Deep learning for photogrammetry and remote sensing

Christian Heipke
IPI - Institut für Photogrammetrie und GeoInformation,
Leibniz Universität Hannover

Table of contents

- Introduction
- Deep learning
  - a little bit of theory
  - examples from our work at IPI
- Conclusions

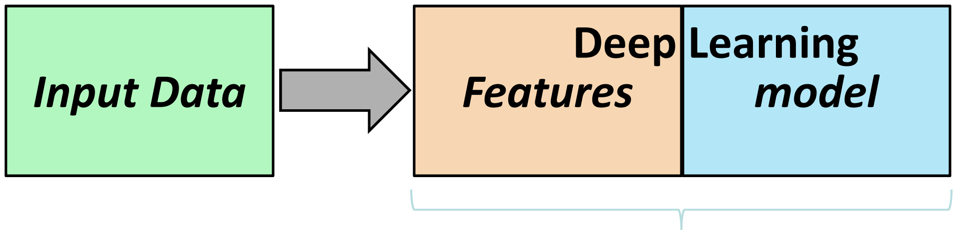
Deep learning

… "learning" an input-output mapping from examples by machines (**supervised classification**) – not so new …

- traditional classification:

most of the "heavy lifting" here, final performance only as good as feature set

- deep learning:

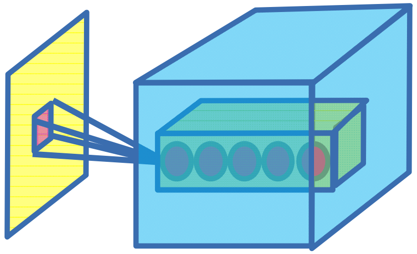


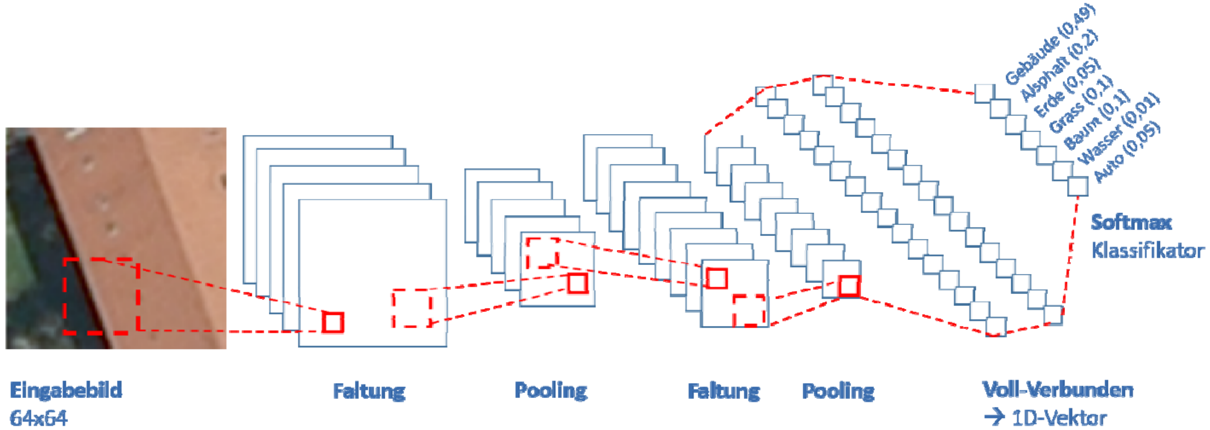
features and model learned together, mutually reinforcing each other

adapted from Michele Catasta

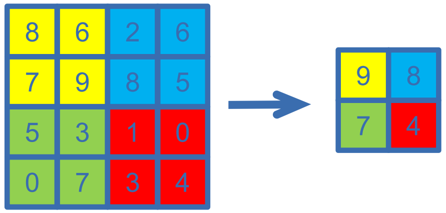
CNN: Convolutional neural networks

- convolution



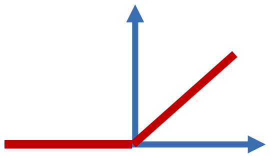


- (max-)pooling



- activation function

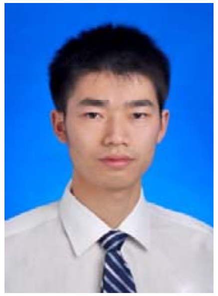
e.g. ReLU(x) = max(0,x)



- for regularly structured data
- repeated execution of those steps based on “some“ architecture followed by classification (one label per patch)
- training of hierarchical feature rep. (filter coefficients) by back prop.

Deep learing at IPI

- **Geometric** applications
  - image orientation of stereoscopic images
  - reliable dense matching



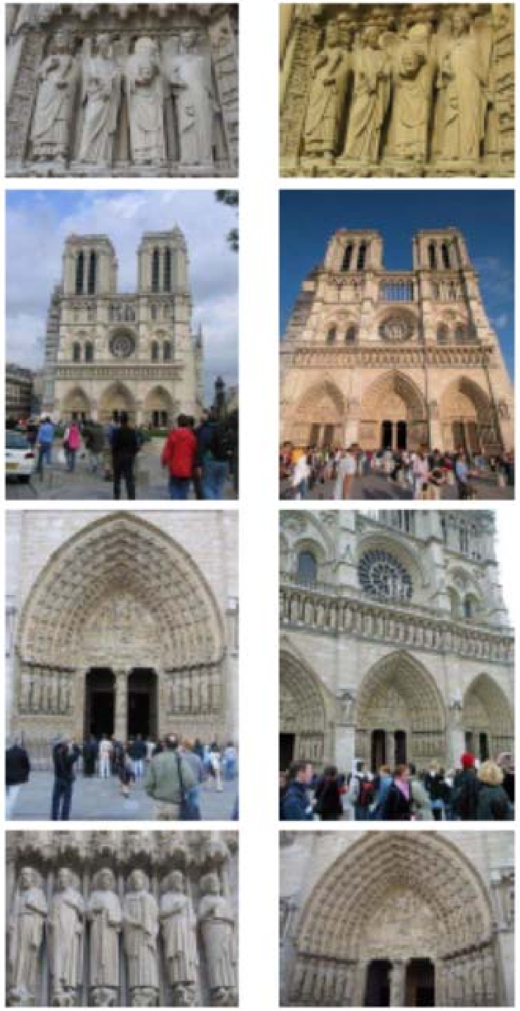
- **Topographic** applications
  - classification and update of topographic databases
  - domain adaptation for semantic segmentation
  - identification of bomb craters in historical aerial images

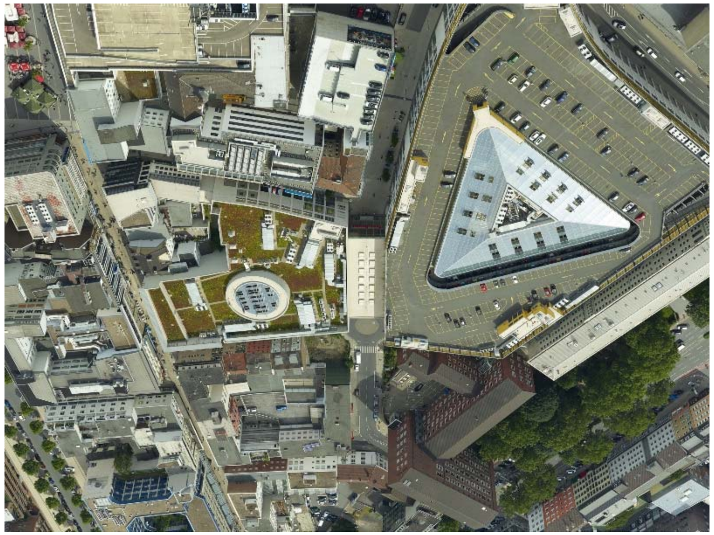
- **Traffic and surveillance** applications
  - pedestrian detection and trackin
  - person re-identification

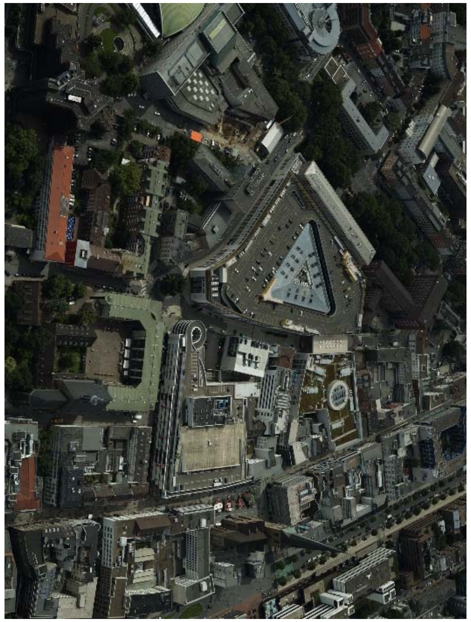
Deep learing at IPI

- **Methodological** questions in deep learning for image analysis
  - use of deep learning for well understood mathematical operations (collinearity equations, image matching, …)
  - geometric accuracy of object delineation
  - transfer learning and domain adaptation in deep learning
    - necessary amount of training data
    - necessary quality of training data
  - time-dependent learning and recurrent neural networks
  - multi-task learning

