# Forests and Jungles for Medical Image Analysis

A. Criminisi

## Overview

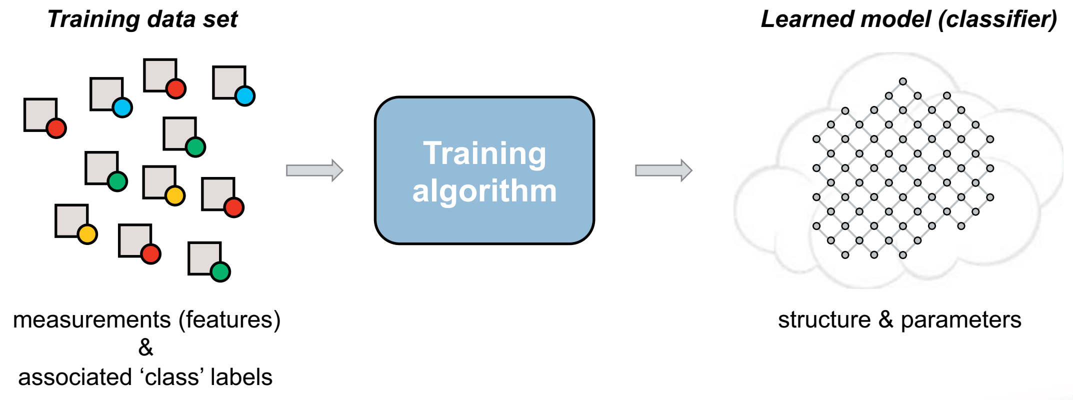
- A brief introduction to machine learning
- Decision forests and jungles
- Applications in medical image analysis
  - Anatomy localization
  - Anatomy segmentation
  - Spine detection
  - Brain tumour segmentation
  - Learned super-resolution

# Overview

- **A brief introduction to machine learning**
- Decision forests and jungles
- Applications in medical image analysis
  - Anatomy localization
  - Anatomy segmentation
  - Spine detection
  - Brain tumour segmentation
  - Learned super-resolution

# Supervised Machine Learning (classification)

**Training phase** (usually offline)

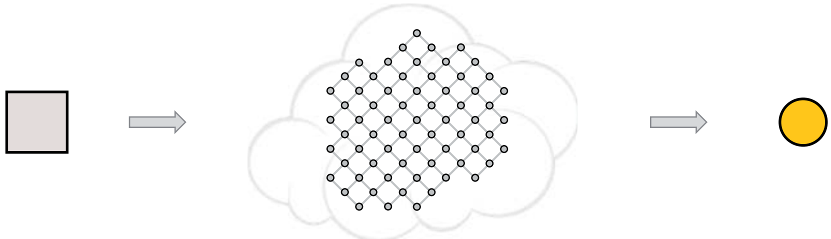



(colors denote class labels)

# Supervised Machine Learning (classification)

**Test phase** (run time, online)

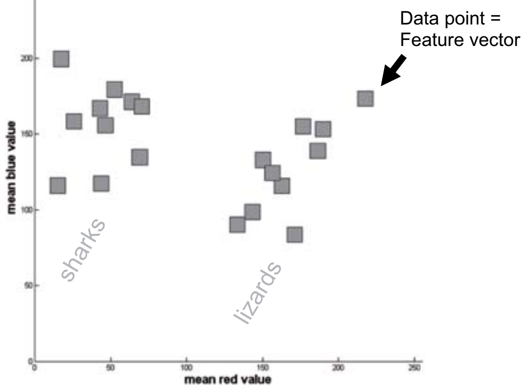
***Input test data point*** | ***Learned model (classifier)*** | ***Output***



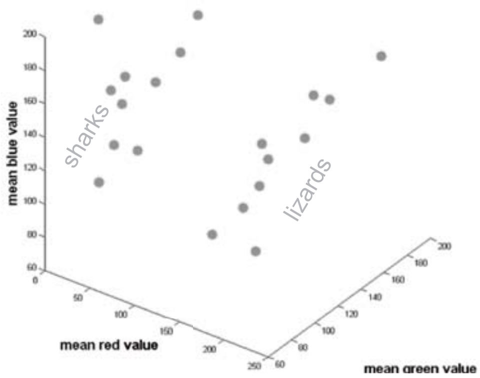
measurements (features) only | structure + parameters | predicted class label

# Data representation, feature vectors and data points

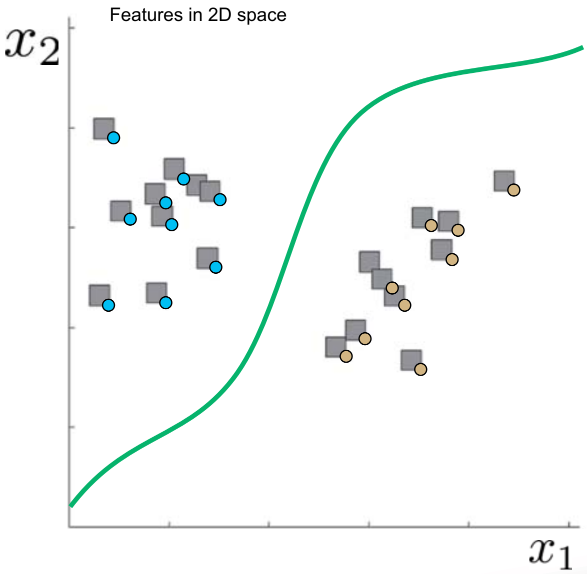
Features in 2D space




Features in 3D space




## Data representation, feature vectors and data points




induction

## Application: Kinect body part recognition

Task: assigning body part labels to each pixel in Kinect depth images

***Input test depth image*** ***Body part segmentation***

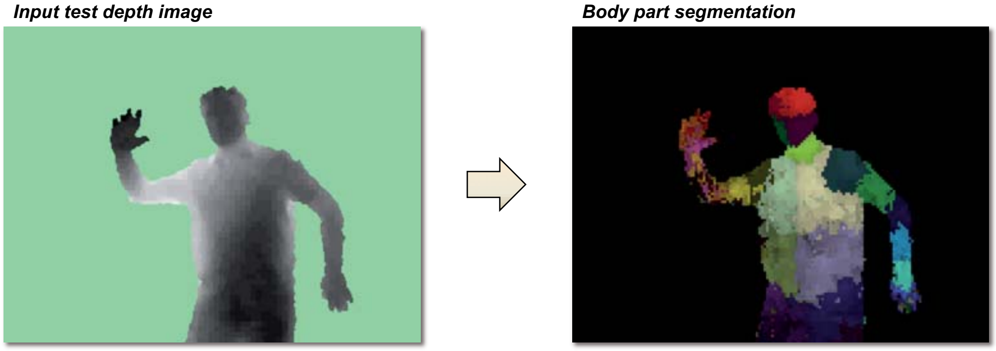
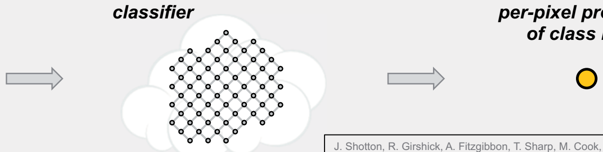


***image measurements made relative to pixel***

e.g. depth, color, neighbors

***classifier***



***per-pixel prediction of class label***

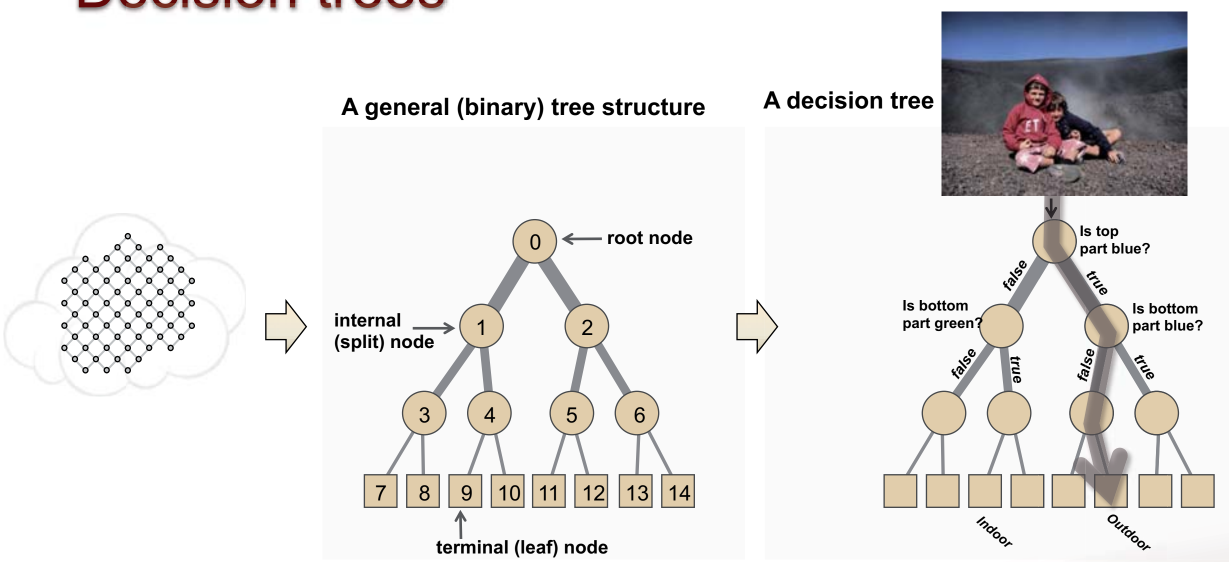
J. Shotton, R. Girshick, A. Fitzgibbon, T. Sharp, M. Cook, M. Finocchio, R. Moore, P. Kohli, A. Criminisi, A. Kipman, and A. Blake, **Efficient Human Pose Estimation from Single Depth Images**, in *Trans. PAMI*, IEEE, 2012

# Overview

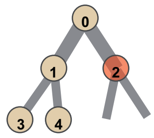
- A brief introduction to machine learning
- **Decision forests** and jungles
- Applications in medical image analysis
  - Anatomy localization
  - Anatomy segmentation
  - Spine detection
  - Brain tumour segmentation
  - Learned super-resolution

A. Criminisi and J. Shotton, **Decision Forests for Computer Vision and Medical Image Analysis**, Springer, February 2013

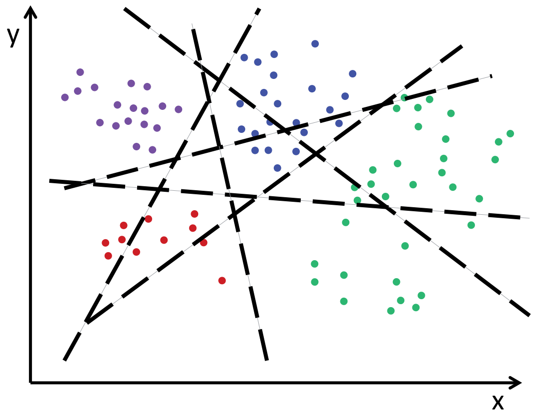
# Decision trees



# Toy Learning Example



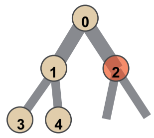

- Try several lines, chosen at random
- Keep line that best separates data
  - information gain
- Recurse

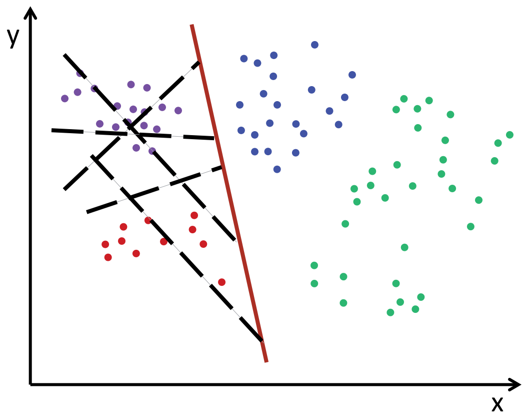




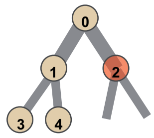
# Toy Learning Example

- Try several lines, chosen at random
- Keep line that best separates data
  - information gain
- Recurse

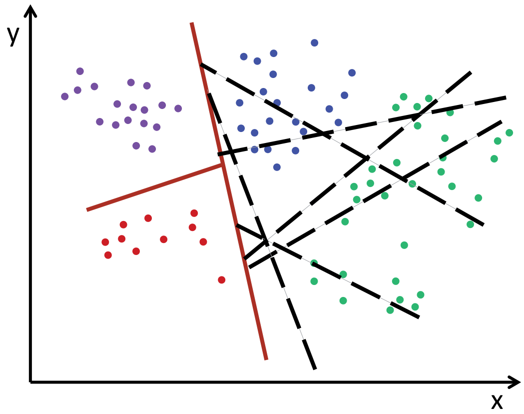



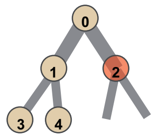
# Toy Learning Example



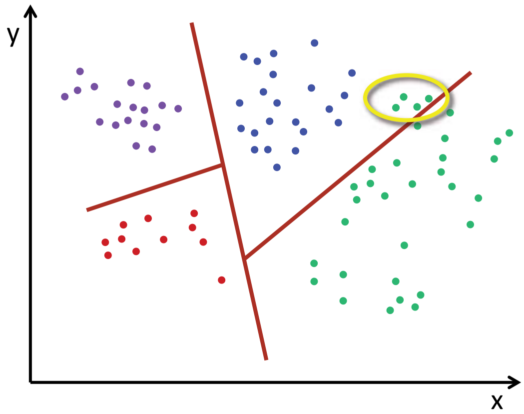

- Try several lines, chosen at random
- Keep line that best separates data
  - information gain
- Recurse



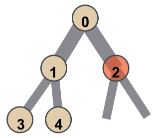

# Toy Learning Example




- Try several lines, chosen at random
- Keep line that best separates data
  - information gain
- Recurse



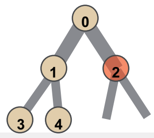
# Training objective function



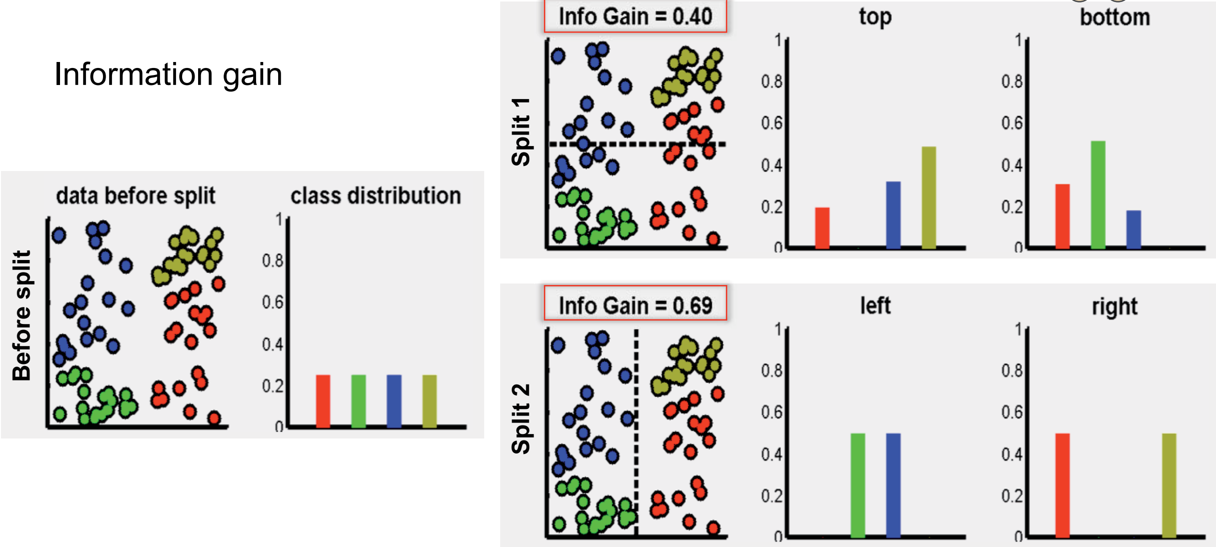

- Used to decide which candidate split function is best
- Typically an “information gain” – a very general and flexible formulation

$$I = \underbrace{H(S_j)}_{\text{entropy of examples at parent node}} - \sum_{i=\mathrm{L,R}} \underbrace{\frac{|S_j^i|}{|S_j|}}_{\text{weighting left/right children}} \underbrace{H(S_j^i)}_{\text{entropy of examples at child nodes}}$$

