Recolored/Fake Image Detection using Convolutional Neural Network

Preethi Kaparaboina[1], Ajay Vinjamuri[2]
[1]Business Technology Analyst, Deloitte Tax India Pvt Lt, Hyderabad, Telangana, India.
[2]Software Developer 2, Micron Technology Operations India LLP, Hyderabad, Telangana, India.

Abstract:
Nowadays, millions of photographs are produced by various devices and distributed by newspapers, television, and websites every day. Many legal, governmental and scientific organizations use digital images as evidence of specific events to make critical decisions. With the development of the sophisticated techniques for digital image forgery and the low cost to obtain a high-quality digital image, anyone can manipulate a digital image easily without leaving visible clues. This inevitably relates to some problems on legal forensics in digital images such as digital image authentication, image media copyright, and so on. Image recoloring is a technique that can transfer image color or theme and result in an imperceptible change in human eyes. As a result, digital images can no longer be believed, and they will not hold the only way as a definitive record of an event. In order to recover people's confidence towards the authenticity of the digital image, it is necessary to develop a set of methods to authenticate digital images. There is no accurate method to detect this kind of forgery. Previous forged image detection approaches focus on statistical relationships of hand-crafted appearance features between the original and tampered images. For example, that pixel value mapping leaves behind artifacts and detect enhancement by observing the intrinsic fingerprints in the pixel value histogram. We attempt to distinguish recolored images from natural images. After which we observe the inter-channel correlation and illumination consistency for natural images which may not hold for recolored images. Using these two features along with the input image we find whether an image is recolored are not.

Keywords: **Recolored Image, Digital Image Forgery, Tampered images, Inter-channel correlation, Illumination consistency.**

1. Introduction

Image recoloring is a technique that can transfer image color or theme and result in an imperceptible change in human eyes. It is one of the most common image operations in photo editing. Usually, satisfying color transfer algorithms apply the color characteristic of a target image to a source image and generate a recolored result that humans cannot distinguish.

The rapid proliferation of image editing technologies has increased both the ease with which images can be manipulated and the difficulty in distinguishing between altered and

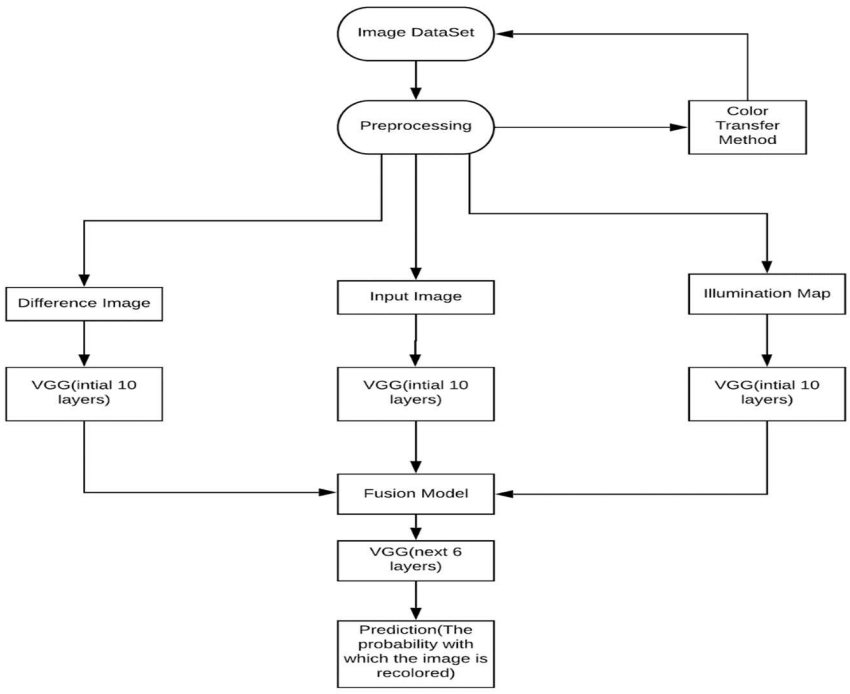
natural images. In addition to the conventional image editing techniques such as splicing, copy-move and retouching, more image editing techniques, such as colorization and image generation, are proposed. Since these types of image editing techniques generate new content with/without references, they are called as the generative image editing techniques.

