

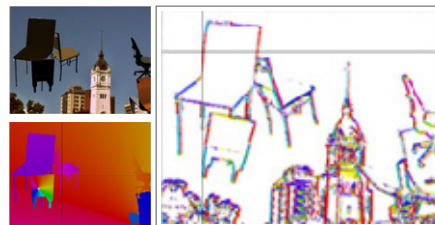


DOI: 10.12086/oe.2018.180027

基于深度卷积神经网络的运动目标光流检测方法

王正来, 黄敏*, 朱启兵, 蒋胜

江南大学轻工过程先进控制教育部重点实验室, 江苏 无锡 214122



摘要: 运动目标检测是物体检测领域的一个重要研究方向, 在目标识别中有着至关重要的作用。针对传统运动检测方法精度不高、无法对运动目标进行检测, 本文将深度卷积神经网络引入到运动目标光流检测中, 将前后帧图像及目标光流场图像作为网络的输入, 自适应地学习运动目标光流, 并通过对网络放大架构的优化及网络的精简, 同时采用数据增广等技术, 设计出精度与实时性兼顾的目标物体光流检测网络。实验结果表明, 本文方法在运动目标的光流场检测中有更好的表现, SS-sp 和 CS-sp 网络相比原网络在检测精度上均提高了约 5.0%, 同时大幅减少了网络的运行时间, 基本满足实时检测的要求。

关键词: 运动目标; 光流检测; 深度卷积神经网络; 网络结构优化

中图分类号: TP391.4

文献标志码: A

引用格式: 王正来, 黄敏, 朱启兵, 等. 基于深度卷积神经网络的运动目标光流检测方法[J]. 光电工程, 2018, 45(8): 180027

The optical flow detection method of moving target using deep convolution neural network

Wang Zhenglai, Huang Min*, Zhu Qibing, Jiang Sheng

Key Laboratory of Advanced Process Control for Light Industry, Ministry of Education, Jiangnan University, Wuxi, Jiangsu 214122, China

Abstract: Moving target detection is an important research direction of object detection, and it plays an important role in target recognition. The accuracy of traditional motion detection methods is low, which are unable to only detect the required moving target. In this study, deep convolutional neural network is introduced into the optical flow detection of moving target. In this method, a pair of images and optical flow fields of target are used as inputs of the network to adaptively study the target optical flow. Furthermore, through optimization of the expanding part of the network and the simplification of the network, and combined with many data augmentation technologies, the optical flow detection network of target object with both accuracy and real-time is designed. Experimental results show that the proposed method has better performance in the optical flow detection of moving target. SS-sp and CS-sp network are improved by about 5.0% compared to the original network on the precision and the runtime of the network is significantly reduced, which meet the requirements of real-time detection.

Keywords: moving target; optical flow detection; deep convolutional neural network; network structure optimization

收稿日期: 2018-01-17; 收到修改稿日期: 2018-05-18

基金项目: 江苏省研究生培养创新工程(SJLX16_0498); 江苏省政策引导类计划(产学研合作)-前瞻性联合研究项目(BY2016022-32)

作者简介: 王正来(1993-), 男, 硕士研究生, 主要从事检测与传感技术的研究。E-mail: wang930508@163.com

通信作者: 黄敏(1974-), 女, 博士, 教授, 主要从事图像处理的研究。E-mail: huangmzqb@163.com

Citation: Wang Z L, Huang M, Zhu Q B, et al. The optical flow detection method of moving target using deep convolution neural network[J]. *Opto-Electronic Engineering*, 2018, 45(8): 180027

1 引言

物体检测是模式识别与机器学习领域热门的研究方向，它在诸如安防、交通、互联网等领域有着广泛的应用，计算机实现物体自动检测能在一定程度上减轻人的负担^[1]。物体检测与图像分类相比更具挑战性，它具有多类别、多物体等特点，其定位加识别的检测模式更贴近于实际生活^[2]。运动目标检测是物体检测的一个重要研究方向，有别于通常的运动目标检测定义，随着深度学习理论的发展，深度卷积神经网络通过对候选区域生成、网络结构、训练方法等方面的改进和优化^[3-4]在物体检测领域取得了卓越成绩，运动目标检测也更加注重于对运动的目标进行检测而不是所有的运动物体，以此来弥补常规卷积神经网络无法提取物体动态特性的不足。

而传统的运动检测方法如帧间差分法、背景减除法、光流法等^[5-6]都无法做到对运动目标进行检测，同时存在着精度不高的问题。帧间差分法对于物体的运动速度有要求，同时易出现“空洞”现象^[7]。背景减除法的检测效果与背景的获取及更新有关，同时受到光线、天气等外界条件变化严重^[8]。光流法对噪声敏感，无法获得运动目标的准确轮廓^[9]，同时光流法亮度恒定的假设不适用于气体和液体检测^[10-11]。

为弥补传统运动检测方法精度不高、无法对运动目标进行检测的缺陷，本文将深度卷积神经网络引入到运动目标检测中，在 FlowNet^[12-13]基础上，通过对网络的优化和精简以及采用数据增广等技术，获得精度

与实时性兼顾的运动目标光流检测网络。最终实验结果表明，本文方法在检测精度和实时性方面均取得更好表现。

2 基于 FlowNet 的目标物体光流检测网络

将深度卷积神经网络应用于光流场检测时，如何对前后帧图像进行处理，完成运动信息的提取、计算仍是一个待解决的问题。如何在提高网络检测光流场精度的同时减少网络的运行时间，满足实时检测的需求也需要结合理论分析和相关试验的验证。

2.1 FlowNet 的网络架构

FlowNet 是深度卷积神经网络应用于光流场检测的典型代表，其对于前后帧图像独特的处理方式使得深度卷积神经网络成功应用于光流场检测，因此本文算法的网络以此为基础进行设计。

