Image-based characterization of laser scribing quality using transfer learning

Mohammad Najjartabar Bisheh[1] [iD] · Xinya Wang[1] · Shing I. Chang[1] · Shuting Lei[1] · Jianfeng Ma[2]

Abstract

Ultrafast laser scribing provides a new microscale materials processing capability. Due to the processing speed and high-quality requirement in modern industrial applications, it is important to measure and monitor quality characteristics in real time during a scribing process. Although deep learning models have been successfully applied for quality monitoring of laser welding and laser based additive manufacturing, these models require a large sample for training and a time-consuming data labelling procedure for a new application such as the laser scribing process. This paper presents a study on image-based characterization of laser scribing quality using a deep transfer learning model for several quality characteristics such as debris, scribe width, and straightness of a scribe line. Images taken from the laser scribes on intrinsic Si wafers are examined. These images are labelled in a large and a small dataset, respectively. The large dataset includes 154 and small dataset includes 21 images. A novel transfer deep convolutional neural network (TDCNN) model is proposed to learn and assess scribe quality using the small dataset. The proposed TDCNN is able to overcome the data challenge by leveraging a convolutional neural network (CNN) model already trained for basic geometric features. Appropriate image processing techniques are provided to measure scribe width and line straightness as well as total scribe and debris area using classified images with 96 percent accuracy. Validating model's performance based on the small data set, the model trained with the large dataset has a similar accuracy of 97 percent. The trained TDCNN model was also applied to a different scribing application. With 10 additional images to retrain the model, the model accuracy performs as well as the original model at 96 percent. Based on the proposed TDCNN classification of debris on a scribed image of straight lines, two algorithms are proposed to compute scribe width and straightness. The results show that all the three quality characteristics of debris, scribe width, and scribe straightness can be effectively measured based on a much smaller set of images than regular CNN models would require.

Introduction

Laser scribing is a laser micromachining technique which uses laser scanning to make a shallow scribe line on a surface. It has been extensively studied using short pulse lasers (i.e. picosecond and nanosecond) for solar cell applications (Ku et al., 2013; Wang et al., 2019; Zhao et al., 2014). Generally, higher pulse energy and lower pulse duration mean higher

productivity, but with some negative effects on scribe quality (Leitz et al., 2011; Wang et al., 2020). On the other hand, lower pulse energy means less energy waste and melting of the process material (Yao et al., 2005). To overcome thermally induced damage due to melting, recast and microcrack formation in laser scribing, process optimization through modeling is a viable approach.

Traditionally, scribing quality issues can be detected using optical and geometrical inspection. Variation in laser parameters such as pulse energy, pulse duration, repetition rate, and scanning speed can occur at any time scale. Those variations may result in several scribe issues such as debris, crack, missing pulse, or un-straight lines (Wang et al., 2021). Hence, it is very important to identify and prevent these defects during a scribing process. Scribing errors are easy to be fixed right after scribing since the defect locations are known (H.

✉ Mohammad Najjartabar Bisheh
mnajjartabar@ksu.edu

[1] Department of Industrial and Manufactruing Systems Engineering, Kansas State University, 2061 Rathbone Hall, 1701B Platt St, Manhattan, KS 66506, USA

[2] Department of Aerospace and Mechanical Engineering, Saint Louis University, 3450 Lindell Blvd., Saint Louis, MO 63103, USA

Roozbahani et al., 2018; Hamid Roozbahani et al., 2019). Note that the usual inspection and testing for the final product (e.g. solar panels) do not help since detected defects will lead to scrapping the product.

Some recent studies deployed image analysis in monitoring different laser based manufacturing processes such as additive manufacturing (Chua et al., 2017; Delli & Chang, 2018; Fotovvati et al., 2018; Grasso et al., 2018; Grasso et al., 2017; Imani et al., 2018a, b, 2019; Najjartabar-Bisheh et al., 2021; Yuan et al., 2019) and laser welding (Gonzalez-Val et al., 2020; Marco Grasso et al., 2017; Mayr et al., 2018; Shevchik et al., 2020; Shevchik et al., 2019; Zhang et al., 2020). For example, Imani et al. (2018b) attempted to relate pore size and location to laser powder bed fusion (LPBF) parameters. In their study, they built nine titanium alloy cylinders on a commercial LPBF machine (EOS M280) at different laser power, hatching spacing, and velocity conditions. Multifractal and spectral graph analysis enabled them to monitor and discriminate process deviations with around 80% statistical fidelity. Later, Imani et al. (2019) used a deep neural network (DNN) for inspection and quality control of 362 regions of interest (ROIs) representing 362 layers of a titanium alloy. A DNN algorithm called AlexNet can detect the lack of fusion flaws with 92 percent accuracy.

Various defects can occur during selective laser melting (SLM) that could be detected during the process using images (e.g. improper heat conduction in overhang features, wrong powder deposition due to a worn recoating blade, or improper heat conduction to the underneath powder at the connection between the bottom layers of the part and the supports) (Grasso et al., 2017). SLM process monitoring might be even much more challenging for difficult-to-process materials (e.g. zinc and its alloys) (Grasso et al., 2018). They compared several image segmentation methods on zinc powder ROIs to detect stable and unstable meting conditions using multivariate control charts. Their study showed that process monitoring of some difficult-to-process materials could be completely automated using suitable image segmentation techniques.

In a laser-induced material melting-solidification process, the quality of welded parts might be deteriorated by porosity, cracks, lack of fusion, and incomplete penetration (Zhang et al., 2020). Even though machine learning has been used and explored in laser welding more than in other applications of laser technology, challenges still remain in making laser welding processes more stable using advanced techniques for quality monitoring (Mayr et al., 2018). Recently Gonzalez-Val et al. (2020) released first large dataset of laser metal decomposition (LMD) and laser welding (Gonzalez-Val et al., 2020). This dataset primarily includes 1.6 million images in which 24,444 of them are labeled as defect.

Based on the general performance of convolutional neural network (CNN) on image data, Mayr et al. (2018) examined a shallow CNN to monitor irregular weld seam, recessed weld seam, undercut, weld bead, and holes and spatters in laser welding. Their combined quality monitoring system was able to detect 209 out of 227 bad parts. Shevchik et al. (2020) used hard X-ray radiography images to train a supervised DNN to reveal the unique signature of sub-surface events in wavelet spectrograms from the laser back-reflection and acoustic emission signals. Using 300 images in training and 100 in test set, their quality classification was able to achieve an accuracy between 71 and 99% (Shevchik et al., 2020). Shevchik et al. (2019) adopted a graph support vector machine with data adaptive kernel approach and 23 laser welds as the dataset to achieve an accuracy ranging between 85.9 and 99.9%.

Current physical models are capable of predicting certain geometric aspects of laser scribing such as scribe width and depth. However, several other important quality measures cannot be obtained from the model. These quality measures include heat affected zone, debris, and micro-cracks (Roozbahani et al., 2017). Roozbahani et al. (2018) defined discontinuance as a kind of defect in laser scribing and tried to detect discontinuance area in copper indium gallium selenide solar panels using a particle analysis algorithm. However, laser scribes might also suffer from several other quality issues.

To the best of our knowledge, no existing models can predict all aspects of laser scribing quality, which can be attributed to the following three main reasons. First, laser scribing using short laser pulses is a very complicated process involving many laser parameters and various physical processes in which mechanisms are not completely understood. Second, ultrafast laser-matter interaction is a highly dynamic process and materials are first pushed to a highly non-equilibrium state followed by a rapid hydrodynamic motion, resulting in material ejection. During this process a material experiences fast phase changes and property changes (e.g. physical, optical, mechanical, electrical, etc.), making it extremely difficult to obtain reliable material data to feed into a model. Finally, the uncertainties associated with physical equipment (e.g., laser power fluctuation) and environment (e.g., temperature, vibration) can derail a model from giving reliable predictions since many of these process variations are treated as noise and thus not being considered in a physical model.

