Exploring and Evaluating Image Restoration Potential in Dynamic Scenes

Cheng Zhang[†,1], Shaolin Su[†,1], Yu Zhu[‡,1], Qingsen Yan[2], Jinqiu Sun[1], Yanning Zhang[1] [*]

[1]School of Computer Science and Engineering, Northwestern Polytechnical University, China
[2]School of Computer Science and Engineering, The University of Adelaide, Australia

https://github.com/Justones/IRP

Abstract

In dynamic scenes, images often suffer from dynamic blur due to superposition of motions or low signal-noise ratio resulted from quick shutter speed when avoiding motions. Recovering sharp and clean results from the captured images heavily depends on the ability of restoration methods and the quality of the input. Although existing research on image restoration focuses on developing models for obtaining better restored results, fewer have studied to evaluate how and which input image leads to superior restored quality. In this paper, to better study an image's potential value that can be explored for restoration, we propose a novel concept, referring to image restoration potential (IRP). Specifically, We first establish a dynamic scene imaging dataset containing composite distortions and applied image restoration processes to validate the rationality of the existence to IRP. Based on this dataset, we investigate several properties of IRP and propose a novel deep model to accurately predict IRP values. By gradually distilling and selective fusing the degradation features, the proposed model shows its superiority in IRP prediction. Thanks to the proposed model, we are then able to validate how various image restoration related applications are benefited from IRP prediction. We show the potential usages of IRP as a filtering principle to select valuable frames, an auxiliary guidance to improve restoration models, and even an indicator to optimize camera settings for capturing better images under dynamic scenarios.

1. Introduction

In real world imaging scenarios with dynamic motions, degradation is a common factor due to moving objects or shaking devices. To avoid blur, the photographer can

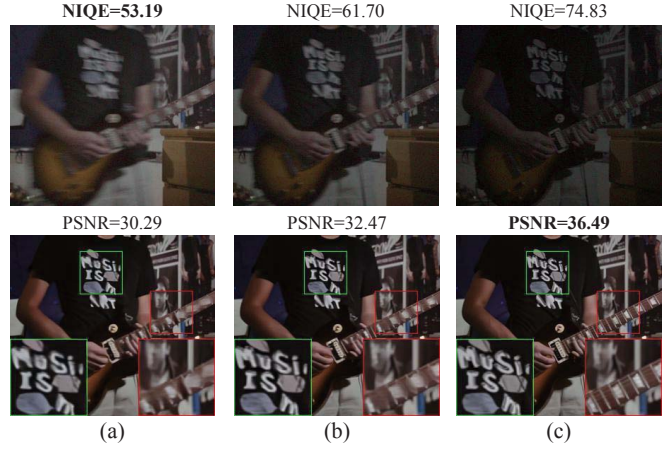| NIQE=**53.19** | NIQE=61.70 | NIQE=74.83 |
| PSNR=30.29 | PSNR=32.47 | **PSNR=36.49** |
| (a) | (b) | (c) |

Figure 1. We show how the proposed concept IRP differs from the traditional image metric visual quality. Given a sequence of images captured under varying exposures in dynamic scenes (first row), both IQA metric NIQE [23] and human observers favorite either (a) or (b) as relatively good shots, but they do not necessarily lead to better restored results (second row). As a comparison, a perceptually poor image (c) leads to the best restored image quality. The result indicates the necessity of developing the IRP metric which predicts if the restored image quality will be good even before it is processed. All images are processed by the same restoration model MPR [40] to ensure fairness.

shorten the exposure time, but results in low illumination regions and notorious noises [5, 6, 32]. With varying camera settings, the type of distortion may change, but hardly diminishes. To alleviate the annoying distortions, different image restoration algorithms are required: one can apply either deblurring approaches to remove motion blur in the image captured from adequate exposure, or denoising methods to alleviate noise artifacts due to shortened exposure. However, among the noise-blur trade-off, different kinds of distorted input lead to different restored results. Naturally, a question comes up that under the trade-off, which kind of image leads to better restored results? The question is fundamental, yet has not been well investigated in the literature. As far as we know, in order to obtain better restored results, most researches focus on developing restoration al-

gorithms [7, 13–15, 39, 40], but few are carried out to evaluate how and which input images lead to superior restored quality. As restored image quality heavily depends on both restoration models and inputs, in this paper, we propose the concept of image restoration potential (IRP), denoting an inherent image attribute that measures the potential value of an image that can be explored for restoration.

Though intuitively, less distorted images with higher quality lead to better restored results, the quality metric here differs from the widely accepted concept of visual perceptual quality. As shown in the top row of Figure 1, giving a sequence of images captured in dynamic scene, both image quality assessment (IQA) metric NIQE [23] and human observers tend to favorite image (a) or (b) as relatively better shots, however, these "better" inputs do not necessarily lead to better restored ones. In contrast, a perceptually poor quality image (c) achieves the best restored result, shown in the bottom row. The phenomenon therefore inspired us the necessity of developing the IRP concept. As traditional IQA metrics [28, 44] aim at measuring image quality at the **present** view and extracting perceptual features which are adapted to the human visual system (HVS), IRP is proposed to forecast if the restored image quality will be good **even before** they are processed, and focus more on the degradation that closely relates to the restoration process.

