

# Vertebrae Localization and Segmentation with Spatial Configuration-Net and U-Net

Christian Payer, Darko Štern, Horst Bischof, **Martin Urschler**



▶ 13th October 2019

[martin.urschler@auckland.ac.nz](mailto:martin.urschler@auckland.ac.nz)

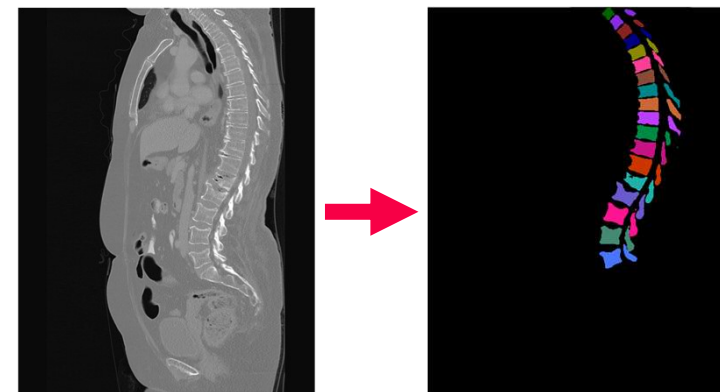
## Vertebrae Localization and Segmentation

Repetitive structures: similar **local appearance**...

... but strong **spatial configuration**

Idea:

1. Localize whole spine
  - To remove background
2. Localize all vertebrae simultaneously
  - With SpatialConfiguration-Net [1]
3. Segment each vertebrae individually
  - With U-Net [2]

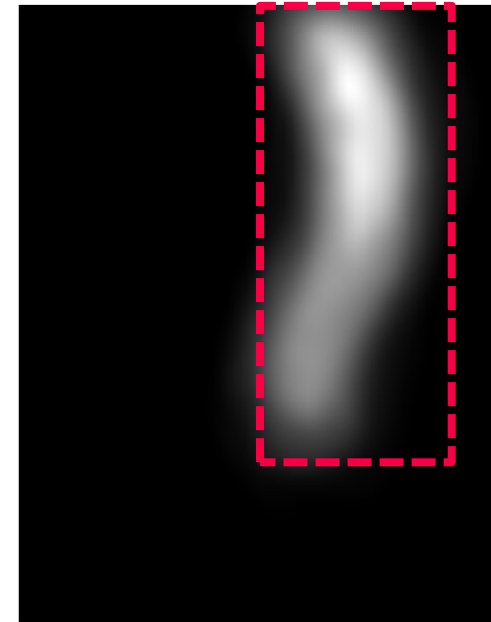


[1] Payer et al. (2019) Integrating Spatial Configuration into Heatmap Regression Based CNNs for Landmark Localization. *MIA*.

[2] Ronneberger et al. (2015) UNet: Convolutional Networks for Biomedical Image Segmentation. *MICCAI*.

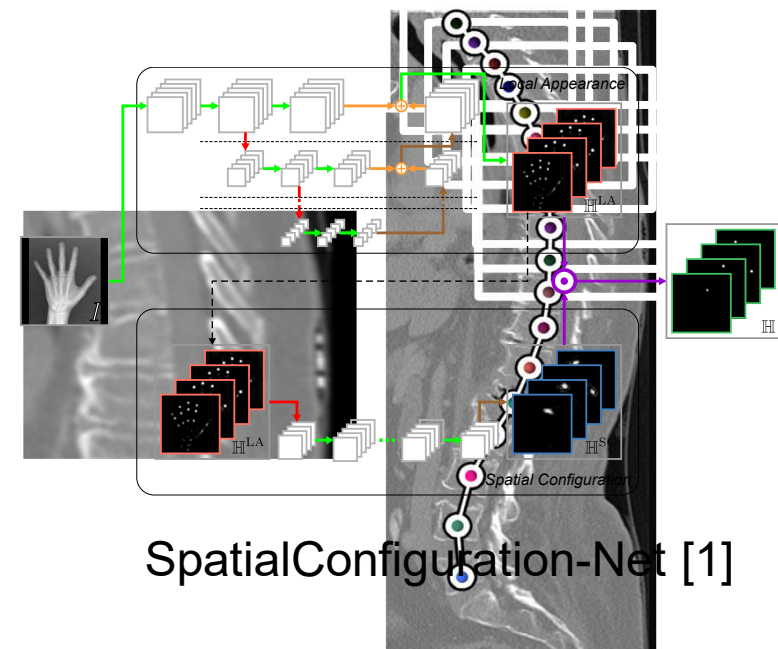
## Spine Localization

- Spine location varies a lot in input volumes
  - Localize and center spine
- Predict heatmap of the spine
  - Groundtruth calculated from individual vertebra locations
- 3D U-Net at low voxel resolution
  - $8 \times 8 \times 8$  mm,  $64 \times 64 \times 128$  voxels



# Vertebrae Localization

- Neighboring vertebrae have similar **local appearance**
  - Lots of repetitive structures
- However, strong **spatial configuration**
  - Order and distance of vertebrae
- 3D SpatialConfiguration-Net [1]
  - Combines **local appearance** with **spatial configuration** of landmarks

