计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.200500158

基于多重差异特征网络的街景变化检测

詹 瑞 雷印杰 陈训敏 叶书函
四川大学电子信息学院 成都 610065
(zhanrray@163.com)

摘 要 街景变化检测对于自然灾害破坏和城市发展变化的研究起着重要作用。其主要目标是将成对的输入图片中变化的区域标注出来,其实质是二分类的语义分割问题。不同时间拍摄的街景图片可能受到如光线、天气、背景噪声、视角误差等诸多干扰因素的影响,这给传统的变化检测方法带来挑战。针对该问题,提出了一种新的神经网络模型(Multiple Difference Features Network,MDFNet)。该模型首先使用孪生网络提取成对输入图片的不同深度特征,并使用差异模块对相同深度特征计算差异,以此有效获得不同尺度的变化信息;然后通过JPU模块融合多重差异特征,在不损失细节信息的情况下提取其深层语义信息;最后使用金字塔池化模块结合全局和局部信息生成二分类的变化检测图像。在PCD数据集上的GSV和TSUNAMI部分分别采用5折交叉验证法对模型进行实验,实验结果表明,MDFNet获得了0.787和0.862的*F-score*,相比排名第二的DOF-CDNet方法,其值提高了约11.9%和2.9%,同时其能够更精准地分割变化细节。因此,所提模型可以有效应对干扰,对于复杂场景也具备优秀的检测能力。

关键词:图像处理;卷积神经网络;变化检测;语义分割;多重差异特征;特征融合
中图法分类号 TP391

Street Scene Change Detection Based on Multiple Difference Features Network

ZHAN Rui,LEI Yin-jie,CHEN Xun-min and YE Shu-han
College of Electronics and Information Engineering,Sichuan University,Chengdu 610065,China

Abstract Street scene change detection plays an important role in the study of natural disaster damage and urban development. Its main goal is to mark out the changing areas in the pair of input images,which is essentially a semantic segmentation problem of binary classification. There may be many interference factors such as light,weather,background noise,viewpoints error and so on when taking street view pictures at different times,which challenges traditional change detection methods. To solve this problem,a new neural network model (Multiple Difference Features Network,MDFNet) is proposed. First,siamese networks are used to extract the different depth features of pairs of input images,and the difference modules are used to calculate the difference of the same depth features to effectively obtain the change information of different depth. Then,by using JPU module to fuse multiple difference features,the deep semantic information can be extracted without losing detail information. Finally,the pyramid pooling module is used to generate the change detection image of the binary classification combined with the global and local information. MDFNet has obtained 0.787 and 0.862 *F-scores* in the GSV and TSUNAMI part on PCD dataset with 5 fold cross-validation,which are 11.9% and 2.9% higher than the second ranked DOF-CDNet,and can segment the change details more accurately. Therefore,the proposed model can effectively deal with interferences and has an excellent detection ability for complex scenes.

Keywords Image processing,Convolution neural network,Change detection,Semantic segmentation,Multiple difference features,Feature fusion

1 引言

变化检测是指针对不同时间拍摄的图像,通过像素级分类来标注出其中变化的区域,其中,输入检测的目标通常为同视角下的两张图片。

本文基于PCD(Panoramic Change Detection)数据集[1]研究街景变化检测方法。街景变化检测主要针对不同时间在同一条街道上所拍摄的RGB全景图像(如图1(a)、图1(b)所示),检测其变化区域,可用于自然灾害破坏和随后恢复过程中的可视化。图1(c)给出了人为标注的真实变化区域。真实的街景变化检测场景中,常见大量的光线变化、气候变化、视角差异、地面建筑物反光和颜色失真等干扰,这都给

到稿日期:2020-05-29 返修日期:2020-08-03 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。
基金项目:国家自然科学基金(61972435)
This work was supported by the National Natural Science Foundation of China (61972435).
通信作者:雷印杰(yinjie@scu.edu.cn)

准确检测带来极大的挑战。

图 1　街景变化检测数据集

Fig. 1　Street scene change detection dataset

针对变化检测问题，研究者们提出了大量方法[2-3]，这些方法大多基于归纳总结的数学模型处理手工提取的特征，存在的固有问题（如鲁棒性低及提取特征缺乏语义信息）导致其无法很好地完成变化检测。

近年来，卷积神经网络取得了巨大成效，VGGnet[4] 由于其强大的图像特征提取能力，被广泛应用于图像处理中。以卷积层代替其网络中原有的全连接层，使得网络输出为矩阵或者图像，可以解决图像像素级语义分割的问题。

