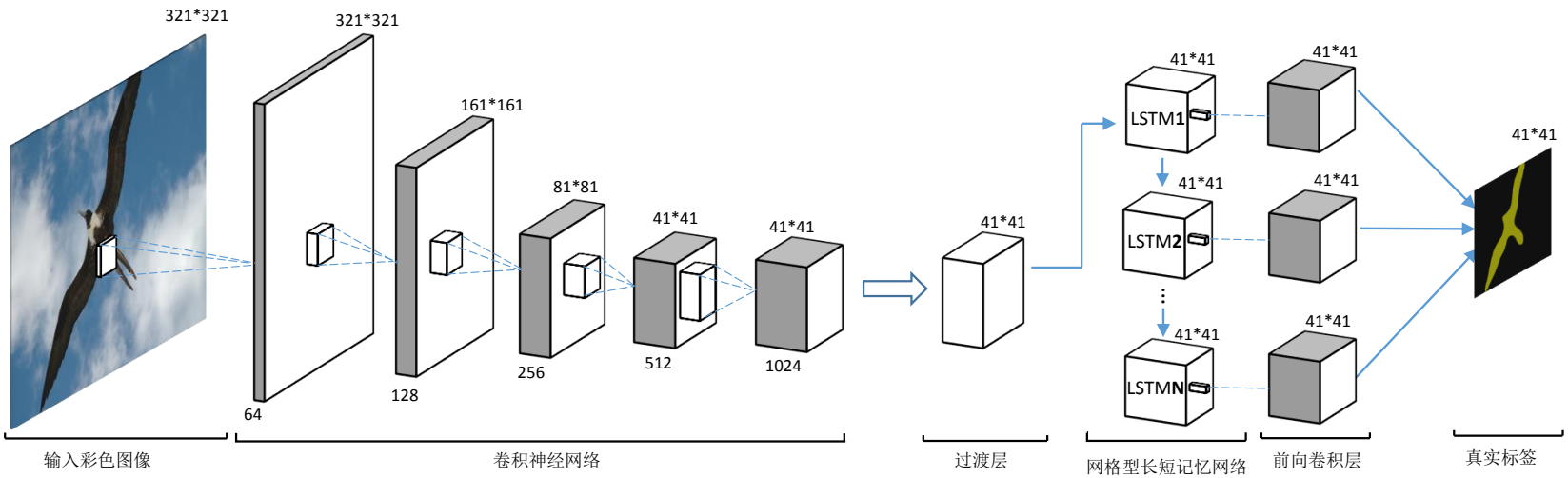
[BoldFont=Adobe Heiti Std,ItalicFont=Adobe Kaiti Std]AdobeSongStd-Light



