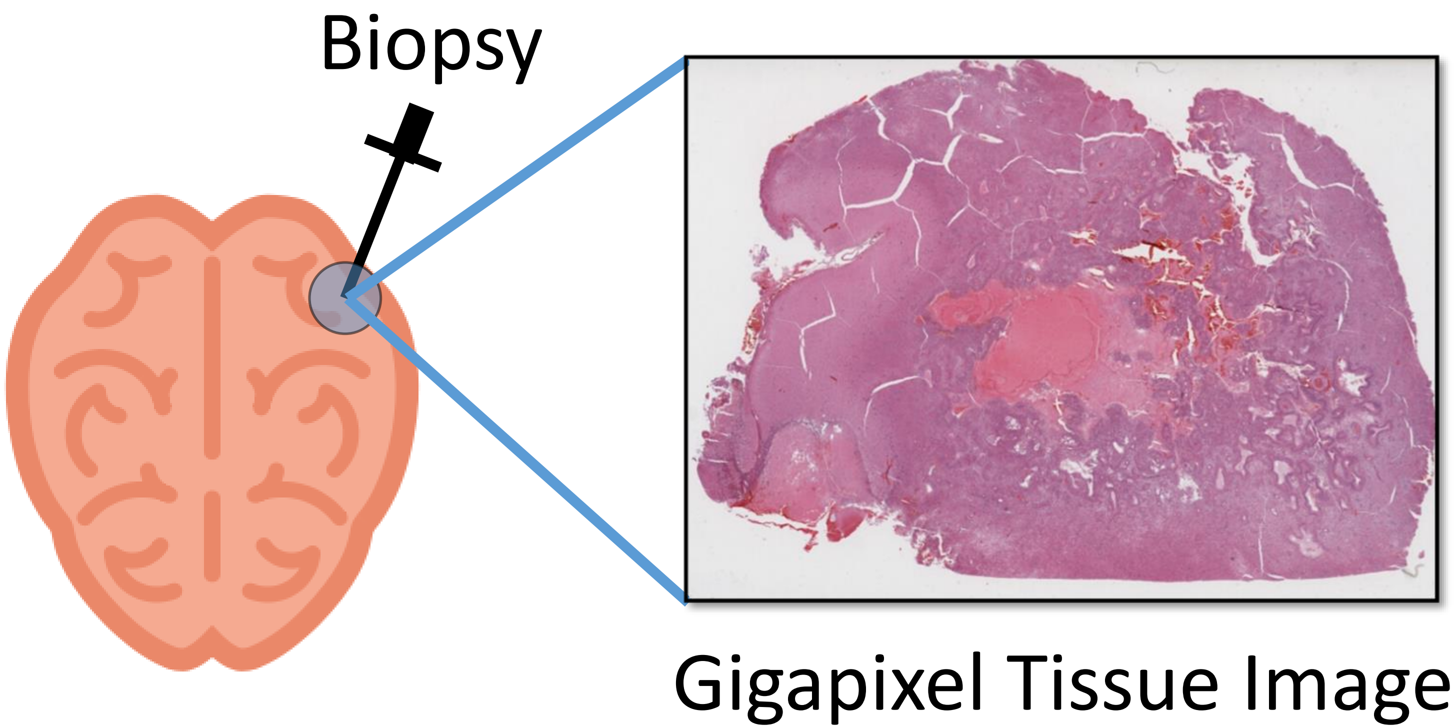
Automatic Histopathology Image Analysis with CNNs

Le Hou, Kunal Singh, Dimitris Samaras, Tahsin M. Kurc, Yi Gao, Roberta J. Seidman, Joel H. Saltz

Stony Brook University

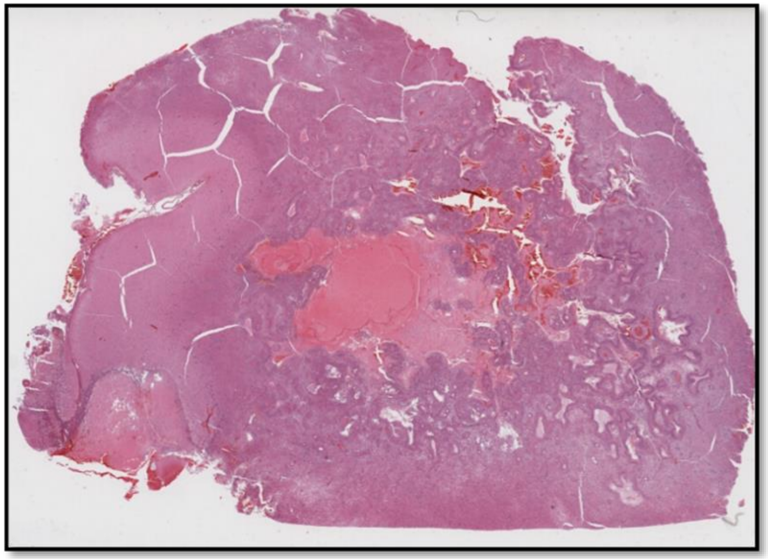
Gagapixel Whole Slide Tissue Image



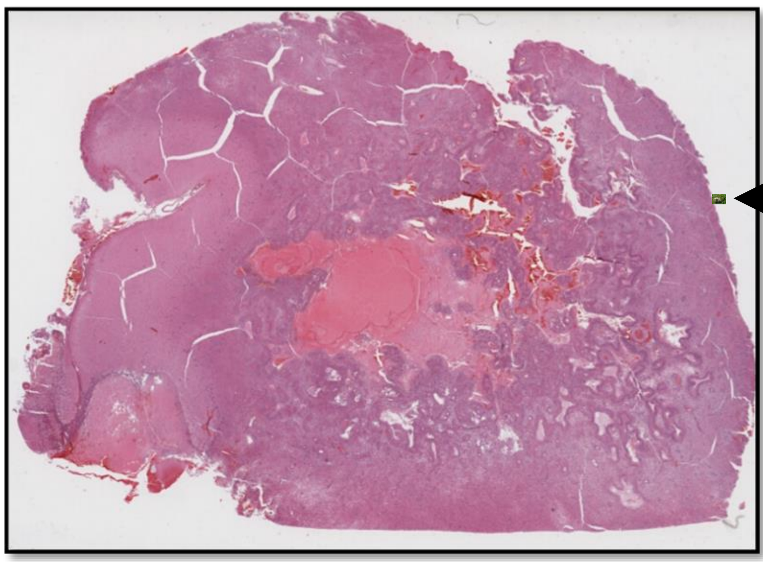
[vector.me]

