Deep Learning Lab

Segmentation of ISPRS Vaihingen dataset including DSM data

Yu-Chuan Cheng, 20.Sep.2022
yu-chuan.cheng@tu-braunschweig.de

Outline

Outline

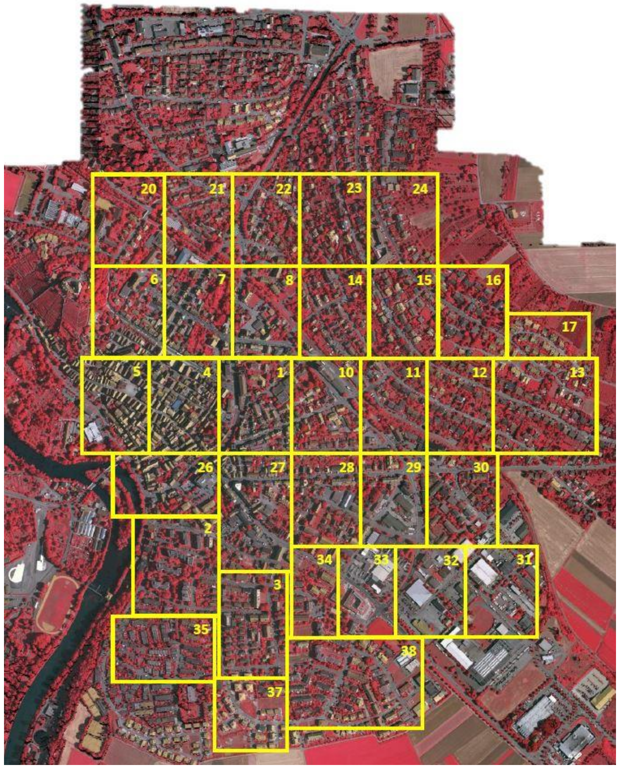
- Introduction
- Dataset
- Experiment Setup
- Evaluation and Result
- Summary

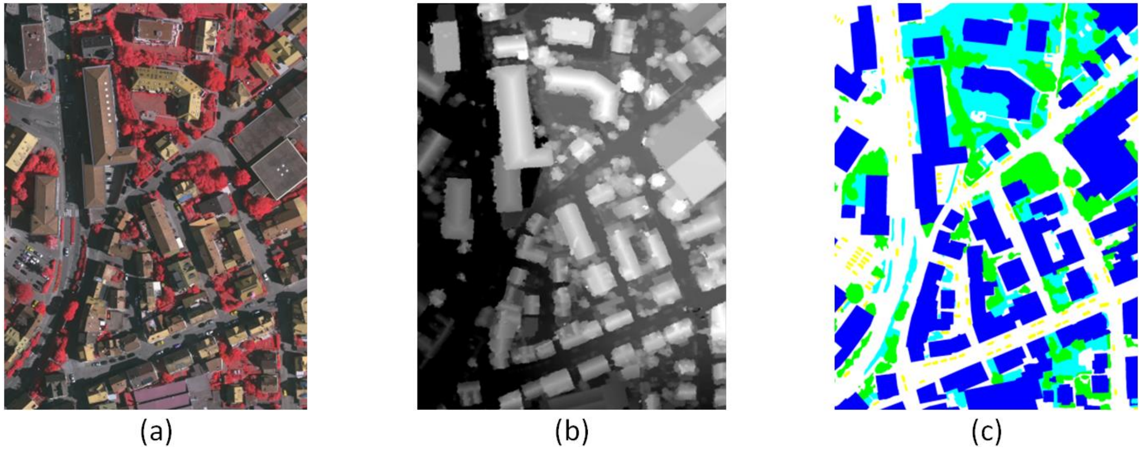
Dataset

128 Training datasets, 38 Validation datasets, 190 Testing datasets

Image size: (2767, 2428, 3)

Patch size: 256





(a) True orthophoto (b) DSM (c) ground truth

Source: 2D Semantic Label. - Vaihingen (isprs.org)

