金属組織観察における深層学習を用いた画像認識技術の研究

Semantic Segmentation using Deep Learning for Microstructure Recognition

粟野友貴　技術開発本部技術基盤センター数理工学グループ
米倉一男　技術開発本部技術基盤センター数理工学グループ　博士（情報理工学）
宮澤優斗　技術開発本部技術基盤センター金属・材料評価グループ

金属ミクロ組織の粒界の長さに対する粒界上に析出する炭化物の比率は，強度や寿命などの重要な機械的特性と関連性があると考えられている．従来，その比率は専門家自らが写真などから手作業で計測していたため，時間とコストを要していた．金属組織の様相は熱処理時間などで異なるため，画一的なしきい値などで粒界を自動検出することは難しい．このような課題には，近年開発された Encoder-Decoder 構造の Convolutional Neural Network (CNN) を用いることで，古典的な画像処理アルゴリズムでは判別が困難であった抽象的な特徴領域の検出が期待できる．本稿では Encoder-Decoder 構造の CNN の一つである U-Net をベースとした構造を用いて粒界の検出を行い，予測精度が先端的な手法である DeepLab v3+ よりも約 2 ポイント高い約 72％で予測できることを確認した．

The aspect of microstructures, such as the ratio of carbon precipitates to grain boundary, is thought to have a certain relation with its mechanical properties. The ratio has been measured by experts, which consumes time and cost. However it is difficult to detect the grain boundary by means of classic image recognition techniques since the aspect of microstructures is easily affected by heat treatment time. Such complex features could be detected by the recently developed architecture, Encoder-Decoder CNN. This paper presents the prediction results of U-Net based model, which IoU (Intersection over Union) was 72% and 2 points higher than state-of-the-art architecture, DeepLab v3+.

1. 緒　　言

金属材料の強度などの機械的特性は，機械製品の設計において健全性や寿命などの評価に関係する重要な特性である．粒界や粒界上に析出する物質の割合などが材料特性に影響を与えるため，これを定量的に評価することは特性を定量的に把握することにつながる．粒界および析出物は SEM（Scanning Electron Microscope，走査型電子顕微鏡）などを用いることで観察できる．部材から断面を切り出して試験片とし，断面を研磨して整え，エッチングで組織にコントラストを付け，顕微鏡で表面画像を取得する．取得した画像から粒界の長さと析出物の割合を計測するが，従来は手作業で計測を行っていたため，観察者の熟練度に依存し，また作業にも時間を要していた．この作業を自動化し，画像に写るものをピクセルごとに析出物かどうかを識別できれば，再現性の高い結果が得られるとともに大幅な作業時間の短縮につながる．

画像のどこに何が写っているかの判別を，機械学習の用語では Semantic Segmentation（もしくは Semantic Image Segmentation）と呼ぶ．近年の深層学習の登場に伴って技術が大きく発展しており，例えば，ある画像から人・車・空などが写っている領域の抽出・分類ができる．特に Google 社（アメリカ）が 2018 年に提案した DeepLab v3+ [1] は，人・車・空などの実写データセットである PASCAL VOC 2012 [2] や Cityscapes [3] の領域分けにおいて，ほかの手法と比較して高精度なスコア（89.0％，82.1％）を記録している．本稿では金属組織観察に Semantic Segmentation を用いて自動で組織の判別を行う手法，および予測精度を DeepLab v3+ と比較した結果を紹介する．なお，本稿の構築には TensorFlow [4] をバックエンドにした Keras [5] を用いた．

