

# 基于Wasserstein生成对抗网络和残差网络的8类蛋白质二级结构预测

李 舜, 马玉明\*, 刘毅慧\*

齐鲁工业大学(山东省科学院)计算机科学与技术学院, 山东 济南

收稿日期: 2023年1月30日; 录用日期: 2023年2月28日; 发布日期: 2023年3月7日

## 摘 要

蛋白质二级结构包含充分的蛋白质信息,而且蛋白质二级结构是研究蛋白质三级结构和药物设计的基础,因此,本文提出了一种基于Wasserstein生成对抗网络(WGAN)和残差网络(ResNet)的蛋白质8态二级结构预测的方法。该方法首先通过Wasserstein生成对抗网络(WGAN)提取蛋白质特征,将其与PSSM结合成新的特征集合,然后将新的特征集合输入到残差网络(ResNet)预测并得到最终结果。经过实验,该方法在测试集CASP10-14和CB513中的Q8预测准确率分别为73.21%, 72.43%, 71.67%, 69.83%, 70.17%和73.89%。通过实验表明, Wasserstein生成对抗网络(WGAN)具有出色的特征提取能力, ResNet能够有效地训练深层网络结构,从而提高蛋白质二级结构的预测精度。

## 关键词

Wasserstein生成对抗网络, 残差网络, 蛋白质二级结构预测

# Protein 8-State Secondary Structure Prediction Based on Wasserstein Generative Adversarial Network and Residual Network

Shun Li, Yuming Ma\*, Yihui Liu\*

College of Computer Science and Technology, Qilu University of Technology (Shandong Academy of Sciences), Jinan Shandong

Received: Jan. 30<sup>th</sup>, 2023; accepted: Feb. 28<sup>th</sup>, 2023; published: Mar. 7<sup>th</sup>, 2023

\*通讯作者。

文章引用: 李舜, 马玉明, 刘毅慧. 基于 Wasserstein 生成对抗网络和残差网络的 8 类蛋白质二级结构预测[J]. 计算生物学, 2023, 13(1): 1-9. DOI: 10.12677/hjcb.2023.131001

## Abstract

Protein secondary structure is the basis for studying protein tertiary structure and drug design, because the 8-state protein secondary structure can provide sufficient protein information for this. Therefore, this paper proposes a method for predicting the 8-state secondary structure of proteins based on Wasserstein generative adversarial network (WGAN) and residual network (ResNet). This method first extracts protein features by Wasserstein generative adversarial network (WGAN), combines them with PSSM to form a new feature set, and then inputs the new feature set to the residual network (ResNet) prediction and obtains the final result. After experiments, the Q8 prediction accuracy of this method in the test set CASP10-14 and CB513 was 73.21%, 72.43%, 71.67%, 69.83%, 70.17% and 73.89%, respectively. Experiments show that the Wasserstein generative adversarial network (WGAN) has excellent feature extraction ability, and ResNet can effectively train the deep network structure, thereby improving the prediction accuracy of protein secondary structure.

## Keywords

Wasserstein Generative Adversarial Network, Residual Network, Protein Secondary Structure Prediction

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

蛋白质在我们的生命活动中发挥着极其重要的作用，具有免疫和细胞信号传导等功能。其不同的功能取决于其不同的结构。因此，为了研究蛋白质的功能并开展相关应用，需要对蛋白质的结构进行预测。蛋白质二级结构是蛋白质多肽链中氨基酸残基的局部空间构象，分别以 8 态的形式出现，即  $\alpha$ -螺旋(H)、310-螺旋(G)、 $\pi$ -螺旋(I)、折叠(E)、分离  $\beta$  转角(B)、转角(T)、弯曲(S)和其他(C) [1] [2] [3]。

1990 年代，Burkhard Rost 和 Chris Sander 首次使用神经网络预测蛋白质的二级结构[4]。该方法在取得优异结果的同时，也在蛋白质结构预测领域具有开创性。随着“人类基因组计划(Human Genome Project, HGP) [5]”的展开，标志着后基因组时代的到来，蛋白质数据的数量也呈爆炸式增长。由于成本高、难度大等缺点，传统的实验测定方法已无法满足日益增长的蛋白质和结构数据分析需求。因此，蛋白质结构预测逐渐成为生物信息学研究的热点。随着 AlphaFold2 [6]的出现改变了蛋白质结构预测的格局，它对于蛋白质三级结构的预测已经取得了非常可靠的结果，尽管如此，二级结构的预测仍然具有重要意义，这是因为三级结构是通过多个二级结构元素在三维空间上的排列所形成的一个蛋白质分子三维结构，了解蛋白质的二级结构将会影响和改善蛋白质三级结构的对齐，从而影响和改变蛋白质的空间形态，这对分析和探索蛋白质的功能和作用有重要的作用，通过结构分析可以进行功能注释，能够更好地知道生物实验，也可以为设计和改造已有的蛋白质提供可靠的依据。

