Data Augmentation with Illumination Correction in Semantic Segmentation

Honghui Liu, Changjian Wang and Yuxing Peng
PDL, Computer School, National University of Defense Technology, Changsha, China
Email: honghuiliu95@outlook.com; hangjianang@nudt.edu.cn; pengyuxing@aliyun.com

Abstract. In the training process of the deep learning network, data augmentation can effectively improve accuracy. Presently, the commonly used data augmentation methods have been mainly divided into space transformation and color transformation. However, these methods do not take into account the effects of the image’s illumination factor. Under natural conditions, the color, angle, and strength of the light will cause the color of the object to change, which is likely to cause segmentation errors. In this article, we start from the angle of illumination and perform data augmentation on the image. By applying the illumination correction algorithm based on Retinex theory, we are able to enhance the data and achieved good results in the experiment. However, there still remains some problems within the current algorithms based on Retinex theory. Especially for images with uniform illumination conditions, which can be overexposed after processed by these algorithms. In response to this problem, we have improved the relevant algorithm. The improved algorithm further improves the accuracy of semantic segmentation on the original basis.

Keywords. Data augmentation; semantic segmentation; illumination correction.

1. Introduction
In the field of computer vision, data augmentation can significantly improve the accuracy of visual tasks, such as segmentation, classification, and detection. As early as 2012, AlexNet [1] was proposed as a milestone in convolutional neural networks, applied data augmentation. The main data augmentation methods of the AlexNet include random cropping, horizontal flipping, and Gaussian perturbation of the principal components analyzed by the PCA method. Almost all subsequent convolutional neural networks have applied data augmentation [2, 3]. Moreover, has been an increasing emergence of data augmentation methods. These methods have been mainly divided into two categories, the first category is geometric transformation, and the second category is color transformation [4]. Data augmentation can increase the number and diversity of data sets, allowing the model to learn stable features. At the same time, it can also increase the difficulty of recognition during training, thereby enhancing the robustness of the model.

At present, there are two urgent problems in the field of semantic segmentation. (1) Intraclass error, that is, two different regions are predicted in the same object. (2) The classification of edge pixels. In order to extract high-level features, convolution neural network needs to reduce the size of feature map through pooling layer to increase the receptive field of convolution. However, this will reduce the resolution of the image and make the image, especially the edge of the object, lack a lot of information. Finally, it affects the accuracy of segmentation.
Illumination is one of the main reasons for the two problems earlier mentioned. For the first problem, there are shadow, color difference and other factors due to the nonuniform illumination in an object, resulting in different parts of the same object being identified as two or more different objects. For the second problem, the edge of the object is the boundary of the object. If the illumination condition here is poor, the edge contrast is not obvious, resulting in segmentation error. It can be seen that the training model enhanced by illumination algorithm can solve the above problems (figure 1).

![Figure 1: The light factor makes the edges of objects blurred and difficult to distinguish. We can clearly distinguish objects and backgrounds in light-corrected images.](image)

At present, there are many algorithms for illumination correction, such as gamma correction, HE equalization and so on. The most classic and widely used algorithms are based on Retinex theory [5]. The main idea of this theory is to divide an image into two parts: illumination and reflection. Illumination represents the illumination property in the image, and reflection represents the color property of the object in the image. Its core task is to calculate and separate illumination. But there are still many problems in all kinds of algorithms, the most obvious problem being the problem of overexposure within the image that has been processed by this kind of algorithm (figure 2).

For the image with poor illumination and low overall brightness, the existing algorithms can deal with the image well. However, for the image with better illumination conditions, the image may be overexposed after the existing algorithm processing. According to the existing algorithm, an improvement has been made so as the image can thus produce better results whilst processing the image with better lighting conditions.

In this paper, our main contributions are as follows: (1) Illumination correction is used as a data augmentation method for semantic segmentation. Extremely satisfying improvement results have been achieved in many models in many data sets. (2) We design a new algorithm based on Retinex theory, which can prevent the phenomenon of overexposure when processing images with better illumination conditions. We named the algorithm MSRBC (Multi Scale Retienx with Brightness Correction).