图 1 是 FlowNet 的网络结构，FlowNet 网络以前后帧图像及目标光流场图像作为网络的输入，分为收缩和放大两部分。网络的收缩部分采用卷积的形式完成特征图的获取，网络的放大部分采用反卷积及双线性插值的方式完成光流场图的放大。前后帧图像的相关性计算由网络的收缩部分完成，光流场的检测及放大由网络的放大部分完成。由于网络没有对输入图像采用尺寸归一化，同时没有设置全连接层，因此输入图像的尺寸受到限制，理论上不能小于 64×64。具体介绍如下。

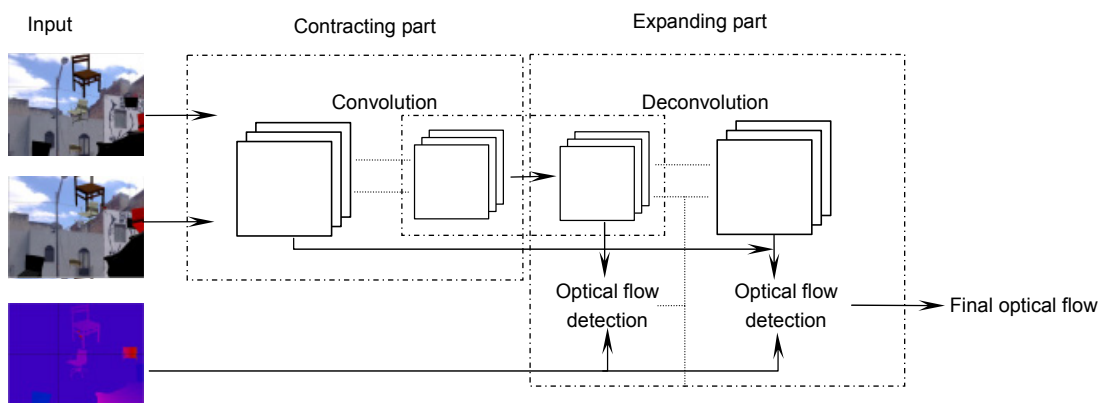


图 1 FlowNet 网络结构

Fig. 1 Network structure of FlowNet

2.2 FlowNet 网络的收缩部分

FlowNet 网络的收缩部分有 2 种基础架构可供选择, 分别为 FlowNet Simple(FN-S)和 FlowNet Correlation(FN-C), 分别对输入的前后帧图像采取不同的关联方法, 以卷积的形式实现图像特征图的获取, 多个卷积层的叠加大幅度地缩小了输入图像的尺寸, 使得网络具备对较大位移的估计。

2.2.1 FN-S 收缩架构

FN-S 收缩架构采用将前后帧图像堆叠的方式作为网络的输入, 即网络的输入为 $c \times w \times h$, c 代表前后帧图像的通道总数, 以 RGB 图像为例, $c=6$, w 、 h 分别为图像的宽和高。FN-S 收缩架构如图 2(a)所示, FN-S 收缩架构中的每个卷积层后都设置了 1 个激活函数单元(ReLU)。

2.2.2 FN-C 收缩架构

FN-C 收缩架构借鉴了光流法中寻找帧与帧之间对应关系的原理^[14], 增加了关联层。FN-C 收缩架构如图 2(b)所示, 前后帧图像分别作为网络的输入, 通过关联层计算前后帧图像的相关性, 其余各层的配置与 FN-S 架构完全一致。

关联层的作用是通过比较前后帧图像的特征图得出之间的联系, 输入为 2 个大小为 $c \times w \times h$ 的特征图(c 、 w 、 h 分别为通道数、宽、高), 设为 f_1 、 f_2 。 x_1 、 x_2 分别是 f_1 、 f_2 上的点, 那么以 x_1 为中心的块和以 x_2 为中心的块之间的联系定义如下(块长为 $K=2k+1$, 大小为 $K \times K$, o 为块中的点, $o \in [-k, k] \times [-k, k]$) :

