

视觉跟踪多转子位移测量

杨荣良¹, 王 森¹, 伍 星², 柳小勤¹, 刘 韬¹

(1. 昆明理工大学机电工程学院, 云南 昆明 650500; 2. 云南机电职业技术学院, 云南 昆明 650023)

摘要: 针对当前传统振动传感器在测量旋转体位移时受限于安装和测点数量等问题, 将高速工业相机作为采集媒介, 在转子振动试验台上进行转子振动视频的采集, 并利用基于多目标跟踪的视觉振动测量方法跟踪多个转子目标的全场振动位移。将注意力机制引入残差神经网络, 结合特征金字塔网络结构建立改进的特征提取骨干网络, 并利用身份重新识别方法来强化相邻帧间目标位移的关联性, 跟踪旋转体全场振动位移信号。在转子振动位移测量数据集上对不同网络模型进行定性和定量的比较。结果表明, 本文构建的网络模型在边界框回归时能够获取更为紧密的贴合度; 将采集的电涡流位移信号作为标准量进行两个转子位移信号的对比实验, 结果表明, 本文多目标跟踪算法拟合的波形和频谱噪声最小, 且能与电涡流信号相匹配; 在目标对象模糊情况下的实验也证明本文算法所具有的泛化性能, 这也体现出视觉测量在旋转体振动位移跟踪领域的工程应用价值。

关键词: 视觉测振; 深度学习; 多目标; 视觉跟踪; 模糊图像; 旋转体位移测量

中图分类号: TH113.2⁺1; TH74; TP391.41 **文献标志码:** A **文章编号:** 1004-4523(2024)01-0113-13

DOI: 10.16385/j.cnki.issn.1004-4523.2024.01.012

引 言

旋转机械是工业生产中的基础性设施, 广泛的应用于航空发动机、动量轮、高速分子泵等旋转设备中。这些设备中的旋转体在复杂工况下长期服役后, 时常出现不对称、不对中、磨损和断裂等故障。基于振动信号测量的结构体健康监测技术在有效避免人员伤亡和经济损失方面发挥了不可替代的作用^[1-2]。目前, 基于高频响振动的触觉式加速度传感器测量是一种常用的旋转体故障检测方法, 但附加质量效应和无法直接吸附被测体而引起的误差必然会影响测量的精度。因此, 基于非接触式的位移测量传感器在特定的工业场景发挥了独特的检测优势, 如利用电磁感应原理测量探头与金属导体之间相对位移变化的电涡流传感器。列举的两种传感器只能独立测量单点的一维信号, 而多点甚至多目标测量时只能依靠数量弥补其缺陷。因此, 研究一种非接触式的多振点、全场测量方法进行可靠的结构体健康监测具有重要的经济价值和意义。

作为一种远距离、非接触、多维的振动位移测量方法, 视觉测量逐步得到广大专家和学者的认可。该方法根据提取的目标特征选择合适的数字图像相关算法, 获取视频中既定目标的类别和位置信息, 通

过映射目标在图像中的像素值回归出目标指定点的位移偏移量信息。但是大部分基于数字图像相关的视觉测量方法因本身固有的缺陷而导致其在目标识别和位移关联方面陷入理论和技术方面的瓶颈。如工业相机在振动视频采集过程中会产生畸变效应, 这种不可避免的物理缺陷会放大目标产生未知几何变形的程度; 基于灰度^[3]、形状^[4]、颜色^[5]和梯度^[6]特征进行视觉测量算法也会因少量的特征参数无法对旋转体有效识别, 尤其在面临高频、微幅和无表面处理的振动测量时略有不足。因此, 尽管视觉测量位移的方法具有较强的适用性, 但是如何增强目标识别和振动位移之间的联系是实现视觉位移可靠性测量的关键所在。

视觉位移测量可根据应用场景、结构对象选择基于序列帧的视频跟踪方法。视频跟踪方法纳入帧间相对运动信息, 将时域信息和空间信息相互融合。与常规图像检测方法相比, 视频追踪方法可以在振动信号位移的处理速度和检测精度之间进行有效平衡, 如基于时空上下文^[7]、支持向量机模糊融合^[8]和遗传算法^[9]等相关搜索算法。而基于亮度不变和小范围运动假定, 用于估计局部位移的速度场的光流法^[10]仍无法解决特征点数量区间不确定的难题。也有学者尝试采用双目视觉^[11]、模式匹配^[12]、远程智能感知^[13]等方法测量结构的位移和几何变形。

收稿日期: 2022-04-06; **修订日期:** 2022-07-24

基金项目: 国家自然科学基金资助项目(52065035); 云南省重大科技专项计划项目(202102AC080002)。

近期,基于深度学习的视觉检测方法开始应用于结构体的位移测量,深度神经网络开始逐渐在视觉振动测量领域展现出普适性优势,但算法仍然存在部分瑕疵。如利用Mask-RCNN网络测量结构体位移偏移量时^[14],像素级分割方法会存在固有的耗时缺陷;将多分辨率深度学习特征框架用于桥梁振动位移监测时^[15],常规卷积中的规格采样会导致网络难以适应几何形变^[16]。因此,尝试将模板匹配、光流法、视觉增强^[17]等算法与深度学习算法相结合^[18],是提高位移测量精度的另外一条主线,或者结合长短期记忆-递归神经网络(LSTM-RNN)和多目标学习^[19]进行振动信号分量预测。除此之外,利用生成对抗网络重建高分辨率边缘信息^[20]以提高人工标志物的清晰度也同样有利于提高振动测量的准确度,但这些方法仍未考虑测量多目标振动信息时因目标相似而发生信息切换的问题。与传统目标检测算法比较,深度学习方法可以自动学习包含上万参数的特征表示,同时深度模型使学习过程效率更高。

虽然深度学习方法在检测精度上具有独特的优势,但是在选定振动对象进行算法框架和模型的迁移、微调和推理过程中仍然存在一些亟待解决的问题。如在考虑检测精度的同时需考虑时间成本,检测的目的在于目标的存在,而振动测量是计算目标位移偏移量;检测模型回归边界框只注重是否覆盖振动目标而非边界框的贴合度;在多目标表现特征相似时,目标身份切换也会引发振动信号混乱;另外,除考虑工业相机受采样定理约束而引发丢帧现象外,高帧率下的目标检测也会因算法缺乏时空关联性而产生“短时失忆”,这些都会导致部分时间节点内的位移偏移量缺失。

因此,本文将转子作为视觉测量的研究对象,以高速工业相机作为采集媒介,利用设计的卷积神经网络模型跟踪视频中多个旋转体目标中心点的振动位移,与相对应电涡流传感器采集的振动信号进行对比,实验结果验证了视觉测量旋转体振动信号的可靠性。

1 多目标视觉跟踪方法

本文在基于先检测后跟踪网络框架Fair-MOT^[21]基础上,将残差网络(ResNet)与注意力机制(SE)相结合,强化学习多目标边缘信息;引入特征金字塔(FPN)组合成骨干网络,精细化局部信息和轮廓全局信息;通过无锚框的检测算法估算旋转体图像的热力映射特征、对象中心偏移量和边

界框大小,并利用身份重新识别(ReID)方法强化旋转体在图像帧之间的位移时空关联性,最后根据视频中每一帧目标的中心点坐标回归出旋转位移偏移量曲线。本文方法的整体流程如图1所示。

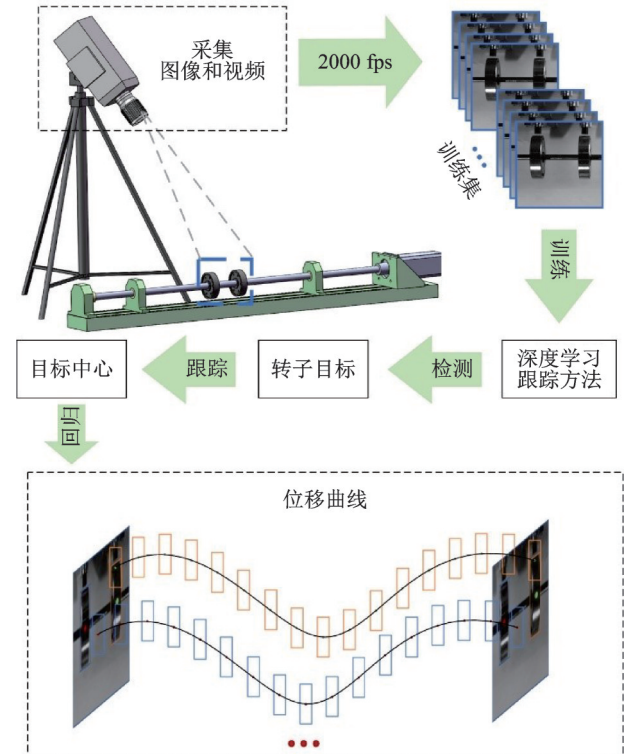


图1 视觉测量振动方法

Fig. 1 Visual measurement of vibration method

1.1 旋转体图像目标检测

目标检测通常是指在图像中利用边界框检测出待识别对象的类别和位置。以本文实验对象转子为例,基于关键点的无锚框算法CenterNet^[22]更加强调获取目标的形状结构和位置特征,对规则的旋转体进行无锚框检测无疑是改善检测精度最好的方式之一。CenterNet通过“目标即是点”的原则,将目标检测问题转变为标准关键点的估计问题,也就是将转子目标检测表达为整个转子中心点的检测。利用真实框中代表转子目标左上和右下的坐标值,每个关键点可表示为:

$$(c_x^i, c_y^i) = \left(\frac{x_1^i + x_2^i}{2}, \frac{y_1^i + y_2^i}{2} \right) \quad (1)$$

