# A general model for detecting digital image fraud using deep learning

[1]G.Venkatesh, [2]R.Ramesh, [3]G.Pavani

[1,2,3]Department of Computer science and Engineering, Kasireddy Narayanreddy College Of Engineering and Research, Hyderabad

## ABSTRACT

Data misuse is being caused by the technological advances that are permeating every element of the modern world. Therefore, researchers must overcome the difficult process of detecting these modified data types and separating the true data from the manipulated. Splicing is one of the most popular methods for manipulating digital images; it involves copying a specific section from one image and pasting it into another. The identification of image forgeries is thought to be a trustworthy method of confirming the veracity of digital photographs. In this paper, we suggested a method based on the cutting-edge ResNet50v2 deep learning architecture. The suggested model uses the ResNet50v2 architecture and YOLO convolutional neural network weights to process image batches as input. In this study, we used the CASIA_v1 and CASIA_v2 benchmark datasets, which contain two distinct categories, original and forgery, to detect image splicing. We used 80% of the data for the training and the remaining 20% for testing purposes. We also performed a comparative analysis between existing approaches and our proposed system. We evaluated the performance of our technique with the CASIA_v1 and CASIA_v2 datasets. Since the CASIA_v2 dataset is more comprehensive compared to the CASIA_v1 dataset, we obtained 99.3% accuracy for the fine-tuned model using transfer learning and 81% accuracy without transfer learning with the CASIA_v2 dataset. The results show the superiority of the proposed system.

**Keywords:** machine learning; deep learning; image forgery; ResNet50; YOLO CNN; CASIA

## INTRODUCTION

Electronic equipment is now widely and affordably available as a result of technical progress and globalization. Digital cameras have become more and more well-liked as a result. We employ the numerous camera sensors that are all around us to capture a large number of photos. Many documents that must be filed online require photographs in the form of soft copies, and a lot of images are shared on social media every day. The beautiful thing about visuals is that even those with little formal education can look at them and deduce information. Images are therefore a fundamental part of the digital world and are crucial for both storing and disseminating data. The photos can be readily edited using a variety of tools [1,2]. These tools were created with the intention of enhancing and improving the images. However, rather than enhancing the image, some people exploit their capabilities to falsify images and propagate falsehoods [3,4]. This is a significant threat, as the damage caused by faked images is not only severe, but also frequently irreversible. There are two basic types of image forgery: image splicing and copy-move, which are discussed below: Electronics 2022, 11, 403. https://doi.org/10.3390/electronics11030403 https://www.mdpi.com/journal/electronics Electronics 2022, 11, 403 2 of 17 • Image Splicing: A portion of a donor image is copied into a source image. A sequence of donor images can likewise be used to build the final forged image. • Copy-Move: This scenario contains a single image. Within the image, a portion of the image is copied and pasted. This is frequently used to conceal other objects. The final forged image contains no components from other images. The primary purpose in both cases of image forgery is to spread misinformation by changing the original content in an image with something else [5,6]. Earlier images were an extremely credible source for the information exchange, however, due to image forgery, they are used to spread misinformation. This is affecting the trust of the public in images, as the forging of images may or may not be visible or recognizable to the naked eye. As a result, it is essential to Detect image forgeries to preventthe spread of misinformation as well as to restore public trust in images. This can be done by exploring the various artifacts left Behind when an image forgery is performed, and they can be identified using various image processing techniques. Researchers have proposed a variety of methods for detecting the presence of image forgeries [7–9]. Conventional image forgery detection techniques detect forgeries by concentrating on the multiple artifacts present in a forged image, such as changes in illumination, contrast, compression, sensor noise, and shadow. CNN's have gained popularity in recent years for various computer vision tasks, including image object recognition, semantic segmentation, and image classification. Two major features contribute to CNN's success in computer vision. Firstly, CNN takes advantage of the significant correlation between adjacent pixels. As a result, CNN prefers locally grouped connections over one-to-one connections between all pixel. Second, each output feature map is produced through a convolution operation by sharing weights. Moreover, compared to the traditional method that depends on engineered features to detect specific forgery, CNN uses learned features from training images, and it can generalize itself to detect unseen forgery. These advantages of CNN make it a promising tool for detecting the presence of forgery in an image. It is possible to train a CNN-based model to learn the many artifacts found in a forged image [10–13]. Thus, we propose a very light CNN-based network, with the primary goal of learning the artifacts that occur in a tampered image as a result of differences in the features of the original image and the tampered region. The major contribution of the proposed technique are as follows: • A lightweight CNN-based architecture is designed to detect image forgery efficiently. The proposed technique explores numerous artifacts left behind in the image tampering process, and it takes advantage of differences in image sources through image recompression. • While most existing algorithms are designed to detect only one type of forgery, our technique can detect both image splicing and copy-move forgeries and has achieved high accuracy in image forgery detection. • Compared to existing techniques, the proposed technique is fast and can detect the presence of image forgery in significantly less time. Its accuracy and speed make it suitable for real-world application, as it can function well even on slower devices. The rest of the paper is organized as follows. Section 2 provides a literature review of image forgery detection methodologies. Section 3 introduces the proposed framework for detecting the presence of forgeries in an image. Section 4 contains a discussion of the experimentation and the results achieved. Finally, in Section 5, we summarize the conclusions.