针对街景变化检测问题，本文首先将成对的图片分别输入到两条支路的孪生网络中来提取特征；然后使用差异模块对不同深度的卷积网络输出特征图进行差异检测，得到三重差异特征；最后将三重差异特征送入 JPU（Joint Pyramid Upsampling）模块进行特征处理，并使用金字塔池化模块融合全局信息生成分割图。

2　相关工作

目前已有大量的深度学习变化检测方法被提出，如文献[1，5-7]采用单级特征融合，将提取的最深层特征进行处理得到变化预测图。Sakurada 等[1] 从成对的卷积网络中计算深层特征的差距，同时采用超像素分割来估计边界，以去除几何背景。Alcantarilla 等[5] 采用单支路的收缩网络和扩张网络对不同时期的变化图片进行密集定位与匹配。Guo 等[6] 采用孪生的 FCN（Fully Convolutional Networks）结构结合均方根值来进行距离度量，并提出阈值对比损失函数来消除常见的噪声影响。DOF-CDNet（Dense Optical Flow Based Change Detection Network）[7] 针对视角差异问题，采用基于密集光流的卷积网络以提高鲁棒性。

相比之下，多级特征融合先提取不同深度的特征信息进行处理，再进行融合以获得分割图，有助于获取不同尺寸的目标信息。Daudt 等[8] 使用对称的编码解码结构比较了单支路与孪生编码网络的效果差异，同时比较了解码网络中将多级特征直接相接与相减后再相接的效果差异。Chen 等[9] 在孪生特征编码结构中加入了多尺度特征卷积单元以提取高维特征，在解码结构中利用一条支路的高维特征与相减的多级特征融合，由此生成变化预测图。Jiang 等[10] 通过全局共同注意力机制获得输入特征对之间的联系，采用不同的注意力机制聚合浅层特征和共同关注特征，从而获得更丰富的目标信息。浅层特征具备更多小尺寸的目标信息，深层特征具备更多大尺寸的目标信息，多级特征融合能够更好地检测不同尺寸的变化目标。MDFNet 采用多级特征融合来处理 3 层不同深度的特征信息。

与街景变化检测相比，遥感图像变化检测几乎不存在视角差异。具体方法有：Wang 等[11] 在孪生网络中加入混合卷积特征提取模块以有效增加网络的深度和宽度，通过不确定性分析，结合多分辨率分割图以及变化决策模块的输出来获得变化检测图；Lv 等[12] 针对图像中的多尺度目标，依次抽取关键点以表述其特征，利用关键点向量距离测量目标局部区域的变化幅度，通过 Otsu 二进制阈值方法获得变化检测图。

变化检测实质上是二分类的语义分割问题，其与普通的语义分割的区别在其输入是成对图片。近年来，许多新的方法运用于语义分割。例如，在注意力机制[13] 中，深层网络生成注意力图而浅层网络利用注意力图去除虚景。FastFCN（Fast Fully Convolutional Networks）[14] 使用 JPU 模块，在不降低分辨率且不增加额外开销的情况下扩大感受野。PSPNet（Pyramid Scene Parsing Network）[15] 采用金字塔池化模块有效融合了低层空间位置信息和高层语义信息。

3　本文方法

有效获取及利用图像的差异是进行准确变化检测的保障。MDFNet 将两条支路共享权重的 VGG16[4] 的前 10 个卷积层作为孪生网络，如图 2 所示。共享权重的孪生网络比单支路的特征提取结构更能凸显成对输入的特征差异。基于浅层的特征包含更多的小尺寸目标及空间位置信息，基于深层的特征包含更多的大尺寸目标、背景及语义信息。因此，MDFNet 首先抽取第 4，7，10 个卷积层输出的特征图并通过差异模块获得多重差异特征；然后采用 JPU 模块融合及提取多重差异特征的语义信息；最后通过金字塔池化模块融合全局和局部信息，经过上采样后生成分割预测图。