$$corr(x_1, x_2) = \sum < f_1(x_1 + o), f_2(x_2 + o) > \quad (1)$$

该公式操作是特征图与特征图之间卷积, 因此无

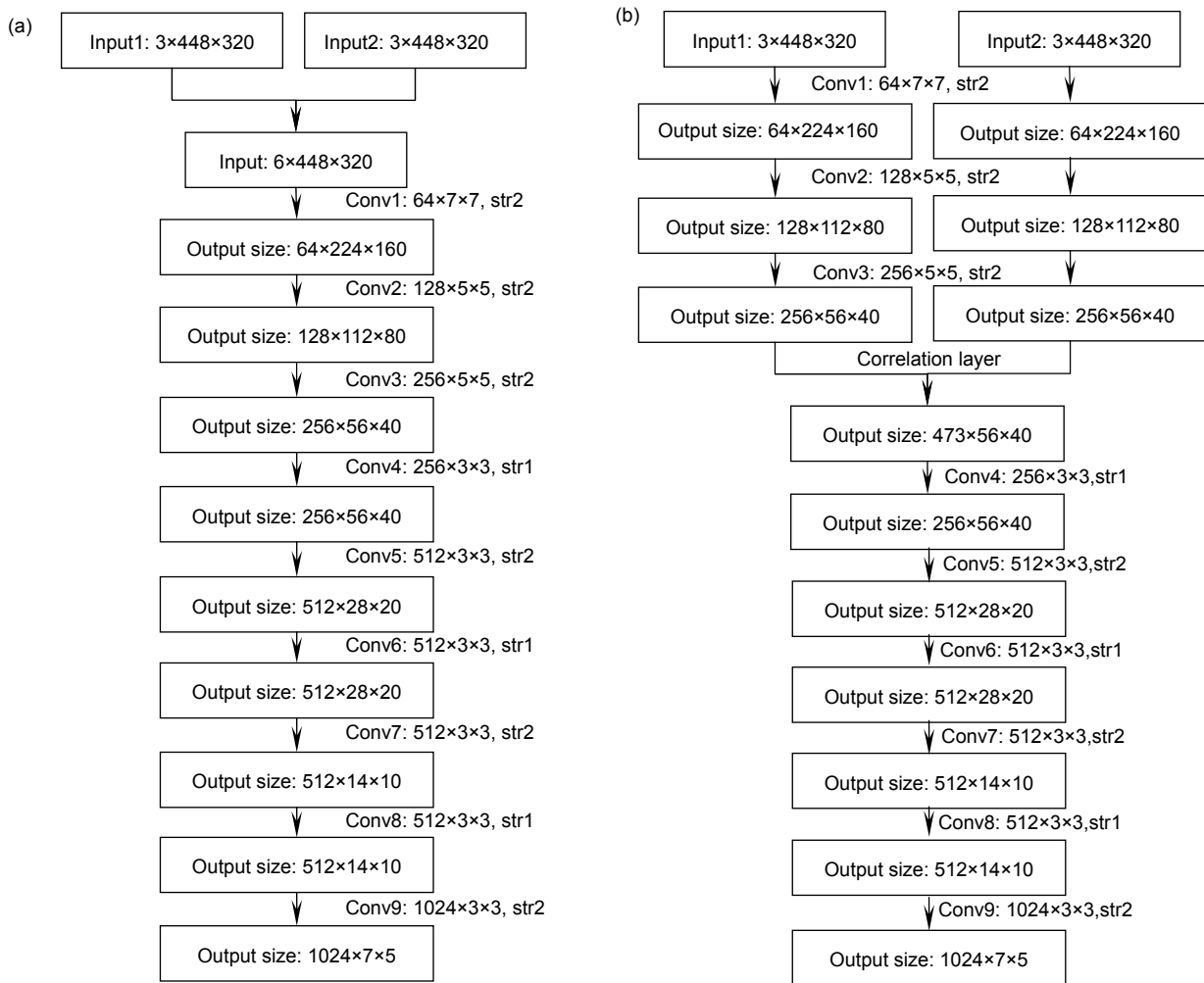


图 2 收缩架构。(a) FN-S 收缩架构; (b) FN-C 收缩架构

Fig. 2 Contracting structure. (a) FN-S contracting structure; (b) FN-C contracting structure

可训练权重。同时为减少计算复杂度，设定了约束位移范围 d ，即当 x_1 确定时，选择 f_2 中对应位置附近距离为 $D=2d+1$ 内的点为计算点，即以 x_1 为中心的块只与 f_2 中以附近 D 范围内的点为中心的块计算联系。关联层步长设置为 1，最后得到 $D^2 \times w \times h$ 个特征图，将前后帧中前帧图像通过 1×1 的卷积后与 D^2 个特征图拼接，得到关联层的输出为 1 个 $(D^2+32) \times w \times h$ 的特征图。

2.3 FlowNet 网络的放大部分

FlowNet 网络放大部分的作用是在特征图上完成光流场检测并逐步精细、放大至输入图像尺寸。FlowNet 网络放大部分的结构示意图如图 3 所示，由反卷积层、拼接层、卷积层和双线性插值层组成。反卷积层的作用是放大输入图像的尺寸，反卷积层的通道数设置与收缩架构的偶数卷积层一致。拼接层将反卷积层得到的特征图、收缩部分对应层的特征图以及光流场检测图拼接，好处在于保留了检测信息的同时结合了特征图信息，有助于检测结果的修正，提高检测精度。卷积层的作用是检测光流场，通道数为 2 对应于光流场的 2 个分量(方向和速度)，并用于损失函

数的计算。放大部分共设有 4 个反卷积-拼接结构，最后得到的光流场检测图宽、高是输入图像的 1/4，通过双线性插值层得到和输入图像相同分辨率的光流场检测图。

2.4 叠加网络

理论上来说，增加网络的深度可提高网络的学习能力，提升网络的性能^[15-16]。文献[12]通过叠加网络的形式提升网络的深度，实现光流场检测的精度提升。图 4 为一个两层叠加网络 CS 的结构图，叠加网络在两个网络中间增加了扭曲(warp)层和拼接层，扭曲层是将输入图像中的后帧图像根据第一个网络输出的光流估计场进行位置变化，使得下一个网络可以专注于第一幅图像与扭曲操作后第二幅图像间剩余增量的检测。

在扭曲操作中 I_1 、 I_2 分别为前帧图像与后帧图像， \tilde{I}_2 为扭曲操作后的后帧图像， (x, y) 为图像上的坐标点， $w_i = (u_i, v_i)^T$ 为 i 点处的估计光流，扭曲操作定义如式(2)、式(3)所示：