近年来，许多学者和研究人员在蛋白质二级结构的 8 态研究领域取得了优异的成绩。文献[7]提出了一种蛋白质序列预测技术，将过去使用卷积神经网络的成功经验与语言建模相结合，取得了很好的效果。

文献[8]设计了一个由卷积滤波器和原始输入组成的局部块来捕获局部序列特征,该方法利用卷积神经网络、残差网络和双向循环神经网络预测的组合协同作用。文献[9]使用最近邻搜索确定蛋白质中残基的估计类成员概率,然后将其输入另一个动态规划算法,在 CASP 数据集上取得了良好的结果。文献[10]提出将 self-attention 机制与 Deep Inception-Inside-Inception (Deep3I)网络相结合,通过 Interaction 跟踪不同距离氨基酸之间的残基。文献[11]设计一个由 1D-Convnet 和改进的循环神经网络组成的模型,该网络具有改进的顺序抛硬币优化器,在 CB513 和 CullPDB 上实现了良好的预测精度。

## 2. 实验模型与实验数据

本文提出了一种基于 Wasserstein 生成对抗网络(WGAN)和残差网络(ResNet)的 8 类蛋白质二级结构预测方法,具体过程是首先通过 WGAN 提取蛋白质的特征信息,然后将这些信息与原始特征信息位置特异性评分矩阵(PSSM) [12]相结合,位置特异性评分矩阵(PSSM)是一种广泛使用的特征信息。PSSM 是通过 NR 数据库中蛋白质的多序列比对,将 PSI-BLAST [13]参数阈值设置为 0.001 和 3 次迭代生成的,其形式为  $20 * L$  矩阵,20 为特征维度,L 代表不同蛋白质的长度。组合后的特征集合形式为  $40 * L$ ,它包含更丰富的蛋白质信息起到特征增强的作用。组合后的特征集合将作为 ResNet 的输入数据,经过深层次网络的训练和学习,最终得到预测结果。

### 2.1. Wasserstein 生成对抗网络

生成对抗网络(GAN)由 Ian Goodfellow 在 2014 年提出[14],已被广泛应用于图像去噪和特征提取[15] [16] [17],并被证明具有良好的性能。GAN 由两部分组成,生成器(G)和鉴别器(D)。G 通过学习真实数据的分布特征来生成相似的模拟数据,而 D 则对模拟数据的真实性进行判断和评分。但是,由于 GAN 的近似最优 D, G 损失面临梯度消失的问题,以及优化目标不明确,梯度不稳定。WGAN 使用 Wasserstein 距离可以改善该问题,其优点是可以反映两个分布之间的距离,即使它们没有任何重叠。WGAN 对 GAN 的目标函数进行改进和优化。GAN 的目标函数如(1)所示。

$$\min_G \max_D (D, G) = E_{x \sim P_{data}(x)} [\log D(x)] + E_{z \sim P_z(z)} [\log(1 - D(G(z)))] \quad (1)$$

在这里,  $x$  和  $z$  分别代表真实数据和随机数据,  $G(z)$  是 G 生成的模拟数据,  $D(x)$  代表数据是否真实的概率。WGAN 将 WGAN 要优化的目标函数(1)分成两个部分:第一部分为保持 G 固定,此时优化 D,则公式(1)改写为公式(2),将公式(1)转化为最小化的形式,可以改写为公式(3);第二部分为当固定 D,优化 G 时,则公式(1)可以改写为最小化,即公式(4)的形式,要使 D 的参数不超过一个固定的常数,则只需要最大化公式(5)。

$$\max_D E_{x \sim P_r} [\log D(x)] + E_{x \sim P_g} [\log(1 - D(x))] \quad (2)$$

