Multi-modal Image Fusion for Multispectral Super-resolution in Microscopy

Neel Dey[a], **Shijie Li**[a], **Katharina Bermond**[b], **Rainer Heintzmann**[c,d], **Christine A. Curcio**[e], **Thomas Ach**[b], **Guido Gerig**[a]

[a]Department of Computer Science and Engineering, New York University Tandon School of Engineering, NY, USA.

[b]Department of Ophthalmology, University Hospital Würzburg, Würzburg, Germany.

[c]Institute of Physical Chemistry, Friedrich-Schiller-University Jena, Germany.

[d]Leibniz Institute of Photonic Technology (IPHT), Jena, Germany.

[e]Department of Ophthalmology and Visual Sciences, University of Alabama at Birmingham, AL, USA.

Abstract

Spectral imaging is a ubiquitous tool in modern biochemistry. Despite acquiring dozens to thousands of spectral channels, existing technology cannot capture spectral images at the same spatial resolution as structural microscopy. Due to partial voluming and low light exposure, spectral images are often difficult to interpret and analyze. This highlights a need to upsample the low-resolution spectral image by using spatial information contained in the high-resolution image, thereby creating a fused representation with high specificity both spatially and spectrally. In this paper, we propose a framework for the fusion of co-registered structural and spectral microscopy images to create super-resolved representations of spectral images. As a first application, we super-resolve spectral images of retinal tissue imaged with confocal laser scanning microscopy, by using spatial information from structured illumination microscopy. Second, we super-resolve mass spectroscopic images of mouse brain tissue, by using spatial information from high-resolution histology images. We present a systematic validation of model assumptions crucial towards maintaining the original nature of spectra and the applicability of super-resolution. Goodness-of-fit for spectral predictions are evaluated through functional R^2 values, and the spatial quality of the super-resolved images are evaluated using normalized mutual information.

Keywords

Corresponding author: Neel Dey. neel.dey@nyu.edu.

1. INTRODUCTION

The chemical nature of a biological sample is often investigated through various microscopic spectral imaging modalities such as fluorescence, mass spectroscopy, and others.[1] These methodologies allow for the spatial localization of bio-chemical sources which may be crucial to specific applications. However, they may have poor spatial resolution due to physics-based limitations, such as spreading a relatively small number of photons over multiple spectral channels for example. Conversely, high resolution structural modalities provide unparalleled spatial detail but can lack spectral specificity. This has led to several lines of research attempting to super-resolve the spectral images by leveraging spatial information in a co-registered structural modality.

Image fusion has been studied extensively in the satellite imaging community, [2] fusing information from cameras of high spatial resolution and high spectral resolution. The medical image analysis literature focuses largely on transferring details between structural modalities measuring disparate properties. This can be either macroscopic such as with MR and CT,[3] or microscopic such as with scanning acoustic microscopy and histology.[4] However, there has been a recent emergence of work for the fusion of microscopic spectral images and structural images. Van de Plas, et al.[5] fuse imaging mass spectrometry (IMS) and high-resolution color histology (H&E) images by using multivariate regression and feature engineering. This method uses spectral channels with high SNR only, and cannot accommodate single channel structural images. Vollnhals, et al.[6] combine electron microscopy with IMS by creating image pyramids and injecting the high-frequency structural image components into the spectral image pyramid. This can cause unwanted artifacts when combining images of different modalities, as reported by the authors.

In this paper, we propose a framework for the super-resolution of multispectral images by using a blind linear degradation model from remote sensing which uses information in a corresponding co-registered structural image. The adopted framework has a secondary denoising effect. We show the potential of the proposed framework to two challenging biomedical applications, where both spatial and spectral information are important. To test the assumptions of the linear degradation model, we validate the predictions of the model using functional R^2 values and perform a residual analysis, finding a high degree of agreement with observed data. As the parameters of the blind model are unintuitive to tune by hand for disparate imaging modalities, we perform a principled blind parameter search using Bayesian optimization.[7] The spatial quality of the resulting images is evaluated by using normalized mutual information with respect to the structural image.