## LITERATURE SERVUY

Plagiarism Detection Approaches Plagiarism detection is a specialized Information Retrieval (IR) task with the objective of comparing an input document to a large collection and retrieving all documents exhibiting similarities above a predefined threshold. PD systems typically follow a two-stage process consisting of candidate retrieval and detailed comparison. For candidate retrieval, the systems commonly employ efficient text retrieval methods, such as n-gram fingerprinting or vector space models. For the detailed comparison, the systems typically apply exhaustive string matching. However, such approaches are limited to finding near copies of a text. To detect disguised forms of academic plagiarism, researchers have proposed a variety of mono-lingual text analysis approaches employing semantic and syntactic features, as well as cross-lingual IR methods. Researchers also showed that hybrid approaches, i.e., the combined analysis of text and other content features, improve the retrieval effectiveness for PD tasks. Alzahrani et al. combined an analysis of text similarity and structural similarity. Gipp and Meuschke showed that the combined analysis of citation patterns and text similarity improves the identification of concealed academic plagiarism. Pertile et al. confirmed the positive effect of combining citation and text analysis and devised a hybrid approach using machine learning. Recently, Meuschke et al. demonstrated the benefit of analyzing the similarity of mathematical expressions and patterns of semantic concepts for improving the identification of academic plagiarism.

Image Analysis for Plagiarism Detection Few studies have investigated the analysis of image similarity for PD. Hurtik and Hodakova use higher degree F-transform to provide a highly efficient and reliable method to identify exact copies of photographs or cropped parts there. However, the method does not consider image alterations aside from cropping. Iwanowski et al. evaluate the suitability of well-established feature point methods, such as SIFT, SURF, and BRISK, to retrieve exact and visually altered copies of photographs. Srivastava et al. address the same task using a combination of SIFT features extracted using SIFT and perceptual hashing. Feature point methods identify and match visually interesting areas of a scene. The methods are insensitive to affine image transformations, such as scaling or rotation, and relatively robust to changes in illumination or the introduction of noise. Perceptual hashing describes a set of methods that map perceived content of images, videos, or audio files to a hash value (pHash. Images perceived as similar by humans also result in similar pHash values, in contrast to cryptographic hashing, in which a minor change in the input results in a drastically different hash value. Thus, the similarity of images can be quantified as the similarity of their pHash values. If image components, such as shapes, are re-arranged, both feature point methods and perceptual hashing often fail. Iwanowski et al. mention that the effectiveness of the feature point approaches they tested decreases if the test images consist of multiple sub-images. We also observed this limitation in our tests. For example, the two compound images shown in Figure 10 in Appendix A consist of six and four sub-images, respectively. The image in the later document omits two of the sub-images present in the compound image from the source document. Applying the combination of SIFT feature extractor and MSAC feature estimator to compare these two compound images correctly identifies a high similarity between the two sub-images at the top in both compound images, but does not establish a similarity for the other sub-image pairs.

Comparing Images for Document Plagiarism Detection The paper presents results of research oriented towards an application of image processing methods into document comparisons in view of their application into plagiarism-detection systems. Among all image processing methods, the feature-point ones, thanks to their invariance to various image transforms, are best suited for computing image similarity. In the paper various combination of feature point detectors and descriptors are investigated as potential tool for finding similar images in document. The methods are tested on the database consisting of scientific papers containing 5 well known image processing test images. Also, an idea is presented in the paper how the algorithms computing the image similarity may extend the functionality of plagiarism detection systems.

