

# A Review on Automatic Image Annotation Techniques

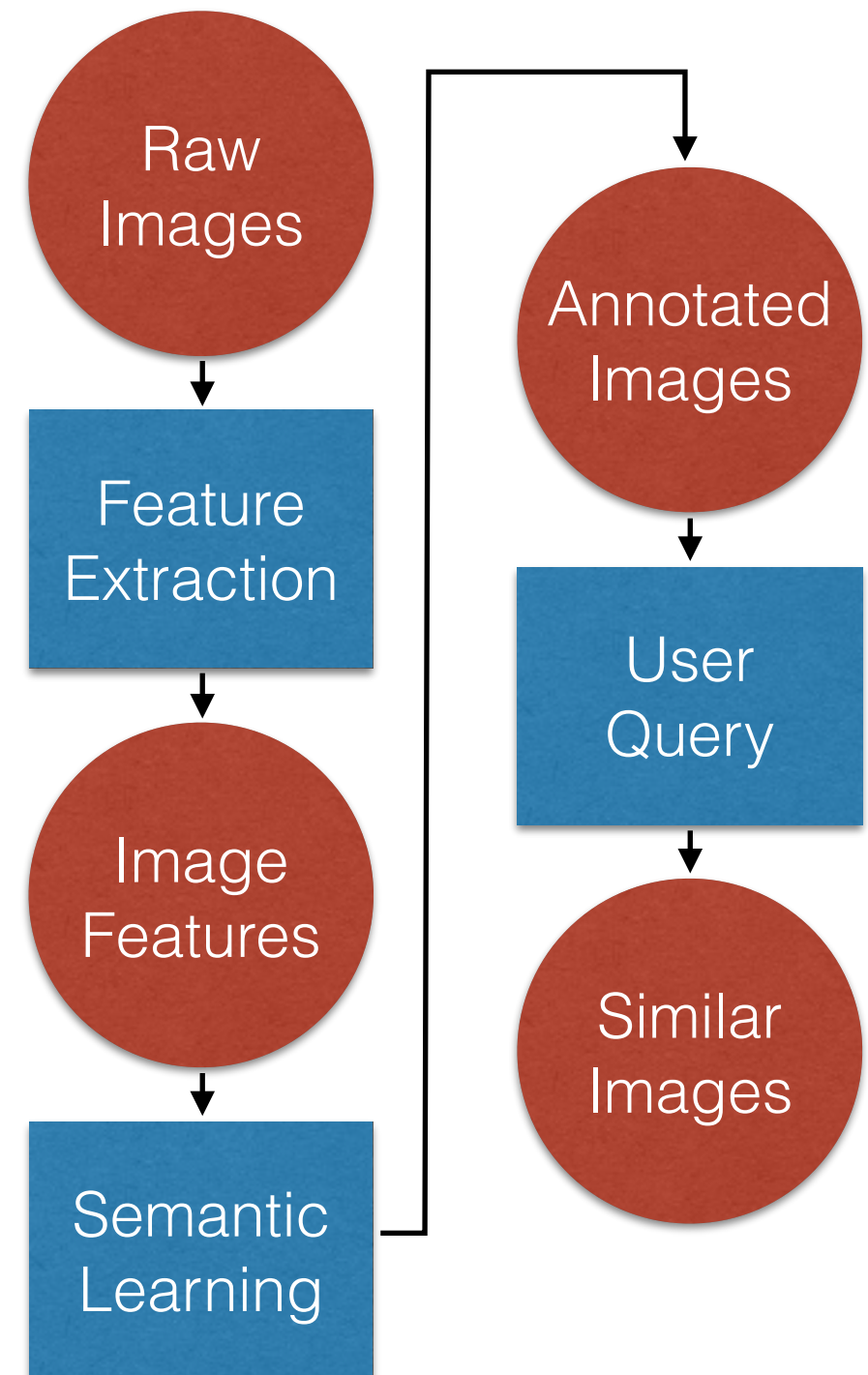
Samuel Jackson (slj11)

# What is AIA?

- In a nut shell: labelling images correctly!
- More formally: Trying to find images that are semantically similar to our input query.
- The key challenges: automation, higher level meaning.

