doi:10.3772/j.issn.1002-0470.2015.04.013

基于关键观测点选择的视频浓缩方法①

祝晓斌② 范芳鑫 徐英瀚 谭 励

(北京工商大学计算机与信息工程学院 北京 100048)

摘 要 为了从海量视频数据中快速提取感兴趣的信息,在研究、分析现有视频浓缩算法性能的基础上,提出了一种基于关键目标选择的视频浓缩方法。该方法用选择的代表性观测点组成的新的运动序列来代表目标原运动序列,从而消除了现有算法不能消除的内容上的冗余,进而提高视频浓缩效率;采用数据驱动方式进行关键观测点选择,通过把这一选择问题转换为最小描述长度(MDL)选取问题来实现自适应选择,从而克服现有算法在视频浓缩中因观测目标极多导致压缩效率下降和影响视觉效果的问题。通过在三个数据库上的试验,证明了该方法的有效性。

关键词 视频浓缩, 关键观测点选择, 最小描述长度(MDL), 数据驱动

0 引 言

随着视频监控技术的发展,遍布大街小巷的监控摄像机不停地摄入海量的视频数据。然而,对海量视频数据的查找、分析,常常会耗用大量的时间和人力。例如,警方为尽快破案,需要动用数百名民警24小时不间断工作,从上万小时的视频中找到仅十几秒的有用片段。由此可见,如何在视频中快速检索到感兴趣的目标,是当前智能视频监控的重要研究课题。

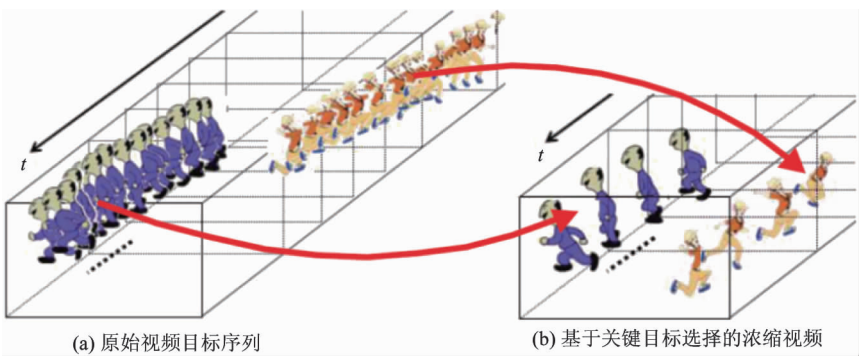
视频摘要(video summarization)技术是从视频中快速查找到感兴趣目标的可行方法。视频摘要技术利用模式识别、机器学习等算法,对视频文件的内容进行分析和处理,提取用户感兴趣的信息,生成一个高压缩率的能代表原始视频文件信息的文件。视频摘要可以极大程度地节省存储空间,同时保留原始视频的关键内容,可方便地实现对视频事件的快速浏览和检索。视频摘要技术大体上可以分为两类:基于关键帧的视频摘要和基于目标运动信息的视频摘要。文献[1]对基于关键帧的动态视频,选取用户定义的感兴趣帧作为关键帧,然后基于关键帧自适应调整视频播放速度。基于关键帧的视频摘要[2]虽然极大地压缩了视频,但是它丢失了视频的动态特性。视频缩略[3]是一种能在一定程度上保留视频动态特性的视频摘要。视频缩略技术从原始视频中提取关键视频片段,然后将这些片段利用淡入淡出等效果链接起来形成视频摘要。这种方法虽然在一定程度上保留了视频的动态变化过程,但是舍弃的视频段中,可能包含重要信息。Smith[3]等人将视频中信息贫乏的段节舍弃,而利用信息丰富的视频段拼接形成新的视频(video skimming),输出类似电影预告片的效果。近几年,以色列希伯来大学的Rav-Acha等人[4]提出了一种基于目标运动信息的视频浓缩技术(video synopsis)。视频浓缩[4-7]打破了传统视频摘要体系,通过优化算法对运动目标的时间上进行改变,保留了运动目标的空间位置。在此工作基础上,Feng等人[5]基于轮盘赌算法,提出了在线视频浓缩技术。Huang等人[8]基于最大后验概率对运动目标进行在线匹配,并基于二维图填充的方式进行在线视频浓缩。视频浓缩算法是一种

