中国光学 (中英文)
Chinese Optics
EI 核心期刊

融合全局和局部信息的铁谱图像自动对焦算法

刘信良 张龙泉 冷晟 王静秋 王晓雷

An autofocus algorithm for fusing global and local information in ferrographic images

LIU Xin-liang, ZHANG Long-quan, LENG Sheng, WANG Jing-qiu, WANG Xiao-lei

引用本文:
刘信良, 张龙泉, 冷晟, 王静秋, 王晓雷. 融合全局和局部信息的铁谱图像自动对焦算法[J]. 中国光学, 2024, 17(2): 423–434. doi: 10.37188/CO.2023-0124
LIU Xin-liang, ZHANG Long-quan, LENG Sheng, WANG Jing-qiu, WANG Xiao-lei. An autofocus algorithm for fusing global and local information in ferrographic images[J]. *Chinese Optics*, 2024, 17(2): 423-434. doi: 10.37188/CO.2023-0124

在线阅读 View online: https://doi.org/10.37188/CO.2023-0124

文章编号　2097-1842(2024)02-0423-12

融合全局和局部信息的铁谱图像自动对焦算法

刘信良, 张龙泉, 冷　晟, 王静秋*, 王晓雷
（南京航空航天大学 直升机传动技术重点实验室, 江苏 南京 210016）

摘要: 针对铁谱图像获取时人工对焦误差大、速度慢等问题, 提出了一种融合全局信息和局部信息的铁谱图像自动对焦方法。此方法分为两个阶段: 全局对焦阶段利用卷积神经网络 (Convolutional Neural Networks, CNN) 提取整幅图像的特征向量, 并利用门控循环单元 (Gate Recurrent Unit, GRU) 融合对焦过程提取的特征, 预测当前全局离焦距离, 起到粗对焦的作用; 局部对焦阶段提取磨粒的特征向量, 利用 GRU 融合当前特征与前一轮对焦提取的特征, 并依据最厚磨粒信息, 预测当前磨粒离焦距离, 起到精对焦的作用。同时, 为了提高对焦准确率, 提出了结合拉普拉斯梯度的对焦方向判定法。实验结果表明, 此算法在测试集上的对焦误差为 2.51 μm, 当景深为 2.0 μm 时对焦准确率为 80.1%, 平均对焦时间为 0.771 s。本文提出的自动对焦方法具有较好的性能, 为铁谱图像自动准确采集提供了技术支持。

关　键　词: 自动对焦; 铁谱图像; 全局信息; 局部信息; 深度学习; 门控循环单元

中图分类号: TP391.4; TH117.2　　**文献标志码:** A　　**doi:** 10.37188/CO.2023-0124

An autofocus algorithm for fusing global and local information in ferrographic images

LIU Xin-liang, ZHANG Long-quan, LENG Sheng, WANG Jing-qiu*, WANG Xiao-lei
(*National Key Laboratory of Science and Technology on Helicopter Transmission*, *Nanjing University of Aeronautics and Astronautics*, *Nanjing* 210016, *China*)
* *Corresponding author*, *E-mail*: *meejqwang@nuaa.edu.cn*

Abstract: To address the issues of large error and slow speed of manual focusing in ferrographic image acquisition, we propose an autofocus method for fusing global and local information in ferrographic images. This method includes two stages. In the first stage, the global autofocus stage, the feature vectors of the whole image is extracted by Convolutional Neural Networks (CNN) , and the features extracted in the focus process is fused by the Gate Recurrent Unit (GRU) to predict global defocusing distance, which serves as coarse focusing. In the local autofocus stage, the feature vector of the wear particle is extracted and the current features is fused with those extracted in the previous focusing process by GRU. The current defocusing distance is predicted by the resulting fused data based on the information of the thickest particle, which facilitates fine focusing. Moreover, we propose a determination method for autofocus direction using Laplacian

收稿日期: 2023-07-26; **修订日期:** 2023-08-24
基金项目: 直升机传动技术重点实验室基金项目 (No. HTL-A-21G03)
Supported by National Key Laboratory of Science and Technology on Helicopter Transmission (No. HTL-A-21G03)

gradient function to improve autofocus accuracy. Experimental results indicate an autofocus error of 2.51 μm on the test set and a focusing accuracy of 80.1% with a microscope depth of field of 2.0 μm. The average autofocus time is 0.771 s. The automatic ferrographic image acquisition system exhibits excellent performance and offers a practical approach for its implementation.

Key words: autofocus; ferrographic images; global information; local information; deep learning; gate recurrent unit

1 引 言

铁谱分析是监测设备运行状态及磨损程度的重要手段，它利用高梯度强磁场将磨粒从润滑油中分离并沉积到玻璃基片上[1]，再通过显微镜和数码相机拍摄铁谱图像。铁谱图像上磨粒的尺寸、数量和形貌等信息，是判定设备当前磨损状态和磨损故障的重要依据[2]。然而，由于不同类型磨粒尺寸差异大，例如，正常滑动磨粒的尺寸范围是 0.5~15 μm，厚度是 0.15~1 μm，而疲劳剥落磨粒尺寸范围是 10~100 μm，厚度是 1~10 μm[3]。因此，当不同类型的磨粒出现在视野范围内时，由于显微镜的景深 (Depth of Field, DoF) 较小，使得磨粒无法同时清晰地呈现在同一幅图像中，而且由于需要手动对焦，因此人工拍摄铁谱图像时存在速度慢、误差大等问题，严重限制了铁谱分析的效率和准确率。

自动对焦技术可以使显微镜快速、精确地定位到对焦平面 (目标物体所在平面)。根据有无额外硬件，其分为主动式对焦和被动式对焦[4]。主动式对焦通过额外加装传感器，如激光测距仪[5-7]、相位传感器[8-11] 等，直接获取对焦平面的位置，准确、快速，但硬件改造复杂且成本高，会对成像系统的其他方面造成限制。被动式对焦方法通过图像处理的方式实现自动对焦，成本低、对硬件改动小，具体包括对焦深度法和离焦深度法。对焦深度法采集成像设备对焦过程中每个位置的图像，通过图像清晰度评价准则[12] 和搜索算法确定对焦平面的位置[13]；离焦深度法依据成像系统当前位置采集图像的失焦参数，如能量函数、梯度函数、点扩散函数 (PSF)[14] 等，预测出当前位置与对焦平面之间的距离[15]。

利用尽可能少的图像获取精准的对焦结果是目前自动对焦算法的挑战。Farnes S A R 等[16] 提出通过单张图像预测离焦距离的方法，此方法将图像分成若干包含简单图案的小块，通过 CNN 提取特征向量，再利用回归器预测出离焦距离。此方法为印刷电路板的自动对焦提供了解决方案，但要求划分的小图像块包含简单的图案或原始形状，如弧形或直线等。Ge Y H 等[17] 使用分类方法，将对焦过程采集的图像分为 7 个类别，分别对应 7 种不同的对焦程度，再将提取的特征向量送入分类器中，计算出当前的对焦程度，从而找到对焦平面的位置，其准确率受对焦程度类别数和每类包含图像数量的影响。Xiang Y 等[18] 提出了结合时序特征的自动对焦算法，从图像中裁剪 5 个固定位置的图像块作为模型的输入，利用 CNN 提取这些图像块的特征向量，同时，融合前一轮对焦提取的特征并将其输入循环神经网络 (Recurrent Neural Network, RNN) 中来预测当前的离焦距离。此方法中，仅凭单幅图像无法准确预测对焦平面的位置，通过图像特征与时序特征相结合，经过多轮迭代才能得到更准确的结果。

