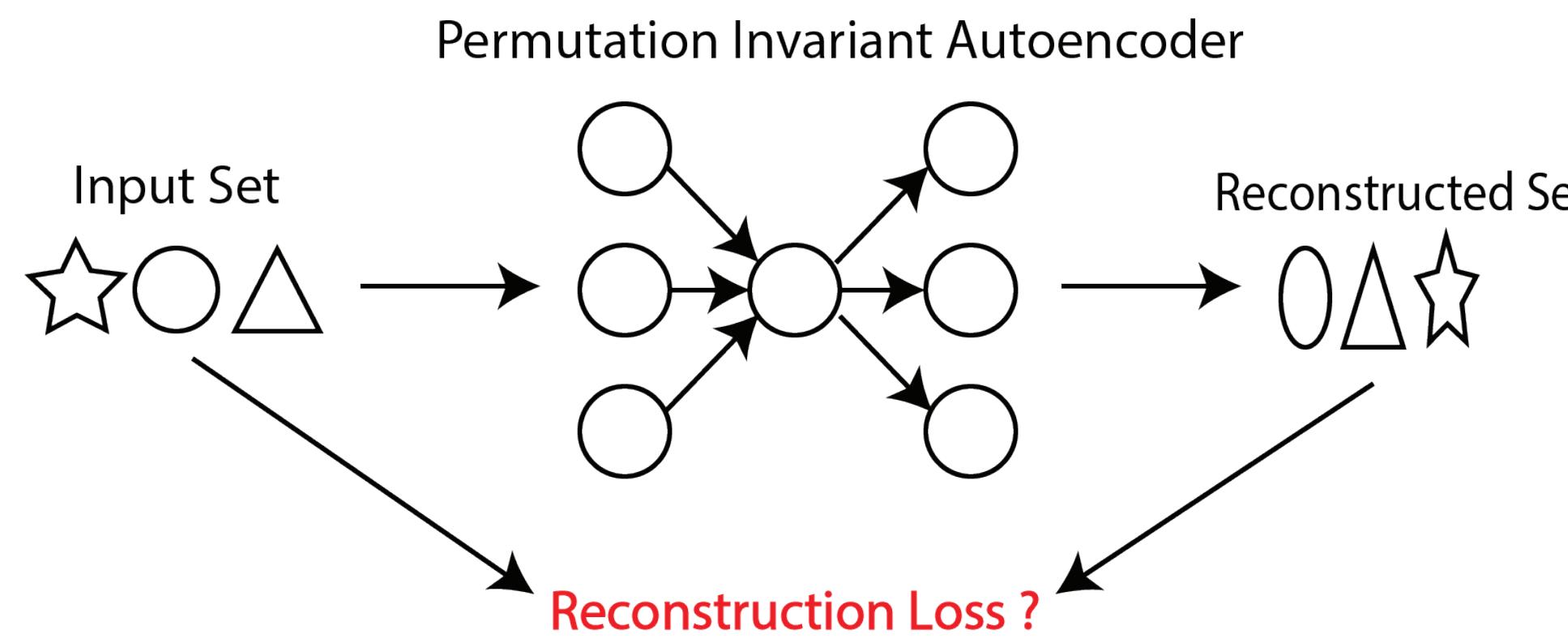