Reducing Computational Effort for Plagiarism Detection by using Citation Characteristics to Limit Retrieval Space This paper proposes a hybrid approach to plagiarism detection in academic documents that integrates detection methods using citations, semantic argument structure, and semantic word similarity with character-based methods to achieve a higher detection performance for disguised plagiarism forms. Currently available software for plagiarism detection exclusively performs text string comparisons. These systems find copies, but fail to identify disguised plagiarism, such as paraphrases, translations, or idea plagiarism. Detection approaches that consider semantic similarity on word and sentence level exist and have consistently achieved higher detection accuracy for disguised plagiarism forms compared to character-based approaches. However, the high computational effort of these semantic approaches makes them infeasible for use in real-world plagiarism detection scenarios. The proposed hybrid approach uses citation-based methods as a preliminary heuristic to reduce the retrieval space with a relatively low loss in detection accuracy. This preliminary step can then be followed by a computationally more expensive semantic and character-based analysis. We show that such a hybrid approach allows semantic plagiarism detection to become feasible even on large collections for the first time. Optical Character Recognition by Open source OCR Tool Tesseract: A Case Study. Optical character recognition (OCR) method has been used in converting printed text into editable text. OCR is very useful and popular method in various applications. Accuracy of OCR can be dependent on text preprocessing and segmentation algorithms. Sometimes it is difficult to retrieve text from the image because of different size, style, orientation, complex background of image etc. We begin this paper with an introduction of Optical Character Recognition (OCR) method, History of Open Source OCR tool Tesseract, architecture of it and experiment result of OCR performed by Tesseract on different kinds images are discussed. We conclude this paper by comparative study of this tool with other commercial OCR tool Transym OCR by considering vehicle number plate as input. From vehicle number plate we tried to extract vehicle number by using Tesseract and Transym and compared these tools based on various parameters.

An Evaluation Framework for Plagiarism DetectionWe present an evaluation framework for plagiarism detection. The framework provides performance measures that address the specifics of plagiarism detection; and the PAN-PC-10 corpus, which contains 64 558 artificial and 4 000 simulated plagiarism cases, the latter generated via Amazon's Mechanical Turk. We discuss the construction principles behind the measures and the corpus, and we compare the quality of our corpus to existing corpora. Our analysis gives empirical evidence that the construction of tailored training corpora for plagiarism detection can be automated, and hence be done on a large scale.

mimPlag: Detecting image plagiarism using hierarchical near duplicate retrieval Plagiarism in any form is a serious offense especially in academia and industry where integrity and royalty from work is of utmost importance. In this work, a novel hierarchical feature extraction as well as an approximate nearest neighbor search is proposed for detecting plagiarism of images. The proposed scheme is applicable for natural images as opposed to specific image classes reported in a previous work. A comprehensive experimental analysis is provided to illustrate the efficacy of the techniques chosen for the scheme. We demonstrate that the scheme shows a lot of promise for a wide variety of attacks and is amenable to scaling.

Comparing and combining Content- and Citation-based approaches for plagiarism detection

The vast amount of scientific publications available online makes it easier for students and researchers to reuse text from other authors and makes it harder for checking the originality of a given text. Reusing text without crediting the original authors is considered plagiarism. A number of studies have reported the prevalence of plagiarism in academia. As a consequence, numerous institutions and researchers are dedicated to devising systems to automate the process of checking for plagiarism. This work focuses on the problem of detecting text reuse in scientific papers. The contributions of this paper are twofold:

We survey the existing approaches for plagiarism detection based on content, based on content and structure, and based on citations and references; and (b) we compare content and citation-based approaches with the goal of evaluating whether they are complementary and if their combination can improve the quality of the detection. We carry out experiments with real data sets of scientific papers and concluded that a combination of the methods can be beneficial.

