Recognizing objects by piecing together the Segmentation Puzzle

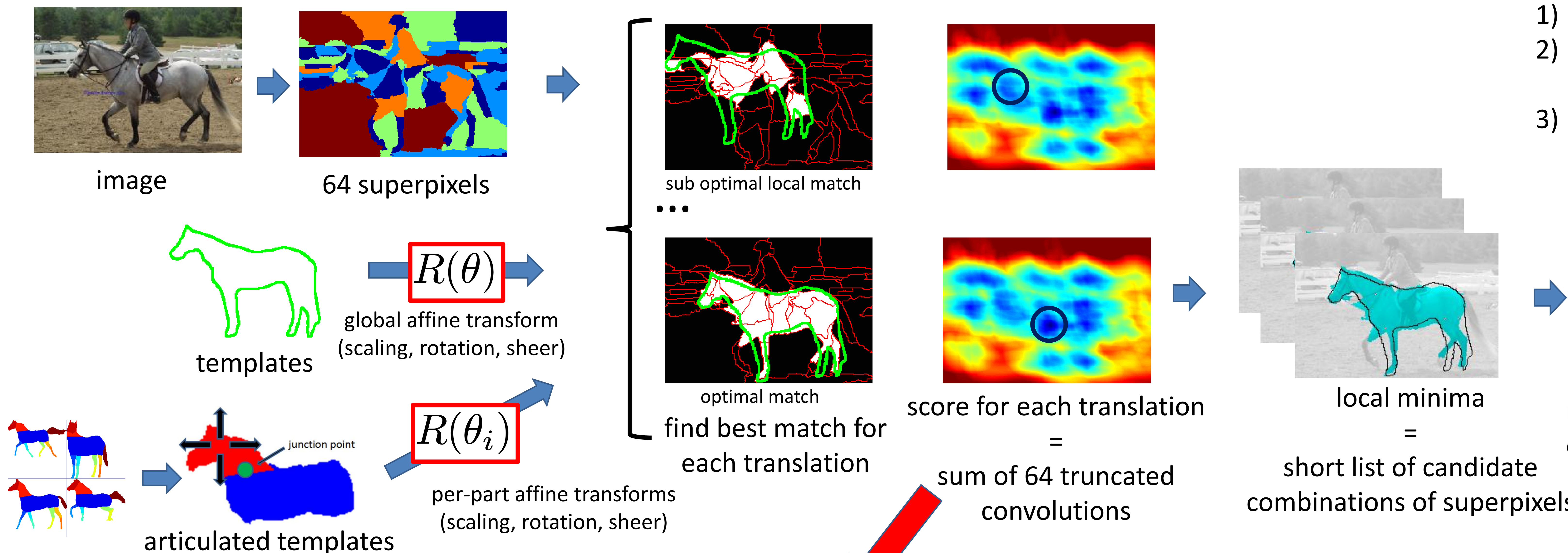
Timothee Cour Jianbo Shi

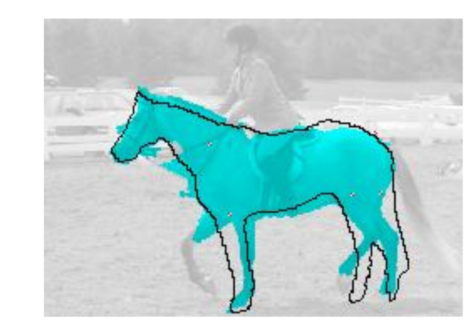
University of Pennsylvania

Goal: object specific segmentation given a few hand-segmented images of that category

Main challenge: integrate bottom-up segmentation cues with top-down model cues

Our approach: exhaustive search in the space of superpixel combinations with separable cost function





reranking using higher complexity methods

Contribution: efficient shape detection/segmentation based on 3 ideas:

1) exhaustive search over superpixel combinations can be decoupled
2) a discrete version of Green's theorem is used to compute the correlation score
3) coarse to fine scheme (with conditions for optimality) allows for linear time detection

space complexity: $O(|\partial Y|)$ using compression based on *Run Length Encoding*
$\Rightarrow$8,000 large templates fit in 2MB

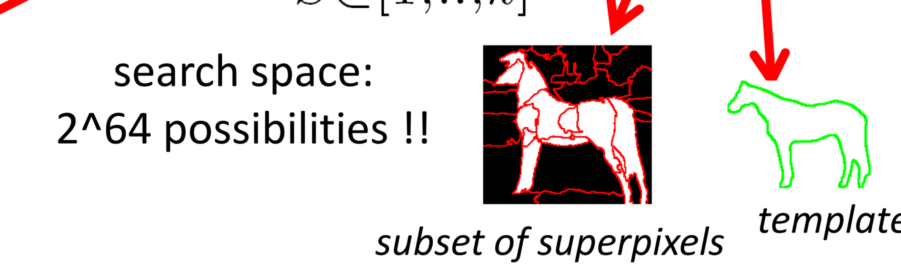
time complexity: $O(|\partial Y|)$ per translation using Green's theorem, and *O(npixels)* per template across translations using coarse to fine
$\Rightarrow$0.01 seconds per template

