文章编号:1671-4598(2021)06-0128-08 DOI:10.16526/j.cnki.11—4762/tp.2021.06.027 中图分类号:TP3 文献标识码:A

基于改进 CRNN 的导弹编号识别算法研究

何伟鑫，邓建球，丛林虎
(海军航空大学 岸防兵学院，山东 烟台 264001)

摘要：海军某部队导弹出入库使用人工登记，导致了出入库过程浪费大量的人力时间；利用深度学习，可实现自动登记；文章介绍了 AA—CRNN 算法，即在 CRNN 中加入非对称卷积，提升宽度感受区域，加入 Attention 机制对特征序列进行加权平均；通过在人工合成的数据集上进行实验对比分析，AA—CRNN 算法识别目标导弹编号模型的准确率以及 LOSS 均达到较好的性能，较其他先进的文本识别算法，其字符准确率达到了 98.9%，同时其平均编辑距离低至 0.92，且经实际测试其均能准确识别出导弹编号；因此，利用改进的 CRNN 算法识别导弹编号，辅助工作人员进行导弹出入库自动登记方案是可行的。

关键词：导弹；卷积神经网络；循环神经网络；文本识别；注意力机制；数据合成；图像变换

Research on Missile Number Recognition Based on Improved CRNN Algorithm

He Weixin, Deng Jianqiu, Cong Linhu
(Naval Aviation University, Yantai 264001, China)

Abstract: A certain naval force used manual check—in for missiles in and out of the warehouse, which caused a lot of manpower and time to be wasted during the warehouse. Using deep learning, automatic registration can be achieved. This article introduces the AA—CRNN algorithm, which adding asymmetric convolution to the CRNN algorithm, the width of the perception area is increased, and the Attention mechanism is added to perform a weighted average of the feature sequence. Through experimental comparative analysis on artificially synthesized data sets, the accuracy and the LOSS of the target missile number model based on the AA—CRNN algorithm have achieved better performance. Compared with other advanced text recognition algorithms, its character accuracy has reached 98.9%, and its average editing distance is as low as 0.92. And it can accurately identify the missile number after actual testing. Therefore, it is feasible to use the improved CRNN algorithm to identify the missile number and assist the staff in the automatic registration of missiles in and out of the warehouse.

Keywords: missile; CNN; RNN; text recognition; attention; data synthesis; image transformation

0 引言

随着智能化时代的发展，大多数行业均向着人工智能化方面发展。部队也在积极寻求着人工智能方面的应用。海航机载弹药大队任务繁忙，弹库装备出入库次数较多，而每一次进出均需进行出入库的登记。而目前却还是人工登统计的方式进行着工作，浪费时间与人力，这与我军的发展目标背道而驰。随着深度学习的发展，可利用自然场景文字识别技术识别装备上的编号，实现弹库自动化登统计方式[1-4]。

识别自然场景中的文字意指从图片中定位到文字部分，而后得出具体文字内容的过程。该过程不仅可单独用于文字的实际识别，同时也可与场景文字检测算法组合成为一个连续的端到端场景文字识别系统。场景文字识别技术是一项难度系数较大的任务，因为自然场景中不仅存在光照的变化以及图像的复杂背景，同时文字存在的多角度、多维、文字的长度等也影响着场景文字识别的准确度。所以，不同于传统图像分类技术，自然场景文字识别技术[5-9]寻求的是从图像中识别出非固定长度序列。

传统文本识别方法局限太大，如模板匹配法在一些应用很凑效，如身份证号码识别，其只存在阿拉伯数字、字体统一清晰，识别难度较低。然而较复杂的场景，传统方法难以满足需求。而 OCR 的通用方法：设计特征，提取特征而后进行分类，得出最终识别结果，但效果也很难达到应用需求。针对传统 OCR 的不足，使得基于深度学习的 OCR 大放异彩[10-13]。

深度学习的出现，让文字识别技术有了新突破，识别率有了质的提升，同时不需耗费较长时间对字符特征进行设计。在文字识别模型中，神经网络的任务是提取图像特征并归类，得到文本的具体结果。其中，在众多基于深度学习的文本识别算法中，CRNN[14]算法无疑是性能最为优秀的，文献 [14] 表明了该算法在多个数据集中的识别准

收稿日期:2020-09-18； **修回日期**:2020-10-15。
基金项目:国家自然科学基金(51605487)。
作者简介:何伟鑫(1995-)，男，福建漳州市人，硕士生，主要从事深度学习方向的研究。
引用格式:何伟鑫，邓建球，丛林虎. 基于改进 CRNN 的导弹编号识别算法研究[J]. 计算机测量与控制，2021，29(6)：128-135.

确率遥遥领先于其他算法，且该算法支持多方向文本识别，对复杂场景下的文本等也拥有较高的识别率[15-20]。

针对包含航空导弹编号的文本识别，介绍基于 CRNN 的导弹编号识别模型 AA－CRNN。AA－CRNN 增加了非对称卷积（asymmetric convolution）和注意力机制（attention），可以提取更丰富的语义信息[21-23]。而基于深度学习的文字识别算法需大量文本图片数据集，而在海军某机载弹药大队中获取到的装备编号数据集数据量不足会使得模型存在过拟合，而文本图像不能够对其进行简单的数据增强，因此本课题研究人工合成装备编号文本数据集供训练使用[24-25]。