(Semantic Image Segmentation) ,

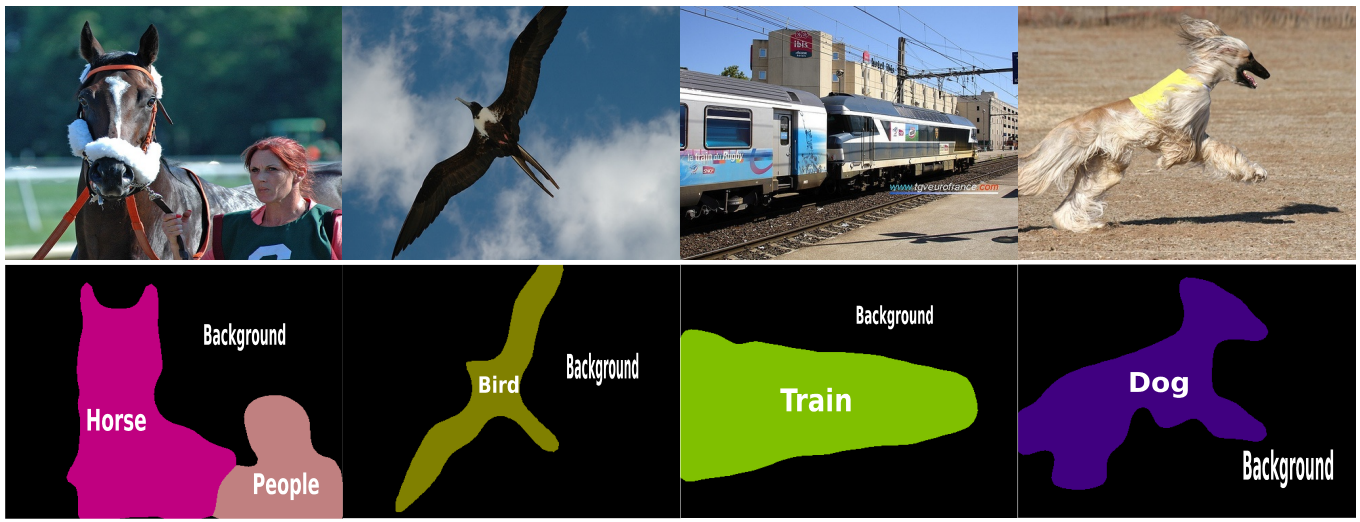


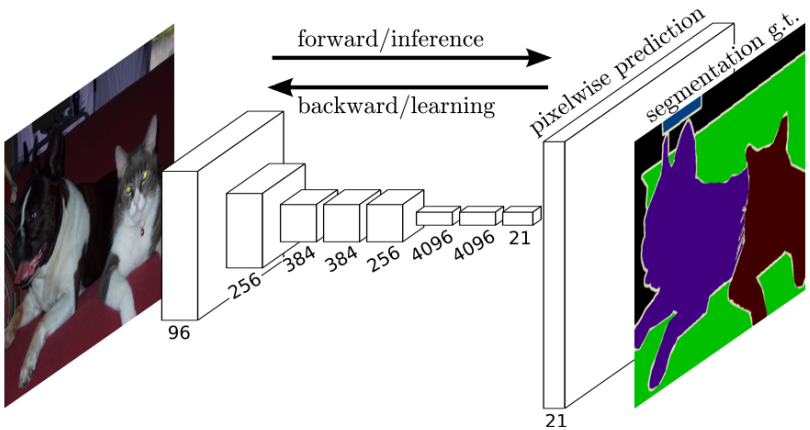
Figure 1: VOC 2012

- (SIFT, HOG)
-
  - GPU
  -
-

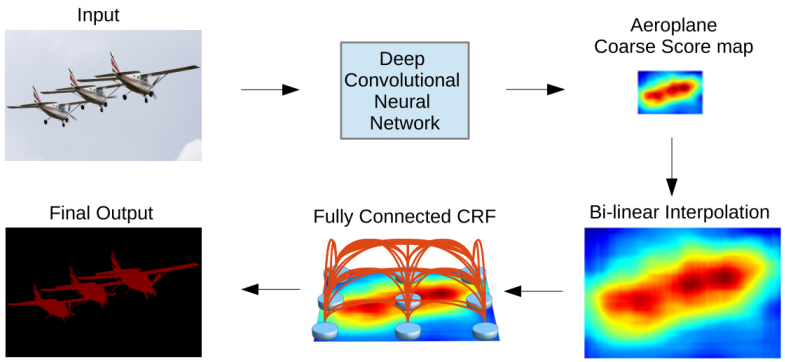
†

-
-
-
-

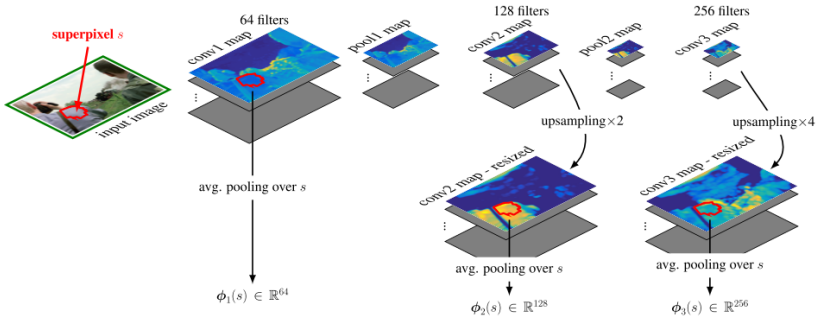
}



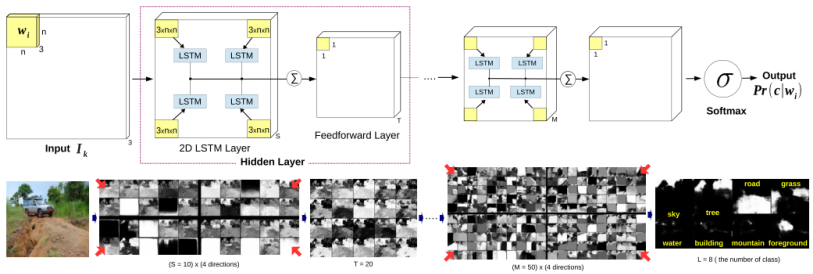
a. [Long et al, CVPR 2015]



b. +[Chen et al, ICLR 2015]



c. +[Mostajabi et al, CVPR 2015]



d. [Byeon et al, CVPR, 2015]