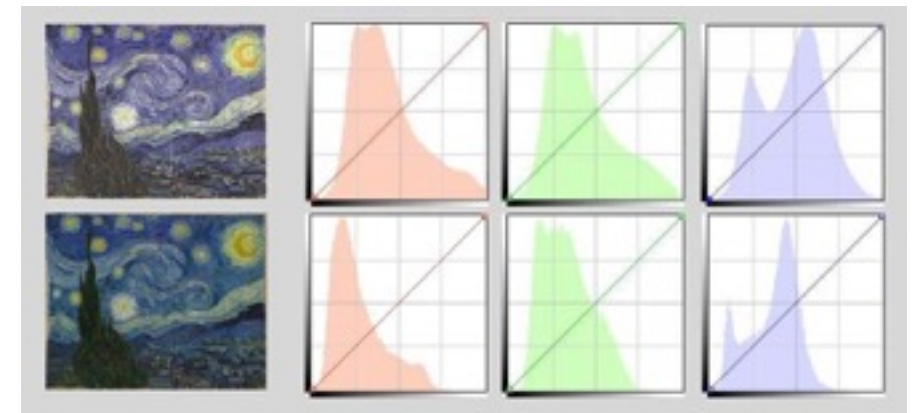
# What Do I Need to Do AIA?

- Feature extraction
  - Segmentation
  - Global vs. Local
  - Colour, Texture, Shape based...
- Semantic learning technique
  - Single label
  - Multi-Labeling
  - Metadata

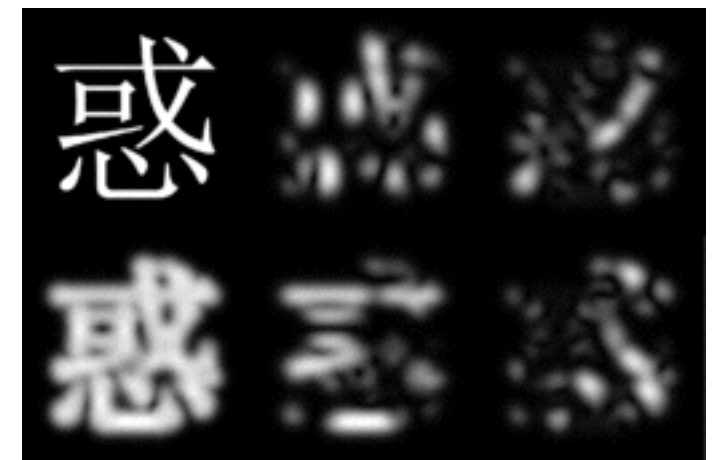


# Feature Extraction

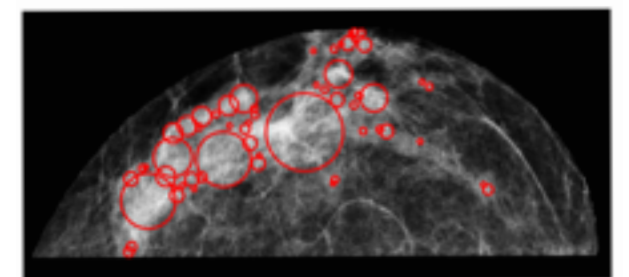
- Low level representation of what characterises the image
- Just a few different approaches:
  - Colour: Moments, Histogram, Correlogram
  - Texture: Texton, GLCM, Gabor
  - Shape: Area, Moments, Circularity
- Global vs. Local?



Colour Histogram of van Gogh's *Starry Night*



Gabor filtered Chinese text



Shape features from mammogram

# Semantic Learning

- Once we've acquired low level features we can use these to represent the image (assumption!)
- Features used as input to a machine learning algorithm.

# Single Label Annotation

- Attempt to assign an image to a single concept/class
- Often uses pretty simple/well known machine learning methods.

# Support Vector Machines

- Finds optimal class boundary
- Train separate SVM for each class.
- Can use multiple groups of classifiers + confidence factor
- Cons: Class imbalance issues, prone to overfitting

