

论文标题

简要说明

图像摘要

Highlights

总结

Introduction

Result

神经网络从单个蛋白质图像中准确预测牵引力

作者解释神经网络架构

用zyxin训练的神经网络效果优于其他的细胞骨架蛋白

经过zyxin蛋白训练的神经网络能够推广到新的细胞类型以及对抗一定的生物干扰

对抗一定生物化学干扰的实验

神经网络识别细胞粘附和形态特征

物理瓶颈神经网络（物理理论限制下）：学习粘连增强有效的弹性模型

模型推导

Green's function neural networks: 不考虑物理学模型构建揭示长度尺度和有效方程

DISCUSSION

method detail

Data processing

Training data

U-Net architecture

Synthetic cells

Physical Bottleneck Forward and Backward Passes

参考文献

论文标题

Machine learning interpretable models of cell mechanics from protein images

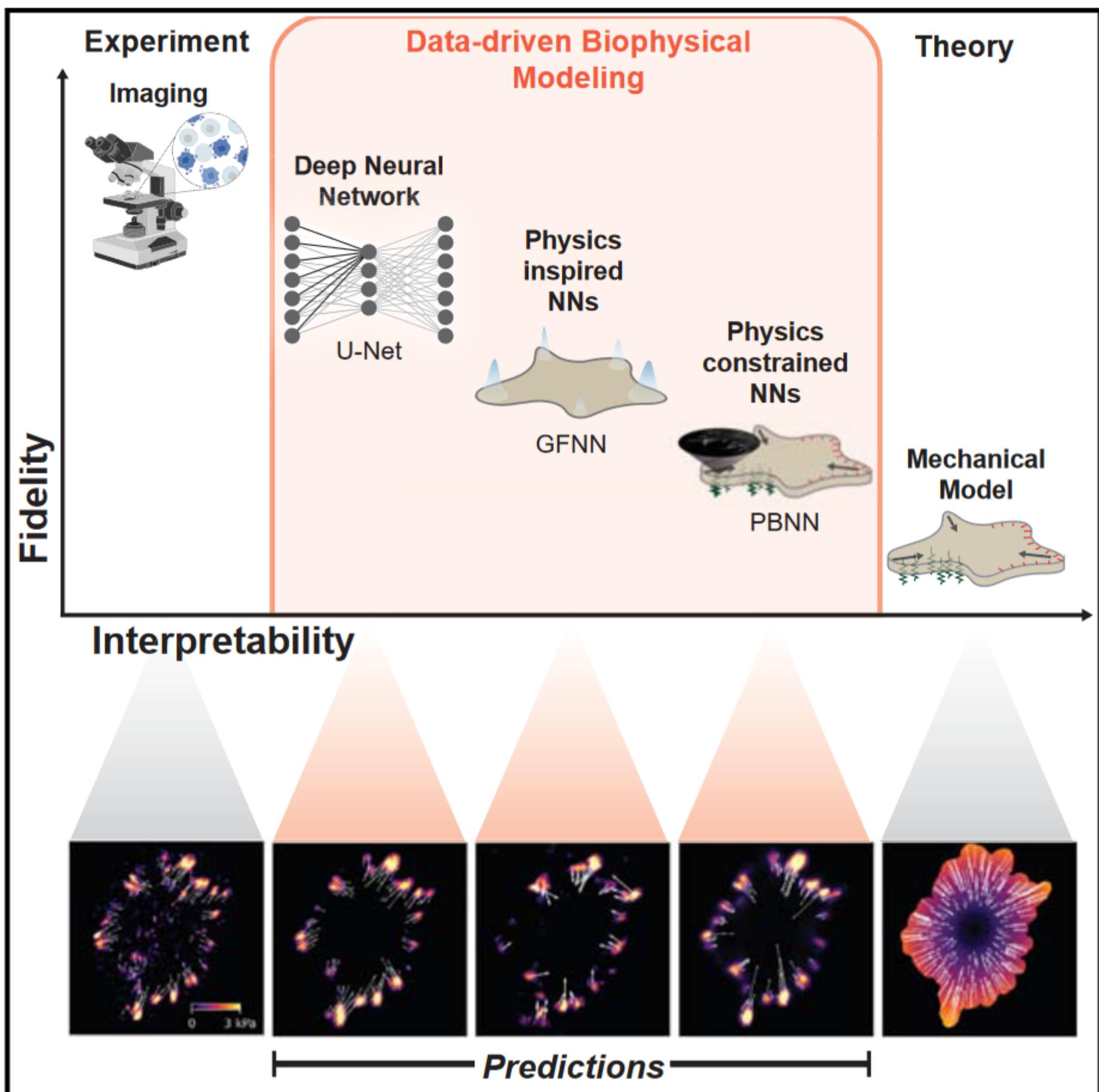
从蛋白质图像中建立机器学习可解释的细胞力学模型。

简要说明

基于图像的深度学习方法可以从单个细胞粘附蛋白（a single focal adhesion protein）的实验图像中预测细胞的机械行为，这些可泛化的细胞力学模型可以推动对细胞粘附的理解和控制。

图像摘要

Graphical abstract



纵轴是表示精度（是机械行为的精度还是图像的精度？）

横轴是表示可解释性。

最左侧表示的是实验方法，最右侧表示的是理论方法，中间部分表示的是数据驱动的生物物理学模型。

（从左往右）从相关的实验中获得图像，此时的图像是精度最高可解释性最低的；第二个是U-Net神经网络模型；第三个是物理学启发的神经网络模型；第三个是物理学相关原理限制的神经网络模型；最后是力学模型。

从神经网络predict的图像可以看出，精度是有所下降的。

Highlights

1. 神经网络预测单个粘附蛋白区域的牵引力情况。
2. 利用神经网络可以泛化到未见过的细胞、细胞类型、以及有生化干扰的情况下。
3. 无论是基于物理约束的方法还是不考虑特定物理规律的方法，都学习出了可解释的规则来预测力（的作用、效果？）。
4. 牵引力通过（根据）粘附斑（FAs）和细胞形状相关的两个长度（范围）尺度进行编码。

总结

- 当前存在的问题：没有存在从细胞分子成分推断细胞的大规模物理性质的系统策略，导致人们在理解细胞粘附和迁移过程存在障碍。

- 理论依据：细胞的形式和功能都源于细胞质内复杂的生化系统。
- 解决方法：作者开发了一种数据驱动的建模方法去学习粘附细胞的机械行为。
- 方法简述：首先训练神经网络，从细胞骨架蛋白的图像中预测细胞力。令人惊讶的是，单焦点粘附蛋白（例如zyxin）的实验图像足以预测力，并且可以推广到看不见的生物状态。利用这样一个观察结果开发了两种方法，一种是受物理学约束的神经网络，另一种是与物理学无关（大概是启发式的）来构建细胞力的数据驱动连续体模型。
- 两种方法都揭示了细胞力是如何由两个不同的长度尺度编码的。除了粘附细胞力学，作者的工作还可以作为将神经网络集成到细胞生物学预测模型中的案例研究。

Zyxin是一种蛋白质，它在细胞内起着重要作用，特别是在细胞骨架的调节和细胞运动方面。这种蛋白通常位于细胞核和细胞质之间的区域，参与细胞的粘附、迁移和信号传导等过程。在细胞粘附斑（focal adhesions）中，Zyxin能够与其他蛋白质相互作用，调节细胞与外界基质之间的相互作用，并参与细胞的力学性质调控。

Introduction

活细胞的结构和动态受到细胞骨架物理特性的控制。然而，细胞骨架本身是复杂生化回路的产物，这些回路调节其动态和空间组织。研究细胞的物理生物学所面临的主要挑战是解开物理学和生物化学之间的相互作用。目前的建模方法在很大程度上依赖于几个世纪古典连续力学工作所形成的直觉，其中对称性和守恒定律不仅决定了这些模型中出现的变量，还决定了它们遵循的方程。然而，细胞显然是非古典的，而是依赖于分布式的酶活性和跨多个尺度层次的非平衡机化学过程。例如，细胞内的力不仅仅是为了恢复局部变形，还因为受到生化信号网络调控的持续重塑而产生。这种物理和生物化学的相互作用使得用少数简单理解的集体变量对细胞动态进行粗粒化和系统参数化变得更加复杂。

机器学习有潜力通过将现有的物理模型与生化信息相结合，甚至直接从数据的统计信息中发现新的模型来解决这个挑战。这些工具在结构生物学中非常成功，可以直接从基因序列预测蛋白质结构。在这里，我们展示了机器学习方法在一个经典的细胞生物学问题中的能力：**细胞骨架蛋白如何影响细胞的力学特性。（Zyxin）**

细胞产生收缩力，这些力是细胞形状、粘附、运动和机械转导的关键调节因子。由肌动蛋白细胞骨架产生的力通过跨膜的焦点粘附（FAs）传递到细胞外基质，在那里可以使用牵引力显微镜（TFM）等技术直接测量这些力。TFM测量结合荧光标记的细胞骨架蛋白的活细胞成像已经帮助开发了许多细胞力学生成和机械感知的生物物理模型。虽然这些模型提供了对各种局部微观机制的洞见，但它们并未捕捉到细胞中广泛的结构和行为的多样性。因此，**它们不能完全解释非局部和细胞尺度的属性，比如细胞形态和FA（focal adhesion，粘附斑）结构以及位置是如何影响甚至主导局部力的。**

作者利用神经网络改进现有的细胞力学模型，并发现新的模型。

流程如下：

首先通过训练深度神经网络，直接从荧光标记的细胞骨架蛋白的图像中预测力，并在这个过程中发现，一个单一的粘附斑蛋白，比如zyxin或paxillin，就足以预测牵引应力。

仅凭单一蛋白分布就能做出准确预测，并不意味着其他蛋白在力的产生中是生化上多余的。相反，这表明**预测细胞力大小和方向所需的最小信息量，以及模型的最小复杂度。**这些预测是稳健的，因为作者发现神经网络可以泛化到先前未见的实验和生物学干扰。

为了理解这种泛化能力，作者探究神经网络以识别影响其预测的特征，并进一步引导两种互补的数学模型的构建。

1. 首先，我们引入了一种物理约束的机器学习方法，增强了现有的机械细胞模型：它直接从数据中学习，了解单个测量的蛋白质分布如何设置有效线性弹性模型的物理含义参数
2. 其次，作者放弃了原有的机械假设，并展示了一个纯粹基于数据的流程，构建了相关领域并提炼了有效方程，预测细胞牵引应力。

尽管包含了不同程度的模型复杂性和先验知识，但作者所有的方法都一致显示出，**力生成模型的特点是局部和非局部特征的相互作用。**

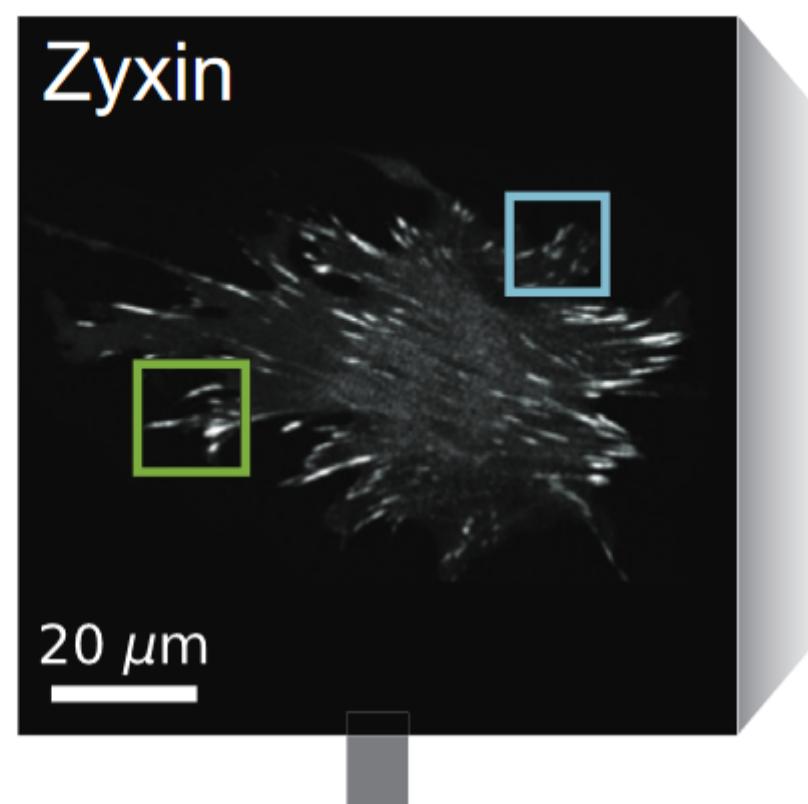
作者的发现说明了粘附斑蛋白如何通过它们在细胞内的分布来编码粘附点的局部力量信息，以及整个细胞的收缩性，并展示了一系列互补的方法来构建新的生物系统模型。

Result

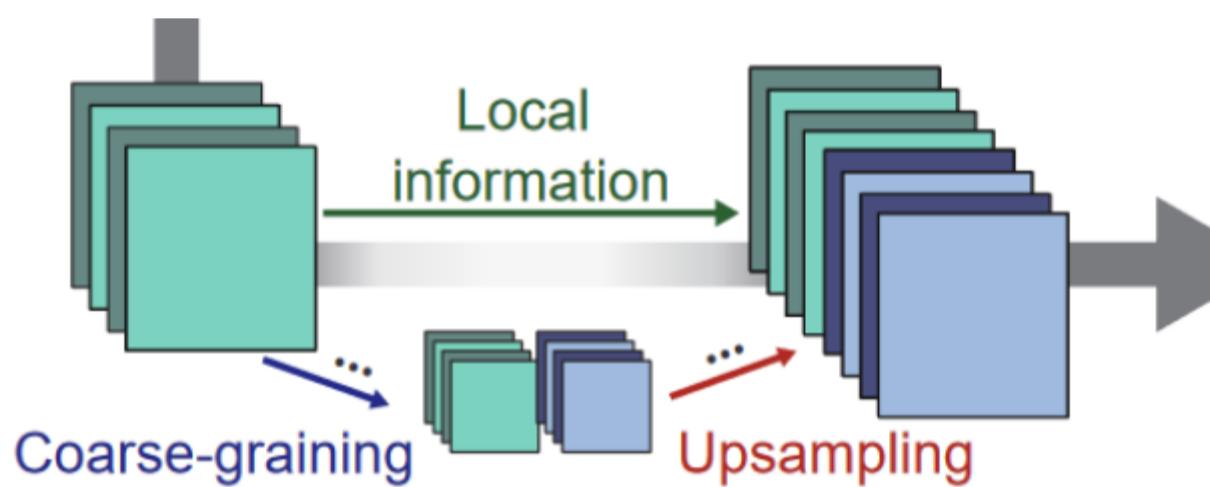
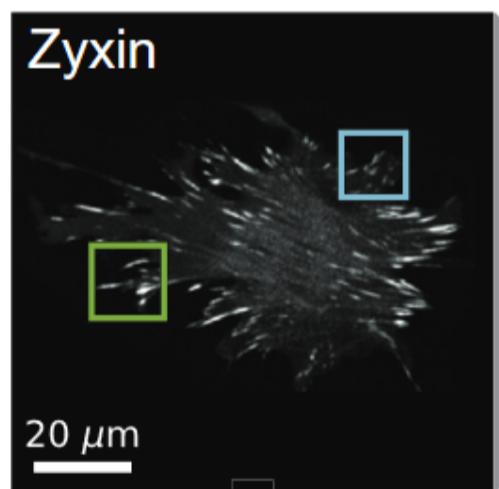
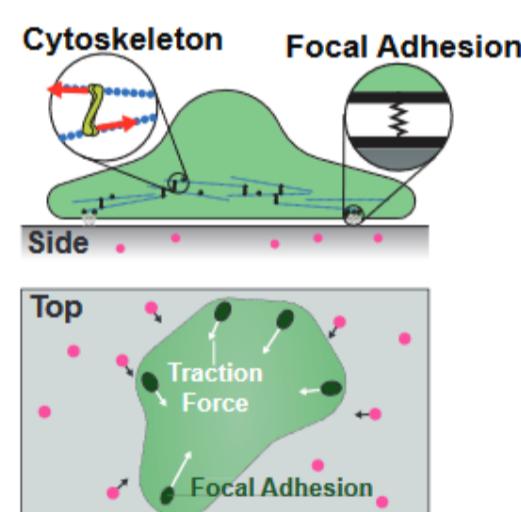
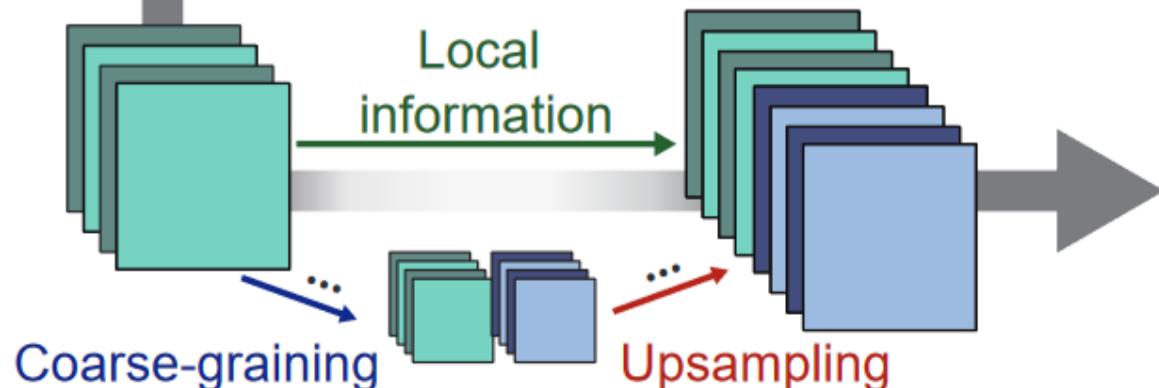
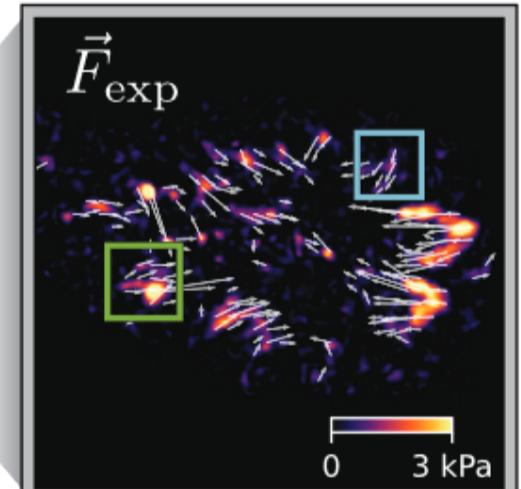
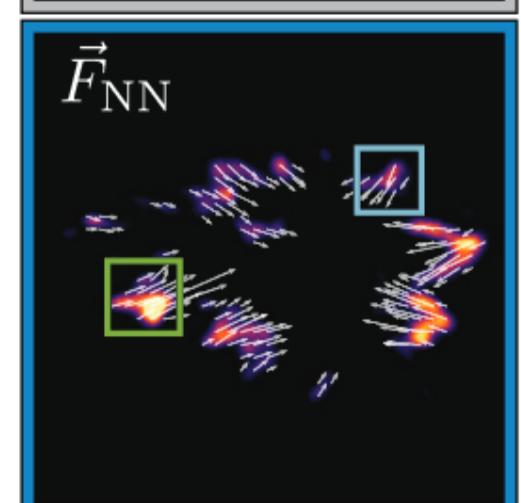
神经网络从单个蛋白质图像中准确预测牵引力

为了评估神经网络是否可以从生化领域进行机械预测，我们将通过将成纤维细胞中的粘附斑蛋白zyxin的荧光显微镜图像与其通过TFM牵引力显微镜直接测量的相应牵引力配对来创建一个库。

- 以下是荧光显微镜图像

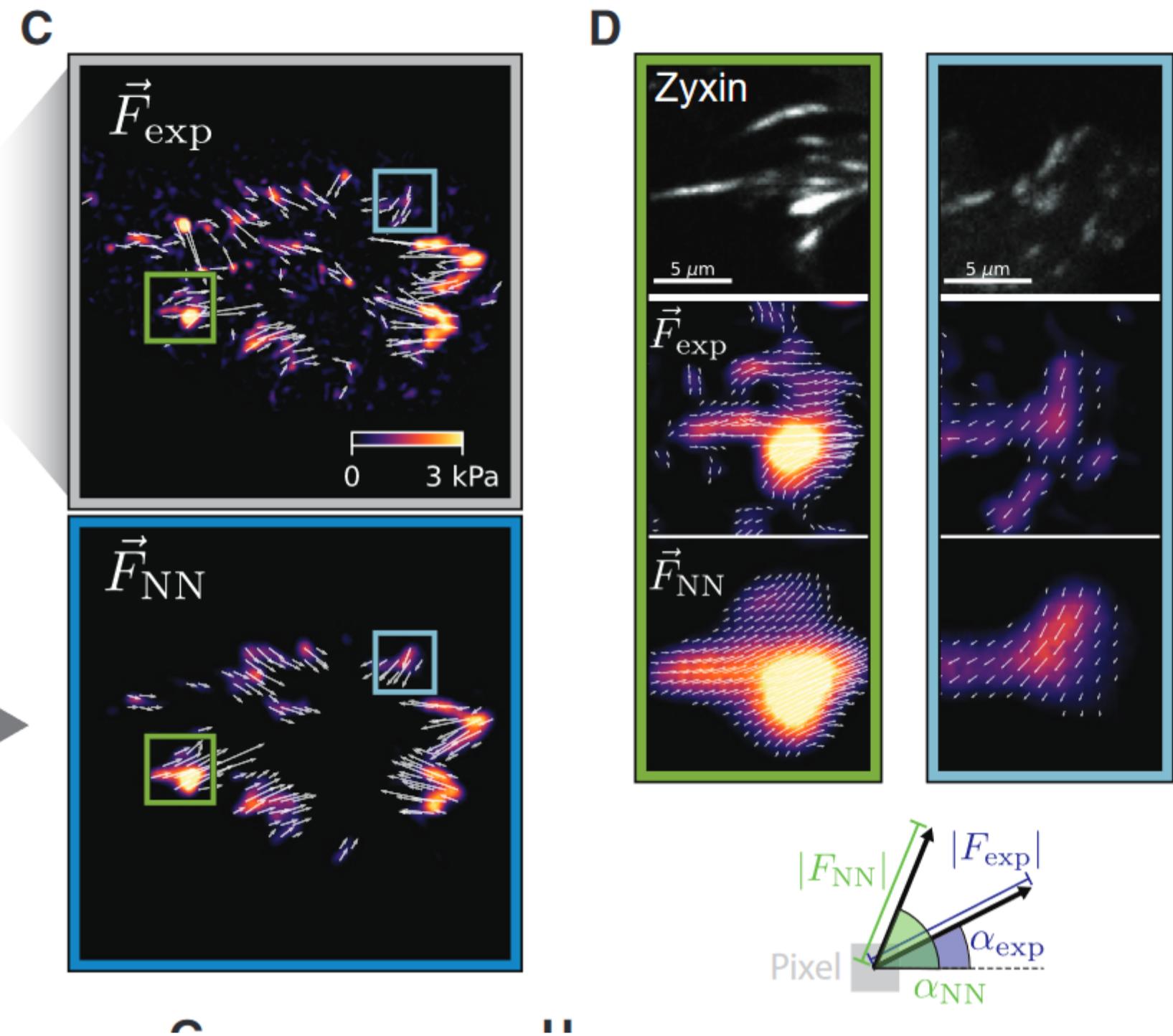
A

- 将蛋白的荧光显微镜图像做Coarse-graining粗粒化处理，然后将处理过的图像再做upsampling处理，已增加其细节或提高其质量，local information应该是TFM原子力显微镜本地测量的力学信息，然后配对形成一个库，最后得到的内容就是Fexp图像所示的内容，其颜色表示不同程度的应力大小。
- B图是示意细胞骨架与焦点粘附的相关牵引力

