NEURAL INFORMATION PROCESSING SYSTEMS

SEAL: Structured Energy network As a Loss function

Jay-Yoon Lee[¶], Dhruvesh Patel[♣], Purujit Goyal[♣],
Wenlong Zhao[♣], Zhiyang Xu[◊], Andrew McCallum[♣]

[¶] Seoul National University
Graduate School of Data Science

[♣] UMassAmherst
Robert and Donna Manning
College of Information
& Computer Sciences

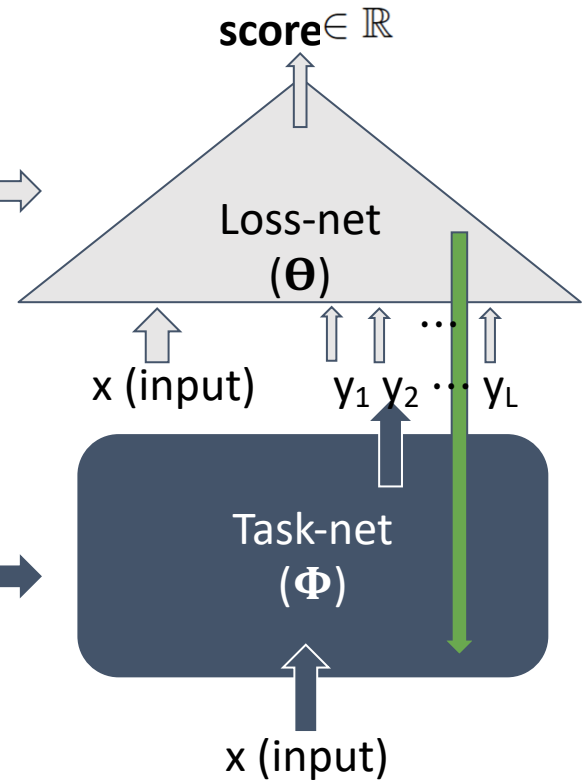
[◊] College of Engineering
Computer Science
Virginia Tech

SEAL:
Structured
Energy network
As a
Loss

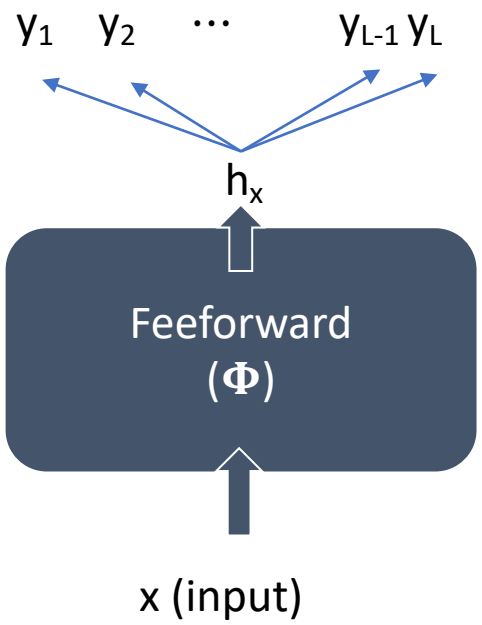
Structured Energy network As a Loss (SEAL)

Trainable loss function

Network used
at inference time.