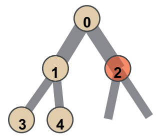
# Training objective function

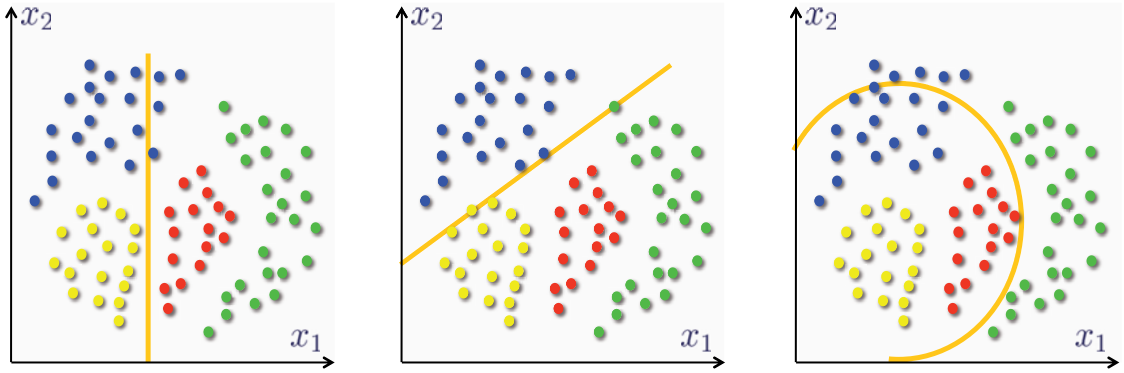



Information gain



# Examples of split functions



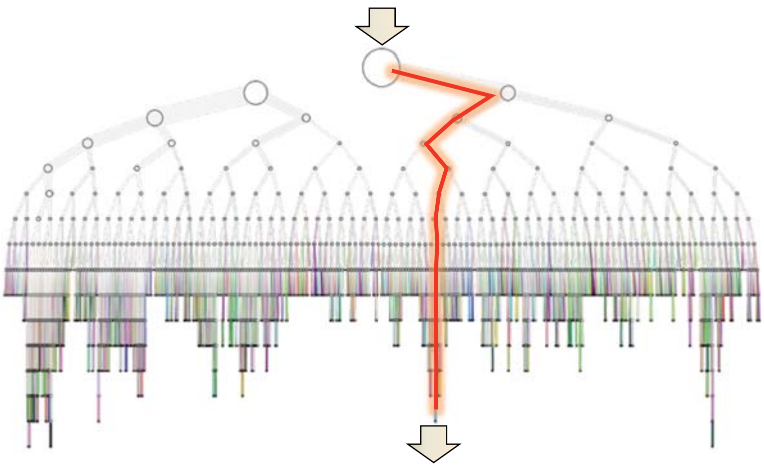




**"Axis aligned"** **"Oriented line"** **"Conic section"**

Particularly efficient

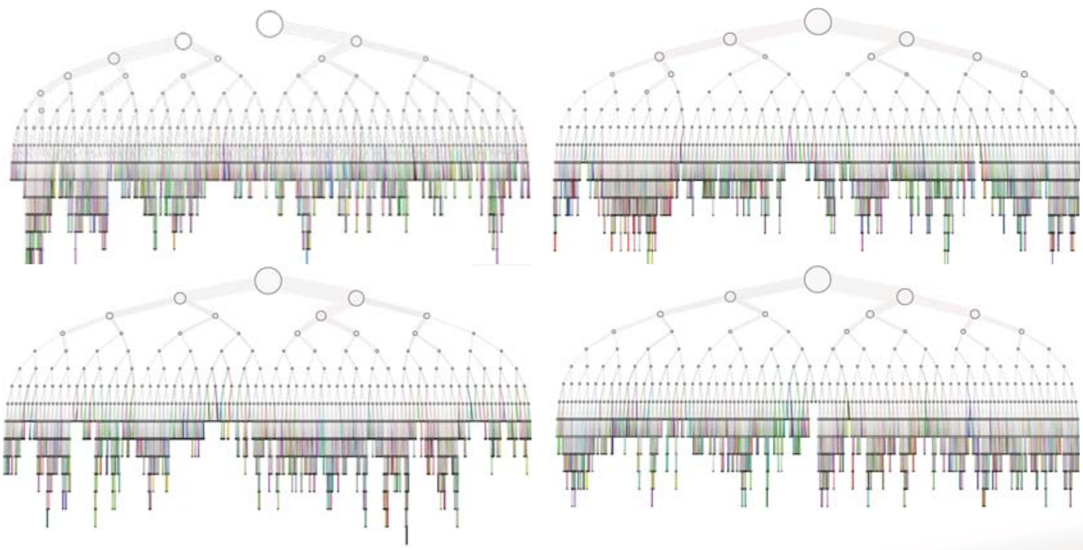
# Decision trees: test time prediction

test input data



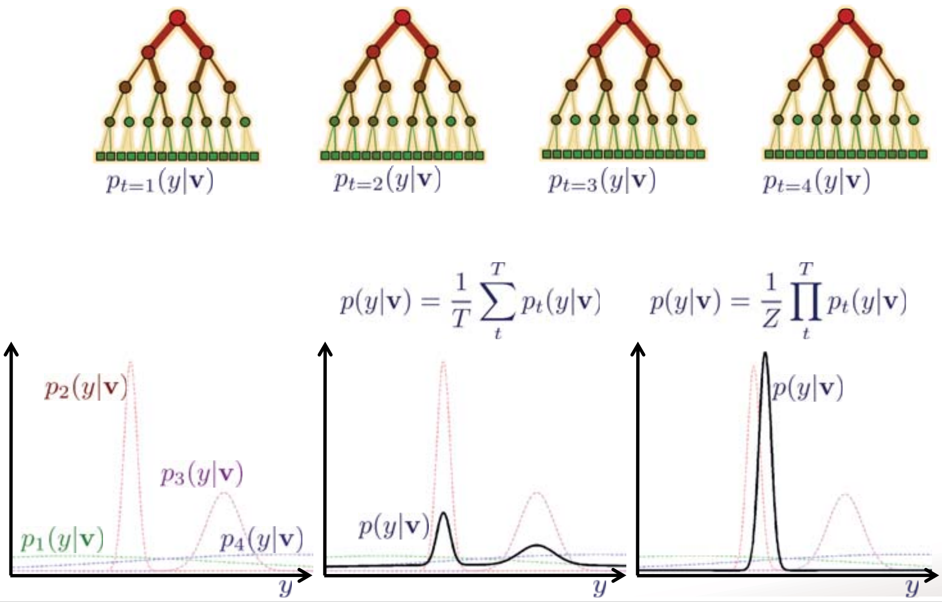
prediction

# Decision forests

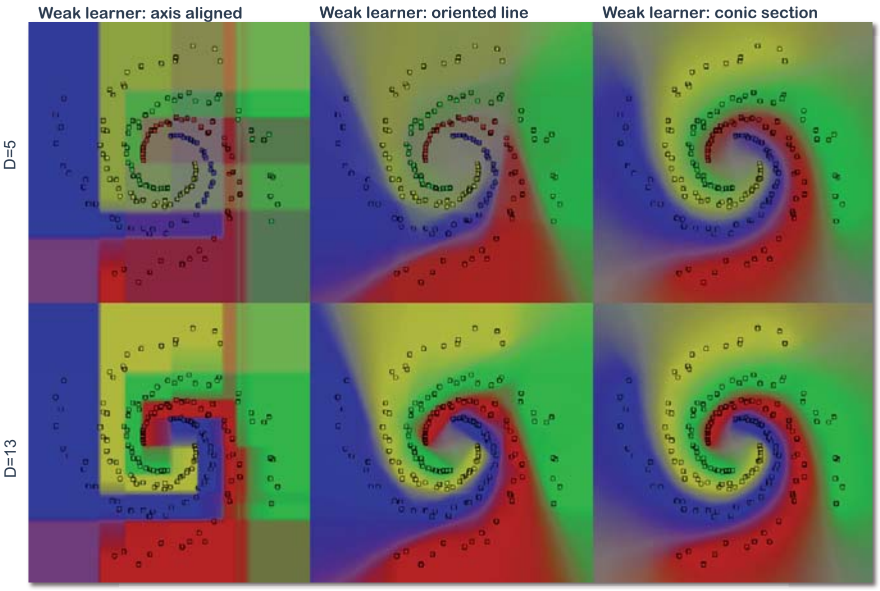


**Forest prediction is an aggregate of the predictions across all trees (*e.g.* average probability)**

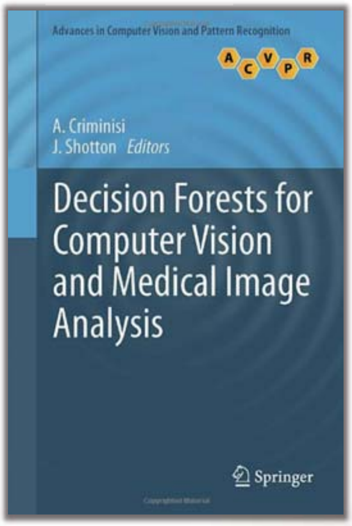
# Aggregating tree predictions



## Effect of tree depth and randomness




**Code available!**



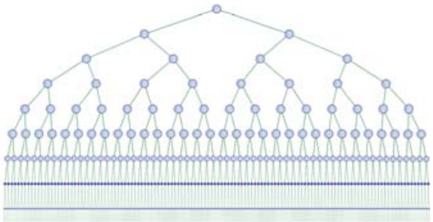

A. Criminisi and J. Shotton, **Decision Forests for Computer Vision and Medical Image Analysis**, Springer, February 2013

## Overview

- A brief introduction to machine learning
- **Decision** forests and **jungles**
- Applications in medical image analysis
  - Anatomy localization
  - Anatomy segmentation
  - Spine detection
  - Brain tumour segmentation
  - Learned super-resolution

J. Shotton, T. Sharp, P. Kohli, S. Nowozin, J. Winn, and A. Criminisi, **Decision Jungles: Compact and Rich Models for Classification**, in *Proc. NIPS*, 2013

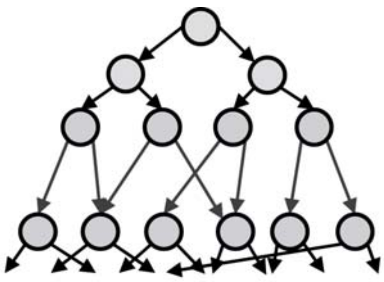
# The Memory Issue



- Given enough data, trees grow exponentially with depth
- Training deeper trees on enough data gives higher test accuracy
- Several real applications (e.g. Kinect) have “infinite” data available
- Memory concerns practically limit accuracy of decision trees

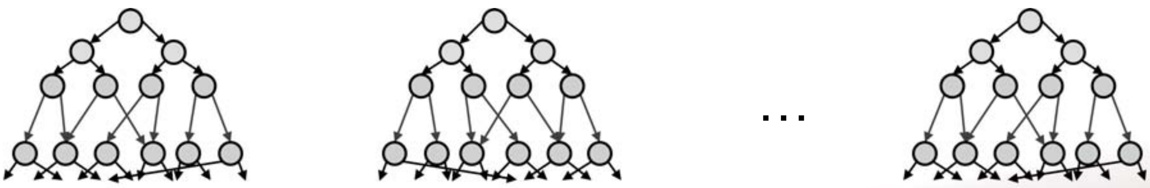
# Node Merging

- Could we find a way to merge similar nodes together?
- Change from a tree to a rooted directed acyclic graph (DAG)
  - same structure, except that nodes can have multiple parents



## Decision Jungles

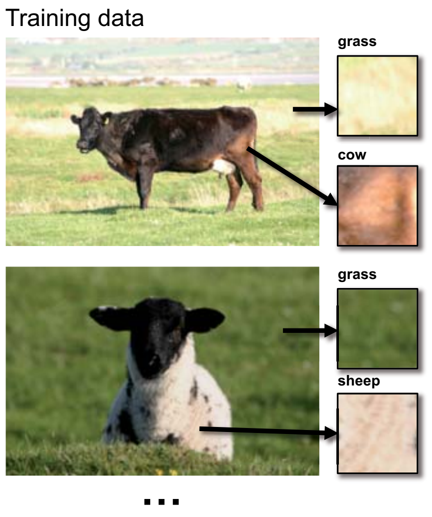
- A "jungle" is an ensemble of rooted decision DAGs
  - just as a "forest" is an ensemble of trees
- We train each DAG layer by layer, jointly optimizing both
  - the structure of the DAG
  - the split node features

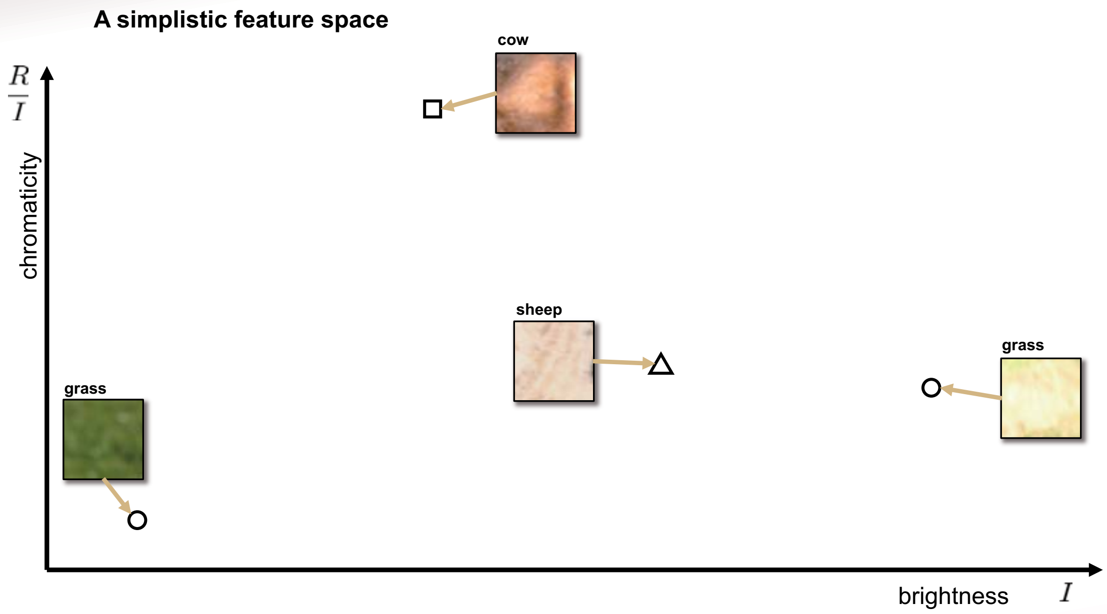


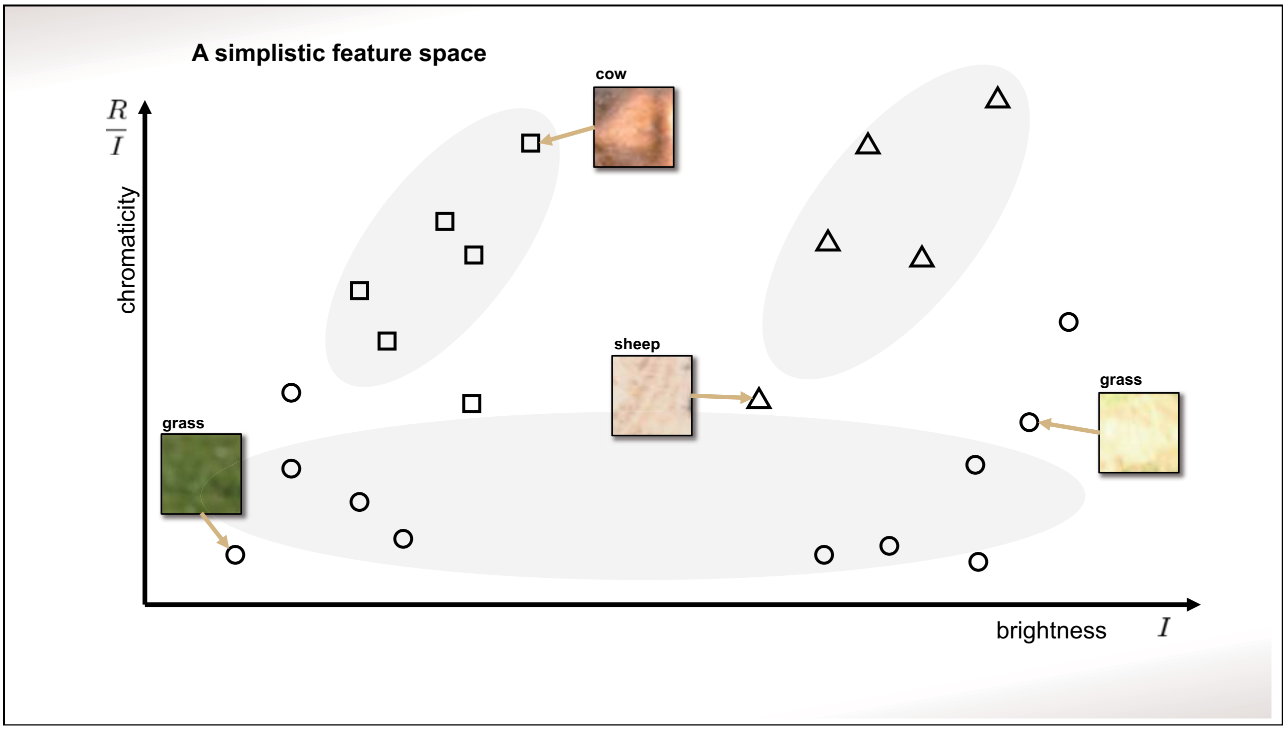
## Properties of Jungles

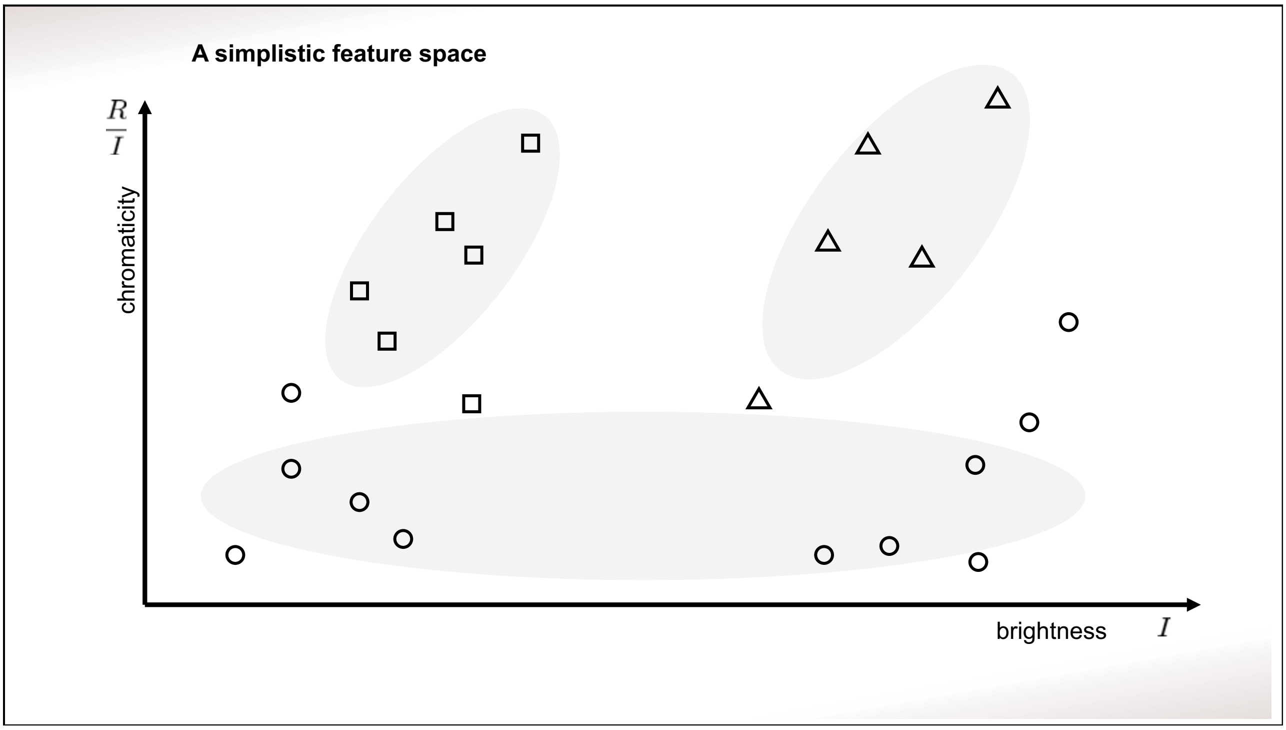
- Limited memory consumption
  - e.g. by specifying a width at each layer in the DAG
- Potentially improved generalization
  - fewer parameters
  - less "dilution" of training data