1　基于改进 CRNN 的文本识别算法

CRNN 对于图像的序列识别任务具有较好效果，特别是场景文字识别问题。图 1 为 CRNN 结构图，可看出共可分为 3 个大层：卷积层、递归层和翻译层[14]。

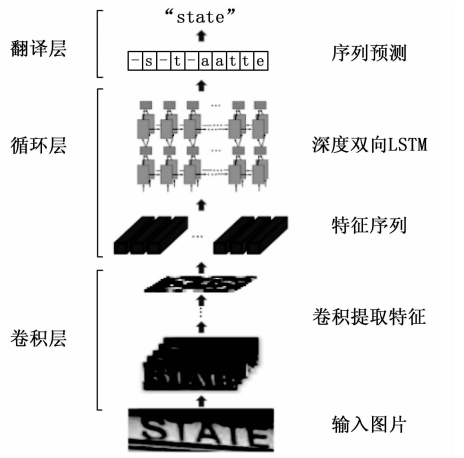


图 1　CRNN 结构图

CNN 层提取输入图像的特征序列。CNN 卷层之后为 RNN 层，用以预测 CNN 层输出的特征序列的每帧。翻译层将 RNN 层的单帧预测转换为标签序列。虽然 CRNN 是由不同类型的网络体系结构组成的，但是该网络可用一个损失函数联合训练 CNN 和 RNN。

1.1　CNN 结构

CNN 部分采用的是 VGG 的结构，并且对 VGG 网络做了一些微调，如表 1 所示。

从表 1 看出，对 VGG 的调整如下：为了将 CNN 提取的特征作为输入，输入到 RNN 网络，将第三和第四个 maxpooling 的步长从 2 * 2 改为了 1 * 2，这样做可以使得特征图的宽度可以被更好地保留。为了加速网络的训练，在第五和第六个卷积层后面加上了 BN 层。

该网络的输入为 W * 32 单通道灰度图，亦即网络对输入图片的宽度无特殊的要求，但高度必须为 32。如一张包含 10 个字符的图片大小为 100 * 32，经上述的卷积神经网络后得到的特征尺度为 25 * 1，这样得到一个序列，每一列特征对应原图的一个矩形区域（如图 2 所示），这样就很方便作为 RNN 的输入进行下一步的计算了，而且每个特征与

表 1　CRNN 网络参数表

TYPE	Configurations
Transcription	—
Bidirectional－LSTM	# hidden inits:256
Bidirectional－LSTM	# hidden inits:256
Map－to－Sequence	—
Convolution	# maps:512,k2 * 2,s1 * 1,p0 * 0,bn
MaxPooling	k2 * 2,s1 * 2
Convolution	# maps:512,k3 * 3,s1 * 1,p1 * 1
Convolution	# maps:512,k3 * 3,s1 * 1,p1 * 1,bn
MaxPooling	k2 * 2,s1 * 2
Convolution	# maps:256,k3 * 3,s1 * 1,p1 * 1
Convolution	# maps:256,k3 * 3,s1 * 1,p1 * 1,bn
MaxPooling	K2 * 2,s2 * 2
Convolution	# maps:128,k3 * 3,s1 * 1,p1 * 1,bn
MaxPooling	K2 * 2,s2 * 2
Convolution	# maps:64,k3 * 3,s1 * 1,p1 * 1,bn
Input	W * 32 单通道灰度图

输入有一个一对一的对应关系，而且 1 * 2 的 pooling stride 使得感受野具备较窄的宽度，有助于识别”i”,”1”等较窄的字符。

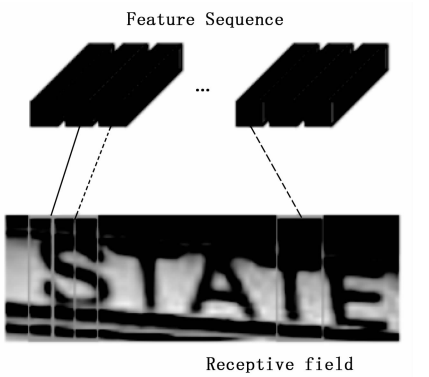


图 2　CNN 序列图

由于卷积神经网络中的 CNN 层以及 pool 层都存在局部性，所以其提取的特征序列同样存在局部性。在图 2 中，每一序列均与输入图像的某一区域相对照。所以 CRNN 的 CNN 层提取图像的序列特征。

对于不同类型的视觉识别任务，深度卷积特征具有鲁棒性、丰富性和可训练性。以前的一些方法已经使用 DCNN 来学习类序列对象的鲁棒表示。然而，这些方法通常通过 DCNN 提取整个图像的整体表示，然后收集局部深度特征来识别类序列对象的每个分量。由于 DCNN 要求将输入图像缩放到固定大小，以满足其固定的输入维数，因此不适合于类序列对象，因为它们的长度变化很大，另一方面，文字的局部细节容易丢失。在 CRNN 中，将深层特征传递到序列表示中，以便对类序列对象的长度变化不变。

与普通图像识别相比，文本识别的任务是识别整个文本行，因此输入的图像宽度一般比高度数值大得多。然而，

普通卷积的感受野具有相同的宽度和高度，可能无法很好地提取文本图像的特征[22]。因此，引入了非对称卷积(asymmetric convolution)，用于适应文本的特征。本文中的非对称卷积运算如图 3 所示。在卷积层中加入 和 非对称卷积核进行卷积运算。该操作相当于增加了图像中水平文本区域的接受野，使得网络增强了从水平文本区域提取特征的能力。