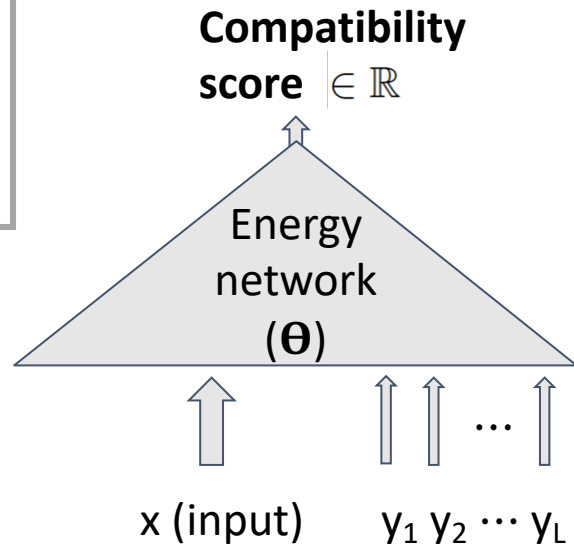


Motivation: Energy network vs. Feedforward

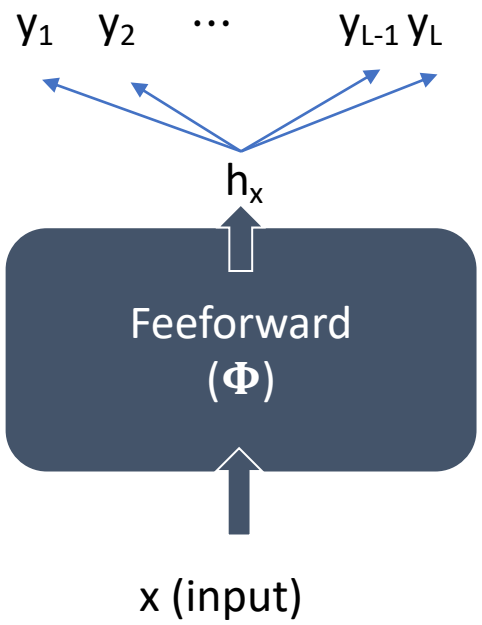


Feedforward model:
Representation of "x" takes care of output dependencies.

Energy-based model:
Captures dependency across multivariate output.



Motivation: Energy network vs. Feedforward

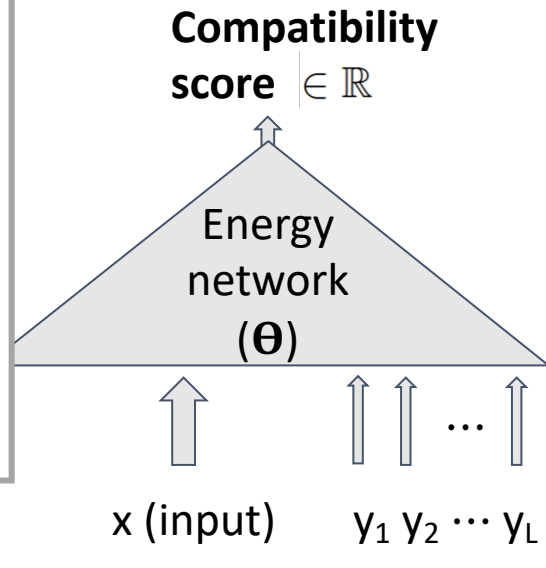


Feedforward model:
Representation of "x" takes care of output dependencies.

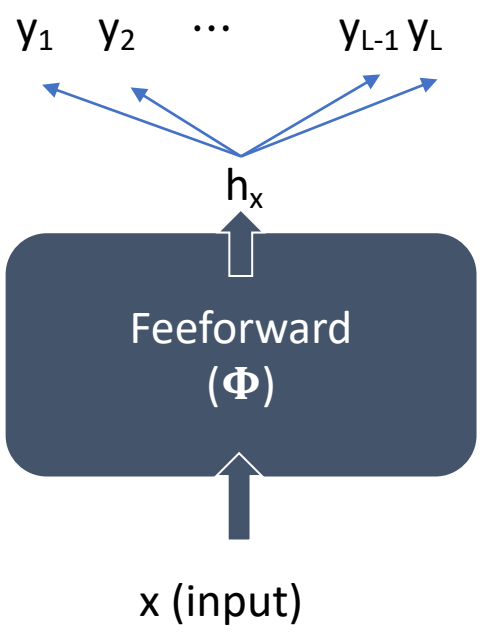
- Light, fast at inference.
- But hard to capture dependencies between labels.

Energy-based model:
Captures dependency across multivariate output

- Captures label dependencies.
- Thus, more accurate.
- Slow, unstable at inference.



Motivation: Energy network vs. Feedforward



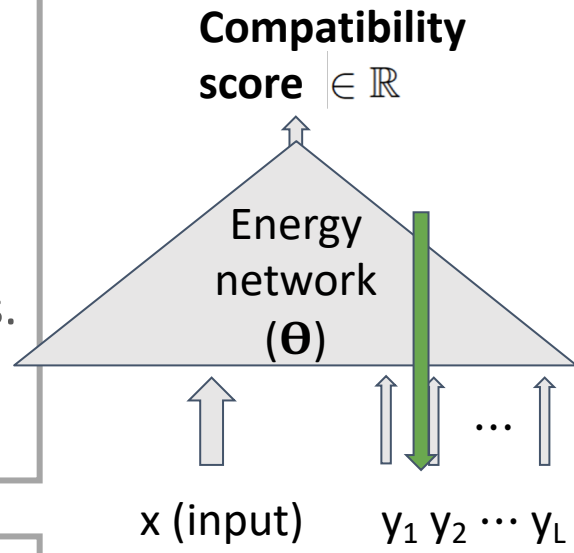
Feedforward model:
Representation of "x" takes care of output dependencies.

