Proceedings of CONF-CDS 2026 Symposium: Machine Learning and Neural Network Applications in Engineering
DOI: 10.54254/2755-2721/2026.AST38671

Mobile Camera Colour Drift and Robustness of Lightweight Visual Models under High-Temperature Conditions: A Narrative Review

Han Shu

Department of Computing, School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou, China
Han.Shu25@student.xjtlu.edu.cn

Abstract. High-temperature operation affects mobile vision at capture as well as during inference. Sunlight and sustained computation heat the lens-sensor assembly, where changes in focus and dark signal are subsequently modified by the image signal processor. This narrative review examines how hardware athermalisation, image restoration and lightweight visual models address the resulting degradation. Chen et al.'s temperature-conditioned recovery provides direct thermal evidence, whereas cross-camera colour constancy and compact task models contribute methods largely tested under other conditions. The distinction matters when judging transfer to a phone in sustained use. Optical compensation can retain detail that software would otherwise have to estimate; learned correction remains dependent on the camera and its calibration data. A faster task model may still make an unreliable decision from the corrected image. Across the reviewed studies, pairwise integration is better established than a complete thermal pipeline spanning capture, correction and task adaptation. Evaluation needs to distinguish device identity from thermal state and measure image quality alongside task accuracy, energy and latency during real heating and cooling. A lightweight, camera-aware design is plausible, but its cross-device reliability remains to be demonstrated.

Keywords: mobile-camera colour drift, high-temperature imaging, colour constancy, lightweight convolutional neural network, vision-language model

1. Introduction and background

1.1. Application context

Mobile cameras support navigation, augmented reality and visual assistants, often with lightweight networks running on the device [1]. Sunlight, poor ventilation and sustained processing expose the camera to heat; prolonged workloads can also trigger thermal throttling [2]. The imaging problem begins before inference. Heating changes the lens and sensor, introducing blur, noise or colour drift that may affect downstream decisions even when computation remains fast enough.

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.AST38671

1.2. Research significance

Heat can shift the focal plane and increase dark current before an image enters the image signal processor (ISP) [3, 4]. Demosaicing, denoising, white balance and tone mapping then reshape these disturbances, so the processed red-green-blue (RGB) image does not directly reveal sensor drift [5]. Models trained on clean images often lose accuracy under blur or noise [6], while colour-dependent questions remain difficult for multimodal systems even without physical heating [7]. An image that looks acceptable may still misrepresent a small warning label or traffic signal. Image quality and task reliability require separate assessment.

1.3. Aim and scope of this review

Gijsenij et al. review computational colour constancy and compare illuminant estimation methods [8], without organising their analysis around thermal degradation in mobile lens-sensor-ISP chains. Li et al.'s low-light review already includes cross-phone testing and downstream face detection [9]. These evaluation ideas are established; the acquisition condition in that review is low illumination rather than controlled heating and cooling.

The gap examined here is the incomplete connection between thermal image formation, camera-specific correction and task reliability during sustained mobile operation. This narrative review compares hardware, image-level and task-level interventions for heat-affected visible-light imaging. It separates direct thermal evidence from adjacent work and asks which combinations remain plausible under mobile power, latency and calibration constraints. Its contribution is this integrated assessment, rather than a new restoration method or a claim of universal thermal robustness.

2. Degradation mechanisms and intervention strategies

Intervention can preserve information at capture, correct a degraded frame or improve a decision from the remaining evidence. These outcomes differ: tolerating blur does not restore lost texture, and a plausible image can still yield an incorrect detection.

2.1. Mechanisms of high-temperature imaging degradation

Heat changes refractive index, lens spacing and support structures. When these effects do not cancel, the focal plane and point-spread function vary with temperature; Chen et al. modelled this behaviour directly [3]. Blur also depends on field position and wavelength, making a single Gaussian kernel a coarse proxy. Lens materials, geometry and barrel expansion help explain why the response differs across modules [10-12].

