基于深度学习的数字病理扫描系统单次曝光自动对焦方法

李强^{1,2}, 刘贤明^{1,2*}, 韩凯歌¹, 江俊君^{1,2}, 季向阳³

1. 哈尔滨工业大学计算学部, 哈尔滨 150001

2. 鹏城实验室, 深圳 518052

3. 清华大学自动化系, 北京 100084

* 通信作者. E-mail: csxm@hit.edu.cn

收稿日期: 2020-11-06; 修回日期: 2021-01-15; 接受日期: 2021-02-22; 网络出版日期: 2021-10-14

国家自然科学基金 (批准号: 61922027, 61827804) 和国家重点研发计划 (批准号: 2019YFE0109600) 资助项目

摘要 全切片数字成像 (whole slide imaging, WSI) 是病理切片数字化的核心技术, 其自动对焦的速度和精度决定了 WSI 系统的性能. 然而, 传统的自动对焦方法需要拍摄多张离焦子图像创建图像堆栈, 或者需要复杂的硬件调制光学系统, 从而限制了 WSI 在实际场景中的应用. 本文设计了基于深度学习的数字病理扫描系统单次曝光自动对焦方法, 对子图像逐个进行网络虚拟自动对焦, 将单张离焦子图像通过网络直接生成准焦子图像. 本方法仅需要在任意离焦距离下的单次曝光拍照即可, 从原理上避免了重复的对焦运动和相机曝光过程. 实验结果表明, 本方法具有高通量、高速度、低成本、实用性强、可线下处理等优点.

关键词 数字病理扫描系统, 自动对焦, 深度学习, 光学显微镜, 计算成像

1 引言

病理诊断被誉为疾病诊断的“金标准”, 其研究水平代表了一所医院的医疗质量水平, 是衡量国家医疗质量的重要标志^[1]. 病理诊断在各种肿瘤、炎症等疾病的诊断准确率上高达 95% 以上, 客观地反映了疾病的发生发展规律, 在疾病的成因、机理、诊断、治疗和预后指导中占有不可替代的地位^[2~4]. 近年来, 随着计算机、光学仪器和自动化技术的不断发展, 病理切片数字化应运而生. 数字病理扫描仪是一种将病理切片转化为全切片数字图像 (whole slide images, WSI)^[5,6], 能够实现大容量图像存储与传输、虚拟病理图像的导航与放缩、疾病远程智能分析与诊断等功能的软硬件一体化综合系统^[7~9], 其工作流程如图 1 所示. 大容量图像的存储与传输实现了病理数据的永久保存、异地实时阅览、远程

引用格式: 李强, 刘贤明, 韩凯歌, 等. 基于深度学习的数字病理扫描系统单次曝光自动对焦方法. 中国科学: 信息科学, 2021, 51: 1675–1689, doi: 10.1360/SSI-2020-0349
Li Q, Liu X M, Han K G, et al. Whole slide imaging via one-shot deep autofocusing (in Chinese). Sci Sin Inform, 2021, 51: 1675–1689, doi: 10.1360/SSI-2020-0349

图 1 (网络版彩图) WSI 工作流程
Figure 1 (Color online) The workflow of WSI

专家会诊、数据资源共享等功能; 虚拟病理图像的导航与放缩, 可通过系统自动对焦、全局扫描实现, 将高分辨率的病理显微图像通过终端展示; 疾病远程智能分析与诊断在基于病理大数据的深度学习临床诊断上开展, 以先进的人工智能技术和计算机辅助诊断技术相结合的方式, 加速了数字病理系统的应用和普及^[10]. 然而, 目前国内病理切片数字化扫描设备数量缺口大, 且主要集中在中低端产品. 高端病理仪器市场被国外公司完全垄断, 制定了严重畸形的市场价格. 数字病理扫描系统的开发涉及光机系统、电控系统、计算机软件等多学科技术结合, 具有很高的技术壁垒.

为了解决上述问题, 需要对数字病理扫描系统相关技术展开研究, 推动国内相关病理设备的发展. 病理切片自动对焦技术^[11], 是病理扫描系统的核心技术: 对焦的精度决定了系统扫描成像质量, 对焦的速度决定了系统扫描策略效率. 显微系统的自动对焦方法分为硬件装置方法和图像处理方法: 硬件方法采用硬件辅助手段, 对光路进行调制, 可以快速实时地实现自动对焦, 但成本高且应用受限; 图像处理方法需要连续采集 N 张图像, 通过评价函数极值确定焦点位置, 但需要大量曝光拍照, 并且系统需要扫描两次, 时间成本较高.

综上所述, 设计一种高成像质量、高扫描效率、低成本易扩展的数字病理扫描系统自动对焦方法具有重要的科学研究意义和临床应用价值^[12]. 不同于传统逐个子图像 (tiles, 单次拍照视场图像) 扫描的机械调焦方法, 本文设计了基于深度学习的数字病理扫描系统单次曝光自动对焦方法, 对不同离焦距离下的离焦子图像进行网络虚拟自动对焦, 将单张离焦子图像直接生成准焦子图像. 实验证明, 本方法具有高通量快速成像、无需硬件系统调制、低成本易迁移、可线下处理等优点.

2 相关工作

本节将简要介绍显微系统自动对焦的两类方法, 分别是基于硬件和辅助装置的自动对焦方法和基于图像处理的自动对焦方法.