式中 x_1^i, y_1^i 代表边界框左上角坐标值; x_2^i, y_2^i 代表边界框右下角坐标值; (c_x^i, c_y^i) 代表目标中心点。

为实现转子的多目标识别并分别估计各个目标的中心位置,CenterNet分别针对图像中每个转子生成单独的热映射图(Hm)。四倍下采样后转子中心点 $(\hat{c}_x^i, \hat{c}_y^i)$ 在特征图中可表示为:

$$(\hat{c}_x^i, \hat{c}_y^i) = \left(\left\lfloor \frac{c_x^i}{4} \right\rfloor, \left\lfloor \frac{c_y^i}{4} \right\rfloor \right) \quad (2)$$

转子目标图像传入骨干网络后输出一幅包含 N 个目标的热力图,该热力图中所属不同转子目标的峰值点即转子目标中心点。计算中心点的热力图公式和损失函数可分别表示为:

$$M_{xy} = \sum_{i=1}^N e^{-\frac{(x-\hat{c}_x^i)^2 + (y-\hat{c}_y^i)^2}{2\sigma_c^2}} \quad (3)$$

$$L_{Hm} = -\frac{1}{N} \sum_{xy} \begin{cases} (1-\tilde{M}_{xy})^\alpha \lg(\tilde{M}_{xy}), M_{xy}=1 \\ (1-M_{xy})^\beta (\tilde{M}_{xy})^\alpha \lg(1-\tilde{M}_{xy}), \text{其他} \end{cases} \quad (4)$$

式中 σ_c 表示标准差, \tilde{M}_{xy} 代表预估的热力图,根据先验知识^[23]设置损失函数预定参数 $\alpha=2$ 和 $\beta=4$ 。

输入转子图像和输出特征图之间存在的采样跨度会使映射终端产生转子目标偏移,即真实边界框与热力图计算中心点存在偏移。为减轻因上采样和降采样操作产生的量化误差,本文通过计算热力图中心点并推理每个位置目标框的宽高,每个像素估计相对于转子目标中心的连续偏移 O^i 、预测偏移 \hat{O}^i 的偏置估计损失 L_{off} 和边界框大小 S^i 可分别表示为:

$$O^i = \left(\frac{c_x^i}{4}, \frac{c_y^i}{4} \right) - \left(\left\lfloor \frac{c_x^i}{4} \right\rfloor, \left\lfloor \frac{c_y^i}{4} \right\rfloor \right) \quad (5)$$

$$L_{off} = \|O^i - \hat{O}^i\|_1 \quad (6)$$

$$S^i = (x_2^i - x_1^i, y_2^i - y_1^i) \quad (7)$$

与预测边界框大小 \hat{S}^i 进行比较,转子目标边界

框的宽高损失 L_{wh} 可表示为:

$$L_{wh} = \|S^i - \hat{S}^i\|_1 \quad (8)$$

本文通过分类任务学习身份重新识别(ReID)功能,为保证跟踪模块准确提取转子目标的表观特征,利用特征更新算法从骨干网络中生成 128 维度特征的实例嵌入(Embedding)。其中,嵌入特征图 $E=R^{128 \times W \times H}$ 可从以 $(\hat{c}_x^i, \hat{c}_y^i)$ 为中心的特征图中提取得到(R 表示特征图大小)。将提取的身份重新识别(ReID)特征向量,学习并映射到类分布向量 $P = \{p(k), k\}$, 实现 ReID 的 k 类提取。将正确 GT 类标签表示为 $L^i(K)$, ReID 损失可表示为:

$$L_{reid} = -\sum_{i=1}^N \sum_{k=1}^K L^i(k) \lg p(k) \quad (9)$$

无锚框检测算法不需要在疑似目标的位置穷举出大量的预测框,这样可以节省大量的推理时间。最为重要的是,转子的振动位移跟踪即是对旋转体中心点的位置跟踪,这正好与无锚框检测目标关键点的方式如出一辙,同时加入身份重新识别(ReID)保证目标时空一致性,所以在旋转体目标跟踪领域有着精度、速度和算法匹配度上的绝对优势。整个网络结构如图 2 所示, Conv_X 将注意力机制(SE)融入在残差结构(ResNet)中组合成新的残差结构;并将每一残差网络层输出特征作为特征金字塔网络(FPN)输入,经过金字塔网络卷积层下采样(C1, C2, C3, C4)、上采样和特征融合输出特征层 P_1 ;其中 P_1 包括上述 Hm, Offset, WH 和 ReID 所需特征,并据此预测对应转子目标的位置、形状和身份等信息。

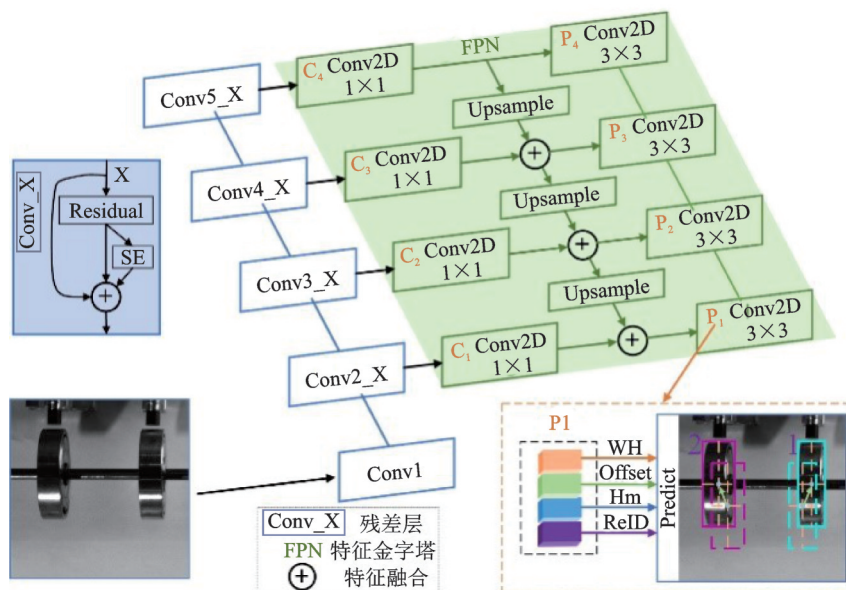


图 2 网络整体结构

Fig. 2 The overall structure of the network

1.1.1 特征提取模块

鉴于转子目标结构特征明显且算法需满足精度和速度要求, 本文将较为浅层的特征提取网络 ResNet18^[24] 结合注意力机制(SE)和特征金字塔(FPN)作为骨干网络, 以提取高速转子振动试验台的转子特征。图 2 中 Conv2_X 至 Conv5_X 分别表示数目为 2 并结合 SE 的残差卷积(Conv3×3)层。Res34 网络 Conv2_X 至 Conv5_X 分别表示数目为 3, 4, 6, 3 的残差卷积网络。同一残差结构层可通过捷径连接和元素相加执行图 3(a) 的同等维度映射, 不同残差结构层可由图 3(b) 的不同维度映射连接。这种跳跃连接方式不会增加额外的参数运算和计算复杂度。两种计算方式分别表示为:

$$y = F(x, W_a) + x \quad (10)$$

$$y = F(x, W_b) + W_i x \quad (11)$$

式中 W_a 和 W_b 分别表示不同层的权重系数。

1.1.2 特征金字塔模块

常规的目标检测算法仅采用顶层特征进行高层语义特征预测, 这种高尺度的全局定位只能粗略地获取转子目标的轮廓位置, 而且多次的降采样和上采样操作使深层网络的定位信息存在逐尺度递增的偏移误差。采用自底向上的金字塔多尺度特征融合方式将处理过的低层特征和高层特征进行累加, 可以综合获取转子目标的精细化局部信息和轮廓全局信息, 提高目标检测的鲁棒性和定位锚框的贴合度。

