Large-Scale Object Recognition using Label Relation Graphs

Jia Deng[1,2] Nan Ding[2], Yangqing Jia[2], Andrea Frome[2], Kevin Murphy[2], Samy Bengio[2], Yuan Li[2], Hartmut Neven[2], Hartwig Adam[2]

University of Michigan[1], Google[2]

Object Classification

- Assign semantic labels to objects

Dog ✓
Corgi ✓
Puppy ✓
Cat ✗

Object Classification

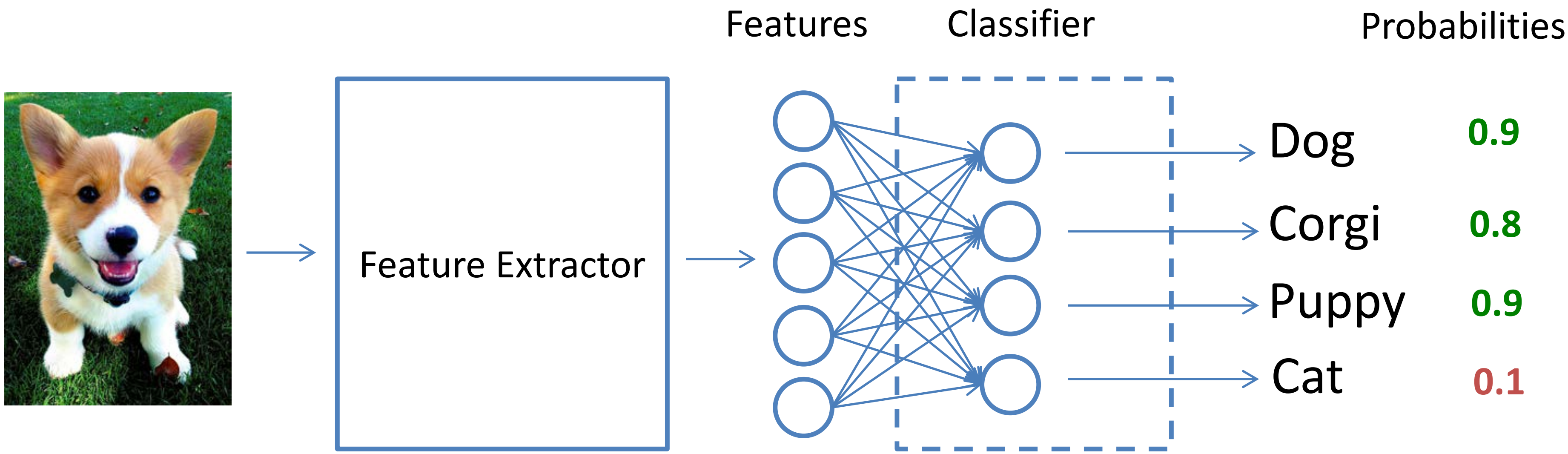
- Assign semantic labels to objects

Probabilities

Dog	0.9
Corgi	0.8
Puppy	0.9
Cat	0.1

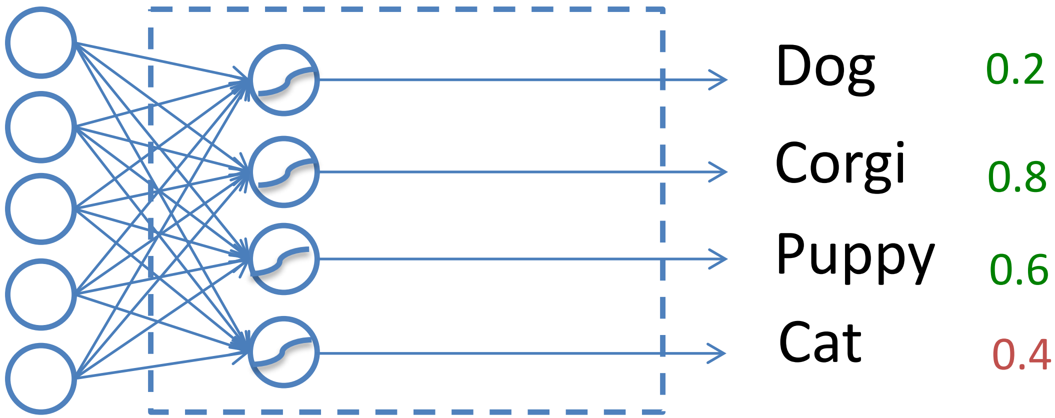
Object Classification

- Assign semantic labels to objects



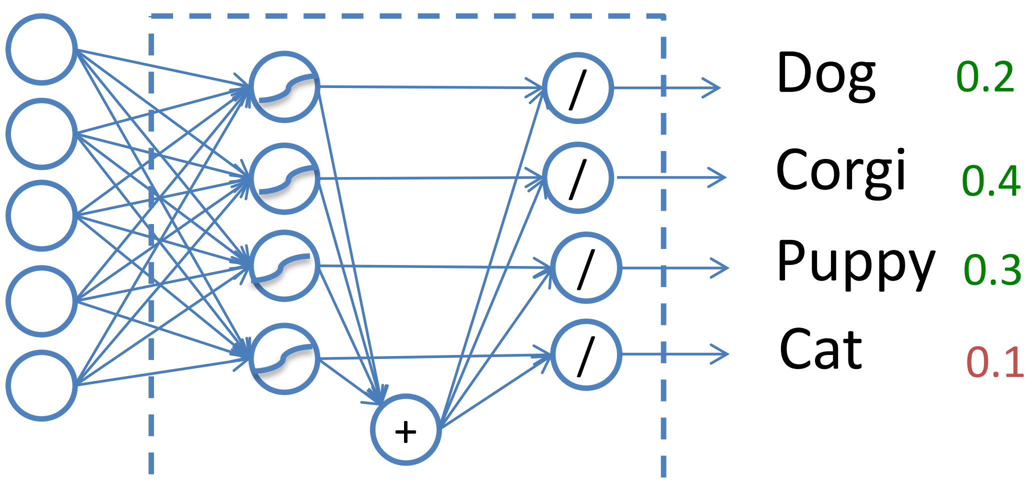
Object Classification

- Independent binary classifiers: Logistic Regression



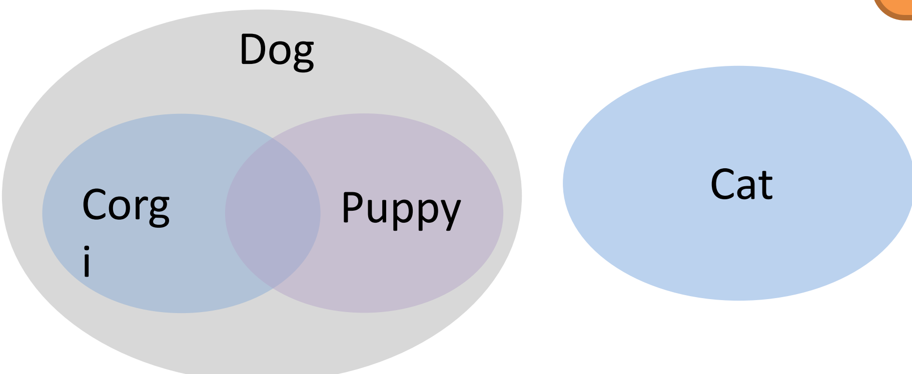
No assumptions about relations.

