ISBN: 978-981-18-5852-9
WCSE 2022 Spring Event: 2022 9th International Conference on Industrial Engineering and Applications
doi: 10.18178/wcse.2022.04.117

Porosity Detection in Powder Bed Fusion Additive Manufacturing with Convolutional Neural Networks

Pascal Becker, Christian Roth, Arne R¨onnau and R¨udiger Dillmann

Intelligent Systems and Production Engineering, FZI Research Center for Information Technology, Karlsruhe, Germany

{fpbecker, roth, roennau, dillmanng}@fzi.de

Abstract

The sector of Additive Manufacturing is growing continuously in recent years, creating a wide range of applications such as medical devices and spacecraft parts. As the industry has high demands on the quality of these printed parts, a proper process monitoring is needed to ensure reliable parts while reducing costs. This approach focuses on the Powder Bed Fusion technology and adds an additional laser microphone to monitor the print process in-situ. Multiple defects can occur while printing, one of these is porosity. Since porosity has a strong influence on part stability, there must be no deviations here. To detect the level of porosity early on, a 2D Convolutional Neural Network was trained on in-situ audio recordings. Within an easy to use tweakable pipeline mel spectrograms were generated and fed into the neural network for classification of the porosity level. A F1-Score of 98,5% proves the concept that porosity defects of printed parts can indeed be effectively detected within production by neural networks fed with audio spectrograms. Porosity can thus be directly derived during the printing process itself, saving costs and material as a porous print can be stopped early and a x-ray after the print is done is not necessary anymore. This approach proves that integrated sensors in the printing process can deliver a huge benefit to the additive manufacturing in production.

Index Terms

Additive Manufacturing, Error Detection, Powder Bed Fusion, Acoustic Anomaly Detection, Porosity, Process Monitoring, Convolutional Neural Network, Machine Learning, Mel Spectrogram

I. Introduction

Additive manufacturing (AM), colloquially known as 3D-Printing, is increasingly growing in industrial production. A few years ago, AM was almost only used in research and development departments for the rapid production of prototypes, but now the processes have improved to such an advanced level that they can also be used for the production of final parts. Next to technologies like Fused Filament Fabrication (FFF) or Stereolithography (SLA), the development of Laser Powder Bed Fusion (L-PBF) has increased in the last years, leading to useful applications in the area of spacecraft, pharmaceutics, as well as medical devices and implants [1]–[4]. This work focuses on the L-PBF technology, which is one of the widest used in industry so far [4]. L-PBF offers inexpensive builds, has a wide range of material options like metals, plastics or even ceramics, and the support structures are already integrated through the powder bed.

Besides the downside that PBF printers require higher power than other AM methods, the PBF printers on the market still lack reliable process monitoring, resulting in many parts not being usable [6]. Defects occurring in PBF include balling, porosity, cracking, geometric defects and spatter [7]. These have a significant effect on the mechanical characteristics and thus the quality of the to be used part [8]. Hence, it is necessary to detect these defects early in the printing process, reducing the effort needed for examining the parts after printing as well as even avoiding their further processing or usage [9]. Especially porosity as a defect has severe effects on the component through crack characteristics and fatigue performances [5], [7]. An example of porosity is shown in 2. Because of its critical impact, this work concentrates only on the detection of porosity.

The further paper is divided into the following sections. Subsection I-A presents what has already been researched in the area of acoustically detecting defects in powder bed fusion by means of neural networks and spectrograms. This papers' approach is outlined in section II. The results achieved by using this approach are presented in section III. At the end, section IV concludes with the paper's actual contribution but also mentions points for future work and improvements like data augmentation.

