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

Michael Maire
TTI Chicago

Takuya Narihira
Sony Corp.

Stella X. Yu
UC Berkeley / ICSI

arXiv:1512.02767 & CVPR 2016

Segmentation and Figure/Ground Results

Figure

Ground

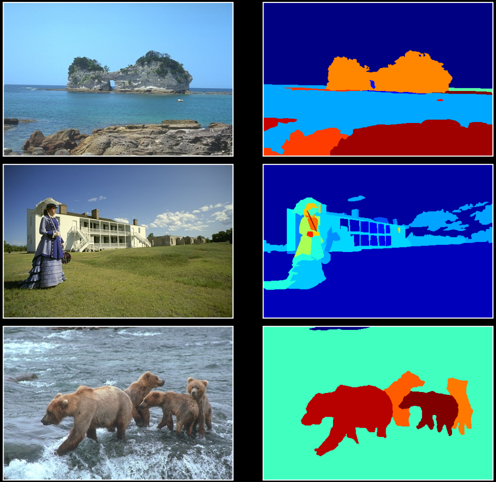


Image Ground-truth

Segmentation and Figure/Ground Results

Figure  Ground

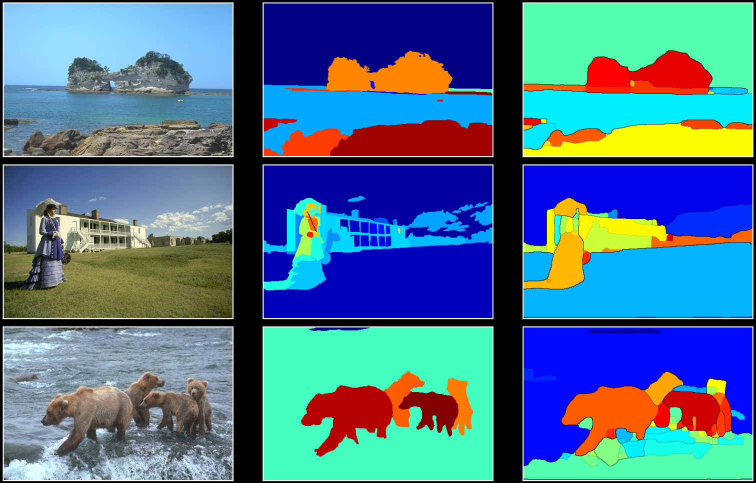


Image | Ground-truth | Our System

Segmentation and Figure/Ground via Embedding

Image

Segmentation and Figure/Ground via Embedding

Image

$\mathbb{R}$

Segmentation and Figure/Ground via Embedding

Image

$\mathbb{R}$

Segmentation and Figure/Ground via Embedding

Image

$\mathbb{R}$

Segmentation and Figure/Ground via Embedding

Image

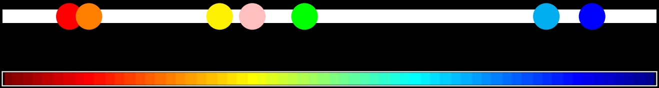
$\mathbb{R}$

Segmentation and Figure/Ground via Embedding

Image

Segmentation and Figure/Ground via Embedding

Image



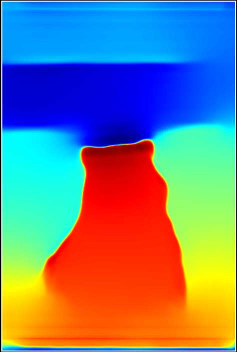
$\mathbb{R}$

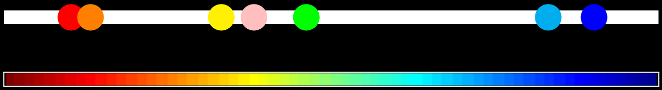
Figure Ground

Segmentation and Figure/Ground via Embedding

Image

Figure/Ground





$\mathbb{R}$

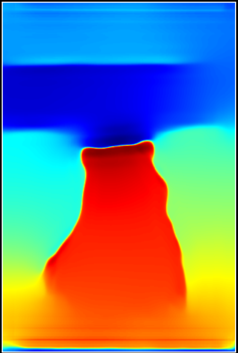
Figure

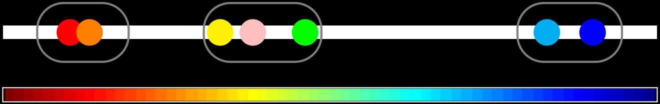
Ground

Segmentation and Figure/Ground via Embedding

Image

Figure/Ground





$\mathbb{R}$

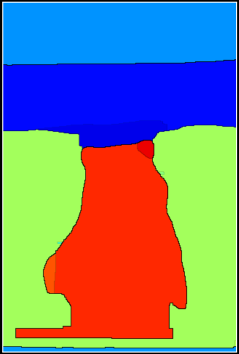
Figure

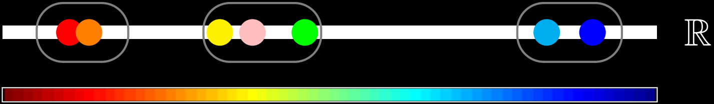
Ground

Segmentation and Figure/Ground via Embedding

Image

Figure/Ground



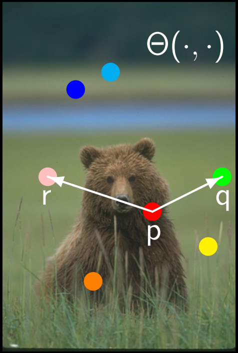


Figure

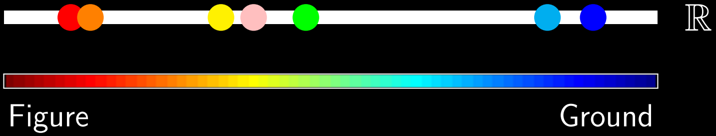
Ground

Embedding Objective: Satisfy Pairwise Affinities

Image

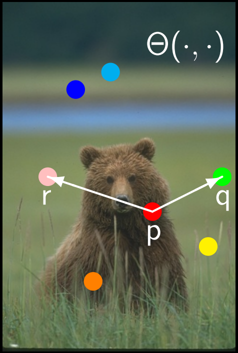


- $\Theta(\cdot,\cdot)$ stores pairwise relationships



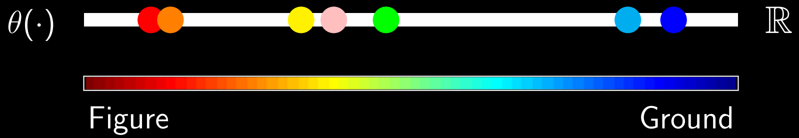
Embedding Objective: Satisfy Pairwise Affinities

