Classification of titanium microstructure with fully convolutional neural networks

S Mongkhonthanaphon and Y Limpiyakorn

Department of Computer Engineering, Chulalongkorn University, 254 Phayathai Road, Pathumwan, Bangkok 10330, Thailand
limpyac@gmail.com

Abstract. Titanium and its alloy exhibit excellent properties for biomedical applications, especially in implant surgery. Classification of Titanium microstructure is the process in material inspection that reveals background of the material. Generally, microstructure classification is manually performed. Due to the complexity of microstructure features, expertise is required for process operation. The traditional classification by humans is time consuming and possibly error prone if the inspection is not performed by titanium microstructure experts. Deep learning is considered the revolution of computer vision to enable computers to see and perceive like humans. The technique is widely used for automatically classifying images with high accuracy. In order to reduce human inspection time during quality control, this research presents the use of a type of deep learning, Fully Convolutional Neural Networks, for pixel-wise classification in the titanium microstructure images. The dataset contains private images of titanium samples taken by SEM microscopes. As the available training dataset is small, data augmentation using elastic deformations is applied for increasing the accuracy of the model. Constructed with the U-net architecture, the network achieves good performance with the pixel accuracy of 92.67% and mean IoU of 71.30%.

1. Introduction
Titanium and its alloys have become very important commercially over the past fifty years due to their low density, good mechanical properties, good strength-to-weight ratio, good corrosion resistance and good biocompatibility. Surgery, especially in implant surgery uses titanium alloys as implant material. The manufacturing process for implant material therefore has to be able to build complex shape. The challenge is how to make this titanium as precise as possible.

Since 2006, selective laser melting (SLM) has been introduced with the ability to make complex shape, reduce fabrication time and reduce material waste from the producing process. Unfortunately, as-built state titanium parts from SLM are unable to achieve the great performance. Their toughness and ductility are not high enough to be used as implant material because they cannot take too much load. To improve the toughness and ductility, heat treatment is needed. To improve some mechanical behaviour, Harr and Becker [1] experimented the heat treatment process on Ti-6Al-4V (6% aluminium, 4% vanadium titanium alloy), produced by SLM method. The result showed a significant increase of %elongation for the samples that have been heated above 750°C (dissolution temperature). An as-built state of Ti-6Al-4V consists only the Alpha phase. The dissolution temperature is the temperature that the Alpha phase is about to start dissolve into Beta phase. The other matters are time and method of cooling that can affect the microstructure morphology. Mostly of microstructure visualization are manually performed by metallurgists including phase classification. In this work, deep learning is
proposed as a means to automate the classification of Ti-6Al-4V microstructure in order to reduce the time spent on inspection for quality control.

Convolutional Neural Networks (CNNs) have been one of the most successful for image classification. In 1989, the first CNN was proposed by LeCun et al. [2] who used the network to recognize patterns on hand-written digits dataset. In 2012, Krizhevsky et al. [3] adapted LeNet into the breakthrough AlexNet with 7-stacked layers that won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in that year. Moreover, AlexNet drastically improved the accuracy of the competition by 10% from 2011. This phenomenon had drawn interest into the field of Deep Learning.

Semantic segmentation is one of the key problems in the field of computer vision, and it can be applied to a variety of applications including self-driving car, medical diagnosis, virtual reality etc. The well-known method in semantic segmentation, so called Fully Convolutional Neural Networks (FCNN), were proposed by Long et al. in 2015 [4]. Compared to the original CNNs, FCNNs are the model containing upsampling layer instead of fully connected layer. Therefore, FCNNs can deliver the output as the same resolution as the input image. Another difference is Skip Connections that skip the information of the shallower layers to the deeper layers. This method improves further accuracy by fusing the information before the final prediction layer.

