Deep learning and energy models for fine dead wood segmentation

Jacquelyn Shelton

Hong Kong Polytechnic University

with Przemyslaw Polewski, Marco Heurich and Wei Yao

Maching Learning for Climate

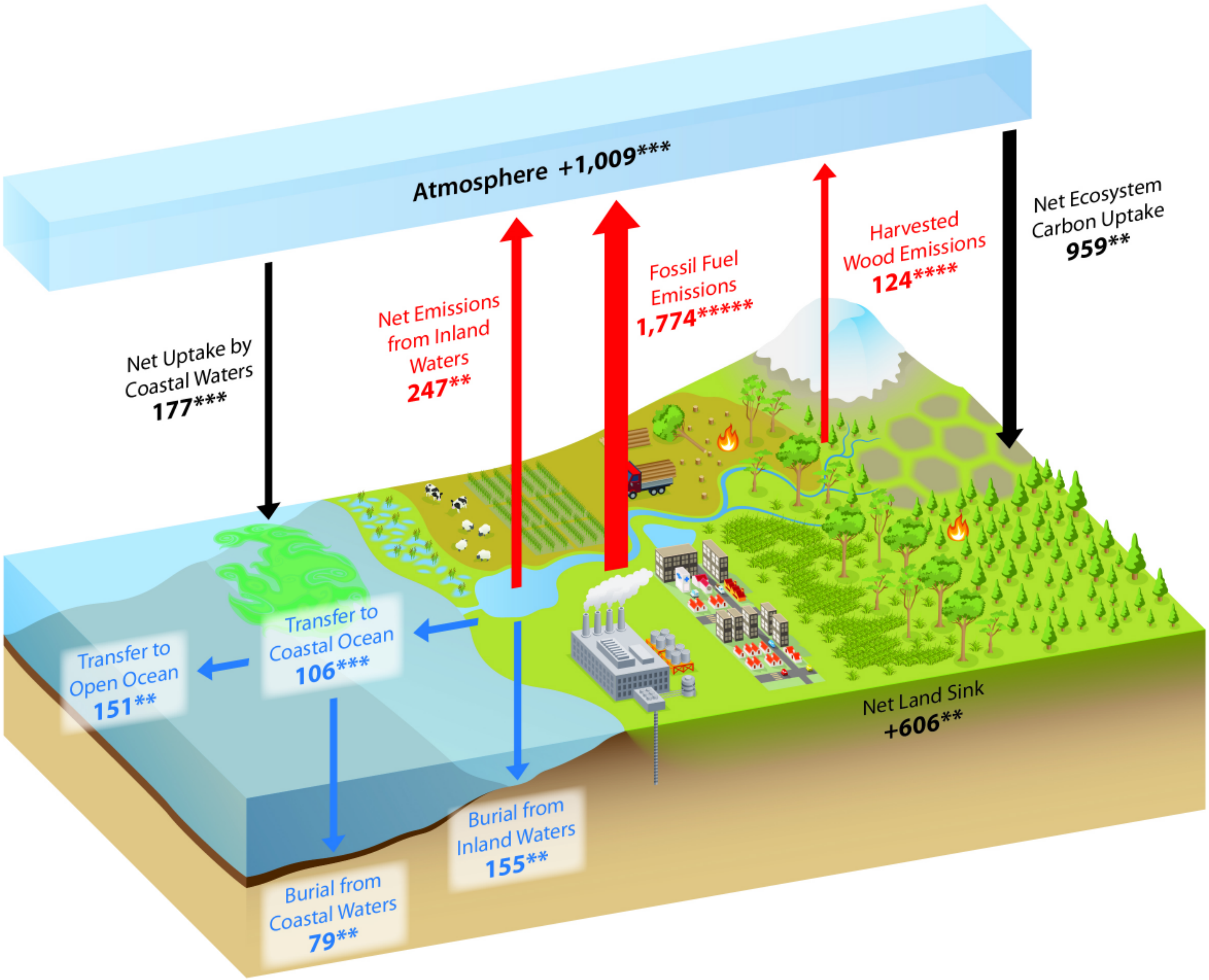
UCSB - KITP

November 4th, 2021

THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

TOMTOM

Introduction - The Carbon Cycle



Introduction - Motivation

Why dead trees?

- **Forests** are a key component in the **global carbon cycle** *and* the most **efficient** CO2 scrubbers
- **Dead wood** is an important part of **forest ecosystems**
  - **Sequesters** 8% of global forest carbon and 8.5% atmospheric carbon
  - **Provides habitat** for $\sim \frac{1}{3}$ of plant and animal species $\rightarrow$ imperative for **biodiversity**
- **Forest biodiversity** helps maintain forest health and shield against **climate change**

Introduction - Motivation

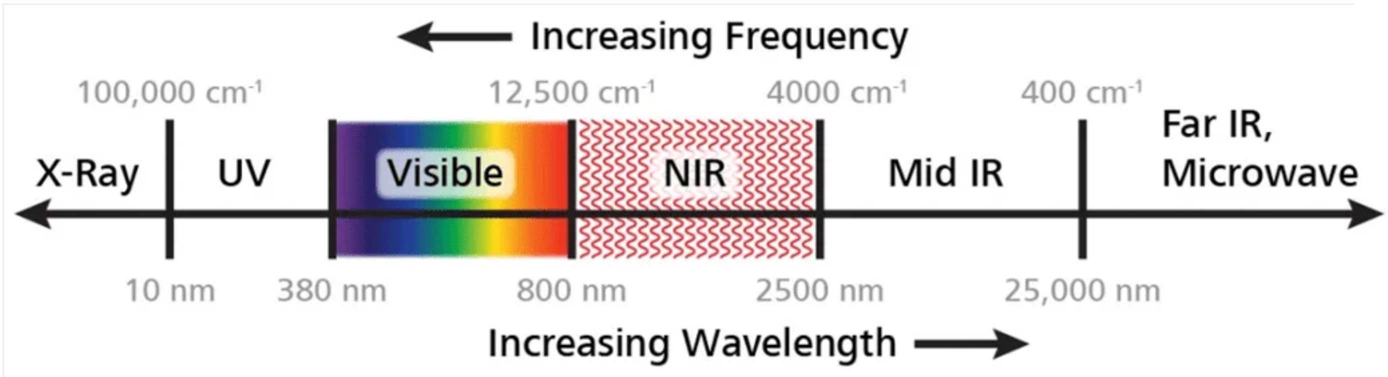
Why dead trees?

- **Forests** are a key component in the **global carbon cycle** *and* the most **efficient** CO^2 scrubbers
- **Dead wood** is an important part of **forest ecosystems**
  - **Sequesters** 8% of global forest carbon and 8.5% atmospheric carbon
  - **Provides habitat** for $\sim \frac{1}{3}$ of plant and animal species $\rightarrow$ imperative for **biodiversity**
- **Forest biodiversity** helps maintain forest health and shield against **climate change**
- **Basically: indicator of overall forest health**...and beyond

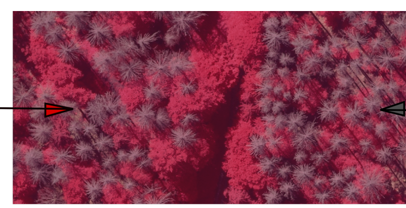
Now what?

- **Need** accurate ways to *measure* and *monitor* dead wood quantities and types...**Machine Learning!**

Data - Color infrared imagery (CIR)

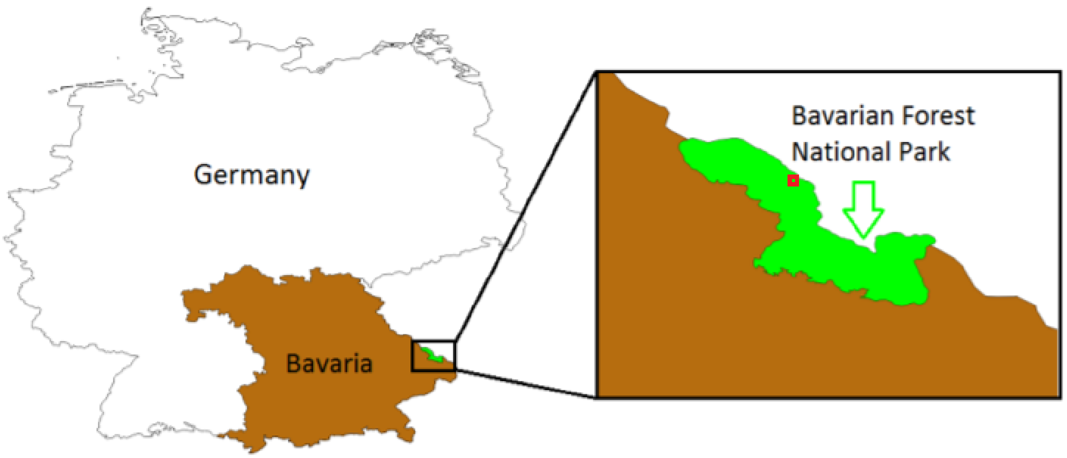


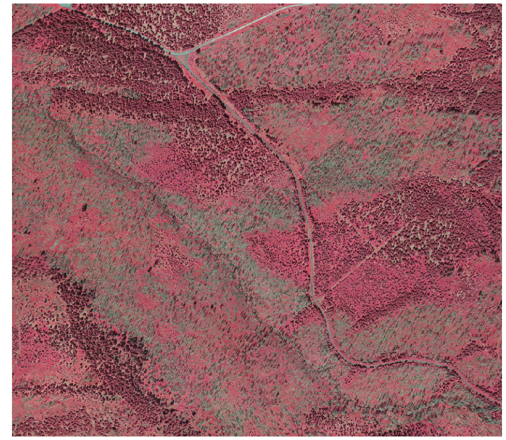
- Distinguish between live and dead vegetation
  $\rightarrow$ chlorophyll has high reflectance in the near-infrared (NIR) spectral band:



- Color infrared imagery (CIR): consists of the NIR, red, and green bands instead of the usual RGB