① 国家自然科学基金(61402023),北京市自然科学基金(4132025)和北京市教师队伍建设青年英才计划(YETP1448)资助项目。
② 男,1981年生,博士,讲师;研究方向:模式识别,视频分析等;联系人,E-mail: buddysoft@sina.com
(收稿日期:2015-02-26)

高效的视频摘要算法，它不仅消除了原视频中时空上的冗余，保留动态变化特性，还能生成高压缩率的摘要视频。但是，现有的视频浓缩算法，忽略了内容上的冗余。此外，在视频浓缩中，太多目标观测点容易降低视频浓缩的压缩率并使浓缩视频杂乱。针对这种情况，本文提出了一种基于关键观测目标选择的视频浓缩方法。

1　方法特点描述

本文中，属于同一目标的完整运动行为，称为对象序列，如图 1 所示。每个对象序列包含很多按时序排列组成的观测点，每个观测点为一个运动目标的前景区域，太多的观测点容易造成内容冗余和目标之间的重叠，如图 1(a)所示。本文方法通过数据驱动方式自适应选择关键观测点，组成新的对象序列来代表原始对象序列。基于关键目标选择的浓缩视频，压缩率更高而且主观视觉更好，如图 1(b)所示。



(a) 原始视频目标序列　　(b) 基于关键目标选择的浓缩视频

图 1　目标序列及浓缩视频

2　关键观测点选择

本研究首先采用混合高斯模型进行背景建模，然后用背景分割法[9]提取目标前景，最后通过跟踪匹配算法[10]得到同一目标的运动序列。同一目标在视频中完整的运动有非常多的观测点，而相邻的观测点通常具有相似的行为和外观。本研究在运动目标完整运动序列中，选择一些关键观测点用来代表原始对象序列[11]。如图 2 所示，由关键观测点组成的新对象序列，可以很好地代表原始的对象序列。

在文献[12]视频摘要工作中，通过聚类的方式选取固定数目的关键测点。但是，即使在同一固定场景中，目标运动方式不一样，关键点的数量不能提前确定。关键观测点选择的标准是选取有明显行为变化的对象。对象序列为一个时间序列 $[t_b^s, t_b^e]$ 时间轴为帧，观测点为序列中的每个点，第 n 条对象序列表示为 $b_n = [O_s, O_{s+1}, O_{s+2}, \cdots, O_e]$，其中 O_i 和 O_j

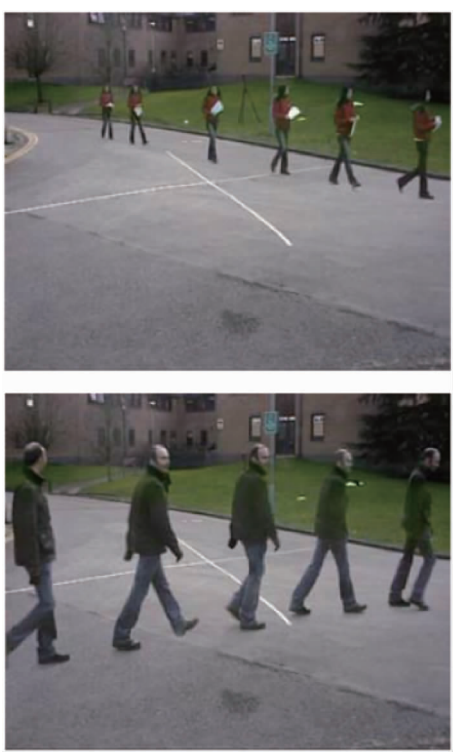


图 2　基于关键观测点的新对象序列展示

代表对象序列中的两个不同观测点。本文算法的目标是选择尽可能少的观测点(压缩性),而且使新的运动序列和原始运动序列尽可能相似(相似性),如图 3 所示。

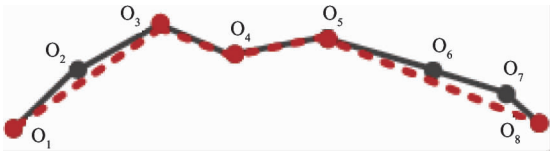


图 3　虚线连接的为关键观测点

但是压缩性和相似性互相矛盾,如果选取所有的点为关键观测点,那相似性最大而压缩性最小。为了解决这个问题,本文算法采用数据驱动的方法来自适应选取关键观测点,并把问题转换为最小描述长度[13](minimum description length, MDL)求解问题。描述长度(description length, DL)通过公式

