CNN based demosaicing for labeled fluorescence cancer intraoperative imaging with visible-NIR sensors

Yifei Jin[a], Borislav Kondov[b], Goran Kondov[b], Sunil Singhal[c], Shuming Nie[a,d,e], and Viktor Gruev[a,d,e,f,*]

[a]Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[b]Department of Thoracic and Vascular Surgery, Ss. Cyril and Methodius University of Skopje
[c]Perelman School of Medicine, Department of Thoracic Surgery, University of Pennsylvania
[d]Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[e]Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[f]Carle Illinois College of Medicine, University of Illinois at Urbana-Champaign, Urbana, IL 61820, USA

ABSTRACT

Single-chip imaging devices with vertically stacked photodiodes and pixelated spectral filters enhance multi-dye imaging techniques for cancer surgeries. However, this advancement sacrifices spatial resolution. To address this issue, we have created a deep convolutional neural network designed to demosaic color and NIR channels, and its effectiveness has been confirmed through testing on both preclinical and clinical datasets.

Keywords: CNN, demosaicing, near infrared imaging, image guided surgery, cancer surgery

1. INTRODUCTION

Almost one in three people globally will have cancer, with surgical treatment being vital to excise localized tumors and identify metastases.[1–5] Surgery guided by near-infrared (NIR) imaging is becoming progressively crucial to improving the accuracy of surgical techniques, driven by the development of highly targeted tumor probes,[6,7] advanced imaging modalities,[8] and the adoption of machine-learning techniques into the surgical workflow.[9] The more recent regulatory approval of tumor-targeted probes based on near-infrared (NIR) imaging such as Cytolux has cemented NIR imaging as a novel and beneficial method for clinical practice.[10–12]

NIR imaging instruments evolved in concert with color imaging technology over the last decade.[13–18] Initial approaches used single NIR filters, or multiplexed signals of color and NIR over time, ultimately causing motion artifacts. Some beam-splitter solutions allowed for simultaneous capture of color and NIR data at the cost of larger and more complex instruments.[19,20] Inspired by the Bayer filter's widespread implementation and success, our team created a compact, bioinspired hexachromatic sensor with color-NIR filters arranged in a pixelated fashion and photodiodes that are vertically stacked to enable real-time imaging using multiple NIR fluorescent markers.[21–23] As with Bayer-patterned sensors, however, pixelation degrades spatial resolution,[24–26] thus there is also a need for efficient demosaicing.

Demosaicing methods vary from simple interpolation to dictionary and learning-based approaches.[27,28] Recently, convolutional neural networks (CNNs) have become state of the art for inferring full-resolution images from mosaic data due to their ability to learn relevant local features and inter-channel relationships.[29–31] In this paper, we present a CNN model specifically designed for our hexachromatic sensor, which recovers resolution for both visible and NIR channels (Fig. 1A-D). We compare its performance with conventional interpolation methods on both a benchmark dataset (the Waterloo dataset[32]) and data acquired from our vertically stacked sensor. Finally, we illustrate the clinical application of our method using preclinical and clinical examples. Comparing to the previous work,[33] we include the three channels reconstruction results for visible and NIR spectrum through preclinical and clincal dataset.

Advanced Biomedical and Clinical Diagnostic and Surgical Guidance Systems XXIII, edited by Caroline Boudoux, James W. Tunnell, Proc. of SPIE Vol. 13306, 133060B · © 2025 SPIE · 1605-7422 · doi: 10.1117/12.3044342

