基于双流卷积神经网络的 RGB-D 图像联合检测

刘帆，刘鹏远，张峻宁，徐彬彬

军械工程学院，河北 石家庄 050003

摘要 当前卷积神经网络结构未能充分考虑 RGB 图像和深度图像的独立性和相关性，针对其联合检测效率不高的问题，提出了一种新的双流卷积网络。将 RGB 图像和深度图像分别输入到两个卷积网络中，两个卷积网络结构相同且权值共享，经过数次卷积提取各自独立的特征后，在卷积层根据最优权值对两个卷积网络进行融合；继续使用卷积核提取融合后的特征，最后通过全连接层得到输出。相比于以往卷积网络对 RGB-D 图像采用的早期融合和后期融合方法，在检测时间相近的情况下，双流卷积网络检测的准确率和成功率分别提高了 4.1%和 3.5%。

关键词 机器视觉；RGB-D；卷积神经网络；多模态信息；联合检测；深度学习

中图分类号 TP391 **文献标识码** A **doi**：10.3788/LOP55.021503

Joint Detection of RGB-D Images Based on Double Flow Convolutional Neural Network

Liu fan, Liu Pengyuan, Zhang Junning, Xu Binbin

Mechanical Engineering College, Shijiazhuang, Hebei 050003, *China*

Abstract The convolutional neural network structure fails to consider the independence and correlation between RGB images and depth images fully, so its detection is not high. A new double flow convolution network is proposed for the joint detection of RGB-D images. The RGB image and depth image are inputted to the two convolutional networks and the two networks have the same structure and weight sharing. After several convolutions, the independent features are extracted. According to the optimal weights in the convolution layer, the two convolutional networks are fused. The fused features are extracted continuously using convolution kernels, and the output is obtained by full connection layer finally. When the detection time is similar, the detection accuracy and the success rate are increased by 4.1% and 3.5% respectively, compared with the previous early and late fusion methods.

Key words machine vision; RGB-D; convolutional neural network; multimodal information; joint detection; depth learning

OCIS codes 150.0155; 150.6910

1 引 言

红绿黄彩色图像和深度(RGB-D)图像是当前计算机视觉领域新的研究热点[1]，主要由深度相机获得，包括红绿黄彩色(RGB)图像和对应的深度图像。RGB 图像包含所拍摄物体的表面颜色和纹理等信息，而深度图像包含物体的空间形状等信息。两种图像对彼此都是一个很好的补充。因此，有效利用 RGB 信息和深度信息能显著提高场景中物体识别检测的准确率。当前，RGB-D 图像的联合检测方法有基于人工设计特征的提取表达法和自动特征学习法[2]两种。

基于人工设计特征的提取表达法是指根据先验知识设计特定的特征，如尺度不变特征变换(SIFT)、加速稳健特性(SURF)、方向梯度直方图(HOG)等。

收稿日期：2017-07-07；**收到修改稿日期**：2017-08-16

基金项目：国家自然科学基金(51205405，51305454)

作者简介：刘帆(1995—)，男，硕士研究生，主要从事计算机视觉方面的研究。E-mail：2434344286@qq.com

导师简介：刘鹏远(1975—)，男，博士，副教授，硕士生导师，主要从事增强现实维修方面的研究。E-mail：lpy_jx@sina.com(通信联系人)

首先采用词袋模型、空间金字塔匹配和稀疏编码等方法进行特征表达，然后对彩色图像和深度图像特征描述进行融合，最后使用分类算法，如支持向量机(SVM)、贝叶斯分类器等，进行图像识别检测。该方法的主要缺点是需要很强的先验知识才能设计好区分性强的特征与融合规则，不具有普适性。

自动特征学习法可以克服人工设计特征通用性不强的缺点，它采用特定的神经网络结构，通过无监督或有监督的训练，自动地学习源图像中的低层或高层特征，并采用一定的方式进行融合。基于卷积神经网络(CNN)的深度学习模型是当前这一领域的优秀代表。该学习模型使用卷积核将相邻像素之间的共同特性较好地提取表达出来，并在卷积核之间进行权值共享，大幅减少了神经网络训练过程中的参数数量，具有识别精度高、训练参数少的优势。Couprie 等[3]将四通道的 RGB-D 图像同时输入到卷积网络中，得到了高于单一 RGB 或深度图像的识别精度。Gupta 等[4]采用多方向梯度算子提取 RGB-D 图像的边缘特征，将提取到的边缘特征融合后送入到卷积神经网络中进行特征学习，在特定物体的识别方面取得了比原始 RGB-D 图像更好的分类效果。Eitel 等[5]将不同模态的数据输入到独立的卷积神经网络中进行训练，经过五层卷积和下采样操作，再经过两个全连接层，最后将 RGB 信息和深度信息融合输入到输出层，得到了比四通道 RGB-D 信息同时输入更优的识别效果。

2 卷积神经网络

2.1 卷积神经网络模型结构

卷积神经网络是图像识别检测领域优秀的深度学习模型[6-7]，与传统的后向传播(BP)神经网络相比增加了可以进行特征提取的卷积层和保证位移不变的池化层。卷积层能将相邻像素之间的共同特性较好地提取出来，池化层用来对图像信息进行降采样。为了降低运算的复杂度，卷积神经网络每层内的神经元权值实行共享，使用不同的卷积核进行卷积可提取前一层特征图的不同特征。这些代表不同特征的特征图共同作为下一层网络的输入数据，其结构如图 1 所示。图 1 中的 C 为卷积层，P 为池化层，FC 为全连接层。

