Flaw Detection in Metal Additive Manufacturing Using Deep Learned Acoustic Features

Wentai Zhang*
Carnegie Mellon University
wentaiz@andrew.cmu.edu

Brandon Abranovic*
Carnegie Mellon University
babranov@andrew.cmu.edu

Jacob Hanson-Regalado
University of California, Berkeley
jhanreg11@gmail.com

Can Koz
canxkoz@gmail.com

Bhavya Duvvuri
Carnegie Mellon University
bduwuri@andrew.cmu.edu

Kenji Shimada
Carnegie Mellon University
shimada@cmu.edu

Jack Beuth
Carnegie Mellon University
beuth@andrew.cmu.edu

Levent Burak Kara
Carnegie Mellon University
lkara@cmu.edu

Abstract

While additive manufacturing has seen rapid proliferation in recent years, process monitoring and quality assurance methods capable of detecting micro-scale flaws have seen little improvement and remain largely expensive and time-consuming. In this work we propose a pipeline for training two deep learning flaw formation detection techniques including convolutional neural networks and long short-term memory networks. We demonstrate that the flaw formation mechanisms of interest to this study, including keyhole porosity, lack of fusion, and bead up, are separable using these methods. Both approaches have yielded a classification accuracy over 99% on unseen test sets. The results suggest that the implementation of machine learning enabled acoustic process monitoring is potentially a viable replacement for traditional quality assurance methods as well as a tool to guide traditional quality assurance methods.

1 Introduction

The fabrication of complex metal parts via additive manufacturing (AM) processes such as laser powder bed fusion (LPBF) has seen rapid proliferation in recent years and has been widely adopted in industries such as medicine and aerospace [1]. Despite this, the process still suffers from several key limitations that hinder further adoption. Chiefly among these, quality assurance remains time consuming and expensive. For this flaw type classification task, basic machine learning classifiers, such as SVMs and decision trees, tend to be computationally expensive when applied to large training datasets. In addition, performance relies heavily on predefined feature extraction. To overcome this challenge, we propose the use of a deep learning framework that allows for the automation of data-driven feature extraction that relies upon a dataset generated with a fixed simple geometry and parameters selected using the process mapping technique to ensure a balanced, high quality dataset.

In metals AM, a key objective of quality assurance efforts is to determine if a built part contains process-induced defects that jeopardize its structural performance. CT scanning requires expensive and large equipment, and adds complex additional steps to the the production process. Additionally,

*Equal contributions.

Workshop on machine learning for engineering modeling, simulation and design @ NeurIPS 2020

the flaw formation mechanism that causes the identified defects must be diagnosed by the machine operator, and therefore requires expertise on the part of the machine operator. Being a vision based approach, the use of high resolution IR techniques is limited by the resolution of the camera used. Existing approaches which apply this method have shown that it is unable to detect defects smaller than $600\,\mu\text{m}$ [2]. High speed camera techniques are limited by both their small field of view in addition to the large digital file sizes which result from the data collection process [3]. Finally, existing acoustic monitoring methods are limited by their ability to detect the flaw formation mechanism responsible for the identified flaw. These methods employ a simple approach to detect the evolution of flaws by comparing the acoustic emission to a "known good" sample signal [4, 5]. Since state of the art methods are hindered either by cost, complexity, robustness, flaw formation mechanism identification, an acoustics based deep learning process monitoring method is proposed. This in-situ method utilizes small, low cost equipment that is capable of detecting not only the evolution of micro scale anomalies, but also the mechanism by which they occur.

This work has far reaching implications in quality assurance which include knowledge of where (what layer) and why flaws occur in addition to the knowledge of when more in-depth quality control is needed and where to target it. The work presented can even be extended to process control to dynamically change processing parameters to correct for flaw formation as it occurs. We accomplished the detection of flaw formation mechanisms via two separate deep learning approaches: A convolutional neural network (CNN) and a long short term memory neural network (LSTM). Both methods allow for fitting a precise non-linear boundary with large amount of training data. The acoustic data are viewed as spectrogram images when feeding to the proposed CNN, and the problem is analyzed in a 2D space, time and frequency. Conversely in the implementation of the LSTM approach, the input data to the model is a sequence of raw time series acoustic emissions. The dependencies of historical data are also taken into account in predicting the flaw formation types.