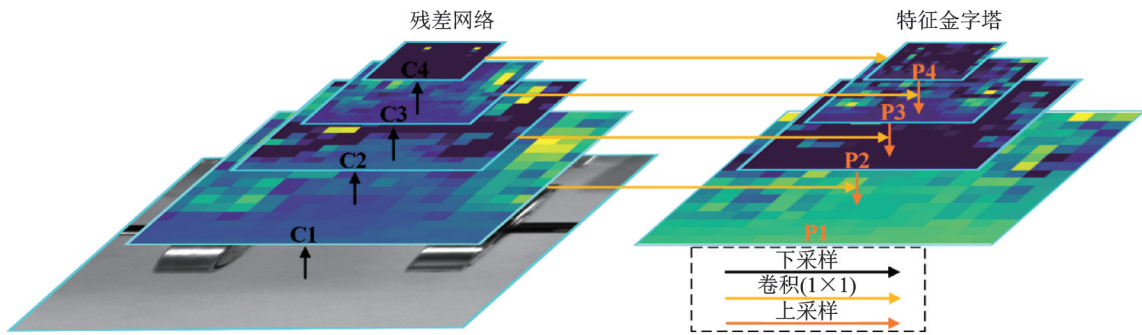


图 4 特征金字塔模块

Fig. 4 Feature pyramid networks module

1.1.3 注意力机制

在检测骨干网络残差结构中加入 SE^[26-27] 注意力机制不仅能够提高网络通道之间的关联性, 而且能够促使模型主动锁定不同通道特征的优先程度。如图 5 所示, 注意力机制首先在输入特征图层上执行全局池化(GAP), 压缩当前特征图的全局特征量; 再通过两层全连接结构激发得到特征图中每个通道的权值, 并将加权(Scale)后的特征图作为下一层网络的输入。图 2 中改进残差结构主要通过加入注意力机制以提高骨干网络在边缘特

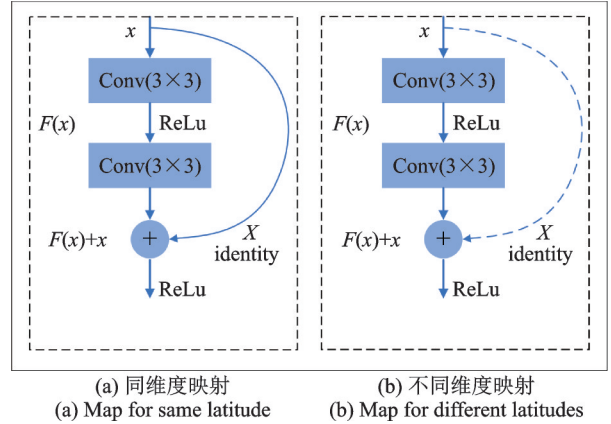


图 3 残差结构的不同连接方式

Fig. 3 Different ways of connecting the residual structure

因此, 构建一个更深层的特征金字塔 (Feature Pyramid Networks, FPN)^[25] 可以有效融合多层特征信息, 提高目标跟踪环节中转子位移偏移量预测的准确性。以图 4 中构建的 4 层 FPN 架构为例, 残差网络提取的转子目标特征通过不同的卷积层下采样 (Downsample) 后可以生成 4 个不同尺度的特征 (C1, C2, C3, C4)。而 FPN 则是通过尺寸为 1×1 的卷积 C4 侧向连接生成的特征层 P4, 经反卷积上采样 (Upsample) 后将其与 1×1 逐点卷积后的 C3 相加生成 P3, 利用同样的运算方式可以得到 P2 和 P1。P3 的计算公式可表示为:

$$P_3 = Conv(1, 1) \odot C_3 + [P_4] \quad (12)$$

征提取的能力, 有效应对转子微幅振动时不明显的边缘变化, 因此能够在位移提取时获得更好的位移信号。

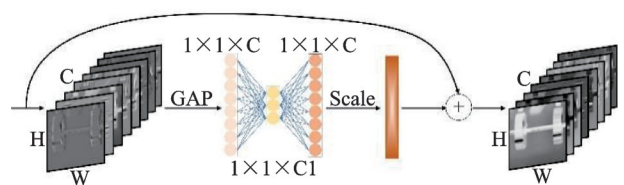


图 5 注意力机制

Fig. 5 Attention mechanism

1.2 旋转体目标帧间匹配与跟踪

提高转子图像中检测目标的匹配相似度和定位精度可以通过跟踪分支强化旋转体图像帧间的位移时空关联性来实现。如图 6 中的跟踪流程图所示,跟踪分支包含边界框(Bbox)、置信度(Confidence,取值 0.6)、身份重新识别(ReID)嵌入的转子特征(Embeddings)。其中,置信度对检测器生成的转子边界框进行筛选,将设定之外的目标框剔除,边界框和 Embeddings 则用于跟踪分支的匹配计算。本文将检测分支并行输出作为跟踪分支的输入,利用跟踪器对第一帧轨迹进行初始化操作。首先利用卡尔曼滤波器进行转子目标的身份预测、级联匹配和迭代更新。其次,级联匹配根据不同优先级将赋予每个转子目标一个单独的跟踪器,即确定目标身份并保持至下一帧。上一帧最先匹配的跟踪器具有匹配的优先权,而未完成匹配的跟踪器则降低优先权。

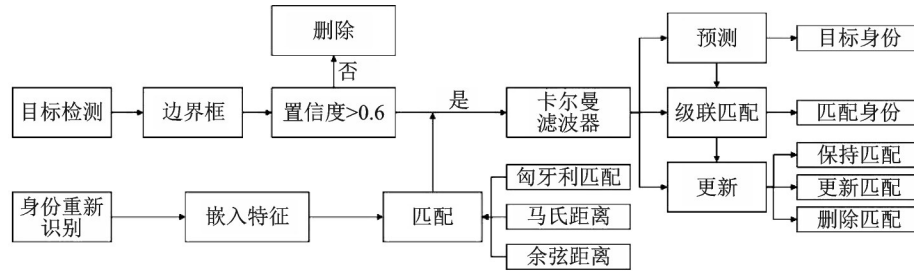


图 6 跟踪流程

Fig. 6 Track process

1.3 旋转振动体位移提取

旋转体视觉振动位移测量本质上是获取振动视频中每一帧每个转子目标的中心点坐标。本文分别提取视频测试帧图像中每个转子目标的中心点,以转子目标在视频第一帧中心点的位置信息作为计算位移偏移量的参考帧,计算公式为:

$$\begin{cases} D1_n = (X_{cn}^1 - X_{c0}^1, Y_{cn}^1 - Y_{c0}^1) \\ D2_n = (X_{cn}^2 - X_{c0}^2, Y_{cn}^2 - Y_{c0}^2) \end{cases} \quad (13)$$

式中 (X_{cn}^1, Y_{cn}^1) 和 (X_{cn}^2, Y_{cn}^2) 代表振动视频中第 n 帧转子 1 和转子 2 中心点坐标, $D1_n$ 和 $D2_n$ 分别代表转子 1 和转子 2 在第 n 帧的振动位移偏移量。

以公式(13)的计算方式分别回归出所有视频帧各转子目标中心点在像素坐标系内的振动位移偏移量,最后分别对其数据归一化处理。振动位移时域信号提取的流程如图 7 所示。图中 T_n 代表视频序列帧。

2 实验

为了验证视觉测量对旋转体振动位移跟踪的可

卡尔曼滤波器根据当前状态进行下一帧目标中心点的状态预测,再将预测结果与测量结果进行迭代更新,若完成匹配和更新,则达到跟踪转子目标的目的。

更新模块部分包括匹配、追踪更新和特征更新。匈牙利匹配算法是一种完成目标与跟踪器匹配的常用算法,但该算法为了完成尽可能多的匹配任务而极易陷入目标身份切换。因此,本文针对运动信息和外观信息的计算特点,利用马氏距离(Mahalanobis)和余弦距离(Cosine)比较目标与跟踪器的相似度,将目标与跟踪器的交并比作为置信度的高低,并通过优化度量方式实现更优匹配。追踪更新包含三个阶段:已被匹配的转子目标在当前帧继续保持匹配状态,并利用卡尔曼滤波器进行更新;未匹配的转子目标将更新一个新的跟踪器进行目标匹配;完全未匹配到目标的跟踪器则直接删除。

