

Heat-assisted detection and ranging

<https://doi.org/10.1038/s41586-023-06174-6>

Received: 6 April 2021

Accepted: 5 May 2023

Published online: 26 July 2023

 Check for updates

Fanglin Bao¹, Xueji Wang¹, Shree Hari Sureshbabu¹, Gautam Sreekumar², Liping Yang¹, Vaneet Aggarwal³, Vishnu N. Boddeti² & Zubin Jacob¹✉

Machine perception uses advanced sensors to collect information about the surrounding scene for situational awareness^{1–7}. State-of-the-art machine perception⁸ using active sonar, radar and LiDAR to enhance camera vision⁹ faces difficulties when the number of intelligent agents scales up^{10,11}. Exploiting omnipresent heat signal could be a new frontier for scalable perception. However, objects and their environment constantly emit and scatter thermal radiation, leading to textureless images famously known as the ‘ghosting effect’¹². Thermal vision thus has no specificity limited by information loss, whereas thermal ranging—crucial for navigation—has been elusive even when combined with artificial intelligence (AI)¹³. Here we propose and experimentally demonstrate heat-assisted detection and ranging (HADAR) overcoming this open challenge of ghosting and benchmark it against AI-enhanced thermal sensing. HADAR not only sees texture and depth through the darkness as if it were day but also perceives decluttered physical attributes beyond RGB or thermal vision, paving the way to fully passive and physics-aware machine perception. We develop HADAR estimation theory and address its photonic shot-noise limits depicting information-theoretic bounds to HADAR-based AI performance. HADAR ranging at night beats thermal ranging and shows an accuracy comparable with RGB stereovision in daylight. Our automated HADAR thermography reaches the Cramér–Rao bound on temperature accuracy, beating existing thermography techniques. Our work leads to a disruptive technology that can accelerate the Fourth Industrial Revolution (Industry 4.0)¹⁴ with HADAR-based autonomous navigation and human–robot social interactions.

The emerging Industry 4.0 of smart technologies¹⁵ calls for a future with scalable human–robot social interactions, as it is expected that one in ten vehicles will be automated by 2030 (ref. 16) and 20 million robot helpers will be serving people¹⁷. Each of these agents will collect information about its surrounding scene through advanced sensors to make decisions without human intervention. However, simultaneous perception of the scene by numerous agents (scalable perception) is fundamentally prohibitive for active modalities because of signal interference and eye safety^{10,11}. Quasi-passive approaches such as cameras are an alternative but they rely on ambient illumination. Furthermore, cameras cannot compete with human perception even though important strides¹⁸ have recently been made on the basis of deep learning^{19,20}. It causes phenomena such as phantom braking⁹ in automated vehicles owing to the visual ambiguity and lack of physical context in perception. A shift of fully passive perception beyond traditional vision is urgently needed that can boost the AI industry (Fig. 1a,b).

An attractive approach to scalable perception is using the fully passive heat signal originating from infrared thermal radiation. Exploiting heat signals for imaging^{21–24} has well-known advantages, such as to see through the darkness or solar glare as well as bad weather²⁵, and—not surprisingly—it has been the natural choice of predators (snakes) when hunting prey (rats) at night²⁶. Nevertheless, fundamental obstacles exist for heat-assisted perception. Physical attributes

of the scene, namely, temperature (T , physical status), emissivity (e , material fingerprint) and texture (X , surface geometry) are mixed in photon streams, as objects and environment constantly emit and scatter thermal radiation. This is manifested as the ghosting effect¹², related to lack of texture in thermal imaging. Ghosting limits thermal imaging only to night-vision enhancement without any specificity even when combined with AI algorithms (see Supplementary Table 3 for a review).

TeX decomposition and TeX vision

We address the ghosting effect with an approach that we call TeX decomposition, which vividly recovers the texture from the cluttered heat signal and also accurately disentangles temperature and emissivity at the Cramér–Rao bound. Representing these decluttered TeX attributes in HSV colour space (hue = e , saturation = T , brightness = X) leads to a shift of TeX vision with physical context for machine perception (Fig. 1b,c). TeX vision empowers AI algorithms to reach information-theoretic bounds, which has—thus far—been elusive for traditional RGB or thermal vision. Figure 1c shows TeX vision for on-road and off-road scenes at night overcoming the ghosting effect (also see Supplementary Videos for video demonstrations). Our demonstrations of HADAR include detection and ranging on the basis of TeX vision,

¹Birck Nanotechnology Center, School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA. ²Michigan State University, East Lansing, MI, USA. ³School of Industrial Engineering and School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA. ✉e-mail: zjacob@purdue.edu

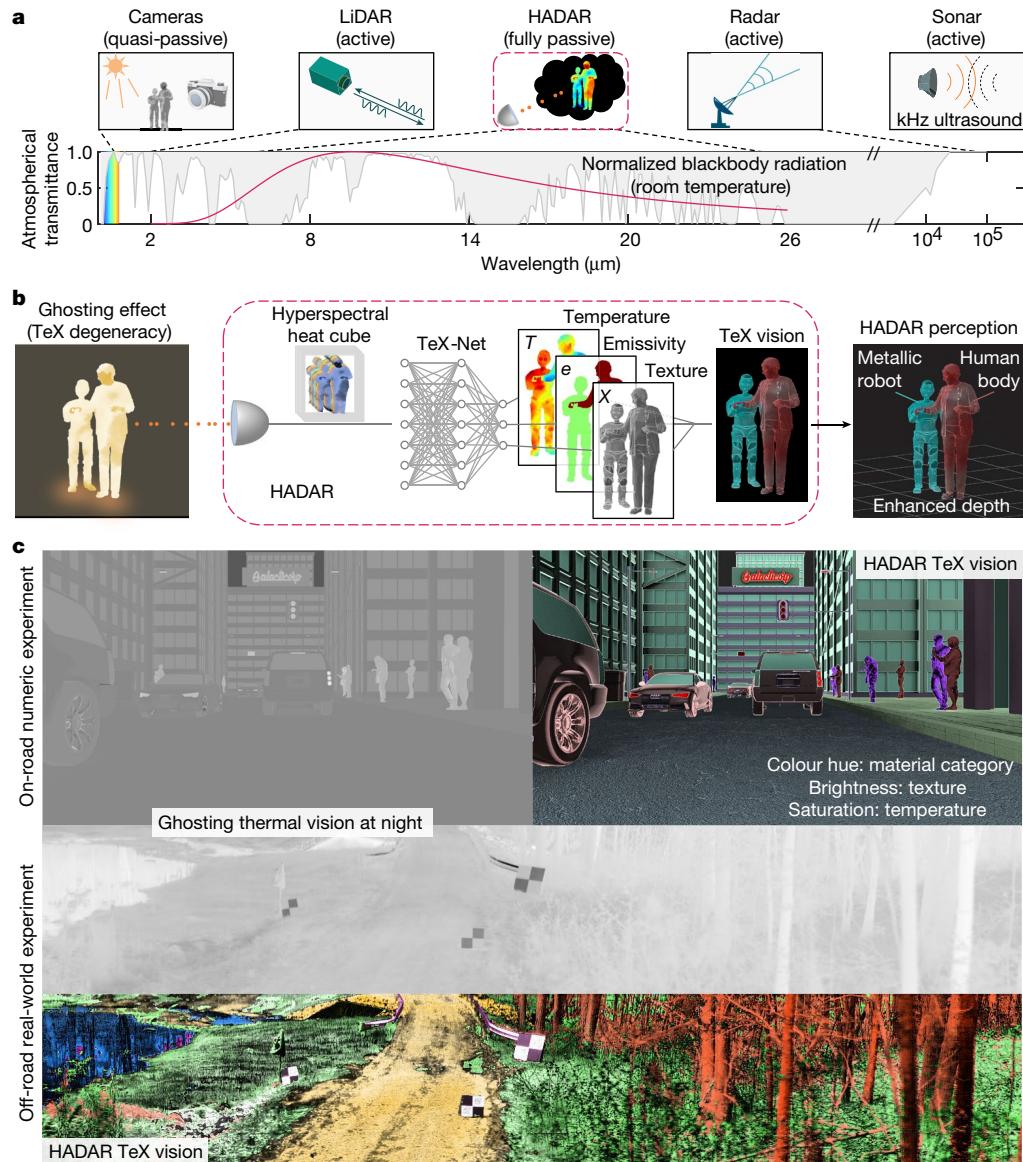


Fig. 1 | HADAR as a shift in machine perception. **a**, Fully passive HADAR makes use of heat signals, as opposed to active sonar, radar, LiDAR and quasi-passive cameras. Atmospheric transmittance window (white area) and temperature of the scene determine the working wavelength of HADAR. **b**, HADAR takes thermal photon streams as input, records hyperspectral-imaging heat cubes, addresses the ghosting effect through TeX decomposition (see Extended Data Fig. 1 for TeX-Net and see Methods for all decomposition methods) and generates TeX vision for improved detection and ranging. **c**, TeX vision demonstrated on our HADAR database and outdoor experiments (see Extended Data Figs. 2–4) clearly shows that HADAR sees textures through the darkness with comprehensive understanding of the scene.

Fig. 1 for TeX-Net and see Methods for all decomposition methods) and generates TeX vision for improved detection and ranging. **c**, TeX vision demonstrated on our HADAR database and outdoor experiments (see Extended Data Figs. 2–4) clearly shows that HADAR sees textures through the darkness with comprehensive understanding of the scene.

for both real-world-level HADAR database and outdoor experiments. We provide detailed comparisons with state-of-the-art AI-enhanced thermal sensing and prove that HADAR provides universal performance enhancement. This can lead to adoption of TeX vision as an industry standard.

For intuitive clarity, we first explain the origin of the ghosting effect using an example of thermal radiation (visible) from a light bulb. Figure 2 shows Monte Carlo path-tracing simulations of rays emanating from a bulb, with reflection of environmental emission taken into account. Geometric textures on the bulb surface can be seen only when the bulb is off. We emphasize that this texture revealed by reflection is completely lost in direct emission when the bulb is switched on, a familiar scenario from daily experience. Because every object in a complex scene emits and scatters thermal radiation, they are thermal light sources with no texture, similar to a shining bulb. The total heat signal leaving an object α has two additive contributions,

$$S_{\alpha v} = e_{\alpha v} B_v(T_\alpha) + [1 - e_{\alpha v}] X_{\alpha v}, \quad (1)$$

in which the first term is direct thermal emission (textureless) and the second term carrying texture is the environmental emission entering the detector after scattering from the object. Here subscript v denotes wavenumber (spectrum) dependence. The key difference with a shining bulb is that blackbody radiation B_v is fundamentally governed by Planck's law and cannot be switched off. Textureless thermal imaging is thus widely regarded as impossible to use for quantitative insight about a scene. The environmental thermal illumination on object α from all other objects β is given by $X_{\alpha v} = \sum_{\beta \neq \alpha} V_{\alpha \beta} S_{\beta v}$ with $V_{\alpha \beta}$ being the thermal lighting factor. The ghosting effect is exacerbated for high-emissivity materials in nature, such as skin and plants ($e \approx 1$), as the total collected signal consists of dominant direct emission and only a weak scattered signal. We note that $S_{\alpha v}$ is invariant under joint transformations of temperature T , emissivity e and texture X (see Methods),

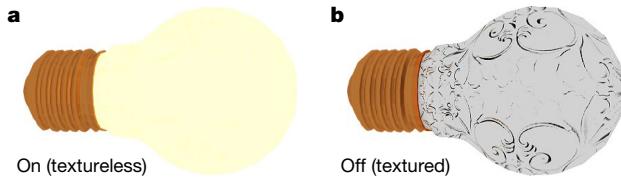


Fig. 2 | Monte Carlo path-tracing simulation of a light bulb to explain the ‘ghosting effect’. Geometric texture on a light bulb can only be seen when the bulb is off, whereas this texture is completely missing when it is glowing. The blackbody radiation can never be turned off, leading to loss of texture for thermal images. This ghosting effect presents the long-standing obstruction for heat-assisted machine perception.

which we address as TeX degeneracy. As well as the ghosting effect, this TeX degeneracy makes the separation of temperature and emissivity a considerable obstruction²⁷ to quantitative thermal sensing.

We recover the texture by breaking TeX degeneracy and discretizing spectral emissivity e_{av} into $e_\nu(m_a)$ in a material library, $\mathcal{M} = \{e_\nu(m)|m=1, 2, \dots, M\}$, that contains all possible spectral emissivity in the scene. This opportunity of dimensional reduction is available naturally in smart applications, in which materials usually have industrial standards²⁸. The material library explains the physics but requires on-site collection/calibration. We have also provided a generalized HADAR theory that does not require an input of material library (see section SV.C of the Supplementary Information). Our approach of TeX-Net uses equation (1) to design physics-based loss and uses a 3D convolutional neural network to learn spatio-spectral features in recovering texture X , temperature T and emissivity e . With general HADAR performance shown in the Extended Data, here we demonstrate the fundamental limits as well as the real-world performance of HADAR.

