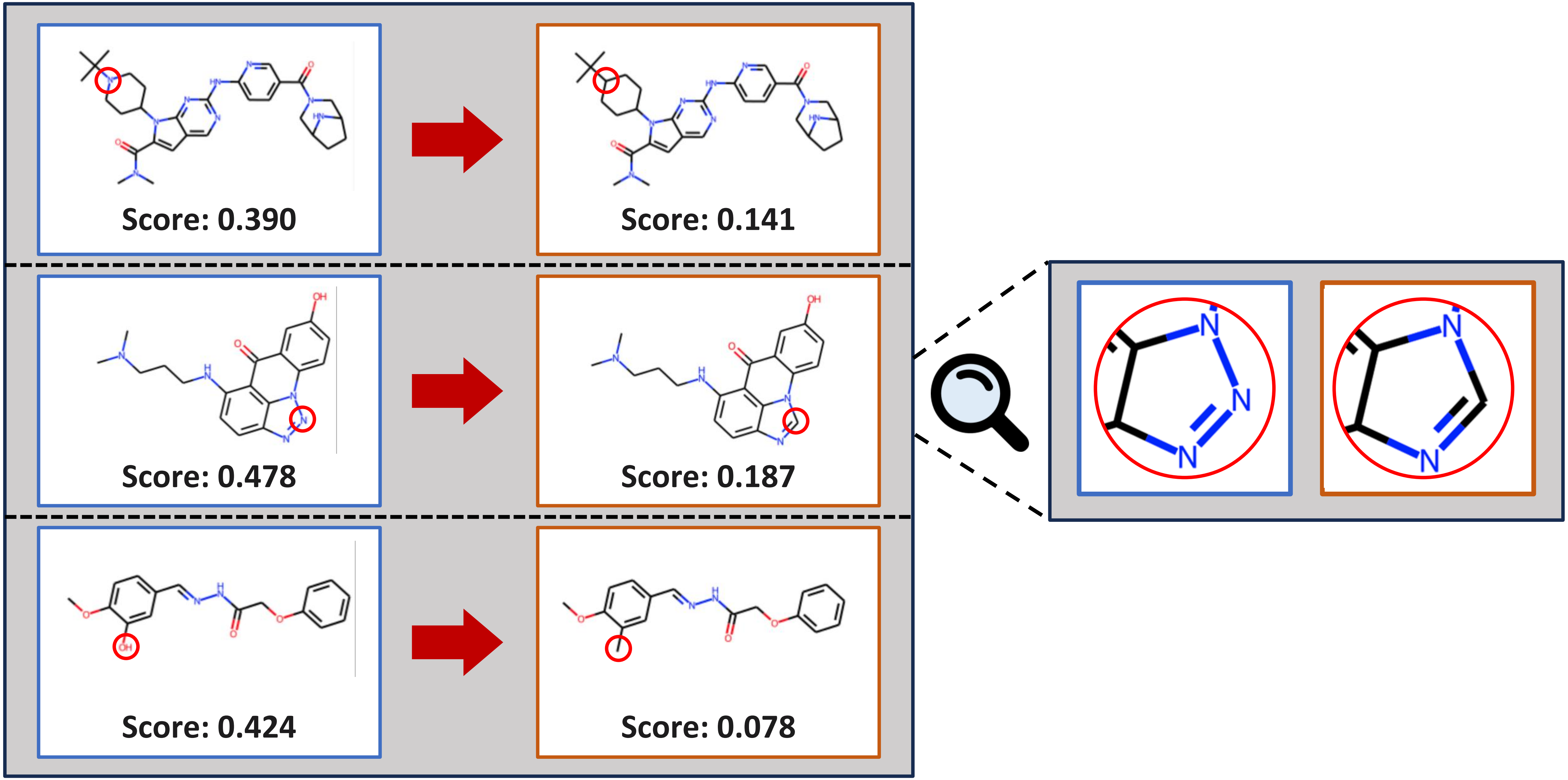
  → Latent Bayesian optimization (LBO)

Latent Bayesian Optimization

- Using a generative model with an **encoder-decoder** structure (e.g., VAE), Bayesian Optimization efficiently **optimizes a black-box function** in the **latent space**.