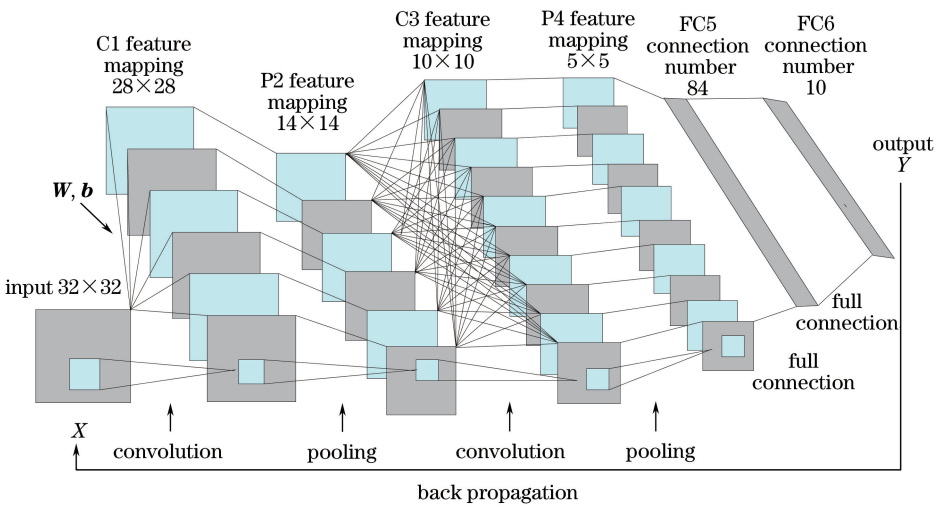

图 1 卷积神经网络结构

Fig. 1 Convolution neural network structure

如图 1 所示，在卷积神经网络中将输入层与卷积层相连接，从输入层获得输入图像 X，使用不同的卷积核可以得到不同的特征图。如在卷积层 C1 中使用 6 种卷积核得到了对应的 6 个特征图，之后对每个特征图进行池化操作得到池化层 P2，P2 中的特征图与 C1 中的特征图是一一对应的。之后对 P2 层再次进行卷积操作，得到对应的 10 个特征图，这两层之间的连接方式类似于全连接，每个卷积核对 P2 层中的 6 个特征图进行卷积操作，相加后得到 C3 中对应的 1 个特征图；之后再对 C3 进行池化操作得到 P4，最后经过数次全连接后得到输出 Y。

2.2 卷积神经网络前向传播过程

卷积神经网络的训练过程可分为前向传播和后向传播两个阶段。在前向传播阶段，信息从输入层开始向前逐层传播，经过各个卷积层和全连接层直至输出层，神经元的基本模型如图 2 如示。在图 2 中，x 为神经元的输入，卷积网络内神经元权值共享，$\boldsymbol{w}$ 为共同的权值，$\boldsymbol{b}$ 为偏置，$f(\cdot)$为激活函数，Y 为输出。

设卷积网络的第 l 层第 j 个神经元的输出为

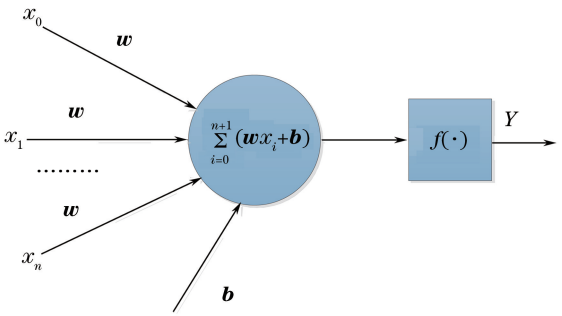

图 2 神经元基本模型

Fig. 2 Basic model of neuron

a_j^l，当第 l 层为卷积层时，该层中第 j 个神经元的输出计算公式为

$$a_j^l = f_C\left[\boldsymbol{W}^l\left(\sum_{i\in M_j^l} a_i^{l-1} * k_{ij}^l\right)+\boldsymbol{b}^l\right], \tag{1}$$

式中 a_i^{l-1} 为上一层的输出；$f_C(\cdot)$ 为卷积层的激活函数，激活函数一般可选取 sigmoid 或 ReLU 函数；k 为卷积核；M 为选择的输入特征图的集合，当 $l=1$ 时，输入的特征图的集合 M 为原始输出图像 X；$\boldsymbol{W}^l$ 为卷积网络第 l 层的权重；$\boldsymbol{b}^l$ 为网络第 l 层的偏置；$*$ 代表卷积运算。图像卷积过程如图 3 所示。

图 3 选用 2 * 2 的卷积核对上一层输入的特征图 a_i^{l-1} 进行卷积操作，得到的输出为 $a_i^{l-1} * k^l$。设 a 为输入的特征图，k 为卷积核，s、h 为卷积核的尺寸，卷积计算 $g=a*k$ 的具体表达式为

$$g(i,j)=\sum_{s,h} a(i-s,j-h)k(s,h)。 \tag{2}$$

当第 l 层为池化层时，该层中第 j 个神经元的输出计算公式为

$$a_j^l = f_P[\boldsymbol{W}^l \cdot P(a_i^{l-1})+\boldsymbol{b}^1], \tag{3}$$

式中 $f_P(\cdot)$ 为池化层的激活函数，$P(\cdot)$ 为池化函数。池化不改变上层神经元的个数，池化层中的每个神经元直接对应上一层的神经元。一般可选取最大池化或平均值池化法进行池化。