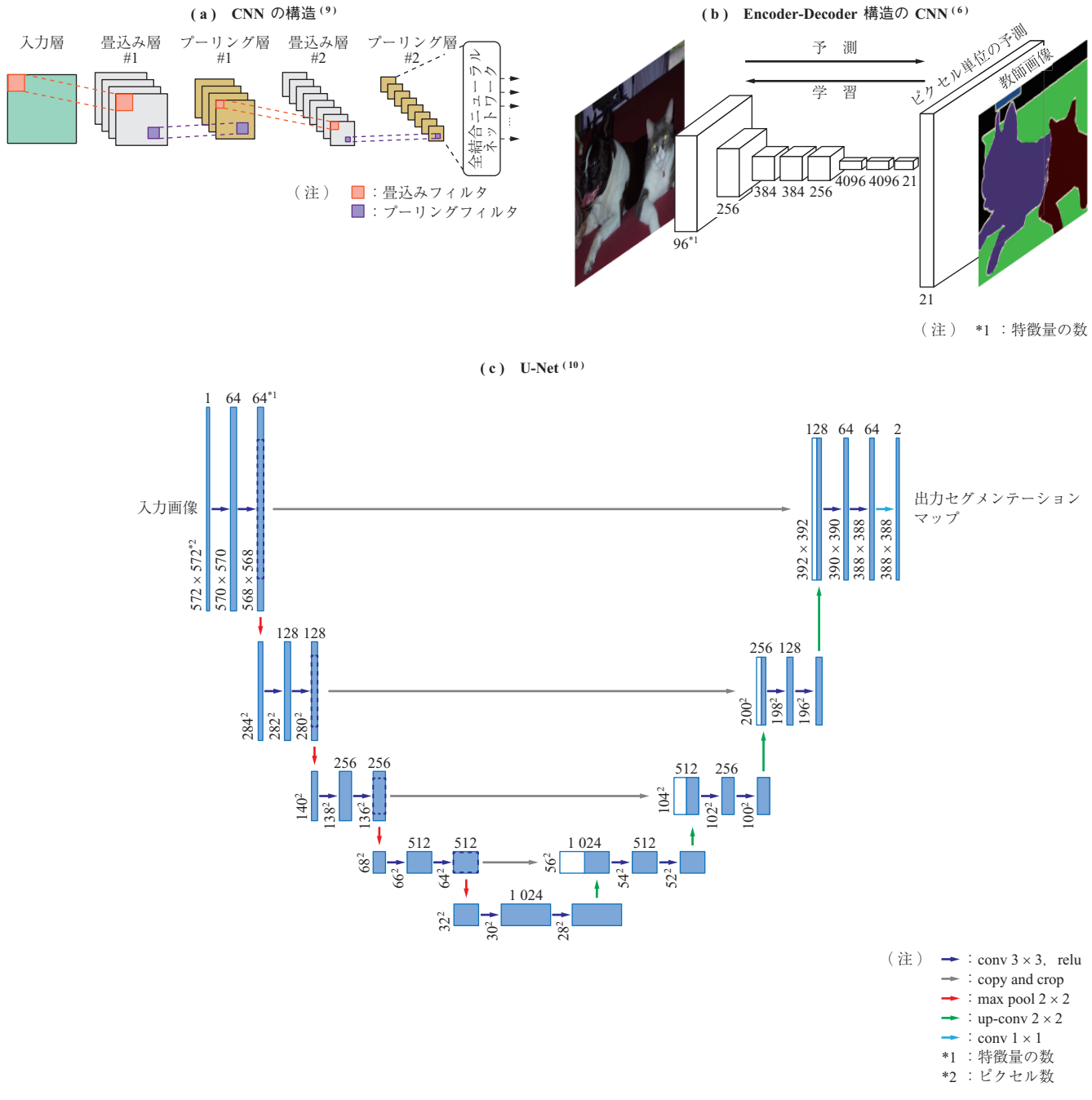
2. 機械学習技術の概要

2. 1　Encoder-Decoder 構造の CNN

Semantic Segmentation は，画像認識の分野では長年難しい問題として扱われてきた．近年，深層学習手法の一つである CNN (Convolutional Neural Network) を利用することで，識別精度が飛躍的に向上することが分かり [6]，CNN を用いた Semantic Segmentation の研究が盛んに行われている．特に CNN を Encoder-Decoder 構造にした手法は，その中でも代表的な手法の一つで，自動運転 [7]，バイオメディカル [8] など幅広い分野で活用が進んでいる．

第 1 図にニューラルネットワークの構造を示す．画像認識などに使われる典型的な CNN では，**第 1 図 - (a)** のように画像の入力に対して特徴量を抽出し，そこから少しずつ層が小さくなる構造をしている．これに対して Encoder-Decoder 構造の CNN は**第 1 図 - (b)** のように，層が途中まで小さくなるものの，その後大きくなる構造をしている．左半分が Encoder と呼ばれる部分で，画像から特徴量を取り出し，どこに人・車・空が写っているかを判断する役割がある．右半分は Encoder から得られる，どこに人・車・空があるという情報を基に，ピクセルに色付けをする役割を果たしており，Decoder と呼ばれる．この Encoder-Decoder 構造は判別精度が高く，これを基にさまざまな工夫を加えた手法として最近数年間で U-Net [10]，SegNet [11]，PSPNet [12]，DeepLab v3+ など多くの構造が提案されている．

本稿では幅広い対象に使われ，実績のある U-Net を使用する．U-Net の構造は**第 1 図 - (c)** のように Encoder と Decoder の同じ大きさの層を結び付け，小さな層を下側にずらした U 字の階段状に変形した形をしている．Encoder と Decoder の同じ階層の層をつなげることで各層



