

利用类视觉问答系统预测药物与蛋白质的相互作用

Shuangjia Zheng^{1,2,3}, Yongjian Li^{2,3}, Sheng Chen², Jun Xu¹✉ and Yuedong Yang¹✉

识别新的药物与蛋白质相互作用对于药物研发至关重要。为此，基于药物描述符和一维蛋白质序列，已开发出许多基于机器学习的方法。然而，蛋白质序列无法准确反映三维空间中的相互作用。但是，由于三维矩阵的稀疏性，直接输入三维结构效率低下，而且可用的共晶结构数量有限也阻碍了这种方法的应用。在此，我们提出了一种端到端的深度学习框架，通过用单体结构的二维距离图（图像）表示蛋白质，用分子线性表示法（字符串）表示药物，遵循视觉问答模式来预测相互作用。为了高效训练该系统，我们引入了一种动态注意力卷积神经网络，从可变长度的距离图中学习固定大小的表示，以及一种自注意力序列模型，自动从线性表示中提取语义特征。大量实验表明，我们的模型在有用的诱饵目录增强版（DUD-E）、人类和 BindingDB 基准数据集上，与最先进的基线模型相比，取得了具有竞争力的性能。进一步的注意力可视化蛋白质和药物分子的突出区域提供了生物学解释。

药物与蛋白质相互作用（DPIs）的预测对于药物的设计和开发至关重要。尽管实验测定仍是确定药物与蛋白质相互作用最可靠的方法，但由于实验成本高昂且工作量巨大，对每一种可能的药物与蛋白质组合进行实验表征是极其困难的。

因此，药物-蛋白质相互作用（DPIs）的计算预测近来取得了迅速进展。总体而言，其大致可分为两类：基于物理的方法和机器学习方法。基于物理的方法，如分子对接，应用受物理启发的预设能量函数来评估原子水平上的药物-蛋白质相互作用。然而，由于难以评估构象熵和溶剂贡献，这些方法通常准确性有限。此外，这些原子级方法对结构波动敏感，难以很好地处理蛋白质的灵活性。

随着近期蛋白质结构数据和蛋白质-配体相互作用数据集的增加，基于机器学习的方法取得了迅速进展³⁻⁵。通常，预测被当作一个二分类任务来处理，在一个统一的框架中整合配体、蛋白质及其相互作用的信息。

药物分子可以通过其线性表示很好地呈现出来，因为大多数药物包含的重原子少于 100 个，因此其结构空间相对较小。近期研究已证明，当前的深度学习技术能够从其线性表示中准确预测结构特性^{6,8}。相比之下，蛋白质分子要大得多，通常包含超过 1000 个重原子。从一维（1D）序列预测三维（3D）结构的过程，即蛋白质折叠，是一个众所周知的难题。因此，传统的通过 1D 蛋白质序列进行表示不足以捕捉决定药物-蛋白质相互作用（DPIs）预测的 3D 空间中的结构特征。尽管在近期的研究中尝试直接输入 3D 结构^{3,9,10}，但由于一些原因，其准确率相对较低。首先

，不规则的蛋白质 3D 结构需要大规模的 3D 矩阵来容纳整个结构。高维、稀疏的矩阵导致了大量繁琐的输入变量。其次，这些研究因高质量的三维结构数量较少而受到影响，因为它们需要蛋白质-配体对的共晶结构，而这种结构通过实验确定起来十分困难。

作为一种平衡，蛋白质也可以通过二维成对距离图来表示。距离图是通过构成蛋白质的氨基酸之间的成对接触来紧凑地表示蛋白质的三维结构。先前的研究表明，距离图可用于生成和比较蛋白质的三维结构。

受这些研究的启发，我们在此利用二维距离图来表示蛋白质，从而将药物-蛋白质相互作用（DPI）任务转化为一个经典的视觉问答（VQA）问题。给定一张图像及其对应的问题，VQA 系统需要给出正确的答案。早期的研究中，VQA 任务涉及回答“是”或“否”或单个单词，被视为与 DPI 预测类似的分类任务。在此框架中，图像为蛋白质的距离图，问题为药物的简化分子线性输入规范（SMILES），答案为它们是否会相互作用。这种框架使得可以在没有蛋白质与其结合配体的共晶结构的情况下，仅基于蛋白质单体结构进行训练，从而极大地扩展了可用于训练的数据集。

然而，视觉问答（VQA）与距离对齐预测（DPI）之间存在差异。首先，在许多 VQA 场景中，图像大小可以调整为固定值，但成对距离图中的每个像素都代表一对氨基酸之间的关系，如果对其进行下采样，会导致信息丢失。其次，SMILES 的语法与自然语言不同，这迫使我们采用定制的分词过程和合适的模型来获取分子线性表示的语义特征。第三，我们的训练集仍然比其他应用小得多，这要求我们精心设计网络。

School of Pharmaceutical Sciences, Sun Yat-sen University, Guangzhou, China. ²National Supercomputer Center in Guangzhou, Sun Yat-sen University, Guangzhou, China. ³These authors contributed equally: Shuangjia Zheng and Yongjian Li. ✉e-mail: junxu@biochemomes.com; yangyd25@mail.sysu.edu.cn