基于硬件和辅助装置的自动对焦方法通常直接快速地实现焦距的测量. Liron 等^[13] 采用额外光源测量参考点的方法实现焦点位置的测量. Montalto 等^[14] 利用了额外的高速相机作为对焦相机, 实现了和主成像相机并行工作的双相机自动对焦. Liao 等^[15] 开发了 LED 编码照明自动对焦方法, 通过设计 LED 阵列编码照明策略, 实现图像自相关分析, 确定图像焦距. 上述基于硬件和辅助装置的方法需要在显微系统上进行硬件的改进, 其成本较高且在当前 WSI 系统中应用较为复杂.

基于图像处理的自动对焦方法, 通常利用多张离焦图像确定焦距. 目前广泛使用的方法是对焦深度法 (depth from focus, DFF)^[16]: 首先将拍摄的图像信号作为反馈, 沿着系统 z 轴采集 N 张等间隔的图像构建 z -stack 图像堆栈; 然后以图像的清晰程度作为评价标准, 构建清晰度评价曲线, 确定曲线

最高点,即清晰度评价指标最高的点为估计的焦距.在数字病理扫描仪中,为平衡扫描速度和成像精度,常采用基于 DFF 的焦点图测量法 (focus map surveying) [4],其工作流程为:首先建立每个子图像的 z -stack 图像堆栈,确定焦点位置;然后将所有子图像的焦点组建为焦点图;最后根据焦点图,进行 x - y 向扫描和 z 向补偿准焦拍照.然而,焦点图测量法需要建立大量的 z -stacks,每个堆栈包含高达 21 张图像,限制了自动对焦的速度,导致了机械误差的积累.尽管可以采用更大采样间距的焦点图,通过插值得到完整的焦点图,但这种焦点下采样方法需要以焦点估计精度降低为代价.另外,病理扫描仪需要 x - y 向扫描两次,第 1 次扫描获得焦点图,第 2 次扫描获得所有准焦子图像.额外的 x - y 向扫描过程降低了准焦子图像的获取速度.

随着人工智能和计算机视觉的快速发展 [17~19],先进的机器学习算法和光学系统成像过程相结合的方法得到了广泛的研究.2018 年, Jiang 等 [20] 首次提出使用深度学习的方法解决自动对焦问题.该方法使用一个 Res-Net50 卷积神经网络,对输入的离焦图像进行处理,输出当前物面与焦面的距离,即离焦距离 ΔD .2019 年, Pinkard 等 [21] 提出了一种基于相关照明光的自准焦方法.这种方法对成像系统的光源进行了修改,利用相干光进行照明,最后使用数据驱动的深度卷积神经网络进行模型训练与学习.但是这种方法需要对显微镜硬件系统进行修改,而且在不同生物样本上的泛化能力一般.2019 年, Wu 等 [22] 提出了一种基于深度学习重对焦算法,能够将二维荧光显微图像重新对焦到指定的成像平面,实现了由二维显微图像到三维图像的重建.然而病理图像比荧光显微图像具有更加复杂的生物组织结构,算法处理较为困难.2020 年, Dastidar 等 [23,24] 提出了将轻量级网络作为自动对焦卷积神经网络结构的方法,将间距固定为 $2\text{ }\mu\text{m}$ 的两张离焦图像的差值图像作为输入,预测当前输入图像的离焦距离.尽管该方法降低了预测离焦距离的误差,但是需要对每个子图像视场进行两次不同 z 向位置的曝光拍照,无法满足 WSI 快速成像的需求.

3 离焦退化模型和病理样本成像模型

传统的自动对焦方法 (如焦点图测量法) 需要不断地重复机械扫描实现调焦,其过程十分繁琐耗时.图 2 展示了不同离焦距离下的离焦病理组织图像和对应的准焦图像.为了降低 WSI 的扫描时间成本,提高系统的便捷性,本文提出了基于深度学习的数字病理扫描系统单次曝光自动对焦方法.该方法从深度学习和图像恢复的角度出发,从原理上避免了重复的机械调焦,通过网络直接输出恢复的准焦图像.本节将讨论 WSI 的显微成像离焦退化模型和病理样本成像模型.

3.1 显微成像离焦退化模型

在光学显微成像中,点扩散函数 PSF 根据 Born & Wolf 模型可知 [25,26]:

$$h(r, \Delta D) = \left| C \int_0^1 J_0 \left(k \frac{\text{NA}}{n} r \rho \right) e^{-\frac{1}{2} i k \rho^2 \Delta D \left(\frac{\text{NA}}{n} \right)^2} \rho d\rho \right|^2, \quad (1)$$

其中 r 为横向半径距离, ΔD 为准焦位置和物面之间的距离,即离焦距离, C 为归一化常量, J_0 为 0 级贝塞尔 (Bessel) 函数第一阶, k 为光源的波数, n 为折射率, i 为虚数, ρ 为归一化出光瞳坐标.随着离焦距离 ΔD 的增大,离焦退化会越来越明显,对应图像的离焦模糊也越来越严重.