- Light, fast at inference.
- But hard to capture dependencies between labels.

Energy-based model:
Captures dependency across multivariate output

- Captures label dependencies.
- Thus, more accurate.
- Slow, unstable at inference.

Gradient-Based Inference (GBI)
to search
combinatorial output space.



Structured **E**nergy network **A**s a **L**oss (SEAL)

Loss-net:

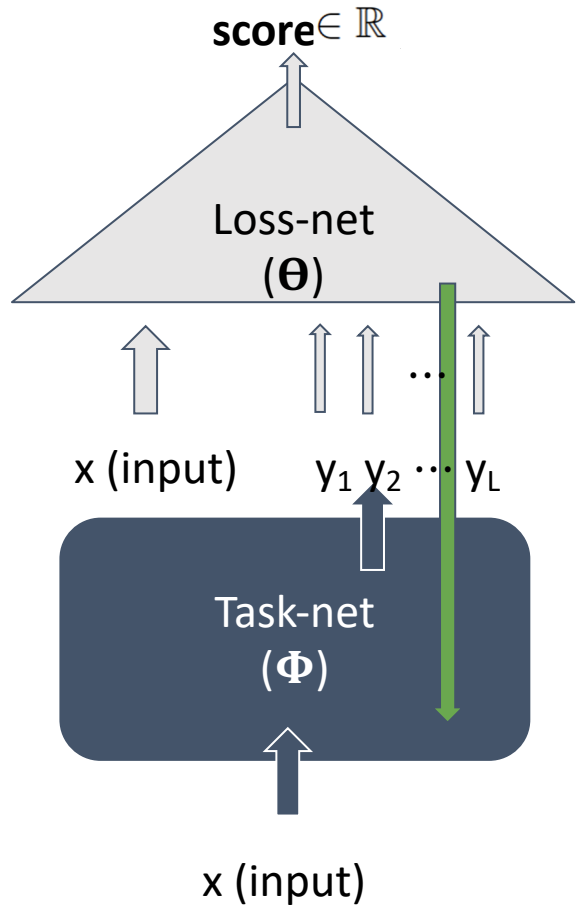
- Structured energy network.

Task-net:

- Fast inference with feedforward.

Two versions:

- SEAL-static
- SEAL-dynamic



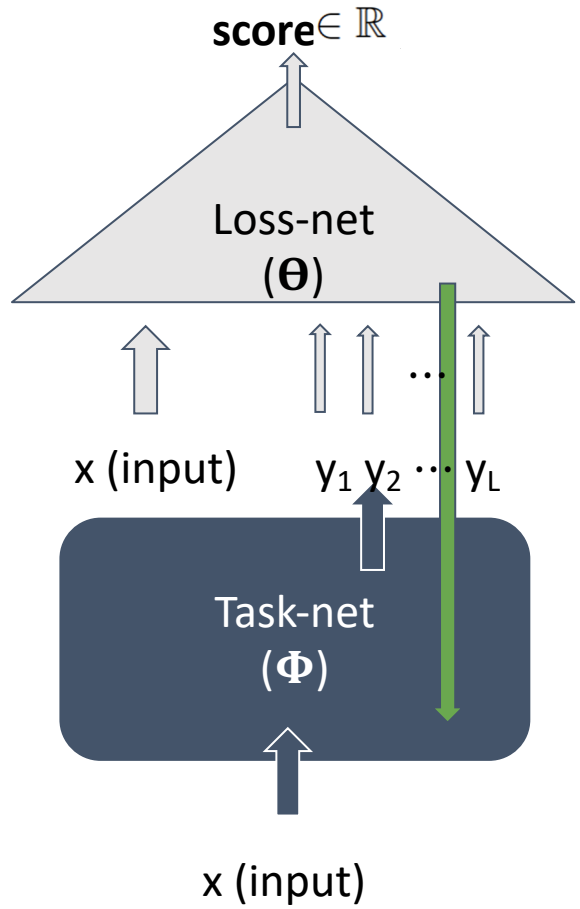
SEAL-static

Loss-net:

- Trained beforehand,
  "fixed" while training task-net.