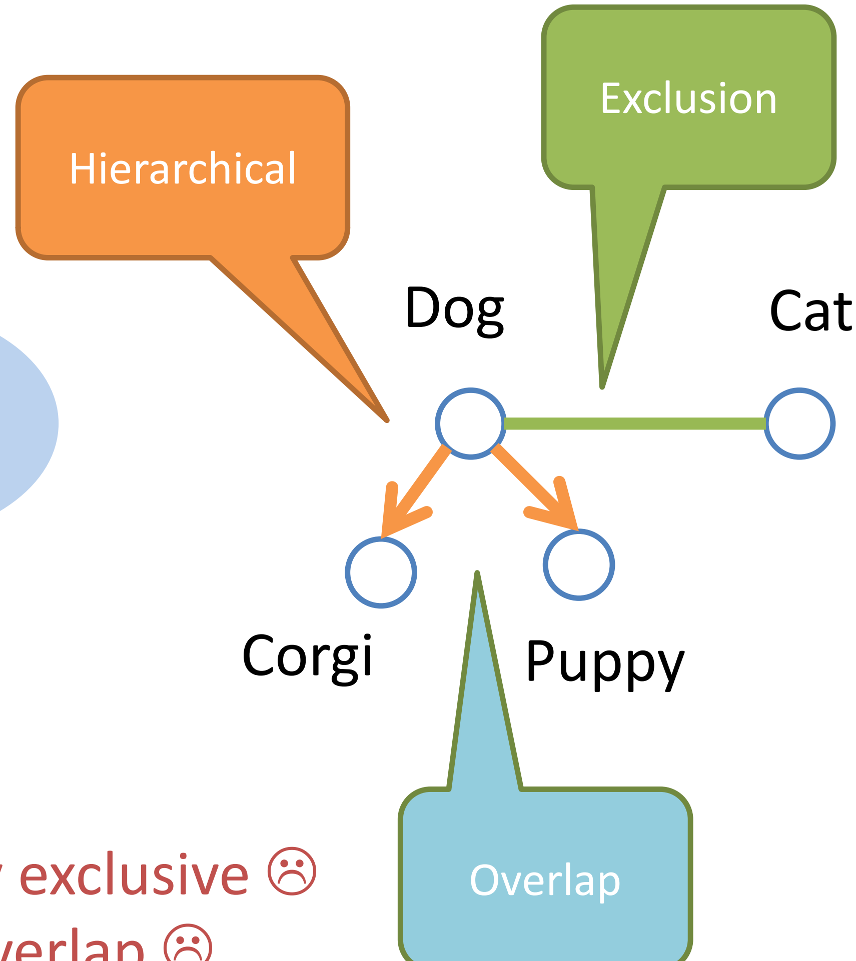
- Multiclass classifier: Softmax



Assumes mutual exclusive labels.

Object labels have rich relations

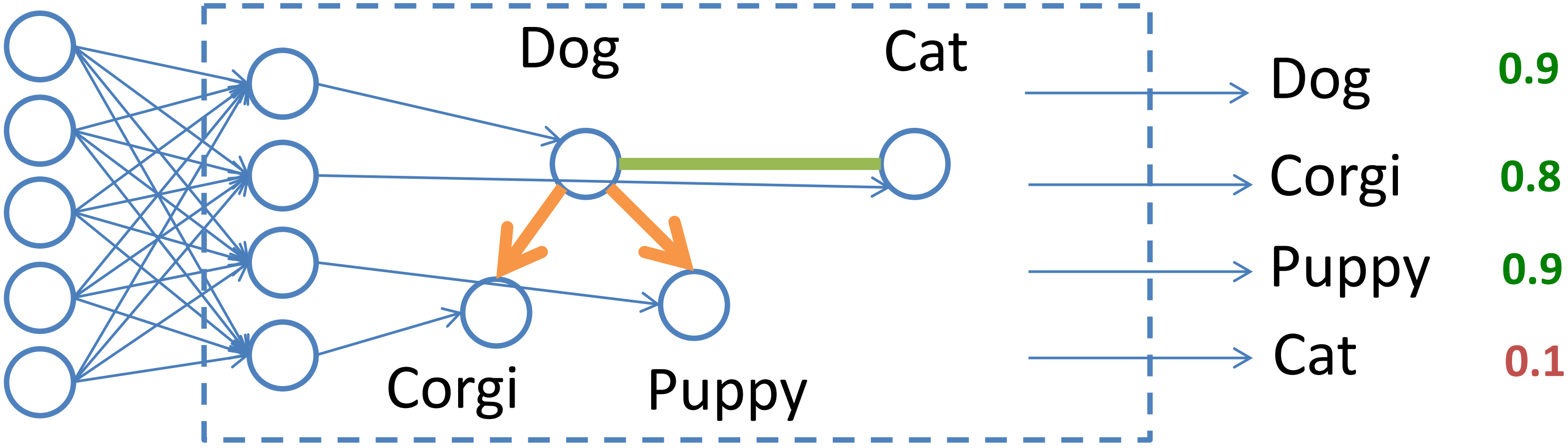




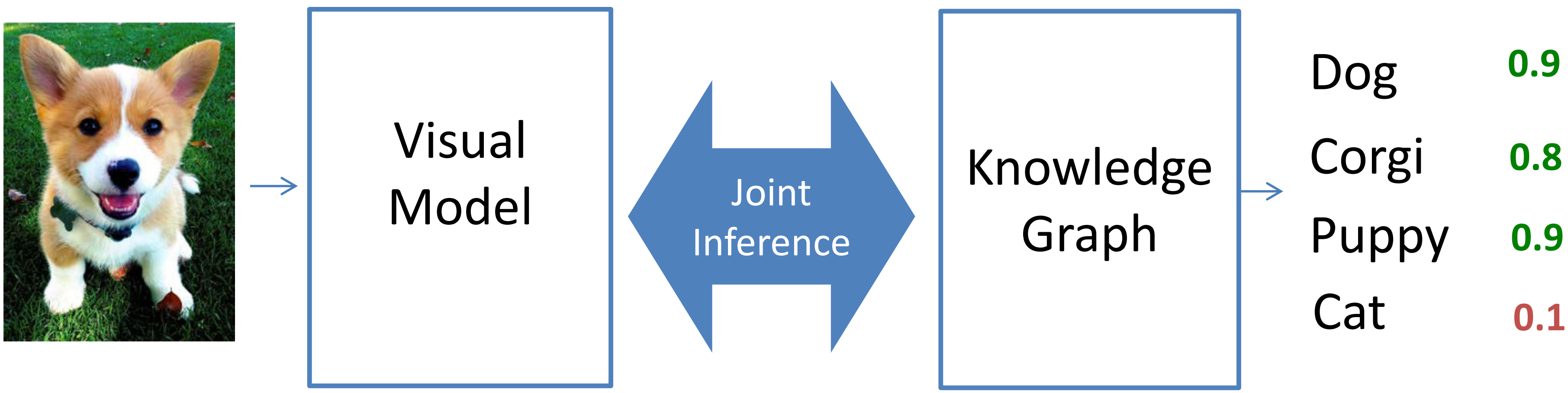
Softmax: all labels are mutually exclusive ☹
Logistic Regression: all labels overlap ☹

Goal: A new classification model

Respects real world label relations



Visual Model + Knowledge Graph



Assumption in this work:
Knowledge graph is given and fixed.

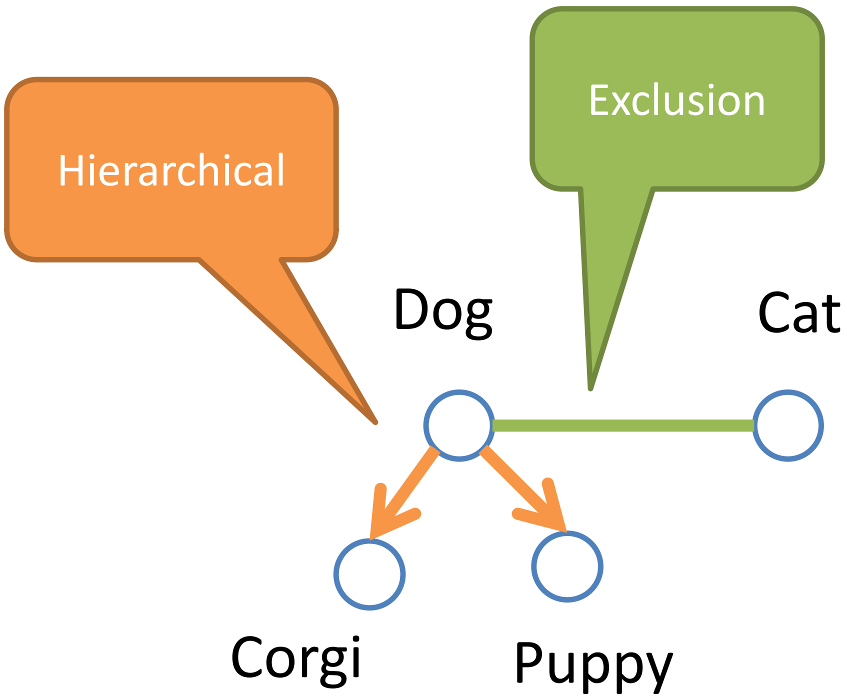
Agenda

- Encoding prior knowledge (HEX graph)
- Classification model
- Efficient Exact Inference
- Experiments
- Conclusion and Future Work