To investigate the proposed IRP, we first established a dynamic scene imaging dataset by simulating 5500 degraded images captured under various camera settings. We then conducted 4 representative restoration algorithms on each of the images in the dataset, and validated that IRP belonging to an inherent image attribute regardless of how concrete restoration algorithms are applied. We further dig into the distortions that exist in dynamic imaging scenarios including noises, blurriness, and low illumination, and proposed a deep model for accurate IRP prediction. By disentangling distortion factors and selective fusing degradation features, the proposed model showed its effectiveness in revealing image potentials for restoration. Lastly, we apply IRP prediction to various kinds of applications, including filtering valuable frames in image sequences for efficiently processing, guiding image processing models for adaptive restoration, and optimizing camera settings for capturing images leading to better restored quality.

To summary, the contributions of this paper include:

- We introduce a novel image attribute, named IRP, as a criterion to measure the potential value of an image that can be explored for restoration. By collecting 5500 images as well as their restored quality labels, we investigate several properties of the proposed IRP.

- We analyze the key factors affecting IRP and develop a deep model for IRP prediction. By gradually distilling image distortions that exist in dynamic scenes and se-

lective fusing the features to form complement representations, we verify the superior prediction accuracy of the proposed model.

- We show potential usages of the proposed IRP to various image restoration related applications. IRP has shown its effectiveness in filtering valuable frames for the restoration process, providing auxiliary guidance to restoration models, and even optimizing camera settings when capturing images under dynamic scenarios.

2. Related work

2.1. Image Restoration

According to the leading type of degradation contained in an image, different kinds of restoration tasks are proposed. In dynamic scenes, when the camera exposure setting is set long to ensure sufficient light, motions will be apparent and deblurring methods are required. Representative models include conventional approaches [33, 36] and CNN based models [7, 13, 19, 24, 29, 40, 42]. Meanwhile, when exposure is set short to avoid the superposition of motions, noise artifacts will be obvious due to insufficient light, and denoising approaches are needed, including conventional techniques [2, 9, 14, 35] and CNN based methods [3, 15, 40, 41, 43]. The above approaches, though achieving improving performances in their own field, are however proposed to deal with separate image restoration tasks. As a comparison, in this paper, we aim at developing a general image measurement IRP, that adapts to both restoration problems, and we are interested in finding out how image restoration related tasks are benefited from IRP predictions.

2.2. Image Quality Assessment

The goal of IQA is to enable machines to perceive the visual quality of images, being consistent with human perceptual results. By assessing an image's visual quality, many vision related tasks could be quantitatively measured and potentially optimized. Existing IQA approaches including full-reference IQA [30, 31, 45], reduced-reference IQA [12, 21, 37], and no-reference IQA [22, 28, 47], according to the accessibility to the pristine reference image. Though many works have been carried out in learning the relationship between image features and their direct visual quality [27, 34, 44], none of the work has studied an image's potential quality that can be explored for restoration. Therefore, in this paper, we propose and investigate the concept of IRP, analyze its properties, show its differences to IQA, and evaluate its potential usages in real world applications.

3. Investigating on IRP

To investigate IRP, we first established a dataset called Dynamic Scene-IRP (DS-IRP). The collection of DS-IRP

mainly includes two stages. In the first stage, we collected 2,500 dynamic scenes and 11 images in each scene under various camera settings. Since in dynamic scenes, it is impractical to acquire ground truth images by shooting in reality, which are however required in stage 2, we thus chose to simulate images following a dynamic imaging formation and finally had 27,500 images in total. In the second stage, we re-trained and tested 4 representative restoration algorithms on each of the images collected from stage 1. The IRP labels are then acquired by calculating restored image quality by referencing ground truth images. With the established DS-IRP dataset, we investigated IRP and revealed several of its properties.

3.1. Dynamic imaging formation

In our dynamic imaging formation, we model the imaging process starting with scene radiance ϕ and ending with image value I in sRGB space, but paying particular attention to dynamic motions existing in scenes. Therefore, the overall dynamic imaging formation can be expressed as the joint result including radiant power ϕ, motion information m and noise n during the exposure time Δt:

$$I = G(\phi, m, n, \Delta t) \tag{1}$$

Specifically, we consider motions during exposure, then convert the scene radiance into linear RGB pixel values, following [5, 10, 25]:

$$I_l = y_p\Big(\int_{t_0}^{t_0+\Delta t} \phi_t m_t \, dt\Big) + n \tag{2}$$

where y_p converts the received photoelectrons expressed by exposed photosite during the exposure time Δt into voltages, which are further recorded by sensor as linear RGB pixel values. m_t denotes motion at t time, and n represents the overall noise.

In practice, motions are considered in an equivalent but simpler way after ϕ are converted into linear RGB signals. Following [13], we apply optical flow to linear RGB signals as motion information. Specifically, we select consecutive image frames containing real scene motions from the vimeo-triplet dataset [38] and estimate the optical flow by ARflow [20], to represent motions m_{t_0} that exist in dynamic scenes. To cover various dynamic imaging results under different camera exposure settings, we sample 11 diverse exposure times in each scene and scale the motions m_{t_0} according to exposure time. In total, 2,500 scenes are selected from the vimeo-triplet dataset, each containing 11 imaging results corresponding to various exposure times. In each scene, we also collect the original image from the vimeo-triplet dataset, which serves as the ground truth image that is used for generating IRP labels in the second stage.