虽然自动对焦方法取得了很大的进步，但将其直接应用到铁谱图像自动对焦时却存在一些问题。由于不同类型磨粒厚度差异大，且显微镜的景深较小，导致对焦平面不唯一；对焦平面两侧等距处采集的图像相似度高，依据单张图像很难准确预测对焦方向；磨粒在谱片上分布位置差异大，固定位置裁剪的图像块可能不包含磨粒。

针对上述问题，本文提出了一种融合全局信息和局部信息的铁谱图像自动对焦算法。该算法包括全局对焦和局部对焦两个阶段，全局对焦阶段通过提取整幅图像的特征来预测全局离焦距离，目的是快速定位到对焦平面附近，起到粗对焦作用；局部对焦阶段提取图像中的磨粒特征，再依据检出的最厚磨粒信息预测对焦平面，起到精对

焦作用。同时, 针对对焦方向不准确的问题, 提出了结合图像清晰度的对焦方向的判定方法。

2　融合全局和局部信息的对焦算法

本文提出的自动对焦算法框架如图 1 所示。在全局对焦阶段, 设镜头初始位置为 x_0, 此位置获取的图像为 I_{x_0}, 首先利用全局自动对焦模块提取整幅图像 I_{x_0} 的特征, 再利用全局离焦模块进行预测, 获得全局离焦距离 d_1', 最后进行对焦方向判断获得修正后的离焦距离 d_1, 以更新镜头位置。同时, 将获取的新图像 $I_{x_0+d_1}$ 作为下次全局对焦过程的输入。全局对焦共进行 3 次, 最后获取的图像为 $I_{x_0+\sum_t d_i}$ (t=3), 即后续局部对焦阶段的输入图像。在局部对焦阶段, 利用局部自动对焦模块提取磨粒特征, 依据最厚磨粒的信息回归出局部离焦距离 d_4', 再进行对焦方向判断获得修正后的离焦距离 d_4, 并更新镜头的位置, 将获取的新图像 $I_{x_0+\sum_t d_i}$ (t=4), 作为下次对焦过程的输入。局部对焦阶段同样进行 3 次, 最终对焦平面的位置为 $x_0+\sum_t d_i$ (t=6)。

下文将详细介绍全局自动对焦模块和局部自动对焦模块。

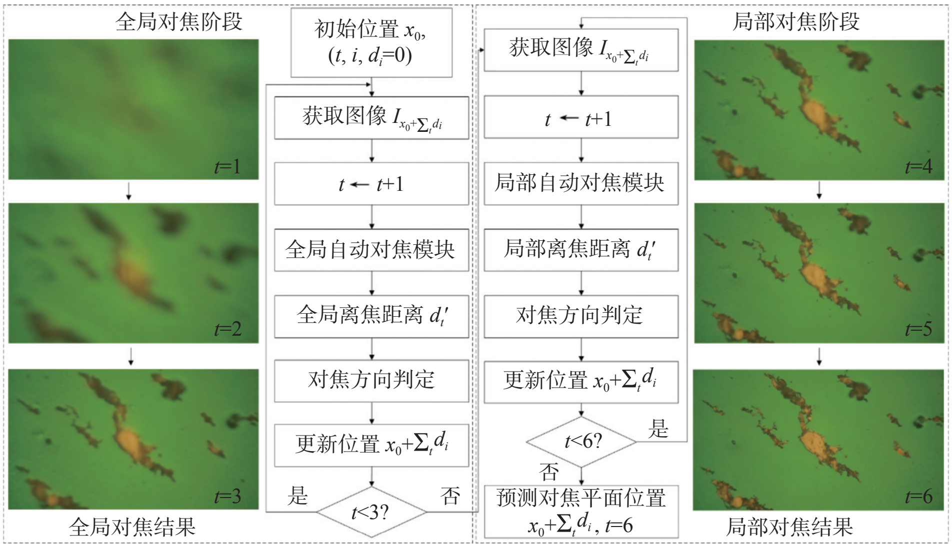


图 1　自动对焦算法框架

Fig. 1　Framework of autofocus algorithm

2.1　全局自动对焦模块

如图 2 (a) 所示, 全局自动对焦模块由全局特征提取网络和全局离焦距离回归网络组成, 分别起到提取图像特征和粗对焦的作用。

全局特征提取网络由 DBL 和 ResBlock 组成, 其结构分别如图 2 (b), 2(c) 所示。输入图像经过 2 个 DBL 和 23 个 ResBlock 处理, 得到尺寸为输入图像 1/32 的特征图; 之后, 使用全局平均值池化 (Global Average Pooling, GAP) 和 Reshape 操作将特征图转化为一维全局特征向量 $\boldsymbol{h}_t$。

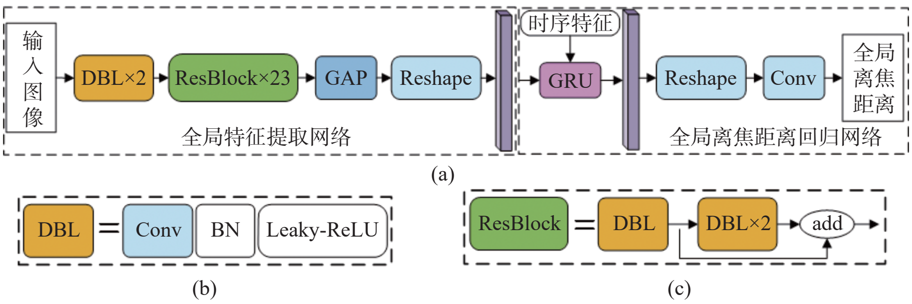


图 2　(a)全局自动对焦模块结构图及(b)DBL 结构和(c)ResBlock 结构

Fig. 2　(a) Schematic diagram of global autofocus module structure, (b) DBL structure and (c) ResBlock structure

为了准确预测离焦距离，在全局离焦距离回归网络中引入 GRU，用于提取时序特征。如图 3(a) 所示，其输入 $\boldsymbol{x}_t$ 为第 t 次对焦提取的图像全局特征，$\boldsymbol{h}_{t-1}$ 为第 t−1 次对焦提取的图像全局特征，输出 $\boldsymbol{h}_t$ 为两者融合后的特征，重置门决定了如何将新的输入信息与前面的记忆相结合，更新门定义了前面记忆保存到当前时间步的量。由于 GRU 能更好地捕捉时序数据中间隔较大的依赖关系，故用于将第 t 次与第 t−1 次对焦获取的图像特征融合，从而更好地预测离焦距离。GRU 单元的计算过程如下：

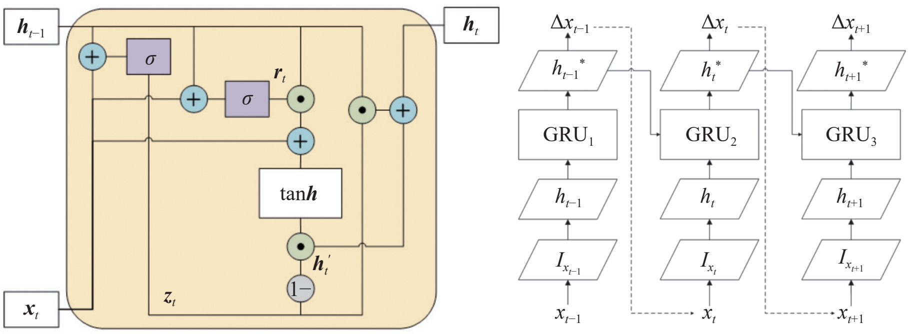
更新门 $$\boldsymbol{z} = \sigma(\boldsymbol{x}_t \cdot \boldsymbol{U}^z + \boldsymbol{b}_{xz} + \boldsymbol{h}_{t-1} \cdot \boldsymbol{V}^z + \boldsymbol{b}_{hz}), \quad (1)$$

