nature portfolio

Peer Review File

Mid-infrared snapshot spectral imaging via nonlinear radial dispersion

Corresponding Author: Professor Kun Huang

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The paper by Fang et al. presents an original implementation for performing mid-infrared (MIR) spectral imaging in a single shot and with a low photon count. The proposed technique extends coded-aperture snapshot spectral imaging (CASSI) to the MIR range by employing an innovative up-conversion process in a χ(2) crystal, which also enables spectral-spatial encoding. The resulting visible, spatially encoded image contains all the information necessary to reconstruct the 2D spatial and 1D spectral hyperspectral image.
The technique is novel, and the paper thoroughly details its implementation, the underlying physics, and its limitations—such as spatial and spectral resolution. The authors provide examples of the technique applied to both static and dynamic test samples.
Major Points:
- While the technical innovation is impressive and warrants publication in a high-impact journal, the paper would benefit from demonstrating more practical or realistic applications. It is somewhat disappointing that the technique is exemplified primarily on artificial samples rather than real-world scenarios where it could serve as a viable alternative to existing MIR imaging methods. Including practical applications would significantly strengthen the paper's impact.
- Mid-infrared photothermal imaging is a rapidly advancing field, with numerous studies—such as those by Ji-Xin Cheng's group at Boston University—highlighting its potential for biological and biomedical imaging. The paper would be enhanced by a discussion comparing the advantages and limitations of the proposed technique relative to MIR photothermal imaging.
Minor Point:
- The speckle encoding appears to cancel spatial information at dark speckle pixels. However, the spatial resolution section does not address this issue explicitly. Clarification from the authors on this point would be valuable.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
The work presented by Fang et al, titled "Mid-infrared single-photon snapshot spectral imaging" is a high-quality, thorough discussion about the novel integration of speckle coded aperture spectral imaging for mid-infrared (MIR) upconversion imaging. It builds upon the PI's previous work on this subject with two key developments -- the use of a "random" aperture coding method using a diffuser, and leveraging the inherent angular dispersion of spectral components of the upconverting nonlinear media to project spectral components radially onto an imager, enabling spectral separation of the image for reconstruction. The most considerable drawback to this approach is that it requires calibration of the speckle pattern across the FOV across wavelengths, which is a lot of data points to calibrate if the bandwidth of each point is sufficiently small! This is somewhat negated by the fact that to ensure inversion stability, the "wavelength resolved speckle aperture codes" must be sufficiently dissimilar. Overall, Fang et al. offer a method to greatly simplify the experimental setup given in previous publications on this subject, and the work done to demonstrate and characterize the imaging system is worthy for publication with some revision to the text.

The revision that requires primacy is the notion of "single-photon" imaging, both in the title and in text. The text states that a

photon flux of "1 photon/pixel/pulse" is achieved through a 5-second integration time. The MIR supercontinuum source in use is a 50 MHz repetition rate system, assumed to be running at 25 MHz to be synchronized with the pump system. At 5 seconds, the calculation then becomes 5 seconds/image * 25E6 pulses/second * 1 photon/pulse = 1.25E8 photons/image. For this claim to be true, the camera must have the same number of pixels in the image, which is impossible since the camera in use, the Andor iXon Ultra 888, has only 1.05E6 pixels per the Supplementary Information. Then, in this case, the authors actually achieve 1.25E8 photons/image / 1.05E6 pixels/image = 119 photons/pixel in the aforementioned 5 second integration time. While per-pulse photon flux is a normal figure of merit in upconversion work, it should not be used to frame the sensitivity of this work as "single-photon", especially when hundreds of millions of pulses are captured in each image. It is recommended that the authors remove any mention of "single-photon" and "ultrahigh" efficiency from the text to avoid overstating the detection efficiency, as well as amend the calculations to instead center this discussion around a "photons/pixel" or "J/pixel" in any given image.

Second, a more thorough discussion of the experimental spectral resolution is needed in the text. What influences the spectral resolution? What factors have constrained the images to a 50 wavenumber resolution? It is implied that this is computational in nature in the text, deferring to the Supplementary, but bears naming explicitly, since it seems that calibrating the system, that is, preparing wavelength selective aperture codes for computational use, corrects a number of experimental difficulties. For example, in Supplementary Note 6, it is discussed that the spectral resolution varies over more than an order of magnitude from 1000 nm to 100 nm from the scaling center to 12 mm radially away from the center. However, experimentally this is not true, and it is measured to be flat across the FOV of the system. The authors propose a "dispersive correction factor" provided by the diffuser, which is not clear physically how this is achieved. What seems more probable is that this is corrected by the calibration and the requirement of sufficient code diversity to instead achieve the "filter bandwidth limit". Finally, Supplementary Note 6 claims the spectral resolution is approximately 55 nm across the field of view, it would be preferable to put this discussion somewhere in the main text and in wavenumbers, as is typical for MIR spectroscopy. Including in the main text some of this discussion would greatly enhance the quality of the publication.

These two points notwithstanding, the article is in general well written, with high quality figures. The simulated images are particularly compelling, showing that the Fang et al. have excellent understanding and control of their system. The discussion of the spatial resolution is thorough and sufficient. The spectral images and videos taken are thoughtful and supportive of the work, especially the "ECNU" filtered plate. The supplemental information is likewise informative enough to fully appreciate the work presented, including the computational details of the coded aperture imaging system. The article can be accepted following minor revisions.

(Remarks on code availability)
N/A

Reviewer #3

(Remarks to the Author)
This is an intriguing manuscript that demonstrates a novel broadband mid-infrared (MIR) snapshot imaging approach based on upconversion and intrinsic nonlinear radial dispersion, enabling extraction of spectral information using a conventional silicon detector rather than a dedicated IR detector. Further the method was demonstrated at extremely low photon doses. Overall, the work presents an interesting approach to snapshot MIR spectral imaging that, after revision, is likely worthy of publication in Nature Communications. That said, the practical impact of the method depends critically on how the tradeoffs among sensitivity, spatial resolution, spectral resolution, and computational complexity compare with existing approaches, as well as on the robustness and reliability of the inverse reconstruction used to recover spatial and spectral information from multiplexed measurements. Clarifying these tradeoffs, demonstrating confidence in the reconstructed results, and benchmarking against alternative silicon-based MIR imaging methods would significantly strengthen the manuscript.

Specific comments and suggestions follow:

1) The performance advantages of the current detection scheme could be made clearer. The current scheme achieves significantly worse spatial and spectral resolution than competing IR microscopy/spectroscopy approaches and that renders the current approach not useful for many applications. For example the spectral resolution of FTIR and tunable IR laser based approaches can be ~1 cm-1, whereas the current approach achieves a spectral resolution 50X worse. The spatial resolution shown is around 80 microns, also an order of magnitude worse than competing approaches. The primary advantage appears to be snapshot acquisition at very low photon flux, which comes at the expense of both spatial and spectral resolution. The authors probably want to make clear what target applications are a good match for the advantages of the current approach. The authors may want to consider adding a table that outlines the performance advantages of the current approach as compared to competitive approaches.

2) Fig. 3 should be improved to show the optical path more clearly. For example I would suggest that the view be rotated forward so that it is more clear that light is being collecting from the object by L1 and then combined with the pump light at D1 and then pass through the crystal and angle dispersed. The current view is too foreshortened in the area of L1, DM1 and the CPLN and with overlapping objects such that it is hard to see what is happening. The authors may wish to consider adding an inset zoom on the beam after the CPLN to show the angle dispersion color separation.

