

Temperature-Robust Learned Image Recovery for Shallow-Designed Imaging Systems

Wei Chen, Bingyun Qi, Xu Liu, Haifeng Li, Xiang Hao, and Yifan Peng*

Imaging systems are widely applied in harsh environments where the performance of shallow-designed systems may deviate from expectation. As a representative scenario, environmental temperature variation may degrade image quality due to thermal defocus and sensor response, resulting in blur and noise. However, extensive athermalization in optics usually requires a complex design process and is limited by materials. Herein, a multibranch computational imaging scheme is developed, using emerging generative adversarial networks as the postprocessing to compensate for degradation of all kinds caused by thermal defocus and noise. In addition, a temperature controllable data acquisition, division, and mixture scheme is described to facilitate effective datasets for model robustness. Experiments on a vehicle lens and a mobile phone lens reveal that the proposed multibranch learned strategy notably increases image quality in the temperature range of 0–80 °C, and outperforms conventional athermalization in most instances, which is beneficial to lowering the design and manufacturing costs of imaging systems.

fidelity of acquired information. Further, in daily life, most low-cost imaging systems such as vehicle or phone cameras depend on ill-conceived optical designs and post-processing algorithms, resulting in poor thermo-optic effects.^[3,4] Due to the intensive usage of plastic lenses, this problem is becoming increasingly critical. As traditional approaches^[5] to handling thermo-optic effects show defects that limit their decentralization to compact machine vision cameras, we develop a temperature-robust learned image recovery scheme, predominantly inspired by emerging machine intelligence advances.

1.1. Related Work

1.1.1. Temperature-Robust Optical Design

Athermalization techniques aim for realizing a stable focal power within a wide range of temperatures. Typically, there are three types: passive optical,^[6] active mechanical,^[7] and passive mechanical methods.^[8] These solutions compensate for thermal defocus, respectively, by the combination of optical materials, thermal expansion of mechanical components, and motorized mechanical parts. However, mechanical methods are often insufficiently robust and unfavorable to miniaturization, while optical ones usually request various types of lens material and complex structural designs. Alternatively, emerging wave front coding approaches^[9,10] alleviate the thermal defocus by increasing the depth of focus and applying proper decoding algorithms, but its viability is compromised by factors including narrow field of view, decoding noise, and the requirement of extra optical elements.

1.1.2. Learned Image Recovery

Intuitively, resolving high-quality images from degraded ones can be regarded as image recovery, which considerably mitigates blur and distortion due to the imperfect imaging and diffraction of the optical system, in tandem with possible sensor noise. When it comes to these issues, deep learning techniques, in particular convolutional neural networks (CNNs), have been proven more effective than conventional iterative deblurring and denoising algorithms. For example, the end-to-end optimization of CNN^[11] allows for solidly holding back the accumulation of errors in the computing process. The DnCNN^[12] leveraged residual learning and batch normalization strategies to handle additive white Gaussian noise and even blind Gaussian noise. Abuolaim et al.^[13] presented a DPDNet using dual pixel data

1. Introduction

The imaging behavior of conventional imaging systems can be affected by a handful of environmental factors. For instance, robots or drones with machine vision cameras are widely applied in polar or desert scientific expeditions, where unexpected sharp temperature variation may alter the refractive index and thickness of their lenses, as well as the sensor response. The resulting defocus blur^[1] and thermal noise^[2] may dramatically degrade the

W. Chen, B. Qi, X. Liu, H. Li, X. Hao
College of Optical Science and Engineering
Zhejiang University
Hangzhou 310027, China

H. Li
Research Center for Intelligent Chips and Devices
Zhejiang Lab
Hangzhou 311121, China

Y. Peng
Department of Electrical and Electronic Engineering
The University of Hong Kong
Hong Kong, Hong Kong
E-mail: evanpeng@hku.hk

The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/aisy.202200149>.

© 2022 The Authors. Advanced Intelligent Systems published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/aisy.202200149

to generate a clear image for handling defocus blur. Although a considerable amount of deblurring and denoising work has been reported yearly,^[14–17] most of them focus on solving one of the problems without considering their joint effects at a wide range of temperatures. Yet, it is mandatory for resolving latent images to consider the peculiar optical properties^[3] and sensor responses to temperature.

1.1.3. Domain-Specific Imaging

A premium recovery relies on a prodigious dataset where more data means more valuable features can be learned in latent space by the network. But concomitantly, datasets containing different scenes and characteristics often showcase various or even conflicting features, which may burden the model learning for domain-specific tasks, e.g., underwater color correction,^[18] high dynamic range imaging,^[19] and vehicle compound lens imaging.^[20] Although there has been a denoising dataset (SSID)^[21] and a defocus blur dataset^[13] dedicated to task-specific imaging, they deviate from our goal in resolving clear images with both noise and defocus blur present due to temperature variance. Notably, these factors could vary dramatically subject to temperatures; thereby, a dataset acquired under a certain condition usually fails to incorporate all kinds of degradation. As such, it is valuable to facilitate a series of domain-specific datasets containing various corrupted features, for the sake of a robust reconstruction network.

1.2. Motivation and Contributions

We develop a temperature-robust image recovery strategy for shallow-designed imaging systems, to address both thermal defocus and noise corruption due to temperature instability. The following technical contributions are investigated: 1) Taking account the varying optical properties and sensor responses at different temperatures, we facilitate a series of datasets to train a robust multibranch reconstruction network model, where each branch corresponds to a unique temperature. 2) To reduce the redundancy of branch networks, we apply a flexible temperature division and mixture scheme to fuse branches according to the degradation levels, and further compress the scale of the network architecture. 3) Theoretically and experimentally, we validate the proposed strategy on two kinds of off-the-shelf camera modules, enabling high-quality imagery at varying temperatures without using tailored optical designs or sensors.

