

走査電子顕微鏡におけるライブ像のノイズ低減

Noise reduction of live image in scanning electron microscope

植松 文徳^{*1}
Fuminori Uematsu

武井 雅彦^{*1}
Masahiko Takei

吉田 光良^{*1}
Mitsuyoshi Yoshida

^{*1} 日本電子株式会社
JEOL Ltd.

A real time display called a live image is used to search for a region to be observed, when observing a sample with a scanning electron microscope (SEM). This live image is usually an image with high noise and poor visibility. This makes it difficult to find an appropriate observation area. Therefore, in this research, a noise reduction model by deep neural network was created with the aim of improving the visibility of live images. We incorporated this model into the SEM and succeeded in obtaining a live image with less noise. This makes it possible to efficiently search for the observation region.

1. 背景と動機

走査電子顕微鏡(SEM)で試料を観察する際には、試料の観察したい領域が電子顕微鏡の表示領域(視野)に含まれるようユーザーが装置を操作する必要がある。これを視野探しと呼び、具体的には試料ステージの移動や観察倍率の変更などを行う。視野の画像がリアルタイムで表示されるので、視野探しで表示される像をライブ像と呼ぶ。

ライブ像では視野探しの操作に対して十分な応答性が得られるよう、高いフレームレートを使用するのが一般的である。すなわち、電子ビームの走査が速く、検出器で観測される時間当たりの電子が少ない。そのため検出信号にノイズが発生しやすい。また、応答性が落ちるという理由から、ライブ像では複数フレームの積算によるノイズ除去を無効にすることも多い。その結果、ライブ像はノイズの多い画像となるのが常態である。

SEMの熟練ユーザーが扱い慣れた試料を観察する場合であれば、ノイズの多いライブ像であっても、ノイズの無い状態にある程度推測しながら視野探しを行える。しかし、そのような推測は経験の浅いユーザーにとって困難であり、視野探しに長い時間を要したり、観察したい対象物を見落としたりする可能性がある。

ライブ像のノイズを低減できれば、視野探しをより効率的に見落としなく行うことが可能となり、高い有用性が期待できる。本研究ではライブ像のノイズ低減にディープラーニングを用いることを提案する。

2. 提案手法

2.1 モデル設計

ノイズの多い SEM の観察像をノイズの少ない観察像に変換するモデル(ノイズ低減モデル)を作成し、それを SEM に組み込んでライブ像のノイズを低減させる。SEM でノイズの少ない観察像を得るには、電子ビームの走査を遅くし、さらに複数フレームの積算によるノイズ除去を用いれば良い。これは静止画を保存するときの観察条件であり、このようにして得られる像をフォト像と呼ぶ。従って本研究ではライブ像からフォト像への変換を学習させることで、ノイズ低減モデルを作成する。

ライブ像はフォト像の一部がノイズによって隠された状態と見

なせるため、ライブ像をノイズ除去した結果は一意に定まらない。そのため Encoder-Decoder モデルを二乗誤差で学習させてノイズ低減モデルを作成しても、ライブ像から推定される多数のフォト像を平均したような、不明瞭な像が得られると予想される。そこで本研究ではノイズ低減モデルを学習させる手法として、Conditional GAN を採用した。この手法ではライブ像をフォト像に変換する生成器と、フォト像が生成された物か本物かを識別する識別器の二つのモデルを相互に競わせて学習する。すなわち、生成器は自身が変換した画像が本物であると識別器に誤認させようとする。一方、識別器は与えられたフォト像を正確に識別しようとする。その際、生成したフォト像が不明瞭であると本物ではないことが容易に判別できてしまうので、明瞭な画像を生成するように生成器の学習が進むと期待できる。

モデル設計には pix2pixHD[Wang 17]で提案されている手法を取り入れた。これは高品質で高解像度の画像が生成可能な Conditional GAN の手法であり、次のような特徴を有している。