2. Related Works

Data augmentation is widely used in convolutional neural networks. In the field of computer vision, such as classification, detection, segmentation and other tasks, the application of data augmentation can greatly improve the accuracy. Previous data augmentation methods mostly use geometric transformation, such as scale, rotation, random crop, image mirroring and so on. In addition, there are many data segmentation methods based on color transformation, such as color shifting/whitening.
Almost every task and data set can be found clearly. Color transformation is mainly used in natural image data sets, such as cifiar-10, cifiar-100, Imagenet, Pascal VOC and so on.

Figure 2. For the image with better illumination condition, the existing illumination correction algorithm will make the image overexposed.

In practical applications, data augmentation mainly has two implementation methods. The first is when the number of data sets is relatively small. The images in the data set are supplemented to the data set by forming new images through the augmentation algorithm. In this way, the data set can be enlarged and the convolutional neural network can learn more abundant parameters. When the number of datasets is very large, the first method will double the number of data sets. Such a long training time is obviously unacceptable. At this time, you need to use the second data augmentation method. In each epoch, an image undergoes one or several random augmentation algorithms to generate an image for model training. This can enhance the robustness of the model and increase the accuracy.

We have found that there are many limitations in the existing data augmentation methods. Especially for images taken in natural light, changes in light intensity lead to changes in the color of image objects, which seriously affects the accuracy of segmentation. Therefore, we propose a data augmentation method based on light correction.

There are many algorithms for illumination correction, among which algorithms based on Retinex theory have become an important research direction of illumination correction algorithms due to their excellent effects. Retinex was proposed by Land and McCann in 1971 [5]. It is an image color theory based on the characteristics of the human visual system. When we observe an object, no matter what kind of lighting conditions the object is in, it is reflected in the same color in our visual system. This means that the color of an object is only related to the reflection characteristics of the object itself. When we analogize this view to an image, we can divide the image into two parts, one part is called the reflection component, this part represents the nature of the light reflected by the object itself, and is represented by \( R \) in this article. The other part is the illumination component, which is the characteristic of the illumination in the image, denoted by \( L \) in this article. According to the above theory, if we can calculate the illumination component, and then separate the illumination component from the overall image, then the remaining image is obviously under uniform lighting conditions. Calculating and separating the illumination component is the main task and main research direction of this type of algorithm.
3. Method

According to Retinex theory, a complete image \(I(x,y)\) can be divided into illumination \(L(x,y)\) and reflection \(R(x,y)\). We combine these two components by multiplication, that is

\[
I(x,y) = L(x,y) \cdot R(x,y)
\]

After Retinex theory was put forward, Land and McCann put forward the first algorithm to calculate the image illumination, namely the path based algorithm [6]. In the follow-up research, there are many algorithms based on this [7]. Later, the center/surround algorithm was proposed and wildly used [8], which mainly uses Gaussian filter to calculate the illumination.

\[
F(x, y) = \lambda e^{-\frac{(x^2 + y^2)}{2c^2}}
\]

where \(c\) is the surround size and \(\lambda\) satisfies the following equation:

\[
\iint F(x, y) dx dy = 1
\]

Algorithms based on center/surround operation include SSR, MSR, MSRCR, MSRCP and so on. SSR (Single Scale Retinex) [8] refers to a single scale algorithm calculated by Gaussian filtering. For the surrounding scale \(c\), the larger the scale, the better the overall effect, and the smaller the scale, the richer the local details. As the picture shows. MSR (Multi Scale Retinex) [9] adopts three different sizes of surround scales, so that it can be well balanced and take into account the whole and the part. Generally take the empirical value [12, 80, 250], from which we can further refine the formula:

\[
I(x,y) = I(x,y) \otimes F(x,y) \cdot R(x,y)
\]

\(\otimes\) represents convolution operation. In order to simplify the calculation, we transform it from the real field to the logarithmic field

\[
\begin{align*}
\log I(x,y) &= \log (I(x,y) \otimes F(x,y)) + \log R(x,y) \\
\log R(x,y) &= \log I(x,y) - \log (I(x,y) \otimes F(x,y))
\end{align*}
\]

In this way, we have completed the task of calculating and separating the light components. When we convert the data from the logarithmic domain back to the real domain, we usually do not use the power operation, that is, the inverse operation of the logarithmic operation. Instead, the following formula is used to map the pixel value from the logarithmic field to the real field, which is called SimplestColorBalance.

\[
R = \frac{\text{max}(\log R) - \text{min}(\log R)}{255}
\]

MSRCR (Multi Scale Retinex with Color Restoration) [10] solves the problem of color difference in MSR algorithm. However, it needs tedious empirical parameters.