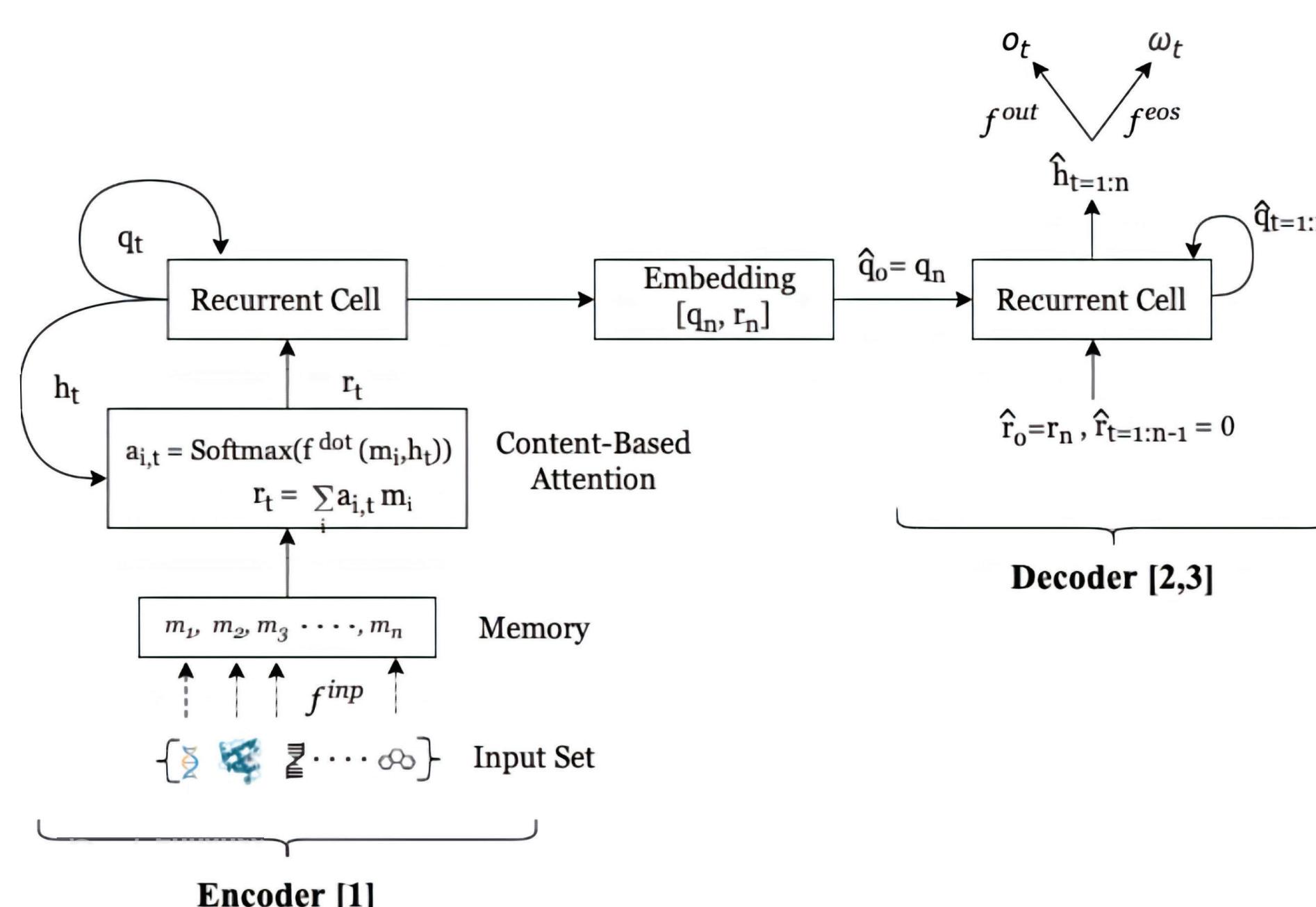


