Interactive Machine Learning for Automated Image Annotation: What can machines and specialists learn from each other?

Alexandre Xavier Falcão

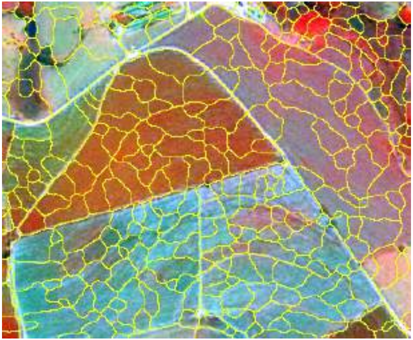
Laboratory of Image Data Science
Institute of Computing — University of Campinas

afalcao@ic.unicamp.br

What is this all about?

Consider the problem of assigning one label to each sample in a given image, where a sample may be a pixel,

- superpixel (connected region with a same texture),
- object (connected region with known shape, usually), or
- subimage (region around some object of interest).



What is this all about?

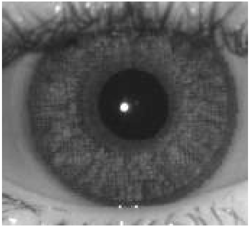
Consider the problem of assigning one label to each sample in a given image, where a sample may be a pixel,

- superpixel (connected region with a same texture),
- object (connected region with known shape, usually), or
- subimage (region around some object of interest).

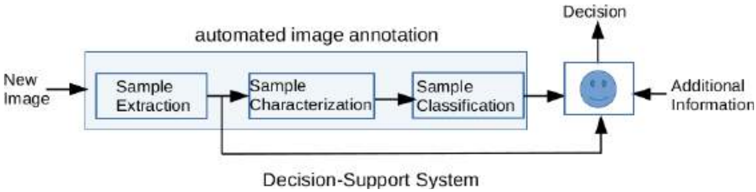
What is this all about?

Consider the problem of assigning one label to each sample in a given image, where a sample may be a pixel,

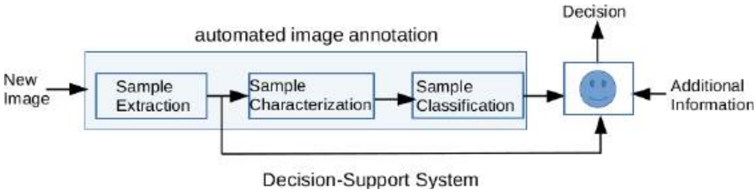
- superpixel (connected region with a same texture),
- object (connected region with known shape, usually), or
- subimage (region around some object of interest).



What is this all about?

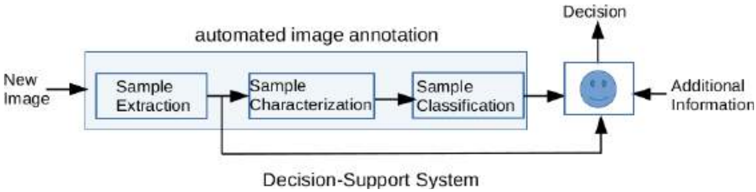


What is this all about?



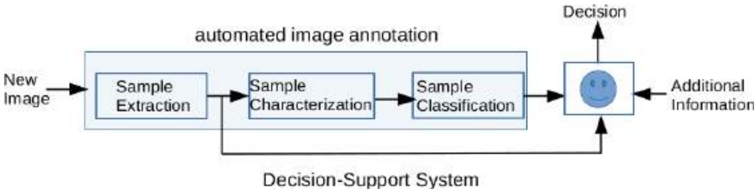
- Machines can learn how to annotate images, but

What is this all about?



- Machines can learn how to annotate images, but
- they require a training set with manually isolated and identified samples.

What is this all about?



- Machines can learn how to annotate images, but
- they require a training set with manually isolated and identified samples.
- As image databases grow large, the procedure becomes infeasible, especially when it requires specialists.

What is this all about?

The prior annotation strategy also leaves unanswered several important questions.

What is this all about?

The prior annotation strategy also leaves unanswered several important questions.

- How many training samples are needed? Which ones are the most effective?

What is this all about?

The prior annotation strategy also leaves unanswered several important questions.

- How many training samples are needed? Which ones are the most effective?
- Can the machines learn from their errors? Are all errors helpful for machine learning?

What is this all about?

The prior annotation strategy also leaves unanswered several important questions.

- How many training samples are needed? Which ones are the most effective?
- Can the machines learn from their errors? Are all errors helpful for machine learning?
- Can the specialists understand the behavior of the machines, explain their actions, and trust on their decisions?

What is this all about?

The prior annotation strategy also leaves unanswered several important questions.

- How many training samples are needed? Which ones are the most effective?
- Can the machines learn from their errors? Are all errors helpful for machine learning?
- Can the specialists understand the behavior of the machines, explain their actions, and trust on their decisions?
- What can machines and specialists learn from each other?

Purpose of the lecture

This lecture

- motivates a change from the traditional to the interactive machine learning paradigm,

Purpose of the lecture

This lecture

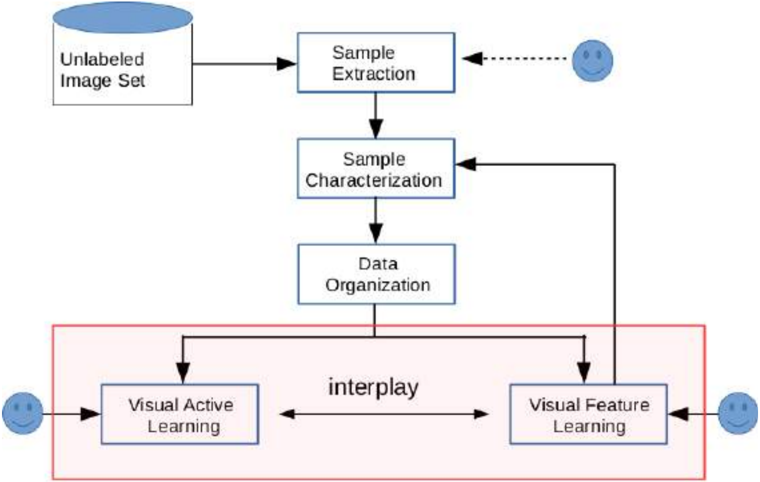
- motivates a change from the traditional to the interactive machine learning paradigm,
- integrates research from several subjects in Visual Computing to address the problem, and

