基于时频特征融合和无锚检测机制的高效话音信号检测框架

李春辉[1]，向 新[1]，杨思力[2]，律国仓[2]，魏景璇[2]，李 桥[1]
（1.空军工程大学航空工程学院，西安，710038；2.94188 部队，西安，710082）

摘要 基于深度学习的宽带信号检测框架将目标检测与时频图结合，能够实现宽带射频系统中多信号的检测、识别和时频定位。然而直接迁移使用原始网络架构难以在实际任务数据集上达到最优信号检测性能。针对这一问题，提出了一种面向话音信号检测任务的网络架构 SignalNet。结合话音信号及任务数据集的特点对网络进行了解耦和任务导向的优化，分别精简了用于特征提取的骨干网络、设计了具有多尺度时频特征上下文融合以及门控注意力组件的颈部网络，同时将传统的有锚检测头替换为无锚机制。实验结果表明，所提网络架构在话音信号检测任务上达到最优检测性能，不仅取得了 97.42% 的 mAP 值，同时具有更少的模型参数和更快的推理速度。

关键词 信号检测识别；深度学习；网络架构设计；任务导向；特征融合
DOI 10.3969/j.issn.2097-1915.2025.03.004
中图分类号 TN92 **文献标志码** A **文章编号** 2097-1915(2025)03-0026-09

An Efficient Voice Signal Detection Framework Based on Time-Frequency Feature Fusion and Anchor-Free Detection Mechanism

LI Chunhui[1], XIANG Xin[1], YANG Sili[2], LYU Guocang[2], WEI Jingxuan[2], LI Qiao[1]
(1. Aviation Engineering School, Air Force Engineering University, Xi'an 710038, China;
2. Unit 94188, Xi'an 710082, China)

Abstract Wideband signal detection framework enables to realize detection, identification, and time-frequency localization of multiple signals in the wideband RF systems with object detection being combined with spectrograms based on deep learning, whereas directly applied original network architecture is difficult to achieve optimal signal detection performance on actual task datasets. For the above-mentioned reasons, this paper proposes a network architecture, SignalNet, for voice signal detection task, which is decoupled for task-oriented optimization according to the characteristics of the voice signals and task dataset. Specifically, the backbone network is streamlined, which is responsible for feature extraction, a neck network that comprises the multi-scale time-frequency feature context fusion and gating attention modules is introduced, and the traditional anchor-based detection head is replaced with an anchor-free one. The experimental results show that the proposed network architecture achieves the optimal detection performance for the voice signal detection task, mAP reaches not only 97.42%, but also is in maintaining fewer model

收稿日期：2024-05-27
作者简介：李春辉(1997—)，男，河南平舆人，博士生，研究方向为智能信号处理。E-mail:929652284@qq.com

引用格式：*李春辉，向新，杨思力，等.基于时频特征融合和无锚检测机制的高效话音信号检测框架[J].空军工程大学学报，2025，26(3)：26-34. LI Chunhui, XIANG Xin, YANG Sili, et al. An Efficient Voice Signal Detection Framework Based on Time-Frequency Feature Fusion and Anchor-Free Detection Mechanism[J]. Journal of Air Force Engineering University, 2025, 26(3): 26-34.*

parameters and faster inference speed.

Key words　signal detection and recognition；deep learning；network architecture design；task-oriented；feature fusion

通信信号检测识别是认知无线电、干扰检测等应用的关键技术，其任务是从接受到的宽带捕获中检测信号，同时估计信号的类别、中心频率、带宽、持续时间等参数。传统的通信信号检测识别技术主要有能量检测法[1]、匹配滤波法[2]以及基于特征的方法[3]，但是能量检测方法鲁棒性较差，匹配滤波法需要有信号的先验知识，设计有效的信号特征依赖专家知识且难以有效泛化到其他射频数据。

近年来，基于深度学习的方法在通信信号处理领域广泛应用，并在调制识别、信号检测和信道估计等关键技术上展现出优异性能[4-5]。其中基于深度学习的宽带信号检测框架利用时频变换将时间序列信号转换为时频图，进而结合计算机视觉领域中的目标检测算法来完成时频图中的信号检测、识别和时频定位。以往的方法采用了 Faster R-CNN[6-7]、SSD[8-9]、YOLO 系列[10-15]等经典的目标检测网络并进行了一定改进。文献[7]在网络规模上进行缩减，用于检测 WIFI 信号。文献[11～12]对损失函数进行改进以提高坐标回归准确度。文献[13]通过聚类来获得更合适的先验锚框。但是这些方法整体上还是保持原有目标检测算法的网络架构，没有结合信号时频图数据集特点进行针对性地改进，在具体任务上难以达到最优检测性能。比如，时频图中的信号检测相比较自然图像中的通用目标检测任务复杂度更低，原有目标检测算法中的骨干网络难以高效提取信号时频特征。此外，原有算法中用于特征融合的颈部网络是针对特定基准通用目标数据集的特点设计的，不能很好地适配信号数据集分布。信号带宽和持续时间的多样性导致信号目标框长宽比变化明显，使得基于锚框的检测网络面临锚框失配问题，进而导致性能下降。针对以上问题，已有研究[16-17]结合时频图中的信号特性来针对性地改进目标检测网络 CenerNet[18]，在高频信号检测和不连续信号检测任务上取得了更优性能。但是文献[16～17]更关注检测头的设计，对网络整体架构的改进有限。

