A reliable and efficient deep learning model integrating convolutional neural network and transformer structure for fine-grained classification of chicken Eimeria species

P. He[1,2,3], Z. Chen[1,2,3], Y. He[1,2,3], J. Chen[1,2,3], K. Hayat[1,2,3], J. Pan[1,2,3] and H. Lin[1,2,3,*]
[1]*College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou, China*
[2]*Key Laboratory of Equipment and Informatization in Environment Controlled Agriculture, Ministry of Agriculture and Rural Affairs, China*
[3]*Key Laboratory of Intelligent Equipment and Robotics for Agriculture of Zhejiang Province, China*
[*]Corresponding author: Hongjian Lin, linhongjian@zju.edu.cn

Abstract

Chicken coccidiosis is a disease caused by Eimeria spp. and costs the broiler industry more than 14 billion dollars per year globally. Different chicken Eimeria species vary significantly in pathogenicity and virulence, so the classification of different chicken Eimeria species is of great significance for the epidemiological survey and related prevention and control. A new hybrid model integrating Transformer structure and the residual module in convolutional neural network (CNN), named Residual-Transformer-Fine-Grained (ResTFG), was proposed and evaluated for fine-grained classification of microscopic images of seven chicken Eimeria species. The results showed that ResTFG achieved the best performance with high accuracy and computationally efficient compared with traditional models. Specifically, the parameters, inference speed and overall accuracy of ResTFG are 1.95M, 256 FPS and 96.9%, respectively, which are 10.9 times lighter, 1.5 times faster and 2.7% higher in accuracy than the benchmark model. In addition, the results of ablation experiments showed that CNN or Transformer alone had model accuracies of only 89.8% and 87.0%. This study invented a reliable, computationally efficient, and promising deep learning model for the automatic fine-grain classification of chicken Eimeria species, which could potentially be embedded in microscopic devices at an affordable price for producers to improve the work efficiency of researchers and to be extended to other parasite ova, and applied to other agricultural tasks as a backbone.

Keywords: chicken Eimeria classification, deep learning, convolutional and transformer structure, complementary effect

Introduction

Chicken coccidiosis is a widespread and economically significant disease caused by parasite of the genus Eimeria (Chapman et al., 2013; Mesa et al., 2021), costing the global broiler industry more than 14 billion dollars per year (Adams et al., 2022). There are seven Eimeria species. Different chicken Eimeria species vary significantly in pathogenicity and virulence, so it is of practical significance to distinguish Eimeria species for epidemiological survey and related prevention and control.

The molecular biological methods are accurate and sensitive but require sophisticated protocols, and the morphological examination is a very challenging task for naked eyes due to the small morphological differences among chicken Eimeria species. Therefore, there is an urgent need to develop an automatic identification process for chicken Eimeria species. In some studies, The morphological characteristics of Eimeria oocysts were extracted and semi-automatic recognition was carried out by machine learning algirithms (Castañón et al., 2007; Kucera and Reznicky, 1991). Castañón et al. (2007) achieved the best overall accuracy of 85.75%. However, the semi-automatic methods requires manually designed features, which is cumbersome. The rapid development of convolutional neural network (CNN) has provided a powerful tool for the image recognition task (Esteva et al., 2017). Due to the superiority of CNN, it has been used for species identification of various parasites with good results and has been embedded in automated devices (Abade et al., 2022; Butploy et al., 2021; Lee et al., 2021; Thevenoux et al., 2021; Yang et al., 2020). Monge and Beltrán (2019) proposed a CNN model to classify chicken Eimeria species and the accuracy was improved to 90.42%, which still has room for improvement.

It is observed that the CNN-based models could achieve better results than traditional models. But these studies did not realize that the Eimeria species recognition is a fine-grained classification task, which focusing on the classifying objects of similar but different subtypes (Zhao et al., 2020). The Transformer structure has been successfully applied in major computer vision tasks (Dosovitskiy et al., 2021; Zheng et al., 2021), and TransFG achieved State-of-The-Art (SOTA) performance on five popular fine-grained classification benchmarks (He et al., 2021). The feature of local region connection makes CNN good at capturing local features, but lacks the ability to capture global features. Transformer can capture global features well, but is less capable of capturing local features. Therefore, theoretically integrating CNN and Transformer structure could improve the model performance, and the results have shown that the combination can indeed achieve good performance in their studies (Dai et al., 2021; Lu et al., 2022).