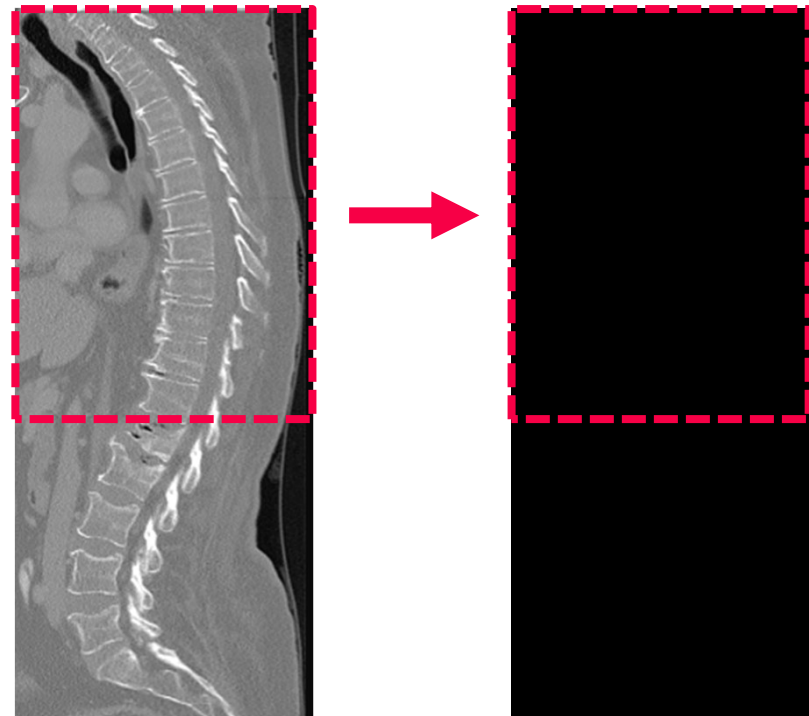


SpatialConfiguration-Net [1]

[1] Payer et al. (2019) Integrating Spatial Configuration into Heatmap Regression Based CNNs for Landmark Localization. *MIA*.

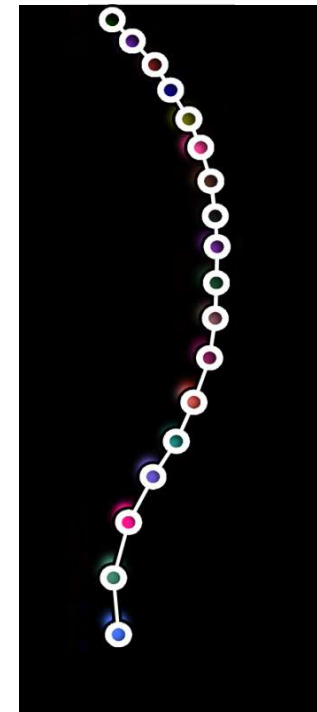
## Vertebrae Localization

- 3D SpatialConfiguration-Net
  - $2 \times 2 \times 2$  mm,  $96 \times 96 \times 128$  voxels
- Predict heatmaps of vertebrae
  - Centered at location of vertebra
- Tiled processing of cropped input volumes



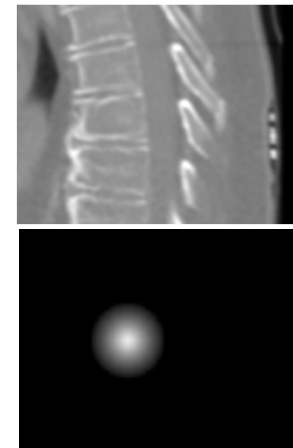
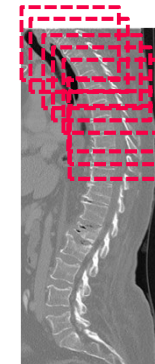
## Vertebrae Localization

- Predicted landmark locations are on heatmap maxima
  - Not all vertebrae are always visible → Postprocessing
- Postprocessing:
  - Valid predictions must have a minimal heatmap value
  - Find valid sequences of consecutive vertebrae
    - Vertebrae order needs to be correct
    - Vertebrae need to be neither too close nor too far apart
    - Final sequence is the sequence with the largest value of summed up heatmap values



# Vertebrae Segmentation

- Segment each vertebra individually
  - Input cropped around the groundtruth (in training) or predicted (inference) location
  - Additional input as focus mechanism: heatmap centered at the location
  
- Predict binary vertebra mask with 3D U-Net
  - $1 \times 1 \times 1$  mm,  $128 \times 128 \times 96$  voxels
  - Single network for all vertebrae
  - Semantic label through localization ID



## Results

Spine Localization n/a (no meaningful metric)

		Distance (in mm)	ID-rate
Vertebrae Localization	CV	5.77 ± 7.18	91.2% (774 of 849)
	Test1	4.55	94.2%

		Dice	Hausdorff (in mm)
Vertebrae Segmentation	CV	0.94 ± 0.11	7.72 ± 13.95
	Test1	0.89	25.5

CV ... Three-fold cross validation: 53 training images, 27 validation images

Test1 ... Public test set: 80 training images, evaluated by organizers on 40 test images



## Conclusion

Three stage pipeline with increasing resolution to...

- ... localize spine
- ... localize individual vertebrae
- ... segment individual vertebrae