WSI 的离焦图像准焦恢复要远比自然图像的去模糊处理困难得多.在自然图像处理领域中,通常认为场景是固定的深度,与相机保持固定的距离,因此在场景的不同横向位置处 PSF 均是相同的.然而在 WSI 成像中,病理样本在不同横向位置的厚度是不同的,其距离的变化导致了 PSF 的不同.为了方便建模分析,本文采用分层景深 (depth of field, DoF) 模型,将连续的深度变化转化为离散的深度

图 2 (网络版彩图) 在不同离焦距离下, 曝光拍照获得的病理组织离焦图像

Figure 2 (Color online) Defocus images at different defocus distances

层, 近似模拟离焦退化成像过程^[27, 28]. 相应地, $h(r, \Delta D)$ 改写成 h_m , 其中 m 代表深度层的位置顺序, h_0 为准焦深度位置下的 PSF. 每个深度层由对应的 PSF 卷积成模糊的深度成像层, 最终通过相机整合后获得真实拍摄的图像. 综上, WSI 的准焦成像模型可以表述为

$$X = \sum_m x_m \otimes h_m, \quad (2)$$

其中, x_m 表示离散化的病理样本深度层, 其数量为 m ; $\otimes$ 代表卷积操作; X 表示相机拍摄的准焦图像, 其对应的准焦物面为 x_0 . 通过公式可知, 准焦图像 X 本质上是三维 PSF 在三维病理样本上积累的结果^[29]. 因此, 数字病理中的准焦图像和自然图像中的清晰图像在成像原理上具有本质的不同, 准焦图像仍然包含离焦成分^[30].

当样本发生 z 向偏移, 即从 x_0 偏移 ΔD 时, 偏移后的准焦物面则为 $x_{\Delta D}$, 对应的第 m 深度层则变为 $x_{m+\Delta D}$. 此时, 相机拍摄的离焦退化图像 Y 可以表述为

$$Y = \sum_m x_{m+\Delta D} \otimes h_m. \quad (3)$$

根据上式离焦退化成像模型可知, 从离焦图像 Y 恢复准焦图像 X , 比自然图像处理的去模糊困难得多, 仅通过离焦退化表达式无法求得准焦图像的闭式解. 因此, 一种合理有效的方法是, 设计数据驱动的深度学习网络模型, 通过学习的方式从离焦图像中恢复对应的准焦图像.

另外, 仿真离焦图像和显微离焦图像的异同点需要明确. 它们的相同点是模糊程度与离焦距离正相关. 不同点包括: (1) 成像原理区别. 自然图像的离焦去模糊采用仿真方法, 通过估计模糊半径进行散焦复原; 自动对焦指硬件成像过程, 即轴向扫描构建 z -stack, 找到准焦图像. 本文用神经网络算法将自动对焦过程进行虚拟化实现. (2) 卷积核区别. 自然图像利用仿真的含模糊半径的模糊核; 自动对焦方法的模糊核为显微系统 PSF. (3) 卷积方式区别. 自然图像利用 2D 信息, 通过仿真实现模糊; 自动对焦采用 3D 病理样本, 在显微镜 3D PSF 的调制下, 在传感器上实现 3D 卷积的 2D 累积. (4) 模糊类型区别. 自然图像通过仿真实现均匀的散焦模糊; 在显微系统离焦时, 折射率的失配导致了球差, 进而产生了明显的伪影效应. 同时, 在球差的影响下, 正、负离焦具有明显非对称的光学特征.

图 3 (网络版彩图) 基于深度学习的数字病理扫描系统单次曝光自动对焦工作流程

Figure 3 (Color online) The workflow of one-shot deep autofocusing in WSI

3.2 基于 WSI 的病理样本成像模型

3.1 小节讨论了在离焦退化过程中, 病理样本厚度的不均匀, 导致了成像过程中对应的 PSF 也不同. 在 WSI 的工作流程中, 需要先确定每个子图像的准焦图像, 进而将其拼接成完整无缝隙的数字病理全切片图像. 因为样本的厚度不均, 所以需要对每个子图像做自动对焦处理. 本文研究发现, 在 WSI 病理样本成像中, 存在两个特有的问题.

(1) 所有子图像的焦点位置, 在样本厚度范围内是不同的. 每个离焦子图像的焦点位置是不同的, 图像包含不同程度的离焦退化模糊. 对于神经网络来说, 同时处理多离焦退化、大离焦范围的离焦图像是困难的. 为此, 本文设计了预准焦策略: 在扫描整个病理样本前, 仅对中心位置的子图像做一次自动对焦即可. 这样能够保证扫描过程中的拍照位置均在病理样本的厚度范围内, 大大缩小了离焦范围. 同时考虑到该预准焦点可能处于临界位置, 根据病理样本的厚度, 经过预准焦操作之后, 整个样本的离焦距离在 $-5 \sim +5 \mu\text{m}$ 之间. 特别地, 当离焦距离为 $+5 \mu\text{m}$ 时, 成像结果虽然模糊, 且包含球差导致的伪影, 但是细胞组织等基本结构仍然保留, 只是小部分细节略有丢失, 仍然在准焦恢复算法的处理范围内, 预期能够获得较好的准焦图像.

(2) 所有子图像的焦点位置, 在样本厚度范围内, 沿着 z 向的分布是不均匀的. 根据统计规律可知^[31], 大部分的焦点集中在病理样本的厚度中心, 且随着焦点逐渐靠近样本表面 (顶面或者底面), 焦点的数量逐渐变少. 为此, 本文设计了单次曝光的扫描成像策略: 首先, 选取样本中心位置的子图像做一次自动对焦, 将该子图像的焦面作为预估的全切片焦面; 然后调节物镜工作距离, 至准焦成像后固定物镜; 最后对所有子图像进行扫描拍照, 拍照位置固定为预估的全切片焦面处. 采用样本中心子图像的焦面作为全切片病理图像的焦面, 能够保证全切片焦面在病理样本厚度中心附近, 并且能够有效降低样本倾斜等制片过程带来的影响.