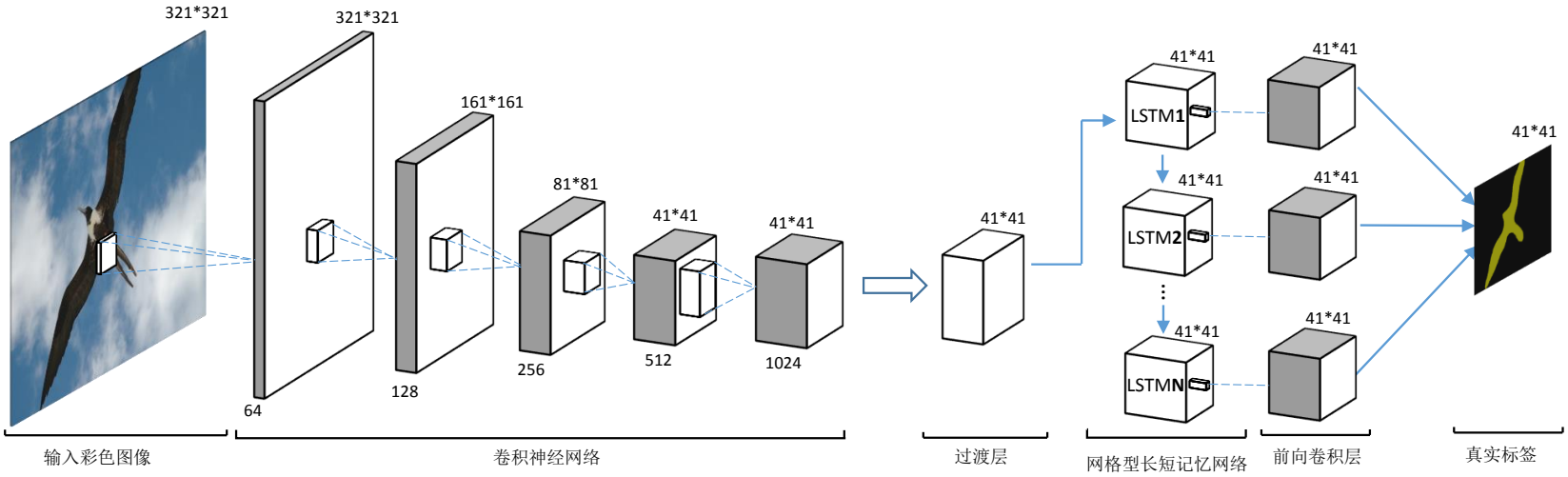

321*321
321*321
161*161
81*81
41*41
41*41
64
128
256
512
1024
41*41
LSTM1
LSTM2
LSTMN
41*41
41*41
41*41
41*41
41*41
41*41
输入彩色图像
卷积神经网络
过渡层
网格型长短记忆网络
前向卷积层
真实标签

Figure 2:

-
- ()
- ()
- ()
- ()

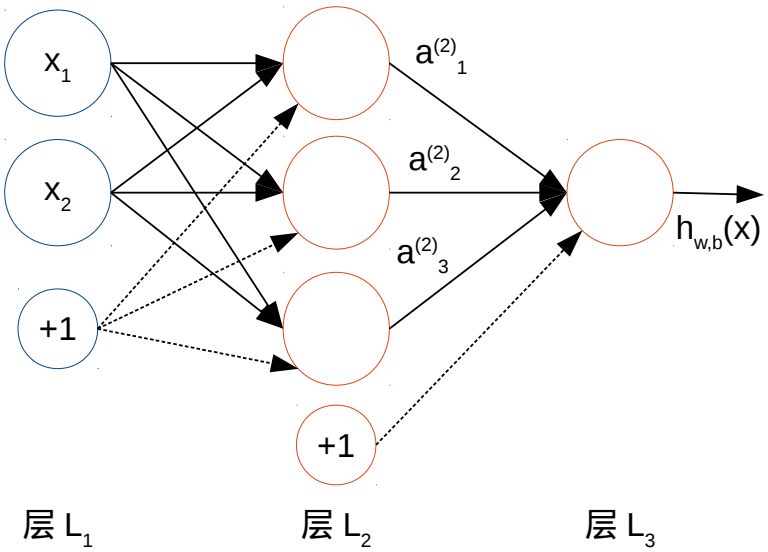


Figure 3:

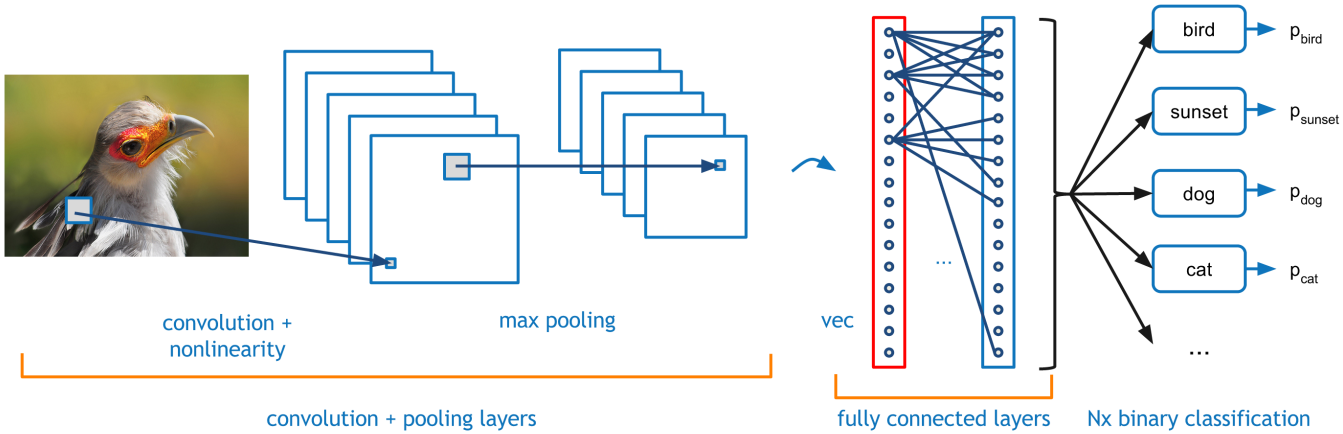
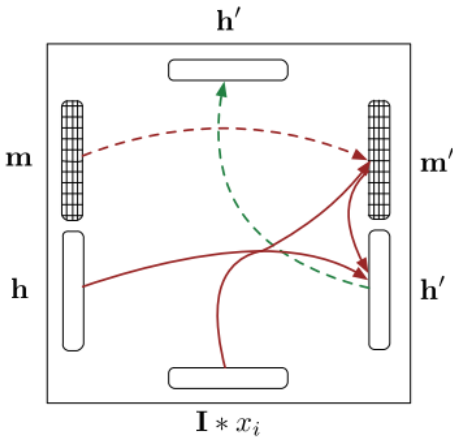

convolution +
nonlinearity
max pooling
vec
bird
sunset
dog
cat
...
p_{bird}
p_{sunset}
p_{dog}
p_{cat}
convolution + pooling layers
fully connected layers
Nx binary classification

Figure 4:

()



Standard LSTM block

Figure 5:

$$
\begin{aligned}
\mathbf{g}^u &= \delta(\mathbf{W}^u * \mathbf{H}) \\
\mathbf{g}^f &= \delta(\mathbf{W}^f * \mathbf{H}) \\
\mathbf{g}^o &= \delta(\mathbf{W}^o * \mathbf{H}) \\
\mathbf{g}^c &= \tanh(\mathbf{W}^c * \mathbf{H}) \\
\mathbf{m}' &= \mathbf{g}^f \odot \mathbf{m} + \mathbf{g}^u \odot \mathbf{g}^c \\
\mathbf{h}' &= \tanh(\mathbf{g}^o \odot \mathbf{m}') \\
\mathbf{H} &= \begin{bmatrix} I * \mathbf{x}_i \\ \mathbf{h} \end{bmatrix}
\end{aligned}
\tag{1}
$$

$$(\mathbf{h}', \mathbf{m}') = \mathrm{LSTM}(\mathbf{H}, \mathbf{m}, \mathbf{W})$$

$\mathbf{W}\mathbf{W}^u, \mathbf{W}^f, \mathbf{W}^o, \mathbf{W}^c$

(N)

