

基于贝叶斯生成对抗网络的背景消减算法

郑文博^{1,2} 王坤峰^{2,3} 王飞跃^{2,3,4}

摘要 背景消减是计算机视觉和模式识别的关键技术之一. 本文提出一种新的背景消减算法, 该算法首先利用中值滤波算法进行背景数据的获取, 然后基于贝叶斯生成对抗网络进行训练, 利用生成对抗网络的特性, 有效地对每个像素进行分类, 解决了光照渐变和突变、非静止背景以及幻影的问题. 本文采用深度卷积神经网络, 来构建贝叶斯生成对抗网络的生成器和判别器. 实验结果表明, 本文提出的算法性能在绝大多数情况下优于现有其他算法. 本文的贡献在于首次将贝叶斯生成对抗网络应用于背景消减, 并且取得了良好的实验效果.

关键词 背景消减, 背景获取, 贝叶斯生成对抗网络, 深度卷积神经网络

引用格式 郑文博, 王坤峰, 王飞跃. 基于贝叶斯生成对抗网络的背景消减算法. 自动化学报, 2018, 44(5): 878–890

DOI 10.16383/j.aas.2018.c170562

Background Subtraction Algorithm With Bayesian Generative Adversarial Networks

ZHENG Wen-Bo^{1,2} WANG Kun-Feng^{2,3} WANG Fei-Yue^{2,3,4}

Abstract Background subtraction is one of the key techniques in computer vision and pattern recognition. A new background subtraction algorithm is proposed, which firstly uses the median filtering algorithm for extracting background and then trains the network based on Bayesian generative adversarial network. The work uses Bayesian generative adversarial network to classify each pixel effectively, thereby addressing the issues of sudden and slow illumination changes, non-stationary background, and ghost. Deep convolutional neural networks are adopted to construct the generator and the discriminator of Bayesian generative adversarial network. Experiments show that the proposed algorithm results in better performance than others in most cases. The contribution of the work is to apply Bayesian generative adversarial network to background subtraction for the first time and achieve good experimental results.

Key words Background subtraction, background extraction, Bayesian generative adversarial network, deep convolutional neural networks

Citation Zheng Wen-Bo, Wang Kun-Feng, Wang Fei-Yue. Background subtraction algorithm with Bayesian generative adversarial networks. *Acta Automatica Sinica*, 2018, 44(5): 878–890

背景消减是底层计算视觉和模式识别领域的关键技术之一, 广泛应用于视频监控、智能交通^[1-2]、体育视频、工业视觉等领域中. 背景消减的作用是将背景部分从视频中减去^[3], 以提取运动前景. 其核

心问题是如何建立一个自适应的背景模型, 以准确描述背景信息. 一个性能良好的背景模型要能够有效描述各种条件下背景在空域和时域所发生的变化. 但这是非常困难的, 因为实际的背景环境变化是非常复杂的, 有光线的变化, 如亮度的渐变、光照的突变、阴影等; 也有运动背景的影响, 如被风吹动的树叶、水纹、变化的显示屏幕等.

过去, 人们提出了很多种统计背景建模方法. Stauffer 等^[4-5] 首次提出了基于混合高斯模型 (Gaussian mixture models, GMM) 的背景建模方法, 该方法较好地解决了复杂背景条件下的背景建模, 能够适应光照的缓慢变化, 但它无法对快速变化的背景进行建模^[6]. Fan 等^[7] 提出的是基于灰度图像或者彩色图像的码书模型的背景建模方法, 该方法对光照变化鲁棒性强, 但是不能对光照长时间缓慢变化的场景进行建模, 而且, 在训练期的初始帧时限制了运动目标的出现. Yan 等^[8] 结

收稿日期 2017-09-30 录用日期 2018-02-26
Manuscript received September 30, 2017; accepted February 26, 2018

国家自然科学基金 (61533019, 91720000) 资助
Supported by National Natural Science Foundation of China (61533019, 91720000)

本文责任编辑 李力
Recommended by Associate Editor LI Li

1. 西安交通大学软件学院 西安 710049 2. 中国科学院自动化研究所复杂系统管理与控制国家重点实验室 北京 100190 3. 青岛智能产业技术研究院平行视觉技术创新中心 青岛 266000 4. 国防科学技术大学军事计算实验与平行系统技术研究中心 长沙 410073

1. The School of Software Engineering, Xi'an Jiaotong University, Xi'an 710049 2. The State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190 3. Innovation Center for Parallel Vision, Qingdao Academy of Intelligent Industries, Qingdao 266000 4. Research Center for Computational Experiments and Parallel Systems Technology, National University of Defense Technology, Changsha 410073

合局部融合信息的策略提出了一种基于变分贝叶斯的背景减法 (Variational Bayesian learning for background subtraction based on local fusion feature, LFVBGM) 算法, 虽然考虑了像素的空间性, 使得背景建模更加可靠, 适应了大多数的动态场景, 但是固定的背景更新率仍然限制了一些特殊场景下的目标检测. MBS (Multimode background subtraction) 方法^[9] 尝试提取每个位置像素变化的主分量和梯度信息描述背景模型, 并提出颜色共生矩阵用来表示相邻帧的背景变化, 但总的来说, 它仍然算是单模式的, 所以处理动态背景的性能不好; BMOG (Boosted Gaussian mixture model with controlled complexity)^[10] 方法集成了传统的 GMM 方法与 MBS 方法, 但没有解决动态背景的缺陷; SSOBS (Simplified self-organized background subtraction) 方法^[11] 只考虑图像空间的一致性, 对于图像之间本身的时间性忽略, 导致动态背景效果较差; ShareModel (Learning sharable models for robust background subtraction)^[12] 方法将 GMM 与像素的相关性、视频帧的时间性相结合, 虽然大部分数据库取得较好的结果, 但是对于突然出现的目标或者突然摆动的背景的情形, 效果欠佳; FTSG (Flux tensor with split Gaussian models) 方法^[9] 分别对像素变化特性、混合高斯模型控制参数进行了分析, 将背景学习分为两个大的阶段, 引入反馈控制机制进行背景建模, 提出了一种自适应学习速率的混合高斯模型算法, 但是仅根据图像背景信息感知, 当图像背景突然变化, 因此, 其感知精度难以保证; WeSamBE (A weight-sample-based method for background subtraction)^[13] 方法借助于背景重建算法设计基于空间分割的分类器, 虽然在一定程度上避免了传统方法的单一性, 但其实质仍然是针对特定分类的分类器, 对于其他特定分类, 未必有好的效果. 基于传统神经网络的算法, 例如 DeepBS (A deep convolutional neural network for video sequence background subtraction)^[14], 需要设计遵循任何种类的因式分解的模型, 只能针对特定的分离器有用, 因此实用性较差.

Saatchi 等于 2017 年提出的贝叶斯生成对抗网络^[15] 具有实现任何复杂非线性映射的功能^[16-17], 并具有自动学习的能力^[18-19]. 与传统神经网络算法相比, 不需要设计遵循任何种类的因式分解的模型, 对任何生成器网络和任何分类器都会有用. 因此, 也无需反复采样^[20-21], 无需在学习过程中进行推断, 回避了近似计算概率的难题^[22-23].

针对传统神经网络算法的弊端及现有多数算法不能完全解决光照渐变和突变、非静止背景以及鬼影的问题, 本文提出了一种在静态摄像机条件下

有效地从背景中提取前景目标轮廓的背景减法, 即基于贝叶斯生成对抗网络的背景减除算法 (Background subtraction algorithm based on Bayesian generative adversarial networks, BSGAN). BSGAN 算法对 t 帧图像中每一个像素的时间序列进行背景建模, 利用中值算法来分离和描述可靠的背景. 综合考虑到噪声和光照对像素的影响, 利用贝叶斯生成对抗网络进行训练, 从而有效地对每个像素进行分类以应对渐进和突然的光照变化以及非静止背景和鬼影的出现等干扰.

本文的主要贡献在于: 1) 将贝叶斯生成对抗网络思想应用于背景消减, 具有鲁棒性高、对场景变化适应能力强等优点; 2) 利用卷积生成对抗网络 (Deep convolutional generative adversarial network, DCGAN) 来建立贝叶斯生成对抗网络的背景消减模型, 建立了基于卷积神经网络的贝叶斯生成对抗网络, 并通过训练生成对抗网络, 获得鲁棒的结果; 3) 将 BSGAN 算法提交到 CDnet2014 数据集平台上进行测试, 获得了较好的性能指标, 超过了许多现有方法.

本文接下来从 4 个方面介绍本文算法, 首先, 在第 1 节介绍贝叶斯生成对抗网络的兴起与原理; 在第 2 节, 详细说明了基于贝叶斯生成对抗网络的背景消减; 并在第 3 节进行该算法的对比实验以及实验结果分析; 最后, 在第 4 节进行总结.

1 贝叶斯生成对抗网络简介

深度学习对海量标签数据的依赖是显而易见的, 这也成为抑制深度学习发展的潜在要素之一. 长久以来, 科学家们都在探索使用尽量少的标签数据的学习方法, 希望实现从监督式学习到半监督式学习再到最后的无监督式学习的转化. 在大多数情况下, 我们是没有带标签的数据的, 带标签的数据往往是以高成本通过人力劳动或通过昂贵的仪器 (如用于自动驾驶的激光雷达) 来实现的. 因此, 半监督式学习已经成为最新的研究热点. 特别自 2015 年出现了将生成对抗网络 (Generative adversarial network, GAN) 用于半监督问题^[19] 以来, 现有的许多 GAN 工作也都表明通过加入少量类别标签, 引入有标签数据的类别损失度量, 不仅在功能上实现了半监督学习, 同时也有助于 GAN 的稳定训练.