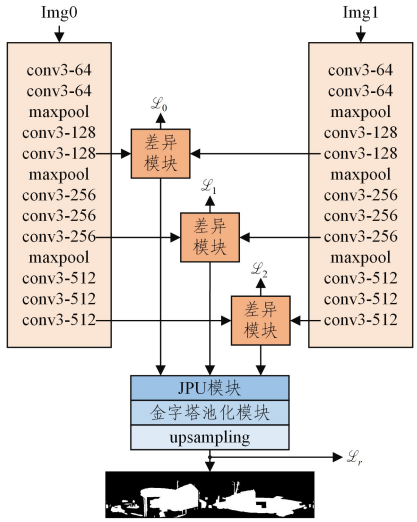

图 2　MDFNet 的结构

Fig. 2　Structure of MDFNet

3.1　差异模块

MDFNet 使用差异模块对输入图片不同深度的特征信息

进行差异特征提取。如图 3 所示，f_n^i 表示成对输入特征；i 取 0,1 表示特征图来自两个不同的支路；n 取 0,1,2,表示第 n 个差异模块。在差异模块中首先对输入的 f_n^i 计算余弦相似度 C_n、欧氏距离 E_n 和绝对差值 D_n。然后将 3 种差异按照通道进行相接获得 M_n。最后以 $channel_n$ 表示输入 f_n^i 的通道数，若 C_n 和 E_n 的通道数都为 1，则 M_n 的通道数为 $channel_n+2$。输出的差异特征 $\underset{c,x,y}{F_n}=\underset{c,x,y}{M_n}*\underset{0,x,y}{m_n}$（其中，$c\in[0,channel_n+2)$，$x\in[0,h_n)$，$y\in[0,w_n)$）表示 M_n 的每一条通道与注意力模块输出注意力图 m_n 的 0 通道在 (x,y) 位置做像素乘积。h_n 和 w_n 表示输入 f_n^i 的尺寸。$\mathcal{L}_n$ 为 m_n 与真实变化图计算的交叉熵。

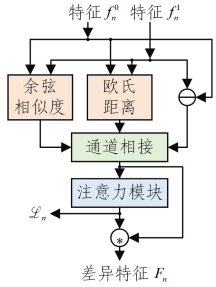

图 3 差异模块

Fig. 3 Difference module

在求 f_n^i 的余弦相似度与欧氏距离时，将每一个像素的全部通道看作一个向量。例如，若将坐标为 (x,y) 的全部通道结合，看作一个 $channel_n$ 维向量 $\underset{x,y}{\boldsymbol{A}_n^i}$，用 $\underset{c,x,y}{a_n^i}$ 表示向量 $\underset{x,y}{\boldsymbol{A}_n^i}$ 的第 c 个元素，即 f_n^i 在 (c,x,y) 的值，则余弦相似度 C_n、欧氏距离 E_n 和绝对差值 D_n 可分别表示为：

$$\underset{x,y}{C_n}(f_n^0,f_n^1)=\frac{\underset{x,y}{\boldsymbol{A}_n^0}\cdot\underset{x,y}{\boldsymbol{A}_n^1}}{\underset{x,y}{\boldsymbol{A}_n^0}\times\underset{x,y}{\boldsymbol{A}_n^1}}=\frac{\sum_{c=0}^{channel_n-1}\left(\underset{c,x,y}{a_n^0}\times\underset{c,x,y}{a_n^1}\right)}{\sqrt{\sum_{c=0}^{channel_n-1}\left(\underset{c,x,y}{a_n^0}\right)^2}\times\sqrt{\sum_{c=0}^{channel_n-1}\left(\underset{c,x,y}{a_n^1}\right)^2}} \tag{1}$$

$$\underset{x,y}{E_n}(f_n^0,f_n^1)=\left\|\underset{x,y}{\boldsymbol{A}_n^0}-\underset{x,y}{\boldsymbol{A}_n^1}\right\|=\sqrt{\sum_{c=0}^{channel_n-1}\left(\underset{c,x,y}{a_n^0}-\underset{c,x,y}{a_n^1}\right)^2} \tag{2}$$

$$\underset{c,x,y}{D_n}(f_n^0,f_n^1)=\left|f_n^0-f_n^1\right|=\left|\underset{c,x,y}{a_n^0}-\underset{c,x,y}{a_n^1}\right| \tag{3}$$

其中，$x\in[0,h_n)$，$y\in[0,w_n)$，$c\in[0,channel_n]$。由此可知，C_n，E_n，D_n 的尺寸与 f_n^i 相同。D_n 独立求取每个像素的绝对差值，其结果只与相同位置的两个像素值有关，因此其保留了各个通道的细节差异，也继承了 f_n^i 原有的结构信息，但是无法获得高维特征图各通道之间的复杂联系。C_n 和 E_n 中的每个数值都由两个向量求得，与各通道相关，考虑了各通道之间的联系，代表了 f_n^i 在同一位置的差异。