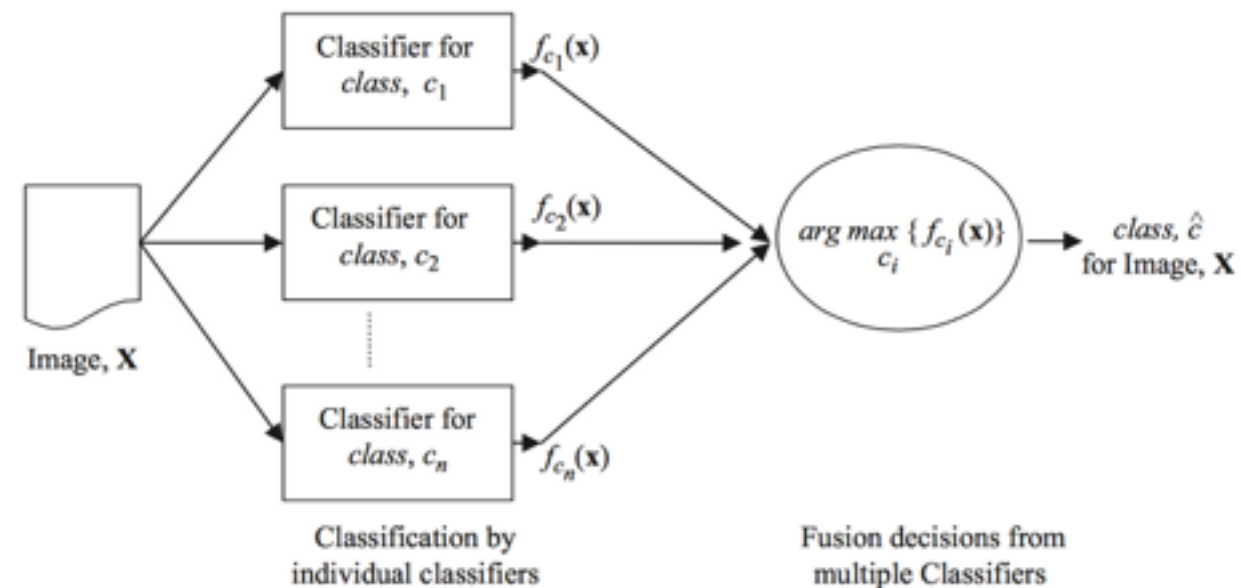


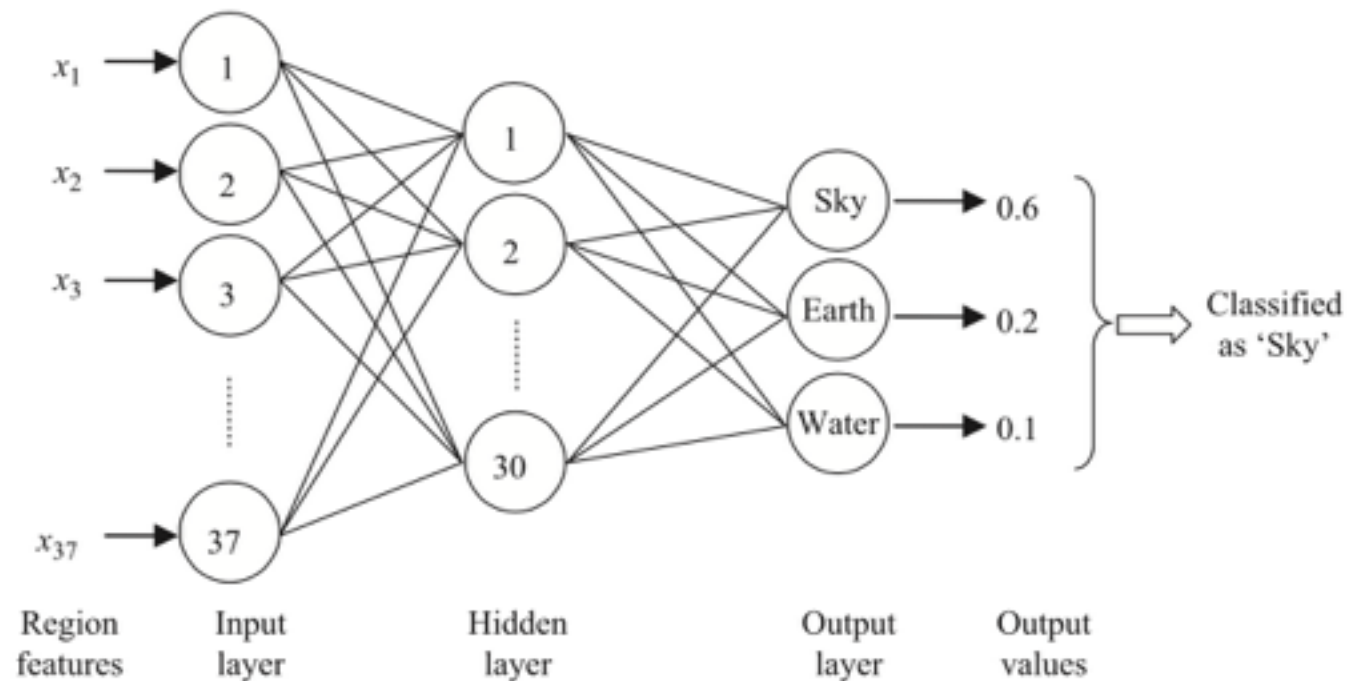
Fig. 2. Multiclass classifier using multiple binary SVM classifiers.



Fig. 3. Image annotation with multiple sets of SVMs.

