ICCV23 PARIS

INTERNATIONAL CONFERENCE ON COMPUTER VISION

WORKSHOP ON COMPUTER VISION FOR AUTOMATED MEDICAL DIAGNOSIS

IEEE COMPUTER SOCIETY

The Medical Image Computing and Computer Assisted Intervention Society MICCAI

NAZARBAYEV UNIVERSITY | School of Engineering and Digital Sciences

A Comparative Study of Vision Transformer Encoders and Few-shot Learning for Medical Image Classification

Maxat Nurgazin and Nguyen Anh Tu

Nazarbayev University, Republic of Kazakhstan

Abstract

Computer vision has been significantly impacted by Vision Transformer (ViT) networks. However, most existing deep learning-based methods primarily rely on a lot of labeled data to train reliable classifiers for accurate prediction. This requirement might be impractical in the medical field.

This study explores the application of ViT in few-shot learning scenarios for medical image analysis, addressing the challenges posed by limited data availability. We evaluate various ViT models alongside few-shot learning algorithms, perform cross-domain experiments, and analyze the impact of data augmentation techniques.

Our findings indicate that when combined with ProtoNets, ViT architectures outperform CNN-based counterparts and achieve competitive performance against SOTA approaches on benchmark datasets.

Motivation

- Vision Transformers (ViTs) have emerged as an alternative to CNNs, showing impressive performance on various tasks.
- CNNs struggle with learning long-range pixel relationships due to locality, which ViTs can handle more effectively.
- Medical imaging often has limited labeled data, making it difficult to train deep learning models.
- Few-shot learning (FSL) is a promising approach for handling limited labeled data.

Goal

To our knowledge, ViT architectures have not been used in the field of medical image classification in few-shot learning scenarios. Therefore, given their success in other areas of computer vision, it is important to assess their performance in this area under various conditions.

Contribution

- Investigate the efficacy of various ViT models for few-shot medical image classification.
- Study how different few-shot learning algorithms impact the performance of ViT models.
- Analyze the impact of advanced data augmentation techniques on ViT models.
- Explore the effect of a cross-domain scenario on the performance of few-shot learners.
- framework through experiments when running on the Spark clusters.
- Our methods achieve state-of-the-art performance on challenging medical datasets of few-shot medical image classification

References

- Zhiyong et al., Pfemed: Few- shot medical image classification using prior guided feature enhancement. Pattern Recognition, 2023.
- Rishav et al., Metamed: Few-shot medical image classification using gradient-based meta-learning. Pattern Recognition 2021.
- Yisheng et al., A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities. arXiv preprint arXiv:2205.06743, 2022

Methodology

Problem definition:

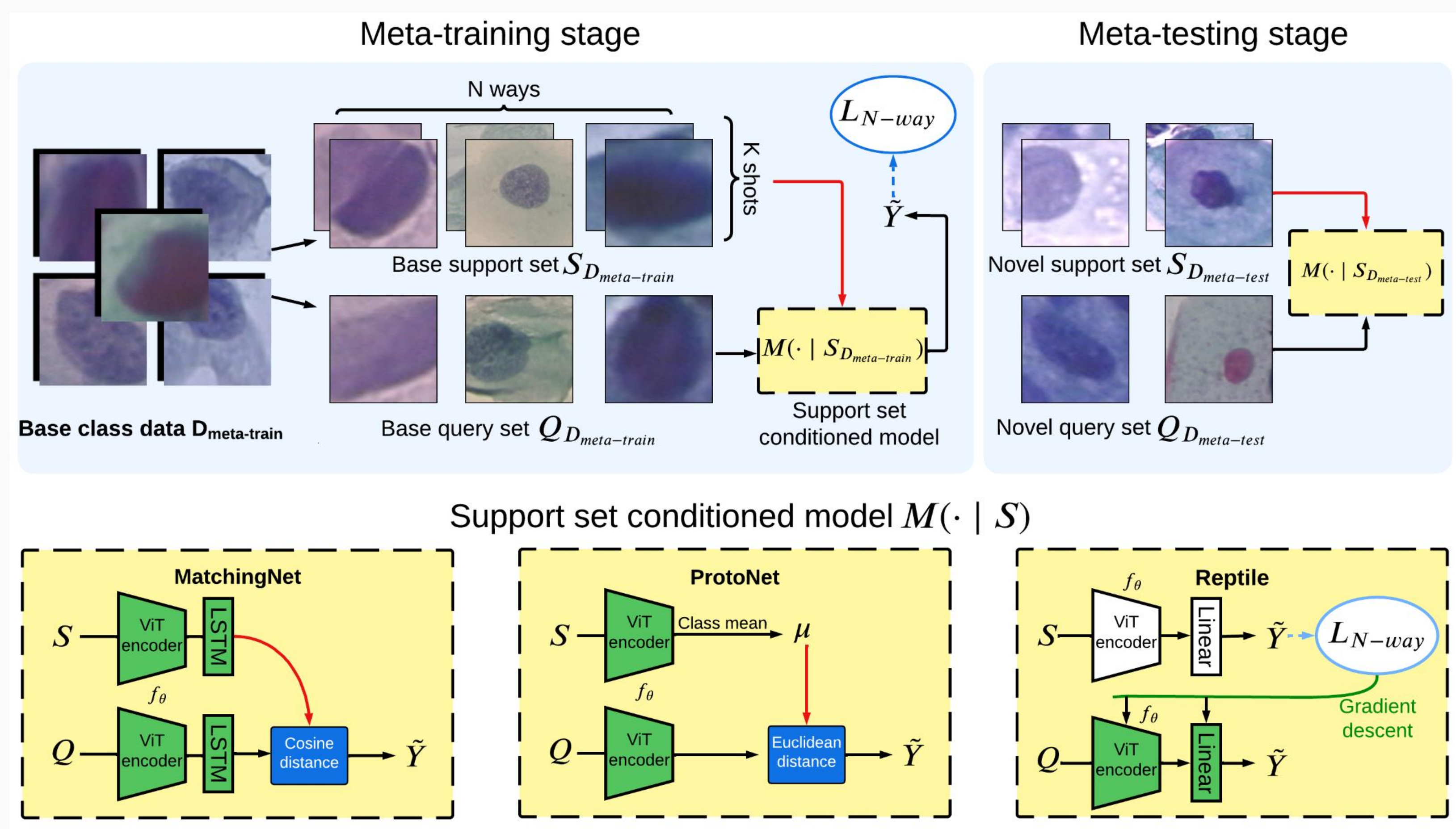
Let D = D_1, D_2, ..., D_n be a collection of n medical datasets, with each dataset D_k consisting of pairs $(x, y)_j$ representing an image and its label.

Datasets are divided into meta-test set ($D_{meta\text{-}test}$) and meta-train set ($D_{meta\text{-}train}$)

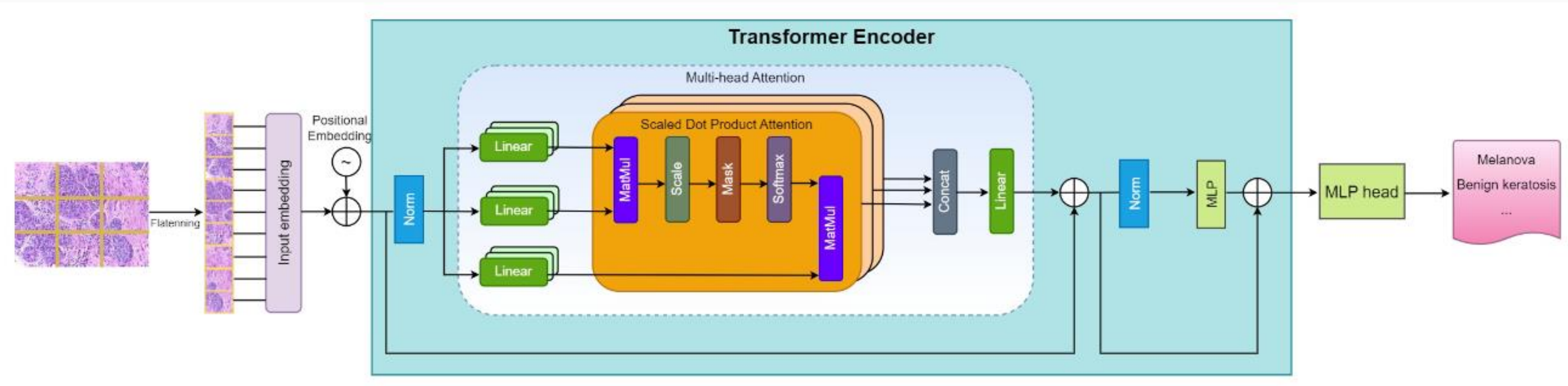
Utilize abundant data in $D_{meta\text{-}train}$ to learn better initial weights (Reptile) or develop effective embedding space (ProtoNet & MatchingNet)

Goal: Improve performance on problems $D_{meta\text{-}test}$ with limited data (novel class data)