In this study, a new hybrid model, named Residual-Transformer-Fine-Grained (ResTFG), was proposed for the classification of chicken Eimeria species based on the residual block (He et al., 2016) and TransFG (He et al., 2021).

Materials and methods

Dataset

Dataset description. The dataset used in this study was from a publicly available website (http://www.coccidia.icb.usp.br/). Figure 1 shows the characteristic morphology of the seven chicken Eimeria species.

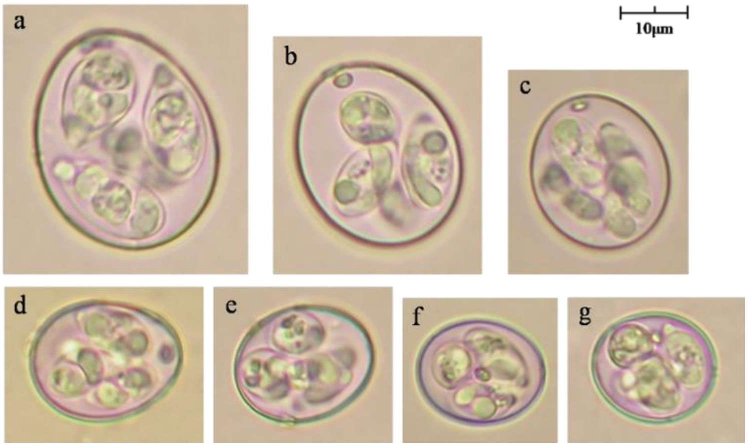

Figure 1: Micrographs of chicken Eimeria oocysts: (a) E. Maxima, (b) E. Brunetti, (c) E. Tenella, (d) E. Necatrix, (e) E. Praecox, (f) E. Acervulina, and (g) E. Mitis.

Dataset augmentation and splitting. There is an originally large difference in the number of different oocyst categories with uneven distribution. All E. Maxima images were flipped horizontally, and 300 E. Brunetti images and 200 E. Necatrix images were randomly selected for horizontal flipping. After balancing the dataset, the total number of images increased from 4243 to 5103. The details of the dataset are shown in Table 1.

Table 1: The number of images of the chicken Eimeria oocyst dataset.

Class label	Species name	Original	After data augmentation	Partitioning of the dataset (7:3)	
				Training	Test
ACE	E. Acervulina	742	742	520	222
BRU	E. Brunetti	442	742	520	222
MAX	E. Maxima	360	720	504	216
MIT	E. Mitis	825	825	578	247
NEC	E. Necatrix	502	702	492	210
PRA	E. Praecox	676	676	474	202
TEN	E. Tenella	696	696	488	208
Total number		4243	5103	3576	1527

Methods

Equipment and Environment. To facilitate intensive computation in model training, a professional deep learning platform, SYS-4029GP-TRT was used, equipped with 2 × Intel© Xeon(R) Gold 6147M CPU @ 2.50GHz, a total of 260 GB memory, and 8 graphics cards including 4 × Nvidia TITAN RTX and 4 × Nvidia GeForce RTX 2080 Ti, a total of 140 GB video memory. The testing and inference speed measurement of models were run on a desktop computer with GeForce RTX 3080 GPU and Inter(R) Core (TM) i9-10900KF CPU @3.70GHz. In terms of the software environment, Python-3.8, PyCharm-Professional-2021.2.3, and Pytorch-GPU-1.8.1 framework were used.

Proposed model. The framework of the proposed ResTFG model is shown in Figure. 2. The left and the right parts are the CNN and Transformer branches, respectively.