Experiment Setup
Model

U-net

- Downsampling
- Upsampling
- Concatenation

DeepLab3+

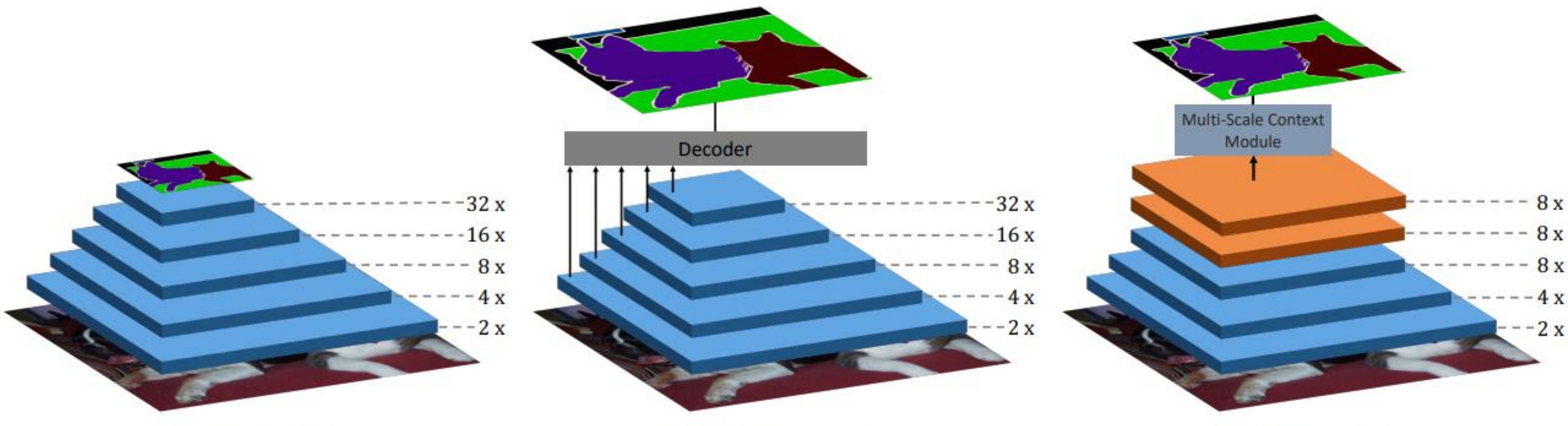
- Atrous Convolution
- Atrous Spatial Pyramid Pooling (ASPP)
- Decoder

Experiment Setup
Model

Original Fully Convolutional networks: lose position information

Encoder-Decoder(U-net): lose position information, but better than FCN

DilatedFCN(DeepLab3+): trade off computing power with strong samentic segmentation

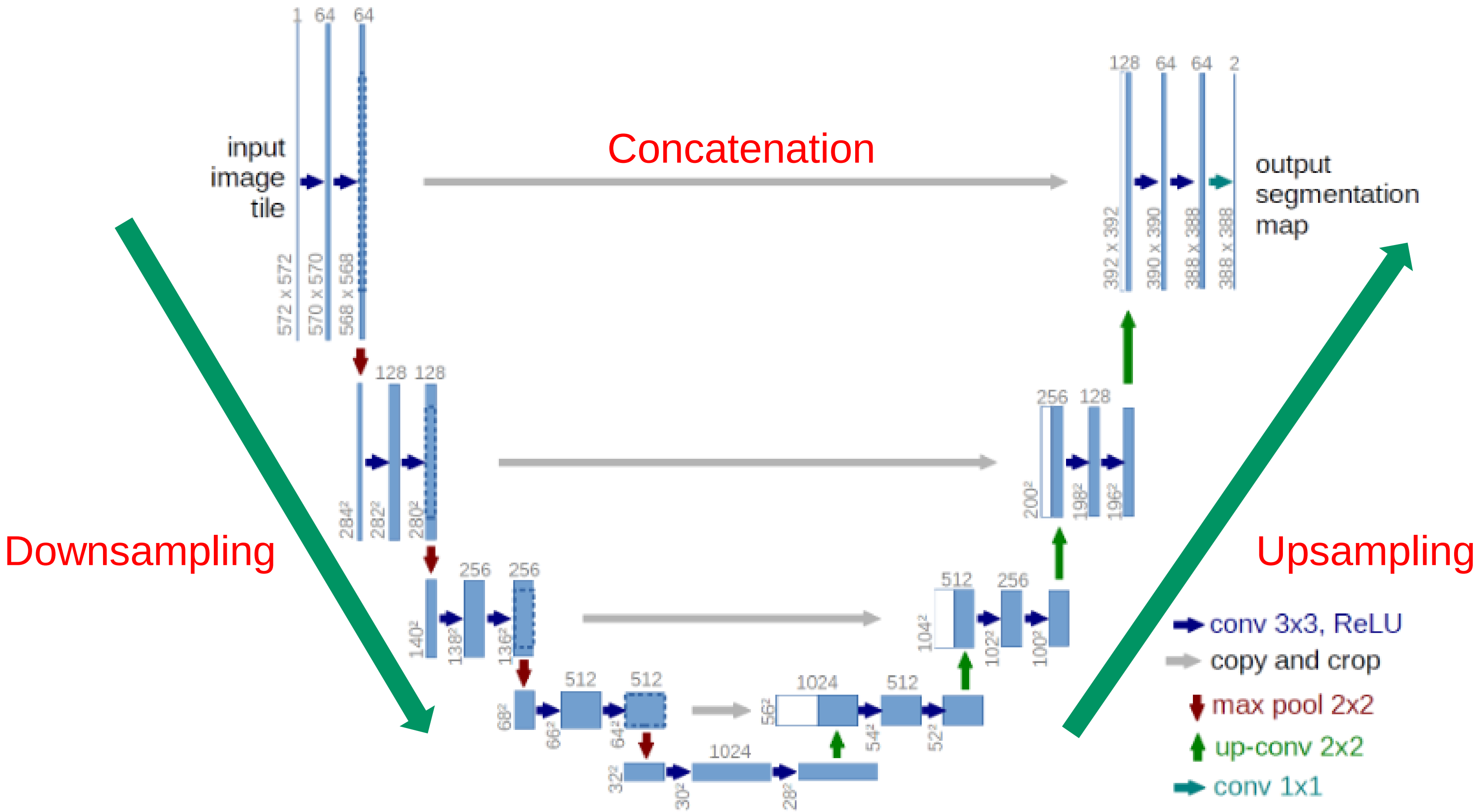


(a) FCN [22] (b) EncoderDecoder (c) DilatedFCN

Different types of networks for semantic segmentation.

