CAPSTONE 3

Synthetic images compared with real photographs

Eunpa Chae

Springboard 2020-2021

DIGITAL IMAGES ARE EASILY MODIFIABLE

- With everything and everyone online a web presence is increasingly vulnerable to interference from hackers, frenemies, competition, etc.
- This is of even greater importance when intruders deface images of professionals whose income is linked to keeping superiority of appearance intact as per original features, etc.
- Intellectual property analogy applies to faces of professionals whose career and reputation is mainly centered on attractive nature of visual presentation
- Several methods of modifying electronic images
  - Copy-Move forgery (copy and paste regions)
  - Copy-Move forgery with modification of copied region (blur, smear, etc.)
  - GANs (Generative Adversarial Networks)

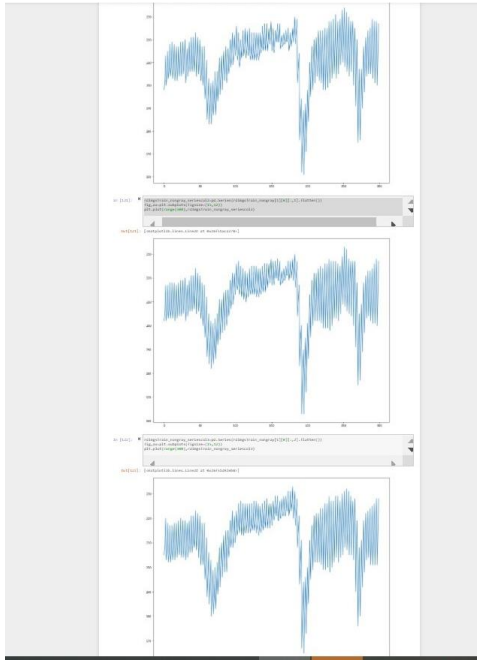
PREPROCESSING OF IMAGES

- Source of data https://www.kaggle.com/xhlulu/140k-real-and-fake-faces
- All 20200 images were standardized to 100x100 pixels in size
- Images were converted to grayscale via skimage's color.rgb2gray
- Datasets labeling each image with right classification were built and converted to numeric Python arrays then tensors via tensorflow.convert_to_tensor

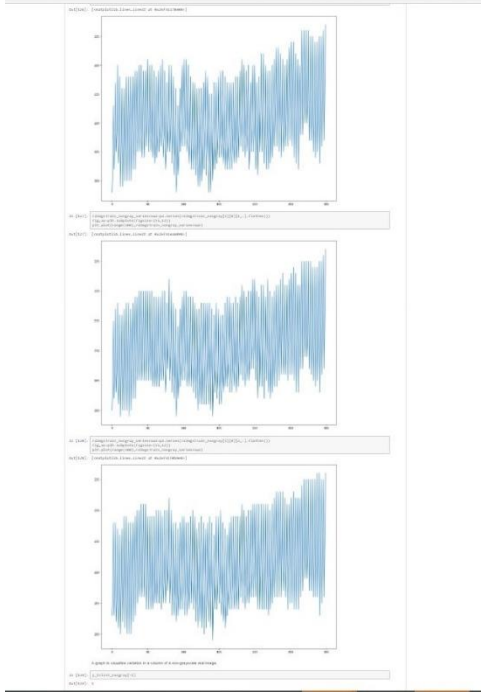
PIXEL VARIATION ANALYSIS

- Detection of images synthesized via GANs compared with real photographs requires identification of a pattern in the mosaic method with which GANs generates images.
- This might be possible by detection of unnatural or abrupt change in pixel variation at edges or other anomalies.
- A quick analysis of variation in pixels by row and column of a sample training image reveals there is no obvious change in pattern of variance between real and GANs images.