MSRCP (Multi Scale Retinex with Chromaticity Preservation) [11] algorithm uses the Intensity channel

\[
Int = R + G + B
\]

The previous algorithms all calculate Gaussian convolution kernels in the three RGB channels. This greatly reduces the amount of calculation, while keeping the values of the three RGB channels at the original ratio. However, the algorithm cannot process images with better lighting conditions.

Through the previous discussion, we propose to use light correction as a method of data augmentation for convolutional neural networks. We found that the illumination correction algorithm based on Retinex theory still has many problems. For poor lighting conditions and the overall image color is dark, the algorithm based on Retinex theory can do a good job of correcting the image lighting conditions. But when this type of algorithm is applied to an image with a good original lighting condition and a brighter
overall image color, the result is not satisfactory. This will cause great color distortion, mainly manifested in the overall brightness of the image is too high.

We propose a new correction algorithm by studying the related algorithms. The core idea is to divide the image into the image with poor illumination condition and the image with better illumination condition. For images with poor illumination conditions, the existing illumination correction algorithms can achieve very good results. For this kind of image, the original algorithm remains unchanged. For images with good lighting conditions, the existing algorithms often produce bad results in the process of processing them. For this kind of algorithm, we add a correction coefficient in the process of its conversion back to the real number field, which can solve the problem of image brightness.

Our core task is to determine whether the brightness of an image is high or low, in other words, what kind of image needs to add a correction coefficient. For an image with better lighting conditions, its brightness value is a very important indicator. But even more significant is the change in the overall average brightness after removing the light component. If the change value is very significant, then we can judge that the original image has poor lighting conditions and the overall color is dark. It can be seen from figure 3 that the difference between the darker image and the brighter image is extremely obvious before and after the light component is removed. We define $D$ to represent the ratio of the average pixel value of the original image $I$ and the image $I'$ obtained after removing the illumination:

$$D = \frac{\text{mean}(I)}{\text{mean}(I')}$$

**Figure 3.** The pixel value of the traditional Retinex algorithm is concentrated in the maximum value. Retinex with brightness correction can solve this problem and prevent overexposure.

If the ratio is close to 1, then it can be determined that the brightness of the image is higher and the lighting conditions are better. If the value is close to zero, it means that after the illumination is removed, the brightness gap is large, and the overall color of the original image can be determined to be dark.

Through this ratio, we can roughly judge the lighting conditions of an image. For images with poor lighting conditions, we still use the original Retinex algorithm. For images with better lighting conditions, we multiply the resulting image pixel values by $D$.

The complete algorithm is shown in Algorithm 1. Flowchart for algorithm is shown in figure 4.
4. Experiment

4.1. Datasets
Our augmentation algorithm has achieved very good results in many datasets, especially in natural conditions. Pascal VOC, a well-known public data set of semantic segmentation, has 21 categories.
(excluding background). At the same time, we also tested our own wound image data set, mainly including diabetic foot and other wounds.

4.2. Models
We used three mainstream semantic segmentation networks, namely Unet [12], deeplabv3+ (ResNet101), and DFN (ResNet101). Unet has important significance in medical image segmentation and is the pioneering work of encoder-decoder. Deeplabv3+ [13] is currently a very powerful network in the field of semantic segmentation, and has achieved very good results in many public data sets. DFN [14] network is a discriminative feature network, which has achieved very good results in the field of semantic segmentation.

Table 1 shows the results we have achieved on PASCAL VOC. Table 2 shows the results achieved on our wound data set.

Table 1. Results on PASCAL VOC. Where T stands for tradition augmentation, IC stands for illumination correction augmentation. We used Deeplabv3+ and DFN for testing. Both networks use ResNet101 as the backbone.

| Network   | Augmentation | Boat  | Bus   | Cat   | Cow   | Dog   | Sofa  | Bird  | Car   | Chair | Train | mIOU  |
|-----------|--------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Deeplabv3+| T            | 0.688 | 0.947 | 0.939 | 0.862 | 0.870 | 0.445 | 0.871 | 0.880 | 0.397 | 0.890 | 0.783 |
| Deeplabv3+| T+IC         | 0.705 | 0.948 | 0.941 | 0.885 | 0.883 | 0.501 | 0.892 | 0.883 | 0.431 | 0.898 | 0.809 |
| DFN       | T            | 0.772 | 0.940 | 0.938 | 0.901 | 0.884 | 0.627 | 0.899 | 0.893 | 0.541 | 0.881 | 0.821 |
| DFN       | T+IC         | 0.825 | 0.964 | 0.936 | 0.940 | 0.913 | 0.695 | 0.925 | 0.920 | 0.600 | 0.929 | 0.856 |