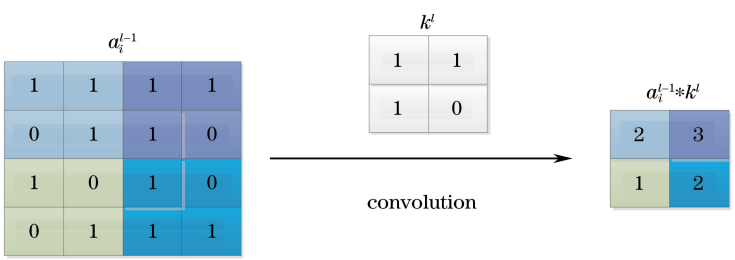

图 3 图像卷积过程

Fig. 3 Image convolution process

图 4 所示的是平均池化操作，对输入的特征图 a_i^{l-1}，将每个 2×2 的方格压缩成一个像素，其值为该方格内所有像素的平均值，得到输出 $P(a_i^{l-1})$；池化后特征图的维数成倍降低，可以有效降低卷积网络的计算复杂度。设 a 为输入的特征图，s_p 和 h_p 为池化核的尺寸，“/”为整除取整操作，则平均池化操作 $g=P_{s_p,h_p}(a)$ 的具体表达式为

$$g(i/s_p,j/h_p)=\frac{\sum_{s_p,h_p} a(i+s_p,j+h_p)}{s_p h_p}。 \tag{4}$$

当第 l 层为全连接层时，该层中第 j 个神经元输出计算公式为

$$a_j^l = f_F\left[\boldsymbol{W}_F^l\left(\sum_{i\in M_j^l} a_i^{l-1}\right)+\boldsymbol{b}_F^l\right], \tag{5}$$

式中 $f_F(\cdot)$ 为全连接层的激活函数，$\boldsymbol{W}_F$ 为全连接层的权重，$\boldsymbol{b}_F$ 为全连接的偏置。经过若干全连接层后，最后一层为 Softmax 输出层，该输出层与普通全连接层的区别在于激活函数为 Softmax 函数[8]。

2.3 卷积神经网络后向传播过程

卷积网络的训练目标是最小化网络的损失函数

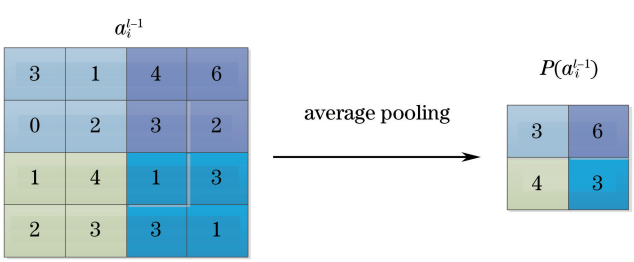

图 4 平均池化操作

Fig. 4 Average pooling operation

L，又称代价函数。该函数的功能是用来评价模型的预测值 $\hat{\boldsymbol{y}}$ 与真实值 $\boldsymbol{y}$ 的不一致程度，它是一个非负实值函数。损失函数 L 越小，模型的性能就越好。损失函数的表达式如下：

$$L=\sum_{i=1}^{N} l(\boldsymbol{y}_i,\hat{\boldsymbol{y}}_i)。\tag{6}$$

卷积神经网络后向传播的实质是根据损失函数 L 的变化情况迭代调整网络的权重 $\boldsymbol{W}$ 和偏置 $\boldsymbol{b}$，公式如下

$$\boldsymbol{W}_{t+1}=\boldsymbol{W}_t-\eta\frac{\partial L(\boldsymbol{W}_t,\boldsymbol{b}_t)}{\partial \boldsymbol{W}_t},\tag{7}$$

$$\boldsymbol{b}_{t+1}=\boldsymbol{b}_t-\eta\frac{\partial L(\boldsymbol{W}_t,\boldsymbol{b}_t)}{\partial \boldsymbol{b}_t},\tag{8}$$

式中 t 为迭代次数，$\boldsymbol{W}_t$ 和 $\boldsymbol{b}_t$ 分别为第 t 次迭代得到的网络权重和偏置向量，$L(\boldsymbol{W}_t,\boldsymbol{b}_t)$ 为在参数向量 $\boldsymbol{W}_t$、$\boldsymbol{b}_t$ 下的损失函数，η 为学习率。

图 5 所示为后向传播模型，其中 $\boldsymbol{y}_{t+1}$ 为神经元当前的输入，L_t 为模型当前的损失函数，$\boldsymbol{w}_{t+1}$ 和 $\boldsymbol{b}_{t+1}$ 分别是由当前损失函数调整得到的新的权值和偏置，x_{t+1} 为反向传播过程中的输出。

与传统识别算法相比，基于卷积神经网络的深度学习模型具有权值共享、模型复杂度低、权值数量少等优点，可避免复杂的手动特征提取和数据重建过程，也可实现自动特征学习，具有较高的研究价值与实用价值。

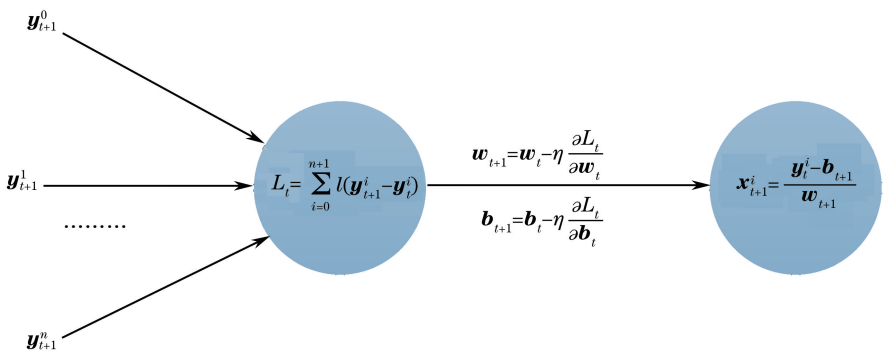

图 5 后向传播模型

Fig. 5 Backward propagation model

3 基于双流卷积神经网络的 RGB-D 信息融合

3.1 早期融合和后期融合策略

为了对 RGB-D 图像进行联合检测，可以通过卷积神经网络对 RGB 和深度图像的信息进行融合。当前基于卷积神经网络的主要的 RGB-D 信息融合模式为早期融合和后期融合[9-10]，其结构分别如图 6 和图 7 所示。卷积网络使用卷积层 C 和池化层 P 交替连接的方式，但在最后一个池化层之前使用连续卷积的方法，如图 6 中的 C6 和 C7 所示，这种策略是为了在池化次数不变的情况下使用多次卷积挖掘图像更深层次的特征。为了方便表述，下文所有方法中 RGB-D 信息的融合层都设为 l 层。