Huikai Wu, Junge Zhang, Kaiqi Huang (2019). FastFCN: Rethinking Dilated Convolution in the Backbone for Semantic Segmentation

Model-Unet



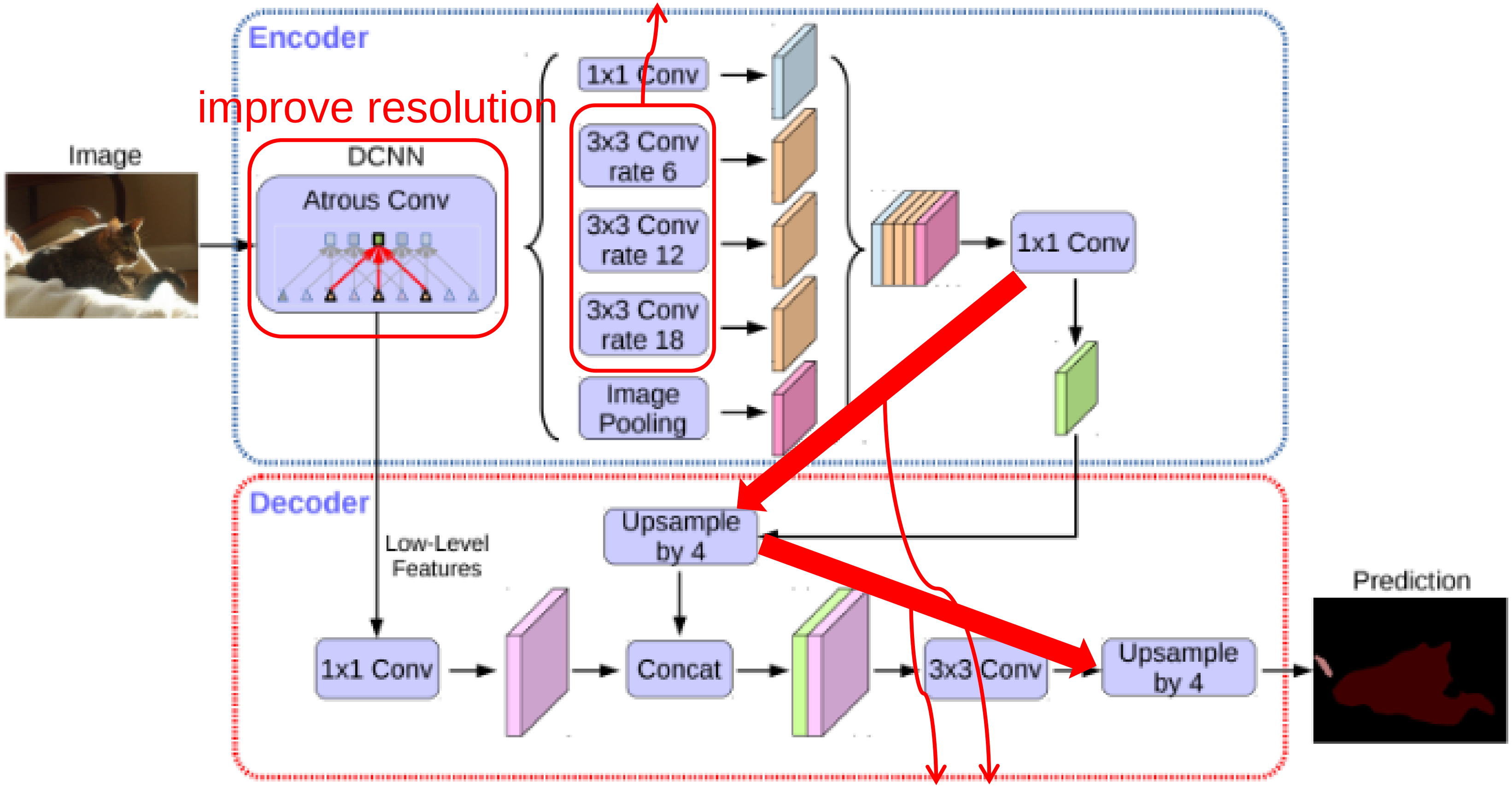
U-net structure

Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

Experiment Setup
Model-Deeplab3+

Be mroe Robust

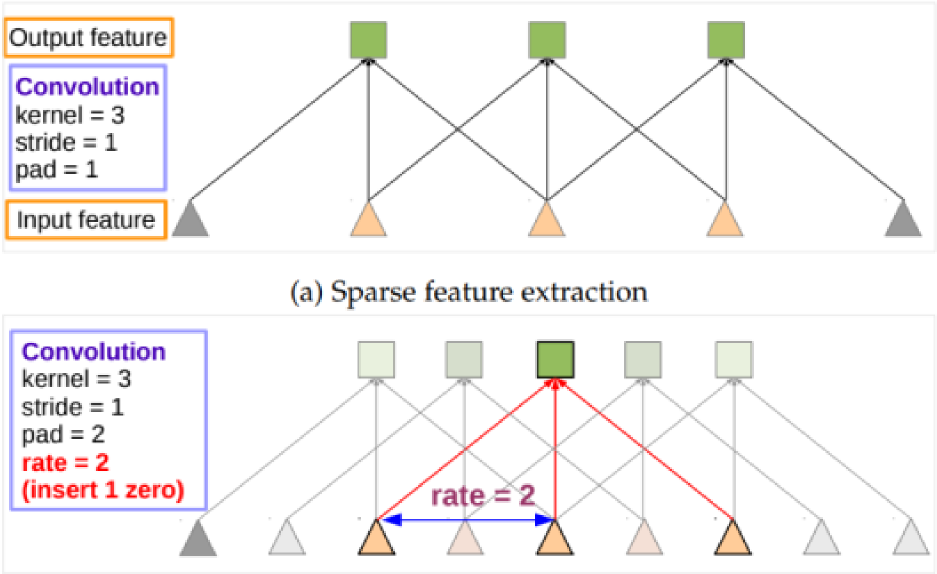
improve resolution



improve results along object boundaries

Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European conference on computer vision (ECCV) (pp. 801-818)..

Experiment Setup
Model-Deeplab3+-DCNN



(a) Sparse feature extraction

(b) Dense feature extraction

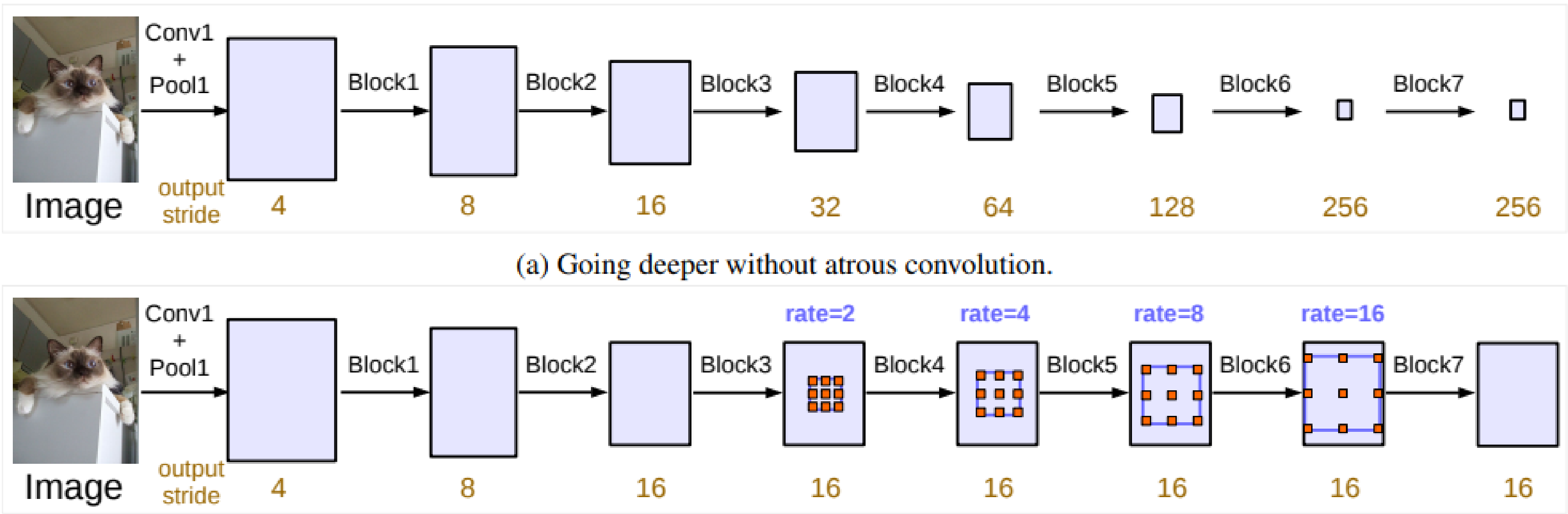
Fig. 2: Illustration of atrous convolution in 1-D. (a) Sparse feature extraction with standard convolution on a low resolution input feature map. (b) Dense feature extraction with atrous convolution with rate $r = 2$, applied on a high resolution input feature map.

Improve feature resolution

Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence, 40(4), 834-848.

Experiment Setup

Model- Deeplab3+-Atrous Convolution



(a) Going deeper without atrous convolution.

(b) Going deeper with atrous convolution. Atrous convolution with $rate > 1$ is applied after block3 when $output\_stride = 16$.

Figure 3. Cascaded modules without and with atrous convolution.

Improve feature resolution

Chen, L. C., Papandreou, G., Schroff, F., & Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation.

