

Article

Image Splicing Location Based on Illumination Maps and Cluster Region Proposal Network

Ye Zhu ^{1,2,†} , Xiaoqian Shen ^{3,†}, Shikun Liu ¹, Xiaoli Zhang ^{3,*} and Gang Yan ^{1,*} 

¹ School of Artificial Intelligence, Hebei University of Technology, Tianjin 300401, China; zhuye@hebut.edu.cn (Y.Z.); 201932804013@stu.hebut.edu.cn (S.L.)

² Guangdong Key Laboratory of Intelligent Information Processing, Shenzhen University, Shenzhen 510900, China

³ College of Computer Science and Technology, Jilin University, Changchun 130012, China; shenxq2118@mails.jlu.edu.cn

* Correspondence: zhangxiaoli@jlu.edu.cn (X.Z.); yangang@hebut.edu.cn (G.Y.)

† Ye Zhu and Xiaoqian Shen are co-first authors of the article.

Abstract: Splicing is the most common operation in image forgery, where the tampered background regions are imported from different images. Illumination maps are inherent attribute of images and provide significant clues when searching for splicing locations. This paper proposes an end-to-end dual-stream network for splicing location, where the illumination stream, which includes Grey-Edge (GE) and Inverse-Intensity Chromaticity (IIC), extract the inconsistent features, and the image stream extracts the global unnatural tampered features. The dual-stream feature in our network is fused through Multiple Feature Pyramid Network (MFPN), which contains richer context information. Finally, a Cluster Region Proposal Network (C-RPN) with spatial attention and an adaptive cluster anchor are proposed to generate potential tampered regions with greater retention of location information. Extensive experiments, which were evaluated on the NIST16 and CASIA standard datasets, show that our proposed algorithm is superior to some state-of-the-art algorithms, because it achieves accurate tampered locations at the pixel level, and has great robustness in post-processing operations, such as noise, blur and JPEG recompression.

Keywords: image splicing blind forensics; illuminant maps; dual-stream; cluster RPN



Citation: Zhu, Y.; Shen, X.; Liu, S.; Zhang, X.; Yan, G. Image Splicing Location Based on Illumination Maps and Cluster Region Proposal Network. *Appl. Sci.* **2021**, *11*, 8437. <https://doi.org/10.3390/app11188437>

Academic Editor: Francesco Bianconi

Received: 11 August 2021

Accepted: 9 September 2021

Published: 11 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the increasing popularity of editing tools, image content can be easily and discretely edited. The forged images are transmitted, usually to pervert the truth, to scenes that require original images, such as news organizations, academics, courts of law, etc. The forgery incidents in modern society have become very frequent and bring serious negative results. Splicing is one of the most common image tampering operations, where one target is placed into another image in order to falsify reality. To cover up the traces of tampering, post-processing is often applied and this is very challenging for splicing blind forensics.

The traditional splicing forensics methods can be divided into two categories: equipment inconsistency and image attributes. The methods based on image equipment inconsistency use the discrepancy of the image device information, such as the Color Filter Array (CFA) [1], Error Level Analysis (ELA) [2], Noise Inconsistency (NI) [3]. With the development of image post-processing technology, the image device information is easy to edit, causing inaccuracy in detection of splicing. The methods based on the image attributes look for the inconsistency features between the original and tampered regions, such as the gray world [4], 3-D lighting environment [5], the two-color reflection model [6]. The principle of these methods are relatively simple, but they have limitations in judging the condition of the image.

Convolutional Neural Networks (CNN) have made significant contributions to the field of computer vision, which also provide a novel idea for image splicing forensics.

Zhang firstly proposed a two-stage deep learning detection approach, which included the stacked autoencoder model to learn the block features, and the fusion model to integrate the context information of each blocks [7]. Inspired by this, a first layer of CNN with the basic high-pass filter to calculate the residual maps is applied in a Spatial Rich Model (SRM), which can detect splicing and copy-move [8]. Bondi considered a CNN-based method to extract camera features of image patches and then to cluster to detect whether an image has been forged [9]. However, the above methods extract the features based on patches, where the location accuracy is patch-level.

In constrained image splicing localization, the Deep Matching and Validation Network (DMVN) [10] and Deep Matching network based on Atrous Convolution (DMAC) [11] were proposed to generate probability estimates and locate constrained splicing regions. Liu further introduced a novel attention-aware encoder-decoder deep matching network with atrous convolution, which had superior performance [12]. A combination of four features has been extracted and trained using a logistic regression classification model [13]. Ringed Residual U-net (RRU-Net) [14], Coarse to Refined Network [15] and Mask R-CNN [16] were directly applied for splicing location. Spatial Pyramid Attention Network (SPAN) is a self-attentive hierarchical structure, which constructs a pyramid of local self-attention blocks to locate tampered regions [17]. These blind forensics methods only extract the feature based on single-stream, where the tampered region location accuracy is relative low. In contrast with the single-stream, the dual-stream methods usually require image preprocessing to extract manipulating traces.

Salloum proposed the Multi-Task Fully Convolutional Network (MFCN) with image-based and edge-based streams to localize the splicing regions, which explored the validity of semantic segmentation framework in forgery detection [18]. Further, the dualstream with different input were proposed, such as spatial and frequency domains [19], RGB and noise images [20], RGB and SRM features [21]. A Skip connection architecture based on Long-Short Term Memory-Encoder Decoder (LSTM-EnDec-Skip) [22] is an improvement of LSTM-EnDec [23], which applied a jump connection structure to the multi-task network. In summary, the input of dual-stream framework always includes the RGB image and the other tampered attribute image. Even though the above methods can locate splicing tampered regions, the location precision still needs improvement and the tampered attribute trace is always hidden through post-processing.