# Neural Networks

- Features form input layer
- Output gives probabilistic estimate of class.
- Cons: Black box, Number of layers?



**Fig. 5.** Classifying a region using ANN.



# Decision Trees

- Grow tree which splits training data by features
- Easy to interpret the decision
- Cons: Discretisation is an issue, Need to prune

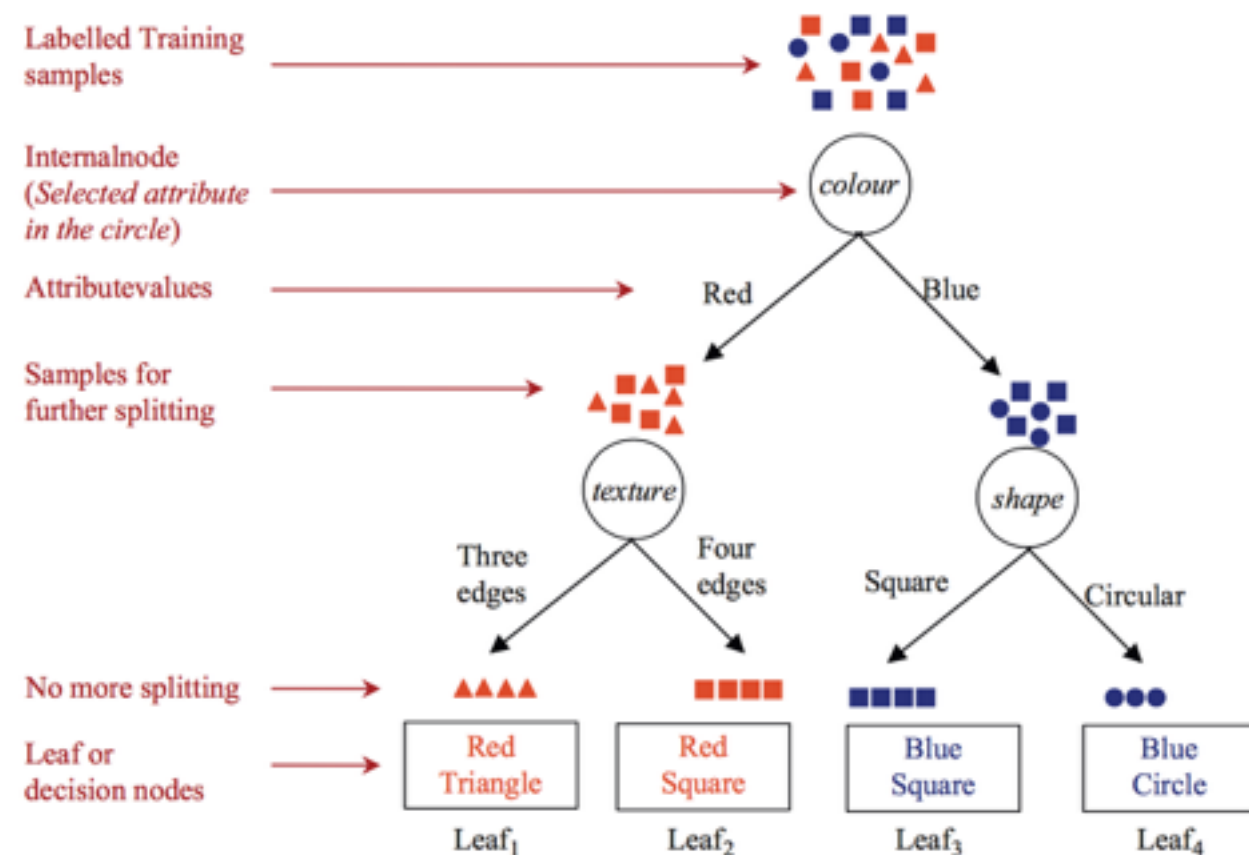


Fig. 6. Decision tree learning.

# Interlude: What Should We Label This?



Car, Lake, Road, Grass, Hills, Sky, Wales, Landscape?  
Google Photos says: Mountains and Car

# Multi-Labellabelling

- Label images with multiple semantic concepts.
- Typically formulated as a probabilistic Bayesian learning problem
- Two major types:
  - Non-Parametric
  - Parametric