Saatchi 等最近发布了一项对无监督和半监督式学习的研究, 名为贝叶斯生成对抗网络 (Bayesian GAN)^[15]. 贝叶斯生成对抗网络的原理, 简单来说, 给定一个变量为 $x^{(i)} \sim p_{\text{data}}(x^{(i)})$ 的数据集, 估计数据集的分布 $p_{\text{data}}(x^{(i)})$. 我们利用参数 θ_g 控制的生成器 $G(z; \theta_g)$ 将白噪声 $z \sim p(z)$ 转化为与样本数据分布近似的候选样本 (生成样本). 我们用另一

个由参数 θ_d 表示的判别器 $D(x, \theta_d)$ 来判定任何来自数据分布的 x 的可靠性 (类别). 这种简单理解显然是对于一般的 D 和 G 而言的, 但是在实践中, G 和 D 通常是具有权重向量 θ_g 和 θ_d 的神经网络. 假定样本分布是能够置于 θ_g 和 θ_d 之间的, 我们在生成器和判别器的无穷大空间中引入一个分布, 并对应于这些权重向量的进行每个可能的设置, 使得生成器生成的候选样本 (生成样本) 可以与样本数据是同一分布. 从数据实例的先验分布中抽样, 如下: 1) 样本 $\theta_g \sim p(\theta_g)$; 2) 样本 $z^{(1)}, \dots, z^{(n)} \sim p(z)$; 3) $\tilde{x}^{(j)} = G(z^{(j)}; \theta_g) \sim p_{\text{generator}}(x)$. 对于之后的推理, 采用无监督和半监督两种形式.

1.1 贝叶斯生成对抗网络的无监督学习

为了推断出后面的 θ_g 和 θ_d , 可以迭代地抽样:

$$p(\theta_g|z, \theta_d) \propto \left(\prod_{i=1}^{n_g} D(G(z^{(i)}; \theta_g); \theta_d) \right) p(\theta_g|\alpha_g)$$

$$p(\theta_d|z, X, \theta_g) \propto \prod_{i=1}^{n_d} D(x^{(i)}; \theta_d) \times \prod_{i=1}^{n_g} (1 - D(G(z^{(i)}; \theta_g); \theta_d)) \times p(\theta_d|\alpha_d) \quad (1)$$

其中, $p(\theta_g|\alpha_g)$ 和 $p(\theta_d|\alpha_d)$ 是生成器和判别器参数的先验, 分别具有超参数 α_g 和 α_d . n_d 和 n_g 分别是生成器和判别器样本的数量.

从数据样本的生成过程开始, 可以直观地了解这一公式. 从先验的 $p(\theta_g|\alpha_g)$ 中采样权重, 再计算由采样后的权重计算该样本的权重, 以形成特定的生成网络. 然后本文从采样来自 $p(z)$ 的白噪声 z , 并通过网络 $G(z; \theta_g)$ 转换该噪声, 以产生候选数据样本 (生成样本). 判别器根据其权重 θ_d 进行调整, 推算出这些来自数据分布是候选样本 (生成样本) 的概率. 在式 (1) 中, 如果判别器输出概率较高, 则后验概率 $p(\theta_g|z, \theta_d)$ 将在 θ_g 的采样设置的附近增加 (其他设置减小). 对于判别器权重 θ_d 的后验, 式 (1) 形成判别分类似然函数, 将实际数据的样本与生成器对比标记分类类别.

1.2 贝叶斯生成对抗网络的半监督学习

将概率的 GAN 形式体系扩展到半监督学习. 在 K 级分类的半监督设置中, 假定训练样本为一组 n 个几乎不带标签集的 $\{x^{(i)}\}$, 以及一组 (通常样本数目小得多) 的 n_s 观察值集合 $\{(x_s^{(i)}, y_s^{(i)})\}_{i=1}^{n_s}$, 其中 $y_s^{(i)} \in \{1, \dots, K\}$. 该网络目标是如果只能使用已分类的训练输入, 那么该网络将从未标记和标记的示例中共同学习统计结构, 以便对新测试样本 X 类标签进行更好的预测.

在这种情况下, 重新定义判别器, 使得 $D(x^{(i)} = y^{(i)}; \theta_d)$ 给出样本 $x^{(i)}$ 属于类 $y^{(i)}$ 的概率. 保留类标签 0 以表示数据样本是生成器的输出. 然后推断出后验概率如下:

$$p(\theta_g|z, \theta_d) \propto \left(\prod_{i=1}^{n_g} \sum_{y=1}^K D(G(z^{(i)}; \theta_g) = y; \theta_d) \right) p(\theta_g|\alpha_g)$$

$$p(\theta_d|z, x, y_s, \theta_g) \propto \prod_{i=1}^{n_d} \sum_{y=1}^K D(x^{(i)} = y; \theta_d) \times \prod_{i=1}^{n_g} D(G(z^{(i)}; \theta_g) = 0; \theta_d) \times \prod_{i=1}^{n_s} (D(G(x_s^{(i)} = y_s^{(i)}; \theta_d)) p(\theta_d|\alpha_d)) \quad (2)$$

在每次迭代中, 使用来自 n_g 个生成器的样本, n_d 个未标记的样本和 n_s 个所有标记的样本, 其中通常为 $n_s \ll n$. 该网络可以使用简单的蒙特卡洛取样法以近似边缘化 z . 借鉴无监督学习思想, 本文可以边缘化后面的 θ_g 和 θ_d , 并且求得其平均值, 用于计算测试样本的分类情况. 在 T 次迭代中, 为了计算测试输入 x_* 时标记分类 y_* 的分布, 使用所有关于后面的收集样本 θ_d 的模型平均值:

$$p(y_*|x_*, D) = \int p(y_*|x_*, \theta_d) d\theta_d \approx \frac{1}{T} \sum_{k=1}^T p(y_*|x_*, \theta_d^{(k)}), \theta_d^{(k)} \sim p(\theta_d|D) \quad (3)$$

并且这种模式平均值对于提升半监督学习能力是有效的.

2 基于贝叶斯生成对抗网络的背景消减

背景减法是利用当前图像与只含有当前背景的图像的差分来提取图像背景的一种技术. 本文背景减法的第一步就是实现背景数据获取, 因为灰度图像的简单中值滤波法已广泛应用于鲁棒的复杂场景背景数据的获取, 快速简单, 而且能适应背景的变化^[24], 所以本文利用基于灰度图像的简单中值滤波法^[25] 获取背景数据; 第二步是生成一个学习网络的场景数据集, 不仅可以规范化每帧的数据, 同时也为第三步的网络训练提供训练样本; 本文第三步的网络训练采用神经网络, 对场景数据集进行训练; 最后, 输出为提取的前景目标. 本文算法流程图如图 1 所示.



图 1 本文算法流程图

Fig. 1 The flow chart of our algorithm

2.1 背景数据获取

我们的算法通过使用从几个视频帧中提取出来的单个灰度图像来模拟背景 (我们观察到使用三个颜色通道仅仅会使我们的方案有一些边际改进). 故此使用以下公式, 将输入图像从 RGB 域转换为灰度级 (以下称为 Y):

$$Y = 0.299 \times R + 0.587 \times G + 0.114 \times B \quad (4)$$

本文以简单的场景为例, 我们通过在短时间内观察场景并计算每个像素的时间中值 Y 值来提取灰度背景图像. 需要特别指出的是, 当背景中的每个像素至少有 50% 的时间是可见的时候, 该方法合适的. 对于混乱的场景, 需要更复杂的背景估计方法, 例如文献 [25] 的复杂背景估计方法.

2.2 场景数据集的生成

本文流程的第二步是生成一个学习网络的场景数据集. 首先通过 T 表示以每个像素为中心的块 (Patch) 的大小, 用于减法运算. 再定义一个 $T \times T$ 的 2 通道图像块 (提取于中值图像且用于背景块的一个通道和用于输入块的一个通道) 为训练样本 x . 相应的目标值由下式给出:

$$\begin{cases} t(x) = 1, & \text{若 } \text{class}(p_c) = FG \\ t(x) = 0, & \text{若 } \text{class}(p_c) = BG \end{cases} \quad (5)$$

其中, p_c 表示图像块的中心像素. 要注意的是, 本文将 T 设置为 27, 并将所有 Y 值标准化使其间隔于 $[0, 1]$. 也就是说, 在假定图像被零填充以避免边界效应的情况下, 一系列 N 个完全标记的输入图像的序列与训练样本的集合等值.

2.3 贝叶斯生成对抗网络的训练原理

假定真实数据集 $S = \{(x_1, y_1), \dots, (x_n, y_n)\}$ 和伪数据集 $S' = \{(x'_1, y'_1), \dots, (x'_n, y'_n)\}$, 其中真实样本数据 $x_i \in \mathbf{R}^{N \times N}$, 真实样本的标签集 $y_i \in \{1, \dots, K\}$. 这种设置似乎只适用于监督学习, 但它实际上适用于 GAN 的半监督和无监督学习问题. 在监督学习设置中, K 是所有数据的类别数. 在我们有一些未标记数据的半监督设置中, 我们可以通过分配标签 $y = K + 1$ 的未标记数据来增加实际集合 S . 在无监督 (即我们所有的都是未标记的数据) 的学习设置中, 真实的集合标记为 $y = 1$, 伪的是 $y = 0$.