Although image editing techniques can provide significant aesthetic or entertainment value, they may also be used with malicious intent. In general, various image editing approaches employ different mechanisms. Splicing and copy-move techniques usually manipulate part of the image and perform object-level changes. Among the generative image editing techniques, image generation usually produces a meaningful image from a noise vector with/without some additional information such as text or a class label. Colorization, on the other hand, usually colorizes images with visually plausible colors, which may cause misjudgment when specific objects/scenes must be identified/tracked.

The problem definition includes the distinguishing of the original images with the recolored images where the recolored images are produced using different recoloring techniques like color transfer method, recoloring based on intrinsic image estimation, copy-move forgery. The significance of this problem is, since the distortion of the various properties of an image such as color, shades etc. would lead to the damage of the data that the image is enclosing. The major objective of this would come under forgery detection which basically deals with the manipulation of any kind of data.

2. Methodologies used in Recolored/Fake Image Detection

2.1 Process diagram



We use three feature extractors and a feature fusion module to learn forgery-relevant features. We adopt the original image as one of the input branches like traditional neural

networks. Additionally, we derive DIs and IM as two pieces of evidence of image recolored detection based on the observations that images may not maintain the inter-channel correlation or illuminant consistency after the recoloring process. These two pieces of evidence are employed as two additional input branches together with the original image. Exploiting the property of inter-channel correlation, difference image is produced and using illumination consistency property illumination map is produced.

2.2 Features

2.2.1 Inter Channel Correlation: Most commercial digital cameras are equipped with an image sensor, charge coupled device (CCD) or complementary metal-oxide-semiconductor (CMOS) and acquire the color information of each pixel using a CFA. For example, the Bayer array, the most frequently used CFA, consists of four channels: red, blue, and two green channels. The green pixels are sampled on a quincunx lattice while the red and blue pixels are sampled on rectilinear lattices. As a result, the captured images by such cameras include specific correlations which are likely to be destroyed during manipulation. Instead of analyzing the property of one special CFA pattern, we focus on the common correlations among a range of CFA algorithms. Gunturk et al. have shown that high-frequency components across image color channels are strongly correlated and similar. For most images, the correlation coefficients range from 0.98 to 1. In addition, this correlation has been widely used in CFA de-mosaicking. We exploit this property to distinguish recolored images by DIs

2.2.1.1 Difference image: Difference Image is developed by exploiting the property of inter-channel correlation. A difference image (DI) from natural images is approximately equivalent to itself after passing through a low-pass filter. Therefore, compared to the original color channels, the DIs are smoother due to the lack of edges or details.

Figure.1 Original Image

This is the original image for which the difference image is produced. Below are the inter channel correlations of R-B, B-G, G-R channels produced for the given image. These images are stacked up to form the difference image.

Figure.2 G-R Difference Image

Figure.3 B-G Difference Image

Figure.4 B-R Difference Image

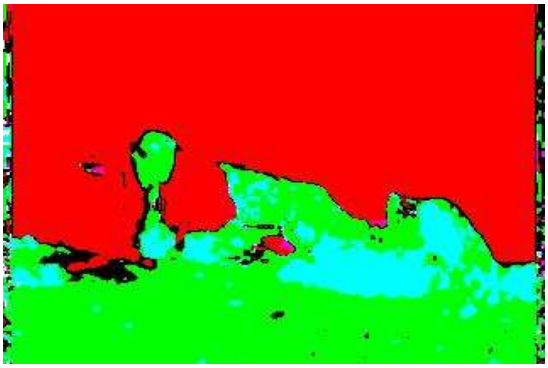


Figure.5 Difference Image

2.2.2 Illumination Consistency: Illumination Map represents the illuminant color of the input image and has the same dimension to the original RGB image. The values at each pixel denote the corresponding estimated illuminant color at this position. In general, the illuminant colors in a neighborhood should be close due to the illuminant consistency. However, image recoloring cannot maintain the illuminant consistency since the changes of pixels are not identical. It has been widely used in forgery detection over a decade, especially for splicing. Illumination-based methods are mainly grouped into two approaches: geometry-based and color-based approaches. Geometry-based methods look for inconsistencies in light source positions and color-based methods focus on inconsistencies in the estimated light color. Since the changes of pixels in one image are not exactly identical during the color transfer process, it may be hard to maintain the illuminant consistency. Therefore, we utilize the illuminant consistency as another property in our discriminative model. Riess and Angelopoulou propose

an estimate strategy, which first segments the image into super-pixels with similar color and then utilizes an illuminant color estimator for local estimation at each super-pixel.
In this work, the same strategy is applied, and a new image called IM is derived. The pixel values of the entire image are used for illuminant color estimation. In this method it focused on low level features. Such as Gray World, Max-RGB, Shades of Gray.