With the advent of machine vision and machine learning (ML), an opportunity arises for an Artificial Intelligence (AI) framework to be used to monitor and characterize a laser scribing process with multiple quality features including debris, scribe width, and scribe straightness. The inputs of the proposed framework are images while the scribing is taking place and the output is a classification reports on the quality characteristics under consideration. We propose a deep learning method for the AI framework. Considering the
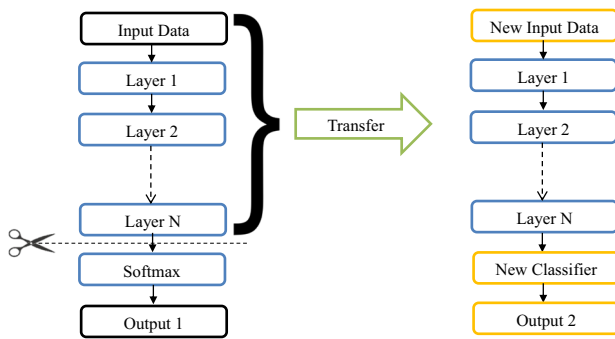
Fig. 1 An example of feature transfer from a pre-trained model in TL

fact that each problem might need a new dataset and high cost of providing labeled dataset for supervised ML that give sufficient accuracy, the main challenge is to use the least amount of images possible to train such a model and able to monitor all aspects of aforementioned scribing quality.

Deep learning (DL) and convolutional neural network (CNN) showed promising performance among other ML methods in recent years in different contexts from autonomous driving to medical image analysis (Badrinarayanan et al., 2017; Sejnowski, 2019). However, these methods need significant amount of data for training a model with adequate accuracy (Bauer & Kohavi, 1999; Bosch et al., 2019; Li et al., 2014). Collecting image data may not be a big challenge, but pre-processing and labeling of these images for training is. This data preparation stage is the most time-consuming and costly step in any machine vision/DL applications. One way to alleviate this problem is the use of transfer learning (Li et al., 2017).

Existing supervised ML methods such as decision tree and other methods based on various trees such as Random Forest (RF) or Gradient Boosting Classifier (GBC) may not need as much training data as the DL/CNN models would but still require a large amount of data. In addition, these traditional ML methods may not be able to handle complicated problems such as semantic segmentation with multiple quality characteristics. In such a complicated problem, there might be several classes of objects to be classified or each might have different geometrical shape and color (Li et al., 2014). A new approach to alleviating the lack of labeled data problem is called Transfer Learning (TL) where the knowledge gained from a different and yet similar problem can be used to solve another problem. Figure 1 demonstrates how the knowledge could be transferred from a pre-trained model to a new model where Softmax layer is an extension version of logistics regression idea into a multi-class world.

In the last few years, several studies used image data to monitor laser-based manufacturing processes. Table 1 summarizes these studies, the research focuses, and their results. However, there is a lack of research for in-process moni-

toring of laser scribing quality. These quality characteristics include scribe width, debris, crack length, scribe depth, width of heat affected zone, and straightness. Also, most of the research is done using a specific experimental condition and one question is whether the results could be applied to other scribing conditions with different imaging systems. In this study we attempt to measure and monitor three important laser scribing characteristics (i.e., scribe width, debris, and straightness) using image data and a state-of-the-art transfer learning method. TDCNN will enable us to leverage existing deep learning models from different domains and accelerate classification with less amount of laser-scribing data.

Research methodology

Experimental setup

The experimental setup is depicted in Fig. 2, and the samples used throughout the study are $<100>$-oriented, 1-mm thick intrinsic Si wafers with a resistivity of $>200\ \Omega$ cm. An IR laser is used to scribe lines on the surface of a silicon wafer. Specifically, the laser source (MWTech, PFL-1550) has a wavelength of $\lambda = 1550$ nm and produces pulses of length $\tau = 3.5$ ns (full width at half-maximum) and can be operated at various repetition rates with a maximum pulse energy of $20\ \mu$J. The output beam has a $1/e^2$ diameter of 6 mm.

A half wave-plate in conjunction with a polarizing beam splitter is used to control the pulse energy by rotation of the wave-plate. The beam is then focused on to the surface of the Si sample by a microscope objective (NA = 0.85, Olympus, Model LCPLN100XIR) that is corrected for spherical aberration. At focus, the beam has a theoretical diameter at $1/e^2$ of $2w_0 = 1.22\lambda/NA = 2.2^-$m, with a Rayleigh length of $y_R = 2.6\ \mu$m in air. Parallel lines are scribed on the surface of the silicon samples. Considering three parameters in control and easy to change, we adopted a 2^3 factorial experimental design. Each factor is experimented on a high and a low value. Specifically, the low level and high level of the pulse energy are 1 and $2\ \mu$J respectively. The low and high levels for petition rates are 20 and 120 kHz. Finally, the low and high setting for scanned speed is 0.5 and 10 mm/s respectively. The scribing conditions are listed in Table 2. The images are obtained by separating long scribe lines from each condition listed in Table 2 into multiple small segments.

Image data and pre-processing

The first step of image processing for object identification is image segmentation. This segmentation task is accomplished by an unsupervised ML model, which does not require a time-consuming labeling process. However, unsupervised ML has limited applications and are not suitable for problems that

Table 1 Applications and features addressed by the most related publications

|	Study	Year	Case	ML/Monitoring approach	Semantic segmentation	Debris	Scribe width	Scribe straightness	Other quality characteristics
1	Grasso et al	2017	SLM	Principal component analysis (PCA) and k-means clustering	No	No	No	No	Local overheating phenomena caused by a wrong heat transfer from the melt pool to the surrounding material
2	Grasso et al	2018	SLM	IsoData, Otsu's, Li's, Huang's and k-means, and multivariate control-charting	No	No	No	No	Detect unstable melting conditions since their early stage
3	Imani et al	2018	LPBF	Multifractal & spectral graph theoretic analysis	No	No	No	No	Size, count, and location of pores around 80% accuracy and images from nine samples
4	Imani et al	2018	LPBF	DNN (AlexNet)	No	No	No	No	Lack of fusion flaws; 362 images from one single part reach accuracy 92.5%
5	Roozbahani et al	2018	Laser scribing	A Particle analysis algorithm	No	No	No	No	Detecting discontinuance area in copper indium gallium selenide solar panels
6	Mayr et al	2018	Laser welding	CNN	No	No	No	No	Irregular weld seam, recessed weld seam, undercut, weld bead, and holes & spatters. 500 operations (images) with 512×512 resolution
7	Shevchik et al	2019	Laser welding	Graph support vector machine with data adaptive kernel	No	No	No	No	23 laser welds for the same events in row 9
8	Shevchik et al	2020	Laser welding	DNN	No	No	No	No	300 images in training and 100 in test set is used to reveal the unique signature of sub-surface events in wavelet spectrograms from the laser back-reflection and acoustic emission signals
9	Zhang et al	2020	Laser welding	DNN	No	No	No	No	Porosity monitoring using grayscale images and achieving 96.1% accuracy
10	Gonzalez-Val et al	2020	Laser welding	CNN	No	No	No	No	The model were able to detect defect parts with 97.5% accuracy
11	This study	2021	Laser scribing	TDCNN	Yes	Yes	Yes	Yes	The proposed model achieved 96% accuracy using a small set and 97% on large set
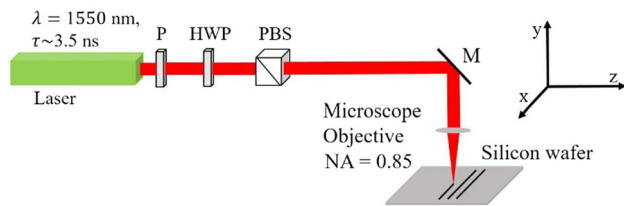
Fig. 2 Experimental setup. *P* Polarizer, *HWP* Halfwave plate, *PBS* Polarized beam splitter, *M* Mirror

need high accuracy. Figure 3(b) shows adaptive thresholding (AT) (Bradley & Roth, 2007) and Fig. 3(c) Otsu's thresholding (OT) (Otsu, 1979) methods, respectively. AT is a local intensity method while OT is a global intensity one. As shown in Fig. 3, AT performs better than OT. However, the segmentation is still very far from desirable for process monitoring. To monitor a process, a high level of segmentation accuracy is required. Neither method achieves this standard. Our goal is to measure and monitor scribe width, debris, and straightness. None of those goals can be achieved using these thresholding methods.

