摘要

细粒度视觉分类（FGVC）

1　背景介绍

人的眼睛是一个很
FGVC在我们的日常

2　研究现状

3　原理和方案

深度学习在机器视觉领域中广泛应用于细粒度分类任务，下边介绍一下它的原理和方案。

4　关键技术

4.1　模型

4.1.1　vision-Transformer

4.1.2　MetaFormer

MetaFormer[19]这是一个对机器视觉任务非常有用的网络架构，它将视觉信息和元信息结合起来。MetaFormer设计了一个五阶段的网络结构。第一阶段S0是一个简单的三层卷积结构。S1和S2是MBConv块[20]，具有挤压式激励。MBConv块是基于一个倒置的残差机制和一个瓶颈设计。倒置残差结构的设计是为了提供更高的非线性表示能力，同时保持模型的轻量级，而瓶颈使用较小的中间层来减少计算量。S3和S4是具有相对位置偏置的变换器块。这种偏置缓解了输入序列中的标记顺序不能用于自我注意操作的问题。MetaFormer的工作流程见图1。
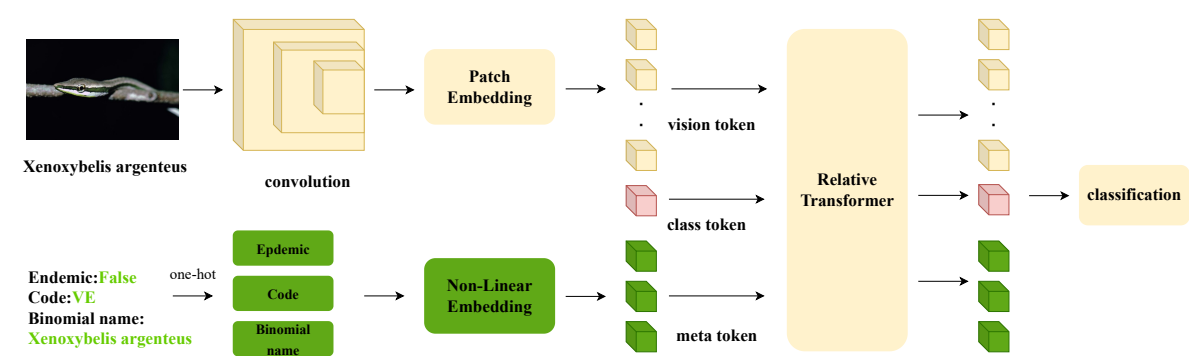

图 1:　该模型采用卷积层来提取视觉特征，然后通过Patch Embedding将图像特征转化为视觉标记。Code、Endemic和Binomial name的元信息被热编码，并通过Non-Liner Embedding获得元标记。视觉令牌、元令牌和类令牌使用相对转换层进行融合。融合后的标记在下面的注意区块中被连续聚合。最终输出的类标记用于类别预测。

4.2 损失函数

参考文献

[1] Anthony Miyaguchi, Jiangyue Yu, Bryan Cheungvivatpant, Dakota Dudley, and Aniketh Swain. Motif mining and unsupervised representation learning for birdclef 2022. *arXiv preprint arXiv:2206.04805*, 2022.

[2] B Cynthia Sherin and Kayalvizhi Jayavel. Effective vehicle classification and re-identification on stanford cars dataset using convolutional neural networks. In *Proceedings of 3rd International Conference on Artificial Intelligence: Advances and Applications: ICAIAA 2022*, pages 177–190. Springer, 2023.

[3] Xing Wang, Xiaoping Sun, Mingchang Cao, Yinlong Zhang, and Mingchang Zhu. A multi-scale approach to investigating the wintering habitat selection of red-crowned cranes in the yancheng nature reserve, china. *Pakistan Journal of Zoology*, 48(2), 2016.

