Weakly Supervised Learning

A. Maier, K. Breininger, L. Mill, N. Ravikumar, T. Würfl
Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg
July 04, 2018

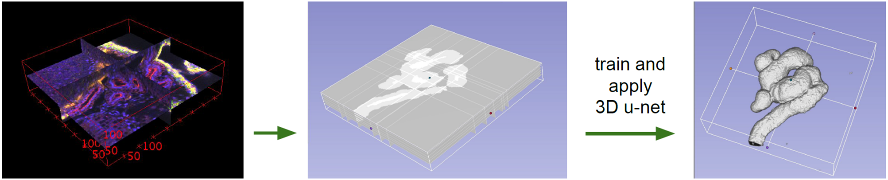
Outline

From 2D Annotation to 3D Annotation

From Labels to Localization

From 2D Annotation to 3D Annotation

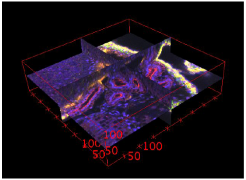
From 2D Annotation to 3D Annotation

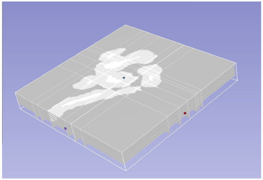


- Only few labeled 2D slices

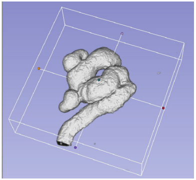
Source: [1]

From 2D Annotation to 3D Annotation





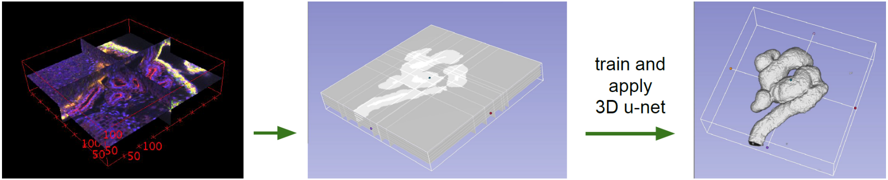
train and apply 3D u-net



- Only few labeled 2D slices
- Generalize to the 3D case

Source: [1]

From 2D Annotation to 3D Annotation



- Only few labeled 2D slices
- Generalize to the 3D case
- Interactive segmentation

Source: [1]

Training with Sparse Labels

"One hot" labels *y* with element *i* being 1

$$L = -\log\left(\frac{e^{s_i}}{\sum_{k=1}^{K} e^{s_k}}\right)$$

Source: [1]

Training with Sparse Labels

"One hot" labels y with element i being 1

$$L = -\log\left(\frac{\mathrm{e}^{s_i}}{\sum_{k=1}^{K}\mathrm{e}^{s_k}}\right)$$

Sparse

➔ Add a label y_{k+1}

$$L' = L \cdot \mathrm{w}(\mathbf{y})$$

Where $\mathrm{w}(\mathbf{y}) = \begin{cases} 0 & \text{if } y_{k+1} = 1 \\ w_i > 0 & \text{otherwise} \end{cases}$

Source: [1]

From Bounding Boxes to Segmentation

Expensively annotated

Fully supervised

- Manual segmentation is tedious

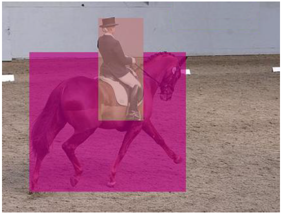
Source: [2]

From Bounding Boxes to Segmentation

Expensively annotated

Fully supervised

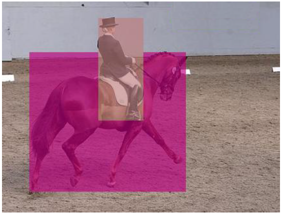
Cheaply annotated



- Manual segmentation is tedious
- Bounding boxes are less tedious

Source: [2]

From Bounding Boxes to Segmentation

Expensively annotated Fully supervised	Cheaply annotated	Cheaply annotated Weakly supervised
		