HADAR identifiability

We develop HADAR estimation theory to address the fundamental limits of object identification from its thermal infrared signature. We believe this will be crucial in guiding public policy for the industrial revolution, in which decision accuracy of machine perception can be bounded by physical laws as opposed to training-data volume. HADAR is distinct from traditional hyperspectral imaging, in which material difference is determined by the Euclidean distance between their reflectance spectra. By stark contrast, HADAR identifiability is determined

by multi-parameter estimation of temperature, emissivity and texture (see Supplementary Fig. 6 and relevant contexts in Supplementary Information). We exploit the multi-parameter Cramér–Rao bound and propose semantic distance to categorize objects based on their intrinsic material properties. Figure 3 shows a pertinent example of human versus robot identification. A human-shaped target (Fig. 3a) could be a human (organic skin or fabrics material) or robot (metallic) with distinct emissivity (top inset), but they will produce a visually indistinguishable incident spectrum on the detector (bottom inset; modelled by FLIR A325sc). We define HADAR identifiability as the maximum Shannon information of the target material that one can retrieve from N incident photons. It holds for all scenes (see Extended Data Fig. 6 for generalization to multi-material scenes) and is given by

$$I = \log_2 \left\{ 1 + \operatorname{erf} \left[\sqrt{\frac{Nd_0^2}{2(1+\gamma)}} \right] \right\}, \quad (2)$$

in which $\gamma = \gamma_1 N + \gamma_0$ is the electronic-noise power of the detector normalized by the photonic shot-noise power. We introduce d_0 as the semantic distance between two materials with known spectral emissivity defined using the single-photon Fisher information matrix (see Methods).

The insight from equation (2) is that the shot-noise limit arising from the discrete nature of photons sets the information-theoretic upper bound to the performance of all identification algorithms. Here we reach the bound with machine-learning-based approaches widely deployed for perception. We generate several spectra for human and robot with Monte Carlo simulation in the shot-noise limit and use machine learning for material classification. Figure 3b shows that machine-learning performance (red circles) is indeed bounded by the theoretical limit (red curve). Our theory also applies to realistic detectors with common noise sources (flicker noise: dashed cyan line; Johnson–Nyquist noise: dash-dotted cyan line; mixed noise: solid cyan line; modelled by FLIR A325sc) and corresponding algorithmic performance (see Supplementary Fig. 7). Figure 3c shows the minimum photon number required to identify the human-shaped target, which is determined by unit statistical distance ($\sqrt{N}d_0=1$, $I \approx 0.75$ bits). The minimum photon number for given semantic distance or vice versa, the minimum semantic distance for given photon number, sets fundamental limits to object identification beyond training volume, providing a theoretical foundation for designing public policies.

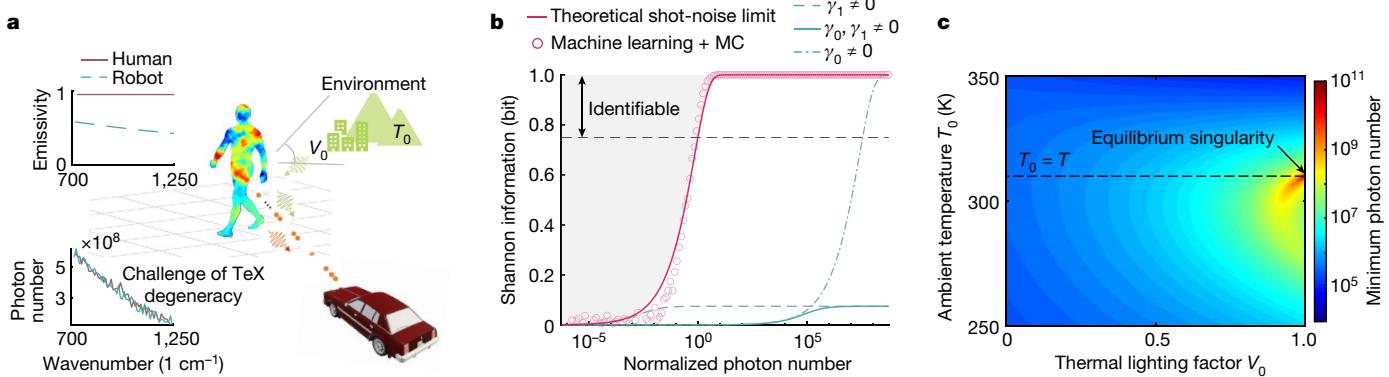


Fig. 3 | Shot-noise limit of HADAR identifiability. **a**, TeX degeneracy limits HADAR identifiability, as in the illustrative human–robot identification problem. Top inset, distinct emissivity of human (grey body) and robot (aluminium). Bottom inset, near-identical incident spectra for human (37 °C, red) and robot (72.5 °C, blue). **b**, HADAR identifiability (Shannon information) as a function of normalized photon number Nd_0^2 . We compare the theoretical shot-noise limit of HADAR (solid red line) and machine-learning performance (red circles) on synthetic spectra generated by Monte Carlo (MC) simulations.

We also consider realistic detectors with Johnson–Nyquist noise ($\gamma_0 = 3.34 \times 10^5$), flicker noise ($\gamma_1 N = 3.34 \times 10^5$) or mixed noise ($\gamma_1 N + \gamma_0 = 3.34 \times 10^5$). Identifiability criterion (dashed grey line) is $\sqrt{N}d_0=1$. **c**, The minimum photon number $1/d_0^2$ required to identify a target is usually large because of the TeX degeneracy, dependent on the scene as well as the thermal lighting factor, as shown for the scene in **a**. Particularly, it diverges at singularity $V_0=1$ and $T_0=T$ when the target is in thermal equilibrium with the environment.

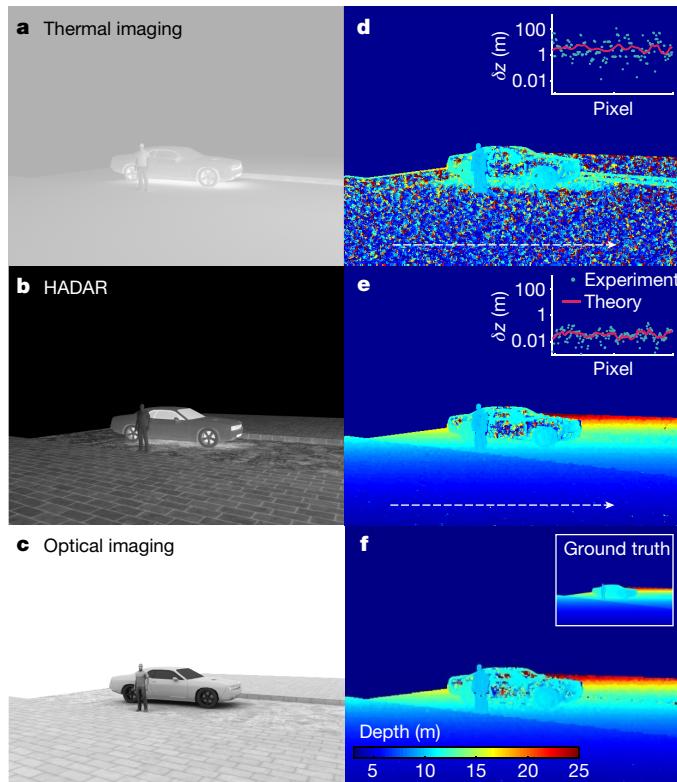


Fig. 4 | Fundamental limit of HADAR ranging. **a,d**, Ranging on the basis of raw thermal images shows poor accuracy owing to ghosting. **b,e**, Recovered textures and enhanced ranging accuracy (approximately 100×) in HADAR as compared with thermal ranging. **c,f**, We also show the optical imaging (**c**) and RGB stereovision (**f**) for comparison. Insets in **d** and **e** show the depth error δz in Monte Carlo experiments (cyan points) in comparison with our theoretical bound (red curve), along the dashed white lines.

HADAR depth resolution

Depth of objects is a critical scene attribute for autonomous navigation. Daylight RGB stereovision already has widespread applications⁸, but infrared thermal ranging is elusive. We demonstrate that HADAR ranging at night beats thermal ranging, with depth accuracy comparable with RGB stereovision in daylight. Our approach of HADAR ranging exploits stereovision based on the TeX vision, but to show the importance of texture in ranging and better capture the physics, here we focus on the scattered signal that can be reconstructed through TeX decomposition. Real-world HADAR ranging will be discussed later. For a concise car/pedestrian scene, thermal imaging loses textures owing to TeX degeneracy (Fig. 4a) and leads to inaccurate ranging (Fig. 4d). HADAR (Fig. 4b) recovers texture comparable with greyscale optical imaging (Fig. 4c). We note that the HADAR ranging result (Fig. 4e) is comparable with RGB stereovision (Fig. 4f). Quantitatively, the absolute ranging error (cyan data points in insets) with respect to the ground truth along the dashed white lines shows approximately 100× accuracy improvement in HADAR versus thermal ranging (the improvement is dependent on the scene).

We derive the fundamental limit on HADAR ranging, providing a rigorous theoretical foundation for future autonomous navigation applications. HADAR ranging error δz of a window (block or feature area) equals the disparity error between corresponding window positions in stereo matching²⁹, up to a dimensionless coefficient. Its fundamental limit is given by

$$\sqrt{N}\delta z \geq \sqrt{2(1+\gamma)(\sigma_d^2 + \sigma_c^2)}, \quad (3)$$

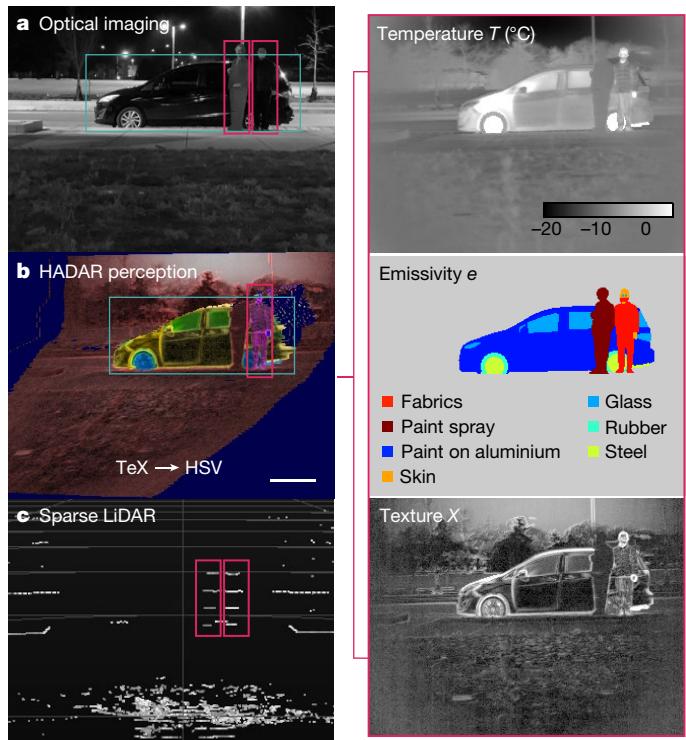


Fig. 5 | Physics-driven HADAR perception in Indiana, USA. For an outdoor scene of a human body, an Einstein cardboard cutout and a black car at night, vision-driven object detection yields two human bodies (error) and one car from optical imaging (**a**) and two human bodies and no car (error) from LiDAR point cloud (**c**). HADAR perception based on TeX physical attributes has comprehensive understanding of the scene and accurate semantics (**b**; one human body and one car) for unmanned decisions. Scale bar, 1 m.

in which σ_d is the diffraction-induced uncertainty in estimating a point-source position from the incident photon distribution. Here σ_c is the photonic correspondence uncertainty in locating the same point source between stereo images in an extended scene with N observed photons. The physical significance of the photonic correspondence uncertainty is the indistinguishability of photons from different point sources. It is given by the Cramér–Rao bound of window-position estimation in the ideal image of the scene (see section SII.C of the Supplementary Information). Theoretical bounds of δz with computed σ_c along the dashed white lines (red curves in insets of Fig. 4d,e) are consistent with numeric experiments (cyan data points), showing a two-orders-of-magnitude accuracy improvement in HADAR.