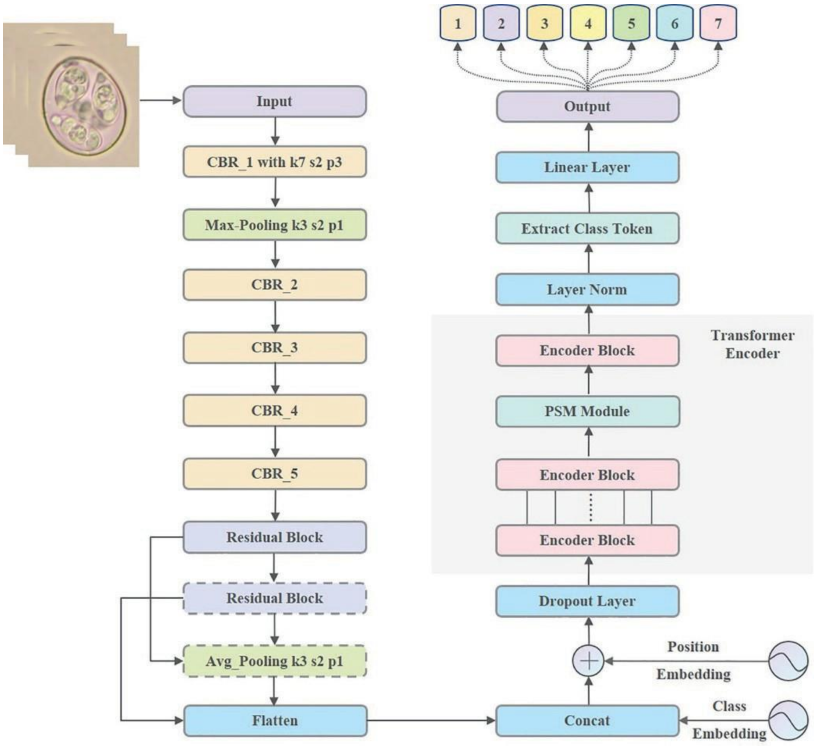

Figure 2: The overview of the proposed Residual-Transformer-Fine-Grained (ResTFG) model.

The CNN branch consists of an input layer, a maximum pooling layer, an average pooling layer, a flatten layer, five CBR modules (acronym for Convolution Batch-Normalization ReLU (rectified linear unit)), and one or two residual blocks. The k, s and p in Figure 2 represent the kernel size, stride, and padding of the convolution layer, respectively. As shown in Figure 3(a), the CBR module consists of three layers, a convolution layer with a kernel size of 3×3, a stride of 1×1 and a padding of 1, followed by a batch normalization (BN) layer and a ReLU layer. The structure of the residual block is shown in Figure 3(b), composed of an input layer, four CBR modules, a downsampling operation and an output layer. The downsampling operation is implemented by a convolution layer with a kernel size of 1×1, a stride of 2×2 and no padding, and is connected to a BN layer afterward. When there is only one residual block, it is then connected to the average pooling layer and the flattening layer. The second residual block is directly connected to the flattened layer when there are two residual blocks. This design aims to match the feature dimensions after the flattened layer with the input dimensions of the Transformer branch.

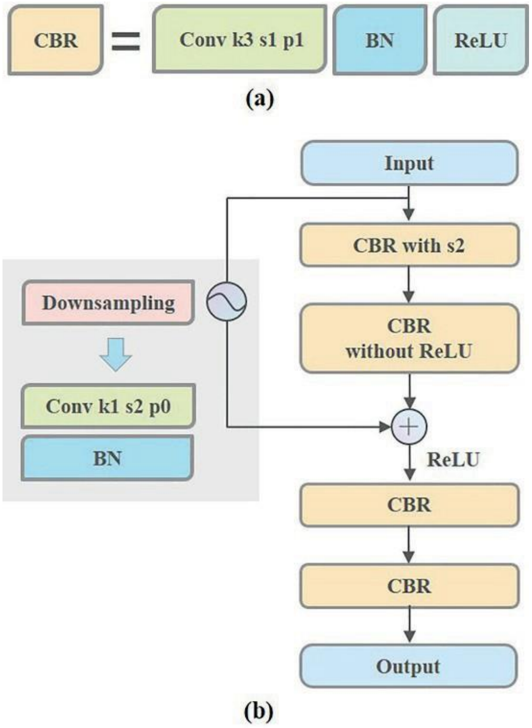

Figure 3: The structure of the Convolution Batch-Normalization ReLU module (CBR) (a), and the residual block (b).

Experimental setting. All models used were trained from scratch. The initial learning rate is 0.001, and the decay rate is 0.5 for every 20 epochs. In addition, the number of epochs is 100, the batch size is 64, and the optimizer is Stochastic Gradient Descent (SGD) with a 0.9 momentum and a 1e-4 weight decay.

Results and discussion

Performance of existing models

Seven existing SOTA models were compared. The results are shown in Table 2. The accuracy of MobilenetNet_V3_Small, MobilenetNet_V3_Large (Howard et al., 2019) and Shufflenet_V2_x1_0 (Ma et al., 2018) was relatively low. It was surprising VGG11(Simonyan and Zisserman, 2014), with the maximum number of parameters, achieved the fastest inference speed of 409 FPS, but its memory consumption is high.

