TUM Data Innovation Lab

Project: "Deep Learning on High-Res Multispectral Aerial Imagery"

Authors:
Felix Buchert, Sarah Dörr,
Filippo Galassi, Max Helleis, Kamilia Mullakaeva

Mentors:
M.Sc. David Dohmen, M.Sc. Felix Horvat (Ocell GmbH)
M.Sc. Oleh Melnyk (Department of Mathematics)

Project Lead:
Dr. Ricardo Acevedo Cabra (Department of Mathematics)

Supervisor:
Prof. Dr. Massimo Fornasier (Department of Mathematics)

Agenda

1 Introduction

2 Setup

3 Approach 1

4 Approach 2

5 Approach 3

6 Further Results

7 Conclusion

Agenda

1 **Introduction**

2 **Setup**

3 **Approach 1**

4 **Approach 2**

5 **Approach 3**

6 **Further Results**

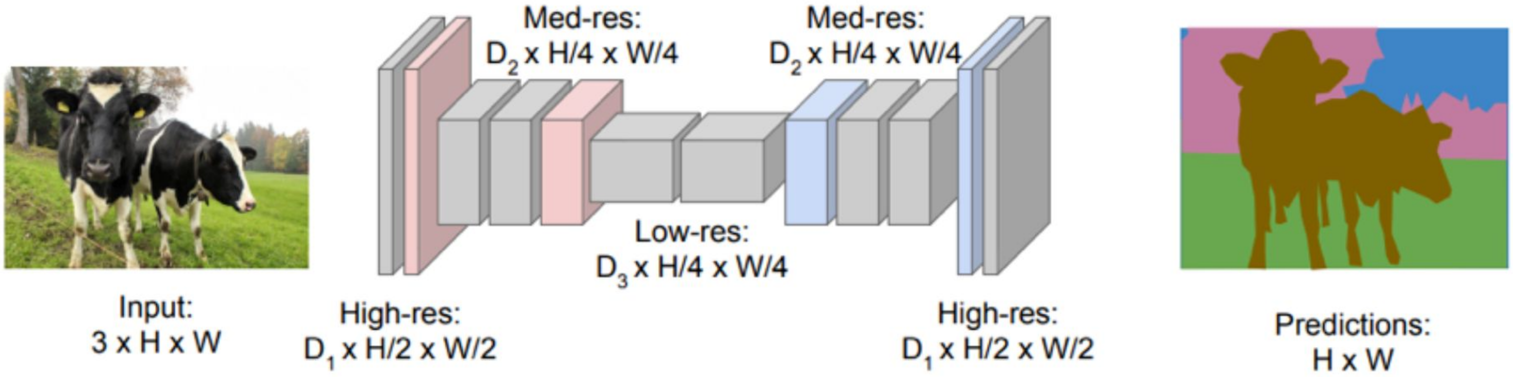
7 **Conclusion**

1.1 Introduction – Motivation

- **Why tree classification?**
  - Determination of the overall forest stock volume
  - Identification of tree species
  - Distribution of tree species
  - Assessment of tree / forest health
- **Why aerial imagery?**
  - Costwise and timewise benefit
  - Very high resolution data
- **Why multispectral imagery?**
  - Most widely used
  - High reflectance of vegetation in near-infrared domain

1.2 Introduction – Problem statement

- **Main goal**
  - Improving OCELL's approach for tree detection and species classification
- **Current approach:**
  - Semantic Segmentation: Generate output segmentation masks using a Fully Convolutional Neural Network (FCNN) from input images
  - Tree localization and classification: Extract center points from output segmentation maps



[10]

1.2 Introduction – Problem statement

- **Potential points of improvement**
  - **Approach 1:**
    Evaluation and comparison of other suitable architectures
  - **Approach 2:**
    Performance analysis under different definitions of ground truth segmentation mask
  - **Approach 3:**
    Integration of height information and Near-Infrared band

Agenda

1 Introduction

2 **Setup**

3 Approach 1

4 Approach 2

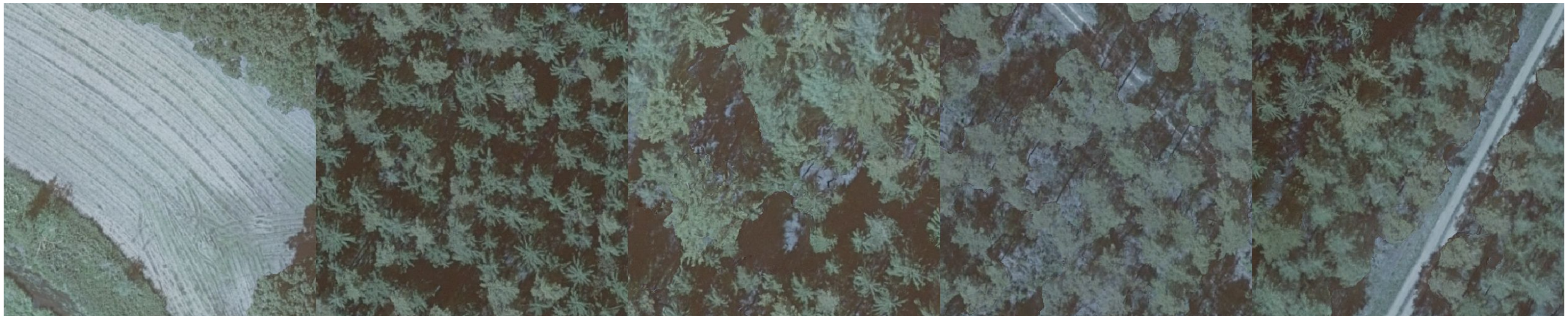
5 Approach 3

6 Further Results

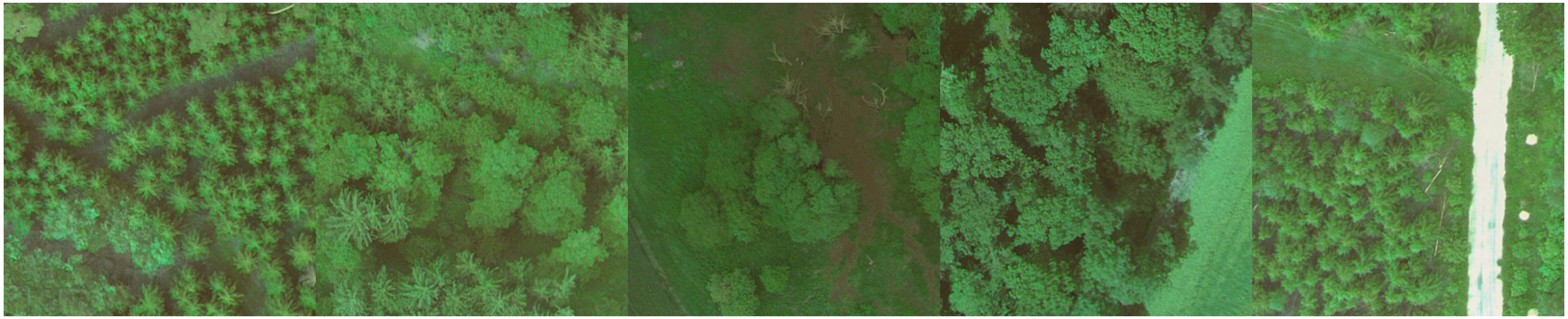
7 Conclusion

2.1 Setup – Data Sets (1)

Data Set A

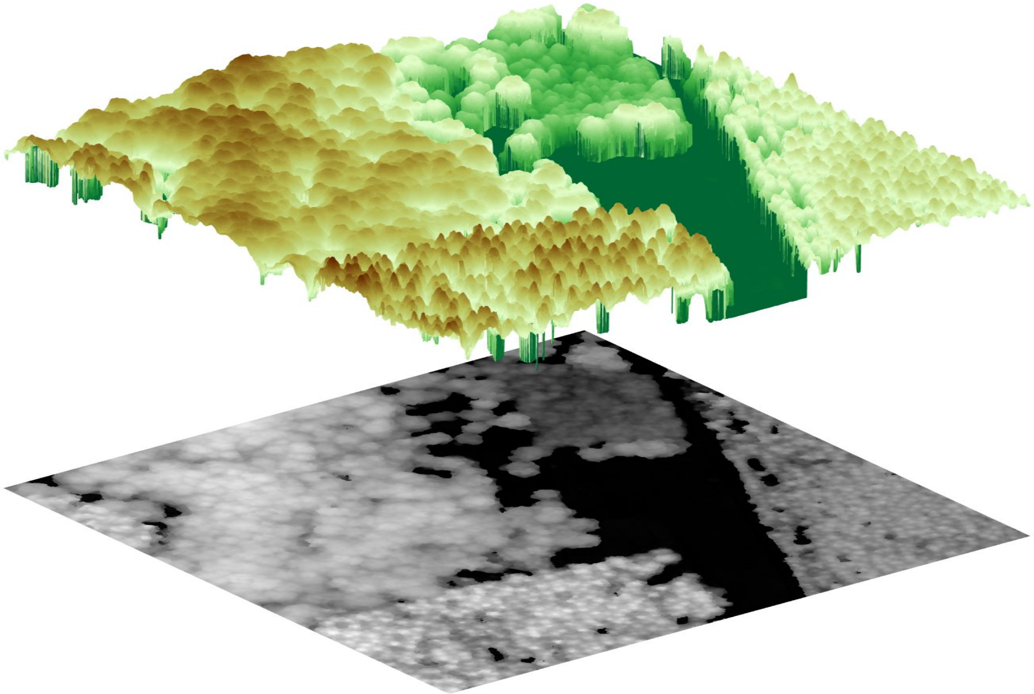


Data Set B



2.1 Setup – Data Sets (2)

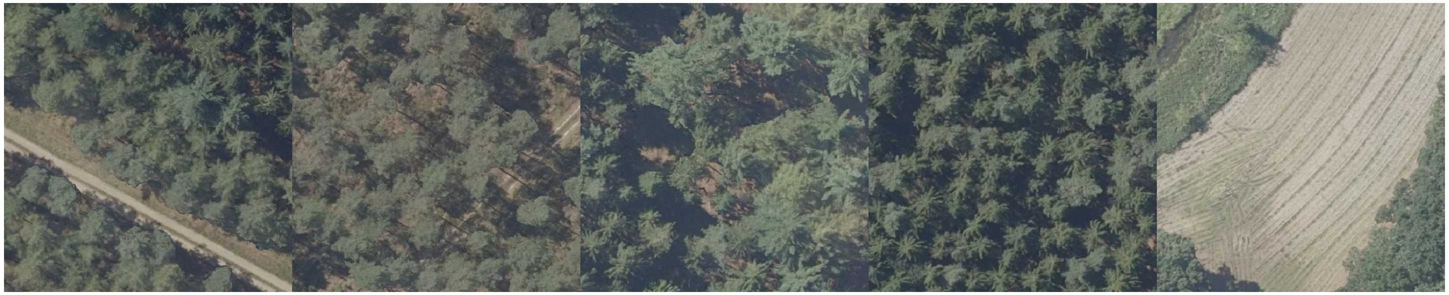
- Acquired with a sensor developed by the company
- Orthorectified images were provided
- Implementation of DSM model
- Implementation of NIR band in data set A



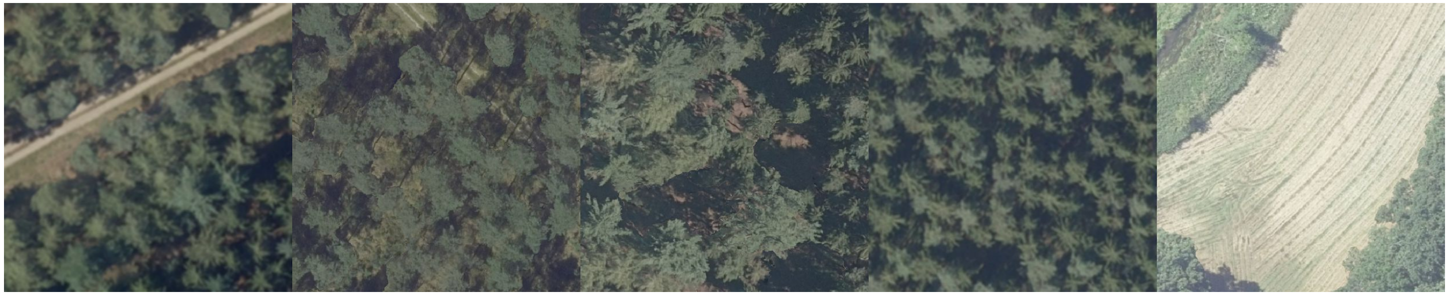
2.2 Setup – Data Preprocessing (1)

- Image tiling
  - Generation of equally sized tiles
  - Tile size: 512 x 512 pixel
- Data augmentation
  - Weak and strong augmentation
  - Augmentation optimized for multispectral images
    - Split → Augment → Recombine → Augment
- Data split
  - Training: 70%
  - Validation: 20%
  - Testing: 10%