第 1 図 ニューラルネットワークの構造
Fig. 1 Neural networks

に情報を伝えることができるため，高い精度が得られる．

2. 2　評価関数

U-Net を含む深層学習では出力した結果と教師データ（正解とする結果）の違いを評価して，ネットワークの重みを変化させる．この評価関数にもさまざまな選択肢があり，適切な関数を選択することが重要である．本稿では，ピクセルごとにラベルの正しさを評価する Binary Cross Entropy (BCE) と正解画像と予測画像がオーバーラップする割合を評価する Dice Loss (DL) [13] という関数を使用した．各関数の特徴を以下に記す．

第 2 図に正解画像と U-Net の出力画像の概念図を示す．前提として，図のように正解画像と U-Net から得られる画像があるものとする．正解画像は i 番目のピクセルに $t_i = 0$ または 1 のラベルが付いている．一方で U-Net の出力は i 番目のピクセルに対してラベルが 1 である確率 y_i ($0 \le y_i \le 1$) が得られる．

BCE ではピクセルごとにラベルの正しさを評価する．

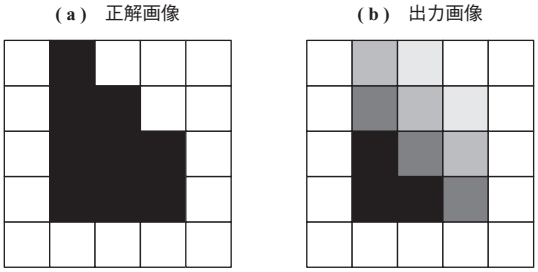
$$BCE(t, y) = -\sum_i \left(t_i \log y_i + (1 - t_i) \log(1 - y_i) \right) \quad \cdots\cdots (1)$$

この関数は，$y_i = t_i = 0$ の場合と $y_i = t_i = 1$ の場合にともに 0 になり，y_i と t_i の一方が 0 でもう一方が 1 の場合に無限大に発散する．したがって，この *BCE* を最小化するように学習を行えば，$y_i = t_i$ となって正しい出力が得られるようになる．*BCE* はピクセル単位での微視的な評価であることが特徴である．

DL は正解画像と予測画像がオーバーラップする割合を評価する．そのため，*BCE* が微視的な指標であるのに対して，*DL* はより画像内にある物体を検出する目的に則した指標と言える．このとき *DL* は (2) 式で定義される．

$$DL = 1 - \frac{2\sum_i t_i y_i + \varepsilon}{\sum_i t_i + \sum_i y_i + \varepsilon} \quad \cdots\cdots (2)$$

ここで ε は 0 で割ることによる発散を回避するための任意の定数項である．*DL* は，与えられた正解に対し，全く同じ形をしていれば 0 になり，不正解のピクセルが増えるほど値が 1 に近づく．



第 2 図　正解画像と U-Net の出力画像の概念図
Fig. 2　Output example of U-Net

2. 3　Residual Block

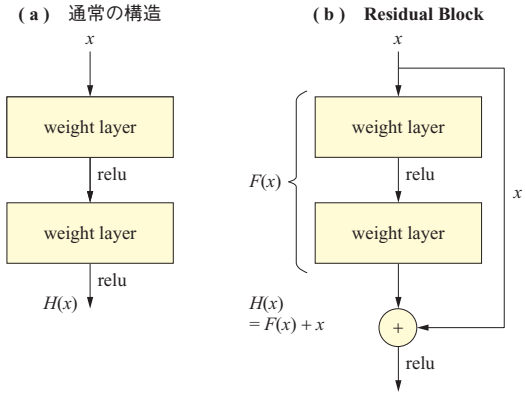
ニューラルネットワークの層をより深くして学習するために ResNet (Residual Network) [14] が Microsoft Research（アメリカ）によって提案された．ResNet が提案される前から層を深くすることでより複雑な特徴を抽出できると考えられていたが，学習を収束させることが難しく，逆に性能が悪化する傾向にあった．**第 3 図**に通常の構造と Residual Block を示す．ResNet では**第 3 図 - (b)** のように入力値を二つ以上下層の層にバイパスするような Residual Block を設け，$H(x) = F(x) + x$ となる計算ネットワークを構成する（バイパスして接続することを skip connection と呼ぶ）．そうすることで，ある層の出力値 $H(x)$ を学習する際，入力値 x との残差 (Residual) $F(x) = H(x) - x$ を学習することになる．これにより，$H(x)$ と x が近い場合に微小な $F(x)$ を学習しやすくなる．

予測値と真値との差を評価する損失関数を可視化した際，Residual Block を用いた場合は，用いなかった場合と比較して緩やかな凸関数に近づき，学習しやすくなったことが報告されている [15]．