In the dynamic imaging formation, we consider noise as another inevitable degradation factor. The overall noise n composes the combination of shot noise n_{shot} and readout noise n_{read}, formulated as:

$$n = n_{shot} + n_{read} \tag{3}$$

where n_{shot} originates from the particle nature of light, and follows a Poisson process as $(I_l + n_{shot}) \sim Poisson(I_l)$ [32], while n_{read} associates with voltage fluctuation in signal processing flow, and follows a zero-mean Gaussian distribution with device-specific standard deviation, termed as $n_{read} \sim Gaussian(0, \sigma^2)$.

Finally, image signals are converted from linear RGB to sRGB space by quantization and bounded transformation, expressed as:

$$I = min(\lfloor (I_l)^\gamma + 0.5 \rfloor, M_{max}) \tag{4}$$

where $\lfloor \cdot \rfloor$, γ and M_{max} indicate floor function, gamma transformation, and the maximum value recorded by the camera sensor respectively.

3.2. IRP Label Generation

After obtaining 27,500 dynamic imaging results, we further collect their IRP labels. Since IRP is proposed to measure the potential that an image can be explored for restoration, we thus first apply image restoration algorithms on the images, then represent IRP value as the restored image quality, which is calculated by referencing the ground truth.

During IRP generation procedure, we select four representative image restoration methods, being effective in processing diverse image distortions, including Unet [26], MIRnet [39] ,MPR [40] and HInet [7] for restoration. Since we collected images under 11 diverse exposure settings, we train and test each restoration model under separate exposure settings across all scenes, leading to a total $4 \times 11 = 44$ times of training and testing procedures. Though it is plausible to mix all images for training and testing, we found that the models become less capable of handling all mixing distortions from noise to blur integrated with one model alone. As a result, we chose to train and test restoration models under each individual exposure settings, expecting that the max amount of "restoration potential" could be explored by the models. Specifically, among 2,500 scenes, we take 2,000 scenes for training restoration models, leaving the rest 500 scenes for testing and obtaining restoration results. To generate IRP values, we calculate both PSNR and LPIPS [45] scores between restored images and the ground truth to balance the trade-off between distortion and perception [4]. After normalizing both criteria to the range of $[0, 1]$, we calculate their mean value as IRP measurement. At last, for each image, we average the 4 IRP values corresponding to 4 restoration methods as the final IRP score. The whole generation process collects $500 \times 11 = 5500$ IRP labels in total. The labels, along with corresponding images collected in stage 1, form the DS-IRP dataset. Examples of data and IRP labels can be found in our supplementary.
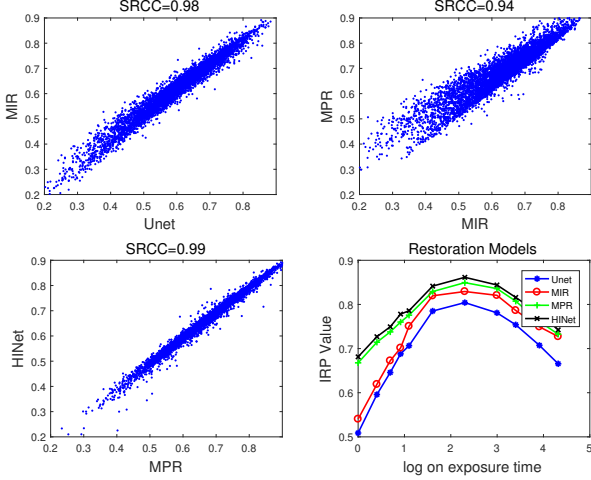
Figure 2. In the first three subfigures, we plot IRP values between two arbitrary restoration models. The restored image quality is highly correlated across different models, indicating IRP as an inherent image attribute. In the fourth subfigure, we plot IRP values generated by four restoration models under one scene, the IRP values corresponding to four methods also keep consistent.

3.3. Investigation on IRP Properties

After establishing the DS-IRP dataset, we are now able to investigate several properties of IRP. Specifically, we explored how IRP is affected by the selection of restoration models, how distortion factors determine IRP values and the difference between IRP and image visual quality.

Influence by Restoration Models. Apparently, even for the same image, the selection of different restoration models results in different restored image quality. Therefore, we are interested in finding out how IRP is influenced by the selection of restoration models. To this end, we take IRP values generated under different restoration models to compare their correlations, and we show the results in Figure 2. From the first three subfigures, it can be observed that despite different restoration models being applied, the restored image quality is highly correlated across all models. In the fourth subfigure, we show IRP values generated by four restoration models in one scene, containing 11 images captured under varying exposure settings. As directly shown, for each image, the relative IRP values keep consistent across restoration models. The above results indicate that even if different restoration models are used, a good image is a good image, *i.e.* its potential that can be explored for restoration keeps consistent. The result also demonstrates that IRP belongs to an inherent image attribute which is majorly mapped from image appearance, similar to existing image attributes including visual quality, brightness, and sharpness, *etc*.