图 6 所示的早期融合过程为采用卷积核对四通道的 RGB 图像和深度图像进行卷积，将卷积后的信息按相等的权值进行相加。设输出的 RGB 原始图像为 X_r，深度原始图像为 X_d，根据(1)式可以得到融

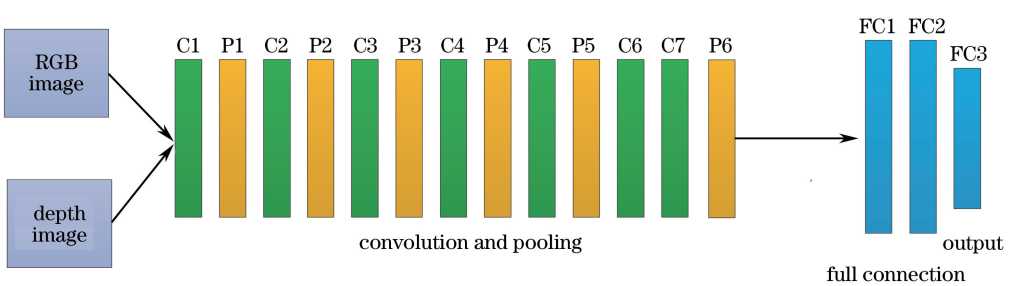

图 6 早期融合结构

Fig. 6 Early fusion structure

合层的第 j 个神经元表达式：

$$a_j^l = f_C[\boldsymbol{W}^l \cdot (0.5 \cdot X_r * k_{ij}^l + 0.5 \cdot X_d * k_{ij}^l) + \boldsymbol{b}^l], \quad (9)$$

式中“+”指对卷积得到的 RGB 特征图和深度特征图进行加法运算，本质是矩阵的加法运算。所有的 RGB-D 信息在第一个卷积层即完成了融合，其融合权值都为 0.5，之后对融合后的 RGB-D 信息进行完整的卷积网络映射，这种方式偏向于对初始图像信息进行融合。

图 7 所示的后期融合过程是将 RGB 图像和深度图像分别输入到两个卷积神经网络中，这两个卷积神经网络具有完全相同的结构，经过卷积、池化和全连接操作后，共同连接到输出层，这种方式偏向于在决策层对 RGB-D 信息进行融合。

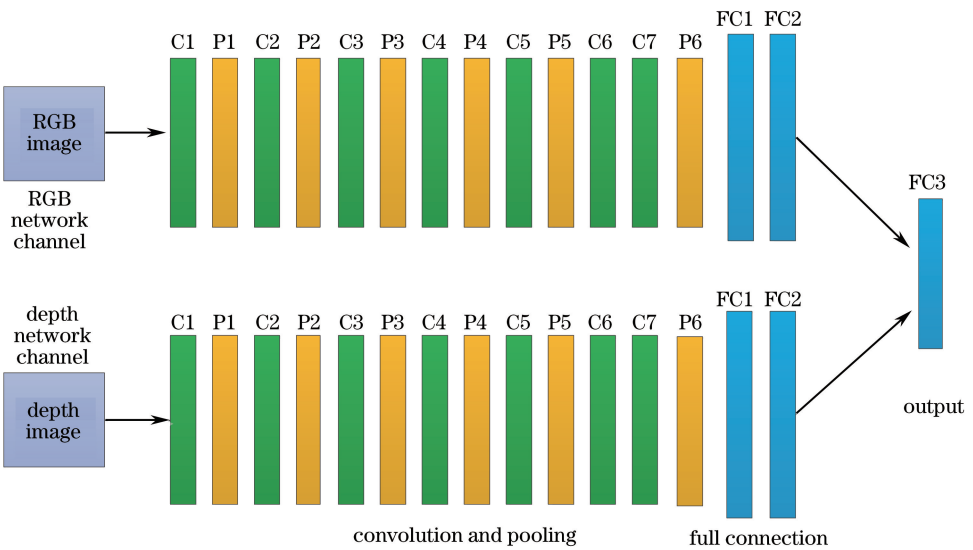

图 7 后期融合结构

Fig. 7 Late fusion structure

在输出层对两种图像进行融合之前，两个独立的卷积神经网络分别对两种图像进行运算，其卷积、池化、全连接运算过程分别如(1)、(3)、(5)式所示。RGB 网络通道中最后一层全连接层中得到的神经元为 a_{rj}^{l-1}，Depth 网络通道中最后一层全连接层中得到的神经元为 a_{dj}^{l-1}，根据(5)式可以得到融合层的第 j 个神经元计算公式：

$$a_j^l = f_F\left\{0.5\left[\boldsymbol{W}_r^l\left(\sum_{i\in M_{rj}^l} a_{rj}^{l-1}\right)+\boldsymbol{b}_r^l\right]+0.5\left[\boldsymbol{W}_d^l\left(\sum_{i\in M_{dj}^l} a_{rj}^{l-1}\right)+\boldsymbol{b}_d^l\right]\right\}, \quad (10)$$

