INSTITUTE OF COMPUTER SCIENCE
UNIVERSITY OF THE PHILIPPINES LOS BAÑOS

Identification of GAN-Generated Images through Frequency Analysis and Machine Learning

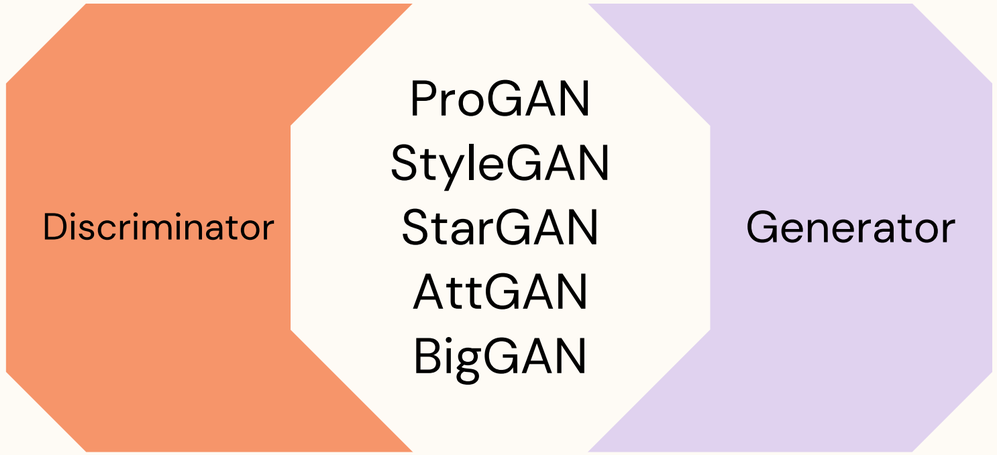
CMSC 190 SPECIAL PROBLEM

Yanna Denise A. Hilario
BS Computer Science

Rodolfo Camaclang III

Background of the Study

- Synthetic media generation through Generative Adversarial Networks (GAN)



- Adversarial Training
- Hyper-realistic images
- Computationally expensive to train

- Race for more advanced detectors to discriminate **fake** vs. **real**

GAN detection

Deep Learning and the use of CNNs

- High computational resources
- Complex Architectures

Concentrated on Fake Faces

- Diverse Fake Face Dataset
- FaceForensics++
- 140k Real and Fake Faces
- CelebA

