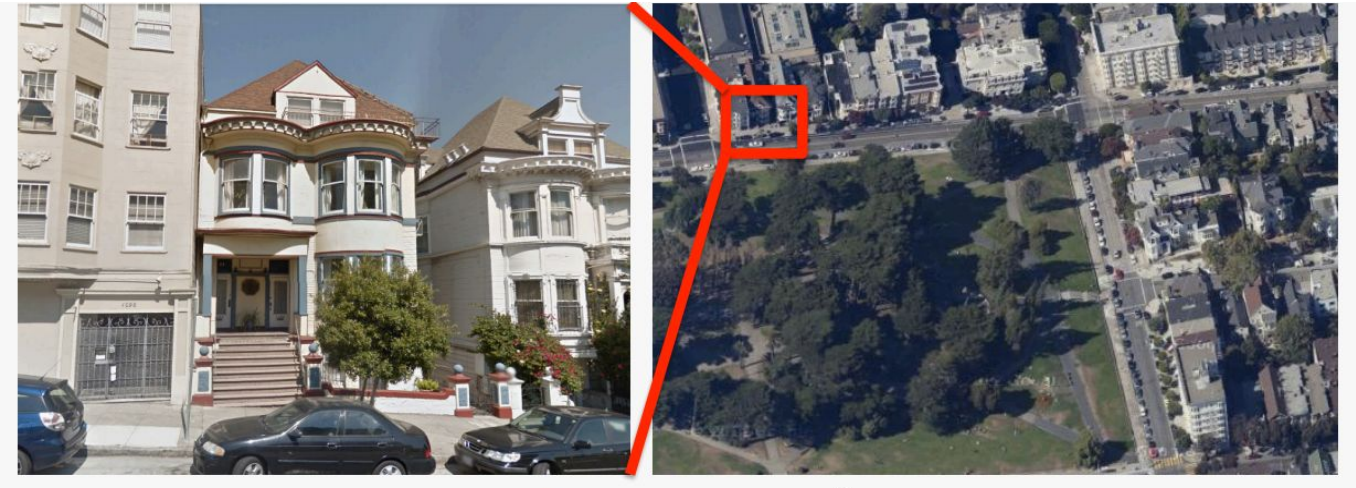
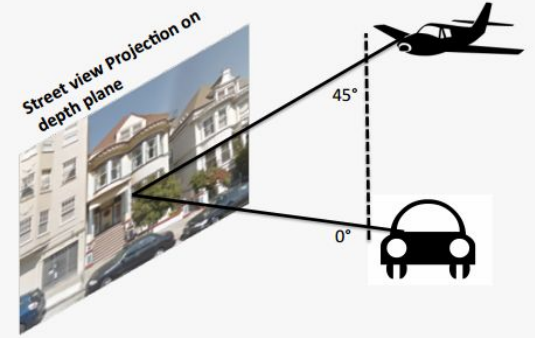
Learning Deep Representations for Ground-to-Aerial Geolocalization

Ground-View and Aerial-View





Match ground-view images to aerial-view ones

- Inspired by deep learning success in facial verification
- Using CNN for processing
- Propose “Where-CNN”
- There are similar approaches(IM2GPS + 3 similar)

Dataset

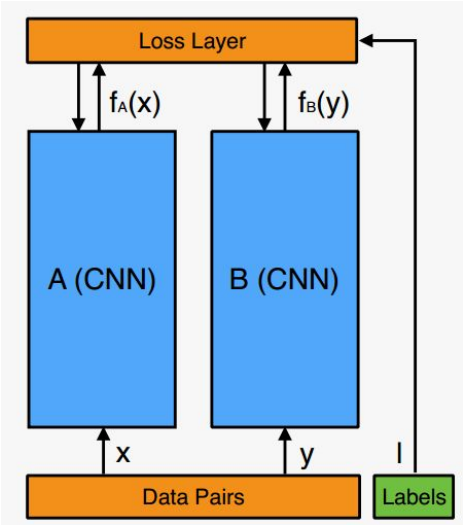
- Google street-view and 45 aerial view images
- Seven cities, different styles
- Including urban and suburban areas

Feature representations

- Hand-crafted features
- Generic deep-learning feature representations
- Learned feature representations from data