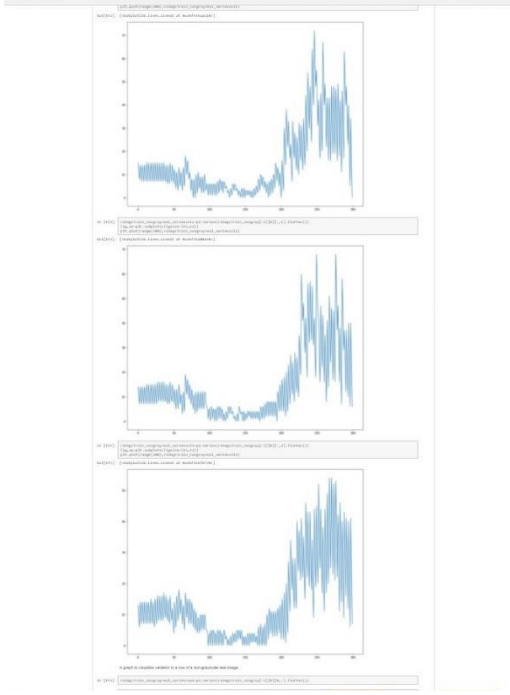
Pixel variation in three adjacent columns of nongrayscale GANs image



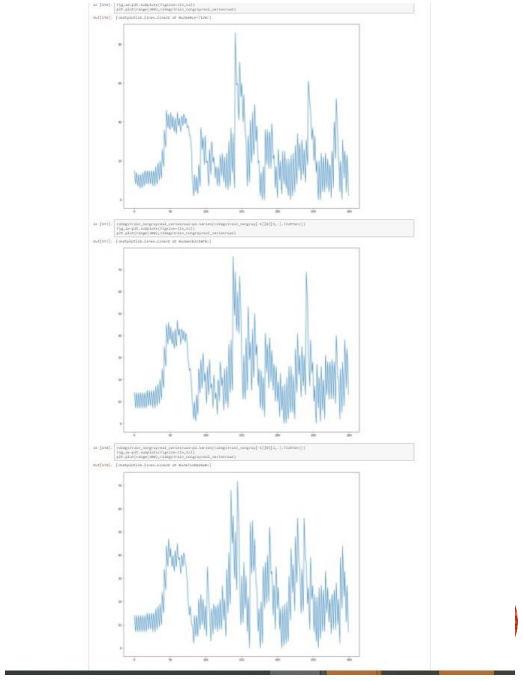
Pixel variation in three adjacent rows of nongrayscale GANs image



Pixel variation in three adjacent columns of nongrayscale real image



Pixel variation in three adjacent rows of nongrayscale real image



MODELS

- Keras neural network models were built with these specifications
  - Layer 1 rescales to standardize images
  - Layer 2 flattens the (100,100,3) structure of images
  - Layer 3 densely connected neural network layer with 128 nodes and relu activation function
  - Layer 4 densely connected neural network layer tapering to 64 nodes [variations with same 128 nodes and double 256 nodes] and relu activation function
  - Layer 5 output with 2 nodes, as this is binary classification, and softmax activation function

- Source of code https://keras.io/getting_started/intro_to_keras_for_engineers/

MODEL COMPILATION

- Adam optimizer
  - Stochastic gradient descent method with greater probability of finding global maximum
  - Applied with default learning rate of 0.001
- Sparse Categorical Cross Entropy loss function
  - Applied with >= 2 labels
  - Relevant with labels as integers
    - GANs images labeled with 0
    - Real images labeled with 1
- SparseCategoricalAccuracy metric
  - Relevant to Sparse Categorical Cross Entropy loss function
  - Calculates percentage of accurately labeled validation images

MODEL TRAINING

This Keras neural network model was trained on 2000 standardized grayscale images and 200 validation test images.

After some experimentation with number of epochs 300 was chosen to prevent overfitting while keeping a reasonable training accuracy level ~85%

fitting models							
model type	epochs (iter'n)	node structure	tr accuracy	valid'n accuracy	# images	batch_size	grayscale
model_g	1	same	57.55	-	2000	2000	y
model_g2	1	taper	50.25	-	2000	2000	y
model_g3	1	double	50.25	-	2000	2000	y
model_c	1	same	50	-	2000	2000	n
model_c2	1	taper	53.2	-	2000	2000	n
model_g	300	same	-	68	2000	2000	y
model_g2	12	taper	59.05	-	2000	2000	y
model_g2	100	taper	69.05	65.5	2000	2000	y
model_g2	300	taper	83.85	67	2000	2000	y
model_g2	500	taper	92.95	67.5	2000	2000	y
model_g2	800	taper	97.8	66	2000	2000	y
model_g2	1000	taper	99.65	65	2000	2000	y
model_c	100	same	64.8	64	2000	2000	n
model_c	300	same	75.95	65	2000	2000	n
model_c	800	same	93.8	65.5	2000	2000	n
model_c2	100	taper	67.75	62.5	2000	2000	n
model_c2	300	taper	69.25	63	2000	2000	n
model_c2	800	taper	91	65.5	2000	2000	n

