Mapping forest tree species in high resolution UAV-based RGB-imagery by means of convolutional neural networks

Felix Schiefer, Teja Kattenborn, Annett Frick, Julian Frey, Peter Schall, Barbara Koch, Sebastian Schmidtlein

PhD-Seminar

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Mapping forest tree species in high resolution UAV-based RGB-imagery by means of convolutional neural networks

Felix Schiefer[a,*], Teja Kattenborn[a,b], Annett Frick[c], Julian Frey[d,e], Peter Schall[f], Barbara Koch[e], Sebastian Schmidtlein[a]

[a] Institute of Geography and Geoecology, Karlsruhe Institute of Technology (KIT), 76131 Karlsruhe, Germany
[b] Remote Sensing Centre for Earth System Research, Leipzig University, 04103 Leipzig, Germany
[c] Luftbild Umwelt Planung GmbH (LUP), Große Weinmeisterstraße 3a, 14469 Potsdam, Germany
[d] Chair of Forest Growth and Dendroecology, University of Freiburg, 79106 Freiburg, Germany
[e] Chair of Remote Sensing and Landscape Information Systems, University of Freiburg, 79106 Freiburg, Germany
[f] Silviculture and Forest Ecology of the Temperate Zones, University of Göttingen, 37077 Göttingen, Germany

ARTICLE INFO

Keywords:
Deep learning
Forest inventory
Convolutional neural networks
Tree species classification
Unmanned aerial systems
Temperate forests

ABSTRACT

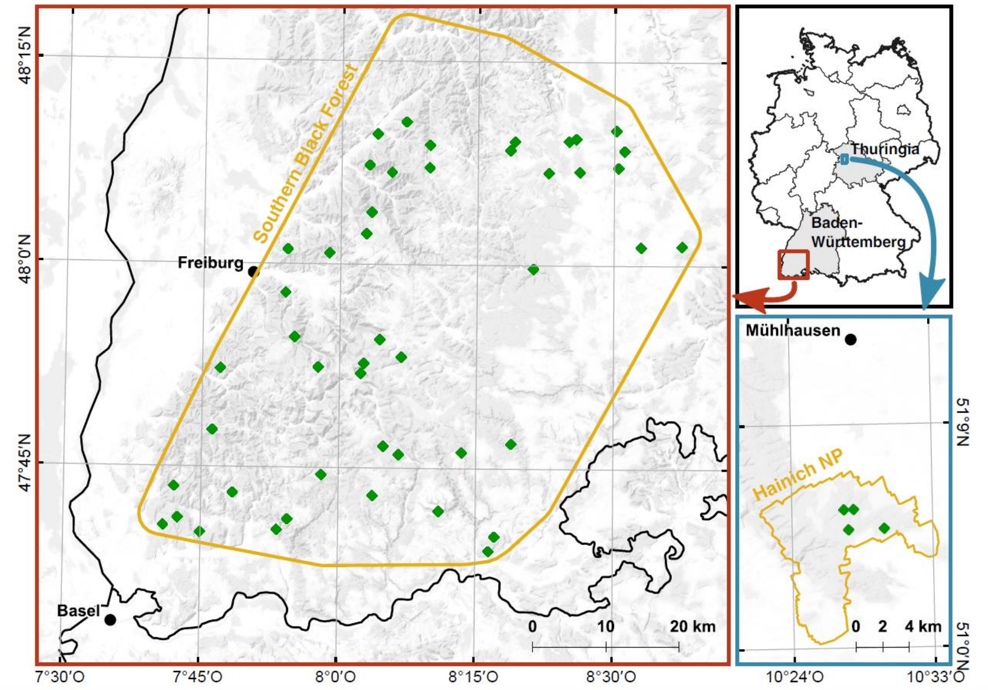
The use of unmanned aerial vehicles (UAVs) in vegetation remote sensing allows a time-flexible and cost-effective acquisition of very high-resolution imagery. Still, current methods for the mapping of forest tree species do not exploit the respective, rich spatial information. Here, we assessed the potential of convolutional neural networks (CNNs) and very high-resolution RGB imagery from UAVs for the mapping of tree species in temperate forests. We used multicopter UAVs to obtain very high-resolution (<2 cm) RGB imagery over 51 ha of temperate forests in the Southern Black Forest region, and the Hainich National Park in Germany. To fully harness the end-to-end learning capabilities of CNNs, we used a semantic segmentation approach (U-net) that concurrently segments and classifies tree species from imagery. With a diverse dataset in terms of study areas, site conditions, illumination properties, and phenology, we accurately mapped nine tree species, three genus-level classes, deadwood, and forest floor (mean F1-score 0.73). A larger tile size during CNN training negatively affected the model accuracies for underrepresented classes. Additional height information from normalized digital surface models slightly increased the model accuracy but increased computational complexity and data requirements. A coarser spatial resolution substantially reduced the model accuracy (mean F1-score of 0.26 at 32 cm resolution). Our results highlight the key role that UAVs can play in the mapping of forest tree species, given that air- and spaceborne remote sensing currently does not provide comparable spatial resolutions. The end-to-end learning capability of CNNs makes extensive preprocessing partly obsolete. The use of large and diverse datasets facilitate a high degree of generalization of the CNN, thus fostering transferability. The synergy of high-resolution UAV imagery and CNN provide a fast and flexible yet accurate means of mapping forest tree species.

