

基于Swin Transformer的皮肤病诊断研究

申雨婧¹, 张仲荣^{1*}, 宋平安²

¹兰州交通大学数理学院, 甘肃 兰州

²兰州石化总医院胸外科, 甘肃 兰州

收稿日期: 2025年1月13日; 录用日期: 2025年2月7日; 发布日期: 2025年2月14日

摘要

皮肤病是世界范围内最常见的疾病之一。准确、及时、有效的皮肤病图像分类研究对皮肤病诊断具有重要意义, 高精度分类算法研究是该领域的热点和难点。近年来, 深度学习算法在皮肤病诊断领域展现出了巨大的发展潜力, 具有广阔的应用前景。文章采用基于Transformer模型改进的Swin Transformer模型, 构建了基于Swin Transformer的皮肤病诊断模型, 并将该模型在HAM10000数据集上进行实验验证。研究证明, 该模型显著提高了皮肤病诊断的准确性, 有望推动皮肤病诊断领域朝着更加高效、精准的方向迈进, 为皮肤病诊断领域的临床实践提供了更有力的实验支撑。

关键词

Swin Transformer, 自注意力机制, 皮肤病诊断

Research on Dermatological Disease Diagnosis Based on Swin Transformer

Yujing Shen¹, Zhongrong Zhang^{1*}, Ping'an Song²

¹School of Mathematics and Physics, Lanzhou Jiaotong University, Lanzhou Gansu

²Department of Thoracic Surgery, Lanzhou Petrochemical General Hospital, Lanzhou Gansu

Received: Jan. 13th, 2025; accepted: Feb. 7th, 2025; published: Feb. 14th, 2025

Abstract

Dermatological diseases are among the most common diseases worldwide. Accurate, timely, and effective classification of dermatological disease images is of great significance for the diagnosis of dermatological diseases, and the research on high-precision classification algorithms is a hot and difficult issue in this field. In recent years, deep learning algorithms have shown great potential for

*通讯作者。

development in the field of dermatological disease diagnosis, with broad application prospects. This paper adopts the Swin Transformer model, an improvement based on the Transformer model, to construct a dermatological disease diagnosis model based on the Swin Transformer. The model is experimentally verified on the HAM10000 dataset. The study proves that the model significantly improves the accuracy of dermatological disease diagnosis and is expected to promote the field of dermatological disease diagnosis in a more efficient and precise direction, providing stronger experimental support for clinical practice in this field.

Keywords

Swin Transformer, Self-Attention Mechanism, Dermatological Diagnosis

Copyright © 2025 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

皮肤最重要的功能是在生物体与环境之间形成有效的屏障,能够防止病原体的入侵,抵御化学和物理攻击,以及水和溶质的无规律流失[1]。皮肤健康对人类的身体健康与心理健康起着至关重要的作用。调查研究显示,全球每年受到皮肤和皮下疾病困扰的人不计其数,目前皮肤病是世界范围内最常见的疾病之一,常见的包括基底细胞癌等[2]。由于皮肤病具有症状多样性和隐匿性的特征,因此依靠皮肤科医生的病史问诊、肉眼观察及临床经验等传统的方式诊断皮肤病,存在着主观性和可变性的问题。随着医疗科技的飞速发展,人们在传统皮肤病诊断方式的基础上,持续探索机器学习、深度学习等技术,为准确、及时、有效的诊断皮肤病提供了有力的支撑。

深度学习(Deep Learning, DL)算法源于人脑处理信息的启发,是从大数据中学习的多层人工神经网络[3]。目前,深度学习已广泛应用于图像识别与处理任务中,在重点领域取得了显著突破,在许多医学重点应用中也展现出巨大的潜力,目前已被广泛应用于放射学[4]、眼科学[5]、皮肤癌[6]等病变诊断中。

综上,基于深度学习的皮肤病诊断是一个具有广阔前景和重大应用潜力的研究领域。因此,持续探索深度学习在皮肤病诊断中的研究与应用具有非常重要的学术价值,对于提升人们的身心健康水平具有重要的现实意义。

Transformer [7]是深度学习领域里一种革命性的模型架构。Transformer 架构自提出以来,已经在多个领域展现出了其强大的性能,尤其是在处理序列数据和图像数据方面。但在处理高分辨率图像时仍具有计算复杂度偏大、缺乏对图像特征信息的有效利用等局限性。因此,本文采用基于 Transformer 模型改进的 Swin Transformer 模型[8],克服了 Transformer 架构中的自注意力机制在处理高分辨率图像时计算复杂度偏大、缺乏对局部结构和位置信息的有效利用等局限性,构建了基于 Swin Transformer 的皮肤病诊断模型, Swin Transformer 模型通过引入窗口多头自注意力机制、移动窗口机制,能够有效降低计算量,提升了模型性能。因此,本文将该模型应用于 HAM10000 皮肤病数据集,推广皮肤病诊断领域的应用。研究证明,该模型提升了皮肤病诊断的准确性,为深度学习技术在医学图像分类方面提供了有效的实验依据。

