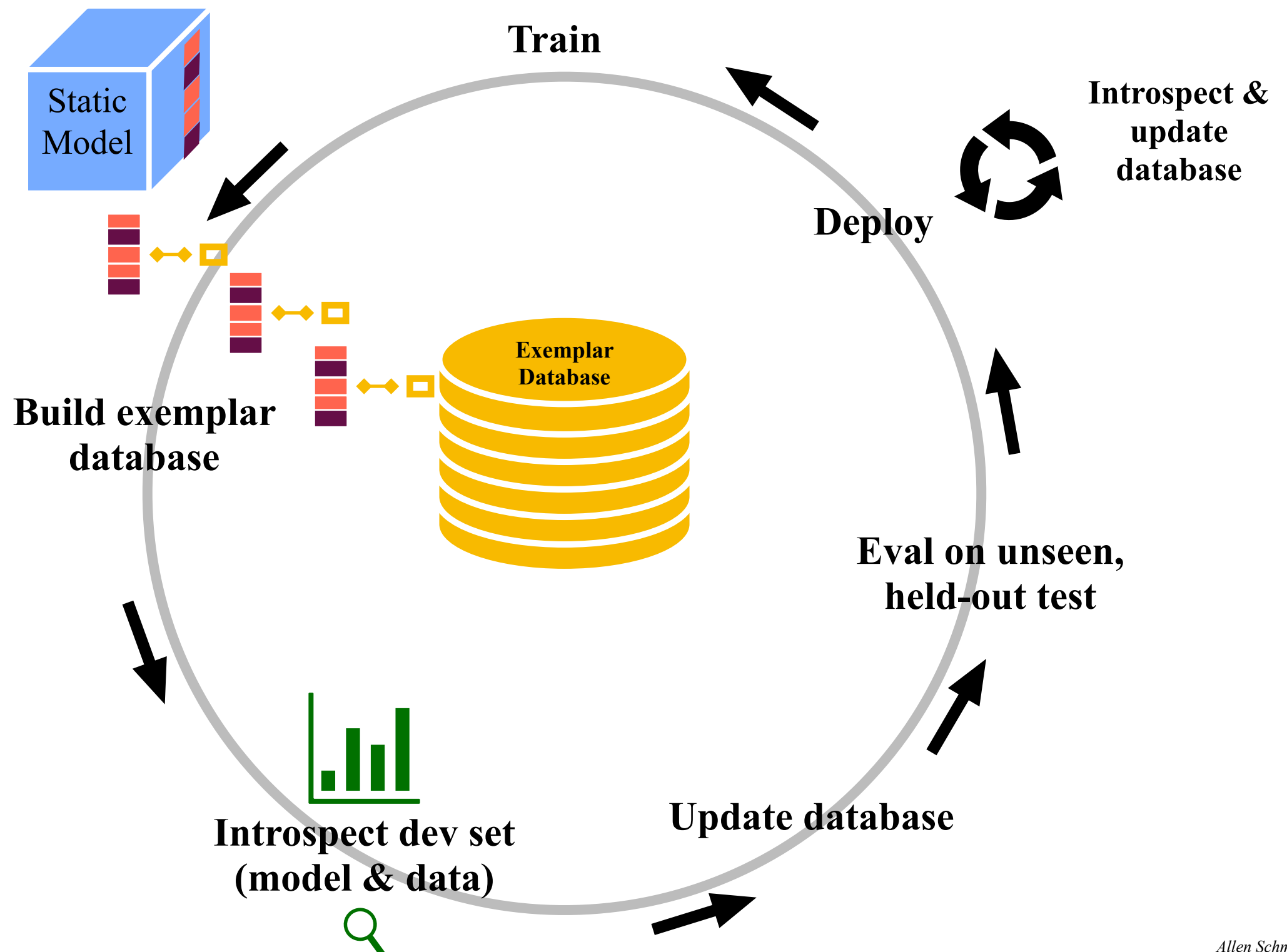


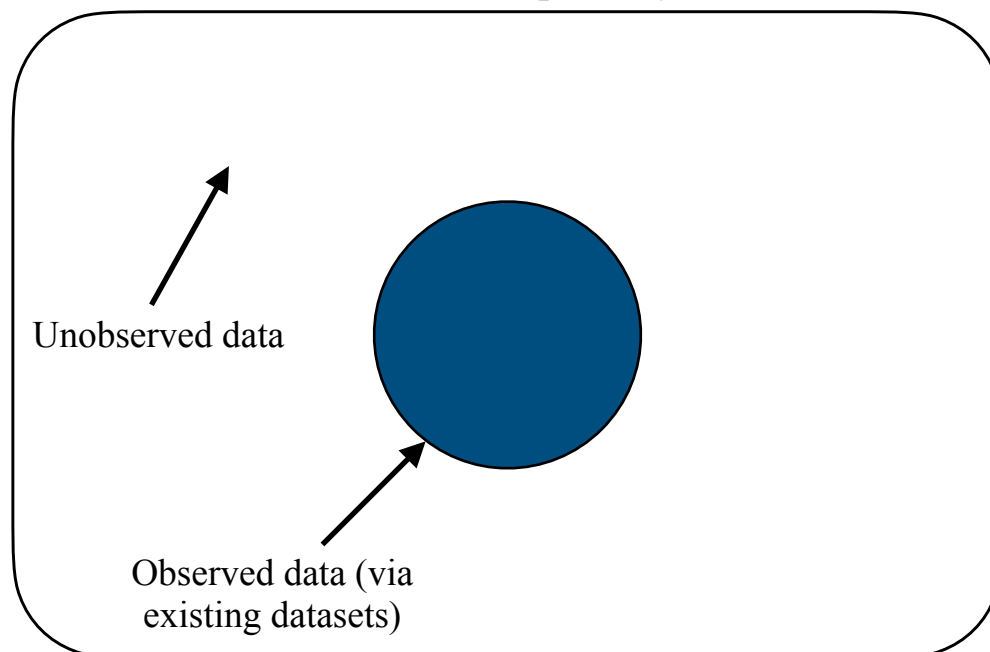
Exemplar Auditing Lifecycle



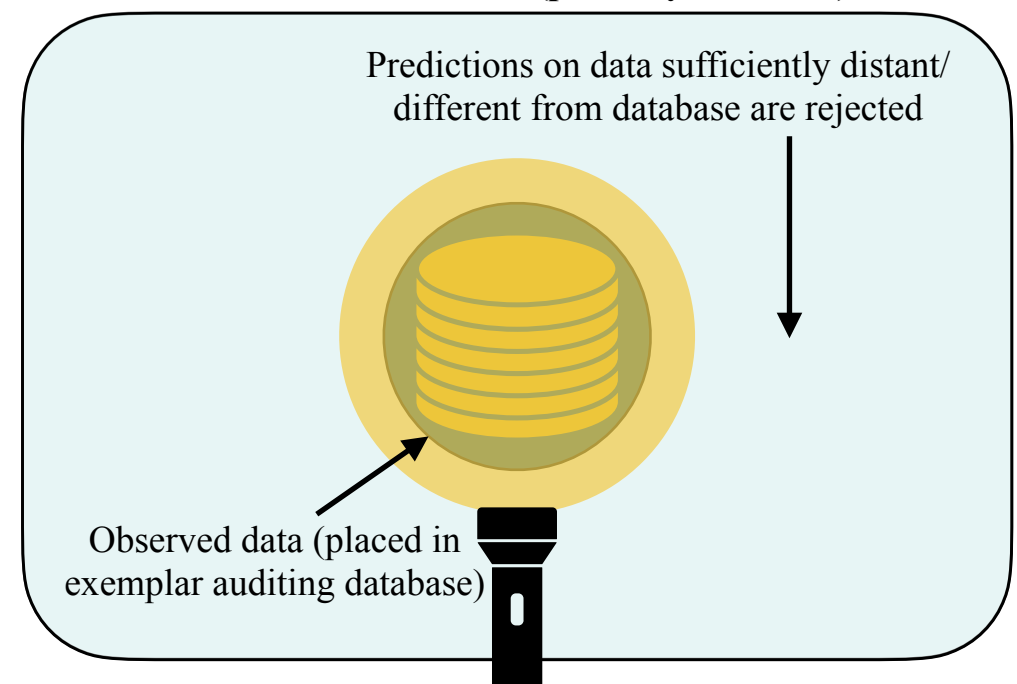
Out-of-Domain Settings

- Pre-train with as much data as possible
- Add as much data as possible to the database
 - Corral the in-domain space, around the ball of the observed data