The illumination maps are an inherent attribute of the image, which is difficult to modify but can help locate the splicing regions. We proposes a novel splicing blind forensics framework, where the image stream is used to extract semantic features, and the illumination stream is used to extract tampered features. The two stages are fused by Multiple Feature Pyramid Network (MFPN). A Cluster Region Proposal Network (C-RPN) is used to locate tampered regions, as shown in Figure 1. In general, our contribution can be summarized into the following three points:

- (1) The illumination maps are applied in the illumination stream to extract inconsistent lighting color features, which can prove the effectiveness of the illumination maps.
- (2) A Multiple Feature Pyramid Network (MFPN) is proposed for deep multi-scale dual-stream features fusion, which provides sufficient tampered features for the tampered region proposal.
- (3) Cluster Region Proposal Network (C-RPN) is proposed, where the spatial attention mechanism retains more position information, and clusters adaptively selects the anchor size.

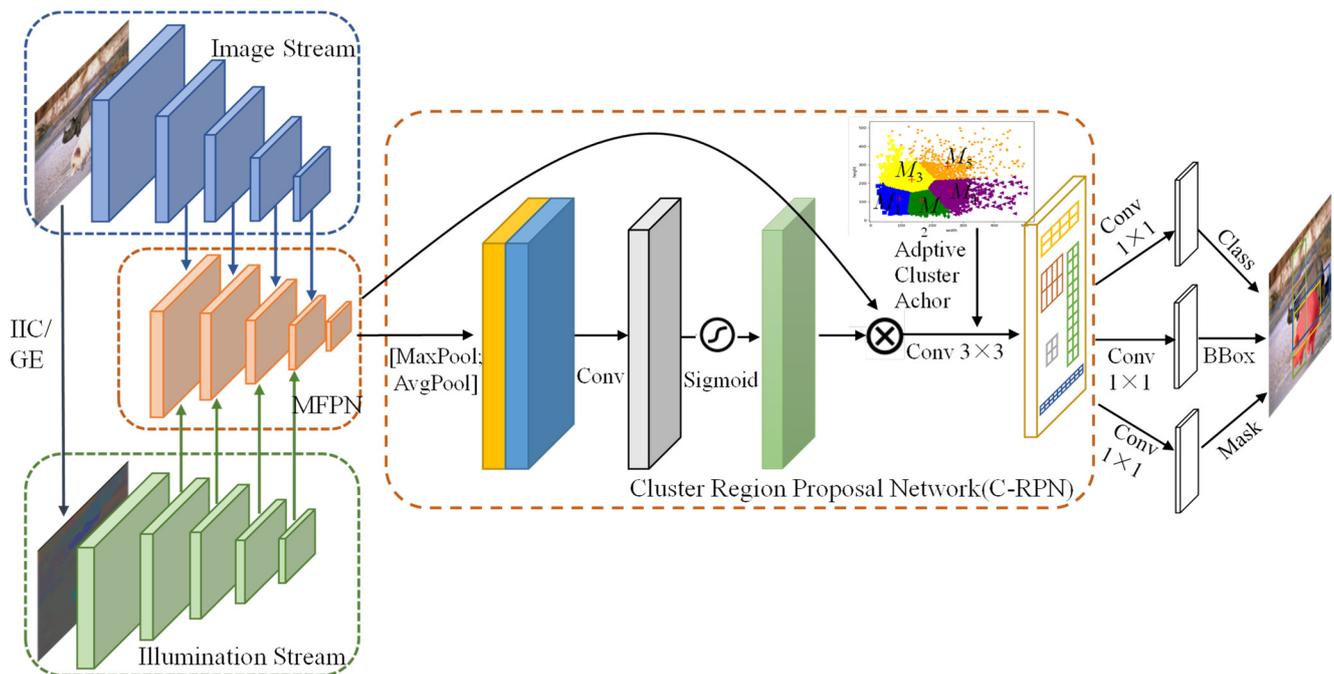


Figure 1. Overview of the proposed framework, where the image stream and the illumination stream extract semantic features and tampered features, respectively. The two stages are fused by MFPN and C-RPN is to locate tampered regions. \otimes denotes matrix multiplication.

The overall structure of this article is as follows: Section 2 briefly describes related work. In Section 3, we introduce the proposed model and the components of the model. Section 4 displays the experimental results and analyses. Section 5 summarizes the full text.

2. Related Works

Illumination maps can establish the illumination source, which are generally categorized into two classes: statistics-based and physics-based methods. Since the pristine and tampered regions are from different images, the inconsistency in illumination maps is a critical clue for splicing detection. The illumination color [24] and transformed spaces [25] are applied to detect image forgery, which can distinguish original and manipulated images in image level. Instead of traditional methods, the illumination maps with CNN can detect the splicing forgery at the pixel level [26,27], where the first step is to classify the pristine or fake image, and the second step is to locate the tampered region. However, the precision in locating the region is low. In order to make full use of illumination maps, we consider Grey-Edge (GE) [28] and Inverse-Intensity Chromaticity (IIC) [29] as the illumination stream, which can extract rich features with the RGB image stream.

Multiple scale feature fusion always consists of merging a low-level feature with a high-level feature map, which can capture global information. U-Net exploited lateral/skip connections that associate low-level feature maps across resolutions and semantic levels [30]. HyperNet concatenated features of multiple layers before predictions, but it brings additional calculations [31]. The Feature Pyramid Networks (FPN) designed horizontal connections to fuse the feature maps in multiple scale in bottom-up and top-down stages, which can obtain more robust semantic information [32]. Since the above multiple scale feature fusion methods are applied in single-stream framework, we propose a Multiple Feature Pyramid Network (MFPN) for the dual-stream feature fusion.

3. The Proposed Framework

With the help of the color inconsistency of the illumination maps and image semantic features, the tampered regions can be classified and located. As shown in Figure 1, the

framework can be divided into the following parts: the dual-stream framework, the multiple feature pyramid network and the cluster region proposal network. The image stream extracts features of unnatural tampering boundaries, while the illumination stream focuses on the inconsistent feature as a supplement. Then, the multiple feature maps of two streams are fused through MFPN. Finally, the Regions of Interest (ROI) are obtained by C-RPN and the tampered regions are located through data training.