The proposed strategy not only paves the way to alternative thermalization imaging solutions from the deep learning perspective, but also enables a division and mixture scheme to reduce model redundancy and computational expenses. With flexibility and robustness, it can promise numerous intelligent machine vision systems in photography, robotics, automatic vehicles, drones, medical devices, and other human vision and sensory applications.

2. Temperature-Robust Computational Imaging Framework

Regarding optics, the fluctuating temperature affects both ray refractive properties and surface form factor of optical elements,

leading to a temperature-sensitive focal power and point spread function (PSF) behavior. Typical passive optical designs tend to compensate for such deviation by using a combination of different materials. However, it is often challenging to reach a great balance with limited material options under the additional premise of aberrations constraint and other inherent design metrics.

Concerning image formation, a degraded observation I can be expressed as

$$I(x, y, T) = \int Q_c(\lambda) \times [p(x, y, d, \lambda, T) \otimes I_c(x, y)] d\lambda + N(x, y, T) \quad (1)$$

where the PSF $p(x, y, d, \lambda, T)$ is dependent on not only the spatial position (x, y) , object distance d , and incident spectral distribution λ , but also temperature T . Q_c is the spectral response of sensor color filters, and I_c is the latent clear image. N represents the sensor noise that exponentially varies concerning temperature due to the electron thermal motion.

As such, a shallow-designed imaging system records images corrupted with different levels of blur, aberration, and noise at various temperatures. Intuitively, it would bring about that an isolated reconstruction model only considering room temperature does not often work well in the postprocessing to resolve clear images at a wide range of temperatures. Therefore, we develop a temperature-robust multibranch image reconstruction model, seeking to yield output images with comparable quality regardless of temperature.

As shown in **Figure 1**, we present datasets obtained at different temperatures to the learned networks and train corresponding branches of the full reconstruction model. For boosting realistic visual effects, each branch model explores a generative adversarial network (GAN)^[22] with a generator based on a variant of the U-net architecture.^[23] In the test or application phase, the final model automatically selects the appropriate branch network with the temperature detector feedback to resolve a photorealistic image.

3. Lightweight, Multibranch Image Recovery Network

We seek to realize the temperature-robust high-quality imaging by leveraging a multibranch GAN-based model, with a risk of redundancy in a sense. For simultaneous lightweight and robustness, we further simplify the network architecture through base unit, as well as explore a proper scheme of dividing temperature ranges and mixing acquired datasets so as to reduce the number of branches.

3.1. Lightweight Base Network Design

In each branch of the reconstruction model, the GAN attempts to minimize the weighted combination of adversarial loss and perceptual loss^[24] to recover degraded images robustly at different temperatures, which can avoid overly blur caused by the pixel-wise average of possible optima.^[25]

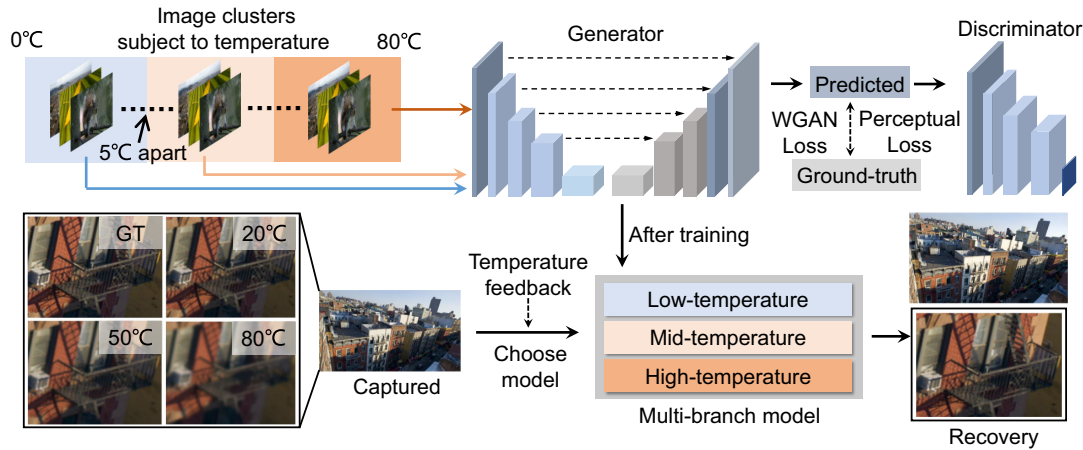


Figure 1. Overview of temperature-robust learned image recovery. Top: we blend a series of datasets acquired at different temperatures (0–80 °C, with 5 °C apart) according to the divided subranges and, respectively, use them to train specific GANs. Eventually, the trained generators constitute the multibranch image reconstruction model. Bottom: given the feedback of environment temperature, a degraded measurement is sent to the corresponding recovery branch of the reconstruction model so as to resolve its latent clear image.

3.1.1. Adversarial Loss

Instead of using a traditional adversarial loss,^[22] we apply one of the variants of Wasserstein GAN^[26] to minimize the Wasserstein distance between predicts and ground truth images for a robust training process. The adversarial loss function can be expressed as

$$\mathcal{L}_{\text{adv}} = E_{I \sim P_r}[D(I)] - E_{\tilde{I} \sim P_g}[D(\tilde{I})] + \lambda E_{\tilde{I} \sim P_g}[(\|\nabla_{\tilde{I}} D(\tilde{I})\|_2 - 1)^2] \quad (2)$$

where P_g and P_r represent distributions of model and data, respectively. In the training, the adversarial loss aims to minimize the structural deviation between the generated image \tilde{I} and the real image I , penalizing missing structures.