DenseNet121 (Huang et al., 2016) had a good accuracy, but very slow inference speed. TransFG_B16 (He et al., 2021) also had good accuracy but high memory consumption. ResNet34 (He et al., 2016) was identified as a balanced model with 21.29M parameters, 166FPS and 94.2% accuracy, which was selected as the benchmark model.

Table 2: The performance of SOTA models.

Model Name	Parameters (M)	Speed (FPS)	Accuracy (%)
VGG11	128.80	409	86.9
MobilenetNet_V3_Small	1.53	157	86.3
MobilenetNet_V3_Large	4.21	107	90.7
ResNet34	21.29	166	94.2
DenseNet121	6.96	56	92.7
Shufflenet_V2_x1_0	1.26	146	80.4
TransFG_B16	85.80	104	93.5

Optimization of transformer branch

The initial value of three hyperparameters in the Transformer branch, i.e., hidden size, MLP dimension and the number of multi-attention heads was 768, 3072 and 12 respectively. The initial model was named ResTFG (C13, H12, L8) (a), where [C] represents the number of convolution layers, [H] represents the number of multi attention head, [L] represents the number of the encoder block layer, and the letter suffixes (a-d) correspond to the different hidden size and MLP dimension. As shown in Table 3, ResTFG (C13, H12, L2) achieved a balanced performance.

Table 3: The performance comparison of the ResTFGs for the Transformer branch with different hyperparameters and the number of the encoder block layer.

Model Name	Hidden size	MLP dimension	Number heads	Number Layers	Parameters (M)	Speed (FPS)	Accuracy (%)
TransFG_B16	768	3072	12	12	85.80	104	93.5
ResTFG (C13, H12, L8) (a)	768	3072	12	8	82.14	105	97.2
ResTFG (C13, H12, L8) (b)	384	1536	12	8	21.50	112	97.0
ResTFG (C13, H12, L8) (c)	288	1024	12	8	11.90	113	95.7
ResTFG (C13, H12, L8) (d)	192	768	12	8	6.12	116	95.7
ResTFG (C13, H8, L8)	384	1536	8	8	21.50	114	96.7
ResTFG (C13, H4, L8)	384	1536	4	8	21.50	114	96.6
ResTFG (C13, H2, L8)	384	1536	2	8	21.50	114	96.0
ResTFG (C13, H12, L6)	384	1536	12	6	17.96	134	96.5
ResTFG (C13, H12, L4)	384	1536	12	4	14.41	165	96.9
ResTFG (C13, H12, L3)	384	1536	12	3	12.63	183	97.0
ResTFG (C13, H12, L2)	384	1536	12	2	10.86	216	97.1

Optimization of CNN branch

As shown in Table 4, ResTFG (C9, H12, L2) (c) obtained the advantages of both high performance and lightweight, with the number of parameters of 1.95M, an inference speed of 256FPS, and an accuracy of 96.9%, which is 10.9 times lighter, 1.5 times faster, and 2.7% higher in accuracy than ResNet34.

Table 4: The performance comparison of the ResTFG for the CNN branch with different kernel parameters and the number of the convolution layer.

Model Name	Hidden size	MLP dimension	Parameters (M)	Speed (FPS)	Accuracy (%)
ResTFG (C13, H12, L2)	384	1536	10.86	216	97.1
ResTFG (C9, H12, L2) (a)	384	1536	10.09	254	96.5
ResTFG (C9, H12, L2) (b)	180	720	2.50	256	96.7
ResTFG (C9, H12, L2) (c)	156	624	1.95	256	96.9
ResTFG (C5, H12, L2)	156	624	1.80	299	96.2

Ablation studies on ResTFG

The test results are shown in Table 5. There is no doubt that the number of parameters would be reduced and the inference speed would increase with only the CNN or Transformer branch. But the accuracy of these two models decreased significantly by 7.1% and 9.9%, respectively. The results showed sufficient evidence that the hybrid model can fully utilize the advantages of CNN and Transformer.

Table 5: The performance of the ResTFG with only CNN or Transformer branch.

