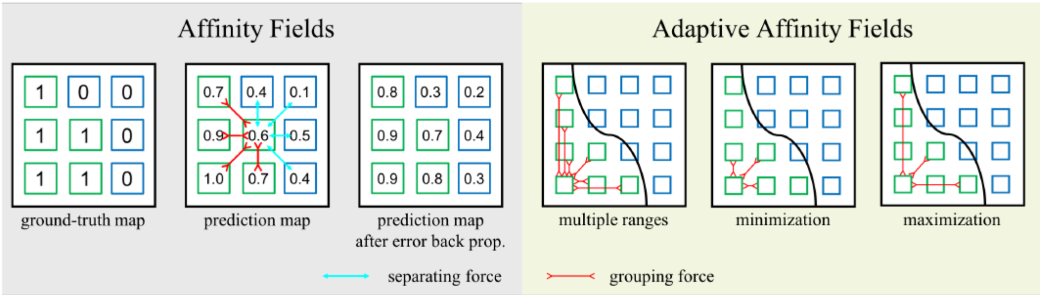
ECCV 2018 长文 “Adaptive Affinity Fields for Semantic Segmentation” from UC Berkeley / ICSI

1.1 Affinity Fields



相似场的方法是一种后处理方法，在已经生成分割结果后，对分割结果的进一步优化。如上图所示，当相似场的 kernel_size = 3 x 3 时，对于第二张图中预测概率为 0.6 的中间点，所有绿色的点和它属于同一类别，蓝色属于其他类别。

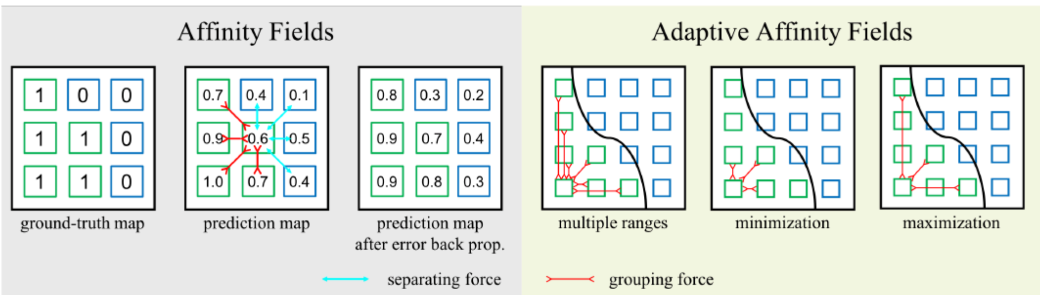
相似场的 LOSS 为：

$$\mathcal{L}^{ic}_{\text{affinity}} = \begin{cases} \mathcal{L}^{i\bar{b}c}_{\text{affinity}} = D_{KL}(\hat{y}_j(c)||\hat{y}_i(c)) & \text{if } y_i(c) = y_j(c) \\ \mathcal{L}^{ibc}_{\text{affinity}} = \max\{0, m - D_{KL}(\hat{y}_j(c)||\hat{y}_i(c))\} & \text{otherwise} \end{cases} \quad (3)$$

经过反向传播后，如果两个点属于同一类则使他们的预测概率更加相近；如果属于不同类则使他们的预测概率差异更大。如上图图 3 所示。

1.2 Adaptive Affinity Fields (AAF)

但是相似场的 kernel size 是固定的，而物体的尺寸却是多样化的。因此 kernel size 能够根据特征而自适应的调整尺寸很重要，自适应相似场由此被提出。



这里使用对抗网络（GAN）的方式，自适应的找出每个区域所适合的相似场的 kernel size，从数学上表达即为对不同区域，给不同 kernel size 以不同的权重得到一个加权的 kernel size。

$$\mathcal{L}_{\text{multiscale}} = \sum_c \sum_k w_{ck} \mathcal{L}^{ck}_{\text{region}} \quad \text{s.t.} \sum_k w_{ck} = 1 \text{ and } w_{ck} \geq 0$$