2. Related work

Thus, it is such a challenge to make computer vision recognize the patterns in microstructure. In the field of metallurgy, there are several instruments for imaging, such as optical microscope (OM) and scanning electron microscope (SEM). The SEM method yields higher resolution in the image compared to the OM method, but the operating cost is more expensive. Gola et al. [5] proposed steel microstructure classification by machine learning methods, using Support Vector Machine (SVM) as classifier to predict on steel microstructure dataset. The dataset consists of steel images that taken by OM method. All images are categorized into three classes: pearlite, martensite and bainite. The result has shown the average test accuracy around 80% on the author’s test dataset. Azimi et al. [6] applied the method of Gola et al. to their dataset, which was categorized into four classes: pearlite, tempered martensite, martensite, and bainite. However, the result showed only the accuracy of 48.89%. Azimi et al. thus proposed steel microstructure classification by deep learning methods, using FCNN [4] to classify every pixel in whole image and predict the class by max-voting scheme. The model achieved 93.94% classification accuracy and 67.84% mean IoU. The publication of Iglovikov and Shvets [7] in 2018 presented the technique of fine-tuning to increase the accuracy with the pretrained network previously trained on a large dataset. The pretrained network was required to freeze the weights for almost the whole of convolutional layers except for the very last one or two layers. Therefore, when training the new classifier, the new data will run through the network to adjust the weights of last convolution block layer. The pretrained network could work on contracting part of the U-net architecture, considered as original CNN.

3. Method

3.1. Architecture

The U-net architecture is built from FCNN concept to perform image segmentation, and it successfully works on biomedical image segmentation as the winner of ISBI cell tracking challenge 2015 [8]. The architecture is named after U alphabet as it consists of two parts concatenating like U-shape as illustrated in figure 1. The architecture can be seen as two divided parts, the part on left-side is contracting path where the architecture encodes the input image into feature map by using convolution and max-pooling operations. The right-side part is expansive path where the architecture decodes the feature map into output segmentation map by using upsampling. The addition idea of this architecture is that the expansive path also has huge feature channels, that is different from the original FCNN [4]. As when the upsampling is used, the feature map channels are halved so that the channels in the deeper layers tend to be smaller. But U-net uses concatenation method to crop and copy the feature maps in contracting
path to merge with the feature maps in the expansive path. In summary, the constructed model has 23 convolutional layers (eighteen 3x3 convolutions, four 2x2 upsampling convolutions, one 1x1 convolution) and 4 max-pooling operations.

![U-net architecture](image)

**Figure 1.** U-net architecture [8].

### 3.2. Dataset

The dataset contains images of titanium samples taken by SEM microscopes. These titanium samples are produced by SLM method and heated to transform the as-built microstructure into dual-phase microstructure. (Alpha-Beta) The dataset contains 96 images with 512x512 pixels. Each image has a corresponding ground-truth image that supervises the classes to the model. The ground-truth images are denoted for Alpha phase (black) and Beta phase (white) in the images. The dataset is split into 66 training images and 30 test images. As the available dataset is small, data augmentation using elastic deformations is applied for increasing data to the size of 264 images. Figure 2 illustrates an example of the augmented training image (on the right) that is the output from pixel shifting by shifting 5% of total width to the left and shifting 5% of total height lower. Associated with each augmented image, the ground truth image is also created using the same pixel shifting. Figure 3 illustrates an example of the augmented ground-truth image (on the right) resulting from pixel shifting. This step can be carried out with Keras [9] using ImageDataGenerator instance.

### 3.3. Training

The network produces an output of a 3D matrix with the number of channels equivalent to the number of classes. Since there are only two classes (Alpha and Beta), ground truth images are thus preprocessed into binary color images by thresholding from grayscale images. Thresholding is manually performed by replacing each pixel in an image with a black pixel if the image intensity is less than some fixed constant, manually chosen by visualizing the histogram color map. The images of titanium alloys taken by SEM microscope are used as input and all binary ground-truth images are used to supervise the network in the output section. For training, Adam optimizer [10] is used with $10^{-4}$ learning rate and no weight decay. The loss function is computed by using binary-cross entropy function. To predict the output of the last layer for every pixel, the sigmoid function is applied to score the posterior probability then choosing the class for that pixel with the highest posterior score.
3.4. Implementation
All of the implementations have been done with Tensorflow and Keras [9] on a single GPU Nvidia GTX1060 (6gb). The network was trained with 15 epochs. Training time is about 5 hours from scratch.

![Figure 2. Example of augmented training image with pixel shifting.](image1)

![Figure 3. Example of augmented ground truth image with pixel shifting.](image2)

4. Result
Once the training has been completed, the training weights are then saved in .hdf5 format. The size of file containing the values of training weight is around 375 MB. To predict the class in the test dataset it is required to load the model architecture and the weight file. The example result is shown in figure 4 where (a) is the original image of Ti-6Al-4V microstructure, (b) is the associated ground truth image, (c) is the predicted image on augmented dataset, and (d) is the predicted image on original dataset.