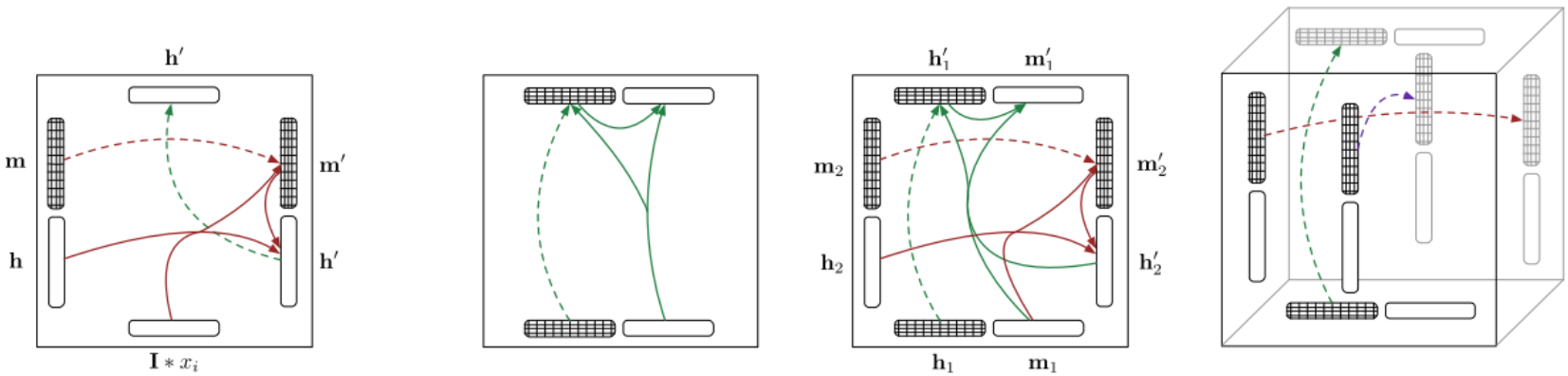


Figure 6: [Kalchbrenner et al, Grid LSTM, ICLR 2016]

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_i \\ \vdots \\ \mathbf{h}_N \end{bmatrix} \tag{2}$$

$$\begin{aligned} (\mathbf{h}'_1, \mathbf{m}'_1) &= \text{LSTM}(\mathbf{H}, \mathbf{m}_1, \mathbf{W}_1) \\ &\vdots \\ (\mathbf{h}'_N, \mathbf{m}'_N) &= \text{LSTM}(\mathbf{H}, \mathbf{m}_N, \mathbf{W}_N) \end{aligned} \tag{3}$$

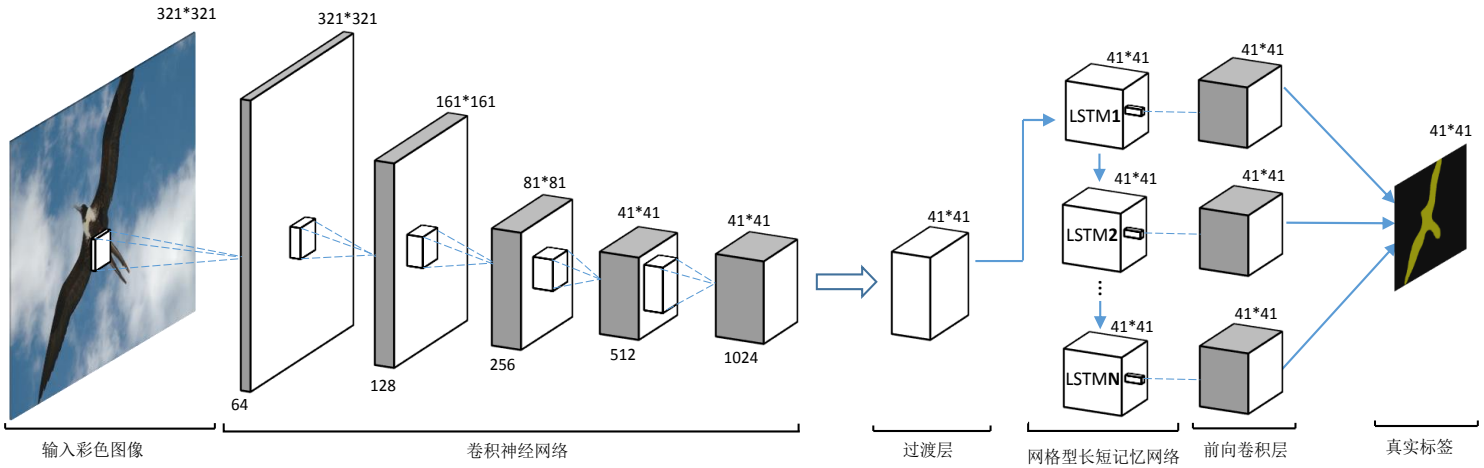

321*321
321*321
161*161
81*81
41*41
41*41
64
128
256
512
1024
41*41
41*41
LSTM1
41*41
41*41
LSTM2
41*41
41*41
LSTMN
41*41
输入彩色图像
卷积神经网络
过渡层
网格型长短记忆网络
前向卷积层
真实标签

Figure 7:

- :
-

- VGG_{16}[1], 16
- 6.5 [2]