Purpose of the lecture

This lecture

- motivates a change from the traditional to the interactive machine learning paradigm,
- integrates research from several subjects in Visual Computing to address the problem, and
- presents a methodology that exploits
  - the superior abilities of humans in knowledge abstraction and
  - the higher capacity of machines for data processing.

Methodology for interactive machine learning



Interactive Machine Learning (IML)

Organization

The lecture is organized as follows.

- Active modeling of object shapes for sample extraction.

Organization

The lecture is organized as follows.

- Active modeling of object shapes for sample extraction.
- Visual feature learning for sample characterization.

Organization

The lecture is organized as follows.

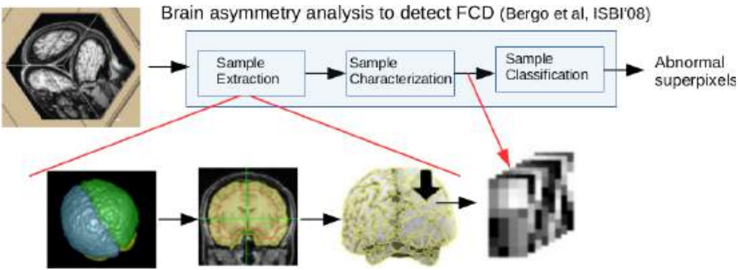
- Active modeling of object shapes for sample extraction.
- Visual feature learning for sample characterization.
- Visual active learning for sample classification.

Organization

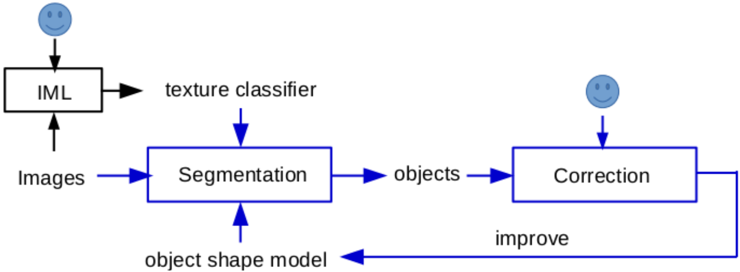
The lecture is organized as follows.

- Active modeling of object shapes for sample extraction.
- Visual feature learning for sample characterization.
- Visual active learning for sample classification.

Example:

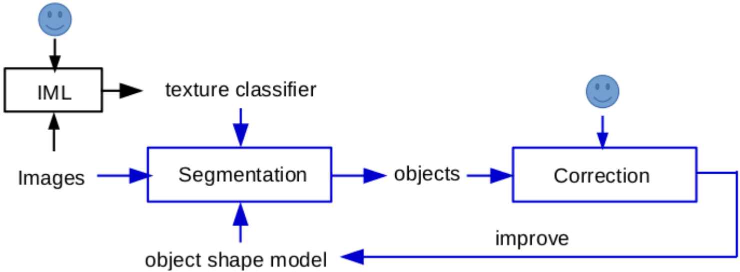


Objects as samples: shape models can be learned from a few examples, applied, and improved by interactive segmentation corrections.



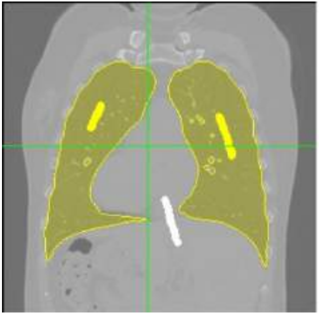
Active modeling of object shapes

Objects as samples: shape models can be learned from a few examples, applied, and improved by interactive segmentation corrections.



The shape and texture models aim at substituting humans in image segmentation.

Image segmentation consists of recognition (humans $\gg$ machines) and delineation (machines $\gg$ humans) tasks.



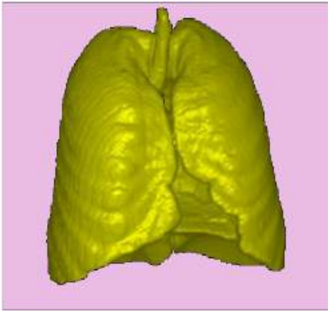
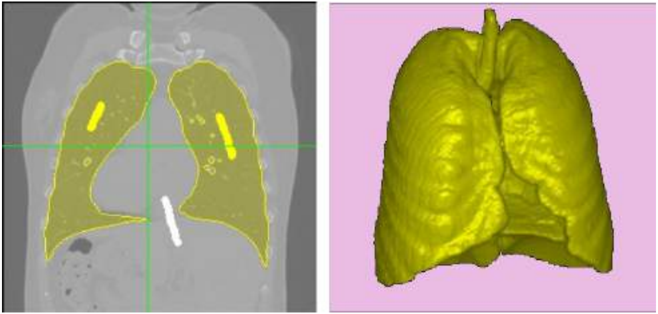


Image segmentation

Image segmentation consists of recognition (humans $\gg$ machines) and delineation (machines $\gg$ humans) tasks.