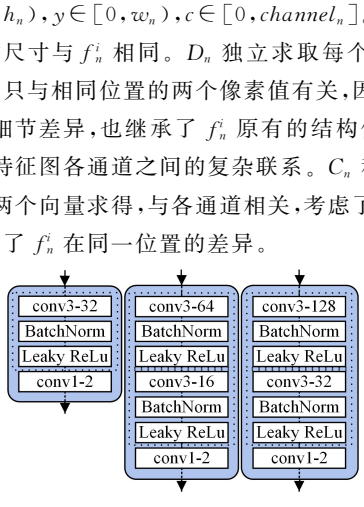

图 4 注意力模块

Fig. 4 Attention module

MDFNet 的注意力模块通过层次化的卷积操作融合 C_n，E_n，D_n 生成注意力图。由于从孪生网络提取特征的深度不同，其输入 f_n^i 的通道数随深度增加以 2 倍数增长。为了使层次化的融合操作更为合理，其在不同的差异模块中采用不同深度的注意力模块。如图 4 所示，从左到右的网络对应由浅层到深层的差异模块中使用的注意力模块，模块分别通过 2 个和 3 个卷积层进行层次化融合生成注意力图。添加注意力图和 $\mathcal{L}_n$ 有助于滤除虚景，使网络更多地关注可能的变化区域。

3.2 JPU 模块

浅层的差异特征虽然保留了更多的位置信息，但容易受环境及视角误差的干扰。利用深层的差异特征虽然可以更为准确地判断区域变化，但容易忽视细节。FastFCN[14] 首次提出了 JPU 模块，该模块通过具有一定规律的网络结构来替代传统 DilatedFCN(Dilated Fully Convolutional Networks)中的空洞卷积层，保留了更多的位置和细节信息，扩大了感受野。为利用多重差异特征，本文采用 JPU 模块将其进行融合，在不损失细节信息的情况下提取深层语义信息。

FastFCN[14] 分解空洞卷积过程是将其映射为具有一定规律的卷积及联合上采样结构(Joint Upsampling)，因此，其将空洞卷积过程拆解为：

$$\begin{aligned}y_d&=x\rightarrow C_r\rightarrow S\rightarrow C_r^n\rightarrow M\\&=y_m\rightarrow S\rightarrow C_r^n\rightarrow M\\&=\{y_m^0,y_m^1\}\rightarrow C_r^n\rightarrow M\end{aligned} \tag{4}$$

其中，y_d 为空洞卷积层输出特征图，x 为输入特征图，C_r 和 C_r^n 分别代表单独的和 n 个连续的普通卷积层。对于单个空洞卷积层，首先将特征图拆分为两部分分别进行普通的卷积操作，然后将其融合恢复为原来的大小。其中，S 和 M 分别代表拆分和融合操作，两者可相互抵消。

FastFCN 采用近似方法替代连续的空洞卷积的过程中，将空洞卷积输出特征图 y_d 近似表示为：

$$\begin{aligned}y_s&=x\rightarrow C_s\rightarrow C_r^n\\&=x\rightarrow C_r\rightarrow R\rightarrow C_r^n\\&=y_m\rightarrow R\rightarrow C_r^n\\&=y_m^0\rightarrow C_r^n\end{aligned} \tag{5}$$

$$y_d'=\{y_m^0,y_m^1\}\rightarrow\hat{h}\rightarrow M \tag{6}$$

上述将空洞卷积过程的拆分工作用步长不到 1 的普通卷积层近似替代，即 C_s。其也可以表示为经过普通卷积层后去掉的一些像素，即 R。其中，$\hat{h}=\arg\min\|y_s-h(y_m^0)\|$，$y_m=x\rightarrow C_r$。借鉴了联合上采样的思想，用 $\hat{h}$ 表示从 y_m^0 到 y_s 的映射，以取代导致大量内存和时间开销的连续空洞卷积层。

图 5 给出了 MDFNet 采用的 JPU 模块，其中，输入的 F_0，F_1 和 F_2 对应 3 个差异模块所生成的差异特征。JPU 模块首先将输入的三重差异特征分别经过一个卷积层和上采样层，使其通道及尺寸统一，并得到 y_m。然后采用 4 个不同扩张率(dilation)的分离层进一步提取特征。其中，扩张率为 1 的分离卷积层用于获取 y_m^0 与 y_m 其余部分的相互关系；扩张率为 2,4,8 的分离卷积层用于学习 y_m^0 到 y_s 的映射，即 $\hat{h}$。

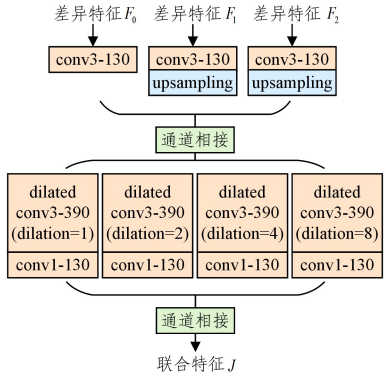

图 5　JPU 模块

Fig. 5　JPU module

3.3　金字塔池化模块

大多数语义分割模型由于感受野有限，无法很好地利用全局上下文信息，导致出现分割误差。例如，文献[15]提到，由于快艇和车辆外观相似，常见的语义分割模型会将快艇识别成车辆，若结合河流、船坞等环境上下文信息便可避免错误分类。对于复杂场景，其上下文关系虽然难以知晓但相当重要，如果能以之引导分割任务，就能有效地改善上述问题。