Task-net:

- Fast inference with feedforward.



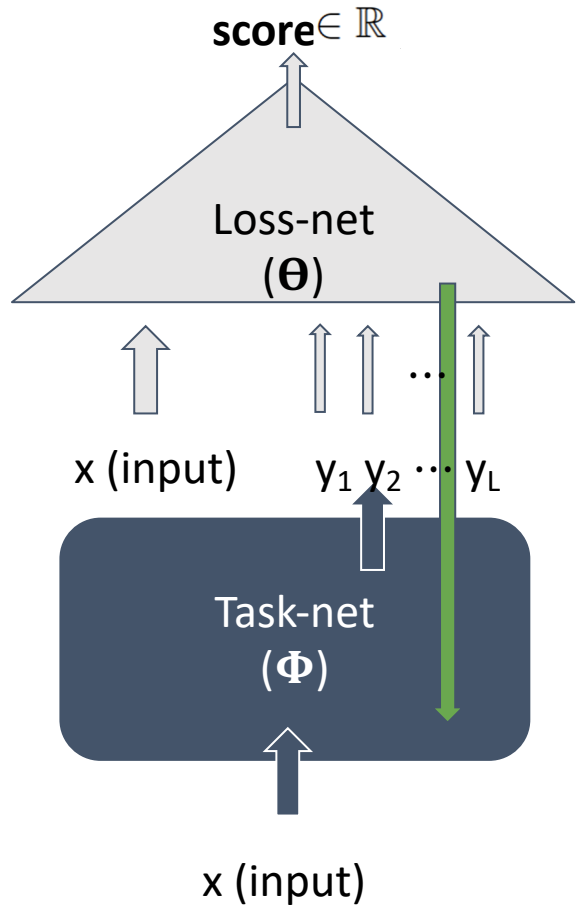
SEAL-dynamic

Loss-net:

- Structured energy network that dynamically adapts to task-net.

Task-net:

- Fast inference with feedforward.