为了避免在实际数据集上直接迁移使用目标检测算法带来的性能损失，本文提出了一种面向话音信号检测任务的网络架构 SignalNet。首先精简了骨干网络以有效提取信号特征，其次设计了新的颈部网络，利用多尺度时频特征上下文融合以及门控注意力组件来实现高效的特征融合。此外还将有锚检测头网络替换为无锚机制，避免计算先验锚框的同时有利于模型参数缩减以及网络推理。实验结果显示 SignalNet 网络架构在构建的话音信号数据集上有着明显更优的检测性能，取得了 97.42% 的 mAP 值，同时在模型参数和推理时间上也比其他深度学习方法更好。信号检测可视化结果从定性的角度体现了所提网络的检测性能良好。

1　基于深度学习的宽带信号检测框架

图 1 为基于深度学习的宽带信号检测框架。其中无线宽带信号捕获根据任务需求设置中心频率和接收带宽以捕获包含目标信号的复基带 I/Q 信号流。由于是宽带接收机，因此接收信号中还有其他信号以及带内噪声的干扰。

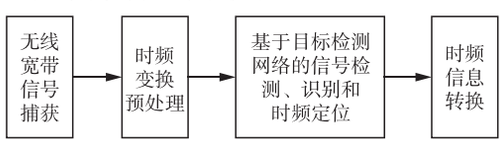


图 1　基于深度学习的宽带信号检测框架

Fig. 1　Deep learning-based framework for the wideband signal detection

时频变换预处理是指使用短时傅里叶变换（short-time Fourier transform，STFT）、小波分析等时频变换将时域信号转换到时频域并生成信号时频图作为深度学习的数据集。信号的表示形式从一维时间序列变为二维时频图像中的前景信号。图 2 为本文所构建话音信号数据集中的一个时频图示例，使用 STFT 生成，黄色背景代表背景噪声干扰，红框内的目标即信号在时频图中的表现形式。

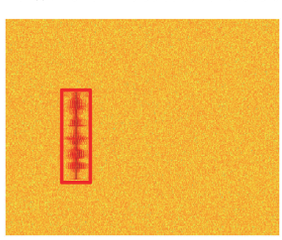


图 2　话音信号时频表示形式

Fig. 2　Time-frequency representation of a voice signal

经过时频变换预处理，时间序列中信号的检测任务就变为时频图中前景信号的检测，这与计算机视觉领域中目标检测任务类似。因此可以使用目标检测网络对时频图中的目标信号进行检测、识别和时频定位。

时频信息转换即根据时频图时间分辨率、频率分辨率将目标检测算法得到的信号框坐标转换为信号起止时间、中心频率、带宽等参数并输出结果。

2 数据集构建

2.1 任务需求

深度学习是数据驱动的，因此构建高质量的数据集是首要且关键的一步。一个领域的数据集大致分为 2 类：一类是大多数算法性能对比的开源基准数据集，比如图像分类领域的 Imagenet[19]、通用目标检测领域的 COCO[20] 数据集；另一类是根据实际任务需求构建的数据集。二者在目标特征、数据规模、分辨率等方面具有不同的特点。

本文面向话音信号检测任务，与之前文献中使用的信号检测仿真数据集相比具有类别单一、目标信号持续时间不定、类内多样性明显等特点。

数据集构建采用真实信号加仿真信号的策略。在数据集中加入仿真信号是因为信号的协议及调制格式已知，在仿真生成信号上训练后的网络权重也可以用于实际信号。此外，仿真生成信号时可以自动化地得到目标的坐标标签，省略人工打标签的繁琐工作。仿真信号生成可以视为一种有效的数据增强方式。

2.2 数据集参数

生成仿真信号首先要录制实际的说话音频信号，采样率 $f_s=200$ kHz，采样精度为 16 bits，录制时长为 4.992 s。对采样音频信号随机截断后以频率调制的方式生成话音信号，这一过程中添加了[0 dB，20 dB]的随机高斯白噪声。

另外考虑到几种时频分析中 STFT 难操作且不会产生交叉项[21]，选择 STFT 作为图 1 中的时频变换预处理，其中 FFT 点数为 4 096，使用 Hamming 窗，窗长 4 096，窗间无重叠。该步骤在 MATLAB 中进行，导出大小为 456×361 的时频图。最终的时频图数据集规模为 8 000 张，按照 8∶1∶1划分为训练集、验证集和测试集。

3 SignalNet 网络架构设计

基于深度学习的通用目标检测近年来快速发展。这些算法可以划分为一阶段以及两阶段目标检测算法，根据是否使用锚框又可以划分为有锚以及无锚的目标检测算法。前期基于深度学习的宽带信号检测算法主要从推理速度和信号处理实时性等方面考虑，因此大多数选择的是一阶段目标检测算法，比如 SSD 以及 YOLO 系列算法。

实际上，很多通用目标检测算法性能优异，不少已经在工业界落地部署[22-23]，也可直接应用在时频图信号检测中。但是由于通用目标和时频图中的信号差异较大，难以在具体任务上达到最优性能。

