A Survey on Image Deblurring

ChuMiao Li
xihuamastermiao@163.com
Xihua University

Abstract

With the improvement of social life quality and the real needs of daily work, images are more and more all around us. Image blurring due to camera shake, human movement, etc. has become the key to affecting image quality. How to remove image blur and restore clear image has gradually become an important research direction in the field of computer vision. After more than half a century of unremitting efforts, the majority of scientific and technological workers have made fruitful progress in image deblurring. This article reviews the work of image deblurring and specifically introduces more classic image deblurring methods, which is helpful to understand current research and look forward to future trends. This article reviews the traditional image deblurring methods and depth-represented image deblurring methods, and comprehensively classifies and introduces the corresponding technical methods. This review can provide some guidance for researchers in the field of image deblurring, and at the same time facilitate their subsequent study and research.

1 introduction

Image deblurring aims to restore blurred pictures into clearer pictures, and has a wide range of application scenarios in human society and industrial production. Because image blurring is widespread in life, the problem of image deblurring has received attention and research since the last century. Since the 1960s, in order to solve the problem of image deblurring, it was first proposed to transform the image problem into the frequency domain to solve it, and classical algorithms such as inverse filtering [1] and Wiener filtering [2] were proposed. However, the image deblurring algorithm based on the frequency domain needs to know the degradation type of the blur accurately and is sensitive to noise, and then an estimation algorithm based on the spatial domain is proposed. Common estimation algorithms based on the spatial domain include: differential restoration algorithm [3], least square algorithm [4], maximum entropy algorithm [5] and so on. In recent years, the image deblurring algorithm has made greater breakthroughs. Over time, many methods and categories have been derived from the problem of deblurring. According to the method of deblurring algorithm, classification can be roughly divided into optimization-based deblurring method and deep learning deblurring method; classification according to input can be divided into: deblurring algorithm based on single frame input and deblurring based on multi-frame input Fuzzy algorithm; classification according to the degradation type can be divided into motion blur, Gaussian blur, defocus blur, turbulence blur, etc.; classification according to whether the blur kernel is known can be divided into blind deblurring and non-blind deblurring; according to whether the degradation is Consistency can be divided into space-changing blur and space-invariant blur.

Image deblurring is a classic problem in computer vision research. In recent years, image deblurring has become a basic topic in the field of computer vision, and it is also one of the key projects that can be applied. As shown in Figure 1, the so-called image deblurring refers to the restoration of the extreme visual experience based on the given blurred picture with poor look and feel. The process of good clear pictures. In recent years, with the rise of deep learning technology, the effect of image
Deblurring has made a qualitative leap. Its applications are mainly focused on space exploration, astronomical observation, material research, remote sensing and telemetry, military science, biological science, and medical imaging, traffic monitoring, criminal investigation and other fields.

Figure 1: Schematic diagram of image deblurring

The image blur process can be simulated as \( b = Ks + n \), where \( b \), \( s \), and \( n \) represent the vectorized blurred image, the clear latent image, and the additional noise, respectively. Therefore, based on the traditional deblurring method, recovering a clear picture under the condition of a known blurred picture is a highly unsuitable problem. However, the blurring can be removed by applying different constraints to the blur kernel and the latent image, or using hand-made regularization items at the same time. Unfortunately, these methods often lead to solving a non-convex optimization problem, so they cannot be well generalized to complex real examples. In recent years, in recent years, in [11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26], the methods based on learning representation have achieved great success in the field of image deblurring, and its objective indicators and subjective feelings have improved significantly. The deblurring network with a large number of learning samples can accurately estimate the fuzzy kernel and recover a clear picture, and this end-to-end model is easy to understand and easy to apply.

In general, this paper mainly introduces two main methods of image deblurring and general data sets for image deblurring. In Section 2, we mainly introduce traditional image deblurring methods, and in Section 3, we mainly introduce image deblurring with depth representation. Methods and their ideas. In Section 4, we mainly introduce several commonly used image deblurring datasets, and in Section 5, we mainly introduce image deblurring evaluation indicators. The main significance of this review is:

- Introduced and summarized several more classic traditional methods and depth representation methods for image deblurring tasks.
- This review provides some guidance for researchers interested in this direction.

## 2 Traditional image deblurring method

Early classic image deblurring algorithms are mainly Wiener filtering algorithm [2], RL filtering algorithm, total variation algorithm, etc. Because these methods fail to make full use of the prior information of natural images, the restored image is not accurate and mainly used Deblurring for non-blind images. In view of the shortcomings of the classic deblurring methods, the later traditional methods make full use of some prior information of the natural image to restore the image based on the classic method, such as the sparsity of the natural image, and overcome the defects of the original method. Generally, the prior information of natural images is mainly used to obtain the prior information related to natural images through a large number of statistical image characteristics, and use it as a regularization item to improve the deblurring effect of the image.