Agenda

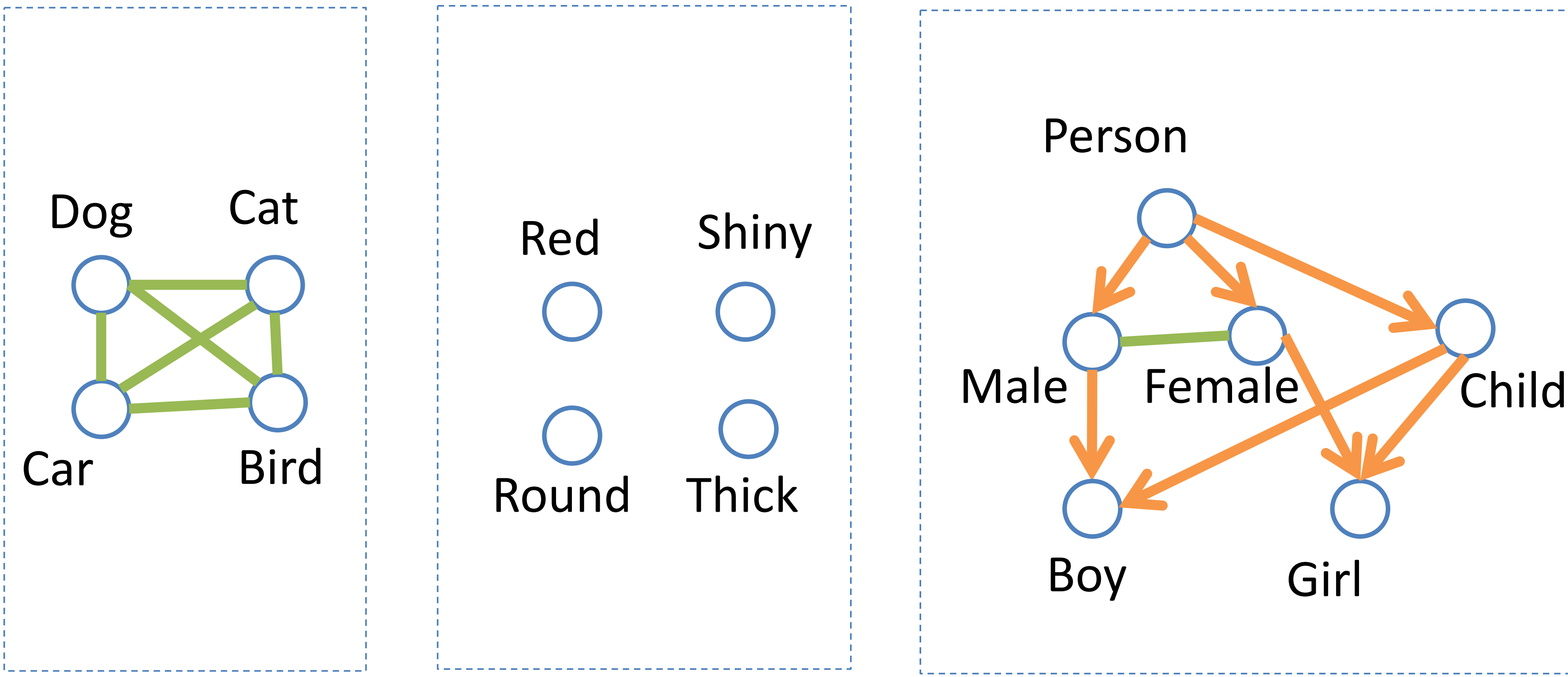
- Encoding prior knowledge (HEX graph)
- Classification model
- Efficient Exact Inference
- Experiments
- Conclusion and Future Work

Hierarchy and Exclusion (HEX) Graph



- Hierarchical edges (directed)
- Exclusion edges (undirected)

Examples of HEX graphs



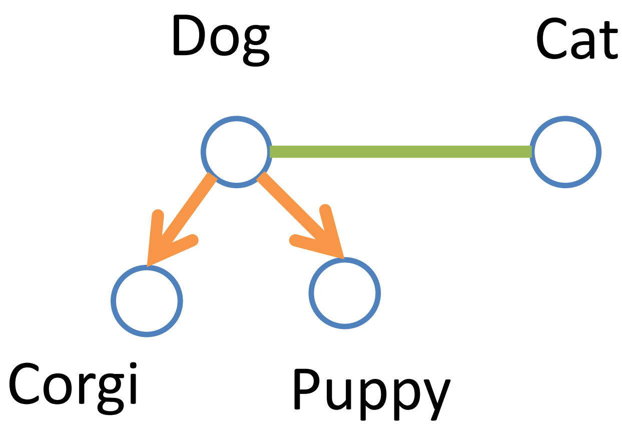
Mutually exclusive

All overlapping

Combination

State Space: Legal label configurations

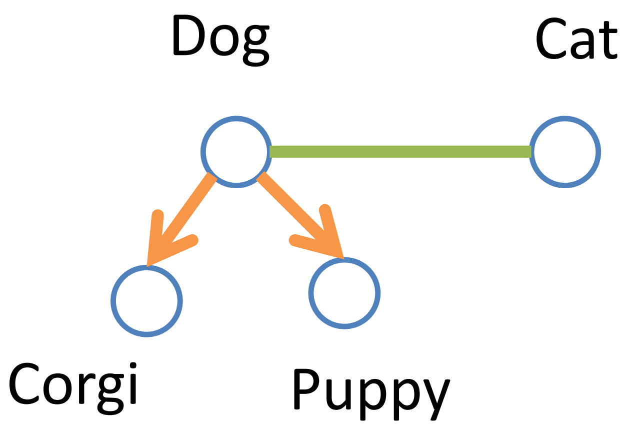
Each edge defines a constraint.



Dog	Cat	Corgi	Puppy
0	0	0	0
0	0	0	1
0	0	1	0
0	0	1	1
1	0	0	0
…			
1	1	0	0
1	1	0	1
…			

State Space: Legal label configurations

Each edge defines a constraint.

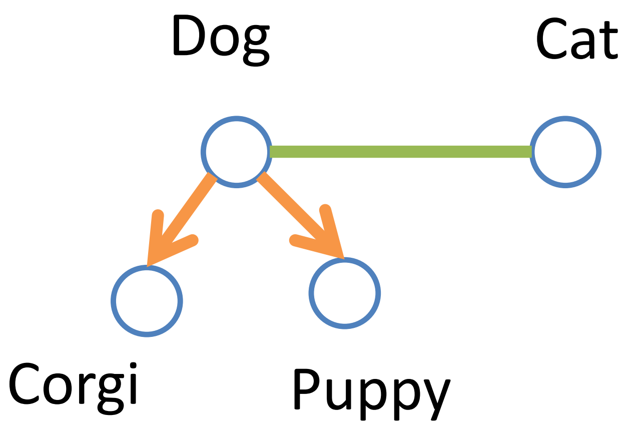


Hierarchy: (dog, corgi) can’t be (0,1)

Dog	Cat	Corgi	Puppy
0	0	0	0
~~0~~	~~0~~	~~0~~	~~1~~
~~0~~	~~0~~	~~1~~	~~0~~
~~0~~	~~0~~	~~1~~	~~1~~
1	0	0	0
…			
1	1	0	0
1	1	0	1
…			

State Space: Legal label configurations

Each edge defines a constraint.