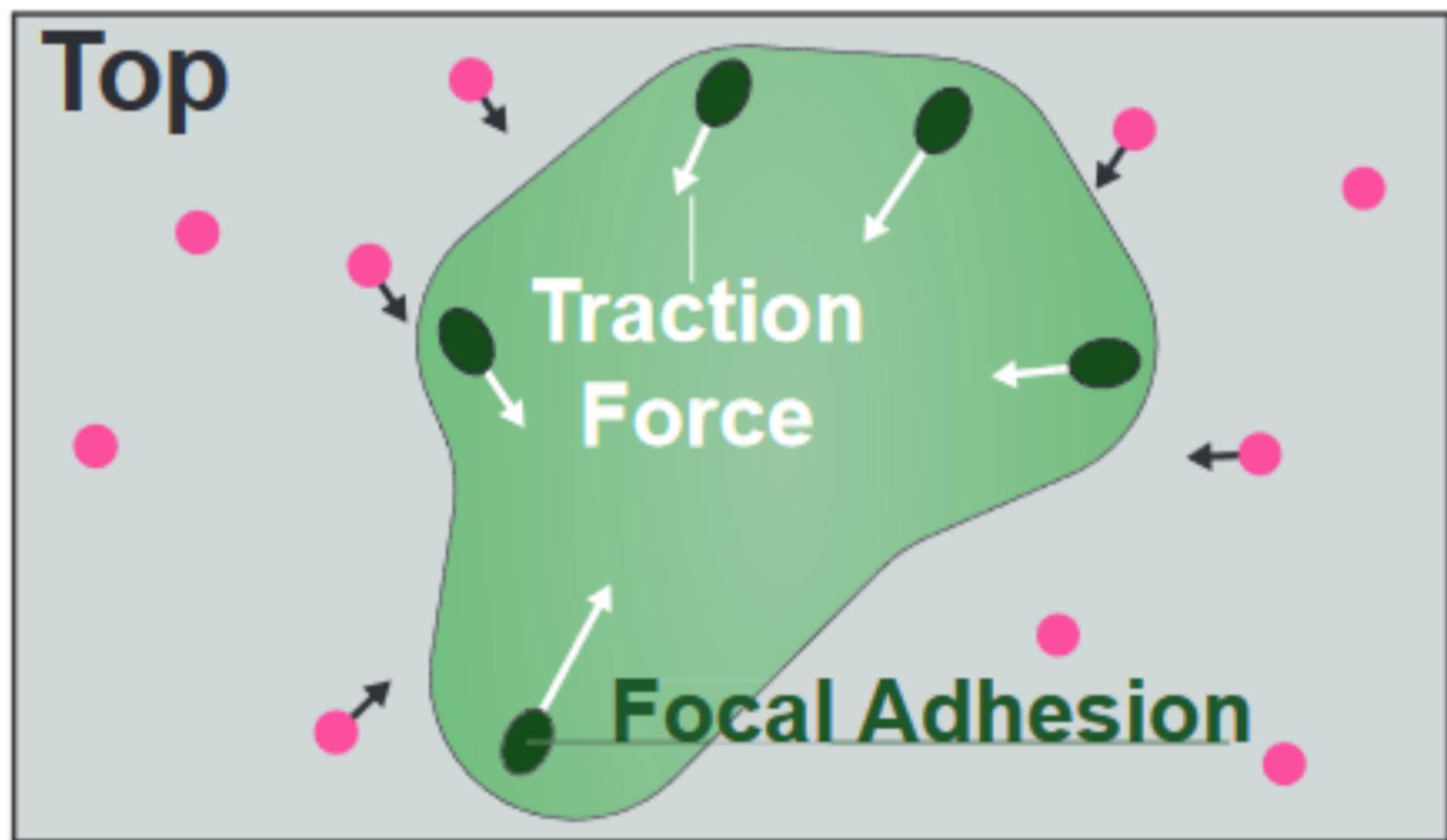
**A****B****C** \vec{F}_{NN} 

总而言之，作者的数据库包含了从31个不同的细胞时间序列中获得的图像，这些图像来自表达zyxin的细胞及其相关的牵引力场（详见STAR Methods）。

在这些细胞中，牵引力主要集中在细胞边界的粘附斑上，这些粘附斑由zyxin聚集标记，并朝向细胞内部指向（见图下图C和D）。



参考如图



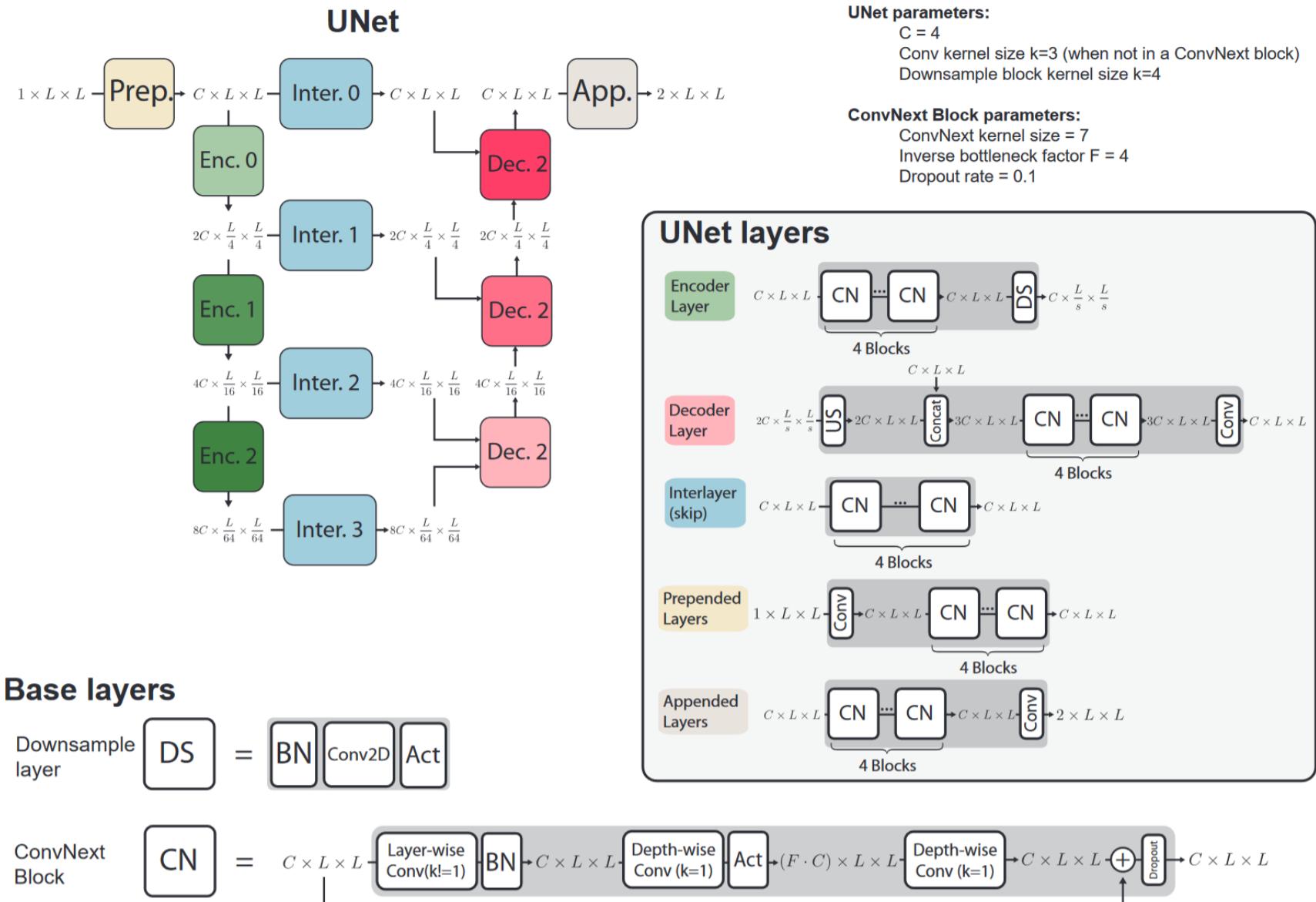
作者解释神经网络架构

选这里U-net结构作为骨干架构，通过连续的步长卷积学习大尺度特征，同时层与层之间的跳跃连接传递细粒度信息，并保留在粗粒化Coarse-graining过程中可能丢失的局部结构。

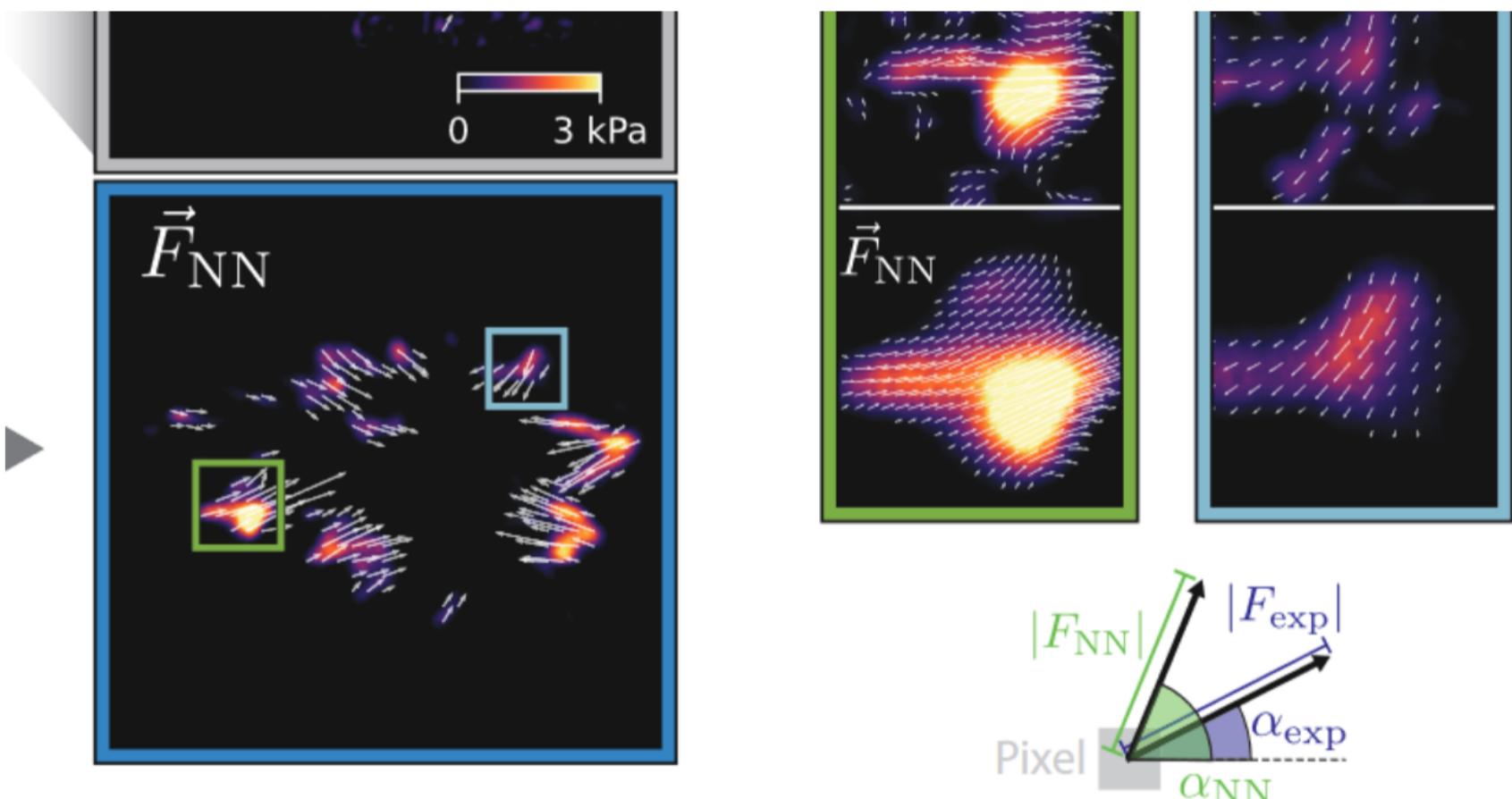
Strided Convolution就是在过滤器移动过程中进行跳格计算，stride具备跳格能力，避免重复计算。stride也会影响feature map的规模。

虽然U-Net在解决TFM inverse problem上已经证明了成功，但在这里，我们推动它们来连接生物化学和力学。我们通过ConvNext模块增强这个骨干结构，以提高准确性和训练效率。

TFM (Traction Force Microscopy, 牵引力显微镜) 的反问题是根据细胞表面的形变信息反推细胞内产生的力的分布情况。通过观察细胞在基质上的变形，可以推断出在细胞与基质之间的牵引力分布情况。这个问题通常是通过测量细胞周围表面的位移场或变形场来解决，然后使用相应的物理模型或计算方法来反推产生这些变形的力的大小和方向。TFM的反问题在细胞力学和生物物理学研究中是一个重要的课题，可以帮助理解细胞对其周围环境施加的力，并揭示细胞行为和功能方面的信息。

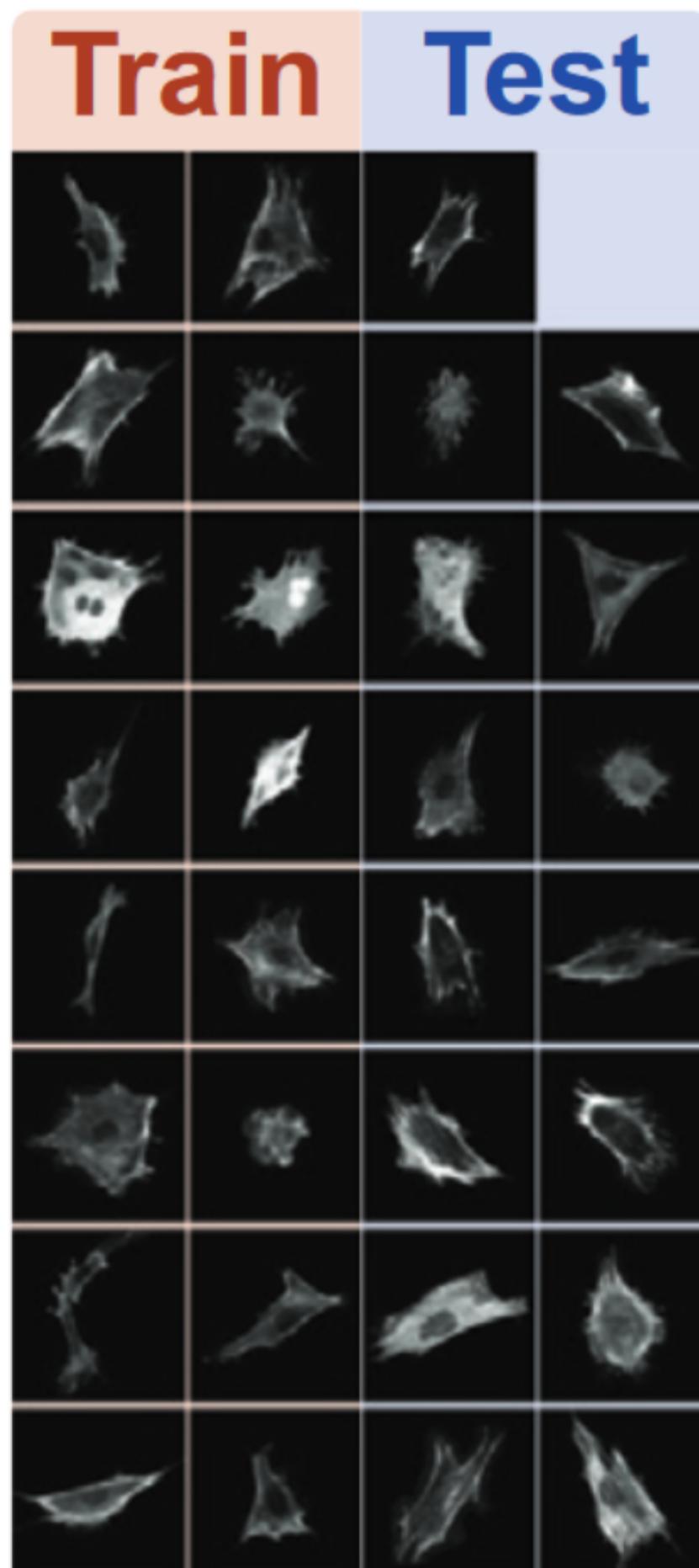


我们训练U-net直接的预测牵引力，使用数据库中配对好的zyxin图像作为输入，即FNN



数据库被分为训练集和测试集，分别包含16个和15个细胞，如图G所示

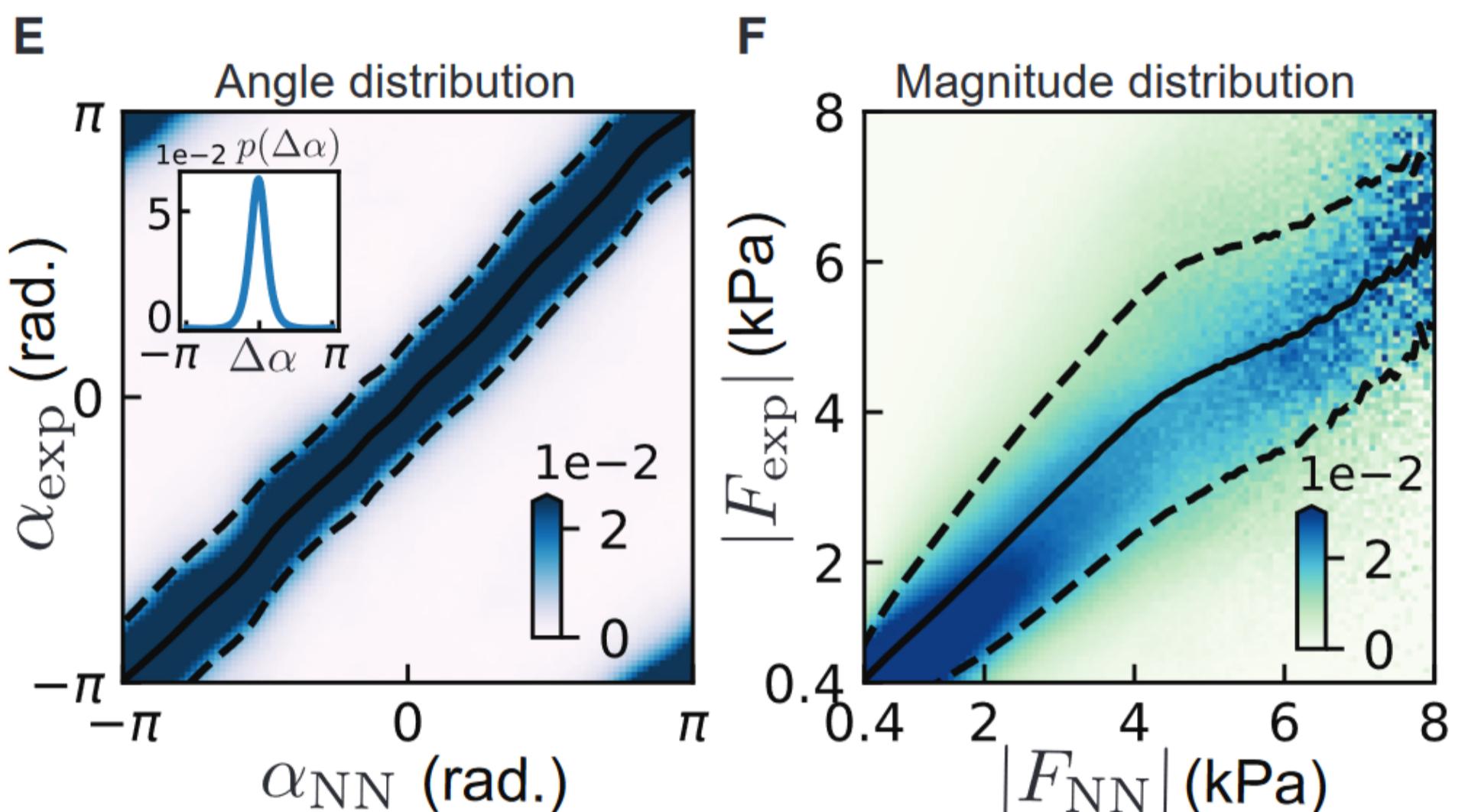
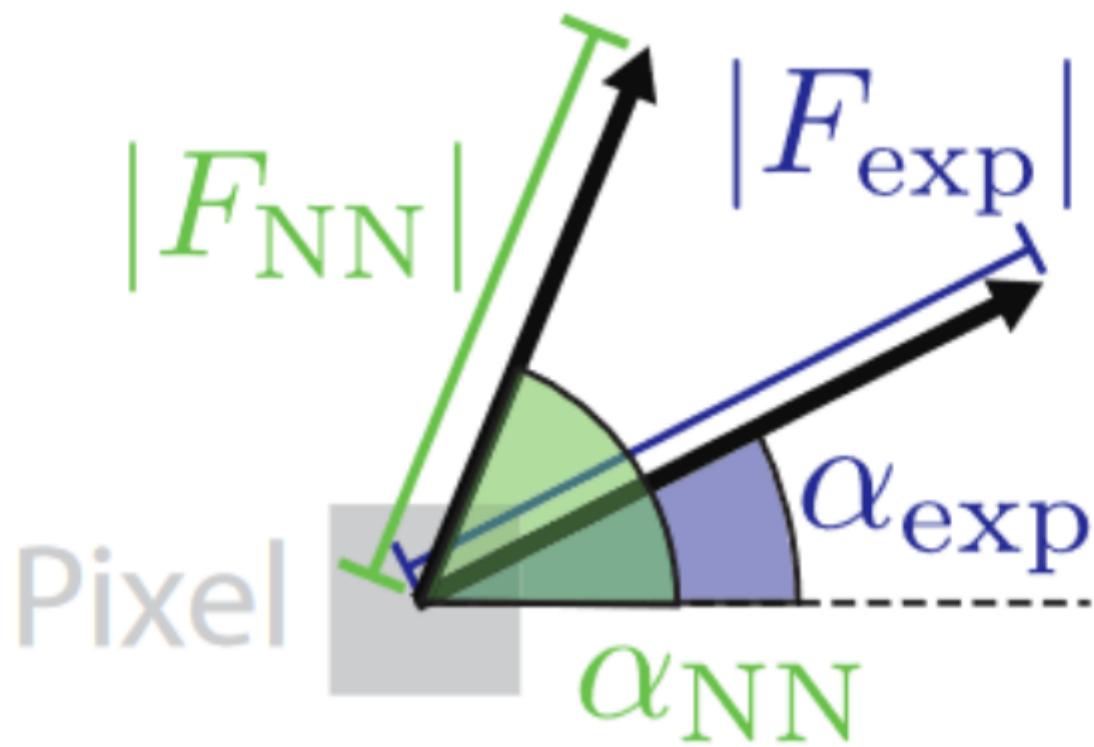
G