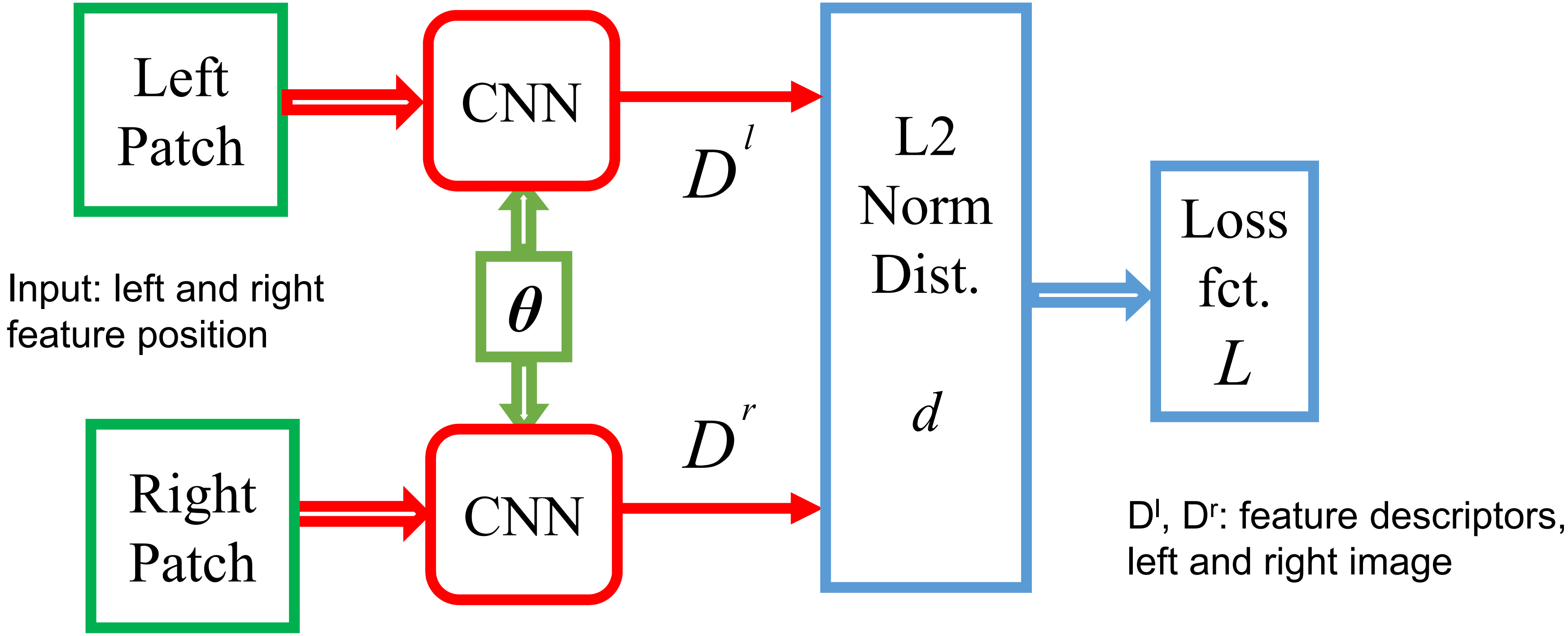
Example – image orientation







Example – image orientation



- Siamese network: two networks with shared parameters θ
- train θ with positive and negative examples using SGD

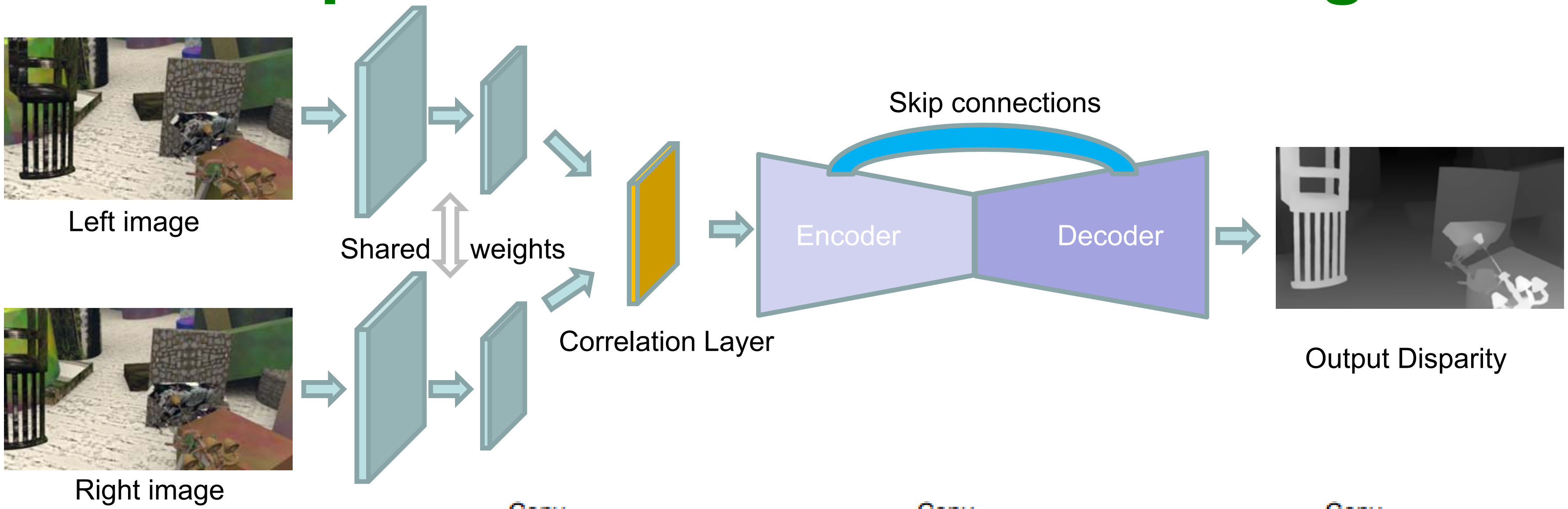
Chen L., Rottensteiner, Heipke C.. "Invariant descriptor learning using a Siamese convolutional neural network." *XXIII ISPRS Annals (3)* 2016.

Example – image orientation

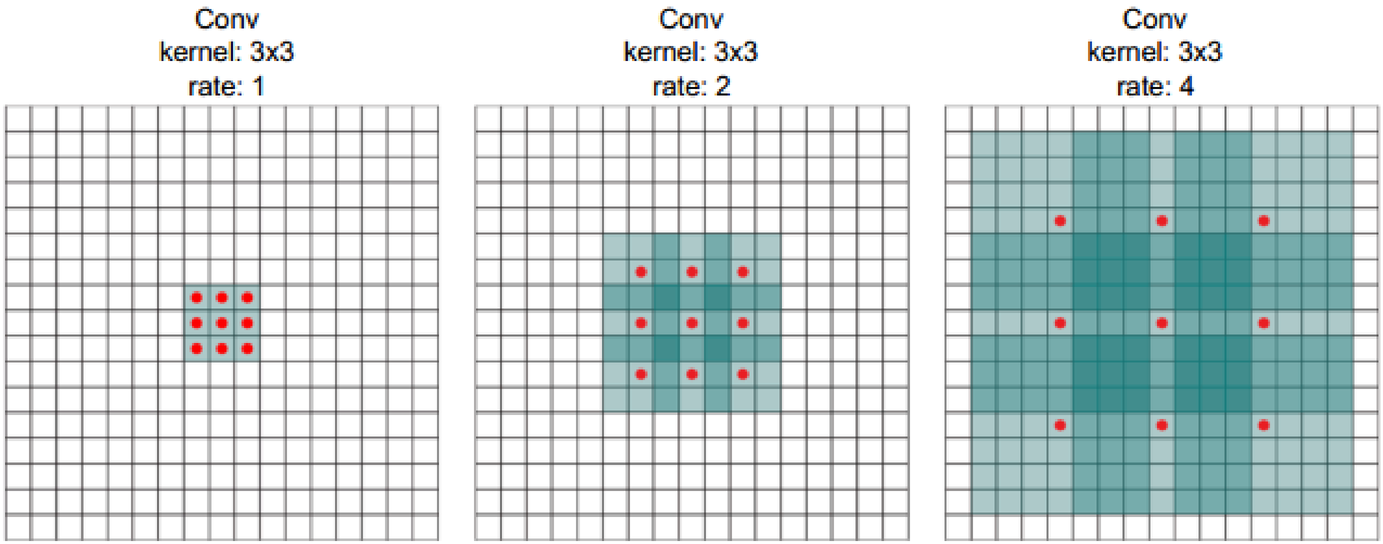
Training	Test	Ours	SIM	BR	TRZ	OS	SIFT
ND	Yos	11.6	**10.1**	13.6	15.9	15.3	29.2
ND	Lib	**11.6**	12.4	16.9	17.9	14.6	36.3
Lib	ND	**6.4**	7.2	-	14.7	10.1	28.1
Lib	Yos	11.3	**11.2**	-	20.9	17.6	29.2
Yos	ND	8.4	**6.8**	18.3	14.8	9.5	28.1
Yos	Lib	**14.4**	14.6	12.0	22.4	17.6	36.3
Mean		10.6	10.4	15.2	17.8	14.1	31.2