- Previous studies using CNN detected **tree individuals** in **relatively simple environments** (plantations, urban areas)
- Studies targeting **forest tree species** relied on **more sophisticated sensors (hyperspectral / LiDAR), intensive preprocessing, or few species**
- Consumer-grade UAVs enable easy and low-cost acquisition of very high-resolution RGB data

→ **Research question**: Is RGB imagery sufficient to accurately map tree species in heterogeneous forests?

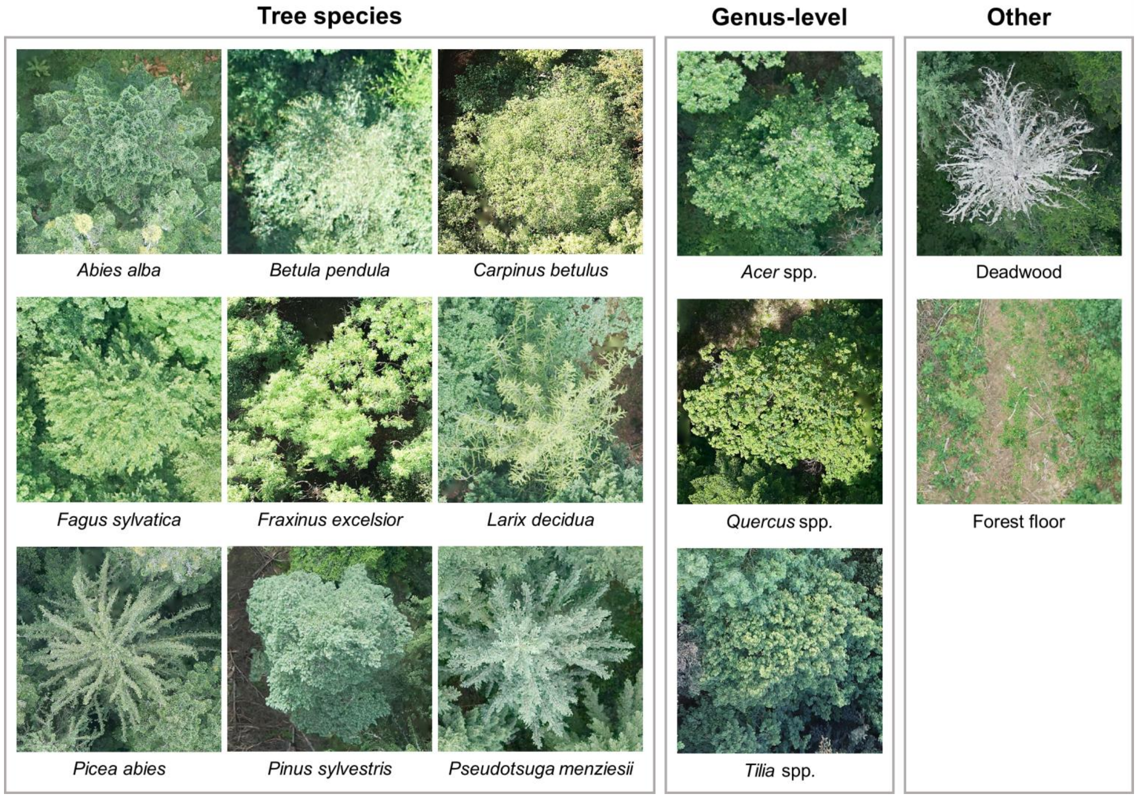
Study area

Southern Black Forest:
- 47 1ha plots within ConFoBi project
- In a mountain range between 120 and 1,492 m a.s.l.
- Mixed and coniferous forests
- Full forest inventory (species, DBH)



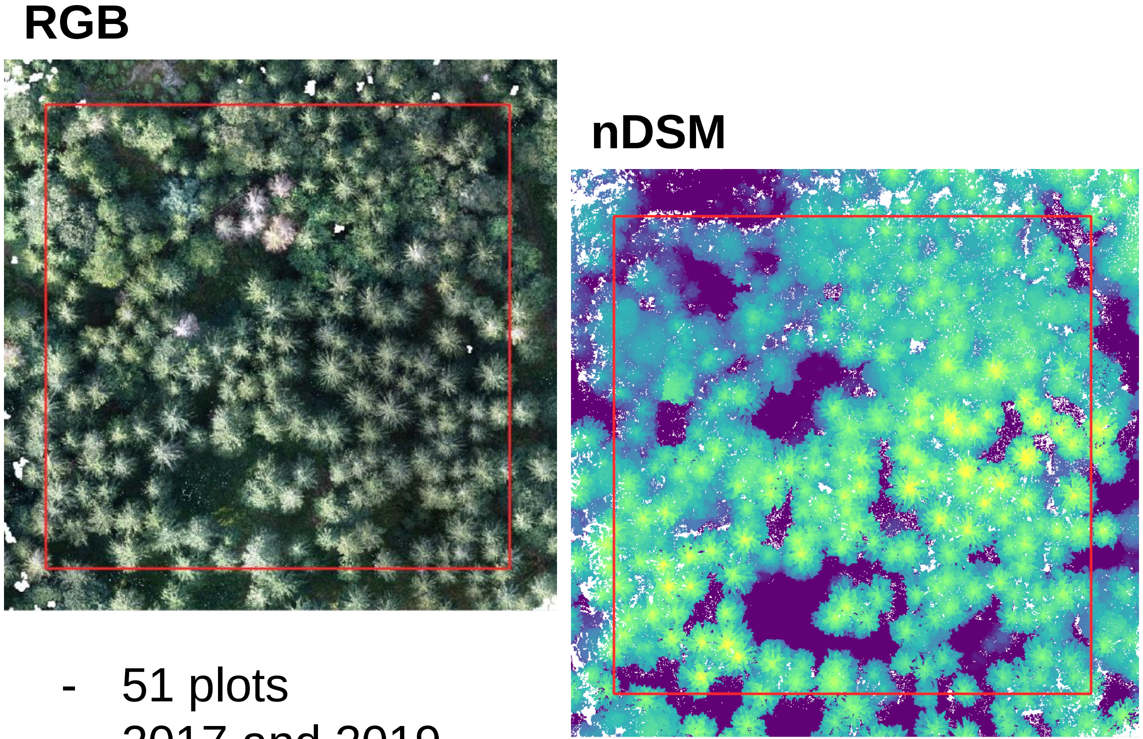
Hainich National Park:
- 4 1ha plots within Biodiversity Exploratories
- NP on a ridge between 225 and 494 m a.s.l.
- Unmanaged mixed deciduous forests
- Full forest inventory (species, DBH, height, stem position)