# E.g. classifying images of cows, sheep and grass

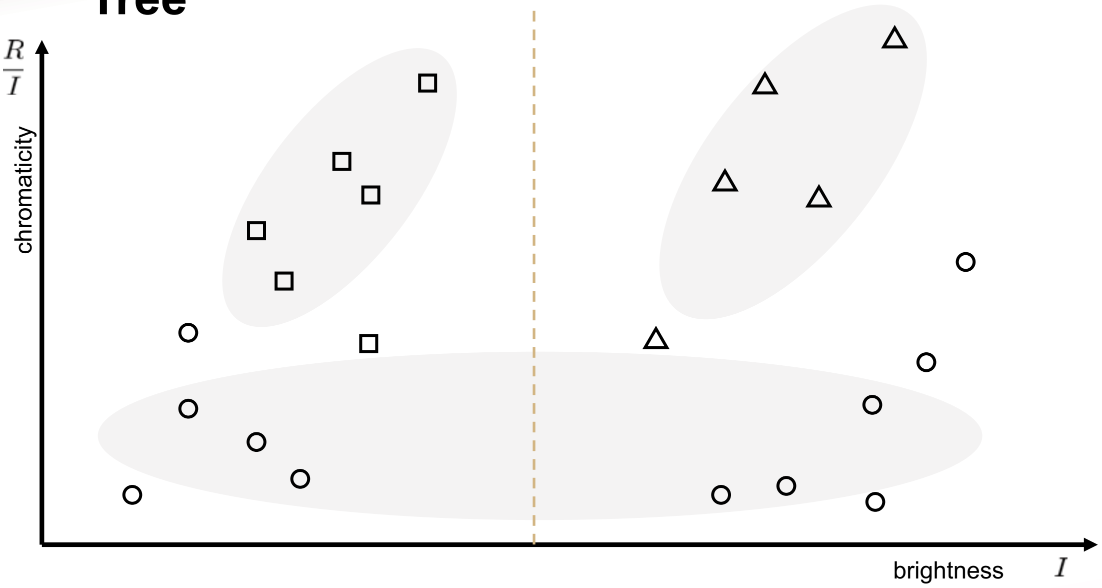



A simplistic feature space
cow
sheep
grass
grass
$\frac{R}{I}$
chromaticity
brightness
$I$