Overview of the system pipeline



ViT Encoder:



Results

- **Datasets**: BreakHis (9109 microscopic images of breast tumor tissues from 82 patients with 8 classes), ISIC 2018 (10,015 dermoscopic images of skin lesions across 7 classes), and Pap Smear (917 microscopic images of cervical smears with 7 classes).
- **Experimental Settings**: Pre-trained models obtained from the timm library.
  - ProtoNet: 20 epochs, 500 episodes per epoch, SGD optimizer, learning rate of 10-5 or 10-6, cosine annealing learning rate schedule.
  - Reptile: SGD optimizer, learning rate of 10-3 for inner optimization, learning rate of 10-1 for outer meta-update, 1000 meta-iterations, batch size of 10 tasks, 5 and 50 adaptation steps for each task
- **Evaluation metric**: Accuracy (%) as evaluation metric. 400 episodes randomly selected from novel categories in the test set. Average accuracy rate for image classification.

Algorithm	Model	2-way			3-way		
		3-shot	5-shot	10-shot	3-shot	5-shot	10-shot
ProtoNet	MViT	74.64	76.94	81.50	60.60	64.23	69.23
	ViT_tiny	81.03	83.61	86.52	67.84	71.82	77.68
	ViT_small	**84.35**	**86.70**	89.72	72.10	76.18	81.45
	ViT_base	83.94	86.02	**90.26**	**72.75**	**77.69**	**81.99**
	DeIT_base	72.17	76.53	81.40	57.86	62.38	69.07
	Swin_base	82.49	84.17	89.12	70.75	74.67	79.92
	ResNet50	66.62	68.65	72.81	51.43	53.83	58.34
	VGG16	74.11	78.17	82.11	60.68	64.58	70.84
Reptile	MViT	62.80	67.00	71.80	53.00	54.47	60.33
	ViT_tiny	75.80	78.40	83.50	64.13	68.67	75.13
	ViT_small	70.30	76.10	80.40	63.13	72.13	78.53
	ViT_base	59.30	67.40	72.70	53.27	62.27	70.53
	DeIT_base	72.80	79.40	83.00	61.73	64.73	73.60
	Swin_base	67.30	74.20	81.10	60.60	69.00	75.93
	ResNet50	71.70	72.70	76.50	47.60	51.60	54.93
	VGG16	64.70	72.90	78.60	56.53	62.40	70.67
MatchingNet	MViT	72.41	75.70	78.59	58.79	62.00	65.62
	ViT_tiny	76.66	79.88	83.41	63.42	66.62	71.95
	ViT_small	78.40	81.61	86.34	65.50	70.00	76.47
	ViT_base	79.81	84.19	88.21	67.70	73.30	79.17
	DeIT_base	73.67	77.11	81.60	58.23	62.54	70.47
	Swin_base	72.84	75.60	80.12	58.66	63.09	68.34
	ResNet50	67.99	71.69	75.66	52.93	56.53	61.60
	VGG16	72.20	75.94	79.90	59.76	61.63	67.98

Performance of models using different meta-learning algorithms for ISIC 2018 dataset

Algorithm	Model	2-way			3-way		
		3-shot	5-shot	10-shot	3-shot	5-shot	10-shot
Protonet	MViT	80.84	84.36	86.88	68.04	73.24	78.37
	ViT_tiny	84.65	86.96	88.86	74.33	77.92	81.17
	ViT_small	**92.40**	**94.05**	**94.90**	**86.38**	**89.09**	**90.62**
	ViT_base	92.05	93.26	93.94	85.21	88.48	89.47
	DeIT_base	88.88	89.38	91.22	78.77	81.70	85.28
	Swin_base	85.42	87.56	89.78	75.73	79.88	82.46
	ResNet50	70.49	71.75	69.61	57.74	58.48	59.60
	VGG16	88.75	89.34	91.76	79.04	82.53	85.63
Reptile	MViT	80.60	80.20	84.30	72.00	73.20	78.87
	ViT_tiny	85.60	88.00	90.10	75.27	82.47	86.00
	ViT_small	86.80	90.70	93.90	77.73	82.27	87.00
	ViT_base	77.10	81.50	88.00	70.53	78.27	88.07
	DeIT_base	84.40	87.30	92.50	76.20	82.33	86.27
	Swin_base	81.40	87.20	87.40	80.47	81.53	87.87
	ResNet50	80.90	82.00	89.30	73.67	75.87	81.73
	VGG16	84.90	88.80	93.20	77.73	81.67	88.60
MatchingNet	MViT	80.10	81.97	85.32	67.61	72.27	76.35
	ViT_tiny	86.00	89.09	90.91	77.34	80.92	83.93
	ViT_small	90.84	92.56	94.27	84.74	87.02	89.23
	ViT_base	89.56	89.50	92.10	78.24	82.53	86.33
	DeIT_base	89.25	89.36	91.70	79.43	82.39	85.16
	Swin_base	82.01	83.58	86.94	70.34	74.38	78.27
	ResNet50	76.01	74.46	77.89	59.66	61.23	64.08
	VGG16	87.66	88.30	89.94	77.07	79.60	83.45