Analyse on Determinant IRP Factors. As shown in Equation (1), the dynamic imaging process is determined by scene radiance ϕ, motion m, noise n and exposure time Δt, reflected as illumination, blurriness and noise on im-
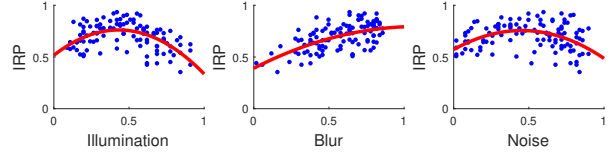

Figure 3. Subjective study on relationship between degradation factors and IRP values.
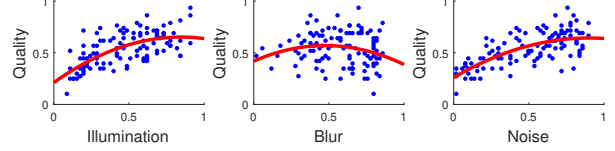

Figure 4. Subjective study on relationship between degradation factors and image visual quality.

age appearances. Therefore, we study on the three factors to analyze how they affect IRP values. Concretely, we conducted a subjective user study on three factors, by labeling the degradation magnitudes corresponding to low illumination, blur, and noise in 110 images randomly selected from the DS-IRP dataset. We then plot the relationship between each degradation magnitude and IRP value in Figure 3. In our experiment, 40 observers are invited to rate the degradation level ranging from $[0, 1]$, where a larger rating indicates less degradation detected. From Figure 3, we observe that when illumination and noise problem is severe, IRP values correlate with the magnitude of distortion. However, when two degradations are becoming less, IRP does not increase accordingly, due to the increasing motion blur resulted from longer exposure time in most scenes. This indicates when camera exposure setting varies, different kinds of distortions are becoming determinant under composite degradation. Meanwhile, we also found that when image blur is diminishing, IRP value consistently improves, suggesting that blurriness always plays a determinant role for IRP.

Comparison between IRP and Image Visual Quality. We further compare IRP with the widely studied image attribute, *i.e.* image visual quality, to investigate their differences. We conducted a similar user study by labeling image quality annotations, and plotting the relationship between each degradation factors and visual quality labels in Figure 4. Compared with Figure 3, it can be easily observed that image degradations, including low illumination, blurriness and noise, affect image visual quality in a different way from IRP. As can be seen, when composite distortions vary, visual quality shows a monotonic correlation with illumination and noise problems, while less correlated to blurriness. According to the above observation, we conclude the differences between IRP and visual quality: under the blur-noise trade-off in dynamic scenes, illumination problem attributes to a major factor determining image visual quality, but a relatively "easy" distortion for IRP as the information could be easily restored. Noise problem also implies relatively minor

affects on IRP attribute, but is determinant to visual quality to some degree. Among the degradation factors, blurriness correlates most with IRP value, and serves as the "hardest" distortion for restoration.

4. Learning to Predict IRP

In this section, we aim at developing a deep model for accurately predicting IRP. The predicting model is expected to serve as an indicator being applied in various dynamic scene imaging or restoration tasks. As we have analysed the major three factors and their mutual interactions in affecting IRP values, we thus propose to gradually distill each of the three factors in our model by individual branches, in order to learn complimentary IRP representations. In addition, to better understand the impact from each kind of degradation among the composite distortions, we propose to selective fuse the degradation features, which are finally regressed to IRP scores by multi-layer perceptrons (MLP).

4.1. Gradually Distilling Degradation Components

Given input images contaminated by the composition of illumination, noise and dynamic blur problems, we propose to gradually distill the degradation components by a series of pre-processing techniques, and extract each component's features one by one.

As image illumination mainly affects image visual presentation but a "easy to eliminate" component for IRP, we thus use the original distorted image to extract illumination features. Specifically, we adopt image histograms in an individual branch to extract illumination statistics. We compute 256 bins of the histogram and apply 1 layer of 1D convolution with a kernel size of 7 to the bins. The histogram is then spatially expanded to fit the size of features extracted from the rest two branches, denoted as $\mathbf{F_i}$. Next, to distill noise and blur features from composite distortions, we scale image signal in linear RGB space to alleviate the effect of illuminations, and extract noise features by another branch. Inside the noise feature extraction branch, we use the feature extractor from stage 1 of the ResNet50 backbone [17] to extract low level features. The features are then fed into 3 ASPP blocks [8] to expand the receptive field. In this way, we extract image features through a shallow branch, which leans to learn low level noise features, denoted as $\mathbf{F_n}$. As last, to distill blur features, we apply guided filtering [16] operation to the scaled image, and extract scene features by the third branch, using the whole ResNet50 encoder. Depth-wise convolution are then applied to reduce channel number. We denote the third part of feature $\mathbf{F_b}$.