- Rate of FP for 95% recall (smaller is better)
- All CNN based approaches are significantly better than SIFT

Example – reliable dense matching

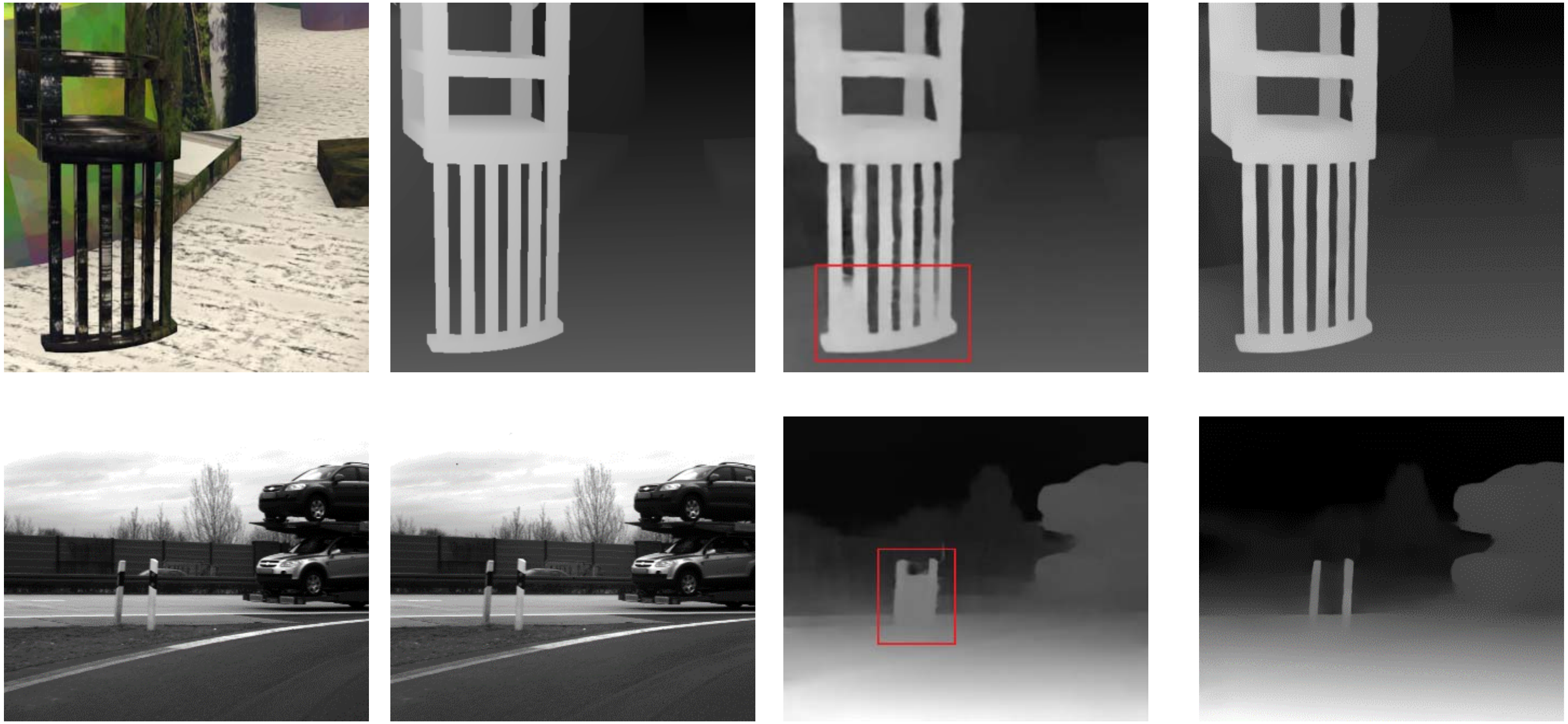


- DispNetC
- Separate
- Correlatio
- New[2]:
  - dilatec
  - L1 gra
  - enhan



[1] Mayer, Nikolaus, et al. "A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation." CVPR, 2016; [2] Kang et al., 2019, in prep.

