SiamFly - A Novel Pre-trained Architecture for Large-Scale Fine-Grained Butterfly Classification

[1]Hari Theivaprakasham, [1]V Sowmya, [2]Vinayakumar Ravi, [1]E.A. Gopalakrishnan and [1]K.P. Soman

[1]Center for Computational Engineering and Networking (CEN), Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India.

[2]Center for Artificial Intelligence, Prince Mohammad Bin Fahd University, Khobar, Saudi Arabia.

CVPR JUNE 19-24 2022 NEW ORLEANS • LOUISIANA

Motivation

• The population of butterflies has been declining in recent decades due to rapid climate change, habitat destruction, and urbanization.

• Butterflies play several vital roles in maintaining biodiversity, the ecological food chain and the sustainable environment.

• Butterflies act as bio-indicators due to their quick and sensitive responses to subtle habitat or climate change, and it is crucial to document their presence for effective conservation.

• Butterfly identification is a fine-grained classification problem due to highly complex features and variability between inter-species and intra-species levels.

• Traditional methods of butterfly identification requires expert intervention and more expensive laboratory techniques.

• Adaptation of artificial intelligence technology will help extract essential features and identify butterflies at the species level.

Objective

• To build a fine-grained intelligent automated butterfly identification system considering complex features and an unbalanced image dataset.