3.1. The Dual-Stream Framework

3.1.1. Illumination Maps

The illumination map is an inherent attribute of the image and difficult to process uniformly, which can be considered as a major indicator for splicing forgery detection. GE and IIC are two state-of-art illumination maps, which are effective for mapping the tampered regions in splicing. GE assumes that the average reflection of the object surface has no color difference, which is easy to obtain and has low computational complexity. The process of GE illumination estimation $GE_{ill}(x)$ of pixel x can be formulated as:

$$GE_{ill}(x) = \frac{1}{k} \left(\int \left| \frac{\partial^n f^\sigma(x)}{\partial x^n} \right|_p dx \right)^{1/p} \quad (1)$$

where k is the scale factor, $|\cdot|$ represents the absolute value, $f^\sigma(x)$ denotes the intensity through Gaussian filter with kernel σ , p is the Minkowski norm, n is the order of the derivative.

The IIC applies the inverse intensity and chromaticity estimation to recolor each pixel, which has fewer surface colors and good robustness. The IIC illumination estimation $IIC_{ill}(x)$ on channel $c \in \{R, G, B\}$, is formulated as:

$$IIC_{ill}(x) = m(x) \frac{1}{\sum_{c \in \{R, G, B\}} p_c(x)} + \gamma_c \quad (2)$$

where $m(x)$ is a parameter that depends on the surface orientation, diffuse chromaticity and specular chromaticity; $p_c(x)$ is the color intensity, γ_c is the chromaticity on channel c .

3.1.2. The Image and Illumination Stream

The image stream with the input of $IMG(x)$ extract features of strong contrast differences, unnatural tampering boundaries, etc. However, the illumination stream with the input of $GE_{ill}(x)/IIC_{ill}(x)$ focus on the inconsistent feature as supplementary. The input of $IMG(x)$ and the $GE_{ill}(x)/IIC_{ill}(x)$ are shown in Figure 2, where the illumination maps of GE and IIC can reveal the tampered regions. The backbone of image and illumination streams are ResNet-50, where the feature maps are recorded as $\{C_1, C_2, C_3, C_4, C_5\}$ and $\{P_1, P_2, P_3, P_4, P_5\}$, respectively.

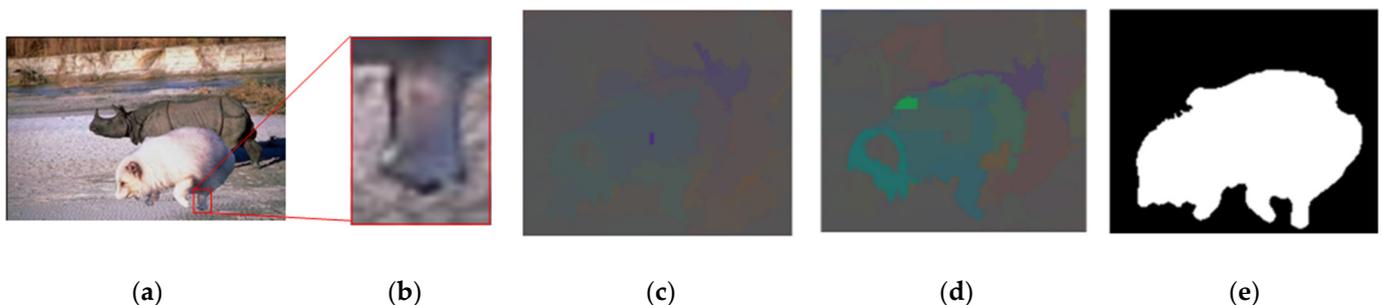


Figure 2. Traces of tampering with image and illumination stream input. (a) splicing image. (b) blown-up. (c) GE. (d) IIC. (e) Ground Truth.

3.2. The Dual-Stream Framework

Feature Pyramid Network (FPN) is a classical method to fuse pyramid features at all scales, which is not accommodated for dual-stream framework. Inspired by FPN, we propose Multiple Feature Pyramid Network (MFPN) for the two streams feature fusion, as shown in Figure 3. The MFPN has three pathways, namely the bottom-up pathway on the left, the bottom-up pathway on the right and the top-down pathway in the middle. Since the feature map of first layer has a larger resolution, the feature maps of second to fifth layers in image and illumination streams are applied. For the top-down pathway in the middle, the feature maps are fused by the image and illumination, recorded as $\{K_2, K_3, K_4, K_5, K_6\}$. The feature map of MFPN K_i is expressed as (3):

$$K_i = \text{Conv}(C_i) + \text{Conv}(P_i) + \text{Up}(K_{i+1}) \tag{3}$$

where $\text{Conv}()$ is the 1×1 convolution function to adjust dimension and connect horizontal, $\text{Up}()$ is the up-sampling function with the factor of 2.