By extracting illumination features $\mathbf{F_i}$ from holistic statistics, noise features $\mathbf{F_n}$ from low level representations, and blurry scene features $\mathbf{F_b}$ from high level extractors, we expect the features form complimentary representations, being effective in accurately predicting IRP values.
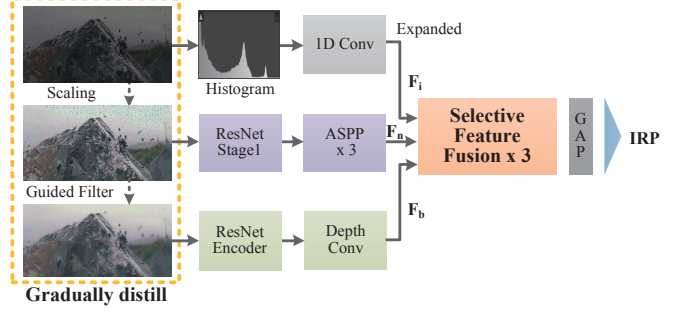

Figure 5. The proposed model architecture for IRP prediction.

4.2. Selective Fusing Degradation Features

As analysed above, when camera exposure setting varies, different kinds of distortions become determinant for IRP among the composite degradation. In order to dynamically adjust to dominant distortion factors, we propose to selective fuse the three parts of distortion features through a self-attention mechanism (also see supplementary for figure illustration), motivated by [39].

Given three parts of extracted feature $\mathbf{F_i}$, $\mathbf{F_n}$ and $\mathbf{F_b}$, we first combine them through element-wise summation and global average pooling, to form a channel-wise feature representation $\mathbf{s}$:

$$\mathbf{s} = \text{GAP}(\mathbf{F_i} + \mathbf{F_n} + \mathbf{F_b}) \qquad (5)$$

where GAP denotes global average pooling, $\mathbf{F_i}, \mathbf{F_i}, \mathbf{F_i} \in \mathbb{R}^{H \times W \times C}$ and $\mathbf{s} \in \mathbb{R}^{1 \times 1 \times C}$.

We then apply a depth-wise convolution with squeeze ratio r to $\mathbf{s}$, resulting a compact representation $\mathbf{z} \in \mathbb{R}^{1 \times 1 \times \frac{C}{r}}$. $\mathbf{z}$ are then fed into three parallel 1×1 convolution layers with expand ratio r to get three feature indicators $\mathbf{u_1}$, $\mathbf{u_2}$ and $\mathbf{u_3}$, which are further re-weighted to attention activations by channel-wise softmax operation:

$$\mathbf{v_i} = \frac{e^{\mathbf{u_i}}}{\sum_j e^{\mathbf{u_j}}} \qquad (6)$$

Finally, the degradation features $\mathbf{F_i}$, $\mathbf{F_n}$ and $\mathbf{F_b}$ are adaptively selected by multiplying $\mathbf{u_1}$, $\mathbf{u_2}$ and $\mathbf{u_3}$ respectively. The selective fusion operation incorporates three kinds of degradation features and refines each of them to adjust to the variation of scene exposures, thus adapting well in the IRP prediction task.

4.3. Regression to IRP Scores

After distilling degradation components and extracting corresponding features, we repeatedly selective fuse the features by 3 times. The output features are then summed over and globally average pooled into a vector representation. Finally, three full connection layers are applied to regress the features into the IRP score. During training, we minimize $L1$ loss for optimization.

Table 1. IRP prediction accuracy comparisons on DS-IRP dataset.

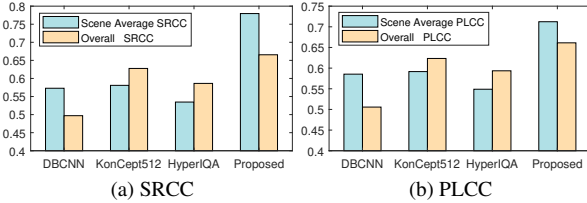| Model	Scene Average		Overall	
	SRCC	PLCC	SRCC	PLCC
BRISQUE [22]	0.3319	0.3560	0.1053	0.1877
IL-NIQE [44]	0.2631	0.3330	0.1869	0.2051
HOSA [34]	0.3360	0.3448	0.2269	0.2014
DBCNN [46]	0.8022	0.8008	0.6903	0.6956
KonCept512 [18]	0.8892	0.8984	0.7536	0.7839
HyperIQA [28]	0.8483	0.8578	0.7383	0.7550
Proposed	**0.9340**	**0.9412**	**0.8461**	**0.8687**

Figure 6. IRP prediction accuracy comparisons on real world data.

5. Experiments

In this section, to demonstrate the superiority of the proposed model, we evaluate IRP prediction accuracy on both synthetic and real world data. Due to the lack of models in the literature used for predicting the newly proposed IRP, we select existing IQA models as competitors.

5.1. Evaluation on the DS-IRP Dataset

We split the proposed DS-IRP dataset into a training subset, a validation subset and a testing subset, each containing 70%, 10% and 20% images according to scene contents. All the competing models are trained following their default settings and the best performing models in validation set are selected for testing. During the evaluation, we calculate Spearman's rank order correlation coefficient (SRCC) and Pearson's linear correlation coefficient (PLCC) for comparison. We compute SRCC and PLCC within each individual scene in DS-IRP and report their average value, denoted as scene average. We also compute the two criteria on the whole test set to evaluate the overall model performance. As shown in Table 1, the proposed model outperforms competitors by a large margin, indicating the effectiveness of the proposed architecture.