$$\min_D - E_{x \sim P_r} [\log D(x)] - E_{x \sim P_g} [\log(1 - D(x))] \quad (3)$$

$$\min_G E_{x \sim P_g} [\log(1 - D(x))] \quad (4)$$

$$L = E_{x \sim P_r} [D(x)] - E_{x \sim P_g} [\log(D(x))] \quad (5)$$

在该方法中,我们首先输入随机数据和 PSSM,经过 G 的不断地学习会不断地模拟 PSSM 并生成一组与 PSSM “相似”的虚拟数据,D 网络将对虚拟网络进行评分,当判定为假时,会反馈给 G 网络,G 网络将继续模拟,直到 D 网络无法判定真伪,此时,G 网络生成的数据就是实验所需要的。WGAN 的工作过程如图 1 所示。

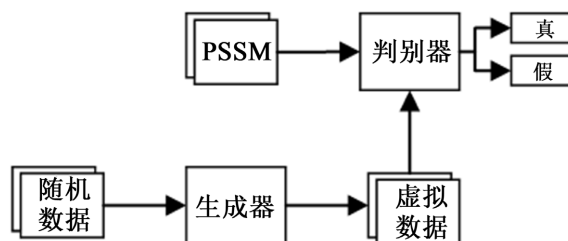


Figure 1. WGAN's workflow  
图 1. WGAN 的工作流程

## 2.2. 卷积神经网络

卷积神经网络(CNN)近年来已广泛应用于图像处理[18]、计算机视觉等领域。在本实验中,我们在WGAN中引入了CNN来协助特征提取。CNN的传统结构由3部分组成,输入层、特征提取层和一个全连接的多层感知器。卷积神经网络中的局部感受野和权重共享操作可以实现位移、缩放和失真不变性。局部感知场就是让每一层中的每个神经元感知上一层输入数据的局部部分,从而减少参数的数量。权重共享在卷积运算时提取特征时做同样的约束,可以减少自由参数的数量,降低网络复杂度,提高学习效率。卷积层卷积运算公式如(5)所示。

$$F_k^i = f\left(\sum_h P_h^{i-1} * W_k^i + b\right) \quad (5)$$

在公式(5)中,  $f$  为激活函数,  $P_h^{i-1}$  表示输入数据与上一层的卷积核得到的特征图,  $W_k^i$  是第  $i$  层的一个卷积核,  $k$  表示卷积核的数量,  $b$  代表偏置参数。

## 2.3. 残差网络

网络模型的深度会对实验结果产生很大的影响,理论上,网络深度越深,训练将会更充分,但事实上深层次的卷积网络模型会产生梯度爆炸和梯度消失的问题,而ResNet [19]引入了残差学习框架来缓解该问题,具体过程是,对于一个堆积层结构,当输入为  $x$  时起学习到的特征记为  $H(x)$ ,我们希望其可以学到的残差  $F(x) = H(x) - x$ ,此时原始特征是  $F(x) + x$ 。之所以这样是因为残差学习相比原始特征直接学习更容易,当残差为 0 时,此时堆积层仅仅做了恒等映射,这使得网络性能不会下降,而实际上残差不会为 0,这也会使得堆积层在输入特征的基础上学习到新的特征,从而使拥有更好的性能。残差学习类似于短路连接,其结构如图 2 所示。