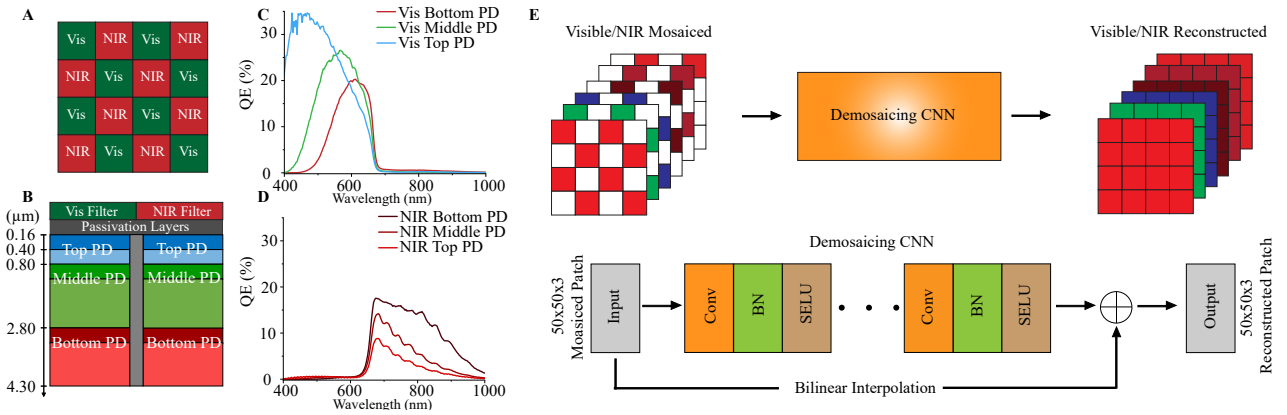

Figure 1. (A-B) illustrates the configuration of our visible-NIR image sensor with pixelated filters and vertically stacked photodiodes. (C-D) presents the quantum efficiency characteristics of the image sensor. (E) showcases the design of the demosaicing CNN for both color and NIR channels.

2. CNN DEMOSAICING IMPLEMENTATION

Figure. 1E shows the architecture of our CNN model for the bio-inspired camera demosaicing. The inpupt of the CNN models contained 3 six channels of mosaiced 2D images, three for the visible image and the other three for the NIR image. Because most of the demosaicing problem happens locally with a small range of pixels, our model residual learning could be taken effectively with a crop of the image instead of a whole image. For each mosaiced channel, the image is cropped to 50x50 pixels patches as an input for our model. Our residual learning-based model contains 20 layers, and each layer contains a convolution, a batch normalization and a SELU activation sublayer. The model follows the below equations:

$$F_n(\mathbf{Y}) = selu(\mathbf{Y} * W_n + B_n), n = 1...N\text{-}1 \tag{1}$$

$$\text{selu}(x) = \begin{cases} \lambda x & \text{if } x > 0 \\ \lambda\alpha(e^x - 1) & \text{if } x \leq 0 \end{cases} \tag{2}$$

$$F(\mathbf{Y}) = F_{N\text{-}1}(\mathbf{Y}) * W_N + B_N \tag{3}$$

We process three visible cropped patches and three NIR cropped patches through two parallel CNN models. The models for visible and NIR images are identical only with different patterns of mosaiced inputs. In the meantime, a bilinear interpolation is implemented on the input cropped mosaiced images as a baseline for residual learning. The convolution layers output is added to the bilinear interpolation outputs as the model's results. Then it calculates the residual difference between the model's results and the ground truth by using an L2 loss function. With this method, the model is deliberately focused on learning the differences between the actual data and the initial interpolated approximation, which tend to represent the more complex elements. As a result, this specialized training strategy improves the model's capacity to learn more effectively and efficiently. Our models are trained on 474,400 image patches from the WED dataset containing 4,744 colorful images.

3. EVALUATION OF SINGLE-MODE VISIBLE IMAGE DEMOSAICING

To assess the CNN-based demosaicing routine's performance of different images with rich colors and textures, we use an unmosaiced image as a reference collected by a commercial camera (DP1x, Sigma) without a color filter array. To bandpass only visible spectrum, a short-pass filter is mounted on the top of the camera while taking a single still image with dimensions of 2640 by 1760 pixels. This image, featured as (Fig.2A), captures the Electrical and Computer Engineering Building at the University of Illinois at Urbana-Champaign. We mosaic the reference image by the checkerboard visible-NIR filter pattern to emulate the visible pixels of the camera