我们将生成器模型看作是随机函数 f_G , f_G 的作用是将输入的 y' 转化为其分布类似于实际数据分布的样本 x . 逻辑结构如图 2 所示. 由本文假定, 可得 $\int S'_i \sim \int p(f_G)$, 由于式 (2) 以及贝叶斯公式, 我们可以得到判别器中生成的一组样本的分布为

$$\begin{aligned} p(S'|f_D) &\propto \int p(\omega) \int p(f_G|\omega) \times \\ &\prod_{i=1}^n p(f_D(f_G(y'_i), y'_i)) df_G d\omega \\ p(S'|\omega, f_D) &= \\ &\int p(f_G|\omega) \prod_{i=1}^n p(f_D(f_G(y'_i), y'_i)) df_G \quad (6) \end{aligned}$$

其中, ω 是生成网络的权重, f_D 是判别器函数.

同理可得, 对于生成器而言:

$$\begin{aligned} p(S) &= \\ &\int p(\theta) \int p(f_D|\theta) \prod_{i=1}^n p(f_D(x, y_i)) df_D d\theta \\ p(f_D|S) &= \int p(\theta) \int p(f_D|\theta) p(S|f_D) d\theta = \\ &\int p(\theta) \int p(f_D|\theta) \prod_{i=1}^n p(f_D(x, y_i)) d\theta \\ p(f_D|S, \theta) &\propto p(f_D|\theta) p(S|f_D) \quad (7) \end{aligned}$$

其中, θ 是判别器网络的权重.

对于式 $p(f_D|S, \theta) \propto p(f_D|\theta) p(S|f_D)$, 应用贝叶斯公式与最大后验估计法, 可得 $p(f_D) \propto p(f_D|S, \theta)$.

使用贝叶斯方法推断参数的优点是我们在我们的方法中包含模型不确定性, 这意味着有能力应对 GAN 的收敛问题^[15]. 这是因为使用后验分布的权重和概率分布函数的优势, 因此可以避免在复杂的参数空间中诸多参数的调优问题, 而且, 可以对 GAN 中由于生成器和判别器交替优化而无法达到鞍点的问题有帮助.

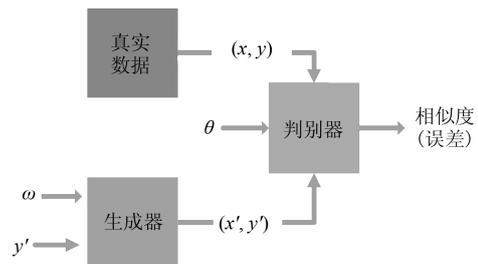


图 2 基于 DCGAN 的贝叶斯生成对抗网络

Fig. 2 Bayesian generative adversarial networks based on the DCGAN

为了实现仅利用概率函数和判别器的不确定性进行生成样本然后进行网络更新的目的, 根据生成对抗网络的原理^[20], 我们对生成器和判别器的推断函数进行定义, 并尽可能地最小化网络的损失, 也就是说, 本文方法策略与在顺序决策中使用的汤姆森抽样相当, 贪心地将每次迭代中都以最小化贝叶斯框架内的预期损失为目的。

根据生成对抗网络的原理^[20], 我们知道判别器的推断函数为

$$-\min_D E_{(x,y) \sim p_{\text{data}}} \log(D(x)) - E_{(x',y') \sim p_{\text{fake}}} \log[1 - D(G(z))] \quad (8)$$

其中, $D(\cdot)$ 和 $G(\cdot)$ 分别为判别器函数和生成器函数。

根据式 (6) 和 (8), 并使用最大后验估计, 假定函数是基于当前观察 (来自伪数据和实际数据的样本) 的一次迭代得到, 最后可以得到判别器推断函数:

$$\min_{\theta} E_{f_D \sim p(f_D | S, \theta)} E_{(x,y) \sim p_{\text{data}}} [l_D(f_D(x), y)] - E_{f_D \sim p(f_D | S, \theta)} E_{(x',y') \sim p_{\text{fake}}} [l_G(f_D(x'), y')] \quad (9)$$

同理可以得到生成器推断函数:

$$\min_{\omega} E_{f_D \sim p(f_D | S, \theta)} [E_{(x',y') \sim p_{\text{fake}}} [l_G(f_D(x'), y')] + E_{(x,y) \sim p_{\text{data}}} E_{(x',y') \sim p(S' | \omega)} [\Delta_{f_D}((x, y), (x', y'))]] \quad (10)$$

其中, l_G 和 l_D 分别是为生成器和判别器的损失函数, $\Delta_{f_D}(S, S')$ 是真实样本与伪样本的衡量函数。

采用蒙特卡洛法, 在真实样本中选择 n 个真实样本, 对应的标签集的样本数量为 m , 在伪样本中选

择 n' 个伪样本, 伪样本对应的标签集的样本数量为 m' . 将函数离散化可得:

判别器推断函数:

$$L_D(\theta) = \frac{1}{n \times m} \sum_i \sum_j l_D(f_D^{(i)}(x_j), y_j) - \frac{1}{n' \times m'} \sum_i l_G(f_D(x'_j), y'_j) \quad (11)$$

$$f_D^{(i)} \sim p(f_D^{(i)} | S, \theta)$$

生成器推断函数:

$$L_G(\omega) = \frac{1}{n' \times m'} \sum_i l_G(f_D(x'_j), y'_j) + \frac{1}{m'} \Delta_{f_D}(S, S') \quad (12)$$

$$f_D^{(i)} \sim p(f_D^{(i)} | S, \theta)$$

$$S' \sim p(S' | f_G^{(i)})$$

$$f_G^{(i)} \sim p(f_G^{(i)} | \omega)$$

考虑到生成器可能生成一些高质量的伪样本, 在这种情况下, 真实样本与伪样本之间的差异很难测量^[26], 因此, 本文采用最大均值差异 (Maximum mean discrepancy, MMD)^[27] 的方法进行评定. 因此可得

$$\Delta_{f_D}(S, S') = \left\| \frac{1}{n} \sum_i f_D^{(i)}(x_i) - \frac{1}{n'} \sum_i f_D^{(i)}(x'_i) \right\|^2 \quad (13)$$