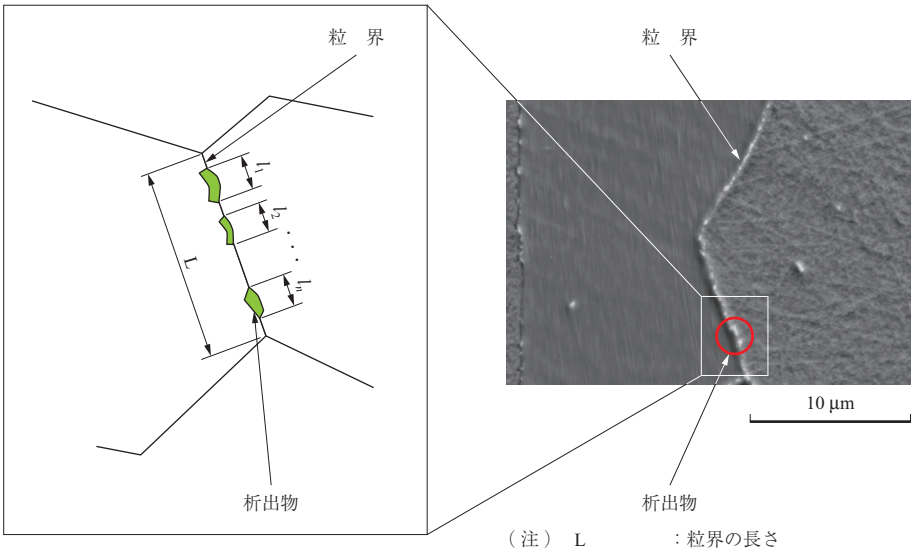
3.　Encoder-Decoder 構造の CNN を用いた粒界検出

3. 1　概　　要

第 4 図に Alloy690 の粒界と析出物を示す．図のようにミクロ組織の粒界上にある炭化物と粒界の比率は機械的特性に関わると考えられているが，これまで手作業で測定されており，作業コストが掛かるとともに，作業者によって結果がばらついていた．そこで本稿では比率を自動計測するための第一歩として，深層学習を用いて Alloy690 の



第 3 図　通常構造と Residual Block
Fig. 3　Normal structure and Residual Block



（注）　L　　：粒界の長さ
l_1, l_2, … l_n　：粒界に沿った析出物の長さ

第 4 図　Alloy690 の粒界と析出物
Fig. 4　Grain boundary and precipitate of Alloy690

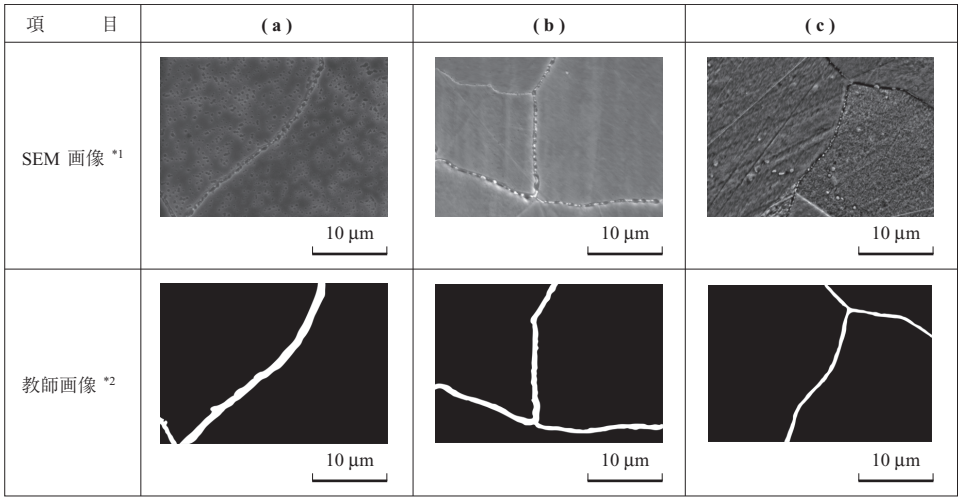
粒界を検出した．Alloy690 は粒界上に炭化物を析出させて，応力腐食割れの耐性を向上させている．炭化物より形状が複雑な粒界の予測性能を評価し，比率取得の可能性について検討した．

3. 2　粒界画像

第 5 図に本稿で使用した教師データである SEM 画像とその粒界を白く塗りつぶした画像（教師画像）の 1 例を示す．**第 5 図 - (a)**，**- (b)**，**- (c)** はそれぞれ熱処理時間が異なることから，断面の模様は大きく異なり，また研磨傷や析出物に似た白い斑点が粒界外にあることが確認できる．このように配色の傾向が画像ごとに異なったり，粒界に近い形状が入り交じったりしており，従来の画像認識手法では識別が困難であるため，本稿では深層学習への適用が適切と考えた．学習枚数は 100 枚とした．