ImageNet Classification with Deep Convolutional Neural Networks

We trained a large, deep convolutional neural network to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes. On the test data, we achieved top-1 and top-5 error rates of 37.5% and 17.0% which is considerably better than the previous state-of-the-art. The neural network, which has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three fully-connected layers with a final 1000-way softmax. To make training faster, we used non-saturating neurons and a very efficient GPU implementation of the convolution operation. To reduce overfitting in the fully-connected layers we employed a recently-developed regularization method called "dropout" that proved to be very effective. We also entered a variant of this model in the ILSVRC-2012 competition and achieved a winning top-5 test error rate of 15.3%, compared to 26.2% achieved by the second-best entry.

## RELATED WORK

[1] CNNs, which are inspired by the human visual system, are designed to be non-linear interconnected neurons. They have already demonstrated extraordinary potential in a variety of computer vision applications, including image segmentation and object detection. They may be beneficial for a variety of additional purposes, including image forensics. With the various tools available today, image forgery is fairly simple to do, and because it is extremely dangerous, detecting it is crucial. When a fragment of an image is moved from one to another, a variety of artifacts occur due to the images' disparate origins. While these artifacts may be undetectable to the naked eye, CNNs may detect their presence in faked images. Due to the fact that the source of the forged region and the background images are distinct, when we recompress such images, the forged is enhanced differently due to the compression difference. We use this concept in the proposed approach by training a CNN-based model to determine if an image is genuine or a fake. A region spliced onto another image will most likely have a statistically different distribution of DCT coefficients than the original region. The authentic region is compressed twice: first in the camera, and then again in the fake, resulting in periodic patterns in the histogram [2]. The spliced section behaves similarly to a singly compressed region when the secondary quantization table is used. As previously stated, when an image is recompressed, if it contains a forgery, the forged

portion of the image compresses differently from the remainder of the image due to the difference between the source of the original image and the source of the forged portion. When the difference between the original image and its recompressed version is analyzed, this considerably emphasizes the forgery component. As a result, we use it to train our CNN-based model for detecting image forgery. Algorithm 1 shows the working of the proposed technique, which has been explained here. We take the forged image A (images shown in Figure 1b tamper images), and then recompress it; let us call the recompressed image as Arecompressed (images shown in Figure 1c are recompressed forged images). Now we take the difference of the original image and the recompressed image, let us call it Adi f f (images shown in Figure 1e are the difference of Figure 1b,c, respectively). Now due to the difference in the source of the forged part and the original part of the image, the forged part gets highlighted in Adi f f (as we can observe in Figure 1d,e, respectively). We train a CNN-based network to categorize an image as a forged image or a genuine one using Adi f f as our input features (we label it as a featured image). Figure 2 gives the pictorial view of the overall working of the proposed method. To generate Arecompressed from A, we use JPEG compression. Image A undergoes JPEG compression and produces Arecompressed as described in Figure 3. When there is a single compression, then the histogram of the dequantized coefficients exhibits the pattern as shown in Figure 4, this type of pattern is shown by the forged part of the image. Moreover, when there is a sort of double compression then, as described in Figure 5, there is a gaping between the dequantized coefficients as shown in Figure 6, this type of pattern is shown by the genuine part of the image. We constructed a very light CNN model with minimal parameters in our proposed model (line number 5 to 13 of Algorithm 1). We constructed a model consisting of 3 convolutional layers after which there is a dense fully connected layer, as described below: • The first convolutional layer consists of 32 filters of size 3-by-3, stride size one, and “relu” activation function. • The second convolutional layer consists of 32 filters of size 3-by-3, stride size one, and “relu” activation function. • The third convolutional layer consists of 32 filters of size 7-by-7, stride size one, and “relu” activation function, followed by max-pooling of size 2-by-2. Electronics 2022, 11, 403 6 of 17 • Then we have the dense layer that has 256 neurons with “relu” activation function, finally which is connected to two neurons (output neurons) with “sigmoid”activation.