- 階層化された生成器
- マルチスケールに対応した識別器
- 複数の誤差関数の組み合わせ

pix2pixHD ではセマンティックラベルマップから写真のようにリアルな画像を生成しているが、本研究は同様のモデルを SEM 観察像のノイズ除去に適用した。

(1) 生成器のアーキテクチャ

本研究では pix2pixHD と同様に粗い解像度の生成器(Global Generator)と密な解像度の生成器(Local Enhancer)を階層的に結合した生成器を用いた。ここでは Johnson ら[Johnson 16]や CycleGAN[Zhu 17]や pix2pixHD で用いられる表記の慣習に従って次のように表現する。

- フィルター数 k でストライド 1 の 7×7 畳み込み層-InstanceNorm[Ulyanov 16]層-ReLU 層を $c7s1-k$ と表現する
- フィルター数 k でストライド 2 の 3×3 畳み込み層-InstanceNorm 層-ReLU 層を dk と表現する
- フィルター数 k の 3×3 畳み込み層を二つ含む Residual ブロックを Rk と表現する
- フィルター数 k でストライド 1/2 の 3×3 転置畳み込み層-InstanceNorm 層-ReLU 層を uk と表現する

生成器のアーキテクチャは、Global Generator を c7s1-32, d64, d128, d256, d512, R512, R512, R512, R512, R512, R512, R512, R512, u256, u128, u64, u32, c7s1-1 とし、Local Enhancer を c7s1-16, d32, R32, R32, R32, u16, c7s1-1 とした。

(2) 識別器のアーキテクチャ

識別器においても pix2pixHD と同様に画像サイズの異なる複数の識別器を結合して用いた。識別器の個数は、Global Generator の学習では 2 個、Local Enhancer を結合した学習では 3 個とした。ここで次のように表現する。

- フィルター数 k でストライド 2 の 4×4 畳み込み層-InstanceNorm 層-LeakyReLU 層を C_k と表現する

いずれの識別器も $C32$ - $C64$ - $C128$ - $C256$ とし、その後にフィルター数 1 でストライド 1 の 4×4 畳み込み層を繋げたアーキテクチャとした。ただし、 $C32$ の InstanceNorm 層は省略した。

2.2 教師データ

教師データには同一視野を観察したライブ像とフォト像のペアが必要となるが、これを大量に取得するには長時間の SEM 操作が必要となり、容易ではない。そこで本研究ではフォト像のみを取得し、それに人工的なノイズを付加することによって、疑似ライブ像を作成して用いた。本手法であれば既に所有しているフォト像を教師データとして活用できる利点がある。SEM の二次電子像において支配的なノイズは、検出器で観測される電子がポアソン分布に従っていることに由来するショットノイズである。そこでフォト像にこのようなショットノイズを付加して疑似ライブ像を作成した。また教師データの数を増やすため、フォト像に拡張・回転・並進・反転といった画像変換を行ってデータ拡張した後に疑似ライブ像を作成した。

3. 実験

3.1 学習

対象とする画像の解像度は、日本電子製の SEM である JSM-7900F のフォト像と同じ 1280×960 ピクセルとした。教師データとして疑似ライブ像とフォト像のペアを約 2 万データ作成し、GeForce GTX 1080(NVIDIA 社製)を 1 基搭載した Ubuntu サーバーで学習を行った。Global Generator を 40 エポック学習した後、Local Enhancer を結合して 20 エポック学習した。フレームワークには Microsoft Cognitive Toolkit を使用した。

3.2 評価

学習を行って作成したノイズ低減モデルを JSM-7900F に組み込み、ノイズ低減されたライブ像を観察して効果を評価した。視野探しに必要な応答性を確保するため、SEM の制御用 PC に GeForce GTX 1080 を 1 基搭載し、画像辺りの変換時間が 100ms 以下となるようにした。

同一視野のライブ像(変換前)、ライブ像(変換後)、フォト像を図 1 に示す。ライブ像のノイズが低減され、フォト像と比較しても自然な画像となっており、視認性が向上していることがわかる。一方で、本手法ではライブ像をフレーム単位で変換しているため、各フレームのノイズの偏りが変換後の画像に僅かに影響を与える。これによって、リアルタイム表示したときに試料表面が波打って見えるなどの違和感を生じる場合があった。