Table 2. Unet uses VGGNet as the backbone, and Deeplabv3+ uses ResNet101 as the backbone.

| Network     | Augmentation | Wound |
|-------------|--------------|-------|
| Unet(VGGNet)| T            | 0.712 |
| Unet(VGGNet)| T+IC         | 0.763 |
| Deeplabv3+(Resnet101)| T    | 0.883 |
| Deeplabv3+(Resnet101)| T+IC | 0.894 |

5. Conclusion
Experiments show that the data augmentation method based on illumination correction can effectively improve the accuracy of semantic segmentation for a variety of tasks. Illumination correction algorithm as the capability to alleviate or remove the uneven illumination in the image. This effectively solves the problem of shadow and color difference caused by different light color and incident angle. And these problems widely exist within the image data set. Convolutional neural network can learn more abundant features from these images. In this way, researchers are able to access a more robust network. We propose a new illumination correction algorithm based on Retinex, which effectively solves the problem of applicability of the original algorithm. With the increasing of data set, the illumination correction algorithm can process all images. It is of great significance to optimize the feature of semantic segmentation network.

Acknowledgements
This work was supported in part by the Advanced Research Project of China (31511010203).

References
[1] Alex K, Ilya S and Geoffrey E H 2017 ImageNet classification with deep convolutional neural networks Commun. ACM 60 6 (June 2017) 84-90.
[2] Liu S and Deng W 2015 Very deep convolutional neural network based image classification using small training sample size 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), (Kuala Lumpur, Malaysia) pp 730-734 doi: 10.1109/ACPR.2015.7486599.

[3] Christian S, et al. 2015 Going deeper with convolutions 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Boston, MA, USA) pp 1-9 doi: 10.1109/CVPR.2015.7298594.

[4] Buslaev A, Iglovikov V, Khvedchenya E, Parinov A, Druzhinin M and Kalinin A 2020 Albumentations: Fast and flexible image augmentations Information 11 125.

[5] Land E H and McCann J J 1971 Lightness and Retinex theory J. Opt. Soc. Am. 61 (1) 1-11 doi: 10.1364/josa.61.000001.

[6] Land E H 1977 The Retinex theory of color vision Scientific American 237 108-128 https://doi.org/10.1038/scientificamerican1277-108

[7] Land E H 1986 Recent advances in Retinex theory Vision Research 26 7-21 http://linkinghub.elsevier.com/retrieve/pii/0016-0032(86)900672.

[8] Jobson D J, Rahman Z and Woodell G A 1997 A multiscale Retinex for bridging the gap between color images and the human observation of scenes IEEE Transactions on Image Processing 6 965-976 http://dx.doi.org/10.1109/83.597272.

[9] Jobson D J, Rahman Z and Woodell G A 1997 Properties and performance of a center/surround Retinex IEEE Transactions on Image Processing 6 (3) 451-462 doi: 10.1109/83.557356.

[10] Buchsbaum G 1980 A spatial processor model for object colour perception Journal of the Franklin Institute 310 1-26 http://dx.doi.org/10.1016/0016-0032(80)90058-7.

[11] Petro A-B, Sbert C and Morel J-M 2014 Multiscale Retinex Image Processing on Line 4 71-88 10.5201/ipol.2014.107.

[12] Ronneberger O, Fischer P and Brox T 2015 U-Net: Convolutional networks for biomedical image segmentation Lecture Notes in Computer Science 9351 https://doi.org/10.1007/978-3-319-24574-4_28.

[13] Liang-Chieh C, Yukun Z, George P, Florian S and Hartwig A 2018 Encoder-decoder with Atrous separable convolution for semantic image segmentation CoRR abs/1802.02611.

[14] Yu C, Wang J, Peng C, Gao C, Yu G and Sang N 2018 Learning a discriminative feature network for semantic segmentation 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (Salt Lake City, UT, USA) pp 1857-1866 doi: 10.1109/CVPR.2018.00199.