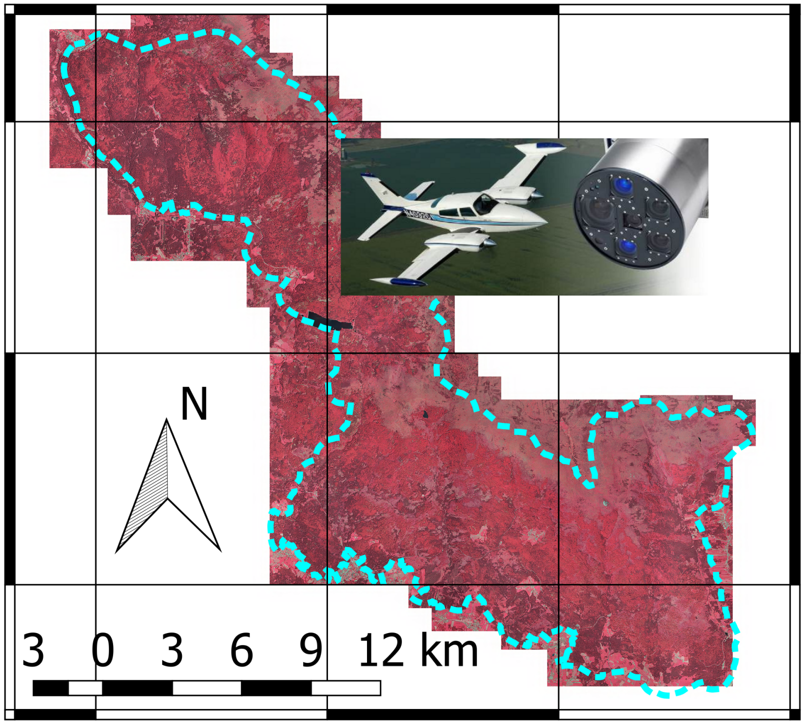
Data - Bavarian Forest National Park

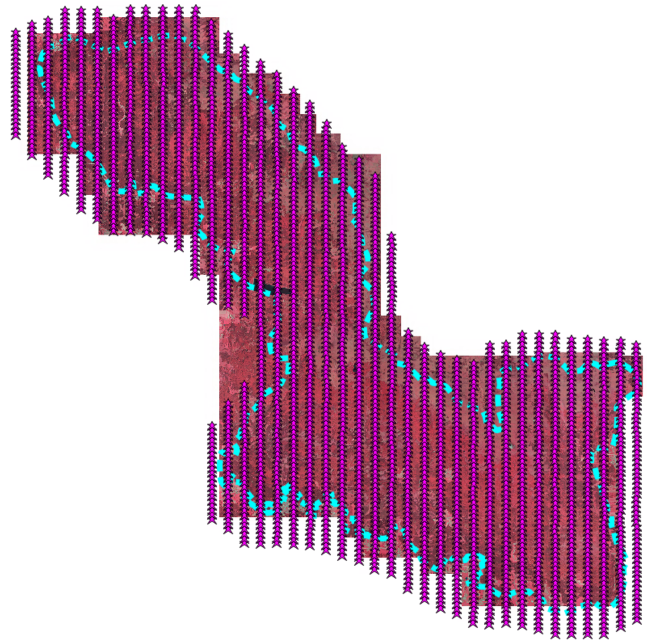




- Located in southeastern Germany, bordering the Czech Republic
- Consists of mostly Norway spruce (Picea abies) and European beech (Fagus sylvatica)
- Suffered bark beetle infestation (Ips typographus): between 1988-2010, total of $5,800$ hectors of the Norway spruce stands died
- Ideal for dead wood studies – decaying wood left undisturbed in forest for scientific purposes

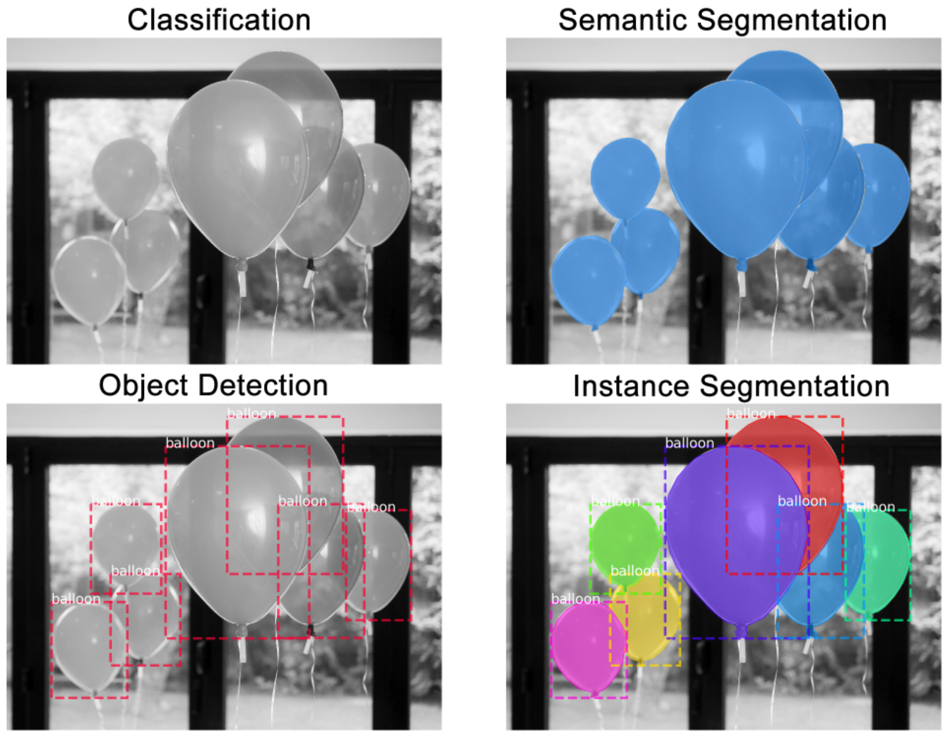
Data - Aerial imagery of Bavarian Forest National Park





- Flight campaign in June 2017 acquired $\sim 2,500$ images
- Mean flight altitude: $2879m$, resulting in a high resolution ground sampling distance of $10cm$ $\rightarrow$ each pixel $= 10 \times 10cm$
- Image resolution of aerial camera: 14592×25728 pixels, with each image covering $\sim 3.7km^2$

Approach – Review of classic segmentation tasks

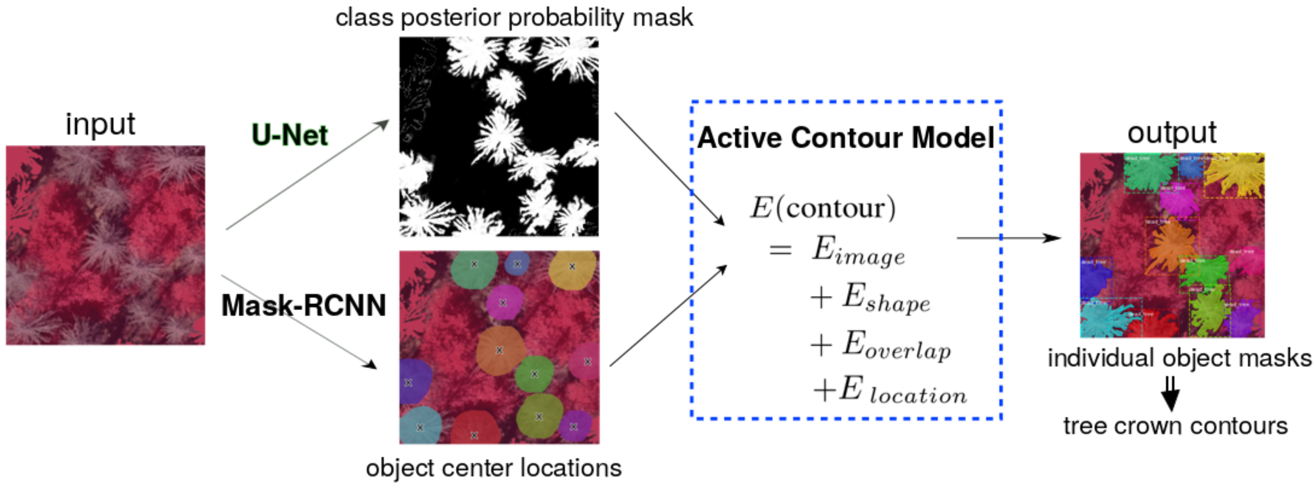


Source: "Splash of color" tutorial by Waleed Abdulla

- Classification: balloon is present in image (or not)
- Semantic Segmentation: find *all* the balloon pixels
- Object Detection: find balloon *quantity* and *location*, e.g. 7 balloons here
- **Instance Segmentation**: identify all pixels belonging to *each* of 7 balloons

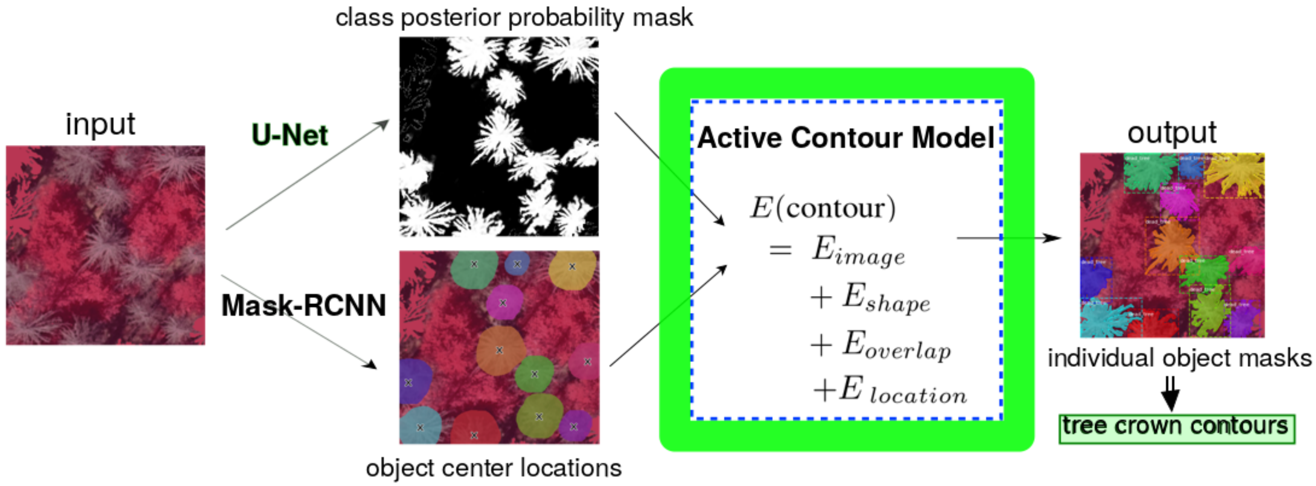
→ Note: **encountering objects that overlap** – our approach **MUST** account for this

Approach – How to attack this problem?



