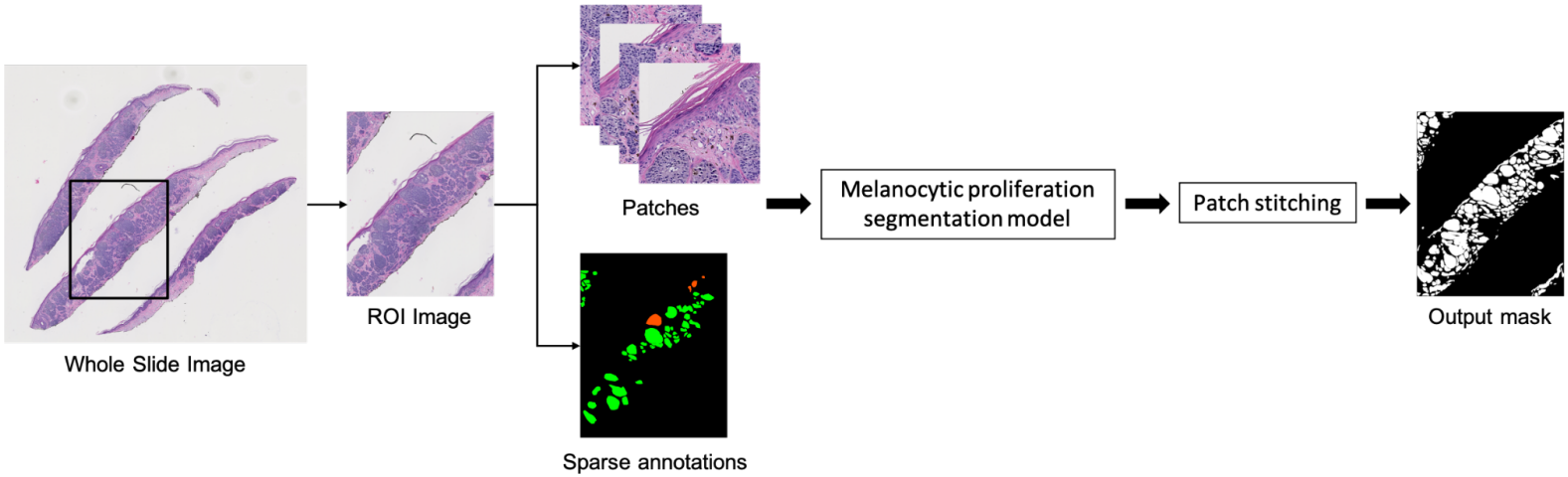
Learning Melanocytic Proliferation Segmentation in Histopathology Images from Imperfect Annotations

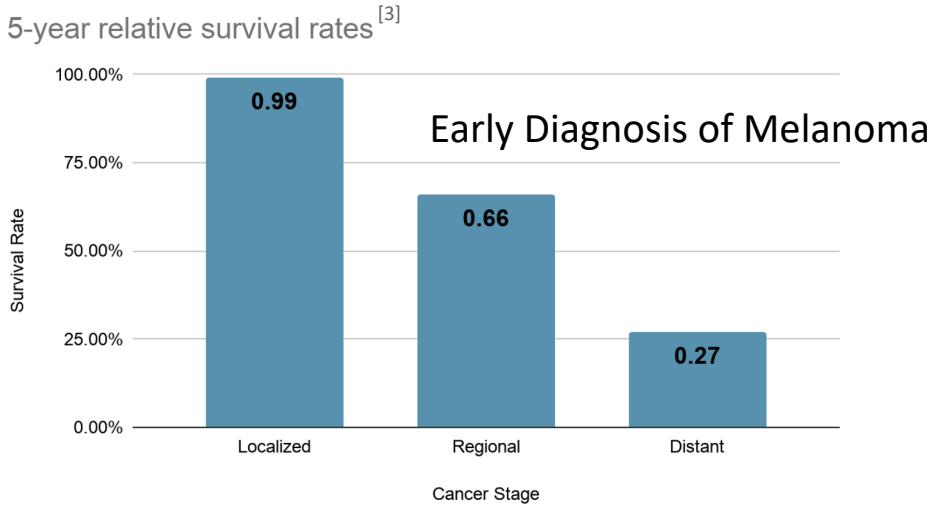
Kechun Liu

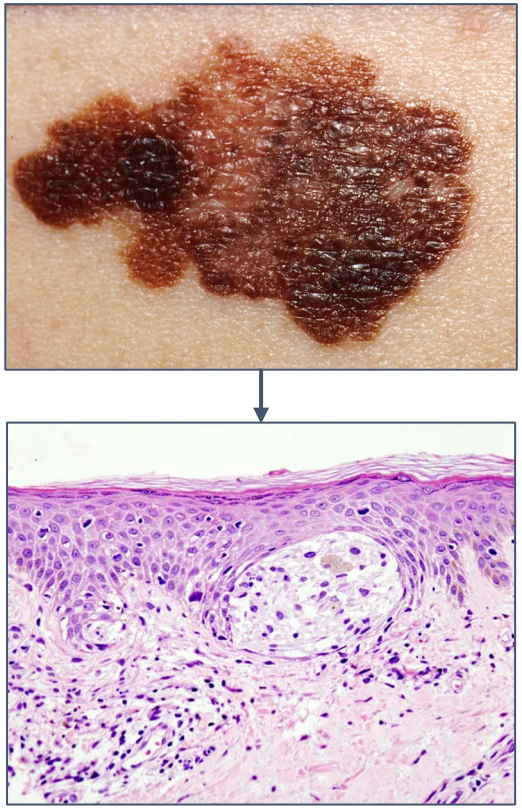
Spring 2021



What is melanoma?