The model-based method is to model the fuzzy process first, and then solve the inverse process with the help of mathematical tools such as optimization methods. As mentioned earlier, the classic forms of traditional deblurring are:

\[
y = k * x + n \tag{1}
\]

Among them, \( n \) represents noise. We know the blurred image \( y \), and then estimate the original image \( x \) and the blur kernel \( k \) (blind deblurring), or know \( y \) and \( k \) to estimate \( x \) (unblind deblurring). The
solution of the inverse process is often an ill-conditioned problem, and the solution to this problem is often not unique.

\[ x, k = \arg\min_{x,k} \left[ \|x \otimes k - y\|_2^2 + \alpha(x) + \beta(k) \right] \] (2)

Later, a method based on regularization was proposed. The method based on regularization technology is to construct an optimized model as shown in equation 2 so that the restored image can make the above equation take the minimum value. Among them, \( x \) represents the final restored image, and \( k \) represents the estimated point spread function. Equation 2 The first term on the right side of the equal sign is called the fidelity term, which is used to ensure that the restored image can be consistent with the original blurred image. The second term \( \alpha(x) \) is called the regularization term of the restored image. The function of this term is to make the restored image meet a certain condition. This term is a function of the restored image \( x \). The third term \( \beta(k) \) is a regularization term about the point spread function, and its function is to make the estimated point spread function meet a certain condition. When the regularization term on the point spread function exists, this type of algorithm estimates the point spread function and the clear image at the same time. This type of algorithm is also called blind deblurring algorithm or blind deconvolution algorithm. When it is assumed that the point spread function is known, it is only necessary to estimate a clear image that minimizes the value of the image deblurring mathematical model. The change of the mathematical model of image deblurring is expressed in the following form:

\[ x = \arg\min_{x} \left[ \|x \otimes k - y\|_2^2 + \alpha(x) \right] \] (3)

Such algorithms are also called non-blind deblurring algorithms or non-blind deconvolution algorithms. The construction of the regularization term is the key part of image deblurring based on regularization. The regular term of the point spread function and the regular term of the restored image will have an important influence on the restoration result.

### 2.1 Maximum posterior method

In order to reduce the solution space of the problem and better approximate the real solution, we need to add prior conditions. Different a priori conditions will lead us into different solution spaces, and thus get different restored images. In the model-based method, the objective function of image deblurring is established using the method of maximum posterior probability under the Bayesian framework. However, model-based deblurring methods also have their limitations. The first is that most methods are based on the classic convolution model, aiming at globally uniform blur. For non-global uniform blur, such as image rotation blur caused by camera rotation around the Z axis, it is difficult to use a single blur kernel to describe. It is even more helpless for the local motion blur...
caused by the motion of objects in the scene. In addition, model-based methods generally use an iterative approach, using image pyramids to estimate blur kernels and remove blurs from small to large, which often requires a lot of time and overhead. Professor Levin believes that it is definitely not good to use the MAP objective function to solve k and x at the same time. There are the following reasons: First, the wrong unit convolution kernel is more likely than the correct sparse convolution kernel (proved in Levin’s paper); Secondly, the number of variables of the above-mentioned MAP objective function has a high degree of asymmetry, which means that the number of known quantities is always smaller than the number of unknown quantities. The above two reasons will lead to the use of this objective function of simultaneously estimating k and x to do MAP always can not get good results. How to solve this asymmetric problem is the key to breaking the game. Levin believes that a better way is to estimate k alone, so that the number of known quantities is much larger than the number of unknowns. To put it simply, for each possible k, we have to find the posterior probability of all possible x, and add these posterior probability values to get \( p(k|y) \).

In [27], the author found in experiments that inaccurately modeled image noise and errors in the estimated point spread equation are the main causes of ringing artifacts, not the Gibbs phenomenon previously thought. The author thinks Gradient-based optimization is the most feasible way to optimize such a log-prior, and gradient-descent is known to be neither efficient nor stable for a complex energy function containing thousands or millions of unknowns. The author proposes a new representation by concatenating two piece-wise continuous functions to fit the logarithmic gradient distribution. Secondly, the author use the blurred image to constrain the gradients of the latent image in a fashion that is very effective in suppressing ringing artifacts. A new smoothness constraint that we impose on the latent image in areas where the observed image has low contrast. This constraint is very effective in suppressing ringing artifacts not only in smooth areas but also in nearby textured ones. The effects of this constraint propagate to the kernel refinement stage, as well. The author combines the above-mentioned global image prior and local image prior to form an image prior.

In [28], the author proposes a unified method for estimating the motion blur filter from the perspective of transparency. At the same time, the author propose an optimization method to estimate the blur filter by solving a maximum a posteriori (MAP) problem only using transparency information. Without directly taking all pixel colors and complex image structures into computation, this method is efficient and robust. The transparency definition in a camera motion blurred image regarding the region blending is different from that on a motion blurred object. Assuming that the moving object occludes static background behind, in the object motion blur, the transparency map can be uniquely determined. The new transparency map can be defined differently, so the author call it the generalized transparency. The author uses this prior knowledge of clarity to solve the maximum posterior probability of the fuzzy kernel.

In [29], the author propose an additional term in the sparse prior that uses a color model built from local color statistics of the sharp latent image. This overcomes over-smoothing as it allows for sharp edges as long as they are consistent with local color statistics. In contrast with a gradient prior, which prefers the lowest intensity edges that are consistent with the observed blurred image, a two-color model can result in higher intensity edges if such edges are more consistent with local color statistics.