Experiment Setup

Model- Deeplab3+-Atrous Spatial Pyramid Pooling (ASPP)

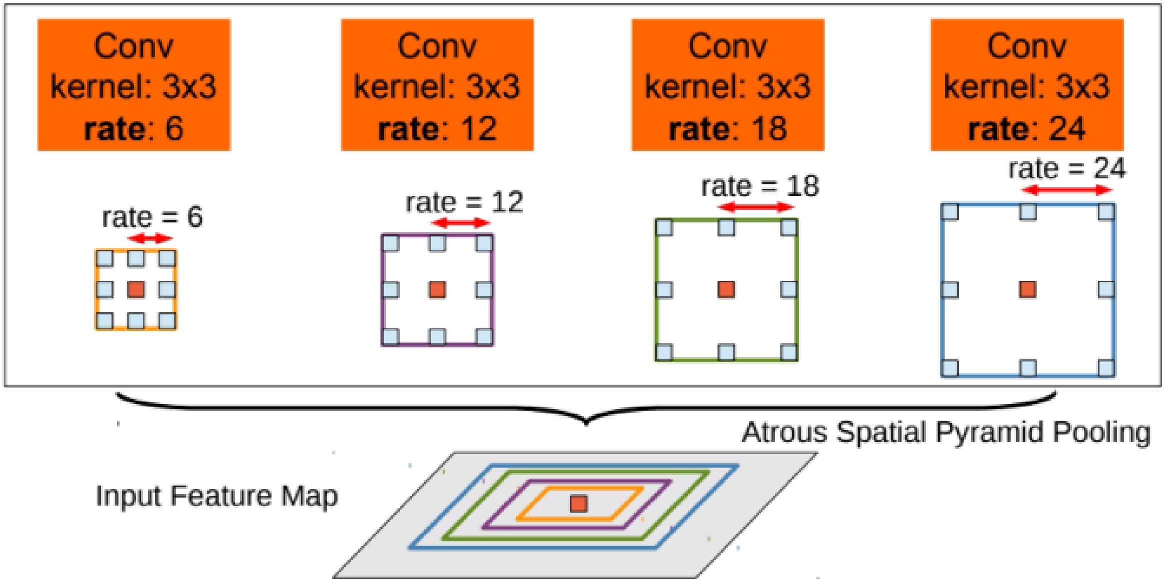


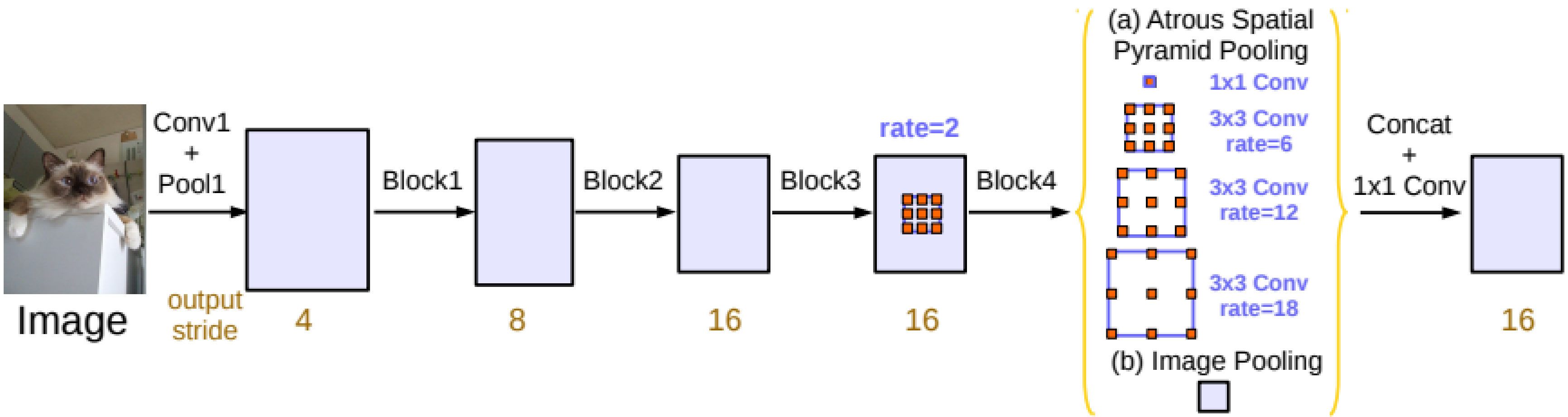
Fig. 4: Atrous Spatial Pyramid Pooling (ASPP). To classify the center pixel (orange), ASPP exploits multi-scale features by employing multiple parallel filters with different rates. The effective Field-Of-Views are shown in different colors.

Robust against objects at multiple scales

Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence, 40(4), 834-848.

