

引用格式:周正,周煦林,刘彬,等. 一种适用于嵌入式平台的 Siamese 网络匹配算法[J]. 微电子学与计算机,2020,37(8):6-9.
[ZHOU Z,ZHOU X L,LIU B,et al. A Siamese network matching algorithm for embedded platform [J]. Microelectronics & Computer,2020,37(8):6-9.]

一种适用于嵌入式平台的 Siamese 网络匹配算法

周 正,周煦林,刘 彬,姚 秦
(西安微电子技术研究所,陕西 西安 710075)

摘 要:在以航天为代表的嵌入式领域,使用传统的模板匹配方法进行目标定位搜索,往往会因为场景复杂,目标物有形变或角度变换而导致匹配的准确率大幅降低.而使用 CNN 网络直接进行目标检测的算法因网络规模和运算量巨大而难以移植到嵌入式平台上.本文提出一种 Siamese 网络匹配算法,一方面利用 CNN 网络的特征提取能力,提高了在复杂场景下的匹配准确性,另一方面利用 Siamese 的对称性,使用子图对匹配过程进行监督,大大缩减了网络规模,使算法可以移植至嵌入式平台,具有很强的实用性.

关键词:图像匹配;卷积神经网络(CNN);深度学习;嵌入式平台

中图分类号: TP183

文献标识码: A

文章编号: 1000—7180(2020)08—0006—04

A siamese network matching algorithm for embedded platform

ZHOU Zheng,ZHOU Xu-lin,LIU Bin,YAO Qin

(Xi'an Microelectronics Technology Institute,Xi'an 710075,China)

Abstract: In the embedded field represented by missile spaceflight, the traditional template matching method is often used to conduct the target location search, and the accuracy of the matching will be greatly reduced due to the complexity of the scene and the deformation or Angle transformation of the target. However, the algorithm that USES CNN network to directly detect the target is difficult to transplant to the embedded platform due to the huge network scale and computation. Siamese network matching algorithm) is presented in this paper, on the one hand, using the network feature extraction ability, improve the matching accuracy under complex scene, on the other hand using symmetry of Siamese, using graph matching process supervision, significantly reduced the size of the network, the algorithm can be transplanted to the embedded platform, have very strong practicability.

Key words: Image matching; CNN; Deep learning; Embedded platform

1 引言

随着航天领域对于精度和目标智能化识别等要求的不断提高,在嵌入式平台下对于图像匹配、目标定位算法的精确性和其再复杂环境下的抗干扰能力的要求也越来越高.

传统的归一化积相关算法实现原理简单,运算量较小,但对于图像的尺寸变换、角度旋转敏感,在图像进行尺寸缩放或角度旋转后,准确率大大降低.另一方面,随着 Alexnet 模型在 ImageNet2012 的比

赛中取第一名^[1],使用卷积神经网络(CNN)的方法处理数字图像,完成对于图像的分类、识别、检测^[2]等任务成为学界与业界热衷的研究方向.而以文献[3]为代表提出的使用 CNN 网络进行目标检测的算法,大大提高了在尺寸、噪声、角度变换等干扰下目标定位的准确性,但计算量极大.算法难以满足在嵌入式硬件平台下尤其是航天领域对于实时性要求.

针对上述算法的不足,本文提出了一种基于 Siamese 网络的模板匹配算法,一方面汲取了卷积神

经网络处理图像的优点,提高了对于图像尺寸和角度变换的抗干扰性;另一方面,模板图像的设计使得整个算法的计算量大大减小,满足了在计算资源有限情况下对于实时性的要求。

2 现有方法的介绍

2.1 传统的灰度图像匹配算法(NCC 算法)

传统的基于灰度的模板匹配算法^[4]在待搜索图与模板图像均未发生失真的情况下,算法性能优异,应用十分广泛。整个算法的流程如下:

在搜索图 S 中,以 (i, j) 为左上角,取 $M \times N$ 大小的子图,通过设定好的相似度函数 $D(i, j)$ 计算其与模板的相似度;调整 i, j 的值遍历整个搜索图,在所有能够取到的子图中,找到与模板图最相似子图作为最终匹配结果。根据 $D(i, j)$ 选取的不同,常用的有:平均绝对差算法 (Mean Absolute Differences, MAD)、平均误差平方和算法 (Mean Square