Example – reliable dense matching

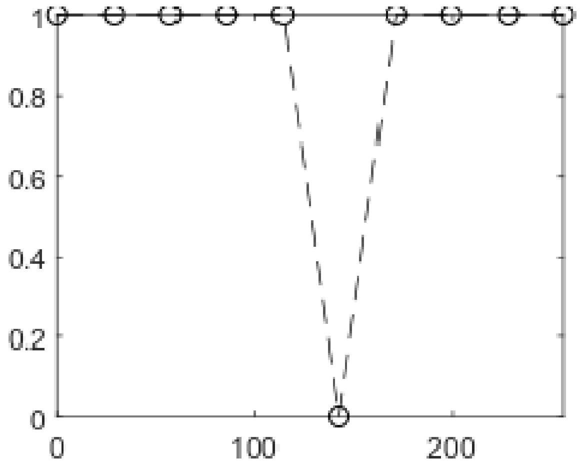


left image gt/right image baseline Our

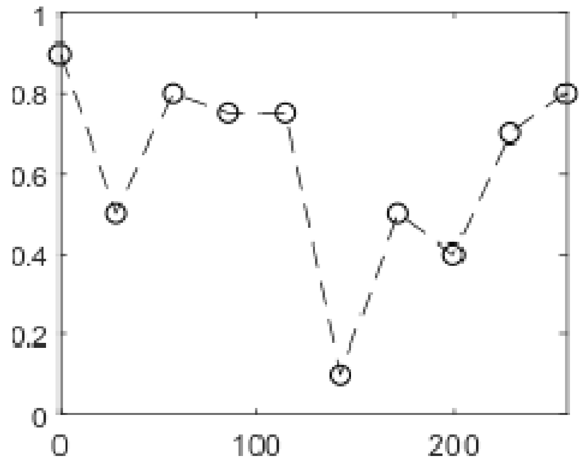
Example – reliable dense matching

High Confidence

Theoretically optimal curve

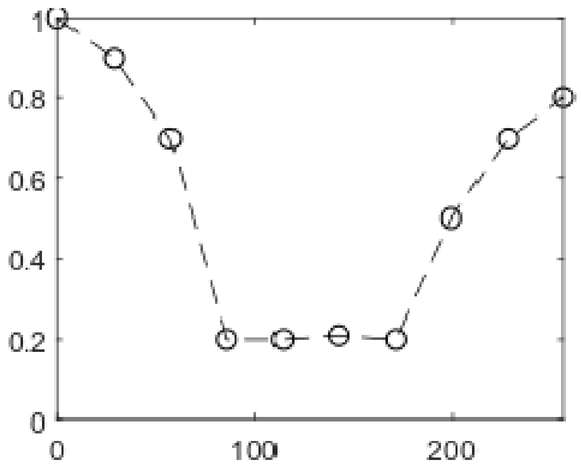


Clear global minimum

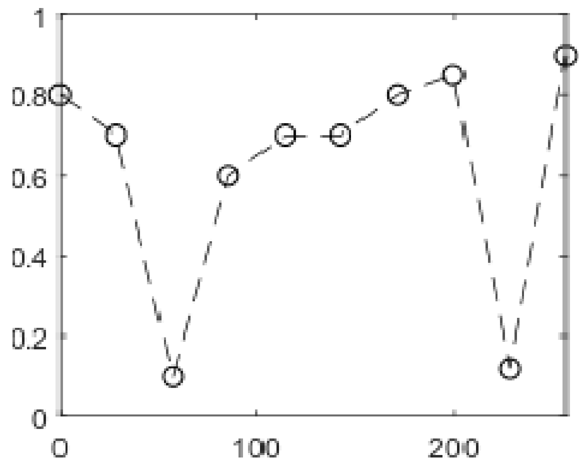


Low Confidence

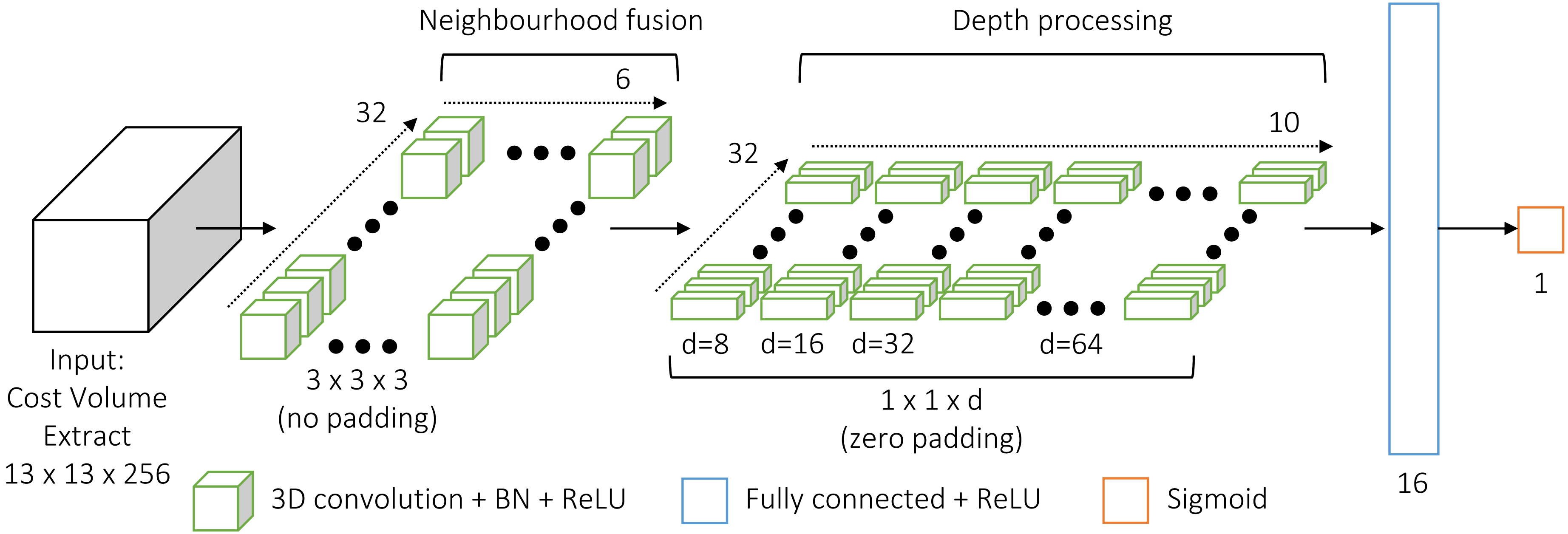
Saddle point



No distinct global minimum



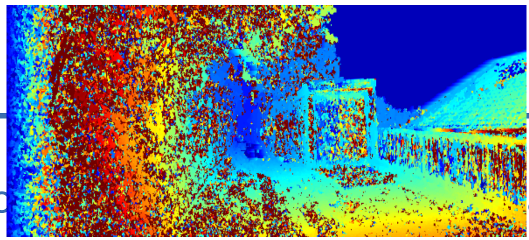
Example – reliable dense matching

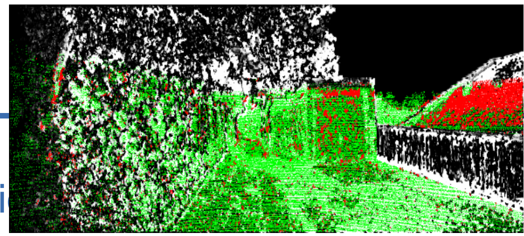


left image

right image

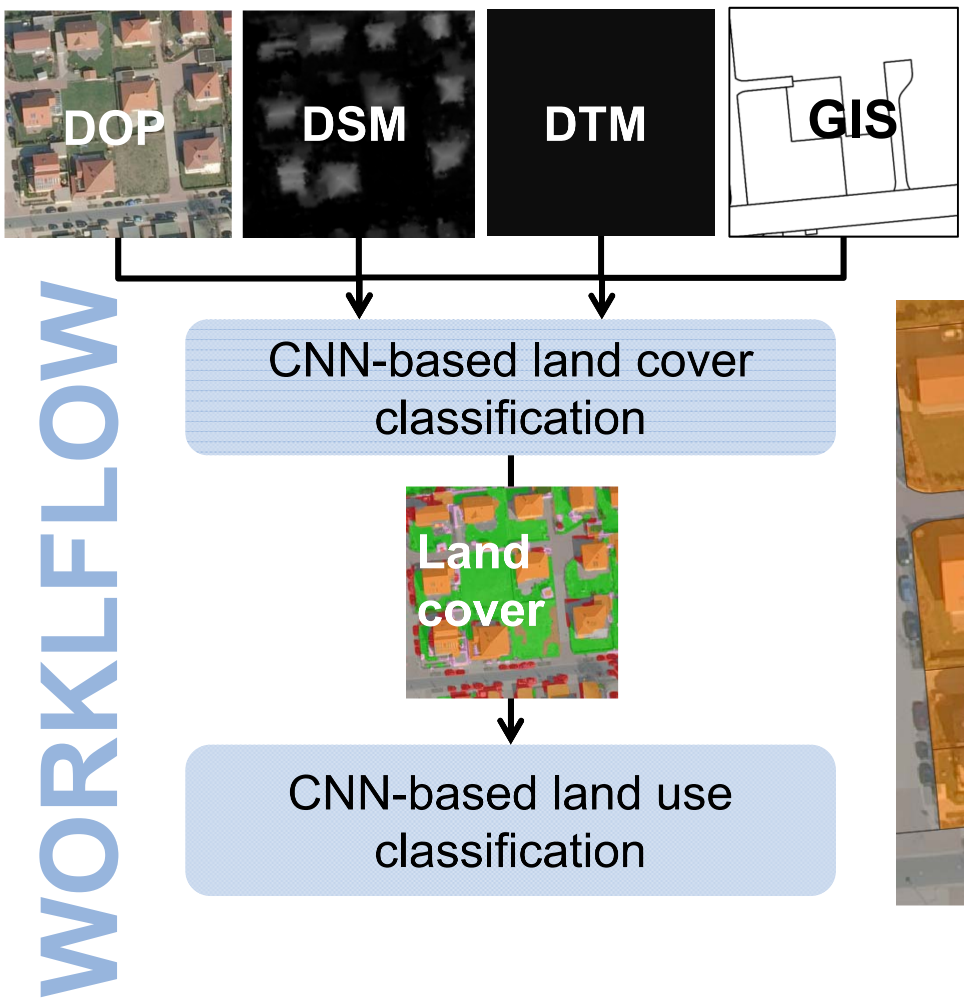
disparity map





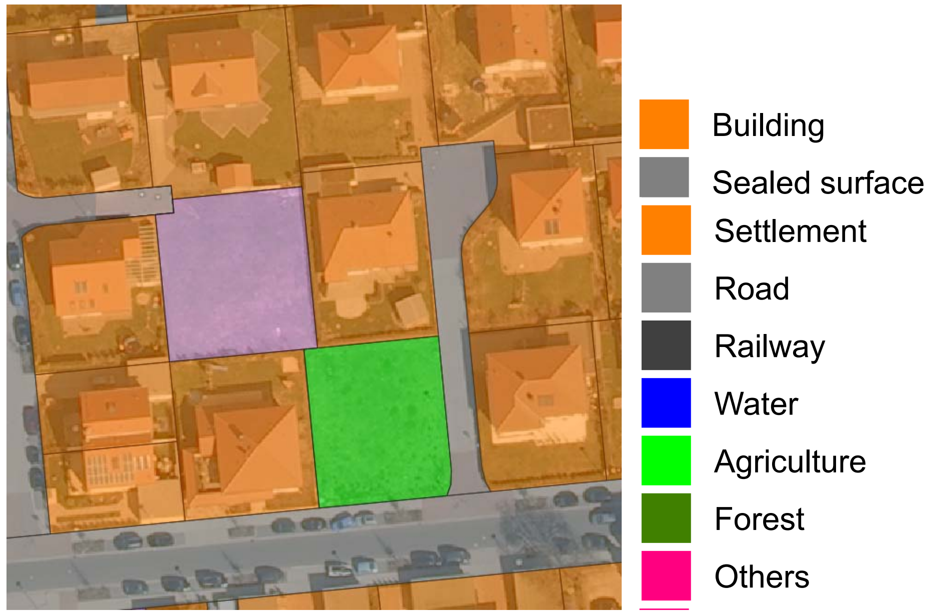
