#2022_Artificial Neural Network and Deep Learning_Term Project Final Presentation

Deep Experiments on Deepfake Detection

Team 1

21510004 이예빈

21510098 한보영

21512023 정주현

Contents

#1, Introduction

#2, Methodology

#3, Experimental Results

#4, Conclusion

Part 1

Introduction

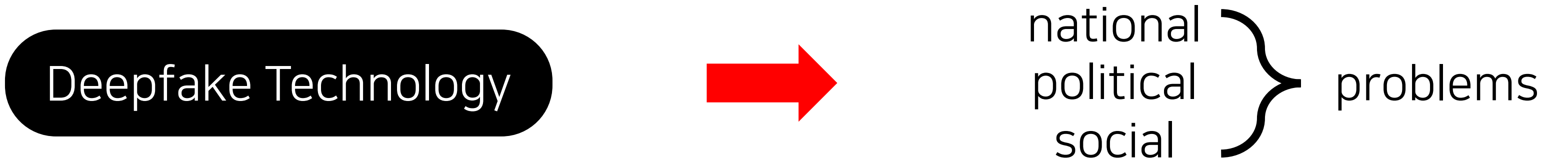
Background & Motivation

- Deepfake = 'deep learning' + 'fake'

⇒ technology that creates a manipulated image by superimposing another image on a video through deep learning [Chawal, 2019]

March 2022 - Ukraine Surrendered to Russia???

- A video clip of the Ukrainian president declaring his surrender to Russia was distributed, drawing much attention. [Wakefield, 2022]
- However, the video is fake video utilized Deepfake technology.



Deepfake Technology → national, political, social } problems

Introduction

Literature Review

Shad, H.S., et al. (2021). Comparative Analysis of Deepfake Image Detection Method Using Convolutional Neural Network. In *Computational Intelligence and Neuroscience*

- Dataset: 70,000 real images (Flickr dataset) + 70,000 fake images (generated by styleGAN)
- Models
  - ✓ DenseNet - DenseNet121, DenseNet169, DenseNet201
  - ✓ VGGNet - VGG16, VGG19
  - ✓ ResNet - ResNet50
  - ✓ VGGFace
  - ✓ custom CNN (build up by authors)
- Performance metrics: accuracy, precision, recall, F1-score, AUROC
- Results
  - VGGFace(99%) > ResNet50(97%) = DenseNet121(97%) > DenseNet201(96%) > DenseNet169(95%) > VGG19(94%) > VGG16(92%) > custom CNN(90%)

Part 1

Introduction

Purpose

- Hany Farid (Professor of California State University at Berkeley)
  - Ukrainian president surrender video = fake

    why?? ⇒ the resolution was lowered to hide the distortion caused by the manipulation process [Metz, 2022]
- investigate the effect of image change on Deepfake detection problem Deepfake detection performance
- original (color) images vs. images with color changes vs. images with changes in saturation