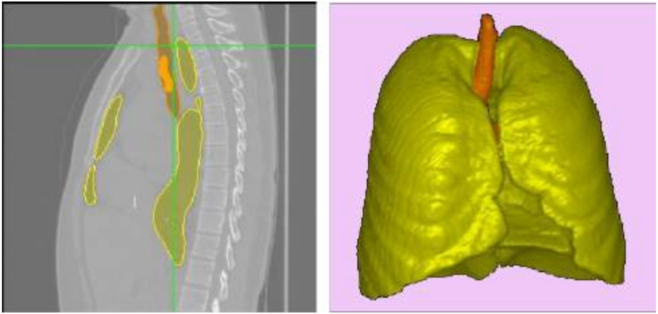


The object is an optimum-path forest rooted at its internal markers

(Falcão et al, IEEE TPAMI'04, Falcão and Bergo, IEEE TMI'04, Miranda et al JMIV'09, Ciesielski et al, JMIV'12, Mansilla et al, SIBGRAPI'16).

Image segmentation

Image segmentation consists of recognition (humans $\gg$ machines) and delineation (machines $\gg$ humans) tasks.

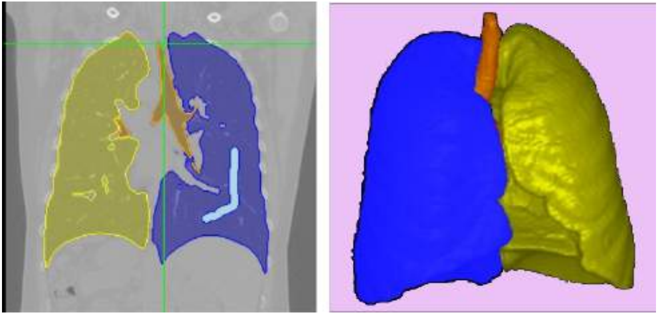


The object is an optimum-path forest rooted at its internal markers

(Falcão et al, IEEE TPAMI'04, Falcão and Bergo, IEEE TMI'04, Miranda et al JMIV'09, Ciesielski et al, JMIV'12, Mansilla et al, SIBGRAPI'16).

Image segmentation

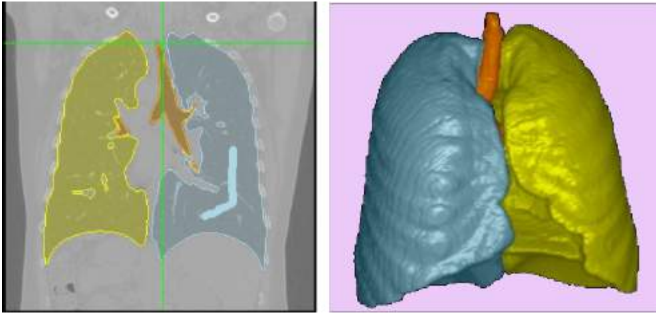
Image segmentation consists of recognition (humans $\gg$ machines) and delineation (machines $\gg$ humans) tasks.



The object is an optimum-path forest rooted at its internal markers

(Falcão et al, IEEE TPAMI'04, Falcão and Bergo, IEEE TMI'04, Miranda et al JMIV'09, Ciesielski et al, JMIV'12, Mansilla et al, SIBGRAPI'16).

Image segmentation consists of recognition (humans $\gg$ machines) and delineation (machines $\gg$ humans) tasks.

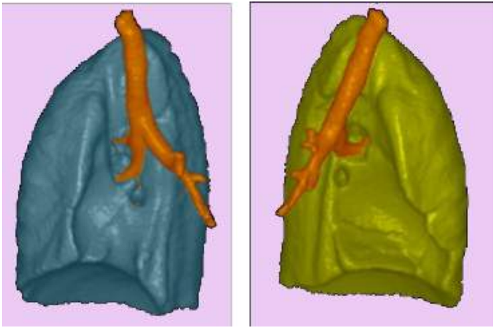


The object is an optimum-path forest rooted at its internal markers

(Falcão et al, IEEE TPAMI'04, Falcão and Bergo, IEEE TMI'04, Miranda et al JMIV'09, Ciesielski et al, JMIV'12, Mansilla et al, SIBGRAPI'16).

Image segmentation

Image segmentation consists of recognition (humans $\gg$ machines) and delineation (machines $\gg$ humans) tasks.



The object is an optimum-path forest rooted at its internal markers

(Falcão et al, IEEE TPAMI'04, Falcão and Bergo, IEEE TMI'04, Miranda et al JMIV'09, Ciesielski et al, JMIV'12, Mansilla et al, SIBGRAPI'16).

Image segmentation

How can texture and shape models substitute a human in object localization?

Image segmentation

How can texture and shape models substitute a human in object localization?

- A shape model estimates markers, while it translates and imposes shape constraints for object delineation.

Image segmentation

How can texture and shape models substitute a human in object localization?

- A shape model estimates markers, while it translates and imposes shape constraints for object delineation.
- A criterion function selects the object at the location of maximum delineation score.

How can texture and shape models substitute a human in object localization?

- A shape model estimates markers, while it translates and imposes shape constraints for object delineation.
- A criterion function selects the object at the location of maximum delineation score.

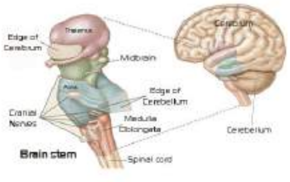
How can texture and shape models substitute a human in object localization?

- A shape model estimates markers, while it translates and imposes shape constraints for object delineation.

- A criterion function selects the object at the location of maximum delineation score.

A texture classifier can reduce candidate locations and avoid wrong marker selection.

Result of model-based segmentation of the brain hemispheres and cerebellum without the brain stem.