- Third most common type of skin cancer[1,2]
- Responsible for most skin cancer deaths[1,2]
- >63,000 diagnosed cases and 9,000 deaths from melanoma each year in US between 2007-2011[3]

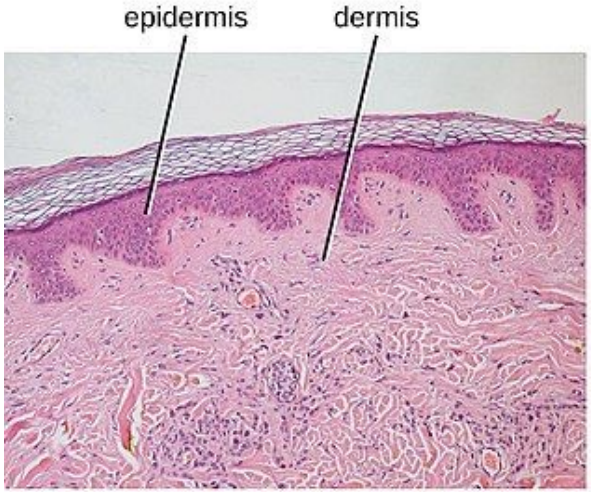


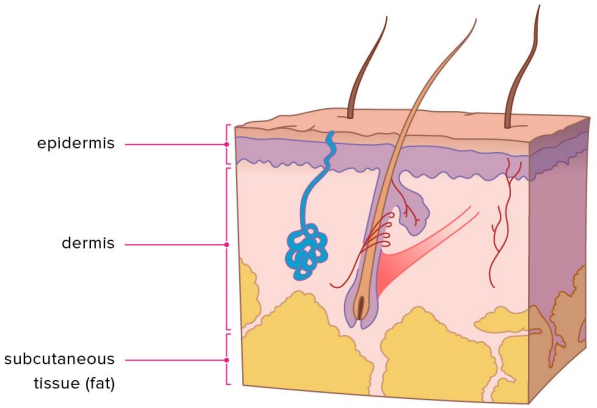


[1] Jemal, Ahmedin, et al. "Recent trends in cutaneous melanoma incidence and death rates in the United States, 1992-2006." *Journal of the American Academy of Dermatology* 65.5 (2011): S17-e1.
[2] Jemal, Ahmedin, et al. "Annual report to the nation on the status of cancer, 1975–2009, featuring the burden and trends in human papillomavirus (HPV)–associated cancers and HPV vaccination coverage levels." *JNCI: Journal of the National Cancer Institute* 105.3 (2013): 175-201.
[3] NNAM Howlader, et al. Seer cancer statistics review, 1975-2016. National Cancer Institute, 2019.

Melanoma Diagnosis

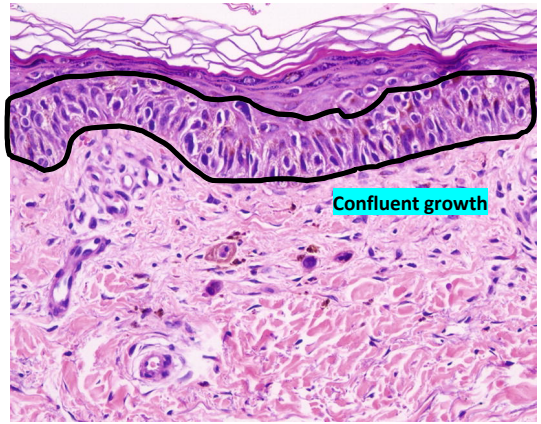
- Microscopic examination of H&E-stained biopsy images
- Assessment of architectural growth patterns
    - where are melanocytes situated? (intraepidermal, dermal-epidermal junction, intradermal)

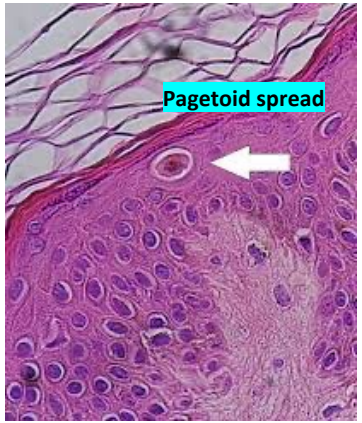




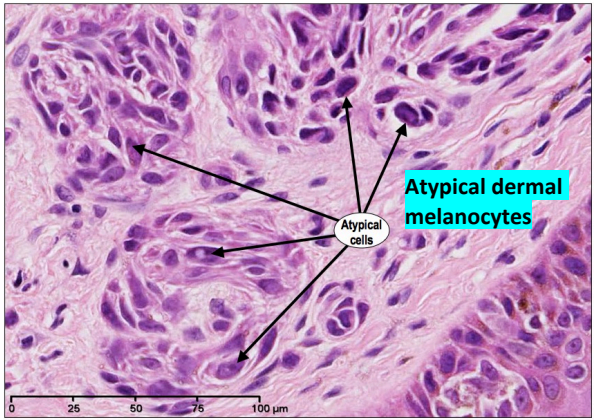
Melanoma Diagnosis

- Microscopic examination of H&E-stained biopsy images
- Assessment of architectural growth patterns
  - where are melanocytes situated? (intraepidermal, dermal-epidermal junction, intradermal)
  - architecture of melanocytic population (confluent growth? pagetoid spread? atypical dermal melanocytes?)