Analyzing Tissue Images

- Is crucial to study disease onset
- To develop targeted treatment
- Is a challenging problem



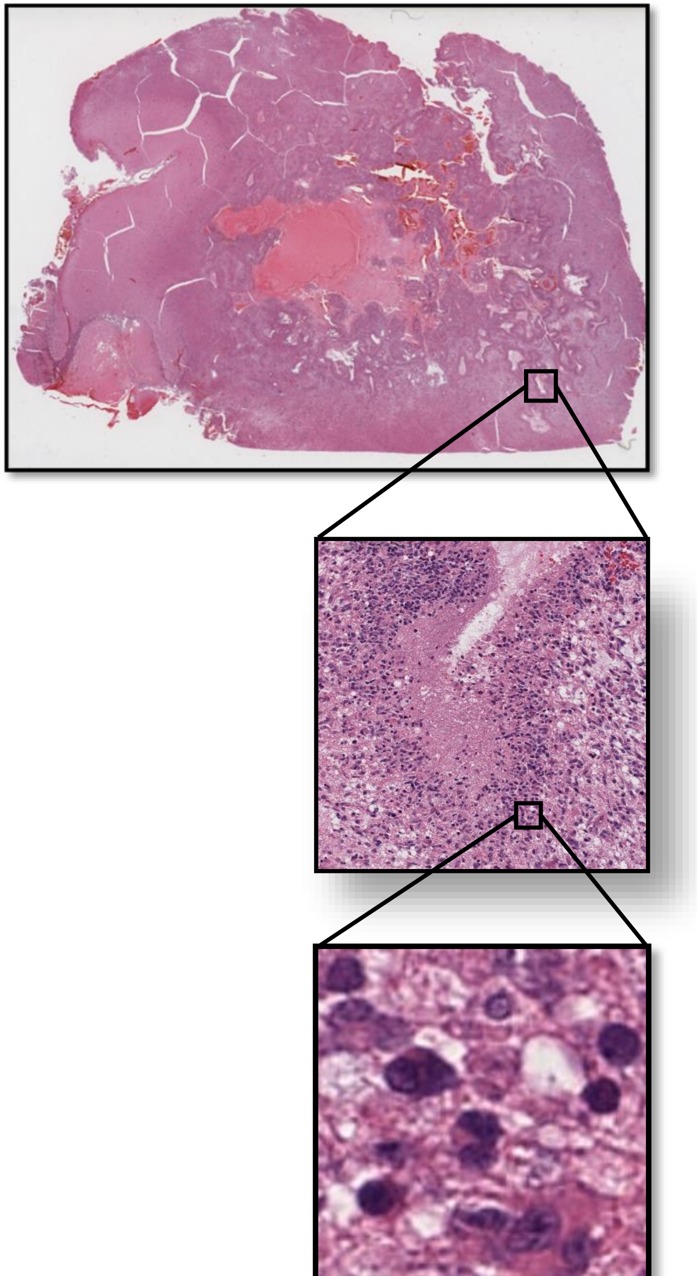