现代目标检测器将网络设计细化为骨干网络、颈部网络和检测头网络 3 个部分，因此根据实际信号检测任务特点对这 3 个网络进行解耦改进，得到任务导向的时频图信号检测网络 SignalNet。网络整体架构如图 3 所示。

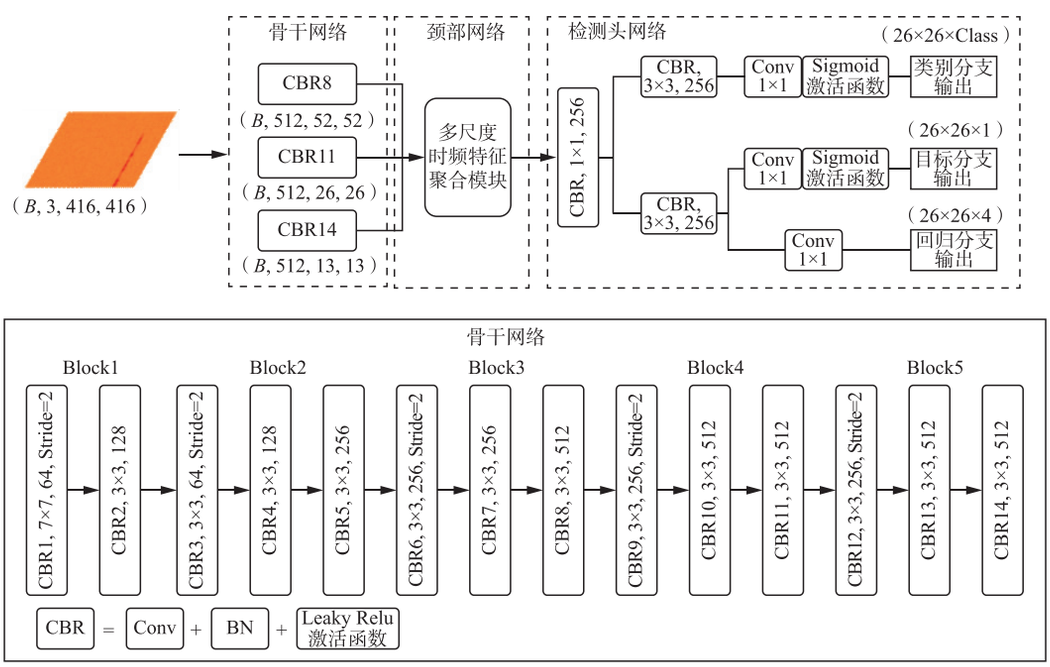


图 3 SignalNet 整体架构及骨干网络

Fig. 3 Overall architecture and backbone network of the SignalNet

$(B,512,52,52)$中的 4 个参数依次代表输入批次大小、特征图的通道数以及特征图的高和宽。CBR 是指由卷积算子(convolution，Conv)、批归一化(batch normalization，BN)[24]和 Leaky Relu(Relu)[25]激活函数串联而成的组件，每个 CBR 组件方框中还描述有卷积核大小以及输出通道数。此外，Stride 和 Class 分别为特征图下采样的步长以及信号类别数。在训练阶段，首先输入批次大小为 B 的 RGB 三通道时频图像，并统一改变尺寸为 416×416，经过骨干网络提取信号时频特征、颈部网络特征融合后在检测头网络得到信号类别、定位回归、目标分数三部分输出。网络输出经过非极大值抑制、时频信息转换等后处理就可以得到信号的类别、起止时间、中心频率及带宽等时频信息。

3.1　骨干网络设计

目标检测算法骨干网络是提取图像特征的关键组件。相比较通用目标检测图像，时频图像背景更简单而且信号内部色彩变化不大，同时该任务数据集中只有一类目标，直接使用已有算法的复杂骨干网络难以高效提取信号特征。因此，设计任务导向的网络宽度和深度不仅有利于后端的颈部网络特征融合和检测头网络分类回归，同时也能缩减网络推理时间、满足信号检测实时性需求。

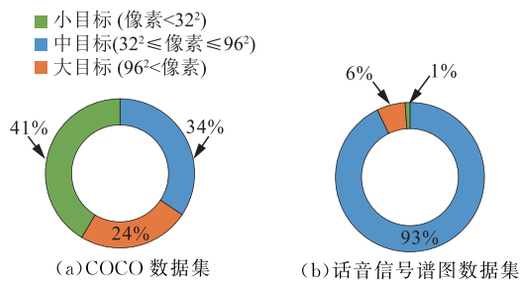
如图 3 所示，所设计骨干网络整体架构分为 5 个卷积块，每个卷积块都是由基础 CBR 组件组成。这样的组件-卷积块设计不仅易于调整网络规模，同时也方便代码复用。

骨干网络包含 14 个 CBR 组件，通过卷积操作不断提高特征图的感受野，随着网络的加深也配置了不同的通道数来得到信号丰富的时频特征。同时在每个块的第 1 个卷积操作中都会设置相应的卷积核大小以及填充参数，使得卷积后的特征图在长宽上变为原来特征图的一半。各个卷积块得到的不同尺度的特征图为下一步特征融合提供支撑。

3.2　颈部网络设计

目标检测算法颈部网络主要实现特征融合。现代目标检测器通常采用特征金字塔结构(feature pyramid networks，FPN)[26]，将不同尺度的目标放到不同特征层次上独立检测同时通过连接使得各层次都有丰富语义信息，这种架构可以方便地集成到现有算法中并有效提高检测性能。