Dataset

Total images: 62,287 images **Total number of species:** 686 butterfly species

• Highly unbalanced Dataset • Research Grade • Complex backgrounds

• Species' sexual dimorphism • Minor inter-class and intra-class variations

• Various seasonal and mimetic forms

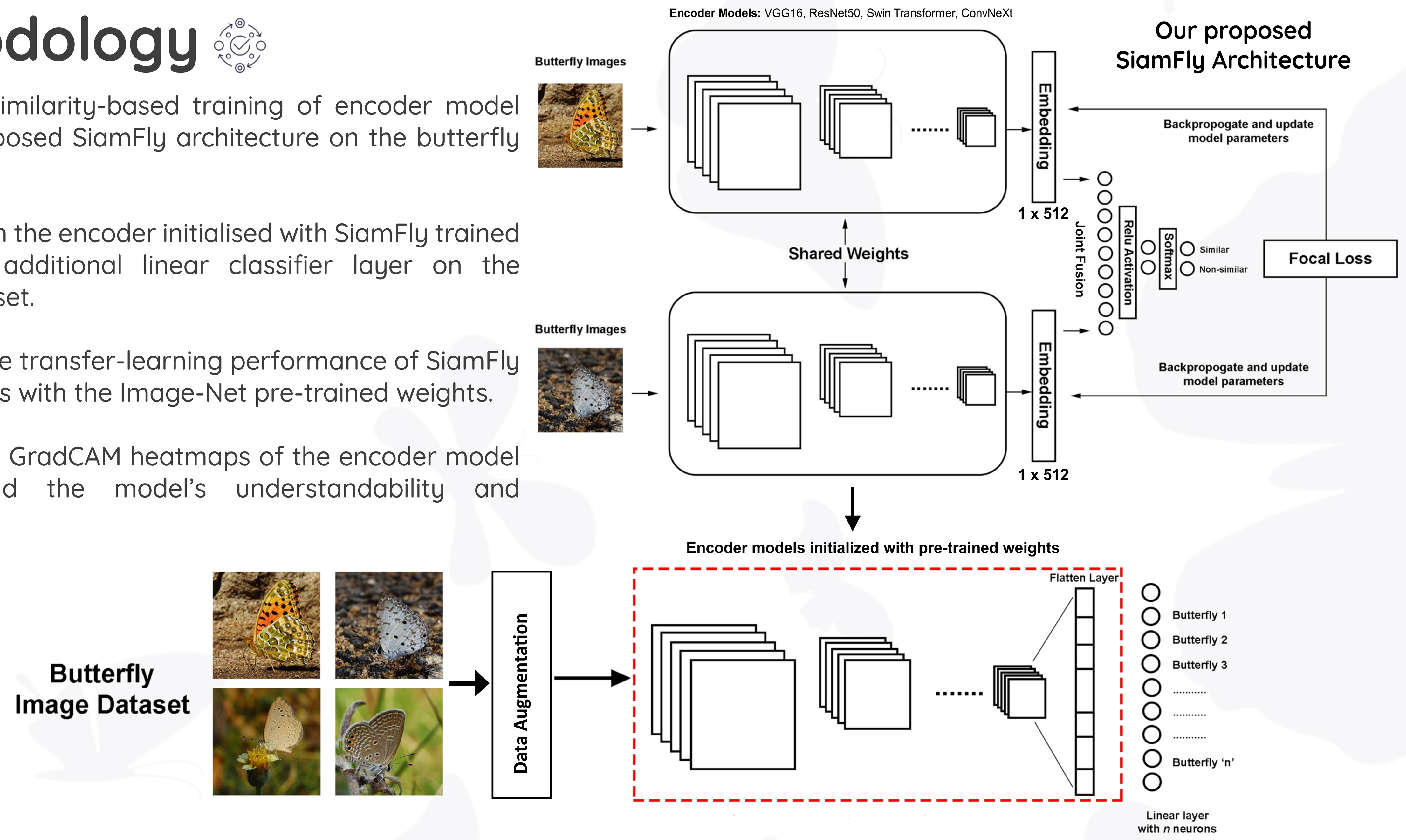
Methodology

1) Butterfly similarity-based training of encoder model using our proposed SiamFly architecture on the butterfly dataset.

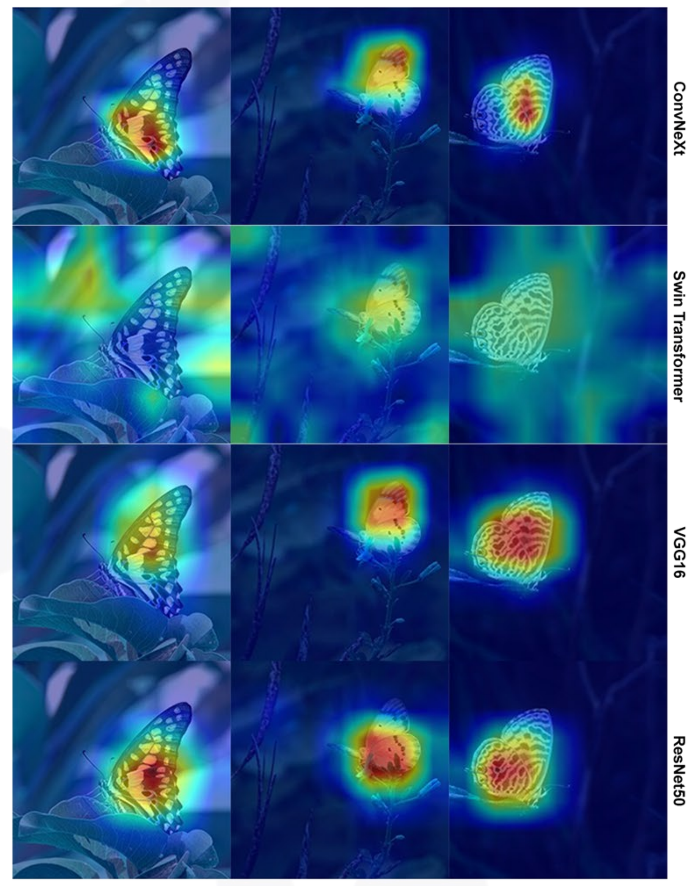
2) Entirely train the encoder initialised with SiamFly trained weights and additional linear classifier layer on the butterfly dataset.

3) Compare the transfer-learning performance of SiamFly trained weights with the Image-Net pre-trained weights.