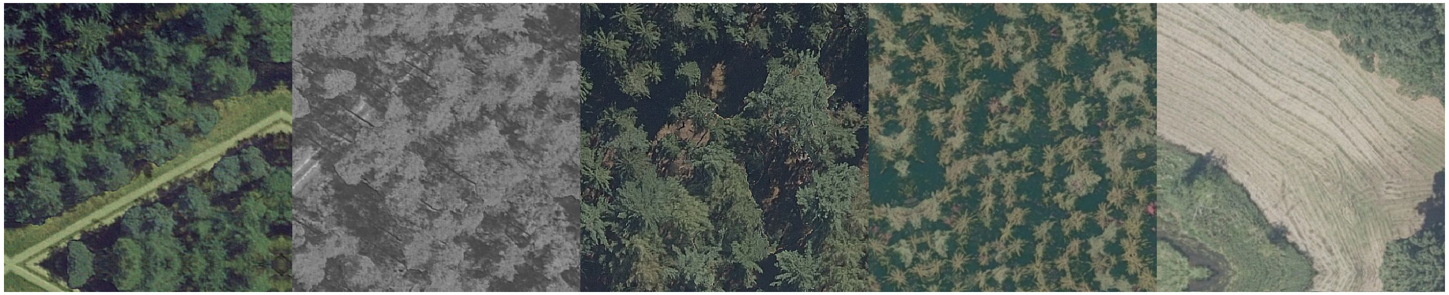
2.2 Setup – Data Preprocessing (2)



without augmentation



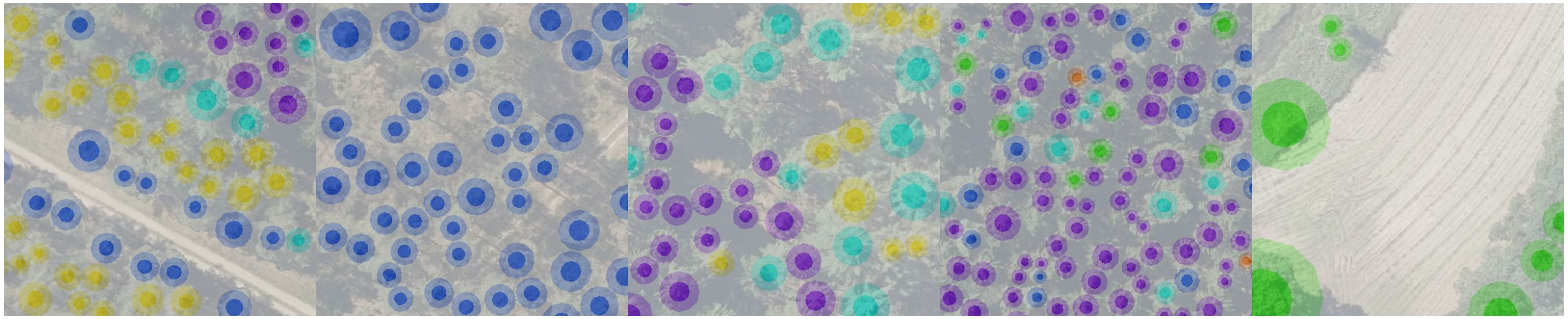
weak augmentation



strong augmentation

2.3 Setup – Training Pipeline

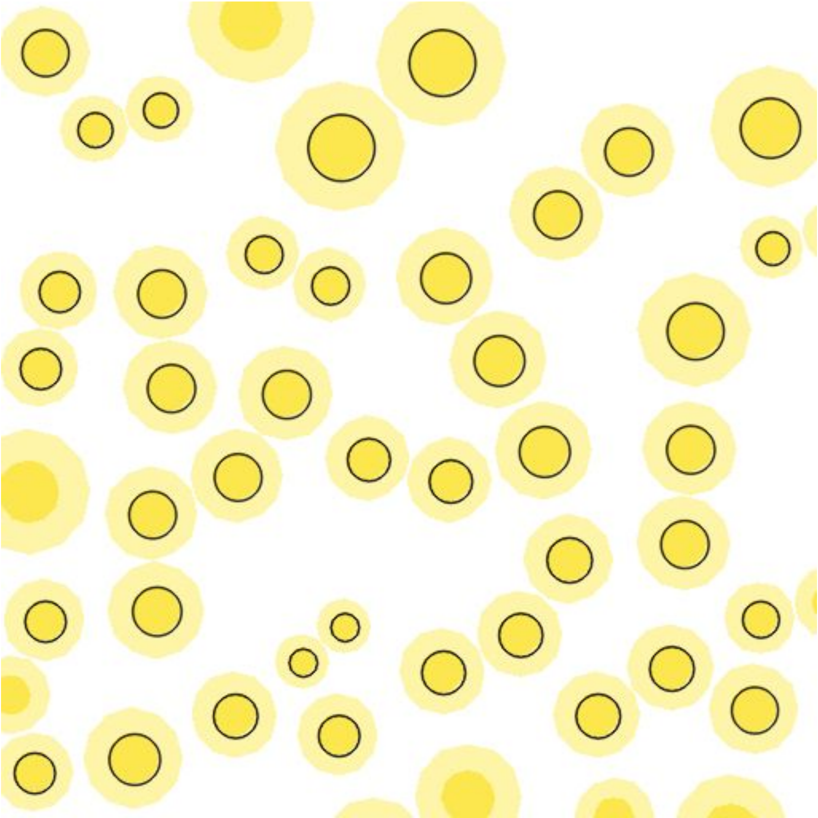
- Each training process has this setup
  - **Choice of best model**
  - **Optimizer:** Adam [1]
  - **Loss function:** Lovász-Softmax loss [2]
- Runs as a sequence of different setups (Architectures, label definitions)



Provided label definition

2.4 Setup – Evaluation Pipeline (1)

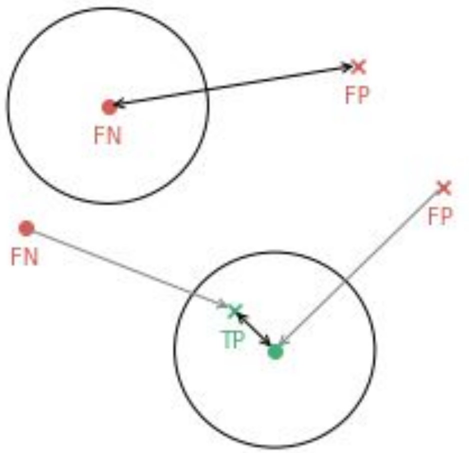
- **Metric Choice:** Pixel-wise metrics are not informative in context of tree detection
- **Point Extraction:** Tree centers and species have to be extracted from output segmentation mask
- **Blob detection:** Extract keypoints (i.e. tree centers) by detecting areas of uniform color
- **Implementation:** OpenCV blob detection algorithm used (based on Border-Following algorithm [3])



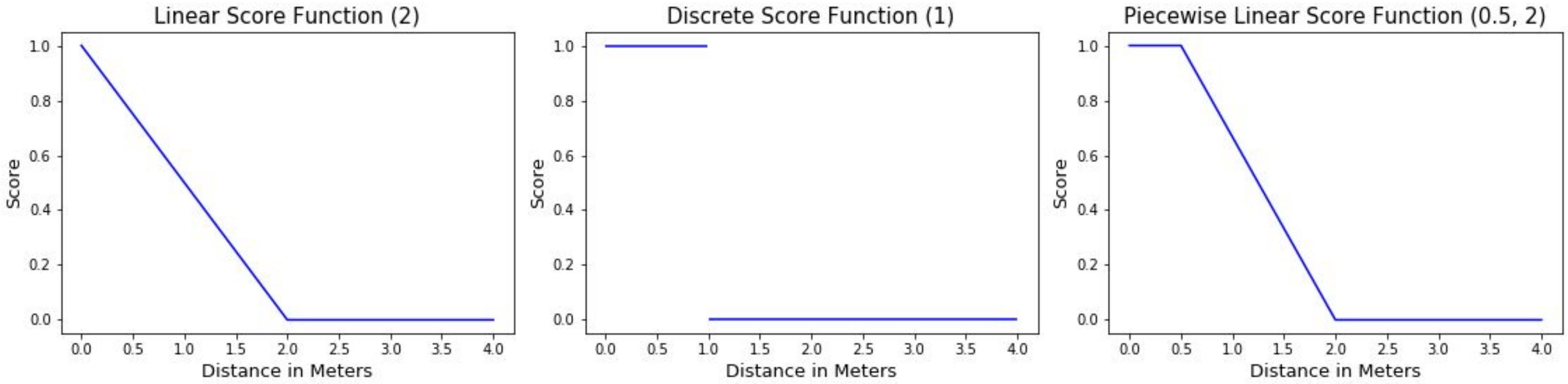
Blob Detection: Detected center points from ground truth segmentation masks

2.4 Setup – Evaluation Pipeline (2)

- **Nearest-Neighbor matching**:
  - Find nearest neighbors for all predictions and labels
  - Only Match if pairwise nearest neighbor
- **Score Definition**
  - **Center Scores:** Measures distance of centers
  - **Sample-Weighted Class Score:** Average score of all class scores w.r.t. correct center predictions (weighted by the number of samples)



Nearest-Neighbor Matching



Agenda

1 Introduction

2 Setup

3 **Approach 1**

4 Approach 2

5 Approach 3

6 Further Results

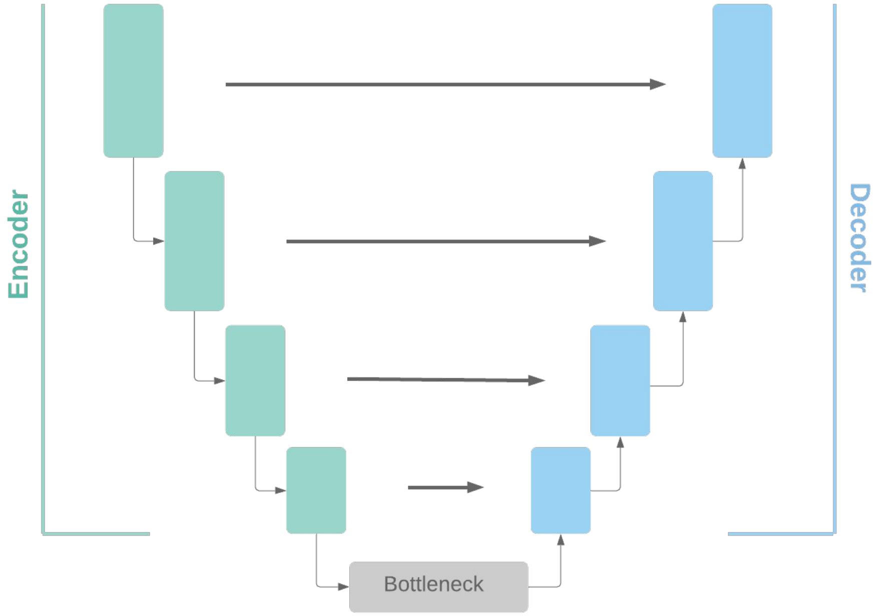
7 Conclusion

3.1 Approach 1 – Architectures

- **Current state:**
  - AlbuNet architecture [4] with pre-trained ResNet-50 as encoder
- **Issues:**
  - No evaluation and comparison to other suitable neural network architectures
    ⇒ Hard to measure how well the current architecture performs
  - Large architecture with a lot of parameters to train
    ⇒ Long training, inference time, requires more GPU memory
- **Goal:**
  - Conduct a comparative analysis of the performance of AlbuNet
  - Evaluate and compare a selection of related architectures

3.2 Approach 1 – General structure

- **Downsampling path:** Capturing the context of the image and extracting feature maps
- **Up-sampling path:** Transforming features back to an output map (same size as the input image)
- **Skip connections:** Reusing feature maps of downsampling path
  ⇒ Helps to recover spatially detailed information



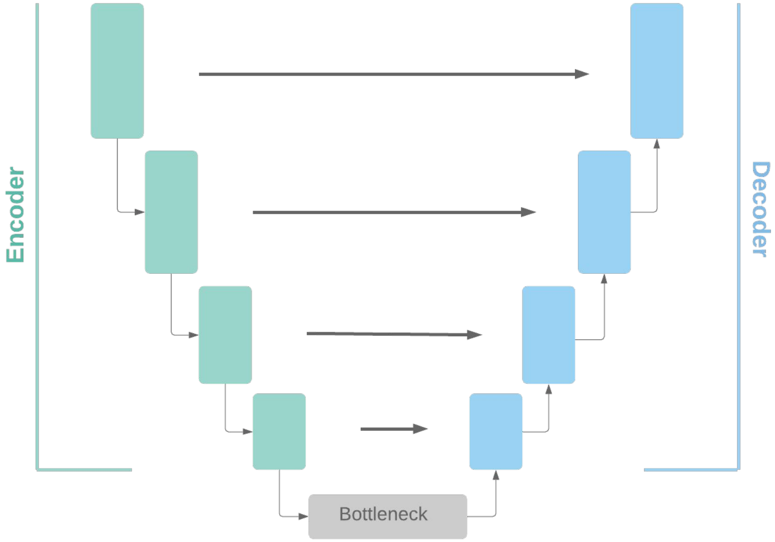
3.2 Approach 1 – U-Net and TernausNet

U-Net [5] (not pre-trained):

- Encoder block: Convolution, ReLU, MaxPool layers
- Decoder block: Convolution, ReLU, Interpolation layers
- Bottleneck: Convolution, Interpolation layers

TernausNet [6]:

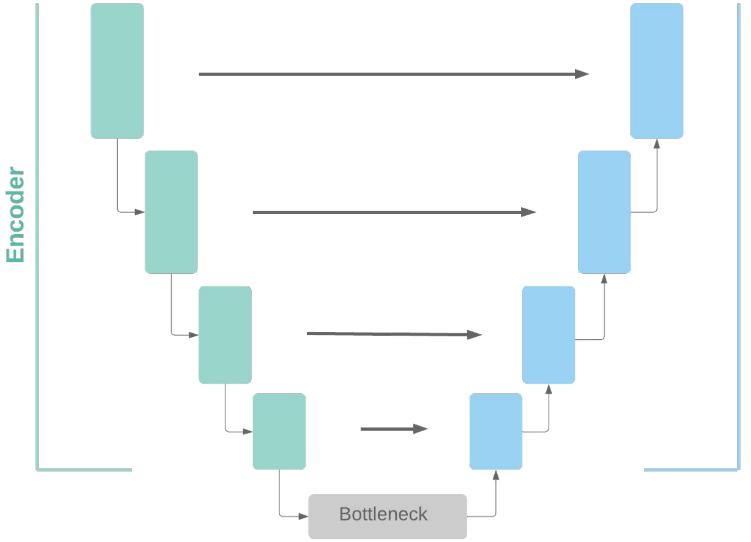
- Encoder: VGG-11, VGG-16
- Pre-trained encoder on ImageNet [9]



3.2 Approach 1 – AlbuNet and Tiramisu

AlbuNet [4]:

- Encoder: ResNet-50, ResNet-34, ResNet-18
- ResNet uses Residual Blocks (skip connection in each block)
- Pre-trained encoder on ImageNet [9]