A. Related Work

Early error detection in AM is essential to maximize the quality of the final component. The earlier an error can be detected, the earlier the operator can react and modify parameters or restart the print in total. Two general approaches exist here, visual and acoustic detection. Research has mostly been focused on visual error detection so far. For example, Okaro *et al*. [10] investigated the automatic detection of faulty parts by using photodiode sensors to measure melt pool properties. This data was used as input for a semi supervised Gaussian mixture model, attaining a correct prediction in 77%. To detect delamination and metal splatter defects, Baumgartl *et al*. [11] trained Inception CNNs with thermographic images of the metal printing process,

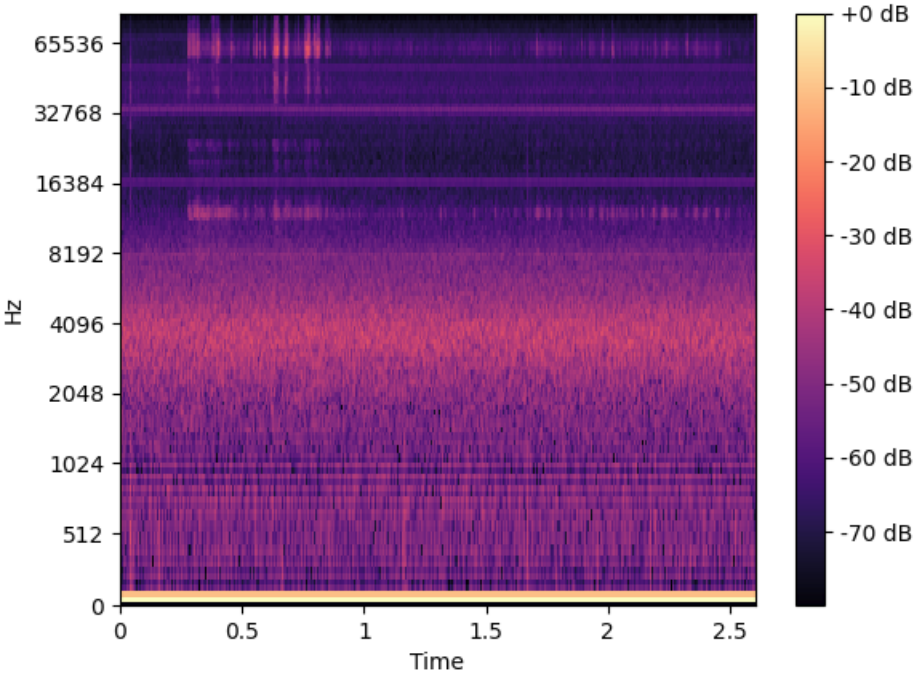

Fig. 1: High porosity in printed specimens [5].

gaining a validation accuracy of 96,8%. In a recent study, Westphal *et al*. [12] recorded images of the powder bed during the training process. As their dataset was small, they apply Transfer Learning. Feeding their images to pretrained state-of-the-art neural networks such as VGG16 and Xception [13], [14], the model performs well with a F1-Score of 95,9%. Figure 3 shows sample images of the powder bed which were fed to the network as well as the respective activation maps, i.e. where the model focused its attention on and found irregularities.

Seeing that visual defect detection is already quite advanced, one can borrow the idea of transforming acoustically measured data to spectrograms, which are in the end images, and feed them to visually learning machine learning models.

Using spectrograms of acoustic signals as input for neural networks is a widely used concept [15], although it may not seem intuitive at first. Gong *et al*. [16] propose a new architecture specifically made for this problem, the Audio Spectrogram Transformer (AST). It is based on the vision transformer architecture [17] and outperforms state-of-the-art networks on various audio benchmarks, such as ESC-50 or AudioSet. But also less sophisticated architectures, namely simple Convolutional Neural Networks (CNN), prove to give good results as Kahl *et al*. [18] have shown. Mel spectrograms of sounds from various bird species were given as input to CNNs and achieved a mAP of 0.791 for their BirdNet network.

As for the area of Additive Manufacturing, specifically powder bed fusion (PBF), Luo *et al*. [19] trained various neural networks with audio spectrograms of the printing process to detect spatter defects. Their best model proved to be a 1D-CNN, achieving a validation accuracy of 85%.