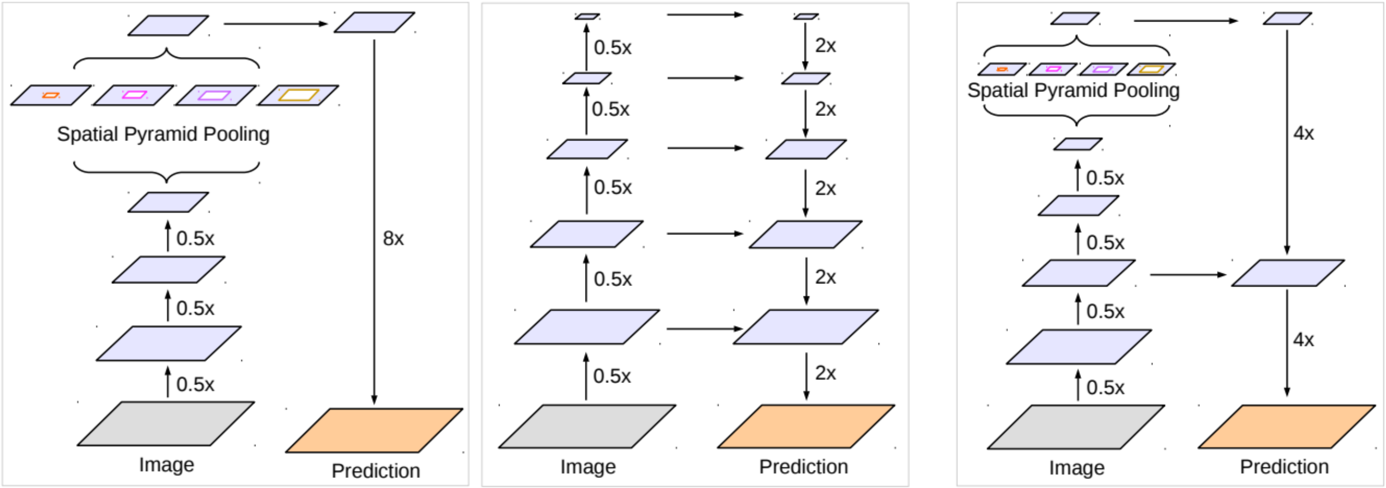
Experiment Setup

Model-Deeplab3+-Atrous Spatial Pyramid Pooling (ASPP)



Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence, 40(4), 834-848.

Experiment Setup
Model- Deeplab3+- Encoder & Decoder



(a) Spatial Pyramid Pooling (b) Encoder-Decoder (c) Encoder-Decoder with Atrous Conv

Fig. 1. We improve DeepLabv3, which employs the spatial pyramid pooling module (a), with the encoder-decoder structure (b). The proposed model, DeepLabv3+, contains rich semantic information from the encoder module, while the detailed object boundaries are recovered by the simple yet effective decoder module. The encoder module allows us to extract features at an arbitrary resolution by applying atrous convolution.

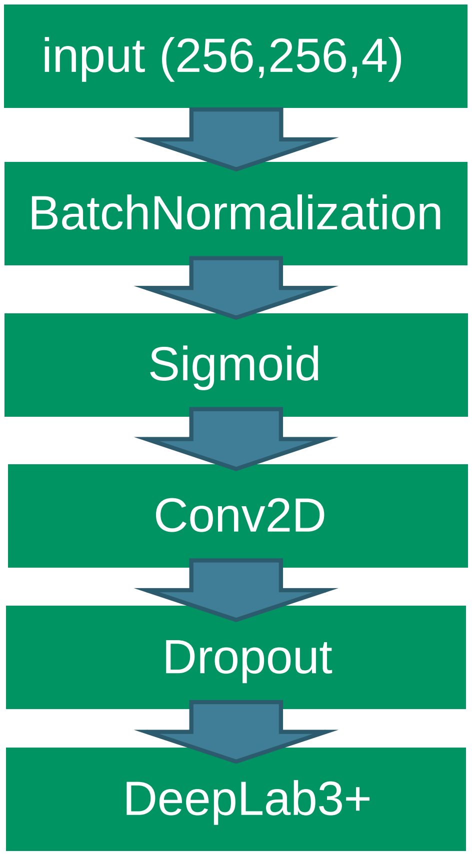
Encoder-Decoder to improve results along object boundaries

Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European conference on computer vision (ECCV) (pp. 801-818)..

Experiment Setup

Model- Deeplab3+ integrate DSM data

Impervious surfaces
Building
Low vegetation
Tree
Car
Clutter/background



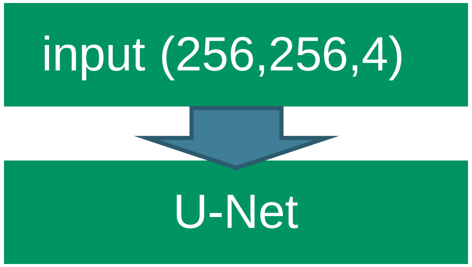
Input image size and preprocess

4 input channels

```
Layer (type)                 Output Shape              Param #
==============================================================
input_2 (InputLayer)         [(None, 256, 256, 4)]     0

batch_normalization_9 (Batc  (None, 256, 256, 4)       16
hNormalization)

tf.math.sigmoid (TFOpLambda  (None, 256, 256, 4)       0
)

conv2d_10 (Conv2D)           (None, 256, 256, 3)       15

dropout (Dropout)            (None, 256, 256, 3)       0

model (Functional)           (None, 256, 256, 6)       11853638

==============================================================
Total params: 11,853,669
Trainable params: 11,820,925
Non-trainable params: 32,744
```

6 output classes

Numbers of DeepLab3+'s parameters



Input image size of U-net

```
===============================
Total params: 8,558,406
Trainable params: 8,558,406
Non-trainable params: 0
```

Numbers of U-net’s Parameters

Experiment Setup
Training protocol

Adam Optimizer with epsilon value: 1e-8
Standard normalization (0 mean, 1 std)
Early stopping patience : 40
Learning rate = 0.001 (by default)
200 Epoch