FPN 是针对通用目标检测数据集的特点设计的。如图 4(a)所示，以 COCO[20]数据集为例，按照其大中小目标的定义，COCO 数据集中的目标尺度分布相对均匀。不同尺度大小的目标会被分配到 FPN 的不同预测层[26]。小目标会在高分辨率预测层检出，因为该层具有更准确的定位信息。而大目标则会被分配到低分辨率预测层检出，因为该预测层具有更丰富的语义信息，这也符合感受野匹配的原则。话音信号时频图数据集里中目标占比很大，仍然利用 FPN 会导致目标信号总是由中分辨率输出层进行预测。这使得其他 2 个分辨率的预测层负责的正样本很少，不能得到有效地训练，网络容量也没有得到充分利用，进而影响检测性能。



(a)COCO 数据集　　(b)话音信号谱图数据集

图 4　不同任务数据集的尺度分布

Fig. 4　Scale distributions for different task datasets

因此，有必要设计任务导向的颈部网络。受 FPN 多尺度特征融合的启发，设计了如图 5(左侧虚线框部分)所示的多尺度时频特征融合模块。该模块由多尺度时频特征上下文融合以及门控注意力组件组成。多尺度时频特征上下文融合首先将骨干网络中第 8、11、14 个 CBR 组件后的 52×52、26×26、13×13 这 3 个不同尺度特征图取出，分别经过图 5(中间虚线框部分)所示的空间金字塔池化(spatial pyramid pooling，SPP)[27]模块得到不同尺度上丰富的上下文特征信息。接下来考虑到上述数据集中目标信号尺度分布情况，将 3 个尺度的特征图通过 2 倍上下采样统一到相同分辨率上。此时可以选择加和、连接等操作来直接融合特征图。但是来自不同卷积阶段的特征图存在较大语义差距，简单的融合方式得到的效果会较差。因此在多尺度时频特征上下文融合之后设计了门控注意力组件。如图 5 所示，对于每一个输入特征图 $\boldsymbol{X}_i$，都可以得到一个门控权重 $\boldsymbol{W}_i$，通过相乘加和的方式可以得到最终的融合特征图，如式(1)所示：

$$\begin{cases}\boldsymbol{Y}=\sum_{i=1}^{3}\boldsymbol{W}_i\odot\boldsymbol{X}_i\\ \boldsymbol{W}_i=\sigma(\mathrm{Conv}_{1\times1}(\mathrm{CBR}(\boldsymbol{X}_i)))\end{cases}\tag{1}$$

式中：σ 为 Sigmoid 激活函数；$\odot$ 表示元素相乘，通过注意力门控可以对特征图进行像素级的融合，随着网络训练，门控权重逐渐抑制无用的信息并融合多尺度特征图中有利于后端检测的信息。

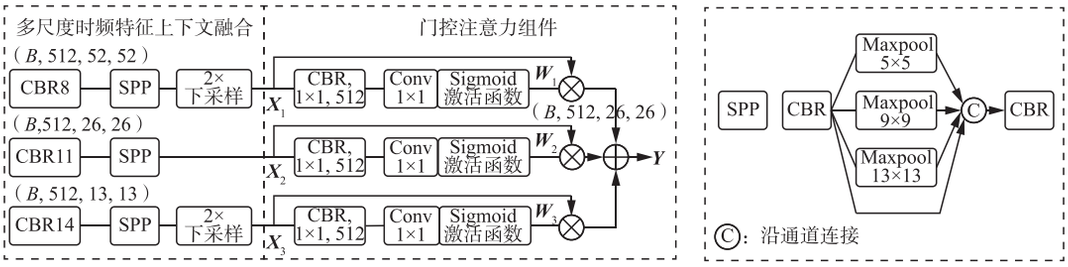


图 5 多尺度时频特征融合模块及 SPP 模块示意图

Fig. 5 Illustration of the multi-scale time-frequency feature fusion module and SPP module

3.3 检测头网络设计

目标检测算法可以分为有锚和无锚两种。锚框是一组先验设置的矩形框，当锚框与真实目标框在大小和长宽比上相似时会有利于目标框的回归，因此锚框的设置一定程度上影响了检测算法性能。文献[10]通过改进 K-means 算法得到更合适锚框，但是这一方面会使得算法时延增加，另一方面在出现前文所述数据集分布问题时难以得到合理的锚框。因此借鉴现代检测器检测头网络[22,28]的设计思路，将有锚机制替换为无锚解码方式加解耦头的结构，其中解耦头是指分类和回归分支并不共用。

无锚输出范式可以分为基于关键点的[18]以及基于特征点的[29]方法。考虑到解码方式的简洁性，选择后一种无锚检测头，这种方法将特征图上的每个特征点视为一个样本。如图 6 所示，在下采样倍数为 16 的 26×26 大小的预测特征图上，每个位置的特征点(x,y)都可以映射回原图上的对应位置$(x_o,y_o)=(16x+8,16y+8)$。假设目标真实框的中心点为(x_c,y_c)，即图 6 中的蓝色点。对于一个特征点，如果它映射回原图后的坐标(x_o,y_o)落在$(x_c-r\times16,\ y_c-r\times16,\ x_c+r\times16,\ y_c+r\times16)$这个方形框内，则定义这个特征点为正样本(图 6 中的绿色点)，否则就是负样本(如图 6 中的黑色点)。其中 r 为超参数，用于控制正样本的定义范围，设置为 1.25。