Real-world HADAR perception

We now experimentally demonstrate HADAR in real-world scenes. Our HADAR prototype-1 for low-end applications is based on a commercial FLIR thermal camera with custom-designed spectral modules (see Extended Data Fig. 10). We propose a shift of physics-driven perception with HADAR TeX attributes, as opposed to existing vision-driven perception¹⁸. We use an outdoor scene at night with a car, a human being and a cutout of Einstein to mimic a human geometrically and illustrate how HADAR addresses phantom braking. Figure 5 shows that both RGB optical imaging (Fig. 5a) and sparse LiDAR point cloud (Fig. 5c; Velodyne Puck VLP-16) cannot distinguish the human body from the real-scale Einstein cardboard cutout. Furthermore, LiDAR has difficulties in detecting the black car because of low reflection, whereas the optical camera cannot see objects in the dark. HADAR detects people in the corresponding material region (skin + fabrics) and clearly distinguishes it from the cardboard, overcoming the phantom-braking

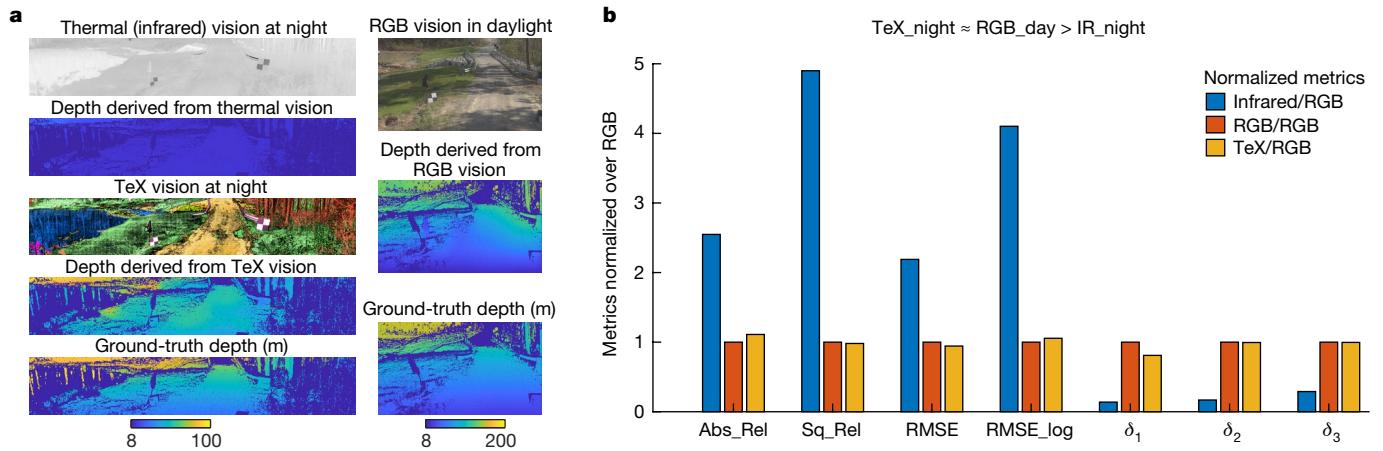


Fig. 6 | HADAR ranging (TeX vision + AI) at night beats state-of-the-art thermal ranging (thermal vision + AI) at night and matches RGB stereovision in daylight, abbreviated as ‘TeX_night ≈ RGB_day > IR_night’. **a**, It can be clearly seen that thermal imaging is impeded by the ghosting effect, whereas HADAR TeX vision overcomes the ghosting effect, providing a fundamental route to extracting thermal textures. This texture is crucial for AI algorithms to function optimally. To prove the HADAR ranging advantage, we used GCNDepth (pre-trained on the KITTI dataset)³⁶ for monocular stereovision, as the state-of-

the-art AI algorithm. Ground-truth depth is obtained through a high-resolution LiDAR. Depth metrics are listed in Table 1. We normalized the depth metrics over that of RGB stereovision. **b**, The comparison of normalized metrics clearly demonstrates that ‘TeX_night ≈ RGB_day > IR_night’, that is, HADAR, sees texture and depth through the darkness as if it were day. See Methods for definitions of the depth metrics used. See sections SIII.A and SV of the Supplementary Information for more details.

problem. See Extended Data Figs. 7 and 8 for more details about HADAR detection and semantics. Key advantages of HADAR perception using physical context will be found in autonomous navigation and wildlife monitoring, in which several physical attributes beyond visual appearance are desired for either safety guarantees³⁰ or scientific purposes³¹.

Our HADAR prototype-2 for high-end applications is based on a push-broom hyperspectral imager (see Methods). We use an off-road scene to demonstrate that TeX vision sees textures through the darkness with physical context and that HADAR ranging at night beats thermal ranging, with accuracy comparable with RGB stereovision in daylight. Real-world TeX vision with material identification and texture recovery has been shown in Fig. 1c and can be found in Extended Data Figs. 3 and 4 with more details. Figure 6 shows the stereovision metric statistics based on TeX vision at night, infrared (IR) thermal vision at night and RGB vision in daylight. The comparison of metrics (normalized by RGB depth metrics) in Fig. 6b clearly demonstrates that HADAR ranging at night beats thermal ranging and matches RGB stereovision in daylight, abbreviated as ‘TeX_night ≈ RGB_day > IR_night’. See Supplementary Fig. 19 for general HADAR ranging performance over various scenes.

HADAR thermography

The COVID-19 pandemic has brought about the urgent need of remote thermography for fever surveillance. Unmanned and high-speed infrared surveillance can substantially relieve the risk to healthcare workers and help limit spread of the virus. However, large-scale temperature screening with existing noncontact infrared thermometers or infrared

thermography is ineffective owing to lack of adaptivity to emissivity (complexion/skin variability), age, gender, circadian variations and distance of the target^{32,33}. As illustrated above, HADAR with TeX vision can identify spectral emissivity, estimate distance and recover textures, promising advanced adaptivity for more accurate temperature estimation. Here we have also experimentally demonstrated that HADAR thermography can automatically recognize emissivity and reach the Cramér–Rao bound on temperature accuracy (see Extended Data Fig. 9). This goal has been elusive owing to TeX degeneracy, which limits temperature accuracy in real-world environments. Unmanned HADAR thermography reaching the Cramér–Rao bound is therefore promising for the smart healthcare industry, including early reliable skin cancer detection³⁴.

Outlook

We proposed and demonstrated HADAR for fully passive and physics-aware machine perception. Our shot-noise limits of detection and ranging set the benchmark and call for heat exploitation in the quantum regime, in which single-photon detectors are being developed beyond the visible spectral range into the thermal infrared³⁵. Practical challenges exist, such as real-time data acquisition, spatio-spectral motion blur and functionality–cost optimization. Nevertheless, we believe that HADAR will lead to a new chapter in the Fourth Industrial Revolution with applications in autonomous navigation, healthcare, agriculture, wildlife monitoring, geosciences and the defence industry.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-023-06174-6>.

1. Rogers, C. et al. A universal 3D imaging sensor on a silicon photonics platform. *Nature* **590**, 256–261 (2021).
2. Floreano, D. & Wood, R. J. Science, technology and the future of small autonomous drones. *Nature* **521**, 460–466 (2015).
3. Jiang, Y., Karpf, S. & Jalali, B. Time-stretch LiDAR as a spectrally scanned time-of-flight ranging camera. *Nat. Photonics* **14**, 14–18 (2020).

Table 1 | Depth metrics statistics associated with Fig. 6 showing HADAR ranging advantage

Real-world performance (depth)	Error		Accuracy (%)				
	Abs_Rel	Sq_Rel	RMSE	RMSE_log	δ_1	δ_2	δ_3
Thermal vision+AI	0.61	13.44	22.96	1.24	7.72	14.82	28.18
RGB vision+AI	0.24	2.74	10.49	0.30	55.88	87.96	97.12
TeX vision+AI	0.27	2.69	9.90	0.32	45.25	87.52	96.77

4. Maccone, L. & Ren, C. Quantum radar. *Phys. Rev. Lett.* **124**, 200503 (2020).
5. Tachella, J. et al. Real-time 3D reconstruction from single-photon lidar data using plug-and-play point cloud denoisers. *Nat. Commun.* **10**, 4984 (2019).
6. Lien, J. et al. Soli: ubiquitous gesture sensing with millimeter wave radar. *ACM Trans. Graph.* **35**, 142 (2016).
7. Kirmani, A. et al. First-photon imaging. *Science* **343**, 58–61 (2014).
8. Geiger, A., Lenz, P. & Urtasun, R. In *Proc. 2012 IEEE Conference on Computer Vision and Pattern Recognition* 3354–3361 (IEEE, 2012).
9. Nassi, B. et al. In *Proc. 2020 ACM SIGSAC Conference on Computer and Communications Security* 293–308 (Association for Computing Machinery, 2020).
10. Popko, G. B., Gaylord, T. K. & Valenta, C. R. Interference measurements between single-beam, mechanical scanning, time-of-flight lidars. *Opt. Eng.* **59**, 053106 (2020).
11. Hecht, J. Lidar for self-driving cars. *Opt. Photon. News* **29**, 26–33 (2018).
12. Gurton, K. P., Yuffa, A. J. & Videen, G. W. Enhanced facial recognition for thermal imagery using polarimetric imaging. *Opt. Lett.* **39**, 3857–3859 (2014).
13. Treble, W. et al. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2961–2969 (IEEE, 2017).
14. Schwab, K. The fourth industrial revolution: what it means, how to respond. *Foreign Affairs* <https://www.foreignaffairs.com/world/fourth-industrial-revolution> (2015).
15. Risteska Stojkoska, B. L. & Trivodaliev, K. V. A review of Internet of Things for smart home: challenges and solutions. *J. Cleaner Prod.* **140**, 1454–1464 (2017).
16. By 2030, one in 10 vehicles will be self-driving globally. *Statista* https://mailchi.mp/statista/autonomous_cars_20200206?e=145345a469 (2020).
17. How robots change the world. What automation really means for jobs and productivity. *Oxford Economics* <https://resources.oxfordeconomics.com/how-robots-change-the-world> (2020).
18. Garcia-Garcia, A. et al. A survey on deep learning techniques for image and video semantic segmentation. *Appl. Soft Comput.* **70**, 41–65 (2018).
19. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
20. Jordan, M. I. & Mitchell, T. M. Machine learning: trends, perspectives, and prospects. *Science* **349**, 255–260 (2015).
21. Gade, R. & Moeslund, T. B. Thermal cameras and applications: a survey. *Mach. Vis. Appl.* **25**, 245–262 (2014).
22. Tang, K. et al. Millikelvin-resolved ambient thermography. *Sci. Adv.* **6**, eabd8688 (2020).
23. Henini, M. & Razeghi, M. *Handbook of Infrared Detection Technologies* (Elsevier, 2002).
24. Haque, A., Milstein, A. & Li, F.-F. Illuminating the dark spaces of healthcare with ambient intelligence. *Nature* **585**, 193–202 (2020).
25. Beier, K. & Gemperlein, H. Simulation of infrared detection range at fog conditions for enhanced vision systems in civil aviation. *Aerospace Sci. Technol.* **8**, 63–71 (2004).
26. Newman, E. & Hartline, P. Integration of visual and infrared information in bimodal neurons in the rattlesnake optic tectum. *Science* **213**, 789–791 (1981).
27. Gillespie, A. et al. A temperature and emissivity separation algorithm for Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) images. *IEEE Trans. Geosci. Remote Sens.* **36**, 1113–1126 (1998).
28. Baldridge, A., Hook, S., Grove, C. & Rivera, G. The ASTER spectral library version 2.0. *Remote Sens. Environ.* **113**, 711–715 (2009).
29. Szeliski, R. *Computer Vision: Algorithms and Applications* (Springer, 2011).
30. Khaleghi, B., Khamis, A., Karay, F. O. & Razavi, S. N. Multisensor data fusion: a review of the state-of-the-art. *Inf. Fusion* **14**, 28–44 (2013).
31. Lopes, C. E. R. & Ruiz, L. B. In *Proc. 2008 1st IFIP Wireless Days* 1–5 (IEEE, 2008).
32. Wright, W. F. & Mackowiak, P. A. Why temperature screening for coronavirus disease 2019 with noncontact infrared thermometers does not work. *Open Forum Infect. Dis.* **8**, ofaa603 (2020).
33. Ghassemi, P., Pfefer, T. J., Casamento, J. P., Simpson, R. & Wang, Q. Best practices for standardized performance testing of infrared thermographs intended for fever screening. *PLoS One* **13**, e0203302 (2018).
34. Magalhaes, C., Tavares, J. M. R., Mendes, J. & Vardasca, R. Comparison of machine learning strategies for infrared thermography of skin cancer. *Biomed. Signal Process. Control* **69**, 102872 (2021).
35. Reddy, D. V., Nerem, R. R., Nam, S. W., Mirin, R. P. & Verma, V. B. Superconducting nanowire single-photon detectors with 98% system detection efficiency at 1550 nm. *Optica* **7**, 1649–1653 (2020).
36. Masoumian, A. et al. GCNDepth: self-supervised monocular depth estimation based on graph convolutional network. *Neurocomputing* **517**, 81–92 (2023).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2023

Methods

TeX degeneracy

For an object α , its spectral radiance $S_{\alpha\nu}$ given by equation (1) is invariant if we change its physical attributes $\{T_\alpha, e_{\alpha\nu}, X_{\alpha\nu}\}$ to $\{T'_\alpha, e'_{\alpha\nu}, X'_{\alpha\nu}\}$, in which T'_α is an arbitrary temperature value, $X'_{\alpha\nu}$ is an arbitrary spectral texture curve and the spectral emissivity curve $e'_{\alpha\nu}$ is given by

$$e'_{\alpha\nu} = \frac{e_{\alpha\nu}[B_\nu(T_\alpha) - X_{\alpha\nu}] + [X_{\alpha\nu} - X'_{\alpha\nu}]}{B_\nu(T'_\alpha) - X'_{\alpha\nu}}.$$

Here ν is the wavenumber and B is the blackbody radiation. Physical states having distinct triplets of TeX attributes but having the same observed heat signal S_ν are addressed as TeX degeneracy.