Cost Function for Shape Detection and Segmentation

shape distance function: $d(X,Y) = \min_{T\in\mathbb{T}} \frac{1}{|T(Y)|} d_H(X,T(Y))$

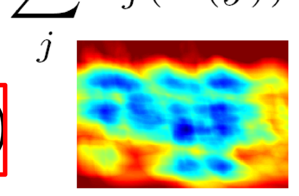
invariant to similarity transforms

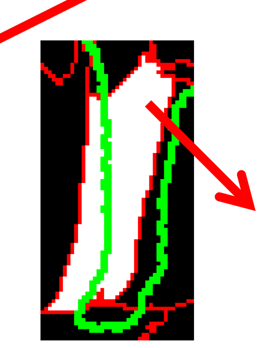
cost function: $\epsilon(Y) = \min_{S\subset[1,..,k]} d(R_S, Y)$



cost can be decoupled w.r.t. each region R_j: $\epsilon(y) = 1 + \min_{T\in\mathbb{T}} \frac{1}{||T(y)||^2} \sum_j \sigma_j(T(y))$

as a sum of truncated convolutions $\sigma_j(y) = \min(0, |R_j| - 2r_j^T y)$

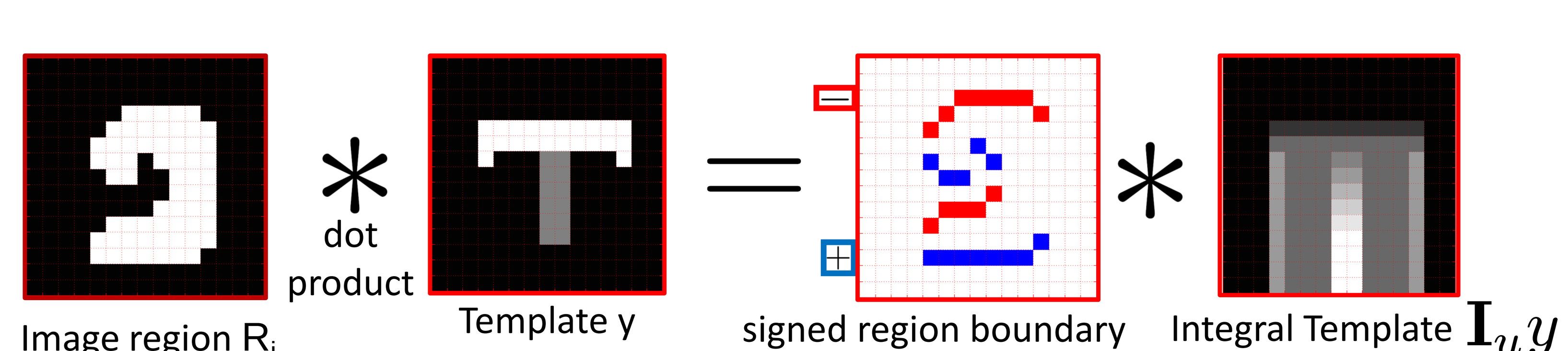




region with $r_j^T y > \frac{1}{2}|R_j|$

how to detect a template anywhere in a large image in *O(npixels)* or 0.01 seconds

discrete version of Green's theorem for fast *random-access* convolution



for each region R_j $\quad r_j^T y = \iint_{R_j} y = \iint_{R_j} \frac{\partial \mathbf{I}_u y}{\partial u} = \int_{\partial R_j} \mathbf{I}_u y$