Hierarchy: (dog, corgi) can’t be (0,1)

Exclusion: (dog, cat) can’t be (1,1)

Dog	Cat	Corgi	Puppy
0	0	0	0
~~0~~	~~0~~	~~0~~	~~1~~
~~0~~	~~0~~	~~1~~	~~0~~
~~0~~	~~0~~	~~1~~	~~1~~
1	0	0	0
…			
~~1~~	~~1~~	~~0~~	~~0~~
~~1~~	~~1~~	~~0~~	~~1~~
…			

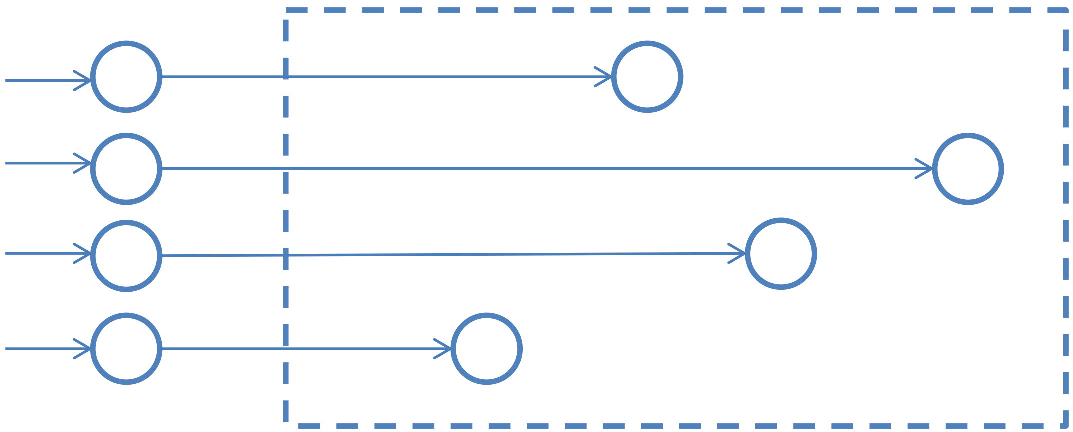
Agenda

- Encoding prior knowledge (HEX graph)
- Classification model
- Efficient Exact Inference
- Experiments
- Conclusion and Future Work

HEX Classification Model

- Pairwise Conditional Random Field (CRF)

$x \in R^n$
Input scores

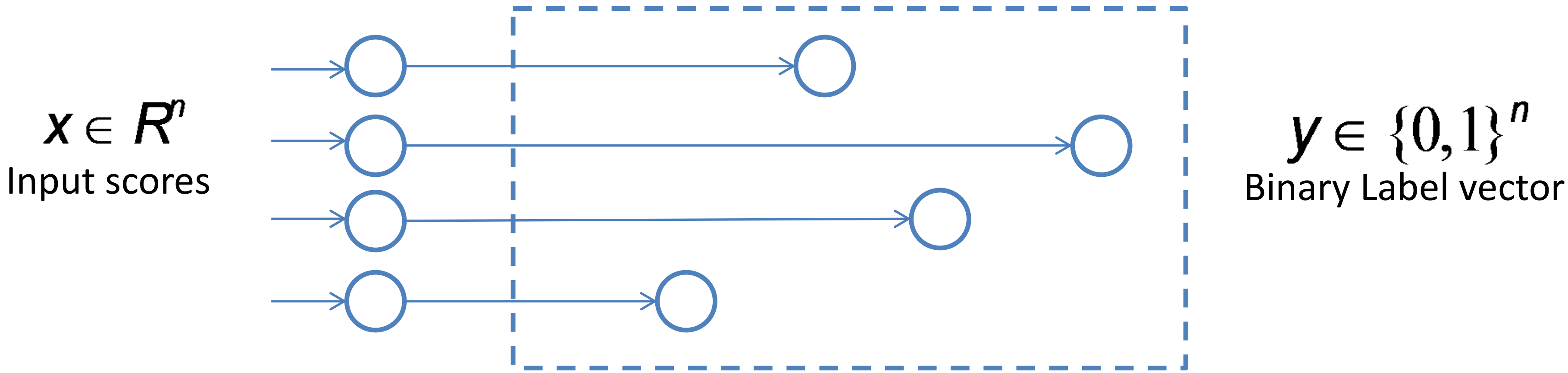


$y \in \{0,1\}^n$
Binary Label vector

$$\Pr(y \mid x) = \frac{1}{Z(x)} \prod_i \phi_i(x_i, y_i) \prod_{i,j} \psi_{i,j}(y_i, y_j)$$

HEX Classification Model

- Pairwise Conditional Random Field (CRF)



$$\Pr(y \mid x) = \frac{1}{Z(x)} \prod_i \phi_i(x_i, y_i) \prod_{i,j} \psi_{i,j}(y_i, y_j)$$

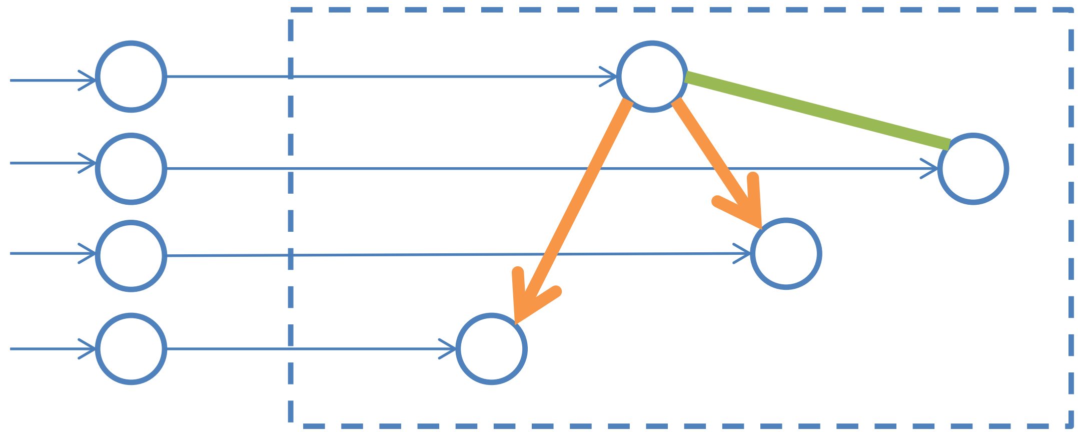
$$\phi_i(x_i, y_i) = \begin{cases} sigmoid(x_i) & if\ y_i = 1 \\ 1 - sigmoid(x_i) & if\ y_i = 0 \end{cases}$$

Unary: same as logistic regression

HEX Classification Model

- Pairwise Conditional Random Field (CRF)

$x \in R^n$
Input scores



$y \in \{0,1\}^n$
Binary Label vector

$$\Pr(y \mid x) = \frac{1}{Z(x)} \prod_i \phi_i(x_i, y_i) \prod_{i,j} \psi_{i,j}(y_i, y_j)$$

$$\phi_i(x_i, y_i) = \begin{cases} sigmoid(x_i) & if\ y_i = 1 \\ 1 - sigmoid(x_i) & if\ y_i = 0 \end{cases}$$

Unary: same as logistic regression

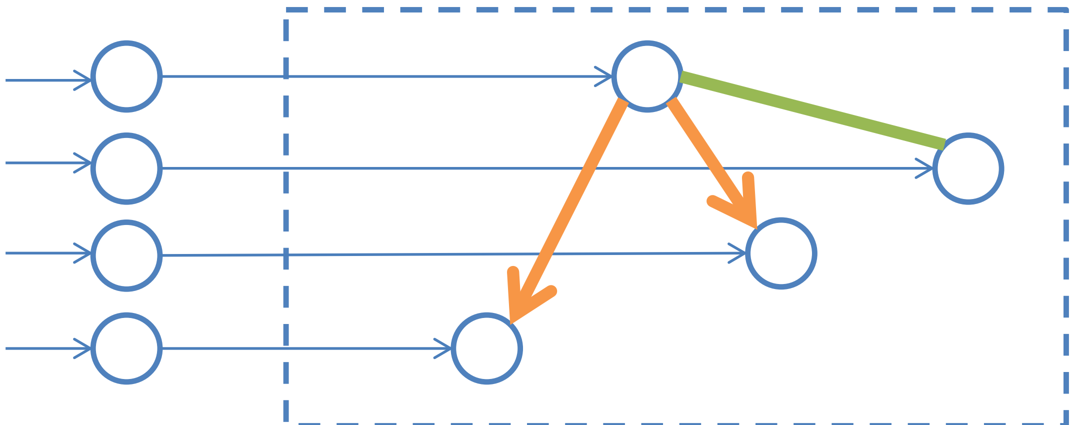
$$\psi_{i,j}(y_i, y_j) = \begin{cases} 0 & \text{If violates constraints} \\ 1 & \text{Otherwise} \end{cases}$$

