Hedging Your Bets: Optimizing Accuracy-Specificity Trade-offs in Large Scale Visual Recognition

Jia Deng[1,2], Jonathan Krause[1], Alexander C. Berg[3], Li Fei-Fei[1]
Stanford University[1], Princeton University[2], Stony Brook University[3]

Proc. of CVPR 2012

Presenter: Lamberto Ballan

MICC
Media Integration and Communication Center

Multimedia and Vision Reading Group

What to recognize?

http://www.flickr.com/photos/sgcallawayimages/3306849049/

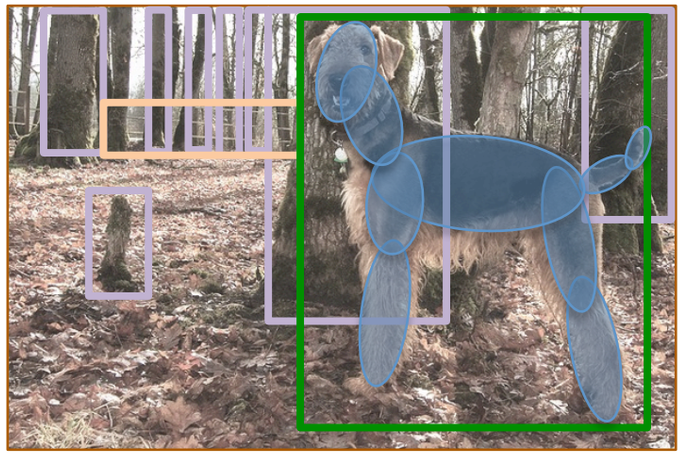
Increasing structural complexity

Dog

Single label

Dog, Tree, Fence, Leaf

Multiple labels



Localization

“Happy shaggy Airedale poses in the autumn forest.”

Description

Where are we? (for large-scale recognition)

http://www.flickr.com/photos/sgcallawayimages/3306849049/

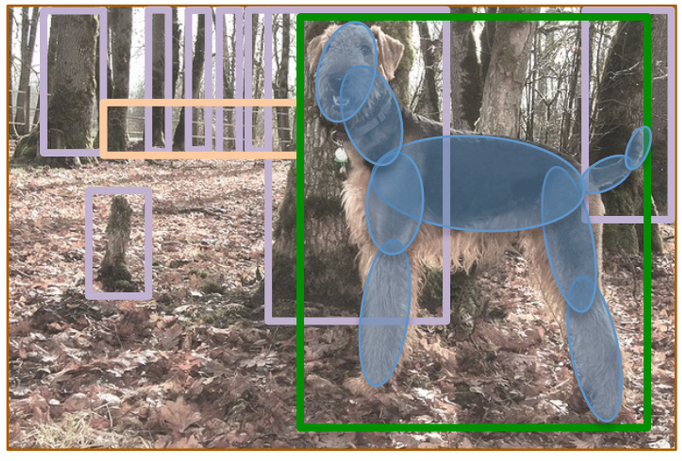
Increasing structural complexity

Dog

Single label

Dog, Tree, Fence, Leaf

Multiple labels



Localization

"Happy shaggy Airedale poses in the autumn forest."

Description

Label space?

~20,000 categories (noun synsets) from WordNet

http://www.flickr.com/photos/sgcallawayimages/3306849049/

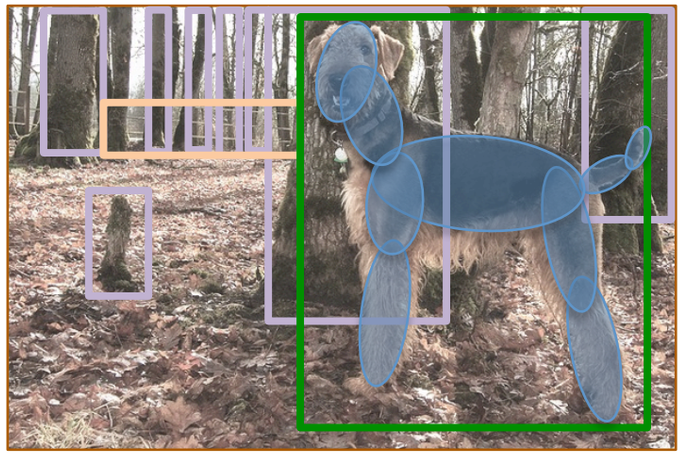
Increasing structural complexity

Dog

Single label

Dog, Tree, Fence, Leaf

Multiple labels

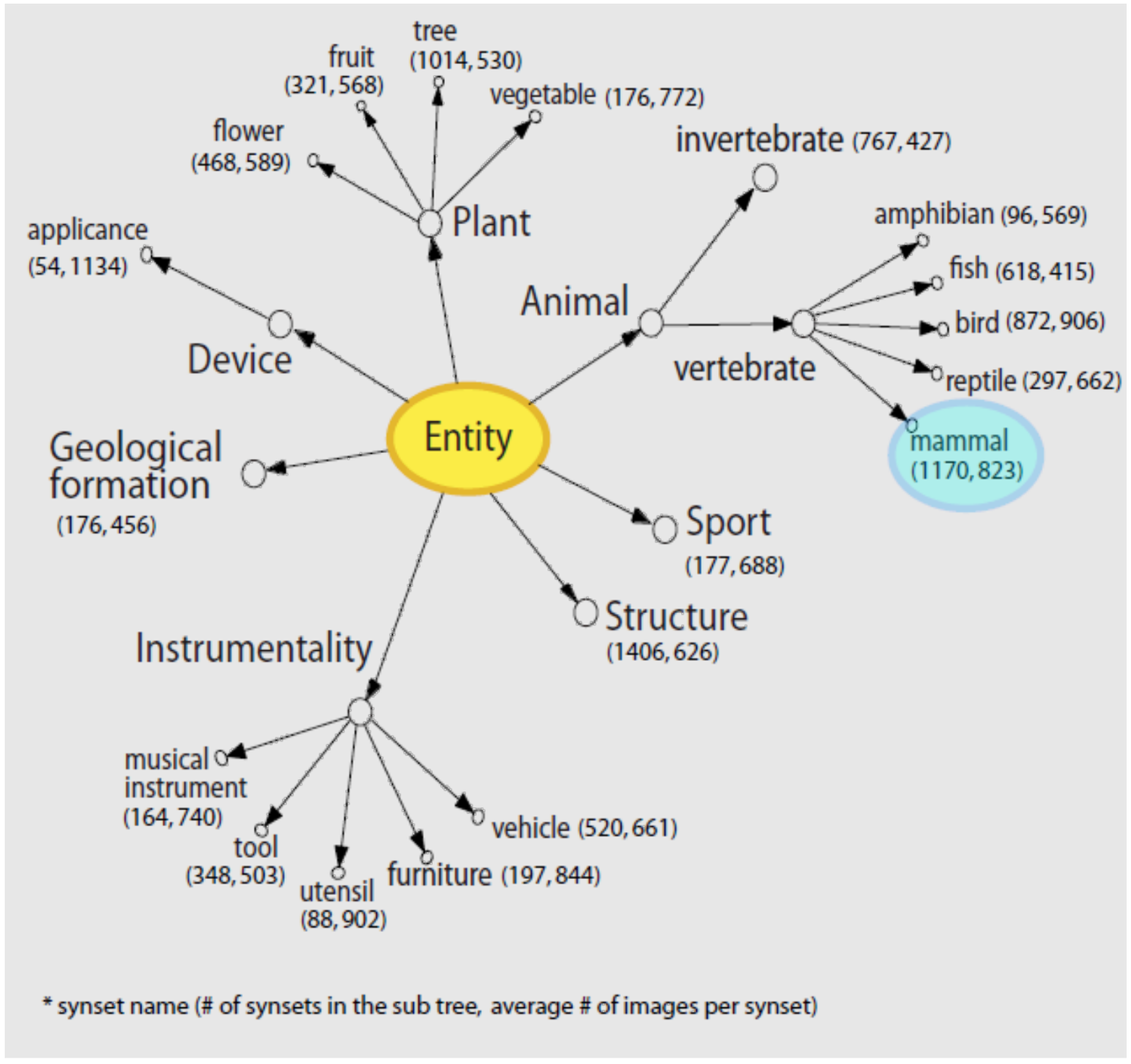


Localization

"Happy shaggy Airedale poses in the autumn forest."