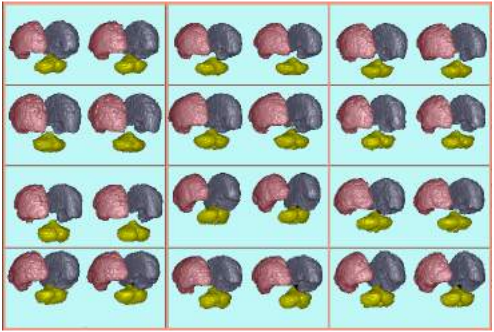
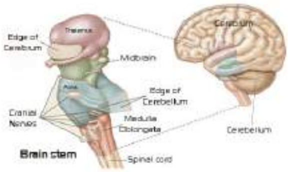
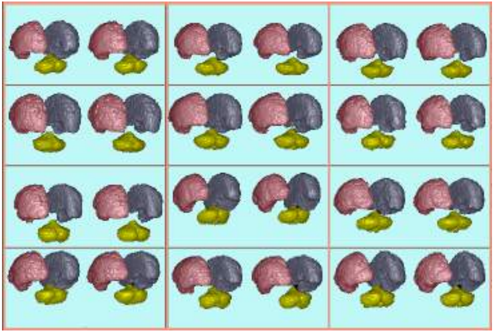


Image segmentation

Result of model-based segmentation of the brain hemispheres and cerebellum without the brain stem.

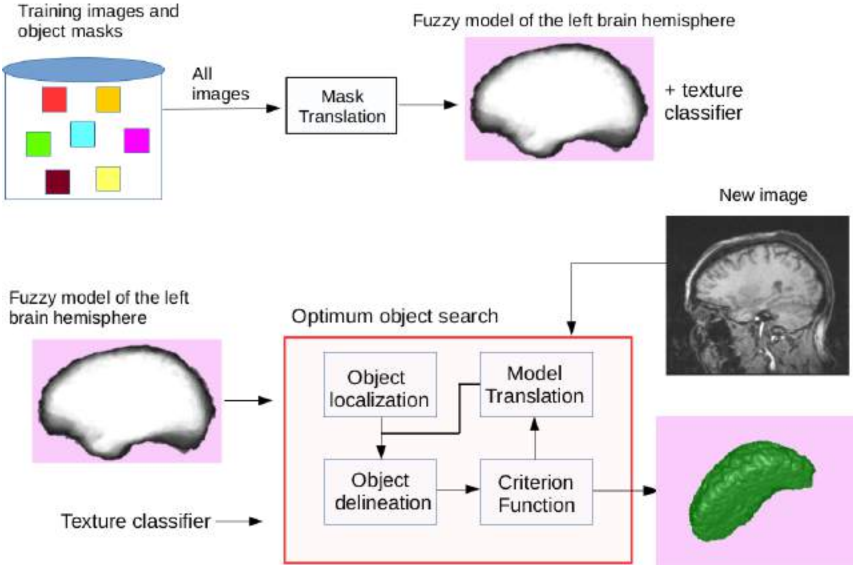




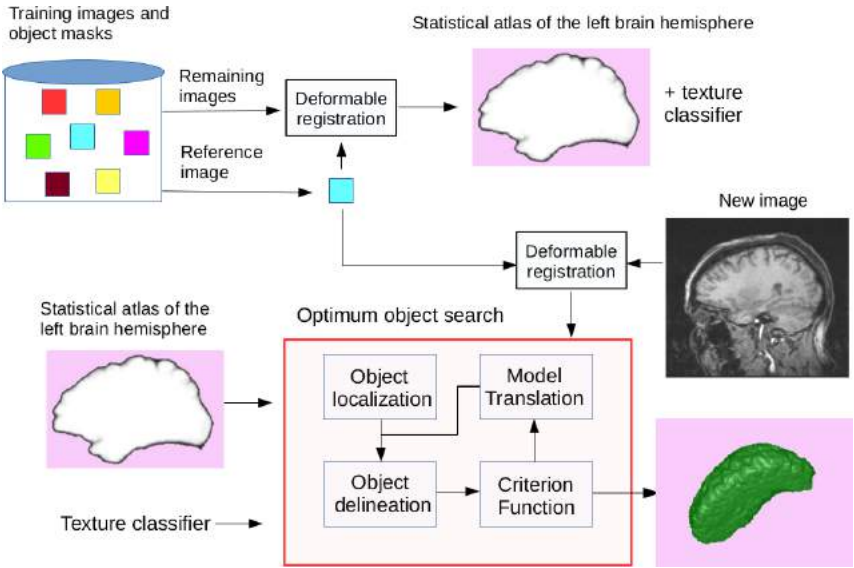
We have developed two types of shape models under this segmentation paradigm, fuzzy models and statistical atlases

(Miranda et al, ISBI09, Udupa et al, MEDIA'14, Phellan et al, Medical Physics'16, Spina et al, SIBGRAPI'16).

Fuzzy object shape models: construction and use

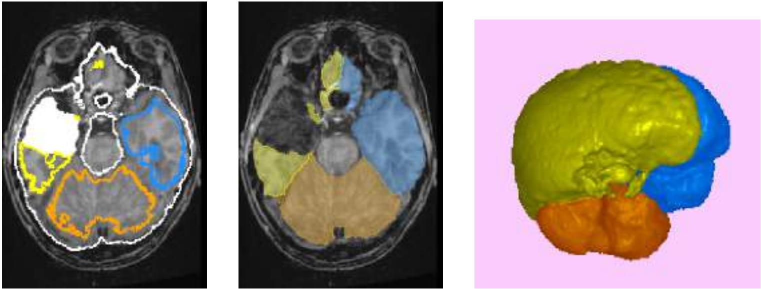


Statistical atlases: construction and use



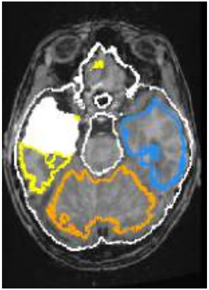
The role of the texture classifier

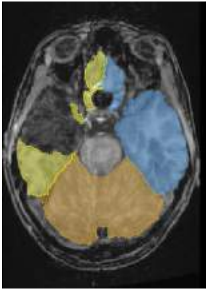
Abnormal brain: the texture classifier avoids internal markers in surgically removed regions of the brain.