Pairwise: set illegal configuration to zero

HEX Classification Model

- Pairwise Conditional Random Field (CRF)

$x \in R^n$
Input scores



$y \in \{0,1\}^n$
Binary Label vector

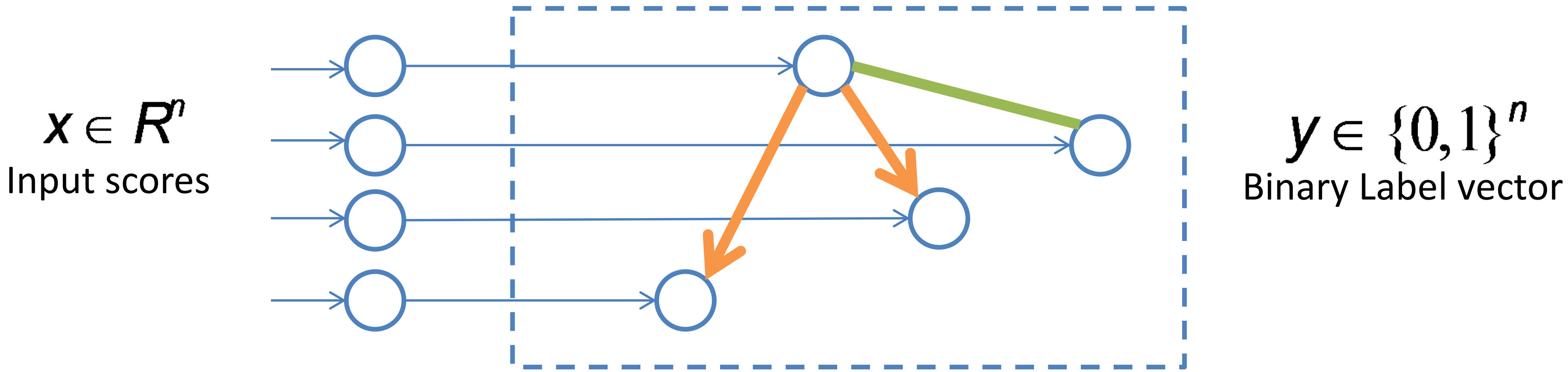
$$\Pr(y \mid x) = \frac{1}{Z(x)} \prod_{i} \phi_i(x_i, y_i) \prod_{i,j} \psi_{i,j}(y_i, y_j)$$

$$Z(x) = \sum_{\bar{y} \in \{0,1\}^n} \prod_{i} \phi_i(x_i, \bar{y}_i) \prod_{i,j} \psi_{i,j}(\bar{y}_i, \bar{y}_j)$$

Partition function: Sum over all (legal) configurations

HEX Classification Model

- Pairwise Conditional Random Field (CRF)



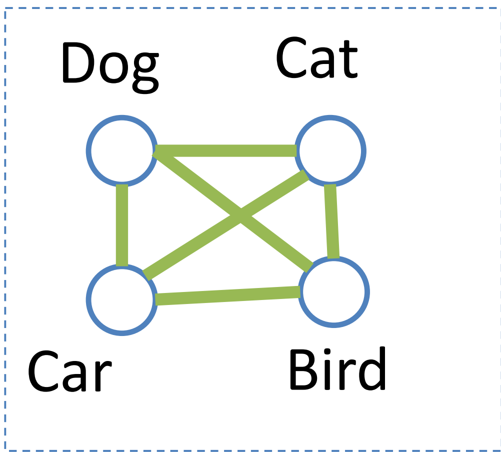
$$\Pr(y \mid x) = \frac{1}{Z(x)} \prod_i \phi_i(x_i, y_i) \prod_{i,j} \psi_{i,j}(y_i, y_j)$$

Probability of a single label: marginalize all other labels.

$$\Pr(y_i = 1 \mid x) = \frac{1}{Z(x)} \sum_{\bar{y}:\bar{y}_i=1} \prod_i \phi_i(x_i, \bar{y}_i) \prod_{i,j} \psi_{i,j}(\bar{y}_i, \bar{y}_j)$$

Special Case of HEX Model

- Softmax



Mutually exclusive

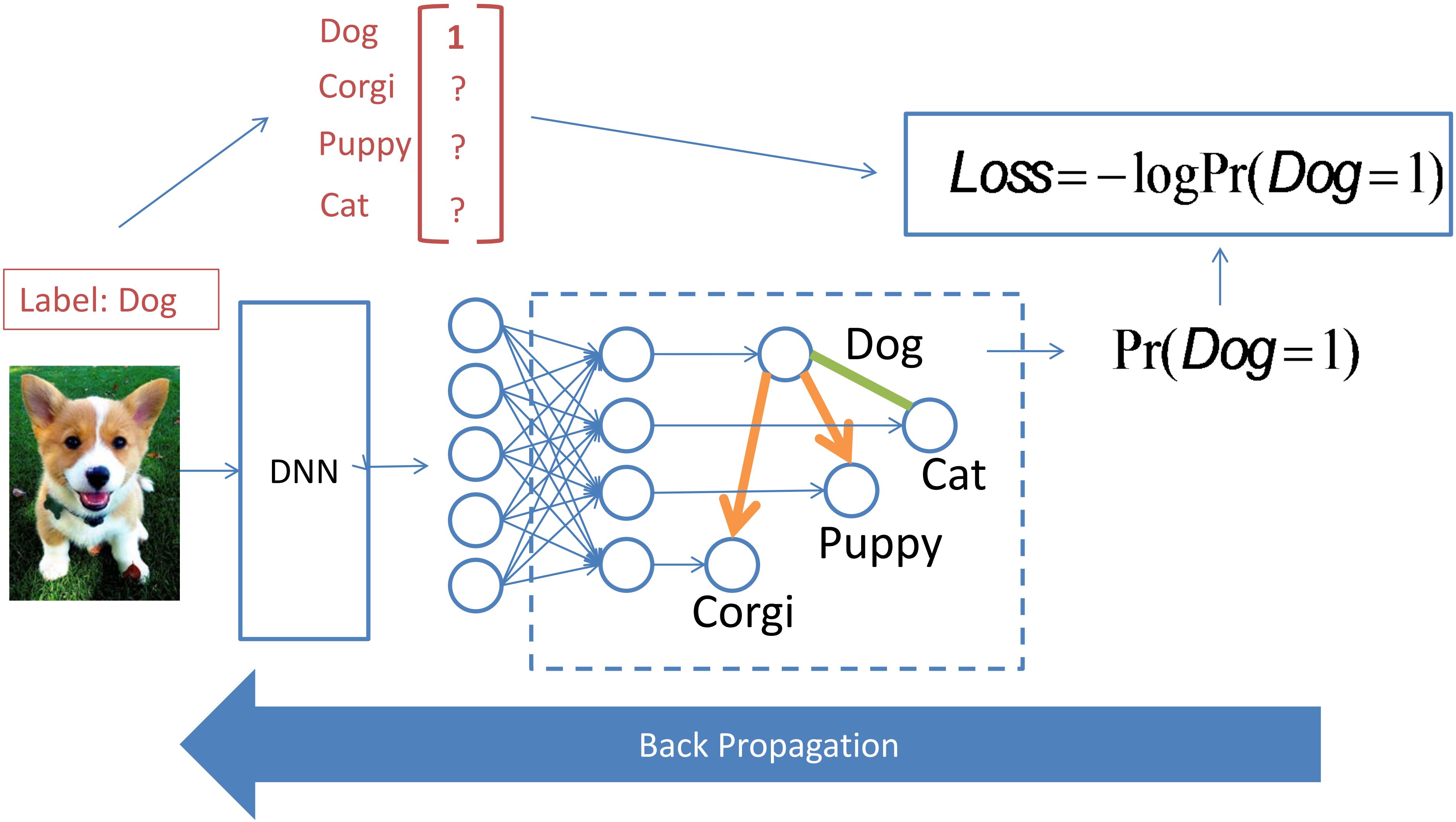
$$\Pr(y_i = 1 \mid x) = \frac{\exp(x_i)}{1 + \sum_j \exp(x_j)}$$

