Malware Classification through Computer Vision

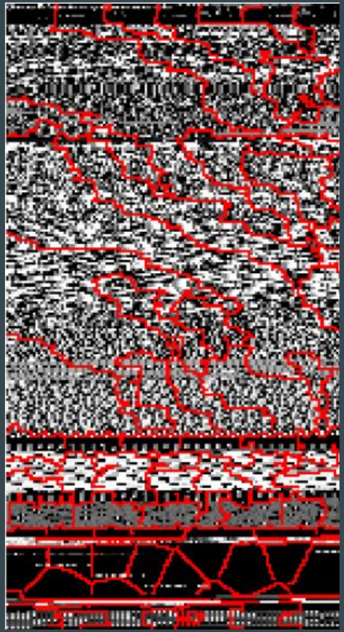
Prasann Viswanathan Iyer
190070047

Introduction

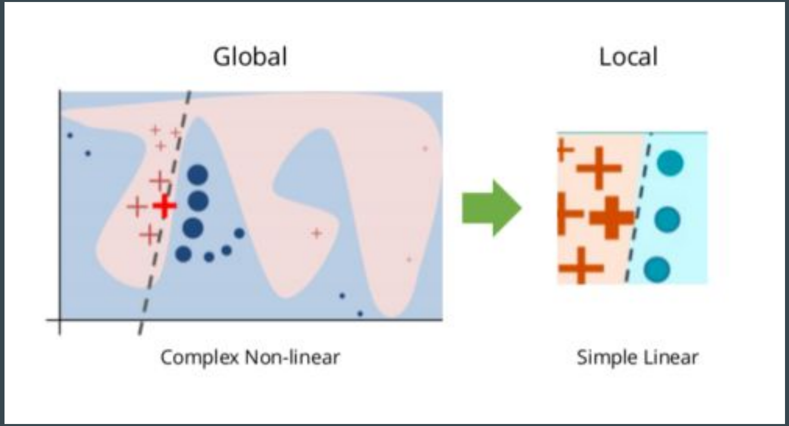
- Static Malware classification - without executing the programme
- Signature matching - Standard Method but with limitations
  - Obfuscations in code
  - Exponentially growing number of signatures
- Machine Learning Approaches are more robust and scalable
- Feature Construction issues
  - Disassembly step
  - Binning the types of function calls, counting loops
  - Large Feature space requires reduction with PCA or K-PCA
- Aim is to consider Malware Classification as a Computer Vision problem

Methodology

- Convert Byte Code of Malware to images
  - Map Binary Code to integers in [0,255]
  - Convert to 2D image with dimensions based on size of Malware file
- Apply Transfer Learning via existing Deep Image Neural Networks
  - Inception
  - ResNet
  - VGG
  - DenseNet
- Classic feature construction method and image method to Decision Forest
- Compare the classification performance



Evaluation and Interpretation



- Classification Metrics
  - Accuracy
  - True Positive Rate
  - False Positive Rate
  - F1 score
- Local Interpretable Model-Agnostic explanation algorithm
  - To understand the reason for our models prediction
- Conclude whether conversion to images assists in Malware Classification
  - With respect to both Classification Metrics and Model Performance Speed

Thank You