Melanoma in situ



Invasive (malignant) melanoma

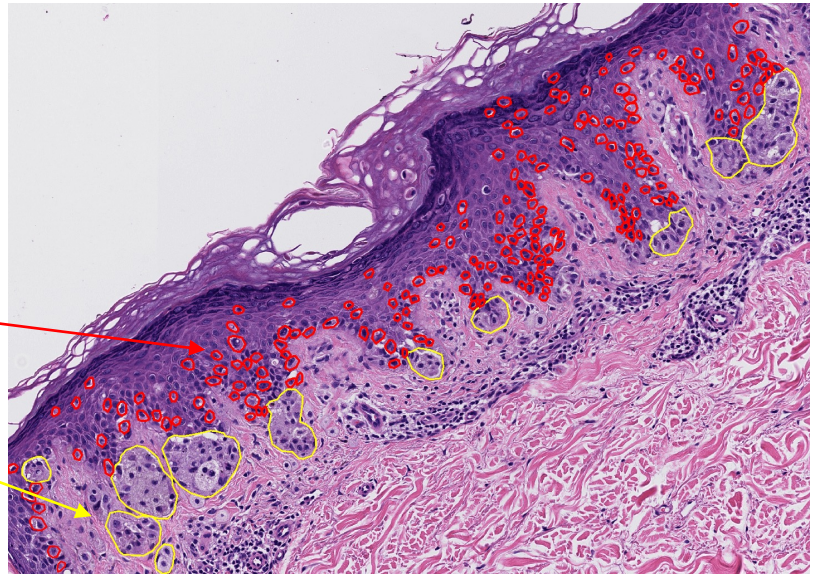
Melanoma Diagnosis

- Microscopic examination of H&E-stained biopsy images
- Assessment of architectural growth patterns
  - where are melanocytes situated? (intraepidermal, dermal-epidermal junction, intradermal)
  - architecture of melanocytic population (confluent growth? pagetoid spread? atypical dermal melanocytes?)

Melanocytic Proliferations

★ Singly dispersed melanocytes

★ Nested melanocytes



Melanoma Diagnosis

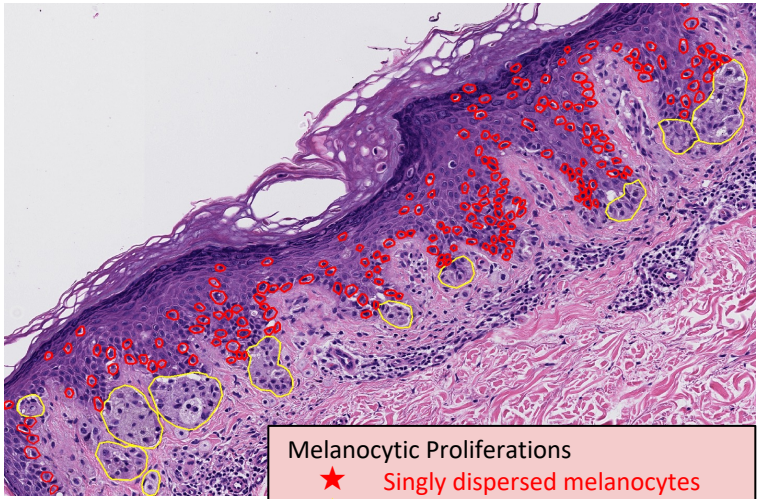
- Microscopic examination of H&E-stained biopsy images
- Assessment of architectural growth patterns
  - where are melanocytes situated? (intraepidermal, dermal-epidermal junction, intradermal)
  - architecture of melanocytic population (confluent growth? pagetoid spread? atypical dermal melanocytes?)

Can we develop a system to automatically point out melanocytic proliferations?

We developed a pipeline to identify image-level melanocytic proliferations with weak supervision.

We leverages sparse and noisy annotations on skin biopsy images and uses weighted loss functions to account for the imperfect labels.

We achieve state-of-the-art performance on segmentation of melanocytic proliferations.

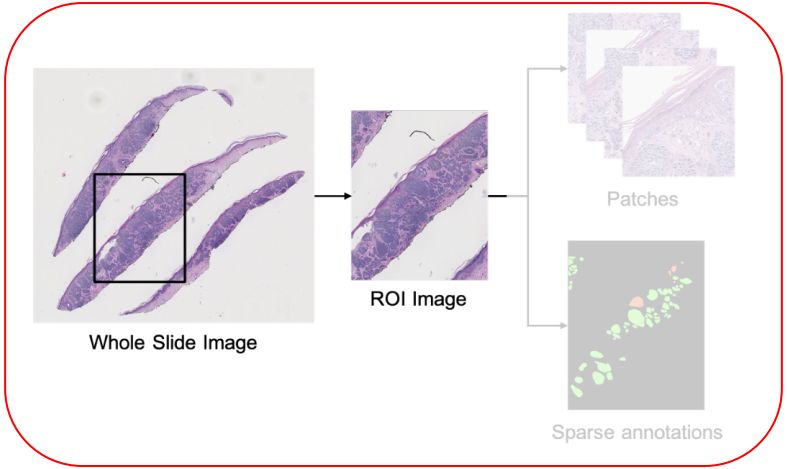


Dataset

H&E stained skin biopsy images, 10x

Consensus under 3 pathologists

227 ROI images[1]



[1] Elmore, Joann G., et al. "Pathologists' diagnosis of invasive melanoma and melanocytic proliferations: observer accuracy and reproducibility study." Bmj 357 (2017).

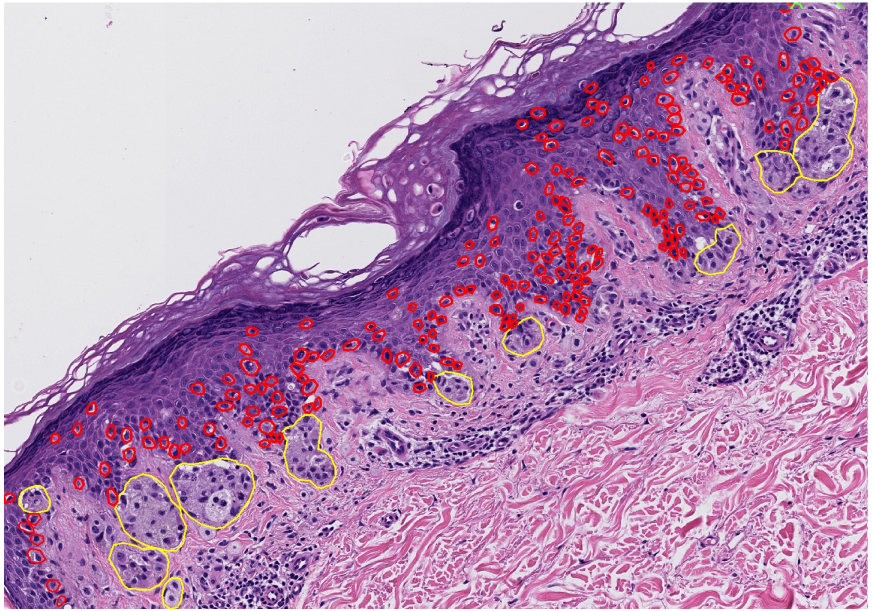
Dataset - Melanocytic Proliferation Annotations

- Difficulties in annotations:
  - Nests come in various sizes and shapes
  - Hundreds of entities
  - Expertise required

Annotation procedure:

- **Partially** mark the 227 ROI images
- Draw **polygons** around many melanocytes
- Two other pathologists check the markings

Save Annotation Time!