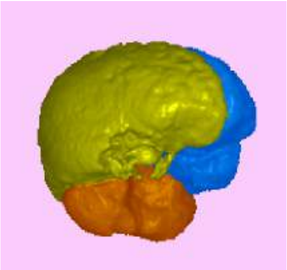


The role of the texture classifier

Abnormal brain: the texture classifier avoids internal markers in surgically removed regions of the brain.



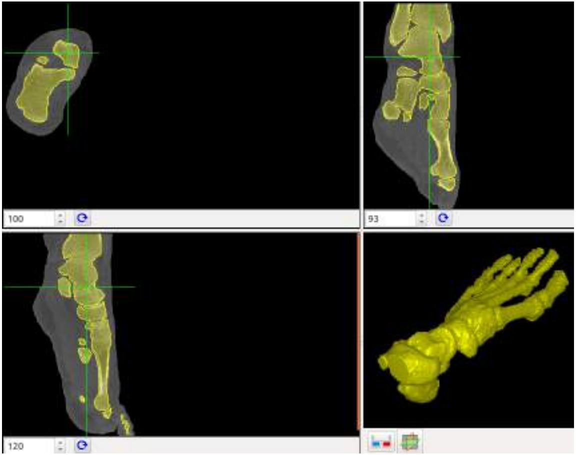




Corrections in this optimum-path forest is difficult, due to the high number of roots (low number of markers).

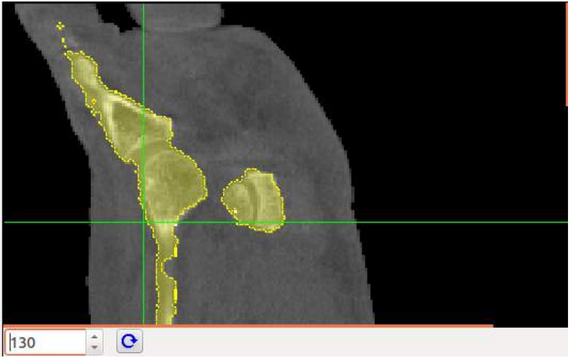
Segmentation correction

Errors in segmentation can be interactively corrected, without starting over, by converting the result into an optimum-path forest rooted at a few markers (Falcão and Bergo, TMI'04, Miranda et al, SIBGRAPI'10 and ISBI'11).



Segmentation correction

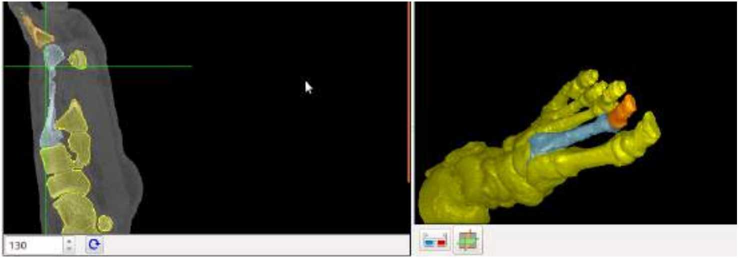
Errors in segmentation can be interactively corrected, without starting over, by converting the result into an optimum-path forest rooted at a few markers (Falcão and Bergo, TMI'04, Miranda et al, SIBGRAPI'10 and ISBI'11).



Segmentation correction

Errors in segmentation can be interactively corrected, without starting over, by converting the result into an optimum-path forest rooted at a few markers (Falcão and Bergo, TMI'04, Miranda et al, SIBGRAPI'10 and ISBI'11).

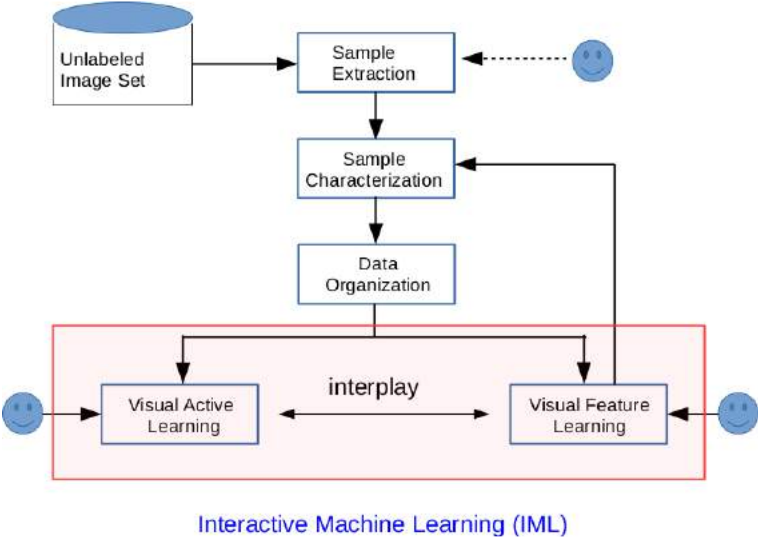
Errors in segmentation can be interactively corrected by converting the result into an optimum-path forest rooted at a few markers (Miranda et al, SIBGRAPI'10 and ISBI'11).



The model can then be improved by adding to it the new example of the object's shape.

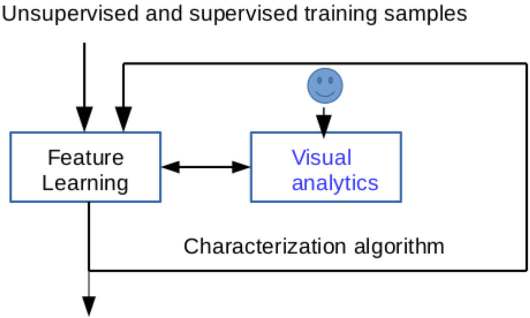
Visual feature learning

For a reasonable training set, visual feature learning aims at considerably improving sample characterization.



