Received: January 3, 2023. Revised: March 17, 2023.

INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Design of a Hybrid GWO CNN Model for Identification of Synthetic Images via Transfer Learning Process

Nupoor M. Yawale[1]* **Neeraj Sahu[1]** **Nikkoo N. Khalsa[2]**

[1]*Computer Science &Engineering Department, G. H. Raisoni University Amravati (Maharashtra), India*
[2]*Electronics and Telecommunication Department, Dean Training & Placement, PRMIT & R, Badnera-Amravati (Maharashtra), India*
* Corresponding author's Email: nupooryawale2021@gmail.com

Abstract: Visual representation of synthetic images is very accurate, due to which it is difficult to differentiate them for their natural counterparts. Existing models that perform this differentiation are either very complex, or cannot be scaled for multidomain image sets. Moreover, the accuracy of these models depends directly upon type of dataset & feature sets used for training & validation purposes. To overcome these limitations, this paper proposes design of a Hybrid GWO CNN Model for identification of Synthetic Images for big data Applications. The proposed model initially extracts multidomain feature sets from input images, that includes wavelet, cosine, fourier & convolutional features. These features are processed via a grey wolf optimization (GWO) Model, that assists in improving inter-class feature variance while minimizing intra-class variance levels. The GWO Model identifies training & validation sets, thereby assisting the classification model to accurately differentiate between different image types. To perform this task, a variance-based fitness function was modelled that covers both inter-class & inter-class variance levels. This classification is performed via use of a transfer learning-based CNN Model, that extends VGG-19 for high-efficiency operations. The proposed model was tested on a large number of datasets including Synthetic Fruit, Unsplash, ESPL Synthetic Image, and Okazaki Synthetic Texture Image (OSTI) databases. Based on these datasets, accuracy, precision, computational delay, recall & AUC (Area Under the Curve) metrics were evaluated & compared with existing models. It was observed that the proposed model showcased 9.5% better accuracy, 8.3% higher precision, 6.5% better recall, and 3.9% faster performance when compared with existing models. Due to such high performance under different datasets, the proposed model is useful for large-scale synthetic-image identification use cases.

Keywords: Natural, Synthetic, Image, Classification, Cosine, Fourier, Wavelet, Convolutional, Transfer, Learning, Accuracy.

1. Introduction

Identification of synthetic images requires design of multi-domain operations that include, but are not limited to, large-scale dataset collection, pre-processing for filtering operations, multi-domain feature representationand identification of features that can improve classification performance. These features are processed via design of accurate classifiers that can categorize images based on these features, development of post-processing models that can pre-empt presence of synthetic regions in natural image sets. A typical synthetic image identification model [1] is depicted in Fig. 1, wherein image processing operations including instance matching, image rendering, partial modelling, feature adaptation, class validation, and continuous learning can be observed. The model uses machine learning methods (MLMs), in order to classify input datasets into different synthetic classes. These models are generally designed via convolutional neural networks (CNNs), gated recurrent units (GRUs), Q-Learning, and other deep learning methods.

Similar models [2-4] along with their functional nuances, application-specific advantages, contextual limitations, and deployment-specific future scopes

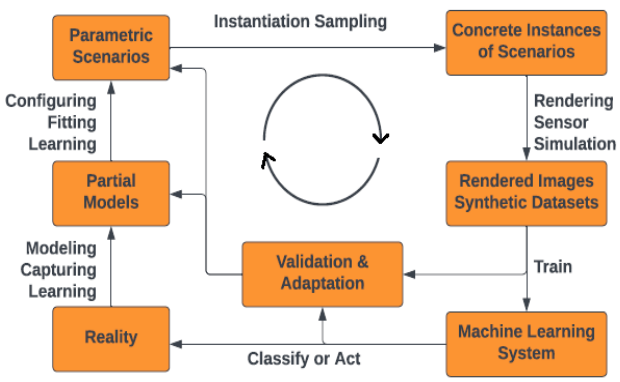

Figure. 1 Design of a typical synthetic image classification model via continuous validation operations

are discussed in the next section of this text. Based on this discussion, it can be observed that existing models for synthetic image classification are either extremely sophisticated or incapable of scaling to multi-domain image sets. Furthermore, the accuracy of these models is closely related to the type of dataset and feature sets utilized for training and validation. To address these limitations, section 3 proposes the development of a Hybrid GWO CNN Model for detection of Synthetic Images in Big Data scenarios. Initially, the proposed model identifies multi-domain feature sets from input images, including wavelet, cosine, fourier, and convolutional features. These features are handled using a grey wolf optimization (GWO) Model, which helps to improve inter-class feature variance while reducing intra-class variation levels. The GWO Model identifies training and validation sets, enabling the classification model in distinguishing between distinct image kinds. To complete this goal, a variance-based fitness function covering both inter-class and inter-class variation levels was created. This categorization is carried out using a CNN Model based on transfer learning, which extends VGG-19 for high-efficiency operations. The model is validated in section 4, wherein its performance is evaluated under multiple datasets, and compared with various existing methods. This paper concludes with some insightful remarks on the proposed model's performance, as well as suggestions for improving it under real-time use cases.

