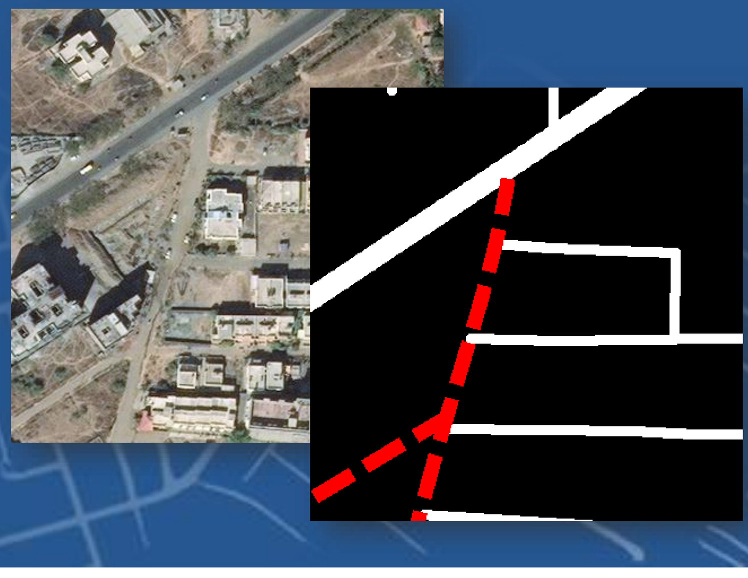
Aerial Road Segmentation in the Presence of Topological Label Noise

ICPR 2020
25th International Conference on Pattern Recognition
Milan, Italy 10 | 15 January 2021

Corentin Henry
German Aerospace Center (DLR)
Oberpfaffenhofen, Germany
corentin.henry@dlr.de

Friedrich Fraundorfer
Institute of Computer Graphics and Vision
Graz University of Technology (TUG)
Graz, Austria

