Thesis Introduction

2020.10.15
Yi Wang

Supervisor: Dr. Ksenia Bittner
Prof. Norbert Haala

Universität Stuttgart
ifp

DLR
Deutsches Zentrum für Luft- und Raumfahrt
German Aerospace Center

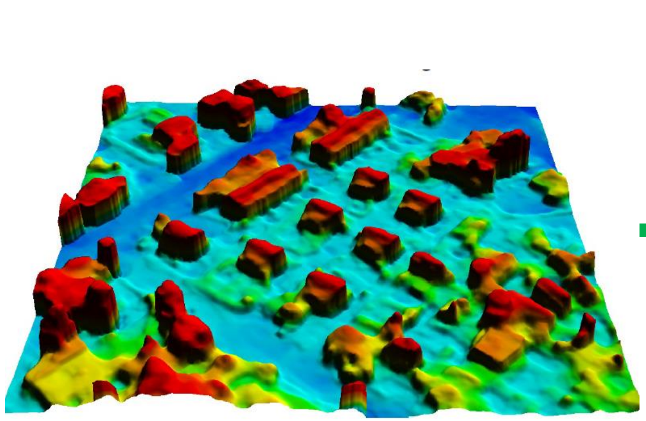
Outline

- General idea
- Methodology
- References

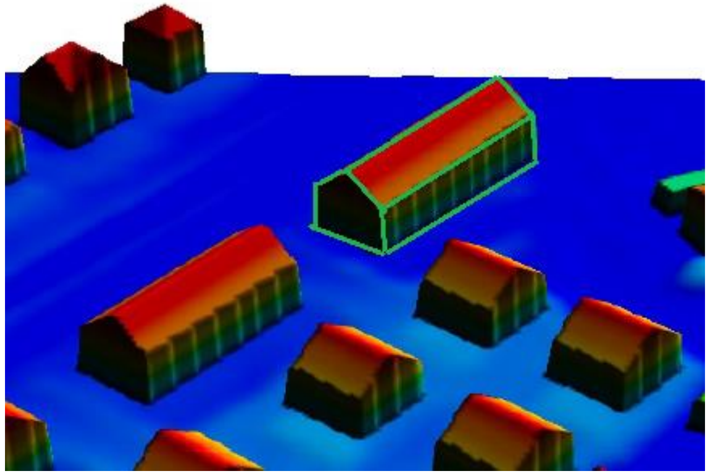
General idea

Core idea: Machine-learnt 3D vectorization of buildings from satellite imagery

- Methodology: multi-task conditional generative adversarial network (cGAN)
- Objective: refined DSM + (instances / edges) + 3D vectorization



[1]



- Source data: stereo DSM + Panchromatic image (WV-1) / RGB image (WV-4)

Methodology

Generative adversarial network:

- Given the training data, two neural networks compete with each other: the Generator learns to generate real-like samples, while the Discriminator learns to distinguish between generated samples and real data.
- In other words,
  - Discriminator: binary classifier (real or fake)
  - Generator: learns to fool the discriminator (learns real data's distribution)
- A min-max problem:

$$\min_G \max_D V(D,G) = \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))]$$ [2]

- In theory, it is proved that this game will go to Nash-equilibrium:

$$P(D(real)) = P(D(fake)) = 0.5$$

Methodology

ifp

Conditional GAN: (Generally) apart from random noise, input also other constraints (label, text, image, etc.) to control the generated outputs.

Pix2pix: train source & destination image pairs to perform image to image style translation.

$$\min_G \max_D \mathcal{L}_{\mathrm{cGAN}}(G, D) = \mathbb{E}_{\boldsymbol{x},\boldsymbol{y}\sim p_{\mathrm{real}}(\boldsymbol{y})}[\log D(\boldsymbol{y}|\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{x},\boldsymbol{z}\sim p_z(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z}|\boldsymbol{x})|\boldsymbol{x}))]$$