$$DL = L(S) + L(D \mid S) \tag{1}$$

计算,式中 S 为关键观测点选择模型,D 为某一对象运动序列。$L(S)$ 代表关键点选择模型的描述长度,而 $L(D \mid S)$ 代表在已定模型下数据的描述长度。针对本文算法,本研究对 $L(S)$ 和 $L(D \mid S)$ 分别做如下定义:

$$L(S) = -\log \frac{N_k}{N} \tag{2}$$

$$L(D \mid S) = -\log \frac{\sum_{i=1}^{N_{k-1}} \left(\sum len(O_j O_{j+1}) - len(O'_i O'_{i+1}) \right)}{R_{\max}} \tag{3}$$

式中 N 代表观测点总数,N_k 代表关键观测点数量,O'_i 代表关键观测点,$R_{\max}$ 为归一化因子(只选起始和结束点),$len(O_j O_{j+1})$ 关键观测点之间的点的距离。

本研究需要选取最佳的方案,使 $L(D,S)$ 最小,即得到最小描述长度(MDL)。但这是个 NP 难题,本文只能通过一种局部最优取代全局最优的近似方法去求解[14]。如图 4 所示,$DL(O_k O_{k+3}) < DL(O_k O_{k+4})$),则 O_{k+3} 被选为 O_k 后的下一个关键观测点。

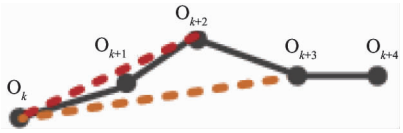


图 4　近似算法例子

具体的算法流程如下:

算法 1　基于最小描述长度(MDL)的关键观测点选择

输入:N,运动序列中观测点数量
输出:由关键观测点组成的新的运动序列 O'
数据:$[O_s, O_{(s+1)}, \cdots, O_e]$.

```
O'_1 = O_s;
p = 2;
for {k = 1; k < N − 1; k ++} do
  j = 1
  for {::} do
    if DL(O_k O_{k+j}) < DL(O_k O_{k+j+1}) then
      O'_p = t_{k+1}, p ++
      continue;
    else
      j ++;
    end if
  end for
  k = k + j − 1;
end for
O'_p = O_e;
```

3　视频浓缩算法

与文献[6]中所述方法一样,本文算法也引入了时间乱序损耗、目标丢失损耗和遮挡损耗,视频浓缩损耗函数定义如下:

$$E = \arg\min_{\forall M} E(M) \tag{4}$$

$$E(M) = \sum_{b_n, b'_n \in B} (\alpha E_\alpha(b_n, b'_n) + \beta E_t(b_n, b'_n) + \gamma E_c(b_n, b'_n)) \tag{5}$$

式中 b_n 和 b'_n 为映射到浓缩视频中的两条目标对象序列,B 为对象序列集合,M 为浓缩视频中对象序列的映射。α, β, γ 为三个损耗权重,根据经验设定。E_a 为目标丢失损耗,$E_t(b_n, b'_n)$ 为时间乱序损耗,$E_c(b_n, b'_n)$ 为遮挡损耗,每项损耗的具体定义如下:

(1)目标丢失损耗

引入目标丢失损耗,是为了尽量不丢失原始视频中的运动目标信息,定义如下:

$$E_a(b_n) = \sum_{b \in B} \chi_b(x, y, t) \tag{6}$$

式中$\chi_b(x, y, t)$代表对象的外观特征函数,定义如下:

$$\chi_b(x, y, t) = \begin{cases} \| I(x,y,t) - B(x,y,t) \|, & t \in t_b \\ 0, & \text{其他} \end{cases} \tag{7}$$

式中 $B(x,y,t)$ 是背景图像的像素, $I(x,y,t)$ 是各自图像的像素, t_b 是目标出现的时间。

(2) 时间一致性损耗

定义时间乱序损耗,是为了尽量保持对象序列之间原有时间顺序。例如,两辆车有先后顺序,优化算法会尽量保持它们的出现顺序。两个对象序列时间上的交叉可以从它们的时空距离上得到,公式如下:

$$d(b_n, b'_n) = \exp(-\min_{t \in t_{b_n} \cap t_{b'_n}} \{d(b_n, b'_n, t)\} / \sigma_{\text{space}}) \tag{8}$$