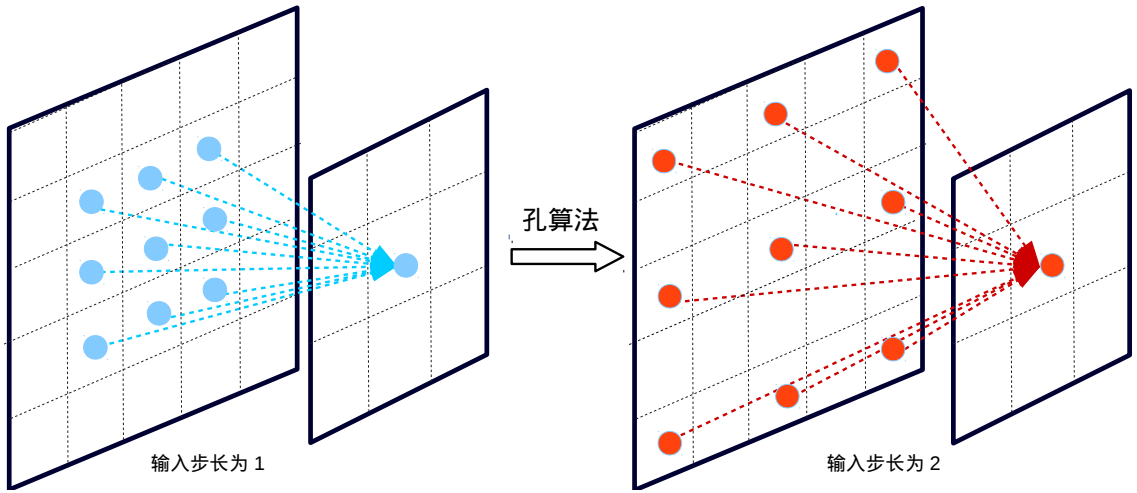


Figure 8: ””

[1] Simonyan & Zissermanet, Very deep Convolutional Networks For Large-scale Image Recognition, ICLR 2015

[2] Chen et al, DeepLab-LargeFOV, ICLR 2015

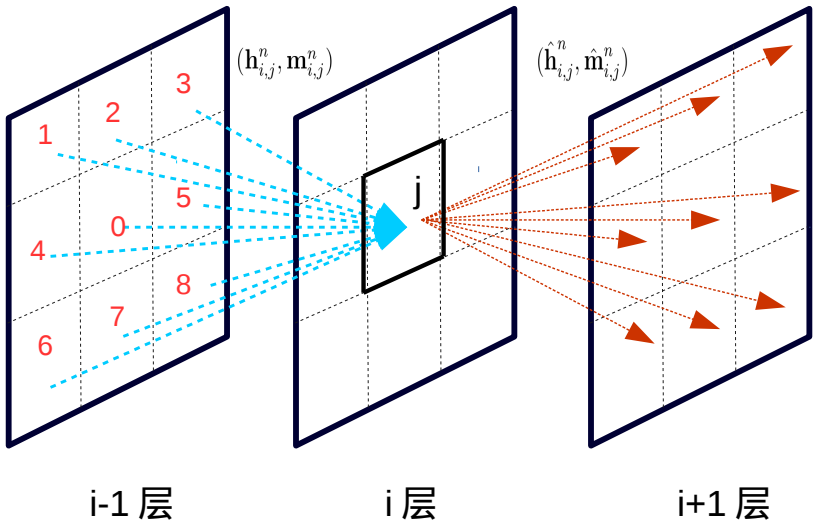


Figure 9:

$$
\begin{aligned}
(\hat{\mathbf{h}}^0_{i,j}, \hat{\mathbf{m}}^0_{i,j}) &= \mathrm{LSTM}(\mathbf{H}_{i,j}, \mathbf{m}^0_{i,j}, \mathbf{W}_i) \\
(\hat{\mathbf{h}}^1_{i,j}, \hat{\mathbf{m}}^1_{i,j}) &= \mathrm{LSTM}(\mathbf{H}_{i,j}, \mathbf{m}^1_{i,j}, \mathbf{W}_i) \\
&\vdots \\
(\hat{\mathbf{h}}^N_{i,j}, \hat{\mathbf{m}}^N_{i,j}) &= \mathrm{LSTM}(\mathbf{H}_{i,j}, \mathbf{m}^N_{i,j}, \mathbf{W}_i) \\
\mathbf{H}_{i,j} &= [\mathbf{h}^0_{i,j}\ \mathbf{h}^1_{i,j}\ \ldots\ \mathbf{h}^N_{i,j}]^T
\end{aligned}
\tag{4}
$$

VOC 2012

SIFT FLOW

Pascal VOC 2012 & SIFT FLOW

Figure 10: VOC 201210582 1464145621

Figure 11: SIFT FLOW248820033

- 321*321
- ,

$$n_{ij}\,ij\,n_{cl}\,t_i = \sum_{j=1}^{n_{cl}} n_{ij}\,i$$

$$= \sum_{i=1}^{n_{cl}} n_{ii} / \sum_{i=1}^{n_{cl}} t_i$$

$$= \frac{1}{n_{cl}} \sum_{i=1}^{n_{cl}} (n_{ii}/t_i) \tag{5}$$

$$\text{Mean IU} = \frac{1}{n_{cl}} \sum_{i=1}^{n_{cl}} \frac{n_{ii}}{t_i + \sum_j^{n_{cl}} n_{ji} - n_{ii}}$$