尽管U-Net只是通过训练集进行训练，但它学会了泛化，能够准确地预测来自测试集中的细胞的牵引力，而这些细胞它以前从未见过。神经网络对牵引力的预测在位置和大小上与实验测量基本一致，在微米尺度上可能存在一些平滑效应。

为了评估U-net准确性，作者将测试集中所有细胞的实验测得的牵引力方向和大小与U-net预测的值作比较。

如下exp表示实验测得、NN表示U-net神经网络预测的。



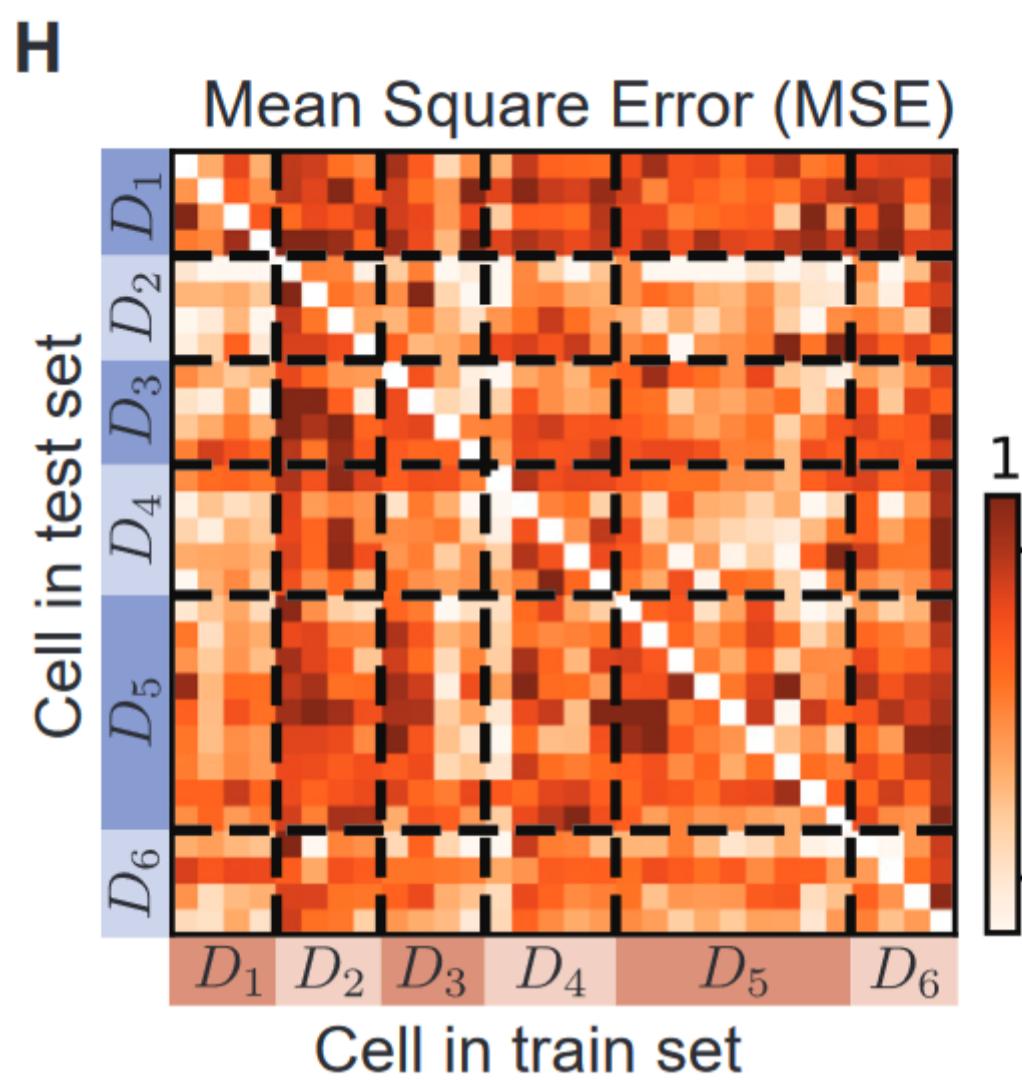
平均值（实线）和标准差（虚线）。

- Magnitude distribution 力的大小分布

神经网络在力的方向和大小上取得了接近最优的准确性，约4kPa，这相当于测试数据集中约99.9%的像素。（非常好的效果）

为了评估神经网络对所使用的测试和训练数据的敏感性，生成了22个随机分割，然后将31个细胞图库分为包含16个细胞的训练集和包含15个细胞的测试集。

在每个分割上分别训练了一个U-net，并评估了力预测的均方误差 (MSE)，如下图所示。



网络性能在一定程度上取决于测试集和训练集中的细胞组成，MSE在测试细胞之间波动不超过正负5%。

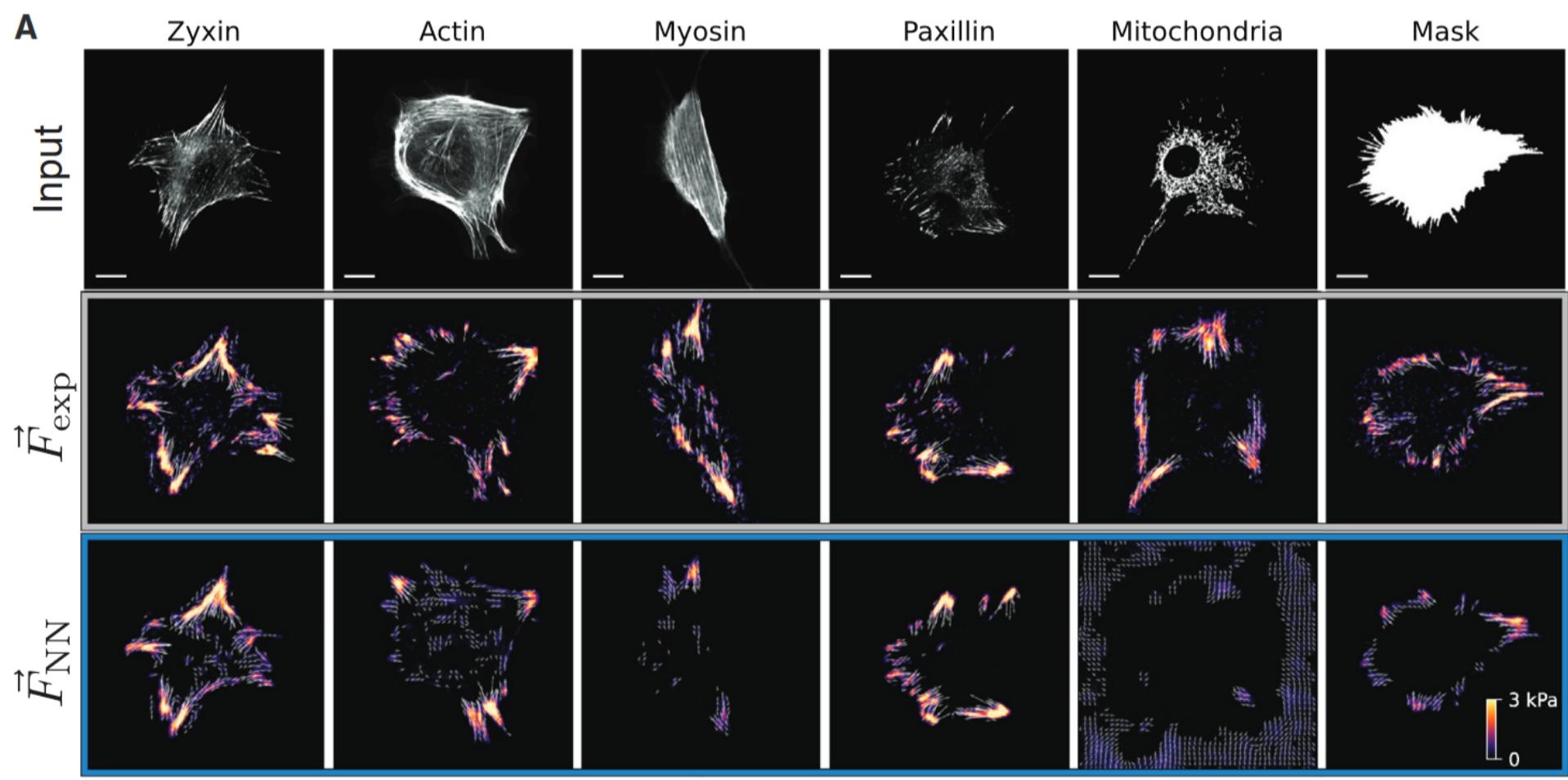
作为比较，不同日期测量的细胞之间的MSE波动为正负2%，在上图H中用D_i表示。

因此，网络对训练数据和测试数据的敏感性与系统性变化的幅度相似，这些系统性变化来自实验准备的差异，而不是来自细胞之间的基本差异。总而言之，这些结果表明，通过相对容易获得的实验数据，U-Net能够稳健地学习从荧光图像中准确预测单个FA蛋白（如zyxin）的牵引力。

用zyxin训练的神经网络效果优于其他的细胞骨架蛋白

利用zyxin训练的神经网络和zyxin和其他蛋白组合训练的神经网络效果几乎没有差别。

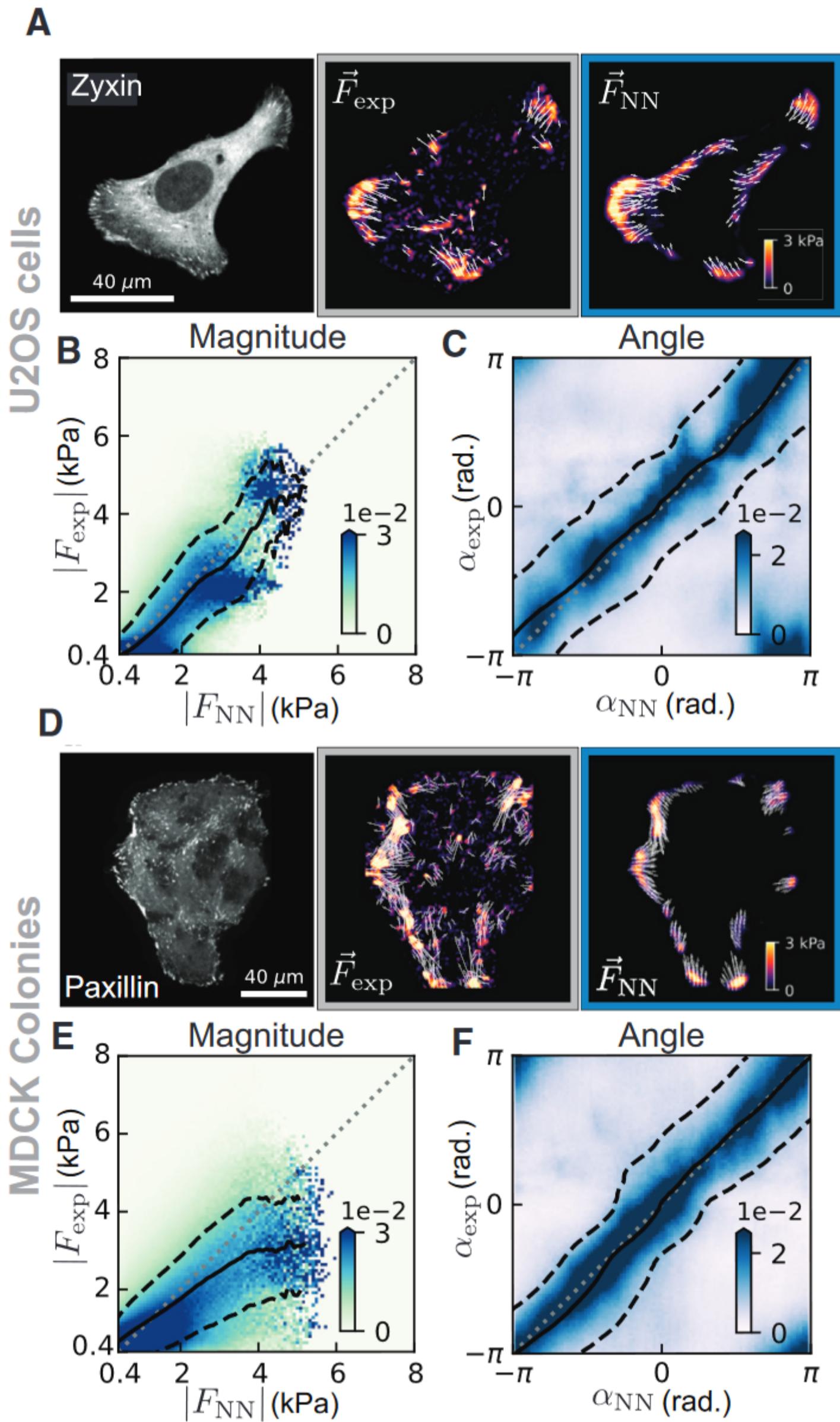
下面是不同蛋白训练的结果。



经过zyxin蛋白训练的神经网络能够推广到新的细胞类型以及对抗一定的生物干扰

作者评估的U-net是在成纤维细胞的图像上训练的，使用这个U-net去在其他粘附类型细胞图像上预测

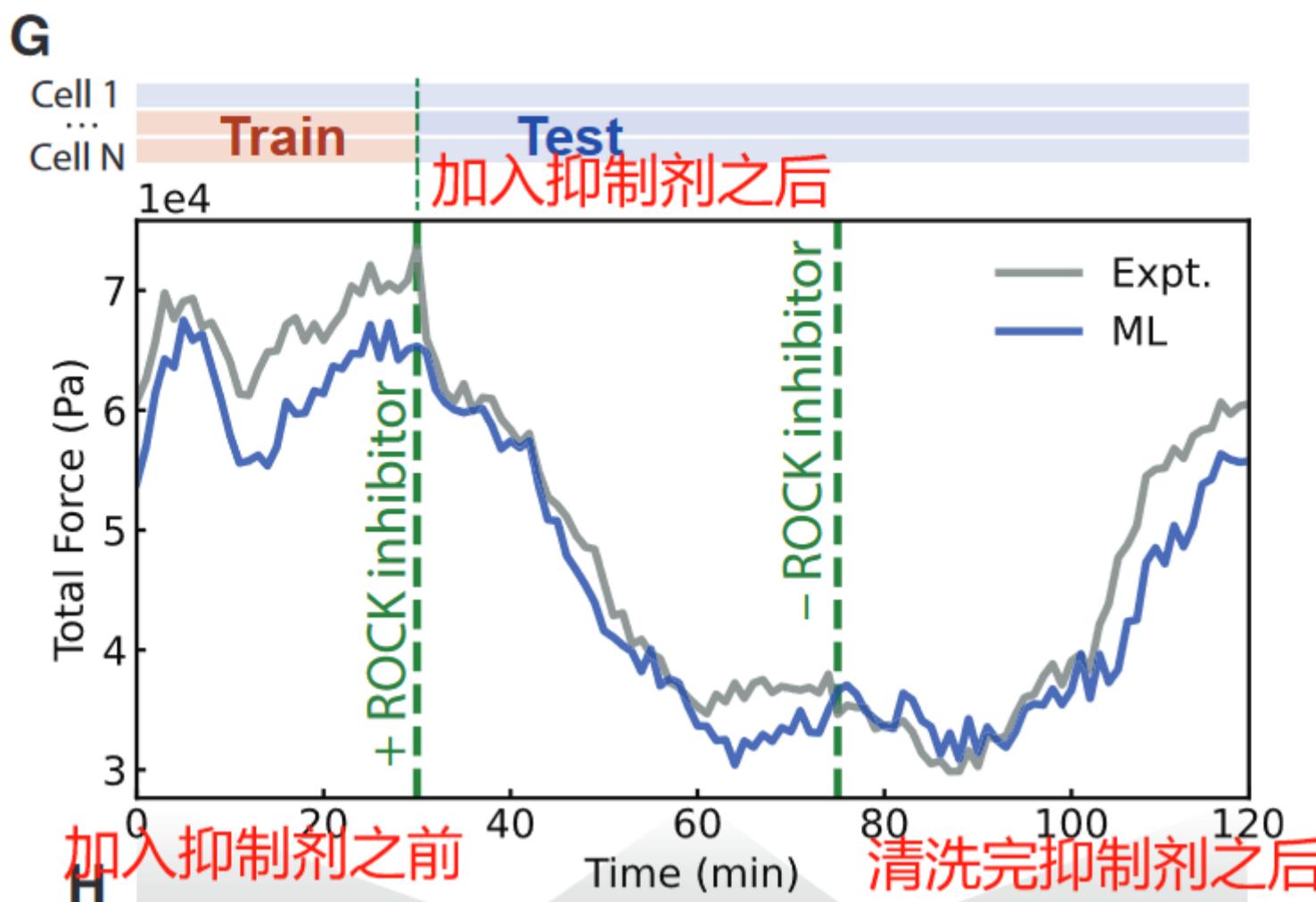
具体如下，



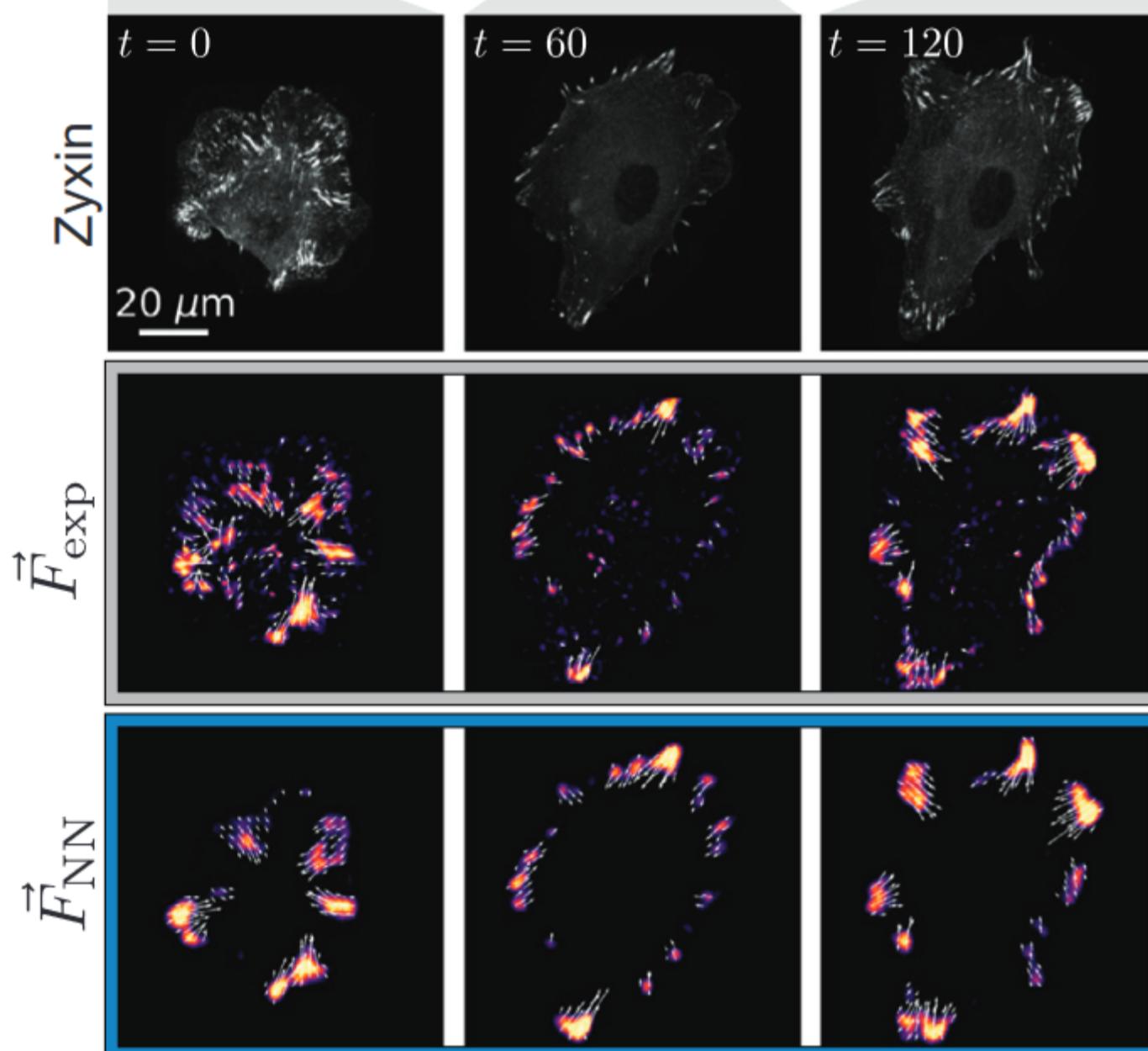
- 犬上皮细胞集落 (MDCK colonies)
- 人类骨肉瘤上皮细胞 (U2OS cells)

对抗一定生物化学干扰的实验

为了进一步探讨这个想法，接下来我们挑战我们的U-Net模型以对生物化学干扰做出预测。我们在基础的收缩状态下成像细胞30分钟，然后添加5um的Rho-kinase (ROCK)抑制剂Y-27632处理45分钟，然后冲洗药物并成像最后45分钟。



H



正如预期那样，添加了抑制剂之后导致了牵引力下降，细胞总面积增加，FA大小减小，而冲洗又扭转了这些趋势。

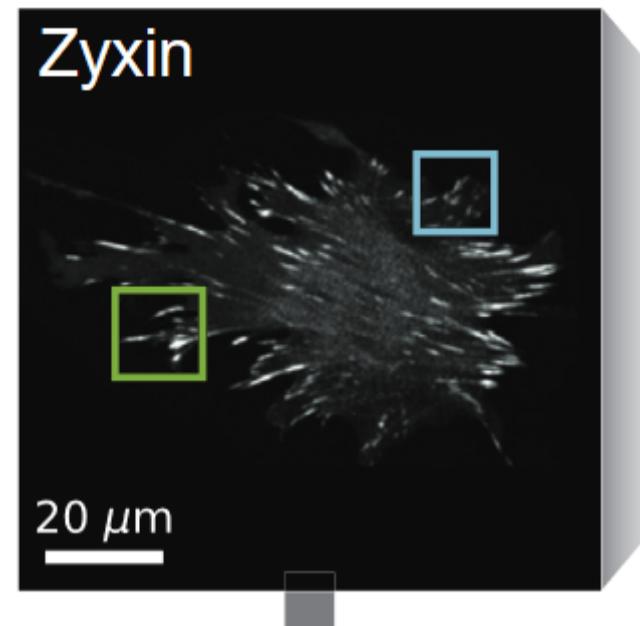
可以发现，尽管神经网络从未见过这些药物的干扰，但是仍然有效的预测了药物影响和随后冲洗后恢复期间全部的牵引力的总体变化和FA的局部变化。