Tree species



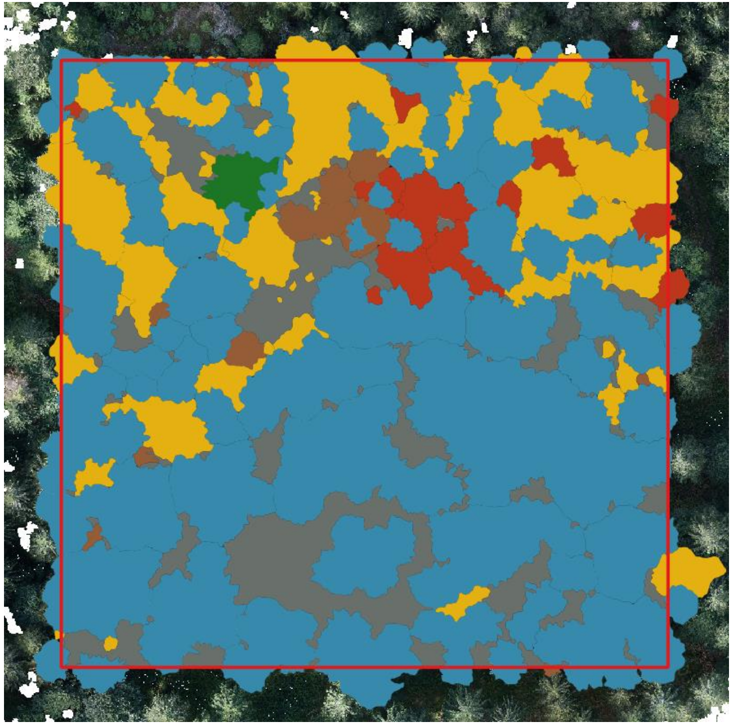
	Area-related share of the class in the dataset [%]	Occurrence of class in number of sites
Picea abies	32,97	45
Fagus sylvatica	29,80	46
Abies alba	10,91	37
Pseudotsuga menziesii	3,89	12
Pinus sylvestris	3,59	19
Acer spp.	2,33	23
Fraxinus excelsior	1,01	14
Larix decidua	0,98	19
Quercus spp.	0,88	10
Carpinus betulus	0,39	4
Tilia spp.	0,24	4
Betula pendula	0,20	8
Forest floor	11,79	50
Deadwood	0,95	44

UAV data



- 51 plots
- 2017 and 2019

Visual interpretation

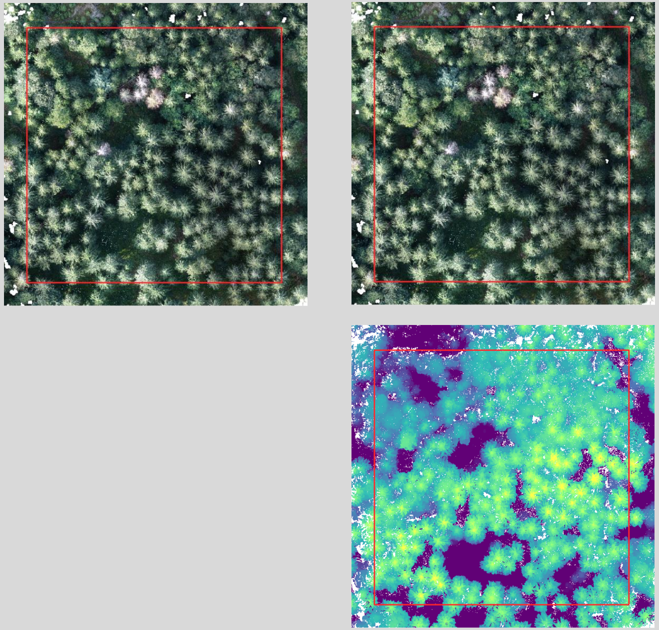


- Acquisition of in-situ data costly, time- and labor-intensive
- Not subject to geolocation errors of GNSS-measurements (especially under dense canopies)
- Spatially explicit link from in-situ data with targeted variable difficult (e.g., tree stems and crowns)