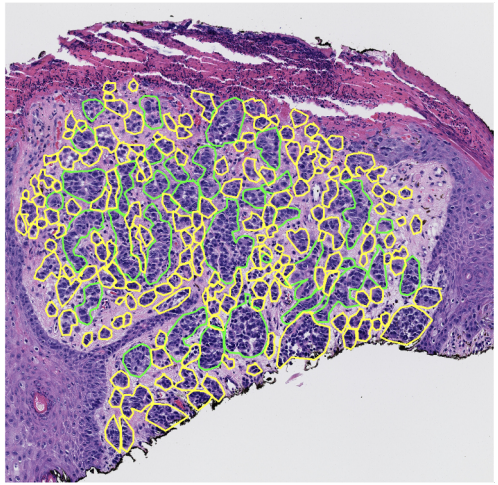


Singly dispersed melanocytes
Nested melanocytes
Annotation polygons

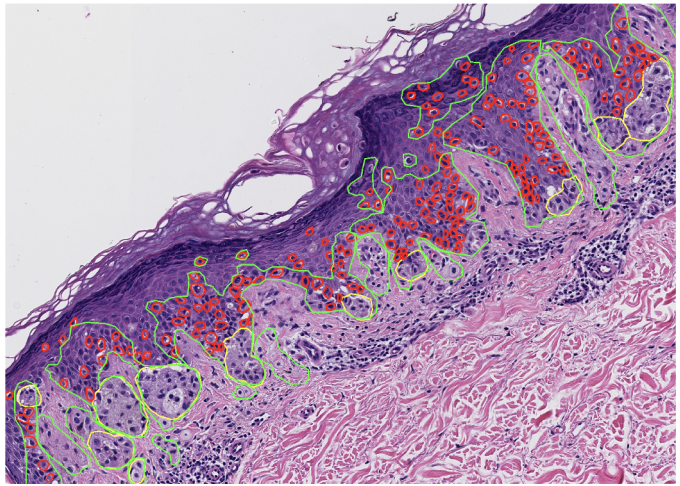
8

Dataset - Annotation Caveats

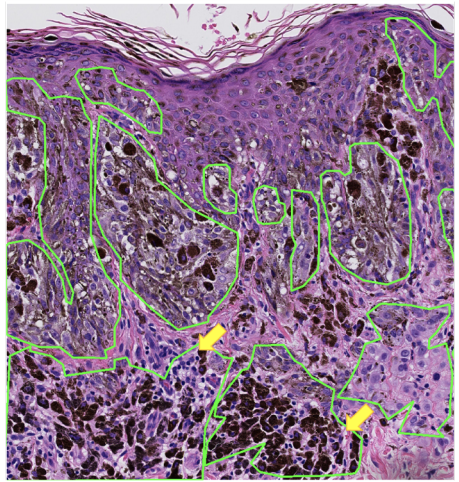
Sparse annotations



Noisy annotations



Human errors



“Silver standard”

Dataset - Preprocessing

1. Data split

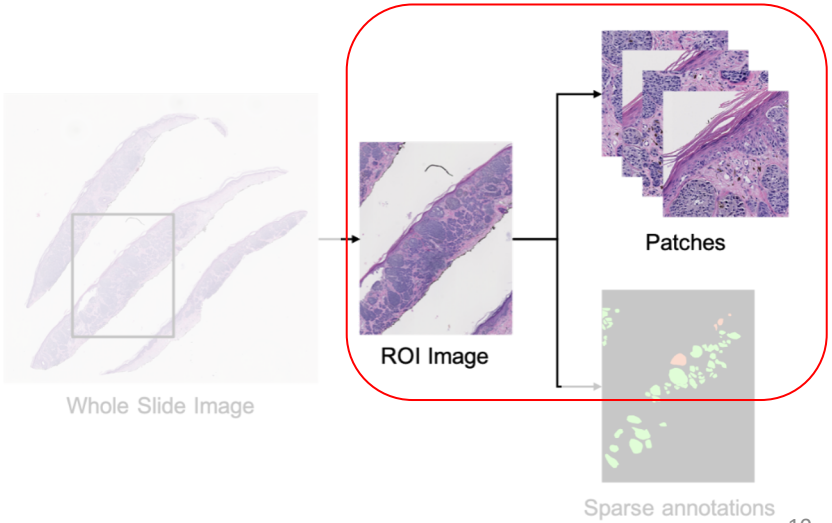
Train: 174 (76%)	Val: 19 (8%)	Test: 34 (15%)