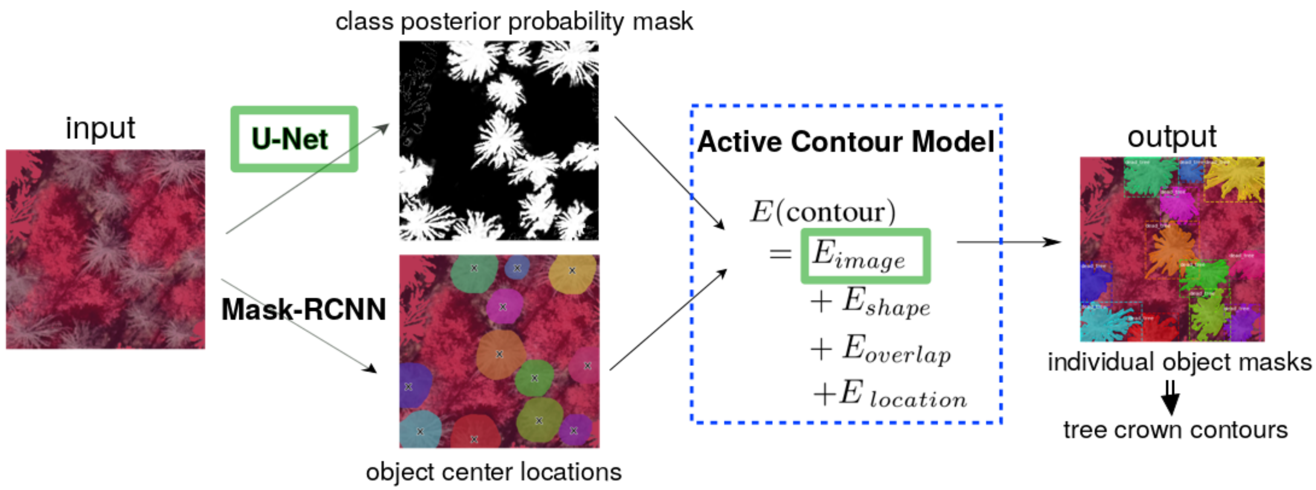
- First, let's break the problem down into basic components
- Next, recognize we can employ state-of-the-art deep learning methods to form the building blocks of our approach

Approach - Build an energy model!



- First, let's break the problem down into basic components
- Next, recognize we can employ state-of-the-art deep learning methods to form the building blocks of our approach
- Combine these methods to formulate a multi-term **energy model** for refined contour segmentation

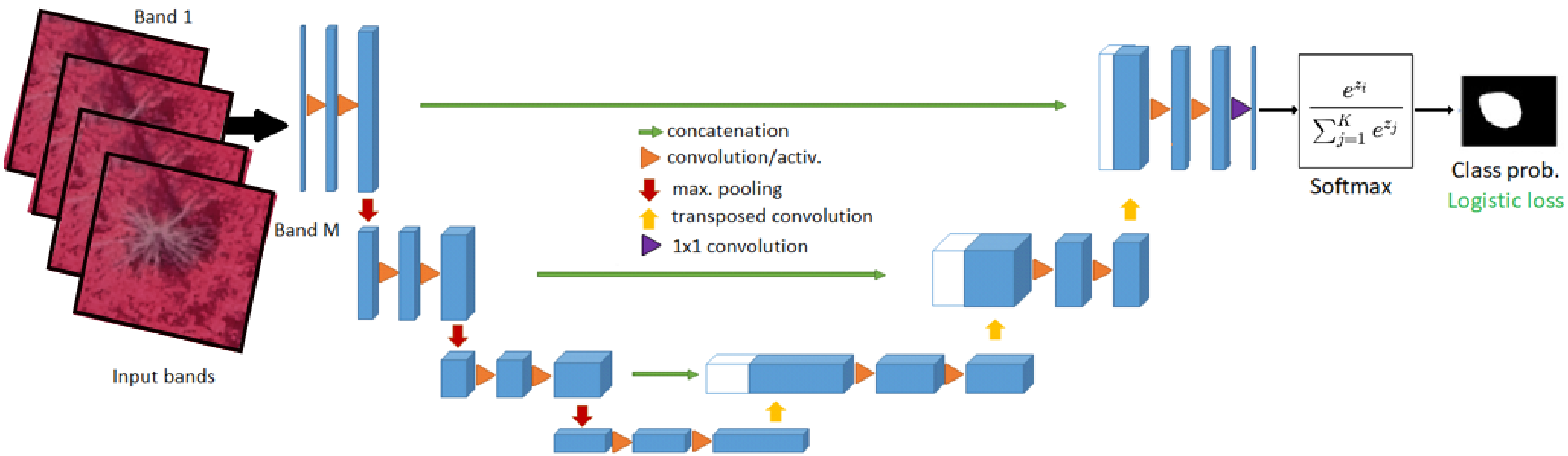
Approach – U-net: energy term for image



Build each energy term: $E_{\,image}(\text{contour})$:

- **U-net:** fully convolutional neural network (FCN) with an *encoder-decoder architecture*
  $\rightarrow$ get class probabilities aka dense ***semantic segmentation***
- Suited for this task: preserves contours, thus fine details of the tree crowns

Approach - U-net: energy term for image

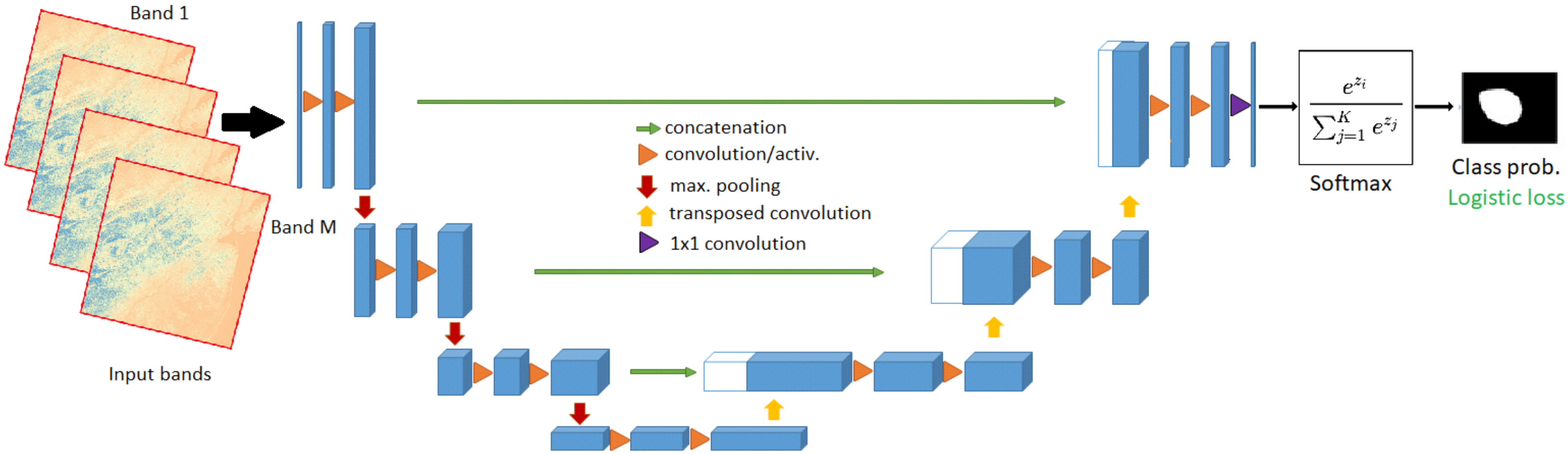


- **U-net:** FCN with encoder-decoder architecture
  - Encoding branch derives convolutional features at successively coarser scales
  - Decoding branch upsamples through transposed convolution

- Preserves detail with high-resolution original feature maps

$\rightarrow$ Additionally, U-nets can also be used for regression...

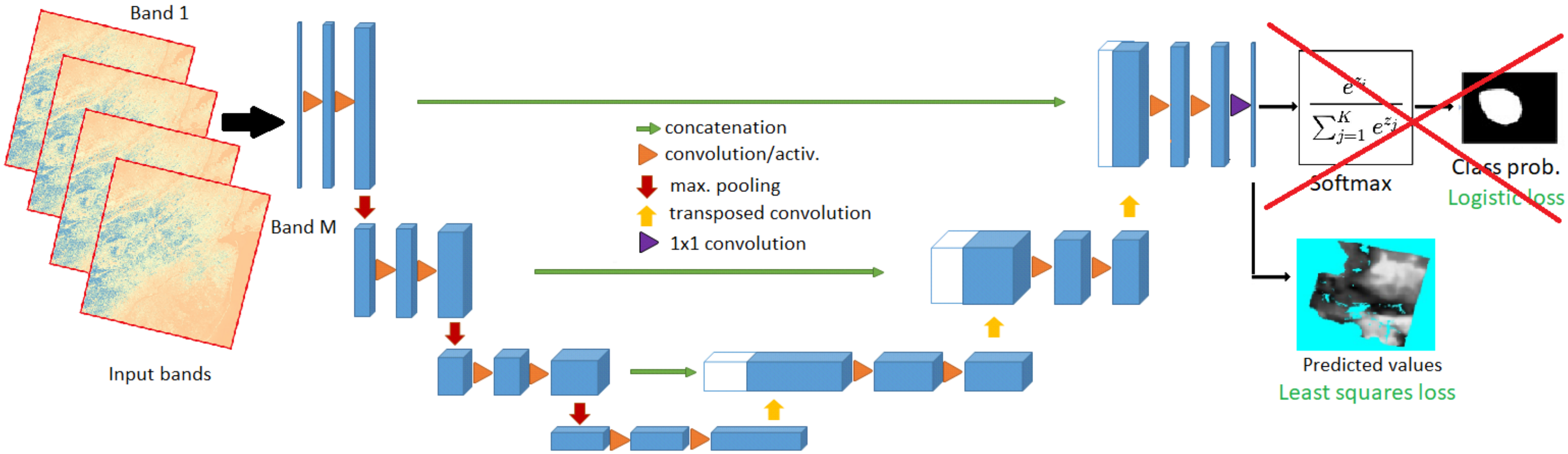
Approach - *Side-bar*: U-net for regression



- **U-net** for regression: remove soft-max layer and logistic loss, which gave us the class probabilities

Shelton, J., Polewski, P., Yao, W., (2020). *U-Net For Learning And Inference Of Dense Representation Of Multiple Air Pollutants From Satellite Imagery. Climate Informatics, Virtual Worldwide (9 2020).*