3) A fundamental issue with the method described in the manuscript is that it convolves spatial and spectral information into the snapshot images and these two effects need to be deconvolved to extract separated spatial and spectral data. Because spatial and spectral information are multiplexed into a single measurement, the fidelity of the reconstructed data cube

depends critically on the conditioning of the inverse problem. The main manuscript does not sufficiently describe the nature/complexity of solving the inverse problem which is a key element regarding the future applicability of this method. I would recommend the following updates to address this issue:

a) I would recommend adding a diagram or flow chart to illustrate the post-processing that occurs on the snapshot images to extract spatial and spectral information, for example perhaps a simplified version of Fig. S2 in the supplemental information.

b) I would recommend adding specific information about the post-processing time required for each data set using current hardware and software approaches and potentially a roadmap to faster processing if this step is currently a major bottleneck. It is worth understanding that in many analytical applications the relevant metric is total time to usable data, i.e. including acquisition and processing time. It is currently unclear whether reconstruction can be performed at or near the reported acquisition rate. This technique may be less advantageous for some applications if the processing time is prohibitive.

c) The authors note “Because the number of measurements is significantly smaller than the number of unknowns, the system is inherently underdetermined, and infinitely many solutions can satisfy the forward model. To obtain a physically meaningful and stable reconstruction, additional regularization must be imposed to restrict the solution space.” This limitation may or may not be a showstopper for the applicability of this method for real world applications. The authors should expand on this issue and explain what sort of prior knowledge and/or additional regularization is imposed and what the implications are of these requirements. A key issue for researchers relying on non-deterministic inverse problems is demonstrating confidence that the arrived at solution represents physical reality. In particular, it would be useful to understand the robustness of the current reconstruction algorithm. The authors should comment on reality checks that have been or can be performed to address these concerns.

d) The authors should make the post-processing code available on an accessible repository. Doing so would make it much easier for other researchers to replicate and build on this work.

4) Fig. 5B is somewhat misleading without further clarification as it appears to show excellent match of a polystyrene spectrum by the current method with the gold standard FTIR approach. The issue is that the FTIR reference spectrum for polystyrene (PS) looks nothing like the solid trace in Fig. 5B. FTIR spectra are typically acquired at 2-8 cm-1 spectral resolution and an FTIR spectrum of PS would show many more peaks in the spectral range shown. The authors have smoothed the FTIR PS spectrum to 50 cm-1 spectral resolution (similar to the spectral resolution of the snapshot imaging approach) to make the match with FTIR look better. This could lead a naïve reader to assume that the current snapshot imaging approach matches the spectral performance of FTIR which it clearly doesn’t. The text should be edited to clarify that the FTIR spectra are typically acquired at higher spectral resolution and that the PS reference spectrum has been smoothed to match that of the snapshot imaging approach.
5) On a related note, the coarse spectral resolution achieved in the current manuscript limits its usefulness for detailed chemical characterization of many materials. The authors may wish to consider identifying specific applications where coarser spectral is an acceptable tradeoff taking into account the other benefits of the current approach. Further the authors may wish to expand on what applications specifically would benefit from the high-speed snapshot imaging at low photon doses.
6) The authors should review the publication “Infrared chemical imaging through non-degenerate two-photon absorption in silicon-based cameras” (https://www.nature.com/articles/s41377-020-00369-6) which describes a high-speed mid-IR imaging approach using silicon cameras to detect mid-IR photons and without the need for complex post-processing. The authors should consider referencing this publication and clarifying how the current method achieves differentiated performance.

(Remarks on code availability)
I did not see a detailed discussion of the post-processing code or any information about code availability. This is an omission mentioned in the review above.

Version 1:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
The revised work presented by Fang et al. has adequately addressed my concerns regarding the manuscript.

However, I would like to offer a suggestion for future claims of "single-photon imaging". In the author's response, they cite a review paper which has only a single reference that makes the same claim, and uses "0.2 photons/spatial element/second" noise figure to assert "single-photon imaging". It would be much more defensible to claim "single-photon imaging" by choosing to demonstrate a photon shot noise limited imaging system, along with a sub-photon noise floor.

Thank you for the time taken for a productive discussion, and congratulations.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
The authors have thoroughly addressed prior review comments and the manuscript is now suitable for publication.

(Remarks on code availability)
I am not qualified to review the code, but code is now available.

Manuscript NCOMMS-26-018107-T
"Mid-infrared single-photon snapshot spectral imaging"
Reply to the Reviewers

We would like to thank the three reviewers for the careful reading of the manuscript and their valuable reports. We give below a detailed response to the reviewers' comments. Excerpts from the original reports are given in blue. Changes in the revised manuscript are indicated in green.

Reviewer #1

> The paper by Fang et al. presents an original implementation for performing mid-infrared (MIR) spectral imaging in a single shot and with a low photon count. The proposed technique extends coded-aperture snapshot spectral imaging (CASSI) to the MIR range by employing an innovative up-conversion process in a $\chi^{(2)}$ crystal, which also enables spectral-spatial encoding. The resulting visible, spatially encoded image contains all the information necessary to reconstruct the 2D spatial and 1D spectral hyperspectral image.
>
> The technique is novel, and the paper thoroughly details its implementation, the underlying physics, and its limitations—such as spatial and spectral resolution. The authors provide examples of the technique applied to both static and dynamic test samples.

We thank the reviewer for the positive assessment of our work and for recognizing the novelty of the technique, the thoroughness of the implementation, and the detailed characterization of the system's performance.

> While the technical innovation is impressive and warrants publication in a high-impact journal, the paper would benefit from demonstrating more practical or realistic applications. It is somewhat disappointing that the technique is exemplified primarily on artificial samples rather than real-world scenarios where it could serve as a viable alternative to existing MIR imaging methods. Including practical applications would significantly strengthen the paper's impact.

We sincerely thank the reviewer for the encouraging assessment of our work and for this constructive suggestion. We agree that demonstrating the technique on more realistic scenarios would greatly enhance its impact. The primary goal of this manuscript is to introduce and validate a fundamentally new physical mechanism, snapshot spectral imaging via intrinsic nonlinear radial dispersion. As such, we initially chose well-controlled, artificial targets to provide a clean, unambiguous proof-of-principle. However, in response to the reviewer's concern, we have performed additional experiments to explore a more practical application: discriminating between two different thin-film materials based on their distinct mid-infrared vibrational signatures.

As shown in Figure R1, we designed a target consisting of the letters "ECNU" engraved on an aluminum plate. The "EC" region is covered by a polystyrene (PS) thin film, while the "NU" region is covered by a polyethylene (PE) thin film. A snapshot measurement was acquired using our system.

From this single exposure, we successfully reconstructed the spectral responses of both materials. As shown in Figure R1(c), the reconstructed spectra for PS and PE show agreement with reference FTIR measurements (solid lines), with the characteristic absorption features of each material clearly resolved (e.g., PS's aromatic C-H stretch near ~3030 cm^{-1} and PE's aliphatic C-H stretch near ~2920 cm^{-1}). Furthermore, we reconstructed monochromatic images at selected wavenumbers, revealing the spatial distribution of each material based on its unique absorption contrast. Specifically, at 3032 cm^{-1} (where PS absorbs strongly but PE does not), the "EC" region appears darker; at 2920 cm^{-1} (where PE absorbs strongly), the "NU" region appears darker.