2.4 贝叶斯生成对抗网络的训练过程

本文算法利用基于贝叶斯的生成对抗网络进行

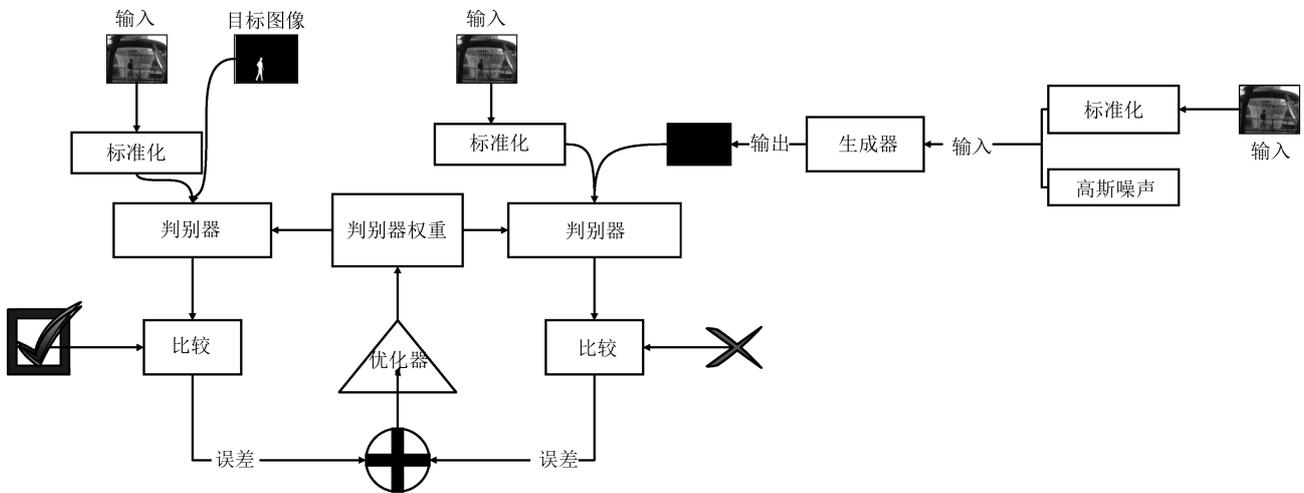


图 3 本文算法工作示意图

Fig. 3 The flow chart of our algorithm works

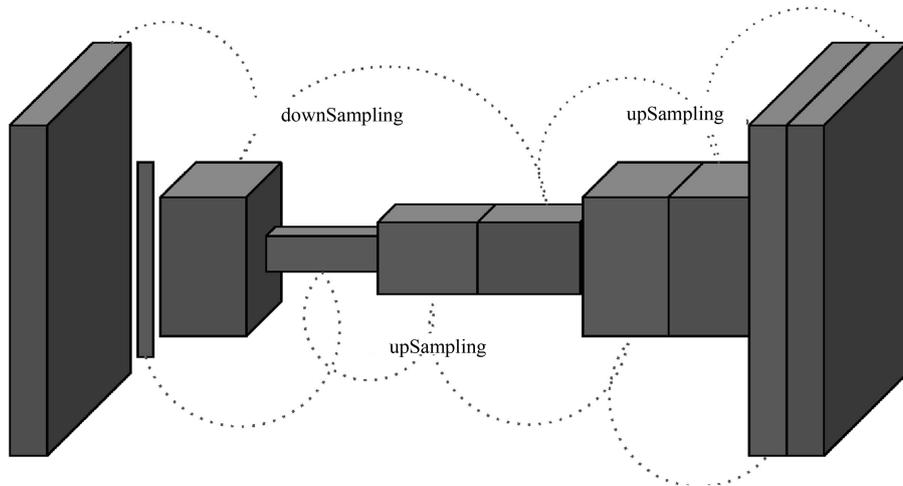
背景消减, 工作流程图如图 3 所示. 背景消减问题, 可以看作是图像像素的分类问题, 即考虑前景、背景 2 种类型, 因此可以看作是 $k = 2$ 的分类问题, 我们的目标是尽可能接近 Ground-truth 的结果. 因此, 我们首先对于每个数据库随机选择 100 帧当前图像进行读入, 再者利用当前图像标准化后的图像与高斯噪声产生的标签集, 作为生成器的输入, 生成器生成背景图像并给判别器裁定; 将当前图像标准化后的图像与生成器生成的背景图像作为判别器的输入, 判别器根据与 Ground-truth 的误差结果进行判定, 并反馈给生成器, 生成器与判别器网络进行权重的更新. 具体来说, 在批量读入原图后, 以一幅图像为例进行说明, 输入一幅图像后, 首先将图像的所有像素值标准化使其间隔于 $[0, 1]$, 然后与高斯噪声一同作为生成器的输入, 经过生成器网络的运算生成一幅背景减除的图像, 生成器将这幅生成的图像交给判别器, 等待裁定. 判别器网络根据标准化后的原图和原图的 Ground-truth 图像, 与生成图像的误差进

行判定. 若判定为真, 则记录这幅生成图像的标签为 1; 否则为 0.

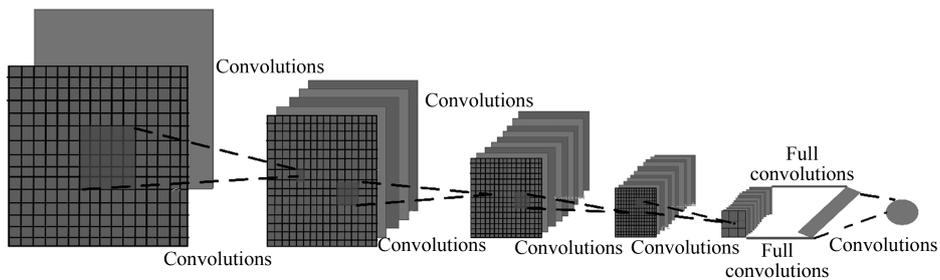
我们采用贝叶斯的卷积生成对抗网络 (DC-GAN) 进行训练, 其网络结构如图 4 所示, 算法如下所示:

初始化: 学习率 η , 生成器和判别器的权重 ω, θ

- 1) 迭代 θ , 随机生成 ω
- 2) **while not converged do**
- 3) $S = \{(x_i, y_i)\}_{i=1}^n$
- 4) **for** $i = 1 \rightarrow m$ **do**
- 5) $\alpha \sim \text{Bernoulli}(\pi_D)$
- 6) $\beta \sim N(0, \sigma_D^2 I), \tilde{\theta} = \theta \odot \alpha + \beta$
- 7) 计算 $\nabla_{\theta} L_D^{(j)}(\tilde{\theta})$
- 8) **end for**
- 9) $\theta = \theta - \eta_t \times (\sum_j \nabla_{\theta} L_D^{(j)}(\tilde{\theta}))$
- 10) 正则化 θ , 使得 $\|\theta\| \leq 1$
- 11) **for** $j = 1 \rightarrow m'$ **do**
- 12) $S' = \text{SampleFake } \omega$



(a) 生成器网络部分结构
(a) The structure of generator network



(b) 判别器网络结构
(b) Discriminator network structure

图 4 贝叶斯卷积生成对抗网络结构

Fig. 4 The structure of Bayesian convolutional generative adversarial network

```

13)   计算  $\nabla_{\omega} L_G^{(j)}(\tilde{\omega})$ 
14)   end for
15)    $\omega = \omega - \eta_t \times (\sum_j \nabla_{\omega} L_G^{(j)}(\tilde{\omega}))$ 
16) end while
17) return  $\theta, \omega$ 
18) procedure SampleFake  $\omega$ 
19)    $\alpha \sim \text{Bernoulli}(\pi_G)$ 
20)    $\beta \sim N(0, \sigma_G^2 I), \tilde{\omega} = \omega \odot \alpha + \beta$ 
21)    $S' = \{(x'_i, y'_i)\}_{i=1}^{n'}$ 
22)   return  $S'$ 
23) end procedure

```

本文生成器和判别器都采用基于 DCGAN 的神经网络, 具体来说, 本文生成器采用先下采样, 后上采样的结构方式, 如图 4(a) 所示, 即网络前半段为卷积层, 后半段为反卷积层, 最后输出背景图像. 其中由噪声生成的过程为上采样过程. 对于噪声的生成网络, 每层网络大小依次变化为 $4 \times 4 \times 1024 \rightarrow 8 \times 8 \times 512 \rightarrow 16 \times 16 \times 256 \rightarrow 32 \times 32 \times 128 \rightarrow 64 \times 64 \times 3$. 生成器网络的每一层的滤波器尺寸为 5×5 . 由图像生成的生成网络每层网络依次变化为 $64 \times 64 \times 3 \rightarrow 32 \times 32 \times 128 \rightarrow 16 \times 16 \times 256 \rightarrow 32 \times 32 \times 128 \rightarrow 64 \times 64 \times 3$, 前三层为卷积层, 后两层为反卷积层. 判别器网络采用最后一层为全连接层的 6 层卷积神经网络, 如图 4(b) 所示, 每一层卷积层的卷积核的大小变化为 $64 \times 64 \times 3 \rightarrow 32 \times 32 \times 128 \rightarrow 16 \times 16 \times 256 \rightarrow 8 \times 8 \times 512 \rightarrow 4 \times 4 \times 1024$. 判别器网络的每一层卷积滤波器尺寸为 5×5 . 对于噪声生成网络中第一层卷积核的大小的含义是将把 1024 的一维向量转化 (Reshape) 成 1024 个通道的 4×4 的特征图 (Feature map), 本文生成对抗网络的其他各层的卷积核的含义与第一层理同. 也就是说, 本文生成对抗网络的卷积层的特征图大小是上一层的 2 倍, 本文生成对抗网络的反卷积层的特征图的大小是上一层的一半.

根据 He 等^[28] 提出的计算卷积神经网络时间复杂度的方法可知,

$$O\left(\sum_{l=1}^d n_{l-1} \times s_l^2 \times n_l \times m_l^2\right) \quad (14)$$

其中, d 为网络卷积层的总层数, n_l 为本层卷积滤波器个数, 也就是大家所说的卷积滤波器的宽度; n_{l-1} 为上一层卷积滤波器个数, 也可以称作 n_l 层的输入通道数; s_l 为滤波器的尺寸; m_l 为输出特征图的尺寸.

例如, 一个输入输出通道数都为 256 的卷积神经网络, 卷积滤波器的尺寸为 3×3 , 可以得到

该层的时间复杂度是 $O(589824 \times m_l^2)$. 相似地, 可以算出本文的生成对抗网络的时间复杂度约为 $O(10^6 \times m_{\min}^2)$, 可见这个时间复杂度与传统机器学习算法^[4-5] 相比是非常大的.

我们在实验的生成器训练中, 可以采取根据每个数据库的难易程度进行随机选择的 100 帧当前图像的增量读入的方式. 当第一次读入第一个数据库 (相对来说简单的数据库, 例如 CDnet 中 Baseline 的数据库) 随机选择的 100 帧图像, 然后输入当前图像标准化后的图像与高斯噪声产生的标签集. 当第二次读入第二个数据库随机选择的 100 帧图像时, 保留第一次训练数据, 以此类推, 形成增量式训练.

我们在实验中发现, 改变生成器的高斯噪声为混合高斯噪声, 并在生成器生成后再次进行中值滤波后, 作为判别器的输入, 效果是比原先要好. 如果随机化不足, 确定性输入噪声可能导致整个生成器网络的崩溃, 因此, 我们在网络的每一层之间增加了高斯噪声. 对于所有实验, 我们将随机梯度下降的批量大小设置为 100. 我们随机选择标记示例的一个子集, 并将其余部分用作未标记的数据进行训练. 我们执行这些实验 5 次, 并记录平均值和标准误差, 我们使用的恒定学习率, 在本文实验中, 设置为 0.001.

具体的说, 我们使用生成器网络生成样本, 在层与层之间, 使用方差为 0.9, 均值为 0.1 的高斯噪声进行增强; 除了输出层外, 其余每一层之后进行批量标准化操作; 对于判别器网络来说, 判别器网络中的卷积层采用 conv + batchNorm + leakyReLU 的方案^[29-30]. 我们使用最后一层为全连接层的 6 层卷积的网络, 相似地, 在层与层之间, 使用方差为 0.9, 均值为 0.05 的高斯噪声进行增强; 在最后一层, 进行权重的标准化操作.

