24

Secured Text-Based CAPTCHA using Style Transfer Approach

PES2UG22CS366	Nishank Bhowal
PES2UG22CS630	Trisha Gupta
PES2UG22CS676	Yash Swarup
PES2UG22CS355	Nihal Satish

OVERVIEW

- Introduction
- Problem Statement
- Literature survey
- Dataset

INTRODUCTION

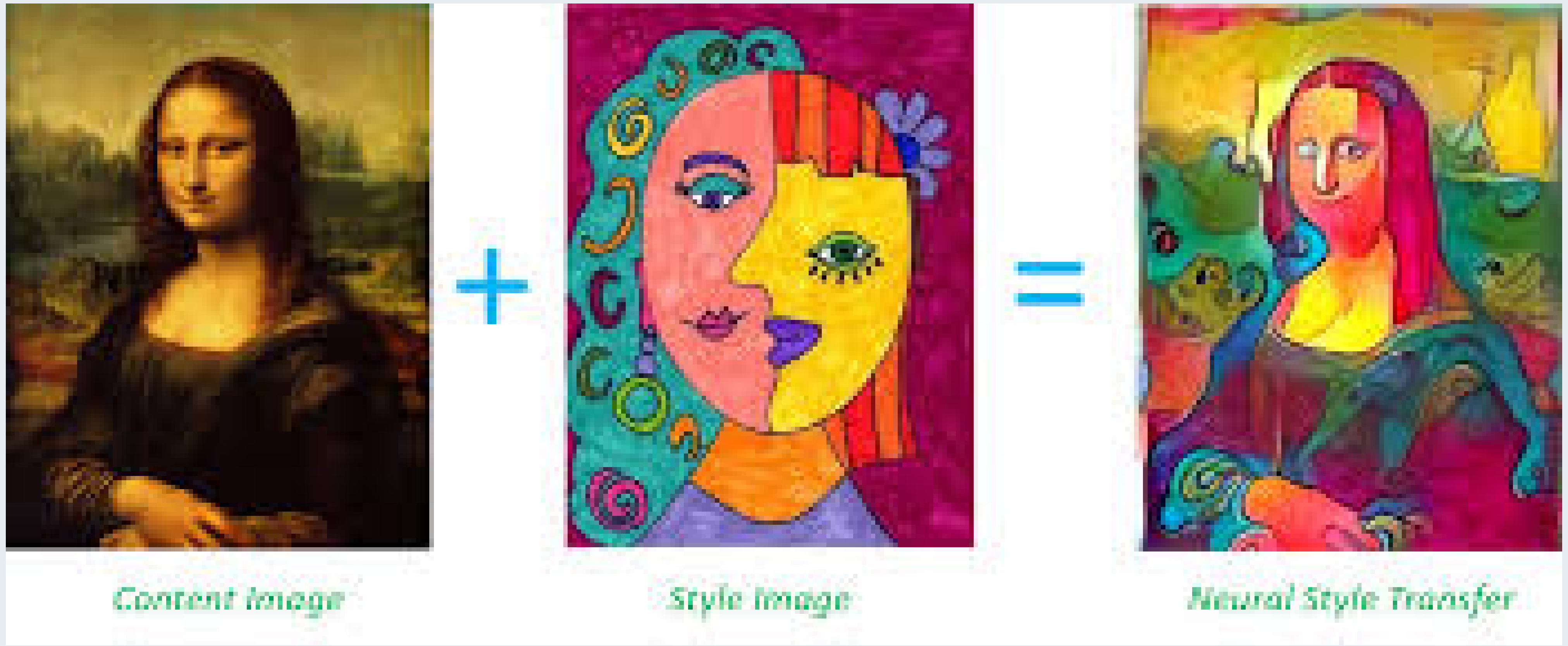
- CAPTCHA is a human-centred test to distinguish a human operator from bots, attacking programs, or any other computerised agent that tries to imitate human intelligence.
- Text-based CAPTCHAs are one of the most widely used security mechanisms.

Challenges with Traditional Text-Based CAPTCHAs

- Easily cracked by deep learning models such as CNNs, OCR (Optical Character Recognition), and GAN-based attacks.
- Adding too much distortion makes CAPTCHAs difficult for humans to read, reducing usability.
- Static CAPTCHAs do not adapt to evolving AI-based solver

PROBLEM STATEMENT

- Proposed Solution: Style Transfer-Based CAPTCHA
- Applying Style Transfer using CNNs to generate CAPTCHAs with complex yet human-readable textures.
- These CAPTCHAs reduce AI attack success rates while ensuring accessibility for human users.



OBJECTIVE

- Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vivamus sed vestibulum nunc, eget aliquam felis. Sed nunc purus, accumsan sit amet dictum in, ornare in dui.