Gray World Hypothesis: In Gray World, Illuminant color is estimated from average pixel values of images. Under a neutral light source or white light source, average reflectance of the entire image is achromatic (having no colors), if any deviation from this condition is due to color of illumination. This average reflected color will be the light source color. The input image f(x) with the assumption that input image is illuminated by single light source. Gijsenij and van deWeijer grey world illuminant estimation in terms of minkowski norm (p-norm).

Max-RGB Hypothesis: In Max-RGB illuminant color estimated from maximum response of Red Green Blue (RGB) channel. Maximum response is obtained from perfect reflectance. A surface having perfect reflectance property will respond (reflect) for the full range of light colors it captures, when light incident on it. Then this reflected color is actually the color of light source. The input image $f(x)$, with the assumption that input image is illuminated by single light source. Computed by maximum response of 3 channel estimated. i.e.

$$\max(f(x))=\max_r(x),\max_g(x),\max_b(x)$$

Shades of Gray: Gray world and the max-RGB illuminant color estimation in terms of Minkowski norm, is called shades of gray. Gijsenij and van deWeijer[14] introduce shades of grey estimation in terms of minkowski norm. The values at each pixel denote the corresponding estimated illuminant color at this position. In general, the illuminant colors in a neighborhood should be close due to the illuminant consistency. However, image recoloring cannot maintain the illuminant consistency since the changes of pixels are not identical.

Figure.6 Illumination Map

2.2.3 Fusion Module: In the testing phase, we have 3 inputs i.e., input image and its difference image and illumination map which are used to improve the accuracy of the model. These three inputs are processed up to 10 layers of the architecture and later are concatenated. This fusion is done by taking the mean of the outputs of the models resulted after processing the 10 layers.

2.3 Dataset

The PASCAL VOC is the dataset used in this project. The dataset provides standardized image data sets for object class recognition. Instance segmentation, semantic segmentation, pose segmentation etc. The VOC challenge encourages two types of participation: (i) methods which are trained using only the provided "trainval" (training + validation) data; (ii) methods built or trained using any data except the provided test data, for example commercial systems. Using this dataset, we generate recolored images using different techniques and train the model. Testing the model also involve the same dataset, but not the same images used for training. There are around 17000 images in this dataset.

Figure.7 Source Image

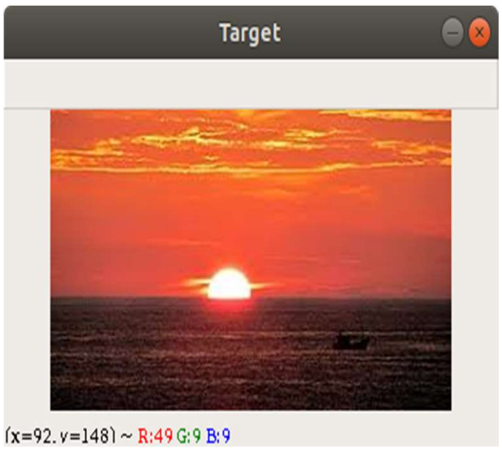


Figure.8 G-R Target Image

The theme of the source image is applied to the target image to produce a recolored image. This process is used to generate the dataset so that the original and the recolored images are in the ratio 1:1.