Gigapixel Resolution

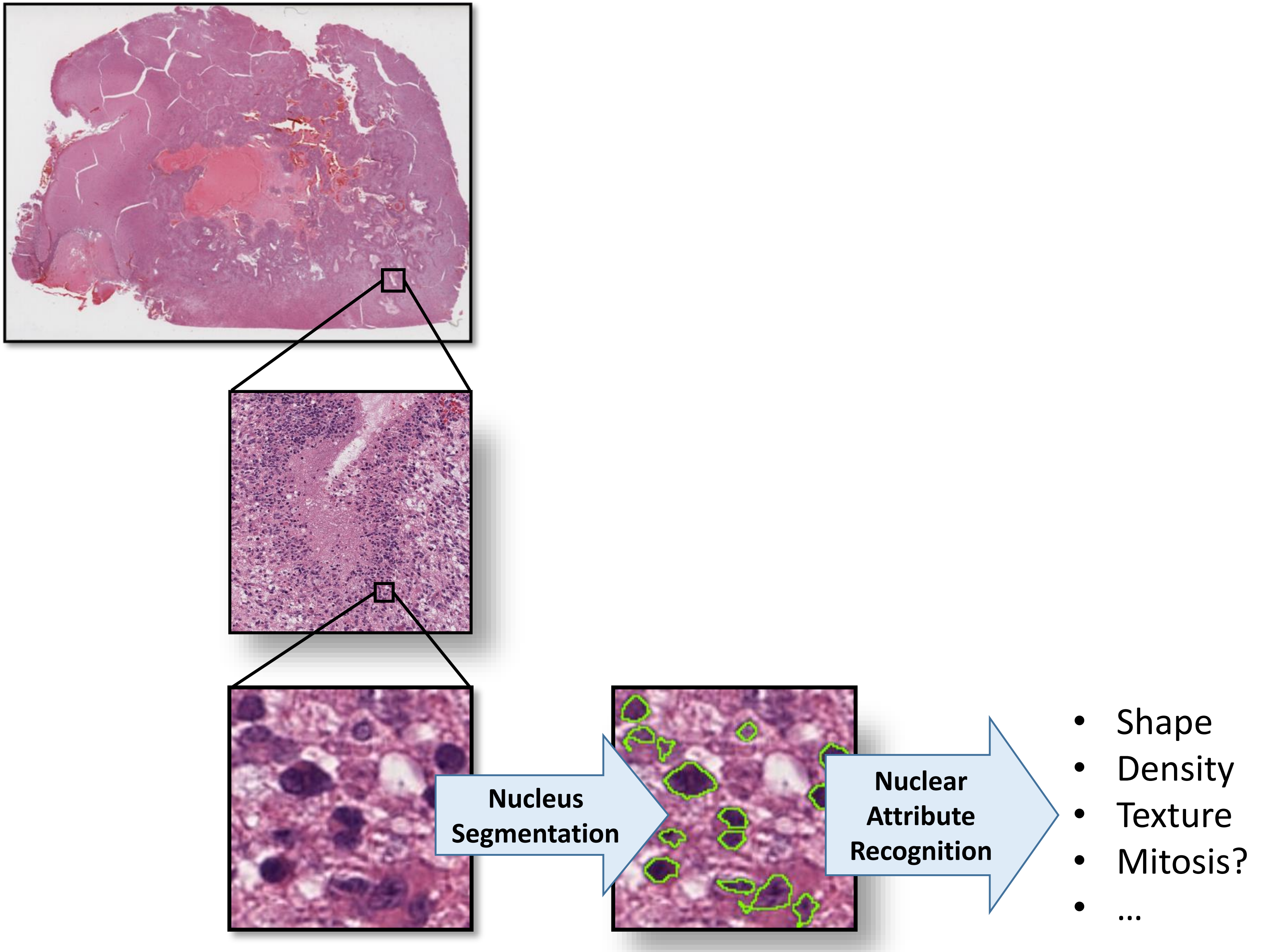


50K × 50K

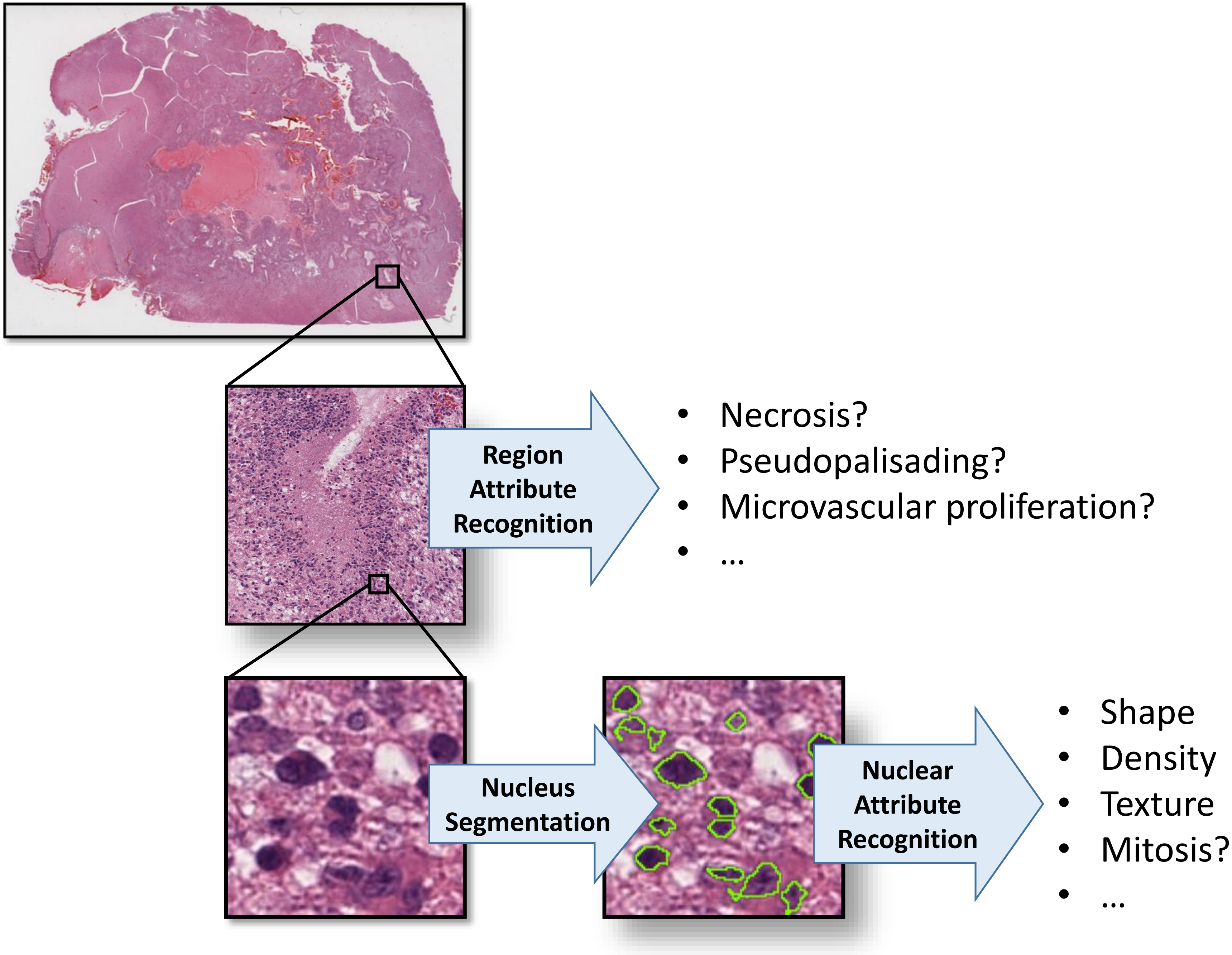
An ImageNet image

350 × 500