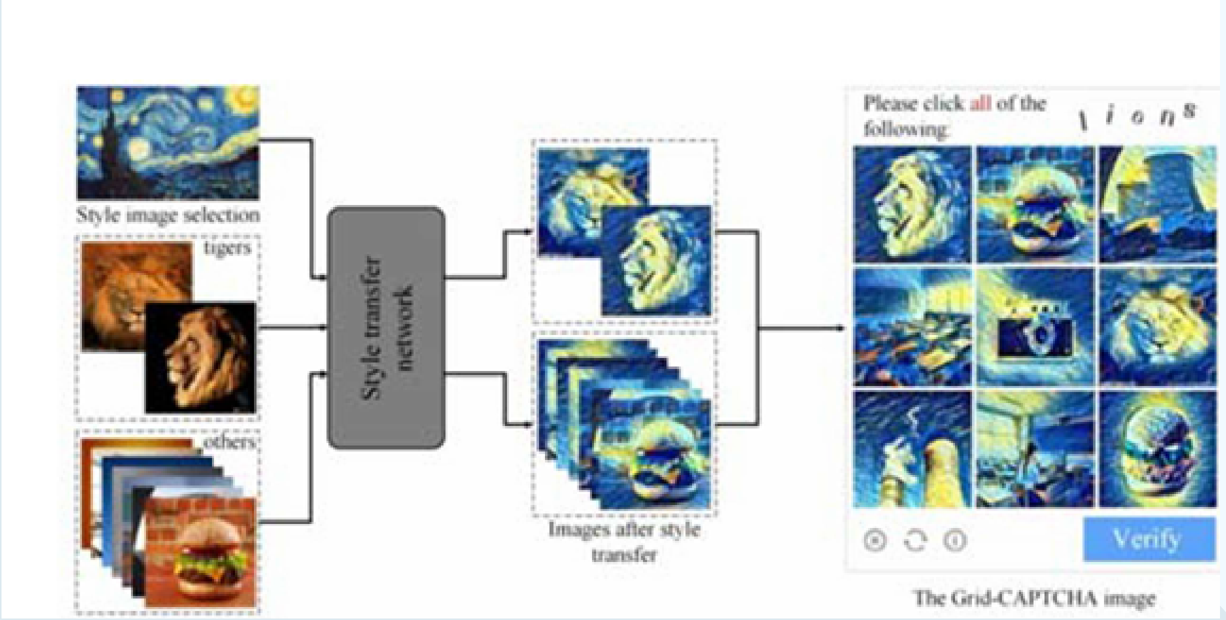
- Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vivamus sed vestibulum nunc, eget aliquam felis. Sed nunc purus, accumsan sit amet dictum in, ornare in dui.

1. IMAGE-BASED CAPTCHAS BASED ON NEURAL STYLE TRANSFER

Paper: IET Information Security, 2019 【9】

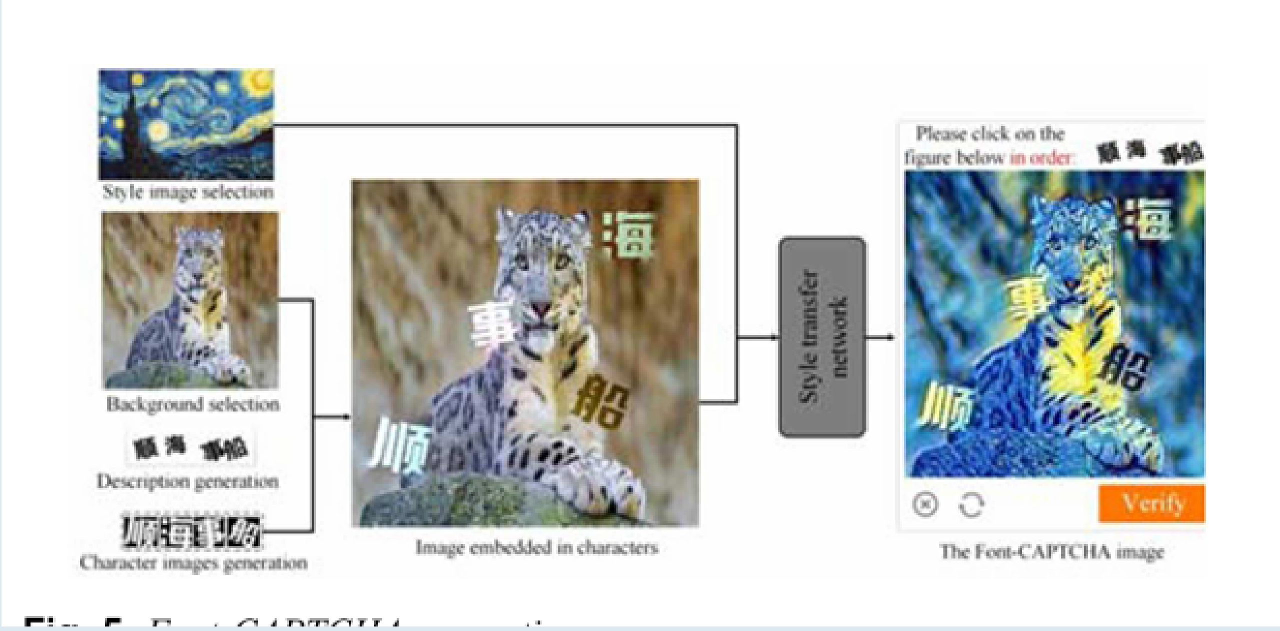
Summary:

This paper explores the use of **neural style transfer (NST) to generate image-based CAPTCHAs** that are difficult for automated solvers but easily recognizable by humans.



Key Implementation:

a. Grid-CAPTCHA
b. Font-CAPTCHA



- Style transfer model based on CNNs was used to generate complex distortions.
- Dataset: Six CAPTCHA datasets were used for evaluation.

1 . IMAGE-BASED CAPTCHAS BASED ON NEURAL STYLE TRANSFER

Results:

Grid-CAPTCHA human success rate: 75.04%

Font-CAPTCHA human success rate: 84.49%

Machine attack success rate: Significantly reduced, showing improved security.

Usability: Higher than traditional CAPTCHAs.

Drawbacks & Future Possibilities:

Limited to image-based CAPTCHAs;

text-based ones are not considered.

More complex NST models can be tested to further enhance security.

Future research could combine NST with adversarial training to resist evolving attacks.

2. REINFORCED PERTURBATION GENERATION FOR ADVERSARIAL TEXT-BASED CAPTCHA

Paper: 2024 IEEE International Conference on Computer Supported Cooperative Work in Design 【10】

Summary:
This paper introduces a Reinforced Perturbation Generation (RPG) framework that applies reinforcement learning to generate adversarial text-based CAPTCHAs. The goal is to create CAPTCHAs that remain human-friendly but highly resistant to automated attacks.

Key Implementation:

Perturbation Initialization (PI) generates an initial distortion on CAPTCHA images.

Perturbation Reinforcement (PR) optimizes distortions using Deep Q-Network (DQN) and Q-learning.

Reward function: Attack model success rates determine reinforcement learning updates.

Multiple perturbation methods used, including warping, blurring, and occlusion.

2. REINFORCED PERTURBATION GENERATION FOR ADVERSARIAL TEXT-BASED CAPTCHA

Results:

RPG-generated CAPTCHAs are more difficult for attack models to solve.

Extensive experiments on 8 datasets showed enhanced resistance to deep learning-based solvers.

Maintains user readability better than static perturbation methods.

Drawbacks & Future Possibilities:

Computationally expensive due to reinforcement learning.

Needs real-time adaptation to counter new CAPTCHA-breaking models.

Future work could integrate generative models (GANs) to create evolving CAPTCHAs.

3. SECURED TEXT-BASED CAPTCHA USING CUSTOMIZED CNN WITH STYLE TRANSFER AND GAN-BASED APPROACH

Paper: 2024 IEEE International Conference on Information and Communication Technology 【11】

Summary:

This study enhances text-based CAPTCHAs by leveraging GANs to create complex backgrounds. The approach aims to reduce machine recognition rates while keeping human usability high.

Key Implementation:

.GAN-based approach is used to create complex and diverse backgrounds.

Evaluation model: Attacker's recognition rate is tested with and without GAN-style transfer.

3. SECURED TEXT-BASED CAPTCHA USING CUSTOMIZED CNN WITH STYLE TRANSFER AND GAN-BASED APPROACH

Results:

Without style transfer: 98.68% CAPTCHA recognition rate

With GAN-style transfer: 2.1% recognition rate

Ensures text clarity for humans while making recognition difficult for AI.

Drawbacks & Future Possibilities:

Might reduce user readability under extreme distortions.

GANs are resource-intensive, requiring optimization.

Future improvements include adaptive CAPTCHAs that change dynamically based on attack success rates.

4. END-TO-END ATTACK ON TEXT-BASED CAPTCHAS USING CYCLE-GAN

Paper: Preprint (2020) 【18】

Summary:

This paper proposes an efficient attack on text-based CAPTCHAs using a Cycle-Consistent Generative Adversarial Network (Cycle-GAN). It reduces the need for labeled training data and increases attack transferability.