式中 $d(b_n, b'_n, t)$ 为 b_n 和 b'_n 在第 t 帧中的对象序列的最近前景目标像素点距离, σ_{space} 定义了对象序列之间的空间互动程度。

如果 b_n 和 b'_n 在浓缩视频中没有时间重叠, b_n 比 b'_n 映射后的时间要早,那么它们的互动程度以时间的指数级减少:

$$d(b_n, b'_n) = \exp(-(t^s_{b'_n} - t^e_{b_n}) / \sigma_{\text{time}}) \tag{9}$$

式中 σ_{time} 定义了对象序列之间的时间互动程度。引入时间一致性损耗,是为了保持对象序列在原始视频上时间顺序设置,对破坏时间关系的映射进行惩罚:

$$E_t(b_n, b'_n) = \begin{cases} 0, & t^s_{b'_n} - t^s_{b_n} = \hat{t}^s_{b'_n} - \hat{t}^s_{b_n} \\ C \times d(b_n, b'_n), & \text{其他} \end{cases} \tag{10}$$

式中 C 为加权因子, $\hat{t}^s_{b'_n}$ 和 $\hat{t}^s_{b_n}$ 为对象序列在原始视频中的起始时间。

(3)目标遮挡损耗

在浓缩视频的映射过程中,经过时间移动的对象序列,如果观测点之间有重叠,则我们定义损耗如下:

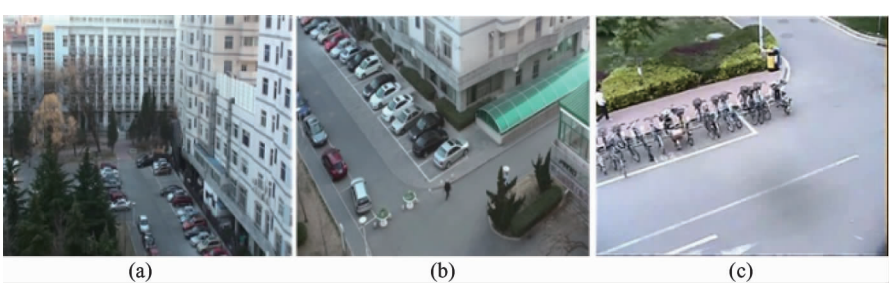
$$E_c(b_n, b'_n) = \sum_{x,y,t \in t_b \cap t'_b} \chi_{b_n}(x,y,t) \chi_{b'_n}(x,y,t) \tag{11}$$

式中 $t_b \cap t'_b$ 为 t_b 和 t'_b 在浓缩视频中的时间交叉。定义目标遮挡损耗是为了尽量避免浓缩视频中目标之间的重叠。

最后,我们用模拟退火法对能量损耗函数进行优化,得到最优排列映射,然后用 Poisson Editing[15] 算法把运动序列融合到背景图像中,得到浓缩视频。

4 实验结果与分析

为了验证算法的有效性,我们录制了 3 个室外场景的视频数据进行试验。第 1 个视频数据总共 31530 帧,代表图像如图 4(a)所示。第 2 个视频数据总共 43685 帧,代表图像如图 4(b)所示。第 3 个视频数据总共 39556 帧,代表图像如图 4(c)所示。视频统一缩放到的分辨率,帧率 15fps。



(a) (b) (c)

图 4 场景代表性图像

本文的算法同文献[6]算法(标记为方法 1)和文献[7]算法(标记为方法 2)进行了比较,详细的数据如表 1、表 2 和表 3 所示。实验中,所有算法损耗都考虑了目标丢失损耗、时间乱序损耗和遮挡损