Research questions

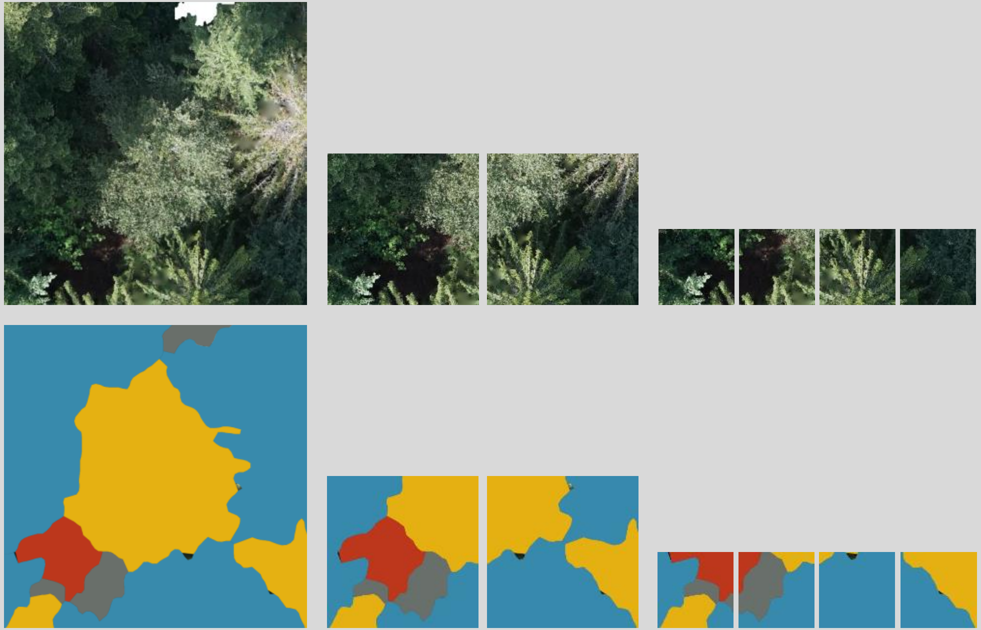
Additional height information

RGB | RGB + nDSM



Different tile sizes

512 px | 256 px | 128 px



Research questions

Different spatial resolutions

2 cm

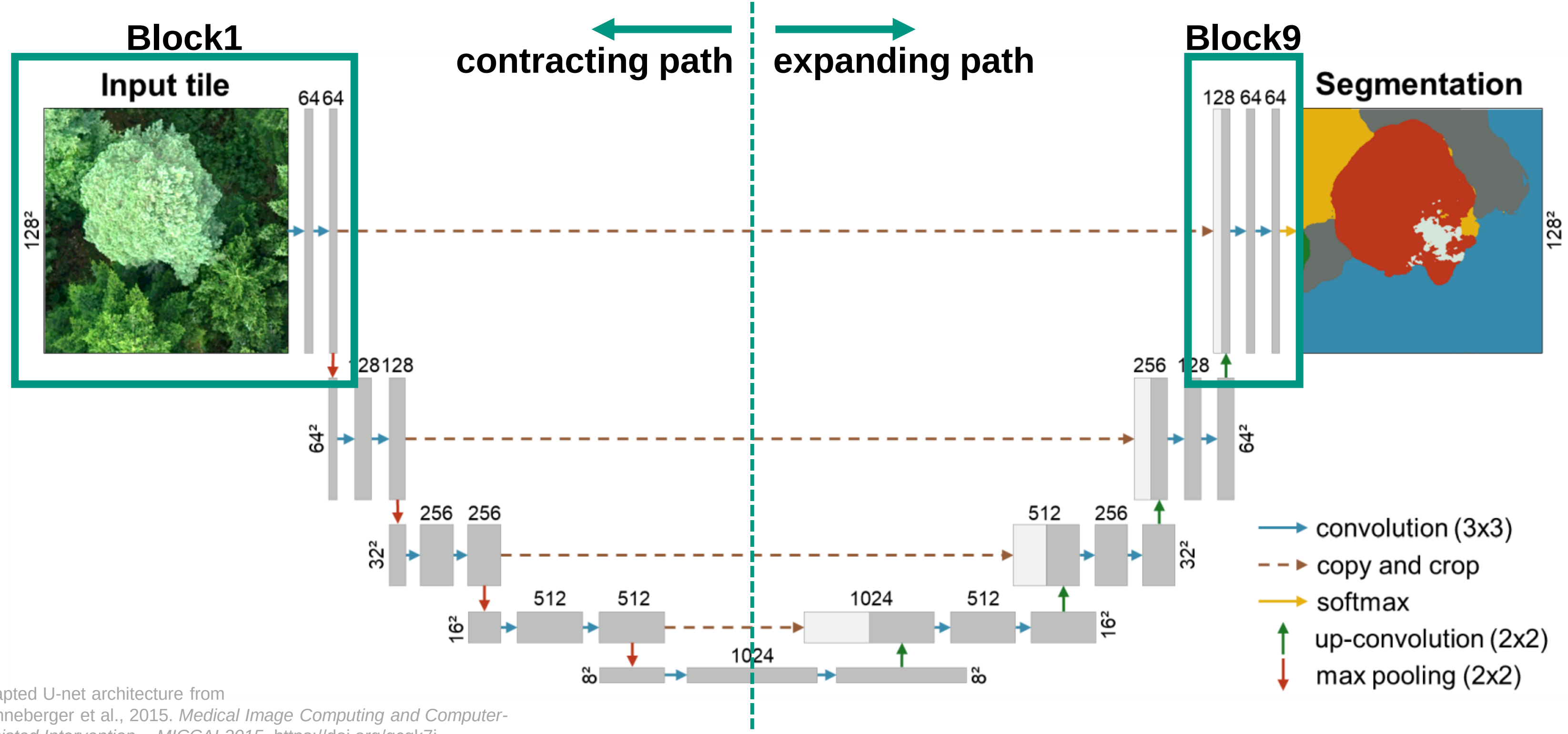
4 cm

8 cm

16 cm

32 cm

CNN-architecture: U-Net



Adapted U-net architecture from Ronneberger et al., 2015. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015.* https://doi.org/gcgk7j

