Learning Spherical Convolution for Fast Features from 360° Imagery

Yu-Chuan Su and Kristen Grauman
The University of Texas at Austin
http://vision.cs.utexas.edu/projects/sphconv

1. Motivation

Strategies for Convolution on 360° Images

Our goal is to apply a pre-trained convolutional neural network (CNN) on 360° images. Existing strategies are either slow or inaccurate.

- Fast – single evaluation
- Inaccurate – distortion
- Slow – repeated projection
- Accurate – exact result

2. Main Idea: Spherical Convolution

Learn a CNN on equirectangular projection that mimics the outputs of an existing target network would produce on tangent plane projections.

- Fast – evaluate SphConv ($N_e$) on single projection
- Accurate – mimic exact result

3. Approach

3.1 Adapt network architecture

**Motivation**

Distortion in equirectangular projection is altitude ($\theta$) dependent.

**Source**

- $\theta = 90°$
- $\theta = 54°$
- $\theta = 18°$

**Untie kernel weights along altitude ($\theta$)**

Adjust kernel size

- Goal – ensure enough information for reproducing outputs
- Criteria – the receptive field in equirectangular projection
- Shape – allow non-square kernels (distortion is directional)

3.2 Kernel-wise pre-training

Require the network to generate the same intermediate representations as the target network.

$$N_l^e(I_l[x, y]) \approx N_l^p(I_l[\theta, \phi]) \quad \forall l \in N_e$$

- Learn each layer using “ground truth” value of previous layer
- Fine-tune the network using the final objective

4. Results

**Learned kernels** – our learned kernels properly account for distortions

- $\phi = 0°$

**Detection accuracy**

- SphConv performs better than Strategy I especially near $\theta=0°$
- SphConv approaches accuracy of Strategy II, but is much faster

**Convolution output errors and computational cost**

- 50% more accurate than Strategy I under similar cost
- Orders of magnitude faster than Strategy II