这些结构表明了zyxin单独的分布是细胞机械状态的忠实代表，足以预测各种条件下的牵引力。

神经网络识别细胞粘附和形态特征

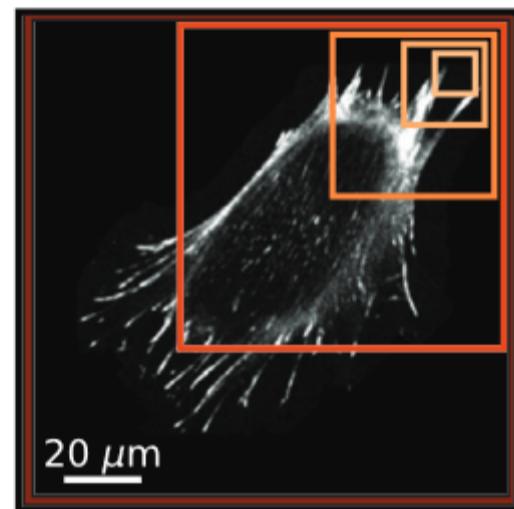
受前面牵引力预测的启发，作者接下来追求识别zyxin蛋白的分布特征。

zyxin蛋白图像的编码是微米级，例如FA也是编码细胞级结构，例如10-100毫米的应力纤维。如下。

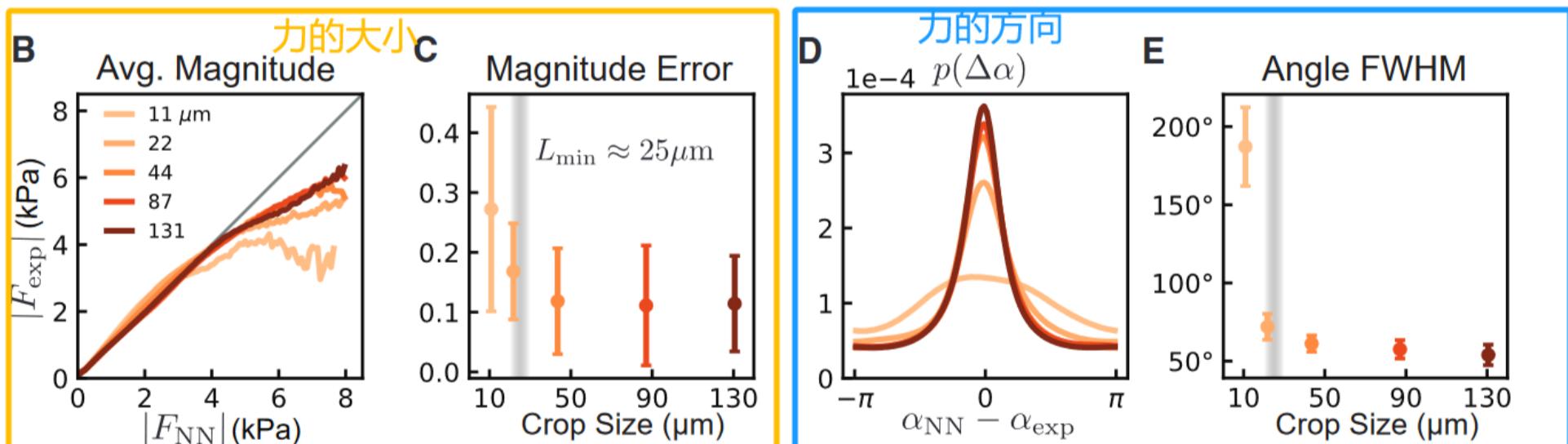


为了探究神经网络如何解释这些特征，对输入数据中从大小10毫米到130毫米的随机图像进行剪裁，然后进行U-net训练。剪裁情况如下。

A



即使只在一小部分细胞上进行训练，这些网络也能学习到力大小和方向平均准确的模型。



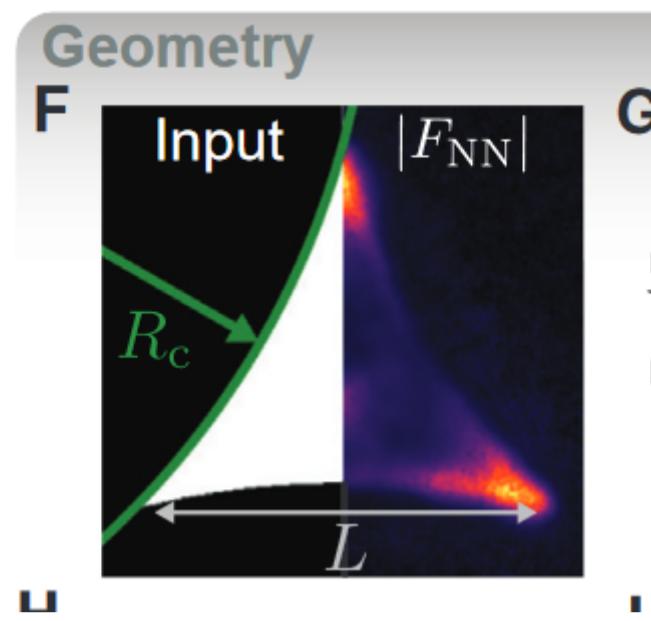
在这两种测量中，随着输入尺寸增加到25mm以上，预测精度的提高可以忽略不记。

这表明u-net不需要知道整个单元的几何形状，并且它可以通过考虑任何给定点周围的较小的邻域来进行准确的预测。

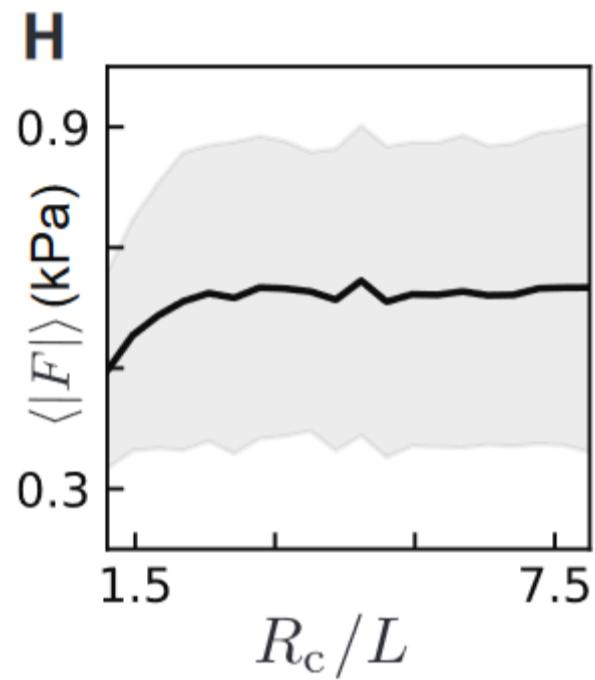
- 思考1：transformer的注意力机制能否在局部的使用有不错的效果
- 思考2：Activating More Pixels in Image Super-Resolution Transformer 混合注意力机制，对于局部超分辨的处理

先前的研究表明，细胞形态和分布都会影响力的作用。为了了解u-net是如何解释这些特征的，作者合成了细胞来系统地改变这些特征，并检查训练模型的反应。

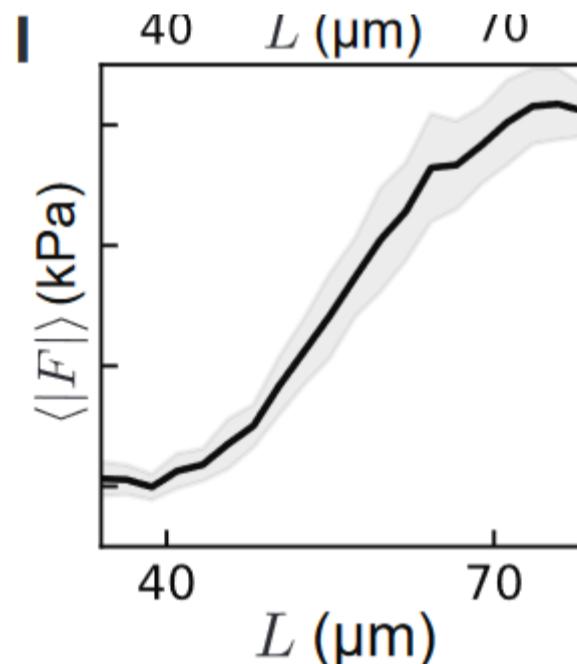
为了探究细胞形态的作用，在宽度为L的三角形细胞上评估了mask-train的u-net，这些细胞的边缘是曲率半径为Rc的圆弧，如图所示。



虽然神经网络没有系统地对细胞边缘曲率的增加做出反应，如图所示，可以看出比例没有发生变化。

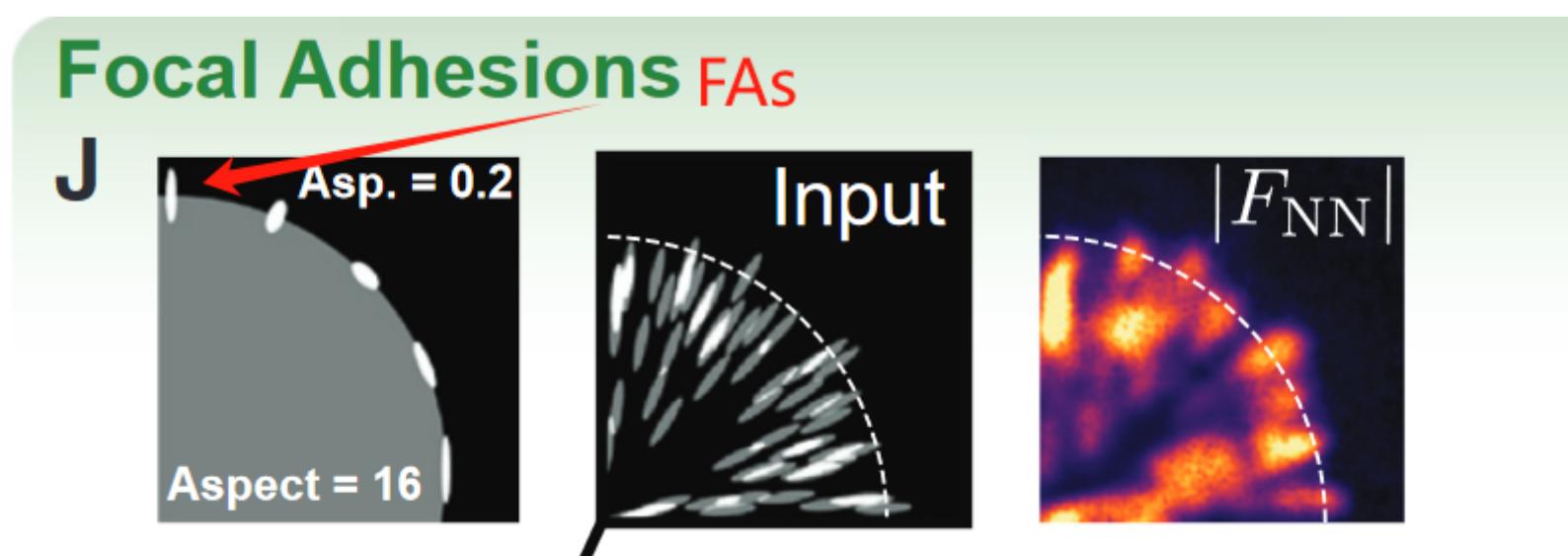


但是发现了力的产生随着细胞总尺寸的增加而增加，如下所示。



这一结果与以前的结果是一致的，以前的实验结果表明，力的产生与单元面积成比例，并且也验证了神经网络对单元几何形状的大规模特征是敏感的。

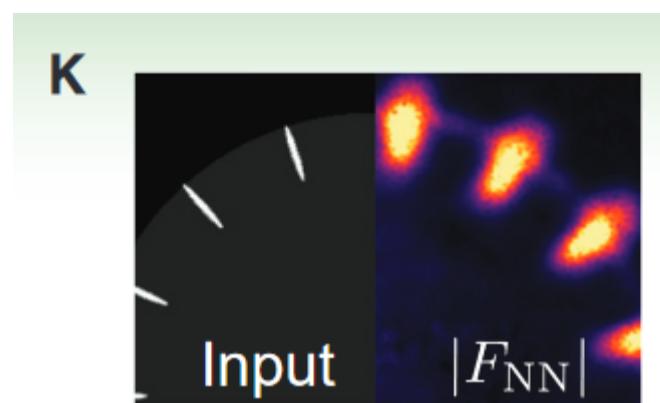
为了探究FA类型特征的作用，作者创建了不同面积和纵横比的椭圆形FAs组成的合成细胞，然后这些细胞随机的分布在整个圆形细胞当中。



宽高比的定义是使网络能够探测对于类似FA结构力的方向和大小。

我们发现，zyxin图像训练的u-net预测了面积约为 $2\mu\text{m}^2$ 横纵比为0: 1的椭圆（即径向指向的椭圆）的最大力，这个与FA的实验结果是一致的。

作者通过创建具有固定强度和长度的均匀分布椭圆的圆形细胞，进一步研究了FA强度的作用，如下图所示。



在增加椭圆强度时，发现了一种非线性响应，其中预测的牵引力大小最初是急剧上升，并在更高的强度下继续以较慢的速率增长，与先前关于zyxin强度随着施加的力而增加的报道是一致的。

这一回顾性分析揭示了神经网络如何将zyxin信号的许多不同的特定特征转化为对细胞力的预测。

u-net没有记忆训练数据中复杂的、无法解释的相关性，而是识别了生物学特征，使其能够准确预测力的产生通过细胞类型和生物力学状态。

物理瓶颈神经网络（物理理论限制下）：学习粘连增强有效的弹性模型

虽然从zyxin预测力的u-net学习规则远远超出了它所训练的领域，但网络如何使用输入数据的特征进行预测并不透明。

相比之下，受经典连续体理论启发的先前模型依赖于允许最大可解释性的简单假设。（？）

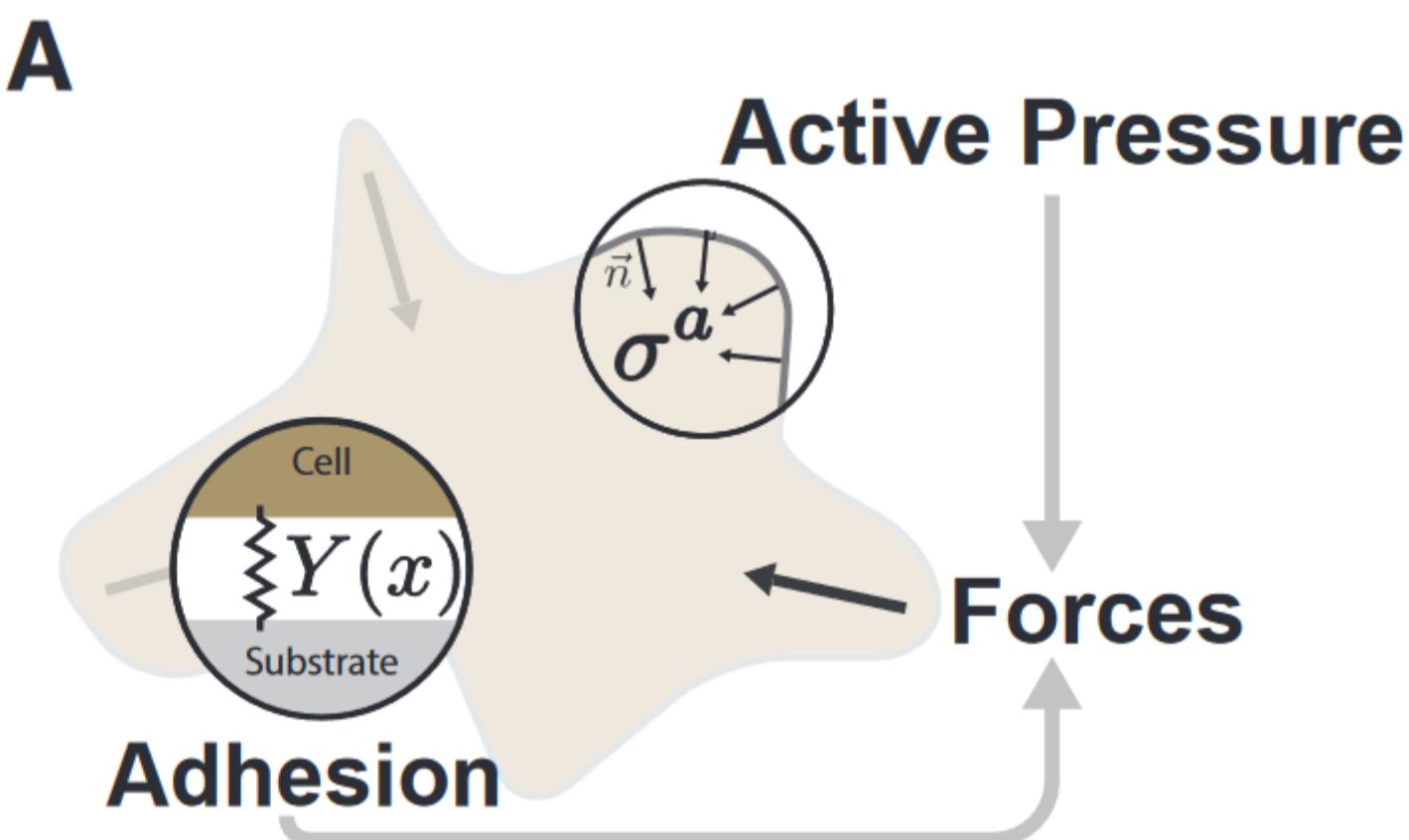
往往上面所述的经典模型缺乏在广泛的细胞形状和局部FA分布下进行预测的能力。而神经网络将zyxin纳入连续力学模型，从而学习蛋白质和物理参数之间的关系，从而提高物理模型的可推广性。

模型推导

作者现在求助于一个现有的模型，该模型将细胞视为粘附在基底上的有效二维（2D）活性弹性凝胶。这个最小模型的吸引力主要在于它的简单性：它仅用两个参数来表示控制细胞粘附和收缩的复杂过程。

均匀粘附强度 Y

整体活性应力 σ^a



力的计算为 $\vec{F}(x) = Y\vec{u}(x)$

其中 $\vec{u}(x)$ 是通过最小化系统的自由能而得到的位移场 (Supplemental information can be found online at <https://doi.org/10.1016/j.cell>. 2023.11.041.)

然后考虑空间变化的粘附场 $Y(x)$ 来扩展该模型，以解释细胞中FA位点的不均匀分布。

受u-net成功的启发，将这两个物理参数，即整体活性应力 σ^a 和力 $\vec{F}(x) = Y\vec{u}(x)$ ，与化学量联系起来，使它们与zyxin相关

$Y[\zeta](x)$ 和 $\sigma^a[\zeta]$ ，其中 $\zeta(x)$ 表示实验确定的zyxin的分布，因此最后力给出的结果为

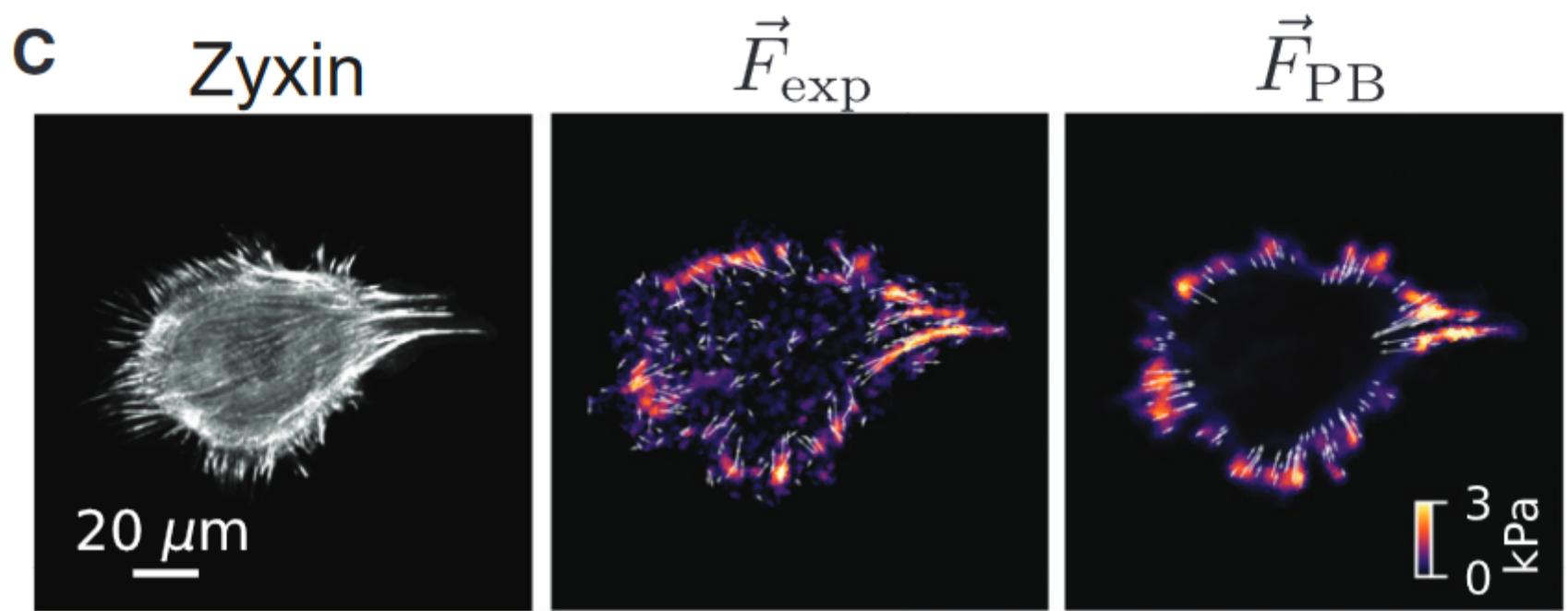
$$\vec{F}(x) = Y[\zeta](x)\vec{u}(x)$$

虽然存在从实验力数据估计模型参数的经典方法，但是它们没有考虑到参数是zyxin函数的额外约束。为克服这一限制，我们引入了一种物理瓶颈神经网络（PBNN）架构。

先前的u-net是通过处理网络潜在层中计算的数百个特征来计算力。

相比之下，物理瓶颈神经网络只计算两个特征，根据这两个特征可以以确定性和众所周知的方式计算力。

具体而言，PBNN使用神经网络计算 $Y[\zeta](x)$ 和 $\sigma^a[\zeta]$ ，并将它们作为参数输入到PDE求解器中，以计算牵引力。如下图所示。