5.2. Evaluation on Real World Data

We further compare model performances on images captured in real world scenarios. We first captured 20 scenes containing dynamic motions in real world, using 8 different exposures under each scene. The collection leads to 160 i-eps-converted-to.pdfmaging results in total, and we use models that are trained on the DS-IRP dataset to predict their IRP values. Due to the lack of ground truth IRP to real world data, several pre-trained restoration methods [7,40,42] are applied to process real world data and we fur-

ther collected subjective perceptual scores from 40 participants to the restored results as substitutions of IRP values. We then calculate SRCC and PLCC among the model prediction scores and subjective scores, and show the results in Figure 6. Similarly, the proposed model also outperformed competitors on the challenging real world data.

Table 2. Ablation study for the proposed model.

Model	Scene Average		Overall	
	SRCC	PLCC	SRCC	PLCC
Baseline	0.8594	0.8608	0.7435	0.7758
w/o illumination	0.9188	0.9222	0.8339	0.8543
w/o blur	0.8886	0.8992	0.7814	0.8188
w/o noise	0.8957	0.9029	0.7920	0.8242
w/o selective	0.9335	0.9385	0.8278	0.8476
Full	**0.9340**	**0.9412**	**0.8461**	**0.8687**

5.3. Ablation Study

We conduct ablation studies to validate the effectiveness of each model component. We remove the illumination, noise and blur branch separately in our model, to observe the effect of individual degradation features. We then remove the selective feature fusion in the model to validate the effectiveness of feature selection among composite distortions. We also evaluate model performance using a ResNet50 baseline, and the results are shown in Table 2.

6. Applications

With the proposed IRP prediction model, in this section, we are able to demonstrate multiple applications that benefit from IRP prediction.

6.1. Filtering Principle for Efficient Processing

There exist circumstances such as autonomous driving and robot vision navigation, where images captured in dynamic scenes require to be processed and restored in real-time. To meet time and computation requirement, image filtering strategy becomes a solution for efficient processing. Specifically, among the continuous frames showing similar scene contents, only valuable frames are selected and processed to avoid overmuch raw data. In this way, both time cost and overall restored image quality could be improved.

Under the circumstance, IRP prediction can serve as a filtering principle, and we evaluate its effectiveness on the real world deblurring dataset GoPro [24]. Specifically, we split the GoPro test set into 105 groups containing 10 continuous frames inside each group. We then select the best frame inside each group according to the predicted IRP values, and process the selected frame by DMPHN [42]. We evaluate overall model complexity, average time consumption, restoration quality and the best frame selection accuracy in Table 3. To make comparisons, we also adopt two IQA models, which are trained on their own proposed IQA datasets as filtering principles, including KonCept512

Figure 7. We show how IRP serves as a filtering principle on the GoPro dataset. Compared with KonCept512 and MT-A, IRP prediction correctly select the most valuable frame in a sequence for efficiently restoring.

trained on the authentically distorted IQA dataset KonIQ-10k [18] , and MT-A trained on the smartphone photography and quality dataset SPAQ [11]. We also show qualitative comparisons in Figure 7.

Table 3. Model evaluations on the GoPro dataset as a frame filtering principle.

Model	FLOPs(G)	Time(s)	PSNR	SSIM	Accuracy
DMPHN	1099.35	0.0521	30.453	0.9022	-
KonCept512	152.234	0.0342	31.849	0.9263	49/105
MT-A	**9.394**	**0.0205**	31.155	0.9116	34/105
IRP	41.895	0.0244	**32.140**	**0.9351**	**57/105**

From Table 3, we found that for all three filtering models, the average restored image quality improved. This demonstrates both the feasibility of the image filtering strategy and the potential extended usages of existing IQA models. Meanwhile, among the competing models, both restored image quality and frame selection accuracy perform best through our IRP prediction model. Although our model is trained on the synthetic DS-IRP dataset, it outperforms competitors which collect real world images and subjective scores for model training. The result further proves the IRP prediction as a superior filtering principle for image restoration applications.

6.2. Auxiliary Guidance for Restoration Models

We also explore if IRP can also provide guidance for training adaptive restoration models. The underlying assumption is that by feeding both the image and its IRP in-

Table 4. Performance comparison on the proposed IRP as an auxiliary guidance for training restoration models.

Model	PSNR	SSIM
CBD baseline	40.227	0.9793
CBD full	40.722	0.9818
CBD + IRP	**40.771**	**0.9821**

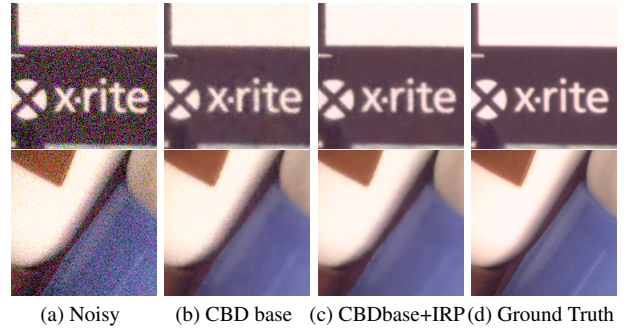(a) Noisy (b) CBD base (c) CBDbase+IRP (d) Ground Truth

Figure 8. Qualitative comparisons of IRP as an auxiliary guidance for restoration models.

formation to restoration models, they are able to distinguish images as easy or hard samples, thus learning adaptive mappings for images showing various restoration potentials.