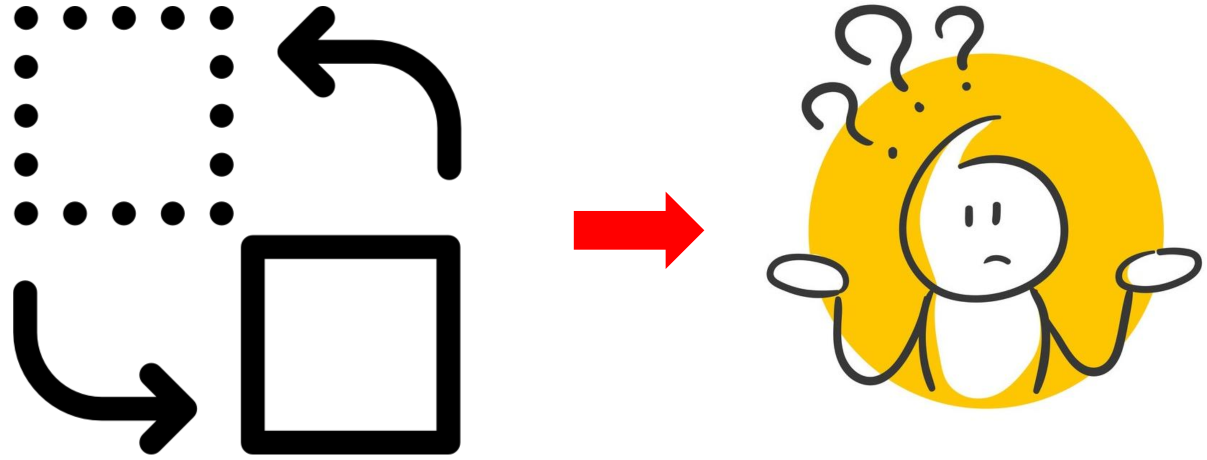


image transformation

Methodology

Data

- Deepfake Detection Challenge(DFDC) dataset from Kaggle [Dolhansky et al. 2019]
- 400 color videos ⇒ using only 20% of whole dataset (by sampling)
- Label: Real / Fake
- Preprocessing
  ① video → image
  ② image transformation
    ✓ RGB value transformation
      - make gray scale image
    ✓ HSV value transformation
      - make high saturation image