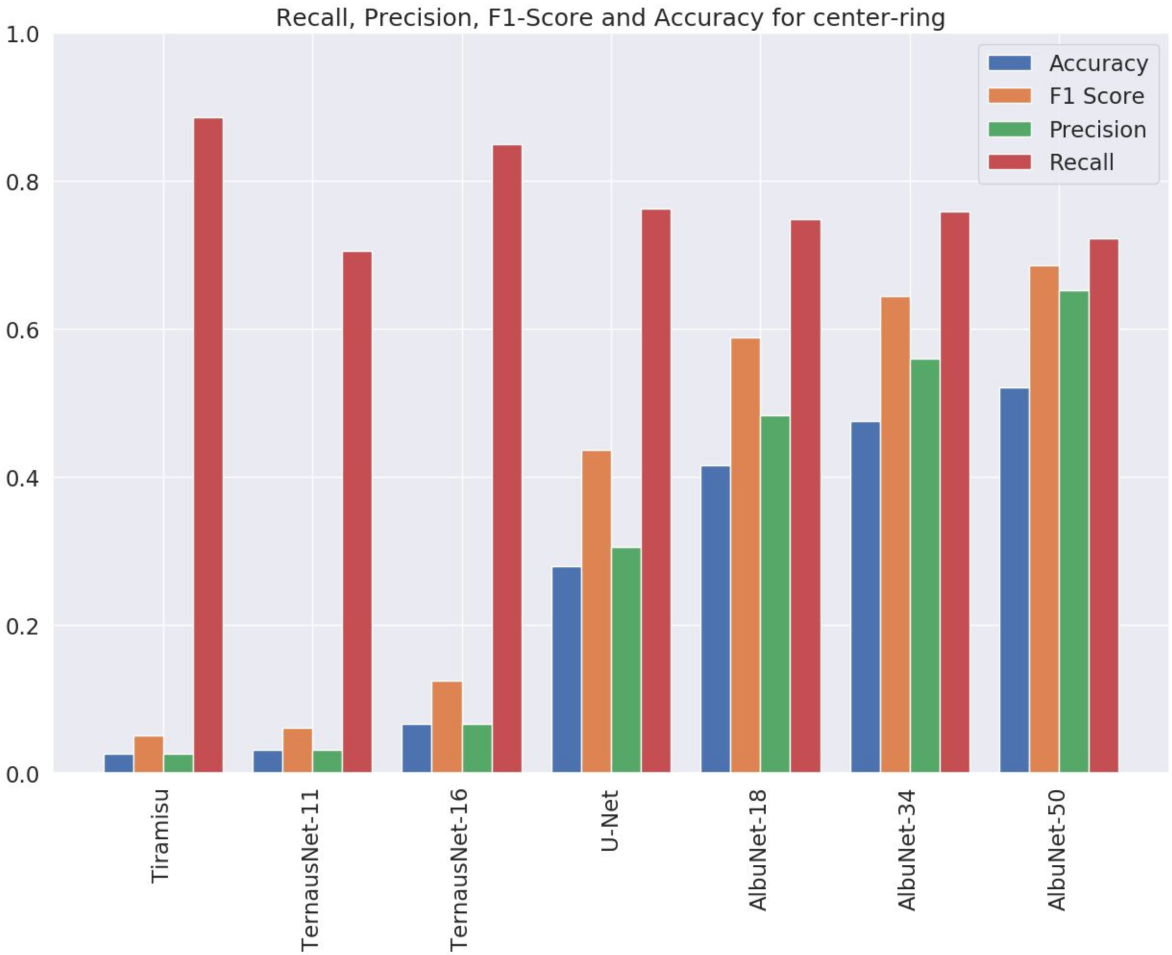


Tiramisu [7] (not pre-trained):

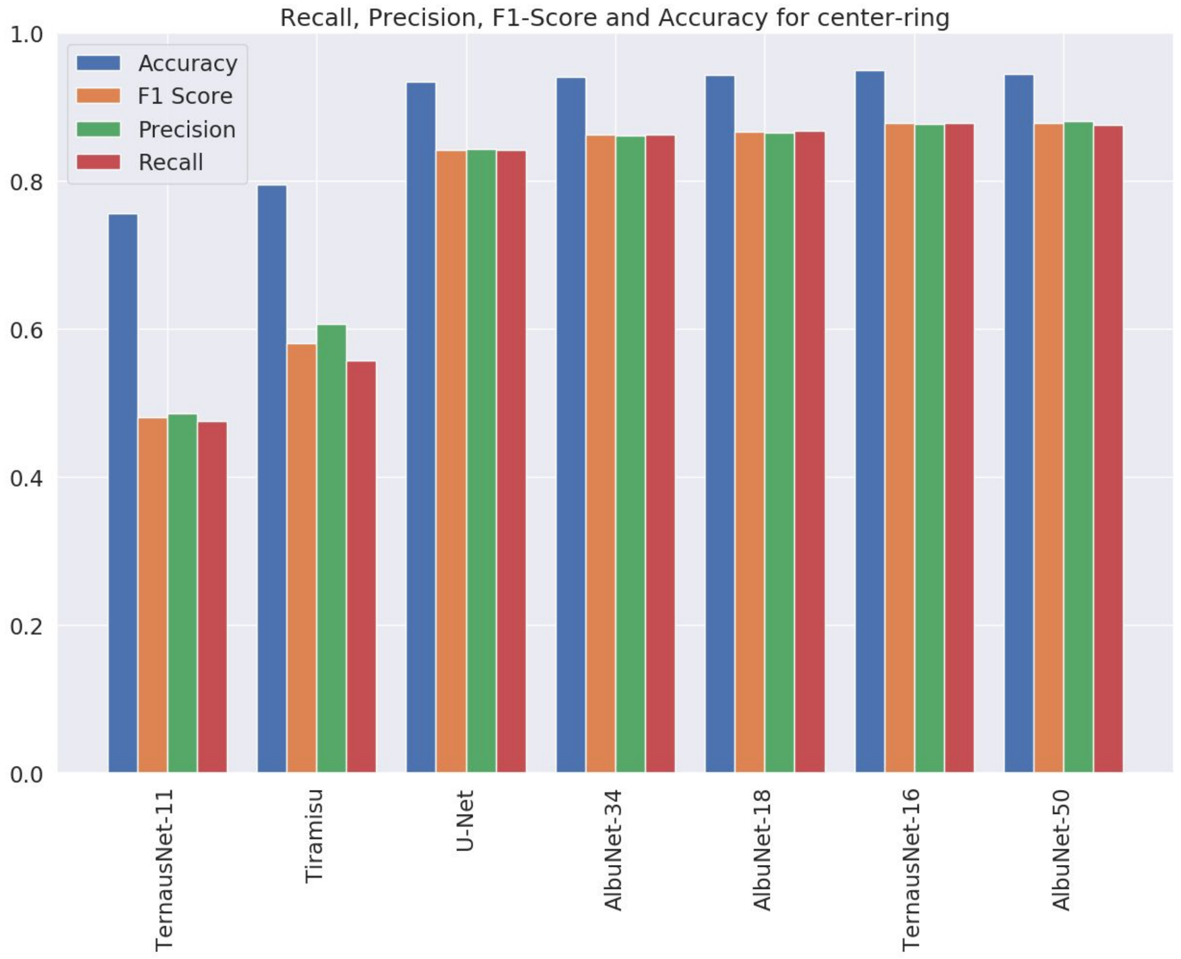
- Encoder is DenseNet-based
- DenseBlocks: Each layer obtains additionally inputs from all preceding layers
- Transition Blocks: Used for downsampling and upsampling

3.3 Approach 1 – Results (1)

Center Prediction Scores



Sample-Weighted Class Scores



3.4 Approach 1 – Conclusion

- The performance of different network architectures was evaluated
- The AlbuNet-50 architecture performs well on all three evaluation methods
- Using AlbuNet-34 or AlbuNet-18 increases efficiency (training time, GPU memory, inference time)
- Further improvements might be:
  - Changing the skip connections between the encoder and the decoder
  - Exploring another architectures like Attention U-Net

Agenda

1 Introduction

2 Setup

3 Approach 1

4 Approach 2

5 Approach 3

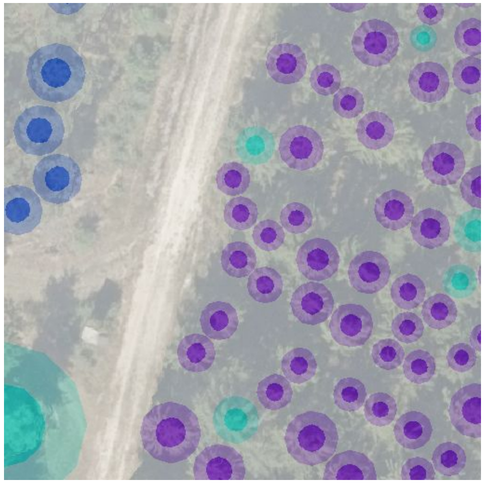
6 Further Results

7 Conclusion

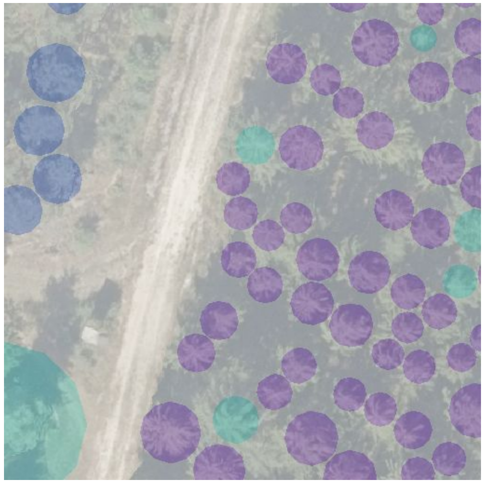
4.1 Approach 2 – Redefining Labels

- **Problem**:
  - **Quality of center predictions:** Strongly varying performance of architectures on center prediction
    BUT: Center point extraction decisive factor for overall performance
- **Goals**:
  - **Label definitions:** Explore different possibilities to define ground truth segmentations masks for tree localization / species classification
  - **Evaluation:** Comparison of models trained on different label definitions

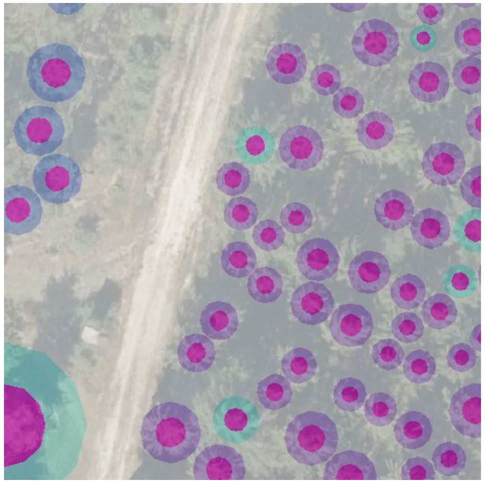
4.2 Approach 2 – Label Definitions



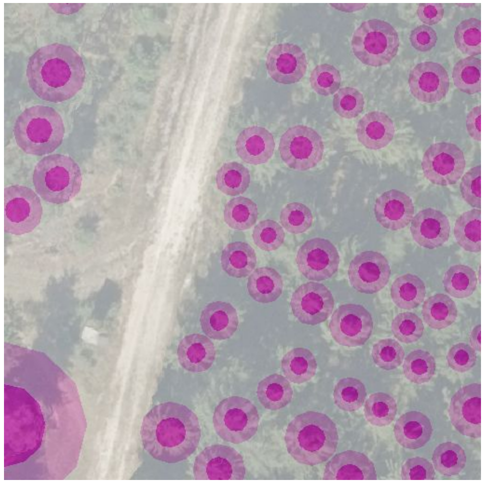
Center-Ring



Ring-only / Two model (class)



Generic-Center

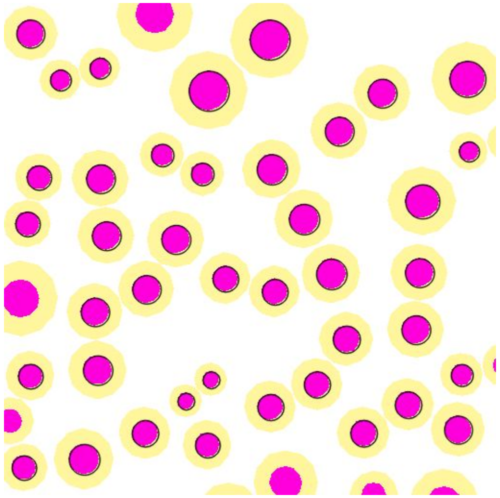


Two model (center)

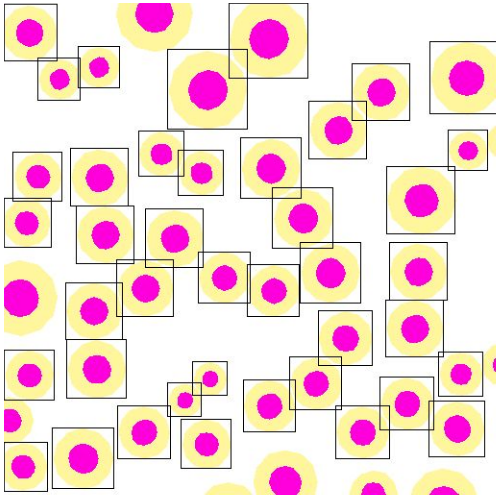
4.3 Approach 2 – Majority-Vote Algorithm

Majority-Vote Algorithm for Species Classification:

- **Tree-center Detection:** Extract tree center points with Blob Detection
- **Enclosing Square:** With the extracted tree center point and the approximated radius a enclosing square is derived
- **Majority-Vote:** Within the enclosing square a majority-vote over all pixels is conducted to derive the species