靠性,本文将常见的转子作为实验对象构建高转速下的转子振动位移测量数据集,并利用该数据集进行不同网络模型的训练和测试比较。本文将同步采集的电涡流位移信号作为标准位移偏移量,对不同深度学习算法视觉位移测量结果进行比较。同时,在目标对象模糊情况下的实验,验证本文算法所具有的泛化性能。

2.1 实验设置和设备

由于旋转体本身难以直接使用接触式传感器测量振动信号,因此实验特意选用同样是非接触式采集的电涡流传感器进行振动信号的采集。由于电涡流采集信号属于连续时间域的近距离振动响应测量,局部振点的测量精度相对较高,因此本文将作为参照的标准振动信号评价视觉振动测量算法的可行性,将其同步采集的多通道局部位移信号进行离散化处理,与基于图像的离散时间域全域空间测量进行多个振点的一一对应。本节将高速工业相机(千眼狼 5F01,分辨率为 512×16 帧率最高可达到 128000 fps)和电涡流传感器(上海欧多-9000, Scale

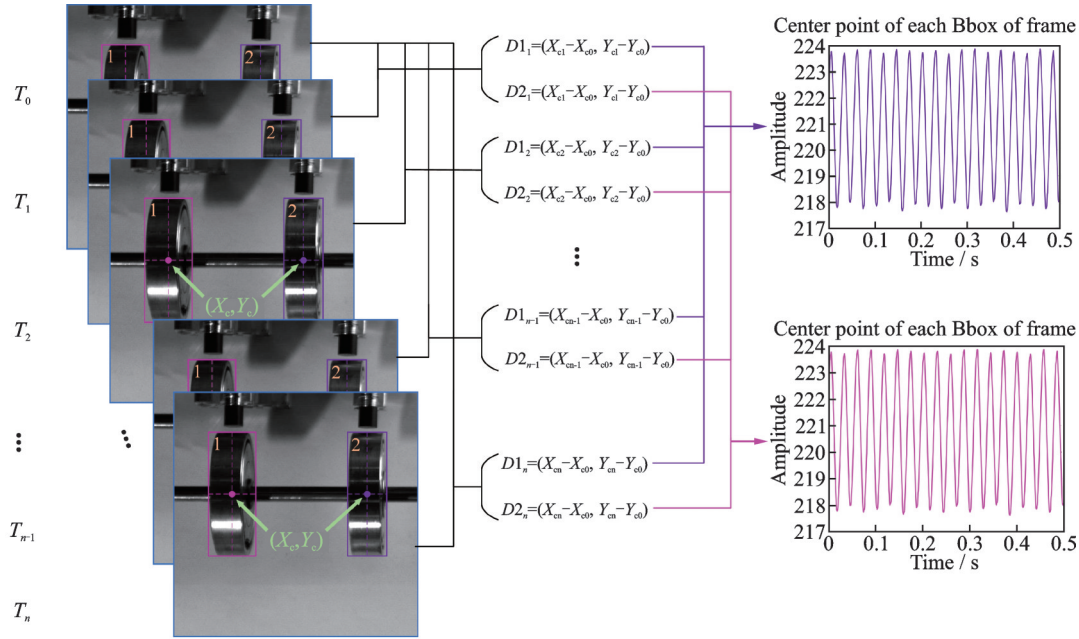


图7 位移信号提取流程

Fig. 7 Displacement signal extraction process

Factory 为 2.5 V/mm, VT 为 15 V) 两种不同采集原理的传感器作为旋转体位移采集的计量器, 同步采集高速转子振动试验台(南京东大 Z-03) 上转子的振动视频和电压位移信号。为了使转子在旋转过程中产生轻微的振动, 在转子的左侧安装了一个紧固螺钉以诱发转子产生不对中现象。光源(金贝 EF-200LED) 明暗度的适当调节可以有效增加图像画面的对比度, 高速相机采集的图像序列以设定的帧率直接存储在自带的设备中。实验时将信号采集卡(NI-9234) 的两端分别与笔记本电脑(Honor Magic book pro) 和电涡流传感器相连接, 利用编译的同步采样模块采集设定时间段内的图像信息和电涡流位移信息。通过模拟台调速器调节不同的电压值, 保证转子转速一定。在振动位移数据采集前, 实验设定高速相机的镜头与转子的垂直距离为 1500 mm, 光源与相机呈 45° 摆放, 并采集不同方向标定板图像对相机参数进行标定; 电涡流传感器的探头方向与转子的水平中心线重合, 两者之间的距离为 2.5 mm。整套实验设备的真实现场图如图 8 所示。设置高速相机采集图像的分辨率大小为 512×512, 像素精度为 0.125 mm/pixel, 帧率为 2000 Fps, 同时设置电涡流传感器的信号采样率为 2000 Hz; 两个电涡流传感器分别测量两个转子振动位移信息。因此, 实验在 5 s 内获取转速为 24 r/s 下的转子图像共 10000 序列帧; 此外, 单独采集同种工况下图像 1000 帧作为训练集, 并利用标注软件 YOLO-mark 制作 GT。因为转子目标在图像序列中目标大小不变, 为减小人工标注误差, 使用统一的尺寸框进行人工标注。

本文的所有实验均在台式机(Intel(R) Core

(TM) i5-10600KF CPU @ 4.10 GHz, 16G 内存, GPU 为 GeForce RTX 2080 super, GPU 显存为 8G) 上进行, 所有深度学习算法全部配置统一的运行环境(Ubuntu18.04, Cuda10.2, Pytorch1.7, torchvision0.8)。

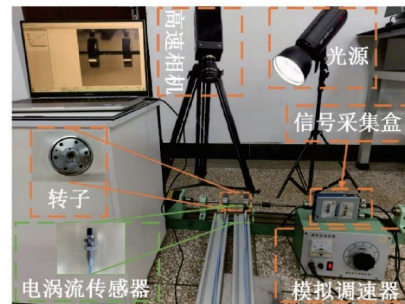


图8 实验设备现场图

Fig. 8 Experimental equipment site map

2.2 模型测试

本次实验共标注顺序帧 300 幅图像, 并在不同骨干网络中各训练 1000 个 epoch, 实验使用多目标转子数据集进行网络模型目标分类精度与定位精度的训练、验证与测试。本文使用的损失函数 Total_loss 如下式所示:

$$T_{Loss} = L_{Hm} + L_{WH} + L_{Off} + L_{ReID} \quad (14)$$

式中 L_{Hm} 表示目标中心点损失, L_{WH} 表示目标边界框宽高损失, L_{Off} 表示目标中心偏置估计损失, L_{ReID} 为目标身份损失。总损失的变化体现算法学习的速度, 总损失值的大小体现了真实值与网络模型预测结果的差异大小。

原始残差网络(Res18)、结合注意力机制的残差网络(Res18_SE)、结合特征金字塔的残差网络(Res18_FPN)、更深层残差网络(Res34)、以及本文网络(结合注意力机制和特征金字塔的残差网络)在进行 1000 次 epoch 训练后获得的总损失变化曲线如图 9 所示。从图 9 中可以看出本文使用算法的 T_{Loss} 下降较快,且损失值小于其他的算法。这也表明本文算法能够更快更好地学习转子表观特征,因此能够在跟踪振动位移时获得更精准的位移信号。深度学习算法对转子目标进行检测的准确性可以通过不同网络对转子预测的 Bbox 贴合度来反映。利用不同算法对图像序列帧中转子推理测试后发现,深度学习方法确实能够实现有效的转子目标检测与跟踪。