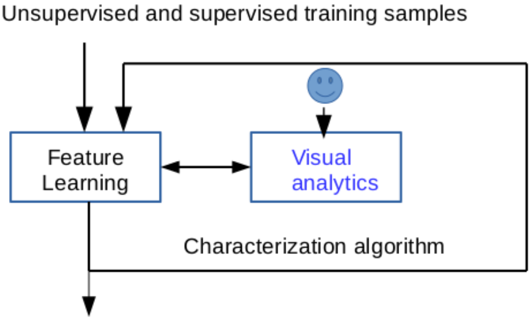
Visual feature learning

Data-driven approaches, such as bag of visual words and deep learning methods, may be applied.



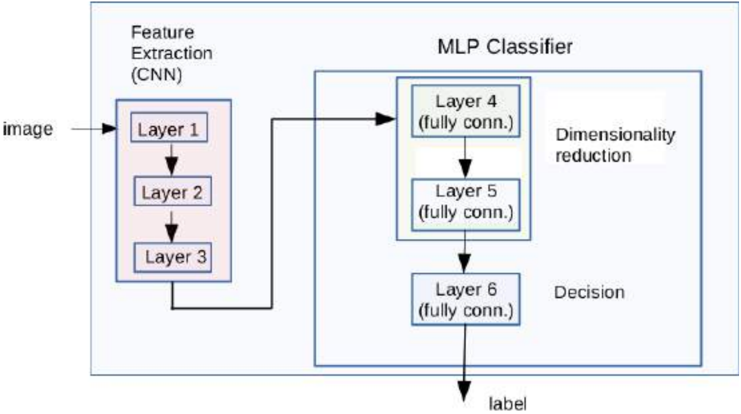
Visual feature learning

Data-driven approaches, such as bag of visual words and deep learning methods, may be applied.

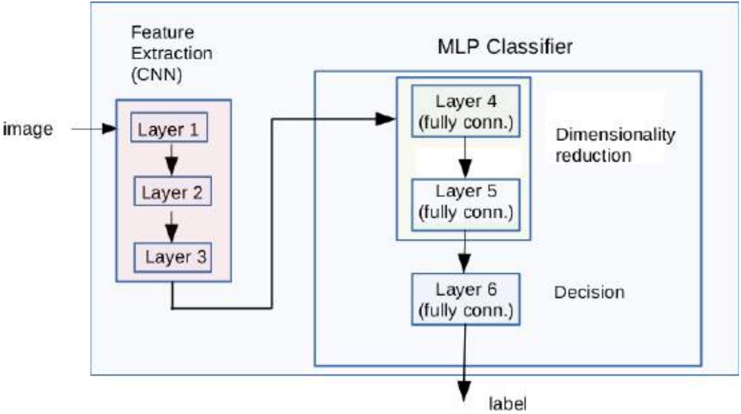


Can visual analytics help the specialist to understand and intervene in the feature learning process?

A deep neural network may be interpreted by parts.

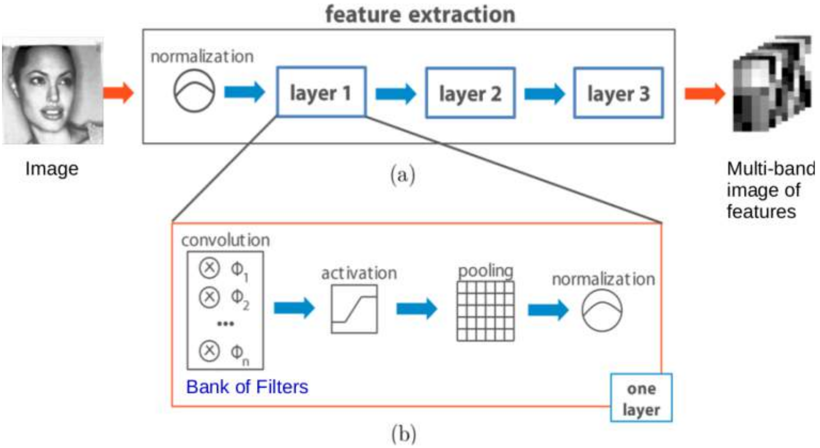


A deep neural network may be interpreted by parts.



Feature learning aims at optimizing the parameters of the network, such that the output of the last hidden layer is an effective feature vector.

Convolutional networks (ConvNets) are very effective for sample characterization.



Visual feature learning

- The best strategy is to use a known network architecture for a similar problem and refine its synaptic weights by back propagation.

Visual feature learning

- The best strategy is to use a known network architecture for a similar problem and refine its synaptic weights by back propagation.

- One may also learn the architecture of the ConvNet by adopting random synaptic weights and replacing the MLP by a SVM classifier (Chiachia et al, IEEE TIFS'14, Menotti et al, IEEE TIFS'15).

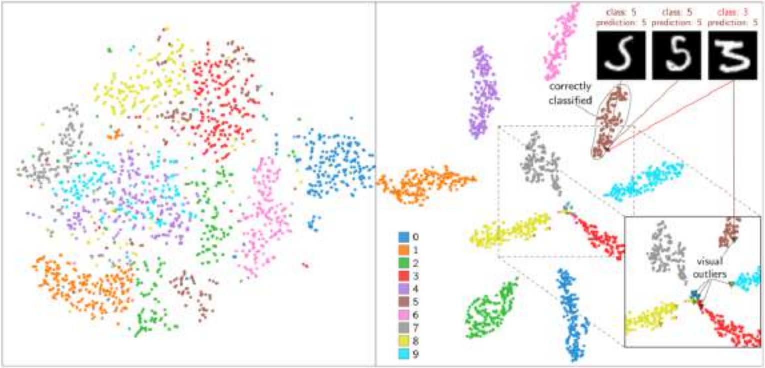
Visual feature learning

- The best strategy is to use a known network architecture for a similar problem and refine its synaptic weights by back propagation.
- One may also learn the architecture of the ConvNet by adopting random synaptic weights and replacing the MLP by a SVM classifier (Chiachia et al, IEEE TIFS'14, Menotti et al, IEEE TIFS'15).
- In any case, visual analytics may help the specialist to assess
  - features before and after dimensionality reduction,
  - features from distinct learning strategies,
  - the most important features for a given class,
  - the evolution of the training process, etc.

and to intervene in the process.