He 等[16]对深层特征进行空间金字塔池化处理，得到尺寸分别为 1×1，2×2，4×4 的特征图来表征全局和局部信息，以增强网络对复杂场景全局语义信息的理解。文献[15]提出金字塔池化模块，对最后一层特征图进行不同规格的平均池化处理，将处理结果运用到分割图生成过程中。

MDFNet 采用空间金字塔池化方法获取全局和局部信息，具体过程如图 6 所示。首先对联合特征 J 并行使用 4 种规格的自适应平均池化层；然后通过一个卷积层融合得到全局和局部信息，将其上采样后运用于层次化的卷积操作中；最后融合得到变化预测图。

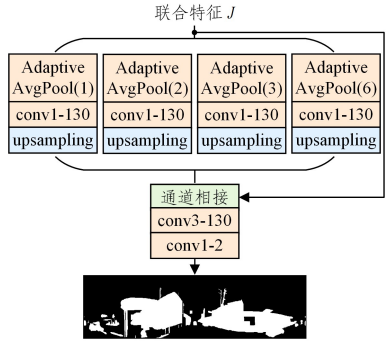

图 6　金字塔池化模块

Fig. 6　Pyramid pooling module

3.4　损失函数

变化检测实质上是二分类的语义分割问题，将图像中的每一个像素分为变与不变两类（即 0 或 1）。MDFNet 采用交叉熵的方法计算损失函数，即：

$$\mathscr{L}=-\frac{1}{xy}\sum_{x,y}\sum_{c=0}^{1}g_{c,x,y}\times\log\ p_{c,x,y} \tag{7}$$

其中，g 表示真实变化图，其在 (c,x,y) 处的值为 0 或 1；P 表示变化预测图经过 softmax 转化为概率形式的输出，$p_{c,x,y}\in[0,1]$。损失函数即为全部像素点交叉熵的平均值。softmax 过程可以表示为：

$$p_{c,x,y}=\frac{\exp(r_{c,x,y})}{\sum_{c=0}^{1}\exp(r_{c,x,y})} \tag{8}$$

其中，r 为网络输出的二通道预测图，不同通道表示不同类别在图中的区域。pytorch 中，交叉熵损失函数为 log softmax 与 nll_loss 的结合，即：

$$\begin{aligned}\mathscr{L}(r,T)&=-\frac{1}{xy}\sum_{x,y}\mathscr{L}_{\text{null\_loss}}(r_{x,y},T_{x,y})\\&=-\frac{1}{xy}\sum_{x,y}\left(\log\frac{\exp(r_{t,x,y})}{\sum_{c=0}^{1}\exp(r_{c,x,y})}\right)\end{aligned} \tag{9}$$

其中，$T_{x,y}$ 表示在 (x,y) 位置的真实类别，其值为 t。

MDFNet 训练过程中的损失函数由两部分构成，即：

$$\mathscr{L}_{\text{final}}=\mathscr{L}_r+\omega\sum_{n=0}^{2}\mathscr{L}_n=\mathscr{L}(r,T)+\omega\sum_{n=0}^{2}\mathscr{L}(m_n,T_n) \tag{10}$$

其中，$\mathscr{L}_r$ 与 $\mathscr{L}_n$ 分别为变化预测图与注意力图和采样的真实变化图所求的交叉熵。ω 为权重常数。m_n 为差异模块中的注意力图。T_n 为下采样至与 m_n 相同尺寸的真实变化图。

4　实验与结果分析

4.1　场景变化检测数据集

本文在 PCD 数据集[1]上测试 MDFNet 的性能。PCD 数据集制作的初衷是检测城市场景的变化，其可被用于自然灾害和随后重建恢复过程的可视化，也可被用于更新城市的三维模型。该数据集中手工标注的变化区域由具有较大差异的两部分组成，分别是 TSUNAMI 和 GSV。这两个部分各自包含 100 对 RGB 全景图像及真实变化图，每对图像是在相同场景下的不同时间拍摄的，尺寸为 $224\times1\,024$。TSUNAMI 数据集记录了海啸前后城市街景的变化，记录了灾难对房屋、街道等产生的较大影响。GSV 数据集采自谷歌街景视图，更多地记录了城市的正常发展变化。

