全自動プロット画像数値化プログラムの公開

Release of Full-Automatic Program for Numerical Conversion of Plot images

物材機構-MaDIS[1]　○吉武 道子[1]，河野　敬[1]，門平 卓也[1]

NIMS-MaDIS[1], ○Michiko Yoshitake[1], Takashi Kono[1], Takuya Kadohira[1]

E-mail: yoshitake.michiko@nims.go.jp

論文には、波長や温度などを変化させて特性値の数値データを取得し、そのデータ点を線でつないだ折れ線グラフが掲載されていることが多い。これらの折れ線グラフは画像の形（プロット画像と呼ぶ）で提供されていることがほとんどで、同一の物質の同じ特性を測定した折れ線グラフであっても論文ごとに図の大きさ・軸のスケールなどが異なっており、画像のままでは数量的解析が難しい。そこで我々は、画像で提供された様々な折れ線グラフ（＝プロット画像）をコンピュータにより自動的に数値データ化するプログラムを開発した[1]。

このプログラムは、NIMS の論文データを多角的に検索する FigResourceMiner というシステム内で、検索したプロット画像の数値データをダウンロードできるようにする目的で開発された。出版社から提供された XML ファイルの画像部分を png フォーマットに変換したものが NIMS のデータプラットフォームに蓄積されており、その png ファイルを読み込んで、全自動で数値データへ変換して出力するよう組み込まれている。これを、png ファイルを読み込んで全自動で数値データへ変換して出力する独立したプログラムとしてソースコードを NIMS-MDR（Material Data Repository）にて公開した[2]。ソースコード中の png ファイルを読み込みに行くディレクトリーを変更するだけで各自の持つ png ファイルから数値データへ変換されたファイルが得られる。

プログラムの処理は大まかに以下のような流れになっている。(1)変換の対象となる画像を選別、(2)グラフ画像の原点位置と横軸縦軸の画像上の長さを求め、(3)プロット画像の色数を求めて色ごとに画像を分離してモノクロ化、(4)(2)の情報を用いてグラフ描画領域のみを抽出し、(5)色の重なりで欠損した部分の補修やアノテーションの削除漏れの除去などの整形処理後、(6)それぞれの色ごとに描画されている折れ線グラフの本数を判別し、(7)本数ごとに折れ線グラフの数値化を行い、最後に元の 1 つの png 画像全体からの折れ線グラフの数値化データを一つの csv ファイルとして出力する。プログラム全体の流れを図に示した。上記のプロセスの内、(1)（図中 A）、(2)（図中 B）、(5)（図中 C）、(6)（図中 D）、(7)（図中 E）、の 5 箇所で深層学習を用いている。

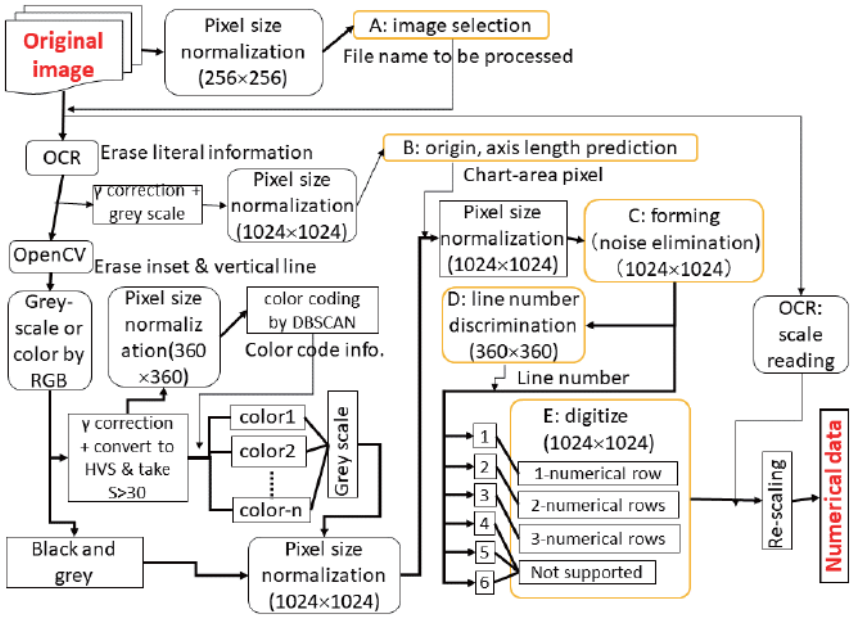

Fig. Brief flow of the data treatments in the program.

[1] Michiko Yoshitake, Takashi Kono, Takuya Kadohira, J. Comput. Chem. Jpn., 19, 25 (2020).
[2] “Program for automatic numerical conversion of a line graph (line plot)“ in https://mdr.nims.go.jp

© 2021年 応用物理学会