Image



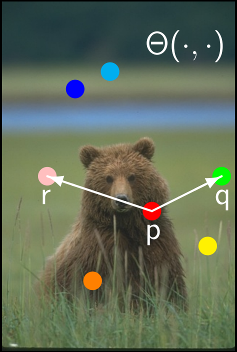
- $\Theta(\cdot,\cdot)$ stores pairwise relationships
- Embedding procedure:
  - Input: $\{\Theta(p,q), \Theta(p,r), \dots\}$
  - Output: $\{\theta(p), \theta(q), \theta(r), \dots\}$

  such that $\Theta(p,q) \approx F(\theta(p), \theta(r))$



Embedding Objective: Satisfy Pairwise Affinities

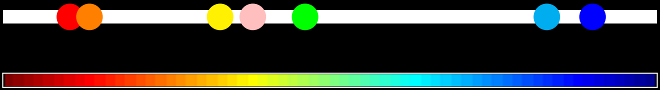
Image



- $\Theta(\cdot,\cdot)$ stores pairwise relationships
- Embedding procedure:
  - Input: $\{\Theta(p,q), \Theta(p,r), \ldots\}$
  - Output: $\{\theta(p), \theta(q), \theta(r), \ldots\}$

  such that $\Theta(p,q) \approx F(\theta(p), \theta(r))$
- $F(\cdot,\cdot)$ is a simple decoding function

$\theta(\cdot)$



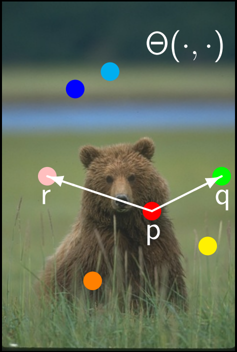
$\mathbb{R}$

Figure

Ground

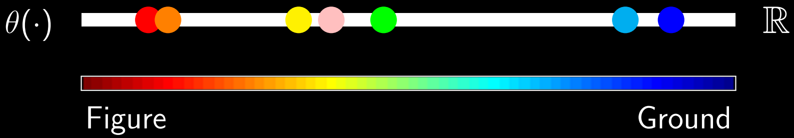
Embedding Objective: Satisfy Pairwise Affinities

Image



- $\Theta(\cdot,\cdot)$ stores pairwise relationships
- Embedding procedure:
  - Input: $\{\Theta(p,q), \Theta(p,r), \ldots\}$
  - Output: $\{\theta(p), \theta(q), \theta(r), \ldots\}$

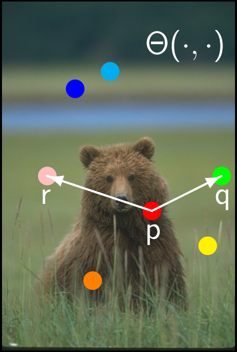
  such that $\Theta(p,q) \approx F(\theta(p), \theta(r))$
- $F(\cdot,\cdot)$ is a simple decoding function
- Embedding $\theta(\cdot)$ is a globally consistent representation of the pairwise local relationships $\Theta(\cdot,\cdot)$

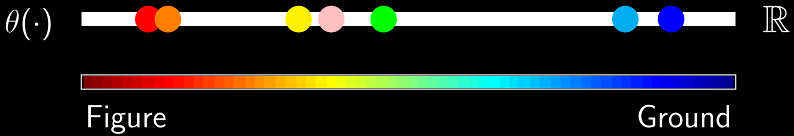


Embedding Objective: Satisfy Pairwise Affinities

Segmentation+Figure/Ground

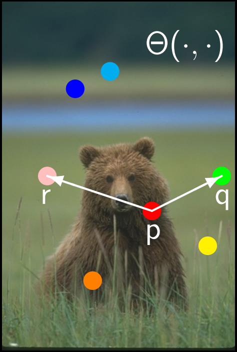
Image





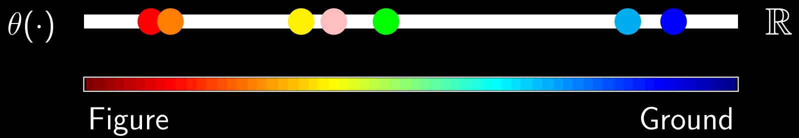
Embedding Objective: Satisfy Pairwise Affinities

Image



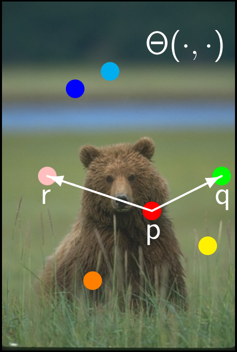
Segmentation+Figure/Ground

- $\Theta(p, q)$ is an offset
  - Is p in the same region as q?
  - Is p in front of/behind q?