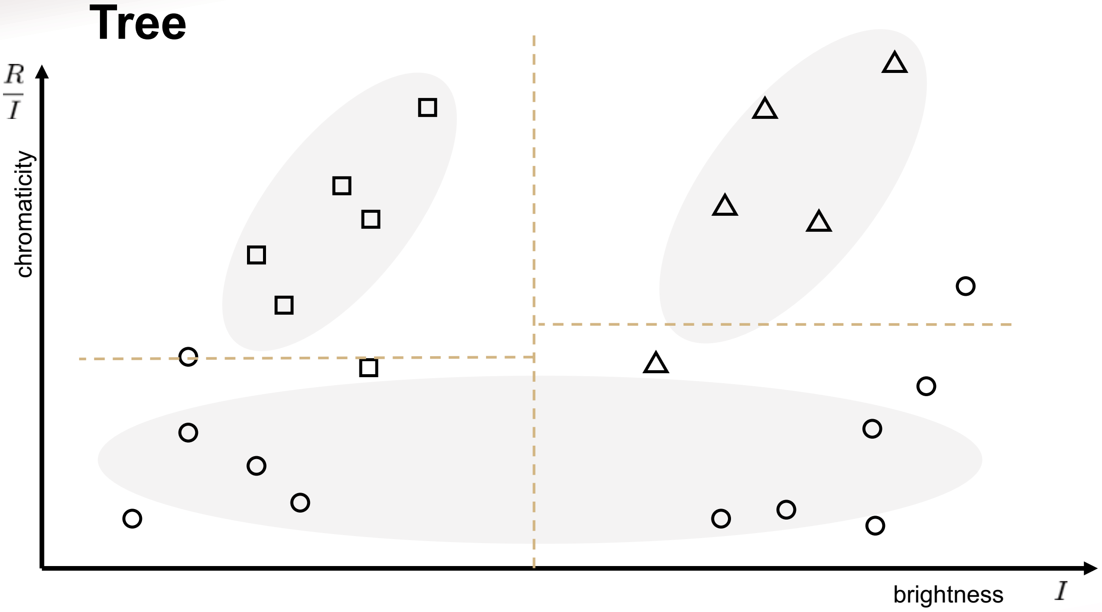

A simplistic feature space
$\frac{R}{I}$
chromaticity
brightness
$I$

# Tree

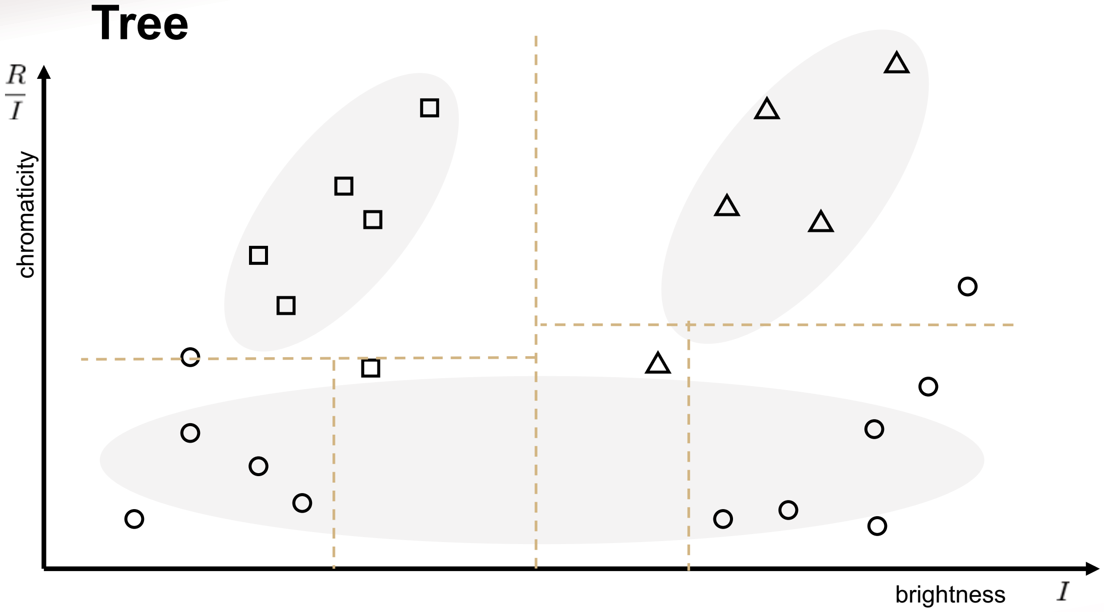



Axis-aligned splits only

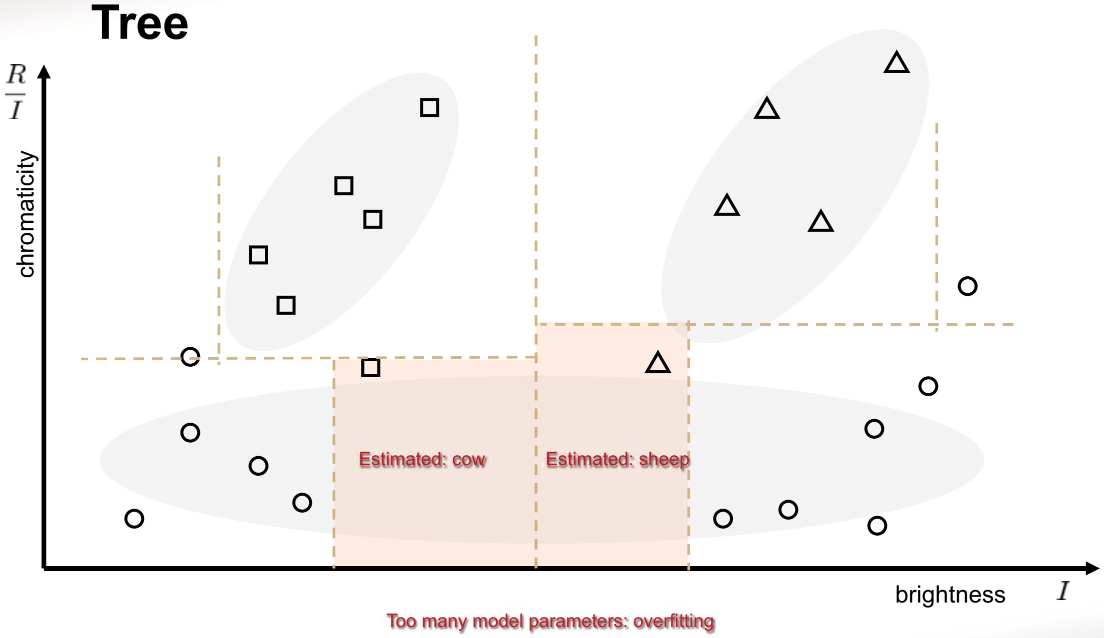
# Tree

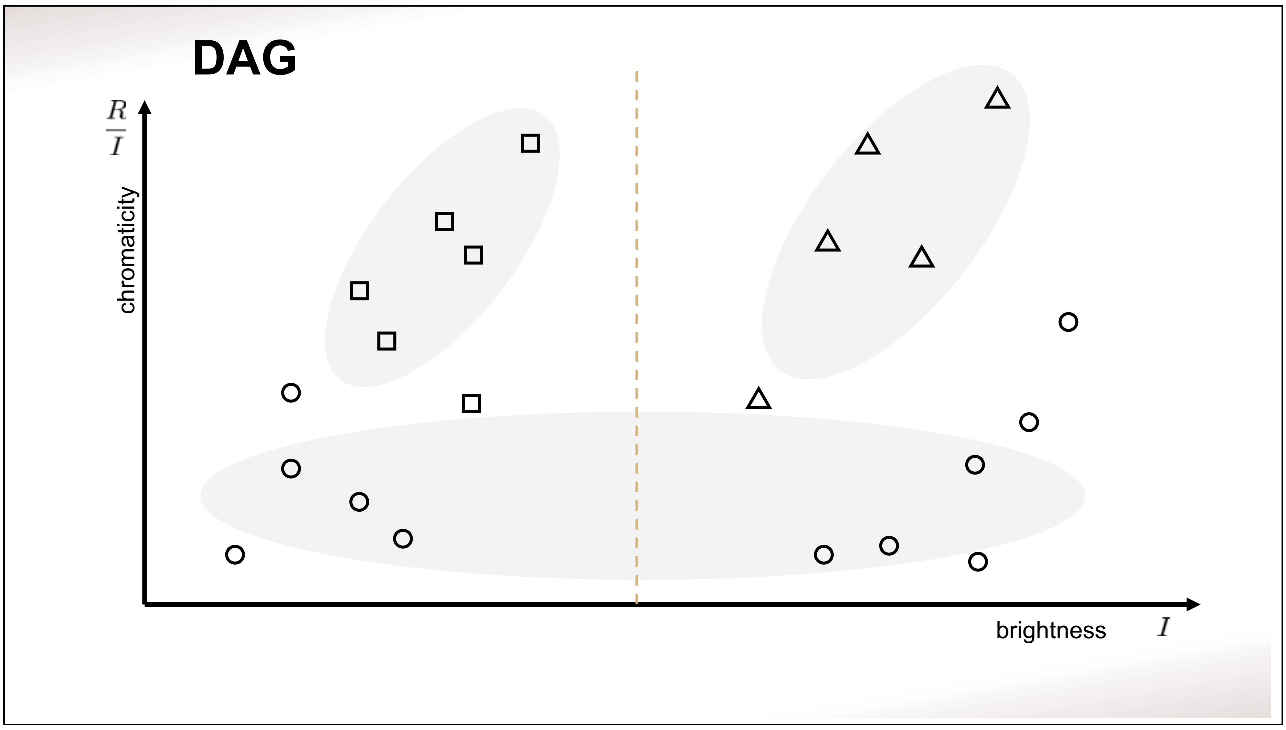


## Tree

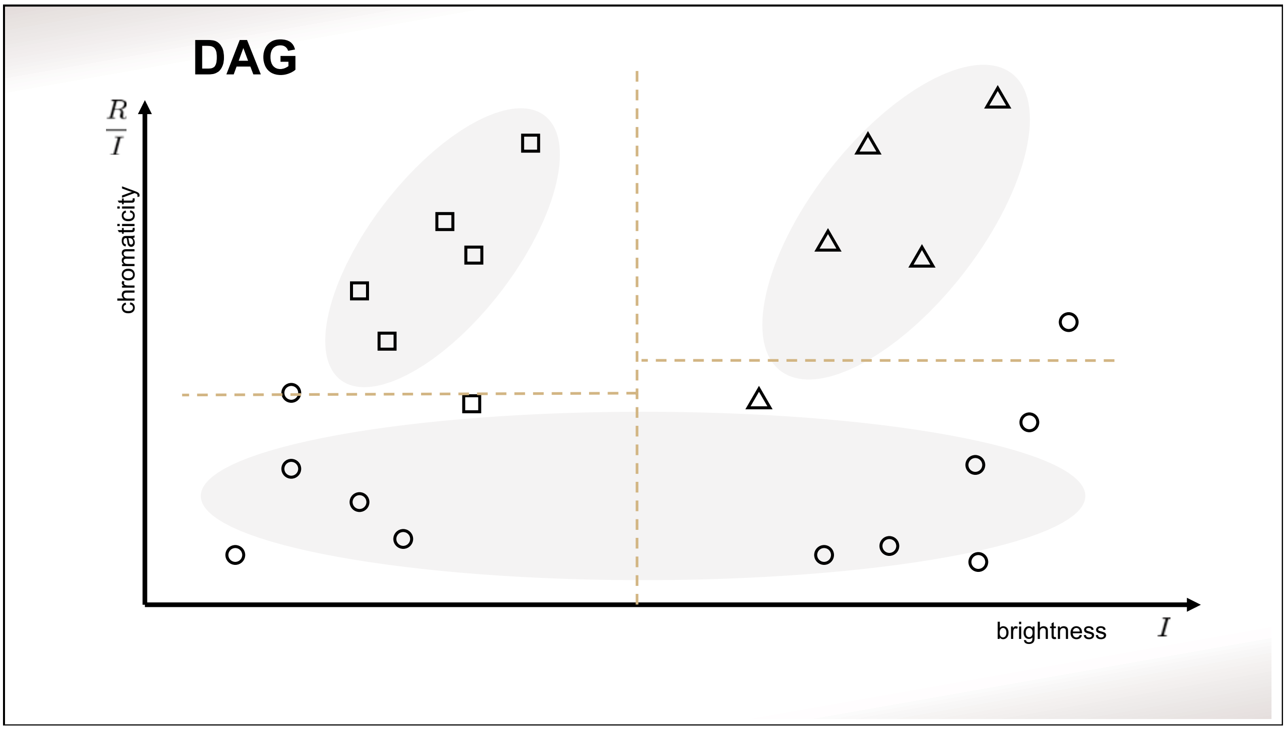

R/I
chromaticity
brightness
I

## Tree

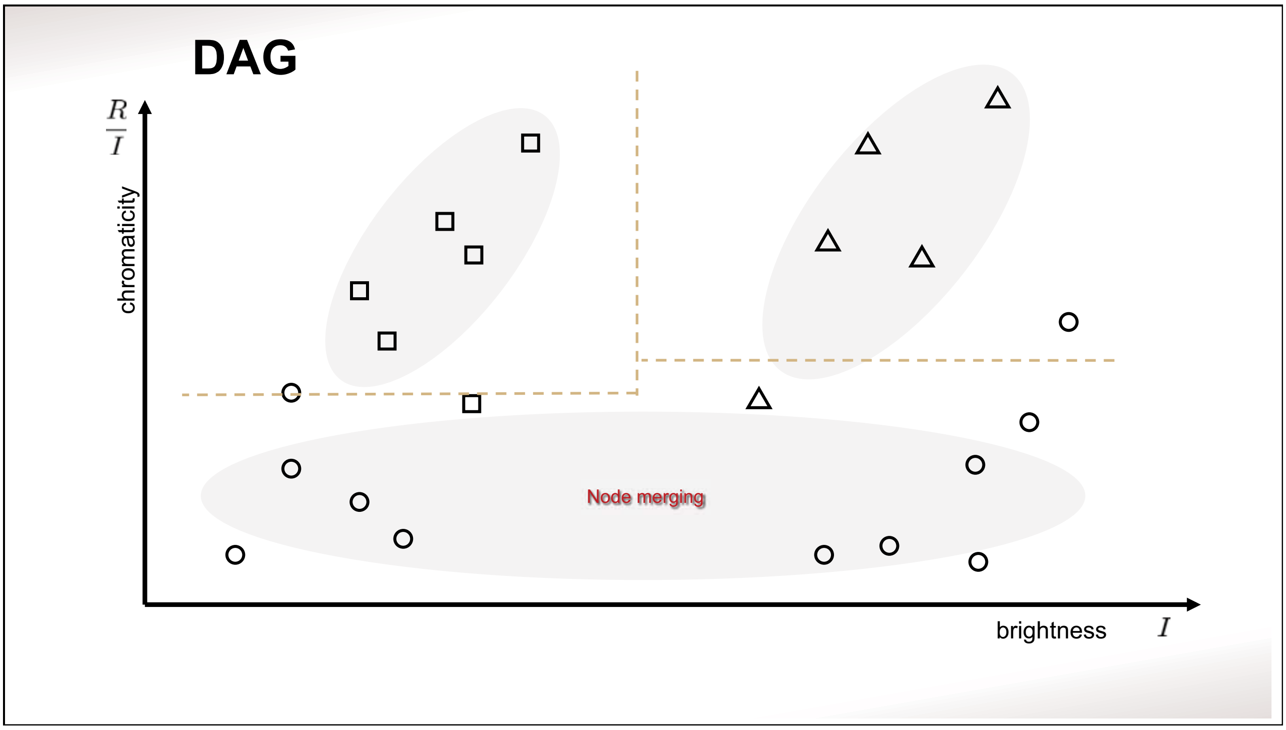

R/I
chromaticity
Estimated: cow
Estimated: sheep
brightness
I
Too many model parameters: overfitting

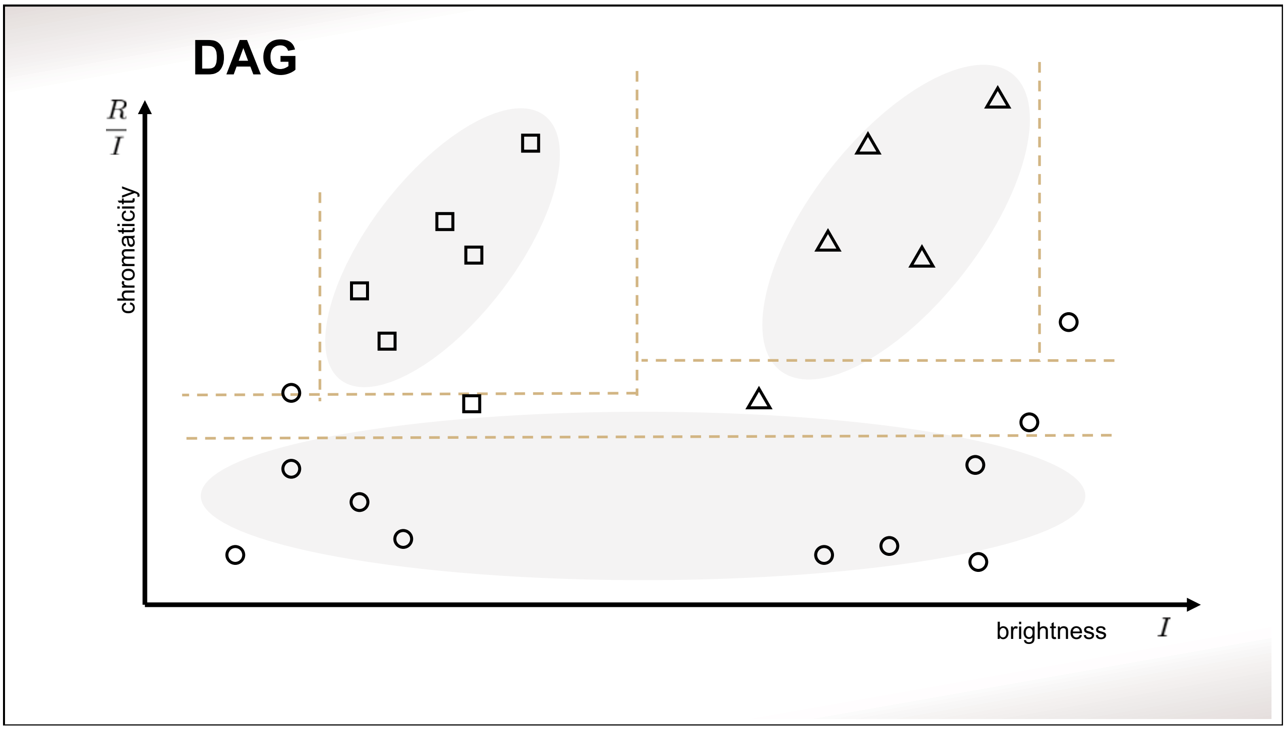


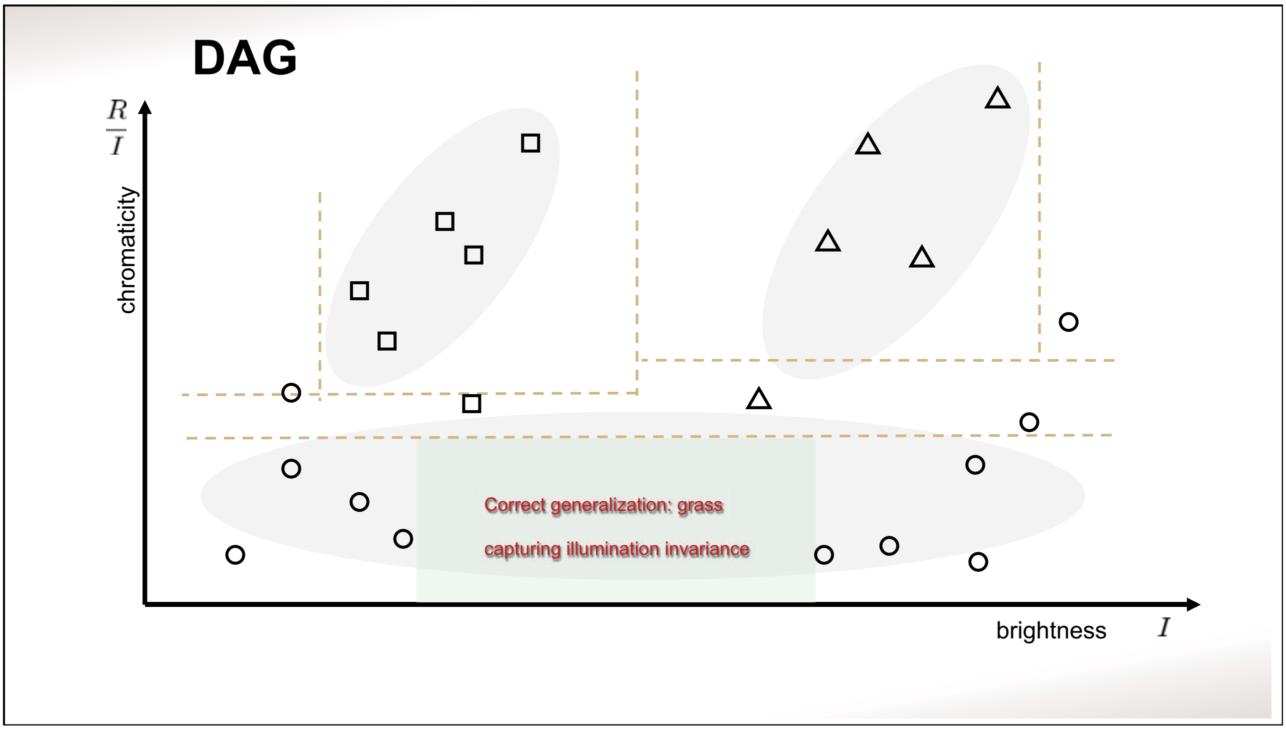
DAG
$\frac{R}{I}$
chromaticity
brightness
$I$

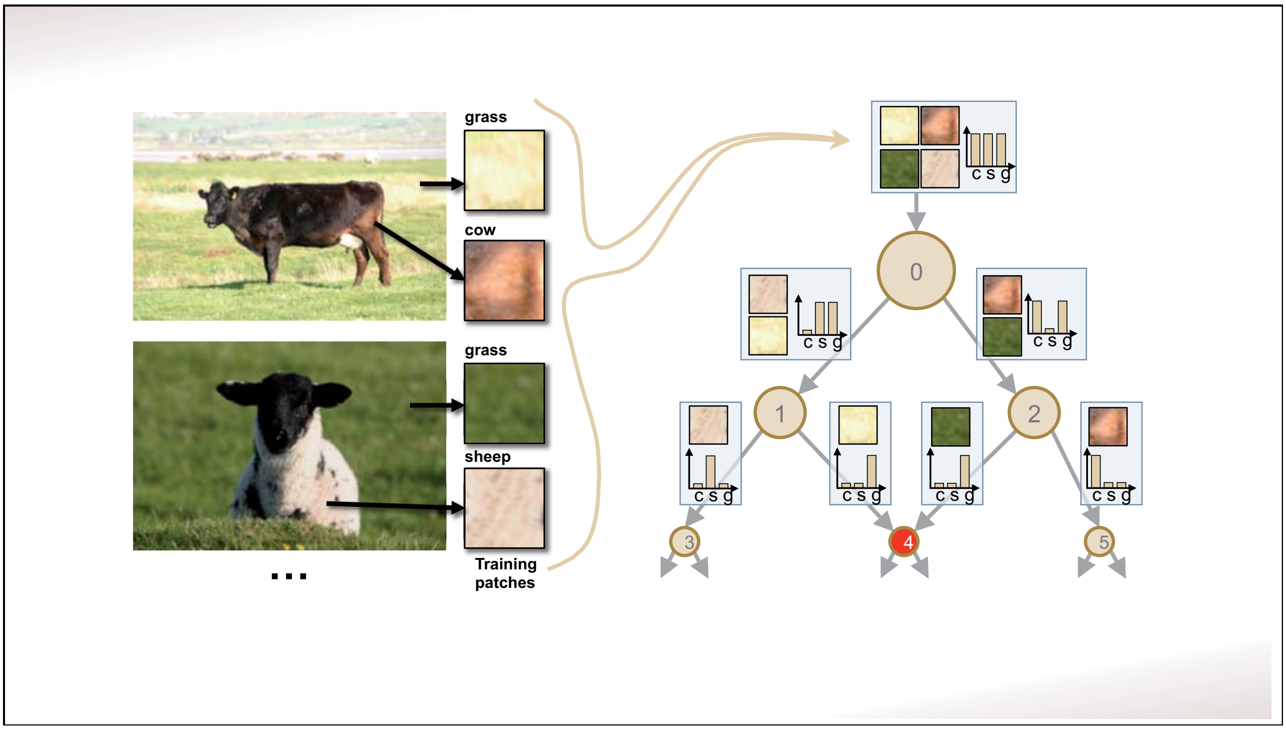



DAG
$\frac{R}{I}$
chromaticity
brightness
$I$

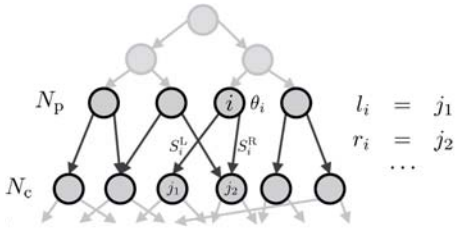

DAG
$\frac{R}{I}$
chromaticity
Node merging
brightness
$I$


DAG
$\frac{R}{I}$
chromaticity
brightness
$I$


DAG
chromaticity
$\frac{R}{I}$
Correct generalization: grass
capturing illumination invariance
brightness
$I$


grass
cow
grass
sheep
...
Training patches
c s g
0
1
2
3
4
5

## Jungles: Training Objective



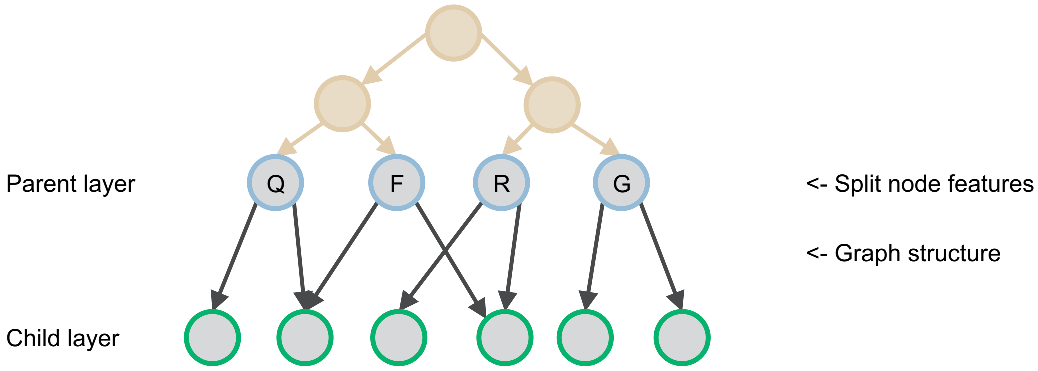

$$\underbrace{E(\{\theta_i\}, \{l_i\}, \{r_i\})}_{\text{features and branches for all parent nodes } i} = \underbrace{\sum_{j \in N_c}}_{\text{sum over child nodes } j} \overbrace{|S_j|}^{\text{number of examples at } j} \underbrace{H(S_j)}_{\text{entropy of examples that reach child node } j}$$

## Jungles: Optimization Algorithm

- Allocate a maximum of $M=|N_c|$ nodes per level
  - allows us to fix memory budget
- Simple "move-making" optimization algorithm
  - start from "feasible" initialization
  - randomly choose a parent node
  - either update its split function (given fixed DAG structure)
  - or update its left or right branch (given fixed split function)

# Algorithm Overview

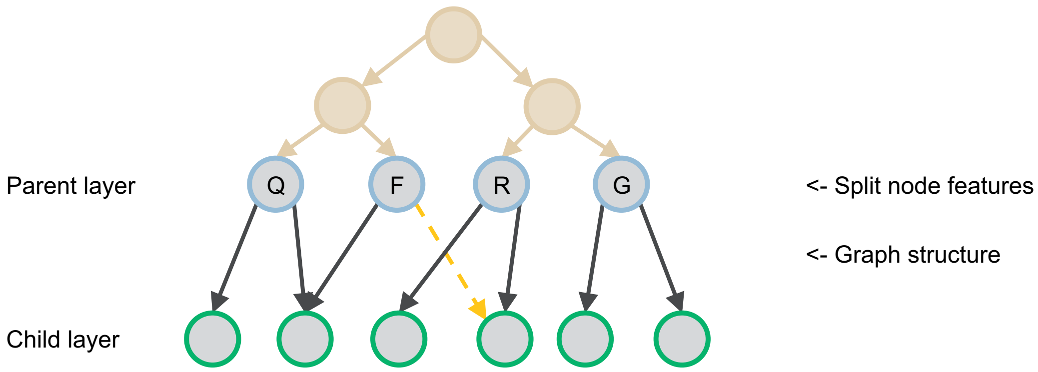
- **Train each DAG layer by layer**, jointly optimizing both
  - the structure of the DAG
  - the split node features themselves




**Current objective score = 2.87**

# Algorithm Overview

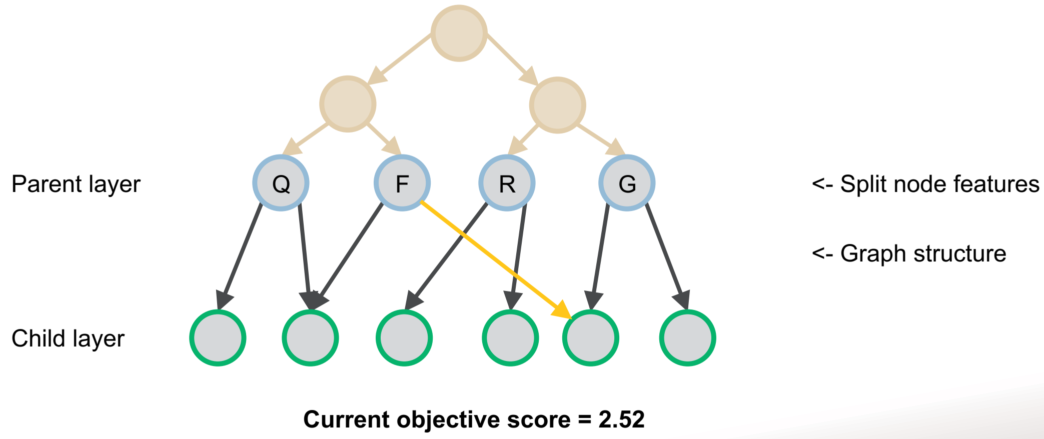
- **Train each DAG layer by layer**, jointly optimizing both
  - the structure of the DAG
  - the split node features themselves