Description

WordNet / ImageNet



IMAGENET

14,197,122 images
21,841 synsets indexed

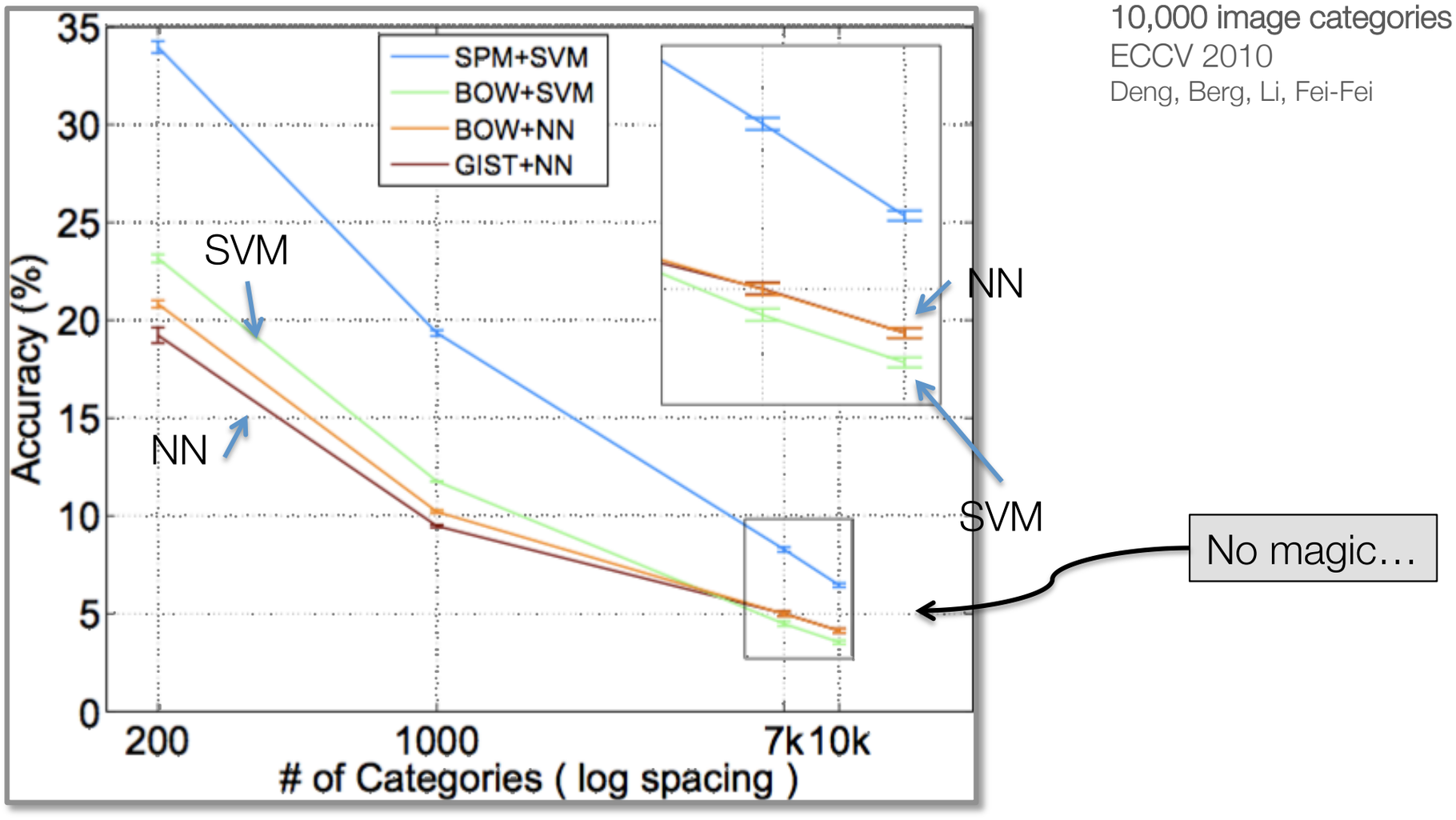
ILSVRC-2013

Task 1: Detection
PASCAL-style detection challenge; 200 categories

Task 2: Classification
Image classification challenge; 1000 categories

Task 2: Classification with localization
Image classification plus object localization challenge; 1000 categories

Why do we want a large label space?



What does classifying more than 10,000 image categories tell us?
ECCV 2010
Deng, Berg, Li, Fei-Fei

- Performances are often weak ... (~10%)
- ... but large-scale allows us to use structure over labels!

What do we mean by similarity?

Query

Airship

Blimp

Retrieved Images




Aquatic Animal

Axlotl

Retrieved Images



Airship

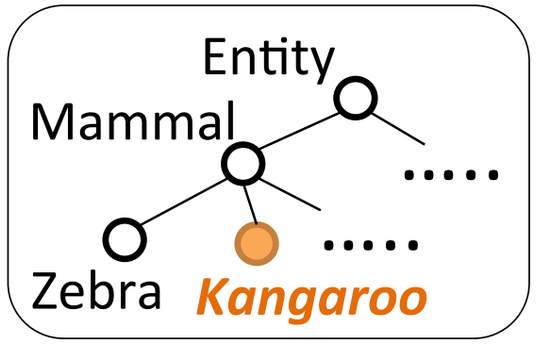
Baloon

How can we do this?

