BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF INFORMATION TECHNOLOGY

Excel@FIT 2026
http://excel.fit.vutbr.cz

# Intrinsic Image Component Decomposition in Scanning Electron Microscopy

Bc. Marek Konečný

**Abstract**

Scanning Electron Microscopes (SEM) are equipped with **multiple detectors** that simultaneously provide distinct modalities for observing the secondary electrons signal. While their combined use enables finer analysis of the specimen's physical properties, each detector also introduces parasitic signal components, such as shot noise and other artifacts. This work proposes a unique approach to multi-detector SEM image fusion via *intrinsic image component decomposition* using an encoder-renderer neural network tuned for the Trinity Detection System™. When applied as a denoising method, the architecture achieves a mean PSNR gain of 6.9 dB at SSIM of 0.726. The proposed pipeline shows potential to enable automated surface analysis and to enhance the interpretability of low-SNR draft scans.

*xkonec86@stud.fit.vut.cz, *Faculty of Information Technology, Brno University of Technology*

## 1. Introduction

Primary motivation for fusing multimodal SEM imagery is to reduce the cognitive load caused by presenting a large volume of raw data [1]. Baseline per-pixel fusion methods are known, e.g., average-intensity mixing, pixel-selection heuristics, etc. Although analytically interpretable and generally enhancing certain signal characteristics (e.g., SNR), they leave no control over qualitative image characteristics, inadvertently suppressing detector-specific features by aggregating their signals.

This work proposes a new multimodal SEM fusion pipeline with fine, yet physically bounded control over the qualitative image characteristics. Instead of aggregating multimodal information on a per-pixel basis, the proposed encoder-renderer neural network predicts intrinsic image components related to the specimen's physical properties and recombines them during reconstruction, while excluding unwanted parasitic signal components. The proposed architecture and training pipeline operate end-to-end, enabling self-supervised training on unannotated data.

## 2. Existing Related Work

Despite abundant multimodal image fusion solutions, literature specific to SEM remains sparse. Publication [2], which inspired this work, approaches multimodal SEM image analysis as a principal component analysis (PCA) problem, decomposing the 4 input channels into underlying signal components:

$$\mathbf{I}_i = \mathbf{I}_{clean,i} + \mathbf{r}_i, \quad \mathbf{I}_{clean,i} = \mathbf{c}\,(1 + \vec{p}_i \cdot \nabla\mathbf{h}),$$

where $I_i$ is one channel of the $i$-modal SEM image, $r_i$ is the detector-specific parasitic residual component, $\vec{p}_i$ is the detector-response vector (characterizing position of the detector inside the chamber and its sensitivity to the secondary-electrons signal), and $\mathbf{c}$ and $\nabla\mathbf{h} = \left(\frac{\partial \mathbf{h}}{\partial x}, \frac{\partial \mathbf{h}}{\partial y}\right)$ are the compositional (material) contrast map and surface topography of the specimen, respectively; $\nabla\mathbf{h}$ is gradient of the specimen surface height function $\mathbf{h}$ in coordinate system $(x, y)$.

Although the authors of [2] presented promising results, their method is tightly coupled to their specific imaging setup (4 symmetrically positioned BSE[1] detectors), and it is unclear how it performs in more challenging imaging scenarios with real samples.

## 3. Proposed Approach

The proposed method is designed for the Trinity Detection System™ of Thermo Fisher Scientific, which consists of 3 detectors of mixed BSE and SE[2] sensitivities. First, it was necessary to define an extended

[1] *BSE* stands for Back-Scattered Electrons.
[2] *SE* stands for Secondary Electrons.