Key Implementation:

Cycle-GAN used to generate adversarial CAPTCHA samples from unlabeled real-world CAPTCHAs.

Convolutional Recurrent Neural Network (CRNN) for sequence recognition.

Active transfer learning to fine-tune the model with minimal labeled data.

4. END-TO-END ATTACK ON TEXT-BASED CAPTCHAS USING CYCLE-GAN

Results:

Successfully broke CAPTCHAs from 10 major websites,

Lower data labeling cost, making large-scale attacks easier.

Demonstrates severe vulnerabilities in existing CAPTCHA schemes.

Drawbacks & Future Possibilities:

Highlights security flaws but does not propose solutions.

Defenses could include dynamic CAPTCHAs or adversarial training.

Future research should explore defensive GANs that evolve against such attacks.

5. ADVERSARIAL CAPTCHAS

Paper: IEEE Transactions on Cybernetics, 2022 【33】

Summary:

This paper presents aCAPTCHA, a framework that generates adversarial CAPTCHAs to counter deep learning-based CAPTCHA attacks by introducing human-tolerable perturbations.

Key Implementation:

- Modular system (aCAPTCHA) includes:12 image preprocessing (IPP) techniques (blurring, noise filtering, binarization).
- Text & image-based CAPTCHA attack models (SVM, CNN, ResNet, VGG).
- Adversarial CAPTCHA generation modules injecting perturbations via FFT and CNNs.

- Frequency-domain perturbations ensure robustness against machine learning attacks.

5. ADVERSARIAL CAPTCHAS

Results:

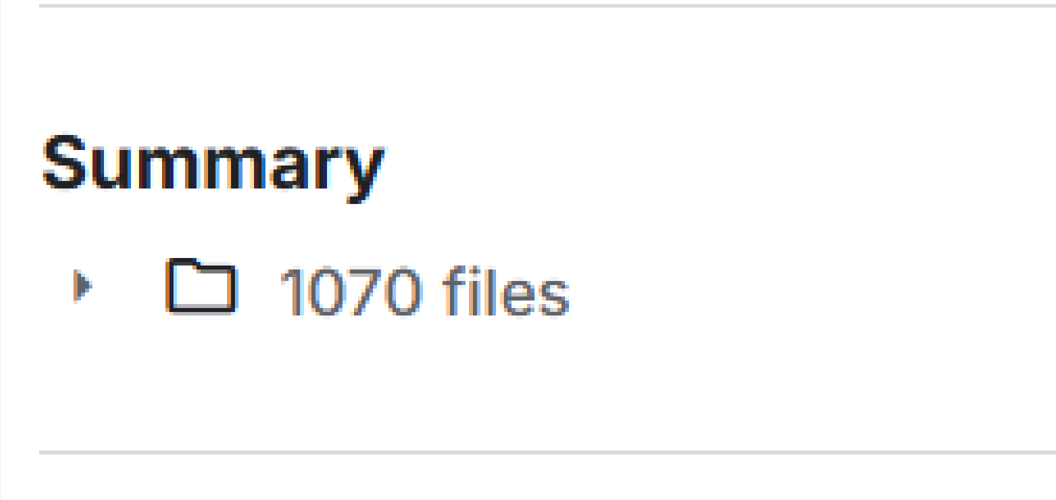
- Normal CAPTCHAs: High success attack rate (95.87% with LeNet).
- Adversarial CAPTCHAs: Attack success rate drops to near 0%.
- High transferability across different attack models and architectures.

Drawbacks & Future Possibilities:

- Security–Usability Trade-off: More perturbations may reduce readability.
- Computational Overhead: Frequency-domain modifications add processing costs.
- Future Work: Adaptive CAPTCHAs that dynamically adjust to real-time attack trends.

DATASET

The images are 5 letter words that can contain numbers. The images have had noise applied to them (blur and a line). They are 200 x 50 PNGs.



Acknowledgements

The dataset comes from [Wilhelmy, Rodrigo & Rosas, Horacio. (2013). captcha dataset.]

CONCLUSION

- Neural Style Transfer (NST) for Dynamic CAPTCHA Styles
- Applies artistic transformations to text-based CAPTCHAs while preserving readability
- Uses diverse style images to generate CAPTCHAs that vary across multiple artistic themes:
- Van Gogh's "Starry Night" – Swirling textures distort text structures.
- Cubism-Inspired Style – Geometric fragmentation disrupts pattern recognition.
- Watercolor & Abstract – Adds random texture effects to interfere with AI solvers.
- Pixelation & Mosaic Styles – Breaks letter continuity, making OCR recognition difficult.

- Generative Adversarial Networks (GANs) for CAPTCHA Enhancement

THANK YOU