耗。图 5 为视频浓缩结果的代表性图像。从表 1 我们可以看出,本文算法得到5.6% 的压缩率,只造成4869 的能量损耗;方法 1 得到7.4% 的压缩率,却造成 5538 的能量损耗;方法 2 得到 6.1% 的压缩率,却造成 5538 的能量损耗。从表 2 我们可以看出,本文算法得到3.9% 的压缩率,只造成 3864 的能量损耗;方法 1 得到 7.3% 的压缩率,却造成 4153 的能量损耗;方法 2 得到6.8% 的压缩率,却造成 4000 的能量损耗。从表 3 我们可以看出,本文算法得到13.8% 的压缩率,只造成 3864 的能量损耗;方法 1 得到 15.3% 的压缩率,却造成 8278 的能量损耗;方法 2 得到14.9% 的压缩率,却造成 7670 的能量损耗。显然,本文算法不仅得到更高的压缩率,还降低了损耗。从比较结果来看,本文算法生成的视频不仅获得主观更好的浓缩视频,还有更好的压缩率和更小的损耗。

表 1　视频 1 实验比较结果

视频 1	能量损耗			帧数	
	E_a	E_t	E_c	原始视频	浓缩视频
本文方法	653	710	3506	31530	1775
方法 1	430	1078	4030	31530	2328
方法 2	423	886	3659	31530	2134

表 2　视频 2 实验比较结果

视频 2	能量损耗			帧数	
	E_a	E_t	E_c	原始视频	浓缩视频
本文方法	521	578	2765	43685	2665
方法 1	512	476	3165	43685	3164
方法 2	498	426	3076	43685	2987

表 3　视频 3 实验比较结果

视频 2	能量损耗			帧数	
	E_a	E_t	E_c	原始视频	浓缩视频
本文方法	521	430	5252	39556	5467
方法 1	526	487	7265	39556	6041
方法 2	498	476	6696	39556	5879

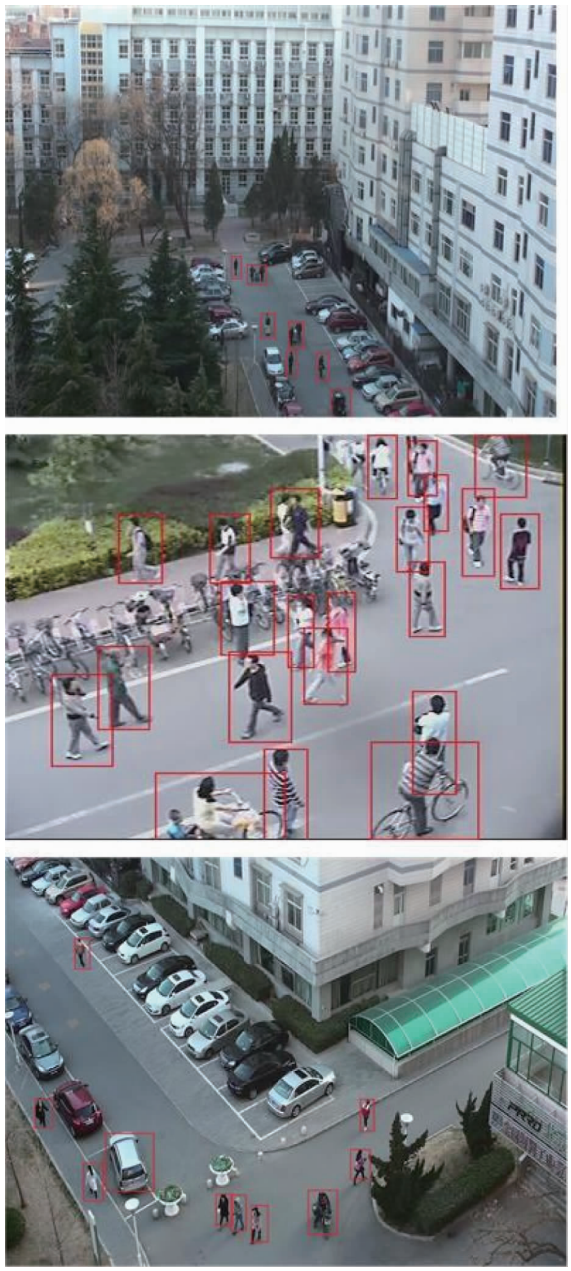


图 5　浓缩视频的代表图像

5　结 论

本文提出并讨论了新颖的基于关键观测点选择的视频浓缩技术。算法采用了数据驱动的方式,把问题转换为最小描述长度求解,进行关键观测点选择。本研究的算法能在视频浓缩中极大地消除内容上的冗余,提高视频浓缩效率。但是在关键观测点选择中,如果采样率过大,会造成视频中目标的空间跳动。一个可能的有效办法是,局部调整观测点的空间位置,并把这个调整结合到视频浓缩优化算法