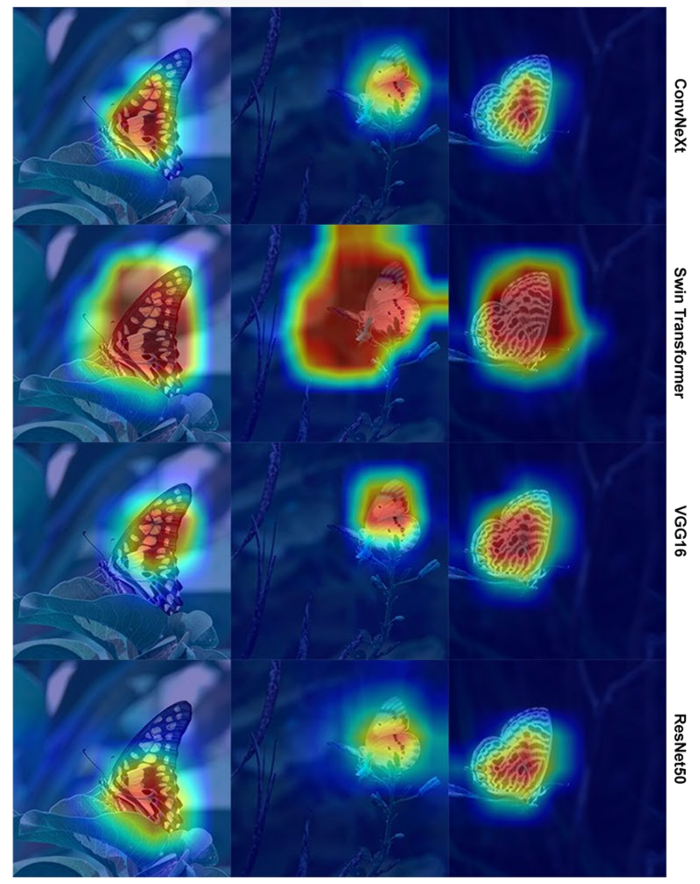
4) Analyse the GradCAM heatmaps of the encoder model to understand the model's understandability and learnability.



Results



GradCAM visualization of various encoder models trained using ImageNet pre-trained weights



GradCAM visualization of various encoder models trained using SiamFly pre-trained weights

Evaluation results on test data using transfer learning with linear classifier output layer

Pre-training Method	Encoder Model	Accuracy	Top-3 accuracy	Precision	Recall	F1-score	Time per epoch (second)
ImageNet	ResNet50	0.671	0.830	0.590	0.480	0.506	64
ImageNet	VGG16	0.590	0.785	0.426	0.374	0.359	90
ImageNet	ConvNeXt	0.714	0.870	0.570	0.500	0.511	74
ImageNet	Swin Transformer	0.781	0.926	0.646	0.598	0.574	73
SiamFly (ours)	ResNet50	0.760	0.916	0.596	0.541	0.539	64
SiamFly (ours)	VGG16	0.592	0.768	0.477	0.388	0.410	90
SiamFly (ours)	ConvNeXt	0.732	0.865	0.592	0.485	0.508	74
SiamFly (ours)	Swin Transformer	0.849	0.948	0.795	0.714	0.734	73

Results of butterfly classification by training entire model using our proposed SiamFly weights and ImageNet weights

Pre-training Method	Percentage of training data	Models	Accuracy	Top-3 Accuracy	Precision	Recall	F1-score
ImageNet	100%	Swin-Transformer	0.869	0.983	0.815	0.761	0.766
		ConvNeXt	0.862	0.981	0.823	0.746	0.764
		ResNet50	0.861	0.981	0.842	0.721	0.763
		VGG16	0.858	0.972	0.816	0.715	0.759
ImageNet	30%	Swin-Transformer	0.718	0.954	0.573	0.483	0.498
		ConvNeXt	0.813	0.966	0.703	0.613	0.634
		ResNet50	0.791	0.963	0.712	0.631	0.671
		VGG16	0.72	0.956	0.617	0.452	0.482
SiamFly	**100%**	Swin-Transformer	0.873	0.982	0.800	0.755	0.757
		ConvNeXt	**0.882**	**0.987**	**0.816**	**0.770**	**0.773**
		ResNet50	0.881	0.987	0.843	0.731	0.771
		VGG16	0.86.1	0.981	0.823	0.746	0.766
SiamFly	30%	Swin-Transformer	0.859	0.974	0.767	0.713	0.716
		ConvNeXt	0.855	0.976	0.821	0.751	0.742
		ResNet50	0.821	0.969	0.731	0.678	0.683
		VGG16	0.815	0.949	0.725	0.668	0.621

Conclusion

• The proposed SiamFly architecture-based pre-training method can achieve better performance on a low-data regime and surpass the ImageNet-based transfer-learning performance by a more significant margin.

• Transfer-learning approach with Siamfly pre-trained weights yielded the best accuracy of 84.9% with Swin transformer as encoder compared to 78.1% when using ImageNet pre-trained weights.

• When completely trained the model with Siamfly pre-trained weights, the ConvNeXt model performed the best with 88.2% accuracy compared to 86.2% with ImageNet weights.

• By using only 30% of the training data, the SiamFly-based model pre-training resulted in the equivalent performance of the model trained with 100% data using ImageNet-based weight.