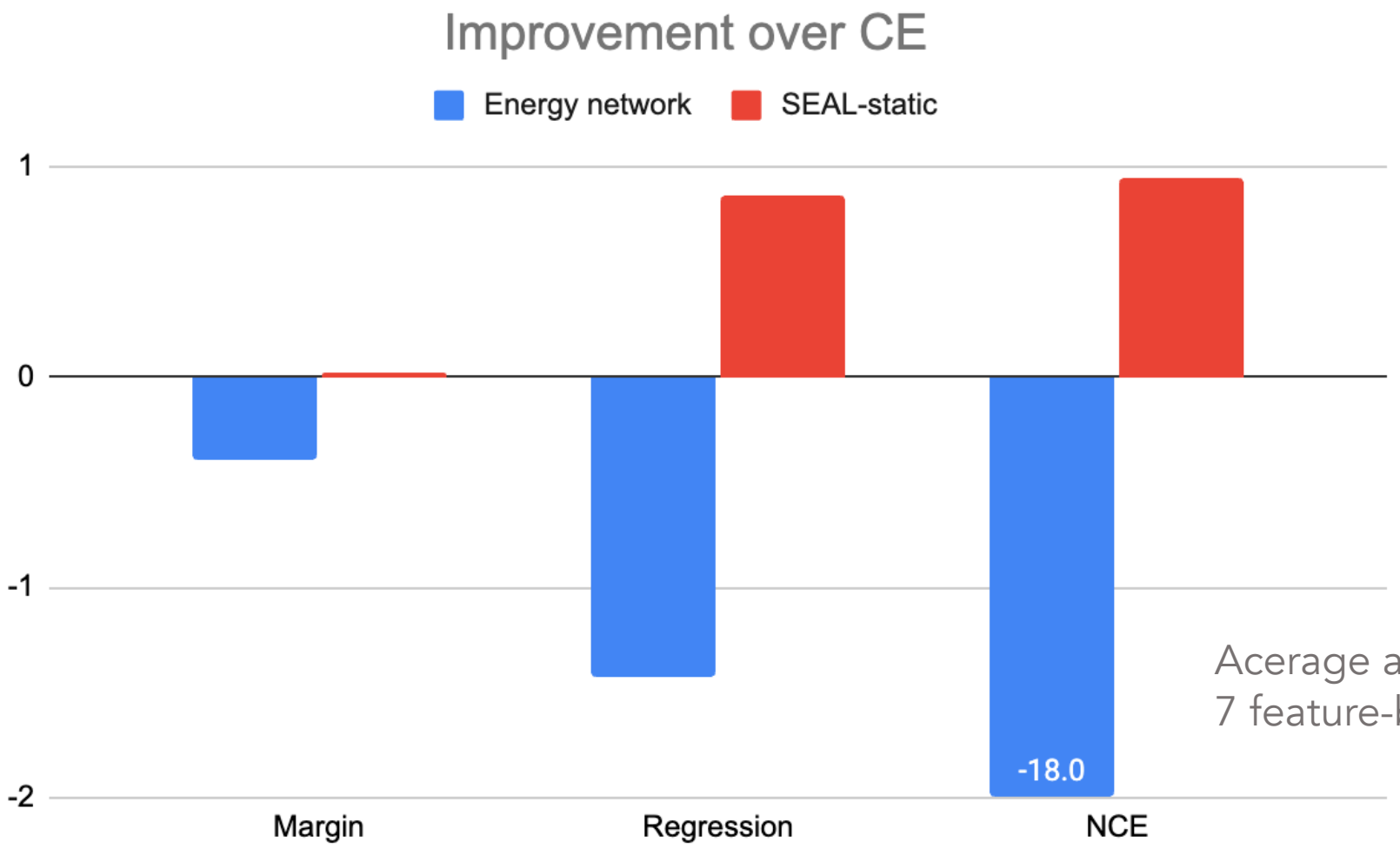


Experiments overview

- Examined three losses for energy netwrok
  - SSVM loss[1]
  - Regression loss[2]
  - Noise-Constrastive Estimation (NCE) ranking loss[3]
- Applications
  - Feature-based MLC
  - Text-based MLC
  - Semantic Role Labeling
  - Binary Image segmentation.