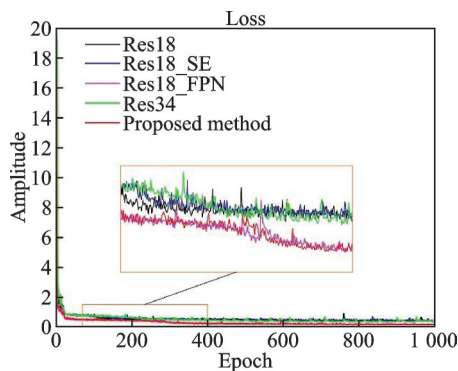


图 9 各算法总损失值下降曲线对比

Fig. 9 Comparison of loss descent curves of each algorithm

以随机挑选图像数据集中的三幅图像为例,本节分别将 Template matching(模板匹配), Res18, Res_SE, Res_FPN, Res34 和本文网络共 6 种方法的跟踪结果列举在图 10 中,并逐个与对应的人工标注结果进行定性比较。从图 10 中可以看出,所有视觉位移测量方法回归的 Bbox 均能实现目标跟踪,各种算法回归的目标区域与人工标注极度相似。从图像目标检测和追踪的角度来看,转子目标的检测已经完全符合目标完整度的要求,但是从振动位移测量的角度来说,这种完整度的测量标准还远远达不到位移测量精确度的严格要求,需要在位移信号曲线拟合度的实验中进行更加详实的实验证明。在图中通过橙色(多框选)、绿色矩形框(欠框选)对不同算法检测两个转子(图像中左右转子分别称为转子 1, 2)的结果进行误差区域标记,可以看出,Res18 网络回归转子 1, 2 的边界框在四个方向上都存在空白区域,说明了此算法并不能充分学习转子边缘特征,即不能精细化回归转子目标边界框。Res18_SE 网络回归转子 1 的边界框在上下两个方向上存在欠框选区域,而在左右两个方向区域拟合较好,回归转子 2

的边界框时贴合紧密。这也表明加入 SE 注意力机制能够有效学习转子的边缘信息,并回归更紧密的边界框。Res18_FPN 检测结果与人工标注的标准框相比,虽然转子 1 边界框存在空白区域,而转子 2 存在欠框选区域,但相比于原网络 Res18 较准确的边界框,加入 FPN 可以有效融合高低层语义信息,提高预测目标的鲁棒性。更深层的 Res34 网络回归转子边界框也存在多框选的情况。与人工标注标准框再次比较时可以发现,结合 FPN, SE 后本文目标跟踪算法回归的边界框能够获取更为紧实的贴合度。根据表 1 中最小损失(MIN_{loss})、权重大小(Weight)、每秒跟踪帧数(Fps)和均值平均精度(MAP)的定量比较结果可知,虽然本文算法的 Fps 略小于 Weight 较小的算法,但在权衡速度和精度前提下,本文算法速度仍高于传统模板匹配算法和较深网络 Res34。结合各网络回归两个转子边界实例可以看出,图像中转子 2 有更规整的表观特征,从而在预测时更易贴合目标边界。

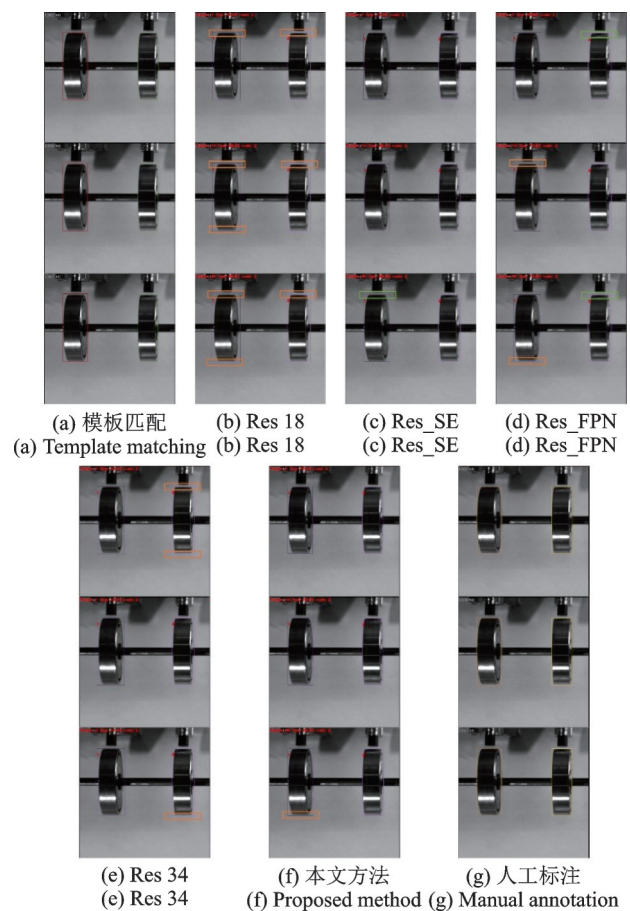


图 10 不同算法回归实例对比

Fig. 10 Comparison of regression examples of different algorithms

2.3 不同传感器信号对比

图 11 和图 12 分别为不同视觉振动测量算法与

表1 视觉传感器各方法定量比较

Tab. 1 Quantitative comparison of visual sensor methods

方法	MIN _{loss}	Weight/MB	Fps	MAP
Template matching	-	-	40.96	1
Res18	0.36851	176.7	90.68	1
Res18_SE	0.32698	176.8	90.40	1
Res18_FPN	0.12864	200.6	71.11	1
Res34	0.28991	298.1	62.94	1
本文算法	0.12528	200.7	71.06	1

电涡流采集信号进行两个转子目标同步比较的时域图。如图所示,除模板匹配外的其他视觉测量算法均在回归的位移轨迹上与电涡流采集的类正弦信号相似,且表现出的近似稳定的周期特性,但不同视觉测量算法获取的位移信号中却存在噪声较大和幅值不对应等问题。从图11和12中显示的结果中可以看出,模板匹配算法得到的结果是像素级位移,因此在回归转子位移曲线时导致所属曲线上的采样点在近似周期内产生频繁或间歇性的阶跃跳动。在图11中,Res18算法在回归转子1的位移信号时存在较大噪声且峰底幅值偏低,很难与电涡流位移信号相拟合,此类异常的原因可以归结于第2.2节Res18回归

转子1边界框时存在空白区域,在表2中峰值也有很好的体现。Res18_SE也因欠框选原因使峰底信号丢失,从而发生信号局部突变现象。Res18_FPN回归的整段信号虽然在幅值和周期上与电涡流信号差异不大,但信号存在较多噪点。Res34网络在峰底时存在信号堆叠,存在一定噪声;而本文算法回归转子1位移信号相比于其他视觉测量方法噪声较小。Res18算法在跟踪转子2的位移时,位移曲线峰底出现信号堆叠且整体存在较多噪声点。由表2的定量分析结果也可以发现,其峰峰值、有效值与标准电涡流值相差较大。图12中Res18_SE整体位移信号噪声较小且在表2中各类值回归均准确,但峰顶信号缺失的现象与转子检测时出现的欠框选区域有极大的关联。Res18_FPN回归的位移曲线要比上述两种算法平滑,但是与电涡流信号相比仍然存在一定的噪声。与较深网络Res34回归转子2时域信号相比,本文算法噪声小,且曲线能与电涡流曲线较好拟合。由图11中回归转子1的位移曲线可知,由于实验时高速工业相机拍摄转子1时引入过多无关信息,因此所有视觉算法在拟合曲线上升和下降阶段存在一定差异,且峰值点与电涡流到达时间也不一致。由图12中回归转子2位移曲线可知,高速工业相机拍摄存