Contributions of the paper

The following are major contributions of this paper,

- The proposed model extracts multidomain feature sets from input images, that includes wavelet, cosine, fourier & convolutional features.
- This paper proposes a grey wolf optimization (GWO) model that assists in improving inter-class feature variance while minimizing intra-class variance levels.
- This classification is performed via use of a transfer learning-based CNN Model, that extends VGG-19 for high-efficiency operations.

Organization of the paper

After referring to the introduction and short review about different image classification models, next section proposes design of a Hybrid GWO CNN model for identification of synthetic images for big data Applications. The model was evaluated under multiple datasets in section 3, and compared with different models for validation purposes. Finally, this text is concluded with some contextual observations about the proposed model, and also recommends methods to improve its performance under real-time scenarios.

2. Literature review

A wide variety of classification models are proposed by researchers, and each of them vary in terms of their internal characteristics. For instance, work in [5, 6] proposes use of latent space learning (LSL), and support vector machine (SVM) with Bayesian classifier, which assists in identification of synthetic & natural images with high accuracy. But these models have higher complexity, which limits their applicability. To overcome this issue, work in [7] proposes design of global-local network structure with CNN, which assists in improving classification accuracy for multiple use cases. It uses vision transformers, which makes it useful for high density feature extraction & classification deployments. Based on this model, work in [8, 9, 10] proposes use of multiple feature weighted sparse graph (MFWSG), self-distillation feature learning network (SDFLN), and transfer fuzzy clustering and active learning-based classification operations, which assists in improving its classification performance under multiple use cases. These models showcase low complexity, but cannot be scaled for multiple numbers of classes. To enhance this performance, work in [11, 12, 13, 14] proposes use of adaptive fuzzy learning (AFL), active ensemble deep learning (AEDL), autoencoder regularization joint contextual attention network (ARJCAN), which assists in improving classification performance for multiple datasets and scenarios. Similar models are discussed in [15, 16, 17], which propose use of Spatial & Semantic Features, Novel Attention Fully Convolutional Network Method (NAFCNN), which

allow the model to augment multiple feature sets for enhancing classification performance.

Models that use hybrid conditional random field model (HCRF) [18], FrequencyAware attention with adaptive feature fusion [19], statistical scattering components (SSCs) [20], and simple linear iterative clustering (SLIC) [21], which assists in extraction of multi-domain feature sets. These models aim at improving classification performance for a wide variety of application scenarios. They can be extended via use of the models discussed in [22, 23, 24], that propose use of residual network (ResNet) withDeep autoencoder (DAE), deep CNN (DCNN), and structure match generative adversarial network (SM GAN), which assist in enhancing feature performance for multiple image analysis scenarios. Models like Ensemble Dual-Branch CNN (EDB CNN) [25], semi supervised complex valued GAN (SSCV GAN) [26], CNN, normalized difference vegetation index (NDVI) [27] and RidgeNets [28] with speckle reduction regularization (RN SRR) [29] propose use of deep learning, thereby improving classification performance for different applications. But these models are either very complex, or cannot be scaled for multi-domain image sets. Moreover, the accuracy of these models depends directly upon type of dataset & feature sets used for training & validation purposes.

Issues with existing techniques

The identification of synthetic images is an active research area that has gained significant attention in recent years. The reviewed techniques used for identification of synthetic images have limitations, some of which include,

1. The current techniques used for identifying synthetic images are not always effective in distinguishing between real and synthetic images [1-3]. Some of the techniques are based on statistical analysis of pixel-level features, which can be easily manipulated by sophisticated generators [4].
2. Some techniques that are effective in identifying synthetic images require significant computational resources, which limits their scalability for real-world applications [7, 8, 10].
3. Most current techniques are tailored to specific types of synthetic images, such as those generated by Generative Adversarial Networks (GANs) [26]. This limits their ability to identify synthetic images generated by other types of models or techniques [11, 12].
4. Many identification techniques are dependent on training data that may not be representative of the full range of synthetic images that may exist [18, 19, 20]. This can lead to inaccurate classification of synthetic images that are not well-represented in the training dataset samples [22, 23, 24].
5. Some identification techniques are vulnerable to adversarial attacks, where a malicious actor can manipulate the input image in a way that the identification technique misclassifies it [13, 15].

Overall, while current techniques have made progress in identifying synthetic images, there are still many models that can be used for improving its performance under real-time scenarios.

3. Design Of the proposed hybrid GWO CNN model for identification of synthetic images via transfer learning process

Based on the literature survey, it was observed that existing models for synthetic & natural image classification are either very complex, or cannot be scaled for multi-domain image sets. Moreover, the accuracy of these models depends directly upon type of dataset & feature sets used for training & validation purposes, which further limits it scalability. To overcome these limitations, this section proposes design of a Hybrid GWO CNN Model for identification of Synthetic Images for big data Applications.

Flow of the proposed model is depicted in Fig. 2, wherein it can be observed that the model initially extracts multi-domain feature sets from input images, that includes wavelet, cosine, fourier & convolutional features. These features are processed via a grey wolf optimization (GWO) Model that assists in improving inter-class feature variance while minimizing intra-class variance levels. The GWO model identifies training & validation sets, thereby assisting the classification model to accurately differentiate between different image types. To perform the classification task, a variance-based fitness function was modelled that covers both inter-class & inter-class variance levels. This classification is performed via use of a transfer learning-based CNN Model, which extends VGG-19 for high-efficiency operations. The model initially collects datasets from multiple sources, and aggregates them to form a corpus of Natural & Synthetic images. These images are initially passed through a feature extraction layer, that extracts Wavelet, Cosine, Fourier & Convolutional feature sets.