重置门 $$\boldsymbol{r} = \sigma(\boldsymbol{x}_r \cdot \boldsymbol{U}^r + \boldsymbol{b}_{xr} + \boldsymbol{h}_{t-1} \cdot \boldsymbol{V}^r + \boldsymbol{b}_{hr}), \quad (2)$$

$$\boldsymbol{n} = \tanh\left(\boldsymbol{x}_t \cdot \boldsymbol{U}^h + \boldsymbol{b}_{xn} + (\boldsymbol{h}_{t-1} \odot \boldsymbol{r})\boldsymbol{V}^h + \boldsymbol{b}_{hn}\right), \quad (3)$$

$$\boldsymbol{h}_t = (1 - \boldsymbol{z}) \odot \boldsymbol{n} + \boldsymbol{z} \odot \boldsymbol{h}_{t-1}. \quad (4)$$

其中，t 代表循环次数，$\boldsymbol{x}_t$ 为 t^{th} 迭代的输入向量，$\boldsymbol{h}_{t-1}$ 保存的是上一次的迭代信息，即 $\boldsymbol{h}_{t-1}$=$\boldsymbol{x}_{t-1}$ (当 t=1 时，$\boldsymbol{h}_{t-1}$=0)，σ 代表 sigmoid 激活函数，$\boldsymbol{z}$ 和 $\boldsymbol{r}$ 分别代表当前迭代中更新门和重置门的参数，$\boldsymbol{n}$ 为当前记忆的特征，$\boldsymbol{h}_t$ 为最终记忆的特征，$\boldsymbol{U}$ 和 $\boldsymbol{V}$ 为权重，$\boldsymbol{b}$ 为偏置。



(a) GRU 结构图
(a) GRU structure diagram
(b) 离焦距离回归网络流程
(b) Defocus distance regression network process

图 3 离焦距离回归网络结构图

Fig. 3 The structure of defocus distance regression network

离焦距离回归网络的结构如图 2 和图 3 (b) 所示。全局特征提取网络输出的特征向量 $\boldsymbol{h}_t$ 与第 t−1 次对焦提取的特征向量 $\boldsymbol{h}_{t-1}^*$一起作为 GRU 的输入，得到新的特征向量 $\boldsymbol{h}_t^*$；再通过 Reshape 操作将维度还原，最后经过卷积层得到全局离焦距离。重复上述过程 3 次，得到最终的全局离焦距离。

2.2 局部自动对焦模块

局部自动对焦模块由磨粒检测网络和局部离焦距离回归网络组成，分别起到磨粒定位与特征提取、精对焦的作用。

如图 4 所示，磨粒检测网络采用 YOLOv3[19] 多尺度目标检测网络，用于检测磨粒和提取磨粒特征，其主干网络采用 ResNet-50[20] 结构。最终磨粒检测网络输出的特征图为 Det，其尺度为 $\frac{W}{2^{2+i}} \times \frac{H}{2^{2+i}} \times[3\times(4+1+1)]$，其中，$i$=1, 2, 3，代表特征图所在尺度分支的序号；$W$ 和 H 分别是输入图像的宽和高；输出特征图中的每个点包含 3 个“锚”(anchor) 的预测信息，每个“锚”包含预测目标的 4 个坐标值、置信度和类别。

局部离焦距离回归网络的输入为磨粒检测网络最后一个 conv 层之前所提取出的特征，每个尺度的特征经过一个 conv 层，得到尺寸为 $\frac{W}{m}\times\frac{H}{m}\times 6$ 的特征图；通过 Reshape 操作将其降为一维特征向量 $\boldsymbol{h}_t$，并与上轮对焦提取的特征向量 $\boldsymbol{h}_{t-1}^*$共同作为 GRU 的输入，得到新特征向量 $\boldsymbol{h}_t^*$；再通过 Reshape 操作将其恢复为原尺寸，最后通过 conv 层得到局部自动对焦的结果为 O-pred，其尺度为 $\frac{W}{m} \times \frac{H}{m} \times[(1+1)\times3]$。其中，输出结果中每个点都包含 3 个“锚”的信息，与磨粒检测网络结果一一对应。每个“锚”包含 2 个通道，分别表示该磨粒为最大厚度磨粒的概率和离焦距离。经过 3 次局

部自动对焦过程, 可以定位对焦平面, 即最厚磨粒的上切面。

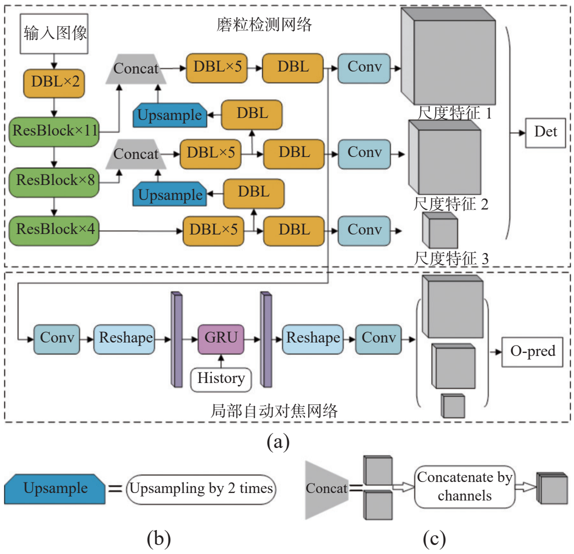


图 4　(a)局部自动对焦模块结构图; (b)Upsample 结构图及(c)Concat 结构图

Fig. 4　(a) Structure diagram of local autofocus module. (b) Upsample structure. (c) Concat structure

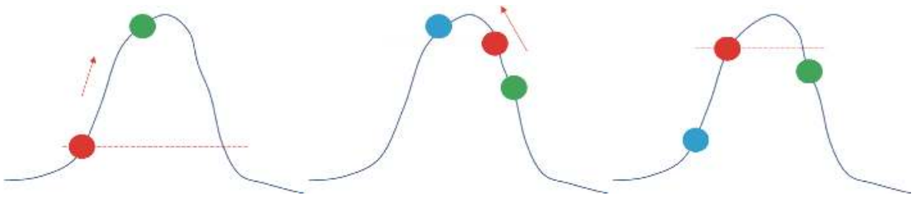
2.3　结合拉普拉斯梯度的对焦方向判定法

由于对焦平面两侧等距处的图像相似度高, 只依据全局和局部对焦过程可能出现预测方向错误、离焦距离偏差过大的问题。为了进一步提高自动对焦的准确率, 提出利用图像清晰度辅助判断对焦方向的方法, 通过对比第 t–1 次和第 t 次采集图像的清晰度, 辅助修正对焦方向。本文采用拉普拉斯算子作为图像清晰度评价准则, 其定义为:

$$\nabla^2 I(x,y) = I(x+1,y) + I(x,y+1) + I(x-1,y) + I(x,y-1) - 4 \cdot I(x,y) \quad ,\tag{5}$$

$$LP(I) = \sum_{x=1}^{W-1}\sum_{y=1}^{H-1} \nabla^2 I(x,y) \quad ,\tag{6}$$

其中, $I(x, y)$ 代表图像中 (x, y) 处的像素值。如图 5 (彩图见期刊电子版)所示, 设 L_c 表示当前位置图像的拉普拉斯值; L_p 表示预测位置的拉普拉斯值; L_r 表示预测位置反向的拉普拉斯值。