confidence map

Example – land cover / land use update



OA (LC) = 89.3 %

OA (LU) = 82.9 %

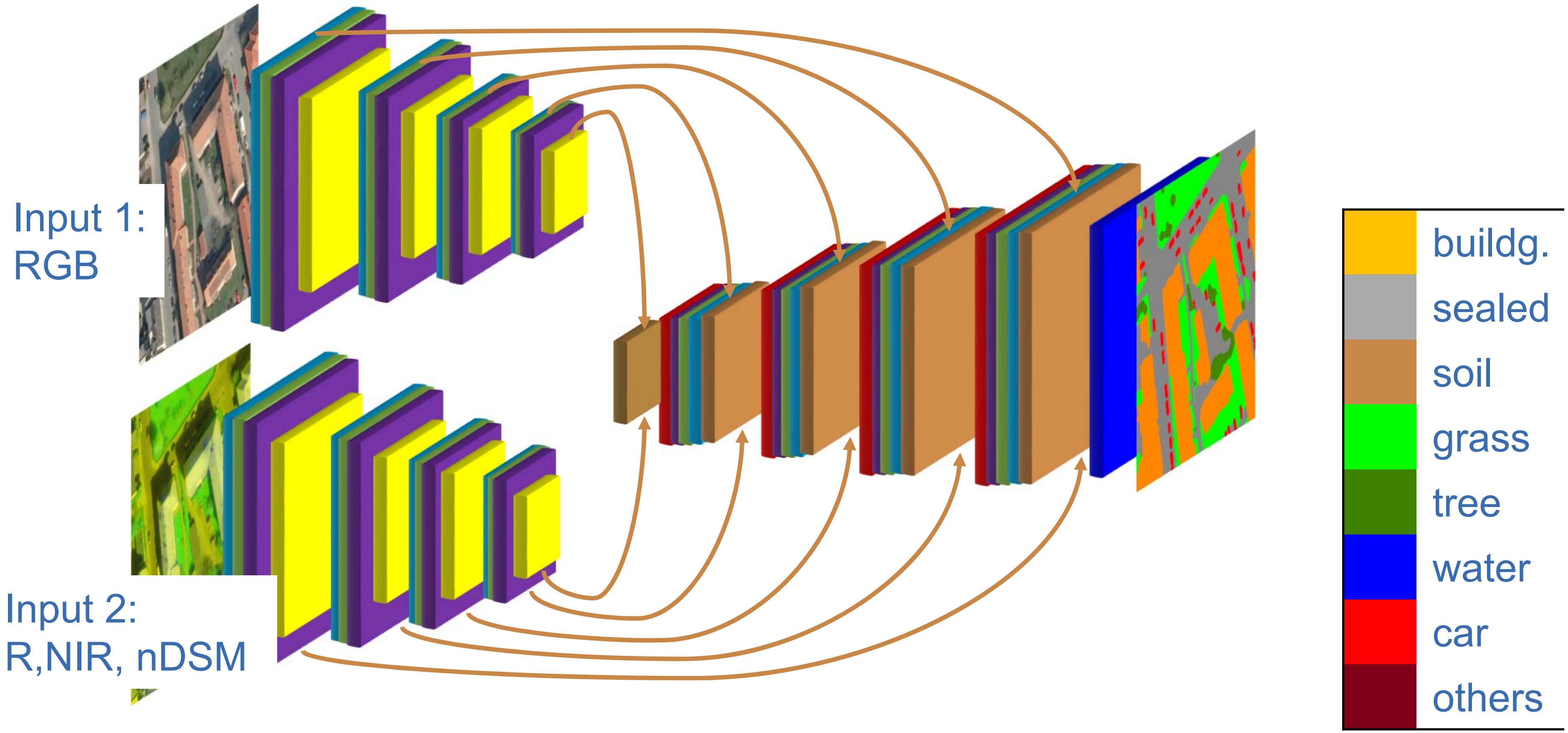


Yang C., Rottensteiner F., Heipke C., 2018: Classification of land cover and land use based on convolutional neural networks. ISPRS Annals of the Ph, RS IV-3, 251-258.

Example – land cover / land use update

LC network: Ensemble classifier: RGB, NIR, nDSM

training from scratch incl. data augmentation

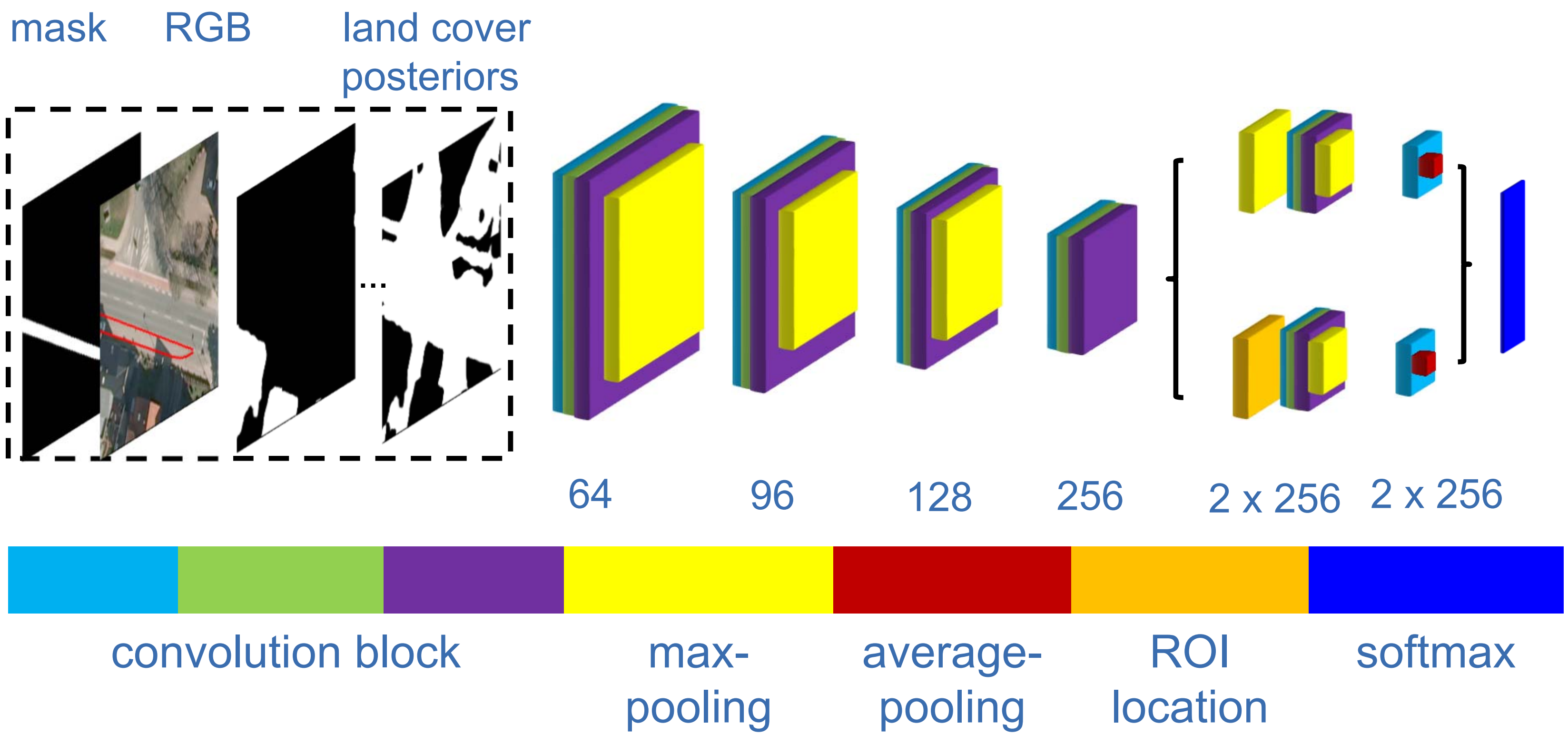


Example – land cover / land use update

LU network

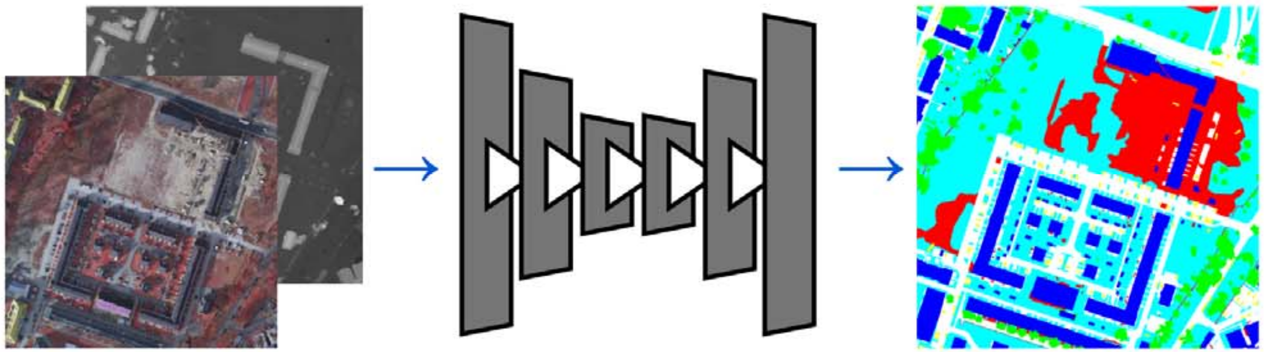
predict one class label per land use object from the database

utilize mask image to locate region of interest (ROI)