- Logistic Regressions

All overlapping

$$\Pr(y_i = 1 \mid x) = \frac{1}{1 + \exp(-x_i)}$$

Learning



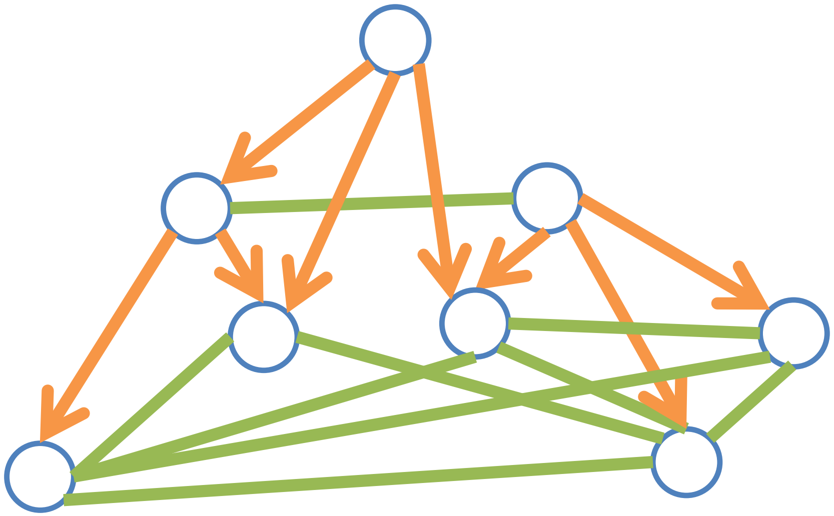
Maximize marginal probability of observed labels

Agenda

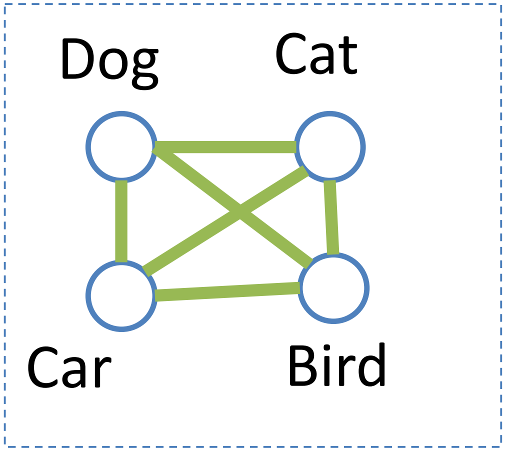
- Encoding prior knowledge (HEX graph)
- Classification model
- Efficient Exact Inference
- Experiments
- Conclusion

Naïve Exact Inference is Intractable

- Inference:
  - Computing partition function
  - Perform marginalization
- HEX-CRF can be densely connected (large treewidth)



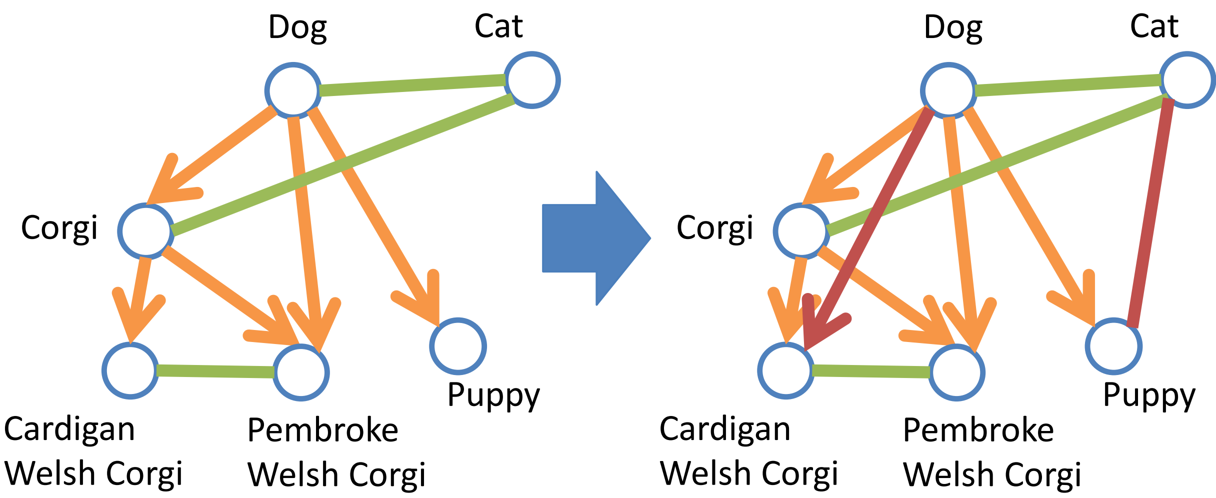
Observation 1: Exclusions are good



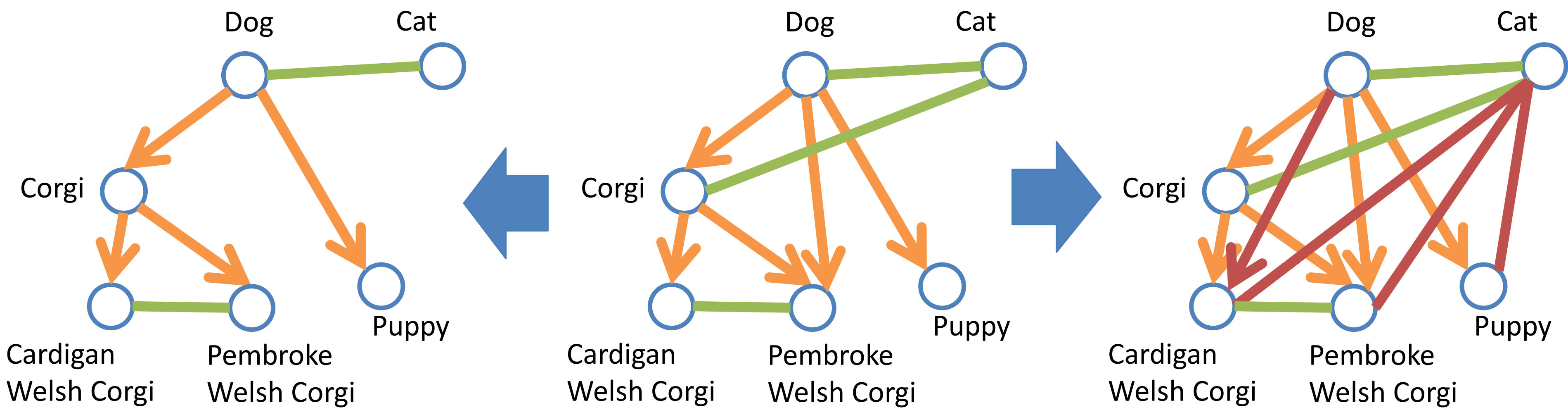
Number of legal states is $O(n)$, not $O(2^n)$.

- Lots of exclusions → Small state space → Efficient inference
- Realistic graphs have lots of exclusions.
- Rigorous analysis in paper.