2. 基于 Swin Transformer 的皮肤病诊断模型

Transformer [7]是一种基于自注意力机制的网络模型,整体架构如图 1 所示。Transformer 模型虽然

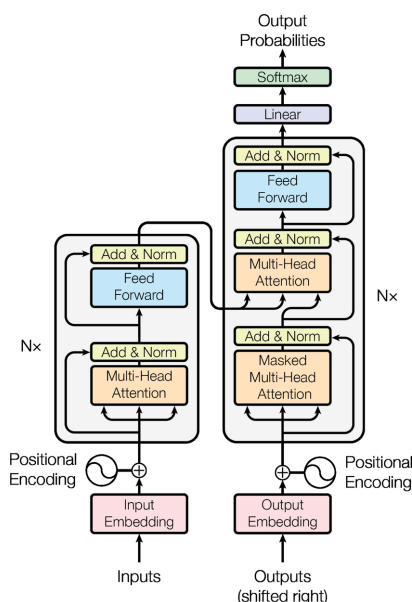


Figure 1. The architecture diagram of Transformer [7]

图 1. Transformer 架构图[7]

已经在较多领域得到了应用,但是 Transformer 架构中的自注意力机制在处理高分辨率图像等长序列数据时,计算复杂度会随着序列长度的增加呈二次方增长,导致计算量偏大,且 Transformer 在一定程度上缺乏对细节特征信息的有效提取,会忽略图像数据中的一些局部特征。针对 Transformer 模型在图像数据处理任务中的局限性,本文采用了基于 Transformer 模型改进的 Swin Transformer 模型,引入了窗口多头自注意力机制、移动窗口机制,在提升模型关注图像细节特征能力的同时降低计算量,并将该模型应用于皮肤病诊断领域。基于 Swin Transformer 的皮肤病诊断模型架构如图 2 所示。

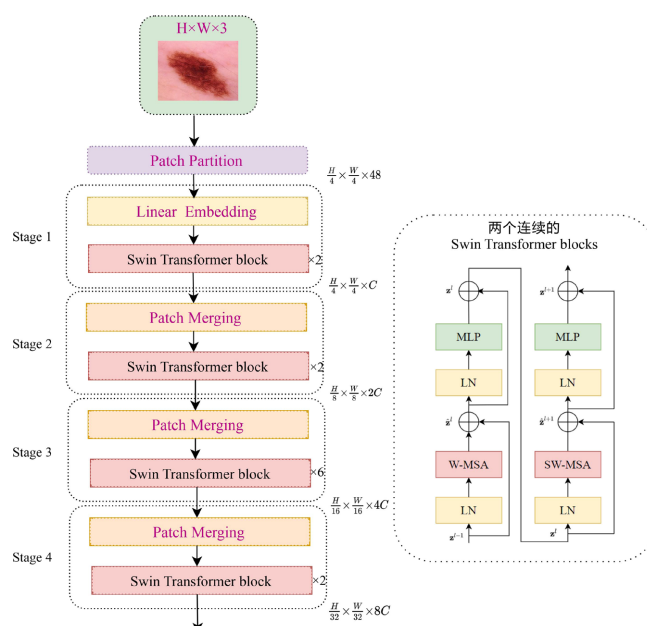


Figure 2. The architecture diagram of the Swin Transformer

图 2. Swin Transformer 架构图

Swin Transformer 模型首先将每张输入图像分割成固定大小的非重叠小块, 然后在这些小块上执行窗口多头自注意力机制, 同时, 通过移动窗口操作, 实现窗口之间的特征连接, 从而获取局部和全局的特征关系。不仅提升了模型对局部特征的关注度, 还大大降低了计算复杂度。Swin Transformer 模型的核心改进在于将 Transformer 中的标准多头自注意力替换为基于窗口的多头自注意力, 并在连续的两个 Swin Transformer block 之间进行了移动窗口操作。