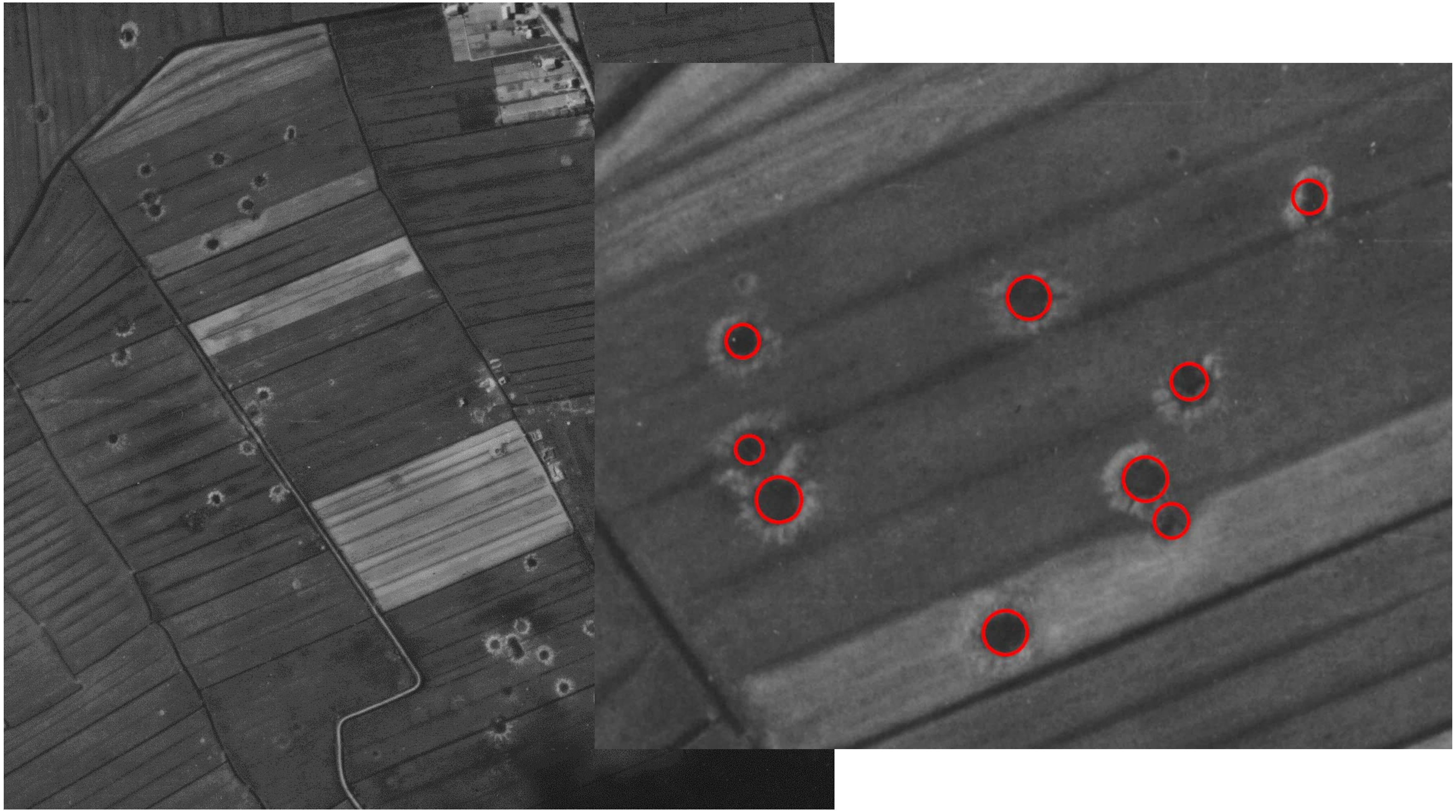


Example – domain adaptation for semantic segmentation

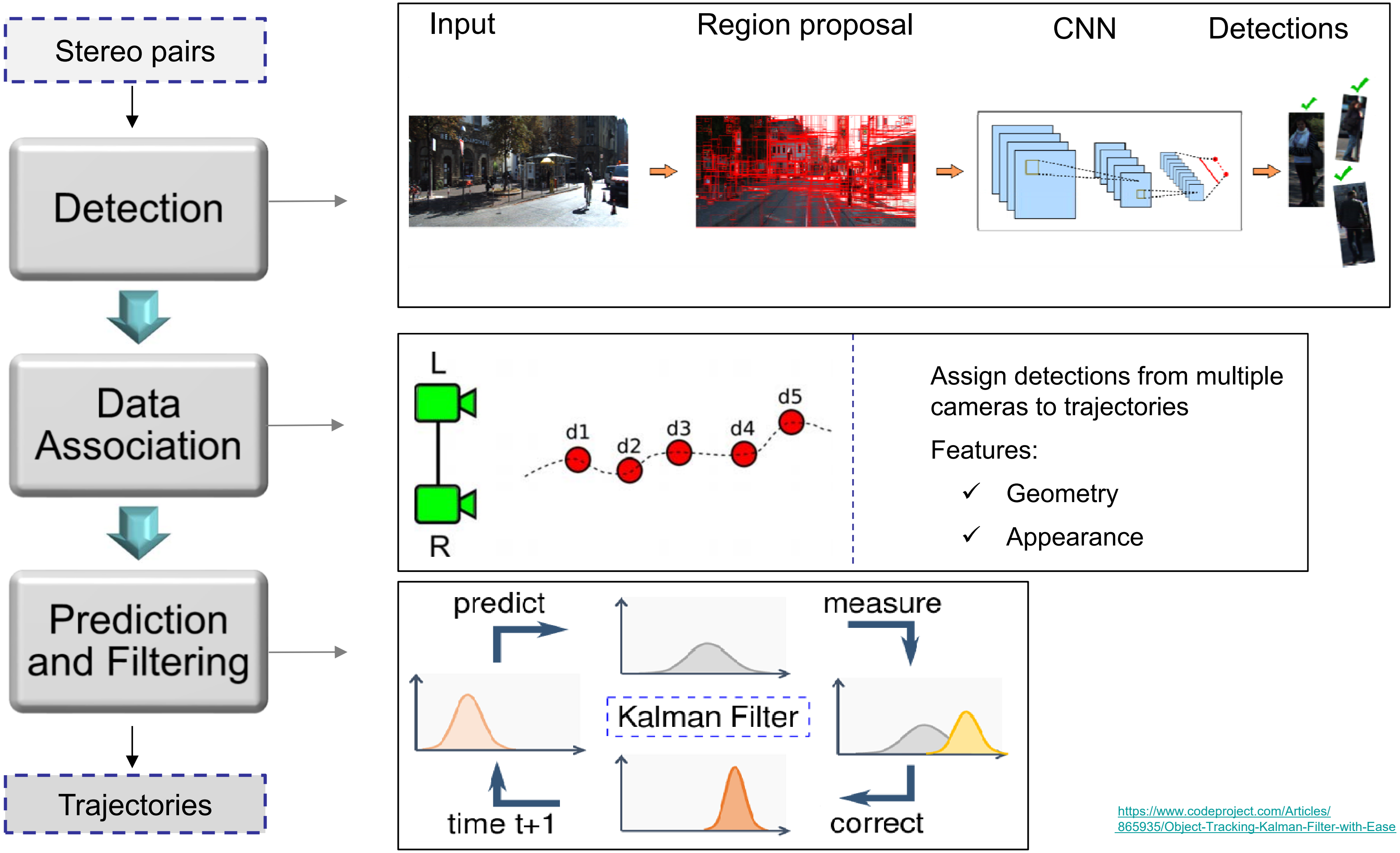
- Pixel-wise classification (semantic segmentation) with very few or no training data using CNN
  - transfer learning: retraining / fine-tuning a trained model
  - domain adaptation: use target domain samples to adapt a model to new domain
    - make samples from target domain „**look**“ like those from source domain
    - match **representations** for samples from both domains