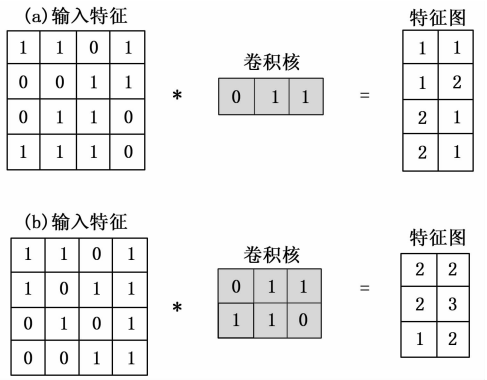


图 3 不对称卷积运算

1.2 RNN 结构

在卷积层的顶部建立了一个深度双向递归神经网络。递归层对每个帧都预测一个标签的概率分布。循环层有三个优点。首先，RNN 具有在序列中捕获上下文信息的强大能力。利用上下文线索进行基于图像的序列识别比独立处理每个符号更加有效。如宽字符需要多个帧进行处理。此外，一些模棱两可的字符在观察其上下文时会更容易区分，例如，通过对比字符高度来识别“il”比单独识别它们中的每一个更容易。其次，RNN 可以反向传播误差至卷积层，使得网络可以进行端到端地训练。第三，RNN 能够预测随机长度的序列。

经典 RNN 单元的输入及输出层间存在一自连接的隐藏层。在序列接收帧时，该层会用一个非线性函数来更新其内部状态，该函数同时以当前输入和过去状态作为输入:，而后其对当前帧做出预测。所以值由过去以及现在的输入决定，所以能够将前文信息进行利用。

但经典 RNN 可能产生梯度消失的情况[20]，这影响其可存储的上下文区间，且加深训练难度。长短期记忆[23](LSTM) 是一种特殊的 RNN，专门解决梯度消失问题。LSTM (图 4) 由一个内存单元以及输入门、输出门还有忘记门组成。内存单元对过往上下文进行储存，输入和输出门允许单元格存储很长一段时间的上下文。同时，细胞中的内存可以通过忘记门进行更新。LSTM 的特殊设计使它能够在多次训练中保持稳定，从而能够获得长距离的上下文。

LSTM 是定向的，它只使用过去的上下文。然而，在基于图像的序列中，来自两个方向的上下文均有用，因此，CRNN 使用两个 LSTM，分别向前以及向后，组合双向 LSTM (BiLSTM)。此外，堆叠多个 BiLSTM，形成深层

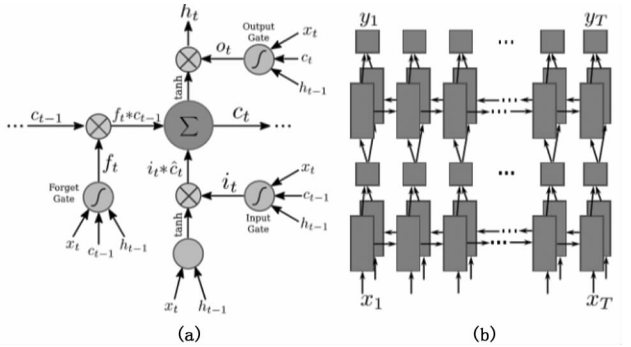


图 4 LSTM 结构

BiLSTM。如图 4 (b) 所示。

误差通过时间反向传播算法在图 4 (b) 所示的箭头的相反方向上传播。在 RNN 的底部，传播的误差序列被连接成图，将特征图转换成特征序列，并反馈到卷积层，这一步骤是通过“Map－to－Sequence”的自定义网络层实现的，该层是卷积层和循环层之间的桥梁。

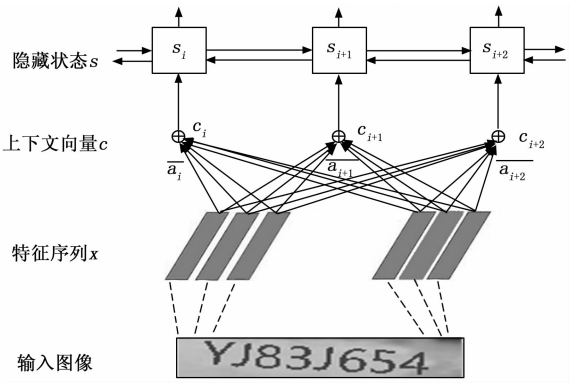


图 5 加入注意力机制的 BiLSTM

在文本识别中，RNN 可以作为解码模块对 CNN 中的特征序列进行解码，并输出最终预测结果。如果使用固定长度的向量对句子进行编码，会导致严重的过拟合问题，特别是对于长输入序列。这是因为输入序列被编码为一个固定长度的向量表示，不管它有多长，但是由于输入的固定长度向量，解码器在解码过程中将受到限制[22]。然后，针对这一问题提出了注意机制。研究将注意力机制添加到双向 LSTM 中。因此，注意机制使得双向 LSTM 能够接收到与当前输出相关的特征序列，并对那些重要的特征序列给予更多的关注，以获得更好的识别结果。本文采用加权注意力机制，根据相关性对特征序列进行加权平均，得到一个上下文向量作为双向 LSTM 的输入。其结构如图 5 所示。其中 c_i 表示为上下文向量，x_j 表示特征序列，a_{ij} 表示在时间 i 时向量 x_j 的相应权重，T_x 表示特征序列的数目，a_{ij} 表示 x_j 被选中的概率，因为 a_{ij} 的和为 1。所以有：