Image pre-processing can include renaming, resizing, denoising, segmenting, edge smoothing, and finally labeling. In this study, the collected image dataset is renamed, resized to 1024×1024, and labeled to the mentioned classes of scribe, debris, and the part background. The initial goal is to train

the model with sufficient accuracy and minimum amount of data. To do this, a total of 21 images from 8 different scribes are collected and labeled to three classes of debris, scribe, and silicon background. Note that this data sets were split into 3 sets of training, validation, and testing with the ratio of 60-20-20. However, a valid concern is testing the model is not reliable based on a handful of images, even if we get very high accuracy and low loss. To make sure the accuracy is reliable, we prepared a large dataset that includes 154 images with the same size and more variety in scribe size, camera zoom, and defects. The purpose of the second dataset is to verify model performance. Specifically, we define the validation set as the dataset held back from training to estimate the model's capability for tuning hyper parameters and the testing set is just as the part of dataset held back from the training set to give an unbiased estimation of the final trained model (Xie et al., 2011). However, by verification we want to ensure that the model won't misbehave on a broader range of circumstances (Ding et al., 2021). Thus, in this research, all the images were labeled carefully using MATLAB R2020a Image Labeler application manually. A pixel-wise region of interest is defined in the MATLAB Image Labeler application where we assigned 1 to all silicon background pixels, 2 to all scribe pixels, and 3 to all debris pixels. The labeled ground truth data was exported from MATLAB environment and then imported to Python for DL and image processing.

Transfer learning model architecture

To solve traditional machine learning issues pointed out in the previous section in image segmentation, we designed our TL-based model. Figure 4 shows the details of the designed architecture where blue part (i.e., the first two and half rows) represents the layers with weights transferred from the pre-trained VGG16 and the orange part (the rest of the rows) is the proposed CNN classifier built on top of the pre-trained model. TL works the best when a related pre-trained CNN can be used to transfer knowledge. There are several well-known pre-trained CNNs such as Xception, VGG16, VGG19, ResNet50, Inception, and MobileNet, which have

Table 2 Scribing conditions used in our experiment

Line number	Pulse energy (μJ)	Repetition rate (kHz)	Scanning speed (mm/s)
1	1	20	0.5
2	1	20	10
3	1	120	0.5
4	1	120	10
5	2	20	0.5
6	2	20	10
7	2	120	0.5
8	2	120	10

Fig. 3 a A sample include 2 scribes, **b** clustering result using adaptive thresholding (AT) method, **c** clustering results using Otsu's thresholding (OT) method
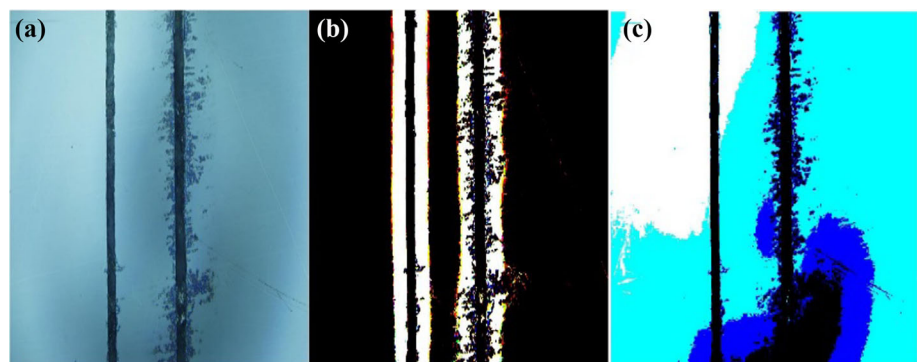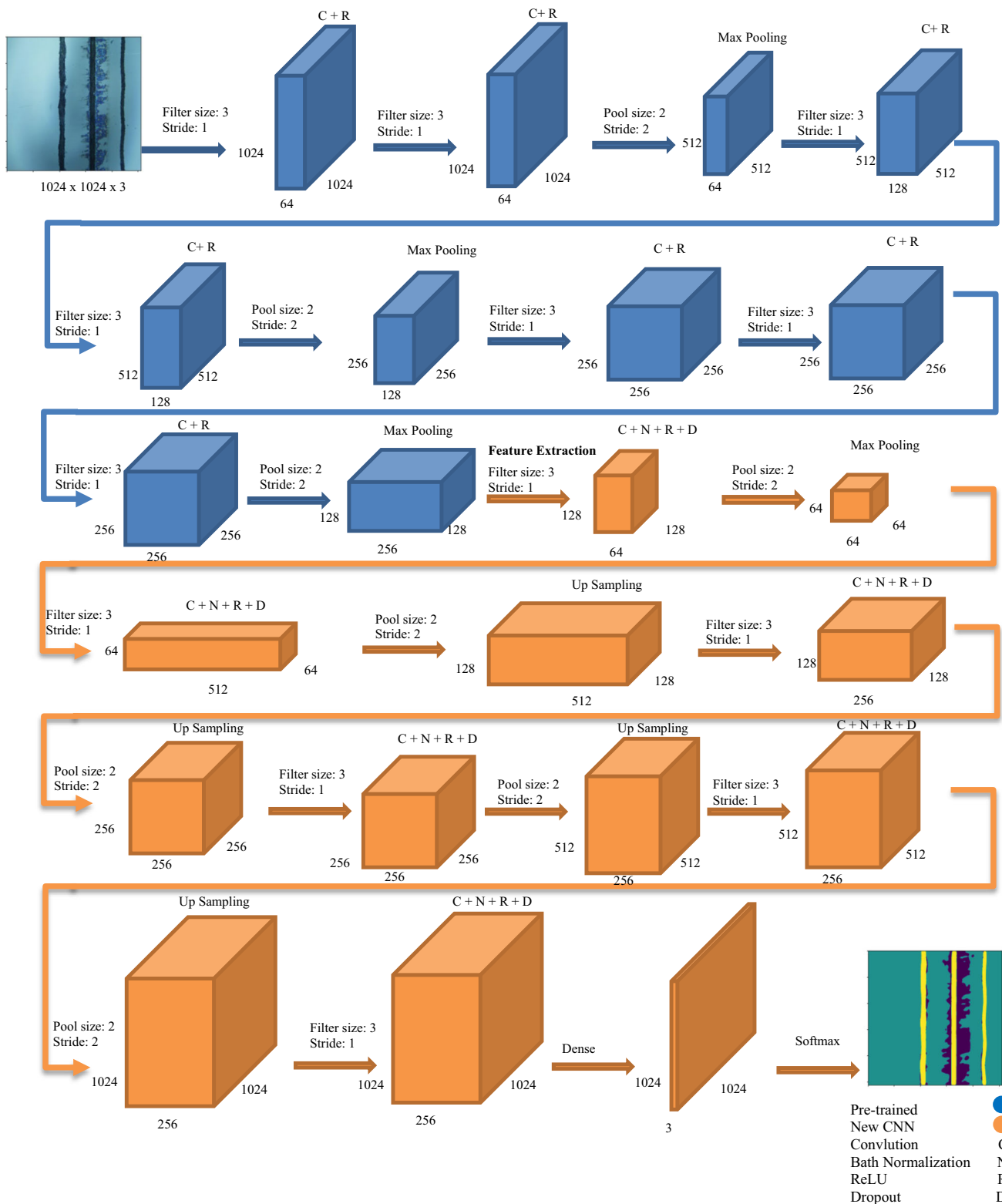
Fig. 4 The proposed TDCNN architecture