As spatter is only one type of defect in PBF, this paper concentrates on another defect type, namely porosity. Porosity immensely affects the mechanical characteristics of the component and can lead to inferior quality [9]. One of the causes of porosity is that the powder does not melt completely because of low energy input [8]. Shevchik *et al*. [20] developed a new sort of neural network, so-called Spectral Convolutional Neural Networks (SCNN). Using SCNN with wavelet spectrograms as input, they managed to differentiate the quality of PBF objects in low, medium and high quality. A higher quality implies less porosity here. Building up on this, Eschner *et al*. [9] extend this approach through the use of structure bound acoustic emission (SAE) sensors, as this promises better results. According to them, air bound AE (AAE) sensors have mostly been used so far, while SAE sensors have been considered rarely for acoustic error detection in PBF. With FFT and spectrograms from their own dataset as input for a Multi Layer Perceptron (MLP), a F1-Score of 83% (spectrograms) and 95% (FFT) were attained.

Using the dataset of Eschner *et al*. [21], this paper aims to improve the prediction metrics and thus the error detection while still retaining a simple model. Additionally, it tries to show that mel spectrograms of printed PBF parts are a meaningful input for convolutional neural networks to reliably detect porosity defects of printed parts. An efficient pipeline is proposed to reduce the manual effort needed to preprocess and train models with mel spectrograms in additive manufacturing. In the preprocessing

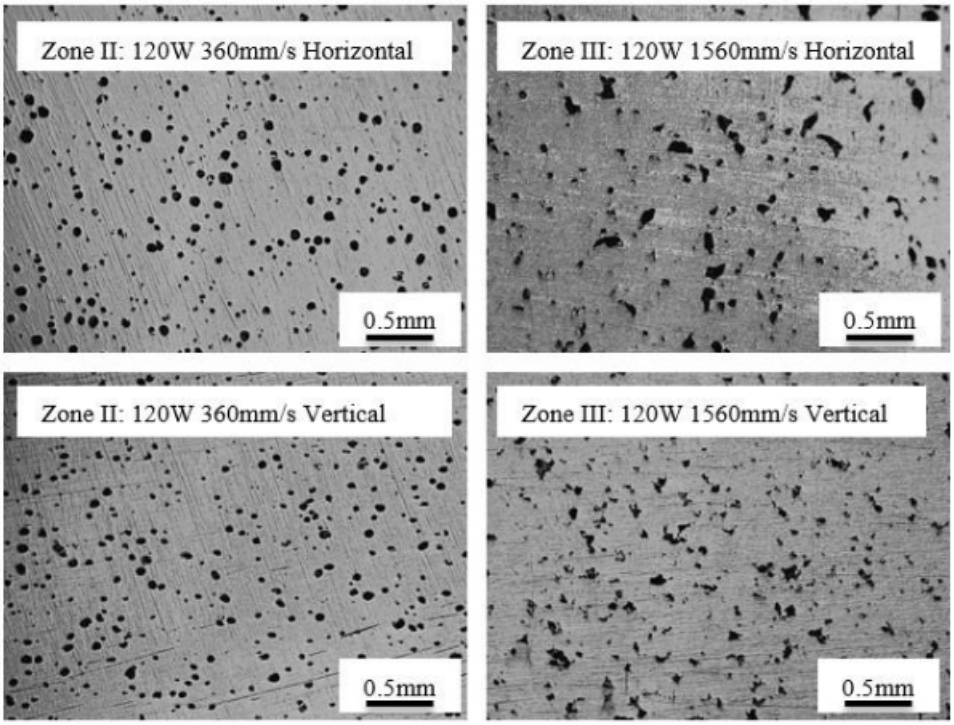

Fig. 2: High porosity in printed specimens. Different settings lead to different porosity characteristics [5].