$$c_i = \sum_{j=1}^{T_x} a_{ij} x_j \tag{1}$$

其中：a_{ij} 根据 RNN 的隐藏状态 s_{i-1} 和特征序列 x_j 计算。公式如下：

$$a_{ij} = f(s_{i-1}, x_j) \tag{2}$$

其中：f 是计算特征序列相关性的函数，f 需要用神经网络建模，因此我们采用三层神经网络来建模，其中 tanh 函数作为激活函数。计算过程是：

$$h_{ij} = \tanh(w_{11} \times x_j + w_{12} \times s_{i-1} + b) \tag{3}$$

$$e_{ij} = h_{ij} \times w_{21} \tag{4}$$

其中：h_{ij} 表示特征序列 x_j 和 s_{i-1} 的第一层输出向量，w_{11} 和 w_{12} 分别表示第一层的权重，b 和 w_{21} 分别表示偏差项，第二层的权重和 e_{ij} 表示神经网络的第二层得分。然后在 e_{ij} 上施加 softmax 层以获得 a_{ij}，如公式：

$$a_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_x} \exp(e_{ik})} \tag{5}$$

然后将每个注意模块输出的上下文向量输入到双向 LSTM 中，得到性能帧预测，并将预测结果反馈给 CTC 层。改进后的网络结构如图 6 所示。

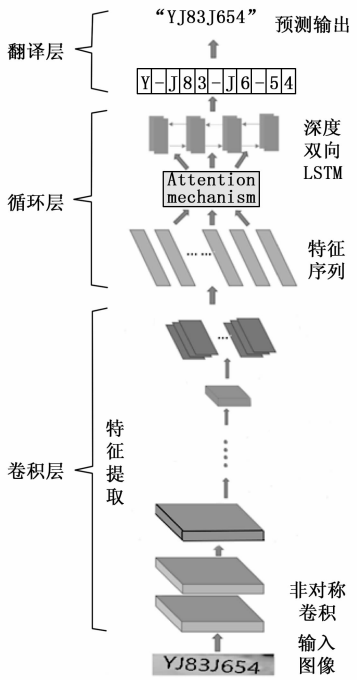


图 6　改进的网络结构

1.3　翻译层

翻译层是将 RNN 所做的每帧预测转换成标签序列的过程。从数学上讲，翻译层是根据每帧预测找到概率最高的标签序列，其共有两种不同模式：无字典及基于词库[23]。无字典的情况下，预测是在没有任何词库的情况下进行的。在基于词库的模式下，通过选择后验概率最高的标签序列进行预测。CRNN 标签序列概率采用 Graves 等人提出的连接时续分类（CTC）进行计算[24]。

1.3.1　序列合并机制

RNN 对时序进行预测时，不可避免地会产生多余信息，可能单一字符被接连预测多次，这需一种去冗余机制。

如识别图 6 文本，RNN 中有 5 个时间步，在正常情况下 t_0, t_1, t_2 映射为“a”，t_3, t_4 映射为“b”，然后将这些字符序列连接起来得到“$aaabb$”，而后将连续重复的字符合并成一个，得最终结果“ab”。但如是 look，hello 等存在连续相同字母的词，则得到 lok 和 helo，产生错误，因此 CTC 提出 blank 机制解决该问题。

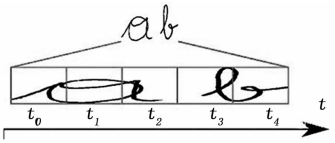


图 7　RNN 预测示意图

以“－”符号代表 blank，RNN 输出序列时，在文本标签中的重复的字符之间插入一个“－”，比如输出序列为“llooo－ookk”，则映射输出“look”，也就是如果有 blank 字符隔开，连续相同字符不合并。即对序列先去掉连续相同字符，接着去掉“－”字符，这个称为解码过程，而编码则是由神经网络来实现。引入 blank 机制，可以很好地解决重复字符的问题。相同的文本标签一定几率具有不同的字符组合如，“11－2”、“1122”及“－122”均表示“12”。也就是说一个文本标签存在一条或多条的路径。

1.3.2　训练阶段

在训练阶段，根据这些概率分布向量和相应的文本标签得到损失函数，从而训练神经网路模型。

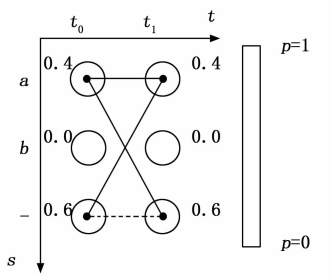


图 8　文本标签概率分布图

图 8 表示时序为 2 的字符识别，有两个时间步长和三个可能的字符为“a”，“b”和“－”，可得两个概率分布向量，如采取最大概率路径解码的方法，则“－－”的概率最大，即真实字符为空的概率为 $0.6 * 0.6 = 0.36$。但是为字符“a”的情况有多种对齐组合，“aa”，“a－“和“－a”都是代表“a”，所以，输出“a”的概率应该为三种之和：$0.4 * 0.4 + 0.4 * 0.6 + 0.6 * 0.4 = 0.16 + 0.24 + 0.24 = 0.64$，因此“a”的概率比空“”的概率高。如果标签文本为“a”，则通过计算图像中为“a”的所有可能的对齐组合（或者路径）的分数之和来计算损失函数。所以最后映射为标签文本的总概率为：

$$p(l \mid x) = \sum_{\pi \in B^{-1}(l)} p(\pi \mid x) \tag{6}$$

其中：$B^{-1}(l)$ 代表从序列到序列的映射函数 B 变换后是文本 l 的所有路径集合，而 π 则是其中的一条路径。每条路径的概率为各个时间步中对应字符的分数的乘积。类似普通的分类，CTC 的损失函数 O 定义为负的最大似然，为方