3.1.2. Perceptual Loss

The perceptual loss encourages outputs to have similar high feature representations with the ground truth images, which are extracted from the 15th layer of the pretrained VGG19.^[27] Minimizing the distance of Gram matrices between the generated images and references, the loss function can be defined as follows

$$\mathcal{L}_c = \Sigma_{15} \|\text{Gram}_{15}(I) - \text{Gram}_{15}(\tilde{I})\|_1 \quad (3)$$

A strong resemblance between high feature representations indicates that the generated images exhibit a high visual quality similar to ground truth.

3.1.3. Network Depth Fine-Tuning

In a sense, the suitable depth of the branch network should be determined by the degradation level. We note that a lightweight model is capable of alleviating slight degradation at room temperature but may not work well at extreme ones. In each branch

of the reconstruction model, one can adopt the generator consisting of different numbers of network blocks subject to degradation level. This insight is conducive to further reducing the computational expense because the extreme-case degradation is usually rare in the wild. Notably, although it may be in practice clever to adopt more network blocks in extreme temperature models, in the following experiments we use the same network architecture (see Section 4.2.1) in each branch model for a fair comparison of recovery performance subject to temperature. In addition, rather than diving into advanced network architectures, it may be more persuasive to adopt a fundamental, well-established architecture to reproduce high-quality images by the superiority of the proposed strategy. The specific network architecture can be found in Section 2 of the Supplement.

3.2. Temperature Division and Data Mixture

As it is superfluous and unaffordable for hardware to specify a model at each temperature, we explore a temperature division and data mixture scheme to reduce the number of branches. Taking a vehicle lens module without the athermalization design as an example, we first assess the blur level of its measured images at uniformly sampled temperatures (5 °C apart) by peak signal-to-noise ratio (PSNR) between the ring-split convolution simulation results and ground truth (see blue curve in Figure 2). In the simulation, we split the ground truth into a sequence of concentric rings to match PSFs in the corresponding field of view, and then convolve them with corresponding PSFs and add the Gaussian noise to simulate sensor responses.

To further confirm the necessity of the multibranch strategy, we apply the basic network trained on the dataset at 20 °C to recover blur images at all sampled temperatures (see Net-20 in Figure 2). We note that it behaves well in low- and medium-temperature ranges where the deviation of blur level is relatively flat, but exhibits a poor performance at high-temperature range where the blur level varies acutely. As such, we divide temperature ranges subject to the PSNR distribution of blur images.

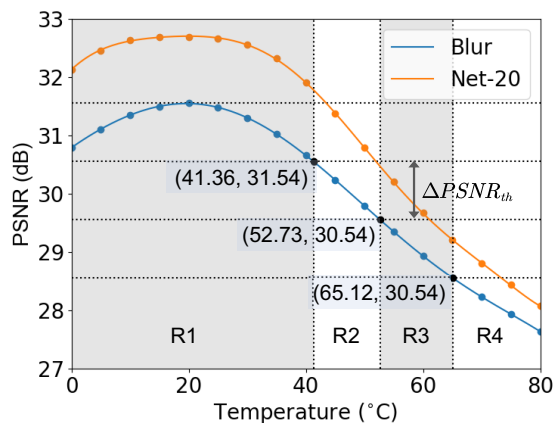


Figure 2. Temperature division scheme. With the $\Delta\text{PSNR}_{\text{th}} = 1$ dB, the entire temperature range is divided into four subranges (R1–R4) according to the PSNR distribution of simulated degradation at different temperatures (blue curve). Besides, the recovery results of the model trained on the 20 °C dataset deteriorate as the temperature increases (orange curve), which proves the necessity of our division and multibranch scheme.

When the PSNR drops beyond a certain threshold, marked as $\Delta\text{PSNR}_{\text{th}}$, we set a demarcation and assume that there exist non-negligible differences in blur images between two adjacent ranges, which are unsuitable to be recovered by the same model. In this work, we set $\Delta\text{PSNR}_{\text{th}}$ as 1 dB and divide 0–80 °C into four ranges marked from R1 to R4. Noted that $\Delta\text{PSNR}_{\text{th}}$ is adjustable and its precision depends on the requirement of the application. Intuitively, a smaller threshold would result in more precise results, while burdening the computational memory due to creating more branches.

Eventually, we blend the blur images of sampled temperatures in the same range and construct a hybrid dataset to train a robust branch model competent for the corresponding range. Taking the R1 as an example, we sequentially assign 68 images in nine sample temperatures to form a training set with 612 data pairs, aka, the total scale of the training dataset. The superiority of this trick is verified in Section 4.2. Notably, the currently divided ranges

guide the initial training process but do not determine the number of branches for the final recovery model, i.e., adjacent ranges can be merged and retrained given that their specific models perform similarly or the hardware resource is strictly limited. The detailed scheme can be found in Section 4.1 of the Supplement.

3.3. Temperature-Controllable Automatic Data Acquisition

As shown in Figure 3, for acquiring substantial datasets automatically, we built a display-capture setup^[28,29] including a high-resolution, calibrated LCD monitor (Asus PA32U) and a shallow-designed phone module (Samsung GN1). Apart from the benefit of acquiring massive datasets without manual effort, it also considers the real noise and blur in both optical design and image signal processing (ISP) together instead of only dealing with one part of them. Therefore, the model trained on the datasets acquired from such a setup has more potential to generalize to complex real-world environments such as extreme temperatures.

To acquire a series of datasets containing accurate temperature features, we build a chamber with the control installation for heating and cooling. The chamber imitates the temperature of environment where the camera is placed, ranging between 0 °C and 60 °C, with a fluctuation (1 °C) much smaller than the specified one. During the shooting, the chamber maintains the temperature at a desired constant value, and the camera captures images shown on the monitor through a piece of high-transparency white glass. Besides, an extra temperature detector (DS18B20) is installed by the camera for precise temperature feedback to the control PC.