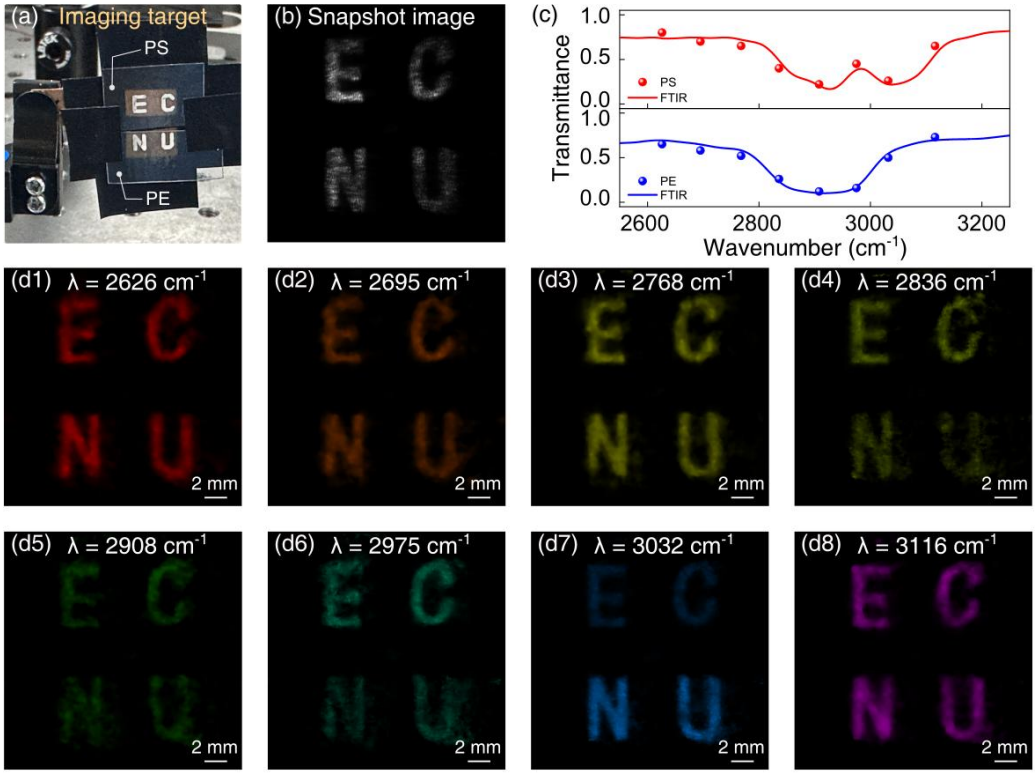

Figure R1. Chemically specific snapshot discrimination of polymer thin films. (a) The target comprises aluminum letters "ECNU" with "EC" covered by polystyrene (PS) and "NU" covered by polyethylene (PE). (b) Single-shot raw measurement. (c) Reconstructed spectra agree with FTIR references, showing distinct PS and PE vibrational fingerprints. (d) Reconstructed monochromatic images, demonstrating clear spatial-spectral contrast.

We believe this result directly addresses the reviewer's concern by demonstrating that our method is capable of tackling a chemically specific, real-world discrimination task, distinguishing different polymer types in a single snapshot without any mechanical scanning. This capability suggests immediate potential for applications such as rapid plastic sorting [R1], material composition identification [R2], and kinetic analysis of liquid mixing processes [R3].

We have added Supplementary Note 7 as an update:

"The mid-infrared spectral region hosts fundamental molecular vibrational resonances that provide chemically specific contrast, making MIR spectroscopy a powerful tool for material identification and discrimination. The snapshot capability of our upconversion imaging system offers

a potential solution for high-speed, non-destructive polymer discrimination. To validate this capability, we applied our system to discriminate between two different polymer thin-film materials based on their distinct MIR vibrational signatures..."

[R1] M. Farries, J. Ward, S. Valle, G. Stephens, P. Moselund, K. V. D. Zanden, and B. Napier, "Mid infrared hyper-spectral imaging with bright super continuum source and fast acousto-optic tuneable filter for cytological applications," Journal of Physics: Conference Series 619, 12032 (2015).

[R2] Y. Zhao, S. Kusama, Y. Furutani, W.-H. Huang, C.-W. Luo, and T. Fuji, "High-speed scanless entire bandwidth mid-infrared chemical imaging," Nat. Commun. 14, 3929 (2023).

[R3] J. Fang, K. Huang, R. Qin, Y. Liang, E Wu, M. Yan and H. Zeng, "Wide-field mid-infrared hyperspectral imaging beyond video rate," Nat. Commun. 15, 1811 (2024).

> Mid-infrared photothermal imaging is a rapidly advancing field, with numerous studies—such as those by Ji-Xin Cheng's group at Boston University—highlighting its potential for biological and biomedical imaging. The paper would be enhanced by a discussion comparing the advantages and limitations of the proposed technique relative to MIR photothermal imaging.

We thank the reviewer for highlighting this important comparison. MIR photothermal imaging [R4, R5] is a powerful technique that achieves excellent spatial resolution (sub-micrometer) by detecting local photothermal effects induced by MIR absorption. This method has been successfully applied to high-resolution chemical imaging of biological and materials science samples. In terms of performance, photothermal imaging offers superior spatial resolution, while our upconversion-based approach provides wide-field snapshot acquisition and operates at much lower photon doses. The two methods thus address different needs: photothermal imaging is well-suited for high-resolution mapping of fine structures, whereas our method prioritizes speed, field of view, and low-light sensitivity for macroscopic dynamic scenes where sub-micrometer resolution is not essential.

In the revised Supplementary Information, we have added related discussion: *"For instance, MIR photothermal imaging (PTI) is implemented by using a visible laser to probe the infrared absorption-induced thermal lensing effect within the sample [16-18]. In wide-field PTI, the required intensive MIR illumination typically relies on an optical parametric oscillator (OPO) with high pulse energy. However, the spectral imaging speed is limited by the slow tuning speed of the OPO, particularly when covering a wide spectrum. Notably, the use of QCLs in MIR photothermal imaging provides access to fast-tuning illumination sources over a wide spectral range. When combined with a fast, high-sensitivity camera, the advent of high-power, high-repetition-rate QCLs could enable wide-field imaging modalities with significantly increased frame rates."*

[R4] X. Ge, Y. Zhu, D. Sun, H. Sun, Y. Li, C. V. P. Dessai and J.-X. Cheng. Fiber laser based stimulated Raman photothermal microscopy towards a high-performance and user-friendly chemical imaging platform. PhotoniX 6, 35 (2025).

[R5] J. Yin, M. Zhang, Y. Tan, Z. Guo, H. He, L. Lan and J.-X. Cheng. Video-rate mid-infrared photothermal imaging by single-pulse photothermal detection per pixel. Sci. Adv. 9, eadg8814 (2023).

> The speckle encoding appears to cancel spatial information at dark speckle pixels. However, the spatial resolution section does not address this issue explicitly. Clarification from the authors on this point would be valuable.

We thank the reviewer for raising this important point. The reviewer is correct that dark speckle pixels locally suppress the signal and thus reduce the information content at those positions for a given wavelength. However, we note that this does not fundamentally prevent accurate reconstruction.

In a typical CASSI system, a coded aperture (binary mask) blocks approximately 50% of the incident light at each spatial position [R6, R7]. Despite this significant signal loss, the spatio-spectral datacube can still be faithfully reconstructed because the encoding pattern is known and the inverse problem is regularized with appropriate priors. The underlying principle is that information is not required to be fully present in every pixel of the raw measurement; rather, it is multiplexed across the entire measurement through the known encoding matrix, and the reconstruction algorithm disentangles it by exploiting spatial and spectral correlations.

Our reconstruction algorithm follows a similar computational framework, where the known speckle encoding patterns serve an analogous role to the coded aperture in CASSI, and the wavelength-dependent radial dispersion provides the spectral multiplexing. The GAP-TV algorithm exploits sparsity and spatial smoothness priors to recover information even from incomplete measurements [R8]. As shown in Supplementary Note 5 (Fig. S5c), the Pearson correlation coefficients between speckle coding matrices at different wavelengths are all below 0.3, confirming sufficient code diversity for reliable reconstruction.

We have added a brief clarification on this point in the revised manuscript: *"Although many pixels are blocked, this sparse structured encoding provides sufficient measurement diversity to enable accurate reconstruction of the full spectral datacube using compressive sensing. [56]"*

[R6] M. E. Gehm, R. John, D. J. Brady, R. M. Willett and T. J. Schulz. Single-shot compressive spectral imaging with a dual-disperser architecture. Opt. Express 15, 14013 (2007).