Introduction



- We build upon previous work to introduce a Fully Differentiable Set Autoencoder (FDFA) that learns permutation-invariant representations of multiple modalities packaged as a set.
- Permutation invariant encoding necessitates post-hoc alignment for loss computation.
- We address this issue by a set matching network that efficiently approximates the combinatorial optimization problem of aligning the input and the reconstructed sets.

Backbone: Overcomplete Autoencoder



- The Set Autoencoder is made up of recurrent units in both encoder and decoder.
- The model achieves permutation invariance in the encoder through content-based attention.
- Subsequently, the outputs of the decoder also show permutation invariance and are aligned by the matching network before computing the loss.

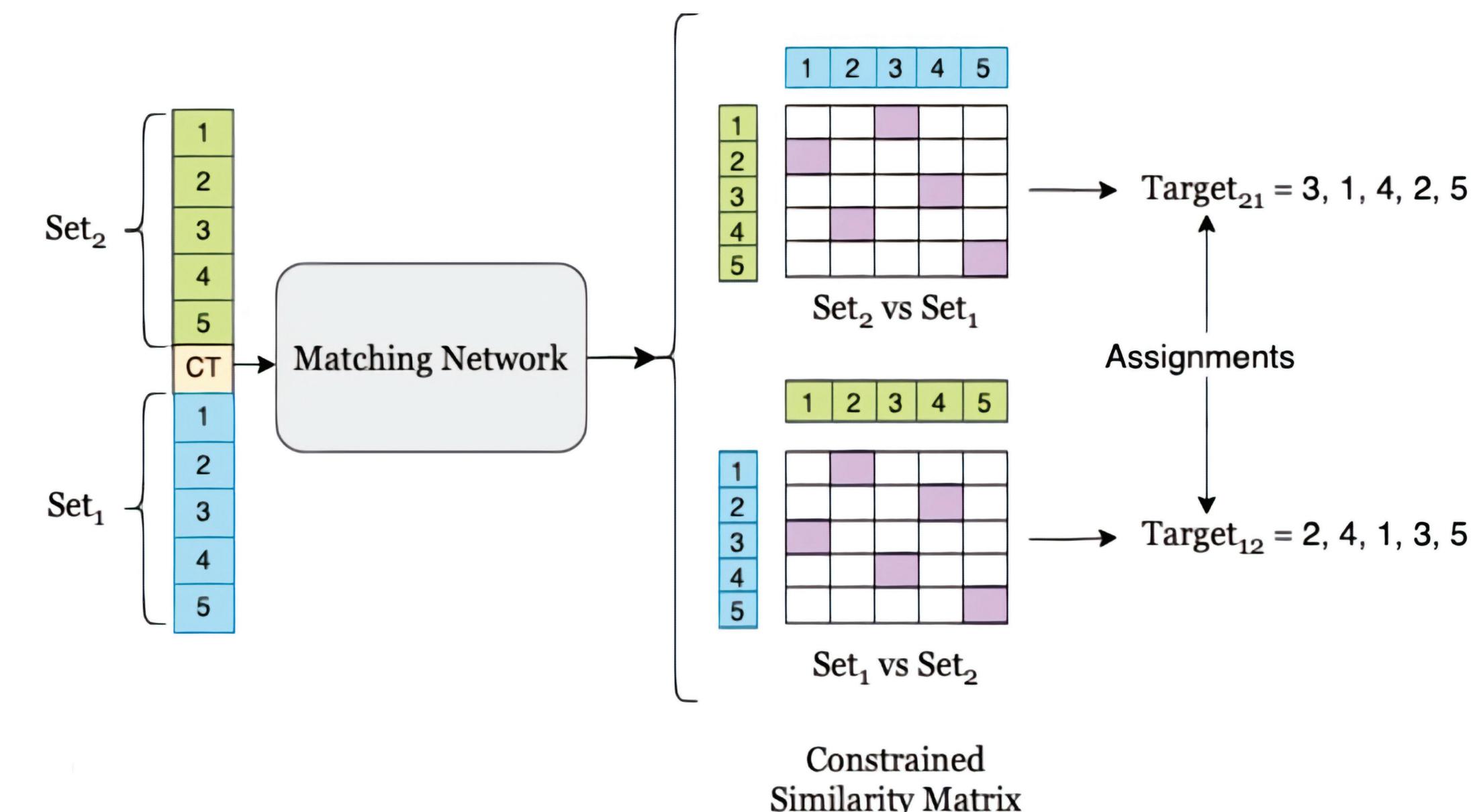
References

- Vinyals, Oriol, Samy Bengio, and Manjunath Kudlur. "Order matters: Sequence to sequence for sets." arXiv preprint arXiv:1511.06391 (2015).
- Probst, Malte. "The Set Autoencoder: Unsupervised Representation Learning for Sets." (2018).
- Vinyals, Oriol, Meire Fortunato, and Navdeep Jaitly. "Pointer networks." Advances in neural information processing systems. 2015.