Dark current and fixed-pattern noise introduce a sensor-level disturbance whose visibility depends on temperature, illumination and exposure [13]. Pixel-level temperature sensing can compensate part of the dark signal, but not thermal defocus [4]. Cameras also have different spectral sensitivities: identical scenes and illuminants need not produce identical channel ratios [14]. White balance, denoising and tone mapping further alter the disturbance in unprocessed sensor data (RAW data) [5]. A yellow cast on one phone is therefore not a universal thermal law. Its direction and magnitude depend on the lens-sensor-ISP chain, and equal ambient temperatures do not imply equal degradation.

DOI: 10.54254/2755-2721/2026.AST38671

2.2. Source-level intervention: hardware athermalisation

Passive optical athermalisation combines materials or diffractive elements with partly cancelling thermal responses [10]. It adds no inference cost, although lens shape, chromatic performance and assembly tolerance remain coupled. Designs reported over -60 to 100 degrees C are still module-specific [11]. Mechanical compensation uses differential expansion to move the lens or detector [12]; space, ageing and thermal hysteresis complicate its use in a phone.

Active compensation can use temperature feedback to move a lens group, adding power and calibration requirements [15]. Wavefront coding instead reduces optical sensitivity to defocus and uses software decoding; one infrared design covered -40 to 60 degrees C [16]. These optical interventions can preserve spatial information at capture, but do not by themselves correct sensor- or ISP-induced colour error. Infrared demonstrations also require caution when extrapolated to visible-light phone modules.

2.3. Image-level intervention: thermal restoration and colour correction

Restoration estimates useful detail from a degraded frame. Colour constancy makes a narrower chromatic correction and cannot recover detail removed by defocus.

2.3.1. Temperature-conditioned learned restoration

Chen et al. use measured temperature to select a branch of a Wasserstein generative adversarial network [3]. In their simulated vehicle-lens study, image-quality differences divide the 0-80 degrees C range into four operating bands. Conditioning on temperature avoids training a model for every temperature while retaining some specificity to the degradation state.

The simulated and captured results need to be distinguished. Mean peak signal-to-noise ratio (PSNR) improved by about 2.29 dB over the blurred input in the vehicle-lens simulation. In a separate display-capture experiment, a camera module was tested in a controlled chamber from 0 to 60 degrees C; mean PSNR improved by about 0.93 dB and learned perceptual image patch similarity (LPIPS) decreased by 0.09 [3]. Network compression reduced the parameter count from approximately 2.15 x 10^8 to 1.20 x 10^7. These are camera-specific image-recovery results, not measurements of downstream task improvement.

Hard branch boundaries may cause discontinuities during heating or cooling, and one temperature reading cannot describe a non-uniform thermal field. The display-capture setup offers limited depth and spectral variation; the paper's qualitative indoor and outdoor examples do not establish transfer across camera models [3]. Calibrated colour recovery and downstream accuracy remain uncertain. Generative restoration adds a further risk: plausible texture can be synthesised without measurement support.

2.3.2. Colour correction and cross-camera adaptation

Convolutional colour constancy estimates illumination in log-chrominance space [17]; FC4 combines local estimates through confidence-weighted pooling [18]. Both depend on the training camera. Multi-domain learning shares most parameters and reserves a small component for each device [19].

C5 adapts a colour-constancy model using unlabelled RAW images from an unseen camera [14]. Its approximately 2 MB model runs in about 7 ms on a graphics processing unit or 90 ms on a

DOI: 10.54254/2755-2721/2026.AST38671

central processing unit, making it a useful calibration baseline. It estimates global illumination, however, without modelling thermal defocus, dark-current growth or ISP drift. Camera adaptation can complement thermal restoration; the two solve different parts of the problem.

2.3.3. Lightweight mobile image processing

Compact convolutional neural networks reproduce aspects of smartphone colour rendition and sharpness [20], while LAN targets efficient RAW-to-RGB processing [21]. These learned ISP approaches make mobile image correction feasible under ordinary conditions. Neither study tests physical heating, and clipped colour or severely lost detail cannot be reliably recovered by software alone.