3 实验与结果分析

为验证本文所提出的方法, 在 Ubuntu 16 平台下, 使用 Python 3.X、TensorFlow 环境, 利用公开数据集 CDnet2014 的视频序列进行了实验, 并与其他检测方法 GMM-Staufier^[4]、GMM-Zivkovic^[5]、LBSP (Background subtraction using local binary similarity patterns)^[6]、IUTIS (In unity there Is strength)^[7]、MBS^[9]、FTSG^[9]、LF-GMM^[8]、LFVBGM^[8]、ArunVarghese^[31]、BMOG^[10]、DeepBS^[11]、ShareModel^[12]、SSOBS^[13]、WeSamBE^[14]、Cascade CNN^[32] 进行比较.

ChangeDetection.net (CDnet) 数据集有 2012 和 2014 两个版本, 其中 2014 版本包含 2012 版本中所有的视频数据, 本文的实验都是基于 CDnet2014 数据集. CDnet2014 数据集具有 11 类挑战性的场景, 大约 16 万帧视频帧图像, 更重要的是它还给

出了所有的视频图像中运动目标区域精确分割的 Ground-truth 标注. 这非常便于我们将自己方法的检测结果与 Ground-truth 进行比较, 从而定量地分析算法是否优越. 本文从精确率 (Pr) 和召回率 (Re) 以及 F-measures 三个评价标准进行对本文算法与其他算法评价. 其中精确率 (Pr) 是针对预测结果而言的, 它表示的是预测为正的样本中有多少是真正的正样本. 召回率 (Re) 是针对原来的样本而言的, 它表示的是样本中的正例有多少被预测正确了. TP 表示将正类预测为正类数的数量, FN 代表将正类预测为负类数的数量, FP 代表将负类预测为正类数的数量, TN 代表将负类预测为负类数的数量. F-measure 指标被定义为精确率和召回率的谐波平均值. 精确率 Pr、召回率 Re 以及 F-measure 的表达式分别为:

$$\begin{aligned} \text{Pr} &= \frac{\text{TP}}{\text{TP} + \text{FP}} \\ \text{Re} &= \frac{\text{TP}}{\text{TP} + \text{FN}} \\ \text{F-measure} &= \frac{2 \text{Pr} \times \text{Re}}{\text{Pr} + \text{Re}} \end{aligned} \quad (15)$$

本文的算法是在每一个数据集中随机选择 100 帧作为训练集, 其余作为测试集. 训练参数学习率设置为 $\eta = 0.001$, 训练轮数设置为 epoch = 300 进行实验, 动量 (参数移动平均数) 设置为 0.5, 图像批大小设置为 batchsize = 64. 将训练集的图像全部批处理为 320×240 , 也就是把输出图像的高度设置为 240, 宽度设置为 320. 以 CDnet 中的 office 数据集为例, 训练时的网络损失函数如图 5 所示. 从图 5 中可以看出, 最终的判别器总体 loss 在 1 左右波动, 而判别器的真实样本的损失函数 (Discrim-

inator real loss) 和判别器的生成样本的损失函数 (Discriminator fake loss) 几乎在一条水平线上波动, 这说明判别器最终对于真伪图像已经没有判别能力, 而是进行随机判断.

在测试阶段, 将经过标准化的测试图像作为生成网络的输入, 生成网络生成该图像的背景减除结果并输出. 具体来说, 以一幅图像为例进行说明, 输入一幅图像后, 将图像的所有像素值标准化使其间隔于 $[0, 1]$, 作为生成器的输入, 经过生成器网络的运算生成一幅背景减除的图像. 为了进行快速的测试, 我们将测试集的图像全部批处理为 320×240 , 然后作为训练好的网络的输入, 进行测试. 最后得到提取出的前景目标效果如图 6 所示, 测试时间大约为 5 帧/秒, 基本满足实时性测试的要求.

从图 6 中, 我们发现 highway、badminton、boulevard、sidewalk、traffic、park、blizzard、street、fall 等数据集的整体背景重建和背景减除效果较好, 而 busyBoulevard 等数据集的某段视频的背景重建和背景减除结果却有较大的误差. 追根溯源, 这是由于本文算法的背景减除结果质量实质上是由第一步的结果——背景重建决定的, 生成对抗网络受背景重建结果的影响. busyBoulevard 数据集是夜晚灯光较暗时的情景, 而在某段视频中, 由于采用的背景重建算法对这种情形缺乏敏感性, 背景无法精确得到, 所以处于黑暗中的物体容易被网络认为是动态的目标. bridgeEntry 数据集是在夜晚中大桥下车来车往的情景, 路灯和车辆的车灯的灯光相互交叉, 在某段视频中, 背景重建算法并不能分清楚哪些是路灯 (真实背景), 哪些是车的灯光 (目标), 所以网络把很多路灯的灯光错认为目标. 在 abandonedBox 和 tramstop 的某段视频中, 目标分别为缓慢行驶中的电车和驻足在马路的行人, 因为重建算法认为摄像

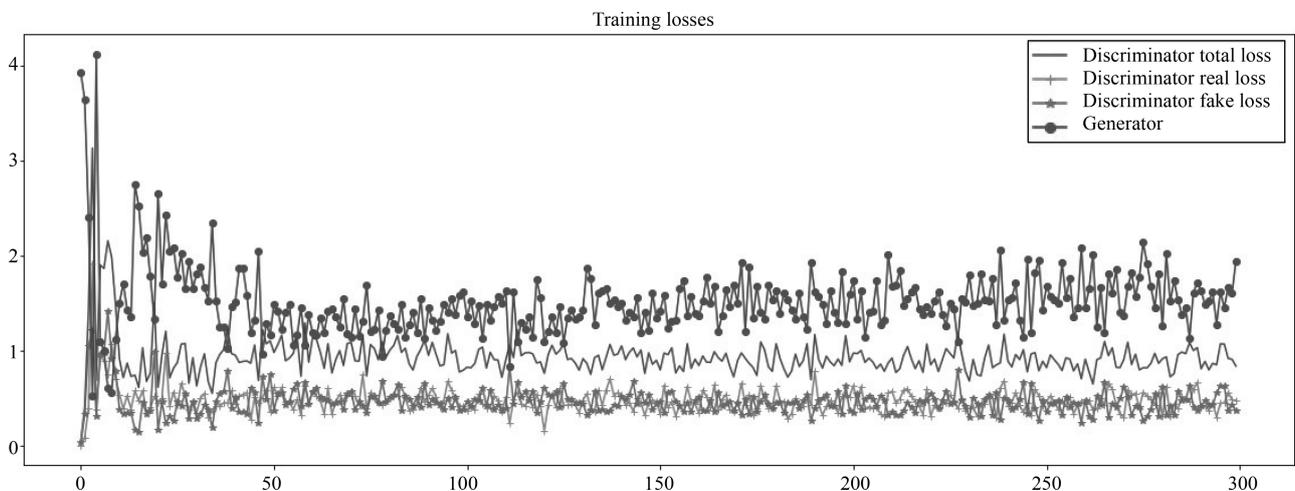


图 5 以 office 训练时的贝叶斯生成对抗网络损失函数图

Fig. 5 The loss function of the Bayesian generative adversarial networks trained base on the office datasets