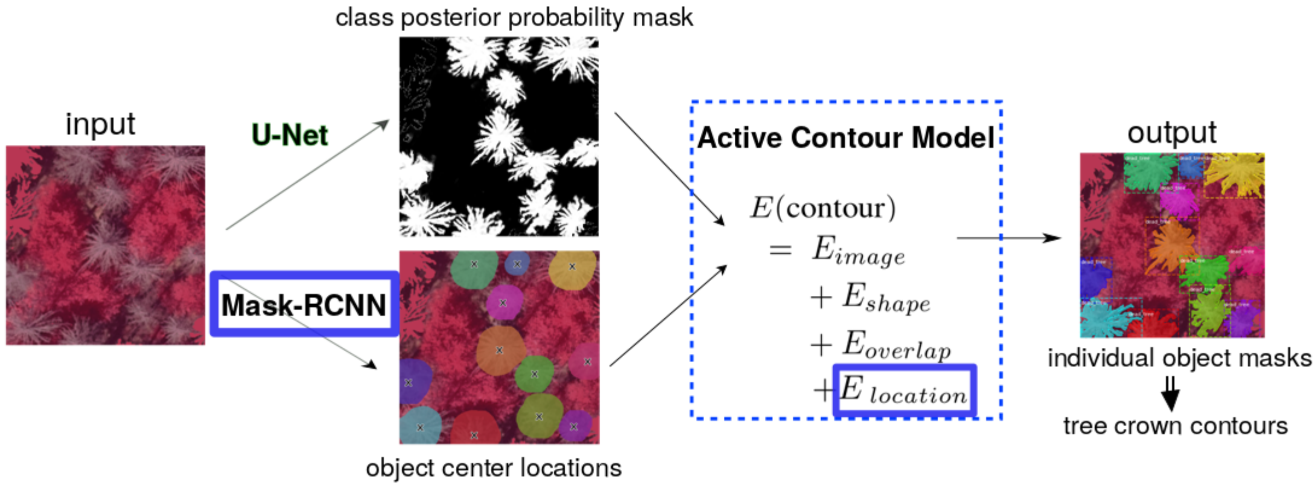
Approach – *Side-bar*: U-net for regression



- **U-net** for regression: remove softmax layer and logistic loss, which gave us the class probabilities
- Instead, train with least-squares loss directly on the 1×1 convolution output (final layer) for dense/pixel predictions
- Our previous work: using a U-net for regression we successfully predicted dense estimates of atmosphere pollutants (NO_3 , NH_4 , $PM_{2.5}$)

Shelton, J., Polewski, P., Yao, W., (2020). *U-Net For Learning And Inference Of Dense Representation Of Multiple Air Pollutants From Satellite Imagery*. *Climate Informatics, Virtual Worldwide (9 2020).*

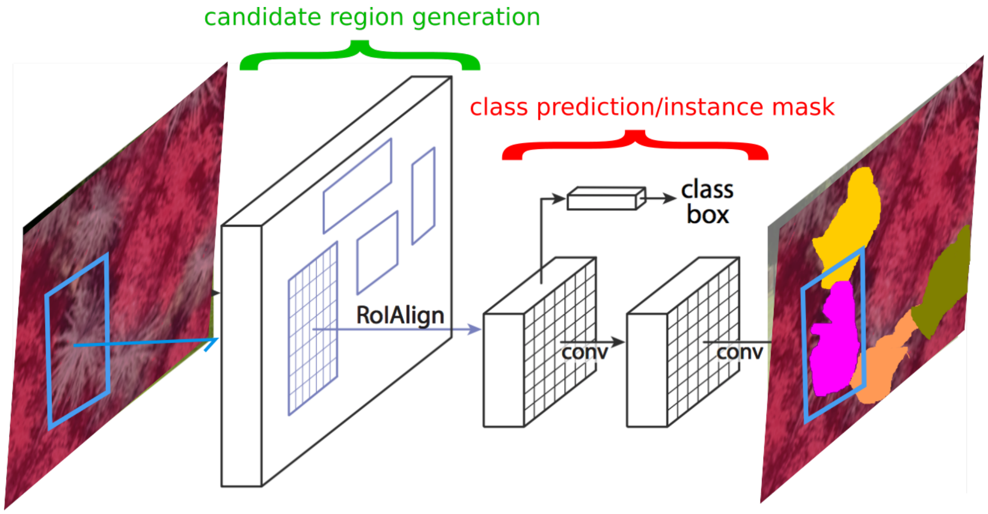
Approach – Mask R-CNN: energy term for location



Build next energy term: $E_{location}(\text{contour})$:

- **Mask R-CNN:** FCN for ***instance segmentation*** – get separate pixel masks for all object instances present in input image:
  (1) region proposal network: select regions of interest (RoI) likely to contain object instances
  (2) fine-grained detection component: evaluate candidate RoI, predict object class, bounding box, and pixel (instance) mask

Approach – Mask R-CNN: energy term for location



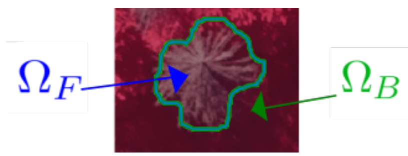
Build next energy term: $E_{\,location}(\text{contour})$:

- **Mask R-CNN:** FCN for ***instance segmentation*** – get separate pixel masks for all object instances present in input image:
  (1) region proposal network: select regions of interest (RoI) likely to contain object instances
  (2) fine-grained detection component: evaluate candidate RoI, predict object class, bounding box, and pixel (instance) mask

Approach - Formalize energy model

Begin building our energy model and its terms/potentials:

- Consider image plane $\Omega \subset \mathbb{R}^2$ and image $I : \Omega \mapsto \mathbb{R}^D$
- Segmentation of a **single** contour C:
  C: evolving shape contour, partitions image into
  foreground Ω_F, background Ω_B $(C \longleftrightarrow \Omega_F, \Omega_B)$



Note: problem **easier** with only **1 contour** and **no overlap**

Express contour model to optimize (probability or energy):

$$\mathcal{P}(\Omega_F, \Omega_B | I) \propto \mathcal{P}(I | \Omega_F, \Omega_B) P(\Omega_F, \Omega_B)$$

$$\Leftrightarrow \quad E(C) = \underbrace{-\log \mathcal{P}(I | \Omega_F, \Omega_B)}_{\text{image term}} \underbrace{-\log P(\Omega_F, \Omega_B)}_{\text{shape term}}$$

Approach - Formalizing energy model: image term

- Reminder: C uniquely identifies Ω_F, Ω_B ($C \longleftrightarrow \Omega_F, \Omega_B$)

$$E(C) = \underbrace{-\log \mathcal{P}(I|\Omega_F, \Omega_B)}_{\text{image term: } E_{img}} \underbrace{-\log \mathcal{P}(\Omega_F, \Omega_B)}_{\text{shape term}}$$

Simplifying assumption:

- Labelings of foreground and background regions are independent: $\mathcal{P}(I|\Omega_F, \Omega_B) = \mathcal{P}(I|\Omega_F)\mathcal{P}(I|\Omega_B)$
- Let $\mathbb{1}_X$ denote the indicator function for $\omega \in X$

$$E_{img}(C) = -\int_{\omega\in\Omega} [\overbrace{\mathbb{1}_{\Omega_F}}^{[\![\omega\in\Omega_F]\!]} \log \mathcal{P}(I(\omega)|\Omega_F) + \overbrace{\mathbb{1}_{\Omega_B}}^{[\![\omega\in\Omega_B]\!]} \log \mathcal{P}(I(\omega)|\Omega_B)]d\omega$$

$$= -\int_{\omega\in\Omega} [\mathbb{1}_{\Omega_F} \log \mathcal{P}(I(\omega)|\Omega_F) + \underbrace{(1 - \mathbb{1}_{\Omega_F})}_{\text{indicator for } \omega\in\Omega_B} \log \mathcal{P}(I(\omega)|\Omega_B)]d\omega$$

Approach – Formalizing energy model: image term

$$E_{img}(C) = -\int_{\omega\in\Omega} [\mathbb{1}_{\Omega_F} \log \mathcal{P}(I(\omega)|\Omega_F) + \underbrace{(1-\mathbb{1}_{\Omega_F})}_{\text{indicator for } \omega\in\Omega_B} \log \mathcal{P}(I(\omega)|\Omega_B)] d\omega$$

$$= -\int_{\omega\in\Omega} [\mathbb{1}_{\Omega_F} [\log \frac{\mathcal{P}(I(\omega)|\Omega_F)}{\mathcal{P}(I(\omega)|\Omega_B)}] d\omega + \text{constant}$$

Simplifying assumption:

- Prior probability of observing foreground/background is equal: $P(\Omega_F) = P(\Omega_B)$

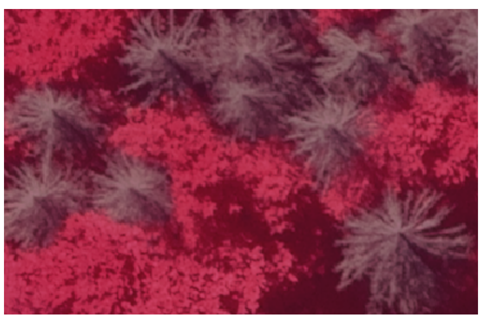
$$\frac{\mathcal{P}(I(\omega)|\Omega_F)}{\mathcal{P}(I(\omega)|\Omega_B)} = \frac{\mathcal{P}(\Omega_F|I(\omega))\mathcal{P}(I(\omega))}{\mathcal{P}(\Omega_B|I(\omega))\mathcal{P}(I(\omega))} \times \frac{\mathcal{P}(\Omega_B)}{\mathcal{P}(\Omega_F)} \text{(Bayes rule)}$$

$$E_{img}(C) = -\int_{\omega\in\Omega} \mathbb{1}_{\Omega_F} \log \frac{\mathcal{P}(\Omega_F|I(\omega))}{\mathcal{P}(\Omega_B|I(\omega))} \frac{\mathcal{P}(\Omega_B)}{\mathcal{P}(\Omega_F)} d\omega + \text{constant}$$

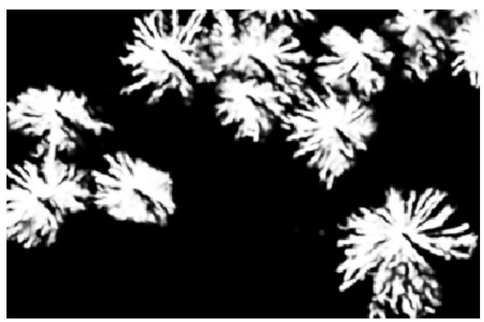
Approach - Formalizing energy model: image term

$$E_{img}(C) = -\int_{\omega\in\Omega} \mathbb{1}_{\Omega_F} \log \frac{\mathcal{P}(\Omega_F|I(\omega))}{\underbrace{\mathcal{P}(\Omega_B|I(\omega))}_{\text{two classes: } 1-\mathcal{P}(\Omega_F|I(\omega))}} d\omega + \text{constant}$$