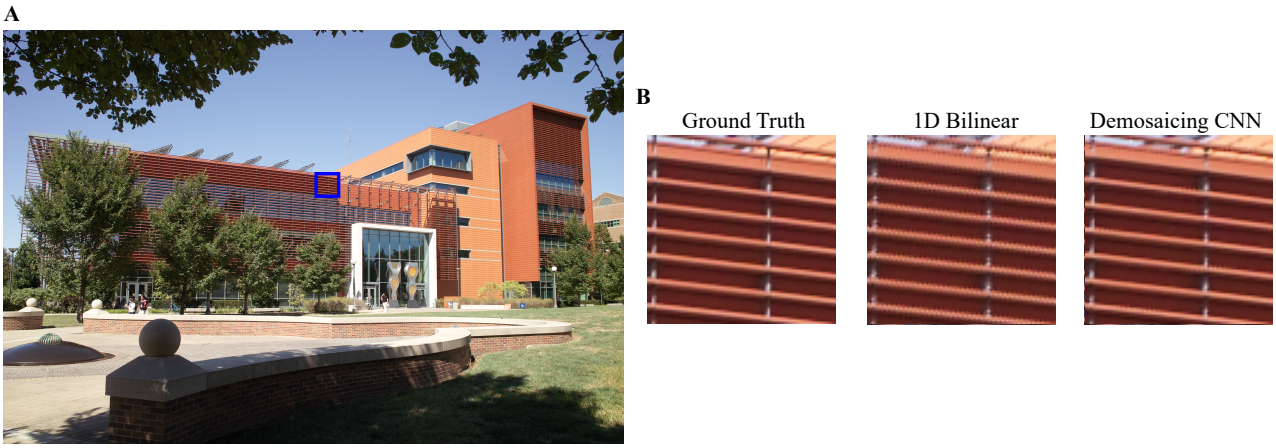

Figure 2. (A) Building image of Electrical and Computer Engineering Building at the University of Illinois at Urbana-Champaign captured by DP1x, Sigma (B) 80x80 pixels patches of building façade. Demosicing CNN shows much less zig-zag in the reconstruction comparing to 1D bilinear results.

frame. The mosaiced reference image is as the input to our model. To evaluate the performances of our model, the original unmosaiced reference image is considered as the ground truth to measure the reconstruction image quality. A 1D bilinear interpolation is also implemented as a baseline to compare its demosaicing performance with our model.

Four measures of the reconstruction image quality for quantitative analysis include peak signal-to-noise ratio (PSNR), mean squared error (MSE), 95th structural dissimilarity index measure (DSSIM), and 95th percentile color difference (ΔE). Compared to the bilinear interpolation's visible spectrum image reconstruction quality, our CNN model greatly improves in different aspects of a whole image. For each color channel (blue, green, and red), the PSNR increases by 2.33 dB, 2.20 dB, and 2.25 dB. The MSE for the CNN model reconstruction improves by 40.6% rather than the baseline. For similarity, the 95% DSSIM advances 18.2% for our model. For the three-color difference, 95% ΔE for our model improves 30.7%.

Rather than the whole visible spectrum image reconstruction quality, we are more interested in the reconstruction quality of small image patches with rich colors and details. We crop one 80x80 pixels patch from the image to analyze, the façade of the building wall as shown in (Fig.2B). For the façade of the building reconstruction quality, our model has a 6.96 dB better PSNR than the baseline. The MSE of our model's output images for three channels average drops 79.5%. For the dissimilarity index, the 95% DSSIM for our model improves 62.3%. The 95% ΔE of color difference shows 61.5% improvements. The bilinear reconstructed image shows plenty of zig-zags at the edges and our model's demosaicing façade is much smoother on those edges. Our model's demosaicing model peaks its performance at the still images with high frequency.

4. EVALUATION OF DUAL-MODE VISIBLE/NIR IMAGE DEMOSAICING

Expanding our model's application to multi-spectrum imaging, we implement our demosaicing model on the 20 resultants still images collected in.[25] Those images showcase the buildings and landmarks on the campus of the University of Illinois at Urbana-Champaign. These images are taken by a custom camera with either a short-pass filter to block the NIR spectrum and a long-pass filter to block the Visible spectrum. The custom camera is an unmosaiced camera with three-photon detectors under each pixel, which could create ground truth unmoaiced images in both Visible and NIR spectrum while mounting the short-pass and long-pass filters with cut-off and cut-on frequency 700nm.