综上所述, 结合 WSI 工作流程的实际情况, 可以先做一次预准焦, 将离焦距离限制在 $-5 \sim +5 \mu\text{m}$ 范围内, 降低图像的离焦模糊范围; 然后根据样本中心子图像位置, 确定对应子图像的焦面位置作为整个病理样本的拍照位置; 最终通过 $x-y$ 向扫描可获得所有离焦的子图像, 作为网络的输入. 上述两步可合并简化为: 仅需对样本中心子图像做预准焦, 确定焦面即可.

4 单次曝光自动对焦方法网络设计

4.1 单次曝光自动对焦方法流程

基于深度学习的数字病理扫描系统单次曝光自动对焦方法流程如图 3 所示. 首先, 对病理样本中心位置的子图像做预准焦, 获得预估的全切片焦面位置; 然后, WSI 系统执行 $x-y$ 向扫描, 拍照获取所

有离焦子图像, 拍照位置固定为全切片的准焦成像位置; 最后, 将离焦子图像输入神经网络, 通过算法处理得到恢复的准焦子图像. 本方法从原理上替代了传统机械扫描对焦方法, 显著提高了病理切片的获取效率.

4.2 网络设计

本文设计的神经网络采用了直接对输入离焦图像进行准焦复原的端对端方式. 由于不同横向位置子图像的焦点是不同的, 即离焦距离是不同的, 所以输入到网络中的离焦图像的离焦程度也是不同的. 因此, 本文所设计的神经网络需要具备同时处理不同离焦程度的离焦图像的能力. 综上, 本文设计了基于深度学习的数字病理扫描系统单次曝光自动对焦的准焦恢复网络 (one-shot autofocusing network, OAF-Net), 其结构如图 4 所示. 目前, U-Net 网络^[32] 已经广泛应用在各类计算成像应用中, 如光学显微成像^[33,34]、无透镜成像^[35]、深度成像^[36] 等. 因此, 本文的网络结构仍以 U-Net 为主体, 网络前半部分进行特征提取, 后半部分进行图像恢复, 通过学习的方式从离焦图像中恢复对应的准焦图像. 深度网络模型公式如下:

$$X = F(Y_{(\Delta D)}, \theta), \quad (4)$$

其中, ΔD 表示离焦距离, 范围是 $-5 \sim +5 \mu\text{m}$, 间隔 $0.5 \mu\text{m}$, F 代表准焦恢复网络, 网络的输出是随离焦距离变化的离焦图像 Y ; 待恢复的准焦图像为 X ; 网络参数为 θ .

针对离焦退化程度不同的离焦子图像进行准焦恢复问题, 本文所设计的网络需要具备处理不同离焦退化图像的能力. 因此对于模糊程度不同的图像, 在图像特征提取阶段和图像准焦恢复阶段所用的特征图应该具有差异性. 因此, 本文在 U-Net 的基础上设计了与离焦距离相关的通道注意力结构, 称其为离焦注意力. 离焦注意力能够实现网络的特征图随着所提取特征的不同而变化, 更好地处理离焦退化程度不同的离焦图像. 离焦注意力 Block 结构为带通道注意力的卷积结构. 在图 4 的网络结构中, 最后一层的卷积核大小为 5×5 , 其余卷积层的卷积核大小均为 3×3 . 网络使用 Relu 作为激活函数. 图 4 展示了很多尺寸不同的 Block, 其中: 在网络的前半部分 (特征提取部分), 每个 Block 的结构类似, 首先通过两个带通道注意力的卷积加权提取特征, 然后通过最大池化操作将特征图的空间尺寸降低一半; 在网络的后半部分 (准焦恢复部分), 通过一个转置卷积, 将特征图的空间尺度扩大一倍, 然后同样通过两个带通道注意力的卷积进一步实现图像重建.

4.3 网络训练

(1) 训练集. 本文使用 Jiang 开源的病理 z-stack 图像数据库 Dataset1^[20] 来训练准焦恢复网络. 数据集包含 35 个研究级别的苏木精和曙红染色的人类病理切片 (Omano OMSK-HP50), 共包含 162 个病理组织的 z-stacks 子图像. 每个子图像的 z-stack 含有 41 张不同离焦距离的图像, 其离焦步长为 $0.5 \mu\text{m}$, 离焦范围为 $-10 \sim +10 \mu\text{m}$. 其中, 准焦图像通过 Brenner 梯度求极值的方法确定^[37,38].

(2) 训练参数设置. 在网络的训练阶段, 模型使用均方误差 (mean square error, MSE) 作为最终的损失函数. 准焦恢复网络使用学习率为 0.00005 的 Adam 优化器对均方误差损失函数进行梯度下降. 所有参数使用标准差为 0.001 的截断正态分布进行初始化. 整个模型在 NVIDIA GTX2080Ti 显卡上训练了 100 个 epoches, 模型的 batch size 大小为 20, 模型训练时长约为 10 个小时, 模型的大小约为 370 M.

图 4 (网络版彩图) 基于深度学习的数字病理扫描仪单次曝光自动对焦的准焦恢复网络 OAF-Net

Figure 4 (Color online) In-focus recovery network for one-shot autofocusing (OAF-Net) of digital pathology system based on deep learning

5 实验分析

为了对准焦恢复网络 OAF-Net 的性能进行验证, 本文在两个公开的病理数据集上进行网络测试^[20]. 测试集 1 是 Dataset1 中除大部分训练图像之外的图像, 包含 6 种具有不同形态特征的病理组织. 测试集 2 包含 HE 染色的皮肤组织切片, 由美国 University of Connecticut 皮肤病研究部门制片. 样片的来源与训练集不同, 共包含 7 种不同形态特征的病理组织.