红色: 当前位置图像的拉普拉斯值 (Lc); 绿色: 预测位置的拉普拉斯值 (Lp); 蓝色: 预测位置反向的拉普拉斯值 (Lr)

图 5　对焦方向判定

Fig. 5　Determination of focus direction

当前清晰度值与预测点的清晰度值之间的关系可以分为以下 3 种情况:

情况 1: $L_c \leqslant L_p$, 预测点的方向正确;

情况 2: $L_c > L_p$, 且 $L_r \geqslant L_c$, 预测点的方向错误, 需要反向;

情况 3: $L_c > L_p$, 且 $L_r < L_c$, 保持当前位置。

通过结合拉普拉斯梯度的对焦方向判定法, 可以决定镜头是前进、后退还是保持当前位置, 从而准确地判定对焦方向。

3　实　验

3.1　数据集制作

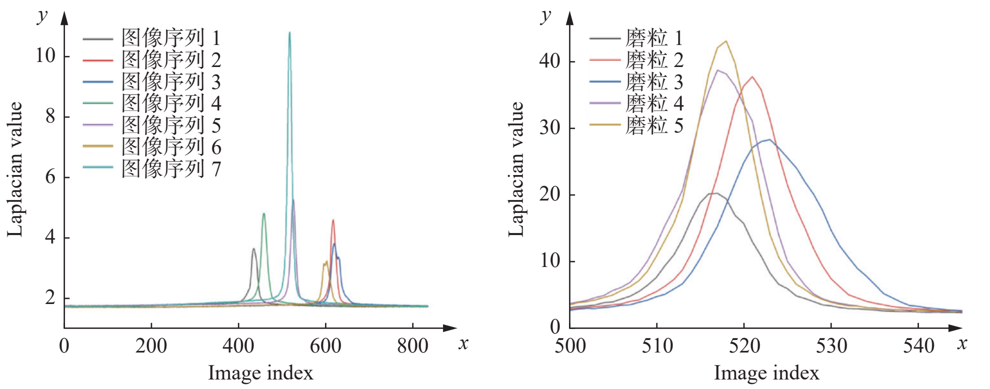
如图 6 所示, 实验平台包括正置式金相显微镜 (XJZ-6)、XYZ 电控位移台、电机控制器、数码照相机和计算机。XYZ 电控位移台与计算机相连, 可以通过指令控制平台的移动速度和距离。相机以 50 帧/秒拍摄图像, 图像大小为 1280×720, 记为 $G=[I_0, I_1, \ldots, I_{N-1}]$, 其中, N=834 为图像序列的帧数。图像序列中的每幅图像都对应不同镜头位置, 相邻两帧图像对应的物理距离 Δd=0.40 μm。为方便表述, 本文以图像在图像序列中的位置 (即图像索引) 代指真实物理镜头的位置, 因此, 可以通过图像索引确定实际镜头的位置。本文共采集 69 组图像序列, 数据集划分为训练集、验证集和测试集, 分别包含 37 组、4 组和 28 组图像序列。为了避免在模型训练时占用内存过大, 所有图像尺寸统一缩放为 352×352。

全局自动对焦数据集：计算图像序列中每帧图像的拉普拉斯梯度，部分图像序列的拉普拉斯曲线如图 7 (a)(彩图见期刊电子版)所示。曲线峰值处对应图像序列中整体最清晰的图像 I_k，将其索引 k 所对应的物理位置作为粗对焦平面位置。因此，当前索引 i 的离焦距离 d 可定义为 $d=(i-k)\cdot\Delta d$。

局部自动对焦数据集：包含图像序列中粗对焦平面前后各 50 帧图像，记为 $L=[I_0, I_1, \ldots, I_{N-1}]$，$N$=100。如图 7 (b)(彩图见期刊电子版)所示，计算图像序列中每个磨粒区域的拉普拉斯梯度，曲线峰值处代表磨粒 P_c 在索引 h 处最清晰，将 h 所对应的物理位置作为该磨粒精对焦平面的位置。对于索引 i 处的同一磨粒，其离焦距离记为 $d=(i-h)\cdot\Delta d$。最后，通过人工辨识，标记图像序列中每张图像中厚度最大的磨粒。

图 6 图像采集平台

Fig. 6 Image acquisition platform



(a) 7 组图像序列的拉普拉斯曲线
(a) Laplacian curves of 7 sets of image sequences

(b) 5 个不同磨粒的拉普拉斯曲线
(b) Laplacian curves of 5 different abrasive particles

图 7 图像序列的拉普拉斯清晰度曲线

Fig. 7 Laplacian sharpness curves of an image sequence

磨粒检测数据集：包含从图像序列中选取的较为清晰的图像和 346 幅静态采集的清晰图像，用于训练和验证磨粒检测网络，并用 labelimg 软件标记图像中所有磨粒。

3.2 实验设置

此自动对焦模型的两个模块采取分开训练的方式，即训练一个模块时冻结另一个模块的参数。

全局自动对焦模块：训练时采用 Momentum 优化器，动量参数设置为 0.9；batch size 设为 8；学习率设为 0.00025；其损失函数定义如下：

$$L=\sum_{i}^{T_c}|\hat{y}_i-y_i|+\lambda L_{\rm reg}\quad, \tag{7}$$

其中，T_c 代表全局自动对焦迭代的次数；$\hat{y}_i$ 代表第 i 次迭代预测的离焦距离，y_i 代表真实的离焦距离；为了防止过拟合、提高鲁棒性，加入 L2 正则化项，$\lambda L_{\rm reg}$，其中 λ 是权重参数。此模块共训练 10 轮。

局部自动对焦模块：磨粒检测网络训练时采用 Momentum 优化器，动量参数设为 0.9；batch size 设为 6；学习率设为 0.00025；损失函数如文献 [19] 所述，共训练 30 轮。

局部自动对焦网络训练时采用相同的优化器；batch size 设为 1；学习率设为 0.001；其损失函数的定义如下：

$$L=\sum_{i}^{T_j}\left(\sum_{n}^{N}t_i^n\cdot\left(\left|\hat{y}_i^{(n)}-y_i^{(n)}\right|-p_i^{(n)}\cdot\log\hat{p}_i^{(n)}\right)\right)+\lambda L_{\rm reg}, \tag{8}$$

其中，T_j 代表局部自动对焦迭代的次数；N 为“锚”的数量；t_i^n 代表第 i 轮第 n 个“锚”是否存在磨粒，有磨粒的 t_i^n 值为 1，否则为 0；$\hat{y}_i^{(n)}$ 代表第 i 次迭代

第 n 个“锚”预测的离焦距离，$y_i^{(n)}$代表真实的离焦距离；$\hat{p}_i^{(n)}$代表第 i 次迭代的第 n 个“锚”预测为最厚磨粒的概率，$p_i^{(n)}$代表是否为厚度最大的磨粒。

3.3　评价指标

本实验采用 3 种评价指标：对焦偏差 $dist$，对焦准确率 Acc_{dof}，平均对焦时间 AT。假设测试集中有 M 个图像序列，每个图像序列包含 N 帧，则测试集可以表示为 φ_t=[[I_1^1, I_1^2, ⋯, I_1^n],[I_2^1, I_2^2, ⋯, I_2^n], ⋯,[I_m^1, I_m^2, ⋯, I_m^n]]。对焦偏差 $dist$ 是预测值$\hat{y}_m^n$与实际值y_m^n之间的偏差，其表达式为：

$$dist = \frac{1}{M} \cdot \frac{1}{N} \cdot \sum_{m=1}^{M} \sum_{n=1}^{N} \left|\hat{y}_m^n - y_m^n\right| \quad . \qquad (9)$$