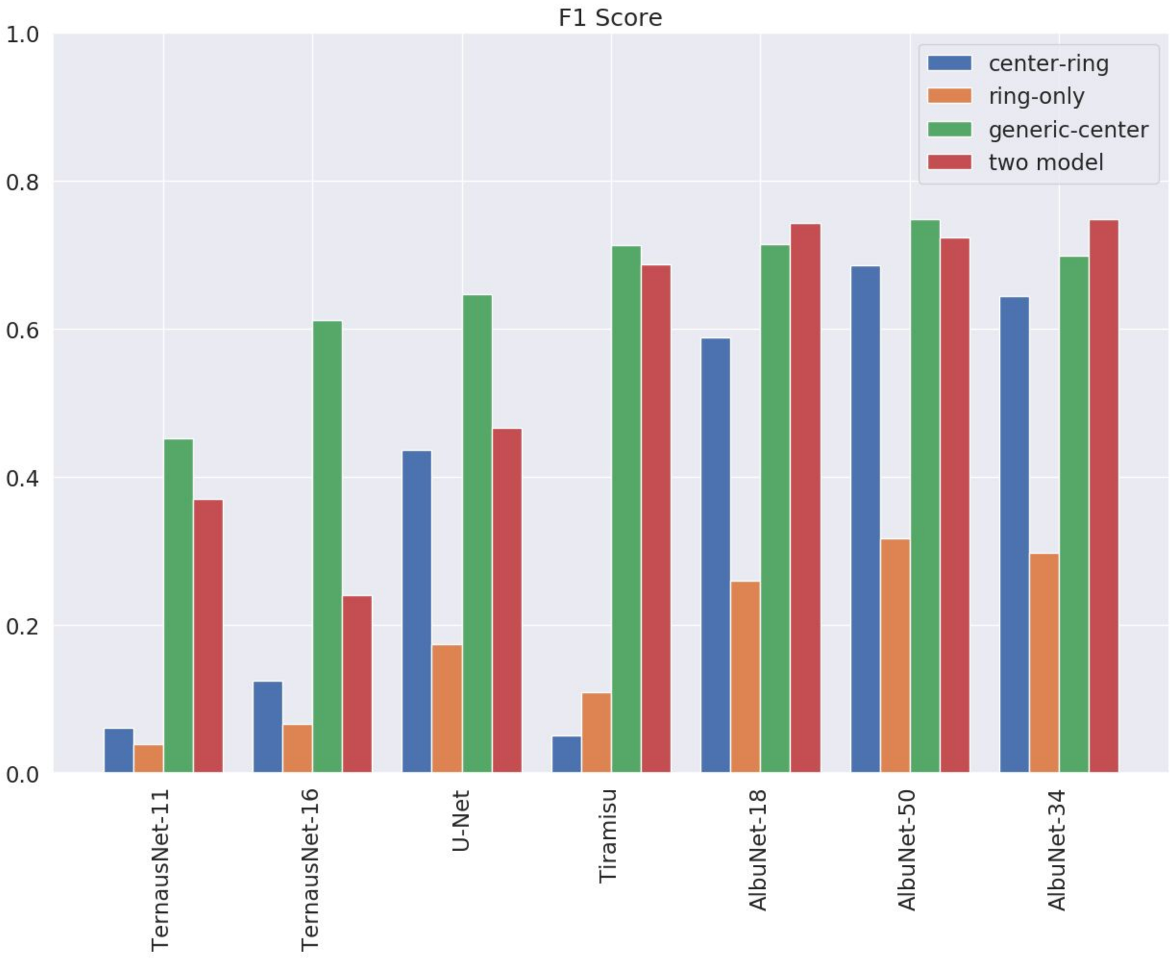
Tree-center Detection



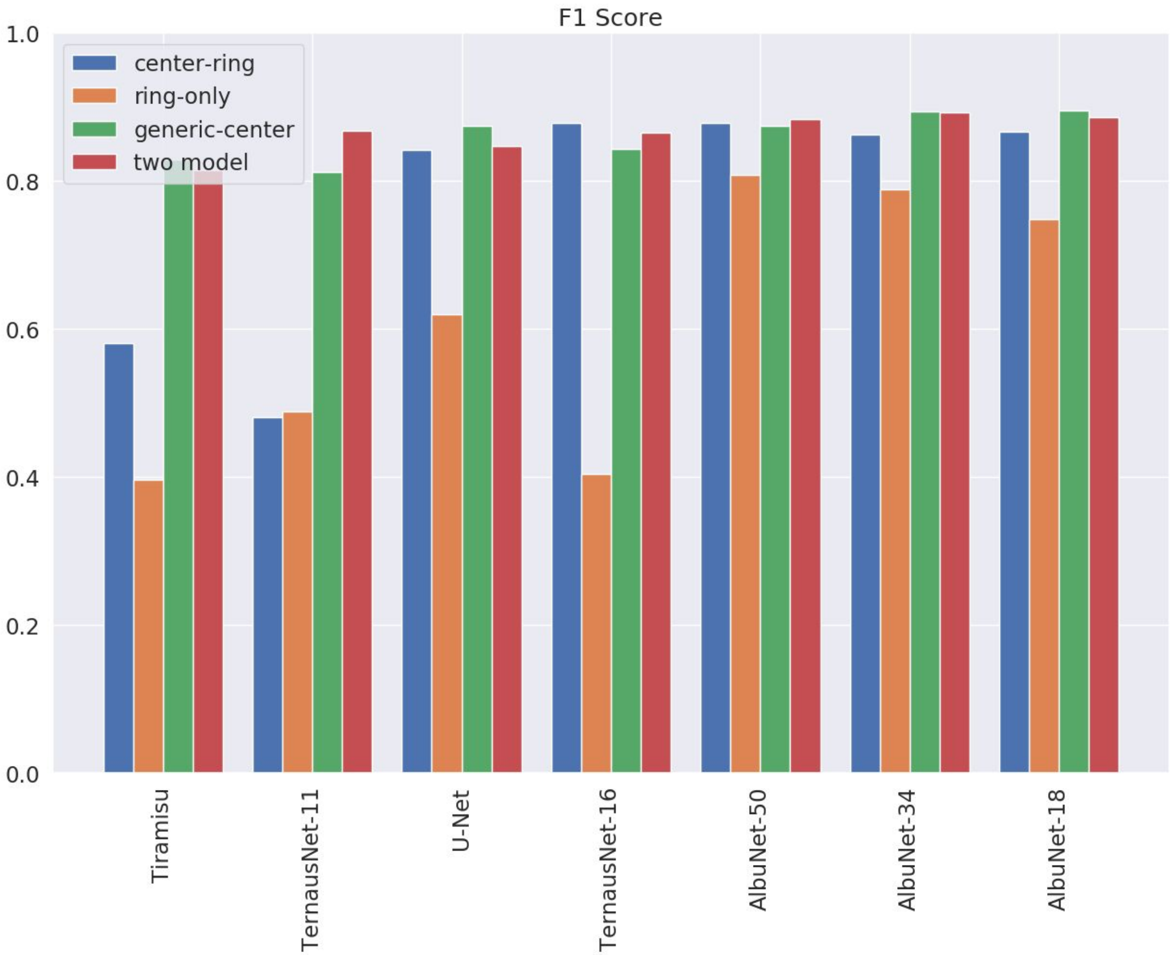
Classification

4.4 Approach 2 – Quantitative Evaluation (1)

Center Prediction Scores

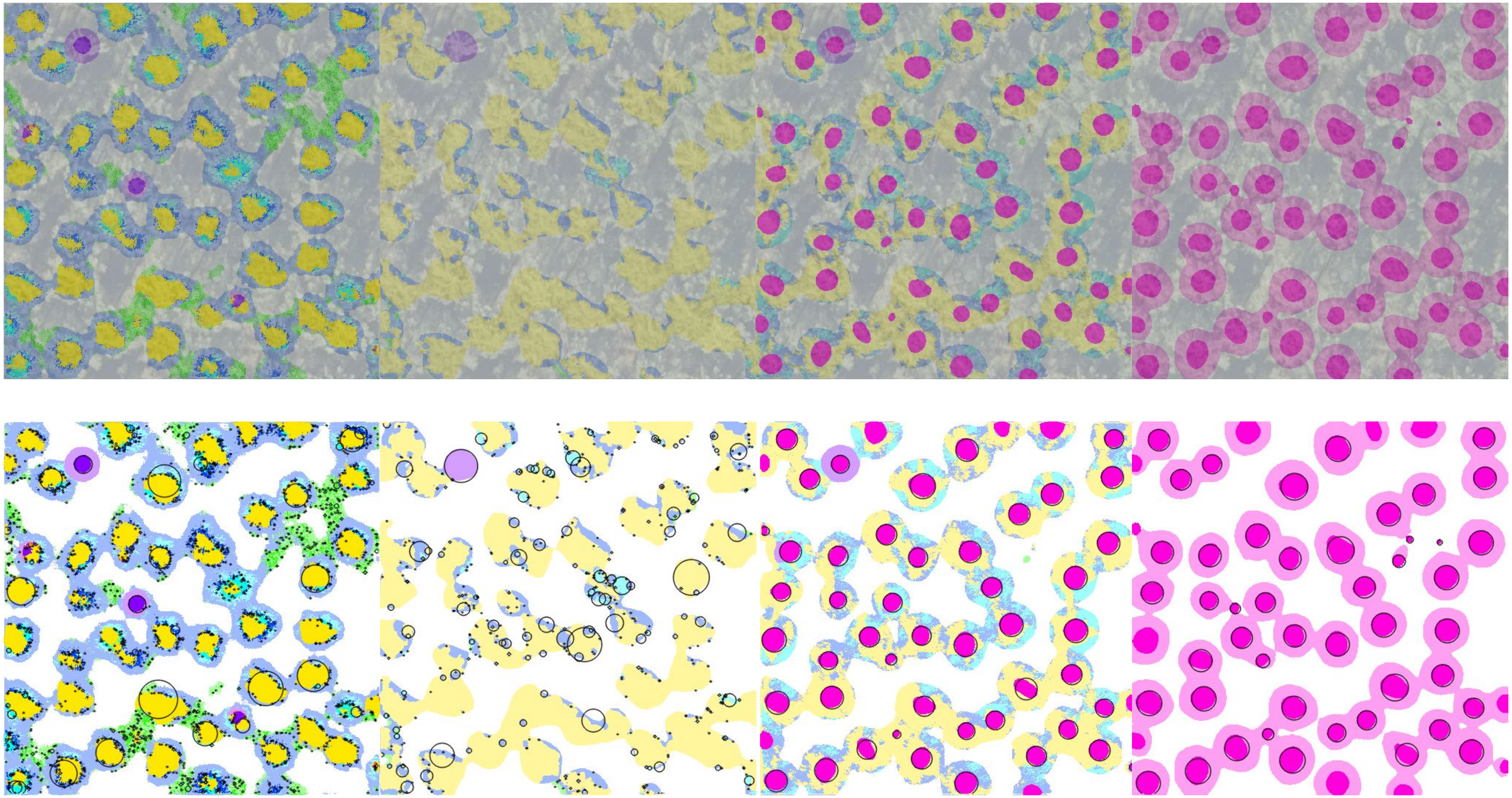


Sample-Weighted Class Scores



4.5 Approach 2 – Qualitative Evaluation (1)

Tiramisu: Tree center extraction



Center-Ring

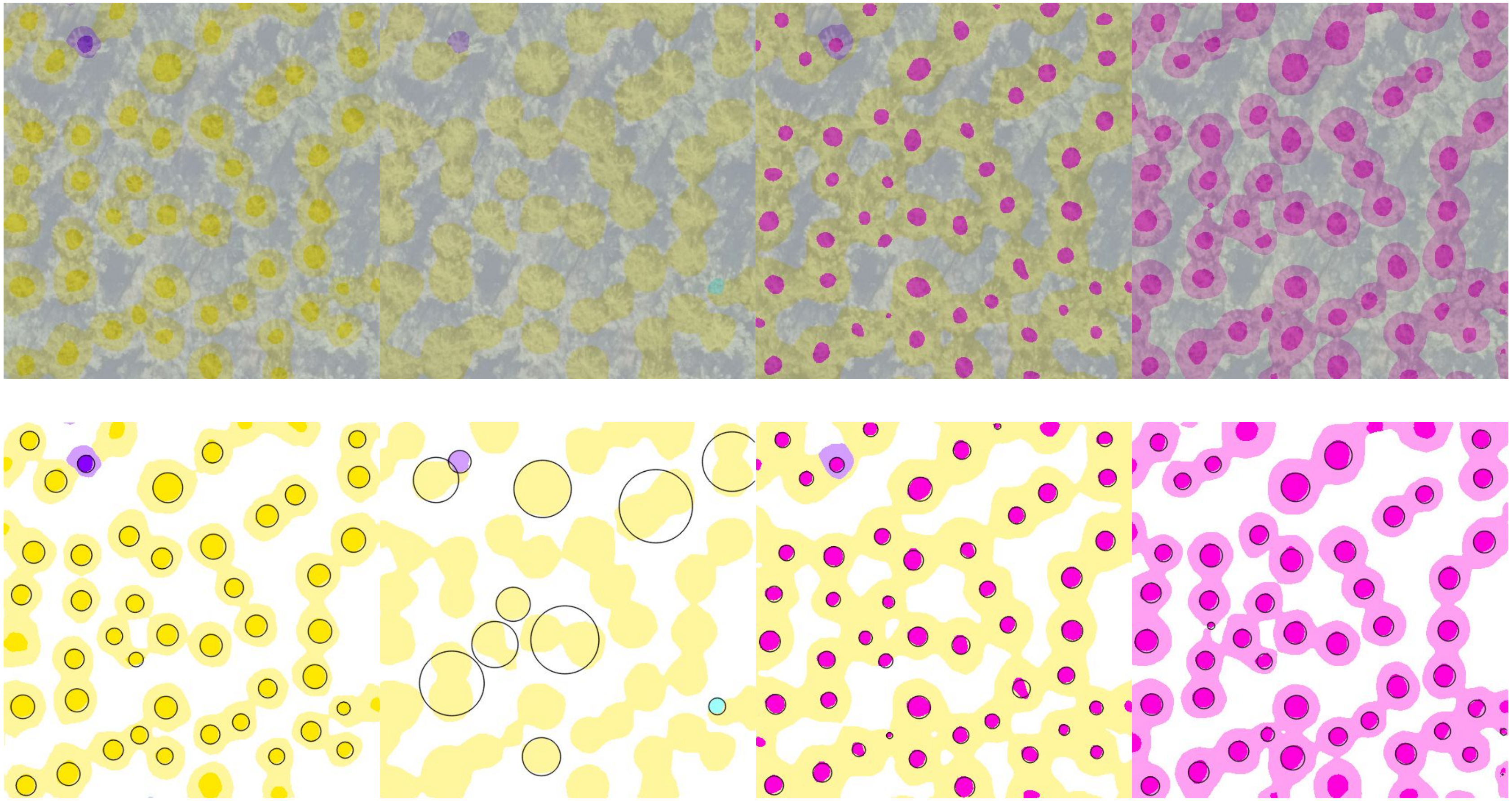
Ring-only

Generic-center

Two model

4.5 Approach 2 – Qualitative Evaluation (2)

AlbuNet: Tree center extraction



4.6 Approach 2 – Conclusion

- **Center / Class Prediction:** Models perform worse on center prediction than on classification
  - **Center point extraction:** Blob detection works reliably for generic-center and two model approach
    ⇒ improves overall performance significantly
  - **Classification:** No significant improvement in species classification

- **Label Definition:** Generic-Center and Two Model approaches yield an improvement of 7-14% for AlbuNet-50.
  - **Generic-center:**
    - + Training of only one model
    - - Less flexibility due to fixed species classes
  - **Two model:**
    - + One generic center model trained on all data
      ⇒ More robust, can be used with different classification models
    - - Training of two models