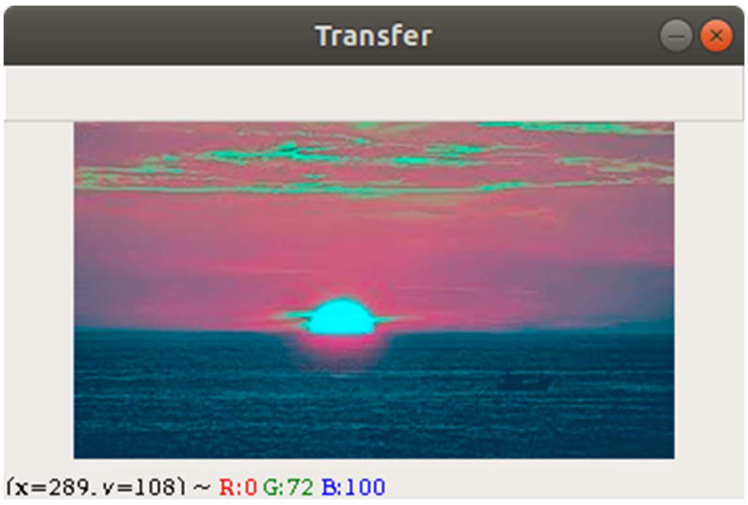


Figure.9 Transfer Image

This is the output of the recolored image produced by applying the color of the source image to the target image.

2.4 Architecture

ConvNet Configuration		
DIs (224*224)	Original image (224*224 RGB)	IM (224*224)
conv3-64	conv3-64	conv3-64
conv3-64	conv3-64	conv3-64
maxpool	maxpool	maxpool
conv3-128	conv3-128	conv3-128
conv3-128	conv3-128	conv3-128
maxpool	maxpool	maxpool
conv3-256	conv3-256	conv3-256
conv3-256	conv3-256	conv3-256
conv3-256	conv3-256	conv3-256
maxpool	maxpool	maxpool
concat		
conv3-512		
conv3-512		
conv3-512		
maxpool		
conv3-512		
conv3-512		
conv3-512		
FC-4096		
FC-4096		
FC-2		
Soft-max		

Figure.10 Architecture

A Convolutional Neural Network (ConvNet) is a Deep Learning algorithm which can take in an input image, assign importance to various aspects in the image and be able to differentiate one from the other. The pre-processing required in a CNN is much lower as compared to other classification algorithms. The architecture of it is similar to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. It is able to successfully capture the Spatial and Temporal dependencies in an image through the application of relevant filters. The architecture performs a better fitting to the image dataset due to the reduction in the number of parameters involved and reusability of weights. In other words, the network can be trained to understand the sophistication of the image better.

2.4.1 Convolution Layer: There are different layers in this architecture. Convolution is the first layer to extract features from an input image. It preserves the relationship between pixels by learning image features using small squares of input data. It is a mathematical operation that takes two inputs such as image matrix and a filter or kernel. Convolution of an image with different filters can perform operations such as edge detection, blur and sharpen by applying filters. It is responsible for capturing the Low-Level features such as edges, color, gradient orientation, etc.

2.4.2 Pooling Layer: The Pooling layer is responsible for reducing the spatial size of the Convolved Feature. This is to decrease the computational power required to process the data through dimensionality reduction. It is useful for extracting dominant features which are rotational and positional invariant, thus maintaining the process of effectively training of the model.

There are two types of Pooling: Max Pooling and Average Pooling. Max Pooling returns the maximum value from the portion of the image covered by the Kernel and also performs as a Noise Suppressant which discards the noisy activations altogether and de-noising along with dimensionality reduction. On the other hand, Average Pooling returns the average of all the values from the portion of the image covered by the Kernel and performs dimensionality reduction as a noise suppressing mechanism.

2.4.3 Fully Connected Layer: Adding a Fully Connected layer is a (usually) cheap way of learning non-linear combinations of the high-level features as represented by the output of the convolutional layer. The Fully Connected layer is learning a possibly non-linear function in that space. Now that we have converted our input image into a suitable form for our Multi-Level Perceptron, we shall flatten the image into a column vector. The flattened output is fed to a feed-forward neural network and backpropagation applied to every iteration of training. Over a series of epochs, the model is able to distinguish between dominating and certain low-level features in images and classify them using the SoftMax Classification technique.

After training the model with dataset of the original and recolored images, the testing takes places with two additional features difference image and illumination map taken as input. These three inputs are processed up to 10 layers of the VGG network simultaneously. Later these layers are concatenated and processed to the remaining layers of the network. The resultant value is the probability of the image to be original.