4.2　评价方法

本文采用变化检测中被广泛使用的 F-Score 作为网络的评价指标。F-Score 由精确率（Precision）和召回率（Recall）组成，适用于二分类的模型，取值为[0,1]。

在计算 *F-Score* 的过程中，引入 4 个变量，即 TP，FP，TN，FN。其中，TP（True Positive）与 FP（False Positive）分别表示预测为变化的像素点中正确预测的与错误预测的像素点数量；TN（True Negative）与 FN（False Negative）分别表示预测为不变的像素点中正确预测的与错误预测的像素点数量。

$$TP=\sum_{x,y}(R_{0,x,y}==0)\bigcap(g_{x,y}==0) \tag{11}$$

$$FP=\sum_{x,y}(R_{0,x,y}==0)\bigcap(g_{x,y}==1) \tag{12}$$

$$TN=\sum_{x,y}(R_{0,x,y}==1)\bigcap(g_{x,y}==1) \tag{13}$$

$$FN=\sum_{x,y}(R_{0,x,y}==1)\bigcap(g_{x,y}==0) \tag{14}$$

其中，R 为二值化的二通道预测图；g 为真实变化图，若其值为 0，则表示 (x,y) 处为变化的像素点，反之为不变的像素点；$\bigcap$ 表示交集，若 $\bigcap$ 两边等式都成立，则返回值为 1，否则为 0。精确率和召回率分别表示预测为变化的像素点和实际变化的像素点中正确预测的比例，即：

$$Precision=\frac{TP}{TP+FP} \tag{15}$$

$$Recall=\frac{TP}{TP+FN} \tag{16}$$

F-Score 可以表示为：

$$\frac{1}{F\text{-}Score}=\frac{1}{(1+\beta^2)Precision}+\frac{\beta^2}{(1+\beta^2)Recall} \tag{17}$$

其中，β 是用来平衡精确率和召回率对 *F-Score* 影响的权重，其值通常为 1。因此，*F-Score* 可被简化为：

$$F\text{-}Score=\frac{(1+\beta^2)TP}{(1+\beta^2)TP+FP+\beta^2FN} \tag{18}$$

由上述公式可知，若错误预测的 FP 与 FN 相对较少，则 *F-Score* 相对较大，模型效果越好。若 FP 与 FN 为 0，则 *F-Score* 值最大，且为 1。

4.3 实验细节

为了训练和测试 MDFNet，本文使用了一个 NVIDIA GTX 1080Ti GPU 以及一个 E5-1650 CPU 在 Python3.6 平台上进行实验。具体实验环境包括 PyTorch0.3.1，CUDA 8.0，CUDNN7.0.5。

训练开始前首先对数据集进行预处理。参考文献[1-2，5-7]，本文采用 5 折交叉验证的方法，分别对 PCD 数据集中的 TSUNAMI 和 GSV 部分进行训练和测试。随机将数据集均分为互不相交的 5 份，依次将其中 1 份作为测试集，其余的作为训练集。最终得到的网络评价结果为每次网络训练及测试后得到的 *F-Score* 平均值。

网络训练过程中对训练集进行数据增广，即首先使用尺寸为 224×224 的滑动窗口以 56 的步长对训练集中的每张图片进行切割，然后将每张子图片进行旋转和水平翻转，从而得到 120 张图片。为了保持测试集的完整性，本文使用 20 对没有经过处理的图片进行测试。

MDFNet 的孪生网络部分使用了在 ImageNet 数据集上预训练的 VGG16[4] 的前 10 个卷积层，其他层采用了随机初始化的方法。训练过程中，采用了 SGD 优化器[17]，设学习率为 2×10^{-2}，动量为 0.95，权值衰减为 1.25×10^{-4}。

4.4 实验结果分析

为了证实本文方法的有效性，增加了两组对照实验。一组对照实验研究注意力模块在使网络关注可能的变化区域时发挥的积极作用，相比 MDFNet，MDFNet-的损失函数不再包括差异模块中的注意力图与真实图的交叉熵。另一组对照实验研究差异模块为提高变化检测精度所做的贡献，MDFNet-在 MDFNet 的基础上去掉了差异模块，对相同层的特征计算绝对差值作为 JPU 模块的输入。所有变化检测模型的 *F-Score* 值如表 1 所列。

表 1 GSV 与 TSUNAMI 数据集上的 *F-Score*

Table 1 *F-Score* on GSV and TSUNAMI datasets

	GSV	TSUNAMI
DenseSIFT[2]	0.528	0.649
DeconvNet[5]	0.614	0.774
CNN-feat[1]	0.639	0.723
CosimNet[6]	0.692	0.806
CDNet[7]	0.695	0.838
DOF-CDNet[7]	0.703	0.838
MDFNet--(Ours)	0.778	0.854
MDFNet-(Ours)	0.782	0.856
MDFNet(Ours)	**0.787**	**0.862**