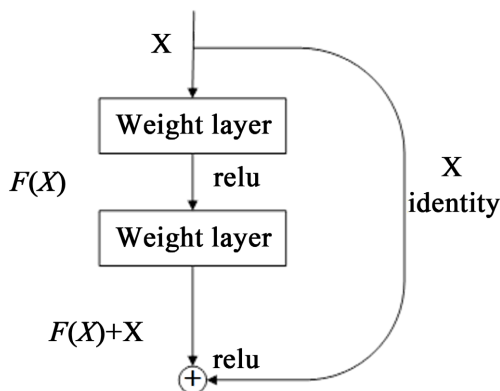


Figure 2. The structure of residual learning  
图 2. 残差学习的结构图

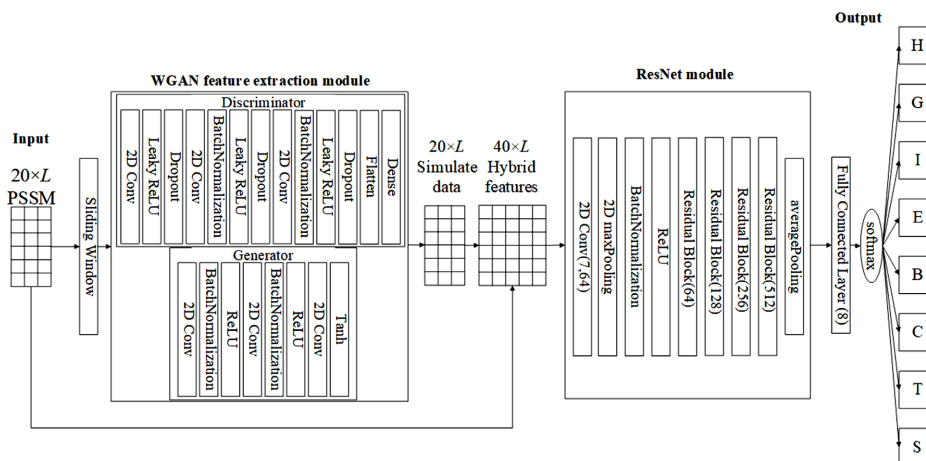
值得注意的是在 ResNet 中, 当特征图的大小降低一半时, 其数量要增加一倍, 以保持网络层复杂度不变。在该实验中使用 ResNet-N 来预测蛋白质的 8 态二级结构。N 是指具有特权值的网络层数, 包括卷积层、全连接层和池化层。我们将权重为 10、18 和 34 的层数设置为实验模型, ResNet-10、ResNet-18 和 ResNet-34 的结构分别如表 1 所示。

**Table 1.** System resulting data of standard experiment  
**表 1.** 标准试验系统结果数据

	ResNet-10	ResNet-18	ResNet-34
Conv_first		$7 \times 7, 64$ , stride 2	
Max_pooling		$7 \times 7$ , stride 2	
Block-1	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$
Block-2	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$
Block-3	/	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$
Block-4	/	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$
Average pool, fully connected, softmax			

## 2.4. 模型的整体结构

我们提出的整体结构如图 3 所示, 可以看出我们的网络模型主要分为 WGAN 和 ResNet 两个模块, 其中 WGAN 的作用是特征增强, 我们使用浅层 CNN 辅助 G 和 D 网络, G 网络使用泄漏 ReLU 作为激活函数, 由于 G 网络的迭代次数很多, 为了防止过度拟合, 我们也使用 Dropout。ResNet 用于局部特征提取和预测工作, 由多个残差学习单元组成, 目的是使用更深层次的卷积网络进行深度训练, 以获得更好的预测结果。



**Figure 3.** Network structure based on WGAN and ResNet  
**图 3.** 基于 WGAN 和 ResNet 方法的网络结构

## 2.5. 实验数据集

本实验中使用的主要公共数据集是 CullPDB [20]数据集和 CASP [21] [22] [23] [24] [25]数据集 CASP10、CASP11、CASP12、CASP13、CASP14 和 CB513 [26]数据集，它们包含的数据的相似性小于 25%。本文去除重复的蛋白质数据集 CullPDB 作为训练集，共有 11,650 个蛋白质。数据集中的蛋白质序列数量如表 2 所示。

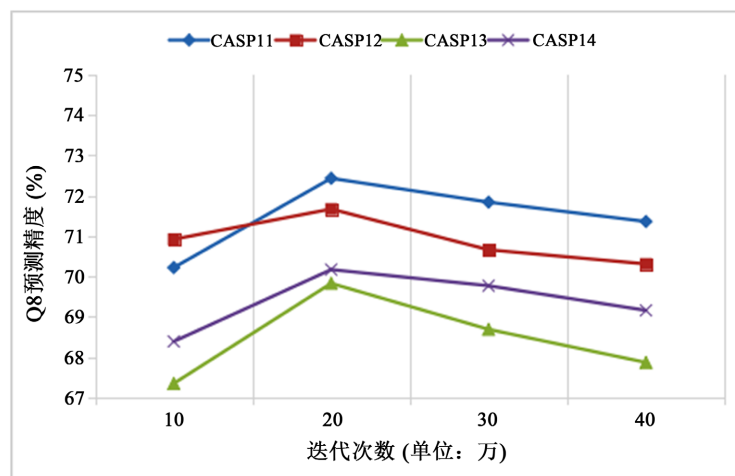
**Table 2.** The number of proteins in the dataset