式中 $\boldsymbol{W}_r$ 和 $\boldsymbol{b}_r$ 分别为 RGB 网络通道的权重和偏置，$\boldsymbol{W}_d$ 和 $\boldsymbol{b}_d$ 分别为 Depth 网络通道的权重和偏置，M_r 为从 RGB 网络通道中选择的输入特征图的集合，M_d 为从 Depth 网络通道中选择的输入特征图的集合，f_F 为 Softmax 激活函数，a_j^l 为神经网络的输出。

3.2 全连接层融合和卷积层融合策略

3.1 节中的两种网络结构都对 RGB-D 图像特征进行了一定的融合，但是早期融合结构偏重考虑 RGB-D 图像之间的相关性，对 RGB 和深度图像的独立特性考虑得不够充分，后期融合结构则偏重考虑 RGB-D 图像的独立性。为了更有效地融合 RGB-D 信息，本课题组提出了全连接层融合和卷积层融合的策略，其结构分别如图 8 和图 9 所示。

图 8 所示的全连接层融合结构与后期融合结构类似，只是 RGB-D 信息的融合不是在输出层上，而是在第一个全连接层上，并且两个卷积神经网络不仅采用相同的结构，还进行了权值共享。这一特点的意义在于网络间权值共享不仅增强了两种模态信息在分别进行特征学习时的联系，而且减少了训练参数，提高了训练识别效率。

RGB 网络通道中最后一层得到的神经元为 a_{rj}^{l-1}，Depth 网络通道中最后一层得到的神经元为 a_{dj}^{l-1}，根据(5)、(10)式可以得到融合层的第 j 个神经元计算公式：

$$a_j^l = f_F\left\{\boldsymbol{W}^l \cdot \left[\sum_{i\in M_{rj}^l}(\alpha \cdot a_{rj}^{l-1})+\sum_{i\in M_{dj}^l}(\beta \cdot a_{dj}^{l-1})+\boldsymbol{b}^l\right]\right\}, \quad (11)$$

式中 α、β 分别为 RGB 图像和深度图像的融合系数。因为两个卷积网络间权值共享，所以共用相同的 $\boldsymbol{W}$ 和 $\boldsymbol{b}$。

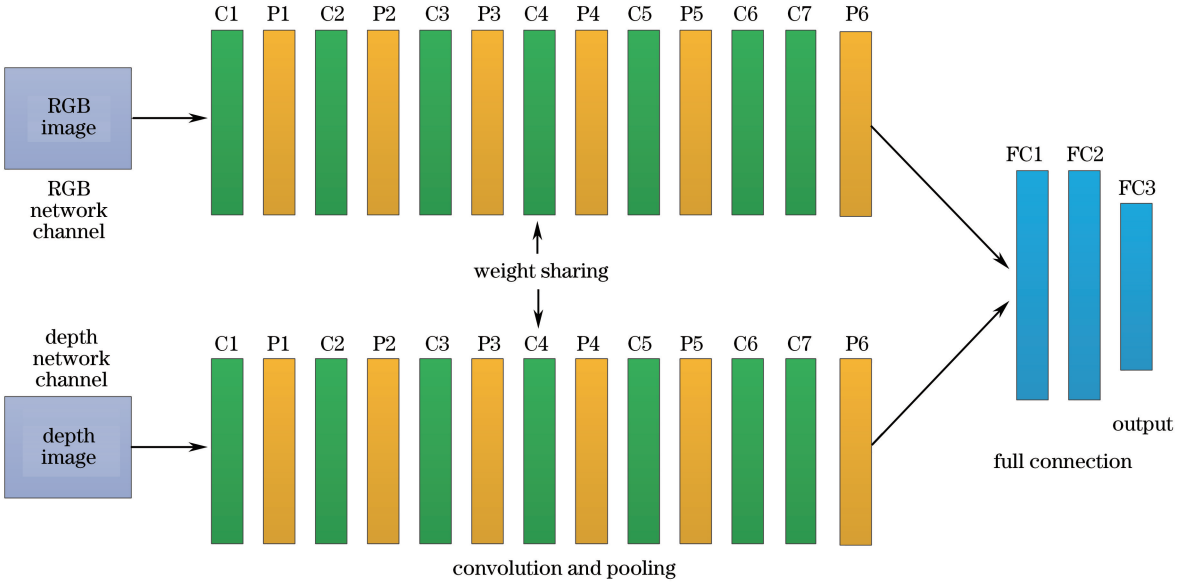

图 8 全连接层融合结构

Fig. 8 Full connection layer fusion structure

图 9 所示的卷积层融合结构如下：首先将 RGB 图像和深度图像分别输入到两个卷积网络中，两个卷积网络间权值共享，经过数次卷积和池化提取各自特征后，对两个卷积网络最后一层的神经元使用相同的卷积核进行卷积，根据最优权值对卷积的结果进行融合，然后继续使用卷积核提取融合后的特征，最后通过全连接层得到输出，由(1)、(11)式可得到融合后的神经元表达式：

$$a_j^l = f_C\left\{\boldsymbol{W}^l \cdot \left[\sum_{i\in M_{rj}^l}(\alpha \cdot a_i^{l-1} * k_{ij}^l) + \sum_{i\in M_{dj}^l}(\beta \cdot a_i^{l-1} * k_{ij}^l)\right] + \boldsymbol{b}^l\right\}。 \tag{12}$$

该网络改进的关键点在于为了保证 RGB 和深度信息的独立性，并没有在最开始时将两种模态的信息输入到同一卷积层中，而是分别在两个卷积通道中进行训练。同时为了保持两种模态信息的联系，在两个卷积网络之间进行权值共享，这样也减少了训练参数，提升了训练和检测速度。经过数层卷积运算挖掘提取图像的内在特征后，为了得到 RGB 信息和深度信息的关联性，再通过相同的卷积核将两种模态信息融合。

