Statistical Classification of Human Histological Images

Yixin Chen
Assistant Professor
Department of Computer Science
University of New Orleans

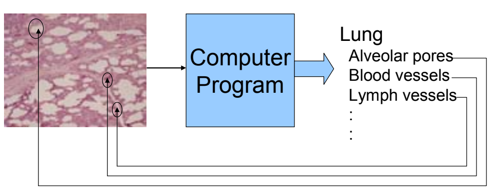
Collaborators: Dehua Zhao and Hernan Correa

Overview

- Why do we choose histological images?
  - Fundamental in medicine
  - Knowledge of normal
  - Clinical research, education, practice
- Interpreting histology slides
  - Expert knowledge
  - Time consuming
  - Human errors
  - Proper training

Overview

- Automatic interpretation of histological images

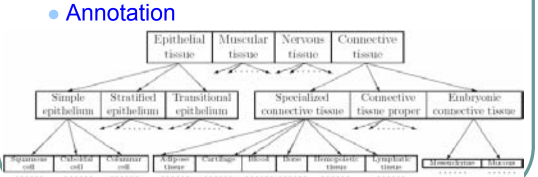


Overview

- Relatively rare research on histological images
  - CT, MRI, X rays, Ultrasound, …
- A challenging problem
  - Variations of colors
  - Object appearances
  - Interpretations at different magnifications
  - Substantial in size

Overview

- Two research problems
  - Classification
    - Identifying the organ or part of the body
  - Annotation

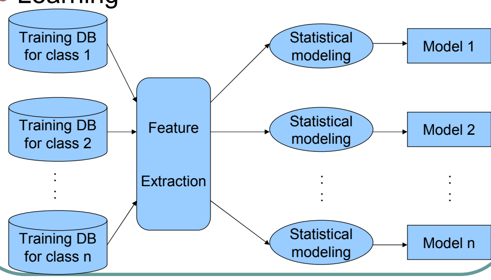


Histological Image Classification

- A road map
  - System architecture
  - Extracting imagery features
  - Statistical modeling
  - Experimental evaluation
  - Discussions

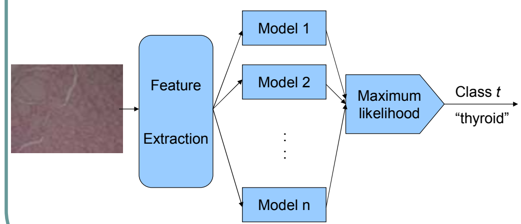
System Architecture

- Learning



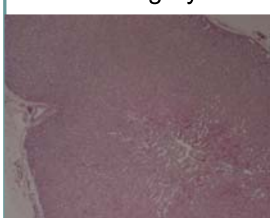
System Architecture

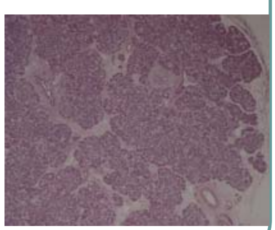
- Classification



Extracting Imagery Features

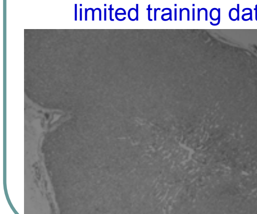
- Color or grayscale?

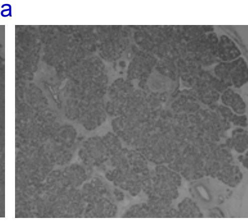

Adrenal (H&E staining)


Pancreas (H&E staining)

Extracting Imagery Features

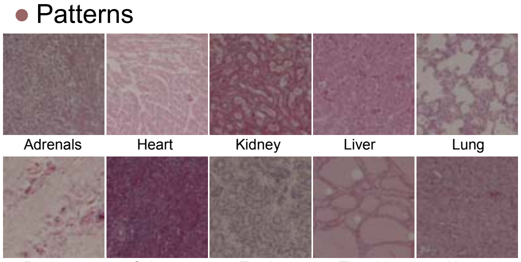
- Grayscale images
  - Pathologists, earlier histology text books, limited training data