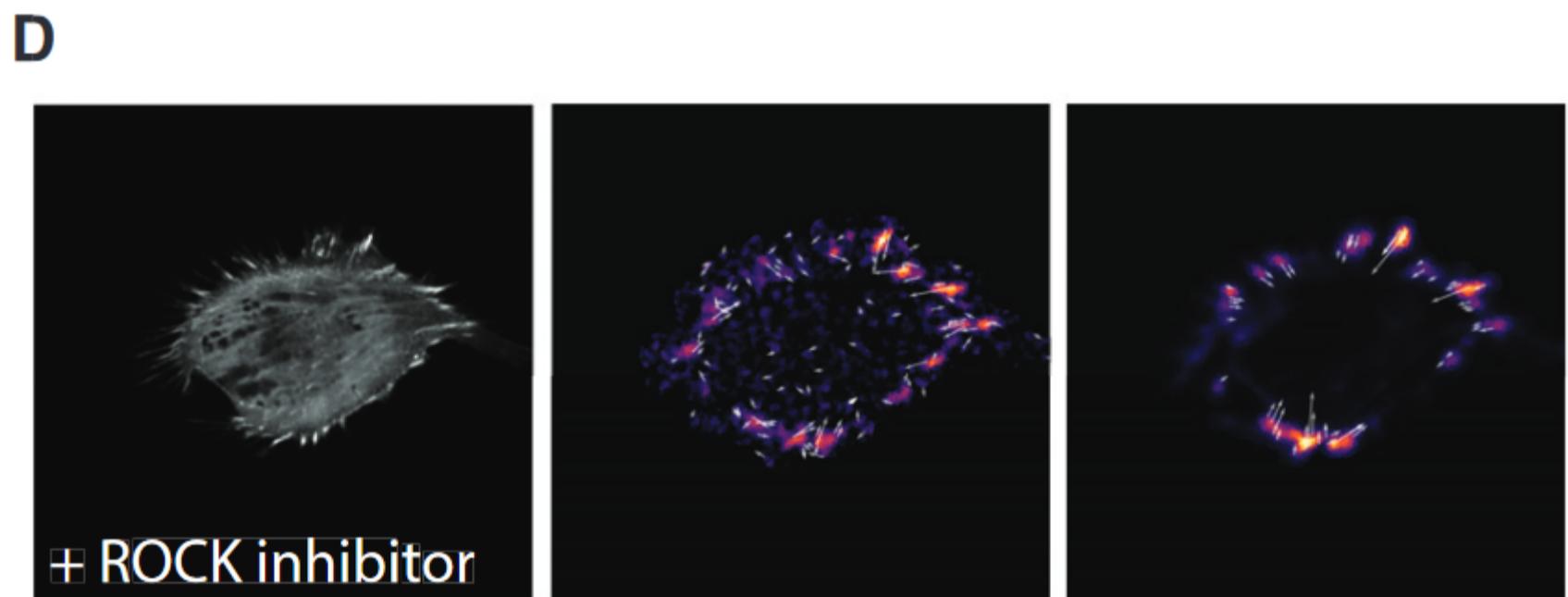
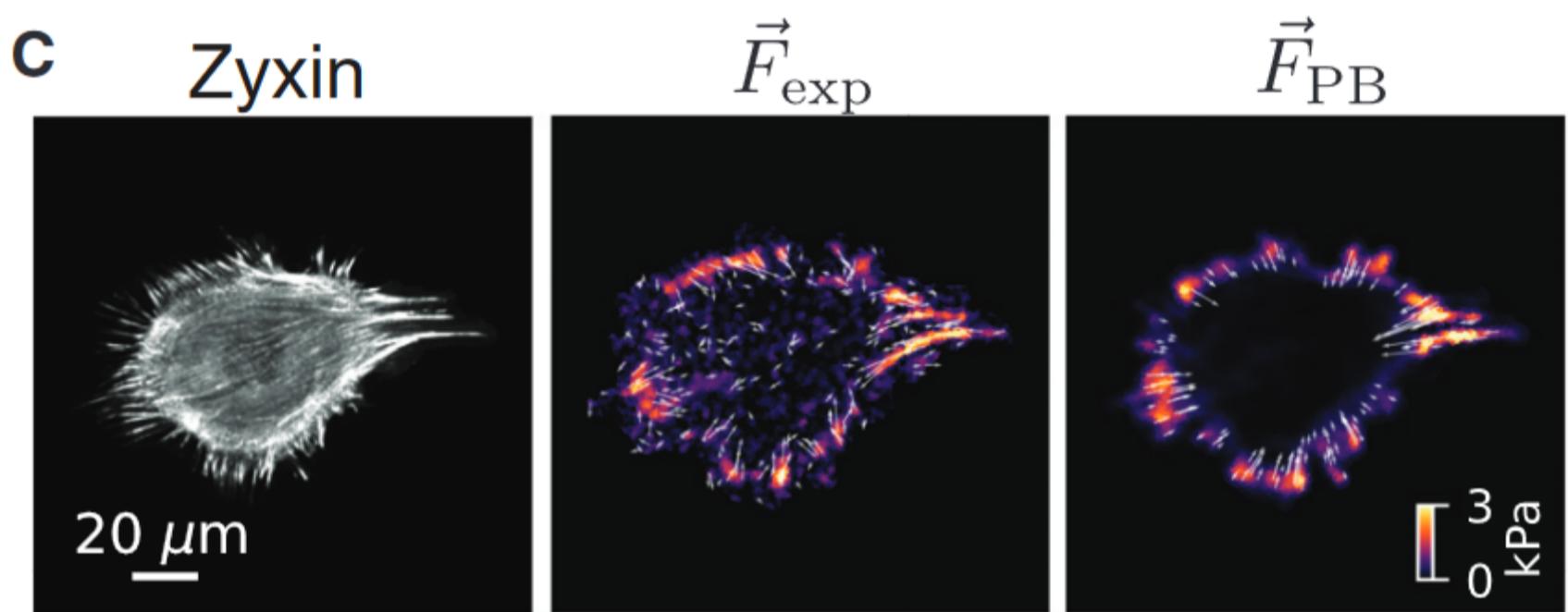


训练PBNN来预测是预测力和实验测量力之间的MSE最小化的参数。

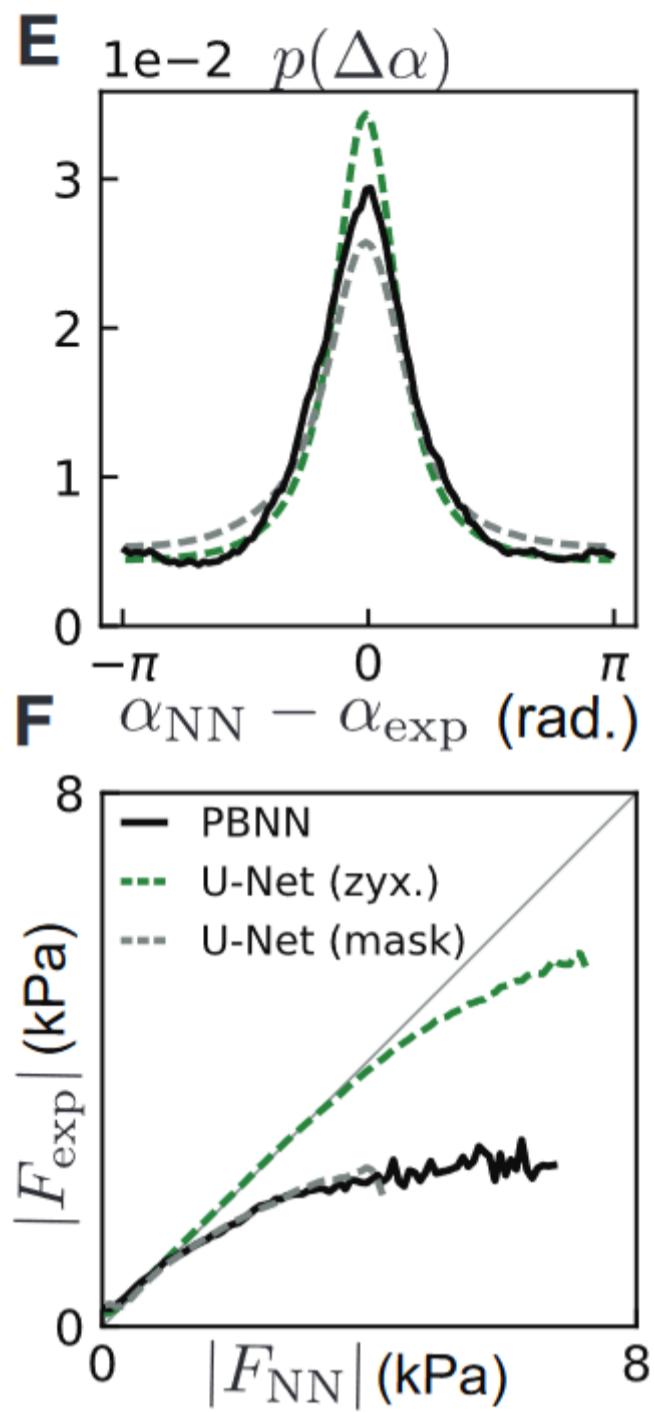
在每个迭代训练步骤中，1（偏微分方程的最优控制：理论、方法和应用）中的方法用于计算物理模型参数的更新，2然后这些参数通过反向传播传递到神经网络。

1、2两步确保了神经网络更新遵守物理模型的严格约束。

PBNN准确预测力，并推广到受Rho-kinase (ROCK)抑制剂Y-27632干扰的细胞。如下。



然而，在整个数据集中，预测的力角和大小的平均精度不如先前无约束条件下zyxin训练的u-net的预测精度。



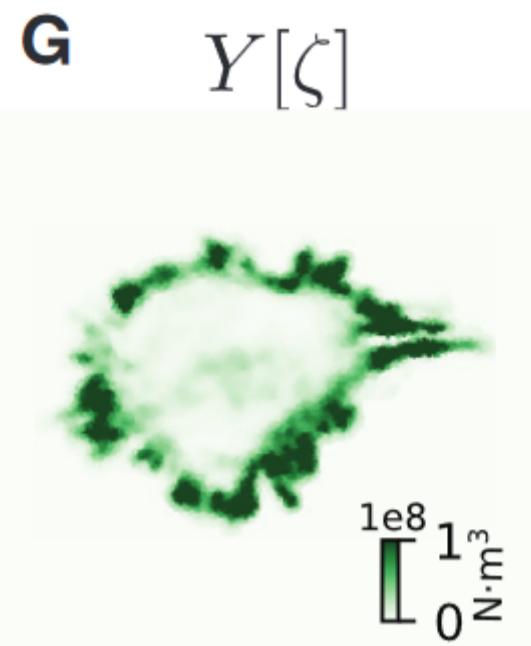
这种情况是由于对PBNN施加的附加约束而预期。

尽管如此，PBNN的预测与无约束u-net不相上下，如上图所示。这表明在物理瓶颈神经网络学习的两个参数包含预测力的相关信息至少与无约束的u-net从细胞形态推断的信息一样多。

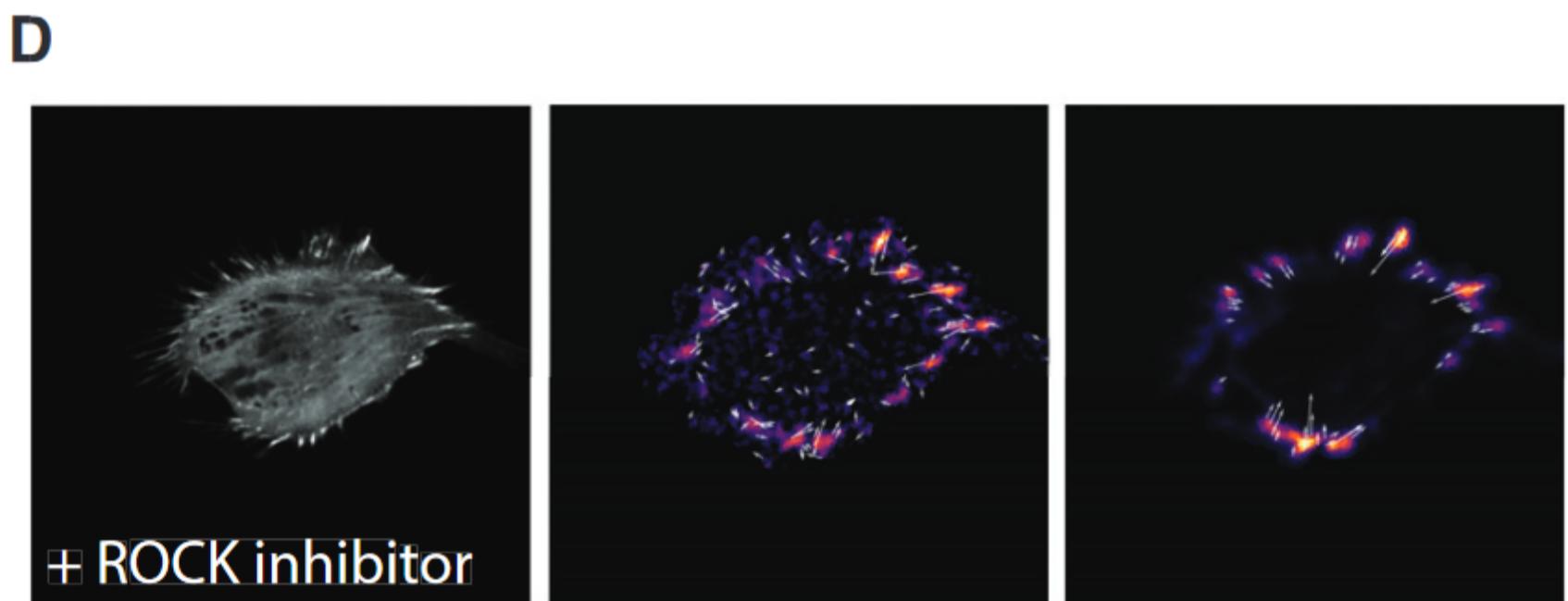
此外，u-net用一个几乎任意复杂的函数处理其潜在特征，而PBNN通过一个简单的微分方程将Y场和 σ^a 处理成力。

依赖zyxin的粘附场 $Y[\zeta](x)$ 的引入就足以使物理约束的网络模型与完全深度学习的u-net竞争。 $Y[\zeta](x)$ 用于表示Y是zyxin场 ζ 的函数，其在空间上变化着。

为了简洁起见，我们偶尔会省略对 ζ 的依赖。我们发现，学习场具有强烈的异质性，并定位于FA位点。



此外， γ 的强度随着抑制剂的作用而降低，反映了FA数量的重组和减少。（如下图明显减少）



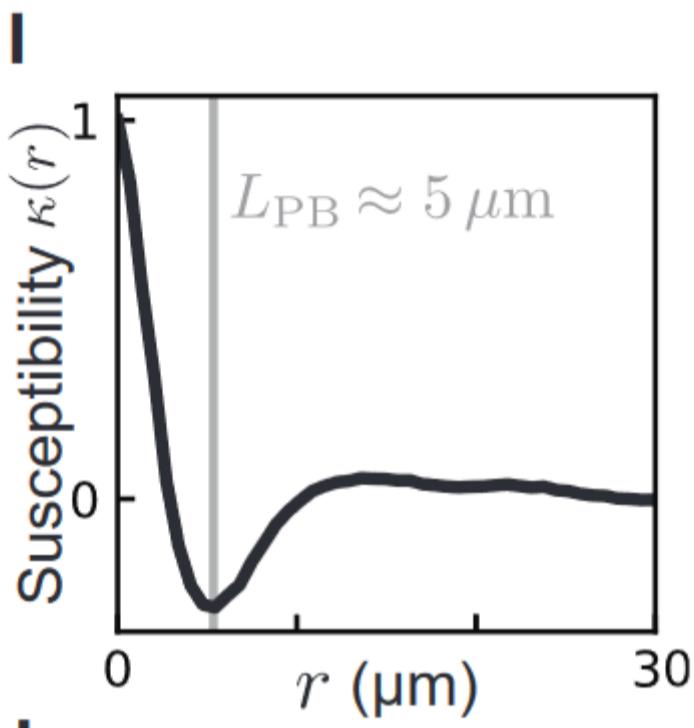
H



然而目前还不清楚PBNN是如何从zyxin $\zeta(x)$ 的空间分布计算 $Y(x)$ 。

为了表征点 X_i 处的粘附如何依赖于点 X_j 处的zyxin, 将网络的易感性或线性响应定义为 $k_{x_i, x_j} = \frac{\partial Y(X_i)}{\partial x_j}$

磁化率曲线呈快速衰减, 最小值约为5um, 如下图所示。

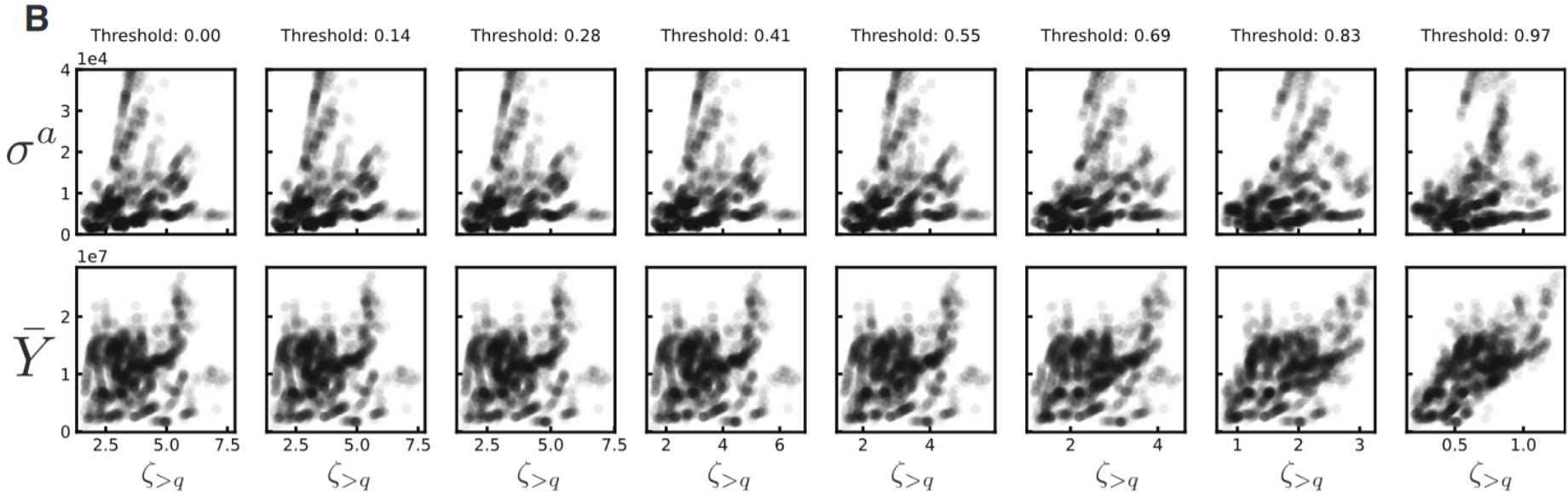


它的形状类似于峰值发现算法 (peak-finding algorithm) 中常用的拉普拉斯滤波器, 表明 $Y(x)$ 与 zyxin 信号中的最大值相关。

通过将每个图像 Y 中得平均粘附力与高于给定阈值得 zyxin 值之和相关联, 进一步探讨了 Y 对 zyxin 的依赖性。

H





上方 σ^a 是整体活性应力

随着阈值的增加，Y与zyxin的相关性显著增强，这表明粘附场的大小主要由最高的zyxin值设定。

综合这些结果表明，粘附场正在编码zyxin强度的高值峰，这对应于粘附斑（FAs）。

PBNN学到的参数受到用于约束它们的模型的假设的影响。弹性模型对细胞内的位移进行预测，这在TFM实验中无法直接获取，而且并不清楚这些位移应该从哪个未变形的参考框架测量。

这是因为与被动的质点和弹簧的晶格不同，细胞即使没有外部变形也会持续发生细胞骨架重塑。尽管如此，PBNN仍然是测试我们假设模型的强大工具，并告诉我们预测牵引力所需的最小必要要素。

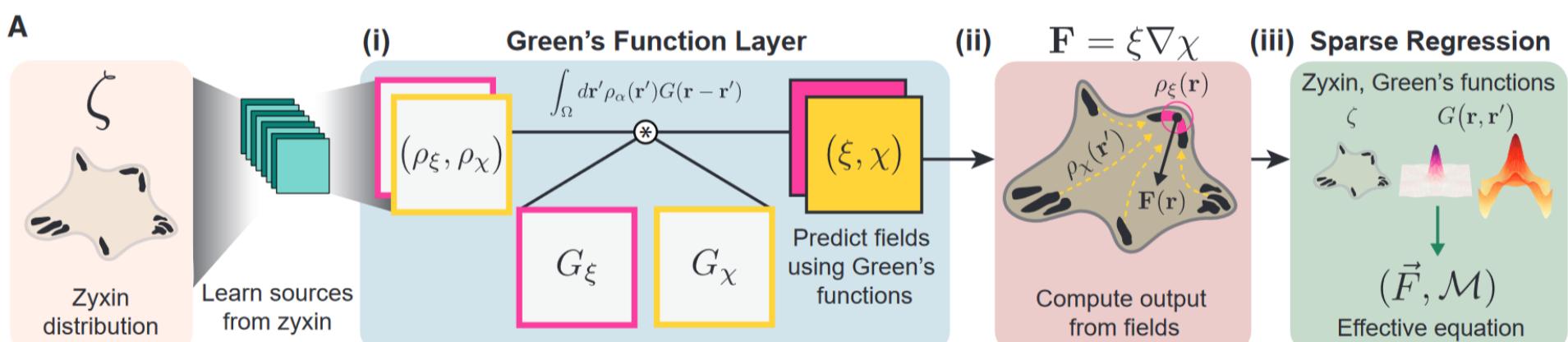
作者展示了细胞形状（编码为边界条件）、全局收缩的“设定点” s_{α} 以及编码粘附斑的场 $Y(x)$ 足以进行预测。此外发现描述中间位移场的线性偏微分方程是足够描述观察到的行为的数学模型。

Green's function neural networks: 不考虑物理学模型构建揭示长度尺度和有效方程

PBNN的成功依赖于生成合理的假设模型，因此它产生的见解是受到具体模型的影响。

于是这一部分就是思考能否放款这些约束，以便在没有强有力的机械假设的情况下获得见解。

为此转向受物理启发的方法，以识别对具体底层物理模型不加选择的机器学习规则。这种方法是通过减少先前那种u-net的复杂度来换取更少、更易解释的操作。



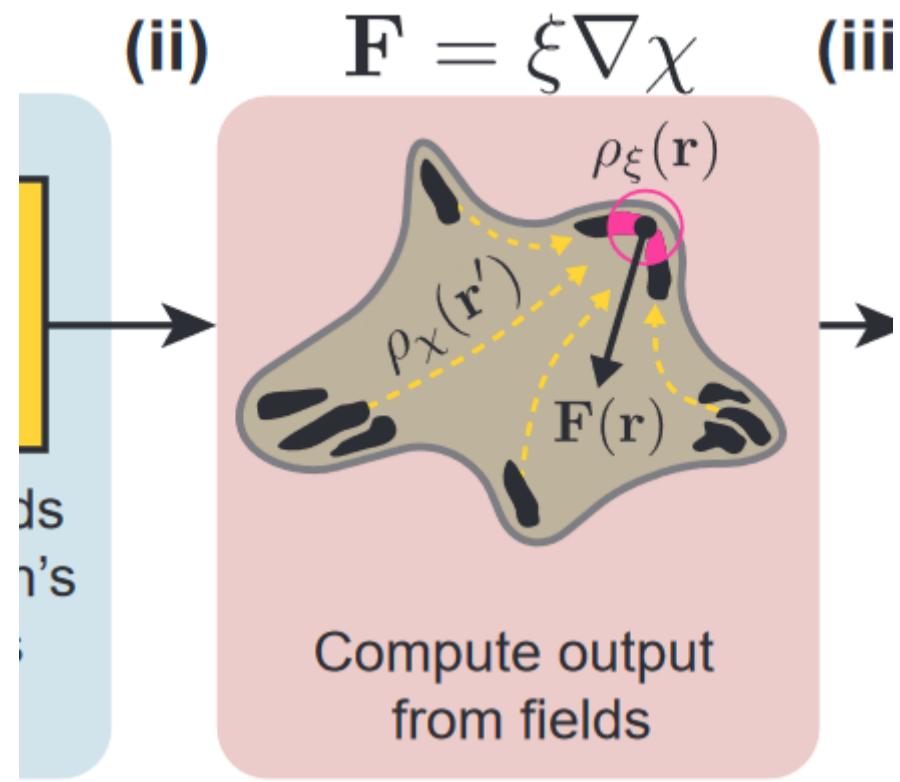
具体如下，

假设力可以写成从zyxin衍生的机器学习场的函数，即上方的黄色和粉色框。虽然这些场类似于PBNN的位移 $u(x)$ 和粘附场 $Y(x)$ ，但不要求这些量遵循线性弹性或任何其他特定的连续理论。