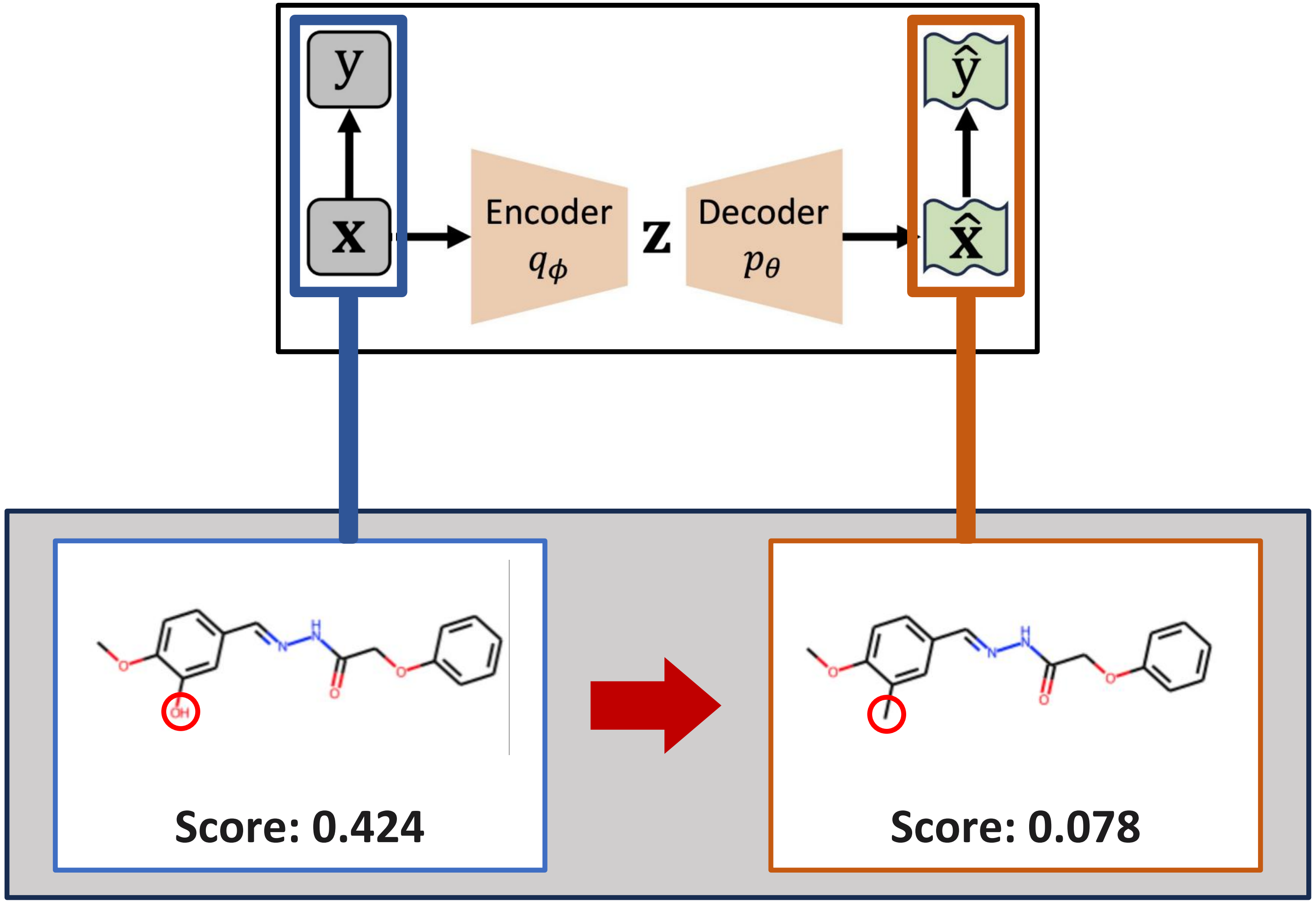


Problem statement – Value discrepancy problem



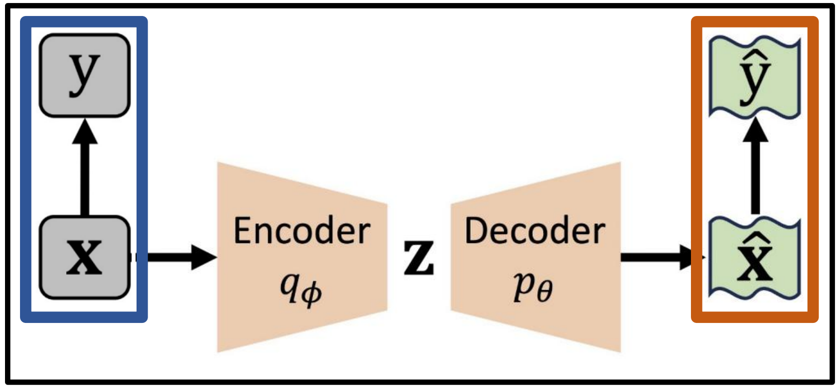
Problem statement – Value discrepancy problem

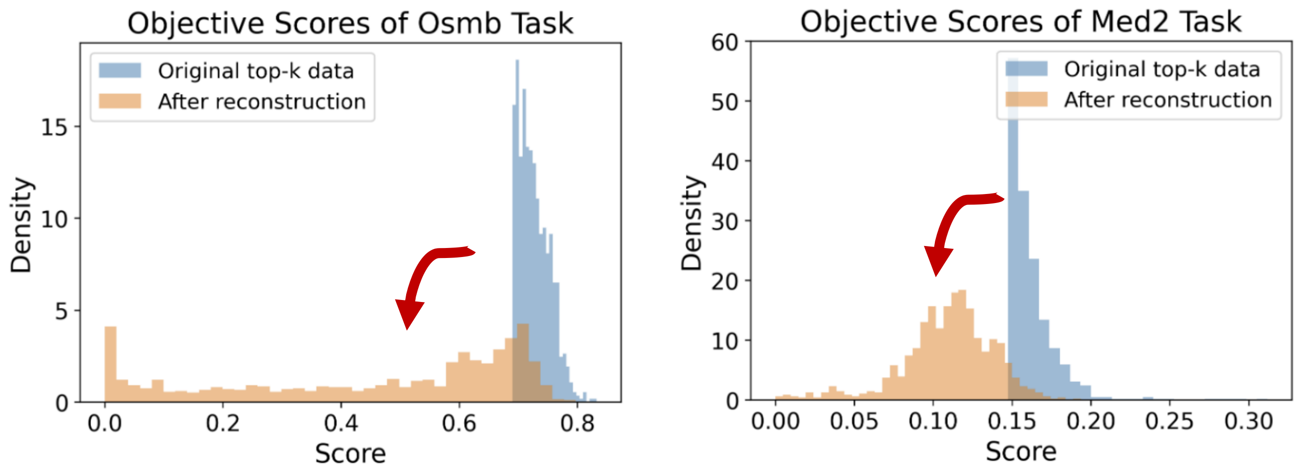
- The **value discrepancy problem ($\mathbf{y} \neq \hat{\mathbf{y}}$)** occurs due to **imperfect reconstruction** by the **encoder-decoder** model.



Problem statement – Value discrepancy problem

- The **value discrepancy problem ($\mathbf{y} \neq \hat{\mathbf{y}}$)** occurs due to **imperfect reconstruction** by the **encoder-decoder** model.





Model: SELFIES VAE[1]

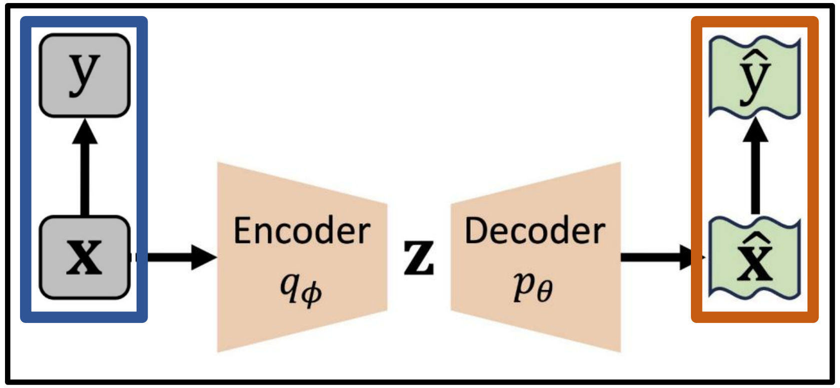
Objective score distribution

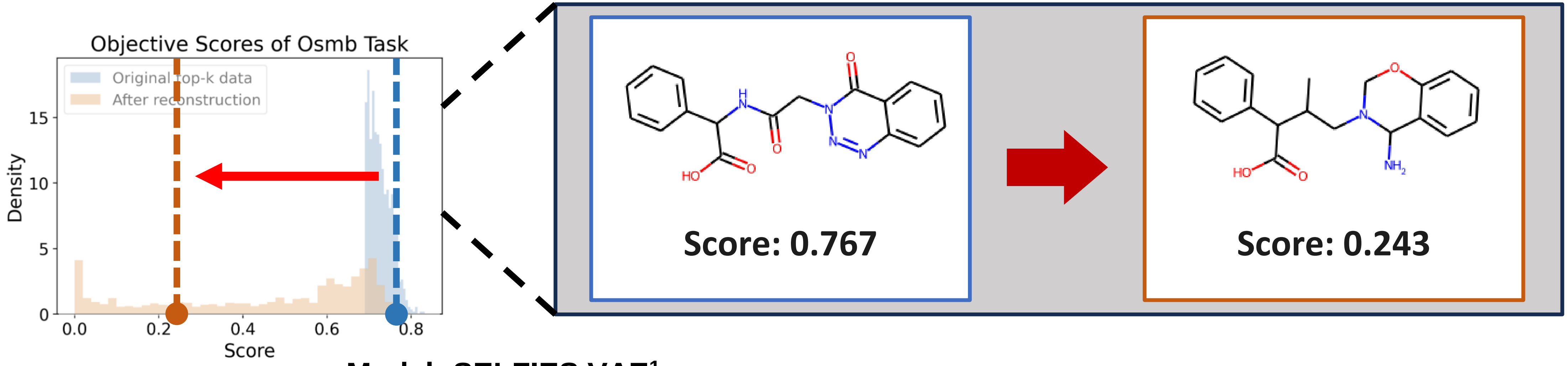
before and after reconstruction

1) Maus, Natalie, et al. "Local latent space bayesian optimization over structured inputs." *Advances in neural information processing systems* 35 (2022): 34505-34518.