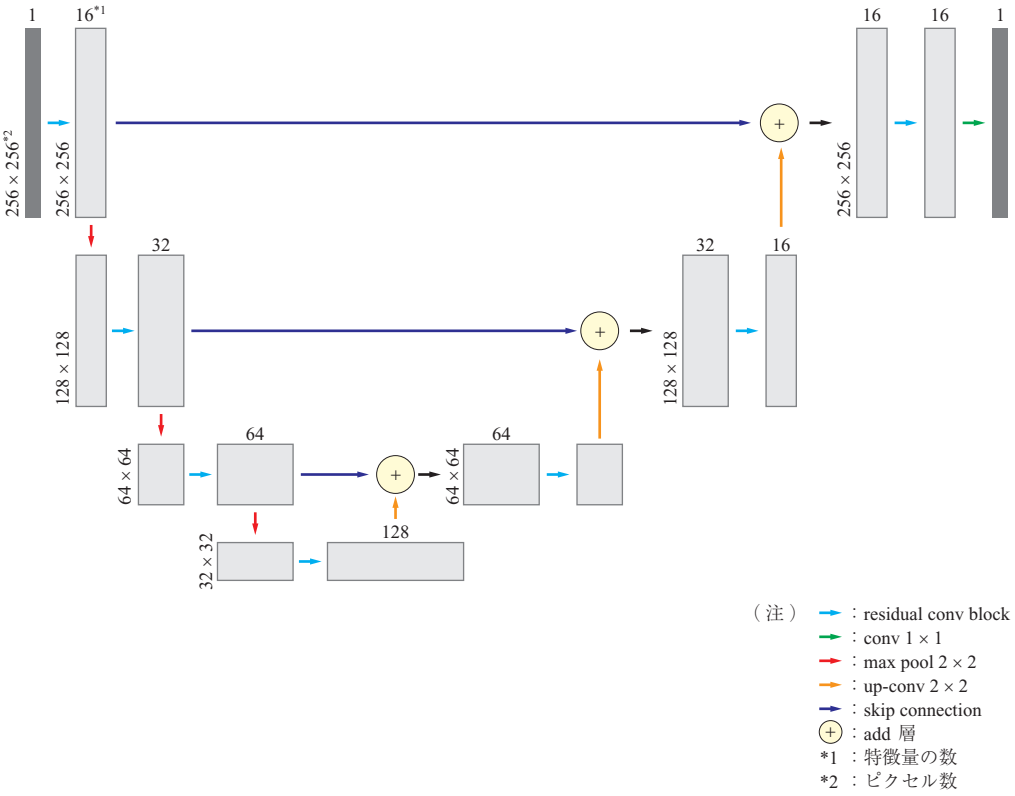
3. 3　ネットワーク構造

ネットワーク構造は U-Net をベースに構築した．**第 6 図**にその構造を示す．従来構造とは異なり，concatenate 層（特徴量同士を連結）の代わりに add 層（特徴量同士を要素ごとに足し合わせる）を入れた．初期検討時に add 層の方が領域を滑らかに検出できたためである．



（注）　拡大率：5 000 倍
*1　　：教師データ
*2　　：粒界を白く塗りつぶした．

第 5 図　識別対象
Fig. 5　Input image and Ground Truth



第 6 図　本稿のネットワーク構造（depth = 4 のとき）
Fig. 6　Our network model (when depth = 4)

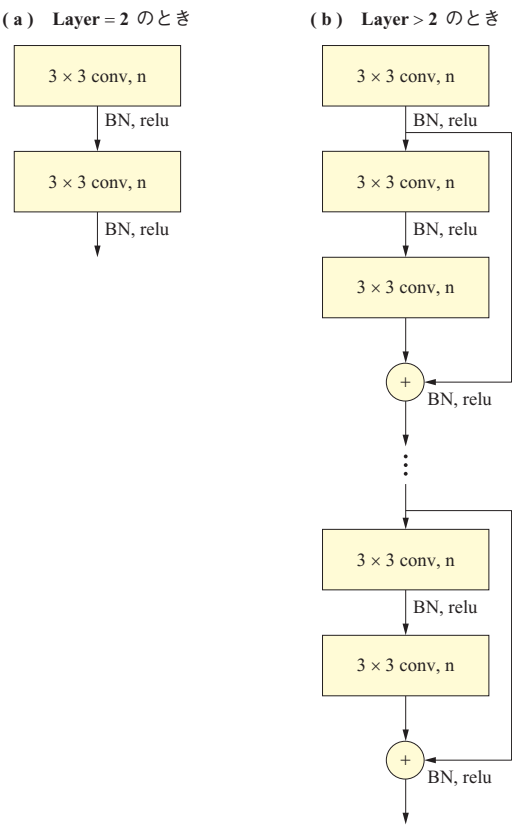
ここで，ネットワークの階層の深さを depth，各階層でのニューラルネットの層の数を layer と呼ぶ．例えば**第 6 図**では depth は 4 である．**第 7 図**に本稿で適用したネットワーク構造を示す．2 層以上の場合は，**第 7 図 - (b)** に示すように畳込み層である conv 層の代わりに，畳込み層を含む Residual Block (residual conv block) 層を入れた．層が深くなっても学習しやすくするためである．**第 7 図 - (a)** に示すように residual conv block 層が入力層を含む 2 層の場合は conv 層とした．