只要求它们与zyxin密度的非局部的机器学习关系由Green函数表示。Green函数方法是计算系统对局部扰动的响应的通用工具。

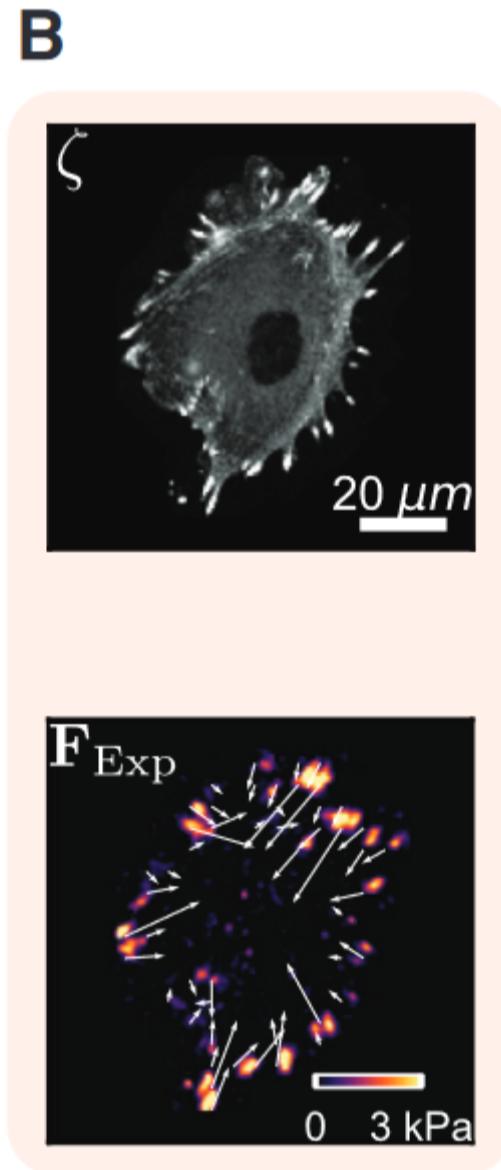
Green函数方法是计算系统对局部扰动的响应的通用工具。

例如，经典静电力学的Green函数是 $\frac{1}{r}$ 电势，它确定了距离 r 处的电荷的影响。借助我们机器学习得到的Green函数，我们类似地寻找确定局部牵引力如何依赖于整个细胞内地zyxin密度。

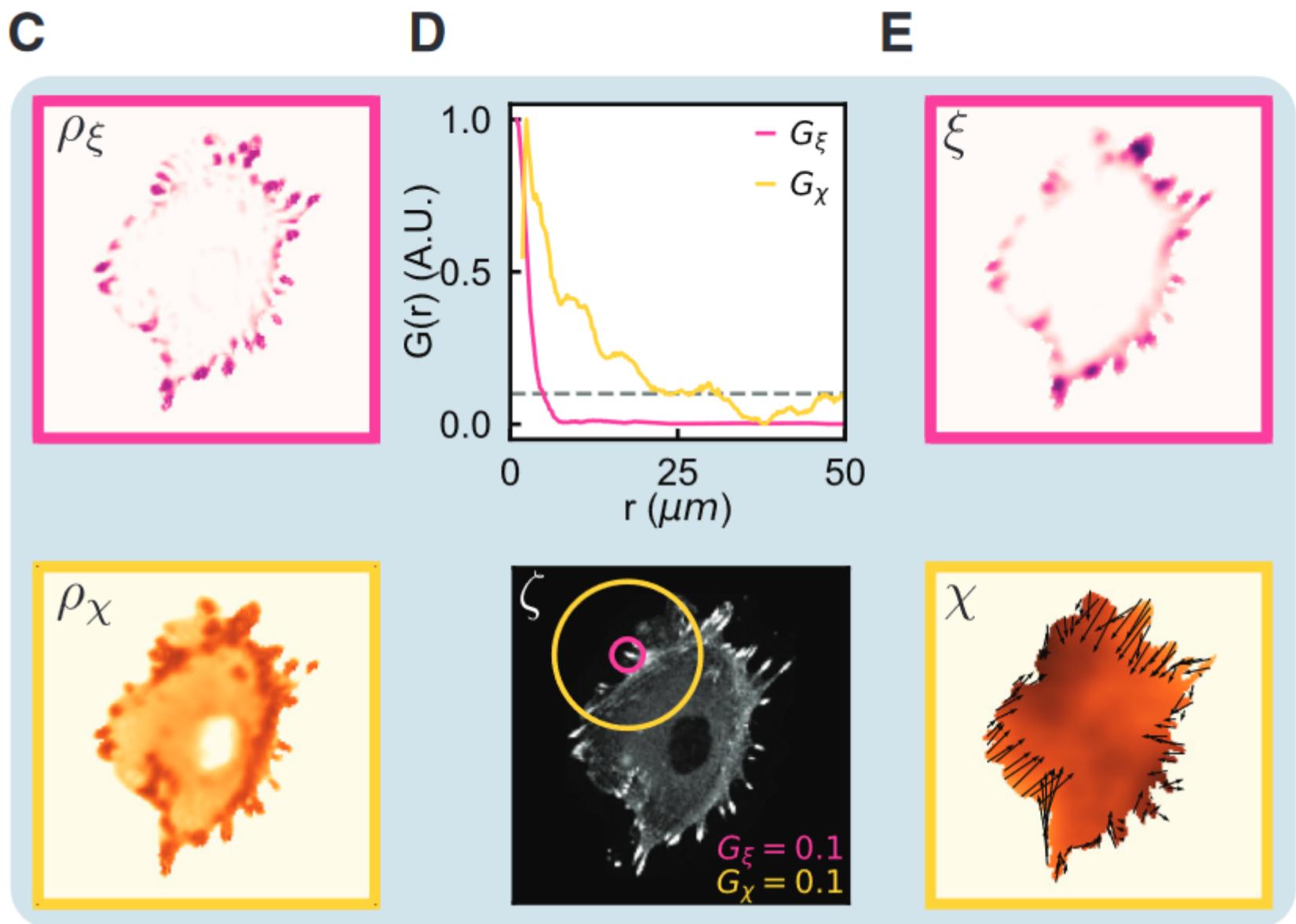


与先前的物理瓶颈神经网络相比，这是一个没有现成公式可用的问题。

使用相同的zyxin图像作为输入，然后训练一个Green函数神经网络（GFNN）来表征输入zyxin图像与它们各自的牵引力之间的空间相互作用。



GFNN从zyxin图像中学习一系列源和场，如下图所示。用于预测牵引力。



虽然原理上GFNN可以学习任意数量的场，但是发现最小复杂模型只需要使用两个就能实现准确预测。

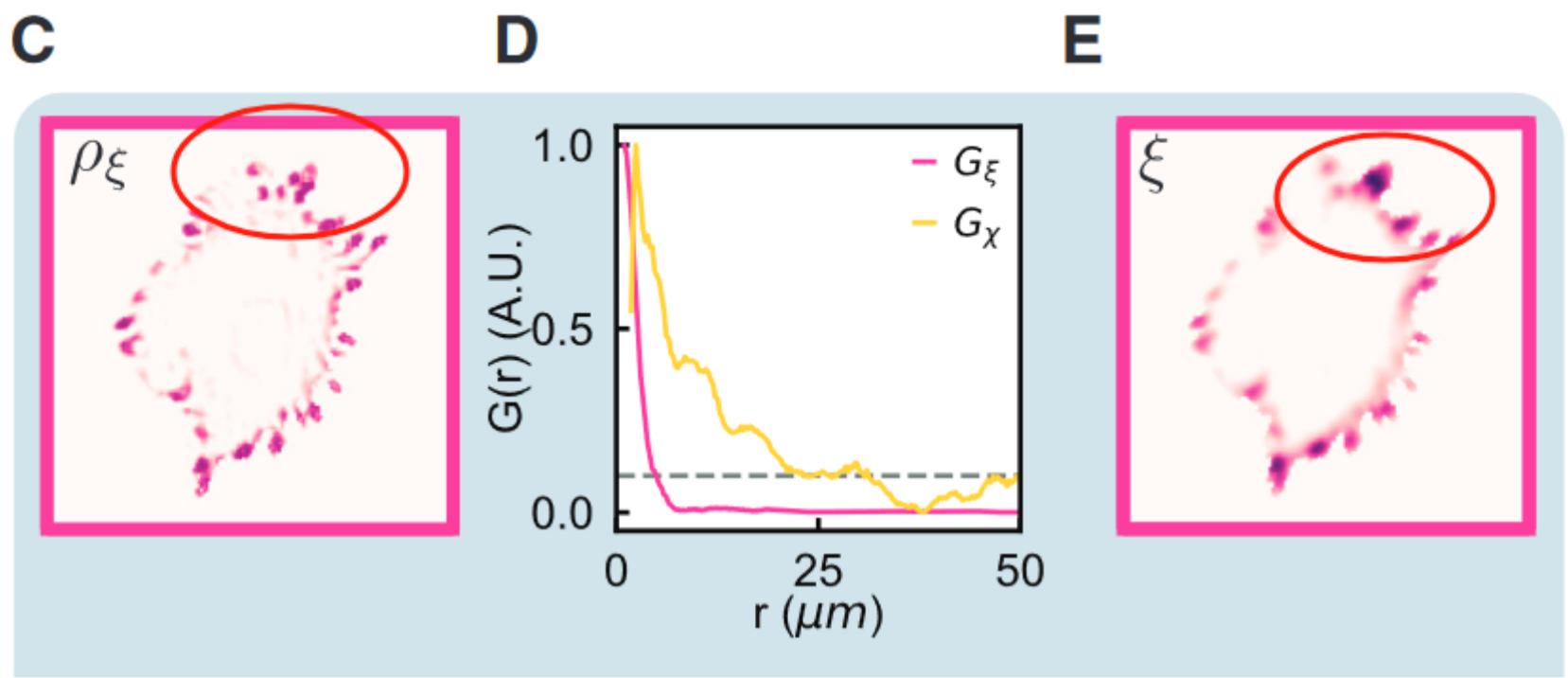
具体而言，GFNN学习两个场，一个是 ξ 一个是 χ ，通过它们可以对牵引力进行预测 $\vec{F} = \xi(x) \vec{\nabla} \chi(x)$

这种表示方式类似于库仑电静力学，其中 ξ 和 χ 分别类似于电荷和电势。

回到机械模型，可以发现机器学习公式在形式上类似于物理信息模型

$$\vec{F} = Y\vec{u}$$

电荷 ξ 识别了类似于FA的zyxin强度的局部峰值。



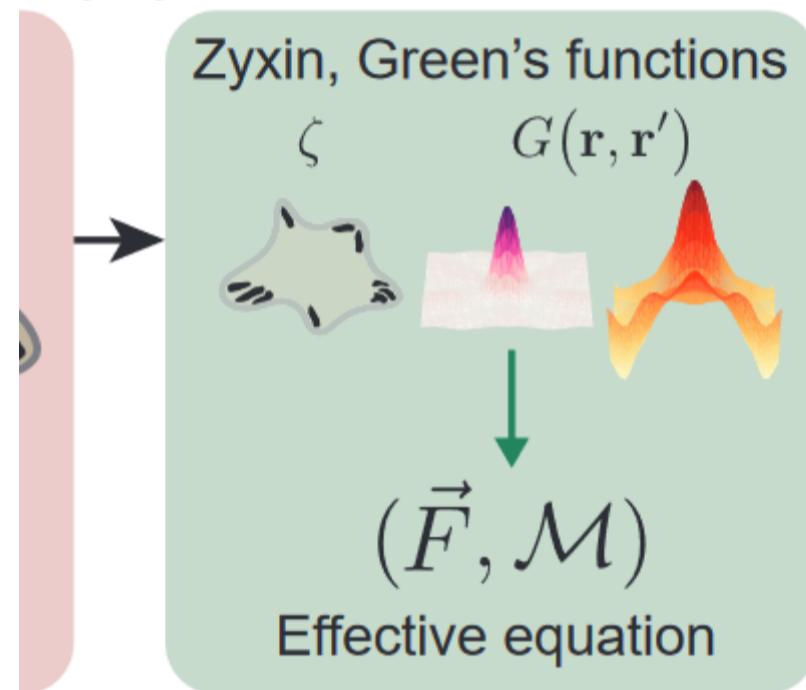
从上图D可以看出， ξ 的Green函数 G_ξ 在非常短的长度尺寸上衰减，表明它是由粘附点处的局部信息确定的。

电势 χ 的低位较差，其Green函数 G_χ 从细胞的更大区域累计zyxin信息。

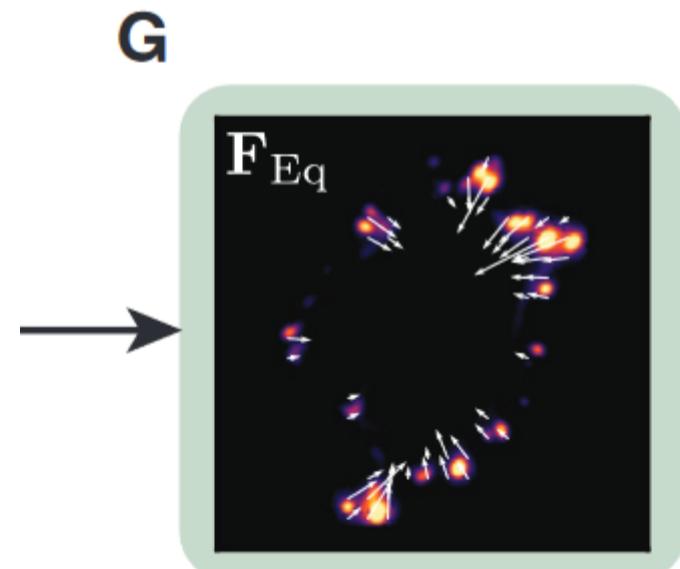
这个较长的衰减长度表明“电势”可以从zyxin分布中推断出细胞形态的一些方面。因此我们的GFNN模型通过FA“电荷”和细胞“电势”之间的相互作用来预测牵引力。

为了进一步简化这个模型，使用稀疏回归来构建近似牵引力的有效方程。

(iii) Sparse Regression

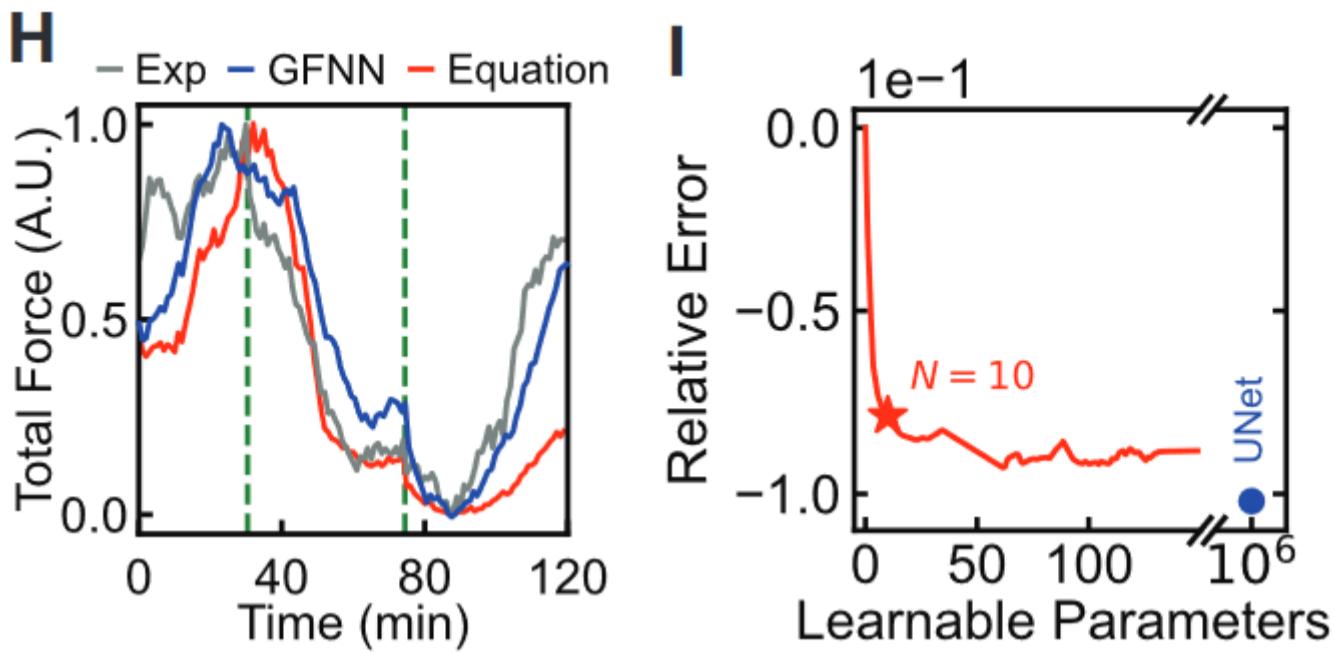


通过GFNN启发的少数项，可以获得一个在定性上准确的分析公式



与包含 10^5 倍参数的u-net相比，这个公式经过了极大的压缩。

同时，它可以捕捉到77%的u-net预测，并且还可以推广到使用抑制剂实验当中，如下图。



这说明了U-Net，一个复杂的黑盒，如何被提炼成一个类似精确的公式，由两个非局部相互作用组成，并且只有少数几个术语参数化。如上有图，在一定参数之后相关性错误基本不在发生改变。

以上的流程展示了如何提取有效的方程，将蛋白质分布映射到牵引力，而无需知道显式的底层关系。

尽管没有使用物理输入来推导它们，但结构上两个方程惊人的相似。

$$\vec{F} = \xi \vec{\nabla} \chi$$

PBNN模型推导出的公式 $\vec{F} = Y \vec{u}$

其中 ξ 和 Y 都是 FA 中累计 zyxin 信息的场，而 $\vec{\nabla} \chi$ 和 \vec{u} 都是传播细胞内信息的矢量场。

DISCUSSION

证明了深度神经网络可以直接从蛋白质分布的图像中预测细胞的收缩力学特性。结果表明，单个FA蛋白质（如zyxin）的图像包含足够的信息来准确预测牵引力。从一个显微镜收集的一种细胞类型的图像上训练的网络可以在各种细胞类型、实验设置和生物力学环境中实现泛化。这些结果说明了机器学习从异质生物数据中提取强健预测的实用性。这种方法可以使用一个容易获取的实验图像数据量来实施。

这使它们特别适用于在蛋白质容易成像但物理测量困难的情况下预测力学行为。

这里引入了三种数据驱动的生物物理建模方法，它们在建模过程的不同阶段利用机器学习来揭示新的见解

Data-driven biophysical modeling

Identify relevant proteins

Learn governing rules

Enhance mechanical models with biochemistry

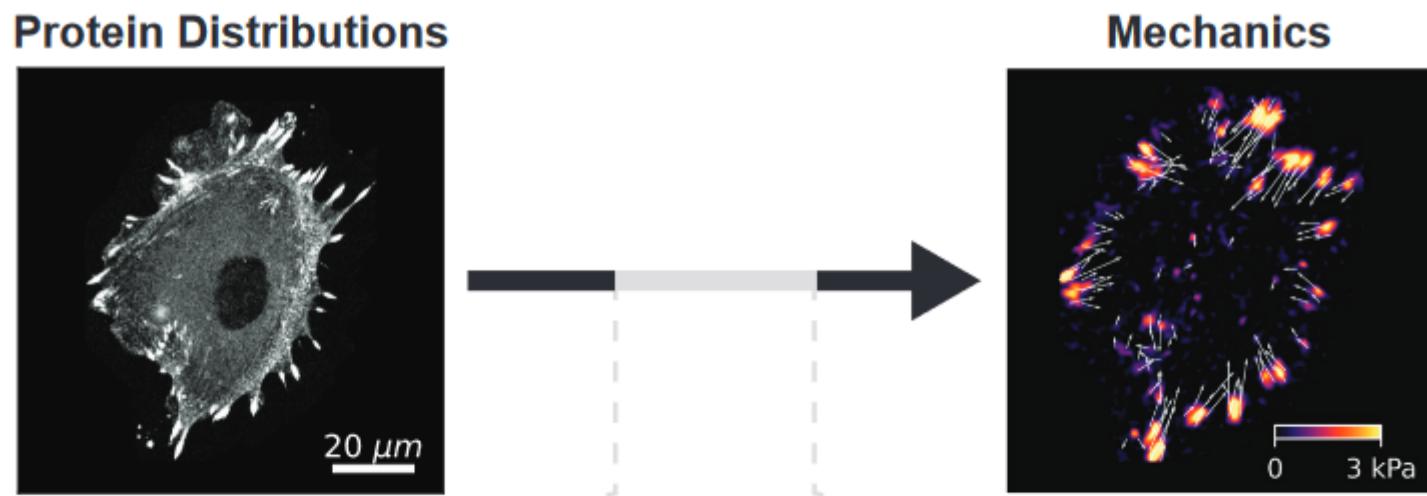
虽然深度学习神经网络本身并不可直接解释，合成数据在识别具有许多相互作用的蛋白质的过程中相关组件的实用性。

接下来引入一种新颖的PBNN来测试和增强现有模型，揭示了有效物理参数与蛋白质的依赖关系。

然后，GFNN方法通过交换一些深度U-Net的复杂性来进行可解释的操作，揭示了长程相互作用，甚至得到了描述系统行为的解析公式。这些方法代表了在数据驱动的生物物理建模框架中进行假设测试和公式制定的替代方法。