由于采集图像时使用了较高的帧率，导致图像序列中两帧之间的差异较小，关键磨粒无明显变化，因此，根据实际情况设置 3 组景深 (分别为 1, 3, 5；对应的 l_{dof} 分别为 0, 1, 2)，分别测试 3 种景深下的对焦准确率 $Acc_{\text{dof-1}}$, $Acc_{\text{dof-3}}$, $Acc_{\text{dof-5}}$，其定义为：

$$Acc_{\text{dof}} = \frac{1}{M} \cdot \frac{1}{N} \cdot \sum_{m=1}^{M} \sum_{n=1}^{N} \phi\left(\left|\hat{y}_m^n - y_m^n\right| \leqslant l_{\text{dof}}\right) \quad , \qquad (10)$$

若最终预测的位置$\hat{y}_m^n$在景深 [y - l_{dof}, y + l_{dof}] 中，函数 $\phi(\cdot)$ 的值为 1，否则为 0。

3.4　实验结果

图 8(彩图见期刊电子版)展示了测试集中一组图像序列的自动对焦过程。在全局自动对焦阶段，输入图像为图像序列的第 97 帧，经过 3 次全局对焦过程，其粗对焦结果分别为第 324 帧、第 402 帧、第 405 帧。可以看出，随着全局对焦阶段的进行，对应位置的图像逐渐清晰，表明逐渐定位至粗对焦平面。在局部自动对焦阶段中，图中红框标记的磨粒为预测的最厚磨粒，利用该磨粒的信息进行 3 次局部对焦，其精对焦结果分别为图像序列的第 409 帧、第 413 帧、第 437 帧，最终获取的精对焦平面为该图像序列的第 437 帧。可以看出，最厚磨粒的边缘、纹理等逐渐清晰，表明镜头已经定位至精对焦平面。

从该图像序列中任取 10 帧作为自动对焦算法的初始图像，其 6 步对焦过程的结果如图 9 (a)(彩图见期刊电子版)所示。可以看出，虽然初始图像的索引差异较大，但随着自动对焦的进行，最终都定位至对焦平面附近。将图像序列的所有帧都作为自动对焦算法的初始图像，其最终对焦结果如图 9 (b)(彩图见期刊电子版)所示，其中，x 代表初始图像的索引，y 代表对焦结果图像的索引。可以看出，当初始图像在 1~70 帧区间内时，镜头距离对焦平面过远，图像完全失焦模糊，而无法准确提取离焦特征，导致全局对焦结果出现偏差，从而影响局部对焦阶段的准确率；同时，当初始图像在 640~720 帧区间时，全局对焦阶段结果出现偏差，导致误检最厚磨粒，从而影响最终对焦准确率；当初始图像在 70~640 帧及 720~834 帧区间时，对焦结果曲线较为平稳，能够准确地定位对焦平面。

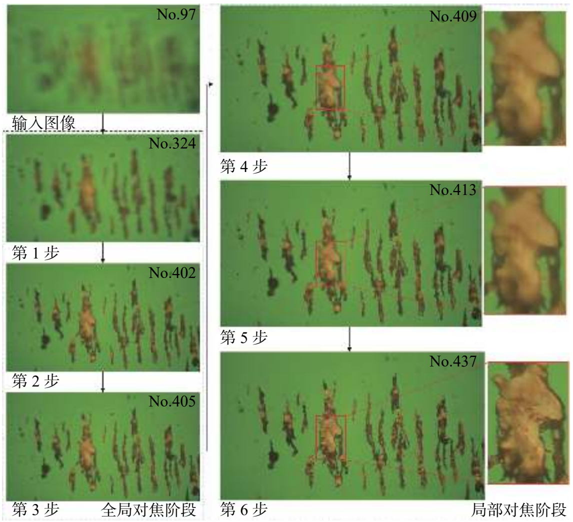
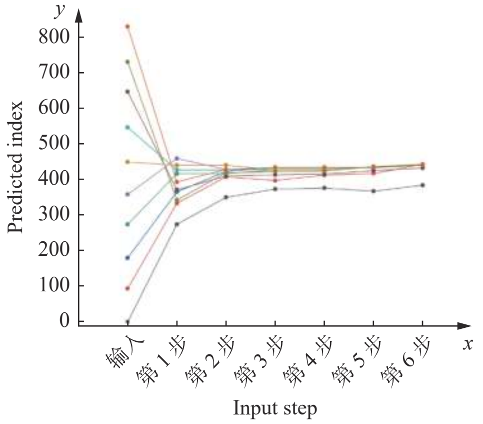


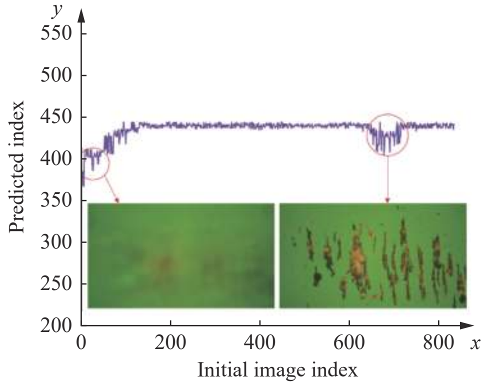
图 8　自动对焦过程

Fig. 8　Autofocus process

图 10(彩图见期刊电子版)展示了测试集中 4 组不同图像序列的对焦结果曲线，其中，x 为初始图像中的索引，y 为对焦结果的索引，6 条曲线分别代 6 步对焦过程的结果。例如，在图 10 (a) 中，当初始图像索引为第 200 帧时，其 6 步对焦结果分别为第 585 帧、第 621 帧、第 630 帧、第 642 帧、第 646 帧、第 648 帧。可以看出，对焦的准确率较高，且随着自动对焦的进行，镜头定位的准确率不断提升，曲线波动程度减弱，最终预测的结果稳定在对焦平面附近，如图 10 (a)、10(b) 所示。然而，当镜头初始位置距离对焦平面过远时，采集图像比较模糊，全局自动对焦模块未能提取准确的离焦信息，预测结果会出现偏差，导致对焦失败；同时，当厚度最大磨粒检测出现错误时，也会导致最终对焦结果出现偏差，如图 10 (c)、10(d) 所示。



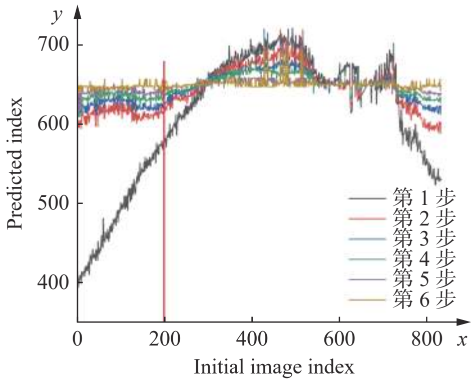
(a) 10 帧初始图像的 6 步对焦结果曲线
(a) Result curves of 6-step focus for ten initial images



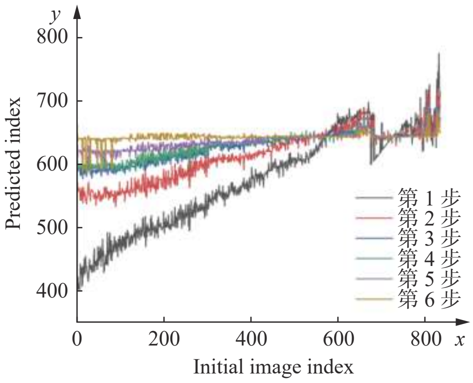
(b) 不同初始图像的最终对焦结果曲线
(b) Final focus curves for different initial images