In [30], the novelty is to use image priors \( P(u) \) that are more heavy-tailed than Laplace distribution and apply a method of augmented Lagrangian to tackle this non-convex optimization problem. In the past, the estimation of fuzzy kernel based on MAP method mostly failed. Although the actual deblurring effect of this method is not satisfactory, the enlightening

\[2.2 \text{ Variational Bayes Method}\]

It is not easy to find the above marginal probability, and the amount of calculation is very complicated. So many scholars have adopted an approximate method to solve the problem. Fergus [32] et al. proposed the first non-parametric motion blur kernel estimation method in 2006, which is also the earliest blind image deblurring method. By learning the mixed Gaussian image prior and the mixed exponential fuzzy kernel prior, the author proposes an image non-parametric deconvolution method based on the VB criterion. In the past, the estimation of fuzzy kernel based on MAP method mostly failed. Although the actual deblurring effect of this method is not satisfactory, the enlightening
significance brought by it is far-reaching. On the one hand, using the posterior point estimation criterion of MAP (which is largely equivalent to the regularization method), the mixed Gaussian image prior is completely unsuitable for the blind deblurring problem, which is not conducive to the effective identification and estimation of saliency edges; another On the one hand, compared with the point estimation of MAP, the posterior mean estimation of VB has natural robustness, which can reduce the risk of "significant edge-non-parametric fuzzy kernel" alternate iteration estimation falling into a local minimum [33]. In addition to the enlightening elaboration of this article, the follow-up literature [34, 35, 36, 37] has proved the superiority of the image deblurring method based on VB compared with the MAP method from the perspective of theory and practice. In the paper, Fergus showed its complete blind deconvolution process: First, preprocess the image. In order to reduce the amount of calculation and get a good result, the user is required to select an image block; secondly, the use of variational Bayes, Estimate the convolution kernel $k$. In order to avoid falling into the local optimum, the author adopts the coarse-to-fine strategy; finally, the standard non-blind deconvolution method is used to reconstruct the clear image $x$.

In [32], the author found that recent research in natural image statistics, which shows that photographs of natural scenes typically obey very specific distributions of image gradients. The author uses a zero-order Gaussian mixture model to represent the distribution of the gradient size. The reason for choosing this representation is that it can provide a good approximation of the empirical distribution and allow the algorithm to have a processable estimation procedure. The author believes that the objective function of MAP tries to minimize all gradients (even larger gradients), and we expect natural images to have some larger gradients. In addition, the author found that the MAP objective function is easily affected by local minima.
In [38], the author further explored the reasons for the failure of the MAP method, and concluded in the $MAP_{x,k}$ problem we can never collect enough measurements because the number of unknowns grows with the image size. The author believes that the number of variables of the MAP objective function has a high degree of asymmetry. It can be seen that the number of known quantities is always less than the number of unknowns. At the same time, the author proved for an increasing image size, a $MAP_{k}$ estimation of $k$ alone (marginalizing over $x$) can recover the true kernel with an increasing accuracy. Moreover, the method of obtaining the variational Bayesian approximation through quantitative analysis is indeed significantly better than the existing methods.

In [33], the author systematically analyzed and demonstrated the basic principles of the VB method and the MAP method, demonstrated that rigorous evaluation of VB and its associated priors cannot be separated from implementation heuristics, and we have meticulously examined the interplay of the relevant underlying algorithmic details employed by practical VB systems. At the same time, it further analyzes the prior optimal selection problem of the VB blind deblurring method theoretically and points out that to some extent, Jeffreys prior can achieve the best blind deblurring effect.

![Figure 4: Classification based on traditional image deblurring methods](image)

### 3 Depth representation method

With the rapid development of deep learning in recent years, using a data-driven approach to solve the problem of blind deblurring of images seems to be another normalized method after VB and MAP methods. Simply put, the focus of current data-driven blind deblurring methods often lies in how to build a supervised deep neural network with good generalization performance. It is obvious that the performance of this method is mainly different from the rationality of the convolutional neural network architecture design and the rationalization of training. Specifically, many highly representative models emerged in the ImageNet ILSVRC Image Recognition Challenge, which largely reflects the relevant progress in the early stage of the development of Deep Representation, including: AlexNet, ZFNet, GoggleNet, VGGNet, ResNet, ResNeXt, etc. The famous international academic conferences such as ICCV, CVPR, ECCV, etc. successively held different computer vision tasks such as facial recognition, object detection, semantic segmentation, etc., also provide a new idea for the network construction of image deblurring. As far as the current development trend is concerned, we mainly introduce multi-scale de-fuzzing networks and the relatively hot generation of confrontation de-fuzzing networks and unsupervised de-fuzzing networks in recent years.

Jian Sun et al. [26] first applied the deep learning method to blind image deblurring, and proposed a deep learning method to predict the probability distribution of motion blur at the patch level using convolutional neural networks. The author uses a well-designed image rotation to further expand the set of candidate motion blur kernels predicted by CNN. Then use the Markov random field model to infer a dense non-uniform motion blur field, thereby enhancing the smoothness of the motion. Finally, the patch-level image prior non-uniform deblurring model is used to remove motion blur.