由表 1 可知，MDFNet 的表现优于其他模型。在 GSV 与 TSUNAMI 上，MDFNet 的 *F-Score* 相比排行第二的 DOF-CDNet 提高了约 11.9% 与 2.9%。MDFNet 在 TSUNAMI 数据集上相比 GSV 数据集上提升较小的原因主要有两点：1)GSV 数据集的变化标注更为精准；2)由于景深不同，GSV 数据集相机视角的差异相比 TSUNAMI 数据集的更为明显。

采用多级特征提取孪生网络有利于图像的对比，采用 JPU 模块有利于融合多尺度差异信息，金字塔池化模块能够结合全局和局部信息生成变化预测图，因此 MDFNet--相比其他模型具有更强的差异信息提取和处理能力。MDFNet-相比 MDFNet--，采用了专门提取差异特征的模块，考虑了不同通道之间的联系，使提取的多重差异特征更好地表述了变化。MDFNet 相比 MDFNet-，将注意力图与下采样的真实图求取交叉熵，使得网络更加关注可能的变化区域，进一步提高了模型预测的能力。

为了对比不同的骨干网络对变化检测结果的影响，表 2 列出了 MDFNet 分别以 VGG16[4] 和 ResNet[18] 作为孪生网络的实验结果。以 VGG16[4] 作为孪生网络提取不同深度的特征，其效果比使用 ResNet[18] 作为孪生网络时更好。这是因为 ResNet[18] 主要用于解决深层网络梯度爆炸与梯度消失的问题，在样本量与复杂度较大的多分类问题中表现优秀，但本实验使用的数据集较小，无法充分训练 ResNet[18]。表 2 最后一列为将 GSV 与 TSUNAMI 数据集合并后进行实验的结果。该合并数据集包含了海啸对城市造成的灾害及城市的正常发展变化，与现实复杂环境下的变化检测需求更为契合，其测试结果与单独对 GSV 和 TSUNAMI 数据集进行实验得到的结果基本一致。

表 2 不同结构的 MDFNet 所得到的 *F-Score* 对比

Table 2 Comparison of *F-Score* from MDFNet with different structures

	GSV	TSUNAMI	GSV&TSUNAMI
MDFNet(VGG16)	**0.787**	**0.862**	**0.813**
MDFNet-(VGG16)	0.782	0.856	0.809
MDFNet--(VGG16)	0.778	0.854	0.804
MDFNet(ResNet18)	0.769	0.849	0.790
MDFNet(ResNet34)	0.769	0.845	0.798
MDFNet(ResNet50)	0.756	0.843	0.795
MDFNet(ResNet101)	0.756	0.826	0.758

图 7 展示了 MDFNet 在 GSV 与 TSUNAMI 数据集上的预测结果。图 7(a)和图 7(b)为测试原图；图 7(c)、图 7(d)和图 7(e)为可视化的注意力掩图；图 7(f)为变化预测图；图 7(g)为真实变化图。由图 7 可知，MDFNet 针对类型、尺寸大不相同的变化目标依然能够进行准确的预测。模型关注不同深度的差异特征所存在的较大差别，如图 7(c)－图 7(e)所示。该差别能反映出不同深度的差异特征对生成变化预测图的贡献。浅层的差异特征更多地关注细节和位置信息，有利于精确预测变化目标的轮廓；中层的差异特征关注可能的变化区域，滤除不变的区域，对局部是否发生变化起指引作用；深层的差异特征关注可靠的变化区域，能保障大尺寸目标的准确预测，有助于削弱相机视角差异的影响。图 7(c)－图 7(e)分别对应可视化的浅、中、深层的注意力图。从图 7 中 GSV 上的实验结果可以明显看出，浅层的注意力图细节信息更为充分，深层的注意力图变化区域更为可靠。图 7 中 TSUNAMI 上的实验结果更为直观地展示出中层和深层的注意力图对于滤除不变区域的重要作用。