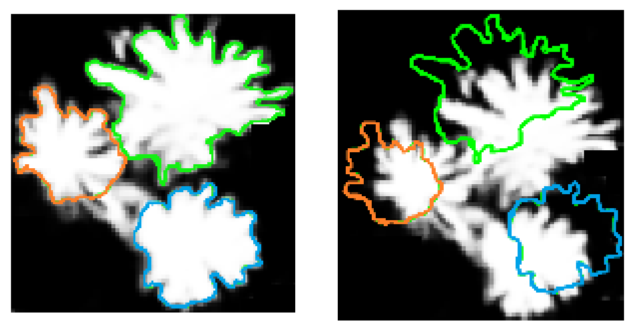
$$= -\int_{\omega\in\Omega} \mathbb{1}_{\Omega_F} \log \frac{\mathcal{P}(\Omega_F|I(\omega))}{1-\underbrace{\mathcal{P}(\Omega_F|I(\omega))}_{\text{arbitrary model of class probs}}} d\omega + \text{constant}$$



Sample CIR image



Posterior probability map



contour positions with:
low energy high energy

Approach – Formalizing energy model: shape term

Denote an approximation of foreground region indicator:

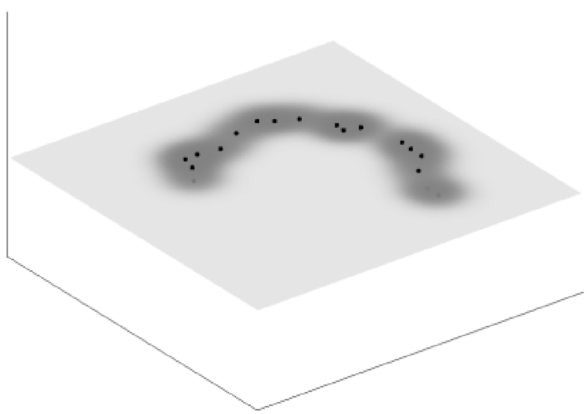
- $\mathbb{1}_{\Omega_F} \approx G(\alpha, x, y)$, where $G(\dots)$ is parameterized by a shape coefficient vector $\alpha \in \mathbb{R}^m$ and center position (x, y)

$$E(C) = E(\alpha, x, y) = \underbrace{-\log \mathcal{P}(I|\alpha, x, y)}_{\text{image term}} \underbrace{-\log \mathcal{P}(\alpha)}_{\text{shape term: } E_{shp}}$$

- $G(\dots)$ instantiates a shape given by coefficients α centered at (x, y) in the image
- Shape term E_{shp} is dependent *only* on the shape coefficients, *not* the position (x, y)
- Standard model: all positions are deemed equally likely – no prior knowledge available/used

- $\mathcal{P}(\alpha)$ can be modeled in a non-parametric fashion in terms of training shape coefficients
- Let $T = \alpha_1, \ldots, \alpha_N$ be a database of training shape coefficients and $K(x, y)$ denote a kernel function quantifying similarity between two vectors $x, y \in \mathbb{R}^m$
- The probability of a shape coefficient vector α is given by:

$$\mathcal{P}(\alpha) \propto \frac{1}{N} \sum_{i=1}^{N} K(\alpha, \alpha_i)$$



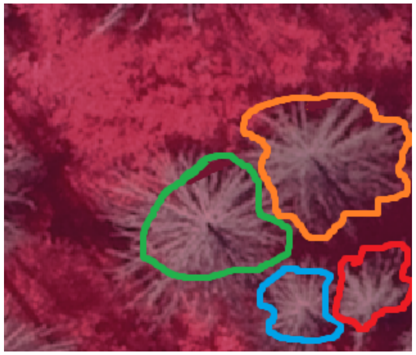
Source: Cremers and Rousson, 2007

$\rightarrow$ bulk of probability mass concentrated in regions of shape coefficient space with training vectors (black dots), regions far away allocated minimal probability

Approach – Multi-contour segmentation

But we **really** want to model **multiple** contours:

- $C_1, \ldots, C_K$: simultaneously evolving shape contours representing different object instances
- Contour C_i is parameterized by shape coefficient vector α_i and position (x_i, y_i)



$$\mathcal{P}(C_1, \ldots, C_K | I) \propto \mathcal{P}(I | C_1, \ldots, C_K) P(C_1, \ldots, C_K)$$

$$\Leftrightarrow E(C_1, \ldots, C_K) = \underbrace{-\log \mathcal{P}(I | C_1, \ldots, C_K)}_{\text{image term}} \underbrace{-\log P(C_1, \ldots, C_K)}_{\text{shape term}}$$

Approach – Multi-contour shape term

$$E(C_1,\dots,C_K) = \underbrace{-\log \mathcal{P}(I|\dots)}_{\text{image term}} \underbrace{-\log \mathcal{P}(C_1,\dots,C_K)}_{\text{shape term}}$$

- The contours are not independent: penalize overlap
- Include location information from Mask R-CNN
- Consider each contour's shape independent of position

$\Rightarrow$ Shape term as an undirected graphical model:

$$\mathcal{P}(C_1,\dots,C_K) = \frac{1}{Z}\prod_{\lambda\in\Lambda}\Psi_\lambda(C_\lambda)$$

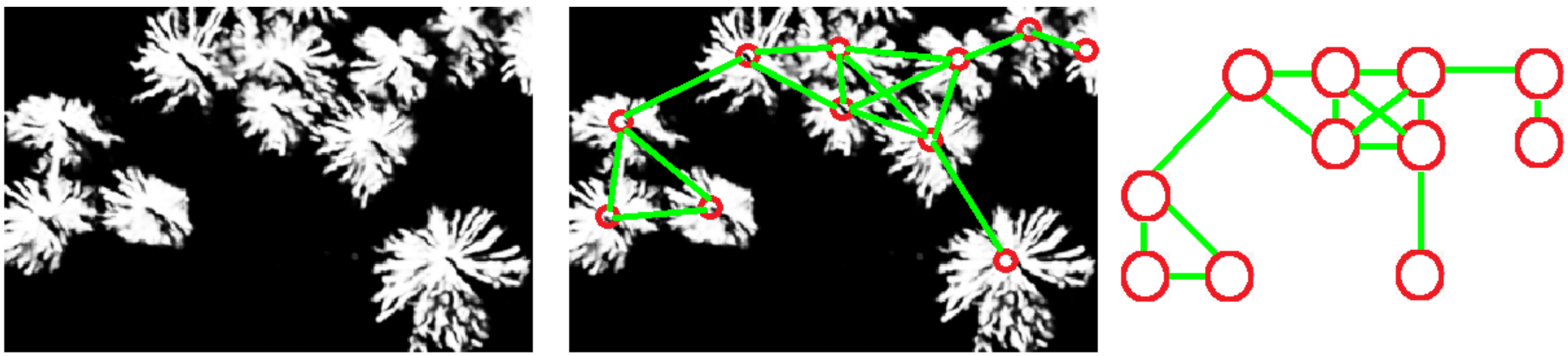
$\lambda \in \Lambda$: cliques of the graph $\mathcal{H}$ induced by object/tree crown centroids from Mask R-CNN and
Ψ_λ: are potential functions defined on subsets of contours

Approach – Shape term as a Graphical model

⇒ Remember:

$$\mathcal{P}(C_1, \ldots, C_K) = \frac{1}{Z} \prod_{\lambda \in \Lambda} \Psi_\lambda(C_\lambda)$$

$\Rightarrow \lambda \in \Lambda$ are cliques of the graph $\mathcal{H} = (\mathcal{V}, \mathcal{E})$ induced by Mask R-CNN found centroids



- Left: class probability map of dead tree scene
- Center: red circles = object centroids detected by Mask R-CNN and green lines connect centroids within d_{thr}
- Right: graph induced by the connections/edges and centroids/nodes

Approach – Graphical model potentials

$$\mathcal{P}(C_1, \ldots, C_K) = \tfrac{1}{Z} \prod_{\lambda \in \Lambda} \Psi_\lambda(C_\lambda)$$

$\Rightarrow$ Our model includes unary and pairwise potentials Ψ:

$$E_C(C_1, \ldots, C_K) = -\log \mathcal{P}(C_1, \ldots, C_K) + \text{constant} =$$

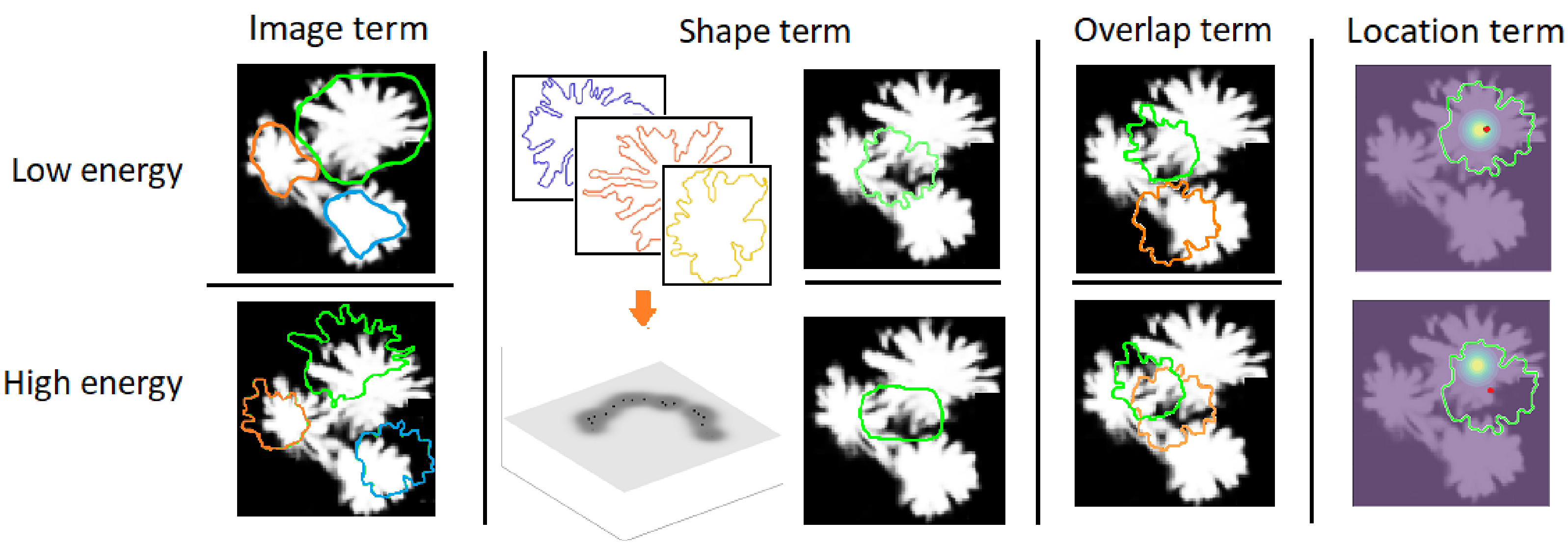
$$-\sum_{k=1}^{K} \log \Psi^{shp}(\alpha_k) - \sum_{k=1}^{K} \log \Psi_k^{loc}(x_k, y_k) - \sum_{(k,l) \in \mathcal{E}} \log \Psi_{ovp}(C_k, C_l)$$