Ayan Chakrabarti et al. [39] applied the complex Fourier coefficients of the output deconvolution filter of the training network to the input patch, and used multi-resolution frequency decomposition to encode the input patch, and based on the locality in frequency (similar to the receiving space Convolutional layer with locality restriction) to limit the connectivity of the initial network layer.
This leads to a significant reduction in the number of weights that need to be learned during the training process, which proves to be crucial because it allows us to successfully train a network that runs on large patches and can therefore explain large fuzzy kernels.

3.1 Multi-scale deblurring network

Multi-scale network design has been very useful in other areas of image tasks. For example, the article [40] and [41], the idea of multi-scale is applied to the field of target detection, such as the [42] and [43], which puts the idea of multi-scale into practice. In the work of removing rain from images, such as [44], the multi-scale method has been successfully applied to the field of image segmentation. The success of multi-scale thinking in other image tasks makes it natural to migrate to the field of image deblurring.

Seungjun et al. [35] based on deep learning methods and traditional optimization methods have proposed to solve the problem of non-uniform blur of the image, but for the blur problem of dynamic scenes, that is, the problem of blurring only partial areas in the image is difficult to solve. Through the analysis and summary of the previous algorithms, the following problems are summarized: First, for the data set, it is difficult to obtain the measured clear and blurred images for training; second, for the deblurring problem in dynamic scenes, it is difficult to obtain the blur of the partial images. Nuclear; Finally, a larger receptive field helps to solve the problem of image deblurring. In order to simulate the traditional optimization method from coarse to fine, the author designed a multi-scale deblurring model. As the name suggests, in the first stage, the author downsampled the original scale of the blurred image to 1/4 of the original size and input it into the network to obtain the largest restoration result; through the up-convolution operation, the result obtained at the 1/4 scale stage was compared with the original. The fuzzy images of the 1/2 scale of the image are combined as the input of the second stage, and the deblurring result of the second stage is obtained through the network; the output of the second stage is convolved and combined with the original size of the blurred image as the input. Get the finest clear image through the network model. This multi-scale strategy is similar to the coarse-to-fine strategy in traditional fuzzy kernel estimation. It decomposes complex problems and restores them step by step. First, restore large-scale information at low resolution, and then high-resolution details can be added progressively.
resolution. Recover detailed information. This design simplifies the problem while increasing the field of perception of the image.

Later, Xin Tao et al. [21] proposed that because each scale solves the same problem, the same model parameters can be used. This cross-scale sharing of network weights can significantly reduce the difficulty of training, while introducing obvious stability benefits. Such a design significantly reduces the number of trainable parameters, and the recycling of shared weights works similarly to using data multiple times to learn parameters.

Hongyun Gao et al. [23] found that not all module parameter sharing is conducive to the improvement of defuzzification performance, and thus proposed the general principle of selective parameter sharing, which can be conducive to the design of future defuzzification systems. Since in resnet [35] and reblocks [34], the short-term jump connection of adding the input to the output after two or more convolutional layers produces superior effects in object detection, deblurring and super-resolution, The author proposes a nested jump connection structure, which corresponds to the high-order residual learning of the conversion module in the defuzzification network.

Hongguang Zhang et al. [22] believe that multi-scale networks use a large number of training parameters to cause too long training time, and increasing the network depth of very low-resolution input does not seem to improve the deblurring performance, so they proposed a deep stacking Multi-patch network. The author proposes an end-to-end CNN hierarchical model similar to spatial pyramid matching. Compared with the scaling of the multi-scale model on the original size, the model directly cuts the original size picture into each small patch, and then the small In the next stage, the patches are spliced two by two into one large patch, which is carried out in stages. The model performs deblurring in fine to coarse meshes, thereby using multi-patch local to coarse operations. Each finer level functions as a residual, by contributing its residual image to a coarser level, allowing the network of each level to focus on a different blur scale.

Zamir SW et al. [9] proposed a multi-stage progressive image deblurring network. The author adds Channel Attention Block (CAB) to the feature extraction at each stage, which effectively extracts useful features. The channel attention module (CAM) has long shown an unparalleled role in various tasks of computer vision. For example, [45] introduces channel attention to image classification tasks, [46, 44, 47] force the introduction of image segmentation tasks. Instead of simply cascading multiple stages, the author merges a supervisory attention module between every two stages. The supervisory attention module rescales the feature map of the previous stage, which can suppress the features with less information in the current stage and only allow useful features to propagate to the next stage. At the same time, the supervisory attention module provides a ground-truth supervisory signal, which is of great significance to the progressive image restoration at each stage. In addition, the author also proposed a cross-stage feature fusion mechanism, which makes the multi-scale features of the previous stage help enrich the features of the next stage, and also makes the network less susceptible to information loss. It is worth noting that the network architecture proposed in this article is not only effective in image de-modulation tasks, but also effective in image de-fogging, image de-raining, and image de-noising tasks.

3.2 Generative adversarial deblurring network

In 2014, Ian Goodfellow [48] and his colleagues at the University of Montreal introduced Generative Adversarial Networks (GAN). This is a new method of learning the basic distribution of data, so that the generated artificial objects can achieve amazing similarity with real objects. The idea behind GAN is very intuitive: two networks of generator and discriminator play against each other. The goal of the generator is to generate an object (such as a photo of a person) and make it look the same as it really is. The goal of the discriminator is to find the difference between the generated result and the real image. Then, the method of adversarial learning is widely used in image tasks, such as the application of GAN network to segmentation tasks in [49].