Scan for GitHub

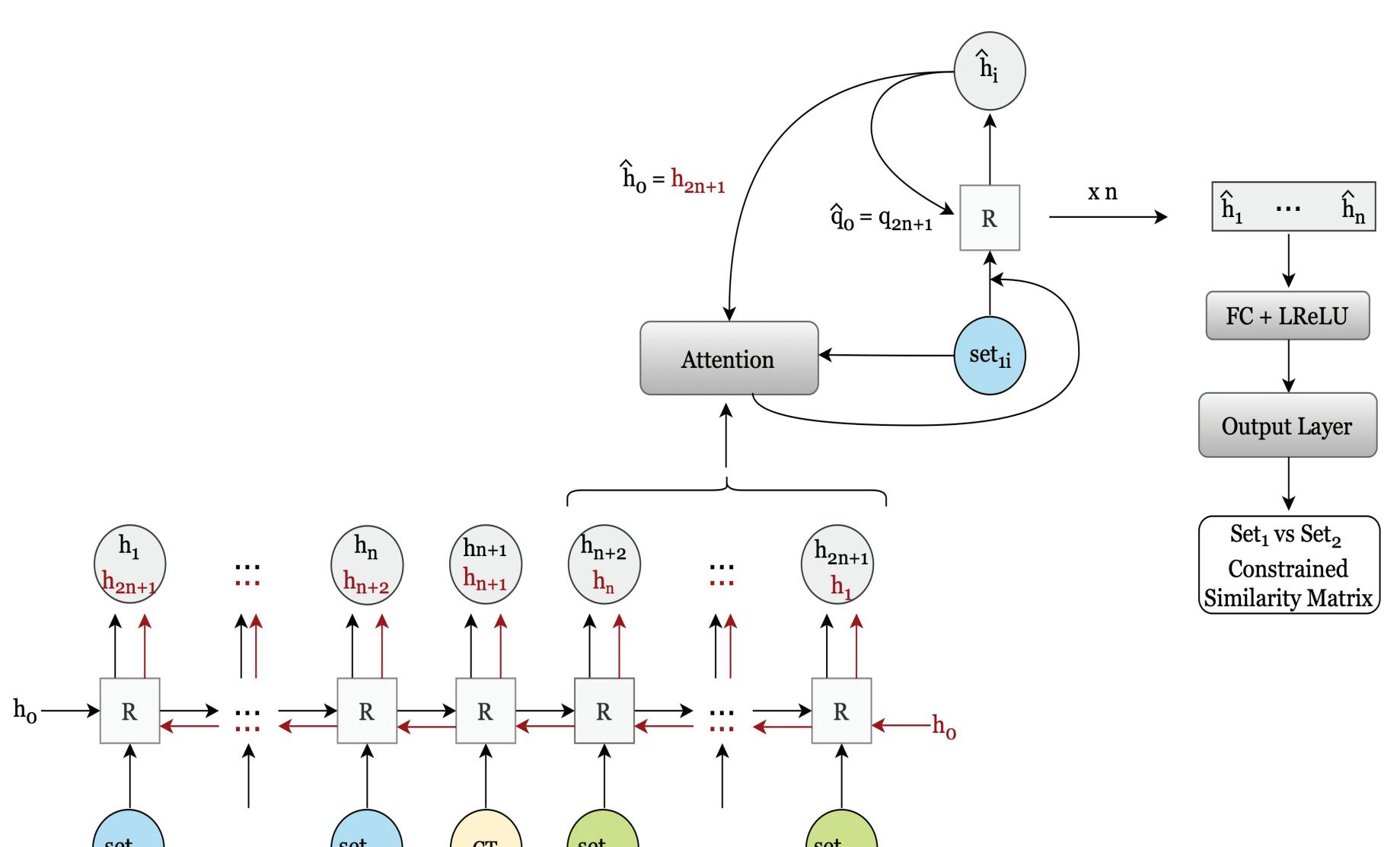


Set matching : A combinatorial optimisation problem



- Goal: Find unique assignments such that the distance between the mapped elements is minimised.