3. Conclusion and Future Work

The dataset of VOC pascal was used. As it is a binary classification the dataset should be in the ratio 1:1. So recolored images are generated using the color transfer between the images. Around 4000 images of dataset id generated. This dataset is used in training the model. The model is built using keras in sequential way. The 16 layers of the VGG network is added one by one in this mode. Dataset generator is used in training the model as large dataset imposes heavy load on the memory usage at a time. It generates epochs where the model is trained with a set of data from the dataset for each epoch.

In the testing phase the features, illumination map and difference image are extracted from the input image. Illumination consistency is the main concept of illumination map. The illumination map is generated from the input image using gray world hypothesis and max RGB hypothesis. The values at each pixel in the illumination map denote the corresponding estimated illuminant color at this position. Inter channel correlation is the main concept of the difference image. For the original image the correlation between the R, G, B channels is 0.98 or 1. But for the recolored image the correlation is not high. Using this correlation property, the difference images between 3 channels are generated. Later the 3 images are stacked up.

The input image along with the two features is sent as input through the model up to 10 layers. The output from these layers are concatenated and processed for the remaining layers of the model to get the probability. And we use the dropouts, optimizers etc to increase the accuracy of the system.

3.1 Future Work

The accuracy of the model is expected to be high as they are used as a mark of evidence in many of the situations. Newer models are developed keeping the same as the goal.

- In this model the training phase takes the more time where large numbers of images are used. It takes hours of time to complete this phase. The speed has to be increased in training the model
- Running the system with illumination map and difference images uses the memory in large which has to be compressed.
- The model can be built using other networks which can improve the accuracy and decrease the building time.
- More efficient architectures can be built.
- Difference recoloring techniques involving different properties can be considered as the part of the features in improving the accuracy.

References

[1] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color transfer between images," IEEE Comput. Graph. Appl., vol. 21, no. 5, pp. 34–41, Sep./Oct. 2001.

[2] J. S. Ho, O. C. Au, J. Zhou, and Y. Guo, "Inter-channel demosaicking traces for digital image forensics," in Proc. IEEE Int. Conf. Multimedia Expo, Jul. 2010, pp. 1475–1480.

[3] S. Gholap and P. K. Bora, "Illuminant colour based image forensics," in Proc. IEEE Region 10 Conf. TENCON, Nov. 2008, pp. 1–5.

[4] Y. Guo, X. Cao, W. Zhang, and R. Wang, "Fake colorized image detection," IEEE Trans. Inf. Forensics Security, vol. 13, no. 8, pp. 1932–1944,Aug. 2018.

[5] Z.-P. Zhou and X.-X. Zhang, "Image splicing detection based on image quality and analysis of variance," in Proc. Int. Conf. Edu. Technol. Comput., Jun. 2010, pp. 242–246.

[6] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," CoRR, vol. abs/1409.1556, Sep. 2014.

[7] S. Beigpour and J. van de Weijer, "Object recoloring based on intrinsic image estimation," in Proc. ICCV, Nov. 2010.

[8] F. Pitié, A. C. Kokaram, and R. Dahyot, "Automated colour grading using colour distribution transfer," Comput. Vis. Image Understand.,vol. 107, nos. 1–2, pp. 123–137, 2007.

[9] H. Chang, O. Fried, Y. Liu, S. DiVerdi, and A. Finkelstein, "Palette-based photo recoloring," ACM Trans. Graph., vol. 34, no. 4, 2015,Art. no. 139.

[10] X. Pan and S. Lyu, "Region duplication detection using image fea- ture matching," IEEE Trans. Inf. Forensics Security, vol. 5, no. 4, pp. 857–867, Dec. 2010.

[11] X. Wang, J. Xue, Z. Zheng, Z. Liu, and N. Li, "Image forensic signature for content authenticity analysis," J. Vis. Commun. Image Represent., vol. 23, no. 5, pp. 782–797, 2012.

[12] Z. Junhong, "Detection of copy-move forgery based on one improved lle method," in Proc. 2nd Int. Conf. Adv. Comput. Control, Mar. 2010, pp. 547–550.

[13] Z.-P. Zhou and X.-X. Zhang, "Image splicing detection based on image quality and analysis of variance," in Proc. Int. Conf. Edu. Technol. Comput., Jun. 2010, pp. 242–246.

[14] https://www.tensorflow.org/tutorials