- Never predict over out-of-domain data in high-risk settings. Instead: Rearrange the deployment to handle non-admitted predictions.

Data distribution for task (partially observed)



Data distribution for task (partially observed)



Implementations

- Binary classification: $f : X \rightarrow \{0,1\}$

Unique side effect: **Sequence labeling:**
 $f : X \rightarrow Y_1, \dots, Y_{|x|}$

- “Detecting Local Insights from Global Labels:
Supervised & Zero-Shot Sequence Labeling via a
Convolutional Decomposition”

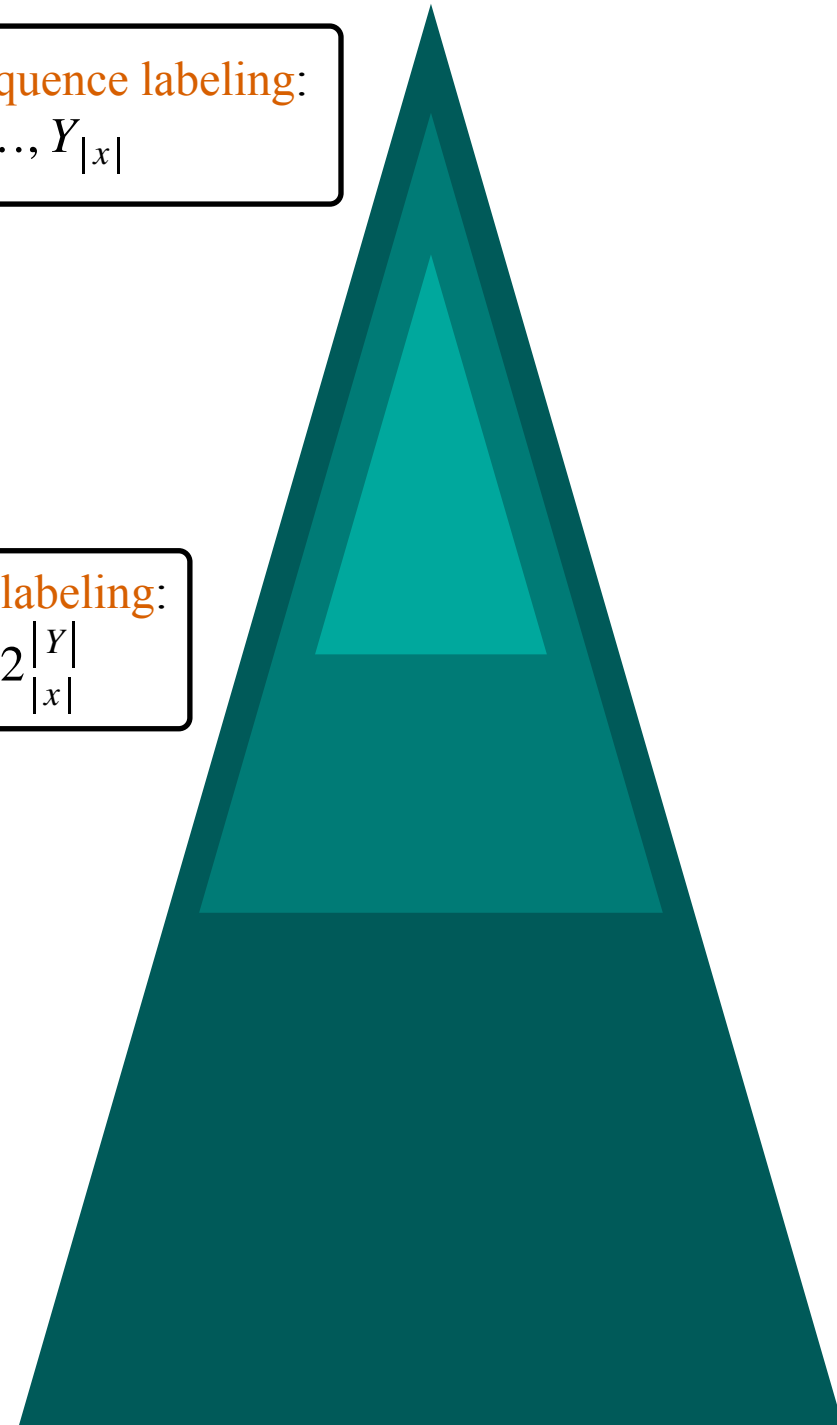
- Multi-label classification: $f : X \rightarrow 2^{|Y|}$