**Current objective score = 2.87**

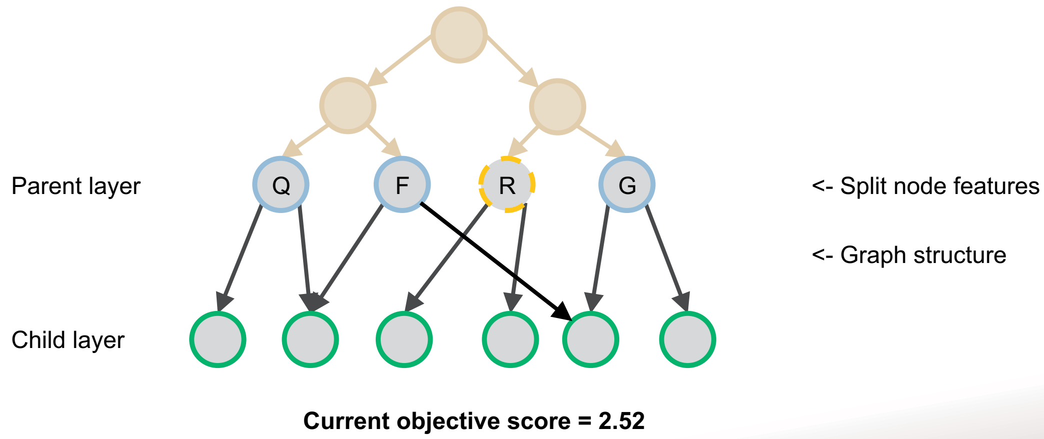
# Algorithm Overview

- **Train each DAG layer by layer**, jointly optimizing both
  - the structure of the DAG
  - the split node features themselves



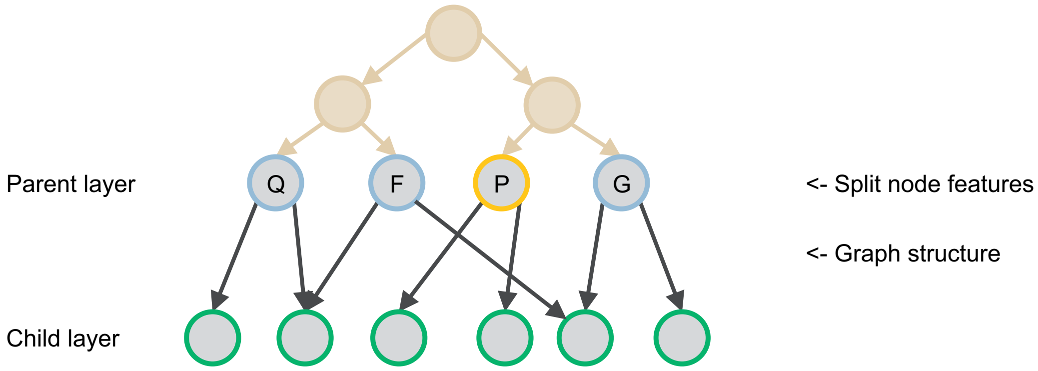

# Algorithm Overview

- **Train each DAG layer by layer**, jointly optimizing both
  - the structure of the DAG
  - the split node features themselves



# Algorithm Overview

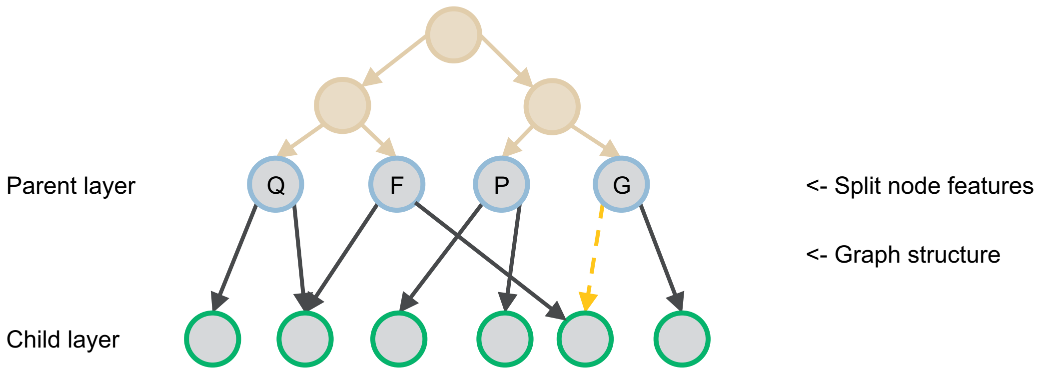
- **Train each DAG layer by layer**, jointly optimizing both
  - the structure of the DAG
  - the split node features themselves




**Current objective score = 2.12**

# Algorithm Overview

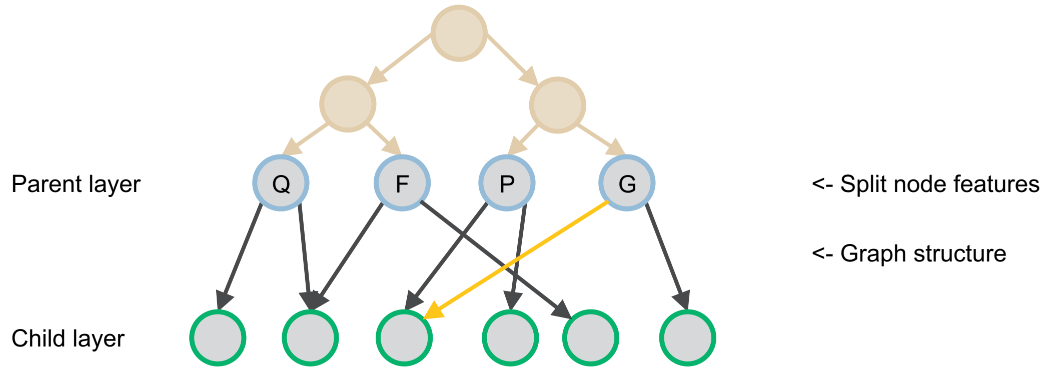
- **Train each DAG layer by layer**, jointly optimizing both
  - the structure of the DAG
  - the split node features themselves




**Current objective score = 2.12**

# Algorithm Overview

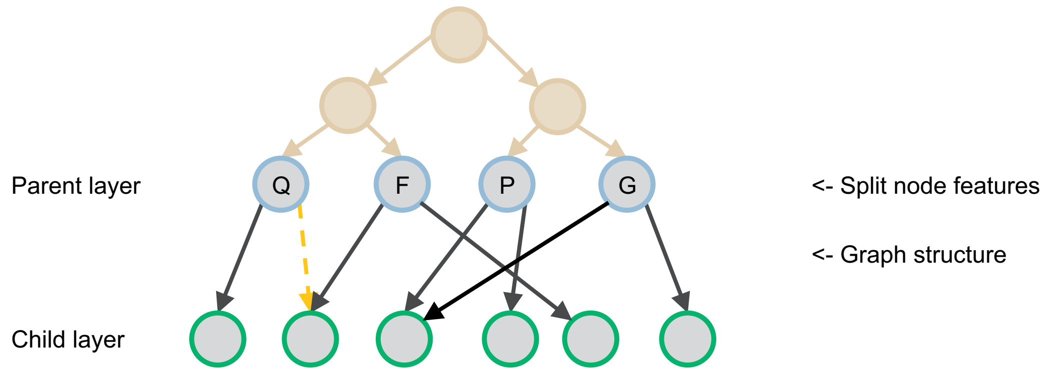
- **Train each DAG layer by layer**, jointly optimizing both
  - the structure of the DAG
  - the split node features themselves




**Current objective score = 1.72**

# Algorithm Overview

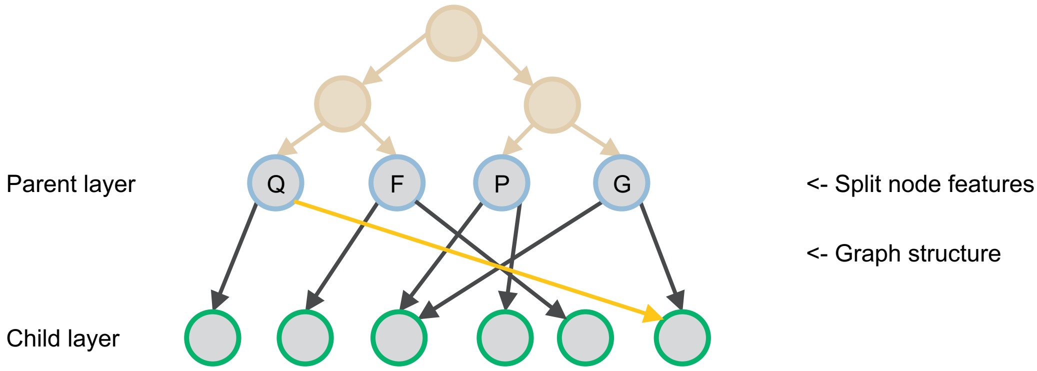
- **Train each DAG layer by layer**, jointly optimizing both
  - the structure of the DAG
  - the split node features themselves




**Current objective score = 1.72**

# Algorithm Overview

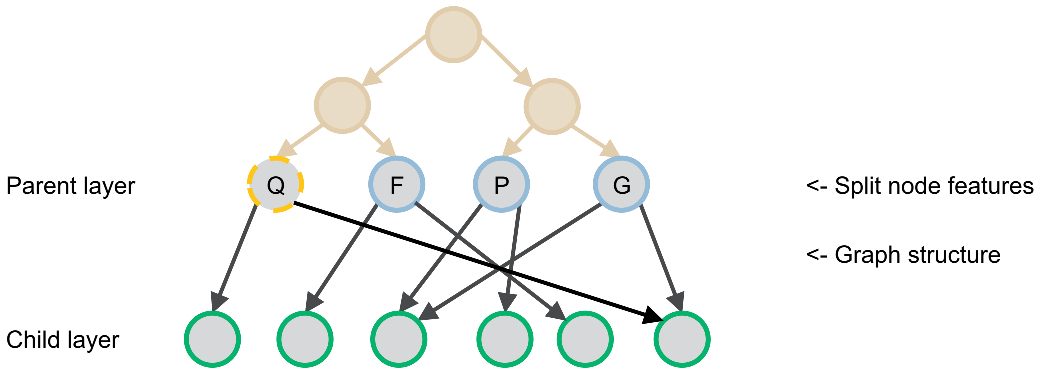
- **Train each DAG layer by layer**, jointly optimizing both
  - the structure of the DAG
  - the split node features themselves




**Current objective score = 1.59**

# Algorithm Overview

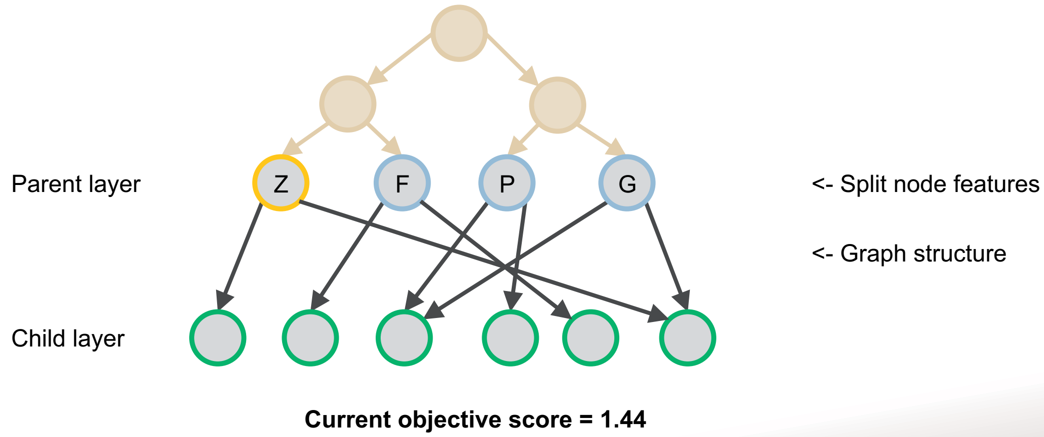
- **Train each DAG layer by layer**, jointly optimizing both
  - the structure of the DAG
  - the split node features themselves




**Current objective score = 1.59**

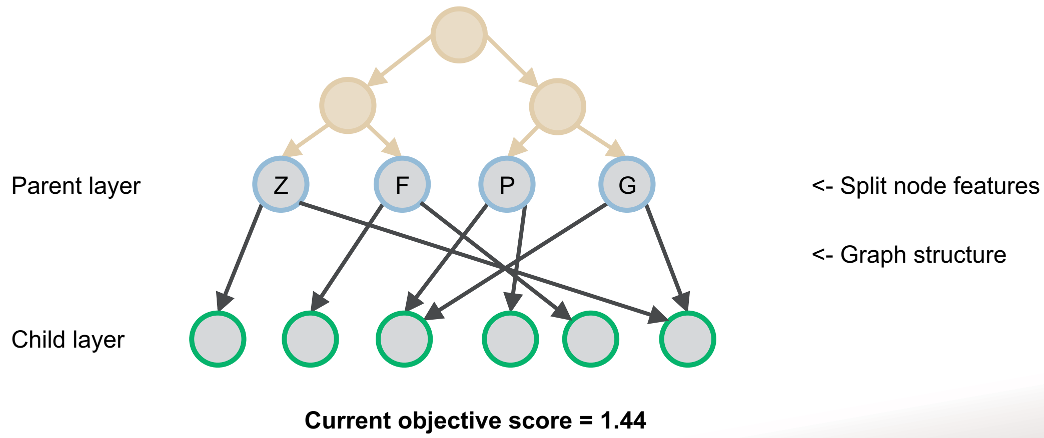
# Algorithm Overview

- **Train each DAG layer by layer**, jointly optimizing both
  - the structure of the DAG
  - the split node features themselves

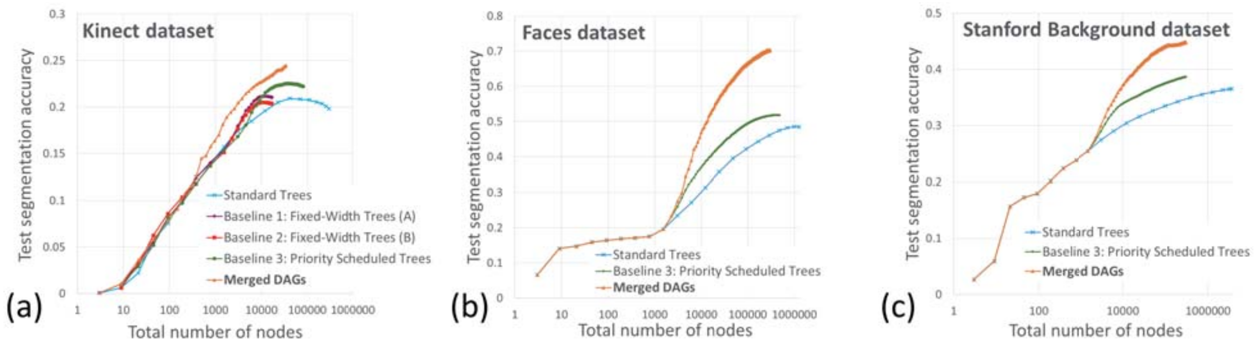



# Algorithm Overview

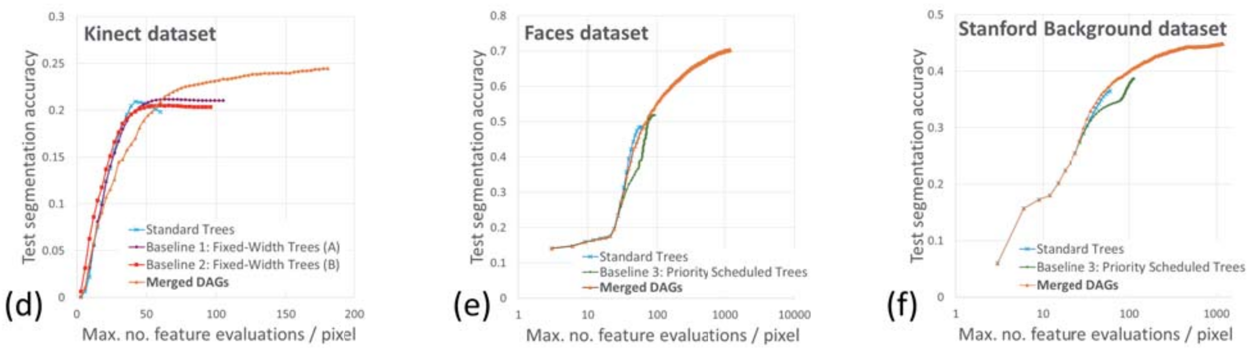
- **Train each DAG layer by layer**, jointly optimizing both
  - the structure of the DAG
  - the split node features themselves