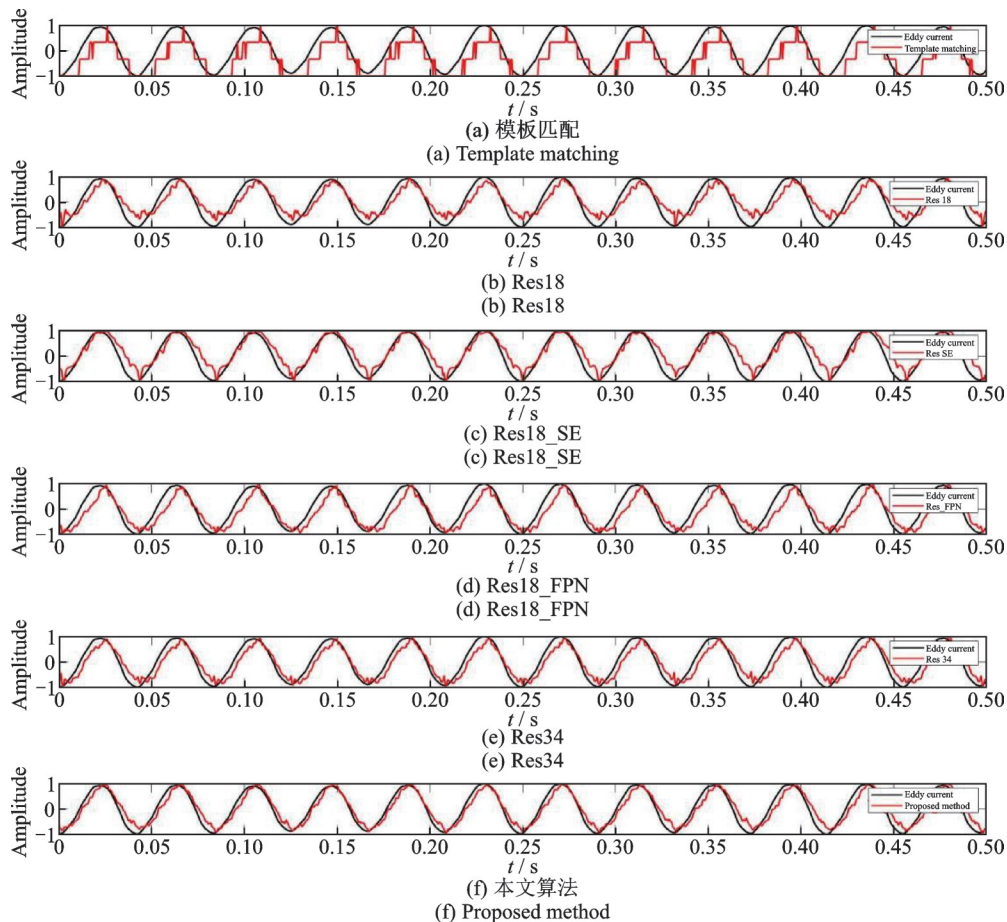


图11 各算法回归转子1位移曲线与电涡流信号对比

Fig. 11 The algorithm regression rotor 1 displacement curve is compared with the eddy current signal

在些许角度偏移导致拟合位移曲线下降阶段滞后电涡流传感器信号。对比图 11 和图 12, 转子 1 相比于转子 2 视觉回归位移信号存在一定的噪声, 证明视觉测量位移拟合会因表观特征造成差异。即使视觉测量方法在检测转子目标时精度都高达 99% 以上, 但并不意味着这些算法具有较好的位移测量性能。由于目标转子属于特征明显的规则结构体, 而网络模型的训练会在很短的训练频次内出现数据快速拟合现象, 这就导致总损失值在较长的一段时间内收敛速度降慢。虽然网络模型的检测精度很高 (MAP 值为 1), 但预测框无法贴合转子的实际边界会引起其他视觉位移测量曲线产生周期性的误差, 而本文算法结合 SE, FPN 能够保证转子目标存在特征差异时, 位移信号拟合精准且噪声较小。

综合图 11 和图 12 的结果后发现, 同类别算法在回归转子 1, 2 曲线时, 转子 2 的位移曲线拟合更好, 这也验证了第 2.2 节中转子 2 因比转子 1 特征明显而在回归时更好地贴合转子目标的定论。如图 13 和图 14 所展示的频谱分析结果, 不同算法学习转子特征时有不同的侧重点, 因此不同算法 2 倍频特征存在明显差异。此外, 在零倍频和 1 倍频上, 频率幅值上有明显的差别, 这与时域信号有关, 如图 11 和图 12 所示, 即时域信号噪声越大, 零倍频幅值越大, 1 倍频幅值越小。本文算法体现出主峰峰值幅值一致性以外且频率噪点小, 而其他算法较多的频率峰值与电涡流相比均有差别。这也能够说明本文的算法在回归位移曲线时, 频率和振幅能与电涡流信号相匹配。

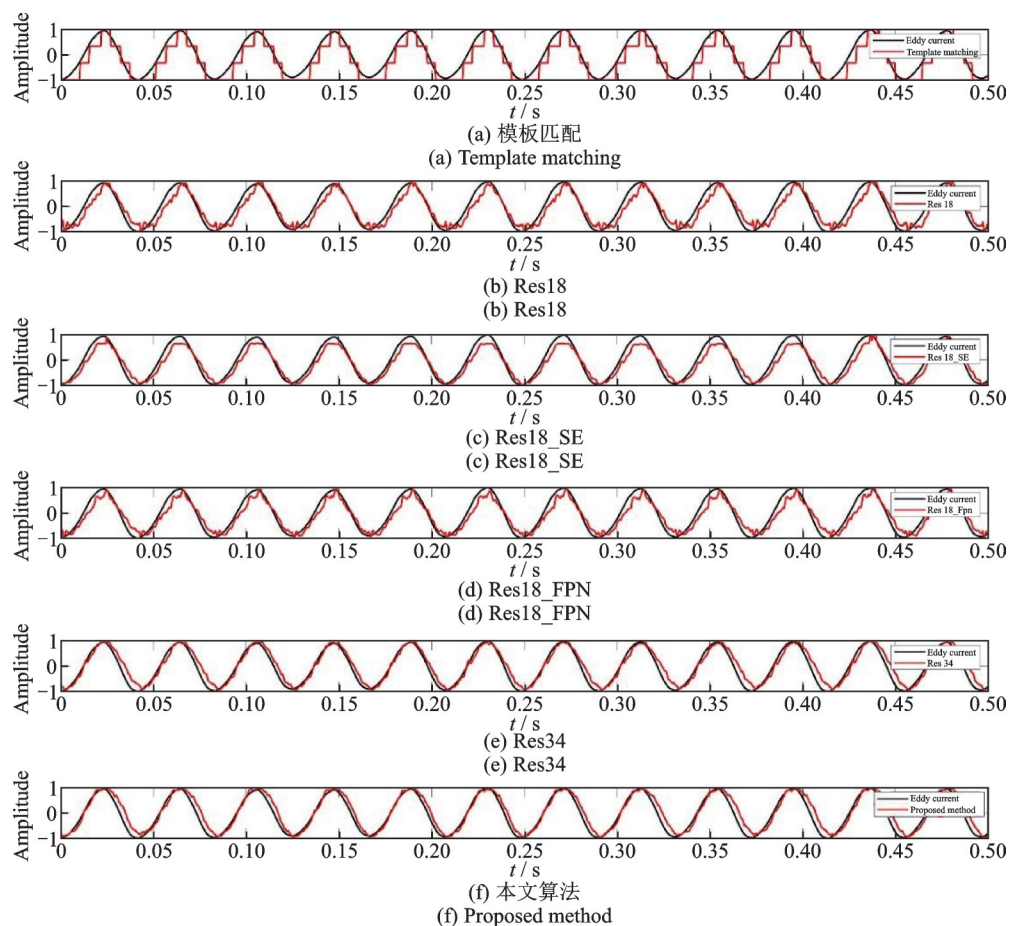


图 12 各算法回归转子 2 位移曲线与电涡流信号对比

Fig. 12 The algorithm regression rotor 2 displacement curve is compared with the eddy current signal

2.4 对象模糊状态下测试

在视觉振动测量中, 过分追求较高帧率和处理速度必然促使采集设备降低图像分辨率, 从而导致采集的低分辨率视频出现模糊现象, 同时也会存在典型背景干扰导致拍摄目标特征不明显的现象。为

了模拟复杂高动态场景下的工业现场, 将采集的图像进行 4 倍下采样后获取低分辨率的图像和目标。采集的原始图像和下采样后对应的低分辨率图像如图 15 所示, 图 15(b) 中两个转子边界存在明显的边缘模糊现象。对视觉振动测量而言, 边缘模糊无疑给振动测量的可靠性带来很大的干扰。对于图 16

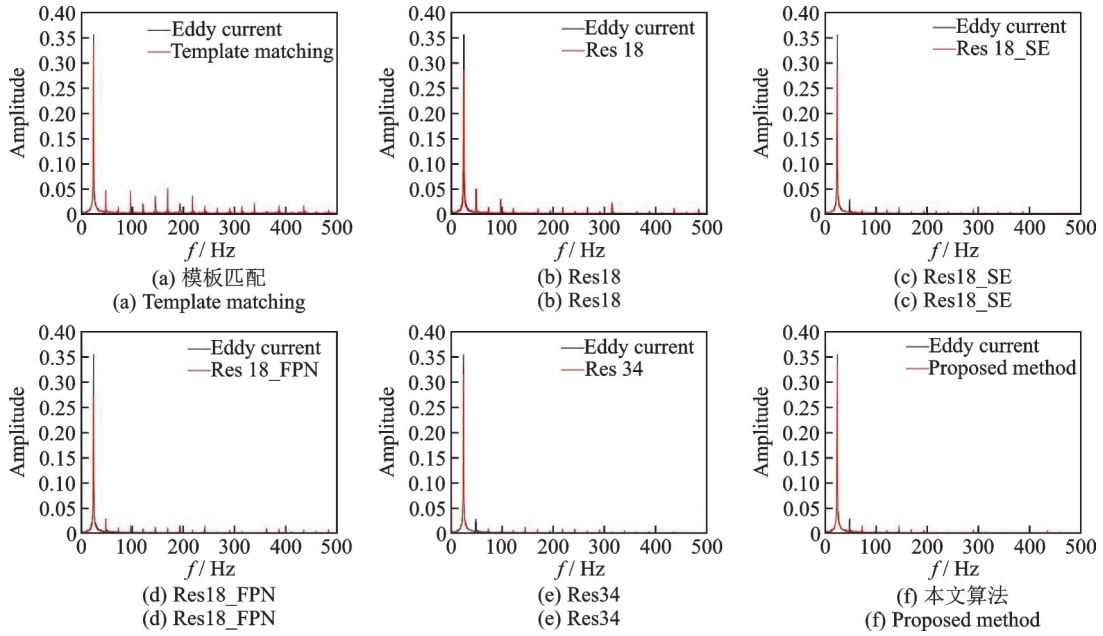


图13 各算法回归转子1位移曲线与电涡流信号的FFT变换对比

Fig. 13 Comparison of the FFT transformation between the displacement curve of rotor 1 and the eddy current signal of each algorithm