2. Patchify

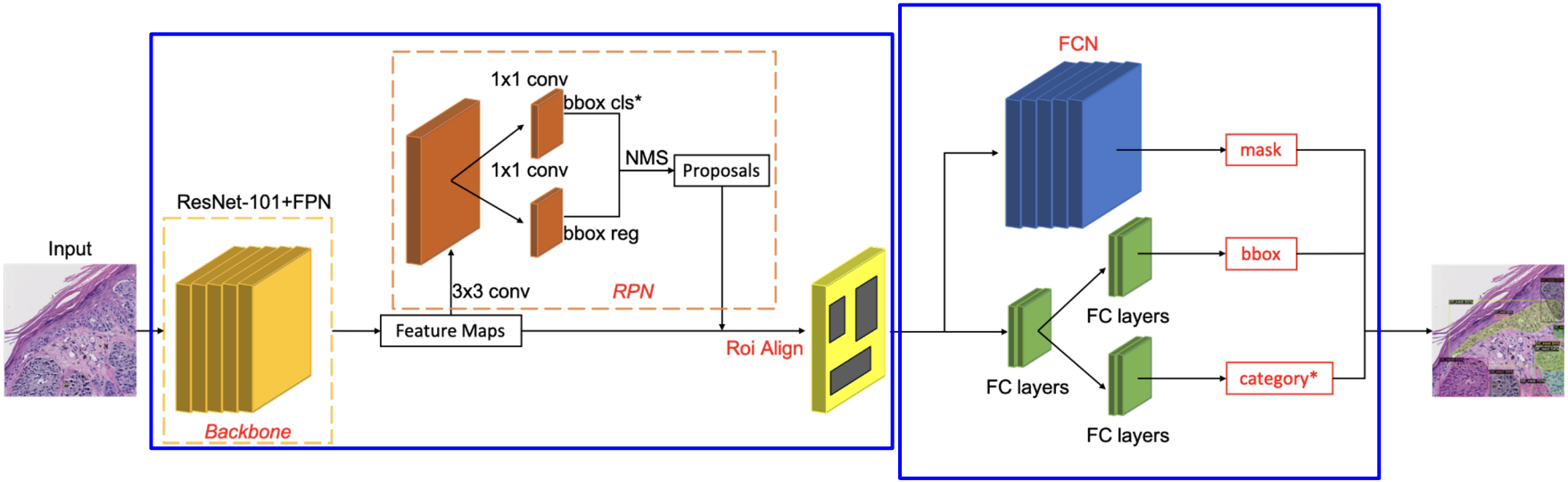
ROI: 428x381 ~ 23691x22401, 10x

Patches: 1000x1000, 5x, 50% overlap

Close to default design
in Mask R-CNN

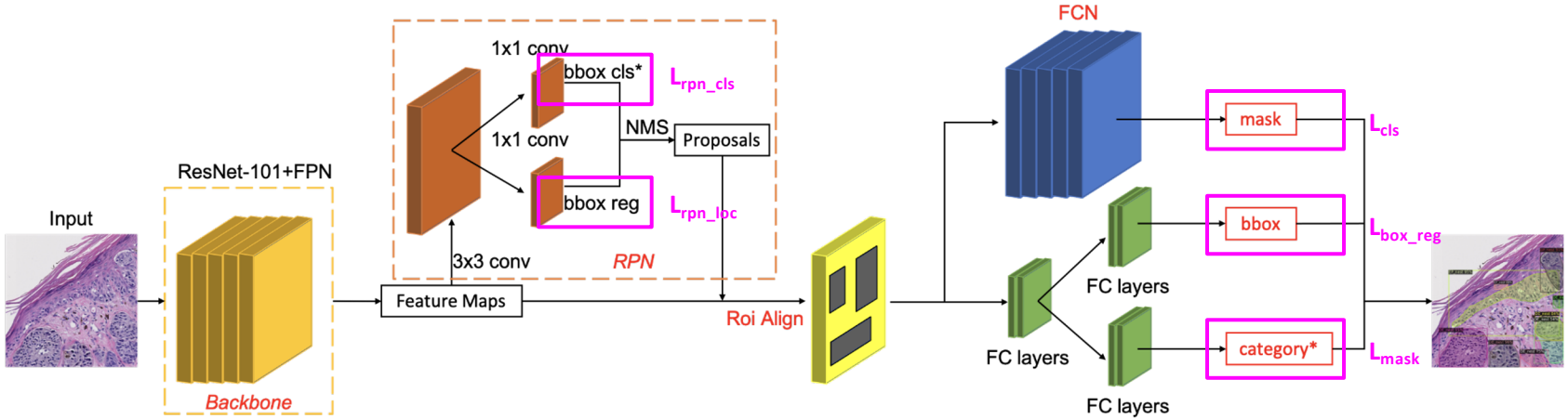


Model - Mask R-CNN

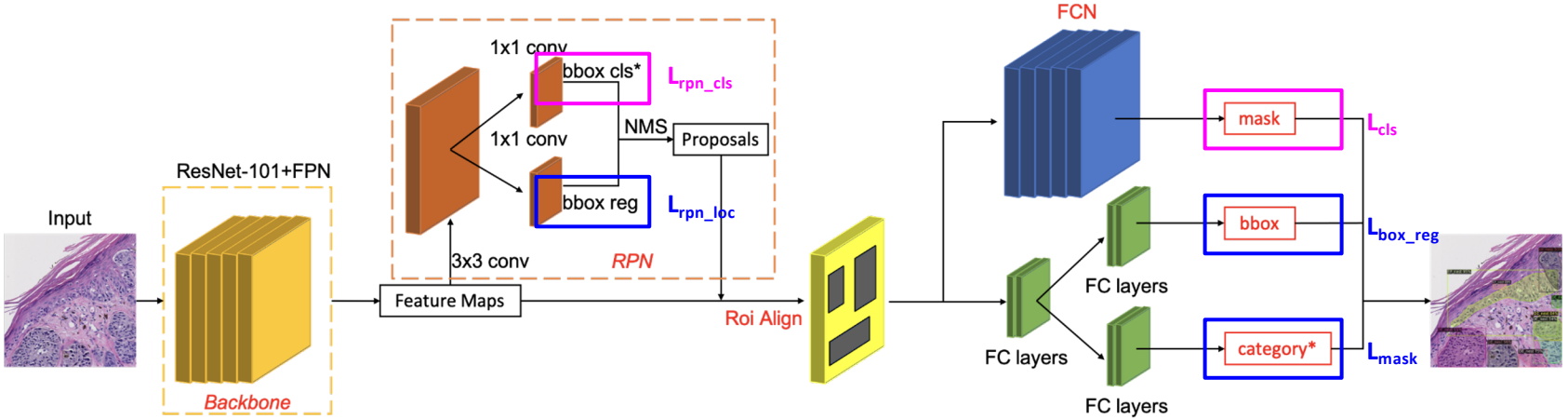


Idea: filter out the majority of the non-target tissues

Model - Loss Function



Model - Loss Function



L_{rpn_loc}, L_{box_reg}, L_{mask} : only back-propagate loss values on positive samples

L_{rpn_cls}, L_{cls} : fully utilize the labeled and unlabeled areas

Binary cross entropy
Categorical cross entropy

⇒

Weighted cross entropy
Focal loss[1]

[1] Lin, Tsung-Yi, et al. "Focal loss for dense object detection." Proceedings of the IEEE international conference on computer vision. 2017.