Implementation

```r
# block 1 - contracting path
down1 <- inputs %>%
  layer_conv_2d(filters = 64, kernel_size = c(3, 3), padding = "same") %>%
  layer_batch_normalization() %>%
  layer_activation("relu") %>%
  layer_conv_2d(filters = 64, kernel_size = c(3, 3), padding = "same") %>%
  layer_batch_normalization() %>%
  layer_activation("relu")
down1_pool <- down1 %>%
  layer_max_pooling_2d(pool_size = c(2, 2), strides = c(2, 2))
```

Backend

TensorFlow

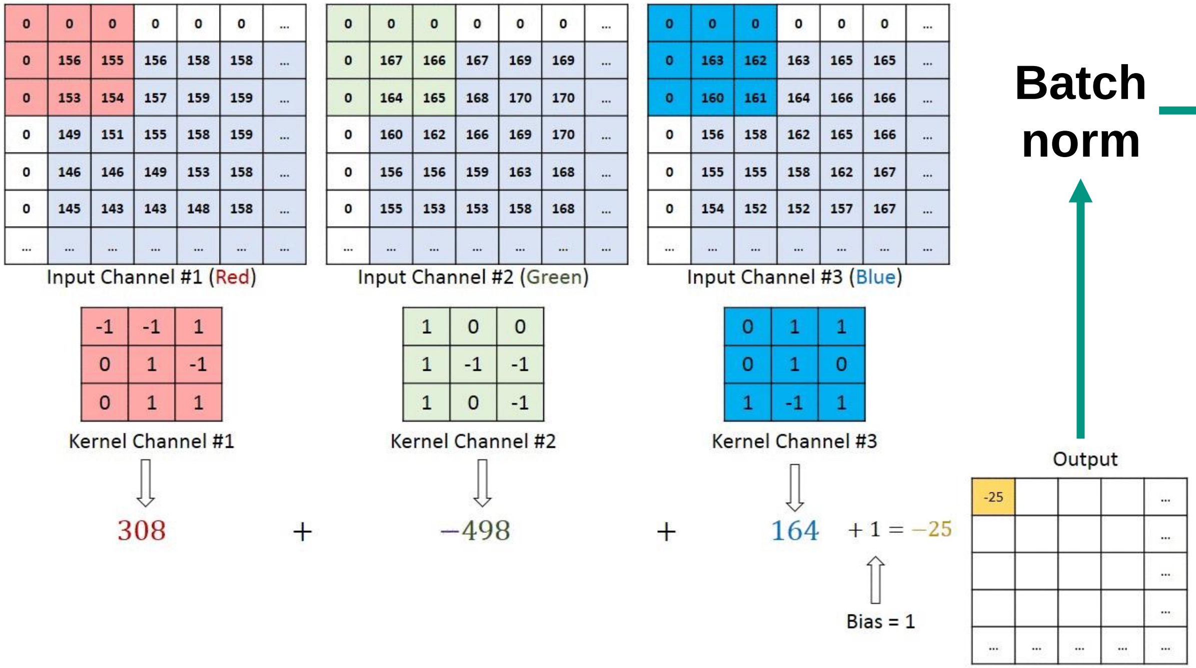
API

Keras

```r
# block 9 - expanding path
up1 <- up2 %>%
  layer_upsampling_2d(size = c(2, 2)) %>%
  {layer_concatenate(inputs = list(down1, .), axis = 3)} %>%
  layer_conv_2d(filters = 64, kernel_size = c(3, 3), padding = "same") %>%
  layer_batch_normalization() %>%
  layer_activation("relu") %>%
  layer_conv_2d(filters = 64, kernel_size = c(3, 3), padding = "same") %>%
  layer_batch_normalization() %>%
  layer_activation("relu") %>%
  layer_conv_2d(filters = 64, kernel_size = c(3, 3), padding = "same") %>%
  layer_batch_normalization() %>%
  layer_activation("relu")
```

Operating principle

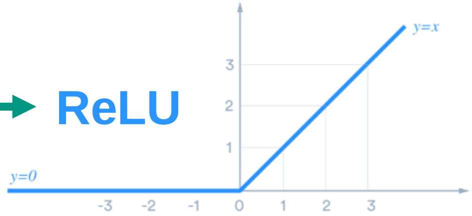
Convolutional filter



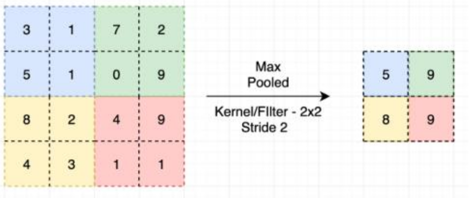
Batch norm

Activation

ReLU



Max-pooling



https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

Model training

1. Draw batch of samples x and corresponding targets y
2. Run CNN on x to obtain y_pred
3. Compute mismatch between y_pred and y → **“loss”**
4. Compute gradient of the loss
5. Adjust parameters in opposite direction from gradient