been trained over different public datasets like ImageNet and MNIST, and CIFAR (Mousavi et al., 2019; Noh et al., 2015a; Unnikrishnan et al., 2019; Uzkent et al., 2019; Zoph & Le, 2019). All these datasets are designed for object detection, of which goal is just to determine whether an image contain a specific object or not. However, to the best of our knowledge

we could not find any pre-trained pixel-wise semantic segmentation CNN. Among several aforementioned pre-trained models for image classification, VGG16 is a very deep CNN that is trained on part of ImageNet dataset with 2 million images and 1000 different class of objects such as animals, furniture, sports, plants, etc. (Ferguson et al., 2018). VGG16 has a high accuracy for the objects it was trained for. Thus, for this research, VGG16 is chosen to transfer the image feature knowledge.

Note that the original VGG16 is trained on images with a 224×224 resolution. The output of this model is one scalar value representing the classified object. However, the input, desired output, and consequently dimensions of all the layers need to be changed for different problems. In each problem, various resolutions for images can be used for training. Thus, the main adjustment needed for the proposed framework is to change the input dimension or image resolution. The topology of the proposed TDCNN is shown in Fig. 4. Working on appropriate fine-tuning and feature extraction, we trained the proposed TDCNN model. This new classifier can be a logistics regression or support vector machine model in case of binary classification, or deep CNN. Since the pre-trained model and the proposed model serve two purposes (the former is for object detection and the latter is for semantic segmentation), thus another deep CNN should be trained based on a new small dataset. To design this deep CNN, the idea of decoding and up-sampling in Unet (Ronneberger et al., 2015) and SegNet (Badrinarayanan et al., 2017) is adopted. In a similar study, pre-trained DeconvNet (Noh et al., 2015b) layers are transferred on top of VGG 16 for off-road autonomous driving without considering any batch normalization or drop outs (Sharma et al., 2019).

The last blue block in Fig. 4 is the output from the feature extraction process. The output of feature extraction from the pre-trained model is a matrix with dimensions of $128 \times 128 \times 256$ where 256 is the number of filters. This output (i.e., the output from the last blue block) is the input of the proposed CNN. Six blocks of convolutional layers are designed. Each block contains a convolutional layer, batch normalization, ReLU, dropout, and Max pooling (or unpooling). The dropout helps to reduce overfitting. There is only one max pooling layer connected right after the first convolutional block followed by four max unpooling layers for up-sampling weights to the desired dimension, which is 1024×1024 in this case. Finally, a fully connected layer followed by a Softmax layer gives the desired classification. Note that the weights in the blue part remain fixed during training. This is because those weights have been obtained with training from millions of images. The rationale of transfer learning is to leverage this knowledge of fundamental geometrical features for various objects. The orange part is the proposed module for the desired classification of a specific problem domain, in this case, the scribed images.

Despite the complex appearance of the proposed architecture, it is significantly less complicated than the other well-known CNNs like the VGG16. As mentioned above, the blue section in the architecture belongs to the transferred layers from the pre-trained VGG16. The blue part does not include all the trainable layers in the original VGG16. The weights in the transferred layers from VGG16 were unchanged during training. In the proposed TDCNN and based on filter size and number of layers, we only trained 5 million parameters in each epoch for the new module with 771 parameters in the last dense layer, which were 3 times less than those of VGG16.

Model training and evaluation

The architecture in Fig. 4 was coded on the Google Colaboratory using a P100 GPU and Tensorflow environment. Model evaluation was done by two sets of image data; a validation set and a testing set. The validation set was the part of the sample data held back from training to estimate model performance during tuning the model's hyper-parameters while the testing set was the part of the sample data held back from training to estimate the model's final performance. We split the entire small dataset with 21 images to 14 images for training, 3 for validation, and 3 for testing. It is necessary to emphasize that the main effort here is to train our model with a few numbers of images and reach desired accuracy. Since the training set was small, the test set was also small, and one might claim the result was not enough. To address this concern, a large data set including 154 images was created and trained later to verify the results and accuracy.

In the training phase, training and validation accuracy and losses were used to access model performance. Thus for the aforementioned model, using Adam optimizer (Jais et al., 2019) with a learning rate of 0.0001, and a mini batch size of 4 the model was trained for 30 epochs. Figure 5 demonstrates the training performance of the designed model. Note that a dropout value of 0.4 was used to prevent possible overfitting during training. Both training and validation indexes (i.e. accuracy and losses) in Fig. 5 are very close to each other in each epoch, especially closer to the final epochs. This observation demonstrates that our model was able to avoid overfitting successfully.

While Fig. 5 represents training and validation performance during the model training process, there is still a need to examine the final trained model accuracy on test set. To do this we calculated F1-score for each class and Overall Accuracy (OA) using confusion matrix. Given a general structure of confusion matrix with True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values in it, F1-score is defined as harmonic average of precision and recall. Equations 1, 2, 3, 4 give some details on how to calculate OA, precision, recall, and F1-score. OA
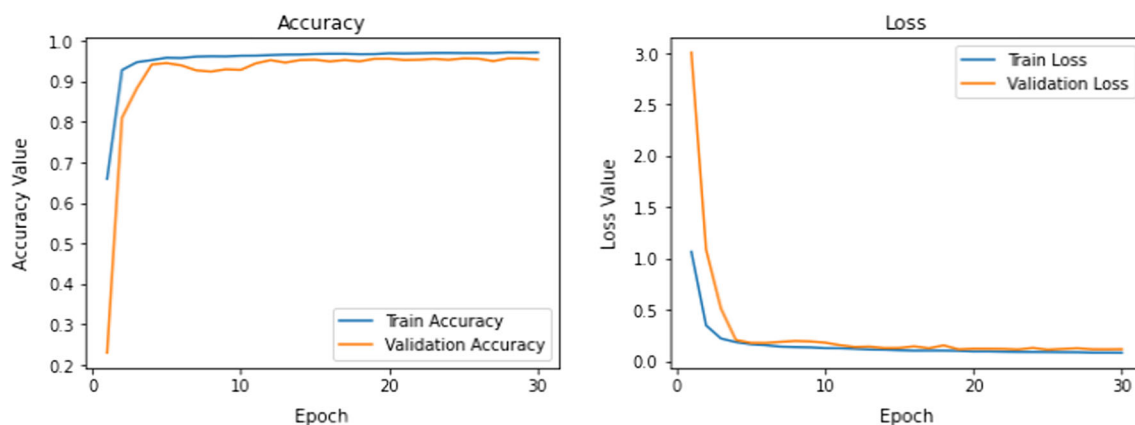
Fig. 5 Training and validation performance of TDCNN

is the correct identification rate. Precision is the rate of correct positive observations out of all observations identified as positive. High precision rates mean low false positive rates. Recall, on the other hand, is the ratio of correctly predicted positive observations out of all positive cases. High recall rates mean low false-negative cases. Finally, F1 score is the weighted average of precision and recall. In short, precision is a measure of false positives while recall is a measure of false negatives. F1 score is an overall measure of both. OA is the F1 score when all correct identifications of all categories are considered as a whole. All 4 metrics used here are the larger the better.