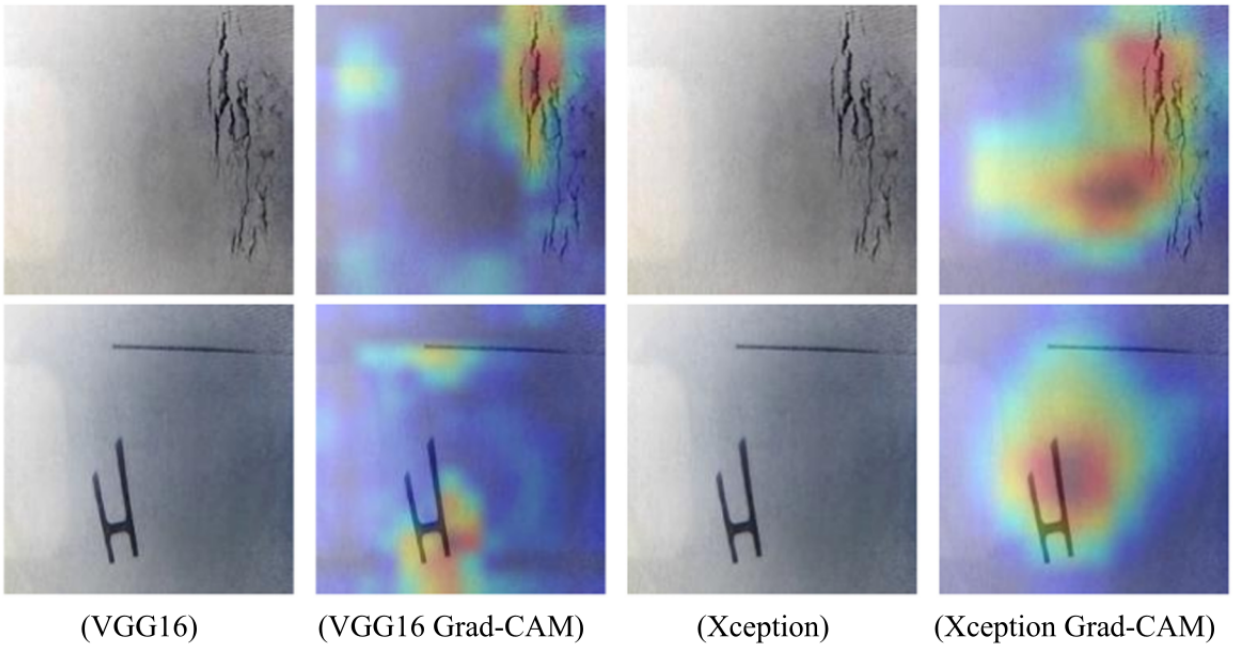

Fig. 3: Images of the powder bed for training and activation maps. Image taken from [12]

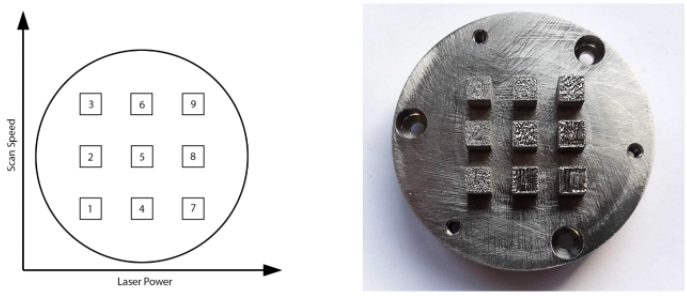

Fig. 4: Multiple cubes with different print settings have been printed to generate the data needed for the CNN, image taken from [22]

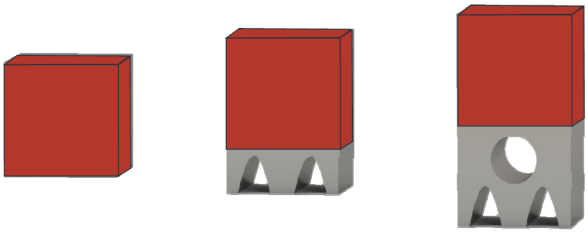

Fig. 5: Complexity classes of the printed objects. From left to right the amount of porous structure increases, starting with class 1 without any porosity to end in class 3 with a high probability of porosity. Image taken from [9]

step, downsampling reduces the dataset size drastically and simultaneously gives high performance with the added benefit of reducing training time.

II. Approach

The original dataset [21] consists of printed object layers sampled with a 4 MHz laser microphone manufactured by OR-lasers. Each object respectively consists of 125 layers. Each layer is equivalent to one sample. This makes up for a total of 6750 samples equaling 90 GB of raw data in total. The parameters varied for collecting the data are laser power (W, $\in \{80, 130, 180\}$), scan speed (mm/s, $\in \{200, 400, 1000\}$) and track distance (μm, $\in \{40, 50\}$). The configurations are illustrated in figure 4.