$$\tilde{I}_2(x + w(x)) = I_2(x + u_i, y + v_i) \quad (2)$$

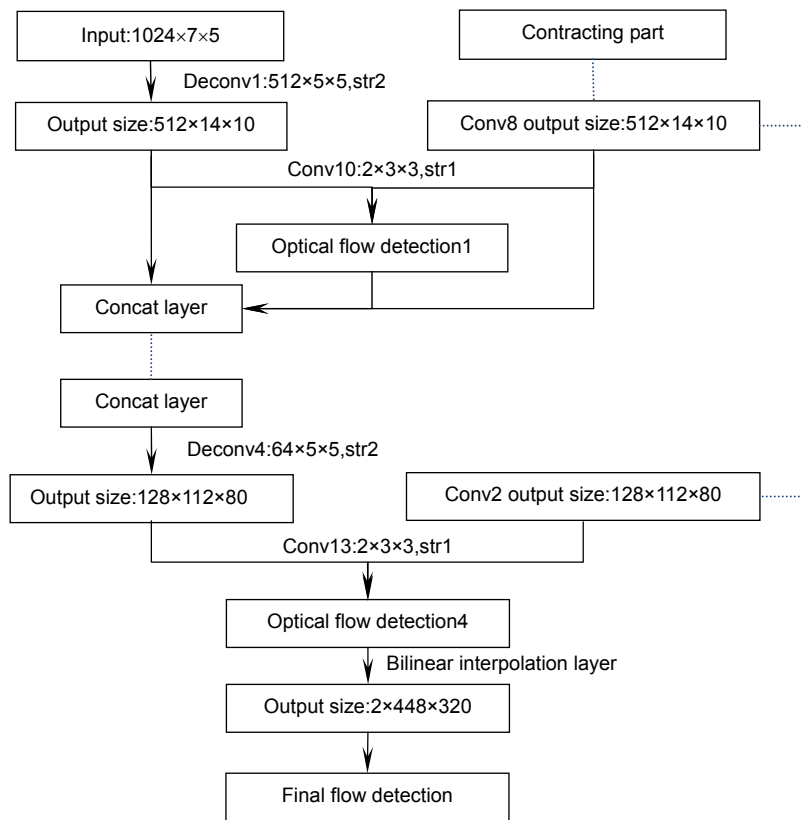


图 3 FlowNet 网络放大部分架构

Fig. 3 Expanding structure of FlowNet network

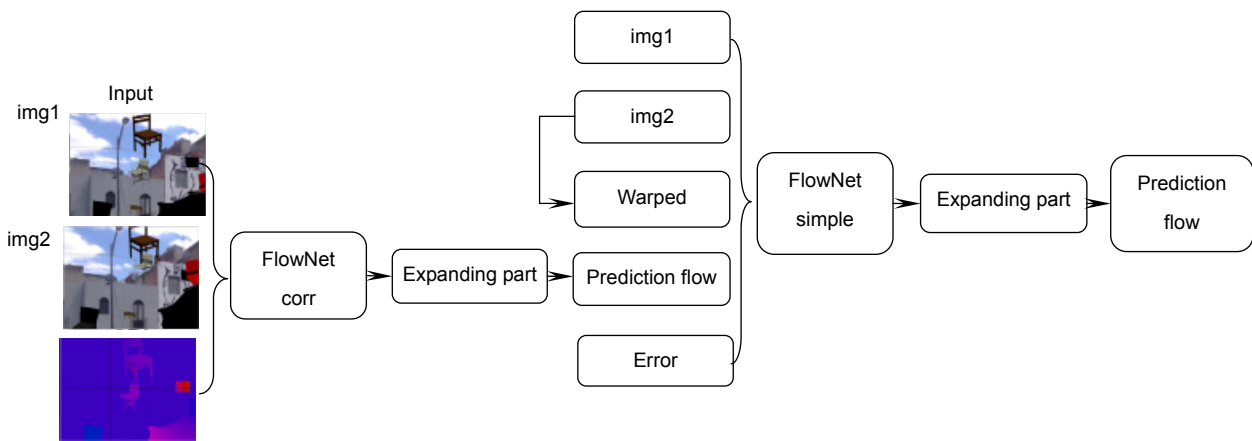


图 4 叠加网络结构图
Fig. 4 Structure of stack network

$$J_{w(x)} = \begin{cases} \tilde{I}_2(x+w(x)) & , x+w(x) \text{ in img} \\ 0 & , \text{otherwise} \end{cases} \quad (3)$$

拼接层将前帧图像、扭曲操作后的后帧图像、误差图像以及检测光流场图像进行拼接，作为下一个网络的输入。误差图像定义如下：

$$e_i = \|\tilde{I}_2 - I_1\| \quad (4)$$

考虑到叠加网络中第一个网络在相同配置下多次训练检测结果可能均不相同、检测效果存在好坏，以及采取了拼接的形式作为下一层网络的输入，FN-C 收缩架构中关联层的形式不再适用于叠加网络中后续的网络，因此本文在配置叠加网络时 FN-C 收缩架构只设置在第一个网络中。

2.5 网络优化及精简

作为应用于物体检测领域的光流场检测方法，不仅网络的检测精度有要求，单对图像的检测时间即网络的实时性也十分重要。本文在 FlowNet 的基础上对网络的放大部分进行了优化，进一步提升了网络的检测精度，同时对优化后网络采取了网络精简的方式，兼顾了网络的检测精度和实时性。

2.5.1 网络优化

在光流神经网络中，网络通过收缩网络获取前后帧图像之间的光流场信息，再通过放大网络对获取的光流场信息进行放大，因此光流场的检测效果依赖于放大网络的学习效果和精细程度，本文对网络的放大部分进行了如下优化：优化后的放大网络缩小了反卷积层的卷积核大小，出于网络运行时间的考虑，卷积核大小缩小为 4×4，这样做的好处是增加了特征图中

卷积核滑动中重复区域的计算，有利于边缘区域的学习区分，反卷积层的其余设置不变；同时每个拼接层后添加了一个卷积层，拼接层的输出原来单独作为光流场检测的卷积层的输入，现在更改为光流场检测的卷积层和新增卷积层的共同输入，新增的卷积层卷积核大小设置为 3×3，步长为 1，通道数与上一层反卷积层一致，并且通过补零的方式没有改变输入前后特征图的尺寸。新增卷积层的作用是减少了特征信息的冗余，由拼接层输出的通道数缩减为反卷积层的通道数，但是增加了网络的深度，同时保留了光流检测图和特征图的信息，这使得网络保留了有效的信息，有利于检测光流场边缘的平滑，提升网络的检测精度。优化后的放大网络基础结构变为反卷积层-拼接层-卷积层，并且比原网络多增设一个基础结构，检测得到的光流检测图为输入图像分辨率的 1/2，进一步提升网络的深度。优化前后放大部分网络设置如表 1 所示。