便计算，对似然函数取对数。

$$O=-\ln\left(\prod_{(x,z)\in S}p(l\mid x)\right)=-\sum_{(x,z)\in S}\ln p(l\mid x) \quad (7)$$

通过对损失函数的计算，就可以对之前的神经网络进行反向传播，神经网络的参数根据所使用的优化器进行更新，从而找到最可能的像素区域对应的字符。这种通过映射变换和所有可能路径概率之和的方式使得 CTC 不需要对原始的输入字符序列进行准确的切分。定义为负的最大似然，为方便计算，对似然函数取对数。

1.3.3 测试阶段

在测试阶段，过程与训练阶段有所不同，用训练好的神经网络来识别新的文本图像。事先不知道任何文本，如过与之前一样计算每一可能文本的所有路径，这样长时间步和长字符序列将产生庞大的计算量。RNN 在每一个时间步的输出为所有字符类别的概率分布，即一个包含每个字符分数的向量，取概率最大字符当做该时间步的输出字符，接着所有时间步的输出组合得一序列路径，即最大概率路径，再根据合并序列方法获得文本的预测。在输出阶段经过 CTC 的翻译，即将网络学习到的序列特征信息转化为最终的识别文本，就可以对整个文本图像进行识别。

如图 9 所示，有 5 个时间步，字符类别为“a”、“b”和“－”(blank)，对于每个时间步的概率分布，取分数最大的字符，得序列路径“aaa－b”，先移除相邻重复的字符得到“a－b”，然后去除 blank 字符得到最终结果：“ab”。

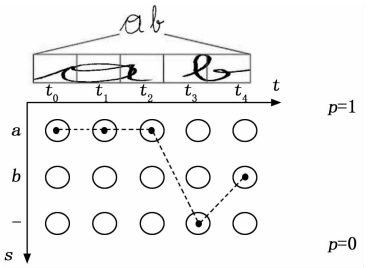


图 9 文本预测图

2 数据集合成

由于军队航空导弹装备的特殊性，当前未有用于导弹编号识别的数据集，而如果仅仅使用通用的文本识别数据集，得到的预测结果准确率将不够高。因此，本课题结合航空导弹编号的文本特征，人工合成可应用于导弹编号识别的文本图像，追求较高的识别准确率。图 10 为人工合成文本数据集的流程。

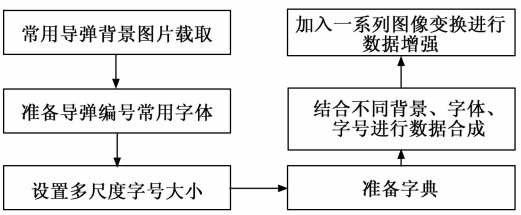


图 10 识别数据集合成流程

2.1 导弹编号背景

为了尽可能真实地获得航空导弹编号的图像场景，实地在海航某部机载弹药大队所拍摄数据集。对于所收集整理到的导弹编号数据集，覆盖海航某部机载弹药大队各项型号装备的不同背景。值得注意的是，CRNN 模型对于输入数据的高度要求是固定的，所以在截取背景时应注意背景图像的大小，否则会影响到训练。而如果通过缩放等图像处理方法来将图像进行或大或小的缩放，可使得背景的像素与真实值产生偏差，从而影响最终的精度。因此在截取航空导弹编号背景图像时，确保其能够满足 CRNN 输入图像的尺寸要求。同时，在实际的航空导弹业务工作中，计算机识别编号的时候并非严格正对编号，或多或少存在一定角度倾斜，因此在截取背景图像时也按照一定角度进行倾斜。图 11 为背景图示例。

图 11 导弹编号背景图

2.2 字体

文本的字体对于在识别任务中占据较高地位。不同字体间的风格存在着差异，航空导弹编号识别模型为了适应不同的装备，也需不同的字体特征，针对已经采集到的航空导弹编号，对其进行分析，得出大体所需字体为：微软雅黑、新罗马以及 Vanta，如图 12 所示。

图 12 字体示意图

2.3 文本尺寸大小

文本图像上的文本尺寸特征对模型非常重要。尽管相同型号航空导弹的编号尺寸相同，但不同型号上的编号大小存在差异，所以在合成图像时，针对性的采用不同尺寸的字符以丰富特征。且 2.2 中背景截取使用相同大小的背景，所以不同文本大小能够匹配相应文本长度，可更加合理粘贴于背景图像上。

2.4 字典

在合成识别数据集时，利用字典检索的方式对生成的背景图像贴上文本内容。字典中包含英文字母以及阿拉伯数字。每一个字符在字典中单独一行，程序可更容易检索到该字符。随机生成一段 4～10 个字符的文本后，检索该文本所有字符在字典中的位置，而后利用该索引位置可获得字符类别，进而可以产生该文本段的标签。

2.5 数据合成

准备工作过后，随机组合航空导弹编号、背景图以及

不同的字体，将航空导弹编号文本粘贴到背景图像上。这一步骤主要使用了 Pillow，在进行合成的时候需设置将航空导弹编号靠近于背景左上角。合成后的航空导弹文本编号图像如图 13 所示。

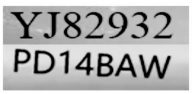


图 13　合成图像示例

2.6　图像变换

由于在航空导弹业务当中，工作场所变换较多，伴随因素影响较大，一些噪声对识别产生干扰的情况无可避免。所以在训练 CRNN 模型过程中需对这种情况进行学习，以提高模型对不同环境的适应能力。所以对合成的数据集应用高斯模糊、灰度拉伸以及透视变换的方式对航空导弹编号文本图像进行增强[25]。