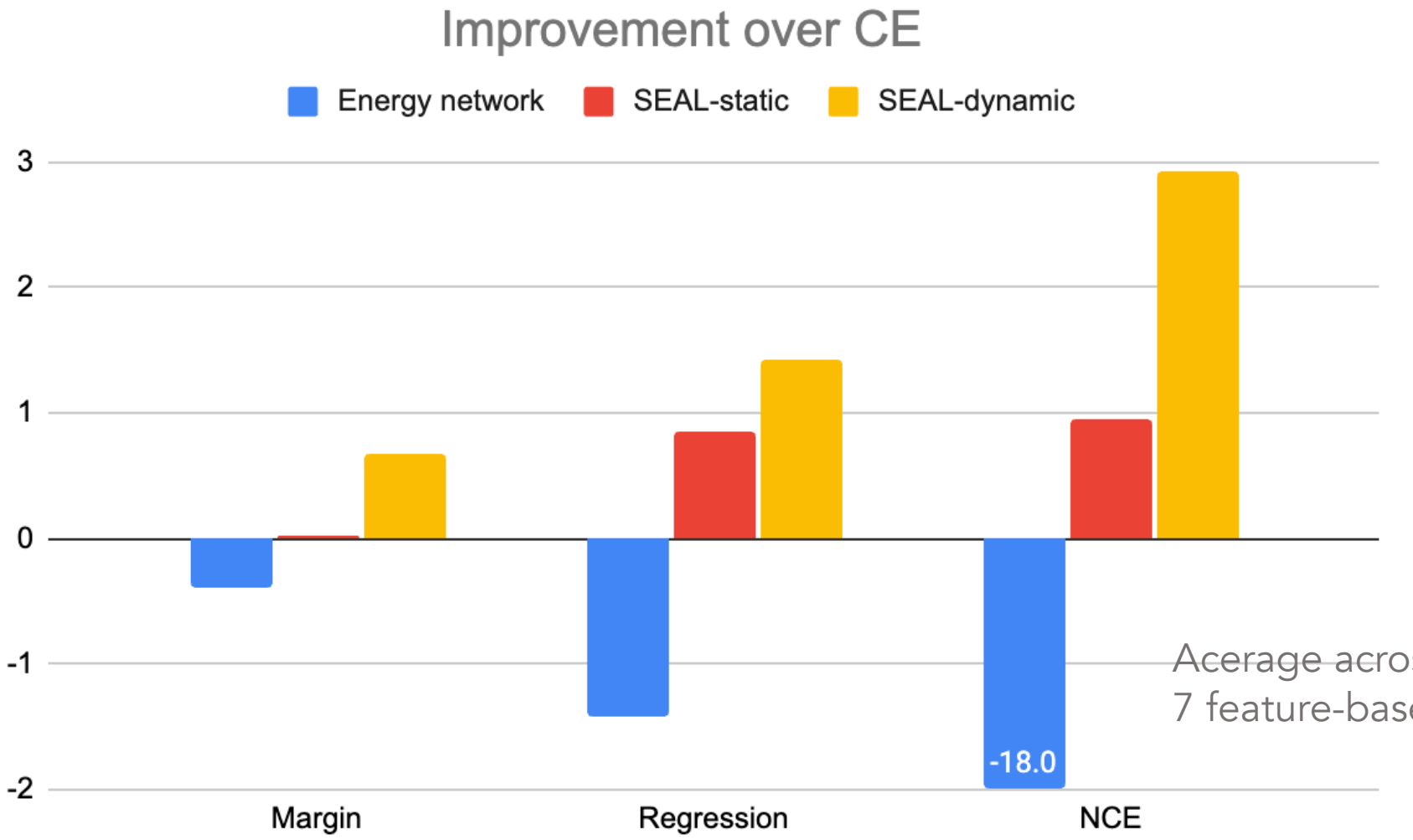
[1] Belanger, D. and McCallum, A. Structured prediction energy networks. *ICML 2016*
[2] Gygli et al., Deep value networks learn to evaluate and iteratively refine structured outputs, *ICML 2017*
[3] Ma, Z. and Collins, M. Noise contrastive estimation and negative sampling for conditional models, EMNLP 2018