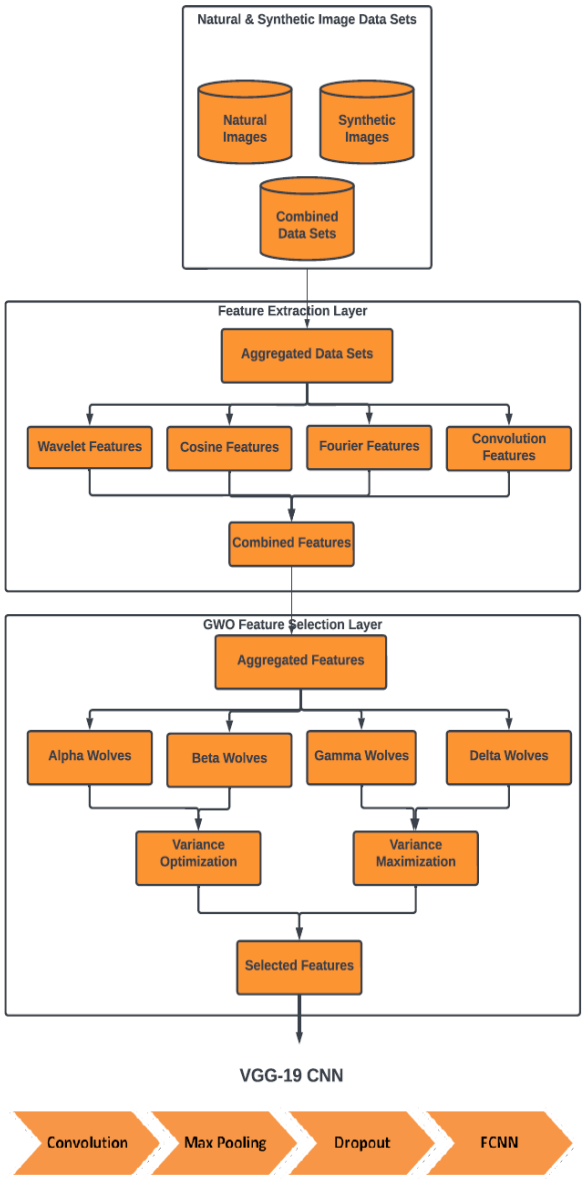

Figure. 2 Overall flow of the proposed synthetic image classification process

The Haar Wavelet features are extracted via Eqs. (1)-(4), which assists in representing input image as approximate, diagonal, horizontal & vertical components.

$$w_{i,j_{approx}} = \frac{x_{i,j}+x_{i+1,j}}{2} \quad (1)$$

$$w_{i,j_{detail}} = \frac{x_{i,j}-x_{i+1,j}}{2} \quad (2)$$

$$w_{i,j_{horizontal}} = \frac{x_{i,j}+x_{i,j+1}}{2} \quad (3)$$

$$w_{i,j_{vertical}} = \frac{x_{i,j}-x_{i,j+1}}{2} \quad (4)$$

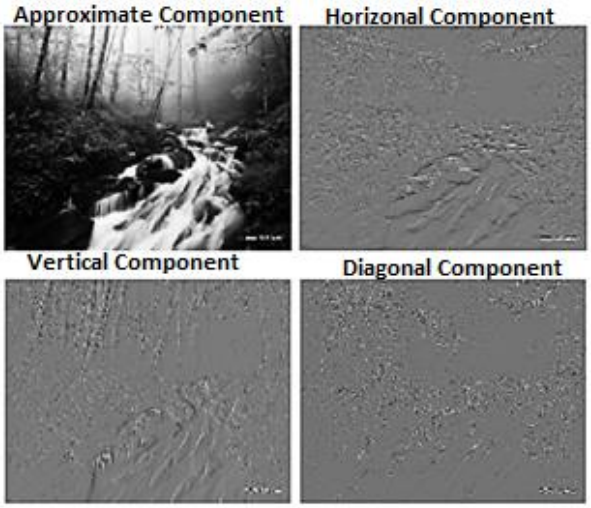

Figure. 3 Extraction of haar wavelet components of natural images

Where, $x_{i,j}$ represents image pixel at i^{th} row, and j^{th} column, while $w_{i,j}$ represents respective wavelet components. These components are extracted for every pixel, and can be visualized from Fig. 3, where natural scene image is used for division analysis. similar to haar wavelets, discrete cosine transform (DCT) is extracted via Eq. (5), thereby representing input image in multiple domains. In Eq. (5), R, C represents rows & columns, while x represents image pixel levels. Similar to DCT components, discrete fourier transform (DFT) components are also extracted via Eq. (6), which assist in identification of different frequency representations in input image sets.