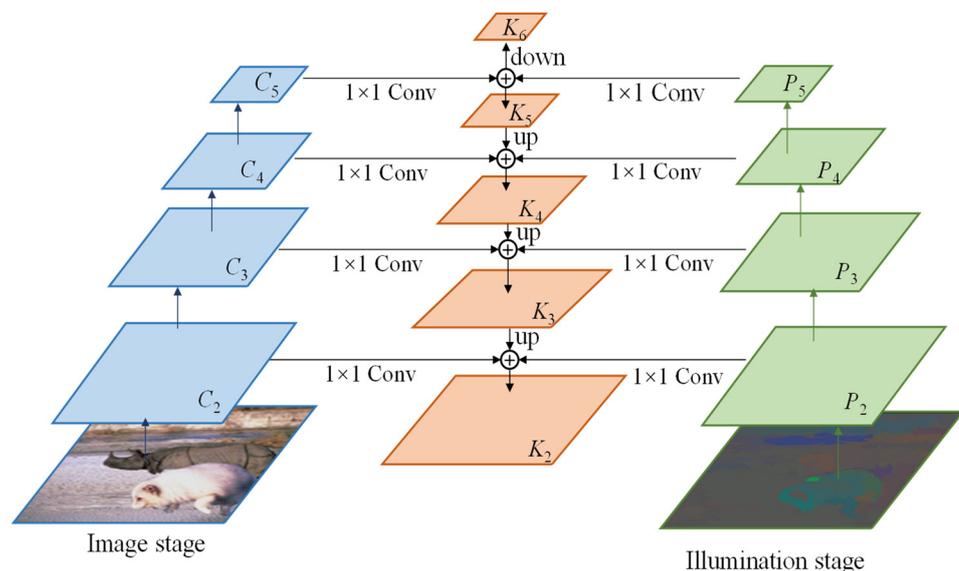


Figure 3. Multiple feature pyramid network.

3.3. Cluster Region Proposal Network (C-RPN)

The default anchor size in Region Proposal Network (RPN) has a large span, which are unreasonable for splicing location. A Cluster Region Proposal Network (C-RPN) is proposed, where the cluster anchor can adapt the varying sizes of tampered region and the attention module can focus on the tampered regions. The spatial attention feature map \bar{K}_i is defined as:

$$\bar{K}_i = K_i \otimes M_s(K_i) \tag{4}$$

where $M_s()$ is the spatial attention weight matrix, \otimes is matrix multiplication. To generate an adaptive anchor, the K-means cluster algorithm is performed to analyze the width and height of tampered region. Since the MFPN feature maps have 5 layers, the K-means cluster centers are initialized as 5. The maximum and minimum tampering regions are marked as M_{\max} and M_{\min} , respectively. The adaptive anchor size set S is expressed as (5):

$$S = \{S_1, S_2, S_3, S_4, S_5\} = \left\{ \frac{M_{\min} + M_1}{2}, M_2, M_3, M_4, \frac{M_5 + M_{\max}}{2} \right\} \tag{5}$$

where M_i is cluster center, the aspect ratios are set as $Q' = \{0.3, 0.5, 1, 2, 3\}$.

3.4. Training Loss

The candidate tampered regions acquire fixed-size feature map through Region of Interest (ROI) alignment. The processing of the network is divided into three branches, which are the classification of the fixed size feature map, the regression of the bounding box and the segmentation of the mask. The loss of classifying the real and fake L_{cls} is expressed as (6):

$$L_{cls} = -\log(p_u) \quad (6)$$

where p_u is confidence coefficient of the fake class. The loss for regression the bounding box L_{bbox} is computed as (7):

$$L_{bbox} = \sum_{i \in \{x, y, w, h\}} \text{smooth}_{L1}(t_i - v_i) \quad (7)$$

where x and y are the center position of ROI, w and h , respectively, indicate the width and height. The t_i and v_i are the label and predict of regression box i . Function $\text{smooth}_{L1}()$ is the improvement of L1 distance, which is computed as (8):

$$\text{smooth}_{L1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases} \quad (8)$$

The loss L_{mask} for the predicted mask y_{pred} and lable mask y_{label} is computed as (9):

$$L_{mask} = -y_{label} * \log(y_{pred}) - (1 - y_{label}) * \log(1 - y_{pred}) \quad (9)$$

In summary, the total loss for dual-streams is defined as Equation (10):

$$L = L_{cls} + L_{bbox} + L_{mask} \quad (10)$$

4. Discussion

4.1. Datasets and Evaluation Metrics

The training set is the Bappy synthetic dataset with 11k splicing tampered images [21], where the tampered regions are scaled and rotated in different factors. To compare with state-of-art methods, we also evaluated on standard datasets NIST16 [33] and CASIA [34]. NIST16 contains three tampered operations: splicing, copy-move and removal. CASIA is a commonly splicing dataset, where the ground-truth masks are obtained by the subtraction of tampered images and corresponding host images. This paper used CASIA2.0 for training and CASIA1.0 for testing. The division of the dataset is shown in Table 1.

Table 1. Comparison of image datasets.

Datasets	Tampered Images	Most Image Size	Number of Train and Test
Synthesized	11k	512 × 512	Train: 10k Test: 1k
NIST16	564	384 × 256	Train: 404 Test: 160
CASIA1.0 and 2.0	6044	384 × 256	Train: 5123 Test: 921

In order to compare the experimental results more objectively, this paper uses the average precision of the image-level index Average Precision (AP) and the pixel-level index F_1 score to quantitatively evaluate the model performance. The definition of the image-level index AP is the same as the evaluation standard of the CoCO dataset [35], which can be used as an evaluation indicator for blind evidence collection for splicing.

The pixel-level F_1 score can be used to evaluate the accuracy of the location of the tampered area. We calculate the F_1 score of each picture and then take the average as the final score for each data set. For each tampered picture, the definition of the F_1 is as follows:

$$F_1(M_{out}, M_{gt}) = \frac{2TP}{2TP + FN + FP} \quad (11)$$

where M_{out} and M_{gt} represent the final predicted mask and ground-truth mask, TP represents the number of tampered pixels that are correctly predicted, FP represents the number of tampered pixels that are incorrectly predicted, FN represents the number of genuine pixels that are incorrectly predicted.

4.2. Training Setting

The optimization of this experimental model adopts the random gradient descent method, and the input image is cropped to a size of 512×512 pixels, so as to avoid excessive image resolution and increase the calculation time. The batch size is 4, while the initial learning rate is 0.001 and reduces to 0.0001 after 25k iterations. The maximum number of iterations is 50k, and the training weight value on ImageNet is used as the initialization weight of the network. All the experiments are conducted on a machine with Intel (R) Xeon W2123 CPU @ 3.60 GHz, 32 GB RAM and a single GPU (GTX 1080Ti). The visualization of training loss for the IIC and GE are shown in Figure 4, which have converged in several iterations.