高斯模糊取图像中每个像素周围像素的平均值，使用正态分布对周围像素值权重进行分配。而目标点像素即为正态分布中心，因此靠近该点则权重较大，远离则小，式（8）为其计算公式：

$$G(x,y)=\frac{1}{2\pi\sigma^2}e^{-(x^2+y^2)/2\sigma^2} \tag{8}$$

由于不确定的环境因素可能会影响后续识别过程中文本特征的提取。航空导弹编号文本图像灰度拉伸将图片的灰度值在更大区间内进行扩展以增强图片对比度。提取图的最大以及最小灰度值 $I_{\max}$ 和 $I_{\min}$，MAX 和 MIN 表示目标灰度最大以及最小值，式（9）表示映射过程：

$$I_{(x,y)}=\frac{I_{(x,y)}-I_{\min}}{I_{\max}-I_{\min}}(MAX-MIN)+MIN \tag{9}$$

透视变换将图像映射到一个新平面，首先将图像从二维平面映射到三维空间，接着映射到另一二维平面，如式（10）所示：

$$\begin{bmatrix}X\\Y\\Z\end{bmatrix}=\begin{bmatrix}a_{11}&a_{12}&a_{13}\\a_{21}&a_{22}&a_{23}\\a_{31}&a_{32}&a_{33}\end{bmatrix}\begin{bmatrix}x\\y\\1\end{bmatrix} \tag{10}$$

式中，$[x\quad y\quad 1]^T$ 表示原图像点在三维空间的值，$[X\quad Y\quad Z]^T$ 表示在三维空间中的点，三维矩阵表示透视变换矩阵[24]。图 14 为经过变换后的航空导弹数据。

(a)对比度拉伸　(b)透视变换　(c)高斯模糊

图 14　图像增强后的合成数据集

3　实验结果与分析

对航空导弹编号文本识别模型进行实验。对比在航空导弹编号识别数据集中 CRNN 算法与其它文本识别算法训练得到深度学习模型的识别效果。

识别实验在 Ubuntu16.04 系统下进行，CPU：酷睿 i5－8400 2.80 GHz，显卡为 GTX1080Ti，显存为 11 GB，计算机内存为 16 G；python 3.6，使用 pycharm 作为实验平台，同时配套 tensorflow，tensorboard 支持实验进行。使用 Momentum 优化器进行优化，初始学习率为 0.01，按训练次数衰减，图 15 为学习衰减曲线图，训练次数为 8 万。

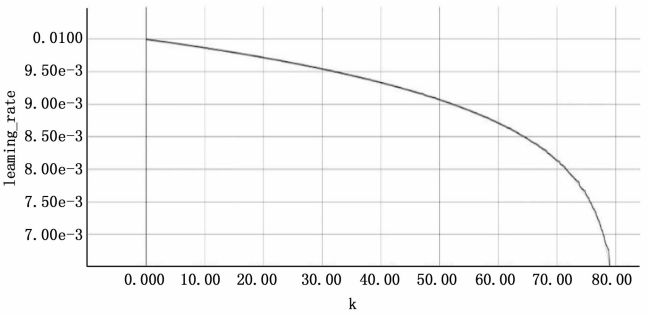


图 15　学习率衰减曲线

3.1　合成数据集以及评价指标

3.1.1　数据集

本实验的训练数据集为前文所合成识别数据集与人工新增入阿拉伯数字图像的公开数据集 Synth90K。本文合成数据集一共 3 万张，包括实地拍摄、合成以及图像增强，标注数据集只需在实地拍摄数据集中按照 Synth90K 数据集进行标注，合成以及图像增强均可以代码形式进行标注。

3.1.2　评价指标

实验中以测试数据的字符识别准确率以及平均编辑距离作为评估标准。字符识别准确率指正确识别的字符数量占总数量的比重，而后对每张图像求得该比重去平均，即为总的字符识别准确率。平均编辑距离是一种度量两个序列（字符串）差异大小的方法。平均编辑距离越小说明识别率越高，可以同时反应识别错，漏识别和多识别的情况。

假设现在两个字符串 A 和 B，其中 A 的长度为 a，B 的长度为 b，现要计算 A 与 B 之间的 Levenshtein distance 可用动态规划的思想解决这个问题：假设 A_i 和 B_i 分别为字符串 A、B 的前 i，j 个字符组成的子串，现将 A_i：A[1] A[2] … A[$i-1$] A[i] 修改为 B_j：B[1] B[2] … B[$j-1$] B[j] 需要的最少编辑次数，即两个子串的编辑距离，下面分别讨论三种操作的操作次数：

1）插入操作：假设将 A[1…i] 修改为 B[1…$j-1$] 需要操作数为 k_1，那么在 A[i] 后插入一个字符 B[j]，这样就可以将 A[1…i] 修改为 B[1…j]，这时所需要的操作数为 k_1+1。

2）删除操作：假设将 A[1…$i-1$] 修改为 B[1…j] 需要操作数 k_2，那么删除 A[i] 就可以将 A[1…i] 修改为 B[1…j]，此时所需要的操作数为 k_2+1。

3）修改操作：假设将 A[1…$i-1$] 修改为 B[1…$j-1$] 需要操作数为 k_3，这时要将 A[1…i] 修改为 B[1…j] 则分为两种情况：一是当 A[i] ≠ B[j] 时，则将 A[i] 替换成 B[j]