Example – bomb craters



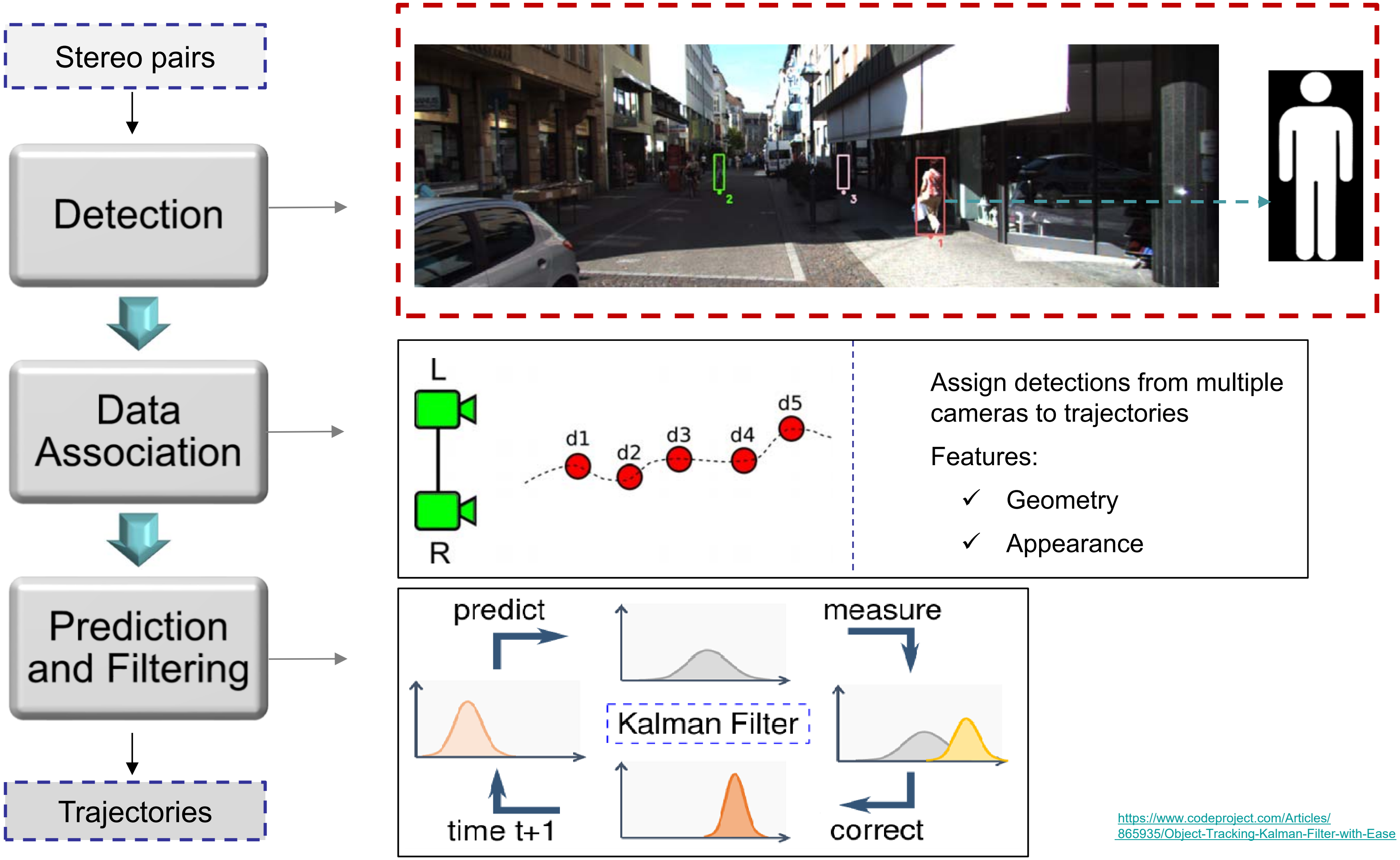
Example – pedestrian detection and tracking



https://www.codeproject.com/Articles/865935/Object-Tracking-Kalman-Filter-with-Ease

Nguyen D.X.U., Rottensteiner F., Heipke C., 2018: Object Proposals for Pedestrian Detection in Stereo Images. In: Kersten T., Gülch E., Schiewe J., Kolbe T., Stilla U. (Eds.), 2018: DGPF (27).

Example – pedestrian detection and tracking



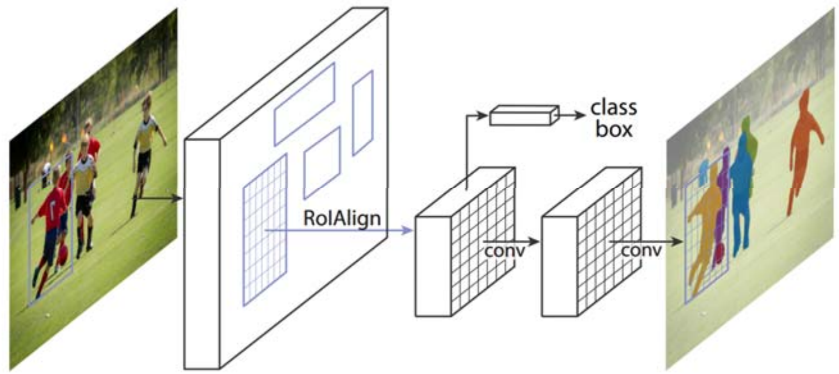
https://www.codeproject.com/Articles/865935/Object-Tracking-Kalman-Filter-with-Ease

Nguyen D.X.U., Rottensteiner F., Heipke C., 2018: Object Proposals for Pedestrian Detection in Stereo Images. In: Kersten T., Gülch E., Schiewe J., Kolbe T., Stilla U. (Eds.), 2018: DGPF (27).

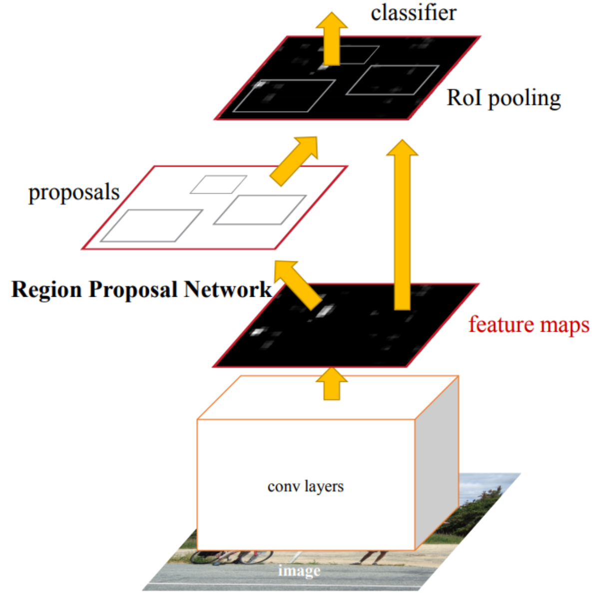
Example – pedestrian detection and tracking

- Detection: Mask-R-CNN [1]
  - depth map from stereo
  - region proposals and delineation from Mask-R-CNN

- Association across time: TriNet
  - (not shown here)



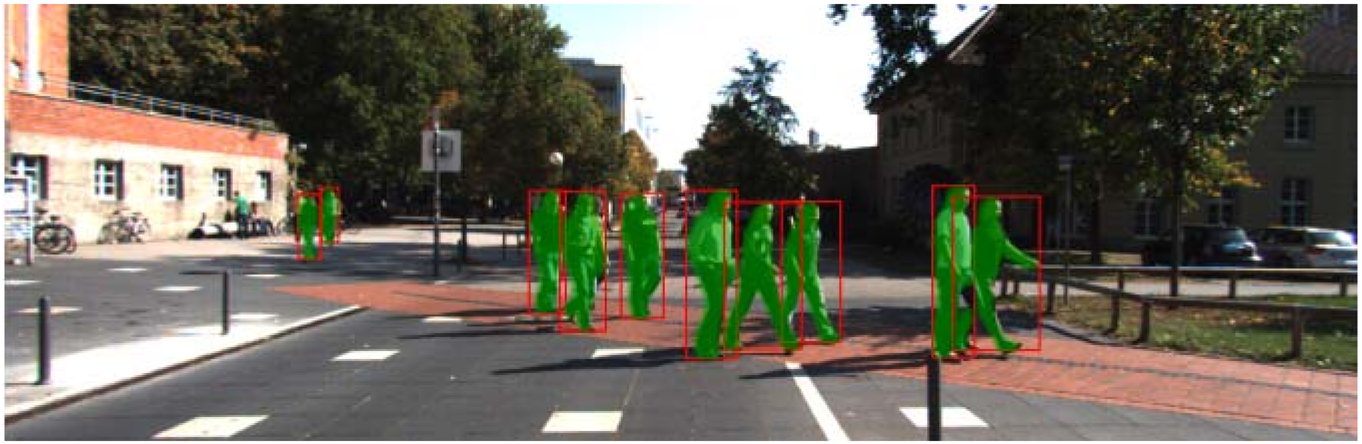
He et al. 2017



[1] He, K., Gkioxari, G., Dollár, P., & Girshick, R.. Mask R-CNN. In *Computer Vision (ICCV 2017)* (pp. 2980-2988).