3. 4　学習手法

学習に使用したハイパーパラメータ値を**第 1 表**に示す．ハイパーパラメータの学習率，モーメント係数，L2 正則化係数は depth = 4，layer = 2 の条件でベイズ最適化 [16] したものを使用し，ベイズ最適化には Optuna [17] を用いた．学習係数の最適化手法は Residual Block と相性が良いとされる Momentum-SGD [18] を用いた．

4.　結果および考察

まず Residual Block 特有の skip connection があるモデルとないモデルの予測精度 IoU (Intersection over Union) を比較した結果を記す．**第 8 図**に layer 数を 5 と固定し

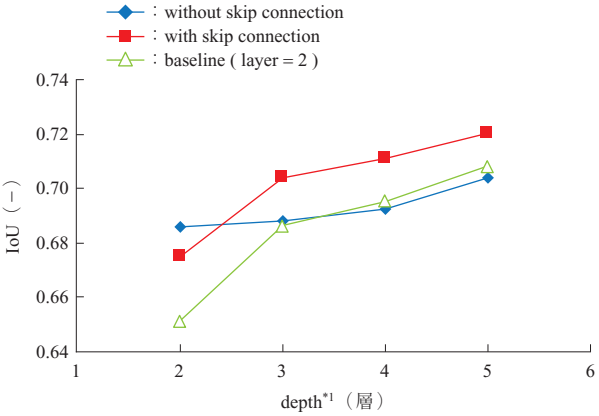


第 7 図　本稿で適用した residual conv block 層
Fig. 7　Residual conv block layer applied in this paper

第 1 表　ハイパーパラメータ
Table 1　Hyper parameters

ハイパーパラメータ	単　位	値
エ　ポ　ッ　ク　数	回	100
バ　ッ　チ　サ　イ　ズ	個	32
フ　ィ　ル　タ　数	個	16
学　　習　　率*1	−	1.195
モ　ー　メ　ン　ト　係　数*1	−	0.929
L2 正　則　化　係　数*1	−	1.130

（注）　*1：depth = 4，layer = 2 の条件でベイズ最適化したものを使用



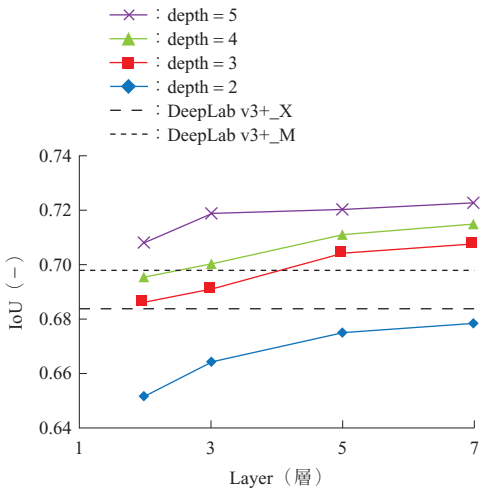
（注）　skip connection の有無
*1：layer = 5 のとき

第 8 図　Residual Block の効果
Fig. 8　Comparison of Regular Block and Residual Block

たとき（skip connection が 2 か所あるとき）に depth 別の予測精度を示す．baseline である layer 数が 2 のときと比較して，depth が 2 のときは，skip connection の有無に関係なく，より layer 数が深い 5 の方が，予測精度が高くなった．一方で，depth が 3 より深いとき，skip connection がないモデルとより浅いモデル（layer 数が 2 のとき）で予測精度がほぼ変わらなかったが，skip connection があるモデルは予測精度がそれらより高くなった．このことから，本稿において depth がある程度大きくなる (depth ≥ 3) と skip connection を入れた方が予測精度は高くなると言える．

つづいて，モデルの深さ (depth，layer) と予測精度の関係を調査した結果を記す．**第 9 図**にネットワークの階層の深さ depth と residual conv block の深さ layer と予測精度の関係を示す．depth が 5 以下（depth をそれ以上大きくする場合は入力解像度を縦横ともに 2 倍以上にする必要がある）かつ layer が 7 以下のときは，depth と layer の両方を大きくするほど予測精度が向上することを確認した．