Agenda

1 Introduction

2 Setup

3 Approach 1

4 Approach 2

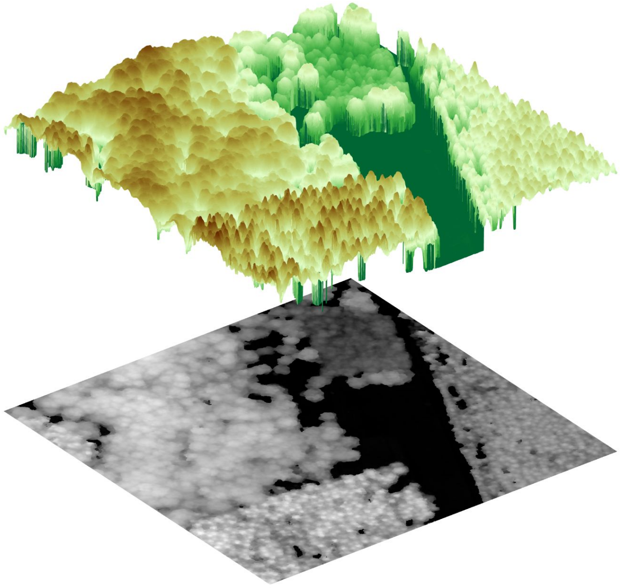
5 Approach 3

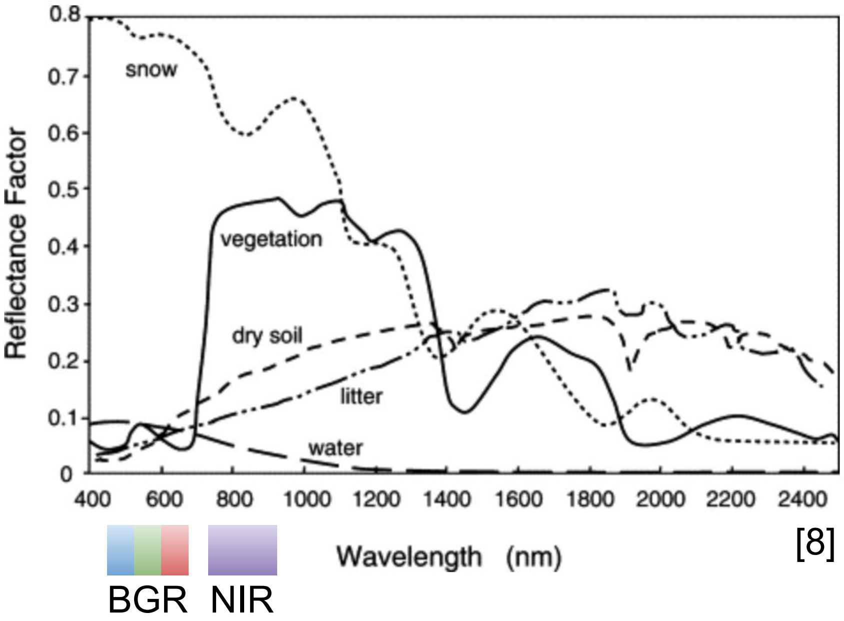
6 Further Results

7 Conclusion

5.1 Approach 3 – Multispectral and 3D Information

- **Goal:** Incorporation of additional information: Near-infrared (NIR) reflectance and Digital Surface Model (DSM)
- **Assumptions:**
  - NIR reflectance provides additional sample from spectral signature and helps with the classification of tree species
  - DSM contains geometric information that helps with the tree center localization and species classification

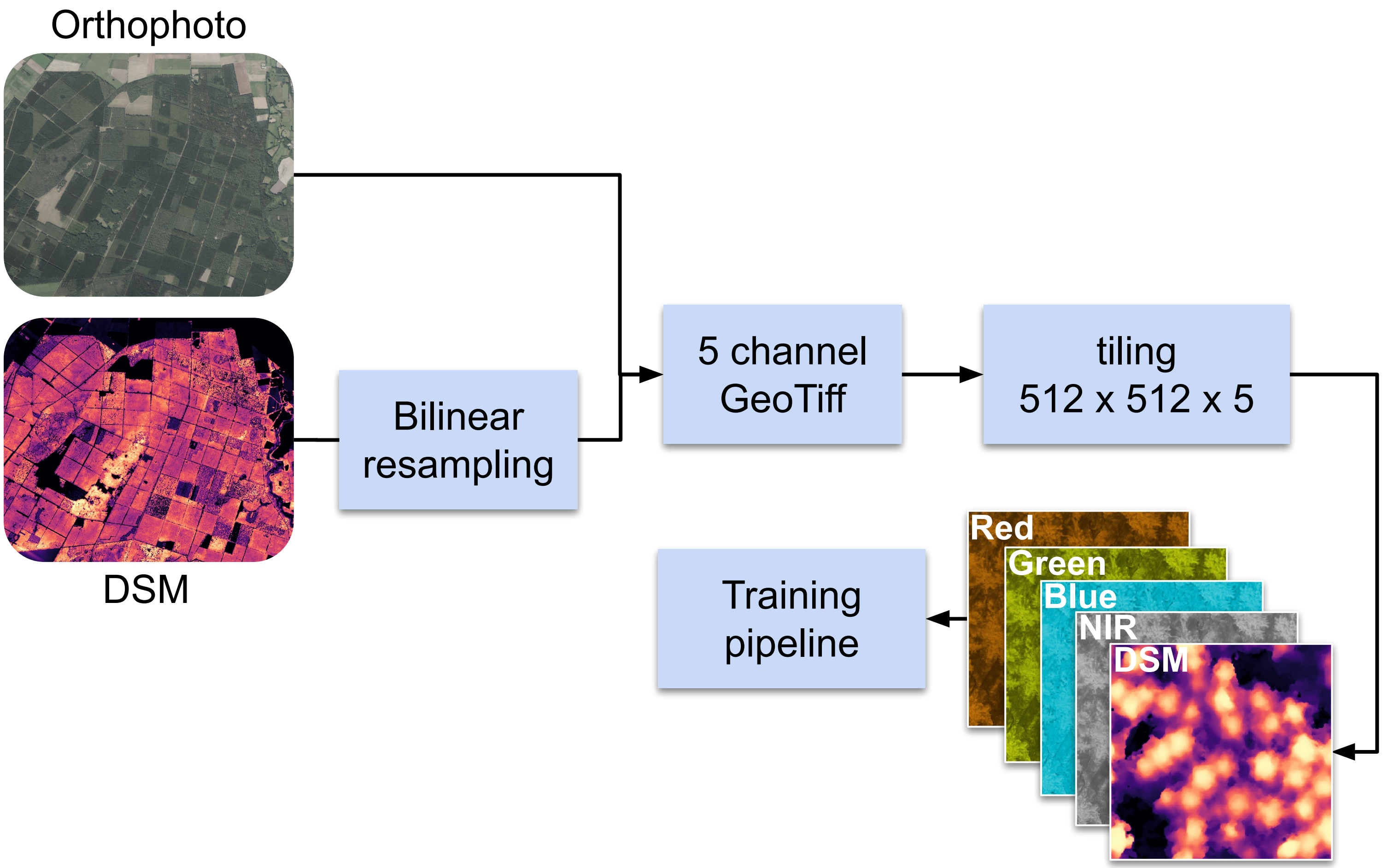




[8]

5.2 Approach 3 – Data fusion

- Fusion of orthophoto and Digital Surface Model
- Adaptation of processing pipeline to work with fused data



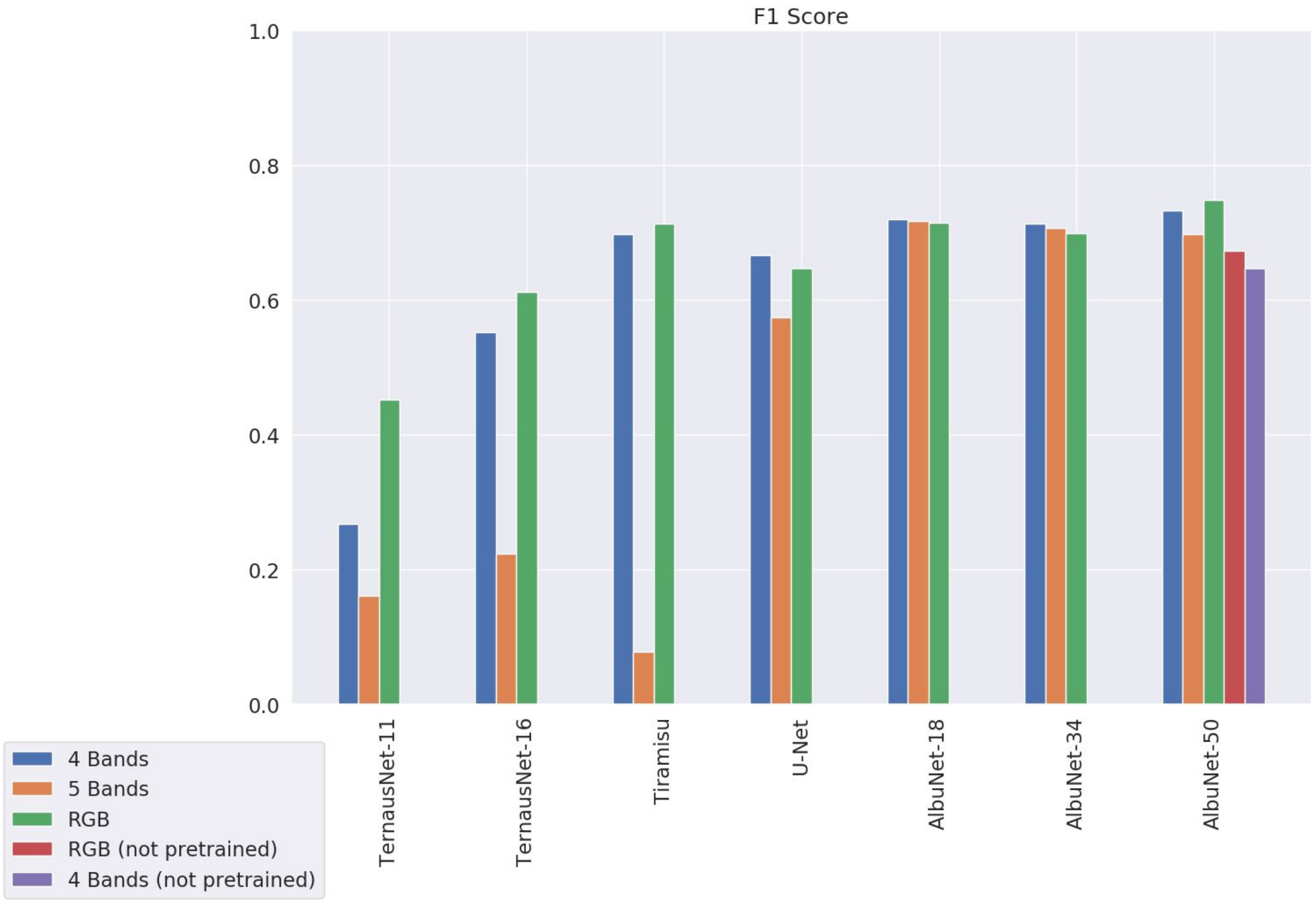
5.3 Approach 3 – Training plan

- All model architectures from the previous approaches were trained

Run	Channels	Models	Epochs	Pre-trained
1	RGB + NIR	all	500	True
2	RGB + NIR + DSM	all	500	True
3a	RGB	AlbuNet-50	2000	False
3b	RGB + NIR	AlbuNet-50	2000	False

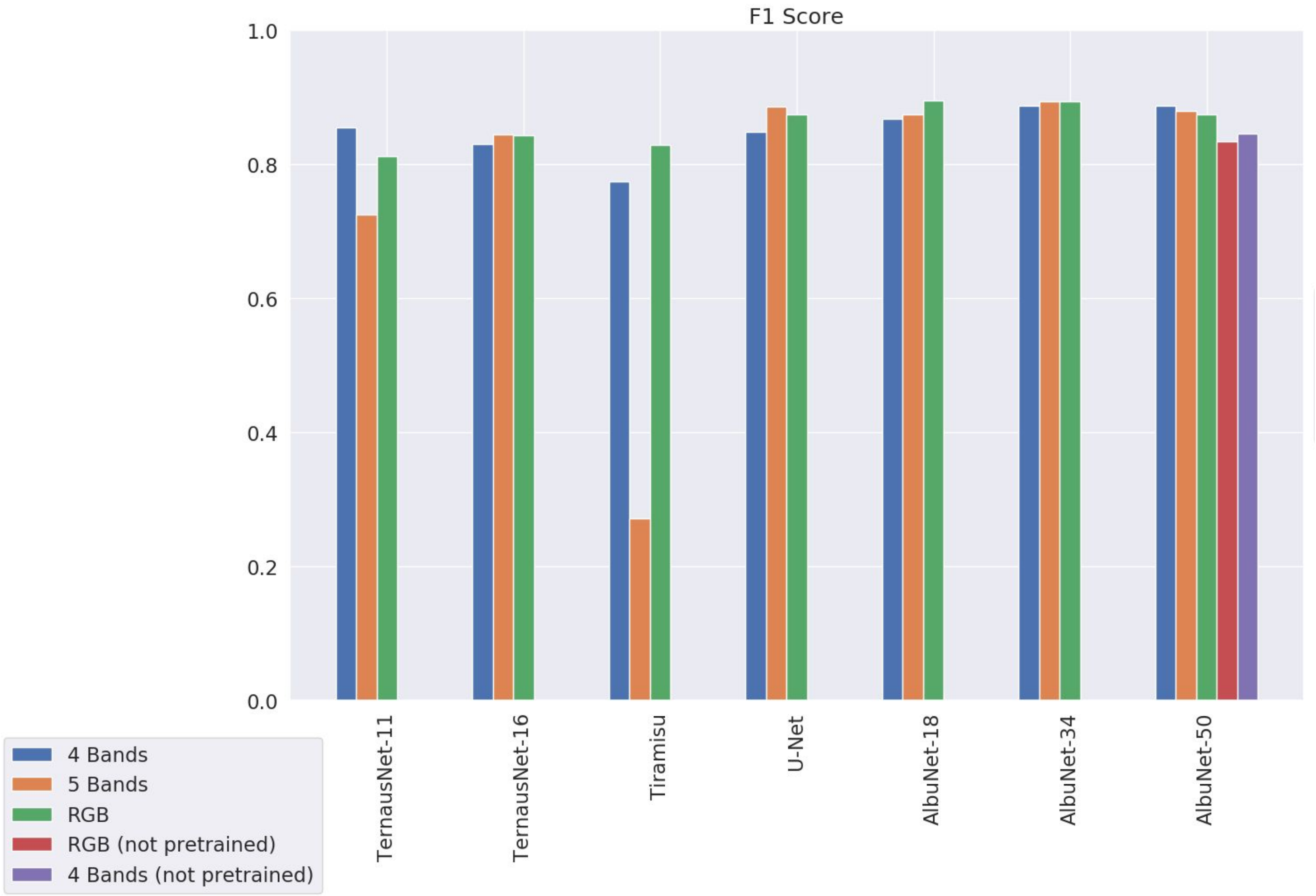
- Transfer learning: To assess influence of transfer learning one model was trained from scratch (AlbuNet-50) on two configurations:
  - RGB
  - RGB + NIR