Performance of models using different meta-learning algorithms for Pap Smear dataset

Algorithm	Model	FSL	2-way			3-way		
			3 shot	5 shot	10 shot	3 shot	5 shot	10 shot
ProtoNet	ViT_small	Standart	**84.35**	**86.70**	**89.72**	**72.10**	76.18	**81.45**
		CutOut	81.73	85.89	89.22	70.55	**76.23**	81.13
		MixUp	-	-	-	-	-	-
		CutMix	-	-	-	-	-	-
	ResNet50	Standart	**66.62**	68.65	**72.81**	**51.43**	**53.83**	**58.34**
		CutOut	65.52	**68.75**	72.18	49.32	53.81	57.74
		MixUp	-	-	-	-	-	-
		CutMix	-	-	-	-	-	-
Reptile	ViT_small	Standart	76.05	80.30	85.55	**67.50**	**73.15**	77.37
		CutOut	75.30	**80.35**	83.95	64.87	69.97	76.53
		MixUp	**77.50**	79.40	**85.75**	66.20	71.33	77.87
		CutMix	74.85	77.75	85.65	67.40	72.57	**79.63**
	ResNet50	Standart	70.28	**75.78**	**78.83**	54.47	58.22	61.58
		CutOut	68.73	73.60	76.58	**55.70**	59.90	64.67
		MixUp	**70.75**	74.15	78.03	55.00	**60.65**	**64.95**
		CutMix	70.10	74.60	77.95	53.70	58.92	63.62

Effect of different Augmentation techniques on Few-shot classification for ISIC 2018 Dataset

Target	Setting	2-way			3-way		
		3-shot	5-shot	10-shot	3-shot	5-shot	10-shot
BreakHis X100	CD + PN	74.12	78.74	84.11	62.72	68.90	74.72
	Non CD	**80.64**	**83.80**	**87.62**	**69.39**	**75.91**	**81.47**
Pap Smear	CD + PN	**92.22**	**94.12**	94.85	86.22	88.82	90.47
	Non CD	92.40	94.05	**94.90**	**86.38**	**89.09**	**90.62**

Cross-domain ISIC2018-to-BreakHis and -Pap smear

Algorithm	Model	2-way			3-way		
		3-shot	5-shot	10-shot	3-shot	5-shot	10-shot
ProtoNet	ViT_small	**84.35**	**86.70**	**89.72**	**72.10**	**76.18**	**81.45**
	ResNet50	66.62	68.65	72.81	51.43	53.83	58.34
MatchingNet	ViT_small	78.40	81.61	86.34	65.50	70.00	76.47
	ResNet50	67.99	71.69	75.66	52.93	56.53	61.60
Reptile	ViT_small	76.05	80.30	85.55	67.50	73.15	77.37
	ResNet50	70.28	75.78	78.83	54.47	58.22	61.58
	MetaMed[20]	72.75	75.62	81.37	54.83	59.33	69.75
-	PFEMed[6]	81.69	83.87	85.14	66.94	69.78	73.81

Comparison with MetaMed and PFEMed on ISIC 2018

Algorithm	Model	2-way			3-way		
		3-shot	5-shot	10-shot	3-shot	5-shot	10-shot
ProtoNet	ViT_small	80.64	83.80	87.62	**69.39**	**75.91**	81.47
	ResNet50	68.62	72.12	73.31	55.80	60.28	61.88
MatchingNet	ViT_small	76.53	82.09	**88.33**	67.13	72.88	**81.80**
	ResNet50	73.45	76.58	79.14	59.53	62.70	66.98
Reptile	ViT_small	68.10	75.60	81.60	54.40	63.13	72.20
	MetaMed[20]	78.75	81.38	83.88	63.08	66.42	74.08
-	PFEMed[6]	**82.16**	**85.28**	86.90	69.21	75.04	78.93

Comparison with MetaMed and PFEMed on BreakHis