Network Architecture

Training

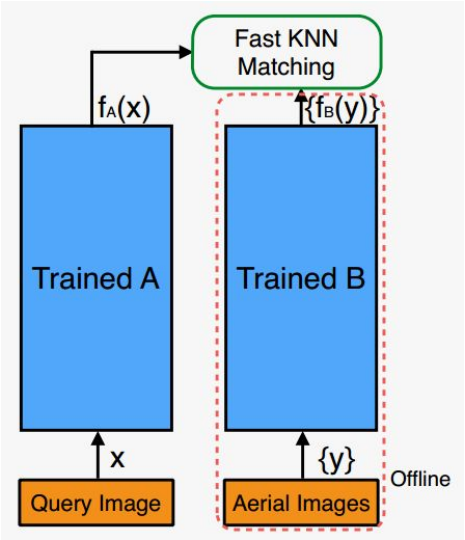


Testing

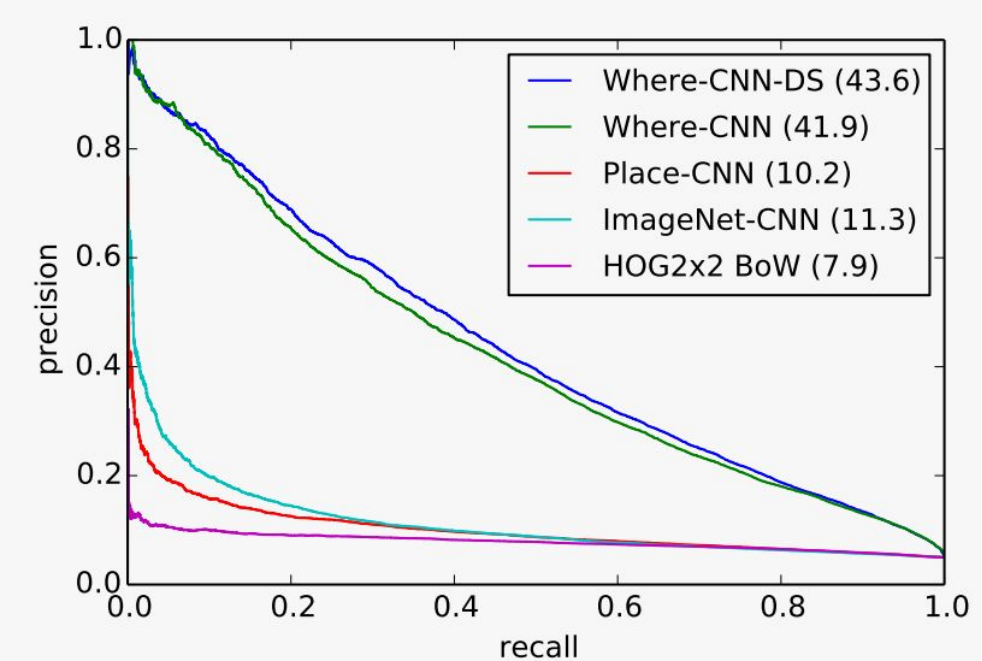
$$f_A(x), f_B(y) \in \mathbb{R}^d$$

$$d \ll n$$

$$x, y \in \mathbb{R}^n$$



Precision comparison with other features