[4] Nagifa Ilma Progga, Noortaz Rezoana, Mohammad Shahadat Hossain, Raihan Ul Islam, and Karl Andersson. A cnn based model for venomous and non-venomous snake classification. In *Applied Intelligence and Informatics: First International Conference, AII 2021, Nottingham, UK, July 30–31, 2021, Proceedings 1*, pages 216–231. Springer, 2021.

[5] Dichao Liu, Longjiao Zhao, Yu Wang, and Jien Kato. Learn from each other to classify better: Cross-layer mutual attention learning for fine-grained visual classification. *Pattern Recognition*, 140:109550, 2023.

[6] Yong Hou, Hangzai Luo, Wanqing Zhao, Xiang Zhang, Jun Wang, and Jinye Peng. Multilayer feature descriptors fusion cnn models for fine-grained visual recognition. *Computer Animation and Virtual Worlds*, 30(3-4):e1897, 2019.

[7] Junwei Han, Xiwen Yao, Gong Cheng, Xiaoxu Feng, and Dong Xu. P-cnn: Part-based convolutional neural networks for fine-grained visual categorization. *IEEE transactions on pattern analysis and machine intelligence*, 44(2):579–590, 2019.

[8] Jianlong Fu, Heliang Zheng, and Tao Mei. Look closer to see better: Recurrent attention convolutional neural network for fine-grained image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4438–4446, 2017.

[9] Sun Fayou, H Ngo, and Y Sek. Combining multi-feature regions for fine-grained image recognition. *Int. J. Image Graph. Signal Process*, 14:15–25, 2022.

[10] Amiza Amir, Nik Adilah Hanin Zahri, Naimah Yaakob, and R Badlishah Ahmad. Image classification for snake species using machine learning techniques. In *Computational Intelligence in Information Systems: Proceedings of the Computational Intelligence in Information Systems Conference (CIIS 2016)*, pages 52–59. Springer, 2017.

[11] Zihan Yang and Richard Sinnott. Snake detection and classification using deep learning. 2021.

[12] Anika Patel, Lisa Cheung, Nandini Khatod, Irina Matijosaitiene, Alejandro Arteaga, and Joseph W Gilkey Jr. Revealing the unknown: real-time recognition of galápagos snake species using deep learning. *Animals*, 10(5):806, 2020.

[13] Xiaohua Zhai, Yuxin Peng, and Jianguo Xiao. Heterogeneous metric learning with joint graph regularization for cross-media retrieval. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 27, pages 1198–1204, 2013.

[14] Grace Chu, Brian Potetz, Weijun Wang, Andrew Howard, Yang Song, Fernando Brucher, Thomas Leung, and Hartwig Adam. Geo-aware networks for fine-grained recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pages 0–0, 2019.

[15] Xiangteng He and Yuxin Peng. Fine-grained image classification via combining vision and language. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5994–6002, 2017.

[16] Louise Bloch, Jan-Frederick Böckmann, Benjamin Bracke, and Christoph M Friedrich. Combination of object detection, geospatial data, and feature concatenation for snake species identification. 2022.

[17] Isabelle Bolon, Lukáš Picek, Andrew M Durso, Gabriel Alcoba, François Chappuis, and Rafael Ruiz de Castañeda. An artificial intelligence model to identify snakes from across the world: Opportunities and challenges for global health and herpetology. *PLoS neglected tropical diseases*, 16(8):e0010647, 2022.

[18] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[19] Qishuai Diao, Yi Jiang, Bin Wen, Jia Sun, and Zehuan Yuan. Metaformer: A unified meta framework for fine-grained recognition. *arXiv preprint arXiv:2203.02751*, 2022.

[20] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520, 2018.

[21] Zhilu Zhang and Mert Sabuncu. Generalized cross entropy loss for training deep neural networks with noisy labels. *Advances in neural information processing systems*, 31, 2018.

[22] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4690–4699, 2019.

[23] Yue Liu, Wenxuan Tu, Sihang Zhou, Xinwang Liu, Linxuan Song, Xihong Yang, and En Zhu. Deep graph clustering via dual correlation reduction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 7603–7611, 2022.

[24] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.