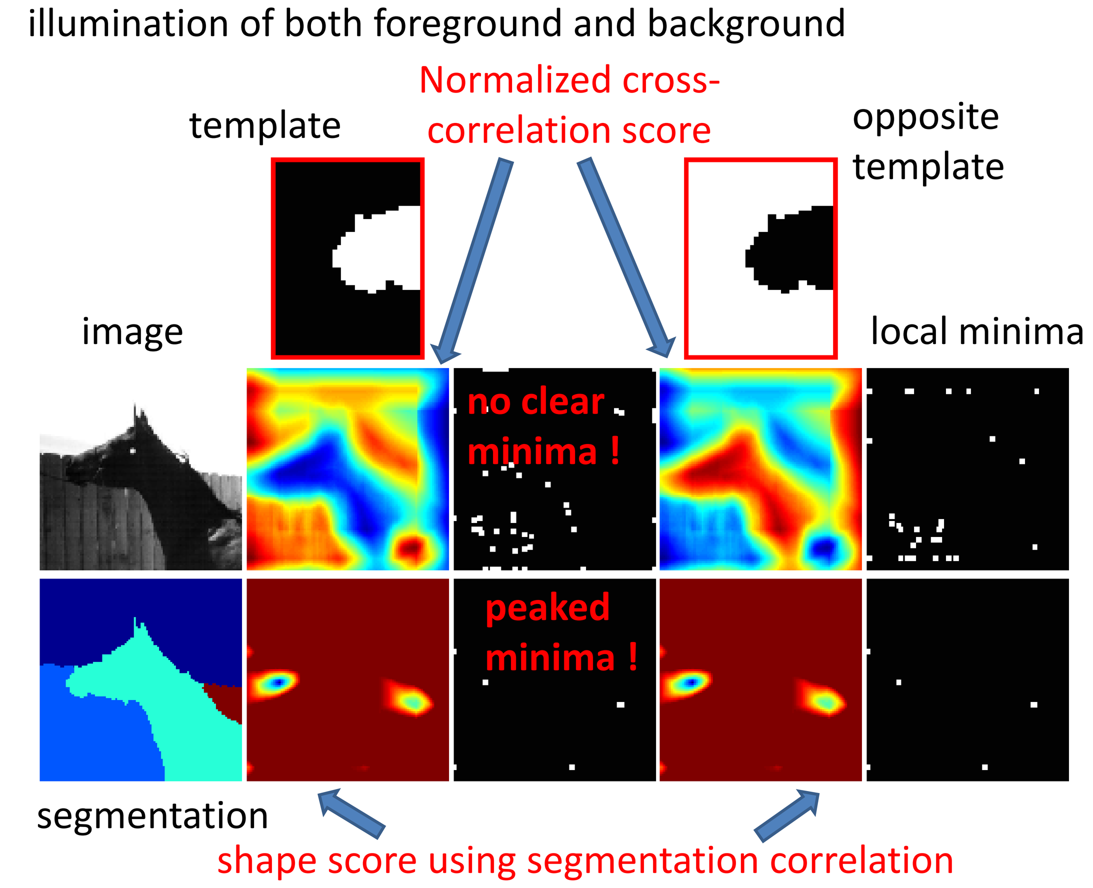
running time per dot product: $O(|\partial Y|)$ instead of $O(|Y|)$

can be coupled with coarse-to-fine approach starting from the local maxima of the score at a coarser resolution projected on the fine resolution.

NOT possible with FFT, which is global !

gray-level correlation vs. segmentation correlation

gray level correlation (or NCC) is very sensitive to variations in illumination of both foreground and background