**表 2.** 数据集中蛋白质的数量

数据集	蛋白质数量
CullPDB	11,650
CASP10	99
CASP11	81
CASP12	19
CASP13	22
CASP14	24
CB513	513

## 3. 实验结果与参数分析

我们在实验中调整了相关参数以测试其对网络模型的影响，首先我们调整了 WGAN 中生成器和判别器的迭代次数，并且在测试集 CASP11-14 上测试了其对于 WGAN 提取特征的影响，其中 G 网络中卷积层数设置为 3，卷积核参数分别为  $3 \times 3 \times 64$ ， $3 \times 3 \times 128$  和  $3 \times 3 \times 256$ ，不同迭代下的实验结果如图 4 所示。



**Figure 4.** Q8 accuracy at different iterations

**图 4.** 不同迭代次数下的 Q8 精度

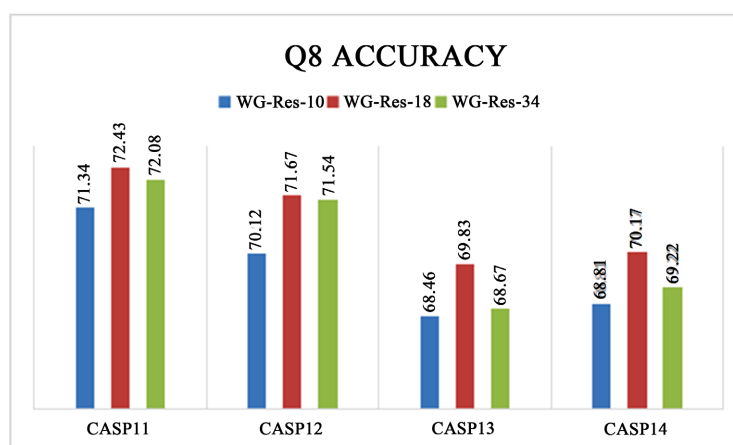
如图所示，当迭代次数为 20 万次时，效果最好，即 G 提取的特征最真实有效。超过 20 多万次迭代以后，D 判断所生成数据真实性的能力下降，模拟特征与真实特征误差较大。此外，为了测试滑动窗口长度对实验结果的影响，我们选择了 13、15、17、19 和 21 进行 Q8 预测。实验结果如表 3 所示，表明当滑动窗口为 19 时，实验结果最好。



**Table 3.** Q8 accuracy at different sliding windows  
**表 3.** 不同滑动窗口下的 Q8 精度

滑动窗口长度	CASP11	CASP12	CASP13	CASP14
13	73.47	68.16	65.88	65.20
15	68.09	69.27	66.67	67.41
17	71.66	70.24	68.18	69.13
19	72.43	71.56	69.83	70.17
21	70.29	70.41	68.42	68.94

使用不同网络层数的 ResNet, 我们对 CASP11-14 进行了测试, 得到了如图 5 所示的实验结果, 可以看出 WG-Res-18 的精度最高, 因为 WG 数据的维度不高, 当层数太深时, 部分数据会丢失, 从而导致精度下降。



**Figure 5.** Q8 accuracy at different residual network depth models  
**图 5.** 不同残差网络深度模型下的 Q8 精度

我们做了消融实验来证明每个结构的重要性。我们用四种网络模型对 CASP11-14 进行了测试, 实验结果如表 4 所示。其中 WGAN + ResNet 是本文提出的模型, WGAN + CNN 是 WGAN 和 CNN 相结合的网络模型, 输入数据采用特征增强后的数据, CNN 的网络模型结构使用 3 个卷积层, 尺度为  $3 \times 3 \times 64$ 、 $3 \times 3 \times 128$  和  $3 \times 3 \times 256$ 。ResNet 是基于最佳 ResNet-18 的残差网络模型, CNN 使用的是 3 层卷积神经网络, 结构为  $3 \times 3 \times 64$ 、 $3 \times 3 \times 128$  和  $3 \times 3 \times 256$ 。ResNet 和 CNN 的输入数据是 PSSM。通过对比表中的实验结果可以看出, 与 CNN 相比, ResNet 深度训练提高了预测精度; WGAN 对 PSSM 的原始数据进行了特征提取和特征增强, 也提高了蛋白质二级结构预测的准确性。