4. Results and Discussion

4.1. System Specification

We assign 612 images as our training dataset and 50 ones as the test set from the Adobe 5 K collection.^[30] For robustness, the distribution of the test set should not be in line with those used for the temperature division in Section 3.2. Implemented with PyTorch, the training process of the recovery networks takes

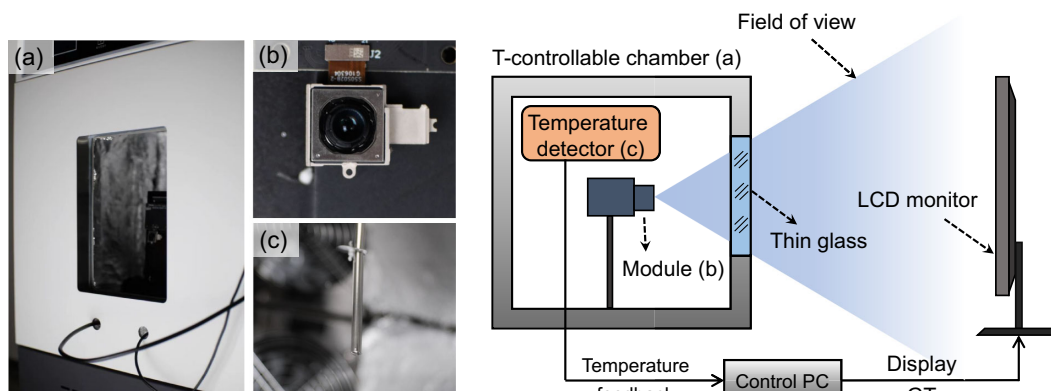


Figure 3. Temperature-controllable display-capture setup. A chamber is adopted to offer a constant temperature environment for the phone module, which is responsible for capturing the dataset displayed on the monitor. In the testing phase, the detector gets back the temperature feedback in real time to the control PC for a suitable branch model.

250 epochs on a single graphics card (NVIDIA GTX 3090), and the simplified architecture allows them to run on any GPU with a memory greater than 6 GB. In the training phase, we crop the 512×512 resolution patches from the blur dataset and their corresponding GT with random flipping and rotation for data augmentation. Adopting the well-known Adam optimizer, the learning rate is initially set to 10^{-5} and linearly decreased to 0. During each iteration, the discriminator D is updated 5 times while the generator G is updated once.

4.2. Simulated Results

For intuition and clarity, we utilize two kinds of vehicle lenses to demonstrate the feasibility of the temperature-robust strategy from 0 °C to 80 °C. One of them with athermalization could provide stable-quality images at different temperatures, while the other suffers from the thermal defocus at extreme ones due to the shallow optical design. For training and test sets, we resize the GT images to 1920×1080 to match the common sensor specifications of these vehicle lenses. We obtain PSFs at various temperatures through the thermal analysis function of the OpticStudio^[31] and simulate blur images via convolution operations.

For each branch model, we report both PSNR and LPIPS^[32] of recovery images and assess by average values, where PSNR exhibits the pixel-wise performance while LPIPS represents the similarity of high-level features. Extra assessment using the structural similarity (SSIM)^[33] can be found in the Supplement, so is the experiment at extremely low temperatures (below 0 °C), which, again, robustly manifests the validity of the proposed scheme.

4.2.1. Ablation Study

To demonstrate the effectiveness of the scheme proposed in Section 3, we conduct several ablation experiments concerning lightweight architectures, multibranch scheme, and temperature division and mixture. For brevity, we refer to these models trained under different strategies as Branch Discrete, Branch Hybrid, and All Hybrid, respectively. Each dataset for Branch Discrete is obtained at only one temperature, while that of Branch Hybrid mixes multiple images under the same divided range. In contrast, All Hybrid simply blends all datasets within the entire temperature region without branches. Above datasets exhibit the same scale but different temperature distributions.

Lightweight Architecture: We note that the depth of the recovery network depends on the degradation level, and images captured by the vehicle lens do not require such a deep network architecture,^[29] termed as the Original model. Therefore, we can reduce the number of network blocks in the generator of the Original and construct a lightweight architecture as Current model. As shown in Table 1, trained on the same dataset from All Hybrid strategy, the number of parameters to optimize in Current is significantly reduced to only one-eighteenth compared to the Original; in the meanwhile, the image quality does not drop noticeably and even improves at several ranges. This significant reduction in parameters offers a possibility to

Table 1. Ablation study of models with different parameters. Numbers indicate the number of the parameters to optimize. We reduce the number of network unit in original model as current model. The PSNR and learned perceptual image patch similarity (LPIPS) metrics showcase that when the number of the parameters drops dramatically, the quality of the recovery does not decline quickly and even increases at some ranges.

Temperature	Metrics	0 °C	20 °C	40 °C	60 °C	80 °C	Average	Numbers
Original	PSNR ↑	32.06	32.18	31.79	30.16	28.85	31.01	2.15×10^8
	LPIPS ↓	0.04	0.02	0.03	0.06	0.11	0.05	
Current	PSNR ↑	32.24	32.51	32.02	30.17	28.70	31.13	1.20×10^7
	LPIPS ↓	0.04	0.02	0.03	0.08	0.14	0.06	

reproduce high-quality images using microprocessors that already exist in many imaging systems.

Multibranch Scheme: The opposite of the multibranch strategy is mixing all datasets under the entire range without considering their temperature deviation, i.e., All Hybrid model. As shown in Table 2, regardless of PSNR and LPIPS, the performance of the multibranch models, Branch Discrete or Branch Hybrid, exhibits better than those of All Hybrid over almost the entire temperature range. Experimental results illustrate that the blind mixture may induce the crosstalk of high features of different corruptions in latent space and burden the learning of the network architecture. And it could be counterproductive to account for thermal corruption by roughly constructing a dataset containing all kinds of degradation. Therefore, the multibranch scheme is beneficial for addressing the degradation at a wide range of temperatures.