图 9 不同输入图像的对焦结果

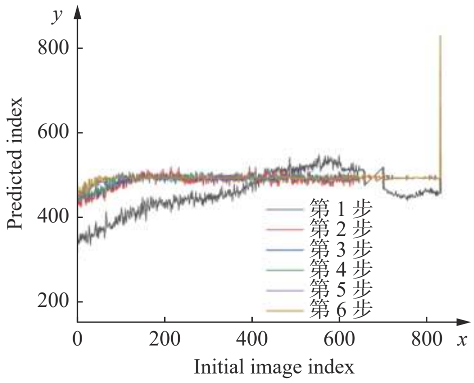
Fig. 9 Focus results for different input images



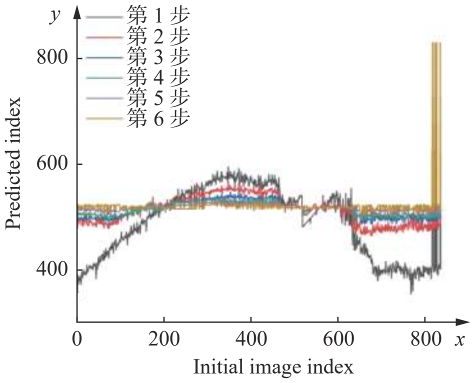
(a) 示例 1
(a) Example 1



(b) 示例 2
(b) Example 2



(c) 示例 3
(c) Example 3



(d) 示例 4
(d) Example 4

图 10 4 组图像序列的对焦结果

Fig. 10 Focusing results of 4 groups of image sequences

将本文提出的自动对焦算法在测试集上进行测试，其真实对焦索引在 414~715 帧之间。计算得到该自动对焦算法在测试集上的结果：对焦偏差 *dist* 为 6.271 帧，即 2.51 μm，对焦准确率 $Acc_{\text{dof-1}}$ 为 36.0%，$Acc_{\text{dof-3}}$ 为 65.1%，$Acc_{\text{dof-5}}$ 为 80.1%，平均对焦时间 *AT* 为 0.771 s。图 11 展示了该自动对焦算法在测试集上预测结果索引与真实对焦索引的关系。可以看出，预测结果都集中在真实对焦索引附近，对焦偏差较小，结果准确率较高且平稳。

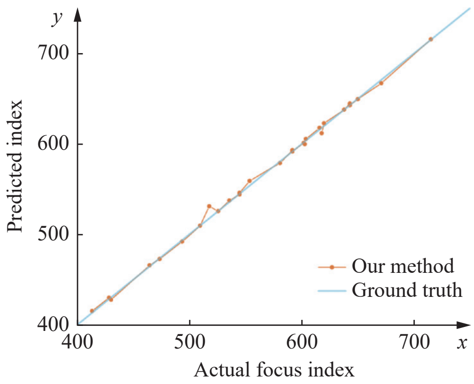


图 11　本方法测试集上自动对焦的结果

Fig. 11　Autofocus results of proposed method on the test set

表 1 展示了自动对焦过程中每一步的结果。可以看出, 随着自动对焦的进行, 平均对焦偏差减小, 平均对焦准确率增大, 表明镜头每一步都朝着真实对焦平面前进。当 DoF 为 1 帧时, 结果准确率最低, 其主要原因是相邻两帧之间的图像特征变化较小, 导致对焦存在微小偏差。随着 DoF 的增大, 其准确率逐渐上升, 当 DoF 为 5 帧时, 对焦过程中每一步的准确率都最高。

表 1　对焦过程中每一步的结果

Tab. 1　Results of each step in the focusing process

i^{th} step	*dist (frame)*	$Acc_{\text{dof-1}}$	$Acc_{\text{dof-3}}$	$Acc_{\text{dof-5}}$	*AT* (*s*)
1	63.649±12.960	0.017±0.009	0.039±0.017	0.061±0.027	0.118±0.034
2	22.678±6.408	0.061±0.026	0.133±0.583	0.202±0.086	0.115±0.027
3	15.404±5.660	0.134±0.062	0.257±0.125	0.346±0.153	0.118±0.041
4	10.891±4.205	0.194±0.076	0.364±0.135	0.474±0.152	0.138±0.027
5	7.393±3.235	0.288±0.102	0.523±0.145	0.666±0.141	0.140±0.019
6	6.271±2.680	0.360±0.130	0.651±0.149	0.801±0.125	0.143±0.018

4　实验分析

4.1　消融实验

利用消融实验可以评价算法中各模块对结果的影响。本文针对时序特征提取单元 GRU、对焦策略和局部自动对焦模块 LAF 进行了 4 组消融实验, 以探究其在本文自动对焦算法中的作用, 实验结果如表 2 和图 12 所示。

消融实验 1 用于分析 GRU 结构对最终结果的影响。可以看出, 两者在 $Acc_{\text{dof-1}}$ 上的差别较小, 但是随着 DoF 增大, 无 GRU 结构的算法因无法结合前一轮对焦提取的特征, 其准确率变低, 且对焦偏差较大, 表明时序特征对算法性能有一定影响。

消融实验 2 用于分析对焦策略对最终结果的影响。可以看出, 由于对焦平面两侧等距处采集的图像相似度高, 无对焦策略的算法由于无法判定当前位置位于对焦平面的哪一侧, 故依据当前图像很难准确预测对焦方向, 导致对焦偏差大, 对焦准确率有所下降, 表明对焦策略对模型准确率的影响较大。使用本文提出的对焦策略可以准确获取对焦方向, 从而提升对焦准确率。

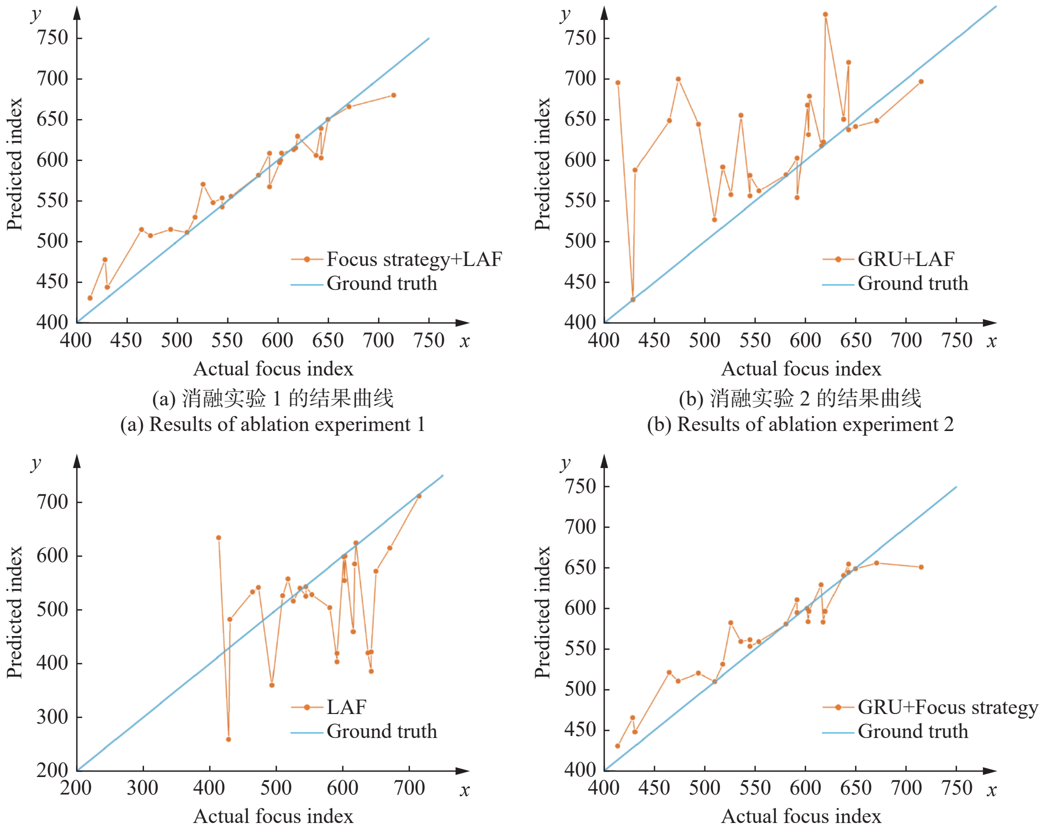
消融实验 3 用于分析未采用 GRU 和对焦策略对最终结果的影响。可以看出, 相较于消融实验 1 和 2, 只采用局部自动对焦模块 LAF 的对焦误差最大, 且对焦准确率最低, 表明 GRU 结构与对焦策略的组合对算法的性能影响较大。