Orest Kupyn et al. [50] were inspired by the work on image super-resolution [36] and the image-to-image conversion in the generative adversarial network [37], and used image deblurring as a special case of this image-to-image conversion. Propose a de-fuzzy network that generates confrontation. The author constructed a generative adversarial network, trained a generator and a discriminator. In the training phase, the author not only trains the generator network but also introduces the discriminator network, and trains the two networks in a confrontational manner. The author hopes that through this
alternate training method, the performance of the generator and the discriminator can be continuously improved, just like the fact that both sides can improve when competing with a good opponent. The author uses perceptual loss and confrontation loss as the total loss. The perceptual loss generally focuses on the recovery of general content and the confrontation loss focuses on the recovery of details. Compared with the L1 loss and L2 loss, the perceptual loss is a simpler L2 Loss is the loss of the feature map. In the test phase, a clear picture can be obtained by inputting the fuzzy picture into the trained generator network.

Orest Kupyn et al. [13] proposed a new framework of deblurgan-v2. Compared with [50], its generator adopts a feature pyramid structure, and the discriminator uses a relativity discriminator. This feature pyramid model rebuilds a higher spatial resolution from the semantically rich layer through a top-down path. The horizontal connection between the compartments complements the bottom-up and top-down paths with high-resolution details and helps to locate objects. The author will take five final feature maps of different scales as output, and integrate all layers of information to improve the deblurring performance. Because the low-level feature semantic information is relatively small, but the target location is accurate; the high-level feature semantic information is richer, but the target location is relatively rough. In addition, although some algorithms use multi-scale feature fusion, they generally use the fused features to make predictions. The difference in FPN is that the predictions are performed independently at different feature layers, so the FPN structure undergoes different times of convolution and other processing. You can get different feature layer information.

3.3 Unsupervised neural network

Whether it is generating confrontational image deblurring networks or multi-scale image deblurring networks, there is a common problem-long training time and large amount of training data. Even if a powerful network model does not have sufficient training data, it cannot have a strong deblurring performance. As we all know, the traditional image deblurring algorithm can still achieve a good deblurring effect without a large amount of data support. How to bridge the gap between traditional algorithms and deep learning, build a bridge between traditional methods and deep learning methods, and build unsupervised neural networks has become a frontier issue. Dongwei Ren et al. [51]
proposed the use of a deep and bright neural network, which can estimate the blur kernel and clear pictures under the condition of a known blur kernel size. Given a fuzzy picture and the size of the blur kernel, the author uses the known image size and the size of the blur kernel, and uses the two noises of the corresponding size as the input of the clear image generation sub-module and the blur kernel generation sub-module. By introducing DIP [52] and FCN respectively to capture the priors of clean images and blur kernels, the clear images and blur kernels can be recovered better. Use the obtained clear image and blur kernel and convolve the two to calculate the loss of the blurred image, and then update the parameters of the two sub-networks. After multiple iterations, a clear image generation sub-module and a blur kernel generation sub-module with excellent performance can be obtained. After a lot of experiments, the author found that the effect of joint optimization of the two sub-modules is significantly better than the effect of alternate optimization, and attributed it to the unconstrained blind deconvolution problem is a highly non-convex problem, so the alternate optimization method will make the optimization fall into Saddle point. This unsupervised method undoubtedly provides us with another brand new idea, an unsupervised idea, to solve the problem of image deblurring. Compared with the previous deep learning methods that require a large amount of data and a longer training time, it is undoubtedly a more advantageous party. However, this method must retrain the network for each single image and need a given blur kernel size, which appears to have certain limitations.

Deep learning methods

Nah et al. [35]
Gao et al. [21]
DMPHN [22]
MPRNet [9]
Gao et al. [23]
DBGAN [50]
DeblurGANv2 [13]
D Ren et al. [51]

Figure 7: An image deblurring classification based on deep learning

4 Image deblurring dataset

In this part, we will introduce some commonly used datasets in the field of image deblurring.

4.1 GoPro dataset

The GoPro dataset [35] is the most frequently used data set in deep learning methods. It includes 3214 pairs of clear pictures and fuzzy pictures, including 2103 pairs in the training set and 1111 pairs in the test set.

4.2 RealBlur dataset

RealBlur dataset [53] consists of two subsets that share the same image content, one of which is generated from the original image of the camera, and the other is generated from the JPEG image processed by the camera’s ISP. Each subset provides 4,556 pairs of blurry and true sharp images of 232 pairs of low-light static scenes.

4.3 HIDE dataset

The HIDE dataset [54] 8422 extensively annotates 65,784 human bounding boxes for clear pictures and blurred pictures. These images are carefully selected from 31 high fps videos, which not only include real outdoor scenes, but also different figures, poses and people appearing at different distances. The data set can also be divided into two categories, including long shots (HIDEI) and ordinary pedestrians (close-up, HIDEII).
4.4 Levin dataset

Levin dataset [38] is the earliest fuzzy image dataset, which contains 32 grayscale pictures with 8 different blur kernels of 4 images and the corresponding clear pictures.