OVERFITTING IN TRAINING OF MODELS

- Given enough iterations some models plateau on a level of accuracy while some models reach 100% training_accuracy
- In this instance 1000 epochs (iter'n) was sufficient to achieve perfect accuracy on training data
- The 'number of iterations' parameter in fitting of models was fine-tuned to 300 at which level training accuracy is a reasonable percentage in 80s while overfitting is minimized

fitting models							
model type	epochs (iter'n)	node structure	tr accuracy	valid'n accuracy	# images	batch_size	grayscale
model_g2	800	taper	97.8	66	2000	2000	y
model_g2	1000	taper	99.65	65	2000	2000	y

GENERALIZABILITY OF MODEL

- As validation_accuracy of model started at 63.5% right away on first iteration there is significant generalizability of model trained on color images without extensive processing

fitting models							
model type	epochs (iter'n)	node structure	tr accuracy	valid'n accuracy	# images	batch_size	grayscale
model_c	100	same	64.8	64	2000	2000	n

MODEL FITTING

- A Keras neural network fitted on 2000 grayscale training images and validated on 200 test images resulted in a maximum validation_accuracy of ~67.5%

- A Keras neural network fitted on 20000 training images and validated on 800 test images resulted in a maximum validation_accuracy of ~73.5%
- Increasing number of images seems to increase validation_accuracy!

fitting models							
model type	epochs (iter'n)	node structure	tr accuracy	valid'n accuracy	# images	batch	grayscale
model_c	800	same	82.25	73.5	20000	10000	n

FUTURE AVENUES OF RESEARCH

- Watermarking
- Convolutional Neural Networks
- Etc.

APPENDIX

- A presentation of additional image-processing analyses covering modification methods in addition to GANs
  - Copy-move-detection
  - Smearing/blurring

DETECTION OF COPY MOVE FORGERY

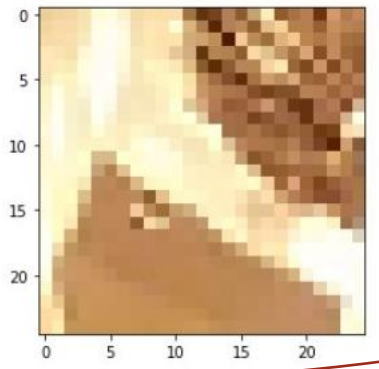
- This analysis covers detection of copy-move forgery
  - Anonymous nature of internet increases probability that intruders might modify electronic images via copying and pasting a region of image to mask or modify the appearance
  - Detectable with data science methodology
- Preliminary Correlation Analyses
  - Divide image into overlapping frames, as matrices, to compare suspected region of modification with original area
  - Compare each frame with original region to verify whether there is a match
  - If a match is found then it is possible that copy-move forgery occurred as probability of identical pixel variation in more than one region is very low

DESCRIBING IMAGES

- 20 images of size 20x20 pixels were modified via copy-move methods in Microsoft Paint application
- Resulting 40 images are the basis of this preliminary analysis
- Functions written in Python to identify copy-move forgery
  - Framed_ep
    - Takes in an image and specification of suspected modified area in pixels (frame-size)
    - Returns a list of indices defining all possible overlapping frames of this size as matrices
  - Identify_cm
    - Takes in an image, frame-size, and optional specification of accuracy level (default 95%)
    - Returns an array of tuples if there is a match on copy-move regions
      - Includes an element that quantifies distance between copied region and original source

GRAPHS ILLUSTRATING VERTICAL VARIATION OF PIXEL INTENSITY

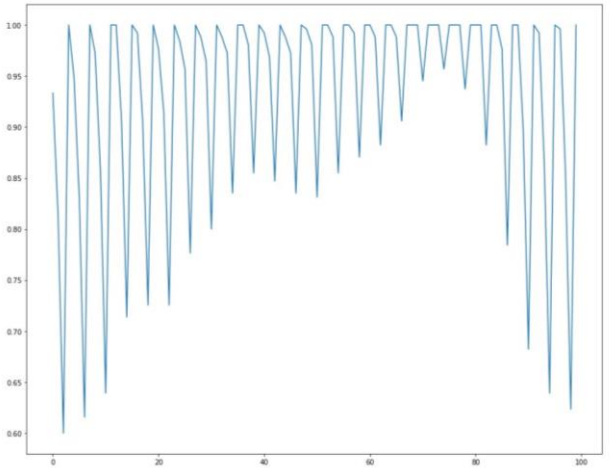
- There is great variability in pixel intensity vertically even in adjacent columns covering regions with little visible difference