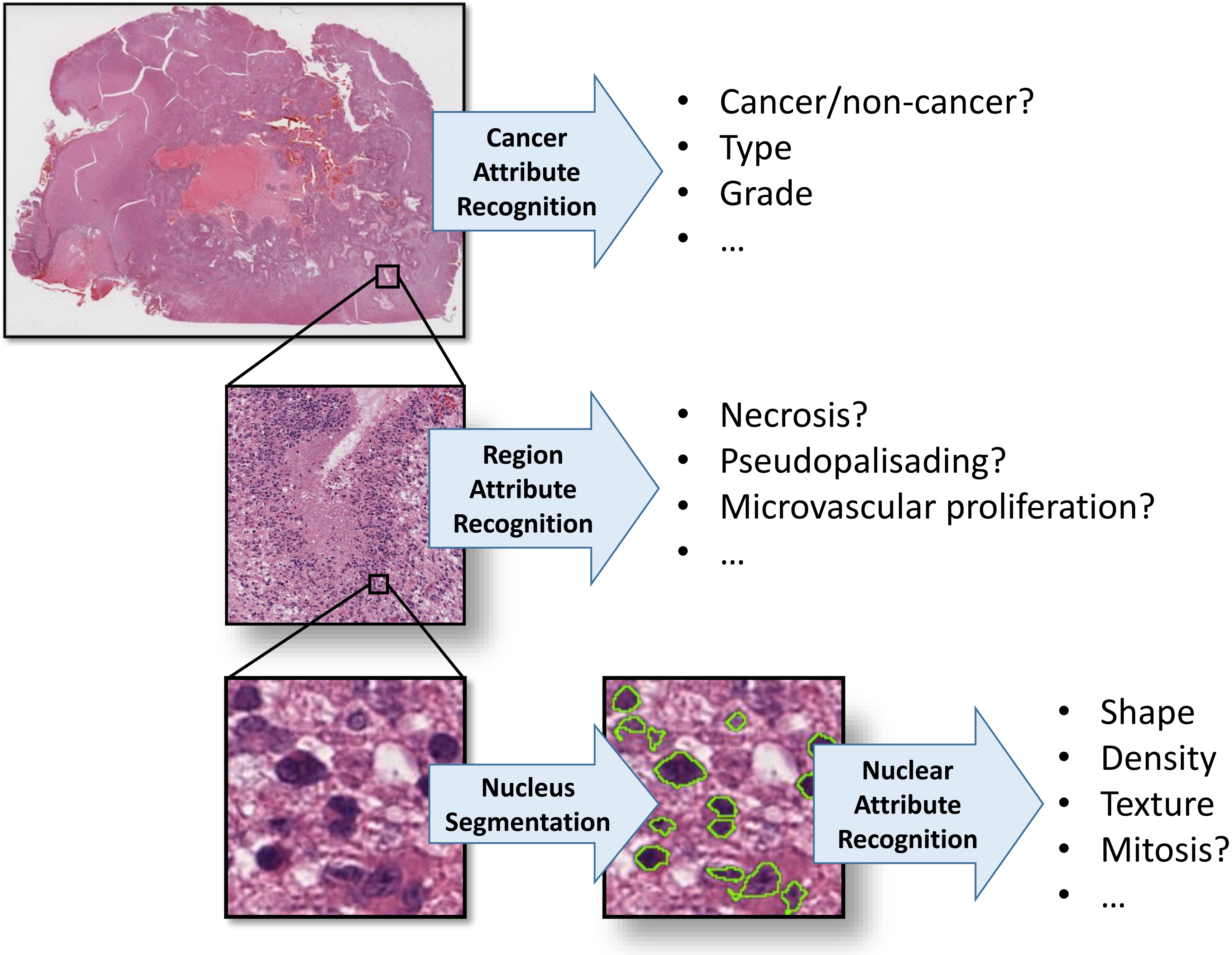


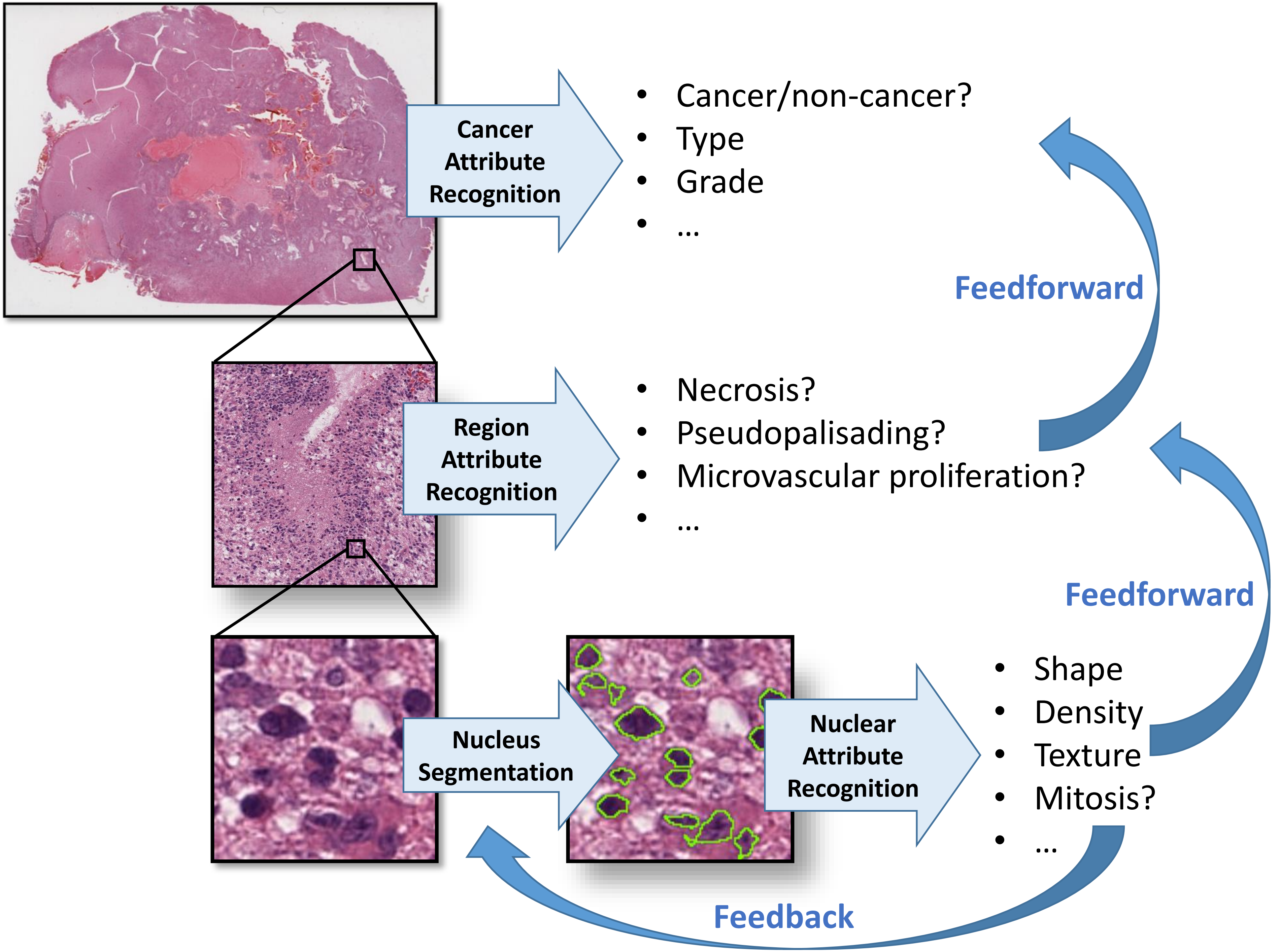


Nucleus Segmentation
Nuclear Attribute Recognition
• Shape
• Density
• Texture
• Mitosis?
• …