$$DCT = \frac{1}{2*\sqrt{R*C}}\sum_{i=1}^{R-1}\sum_{k=1}^{C-1} x_{i,k} * \cos\left[\frac{(2*i+1)*j*pi}{2*R}\right] * \cos\left[\frac{(2*k+1)*j*pi}{2*C}\right] \quad (5)$$

$$F_{j,k} = \frac{1}{R*C}\sum_{j=0}^{R-1}\sum_{k=0}^{C-1} x_j * \exp\left(-i*2*pi*\frac{j}{R}\right) * \exp\left(-i*2*pi*\frac{k}{C}\right) \quad (6)$$

All these components are combined with convolutional features, which are evaluated via Eq. (7), to form a super feature vector set (SFVS), and is processed by the GWO Model for identification of variant feature sets.

$$Conv_{out_{i,j}} = \sum_{a=-\frac{m}{2}}^{\frac{m}{2}}\sum_{b=-\frac{n}{2}}^{\frac{n}{2}} I_{2D}(i-a, j-b) * LReLU\left(\frac{m}{2}+a, \frac{n}{2}+b\right) \quad (7)$$

Where, I_{2D} represents input image, while m, n represents number of rows & columns of the image,

and a, b represents window size for extraction of convolutional features. These features are activated by a leaky rectilinear unit (LReLU) operation, which is represented via Eq. (8) as follows,

$$LReLU(x,y) = l_a * x + l_b * y, when x < 0 \; or \; y < 0, else \; LReLU(x,y) = x + y \quad (8)$$

To further augment these features, window sizes are varied between 32x32, 64x64, 128x128, and 256x256 ranges, which assists in extraction of a large number of feature sets. Due to which redundancies are inherent, which are removed via use of a GWO based feature selection layer. This layer works via the following process,

- To initialize the Grey Wolf Optimizer, setup the following optimization constants,
  - Optimization Iterations = N_i
  - Optimization Wolves = N_w
  - Rate at which the model is learning = L_r
- All Wolves must be setup as 'Delta' Wolves, which will assist in modifying their internal configurations.
- Scan all Wolves for N_i iterations, via following process,
  - Check Wolf status, and if it is marked as 'Alpha', 'Beta', or 'Gamma', then skip it and go to the next Wolf in sequence
  - Else, modify internal configuration of the Wolf via following process,
    - Select N_f stochastic features, via Eq. (9),

$$N_f = STOCH(L_r * N_{SFV}, N_{SFV}) \quad (9)$$

Where, N_{SFV} represents number of features present in the Super Feature Vector, which is the total features extracted via combination of DCT, DFT, Wavelet and Convolutional operations.

- Based on these feature sets, identify solution fitness via Eq. (10),

$$f = \sqrt{\frac{\sum_{a-1}^{N_f}\left(x_a - \frac{\sum_{i=1}^{N_f}\sqrt{\frac{\sum_{j=1}^{N_{SFV}-N_f}\left(x_j - \frac{\sum_{k=1}^{N_{SFV}-N_f} x_k}{n}\right)^2}{N_{SFV}-N_{f-1}}}}{N_{SFV}-N_f}\right)^2}{N_f - 1}} \quad (10)$$

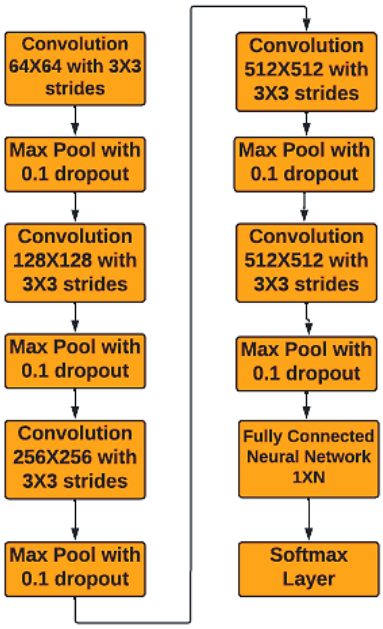

Figure. 4 Internal flow of the VGGNet-19 model for classification of selected feature sets into synthetic & natural images

Where, $a, i\&j$ represents number of features in current class, number of features common between current & other classes and number of features in other classes, while x represents value of extracted features.

- Due to this fitness, an interclass variance is extracted between different classes, which assist in identification of features that are highly variant across Multiple Natural & Synthetic class sets.
  - Similar operations are performed for all 'Delta' wolves, and at the end of final iteration, a fitness threshold is calculated via Eq. (11),

$$f_{th} = \sum_{i=1}^{N_s} f_i * \frac{L_r}{N_s} \quad (11)$$

- Once an iteration is completed, then re-evaluate Wolf status via following process,
  - Wolf is marked as 'Alpha', if $f > 2 * f_{th}$
  - Else, it is marked as 'Beta', if $f > f_{th}$
  - Else, it is marked as 'Gamma', if $f > L_r * f_{th}$
  - Else, it is marked as 'Delta', if $f \leq L_r * f_{th}$

At the end of final iteration, select wolf with maximum fitness levels, and use its features for classification process.

To perform this classification, the VGGNet-19 model is used, which consists of a combination of convolutional, max pooling, drop out and fully connected neural network (FCNN) layers. The internal layer flow of the VGGNet-19 model is

depicted in Fig. 4, where these layers along with an output Soft Max Layer are depicted for final classification operations. Convolutional features are extracted via Eq. (7), which results in identification of multidomain feature sets. These feature sets are reduced via a Max Pooling layer, that evaluates a variance threshold via Eq. (12) for identification of variant feature sets.

$$f_{th} = \left(\frac{1}{X_k} * \sum_{x \in X_k} x^{p_k}\right)^{1/p_k} \quad (12)$$

Where, X_k represents extracted convolutional features, while p_k represents variance of these feature sets.