Temperature Division and Mixture: With the comparison between the Branch Discrete and Branch Hybrid shown in Table 2, the PSNR values of the Branch Hybrid recovery increase noticeably and outperform other models at each temperature. Although the Branch Hybrid model cannot achieve the best LPIPS at all temperatures, it still produces comparable values to Branch Discrete in most ranges and significantly outperforms All Hybrid. With the significant reduction of the branch number, as well as yielding the best PSNR and comparable LPIPS, our multibranch model with appropriate temperature division and mixture is proven the promising scheme to deliver high-quality images at a wide range of temperatures.

4.2.2. Temperature-Robust Image Recovery

As the temperature increases, the textures and outlines of objects become indistinguishable beyond 60 °C, such as stone carvings on the wall, edges of tree branches, and patterns on the lifebuoy in Figure 4. Benefiting from the division and mixture scheme, images generated by our multibranch model represent superior visual quality in clarity, sharpness, and details as opposed to the blur ones. In addition, as shown in Figure 5, the PSNR values of images recovered by our model all exhibit over 2.0 dB higher than the blur ones at 0–80 °C. Meanwhile, we report the recovery results below 0 °C, revealing that our multibranch strategy is compatible at extremely low temperatures. The detailed experiments can be found in Section 4.2 of the Supplement. To conclude, the proposed multibranch model significantly improves

Table 2. Ablation study of models with various mixture schemes. The comparison between All Hybrid and Branch Discrete or Branch Hybrid suggests the inadequacy of the blind mixture and the necessity of applying multiple branches. With leading to the best PSNR and comparable LPIPS in quality of recovery, Branch Hybrid further reveals the superiority of the multibranch strategy with appropriate temperature division and mixture. Bold values indicate the best performance.

Metrics	Temperature	0 °C	10 °C	20 °C	30 °C	40 °C	50 °C	60 °C	70 °C	80 °C	Average
PSNR ↑	All Hybrid	32.24	32.57	32.51	32.42	32.02	31.18	30.17	29.49	28.70	31.26
	Branch Discrete	32.53	32.69	32.77	32.34	32.02	31.20	30.19	29.64	28.71	31.34
	Branch Hybrid	32.65	33.06	32.99	32.96	32.31	31.33	30.26	29.66	29.19	31.60
LPIPS ↓	All Hybrid	0.04	0.02	0.02	0.02	0.03	0.06	0.08	0.10	0.14	0.055
	Branch Discrete	0.02	0.01	0.01	0.02	0.03	0.04	0.06	0.07	0.10	0.039
	Branch Hybrid	0.03	0.02	0.02	0.02	0.04	0.04	0.06	0.08	0.10	0.044

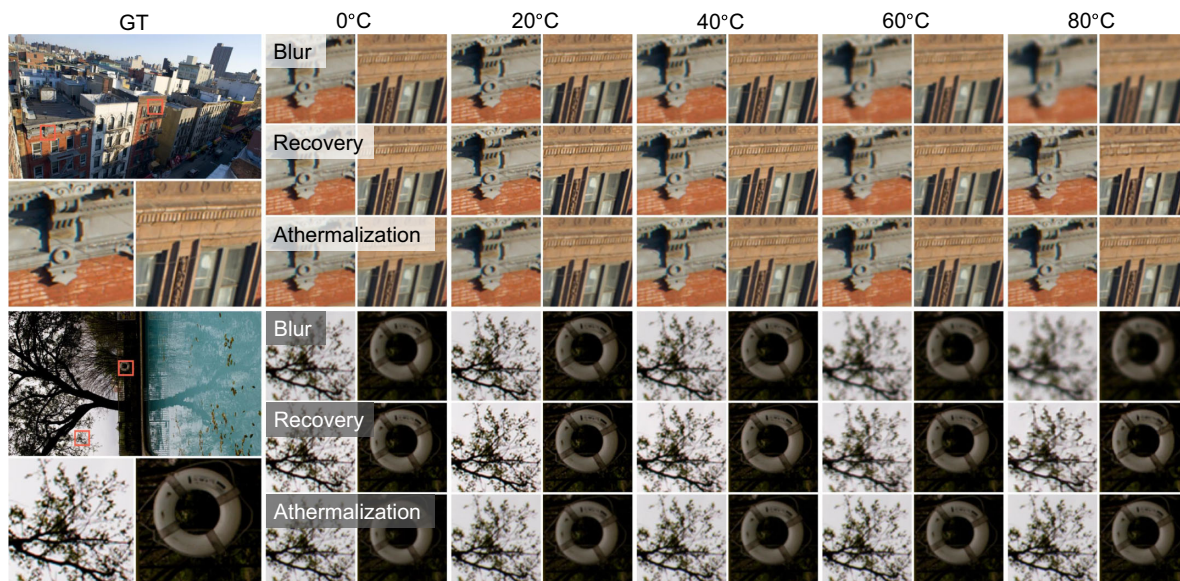


Figure 4. Recovery results of the multibranch model at different temperatures. Although the blur (row# 1, 4) aggravates as the temperature moves away from the design one, the recovery results (row# 2, 5) are relatively robust and exhibit similar to or even slightly better than the images directly obtained by the vehicle lens with athermalization (row# 3, 6).

the quality of corrupted images, both visually and quantitatively, and makes them as clear and photorealistic as possible.

Although the traditional passive optical athermalization offers relatively stable image quality over the entire temperature range, it trades off the performance at 0–40 °C, which is the most commonly used perception scope in daily life. By comparison, our model produces better results at –20 to 60 °C, which can cope with most application scenarios. We note that the recovery PSNR values are slightly inferior to athermalization beyond 60 °C, where the multibranch model still improves the blur images by 2.36 dB on average.

