Note:Crafting Better Contrastive Views for Siamese Representation Learning

Sinkoo

April 10, 2022

1 Abstract

For Siamese contrastive learning, it is important to generate good pairs, and mostly used today is random crop. But random crop have 2 major problems: 1, it may produce pairs from "helpless" sample (Like background) which will decrease the performance; 2, produce similar pairs that can hardly help.
This paper proposed ContrastiveCrop to generate positive pairs through Semantic-aware Localization and Center-suppressed Sampling.

2 Model

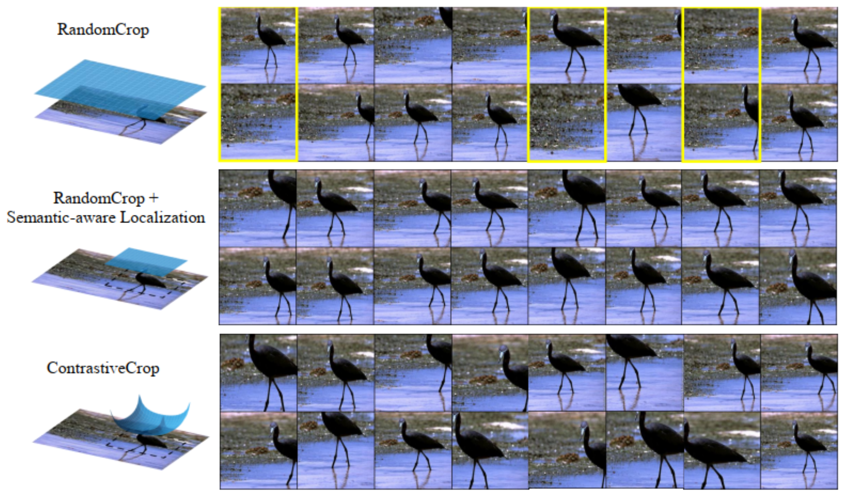


Figure 3. Visualization of *RandomCrop*, *RandomCrop* + *Semantic-aware Localization* and our *ContrastiveCrop*. We show the sampling distributions and operable regions for three settings on the left, and correspondent sampled pairs on the right. Pairs made by *RandomCrop* include several false positives that totally miss the object (marked in yellow box). Using *RandomCrop* with *Semantic-aware Localization* reduces false positives, but introduces easy positive pairs that share large similarity. Last, our *ContrastiveCrop* could reduce false positive pairs while increasing variance at the same time.

2.1 Semantic-aware localization(Avoiding the 1st problem)

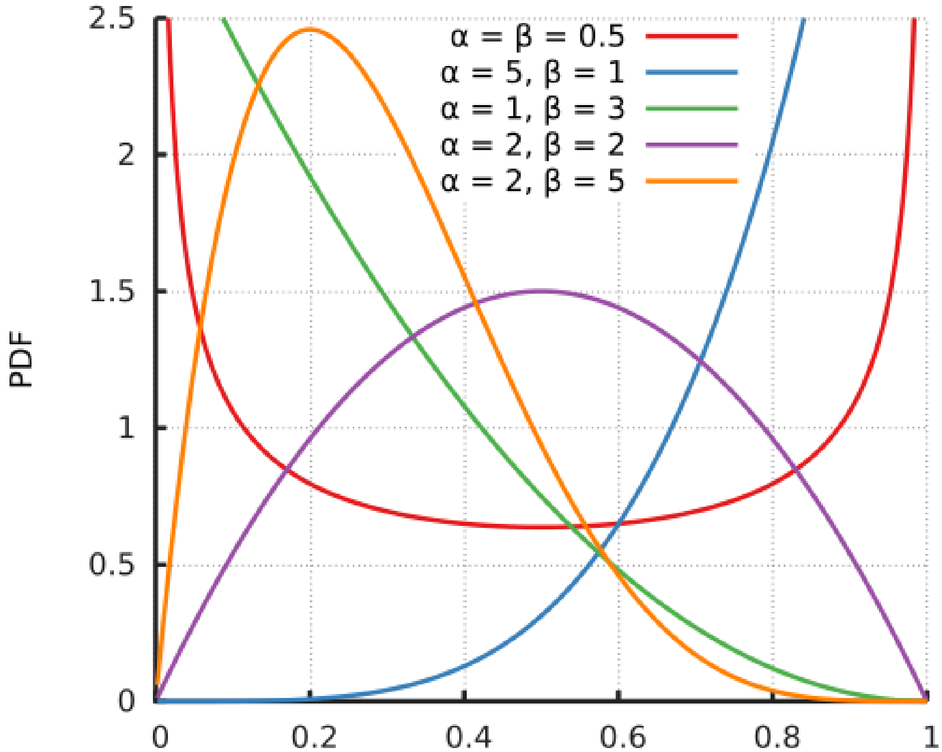
Derive the heatmap by summing the features of last convolutional layer across the channel dimension and normalizing it to [0, 1]. Use this heatmap(represents the probably position of target object) to obtain the bounding box of the object, which can be written as $B = L(\mathbf{1}[M > k])$ $k \in [0, 1]$ is the threshold.
After get bounding box B the semantic crops can be obtained as follows:

$$(\dot{x}, \dot{y}, \dot{h}, \dot{w}) = \mathbf{R}_{crop}(s, r, B)$$

3 Center-suppressed Sampling(Avoiding 2nd problem)

$$(\dot{x}, \dot{y}, \dot{h}, \dot{w}) = \mathbf{C}_{crop}(s, r, B)$$

where $\mathbf{C}_{crop}$ applies center-suppressed sampling. And center-suppress is achieved by beta distribution with $\alpha < 1$, and in order to get symmetric results, α is always equal to β.



Algorithm 1 *ContrastiveCrop* for Siamese Representation Learning

Input: Image I, Crop Scale s, Crop Ratio r, Threshold of Activations k, Parameter of β Distribution α.

$h = \sqrt{s \cdot r}$ ▷ Height of the crop

$w = \sqrt{s/r}$ ▷ Width of the crop

$F = Forward(I)$ ▷ Features of last layer

$M = Normalize(F)$ ▷ Heatmap after normalizing

$B = L(\mathbb{1}[M > k])$ ▷ Bounding box by Eq. 2.

$x = B_{x0} + (B_{x1} - B_{x0}) \cdot u, u \sim \beta(\alpha, \alpha)$

$y = B_{y0} + (B_{y1} - B_{y0}) \cdot v, v \sim \beta(\alpha, \alpha)$

▷ Sample crop center x and y from β distribution

Output: Crop $C = (x, y, h, w)$

4 Summary

This paper proposed a crop method to provide more useful pairs for Siamese Network and solved two major problems in previous methods by Semantic-aware localization and Center-suppressed Sampling.