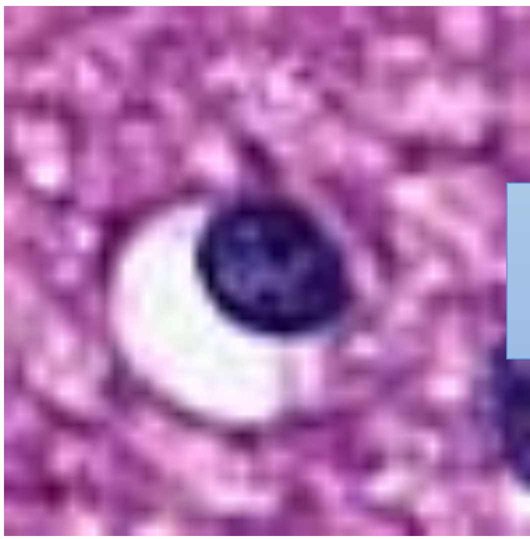
Region Attribute Recognition
Necrosis?
Pseudopalisading?
Microvascular proliferation?
...
Nucleus Segmentation
Nuclear Attribute Recognition
Shape
Density
Texture
Mitosis?
...


Cancer Attribute Recognition
• Cancer/non-cancer?
• Type
• Grade
• …
Region Attribute Recognition
• Necrosis?
• Pseudopalisading?
• Microvascular proliferation?
• …
Nucleus Segmentation
Nuclear Attribute Recognition
• Shape
• Density
• Texture
• Mitosis?
• …



Cancer Attribute Recognition
Cancer/non-cancer?
Type
Grade
…
Feedforward
Region Attribute Recognition
Necrosis?
Pseudopalisading?
Microvascular proliferation?
…
Feedforward
Nucleus Segmentation
Nuclear Attribute Recognition
Shape
Density
Texture
Mitosis?
…
Feedback

Recognizing Attributes of Glioma Nuclei

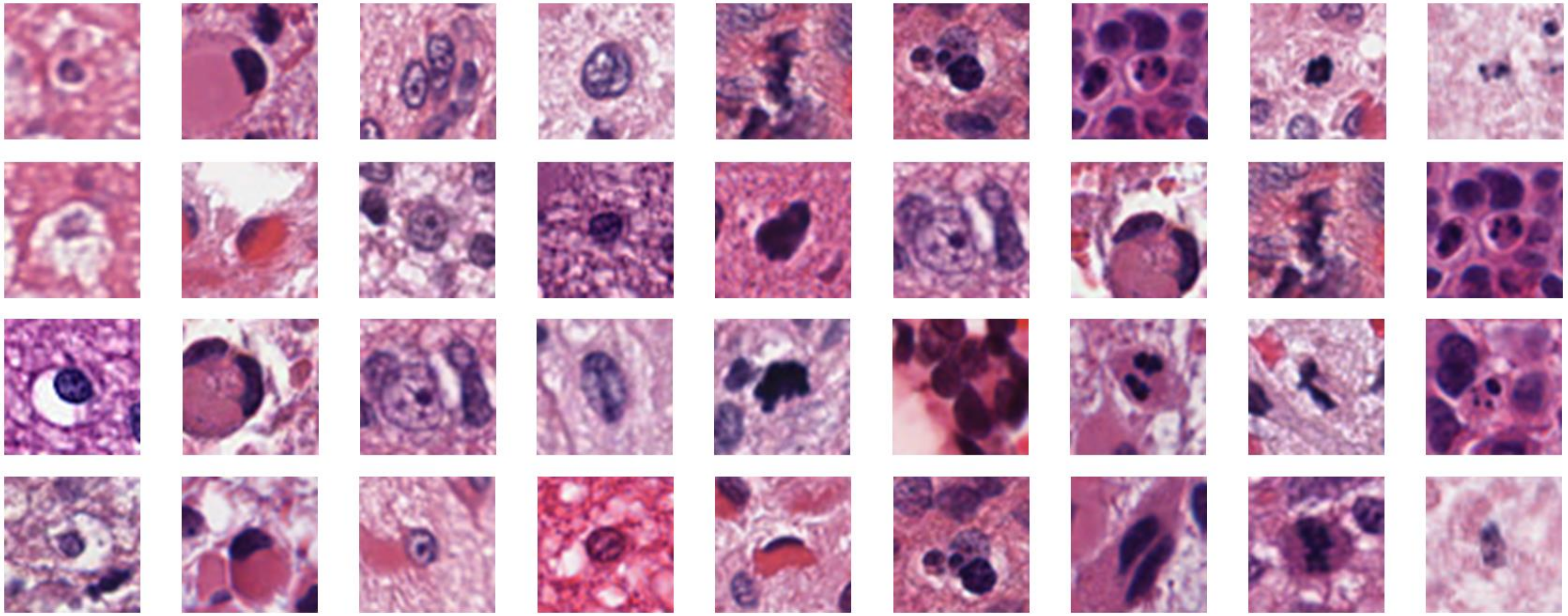


Automatic Recognition

- ✓ Perinuclear Halo
- × Gemistocyte
- × Nucleoli
- ✓ Hyperchromasia
- × Mitosis
- …

Our Dataset

Morphological Attributes	#. Present	#. Absent
Perinuclear halos	78	2000
Gemistocyte	51	2027
Nucleoli	77	2001
Grooved	14	2064
Hyperchromasia	505	1573
Overlapping nuclei	105	1973
Multinucleation	43	2035
Mitosis	53	2025
Apoptosis	20	2058
No nucleus	545	1533



A Multi-label Problem

- There can be multiple classes (nuclear attributes) for each instance (glioma nuclear image).
- Existing approaches [Thibault, 2008] [Kong, 2011] ignored the multi-label nature.