5.1 主观效果对比分析

5.1.1 主观效果横向对比分析

图 5 展示了在不同离焦距离的测试离焦图像下, 不同方法之间的效果对比, 分别为基于优化方法的 Dark Channel Prior (Dark)^[39]、Graph Based Deblur (Graph)^[40]、基于学习方法的尺度递归网络 (SRN)^[41] 和本文提出的准焦恢复网络 OAF-Net. 其中, Defocus 代表输入的离焦图像, 其离焦距离如左侧所示; in-focus 代表 GT, 即准焦图像, 对应的离焦距离为 0. 通过实验结果可知, 基于数据驱动的深度学习方法的效果优于传统基于优化迭代的方法: (1) 在传统优化的方法中, Dark Channel Prior 和 Graph Based Deblur 的效果类似. 对于小离焦距离的退化图像 (如 in-focus, 1 μm 图像), 传统方法仅对图像细节进行了增强, 这种过度增强导致了显微图像失真, 不符合实际准焦成像的情况 (如输入为 in-focus 图像下的结果). 对于大离焦距离的退化图像 (如 3 μm , 5 μm 图像), 传统方法并不能获得较好的恢复效果, 离焦模糊严重且不均匀. 由此可知, 在自然图像处理中获得很好效果的传统方法, 在 WSI 的离焦图像准焦恢复问题中并不适用. (2) 在深度学习的方法中, 去模糊网络 SRN 在小离焦距离退化图像的局部细节上, 有过度增强的倾向; 在大离焦距离退化图像上, 有一定的效果, 但仍然存在较明显的离焦退化模糊. 本文提出的准焦恢复网络 OAF-Net, 在小离焦距离退化的图像上能够很好地保持图像细节清晰, 在大离焦距离退化的图像上保留了病理组织全部结构信息, 实现了较好的准焦恢复效果. 另外, 由于显微离焦成像导致了光学像差 (主要为球差) 的非对称现象, 因此正、负离焦退化图

图 5 (网络版彩图) 准焦恢复主观效果横向对比

Figure 5 (Color online) Subjective performance comparison of in-focus recovery methods

像具有非对称的结构特征, 负离焦图像的模糊程度较为均匀, 而正离焦图像的模糊却伴随着大量伪影特征 (如横向条纹等). 对于正离焦退化图像和伪影特征, 上述 3 种对比的方法不能很好地实现准焦恢复. 然而, 本文提出的方法可以在实现准焦恢复的同时, 较好地去除伪影特征. 综上所述, 本文提出的准焦恢复网络 OAF-Net 能够实现多离焦退化、大离焦范围的准焦图像恢复, 并可抑制光学像差导致的伪影现象.

图 6 (网络版彩图) 不同样本下的准焦恢复主观效果对比

Figure 6 (Color online) Subjective in-focus recovery performance comparison of different samples

5.1.2 不同类别病理样本下主观效果分析

图 6 展示了在不同种类的病理样本下, 本文提出的准焦恢复网络 OAF-Net 的对比效果. 不同种类的病理图像分别为 Sample1~Sample6: Sample1 具有明显的细胞核结构; Sample2 染色较浅, 病理组织更透明; Sample3 具有聚集的细胞结构; Sample4 具有明显的环状组织结构; Sample5 具有明显的组织边缘结构; Sample6 具有明显颜色区别的组织结构. 根据实验结果对比可知, 本文提出的准焦恢复网络 OAF-Net 在不同种类病理图像中均取得较好的准焦恢复效果, 对不同结构、颜色、组织的离焦退化图像均能保留比较完整的细节信息. 将网络恢复的准焦图像与实际准焦图像的结果相比, 恢复的准焦图像在成像质量上没有明显的差异, 不影响后续病理组织诊断与分析. 综上所述, 实验证明了本文提出的准焦恢复网络 OAF-Net 在不同种类病理图像上的有效性.

5.1.3 全切片数字病理子图像主观效果分析

为了更全面地评估本文方法的效果, 图 7 展示了不同种类病理样本下的全切片数字病理子图像主观效果. 图 7(a) 表示离焦退化的子图像, 且离焦距离为 $5\ \mu\text{m}$, 即离焦距离最远的图像; 图 7(b) 表示 GT, 即准焦子图像; 图 7(c) 表示离焦退化子图像经过算法处理后恢复的准焦子图像. 在上述图像中, 大框中的图像代表网络的输入 patch, 是小框图像的局部放大. 通过对比可知, 本文提出的准焦恢复网络 OAF-Net 在全切片数字病理子图像上获得了较好的实验效果.

图 7 (网络版彩图) 数字病理子图像准焦恢复主观效果对比 (样本 4)

Figure 7 (Color online) Subjective in-focus recovery performance comparison on the tile of WSI (Sample4). (a) Defocus blur image; (b) focused clear image; (c) deblur result

图 8 (网络版彩图) 数字病理子图像准焦恢复客观效果对比 (PSNR)

Figure 8 (Color online) Objective in-focus recovery performance comparison on the tile of WSI (PSNR)

5.2 客观效果对比分析

为了对准焦恢复效果进行客观评价, 本文使用相同数据集上的 PSNR 作为评价指标. 图 8 展示了测试集上不同方法的 PSNR 结果.