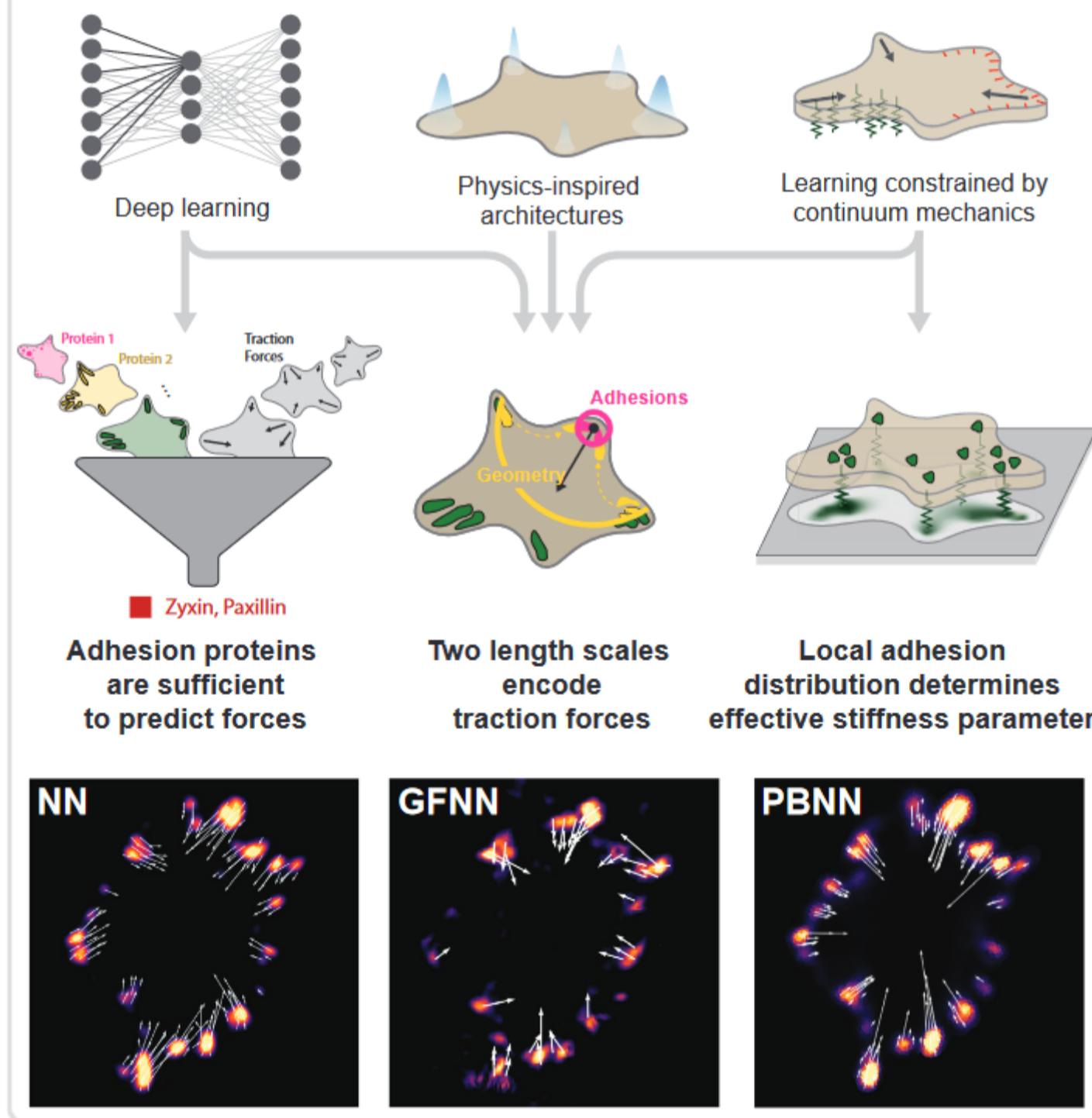
"长程相互作用"指的是物理系统中远距离的相互作用，其影响可以在空间较远的地方产生显著的效果。

此处所指蛋白质的相互作用



Data-driven biophysical modeling

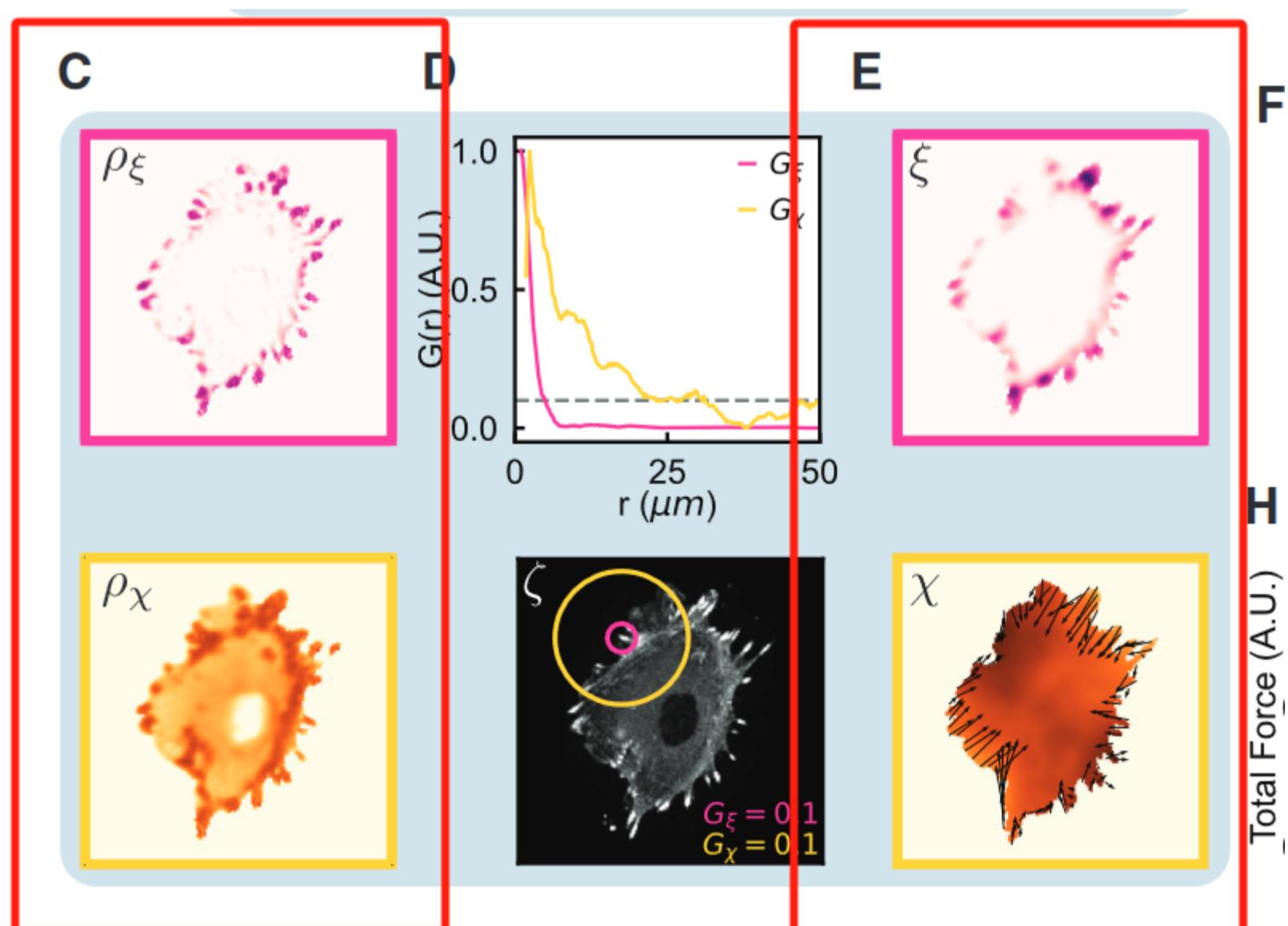
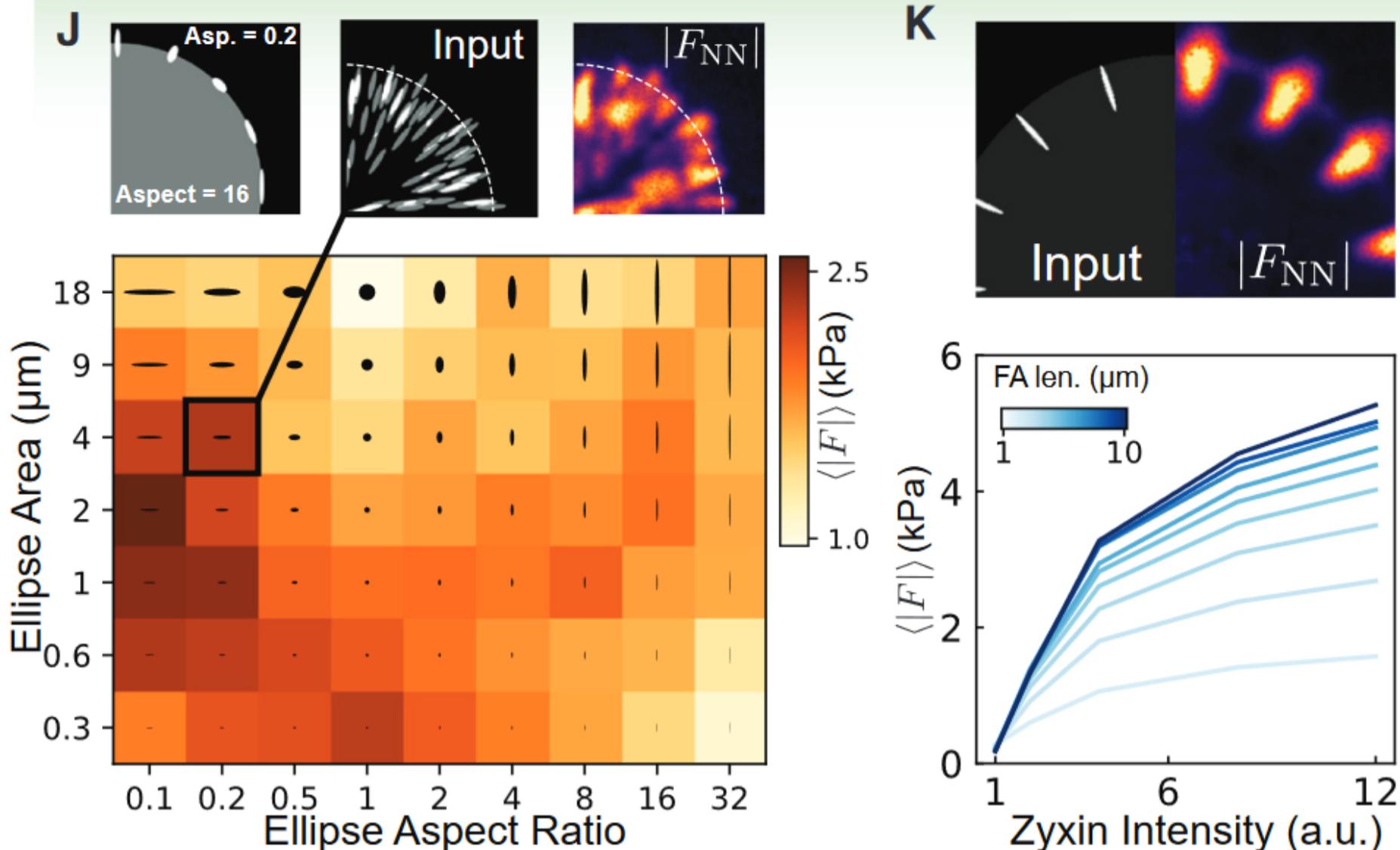
Identify relevant proteins **Learn governing rules** **Enhance mechanical models with biochemistry**



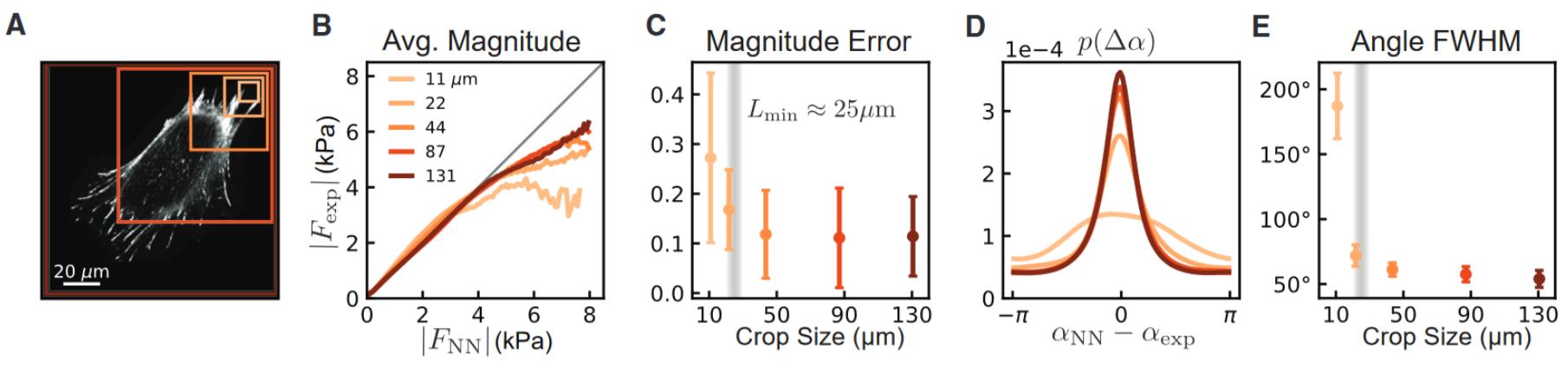
尽管三种方法都受到一定的限制和约束，但它们都揭示了两个重要的长度尺度。

几微米的一个长度尺度与单个FA的大小一致，描述了力大小与局部zyxin强度 (Intensity) 之间的关系。

Focal Adhesions



然而，预测力的方向需要编码在一个较大长度尺度上的信息。在GFNN和U-Net中，这个数十微米的尺度与细胞形态的各个方面有关，而在PBNN中，它在偏微分方程的边界条件中得到考虑。



此外，在PBNN和GFNN中，对应于长长度尺度(χ, \vec{u})和短长度尺度(ξ, Y)的场以惊人相似的方式耦合 ($\vec{F} = \xi \vec{\nabla} \chi$ 和 $\vec{F} = Y \vec{u}$)。神经网络是一个复杂的黑盒，分析其行为时必须小心。通过使用产生一致结果的多种方法，我们变得更加确信通过我们的数据驱动框架学到的规律是可泛化的，并且独立于我们用来推导它的方法。

从这些规律中，我们还可以推测为什么单个FA蛋白就足以进行准确预测。这些模型识别的较短长度尺度似乎直接由FA本身编码，而与细胞几何有关的较长长度尺度可以通过整合多个FA进行推断。额外的信息可能还可以从FA的几何形状和方向中获取，这些形状和方向是由它们耦合的应力纤维决定的。这种信息存在于多个FA蛋白的分布中，这可能解释了在我们对MDCK细胞的分析中所揭示的zyxin和paxillin之间的某种程度的可互换性。至于为什么zyxin似乎略优于paxillin，我们推测这可能与其对肌动蛋白和FA的力敏感吸引有关，但需要进一步的研究。

可解释的机器学习方法可以提高对调控生物系统时空行为的规则和方程的理解。这些方法可用于测试和改进现有模型，也可以在生物物理学的第一原理方法无法成功的领域学习全新的模型。作者仅考虑了从蛋白质预测力的情况，但一个自主的动态模型需要通过一个关系来封闭，该关系预测蛋白质分布随时间的演变。

利用插值的方法可以在随时间变化的预测上有不错的效果

作者的研究表明，仅考虑有效的黏附场的动态可能就足够了，而无需考虑细胞骨架重排的精确细节。这些方法可能有助于发展关于多种系统的机械-化学描述，如迁移细胞、上皮组织动力学和形态发生等。它们还可以与最近描述细胞器定位和相互作用的大量定量数据集相结合，以获取额外的见解。这些方法代表了利用机器学习的多才多艺来解决生命系统复杂性的一步。

method detail

Data processing

对于所有考虑的不同蛋白质，荧光图像都被规范化为在所有细胞中具有相似的数值。（归一化处理）

对于每个细胞，计算细胞掩模 (`mcell in`) 内荧光信号 f 的平均值，其中 $\mu_{in}^{cell} = \langle\langle f(x, t)_{x \in mask} \rangle\rangle$ ，以及掩模外 (`mcell out`) 荧光信号的平均值，其中 $\mu_{out}^{cell} = \langle\langle f(x, t)_{x \notin mask} \rangle\rangle$ 非掩模部分 `maskDt`。

然后，信号被规范化为 $f_{norm}(x, t) = (f(x, t) - \mu_{out}^{cell}) / (\mu_{in}^{cell} - \mu_{out}^{cell})$ ，并将任何负值（对应于低于空白空间噪声值的值）设置为 0。这确保 `fnorm` 的平均值约为 1。

细胞掩模是二进制的，通过对每个图像中的 zyxin 通道进行阈值处理并填充任何出现的空洞来生成。

由于基质准备的变化，每个细胞测得的力在一定程度上取决于它们所属的实验轮次。

因此，对每个细胞的力进行归一化，方法是通过其数据集内的平均值 $\mu_F^{day} = \langle\langle |F(x, t)| \rangle\rangle_{x, t} \text{ for } cell \in day$ 进行归一化，即 $\vec{F}_{norm}^{cell} = \vec{F}^{cell} / \mu_F^{day}$ ，确保每个细胞在特定实验轮次内得到归一化。在这项工作中，规范化的荧光信号和力被广泛使用。

Training data

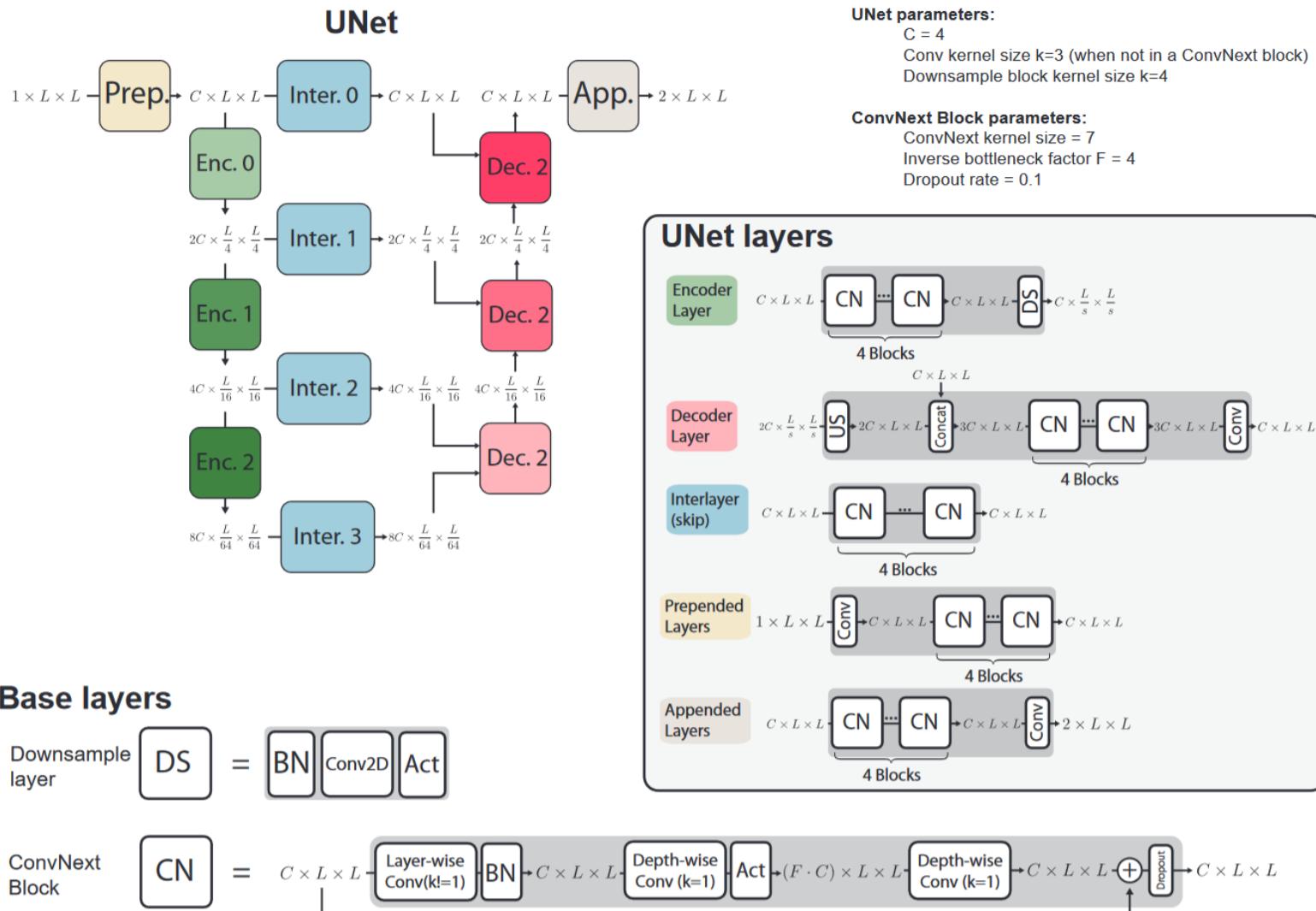
The training and testing data used for all networks in this work is shown below.

Overview table of the training and testing data used in this work. “D” stands for “day”, corresponding to datasets in the previous table. For the protein experiments (U-Net²), separate networks were, for each protein input, trained on all but one cell that was reserved for testing. For example, the dataset in row 5 (“D3-4”) contains 8 cells. We therefore train 8 identical U-Nets on the zyxin channel from 7 cells, withholding a different cell for testing each time.

Network	Figures	Trained on	Evaluated on
U-Net ¹	1, 4	(16 cells) D1-6, zyxin (first 30 frames of each cell)	D1: {2}, D2: {1, 2}, D3: {2, 3, 5} D4: {4, 5}, D5: {2, 4}, D6: {1, 3, 4, 5, 6}
	3A-3C	see above	U2OS
	3D-3F	see above	MDCK
	3G and 3H	see above	D4: {4}
U-Net ²	2	(8 cells) D3-4, zyxin OR mask OR actin OR zyxin and actin	(1 cell; each cell chosen once)
U-Net ²	2	(10 cells) D5, 7, zyxin OR mask OR paxillin OR zyxin and paxillin	(1 cell; each cell chosen once)
U-Net ²	2	(9 cells) D6, zyxin OR mask OR myosin OR zyxin and myosin	(1 cell; each cell chosen once)
U-Net ²	2	(6 cells) D8, zyxin OR mitoch. OR myosin OR zyxin and mitoch.	(1 cell; each cell chosen once)
U-Net ²	2	(10 cells) D9, paxillin OR actin OR paxillin and actin	(1 cell; each cell chosen once)
PBNN	5	D1: {3, 5}, D2: {2, 4}, D3: {1, 3}, D4: {1, 2, 4}, D6: {1}, zyxin	D1: {2, 4}, D2: {3, 5}, D3: {2, 5}, D4: {3, 5}, D6: {2, 3, 4, 5},
GFNN	6	D1 {1, 2} D2 {3}	D3 {1}

U-Net architecture

[www.github.com/schmittms/cell_force_prediction](https://github.com/schmittms/cell_force_prediction)



Channel structure for the U-Net used in Figures 1, 3, and 4. We set $C=4$, while L varies depending on input image size. Strided convolutions in the encoder layers have a stride of 4. The U-Nets in Figure 2 are the same, but do not have encoder block 2, skip block 3, or decoder block 2. They also have only 3 ConvNext blocks everywhere instead of 4.