[R7] H. Song, Y. Feng, X. Dai, L. Bian. Single-photon CASSI: towards ultralow-light spectral imaging. Opt. Lett. 50, 6401 (2025).

[R8] X. Yuan. Generalized alternating projection based total variation minimization for compressive sensing. Proc. IEEE Int. Conf. Inf. Process., 2539-2543 (2016).

Reviewer #2

> The work presented by Fang et al, titled "Mid-infrared single-photon snapshot spectral imaging" is a high-quality, thorough discussion about the novel integration of speckle coded aperture spectral imaging for mid-infrared (MIR) upconversion imaging. It builds upon the PI's previous work on this subject with two key developments - the use of a "random" aperture coding method using a diffuser, and leveraging the inherent angular dispersion of spectral components of the upconverting nonlinear media to project spectral components radially onto an imager, enabling spectral separation of the image for reconstruction. The most considerable drawback to this approach is that it requires calibration of the speckle pattern across the FOV across wavelengths, which is a lot of data points to calibrate if the bandwidth of each point is sufficiently small! This is somewhat negated by the fact that to ensure inversion stability, the "wavelength resolved speckle aperture codes" must be sufficiently dissimilar. Overall, Fang et al. offer a method to greatly simplify the experimental setup given in previous publications on this subject, and the work done to demonstrate and characterize the imaging system is worthy for publication with some revision to the text.

We thank the reviewer for the thorough and insightful evaluation of our work, and for recognizing the two key developments: speckle-based aperture coding and the exploitation of intrinsic radial dispersion for spectral separation.

> The revision that requires primacy is the notion of "single-photon" imaging, both in the title and in text. The text states that a photon flux of "1 photon/pixel/pulse" is achieved through a 5-second integration time. The MIR supercontinuum source in use is a 50 MHz repetition rate system, assumed to be running at 25 MHz to be synchronized with the pump system. At 5 seconds, the calculation then becomes 5 seconds/image * 25E6 pulses/second * 1 photon/pulse = 1.25E8 photons/image. For this claim to be true, the camera must have the same number of pixels in the image, which is impossible since the camera in use, the Andor iXon Ultra 888, has only 1.05E6 pixels per the Supplementary Information. Then, in this case, the authors actually achieve 1.25E8 photons/image / 1.05E6 pixels/image=119 photons/pixel in the aforementioned 5 second integration time. While per-pulse photon flux is a normal figure of merit in upconversion work, it should not be used to frame the sensitivity of this work as "single-photon", especially when hundreds of millions of pulses are captured in each image. It is recommended that the authors remove any mention of "single-photon" and "ultrahigh" efficiency from the text to avoid overstating the detection efficiency, as well as amend the calculations to instead center this discussion around a "photons/pixel" or "J/pixel" in any given image.

We thank the reviewer for this careful reading and for raising an important point regarding the clarity of our sensitivity description. We would like to clarify that the term “1 photon/pixel/pulse” is a standard figure of merit in the field of pulsed nonlinear upconversion detection. Demonstrating detection at an incident flux as low as 1 photon/pixel/pulse indicates that the system operates in the

quantum-noise-limited regime, and this terminology is widely used in the community to describe "single-photon level" sensitivity.

However, we agree that for the broad, interdisciplinary readership of Nature Communications, additional metrics may be helpful to avoid misinterpretation. Therefore, while we retain the "1 photon/pixel/pulse" description as the primary metric given its relevance to the nonlinear upconversion process, we also provide the equivalent incident power per pixel. Specifically, with a pulse repetition rate of 25 MHz, an incident flux of 1 photon/pixel/pulse corresponds to 2.5×10^{7} photons/pixel/second. At a representative wavelength of 3000 nm, the photon energy is approximately 6.63×10^{-20} J, giving an incident power per pixel of 1.65 pW.

We note that this incident power refers to the MIR illumination power incident on the target or the upconversion crystal. After upconversion to the visible band, the signal power detected by the EMCCD is significantly lower (typically at the femtowatt per pixel level) due to the finite upconversion efficiency. Nevertheless, the EMCCD itself is a highly sensitive detector capable of operating at the single-photon level in the visible range, meaning that the signal we detect after upconversion indeed approaches single-photon detection at the sensor side. Therefore, the "single-photon" terminology is not only justified by the incident MIR flux metric but also reflects the actual detection regime of our system. We believe these clarifications provide a more accurate representation of the system's sensitivity.

In the revised manuscript, we have made the following changes:

(1) We have modified the title from "Mid-infrared single-photon snapshot spectral imaging" to *"Mid-infrared snapshot spectral imaging via nonlinear radial dispersion"* to better reflect the central contribution of this work. While the original title emphasized the high sensitivity ("single-photon") aspect, the revised title shifts the focus to the key enabling mechanism—namely, the intrinsic nonlinear radial dispersion in the upconversion process. This mechanism is the core innovation that allows snapshot spectral imaging without external dispersive elements, and it is the main subject of our theoretical and experimental investigations. The new title more accurately and justifiably highlights the conceptual advance of the work.

(2) The sensitivity is now reported as *"high-quality snapshot spectral images are obtained at an incident MIR power of approximately 66 pW/pixel, equivalent to a photon flux of 40 photons/pixel/pulse"* and *"The sensitivity limit of the system is further assessed under the condition of a 5-s camera integration time by reducing the incident photon flux to approximately 1 photon/pixel/pulse (about 1.65 pW/pixel)".* We retain the per-pulse metric as a standard figure of merit in the upconversion imaging community (as used in R9, R10) to facilitate comparison with prior work.

(3) We have added a quantitative system noise analysis in the revised Supplementary Information Note 4, *"The dark noise is an important figure of merit for sensitive optical imaging. In contrast to conventional MIR imagers based on HgCdTe and InSb, the silicon-based cameras are characterized by an extremely low dark current due to the much larger bandgap..."* We emphasize that the photon budget currently employed by this system is several orders of magnitude lower than that required by conventional room-temperature mid-infrared detectors, highlighting the practical advantage of

upconversion-based detection.

[R9] A. Barh, P. J. Rodrigo, L. Meng, C. Pedersen, and P. Tidemand-Lichtenberg. Parametric upconversion imaging and its applications. Adv. Opt. Photon. 11, 952-1019 (2019).

[R10] K. Huang, J. Fang, M. Yan, E Wu and H. Zeng. Wide-field mid-infrared single-photon upconversion imaging. Nat. Commun. 13, 1077 (2022).

> Second, a more thorough discussion of the experimental spectral resolution is needed in the text. What influences the spectral resolution? What factors have constrained the images to a 50 wavenumber resolution? It is implied that this is computational in nature in the text, deferring to the Supplementary, but bears naming explicitly, since it seems that calibrating the system, that is, preparing wavelength selective aperture codes for computational use, corrects a number of experimental difficulties. For example, in Supplementary Note 6, it is discussed that the spectral resolution varies over more than an order of magnitude from 1000 nm to 100 nm from the scaling center to 12 mm radially away from the center. However, experimentally this is not true, and it is measured to be flat across the FOV of the system. The authors propose a "dispersive correction factor" provided by the diffuser, which is not clear physically how this is achieved. What seems more probable is that this is corrected by the calibration and the requirement of sufficient code diversity to instead achieve the "filter bandwidth limit". Finally, Supplementary Note 6 claims the spectral resolution is approximately 55 nm across the field of view, it would be preferable to put this discussion somewhere in the main text and in wavenumbers, as is typical for MIR spectroscopy. Including in the main text some of this discussion would greatly enhance the quality of the publication.