In comparison to the image reconstruction quality achieved by bilinear interpolation in the visible spectrum, our CNN model exhibits significant improvements across various aspects when applied to the entire image. Specifically, for each of the color channels (blue, green, and red), there is an increase in PSNR of 2.41 dB, 2.50 dB, and 2.50 dB. The CNN model's reconstruction yields a 41.4% reduction in MSE compared to the baseline. In terms of similarity metrics, our model demonstrates notable advancements. The 95% DSSIM increases by

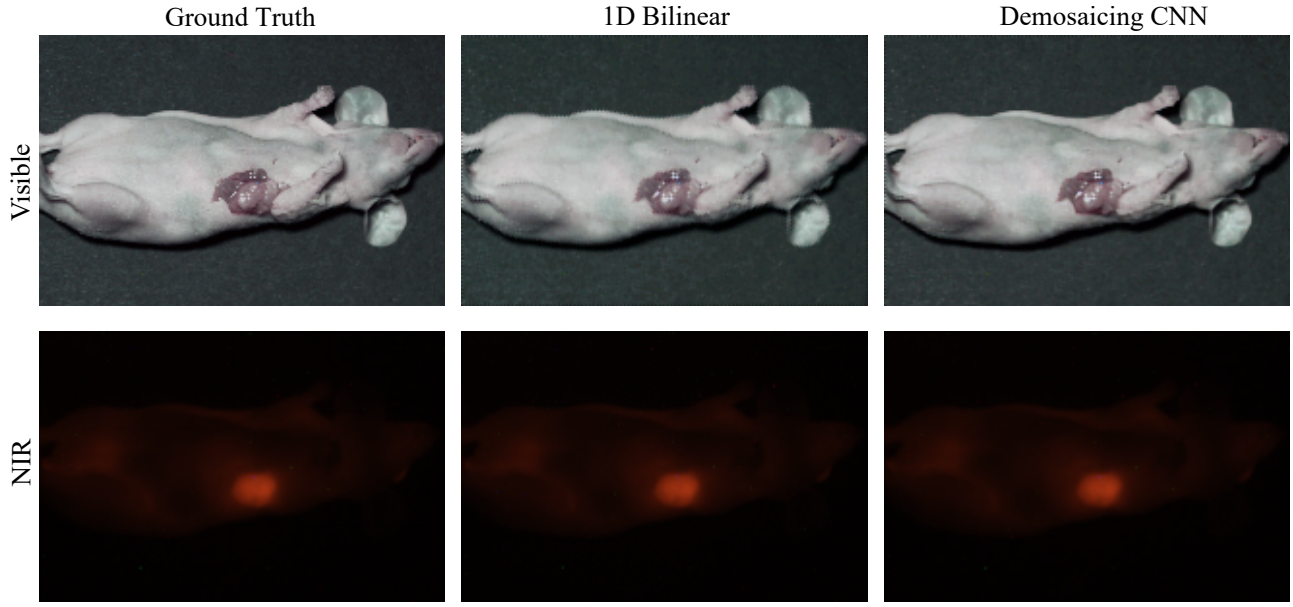

Figure 3. Preclinical mice visible and NIR fluorescence images reconstructed results comparing 1D bilinear and demosaicing CNN.

24.0% for our model compared to the baseline. Additionally, when evaluating the color differences in the three visible channels. The 95% ΔE shows a 14.2% improvement in favor of our model. When compared to the image reconstruction quality of bilinear interpolation in the NIR spectrum of the dual-mode campus images, for each of the NIR channels, we observe a respective PSNR increase of 1.80 dB, 1.67 dB, and 1.37 dB. Furthermore, the CNN model's reconstruction results in a substantial 33.0% reduction in Mean Squared Error (MSE) compared to the baseline. The 95% DSSIM experiences a 10.8% improvement for our model. Considering the color differences in the NIR channels, 95% ΔE of the reconstructed image indicates 16.6% enhancement.