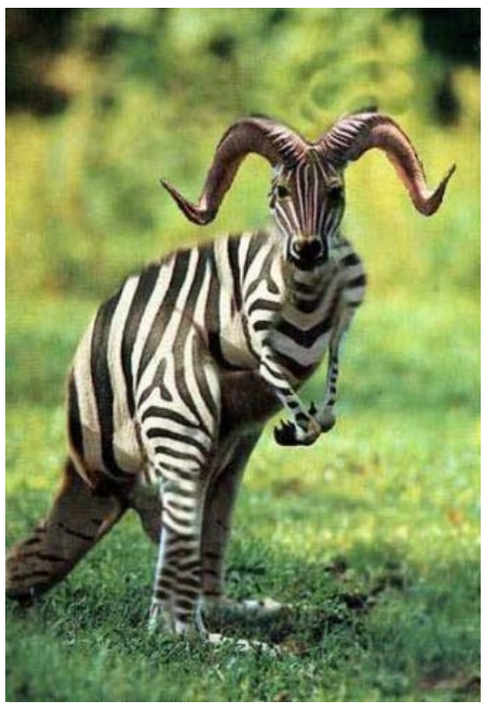
traditional
flat classifier

Kangaroo ✓

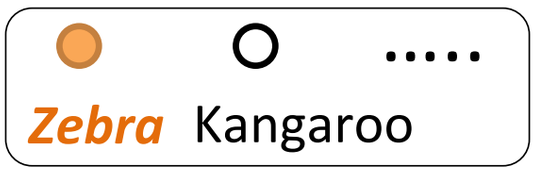
our
proposal



Kangaroo ✓

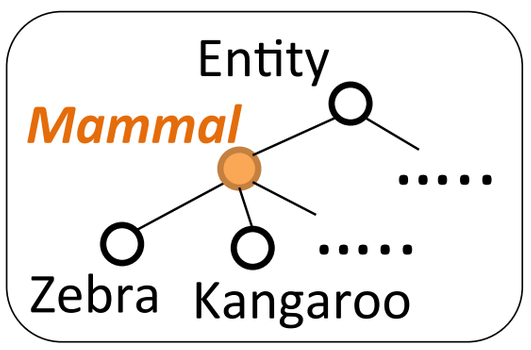


traditional
flat classifier



Zebra ✗

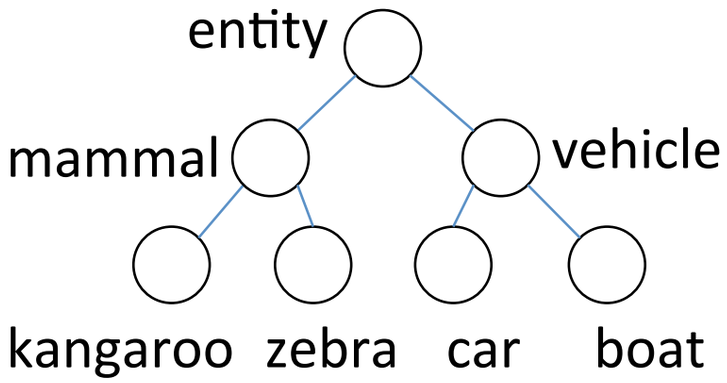
our
proposal



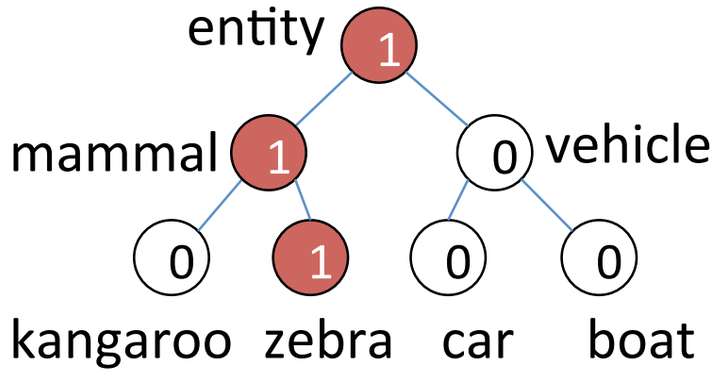
Mammal ✓

Formulation

Semantic hierarchy

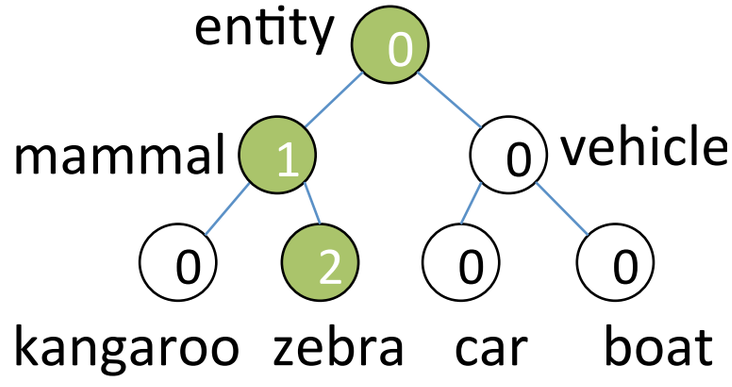


Accuracy



ground truth

Reward

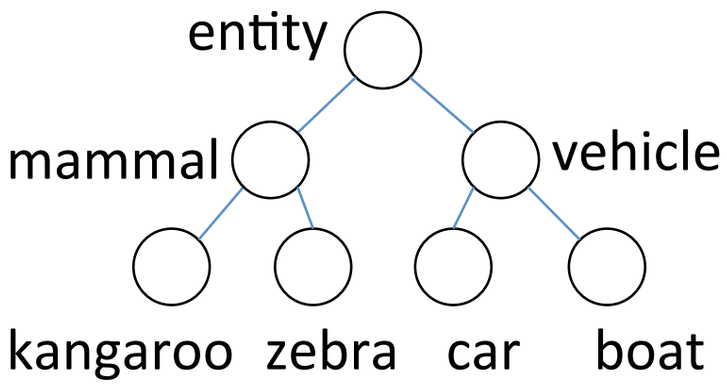


ground truth

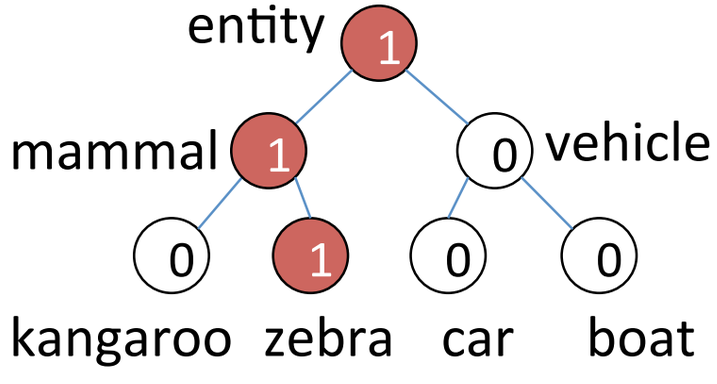
Key idea: automatically select the appropriate level of abstraction to *optimize* the accuracy/specificity trade-off