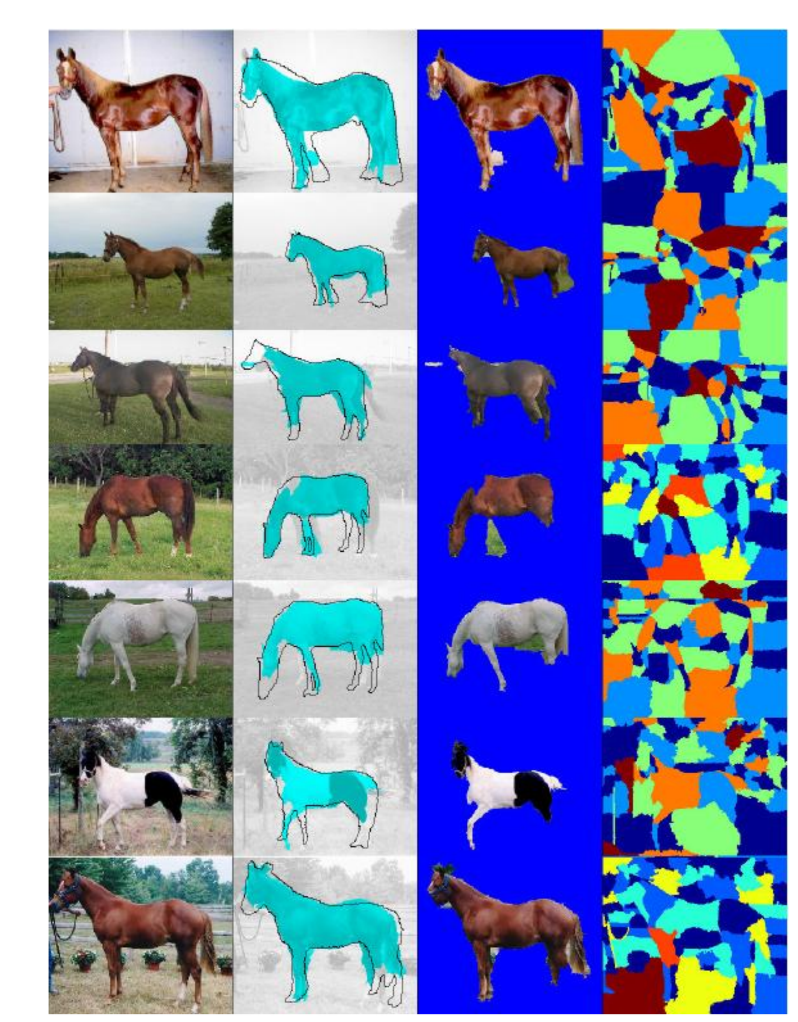


Results on Weizmann horse dataset

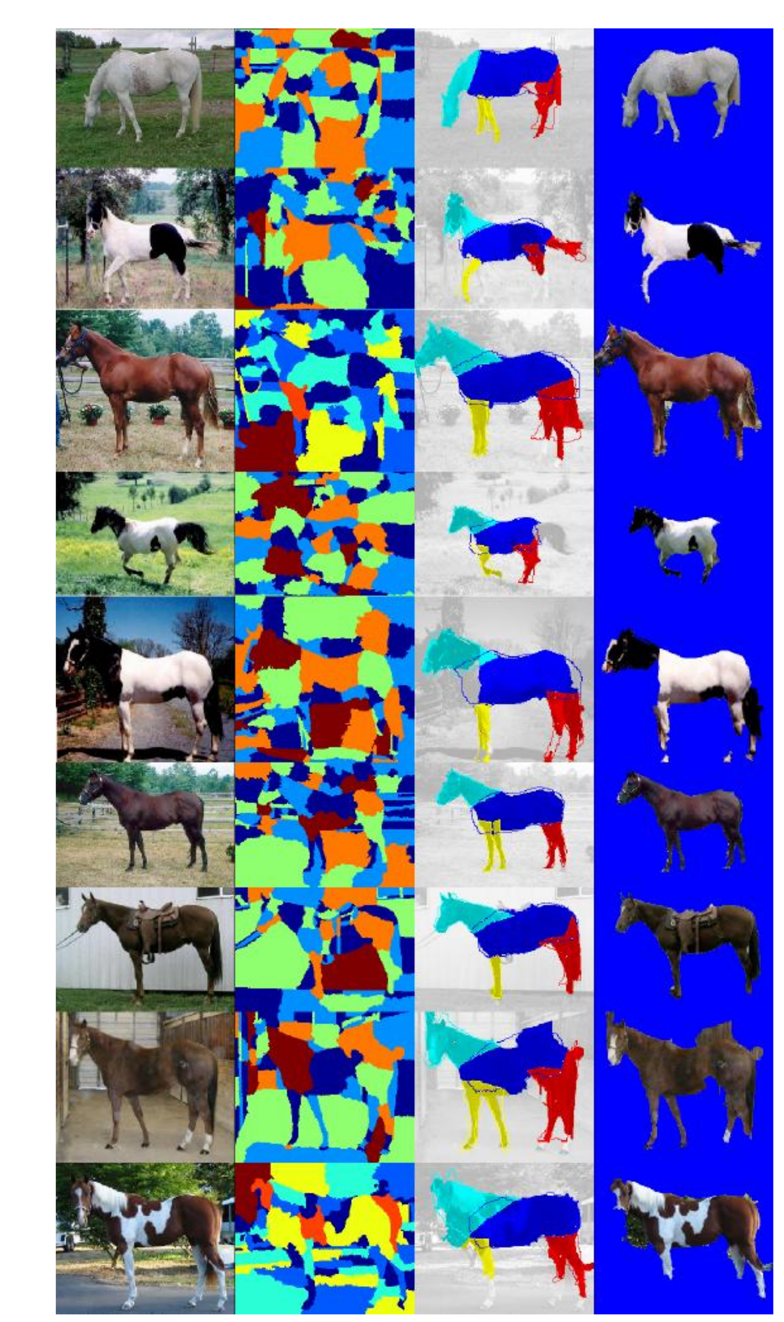
training/testing: 20/308 images

average pixel consistency: 94.2% (using the oracle best out of 10)

same with articulated model: 95.2% (using the oracle best out of 10 for each part)



segmentation/detection with whole template



segmentation/detection with articulated template