$$OA = \frac{TP + TN}{TP + TN + FN + FP} \tag{1}$$

$$Precision = \frac{TP}{FP + TP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{4}$$

Table 3 shows the OA and details of F-1 scores over the test set. The last column in Table 3 is the number of pixels that support each class. Precision here means, for example, 64 percent of those pixels classified as debris are actually debris. This outcome is expected because debris pixels are only 6.7% of all pixels examined and there are many none debris pixels mistakenly identified as debris. The recall value for debris means our model was able to find and classify 85 percent of all debris pixels. The recall for "Part background" is over 97 percent while the recall for scribed part is 94 percent. This means that the proposed model is able to find and classify all the pixels related to the part correctly with very high rates. This outcome also suggests that most of misclassifications are related to debris and scribe pixels. Debris has the lowest F-1 score. The fact of low precision and high recall values for the

Table 3 Overall accuracy and F1-score of testing the model using the first data set

	Precision	Recall	F1-n score	Support
Debris	0.64	0.85	0.73	281,939
Part background	0.99	0.97	0.98	3,574,063
Scribe	0.95	0.93	0.94	338,302
OA (Overall accuracy)			0.96	4,194,304

debris means the model was able to find most of the debris; however, there are also many non-debris pixels that were classified as debris. We suspect that the misclassification is from the scribe. Figure 7 provides the evidence since many debris are connected to the scribed line.

Since using 3 images in the test set (despite the high image resolution) might not be very reliable, we trained this model with 60 percent of images in the large dataset and the rest of them is used for testing and validation. Interestingly, this time accuracy even improved to 97 percent and loss value went below 0.1 for both training and validation. Table 4 shows the results of training the model and testing it. As can be seen, performance is just slightly better than training with the small set. The F-1 score for debris is still less than 80 percent. This is probably because of the high unbalance ratio. Also, considering Fig. 5 and the same graph in training the large dataset, there is not a big gap between the training and validation loss value. This observation confirms that our model was able to avoid overfitting.

Generation of scribe width and straightness of laser scribes

Given a sample scribe in Fig. 6, we can measure debris as well as scribe width and straightness. Trained models over the small and large sets had almost the same performance. In this section we use the model trained by the small set.

Table 4 Overall accuracy and F1-score of testing the model using the large dataset

	Precision	Recall	F1-score	Support
Debris	0.69	0.83	0.79	663,595
Part background	0.98	0.99	0.98	17,048,192
Scribe	0.96	0.98	0.97	3,259,733
OA (Overall accuracy)			0.97	20,971,520

Fig. 6 A scribed sample part showing measurement parameters
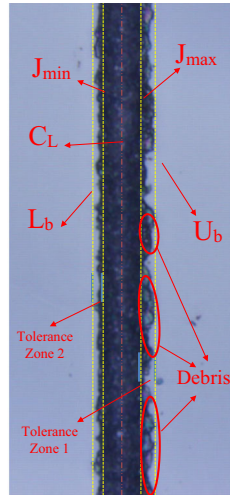
The output from TDCNN is a 2D image that three categories of scribe, debris, and sample part are classified in it (see Fig. 7(b)). Thus, debris can be measured directly by measuring the number of pixels that classified as debris. Using the same classified image, scribe width and straightness also could be measured using geometric dimensioning & tolerancing (GD&T) (*GD&T* Straightness, 2014; NADCA, 2015) with appropriate adjustments on the formulas based on the classified images.

After quantifying debris, the whole image will be segmented to scribe and no-scribe pixels. Figure 7 shows the image and pixel values for an example image of 3 scribe lines. In the initial classification in Fig. 7(b), the model assigns ones to all scribed pixels, zeros to all background pixel and twos to all debris. After measuring debris area and monitoring scribe width and straightness, all the debris pixels' value will be replaced with zeros to have a binary classification of scribe and no-scribe.

In Fig. 7(c) a pixel matrix is presented, in which one's (1's) represent all pixels that are classified as scribed, and zeros represent the background and debris. Let x_{ij} be the value of a pixel in row i and column j. In order to determine the width and straightness, the tolerance zones and center line C_L need to be identified. The width (W) is the distance between J_{max} and J_{min} which are the maximum and minimum positions in the effective area, respectively. The effective area consists of all column j's where the row sum of $j = \sum_{i=0}^{1023} x_{ij}$ is greater than 1024*(1-α). Here α is the classification error and 1024

is the image resolution. Thus, the effective area's boundary and width are:

$$J_{min} \leq \text{Effective area} \leq J_{max} \qquad (5)$$

And,

$$W = J_{max} - J_{min} \qquad (6)$$

Let J be all the columns that with value 1 (i.e., contain scribe). Then the center line (C_L) will be:

$$CL = \sum_{j_{min}}^{j_{max}} j / J \qquad (7)$$

Now define upper bound (U_b) max j that contains at least 1 pixel contain scribe and lower bound (L_b) min j that contains at least 1 pixel contain scribe. Thus:

Tolerance Zone 1 is between line

$$U_b \text{ and } J_{max} \qquad (8)$$

Tolerance Zone 2 is between line

$$J_{min} \text{ and } L_b \qquad (9)$$

Finally, if the total number of columns in the tolerance zone is n then for each row, the error (e_i) in the tolerance zone is:

$$ei = \sum_{all\ j\ in\ tolerance\ zone} |x_{ij} - \bar{x}_{ij}| / n \qquad (10)$$

and $\bar{x}_{ij}$ is the average x_{ij} in row i. Plotting e_i gives a clear idea about the straightness of the scribe.

Results and discussion

The proposed study uses different scribing conditions causing different quality issues such as debris, fluctuation, and very thin or very thick lines. The proposed TDCNN method was able to capture and help quantify the scribed lines. In the following sections, we will discuss these findings.

