Supplementary material MixSIS: Mixed-supervised instance segmentation

Christian Wilms
christian.wilms@uni-hamburg.de
Emre Gülsoylu
emre.guelsoylu@uni-hamburg.de
Sven Schreiber
sven.schreiber@uni-hamburg.de
Simone Frintrop
simone.frintrop@uni-hamburg.de

Computer Vision Group
University of Hamburg
Hamburg, Germany

1 Object Properties for Efficient Annotation Study

For our study on efficient annotation schemes, we split the annotations in the COCO dataset according to several object properties representing shape and color features. Below, we define each object property.

1.1 Compactness Properties

Boxness is the ratio between the number of pixels inside the mask and inside the axis-aligned bounding box. It assesses the similarity of the object's shape with an axis-aligned rectangle.

Oriented Boxness is similar to boxness, but uses an oriented bounding box. To produce the oriented bounding box, the second-order moments of a given object mask are calculated, leading to the orientation as an angle. Oriented boxness assesses the similarity of the object's shape with a rectangle of the same orientation.

Center Distance, proposed by [4], is the average distance between the points on the object's contour and the object's centroid. To gain independence from the object's size, it is normalized by the number of points on the object's contour. Additionally, it is normalized by the maximum center distance within the contour. Intuitively, objects with a strong variation in their contour point's center distances will be assigned low values and a low compactness.

Circularity [1] measures the similarity of an object's shape with a circle based on the object's perimeter p and area a. The perimeter is the length of the contour. Circularity is defined as

$$\rho_{\text{circularity}} = \frac{4\pi a}{p^2}. \tag{1}$$

Eccentricity [1] is based on the second-order moments of an object's mask. Given these moments, an ellipse with the same moments is constructed. Subsequently, eccentricity of the object is defined as the ratio between the minor axis and the major axis of the constructed ellipse. Hence, the eccentricity assesses the elongation of the object, while being rotation invariant.

Centroid Location describes whether the object's centroid is within the object or outside. Hence, it singles out objects with a ring-like shape or a central hole. Such objects are likely more difficult to segment based on a single point, the centroid.

1.2 Convexity Properties

Distance, proposed by [2], assesses the average minimal distance between each point on the object's contour and its respective closest point on the object's convex hull. Subsequently, it is normalized by the length of the object's contour and the largest minimal distance observed.

Hull assesses the ratio between the perimeter of the object and its convex hull.

Solidity is the ratio between the size of the object and its convex hull.

1.3 Compactness Properties

Good Continuation measures how smooth an object's contour is. The smoother the contour, the more likely a human will recognize the object as such, according to Gestalt principles [5]. We follow [3] and measure good continuation by examining the angular changes between neighboring contour pixels and calculating the mean angular change across an object's contour. Examples for objects with high and low good continuation values are visualized in Fig. **??**.

Symmetry [2, 4] is defined based on the object's major and minor axes derived from the second-order moments. First, the object is flipped along each axis individually. Subsequently, the overlap of each flipped version and the original is calculated. The average of the two overlaps is used to describe the level of symmetry.

Connectedness determines if an object consists of one or multiple connected components. Consisting of multiple connected components might give rise to segmentation errors, as some components could be missed.

1.4 Other Properties

Centroid Proximity describes the distance between an object's centroid and the closest other object centroid in the image. In case no other object exists, it is set to the length of the image diagonal. To make this property invariant to the image size, the resulting distance is normalized to the length of the image diagonal. This measure will allow us to determine objects in crowded environments, which might be more difficult to discern.

Color Contrast assesses how different an object's colors are w.r.t. the object's direct surrounding. Hence, it measures how much an object sticks out from it's surrounding, possibly aiding segmentation. To measure color contrast, we define the object as the center region and a 10-pixel-wide area around the object as the surrounding area. Subsequently, we calculate a normalized 3D histogram in RGB space for each region and compute their distance using intersection similarity.

Object Color determines how many bins in a $16 \times 16 \times 16$ 3D histogram in RGB space contain more the 5% of the object's pixels. Hence, the property determines if an object is dominated by only a few colors or if it is highly multicolored. Multicolored objects might be challenging to segment due to stronger intra-object diversity.

References

[1] R. Gonzalez and R. Woods. *Digital Image Processing*. Peason, 4th edition, 2014.

[2] A. Karpathy, S. Miller, and L. Fei-Fei. Object discovery in 3D scenes via shape analysis. In *ICRA*, 2013.

[3] G. Kootstra and D. Kragic. Fast and bottom-up object detection, segmentation, and evaluation using Gestalt principles. In *ICRA*, 2011.

[4] T. Werner, G. Martín-García, and S. Frintrop. Saliency-guided object candidates based on gestalt principles. In *ICVS*, 2015.

[5] M. Wertheimer. Untersuchungen zur Lehre von der Gestalt II. *Psychologische Forschung*, 1923.