2.4. Task-level intervention with lightweight visual models

Task models seek a reliable label, box, track or answer. This can reduce reconstruction demands, but sensitivity depends on the task: a mild colour cast may leave shape recognition intact while corrupting a traffic-signal decision.

2.4.1. Lightweight convolutional networks under colour and thermal shift

The MobileNet family establishes the feasibility of efficient mobile inference [1, 22, 23]. MobileNetV3-Large reported 3.2% higher ImageNet accuracy at 20% lower latency than MobileNetV2 [1]. Efficiency does not establish robustness to heated-camera input: blur and noise reduce clean-trained accuracy [6], and synthetic colour shifts can substantially change predictions [24].

Training could combine realistic colour drift, blur and signal-dependent noise with feature consistency or a small camera adapter. Invariance must remain selective, preserving colour that carries task meaning. Runtime requires separate attention. Thermal-aware control reduced selected smartphone workload latency by up to 18% while avoiding throttling [2], and mobile graphics processors support on-device inference [25]. These computing results do not measure thermal camera degradation.

2.4.2. Small vision-language models

Vision-language models (VLMs) connect a visual encoder to language processing, as in LLaVA [26]. MobileVLM and TinyLLaVA extend this approach to smaller systems [27, 28]; MobileVLM reported 21.5 generated tokens/s on Snapdragon 888 [27]. Neither study evaluated heated-camera frames. Defocus can erase text or small symbols before encoding, while camera-specific colour drift alters cues that later reasoning cannot reconstruct.

ColorBench found uneven colour perception and reasoning across 32 VLMs [7]. Robust-LLaVA addresses adversarial robustness, without testing thermal defocus, sensor noise or ISP drift [29]. Within these studies, dedicated high-temperature mobile-camera evaluation remains absent. Joint degradations during training and an uncertainty mechanism that can request another frame are plausible safeguards, but require thermal validation.

Table 1 compares the three intervention layers. Combining them is a practical design hypothesis; superiority over single-layer alternatives has not been established across heated mobile cameras.

DOI: 10.54254/2755-2721/2026.AST38671

Table 1. Comparison of the three intervention layers for high-temperature mobile vision

Comparison dimension	Hardware-level intervention	Image-level intervention	Task-level intervention
Point and target	Before or during capture; limits defocus and selected sensor effects.	After capture; corrects blur, noise or colour in RAW or linear RGB.	During inference; improves decisions under residual degradation.
Methods and use	Optical or mechanical compensation, active focus and dark-signal correction [4, 10-12, 15, 16]; suited to stable modules.	Thermal restoration and colour calibration [3, 14, 17-21]; suited to updatable cameras with calibration data.	Lightweight networks and corruption-aware training [1, 6, 22-29]; suited to resource-limited recognition or reasoning.
Main strength	Preserves captured information; passive designs need little runtime computation.	Can be updated after manufacture and adapted to a camera.	Targets task accuracy with lower memory and latency.
Main limitation	Module-specific cost, space and calibration; colour errors can remain.	Limited transfer; clipping and severe information loss are not reliably reversible; generated detail may be unsupported.	Does not restore the image; task-relevant colour may be suppressed; heated-camera evidence remains limited.
Evaluation	Focus stability, optical response, sensor noise and thermal repeatability.	PSNR, LPIPS, colour and illuminant error; preservation of task cues.	Task accuracy, colour sensitivity and uncertainty; sustained latency, energy and temperature.

2.5. Joint optimisation and hybrid intervention pipelines

Existing integration mainly couples two layers. Wavefront coding joins optical compensation with computational decoding in an infrared system [16], while Chen et al. pair shallow-designed optics with temperature-conditioned recovery, without downstream task training [3]. Dirty Pixels jointly learns RAW processing and classification under low light, without evaluating a controlled thermal sweep [30]. C5 and multi-domain colour constancy offer camera-adaptation components [14, 19]. These precedents have not established a fully optimised, thermally validated three-layer pipeline across devices.