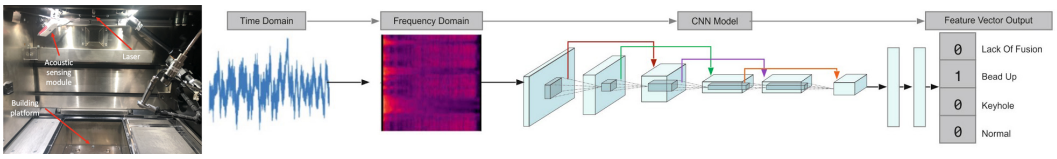


Figure 1: The pipeline of our proposed method.

The developed methods are effective and take advantage of the benefits inherent to acoustic monitoring. These benefits include, but are not limited to, high speed, manageable file sizes, and easy set-up. Our main contributions to the advancement of the field include an efficient data generation method that results in a large balanced dataset, a deployable, novel, potentially in-situ, fast laser powder bed monitoring platform demonstrated in Fig. 1. A convolutional neural network and a long short term memory neural network are trained to classify the flaw types with high accuracy.

2 Background

This section presents existing research focused on part quality assurance through control of process parameters, and the methods developed which are central to method of analysis of acoustic data presented in this work.

2.1 Process mapping

In metals AM, process mapping correlates the primary variables of the LPBF process such as laser power and laser speed to resulting part characteristics such as melt pool dimensions and the evolution of micro scale flaws under transient and steady-state conditions [6]. As such, the technique offers a means to understand limits of relevant processing variables that yield desirable outcomes and can serve to establish regions in processing space where flaws such as lack of fusion porosity (LOF), keyhole porosity (KH), and bead-up (BU) are likely to occur [7, 8]. In fixing other relevant processing variables such as geometry, preheat, and beam diameter, the dependence of melt pool dimensions can be simplified to depend on only absorbed power and laser travel velocity. The resultant melt pool dimensions can then be correlated to flaw formation for reliable flaw prediction in the simple case earlier described. Flaw formation regions in processing space were defined based on the work of

Francis et al. [8] and these specified regions were then used to guide the selection of parameters for the work described here. The relevant processing map is shown in Fig. 2 below.

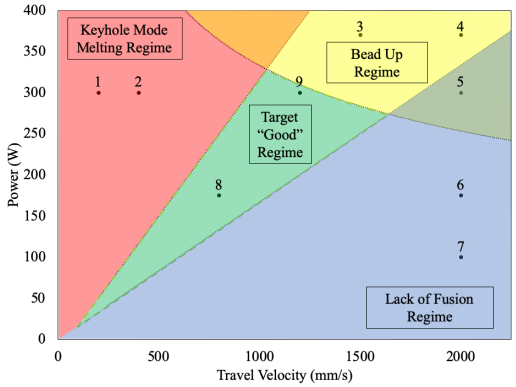


Figure 2: Selected parameter sets mapped to Ti-6Al-4V processing space.

2.2 CNN and LSTM models

Deep learning methods, specifically convolutional neural network (CNN) architectures, have proven to be very effective in various audio classification tasks. Guzhov et al. [9] show that CNN architectures such as ResNet [10] can be powerful in classifying environmental audio signals by first converting them to spectrograms. Feng et al. [11] utilize the same approach with ResNet when classifying background noise in audio scenes. Bian et al. [12] show that Densely Connected Neural Networks (DenseNet [13]) along with ResNets can also be strong classifiers for musical genres given audio. Finally, Palanisamy et al. [14] show that CNN architectures pretrained on the ImageNet dataset can serve as strong starter models for audio classification tasks. We build on these ideas and utilize ResNet and DenseNet pretrained on the ImageNet database to classify audio spectrograms for acoustic monitoring.

Apart from CNN, long short-term memory networks (LSTMs) have been increasingly popular in the acoustic analysis tasks. LSTMs have been applied to a wide range of time series tasks like accelerometer studies [15], human activity recognition [16, 17, 18, 19], diagnosis classification based on clinical data [20], process monitoring [21] etc. While approaches such as Markov models, conditional random fields, and Kalman filters have been extensively used for learning from sequential data, they are ill-equipped to learn long-range dependencies. LSTMs can directly learn from the raw time series, hence can avoid dependence on domain expertise to manually engineer input features. The model can learn an internal representation of the time series data and ideally achieve comparable performance to models fit on a version of the dataset with engineered features [21]. Therefore, LSTM is introduced in the analysis to explore the dependencies among the data in time series.