A Multi-label Problem

- There can be multiple classes (nuclear attributes) for each instance (glioma nuclear image).
- Existing approaches [Thibault, 2008] [Kong, 2011] ignored the multi-label nature.

- Our contribution:

  first multi-label modeling on this problem.

A Multi-label Problem

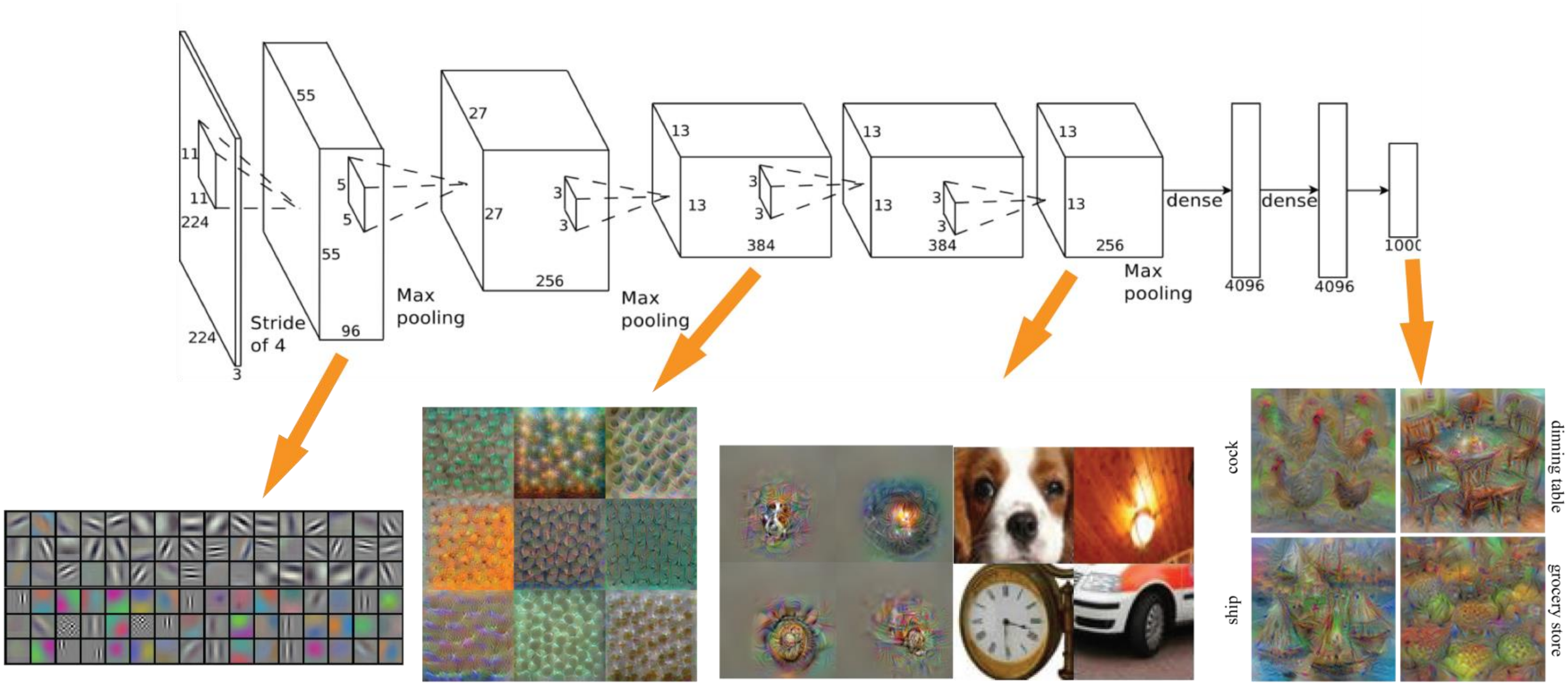
- Existing approaches [Thibault, 2008] [Kong, 2011] only focus on a subset of classes at a time.
- Our contribution:

  Recognizing nine subtle and important attributes with good accuracy

Convolutional Neural Network (CNN)

- A popular image classification method
- Input training set:
  Images with ground truth labels
- Output:
  Predicted class(es)

CNN for Image Classification



Conv 1: Edge+Blob

Conv 3: Texture

Conv 5: Object Parts

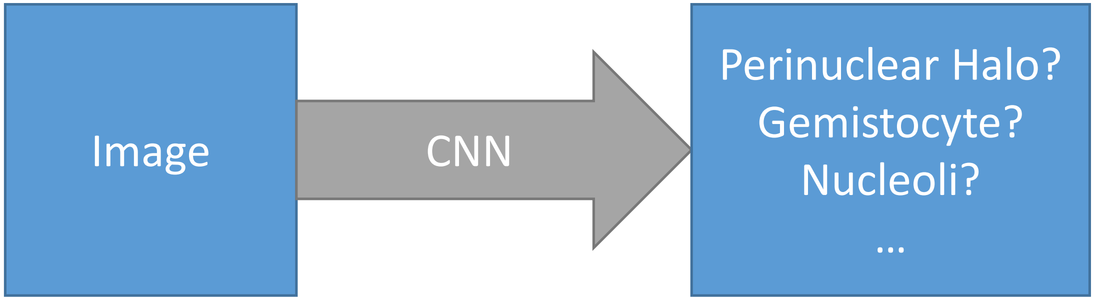
Fc8: Object Classes

[James Hays]