5. RESULT OF PRECLINICAL AND CLINICAL DATASET

We also examine our model on a public preclinical dataset.[26] The dataset contains 3 pairs of unmosaiced visible and NIR in-vivo images of female mice with breast tumors (4T1, American Type Culture Collection; delivered via subcutaneous injection and grown to 1 cm diameter). To provide NIR fluorescence on the breast tumors, the three mice are injected IRDye 800CW Maleimide (100 μL at 11.91 μg per mL phosphate-buffered saline) into the retro-orbital sinus and allowed to accumulate in the tumor over a 24-h period. The images are collected by a custom camera with three stacked photon detectors mounting with an excitation filter to block the NIR excitation light. The visible images is taken under a white light source illumination and the NIR images are taken under an infrared source (I0785MU6000M4S, Innovative Photonic Solutions) illumination.

We implement our model on the three pairs of mice with breast tumor images and discover our CNN model reveals significant improvements in several key metrics for both visible and NIR images. For each color channel, the PSNR increases by 2.15 dB, 2.26 dB, and 2.00 dB. The MSE for the CNN model reconstruction improves by 34.7% rather than the baseline. For similarity, the average DSSIM decreased by 37.0% for our model. For the three-color difference, ΔE for our model improves by 13.6%. Reconstructing the image in the NIR spectrum, for each NIR channel the PSNR increases by 78 dB, 0.17 dB, and 0.11 dB. The average MSE for the CNN model reconstruction improves by 10.1% rather than the baseline. For dissimilarity, the 95% DSSIM decreased by 7.0% of our reconstruction. For the NIR color difference, 95% ΔE for our model improves 0.2%.

Clinical data were collected using our hexachromatic image sensor in two distinct scenarios (Fig.4). In the first scenario, during breast cancer surgery, patients received a peritumoral injection of indocyanine green (ICG) to delineate sentinel lymph nodes. The sensor was mounted above the surgical field, enabling real-time capture of both color and NIR fluorescence videos as surgeons performed lymph node mapping. In the second scenario,