Layer	Size in	Size out	Details
Prepended block	$1 \times L \times L$	$C \times L \times L$	Conv2d, 4× ConvNext blocks
Skip block 0	$C \times L \times L$	$C \times L \times L$	4× ConvNext blocks
Encoder block 0	$C \times L \times L$	$2C \times \frac{L}{4} \times \frac{L}{4}$	4× ConvNext, BN, Strided Conv2d, GELU
Skip block 1	$2C \times \frac{L}{4} \times \frac{L}{4}$	$2C \times \frac{L}{4} \times \frac{L}{4}$	(cf. skip 0)
Encoder block 1	$2C \times \frac{L}{4} \times \frac{L}{4}$	$4C \times \frac{L}{16} \times \frac{L}{16}$	(cf. encoder 0)
Skip block 2	$4C \times \frac{L}{16} \times \frac{L}{16}$	$4C \times \frac{L}{16} \times \frac{L}{16}$	(cf. skip 0)
Encoder block 2	$4C \times \frac{L}{16} \times \frac{L}{16}$	$8C \times \frac{L}{64} \times \frac{L}{64}$	(cf. encoder 0)
Skip block 3	$8C \times \frac{L}{64} \times \frac{L}{64}$	$8C \times \frac{L}{64} \times \frac{L}{64}$	(cf. skip 0)
Decoder block 2	$2C \times \frac{L}{64} \times \frac{L}{64}$	$C \times \frac{L}{16} \times \frac{L}{16}$	Upsample, Concat, 4× ConvNext, Conv2d
Decoder block 1	$6C \times \frac{L}{16} \times \frac{L}{16}$	$2C \times \frac{L}{4} \times \frac{L}{4}$	(cf. decoder 2)
Decoder block 0	$3C \times \frac{L}{4} \times \frac{L}{4}$	$C \times L \times L$	(cf. decoder 2)
Appended block	$C \times L \times L$	$2 \times L \times L$	4× ConvNext blocks, Conv2d

作者选择了卷积神经网络 (CNN)，因为它们的实现和训练过程相对直观。CNN还利用数据的空间结构来限制可训练参数的数量。虽然基于注意力的网络（如视觉变换器）在图像分析任务中也取得了成功，但它们带来了更高的计算和训练成本，并且需要非常庞大的数据集。为了在保持效率的同时实现与视觉注意网络相当的性能，因此采用了ConvNext设计。

发现 U-Net 结构取得了成功，该结构结合了激进的粗粒化和保留细粒度特征的跳跃连接，并且关键地能够从有限的数据中学到很好的泛化性能。

U-Net的通道结构如上所示。大多数层由具有ConvNext结构的块组成。简而言之，它们包括层内卷积、批量归一化、反向瓶颈深度卷积、激活函数，最后是深度卷积。我们的ConvNext块具有7的层内卷积核大小，并通过4的倍数增加反向瓶颈中的通道数。对于所有其他卷积，我们使用3的卷积核大小。使用10%的丢失概率进行丢失。架构的详细说明如图S2所示。编码（粗粒化）层的数量由我们在论文中处理的最小图像大小确定。由于最小图像大小为 64×64 ，有4个 3×4 的下采样卷积，这限制了我们只能使用3个编码层。

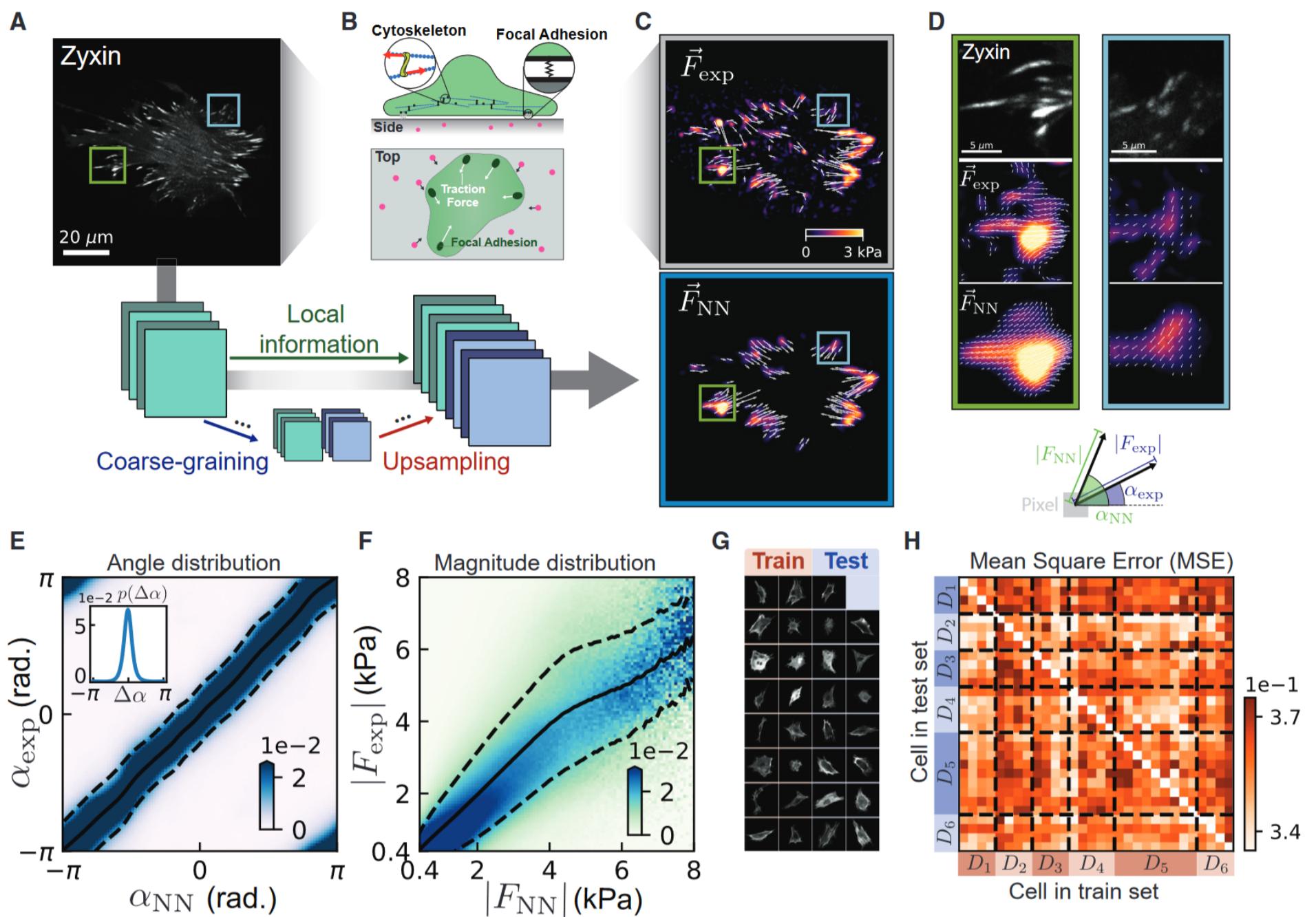
ConvNext块的超参数，包括卷积核大小、反向瓶颈宽度和归一化层，根据在文献中找到的最佳值进行选择。（？）

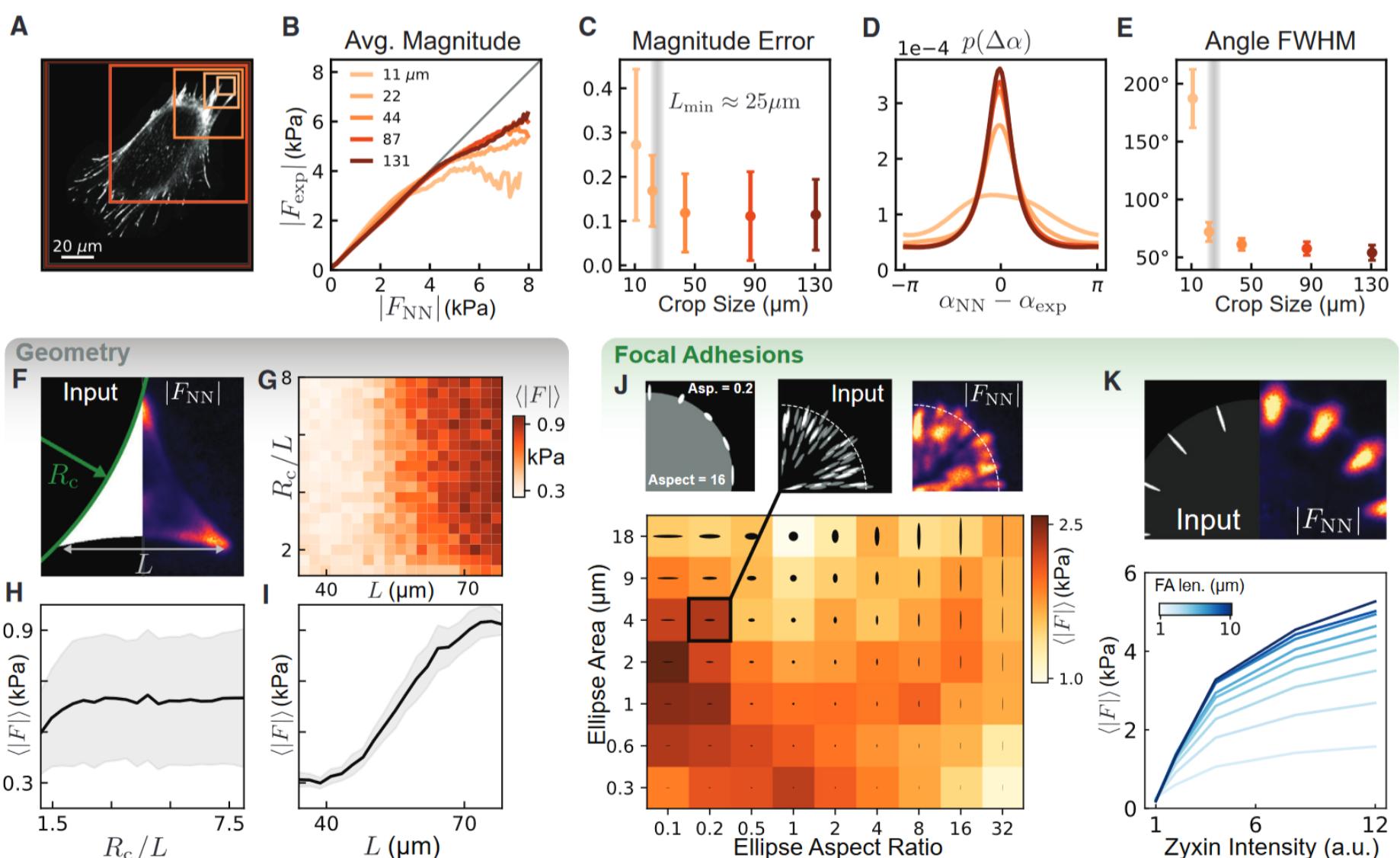
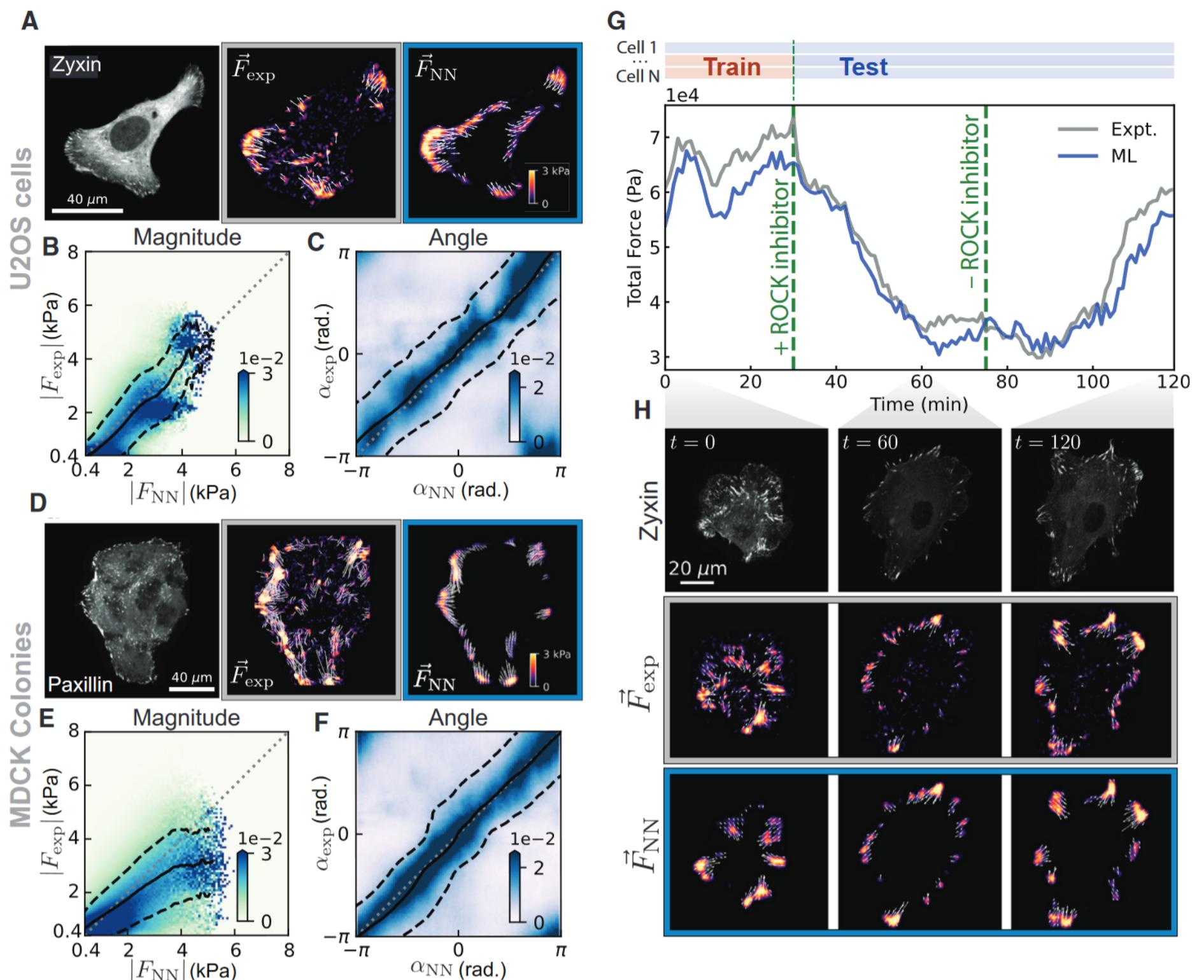
没通过调参，很大概率效率和性能是否存在问题呢？

U-Net使用Adam优化器进行训练，带有权重衰减，初始学习率为0.001。

学习率以0.99的速率指数衰减。我们使用批量大小为8。

在下三图使用的U-Net的训练数据包括从16个细胞的时间序列中随机抽取的480个帧（总计31个细胞）。





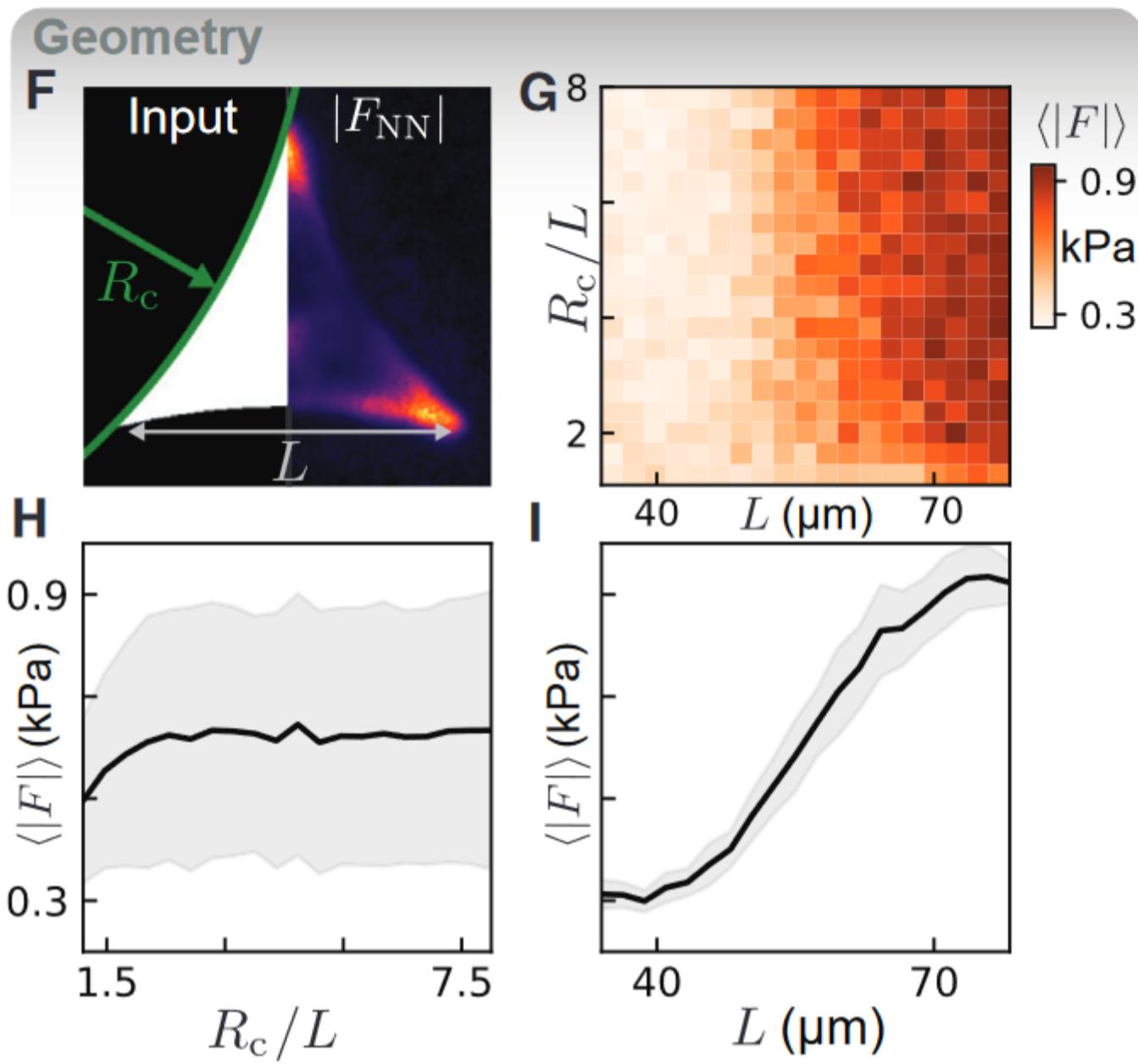
使用的U-Net的训练数据包括从不同数量细胞的随机抽取的600个帧。每个数据样本包含一个输入图像 (zyxin、另一个蛋白质、掩膜或双通道 zyxin + 蛋白质图像) 以及通过TFM测得的相应牵引力地图。牵引力地图有两个通道，我们表示为幅度和角度，而不是x和y分量。

在所有情况下，网络训练了300个epochs。作为损失函数，我们采用幅度分量的MSE和角度的 2π 周期性MSE。

Synthetic cells

考虑了三种合成细胞的变体

第一种，变体捕捉了细胞几何形状的大尺度特征。

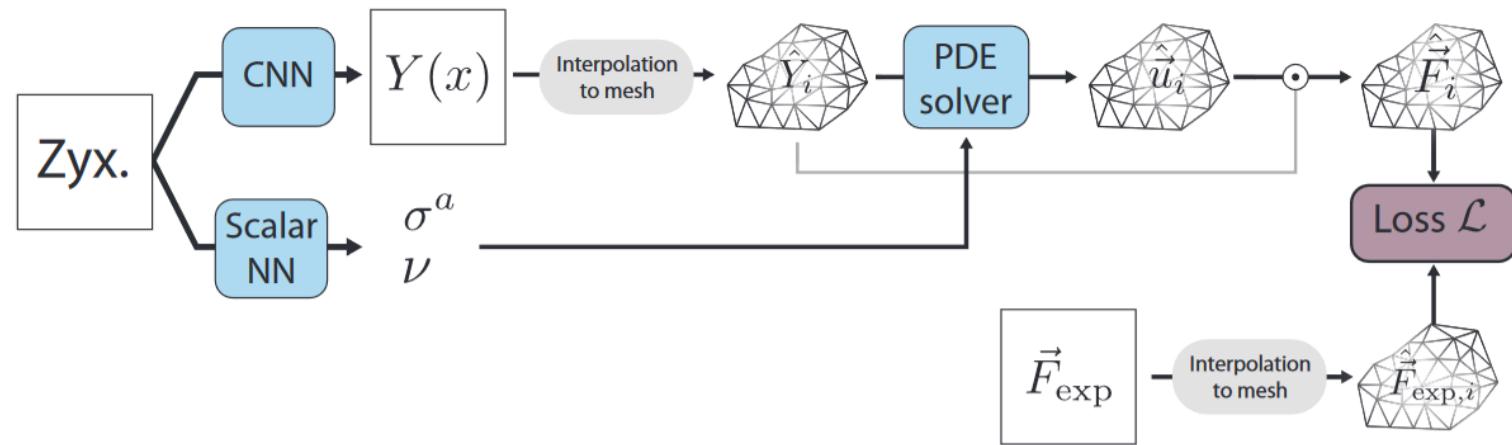


第二种，合成细胞模拟了细胞中类似黏附点的物体的分布。

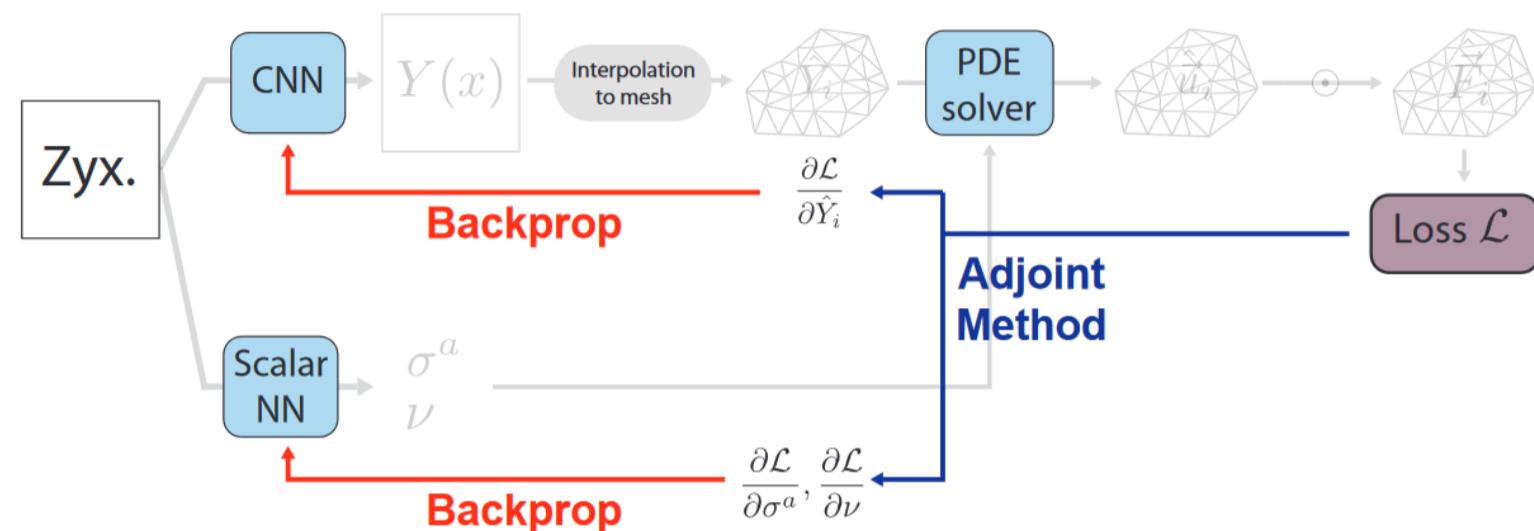
第三种，利用神经网络合成细胞。（基于长度尺寸依赖性）

Physical Bottleneck Forward and Backward Passes

Forward Pass



Backward Pass



参考文献

1. U-net
2. convNext <https://zhuanlan.zhihu.com/p/458016349>