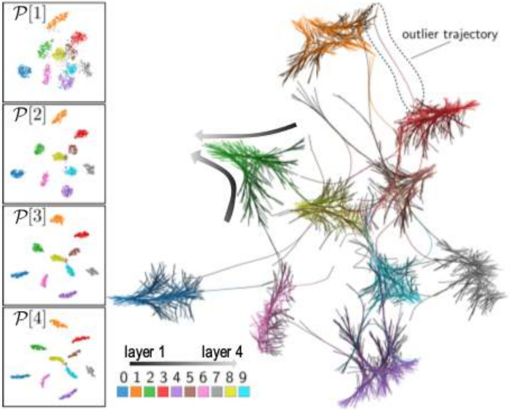
Visual analytics

Higher separability implies higher accuracy (Rauber et al, EuroVis'15).

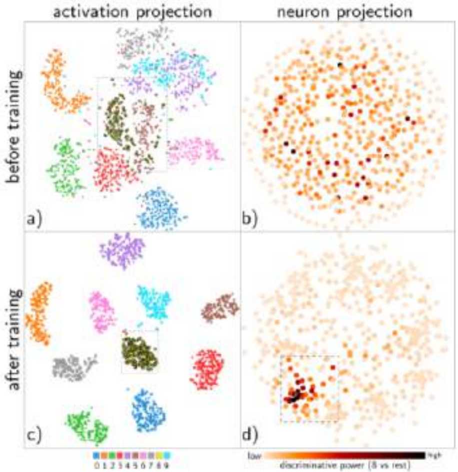


Sample projections (t-SNE) from $\Re^{1000}$ of the last HL of a MLP (with 4 HLs), before (left) and after (right) training (MNIST test set).

Sample projections of the inter-layer evolution of the MLP (with 4 HLs) after training (MNIST test set).

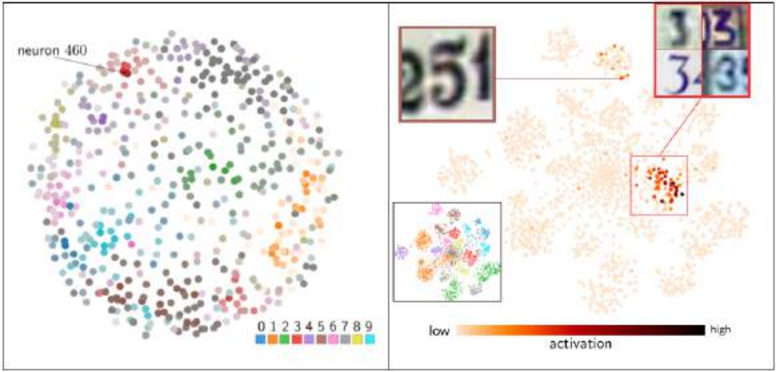


It shows compact clusters and a few outliers (Rauber et al, EuroVis'16).



Sample projections (left) from $\Re^{512}$ of the last CNN HL (MNIST test set). Neuron projections (right) colored by their discriminative power for class 8 versus the rest. (Rauber et al, IEEE TVCG'16)

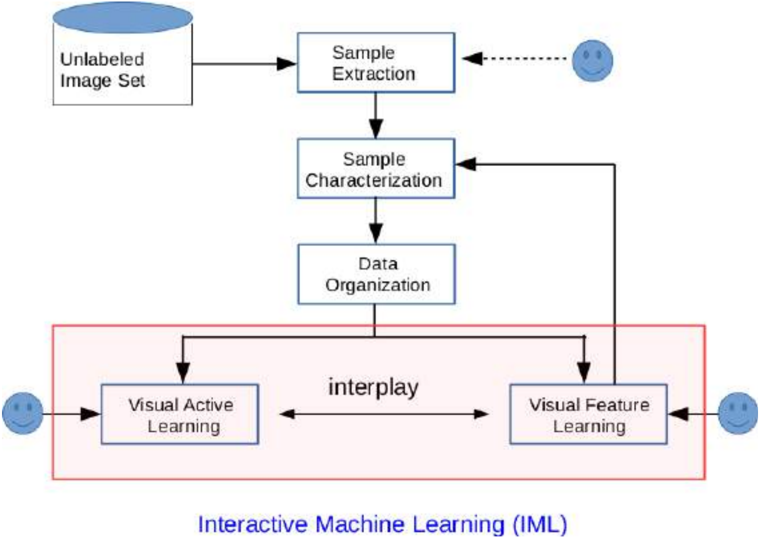
Specialized neurons fire on confusing inputs (SVHN test set).



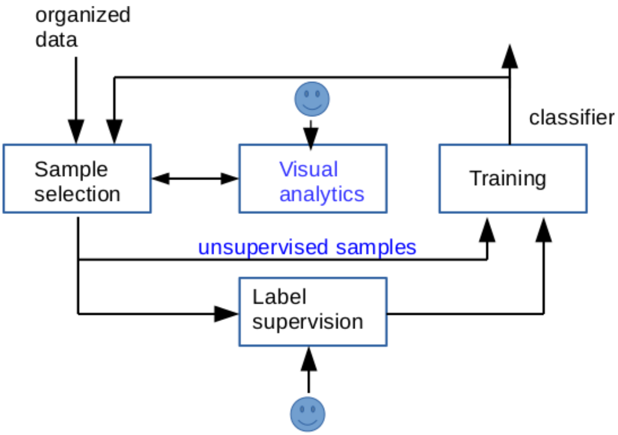
Last CNN HL after training. Discriminative neuron map (left) and sample projection (right) colored by the activation of neuron 460, which is related to class 3. (Rauber et al, IEEE TVCG'16)

Visual active learning

Visual active learning must select representative samples from all classes, and then evolve by selecting informative samples.