Embedding Objective: Satisfy Pairwise Affinities

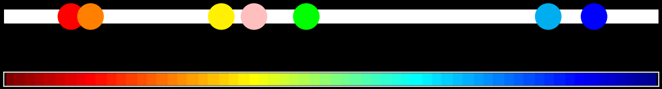
Image



Segmentation+Figure/Ground

- $\Theta(p,q)$ is an offset
  - Is p in the same region as q?
  - Is p in front of/behind q?
- $C(p,q)$: confidence on $\Theta(p,q)$

$\theta(\cdot)$



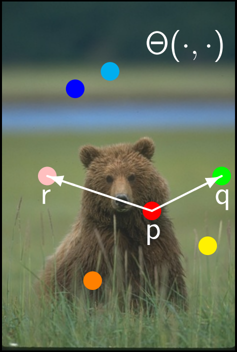
$\mathbb{R}$

Figure

Ground

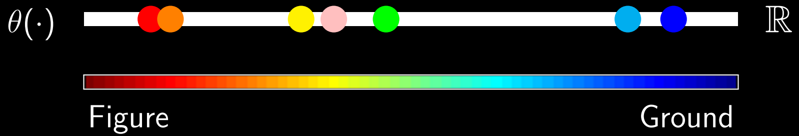
Embedding Objective: Satisfy Pairwise Affinities

Image



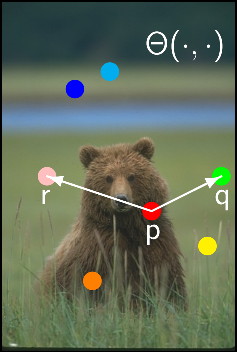
Segmentation+Figure/Ground

- $\Theta(p, q)$ is an offset
  - Is p in the same region as q?
  - Is p in front of/behind q?
- $C(p, q)$: confidence on $\Theta(p, q)$
- Angular Embedding: $(C, \Theta) \to \theta$



Embedding Objective: Satisfy Pairwise Affinities

Image

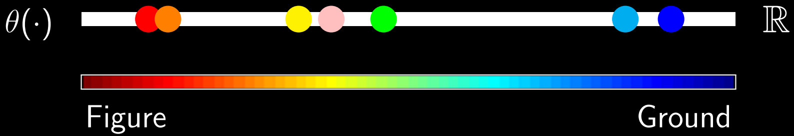


Segmentation+Figure/Ground

- $\Theta(p, q)$ is an offset
  - Is p in the same region as q?
  - Is p in front of/behind q?
- $C(p, q)$: confidence on $\Theta(p, q)$
- Angular Embedding: $(C, \Theta) \rightarrow \theta$

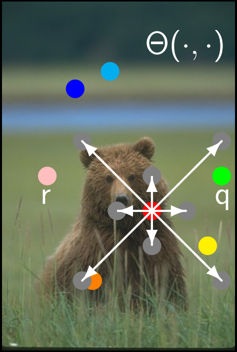
Implementation

- CNN estimates $\Theta(p, q)$, $C(p, q)$



Embedding Objective: Satisfy Pairwise Affinities

Image



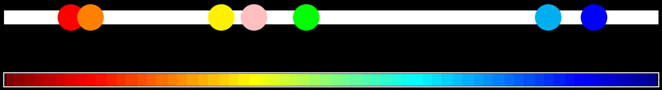
Segmentation+Figure/Ground

- $\Theta(p, q)$ is an offset
  - Is p in the same region as q?
  - Is p in front of/behind q?
- $C(p, q)$: confidence on $\Theta(p, q)$
- Angular Embedding: $(C, \Theta) \to \theta$

Implementation

- CNN estimates $\Theta(p, q)$, $C(p, q)$
- Sparse multiscale connectivity:
  each p connects to k neighbors

$\theta(\cdot)$



$\mathbb{R}$

Figure

Ground

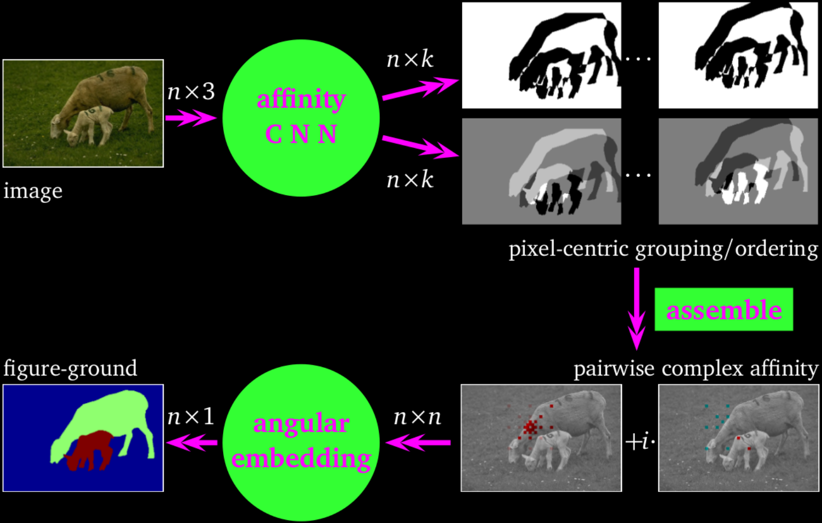


image
$n\times3$
affinity CNN
$n\times k$
$n\times k$
pixel-centric grouping/ordering
assemble
pairwise complex affinity
$+i\cdot$
$n\times n$
angular embedding
$n\times1$
figure-ground

Angular Embedding

Given Pairwise:

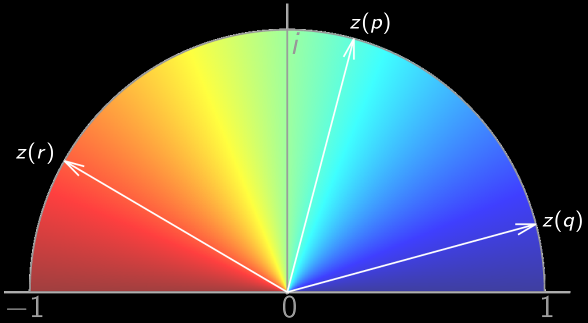
Ordering $\Theta(\cdot,\cdot)$

Confidence $C(\cdot,\cdot)$

Recover:

Global ordering $\theta(p)$

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



[Yu, PAMI 2012]

Angular Embedding

Given Pairwise:

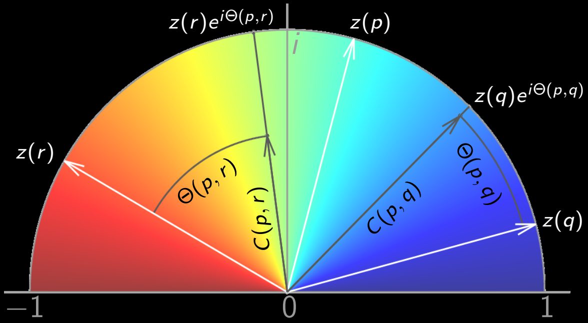
Ordering $\Theta(\cdot,\cdot)$

Confidence $C(\cdot,\cdot)$

Recover:

Global ordering $\theta(p)$

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



[Yu, PAMI 2012]

Angular Embedding

Given Pairwise:

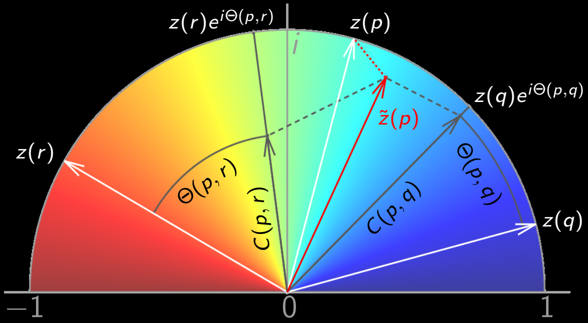
Ordering $\Theta(\cdot,\cdot)$

Confidence $C(\cdot,\cdot)$

Recover:

Global ordering $\theta(p)$

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



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 2012]

Angular Embedding

Given Pairwise:

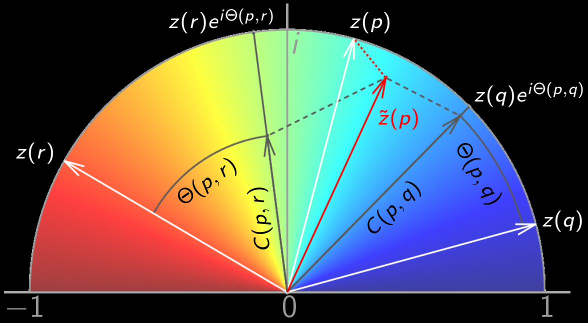
Ordering $\Theta(\cdot,\cdot)$

Confidence $C(\cdot,\cdot)$

Recover:

Global ordering $\theta(p)$

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



Relax to generalized eigenproblem $Wz = \lambda Dz$ where:

$$D = \mathrm{Diag}(C1_n)$$

$$W = C \bullet e^{i\Theta}$$

Angular Embedding

Given Pairwise:

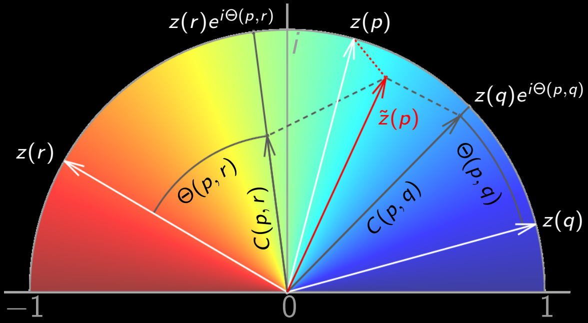
Ordering $\Theta(\cdot,\cdot)$

Confidence $C(\cdot,\cdot)$

Recover:

Global ordering $\theta(p)$

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

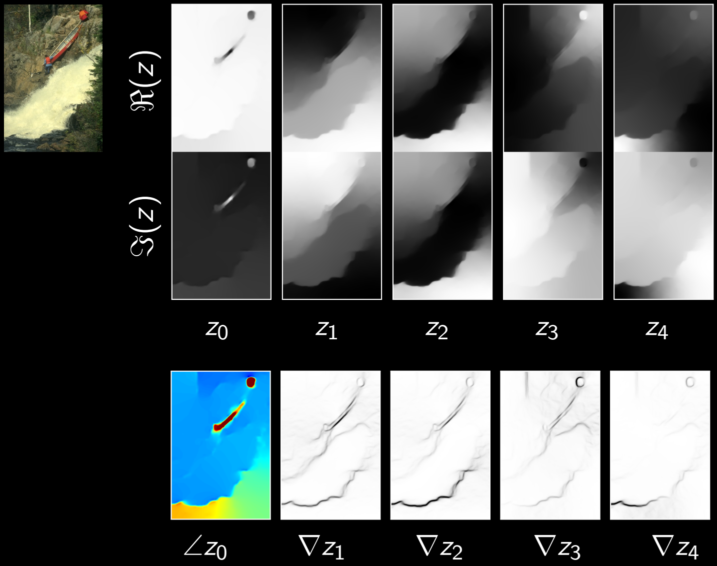


Relax to generalized eigenproblem $Wz = \lambda Dz$ where:

$$D = \mathrm{Diag}(C1_n)$$
$$W = C \bullet e^{i\Theta}$$

Normalized Cuts is the special case: $\Theta = 0$

Eigenvectors



Angular Embedding

Given Pairwise:

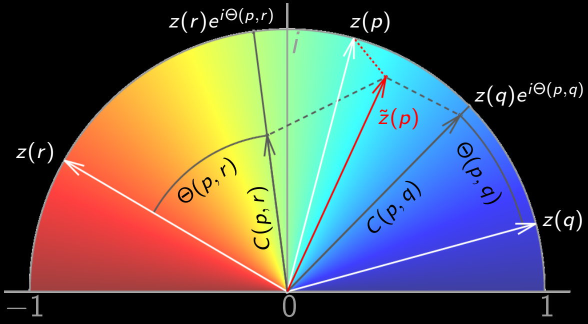
Ordering $\Theta(\cdot,\cdot)$

Confidence $C(\cdot,\cdot)$

Recover:

Global ordering $\theta(p)$

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

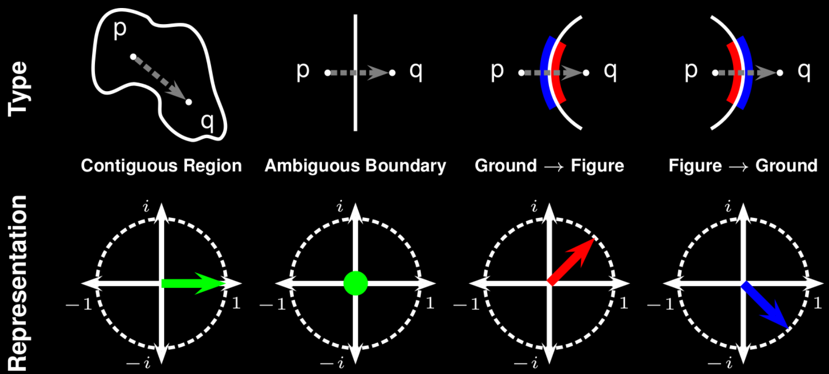


Squishing trick: rescale Θ by

$$\frac{\pi}{2}(1_n^T|\Theta|1_n)^{-1}$$

prior to embedding

Complex-Valued Affinities



Pairwise Pixel Interaction Model

Pairwise Pixel Interaction Model

- Probability estimates:

$$
\begin{aligned}
e(p) &= Pr(p \text{ lies on a boundary}) \\
b(p, q) &= Pr(\mathrm{seg}(p) \neq \mathrm{seg}(q)) \\
f(p, q) &= Pr(\mathrm{figural}(p, q) \mid \mathrm{seg}(p) \neq \mathrm{seg}(q)) \\
g(p, q) &= Pr(\mathrm{figural}(q, p) \mid \mathrm{seg}(p) \neq \mathrm{seg}(q))
\end{aligned}
$$

Pairwise Pixel Interaction Model

- Probability estimates:

$$
\begin{aligned}
e(p) &= Pr(p \text{ lies on a boundary}) \\
b(p,q) &= Pr(\mathrm{seg}(p) \neq \mathrm{seg}(q)) \\
f(p,q) &= Pr(\mathrm{figural}(p,q) \mid \mathrm{seg}(p) \neq \mathrm{seg}(q))) \\
g(p,q) &= Pr(\mathrm{figural}(q,p) \mid \mathrm{seg}(p) \neq \mathrm{seg}(q)))
\end{aligned}
$$

- Transition error probabilities:

$$
\begin{aligned}
&E_B(p,q) = b(p,q) \\
&E_F(p,q) = 1 - (1 - e(p))b(p,q)(1 - e(q))f(p,q) \\
&E_G(p,q) = 1 - (1 - e(p))b(p,q)(1 - e(q))g(p,q)
\end{aligned}
$$

B: binding F: ground $\rightarrow$ figure G: figure $\rightarrow$ ground

Pairwise Pixel Interaction Model

- Probability estimates:

$$
\begin{aligned}
e(p) &= Pr(p \text{ lies on a boundary}) \\
b(p,q) &= Pr(\mathrm{seg}(p) \neq \mathrm{seg}(q)) \\
f(p,q) &= Pr(\mathrm{figural}(p,q) \mid \mathrm{seg}(p) \neq \mathrm{seg}(q))) \\
g(p,q) &= Pr(\mathrm{figural}(q,p) \mid \mathrm{seg}(p) \neq \mathrm{seg}(q)))
\end{aligned}
$$

- Transition error probabilities:

$$
\begin{aligned}
E_B(p,q) &= b(p,q) \\
E_F(p,q) &= 1 - (1 - e(p))b(p,q)(1 - e(q))f(p,q) \\
E_G(p,q) &= 1 - (1 - e(p))b(p,q)(1 - e(q))g(p,q)
\end{aligned}
$$

B: binding F: ground $\rightarrow$ figure G: figure $\rightarrow$ ground

Pairwise Pixel Interaction Model

- Probability estimates:

$$
\begin{aligned}
e(p) &= Pr(p \text{ lies on a boundary}) \\
b(p,q) &= Pr(\mathrm{seg}(p) \neq \mathrm{seg}(q)) \\
f(p,q) &= Pr(\mathrm{figural}(p,q) \mid \mathrm{seg}(p) \neq \mathrm{seg}(q)) \\
g(p,q) &= Pr(\mathrm{figural}(q,p) \mid \mathrm{seg}(p) \neq \mathrm{seg}(q))
\end{aligned}
$$

- Transition error probabilities:

$$
\begin{aligned}
E_B(p,q) &= b(p,q) \\
E_F(p,q) &= 1 - (1 - e(p))b(p,q)(1 - e(q))f(p,q) \\
E_G(p,q) &= 1 - (1 - e(p))b(p,q)(1 - e(q))g(p,q)
\end{aligned}
$$

B: binding **F**: ground $\to$ figure **G**: figure $\to$ ground

Pairwise Pixel Interaction Model

- Convert error probabilities to confidence :

$$
\begin{aligned}
C_B(p,q) &= \exp(-E_B(p,q)/\sigma_b) \\
C_F(p,q) &= \exp(-E_F(p,q)/\sigma_f) \\
C_G(p,q) &= \exp(-E_G(p,q)/\sigma_g)
\end{aligned}
$$

Pairwise Pixel Interaction Model

- Convert error probabilities to confidence :

$$
\begin{aligned}
C_B(p,q) &= \exp(-E_B(p,q)/\sigma_b) \\
C_F(p,q) &= \exp(-E_F(p,q)/\sigma_f) \\
C_G(p,q) &= \exp(-E_G(p,q)/\sigma_g)
\end{aligned}
$$

- Apply rotational action of figure/ground transitions:

$$
\begin{aligned}
W_B(p,q) &= C_B(p,q) \\
W_F(p,q) &= C_F(p,q)\exp(i\phi) \\
W_G(p,q) &= C_G(p,q)\exp(-i\phi)
\end{aligned}
$$