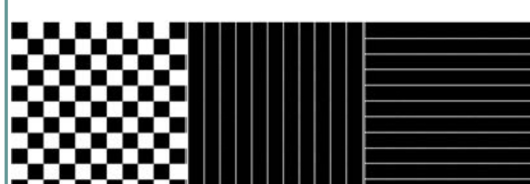
Extracting Imagery Features

- Patterns



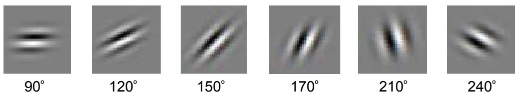
Extracting Imagery Features

- Characterizing intensity variations

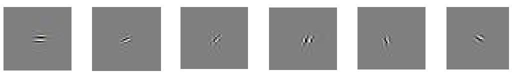

Horizontal Vertical

Extracting Imagery Features

- Gabor filter bank
  - Multiple scales and orientations



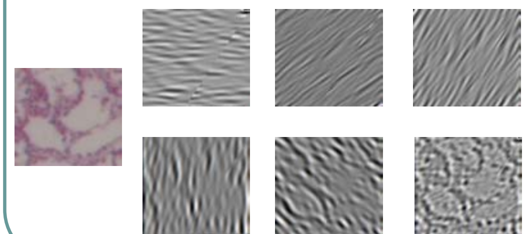
Low resolution



High resolution

Extracting Imagery Features

- Example



Extracting Imagery Features

- 3 scales x 6 orientations
- Image size: 3072x3840
- Over 212 million Gabor responses
- Statistic measure of Gabor responses

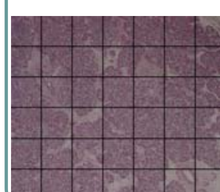
$$\mu_g = mean(|response \;\; for \;\; g|)$$

$$\sigma_g = std(|response \;\; for \;\; g|)$$

3x6x2=36 features

Extracting Imagery Features

- Compromise between pattern effectiveness and computation efficiency



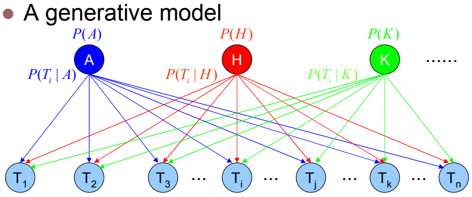
6x7 blocks

3x6 Gabor Filter Bank

$$\{\vec{f}_i \in R^{36} : i = 1, \ldots, 42\}$$

Statistical Modeling

- A generative model

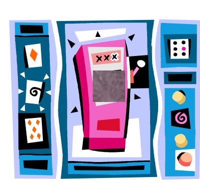


$$P(T, L) = P(L)P(T \mid L)$$

  - Generating a label, L, with probability *P*(*L*)
  - Generating a block, T, with probability *P*(*T*|*L*)

Statistical Modeling

- An illustration



A virtual image of lung

Statistical Modeling

- Class conditional probability density

$$p(\vec{f} \mid Class_i)$$

- The probability of generating an image given class label

$$p(image \mid Class_i) = p(\{\vec{f}_j : j = 1,\ldots,N\} \mid Class_i)$$

$$= \prod_{j=1}^{N} p(\vec{f}_j \mid Class_i)$$

Statistical Modeling

- How to decide the label of an image?
  - Minimal error probability
  - Bayes formula

$$p(Class_i \mid image) = \frac{P(Class_i)\prod_{j=1}^{N} p(\vec{f}_j \mid Class_i)}{\sum_{i=1}^{M}\left[P(Class_i)\prod_{j=1}^{N} p(\vec{f}_j \mid Class_i)\right]}$$

- $p(\vec{f} \mid Class_i)$ is unknown

Statistical Modeling

- Gaussian mixture model

$$p(\vec{f} \mid Class_i) = p(\vec{f} \mid \theta_i) = \sum_{k=1}^{K} p(\vec{f} \mid \mu_{ki}, \Sigma_{ki}) P_{ki}$$

$$p(\vec{f} \mid \mu_{ki}, \Sigma_{ki}) \sim Normal(\mu_{ki}, \Sigma_{ki})$$

$$\theta_i = \{\mu_{1i}, \Sigma_{1i}, \mu_{2i}, \Sigma_{2i}, \ldots, \mu_{Ki}, \Sigma_{Ki}, P_{1i}, P_{2i}, \ldots, P_{Ki}\}$$

$$\sum_{k=1}^{K} P_{ki} = 1$$

Statistical Modeling

- Parameter estimation

$$Likelihood(\theta_i) = \prod_{l=1}^{n} p(image_l \mid \theta_i)$$

$$\theta_i^{ML} = \arg\max_{\theta_i} Likelihood(\theta_i)$$

- Expectation Maximization (EM algorithm)
- Minimum Description Length model selection
- Substitution rule $p(\vec{f} \mid Class_i) = p(\vec{f} \mid \theta_i^{ML})$