Eleonora Vig
Amazon
Berlin, Germany



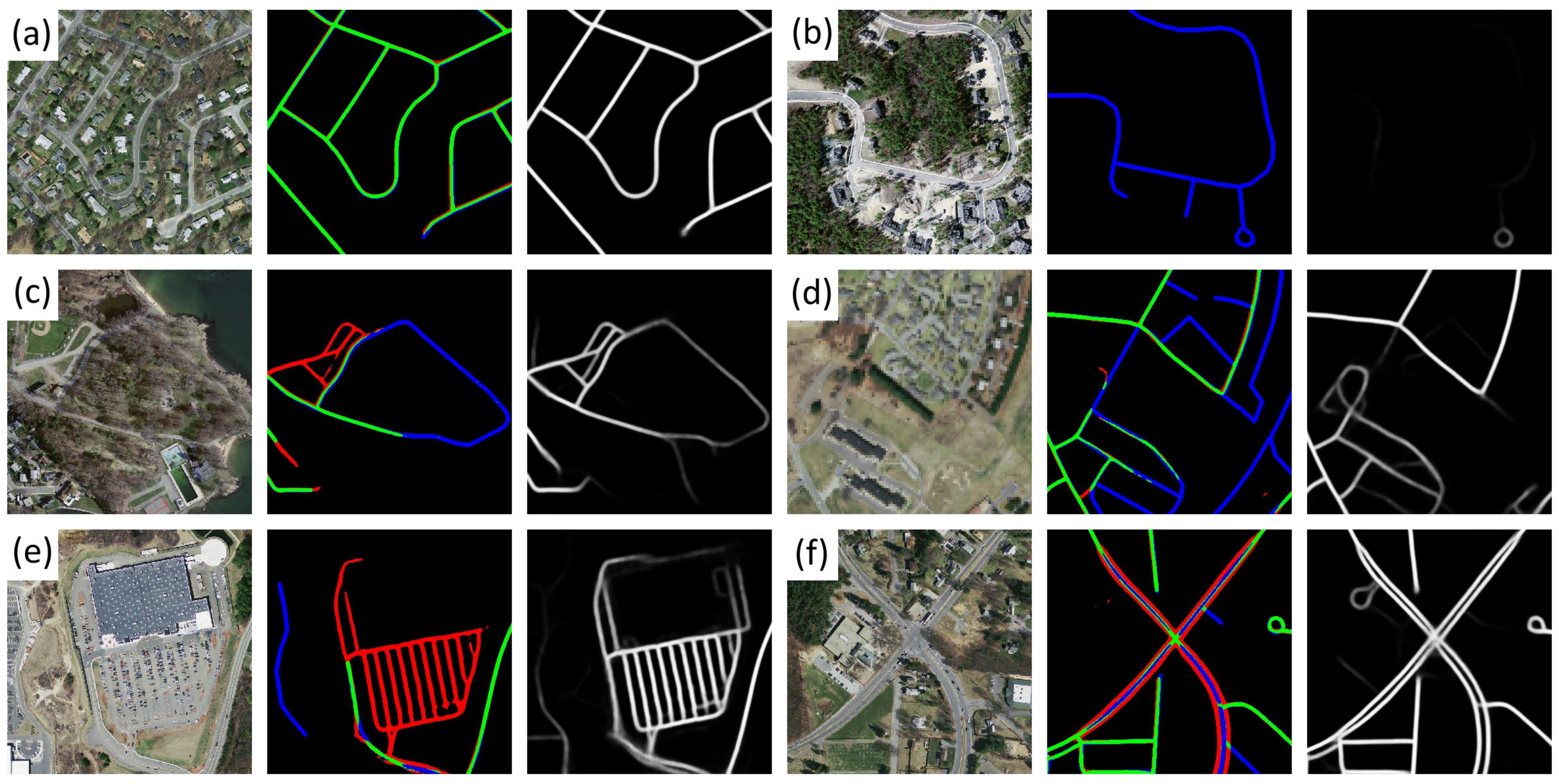
Labeling Roads in Aerial Images

High-resolution aerial images are expensive to annotate
- Large scope or high definition: from 1m/pixel to 5cm/pixel
- Many objects: cars, trees, parking places
- Small details: lane-markings, walls, danger-areas
- Frequent occlusion: vegetation, buildings, shadows

Roads are complex objects
- Types diversity: streets, highways, dirt paths
- Topology complexity: shape, width, connectivity
- Similarity to other objects: parking lots, sidewalks, bikeways

Large-scale datasets trade-off quality for quantity
- With inconsistencies from OpenStreetMap [1]
- With inaccurate polylines [2]
- With incomplete ground truths [3]

Examples of label noise and its effect on predictions
- (a-f) Triplets of RGB, confusion and probability maps of our baseline model
- RGB: 1m/pixel satellite images from [1]
- Confusion Maps: **true positives**, **false negatives**, **false positives**
- Probability Maps: pixel-wise detection of roads in 0-100%



The training labels are not entirely trustworthy

Extracting Roads in Aerial Images

Fully-convolutional neural networks (FCNNs)
- Fine-grained segmentation: need for dedicated architectures
- U-Nets [4] are state-of-the-art
- Shallow architectures used for fast training in challenges
- Deep architectures used to leverage large-scale data