To validate the assumption, we select small RGB data from the image denoising dataset SIDD [1] for evaluation. Given an input image, we extract its IRP features through the proposed model trained in Section 5.1 and use a 1×1 convolution layer to adjust the feature map into 3 channels. The feature map, concatenated with the original image, is

| (a) Auto-exposure | (b) IRP-exposure | (c) Auto-exposure restored | (d) IRP-exposure restored |

Figure 9. Comparison of IRP for optimizing exposure settings against conventional auto-exposure settings of camera.

then fed into the restoration model for either training or testing. We compare model performance with CBDnet [15], which also uses an auxiliary subnetwork, especially trained on the mixture of synthetic and real noisy images, to estimate image noise levels for restoration guidance. We also compare model performance when trained without any auxiliary guidance, denoted as CBDnet baseline, and show the results in Table 4, Figure 8. It can be found that by adding an auxiliary guidance, both CBDnet and IRP boost the image restoration baseline. Moreover, by extracting IRP features, restoration models are even more benefited than CBDnet. Though IRP is proposed to deal with composite distortions and we do not explicitly train it on real noisy images, we found it achieved impressive performance on the denoising task. Furthermore, since IRP features are extracted separately from restoration models, it is expected that IRP prediction can serve as a plug-and-play module in improving many other image restoration tasks.

6.3. Indicator for Optimizing Camera Settings

In the above subsections, we show IRP applications on single deblurring or denoising tasks, in this subsection, we further show its usages under real world composite distortions. We show the potential usage that applying IRP in optimizing camera exposure settings in real world dynamic imaging scenarios. As conventional auto-exposure settings in existing camera devices guarantee sufficient illumination as a priority, in dynamic scenes, whether the captured image

suits its best for restoration cannot be promised. We illustrate the application in Figure 9, where we show real world images collected from Section 5.2, selected by camera auto exposures and by IRP predictions. Both originally distorted images and the restored results are presented for qualitative comparisons. As can be seen, the conventional auto-exposure setting tends to capture images with sufficient illumination, they do not lead to satisfying restored quality. As a comparison, by predicting the origin images' IRP values, the selection leads to restoration results showing more satisfying visual quality.

7. Conclusion

In this paper, we propose IRP, a novel image attribute measuring its potential power that can be explored for restoration. We first established a DS-IRP dataset and explored the properties of IRP. Based on the analysis, we further proposed a deep model which gradually distills and selective fuse degradation features to accurately predict IRP. Experimental evaluations demonstrate the effectiveness of the proposed model architecture. Finally, with the IRP prediction model, we are able to apply it in various image restoration related tasks. The IRP has shown its potential usages in filtering valuable frames for efficient processing, providing extended guidance for restoration models, and even optimizing camera settings when capturing images under dynamic scenarios.

References

[1] Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown. A high-quality denoising dataset for smartphone cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1692–1700, 2018. 7

[2] Michal Aharon, Michael Elad, and Alfred Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing*, 54(11):4311–4322, 2006. 2

[3] Saeed Anwar and Nick Barnes. Real image denoising with feature attention. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3155–3164, 2019. 2

[4] Yochai Blau and Tomer Michaeli. The perception-distortion tradeoff. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6228–6237, 2018. 3

[5] Meng Chang, Huajun Feng, Zhihai Xu, and Qi Li. Low-light image restoration with short-and long-exposure raw pairs. *IEEE Transactions on Multimedia*, 2021. 1, 3

[6] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3291–3300, 2018. 1

[7] Liangyu Chen, Xin Lu, Jie Zhang, Xiaojie Chu, and Chengpeng Chen. Hinet: Half instance normalization network for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 182–192, 2021. 2, 3, 6

[8] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017. 5

[9] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8):2080–2095, 2007. 2

[10] Omer Dahary, Matan Jacoby, and Alex M Bronstein. Digital gimbal: End-to-end deep image stabilization with learnable exposure times. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11936–11945, 2021. 3

[11] Yuming Fang, Hanwei Zhu, Yan Zeng, Kede Ma, and Zhou Wang. Perceptual quality assessment of smartphone photography. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3677–3686, 2020. 7

[12] SeyedAlireza Golestaneh and Lina J Karam. Reduced-reference quality assessment based on the entropy of dwt coefficients of locally weighted gradient magnitudes. *IEEE Transactions on image processing*, 25(11):5293–5303, 2016. 2

[13] Dong Gong, Jie Yang, Lingqiao Liu, Yanning Zhang, Ian Reid, Chunhua Shen, Anton Van Den Hengel, and Qinfeng Shi. From motion blur to motion flow: A deep learning solution for removing heterogeneous motion blur. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2319–2328, 2017. 2, 3

[14] Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng. Weighted nuclear norm minimization with application to image denoising. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2862–2869, 2014. 2

[15] Shi Guo, Zifei Yan, Kai Zhang, Wangmeng Zuo, and Lei Zhang. Toward convolutional blind denoising of real photographs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1712–1722, 2019. 2, 8