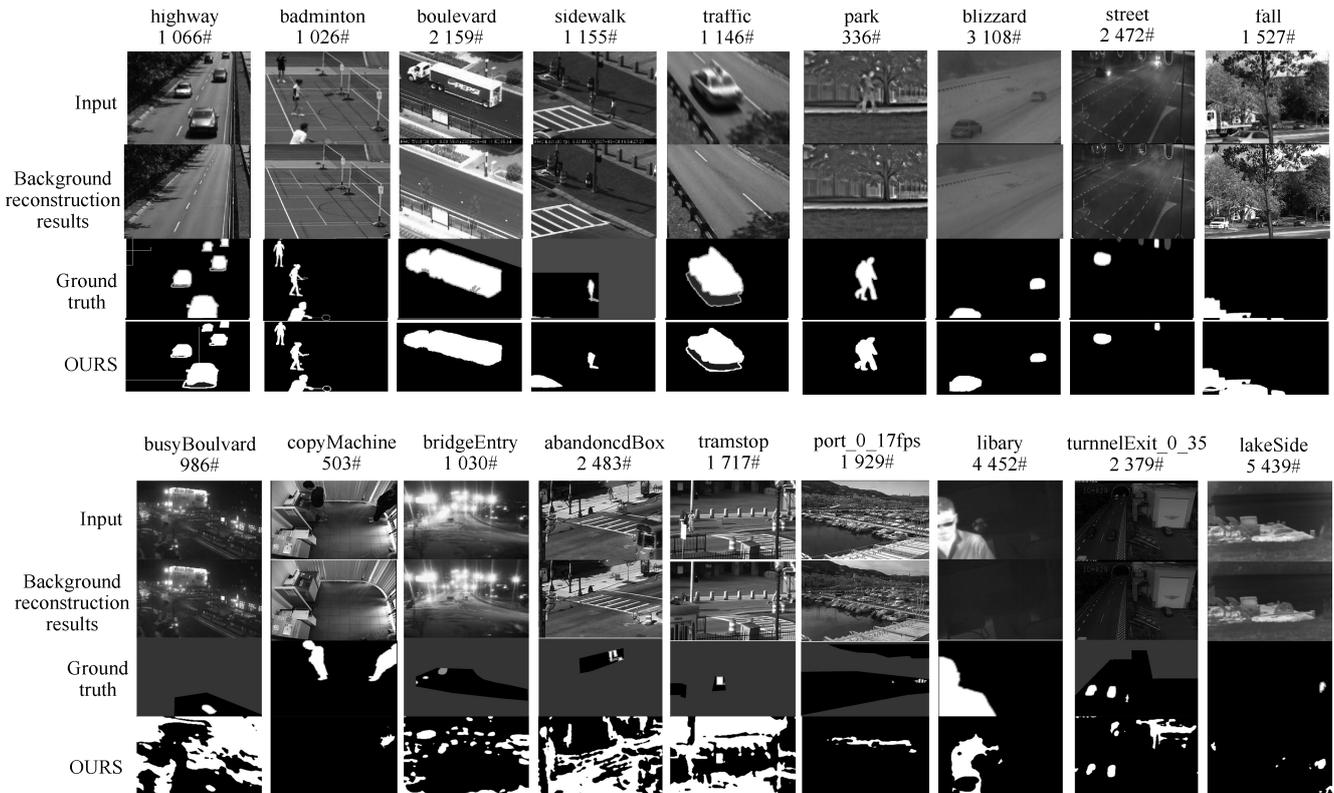


图6 背景重建与背景减除结果图

Fig. 6 The results of the background reconstruction and background subtraction

头是静止的, 所以网络也认为摄像头是静止的. 因此, 网络分别作出电车和电车所在的道路、行人和行人所在的道路都是动态目标的结果. 可见, 本文这种采用生成对抗网络的算法无法突破训练图像的限制, 而本文背景训练图像的来源是背景重建算法的结果, 所以本文生成对抗网络无法摆脱重建算法的影响.

图7中 highway 为在高速路上的车流, 大多数算法都能准确进行背景减除; 图7中 badminton 为人们在网球场练习网球, 多数算法对于前景中最近的人做不到准确判断, 而本算法明显优于其他算法, 这说明, 本文算法具有较强的场景适应的能力; 图7中 boulevard 为林荫大道中有车流和人流的情形, 可以看出本算法和 WeSamBE 对于这种复杂情形干扰较少, 因此可以准确进行背景减除. 图7中 sidewalk 是夕阳西下时人行道的情况, 光线较暖, 车辆经过时, 反光不强烈, 从结果可以看出本算法与 LFVBGM 优于其他算法, 这说明, 本文算法具有较强的场景适应的能力. 图7中 street 为光照发生突变的情况. 在夜空中车辆由黑暗处驶来, 此时由图7中结果可以看出, 其他方法将大部分场景尤其是汽车前灯区域误检测为背景, 而文中方法检测结果基本正确, 效果明显好于其他方法结果, 这说明, 本文算法整体具有鲁棒性好, 有着较强的场景适应的能力; 图7中 fall 为非静止背景的情况, 树叶、树枝一

直在摇晃, 当运动目标出现在视频中时, 由图7结果可以看出, 其他方法检测结果中都存在将摇曳的树枝误检测为前景点, 而本文方法检测结果虽然没有区分得足够清晰, 但是基本没有受到非静止背景的干扰, 整体效果优于其他方法. 这说明, 整体上本文算法具有鲁棒性好, 有着较强的场景适应的能力; 图7中 traffic 为偶然抖动的相机拍摄运动目标的情况, 其他方法检测结果中明显将背景显露区误判为前景出现了鬼影, LFVBGM 方法虽然判别出了因抖动出现的情况但是依然没有彻底消除, 而本文方法检测结果稍微有一些拖影, 但是将背景误认做目标的情况, 整体效果好于其他方法, 这说明, 本文算法具有较强的场景适应的能力. 因此, 从测试结果和对比结果整体来看, 算法具有鲁棒性高、对场景变化适应能力强的优点.

从召回率对比来看, 根据表1可知, 本文的检测算法相比其他检测方案在不同场景下, 在 baseline 系列数据集中, 本文算法比 Cascade CNN 算法低 0.8%, 排在第二位; 在 Shadow 系列数据集中, 本文算法比 Cascade CNN 算法低 5.2%, 排在第二位; 在 Low frame-rate 系列数据集, Bad weather 系列数据集, Air turbulence 数据集中都不同程度地略小于 Cascade CNN 算法, 排第二位; 从 Average 可以看出, 本文的检测方案所对应的召回率

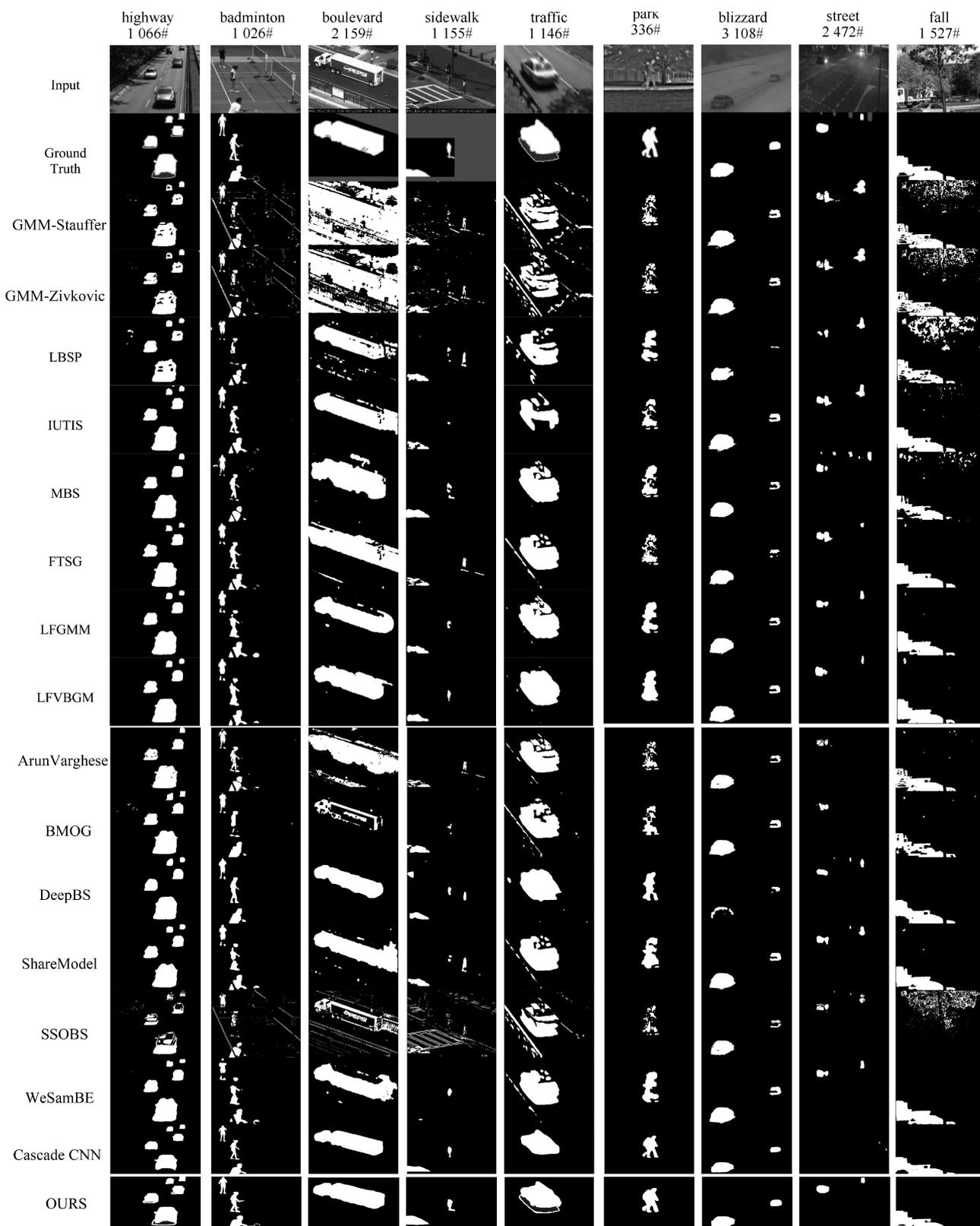


图 7 背景减法算法结果对比图

Fig. 7 Background subtraction algorithm results in comparison

排在第二位, 在不同检测方法仅次于 Cascade CNN 算法, 略逊于 0.2%。总体而言, 本文提出的检测方法在众多检测算法中整体效果表现优秀。另外, 本文算法在 Shadow、Inter.ob.motion、Dynamic background 数据库上的召回率的平均值是所有算法中最高的, 这说明本文算法在解决光照渐变和突变、非静止背景以及鬼影等问题是优秀的。

从精准率对比来看, 根据表 2 可知, 本文的检测算法相比其他检测方案在不同场景下, 在 Inter.ob.motion 系列数据集中, 本文算法比 Cascade CNN 算法低 4.0%, 排在第二位; 在 Low frame-rate 系列数据集中, 本文算法比 WeSamBE 算法低 7.5%, 差距较大, 排在第二位; 在其他的系列数据集中, 本文算法优于其他算法; 从 Average 可以看出, 本文的检测方案所对应的精确率排在第一位, 在不同检测方法中是最优的, OURS 相比排在第二位 Cascade CNN 算法, 精确率提升了 2.1%。总体而言, 本文提出的检测方法在众多检测算法中整体效果为优。另外, 本文算法在 Shadow、Inter.ob.motion、Dynamic

background 数据库上的精准率的平均值是所有算法中最高的, 这说明本文算法在解决光照渐变和突变、非静止背景以及鬼影等问题是优秀的。

从 F-measure 来看, 根据表 3 可知, 本文的检测算法相比其他检测方案在不同场景下, 在 Thermal 系列数据集中, 本文算法比 Cascade CNN 算法低 0.1%, 排在第二位; 在 Air turbulence 系列数据集中, 本文算法比 Cascade CNN 算法低 0.1%, 排在第二位; 从 Average 看, OURS 的 F-measure 在所有检测方案中是数值最高的, 比排在第二的 Cascade CNN 算法升高了 2.2%, 检测性能略提升。另外, 在 Shadow、Inter.ob.motion、Dynamic background 数据库上, 比第二位的 Cascade CNN 算法分别高 12.5%、1.5% 和 1.1%。这表明本文算法在解决光照渐变和突变、非静止背景以及鬼影等问题是优秀的。因此总体来说, 本文提出的算法具有鲁棒性高、对场景变化适应能力强的优点, 并且在解决光照渐变和突变、非静止背景以及鬼影等问题有较好的背景减除作用。