TeX decomposition

We exploited a material library \mathcal{M} and the mathematical structure of X to overcome TeX degeneracy. In this paper, we proposed several approaches to fully solve TeX decomposition, depending on the specific problem. Learning-based TeX-Net (see Extended Data Fig. 1 for the architecture) using both spatial and spectral information in decomposing TeX is our general solution, whereas we have also provided the analytical inverse function, least-squares estimator and TeX-SGD (semi-global decomposition) as non-machine-learning baselines (see section SIII.B of the Supplementary Information). For the HADAR database and HADAR prototype-2 experiments, we used TeX-Net and TeX-SGD. For human–robot identification in Fig. 3, we used a 1D three-layer convolutional neural network (with ReLU activation) followed by a softmax classifier to recognize material category m_α from collected radiation spectrum $S_{\alpha\nu}$. Categorical cross entropy was used as the loss function and the Adam optimizer was used. For Extended Data Fig. 7, we used the analytical inverse function of T and X following a material classifier as mentioned above. In HADAR prototype-1 experiments, we used the least-squares estimator for TeX decomposition. In Fig. 5, the material library was drawn from the NASA JPL ECOSTRESS spectral library²⁸, with emissivity shown in section SIV.D of the Supplementary Information. In HADAR prototype-2 experiments, we used a semantic library, instead of the material library, estimated from the data itself; see section SV.B of the Supplementary Information. With the least-squares estimator, we verified that TeX decomposition is crucial for vision applications and goes beyond the traditional temperature–emissivity separation approach; see Supplementary Fig. 20.

TeX vision and pseudo-Tex vision

Motivated by colouring convention in the existing literature, in which different colours represent different categories, we use HSV format to represent TeX with mapping $H = e, S = T, V = X$. In this TeX vision, different hues of colour represent different materials, saturation indicates temperature and brightness gives textures. The texture recovered in TeX vision is from increased information in sensor data, in contrast to state-of-the-art approaches such as automatic gain control. Automatic gain control is also applied to TeX to obtain better visualization. See section SIII.C of the Supplementary Information for more details about TeX vision and how the TeX vision image is formed.

As TeX vision requires the input of hyperspectral heat cubes, we also propose pseudo-Tex vision to extend its applications to common thermal datasets without spectral resolution. See section SIII.D of the Supplementary Information for details.

Eye safety restricts the scalability of LiDAR

Although laser has been successfully used to measure the lunar distance, a LiDAR equipped on an automated vehicle can only detect objects within a few hundred metres, and this is because of the fact that the LiDAR emitting power is restricted according to eye safety standards. When the number of intelligent agents scales up, eye safety

further requires the emitting power of an individual agent to scale down as the inverse of the number of agents. This scaling law makes active modalities such as LiDAR impractical to use in multi-agent applications.

Monte Carlo path tracing

The HADAR database is an long-wave infrared stereo-hyperspectral database mostly synthesized by exploiting Planck's law and Kirchhoff's law in the Blender Cycles renderer. The database has been made public and is available at <https://github.com/FanglinBao/HADAR>, in which detailed descriptions can be found. In this paper, Lambertian (diffusive) reflectance was used for simplicity. Samples per pixel was 2,000. We also implemented path tracer according to equation (1), per wavenumber, with OpenGL (version 4.6) and Compute Unified Device Architecture (CUDA, version 10.2) on a GPU. For Fig. 4, ray depth was 8 with (thermal imaging) or without (HADAR perception) direct emission at the final step. Rendering wavenumber was 769 cm⁻¹. The ground and the sky were at 20 °C. The emissivity pattern of the ground was generated by mapping a normal road image to emissivity between 0.8 and 1, to maximize texture loss in thermal vision. Synthetic textures for all other scenes were surface-normal textures. The optical image was rendered without direct emission from objects but with sky illumination. The image size was 640 × 480 pixels. For Fig. 2, ray depth was 8 with normal texture on an opaque glass bulb. For Extended Data Fig. 7, ray depth was 1 for 11 discrete wavenumbers within 715–1,250 cm⁻¹.

HADAR estimation theory

Here we provide a short answer to the question ‘How many photons are needed to identify the target material’. Full derivations of fundamental bounds for both detection and ranging are given in section SII of the Supplementary Information. Identifying the target between two candidate materials $e_{1\nu}$ and $e_{2\nu}$ is mapped to estimating the fraction g of a mixture of these two materials, $e_\nu = (1-g)e_{1\nu} + ge_{2\nu}$, with $g=0$ indicating one material and $g=1$ indicating the other. The normalized spectrum $p_{\alpha\nu} = S_{\alpha\nu}/\int S_{\alpha\nu} d\nu$ describes the spectral probability density for one incident photon. The Fisher information matrix for unknown parameters $\Theta = \{g, T, V_0\}$ reads $J_{ij}^0 = NJ_{ij}^0/(1+\gamma)$, in which $J_{ij}^0 = \int \frac{(\partial_i p_{\alpha\nu})(\partial_j p_{\alpha\nu})}{p_{\alpha\nu}} d\nu$ is the single-photon Fisher information matrix, N is the total number of photons, T is the temperature, V_0 is the thermal lighting factor of the environment, $i, j \in \Theta$, and $\gamma = \gamma_1/N + \gamma_0$ is the electronic-noise power normalized by the photonic shot-noise power. The Cramér–Rao bound $\sigma^2 \equiv [1/J]_{gg}$ puts a lower bound on the variance of any unbiased estimator of g and the statistical distance is $d \equiv 1/2\sigma$. Only depending on material, semantic distance $d_0 \equiv 1/2\sigma_0$ with $\sigma_0^2 \equiv [1/J]_{gg}$ describes how different two materials are to each other, under the TeX degeneracy. The related but distinct concept of statistical distance depicts the overall distinguishability of two spectra but depends on the detector and measurement time. Our semantic-distance approach captures the intrinsic identifiability of objects from the scene alone. The detection probability (true positive rate) is given by $P = [1 + \text{erf}(d/\sqrt{2})]/2$. The Shannon information of material is given by $I = \log_2 P - \log_2(1/2)$.

In evaluating the theoretical bound on HADAR identifiability in Fig. 3, we used $V_0 = 0.5$ (α suppressed) and $T_0 = 20$ °C. The target distance was 30 m when the input signal was evaluated. In the Monte Carlo simulations in Fig. 3b, we first found the nearest robot condition ($T = 83.46$ °C, $V_0 = 0.17$) to human ($T = 37$ °C, $V_0 = 0.5$). We used 150 samples of the normalized photon number, and then for each normalized photon number we generated 5,000 spectra (715–1,250 cm⁻¹, $\Delta\nu = 1$ cm⁻¹) for each of the two candidates with Monte Carlo simulation in the shot-noise limit. Finally, we used machine learning (25% spectra for training, 25% for validation and 50% for testing) for material classification, and the test accuracy was used to compute the Shannon information for each normalized photon number. The dimensionality curse for high spectral resolution (536 bands used) leads machine learning to overfitting, and slight deviation between Monte Carlo simulation and theoretical prediction can be observed in Fig. 3b. Once the dimensionality curse

Article

is relieved, perfect agreement can be reached; see Supplementary Fig. 7, in which all spectra are downsampled into three spectral bands (dimension = 3) for both theory and machine learning.

In evaluating the theoretical bound on ranging error in Fig. 4 without photon number, we used $J_x = (\partial_x N_{iq})^2 / (N_{iq} + \sigma^2)$ (equation (S37) in the Supplementary Information) instead. Variance was estimated by matching corresponding pixels according to the ground-truth disparity and computing the signal fluctuation. Finite difference was used to approximate the derivative. $b = 0.2$ m and $f = 1.4$ cm. The block size in correlation-based sub-pixel block matching was 5×5 (see Supplementary Fig. 10 for AI results).

Guiding public policy

The HADAR identifiable criterion is $\frac{Nd_0^2}{1+y} = 1$, which means that we can identify the target material if $\frac{Nd_0^2}{1+y} \geq 1$. The semantic distance between human body (skin) and robot (aluminium) in Fig. 3 is calculated to be $d_0 \approx 0.001$. This requires $N/(1+y) \geq 10^6$ to identify the target if the environment is at $T_0 = 20$ °C and $V_0 = 0.5$ (see Fig. 3c). The observed photon number N is related to the human–robot scene, as well as the f -number (focal length f over the aperture size D), exposure time t and pixel size A_p ; see the heat-signal model in section SI of the Supplementary Information. Eventually, the above identifiable criterion leads to the minimum requirement of the hardware configurations, $\frac{tA_p}{(1+y)(f/D)^2} \geq 5 \times 10^{-16}$. This minimum requirement of the hardware will guide public policies in the AI industry. For example, the lowest detectivity (or highest noise equivalent power), the smallest aperture size, the highest frame rate and hence the maximum travelling speed etc., must meet the above inequality so as to be able to identify human versus robot. If the detector does not meet the above requirement, its collected data will be insufficient in information. No matter how much data are collected and used to train a neural network (how large the training volume is), machine learning cannot perform well (see the machine-learning performance in Fig. 3b when the photon number is insufficient, that is, the normalized photon number is below 1). If the detector is given, for example, the FLIR A325sc camera, we have $\frac{tA_p}{(1+y)(f/D)^2} = 8.16 \times 10^{-18}$ in one image frame. To meet the criterion, we must have $d_0 > 0.0078$, which means that the FLIR A325sc camera can only distinguish sufficiently different material pairs in one image frame, such as organic skin versus glass mannequin ($d_0 = 0.049$). This minimum semantic distance identifiable by the given detector will also guide public policies in the AI industry, for example, in which scenario the given camera can be used and in which scenario the camera cannot be used. Likewise, our fundamental limit of HADAR ranging accuracy can also put requirements on hardware configurations or restrict travelling speed etc. and guide public policies.

Our fundamental limits bound the average machine-learning performance owing to the shot noise and detector noise. Lucky evaluation events could occur, but they will fluctuate around the average bounds, as can be seen in Fig. 3b and insets of Fig. 4. Human error or software bugs are not considered in our bounds, but our bounds are useful because they depict the optimal performance of machine learning when human error and software bugs are completely corrected. Therefore, our bounds related to physical laws of thermal photonic information theory can be used as a guidance to public policies.

Thermal camera specifications

Our FLIR A325sc thermal camera is a science-grade high-performance radiometric camera (price around \$10,000). It is equipped with an uncooled vanadium oxide (VO_x) microbolometer detector that produces thermal images of 320×240 pixels. The detector pitch is $25\text{ }\mu\text{m}$, pixel size is approximated as $12\text{ }\mu\text{m}$, time constant is 12 ms , focal length is 18 mm and the f -number is 1.3 . The noise-equivalent temperature difference (NETD) is typically $<50\text{ mK}$ and characterized to be 47.8 mK . The FLIR A325sc camera was available when the experiments in this paper were designed and conducted. We note that it has now been

discontinued and replaced by a more advanced model, FLIR A655sc. The latter has a 640×480 -pixel array with typical NETD $< 30\text{ mK}$, but it is twice as expensive. A better camera will give better HADAR data. Because the advantage of HADAR over traditional thermal vision comes from the spectral resolution and the theory we used to interpret the hyperspectral data, FLIR A325sc presents a better functionality–cost balance.

Fourier transform infrared specifications

Our Nicolet iS50 FTIR Spectrometer is equipped with a cooled (liquid nitrogen) MCT-A detector, with the sensor element size being 1 mm . Its special detectivity is $4.7 \times 10^{10}\text{ cm Hz}^{1/2}\text{ W}^{-1}$, with the preamplifier bandwidth being 175 kHz . Spectral resolution $\Delta\nu = 0.48\text{ cm}^{-1}$ within $769\text{--}1,332\text{ cm}^{-1}$ is used in this paper. The aperture of external optics is 2 inches and the focal length of external optics is about 10 cm . Optical efficiency is approximated as 0.9 in deriving the Cramér–Rao bound on temperature accuracy in Extended Data Fig. 9.