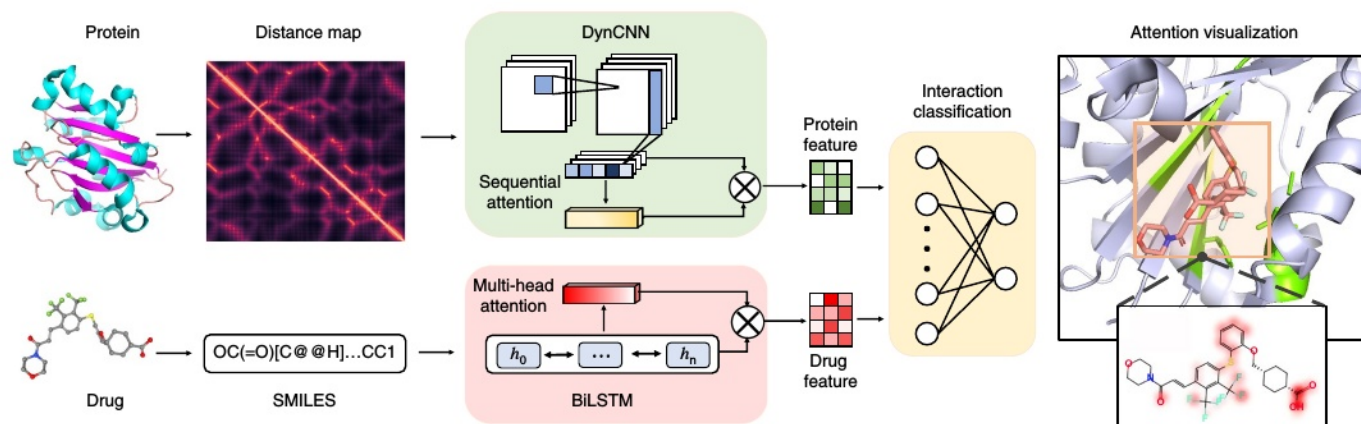


图 1 | 所提出的 DrugVQA 模型框架。该模型由两个主要组件构成：具有顺序注意力机制的动态卷积神经网络（CNN）和具有多头自注意力机制的双向长短期记忆网络（BiLSTM）。所学习到的注意力权重能够实现对每个蛋白质残基和配体成分对分类决策的个体贡献的可视化。

为了解决上述问题，我们提出了一种受视觉问答（VQA）启发的可解释模型，该模型直接从蛋白质距离图和分子 SMILES 预测药物-蛋白质相互作用（DPI）。距离图和 SMILES 分别通过动态卷积神经网络（DynCNN）和带有注意力机制的双向长短期记忆网络（BiLSTM）进行编码，其输出被连接到全连接层以进行预测。所提出的模型在三个公开的 DPI 数据集上优于最先进的方法。此外，所学的注意力机制能够可视化蛋白质和配体结合区域之间的个体贡献，这对于配体优化非常重要。

相关工作

在本节中，我们将介绍两个领域的相关工作。首先，我们简要回顾当前用于视觉问答（VQA）任务的模型。其次，我们将重点总结当前用于深度疼痛指数（DPIs）预测的方法。

视觉问答。视觉问答（VQA）的任务是根据给定的图像回答相关问题。早期的 VQA 系统倾向于通过直接学习卷积神经网络（CNN）和循环神经网络（RNN）提取的图像和文本特征的联合表示来推断答案^{18,21}。后来，基于注意力的模型^{22,23}学会关注为问题提供信息证据的视觉组件，取得了巨大成功。最近，Schwartz 等人²⁴提出了多模态注意力机制，该机制不仅关注图像中的对象，还关注问题。在此，我们采用多模态注意力框架来捕捉药物分子和蛋白质的关键成分。

药物-蛋白质相互作用（DPI）预测。基于对接的方法，如参考文献^{1,2}所述，被广泛用于根据药物化合物和蛋白质的三维结构输入来预测结合模式和亲和力。这些方法应用预定义的力场来估算结合分数，从而在原子水平上评估药物-蛋白质相互作用。

基于机器学习的方法也被用于预测药物-蛋白质相互作用。例如，Bleakley 和 Yamanishi 提出了二分图局部模型，通过从化学结构相似性和蛋白质序列相似性训练局部支持向量机分类器来实现。Ballester 和 Mitchell 利用随机森林算法来捕捉分子对接过程中的结合效应。Durrant 和 McCammon 提出了一种基于全连接神经网络的评分函数，用于表征蛋白质-配体复合物的结合亲和力。Tabei 和 Yamanishi 进一步通过使用具有更紧凑的化合物-蛋白质对指纹的哈希算法来改进药物-蛋白质相互作用的预测。

近期，深度学习技术已被引入，通过直接使用 3D 蛋白质-化合物复合物来预测药物-蛋白质相互作用（DPIs）^{3,9,10}。由于输入特征基于围绕口袋-配体复合物定义的 3D 矩阵，这些方法产生了大量的输入变量，并且由于训练集数量有限而受到影响。此外，尽管一些基于图神经网络的表示学习研究基于分子拓扑结构和蛋白质序列或其功能注释来预测 DPIs^{4,5,7}，但由于缺乏蛋白质结构信息，其准确性有限。

深度学习框架

在本节中，我们将介绍基于视觉问答（VQA）的药物蛋白质相互作用（DPIs）预测模型（DrugVQA）的框架。

问题表述。我们的任务是预测药物化合物与目标蛋白质之间的相互作用。具体来说，药物化合物以 SMILES 格式表示，这是一种基于化学键规则的拓扑信息文本字符串。例如，苯环可以编码为 ‘c1ccccc1’。每个小写的 ‘c’ 表示一个芳香碳原子，‘1’ 表示一个环的开始和结束。由于氢原子可以通过简单规则推断出来，所以它们未显示。为了保留重要的化学特征，我们使用以下正则表达式对药物分子进行分词，该表达式借鉴了参考文献²⁹的工作：

$$\text{token}_{\text{regex}} = "([\{ 1, 6 \}])" \quad (1)$$

此外，我们按照以下规则替换了多字符符号：“Br” : R, “Cl” : L, “Si” : A, “Se” : Z。

假设我们有一个包含 n 个标记的药物分子线性表示，该分子可以用分子嵌入序列 $M = (t_1, \dots, t_n)$ 来表示，其中 t_i 是第 i 个标记的 d 维标记嵌入向量。因此， $M \in \mathbb{R}_n \times d$ 是通过将所有标记嵌入向量连接在一起而形成的 2D 矩阵表示。