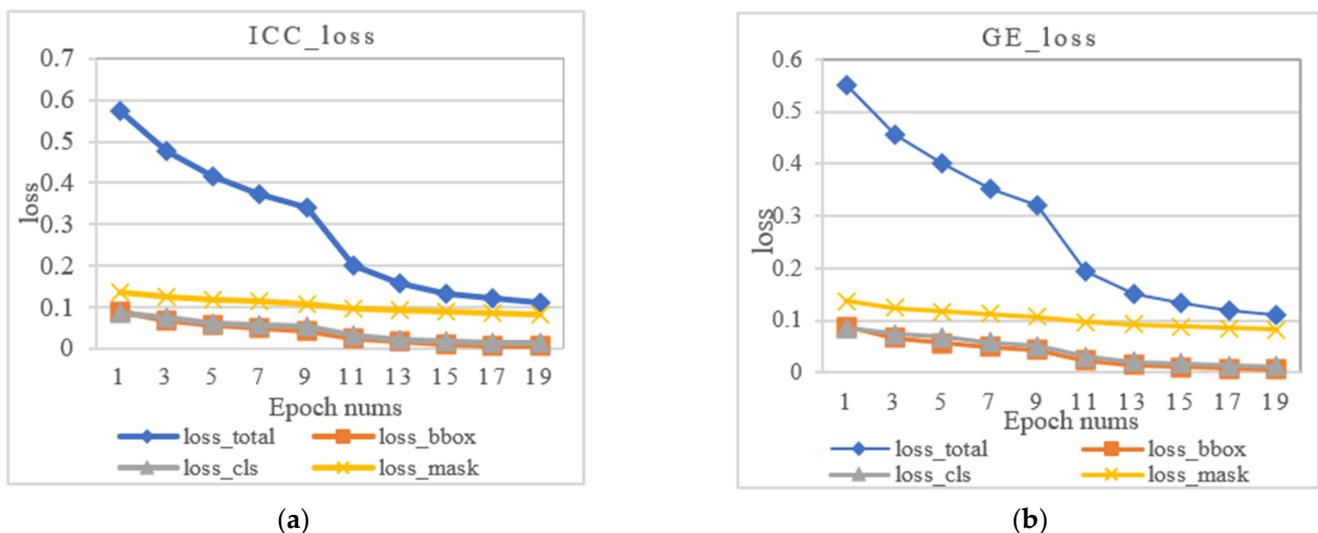


Figure 4. Visualization training loss in (a) IIC and (b) GE.

4.3. Experiments and Comparative Analysis

4.3.1. Ablation Experiments

In order to verify the effectiveness of the illumination maps in dual-stream and C-RPN, the ablation experiments with various frameworks are evaluated on synthetic dataset. The 'GE stream', 'IIC stream' and 'image stream' are, respectively, single stream for Mask R-CNN network with the input of GE, IIC and image. The 'Dual-stream (GE + Image)' and 'Dual-stream (IIC + Image)' are double streams with MFPN. The 'Ours (GE)' and 'Ours (IIC)' are proposed framework with the input of GE and IIC. The comparison results are shown in Table 2, where the bold entities denote the greatest performance. The AP score of dual-stream network with the input of 'IIC + Image' is almost 8.1%, 7.7% and 2.6% higher than the single-stream with the input of GE, IIC and image, which indicates that the illumination inconsistent feature and the image feature are complementary. In addition, our proposed method with 'IIC' is 3.6% higher than dual-stream with 'IIC + image', which indicates that

the application of MFPN fuses the low-level and high-level features to provide sufficient tampering features. In addition, the adaptive anchor size by applied C-RPN is conducive to return appropriate bounding boxes for improving splicing forgery location.

Table 2. Ablation experiments results of *AP* on synthetic dataset (%).

Synthetic Dataset	<i>AP</i>
GE stream	81.7
IIC stream	82.1
Image stream	87.2
Dual-stream(GE + Image)	89.1
Dual-stream(IIC + Image)	89.8
Ours (GE)	92.9
Ours (IIC)	93.4

4.3.2. Robustness Analysis

In this part, the robustness experiments of post-processing, such as noise, blur and JPEG compression, are evaluated on the CASIA and NIST16 dataset. The post-processing parameters are set as follows: the mean value is zero, and variances are 0.001, 0.005, 0.01 and 0.02 in Gaussian noise; the window size is 3×3 , and variance are 0.01, 0.1, 0.3 and 0.5 in Gaussian blur; the quality factors are 80, 70, 60 and 50 in the JPEG recompression, respectively. The robustness results of F_1 scores on noise, blur and JPEG are shown in Figure 5. Since the tampered trace features are hidden in JPEG compression and noise, the F_1 score has slightly decrease. However, the proposed methods with GE and IIC are robust to JPEG compression and noise within a certain range. For Gaussian blur, our network framework maintains great robustness. The effects of different robustness experiments on the IIC and GE are shown in Figure 6, which can be seen that the proposed dual-streams framework has great robustness on noise, blur, and JPEG compression.