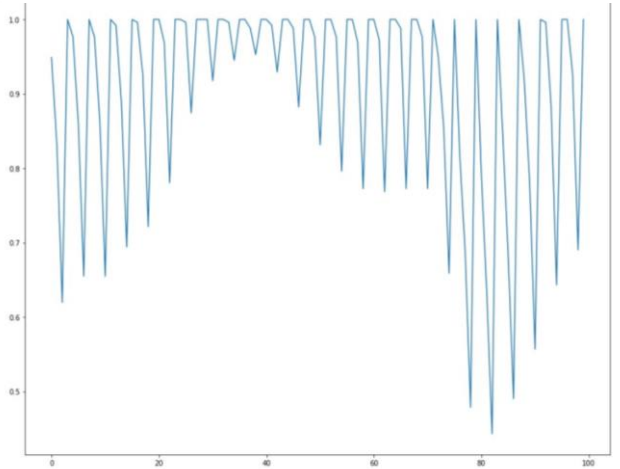


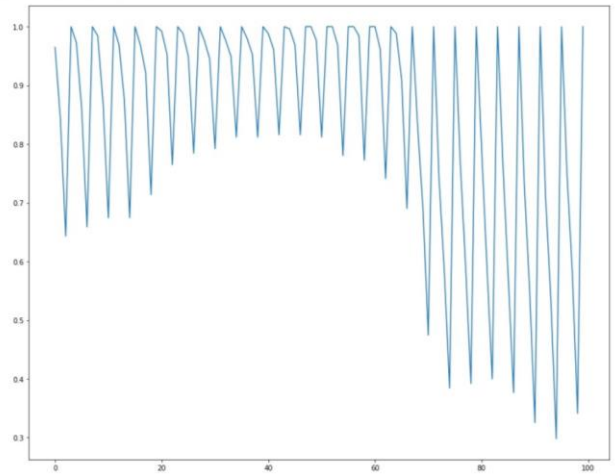
First column of pixels on left side

Second column of pixels on left side

Third column of pixels on left side







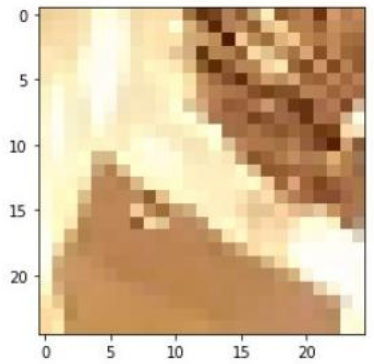
GRAPHS ILLUSTRATING HORIZONTAL VARIATION OF PIXEL INTENSITY

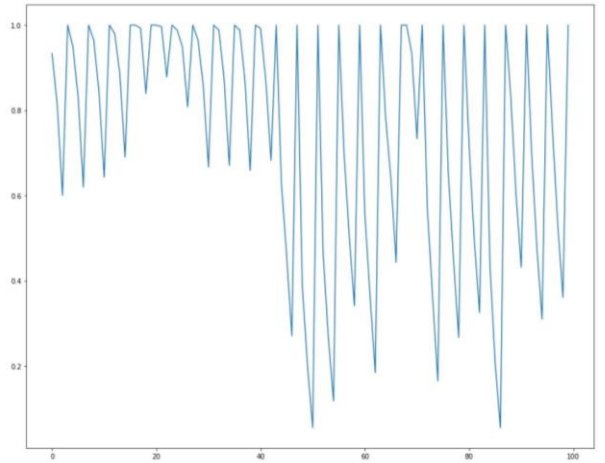
- There is variability in pixel intensity horizontally even in adjacent rows covering regions with little visible difference