Multi-label sequence labeling:
 $f : X \rightarrow 2_1^{|Y|}, \dots, 2_{|x|}^{|Y|}$

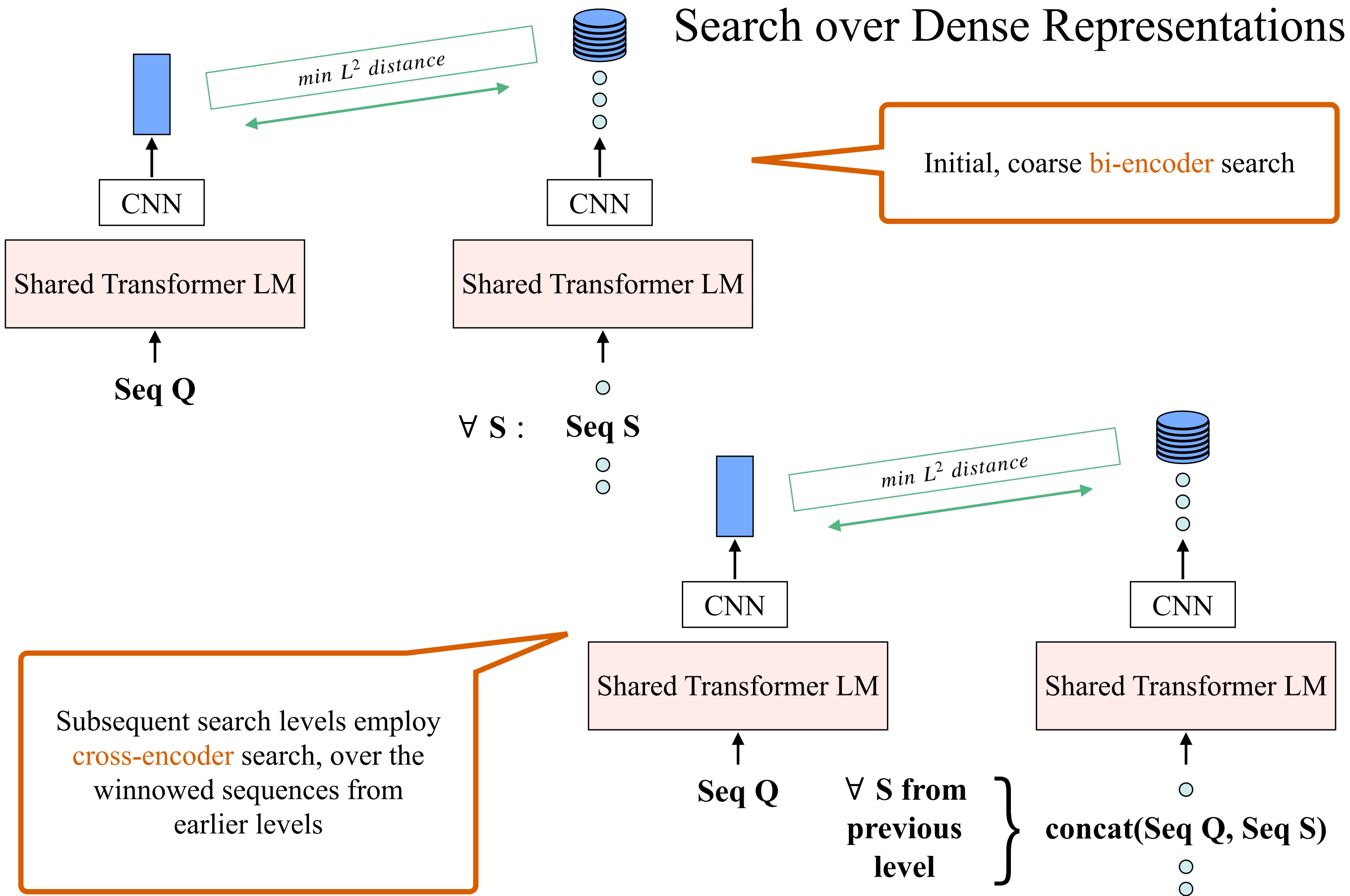
- “Exemplar Auditing for Multi-Label Biomedical
Text Classification”