根据图 8 可知, Dark 和 Graph 是传统的基于优化迭代的去模糊算法, 从客观实验效果上看, 其 PSNR 结果偏低, 证明了这些方法并不适用于处理显微图像的离焦退化模糊. SRN 和 Ours 都是基于数据驱动的深度学习方法, 针对显微图像数据集进行训练, 无需对离焦退化成像过程进行数学建模, 因此其结果优于传统基于优化迭代的方法. 根据实验结果可知, 本文提出的准焦恢复网络 OAF-Net 具有最高的 PSNR 值, 证明了本文的方法在处理多离焦退化、大离焦范围和光学像差伪影问题上具有优异的性能. 另外, 算法在测试集 1 上的运行时间分别为: (1) Dark: 8 days; (2) Graph: 18.9 h; (3) SRN: 3.7 min; (4) Ours: 2.1 min. 因此, 本文方法具有更快的推理速度, 能够满足病理扫描系统对高速成像的需求.

5.3 消融实验分析

5.3.1 网络结构分析

为了对本文提出的准焦恢复网络 OAF-Net 结构设计策略的有效性进行验证, 表 1 展示了消融分析实验结果. 其中, Baseline 表示没有加入任何特殊的网络结构, 不含离焦注意力; Baseline+Attention 表示在 Baseline 基础上增加了离焦注意力. 根据表 1 第 1 列和第 2 列可知, 在测试集 1 上, 加入离焦注意力之后, 网络的整体效果均得到了提升. 另外, 表 1 也展示了测试集 2 上的结果. 实验证明, 加入

表 1 测试集 1 和 2 上的消融实验结果对比 (PSNR/SSIM)
Table 1 Ablative analysis on test sets 1 and 2 (PSNR/SSIM)

Defocus distance (μm)	Test set 1		Test set 2	
	Baseline	Baseline+Attention	Baseline	Baseline+Attention
	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
-5	34.53/0.8914	34.92/0.8960	31.07/0.8328	31.58/0.8542
-4	35.26/0.9076	35.66/0.9114	31.89/0.8588	32.42/0.8784
-3	36.21/0.9260	36.55/0.9281	32.99/0.8875	33.61/0.9059
-2	37.23/0.9422	37.50/0.9423	33.85/0.9073	34.59/0.9235
-1	38.51/0.9574	38.90/0.9581	34.29/0.9196	35.20/0.9340
0	42.12/0.9781	44.09/0.9813	34.79/0.9341	36.27/0.9474
+1	39.11/0.9593	39.91/0.9637	33.67/0.9162	35.03/0.9316
+2	37.22/0.9312	37.94/0.9382	32.41/0.8867	34.02/0.9100
+3	35.67/0.9029	36.30/0.9116	31.24/0.8537	33.01/0.8871
+4	34.49/0.8814	35.02/0.8902	30.35/0.8245	32.16/0.8650
+5	33.62/0.8658	34.11/0.8746	29.26/0.7903	30.60/0.8246
Average	36.99/0.9265	37.64/0.9310	32.41/0.8757	33.59/0.8985

图 9 不同离焦距离下含注意力机制的特征图主观效果对比

Figure 9 Subjective performance comparison on features with channel attention at different defocus distances

离焦注意力之后, 网络在新的数据集上的泛化能力也有所提高. 另外, 图 9 展示了不同离焦距离下的特征图, 分别为第 3 ($56 \times 56 \times 256$), 5 ($28 \times 28 \times 512$) 和 7 ($14 \times 14 \times 1024$) 层的带注意力的卷积层输出. 通过对比可知, 注意力机制在准焦、离焦及不同离焦距离下, 可以较好地实现离焦特征的加权提取. 综上所述, 实验证明了本文设计的离焦注意力的有效性.

5.3.2 泛化能力分析

通过表 1 可知, 网络性能在测试集 2 上, 具有一定程度的下降. 测试集 1 为购置的病理公司切片,

图 10 (网络版彩图) 在不同数据集和不同损失函数 (MAE/MSE) 上的准焦恢复客观效果对比

Figure 10 (Color online) Objective in-focus recovery performance comparison on different datasets and loss functions (MAE/MSE)

测试集 2 为学校自制的病理切片. 两种测试集均在相同显微镜、相同条件下 (如光照强度) 拍摄获得. 两种测试集在病理样本结构上 (均为不相同的组织)、H&E 染色上 (不同公司染色程度不同) 以及病理切片制片上 (封蜡等效果不同) 均存在较大区别, 因此网络性能在测试集 2 上不可避免地存在一定程度的下降. 然而, 病理组织切片具有结构特征明显、种类固定有限、纹路分布相似等特点, 因此可以通过适当扩展训练集等方式, 进一步提升网络的泛化能力.

5.3.3 损失函数分析

为了对本文提出的准焦恢复网络 OAF-Net 损失函数设计策略的有效性进行验证, 图 10 展示了不同损失函数 MAE 和 MSE 在不同测试集上的效果对比. 通过对比可知, 在测试集 1 上, MAE 和 MSE 的效果近似相等; 在测试集 2 上, MSE 的效果明显优于 MAE. 实验证明, 在神经网络自动对焦任务上, 采用 MSE 损失函数可以获得更好的效果.

6 结论

本文设计了一种基于深度学习的数字病理扫描系统单次曝光自动对焦方法. 该算法的输入为离焦模糊图像, 通过算法处理后直接获得对应的准焦图像, 从原理上替代了传统方法对逐个子图像进行机械扫描调焦. 本文首先分析了显微成像离焦退化模型和基于 WSI 的病理样本成像模型, 设计了预准焦策略; 然后, 对准焦恢复网络进行设计, 针对不同离焦距离下的离焦图像模糊程度不同, 设计了包含离焦注意力的 OAF-Net; 最后, 主、客观实验结果证明, 本文设计的准焦恢复网络, 在保证成像质量的前提下, 能够快速处理离焦模糊图像, 获得准焦恢复结果.