Problem statement – Value discrepancy problem

- The **value discrepancy problem ($\mathbf{y} \neq \hat{\mathbf{y}}$)** occurs due to **imperfect reconstruction** by the **encoder-decoder** model.



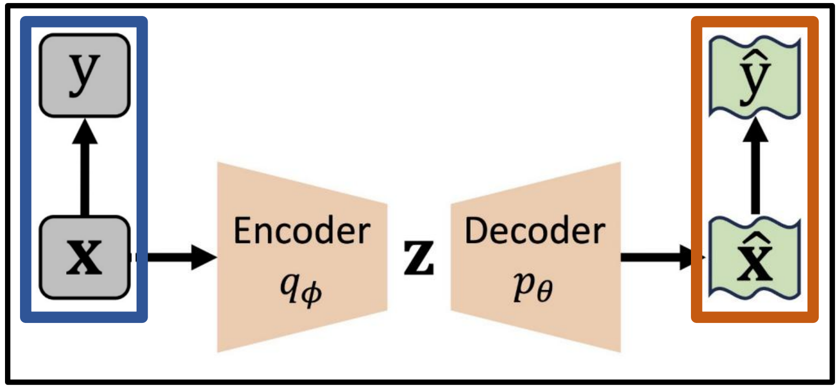


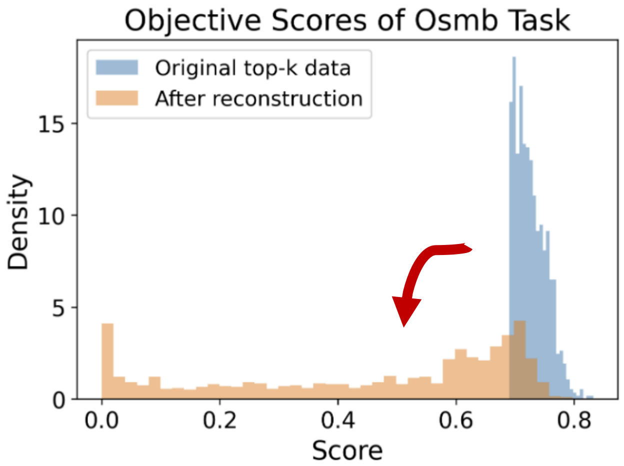
Model: SELFIES VAE[1]

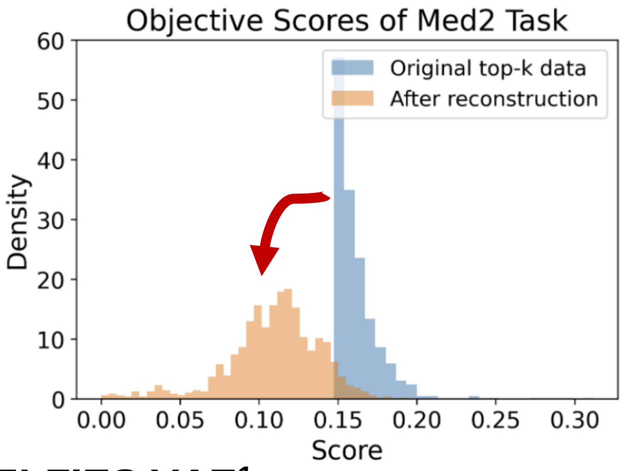
1) Maus, Natalie, et al. "Local latent space bayesian optimization over structured inputs." *Advances in neural information processing systems* 35 (2022): 34505-34518.

Problem statement – Value discrepancy problem

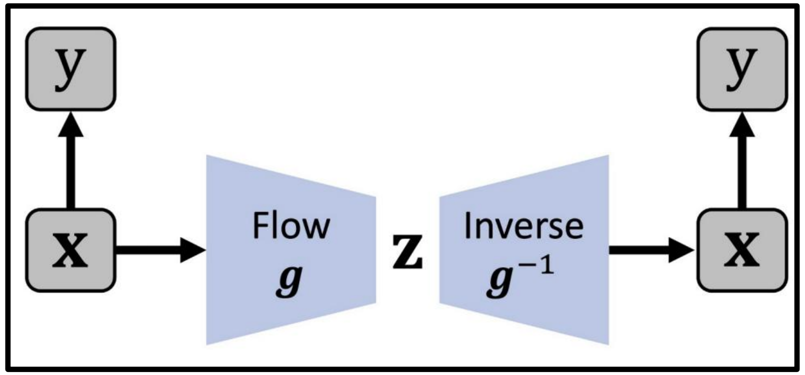
- The **value discrepancy problem ($\mathbf{y} \neq \hat{\mathbf{y}}$)** occurs due to **imperfect reconstruction** by the **encoder-decoder** model.