Experimental Evaluation

- H&E stained, 40x
- Original size 3072x3840≈11MB
- Down sampling to 1536x1920
- Block size 64x64 720 blocks
- Gabor filter bank

ID	Category Name	Number of Images
C_1	Adrenals	25
C_2	Heart	114
C_3	Kidney	20
C_4	Liver	107
C_5	Lung	288
C_6	Pancreas	120
C_7	Spleen	18
C_8	Testis	25
C_9	Thyroid	39
C_{10}	Uterus	22

Experimental Evaluation

- 5 fold cross validation (training)

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}
	80.0%	0.0%	0.0%	3.0%	7.0%	$\boxed{8.0\%}$	0.0%	$\boxed{2.0\%}$	0.0%	0.0%
C_2	0.4%	**84.4%**	0.0%	0.0%	2.2%	5.5%	0.0%	0.0%	0.0%	7.5%
C_3	1.3%	2.5%	**82.5%**	0.0%	1.3%	1.3%	0.0%	0.0%	0.0%	0.0%
C_4	5.4%	0.2%	0.0%	**84.8%**	0.0%	4.7%	4.0%	0.5%	0.0%	0.5%
C_5	4.5%	0.0%	0.0%	0.0%	**90.0%**	4.5%	0.0%	0.0%	1.0%	0.0%
	$\boxed{16.7\%}$	1.5%	1.9%	4.8%	6.0%	**64.6%**	0.0%	0.2%	0.8%	3.5%
C_7	4.2%	0.0%	0.0%	1.4%	0.0%	1.4%	**93.1%**	0.0%	0.0%	0.0%
	$\boxed{11.0\%}$	0.0%	0.0%	1.0%	0.0%	9.0%	0.0%	**75.0%**	0.0%	4.0%
C_9	0.6%	0.0%	0.0%	0.0%	1.3%	2.6%	0.0%	0.0%	**94.2%**	1.3%
C_{10}	0.0%	8.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	**92.1%**

On average 84.1%

Experimental Evaluation

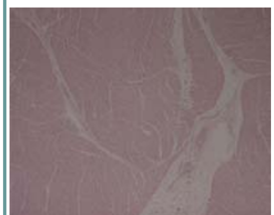
- 5-fold cross validation (testing)

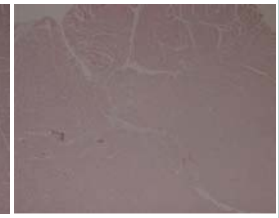
	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}
	48.0%	0.0%	0.0%	16.0%	8.0%	$\boxed{28.0\%}$	0.0%	$\boxed{0.0\%}$	0.0%	0.0%
C_2	17.5%	**80.7%**	2.6%	0.9%	0.9%	7.0%	0.0%	0.0%	0.0%	6.1%
C_3	0.0%	10.0%	**65.0%**	0.0%	5.0%	10.0%	0.0%	5.0%	0.0%	5.0%
C_4	6.5%	0.0%	0.0%	**81.3%**	0.0%	4.7%	5.6%	1.9%	0.0%	0.0%
C_5	6.3%	0.0%	0.0%	0.0%	**87.2%**	5.6%	0.0%	0.0%	1.0%	0.0%
	$\boxed{15.8\%}$	2.5%	2.5%	7.5%	5.0%	**60.8%**	0.0%	1.7%	0.8%	3.3%
C_7	5.6%	0.0%	0.0%	0.0%	0.0%	0.0%	**94.4%**	0.0%	0.0%	0.0%
	$\boxed{20.0\%}$	0.0%	8.0%	12.0%	0.0%	12.0%	0.0%	**44.0%**	0.0%	4.0%
C_9	0.0%	2.6%	0.0%	0.0%	2.6%	5.1%	0.0%	0.0%	**89.7%**	0.0%
C_{10}	0.0%	22.7%	0.0%	0.0%	0.0%	4.6%	0.0%	0.0%	0.0%	**72.7%**