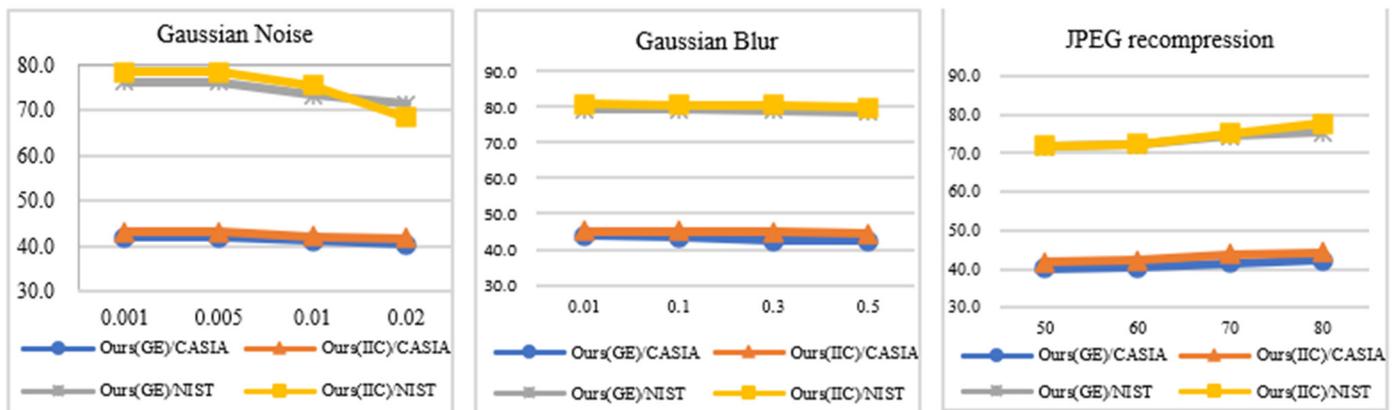


Figure 5. Robustness validation of F_1 scores on CASIA and NIST16 dataset.

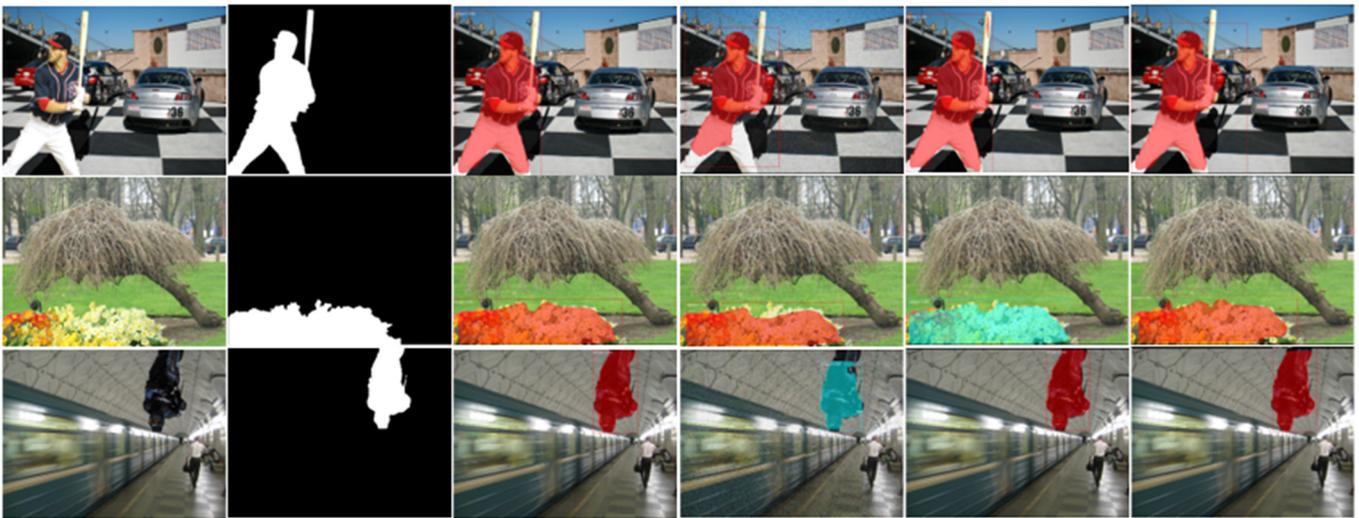


Figure 6. Exemplar visual results on NIST16 dataset. From left to right: the first two columns are tampered images and Ground-Truth, the third to fifth columns are results of the tampered images, tampered images with noise (0, 0.02), blur (3×3 , 0.5) and JPEG (50).

4.4. Experiments and Comparative Analysis

We compare the proposed framework with traditional methods CFA [1], ELA [2], NI [3] and CNN-based methods SPAN [17], RGB-N [19], LSTM-EnDec [21], LSTM-EnDec-Skip [22]. The $F1$ score comparison results are shown in Table 3, where the results of CFA, ELA, NOI1 comes from [19]. Among them, the dual-stream R-CNN uses the GE as ‘Ours (GE)’, and the as ‘Ours (IIC)’. The top performance values in the table are all expressed in bold font.

Our method is significantly better than other methods, and even reaches 81.0% on the NIST16 dataset when using IIC. Compared with RGB-N [19], the $F1$ score is increased by at least 2.9% using GE and 4.4% using ICC. This shows that the illumination maps (GE and IIC) can better represent the tampered area compared to the noise feature map. Since the LSTM-EnDec [21] and LSTM-EnDec-Skip [22] have no experiments on Columbia and NIST16 datasets, we compare the results on CASIA, which demonstrates that our method on IIC and GE has a better performance. However, our method does not reach the best performance on the Columbia dataset, since most tampering on Columbia occur on background. The visualization results are shown in Figure 7, where the first two rows are from the CASIA dataset and the last two rows are from the NIST16 dataset. It shows better localization of pixel-level by using illumination maps (IIC and GE) on the CASIA and NIST16 datasets.

Table 3. $F1$ score comparison on NIST16 and CASIA datasets (%).