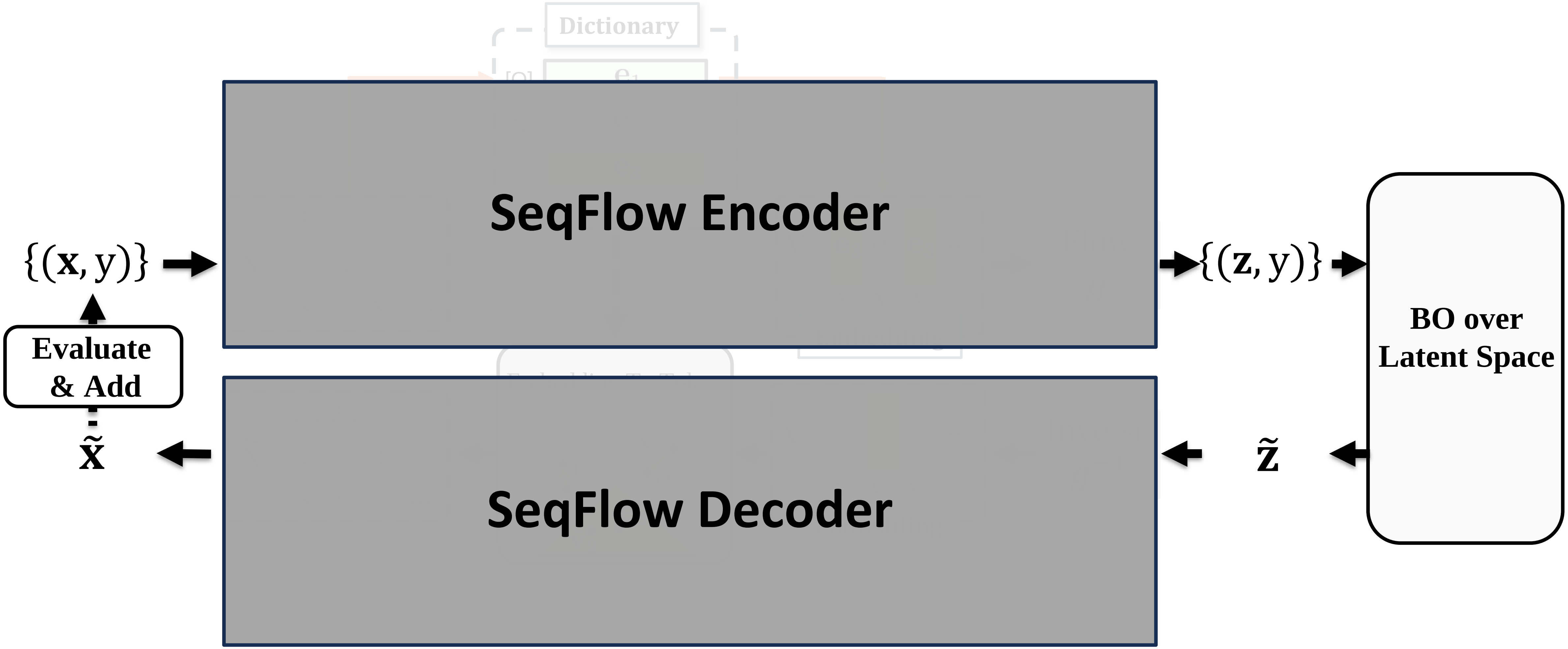
Model: SELFIES VAE[1]



Normalizing Flow (NF) can resolve it!

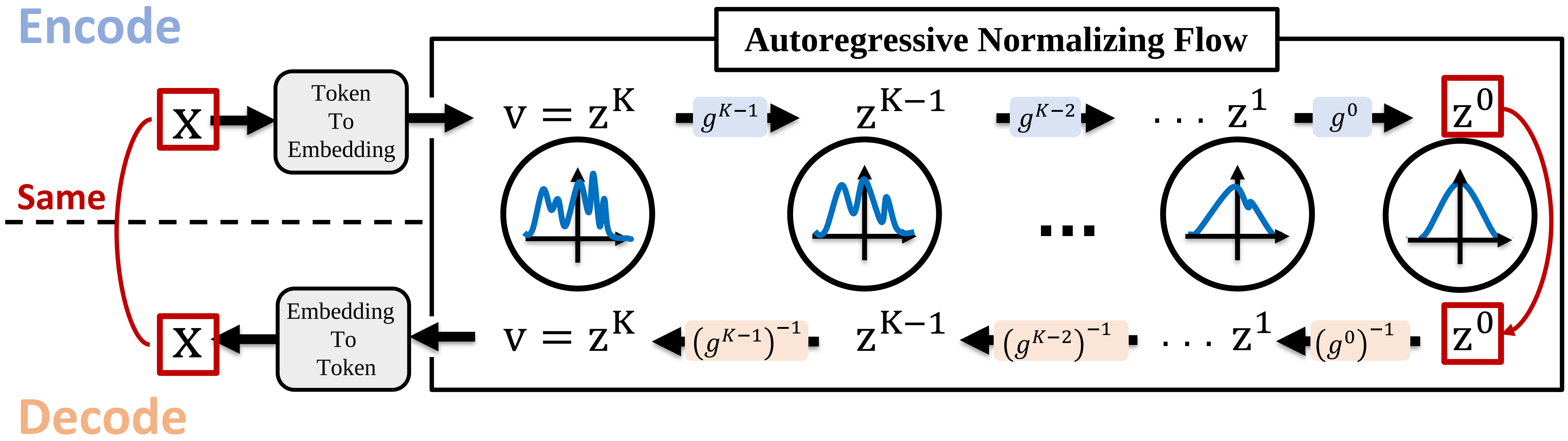
1) Maus, Natalie, et al. "Local latent space bayesian optimization over structured inputs." *Advances in neural information processing systems* 35 (2022): 34505-34518.

NF-BO: Overview

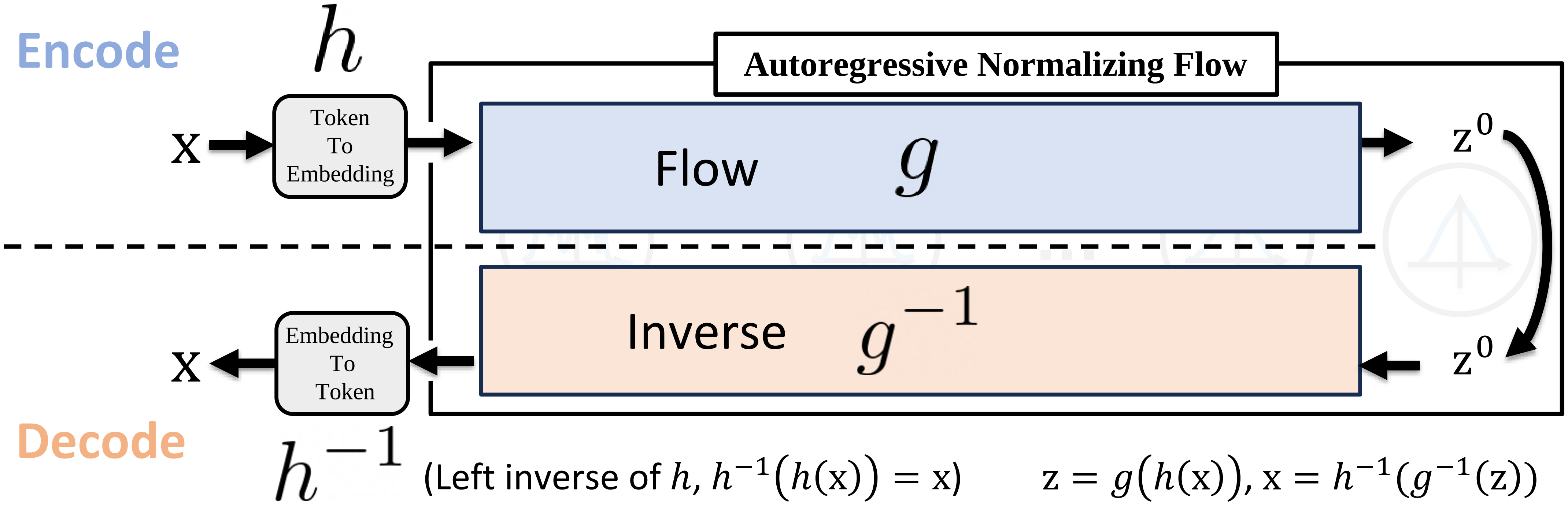


NF-BO: SeqFlow

- Prior **Autoregressive Normalizing Flows** are applied to continuous data or used with **imperfect reconstruction on discrete sequences**.
- We propose **SeqFlow**, which **perfectly reconstruct** discrete sequence data (e.g., text, SMILES).