- Retrieval-classification: $f : X \times \mathcal{D} \rightarrow \langle \{0,1,2\}, 2^{|D|} \rangle$

- “Coarse-to-Fine Memory Matching for Joint
Retrieval and Classification”



An End-to-End Retrieval-Classification Model via a Coarse-to-Fine Search over Dense Representations



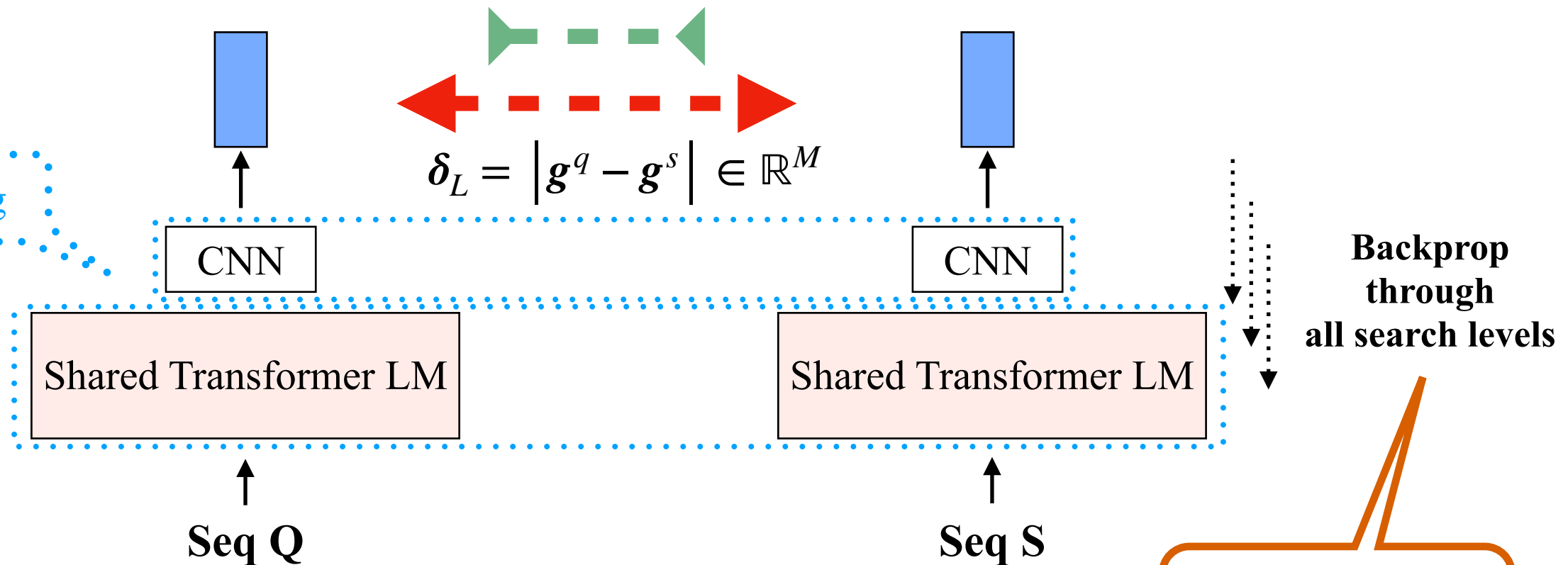
Joint Retrieval and Classification Training