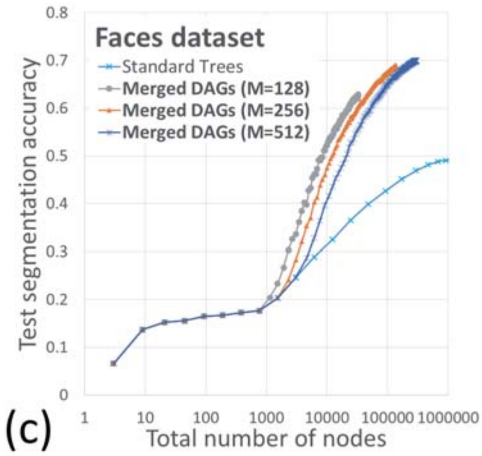
# Jungles: Results – accuracy *vs.* memory

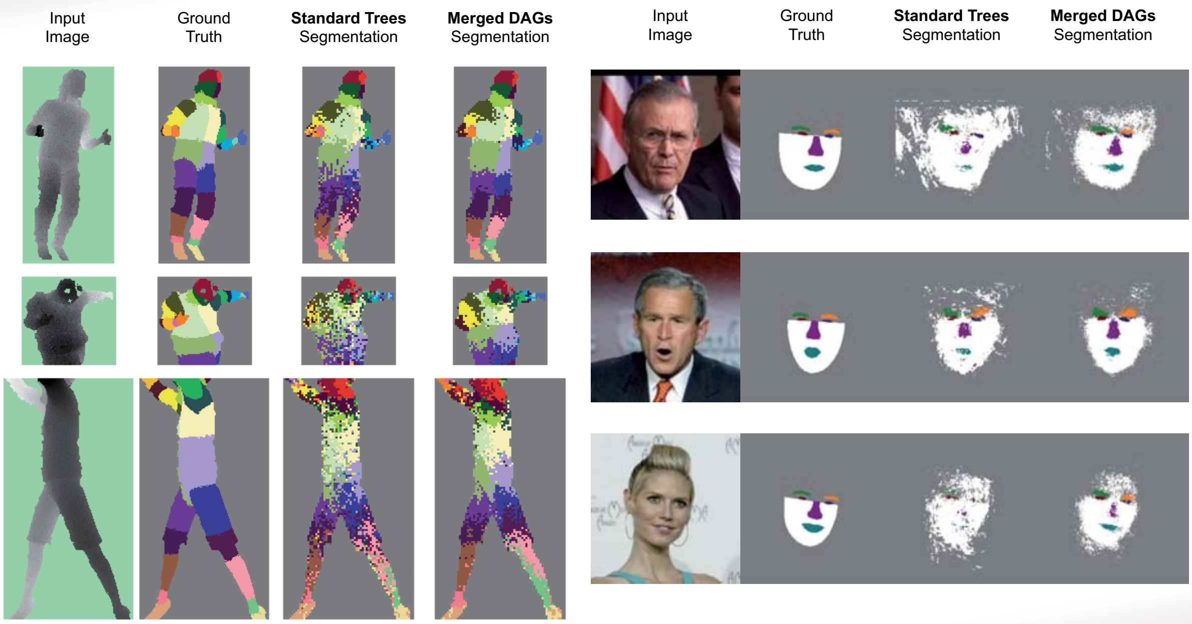


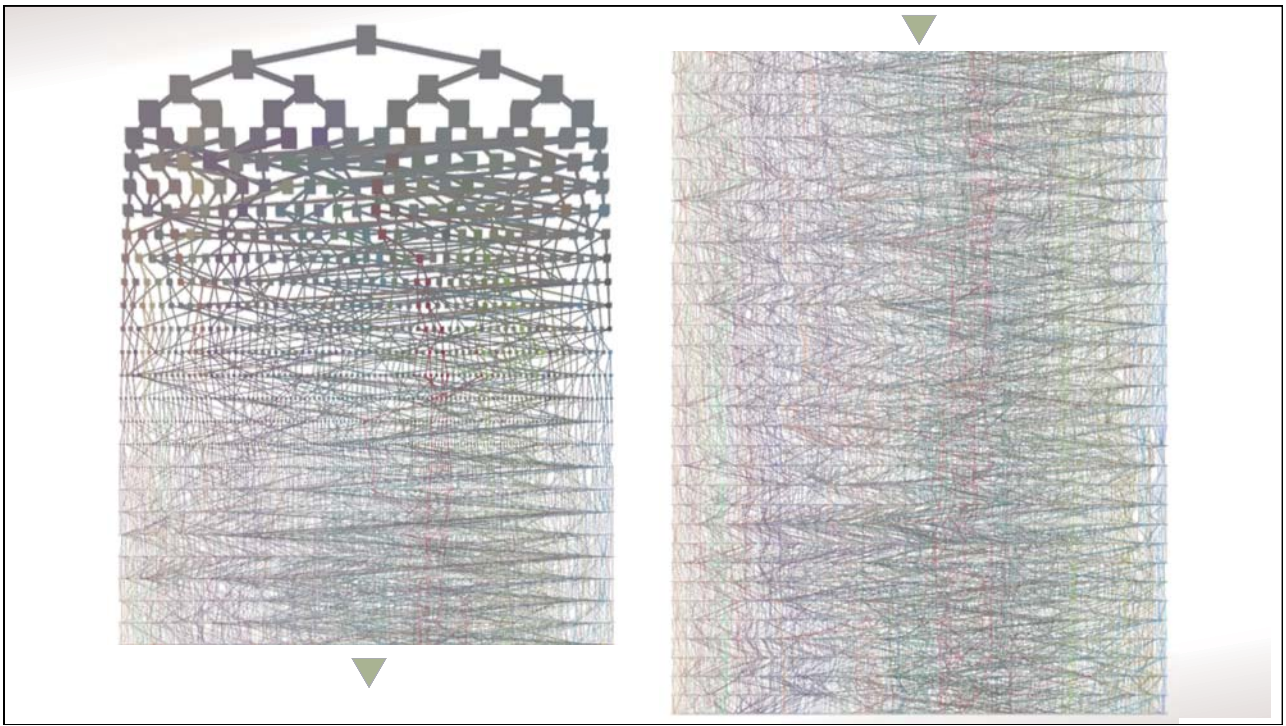

# Jungles: Results – accuracy *vs.* compute time



# Jungles: Results – node budget *M*





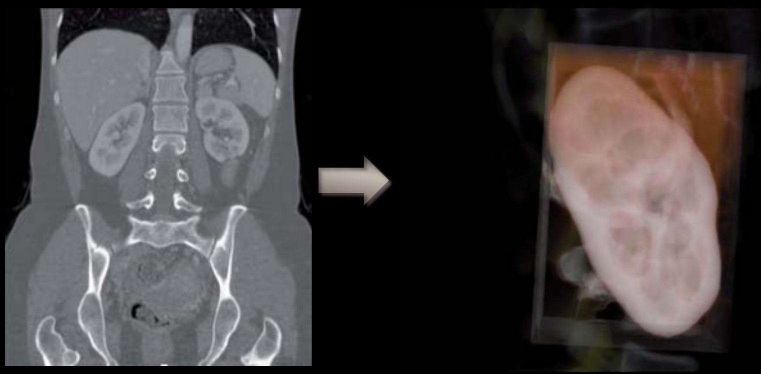
# Overview

- A brief introduction to machine learning
- Decision forests and jungles
- **Applications in medical image analysis**
  - **Anatomy localization**
  - Anatomy segmentation
  - Spine detection
  - Brain tumour segmentation
  - Learned super-resolution

A. Criminisi, D. Robertson, E. Konukoglu, J. Shotton, S. Pathak, S. White, and K. Siddiqui, **Regression Forests for Efficient Anatomy Detection and Localization in Computed Tomography Scans**, in *Medical Image Analysis (MedIA)*, Elsevier, 2013

## Anatomy Localization in 3D Computed Tomography Scans

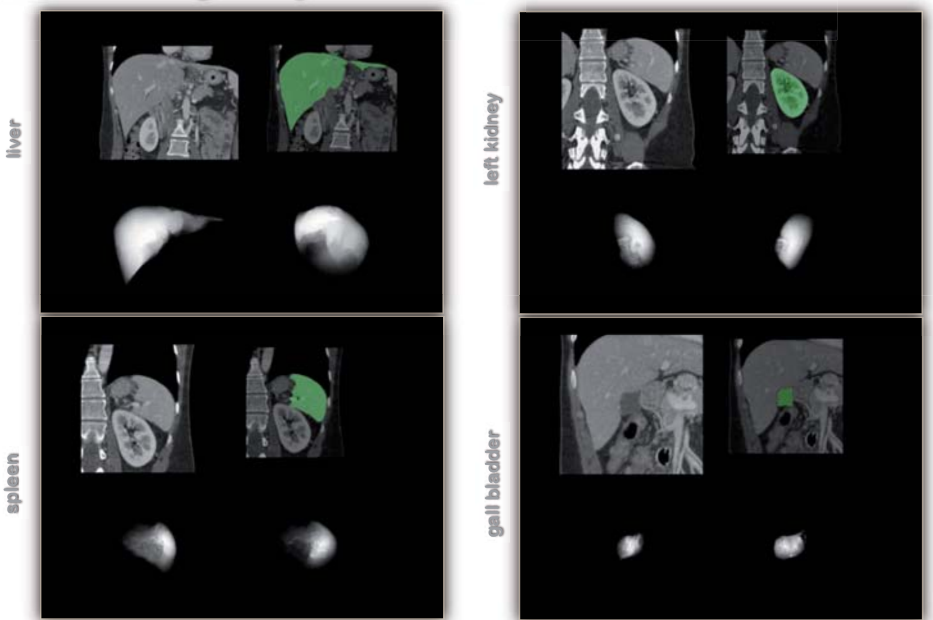
- Direct mapping of voxels to organ bounding boxes.
- No search, no sliding window.
- No atlas registration.



*Input CT scan* *Output anatomy localization*

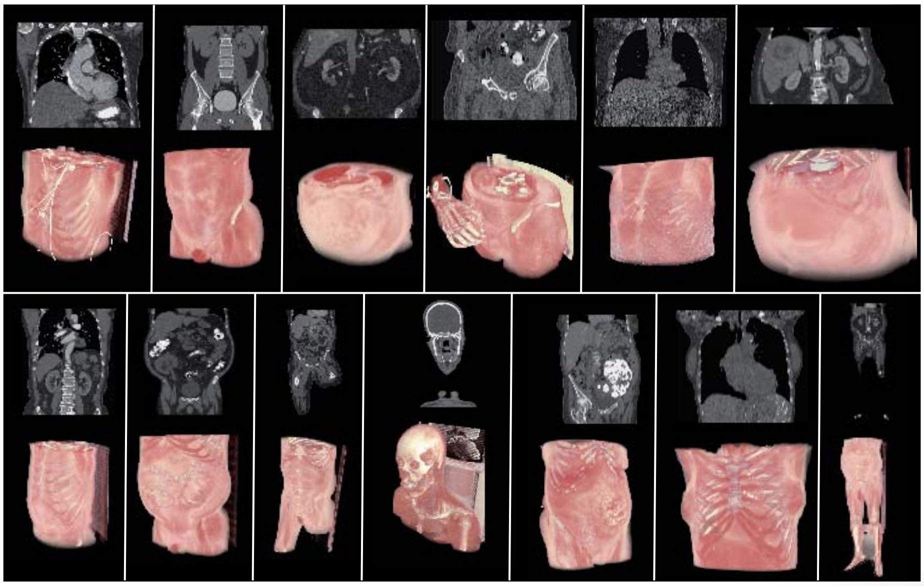
***Key idea: each voxel votes (probabilistically) for the position of each organ's bounding box.***

## Organ labelling: why is it hard?



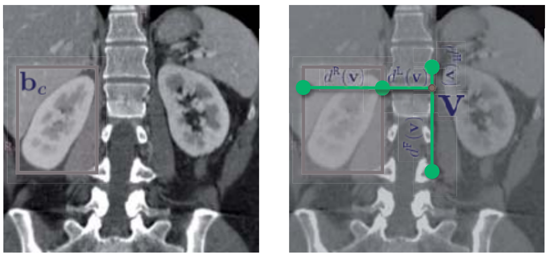

High variability in appearance, shape, location, resolution, noise, pathologies …

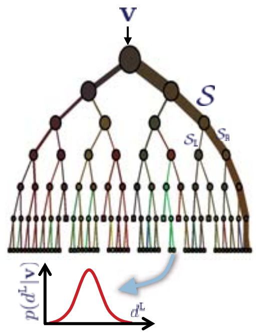
## Organ labelling: the ground-truth database



Different image cropping, noise, contrast/no-contrast, resolution, scanners, body shapes/sizes, patient position…

## Organ labelling: regression forest




- **Each voxel** in the volume votes for the position of the 6 box sides
- We wish to learn a set of **discriminative points** (landmarks, clusters) which can predict the kidney position with **high confidence.**

**Input data point** $\mathbf{v} = (v_x\ v_y\ v_z)$ *(voxel position in volume)*

**Output** $\mathbf{b}_c = (b_c^L\ b_c^R\ b_c^A\ b_c^P\ b_c^H\ b_c^F)$ *(bound. box continuous pos.)*

**Multiple organs** $c \in \{\texttt{liver}, \texttt{spleen}, \texttt{l.-kidney}, \texttt{r.-kidney}...\}$

**Node split function** $\xi_j > f(\mathbf{v};\theta_j) > \tau_j$

**Node optimization** $IG = \frac{1}{2}\left(\sum_c p(c;\mathcal{S}) \log|\Lambda_c(\mathcal{S})| - \sum_{i\in\{L,R\}} \omega_i \sum_c p(c;\mathcal{S}_i)\log|\Lambda_c(\mathcal{S}_i)|\right)$

**Node training** $(\theta^*, \xi^*, \tau^*) = \max_{\theta,\xi,\tau} IG$

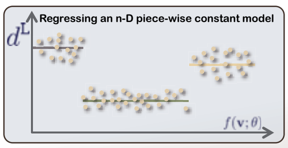
**Feature response** $f(\mathbf{v},\theta_j) = \frac{1}{|B_j|}\sum_{\mathbf{p}\in B_j} I(\mathbf{p})$ *(mean over displaced 3D boxes)*

*Error in model fit*

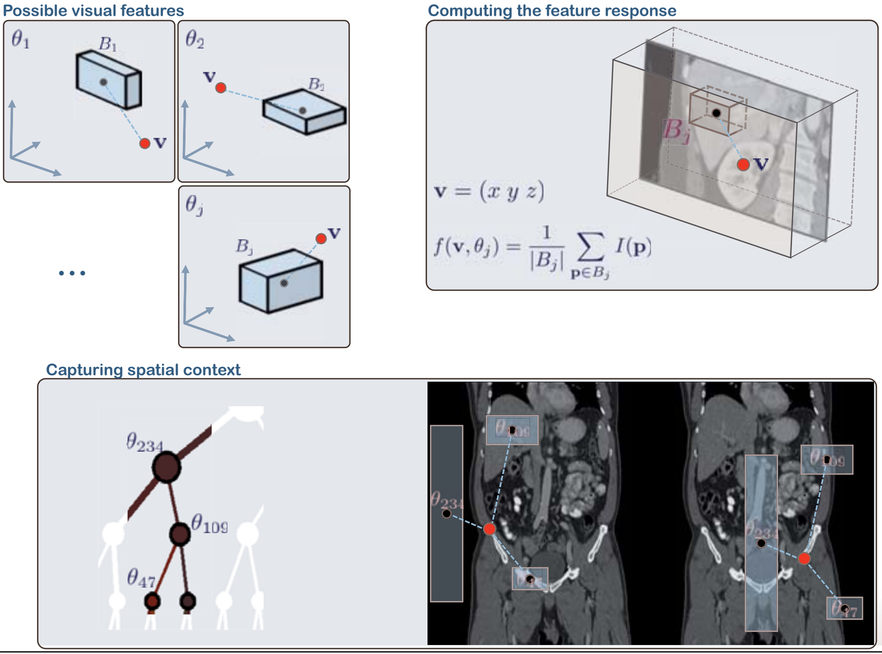
$\sum_c p(c;\mathcal{S}) \log|\Lambda_c(\mathcal{S})|$ *(weighted uncertainty for all organs)*

$\mathbf{d}_c(\mathbf{v}) = (d_c^L\ d_c^R\ d_c^A;\ d_c^P\ d_c^H;\ d_c^F)$ *(relative displacement)*

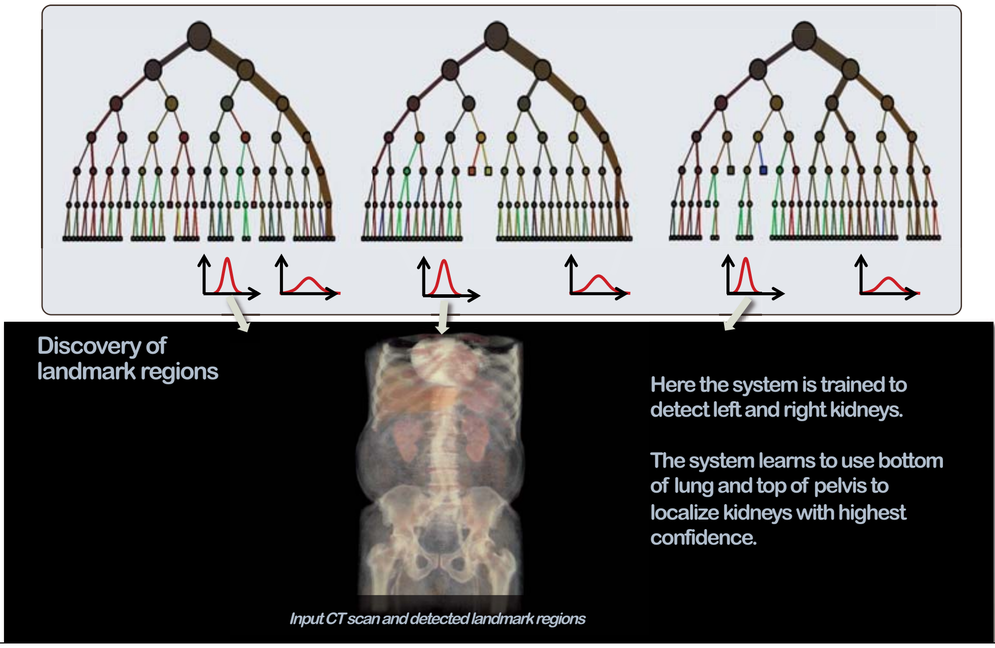
$p(\mathbf{d}) = \mathcal{N}(\mathbf{d}; \overline{\mathbf{d}}, \Lambda)$ *(Gaussian repres. of distribs)*



