

Spatio-temporal feature gating

实验结果

CBAM(ECCV2018)

基于 SE-Net [5]中的 Squeeze-and-Excitation module 来进行进一步拓展

Convolutional Block Attention Module and Spatial attention module

CoST(CVPR2019)

卷积操作:

Collaborative SpatioTemporal (CoST)

Connection to C2D and C3D

TSM(CVPR2019)

NLNet(CVPR2018)

SENet: (CVPR2019)

DANet(CVPR2019)

Position attention module:

Channel Attention Module

[行为识别链接](#)

主要参考 ¹

行为识别主要包含两个方向,

1. 动作分类, 给出一个视频截断, 判断视频的动作类别, 或者称为offline。
2. 动作识别, 给出一个自然视频, 没有进行任何的裁剪, 这个时候需要先知道动作的开始时间和结束时间, 然后还要知道动作的类别

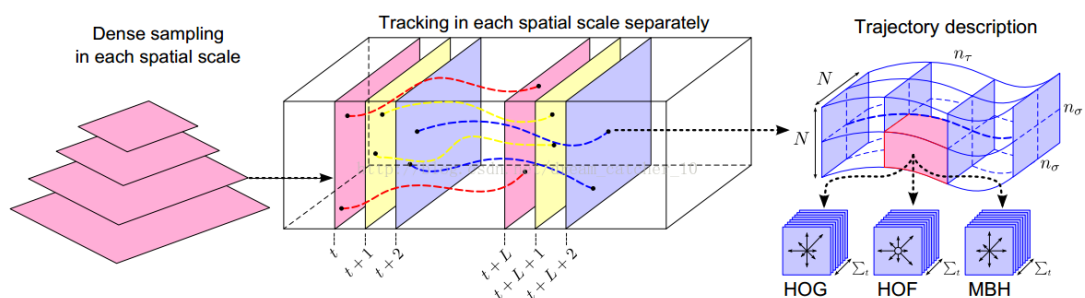
1. iDT算法(2013)

Dense Trajectories and Motion Boundary Descriptors for Action Recognition

Action Recognition with Improved Trajectories

密集轨迹算法(Dense Trajectories算法):Improve dense trajectory简称iDT, 是一种用来提取视频密集跟踪轨迹的算法; 通常基于该轨迹进行取块计算descriptor

算法基本框架



如图所示即为算法的基本框架，包括：

1. 密集采样特征点，
2. 特征点轨迹跟踪
3. 基于轨迹的特征提取

1.1. 密集采样

1. DT方法通过网格划分的方式在图片的**多个尺度上分别密集采样特征点**。

在多个空间尺度上采样能保证采样的特征点覆盖了所有空间位置和尺度，通常8个空间尺度已经非常足够了，若图像很大，可以适当增加。后续的特征提取也是在各个尺度上分别进行的。

2. 特征点采样的间隔(即网格的大小)W通常取W=5。

多尺度的解释



对图片进行缩放以便对各种尺度的对象都能进行识别。

1.2. 特征点轨迹跟踪

在视频序列中对每一帧的兴趣点进行跟踪就形成trajectory，若是对每一帧密集采样兴趣点进行跟踪就形成dense trajectory；

1. 对采样点进行跟踪：由光流判断跟踪点在下一帧的位置；

$$P_{t+1} = (x_{t+1}, y_{t+1}) = (x_t, y_t) + (M * \omega)|_{(\bar{x}_t, \bar{y}_t)}, \quad M$$

为均值滤波器，w为光流场

2. 对每个点跟踪都会形成一条trajectory，**为了避免长时间跟踪而产生的跟踪点漂移现象**，可以对跟踪的长度L进行约束 (L=15) $(P_t, P_{t+1}, P_{t+2}, \dots)$

这里也就暴露了该方法不能处理长视频的原因

3. 现实视频中存在摄像头运动的缺陷，因此需要相应算法消除摄像头影响，得到最终的iDT；

1.3. 基于iDT计算descriptor

四种特征：

1. trajectory轨迹特征：

每条trajectory都可以提取一个轨迹特征向量 S' （当 $L=15$ ， S' 为30维），对局部动作模式进行编码

$$\Delta P_t = (P_{t+1} - P_t) = (x_{t+1} - x_t, y_{t+1} - y_t).$$

$$S = (\Delta P_t, \dots, \Delta P_{t+L-1})$$

$$S' = \frac{(\Delta P_t, \dots, \Delta P_{t+L-1})}{\sum_{j=t}^{t+L-1} \|\Delta P_j\|}.$$

2. HOG特征：

HOG特征计算的是**灰度图像梯度的直方图**。对视频块进行表面特征描述。直方图的bin数目为8。所以HOG特征的长度为 $2 \times 2 \times 3 \times 8 = 96$ 。

3. HOF特征：

HOF计算的是**光流的直方图**。直方图的bin数目取为 $8+1$ ，前8个bin与HOG都相同。额外的一个用于统计光流幅度小于某个阈值的像素。故HOF的特征长度为 $223 \times 9 = 108$ 。

4. MBH特征：

MBH计算的是**光流图像梯度的直方图**，也可以理解为在光流图像上计算的HOG特征。由于光流图像包括X方向和Y方向，故分别计算MBHx和MBHy。MBH总的特征长度为 $2 \times 96 = 192$ 。最后进行特征的归一化，DT算法中对HOG，HOF和MBH均使用L2范数进行归一化。

1.4. 特征编码和分类过程

1. 特征编码—Bag of Features
2. 分类-SVM

2. Two-Stream(NIPS2014)