Significance of the Study

Frequency Analysis

Discrete Cosine Transform

Discrete Wavelet Transform

+

Machine Learning

Support Vector Machine

GAN traces in different object classes

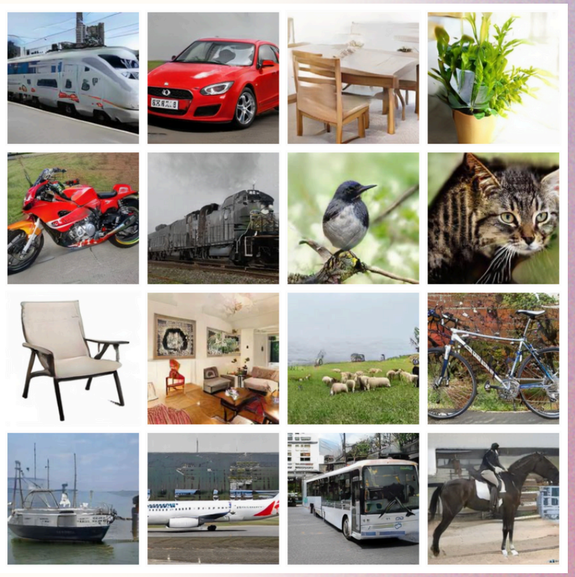
Reduced computational demands

Research Objectives

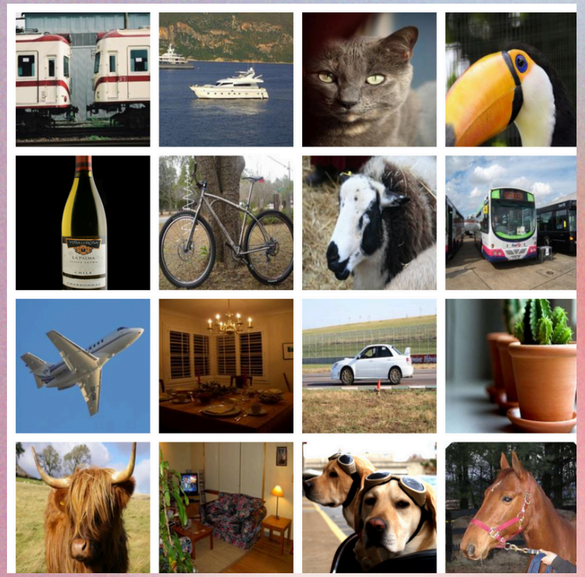
- Collect synthetic images from the ProGAN dataset.
- Apply DCT and DWT as pre-processing methods on the input images.
- Determine which spectral features can be used to classify fake from real images.
- Implement a Support Vector Machine classification model for this problem
- Develop a Graphical User Interface (GUI) for the classification problem
- Evaluate the accuracy of the classification model using the proposed methods.