Each object is mapped on a complexity class, which equals the class that is used for machine learning later. The classes are balanced evenly, each class having 2250 samples. Figure 5 illustrates the complexity classes.

The higher the class, the higher the probability that the object has a higher porosity. The aim of a possible classifier would be to correctly make a connection between the acoustic signal and the porosity of an object. This raises multiple issues. Firstly, it is important to choose meaningful features. In its simplest form, this can be the raw acoustic signal, but more processed and sophisticated features such as spectrograms are also possible. In addition to this, the features also have to be preprocessed accordingly. As the sensor system introduces an offset, this factor has to be eliminated. Eventually, they have to be fit to the neural network. A 2D-CNN for example cannot process one-dimensional raw data. Even with good features, the training process can take up multiple days such as in the case of a high sampling rate and spectrograms. The goal here is to find samples that take up less space and provide faster training while still having a high performance.

The machine learning approach is differentiated into two phases, the preprocessing and training phase. The preprocessing phase has the purpose to preprocess the individual layers from float timeseries to images as well as to speed up the whole process. The speedup occurs because preprocessing is done only once and then saved on disk, compared to performing the preprocessing before every training. In the training phase, the images are loaded from the preprocessing folder, resized and fed batchwise to the neural network. The preprocessing was done to improve the iteration speed while creating and training the network. Both phases are illustrated as a common pipeline in figure 6.

Beginning with the *preprocessing phase*, the layers are loaded sequentially as .h5py files and flattened into a single array. From this array, an offset is subtracted. The offset is calculated as the global minimum over all layers. This has to be done as the sensor adds an offset automatically to all data points. To reduce the number of samples, downsampling was performed. While this may seem counter-intuitive at first, it proved to be effective later as even with a low sample rate the prediction metrics were high including the added bonus of having a much shorter training and preprocessing duration. Experiments are conducted with the following sample rates (in Hz): $F_S \in \{176400, 352800, 705600, 1411200, 2822400\}$. Experiments were first made with 44,1kHz to set a bare minimum for the downsampling rate. The model did not train well with this rate. Thus,