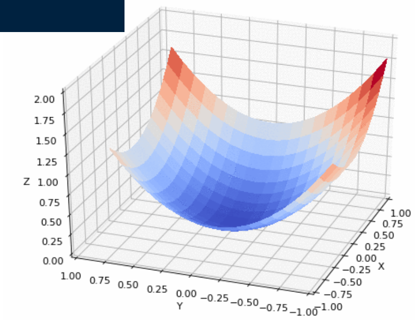
   → **“gradient descent”**

```r
# output layer
classify <- layer_conv_2d(up1,
                          filters = numClasses,
                          kernel_size = c(1, 1),
                          activation = "softmax")

# build model
model <- keras_model(
  inputs = inputs,
  outputs = classify
)

# compile model
model %>% compile(
  optimizer = tf$keras$optimizers$RMSprop(0.0001),
  loss      = weightedCategoricalCrossentropy,
  metrics   = c("accuracy", "categorical_crossentropy")
)
```

```
================================
Total params: 34,541,582
Trainable params: 34,527,886
Non-trainable params: 13,696
________________________________
```



Data splitting + Accuracy assessment

Training
75%

Validation
25%

Test
10%
+1 plot

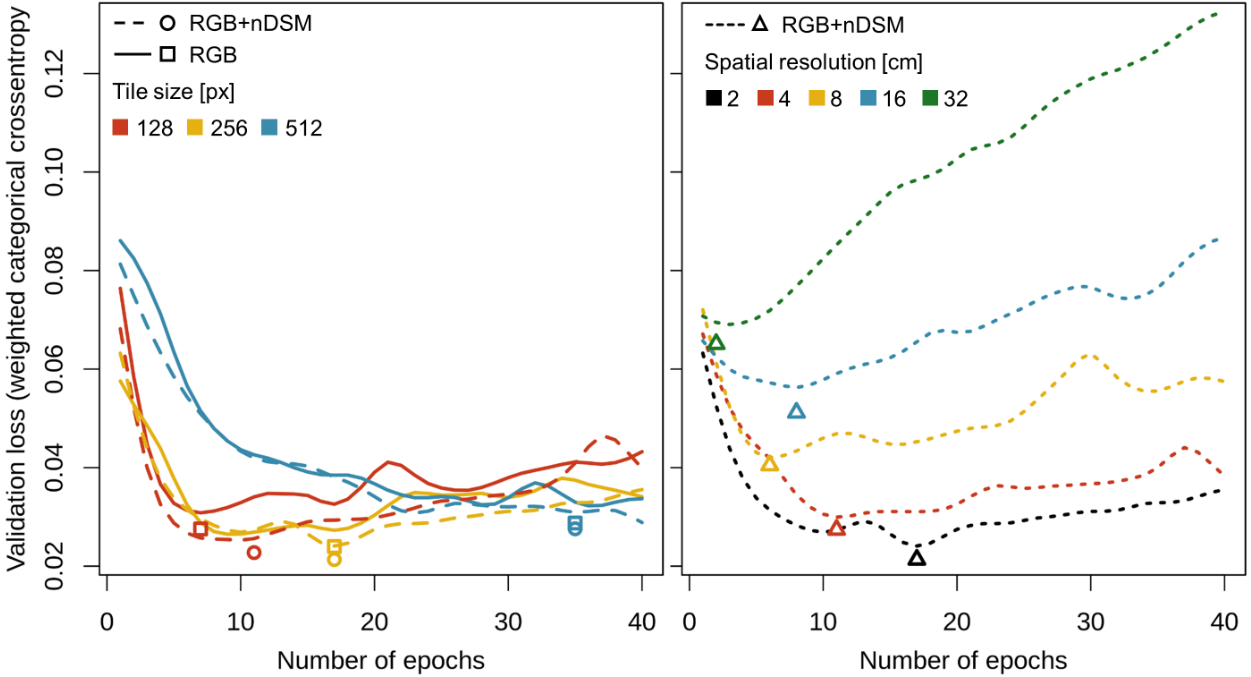
$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

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

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

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Results



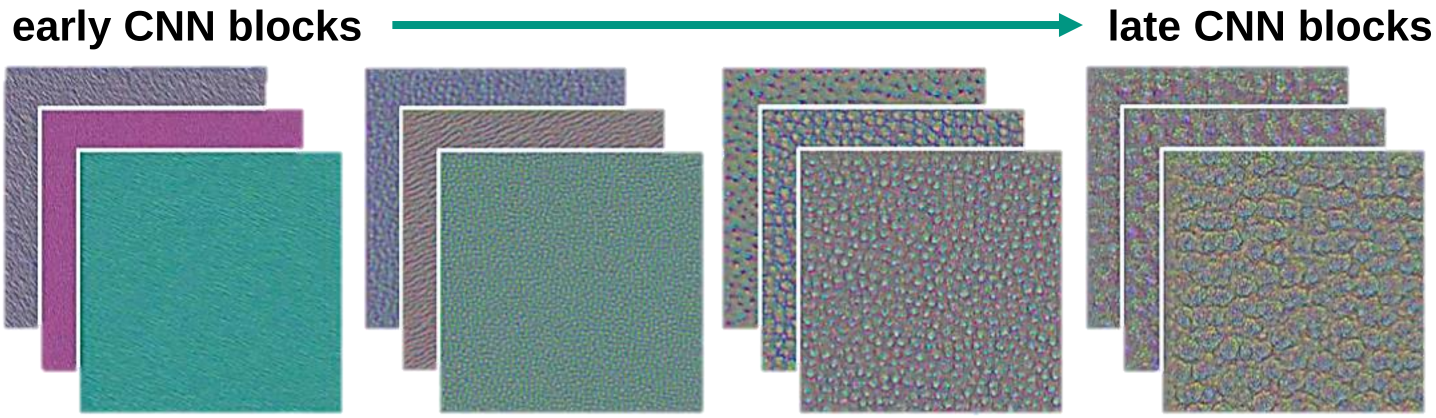
RGB+nDSM
RGB
Tile size [px]
128
256
512
Validation loss (weighted categorical crossentropy)
0.02
0.04
0.06
0.08
0.10
0.12
0
10
20
30
40
Number of epochs
RGB+nDSM
Spatial resolution [cm]
2
4
8
16
32
0
10
20
30
40
Number of epochs