We thank the reviewer for this rigorous analysis. The achievable spectral resolution of the snapshot imaging system is determined by three factors: the calibration channel spacing, the bandwidth of the upconversion process, and the wavelength-dependent spatial separability imposed by the nonlinear radial dispersion. The reviewer is correct that the original explanation of a “translational dispersion correction factor” from the diffuser was insufficiently clear. We have since investigated this effect in detail and can now provide a concrete physical explanation, which we thank the reviewer for motivating.

The diffuser used in our experiment (Thorlabs DG10-1500-M01) is a reflective diffuser with a ground-glass surface (1500-grit) coated with a protective gold film. While such diffusers are often treated as Lambertian scatterers, the 1500-grit surface has a characteristic microstructure with feature sizes on the order of a few microns. This microstructure acts as a grating. When a collimated beam (particularly from a laser source) illuminates this surface, diffraction occurs: different wavelengths are diffracted at different angles, producing a wavelength-dependent spatial dispersion. Importantly, the grinding process can introduce a preferred orientation in the surface structure (e.g., a “brushed” or anisotropic pattern), causing the diffuser to behave more like a one-dimensional grating than an isotropic scatterer. When the incident beam strikes the diffuser at 45°, the projection of the beam onto the surface elongates, modifying the effective interaction period and further enhancing the directional dispersion.

Crucially, this diffuser-induced wavelength-dependent angular dispersion is calibrated into our sensing matrix **Φ** during the AOTF-based calibration procedure. The reconstruction algorithm therefore has access to a rich set of wavelength-specific encoding patterns that include contributions from both the nonlinear radial dispersion from the CPLN crystal, and the diffuser-induced diffraction pattern (wavelength-dependent angular/directional dispersion). As shown in Figure R2, the wavelength-dependent translation coefficient is experimentally characterized. By measuring the lateral displacement of characteristic speckle features as the wavelength varies, the translation coefficient is determined to be $\alpha = 2.77$ μm/nm, which quantifies the angular dispersion introduced by the diffuser.

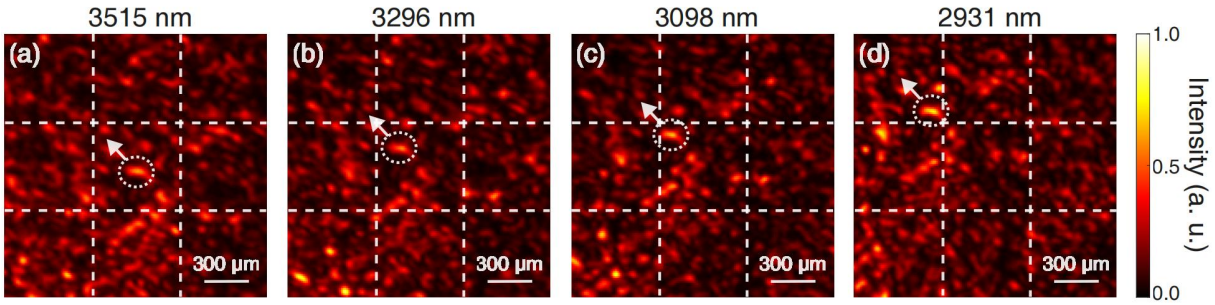

Figure R2. Wavelength-dependent translation dispersion induced by the diffuser. (a-d) show upconverted images of the speckle pattern at selected wavelengths (3515, 3296, 3098, and 2931 nm). The lateral shift of characteristic speckle features across the field of view reveals the dispersion introduced by the diffuser. These shifts are used to determine the wavelength-dependent translation dispersion coefficients for calibration of the sensing matrix.

The reviewer correctly notes that the radial dispersion alone would predict a strongly position-dependent spectral resolution (varying by >10× from center to edge). However, because the diffuser-induced encoding is spatially varying and wavelength-specific, the calibrated sensing matrix provides sufficient code diversity (Pearson correlation coefficients <0.3 for adjacent spectral channels, SI Fig. S5c) for the reconstruction algorithm to disentangle spectral components irrespective of their radial position. As analyzed in Supplementary Note 6, the AOTF used for calibration has a bandwidth of $\Delta\lambda_u \approx 2.9$ nm at the upconverted wavelength. Through energy conservation, this translates to an MIR wavelength uncertainty of $\Delta\lambda_s \approx (\lambda_s^2/\lambda_u^2)\Delta\lambda_u \approx 50$ nm across our 2.5-4.0 μm range, corresponding to $\Delta\nu \approx 50$ cm^{-1}. This is precisely the spectral resolution we observe experimentally, and it remains uniform across the field of view.

We have added a brief clarification in the revised manuscript in the section on Single-photon snapshot spectral imaging: *"The current spectral resolution is primarily limited by the calibration bandwidth of the AOTF. Thanks to the encoding diversity introduced by the diffuser, the achieved resolution surpasses that predicted by a simple model based solely on nonlinear radial dispersion (see Supplementary Note 6)."* And we have clarified this physical mechanism in the revised Supplementary Note 6: *"Physically, this diffuser-induced translational dispersion originates from the ground-glass surface microstructure of the diffuser (Thorlabs DG10-1500-M01). The 1500-grit surface has micron-scale feature sizes that act as a grating. When the broadband MIR beam illuminates this surface at a 45° incident angle, different wavelengths are diffracted at slightly different angles, producing a wavelength-dependent spatial shift in the reflected beam. This*

diffuser-induced angular dispersion is calibrated into the sensing matrix $\boldsymbol{\Phi}$ *alongside the radial dispersion from the nonlinear crystal, and together they provide sufficient wavelength-specific encoding diversity for the reconstruction algorithm to achieve uniform spectral resolution across the field of view. To this end, we selected calibration speckle images at four wavelengths, as shown in Figure S8. Based on the lateral shift of characteristic speckle features across the field of view, the corresponding translational dispersion coefficient is calculated to be* $\alpha = 2.77\ \mu/nm$.”

> These two points notwithstanding, the article is in general well written, with high quality figures. The simulated images are particularly compelling, showing that the Fang et al. have excellent understanding and control of their system. The discussion of the spatial resolution is thorough and sufficient. The spectral images and videos taken are thoughtful and supportive of the work, especially the "ECNU" filtered plate. The supplemental information is likewise informative enough to fully appreciate the work presented, including the computational details of the coded aperture imaging system. The article can be accepted following minor revisions.

We greatly appreciate the reviewer’s time and expertise, and we are grateful for the positive and encouraging evaluation of our work.

Reviewer #3

> This is an intriguing manuscript that demonstrates a novel broadband mid-infrared (MIR) snapshot imaging approach based on upconversion and intrinsic nonlinear radial dispersion, enabling extraction of spectral information using a conventional silicon detector rather than a dedicated IR detector. Further the method was demonstrated at extremely low photon doses. Overall, the work presents an interesting approach to snapshot MIR spectral imaging that, after revision, is likely worthy of publication in Nature Communications. That said, the practical impact of the method depends critically on how the tradeoffs among sensitivity, spatial resolution, spectral resolution, and computational complexity compare with existing approaches, as well as on the robustness and reliability of the inverse reconstruction used to recover spatial and spectral information from multiplexed measurements. Clarifying these tradeoffs, demonstrating confidence in the reconstructed results, and benchmarking against alternative silicon-based MIR imaging methods would significantly strengthen the manuscript.

We thank the reviewer for the thorough and constructive evaluation, and for recognizing the novelty and potential of our approach. We address each specific comment below.