Prototype HADAR calibration and data collection

Our HADAR prototype-1 was built with FLIR A325sc as the detector, plus ten thermal infrared filters (price about \$10,000) from Spectrogon to retrieve spectral resolution. A gold mirror was also mounted on the filter wheel to monitor the status of the detector in real time. Once started, the detector was left to stabilize for at least 30 min to warm up. In experiments in which the detector exchanged heat with the scene, the mirror signal was checked so that data collected with very different detector status was ignored. The filter transmittance curves were characterized by the Nicolet iS50 FTIR Spectrometer. The spectral response curve of the camera was calibrated with a standard blackbody source (EOI Inc. DCN1000N7). In experiments, a uniform reference object was used to further calibrate the self-radiation pattern of the camera, as well as the side effect of the filter wheel acting as an out-of-focus diaphragm. The experimental data collected were left and right heat cubes of dimensions height \times width \times channel = $240 \times 320 \times 10$. The number of channels was the number of filters. See Extended Data Fig. 10 for HADAR prototype-1 calibration and data collection.

The HADAR prototype-2 is based on a pushbroom hyperspectral imager with a cooled HgCdTe sensor. To collect the real-world experimental data, we formed a partnership with the Defense Advanced Research Projects Agency (DARPA), through the Invisible Headlights project, and the Army night-vision team (Infrared Camera Technology Branch, DEVCOM C5ISR Center, U.S. Army). The pushbroom hyperspectral imager gives 256 spectral bands, but its price is over a million dollars. The focal length is 50 mm and the f -number is $f/0.9$. It uses a 256×256 focal plane array with $40\text{-}\mu\text{m}$ -pitch pixels. The typical noise of the sensor is around 1 ‘microflick’, which—at $10\text{ }\mu\text{m}$ wavelength—corresponds to about a $1,000:1$ signal-to-noise ratio. Explicitly, for a 300K -temperature scene at $10\text{ }\mu\text{m}$, the NETD is around 63 mK . Denoising and extrinsic calibrations can be found in section SV of the Supplementary Information.

In our proof-of-concept experiments, we used the filter-wheel approach to demonstrate the HADAR prototype-1. The filter-wheel approach is time consuming but cost effective and suitable for low-end HADAR applications. By contrast, HADAR prototype-2 with a pushbroom sensor was demonstrated for high-end HADAR applications. HADAR can also be implemented by other approaches with mosaic sensors, gratings, prisms, interferometers or Fabry–Pérot cavities, depending on the desired spectral resolution, spatial resolution, data-acquisition speed or functionality–cost balance.

Computational efficiency and deployability

(1) TeX-Net has about half a million weights in total. The evaluation of TeX-Net (GPU NVIDIA RTX A6000 48 GB) takes 42.4 ms . Data collection of the pushbroom hyperspectral imager used at present takes around 1 s , but the filter-wheel approach can be optimized down to around

10 ms with high-speed filter wheel (for example, Telops multi-spectral cameras). Overall, our results show that HADAR data collection and processing can support up to 20 Hz TeX vision frame rate. Pursuing higher frame rates motivates further research on new hyperspectral imaging sensors to collect thermal infrared data and photonic neural networks for TeX decomposition. (2) Our generalized HADAR theory does not require the input of a material library and, hence, is free of on-site library collection/calibration. This enables real-time HADAR applications. Our HADAR prototype-2 experiment is a field test with the HADAR sensor mounted on a car. Corresponding TeX vision results on the DARPA Invisible Headlights test data show the deployability of HADAR; see Extended Data Figs. 3 and 4.

Standard depth metrics

Let pred and gt denote the predicted and ground-truth depth, respectively. D represents the set of all predicted depth values, $|\cdot|$ returns the number of elements and $\|\cdot\|$ returns the absolute value. The standard depth metrics used in Fig. 6 are defined as below:

Absolute and relative error,

$$\text{Abs_Rel} = 1/|D| \cdot \sum_{\text{pred} \in D} \|\text{pred} - \text{gt}\|/\text{gt},$$

Squared relative error,

$$\text{Sq_Rel} = 1/|D| \cdot \sum_{\text{pred} \in D} \|\text{pred} - \text{gt}\|^2/\text{gt},$$

Root mean square error,

$$\text{RMSE} = \sqrt{1/|D| \cdot \sum_{\text{pred} \in D} \|\text{pred} - \text{gt}\|^2},$$

Root mean square log error,

$$\text{RMSE_log} = \sqrt{1/|D| \cdot \sum_{\text{pred} \in D} \|\log(\text{pred}) - \log(\text{gt})\|^2},$$

δ_t accuracy,

$$\delta_t = \frac{1}{|D|} \left| \left\{ \text{pred} \in D \mid \max \left(\frac{\text{gt}}{\text{pred}}, \frac{\text{pred}}{\text{gt}} \right) < 1.25^t \right\} \right|.$$

HADAR thermography

Note that existing thermal imaging measures the total radiance and approximates emissivity e , as a default parameter or manually input constant e . This causes the temperature readout to be biased and incorrect. Furthermore, when two different materials at different temperatures happen to emit the same total radiance, thermal imaging predicts the same temperature, leading to the thermal camouflage effect^{37,38} (the integral version of TeX degeneracy; Extended Data Fig. 9b). We use two strips of tapes on plastics (not shown) and three patches of tapes on silicon (three rectangles in Extended Data Fig. 9a–d) to read out the temperature without influence from non-trivial emissivity of sample materials. Non-uniform heating effect and self-radiation of the thermal camera (FLIR A325sc) are calibrated and removed from data in Extended Data Fig. 9b before TeX decomposition. In HADAR TeX decomposition, two materials are identified automatically and the temperature is estimated accordingly, revealing an otherwise hidden HADAR alphabet pattern.

In statistical analysis (Extended Data Fig. 9e,f, based on the Nicolet iS50 FTIR Spectrometer), the response curve and environmental radiation are calibrated, as well as the dark noise caused by self radiation of the device, before experiments. See section SVI of the Supplementary Information for calibration details. After calibration, 20 measurements are taken for each of 16 heating powers. Conventional thermal imaging with default emissivity ($e = 0.95$) severely deviates from the ground truth obtained by the thermocouple. Manually input emissivity is calibrated

at the lowest heating power, but this approach also deviates as the heating power increases, as the calibration is inaccurate owing to, once again, thermal camouflage. These issues are overcome with HADAR thermography, which estimates temperature unbiasedly. We note that the root mean square error beats both infrared as well as contact thermography. This is not surprising as even contact thermography using thermocouples have inevitable errors arising from noisy physical contacts.

The temperature difference between tumour cells and normal cells in skin cancer could be as high as 0.25 °C. However, the signal captured by a thermal camera is the radiance S that includes scattering contributions from the environment (X), along with direct emission from the tumour cells. Having a hot object (other people, instruments) in the patient room (that is, considering X or not) makes a marked difference in estimated temperatures. As an example, the emissivity of skin can be well approximated as a constant of 0.95, and we assume that the environment is a blackbody ($X = B$) to approximately see the errors arising from ignoring the environment. The presence/absence of X is equivalent to a 5% relative difference of $B(T)$, which corresponds to 3 °C temperature variation around the standard 37 °C temperature. This error arising from ignoring the environmental signal is much larger than the temperature difference caused by tumour cells. To minimize this effect, accurate thermography is limited to “either in an open-area, outdoor environment under clear sky (cloud free), or using a cold-plate setup”, which restricts the indoor applications for fever surveillance. Because TeX vision decomposes S , HADAR can reach the Cramér–Rao bound of temperature by properly estimating e and X and, hence, is promising for reliable skin cancer detection.

Data availability

The data supporting the findings of this study are available in the paper. The HADAR database is available at <https://github.com/FanglinBao/HADAR>.

Code availability

The custom-designed codes are provided along with the HADAR database at <https://github.com/FanglinBao/HADAR>.

- 37. Qu, Y. et al. Thermal camouflage based on the phase-changing material GST. *Light Sci. Appl.* **7**, 26 (2018).
- 38. Li, M., Liu, D., Cheng, H., Peng, L. & Zu, M. Manipulating metals for adaptive thermal camouflage. *Sci. Adv.* **6**, eaaba3494 (2020).
- 39. Li, H., Xiong, P., An, J. & Wang, L. Pyramid Attention Network for semantic segmentation. Preprint at <https://arxiv.org/abs/1805.10180> (2018).
- 40. Fu, J. et al. in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 3146–3154* (IEEE, 2019).

Acknowledgements This work was supported by the Invisible Headlights project from the Defense Advanced Research Projects Agency (DARPA). We thank the Army night-vision team (Infrared Camera Technology Branch, DEVCOM C5ISR Center, U.S. Army) for the help in collecting HADAR prototype-2 experimental data. We thank Z. Yang for her help in experiments.

Author contributions F.B. and Z.J. conceived the idea. F.B. led and Z.J. supervised the project. F.B. developed and L.Y. contributed to the HADAR estimation theory. F.B. generated the HADAR database and designed experiments. X.W. built HADAR prototype-1 and conducted experiments. F.B. analysed experimental data. S.H.S., G.S., F.B., V.A. and V.N.B. contributed to machine learning. F.B. prepared and all authors revised the manuscript.

Competing interests The authors declare no competing interests.

Additional information

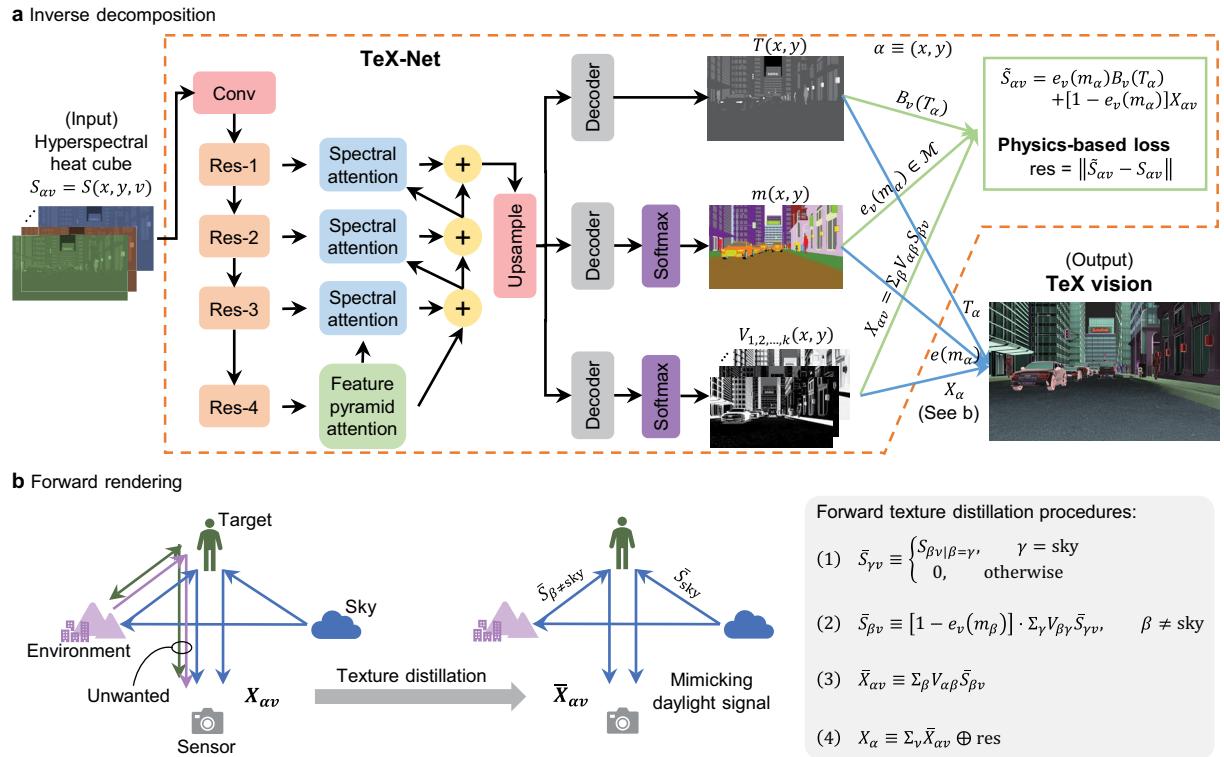
Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-023-06174-6>.

Correspondence and requests for materials should be addressed to Zubin Jacob.