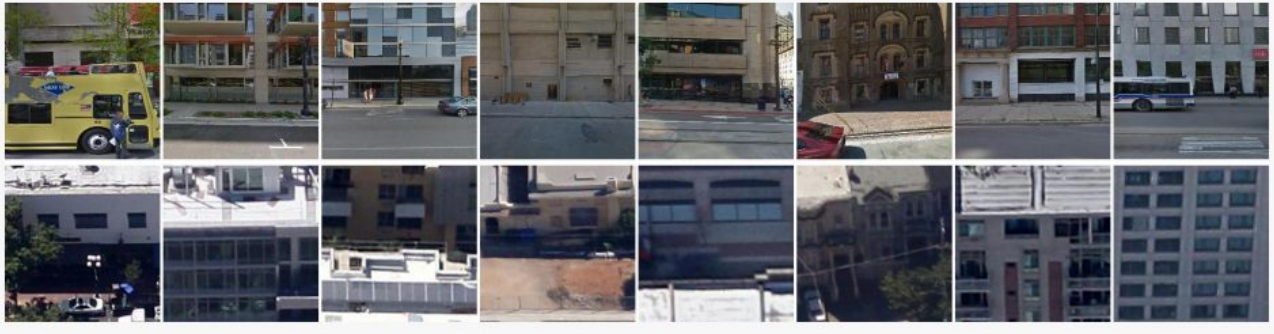


Comparison between different initialization

Where-CNN	ImageNet init.	Places init.
AP	41.9%	41.4%

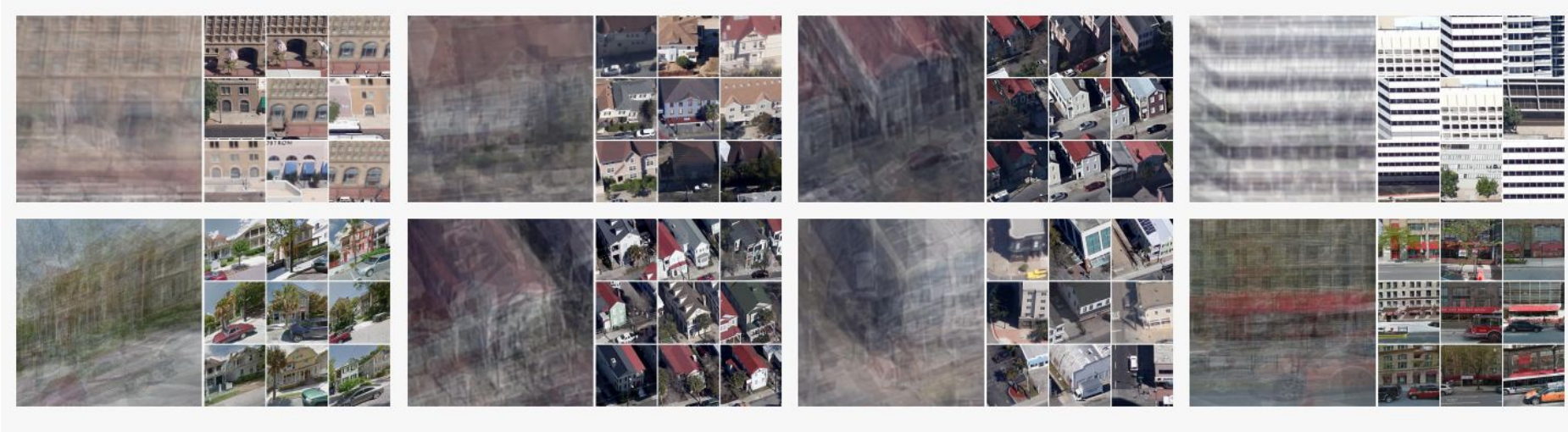
Output example

(a) Easy positive pairs.