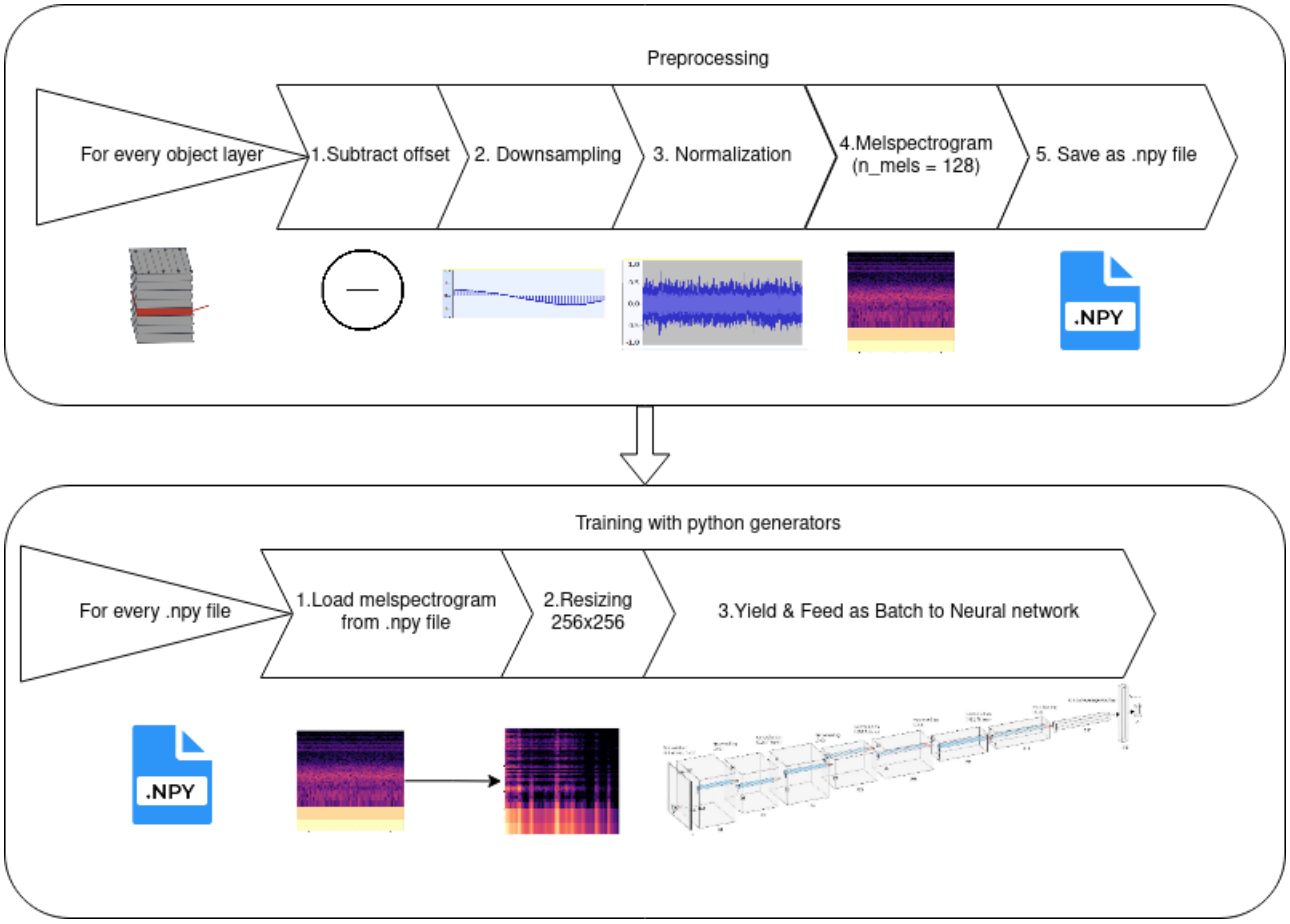

Fig. 6: Processing pipeline: From the acoustically recorded data of each layer, an offset is subtracted, downsampling and normalization applied. After transforming the data to mel spectrograms, they are saved as .npys. For the training phase, the .npys are loaded again, resized and fed to the CNN.

this rate was doubled until the results were appropriate. Subsequently, normalization is performed to bring the values in a common range and center them around zero. Afterwards, either the spectogram or mel spectrogram is created with the use of the *LibROSA* library functions [23]. A spectrogram is the graphical representation of a frequency spectrum regarding time. This aims to find out, which image type is better for error detection. Both spectrograms and mel pectrograms are used often in acoustical error detection, as a study of Nunes *et al.* [15] demonstrates, resulting in them being tested as an input to the model. Ultimately, the mel spectrograms proved to be decisively better as the result metrics were not only higher but also training time shorter.

An example from the dataset is shown in 7.

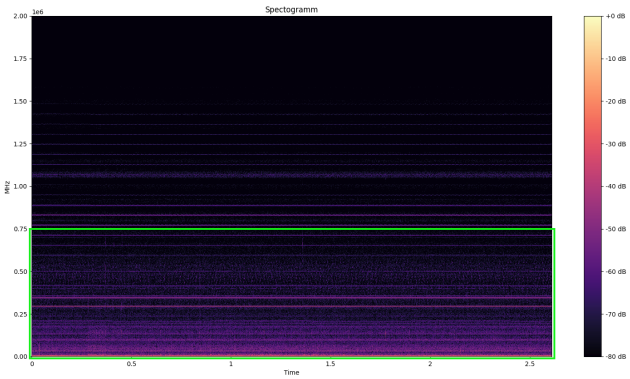

Fig. 7: A spectrogram sampled with 4 MHz and $n_fft = 2048$. Spectral energies in the lower frequency range ($<$ 500kHz, highlighted in green) are clearly visible.

The mel scale better resembles human hearing and is thus applied on many challenges, where anomalies or errors are actually hearable. These challenges include sound event detection [24], [25], anomaly detection [26] and music information retrieval (MIR) [27]. An example of a mel spectrogram from the dataset is envisioned in 8. The x-axis describes the time in seconds, the y-axis the logarithmically scaled frequency in Hz. The mel energy is quite constant in the lower spectrum, in the higher range spectral patterns can be seen around 0,5 s.