## Organ labelling: context-rich visual features




## Organ labelling: automatic landmark discovery

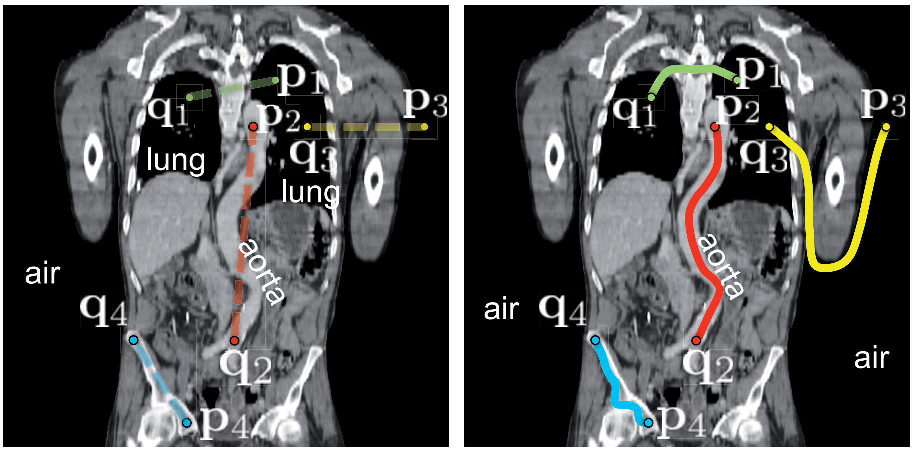


# Overview

- A brief introduction to machine learning
- Decision forests and jungles
- **Applications in medical image analysis**
  - Anatomy localization
  - **Anatomy segmentation**
  - Spine detection
  - Brain tumour segmentation
  - Learned super-resolution

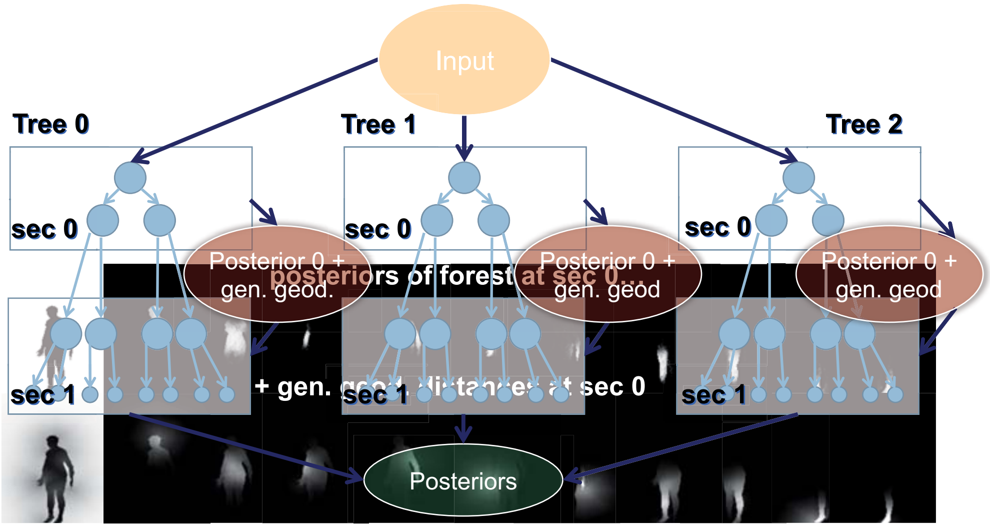
P. Kontschieder, P. Kohli, J. Shotton, and A. Criminisi, **GeoF: Geodesic Forests for Learning Coupled Predictors**, in *Proc. Computer Vision and Pattern Recognition (CVPR)*, IEEE, June 2013

# Entangled geodesic forests for semantic segmentation




- Using soft connectivity features efficiently
- Capturing semantic context
- No need for Markov-, Conditional Random Field post-processinf

# Entangled geodesic forests for semantic segmentation




# Entangled geodesic forests for semantic segmentation

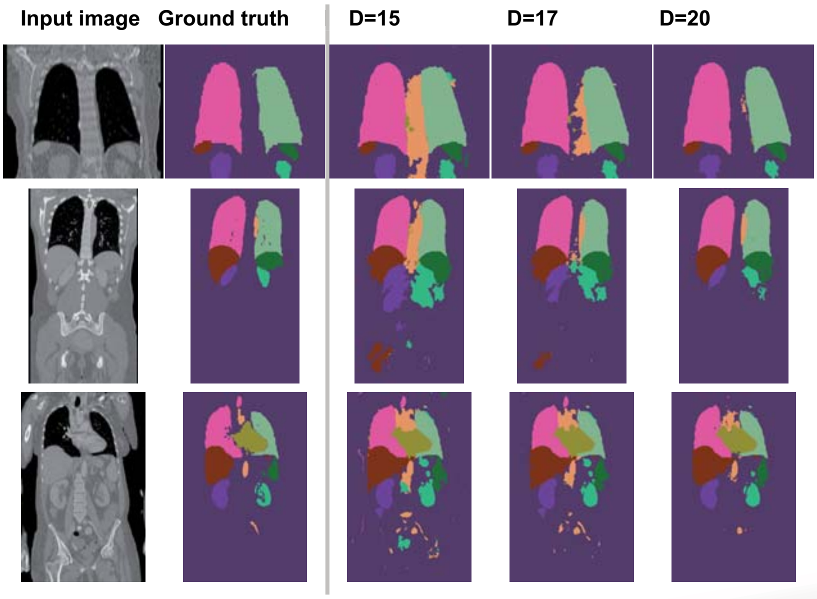
| Algorithm | Jaccard |
|---|---|
| | |
| Conventional Classification Forest | 53.2 |
| Classification forest + (CRF) | 68.3 |
| Auto-context classification forest | 65.9 |
| Entangled classification forest | 58.3 |
| Auto-context geodesic forests | 69.2 |
| | |
| **Entangled geodesic forests** | **72.3** |
| | |
| | |

Input

Ground truth

Our result

## Entangled geodesic forests for semantic segmentation

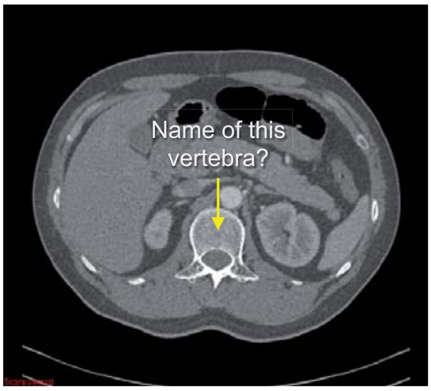


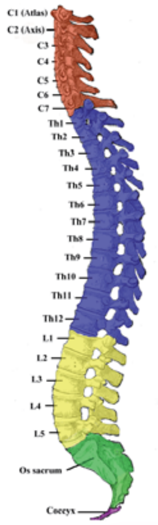

## Overview

- A brief introduction to machine learning
- Decision forests and jungles
- **Applications in medical image analysis**
  - Anatomy localization
  - Anatomy segmentation
  - **Spine detection**
  - Brain tumour segmentation
  - Learned super-resolution

B. Glocker, D. Zikic, E. Konukoglu, D. R. Haynor, and A. Criminisi, **Vertebrae Localization in Pathological Spine CT via Dense Classification from Sparse Annotations**, in *MICCAI 2013*, Springer, September 2013

# Vertebrae Localization and Classification

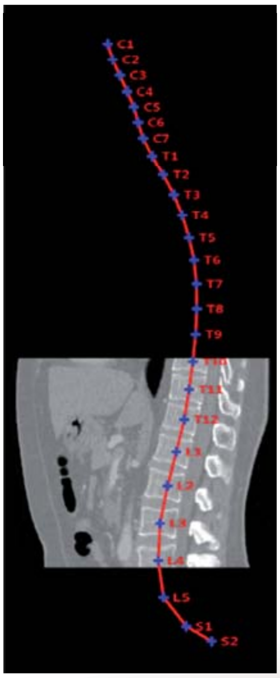



# Clinical motivation

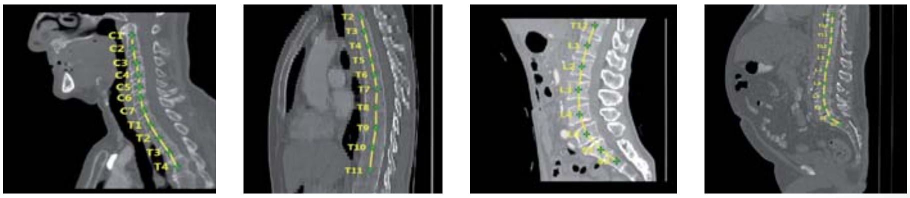
### Patient-specific coordinate system

- Guided visualization/navigation in diagnostic tools
- Longitudinal assessment after surgical Intervention
- Shape/population analysis for disease modelling

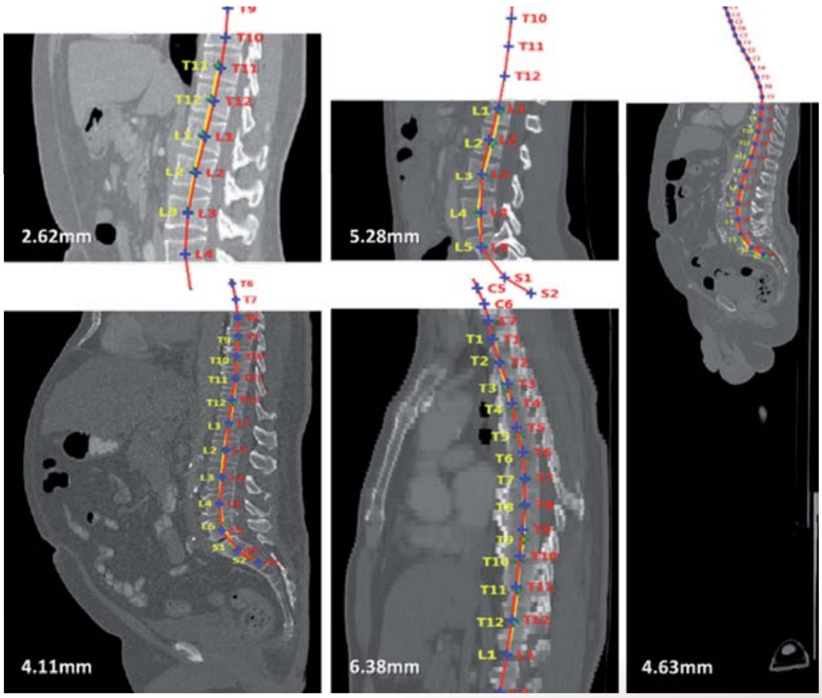


# Challenges

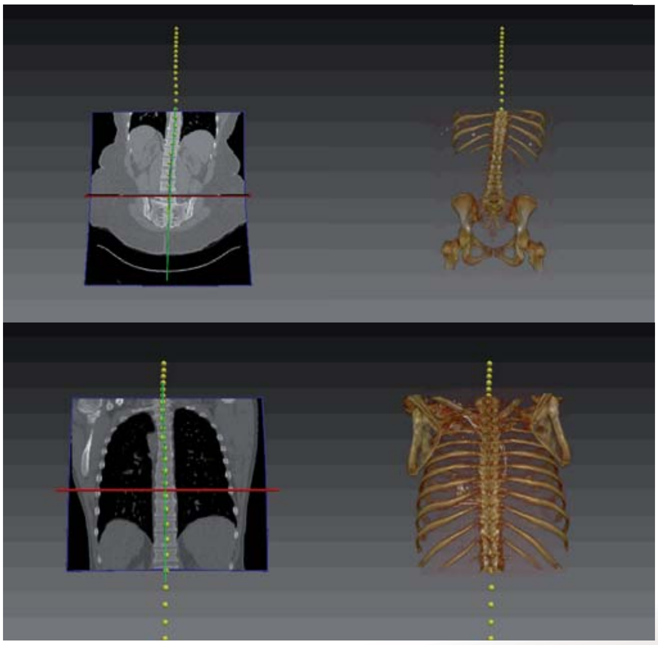
- Repetitive nature of structures
- Variability of normal anatomy
- Presence of pathologies
- Varying image acquisition (FOV, noise level, resolution, …)



# Some results



# Some results

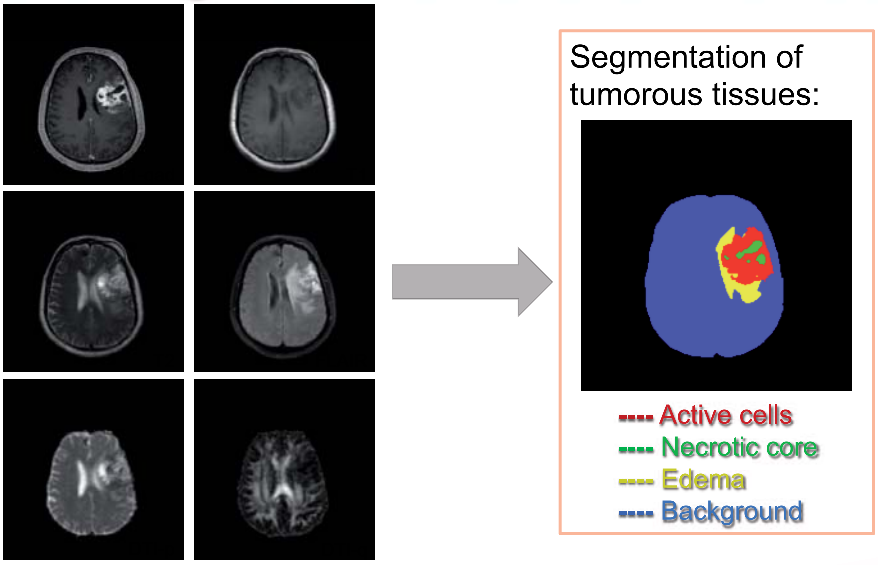


# Overview

- A brief introduction to machine learning
- Decision forests and jungles
- **Applications in medical image analysis**
  - Anatomy localization in CT scans
  - Anatomy segmentation in CT scans
  - Spine detection in CT scans
  - **Brain tumour segmentation in MR scans**
  - Learned super-resolution in diffusion MRI

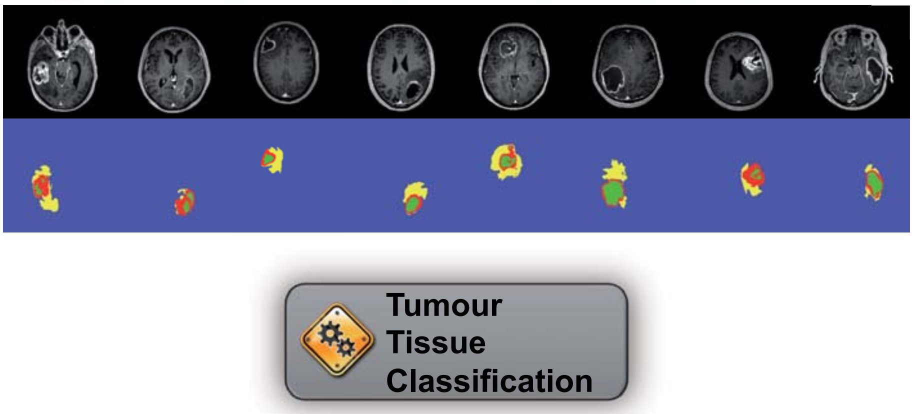
D. Zikic, B. Glocker, E. Konukoglu, A. Criminisi, J. Shotton, C. Demiralp, O. Thomas, T. Das, R. Jena, and S. Price, **Decision Forests for Tissue-specific Segmentation of High-grade Gliomas in Multi-channel MR**, in *MICCAI 2012*, Springer, October 2012

# Automatic Segmentation of Brain Tumour

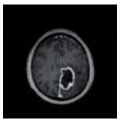



3D MRI input data

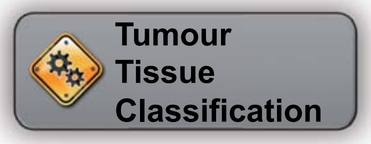
# Training a Pixel-Wise Forest Classifier