2.5.2 网络精简

通过对光流神经网络采取叠加网络的方式以及对于网络中放大网络的优化，在增加网络的深度提升网络性能的同时，大幅度增加了网络所需要学习的参数，使得网络的检测时间增加。而在实际应用中，网络检测的实时性也是一个重要的参考标准，本课题参考文献[12]对于网络宽度、深度与性能所做的实验，对优化后网络的宽度进行了精简，在系数 $\sigma=0.375$ 时，网络的宽度精简后实时性与检测精度兼顾得最好，优化后网络的最大通道数为 $1024 \times 0.375 = 384$ ，收缩网络的其余设置依次对应进行调整，精简后的网络放大网络设置如表 2 所示。

表 1 优化前后放大部分网络比对

Table 1 Comparison between original expanding network and network after optimization

Network name	Network configurations					
	Deconv1	Conv10	Deconv2	Conv11	Deconv3	Conv12
Expand	512×5×5	-	256×5×5	-	128×5×5	-
	str2		str2		str2	
Expand-c	512×4×4	512×3×3	256×4×4	256×3×3	128×4×4	128×3×3
	str2	str1	str2	str1	str2	str1

Network name	Deconv4	Conv13	Deconv5	Conv14	Bilinear interpolation	
	Expand	64×5×5	-	-	-	√
str2						
Expand-c	64×4×4	64×3×3	32×4×4	32×3×3	√	
	str2	str1	str2	str1		

表 2 精简后放大部分网络

Table 2 Expanding network after simplification

Network name	Network configurations					
	Deconv1	Conv10	Deconv2	Conv11	Deconv3	Conv12
Expand-sp	192×4×4	192×3×3	96×4×4	96×3×3	48×4×4	48×3×3
	str2	str1	str2	str1	str2	str1

Network name	Deconv4	Conv13	Deconv5	Conv14	Bilinear interpolation	
	Expand-sp	24×4×4	24×3×3	12×4×4	12×3×3	√
str2		str1	str2	str1		

3 网络的训练与分析

3.1 数据与数据增广

光流场检测精度一般通过比较估计光流值与真实光流值进行评估，因此评估前需要预先知道真实光流值。本文选择了 FlyingChairs 数据库^[13]用于网络的训练和测试，数据库含有 22872 对图像及对应真实光流场，图像大小为 512×384，同时选择了椅子作为检测目标，符合本文与传统光流法比较的需求，也适用于大量数据进行训练的卷积神经网络。本文选择了其中 22000 对图像用于网络的训练和测试，其中训练集 20000 对，验证集 1000 对，测试集 1000 对。

在深度卷积神经网络的训练过程中通常会遇到过拟合的问题，即在训练集中损失函数值很小而在测试集中很大。数据增广技术是指对原图像进行各种变换，以增加样本的多样性，从而达到防止过拟合的目的，增强模型的鲁棒性。在本文算法中采取了多种数据增

广方法。

1) 图像的缩放及随机旋转：将训练集中 512×384 的原图裁剪 5 个 448×320 大小的块，每个角落一个，中心一个，然后随机旋转一定角度(0~180°)形成新的图像，训练集样本可扩增至原来的(4+1)×2=10 倍。

2) 噪声扰动：对图像的每个像素添加随机扰动，采用均值为 0，方差为 1 的高斯噪声。

3.2 网络训练方法

实验中各网络参数设置一致，如下：网络参数更新方法为随机梯度下降法(stochastic gradient descent, SGD)。每一批量样本的数量为 8，momentum 系数设置为 0.9。本文采用均值为 0、方差为 2/n 的高斯分布对网络所有层的权重进行初始化^[17-18]，权重衰减系数为 0.0004，偏置初始化为常量 0。网络的初始学习率为 10⁻⁵，迭代总次数为 120 万次，学习率改变策略为：multistep，分别在 40 万、60 万、80 万、100 万次时重新计算学习率，gamma 设置为 0.5。

计算机 CPU 配置为 Intel i7-6850k 3.6 G ,显卡为 2 个 NVIDIA GTX 1080 Ti , 操作系统为 Linux 系统 , 在 caffe 框架基础上开展相关实验研究。

3.3 损失函数及质量评价

损失函数的计算在网络的放大部分通过检测光流场与目标光流场的比对完成 , 本文采用端点误差 (endpoint error, EPE , 用 E_{EP} 表示) 作为网络的损失函数 , 定义如下 :

$$E_{EP} = \sum_{i=1}^N \sqrt{(u_i^{gt} - u_i^e)^2 + (v_i^{gt} - v_i^e)^2} \quad (5)$$

其中 : N 为像素点总数 , (u_i^{gt}, v_i^{gt}) 为像素点 i 处光流真实值 , (u_i^e, v_i^e) 为光流估计值。同时采用平均端点误差 (average endpoint error, AEE , 用 E_{AEP} 表示) 和平均角误差 (average angular error, AAЕ , 用 E_{AA} 表示) 来评估整体光流。定义 :

$$u = \sqrt{(u_i^{gt})^2 + (u_i^e)^2 + 1.0} \quad ,$$

$$v = \sqrt{(v_i^{gt})^2 + (v_i^e)^2 + 1.0} \quad .$$

则 :

$$E_{AEP} = \frac{1}{N} E_{EP} \quad , \quad (6)$$

$$E_{AA} = \frac{1}{N} \sum_{i=1}^N \arccos \left(\frac{u_i^{gt} \times u_i^e + v_i^{gt} \times v_i^e + 1.0}{u \times v} \right) \quad . \quad (7)$$

同时本文还采用了光流场可视化技术 , 通过直观显示的方式对检测结果进行比较。本文基于孟塞尔颜色系统 (Munsell color system) 实现对二维向量场——