**Table 4.** Q8 accuracy at different methods  
**表 4.** 不同模型下的 8 态预测精度

方法	CASP11	CASP12	CASP13	CASP14
WGAN + ResNet	73.43	73.67	71.83	72.17
WGAN + CNN	71.47	71.79	68.33	68.24
ResNet	69.76	68.84	66.57	67.19
CNN	66.62	66.29	64.69	66.71

## 4. 结论和未来工作

蛋白质二级结构的预测是全面认识和探索蛋白质多样化功能和空间结构的重要工作。本文首次将 Wasserstein 生成对抗网络和残差网络相结合, 用于预测蛋白质的 8 态二级结构。首先利用 WGAN 依照 PSSM 提取蛋白质序列特征, 其中 G 和 D 模型都使用卷积神经网络辅助, 然后将 PSSM 与提取的特征组合成一个新的特征矩阵, 该矩阵将包含更丰富的蛋白质特征信息, 最后将该矩阵输入不同层次的残差网络训练, 残差网络可以训练更深层次的网络结构, 以获得更好的预测结果。经过我们的实验证明, Wasserstein 生成对抗网络具有较强的特征提取能力, 残差网络也可以更好地训练数据, 该方法下的蛋白质 8 态二级结构预测精度有更显著的结果。在未来的工作中, 我们将继续优化残差网络, 该实验使用的残差网络模型虽然在一定程度上使得深层次的网络模型得到了充分训练, 但是其深度同时也预示着残差网络的训练参数巨大, 训练时间也会相应的增加, 这给实验环境带来了一定的压力。多尺度卷积理论上是通过不同大小的卷积核经过卷积运算得到不同大小的特征图, 在不增加网络深度的基础上拓宽了视野宽度, 也起到了特征增强的作用, 在未来我们期望将多尺度卷积神经网络融合到我们的模型中, 使得训练模型可以从深度和宽度两个方向着力, 在加深网络层数的基础上尽可能地减少训练参数。

## 基金项目

国家自然科学基金(No. 61375013), 山东省自然科学基金(No. ZR2013FM020)。

## 参考文献

- [1] Zhou, J. and Troyanskaya, O. (2014) Deep Supervised and Convolutional Generative Stochastic Network for Protein Secondary Structure Prediction. *Proceedings of the 31st International Conference on Machine Learning*, Beijing, 22-24 June 2014, 745-753.
- [2] Yaseen, A. and Li, Y. (2014) Template-Based C8-Scorpion: A Protein 8-State Secondary Structure Prediction Method Using Structural Information and Context-Based Features. *BMC Bioinformatics*, **15**, Article No. S3. <https://doi.org/10.1186/1471-2105-15-S8-S3>
- [3] Kabsch, W. and Sander, C. (1983) Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers*, **22**, 2577-2637. <https://doi.org/10.1002/bip.360221211>
- [4] Senior, A.W., Evans, R., Jumper, J., et al. (2020) Improved Protein Structure Prediction Using Potentials from Deep Learning. *Nature*, **577**, 706-710. <https://doi.org/10.1038/s41586-019-1923-7>
- [5] 骆建新, 郑崛村, 马用信, 张思仲. 人类基因组计划与后基因组时代[J]. 中国生物工程杂志, 2003, 23(11): 87-94.
- [6] Jumper, J., Evans, R., Pritzel, A., et al. (2021) Highly Accurate Protein Structure Prediction with AlphaFold. *Nature*, **596**, 583-589. <https://doi.org/10.1038/s41586-021-03819-2>
- [7] Busia, A. and Jaitly, N. (2017) Next-Step Conditioned Deep Convolutional Neural Networks Improve Protein Secondary Structure Prediction. ArXiv Preprint ArXiv: 1702.03865.
- [8] Zhang, B., Li, J. and Lü, Q. (2018) Prediction of 8-State Protein Secondary Structures by a Novel Deep Learning Architecture. *BMC Bioinformatics*, **19**, Article No. 293. <https://doi.org/10.1186/s12859-018-2280-5>
- [9] Krieger, S. and Kececioglu, J. (2020) Boosting the Accuracy of Protein Secondary Structure Prediction through Nearest Neighbor Search and Method Hybridization. *Bioinformatics*, **36**, i317-i325. <https://doi.org/10.1093/bioinformatics/btaa336>
- [10] Uddin, M.R., Mahbub, S., Rahman, M.S. and Bayzid, M.S. (2020) SAINT: Self-Attention Augmented Inception-Inside-Inception Network Improves Protein Secondary Structure Prediction. *Bioinformatics*, **36**, 4599-4608. <https://doi.org/10.1093/bioinformatics/btaa531>
- [11] Sonsare, P.M. and Gunavathi, C. (2021) Cascading 1D-Convnet Bidirectional Long Short Term Memory Network with Modified COCOB Optimizer: A Novel Approach for Protein Secondary Structure Prediction. *Chaos, Solitons & Fractals*, **153**, Article ID: 111446. <https://doi.org/10.1016/j.chaos.2021.111446>
- [12] Zvelebil, M. and Baum, J.O. (2007) Understanding Bioinformatics. Garland Science, London. <https://doi.org/10.1201/9780203852507>
- [13] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped



- BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acids Research*, **25**, 3389-3402. <https://doi.org/10.1093/nar/25.17.3389>
- [14] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y. (2020) Generative Adversarial Nets. *Communications of the ACM*, **63**, 139-144. <https://doi.org/10.1145/3422622>
- [15] Wang, R., Xiao, X., Guo, B., Qin, Q. and Chen, R. (2018) An Effective Image Denoising Method for UAV Images via Improved Generative Adversarial Networks. *Sensors*, **18**, Article No. 1985. <https://doi.org/10.3390/s18071985>
- [16] Yu, S., Chen, H., Reyes, E.B.G. and Poh, N. (2017) GaitGAN: Invariant Gait Feature Extraction Using Generative Adversarial Networks. 2017 *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, 21-26 July 2017, 532-539. <https://doi.org/10.1109/CVPRW.2017.80>
- [17] 赵亚武, 张华兰, 刘毅慧. 基于生成对抗和卷积神经网络的蛋白质二级结构预测[J]. 计算生物学, 2020, 10(4): 49-56.
- [18] Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2017) ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*, **60**, 84-90. <https://doi.org/10.1145/3065386>
- [19] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- [20] Wang, G. and Dunbrack, R.L. (2005) PISCES: Recent Improvements to a PDB Sequence Culling Server. *Nucleic Acids Research*, **33**, W94-W98. <https://doi.org/10.1093/nar/gki402>
- [21] Moulton, J., Fidelis, K., Kryshtafovych, A., Schwede, T. and Tramontano, A. (2014) Critical Assessment of Methods of Protein Structure Prediction (CASP)—Round X. *Proteins: Structure, Function, and Bioinformatics*, **82**, 1-6. <https://doi.org/10.1002/prot.24452>
- [22] Moulton, J., Fidelis, K., Kryshtafovych, A., Schwede, T. and Tramontano, A. (2016) Critical Assessment of Methods of Protein Structure Prediction: Progress and New Directions in Round XI. *Proteins: Structure, Function, and Bioinformatics*, **84**, 4-14. <https://doi.org/10.1002/prot.25064>
- [23] Moulton, J., Fidelis, K., Kryshtafovych, A., Schwede, T. and Tramontano, A. (2018) Critical Assessment of Methods of Protein Structure Prediction (CASP)—Round XII. *Proteins: Structure, Function, and Bioinformatics*, **86**, 7-15. <https://doi.org/10.1002/prot.25415>
- [24] Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K. and Moulton, J. (2019) Critical Assessment of Methods of Protein Structure Prediction (CASP)—Round XIII. *Proteins: Structure, Function, and Bioinformatics*, **87**, 1011-1020. <https://doi.org/10.1002/prot.25823>
- [25] Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K. and Moulton, J. (2021) Critical Assessment of Methods of Protein Structure Prediction (CASP)—Round XIV. *Proteins: Structure, Function, and Bioinformatics*, **89**, 1607-1617. <https://doi.org/10.1002/prot.26237>
- [26] Cuff, J.A. and Barton, G.J. (1999) Evaluation and Improvement of Multiple Sequence Methods for Protein Secondary Structure Prediction. *Proteins: Structure, Function, and Bioinformatics*, **34**, 508-519. [https://doi.org/10.1002/\(SICI\)1097-0134\(19990301\)34:4<508::AID-PROT10>3.0.CO;2-4](https://doi.org/10.1002/(SICI)1097-0134(19990301)34:4<508::AID-PROT10>3.0.CO;2-4)