Model Name	Parameters (M)	Speed (FPS)	Accuracy (%)
CNN Only	0.92	549	89.8
Transformer Only	1.03	403	87.0
ResTFG (C9, H12, L2) (c)	1.95	256	96.9

Balance of accuracy and inference speed

Figure 4 shows the bubble plots of seven SOTA models and our model, with larger bubbles representing more parameters. Our model is a balanced model with the advantages of both high performance and lightweight.

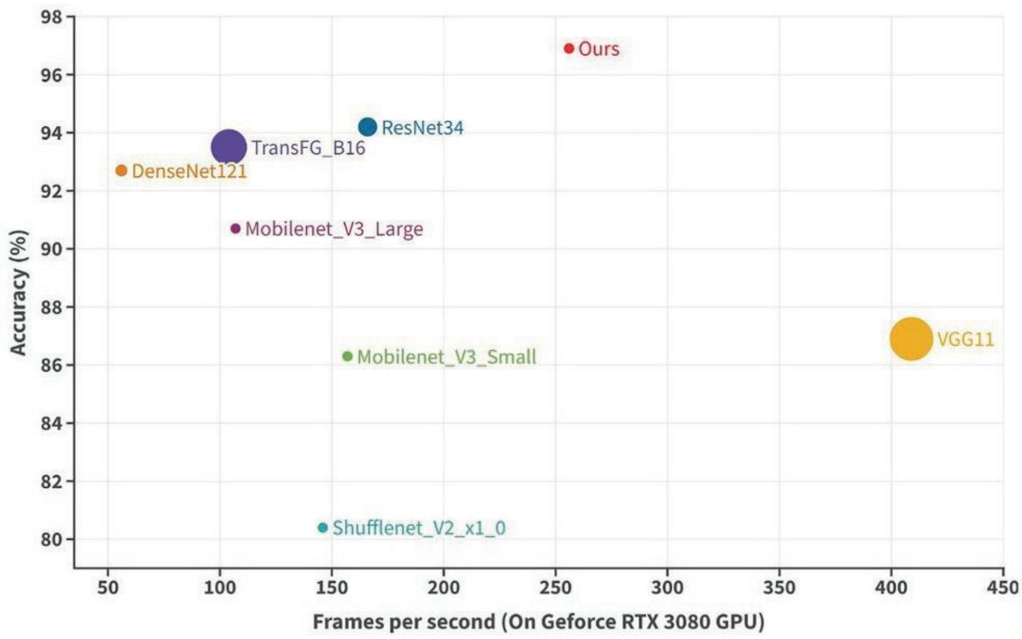

Figure 4: The bubble plots of seven SOTA models and our model.

Conclusions

A new hybrid deep-learning model, named ResTFG, which integrates the advantages of the CNN and Transformer structure, was proposed in this study. The CNN structure containing residual blocks was used as the backbone, which has a powerful feature extraction ability, and compensated for the defect of CNN lacking a global receptive field through the deployment of the multi-head attention mechanism in Transformer. The ablation experiments proved the synergistic effect of integrating the CNN and Transformer structure. Overall, the proposed ResTFG model performs well, achieving an accuracy of 96.9%, an inference speed of 256 FPS, and a memory consumption of 1.95M, which has the advantages of both high accuracy and computationally efficient. This model can improve the work efficiency of researchers. More importantly, for people who do not have the ability to identify Eimeria species with the naked eye, they can obtain species distribution information to infer the severity of the disease with the help of this automatic identification system, which can provide guidance for subsequent medication and the basis for effective control measures.

Acknowledgments

This work was supported by the Key R&D Program of Zhejiang Province (2021C02026) and the Agriculture Research System of China (CARS-40).

References

Abade, A., Porto, L.F., Ferreira, P.A., and de Barros Vidal, F. (2022) Nemanet: A convolutional neural network model for identification of soybean nematodes. *Biosystems Engineering* 213, 39-62.

Adams, D.S., Kulkarni, R.R., Mohammed, J.P., and Crespo, R. (2022) A flow cytometric method for enumeration and speciation of coccidia affecting broiler chickens. *Veterinary Parasitology* 301, 109634.

Butploy, N., Kanarkard, W., and Maleewong Intapan, P. (2021) Deep learning approach for ascaris lumbricoides parasite egg classification. *Journal of Parasitology Research* 2021.