光流场的彩色图像显示。

3.4 实验结果与分析

针对经典的光流算法 Horn-Schunck^[19]、Lucas-Kanade^[20]、FlowNet 经典网络以及本文优化和精简后的网络 , 本文开展了相关实验。Horn-Schunck 算法是基础的亮度恒常及全局平滑假设算法 , 是变分偏微分光流算法的起点。Lucas-Kanade 算法是局部光流优化算法 , 假设相邻像素间光流相等 , 充分利用了窗口领域约束。实验结果比对及误差列表如图 5、图 6 和表 3 所示。

从图 5 中可以看到 , 传统的光流方法漏检了右上角的椅子。结合图 5、图 6 可以看出 , 传统的光流方法无法做到只针对检测目标——椅子进行光流检测 , 同时对于背景的光流变化区分也略有欠缺 , 在实际使用中存在着精度不高的缺点。而本文算法能做到对目标光流场的准确检测 , 同时能做到背景的光流变化的精确显示 , 但是还需要在细节方面提升网络的检测性能 , 如椅子的脚有漏检。同时从表 3 中也可以看出 , 传统的光流方法在误差和运行时间上也要远远高于本文算法。

在表 3 中 , 网络 S、C 是分别基于 FlowNet 两种收缩架构和放大架构建立的网络 , S 采用了 FN-S 收缩架构 , C 采用了 FN-C 收缩架构 , SS 和 CS 网络是基于 FlowNet 的两层叠加网络 , SS 网络是两个 S 网络的叠加 , CS 网络是一个 C 网络和一个 S 网络的叠加。表中

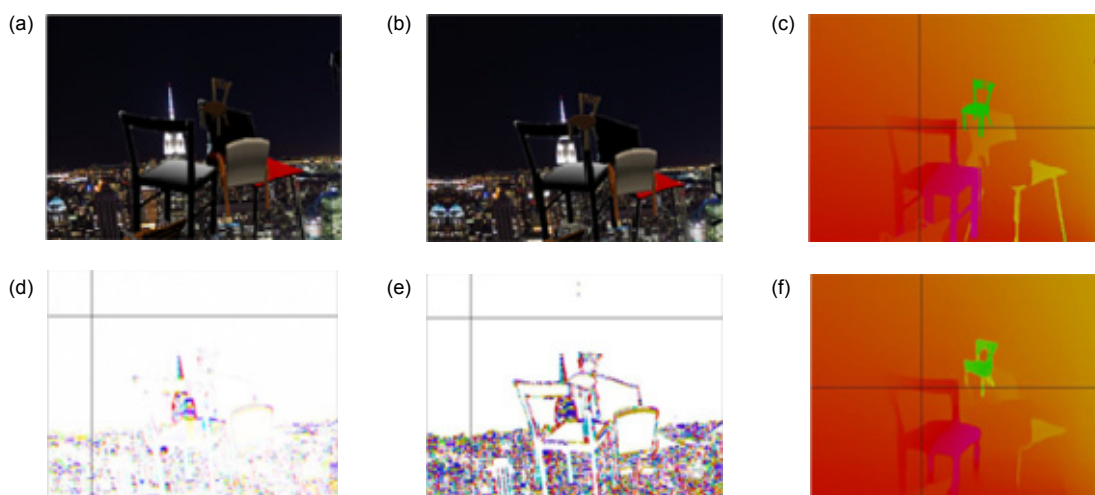


图 5 部分实验结果 1。(a) 原图 1; (b) 原图 2; (c) 目标光流场图; (d) Horn-Schunck 算法结果; (e) Lucas-Kanade 算法结果; (f) 本文 CS-sp 算法结果

Fig. 5 Part of experimental results 1. (a) Original image 1; (b) Original image 2; (c) Optical flow image of target; (d) Result of Horn-Schunck algorithm; (e) Result of Lucas-Kanade algorithm; (f) Result of CS-sp algorithm

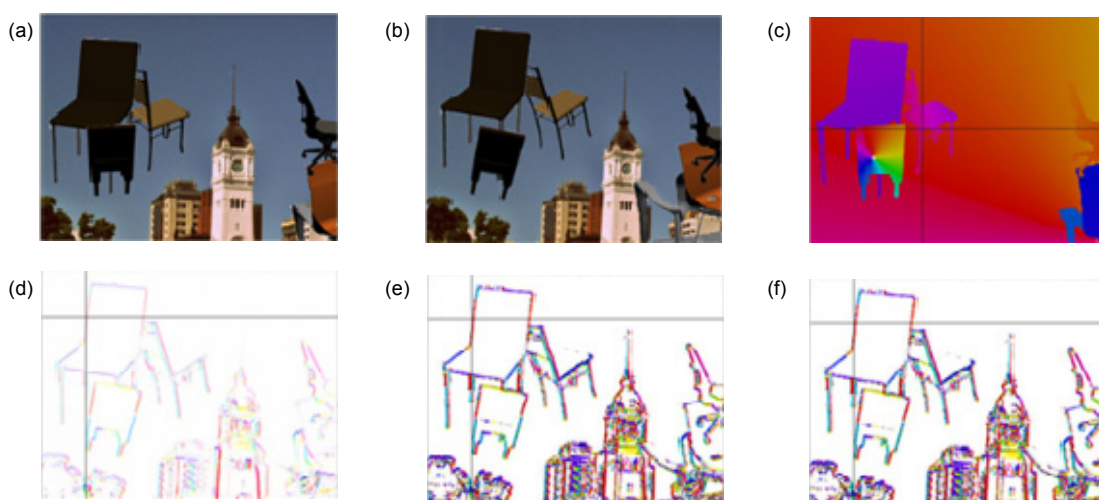


图 6 部分实验结果 2。(a) 原图 1; (b) 原图 2; (c) 目标光流场图; (d) Horn-Schunck 算法结果; (e) Lucas-Kanade 算法结果; (f) 本文 CS-sp 算法结果

Fig. 6 Part of experimental results 2. (a) Original image 1; (b) Original image 2; (c) Optical flow image of target; (d) Result of Horn-Schunck algorithm; (e) Result of Lucas-Kanade algorithm; (f) Result of CS-sp algorithm