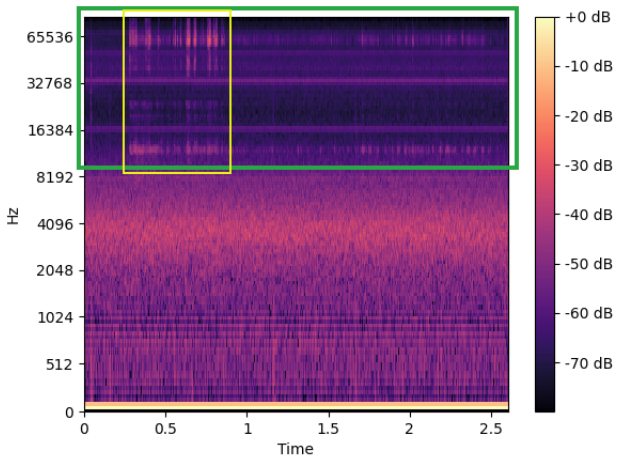

Fig. 8: Mel spectrogram sampled with 176400 Hz, axis is scaled logarithmically. Spectral patterns highlighted in yellow in upper frequency range ($>$ 10 kHz, green highlighted area). The rest is mostly constant.

Both image types use a FFT window length of $n_{fft} = 2048$. For the mel spectrogram, a minimum frequency bin of $f_{min} = 8000$ Hz and $n_{mels} = 128$ as the number of mel frequency bins is used. Those (mel-)spectrograms are saved finally as .npy files for further computing.

In the upcoming *training phase*, first a batch of layers, represented by the (mel-)spectrograms, is fetched by loading the respective .npy files. By using the .npy files, the approach is speeded up, as the files only have to be preprocessed once. Additionally, .npy files are loaded faster than .h5py files. While iterating, this speed-up sums up and saved multiple hours of computation time. The spectrograms are then resized to either 256x256 or 512x512, serving as uniform input size, before being forwarded to the network. With a bigger input size, more information can be carried, but also the network has to be good enough to extract the information inside. Because of this, it will be tested in the experiments whether 512x512 yields higher result metrics than 256x256. The batchwise loading is done by python generators using the keyword `yield`, as not too many images fit simultaneously in the RAM. The generators iteratively only fetch the currently needed images for the batch, feed them to the network and discard them afterwards from RAM.

For constructing and training the neural networks, Tensorflow is used [28]. A NVIDIA GeForce RTX 2080 Ti is used for training the models. The convolutional neural network (CNN) topology is shown in Figure 9.

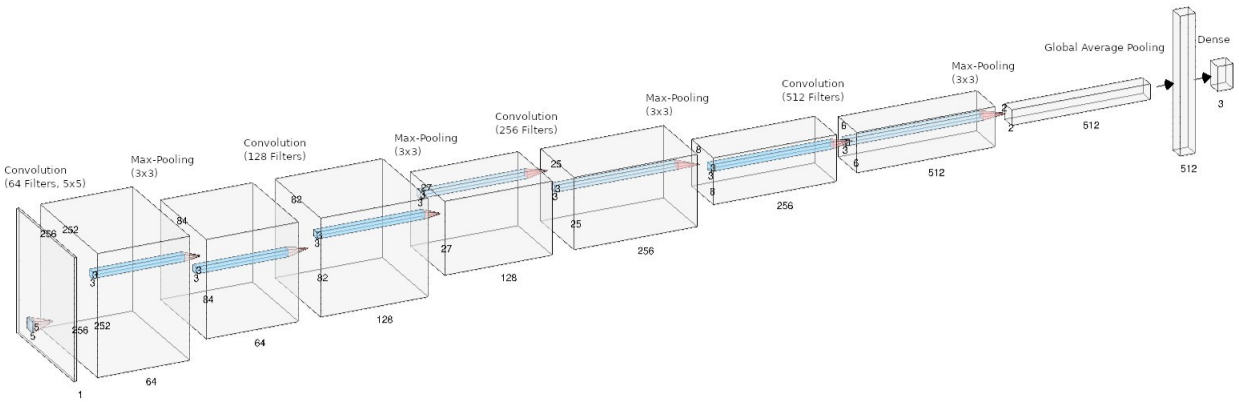

Fig. 9: Convolutional neural network architecture. It has in total 1,8 M trainable parameters and consists of multiple pooling and convulutional layers tom improve the classification result.