Center Prediction Scores



5.4 Approach 3 – Results (2)

Sample-Weighted Class Scores



5.5 Approach 3 – Conclusion

- **Center prediction**
  - Center prediction seems not to profit from additional information
  - No significant difference for AlbuNet family
    - Still performs best
  - DSM decreases performance for other architectures
  - NIR has a smaller impact on scores than DSM
- **Class prediction**
  - No significant change for AlbuNet family
  - Impact of NIR and DSM channel weaker than for center prediction
  - Could be valuable for different set of tree species
- **Transfer learning**
  - Transferability of knowledge obtained from ImageNet can be seen
  - Class prediction: Almost caught up with pre-trained models
  - Training from scratch should be considered for future tests
    - Might improve performance with different set of tree species

Agenda

1 Introduction

2 Setup

3 Approach 1

4 Approach 2

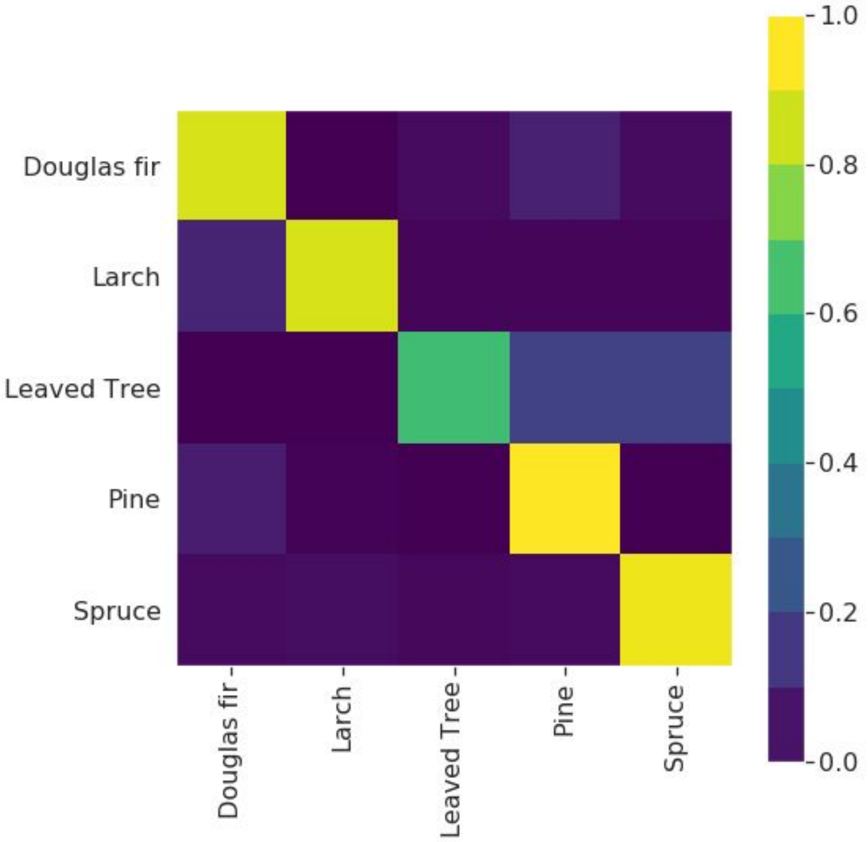
5 Approach 3

6 **Further Results**

7 Conclusion

Evaluation based on tree species:

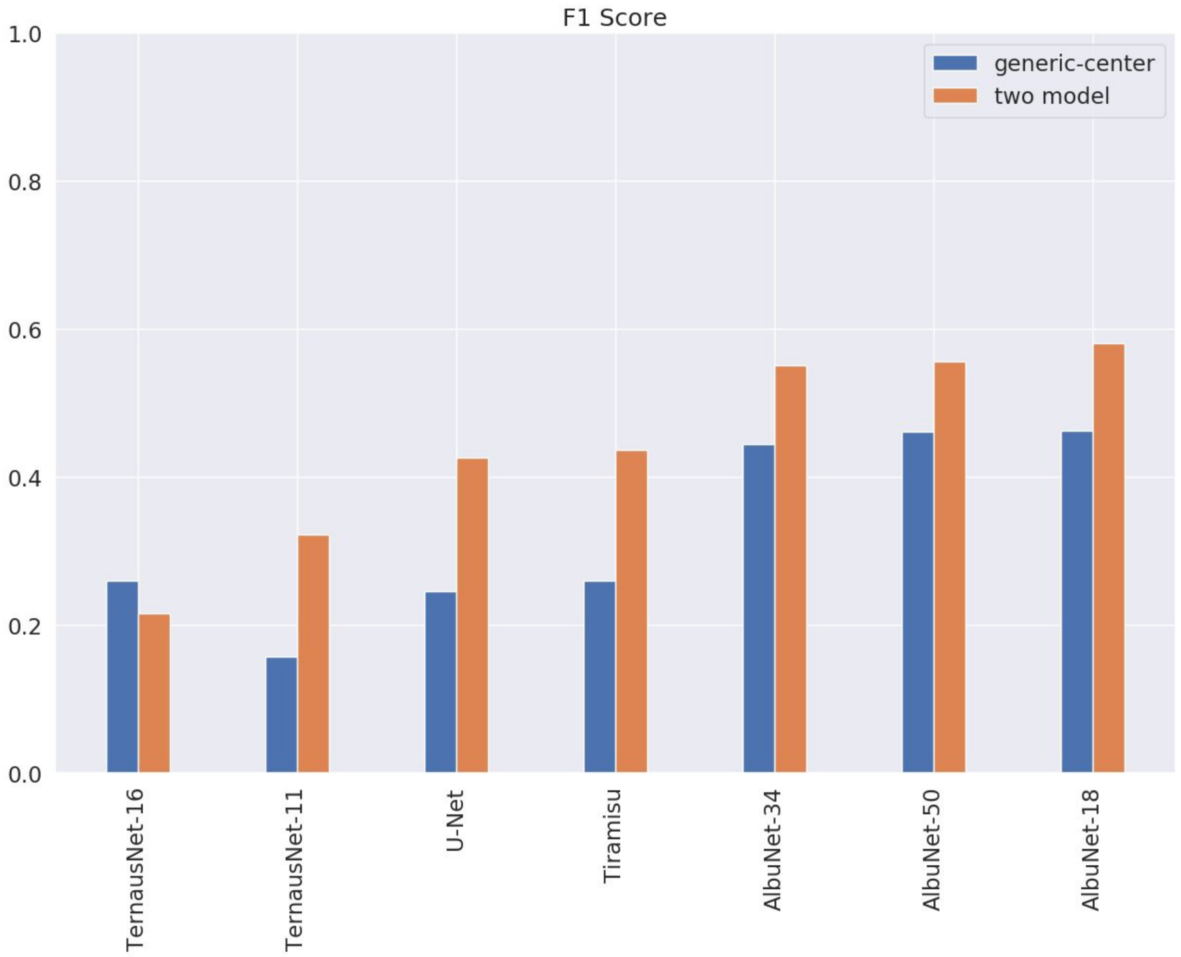
- Confusion between conifers
- *Leaved Tree* performs the worst
  - **Not** important for foresters
  - Only few samples
  - Tree centers hard to predict
- *Spruce* and *Pine* perform the best
  - **Most** important tree species for foresters
  - Lots of samples



Dead Tree	Douglas fir	Larch	Leaved Tree	Pine	Spruce
0	82	224	33	146	158

6.2 Further results – Evaluation on Data Set B

Center Prediction Scores



Agenda

1 Introduction

2 Setup

3 Approach 1

4 Approach 2

5 Approach 3

6 Further Results

7 **Conclusion**

7 Conclusion

- **Best performing model:** AlbuNet-based architectures
  - No significant difference between AlbuNet-50 and AlbuNet-34
  - AlbuNet-34 has less trainable parameters
    ⇒ Decreases training and inference time, but also GPU resources
- **Best label definition:** Generic-Center and Two Model
  - Generic-Center only needs training of one model
  - Two Model generalizes better on unseen data
- **Use of multispectral data:**
  - No significant difference in performance for best models
  - Still worth to test if set of tree species changes
  - May be helpful for detecting unhealthy trees

References

[1] Kingma DP and Ba J. Adam: A method for stochastic optimization. arXiv, abs/1412.6980, 2014.

[2] Berman M and Blaschko MB. Optimization of the jaccard index for image segmentation with the lovász hinge. Computing Research Repository, abs/1705.08790, 2017.

[3] Keiichi AB Suzuki S. Topological structural analysis of digitized binary images by border following. Computer vision, graphics, and image processing, 30(1):32–46, 1985.

[4] A. Shvets, V. Iglovikov, A. Rakhlinand, and A. Kalinin. Angiodysplasia detection and localization using deep convolutional neural networks. 17th IEEE International Conference on Machine Learning and Applications, pages 612–617, 04 2018.

[5] Ronneberger O, Fischer P, and Brox T. U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention, pages 234 – 241. Springer International Publishing, 2015.

[6] Iglovikov V and Shvets A. Ternausnet: U-net with VGG11 encoder pre-trained on imagenet for image segmentation. Computing Research Repository, abs/1801.05746, 2018.

[7] Jégou S, Drozdzal M, Vázquez D, Romero A, and Bengio Y. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. Computing Research Repository, 2016.

[8] Knipling EB (1970): Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation, Remote Sensing of Environment, 1(3): 155-159

[9] Deng J, Dong Wand Socher R, Li LJ, Li K, and Fei-Fei L. (2009): ImageNet: A large-scale hierarchical image database. In 2009 IEEE Conference on Computer Vision and Pattern Recognition, pages 248–255

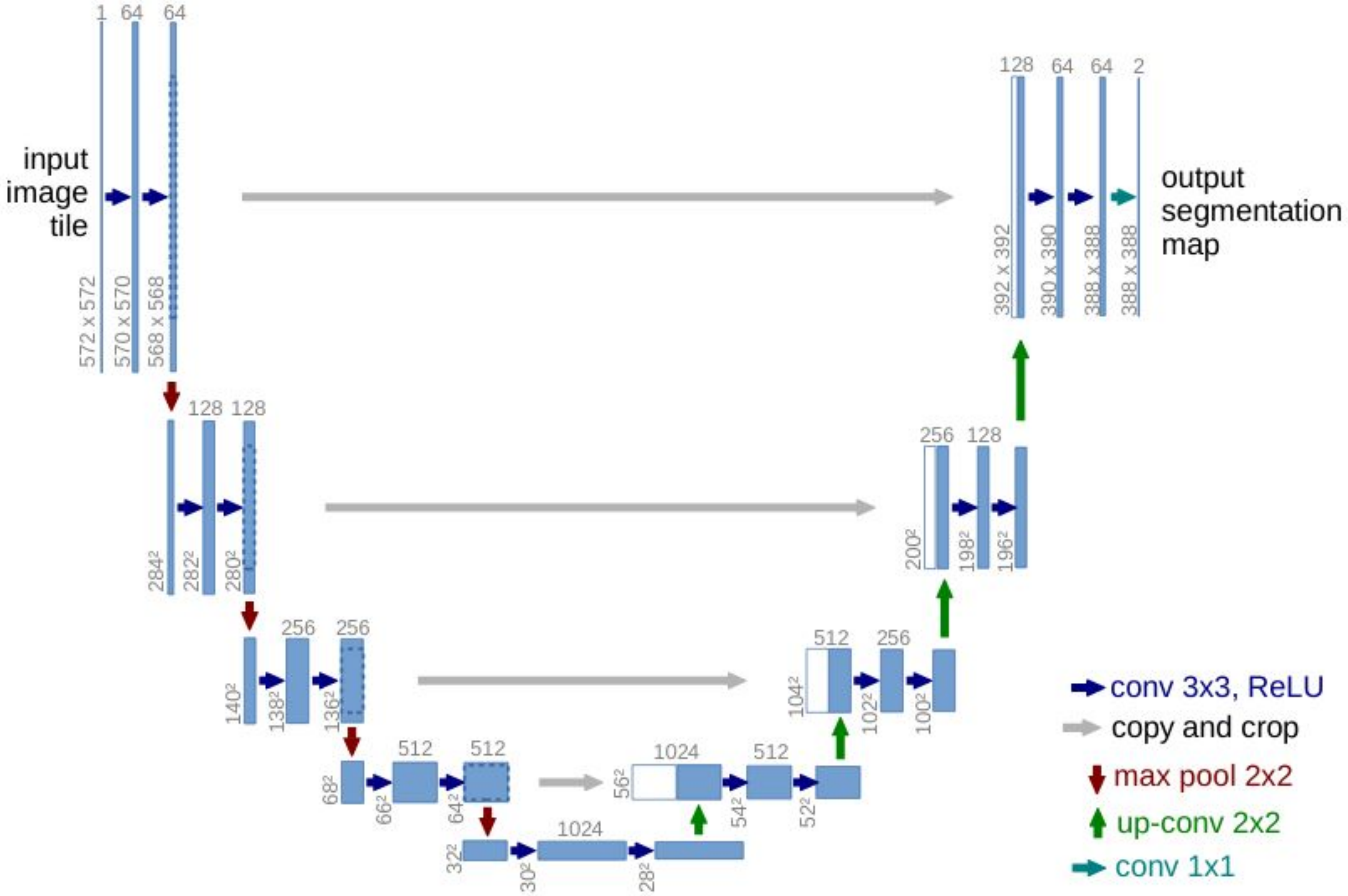
[10] Image: FCNN. Online available:
https://www.jeremyjordan.me/semantic-segmentation/#advanced_unet [last access: 17.02.2020]

Thank you for your attention!