同样地，蛋白质可以简单地描述为一个线性序列，由一系列氨基酸残基组成 $P = (r_1, \dots, r_l)$ ，其中 r_i 是一个长度为 20 的一热编码向量，代表位置 i 处的氨基酸类型， l 为序列长度。为了捕捉结构信息，我们转而将蛋白质描述为通过以下公式计算得出的二维成对距离图：

$$\hat{s}(r_i, r_j) = \frac{1}{1 + d(r_i, r_j)/d_0}, r_i, r_j \in P \quad (2)$$

表 1 | 在人类、BindingDB 和 DUD-E 数据集上的消融实验结果

消融试验	人类	BindingDB	杜德-E
长短期记忆网络+卷积神经网络	0.005	0.913±0.003	0.937 ± 0.003
注意力机制长短期记忆网络+卷积神经网络	0.005	0.928±0.002	0.951±0.005
长短期记忆网络+注意力机制卷积神经网络	0.005	0.921 ± 0.003	0.946±0.004
VQA-seq	0.964±0.005	0.897 ± 0.004	0.948 ± 0.003
DrugVQA	0.979 ± 0.003	0.936 ±0.002	0.972 ± 0.003

前三个模型展示了顺序注意力和自注意力模块的有效性。VQA-seq 表示用蛋白质序列替换蛋白质距离图的版本。我们报告了 DrugVQA 在使用不同随机种子的三次运行中的平均 AUC 值和标准差。Att, 注意力模块。

其中 $d(r_i, r_j)$ 为残基 i 和 j 的 Ca 原子之间的距离, d_0 设为 3.8 埃, 这是相邻 Ca 原子之间的距离。令 $\hat{s}_i = \frac{1}{2} \sqrt{\frac{1}{d_0} \sum_{j \in P} d(r_i, r_j)^2}$ 为 r_i 中 P 内所有残基的 l 维距离向量, 蛋白质可以表示为一个距离矩阵:

$$P = [\hat{s}_1, \hat{s}_2, \dots, \hat{s}_l]_{l \times l} \quad (3)$$

深度蛋白质相互作用 (DPI) 预测的目标是学习一个系统, 该系统以一对 $(M; P)$ 作为输入, 并输出标签 $y \in \{0, 1\}$, 其中 $y = 1$ 表示 M 和 P 之间存在相互作用。

如图 1 所示, 我们的 DrugVQA 模型由两个主要组件构成: 具有顺序注意力机制的动态卷积神经网络 (“动态注意力卷积神经网络”) 和具有多头自注意力机制的双向长短期记忆网络 (“自注意力双向长短期记忆网络”)。

动态注意力卷积神经网络。动态处理过程。在我们的模型中, 采用了一种经过调整的卷积神经网络 (CNN) 将蛋白质距离图编码为固定大小的向量表示。该 CNN 模块由堆叠的残差块和一个顺序自注意力块组成。我们使用的残差块借鉴自 Resnet30, 每个残差单元由一个 5×5 的卷积层和一个 3×3 的卷积层堆叠而成。我们采用指数线性单元 (ELU) 而非传统使用的修正线性单元 (ReLU)。与通常将图像预处理为相同大小的视觉问答 (VQA) 任务不同, 现实世界中的蛋白质由不同长度的氨基酸组成, 无法进行缩放。因此, 我们希望设计一种动态神经网络, 能够 (1) 处理不同长度的输入, 以及 (2) 预测每个氨基酸的重要性。为此, 我们首先移除了残差块之间的池化层, 并在输入的两侧使用零填充, 以确保残差块的结果与输入具有相同的大小。具体来说, 给定一个蛋白质距离图 $P \in \mathbb{R}^{l \times l \times N_f}$, 最后一个残差块的输出维度仍为 $l \times l \times N_f$, 其中 N_f 是最后一个卷积层的滤波器数量。之后, 我们使用平均池化来压缩残差块的信息丰富的输出, 以供下游处理。

顺序注意力机制。通过平均池化, 我们获得一个卷积蛋白质特征图 $P_c \in \mathbb{R}^{l \times N_f}$ 。实际上, P_c 可以被视为蛋白质的顺序表示, 其中 l 是蛋白质中的氨基酸 (位点) 数量, N_f 表示每个位点的空间特征。由于大多数位点与药物结合并无直接关联, 识别出小部分结合位点对于准确预测药物-蛋白质相互作用至关重要。为了处理卷积层中特征图大小的变化, 并突出重要的结合位点, 我们采用了一种顺序自注意力机制³³来充分利用这些特征进行分类。具体而言, 注意力机制以 P_c 作为输入, 并输出一个权重向量 a_p (蛋白质的注意力矩阵):

表 2 | 提出模型与基准模型在人类数据集上的对比结果

方法	AUC	召回	精度
k 近邻算法	0.860	0.927	0.798
射频	0.940	0.897	0.861
L2	0.911	0.913	0.861
图神经网络	0.970	0.918	0.923
DrugVQA	0.979 ± 0.003	0.961 ±0.002	0.954 ±0.003

$$a^p = \text{softmax}(W_{p2} \tanh(W_{p1} P_c^T)) \quad (4)$$

$$\sum_{i=1}^l (a_i^p) = 1, \forall i, 1 \leq i \leq l$$

其中 $W_{p1} \in \mathbb{R}^{d_p \times N_f}$, 而 W_{p2} 是一个参数向量, 具有 d_p 尺寸 d_p , 其中 d_p 是一个可调的超参数。这种向量表示通常关注蛋白质序列中的一组连续位点。由于蛋白质结合口袋由空间上相邻的多个连续位点组成, 我们进一步将 W_{p2} 扩展为一个 r_p -by- d_p 矩阵, 记为 W_{p2} , 以捕获结合口袋的整体结构信息。因此, a_p 被转换为一个多头注意力权重 $A_p \in \mathbb{R}^{r_p \times l}$, 如下所示

$$A^p = \text{softmax}(W_{p2} \tanh(W_{p1} P_c^T)) \quad (5)$$

实际上, 方程 (5) 可以被视为一个两层的多层感知机 (MLP), 且没有偏置项, 其隐藏单元数量为 d_p , 参数为 $\{W_{p1}$ 和 $W_{p2}\}$ 。我们通过将注释矩阵 A^p 与特征图 P_c 相乘来计算 r_p 加权和。

$$P_a = A^p P_c \quad (6)$$

其中 P_a 是一个注意力特征图, 包含各站点对交互作用贡献之间的潜在关系。 P_a 的大小为 r_p -by- N_f , 其中 r_p 是一个可调的超参数, 代表注意力向量的数量。