Formulation

Semantic hierarchy

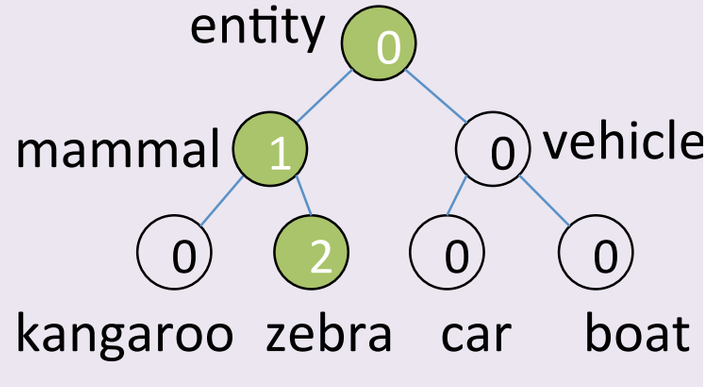


Accuracy



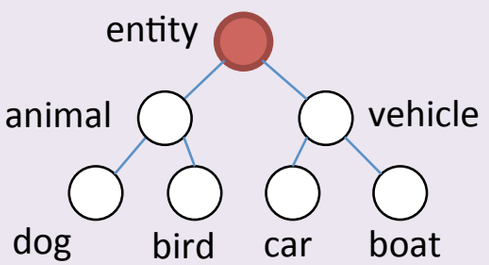
ground truth

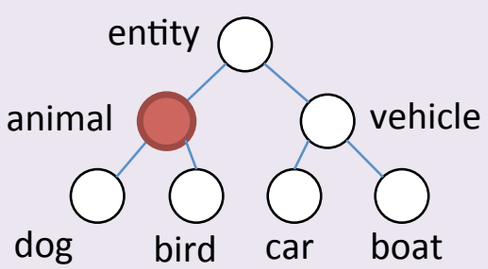
Reward

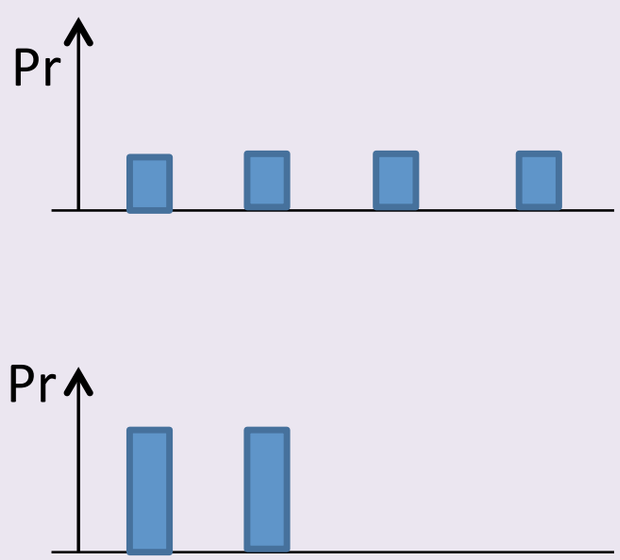


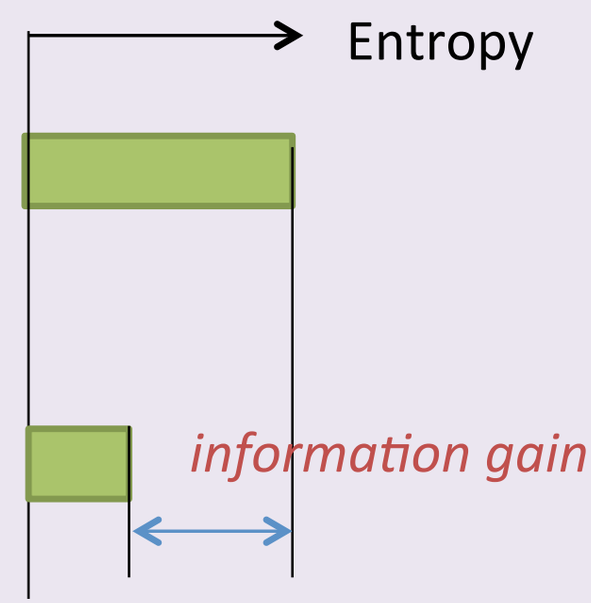
ground truth

Reward: amount of correct *information gain (i.e. decrease of uncertainty)*



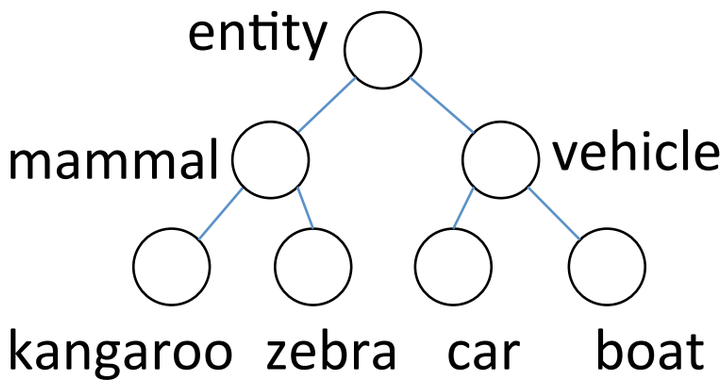




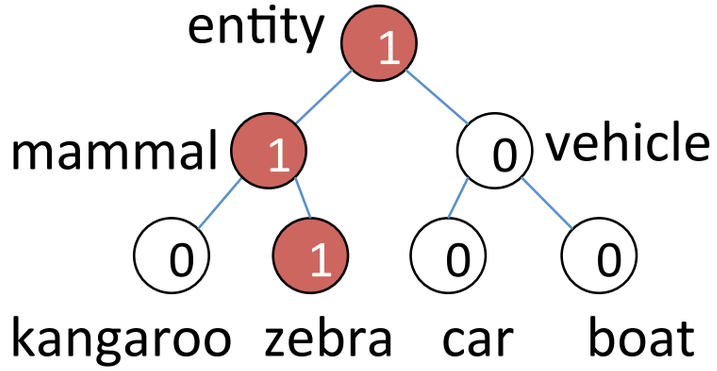


Formulation

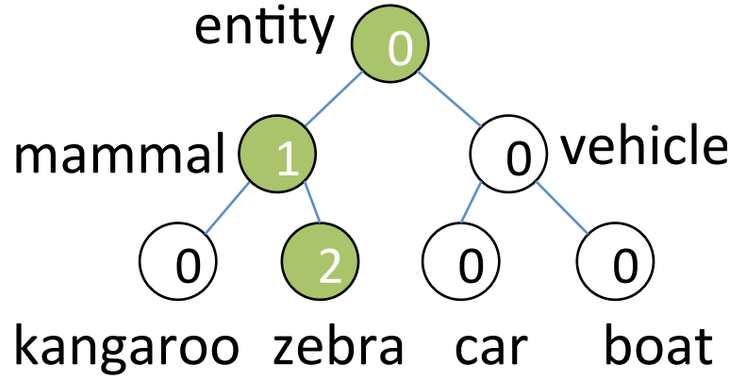
Semantic hierarchy



Accuracy



Reward



Goal: maximize the reward given an arbitrary accuracy guarantee

Training images

Reward of classifier

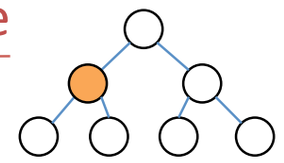
Rewards on nodes

$$\text{Maximize over } f, \quad R(f,r)$$
$$\text{Subject to} \quad \Phi(f) \geq 1-\epsilon$$

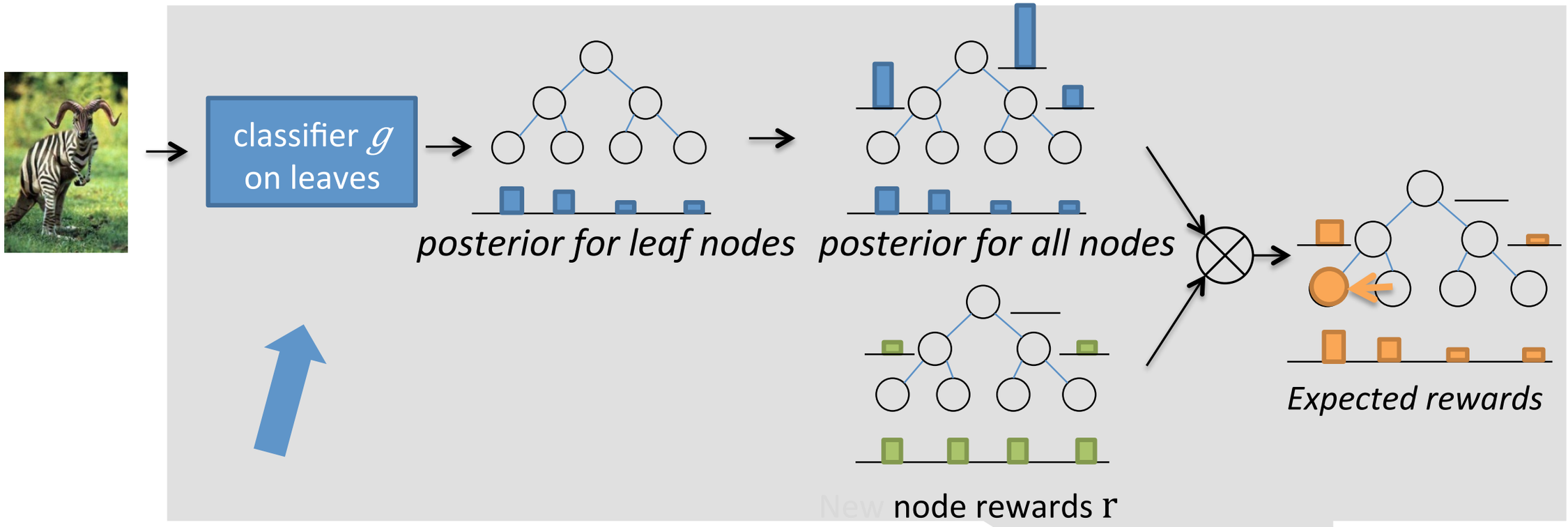
Accuracy of classifier

Accuracy guarantee

classifier f



For now, assume no accuracy guarantee ...