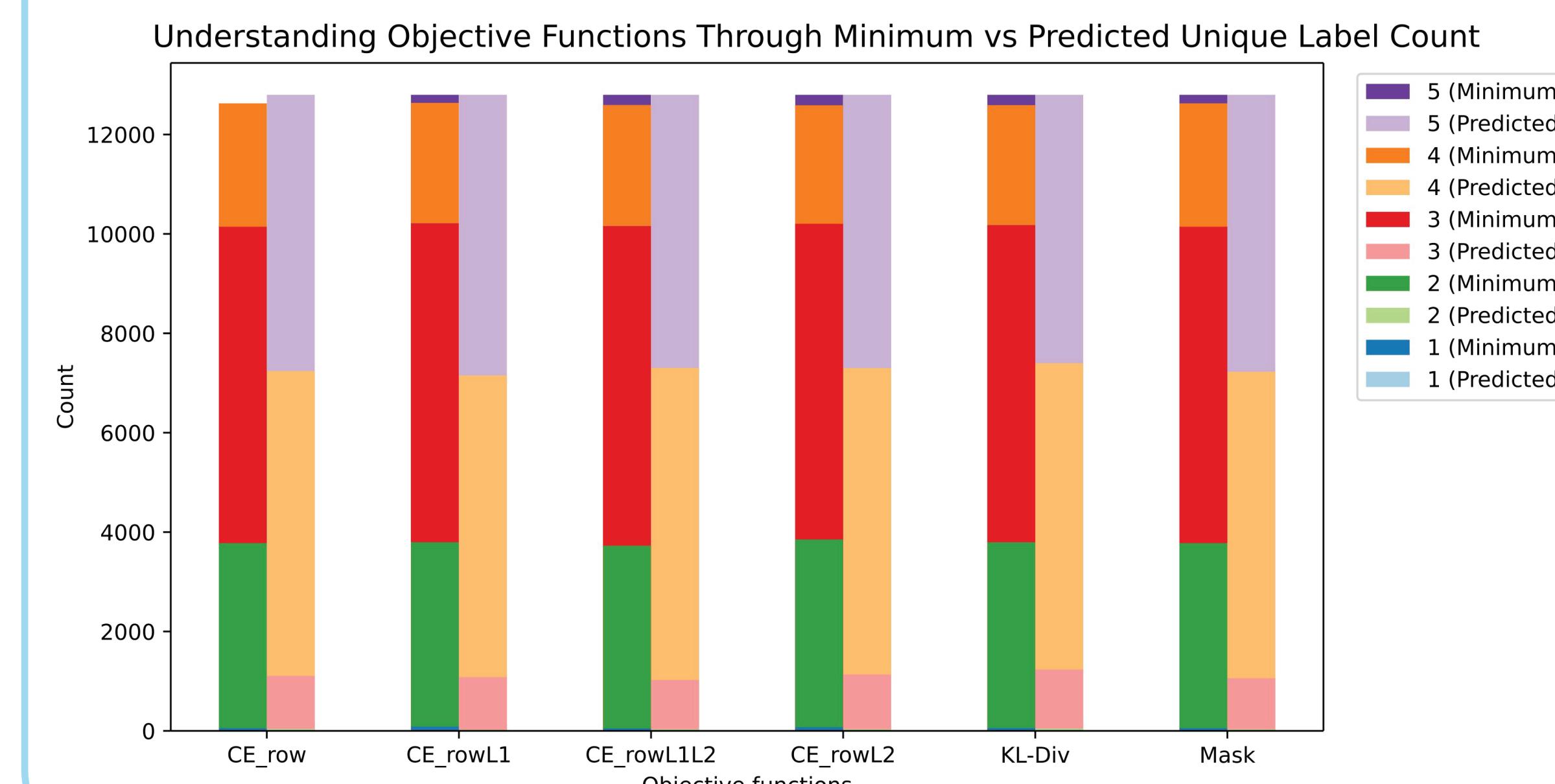
Seq2Seq model with attention has the best accuracy in matching sets



- The table below summarises the best accuracy achieved by each model in matching 128D homogeneous sets with lengths varying from 2-5 and the loss function (10 objectives tested) that contributed to it. KL abbreviates KL Divergence and CE is Cross Entropy.

Model	RNN	BiRNN	Seq2Seq			
Cell	Acc.	Loss	Acc.	Loss	Acc.	Loss
GRU	0.9843	KL _{col}	0.9864	KL _{col}	0.9995	KL _{col}
LSTM	0.9985	KL _{col}	0.9979	KL _{col}	0.9999	KL _{col}
nBRC	0.9856	KL _{col}	0.9821	KL _{col}	0.9990	KL _{col}
BRC	0.2694	CE _{row,L1L2}	0.3370	KL _{col}	0.9902	KL _{col}