- Shape potential: $\Psi^{shp}(\alpha_k) = \mathcal{P}(\alpha_k)$ - global KDE model
- Location potential: Gaussian centered on k-th centroid μ_k: $\Psi_k^{loc}(z) \propto \exp(-\frac{1}{2}(z - \mu_k)^T \Sigma^{-1} (z - \mu_k))$
- Overlap potential: overlap area between contours C_k, C_l: $\log \Psi_{ovp}(C_k, C_l) =$ $-\int_\omega G(\alpha_k, x_k, y_k)[\omega] \cdot G(\alpha_l, x_l, y_l)[\omega] d\omega$

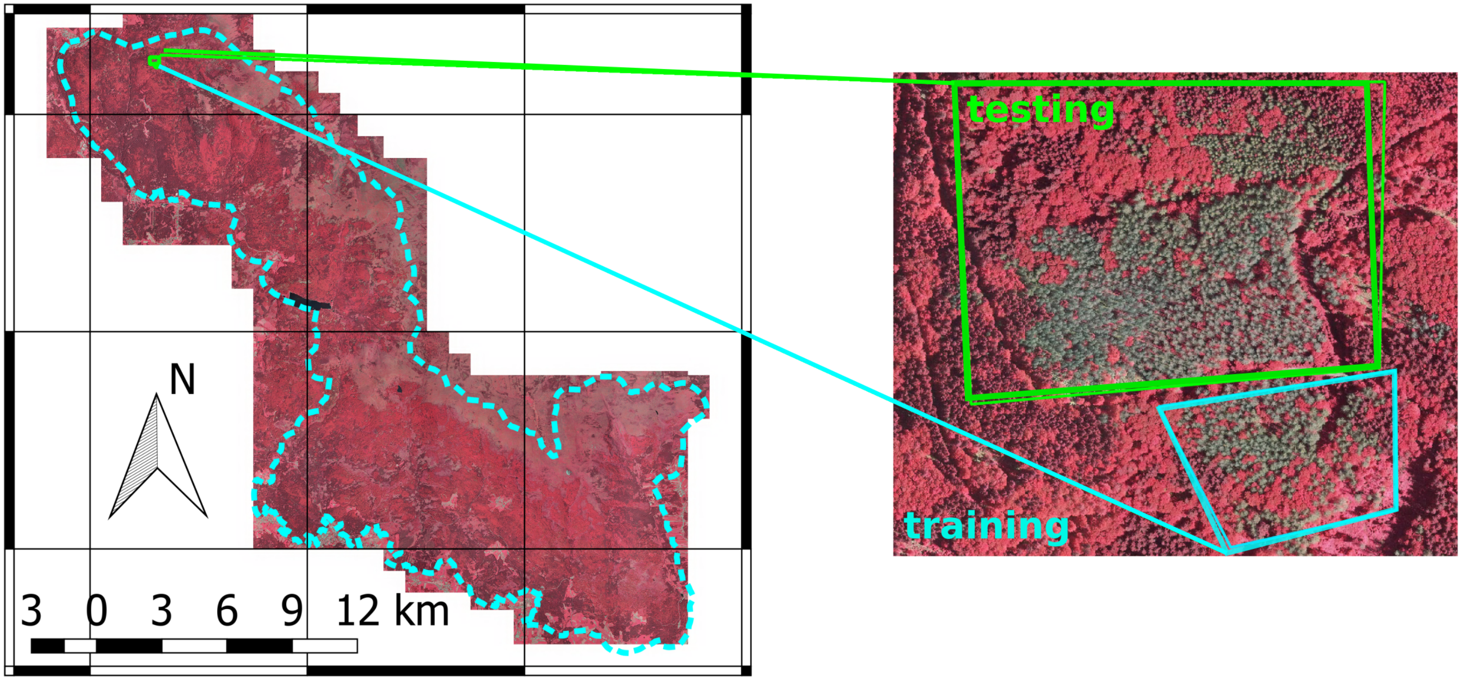
Approach – Put all the pieces together

$$E_{\text{total}}(C_1,\dots,C_K) = \underbrace{-\log \mathcal{P}(I|C_1,\dots,C_K)}_{\text{image term}} \underbrace{- \sum_{k=1}^{K} \log \Psi^{shp}(\alpha_k)}_{\text{shape term}}$$

$$\underbrace{- \sum_{(k,l)\in\mathcal{E}} \log \Psi^{ovp}(C_k, C_l)}_{\text{overlap term}} \underbrace{- \sum_{k=1}^{K} \log \Psi_k^{loc}(x_k, y_k)}_{\text{location term}}$$



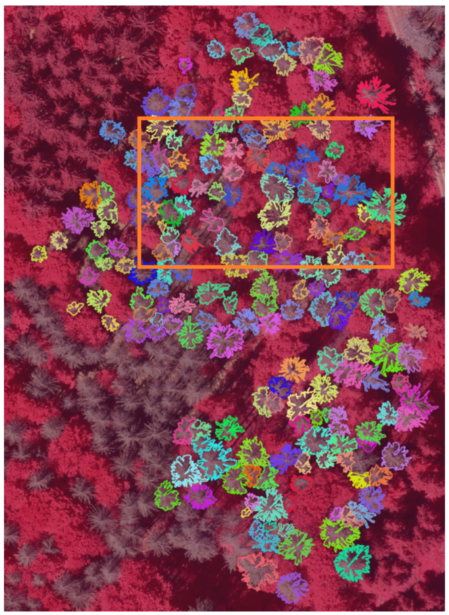
Experiments - Data: train and test areas

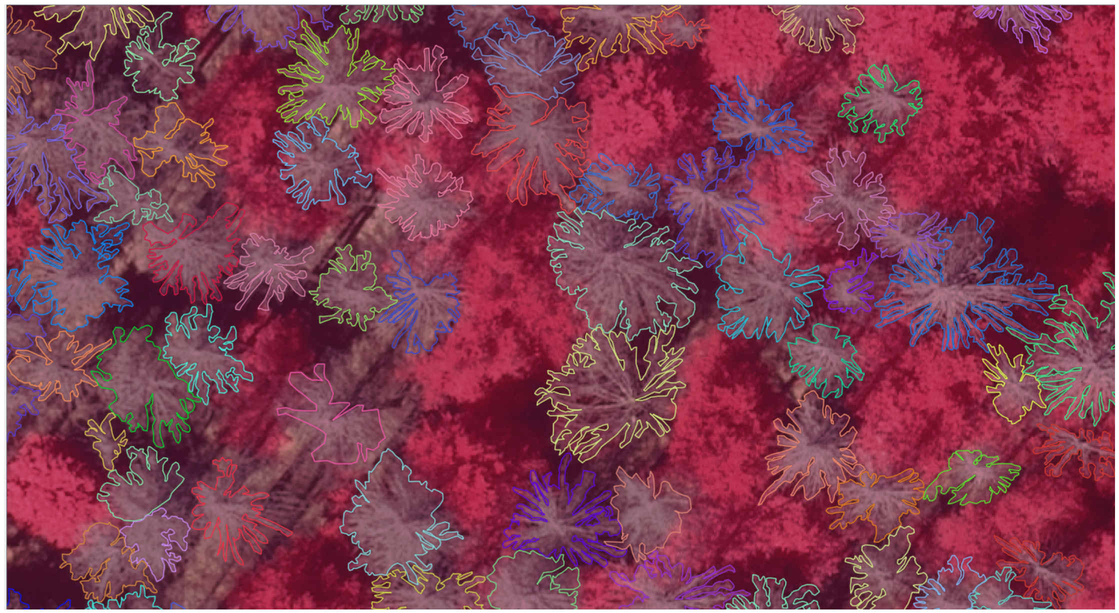


- We labeled tree crown polygons in forest areas for datasets:

$$\text{training} = 201 \text{ and } \text{testing} = 750$$

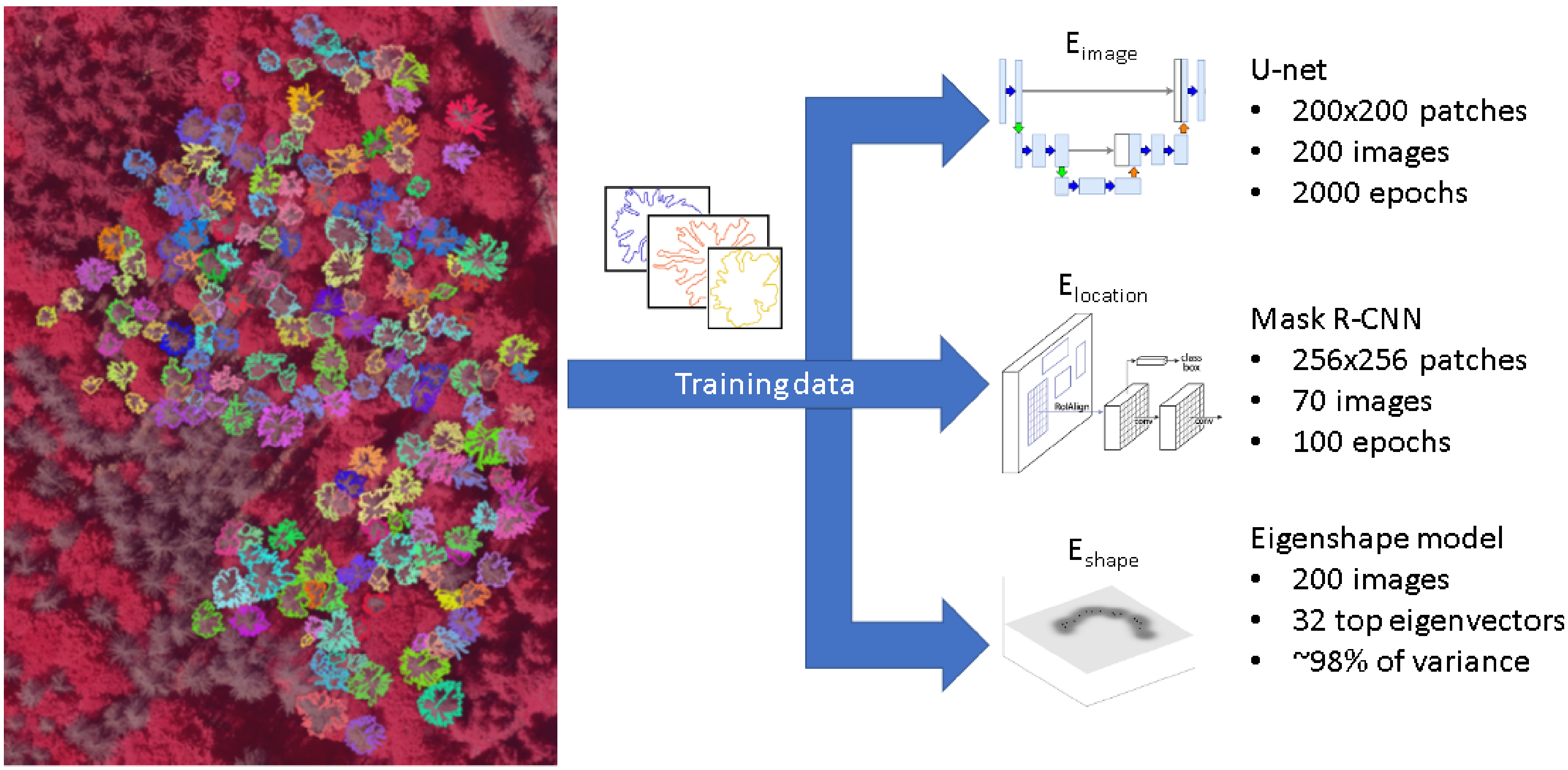
Experiments - Data: labels for training data