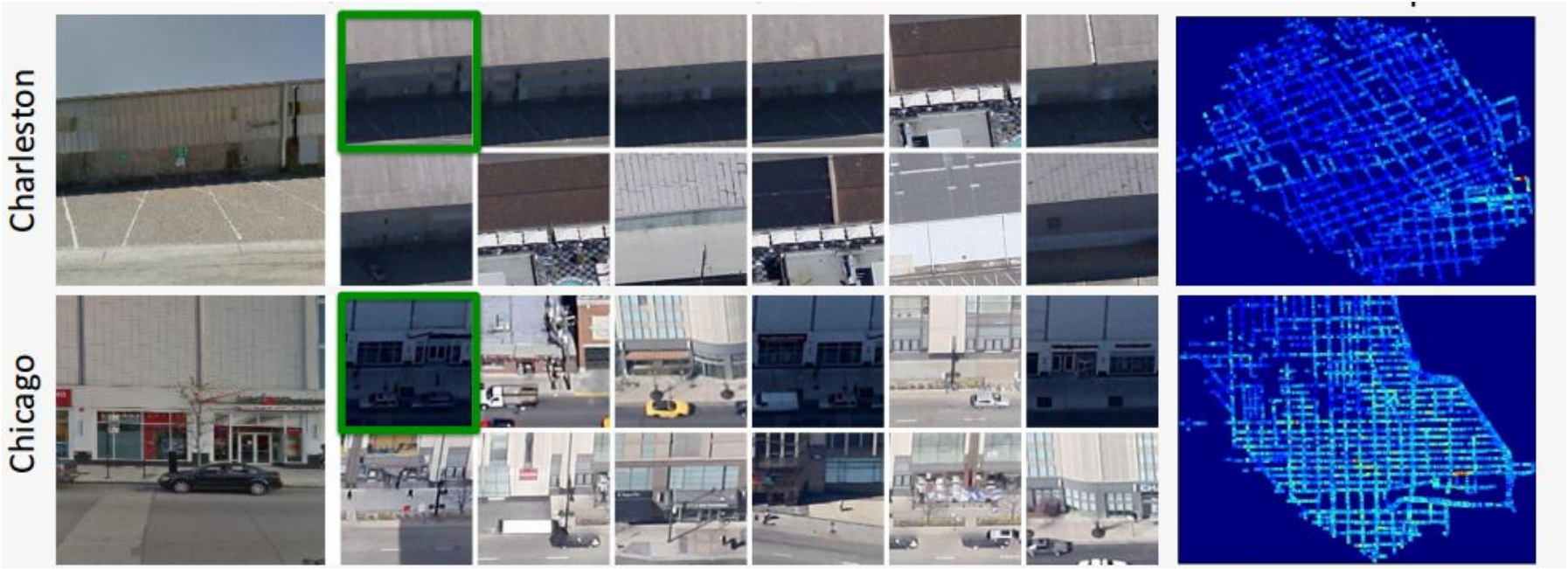


(b) Hard negative pairs.

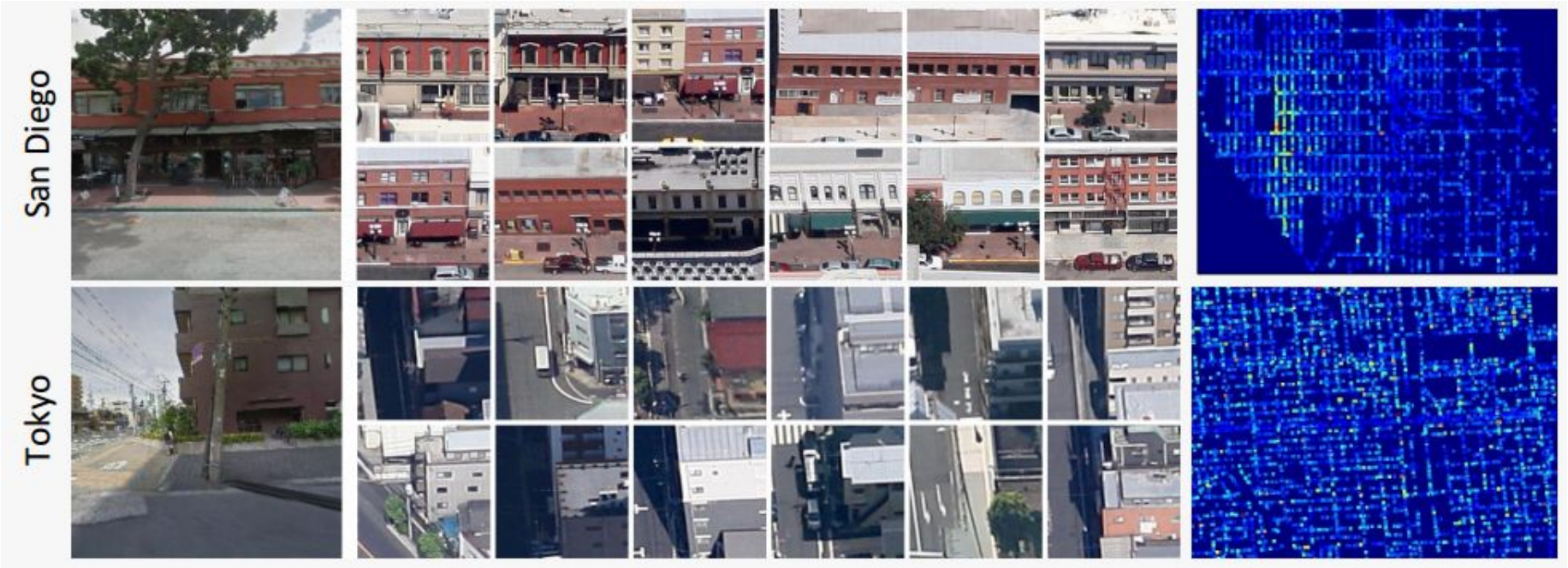
Strong activations



Possible matches and locations



Possible matches and locations



Geolocalization task