- Manual segmentation is tedious
- Bounding boxes are less tedious

→ Learn segmentation from boxes

Source: [2]

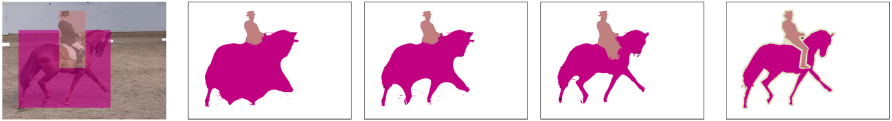
Naive Recursive Training

Observation: Bounding boxes as target lead to basic segmentations.

Source: [2]

Naive Recursive Training

Observation: Bounding boxes as target lead to basic segmentations.



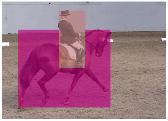
Example input rectangles	Output after 1 training round	After 5 rounds	After 10 rounds	Ground truth

→ Smaller box is foreground

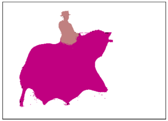
Source: [2]

Naive Recursive Training

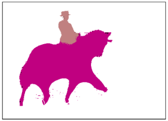
Observation: Bounding boxes as target lead to basic segmentations.



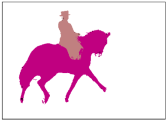
Example
input rectangles



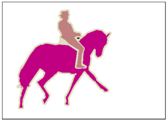
Output after
1 training round



After
5 rounds



After
10 rounds



Ground
truth

→ Smaller box is foreground

- **Suppress** detections

  ...of wrong class

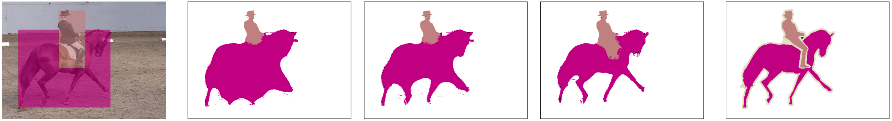
  ...outside of box

  ...<% box area

Source: [2]

Naive Recursive Training

Observation: Bounding boxes as target lead to basic segmentations.



Example input rectangles	Output after 1 training round	After 5 rounds	After 10 rounds	Ground truth

→ Smaller box is foreground

- **Suppress** detections

  ...of wrong class

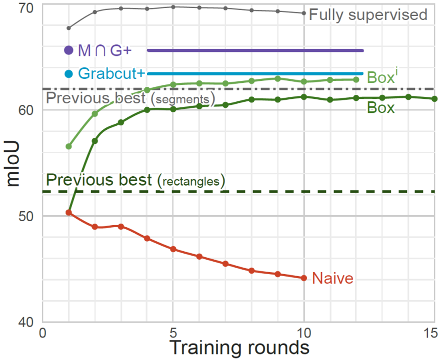
  ...outside of box

  ...<% box area

- Apply CRF on output

Source: [2]

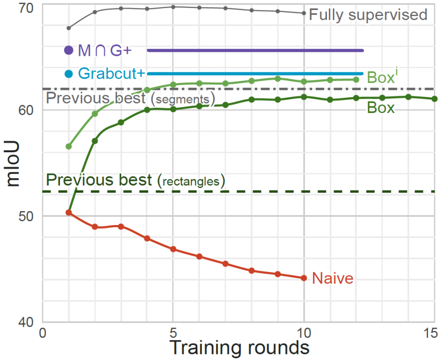
Improved Recursive Training



➔ Shrink box - beats state of the art

Source: [2]

Improved Recursive Training

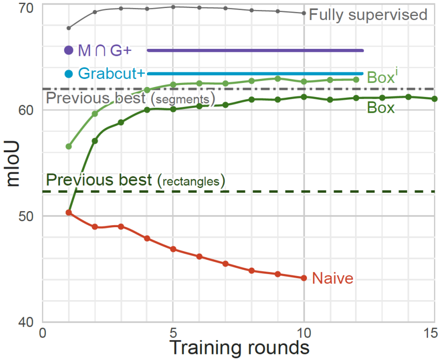


→ Shrink box - beats state of the art

- Use Grabcut

Source: [2]