表 3 经典光流法、FlowNet 以及本文网络的检测误差

Table 3 The detection errors of classical optical flow, FlowNet and our network

Network name	AEE	AAE	Runtime/ms	Network name	AEE	AAE	Runtime/ms
Horn-Schunck	9.7983	1.1958	1260	Lucas-Kanade	9.7813	1.1744	1220.4
S	2.9283	0.2754	91	SS	2.6482	0.2560	170
S-c	2.7173	0.2506	105	SS-c	2.4224	0.2343	190
S-sp	2.7852	0.2609	18	SS-sp	2.5146	0.2427	33
C	2.8451	0.2619	95	CS	2.5596	0.2535	179
C-c	2.7078	0.2527	108	CS-c	2.3438	0.2287	195
C-sp	2.8375	0.2650	20	CS-sp	2.4325	0.2376	34

c 结尾命名的网络采用了本文对于放大部分的优化，而 sp 结尾命名的网络同时采用了本文优化和精简的方法。

由表 3 可知，经放大部分优化后的网络比起原网络在检测误差方面均有不同程度的降低，在 CS 网络表现得最好，AEE 和 AAE 误差分别降低了 8.4%和 9.8%；在 C 网络上表现略有不足，AEE 和 AAE 误差分别降低 4.8%和 3.5%。在放大部分优化的基础上对网络采取了精简的措施，网络的运行时间均有大幅度下降，基本达到了实时检测的需求，并且网络的性能仍然优于 FlowNet 网络，CS-sp 和 SS-sp 网络在 AEE 误差上分别降低了约 5.0%，在 AAE 误差上分别降低了 6.3%和 5.2%，说明本文方法兼顾了检测精度和实时性。

本文在表 3 的基础上对三层叠加网络和更多层叠

加网络进行了实验，在三层网络上取得了更高的检测精度，但是更多层的叠加网络训练涉及到了网络性能退化问题，即深度网络在训练更深层次网络时出现了网络训练收敛问题，需要通过进一步的研究予以解决。

通过以上实验可以得出，优化和精简后的光流场检测网络保留了深度卷积神经网络针对目标检测光流的优点，更符合物体检测在实际中的应用，同时兼顾了网络的检测精度和实时性，大幅度减少网络运行时间的同时提升了网络的检测精度，但是在更深层叠加网络的训练及性能问题还需要更多的研究、测试。

4 结 论

本文对深度卷积神经网络在光流场的检测应用方面展开了相关试验。在 FlowNet 基础上，对网络放大

部分进行了优化,同时对网络进行了精简,在训练样本上通过多种数据增广技术进行训练,并且对测试结果采取了可视化手段,更直观地进行光流场评估。试验结果证明,基于深度卷积神经网络的运动目标光流检测网络在检测精度和实时性方面均取得了更好表现。