NF-BO: SeqFlow

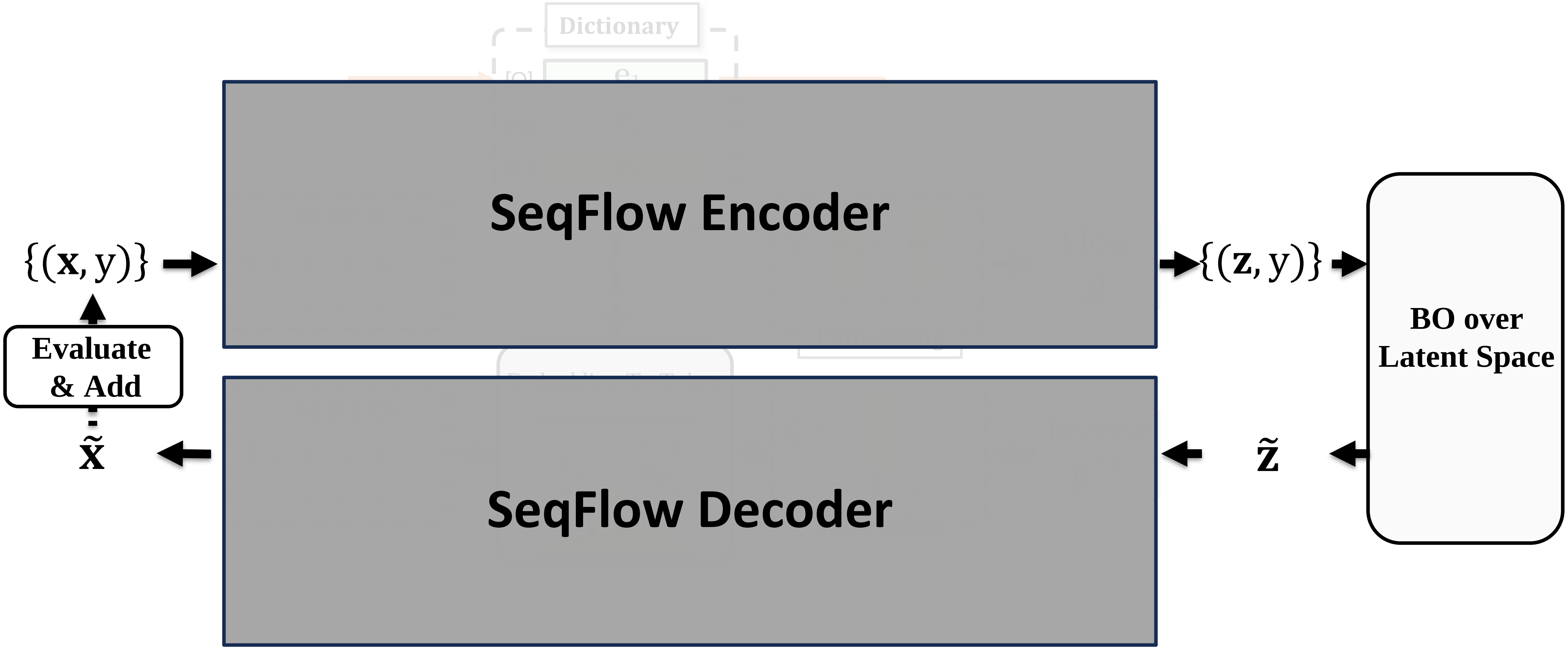


(Left inverse of h, $h^{-1}(h(\mathrm{x})) = \mathrm{x}$) $\mathrm{z} = g(h(\mathrm{x})), \mathrm{x} = h^{-1}(g^{-1}(\mathrm{z}))$

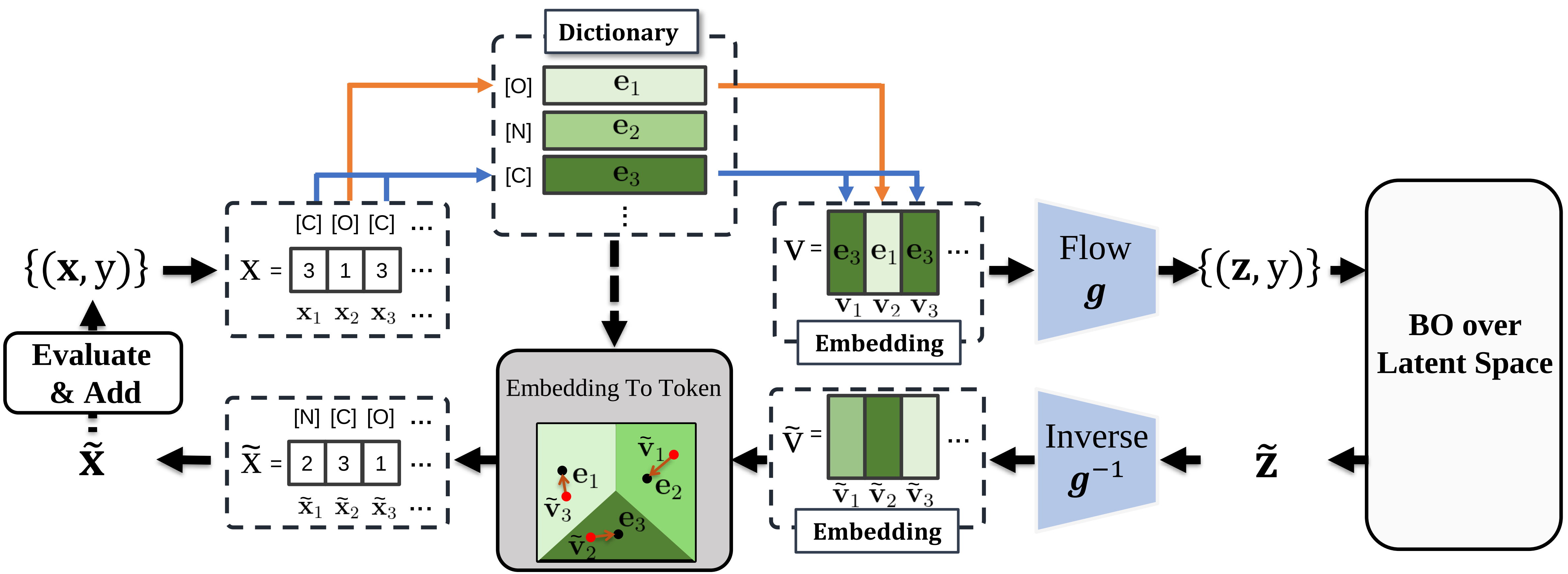
Proposition 1. *If g is invertible and h is injective, then $f = g \circ h$ is left-invertible, i.e., $f^{-1} \circ f = id_X$, where $f^{-1} := h^{-1} \circ g^{-1}$ and h^{-1} is the left inverse of h.*

- This means x can be reconstructed exactly through these operations.