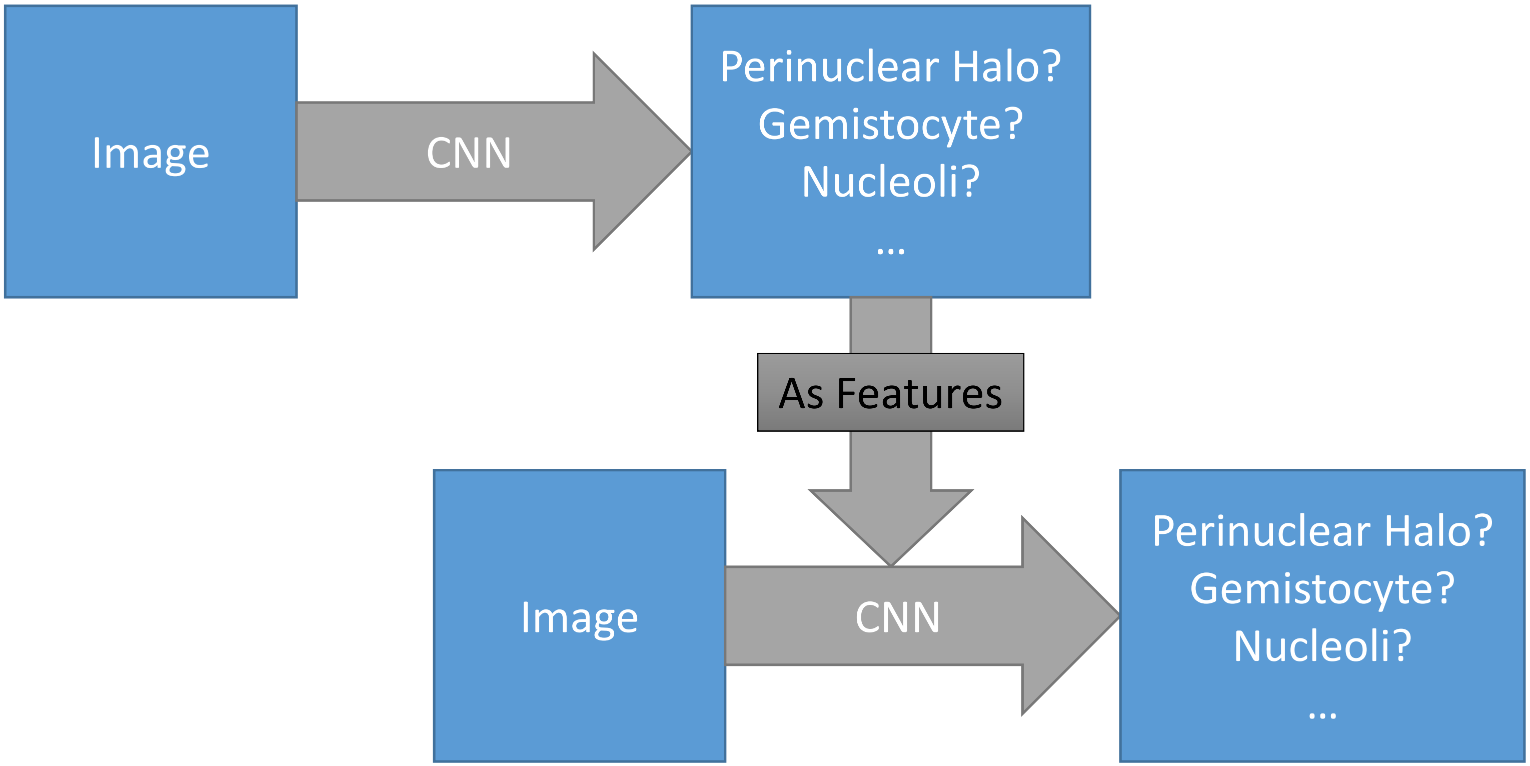
Multi-label CNN

- Approach 1:
  - Predict each class independently
  - Drawback: do not capture inter-class dependency
    Example: mitosis are always hyperchromasia
- Approach 2:
  - A chain of CNNs [Read, 2009]

Multi-label CNN



Multi-label CNN

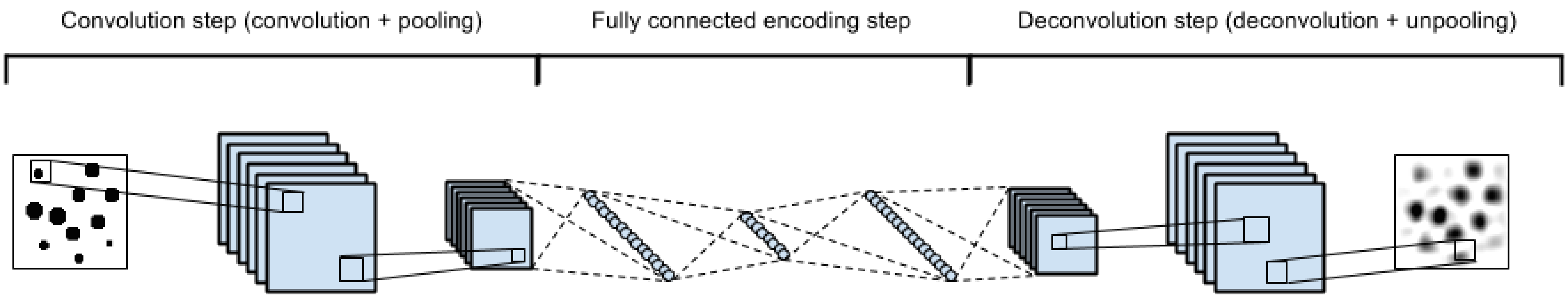


Semi-supervised CNN

- Getting ground truth labels is laborious
- Tissue Images have billions of unlabeled nuclei
- To utilize unlabeled nuclear images: Semi-supervised CNN