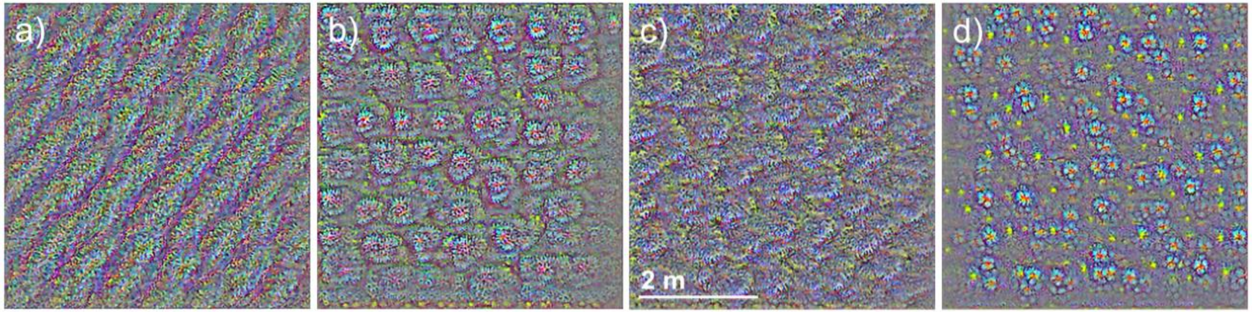
Results

	Tile size [pixel]						Spatial resolution [cm]					Area-related share[a]
Input data	RGB			RGB + nDSM			RGB + nDSM					
Tile size/resolution	128	256	512	128	256	512	2	4	8	16	32	
F1-Score												
Picea abies	0.89	0.93	0.91	0.93	**0.93**	0.93	0.93	0.91	0.86	0.81	0.70	32.97
Fagus sylvatica	0.89	0.90	0.87	**0.90**	0.90	0.86	0.90	0.86	0.79	0.75	0.66	29.80
Abies alba	0.79	0.85	0.86	0.86	**0.87**	0.86	0.87	0.83	0.60	0.60	0.34	10.91
Pseudotsuga menziesii	0.84	0.89	0.74	0.89	**0.91**	0.88	0.91	0.86	0.79	0.77	0.36	3.89
Pinus sylvestris	0.89	0.90	0.89	**0.91**	0.91	0.87	0.91	0.81	0.78	0.60	0.24	3.59
Acer spp.	0.70	0.72	0.53	**0.80**	0.73	0.40	0.73	0.60	0.40	0.37	0.12	2.33
Fraxinus excelsior	0.75	0.79	0.16	**0.87**	0.82	0.52	0.82	0.59	0.28	0.15	–	1.01
Larix decidua	0.80	0.82	0.80	0.83	**0.89**	0.82	0.89	0.65	0.21	0.17	–	0.98
Quercus spp.	**0.64**	0.49	0.28	0.58	0.39	0.02	0.39	0.38	0.00	–	–	0.88
Carpinus betulus	**0.45**	0.33	–	0.38	0.36	0.00	0.36	0.24	0.08	0.06	–	0.39
Tilia spp.	0.26	0.20	–	**0.50**	0.02	–	0.02	0.01	–	–	–	0.24
Betula pendula	0.07	**0.33**	–	0.27	–	–	–	–	–	–	–	0.20
Forest floor	0.78	0.83	0.82	0.83	**0.84**	0.84	0.84	0.82	0.80	0.77	0.72	11.79
Deadwood	0.71	0.73	0.68	**0.72**	0.75	0.69	0.75	0.70	0.53	0.57	0.44	0.95
Mean F1-Score	0.68	0.69	0.54	**0.73**	0.67	0.55	0.67	0.59	0.44	0.40	0.26	
Overall Accuracy	0.86	0.88	0.86	**0.89**	0.89	0.87	0.89	0.85	0.78	0.73	0.62	

[a] Area-related share of the class in the dataset [%].
[b] Occurrence of class in number of sites.