即可完成修改，此时操作数为 k_3+1；另一种情况是当 A[i] == B[j]时，则将不需要进行修改操作，操作数仍然为 k_3。最后可得状态转移方程：

$$lev_{a,b}(i,j)=\begin{cases} max(i,j), & \text{if} \quad min(i,j)=0 \\ min\begin{cases} lev_{a,b}(i-1,j)+1 \\ lev_{a,b}(i,j-1)+1 \\ lev_{a,b}(i-1,j-1)+1_{a_i\neq b_j} \end{cases} & \text{otherwise} \end{cases} \tag{11}$$

其中：$1_{a_i\neq b_j}$ 表示 $a_i\neq b_j$ 表达式取 0，否则取 1。

3.2 实验结果评价与分析

首先模型 AA－CRNN 上进行了实验，然后与其他经典的文本识别算法 CRNN，CNN＋CRF 以及 ESIR[25] 为对比对象，三者均为当前较为先进的文本识别算法。对训练得到的各个模型逐一测试，使用前文所提的两个评价指标进行评估，结果如表 2 所示。

表 2 不同模型效果对比

深度学习模型	字符准确率	平均编辑距离
AA－CRNN	0.989	0.92
CRNN	0.977	1.04
CNN＋CRF	0.956	1.43
ESIR	0.968	1.36

由表 2 可以明显看出 AA－CRNN 的字符准确率以及平均编辑距离均优于另外的算法，因此说明本课题改进的 AA－CRNN 算法作为航空导弹编号识别的算法是较优的。图 16 以及图 17 表示训练过程中 AA－CRNN 的 train loss 曲线以及 val loss。

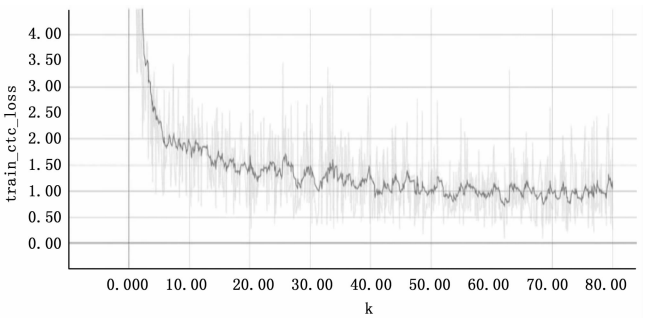


图 16 AA－CRNN train LOSS 曲线

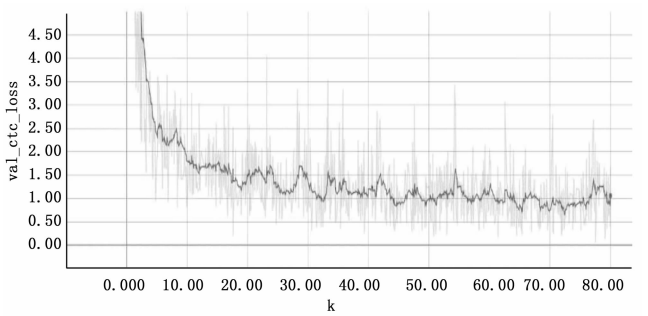
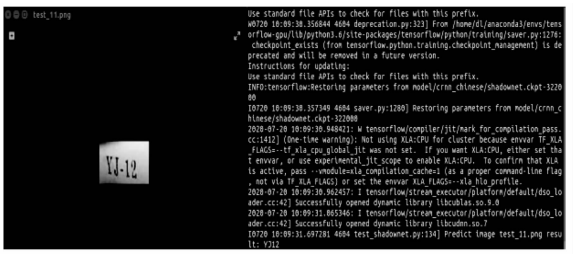


图 17 AA－CRNN val LOSS 曲线

由表 2 以及图 16、17 可以看出 AA－CRNN 模型具备较好的性能，接下来使用图像对其进行测试，如图 18 所示，该模型均准确地将导弹编号识别出来。



(a) 得到结果：YJ12

(b) 得到结果PL519F7

图 18 模型测试

因此，本课题改进的 CRNN 模型在合成的航空导弹编号数据集上训练能够得到性能较好的模型，在实际图片测试中均可将导弹编号准确的识别出来，说明本课题对航空导弹编号识别的研究是可行的。

4 结束语

文章首先分析了文本识别模型在航空导弹业务应用中的地位，介绍了 AA－CRNN 模型。由于缺乏航空导弹编号数据集，因此对照实际编号进行人工合成数据集。通过训练对比，发现 AA－CRNN 的模型性能能够由于当前较为优秀的深度学习模型，且实际测试均正确识别出航空导弹编号，因此 AA－CRNN 模型应用于航空导弹编号识别应用中可行且优秀。

但是同样存在巨大的继续研究的空间，因为工作人员需手持摄像头对准导弹编号进行操作。在未来的时间里，笔者将研究检测与识别结合，且是端到端的训练，而非文本检测与识别分开，如此使得模型运行速度更快，且进一步减轻弹库工作人员的工作量。

参考文献：

[1] Shi B, Wang X, Lyu P, et al. Robust scene text recognition with automatic rectification [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. 2016: 4168－4176.

[2] Cheng Z, Bai F, Xu Y, et al. Focusing attention: Towards accurate text recognition in natural images [A]. Proceedings of the IEEE International Conference on Computer Vision [C]. 2017: 5076－5084.