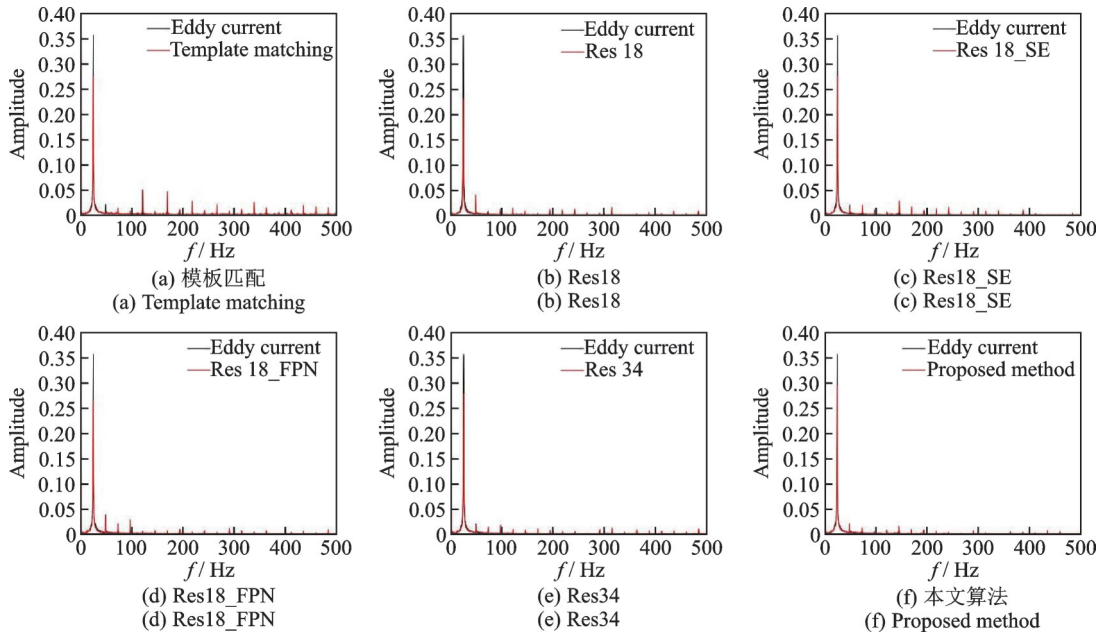


图14 各算法回归转子2位移曲线与电涡流信号的FFT变换对比

Fig. 14 Comparison of the FFT transformation between the displacement curve of rotor 2 and the eddy current signal of each algorithm

中由于拍摄角度引入过多无关信息的转子1而言,模糊后回归的位移曲线与原始图像回归的位移曲线相比虽然在单个周期内的位移曲线上均产生了微弱的偏差,但仍然保持近似的周期性。而因本文算法稳健的鲁棒性测试转子2目标的位移输出依旧,在曲线拟合度方面保持近似一致的周期性。因此,本文网络对于模糊场景下的视觉振动测量具有一定的优势,且频域图17中降采样后的模糊图像测量数据与原始图像、电涡流信号的1倍频峰值以及频率都

能有效对应,这也验证了本文网络在实际应用场景的泛化性。

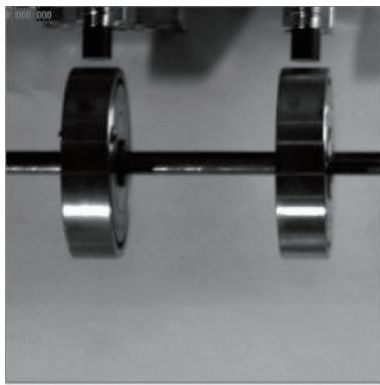
3 结论

视觉位移测量因具有远距离、非接触和多点测量等特性,近年来开始逐渐应用于结构体健康监测领域。本文在提高旋转目标检测精度的同时增强振动位移的关联性,利用深度卷积神经网络进行视觉

表 2 视觉信号与电涡流信号定量比较

Tab. 2 Quantitative comparison of visual and eddy current signals

方法	目标	峰值	峰峰值	有效值
Eddy current	转子 1	1.0000	1.9849	0.6817
	转子 2	1.0000	1.9733	0.6798
Template matching	转子 1	1.0000	2.0000	0.7072
	转子 2	1.0000	2.0000	0.7072
Res18	转子 1	0.8862	1.8191	0.5958
	转子 2	0.8898	1.5860	0.4831
Res18_SE	转子 1	0.6425	1.6245	0.5859
	转子 2	0.9560	1.9438	0.5901
Res18_FPN	转子 1	0.9540	1.9023	0.5720
	转子 2	0.9107	1.8172	0.5683
Res34	转子 1	0.8548	1.7616	0.6412
	转子 2	0.9258	1.8868	0.6546
本文方法	转子 1	0.9445	1.9151	0.6693
	转子 2	0.9305	1.8402	0.5922



(a) 原始图像
(a) Original image

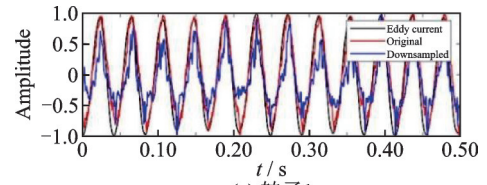


(b) 降采样后图像
(b) Downsampled image

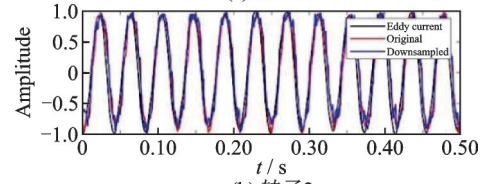
图 15 原始和下采样后图像对比

Fig. 15 Comparison of original and downsampled images

跟踪测量多目标振动位移。将视频信息和电压位移信息进行不同视觉测量算法性能比较,本文网络在旋转体视觉位移多目标振动测量时具有较高的精度;对比不同振动测量方式后发现,视觉在测量点数、距离有独特的优势,但会因目标特征造成信号拟



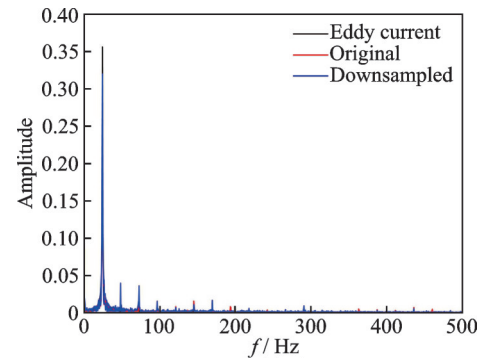
(a) 转子 1
(a) Rotor 1



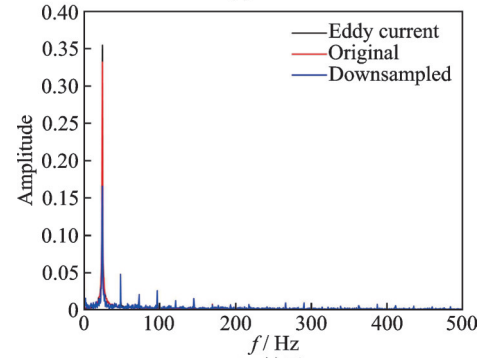
(b) 转子 2
(b) Rotor 2

图 16 两个转子的电涡流、原始图像和降采样后图像时域信号对比

Fig. 16 Comparison of eddy current, original image and downsampled image time domain signal of two rotors



(a) 转子 1
(a) Rotor 1



(b) 转子 2
(b) Rotor 2

图 17 两个转子的电涡流、原始图像和降采样后图像的 FFT 对比

Fig. 17 FFT comparison of eddy current, original image and downsampled image of two rotors