../image/result/combine-eps-converted-to.pdf

† 57.5%

Figure 12:

Method	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	shep	sofa	train	tv	mIoU.
SDS[3]	63.3	25.7	63.0	39.8	59.2	70.9	61.4	54.9	16.8	45.0	48.2	50.5	51.0	57.7	63.3	31.8	58.7	31.2	55.7	48.5	51.6
FCN-8s[4]	76.8	34.2	68.9	49.4	60.3	75.3	74.7	77.6	21.4	62.5	46.8	71.8	63.9	76.5	73.9	45.2	72.4	37.4	70.9	55.1	62.2
TTI-zoomout-16[5]	**81.9**	35.1	**78.2**	**57.4**	56.5	80.5	74.0	79.8	22.4	69.6	53.7	74.0	**76.0**	76.6	68.8	44.3	70.2	40.2	68.9	55.3	64.4
DeepLab-CRF[6]	78.4	33.1	**78.2**	55.6	**65.3**	81.3	75.5	78.6	**25.3**	69.2	52.7	**75.2**	69.0	**79.1**	**77.6**	**54.7**	78.3	45.1	73.3	56.2	66.4
CNN+**5**LSTM	80.2	**35.3**	74.1	54.4	64.4	**87.3**	**81.1**	**80.6**	22.7	**73.6**	**58.8**	73.9	73.7	78.7	77.4	50.2	**80.0**	**47.9**	**76.5**	**63.1**	**67.9**

Table 1: VOC2012

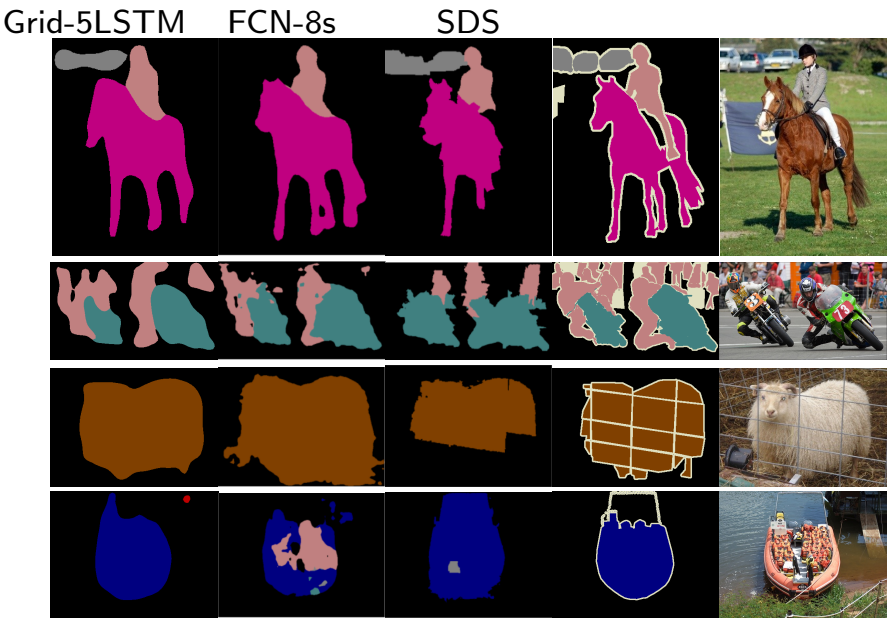
†

[3] Simultaneous Detection and Segmentation, ECCV 2014

[4] Fully convolutional networks for semantic segmentation, CVPR 2015

[5] Feedforward semantic segmentation with zoom-out features, CVPR 2015

[6] Semantic image segmentation with deep convolutional nets and fully connected crfs, ICLR 2015



(a)

(b)