The evaluation of the performance of semantic segmentation includes four widely used metrics: 1) pixel accuracy, 2) mean accuracy, 3) mean of intersection over union (IoU), and 4) frequency weighted of IoU. Given that $n_{ij}$ denotes the number of pixels of class i-th predicted to belong to class j; $n_{cl}$ is the number of different classes included in ground truth segmentation; and $t_i$ is the total number of pixels of class i-th in ground truth segmentation. The calculation of the four metrics is defined as in equation (1), (2), (3), and (4), respectively.

\[
\text{pixel accuracy} = \frac{\sum_i n_{ii}}{\sum_i t_i} 
\]
\[
\text{mean accuracy} = \frac{1}{n_{cl}} \sum_i \frac{t_i}{n_{ii}}
\]
\[
\text{mean IoU} = \frac{1}{n_{cl}} \sum_i \frac{t_i}{t_i + \sum_j (n_{ji} + n_{ij} - n_{ii})}
\]
\[
\text{frequency weighted IoU} = \frac{1}{n_{cl}} \sum_i \frac{t_i n_{ii}}{t_i + \sum_j (n_{ji} + n_{ij} - n_{ii})}
\]

The summary of model performance is detailed in table 1 reporting all the measures on two datasets: with or without data augmentation. In case of using data augmentation and Adam optimizer strategy, the network achieves the best pixel accuracy of 92.67%, mean accuracy of 80.39%, mean IoU of 71.30% and frequency weighted IoU of 87.24%, computed by comparing the prediction images with the ground-truth images prepared for testing. The results on non-data augmentation achieve lower performance, that is, the pixel accuracy of 83.09%, mean accuracy of 67.27%, Mean IoU of 65.12% and frequency weighted IoU of 82.03%. Observing that data augmentation significantly increases the accuracy of the network.

| Training strategy | pixel accuracy | mean accuracy | mean IoU | frequency weighted IoU |
|-------------------|----------------|---------------|----------|-----------------------|
| no data augmentation | 0.8309         | 0.6727        | 0.6512   | 0.8203                |
| data augmentation  | **0.9267**     | **0.8039**    | **0.7130**| **0.8724**            |

Table 1. The effect of data augmentation on FCNN.
5. Conclusion
The unique properties of Titanium alloys have a potential to solve problems encountered by current biomedical materials. Improvements in both the quality and quantity of the implant material have made this treatment modality very promising and highly practiced in today’s era. Since manual inspection of Titanium microstructure is resource consumption, this paper presents an approach to applying deep learning for the classification of the Ti-6Al-4V microstructure. The FCNN classifier is built on U-net architecture to classify the phases (Alpha or Beta) in titanium microstructure images. As the small dataset is available, data augmentation with elastic deformations is performed for increasing data to prevent model overfitting. The performance of the network achieves the pixel accuracy of 92.67% and mean IoU of 71.30%. Further investigation of other techniques would be carried out to improve the network performance. For example, integrating Fine-tuning into the U-net architecture. The technique requires the pretrained network previously trained on a large dataset.

References
[1] Terhaar, G., M., Becker, T., H. (2018) Selective Laser Melting Produced Ti-6Al-4V: Post-Process Heat Treatments to Achieve Superior Tensile Properties. Journal of Physics, Volume 371, conference 1
[2] LeCun, Y., Boser, B., Denker, J., S., Henderson, D., Howard, R., E., Hubbard, W., Jackel, L., D. (1989) Backpropagation applied to handwritten zip code recognition. Neural Computation 1(4), 541-551
[3] Krizhevsky, A., Sutskever, I., Hinton, G., E., (2012) Imagenet classification with deep convolutional neural networks. In NIPS, 1106-1114
[4] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
[5] Gola, J., Britz, D., Staudt, T., Winter, M., Schneider, S., A., Ludovici, M., Mucklich, F. (2016) Advanced microstructure classification by data mining methods. International Conference on ThermoMechanical Processing (TMP).
[6] Azimi, S., Britz, D., Engstler, M., Fritz, M. and Mücklich, F. (2018) Advanced Steel Microstructural Classification by Deep Learning Methods. Nature-Scientific Reports
[7] Iglovikov, V., Shvets, A. (2018) TernausNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet for Image Segmentation. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
[8] Ronneberger, O., Fischer, P., Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation, MICCAI 2015
[9] Chollet, F. (2017) Deep learning with Python. New York: Manning
[10] Kingma, P., D., Ba, L., J. (2015) Adam: A method for stochastic optimization. In ICLR 2015