Predicting Ground-Level Scene Layout from Aerial Imagery

Menghua Zhai, Zachary Bessinger, Scott Workman, Nathan Jacobs
http://cs.uky.edu/~ted

IEEE 2017 Conference on Computer Vision and Pattern Recognition
CVPR July 21-26 2017 HONOLULU

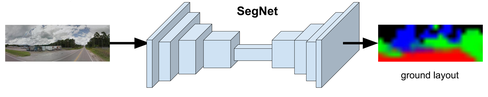
Big Picture

Key Idea:
Train a network to predict the semantic layout of a ground-level panorama from an overhead image centered on the camera location.

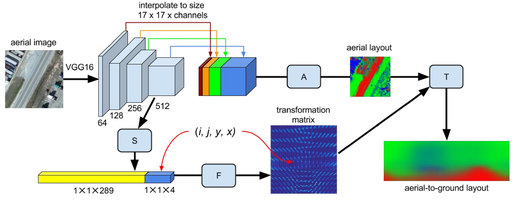
Motivations: This enables several different applications:
- Weakly supervised learning for aerial image classification/segmentation
- Train a network to predict the semantic layout of a ground-level panorama from an overhead image centered on the camera location
- Orientation estimation and fine-grained geocalibration for ground-level images
- Estimating the ground-level appearance of an arbitrary location

Algorithm Overview

Ground-level Layout as Ground Truth:
Collect the ground layout ground truth using SegNet (Badrinarayanan & Kendall 2015).



Aerial-to-Ground Layout Prediction:
1. Extract hypercolumns (Bansal & Chen 2016) to predict aerial layout
2. Estimate the transformation matrix using the aerial image
3. Transform the aerial layout to the ground layout using the transformation matrix



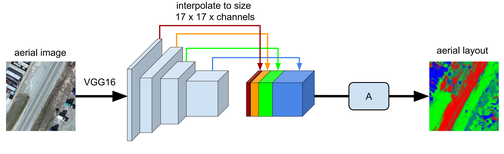
Layouts Distance Minimization:
Minimize the pixel distance between the ground layout ground truth and the aerial-to-ground layout prediction.

Cross-view Dataset

Example of an aligned aerial/ground image pair from our dataset. In the aerial images, north is the up direction. In the ground images, north is the central column.

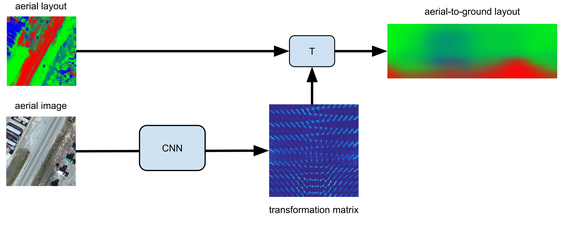
Aerial-Level Layout Prediction

- VGG16 ($conv_1$ to $conv_4$) + Hypercolumn approach for segmentation.
- Use of hypercolumns: Flexible output resolution control.



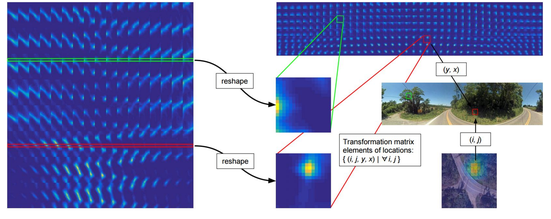
Aerial-to-Ground Transformation

1. Estimate the transformation matrix, M, from the aerial image
2. The transformation from the aerial layout, L_a, to ground layout, L_g, is a matrix multiplication: $L_a = M \cdot L_g$.



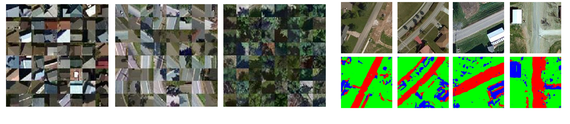
Transformation Matrix

The transformation matrix encodes the relationship between the aerial-level pixels and the ground-level pixels.



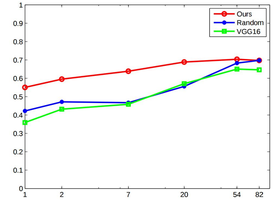
Experiments

Weakly Supervised Learning for Aerial Image Classification and Segmentation:
The resulting network learns to extract semantic features from an aerial-level image, all without any manual annotated aerial imagery.



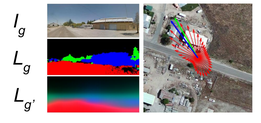
Network Pre-training for Aerial Image Segmentation:
Our proposed technique can be used as a pre-training strategy for semantic-pixel labeling of the aerial-level imagery.

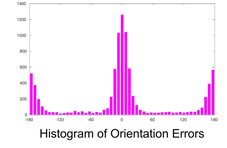
We carried out the pre-training experiment using three different initializations of the VGG16 convolutional layers on the ISPRS dataset (Rottensteiner & Sohn 2013).



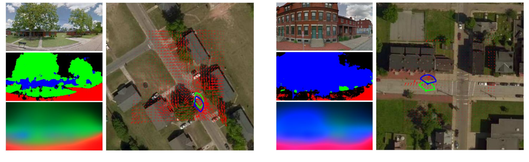
Other Applications

Orientation Estimation:





Fine-Grained Geocalibraion:



Ground-level Image Synthesization:

Conclusion

Main innovations:
- A novel strategy to relate aerial-level imagery and the ground-level imagery
- A novel strategy to exploit the automatically labeled ground images as a form of weak supervision for aerial imagery understanding
- Show the potential of our method in geocalibraion and ground appearance synthesization using algorithm

Acknowledgement: We gratefully acknowledge the support of NSF CAREER grant (IIS-1553116), a Google Faculty Research Award, and an AWS Research Education grant.