自注意力双向长短期记忆网络。每个药物分子的 SMILES 字符串都被编码为一个二维嵌入矩阵 $M \in \mathbb{R}^{n \times d}$ 。分子矩阵 M 中的标记向量彼此独立。为了在分子内相邻标记之间获得一定的依赖关系, 使用双向长短期记忆网络 (BiLSTM) 来处理分子:

$$\vec{h}_i = \vec{\text{LSTM}}(t_i, \vec{h}_{i-1}) \quad (7)$$

$$\overleftarrow{h}_i = \overleftarrow{\text{LSTM}}(t_i, \overleftarrow{h}_{i+1}) \quad (8)$$

\vec{h}_i 与 \overleftarrow{h}_i 相连接, 并得到一个隐藏状态 h_i is obtained to 替换词嵌入 t_i , 因此 h_i 成为一个信息更丰富的向量, 在分子中相邻标记之间获得了一些依赖关系。为简单起见, 我们将每个时间步 i 中的所有 h_i 记为 H 。

$$h_i = (\vec{h}_i, \overleftarrow{h}_i) \quad (9)$$

$$H = (h_0, h_1 \dots h_n) \quad (10)$$

如果每个单向 LSTM 的隐藏单元数设为 u , 则 H 的形状将是 $n \times 2u$ 。

下一个目标是了解分子的哪一部分对相互作用的预测贡献最大。换句话说, 我们想要

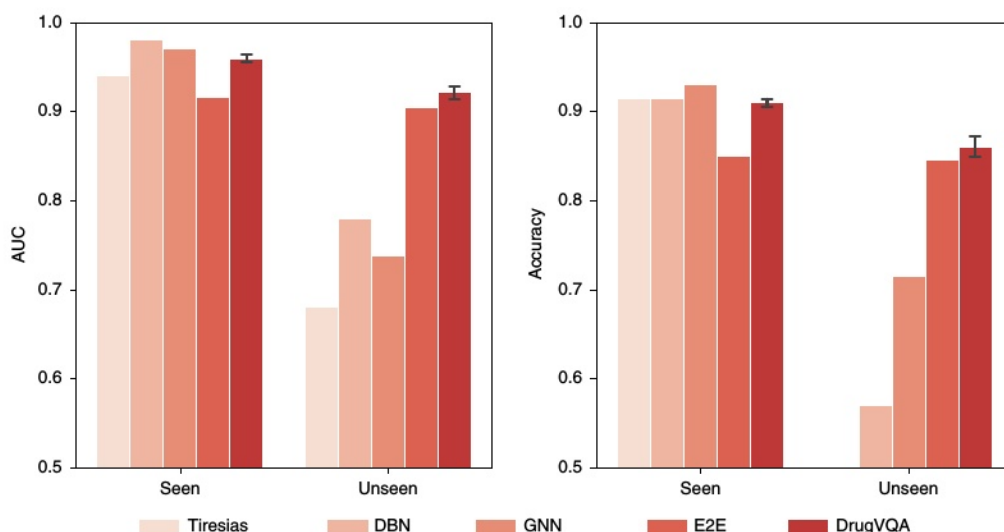


图 2 | 我们所提出的方法与基准方法在来自 BindingDB 数据集的已见和未见蛋白质靶点上的性能比较。误差线表示标准偏差。

表 3 | 在 DUD-E 数据集的不同靶点上，所提出的 DrugVQA 模型与不同类型的基准模型的比较

类别	模型	AUC	0.5% RE	1.0% 铈 (RE 代表稀土元素)	1.5% 铈 (RE 代表稀土元素)	2.0% 铈 (RE 代表稀土元素)
机器学习评分函数	NN得分	0.584	4.166	2.980	2.460	1.891
	RF 分数	0.622	5.628	4.274	3.499	2.678
基于对接的	维纳	0.716	9.139	7.321	5.811	4.444
	Smina	0.696	-	-	-	-
基于深度学习的	三维卷积神经网络 (3D-CNN)	0.888	42.559	26.655	19.363	10.710
	原子网络	0.895	-	-	-	-
	掌上GameCube	0.886	44.406	29.748	19.408	10.735
	图神经网络	0.940	-	-	-	-
提议的	DrugVQA	0.972 ± 0.003	88.17 ± 4.8	58.71 ± 2.74	35.06 ± 1.91	17.39 ± 0.94

DrugVQA 的结果是三折交叉验证中 AUC 和 RE 的平均值和标准差。

确定标记与相互作用之间的关系，这可用于化学家设计或改进化学化合物。同样，我们通过将多头自注意力机制引入双向长短期记忆网络 (BiLSTM) 来实现这一目标。注意力机制将整个 LSTM 隐藏状态 H 作为输入，并输出一个权重向量 A^m (分子的注意力矩阵) 作为

$$A^m = \text{softmax}(\text{MLP}(H^T)) \quad (11)$$

其中多层感知机 (MLP) 的隐藏单元数量为 d_m ，参数为 $\{W_{m1}, W_{m2}\}$ 。我们通过将注释矩阵 A_m 与 LSTM 隐藏状态 H 相乘来计算加权和，所得矩阵即为自注意力分子嵌入：

$$M_a = A^m H \quad (12)$$

其中 M_a 是一个自注意力药物分子特征图，它包含了各标记交互贡献之间的潜在关系。 M_a 的大小为 $r_m \times \text{by-}2u$ 。

对于 P_a 和 M_a ，我们对所有注意力向量求和，然后对得到的权重向量进行归一化处理，使其总和为 1。这一过程使我们能够获得两个信息丰富的 1D 向量 Pb_a 和 Mc_a ，它们将被输入到分类层。我们