> Specific comments and suggestions follow:
>
> 1) The performance advantages of the current detection scheme could be made clearer. The current scheme achieves significantly worse spatial and spectral resolution than competing IR microscopy/spectroscopy approaches and that renders the current approach not useful for many applications. For example the spectral resolution of FTIR and tunable IR laser based approaches can be ~1 cm^{-1}, whereas the current approach achieves a spectral resolution 50X worse. The spatial resolution shown is around 80 microns, also an order of magnitude worse than competing approaches. The primary advantage appears to be snapshot acquisition at very low photon flux, which comes at the expense of both spatial and spectral resolution. The authors probably want to make clear what target applications are a good match for the advantages of the current approach. The authors may want to consider adding a table that outlines the performance advantages of the current approach as compared to competitive approaches.

We thank the reviewer for this direct and constructive critique. The reviewer is absolutely correct. Compared to high-end FTIR spectrometers (which can achieve ~1 cm^{-1} spectral resolution) and IR microscopes (which can achieve diffraction-limited spatial resolution), our current snapshot method exhibits modest spectral resolution (50 cm^{-1}) and spatial resolution (177 μm).

However, as the reviewer rightly notes, the primary advantage of our method is snapshot acquisition at very low photon flux (approaching 1.65 pW/pixel). This combination of high-speed and high sensitivity is not easily achieved by conventional scanning methods, which typically require either strong illumination (which may damage photosensitive samples) or long integration times (which preclude dynamic scene capture).

Table S1: Performance comparison of wide-field MIR spectral imaging systems.

Type	Ref.	Source	Range (cm^{-1})	Res. (cm^{-1})	Band Num.	Pixel Num.	Time (per cube)	Power ($/nm/cm^2$)	Detector
SFG	This	SC	2500-4000	50	24	1024 × 1024	50 ms[a] 5 s[b]	0.8 μW 0.2 nW	Si
	[5]	SC	2600-4085	50	100	1024 × 1024	16 ms[c]	7 μW	Si
	[20]	OPO	2500-4348	3.6-7.6	62	658 × 496	25 s[d]	20 mW	Si
	[21]	DFG	640-3015	2.6-3.7	1069	640 × 480	8 s	1 μW	Si
NTA	[19]	OPO	2735-3265	8.4	63	1280 × 1024	1.1 s	17 μW	InGaAs
PTI	[16]	OPO	2830-2950	10	13	2016 × 2016	260 ms	4 W	Si
CS	[23]	Globar	2083-2703	10	111	640 × 512	20 ms	/[f]	InSb
	[22]	Globar	2000-3333	20	100	64 × 48	60 s	/[f]	InSb
Direct	[10]	Globar	988-1950	16	FTIR[e]	96 × 96	93 ms	/[f]	MCT
	[11]	Globar	1850-6667	1	FTIR[e]	256 × 160	60 s	/[f]	InSb
	[12]	QCL	952-1351	2	137	128 × 128	95 ms	18 μW	MCT
	[13]	QCL	777-1904	2	282	128 × 128	120 s	121 mW	MCT
	[14]	QCL	1160-1320	1	565	640 × 480	11.3 s	75 mW	Bolometer
	[15]	SC	2222-5000	1-5	100	640 × 512	1.2 s	7 mW	InSb

[a] Continuous shooting mode for dynamic scene capture at a frame rate of 20 Hz.

[b] Mid-infrared snapshot images collected under low illumination were recorded using an EMCCD set to -80 ℃ with a gain of 1000.

[c] The acquisition time is determined by the maximum frame rate of 6.4 kHz for the full spatial format of the CMOS camera used in our experiment.

[d] Although a wide-field image at a single wavelength can be recorded in 2.5 ms, the exposure time of the camera is set to be 400 ms for improving the image quality in the spectral imaging demonstration.

[e] The spectral imaging is realized based on the FTIR technique.

[f] The illumination power for the used thermal source is not given in these works.

In response to the reviewer's suggestion, we have added a performance comparison table (Supplementary Table S1) to clearly outline the trade-offs between our method and competing MIR imaging/spectroscopy approaches. We have also added Supplementary Note 8, clearly indicating in which specific application areas our adopted methods have advantages that outweigh their limitations:

"..., we design and implement ultrasensitive MIR snapshot spectral imaging at video frame rates based on a nonlinear radial dispersion process. In this scheme, the full MIR information in both spatial and spectral domains experiences simultaneous wavelength conversion and spatial dispersion during the upconversion process, while speckle illumination provides spatial encoding,

eliminating the need for additional spatial modulation elements. Notably, controlling illumination intensity is critical in biochemical imaging to prevent sample damage. Here, the illumination intensity is approximately 0.2 nW/nm/cm^2, which is significantly lower than that of other spectral imaging systems. The illumination power could be further reduced by improving conversion efficiency. Therefore, the achieved features of high acquisition rates, wide-field operation, and broadband spectral coverage open up new possibilities for the high-throughput characterization of dynamic processes in chemical, medical, and bio-related fields."

2) Fig. 3 should be improved to show the optical path more clearly. For example I would suggest that the view be rotated forward so that it is more clear that light is being collecting from the object by L1 and then combined with the pump light at D1 and then pass through the crystal and angle dispersed. The current view is too foreshortened in the area of L1, DM1 and the CPLN and with overlapping objects such that it is hard to see what is happening. The authors may wish to consider adding an inset zoom on the beam after the CPLN to show the angle dispersion color separation.

We thank the reviewer for this constructive suggestion regarding Fig. 3. We agree that the optical layout, particularly around the relay lenses, dichroic mirror, and the CPLN crystal, appears somewhat compressed in the current view.

However, we would like to note that the conceptual principle of the angle dispersion and spectral multiplexing is already illustrated in Fig. 1(B) and Fig. 1(C) of the main manuscript, where the wavelength-dependent angular separation and radial scaling are shown schematically without the clutter of actual optical components.

To help the reviewer and readers better understand the experimental layout, we have made slight adjustments to the positions of the optical components to avoid overlapping and improve clarity, while keeping the overall layout unchanged. In addition, we have added a short clarifying sentence in the figure caption of Fig. 3 to explicitly guide the reader's attention to the beam path: *"The wavelength-dependent phase matching in the nonlinear crystal gives rise to an intrinsic nonlinear radial dispersion, which directly encodes spatio-spectral information into the upconverted image. Consequently, MIR spectral imaging can be achieved in a single shot."* And we have also added a cross-reference in the main text to remind readers that the spectral dispersion mechanism is conceptually shown in Fig. 1(B): *"In contrast, our approach replaces the external dispersive element with an intrinsic nonlinear mechanism. As illustrated in Fig. 1(B), nonlinear frequency upconversion in a phase-matched crystal simultaneously performs spectral translation and spectral dispersion."* We hope that these textual clarifications, together with the existing conceptual illustration in Fig. 1, will sufficiently address the reviewer's concern without requiring a complete redrawing of Fig. 3.

3) A fundamental issue with the method described in the manuscript is that it convolves spatial and spectral information into the snapshot images and these two effects need to be deconvolved to extract separated spatial and spectral data. Because spatial and spectral information are multiplexed into a single measurement, the fidelity of the reconstructed data

cube depends critically on the conditioning of the inverse problem. The main manuscript does not sufficiently describe the nature/complexity of solving the inverse problem which is a key element regarding the future applicability of this method.

We thank the reviewer for raising this fundamental and important issue regarding the inverse problem in computational spectral imaging. The reviewer is absolutely correct that the fidelity of the reconstructed datacube depends critically on the conditioning of the sensing matrix and the regularization strategies employed. Below we respond to each of the reviewer's specific recommendations (a–d) in detail.

I would recommend the following updates to address this issue:

a) I would recommend adding a diagram or flow chart to illustrate the post-processing that occurs on the snapshot images to extract spatial and spectral information, for example perhaps a simplified version of Fig. S2 in the supplemental information.

We thank the reviewer for this suggestion. We agree that illustrating the post-processing workflow would be helpful for readers.