Peer review information *Nature* thanks Manish Bhattarai, Moongu Jeon and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

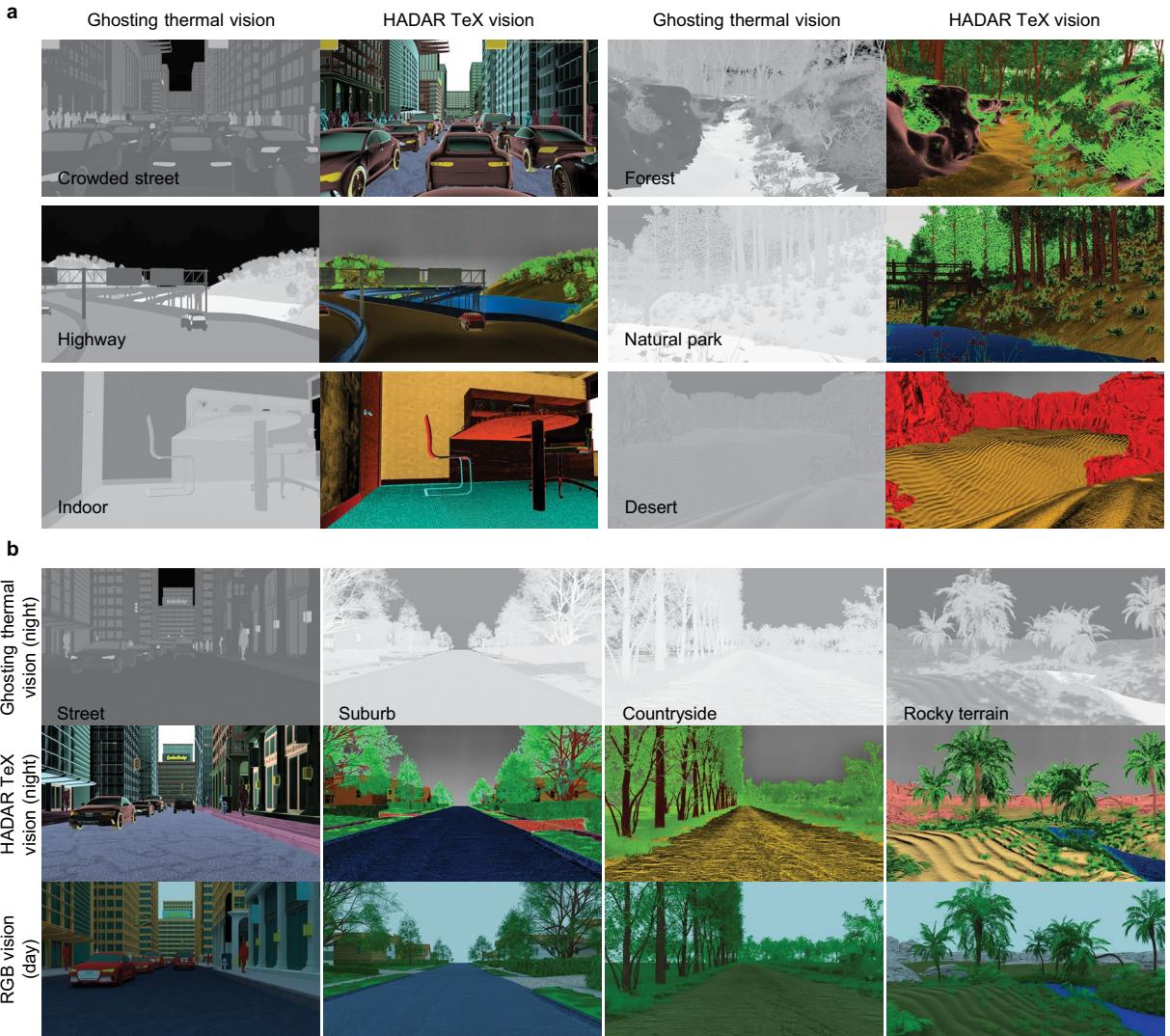
Reprints and permissions information is available at <http://www.nature.com/reprints>.

Article



Extended Data Fig. 1 | HADAR TeX vision algorithms. **a**, Architecture of TeX-Net for inverse TeX decomposition. TeX-Net is physics-inspired for three aspects. First, TeX decomposition of heat cubes relies on both spatial patterns and spectral thermal signatures. This inspires the adoption of spectral and pyramid (spatial) attention layers³⁹ in the UNet model. Second, owing to TeX degeneracy, the mathematical structure, $X_{\alpha v} = \sum_{\beta \neq \alpha} V_{\alpha\beta} S_{\beta v}$, has to be specified to ensure the uniqueness of inverse mapping and hence it is essential to learn thermal lighting factors V instead of texture X . That is, TeX-Net cannot be trained end to end. Here α, β and γ are indices of objects and v is the wavenumber. X_α is constructed with V and S_β , indirectly, in which S_β is the downsampled $S_{\alpha v}$ to approximate k most notable environmental objects. Third, the material library \mathcal{M} and its dimension are key to the network. TeX-Net can be trained with either ground-truth T , m and V in supervised learning or, alternatively, with material library \mathcal{M} , Planck's law $B_v(T_\alpha)$ and the mathematical structure of $X_{\alpha v}$ in unsupervised learning. In supervised learning, the loss function is a combination of individual losses with regularization hyperparameters. In unsupervised learning, the loss function defined on the reconstructed heat cube is based on physics models of the heat signal. In practice, a hybrid loss function with T , e and V contributions (50%), as well as the physics-based loss (50%), is used. In this work, we have also proposed a non-machine-learning approach, the TeX-SGD, to generate TeX

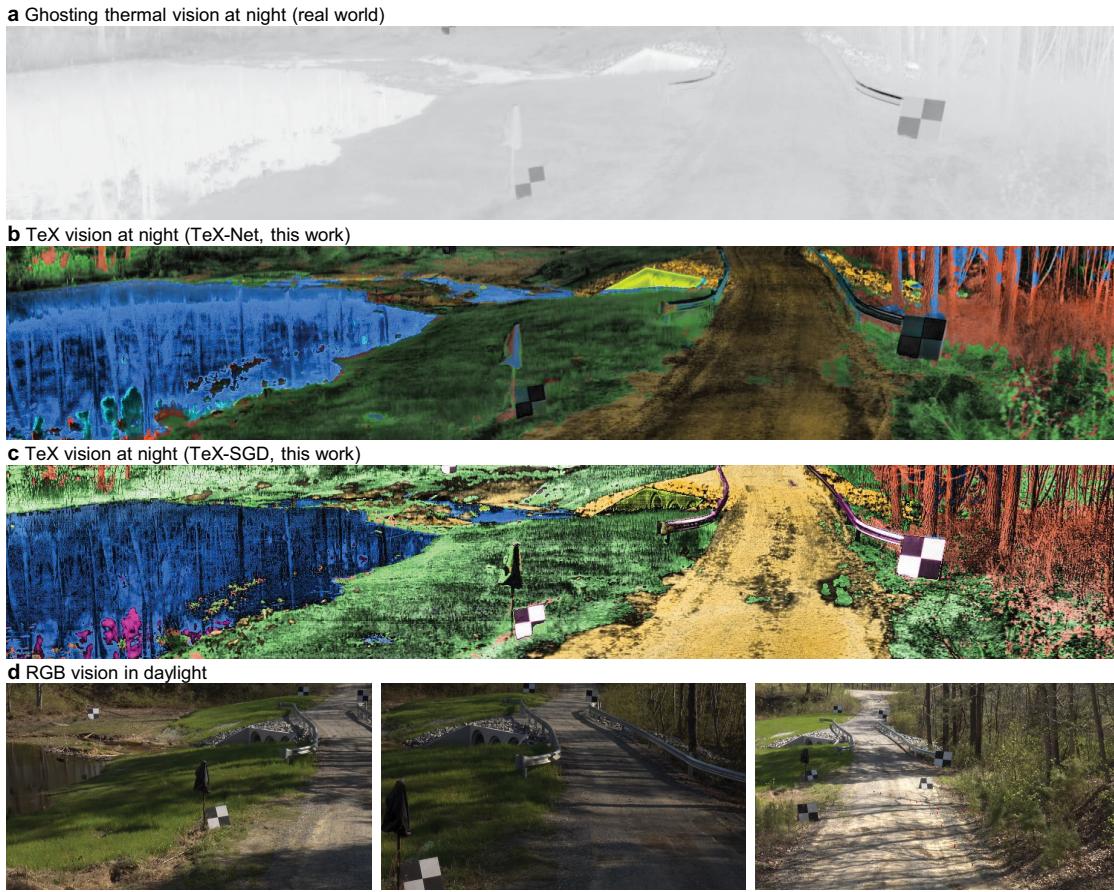
vision. TeX-SGD decomposes TeX pixel per pixel, based on the physics loss and a smoothness constraint; see sections SIII.A and SIII.B of the Supplementary Information for more details. Res-1/2/3/4 are ResNet50 with downsampling. The plus symbol is the addition operation followed by upsampling. **b**, Texture distillation reconstructs the part of scattered signal that originates only from sky illuminations. The texture-distillation process is to mimic daylight signal as X to form TeX vision and it is done by evaluating the HADAR constitutive equation in a forward way, with the physical attributes solved out in TeX-SGD or TeX-Net. It removes the unwanted effect of other environmental objects being the light source, which is unfamiliar in daily experience. The process can be described in four steps. Here step 1 is the initialization that keeps only the sky illumination on and turns other radiations off. Step 2 is the iterative HADAR constitutive equation without direct emission. Evaluating it several times gives the multiple-scattering effect. Note that the ground-truth texture partly remains in the physics-based loss, res , owing to cutoffs on scattering and/or number of environmental objects. The final estimated texture in step 4 is a fusion of distilled scattered signal $\bar{X}_{\alpha v}$ and the physics-based loss res . Arrows in **b** indicate thermal radiation emitted/scattered along the arrow direction. The TeX-Net code, pre-trained weights and a sample implementation of texture distillation is available at <https://github.com/FanglinBao/HADAR>.



Extended Data Fig. 2 | HADAR database and demonstrated TeX vision show that HADAR overcomes the ghosting effect in traditional thermal vision and sees through the darkness as if it were day. TeX vision (colour hue H = material; saturation S = temperature T ; brightness V = texture X) provides intrinsic attributes and enhanced textures of the scene to enable comprehensive understanding. Our HADAR database consists of 11 dissimilar night scenes covering most common road conditions that HADAR may find applications in. Particularly, the indoor scene is designed for robot helpers in smart-home

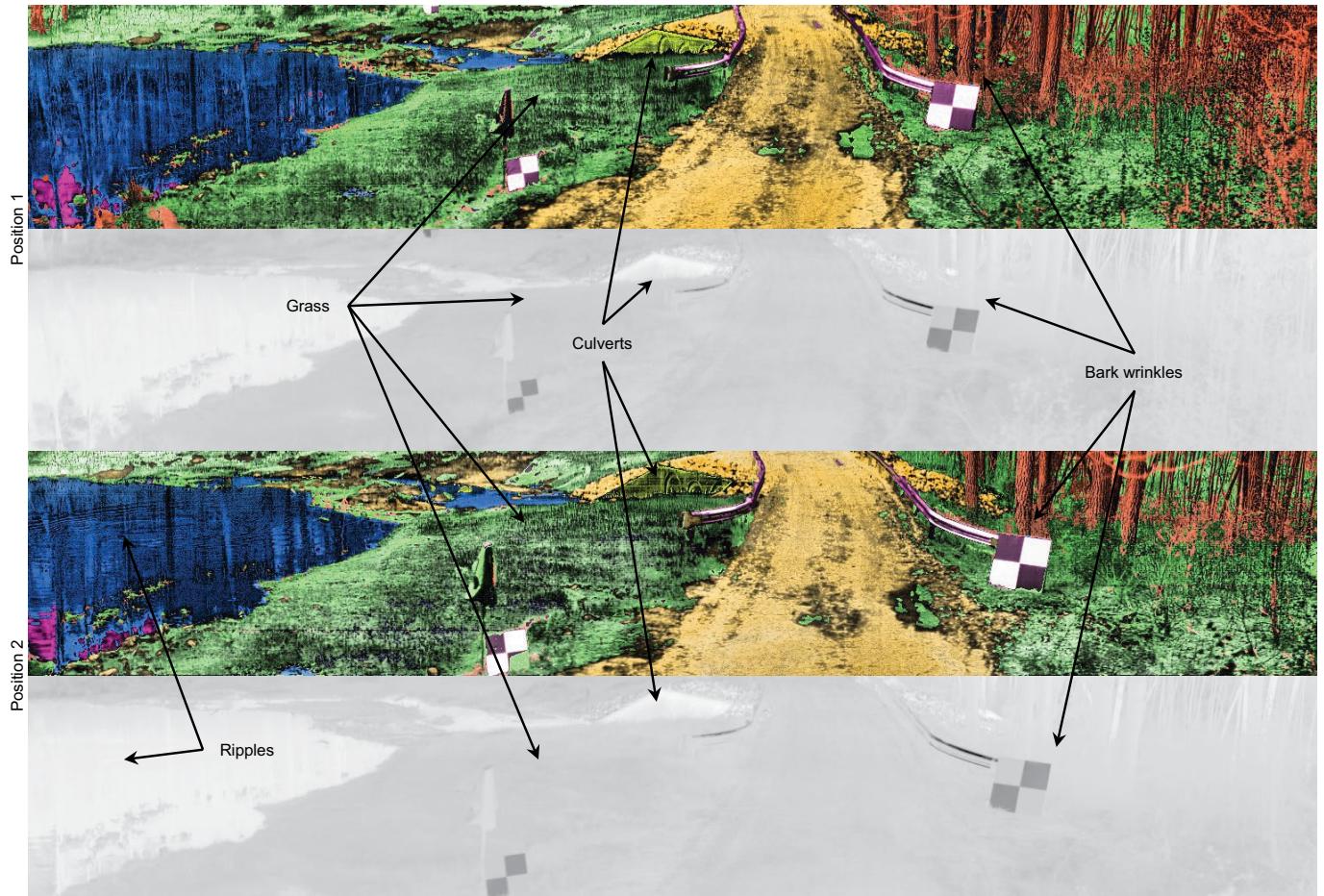
applications, whereas others are for various self-driving applications. Scene-11 is a real-world off-road scene and shall be shown in Extended Data Fig. 3. The database is a long-wave infrared stereo-hyperspectral database with crowded (for example, crowded street) and complicated (for example, forest) scenes, having several frames per scene and 30 different kinds of material. The database is available at <https://github.com/FanglinBao/HADAR>. See Supplementary Fig. 18 for the TeX-Net performance on the HADAR database.