参考文献

- 1 Meng G. Discussion on the systemic problems and countermeasures in the development of pathology industry in mainland China. *Chin J Clin Exp Pathol*, 2012, 28: 5–6 [孟刚. 中国大陆病理行业发展存在的体制性问题及对策探讨. *临床与实验病理学杂志*, 2012, 28: 5–6]
- 2 Pantanowitz L, Valenstein P N, Evans A J, et al. Review of the current state of whole slide imaging in pathology. *J Pathol Inform*, 2011, 2: 36
- 3 Weinstein R S, Graham A R, Richter L C, et al. Overview of telepathology, virtual microscopy, and whole slide imaging: prospects for the future. *Human Pathol*, 2009, 40: 1057–1069

- 4 Higgins C. Applications and challenges of digital pathology and whole slide imaging. *Biotechnic Histochem*, 2015, 90: 341–347
- 5 Abels E, Pantanowitz L, Aeffner F, et al. Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the digital pathology association. *J Pathol*, 2019, 249: 286–294
- 6 Zhang B. Pathology development in the era of precision medicine. *Medical J Peking Union Medical College Hospital*, 2017, 8: 117–121 [张波. 精准医疗时代的病理学发展. *协和医学杂志*, 2017, 8: 117–121]
- 7 Chen M W. Analysis of digital slice scanning technology. *Opt Instr*, 2017, 39: 7–11 [陈木旺. 数字切片扫描技术浅析. *光学仪器*, 2017, 39: 7–11]
- 8 Abels E, Pantanowitz L. Current state of the regulatory trajectory for whole slide imaging devices in the USA. *J Pathol Inform*, 2017, 8: 23
- 9 Langehanenberg P, von Bally G, Kemper B. Autofocusing in digital holographic microscopy. *3D Res*, 2011, 2: 4
- 10 Kohlberger T, Liu Y, Moran M, et al. Whole-slide image focus quality: automatic assessment and impact on AI cancer detection. *J Pathol Inform*, 2019, 10: 39
- 11 Bian Z C, Guo C F, Jiang S W, et al. Autofocusing technologies for whole slide imaging and automated microscopy. *J Biophotonics*, 2020, 13: e202000227
- 12 Gurcan M N, Boucheron L E, Can A, et al. Histopathological image analysis: a review. *IEEE Rev Biomed Eng*, 2009, 2: 147–171
- 13 Liron Y, Paran Y, Zatorsky N G, et al. Laser autofocus system for high-resolution cell biological imaging. *J Microsc*, 2006, 221: 145–151
- 14 Montalto M C, McKay R R, Filkins R J. Autofocus methods of whole slide imaging systems and the introduction of a second-generation independent dual sensor scanning method. *J Pathol Inform*, 2011, 2: 44
- 15 Liao J, Jiang Y T, Bian Z C, et al. Rapid focus map surveying for whole slide imaging with continuous sample motion. *Opt Lett*, 2017, 42: 3379–3382
- 16 Subbarao M, Tyan J K. Selecting the optimal focus measure for autofocus and depth-from-focus. *IEEE Trans Pattern Anal Machine Intell*, 1998, 20: 864–870
- 17 Kou Z, Wu J F, Wang H L, et al. Obstacle visual sensing based on deep learning for low-altitude small unmanned aerial vehicles. *Sci Sin Inform*, 2020, 50: 692–703 [寇展, 吴健发, 王宏伦, 等. 基于深度学习的低空小型无人机障碍物视觉感知. *中国科学: 信息科学*, 2020, 50: 692–703]
- 18 Wang G F, Huang Y D. Medical-image-fusion algorithm based on a detail-enhanced and pulse-coupled neural-network model stimulated by parallel features. *Sci Sin Inform*, 2020, 50: 239–260 [王国芬, 黄永东. 基于细节增强和平行特征刺激脉冲耦合神经网络的医学图像融合算法. *中国科学: 信息科学*, 2020, 50: 239–260]
- 19 Xu H Y, Yao N M, Peng X L, et al. A multi-pose face frontalization method based on encoder-decoder network. *Sci Sin Inform*, 2019, 49: 450–463 [徐海月, 姚乃明, 彭晓兰, 等. 基于编解码网络的多姿态人脸图像正面化方法. *中国科学: 信息科学*, 2019, 49: 450–463]
- 20 Jiang S W, Liao J, Bian Z C, et al. Transform- and multi-domain deep learning for single-frame rapid autofocus in whole slide imaging. *Biomed Opt Express*, 2018, 9: 1601–1612
- 21 Pinkard H, Phillips Z, Babakhani A, et al. Deep learning for single-shot autofocus microscopy. *Optica*, 2019, 6: 794–797
- 22 Wu Y C, Rivenson Y, Wang H D, et al. Three-dimensional virtual refocusing of fluorescence microscopy images using deep learning. *Nat Methods*, 2019, 16: 1323–1331
- 23 Dastidar T R. Automated focus distance estimation for digital microscopy using deep convolutional neural networks. In: *Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019. 1049–1056
- 24 Dastidar T R, Ethirajan R. Whole slide imaging system using deep learning-based automated focusing. *Biomed Opt Express*, 2020, 11: 480–491
- 25 Born M, Wolf E. Principles of optics: electromagnetic theory of propagation, interference and diffraction of light. *Phys Today*, 2000, 53: 77–78
- 26 Hosseini M S, Brawley-Hayes J A Z, Zhang Y, et al. Focus quality assessment of high-throughput whole slide imaging in digital pathology. *IEEE Trans Med Imag*, 2020, 39: 62–74
- 27 Scofield C. 1.8-2 $\frac{1}{2}$ -d depth-of-field simulation for computer animation. *Graph Gems III (IBM Version)*, 1992: 36–38