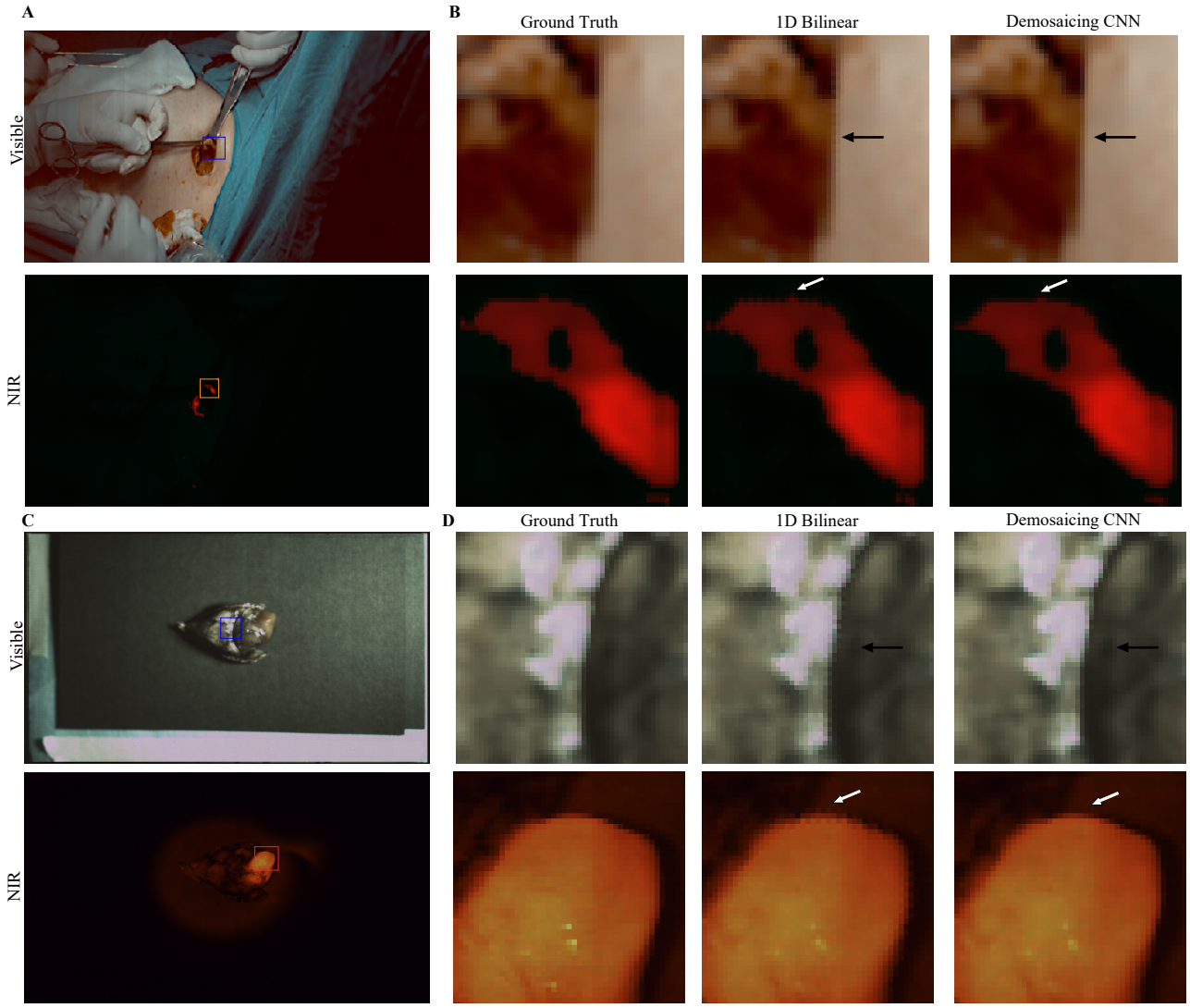

Figure 4. (A-B) Color and NIR images captured *in vivo* during breast cancer surgery in the operating room. (C-D) Color and NIR images captured *ex vivo* on the backtable during lung cancer surgery. Arrows indicate zipper artifacts at the image edges, which are prominent when processed with bilinear interpolation but significantly reduced using our CNN model.

lung cancer patients undergoing surgical procedures received VGT-309, a cathepsin-targeting ICG agent designed to illuminate tumor microenvironments through fluorescence in regions with heightened cathepsin activity. Ex vivo tissues were subsequently imaged directly in the operating room using our hexachromatic sensor, allowing for detailed observation of tumor-specific fluorescence signals. Figure 4 presents the original images, those processed via bilinear interpolation, and images enhanced using CNN-based interpolation. The CNN-enhanced images exhibit sharper edges and significantly reduced zipper artifacts compared to bilinear interpolation, as highlighted by the arrows in the figure.

6. CONCLUSION

In this paper, we demonstrated a deep learning-based model for a bio-inspired visible-NIR camera demosaicing. Each pixel of the camera has three stacked photon detectors and is mounted by a pixelated checkboard visible-NIR spectrum filter array. Compared to the bilinear routine for demosaicing, our model's reconstruction quality

of the campus building images has around 2.5 dB better performance in the visible spectrum and 1.6 dB better in the NIR spectrum.

This model shows its capability in preclinical image demosaicing which has an average 2.1 dB better PSNR than the baseline for the visible spectrum. The quality of the tumor and mice edge reconstruction is advanced with fewer mosaiced spots. However, the NIR images of mice with breast cancers have quite low frequency, which is not ideal for evaluating our model's demosaicing in the preclinical field for the NIR range. We also see reconstruction quality advances in the clincial dataset for both in vivo and ex vivo. The 1D bilinear reconstruction's zig-zags defects on the tumor edges have been avoided through our CNN model for both visible and NIR images.