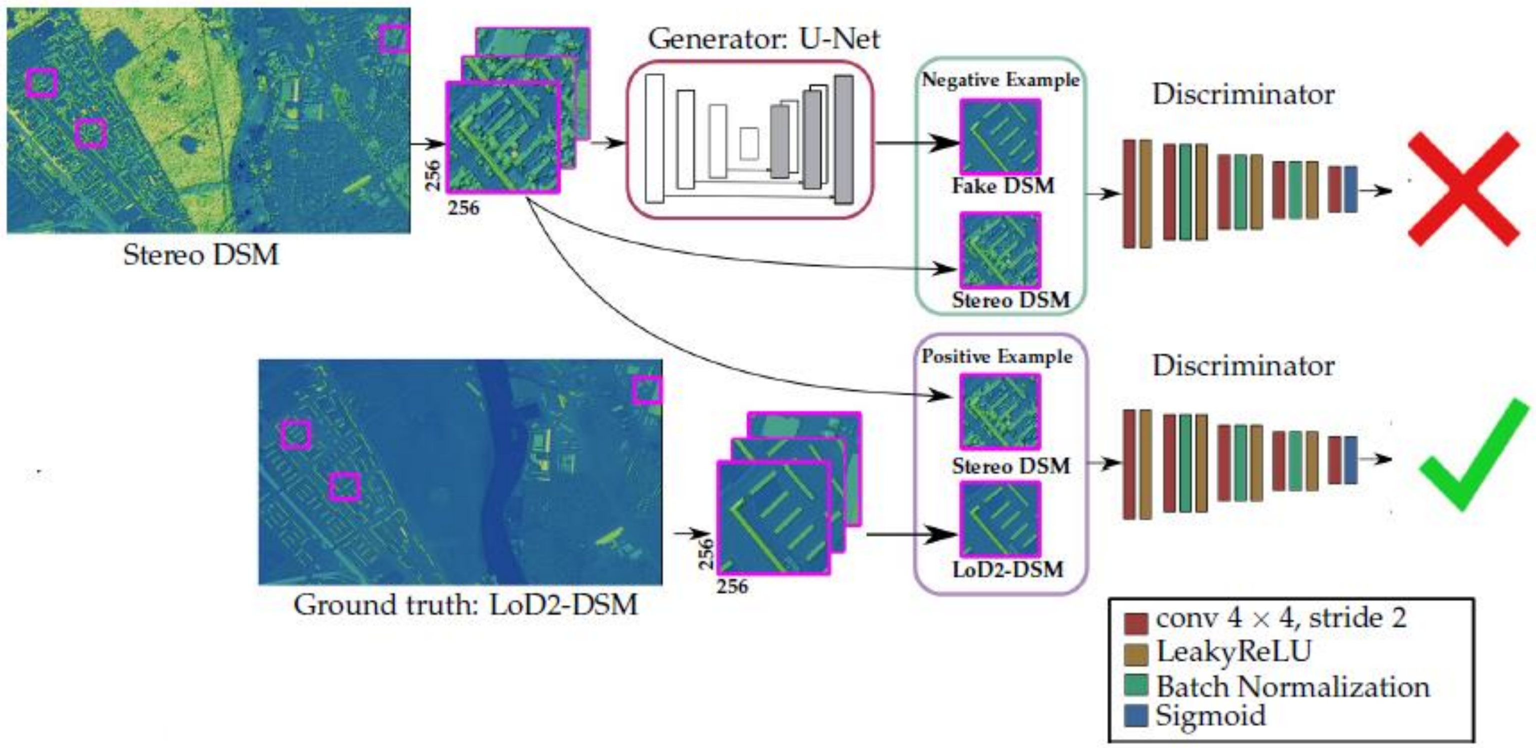
[3]

During training, we build generator (usually ResNet/UNet structure) and discriminator (usually simple CNN) as two independent networks, train and optimize them alternately.

DLR

Methodology

cGAN -> DSM refinement (base):



[4]

Methodology

Building vectorization (new):
- previous research --- 2D polygonization; my task --- 3D vectorization
- Step1: Extract corners and edges from satellite image (DSM+orthophoto)
- Step2: Connect the corners and edges as polygons and combine these polygons for vectorized building

Multi-task learning (new):
- multiple learning tasks are solved at the same time, while exploiting commonalities and differences across tasks.
- In our case, the building instance masks (optional), the edge extraction and the DSM refinement should have useful commonalities, thus we expect the combination of all tasks can improve each other.