Data augmentation:
- Horizontal Flip
- Vertical Flip
- Random Rotate 90
- Shift Scale Rotate

Evaluation and Result

U-Net

- with CE loss: 51.17% mIOU without DSM
- with Focal loss: 53.20% mIOU without DSM

DeepLab3+

- with CE loss: **64.31**% mIOU without DSM
- with Focal loss: 63.56% mIOU without DSM

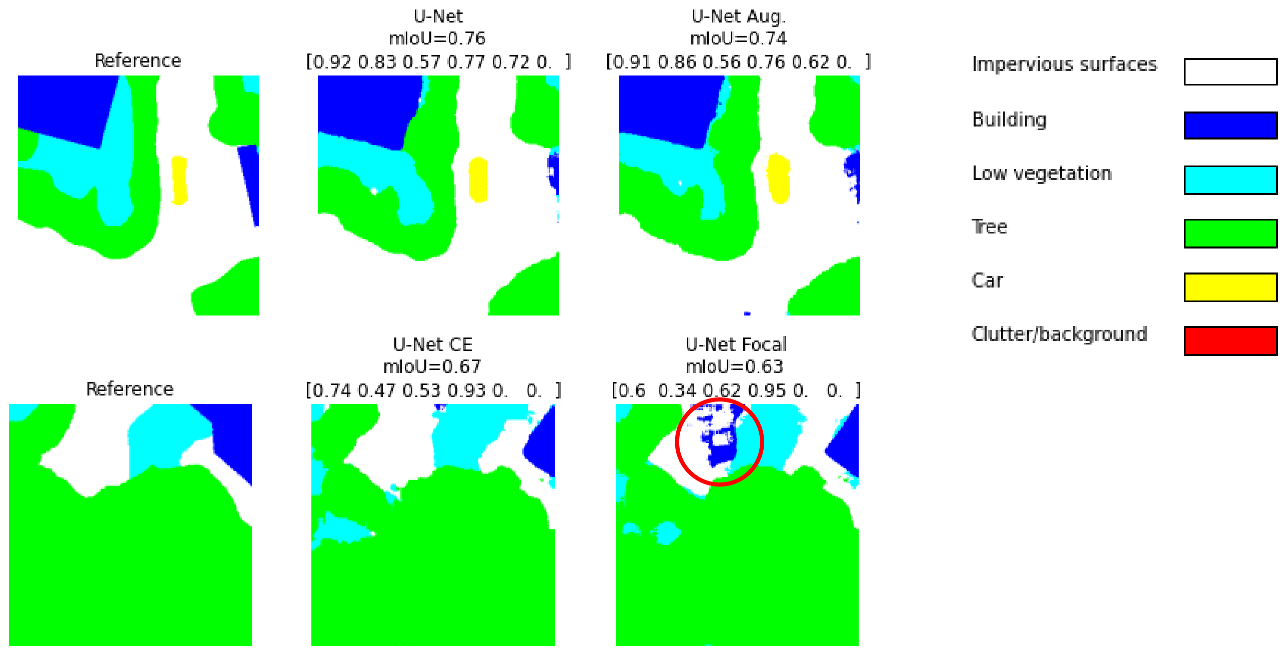
Technische Universität Braunschweig

igp

Evaluation and Result

U-Net

- With CE loss: **55.03**%mIOU 51.17% mIOU without DSM
- With Focal loss: 51.2%mIOU **53.20**% mIOU without DSM

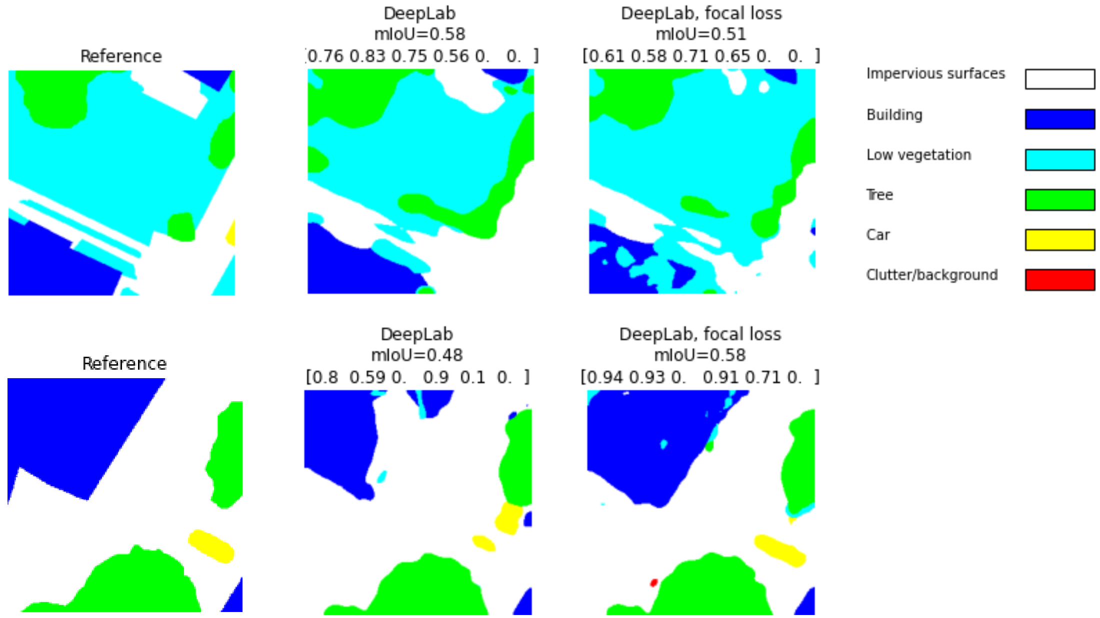


The semantic segmentation result of U-Net

Evaluation and Result

DeepLab3+

- with CE loss: 52.78% mIOU **64.31%** without DSM
- with Focal loss: 58.65% mIOU **63.56%** without DSM