Article



Extended Data Fig. 3 | HADAR TeX vision demonstrated in real-world experiments (Scene-11 of the HADAR database) overcomes the ghosting effect in traditional thermal vision and sees through the darkness as if it were day. Here TeX vision was generated by both TeX-Net and TeX-SGD for comparison. We used a semantic library instead of the exact material library for the TeX vision; see section SV.C of the Supplementary Information for more details. The semantic library consists of tree (brown), vegetation (green), soil (yellow), water (blue), metal (purple) and concrete (chartreuse). Water gives mirror images of trees and part of the sky beyond the view. Most of the water pixels can be correctly estimated as ‘water’, except for a small portion corresponding to sky image that has been estimated as ‘metal’, as metal also

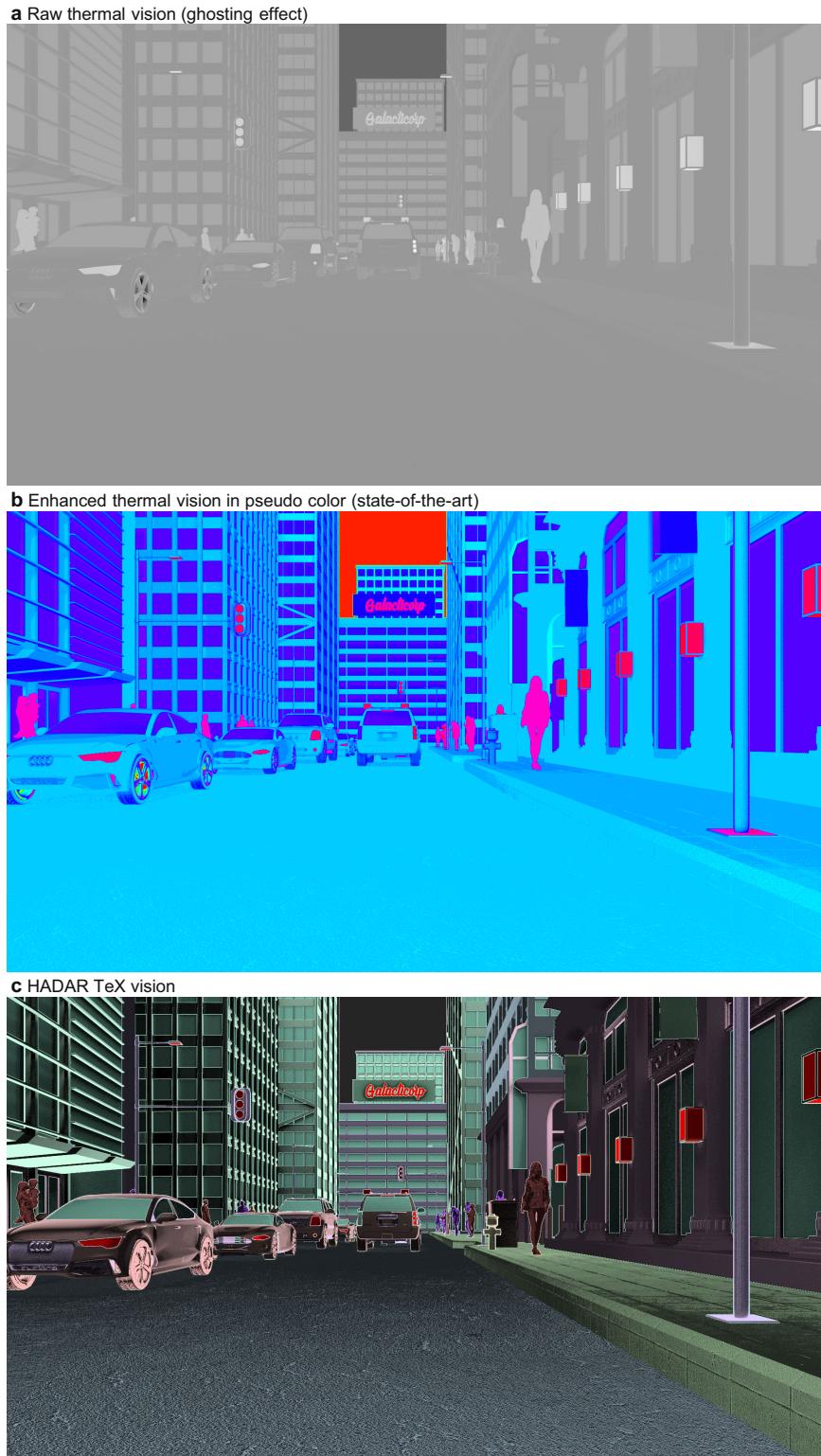
reflects the sky signal. TeX-Net uses both spatial information and spectral information for TeX decomposition and hence its TeX vision is spatially smoother. By contrast, TeX-SGD mainly makes use of spectral information and decomposes TeX pixel per pixel. Compared with TeX-Net, we observed that TeX-SGD is better at material identification and texture recovery for fine structures, such as the fence of the bridge, bark wrinkles and culverts. Note that the current TeX-Net was trained partially with TeX-SGD outputs. The above observations are not used to claim performance ranking between TeX-SGD and TeX-Net. Both TeX-Net and TeX-SGD confirm that HADAR TeX vision has achieved a semantic understanding of the night scene with enhanced textures comparable with RGB vision in daylight.



Extended Data Fig. 4 | HADAR TeX vision recovers textures and overcomes the ghosting effect. Here TeX vision is generated by TeX-SGD. From top to bottom are TeX/thermal/TeX/thermal vision of an off-road night scene at two different positions. HADAR recovers fine textures such as water ripples, bark wrinkles, culverts, as well as the great details of the grass lawn. The HADAR

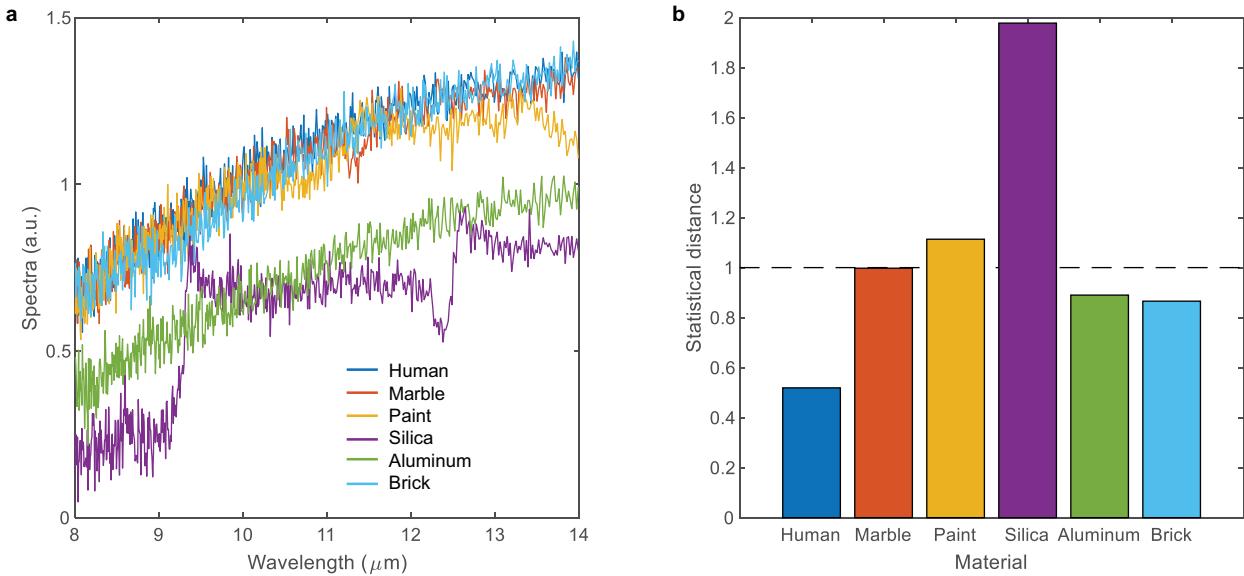
prototype-2 sensor is a focal plane array focusing on infinity. Close objects exhibit focus blur, whereas distant objects are beyond the spatial resolution to show fine details. Therefore, fine textures are mostly observed in a certain distance range.

Article



Extended Data Fig. 5 | HADAR TeX vision overcomes the ghosting effect in traditional thermal vision and beats the state-of-the-art approach to enhance visual contrast. This scene consists of several humans (dark red in TeX vision), robots (purple), cars and buildings on a summer night. Geometric textures of the road and pavements are vivid in TeX vision but invisible in raw

thermal vision and poor in enhanced thermal vision. The mean texture density (standard deviation; see section SII.D of the Supplementary Information for more details) in TeX vision is 0.0788, about 4.6-fold more than the texture density of 0.0170 in the state-of-the-art enhanced thermal vision. This scene is the Street-Long-Animation in the HADAR database.

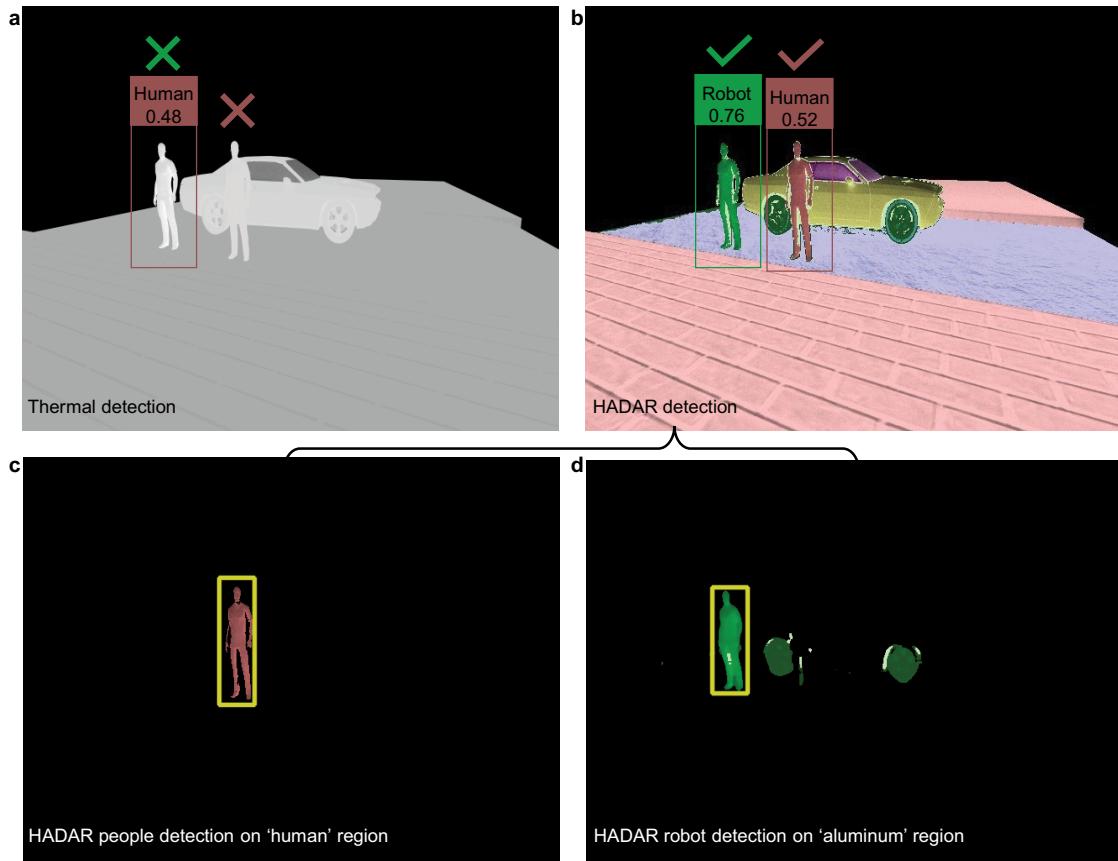


Extended Data Fig. 6 | HADAR estimation theory for multi-material library.

a, Sample incident spectra of five materials generated by Monte Carlo simulations. $T = 60^\circ\text{C}$, $T_0 = 20^\circ\text{C}$ and $V_0 = 0.5$. **b**, Minimum statistical distance of each material. Spectra of silica and paint have non-trivial features that are distinct from other materials in the library. Statistical distance larger than 1 (dashed line) consistently indicates that silica and paint are identifiable. Note that aluminium is similar to human skin under TeX degeneracy and non-identifiable, as discussed in Fig. 3,

even though with the same temperature its spectrum is much weaker than human skin. Emissivity of human skin was approximated as a constant 0.95. Other emissivity profiles were drawn from the NASA JPL ECOSTRESS spectral library. This figure intuitively shows that HADAR identifiability based on semantic/statistical distance is an effective figure of merit to describe identifiability. For more details on generalizing HADAR estimation theory to several materials, see section SII.B of the Supplementary Information.