3 Methodology

3.1 Spectral Image Approach

Using the methods to record and process audio data detailed in previous sections, we apply a convolutional neural network (CNN) architecture to classify the flaw formation mechanisms present. First, we transform the audio signal to spectrograms resulting in a 2D representation of the audio signal that encodes the audio intensity of across different frequencies over time. To do this, we first slice the audio data into 1.92 second segments using a sliding window with a step size of 96 milliseconds. Then, we use Librosa's spectrogram generation utilities to create a 128 x 80 spectrogram matrix with decibel units for each audio slice. Finally, we use cubic interpolation and Seaborn's heatmap tool to save each spectrogram matrix as an image file of size 1024 x 1024 pixels. Samples of these spectrograms for each of our target classes are shown in Figure 3.

CNNs are a family of machine learning models used in vision processes that take in images (or any other >1D matrix), and perform convolutions on them to understand them spatially prior to

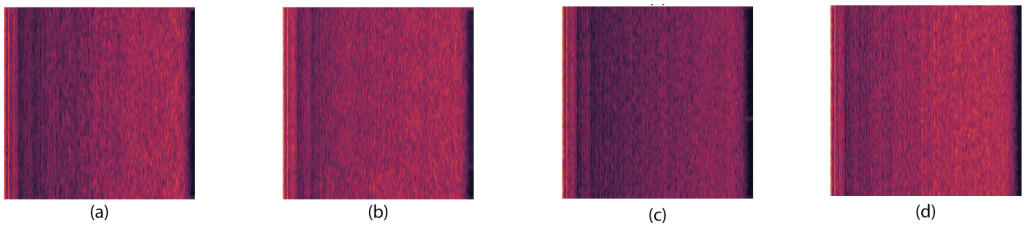


Figure 3: Spectrograms for all 3 flaw formation mechanisms and the normal case. (a)Keyhole, (b)Normal, (c)Lack of Fusion, (d)Bead Up.

completing tasks such as classification or regression. Residual Neural Networks (ResNets) and Densely Connected Neural Networks (DenseNets) are very large CNN architectures that utilize skip connections between convolutional layers to improve their performance. Using the previously generated spectrograms, we used PyTorch to load pre-trained versions of ResNet18, ResNet34, and DenseNet121 CNN models (originally trained on the 1000-class database, ImageNet). We then replaced each network's final fully-connected layer with a 4 node fully-connected layer, with each node representing 1 of our 4 output classes. For training, we utilized the Adam optimizer with a learning rate of 0.0001. Each model was trained for 10 epochs, and was compared against each other based on their resultant confusion matrices and test accuracy scores.

3.2 Time Series Approach

Unlike the spectral image approach, the LSTM model can directly make a prediction based solely on time series raw acoustic data. In this study, training samples are prepared using a sliding window approach across the raw acoustic data, with a set window size and corresponding window class label. To complete a parametric study, the window size ranged from 50 to 50000, which roughly corresponds to 0.0023 to 2.3 seconds in physical time.

The architecture of our LSTM model consists of the input layer, the memory layers and the output layer. The size of the input layer is the same as the window size specified in the data preparation step if the window size is less than 500. For data with a larger window size, we downsampled the data into a series of length 500 by interpolation. There are 3 LSTM memory layers with 190 LSTM cells in each layer. After passing through the recurrent layers, the intermediate output is then passed through a dropout layer with 35% drop probability. The output layer size is 4, which corresponds to the four classes in this task. A Softmax activation function was applied to the output layer. For training, we utilized the Adam optimizer with a learning rate of 0.005. All models were trained for 100 epochs.

4 Results

We find that in each approach, models can successfully identify the flaw formation mechanism present in a sample with comparable accuracy. Apart from accuracy, prediction speed is another important metric we consider for effective real-world applications of these models. For real time classification, the model must be capable of performing on-the-fly prediction with very low latency. On average, CNN models outperformed the LSTM models both in classification accuracy and in their total speed in converting raw audio samples to a classification vector.