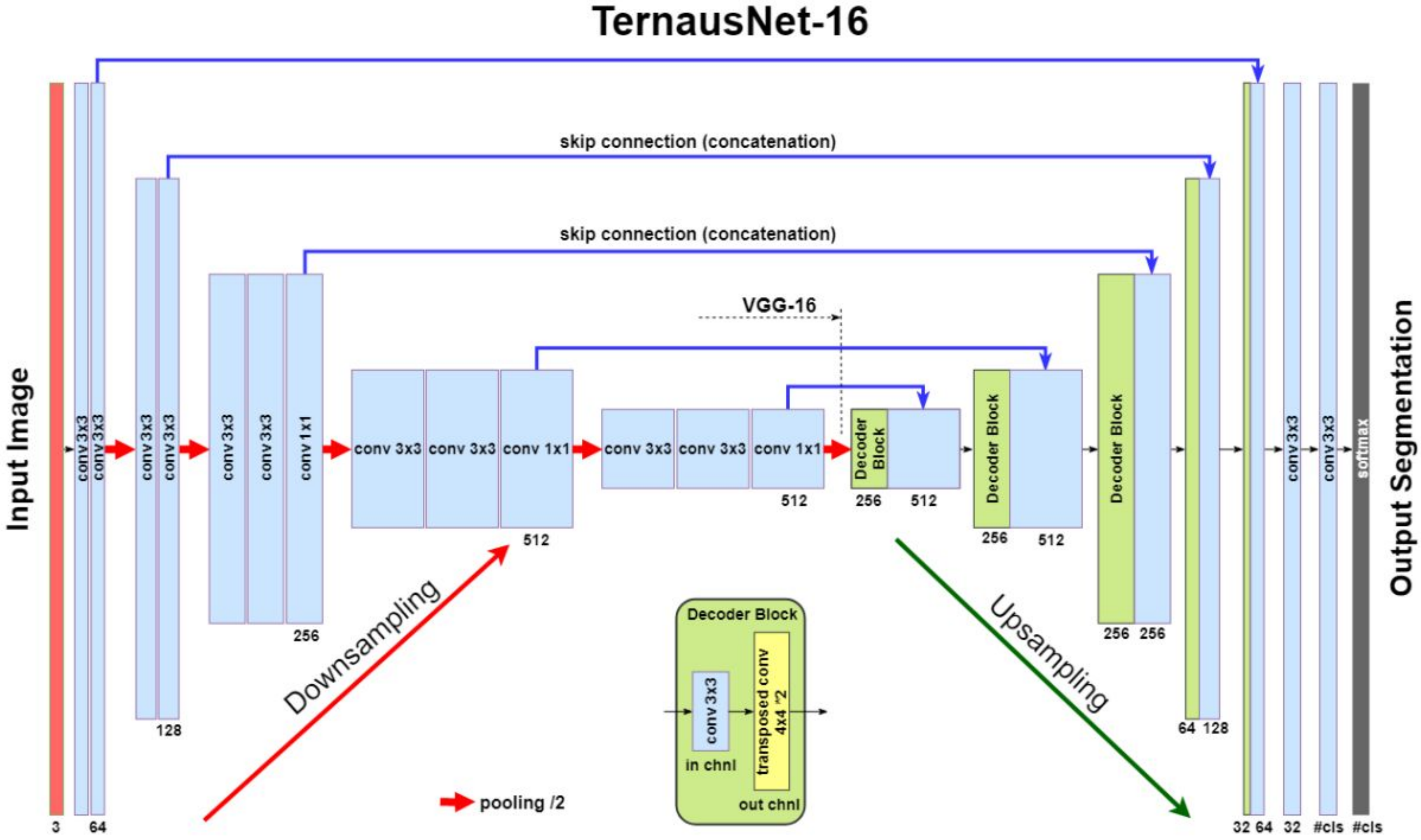
Backup – Future Research

- **Labeling techniques**
  - Two model approach: Update *ring-only* labels to area segmentation
  - Training different models for different selection of models (i.e. use generic class for all other species)
- **Task specific model development**
  - Two model approach: combine different architectures
  - Feed classification model with center prediction confidence mask
- **Multispectral information**
  - Evaluate performance on bigger data set
  - Use NIR channel to predict diseases or water-stress
- **Blob detection and species classification**
  - Improve performance on image borders
  - Conditional Random Fields for post-processing
  - Majority-Vote: weight input of pixel by distance to center
- **Improving architectures**
  - New architectures: Attention U-Net, QuickNat
  - Regularization during training: Dropout, Weight Regularization

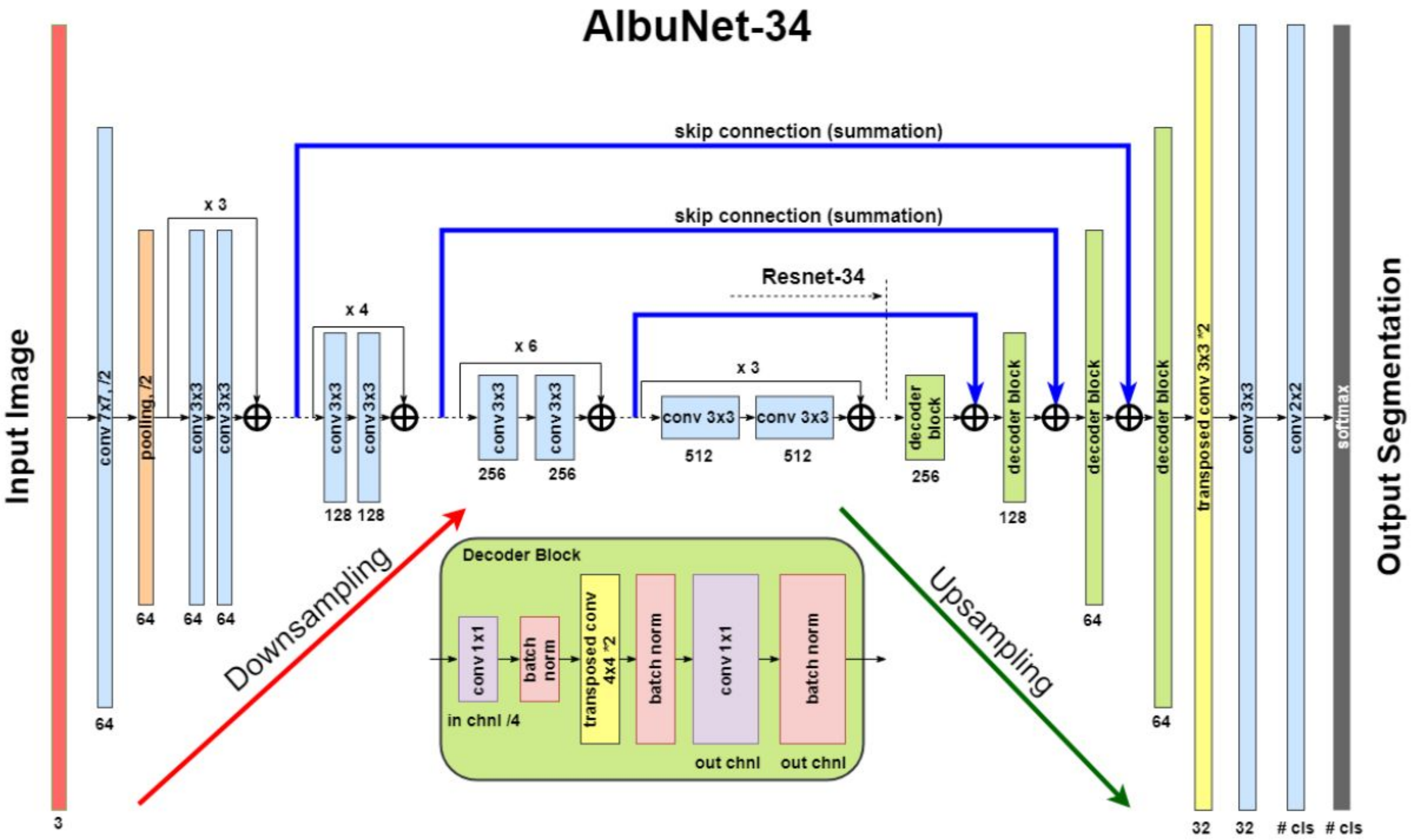
Backup – Architectures (1)



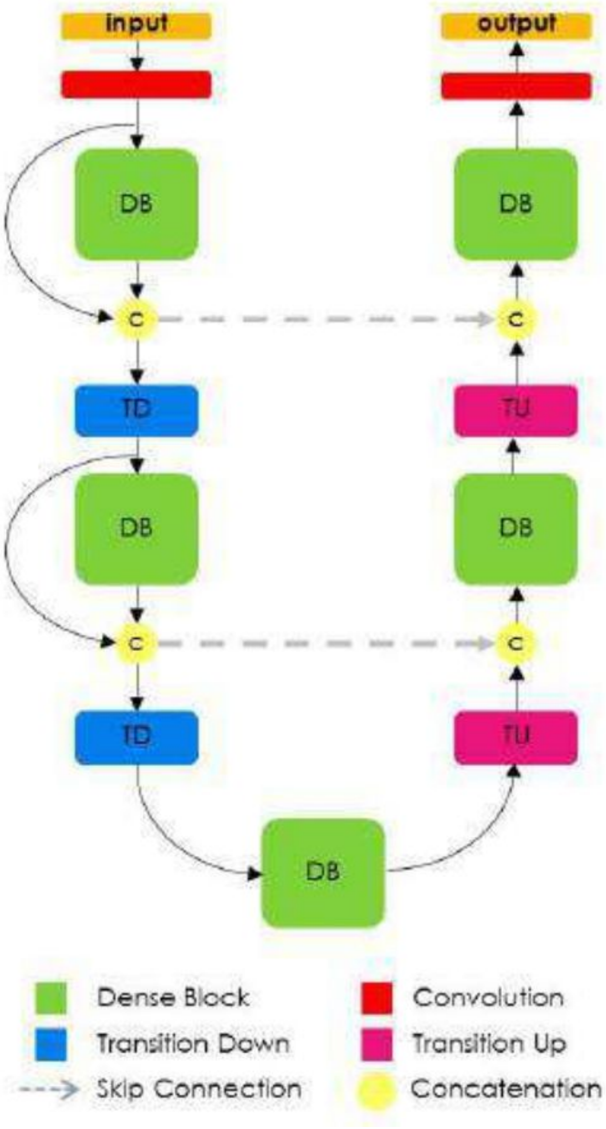
Backup – Architectures (2)

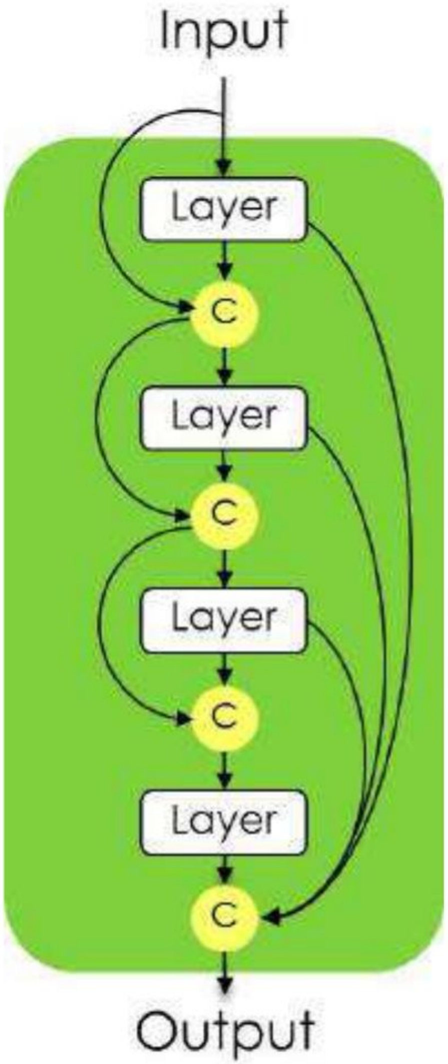


Backup – Architectures (3)

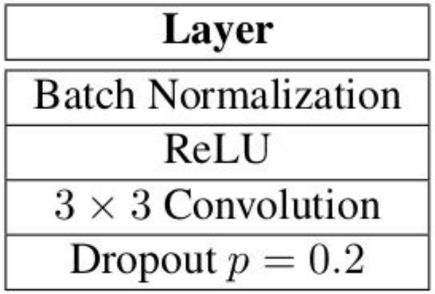


Backup – Architectures (4)

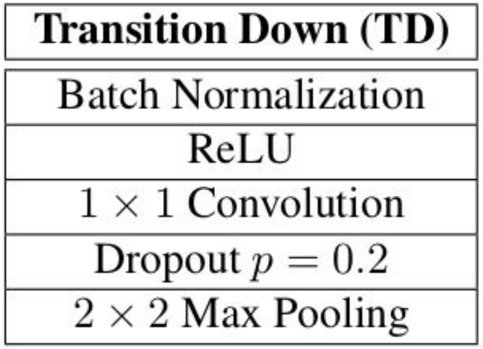




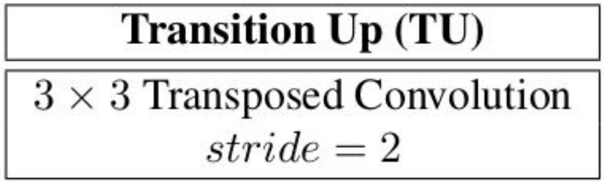
Backup – Architectures (5)