Notably, we adopt the same network architecture, hyperparameters (e.g., learning rate, feature channels, etc.), and scale of the training dataset for a fair comparison at all temperatures, leading to inferior results than athermalization at extremely high temperatures. In Section 4.3 of the Supplement, we further explore the insight of fine-tuning parameters to enable resolved images to outperform those captured by athermalization at extreme temperatures. For instance, with the data augmentation, our

multibranch model could significantly increase the improvement in PSNR, marked as the yellow star shown in Figure 5. Nevertheless, our multibranch strategy is an effective yet simple solution to free optical systems from the troublesome design process.

4.3. Real-Capture Experimental Results

To verify the practicability of our strategy, we test on a phone camera module consisting of several plastic aspheric lenses which lead to unneglectable thermo-optic effects. For acquiring massive datasets at different temperatures, we introduce the fine-tuning display-capture setup where the module is fixed in the chamber (see Figure 3). Based on the daily usage of phones, we adopt 0–60 °C as the experimental region and calculate the PSNR of the blur test sets 10 °C apart. Analyzing the PSNR-temperature plot, we divide this range into three subranges of 0–10 °C, 20–50 °C, and 60 °C. Subsequently, we blend the datasets inside

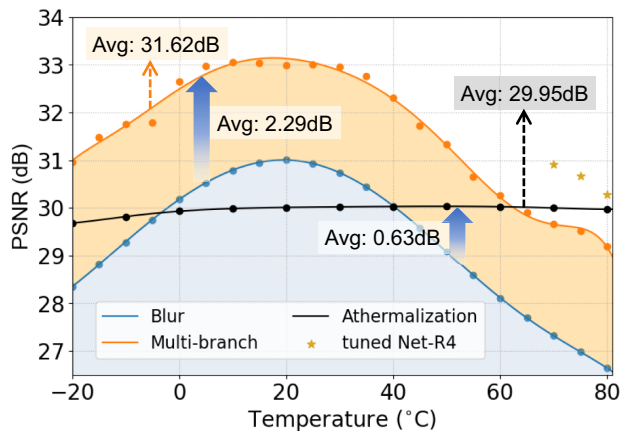


Figure 5. Improvement comparison between the multibranch model and athermalization at -20 to 80°C . The multibranch model (orange curve) increases the image quality of the captured ones (blue curve) by 2.29 dB on average, which is significantly superior to that of the lens with athermalization (black curve) at -20 to 60°C . With the extra data augmentation, our multibranch strategy further improves the recovery quality at extremely high temperatures (marked as yellow stars) and noticeably outperforms the conventional passive athermalization technique.

the subranges and train a model separately to combine the final multibranch reconstruction model. As noted, the display-captured setup also records the thermal noise in the dataset, which is consistent with the actual situation (see division details of the phone module in Section 3 of the Supplement).

4.3.1. Display-Capture Verification

We, respectively, reserve 50 display-capture images at different temperatures as the test set. From the results shown in

Figure 6, we observe that the image quality drops significantly when the temperature deviates away from 20°C , especially at 0°C and 60°C . Through our multibranch model retrained on the display-capture dataset, the PSNR shows an increase of 0.93 dB on average and 1.5 dB at extremely low temperatures. The LPIPS also showcases the same tendency with an average drop of 0.09 . Visually, the recovered images reproduce more identified details even at high frequency, e.g., the pattern on the parrot's face and the zebra crossing on the overpass in Figure 6. In addition, the captured images exhibit a bit of color bias due to the limited spectrum of the LCD monitor and built-in ISP. Our model also corrects tone automatically and leads to more photorealistic results.

4.3.2. Generalization Analysis

Although images displayed on the monitor cannot faithfully match reality due to the limited dynamic range and spectrum of the screen, our model trained on such datasets still works in real scenes to some extent. We demonstrate the generalization with the recovery results in several outer and indoor scenarios at the low, mid, and high temperatures, as shown in Figure 7. When captured in the wild, the detector delivers the temperature feedback for automatic branch selection. Experimental results suggest that our model can perform fairly well in practice, e.g., clearer textures of the stone lion, sharper edges of the succulent plant, and more comfortable color of the overall images.

4.4. Limitations and Future Work

We are aware that several limitations in our work hinder us from further exploring the temperature impact on shallow-designed imaging systems. For instance, it is straightforward to facilitate the image degradation at extreme temperatures by simulation,

