Supplementary Materials



Figure A - 1. The overall evolutionary algorithm framework of our search method for learning data augmentation policies.

A. Search framework review

We provide a review of our overall search method framework in Fig. A - 1. We adopt the evolutionary algorithm for search, where a population of data augmentation policies are randomly initialized and then evolved in iterations. During search, policies are sampled from the search space. Then, they are trained and evaluated by our estimation metric. The computed metrics serve as feedback to update. Better policies are generated in this framework over time.

B. Derivation of Gaussian deviation

The standard deviation of the Gaussian map can be derived as the following. Given the Gaussian map

$$f(x,y) = \exp\left(-\left(\frac{(x-x_c)^2}{2\sigma_x^2} + \frac{(y-y_c)^2}{2\sigma_y^2}\right)\right), \quad (1)$$

its integration among the image can be calculated as

$$V = \int_0^H \int_0^W f(x, y) \, \mathrm{d}x \mathrm{d}y \approx 2\pi \delta_x \delta_y. \tag{2}$$

With the defination of the *area ratio* $r = V/s_{\text{box}}$ and $s_{\text{box}} = hw$, we can formulate their relationship as

$$r = \frac{2\pi}{hw} \delta_x \delta_y. \tag{3}$$

Without loss of generality, the variance factors δ_x and δ_y should be correlated with the ratio of box height (width) and image height (width) to make the Gaussian map match the box aspect ratio. This can be represented as

$$\delta_x/\delta_y = \left(\frac{h}{H}\right) \left/ \left(\frac{w}{W}\right).$$
 (4)

Table A - 1. Comparison with methods on test-dev.

Method	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
Res101:						
PSIS	40.2	61.1	44.2	22.3	45.7	51.6
InstaBoost	43.0	64.3	47.2	24.8	45.9	54.6
Ours	44.4	66.1	48.8	27.1	47.4	55.3
Res101-DCN-C4:						
$SNIP^{\dagger}$	44.4	66.2	49.9	27.3	47.4	56.9
SNIPER[†]	46.1	67.0	51.6	29.6	48.9	58.1
Ours [†]	46.9	68.8	51.7	30.6	48.1	58.4

Table A - 2. Comparison on larg-scale jittering.

Method	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
RetinaNet Res50:						
Baseline	40.1	59.7	43.0	23.7	44.1	54.4
Ours	41.6	61.6	44.4	25.4	45.4	55.6

Combining the above two equations, Eq. (3) and Eq. (4), we can obtain the variance factors as

$$\sigma_x = h \sqrt{\frac{W/H}{2\pi}} r, \quad \sigma_y = w \sqrt{\frac{H/W}{2\pi}} r.$$
 (5)

C. Augmentation operations details

We list the details about all box-level operations in Tab. A - 3 with their description and magnitude ranges. Besides, we provide the visualization example of these augmentations in Fig. A - 2.

D. Removing context pixels

In Tab. 1 in the main paper, we evaluate well-trained models on validation images whose context (background) pixels are removed. For better understanding, we provide an example image as shown in Fig. A - 3.

E. Other Comparisons

Some methods in Tab. 8 are reported on test-dev. We show our counterparts on test-dev in Tab. A - 1. † denotes that the multi-scale testing technique has been used.

We also perform experiments upon the large scale jittering, *i.e.*, [0.5, 2.0] as in Tab. A - 2. The baseline is enhanced to 40.1% AP, while ours achieves 41.6% AP.

Operation	Description	Magnitude range
Brightness	Control the object brightness. Magnitude = 0 represents the black, while magni-	[0.1, 1.9]
	tude = 1.0 means the original.	
Color	Control the color balance. Magnitude = 0 represents a black & white object, while	[0.1, 1.9]
	magnitude = 1.0 means the original.	
Contrast	Control the contrast of the object. Magnitude = 0 represents a gray object, while	[0.1, 1.9]
	magnitude = 1.0 means the original object.	
Cutout	Randomly set a square area of pixels to be gray. Magnitude represents the side length.	[0, 60]
Equalize	Equalize the histogram of the object area.	-
Sharpness	Control the sharpness of the object. Magnitude = 0 represents a blurred object, while	[0.1, 1.9]
	magnitude = 1.0 means the original object.	
Solarize	Invert all pixels above a threshold value. Magnitude represents the threshold.	[0,256]
SolarizeAdd	For pixels less than 128, add an amount to them. Magnitude represents the amount.	[0,110]
Hflip	Flip the object horizontally.	-
Rotate	Rotate the object to a degree. Magnitude represents the degree.	[-30,30]
ShearX/Y	Shear the object along the horizontal or vertical axis with a magnitude.	[-0.3, 0.3]
TranslateX/Y	Translate the object in the horizontal or vertical direction by magnitude pixels.	[-150, 150]

Table A - 3. Details about box-level operations with their description and magnitude ranges.



Figure A - 2. Examples on different box-level operations with magnitudes random sampled.



(a) Original image

(b) Image with context pixels removed

Figure A - 3. An example image of removing context.