IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

FACIAL MANIPULATION DETECTION USING U-NET

K.Varalakshmi[1] Dalish Princa William W[2,]

Assistant Professor[1] ,Student[2]

Department of Computer Science and Engineering

PERI INSTITUTE OF TECHNOLOGY

ABSTRACT

As advanced face synthesis and manipulation methods are made available, new types of fake face representations are being created which have raised significant concerns for their use in day-to-day life. Hence, it is crucial to identify the digitally altered facial images. A new method to identify the manipulated human faces and locate the region of manipulation is done using U-net Architecture. This manipulated image, created by altering the original image, along with the mask of alteration is used to train the model. Image padding, rectified linear activation, max pooling, up convolution are performed and thus output mask is obtained. The accuracy and dice score are calculated. Thus the area of manipulation done on the human face is found by using U-net.

Key words: Facial manipulation, U-net, area of manipulation, multi-class classification

INTRODUCTION

The use of manipulated images in social media and on the internet is growing rapidly. Attackers manipulate the facial area of an image digitally and create a new fake image. This fake image, may be created by altering the original image by image splicing – copy a part of image and pasting it onto another image – by face swap –swapping the faces of individuals in an image- by deep fakes replacing a face with another face based on deep learning. It's crucial to find the digitally manipulated facial images and videos surfing on the internet to avoid spread of false news. There are many existing facial manipulation detection methods. However, these detection methods results in binary classification and are feasible to only limited manipulation techniques.

In this paper, a novel method for facial manipulation detection using the U-net Architecture is proposed. U-net is an architecture used for multi class segmentation of images, which was first designed and applied to bio-medical images in 2015. Using U-net architecture, the model is trained with datasets of real

and fake faces. The images are padded, max pooled, up convoluted to obtain the region of manipulation. Thus the multi-class detection of altered human faces is done and the part of alteration is found.

RELATED WORK

In the past few decades, the virtual face manipulation has emerged rapidly. Many detection methods have been proposed and implemented. Most of the implementations are based on the binary classification of images as true or fake. In Face X ray for more general face forgery detection[1] paper, a novel method of using face X ray to detect the facial manipulation is performed. The face X-ray of an input face image is a greyscale image that reveals whether the input image can be decomposed into the blending of two images from different sources. It does so by showing the blending boundary for a forged image and the absence of blending for a real image

The most popular implemenation of facial manipulation detection is the Face Forensics ++: Learning to Detect Manipulated Facial Images [2] paper by Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, Matthias Neibner. The facial images are manipulated based on DeepFakes, FaceSwap, Face2Face, NeuralTextures. The model detects based on Stegnalysis features, Learned features, and compared with forgery detection variants and GAN based methods, and then evaluated based on training corpus size.

Generalized Facial Manipulation Detection with Edge Region Feature Extraction [3] paper by Dong-Keon Kim, Kwangsu Kim presents a generalized manipulation detection based on the edge region features appearing in the images. A facial forensic framework that utilizes pixel-level color features appearing in the edge region of the whole image is used along with a 3D-CNN classification model.

A compact architecture based on fully convolutional neural network which is used to detect manipulated human faces, is proposed on the paper, Using Fully Convolutional Neural Networks to detect manipulated images in videos [4] by Michail Tarasiou, Stefanos Zafeiriro. This paper shows that the network's classification performance improves significantly by the addition of pixel level classifiaction loss. For the automatic detection of manipulated face images, local image features that are shared across manipulated regions are the key element used.

DeepFake Detection Based on the Discrepancy Between the Face and its Context [5] by Yuval Nirkin, Lior Wolf, Yosi Keller, Tal Hassner. In this paper, a method for detecting face swapping and other identity manipulations in single images is proposed. This approach involves two networks: (i) a face identification network that considers the face region bounded by a tight semantic segmentation, and (ii) a context recognition network that considers the face context (e.g., hair, ears). A method which uses the recognition signals from these two networks to detect such discrepancies, providing a complementary detection signal that improves conventional real vs. fake classifiers is used.

Forensic Transfer is used to detect unseen image manipulation approaches. A forensic embedding that is used to distinguish between real and fake imagery. A new auto encoder-based architecture which enforces activations in different parts of a latent vector for the real and fake classes is used. This is explained in the paper, ForensicTransfer: Weakly-supervised Domain Adaptation for Forgery Detection[6] by Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, Matthias Neibner.

In the paper, On the Detection of Digital Face Manipulation [7] by Hao Dang, Feng Liu, Joel Stehouwer, Xiaoming Liu, Anil Jain, an attention mechanism to process and improve the feature maps of the classifier model is proposed. The learned attention maps highlight the informative regions to further improve the binary classification power, and also visualize the manipulated regions. The use of an attention mechanism improves manipulated facial region localization and fake detection.