Figure 8: Classification of datasets

5 Image quality evaluation

Image evaluation index has been widely used in other fields of image work. Image quality evaluation is one of the basic techniques of image processing. Image quality evaluation is one of the basic techniques of image processing. It is an important indicator to reflect the quality of an image. It is mainly divided into subjective quality evaluation and objective quality evaluation. The subject of subjective quality evaluation is people, who score according to their viewing experience. This evaluation method is the most accurate and in line with human perception. The disadvantage is that the evaluation is troublesome and inefficient. Objective quality evaluation is based on image statistical information, so it has relatively many applications and can be divided into full reference, semi-reference and no reference.

5.1 Subjective quality assessment

The subjective quality scoring method is the most representative subjective evaluation method of image quality, which judges the image quality by normalizing the scores of the observers. The subjective quality scoring method can be divided into two types: absolute evaluation and relative evaluation.

The International Telecommunication Union (ITU) has proposed a variety of criteria for subjective evaluation methods, three of which are most commonly used, namely, the dual-stimulus damage grading method, the dual-stimulus continuous quality grading method, and the single-stimulus continuous quality grading method: Double stimulus damage grading method, given two sets of images of the original image (undistorted reference image) and the image to be tested (with certain distortion), compare the two sets of images to observe the damage of the image to be tested, according to the subjective quality of the image A 5-level scoring table to select the level of the image to be tested. The dual-stimulus continuous quality grading method is also given two sets of images, the difference is that the observer has no idea which is the reference image and which is the distorted image. The observer only needs to score the reference image and the test image separately according to the scoring table. Finally, the average value of the reference image and the image to be tested is calculated, and the difference between the two is calculated. The smaller the difference, the better the quality of the image to be tested. The single-stimulus continuous quality grading method is similar to the absolute subjective evaluation. The observer continuously scores the images to be tested according to the scoring table, and obtains the quality evaluation of the images to be tested according to the scoring and scoring time.

5.2 Objective quality assessment

Due to the hot development of deep learning in the field of image deblurring, objective image quality evaluation based on full reference has become the mainstream. The objective evaluation method of image quality is to establish a mathematical model based on the subjective visual system of the human eye, and calculate the image quality through a specific formula. There are many traditional image objective quality evaluation indicators, among which MSE and PSNR are the most well-known.
5.2.1 MSE

Mean squared error is generally used to detect deviations between predicted and true values.

\[ MSE = \frac{1}{m} \sum_{1}^{m} (y_i - \hat{y}_i)^2 \]  

(4)

where \( m \) represents the number of pixels, \( y_i \) represents the real value of \( i \) pixel, and \( \hat{y}_i \) represents the predicted value of \( i \) pixel. The smaller the mean square error, the smaller the difference between the predicted value and the real value, and the more realistic the predicted image. However, in the image deblurring of today’s deep methods, mse is often used as the loss during training, and PSNR and SSIM are more regarded as objective evaluation indicators for testing the deblurring effect.

5.2.2 PSNR

Peak signal-to-noise ratio (PSNR) is an objective standard for evaluating images, and there are many application scenarios.

\[ PSNR = 10 \log_{10} \left( \frac{(2^n - 1)^2}{MSE} \right) \]  

(5)

\( n \) is the number of bits per pixel, generally 8, that is, the number of pixel grayscale is 256. The unit of PSNR is dB, and the larger the value, the smaller the distortion. The smaller the MSE value, the closer the two pictures are, the larger the PSNR value, and the smaller the distortion. PSNR is the most common and widely used image objective evaluation index, but it is based on the error between corresponding pixels, that is, based on error-sensitive image quality evaluation. Since the visual characteristics of the human eye are not considered (the human eye is more sensitive to contrast differences with lower spatial frequencies, and the human eye is more sensitive to luminance contrast differences than chromaticity, the human eye’s perception of an area will be affected by the results. The influence of the surrounding area, etc.), the evaluation results are often inconsistent with people’s subjective feelings.

5.2.3 SSIM

In order to overcome the difference between objective indicators and subjective feelings, et al. proposed a new image quality evaluation indicator, namely structural similarity (SSIM). The system divides the tasks of similarity measurement into three comparisons: brightness, contrast, and structure.

\[ l(x,y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \]  

(6)

\[ c(x,y) = \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \]  

(7)

\[ s(x,y) = \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3} \]  

(8)

Generally take \( c_3 = \frac{2}{\sigma^2} \), \( \mu_x \) is the mean of \( x \), \( \mu_y \) is the mean of \( y \), \( \sigma_x^2 \) is the variance of \( x \), \( \sigma_y^2 \) is the variance of \( y \), \( \sigma_{xy} \) is the covariance of \( x \) and \( y \), \( c_1 = (k_1L)^2 \) and \( c_2 = (k_2L)^2 \) are two constants, \( L \) is the range of pixel values, usually \( k_1 = 0.01 \), \( k_2 = 0.03 \) is the default value. So,

\[ SSIM(x,y) = \left[ l(x,y)^\alpha \cdot c(x,y)^\beta \cdot s(x,y)^\gamma \right] \]  

(9)

Usually take \( \alpha, \beta, \gamma \) as 1. The higher the SSIM value, the higher the similarity between the two images.