表 1 不同检测算法的召回率对比

Table 1 The recall rate of different detection algorithms are compared

Method	GMM- Stauffer	GMM- Zivkovic	L BSP	IUTIS	MBS	FTSG	LFGMM	LFVBGM	Arun Varghese	BMOG	DeepBS	Share Model	SSOBS	WeSamBE	Cascade CNN	BSGAN
database	Re (%)	Re (%)	Re (%)	Re (%)	Re (%)	Re (%)	Re (%)	Re (%)	Re (%)	Re (%)	Re (%)	Re (%)	Re (%)	Re (%)	Re (%)	Re (%)
Baseline	81.8	80.9	89.6	97.1	91.6	95.1	94.1	95.2	94.1	85.5	94.2	95.4	49.1	94.5	99.9	99.1
Dynamic background	83.8	80.2	76.7	87.8	76.4	86.9	77.7	88.7	88.7	90.1	85.4	75.9	52.3	67.9	84.8	98.8
Camera jitter	73.3	69.0	67.4	79.2	83.2	77.2	82.4	80.4	78.4	83.6	87.9	79.6	58.1	77.8	91.4	99.1
Shadow	79.6	77.7	87.8	94.8	79.2	92.1	94.2	92.7	74.8	50.9	57.4	71.8	21.6	74.7	96.6	91.4
Inter.ob.motion	51.4	54.7	55.9	69.9	75.3	76.2	81.4	72.0	91.7	85.9	95.8	94.5	50.2	94.0	93.0	98.1
Thermal	56.9	55.4	81.4	78.3	81.6	73.6	81.6	85.0	85.1	52.4	66.3	86.2	30.1	77.2	98.9	95.3
Bad weather	71.8	68.6	70.4	74.8	83.4	74.6	82.1	78.8	71.8	76.4	75.2	84.3	58.2	81.7	97.9	93.4
Low frame-rate	58.2	53.0	59.7	82.1	67.7	75.2	85.4	81.4	77.3	63.8	59.2	84.3	53.1	88.4	96.4	86.1
Night videos	52.6	48.0	51.0	56.6	55.4	61.1	65.7	66.1	36.1	64.9	53.2	59.9	44.7	63.7	94.2	91.4
PTZ	64.8	61.1	54.8	66.4	59.7	67.3	83.1	87.8	69.8	76.7	74.6	79.7	68.8	81.5	96.2	96.8
Air turbulence	79.1	77.9	76.1	68.6	60.4	61.1	80.5	81.2	81.2	68.7	79.8	79.1	74.4	71.8	96.1	93.1
Average	68.5	66.0	70.1	77.8	74.0	76.4	82.6	82.7	77.2	72.6	75.4	81.0	51.0	79.4	95.0	94.8

表 2 不同检测算法的精确率对比

Table 2 The precision rate of different detection algorithms are compared

Method	GMM- Stauffer	GMM- Zivkovic	L BSP	IUTIS	MBS	FTSG	LFGMM	LFVBGM	Arun Varghese	BMOG	DeepBS	Share Model	SSOBS	WeSamBE	Cascade CNN	BSGAN
database	Re (%)	Re (%)	Re (%)	Re (%)	Re (%)	Re (%)	Re (%)	Re (%)	Re (%)	Re (%)	Re (%)	Re (%)	Re (%)	Re (%)	Re (%)	Re (%)
Baseline	84.6	89.9	95.6	93.9	94.3	91.7	93.3	95.0	93.9	81.9	96.6	95.0	94.2	91.7	97.8	98.8
Dynamic background	59.9	62.1	59.2	92.8	86.5	91.3	89.2	90.4	90.4	75.8	90.8	91.9	10.1	89.3	96.7	96.8
Camera jitter	51.3	48.7	83.7	85.2	84.4	76.5	81.2	86.6	86.6	72.9	93.1	83.8	34.1	83.9	96.3	97.5
Shadow	71.6	72.3	87.7	85.8	82.6	85.3	86.5	87.1	83.9	68.2	82.5	75.8	65.3	78.8	78.2	98.8
Inter.ob.motion	66.9	64.4	71.0	81.5	74.2	78.1	65.8	74.9	48.2	53.7	47.4	59.3	36.9	55.3	94.4	90.4
Thermal	86.5	87.1	75.8	89.2	82.7	90.9	83.3	82.8	82.8	90.1	92.6	80.7	72.8	85.6	85.7	87.2
Bad weather	77.0	81.4	86.6	89.6	78.3	92.3	90.9	94.7	94.8	81.5	96.8	85.7	85.1	91.3	95.5	95.9
Low frame-rate	68.9	66.9	65.8	70.0	60.0	65.5	67.4	69.7	64.1	69.5	70.1	68.4	64.4	91.3	82.8	83.8
Night videos	41.3	42.3	44.9	51.3	49.0	49.0	53.4	55.4	65.4	46.1	83.7	58.5	51.5	58.3	88.1	88.0
PTZ	11.9	68.3	20.4	34.7	54.0	28.6	28.4	30.3	47.2	20.9	28.5	31.2	10.2	31.2	87.3	88.6
Air turbulence	42.9	34.9	59.7	92.6	62.0	90.4	78.1	78.4	68.1	76.8	90.8	75.6	9.8	83.7	89.3	89.5
Average	60.3	65.3	68.2	78.8	73.5	76.3	74.3	76.8	75.0	67.0	79.4	73.3	48.6	76.4	90.2	92.3

表3 不同检测算法的 F-measure
Table 3 F-measure of different detection algorithms

Method	GMM-1	GMM-2	LBSP	IUTIS	MBS	FTSG	LFGMM	LFVBGM	Arun	BMOG	DeepBS	Share	SSOBS	WeSamBE	Cascade	BSGAN
database	Re (%)	Varghese	Re (%)	Re (%)	Model	Re (%)	Re (%)	CNN	Re (%)							
Baseline	83.2	85.2	92.5	95.5	92.9	93.4	93.7	95.1	93.9	83.0	95.1	95.2	60.8	93.1	97.8	98.1
Dynamic background	69.9	70.0	66.8	90.2	81.1	89.0	83.1	89.5	89.4	79.3	87.6	82.2	16.1	74.4	96.5	97.6
Camera jitter	60.4	57.1	74.7	82.1	83.8	76.8	81.8	83.4	91.3	74.9	89.9	81.4	41.5	79.7	97.6	98.8
Shadow	75.4	74.9	87.7	90.1	80.9	88.6	90.2	89.8	77.6	52.9	60.9	67.3	30.2	73.9	85.1	97.6
Inter.ob.motion	58.1	59.2	62.6	75.3	74.7	77.1	72.8	73.4	87.1	83.9	89.9	84.6	75.2	86.9	94.1	95.6
Thermal	68.6	67.7	78.5	83.4	82.1	81.3	82.4	83.9	83.3	63.4	75.8	83.1	40.9	79.6	90.7	90.6
Bad weather	74.3	74.5	77.7	81.5	80.8	82.5	86.3	86.0	81.5	78.4	83.0	84.8	68.5	86.1	94.3	95.6
Low frame-rate	63.1	59.1	62.6	75.6	63.6	70.0	75.3	75.1	65.8	61.0	60.1	72.9	46.4	66.0	83.7	85.7
Night videos	46.3	45.0	47.8	53.8	52.0	54.4	58.9	60.3	41.5	49.8	58.4	54.2	44.6	59.3	89.6	90.6
PTZ	20.1	64.5	29.7	45.6	56.7	40.1	42.3	45.1	46.1	23.5	31.3	38.6	13.8	38.4	91.6	93.6
Air turbulence	55.6	48.2	66.9	78.8	61.2	72.9	79.3	79.8	64.5	69.3	84.6	73.4	15.2	75.4	91.8	91.7
Average	61.4	64.1	68.0	77.4	73.6	75.1	76.9	78.3	74.7	65.4	74.2	74.3	41.2	73.9	92.1	93.4

4 结论

本文提出了一种基于贝叶斯生成对抗网络的背景消减算法. 实验结果表明本文的算法在绝大多数情景下, 算法效果好于其他算法. 可见, 本文算法具有鲁棒性高、对场景变化适应能力强等优点. 但算法由于依赖于现有背景重建算法的结果, 所以本文算法有着无法突破背景重建算法提供的背景图像的局限性. 另外, 算法也不能够满足实时增量训练的需求. 在未来的研究中, 我们一方面考虑与基于贝叶斯理论的背景重建算法相结合, 摆脱对现有算法重建结果的依赖, 另一方面将要研究如何在并行化平台^[33] 下进行算法重构.