Differences, MSD)、归一化积相关算法 (Normalized Cross Correlation, NCC) 等。

以 NCC 算法为例,其相似度函数为:

$$D(i, j) = \frac{\sum_{s=1}^M \sum_{t=1}^N (S(i+s-1, j+t-1) - \bar{S}) \cdot |T(s, t) - \bar{T}|}{\sqrt{\sum_{s=1}^M \sum_{t=1}^N [(S(i+s-1, j+t-1) - \bar{S})^2] \cdot \sum_{s=1}^M \sum_{t=1}^N [T(s, t) - \bar{T}]^2}} \quad (1)$$

式中, $1 \leq i \leq m - M + 1; 1 \leq j \leq n - N + 1$ 。本公式计算归一化系数的 $D \in [0, 1]$, 与 MAD 和 MSD 不同, D 值越大, 说明子图与模板图越接近。整个算法需要找到使得 D 值最大的对应子图坐标。

2.2 SSD 检测算法

SSD 算法为单阶段的 CNN 目标检测算法, 在检测精度和检测速度上都有不错的效果, 是使用 CNN 网络进行目标检测的典型算法, 其算法设计如下:

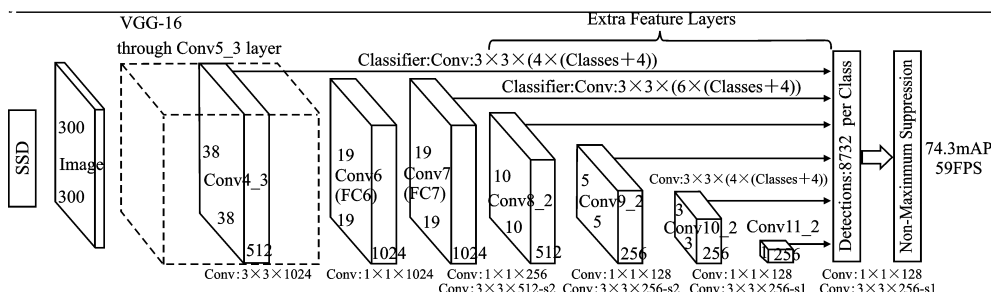


图 1 SSD 算法的网络结构

如图 1 所示, 整个网络中, 输入为 300×300 的图片, 首先将图片传入经典的 vgg16 网络, 经过 vgg16 网络提取特征, 将图片经过卷积层 Conv4 后, 尺寸“压缩”为 $38 \times 38 (\times 512)$ 时, 通过一个 $3 \times 3 \times (4 + \text{class})$ 的 (class 为待预测的目标种类, 例如在本项目中为 8) 卷积核, 得到第一个用于检测的特征图, 尺寸大小为 38×38 。后面新增的卷积层中提取 Conv7, Conv8_2, Conv9_2, Conv10_2, Conv11_2 作为检测所用的特征图, 加上 Conv4_3 层, 共提取了 6 个特征图, 其大小分别是 $(38, 38), (19, 19), (10, 10), (5, 5), (3, 3), (1, 1)$ 。得到特征图后, 需要对特征图进行卷积得到检测结果, 最终的检测值包括两部分: 类别置信度和边界框位置, 分别采用一次 3×3 的卷积完成。其中, Conv4_3, Conv10_2, Conv11_2d 先验框为 4 个, 其余特征图先验框为 6 个, 整个 SSD300 会预测共计:

$38 \times 38 \times 4 + 19 \times 19 \times 6 + 10 \times 10 \times 6 + 5 \times 5 \times 6 + 3 \times 3 \times 4 + 1 \times 1 \times 4 = 8732$ (个) 待选框图, 从中选

出置信度最高的目标位置。

按照文献[3] 的描述, 计算可得整个 SSD 算法的参数规模约为 28 M, 单幅图像的运算量为 30 Gflops。

3 基于 Siamese 网络的图像匹配算法

和一般的目标检测不同, 航天等领域对于目标的检测识别之前往往已经确定了要检测的目标 (拥有模板子图), 只是在传感器传回的图片上对目标位置进行锁定。但如果只是如 NCC 算法一样通过子图与原图的相似度来检测目标位置, 则在目标图片有旋转、形变、噪点等干扰因素下, 准确率会大幅度下降。另一方面, 如果使用如 SSD 算法通过神经网络来进行目标检测, 又没有充分利用到模板子图所提供的待检测目标的特征信息, 导致待检测目标的特征信息必须存储在 SSD 的网络模板参数中, 使得网络结构和运算量都极大, 无法满足嵌入式系统下的实时性要求。

本文提出的基于 Siamese 网络的模板匹配算法既通过 CNN 网络提取了图片的特征,提高了匹配的鲁棒性,又通过模板子图的设计,大大缩减了网络结构,降低了算法的运行时间。