Based on this variance, features with $f > f_{th}$ are passed to the next convolutional layers, while others are removed from the convolution process. This process is repeated for all layers, and finally a fully connected neural network (FCNN) is used to identify final image class. This layer uses a Soft Max based activation function and optimizes weights w& biases b to identify output class via Eq. (13) as follows,

$$c_{out} = SoftMax\left(\sum_{i=1}^{N_f} f_i * w_i + b\right) \quad (13)$$

The output class is compared with ground truth values, and parameters like accuracy, precision, recall, & delay needed for evaluation are calculated & compared with existing models, which assists in performance estimation for multiple use cases. This performance can be observed from the next section of this text.

4. Result analysis & comparison

The proposed model uses a combination of GWO with multi-domain feature sets & CNN in order to efficiently classify images into natural & synthetic classes. To validate the proposed model, its accuracy (A), precision (P), recall (R) & classification delay (D), and was evaluated on the following datasets,

- IEEE synthetic image database which is available at https://ieee-dataport.org/keywords/synthetic-image-database
- ESPL synthetic image database which is available at http://signal.ece.utexas.edu/~bevans/synthetic/
- CVOnline Database which is available at https://homepages.inf.ed.ac.uk/rbf/CVonline/Imagedbase.htm

Table 1. Average accuracy of classification for different image types

TIS	A (%) CNN [3]	A (%) SDFL [9]	A (%) FWE NET [14]	A (%) HGC MSI TL
1875	79.93	62.93	64.97	98.81
2813	80.02	63.00	65.04	98.92
3750	80.07	63.04	65.08	98.98
4688	80.10	63.07	65.11	99.01
5625	80.11	63.08	65.12	99.03
6563	80.13	63.09	65.13	99.05
7500	80.15	63.11	65.15	99.08
8438	80.19	63.13	65.18	99.12
9375	80.24	63.17	65.22	99.18
10313	80.30	63.23	65.27	99.27
11250	80.37	63.28	65.32	99.34
12188	80.43	63.32	65.37	99.42
13125	80.48	63.37	65.42	99.49
14063	80.53	63.41	65.46	99.55
15000	80.58	63.44	65.49	99.60

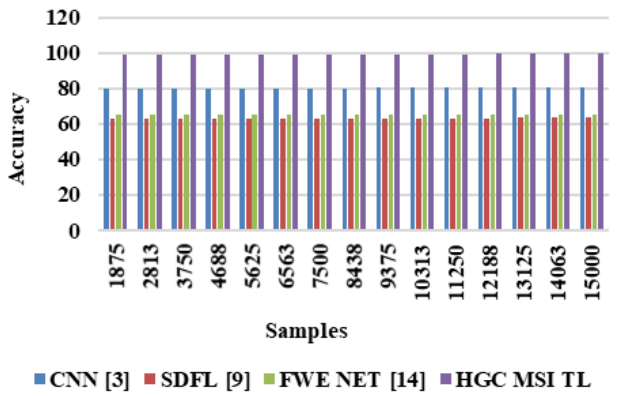

Figure. 5 Average accuracy of classification for different image types

When aggregated, these datasets constitute of over15k images of different categories, making them ideal for validation of different classifier sets. The dataset was divided into a ratio of 65:15:20, out of which 65% images were used for training, 15% were used for testing while 20% were used for validation purposes. To further facilitate this process, CNN [3], SDFL [9], &FWE NET [14] were evaluated with these Test Image Sets (TIS). The NAFCNN [16] model uses a high-efficiency CNN Model for identification of Synthetic images;thus, it is used for comparisons. While, the SDFL [9] and FWE NET [14] Models showcased use of GAN, which is a highly effective classification technique, and thus is used for comparison with our proposed model, under real-time dataset samples.The accuracy performance can be observed from Table 1 as follows,

Table 2. Average precision of classification for different image types

TIS	P (%) CNN [3]	P (%) SDFL [9]	P (%) FWE NET [14]	P (%) HGC MSI TL
1875	64.56	59.69	50.05	79.79
2813	64.63	59.75	50.11	79.88
3750	64.67	59.79	50.14	79.93
4688	64.69	59.81	50.16	79.96
5625	64.70	59.82	50.16	79.97
6563	64.71	59.83	50.17	79.99
7500	64.73	59.85	50.19	80.01
8438	64.76	59.88	50.21	80.05
9375	64.80	59.91	50.24	80.10
10313	64.85	59.96	50.28	80.16
11250	64.90	60.01	50.32	80.23
12188	64.95	60.06	50.36	80.29
13125	65.00	60.10	50.39	80.34
14063	65.04	60.14	50.43	80.39
15000	65.07	60.17	50.45	80.44