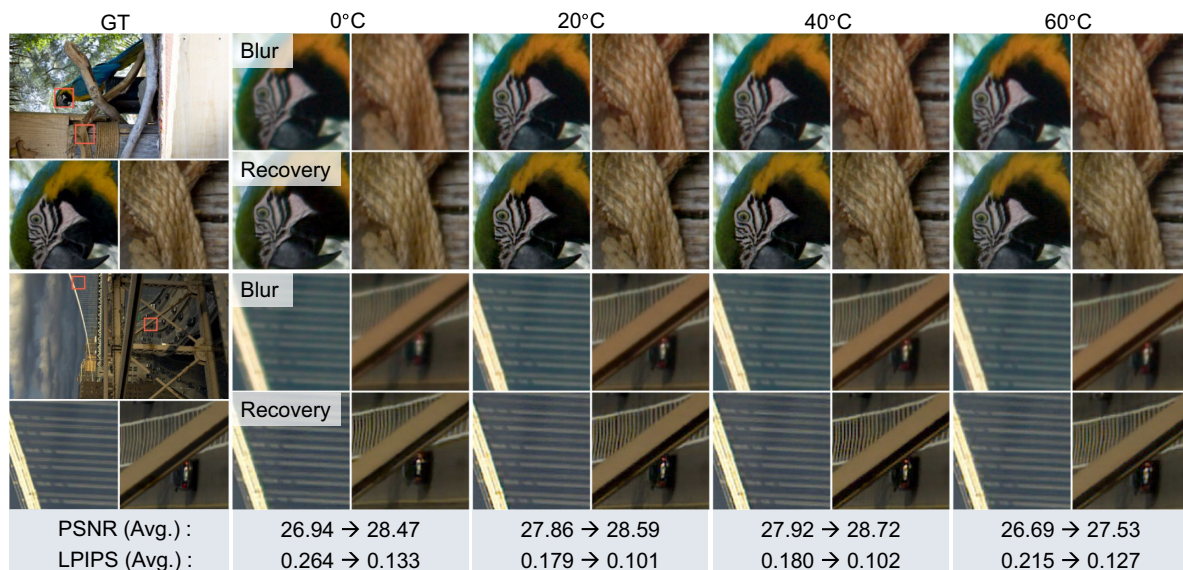


Figure 6. Experimental recovery results based on the dataset acquired from the display-capture setup. Extremely low or high temperatures aggravate image degradation while our model still realizes relatively robust high-quality results with details and true color. The labels illustrate the average improvement in PSNR and LPIPS of images at different temperatures.

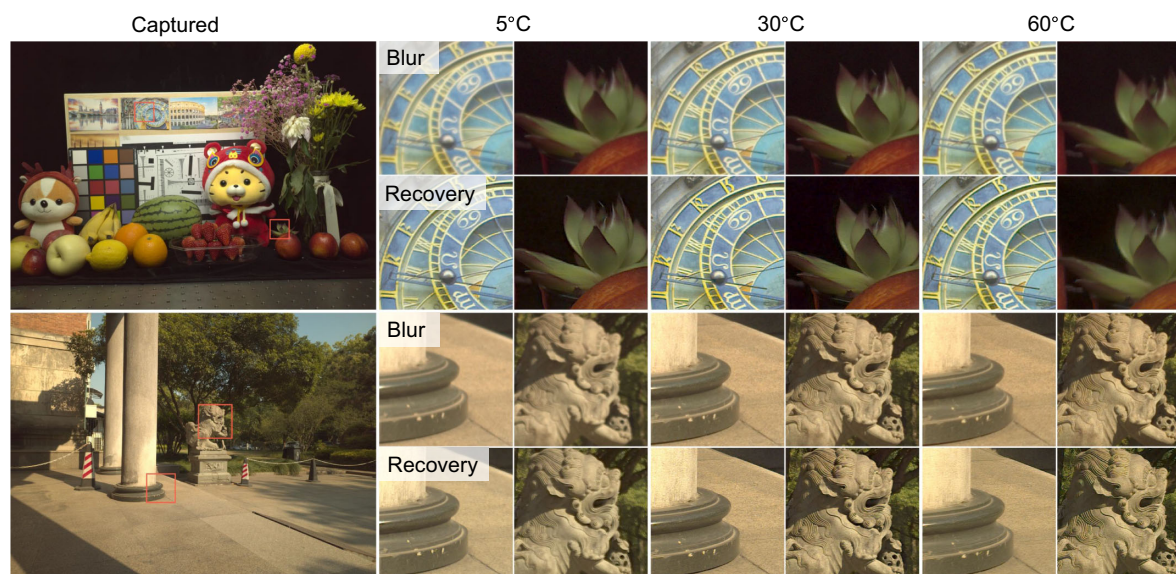


Figure 7. Real-world generalization results. We show the local details and their recovery results at three temperatures in different ranges. Although the training dataset cannot fully match the reality, our model offers robust clear output images against blur and noise.

but cumbersome and unruly in the real world. During the time the temperature rises in the chamber, lighting conditions and content features of the scene are likely to change, leading to the indistinct comparison under different temperatures. In addition, the depth of field brings blur bokeh which shows similar visual effects with defocus blur, raising difficulties for our model to recover. The further discussion can be found in Section 4.4 of the Supplement.

In the next step, more advanced network architectures may be worth exploring to address blur images with excessive degradation. To this point, our model can still be valuable in constructing complex but efficient image recovery networks at extreme temperatures, and remaining lightweight architecture at room temperature regions due to the independent branches. Meanwhile, the enormous network architecture requirements urge us to seek an acceleration for the network processing, which is crucial for applications like vehicle cameras. Last but not least, the proposed temperature-robust imaging scheme can be elaborated to combat with the impact of other damp or dusty environments, where there is to date no proper metrics, through emerging transfer learning strategies.

5. Conclusion

We have developed a temperature-robust learned image recovery for shallow-designed imaging systems and verified its superiority both theoretically and experimentally. To tackle feature crosstalk due to temperature deviations, a series of datasets are facilitated to train the corresponding recovery network as a branch of the full model. To reduce computational expenses, a scheme of temperature division and mixture of multiple datasets is particularly exploited. Experimental results showcase the feasibility of our temperature-robust learned model, which outperforms traditional athermalization approaches both visually and quantitatively. Overall, this is a novel approach for the athermalization

technology from the emerging deep learning perspective. Without complex optical designs and expensive materials, it allows shallow-designed imaging systems to acquire high-quality images at unconventional temperatures and lowered manufacturing costs.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

W.C. and B.Q. contributed equally to this work. The authors gratefully acknowledge Tao Wang for assisting in the vehicle lens design without athermalization; Yanru Shu, Xiangyu Meng, and Rui Wang for technical support in algorithmic implementation; and Sunny Optical Technology Co. for offering the tested phone lens module. This work was supported by Zhejiang University Education Foundation Global Partnership Fund (grant no. 100000-11320), Fundamental Research Funds for the Central Universities (grant no. 226-2022-00137), and Faculty Start-up Fund of the University of Hong Kong.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

computational imaging, deep learning, generative adversarial networks (GANs), multibranch models, temperature-robust imaging