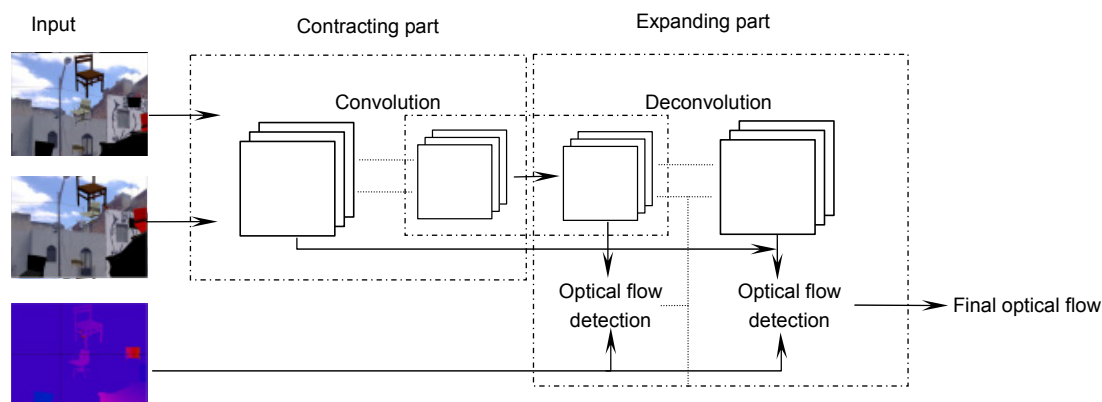
参考文献

- [1] Huang K Q, Ren W Q, Tan T N. A review on image object classification and detection[J]. *Chinese Journal of Computers*, 2014, **37**(6): 1225–1240.
黄凯奇, 任伟强, 谭铁牛. 图像物体分类与检测算法综述[J]. *计算机学报*, 2014, **37**(6): 1225–1240.
- [2] Lu H T, Zhang Q C. Applications of deep convolutional neural network in computer vision[J]. *Journal of Data Acquisition and Processing*, 2016, **31**(1): 1–17.
卢宏涛, 张秦川. 深度卷积神经网络在计算机视觉中的应用研究综述[J]. *数据采集与处理*, 2016, **31**(1): 1–17.
- [3] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, **39**(6): 1137–1137.
- [4] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//*Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition*, 2015: 3431–3440.
- [5] Wang S F, Yan J H, Wang Z G. Improved moving object detection algorithm based on local united feature[J]. *Chinese Journal of Scientific Instrument*, 2015, **36**(10): 2241–2248.
王顺飞, 闫钧华, 王志刚. 改进的基于局部联合特征的运动目标检测方法[J]. *仪器仪表学报*, 2015, **36**(10): 2241–2248.
- [6] Liu W, Zhao W J, Li C, et al. Detecting small moving target based on the improved ORB feature matching[J]. *Opto-Electronic Engineering*, 2015, **42**(10): 13–20.
刘威, 赵文杰, 李成, 等. 基于改进 ORB 特征匹配的运动小目标检测[J]. *光电工程*, 2015, **42**(10): 13–20.
- [7] Huang C Q. Motion detection and tracking based on gaussian mixture model and kalman filter[D]. Kunming: Yunnan University, 2010.
黄超群. 基于混合高斯模型和 Kalman 滤波器的运动目标检测与跟踪[D]. 昆明: 云南大学, 2010.
- [8] Zhang J M, Wang B. Moving object detection under condition of fast illumination change[J]. *Opto-Electronic Engineering*, 2016, **43**(2): 14–21.
张金敏, 王斌. 光照快速变化条件下的运动目标检测[J]. *光电工程*, 2016, **43**(2): 14–21.
- [9] Yuan G W, Chen Z Q, Gong J, et al. A moving object detection algorithm based on a combination of optical flow and three-frame difference[J]. *Journal of Chinese Computer Systems*, 2013, **34**(3): 668–671.
袁国武, 陈志强, 龚健, 等. 一种结合光流法与三帧差分法的运动目标检测算法[J]. *小型微型计算机系统*, 2013, **34**(3): 668–671.
- [10] Luo S, Jiang Y Z. State-of-art of video based smoke detection algorithms[J]. *Journal of Image and Graphics*, 2013, **18**(10): 1225–1236.
罗胜, Jiang Y Z. 视频检测烟雾的研究现状[J]. *中国图象图形学报*, 2013, **18**(10): 1225–1236.
- [11] Shi L F, Long F, Zhan Y J, et al. Video-based fire detection with spatio-temporal SURF and color features[C]//*Proceedings of 2016 12th World Congress on Intelligent Control and Automation*, 2016: 258–262.
- [12] Dosovitskiy A, Fischer P, Ilg E, et al. FlowNet: learning optical flow with convolutional networks[C]//*Proceeding of 2015 IEEE International Conference on Computer Vision*, 2015: 2758–2766.
- [13] Ilg E, Mayer N, Saikia T, et al. FlowNet 2.0: evolution of optical flow estimation with deep networks[C]//*Proceeding of 2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 1647–1655.
- [14] Deng H. Research on neuron image recognition[D]. Wuhan: Hubei University, 2016.
邓昊. 神经元图像识别相关方法探究[D]. 武汉: 湖北大学, 2016.
- [15] Ouyang P, Hu H, Shi Z Z. Plankton classification with deep convolutional neural networks[C]//*Proceeding of 2016 IEEE Information Technology, Networking, Electronic and Automation Control Conference*, 2016: 132–136.
- [16] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//*Proceeding of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 770–778.
- [17] He K M, Zhang X Y, Ren S Q, et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification[C]//*Proceeding of 2015 IEEE International Conference on Computer Vision*, 2015: 1026–1034.
- [18] He K M, Sun J. Convolutional neural networks at constrained time cost[C]//*Proceeding of 2015 IEEE Conference on Computer Vision and Pattern Recognition*, 2015: 3992–4000.
- [19] Horn B K P, Schunck B G. Determining optical flow[J]. *Artificial Intelligence*, 1981, **17**(1–3): 185–203.
- [20] Lucas B D, Kanade T. An iterative image registration technique with an application to stereo vision (DARPA)[C]//*Proceedings of the 1981 DARPA Image Understanding Workshop*, 1981: 121–130.

The optical flow detection method of moving target using deep convolution neural network

Wang Zhenglai, Huang Min*, Zhu Qibing, Jiang Sheng

Key Laboratory of Advanced Process Control for Light Industry, Ministry of Education,
Jiangnan University, Wuxi, Jiangsu 214122, China



Optical flow detection network of moving target

Overview: Object detection is a hot research direction in the fields of pattern recognition and machine learning. It is used widely in areas such as security, traffic and the Internet. Computer realizes object detection automatically, which can reduce the burden of people in a certain extent. Moving target detection is an important research direction of object detection and plays an important role in target recognition. With the development of deep learning theory, the deep convolutional neural network has made remarkable achievements in the field of object detection by improving the generation strategy of candidate regions and by optimizing the network structure and training methods. Different from the usual definition, moving target detection is turned into focus on the detection of moving target that is needed rather than all of the moving targets. However, conventional convolution neural network cannot extract the dynamic characteristics of object and the accuracy of traditional motion detection methods is low and unable to only detect the required moving target. In this study, deep convolutional neural network is introduced into the optical flow detection of moving target. In this method, a pair of images and optical flow field of target are used as inputs of the network to adaptively study the target optical flow. Furthermore, through optimization of the expanding part of the network and the simplification of the network, and combined with many data augmentation technologies, the optical flow detection network of target object with both accuracy and real time is designed. Visualization technology is also applied to the network, which can directly compare the results with visual image. Experimental results show that the proposed method has better performance in the optical flow detection of moving target. The proposed network in this study retains the advantage of the deep convolutional neural network that only detect the optical flow of required target, which conforms to the practical application of object detection. Additionally, the proposed network achieves better performance in the both accuracy and the real-time of network. S-sp is improved by about 4.9% and 5.3%, respectively, on the AAE and AEE indicators compared to the original network, and SS-sp and CS-sp network are improved by about 5.0% on the AAE indicators compared to the original network. The runtime CS-sp and SS-sp network can be significantly reduced. The runtime of CS-sp network is reduced from about 179 ms to 34 ms and the runtime of the CS-sp network is reduced from 170 ms to 33 ms, which meet the requirements of real-time detection.

Citation: Wang Z L, Huang M, Zhu Q B, *et al.* The optical flow detection method of moving target using deep convolution neural network[J]. *Opto-Electronic Engineering*, 2018, 45(8): 180027

Supported by the Research Innovation Program for College Graduates of Jiangsu Province (SJLX16_0498) and the Prospective Joint Research Foundation of Jiangsu Province of China (BY2016022-32)

* E-mail: huangmzqb@163.com