4. 結論と今後の課題

本研究ではディープラーニングを用いることで走査電子顕微鏡におけるライブ像のノイズ低減が可能であることを確認し、効率的な視野探しへの寄与が期待できる結果が得られた。

今回は効果の測定を目視による主観に頼ったが、今後は非参照メトリックなどを用いた定量的な知覚品質の測定に取り組む必要がある。また、vid2vid[Wang 18]で行われているように、変換結果に時系列での一貫性を持たせるなどして、試料表面が波打って見える現象を解消することが今後の課題となる。

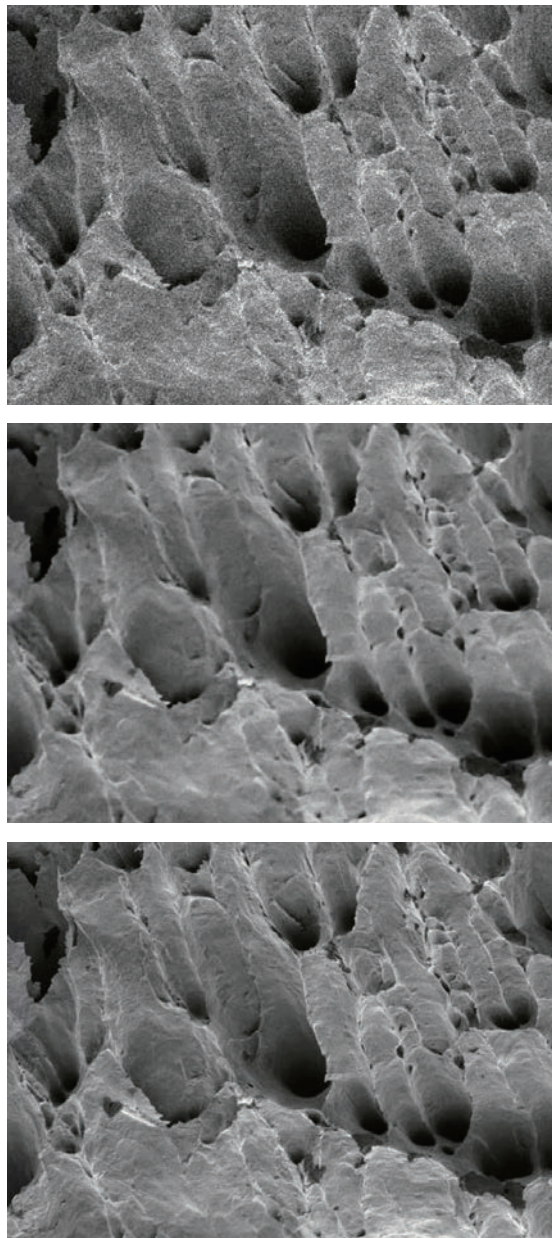


図 1 SEM ライブ像の変換例。上段:ライブ像(変換前)、中段:ライブ像(変換後)、下段:フォト像

参考文献

- [Wang 17] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz and Bryan Catanzaro: High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs, arXiv:1711.11585, 2017.

- [Johnson 16] Justin Johnson, Alexandre Alahi and Li Fei-Fei: Perceptual losses for real-time style transfer and super-resolution, arXiv:1603.08155, 2016.
- [Zhu 17] Jun-Yan Zhu, Taesung Park, Phillip Isola and Alexei A. Efros: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, arXiv:1703.10593, 2017.
- [Ulyanov 16] Dmitry Ulyanov, Andrea Vedaldi and Victor Lempitsky: Instance Normalization: The Missing Ingredient for Fast Stylization, arXiv:1607.08022, 2016.
- [Wang 18] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Guilin Liu, Andrew Tao, Jan Kautz and Bryan Catanzaro: Video-to-Video Synthesis, arXiv:1808.06601, 2018.