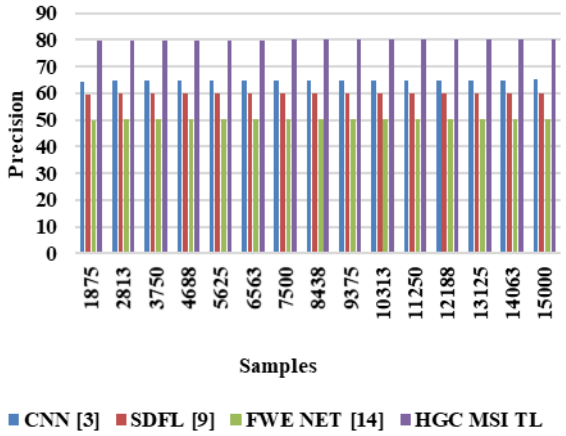

Figure. 6 Average precision of classification for different image types

The results of this assessment and the comparison shown in Fig. 5 show that the suggested model is capable of obtaining anaccuracy that is 18.4 percent higher than CNN [3], 25.5 percent higher than SDFL [9], and 23.9 percent higher than FWE NET [14] for a variety of test imagesizes. This is because GWO was combined with multiple feature sets & CNN, which helps improve search performance across a variety of use cases. Evaluations for the classification precision that are comparable to these are observed in Table 2 as follows,

The results of this assessment and the comparison shown in Fig. 6 show that the suggested model is capable of obtaining an accuracy that is 14.5 percent higher than CNN [3], 18.5 percent higher than SDFL [9], and 23.4 percent higher than FWE NET [14] for a variety of test image sizes. This is because GWO was combined with multiple feature sets & CNN, which helps improve search performance across a variety of use cases. Evaluations for the classification recall that are comparable to these are observed in Table 3 as follows,

Table 3. Average recall of classification for different image types

TIS	R (%) CNN [3]	R (%) SDFL [9]	R (%) FWE NET [14]	R (%) HGC MSI TL
1875	63.75	58.94	49.43	78.80
2813	63.82	59.01	49.49	78.89
3750	63.86	59.05	49.52	78.94
4688	63.88	59.07	49.53	78.97
5625	63.89	59.08	49.54	78.98
6563	63.90	59.09	49.55	78.99
7500	63.92	59.11	49.56	79.02
8438	63.95	59.13	49.58	79.05
9375	63.99	59.17	49.62	79.10
10313	64.05	59.22	49.66	79.17
11250	64.10	59.27	49.69	79.23
12188	64.14	59.31	49.73	79.29
13125	64.19	59.35	49.77	79.34
14063	64.23	59.39	49.80	79.39
15000	64.27	59.42	49.82	79.44

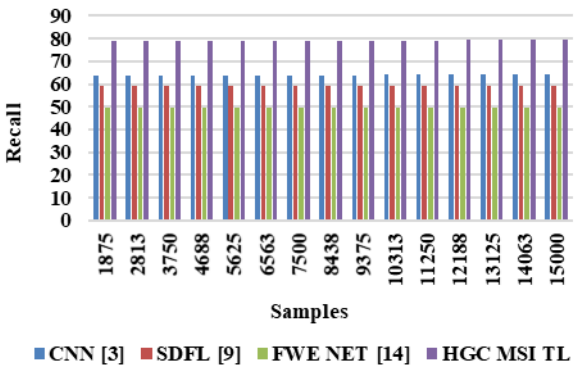

Figure. 7 Average recall of classification for different image types

The results of this assessment and the comparison shown in Fig. 7 show that the suggested model is capable of obtaining a recall that is 15.5 percent higher than CNN [3], 19.4 percent higher than SDFL [9], and 25.9 percent higher than FWE NET [14] for a variety of test imagesizes.

This is because GWO was combined with multiple feature sets & CNN, which helps improve search performance across a variety of use cases. Evaluations for the classification delay that are comparable to these are observed in Table 4 as follows,

Table 4. Average delay of classification for different image types

TIS	D (ms) CNN [3]	D (ms) SDFL [9]	D (ms) FWE NET [14]	D (ms) HGC MSI TL
1875	161.24	126.95	131.05	74.74
2813	161.42	127.09	131.20	74.83
3750	161.52	127.17	131.28	74.87
4688	161.58	127.21	131.33	74.90
5625	161.60	127.23	131.35	74.91
6563	161.63	127.26	131.37	74.92
7500	161.68	127.29	131.41	74.95
8438	161.75	127.35	131.47	74.98
9375	161.85	127.43	131.55	75.03
10313	161.99	127.54	131.66	75.09
11250	162.11	127.63	131.76	75.15
12188	162.24	127.73	131.86	75.20
13125	162.35	127.82	131.96	75.26
14063	162.45	127.90	132.04	75.30
15000	162.54	127.97	132.11	75.34

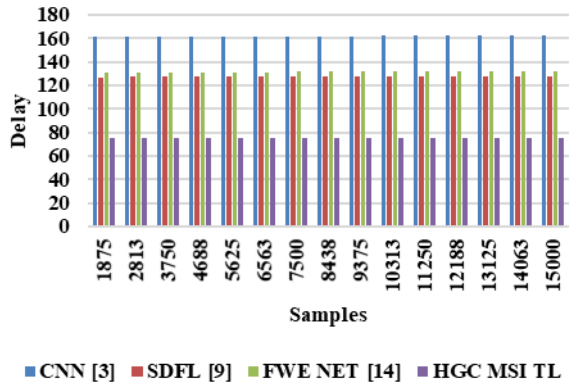

Figure. 8 Average Delay of Classification for different image types

The results of this assessment and the comparison shown in Fig. 8 show that the suggested model is capable of obtaining an accuracy that is 18.5 percent faster than CNN [3], 8.3 percent faster than SDFL [9], and 8.5 percent faster than FWE NET [14] for a variety of test image sizes. This is because GWO was combined with multiple feature sets & CNN, which helps improve search performance across a variety of use cases. Due to these enhancements the proposed model is capable of deployment for a wide variety of real-time Natural & Synthetic classification application scenarios.