TTI-zoomout-16,DeepLab-CRFGrid-5LSTM

Figure 14: Grid-5LSTMVOC 2012

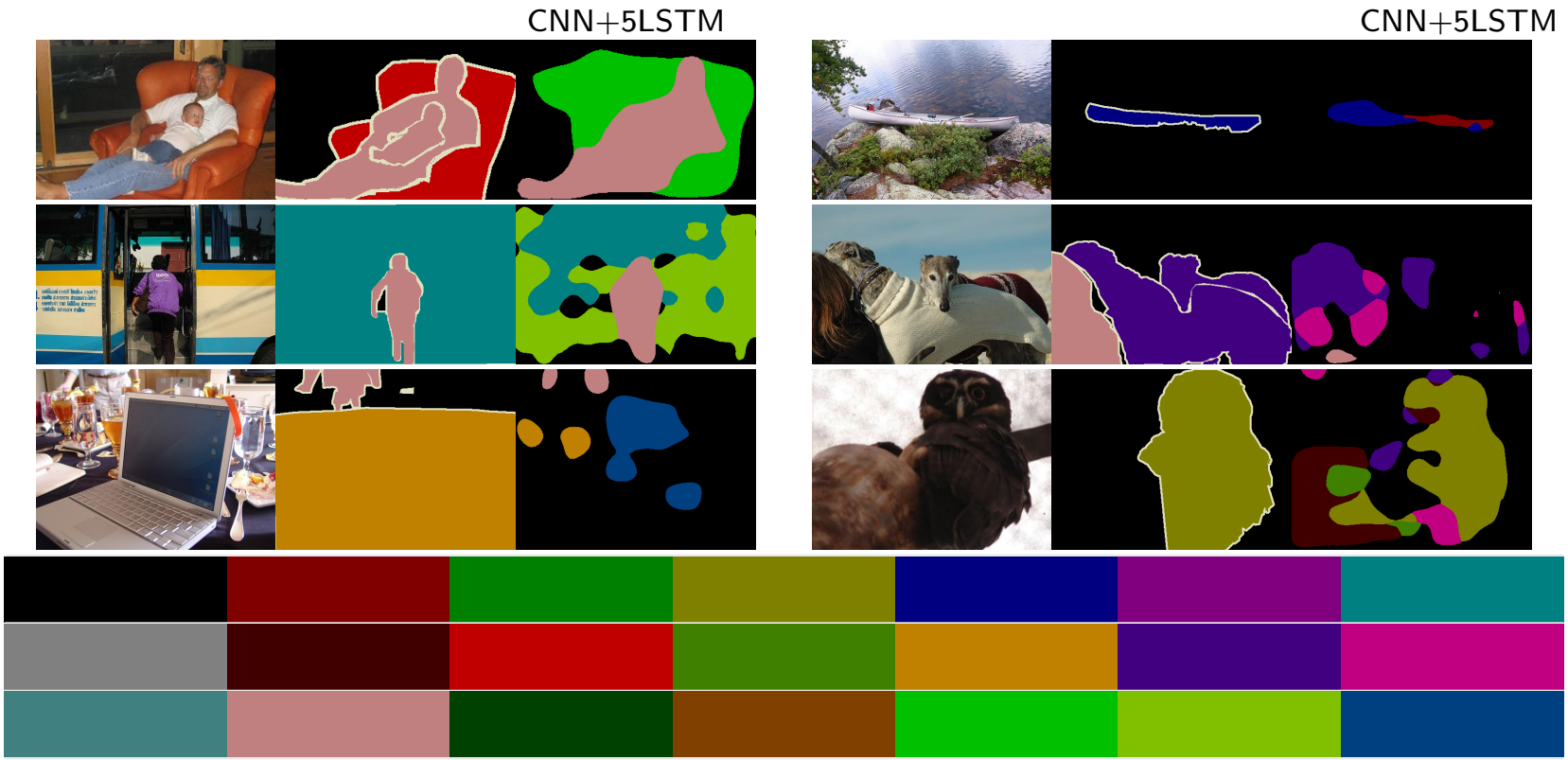


Figure 15: CNN+5LSTM

SIFT FLOW

Method	Pixel Acc.	Mean Acc.	Mean IU.
Liu et al.[7]	76.7	-	-
Tighe et al.[8]	78.6	39.2	-
FCN-16s[9]	85.2	**51.7**	39.5
Deeplab-LargeFOV[10]	85.6	51.2	39.7
Grid-5LSTM	**86.2**	51.0	**41.2**

Table 2: SIFT FLOWTigheSVM+MRFDeeplab-LargeFOV

[7] Sift flow: Dense correspondence across scenes and its applications, PAMI 2011

[8] Finding things: Image parsing with regions and per-exemplar detectors, CVPR 2013

[9] Fully convolutional networks for semantic segmentation, CVPR 2015

[10] emantic image segmentation with deep convolutional nets and fully connected crfs, ICLR 2015

†

†

†

- † (He et al. ResNet, CVPR 2016)
- † (Han et al. Deep Compression, ICLR 2016 Best Paper)(Courbariaux et al. Binaryconnect, NIPS 2015)
- † (Papandreou et al. Weakly-and semi-supervised learning, ICCV 2015)

Q & A

Questions?

Thank you!