A candidate workflow would limit degradation through optical or focus compensation, correct residual blur and colour using camera-state information, then adapt a lightweight task model. It remains a proposed direction. Joint optimisation needs real thermal data with task labels and a model of the residual degradation after hardware compensation. Restoration must preserve measured evidence rather than invent detail, and switching or adaptation must remain stable through heating

DOI: 10.54254/2755-2721/2026.AST38671

and cooling. Added computation itself generates heat [2], so fidelity, task accuracy, latency and energy cannot be optimised independently of thermal state. Matched-budget ablations of single layers, pairwise combinations and the full pipeline are needed to test when integration helps.

3. Challenges and future directions

3.1. Limited real high-temperature data

Real heated-camera data remain limited in the reviewed work [3]. Synthetic RAW pipelines reproduce noise and processing, but not physical warm-up [5]. Useful comparisons require several phone models and repeated units, paired RAW and processed frames where available, neutral targets and natural scenes, and controlled camera-state and thermal metadata. Sunlight and processor-driven heating should be recorded separately because their internal temperature fields differ.

3.2. Cross-camera generalisation

A model trained on one phone may absorb its ISP or white-balance behaviour and fail when another camera produces a similar cast for a different reason. Thermal transfer confounds device identity with changing thermal state. A shared model with a lightweight device-conditioned adapter could separate them, drawing on multi-domain learning and C5 [14, 19]. Their camera adaptation is a precedent, not proof of thermal transfer.

Calibration should span a reference temperature and a thermal sweep, with paired RAW and processed frames helping to distinguish capture error from ISP effects. Camera-state and thermal metadata are needed to control exposure, focus and workload; ambient temperature alone is insufficient. Device-specific forward models can generate training perturbations, but require checks against real measurements. Validation should hold out both camera models and physical units, include actual heating and cooling, and report colour and task errors for each device rather than only pooled means.

3.3. Disconnect between image quality and task performance

Better PSNR, LPIPS or colour error does not automatically imply better classification, detection or visual question answering [3, 6, 7]. Smoothing may raise PSNR while erasing a small coloured cue; a visible cast may leave a shape-based decision unchanged. Image and task metrics should be reported together, using both colour-dependent and colour-independent tasks. This separates improved appearance from preserved decision-relevant evidence.

3.4. Continuous adaptation and mobile co-design

Continuous state conditioning could reduce the discontinuities of discrete temperature branches [3]. Correction need not run on every frame: metadata or image statistics could trigger an update, with uncertainty reported by the task model. These choices require testing during warm-up and cooling, including whether switching introduces flicker or unstable predictions.

Adaptation also consumes thermal headroom. A larger model may improve one frame yet increase power demand and bring throttling forward [1, 2]. Measurements after warm-up should report task accuracy, colour error, energy, latency and device temperature separately. Cold-start speed alone gives an incomplete account of sustained operation.

DOI: 10.54254/2755-2721/2026.AST38671

4. Conclusion

Thermal degradation is camera-dependent: optical drift, sensor response and ISP processing jointly affect the evidence available to a visual model. Temperature-conditioned restoration improves image quality on particular cameras [3], while lightweight architectures establish mobile inference feasibility [1, 27]. Neither finding alone demonstrates dependable downstream performance under heat.

Cross-layer co-design is a plausible route, provided capture degradation remains recoverable and correction preserves task-relevant detail. The decisive tests are cross-device measurements during real heating and cooling, with colour-sensitive tasks and sustained power and latency constraints. Such evaluation is still needed before a hybrid pipeline can be treated as a general mobile solution.

References

[1] Howard, A., Sandler, M., Chu, G., et al. (2019). Searching for MobileNetV3. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul, Republic of Korea. pp. 1314–1324. https://doi.org/10.1109/ICCV.2019.00140.

[2] Srinivasa, G. P., Werner, D., Hempstead, M., et al. (2021). Thermal-aware overclocking for smartphones. In: *Proceedings of the IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS)*. Virtual conference. pp. 220–222. https://doi.org/10.1109/ISPASS51385.2021.00039.