其中, 连续的 Swin Transformer block 计算过程如下:

$$\begin{aligned}\hat{\mathbf{z}}^l &= W - \text{MSA}\left(\text{LN}\left(\mathbf{z}^{l-1}\right)\right) + \mathbf{z}^{l-1}, \\ \mathbf{z}^l &= \text{MLP}\left(\text{LN}\left(\hat{\mathbf{z}}^l\right)\right) + \hat{\mathbf{z}}^l, \\ \hat{\mathbf{z}}^{l+1} &= SW - \text{MSA}\left(\text{LN}\left(\mathbf{z}^l\right)\right) + \mathbf{z}^l, \\ \mathbf{z}^{l+1} &= \text{MLP}\left(\text{LN}\left(\hat{\mathbf{z}}^{l+1}\right)\right) + \hat{\mathbf{z}}^{l+1},\end{aligned}$$

其中, $\hat{\mathbf{z}}^l$ 和 $\mathbf{z}^l$ 为 block l 的(移动)窗口多头自注意力机制及 MLP 层的输出特征。

此外, 假设每个窗口包含 $M \times M$ 个小块, 对于一幅具有 $h \times w$ 个小块的皮肤图像来说, 全局 MSA 和基于窗口的 W-MSA 计算复杂度分别为:

$$\begin{aligned}\Omega(\text{MSA}) &= 4hwC^2 + 2(hw)^2C, \\ \Omega(W - \text{MSA}) &= 4hwC^2 + 2M^2hwC,\end{aligned}$$

与全局 MSA 相比, W-MSA 的计算复杂度显著降低, 尤其是在高分辨率图像上, 能够有效减少计算成本。

3. 实验及结果分析

本文采用的数据集为 HAM10000 [9], 该数据集为尺寸为 450×600 , 包含了 10015 张来自不同人群的皮肤镜图像集合, 共包括七种类别: Actinic Keratosis/Intraepithelial Carcinoma (AKIEC), Basal Cell Carcinoma (BCC), Benign Keratosis (BKL), Dermatofibroma (DF), Melanoma (MEL), Melanocytic Nevi (NV) 和 Vascular Lesions (VASC)。这七种皮肤病类别可视化如图 3 所示。

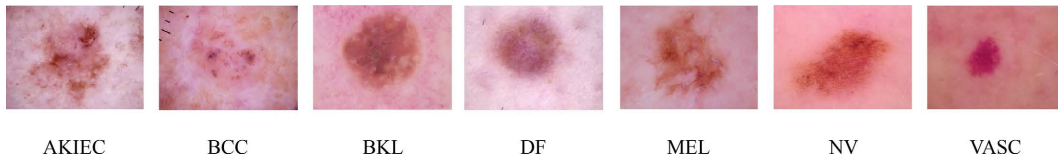


Figure 3. Examples of dermatological disease categories in the HAM10000 dataset

图 3. HAM10000 数据集中的皮肤病类别示例

本文采用分类准确率、分类精确率指标对模型性能进行评估。计算公式分别如下:

$$\begin{aligned}\text{Accuracy} &= \frac{TP + TN}{T} \\ \text{Precision} &= \frac{TP}{TP + FP}\end{aligned}$$

其中, TP 、 TN 、 FP 、 T 分别为真正例、真负例、假正例和样本总数。

将基于 Swin Transformer 的皮肤病诊断模型与 GoogLeNet、Inception V3 模型在 HAM10000 数据集

上的总体准确率进行对比，如表 1 所示，GoogLeNet、Inception V3 在该数据集上的总体精确率分别达到了 83.94%、86.82%，而本实验模型分类准确率达到 90.20%。由此证明，在分类性能评估指标上，本实验模型表现较为出色，能够准确地识别皮肤病。

Table 1. Comparison of accuracy rates between the dermatological disease diagnosis model based on Swin Transformer and other models on the HAM10000 dataset

表 1. 基于 Swin Transformer 的皮肤病诊断模型与其他模型在 HAM10000 数据集上的准确率对比

方法	准确率(%)
GoogLeNet [10]	83.94
Inception V3 [10]	86.82
Swin Transformer	90.20

除了验证总体准确率的表现，本实验还列出了基于 Swin Transformer 的皮肤病诊断模型在 HAM10000 数据集上对于七种不同的皮肤病类别的识别精确率，如表 2 所示。

Table 2. Classification precision of the dermatological disease diagnosis model based on Swin Transformer for seven dermatological disease categories on the HAM10000 dataset