[3] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [A]. Proceedings of the IEEE Conference on Computer Vision

and Pattern Recognition [C]. 2014：580－587.
[4] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks [J]. Advances in Neural Information Processing Systems, 2012：1097－1105.
[5] Bissacco A, Cummins M, Netzer Y, et al. Photoocr：Reading text in uncontrolled conditions [A]. Proceedings of the IEEE International Conference on Computer Vision [C]. 2013：785－792.
[6] Jaderberg M, Simonyan K, Vedaldi A, et al. Reading text in the wild with convolutional neural networks [J]. Computer Vision, 2016, 116 (1)：1－20.
[7] Su B, Lu S. Accurate scene text recognition based on recurrent neural network [A]. Asian Conference on Computer Vision [C]. Springer, Cham, 2014：35－48.
[8] Ye Q, Doermann D. Text detection and recognition in imagery：A survey [J]. IEEE transactions on Pattern Analysis and Machine Intelligence, 2014, 37 (7)：1480－1500.
[9] Zhu Y, Yao C, Bai X. Scene text detection and recognition：Recent advances and future trends [J]. Frontiers of Computer Science, 2016, 10 (1)：19－36.
[10] Almazán J, Gordo A, Fornés A, et al. Word spotting and recognition with embedded attributes [J]. IEEE transactions on pattern analysis and machine intelligence, 2014, 36 (12)：2552－2566.
[11] Rodriguez－Serrano J A, Gordo A, Perronnin F. Label embedding：a frugal baseline for text recognition [J]. Computer Vision, 2015, 113 (3)：193－207.
[12] Lee C Y, Bhardwaj A, Di W, et al. Region－based discriminative feature pooling for scene text recognition [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. 2014：4050－4057.
[13] Shi B, Bai X, Yao C. An end－to－end trainable neural network for image－based sequence recognition and its application to scene text recognition [J]. IEEE Trans. Pattern Anal, 2017, 39：2298.
[14] Bai X, Yao C, Liu W. Strokelets：A learned multi－scale mid－level representation for scene text recognition [J]. IEEE Transactions on Image Processing, 2016, 25 (6)：2789－2802.
[15] Ioffe S, Szegedy C. Batch normalization：Accelerating deep network training by reducing internal covariate shift [A]. International Conference on Machine Learning [C]. PMLR, 2015：448－456..
[16] Gordo A. Supervised mid－level features for word image representation [A]. Proceedings of the IEEE conference on computer vision and pattern recognition [C]. 2015：2956－2964.
[17] Yao C, Bai X, Liu W. A unified framework for multioriented text detection and recognition [J]. IEEE Transactions on Image Processing, 2014, 23 (11)：4737－4749.
[18] Cong F, Hu W, Huo Q, et al. A comparative study of attention－based encoder－decoder approaches to natural scene text recognition [A]. 2019 International Conference on Document Analysis and Recognition (ICDAR) [C]. IEEE, 2019：916－921.
[19] Chen L, Su H, Ji Q. Deep structured prediction for facial landmark detection [J]. Advances in Neural Information Processing Systems, 2019, 32：2450－2460.
[20] Jaderberg M, Vedaldi A, Zisserman A. Deep features for text spotting [A]. European Conference on Computer Vision [C]. Springer, Cham, 2014：512－528.
[21] Graves A, Mohamed A, Hinton G. Speech recognition with deep recurrent neural networks [A]. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing [C]. IEEE, 2013：6645－6649.
[22] Tong G F, et al. MA－CRNN：a multi－scale attention CRNN for Chinese text line recognition in natural scenes [J]. Document Analysis and Recognition, 2019：1－12.
[23] Le T A, Baydin A G, Zinkov R, et al. Using synthetic data to train neural networks is model－based reasoning [A]. 2017 International Joint Conference on Neural Networks (IJCNN) [C]. IEEE, 2017：3514－3521.
[24] 姜典转. 基于深度学习的票据文本定位与识别研究 [D]. 北京：北京交通大学，2019.
[25] Zhan F N, Lu S J. ESIR：End－to－end scenetext recognition via iterative image rectification [A]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [C]. 2016：2059－2068.

(上接第 59 页)

[4] 赵海生，胥效文. 小型无人机飞行姿态测量系统的设计 [J]. 计算机测量与控制，2012，20 (3)：583－585.
[5] 周　亢，闰建国，屈耀红. 捷联惯导系统姿态测量算法研究 [J]. 计算机测量与控制，2008，16 (6)：763－765.
[6] 曹景伟，朱宝全. 应用 MEMS 陀螺仪和加速度计的汽车运动姿态测量 [J]. 重庆理工大学学报 (自然科学)，2018，32 (4)：48－54.
[7] Zega V, Comi C, Minotti P, et al. A new MEMS three－axial frequency－modulated (FM) gyroscope：a mechanical perspective [J]. European Journal of Mechanics/A Solids, 2018：70.
[8] 陈　晨，赵文宏，徐慧鑫，等. 基于卡尔曼滤波的 MEMS 陀螺仪漂移补偿 [J]. 机电工程，2013，30 (3)：311－313.
[9] 吴冬冬. 基于 MEMS 陀螺仪姿态检测系统研制 [D]. 杭州：浙江理工大学，2014.
[10] 盛娟红，张志安，邢炳楠. 基于 MEMS 陀螺仪和加速度计的自适应姿态测量算法 [J]. 测试技术学报，2018，32 (4)：277－284.
[11] 马培圣，肖前贵，杨柳庆. 基于单天线 GPS 的无陀螺姿态测量系统设计 [J]. 计算机测量与控制，2013，21 (6)：1465－1466.
[12] 杜海龙，张荣辉，刘　平，等. 捷联惯导系统姿态解算模块的实现 [J]. 光学精密工程，2008 (10)：1956－1962.