Minimize/maximize difference
to
correct/incorrect matches



$$\delta_L = |g^q - g^s| \in \mathbb{R}^M$$

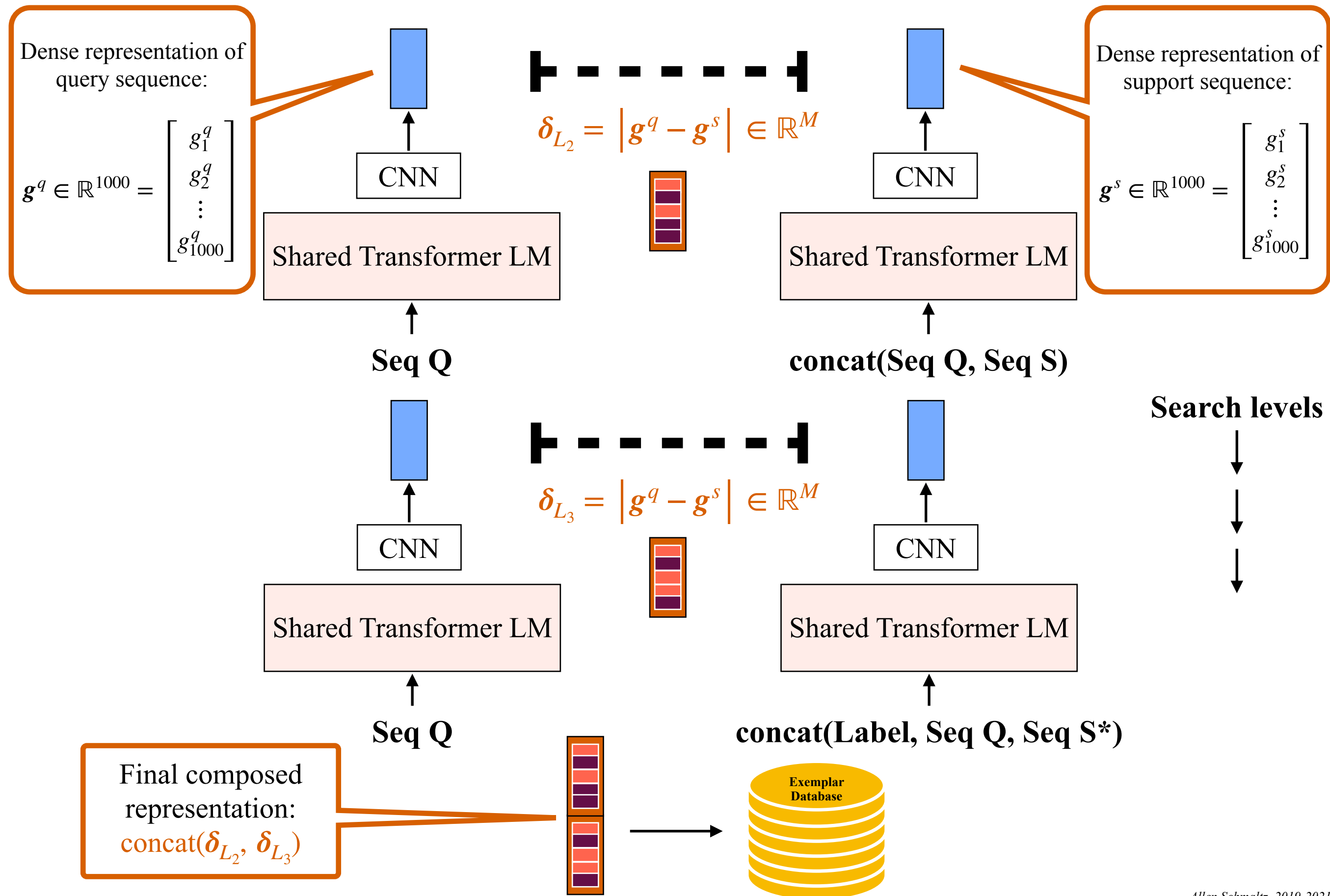
Iterative freezing



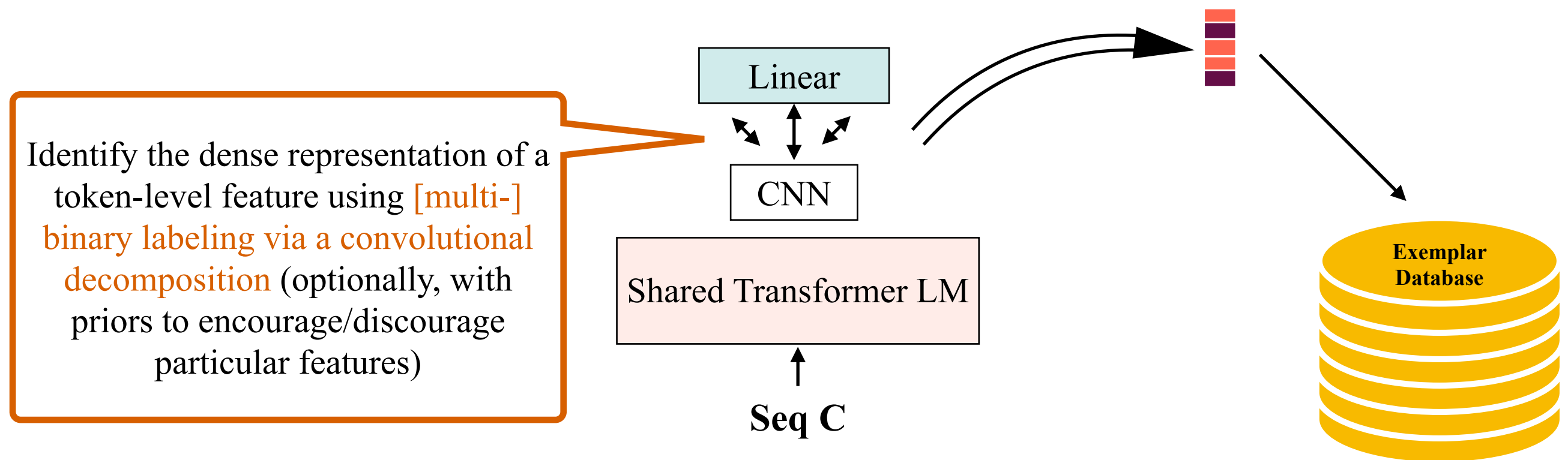
The training set is dynamically created via coarse-to-fine search to find hard negatives, as well as prediction sequences that emulate inference

Yields a single model for both retrieval and classification

Multi-Sequence Representation Composition for Exemplar Auditing



Token-Level Representations for Exemplar Auditing



Prospective Outlook: Interlocking distance constraints across input modalities and tasks via a single, shared model and a dense database...