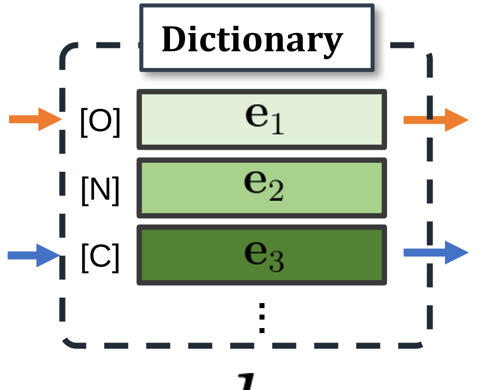
NF-BO: Overview

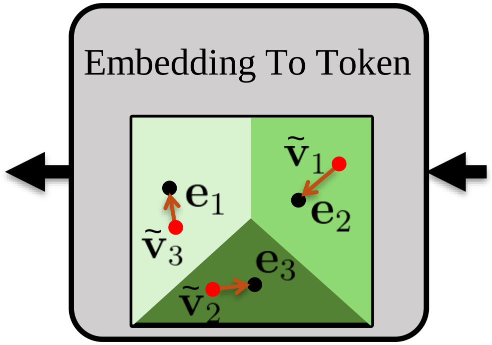


NF-BO: Overview



NF-BO: Overview



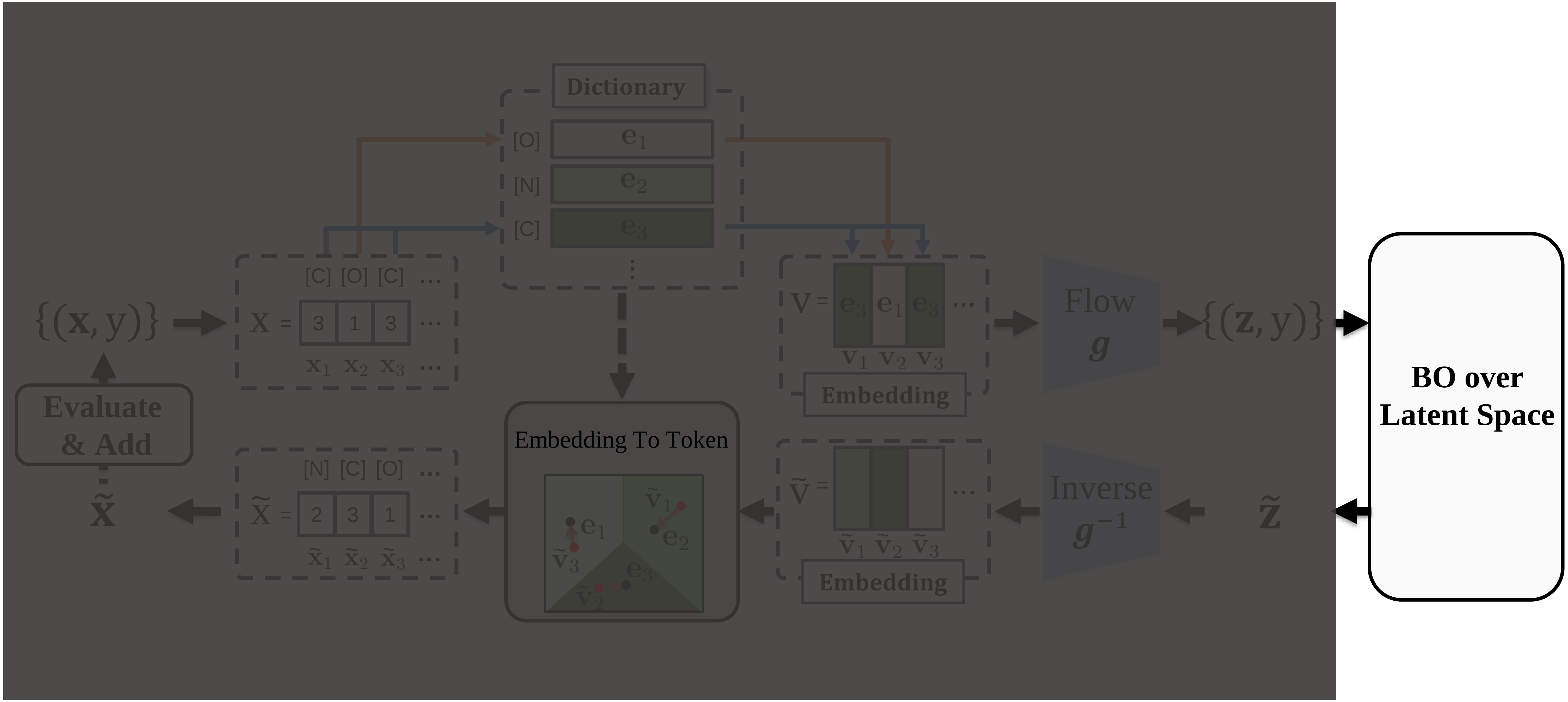


h^{-1} (Left inverse of h)

Proposition 2. *Let the elements of embedding set* $\mathcal{E} = \{\mathrm{e}_1, \mathrm{e}_2, \dots, \mathrm{e}_{|\mathcal{E}|}\}$ *L2-normalized. Then* $h(\mathrm{x}) \coloneqq \mathrm{e}_{\mathrm{x}}$ *is injective and its left-inverse is* $h^{-1}(\mathrm{v}) = [\arg\max_j \mathrm{sim}(\mathrm{v}_i, \mathrm{e}_j)]_{i=1}^{L}$*, where* $\mathrm{sim}(\mathrm{e}_i, \mathrm{e}_j) = \mathrm{e}_i^T \mathrm{e}_j$*, i.e.,* $h^{-1}(h(\mathrm{x})) = \mathrm{x}$.