合的差异。不同转子测量结果也体现出视频帧间的目标重复识别有利于相邻位移点之间的位置关联。否则即使是识别精度更高的图像检测算法,在振动位移信号提取时还是存在较大的抖动误差。另一方面,由于转子具有较强的通用性和互换性,其形态规则、特征明显,所以基于关键点的追踪网络尤其适用于旋转体的视觉振动位移测量。考虑到在对象模糊

的情况下,本文将原始采集图像进行四倍下采集操作,并与原始图像测试比较,实验结果验证本文算法在目标不清晰的情况下仍然有较好的振动测量能力。

本文目前的实验主要关注简单背景下多目标转子的位移测量,后期工作会将逐渐增加背景的复杂度和单个转子的跟踪维度,关注图像拍摄角度对振动测量的影响,以构建出适用于复杂工业场景下的多目标视觉位移追踪网络。

参考文献:

- [1] Lu S L, He Q B, Zhao J W. Bearing fault diagnosis of a permanent magnet synchronous motor via a fast and online order analysis method in an embedded system [J]. *Mechanical Systems and Signal Processing*, 2018, 113: 36-49
- [2] Jia F, Lei Y G, Guo L, et al. A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines[J]. *Neurocomputing*, 2018, 272: 619-628.
- [3] Dong C Z, Ye X W, Jin T. Identification of structural dynamic characteristics based on machine vision technology[J]. *Measurement*, 2018, 126: 405-416.
- [4] Ehrhart M, Lienhart W. Development and evaluation of a long range image-based monitoring system for civil engineering structures[C]//*Structural Health Monitoring and Inspection of Advanced Materials, Aerospace, and Civil Infrastructure 2015*. San Diego, California, USA, 2015: 94370K.
- [5] Ye X W, Dong C Z, Liu T. Image-based structural dynamic displacement measurement using different multi-object tracking algorithms[J]. *Smart Structures and Systems*, 2016, 17(6): 935-956.
- [6] Celik O, Dong C Z, Catbas F N. A computer vision approach for the load time history estimation of lively individuals and crowds[J]. *Computers & Structures*, 2018, 200: 32-52
- [7] Dong C Z, Celik O, Catbas F N. Long-term structural displacement monitoring using image sequences and spatio-temporal context learning [C]//*Structural Health Monitoring 2017*. Stanford, CA, 2017:3217-3223.
- [8] 钱承, 鄂加强. 基于 FSVM 的起重机臂筒焊接变形测量研究[J]. *振动、测试与诊断*, 2018, 38(6): 1229-1233.
Qian Cheng, E Jiaqiang. Measurement research on welding distortion of crane arm tube based on Fusion of SVM and fuzzy logical theory[J]. *Journal of Vibration, Measurement and Diagnosis*, 2018, 38(6): 1229-1233.
- [9] 葛朋祥, 叶沛, 李桂华. 基于遗传算法的数字图像相关法在微位移测量中的应用[J]. *光学学报*, 2018, 38(6): 198-203.
Ge Pengxiang, Ye Pei, Li Guihua. Application of digital image correlation method based on genetic algorithm in micro-displacement measurement[J]. *Acta Optica Sinica*, 2018, 38(6): 198-203.
- [10] Khaloo A, Lattanzi D. Pixel-wise structural motion tracking from rectified repurposed videos[J]. *Structural Control and Health Monitoring*, 2017, 24(11): e2009.
- [11] 邱志成, 汪先锋. 基于双目视觉的振动测量及控制[J]. *振动、测试与诊断*, 2018, 38(1): 51-58.
Qiu Zhicheng, Wang Xianfeng. Vibration measurement and control based on binocular vision[J]. *Journal of Vibration, Measurement and Diagnosis*. 2018, 38(1): 51-58.
- [12] Lydon D, Lydon M, Taylor S, et al. Development and field testing of a vision-based displacement system using a low cost wireless action camera[J]. *Mechanical Systems and Signal Processing*, 2019, 121: 343-358.
- [13] 邵帅, 周志祥, 邓国军, 等. 基于非接触远程智能感知的桥梁形态监测试验[J]. *中国公路学报*, 2019, 32(11): 91-102.
Shao Shuai, Zhou Zhixiang, Deng Guojun, et al. Experiment of structural morphology monitoring for bridges based on non-contact remote intelligent perception method [J]. *China Journal of Highway and Transport*, 2019, 32(11): 91-102.
- [14] Zhang Y, Liu P, Zhao X F. Structural displacement monitoring based on mask regions with convolutional neural network [J]. *Construction and Building Materials*, 2021, 267: 120923.
- [15] Zhu J S, Zhang C, Lu Z Y, et al. A multi-resolution deep feature framework for dynamic displacement measurement of bridges using vision-based tracking system [J]. *Measurement*, 2021, 183:109847.
- [16] Dai J F, Qi H Z, Xiong Y W, et al. Deformable convolutional networks[C]//*Proceedings of the IEEE International Conference on Computer Vision*. Venice, Italy, 2017: 764-773.
- [17] Huang M F, Zhang B Y, Lou W J, et al. A deep learning augmented vision-based method for measuring dynamic displacements of structures in harsh environments [J]. *Journal of Wind Engineering and Industrial Aerodynamics*, 2021, 217:104758.
- [18] Xu Y, Zhang J, James B. An accurate and distraction-free vision-based structural displacement measurement method integrating Siamese network based tracker and correlation-based template matching[J]. *Measurement*,

- 2021, 179:109506.
- [19] Liu J T, Yang X X, Li L. VibroNet: recurrent neural networks with multi-target learning for image-based vibration frequency measurement[J]. Journal of Sound and Vibration, 2019, 457: 51-66.
- [20] Son K S, Jeon H S, Chae G S, et al. A fast high-resolution vibration measurement method based on vision technology for structures[J]. Nuclear Engineering and Technology, 2021, 53:294-303.
- [21] Zhang Y F, Wang C Y, Wang X G, et al. A simple baseline for multi-object tracking[J]. arXiv: 2004.01888, 2020.
- [22] Zhou X Y, Wang D Q, Krähenbühl P. Objects as points [J]. arXiv:1904.07850, 2019.
- [23] Law H, Deng J. Cornernet: detecting objects as paired keypoints[J]. International Journal of Computer Vision, 2020,128:642-656.
- [24] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks [J]. arXiv preprint arXiv:1609.02907, 2016.
- [25] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, Hawaii, USA, 2017: 936-944..
- [26] Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019,42(8):2011-2023.
- [27] Wang Q L, Wu B G, Zhu P F, et al. ECA-Net: efficient channel attention for deep convolutional neural networks[J]. arXiv: 1910.031512019.

Vision tracking multi-rotor displacement measurement

YANG Rong-liang¹, WANG Sen¹, WU Xing², LIU Xiao-qin¹, LIU Tao¹

(1.School of Mechanical and Electrical Engineering, Kunming University of Science and Technology, Kunming 650500, China;

2.Yunnan Mechanical and Electrical Vocational and Technical College, Kunming 650023, China)

Abstract: Aiming at the problem that the current traditional vibration sensor is limited by the installation and the number of measuring points when measuring the displacement of the rotating body, this paper uses a high-speed industrial camera as the acquisition medium, and collects the rotor vibration video on the rotor vibration test bench. The visual vibration measurement method tracks the full-field vibration displacement of multiple rotor targets. The feature pyramid network structure is introduced into the residual neural network, and the improved feature extraction backbone network is established by combining the attention mechanism. The identity re-identification method is used to strengthen the correlation of target displacement between adjacent frames, and to track the full-field vibration displacement signals of the rotator. Qualitative and quantitative comparisons of different network models on the rotor vibration displacement measurement dataset show that the network model proposed in this paper can obtain a tighter fit when the bounding box is regressed. The collected eddy current displacement signal is used as the standard value to compare the two rotor displacement signals, and the experimental results show that the waveform and spectral noise fitted by the multi-target tracking algorithm in this paper is the smallest and can match the eddy current signal. The experiments also prove the generalization performance of the algorithm in this paper, which reflects the engineering application value of visual measurement in the field of vibration displacement tracking of rotating bodies.

Key words: visual vibration measurement; deep learning; multi-target; visual tracking; blurred image; rotational body displacement measurement

作者简介: 杨荣良(1997—),男,硕士研究生。E-mail: yr19797@126.com。

通讯作者: 王 森(1983—),男,博士,副教授。E-mail: wangsen0401@126.com。