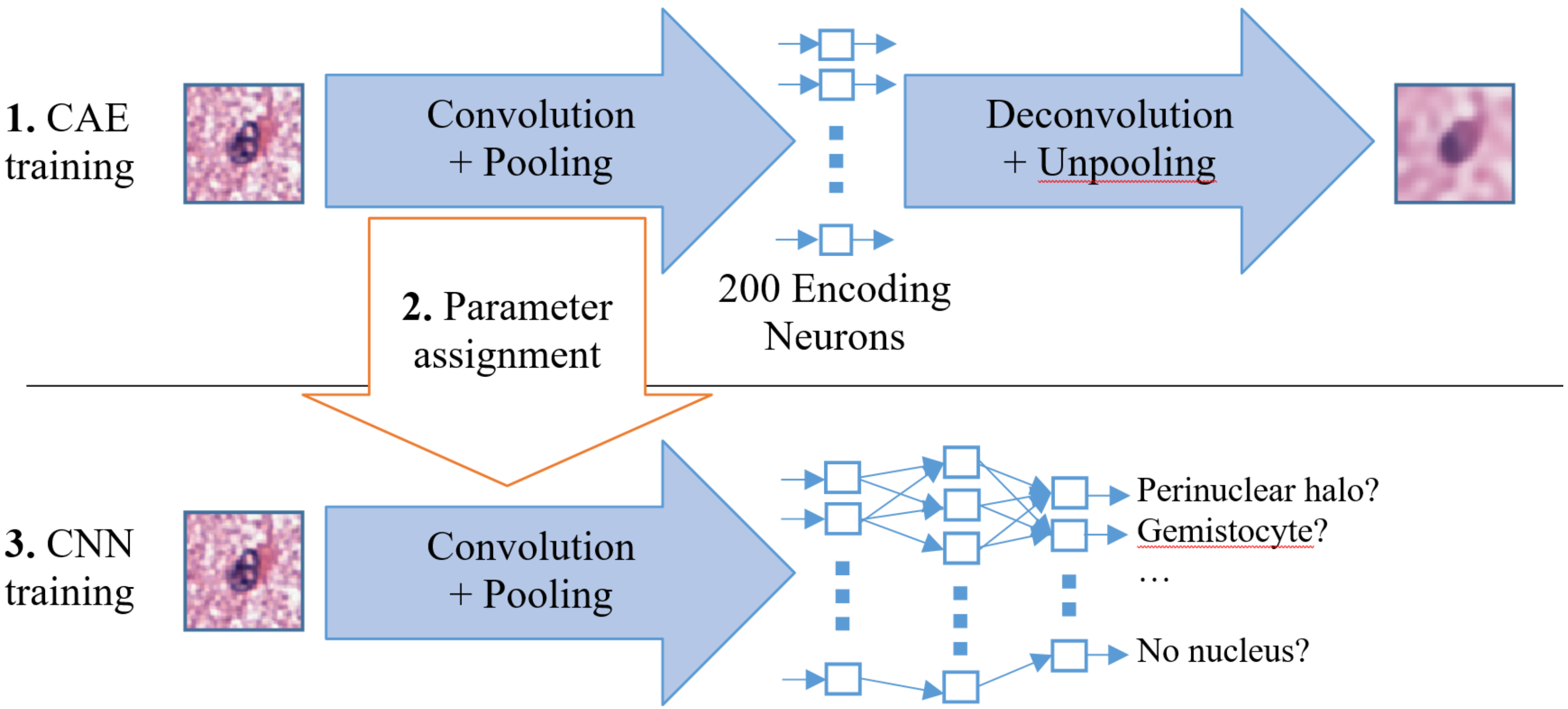
Training a CNN without Ground Truth Labels

Convolutional AutoEncoder (CAE)

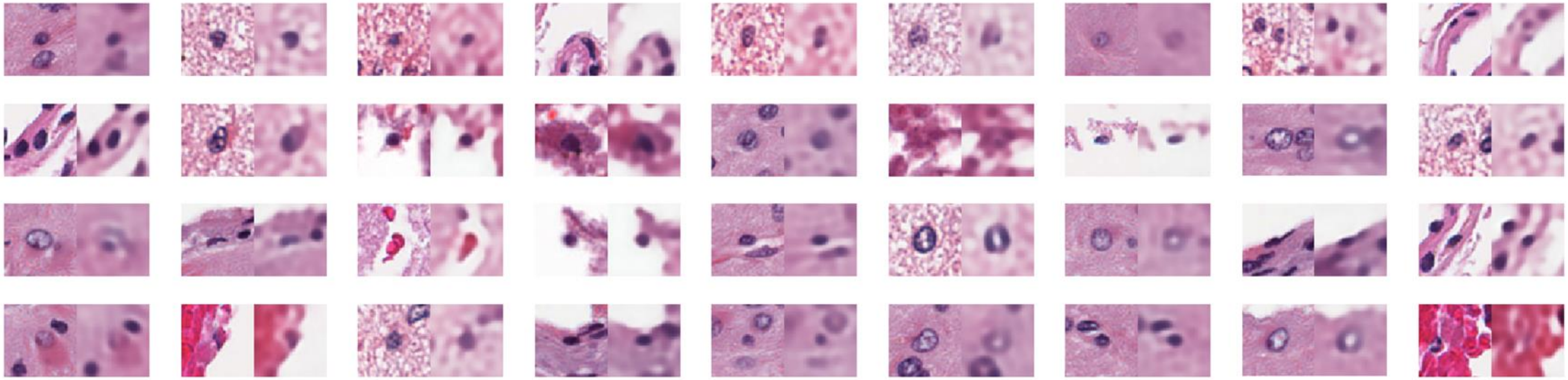


[Mike Swarbrick Jones]

Semi-supervised CNN with CAE



Reconstructed Images by CAE



Left: original nuclear images.
Right: CAE reconstructed images.

Training a CNN on Another Dataset

- Alternative:
  - Train a CNN on a different dataset that has ground truth labels.
  - Use this CNN as a feature extractor.
  - Use Support Vector Machine (SVM) as a classification model.
- We used the CNN trained by the Visual Geometry Group (VGG) as a feature extractor.

Results

	Area Under the ROC Curve (AUC)		
Morphological Attributes	Semi-supervised CNN	VGG16 + SVM	Best of two (per attribute)
Perinuclear halos	0.8789	**0.9257**	0.9257
Gemistocyte	0.8026	**0.9548**	0.9548
Nucleoli	0.8366	**0.9076**	0.9076
Grooved	**0.8956**	0.7296	0.8956
Hyperchromasia	**0.9450**	0.8854	0.9450
Overlapping nuclei	**0.8969**	0.8305	0.8969
Multinucleation	0.7329	**0.7507**	0.7507
Mitosis	**0.8731**	0.8559	0.8731
Apoptosis	0.8676	**0.9767**	0.9767
No nucleus	**0.9828**	0.9639	0.9828
Averaged AUC	0.8712	0.8616	**0.9109**

Both methods perform well on some but not all morphological attributes and are complementary with each other.

Summary

- Automatically classify nuclei is important
- We model it as a multi-label problem
- We achieved promising results classifying nine subtle nuclear attributes
- Future work: combining two CNN-based methods. R-CNN