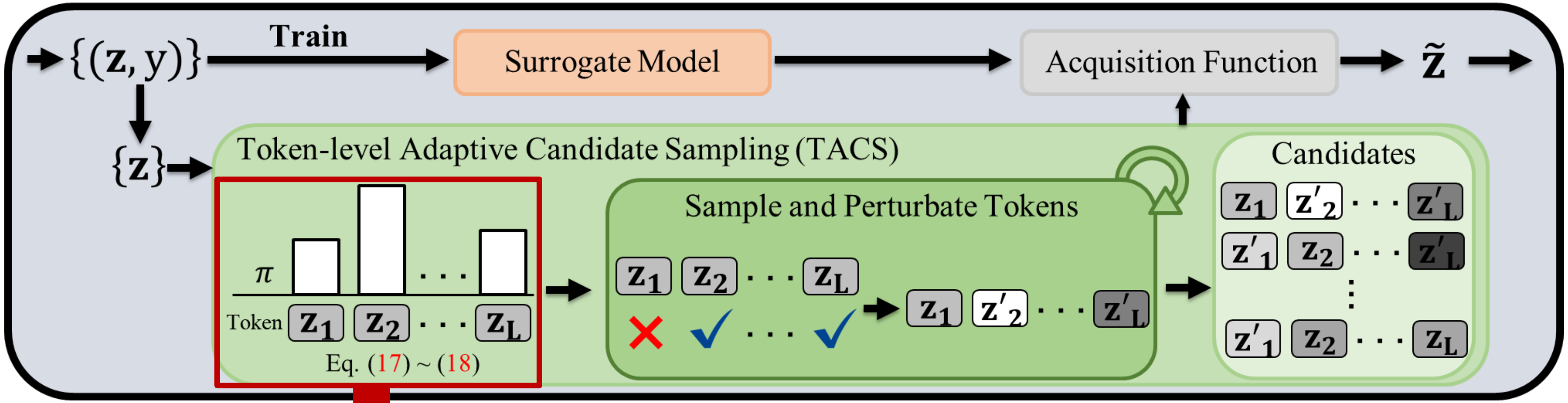
- Our **token-to-embedding** function and its **left-inverse** enable **perfect reconstruction**.

NF-BO: Overview



NF-BO: TACS - Token-level Adaptive Candidate Sampling

- We present a **Token-level Adaptive Candidate Sampling (TACS)** that utilize the **Pointwise Mutual Information (PMI)** between each token $\mathbf{z}_i$ and the sequence $\mathbf{x}$.

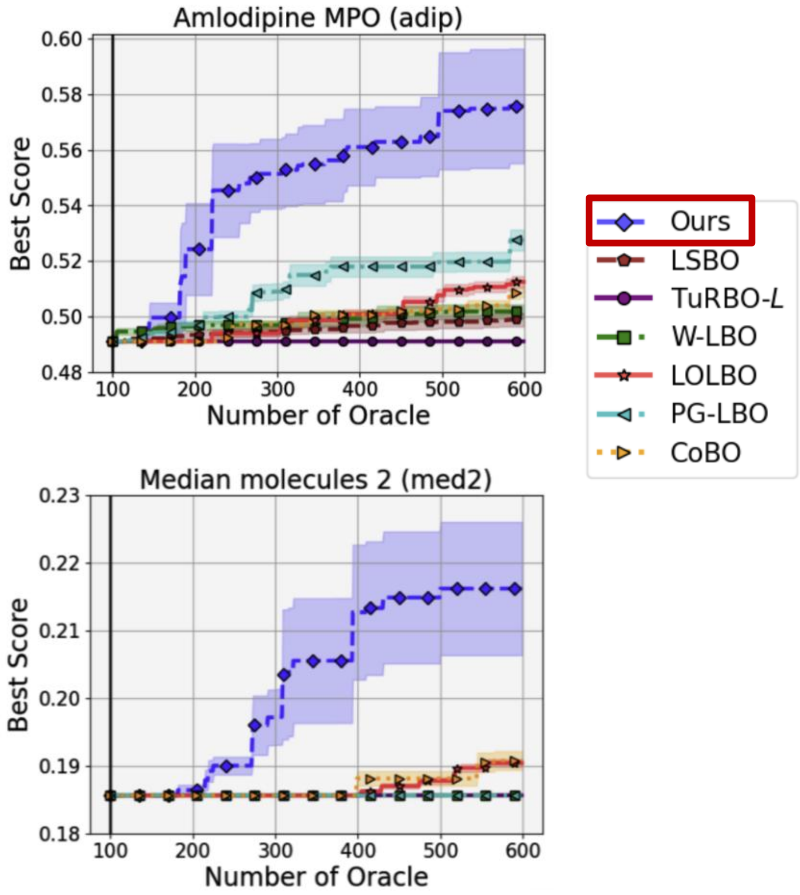


$$\omega_i(\mathbf{z}) = \mathrm{PMI}(\mathbf{x}, \mathbf{z}_i | \mathbf{z}_{-i}) = \log \frac{p(\mathbf{x}|\mathbf{z})}{p(\mathbf{x}|\mathbf{z}_{-i})} = \log \frac{p(\mathbf{x}|\mathbf{z})}{\mathbb{E}_{\mathbf{z}_i \sim \mathcal{N}(\mathbf{0}, I)}(p(\mathbf{x}|\mathbf{z}))},$$

$$\pi_i(\mathbf{z}) = \min\left(\kappa s_i(\mathbf{z}), 1\right), \quad s_i(\mathbf{z}) = \frac{\exp\left(\omega_i(\mathbf{z})/\tau\right)}{\sum_j \exp\left(\omega_j(\mathbf{z})/\tau\right)},$$

NF-BO: Experiments - Molecule design tasks

Methods	Assembly	Top-1 Score (Rank)	Top-10 Score (Rank)	Top-100 Score (Rank)	AUC Top-1 Score (Rank)	AUC Top-10 Score (Rank)	AUC Top-100 Score (Rank)
Bayesian Optimization							
NF-BO	**SELFIES**	**18.095 (1)**	**17.692 (1)**	**17.037 (1)**	**15.539 (1)**	**14.737 (1)**	13.423 (2)
GP BO	Fragments	15.345 (7)	14.940 (6)	14.365 (6)	13.798 (5)	13.156 (5)	12.122 (6)
VAE BO	SELFIES	11.423 (17)	9.788 (19)	7.622 (22)	10.589 (17)	8.887 (19)	6.899 (22)
VAE BO	SMILES	10.926 (21)	9.435 (21)	7.623 (21)	10.197 (19)	8.587 (21)	6.909 (21)
JT-VAE BO	Fragments	10.296 (23)	8.671 (24)	7.037 (24)	9.973 (22)	8.358 (24)	6.740 (23)
Reinforcement Learning							
REINVENT	SMILES	16.772 (2)	16.654 (2)	16.297 (2)	14.711 (2)	14.196 (2)	**13.445 (1)**
REINVENT	SELFIES	16.059 (5)	15.889 (4)	15.377 (3)	14.077 (4)	13.471 (4)	12.475 (5)
MolDQN	Atoms	7.143 (26)	6.495 (26)	5.435 (26)	6.332 (26)	5.620 (26)	4.528 (26)
Genetic Algorithm							
Graph GA	Fragments	16.244 (4)	15.946 (3)	15.342 (4)	14.356 (3)	13.751 (3)	12.696 (3)
STONED	SELFIES	14.257 (8)	14.201 (8)	14.017 (7)	13.256 (7)	13.024 (6)	12.518 (4)
SMILES GA	SMILES	13.123 (11)	12.997 (9)	12.824 (9)	12.357 (10)	12.054 (8)	11.598 (7)
SynNet	Synthesis	13.105 (12)	12.279 (12)	10.768 (15)	12.425 (9)	11.498 (9)	9.914 (9)
GA+D	SELFIES	11.942 (16)	11.696 (15)	11.230 (13)	9.387 (24)	8.964 (18)	8.280 (15)
Hill Climbing							
LSTM HC	SMILES	16.754 (3)	15.880 (5)	14.621 (5)	13.611 (8)	12.223 (7)	10.365 (8)
LSTM HC	SELFIES	13.770 (9)	12.894 (10)	11.657 (12)	11.441 (14)	10.246 (15)	8.595 (13)
DoG-Gen	Synthesis	15.633 (6)	14.772 (7)	13.653 (8)	12.721 (8)	11.456 (10)	9.635 (12)
MIMOSA	Fragments	12.524 (15)	12.223 (13)	11.717 (11)	11.378 (15)	10.651 (13)	9.708 (11)



- Practical Molecular Optimization (PMO) benchmarks (Gao, Wenhao, et al.)
- Guacamol benchmarks (Brown, Nathan, et al.)