2. MATERIALS AND METHODS

2.1 Data Description

In our first dataset, human tissue from the retinal pigment epithelium was imaged by two different fluorescence microscopy modalities excited at 488 nm. The first is Structured Illumination Microscopy (SIM), a form of 3D fluorescence microscopy with twice the resolution of diffraction-limited instruments. The second is confocal Laser Scanning Microscopy (LSM), which in addition to volumetric data, can capture 24 spectral channels

over the visible spectrum for each voxel, thus acquiring 4D data. The first and last spectral channels of the LSM data are discarded due to noise. No external fluorescent label was added, and thus these images correspond to autofluorescence measurements. For more details, see Ach, et al.[8] In this study, we focus on spectral super-resolution at the slice level, leaving a fully 3D treatment as future work.

For the second experiment, we use cropped sample data provided by Van de Plas, et al.[5] This consists of a high-resolution RGB image of a mouse brain H&E section and a corresponding IMS image. The IMS image is 20 times smaller than the RGB image, with 7945 spectral channels corresponding to mass over charge (*m/z*) bins.

2.2   Image Registration

For LSM-SIM registration, a reference spectral channel was chosen (600 nm) from the LSM volume, and 3D affine registration with normalized cross correlation as an image matching metric was performed. Corresponding 2D slices out of 4D and 3D data are used in this paper. For IMS-histology registration, scalar images of both modalities are formed via a mean projection and registered with 2D affine registration using normalized cross-correlation. All registrations were computed with the SimpleITK library.[9] In both cases, non-rigid registration was skipped so as to not affect the estimated blurring kernels.

2.3   Fusion Methodology

We adopt the well-established linear degradation model used in remote sensing for our problem.[2] Let the multispectral image have p channels and m total pixels in each channel. Let the high-resolution image have c channels and n pixels in each channel. Each channel in both modalities is vectorized, thereby converting an image with two dimensions of space and one of spectra into a matrix. The model is as follows,

$$\mathbf{Y_M} = \mathbf{Z}BS + \mathbf{N_M}$$
$$y_P = rZ + n_P$$

where,

$\mathbf{Y_M} \in \mathbb{R}^{p \times m}$ is the original spectral image,

$\mathbf{Z} \in \mathbb{R}^{p \times n}$ is the target fused image,

$\mathbf{B} \in \mathbb{R}^{n \times n}$ is the blurring matrix,

$\mathbf{S} \in \mathbb{R}^{n \times m}$ is a downsampling matrix,

$\mathbf{y_P} \in \mathbb{R}^{c \times n}$ is the original structural image,

$\mathbf{r} \in \mathbb{R}^{c \times p}$ is the spectral response of the structural image,

$\mathbf{N_M} \in \mathbb{R}^{p \times m}$ is the reconstruction noise for the spectral image,

$\mathbf{n_P} \in \mathbb{R}^{c \times n}$ is the reconstruction noise flor the structural image .

Using the approach of Wei, et al.,[10] we break the problem of estimating **B**, **r** and **Z** for both modalities into two separate stages. First, we estimate **B**, **r** by using the method proposed by Simões, et al.[11] via the convex minimization of the following objective function,

$$\min_{\mathbf{B}, \mathbf{r}} \left\| \mathbf{r}\mathbf{Y}_M - \mathbf{y}_P \mathbf{B}\mathbf{S} \right\|_2^2 + \lambda_b \phi_b(\mathbf{B}) + \lambda_r \phi_r(\mathbf{r}),$$

where ϕ_b and ϕ_r are quadratic regularizers and λ_b and λ_r are their respective weights. **S** is easily estimated from the dimensions of the co-registered data. We then estimate the target fused image using the method proposed by Wei, et al.[12] for robust fast fusion via solving the following problem,

$$\underset{\mathbf{Z}}{\arg\min} \frac{1}{2} \left\| \mathbf{Y}_M - \mathbf{Z}\mathbf{B}\mathbf{S} \right\|_F^2 + \frac{1}{2} \left\| \mathbf{y}_P - \mathbf{r}\mathbf{Z} \right\|_2^2 + \lambda \left\| \mathbf{Z} - \overline{\mathbf{Z}} \right\|_F^2,$$

where the third term acts as a regularizer, corresponding to a Gaussian prior over the target fused image where $\bar{Z}$ is the prior mean. Importantly, these methods rely on the assumption that the target fused image **Z** lies in a low-dimensional subspace.