However, we note that the essential steps of the computational pipeline are already presented in the Supplementary Information. Specifically, Supplementary Fig. S1 schematically illustrates the snapshot spectral encoding process via intrinsic nonlinear radial dispersion, showing how the scene is spatially encoded and then spectrally multiplexed into a single two-dimensional measurement. Supplementary Fig. S2 shows the vectorization of the spectral datacube and the construction of the sensing matrix $\mathbf{\Phi}$ from the wavelength-dependent encoding patterns. Together, these two figures provide a complete picture of the forward model and the inverse problem formulation.

We have now added explicit cross-references to these figures in the main text to guide readers to the relevant details in the Basic principle section: *"A detailed illustration of the vectorization process and the construction of the sensing matrix is provided in Supplementary Notes 1 and 2."*

We believe that with these cross-references, the existing figures in the Supplementary Information are sufficient to convey the post-processing workflow without requiring an additional flowchart in the main text.

b) I would recommend adding specific information about the post-processing time required for each data set using current hardware and software approaches and potentially a roadmap to faster processing if this step is currently a major bottleneck. It is worth understanding that in many analytical applications the relevant metric is total time to usable data, i.e. including acquisition and processing time. It is currently unclear whether reconstruction can be performed at or near the reported acquisition rate. This technique may be less advantageous for some applications if the processing time is prohibitive.

The reviewer raises an important practical consideration: the total time to usable data, including both acquisition and post-processing. We have now quantified the processing time for our current implementation. Current processing time: Using a standard desktop computer (Intel Core

i9-14900HK, 64 GB RAM, Python 3.9, without GPU acceleration), the reconstruction of a single 1024×1024×24 spectral datacube (24 spectral channels) takes approximately 10 minutes using the GAP-TV algorithm with 100 iterations. The acquisition time per snapshot is 50 ms (for video-rate imaging at 20 Hz) or 5 s (for high sensitivity scenarios). Thus, while the system can acquire snapshots at video rate, the current post-processing time is significantly longer than the acquisition time. We acknowledge that for applications requiring real-time feedback, this processing delay would be a limitation.

We would like to clarify that the primary goal of this manuscript is to introduce and validate a fundamentally new mechanism, snapshot spectral imaging via intrinsic nonlinear radial dispersion. The demonstration of real-time acquisition (20 Hz video-rate imaging of a rotating target, as shown in Fig. 6 and Supplementary Video 1) serves to illustrate the system's capability to capture dynamic scenes without motion artifacts. We do not claim that real-time reconstruction has been achieved in the current implementation.

We fully agree with the reviewer that for many practical applications, the total time to usable data is the relevant metric, and we acknowledge that the current processing time is a bottleneck. However, we also note that there is substantial room for improvement:

(1) GPU acceleration: The GAP-TV algorithm is highly parallelizable. Preliminary tests using PyTorch with a high-performance GPU demonstrate a substantial speedup, potentially reducing the reconstruction time to well under a minute per datacube.

(2) Deep learning-based reconstruction: Learned reconstruction methods (e.g., U-Net or transformer-based architectures) can perform inference in milliseconds once trained. This is an active direction in compressive spectral imaging and could enable true real-time reconstruction.

(3) Algorithmic optimization: The current Python implementation uses non-optimized matrix operations. Further optimizations (e.g., sparse matrix representations, just-in-time compilation with Numba, or better preconditioning) would further reduce processing time.

We have added the following discussion to the Methods (Reconstruction algorithm) section in the revised manuscript: *"Reconstruction of a single 1024 × 1024 × 24 spectral datacube requires approximately 10 minutes on a standard desktop computer (Intel Core i9-14900HK, 64 GB RAM, Python 3.9, without GPU acceleration) using the GAP-TV algorithm with 100 iterations."* And we have further added the time information for acquisition and processing as well as the relevant conditions in the revised Supplementary Note 2: *"On a standard desktop computer (Intel Core i9-14900HK, 64 GB RAM, Python 3.9, without GPU acceleration), the GAP-TV algorithm with 100 iterations reconstructs a single 1024×1024×24 spectral datacube (24 spectral channels) in approximately 10 minutes. While the system acquires snapshots at 20 Hz, post-processing remains a bottleneck for real-time applications. However, significant acceleration is feasible: GPU acceleration offers substantial speedup, reducing reconstruction time to well under a minute; deep learning-based methods can achieve millisecond inference; and further algorithmic optimizations (sparse matrices, Numba JIT compilation) would provide additional gains. The reconstruction code is provided as open source to support such optimizations."*

c) The authors note "Because the number of measurements is significantly smaller than the number of unknowns, the system is inherently underdetermined, and infinitely many solutions can satisfy the forward model. To obtain a physically meaningful and stable reconstruction, additional regularization must be imposed to restrict the solution space." This limitation may or may not be a showstopper for the applicability of this method for real world applications. The authors should expand on this issue and explain what sort of prior knowledge and/or additional regularization is imposed and what the implications are of these requirements. A key issue for researchers relying on non-deterministic inverse problems is demonstrating confidence that the arrived at solution represents physical reality. In particular, it would be useful to understand the robustness of the current reconstruction algorithm. The authors should comment on reality checks that have been or can be performed to address these concerns.

We thank the reviewer for this crucial question. The reviewer is correct that for an underdetermined inverse problem, the solution is not unique without additional constraints. Here we explain the prior knowledge we impose, how we validate the reconstruction fidelity, and the robustness of our approach.

In our GAP-TV formulation, we impose two forms of prior knowledge. The first one is spatial sparsity in the gradient domain (Total Variation, TV). TV regularization assumes that the spatial distribution of the scene is piecewise smooth with sparse gradients. This is a weak but widely applicable prior for natural scenes and is particularly suitable for imaging targets with sharp edges (e.g., the ECNU letters, USAF target, and airplane pattern in our experiments). It does not assume that the scene is sparse in the pixel domain, only that its gradient is sparse. We also employed non-negativity constraints. The spectral datacube is physically constrained to have non-negative intensity values at all spatial and spectral positions. This simple constraint significantly reduces the solution space. We do not assume any prior knowledge of the spectral signatures (e.g., specific absorption peaks), which makes the method applicable to unknown samples.

In Fig. 2 of the main manuscript, we validated our reconstruction algorithm against a fully known synthetic hyperspectral scene. The excellent agreement (PSNR >30 dB across all channels, high spectral fidelity as shown in Figs. 2(G,H)) demonstrates that the algorithm can accurately recover the ground truth when the forward model is known and the scene satisfies the prior assumptions. In Figs. 5(B) and 6(B), we compared the reconstructed spectra from our snapshot system with reference FTIR measurements. The close agreement (typically within the experimental uncertainty) confirms that the reconstructed spectral features are physically meaningful and not artifacts of the reconstruction. The reconstructed monochromatic images (Figs. 5(D-F), 6(D1-D4), and Supplementary Fig. S8) show spatially contiguous features that match the expected morphology of the targets. If the reconstruction were dominated by artifacts, we would expect random or checkerboard-like patterns rather than coherent spatial structures.

We also wish to be transparent about the limitations. The reconstruction quality degrades when: (i) the scene violates the piecewise-smooth assumption (e.g., highly textured or random patterns), (ii) the spectral features are narrower than the calibration channel spacing. Therefore, we have added the following discussion to the revised Supplementary Note 2: *"Several factors can degrade the reconstruction fidelity of our current algorithm. First, the GAP-TV reconstruction assumes that the*

spatial distribution of the scene is piecewise smooth with sparse gradients. This assumption holds well for targets with distinct spatial features (e.g., letters, resolution targets, and patterned thin films), but may lead to reduced fidelity for highly textured or random scenes (e.g., biological tissues with fine structures or rough surfaces). Second, the spectral resolution is primarily limited by the calibration channel spacing (approximately 50 cm^{-1} in our implementation). Spectral features narrower than this spacing may not be fully resolved and could be smeared across adjacent channels.

