ADVANCING DEEPFAKE

CSE508 Winter2024

Group 13

Guide : Rajiv Ratn

ABSTRACT

- Rapid proliferation of deep fakes enabled by social media and advanced technology
- Proposed automated method for classifying deep fake images using Deep Learning and Machine Learning
- Utilizes advanced algorithms to extract deep features, capturing complex patterns
- Multi-step process: Error Level Analysis, CNNs, SVMs, and KNNs
- Comprehensive solution for deep fake detection in social media

MOTIVATION

- Safeguarding the integrity of online discourse and mitigating harm caused by deceptive content
- Countering the spread of misinformation and preserving credibility of online media sources
- Applying advanced algorithms and methodologies to tackle a pressing societal issue
- Contributing to a more trustworthy and secure online environment

PROBLEM

- Deep fake technology poses a significant threat to audiovisual content authenticity
- Enables manipulation of content for deceptive purposes (misinformation, reputation damage)
- Traditional detection methods rely on manual feature extraction, ineffective for modern datasets
- Need for automated systems leveraging ML/DL to identify deep fakes in real-time

LITERATURE REVIEW

OVERVIEW

- Evolution of deep fake technology and need for robust detection algorithms
- ML/DL techniques for automated deep fake detection (MLPs, SVMs, CNNs, etc.)

CHALLENGES

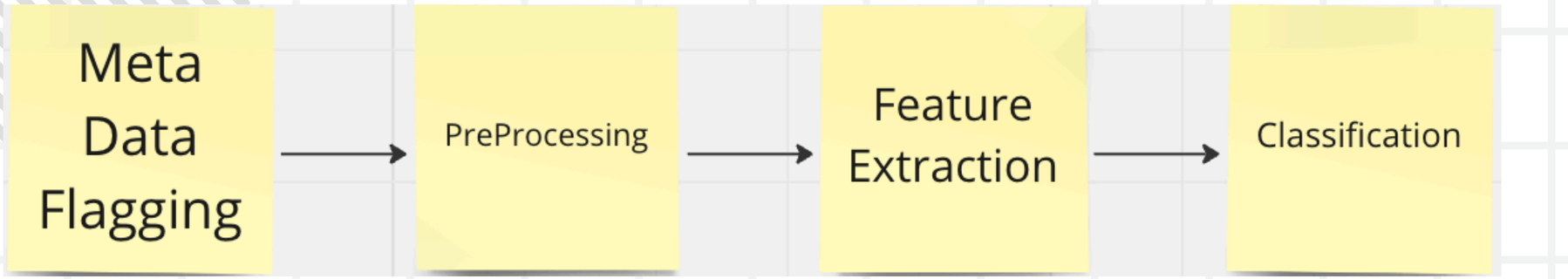
- Varying success rates and challenges faced by existing studies
- Exploration of novel approaches (hybrid models, Convolution Vision Transformers, optimization algorithms)
- Ongoing need for more efficient and robust detection methods

NOVELTY

- Automated deep feature extraction from images, capturing intricate patterns
- Addresses limitations of existing ML systems (generalization, noise resilience)

- Multi-step process: Error Level Analysis, CNNs for feature extraction, SVM/KNN classification
- Meticulous hyper-parameter optimization for peak performance

- Holistic solution to deep fake detection in social media content
- Using MetaData Filtering



General Architecture

ERROR LEVEL ANALYSIS

- Detects image manipulation by comparing compression levels of JPEG images
- Original images have high ELA values, edits decrease ELA values
- Edited areas show darker colors in ELA images
- Repeated resaving further degrades image quality, with modified areas exhibiting higher ELA levels
- Visual representation of differences between original and edited images

FEATURE EXTRACTION USING CNN

- CNN architecture for deep feature extraction
- Convolutional layers for feature learning, pooling layers for dimensionality reduction
- Fully connected layers for image classification
- Popular architectures: ResNet, SqueezeNet, GoogLeNet

