Affinity CNN: Learning Pixel-Centric Pairwise Relations for Figure/Ground Embedding

Goal: Segmentation and figure/ground from a single image, without joint training

Overview

Complex Affinity Model

Segmentation and Figure/Ground via Embedding

Embedding:
- \( (p, q) \) is in effect
- \( p \) and \( q \) lie in the same region as \( r \)
- \( p \) is in more off-field than \( q \)
- \( C(p, q) \) is with \( r \) at a boundary
- \( p \) and \( q \) lie in different regions
- \( p \) in local area for relative displacements
- Low confidence \( (C(p, q) = 0) \)
- Positive displacement \( (p < q) \)
- Ground transition
- Negative displacement \( (p > q) \)

Angular Embedding [2]

Affinity Learning from Globalized Ground-truth

Annotated Ground-truth
- Human-drawn segmentation (boundary map)
- Figure-ground and image boundaries with globalized ground-truth

Affinity from Interactions:
- Context probabilities to consider:
  \( C(p, q) \) over \( r \) at \( (c, d) \)
  \( C(p, q) = \text{argmax} C(p, q) \)
  \( C(p, q) \) over \( r \) at \( (c, d) \)
  \( C(p, q) = \text{argmax} C(p, q) \)
- Apply contextual action:
  \( W(p, q) = C(p, q) + r \)
  \( W(p, q) = C(p, q) \)
  \( W(p, q) = W(p, q) + r \)
  \( W(p, q) = W(p, q) + r \)
- Sum transition forces and symmetries:
  \( W + W + W + W \)

Results and Evaluation on Berkeley Segmentation Dataset

Image

Cross-Domain Generalization

Learning:
- CNN output \( s_{c} \), \( s_{d} \) : maps per neighbor
- Mixed with soft labels \( s_{d} \), \( s_{d} \) (neighbors)
- SGD with log loss average-pixel-wise:
  \( L = \sum_{c} \sum_{d} \log(s_{c}(p, q)) \)
- Data augmentation: translate, left-right mirror

Michael Maire
TTI Chicago

Takuya Narihira
Sony Corp

Stella X. Yu
UC Berkeley / HKSI

Basic Relationship Types:
- Continuous region:
  \( p \) and \( q \) lie in the same region as \( r \)
- High confidence \( (C(p, q) = 1) \)
- Low confidence \( (C(p, q) = 0) \)
- Ambiguous boundary:
  \( p \) and \( q \) lie in different regions
  \( C(p, q) = 0 \)
- Positive displacement:
  \( p \) and \( q \) lie in different regions
  \( C(p, q) = 0 \)
- Ground transition:
  \( p \) and \( q \) lie in different regions
  \( C(p, q) = 0 \)

Pairwise Pixel Instructions:
- Probability estimates:
  \( C(p, q) = \text{argmax} C(p, q) \)
  \( C(p, q) = \text{argmax} C(p, q) \)
- Transition error probabilities:
  \( C(p, q) = \text{argmax} C(p, q) \)
  \( C(p, q) = \text{argmax} C(p, q) \)
  \( C(p, q) = \text{argmax} C(p, q) \)
  \( C(p, q) = \text{argmax} C(p, q) \)
  \( C(p, q) = \text{argmax} C(p, q) \)
  \( C(p, q) = \text{argmax} C(p, q) \)
  \( C(p, q) = \text{argmax} C(p, q) \)
  \( C(p, q) = \text{argmax} C(p, q) \)
  \( C(p, q) = \text{argmax} C(p, q) \)

Affinity from Ground:
- Context probabilities to consider:
  \( C(p, q) = \text{argmax} C(p, q) \)
  \( C(p, q) = \text{argmax} C(p, q) \)
  \( C(p, q) = \text{argmax} C(p, q) \)
  \( C(p, q) = \text{argmax} C(p, q) \)
  \( C(p, q) = \text{argmax} C(p, q) \)
  \( C(p, q) = \text{argmax} C(p, q) \)
  \( C(p, q) = \text{argmax} C(p, q) \)
  \( C(p, q) = \text{argmax} C(p, q) \)
  \( C(p, q) = \text{argmax} C(p, q) \)

Affinity from Ground:
- Context probabilities to consider:
  \( C(p, q) = \text{argmax} C(p, q) \)
  \( C(p, q) = \text{argmax} C(p, q) \)
  \( C(p, q) = \text{argmax} C(p, q) \)
  \( C(p, q) = \text{argmax} C(p, q) \)
  \( C(p, q) = \text{argmax} C(p, q) \)
  \( C(p, q) = \text{argmax} C(p, q) \)
  \( C(p, q) = \text{argmax} C(p, q) \)
  \( C(p, q) = \text{argmax} C(p, q) \)
  \( C(p, q) = \text{argmax} C(p, q) \)

M. Maire. Simultaneous Segmentation and Figure/Ground Organization using Angular Embedding. CVPR 2010

• \( k \) and \( q \) lie in the same region as \( r \)
• \( k \) is more off-field than \( q \)
• \( k \) is globally consistent representation of the pairwise local grouping/ordering relationships \( p, q \)
• \( r \) is globally consistent representation of the pairwise local grouping/ordering relationships \( p, q \)

Implementation:
- Learn to classify \( C(k, r) \)
- Sparse multiscale stroke each pixel connects to k neighbors

Image

Figure/Ground

Segmentation