在实验中，作者设定了 3x3，5x5 和 7x7 三种尺寸。下图即为对于不同类别，在 AAF 的处理过程中的不同 kernel size 权重不同。

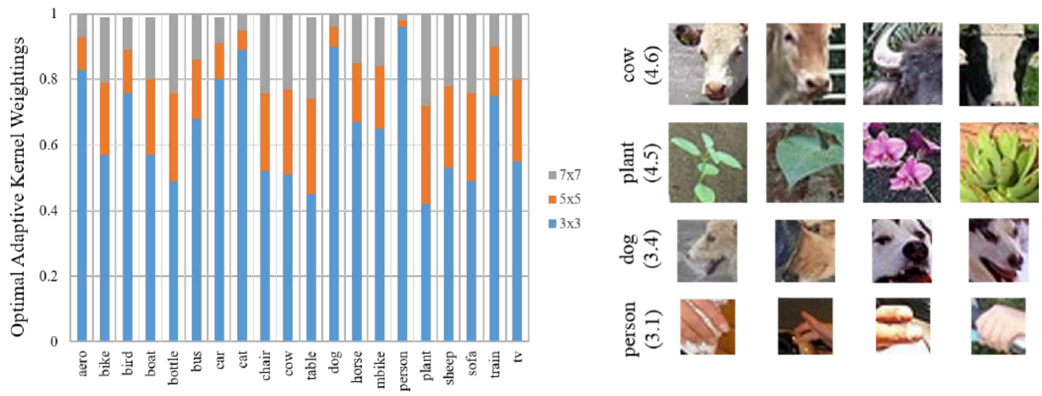


Fig. 4. Left: The optimal weightings for different kernel sizes of the edge term in AAF for each category on PASCAL VOC 2012 validation set. **Right:** Visualization of image patches with corresponding effective receptive field sizes, suggesting how kernel sizes capture the shape complexity in critical regions of different categories.

实验结果表明，这种尺寸自适应的方式比单纯的固定尺寸方式，在 cityscapes 验证集上提高了 0.52%的 mIoU。

Method	road	swalk	build.	wall	fence	pole	tlight	tsign	veg.	terrain	sky	person	rider	car	truck	bus	train	mbike	bike	mIoU
FCN	97.31	79.28	89.52	38.08	48.63	49.70	59.37	69.94	90.86	56.58	92.38	75.91	46.24	92.26	50.41	64.51	39.73	54.91	73.07	66.77
PSPNet	97.96	83.89	92.22	57.24	59.31	58.89	68.39	77.07	92.18	63.71	94.42	81.80	63.11	94.85	73.54	84.82	67.42	69.34	77.42	76.72
Affinity	97.52	80.90	90.42	40.45	49.81	55.97	63.92	73.37	91.49	59.01	93.30	78.17	52.16	92.85	52.53	65.78	39.28	52.88	74.53	68.65
AAF	97.58	81.19	90.50	42.30	50.34	57.47	65.39	74.83	91.54	59.25	93.11	78.65	52.98	93.15	53.10	67.58	38.40	51.57	74.80	69.14
CRF	97.96	83.82	92.14	57.16	59.28	57.48	67.71	76.61	92.09	63.67	94.35	81.62	62.98	94.81	73.59	84.81	67.49	69.22	77.28	76.53
GAN	97.95	83.59	92.01	56.92	60.17	58.63	68.37	77.36	92.28	62.70	94.42	81.59	62.27	94.94	78.09	82.79	56.75	69.19	77.78	76.20
Affinity	98.08	85.58	92.60	58.33	61.45	**66.80**	**74.19**	81.29	92.90	65.34	94.87	84.00	65.84	95.50	76.84	85.80	64.19	72.32	79.83	78.72
AAF	98.18	85.35	92.86	58.87	61.48	66.64	74.00	80.98	92.95	65.31	94.91	**84.27**	**66.98**	95.51	79.39	87.06	67.80	72.91	80.19	**79.24**

Table 3. Per-class results on Cityscapes validation set. Gray colored background denotes using FCN as the base architecture.