对于融合系数 α、β 的计算，本课题组提出了一种最优权值算法。该方法首先使用单通道的卷积神

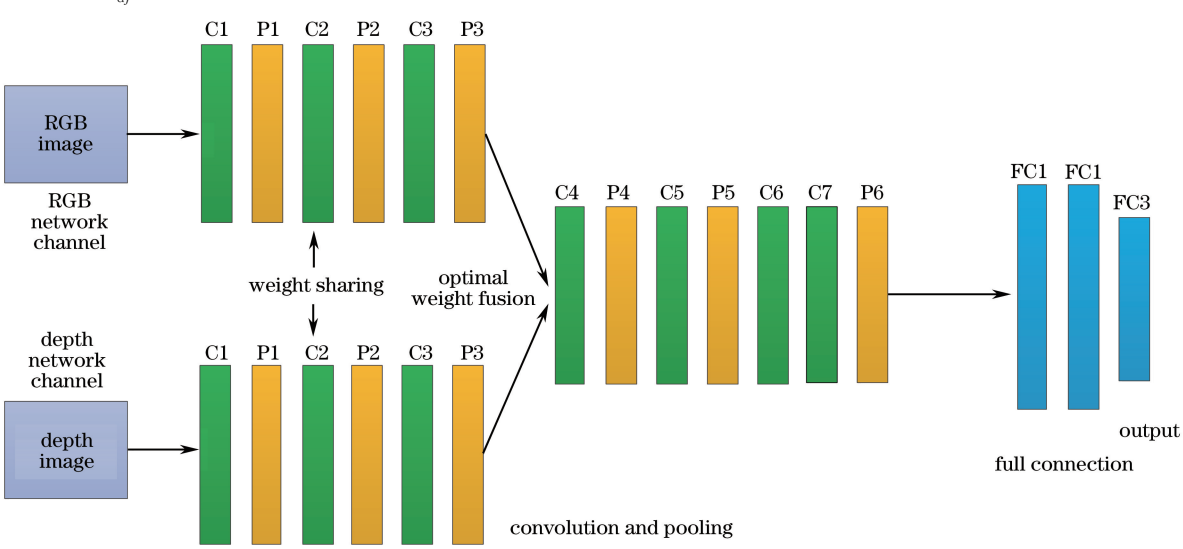

图 9 卷积层融合结构

Fig. 9 Convolution layer fusion structure

经网络对 RGB 图像和深度图像分别进行训练，得到对应的检测准确度之后，按照(10)式计算 RGB 图像和深度图像融合的最优权重。其中，

$$\frac{\alpha}{\beta}=\frac{R_{\mathrm{RGB}}}{R_{\mathrm{d}}}, \tag{13}$$

$$\alpha+\beta=1, \tag{14}$$

式中 R_{RGB} 为单独对 RGB 图像进行检测的准确度，R_{d} 为单独对深度图像进行检测的准确度。

4 实验实现及结果分析

4.1 实验平台和数据集选取

选取 Tensorflow[11-12] 作为实验平台，因其在搭建深度学习模型方面具有独特优势。它可提供 Python 和 C++编程接口，通过把节点分配给多个图像处理器(GPU)可便捷地实现并行计算，加快运算效率。

实验数据库选用华盛顿大学采集构建的 RGB-D 数据库[13]，该数据库利用 Kinect 传感器同步拍摄物体的 RGB 和深度信息，是一个大规模、多层次、多视角的物体和场景数据库。数据集分为两大部分：300 个家用物体构成的 51 个类别数据集以及 8 个办公室和厨房视频 RGB-D 场景数据集。本实验从 8 个视频 RGB-D 场景数据集中抽取部分数据作为训练样本。

4.2 评价指标

使用中心位置误差、准确率和成功率[14]对识别检测结果进行评价。中心位置误差 ε 是检测得到的目标中心点坐标与实际目标中心点坐标的二次方根差，其定义为

$$\varepsilon=\sqrt{(x_1-x_0)^2+(y_1-y_0)^2}, \tag{15}$$

式中 (x_1,y_1) 为识别出的目标中心点的位置，(x_0,y_0) 为标记中心点的位置。本研究对中心位置误差进行了归一化处理。

准确率 P 的定义为中心位置误差小于设定阈值的检测目标个数占检测总目标个数的百分比，其表达式如下：

$$P=\frac{N_{\mathrm{a}}}{N_{\mathrm{z}}}, \tag{16}$$

式中 N_{a} 为中心位置误差小于设定阈值的检测目标个数，N_{z} 为检测总目标数。本研究以 20 个像素作为阈值。

成功率 S 的定义为

$$S=\frac{|R_{\mathrm{t}}\cap R_{\mathrm{a}}|}{|R_{\mathrm{t}}\cup R_{\mathrm{a}}|}, \tag{17}$$

式中 R_{t} 为检测出的目标边界区域，R_{a} 为目标真实的边框区域，$\cap$、$\cup$ 分别表示对 R_{t} 和 R_{a} 作集合的交运算和并运算，$|\cdot|$ 表示集合中像素的个数。当 $S>0.6$ 时，认为检测成功。

4.3 网络参数设置

选用均方误差函数作为网络的损失函数[15]：

$$\mathrm{MSE}(\boldsymbol{W},\boldsymbol{b})=\frac{1}{|\boldsymbol{y}|}\sum_{i=1}^{|\boldsymbol{y}|}[\boldsymbol{y}_{(\boldsymbol{W},\boldsymbol{b})}(i)-\hat{\boldsymbol{y}}(i)]^2, \tag{18}$$