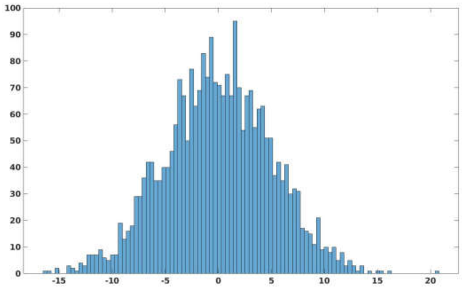


**Figure 4.** The histogram of DCT coefficients of the forged region in the image, which behaves as singly compressed.

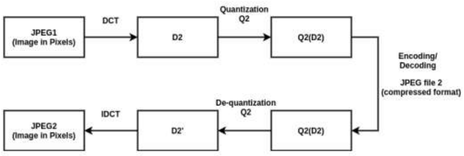


**Figure 5.** The recompression of the genuine region of the image.

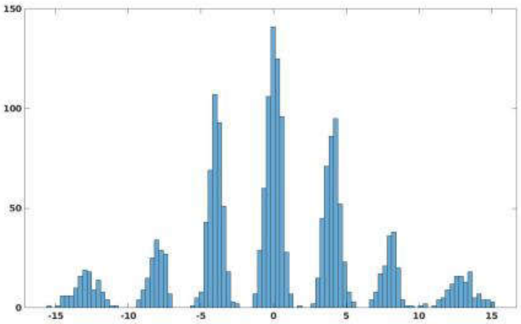
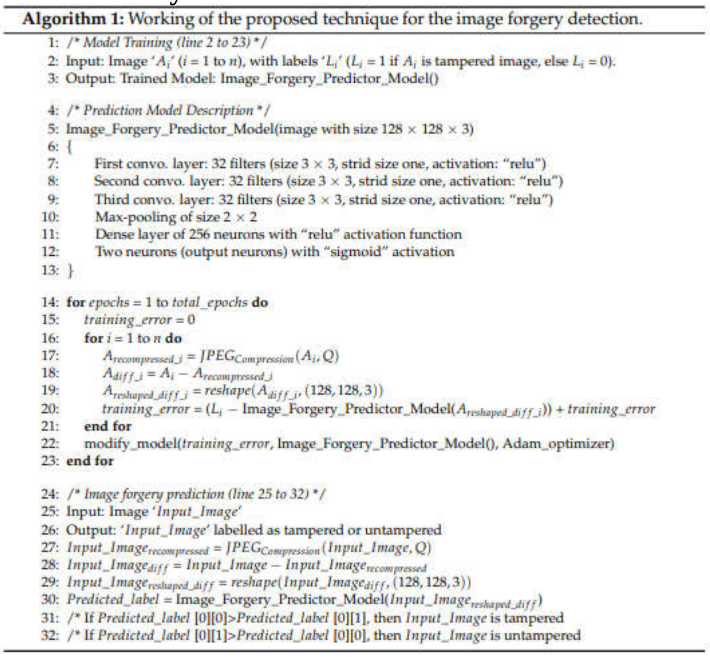


**Figure 6.** The histogram of DCT coefficients of the genuine region of an image which gets double compressed.



**Algorithm 1:** Working of the proposed technique for the image forgery detection.

```
1: /* Model Training (line 2 to 23) */
2: Input: Image '$A_i$' ($i = 1$ to $n$), with labels '$L_i$' ($L_i = 1$ if $A_i$ is tampered image, else $L_i = 0$).
3: Output: Trained Model: Image_Forgery_Predictor_Model()

4: /* Prediction Model Description */
5: Image_Forgery_Predictor_Model(image with size 128 × 128 × 3)
6: {
7:     First convo. layer: 32 filters (size 3 × 3, strid size one, activation: "relu")
8:     Second convo. layer: 32 filters (size 3 × 3, strid size one, activation: "relu")
9:     Third convo. layer: 32 filters (size 3 × 3, strid size one, activation: "relu")
10:    Max-pooling of size 2 × 2
11:    Dense layer of 256 neurons with "relu" activation function
12:    Two neurons (output neurons) with "sigmoid" activation
13: }

14: for $epochs = 1$ to $total\_epochs$ do
15:    $training\_error = 0$
16:    for $i = 1$ to $n$ do
17:        $A_{recompressed\_i} = JPEG_{Compression}(A_i, Q)$
18:        $A_{diff\_i} = A_i - A_{recompressed\_i}$
19:        $A_{reshaped\_diff\_i} = reshape(A_{diff\_i}, (128, 128, 3))$
20:        $training\_error = (L_i -$ Image_Forgery_Predictor_Model($A_{reshaped\_diff\_i}$)) $+ training\_error$
21:    end for
22:    modify_model($training\_error$, Image_Forgery_Predictor_Model(), Adam_optimizer)
23: end for

24: /* Image forgery prediction (line 25 to 32) */
25: Input: Image '$Input\_Image$'
26: Output: '$Input\_Image$' labelled as tampered or untampered
27: $Input\_Image_{recompressed} = JPEG_{Compression}(Input\_Image, Q)$
28: $Input\_Image_{diff} = Input\_Image - Input\_Image_{recompressed}$
29: $Input\_Image_{reshaped\_diff} = reshape(Input\_Image_{diff}, (128, 128, 3))$
30: $Predicted\_label$ = Image_Forgery_Predictor_Model($Input\_Image_{reshaped\_diff}$)
31: /* If $Predicted\_label$ [0][0]>$Predicted\_label$ [0][1], then $Input\_Image$ is tampered
32: /* If $Predicted\_label$ [0][1]>$Predicted\_label$ [0][0], then $Input\_Image$ is untampered
```