Improved Recursive Training



→ Shrink box - beats state of the art

- Use Grabcut
- Combine Grabcut and MCG

Source: [2]

From Labels to Localization

From Labels to Localization

Four ingredients:

From Labels to Localization

Four ingredients:

- **Pretrained** filters

From Labels to Localization

Four ingredients:

- **Pretrained** filters
- Whole **arbitrary size** image as input

From Labels to Localization

Four ingredients:

- **Pretrained** filters
- Whole **arbitrary size** image as input
- **Localize** class

From Labels to Localization

Four ingredients:

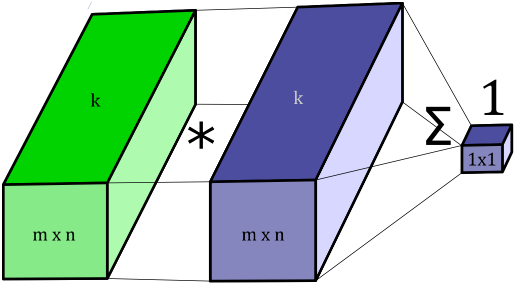
- **Pretrained** filters
- Whole **arbitrary size** image as input
- **Localize** class
- Classify **individually**

Arbitrary input size

➔ **Fully connected** layer fixes the size

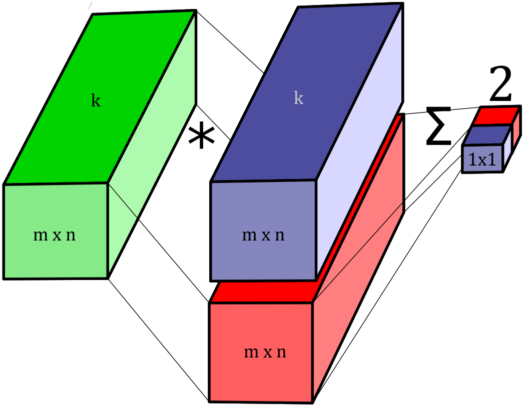
Arbitrary input size

- ➔ **Fully connected** layer fixes the size
- ➔ Replace by **convolutional** layer



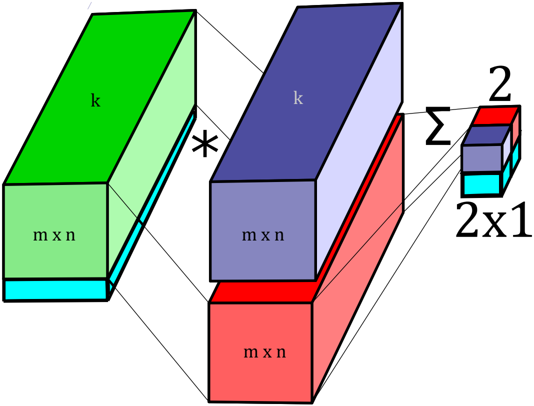
Arbitrary input size

- → **Fully connected** layer fixes the size
- → Replace by **convolutional** layer



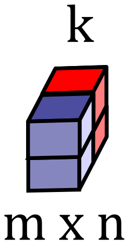
Arbitrary input size

- ➔ **Fully connected** layer fixes the size
- ➔ Replace by **convolutional** layer



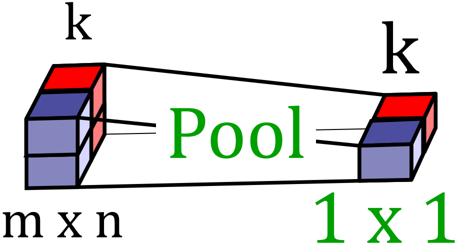
Arbitrary input size

- How to localize?