On average 72.4%

Experimental Evaluation

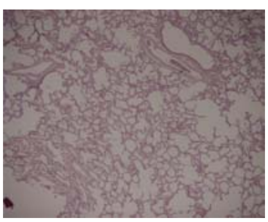
- Observations

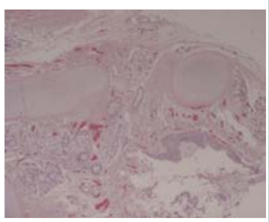

A "good" Heart image, always classified correctly


A boundary Heart image, misclassified in all 5 folds

Experimental Evaluation

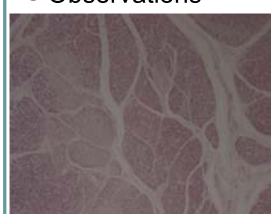
- Observations

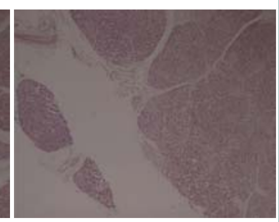

A "good" Lung image, always classified correctly


A boundary Lung image, misclassified in all 5 folds

Experimental Evaluation

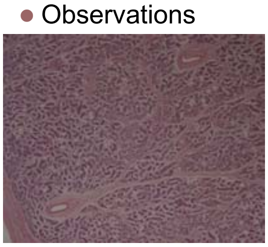
- Observations

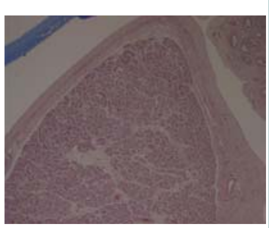

A "good" Pancreas image, always classified correctly


A boundary Pancreas image, misclassified in all 5 folds

Experimental Evaluation

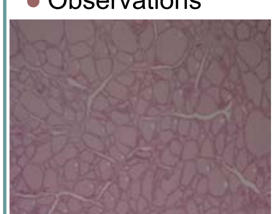
- Observations

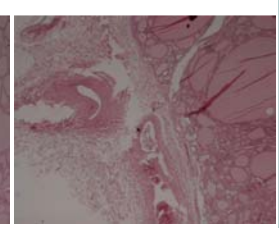

A "good" Testis image, always classified correctly


A boundary Testis image, misclassified in all 5 folds

Experimental Evaluation

- Observations


A "good" Thyroid image, always classified correctly


A boundary Thyroid image, misclassified in all 5 folds

Experimental Evaluation

- 75 images are misclassified in all 5 folds
- 53 images are boundary images (70.7%)
- Around 14% of the images in the dataset are boundary images

Experimental Evaluation

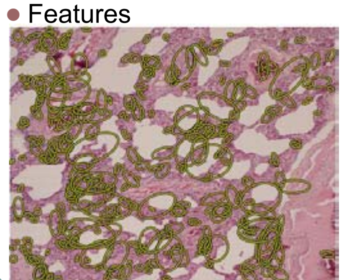
- Software
  - Written in C programming language
- Speed
  - SunFire V800z, Solaris, on a single P4 1.9GHz CPU
  - Training time varies from 30 minutes to 15 hours for different categories
  - Testing time ≈ 7 minutes/image
- Scalability

Performance Improvements

- Data set
- Tuning up
- Features
- Model

Performance Improvements

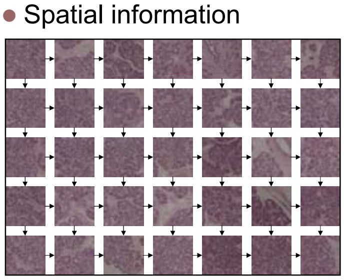
- Features



Affine invariant region patches

Performance Improvements

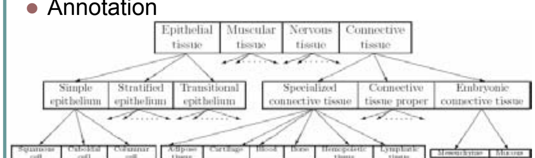
- Spatial information



Two-dimensional Hidden Markov Model

Future Work

- A larger scale
- Annotation



- Pathology images
- Retrieval

Acknowledgements

- Supported by
  - University of New Orleans
  - The Research Institute for Children
  - Louisiana NASA EPSCoR DART
- Thank
  - Dr. Seth Pincus (RIC)
  - Dr. Hernan Correa (LSUHSC)
  - Dehua Zhao (UNO)
  - Dr. Giovanni D. Lorusso (LSUHSC)
  - Dr. Mahdi Abdelguerfi (UNO)
  - Venkata Mahadevan (UNO)