Methods	Columbia	NIST16	CASIA
CFA [1]	46.7	17.4	20.7
ELA [2]	47.0	23.6	21.4
NI [3]	57.4	28.5	26.3
SPAN [17]	81.5	58.2	38.2
RGB-N [19]	61.2	72.2	40.8
LSTM-EnDec [21]	–	–	39.1
LSTM-EnDec-Skip [22]	–	–	43.2
Ours (GE)	73.1	79.4	43.9
Ours (IIC)	73.2	81.0	45.2

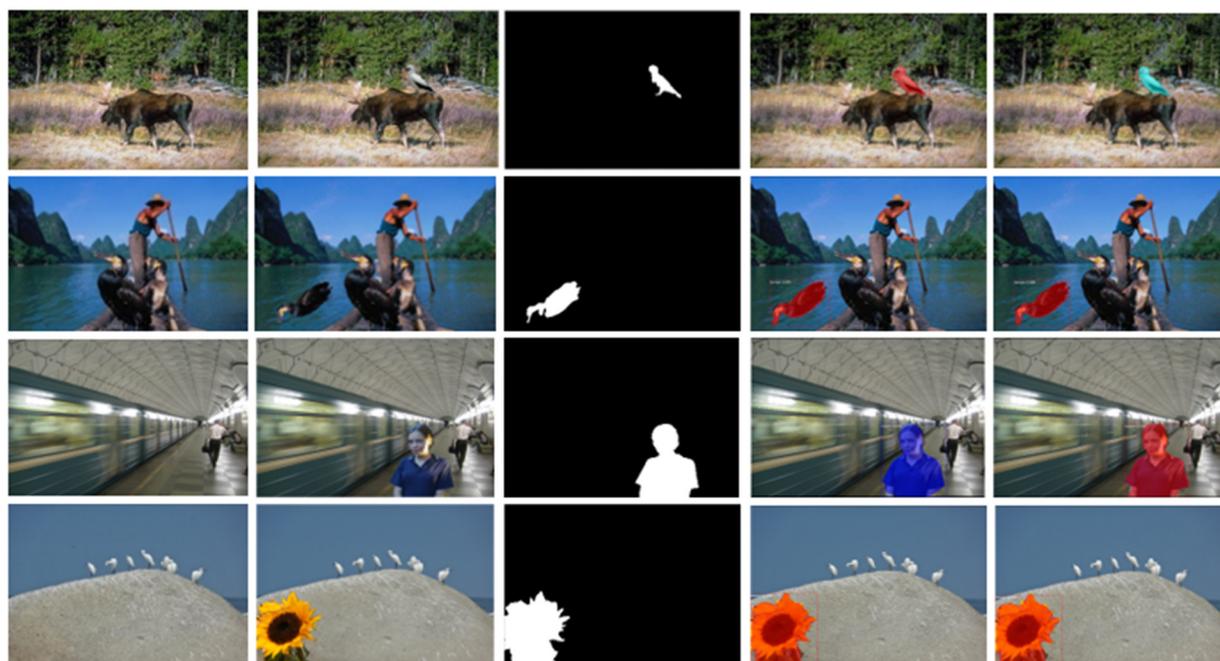


Figure 7. Examples of visual results on the CASIA and nist16 datasets. From left to right: the first three columns are original image, tampered images and Ground-Truth, the last two columns are the results of Ours (GE) and Ours (IIC).

5. Conclusions

We propose an end-to-end splicing location network, which includes the image stream and the illumination stream. The image stream extracts global features such as strong contrast differences and unnatural tampered boundaries. The illumination stream extracts inconsistent color illumination features in the IIC and GE illumination maps. In addition, the MFPN is used to fuse the multi-scale features of dual-stream network and C-RPN is proposed to generate candidate tampered regions with greater retention of location information. Extensive experiments, evaluated on the NIST16 and CASIA datasets, show that our proposed algorithm is superior to some state-of-the-art algorithms, which achieves accurate tampered location at the pixel level, and has good robustness to post-processing operations, such as noise, blur and JPEG recompression. However, the illumination maps are also effective in detecting splicing forgery. In the future, we will try to explore common features for image forensics.