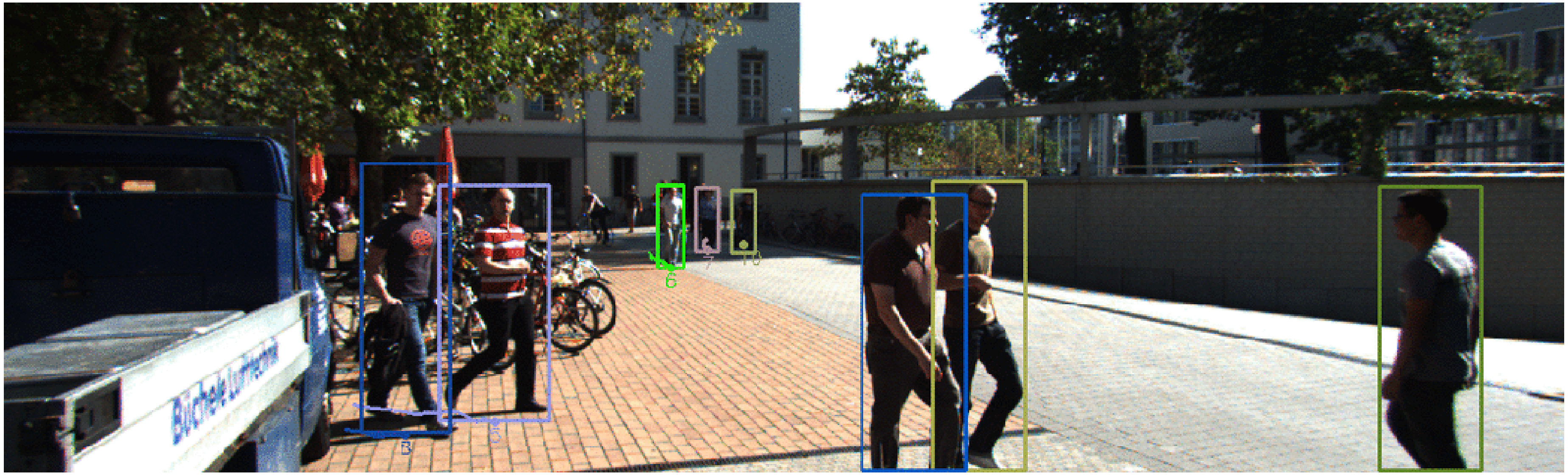
Example – pedestrian detection and tracking

Mask R-CNN results



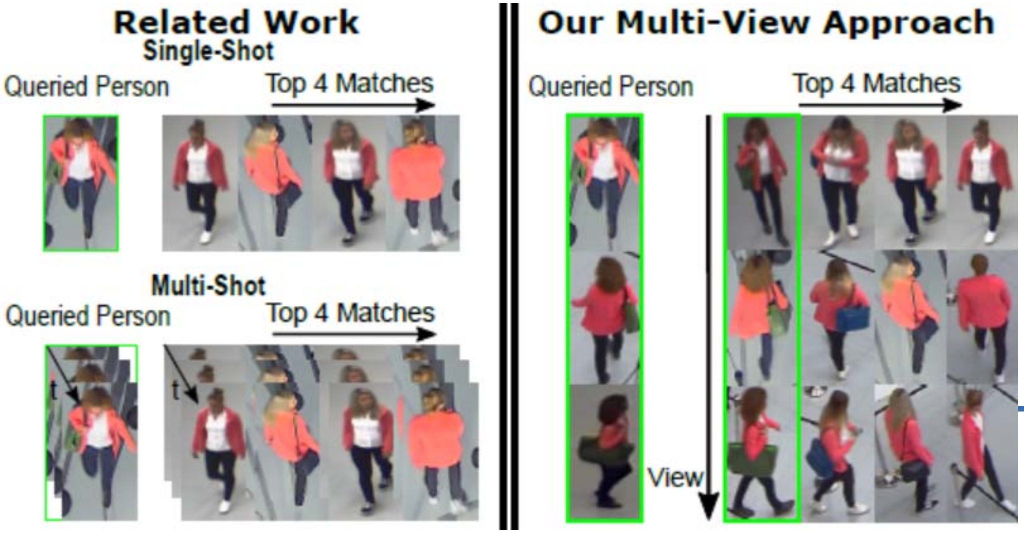
Example – pedestrian detection and tracking

- Sample results for tracking



Example – tracking across views (PRID)

- Multi-view approach: one and the same person from front, side and back
  - different views from fish eye camera
- Detection and view classification
  - VGG 16, Resnet 50
- Triplet matching with gallery sample
  - TriNet



Blott G, Yu J, Heipke C., 2018: View-aware person re-identification, DAGM/GCPR Stuttgart, 2018.

Example – multitask learning

Genre?

Epoch?

Artist?

Location?

Technique?

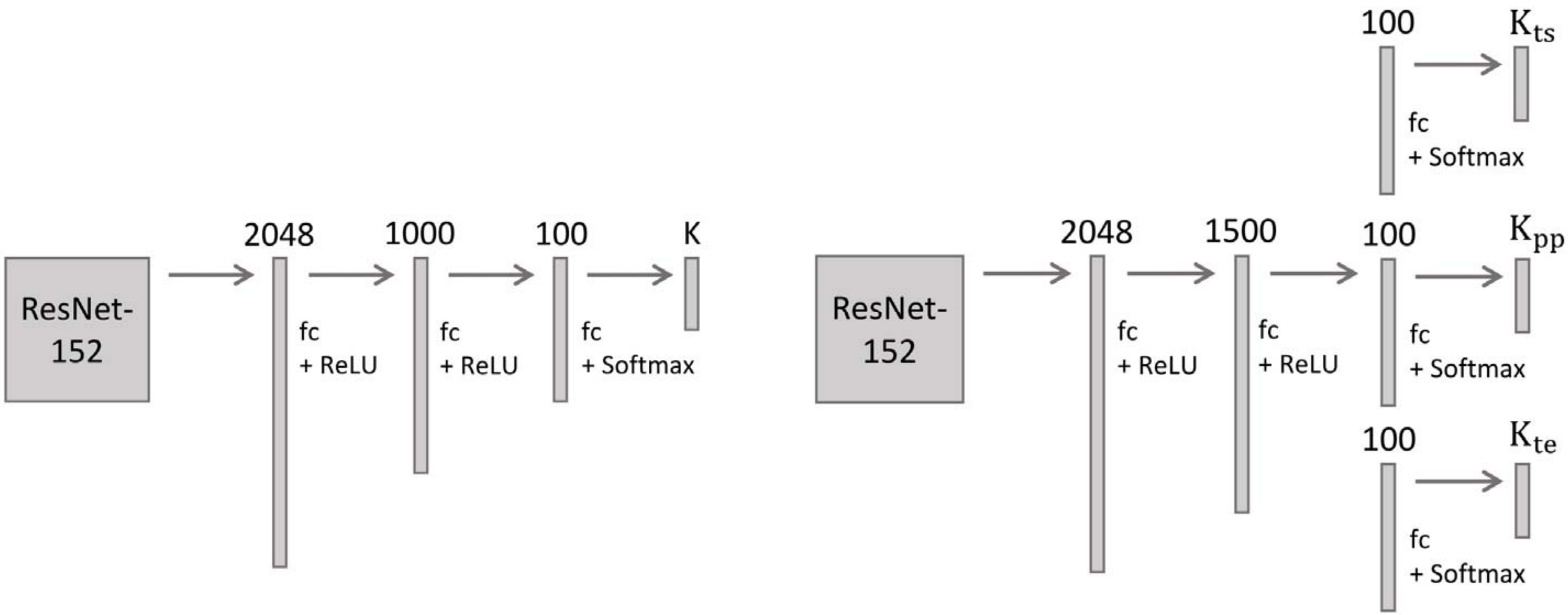
Material?

Dorozyinski et al., 2019, in prep.

Example – multitask learning

Single-task Learning

Multi-task Learning



Conclusions

- many other examples in the last few years (see e.g. review Zhu et al., 197 reference entries)
  - **more flexible** wrt. different object characteristics
  - strength: **learning features**, classifier not so important
  - key to good performance: **network depth, no. of training data**
  - has **outperformed just about all classical algorithms** by a large margin, provided enough training data is available
  - is about to do the same in geometric tasks
- no expert-free totally automatic processing chain, but **integration with human** capabilities

---

Zhu X., Tuia D., Mou L., Xia G., Zhang L., Xu F., Fraundorfer F., 2017. Deep Learning in Remote Sensing: A comprehensive review and list of resources. IEEE GRSS Magazine 5(4): 8-36.

Conclusions

- CNN is not magic - fundamentally a classifier
  - based on **correlation** between data sets
  - an enormous **amount of unknowns** to be estimated
    - lots of training data, long training times (days … weeks)
  - sensitive to
    - **overfitting** (“curse of dimensionality”)
    - **non-rep. training data** (incorrect / biased / unbalanced / …)
    - **user choices**, incl. hyper-parameters
      - network architecture, (non-lin.) activation function
      - design of loss function (function to be optimised)
      - training parameters (weight initialisation, learning rate, drop out rate, “momentum”, batch size, regularisation, …)
  - **generalisation and prediction capabilities unclear**
    - the system can’t learn what it never saw

Conclusions – to do's

- integrate **prior knowledge**, e.g. physical models
  - don't try to learn what we already know (e.g. laws of nature)
- CNNs for **point clouds** and other irregularly rep. data
- tackle **training data shortage**
  - weakly, semi-, unsupervised learning, reinforcement learning
  - data augmentation and simulation
  - deal with errors in training data, e.g. when using outdated DB
  - combination with crowd sourcing and social media data
  - investigate limits of pre-training + fine tuning
- pay attention to **geometric accuracy** of object delineation
- sequential data, **time series** processing - RNNs
  - streaming (big) data, online CNN learning