- Labeled 201 individual tree crown polygons for training area

Experiments - Training of the models



Experiments - Setting and evaluation metrics

Setting

- Data: $N = 750$ contours from test area in Bavarian National Forest
- Goal: experiment to compare polygons learned by **(1) Mask R-CNN** against contours refined by our active **(2) multi-contour model**
- Evaluation: both pixel level and object level of detected vs. reference polygons

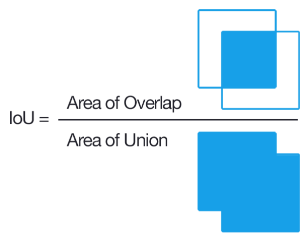
Experiments – Setting and evaluation metrics

Setting

- Data: $N = 750$ contours from test area in Bavarian National Forest
- Goal: experiment to compare polygons learned by **(1) Mask R-CNN** against contours refined by our active **(2) multi-contour model**
- Evaluation: both pixel level and object level of detected vs. reference polygons

Metrics

- Pixel level:

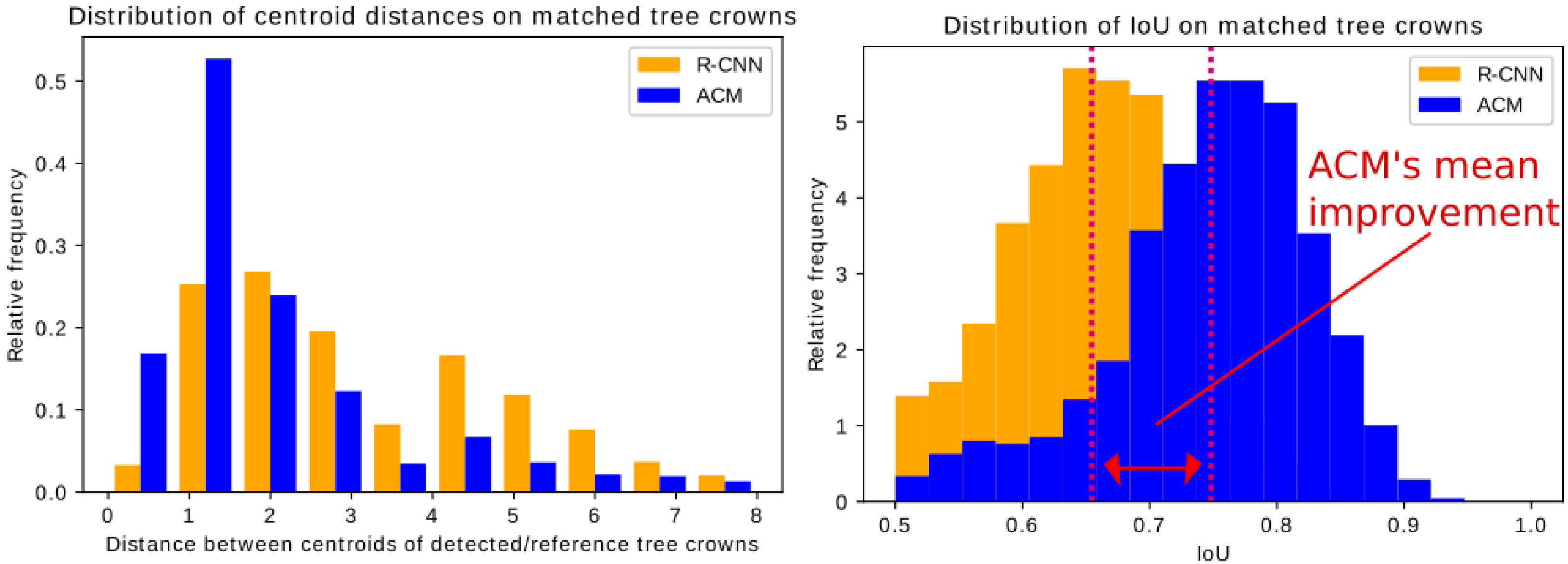


$\rightarrow$ intersection over union (IoU):
ratio intersection area to union area of compared shapes:
$IoU(A, B) = \frac{|A \cap B|}{|A \cup B|}$

- Object level:
  - mean distance between centroids of reference & detected polygons at $IoU \geq 0.5$
  - precision and recall at $IoU \geq 0.5$:

$$\text{precision} = \frac{\#\text{matched polygons}}{\#\text{detected polygons}} \qquad \text{recall} = \frac{\#\text{matched polygons}}{\#\text{reference polygons}}$$

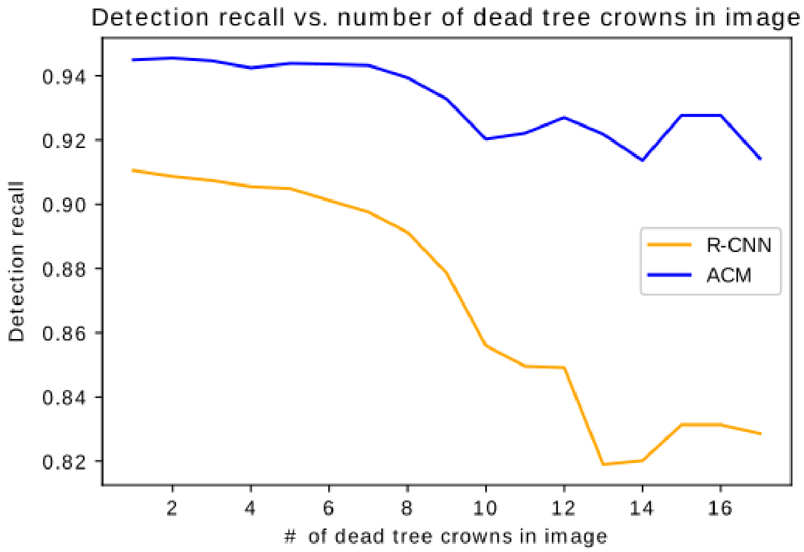
Experiments - Results: centroid distance and IoU

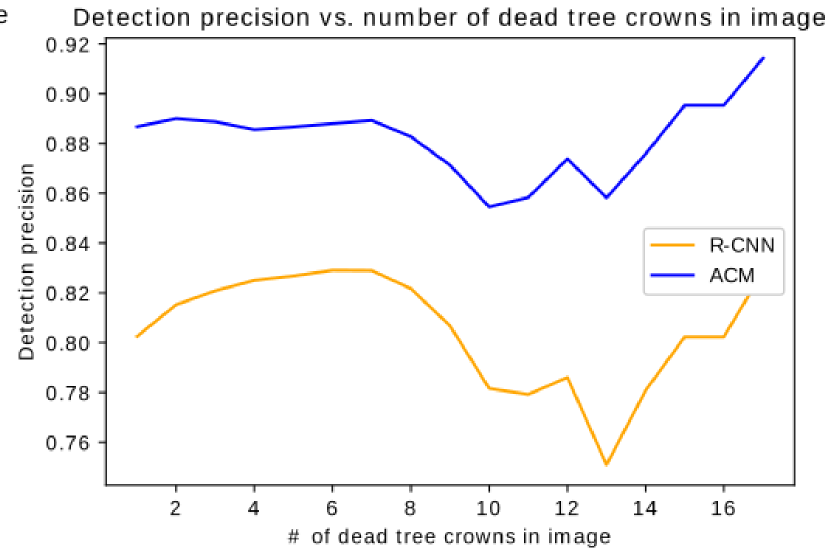


Shown: our active contour model (ACM) outperforms Mask R-CNN both in terms of identifying tree crown centers and discerning overlapping tree crowns

→ increased mean reference and detected centroid distance from 3.4 to 2.4 pixels (left)
→ increased mean matched IoU from 0.66 to 0.75 (right)