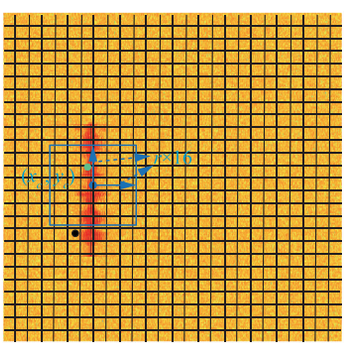


图 6 正负样本分配策略

Fig. 6 Assignment strategy for positive and negative samples

如图 3 所示，在得到颈部网络融合后的特征后，通过 2 个卷积核为 3×3 的 CBR 组件将特征解耦，进而通过 1×1 卷积及 Sigmoid 激活函数得到类别、目标、回归 3 个分支输出。其中类别分支输出 26×26×Class 的特征图用于预测每个特征点上的信号类别。目标分支输出 26×26×1 的特征图，用于额外得到一个预测框的分数[29]来抑制远离中心目标的低质量预测框。回归分支输出 26×26×4 的特征图，用于回归每个特征点到信号包围框 4 个边的垂直距离，进而得到信号包围框坐标。

3.4 损失函数

考虑到检测头输出的设计，损失函数也由 3 个部分组成，如式(2)所示：

$$L_{\mathrm{det}}=\frac{1}{N_{\mathrm{pos}}}\sum_{x,y}L_{\mathrm{cls}}+\frac{1}{N_{\mathrm{pos}}}\sum_{x,y}\Gamma_{\{c_{(x,y)}=1\}}(\lambda L_{\mathrm{reg}}+\mu L_{\mathrm{obj}}) \tag{2}$$

式中：N_{pos} 为正样本点个数；$\Gamma_{\{c_{(x,y)}=1\}}$ 为指示函数，当该特征点是正样本时函数值为 1 否则为 0；λ 和 μ 为超参数，用于平衡损失值；L_{det} 为总的代价函数；L_{cls} 和 L_{obj} 为类别损失和目标损失，分别使用焦点损失函数[30]和二分类交叉熵函数来计算；L_{reg} 为定位回归函数，使用 L1 损失函数来计算。

4 实验结果与分析

4.1 仿真参数设置

实验在 4 块 NVIDIA Tesla K80 GPU 上进行，网络训练了 90 个周期，批大小为 64。采用 SGD 动量优化器，初始学习率为 0.01，在第 70～90 个周期以 0.000 1 线性衰减学习率，同时在前 10 个周期使用了预热训练策略。损失函数中 λ 和 μ 分别设置为 5 和 3。

4.2 评价指标及基准算法

评价指标使用了计算机视觉目标检测领域广泛使用的平均查准率(average precision，AP)以及全类平均精度(mean average precision，mAP)准则。召回率 Recall 和精确率 Precision 定义如下：

$$\mathrm{Precision}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{3}$$

$$\mathrm{Recall}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{4}$$

式中：TP、FN 为实际是正例而预测分别是正例和反例；FP、TN 为实际是反例而预测分别是正例和反例。

召回率和精确率构成的 PR 曲线能够有效反映目标信号检测性能，同时也具有信号检测领域中发现概率和虚警概率的含义。由于数据集中只有一类，所以文中 mAP 和 AP 值相等。

为了验证所提架构的有效性，选择 SSD 算法[9]和改进的 YOLOv3 算法[10]，文献[10]中改进的 YOLOv3 算法，改进点是使用聚类算法优化了先验锚框同时采用了 CIOU[31]损失函数。

4.3　检测性能分析

图 7 显示了所提网络训练过程中的损失函数变化曲线，其中 L_{det}、L_{cls}、L_{reg}、L_{obj} 分别为式(2)中总的代价函数、类别损失函数、时频坐标回归损失函数、目标损失函数。从图 7 可以看到损失函数随着训练不断减小并最终收敛。

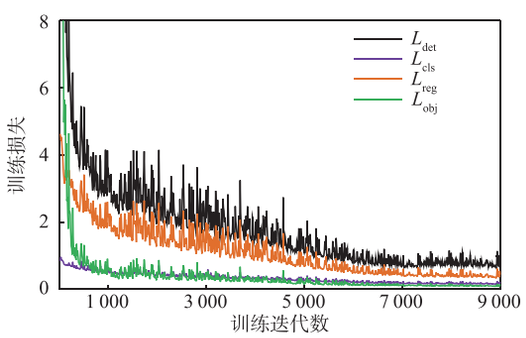


图 7　训练过程损失函数变化曲线

Fig. 7　Curve of loss function value during training

图 8 为网络训练过程中 3 种算法 mAP 值的变化曲线。从图 8 可以看到，在前 30 个周期内 3 种算法的 mAP 快速上升，检测性能大幅提高。但是随着进一步训练，SSD 算法趋于收敛，mAP 值基本不再提高。而 YOLOv3 算法的性能随着后续周期的网络训练仍能改善，mAP 值呈上升趋势并最终收敛。而所提的 SignalNet 网络在训练中的 mAP 值稳步提升，且在训练结束时明显高于其他 2 种算法。