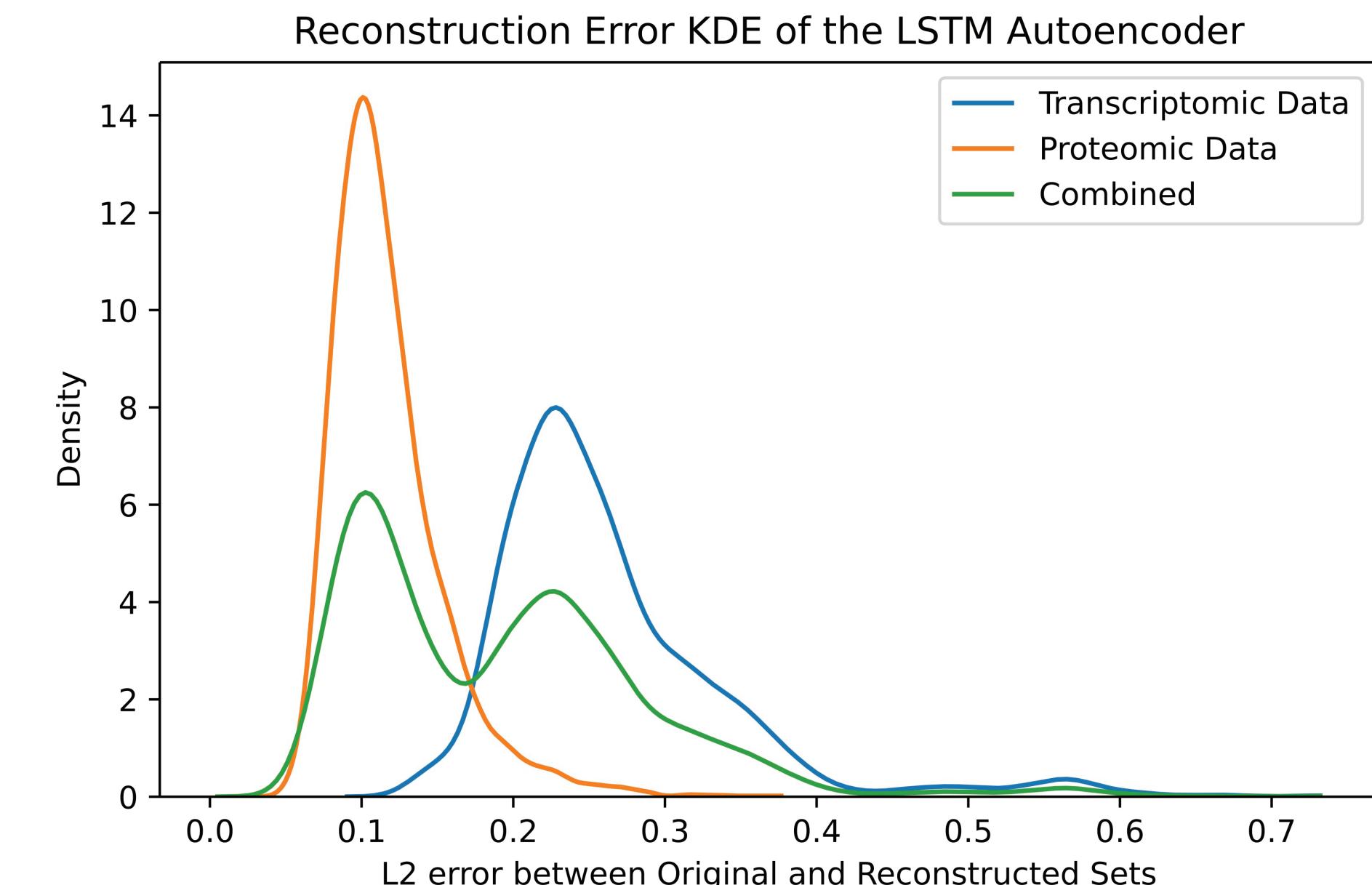
Objective Functions drive the model towards unique label generation