6 Conclusion

This article reviews the work of image deblurring. The traditional image deblurring method and the depth-represented image deblurring method are comprehensively summarized respectively. In the traditional image deblurring method, the method of maximum a posteriori and the variational Bayesian method are specifically introduced and analyzed; in the image deblurring method of depth representation, the multi-scale deblurring network and generation Confronting de-fuzzing networks and unsupervised neural networks. Finally, several widely used image deblurring datasets are also proposed. In general, this review provides considerable guidance for relevant researchers.
References

[1] Y. Yamada, K. Nakamae, and H. Fujioka, “Image deblurring by the combined use of a super-resolution technique and inverse filtering,” in Computational Imaging IV, 2006.

[2] P. Bojarczak and Z. Łukasik, “Image deblurring – wiener filter versus tsvd approach,” Advances in Electrical and Electronic Engineering, vol. 6, no. 2, 2007.

[3] P. Anandan, “A computational framework and an algorithm for the measurement of visual motion,” International Journal of Computer Vision, vol. 2, no. 3, pp. 283–310, 1989.

[4] Y. Engel, S. Mannor, and R. Meir, “The kernel recursive least-squares algorithm,” IEEE Trans Signal Process, vol. 52, no. 8, pp. 2275–2285, 2004.

[5] M. K. Singh, S.-H. Kim, Y.-H. Kim, and U. Tiwary, “Passive millimeter-wave image deblurring using adaptively accelerated maximum entropy method,” in Proceedings of the KSRS Conference. The Korean Society of Remote Sensing, 2007, pp. 414–417.

[6] M. Zaharescu and C. A. Boianguie, “An investigation of image deblurring techniques,” International Journal of Mathematical Models and Methods in Applied Sciences, vol. 8, no. 1, pp. 75–83, 2014.

[7] F. D. Vieilleville, P. Weiss, V. Lobjois, and D. Kouame, “Alternating direction method of multipliers applied to 3d light sheet fluorescence microscopy image deblurring using gpu hardware,” Conference proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference, pp. 4872–4875, 2011.

[8] S. Aliyan and A. Broumandnia, “A new machine learning approach to deblurring license plate using k-means clustering method,” International Journal of Advanced Research in Artificial Intelligence, vol. 1, no. 2, 2012.

[9] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, “Multi-stage progressive image restoration,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 14 821–14 831.

[10] F.-J. Tsai, Y.-T. Peng, Y.-Y. Lin, C.-C. Tsai, and C.-W. Lin, “Banet: Blur-aware attention networks for dynamic scene deblurring,” arXiv preprint arXiv:2101.07518, 2021.

[11] Z. Yong, Y. M. Shao, Z. Xi, L. Li, W. H. Ip, and L. Y. Kai, “Edgan: motion deblurring algorithm based on enhanced generative adversarial networks,” The Journal of Supercomputing, 2020.

[12] G. Huang, Z. Liu, V. Laurens, and K. Q. Weinberger, “Densely connected convolutional networks,” IEEE Computer Society, 2016.

[13] O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang, “Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better,” in 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019.

[14] J. Lee, S. W. Ji, S. J. Cho, J. P. Hong, and S. J. Ko, “Deep learning-based deblur using gyroscope data,” in 2020 IEEE International Conference on Consumer Electronics - Asia (ICCE-Asia), 2020.

[15] Q. Hou, L. Zhang, M. M. Cheng, and J. Feng, “Strip pooling: Rethinking spatial pooling for scene parsing,” in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

[16] S. Lin, J. Zhang, J. Pan, Z. Jiang, D. Zou, Y. Wang, J. Chen, and J. Ren, “Learning event-driven video deblurring and interpolation,” in Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VIII 16. Springer, 2020, pp. 695–710.

[17] J. Pan, H. Bai, and J. Tang, “Cascaded deep video deblurring using temporal sharpness prior,” in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
[18] K. Purohit and A. Rajagopalan, “Region-adaptive dense network for efficient motion deblurring,” in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 07, 2020, pp. 11 882–11 889.

[19] K. Zhang, W. Luo, Y. Zhong, L. Ma, B. Stenger, W. Liu, and H. Li, “Deblurring by realistic blurring,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 2737–2746.

[20] Z. Zhong, Y. Gao, Y. Zheng, and B. Zheng, “Efficient spatio-temporal recurrent neural network for video deblurring,” in European Conference on Computer Vision. Springer, 2020, pp. 191–207.

[21] X. Tao, H. Gao, Y. Wang, X. Shen, J. Wang, and J. Jia, “Scale-recurrent network for deep image deblurring,” IEEE, 2018.

[22] H. Zhang, Y. Dai, H. Li, and P. Koniusz, “Deep stacked hierarchical multi-patch network for image deblurring,” IEEE, 2019.

[23] H. Gao, X. Tao, X. Shen, and J. Jia, “Dynamic scene deblurring with parameter selective sharing and nested skip connections,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 3848–3856.

[24] D. Park, D. U. Kang, J. Kim, and S. Y. Chun, “Multi-temporal recurrent neural networks for progressive non-uniform single image deblurring with incremental temporal training,” in European Conference on Computer Vision. Springer, 2020, pp. 327–343.