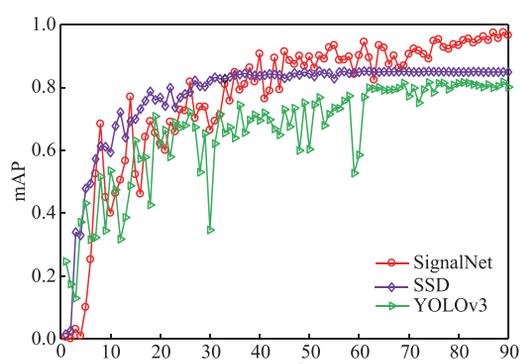


图 8　训练过程 mAP 值变化曲线对比

Fig. 8　Curve of mAP value during training

对比 3 种算法 mAP 数值可以看到 YOLOv3 算法整体上低于其他 2 种算法。SSD 算法在训练周期数较少时性能较优，但是其检测性能存在上限。相比之下所提网络架构经过有效训练后 mAP 明显高于其他算法，检测性能有较大优势。

为了进一步分析 3 种算法的性能差异，以 mAP 为指标在训练过程中保存了各自的最优权重，训练结束后在测试集上进行评估。需要说明的是图 8 是训练过程中在验证集上的评估结果，与测试集没有重叠。

3 种算法的 mAP 值如表 1 所示。从表中可以看到，所提网络在测试集上的 mAP 比 SSD 算法高 12.22%，比 YOLOv3 算法高 15.66%。图 9 为 3 种算法的 PR 曲线。

表 1　3 种算法平均精度对比

Tab. 1　Comparisons of the three algorithms in terms of mAP value

算法	mAP
SignalNet	0.974 2
SSD	0.852 0
YOLOv3	0.817 6

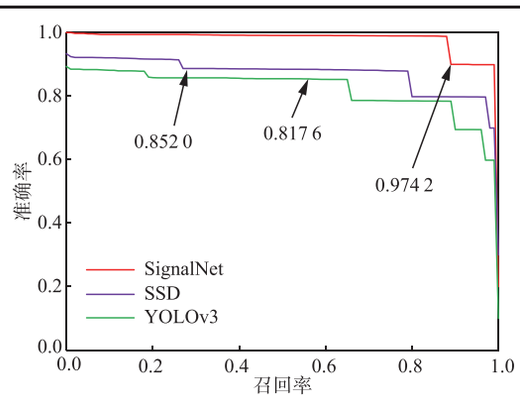


图 9　3 种算法 PR 曲线对比

Fig. 9　Comparisons of the three algorithms in terms of PR curve

PR 曲线越靠近右上方代表检测性能越优，这是由于召回率高代表判定预测框时设置的阈值低，会引入更高的虚警，如果检测算法能够有效判别时频图中信号前景和噪声背景，则 PR 曲线就会接近坐标(1,1)。从图 9 可以看到，所提网络明显优于其他 2 种算法，精确率更高，在检测概率和虚警概率上有着更好的均衡，信号检测性能更优。

结合图 9 以及所提 SignalNet 网络架构改进的出发点可以分析得出，YOLOv3 算法使用的 Darknet53[10]相对于该数据集过于复杂，而 SSD 使用的 VGG16[8]以及所设计的任务导向骨干网络在时频图信号特征提取上更高效，因此 YOLOv3 算法检测

性能略差而且训练过程中 mAP 值有较大波动。另外，所提网络架构通过颈部网络的时频特征上下文融合以及门控注意力组件有效融合了多尺度特征，因此得到比其他 2 种算法明显更优的 mAP 值。

为了分析所提网络在骨干网络精简以及无锚检测头上的优势，对网络模型参数以及推理时间进行对比。其中 3 种算法均在单块 NVIDIA Tesla K80 GPU 上进行推理并计算单张图片所需时间。对比结果如表 2 所示。

表 2 3 种算法复杂度对比

Tab. 2 Comparisons of the three algorithms in terms of complexity

算法	模型参数	推理时间/ms
SignalNet	5.03×10^6	29.4
SSD	20.31×10^6	61.5
YOLOv3	61.52×10^6	132.7

从表 2 中可以看出，所提网络不仅模型参数量明显更少，而且在推理时间上也优于其他 2 种算法。这是因为所提架构在骨干网络规模上更精简，同时结合数据集信号尺度设计了单输出检测头。复杂度上的优势也使得所提网络在实际部署时更有前景。

此外，还对 SignalNet 进行了消融实验，结果如表 3 所示。表中完全实施的 SignalNet 表示模型。SignalNet[1] 表示将骨干网络替换为 Darknet53。SignalNet[2] 表示将所设计的颈部网络替换为 FPN。SignalNet[3] 表示将无锚机制替换为与 YOLOv3 相同的有锚检测头。实验结果验证了 SignalNet 针对骨干网络、颈部网络和检测头所做的任务导向的改进的有效性。具体来说，替换复杂度更高的骨干网络使得网络难以获得有效的特征表示。FPN 在融合信号多尺度时频特征上不如所提的多尺度时频特征融合模块。有锚检测头在检测长宽比变化明显的话音信号时也存在锚框失配等不足，这些都导致模型难以在具体任务上达到最优性能。