式中 $\boldsymbol{W}$ 和 $\boldsymbol{b}$ 分别为网络的权重和偏置向量，$\mathrm{MSE}(\boldsymbol{W},\boldsymbol{b})$ 为在当前参数向量 $\boldsymbol{W}$ 和 $\boldsymbol{b}$ 下的损失函数值，$\boldsymbol{y}_{(\boldsymbol{W},\boldsymbol{b})}$ 为在当前参数向量 $\boldsymbol{W}$ 和 $\boldsymbol{b}$ 下的实际输出，$\hat{\boldsymbol{y}}$ 为理想输出。

采用随机梯度下降法求取损失函数下降最大的方向。学习率设置为 0.01，激活函数选用 Sigmoid 函数，其形式如下

$$f(x)=\frac{1}{1+\mathrm{e}^{-\alpha x}},\ 0<f(x)<1, \tag{19}$$

式中 x 是(1)、(3)、(5)式中激活函数的输入。Sigmoid 函数的导数为

$$f'(x)=\frac{\alpha\mathrm{e}^{-\alpha x}}{(1+\mathrm{e}^{-\alpha x})^2}=\alpha f(x)[1-f(x)]。 \tag{20}$$

损失函数和测试中心位置误差随训练步数的变化情况如图 10 所示，在 200 步训练步数之内，损失函数和测试中心位置误差随着训练步数降低而迅速下降，之后趋于平缓，在 700 步左右时损失函数达到一个极小值，因此本实验训练步数选 700。

由构建的卷积网络结构分别对 RGB 图像和深度图像进行训练，得到几类不同物体分别在 RGB 和深度图像下的检测精度，然后由(10)式计算出 RGB 和深度图像的融合权重，结果如表 1 所示。

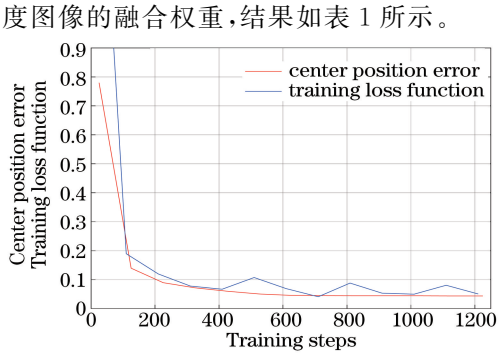

图 10 测试中心位置误差和训练损失函数随训练步数的变化曲线

Fig. 10 Change curves of center position error and training loss function with the training steps

表 1 不同检测物体的融合权重

Table 1 Fusion weight of different detection objects

	RGB-accuracy	D-accuracy	RGB-weight	D-weight
Flashlight	82.8	77.2	0.518	0.482
Coffee cup	80.4	75.8	0.514	0.486
Cereal boxes	83.2	78.6	0.513	0.487
Bowl	78.4	75.1	0.511	0.489

4.4 实验结果分析

使用图 9 中提出的网络结构和 4.2 节得到的网络参数对 RGB-D 图像集进行训练，RGB-D 信息的融合层选在第三层卷积层之后；采用交叉验证法进行实验[16]，即从每组数据集中随机选取 100 对 RGB 和深度图像作为测试数据集，其余的图像作为训练数据集；反复选取五次，进行五次实验，取五次实验结果的平均值。图 11 所示是对一只白碗的部分检测结果。

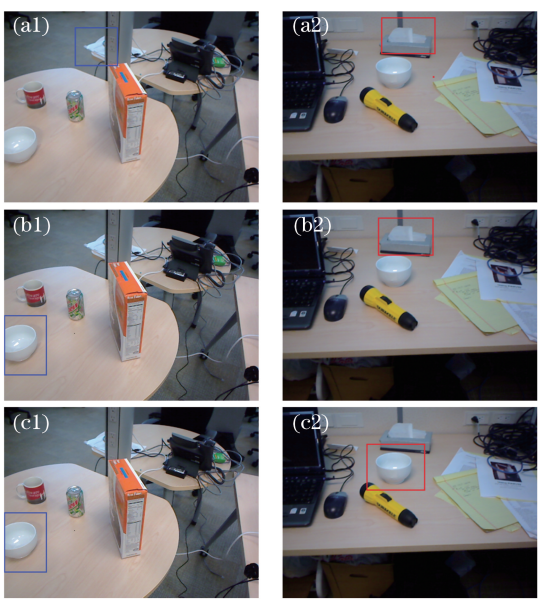

图 11 不同算法下的检测结果。(a)基于 RGB 图像的检测；(b)基于后期融合的 RGB-D 联合检测；(c)基于卷积层融合的 RGB-D 联合检测

Fig. 11 Detection results at different algorithms. (a) Detection based on RGB images; (b) joint detection of RGB-D based on late fusion; (c) joint detection of RGB-D based on convolution layer fusion

在图 11(a)中可以看出，基于 RGB 的检测在这两幅图片中都出现了误检的情况，这是因为只利用 RGB 信息很难区分与目标颜色相近的干扰物体。基于后期融合的 RGB-D 联合检测在一定程度上避免了这种干扰，正确地检测出了图 11(b)中左侧碗的位置。但是当目标与干扰物体相距较近且颜色相似时，仍然存在误检，这是因为 RGB-D 联合检测对深度信息的融合不够充分。所提 RGB-D 信息卷积层融合网络结构由于更有效地融合了 RGB-D 图像的互补特征信息，较好地避免了误检现象，取得了更好的检测效果。不同方法下的检测结果如表 2 所示。

表 2 不同方法的检测结果

Table 2 Detection results by different methods

Method	Central error	Accuracy rate /%	Success rate /%	Detection time /s
RGB image	0.0324	81.2	75.4	0.228
Depth image	0.0371	76.7	71.9	0.177
Early fusion	0.0292	85.6	79.4	0.248
Late fusion	0.0277	87.1	81.3	0.325
FC-fusion	0.0258	88.3	82.2	0.306
C-fusion	0.0235	91.2	84.8	0.288