Towards Generalizable and Robust Face Manipulation Detection via Bag-of-Local-Feature [8] by Changtao Miao, Qi Chu, Weihai Li, Tao Gong, Wanyi Zhuang, Nenghai Yu, paper extends Transformers using bag-of-feature approach to encode inter-patch relationships, allowing it to learn local forgery features without any explicit supervision. Bag-of-local-feature representations are based on a vocabulary of visual words which represents cluster of local image features.

In the paper, Extracting deep local features to detect manipulated images of human faces[9], by Michail Tarasiou, Stefanos Zafeiriro, a lightweight architecture with the correct structural bias for extracting local image features and derive a multi-task training scheme that consistently outperforms image class supervision alone is proposed.

U-Net: Convolutional Networks for Biomedical Image Segmentation [10]by Olaf Ronneberger, Philipp Fischer, Thomas Brox, paper presents a network and training strategy that relies on the strong use of data augmentation to use the available annotated samples more efficiently. The U-net architecture consists of a contracting path to capture context and a symmetric expanding path that enables precise localization. We show that such a network can be trained end-to-end from very few images and outperforms the prior best method a sliding-window convolutional network.

A generic deep convolutional neural network (DCNN) for multi-class image segmentation is presented in the paper FU-net: Multi-class Image Segmentation Using Feedback Weighted U-net [11] byMina Jafari, Ruizhe Li, Yue Xing, Dorothee Auer, Susan Francis, Jonathan Garibaldi, Xin Chen. It is based on a well-established U-net. U-net is firstly modified by adding widely used batch normalization and residual block (named as BRU-net) to improve the efficiency of model training. Based on BRU-net, a dynamically weighted cross-entropy loss function is introduced. The weighting scheme is calculated based on the pixel-wise prediction accuracy during the training process. Assigning higher weights to pixels with lower segmentation accuracies enables the network to learn more from poorly predicted image regions.

ARCHITECTURE

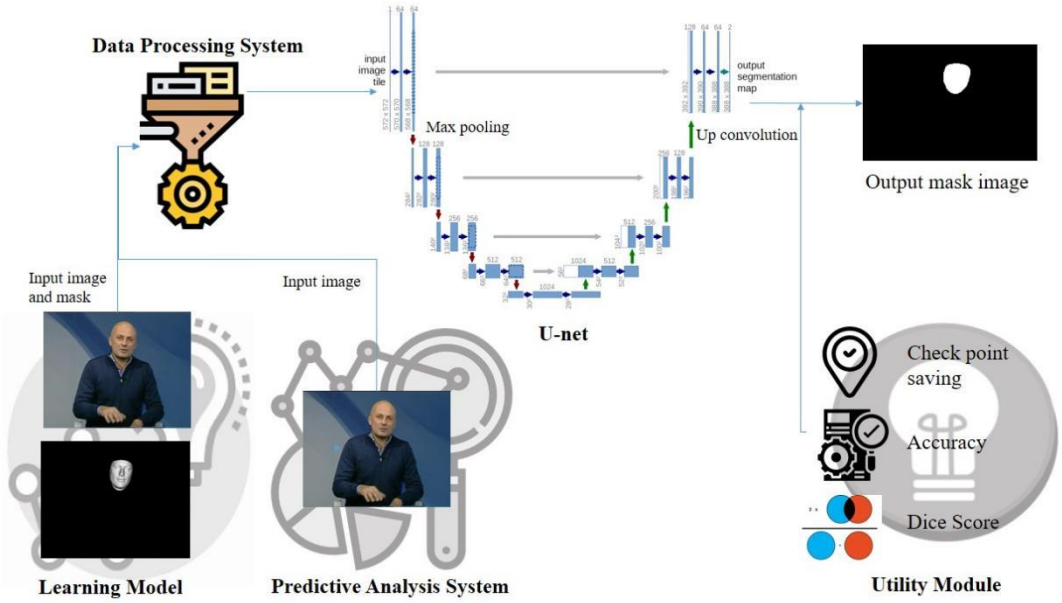

Fig 1: system architecture

- **Data Processing System**

The data processing system fetches the data from the input folder and changed it into the RGB format and resizes it as per requirement. Then, it returns the mask and the image to the learning model. It has the images of human faces, their masks for training purpose. The data folder has train_images folder to contain the images used for training, train_mask folder to contain the masks of the altered images. val_images and val_mask folders contains the images and masks used for validation respectively.

