Object Detection and Segmentation from Joint Embedding of Parts and Pixels

Michael Maire\textsuperscript{1}, Stella X. Yu\textsuperscript{2}, Pietro Perona\textsuperscript{1}

\textsuperscript{1}California Institute of Technology - Pasadena, CA 91125
\textsuperscript{2}Boston College - Chestnut Hill, MA 02467
Segmentation

Detection
Segmentation

Detection

Perceptual Grouping Framework
Ingredients

Plug in state-of-the-art components:
Ingredients

Plug in state-of-the-art components:

low-level cues:
- color
- texture
- edges

[Arbeláez, Maire, Fowlkes, Malik, PAMI 2011]
Ingredients

Plug in state-of-the-art components:

low-level cues: color, texture, edges

[Arbeláez, Maire, Fowlkes, Malik, PAMI 2011]

top-down parts: poselets for person detection

[Bourdev, Maji, Brox, Malik, ECCV 2010]
Ingredients

Plug in state-of-the-art components:

**PASCAL VOC 2010 Person Category:** Improved Detection and Segmentation

- **low-level cues:** color, texture, edges

  [Arbeláez, Maire, Fowlkes, Malik, PAMI 2011]

- **top-down parts:** poselets for person detection

  [Bourdev, Maji, Brox, Malik, ECCV 2010]
Grouping Relationships
Grouping Relationships
Pixel Affinity: Color, Texture Similarity
Pixel Affinity: Color, Texture Similarity
Pixel Affinity: Color, Texture Similarity
Part Affinity: Geometric Compatibility
Part Affinity: Geometric Compatibility
parts

surround

pixels
Angular Embedding

- parts
- surround
- pixels

⇒

⇒

⇒

⇒

objects

figure/ground

segmentation
Angular Embedding

Given:

- Relative ordering $\Theta(\cdot, \cdot)$
- Confidence on relationships $C(\cdot, \cdot)$

Compute:

- Global ordering $\theta(\cdot)$
- Embed into unit circle: $p \rightarrow z(p) = e^{i \theta(p)}$

Subject to:

- Linear constraints on embedding solution in columns of $U$
Angular Embedding

Given:
- Relative ordering $\Theta(\cdot, \cdot)$
- Confidence on relationships $C(\cdot, \cdot)$

Compute:
- Global ordering $\theta(\cdot)$
- Embed into unit circle: $p \rightarrow z(p) = e^{i\theta(p)}$

Subject to:
- Linear constraints on embedding solution in columns of $U$
Angular Embedding

Given:
- Relative ordering $\Theta(\cdot, \cdot)$
- Confidence on relationships $C(\cdot, \cdot)$

Compute:
- Global ordering $\theta(\cdot)$
- Embed into unit circle: $p \rightarrow z(p) = e^{i\theta(p)}$

Subject to:
- Linear constraints on embedding solution in columns of $U$
Angular Embedding

Given:

- Relative ordering $\Theta(\cdot, \cdot)$
- Confidence on relationships $C(\cdot, \cdot)$

Compute:

- Global ordering $\theta(\cdot)$
- Embed into unit circle: $p \rightarrow z(p) = e^{i \theta(p)}$

Subject to:

- Linear constraints on embedding solution in columns of $U$
Angular Embedding

Given:

- Relative ordering $\Theta(\cdot, \cdot)$
- Confidence on relationships $C(\cdot, \cdot)$

Compute:

- Global ordering $\theta(\cdot)$
- Embed into unit circle:

$$p \rightarrow z(p) = e^{i\theta(p)}$$
Angular Embedding

Given:
- Relative ordering $\Theta(\cdot, \cdot)$
- Confidence on relationships $C(\cdot, \cdot)$

Compute:
- Global ordering $\theta(\cdot)$
- Embed into unit circle:

$$p \rightarrow z(p) = e^{i\theta(p)}$$

Subject to:
- Linear constraints on embedding solution in columns of $U$
minimize: $\epsilon = \sum_{p} \sum_{q} C(p, q) \cdot \left| z(p) - \tilde{z}(p) \right|^2$

[Yu, PAMI 2011]
\[
\minimize \varepsilon = \sum_p \sum_q C(p, q) C(p, r) \cdot \left| z(p) - \tilde{z}(p) \right|^2
\]

[Yu, PAMI 2011]
minimize: $\varepsilon = \sum_p \frac{\sum_q C(p,q)}{\sum_{p,q} C(p,q)} \cdot |z(p) - \tilde{z}(p)|^2$

[Yu, PAMI 2011]
$U$  

$C_q$  

$(C_s, \Theta_s)$  

$(C_f, \Theta_f)$  

$C_p$
\[ C = \begin{bmatrix}
  C_p & 0 & 0 & 0 \\
  0 & \alpha \cdot C_q & \beta \cdot C_s & \gamma \cdot C_f \\
  0 & \beta \cdot C_s^T & 0 & 0 \\
  0 & \gamma \cdot C_f^T & 0 & 0 \\
\end{bmatrix} \]

\[ \Theta = \sum^{-1} \begin{bmatrix}
  0 & 0 & 0 & 0 \\
  0 & 0 & \Theta_s & \Theta_f \\
  0 & -\Theta_s^T & 0 & 0 \\
  0 & -\Theta_f^T & 0 & 0 \\
\end{bmatrix} \]
Angular Embedding

Relax to generalized eigenproblem \( QPQz = \lambda z \):

\[
P = D^{-1}W \\
Q = I - D^{-1}U(U^T D^{-1} U)^{-1} U^T
\]

with \( D \) and \( W \) defined as:

\[
D = \text{Diag}(C_{1n}) \\
W = C \cdot e^{i\Theta}
\]