Methodology

ifp

Whole Process:

- For the GAN network,
  - Input: Panchromatic/RGB image + stereo DSM
  - Output: refined DSM + extracted core points + (optional, building instance mask)
- Afterwards,
  - Connect the core points with the help of all outputs to get vectorized 3D buildings

- To be mentioned, the ground truth consists of:
  - Ground truth DSM
  - Ground truth core points
  - Ground truth instance mask (optional)

  All these three are to be generated from cityGML data

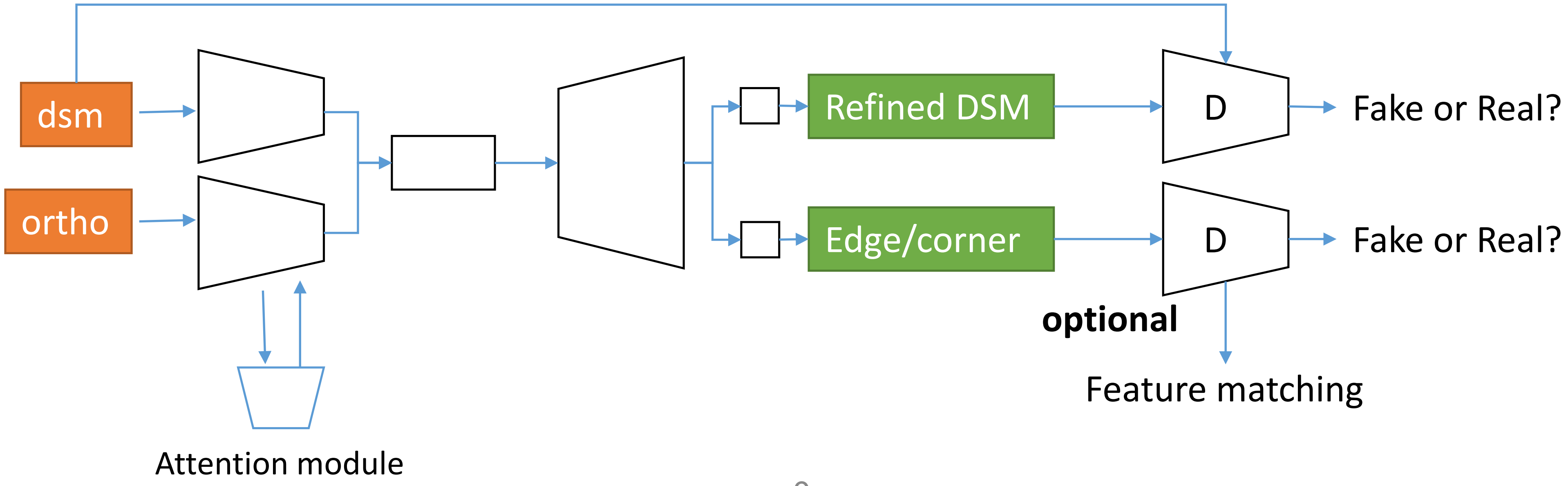
DLR

Methodology

ifp

Ideas for the network structure:

- Generator: U-Net with residual block, Discriminator: PatchGAN
- Several 1x1 conv layers for edge detection task
- Attention module in encoder → importance weight → attention feature map
- Fine tune GAN based on pretrained CNN: Network interpolation [5]
- Optional: Give a Discriminator to edge-detection (segmentation GAN)?



DLR

Methodology

ifp

Ideas for the objective function:

- DSM refinement: LSGAN loss + L1 loss + Surface Normal loss [1]
- Edge detection:
  - CrossEntropy (weighted?)
  - multi-scale edge-detection loss
    - Encoder? → HED[6]
    - Discriminator? → Feature Matching Loss [7]
- Loss weights: Learnable multi-task loss weights [8]

DLR

References

ifp

[1] Bittner, Ksenia, Peter Reinartz, and Marco Korner. "Late or Earlier Information Fusion from Depth and Spectral Data? Large-Scale Digital Surface Model Refinement by Hybrid-cGAN.“ Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2019.

[2] Goodfellow, Ian, et al. "Generative adversarial nets." Advances in neural information processing systems. 2014.

[3] Isola, Phillip, et al. "Image-to-image translation with conditional adversarial networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[4] Bittner, Ksenia, et al. "Dsm-to-lod2: Spaceborne stereo digital surface model refinement." Remote Sensing 10.12 (2018): 1926.

DLR

References

[5] Wang, Xintao, et al. "Esrgan: Enhanced super-resolution generative adversarial networks." Proceedings of the European Conference on Computer Vision (ECCV). 2018.

[6] Xie, Saining, and Zhuowen Tu. "Holistically-nested edge detection." Proceedings of the IEEE international conference on computer vision. 2015.

[7] Nazeri, Kamyar, et al. "Edgeconnect: Generative image inpainting with adversarial edge learning." arXiv preprint arXiv:1901.00212 (2019).

[8] Kendall, Alex, Yarin Gal, and Roberto Cipolla. "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.