Semantic Segmentation for Real-World Data by Jointly Exploiting Supervised and Transferrable Knowledge

Li-Hsien Lu and Chiou-Ting Hsu

Dept. of Computer Science, National Tsing Hua University, Taiwan

Introduction

Parametric methods

- Achieve good performance
  - on datasets with a moderate size of labels

Nonparametric methods

- More efficient and adaptive to real-world data

Idea

Take advantage of both methods for increasing real-world data

Proposed scenario

Given one FCN-based parametric model

- Off-line trained with a fixed set of labels

Given an annotated real-world dataset

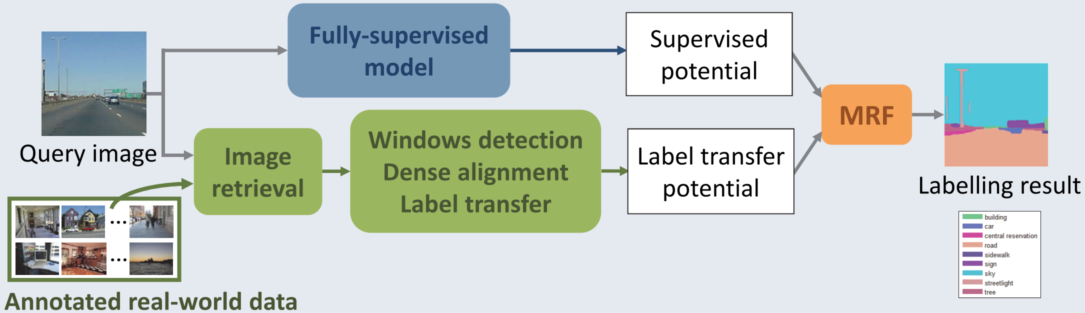
- with labels unknown to parametric model

$C_{fs} \neq C_r$

Label set already known to parametric model

Real-world dataset

Markov Random Field (MRF) Framework



$$E(\boldsymbol{c}) = -\sum_{\boldsymbol{p}\in I}\left[(1-\alpha(I))\cdot\psi_{fs}(c,\boldsymbol{p}) + \alpha(I)\beta(c,I)\cdot\psi_{trans}(c,\boldsymbol{p})\right] + \lambda\sum_{(\boldsymbol{p},\boldsymbol{q})\in\varepsilon}\theta(c(\boldsymbol{p}),c(\boldsymbol{q}))$$

Supervised potential

- Off-line trained FCN

$$\psi_{fs}(c,\boldsymbol{p}) = \begin{cases} \dfrac{e^{\mathbf{M}_c(\boldsymbol{p})}}{\sum_{d\in\{1,\dots,|C_{fs}|\}} e^{\mathbf{M}_d(\boldsymbol{p})}}, & \text{for } c\in C_{fs} \\ 0, & \text{otherwise} \end{cases}$$

λ	: a smoothing constant
ε	: the set of adjacent pixels
$c(\boldsymbol{p})$	: the label of the pixel $\boldsymbol{p}$
W	: the set of query windows
W_{rs}	: the set of reference windows
w	: a query window
w_i	: a matched window of w
$\widetilde{w}_i$	: the resized w_i
$\widehat{\boldsymbol{p}}$	: the window-centric coordinates of $\boldsymbol{p}$ in w
$\boldsymbol{f}_i$	: the SIFT flow vector from w to $\widetilde{w}_i$
$L(\cdot,\cdot)$	: transfers the label from $\widetilde{w}_i$ to w

Label transfer potential

- Window-based label transfer
  - Fine-tuned Faster R-CNN
- Dense alignment between similar windows

Penalize large matched windows

Reflect the rareness of the label c

$$\psi_{trans}(c,\boldsymbol{p}) = \begin{cases} \sum_{w\in W}\sum_{w_i\in W_{rs}} \delta[L(\widetilde{w}_i(\boldsymbol{p}+\boldsymbol{f}_i),\widehat{\boldsymbol{p}}) = c]\boxed{\phi_{size}(w_i)}\boxed{\phi_{idf}(c)}, & \text{for } c\in C_r \\ 0, & \text{otherwise} \end{cases}$$