4.1 Results for Spectral Image Approach

Out of the 3 CNN models we trained, we observed extremely similar results. Each model converged to near 100% accuracy after less than 7 training iterations over 4800 training images. Figure 4 shows the accuracy (left) and average cross entropy loss per batch (right) after each training epoch. The solid lines represent each models performance on the training set itself, while the dashed lines represent each model's performance on a validation set of 300 images that the model cannot optimize itself with.

After training each model for 10 epochs, we evaluated its final accuracy on another unseen data set consisting of 225 randomly selected images from each class summing up to 900 images in total.

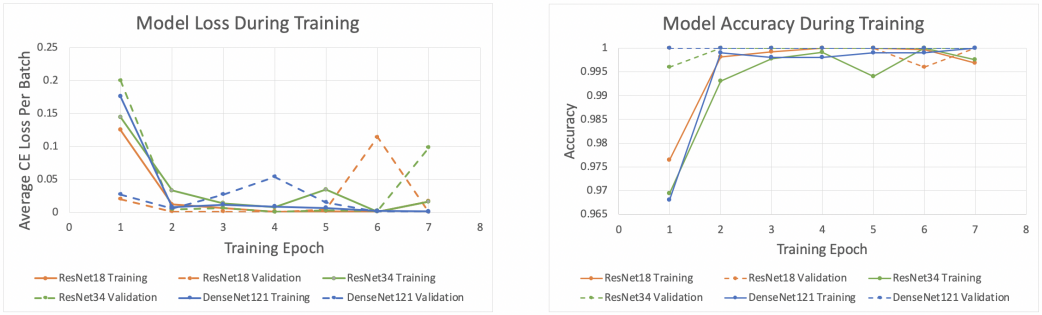


Figure 4: Loss/Accuracy On the Training/Validation data sets for ResNet18, ResNet34, and DenseNet121

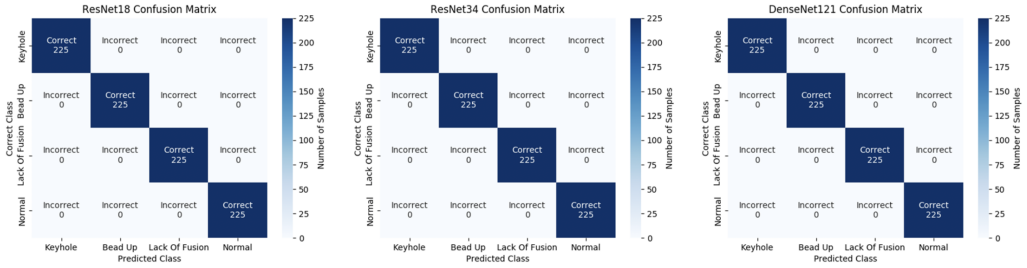


Figure 5: Confusion Matrices for each model on a 900 image test set. From left to right, the confusion matrices are for ResNet18, ResNet34, DenseNet121.

Figure 5 shows the result of these tests in the form of a confusion matrix where each row represents the correct output class for a given image and each column represents the predicted class. As can be seen, each model scores perfect accuracy, suggesting that each has very strong classification abilities.

In an effective real world implementation of this algorithm, our CNN pipeline would need to make a prediction within the signal sampling time. Since the time demand of the pre-processing step is the same no matter what model is used, the most effective way to decrease time duration would be to use the most efficient model that still achieves strong results. As such, we conducted a parametric study providing each model 1000 randomly generated 512x512 images and compared the average amount of time it took to come up with a prediction using a Nvidia GeForce 1080 GPU. Since the size of a model would also play a role in situations where memory is limited, we also compared the number of trainable parameters in each model to get a sense of which one consumed the least memory. Table 1 shows the prediction speed and number of trainable parameters for each model studied. While the number of parameters in DenseNet121 may be less than others, its prediction speed is still heavily negatively impacted because it includes many other model components such as extra skip connections that take time/space.