Training images

Reward of classifier

Rewards on nodes

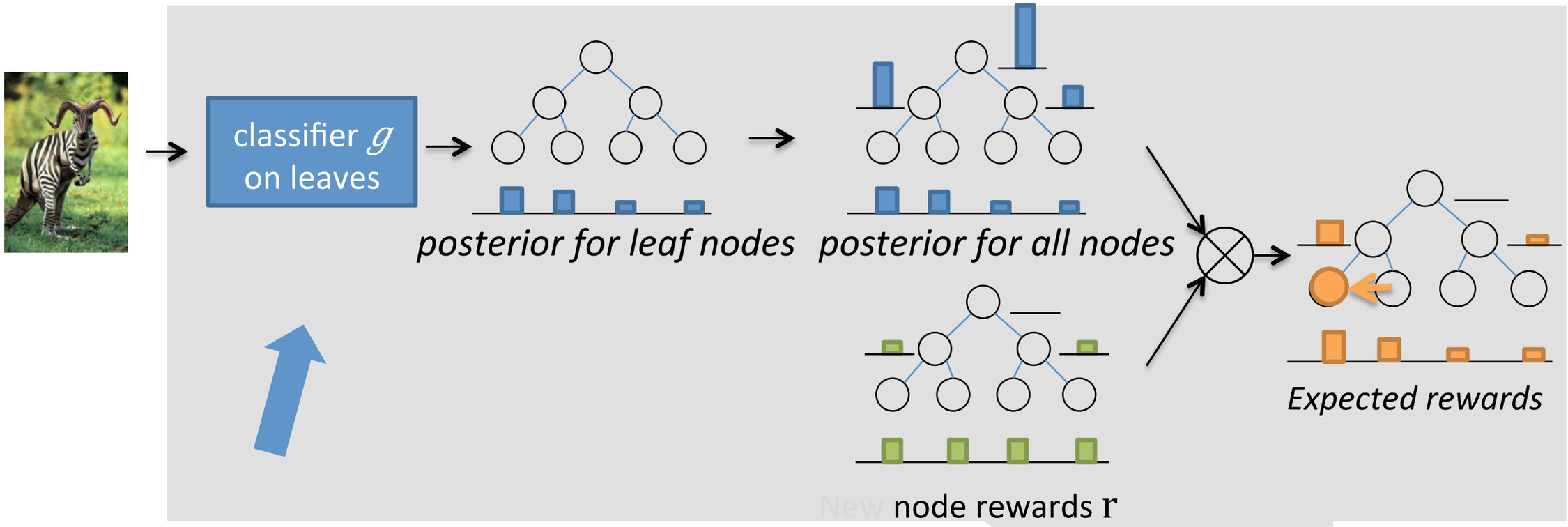
$$\text{Maximize over } f, \quad R(f,r)$$
$$\text{Subject to} \quad \Phi(f) \geq 1-\epsilon$$

Accuracy of classifier

Accuracy guarantee

classifier f

What about the accuracy guarantee $1-\epsilon$?



Training images

Reward of classifier

Rewards on nodes

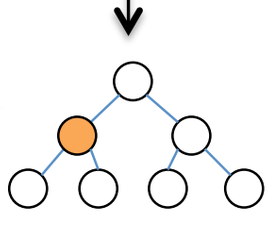
Maximize over f, $R(f,r)$

Subject to $\Phi(f) \geq 1-\epsilon$

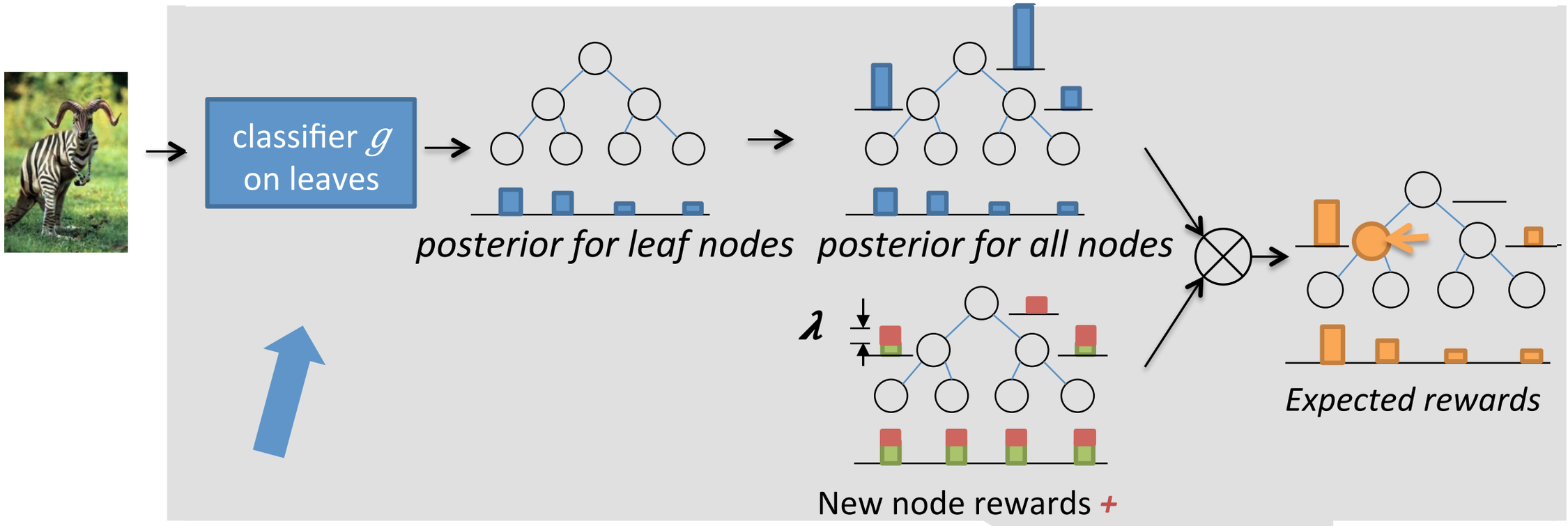
Accuracy of classifier

Accuracy guarantee

classifier f



What about the accuracy guarantee $1-\epsilon$?