Received: May 31, 2022
Revised: July 10, 2022
Published online: September 23, 2022

- [1] C. Zhang, Z. Shi, B. Xu, B. Feng, *Infrared Phys. Technol.* **2016**, *74*, 72.
- [2] E. R. Fossum, D. B. Hondongwa, *IEEE J. Electron Devices Soc.* **2014**, *2*, 33.
- [3] S. Chen, H. Feng, D. Pan, Z. Xu, Q. Li, Y. Chen, *ACM Trans. Graph.* **2021**, *40*, 5.
- [4] A. Abdelhamed, S. Lin, M. S. Brown, in *2018 IEEE/CVF Conf. Computer Vision and Pattern Recognition*, IEEE, Piscataway, NJ **2018**, pp. 1692–1700.
- [5] A. R. Nasyrov, *J. Opt. Technol.* **2017**, *84*, 205.
- [6] J. Wang, C. Xue, *Appl. Opt.* **2015**, *54*, 9665.
- [7] C. W. Kuo, C. L. Lin, C. Y. Han, *Appl. Opt.* **2010**, *49*, 3691.
- [8] H. S. Yang, H. Kihm, I. K. Moon, G. J. Jung, S. C. Choi, K. J. Lee, H. Y. Hwang, S. W. Kim, Y. W. Lee, *Appl. Opt.* **2011**, *50*, 6206.
- [9] S. Chen, Z. Fan, Z. Xu, B. Zuo, S. Wang, H. Xiao, *Opt. Lett.* **2011**, *36*, 3021.
- [10] B. Feng, Z. Shi, Y. Zhao, H. Liu, L. Liu, *Infrared Phys. Technol.* **2017**, *87*, 11.
- [11] S. Nah, T. H. Kim, K. M. Lee, in *2017 IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, IEEE, Piscataway, NJ **2017**, pp. 257–265.
- [12] K. Zhang, W. Zuo, Y. Chen, D. Meng, L. Zhang, *IEEE Trans. Image Process.* **2017**, *26*, 3142.
- [13] A. Abuolaim, M. S. Brown, in *European Conf. Computer Vision*, Springer **2020**, Cham, pp. 111–126.
- [14] P. Liu, H. Zhang, J. Wang, Y. Wang, D. Ren, W. Zuo, <https://arxiv.org/abs/2206.03691> (accessed: 2022).
- [15] L. Ruan, B. Chen, J. Li, M. Lam, <https://arxiv.org/abs/2204.00367> (accessed: 2022).
- [16] T. Remez, O. Litany, R. Giryes, A. M. Bronstein, Deep convolutional denoising of low-light images **2017**, *arXiv:1701.01687*.
- [17] Z. Hu, S. Cho, J. Wang, M.-H. Yang, in *2014 IEEE Conf. Computer Vision and Pattern Recognition*, IEEE, Piscataway, NJ **2014**, pp. 3382–3389.
- [18] J. Deng, G. Luo, C. Zhao, *IET Image Process.* **2020**, *14*, 3613.
- [19] S. W. Hasinoff, D. Sharlet, R. Geiss, A. Adams, J. T. Barron, F. Kainz, J. Chen, M. Levoy, *ACM Trans. Graph.* **2016**, *35*, 6.
- [20] E. Tseng, A. Mosleh, F. Mannan, K. St-Arnaud, A. Sharma, Y. Peng, A. Braun, D. Nowrouzezahrai, J.-F. Lalonde, F. Heide, *ACM Trans. Graphics* **2021**, *40*, 1.
- [21] A. Abdelhamed, S. Lin, M. Brown, in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, IEEE, Piscataway, NJ **2018**, pp. 1692–1700.
- [22] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, *Commun. ACM* **2020**, *63*, 139.
- [23] *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (Eds: O. Ronneberger, P. Fischer, T. Brox, I. N. Navab, J. Hornegger, W. M. Wells, A. F. Frangi), Springer International Publishing, Cham **2015**, pp. 234–241.
- [24] *Computer Vision – ECCV 2016* (Eds: J. Johnson, A. Alahi, L. Fei-Fei, I. B. Leibe, J. Matas, N. Sebe, M. Welling), Springer International Publishing, Cham **2016**, pp. 694–711, ISBN 978-3-319-46475-6.
- [25] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, W. Shi, in *2017 IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, IEEE, Piscataway, NJ **2017**, pp. 105–114.
- [26] M. Arjovsky, S. Chintala, L. Bottou, *Wasserstein Gan* **2017**, *arXiv:1701.07875*.
- [27] K. Simonyan, A. Zisserman, *Very Deep Convolutional Networks for Large-Scale Image Recognition* **2015** *arXiv:1409.1556*.
- [28] Y. Peng, Q. Sun, X. Dun, G. Wetzstein, W. Heidrich, F. Heide, in *ACM Transactions on Graphics*, New York, USA, vol. **38** **2019**, ISSN 15577368.
- [29] B. Qi, W. Chen, X. Dun, X. Hao, R. Wang, X. Liu, H. Li, Y. Peng, *Appl. Opt.* **2022**, *61*, 1097.
- [30] V. Bychkovsky, S. Paris, E. Chan, F. Durand, in *The Twenty-Fourth IEEE Conf. Computer Vision and Pattern Recognition*, IEEE, Piscataway, NJ **2011**.
- [31] J. M. Geary, *Introduction to Lens Design: With Practical ZEMAX Examples*, Willmann-Bell, Richmond **2002**.
- [32] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, O. Wang, in *2018 IEEE/CVF Conf. Computer Vision and Pattern Recognition*, IEEE, Piscataway, NJ **2018**, pp. 586–595.
- [33] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, *IEEE Trans. Image Process.* **2004**, *13*, 600.