中,用来进一步保证目标的空间一致性,这在今后的工作中,会继续深入研究。综上所述,本文算法在智能监控应用领域,特别是在视频检索和浏览中,具有很好的应用前景。

参考文献

[1] Petrovic N, Jojic N, Huang T. Adaptive video fast forward. *Multimedia Tools and Applications*, 2005, 26(3): 327-344

[2] Ma Y F, Lu L, Zhang H J, et al. A user attentionmodel for video summarization. In: ACM Multimedia, New York, USA, 2003. 533-542

[3] Smith M A, KanadeT. Video skimming and characterization throughthe combination of image and language understanding. In: Proceedings of the IEEE Conference on Computer Vision and Pattem Recognition, San Juan, Puerto Rico, 1997. 775-781

[4] Rav-Ach A, Pritch Y, Peleg S. Making a long video short: Dynamic video synopsis. In: Proceedings of the IEEE Conference on Computer Vision and Pattem Recognition, New York, USA, 2006. 435-441

[5] Feng S K, Li S Z, Yi D, et al. Online content-aware videocondensation. In: Proceedings of the IEEE Conference on Computer Vision and Pattem Recognition, Rhode Island, Providence, 2012. 2082-2087

[6] PritchY, Rav-Ach A, Peleg S. Webcam synopsis: Peeking around theworld. In: Proceedings of the IEEE International Conference on Computer Vision, Rio de Janeiro, Brazil, 2007. 1-8

[7] Pritch Y, Ratovitch S, Hendel A, et al. Clustered synopsis ofsurveillance video. In: Proceedings of the IEEE International Conference on Advanced Video and Signal based Surveillance, Genoa, Italy, 2009. 195-200

[8] Huang C R, Chung P C. Maximum a Posteriori Probability Estimationfor Online Surveillance Video Synopsis. *IEEE Transactions on Circuits and Systems for Video Technology*, 2014, 24(8):1417-1429

[9] Sun J, Zhang W, Tang X, et al. Background cut. In: Proceedings of the European Conference on Computer Vision, Graz, Austria, 2006. 628-641

[10] Taj M, Maggio E, Cavallaro A. Multi-feature graph-based object tracking. In: Proceedings of the 1st International Evaluation Conference on Classification of Events, Activities and Relationships, Southampton, UK, 2006. 190-199

[11] Zhang X Y, Wang S, Yun X. Bidirectional active learning, a two-way exploration into unlabeled and labeled dataset. *IEEE Transactions on Neural Networks and Learning Systems*, 2015, PP(99):(ahead-of-print)

[12] Tian Z Q, Xue J R, Lan X G, et al. Key object-basedstatic video summarization. In: ACM Multimedia, Arizonna, USA, 2011, 1301-1304

[13] Lee J, Han J. Trajectory clustering: Apartition-and-group framework. In: Proceedings of the ACM Special Interest Group on Management Of Data, Beijing China, 2007. 593-604

[14] Zhang X. Interactive patent classification based on multi-classifier fusion and active learning. *Neurocomputing*, 2014, 127(3):200-205

[15] Gangnet M, Perez P, Blake A. Poisson image editing. In: Proceedings of the ACM Special Interest Group for Computer GRAPHICS, San Diego, USA, 2003. 313-318

Video synopsis based on key observation selection

Zhu Xiaobin, Fan Fangxin, Xu Yinghan, Tan Li

(School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048)

Abstract

To quickly find the useful information from a vast amount of video data, a novel video synopsis method based on key observation selection was presented after the analysis of the performance of existing video synopsis algorithms. The method uses the new motion sequence composed by the selected representative observations to represent targets' original motion sequence to eliminate the content redundancy existing video synosis algorithms can not eliminate, so the video synopsis efficiency can be improved. In addition, the method adopts a data-driven mode to select key observations, and achieves the adaptive selection by transforming the key observation selection into the optimization of the minimum description length (MDL) to overcome the synopsis efficiency degrading and the synopsis video confusing of esisting video synopsis algorithms caused by too many observations. The experiments on three real surveillance videos were conducted to validate the effectiveness of the proposed approach.

Key words: video synopsis, key observation selection, minimum description length (MDL), data-driven