[25] X. Wang, R. Girshick, A. Gupta, and K. He, “Non-local neural networks,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7794–7803.

[26] J. Sun, W. Cao, Z. Xu, and J. Ponce, “Learning a convolutional neural network for non-uniform motion blur removal,” IEEE, 2015.

[27] S. Qi, J. Jia, and A. Agarwala, “High-quality motion deblurring from a single image,” Acm Transactions on Graphics, vol. 27, no. 3, pp. 1–10, 2008.

[28] J. Jia, “Single image motion deblurring using transparency,” in 2007 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2007, pp. 1–8.

[29] N. Joshi, C. L. Zitnick, R. Szeliski, and D. J. Kriegman, “Image deblurring and denoising using color priors,” in 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009, pp. 1550–1557.

[30] J. Kotera, F. Šroubek, and P. Milanfar, “Blind deconvolution using alternating maximum a posteriori estimation with heavy-tailed priors,” in International Conference on Computer Analysis of Images and Patterns. Springer, 2013, pp. 59–66.

[31] D. Krishnan, T. Tay, and R. Fergus, “Blind deconvolution using a normalized sparsity measure,” in CVPR 2011. IEEE, 2011, pp. 233–240.

[32] R. Fergus, B. Singh, A. Hertzmann, S. Roweis, and W. Freeman, “Removing camera shake from a single photograph,” Acm Transactions on Graphics, vol. 25, no. 3, pp. 787–794, 2006.

[33] D. Wipf and H. Zhang, “Revisiting bayesian blind deconvolution,” Journal of Machine Learning Research (JMLR), 2014.

[34] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” IEEE, 2016.

[35] S. Nah, T. H. Kim, and K. M. Lee, “Deep multi-scale convolutional neural network for dynamic scene deblurring,” IEEE Computer Society, 2016.

[36] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang et al., “Photo-realistic single image super-resolution using a generative adversarial network,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4681–4690.
[37] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” *IEEE*, 2016.

[38] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, “Understanding and evaluating blind deconvolution algorithms,” *Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1964–1971, 2009.

[39] A. Chakrabarti, “A neural approach to blind motion deblurring,” in *European Conference on Computer Vision*, 2016.

[40] H. Qiu, H. Li, Q. Wu, F. Meng, and H. Shi, “A2rmnet: Adaptively aspect ratio multi-scale network for object detection in remote sensing images,” *Remote Sensing*, vol. 11, no. 13, p. 1594, 2019.

[41] T. Zhao, H. Li, H. Qiu, Q. Wu, and K. N. Ngan, “A multi-scale language embedding network for proposal-free referring expression comprehension,” in *Proceedings of the 2nd ACM International Conference on Multimedia in Asia*, 2021, pp. 1–7.

[42] H. Luo, Q. Wu, K. N. Ngan, H. Luo, and L. Xu, “Multi-scale shape adaptive network for raindrop detection and removal from a single image,” *Sensors*, vol. 20, no. 23, p. 6733, 2020.

[43] H. Luo, H. Luo, Q. Wu, K. N. Ngan, and L. Xu, *Single Image Deraining via Multi-scale Gated Feature Enhancement Network*. Digital TV and Wireless Multimedia Communication, 2021.

[44] Y. Yang, F. Meng, H. Li, K. N. Ngan, and Q. Wu, “A new few-shot segmentation network based on class representation,” *2019 IEEE Visual Communications and Image Processing (VCIP)*, 2019.

[45] Q. Cheng, H. Li, Q. Wu, F. Meng, L. Xu, and K. N. Ngan, “Learn to pay attention via switchable attention for image recognition,” in *2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*. IEEE, 2020, pp. 291–296.

[46] F. Meng, K. Luo, H. Li, Q. Wu, and X. Xu, “Weakly supervised semantic segmentation by a class-level multiple group cosegmentation and foreground fusion strategy,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 12, pp. 4823–4836, 2019.

[47] L. Yang, F. Meng, Q. Wu, and H. Li, “Incorporating non-local and task-specific features for instance segmentation,” in *2019 IEEE 21st International Workshop on Multimedia Signal Processing (MSMP)*, 2019.

[48] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,” *Advances in Neural Information Processing Systems*, vol. 3, pp. 2672–2680, 2014.

[49] L. Yang, H. Li, F. Meng, Q. Wu, and K. N. Ngan, “Task-specific loss for robust instance segmentation with noisy class labels,” *IEEE Transactions on Circuits and Systems for Video Technology*, 2021.

[50] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, “Deblurgan: Blind motion deblurring using conditional adversarial networks,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[51] D. Ren, K. Zhang, Q. Wang, Q. Hu, and W. Zuo, “Neural blind deconvolution using deep priors,” in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[52] D. Ulyanov, A. Vedaldi, and V. Lempitsky, “Deep image prior,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 9446–9454.

[53] J. Rim, L. Haeyun, J. Won, and S. Cho, “Real-world blur dataset for learning and benchmarking deblurring algorithms,” in *European Conference on Computer Vision (ECCV)* 2020, 2020.

[54] Z. Shen, W. Wang, X. Lu, J. Shen, and L. Shao, “Human-aware motion deblurring,” *ICCV2019*, 2019.