Weighted Cross Entropy (WCE)

$$L_{\mathrm{WCE}} = -\sum_{i} (w * y_i * \log(\hat{p}_i) + (1 - y_i) * \log(1 - \hat{p}_i))$$

$y_i \in \{0,1\}$: ground-truth label whether the object belongs to class i.
$\widehat{p_i} \in [0,1]$: probability of the object being in class i.
w: weight given to the categories.

Focal Loss (FL)[1]

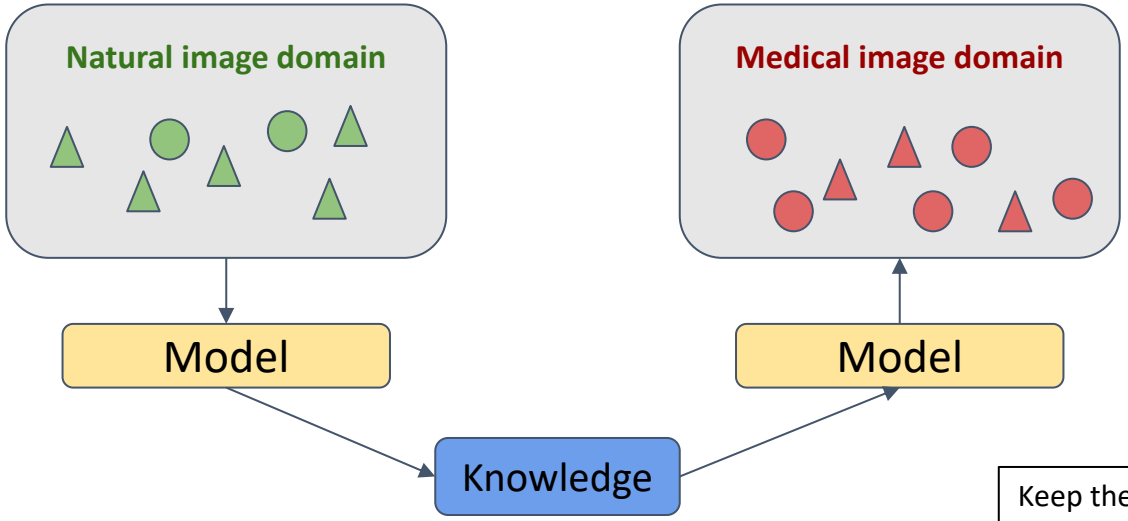
$$L_{\mathrm{WFL}} = -\sum_{i}(w * y_i * (1-\hat{p}_i)^{\lambda} * \log(\hat{p}_i) + (1-y_i) * \hat{p}_i{}^{\lambda} \log(1-\hat{p}_i))$$

$y_i \in \{0,1\}$: ground-truth label whether the object belongs to class i.
$\widehat{p_i} \in [0,1]$: probability of the object being in class i.
w: weight given to the categories.
λ: the larger λ is, the more the model focuses on hard examples. (λ=2)

[1] Lin, Tsung-Yi, et al. "Focal loss for dense object detection." Proceedings of the IEEE international conference on computer vision. 2017.

Model - Transfer Learning

Lack of accurately annotated training data: 130 images in train set!

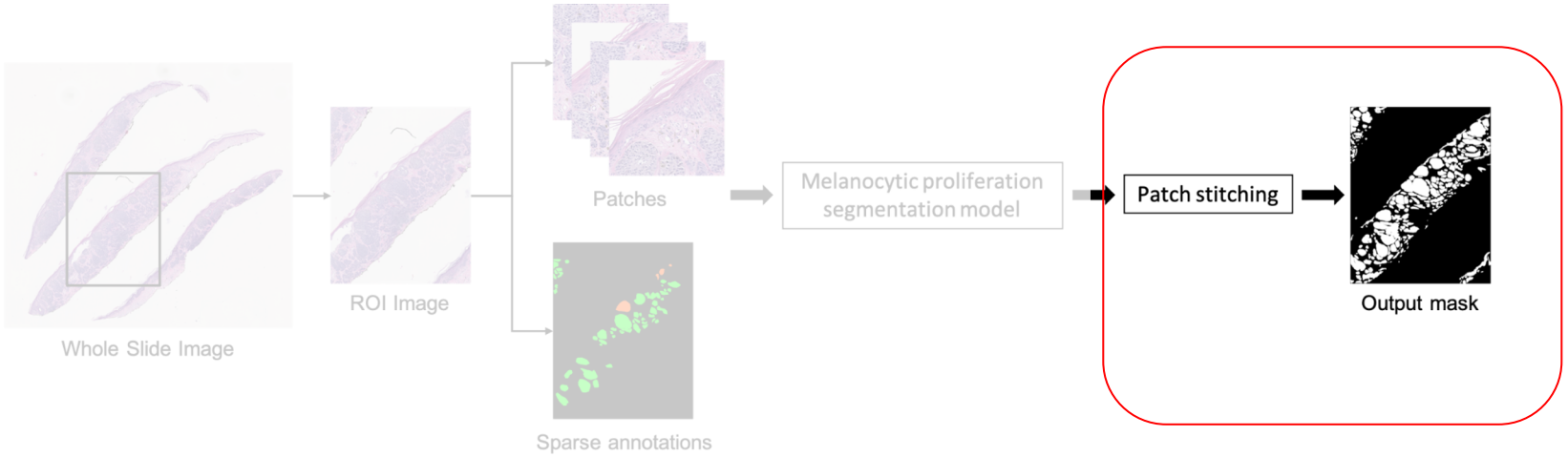


Keep the pretrained parameters as much as possible!