Training images

Reward of classifier

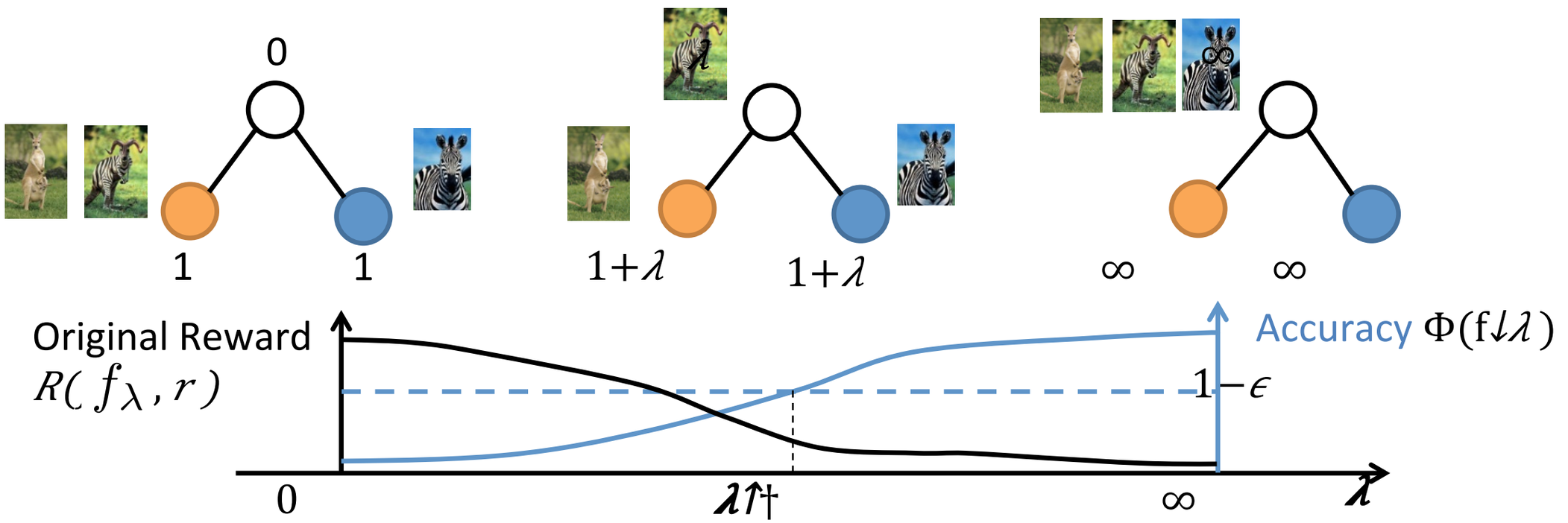
Rewards on nodes

$$\text{Maximize over } f,\ R(f,r)$$
$$\text{Subject to } \Phi(f)\geq 1-\epsilon$$

Accuracy of classifier

Accuracy guarantee

classifier f



The optimal $\lambda\uparrow\dagger$ is where the accuracy is exactly $1-\epsilon$: binary search

Training images

Reward of classifier

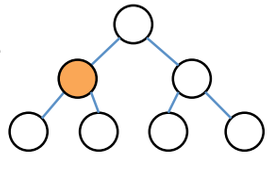
Rewards on nodes

$$\text{Maximize over } f, \quad R(f,r)$$
$$\text{Subject to} \quad \Phi(f) \geq 1-\epsilon$$