[3] Chen, W., Qi, B., Liu, X., et al. (2022). Temperature-robust learned image recovery for shallow-designed imaging systems. *Advanced Intelligent Systems, 4*(10): 2200149. https://doi.org/10.1002/aisy.202200149.

[4] Xie, S., Abarca Prouza, A., Theuwissen, A. J. P. (2020). A CMOS-imager-pixel-based temperature sensor for dark current compensation. *IEEE Transactions on Circuits and Systems II: Express Briefs, 67*(2): 255–259. https://doi.org/10.1109/TCSII.2019.2914588.

[5] Brooks, T., Mildenhall, B., Xue, T., et al. (2019). Unprocessing images for learned RAW denoising. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA. pp. 11036–11045. https://openaccess.thecvf.com/content_CVPR_2019/html/Brooks_Unprocessing_Images_for_Learned_Raw_Denoising_CVPR_2019_paper.html.

[6] Hendrycks, D., Dietterich, T. (2019). Benchmarking neural network robustness to common corruptions and perturbations. *International Conference on Learning Representations (ICLR)*, New Orleans, LA, USA. https://arxiv.org/abs/1903.12261.

[7] Liang, Y., Li, M., Fan, C., et al. (2025). ColorBench: Can VLMs see and understand the colorful world? A comprehensive benchmark for color perception, reasoning, and robustness. In *Advances in Neural Information Processing Systems 38 (NeurIPS 2025), Datasets and Benchmarks Track*. San Diego, CA, USA and Mexico City, Mexico. https://proceedings.neurips.cc/paper_files/paper/2025/hash/d34497330b1fd6530f7afd86d0df9f76-Abstract-Datasets_and_Benchmarks_Track.html.

[8] Gijsenij, A., Gevers, T., & van de Weijer, J. (2011). Computational color constancy: Survey and experiments. *IEEE Transactions on Image Processing, 20*(9), 2475–2489. https://doi.org/10.1109/TIP.2011.2118224.

[9] Li, C., Guo, C., Han, L., et al. (2021). Low-light image and video enhancement using deep learning: A survey. *arXiv preprint arXiv:2104.10729*. https://arxiv.org/abs/2104.10729.

[10] Wang, J., & Xue, C. (2015). Athermalization and thermal characteristics of multilayer diffractive optical elements. *Applied Optics, 54*, 9665–9670. https://doi.org/10.1364/AO.54.009665.

[11] Xu, S., Peng, X., Tie, G., et al. (2019). Athermalization of an optical system based on lens shape and assembly method. *Current Optics and Photonics, 3*(5), 429–437. https://doi.org/10.3807/COPP.2019.3.5.429.

[12] Yang, H. S., Kihm, H., Moon, I. K., et al. (2011). Three-shell-based lens barrel for the effective athermalization of an IR optical system. *Applied Optics, 50*(33), 6206–6213. https://doi.org/10.1364/AO.50.006206.

[13] Lin, D. L., Wang, C. C., & Wei, C. L. (2010). Quantified temperature effect in a CMOS image sensor. *IEEE Transactions on Electron Devices, 57*(2), 422–428. https://doi.org/10.1109/TED.2009.2037389.

[14] Afifi, M., Barron, J. T., LeGendre, C., et al. (2021). Cross-camera convolutional color constancy. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. Montreal, QC, Canada. pp. 1981–1990.

