

Use soft embeddings from  $S$  to update all parameters

Learn task-dependent  
similarity matrix  $S$  on  
 $D^{\text{B-val}}$

Train Model weights ( $W$ )  
on  $D^{\text{B-train}}$

Pass task dependent knowledge to learn  $S$

Search Phase using Bi-level Optimization

Use soft embeddings from learned  $S$  to update all parameters

Learned  $S$  from  
search phase

Train for  $W^*$   
on  $D^{\text{train}}$

Fine-tuning phase