Proposing U-Nets with ResNet [5] and DenseNet [6] backbones

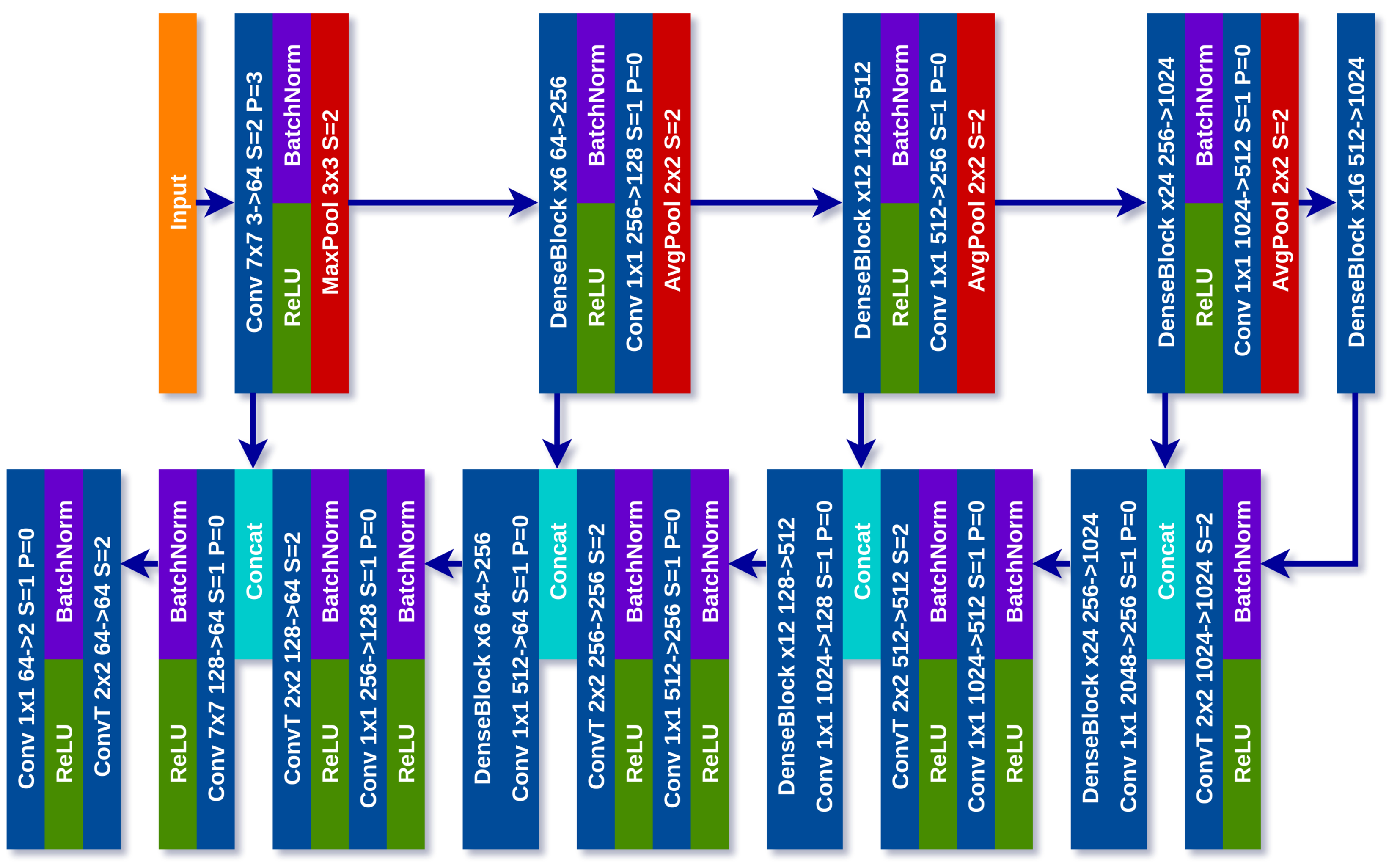


Diagram of the Dense-U-Net-121 architecture

Noise-Aware Loss Functions

Predictions are often more consistent and better localized than labels
- Injecting predictions as ground truth in losses using bootstrapping [7]

$$BCE(y,p) = -\sum_k^C \sum_i^N y_{ik} log(p_{ik})$$

$$SoftBootBCE(y,p) = -\sum_k^C \sum_i^N [\beta y_{ik} + (1-\beta) p_{ik}] log(p_{ik})$$

Reinforcing Noise-Resilience

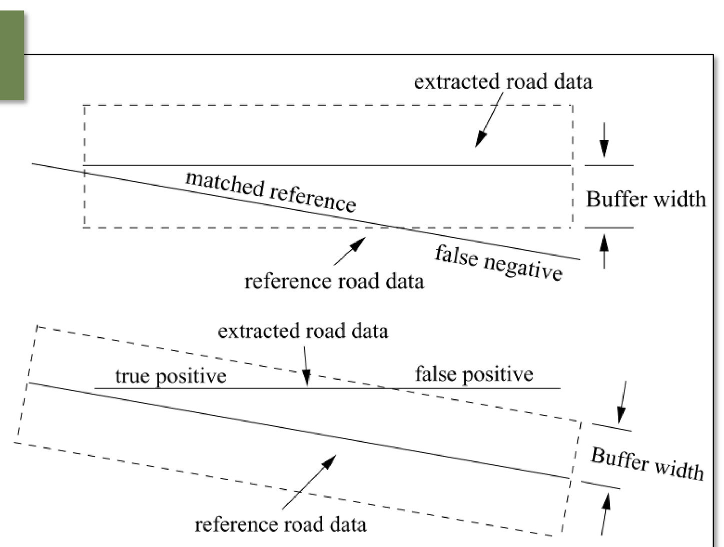
Introduction of synthetic noise as data augmentation during training
- Improving the resistance towards label errors
- Using different noise types with gradual amplitudes:

Annotation *Translated* *Duplicated* *Ablated*

Measuring Road Segmentation Quality

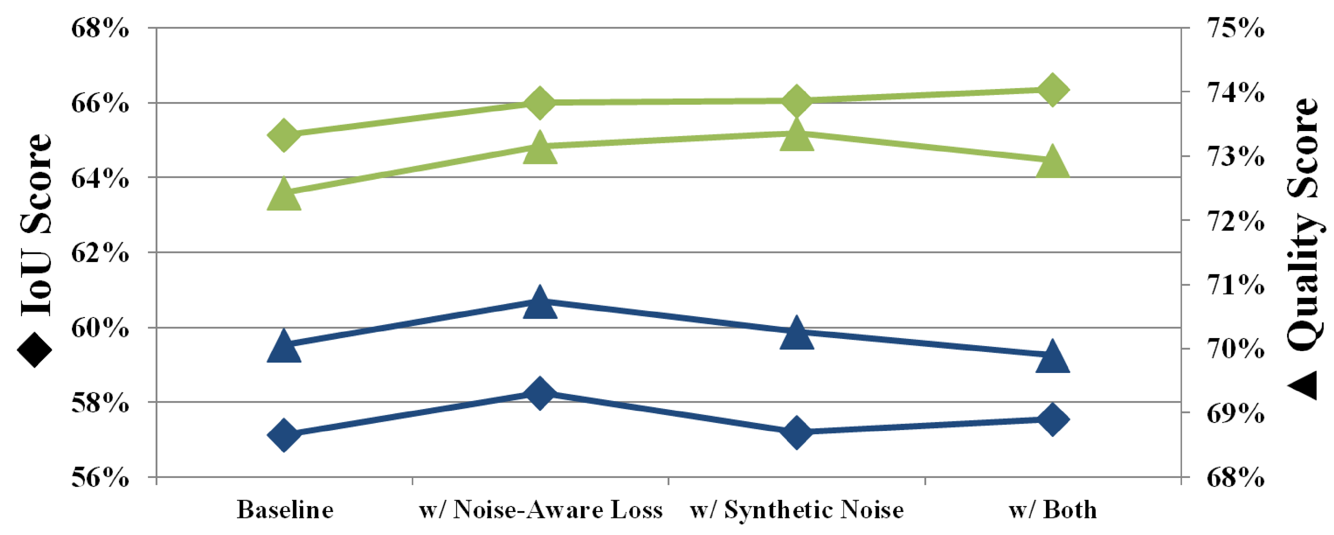
Pixel-wise accuracy
- IoU, F1-DICE, Precision, Recall

Topological accuracy
- Completeness, Correctness, Quality [8]