3.1 网络结构

如图 2 所示,将待检测图 S 与待检测目标的模板图 T 预处理为(缩放为) $128 \times 128, 64 \times 64$ 的灰度图,传入训练好的 CNN 网络,网络通过提取待检测图与模板图的特征,输出图片的特征向量图(feature map),将两个特征向量进一步卷积,得到 9×9 的位置置信度图,其中每一点的数值代表着这一点可能存在目标物的概率。最终再通过一个全连接层,网络将输出待检测目标在待检测图上的预测坐标 (x, y, h, w) 。

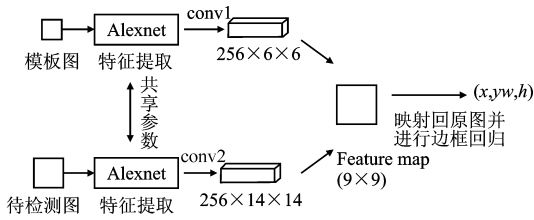


图 2 基于模板的深度学习目标匹配算法

整个检测网络中,特征提取层的网络可以基于典型的目标分类卷积神经网络,如 Alexnet, vgg^[5] 等。本文在实践中采取的是修改过的 alexnet 网络模型,该网络模型共有五层卷积网络,每一层的卷积核设计如表 1。

表 1 特征提取层的卷积核设计

卷积核尺寸 (size)	卷积核通道数 (channel)	卷积核个数 (num)
3×3	1	64
5×5	64	192
3×3	192	256
3×3	256	128
3×3	128	128

可以看出相对于 SSD 算法,由于充分利用了模板子图的特征,整个网络深度浅,参数少,运算量小,便于移植于嵌入式结构。

计算可得该网络的参数量为 0.57 M,计算量约为 90 Mflops。

3.2 训练与损失函数的设计

对网络进行训练前,先对训练数据集进行预处理。主要包括:

(1) 对图像尺寸进行裁剪,使得网络对于图像尺寸变换具有抗干扰性。

(2) 对图像进行随机角度的旋转。这一步主要为了使得网络对于图像的角度旋转具有抗干扰性。

(3) 对图像进行添加噪点,这一步主要为了使得网络对于噪声具有抗干扰性。

训练过程中,在初始化图 2 所示的网络参数后,采取下式为损失函数,使用梯度下降法对参数进行训练。

$$L_{\text{总}} = \lambda L_1 + L_2 \quad (2)$$

$$L_1 = \sum (t_* - P_*)^2 \quad (3)$$

$$L_2 = \sum \max(0, 1 - T \times P) \quad (4)$$

式中, L_1 计算的是输出的边框预测值与人工标记的真实值的差值; $*$ 表示 x, y, h, w 四个下标; t_*, P_* 分别表示真实值与预测值; L_1 的设计主要是为了衡量边框(目标边界)的预测精度。

L_2 计算的是特征图中每一点概率的预测值与真值的差值,使用 Hinge Loss 函数进行计算。 L_2 的设计主要是为了衡量目标位置的预测精度。

总的损失函数为 L_1 和 L_2 的线性组合,通过系数 λ 的调整,使得整个训练过程兼顾了边框大小与边框位置,提高了网络的精准性。

3.3 边框回归设计

为使得网络的匹配更加精确,在网络的末端,通过如下公式进行边框回归,调整网络的输出:

$$\sigma(t) = \frac{1}{1 + e^{-t}} \quad (5)$$

$$b_c = \sigma(t_c) \quad (6)$$

$$b_x = \sigma(t_x) + x \quad (7)$$

$$b_y = \sigma(t_y) + y \quad (8)$$

$$b_w = w e^{t_w} \quad (9)$$

$$b_h = h e^{t_h} \quad (10)$$

式中, σ 表示 sigmoid 函数; t_c 代表位置置信度图的值,通过 sigmoid 函数限制输出值,即 $b_c \in (0, 1)$; t_x, t_y, t_w, t_h 代表坐标偏置 B 的值,同样通过 sigmoid 函数限制的 t_x, t_y 输出值; x, y, w, h 代表网络预测的边框位置值。

4 实验结果与分析

本文的实验以寻找海上船只为背景,将寻找的目标物定为船舶。在此应用背景下,采取了包含船舶目标物的图片共 103 张,首先人工使用矩形框标记目标物在图中的位置,记录具体位置信息以及被框定的子图。如图 3 所示,将子图进行随机的加噪声、旋转、缩放操作。然后同时用不同算法在待检测图中寻找子图位置。