## Conclusions and Future Work

Recent years have seen a rise in popularity for photography due to the accessibility of cameras. Images play a significant part in our lives and have developed into a critical tool for information dissemination since they are easily understood by the general population. There are several tools available to edit photographs; while their primary purpose is to enhance them, these technologies are routinely abused to alter the images in order to distribute false information. Image forgery has consequently grown to be a serious issue and cause for concern. In this study, we provide a novel neural network- and deep learning-based image fraud detection system, with a focus on CNN architecture. The proposed method makes use of a CNN architecture with modifications in picture compression to produce good results. We use the difference between the original and recompressed images to train the model. The proposed technique can efficiently detect image splicing and copy-move types of image forgeries. The experiments results are highly encouraging, and they show that the overall validation accuracy is 92.23%, with a defined iteration limit. We plan to extend our technique for image forgery localization in the future. We will also combine the suggested technique with other known image localization techniques to improve their performance in terms of accuracy and reduce their time complexity. We will enhance the proposed technique to handle spoofing [50] as well. The present technique requires image resolution to be a minimum of 128 × 128, so we will enhance the proposed technique to work well for tiny images. We will also be developing a challenging extensive image forgery database to train deep learning networks for image forgery detection. Author Contributions: S.S.A. and I.I.G. designed and performed the experiments, conceptualization and methodology, and analyzed the data. S.S.A. and I.I.G. wrote the manuscript in consultation with N.-S.V., N.S., S.D.A. and N.W. All authors have read and agreed to the published version of the manuscript. Funding: This research has been funded by CY Cergy Paris Université (CY initiative via contract number 2019-067-C01-A0) and Khalifa University. Acknowledgments: We are thankful to CNRS for providing us the

GPU Cluster Jean Zay (http: //www.idris.fr/jean-zay/ accessed on 10 January 2022) to train our network. Conflicts of Interest: The authors declare no conflict of interest.

## REFERENCES

1. Xiao, B.; Wei, Y.; Bi, X.; Li, W.; Ma, J. Image splicing forgery detection combining coarse to refined convolutional neural network and adaptive clustering. Inf. Sci. 2020, 511, 172–191. [CrossRef]
2. Kwon, M.J.; Yu, I.J.; Nam, S.H.; Lee, H.K. CAT-Net: Compression Artifact Tracing Network for Detection and Localization of Image Splicing. In Proceedings of the 2021 IEEE

Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 5–9 January 2021; pp. 375–384.

3. Wu, Y.; Abd Almageed, W.; Natarajan, P. ManTra-Net: Manipulation Tracing Network for Detection and Localization of Image Forgeries With Anomalous Features. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 9535– 9544.
4. Ali, S.S.; Baghel, V.S.; Ganapathi, I.I.; Prakash, S. Robust biometric authentication system with a secure user template. Image Vis. Comput. 2020, 104, 104004. [CrossRef]
5. Castillo Camacho, I.; Wang, K. A Comprehensive Review of Deep-Learning-Based Methods for Image Forensics. J. Imaging 2021, 7,[CrossRef] [PubMed] Electronics 2022, 11, 403 16 of 17
6. Zheng, L.; Zhang, Y.; Thing, V.L. A survey on image tampering and its detection in real-world photos. J. Vis. Commun. Image Represent. 2019, 58, 380–399. [CrossRef]
7. Jing, L.; Tian, Y. Self-supervised Visual Feature Learning with Deep Neural Networks: A Survey. IEEE Trans. Pattern Anal. Mach. Intell. 2020, 43, 1