# Testing the Pixel-Wise Forest Classifier

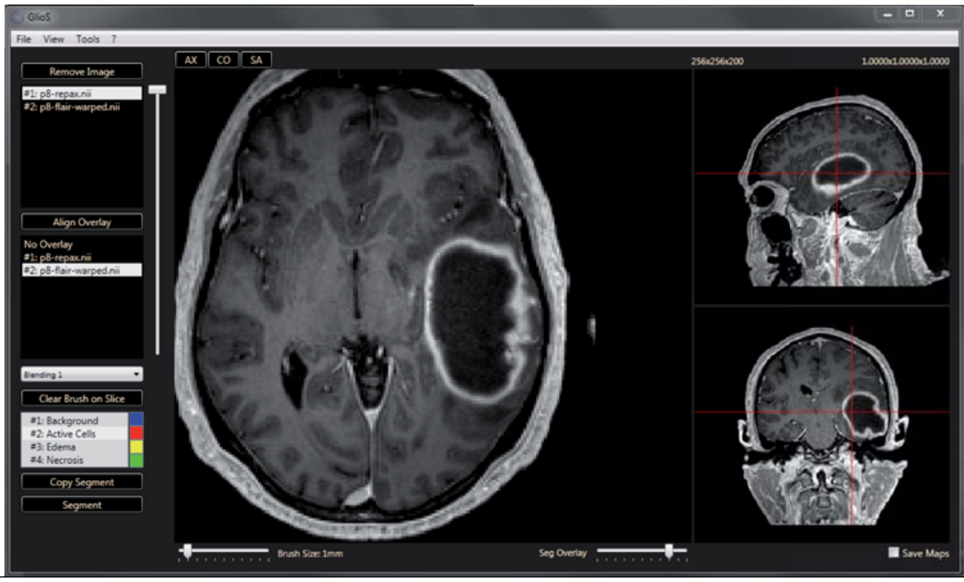


New Patient,
previously unseen



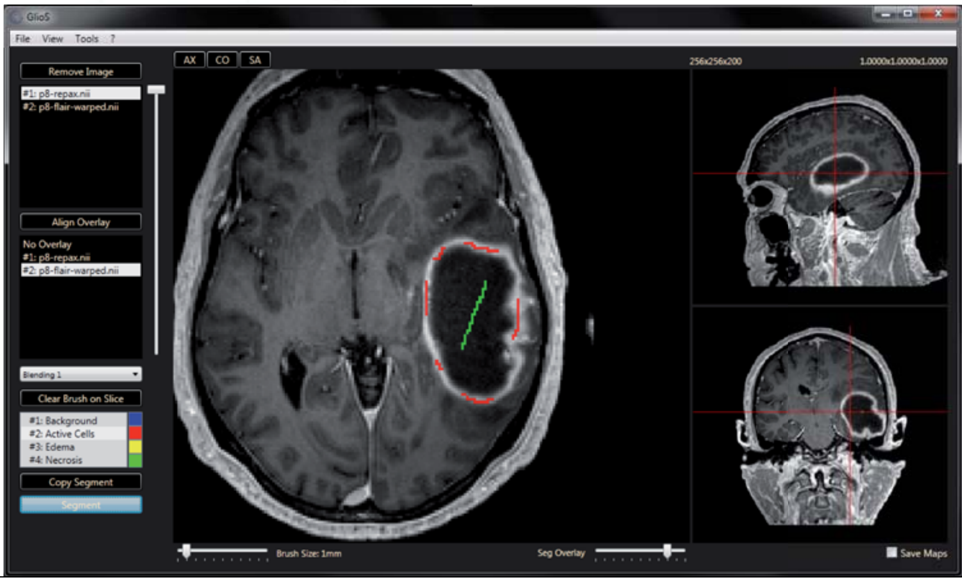

# Building the Training Database of Patients' Images

1st Step: Obtain Expert Segmentation



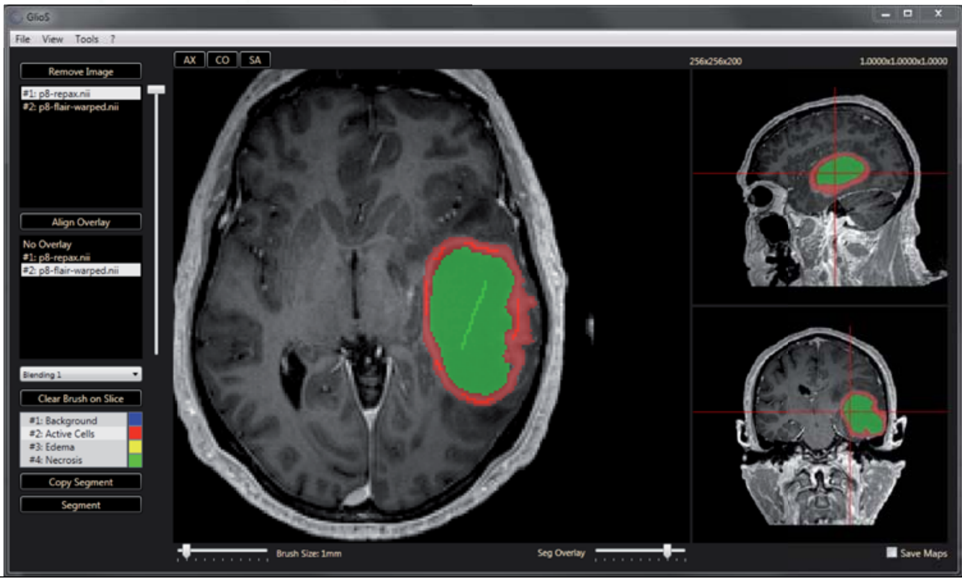
# Building the Training Database of Patients' Images

1st Step: Obtain Expert Segmentation

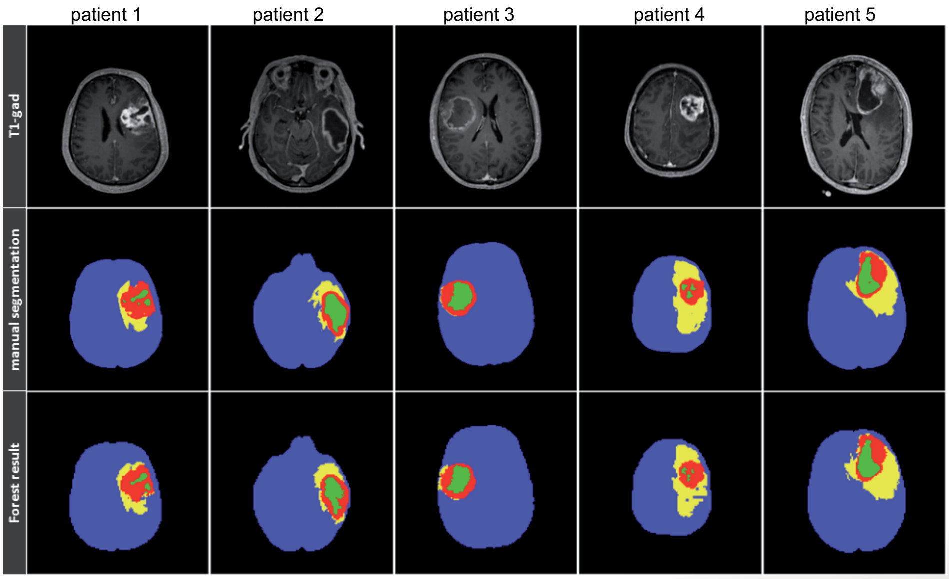



# Building the Training Database of Patients' Images

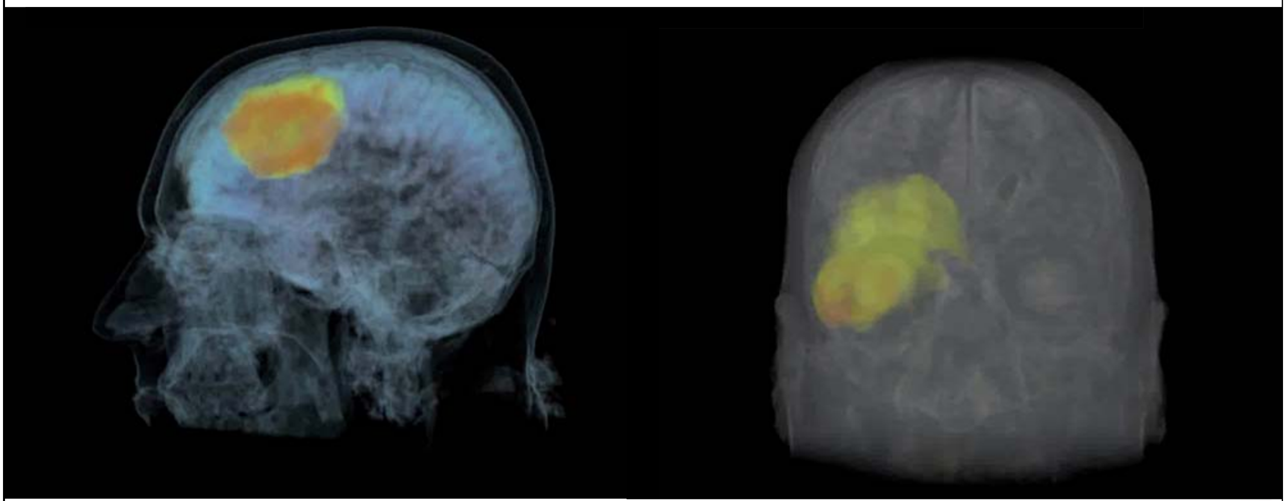
1st Step: Obtain Expert Segmentation



# Glioblastoma Segmentation




# Glioblastoma Segmentation: results

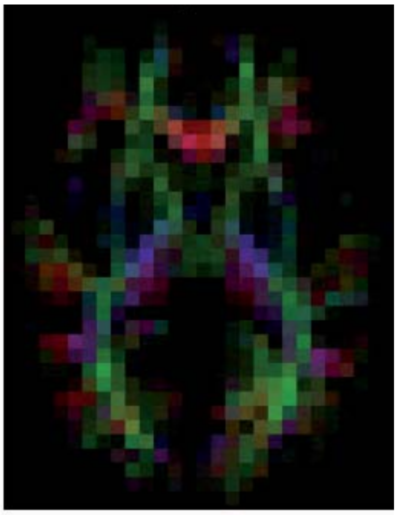


# Overview

- A brief introduction to machine learning
- Decision forests and jungles
- **Applications in medical image analysis**
  - Anatomy localization
  - Anatomy segmentation
  - Spine detection
  - Brain tumour segmentation
  - **Learned super-resolution**

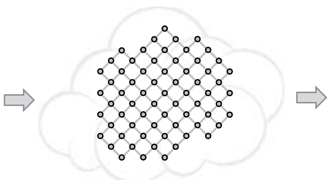
D. Alexander, D. Zikic, J. Zhang, H. Zhang, and A. Criminisi, **Image Quality Transfer via Random Forest Regression: Applications in Diffusion MRI**, in *MICCAI 2014 - Intl Conf. on Medical Image Computing and Computer Assisted Intervention*, Springer, 2014

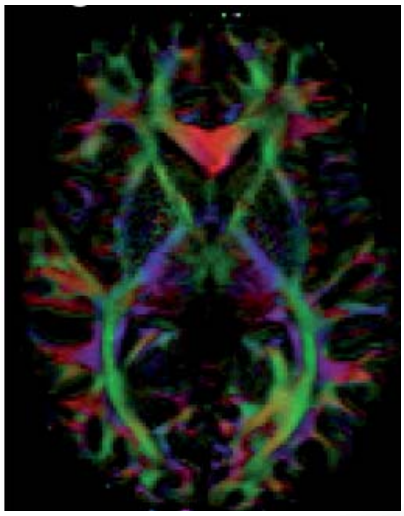
# Learned image super-resolution



*Low-res diffusion MRI*
*(faster acquisition, cheaper)*

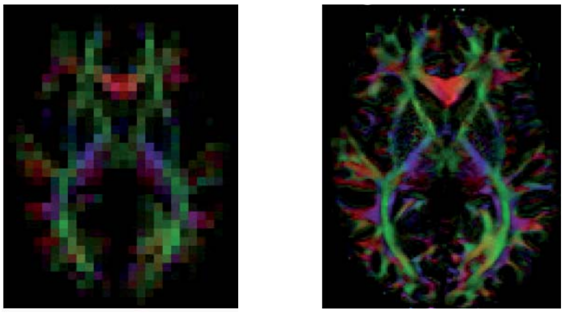
***Learned voxel predictor***

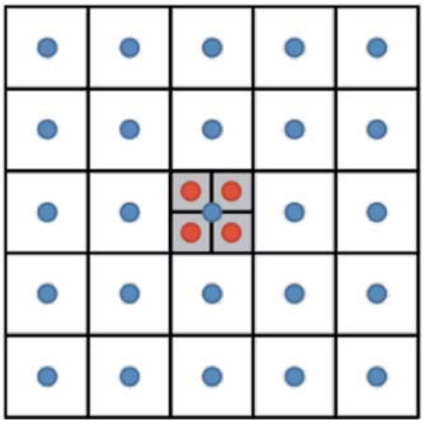




*High-res diffusion MRI*

# Learned image super-resolution



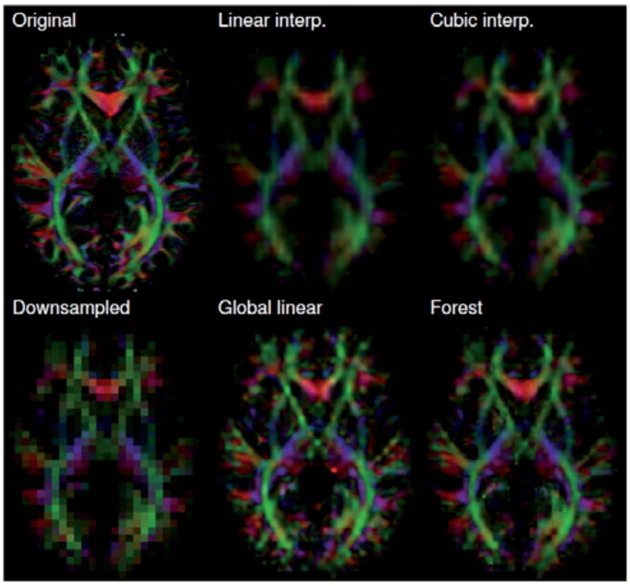


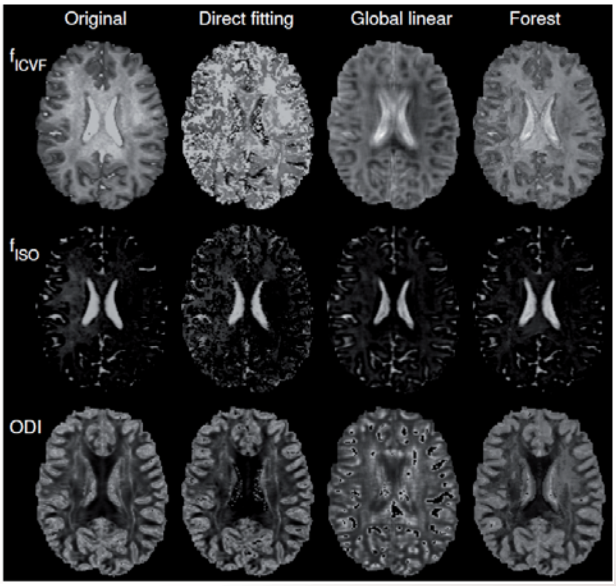
***Goal***

learning to predict the value of the **high-res voxels** from the **low-res voxels.**

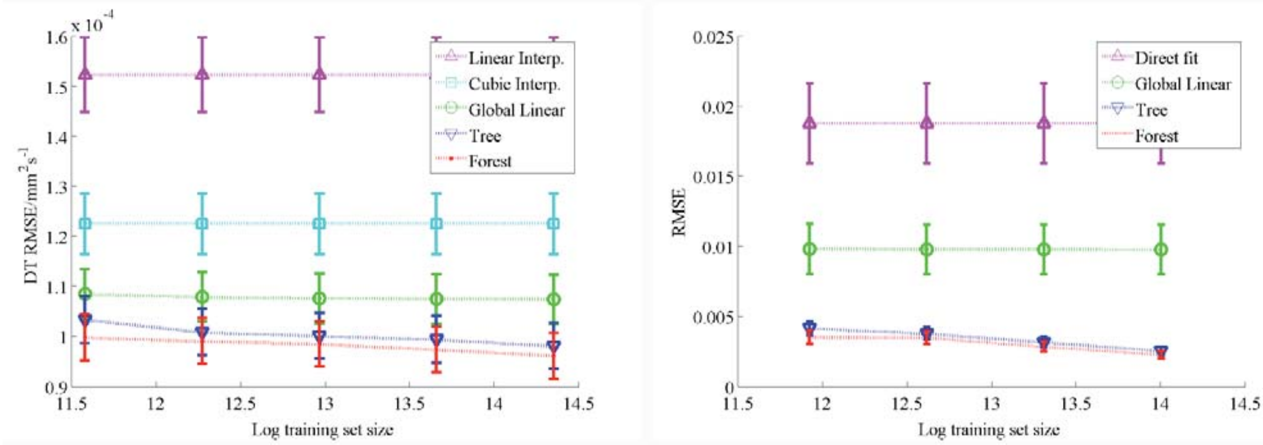
- Training data can be easily obtained
- Well defined accuracy measure

# Learned image super-resolution



# Learned image super-resolution




D. Alexander, D. Zikic, J. Zhang, H. Zhang, and A. Criminisi, **Image Quality Transfer via Random Forest Regression: Applications in Diffusion MRI**, in *MICCAI 2014 - Intl Conf. on Medical Image Computing and Computer Assisted Intervention*, Springer, 2014

Modern, efficient machine learning has the potential to revolutionize medicine!