2.4 Bayesian Parameter Optimization

As opposed to remote sensing, we do not have any heuristics on parameter tuning in the image fusion optimization for vastly different microscopy modalities. Therefore, we take a principled approach to blind parameter selection. For both sets of images, the three regularization parameters and blurring kernel widths were tuned by using Bayesian optimization for parameter selection,[7] by maximizing the normalized mutual information (NMI) between the fused image and the high-resolution structural image. NMI is a common measure of image matching when comparing co-registered multi-modal images. By selecting parameters that maximize it, we ensure that our fused image is close to the quality of the high-resolution structural image. All results reported have had their parameters tuned in this manner, with thirty evaluations via Bayesian optimization.

2.5 Validation Methodology

A common goodness-of-fit measure is R^2. We create prediction and observation pairs by mapping each pixel of the original spectral image to the center of the corresponding patch in the generated image. As each pixel has a spectrum, we consider each pixel as an observation and each spectrum as a functional measurement. Inspired by common practice in functional data analysis,[13] we report the per-pixel R^2 measure for an pixel spectrum, defined as,

$$R_i^2 = 1 - \frac{\sum_{j=1}^p \left(\hat{y}_i(j) - y_i(j) \right)^2}{\sum_{j=1}^p \left(y_i(j) - \bar{y}_i(j) \right)^2}$$

where i is the pixel index, j is the spectral channel index, y_i is the observation, $\hat{y}_i$ is the predicted value and $\bar{y}_i$ is the mean of the observed values. We also report the per-channel R^2 as,

$$R^2(j) = 1 - \frac{\sum_{i=1}^{n}\left(\hat{y}_i(j) - y_i(j)\right)^2}{\sum_{j=1}^{n}\left(y_i(j) - \bar{y}_i(j)\right)^2}.$$

To inspect whether the model misses systematic variation, we perform a residual analysis. We finally report a qualitative measure of the super-resolved image by using NMI with respect to the high-resolution image.

3. RESULTS

3.1 LSM and SIM

A $328 \times 328 \times 22$ LSM image stack is upsampled by a co-registered 1312×1312 SIM image via the proposed image fusion framework. Using dataset 1 in Fig. 1 as an example, the average per-channel NMI between the original LSM and SIM images is 0.2205, whereas the super-resolved image has an NMI of 0.6874 indicating higher spatial similarity to the high-resolution image. We find a per-channel R^2 of 0.3853 and a per-pixel R^2 of 0.4973. While these values are lower than the subsequent experiment, this is easily explained as the original spectra are much noisier (see Fig. 1, Col. 4 for an overlay of raw spectra). Hence, as predictions are denoised, they explain a smaller portion of the total variation which consisted of the noise as well. Importantly, the mean prediction matches the mean observation (Fig. 1, Col. 4).

3.2 IMS and H&E

We super-resolve a $20 \times 20 \times 7945$ IMS image stack to a $400 \times 400 \times 7945$ image stack by using spatial information from a $400 \times 400 \times 3$ histology image. We find a pixel-wise average R^2 of 0.6943 and a channel-wise average R^2 of 0.8493, indicating excellent agreement with the data overall. The average per-channel residuals have a mean and standard deviation of -5.36×10^{-4} and 1.7×10^{-3}. The per-channel statistics of the data itself have a mean of 0.1451 and standard deviation of 0.2233, which are orders of magnitude higher than the residual statistics, signifying small error. The average per-channel NMI between the interpolated IMS and histology images is 0.2394, whereas the super-resolved image has an NMI of 0.2401. Instead of using 7945 channels, we can compute NMI on only high SNR channels.[5] This computes an NMI to the structural image of 0.4072 via cubic interpolation and 0.4322 via super-resolution, indicating higher average structural similarity with the super-resolution.

4. DISCUSSION

The presented model uses a linear approach for estimating spectral responses. It is possible that there may be non-linear models that better fit specific image modality-pairs. Future work will explore and validate the various nonlinear models that might better describe the

relationships between the specific modalities presented in this paper. Specific to IMS images, the number of features (m/z bins) exceeds the number of observations (pixels). In this case, more robust super-resolution results may be obtained by incorporating Elastic Net-style regularization[14] into the optimization process. Finally, there has been a recent emergence of deep learning based approaches for spectral image super-resolution via image fusion in remote sensing. Currently, these methods are either supervised,[15] or require a known spectral response,[16] precluding their direct application to this work.

