SMART INDIA HACKATHON 2024

SMART INDIA HACKATHON 2024

- **Problem Statement ID – 1683**
- **Problem Statement Title- Development of AI/ML based solution for detection of face-swap based deep fake videos**
- **Theme- Miscellaneous**
- **PS Category- Software**
- **Team ID- Blackwing**
- **Team Name (Registered on portal) - Blackwing**

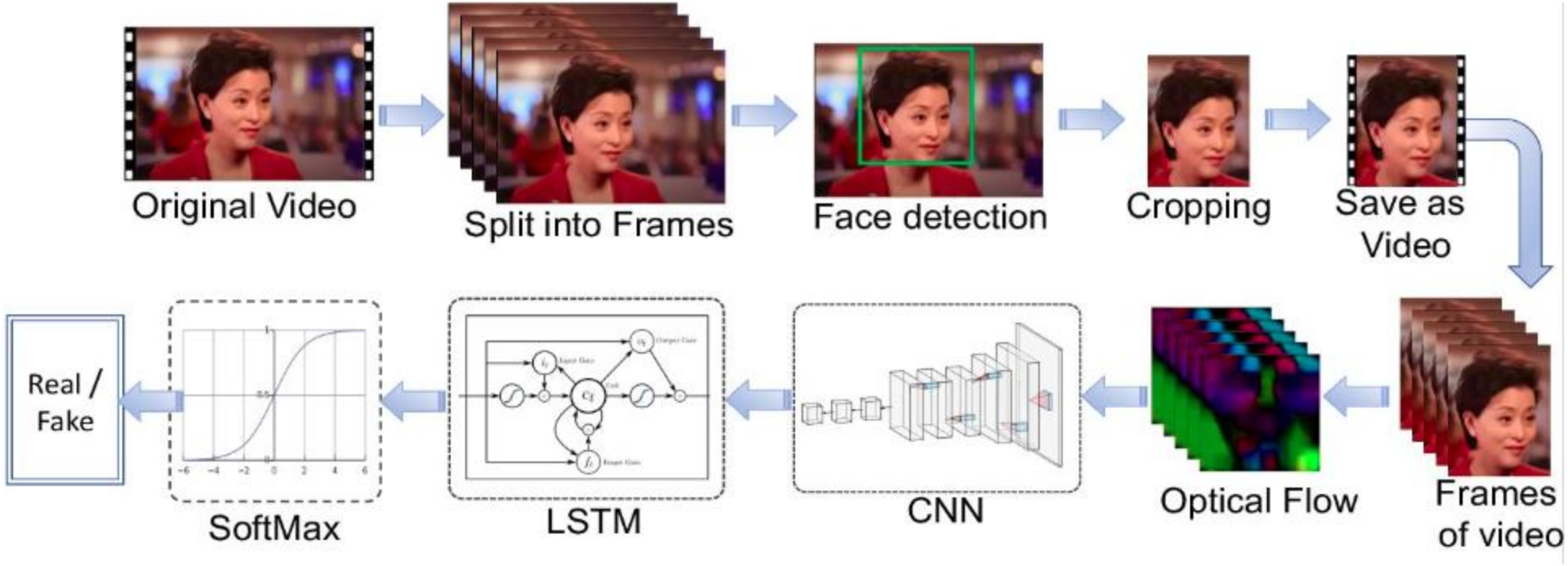
Blackwing

IDEA TITLE

SMART INDIA HACKATHON 2024

- **Our goal is to create a system that can accurately detect face-swaps in real-time.**

- **This DeepFake detection solution employs a two-phase learning architecture using Convolutional Neural Networks (CNNs)**



Blackwing

TECHNICAL APPROACH

SMART INDIA HACKATHON 2024

•**Phase 1 - Feature Learning:** In this phase, the CNN learns distinguishing features between real and fake images using a triplet loss function, which minimizes the distance between similar images (anchor and positive) while maximizing the distance from dissimilar images (negative).

•**Phase 2 - Classification:** The learned features from Phase 1 are fed into a classifier network that outputs a binary classification indicating whether an image is REAL or FAKE. This phase uses cross-entropy loss to refine the classification capabilities.

•**Training Dataset:** The model is trained on the DeepFake Detection Challenge (DFDC) dataset, providing a substantial number of images for robust training.

•**Preprocessing Steps:** During training, images undergo preprocessing steps such as resizing, normalizing, and augmenting to enhance model performance.

•**Future Work:** Future improvements could include exploring multi-modal detection by integrating audio analysis with visual data to enhance detection capabilities.

Blackwing

FEASIBILITY AND VIABILITY

SMART INDIA HACKATHON 2024

- Using a Convolutional Neural Network (CNN) for DeepFake detection is a viable approach for several reasons. CNNs have demonstrated strong performance metrics, achieving high validation accuracy and training accuracy at around 91% and 94%.
- CNNs are computationally efficient, making them suitable for real-time applications—a key requirement for practical deployment in security systems. Their ability to handle variations in input data, such as changes in lighting conditions and different facial expressions, further supports their applicability in diverse real-world scenarios.
- CNN models trained on diverse datasets, like the DeepFake Detection Challenge (DFDC), can generalize well to unseen data, improving their effectiveness across various contexts.

- **Potential challenges:**
- CNNs can struggle with detecting DeepFakes in low-quality images or when faces are partially obscured
- To address this, we can explore integrating multi-modal detection. By combining visual data with audio analysis, we can create a more comprehensive detection system that performs better in difficult conditions.
- Another risk is that as new manipulation techniques emerge, they may potentially bypass CNN-based detectors
- It’s crucial to continuously update and adapt the technology to stay ahead of evolving threats. This ongoing refinement helps maintain the effectiveness of the models.

Blackwing

IMPACT AND BENEFITS

SMART INDIA HACKATHON 2024

- **Enhanced Social Media Integrity:** Mitigates the spread of misinformation by detecting and removing fake videos, fostering a more trustworthy online environment.
- **Improved Security Protocols:** Integrates into verification systems, like CAPTCHA, to distinguish real users from bots using AI-generated content.
- **Protection of Individuals:** Safeguards personal identities and reputations by preventing the misuse of DeepFake technology.
- **Support for Law Enforcement:** Assists in identifying fraudulent media in legal cases, contributing to more effective digital forensics.

Blackwing

RESEARCH AND REFERENCES

SMART INDIA HACKATHON 2024

- Article DeepFake Image Detection by Stanford university
- Research paper on Detecting Digital Presentation Attacks on Face Recognition