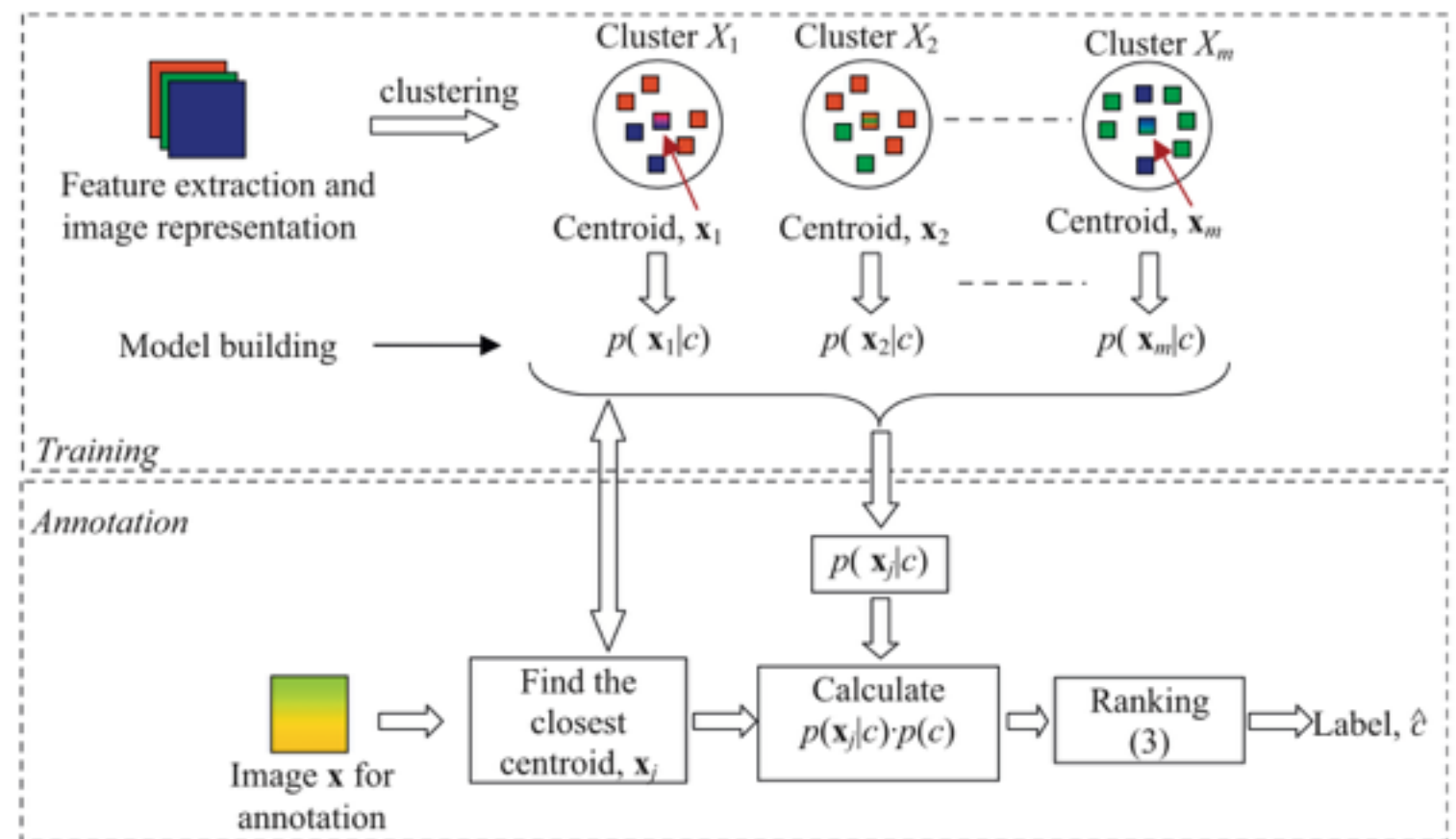


Fig. 7. The general Bayesian annotation model.

# Parametric vs. Non-Parametric

- Conditional probabilities are calculated without prior assumption
  - Pros:
    - Model free
    - Quick to train
  - Cons:
    - Lots of Parameters
    - Sensitive to noise
- Features space assumed to follow a known distribution (often Gaussian)
  - Pros:
    - Good approximation of unknown distribution
    - Small samples
  - Cons:
    - Predefined distribution
    - Expensive training

# Metadata

- But isn't all this image processing is computationally expensive?
- Use content from HTML, URLs, text etc.
- Can use hybrid approach
- Issues:
  - Often noisy/unreliable content
  - The “semantic gap”
  - Lexicon size
  - Mapping to visual features



Kangaroo or Australia?

Image source: Sinead Friel <http://www.flickr.com/people/53005672@N04>

# Final Thoughts

- Generally a well written & interesting paper
- Covers an awful lot of content (perhaps too much?)
- Gives the reader a good flavour of the methods
- Perhaps out of date? Are Deep NNs used more now?

# Thank You

Any Questions?