Multispectral super-resolution via image fusion can potentially be used in biomedical research for better spatial localization of chemical signatures, past the resolving ability of spectral imaging systems. Furthermore, it can benefit image analysis pipelines by providing a high-quality reference for segmentation of features which might not be present in the structural image and vice-versa.

ACKNOWLEDGEMENTS

REFERENCES

[1]. da Cunha MML, Trepout S, Messaoudi C, Wu T-D, Ortega R, Guerquin-Kern J-L, and Marco S, "Overview of chemical imaging methods to address biological questions," Micron 84, 23–36 (2016). [PubMed: 26922256]

[2]. Loncan L, de Almeida LB, Bioucas-Dias JM, Briottet X, Chanussot J, Dobigeon N, Fabre S, Liao W, Licciardi GA, Simoes M, Tourneret J, Veganzones MA, Vivone G, Wei Q, and Yokoya N, "Hyperspectral Pansharpening: A Review," IEEE Geoscience and Remote Sensing Magazine 3(3), 27–46 (2015).

[3]. James AP et al., "Medical image fusion: A survey of the state of the art," Information Fusion (2014).

[4]. Khalilian-Gourtani A, Wang Y, and Mamou J, "Scanning Acoustic Microscopy Image Super-Resolution using Bilateral Weighted Total Variation Regularization," in [2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)], 5113–5116 (July 2018).

[5]. Van de Plas R, Yang J, Spraggins J, and Caprioli RM, "Image fusion of mass spectrometry and microscopy: a multimodality paradigm for molecular tissue mapping," Nature methods 12(4), 366 (2015). [PubMed: 25707028]

[6]. Vollnhals F, Audinot J-N, Wirtz T, Mercier-Bonin M, Fourquaux I, Schroeppel B, Kraushaar U, Lev-Ram V, Ellisman MH, and Eswara S, "Correlative microscopy combining secondary ion mass spectrometry and electron microscopy: comparison of intensity–hue–saturation and Laplacian pyramid methods for image fusion," Analytical chemistry 89(20), 10702–10710 (2017). [PubMed: 28901122]

[7]. Brochu E, Cora VM, and De Freitas N, "A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning," arXiv preprint arXiv:1012.2599 (2010).

[8]. Ach T, Hong S, Heintzmann R, Hillenkamp J, Sloan KR, Dey NS, Gerig G, Smith T, Curcio C, and Bermond K, "High-resolution and multispectral imaging of autofluorescent retinal pigment epithelium (RPE) granules," Investigative Ophthalmology & Visual Science 58(8), 3382–3382 (2017).

[9]. Yaniv Z, Lowekamp BC, Johnson HJ, and Beare R, "SimpleITK Image-Analysis Notebooks: a Collaborative Environment for Education and Reproducible Research," Journal of Digital Imaging 31(3), 290–303 (2018). [PubMed: 29181613]

[10]. Wei Q, Bioucas-Dias J, Dobigeon N, Tourneret J-Y, and Godsill S, "Blind model-based fusion of multi-band and panchromatic images," in [Multisensor Fusion and Integration for Intelligent Systems (MFI), 2016 IEEE International Conference on], 21–25, IEEE (2016).

[11]. Simões M, Bioucas-Dias J, Almeida LB, and Chanussot J, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," IEEE Transactions on Geoscience and Remote Sensing 53(6), 3373–3388 (2015).

[12]. Wei Q, Dobigeon N, Tourneret J-Y, Bioucas-Dias J, and Godsill S, "R-FUSE: Robust fast fusion of multiband images based on solving a Sylvester equation," IEEE signal processing letters 23(11), 1632–1636 (2016).

[13]. Ramsay J, "Functional data analysis," Encyclopedia of Statistics in Behavioral Science (2005).

[14]. Zou H and Hastie T, "Regularization and variable selection via the elastic net," Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67(2), 301–320 (2005).