Methodology

Data Gathering



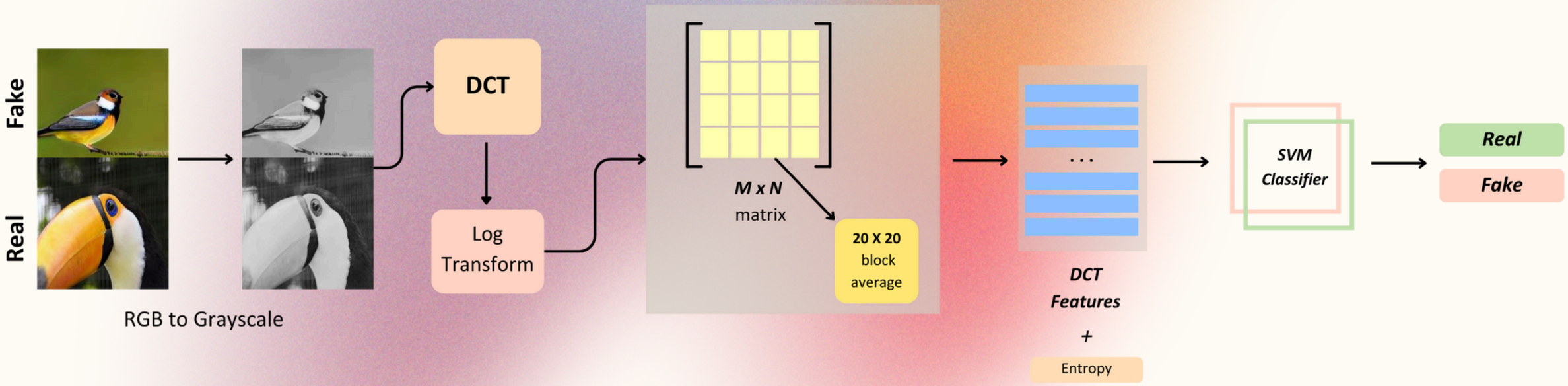
Fake Images from the ProGAN dataset



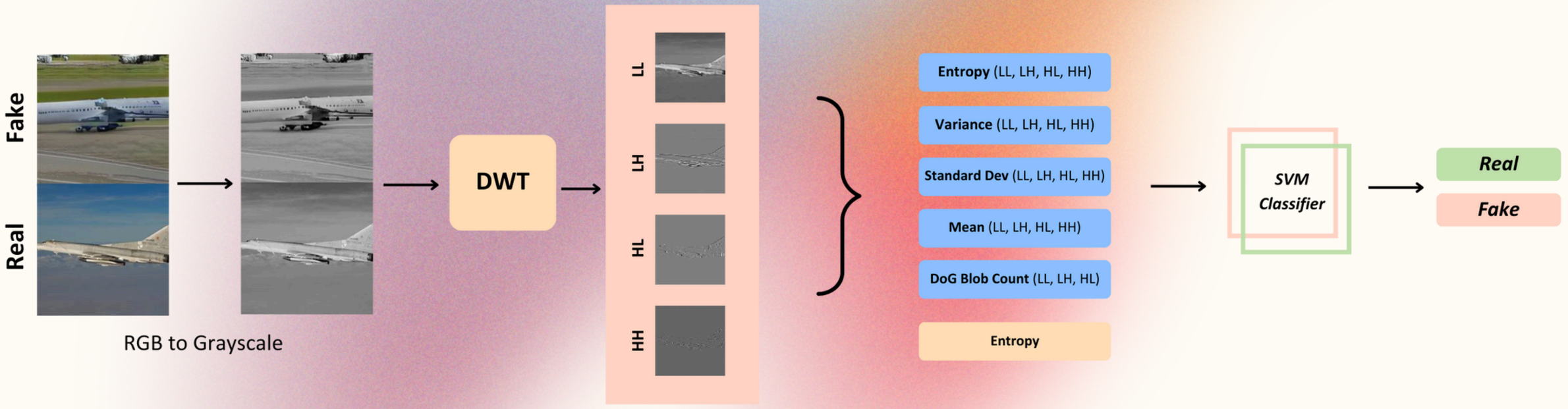
Real Images from the Pascal VOC dataset

- 20 object classes
- 4000 fake images
- 4000 real images
- 200 x 200 dimension
- 60% training
- 40% testing

DCT Feature Extraction



DWT Feature Extraction



Support Vector Machine Model

RBF Kernel

Grid Search

10-fold Cross Validation

Results and Discussion

Initial Testing Results

N	Accuracy	CV Mean Accuracy	C		Features
8	96.9063	96.975	10	0.001	625
10	96.9688	96.9875	10	0.001	400
20	**97.1875**	**97.2375**	**10**	**0.01**	**100**
25	96.8438	97.00	10	0.01	64
40	97.0938	96.825	100	0.01	25
50	96.2188	96.1	100	0.01	16
100	75.4375	75.125	1	1	4
200	60.25	60.3625	1	1	1

DWT Level	Accuracy	CV Mean Accuracy	Image Size	C	γ
1	**79.0313**	**68.8125**	**100 x 100**	**100**	**0.01**
2	66.0937	59.25	50 x 50	1	0.1
3	67.0313	65.5	25 x 25	100	0.01

Final Testing Results

Method	Accuracy	C	γ
DCT	97.0833	10	0.01
DWT	79.8333	100	0.01
DCT + DWT	**97.125**	**100**	**0.001**

Method	Class	Precision	Recall	F1-Score
DCT	Real	0.9716	0.9700	0.9708
	Fake	0.9700	0.9717	0.9709
DWT	Real	0.8003	0.7950	0.7977
	Fake	0.7964	0.8017	0.7990
DCT + DWT	Real	**0.9724**	**0.9700**	**0.9712**
	Fake	**0.9701**	**0.9725**	**0.9713**

Conclusion

- GAN-generated images have discernable differences in the frequency domain
- Generalizable features across various object classes
- DCT coefficient slices prove to be informative factors for GAN image classification

Future Works

- Improve DWT feature extraction
- Use datasets from other GAN architectures with various object classes
- Experiment on image dimensions
- Test for robustness against adversarial attacks

Application Demo

INSTITUTE OF COMPUTER SCIENCE
UNIVERSITY OF THE PHILIPPINES LOS BAÑOS

Identification of GAN-Generated Images through Frequency Analysis and Machine Learning

CMSC 190 SPECIAL PROBLEM

Yanna Denise A. Hilario
BS Computer Science

Rodolfo Camaclang III