SEM image model, constituting the **Differentiable Renderer** (DR) module of the proposed architecture:

$$\begin{aligned}\mathbf{I}_i &= g_i\big((1-\alpha_i)\mathbf{S}_{\mathrm{BSE,i}} + \alpha_i\mathbf{S}_{\mathrm{SE,i}} + \mathbf{r}_{\mathrm{lf,i}}\big) + \mathbf{r}_{\mathrm{hf,i}}\\ \mathbf{S}_{\mathrm{BSE,i}} &= \mathbf{c}\,(\vec{p}_i \cdot \vec{\mathbf{n}}),\\ \mathbf{S}_{\mathrm{SE,i}} &= \big(w_i\,\mathbf{c} + b_i\big)\,\exp\big(\beta_i(1-\mathbf{n}_z)\big),\end{aligned}$$

where $\alpha_i$, $w_i$, $b_i$ and $\beta_i$ are learned per-detector parameters, $\mathbf{r}_{\mathrm{lf,i}}$ and $\mathbf{r}_{\mathrm{hf,i}}$ are inferred per-detector low-frequency and high-frequency residual signal components, respectively; $\vec{\mathbf{n}} = (\mathbf{n}_x, \mathbf{n}_y, \mathbf{n}_z)$ is the specimen surface normal map and $g_i$ is inferred image gain. The $\mathbf{S}_{\mathrm{BSE,i}}$ and $\mathbf{S}_{\mathrm{SE,i}}$ maps then denote the BSE- and SE-sensitive shading. Since the $\mathbf{r}_{\mathrm{hf,i}}$ component should primarily model the electron shot noise, which is expected to be Poisson-like [3], it is comprised of predicted *white-noise* map $\boldsymbol{\eta}_i$ scaled by the noise-free image intensity at a given point:

$$\mathbf{r}_{\mathrm{hf,i}} = \boldsymbol{\eta}_i \sqrt{g_i\big((1-\alpha_i)\mathbf{S}_{\mathrm{BSE,i}} + \alpha_i\mathbf{S}_{\mathrm{SE,i}} + \mathbf{r}_{\mathrm{lf,i}}\big)}.$$

### 3.1 Intrinsic Image Decomposition Module

Extraction of the intrinsic image components themselves takes place in the **Intrinsic Image Decomposition** (IID) module, depicted in the diagram in Figure 1. In its early stages, the module re-encodes the 3-channel input to a 96-channel latent-space tensor. Finally, the tensor is processed by the final *convolutional heads*: the $\mathbf{r}_{\mathrm{hf}}$ and $\mathbf{r}_{\mathrm{lf}}$ heads isolate the residual signal components, while the specimen surface head extracts the surface compositional map $\mathbf{c}$ and surface normals map $\vec{\mathbf{n}}$.

### 3.2 Training the Model

Due to the unavailability of ground-truth data, the model was trained end-to-end in a self-supervised manner, with the primary loss function component being the **absolute error** of the channel $\mathbf{I}_i$ **reconstruction**. Other (minimized) regularization parameters include:

$\|\nabla\mathbf{c}\|^2$ – Minimizes high-frequency variation in $\mathbf{c}$.
$\|\nabla\vec{\mathbf{n}}\|^2$ – Minimizes high-frequency variation in $\vec{\mathbf{n}}$.
$\mathbf{n}_x + \mathbf{n}_y$ – Enforces minimal overall surface tilt.
$\|\mathbf{r}_{\mathrm{lf}}\|$ – $L_1$ regularization over low-frequency residuals.
**Autocorrelation of $\boldsymbol{\eta}$** – Enforces "noise-whiteness".
**Inter-detector correlation of $\boldsymbol{\eta}$** – Enforces noise independence across detectors.

Supplemental ground-truth surface normals $\vec{\mathbf{n}}_{\mathrm{gt}}$ were also synthesized from samples containing spherical powder particles (see poster Section "Training Ground-Truth Synthesis"). Where available, this synthetic data was used to guide the extraction of the surface normals map $\vec{\mathbf{n}}$ inside the IID module via masked mean cosine distance: $1 - \vec{\mathbf{n}} \cdot \vec{\mathbf{n}}_{\mathrm{gt}}$.