Accuracy of classifier

Accuracy guarantee

classifier f

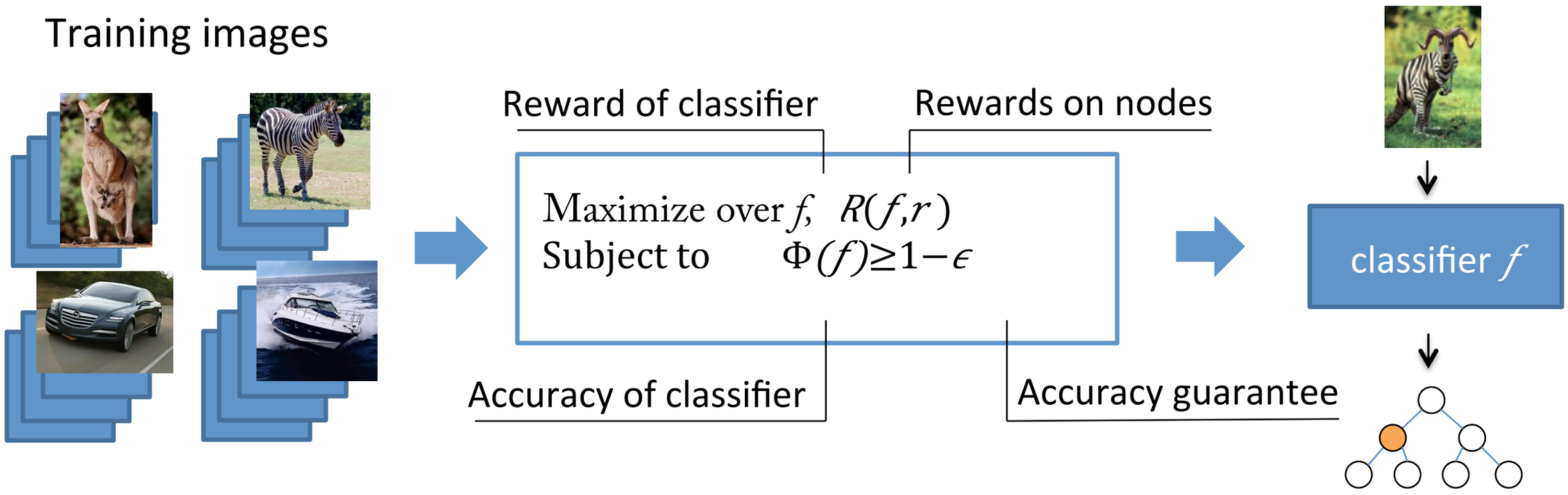


The DARTS algorithm

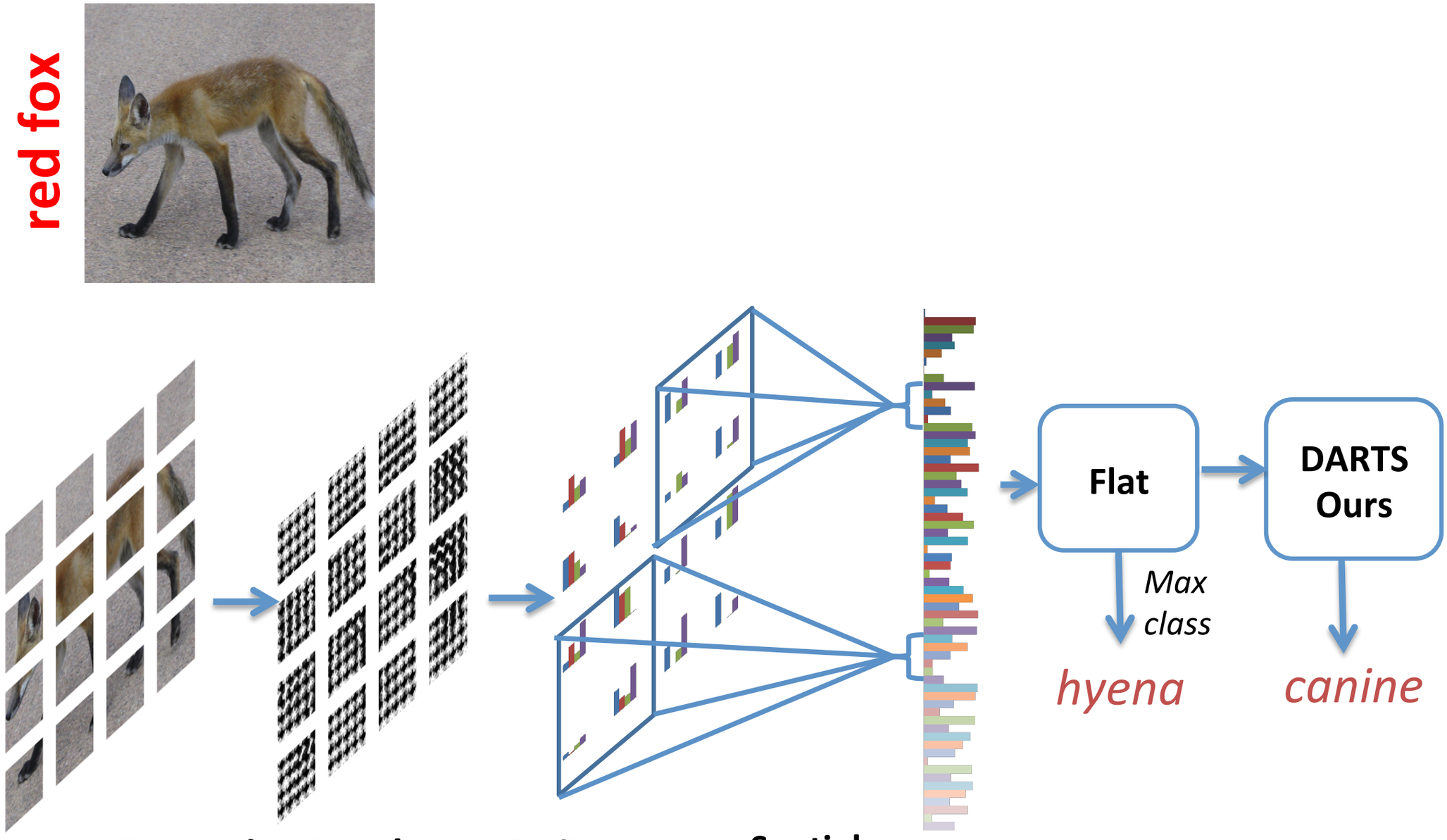
Dual **A**ccuracy **R**eward **T**rade-off **S**earch

- Train a flat classifier that gives probability estimates on the leaf nodes.
- $f_\lambda \leftarrow$ a classifier that maximizes the expected *new node rewards* ($r + \lambda$)
- Binary search to find the optimal f_λ such that f_λ is $1-\epsilon$ accurate

- λ is the dual variable in the Lagrange function
- Theorem: for any $1-\epsilon$, DARTS converges to an optimal solution except for artificial cases (no worries in practice).



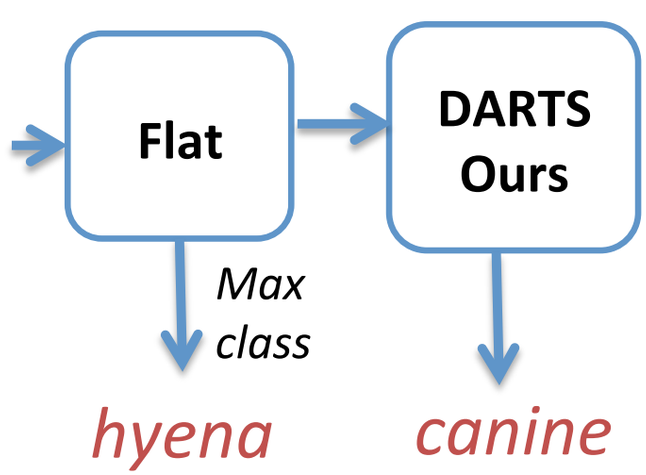
Recognition Pipeline



Recognition Pipeline

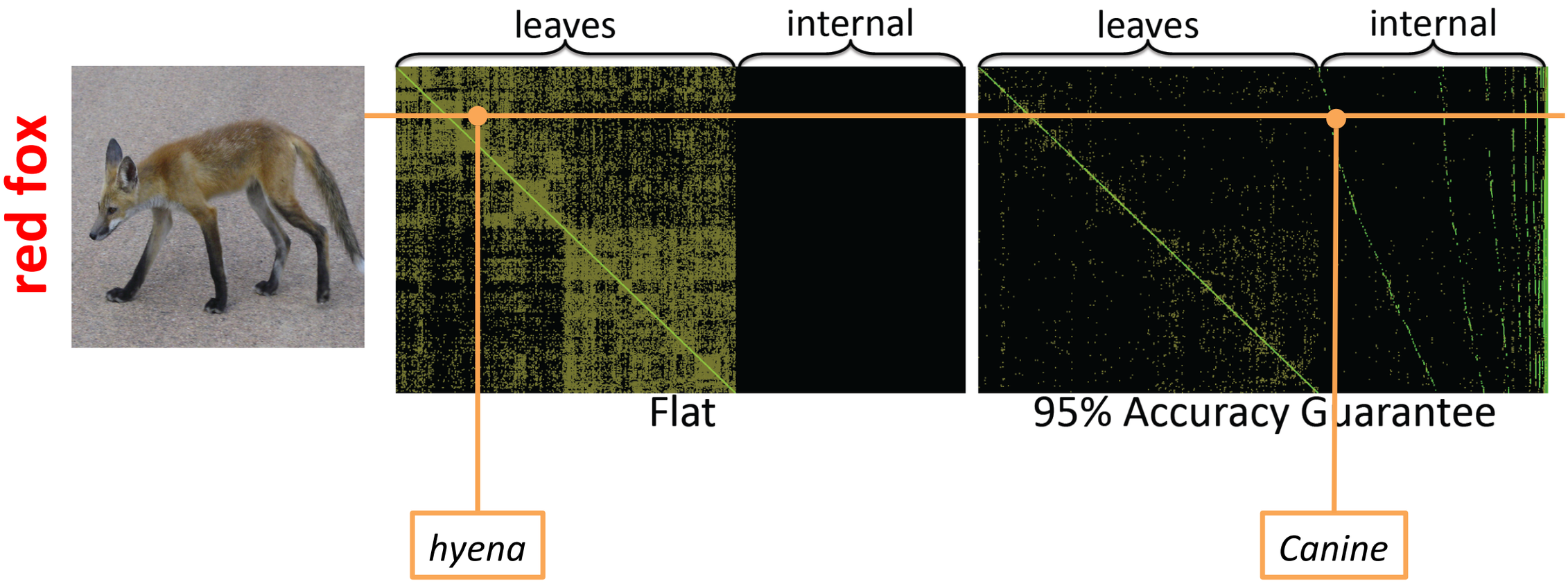
red fox

Black box



Classification **DARTS**

Results

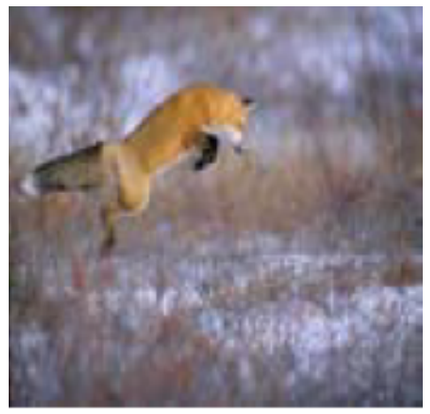

red fox
leaves
internal
leaves
internal
Flat
95% Accuracy Guarantee
hyena
Canine

Some examples

red fox



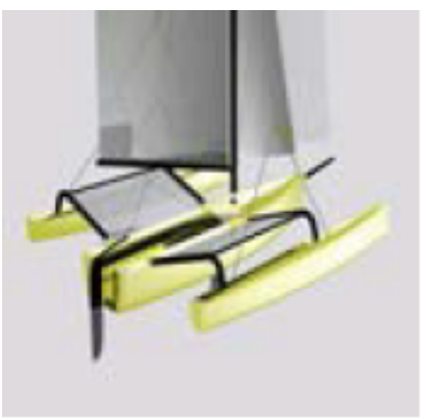




Flat	hyena	Egyptian cat	orangutan	mantis	jelly fungus
Ours	**canine**	**carnivore**	**mammal**	**animal**	**living thing**