Mask R-CNN from detectron2[Wu et al.]: pretrained on MSCOCO

Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. Detectron2. https://github.com/facebookresearch/detectron2, 2019.

Model - post processing



Patch-level segmentation results ⇒ Image-level segmentation result

Model - Implementation details

- SGD optimizer
  - Initial learning rate: 0.001; learning rate warm-up; 0.5 decay every 4 epochs
  - Total 40 epochs
- Loss
  - Weighted cross entropy
  - Focal loss
    - Weight: 1, 2, 3, 5, 8, 12
- Run each model 10 times with different randomization

Evaluation metrics

$$\text{Dice} = \frac{2 \times TP}{2 \times TP + FP + FN}$$

$$\text{mIOU} = \frac{1}{2} \times (\mathcal{J}_{\text{nest}} + \mathcal{J}_{\text{bg}})$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

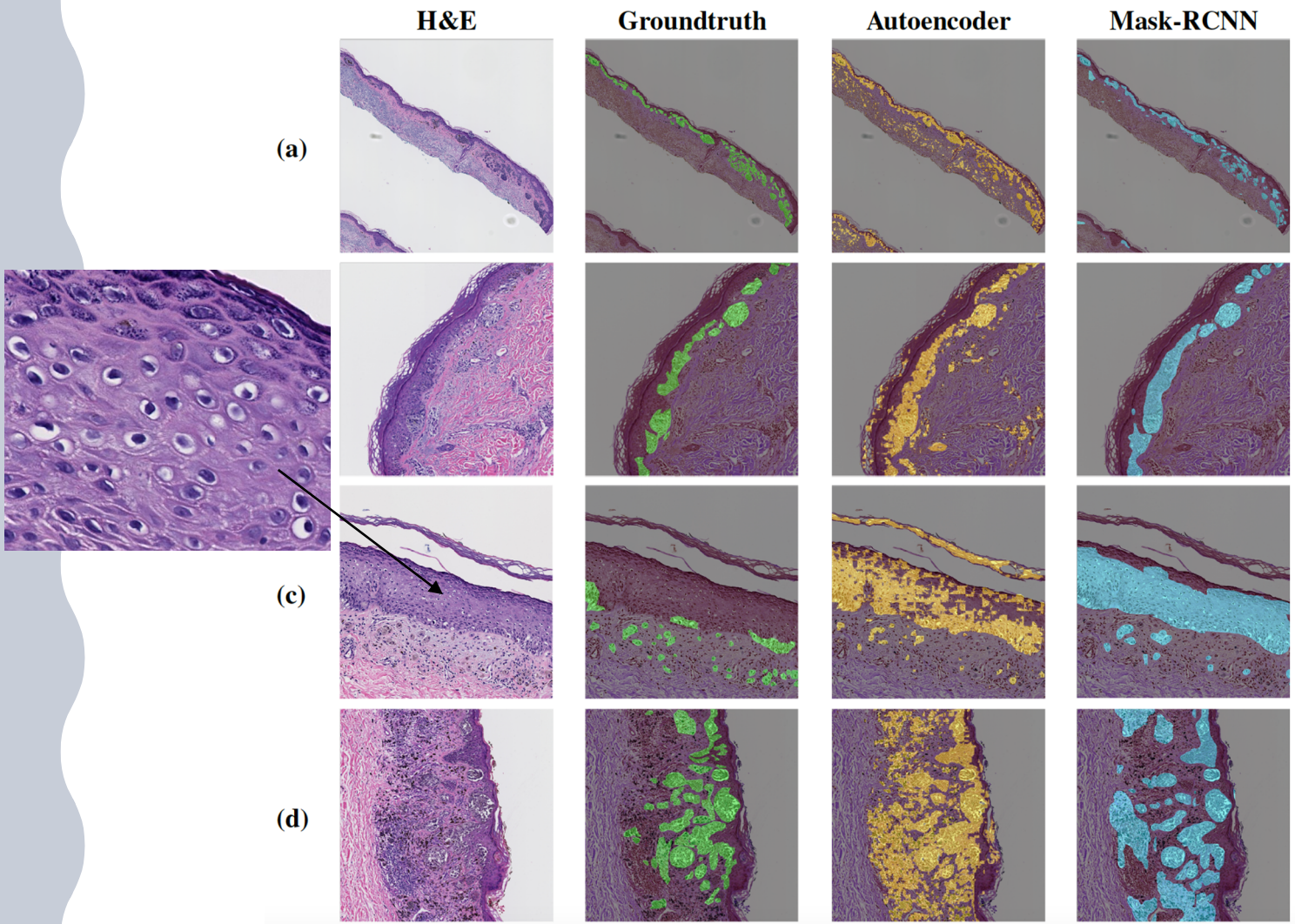
$$\text{Specificity} = \frac{TN}{TN + FP}$$

All metrics are reported in **mean** and **standard deviation**.

Experimental results

- We fully label the melanocytic nests in our test set (34 ROI images).
- We re-implemented the convolutional autoencoder (previous SOTA work[Kucharski et al.]).
- We achieve better performance in Dice score, mIOU, accuracy and specificity.