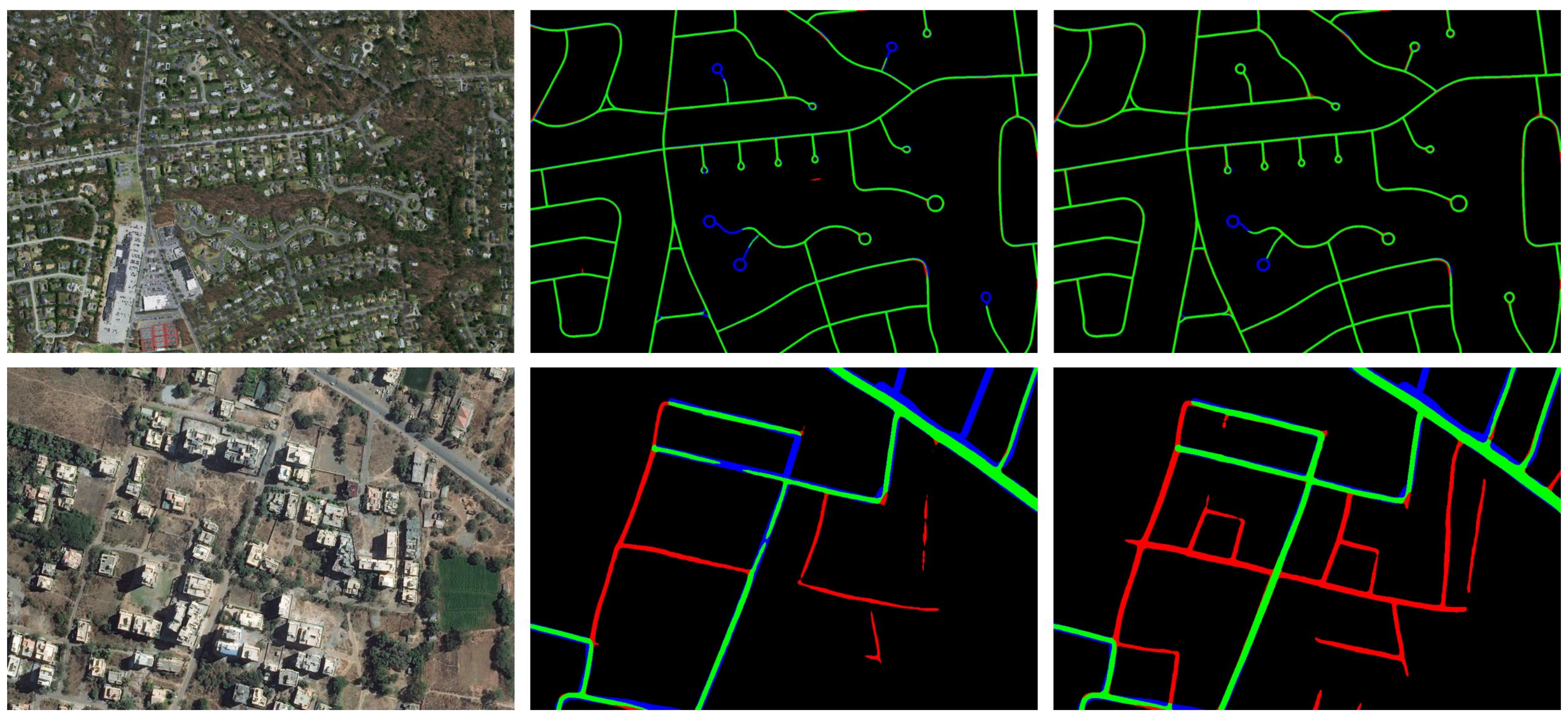


Experiments on Road Datasets

Results on noisy datasets: Massachusetts Roads [1] and DeepGlobe [3]



Qualitative prediction improvements on MA (top) and DG (bottom)



RGB *Baseline confusion* *Noise-aware confusion*

The test labels are also noisy, it affects the metrics evaluation

References

[1] Mnih, "Machine Learning for Aerial Image Labeling," PhD Thesis, University of Toronto, 2013
[2] SpaceNet on Amazon Web Services (AWS). Datasets. The SpaceNet Catalog, 2018
[3] Demir et al., "Deepglobe 2018: A challenge to parse the earth through satellite images," in CVPRW, 2018
[4] Ronneberger et al., "U-Net: Convolutional Networks for Biomedical Image Segmentation," in Medical Image Computing and Computer-Assisted Intervention, 2015
[5] He et al., "Deep Residual Learning for Image Recognition," in CVPR, 2016
[6] Huang et al., "Densely Connected Convolutional Networks," in CVPR, 2017
[7] Reed et al., "Training Deep Neural Networks on Noisy Labels with Bootstrapping," in arXiv, 2014
[8] Wiedemann et al., "Empirical Evaluation Of Automatically Extracted Road Axes," in Empirical Evaluation Techniques in Computer Vision, 1998

REMOTE SENSING TECHNOLOGY INSTITUTE