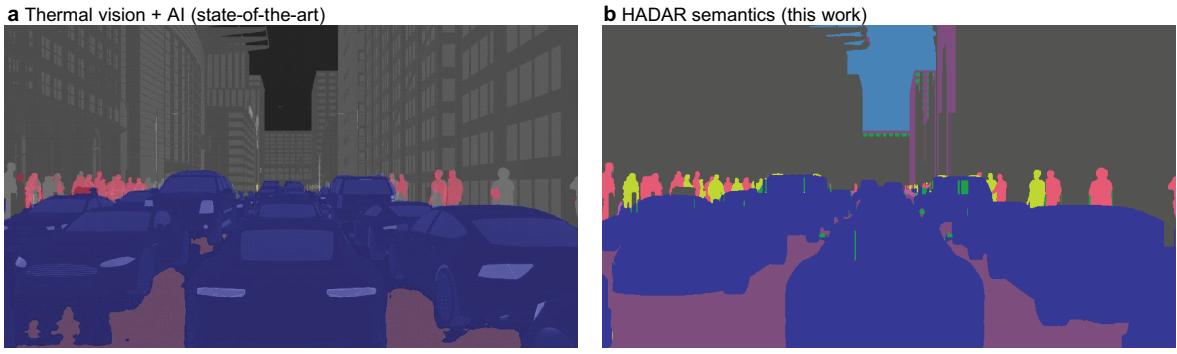
Article



Extended Data Fig. 7 | HADAR detection (TeX vision + AI) beats widely used state-of-the-art thermal detection (conventional thermal vision + AI).

a, Human body detection results based on thermal imaging. **b**, Human and robot identification results based on HADAR. Detection is performed by thermal-YOLO (YOLO-v5 fine-tuned on the thermal automotive dataset; <https://github.com/MALi-Farooq/Thermal-YOLO-And-Model-Optimization-Using-TensorFlowLite>), with detection score/confidence shown together with the bounding box. Owing to TeX degeneracy and the ghosting effect, human body, robot (aluminium at 72.5 °C) and the car (paint at 37 °C) emit similar amounts of thermal radiation and, hence, the human body visually merges into the car in thermal imaging, whereas the robot is mis-recognized as a human body. With our proposed TeX

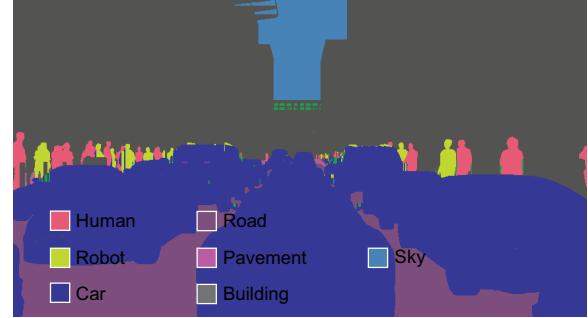
vision, which captures intrinsic attributes, HADAR can distinguish them clearly and yield correct detection. Explicitly, we first extract the material regions corresponding to human (**c**) and robot (**d**), then we perform people detection individually and, finally, we combine detection results to form the final HADAR detection (**b**). We observed that the above results showing the advantage of HADAR TeX vision versus thermal vision is robust and independent of the AI algorithms. Standard computer vision toolbox (people detector in Matlab R2021b) also confirms the results. We also observed that HADAR detection is robust against wrong material predictions, even though a few road pixels under the car and around the human leg are predicted as 'aluminium' in **b** and **d**.



c Ground truth HADAR material map e(m)



d Ground truth semantics



mIoU %	Human	Robot	Car	Road	Pavement	Building	Sky
Thermal vision + AI	33	0	90	67	25	69	16
TeX vision + non-AI	94	84	98	92	87	84	90

e HADAR TeX vision



g HADAR semantics (this work)



f Thermal vision + AI (state-of-the-art)



h Ground truth semantics

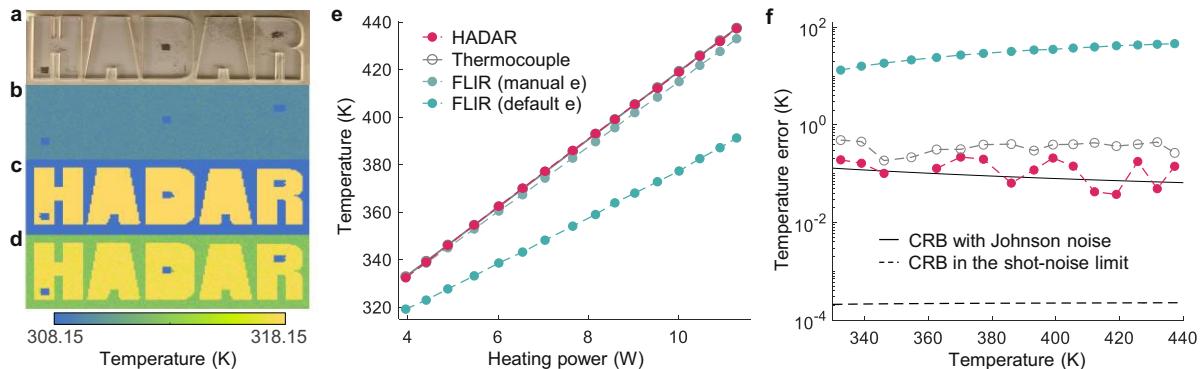


mIoU %	Road	Obstacles	Water	Vegetation
Thermal vision + AI	0.0	0.4	0	0
TeX vision + non-AI	77.4	66.4	89.7	90.0

Extended Data Fig. 8 | HADAR physics-driven semantic segmentation beats state-of-the-art vision-driven semantic segmentation (thermal vision + AI). **a**, Thermal semantic segmentation with DANet (pre-trained on the Cityscapes dataset)⁴⁰. **b**, HADAR semantic segmentation transformed from the material map in estimated TeX vision. **c**, Ground-truth material map in the ground-truth TeX vision. **d**, Semantic segmentation transformed from **c** to approximate the ground-truth segmentation; see section SIII.E of the Supplementary Information for more details of the non-machine-learning transformation. Statistics in the upper table were done on the first four on-road scenes in the HADAR database with fivefold cross validation. **e–h** and the lower table show the typical performance comparison between HADAR versus thermal semantics, in which

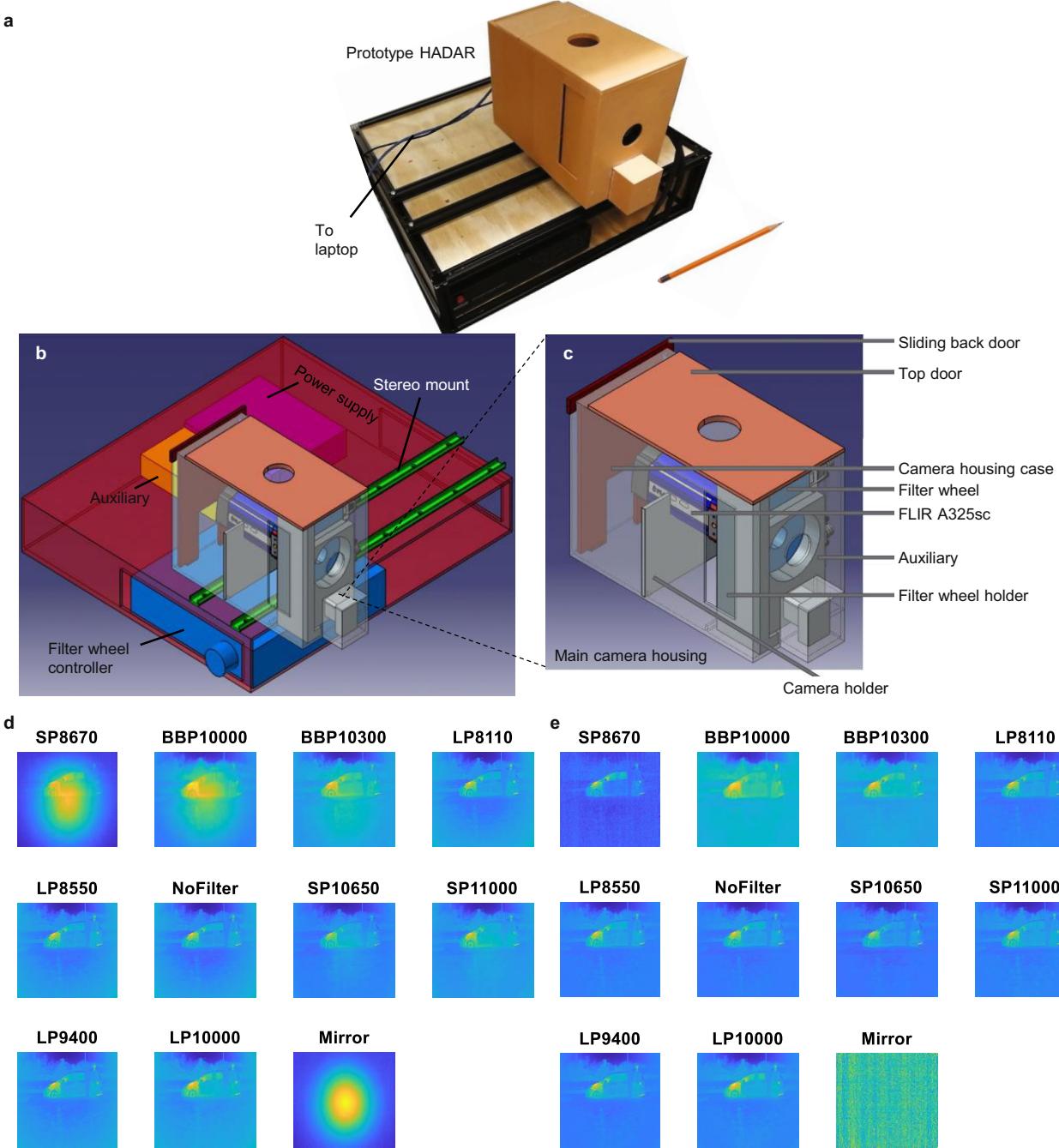
the off-road scene is beyond the training set of DANet. We have also observed consistent results on other non-city scenes in the HADAR database (not shown). This real-world off-road scene is a general example to show the importance of material fingerprint in detection/segmentation. Because AI enhancement is only used in thermal semantics, the advantage of HADAR semantics is clear from TeX vision with physical attributes. In the future, learning-based approaches to convert the material map to semantic segmentation with the help of spatial information may further improve HADAR semantics. mIoU, pixelwise mean intersection over union. Ground truths of the real-world scene were manually annotated.

Article



Extended Data Fig. 9 | Unmanned HADAR thermography reaching the Cramér–Rao bound. By exploiting spectral information and automatically identifying the target, HADAR maximizes temperature accuracy beyond traditional methods. Demonstrated in **a–d** is a HADAR alphabet sample made of plastics at 312.15 K on an unpolished silicon wafer at 317.15 K. **a**, Optical image. **b**, Thermograph using FLIR A325sc shows camouflage and lack of information. **c**, HADAR material readout. **d**, HADAR temperature readout.

b and **d** sharing the same colour bar clearly demonstrate the HADAR advantage. Shown in **e–f** is the measurement of a uniform n-type SiC sample kept on a heating plate with varying heating power. **e**, Mean temperature readout shows that HADAR is unbiased and beats commercial infrared thermograph. **f**, Root mean square error shows that HADAR reaches the Cramér–Rao bound (CRB) for the given detector and imaging system. HADAR also beats commercial thermocouple in precision.



Extended Data Fig. 10 | Prototype HADAR calibration and data collection.
a, Experimental setup of our HADAR prototype-1. **b,c**, 3D schematics of our prototype HADAR. In our prototype HADAR, we used a sturdy stereo mount to take stereo heat cube pairs. **d**, Raw HADAR data with self-radiation of the

detector reflected by filters. **e**, HADAR signal calibrated with a uniform reference object, to remove self-radiation of the detector. For more details of calibration, see section SIV of the Supplementary Information.