REFERENCES

[1] Siegel, R. L., Miller, K. D., Fuchs, H. E., Jemal, A., et al., "Cancer statistics, 2021," *Ca Cancer J Clin* **71**(1), 7–33 (2021).

[2] Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F., "Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians* **71**(3), 209–249 (2021).

[3] Crosby, D., Bhatia, S., Brindle, K. M., Coussens, L. M., Dive, C., Emberton, M., Esener, S., Fitzgerald, R. C., Gambhir, S. S., Kuhn, P., et al., "Early detection of cancer," *Science* **375**(6586), eaay9040 (2022).

[4] Wyld, L., Audisio, R. A., and Poston, G. J., "The evolution of cancer surgery and future perspectives," *Nature reviews Clinical oncology* **12**(2), 115–124 (2015).

[5] Sullivan, R., Alatise, O. I., Anderson, B. O., Audisio, R., Autier, P., Aggarwal, A., Balch, C., Brennan, M. F., Dare, A., D'Cruz, A., et al., "Global cancer surgery: delivering safe, affordable, and timely cancer surgery," *The lancet oncology* **16**(11), 1193–1224 (2015).

[6] Hong, G., Antaris, A. L., and Dai, H., "Near-infrared fluorophores for biomedical imaging," *Nature biomedical engineering* **1**(1), 0010 (2017).

[7] Chen, Y., Wang, S., and Zhang, F., "Near-infrared luminescence high-contrast in vivo biomedical imaging," *Nature Reviews Bioengineering* **1**(1), 60–78 (2023).

[8] DSouza, A. V., Lin, H., Henderson, E. R., Samkoe, K. S., and Pogue, B. W., "Review of fluorescence guided surgery systems: identification of key performance capabilities beyond indocyanine green imaging," *Journal of biomedical optics* **21**(8), 080901–080901 (2016).

[9] Hollon, T., Jiang, C., Chowdury, A., Nasir-Moin, M., Kondepudi, A., Aabedi, A., Adapa, A., Al-Holou, W., Heth, J., Sagher, O., et al., "Artificial-intelligence-based molecular classification of diffuse gliomas using rapid, label-free optical imaging," *Nature medicine* **29**(4), 828–832 (2023).

[10] Van Dam, G. M., Themelis, G., Crane, L. M., Harlaar, N. J., Pleijhuis, R. G., Kelder, W., Sarantopoulos, A., De Jong, J. S., Arts, H. J., Van Der Zee, A. G., et al., "Intraoperative tumor-specific fluorescence imaging in ovarian cancer by folate receptor-α targeting: first in-human results," *Nature medicine* **17**(10), 1315–1319 (2011).

[11] Sega, E. I. and Low, P. S., "Tumor detection using folate receptor-targeted imaging agents," *Cancer and Metastasis Reviews* **27**, 655–664 (2008).

[12] Bou-Samra, P., Muhammad, N., Chang, A., Karsalia, R., Azari, F., Kennedy, G., Stummer, W., Tanyi, J., Martin, L., Vahrmeijer, A., et al., "Intraoperative molecular imaging: 3rd biennial clinical trials update," *Journal of biomedical optics* **28**(5), 050901–050901 (2023).

[13] Fossum, E. R. and Hondongwa, D. B., "A review of the pinned photodiode for ccd and cmos image sensors," *IEEE Journal of the electron devices society* (2014).

[14] Dillon, P. L., Lewis, D. M., and Kaspar, F. G., "Color imaging system using a single ccd area array," *IEEE Journal of Solid-State Circuits* **13**(1), 28–33 (1978).

[15] Frangioni, J. V., "In vivo near-infrared fluorescence imaging," *Current opinion in chemical biology* **7**(5), 626–634 (2003).

[16] Troyan, S. L., Kianzad, V., Gibbs-Strauss, S. L., Gioux, S., Matsui, A., Oketokoun, R., Ngo, L., Khamene, A., Azar, F., and Frangioni, J. V., "The flare™ intraoperative near-infrared fluorescence imaging system: a first-in-human clinical trial in breast cancer sentinel lymph node mapping," *Annals of surgical oncology* **16**, 2943–2952 (2009).