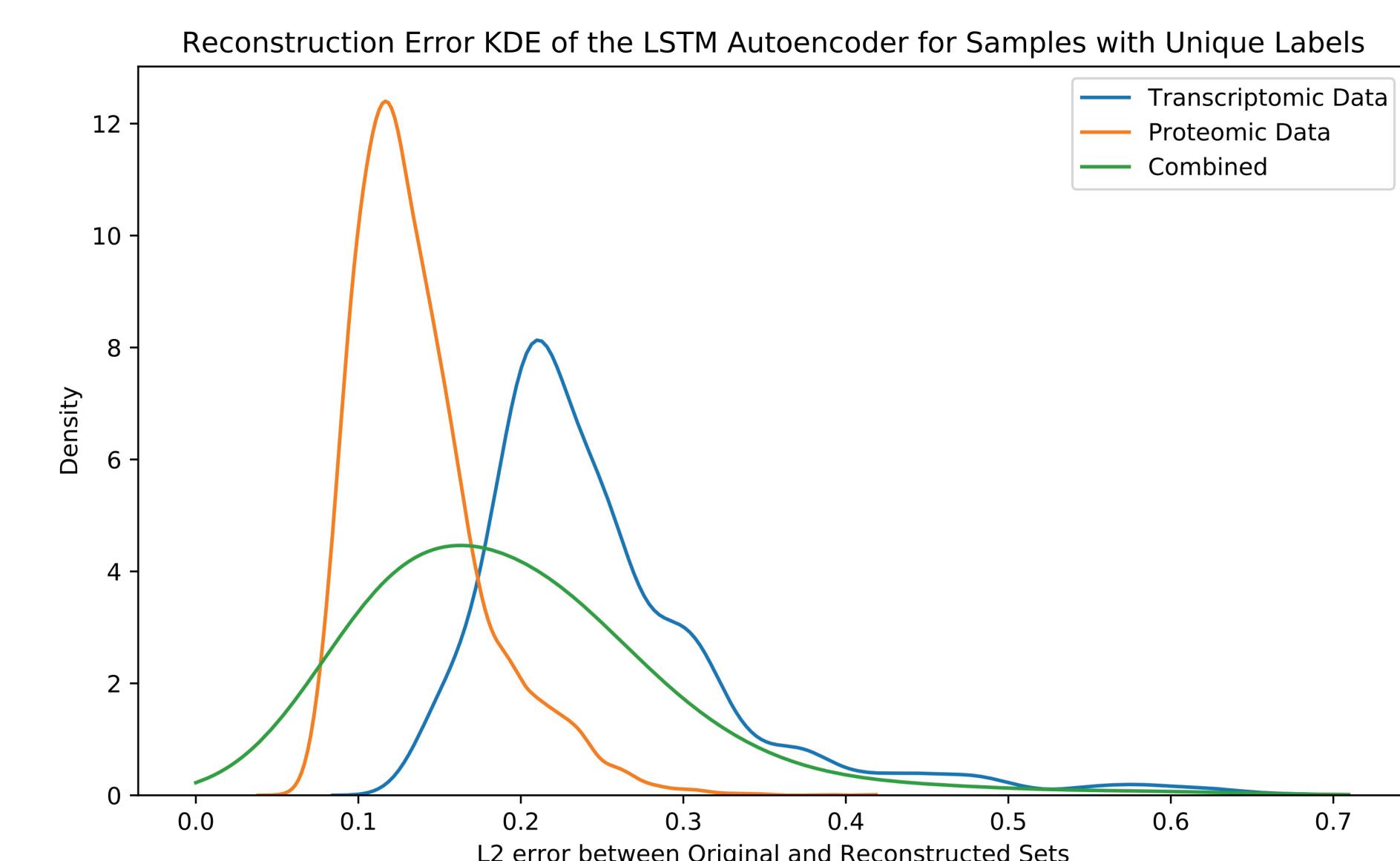
FDFA: Reconstructing bi-modal cancer data

- Data: Bi-modal 128-D cancer dataset containing transcriptomic profiles and associated proteins.

- Baseline : Set autoencoder with the Hungarian algorithm



- Fully Differentiable Set Autoencoder



- The table below summarises the average training time per epoch spent by the Fully Differentiable Set Autoencoder variants versus the baseline in matching sets.

Autoencoder	Network (sec.)	Hungarian (sec.)
GRU	77.44	456.80
LSTM	78.48	518.35
pLSTM	75.74	494.01
nBRC	75.14	476.82

Conclusion

- FDSA fulfils our vision of a multi-modal representation where the mode-specific properties are preserved, the mode-order does not matter (permutation-invariance) and lastly, additional available modes are easy to integrate.
- Replacing an accurate matching algorithm with an approximating network results in a 7x improvement in runtime.
- Need more rigorous testing for the matching network and FDSA to test potential and limits.