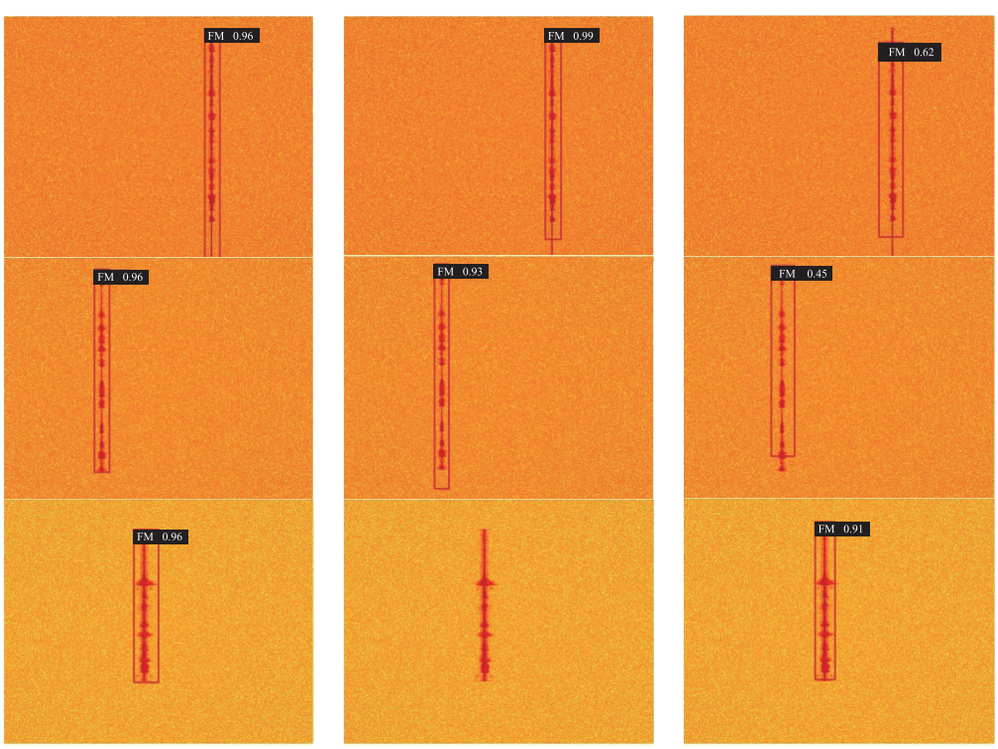
表 3 消融实验结果

Tab. 3 Results of ablation experiments

方法	mAP
SignalNet	0.974 2
SignalNet[1]	0.891 3
SignalNet[2]	0.926 5
SignalNet[3]	0.934 7

4.4 可视化结果

图 10 为可视化后的检测结果，可以定性地分析 3 种算法的检测性能。需要说明的是在制作数据集标签时以话音信号的调制方式作为其类别，即图中显示的 FM。可以清晰地看到 YOLOv3 存在回归不准以及预测置信度低的情况，而 SSD 也有信号包围框不准确以及漏检的不足。相比之下所提网络得到的信号预测框更准确且置信度都在 0.9 以上，检测性能明显更优。



(a) SignalNet (b) SSD (c) YOLOv3

图 10 3 种算法可视化结果对比

Fig. 10 Comparisons of the three algorithms in terms of the visualization results

5　结语

本文提出了一种话音信号检测任务导向的网络架构，通过解耦分别改进了骨干网络以及颈部网络，并将有锚检测头网络替换为无锚机制，不仅有效提取了时频图中信号特征，还利用多尺度时频特征上下文融合以及门控注意力组件有效融合了特征，无锚检测头的机制也使得网络模型参数更少且避免复杂的先验锚框计算过程。仿真结果表明所提架构具有更优的检测性能以及更少的模型参数和推理时间，可视化结果也表明了所提网络具备良好的信号检测性能。

参考文献

[1] CHEN Y F. Improved Energy Detector for Random Signals in Gaussian Noise[J]. IEEE Transactions on Wireless Communications，2010，9(2)：558-563.

[2] THEILER J，FOY B R. Effect of Signal Contamination in Matched-Filter Detection of the Signal on a Cluttered Background[J]. IEEE Geoscience and Remote Sensing Letters，2006，3(1)：98-102.

[3] SALAHDINE F，KAABOUCH N，EL GHAZI H. A Survey on Compressive Sensing Techniques for Cognitive Radio Networks[J]. Physical Communication，2016，20：61-73.

[4] PHAM Q V，NGUYEN N T，HUYNH-THE T，et al. Intelligent Radio Signal Processing：A Survey[J]. IEEE Access，2021，9：83818-83850.

[5] 周静雷，王晓明，李丽敏. 融合注意力机制卷积神经网络的扬声器异常声分类[J]. 西安工程大学学报，2024，38(2)：101-108.
ZHOU J L，WANG X M，LI L M. [J]. Diagnosis of abnormal sound in loudspeakers by Integrated Attention Mechanism Convolutional Nerual Network[J]. Journal of Xi'an Polytechnic University，2024，38(2)：101-108. (in Chinese)

[6] 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)：1137-1149.