Author Contributions: Conceptualization: Y.Z., X.S., S.L., X.Z. and G.Y.; methodology: Y.Z. and X.S.; software: Y.Z., X.S. and S.L.; writing—original draft preparation: Y.Z. and X.S.; writing—review and editing: X.Z. and G.Y.; supervision: X.Z. and G.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Major Program of the National Natural Science Foundation of China (Grant No. 91746207), the National Natural Science Foundation of China (Grant No. 6210071784 and Grant No. 61806071), the Natural Science Foundation of Hebei Province, China (Grant No. F2021202030, Grant No. F2019202381 and Grant No. F2019202464), the Sci-tech Research Projects of Higher Education of Hebei Province, China (Grant No. QN2019207 and Grant No. QN2020185), Key Research and Development Program of Xinjiang Province (Grant No. 2020B03001).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available on request due to restrictions. The data presented in this study are available on request from the corresponding authors.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ferrara, P.; Bianchi, T.; De Rosa, A.; Piva, A. Image Forgery Localization via Fine-Grained Analysis of CFA Artifacts. *IEEE Trans. Inf. Forensics Secur.* **2012**, *7*, 1566–1577. [[CrossRef](#)]
2. Krawetz, N.; Solutions, H.F. A pictures worth. *Hacker Factor Solut.* **2007**, *6*, 1–31.
3. Mahdian, B.; Saic, S. Using noise inconsistencies for blind image forensics. *Image Vis. Comput.* **2009**, *27*, 1497–1503. [[CrossRef](#)]
4. Youseph, S.N.; Cherian, R.R. Pixel and edge based illuminant color estimation for image forgery detection. *Procedia Comput. Sci.* **2015**, *46*, 1635–1642. [[CrossRef](#)]
5. Kee, E.; Farid, H. Exposing digital forgeries from 3-D lighting environments. In Proceedings of the 2010 IEEE International Workshop Information Forensics and Security, Seattle, WA, USA, 12–15 December 2010; pp. 1–6.
6. Francis, K.; Gholap, S.; Bora, P.K. Illuminant colour based image forensics using mismatch in human skin highlights. In Proceedings of the 2014 International Conference on Communications, Kanpur, India, 28 February–2 March 2014; pp. 1–6.
7. Zhang, Y.; Goh, J.; Win, L.L.; Thing, V. Image region forgery detection: A deep learning approach. *SG-CRC* **2016**, *2016*, 1–11.
8. Rao, Y.; Ni, J. A deep learning approach to detection of splicing and copy-move forgeries in images. In Proceedings of the 2016 International Workshop on Information Forensics and Security, Abu Dhabi, United Arab Emirates, 4–7 December 2016; pp. 1–6.
9. Bondi, L.; Lameri, S.; Guera, D.; Bestagini, P.; Delp, E.J.; Tubaro, S. Tampering detection and localization through clustering of camera-based CNN features. In Proceedings of the Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 1855–1864.
10. Wu, Y.; Abd-Almageed, W.; Natarajan, P. Deep matching and validation network: An end-to-end solution to constrained image splicing localization and detection. In Proceedings of the 25th ACM International Conference on Multimedia, Mountain View, CA, USA, 23–27 October 2017; pp. 1480–1502.
11. Liu, Y.; Zhu, X.; Zhao, X.; Cao, Y. Adversarial Learning for Constrained Image Splicing Detection and Localization Based on Atrous Convolution. *IEEE Trans. Inf. Forensics Secur.* **2019**, *14*, 2551–2566. [[CrossRef](#)]
12. Liu, Y.; Zhao, X. Constrained Image Splicing Detection and Localization with Attention-Aware Encoder-Decoder and Atrous Convolution. *IEEE Access* **2020**, *8*, 6729–6741. [[CrossRef](#)]
13. Jaiswal, A.K.; Srivastava, R. A technique for image splicing detection using hybrid feature set. *Multimedia Tools Appl.* **2020**, *79*, 11837–11860. [[CrossRef](#)]
14. Bi, X.; Wei, Y.; Xiao, B.; Li, W. Rru-net: The ringed residual u-net for image splicing forgery detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–17 June 2019; pp. 1–10.
15. 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](#)]
16. Ahmed, B.; Gulliver, T.A.; Alzahir, S. Image splicing detection using mask-RCNN. *Signal Image Video Process.* **2020**, *14*, 1035–1042. [[CrossRef](#)]
17. Hu, X.; Zhang, Z.; Jiang, Z.; Chaudhuri, S.; Yang, Z.; Nevatia, R. SPAN: Spatial Pyramid Attention Network for Image Manipulation Localization. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 312–328.
18. Salloum, R.; Ren, Y.; Kuo, C.C.J. Image splicing localization using a multi-task fully convolutional network (MFCN). *J. Vis. Commun. Image R.* **2018**, *51*, 201–209. [[CrossRef](#)]
19. Shi, Z.; Shen, X.; Kang, H.; Lv, Y. Image Manipulation Detection and Localization Based on the Dual-Domain Convolutional Neural Networks. *IEEE Access* **2018**, *6*, 76437–76453. [[CrossRef](#)]
20. Zhou, P.; Han, X.; Morariu, V.I.; Davis, L.S. Learning rich features for image manipulation detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 1053–1061.
21. Wu, Y.; Abd Almageed, W.; Natarajan, P. ManTra-net: Manipulation tracing net-work for detection and localization of image forgeries with anomalous features. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 9543–9552.
22. Mazaheri, G.; Mithun, N.C.; Bappy, J.H.; Roy-Chowdhury, A.K. A skip connection architecture for localization of image manipulations. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–17 June 2019; pp. 119–129.
23. Bappy, J.H.; Simons, C.; Nataraj, L.; Manjunath, B.S.; Roy-Chowdhury, A.K. Hybrid LSTM and Encoder–Decoder Architecture for Detection of Image Forgeries. *IEEE Trans. Image Process.* **2019**, *28*, 3286–3300. [[CrossRef](#)] [[PubMed](#)]
24. Riess, C.; Angelopoulou, E. Scene illumination as an indicator of image manipulation. In Proceedings of the International Workshop on Information Hiding, Calgary, AB, Canada, 28–30 June 2010; pp. 66–80.
25. Carvalho, T.; Faria, F.A.; Pedrini, H.; Torres, R.D.S.; Rocha, A. Illuminant-Based Transformed Spaces for Image Forensics. *IEEE Trans. Inf. Forensics Secur.* **2015**, *11*, 720–733. [[CrossRef](#)]
26. Pomari, T.; Ruppert, G.; Rezende, E.; Rocha, A.; Carvalho, T. Image splicing detection through illumination inconsistencies and deep learning. In Proceedings of the 2018 IEEE International Conference on Image Processing, Athens, Greece, 7–10 October 2018; pp. 3788–3792.
27. Mazumdar, A.; Bora, P.K. Deep learning-based classification of illumination maps for exposing face splicing forgeries in images. In Proceedings of the 2019 IEEE International Conference on Image Processing, Taipei, Taiwan, 22–25 September 2019; pp. 116–120.

28. Van De Weijer, J.; Gevers, T.; Gijsenij, A. Edge-based color constancy. *IEEE Trans. Image Process.* **2007**, *16*, 2207–2214. [[CrossRef](#)] [[PubMed](#)]
29. Tan, R.T.; Nishino, K.; Ikeuchi, K. Color constancy through inverse-intensity chromaticity space. *J. Opt. Soc. Am. A* **2004**, *21*, 321–334. [[CrossRef](#)] [[PubMed](#)]
30. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; Springer: Cham, Switzerland, 2015; pp. 234–241.
31. Kong, T.; Yao, A.; Chen, Y.; Sun, F. HyperNet: Towards Accurate Region Proposal Generation and Joint Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 845–853.
32. Lin, T.-Y.; Dollar, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 936–944.
33. Nist Nimble 2016 Datasets. Available online: <https://www.nist.gov/itl/iad/mig/nimble-challenge-2017-evaluation/> (accessed on 10 September 2020).
34. Dong, J.; Wang, W.; Tan, T. Casia image tampering detection evaluation database. In Proceedings of the 2013 IEEE China Summit and International Conference on Signal and Information Processing, Beijing, China, 6–10 June 2013; pp. 422–426.
35. Lin, T.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 740–755.