Castañón, C.A., Fraga, J.S., Fernandez, S., Gruber, A., and Costa, L.D.F. (2007) Biological shape characterization for automatic image recognition and diagnosis of protozoan parasites of the genus eimeria. *Pattern Recognition* 40(7), 1899-1910.

Chapman, H.D., Barta, J.R., Blake, D., Gruber, A., Jenkins, M., Smith, N.C., Suo, X., and Tomley, F.M. (2013) A selective review of advances in coccidiosis research. *Advances in Parasitology* 83, 93-171.

Dai, Y., Gao, Y., and Liu, F. (2021) Transmed: Transformers advance multi-modal medical image classification. *Diagnostics* 11(8), 1384.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., and Gelly, S. (2020) An image is worth 16x16 words: Transformers for image recognition at scale. *ArXiv* 2010, 11929.

He, J., Chen, J., Liu, S., Kortylewski, A., Yang, C., Bai, Y., and Wang, C. (2022) Transfg: A transformer architecture for fine-grained recognition. In: *Proceedings of the AAAI Conference on Artificial Intelligence* 852-860.

He, K., Zhang, X., Ren, S., and Sun, J. (2016) Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 770-778.

Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., and Vasudevan, V. (2019) Searching for mobilenetv3. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision* Seoul, Korea (South), 1314-1324.

Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K.Q. (2017) Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 4700-4708.

Kucera, J., and Reznický, M. (1991) Differentiation of species of eimeria from the fowl using a computerized image-analysis system. *Folia Parasitol (Praha)* 38(2), 107-113.

Lee, C.-C., Huang, P.-J., Yeh, Y.-M., Li, P.-H., Chiu, C.-H., Cheng, W.-H., and Tang, P. (2022) Helminth egg analysis platform (heap): An opened platform for microscopic helminth egg identification and quantification

based on the integration of deep learning architectures. *Journal of Microbiology, Immunology and Infection* 55(3), 395-404.
Lu, X., Yang, R., Zhou, J., Jiao, J., Liu, F., Liu, Y., Su, B., and Gu, P. (2022) A hybrid model of ghost-convolution enlightened transformer for effective diagnosis of grape leaf disease and pest. *Journal of King Saud University-Computer and Information Sciences* 34(5), 1755-1767.
Ma, N., Zhang, X., Zheng, H.-T., and Sun, J. (2018) Shufflenet v2: Practical guidelines for efficient cnn architecture design. In: *Proceedings of the European conference on computer vision (ECCV)* 116-131.
Mesa, C., Gómez-Osorio, L., López-Osorio, S., Williams, S., and Chaparro-Gutiérrez, J. (2021) Survey of coccidia on commercial broiler farms in colombia: Frequency of eimeria species, anticoccidial sensitivity, and histopathology. *Poultry Science* 100(8), 101239.
Monge, D.F., and Beltrán, C.A. (2019) Classification of eimeria species from digital micrographies using cnns. In: *10th International Conference on Pattern Recognition Systems (ICPRS-2019)* Tours, France, 88-91.
Shirley, M. (1997) Eimeria spp. From the chicken: Occurrence, identification and genetics. *Acta Veterinaria Hungarica* 45(3), 331-347.
Simonyan, K., and Zisserman, A. (2014) Very deep convolutional networks for large-scale image recognition. *ArXiv* 1409, 1556.
Thevenoux, R., Van Linh, L., Villessèche, H., Buisson, A., Beurton-Aimar, M., Grenier, E., Folcher, L., and Parisey, N. (2021) Image based species identification of globodera quarantine nematodes using computer vision and deep learning. *Computers and Electronics in Agriculture* 186, 106058.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., and Polosukhin, I. (2017) Attention is all you need. *Advances in Neural Information Processing Systems* 30, 5998-6008.
Yang, F., Poostchi, M., Yu, H., Zhou, Z., Silamut, K., Yu, J., Maude, R.J., Jaeger, S., and Antani, S. (2019) Deep learning for smartphone-based malaria parasite detection in thick blood smears. *IEEE Journal Of Biomedical And Health Informatics* 24(5), 1427-1438.
Zhao, J., Peng, Y., and He, X. (2020) Attribute hierarchy based multi-task learning for fine-grained image classification. *Neurocomputing* 395, 150-159.
Zheng, S., Lu, J., Zhao, H., Zhu, X., Luo, Z., Wang, Y., Fu, Y., Feng, J., Xiang, T., and Torr, P.H. (2021) Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 6881-6890.