Eigenvectors \( \{z_0, z_1, \ldots, z_{m-1}\} \) embed pixels and parts into \( \mathbb{C}^m \).
Angular Embedding

$\angle z_0$ encodes global ordering

$z_1, z_2, \ldots, z_{m-1}$ encode grouping
Angular Embedding

\[ \angle z_0 \text{ encodes global ordering} \]

\[ z_1, z_2, \ldots, z_{m-1} \text{ encode grouping} \]

if \( \Theta = 0 \) \implies \text{Normalized Cuts}

\( \text{(grouping without ordering)} \)
Decoding Eigenvectors: Object Detection
Decoding Eigenvectors: Object Detection

\[ \mathbb{R}(z_0) \quad \mathbb{R}(z_1) \quad \mathbb{R}(z_2) \]

\[ \mathbb{I}(z_0) \quad \mathbb{I}(z_1) \quad \mathbb{I}(z_2) \]
Decoding Eigenvectors: Object Detection

\[ \Re(z_0), \Re(z_1), \Re(z_2), \Im(z_0), \Im(z_1), \Im(z_2) \]
Decoding Eigenvectors: Object Detection

\[ \mathbb{R}(z_0) \mathbb{R}(z_1) \mathbb{R}(z_2) \mathbb{I}(z_0) \mathbb{I}(z_1) \mathbb{I}(z_2) \]

Ordering Grouping
Decoding Eigenvectors: Object Detection

- $\Re(z_0)$
- $\Re(z_1)$
- $\Re(z_2)$
- $\Im(z_0)$
- $\Im(z_1)$
- $\Im(z_2)$

Ordering Grouping
Decoding Eigenvectors: Object Detection

\[ \Re(z_0) \quad \Re(z_1) \quad \Re(z_2) \]
\[ \Im(z_0) \quad \Im(z_1) \quad \Im(z_2) \]

Ordering

Grouping
Decoding Eigenvectors: Object Detection

Ordering Grouping
Decoding Eigenvectors: Object Detection

Ordering

Grouping

ℜ(z_0) ℜ(z_1) ℜ(z_2)
ℑ(z_0) ℑ(z_1) ℑ(z_2)
Decoding Eigenvectors: Figure/Ground

\[ \Re(z) \]

\[ \Im(z) \]

\[ z_0 \]

\[ z_1 \]

\[ z_2 \]

\[ z_3 \]

\[ z_4 \]
Decoding Eigenvectors: Figure/Ground

\[\Re(z)\]

\[\Im(z)\]

\[z_0 \quad z_1 \quad z_2 \quad z_3 \quad z_4\]

\[\angle z_0 \quad \nabla z_1 \quad \nabla z_2 \quad \nabla z_3 \quad \nabla z_4\]
Decoding Eigenvectors: Segmentation

Figure/Ground

Hierarchical Segmentation
[Arbeláez, Maire, Fowlkes, Malik, PAMI 2011]
Decoding Eigenvectors: Object Segmentation

Assign pixels $p_k$ to objects $Q_i$ via parts $q_j$:

$$p_k \rightarrow \arg\min_{Q_i} \left\{ \min_{q_j \in Q_i} \{ \text{Dist}(p_k, q_j) \} \right\}$$
Decoding Eigenvectors: Object Segmentation

Assign pixels $p_k$ to objects $Q_i$ via parts $q_j$:

$$p_k \rightarrow \arg \min_{Q_i} \left\{ \min_{q_j \in Q_i} \{ \text{Dist}(p_k, q_j) \} \right\}$$
Decoding Eigenvectors
Results: PASCAL 2010 Person Category

Detections  Poselet Mask  F/G Mask  Segmentation
Results: PASCAL 2010 Person Category

Detections  Poselet Mask  F/G Mask  Segmentation
Results: PASCAL 2010 Person Category

- Segmentation task score: 41.1 (35.5 for poselet baseline)
Results: PASCAL 2010 Person Category

- Segmentation task score: 41.1 (35.5 for poselet baseline)
- 11% relative improvement due to better detection
Summary

- Simultaneous segmentation and detection:
  - Part detectors $\rightarrow$ figure pop-out, object grouping
  - Color, texture $\rightarrow$ pixel grouping
Summary

- Simultaneous segmentation and detection:
  - Part detectors → figure pop-out, object grouping
  - Color, texture → pixel grouping

- Graph:
  - Parts and pixels as nodes
  - Links encode multiple relationship types

Better person detection and segmentation on PASCAL
Summary

- Simultaneous segmentation and detection:
  - Part detectors → figure pop-out, object grouping
  - Color, texture → pixel grouping

- Graph:
  - Parts and pixels as nodes
  - Links encode multiple relationship types

- Embedding: graph nodes $\rightarrow \mathbb{C}^m$
Summary

- Simultaneous segmentation and detection:
  - Part detectors → figure pop-out, object grouping
  - Color, texture → pixel grouping
- Graph:
  - Parts and pixels as nodes
  - Links encode multiple relationship types
- Embedding: graph nodes → $\mathbb{C}^m$
- Decode:
  - Figure/ground
  - Image segmentation
  - Detected objects
  - Segmentation of each object instance

Better person detection and segmentation on PASCAL
Summary

- Simultaneous segmentation and detection:
  - Part detectors → figure pop-out, object grouping
  - Color, texture → pixel grouping

- Graph:
  - Parts and pixels as nodes
  - Links encode multiple relationship types

- Embedding: graph nodes $\rightarrow \mathbb{C}^m$

- Decode:
  - Figure/ground
  - Image segmentation
  - Detected objects
  - Segmentation of each object instance

- Better person detection and segmentation on PASCAL