将 Pb_a 和 Mc_a 拼接起来，即 $[Pb_a; M]$ ，并得到一个输出 $o \in R$ ：

其中， $W_o \in \mathbb{R}^{(2r_m + 2u) \times 1}$ 是权重矩阵， $b_o \in \mathbb{R}^1$ 是偏置。

给定数据集 $D = \{(m_i, p_i, y_i)\}$ ，训练目标是将交叉熵最小化，具体如下：

$$\mathcal{L}_{CE}(\theta) = - \sum_{i=1}^N (y_i \log(\sigma(o_i)) + (1 - y_i) \log(1 - \sigma(o_i))) + \frac{\lambda}{2} \|\theta\|_2^2 \quad (14)$$

其中， Θ 表示系统中所有权重矩阵和偏置向量的集合， N 表示训练数据集中药物 - 蛋白质对的总数， σ 表示 S 型函数。 λ 是 L2 正则化超参数。

实验

在本节中，我们将介绍所使用的数据集和实验，以表明我们的 DrugVQA 模型的性能。

数据集。为了能够将 DrugVQA 与现有的基于机器学习的方法 and 对接程序进行直接对比，

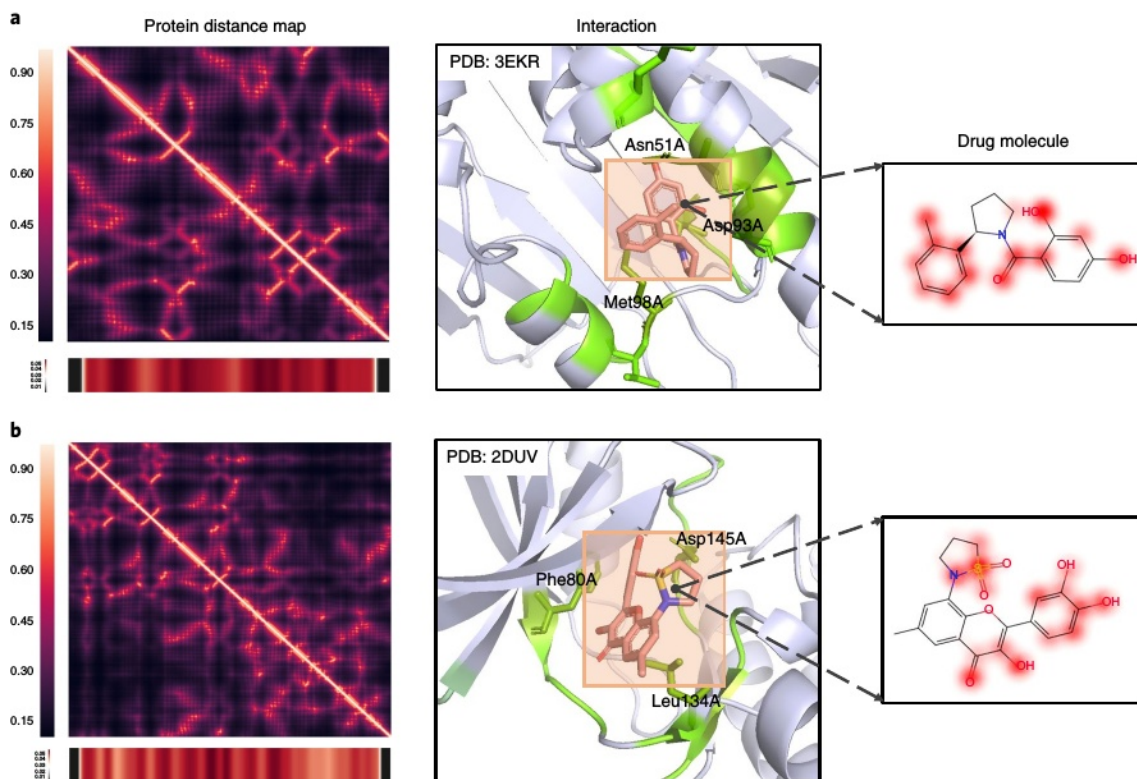


图 3 | 药物口袋与配体对的重要性可视化。 **a**, 3EKR。 **b**, 2DUV。相应的序列注意力条位于成对图的下方。绿色突出显示了结合口袋内部或周围的高注意力位点，红色云团表示具有注意力系数的药物原子。颜色越深表示注意力系数越高。

我们在三个公开的药物-蛋白质相互作用（DPI）数据集上评估了我们提出的模型：有用的诱饵目录增强版（DUD-E）数据集、人类数据集和 BindingDB 数据集。

DUD-E 数据集是一个著名的基准数据集，包含 8 个蛋白质家族的 102 个靶点。每个靶点平均有 224 个活性分子和超过 10,000 个非活性分子。这些非活性分子在物理性质上与活性分子相似，但在拓扑结构上不同。最终的数据集包含 22,645 个正例和 1,407,145 个负例。我们采用三折交叉验证策略在 DUD-E 数据集上训练和评估我们的模型，遵循参考文献³的方法。折叠是在靶点之间进行划分的，即同一靶点的所有配体都属于同一折叠。为避免同源蛋白质的影响，属于同一蛋白质家族的靶点严格保持在同一折叠中。为了快速训练模型，我们使用平衡集（每个靶点的所有正例和随机选择的等量负例）进行训练，但在评估时仍使用整个数据集（不平衡的）。

人类。该数据集由参考文献³⁵创建，其中包含通过系统筛选框架获得的高可信度的化合物-蛋白质配对的负样本。依照参考文献⁴，我们使用了一个平衡数据集，其中正负样本的比例为 1:1。最终，人类数据集包含 5423 个相互作用和 1803 种独特的蛋白质。同样依照参考文献⁴，我们采用 80%/10%/10% 的随机划分进行训练/验证/测试。