trimaran

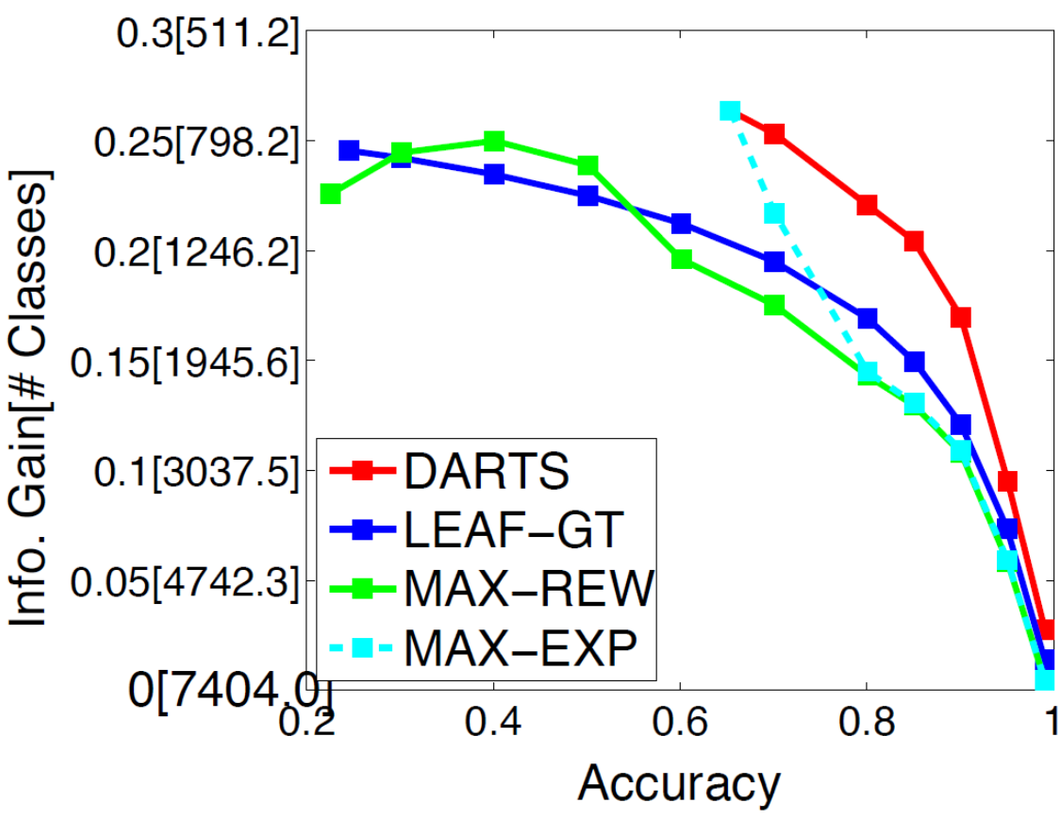


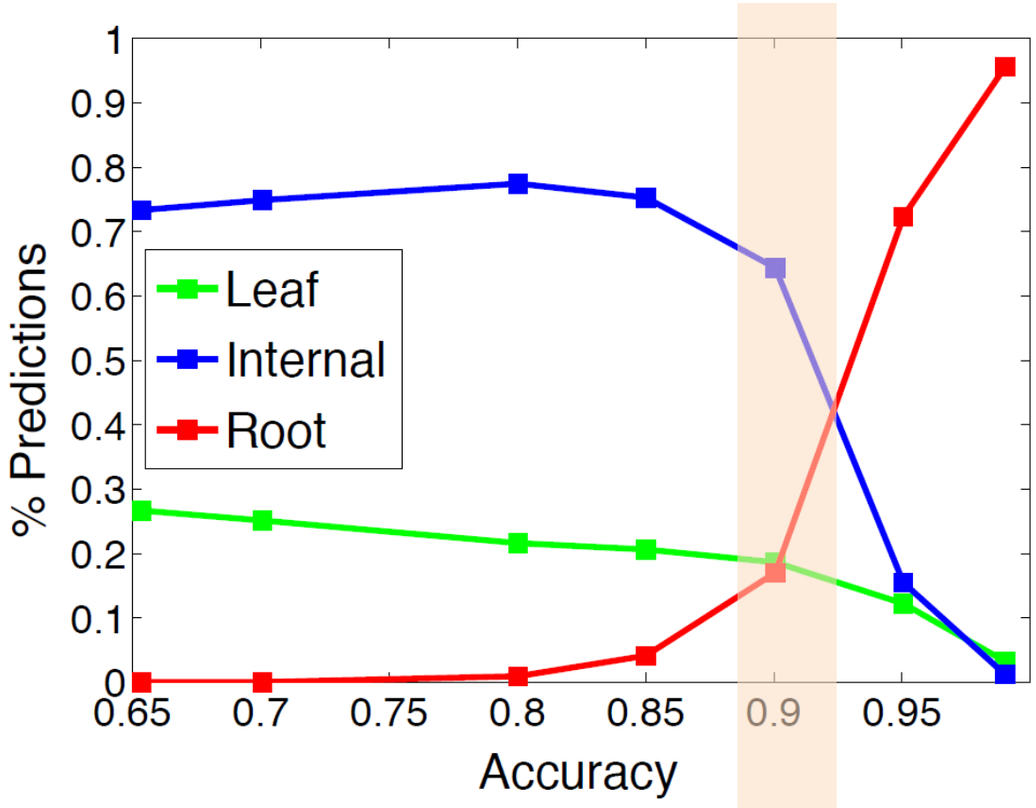




Flat	catamaran	submarine	airship	iron	electric guitar
Ours	**sailboat**	**watercraft**	**craft**	**artifact**	**artifact**

Results

Datasets: 10,000 image classes from ImageNet (~9million images)
Baselines: Flat classifier with a reject option, etc.





10K classes: 90% accurate, 19% on leaf nodes, 64% non-root internal nodes, 17% "entity"

Some more (extreme) examples




Flat	bobsled	pheasant	mortar	canoe
Ours	**vehicle**	**animal**	**edible fruit**	**watercraft**




Flat	loggerhead	cannon	Bouvier des Flandres	grapefruit
Ours	**animal**	**animal**	**living thing**	**citrus fruit**

References

- Alex Berg, "Toward richer targets in large-scale recognition", *invited talk @ NIPS'12 BigVision Workshop* *

- Deng et al, "Hierarchical Semantic Indexing for Large Scale Image Retrieval", *Proc. of CVPR 2011*

- Deng et al, "What Does Classifying More Than 10,000 Image Categories Tell Us?, *Proc. of ECCV 2010*

- Deng et al, "ImageNet: A Large-Scale Hierarchical Image Database", *Proc. of CVPR 2009*

* Available online: http://techtalks.tv/talks/toward-richer-targets-in-large-scale-recognition/57860/