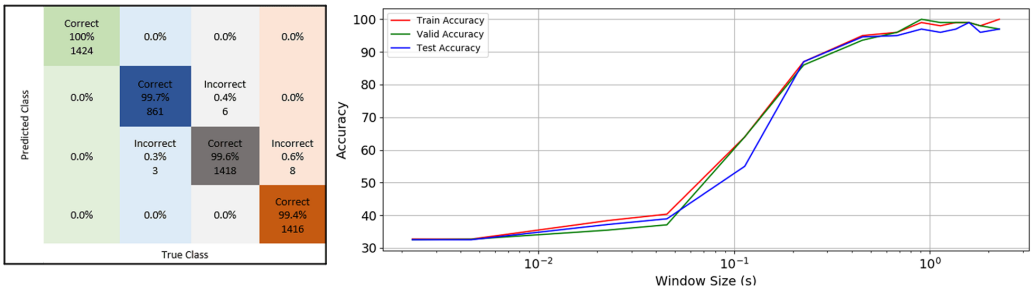
Model	Average Prediction Speed (seconds)	Number of Trainable Parameters
ResNet18	0.004	11689512
ResNet34	0.007	21797672
DenseNet121	0.02	7978856

Table 1: Average and number of trainable parameters of ResNet18, ResNet34, and DenseNet121.

As is shown, the ResNet18 model provides the best tradeoff between model efficiency and accuracy. To further understand the model's strength as a classifier we also tested how confident it was in its predictions. To do this, we used our results from the test data set and applied the softmax function to each class vector so that each element would represent a probability summing to 1. We then examined the average and standard deviation of the prediction probability for the ground truth class. Table 2 shows the results of this study for the ResNet18 model.

Ground Truth Class	Average Probability	Standard Deviation
Keyhole	0.99997	0.00005
Bead Up	0.99964	0.004
Lack of Fusion	0.99999	0.00001
Normal	0.99998	0.00003

Table 2: Average/Standard Deviation of Prediction Confidence for ResNet18 on unseen test data set



(a) (b)

Figure 6: (a) The confusion matrix of the best LSTM model. (b) The accuracy curves for LSTM models with different windows size for the input data

4.2 Results for Time Series Approach

Through the experiments with the LSTM models, trained with acoustic data of various durations, we wanted to explore a minimum characteristic window size that yields good performance of our proposed model. This window size was selected because it is substantial enough to provide context for the acoustic signal of the induction of the flaw. As is shown in figure 6(b), the test accuracy generally increases as the window size increases as well. However, there are three distinct phases in the curve. In the first phase (window size$< 0.05s$), the dependency in the data is greatly affected by the high frequency noise and the accuracy is extremely low($< 40\%$). In the second phase ($0.05s <$window size$< 1.0s$), more context information is provided to the classifier and contributes to the distinction of the induction of the flaws. The test accuracy increases rapidly and reaches over 95%. In the third phase, the further increase in the window size makes little difference in model performance, but it does result in much higher latency due to increased data processing time. So we identify that the characteristic window size should be between 1 to 2 seconds in order to maintain a good balance between the prediction accuracy and model latency. For all trained LSTM models, the average prediction time for a training sample is less than 0.09 seconds. The best model in terms of the test accuracy is shown in Fig. 6 (a), which exceeds 99% with window size of 1.5 seconds. Both CNN and LSTM models achieve comparably decent performance on the classification of the flaw types, while LSTM models are not as computationally efficient and take longer time to make each prediction. As such, our final proposed model is the CNN model with ResNet18 architecture demonstrated above.

5 Conclusion

Two different deep learning process monitoring techniques, including a spectral image approach and a time series approach, were developed and evaluated in this work. Both models achieved a high level of accuracy on unseen test sets. In consideration of model complexity and processing latency, the spectral image approach with ResNet18 is our final proposed model for identifying the flaw types for the LPBF AM of Ti-6Al-4V. The time dependency in the data was also explored through a parametric study on the effect of window size for the data preprocessing on the final LSTM model performance. The results indicated that the minimum characteristic time for this flaw detection task is about 1 second. Extensively, we want to test our model performance on more practical builds with complex geometries. The robustness and reliability for long-term in-situ monitoring can be tested and evaluated through these experiments. Although the current experiments are relatively simple in terms of the build geometry and laser scan strategy, the presented work can serve as the foundation for a practical, robust monitoring system for detecting flaw formation in LPBF AM builds.