Debris measurement

Figure 7(a) is an original image from scribing 3 lines and Fig. 7 (b) shows the classified image after the use of the proposed TDCNN model. The green region in the classified image represents the background of the part, the yellow regions are scribes, and the purple ones are debris. Figure 7(c) shows the classification values around the middle line. Note
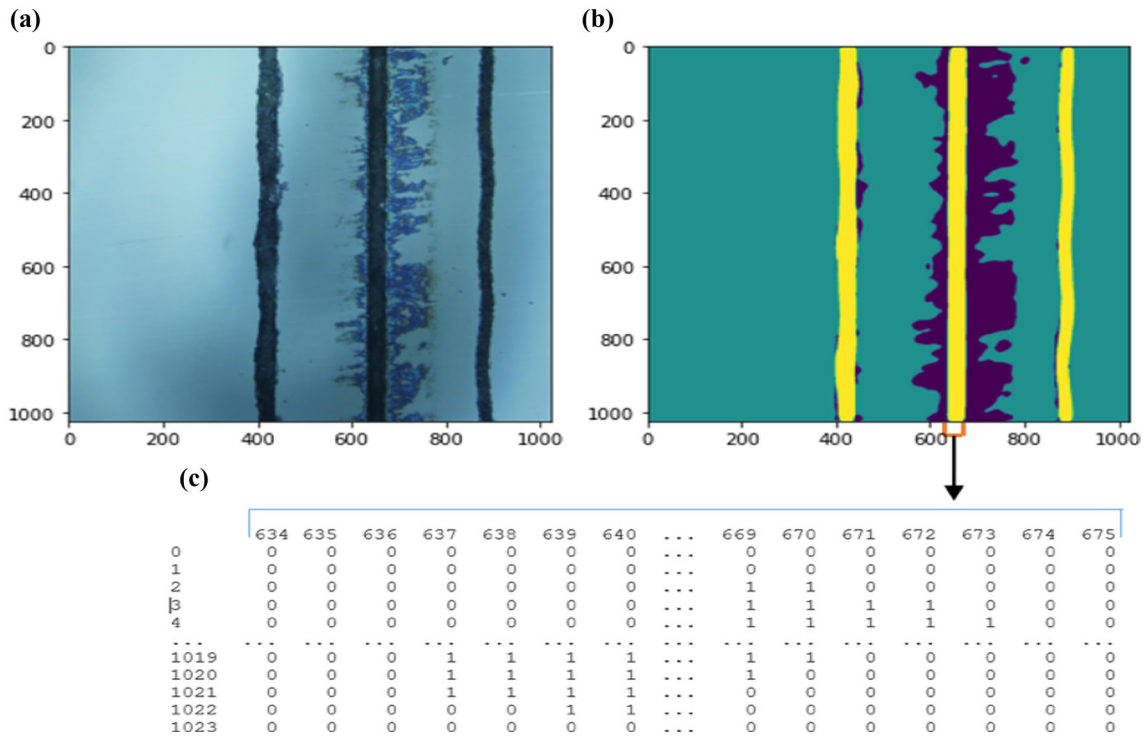
Fig. 7 **a** original scribed sample on top left; **b** classification using TDCNN on top right. **c** scribe classification values

that there are $1024 * 1024 = 2^{20}$ pixels in the image. Based on this classification, pixel counting shows that the 110,893 purple pixels are debris, the 833,790 greens are background, and the 103,893 yellows are scribes. This means 79.5 percent of the shape is background, 9.9 percent is scribe, and 10.6 percent is debris (mostly because of the second scribe line).

Figure 8 show straightness measurement plots for a normal (straight) line and a fluctuating line, respectively. Those plots were obtained after classifying the line using TDCNN and then quantifying straightness using Eqs. (5, 6, 7, 8, 9, 10). On top of each figure the original scribe can be seen and e_i is plotted for the first tolerance zone of each scribe.

The error value range for a straight line (Fig. 8(a)) was between 0 and 1. However, this range for a fluctuating line (Fig. 8(b)) was between 0 and 6. In addition, comparing Fig. 8 (a) and (b) shows that a straight line has a smoother e_i plot with lower variations.

Transferability to a new case

Most DL models are designed and trained based on the assumption that experimental and environmental conditions are consistent. However, these assumptions might not hold in real life. Changes in scribing parameters and imaging conditions potentially make significant differences in the final picture. To solve this problem, the proposed model needs to be retrained in order to obtain knowledge from a new

Table 5 Scribing conditions of the second group of lines

Line number	Pulse energy (mJ)	Spot size(mm)	Scanning speed (mm/s)
1	202	0.7	1.5
2	202	0.7	3.0
3	202	1.2	1.5
4	202	1.2	3.0
5	415	1.2	1.5
6	415	1.2	3.0
7	415	0.7	1.5
8	415	0.7	3.0

environment. In a model without transfer learning, we might need to use a big data set that includes new conditions to get appropriate accuracy. However, the proposed model can accomplish this task with only a handful of images.

In our research, to test this hypothesis, we provided a new dataset of images from a different laser and imaging conditions. This new dataset is collected from a group of lines scribed by a nanosecond laser with a wavelength of 1064 nm and a repetition rate of 10 Hz. Eight lines with different conditions were scribed. Table 5 shows the technical details about the second group of scribes.

To examine the trained TDCNN, we tried to classify the new images using the current trained model. Figure 9 shows the classification result on new samples without any change
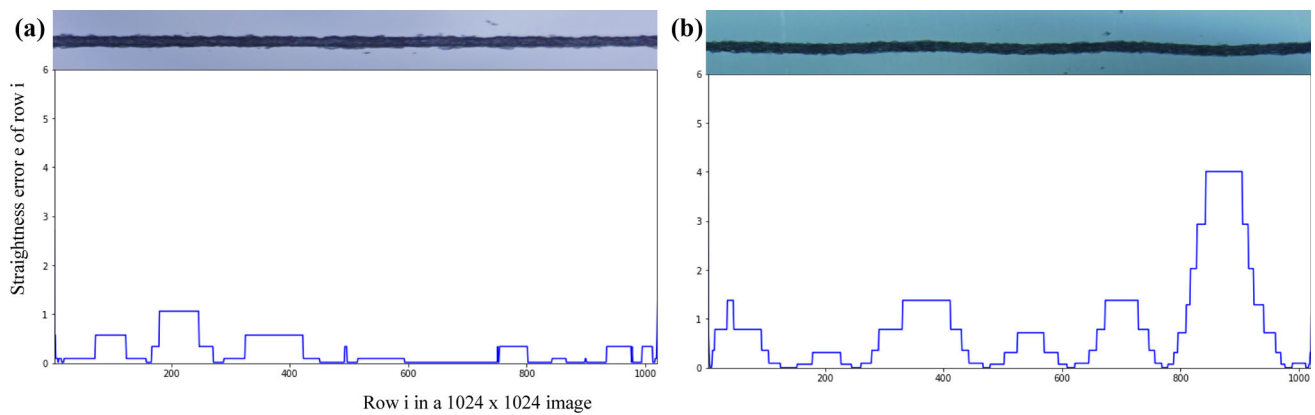
Fig. 8 Error (e_i) plot of **a** normal line and **b** fluctuating line
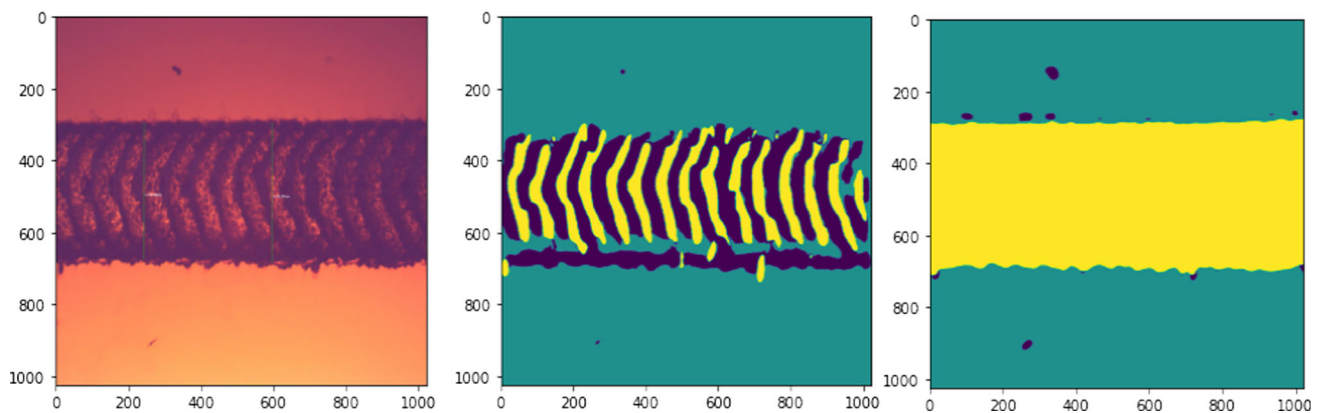

Fig. 9 Image classification on new dataset **a** original image with different imaging setup, **b** classification result without retraining **c** classification result with retraining

or retraining. Figure 9(a) is the original scribe image while Fig. 9(b) is the classification result. Note that the yellow part represents the scribe. Interestingly, the model was relatively successful in finding the background. However, we have a high rate of misclassification for the scribe. This misclassification presents a challenge for achieving our goal of automatic characterization of debris, scribe width, and scribe straightness.