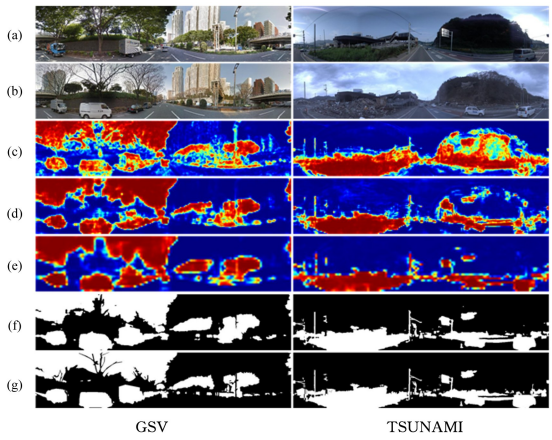

图 7　可视化结果

Fig. 7　Visualization results

结束语　街景图像变化检测任务的关键是差异特征的提取与利用。本文提出了一个新的深度神经网络模型 MDFNet。为了提取差异特征，MDFNet 采用了孪生网络作为特征提取结构，并使用差异模块对不同深度的特征进行处理以获得多重差异特征。针对多重差异特征，本文采用 JPU 模块融合及提取多重差异特征的语义信息，由金字塔池化模块结合全局和局部信息生成变化预测图。为了验证本文方法的有效性，我们在 PCD 数据集上进行了实验。GSV 和 TSUNAMI 数据集上的实验结果表明，MDFNet 针对街景图像变化检测具备优秀的性能。

本文方法能够精准地预测变化细节，但针对超大尺寸变化目标的完整分割尚有一定的提升空间，未来可以考虑采用深层感受野更大的孪生网络，以增强网络对大尺寸目标完整语义信息的提取能力。

参 考 文 献

[1] SAKURADA K，OKATANI T. Change detection from a street image pair using cnn features and superpixel segmentation[C]// Proceedings of the British Machive Vision Conference. Swansea，UK：BMVA Press，2015，61：1-12.

[2] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision，2004，60(2)：91-110.

[3] TANEJA A，BALLAN L，POLLEFEYS M. City-scale change detection in cadastral 3d models using images[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE，2013：113-120.

[4] SIMONYAN K，ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv：1409. 1556，2014.

[5] ALACANTARILLA P F，STENT S，ROS G，et al. Street-view change detection with deconvolutional networks[J]. Autonomous Robots，2018，42(7)：1301-1322.

[6] GUO E，FU X，ZHU J，et al. Learning to measure change：fully convolutional Siamese metric networks for scene change detection[J]. arXiv：1810. 09111，2018.

[7] SAKURADA K，WANG W，KAWAGUCHI N，et al. Dense optical flow based change detection network robust to difference of camera viewpoints[J]. arXiv：1712. 02941，2017.

[8] DAUDT R C，LE SAUX B，BOULCH A. Fully convolutional siamese networks for change detection[C]// 2018 25th IEEE International Conference on Image Processing (ICIP). IEEE，2018：4063-4067.

[9] CHEN H，WU C，DU B，et al. Deep siamese multi-scale convolutional network for change detection in multi-temporal vhr images[C]// 2019 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp). IEEE，2019：1-4.

[10] JIANG H，HU X，LI K，et al. Pga-siamnet：pyramid feature-based attention-guided siamese network for remote sensing orthoimagery building change detection[J]. Remote Sensing，2020，12(3)：484.

[11] WANG M，TAN K，JIA X，et al. A deep siamese network with hybrid convolutional feature extraction module for change detection based on multi-sensor remote sensing images[J]. Remote Sensing，2020，12(2)：205.

[12] LV Z，LIU T，BENEDIKTSSON J A，et al. Object-oriented key point vector distance for binary land cover change detection using vhr remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing，2020，58(9)：6524-6533.

[13] LIN H，SHI Z，ZOU Z. Fully convolutional network with task partitioning for inshore ship detection in optical remote sensing images[J]. IEEE Geoscience and Remote Sensing Letters，2017，14(10)：1665-1669.

[14] WU H，ZHANG J，HUANG K，et al. Fastfcn：rethinking dilated convolution in the backbone for semantic segmentation[J]. arXiv：1903. 11816，2019.

[15] ZHAO H，SHI J，QI X，et al. Pyramid scene parsing network[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017：2881-2890.

[16] HE K，ZHANG X，REN S，et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2015，37(9)：1904-1916.

[17] KRIZHEVSKY A，SUTSKEVER I，HINTON G E. Imagenet classification with deep convolutional neural networks[C]// Advances in Neural Information Processing Systems. 2012：1097-1105.

[18] HE K，ZHANG X，REN S，et al. Deep residual learning for image recognition[C]// Computer Vision and Pattern Recognition. 2016：770-778.

ZHAN Rui，born in 1996，postgraduate. His main research interests include deep learning and computer vision.

LEI Yin-jie，born in 1983，Ph.D，associate professor. His main research interests include machine learning，multimedia communication，pattern recognition and image processing.