NF-BO: Analysis – SeqFlow

Compare SeqFlow to TextFlow[1]

- We compare our base flow model, **TextFlow**, with our **SeqFlow** in two aspects:
- **(1) Reconstruction ability,** and **(2) Optimization performance.**
- Both models share the **same Normalizing Flow (NF) component.**

1) Ziegler, Zachary, and Alexander Rush. "Latent normalizing flows for discrete sequences." International Conference on Machine Learning. PMLR, 2019.

NF-BO: Analysis – SeqFlow

(1) Reconstruction ability.

- We evaluate **reconstruction ability** by measuring the **ratio of $y \neq \hat{y}$.**
- **TextFlow[1] fails to reconstruct accurately** due to its BiLSTM-based mapping, whereas **our SeqFlow** achieves **perfect reconstruction** with an invertible function**.**

Task	SeqFlow	TextFlow
Adip	**0.000**	0.548
Med2	**0.000**	0.609
Osmb	**0.000**	0.630
Pdop	**0.000**	0.502
Rano	**0.000**	0.814
Zale	**0.000**	0.750

The ratio of $y \neq \hat{y}$

1) Ziegler, Zachary, and Alexander Rush. "Latent normalizing flows for discrete sequences." International Conference on Machine Learning. PMLR, 2019.

NF-BO: Analysis – SeqFlow

(2) Optimization performance

- Despite having **more parameters, TextFlow[1] performs similarly to the baselines**, as inaccurate reconstruction limits its effectiveness.
- **SeqFlow**, with fewer parameters and exact reconstruction, **outperforms TextFlow** in Bayesian Optimization tasks.

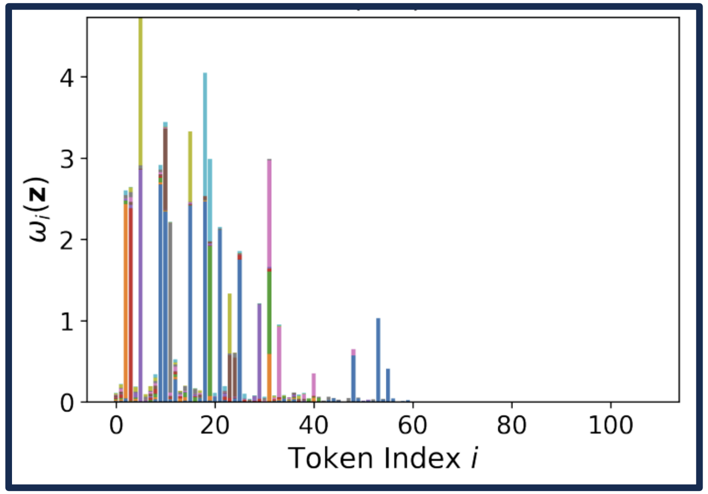
Method	SeqFlow	TextFlow
Base NF	Autoregressive NF	Autoregressive NF
Perfect Reconsturction	**O**	X
# Params	**31M**	54M
Adip Score	**0.778**	0.716
Med2 Score	**0.372**	0.347

1) Ziegler, Zachary, and Alexander Rush. "Latent normalizing flows for discrete sequences." International Conference on Machine Learning. PMLR, 2019.

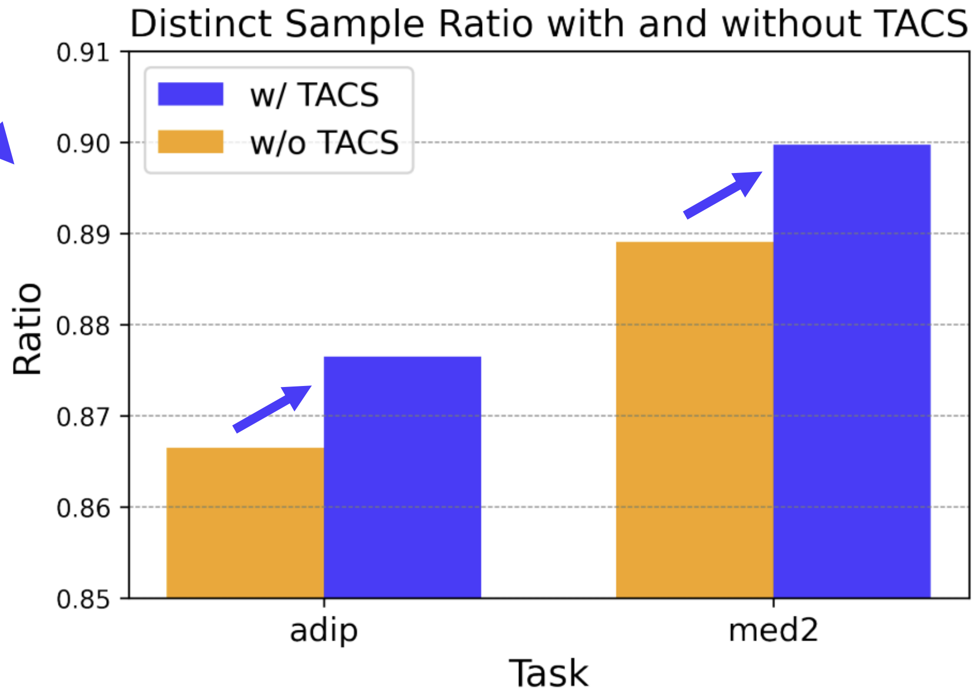
NF-BO: Analysis - TACS

Impact of TACS (Token-level Adaptive Candidate Sampling)

- Model **with TACS** makes a higher ratio of **distinct samples** compared to those without TACS.



PMI (Pointwise Mutual-information) values on each token i.



Conclusion

✓**Normalizing Flow-based Latent Bayesian Optimization (NF-BO)**

✓**Perfect reconstruction** via **invertible normalizing flows**.

✓**Resolves value discrepancy problem** in **latent Bayesian optimization**.

✓**Token-level Adaptive Candidate Sampling** improves **local search efficiency**.

Thank you.

Paper