[17] Zhu, B. and Sevick-Muraca, E., "A review of performance of near-infrared fluorescence imaging devices used in clinical studies," *The British journal of radiology* **88**(1045), 20140547 (2015).

[18] Mondal, S. B., Gao, S., Zhu, N., Liang, R., Gruev, V., and Achilefu, S., "Real-time fluorescence image-guided oncologic surgery," *Advances in cancer research* **124**, 171–211 (2014).

[19] Frangioni, J. V., "New technologies for human cancer imaging," *Journal of clinical oncology* **26**(24), 4012 (2008).

[20] Garcia-Allende, P. B., Glatz, J., Koch, M., Tjalma, J. J., Hartmans, E., Van Scheltinga, A. G. T., Symvoulidis, P., van Dam, G. M., Nagengast, W. B., and Ntziachristos, V., "Towards clinically translatable nir fluorescence molecular guidance for colonoscopy," *Biomedical optics express* **5**(1), 78–92 (2014).

[21] Garcia, M., Edmiston, C., York, T., Marinov, R., Mondal, S., Zhu, N., Sudlow, G. P., Akers, W. J., Margenthaler, J., Achilefu, S., et al., "Bio-inspired imager improves sensitivity in near-infrared fluorescence image-guided surgery," *Optica* **5**(4), 413–422 (2018).

[22] Blair, S., Garcia, M., Davis, T., Zhu, Z., Liang, Z., Konopka, C., Kauffman, K., Colanceski, R., Ferati, I., Kondov, B., et al., "Hexachromatic bioinspired camera for image-guided cancer surgery," *Science Translational Medicine* **13**(592), eaaw7067 (2021).

[23] George, M. B., Lew, B., Blair, S., Zhu, Z., Liang, Z., Srivastava, I., Chang, A., Choi, H., Kim, K., Nie, S., et al., "Bioinspired color-near infrared endoscopic imaging system for molecular guided cancer surgery," *Journal of biomedical optics* **28**(5), 056002 (2023).

[24] Bayer, B., "Color imaging array," *United States Patent, no. 3971065* (1976).

[25] Blair, S. and Gruev, V., "Interpolant-based demosaicing routines for dual-mode visible/near-infrared imaging systems," *Optics Express* **30**(19), 34201–34217 (2022).

[26] Blair, S., Garcia, M., Zhu, Z., Liang, Z., Lew, B., George, M., Kondov, B., Stojanoski, S., Todorovska, M. B., Miladinova, D., et al., "Decoupling channel count from field of view and spatial resolution in single-sensor imaging systems for fluorescence image-guided surgery," *Journal of biomedical optics* **27**(9), 096006–096006 (2022).

[27] Menon, D. and Calvagno, G., "Color image demosaicking: An overview," *Signal Processing: Image Communication* **26**(8-9), 518–533 (2011).

[28] Mairal, J., Elad, M., and Sapiro, G., "Sparse representation for color image restoration," *IEEE Transactions on image processing* **17**(1), 53–69 (2007).

[29] Krizhevsky, A., Sutskever, I., and Hinton, G. E., "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems* **25** (2012).

[30] Simonyan, K. and Zisserman, A., "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556* (2014).

[31] Syu, N.-S., Chen, Y.-S., and Chuang, Y.-Y., "Learning deep convolutional networks for demosaicing," *arXiv preprint arXiv:1802.03769* (2018).

[32] Ma, K., Duanmu, Z., Wu, Q., Wang, Z., Yong, H., Li, H., and Zhang, L., "Waterloo exploration database: New challenges for image quality assessment models," *IEEE Transactions on Image Processing* **26**(2), 1004–1016 (2017).

[33] Jin, Y., Kondov, B., Kondov, G., Singhal, S., Nie, S., and Gruev, V., "Convolutional neural network advances in demosaicing for fluorescent cancer imaging with color–near-infrared sensors," *Journal of biomedical optics* **29**(7), 076005–076005 (2024).