References

- Wang K F, Liu Y Q, Gou C, Wang F Y. A multi-view learning approach to foreground detection for traffic surveillance applications. *IEEE Transactions on Vehicular Technology*, 2016, **65**(6): 4144–4158
- Liu Y Q, Wang K F, Shen D Y. Visual tracking based on dynamic coupled conditional random field model. *IEEE Transactions on Intelligent Transportation Systems*, 2016, **17**(3): 822–833
- Zhang Hui, Wang Kun-Feng, Wang Fei-Yue. Advances and perspectives on applications of deep learning in visual object detection. *Acta Automatica Sinica*, 2017, **43**(8): 1289–1305 (张慧, 王坤峰, 王飞跃. 深度学习在目标视觉检测中的应用进展与展望. *自动化学报*, 2017, **43**(8): 1289–1305)
- Stauffer C, Grimson W E L. Adaptive background mixture models for real-time tracking. In: Proceedings of the 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). Fort Collins, CO, USA: IEEE, 1999. 252
- Zivkovic Z. Improved adaptive Gaussian mixture model for background subtraction. In: Proceedings of the 17th International Conference on Pattern Recognition (ICPR). Cambridge, UK: IEEE, 2004. 28–31
- St-Charles P L, Bilodeau G A. Improving background subtraction using local binary similarity patterns. In: Proceedings of the 2014 IEEE Winter Conference on Applications of Computer Vision (WACV). Steamboat Springs, CO, USA: IEEE, 2014. 509–515
- Fan J J, Xin Y Z, Dai F L, Hu B, Zhang J Q, Lu Q Y, et al. Distributed multi-camera object tracking with Bayesian Inference. In: Proceedings of the 2011 IEEE International Symposium on Circuits and Systems (ISCAS). Rio de Janeiro, Brazil: IEEE, 2011. 357–360
- Yan J H, Wang S F, Xie T X, Yang Y, Wang J Y. Variational Bayesian learning for background subtraction based on local fusion feature. *IET Computer Vision*, 2016, **10**(8): 884–893
- Wang Y, Jodoin P M, Porikli F, Konrad J, Benezeth Y, Ishwar P. CDnet 2014: an expanded change detection benchmark dataset. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Columbus, OH, USA: IEEE, 2014. 393–400
- Martins I, Carvalho P, Corte-Real L, Alba-Castro J L. BMOG: boosted Gaussian mixture model with controlled complexity. In: Pattern Recognition and Image Analysis. Faro, Portugal: Springer, 2017. 50–57
- Sehairi K, Chouireb F, Meunier J. Comparative study of motion detection methods for video surveillance systems. *Journal of Electronic Imaging*, 2017, **26**(2): 023025
- Chen Y Y, Wang J Q, Lu H Q. Learning sharable models for robust background subtraction. In: Proceedings of the 2015 IEEE International Conference on Multimedia and Expo (ICME). Turin, Italy: IEEE, 2015. 1–6
- Jiang S Q, Lu X B. WeSamBE: a weight-sample-based method for background subtraction. *IEEE Transactions on Circuits and Systems for Video Technology*, <https://ieeexplore.ieee.org/document/7938679/>
- Babae M, Dinh D T, Rigoll G. A deep convolutional neural network for video sequence background subtraction. *Pattern Recognition*, 2018, **76**: 635–649
- Saatchi Y, Wilson A G. Bayesian GAN. arXiv: 1705.09558, 2017
- Li Li, Lin Yi-Lun, Cao Dong-Pu, Zheng Nan-Ning, Wang Fei-Yue. Parallel learning — a new framework for machine learning. *Acta Automatica Sinica*, 2017, **43**(1): 1–8 (李力, 林懿伦, 曹东璞, 郑南宁, 王飞跃. 平行学习 — 机器学习的一个新型理论框架. *自动化学报*, 2017, **43**(1): 1–8)

- 17 Wang F Y, Wang X, Li L X, Li L. Steps toward parallel intelligence. *IEEE/CAA Journal of Automatica Sinica*, 2016, **3**(4): 345–348
- 18 Wang Kun-Feng, Gou Chao, Duan Yan-Jie, Lin Yi-Lun, Zheng Xin-Hu, Wang Fei-Yue. Generative adversarial networks: the state of the art and beyond. *Acta Automatica Sinica*, 2017, **43**(3): 321–332
(王坤峰, 苟超, 段艳杰, 林懿伦, 郑心湖, 王飞跃. 生成式对抗网络 GAN 的研究进展与展望. *自动化学报*, 2017, **43**(3): 321–332)
- 19 Wang K F, Gou C, Duan Y J, Lin Y L, Zheng X H, Wang F Y. Generative adversarial networks: introduction and outlook. *IEEE/CAA Journal of Automatica Sinica*, 2017, **4**(4): 588–598
- 20 Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. arXiv: 1406.2661, 2014
- 21 Wang Kun-Feng, Lu Yue, Wang Yu-Tong, Xiong Zi-Wei, Wang Fei-Yue. Parallel imaging: a new theoretical framework for image generation. *Pattern Recognition and Artificial Intelligence*, 2017, **30**(7): 577–587
(王坤峰, 鲁越, 王雨桐, 熊子威, 王飞跃. 平行图像: 图像生成的一个新理论框架. *模式识别与人工智能*, 2017, **30**(7): 577–587)
- 22 Wang Kun-Feng, Gou Chao, Wang Fei-Yue. Parallel vision: an ACP-based approach to intelligent vision computing. *Acta Automatica Sinica*, 2016, **42**(10): 1490–1500
(王坤峰, 苟超, 王飞跃. 平行视觉: 基于 ACP 的智能视觉计算方法. *自动化学报*, 2016, **42**(10): 1490–1500)
- 23 Wang K F, Gou C, Zheng N N, Rehg J M, Wang F Y. Parallel vision for perception and understanding of complex scenes: methods, framework, and perspectives. *Artificial Intelligence Review*, 2017, **48**(3): 299–329
- 24 Laugraud B, Piérard S, Braham M, Van Droogenbroeck M. Simple median-based method for stationary background generation using background subtraction algorithms. In: *New Trends in Image Analysis and Processing*. Italy: Springer, 2015. 477–484
- 25 Braham M, van Droogenbroeck M. Deep background subtraction with scene-specific convolutional neural networks. In: *Proceedings of the 2016 International Conference on Systems, Signals and Image Processing*. Bratislava, Slovakia: IEEE, 2016. 1–4
- 26 Zhao J, Meng D Y. FastMMD: ensemble of circular discrepancy for efficient two-sample test. *Neural Computation*, 2015, **27**(6): 1345–1372
- 27 Gangeh M J, Sadeghi-Naini A, Diu M, Tadayyon H, Kamel M S, Czarnota G J. Categorizing extent of tumor cell death response to cancer therapy using quantitative ultrasound spectroscopy and maximum mean discrepancy. *IEEE Transactions on Medical Imaging*, 2014, **33**(6): 1390–1400
- 28 He K M, Sun J. Convolutional neural networks at constrained time cost. In: *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, MA, USA: IEEE, 2015. 5353–5360
- 29 Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A, Chen X. Improved techniques for training GANs. arXiv: 1606.03498, 2016
- 30 Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville A. Improved training of Wasserstein GANs. arXiv: 1704.00028, 2017
- 31 Varghese A, Sreelekha G. Sample-based integrated background subtraction and shadow detection. *IPSJ Transactions on Computer Vision and Applications*, 2017, **9**: 25
- 32 Wang Y, Luo Z M, Jodoin P M. Interactive deep learning method for segmenting moving objects. *Pattern Recognition Letters*, 2017, **96**: 66–75
- 33 Bai Tian-Xiang, Wang Shuai, Shen Zhen, Cao Dong-Pu, Zheng Nan-Ning, Wang Fei-Yue. Parallel robotics and parallel unmanned systems: framework, structure, process, platform and applications. *Acta Automatica Sinica*, 2017, **43**(2): 161–175
(白天翔, 王帅, 沈震, 曹东璞, 郑南宁, 王飞跃. 平行机器人与平行无人系统: 框架、结构、过程、平台及其应用. *自动化学报*, 2017, **43**(2): 161–175)



郑文博 西安交通大学软件学院院长学制研究生. 主要研究方向为平行视觉, 平行学习, 机器学习.

E-mail: zwb2017@stu.xjtu.edu.cn

(**ZHENG Wen-Bo** Master-doctor combined candidate at the School of Software Engineering, Xi'an Jiaotong University. His research interest covers parallel vision, parallel learning, and machine learning.)



王坤峰 中国科学院自动化研究所复杂系统管理与控制国家重点实验室副研究员. 主要研究方向为智能交通系统, 智能视觉计算, 机器学习.

E-mail: kunfeng.wang@ia.ac.cn

(**WANG Kun-Feng** Associate professor at The State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences. His research interest covers intelligent transportation systems, intelligent vision computing, and machine learning.)



王飞跃 中国科学院自动化研究所复杂系统管理与控制国家重点实验室研究员. 国防科技大学军事计算实验与平行系统技术研究中心主任. 主要研究方向为智能系统和复杂系统的建模、分析与控制. 本文通信作者.

E-mail: feiyue.wang@ia.ac.cn

(**WANG Fei-Yue** Professor at The State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences. Director of the Research Center for Computational Experiments and Parallel Systems Technology, National University of Defense Technology. His research interest covers modeling, analysis, and control of intelligent systems and complex systems. Corresponding author of this paper.)