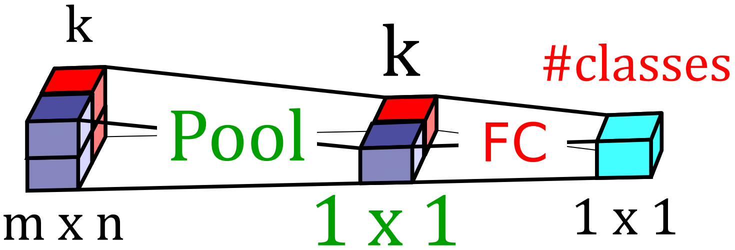
Arbitrary input size

- How to localize?
- Global pooling
- Oquab et al., 2015: Max pooling; Bolei et al., 2016: Average pooling



Arbitrary input size

- How to localize?
- Global pooling
- Oquab et al., 2015: Max pooling; Bolei et al., 2016: Average pooling



Individual Classification

➔ **Multiple objects** could be visible

$$L_i = -\log\left(\frac{\mathrm{e}^{s_i}}{\sum_{k=1}^{K} \mathrm{e}^{s_k}}\right)$$

Individual Classification

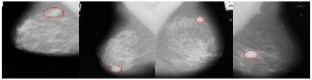
→ **Multiple objects** could be visible

$$L_i = -\log\left(\frac{e^{s_i}}{\sum_{k=1}^{K} e^{s_k}}\right)$$

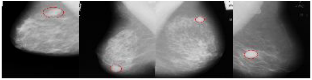
Change "one-hot" labels y_k from 0 to -1 indicating "absence"

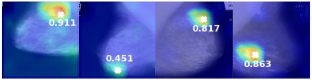
$$\begin{aligned} L_i &= \sum_{k}^{K} -\log\left(\frac{1}{1+e^{-y_k s_k(\mathbf{x})}}\right) \\ &= \sum_{k}^{K} \log\left(1+e^{-y_k s_k(\mathbf{x})}\right) \end{aligned}$$

What if we can't get an appropriate pretrained net?

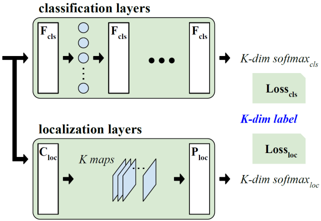


What if we can’t get an appropriate pretrained net?





→ Multi-task learning



$$\text{Loss}_{total} = (1 - \alpha)\,\text{Loss}_{cls} + \alpha\,\text{Loss}_{loc}$$

Also significantly **boosts** the **localization**!

Source: Hwang et al., Self-Transfer Learning for Fully Weakly Supervised Object Localization, 2016

FRIEDRICH-ALEXANDER-
UNIVERSITÄT
ERLANGEN-NÜRNBERG
SCHOOL OF ENGINEERING

References

References I

[1] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, et al. "3d u-net: learning dense volumetric segmentation from sparse annotation". In: International Conference on Medical Image Computing and Computer-Assisted Springer. 2016, pp. 424–432.

[2] Anna Khoreva, Rodrigo Benenson, Jan Hosang, et al. "Simple Does It: Weakly Supervised Instance and Semantic Segmentation". In: arXiv preprint arXiv:1603.07485 (2016).

[3] Maxime Oquab, Léon Bottou, Ivan Laptev, et al. "Is object localization for free?-weakly-supervised learning with convolutional neural networks". In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recogniti 2015, pp. 685–694.

References II

[4] Bolei Zhou, Aditya Khosla, Agata Lapedriza, et al. "Learning deep features for discriminative localization". In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recogniti 2016, pp. 2921–2929.

[5] Sangheum Hwang and Hyo-Eun Kim. "Self-Transfer Learning for Weakly Supervised Lesion Localization". In: International Conference on Medical Image Computing and Computer-Assisted Springer. 2016, pp. 239–246.

[6] Richard O. Duda, Peter E. Hart, and David G. Stork. Pattern classification. 2nd ed. New York: Wiley-Interscience, Nov. 2000.