4.1 算法准确性对比

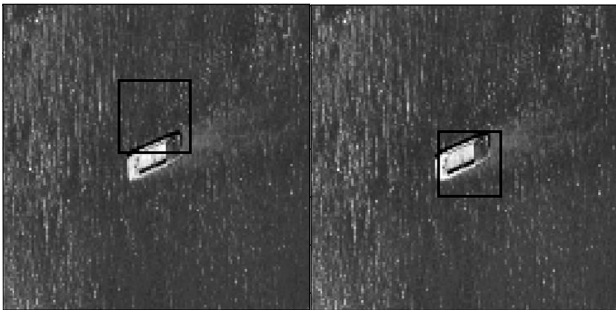
分别在对原图像不进行处理,进行尺寸变换,对

图像进行角度变换三种处理方式下,得到 NCC 与本文提出方法的准确性如表 2.

表 2 不同方法准确性测试结果

不同的图片 处理方法	NCC 算法的 准确性	SSD 算法	本文方法 的准确性
未处理	0.98	0.961	0.951
尺寸变换	0.117	0.942	0.951
角度变换	0.272	0.942	0.932
尺寸加角度变换	0.117	0.903	0.922

以尺寸变化后的实验为例,具体的匹配效果如图 3.



NCC 算法 基于 CNN 网络的算法

图 3 算法的效果图

可以看出,在图片未进行任何处理的情况下,NCC 的准确性极高,但在进行尺寸变换或角度变换后,NCC 算法的准确性大幅度下降,低于 50% 的准确性几乎没有实用价值.而本文提出的方法得益于神经网络对图片特征的提取,准确性一直保持在 80% 以上,准确性与 SSD 算法相当.此实验表明了本文提出的方法在图像有尺寸或角度变换时,仍具有很高的实用价值.

4.2 算法检测速度对比

本文提出的方法与典型的 CNN 网络进行目标检测的 SSD 方法对于船舶的检测,正确率均能达到 90% 以上,但在上述硬件平台下,运算速度大幅度减少,如下表 3.

表 3 不同算法的运行时间结果

算法	单幅图片运行时间(ms)
SSD 算法	≥1 500
本文算法	15.6
NCC 算法	≤10

可以看出,在普通 CPU 上运行 SSD 算法所需要的计算时间是巨大的,而本文的算法在运行时间和参数量上远远小于 SSD 算法.考虑到实验所用 CPU 的主频

仅为 1.6 GHz,而目前主流的 DSP 芯片主频达到 1 GHz,本算法已经具备了朝嵌入式 DSP 上移植的可行性.从本实验可以看出,本算法在导弹航天等计算资源有限的嵌入式平台下,具有实用价值.

5 结束语

图像定位与图像匹配作为计算机视觉与数字图像领域研究的焦点问题,有诸多广泛的应用.然而传统的图像匹配算法对于噪声过于敏感的特点,造成了其鲁棒性差,对同一目标物在不同场景下匹配准确率低的问题.而现有的基于深度学习的检测方法有由于缺少目标模板图的监督,往往网络规模很大.

本文利用了 CNN 网络对图像进行特征提取,在特征图上对子图与待检测图进行匹配.改变了传统的图像匹配仅在像素层面进行匹配的束缚,大大降低了先前方法对于噪声过于敏感的缺点,提高了在有噪声、形变等情况下匹配的准确度,拓宽了匹配算法的应用场景

参考文献:

[1] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet Classification with Deep Convolutional Neural Networks[C]// International Conference on Neural Information Processing Systems, 2012:1097-1105.

[2] MERCIER JP, TROTTIER L, GIGUERE P, et al. Deep Object Ranking for Template Matching[C] // IEEE Winter Conference on Applications of Computer Vision, 2017: 734-742.

[3] LIU WEI, DRAGOMIR A, DUMITRU et al. SSD: Single Shot MultiBox Detector [EB/OL] //(2016-4), <https://arxiv.org/pdf/1611.10012.pdf>.

[4] LIU J, YAN L. Fast realization of image correlation matching algorithm[J]. Geomatics and Information Science of Wuhan University, 2007, 32(8): 684-687.

作者简介:

周 正 男, (1995-), 硕士研究生. 研究方向为嵌入式高速计算机技术. E-mail: 2558796299@qq.com

周煦林 男, (1970-), 硕士, 研究员, 硕士生导师. 研究方向为嵌入式高速计算机技术.

刘 彬 女, (1983-), 硕士, 工程师, 研究方向为嵌入式计算机.

姚 秦 男, (1988-), 硕士, 工程师, 研究方向为嵌入式计算机.