5. Conclusion

The proposed model is useful for classification of multiple natural & synthetic image datasets, which is done via combination of multi-domain feature sets along with GWO & VGGNet based classification operations. The proposed model was observed to be 18.4 percent accurate than CNN [3], 25.5 percent accurate than SDFL [9], and 23.9 percent accurate than FWE NET [14], while 14.5 percent precise than CNN [3], 18.5 percent precise than SDFL [9], and 23.4 percent precise than FWE NET [14], it was also observed to be having 15.5 percent higher recall than CNN [3], 19.4 percent higher recallthanSDFL [9], and 25.9 percent higher recallthanFWE NET [14] for a variety of test image sizes. This is because GWO was combined with multiple feature sets & CNN, which helps improve search performance across a variety of use cases. The model was observed to be 18.5 percent faster than CNN [3], 8.3 percent faster than SDFL [9], and 8.5 percent faster than FWE NET [14] for a variety of test image sizes. This is due to use of GWO which assists in reducing number of features used for classification process. Due to these enhancements the proposed model is capable of deployment for a wide variety of real-time Natural & Synthetic classification application scenarios. In future, the proposed model must be validated on larger datasets, and can be extended via integration of deep learning models like auto encoders, gated recurrent units (GRUs), and recurrent neural networks (RNNs) for better performance under multiple scenarios. Moreover, this performance can be further enhanced via use of hybrid bio-inspired models which will further assist in improving their classification performance under multiple use cases.

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

The paper conceptualization, NMY and NNK; methodology, NMY, NS and NNK; software, NMY; validation, NMY, NS, and NNK; formal analysis, NMY and NNK; investigation, NNK; resources, NMY; data curation, NMY; writing—original draft preparation, NMY and NNK; writing—review and editing, NMY and NNK; visualization, NMY; supervision, NMY and NS; project administration, NS and NNK.

References

[1] Y. Zhang, F. Liu, L. Jiao, S. Yang, L. Li, and M. Yang, “Discriminative Sketch Topic Model With Structural Constraint for SAR Image Classification”, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 13, pp. 5730-5745, 2020.

[2] H. Zhu, R. Leung, and M. Hong, "Shadow Compensation for Synthetic Aperture Radar Target Classification by Dual Parallel Generative Adversarial Network", *IEEE Sensors Letters*, Vol. 4, No. 8, pp. 1-4, 2020.

[3] H. Bi, J. Deng, T. Yang, J. Wang, and L. Wang, "CNN-Based Target Detection and Classification When Sparse SAR Image Dataset is Available", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 14, pp. 6815-6826, 2021.

[4] H. Zhu, W. Wang, and R. Leung, "SAR Target Classification Based on Radar Image Luminance Analysis by Deep Learning", *IEEE Sensors Letters*, Vol. 4, No. 3, pp. 1-4, 2020.

[5] Q. Wu, B. Hou, Z. Wen, Z. Ren, and L. Jiao, "Cost-Sensitive Latent Space Learning for Imbalanced PolSAR Image Classification", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 59, No. 6, pp. 4802-4817, 2021.

[6] F. E. A. Nogueira, R. C. P. Marques, and F. N. S. Medeiros, "SAR Image Segmentation Based on Unsupervised Classification of Log-Cumulants Estimates", *IEEE Geoscience and Remote Sensing Letters*, Vol. 17, No. 7, pp. 1287-1289, 2020.

[7] X. Liu, Y. Wu, W. Liang, Y. Cao, and M. Li, "High Resolution SAR Image Classification Using Global-Local Network Structure Based on Vision Transformer and CNN", *IEEE Geoscience and Remote Sensing Letters*, Vol. 19, pp. 1-5, 2022.

[8] J. Gu, L. Jiao, F. Liu, X. Zhang, X. Tang, and P. Chen, "Multi-Feature Weighted Sparse Graph for SAR Image Analysis", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 58, No. 2, pp. 881-891, 2020.

[9] D. Quan, H. Wei, S. Wang, R. Lei, B. Duan, Y. Li, B. Hou, and L. Jiao, "Self-Distillation Feature Learning Network for Optical and SAR Image Registration", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 60, pp. 1-18, 2022.

[10] P. Qian, Y. Chen, J. W. Kuo, Y. Zhang, Y. Jiang, K. Zhao, R. Helo, H. Friel, A. Baydoun, F. Zhou, J. Heo, N. Avril, K. Herrmann, R. Ellis, B. Traughber, R. S. Jones, S. Wang, K. Su, and R. F. Muzic, "mDixon-Based Synthetic CT Generation for PET Attenuation Correction on Abdomen and Pelvis Jointly Using Transfer Fuzzy Clustering and Active Learning-Based Classification", *IEEE Transactions on Medical Imaging*, Vol. 39, No. 4, pp. 819-832, 2020.