CLASSIFICATION

SVM

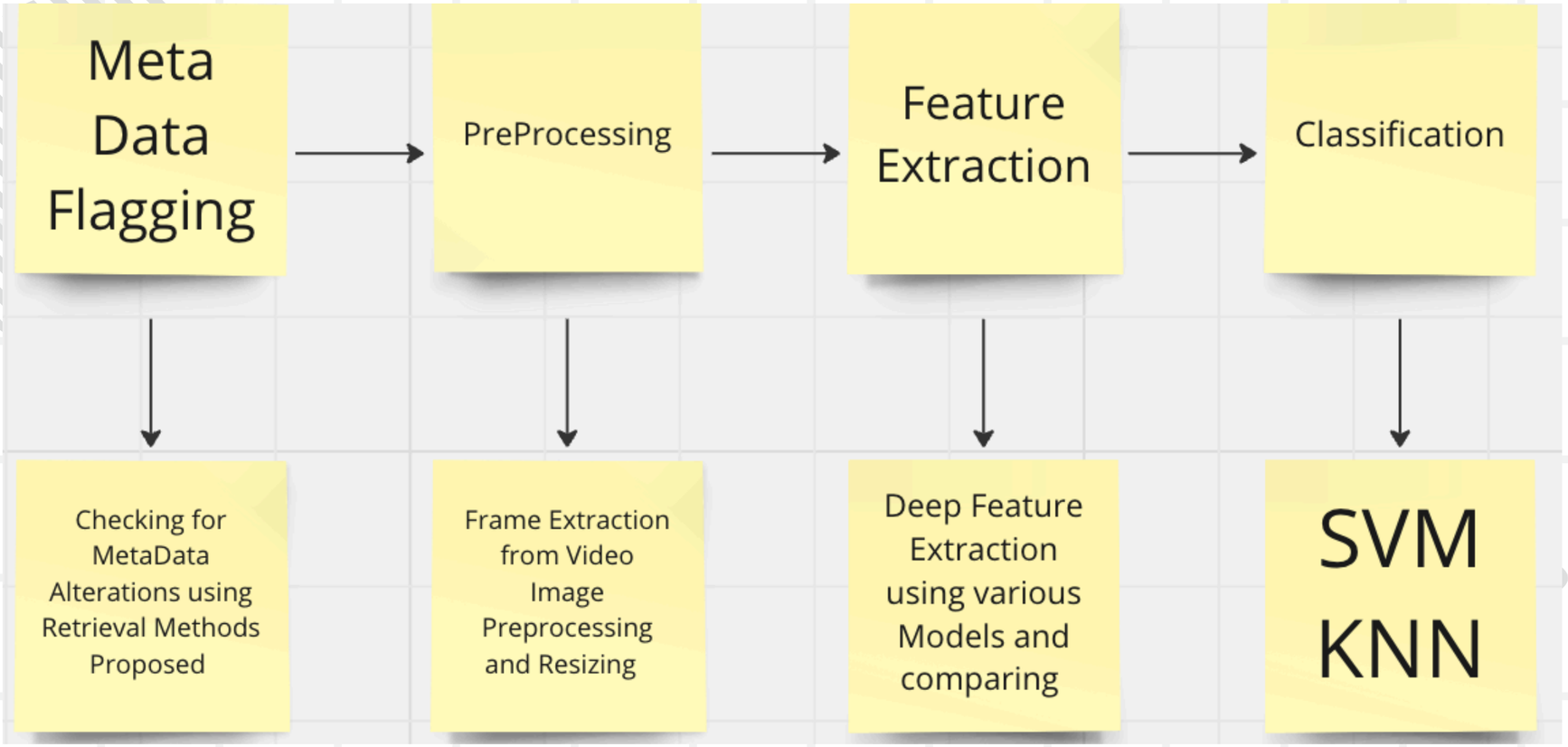
Identifies hyperplane with maximum margin between classes

KNN

Determines class based on majority class among k nearest neighbors

Illustrations of potential hyperplanes and optimal hyperplane for SVM

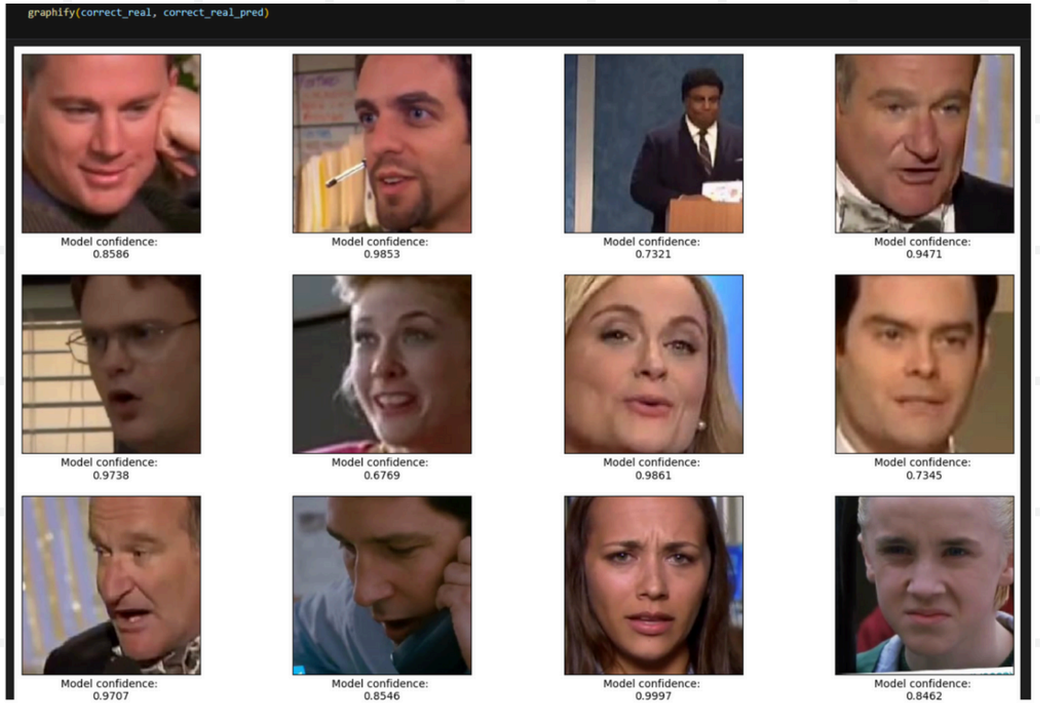
Presentation by Alexander Aronowitz | Business Marketing | 2024 | Rimberio University