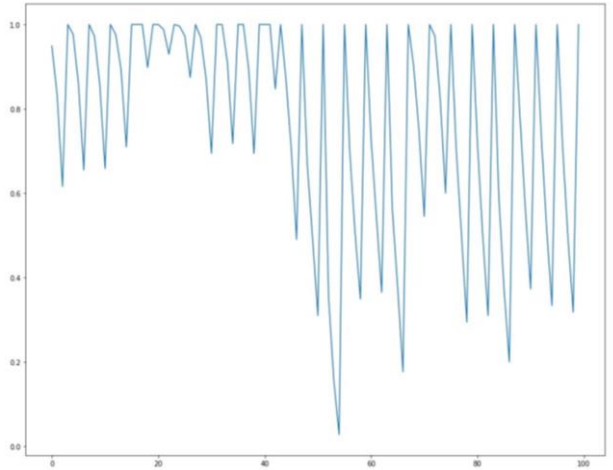
First row of pixels on top

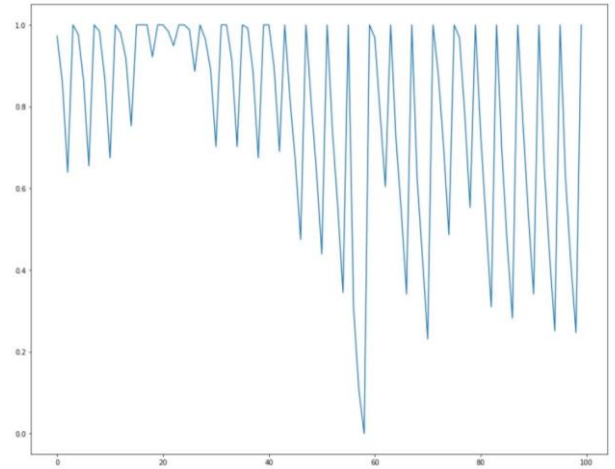
Second row of pixels at top

Third row of pixels at top









RESULTS OF COPY-MOVE DETECTION (CMD) VIA CORRELATION ANALYSES

Image	Copy-Move Forgery (Y/N)	Accuracy (default of 0.95 or specific level)	Results returned based on written Copy-Move detection functions
	Y	0.95 default	Identified greater than 100 regions of similarity at this accuracy level
	N	0.95 default	Identified greater than 100 regions of similarity at this accuracy level
	N	1.0 specified level	Empty array verifying original status of unmodified image
	Y	1.0 specified level	Returns one tuple correctly identifying one region of copy-move modification and distance from original area
1	N	1.0 specified level	Empty array verifying original status of unmodified image
2	Y	1.0 specified level	Returns one tuple correctly identifying one region of copy-move modification and distance from original area
…	CMD analyses between these rows omitted to streamline		
30	N	1.0 specified level	Empty array verifying original status of unmodified image
32	N	1.0 specified level	Empty array verifying original status of unmodified image

With personally written functions to detect copy-move forgeries on proprietary data accuracy is 100% at specified level of 1.0

SUBSEQUENT ANALYSES COVER COMPLICATED CMD

- Some types of copy-move-detection involve modification of copied area
  - Blurred edges of copied region
  - Lightened or darkened copied region to blend in with surrounding area, etc.
  - Smeared parts of image

- Neural networks were applied to detect more complicated copy-move forgeries

SUMMARY OF NEURAL NETWORKS ANALYSES

- 121 images of 10x10 pixel size were taken based on proprietary data
  - 71 original unmodified digital .png photographs
  - 50 digital .png photographs modified with one of following methods
    - Pasting a smeared 2x2 pixel region via 2017 Autodesk Inc. Sketchbook
    - Copying and pasting a region ranging between 2x2 and 3x5 via 2004 Microsoft Inc. Paint
    - Copying and pasting a rotated region ranging between 2x2 and 3x5 via 2004 MS Inc. Paint

- All images were randomly shuffled and analyzed via keras module in tensorflow
  - Preprocessing included separation of data into train and test data
    - test_size set at 0.18 to ensure a model trained on at least 100 images as each photo was 100x100 pixels
    - Neural Network model with 3-4 layers, 128 nodes, and epochs=1 gave training accuracy of 38.1%
    - Neural Network model with 3-4 layers, 100 nodes, and epochs=20 gave training accuracy of 60%
    - Neural Network model with 3-4 layers, 100 nodes, and epochs=20 gave prediction accuracy of 52.38%
    - Neural Network model with 3-4 layers, 100 nodes, and epochs=50 gave prediction accuracy of 52.38%

FUTURE AVENUES OF RESEARCH

- Watermarking
- Convolutional Neural Networks
- Etc.