Filter visualizations





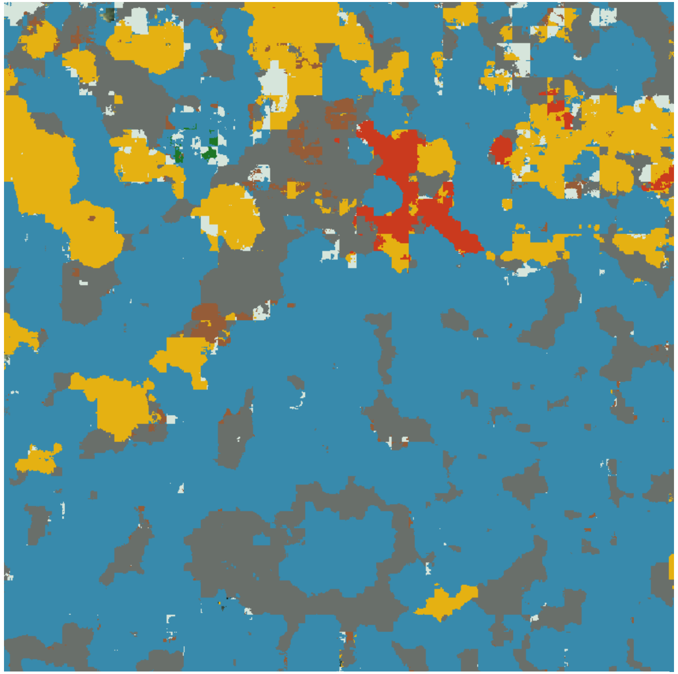
Kattenborn et al., 2021. *ISPRS Journal of Photogrammetry and Remote Sensing.* https://doi.org/ghtrws

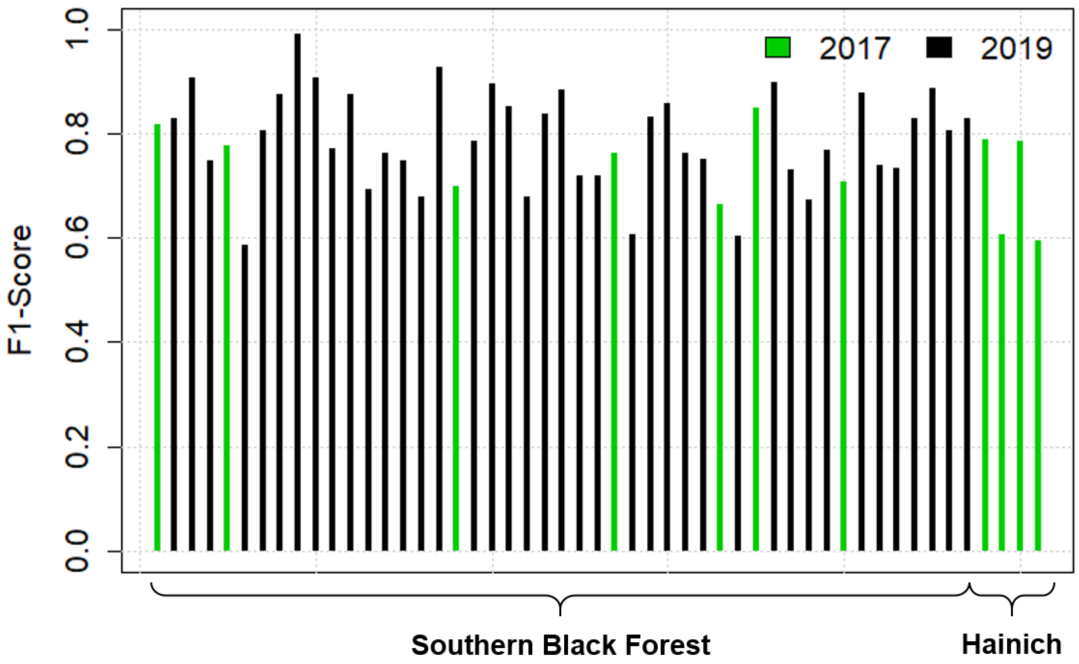
Results

RGB

Reference data

Prediction

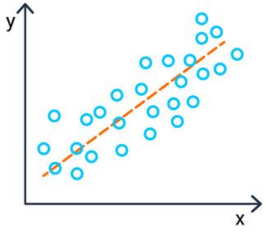




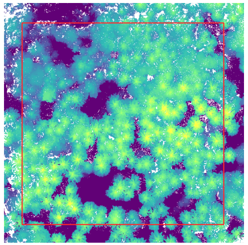
2017
2019
F1-Score
1.0
0.8
0.6
0.4
0.2
0.0
Southern Black Forest
Hainich

Key findings

Model performance



- Comparable with literature, but instead of hyperspectral or LiDAR data only RGB-imagery
- No feature engineering and no tree segmentation or localization steps prior to model inference required → **end-to-end learning**



Additional height information

- No consistent positive effect → *other studies even report negative effects*
- Additional computational cost
- Terrain model required

Key findings

Tile size

- With smaller tiles rare species larger in tile, thus contribute more; with large tiles lost in surrounding species
- Except for classes that feature distinct characteristics (*L. decidua* and deadwood)
- CNN suffer from edge effects → If enough reference data larger tile size preferred

Spatial resolution

- Very-high spatial resolution essential → key role for UAVs in remote sensing-based forest assessment
- Smaller share decreased accuracies (even with species weighted loss function)
- Except deadwood (prominent features)

Future directions

- **Transfer learning**: already trained CNN can be updated and refined for specific use-case
- **Universal models**: one model trained over a variety of landscapes and many species
- **Large synergies** of UAV (high-resolution) + CNN (harness spatial detail) → implications for ecological research

- **Next steps:**
  - use CNN predictions as reference for satellite-scale models
  - Alternative applications, e.g. biomass