我们进一步选择 BindingDB 数据集³⁶作为真实世界数据集来评估我们的模型。BindingDB 是一个公开的实验测量结合亲和力和数据库，主要关注小分子与蛋白质之间的相互作用。在我们的实验中，我们使用了参考文献⁵构建的定制化 BindingDB 数据集

来进行直接对比。该数据集包含来自 BindingDB 的 39,747 个正例和 31,218 个负例。数据集被分为三部分：训练集（50,155 个相互作用）、验证集（5,607 个相互作用）和测试集（5,508 个相互作用）。为了进一步验证我们的模型对未见过的蛋白质的泛化能力，我们将测试相互作用分为两部分，一部分是训练集中出现过的蛋白质，另一部分是未出现过的蛋白质。

实施与评估策略。实施细节。

我们使用 Pytorch^{0.4.037} 实现了所提出的模型。在所有数据集上，训练过程最多持续 50 个周期，采用 Adam 优化器³⁸，学习率为 0.001，批处理大小为 1。考虑到所用 GPU（GTX1080Ti 12GB）的内存限制，我们采用了 32 个残差块，分别具有 16 和 32 个滤波器。双向长短期记忆网络（BiLSTM）的隐藏状态设置为 64（即“Self-attentive BiLSTM”中的 u ），并且在 BiLSTM 上应用了 0.2 的丢弃率。此外，卷积神经网络（CNN）和 BiLSTM 中的注意力多层感知机（MLP）都具有一个 100 个单元的隐藏层（即 d_p 和 d_m ），并且我们选择矩阵嵌入具有 10 行（蛋白质的 r_p ）和 20 行（药物的 r_m ）。L2 正则化的系数为 0.001。我们在广泛的范围内探索了超参数，并发现上述超参数组合在人类数据集上表现最佳。超参数空间和人类验证集上不同超参数的学习曲线见补充表 1 和补充图 1。我们注意到，在其他两个数据集上，其余的超参数未进行调整，因为我们发现模型性能对合理的设置不敏感。所有实验均重复了三次，每次使用不同的随机数种子。

评估指标。性能通过受试者工作特征曲线下的面积（AUC）进行评估。此外，

对于人类数据集，我们按照参考文献⁴报告了准确率和召回率的值。对于 DUD-E 数据集，我们按照参考文献³的工作报告了 ROC 富集度指标 (RE)。具体而言，RE 得分定义为在给定的假阳性率 (FPR) 阈值下，真阳性率与假阳性率的比值。在此，我们报告了 0.5%、1%、2% 和 5% FPR 阈值下的 RE 得分。对于 BindingDB 数据集，我们还按照参考文献⁵报告了准确率。

消融实验。我们在三个基准数据集上进行了消融实验，以探究影响所提出的 DrugVQA 框架性能的因素。结果如表 1 所示。我们首先研究了注意力模块的影响。如前 3 行所示，移除注意力模块会使 AUC 降低 2.3% - 3.9%。这表明两个注意力模块的有效性。为了验证蛋白质距离图的影响，我们还纳入了一个版本的模型，用蛋白质序列替换了蛋白质距离图（记为 VQA-seq）。蛋白质序列通过自注意力双向长短期记忆网络 (BiLSTM) 处理为药物简化分子线性输入规范 (SMILES)。如最后两行所示，使用距离图在所有三个数据集上都带来了稳定的提升。这与我们的预期相符，即距离图比序列包含更多用于预测药物 - 蛋白质相互作用 (DPI) 的信息。

在人类数据集上的比较。对比模型。在本节中，我们在人类数据集上将我们的 DrugVQA 与最先进的药物-蛋白质相互作用 (DPI) 方法进行了比较。我们将其与 k 近邻 (k-NN)、随机森林 (RF)、L2-逻辑回归 (L2)（结果来自参考文献³⁵）以及图神经网络 (GNN)⁴进行了对比（我们使用与原始论文相同的参数设置重新训练了该模型）。我们报告了 DrugVQA 在三次随机种子运行中的平均 AUC、召回率、精确率及其标准差。

结果。如表 2 所示，我们的 DrugVQA 系统优于当前最先进的 GNN 模型，AUC、召回率和准确率分别提高了 0.9%、4.3% 和 3.1%。这一现象符合我们的预期。我们的系统包含蛋白质结构信息，这有助于学习更丰富的蛋白质表示。其他基于描述符的机器学习技术表现不佳，AUC 在 0.86 至 0.94 之间，这表明端到端学习的表示能够从蛋白质和药物中学习到的对药物 - 蛋白质相互作用预测重要的信息。

在 BindingDB 数据集上的比较。对比模型。我们进一步在 BindingDB 数据集上评估我们的模型。我们将我们的模型与四个基准模型进行了比较：（1）基于相似性的方法 Tiresias³⁹；（2）深度信念网络 (DBN)⁴⁰，这是一种使用来自预定义分子指纹和蛋白质描述符的中级特征的深度学习方法；（3）端到端神经网络模型 E2E⁵，该模型分别使用 GCN 和 LSTM 处理药物分子和蛋白质的高级信息（基因本体论注释）；（4）GNN⁴。

结果。图 2 展示了 BindingDB 数据集的实验结果。我们的方法在所有测试集和所有指标上均表现良好。当测试的蛋白质在训练数据中出现过（已见）时，DrugVQA 的 AUC 为 0.955 ± 0.005 ，准确率为 0.916 ± 0.005 。当测试的蛋白质未在训练数据中出现过（未见）时，AUC 和准确率分别为 0.922 ± 0.007 和 0.861 ± 0.009 。三个基线模型 (Tiresias、DBN 和 GNN) 在已见蛋白质上表现良好，但在未见蛋白质上的表现则差很多。这表明存在对训练所用蛋白质的过拟合现象。然而，E2E 对已见和未见蛋白质的表现一致，但其 AUC 和准确率平均分别比 DrugVQA 低 2.6% 和 2.3%。