[11] Y. Guo, L. Jiao, R. Qu, Z. Sun, S. Wang, S. Wang, and F. Liu, "Adaptive Fuzzy Learning Superpixel Representation for PolSAR Image Classification", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 60, pp. 1-18, 2022.

[12] S. J. Liu, H. Luo, and Q. Shi, "Active Ensemble Deep Learning for Polarimetric Synthetic Aperture Radar Image Classification", *IEEE Geoscience and Remote Sensing Letters*, Vol. 18, No. 9, pp. 1580-1584, 2021.

[13] Z. Wu, B. Hou, and L. Jiao, "Multiscale CNN With Autoencoder Regularization Joint Contextual Attention Network for SAR Image Classification", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 59, No. 2, pp. 1200-1213, 2021.

[14] J. Wang, S. Wang, F. Wang, Y. Zhou, Z. Wang, J. Ji, Y. Xiong, and Q. Zhao, "FWENet: a deep convolutional neural network for flood water body extraction based on SAR images", *International Journal of Digital Earth*, Vol. 15, No. 1, pp. 345-361, 2022.

[15] B. Zou, X. Xu, and L. Zhang, "Object-Based Classification of PolSAR Images Based on Spatial and Semantic Features", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 13, pp. 609-619, 2020.

[16] R. W. Jansen, M. A. Sletten, T. L. Ainsworth, and R. G. Raj, "Multi-Channel Synthetic Aperture Radar Based Classification of Maritime Scenes", *IEEE Access*, Vol. 8, pp. 127440-127449, 2020.

[17] Z. Yue, F. Gao, Q. Xiong, J. Wang, A. Hussain, and H. Zhou, "A Novel Attention Fully Convolutional Network Method for Synthetic Aperture Radar Image Segmentation", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 13, pp. 4585-4598, 2020.

[18] W. Liang, Y. Wu, M. Li, and Y. Cao, "High-Resolution SAR Image Classification Using Context-Aware Encoder Network and Hybrid Conditional Random Field Model", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 58, No. 8, pp. 5317-5335, 2020.

[19] Y. Cao, Y. Wu, M. Li, W. Liang, and X. Hu, "DFAF-Net: A Dual-Frequency PolSAR Image Classification Network Based on Frequency-Aware Attention and Adaptive Feature Fusion", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 60, pp. 1-18, 2022.

[20] R. Gui, X. Xu, R. Yang, L. Wang, and F. Pu, "Statistical Scattering Component-Based Subspace Alignment for Unsupervised Cross-Domain PolSAR Image Classification", *IEEE*

Transactions on Geoscience and Remote Sensing, Vol. 59, No. 7, pp. 5449-5463, 2021.
[21] Q. Chen, W. Cao, J. Shang, J. Liu, and X. Liu, "Superpixel-Based Cropland Classification of SAR Image With Statistical Texture and Polarization Features", *IEEE Geoscience and Remote Sensing Letters*, Vol. 19, pp. 1-5, 2022.
[22] R. Yang, Z. Hu, Y. Liu, and Z. Xu, "A Novel Polarimetric SAR Classification Method Integrating Pixel-Based and Patch-Based Classification", *IEEE Geoscience and Remote Sensing Letters*, Vol. 17, No. 3, pp. 431-435, 2020.
[23] B. Hou, J. Guan, Q. Wu, and L. Jiao, "Semisupervised Classification of PolSAR Image Incorporating Labels' Semantic Priors", *IEEE Geoscience and Remote Sensing Letters*, Vol. 17, No. 10, pp. 1737-1741, 2020.
[24] Z. Ren, B. Hou, Q. Wu, Z. Wen, and L. Jiao, "A Distribution and Structure Match Generative Adversarial Network for SAR Image Classification", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 58, No. 6, pp. 3864-3880, 2020.
[25] W. Hua, C. Zhang, W. Xie, and X. Jin, "Polarimetric SAR Image Classification Based on Ensemble Dual-Branch CNN and Superpixel Algorithm", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 15, pp. 2759-2772, 2022.
[26] X. Li, Q. Sun, L. Li, X. Liu, H. Liu, L. Jiao, and F. Liu, "SSCV-GANs: Semi-Supervised Complex-Valued GANs for PolSAR Image Classification", *IEEE Access*, Vol. 8, pp. 146560-146576, 2020.
[27] R. Cao, Y. Wang, B. Zhao, and X. Lu, "Ship Target Imaging in Airborne SAR System Based on Automatic Image Segmentation and ISAR Technique", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 14, pp. 1985-2000, 2021.
[28] X. Zhao, J. Wu, H. Wang, X. Gao, and L. Zhao, "Injecting spectral indices to transferable convolutional neural network under imbalanced and noisy labels for Landsat image classification", *International Journal of Digital Earth2022*, Vol. 15, No. 1, pp. 437-462, 2022.
[29] X. Qian, F. Liu, L. Jiao, X. Zhang, Y. Guo, X. Liu, and Y. Cui, "Ridgelet-Nets With Speckle Reduction Regularization for SAR Image Scene Classification", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 59, No. 11, pp. 9290-9306, 2021.