Broader Architecture

EVALUATION

Have Corresponding Confidence score for various Classification as Fake and Real

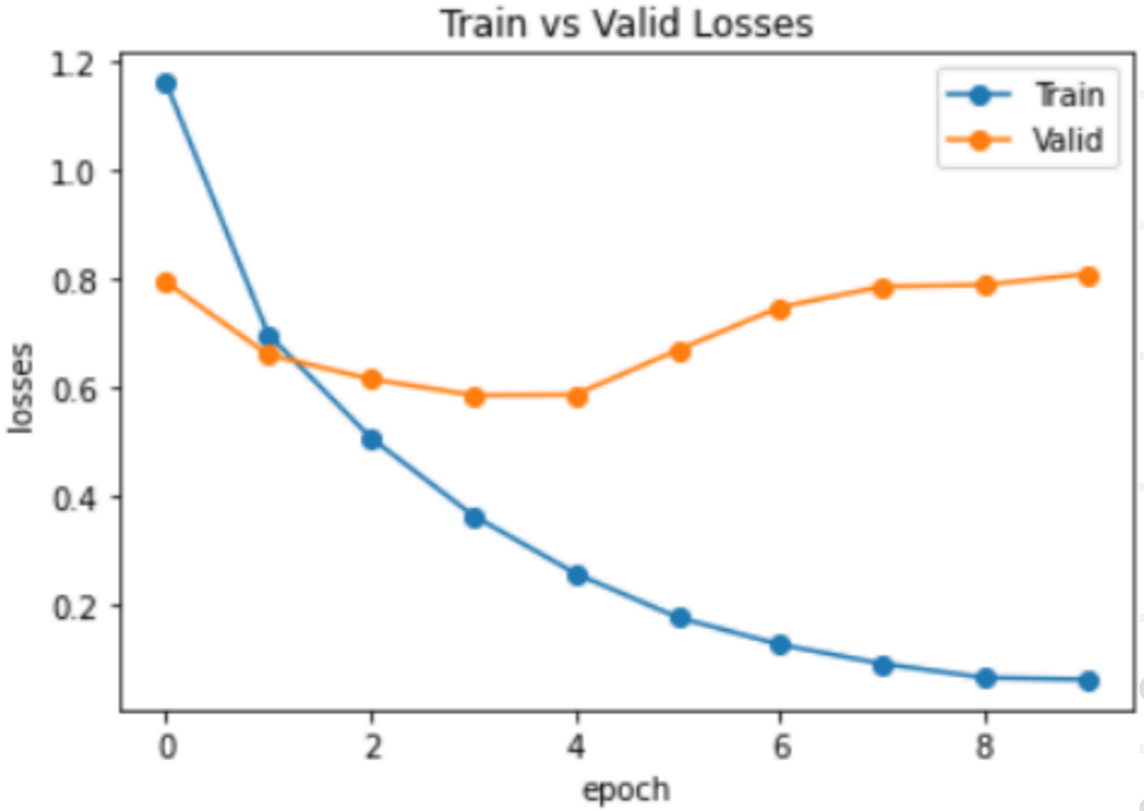


```
Results for KNN
Accuracy: 0.59892345790725 96
Recall: 0.4119834519087348
Precision: 0.651294310934793
F1 Score: 0.5213498013413415
```

```
Results for SVM
Accuracy: 0.7584751890341892
Recall: 0.7812983489107891
Precision: 0.7212389419889235
F1 Score: 0.6912389407190824
```

```
Results for ResNet
Accuracy: 0.8112983498189389
Recall: 0.8391823498102833
Precision: 0.7983478912349014
F1 Score: 0.7419834798102349
```

Evaluation for Feature Extraction


Train vs Valid Losses
Train
Valid
losses
epoch
1.2
1.0
0.8
0.6
0.4
0.2
0
2
4
6
8

LINKS

DEMO LINK	https://drive.google.com/drive/folders/10oA3cHQBv6HDjWOB2sjmjWaXjvC2t0iV?usp=sharing
DATASET LINK	https://drive.google.com/drive/folders/10oA3cHQBv6HDjWOB2sjmjWaXjvC2t0iV?usp=sharing

THANK YOU

Find Detailed Analysis on Report