[15]. Xie Q, Zhou M, Zhao Q, Meng D, Zuo W, and Xu Z, "Multispectral and Hyperspectral Image Fusion by MS/HS Fusion Net," arXiv preprint arXiv:1901.03281 (2019).

[16]. Qu Y, Qi H, and Kwan C, "Unsupervised Sparse Dirichlet-Net for Hyperspectral Image Super-Resolution," in [The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)], (June 2018).
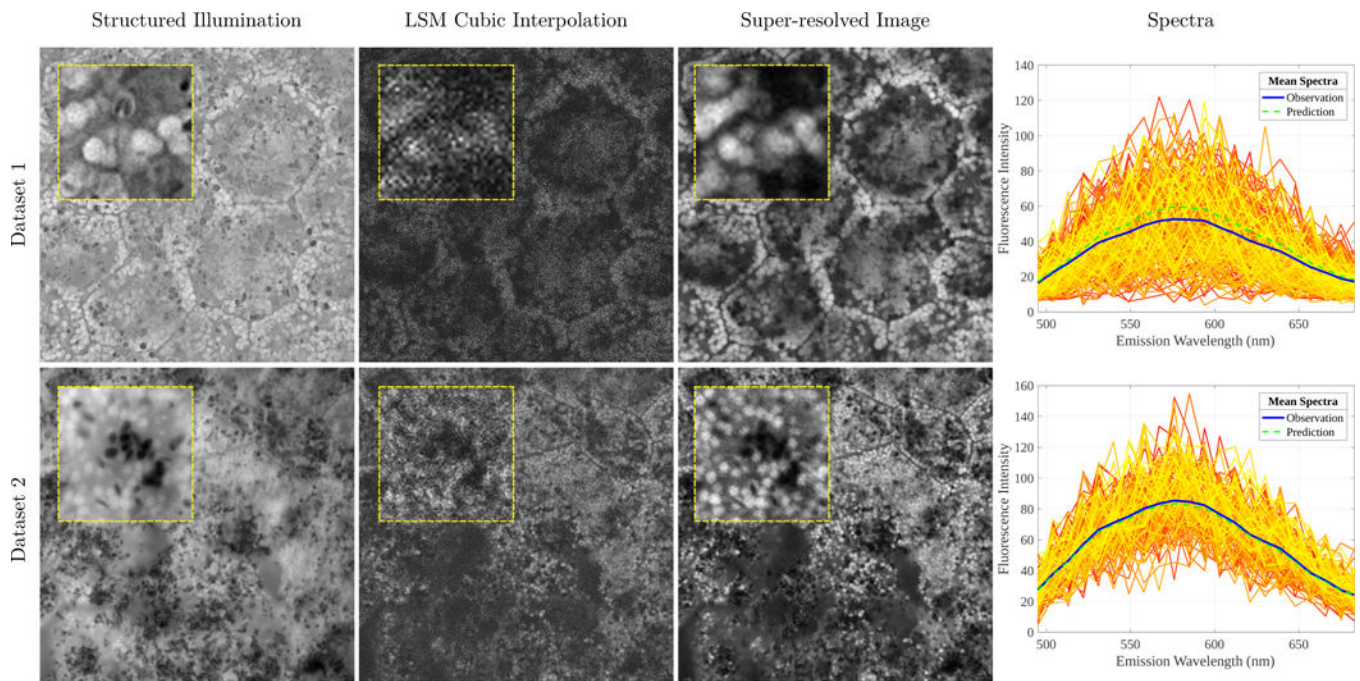
Figure 1.
Results from two datasets (rows 1 and 2) of co-registered LSM and SIM images. **Column 1:** The high-resolution SIM image. **Column 2:** A reference channel of the LSM image upsampled with cubic interpolation. **Column 3:** The super-resolved image using the proposed framework. Best case reference channels are shown. Yellow-framed insets show improved