DOI: 10.54254/2755-2721/2026.AST38671

https://openaccess.thecvf.com/content/ICCV2021/html/Afifi_Cross-Camera_Convolutional_Color_Constancy_ICCV_2021_paper.html.
[15] Kuo, C. W., Lin, C. L., & Han, C. Y. (2010). Dual field-of-view midwave infrared optical design and athermalization analysis. *Applied Optics, 49*(19), 3691–3700. https://doi.org/10.1364/AO.49.003691.
[16] Chen, S., Fan, Z., Xu, Z., et al. (2011). Wavefront coding technique for controlling thermal defocus aberration in an infrared imaging system. *Optics Letters, 36*(16), 3021–3023. https://doi.org/10.1364/OL.36.003021.
[17] Barron, J. T. (2015). Convolutional color constancy. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. Santiago, Chile. pp. 379–387. https://openaccess.thecvf.com/content_iccv_2015/html/Barron_Convolutional_Color_Constancy_ICCV_2015_paper.html.
[18] Hu, Y., Wang, B., & Lin, S. (2017). FC4: Fully convolutional color constancy with confidence-weighted pooling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI, USA. pp. 4085–4094. https://openaccess.thecvf.com/content_cvpr_2017/html/Hu_FC4_Fully_Convolutional_CVPR_2017_paper.html.
[19] Xiao, J., Gu, S., & Zhang, L. (2020). Multi-domain learning for accurate and few-shot color constancy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA. pp. 3258–3267. https://openaccess.thecvf.com/content_CVPR_2020/html/Xiao_Multi-Domain_Learning_for_Accurate_and_Few-Shot_Color_Constancy_CVPR_2020_paper.html.
[20] Ignatov, A., Kobyshev, N., Timofte, R., et al. (2017). DSLR-quality photos on mobile devices with deep convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. Venice, Italy. pp. 3277–3285. https://openaccess.thecvf.com/content_iccv_2017/html/Ignatov_DSLR-Quality_Photos_on_ICCV_2017_paper.html.
[21] Raimundo, D. W., Ignatov, A., & Timofte, R. (2022). LAN: Lightweight attention-based network for RAW-to-RGB smartphone image processing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. New Orleans, LA, USA. pp. 808–816. https://openaccess.thecvf.com/content/CVPR2022W/NTIRE/html/Raimundo_LAN_Lightweight_Attention-Based_Network_for_RAW-to-RGB_Smartphone_Image_Processing_CVPRW_2022_paper.html.
[22] Howard, A. G., Zhu, M., Chen, B., et al. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861. https://arxiv.org/abs/1704.04861.
[23] Sandler, M., Howard, A., Zhu, M., et al. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Salt Lake City, UT, USA. pp. 4510–4520. https://doi.org/10.1109/CVPR.2018.00474.
[24] De, K., & Pedersen, M. (2021). Impact of colour on robustness of deep neural networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*. Montreal, QC, Canada. pp. 21–30. https://doi.org/10.1109/ICCVW54120.2021.00009.
[25] Lee, J., Chirkov, N., Ignasheva, E., et al. (2019). On-device neural net inference with mobile GPUs. *Efficient Deep Learning for Computer Vision Workshop, CVPR*, Long Beach, CA, USA. https://research.google/pubs/on-device-neural-net-inference-with-mobile-gpus/.
[26] Liu, H., Li, C., Wu, Q., et al. (2023). Visual instruction tuning. In: *Advances in Neural Information Processing Systems 36 (NeurIPS 2023)*. New Orleans, LA, USA. pp. 34892–34916. https://proceedings.neurips.cc/paper_files/paper/2023/hash/6dcf277ea32ce3288914faf369fe6de0-Abstract-Conference.html.
[27] Chu, X., Qiao, L., Lin, X., et al. (2023). MobileVLM: A fast, strong and open vision language assistant for mobile devices. arXiv preprint arXiv:2312.16886. https://arxiv.org/abs/2312.16886.
[28] Zhou, B., Hu, Y., Weng, X., et al. (2024). TinyLLaVA: A framework of small-scale large multimodal models. arXiv preprint arXiv:2402.14289. https://arxiv.org/abs/2402.14289.
[29] Malik, H. S., Shamshad, F., Naseer, M., et al. (2025). Robust-LLaVA: On the effectiveness of large-scale robust image encoders for multimodal large language models. arXiv preprint arXiv:2502.01576. https://arxiv.org/abs/2502.01576.
[30] Diamond, S., Sitzmann, V., Julca-Aguilar, F., et al. (2021). Dirty pixels: Towards end-to-end image processing and perception. *ACM Transactions on Graphics, 40*(3): Article 23, 1-15. https://doi.org/10.1145/3446918.