Pairwise Pixel Interaction Model

- Convert error probabilities to confidence :

$$C_B(p,q) = \exp(-E_B(p,q)/\sigma_b)$$
$$C_F(p,q) = \exp(-E_F(p,q)/\sigma_f)$$
$$C_G(p,q) = \exp(-E_G(p,q)/\sigma_g)$$

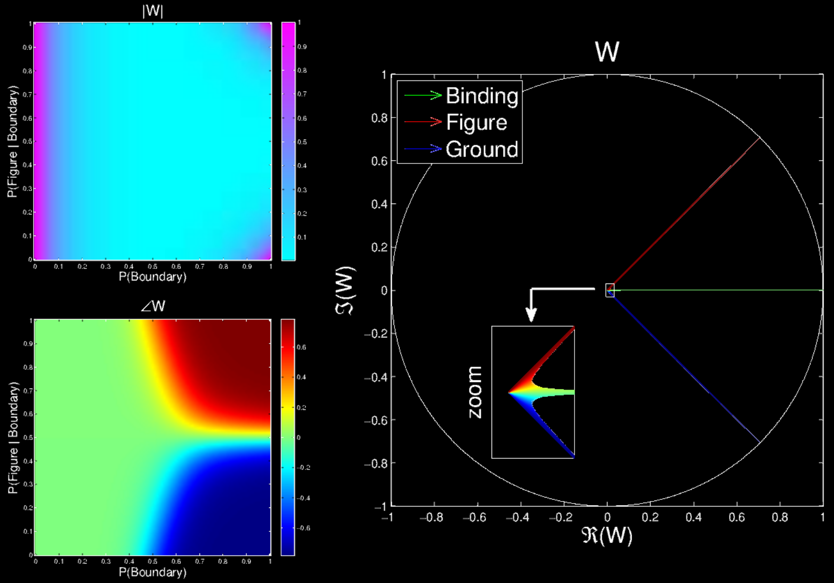
- Apply rotational action of figure/ground transitions:

$$W_B(p,q) = C_B(p,q)$$
$$W_F(p,q) = C_F(p,q)\exp(i\phi)$$
$$W_G(p,q) = C_G(p,q)\exp(-i\phi)$$

- Sum transition forces and symmetrize:

$$W(p,q) = W_B(p,q) + W_F(p,q) + W_G(p,q)$$
$$W \leftarrow (W + W^*)/2$$

Generalized Affinity



Affinity Learning

Ground-truth: Berkeley Segmentation Dataset

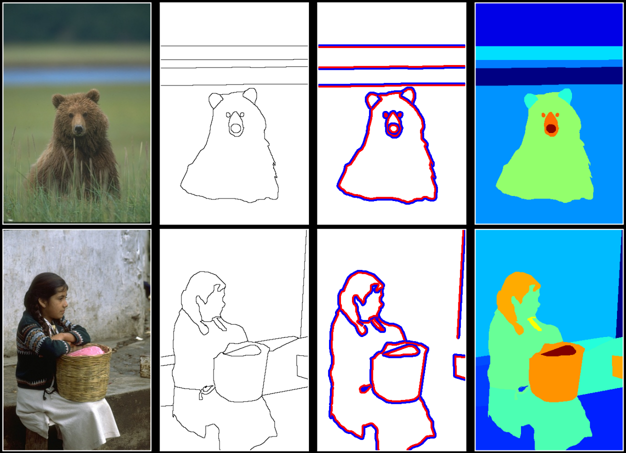
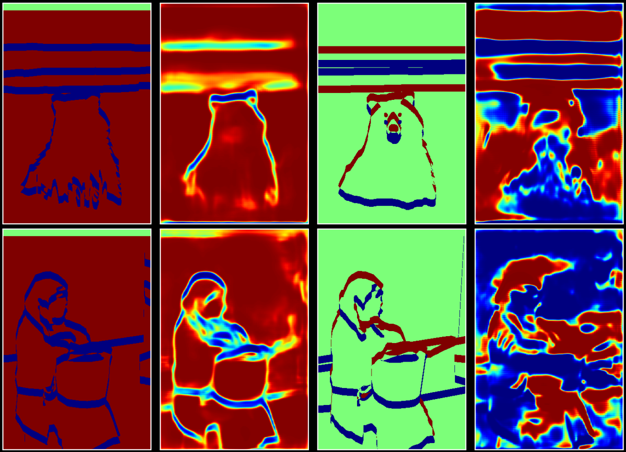


Image **Segmentation** **Local F/G** **Globalized F/G**

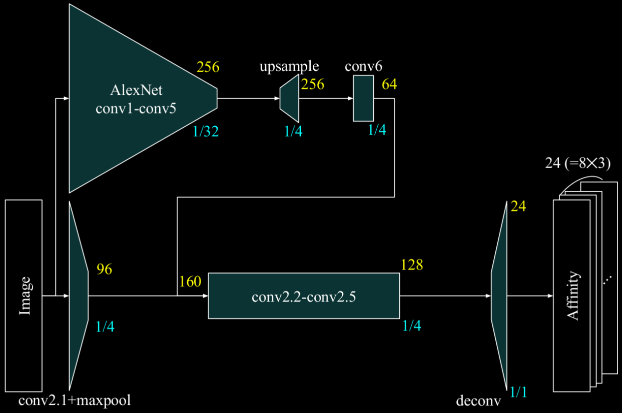
Affinity Learning

Target $(\tilde{b}, \tilde{f})$ and learned (b, f) signals



$\tilde{b}(p, p+d)$ $b(p, p+d)$ $\tilde{f}(p, p+d)$ $f(p, p+d)$

Affinity Learning - CNN Architecture



- Trained using log loss between target and prediction
- Left/right mirroring of examples