Part 2

Methodology

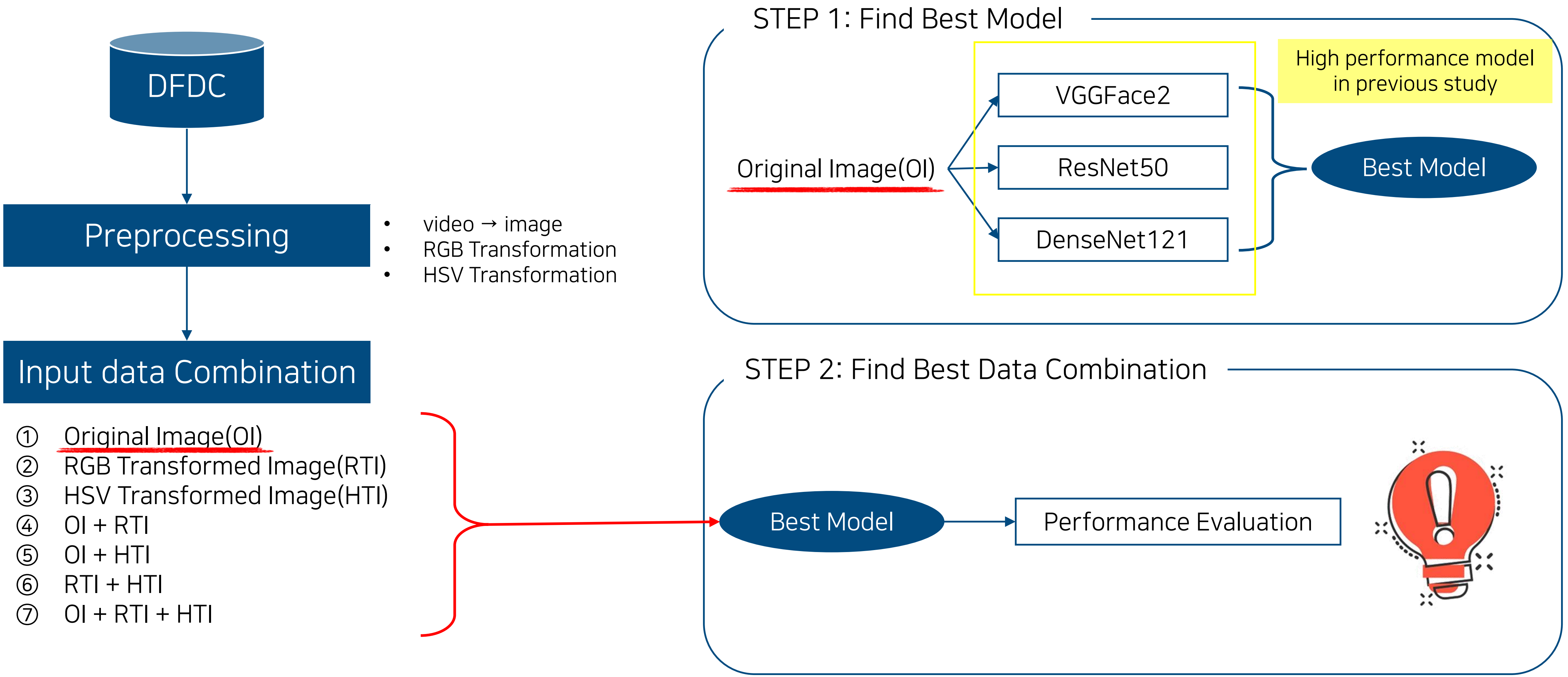
Data Examples

Real Image Example

Fake Image Example

Part 2

Methodology



DFDC

Preprocessing

- video → image
- RGB Transformation
- HSV Transformation

Input data Combination

① Original Image(OI)
② RGB Transformed Image(RTI)
③ HSV Transformed Image(HTI)
④ OI + RTI
⑤ OI + HTI
⑥ RTI + HTI
⑦ OI + RTI + HTI

STEP 1: Find Best Model

Original Image(OI)

VGGFace2

ResNet50

DenseNet121

High performance model in previous study

Best Model

STEP 2: Find Best Data Combination

Best Model

Performance Evaluation

Part 3

Experimental Results

Step1: Find Best Model

- Performance of models with original image(OI)

	VGGFace2	ResNet50	DenseNet121
Train Accuracy	0.9658	0.9863	0.9995
Test Accuracy	0.9714	0.7229	0.9983

- Train Accuracy → high in all models
- Test Accuracy → low in ResNet50, high in DenseNet121

► Using DenseNet121!!

Part 3

Experimental Results

Step2: Find Best Data Combination

- Performance of DenseNet121 with input data combination

	Label	Accuracy	Precision	Recall	F1-Score
OI	Fake(0)	0.9983	0.9984	0.9995	0.9989
	Real(1)		0.9981	0.9944	0.9963
RTI	Fake(0)	0.9842	0.9805	0.9995	0.9899
	Real(1)		0.9980	0.9315	0.9636
HTI	**Fake(0)**	**0.9998**	**1.0000**	**0.9997**	**0.9999**
	Real(1)		**0.9991**	**1.0000**	**0.9995**
OI + RTI	Fake(0)	0.9995	0.9996	0.9998	0.9997
	Real(1)		0.9994	0.9985	0.9990
OI + HTI	**Fake(0)**	**0.9997**	**1.0000**	**0.9996**	**0.9998**
	Real(1)		**0.9987**	**1.0000**	**0.9994**
RTI + HTI	**Fake(0)**	**0.9998**	**1.0000**	**0.9997**	**0.9998**
	Real(1)		**0.9989**	**1.0000**	**0.9994**
OI + RTI + HTI	**Fake(0)**	**0.9997**	**1.0000**	**0.9996**	**0.9998**
	Real(1)		**0.9987**	**1.0000**	**0.9993**

Using only 1 type of data

- best: model trained with HTI data
- In model trained with RTI data, performance has been slightly reduced.

Using data combination

- overall performance ↑ 🡐 data augmentation
- Especially, data combinations which contain HTI data have better performance.

► data augmentation → performance ↑

► increasing the image saturation → performance ↑

Part 4

Conclusion

Summary

- Deepfake video Ukrainian president's declaration of surrender

  ➔ it can be a threat to national security
- Comparison VGGFace2, ResNet50, and DenNet121 ➔ Best performance model : DensNet121
- Color changed data
  - Original Image
  - RGB Transformed Image
  - HSV Transformed Image
- Results
  - Data augmentation → performance ↑
  - HSV Transformed Image → performance ↑

Part 4

Conclusion

Future Research

- Larger training data
- Face-centered image data

Thank You