[16] Kaiming He, Jian Sun, and Xiaoou Tang. Guided image filtering. In *European conference on computer vision*, pages 1–14. Springer, 2010. 5

[17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 5

[18] Vlad Hosu, Hanhe Lin, Tamas Sziranyi, and Dietmar Saupe. Koniq-10k: An ecologically valid database for deep learning of blind image quality assessment. *IEEE Transactions on Image Processing*, 29:4041–4056, 2020. 6, 7

[19] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiří Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8183–8192, 2018. 2

[20] Liang Liu, Jiangning Zhang, Ruifei He, Yong Liu, Yabiao Wang, Ying Tai, Donghao Luo, Chengjie Wang, Jilin Li, and Feiyue Huang. Learning by analogy: Reliable supervision from transformations for unsupervised optical flow estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6489–6498, 2020. 3

[21] Yutao Liu, Guangtao Zhai, Ke Gu, Xianming Liu, Debin Zhao, and Wen Gao. Reduced-reference image quality assessment in free-energy principle and sparse representation. *IEEE Transactions on Multimedia*, 20(2):379–391, 2017. 2

[22] A Mittal, A. K. Moorthy, and A. C. Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, 21(12):4695, 2012. 2, 6

[23] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012. 1, 2

[24] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3883–3891, 2017. 2, 6

[25] Emmanuel Onzon, Fahim Mannan, and Felix Heide. Neural auto-exposure for high-dynamic range object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7710–7720, 2021. 3

[26] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 3

[27] Shaolin Su, Vlad Hosu, Hanhe Lin, Yanning Zhang, and Dietmar Saupe. KonIQ++: Boosting no-reference image quality assessment in the wild by jointly predicting image quality and defects. *The 32nd British Machine Vision Conference*, 2021. 2

[28] Shaolin Su, Qingsen Yan, Yu Zhu, Cheng Zhang, Xin Ge, Jinqiu Sun, and Yanning Zhang. Blindly assess image quality in the wild guided by a self-adaptive hyper network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3667–3676, 2020. 2, 6

[29] Pei Wang, Wei Sun, Qingsen Yan, Axi Niu, Rui Li, Yu Zhu, Jinqiu Sun, and Yanning Zhang. Non-uniform motion deblurring with blurry component divided guidance. *Pattern Recognition*, 120:108082, 2021. 2

[30] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 2

[31] Zhou Wang and Qiang Li. Information content weighting for perceptual image quality assessment. *IEEE Transactions on image processing*, 20(5):1185–1198, 2010. 2

[32] Kaixuan Wei, Ying Fu, Jiaolong Yang, and Hua Huang. A physics-based noise formation model for extreme low-light raw denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2758–2767, 2020. 1, 3

[33] Oliver Whyte, Josef Sivic, Andrew Zisserman, and Jean Ponce. Non-uniform deblurring for shaken images. *International journal of computer vision*, 98(2):168–186, 2012. 2

[34] Jingtao Xu. Blind image quality assessment based on high order statistics aggregation. *IEEE Transactions on Image Processing*, 25(9):4444–4457, 2016. 2, 6

[35] Jun Xu, Lei Zhang, David Zhang, and Xiangchu Feng. Multi-channel weighted nuclear norm minimization for real color image denoising. In *Proceedings of the IEEE international conference on computer vision*, pages 1096–1104, 2017. 2

[36] Li Xu, Shicheng Zheng, and Jiaya Jia. Unnatural l0 sparse representation for natural image deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1107–1114, 2013. 2

[37] Yong Xu, Delei Liu, Yuhui Quan, and Patrick Le Callet. Fractal analysis for reduced reference image quality assessment. *IEEE Transactions on Image Processing*, 24(7):2098–2109, 2015. 2

[38] Tianfan Xue, Baian Chen, Jiajun Wu, Donglai Wei, and William T Freeman. Video enhancement with task-oriented flow. *International Journal of Computer Vision*, 127(8):1106–1125, 2019. 3

[39] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for real image restoration and enhancement. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16*, pages 492–511. Springer, 2020. 2, 3, 5

[40] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14821–14831, 2021. 1, 2, 3, 6

[41] Cheng Zhang, Qingsen Yan, Yu Zhu, Xianjun Li, Jinqiu Sun, and Yanning Zhang. Attention-based network for low-light image enhancement. In *2020 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2020. 2

[42] Hongguang Zhang, Yuchao Dai, Hongdong Li, and Piotr Koniusz. Deep stacked hierarchical multi-patch network for image deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5978–5986, 2019. 2, 6

[43] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7):3142–3155, 2017. 2

[44] Lin Zhang, Lei Zhang, and Alan C Bovik. A feature-enriched completely blind image quality evaluator. *IEEE Transactions on Image Processing*, 24(8):2579–2591, 2015. 2, 6

[45] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 2, 3

[46] Weixia Zhang, Kede Ma, Jia Yan, Dexiang Deng, and Zhou Wang. Blind image quality assessment using a deep bilinear convolutional neural network. *IEEE Transactions on Circuits and Systems for Video Technology*, 2018. 6

[47] Hancheng Zhu, Leida Li, Jinjian Wu, Weisheng Dong, and Guangming Shi. Metaiqa: Deep meta-learning for no-reference image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14143–14152, 2020. 2