在 DUD-E 数据集上的比较。对比模型。我们在 DUD-E 数据集上将我们的 DrugVQA 与最先进的药物-蛋白质相互作用 (DPI) 方法进行了比较，这些方法可分为三类：（1）传统的对接方法 Vina2⁴¹ 和 Smina1⁴²；（2）机器学习评分函数 NNScore27⁴³ 和 RFscore26⁴⁴；（3）基于深度学习的方法 3D-CNN3、AtomNet9、PocketGCN41 和 GNN4。

结果。如表 3 所示，DrugVQA 相较于基准模型实现了数量级的提升，达到了对药物研发有用的准确率水平。请注意，我们基于每个靶点分别计算结果，然后报告了在总共 102 个靶点上的平均结果（基于三折交叉验证）。实验重复了三次以获得标准差。在完整的 DUD-E 数据集上，DrugVQA 的平均 AUC 为 0.972，优于最先进的 GNN 模型（其平均 AUC 为 0.94）。尽管 3D-CNN 在训练时使用了 3D 结构，但在三种深度学习方法中其表现最差。这可能是由于 3D 空间中的数据较为稀疏，而 2D 成对距离图则提供了良好的平衡。因此，在 DUD-E 数据集的每个靶点上，DrugVQA 超过 3D-CNN 的比例达到了 96%。我们还在补充图 1 中展示了 DrugVQA 模型在 DUD-E 基准上的交叉验证性能以及 Vina 评分函数和 3D-CNN 的对比情况。

注意力可视化。我们模型的另一个优势在于其可解释性。为了说明这一点，我们在 DUD-E 数据集中选取了两个预测得分最高的相互作用。通过输入蛋白质距离图和化合物的 SMILES 表示，模型为蛋白质和化合物生成了多头注意力。我们将示例蛋白质中权重最高的 15 个残基用绿色标注，并将其与从蛋白质数据库 (PDB)⁴² 中检索到的蛋白质结构图和相互作用（由专家标注）进行比较。我们发现，权重最高的氨基酸（绿色）和化合物原子（红色）与真实的相互作用位点有很大程度的重叠。对于蛋白质 Hsp90（图 3a），注意力条形图突出显示了残基 Asn51A、Asp93A 和 Met98A，这些残基与共晶复合物 (PDB: 3EKR) 中观察到的关键口袋残基高度重叠。对于蛋白质 CDK2（图 3b），重要性图中突出显示的残基 (Phe80A、Asp145A、Leu134A) 和配体官能团与 2DUV 中观察到的相互作用显示出高度相似性。因此，我们的模型为结合因素提供了合理的线索，这可能具有广泛的生物医学应用价值。更多示例见补充图 3 - 5。

结论

在此，我们提出了一种新颖的端到端深度学习框架，类似于视觉问答 (VQA) 任务，用于预测蛋白质相互作用距离 (DPI)。在 DPI 研究中，我们使用了自注意力卷积和循环结构，同时从蛋白质二维距离图和分子语言中提取特征。实验评估表明，我们的模型在三个公共数据集上始终具有最佳性能。此外，该模型能够为理解分子相互作用的本质提供生物学见解。在 DPI 任务上的显著改进以及模块化系统表明，将 VQA 与生物问题相结合是一种新的策略，包括所有基于 3D 结构的预测，如蛋白质相互作用、蛋白质功能和蛋白质设计。

数据可用性

本文中使用的的数据均为公开数据，可从以下网址获取：DUD-E 数据集：<http://dude.docking.org>；BindingDB-IBM 数据集：<https://github.com/IBMInterpretableDTIP>；人类数据集：https://github.com/masashitsubaki/CPI_prediction/tree/master/dataset；蛋白质晶体结构：<https://www.rcsb.org>。