To solve this problem, we retrained the same TDCNN but used additional 6 images to create a new dataset to train a new classification model. This way the model will learn extra knowledge in addition to the knowledge it already has. Network architecture and all the training hyper parameters (e.g., learning rate) were kept the same as the first training. We manually labeled the middle of the scribe as scribe rather than background or debris in the pixels in Fig. 9(a). Training and validation accuracy in retraining is over 98 percent which is even better than that of the first training result. Figure 9(c) shows the retraining performance on a new image. From this point, all the steps in "Generation of scribe width and straightness of laser scribes" section can be repeated and

Eqs. (5, 6, 7, 8, 9, 10) can be used to calculate scribe width and straightness.

Conclusions

In this study we proposed a novel deep transfer learning model to classify images from laser scribing with balanced performance for high accuracy and low overfitting. The classified images contain identified debris and scribes. Further image processing and algorithms are developed to quantify scribe width and straightness based on the classified images.

The proposed TDCNN has big advantages over a regular deep learning algorithm without transfer learning. First, while all regular deep learning models require a huge amount of image data and long training time for model accuracy, the proposed model requires only a few image samples for adapting new situations. This transfer learning feature in the proposed framework saves substantial effort and time in data preparation and labeling and leads to a much shorter time in the modeling and training phase. With the training of only 5 million parameters compared to the current well-known

architecture (e.g., VGG16 with more than 15 million trainable parameters), the proposed model has less complexity in the newly developed portion. Second, the proposed TL architecture is flexible in that adding new information based on new scribing conditions can be achieved with minimum effort and still retains the same performance. Usually working with a small dataset increases the chance of overfitting. In the proposed TDCNN architecture, we used several layers of batch normalization and dropout to overcome this challenge. These layers helped reduce (if not remove) overfitting.

The main idea in re-using pre-trained layers is transferring general knowledge and patterns such as edges, corners, dots etc. from other labeled images. Thus, only most common features are needed from the transferred layers, specifically, the number of filters were reduced to 64. Facing a new and specific problem, 512 filters were added to capture larger combinations of patterns. For future research, more complicated pattern combinations may be captured and studied by adding new layer modules.

The proposed TDCNN model enables the characterization of laser scribing quality using just a small number of sample images and reaches the accuracy as high as 96 percent. The scribe images in this study had mainly three features of concern: debris, width variation, and straightness. The quality measures on these characteristics pave the way to track all these features automatically and enable the possibility of real-time process control as a logical next step. Although the proposed method is able to measure and quantify debris with any scribe shape, quantifying straightness and width is only limited to vertical lines and further study is needed for non-vertical lines. This study only focuses on laser scribing on silicon wafers. We expect the same framework can be extended to different materials such as solar photovoltaic thin film. However, different quality issues such as cracks may arise that also require future research. Highly unbalanced data was a significant limitation in this study. We will tackle this issue using the other state-of-the-art methods such as Differentially Private Generative Adversarial Networks in future studies as well. Finally, the case study in "Transferability to a new case" section demonstrates that the trained TDCNN model can be transferred to a new case to improve its original performance. We expect the trained models may also be transferred to other laser related processes. Future research is much needed to confirm this hypothesis.

References

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet : A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(12), 2481–2495. https://doi.org/10.1109/TPAMI.2016.2644615

Bauer, E., & Kohavi, R. (1999). Empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine Learning, 36*(1), 105–139. https://doi.org/10.1023/a:1007515423169

Bosch, M., Foster, K., Christie, G., Wang, S., Hager, G. D., & Brown, M. (2019). Semantic stereo for incidental satellite images. *IEEE Winter Conference on Applications of Computer Vision (WACV), 2019*, 1524–1532. https://doi.org/10.1109/WACV.2019.00167

Bradley, D., & Roth, G. (2007). Adaptive thresholding using the integral image. *Journal of Graphics Tools, 12*(2), 13–21. https://doi.org/10.1080/2151237X.2007.10129236

Chua, Z. Y., Ahn, I. H., & Moon, S. K. (2017). Process monitoring and inspection systems in metal additive manufacturing: Status and applications. *International Journal of Precision Engineering and Manufacturing-Green Technology, 4*(2), 235–245. https://doi.org/10.1007/s40684-017-0029-7

Delli, U., & Chang, S. (2018). Automated process monitoring in 3D printing using supervised machine learning. *Procedia Manufacturing, 26*, 865–870. https://doi.org/10.1016/j.promfg.2018.07.111

Ding, J., Hu, X.-H., & Gudivada, V. (2021). A machine learning based framework for verification and validation of massive scale image data. *IEEE Transactions on Big Data, 7*(2), 451–467. https://doi.org/10.1109/TBDATA.2017.2680460

Ferguson, M., Ak, R., Lee, Y.-T. T., & Law, K. H. (2018). Automatic localization of casting defects with convolutional neural networks. *2017 IEEE International Conference on Big Data (BIG-DATA), December*, 1726–1735. https://doi.org/10.1109/bigdata.2017.8258115

Fotovvati, B., Wayne, S. F., Lewis, G., & Asadi, E. (2018). A Review on melt-pool characteristics in laser welding of metals. *Advances in Materials Science and Engineering, 2018*, 1–18. https://doi.org/10.1155/2018/4920718

GD&T Straightness. (2014). Geometric dimensioning and tolerancing (GD&T). https://www.gdandtbasics.com/straightness/

Gonzalez-Val, C., Pallas, A., Panadeiro, V., & Rodriguez, A. (2020). A convolutional approach to quality monitoring for laser manufacturing. *Journal of Intelligent Manufacturing, 31*(3), 789–795. https://doi.org/10.1007/s10845-019-01495-8

Grasso, M., Demir, A. G., Previtali, B., & Colosimo, B. M. (2018). In situ monitoring of selective laser melting of zinc powder via infrared imaging of the process plume. *Robotics and Computer-Integrated Manufacturing, 49*, 229–239. https://doi.org/10.1016/j.rcim.2017.07.001

Grasso, M., Laguzza, V., Semeraro, Q., & Colosimo, B. M. (2017). In-Process monitoring of selective laser melting: spatial detection of defects via image data analysis. *Journal of Manufacturing Science and Engineering*. https://doi.org/10.1115/1.4034715

Imani, F., Chen, R., Diewald, E., Reutzel, E., & Yang, H. (2019). Deep learning of variant geometry in layerwise imaging profiles for additive manufacturing quality control. *Journal of Manufacturing Science and Engineering*. https://doi.org/10.1115/1.4044420

Imani, F., Gaikwad, A., Montazeri, M., Rao, P., Yang, H., & Reutzel, E. (2018a). Layerwise in-process quality monitoring in laser powder bed fusion. *Additive Manufacturing Bio and Sustainable Manufacturing*. https://doi.org/10.1115/MSEC2018a-6477

Imani, F., Gaikwad, A., Montazeri, M., Rao, P., Yang, H., & Reutzel, E. (2018b). Process mapping and in-process monitoring of porosity in laser powder bed fusion using layerwise optical imaging. *Journal of Manufacturing Science and Engineering, Transactions of the ASME*. https://doi.org/10.1115/1.4040615

Jais, I. K. M., Ismail, A. R., & Nisa, S. Q. (2019). Adam optimization algorithm for wide and deep neural network. *Knowledge Engineering and Data Science, 2*(1), 41.