- 28 Wu Y C, Boominathan V, Chen H J, et al. PhaseCam3D-learning phase masks for passive single view depth estimation. In: Proceedings of 2019 IEEE International Conference on Computational Photography (ICCP), 2019. 1–12
- 29 Gao Z P, Zhai G T, Deng H W, et al. Extended geometric models for stereoscopic 3D with vertical screen disparity. Displays, 2020, 65: 101972
- 30 Zhai G T, Min X K. Perceptual image quality assessment: a survey. Sci China Inf Sci, 2020, 63: 211301
- 31 Hart M, Barkhouser R H, Carr M, et al. Focal plane alignment and detector characterization for the Subaru prime focus spectrograph. In: Proceedings of High Energy, Optical, and Infrared Detectors for Astronomy VI, 2014. 9154
- 32 Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Proceedings of International Conference on Medical Image Computing and Computer-assisted Intervention, 2015. 234–241
- 33 Rivenson Y, Göröcs Z, Günaydin H, et al. Deep learning microscopy. Optica, 2017, 4: 1437–1443
- 34 Li Q, Liu X M, Jiang J J, et al. Rapid whole slide imaging via dual-shot deep autofocusing. IEEE Trans Comput Imag, 2021, 7: 124–136
- 35 Barbastathis G, Ozcan A, Situ G. On the use of deep learning for computational imaging. Optica, 2019, 6: 921–943
- 36 Chang J, Wetzstein G. Deep optics for monocular depth estimation and 3D object detection. In: Proceedings of the IEEE International Conference on Computer Vision, 2019. 10193–10202
- 37 Sun Y, Duthaler S, Nelson B J. Autofocusing algorithm selection in computer microscopy. In: Proceedings of 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2005. 70–76
- 38 Yazdanfar S, Kenny K B, Tasimi K, et al. Simple and robust image-based autofocusing for digital microscopy. Opt Express, 2008, 16: 8670–8677
- 39 Pan J S, Sun D Q, Pfister H, et al. Blind image deblurring using dark channel prior. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016. 1628–1636
- 40 Bai Y C, Cheung G, Liu X M, et al. Graph-based blind image deblurring from a single photograph. IEEE Trans Image Process, 2019, 28: 1404–1418
- 41 Aittala M, Durand F. Burst image deblurring using permutation invariant convolutional neural networks. In: Proceedings of the European Conference on Computer Vision (ECCV), 2018. 731–747

Whole slide imaging via one-shot deep autofocusing

Qiang LI^{1,2}, Xianming LIU^{1,2*}, Kaige HAN¹, Junjun JIANG^{1,2} & Xiangyang JI³

1. Faculty of Computing, Harbin Institute of Technology, Harbin 150001, China;

2. Peng Cheng Laboratory, Shenzhen 518052, China;

3. Department of Automation, Tsinghua University, Beijing 100084, China

* Corresponding author. E-mail: csxm@hit.edu.cn

Abstract Whole slide imaging (WSI) is the core technology of pathology digitization, and the performance of WSI is determined by the speed and accuracy of autofocusing. However, the conventional autofocusing methods need to shoot multiple defocus tiles to create an image stack, or require a complex hardware modulation optical system, thereby restricting the applications in the real scenario of WSI. In this paper, we develop a deep one-shot autofocusing network of WSI for tile-wise virtual autofocusing to generate the in-focus image directly from one possible defocus image. Our method works with one image taken at any defocus distance, avoiding repeated focusing movements and camera shooting processes in principle. Experimental results demonstrate that our scheme enjoys the merits of high throughput, high speed, low cost, compatibility and offline processing.

Keywords digital pathological systems, autofocusing, deep learning, optical microscopes, computational imaging

Qiang LI received his B.S. and M.S. degrees from the School of Electrical Engineering and Automation, Harbin Institute of Technology (HIT), Harbin, China, in 2015 and 2018, respectively. From 2018, he is studying for a doctorate in computer science at HIT. His research interests include computational imaging, optical microscopy, and image processing.

Xianming LIU is a professor, a Ph.D. advisor at the Department of Computer Science, Harbin Institute of Technology (HIT), Harbin, China. He received his Ph.D. degree in computer science from HIT in 2012. From December 2012 to December 2013, he worked as a post-doctoral fellow at McMaster University, Hamilton, Canada. From January 2014 to March 2017, he worked as a project researcher at National Institute of Informatics (NII), Tokyo, Japan. His research interests include image/video coding, image/video processing, and graph signal processing.

Junjun JIANG received his B.S. degree from Department of Mathematics, Huaqiao University, Quanzhou, China, in 2009, and his Ph.D. degree from School of Computer, Wuhan University, Wuhan, China, in 2014. He is currently a professor at the School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China. His research interests include image processing and computer vision.

Xiangyang JI received his B.S. degree in materials science and his M.S. degree in computer science from Harbin Institute of Technology, Harbin, China, in 1999 and 2001, respectively, and his Ph.D. degree in computer science from Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. He is currently a professor at the Department of Automation, Tsinghua University. His current research interests include signal processing, image/video compressing, and intelligent imaging.