代码可用性

DrugVQA 的演示、说明和代码可在 <https://github.com/prokia/drugVQA> 获取。

收稿日期：2019 年 9 月 19 日；接受日期：2020 年 1 月 13 日；在线发表日期：2020 年 2 月 14 日

参考文献

1. Koes, D. R., Baumgartner, M. P. & Camacho, C. J. Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. *J. Chem. Inf. Model.* **53**, 1893–1904 (2013).
2. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).
3. Ragoza, M., Hochuli, J., Idrobo, E., Sunseri, J. & Koes, D. R. Protein–ligand scoring with convolutional neural networks. *J. Chem. Inf. Model.* **57**, 942–957 (2017).
4. Tsubaki, M., Tomii, K. & Sese, J. Compound–protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. *Bioinformatics* **35**, 309–318 (2018).
5. Gao, K. Y., Fokoue, A., Luo, H., Iyengar, A., Dey, S. & Zhang, P. Interpretable drug target prediction using deep neural representation. In *Int. Joint Conf. on Artificial Intelligence* 3371–3377 (IJCAI, 2018).
6. Zheng, S., Yan, X., Yang, Y. & Xu, J. Identifying structure–property relationships through SMILES syntax analysis with self-attention mechanism. *J. Chem. Inf. Model.* **59**, 914–923 (2018).
7. Öztürk, H., Özgür, A. & Ozkirimli, E. DeepDTA: deep drug–target binding affinity prediction. *Bioinformatics* **34**, i821–i829 (2018).
8. Jastrzebski, S., LeNiaik, D. & Czarnecki, W. M. Learning to SMILE(S). Preprint at <https://arxiv.org/abs/1602.06289> (2016).
9. Wallach, I., Dzamba, M. & Heifets, A. AtomNet: a deep convolutional neural network for bioactivity prediction in structure-based drug discovery. Preprint at <https://arxiv.org/abs/1510.02855> (2015).
10. Stepniewska-Dziubinska, M. M., Zielenkiewicz, P. & Siedlecki, P. Development and evaluation of a deep learning model for protein–ligand binding affinity prediction. *Bioinformatics* **34**, 3666–3674 (2018).
11. Skolnick, J., Kolinski, A. & Ortiz, A. R. MONSSTER: a method for folding globular proteins with a small number of distance restraints. *J. Mol. Biol.* **265**, 217–241 (1997).
12. Namrata, A. & Possu, H. Generative modeling for protein structures. *Adv. Neural Inf. Process. Syst.* **31**, 7494–7505 (2018).
13. Bepko, T. & Berger, B. Learning protein sequence embeddings using information from structure. Preprint at <https://arxiv.org/abs/1902.08661> (2019).
14. Yang, Z., He, X., Gao, J., Deng, L. & Smola, A. Stacked attention networks for image question answering. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 21–29 (2016).
15. Xu, K. et al. Show, attend and tell: neural image caption generation with visual attention. In *Int. Conf. on Machine Learning* **37**, 2048–2057 (PMLR, 2015).
16. Noh, H., Seo, P. H. & Han, B. Image question answering using convolutional neural network with dynamic parameter prediction. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 30–38 (IEEE, 2016).
17. Agrawal, A. et al. VQA: visual question answering. *Int. J. Comput. Vis.* **123**, 4–31 (2017).
18. Antol, S. et al. VQA: Visual Question Answering. In *Proc. IEEE International Conference on Computer Vision* 2425–2433 (IEEE, 2015).
19. Weininger, D. et al. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **28**, 31–36 (1988).
20. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).
21. Ma, L., Lu, Z. & Li, H. Learning to answer questions from image using convolutional neural network. In *Thirtieth AAAI Conference on Artificial Intelligence* (AAAI, 2016).
22. Shih, K. J., Singh, S. & Hoiem, D. Where to look: focus regions for visual question answering. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 4613–4621 (IEEE, 2016).
23. Xu, H. & Saenko, K. Ask, attend and answer: exploring question-guided spatial attention for visual question answering. In *European Conference on Computer Vision* (Springer, 2016).
24. Schwartz, I., Schwing, A. & Hazan, T. High-order attention models for visual question answering. *Adv. Neural Inf. Process. Syst.* 3664–3674 (2017).
25. Bleakley, K. & Yamanishi, Y. Supervised prediction of drug–target interactions using bipartite local models. *Bioinformatics* **25**, 2397–2403 (2009).
26. Ballester, P. J. & Mitchell, J. B. O. A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking. *Bioinformatics* **26**, 1169–1175 (2010).
27. Durrant, J. D. & McCammon, J. A. NNScore 2.0: a neural-network receptor–ligand scoring function. *J. Chem. Inf. Model.* **51**, 2897–2903 (2011).
28. Tabei, Y. & Yamanishi, Y. Scalable prediction of compound–protein interactions using minwise hashing. *BMC Syst. Biol.* **7**, S3 (2013).
29. Olivecrona, M., Blaschke, T., Engkvist, O. & Chen, H. Molecular de-novo design through deep reinforcement learning. *J. Cheminform.* **9**, 48 (2017).
30. He, K., Zhang, X., Ren, S. & Sun, J. Identity mappings in deep residual networks. In *European Conference on Computer Vision* 630–645 (Springer, 2016).
31. D.-A. Clevert, T. Unterthiner, and S. Hochreiter. Fast and accurate deep network learning by exponential linear units (ELUs). Preprint at <https://arxiv.org/abs/1511.07289> (2015).
32. Nair, V. & Hinton, G. E. Rectified linear units improve restricted Boltzmann machines. In *Proc. 27th International Conference on Machine Learning* 807–814 (ICML, 2010).
33. Lin, Z. et al. A structured self-attentive sentence embedding. Preprint at <https://arxiv.org/abs/1703.03130> (2017).
34. Mysinger, M. M., Carchia, M., Irwin, J. J. & Shoichet, B. K. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J. Med. Chem.* **55**, 6582–6594 (2012).
35. Liu, H., Sun, J., Guan, J., Zheng, J. & Zhou, S. Improving compound–protein interaction prediction by building up highly credible negative samples. *Bioinformatics* **31**, i221–i229 (2015).
36. Gilson, M. K. et al. BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res.* **44**, D1045–D1053 (2015).
37. Paszke, A. et al. Automatic differentiation in PyTorch. In *Neural Information Processing Systems Workshop Autodiff* (NeurIPS, 2017).
38. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. Preprint at <https://arxiv.org/abs/1412.6980> (2014).
39. Fokoue, A., Sadoghi, M., Hassanzadeh, O. & Zhang, P. Predicting drug–drug interactions through large-scale similarity-based link prediction. In *European Semantic Web Conference* 774–789 (Springer, 2016).
40. Wen, M. et al. Deep-learning-based drug–target interaction prediction. *J. Proteome Res.* **16**, 1401–1409 (2017).
41. Torng, W. & Altman, R. B. Graph convolutional neural networks for predicting drug–target interactions. *J. Chem. Inf. Model.* **59**, 4131–4149 (2019).
42. Burley, S. K. et al. RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Res.* **47**, D464–D474 (2018).

致谢

本研究部分得到了中国国家重点研发计划（2018YFC0910500）、广东省前沿与关键技术创新计划（2018B010109006、2019B020228001）、国家自然科学基金（61772566、U1611261、81801132、81903540）以及广东省创新创业团队引进计划（2016ZT06D211）的支持。

作者贡献

S.Z.、Y.L. 和 Y.Y. 负责概念设计和实施。S.Z. 和 Y.L. 共同设计实验。S.Z. 和 Y.L. 负责编程。所有作者都参与了结果的解读。S.Z. 和 Y.Y. 撰写了论文。所有作者都审阅并批准了最终的论文。

利益冲突

作者声明不存在利益冲突。

附加信息

有关本文的补充信息可在 <https://doi.org/10.1038/s42256-020-0152-y> 获取。

来信及索取资料请致函 J.X. 或 Y.Y.。有关重印及转载事宜，请访问 www.nature.com/reprints。出版者注：施普林格·自然对于所发表的地图中涉及的管辖权主张以及机构隶属关系保持中立。

© 作者（们），2020 年独家授权于施普林格·自然有限公司