本稿の精度を DeepLab v3+ と比較したところ，depth = 5，layer = 7 のとき，本稿の予測精度が 2 ポイント以上高いことを確認した（本稿：72.2%，DeepLab v3+：69.8%）．DeepLab v3+ はネットワークの規模が大きく，大規模なデータセットから多クラス領域を予測することに長けている一方で，本稿のように粒界に限定し，小規模なデータセット（100 枚）に対して予測する場合は専用のモデルを作成した方が良い精度が得られることを示唆している．

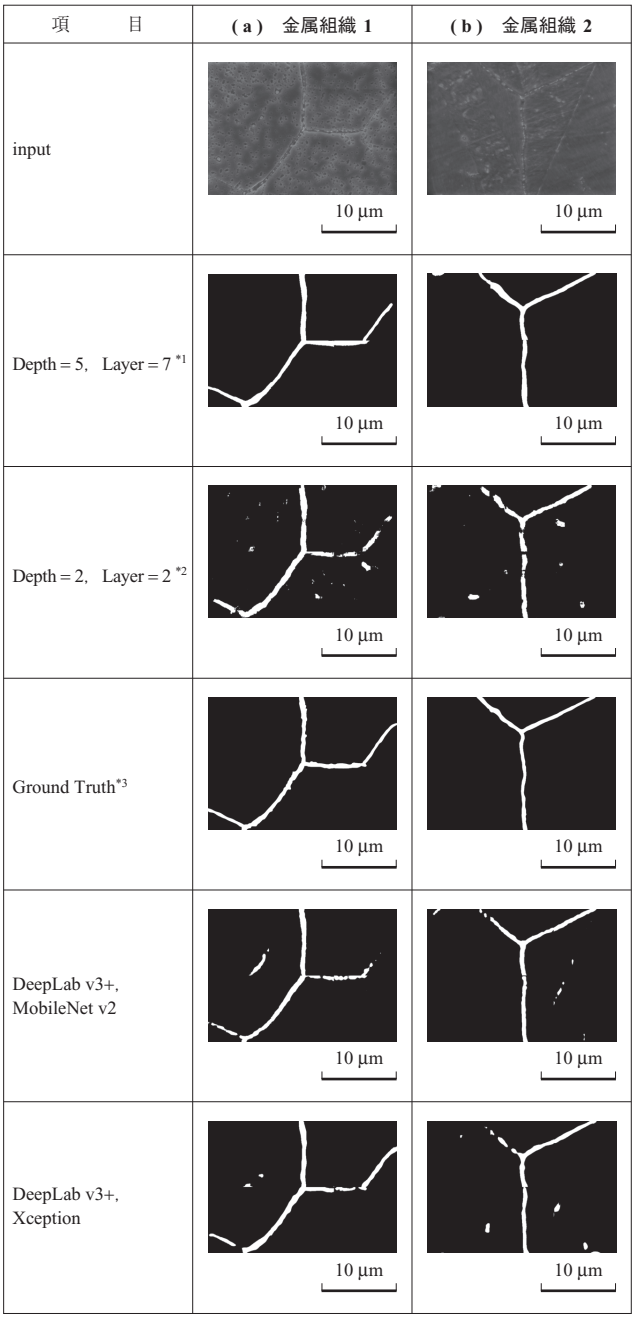


（注）　DeepLab v3+_X はネットワークのバックボーンとして Xception を採用，DeepLab v3+_M は MobileNet v2 を採用

第 9 図　層の深さ (layer，depth) と検出精度の関係
Fig. 9　Result of each model

第 10 図に本稿で最も精度の高かったモデル (Depth = 5, Layer = 7)，最も精度の低かったモデル (Depth = 2, Layer = 2)，と DeepLab v3+（バックボーンが MobileNet v2，Xception の 2 種）の予測結果を示す．最も精度の高かったモデル (Depth = 5，Layer = 7) とほかの予測結果を比較すると粒界以外の領域が検出されたり，粒界が 1 本の線ではなく，ぶつ切りに検出されたりしたことを確認した．最終的に求めたい析出物との比率は粒界の長さから求める．粒界の長さは予測領域の中心軸を求める medial axis（物体形状の基本的な表現方法の一つ）を用いて測定する．そのため，Depth = 5，Layer = 7 の予測結果では，粒界が途切れずに正しい幾何形状で予測できており，粒界全体の長さが正しく測定可能である．予測精度を落とした要因として，真値である Ground Truth と比べて粒界が太く検出されていることが挙げられるが，長さを測定する場合はほとんど影響しない．