- **Utility Module**

It imports the image dataset. This file saves the checkpoint at each execution. A checkpoint is an intermediate dump of a model's entire internal state (its weights, current learning rate, etc.) so that the framework can resume the training from this point whenever desired. It loades the images for training, calculates the dicescore. Dice score is used to quantify the performance of image segmentation methods. It is calculated by,

$$\textbf{Dice Score} = 2^* X \cap Y / X+Y$$

where,

$X \cap Y$ is overlapping of image and mask

$X+Y$ is number of pixels in image and mask

The utils.py file also checks accuracy and saves the predictions as image in the output folder.

- **Learning Model**

The learning model uses the data of facial images and their masks from the data processing system and gives it to the U-net model as input. Then the output of the U-net which is the area of manipulation is stored in the output folder. The learning model also saves checkpoint, checks accuracy, saves the predictions as images in a loop based on the number of epochs. It prints the accuracy, loss, progress bar, dice score on the screen/terminal. It saves the predictions as image in the path specified. The cross entropy loss is calculated by,

$$H(p,q) = - \sum p(x) * \log (q(x))$$

where,

p(x) = true probablity distribution

q(x)= model's predicted probablity distribution

- **Unet**

UNet, a convolutional neural network was first designed, and applied in 2015 for biomedical image segmentation, by Ronneberger, Fischer and Brox.

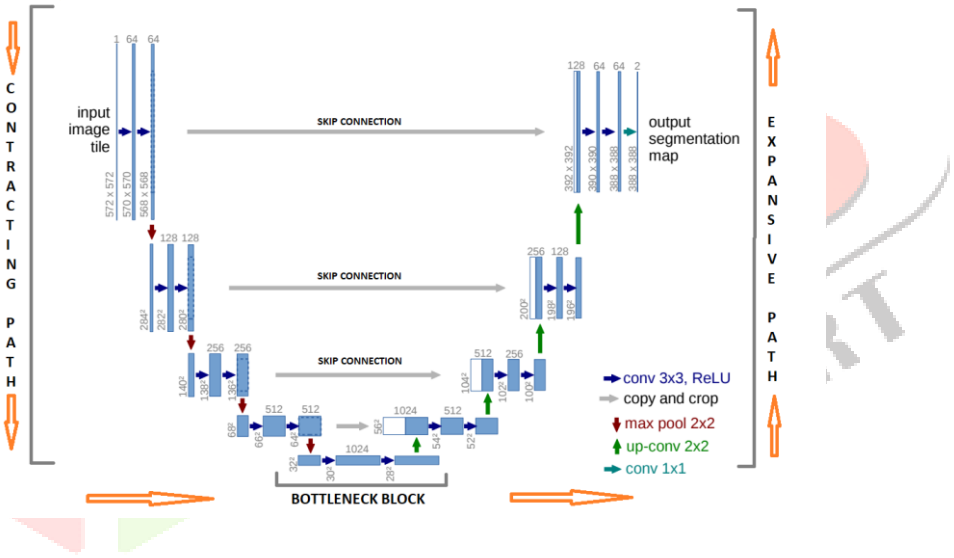

Fig 2: U-net

A typical convolutional neural network focuses on image classification tasks, where the output to an image is a single class label. But Unet focuses on localize the area i.e., a class label is supposed to be assigned to each pixel. The model.py file uses *torch* for building U-net.

The UNet network model has 3 parts:

• The Contracting/Downsampling Path.

• Bottleneck Block.

• The Expansive/Upsampling Path.

- **Contracting Path**

The contracting path starts with the input image undergoing two 3x3 padded convolutions (denoted by the blue right headed arrows) in sequence.

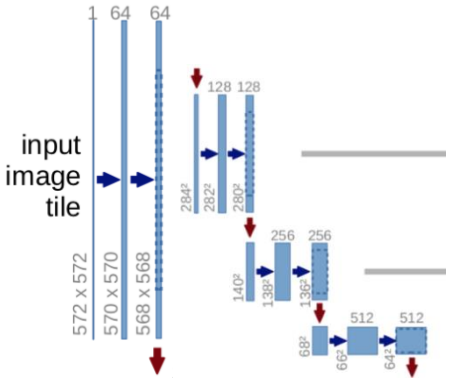

Fig 3: Input layer Fig. Contracting path

Each followed by a rectified linear unit and a 2x2 max pooling operation with stride 2 (denoted by the downward red arrow) for downsampling. At each downsampling step the depth of the image is increased by doubling the number of feature channels.

- **Bottleneck Block**

The bottleneck block connects the contracting and the expansive paths. This block performs two padded convolutions each with 1024 filters and prepares for the expansive path. There is no pooling operation involved in this part of the network.

Fig 4: bottle neck block

- **Expansive Path**

Every step in the expansive path consists of an up sampling of the feature map followed by a 2x2 up-convolution using transposed convolutions (denoted by the upward green arrows), a concatenation with the correspondingly feature map from the contracting path.