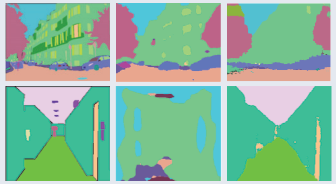
Pairwise potential term

$$\theta(c(\boldsymbol{p}),c(\boldsymbol{q})) = -\log[(P(c(\boldsymbol{p})|c(\boldsymbol{q})) + P(c(\boldsymbol{q})|c(\boldsymbol{p})))/2]\cdot\delta[c(\boldsymbol{p})\neq c(\boldsymbol{q})]$$

Challenge

- FCN model is unaware of unknown labels $c\in(C_r - C_{fs})$

Observation



Consistent: ψ_{fs} is reliable

Inconsistent: ψ_{fs} is unreliable

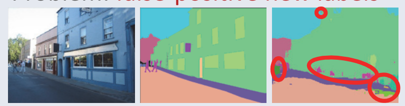
Ground truth | Estimation by $\psi_{fs}(c,\boldsymbol{p})$ | Estimation by $\psi_{trans}(c,\boldsymbol{p})$

Content-adaptive parameter $\alpha(I)$

- Adaptive to different query image

$$\alpha(I) = \frac{|\hat{C}_r(I) - \hat{C}_{fs}(I)|}{|\hat{C}_r(I) \cup \hat{C}_{fs}(I)|} \quad (4)$$

- Problem: false-positive new labels



Number of pixels estimated as new labels

$$\alpha(I) = \frac{\ln(\#pixels(\hat{C}_r(I) - \hat{C}_{fs}(I)))}{\ln(m\times n)} \quad (5)$$

Label-aware parameter $\beta(I)$

- Inverse proportion to # of pixels estimated as the label c

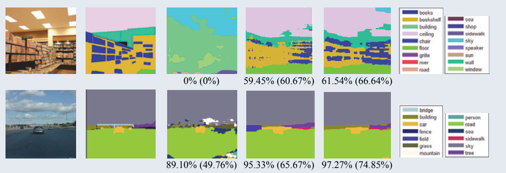
$$\beta(c,I) = \begin{cases} 2 - \dfrac{\#pixels(c)}{\max_{\forall c\in\hat{C}_r(I)}\#pixels(c)}, & \text{if } (c\in\hat{C}_r(I)) \text{ and } \left(0.01 \leq \dfrac{\#pixels(c)}{m\times n} \leq 0.2\right) \\ 1, & \text{otherwise} \end{cases}$$

Noisy new labels

Pareto principle

Experimental Results

Method	# labels	α	β	Per-pixel acc. (%)	Per-class acc. (%)
LMSun dataset $|C_r| = 232$					
FCN-16s-siftflow	$|C_{fs}| = 33$, $|C_{fs}\cap C_r| = 32$	-	-	53.1	5.2
Proposed method *FCN-16s-siftflow*		0.5		55.2	7.1
		Eq. (4)		65.2	14.3
		Eq. (5)		**66.0**	16.3
		Eq. (5)	✓	65.9	**16.4**



Method		Per-pixel acc. (%)	Per-class acc. (%)
SIFT Flow dataset $|C_r| = 33$			
Nonparametric method	C. H. Ma et al. [11]	78.3	46.1
	F. Tung et al. [14]	77.1	41.1
	F. Tung et al. [29]	79.9	49.3
Parametric + nonparametric method	B. Shuai et al. [10]	80.1	39.7
	M. George [16]	81.7	50.1
Proposed method ($C_{fs}\neq C_r$)	*FCN-8s-pascal*	81.6	50.1
	FCN-8s-pascal-context	79.9	47.4
Parametric method	A. Sharma et al. [2]	75.5	**52.8**
	J. Long et al. [1]	**85.6**	50.1
Proposed method ($C_{fs} = C_r$)	*FCN-16s-siftflow*	85.2	52.4
LMSun dataset $|C_r| = 232$			
Nonparametric method	F. Tung et al. [29]	60.8	**19.3**
Parametric + nonparametric method	M. George [16]	61.2	16.0
	J. Yang et al. [15]	60.6	18.0
Proposed method ($C_{fs}\neq C_r$)	*FCN-16s-siftflow*	**65.9**	16.4
	FCN-32s-pascal-context	65.0	17.0
	FCN-8s-pascal-context	65.3	17.3

$|C_r - C_{fs}| = 26$

$|C_r - C_{fs}| = 200$

$|C_r - C_{fs}| = 177$