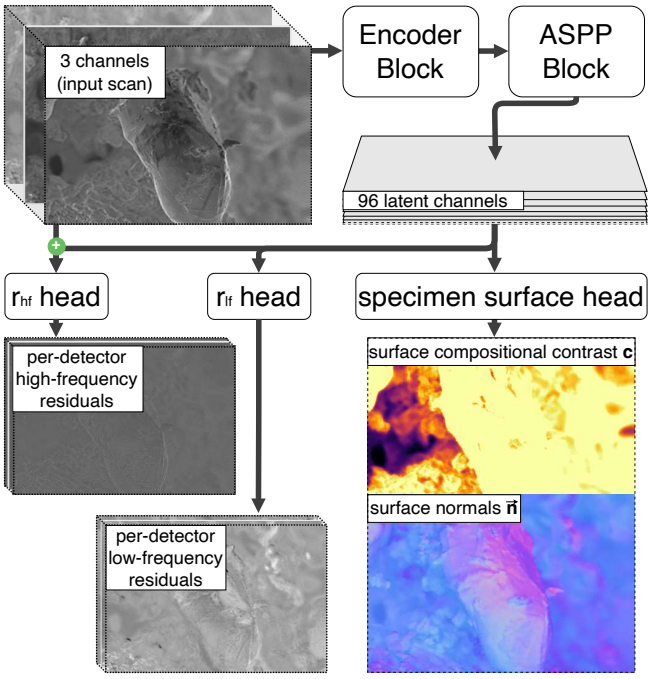


**Figure 1.** Top-level diagram of the IID module.

Such a pre-trained model was then fine-tuned in the final pseudo-labeling phase to improve residual components isolation: scan triplets capturing an identical sample were identified inside the dataset, their registration transformations were calculated, and the IDD module was guided to reconstruct the previously inferred $\mathbf{c}$ and $\vec{\mathbf{n}}$ maps of the scan triplet with best SNR estimate [4] of the group. Both training phases ran for 4000 epochs, where each epoch consisted of 4000 sampled patches of size $256 \times 256$ px.

## 4. Evaluation and Conclusions

On GPU, sample inference takes approximately 1.5 s and in a controlled environment, reconstructions with the $\mathbf{r}_{\mathrm{hf,i}}$ component excluded achieve a mean PSNR gain of 6.9 dB at SSIM of 0.726 relative to synthesized ground-truth high-SNR images. The primary mode of evaluation is ongoing subjective quality assessment via trials with public participants and Thermo Fisher Scientific personnel.

Although this work does not aim to replace fundamental high-quality SEM imaging practices (e.g., higher dwell times), its value lies in the potential for automated specimen surface analysis or to enhance the interpretability of rapid draft scans, which are inherently limited by low signal-to-noise ratios. A primary direction for future development would be to expand the training dataset and to synthesize more comprehensive annotations, thereby accelerating the performance of the IID module.

## References


[1] Elizabeth L. Fox and Joseph W. Houpt. The perceptual processing of fused multi-spectral imagery. *Cognitive Research: Principles and Implications*, 1(1):31, dec 2016.

[2] Jan Neggers, Eva Héripré, Marc Bonnet, Denis Boivin, Alexandre Tanguy, Simon Hallais, Fabrice Gaslain, Elodie Rouesne, and Stéphane Roux. Principal image decomposition for multi-detector backscatter electron topography reconstruction. *Ultramicroscopy*, 227:113200, aug 2021.

[3] F. Timischl, M. Date, and S. Nemoto. A statistical model of signal–noise in scanning electron microscopy. *Scanning*, 34(3):137–144, 2012.

[4] Dominic Chee Yong Ong and Kok Swee Sim. Single image signal-to-noise ratio (snr) estimation techniques for scanning electron microscope: A review. *IEEE Access*, 12:155747–155772, 2024.