The semantic segmentation result of DeepLab3+

Summary

The main parts of my presentation

Integration the 4th channel

Comparison between Cross Entropy loss (CE) and focal loss

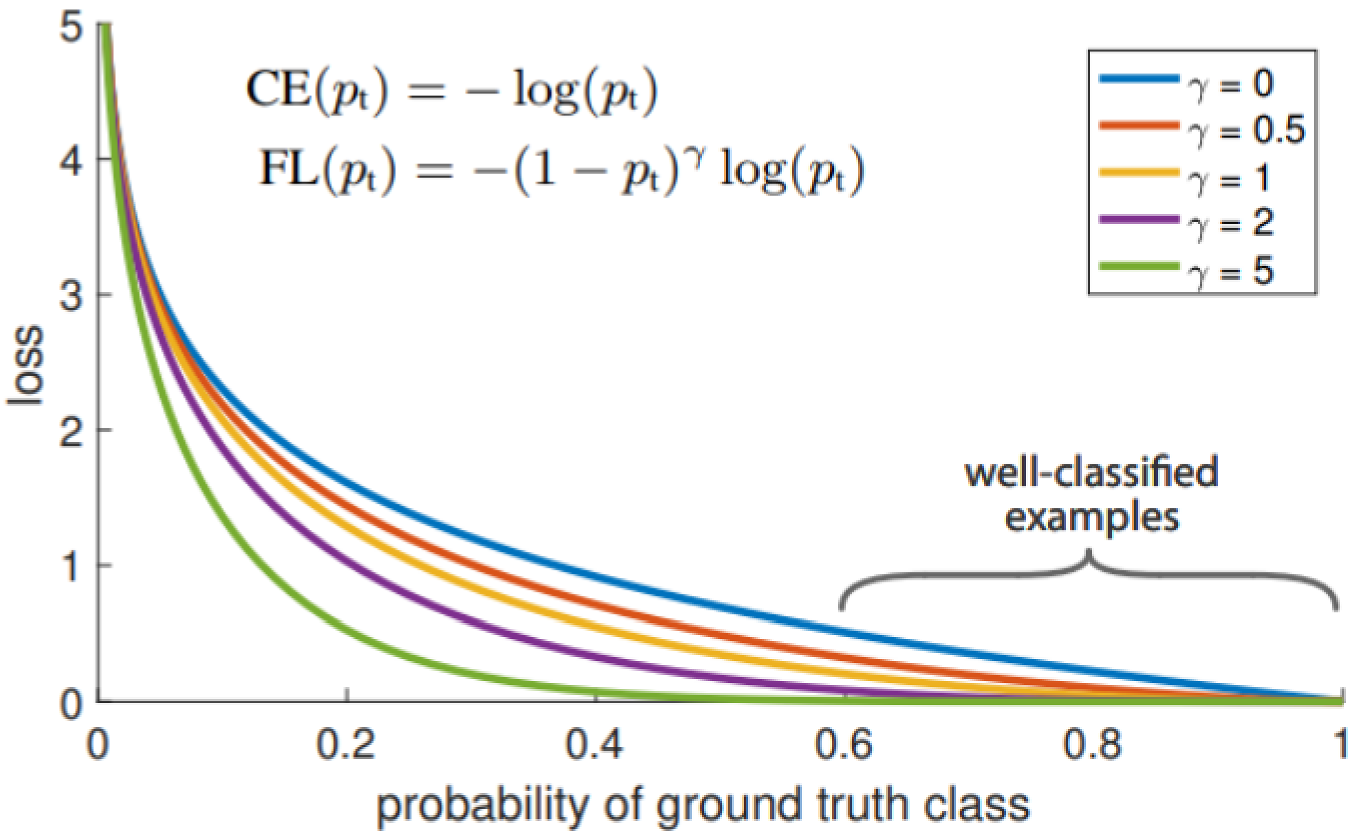
Comparison between U-net and Deeplab3+

- Discussion
  - The best performance is Deeplab3+ with focal loss 58.65% mIOU, but still lower than the one using CE with only RGB channel (64.31%).
  - Focal loss tends to make false semantic segmentations on impervious surfaces.
  - Future work

tuning the gamma value of focal loss

try SGD or RMSProp optimizer

Focal loss



Tsung-Yi Lin, Priyal Goyal, Ross Girshick, et al. „Focal Loss for Dense Object Detection“. In: IEEE Transactions on Pattern Analysis and Machine Intelligence 42 (2020), pages 318–327.