To address these limitations, we envision several strategies. For highly textured scenes, alternative regularization schemes (e.g., wavelet-based sparsity or learned priors from deep generative models) could replace or complement total variation. For sub-calibration spectral features, a finer calibration step (e.g., using an AOTF with narrower bandwidth or a tunable laser source) could improve spectral resolution at the cost of increased calibration time. Alternatively, advanced super-resolution algorithms or dictionary learning methods could potentially extract sub-channel spectral information if the target spectra are known to lie within a low-dimensional manifold. These improvements are beyond the scope of the present work but represent promising directions for future research."

> d) The authors should make the post-processing code available on an accessible repository. Doing so would make it much easier for other researchers to replicate and build on this work.

We thank the reviewer for this constructive suggestion. In response, we have uploaded the complete Python implementation of our reconstruction algorithm (including GAP-TV, sensing matrix construction, and calibration routines) as code files on GitHub:

https://github.com/HKLAB-ECNU/MIR-snapshot-spectral-imaging.

The code is accessible to the reviewers and will be made publicly available upon publication.

> 4) Fig. 5B is somewhat misleading without further clarification as it appears to show excellent match of a polystyrene spectrum by the current method with the gold standard FTIR approach. The issue is that the FTIR reference spectrum for polystyrene (PS) looks nothing like the solid trace in Fig. 5B. FTIR spectra are typically acquired at 2-8 cm-1 spectral resolution and an FTIR spectrum of PS would show many more peaks in the spectral range shown. The authors have smoothed the FTIR PS spectrum to 50 cm-1 spectral resolution (similar to the spectral resolution of the snapshot imaging approach) to make the match with FTIR look better. This could lead a naïve reader to assume that the current snapshot imaging approach matches the spectral performance of FTIR which it clearly doesn't. The text should be edited to clarify that the FTIR spectra are typically acquired at higher spectral resolution and that the PS reference spectrum has been smoothed to match that of the snapshot imaging approach.

We thank the reviewer for pointing out this critical issue regarding the presentation of the FTIR reference data in Fig. 5B. The reviewer is absolutely correct. The raw FTIR spectrum of polystyrene (PS) measured at high resolution (e.g., 2-8 cm^{-1}) contains numerous sharp vibrational peaks. The solid trace shown in our original Fig. 5B was indeed a smoothed version of the raw FTIR data,

convolved to a spectral resolution of approximately 50 cm^{-1} to match the native resolution of our snapshot imaging system. We sincerely apologize for not making this preprocessing step clear in the original figure and caption, as this could indeed mislead a reader into overstating the performance of our method.

Our intention was to provide a fair, like-with-like comparison between the reconstructed spectrum from our snapshot method and a reference spectrum at the same effective spectral resolution. Since our current snapshot system operates at a resolution of ~50 cm^{-1}, comparing it directly to a high-resolution FTIR spectrum would be inappropriate and would incorrectly suggest that our method fails to resolve peaks that it is fundamentally not designed to resolve. Therefore, we convolved the high-resolution FTIR data with a Gaussian kernel to match the 50 cm^{-1} resolution of our system.

We have made the following changes in the revised manuscript to avoid any misinterpretation, *"The reconstructed spectrum of the polystyrene film [red symbols in Fig. 5(B)] shows good agreement with the FTIR reference [black curve in Fig. 5(B)], where the FTIR spectrum has been convolved to a spectral resolution of about 50 cm^{-1} for the sake of direct comparison."* We will also add a note in the figure caption stating: *"Note that the FTIR reference spectrum (the black curve) has been convolved to a spectral resolution of 50 cm^{-1} for the sake of direct comparison."*

> 5) On a related note, the coarse spectral resolution achieved in the current manuscript limits its usefulness for detailed chemical characterization of many materials. The authors may wish to consider identifying specific applications where coarser spectral is an acceptable tradeoff taking into account the other benefits of the current approach. Further the authors may wish to expand on what applications specifically would benefit from the high-speed snapshot imaging at low photon doses.

We thank the reviewer for this constructive suggestion. The reviewer is correct that the current spectral resolution is coarser than that of a typical high-end FTIR spectrometer. We fully acknowledge this limitation and agree that for applications requiring detailed vibrational fingerprinting (e.g., identifying subtle chemical differences or resolving overlapping peaks), higher spectral resolution is necessary.

However, we would like to emphasize that many real-world applications do not require ultrahigh spectral resolution. Instead, they benefit from the unique combination of snapshot acquisition, wide-field imaging, and single-photon sensitivity offered by our method, even at moderate spectral resolution. For rapid material sorting (e.g., distinguishing PS from PE in recycling streams), this resolution readily discriminates broad polymer classes. For biomedical imaging, the low photon dose (~1 photon/pixel/pulse) minimizes photodamage, making the method attractive for live-cell or tissue imaging where phototoxicity is a concern. Additionally, the snapshot acquisition (20 Hz) enables real-time monitoring of dynamic processes such as polymer curing or thermal decomposition, where spectral features broaden at elevated temperatures and high-speed capture is prioritized over high resolution.

We have added the following discussion to the revised manuscript: *"The achieved modest spectral*

resolution is sufficient for material identification and sorting [42, 53], as exemplified by the chemical film identification demonstration in Supplementary Note 7."

We believe these additions appropriately frame the capabilities and limitations of our method and will help readers identify whether our approach is suitable for their specific needs. We thank the reviewer for prompting this important clarification.

> 6) The authors should review the publication "Infrared chemical imaging through non-degenerate two-photon absorption in silicon-based cameras" (https://www.nature.com/articles/s41377-020-00369-6) which describes a high-speed mid-IR imaging approach using silicon cameras to detect mid-IR photons and without the need for complex post-processing. The authors should consider referencing this publication and clarifying how the current method achieves differentiated performance.

We thank the reviewer for drawing our attention to this relevant and important work by David Knez et al. (Light Sci. Appl. 9, 125, 2020) [R11], which demonstrates mid-infrared chemical imaging via non-degenerate two-photon absorption (NTA) directly in a silicon-based CCD camera. We have now cited this paper in our revised manuscript (Ref. [41]): *"Recently, two-photon absorption has enabled MIR spectral imaging on a Si-based chip [41, 42] albeit with limited detection efficiency due to the high-order nonlinearity."*

While both that work and our approach enable MIR imaging using silicon-based detectors, the underlying physical mechanisms and resulting system characteristics are fundamentally different. We have added a comparison in Supplementary Note 8 to clarify the differentiated performance of our method: *"Recently, NTA-based spectral imaging was implemented using a high-definition InGaAs camera [19]. While the spectral scanning rate can be fast during a single sweep, the dwell time between sequential sweeps is relatively long due to the mechanical inertia of the translational stage. Consequently, the demonstrated refresh rate for datacube acquisition is limited to the Hz level in continuous operation mode. It is worth mentioning that the refreshing rate in the NTA-based scheme could be substantially improved by using a voice coil actuator as the delay line, which would facilitate a real-time hyperspectral imaging at high definition."*

[R11] D. Knez, A. M. Hanninen, R. C. Prince, E. O. Potma, and D. A. Fishman. Infrared chemical imaging through non-degenerate two-photon absorption in silicon-based cameras. Light Sci. Appl. 9, 125 (2020).

> I did not see a detailed discussion of the post-processing code or any information about code availability. This is an omission mentioned in the review above.

We thank the reviewer for this constructive suggestion. In response, we have provided a detailed discussion of the post-processing procedure in Supplementary Notes 1 and 2, covering the forward model and the reconstruction algorithm. Additionally, to facilitate use by reviewers and readers, we have uploaded the reconstruction code to GitHub:

https://github.com/HKLAB-ECNU/MIR-snapshot-spectral-imaging.