本稿で作成したモデルによって人間の判断に近い予測結果が得られた．その一方で，粒界が一部途切れている場合もあることを確認した．途切れないようにするために，トポロジーを保存するような関数を損失関数に入れることが



（注） *1：本稿で最も精度の高かったモデル
*2：本稿で最も精度の低かったモデル
*3：真　値

第 10 図　予測結果
Fig. 10　Predicted results

一案として挙げられる．粒界の場合，粒界が途切れたり，欠損したりすると粒界のエッジ数が増減するため，粒界のエッジ数が正解画像と一致しない場合はペナルティを加えることで途切れが抑制されると考えられる．粒界のエッジ数を medial axis を用いて求め，真のエッジ数との差分を損失関数に与えることが 1 例として挙げられる．

5. 結　　言

Encoder-Decoder 構造の CNN を用いて合金断面の粒界を検出した．さらに，U-Net をベースとした構造に Residual Block を加えたネットワーク構造を作成し，ネットワークの深さを調整しながらその予測精度を調査した．結果としてネットワークの階層 depth を深くするほど，各階層の層 (Residual Block) を深くするほど予測精度が向上することを確認した．また，先端的な手法である DeepLab v3+ と比較して約 2 ポイント精度が向上した．これらの検証結果から本稿は粒界の長さを計測するための精度を担保していることを確認した．

その一方で，粒界が一部途切れて検出される傾向にあり，トポロジーを維持できるように工夫する余地がある．具体的には粒界のエッジ数が正解画像と一致しないときに損失関数にペナルティを加えることが挙げられる．

── 謝　辞 ──

本稿は 2 人の海外研修生である富田 Huuskonen 祐介氏（ENSEIRB-MATMECA，フランス），Yanis Basso-Bert 氏（École Centrale de Lyon，フランス）の協力の下，実施された．積極的に課題に取り組み，多くの試行錯誤を重ねてくれたことに厚く感謝を申し上げる．

参 考 文 献

(1) L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff and H. Adam：Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation, arXiv:1802.02611v3，2018. 8

(2) Visual Object Classes Challenge 2012 (VOC2012), http://host.robots.ox.ac.uk/pascal/VOC/voc2012/,（参照 2019. 10. 1 ）

(3) The Cityscapes Dataset，https://www.cityscapes-dataset.com/,（参照 2019. 10. 1 ）

(4) TensorFlow：https://www.tensorflow.org/,（参照 2019. 10. 1 ）

(5) Keras Documentation：Keras: The Python Deep Learning library，https://keras.io,（参照 2019. 10. 1 ）

(6) J. Long, E. Shelhamer and T. Darrell：Fully Convolutional Networks for Semantic Segmentation, CVPR，2015

(7) M. Treml, J. A. Medina, T. Unterthiner, R. Durgesh,

F. Friedmann, P. Schuberth, A. Mayr, M. Heusel, M. Hofmarcher, M. Widrich, B. Nessler and S. Hochreiter：Speeding up Semantic Segmentation for Autonomous Driving，NIPS，2016

(8) X. Hu, L. Fuxin, D. Samaras and C. Chen：Topology-Preserving Deep Image Segmentation，NeurIPS，arXiv:1906.05404v1，2019. 6

(9) 斉藤弘樹，服部　均，米倉一男：機械学習とCAE を利用したターボ機械の設計支援技術，IHI 技報，Vol. 59，No. 1，2019 年 3 月，pp. 30 － 43

(10) O. Ronneberger, P. Fischer and T. Brox：U-Net: Convolutional Networks for Biomedical Image Segmentation，arXiv:1505.04597v1，2015

(11) V. Badrinarayanan, A. Kendall and R. Cipolla：SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation，IEEE，2016

(12) H. Zhao, J. Shi, X. Qi, X. Wang and J. Jia：Pyramid Scene Parsing Network，CVPR，2017

(13) C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin and M. J. Cardoso：Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations，Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support，LNCS，Vol. 10 553，2017. 7，pp. 240 － 248

(14) K. He, X. Zhang, S. Ren and J. Sun：Deep Residual Learning for Image Recognition，CVPR，2015. 12

(15) H. Li, Z. Xu, G. Taylor, C. Studer and T. Goldstein：Visualizing the Loss Landscape of Neural Nets，NIPS，2018

(16) J. Snoek, H. Larochelle and R. P. Adams：Practical Bayesian Optimization of Machine Learning Algorithms，NIPS，2012

(17) Optuna：A hyperparameter optimization framework，https://github.com/optuna/optuna，（参照 2019. 10. 1）

(18) N. Qian：On the Momentum Term in Gradient Descent Learning Algorithms, Neural Networks，Vol. 12，Iss. 1，1999，pp. 145 － 151