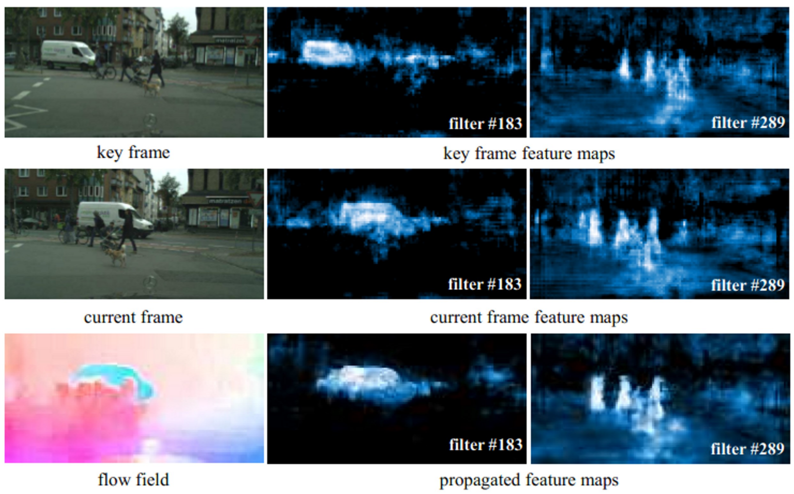
summary

2018年11月14日 15:26

deep feature flow

Motivation 来自于一个实验现象.



一般把目标检测任务分2段.

$f_i = N_{feat}(I_i)$ 提特征. (1)

$y_i = N_{task}(f_i)$ 检测/分割 (2)

过程(1) 提特征是较慢的.

实验发现 用光流场信息和关键帧融合.

可以用小开销在非关键帧上代替(1)

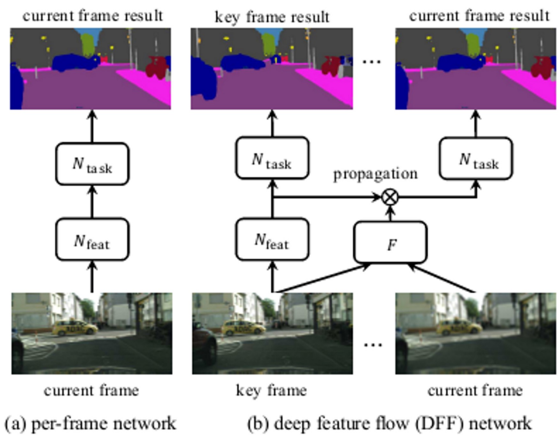


Figure 2. Illustration of video recognition using per-frame network evaluation (a) and the proposed deep feature flow (b).

Algorithm 1 Deep feature flow inference algorithm for video recognition.

input: video frames $\{\mathbf{I}_i\}$
$k = 0$; ▷ initialize key frame
$\mathbf{f}_0 = \mathcal{N}_{feat}(\mathbf{I}_0)$
$\mathbf{y}_0 = \mathcal{N}_{task}(\mathbf{f}_0)$
for $i = 1$ **to** ∞ **do**
  if $is\_key\_frame(i)$ **then** ▷ key frame scheduler
    $k = i$ ▷ update the key frame
    $\mathbf{f}_k = \mathcal{N}_{feat}(\mathbf{I}_k)$
    $\mathbf{y}_k = \mathcal{N}_{task}(\mathbf{f}_k)$
  else ▷ use feature flow
    $\mathbf{f}_i = \mathcal{W}(\mathbf{f}_k, \mathcal{F}(\mathbf{I}_k, \mathbf{I}_i), \mathcal{S}(\mathbf{I}_k, \mathbf{I}_i))$ ▷ propagation
    $\mathbf{y}_i = \mathcal{N}_{task}(\mathbf{f}_i)$
  end if
end for
output: recognition results $\{\mathbf{y}_i\}$

The feature network N_{feat} only runs on sparse key frames.

The feature maps of a non-key frame I_i are propagated from its preceding key frame I_k.

The feature in the deep conv layers encode the semantic concepts and correspond to spatial locations in the image [47].
Examples are illustrated in Figure 1. Such spatial correspondence allows us to cheaply propagate the feature maps by the manner of spatial warping.

The feature maps of non-key frame I_i is propagated from the key frame to current frame via flow field.

当前帧的特征图通过前一个关键帧和光流场 $F\{I_i, I_k\}$ 融合，传播得到.

首先估计流场 $M_{i\to k} = F(I_k, I_i)$，

对于每一点 p 有 $\delta p = M_{i\to k}(p)$

然后通过双线性插值进行融合

$$f_i^c(p) = \sum_q G(q, p+\delta p) f_k^c(q) \qquad (1a)$$

$$G(q, p+\delta p) = g(q_x, p_x+\delta p_x)\cdot g(q_y, p_y+\delta p_y) \qquad (1b)$$

$$\text{where } g(a,b) = \max(0, 1-|a-b|) \qquad (1c)$$

这个操作参考 STN.

(1a) 就是 f_k 到 f_i 根据 flow 插值 (详细推导见. propagation)

融合过后 作者又说.

"To better approximate the features, their amplitudes are modulated by a "scale field" $S_{i\to k}$, which is of the same spatial and channel dim as the feature maps.

$S_{i\to k} = S(I_k, I_i)$ 这么神密，究竟是什么？

大概也能猜到就是一个标量矩阵

看了源码 就是用$\underbrace{\text{一个卷积}}_{1\times 1}$实现的.

感兴趣的看pdf文件conv_feats