Observation 2: Equivalent graphs



Observation 2: Equivalent graphs



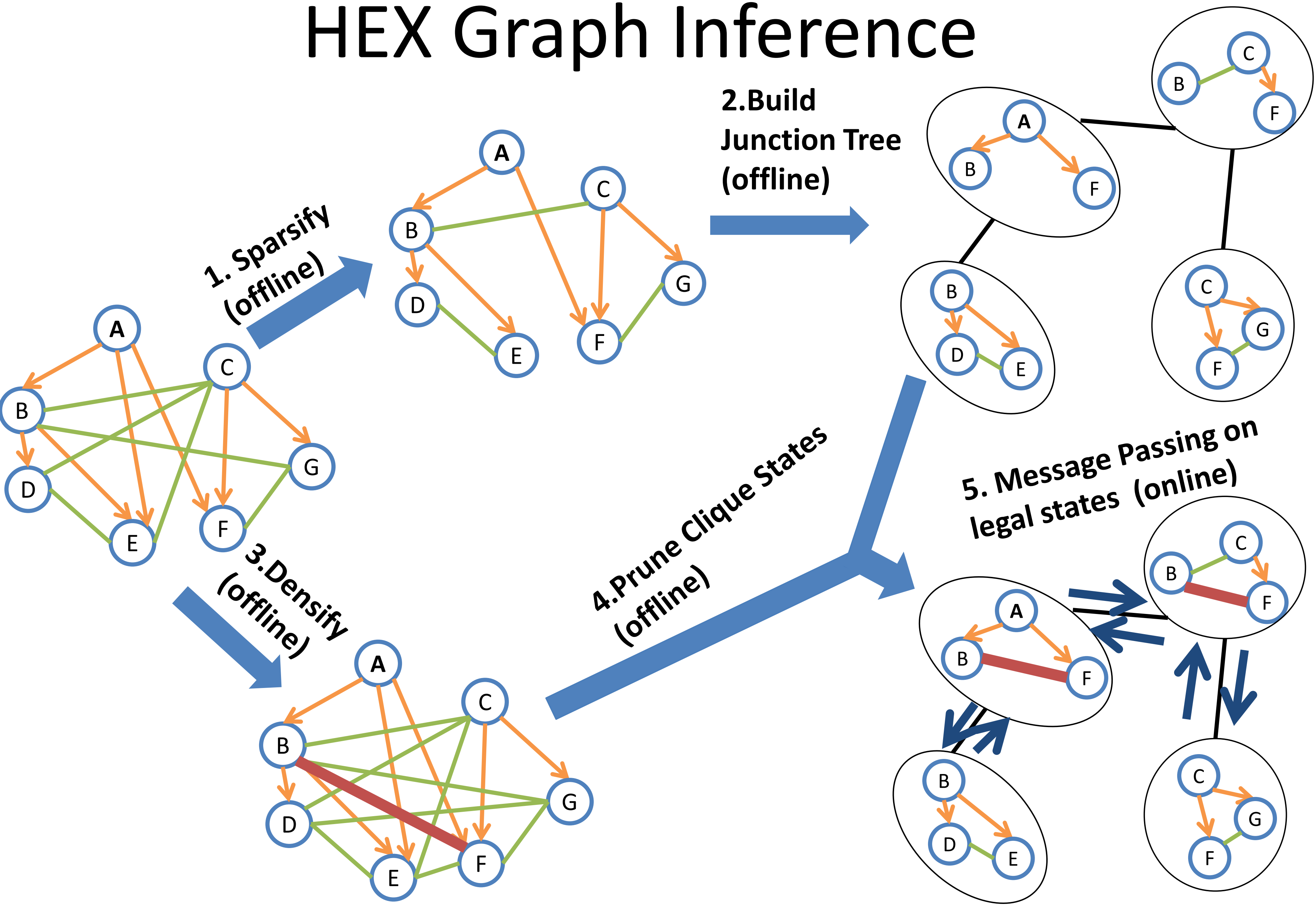
Sparse equivalent

- Small Treewidth ☺
- Dynamic programming

Dense equivalent

- Prune states ☺
- Can brute force

HEX Graph Inference

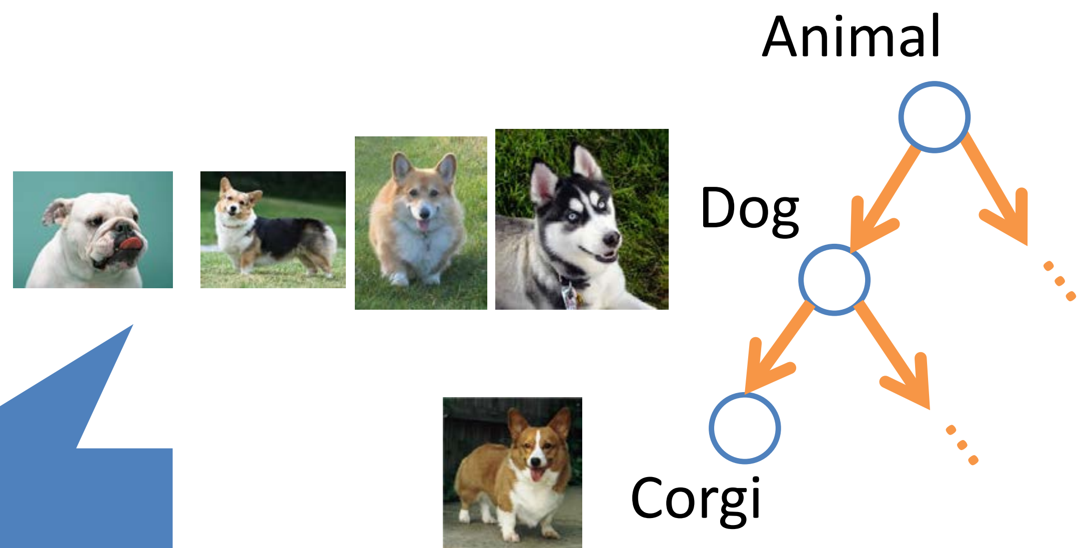


Agenda

- Encoding prior knowledge (HEX graph)
- Classification model
- Efficient Exact Inference
- Experiments
- Conclusion and Future Work

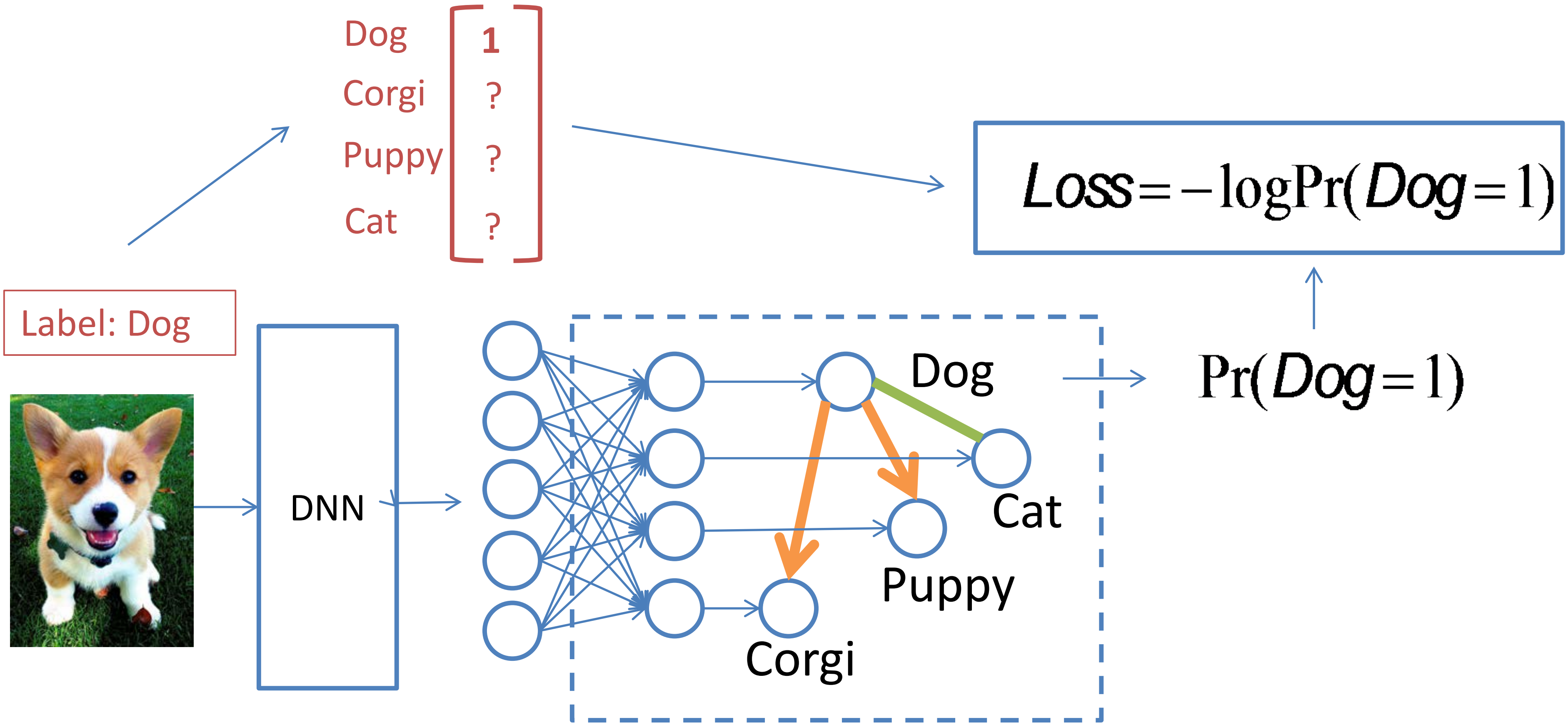
Exp 1: Learning with weak labels

- Many basic category labels
- Few fine-grained labels



Weak labels:
No information on
subcategories.