References

[1] S. K. Everton, M. Hirsch, P. I. Stavroulakis, R. K. Leach, A. T. Clare, Review of in-situ process monitoring and in-situ metrology for metal additive manufacturing, Materials and Design 95 (2016) 431–445. doi:10.1016/j.matdes.2016.01.099.
URL http://dx.doi.org/10.1016/j.matdes.2016.01.099

[2] J. Mireles, S. Ridwan, P. A. Morton, A. Hinojos, R. B. Wicker, Analysis and correction of defects within parts fabricated using powder bed fusion technology, Surface Topography: Metrology and Properties 3 (3) (2015) 034002.

[3] L. Scime, B. Fisher, J. Beuth, Using coordinate transforms to improve the utility of a fixed field of view high speed camera for additive manufacturing applications, Manufacturing Letters 15PB (2018) 104–106. doi:10.1016/j.mfglet.2018.01.006.

[4] S. A. Gold, T. G. Spears, Acoustic monitoring method for additive manufacturing processes., patent No.: US20170146488A1 (2017).

[5] M. R. Redding, S. A. Gold, T. G. Spears, Non-contact acoustic inspection method for additive manufacturing processes, patent No.: US20170146489A1 (2017).

[6] J. Beuth, J. Fox, J. Gockel, C. Montgomery, R. Yang, H. Qiao, E. Soylemez, P. Peeseewatt, A. Anvari, S. Narra, N. Klingbeil, Process Mapping for Qualification Across Multiple Direct Metal Additive Manufacturing Processes, in: Solid Freeform Fabrication Proceedings, 2013, pp. 655–665. arXiv:arXiv:1011.1669v3, doi:10.1017/CBO9781107415324.004.

[7] A. Vasinonta, J. Beuth, M. Griffith, Process Maps for Laser Deposition of Thin-Walled Structures, Proceedings of the 10th Solid Freeform Fabrication Symposium (1999) 383–392.

[8] Z. R. Francis, The effects of laser and electron beam spot size in additive manufacturing processes, PhD Thesis.

[9] A. Guzhov, F. Raue, J. Hees, A. Dengel, Esresnet: Environmental sound classification based on visual domain models, arXiv preprint arXiv:2004.07301.

[10] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[11] D. Feng, K. Xu, H. Mi, F. Liao, Y. Zhou, Sample dropout for audio scene classification using multi-scale dense connected convolutional neural network, in: Pacific Rim Knowledge Acquisition Workshop, Springer, 2018, pp. 114–123.

[12] W. Bian, J. Wang, B. Zhuang, J. Yang, S. Wang, J. Xiao, Audio-based music classification with densenet and data augmentation, in: Pacific Rim International Conference on Artificial Intelligence, Springer, 2019, pp. 56–65.

[13] G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4700–4708.

[14] K. Palanisamy, D. Singhania, A. Yao, Rethinking cnn models for audio classification, arXiv preprint arXiv:2007.11154.

[15] H. X. Tan, N. N. Aung, J. Tian, M. C. H. Chua, Y. O. Yang, Time series classification using a modified lstm approach from accelerometer-based data: A comparative study for gait cycle detection, Gait & posture 74 (2019) 128–134.

[16] Y. Tang, J. Xu, K. Matsumoto, C. Ono, Sequence-to-sequence model with attention for time series classification, in: 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), IEEE, 2016, pp. 503–510.

[17] Y. Chen, K. Zhong, J. Zhang, Q. Sun, X. Zhao, Lstm networks for mobile human activity recognition, in: 2016 International Conference on Artificial Intelligence: Technologies and Applications, Atlantis Press, 2016.

[18] Y. Guan, T. Plötz, Ensembles of deep lstm learners for activity recognition using wearables, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1 (2) (2017) 1–28.

[19] Y. Zhao, R. Yang, G. Chevalier, X. Xu, Z. Zhang, Deep residual bidir-lstm for human activity recognition using wearable sensors, Mathematical Problems in Engineering 2018.

[20] Z. C. Lipton, D. C. Kale, C. Elkan, R. Wetzel, Learning to diagnose with lstm recurrent neural networks, arXiv preprint arXiv:1511.03677.

[21] N. Mehdiyev, J. Lahann, A. Emrich, D. Enke, P. Fettke, P. Loos, Time series classification using deep learning for process planning: A case from the process industry, Procedia Computer Science 114 (2017) 242–249.