Energy network vs. SEAL-static



Acerage across
7 feature-based MLC dataset

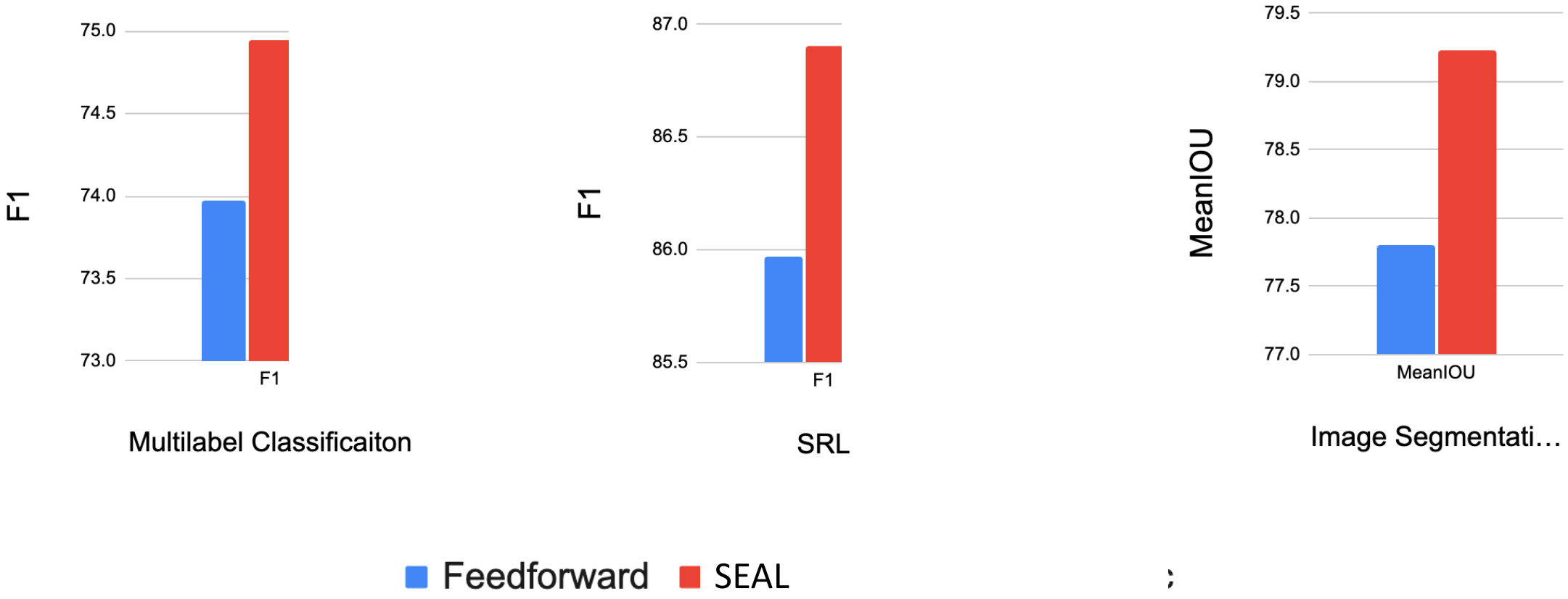
SEAL-static vs. SEAL-dynamic



Acerage across
7 feature-based MLC dataset

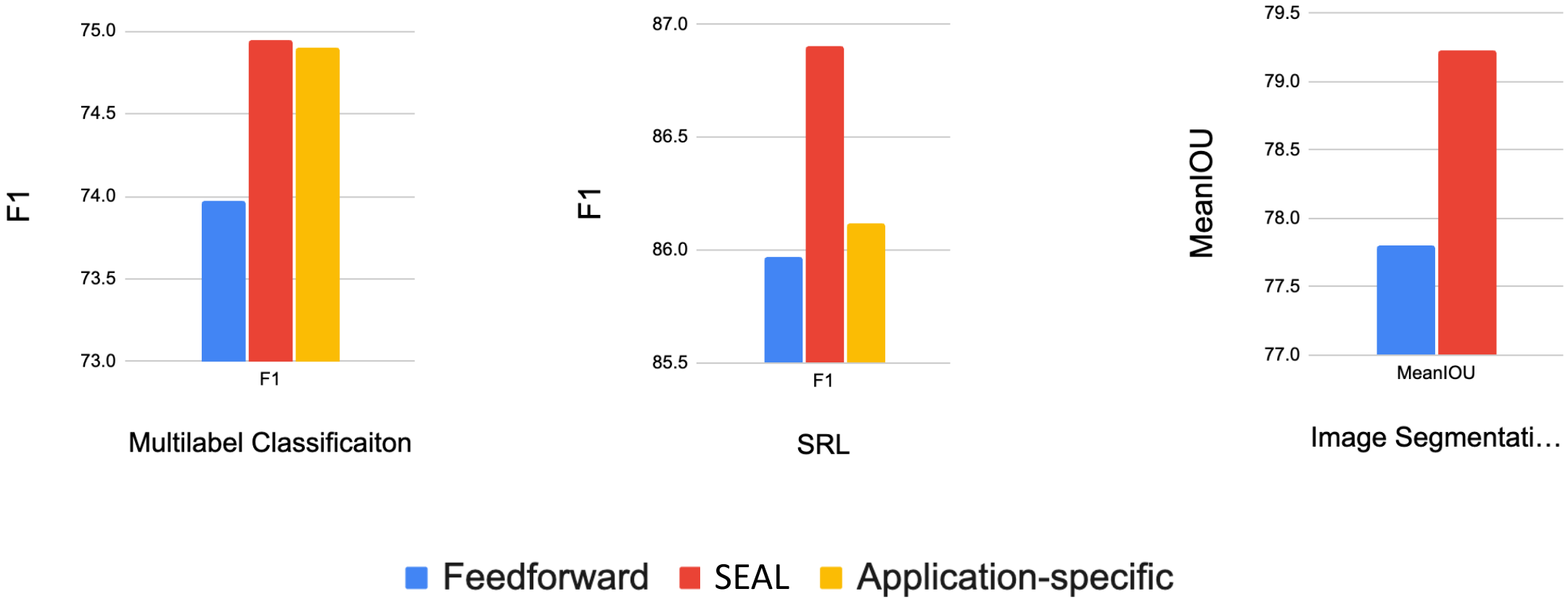
Cross-entropy vs. SEAL

Experiment results



SEAL vs. Application-specific approach

Experiment results



Summary

- SEAL provides
  - better performance than both energy network & feedforward.
  - Faster than energy network, as fast as feedforward.
- Energy network < SEAL-static < SEAL-dynamic

Please visit our poster for more details!