At each iteration, the classifier labels and selects samples for label supervision and data augmentation (Saito et al, ICPR'14, Amorim et al, Patt.Recog.'16).



The increased semi-supervised training set is expected to improve the classifier.

Visual analytics allows the specialist to

- assess the organized data,
- observe samples selected by the classifier,
- intervene on label propagation to unsupervised samples, etc.

Visual analytics

Visual analytics allows the specialist to

- assess the organized data,
- observe samples selected by the classifier,
- intervene on label propagation to unsupervised samples, etc.

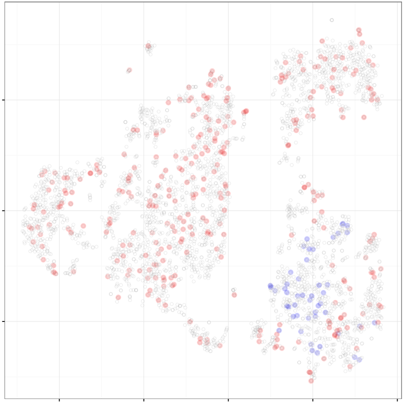
Can it help to improve the design of the classifier?

Visual analytics

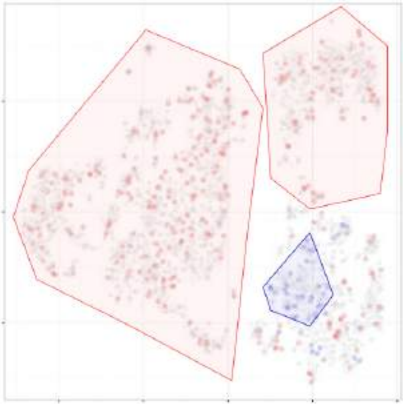
Consider a real example from the automated diagnosis of intestinal parasites (Saito et al, Patt. Recogn.'15).

It is easy to isolate larvae of helminths and impurities (large being some similar to larvae) from other classes by simple shape features.

By projecting (t-SNE) larvae of helminths (blue), impurities (red), and the unsupervised samples (black) from $\Re^{256}$, the specialist can intervene in label propagation for data augmentation.

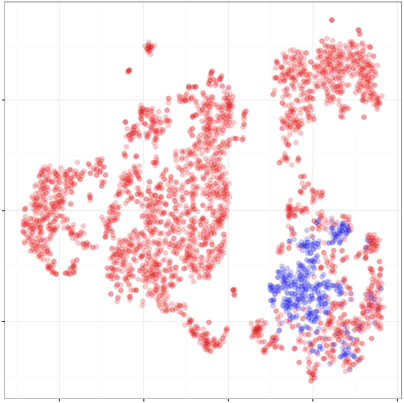


By projecting (t-SNE) larvae of helminths (blue), impurities (red), and the unsupervised samples (black) from $\Re^{256}$, the specialist can intervene in label propagation for data augmentation.



Visual analytics

By projecting (t-SNE) larvae of helminths (blue), impurities (red), and the unsupervised samples (black) from $\Re^{256}$, the specialist can intervene in label propagation for data augmentation.



Conclusive remarks

- Interactive machine learning is the way to build explainable image annotation systems.

Conclusive remarks

- Interactive machine learning is the way to build explainable image annotation systems.
- Active modeling of object shapes, visual active learning, and visual feature learning are new topics for research.

Conclusive remarks

- Interactive machine learning is the way to build explainable image annotation systems.
- Active modeling of object shapes, visual active learning, and visual feature learning are new topics for research.
- In all cases, the challenges involve
  - to minimize human effort in creating training sets,

Conclusive remarks

- Interactive machine learning is the way to build explainable image annotation systems.
- Active modeling of object shapes, visual active learning, and visual feature learning are new topics for research.
- In all cases, the challenges involve
  - to minimize human effort in creating training sets,
  - to provide response to the humans' actions at interactive time, and

Conclusive remarks

- Interactive machine learning is the way to build explainable image annotation systems.
- Active modeling of object shapes, visual active learning, and visual feature learning are new topics for research.
- In all cases, the challenges involve
  - to minimize human effort in creating training sets,
  - to provide response to the humans' actions at interactive time, and
  - to speed up convergence from human intervention (i.e., without losing control over the process).

Acknowledgments

Financial support

FAPESP

Ubbo Emmius

CNPq

CAPES

UNICAMP

Image Segmentation

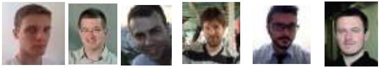
T. Spina S. Martins R. Phellan

P. Miranda J. Udupa K. Ciesielski

L. Mansilla F. Bergo F. Cappabianco

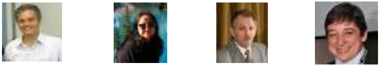
R. Lotufo J. Stolfi

Feature Learning and Visual Analytics



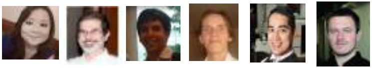
P. Rauber D. Cox D. Menotti G. Chiachia A. Peixinho A. Telea

Parasitologist Neurologist Pneumologist Radiologist



J.Gomes C. Yasuda E. Bagatin K. Irion

Active Learning, Classification, and Visual Analytics



P. Saito P. Rezende D. Neves J.P. Papa W. Amorim A. Telea