Results Comparison

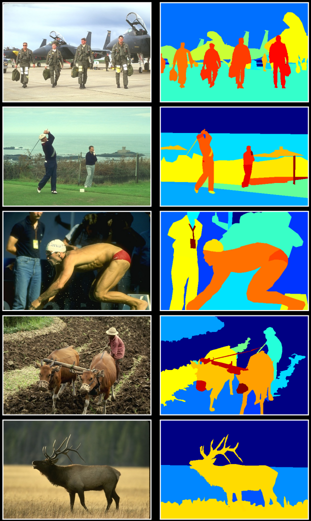


Image **Ground-truth F/G**

Results Comparison

Image | Ground-truth F/G | F/G: [Maire, ECCV 2010]

Results Comparison

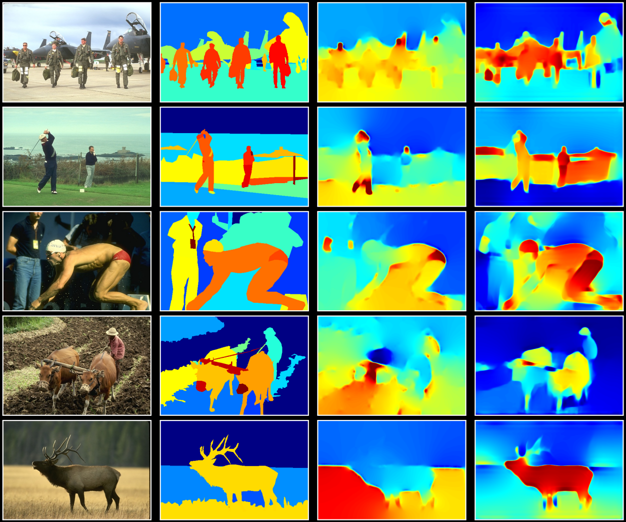


Image | Ground-truth F/G | F/G: [Maire, ECCV 2010] | F/G: Ours

Benchmark: Project onto Ground-truth Regions

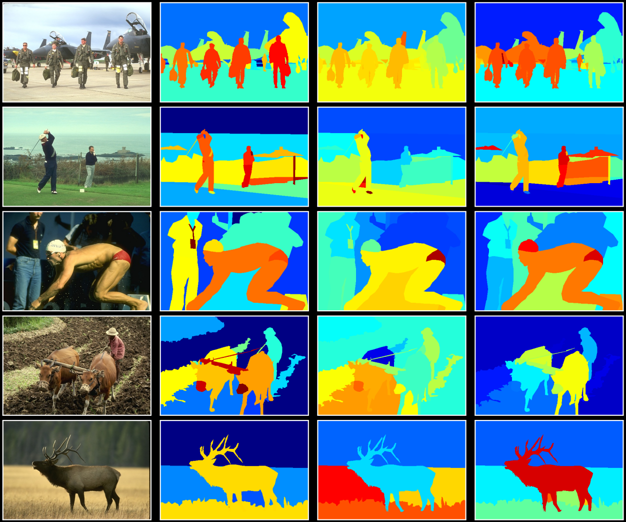


Image | Ground-truth F/G | F/G: [Maire, ECCV 2010] | F/G: Ours

Benchmark: Boundary Ownership Correctness

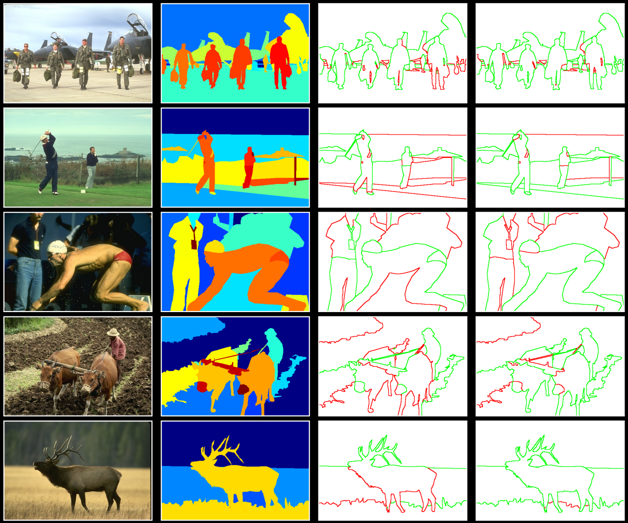


Image | Ground-truth F/G | F/G: [Maire, ECCV 2010] | F/G: Ours

Benchmark: Project onto Our Regions

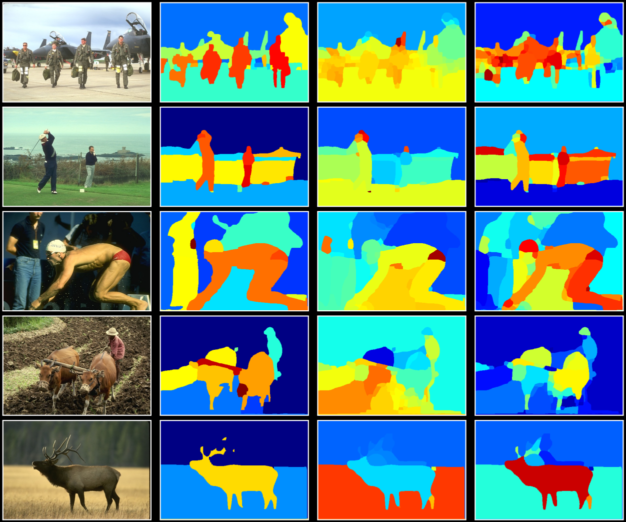


Image | F/G: Ground-truth | F/G: [Maire, ECCV 2010] | F/G: Ours

Benchmark: Boundary Ownership Correctness

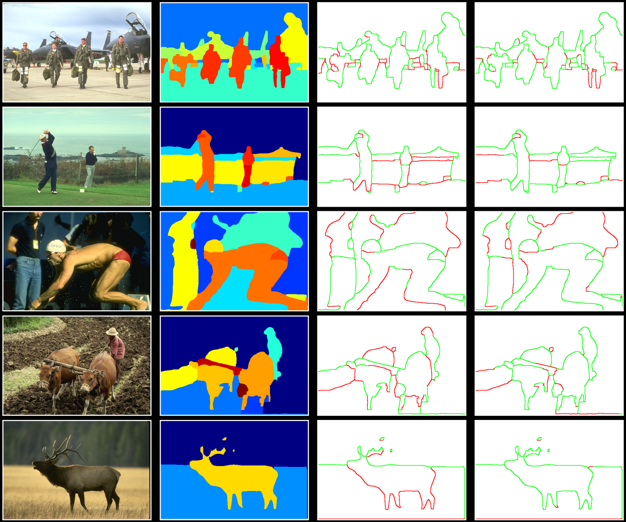


Image | **F/G: Ground-truth** | **F/G: [Maire, ECCV 2010]** | **F/G: Ours**

Benchmarks: Quantitative Performance