resolution using the proposed framework. **Column 4:** Raw observed spectra from segmented granules showing a high degree of noise. Their mean spectra are reported in blue and the mean prediction of the model is shown in green.
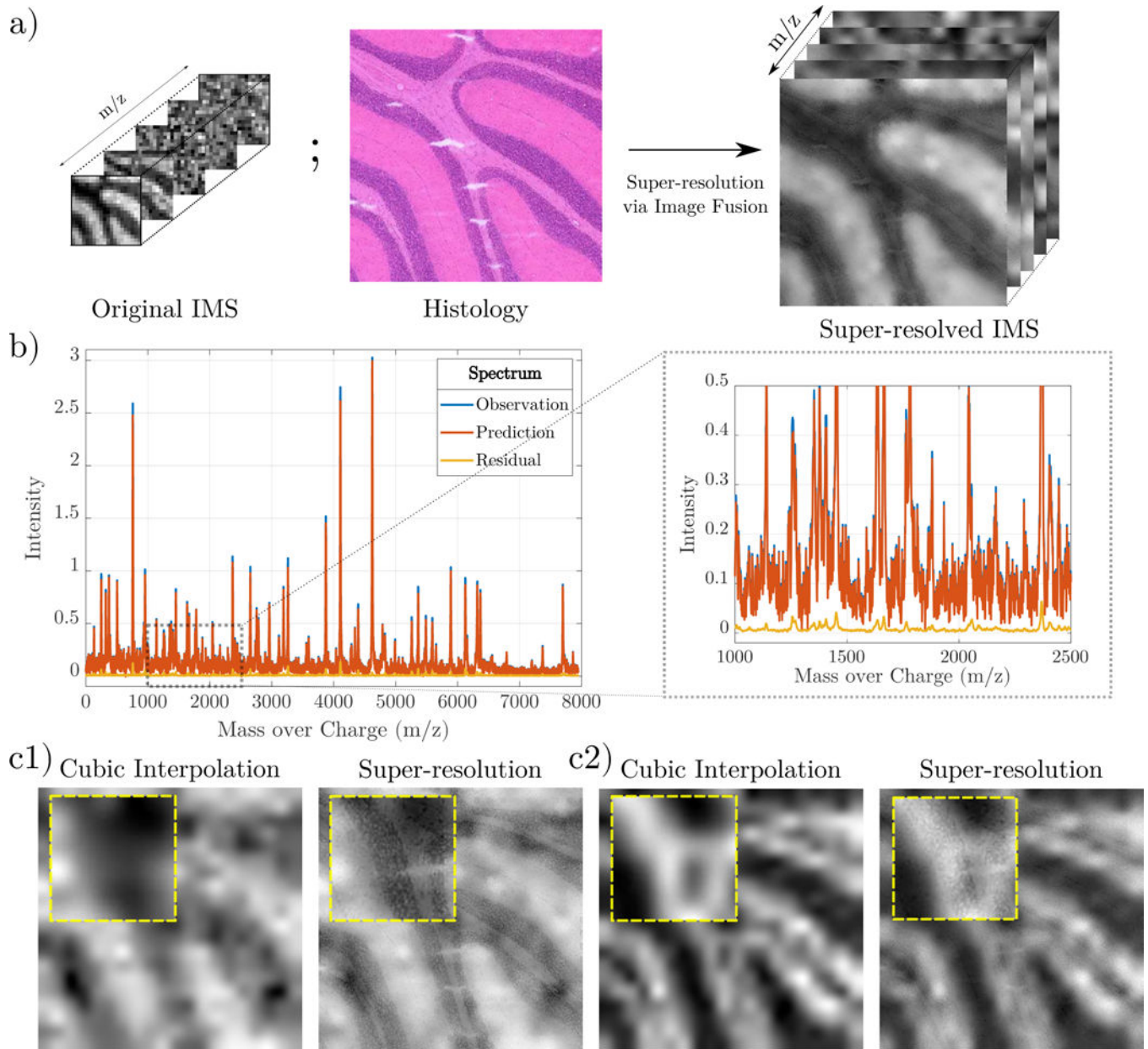
Figure 2.
a) An overview of the super-resolution pipeline. Spectral information from IMS and spatial information from H&E is used to create a super-resolved IMS image. **b)** An overlay of the original, predicted and residual spectrum for an arbitrary pixel. A zoomed section is shown in the adjacent box. Red and blue lines overlap significantly. **c1)** and **c2)** Two examples of super-resolved image channels compared to naive cubic interpolation. Zoom-boxes are shown in yellow.