Experiments - Results: object-level precision and recall



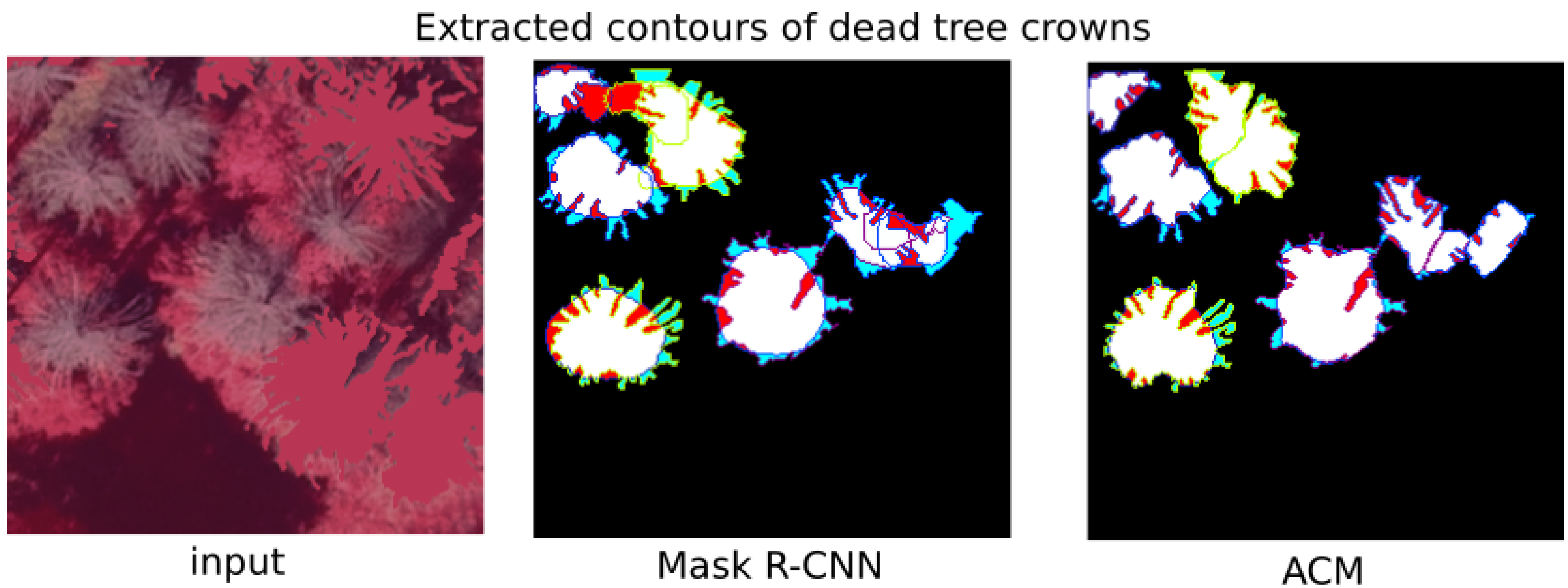


Shown: our active contour model (ACM) outperforms Mask R-CNN when shown *any* # of tree crowns

→ increased recall by 3.5 perentage points (left)
→ inreased precision by 8 percentage points (right)

Note: in images with **many** adjacent tree crowns → active contour model (ACM) can handle complex overlapping objects **far better** than Mask R-CNN

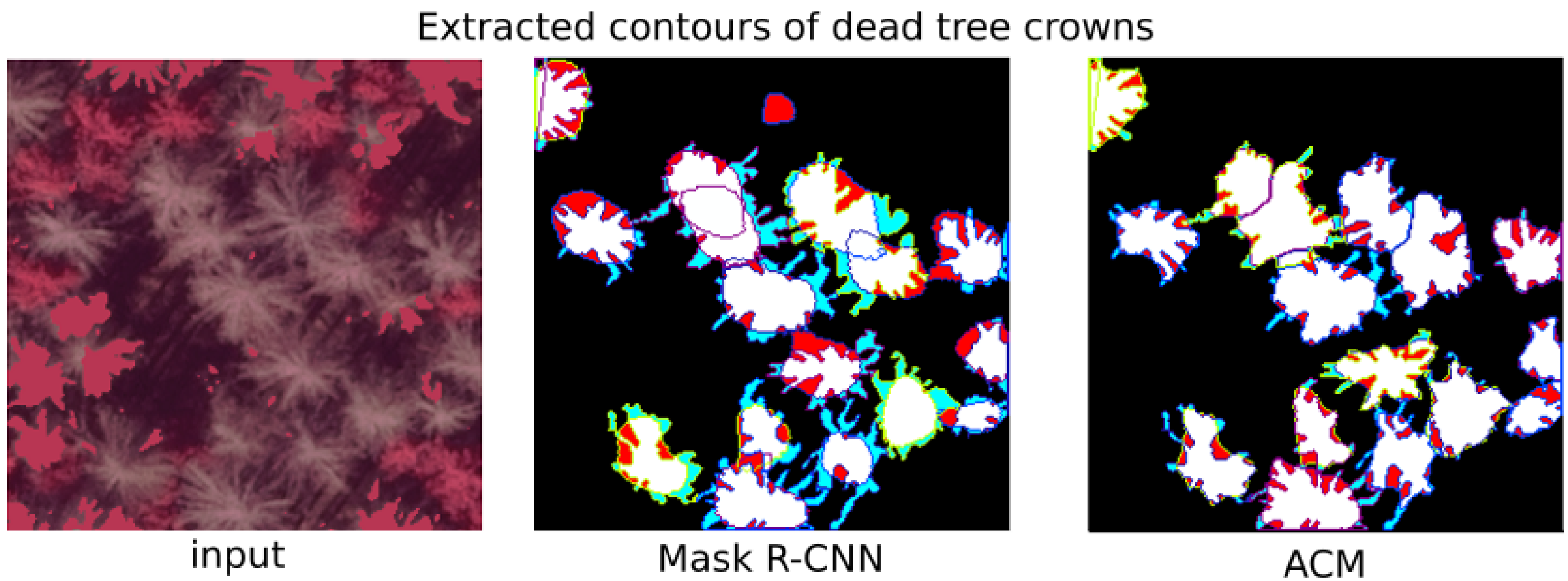
Experiments - Results: extracted contours



Shown: tree contours extracted from a CIR input image by our active contour model (ACM) are more refined and visually match the true contours than Mask R-CNN

→ Correctly identified tree crown pixels are shown in white
→ False positives are shown in red
→ False negatives are shown in cyan

Experiments - Results: extracted contours



Shown: tree contours extracted from a CIR input image by our active contour model (ACM) are more refined and visually match the true contours than Mask R-CNN

→ Correctly identified tree crown pixels are shown in white
→ False positives are shown in red
→ False negatives are shown in cyan

To **summer**-ize...

- Dead wood sequesters 8% of global forest carbon, houses $\frac{1}{3}$ of forests biodiversity, and is a **key indicator of overall forest health** ... but good wood models necessary/lacking

- This work:
  - Combines deep learning and instance segmentation
  - Leverages prior knowledge of crown shape and appearance
  - Constructs comprehensive energy functional for **multiple, overlapping contours**

$\rightarrow$ **Our approach discovers improved, localized and refined contours of dead trees**

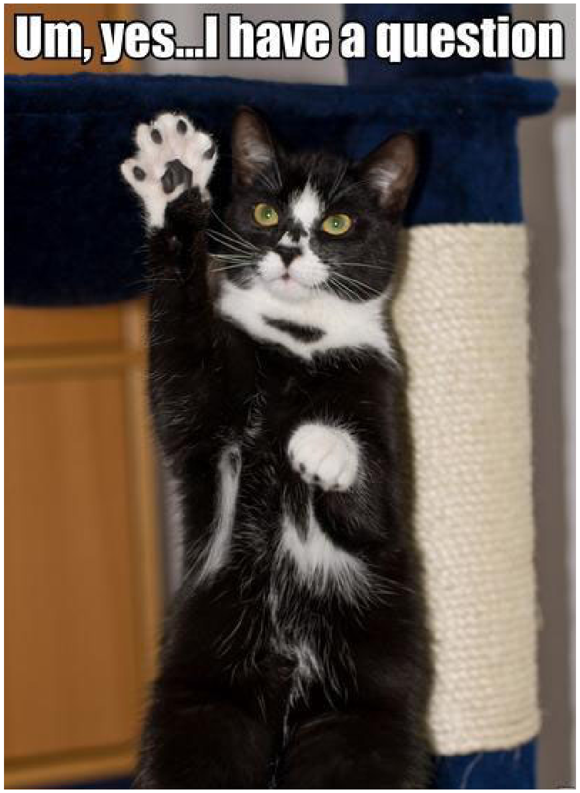
Outlook

Future/current work:

- **Goals (always):** efficient, robust, scalable ML methods, critical to:
  $\rightarrow$ Exploit modern remote sensing data – cheaper, higher quality, larger quantity, freely available
  $\rightarrow$ Better understand biodiversity and the role of dead wood, and impacts of climate
- **Change detection:** high quality contour delineation improves estimates of dead wood
  $\rightarrow$ Model of decay dynamics, tree type and tree counts
- **Advanced priors:** replace vanilla eigenshape prior with GAN generated shapes
  $\Rightarrow$ **preliminary results show contour improvements**
- **Sparsity and approximations:** GANs to develop contours using only pixels on the outline $\rightarrow$ save computational resources

Thanks for your attention!

Questions, feedback, suggestions, cats, ideas, musings?
Very eager for all of the above!



Obligitory morbid satire...

Where is Wiley going!?

"We have built a civilization based on a world that doesn't exist anymore."

– Katharine Hayhoe, climate scientist at Texas Tech University and chief scientist at the Nature Conservancy

Appendix - References

1. A. R. Martin, G. M. Domke, M. Doraisami, and S. C. Thomas: Carbon fractions in the world's dead wood, Nature Communications, vol. 12, 2021.

2. Y. Pan, R. A. Birdsey, J. Fang, ... and D. Hayes: A large and persistent carbon sink in the world's forests, Science, vol. 333, no. 6045, 2011.

3. S. Seibold, W. Rammer, T. Hothorn, ... and J. Mueller: The contribution of insects to global forest deadwood decomposition, Nature, vol. 597, Sep 2021.

4. M. A. Bradford, R. J. Warren II, P. Baldrian, T. W. Crowther, D. S. Maynard, E. E. Oldfield, W. R. Wieder, S. A. Wood, and J. R. King: Climate fails to predict wood decomposition at regional scales, Nature Climate Change, 2014.

5. P. Polewski, J. Shelton, W. Yao, and M. Heurich: Segmentation of single standing dead trees in high- resolution aerial imagery with generative adversarial network-based shape priors, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2020.

6. D. Cremers, M. Rousson, and R. Deriche: A review of statistical approaches to level set segmentation: Integrating color, texture, motion and shape, Int. J. Comput. Vision, vol. 72, no. 2, 2007.

7. D. Cremers and M. Rousson: Efficient kernel density estimation of shape and intensity priors for level set segmentation, Deformable Models, Topics in Biomedical Engineering. International Book Series, Springer, 2007.

8. O. Ronneberger, P. Fischer, and T. Brox: U-Net: Convolutional Networks for Biomedical Image Segmentation, CoRR, vol. abs/1505.04597, 2015.

9. Leventon, M., Grimson, W. and Faugeras, O.: Statistical shape influence in geodesic active contours, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Vol. 1, 2000.

10. Tsai, A., Yezzi, A., III, W. W., Tempany, C., Tucker, D., Fan, A., Grimson, W. and Willsky, A.: A shape-based approach to the segmentation of medical imagery using level sets, IEEE T. Med. Imaging 22(2), 2003.