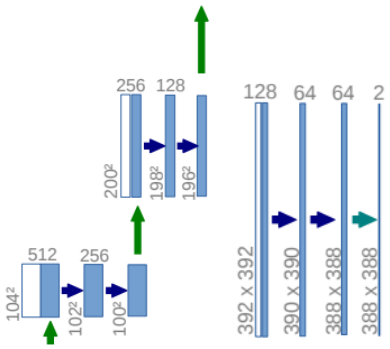

Fig 5: Expansive path Fig 4.8 Output layer

The skip connections provide features from earlier layers that are sometimes lost due to the depth of the network. It is followed by two 3x3 convolutions, each followed by a ReLU (denoted by the blue right headed arrows). Transposed convolution is an up sampling technique to expand the size of images.

Output Layer

The final (output) layer is a 1x1 convolution is used to map each (64 component) feature vector to the desired number of classes.

- **Prediction System:**

The prediction system uses the trained U-net model to predict and localize the area of manipulation done on the facial image provided.

The data provided to this system is the image that may be digitally manipulated. This image is processed in data processing system and then given to the trained U-net model for prediction, and the output mask is saved in the output folder.

IMPLEMENTATION

The input data is processed before it is fed to the U-net. The input images are resized and converted into RGB format for facial images and grayscale for mask images. The conversion to RGB format of images is done using the predefined functions in python

The U-net model is trained with the training dataset and then the ouput masks are saved in the output folder. The checkpoints are saved in each epochs and the accuracy, loss, dice scrore are calculated. This is repeated for a predefined number of epochs. The loss is calcluated by using the predefined loss funtion in python.

For prediction, the untrained dataset of only facial images is used. This is processed in Data processing System and then given into the trained U-net model. The area of manipulation is located and then the output grayscale mask image is saved in the output folder. The accuracy is calculated along with the loss in each epoch. On implementing the loss was found to be 25.2 with dice score 0.0

CONCLUSION

The usage of fake face representations is growing at a fast rate in social media and on the internet. Digital alteration methods and the usage of manipulated facial images are evolving rapidly. False representations lead to disbelief in the digital content. We are in tire need to find solution for this issue to prevent the social evils like false news spread, imitation of persons.

This project will help us to not only differentiate the fake faces from pristine, but also shows the area of manipulation or alteration. Thus the false representations of human faces can be avoided in social media and on the internet.

REFERENCES

1. *Face X ray for more general face forgery detection* paper by Paper URL: https://arxiv.org/abs/1912.13458

2. *Face Forensics ++: Learning to Detect Manipulated Facial Images* paper by Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, Matthias Neibner. Paper URL: https://arxiv.org/abs/1901.08971

3. *Generalized Facial Manipulation Detection with Edge Region Feature Extraction* paper by Dong-Keon Kim, Kwangsu Kim. Paper URL: https://arxiv.org/abs/2102.01381

4. *Using Fully Convolutional Neural Networks to detect manipulated images in videos* by Michail Tarasiou, Stefanos Zafeiriro. Paper URL: https://arxiv.org/abs/1911.13269

5. *DeepFake Detection Based on the Discrepancy Between the Face and its Context* by Yuval Nirkin, Lior Wolf, Yosi Keller, Tal Hassner. Paper URL: https://arxiv.org/abs/2008.12262

6. *ForensicTransfer: Weakly-supervised Domain Adaptation for Forgery Detection* by Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, Matthias Neibner. Paper URL: https://arxiv.org/abs/1812.02510

7. *On the Detection of Digital Face Manipulation* by Hao Dang, Feng Liu, Joel Stehouwer, Xiaoming Liu, Anil Jain. Paper URL: https://arxiv.org/abs/1812.02510

8. *Towards Generalizable and Robust Face Manipulation Detection via Bag-of-Local-Feature* by Changtao Miao, Qi Chu, Weihai Li, Tao Gong, Wanyi Zhuang, Nenghai Yu. Paper Link:https://arxiv.org/abs/2103.07915

9. *Extracting deep local features to detect manipulated images of human faces*, by Michail Tarasiou, Stefanos Zafeiriro. Paper Link: https://arxiv.org/abs/1911.13269

10. *U-Net: Convolutional Networks for Biomedical Image Segmentation*by Olaf Ronneberger, Philipp Fischer, Thomas Brox. Paper Link:https://arxiv.org/abs/1505.04597

11. *FU-net: Multi-class Image Segmentation Using Feedback Weighted U-net* byMina Jafari, Ruizhe Li, Yue Xing, Dorothee Auer, Susan Francis, Jonathan Garibaldi, Xin ChenPaper URL: https://arxiv.org/abs/2004.13470