Segmentation: Ground-truth	Figure/Ground Prediction Accuracy R-ACC	B-ACC	B-ACC-50	B-ACC-25
F/G: Ours	**0.62**	**0.69**	**0.72**	**0.73**
F/G: [Maire, ECCV 2010]	0.56	0.58	0.56	0.56

Segmentation: Ours	Figure/Ground Prediction Accuracy R-ACC	B-ACC	B-ACC-50	B-ACC-25
F/G: Ours	**0.66**	**0.70**	**0.69**	**0.67**
F/G: [Maire, ECCV 2010]	0.59	0.62	0.61	0.58

R-ACC: Pairwise region accuracy
B-ACC: Boundary ownership accuracy
B-ACC-25: B-ACC on 25% most foreground regions in each image
B-ACC-50: B-ACC on 50% most foreground regions in each image

Cross-Domain Generalization: Horses

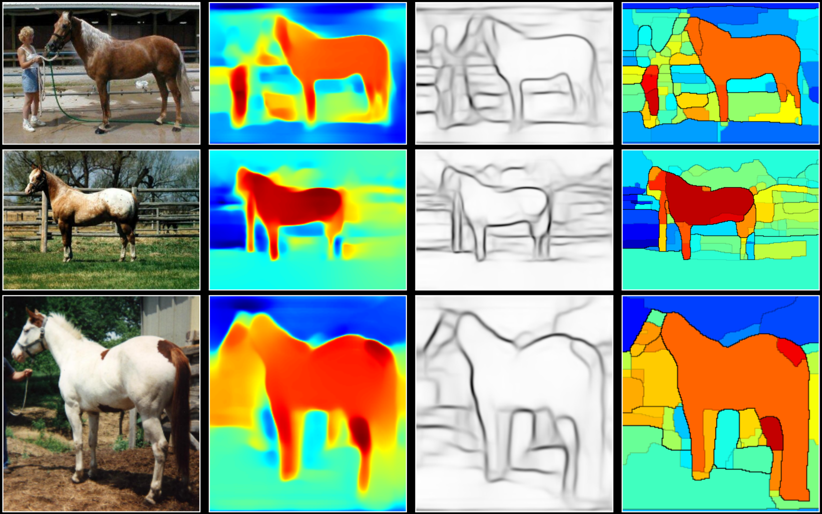


Image **F/G** **Boundaries** **Seg + F/G**

Cross-Domain Generalization: PASCAL

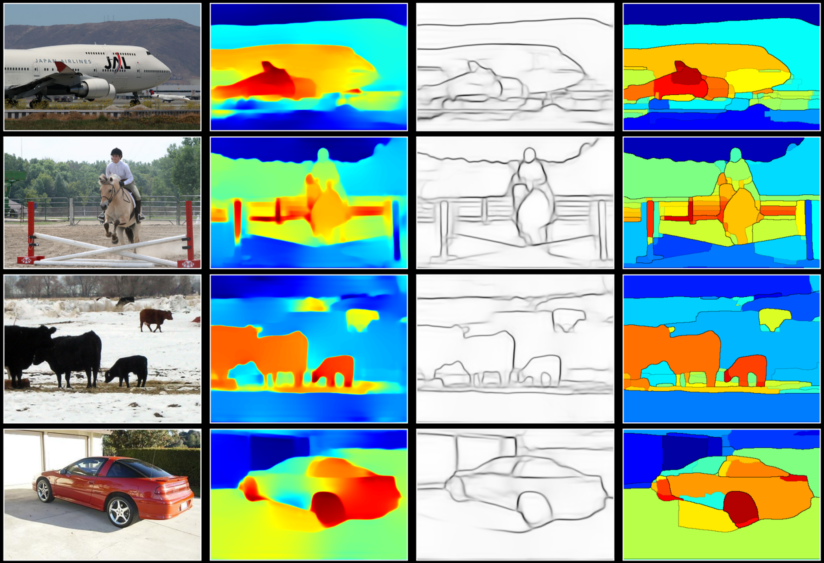
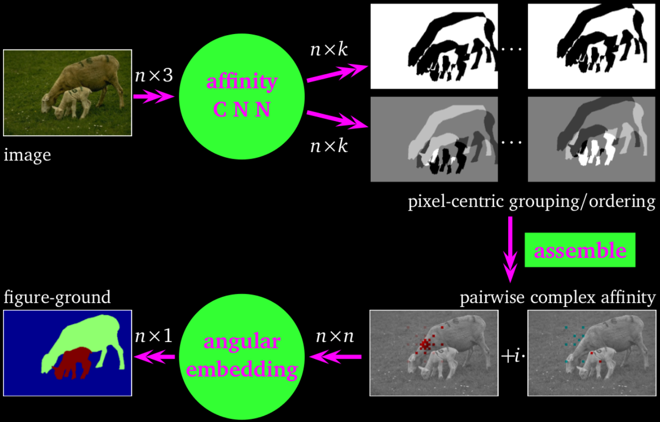


Image | **F/G** | **Boundaries** | **Seg + F/G**


n×3
affinity
CNN
n×k
n×k
image
pixel-centric grouping/ordering
assemble
pairwise complex affinity
figure-ground
n×1
angular
embedding
n×n
+i·

Thank You!