消融实验 4 用于分析局部自动对焦模块对最

表 2　消融实验的结果

Tab. 2　Results of ablation experiments

消融实验序号	GRU	Focus strategy	LAF	*dist (frame)*	$Acc_{\text{dof-1}}$	$Acc_{\text{dof-3}}$	$Acc_{\text{dof-5}}$
消融实验1		√	√	29.496±16.882	0.345±0.150	0.578±0.163	0.663±0.157
消融实验2	√		√	81.259±71.561	0.032±0.047	0.068±0.092	0.111±0.123
消融实验3			√	101.528±71.457	0.023±0.029	0.056±0.059	0.084±0.086
消融实验4	√	√		28.046±20.225	0.253±0.146	0.484±0.202	0.626±0.216
本文算法	√	√	√	6.271±2.680	0.360±0.130	0.651±0.149	0.801±0.125



(a) 消融实验 1 的结果曲线
(a) Results of ablation experiment 1

(b) 消融实验 2 的结果曲线
(b) Results of ablation experiment 2

(c) 消融实验 3 的结果曲线
(c) Results of ablation experiment 3

(d) 消融实验法 4 的结果曲线
(d) Results of ablation experiment 4

图 12　4 组消融实验的对焦结果

Fig. 12　Focusing results of four ablation experiments

终结果的影响。可以看出，未采用局部自动对焦模块时，仅依据整幅图像的特征预测离焦距离，对焦偏差变大，且对焦准确率降低。表明只用全局对焦过程的算法，无法检测到最厚磨粒，只能将镜头定位至对焦平面附近，不能实现精确对焦。

从上述对比实验可以看出，GRU、对焦策略和局部自动对焦模块分别起到结合时序信息、修正对焦方向和磨粒精对焦的作用，本文提出的算法将三者结合，为准确对焦提供了保障。

图 12 展示了上述 4 组消融实验的结果。对比图 11 中的本文算法结果和图 12 结果。可以看出：本文算法的准确率高，预测的平稳性好；其次为未采用 GRU 的算法 (消融实验 1) 和未采用局部自动对焦模块的算法 (消融实验 4)，它们的预测结果与真实对焦索引较为接近，且波动较小；而未采用对焦策略的算法 (消融实验 2) 和只采用局部自动对焦模块的算法 (消融实验 3) 表现相对较差，它们的预测的对焦结果与真实标签之间的偏差较大。

4.2　与其他自动对焦算法对比分析

将本文提出的自动对焦算法分别与 5 种已有对焦算法进行比较。5 种算法分别为整图全局搜索法、图像块全局搜索法 (取图像 4 个角及中心位置处 700×400 大小的图像块)、爬山法和基于深度学习的两种算法，即 HH-Net[17] 与 Autofocus - RNN[18] 算法。所有算法在测试集上的最终结果如表 3 所示。

表 3　不同自动对焦算法的结果

Tab. 3　Results of different autofocus algorithms

序号	算法	*dist (frame)*	$Acc_{dof\text{-}1}$	$Acc_{dof\text{-}3}$	$Acc_{dof\text{-}5}$	*AT* (*s*)
1	整图全局搜索法	8.647	0.107	0.321	0.536	17.856
2	图像块全局搜索法	8.603	0.179	0.357	0.607	22.068
3	爬山法	12.926	0.286	0.464	0.500	1.459
4	HH-Net	27.177	0.036	0.179	0.429	0.119
5	Autofocus-RNN	31.839	0.321	0.429	0.571	0.419
6	本文算法	6.271	0.360	0.651	0.801	0.771

由表 3 可以看出：整图全局搜索法和图像块全局搜索法表现相对较好。全局搜索算法会计算

镜头行程内每个位置的图像清晰度值, 找出清晰度值最大的图像, 再将镜头移至此位置。此算法的对焦误差 *dist* 为 8.647 帧, 对焦准确率 $Acc_{\text{dof-1,3,5}}$ 分别为 0.107, 0.321, 0.536; 由于需要遍历镜头行程的每个位置, 其平均对焦时间 *AT* 为 17.856 s, 对焦速度远慢于本文提出的算法。图像块全局搜索算法的对焦误差 *dist* 为 8.603 帧, 对焦准确率 $Acc_{\text{dof-1,3,5}}$ 分别为 0.179, 0.357, 0.607, 平均对焦时间 *AT* 为 22.068 s。可以看出, 基于图像块的算法相较基于整图的算法结果更精确, 但速度相对较慢。这是由于图像块中的磨粒分布比整图的更集中, 梯度更明显, 故以此图像块为搜索目标可以得到更准确的结果。然而, 无论是以整幅图像还是图像块为搜索目标, 都无法获取最厚磨粒的准确位置, 从而导致对焦出现偏差。

爬山算法改进了搜索准则, 镜头从初始位置移动, 每前进一次就计算当前位置的图像清晰度值, 并与前一次计算结果做比较, 从而确定下一次的步长和前进方向。相较前两种全局搜索算法, 爬山算法提升了搜索效率, 其平均对焦时间 *AT* 缩短至 1.459 s, 并且 $Acc_{\text{dof-1}}$ 与 $Acc_{\text{dof-3}}$ 分别提升至 0.286 与 0.464。然而, 此算法的 *dist* 与 $Acc_{\text{dof-5}}$ 则分别下降为 12.926 帧与 0.500。产生这种情况的主要原因是图像序列的拉普拉斯曲线可能呈现双峰或多峰特性, 导致爬山算法容易陷入局部最优解, 从而使一些图像组对焦失败。

HH-Net 算法将对焦问题转化为对图像清晰度的分类, 通过构建由 13 个卷积层和 3 个全连接层组成的简单分类模型, 识别当前图像清晰度的等级, 从而决定镜头前进的方向与步长。此算法的平均对焦时间 *AT* 为 0.119 s, 对焦速度最快。然而, 对焦误差 *dist* 为 27.177 帧, 对焦准确率 $Acc_{\text{dof-1,3,5}}$ 分别为 0.036, 0.179, 0.429。相对而言, 其对焦误差较大、对焦准确率最低。这主要是由于对焦过程中非清晰图像占多数, 各类图像的占比差别大, 类别不均衡, 从而影响分类及对焦的准确率。