[7] PRASAD S V，D'SOUZA K B，BHARGAVA V K. A Downscaled Faster-RCNN Framework for Signal Detection and Time-Frequency Localization in Wideband RF Systems[J]. IEEE Transactions on Wireless Communications，2020，19(7)：4847-4862.

[8] LIU W，ANGUELOV D，ERHAN D，et al. SSD：Single Shot MultiBox Detector[M]//Computer Vision-ECCV 2016. Cham：Springer International Publishing，2016：21-37.

[9] ZHA X，PENG H，QIN X，et al. A Deep Learning Framework for Signal Detection and Modulation Classification[J]. Sensors，2019，19(18)：4042.

[10] LI R D，HU J H，LI S Q，et al. Blind Detection of Communication Signals Based on Improved YOLO3[C]//2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP). Xi'an：IEEE，2021：424-429.

[11] 李润东. 基于深度学习的通信信号智能盲检测与识别技术研究[D]. 成都：电子科技大学，2021.
LI R D. Research on Intelligent Blind Detection and Recognition of Communication Signals Based on Deep Learning[D]. Chengdu：University of Electronic Science and Technology of China，2011. (in Chinese)

[12] 周鑫，何晓新，郑昌文. 基于图像深度学习的无线电信号识别[J]. 通信学报，2019，40(7)：114-125.
ZHOU X，HE X X，ZHENG C W. Radio Signal Recognition Based on Image Deep Learning[J]. Journal on Communications，2019，40(7)：114-125. (in Chinese)

[13] 杨晓乐，付天晖，王永斌. 基于改进型 YOLOV3-Tiny 的通信干扰检测算法[J]. 舰船电子工程，2021，41(2)：60-63，76.
YANG X L，FU T H，WANG Y B. Communication Interference Detection Based on Improved YOLOV3-Tiny Algorithm [J]. Ship Electronic Engineering，2021，41(2)：60-63，76. (in Chinese)

[14] ZHAO R Y，RUAN Y H，LI Y Z. Cooperative Time-Frequency Localization for Wideband Spectrum Sensing with a Lightweight Detector[J]. IEEE Communications Letters，2023，27(7)：1844-1848.

[15] LIN M Y，TIAN Y，ZHANG X X，et al. Parameter Estimation of Frequency-Hopping Signal in UCA Based on Deep Learning and Spatial Time-Frequency Distribution[J]. IEEE Sensors Journal，2023，23(7)：7460-7474.

[16] LI W H，WANG K R，YOU L，et al. A New Deep Learning Framework for HF Signal Detection in Wideband Spectrogram[J]. IEEE Signal Processing Letters，2022，29：1342-1346.

[17] CHENG T，SUN L，ZHANG J N，et al. A Start-Stop Points CenterNet for Wideband Signals Detection and Time-Frequency Localization in Spectrum Sensing [J]. Neural Networks，2024，170：325-336.

[18] ZHOU X Y，WANG D Q，KRÄHENBÜHL P. Objects as Points [EB/OL]. (2019-04-25) [2024-05-01]. https://arxiv.org/abs/1904.07850.

[19] DENG J，DONG W，SOCHER R，et al. ImageNet：A Large-Scale Hierarchical Image Database[C]//2009

IEEE Conference on Computer Vision and Pattern Recognition. Miami, FL: IEEE, 2009: 248-255.

[20] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: Common Objects in Context[C]//Computer Vision-ECCV 2014. Cham: Springer International Publishing, 2014: 740-755.

[21] 卫俊平. 时频分析技术及应用[D]. 西安: 西安电子科技大学, 2005.
WEI J P. Technique and Application of Time-Frequency Analysis[D]. Xi'an: Xidian University, 2005. (in Chinese)

[22] LI C, LI L, JIANG H, et al. YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications[EB/OL]. (2022-09-07)[2024-05-01]. https://arxiv.org/abs/2209.02976.

[23] CARION N, MASSA F, SYNNAEVE G, et al. End-to-End Object Detection with Transformers[C]// Computer Vision-ECCV 2020. Cham: Springer International Publishing, 2020: 213-229.

[24] IOFFE S, SZEGEDY C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[C]// Proceedings of the 32nd International Conference on International Conference on Machine Learning. New York, NY: ACM, 2015: 448-456.

[25] NAIR V, HINTON G E. Rectified Linear Units Improve Restricted Boltzmann Machines[C]// ICML'10: Proceedings of the 27th International Conference on International Conference on Machine Learning. New York, NY: ACM, 2010: 807-814.

[26] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature Pyramid Networks for Object Detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI: IEEE, 2017: 936-944.

[27] BOCHKOVSKIY A, WANG C Y, LIAO H M. YOLOv4: Optimal Speed and Accuracy of Object Detection[EB/OL]. (2020-04-23)[2024-05-01]. https://arxiv.org/abs/2004.10934.

[28] GE Z, LIU S T, WANG F, et al. YOLOX: Exceeding YOLO Series in 2021[EB/OL]. (2021-08-06)[2024-05-01]. https://arxiv.org/abs/2107.08430.

[29] TIAN Z, SHEN C H, CHEN H, et al. FCOS: Fully Convolutional One-Stage Object Detection[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE, 2019: 9010746.

[30] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal Loss for Dense Object Detection[C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE, 2017: 2999-3007.

[31] ZHENG Z H, WANG P, LIU W, et al. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.

(编辑: 徐楠楠)