表 2. 基于 Swin Transformer 的皮肤病诊断模型在 HAM10000 数据集上针对七种皮肤病类别的分类精确率

类别	精确率
AKIEC	0.79
BCC	0.82
BKL	0.74
DF	0.85
MEL	0.79
NV	0.96
VASC	0.90

表 2 展示了基于 Swin Transformer 的皮肤病诊断模型在 HAM10000 数据集上对七种皮肤病类别的分类精确率，AKIEC、BCC、BKL、DF、MEL、NV、VASC 分类精确率分别为 79%、82%、74%、85%、79%、96%、90%。总体来看，基于 Swin Transformer 的皮肤病诊断模型在 HAM10000 数据集上对不同皮肤病类别的分类精确率有所差异。在识别基底细胞癌、皮肤纤维瘤、黑色素细胞痣和血管病变方面表现较好，而在识别光化性角化病、黑色素瘤、良性角化样病变方面的精确率相对较低。提升 Swin Transformer 模型在光化性角化病、黑色素瘤和良性角化样病变等类别的分类精度将成为我们之后研究的重点。

4. 结论

本文采用一种基于 Swin Transformer 的皮肤病诊断模型，针对 Transformer 在处理高分辨率图像等长序列数据时，计算复杂度会随着序列长度的增加呈二次方增长，导致计算量偏大及缺乏对局部结构和位

置信息的有效利用,并且在长序列数据中可能会忽略一些局部细节特征等问题,通过引入窗口多头自注意力机制、移动窗口机制,来提升模型提取细节特征的能力和降低计算量,并在 HAM10000 皮肤病数据集上进行了实验验证,展现了模型优秀的分类性能,为医学图像识别及皮肤病诊断领域做出了新的贡献,为未来开发更精准的智能皮肤病诊断系统助力。然而,在这项工作中仍然有一些方面需要改进,这些将成为我们下一步研究的重点。

致 谢

感谢甘肃省计算中心提供的计算资源和技术支持。本文中的大量计算工作均在该中心进行,这对于本论文的顺利完成起到了至关重要的作用。

基金项目

本研究受甘肃省科技计划项目(24YFFA055, 22JR5RA797)、甘肃省云计算重点实验室开放课题(2023KFKT-005)与甘肃省重点人才项目(“东数西算”场景下的后量子数据加密传输机制研究)的资助。

参考文献

- [1] Proksch, E., Brandner, J.M. and Jensen, J. (2008) The Skin: An Indispensable Barrier. *Experimental Dermatology*, **17**, 1063-1072. <https://doi.org/10.1111/j.1600-0625.2008.00786.x>
- [2] Brown, M., Williams, A., Chilcott, R.P., Brady, B., Lenn, J., Evans, C., *et al.* (2024) Topically Applied Therapies for the Treatment of Skin Disease: Past Present and Future. *Pharmacological Reviews*, **76**, 689-790. <https://doi.org/10.1124/pharmrev.123.000549>
- [3] LeCun, Y., Bengio, Y. and Hinton, G. (2015) Deep Learning. *Nature*, **521**, 436-444. <https://doi.org/10.1038/nature14539>
- [4] Krit, D., Michael, S., Samer, A., Langlotz, C.P., Swift, A.J. and Bluethgen, C. (2023) External Validation, Radiological Evaluation, and Development of Deep Learning Automatic Lung Segmentation in Contrast-Enhanced Chest CT. *European Radiology*, **34**, 2727-2737. <https://doi.org/10.1007/s00330-023-10235-9>
- [5] Bernstein, I.A., Koornwinder, A., Hwang, H.H. and Wang, S.Y. (2024) Automated Recognition of Visual Acuity Measurements in Ophthalmology Clinical Notes Using Deep Learning. *Ophthalmology Science*, **4**, Article 100371. <https://doi.org/10.1016/j.xops.2023.100371>
- [6] Price, E. (2018) AI Is Better at Diagnosing Skin Cancer than Your Doctor, Study Finds. Yahoo Finance. <https://finance.yahoo.com/news/ai-better-diagnosing-skin-cancer-182057563.html>
- [7] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., *et al.* (2017) Attention Is All You Need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 6000-6010.
- [8] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., *et al.* (2021) Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 9992-10002. <https://doi.org/10.1109/iccv48922.2021.00986>
- [9] Tschandl, P., Rosendahl, C. and Kittler, H. (2018) The HAM10000 Dataset, a Large Collection of Multi-Source Dermatoscopic Images of Common Pigmented Skin Lesions. *Scientific Data*, **5**, Article No. 180161. <https://doi.org/10.1038/sdata.2018.161>
- [10] Lan, Z., Cai, S., He, X. and Wen, X. (2022) FixCaps: An Improved Capsules Network for Diagnosis of Skin Cancer. *IEEE Access*, **10**, 76261-76267. <https://doi.org/10.1109/access.2022.3181225>