由得到的检测结果可以看出，无论采用哪种融合方式，基于 RGB-D 图像的联合检测均可显著提高检测精度。RGB 信息的全连接层融合比早期融合、后期融合具有更高的检测准确率和成功率，而且由于双流卷积网络之间的权值共享，检测时间比后期融合方式也有一定缩短。卷积层的融合方式充分考虑了 RGB 图像和深度图像的独立性和关联性，进一步提高了联合检测的效率，其识别准确率和成功率相比于之前最好的融合方法分别提高了 4.1%和 3.5%。

5 结 论

针对 RGB-D 图像联合检测效率不高的问题提出了一种新的双流卷积神经网络结构。该网络结构有效地融合了 RGB 图像和深度图像的互补信息，提高了识别检测精度和成功率，为 RGB-D 图像的联合检测提供了一条新途径。

将 RGB-D 图像分别输入到两个结构相同、权值共享的卷积网络中进行特征学习训练，之后根据一定的权值在卷积层中进行融合，对融合后的特征进行二次卷积训练，有效地提升了 RGB-D 信息的融合利用程度。提出了一种最优融合权值算法，根据各类识别目标在单一模态信息下的识别准确率确定融合时所占的权重，确保得到最高效的融合结果。相比于已有的融合方法，所提 RGB-D 信息融合策略更好地挖掘了 RGB 图像和深度图像之间的内在联系，提高了互补信息的融合效率，其识别准确率和成功率分别提高了 4.1%和3.5%。

参 考 文 献

[1] Cui B. Significant object detection based on RGB-D

information[D]. Tianjin: Tianjin University, 2014: 17-26.
崔波. 基于RGB-D信息的显著物体检测[D]. 天津: 天津大学, 2014: 17-26.

[2] Gupta S, Girshick R, Arbeláez P, *et al*. Learning rich features from RGB-D images for object detection and segmentation [C]. European Conference on Computer Vision, 2014, 8695: 345-360.

[3] Couprie C, Farabet C, Najman L, *et al*. Indoor semantic segmentation using depth information [J]. arXiv: 1301.3572v2.

[4] Gupta S, Arbelaez P, Malik J. Perceptual organization and recognition of indoor scenes from RGB-D images [C]. IEEE Computer Vision and Pattern Recognition, 2013: 564-571.

[5] Eitel A, Springenberg J T, Spinello L, *et al*. Multimodal deep learning for robust RGB-D object recognition[C]. IEEE/RSJ International Conference on Intelligent Robots and Systems, 2015: 681-687.

[6] Rui T, Fei J C, Zhou Y, *et al*. Pedestrian detection based on deep convolutional neural network [J]. Computer Engineering and Application, 2016, 52 (13): 163-168.
芮挺, 费建超, 周遊, 等. 基于深度卷积神经网络的行人检测[J]. 计算机工程与应用, 2016, 52(13): 163-168.

[7] Lu H T, Zhang Q C. Overview of application of depth convolutional neural network in computer vision[J]. 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.

[8] Song S, Xiao J X. Deep sliding shapes for amodal 3D object detection in RGB-D images [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 808-816.

[9] Tu S Q, Xue Y J, Liang Y, *et al*. RGB-D image classification methods[J]. Laser & Optoelectronics Progress, 2016, 53(6): 060003.
涂淑琴, 薛月菊, 梁云, 等. RGB-D图像分类方法研究综述[J]. 激光与光电子学进展, 2016, 53(6): 060003.

[10] Lu L F, Xie Z J, Ye H W. Object recognition algorithm based on RGB-D feature and depth feature fusion[J]. Computer Engineering, 2015, 42(5): 187-192.
卢良锋, 谢志军, 叶宏武. 基于RGB-D特征与深度特征融合的物体识别算法[J]. 计算机工程, 2015, 42(5): 187-192.

[11] Zhang J, Li X. Handwritten character recognition based on TensorFlow platform [J]. Computer Knowledge, 2016, 12(16): 199-201.
张俊, 李鑫. TensorFlow平台下的手写字符识别[J]. 电脑知识, 2016, 12(16): 199-201.

[12] Zhang W. Design and implementation of intelligent home system based on machinelearning[D]. Jilin: Jilin University, 2016: 25-37.
张炜. 基于机器学习的智能家居系统设计与实现[D]. 吉林: 吉林大学, 2016: 25-37.

[13] Lai K, Bo L, Ren X, *et al*. A large-scale hierarchical multi-view RGB-D object dataset [C]. IEEE International Conference on Robotics and Automation, 2011: 1817-1824.

[14] Mao N, Yang D D, Yang F C, *et al*. Adaptive target tracking based on hierarchical convolution[J]. Laser & Optoelectronics Progress, 2016, 53(12): 121501.
毛宁, 杨德东, 杨福才, 等. 基于分层卷积特征的自适应目标跟踪[J]. 激光与光电子学进展, 2016, 53(12): 121501.

[15] Jia Y, Shelhamer E, Donahue J, *et al*. Caffe: Convolutional architecture for fast feature embedding [C]. ACM International Conference on Multimedia, 2014: 675-678.

[16] Cai Q, Wei L W, Li H S, *et al*. Target detection of RGB-D images based on ANNet networks[J]. Journal of Systems Simulation, 2016, 28(9): 2260-2266.
蔡强, 魏立伟, 李海生, 等. 基于ANNet网络的RGB-D图像的目标检测[J]. 系统仿真学报, 2016, 28(9): 2260-2266.