Autofocus-RNN 将对焦问题转化为离焦距离预测, 利用 CNN 提取图像特征, 并通过 RNN 融合前一轮对焦提取的特征来预测当前的离焦距离。此算法的对焦误差 *dist* 为 31.839 帧, 对焦准确率 $Acc_{\text{dof-1,3,5}}$ 分别为 0.321, 0.429, 0.571, 平均对焦时间 *AT* 为 0.419 s。可以看出, 其对焦准确率较高, 速度较快, 但对焦误差较大。这主要由于部分图像块中不包含磨粒, 且当镜头远离对焦平面时图像完全失焦模糊, 导致离焦特征提取时出现偏差; 同时, 此算法难以精准定位最厚磨粒, 难以精准对焦。

本文提出算法的对焦误差 *dist* 为 6.271 帧, 对焦准确率 $Acc_{\text{dof-1,3,5}}$ 分别为 0.360, 0.651, 0.801, 平均对焦时间 *AT* 为 0.771s。可以看出, 本文提出算法的对焦误差最小, 对焦准确率最高。然而, 由于此算法分为全局和局部对焦两个阶段, 且每个阶段分别对焦 3 次, 算法复杂度相较 HH-Net 和 Autofocus-RNN 更高, 因此平均对焦时间相对稍长, 但显著短于前 3 种算法, 满足铁谱图像获取的需求。

5　结　论

本文提出了一种融合全局信息与局部信息的铁谱图像自动对焦算法。该算法包含全局对焦和局部对焦 2 个阶段, 分别起到粗对焦和精对焦的作用。在全局自动对焦阶段, 利用 CNN 提取整幅图像的特征向量, 并用 GRU 融合上轮对焦提取的全局特征, 预测当前位置的全局离焦距离。在局部自动对焦阶段, 利用 YOLOv3 网络提取磨粒特征, 用 GRU 融合上轮对焦提取的磨粒特征, 依据最厚磨粒的信息, 预测当前位置的局部离焦距离。同时, 针对对焦方向不准确的问题, 提出了结合拉普拉斯的对焦方向判定法, 以获取准确的对焦方向。

该算法将 CNN 提取的图像特征与 GRU 提取的时序特征结合, 通过全局对焦与局部对焦两个阶段, 可以实现快速、准确的自动对焦。实验结果表明, 在测试集上, 该算法的对焦误差为 6.271 帧, 即 2.51 μm, 对焦准确率 $Acc_{\text{dof-5}}$ 为 80.1%, 平均对焦时间为 0.771 s, 性能上优于传统自动对焦算法和当前性能较好的深度学习算法, HH-Net 与 Autofocus-RNN。该算法提升了铁谱图像采集的自动化程度, 方便后续智能化铁谱分析。

参考文献:

[1]　ROYLANCE B J. Ferrography—then and now[J]. *Tribology International*, 2005, 38(10): 857-862.

[2]　WANG J Q, WANG X L. The segmentation of ferrography images: a brief survey[J]. *Materials Science Forum*, 2013,

770: 427-432.

[3] 卿华, 王新军. 飞机油液监控技术[M]. 北京: 航空工业出版社, 2011.

QING H, WANG X J. *Aircraft oil Monitoring Technology*[M]. Beijing: Aviation Industry Press, 2011. (in Chinese).

[4] 叶一青, 易定容, 张勇贞, 等. 基于倾斜摄像头的显微自动对焦方法[J]. 光学学报, 2019, 39(12): 1218001.

YE Y Q, YI D R, ZHANG Y ZH, *et al*.. Microscopy autofocus method using tilt camera[J]. *Acta Optica Sinica*, 2019, 39(12): 1218001. (in Chinese).

[5] BATHE-PETERS M, ANNIBALE P, LOHSE M J. All-optical microscope autofocus based on an electrically tunable lens and a totally internally reflected IR laser[J]. *Optics Express*, 2018, 26(3): 2359-2368.

[6] ZHANG X, ZENG F, LI Y, *et al*.. Improvement in focusing accuracy of DNA sequencing microscope with multi-position laser differential confocal autofocus method[J]. *Optics Express*, 2018, 26(2): 895-904.

[7] 唐凌宇, 葛明锋, 董文飞. 全自动推扫式高光谱显微成像系统设计与研究[J]. 中国光学, 2021, 14(6): 1486-1494.

TANG L Y, GE M F, DONG W F. Design and research of fully automatic push-broom hyperspectral microscopic imaging system[J]. *Chinese Optics*, 2021, 14(6): 1486-1494. (in Chinese).

[8] JANG J, YOO Y, KIM J, *et al*.. Sensor-based auto-focusing system using multi-scale feature extraction and phase correlation matching[J]. *Sensors*, 2015, 15(3): 5747-5762.

[9] GUO K K, LIAO J, BIAN Z CH, *et al*.. InstantScope: a low-cost whole slide imaging system with instant focal plane detection[J]. *Biomedical Optics Express*, 2015, 6(9): 3210-3216.

[10] GAN Y H, YE Z T, HAN Y B, *et al*.. Single-shot autofocusing in light sheet fluorescence microscopy with multiplexed structured illumination and deep learning[J]. *Optics and Lasers in Engineering*, 2023, 168: 107663.

[11] HOU S B, ZHANG H Y, MA B L, *et al*.. Extended autofocusing in dual-wavelength digital holography[J]. *Applied Optics*, 2023, 62(22): 5959-5968.

[12] 郭立强, 刘恋. 结合斜变换与方差的图像聚焦测度[J]. 光学 精密工程, 2021, 29(7): 1731-1739.

GUO L Q, LIU L. Image focus measure based on slant transform and variance[J]. *Optics and Precision Engineering*, 2021, 29(7): 1731-1739. (in Chinese).

[13] LIAO Y, XIONG Y H, YANG Y H. An auto-focus method of microscope for the surface structure of transparent materials under transmission illumination[J]. *Sensors*, 2021, 21(7): 2487.

[14] NOEK R, KNOERNSCHILD C, MIGACZ J, *et al*.. Multiscale optics for enhanced light collection from a point source[J]. *Optics Letters*, 2010, 35(14): 2460-2462.

[15] SHAJKOFCI A, LIEBLING M. Semi-blind spatially-variant deconvolution in optical microscopy with local point spread function estimation by use of convolutional neural networks[C]. *Proceedings of 25th IEEE International Conference on Image Processing*, IEEE, 2018.

[16] FARNES S A R, TSAI D M, CHIU W Y. Autofocus measurement for electronic components using deep regression[J]. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, 2021, 11(4): 697-707.

[17] GE Y H, LI B, ZHAO Y ZH, *et al*. . HH-Net: image driven microscope fast auto-focus with deep neural network[C]. *Proceedings of the 9th International Conference on Biomedical Engineering and Technology*, Association for Computing Machinery, 2019: 180-185.

[18] XIANG Y, HE ZH J, LIU Q, *et al*.. Autofocus of whole slide imaging based on convolution and recurrent neural networks[J]. *Ultramicroscopy*, 2021, 220: 113146.

[19] REDMON J, FARHADI A. YOLOv3: an incremental improvement[J/OL]. arXiv: 1804.02767, 2018.

[20] HE K M, ZHANG X Y, REN SH Q, *et al*.. Deep residual learning for image recognition[C]. *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2016.

通讯作者：

王静秋 (1972—), 女, 辽宁抚顺人, 博士, 教授, 1994 年、1999 年、2014 年于南京航空航天大学分别获得学士、硕士、博士学位, 主要从事摩擦学、计算机图像处理、分子动力学模拟等方向的研究。E-mail: meejqwang@nuaa.edu.cn