Method	Dice	mIOU	Accuracy	Sensitivity	Specificity
Autoencoder [21]	0.679	0.705	0.905	**0.814**	0.918
Mask R-CNN with CE loss	0.685	0.715	0.917	0.726	0.944
Mask R-CNN with WCE loss	0.705	0.726	0.917	0.792	0.935
Mask R-CNN with FL loss	**0.719**	**0.740**	**0.927**	0.751	**0.952**



Good result

Imperfect results

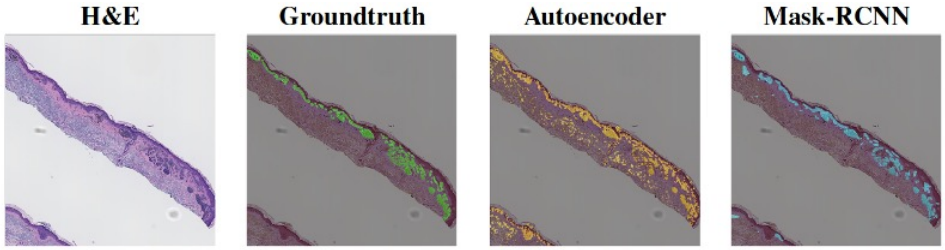
Ablations

Loss function	Weight	Dice	mIOU	Accuracy	Sensitivity	Specificity
Weighted Cross Entropy (WCE)	$w = 1$	0.685(0.013)	0.715(0.008)	0.917(0.002)	0.726(0.041)	0.944(0.006)
	$w = 2$	**0.705(0.003)**	**0.726(0.003)**	**0.917(0.003)**	**0.792(0.027)**	**0.935(0.007)**
	$w = 3$	0.701(0.009)	0.723(0.006)	0.915(0.003)	0.792(0.021)	0.933(0.005)
	$w = 5$	0.701(0.008)	0.722(0.006)	0.914(0.002)	0.813(0.028)	0.928(0.005)
	$w = 8$	0.700(0.007)	0.718(0.007)	0.909(0.005)	0.850(0.022)	0.918(0.008)
	$w = 12$	0.700(0.005)	0.716(0.003)	0.908(0.002)	0.847(0.021)	0.917(0.005)
Focal Loss (FL)	$w = 1$	0.717(0.018)	0.740(0.011)	0.928(0.002)	0.740(0.053)	0.954(0.007)
	$w = 2$	0.703(0.022)	0.731(0.014)	0.926(0.003)	0.710(0.053)	0.956(0.006)
	$w = 3$	0.702(0.021)	0.730(0.014)	0.926(0.003)	0.705(0.045)	0.957(0.004)
	$w = 5$	0.711(0.014)	0.735(0.008)	0.926(0.002)	0.730(0.044)	0.954(0.006)
	$w = 8$	**0.719(0.011)**	**0.740(0.007)**	**0.927(0.003)**	**0.751(0.027)**	**0.952(0.005)**
	$w = 12$	0.710(0.023)	0.734(0.015)	0.925(0.004)	0.742(0.056)	0.951(0.007)

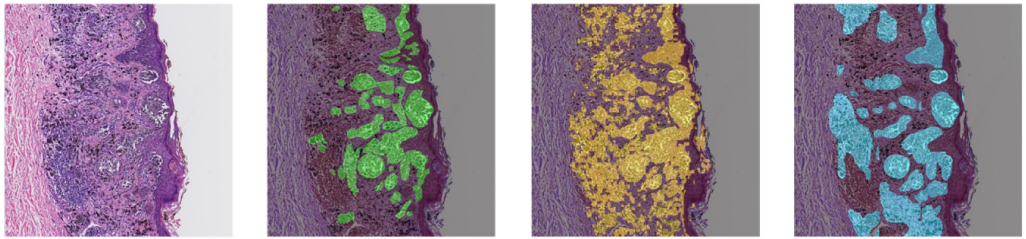
Larger STD

- Adding weights helps improve performance.
- Noise is also amplified when using focal loss.

Discussions

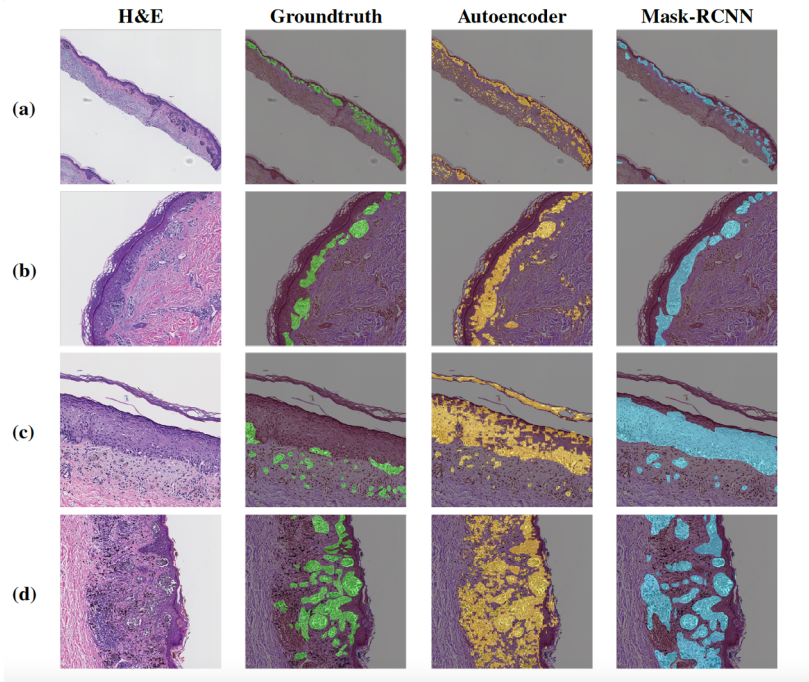


- Why Mask R-CNN?
  - Robust to noise
- How does this work serve to help diagnosis?
  - First step of an automated diagnosis pipeline
  - Combine features with classification techniques to create a diagnosis tool
- How does annotation quality affect the performance?
  - Reduce human errors by leveraging our model's output



Conclusion

- We propose a weakly-supervised Mask R-CNN-based model for melanocytic proliferations segmentation.
- Our model only requires partially labeled datasets by leveraging weak supervision.
- Our approach achieves state-of-the-art accuracy on identification of melanocytic proliferations.



Thank you!