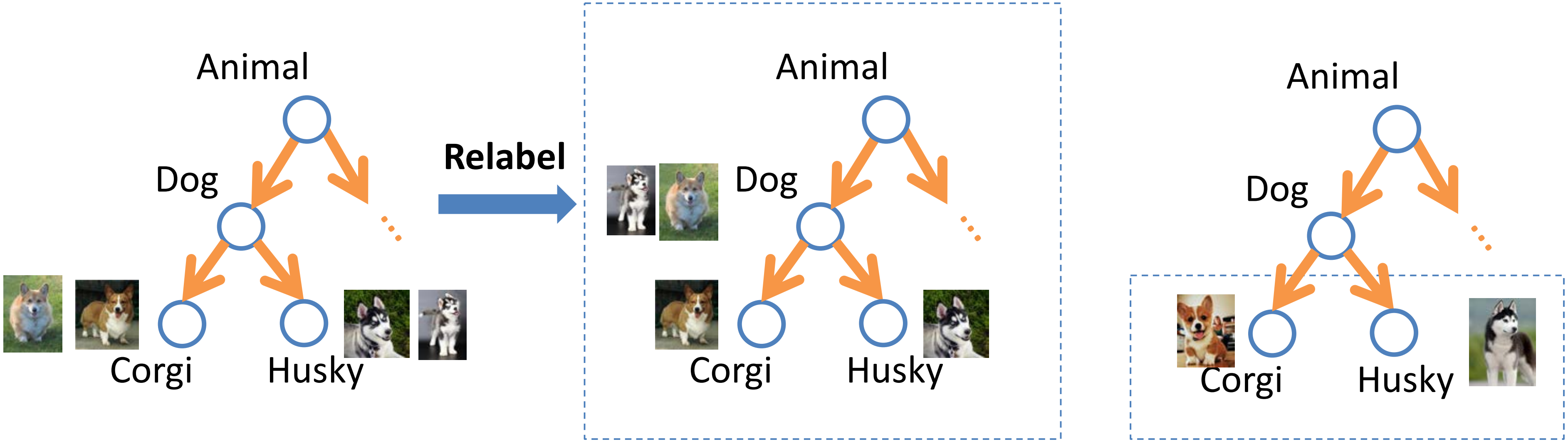
Exp 1: Learning with weak labels



Hypothesis: HEX models can improve fine-grained recognition using basic level labels.

Exp 1: Learning with weak labels

- ILSVRC 2012: "relabel" or "weaken" a portion of fine-grained leaf labels to basic level labels.
- Evaluate on fine-grained recognition



Original ILSVRC 2012 (leaf labels)

Training ("weakened" labels)

Test

Exp 1: Learning with weak labels

- ILSVRC 2012: "relabel" or "weaken" a portion of fine-grained leaf labels to basic level labels.
- Evaluate on fine-grained recognition.
- **Consistently outperforms baselines.**

relabeling	softmax-leaf	softmax-all	logistic	**ours**
50%	50.5(74.7)	56.4(79.6)	21.0(45.2)	**58.2(80.8)**
90%	26.2(47.3)	52.9(77.2)	9.3(27.2)	**55.3(79.4)**
95%	16.0(32.2)	50.8(76.0)	5.6(17.2)	**52.4(77.2)**
99%	2.5 (7.2)	**41.5**(68.1)	1.0(3.8)	**41.5(68.5)**

Top 1 accuracy (top 5 accuracy)

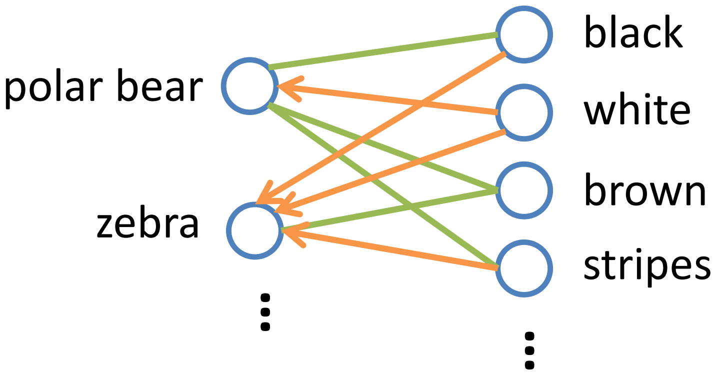
Exp 2: Zero-Shot Recognition using Object-Attribute Knowledge

polar bear

black: no
white: yes
brown: no
stripes: no

zebra

black: yes
white: yes
brown: no
stripes: yes



- Animals with Attribute (AwA) dataset (Lampert et al. 2009)
- Training:
  - Observe only a subset of animal labels.
  - Given all animal-attribute relations
  - Indirectly learns attributes.
- Test: predict new classes with no images in training.

DAP (Lampert et al.)	IAP (Lampert et al.)	Ours
40.5%	27.8%	38.5%

Related Work

- **Multilabel Annotation & Hierarchy**

  [Lampert et al. NIPS'11] [Hwang et al. CVPR'11]
  [Chen et al. ICCV'11] [Kang et al. CVPR'06]
  [Bi & Kwok, NIPS'12] [Marszalek & Schmid CVPR'07]
  [Bucak et al. CVPR'11] [Zweig & Weinshall CVPR'07]

  Ours: Unifies hierarchy and exclusion.

- **Transfer learning & Attributes**

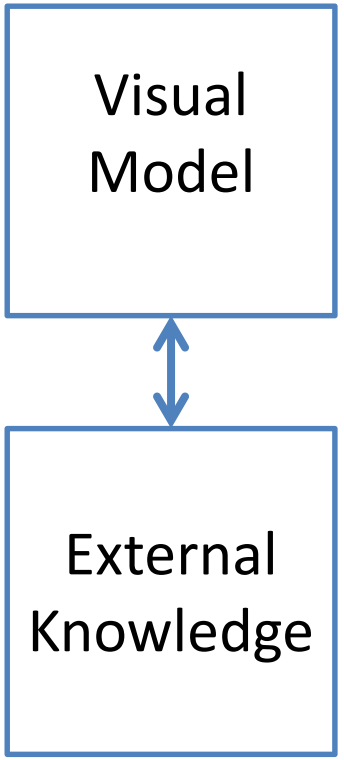
  [Rohrbach et al. CVPR'10] [Farhadi et al. CVPR'10]
  [Lampert et al. CVPR'09] [Lim et al. NIPS'11]
  [Kuettel et al. ECCV'12] [Yu et al. CVPR'13]
  [Akata et al. CVPR'13] [Fergus et al. ECCV'10]

  Ours: A classification model that allows transferring.

- **Extracting Common Sense Knowledge**

  [Chen et al. ICCV'13] [Zhu et al. ECCV'14]
  [Zitnick & Parikh CVPR'13] [Fouhey & Zitnick CVPR'14]

  Ours: Assumes knowledge is given.



Conclusions

- A unified framework for single object classification
  - Generalizes standard classification models
  - Leverages a knowledge graph
  - Efficient exact inference
- Future work
  - Non-absolute relations
  - Spatial relations between object instances