Layer
Batch Normalization
ReLU
3×3 Convolution
Dropout $p = 0.2$

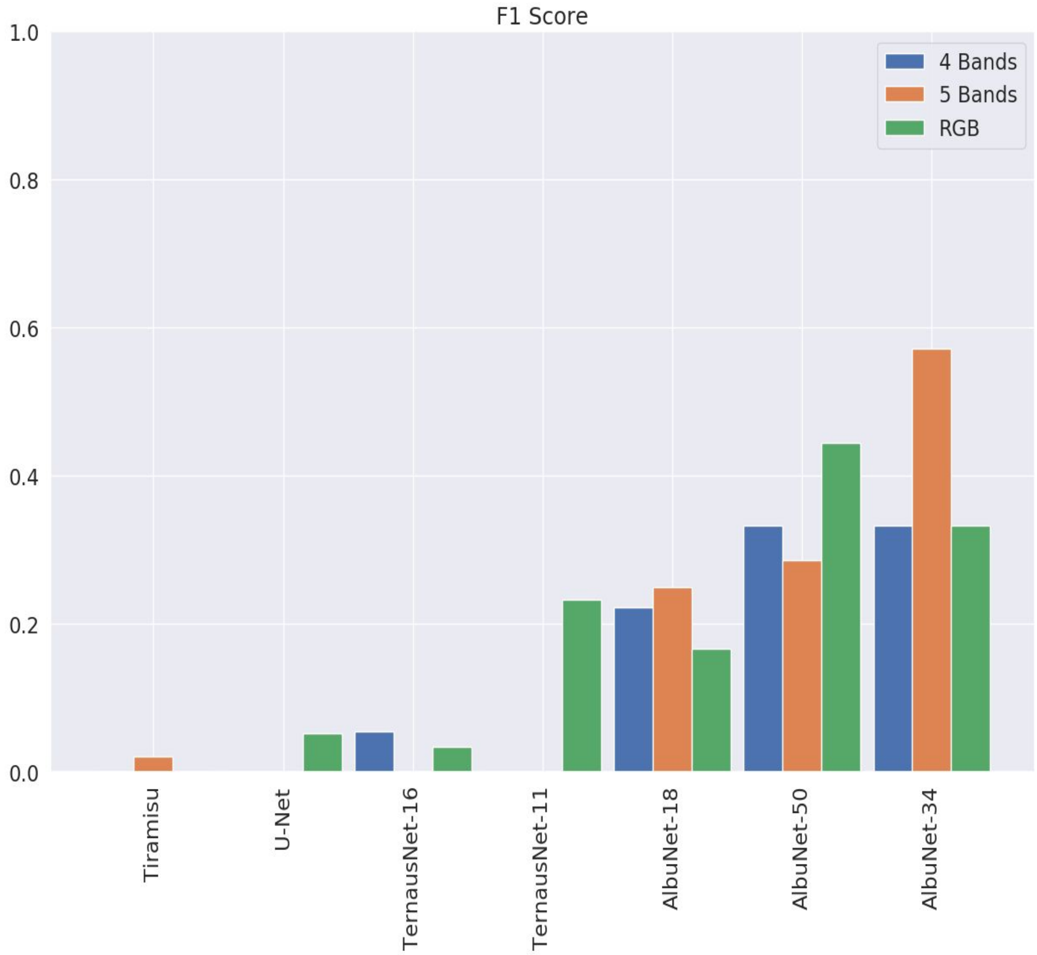


Transition Down (TD)
Batch Normalization
ReLU
1×1 Convolution
Dropout $p = 0.2$
2×2 Max Pooling



Transition Up (TU)
3×3 Transposed Convolution $stride = 2$

Dead Tree Classification



Backup – Performance metrics

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

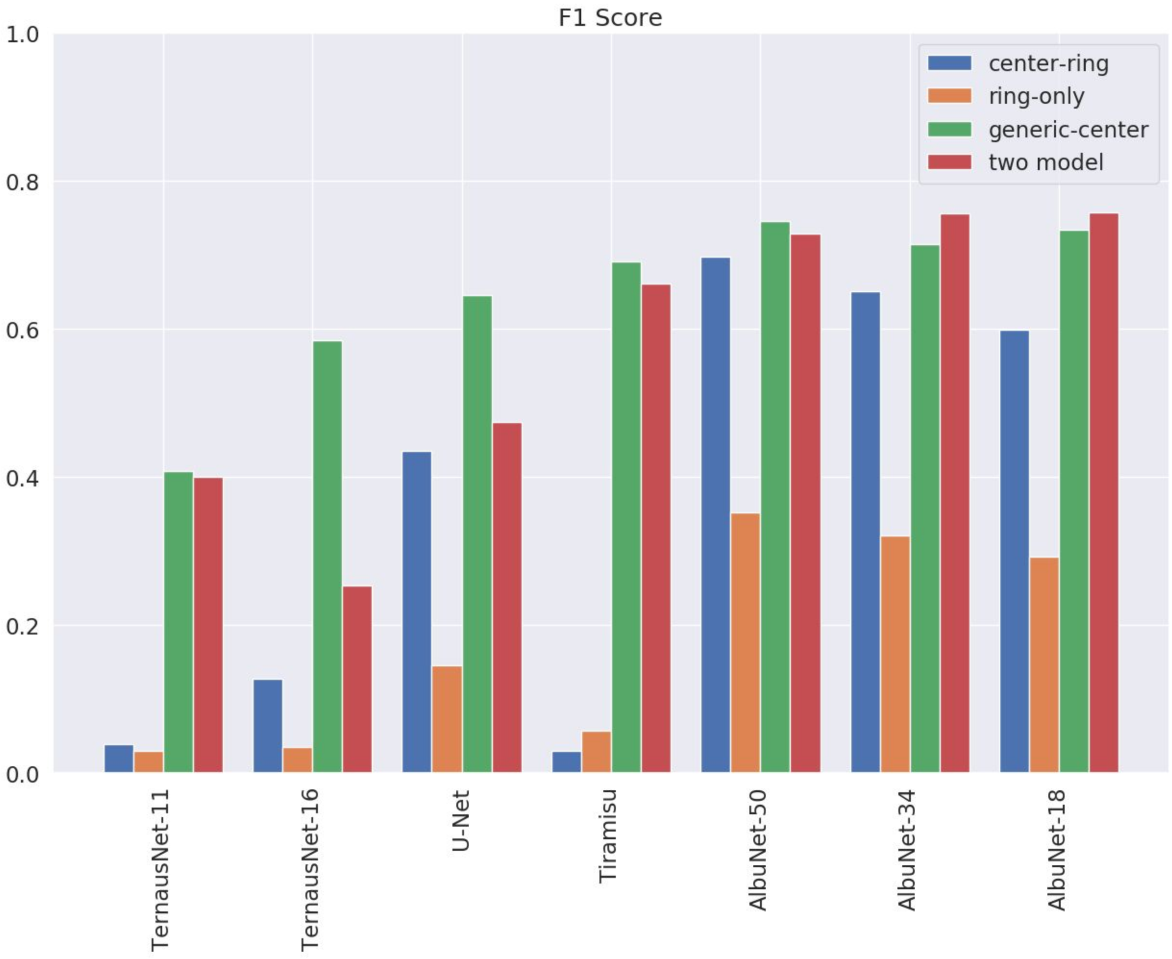
$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{F1} = 2 \cdot \frac{\text{Recall} \cdot \text{Precision}}{\text{Recall} + \text{Precision}}$$

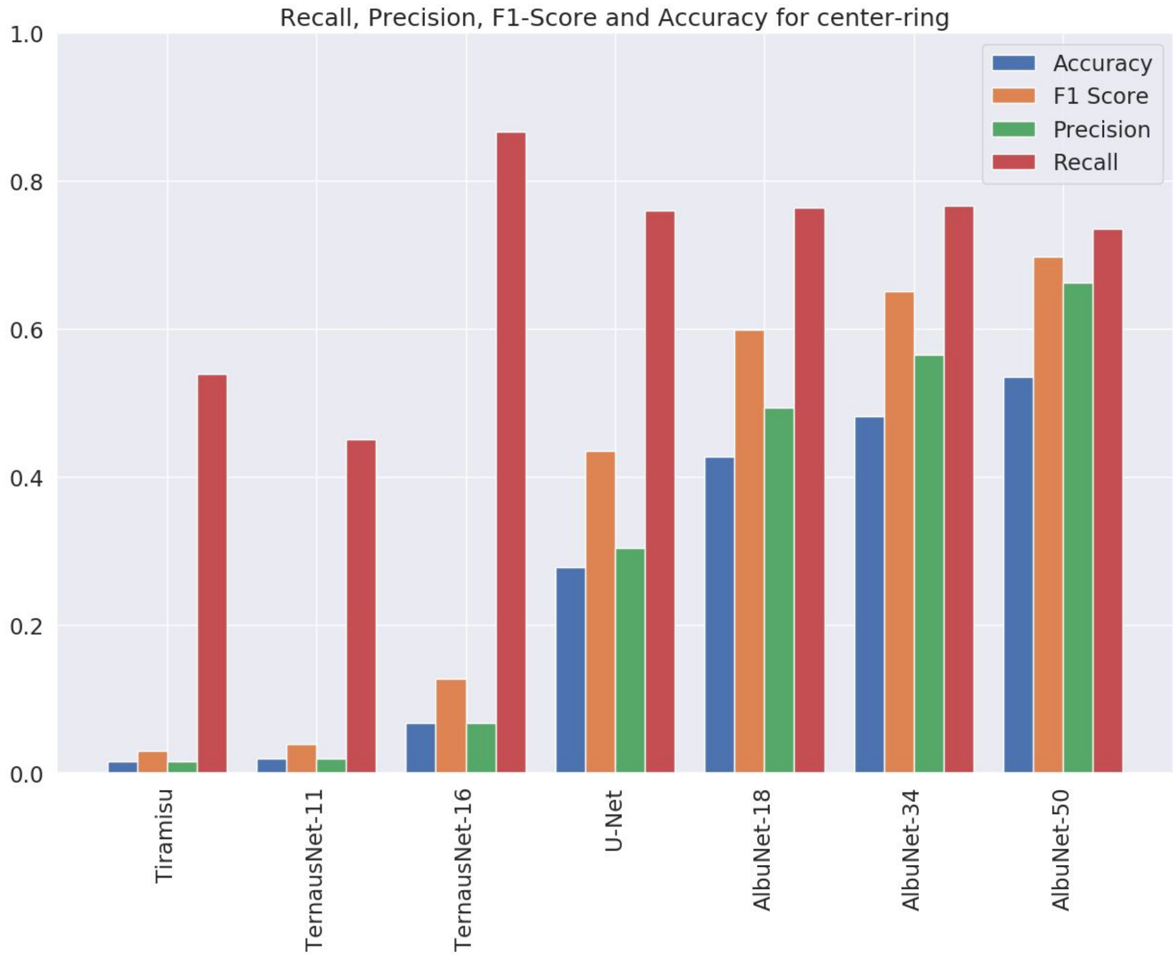
Backup - Approach 2 (1)

Aggregate Class-Center Scores



Backup - Approach 2 (2)

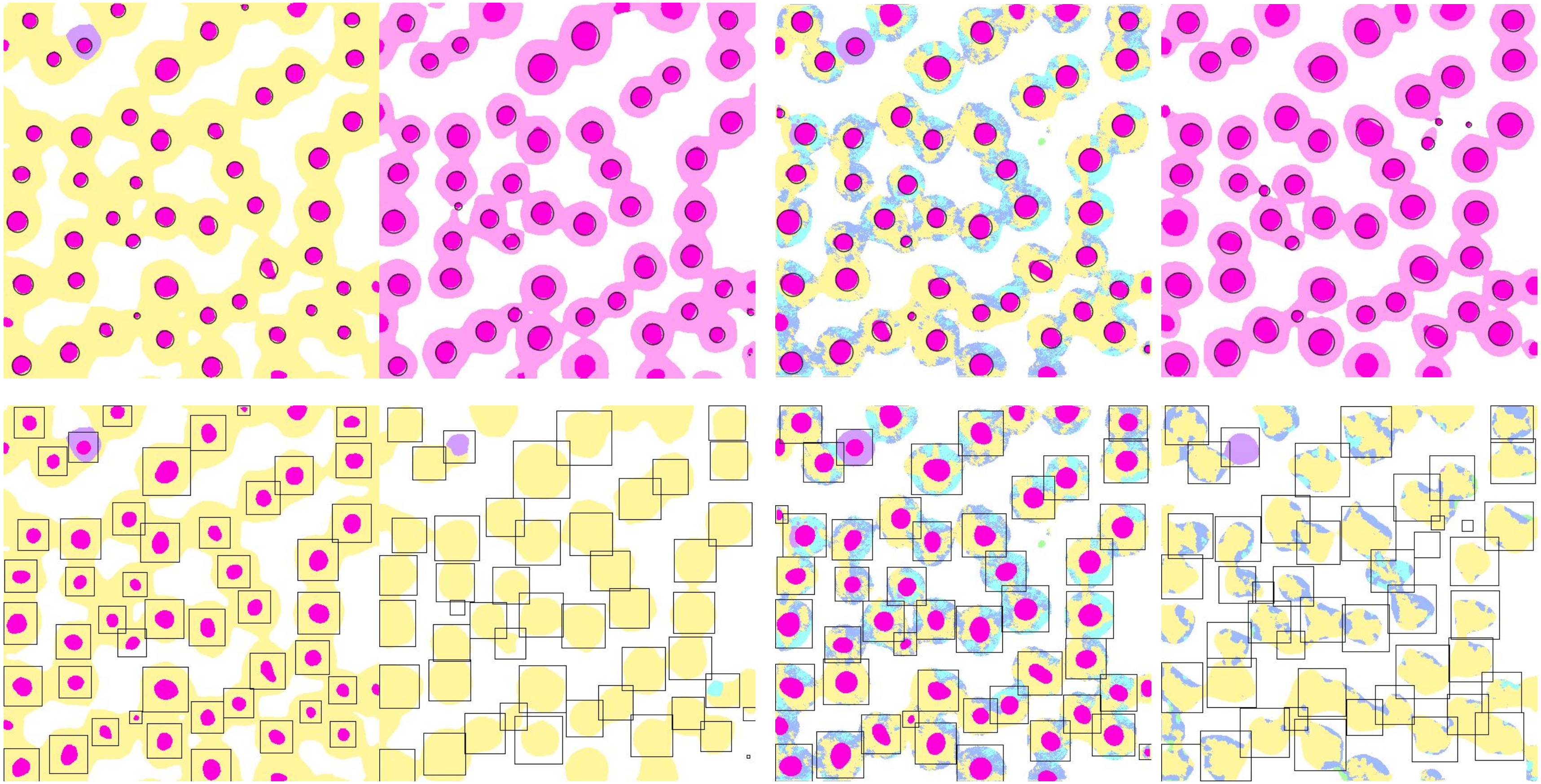
Aggregate Class-Center Scores



Backup - Approach 2 (3)

AlbuNet: Majority-Vote

Tiramisu: Majority-Vote



Generic-Center

Two model

Generic-Center

Two model