Ku, S., Pieters, B. E., Haas, S., Bauer, A., Ye, Q., & Rau, U. (2013). Electrical characterization of P3 isolation lines patterned with a UV laser incident from the film side on thin-film silicon solar cells. *Solar Energy Materials and Solar Cells, 108*, 87–92. https://doi.org/10.1016/j.solmat.2012.09.017

Leitz, K.-H., Redlingshöfer, B., Reg, Y., Otto, A., & Schmidt, M. (2011). Metal ablation with short and ultrashort laser pulses. *Physics Procedia, 12*, 230–238. https://doi.org/10.1016/j.phpro.2011.03.128

Li, L., Wu, Y., & Ye, M. (2014). Multi-class image classification based on fast stochastic gradient. *Informatica, 38*(145), 153.

Li, X., Zhang, L., Du, B., Zhang, L., & Shi, Q. (2017). Iterative reweighting heterogeneous transfer learning framework for supervised remote sensing image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10*(5), 2022–2035. https://doi.org/10.1109/JSTARS.2016.2646138

Mayr, A., Lutz, B., Weigelt, M., Glabel, T., Kibkalt, D., Masuch, M., Riedel, A., & Franke, J. (2018). Evaluation of machine learning for quality monitoring of laser welding using the example of the contacting of hairpin windings. *2018 8th International Electric Drives Production Conference (EDPC)*, 1–7. https://doi.org/10.1109/EDPC.2018.8658346

Mousavi, H. K., Nazari, M., Takac, M., & Motee, N. (2019). Multi-agent image classification via reinforcement learning. *IEEE International Conference on Intelligent Robots and Systems*. https://doi.org/10.1109/IROS40897.2019.8968129

NADCA. (2015). NADCA Product Specification Standarts for Die Casting. In *North American Die Casting Association*. http://www.caldiecast.com/docs/Zinc-and-ZA-Alloy-Data.pdf

Najjartabar-Bisheh, M., Chang, S. I., & Lei, S. (2021). A layer-by-layer quality monitoring framework for 3D printing. *Computers & Industrial Engineering*. https://doi.org/10.1016/j.cie.2021.107314

Noh, H., Hong, S., & Han, B. (2015a). Learning deconvolution network for semantic segmentation. Proceedings of the IEEE International Conference on Computer Vision, 2015a Inter, 1520–1528. https://doi.org/10.1109/ICCV.2015a.178

Noh, H., Hong, S., & Han, B. (2015b). Learning deconvolution network for semantic segmentation. *IEEE International Conference on Computer Vision (ICCV), 2015*, 1520–1528. https://doi.org/10.1109/ICCV.2015.178

Otsu, N. (1979). Threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*. https://doi.org/10.1109/TSMC.1979.4310076

Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-Net: convolutional networks for biomedical image segmentation*. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

Roozbahani, H., Marttinen, P., & Salminen, A. (2018). Real-time monitoring of laser scribing process of cigs solar panels utilizing high-speed camera. *IEEE Photonics Technology Letters, 30*(20), 1741–1744. https://doi.org/10.1109/LPT.2018.2867274

Roozbahani, H., Salminen, A., & Manninen, M. (2017). Real-time online monitoring of nanosecond pulsed laser scribing process utilizing spectrometer. *Journal of Laser Applications, 29*(2), 022208. https://doi.org/10.2351/1.4983520

Roozbahani, Hamid, Salminen, A., & Manninen, M. (2019). *Real-time monitoring of laser scribing process utilizing high-speed camera*. *2310*, 2310. https://doi.org/10.2351/1.5118570

Sejnowski, T. J. (2019). Neural Information Processing Systems. *The Deep Learning Revolution*. https://doi.org/10.7551/mitpress/11474.003.0014

Sharma, S., Ball, J. E., Tang, B., Carruth, D. W., Doude, M., & Islam, M. A. (2019). Semantic segmentation with transfer learning for off-road autonomous driving. *Sensors (switzerland), 19*(11), 1–21. https://doi.org/10.3390/s19112577

Shevchik, S. A., Le-Quang, T., Farahani, F. V., Faivre, N., Meylan, B., Zanoli, S., & Wasmer, K. (2019). Laser welding quality monitoring via graph support vector machine with data adaptive kernel. *IEEE Access, 7*, 93108–93122. https://doi.org/10.1109/ACCESS.2019.2927661

Shevchik, S., Le-Quang, T., Meylan, B., Farahani, F. V., Olbinado, M. P., Rack, A., Masinelli, G., Leinenbach, C., & Wasmer, K. (2020). Supervised deep learning for real-time quality monitoring of laser welding with X-ray radiographic guidance. *Scientific Reports, 10*(1), 3389. https://doi.org/10.1038/s41598-020-60294-x

Unnikrishnan, A., Sowmya, V., & Soman, K. P. (2019). Deep learning architectures for land cover classification using red and near-infrared satellite images. *Multimedia Tools and Applications, 78*(13), 18379–18394. https://doi.org/10.1007/s11042-019-7179-2

Uzkent, B., Sheehan, E., Meng, C., Tang, Z., Burke, M., Lobell, D., & Ermon, S. (2019). Learning to interpret satellite images using wikipedia. *IJCAI International Joint Conference on Artificial Intelligence, 2019-Augus*, 3620–3626. https://doi.org/10.24963/ijcai.2019/502

Wang, X., Yu, X., Berg, M. J., Chen, P., Lacroix, B., Fathpour, S., & Lei, S. (2021). Curved waveguides in silicon written by a shaped laser beam. *Optics Express, 29*(10), 14201. https://doi.org/10.1364/OE.419074

Wang, X., Yu, X., Berg, M., DePaola, B., Shi, H., Chen, P., Xue, L., Chang, X., & Lei, S. (2020). Nanosecond laser writing of straight and curved waveguides in silicon with shaped beams. *Journal of Laser Applications, 32*(2), 022002. https://doi.org/10.2351/1.5139973

Wang, X., Yu, X., Shi, H., Tian, X., Chambonneau, M., Grojo, D., DePaola, B., Berg, M., & Lei, S. (2019). Characterization and control of laser induced modification inside silicon. *Journal of Laser Applications, 31*(2), 022601. https://doi.org/10.2351/1.5096086

Xie, X., Ho, J. W. K., Murphy, C., Kaiser, G., Xu, B., & Chen, T. Y. (2011). Testing and validating machine learning classifiers by metamorphic testing. *Journal of Systems and Software, 84*(4), 544–558. https://doi.org/10.1016/j.jss.2010.11.920

Yao, Y. L., Chen, H., & Zhang, W. (2005). Time scale effects in laser material removal: A review. *The International Journal of Advanced Manufacturing Technology, 26*(5–6), 598–608. https://doi.org/10.1007/s00170-003-2026-y

Yuan, B., Giera, B., Guss, G., Matthews, I., & Mcmains, S. (2019). Semi-supervised convolutional neural networks for in-situ video monitoring of selective laser melting. *IEEE Winter Conference on Applications of Computer Vision (WACV), 2019*, 744–753. https://doi.org/10.1109/WACV.2019.00084

Zhang, B., Hong, K.-M., & Shin, Y. C. (2020). Deep-learning-based porosity monitoring of laser welding process. *Manufacturing Letters, 23*, 62–66. https://doi.org/10.1016/j.mfglet.2020.01.001

Zhao, X., Cao, Y., Nian, Q., Shin, Y. C., & Cheng, G. (2014). Precise selective scribing of thin-film solar cells by a picosecond laser. *Applied Physics A, 116*(2), 671–681. https://doi.org/10.1007/s00339-014-8330-6

Zoph, B., & Le, Q. V. (2019). Neural architecture search with reinforcement learning. *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*, 1–16.