The individual network layers have the structure Conv2D $\in \{64, 128, 256, 512\} \rightarrow$ MaxPooling $\rightarrow$ Batch-Normalization $\rightarrow$

Dropout $\in \{0.2, 0.5\}$
The Conv2D layers uses "relu" (Rectified Linear Unit) activation functions. In the final classification layers, a softmax function is used, Adam as optimizer and categorical crossentropy as loss function. The model possesses ~1,8 M parameters. Every training is iterated over 5 epochs with a batch size of 8. To save precious time and prevent overfitting, early stopping with a patience of two is applied, which means that the training will stop as soon as the models' validation accuracy has not increased for two epochs. To lower the bias and to get the best model, k-fold cross validation with $k = 5$ is applied and k models are constructed. The value k = 5 was chosen because it is a good medium between a single run and the recommended value k = 10 [29], which would take long to conduct. Table I illustrates how the cross validation works. With this approach, every sample in the dataset will be validated against.

TABLE I: 5-fold cross validation. 5 folds are created every iteration, from which one fold acts as the validation set, the other four belong to the training set.

Iteration	Training set				Validation set
1	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
2	Fold 5	Fold 1	Fold 2	Fold 3	Fold 4
3	Fold 4	Fold 5	Fold 1	Fold 2	Fold 3
4	Fold 3	Fold 4	Fold 5	Fold 1	Fold 2
5	Fold 2	Fold 3	Fold 4	Fold 5	Fold 1

From the computed models, the best one regarding the validation accuracy is chosen for the prediction. The prediction metric used to compare the experiments is the F1-Score. It is defined as

$$2 * \frac{\text{precision * recall}}{precision + recall}$$

and one of the most widespread metrics in the Machine Learning area for multiclass classification [30].

Concerning the labels, Eschner *et al.* [9] created the respective labels according to their estimation of the porosity in each object. This means, each layer of the same object has the same class label. Figure 10 shows this mapping. The class borders are at 1% and 6%. The estimated porosity is calculated after equation II, where ρ is the density, m the mass and V the volume of the object.

$$\rho = \frac{m}{V} \tag{1}$$

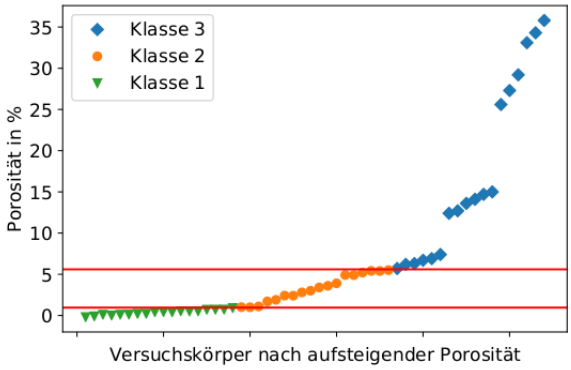

Fig. 10: Mapping of porosity to class label. The estimated porosity (in %) is on the y-axis, the objects ordered after increasing porosity on the x-axis. Taken from [9].

A ML pipeline was proposed where preprocessing and training properties can be easily controlled. With a CNN, the preprocessed PBF layers get classified accordingly.

III. Experiments and Results

The offset O subtracted from every sample is $O = 8220261$. Experiments were made with spectrograms and mel spectrograms as image types. Input sizes were 256x256 and 512x512 to bring the images to uniform size. The latter one is tested in case the downsampling to 256x256 includes losing too much information. Provided the model is complex enough, it could also attain better prediction metrics than the former one. Since a 4 MHz sensor is used, it can be expected that the results in the higher

sampling rates are better as they could contain more information. It is also anticipated that the spectrograms perform better than mel spectrograms as the sounds heard during the printing process are far out of the acoustic spectrum of humans. The dataset was split into 10% test/prediction data, the remaining 90% was used for cross validation. For each different sampling rate F_S, 5-fold cross validation was used with each CNN model being trained for 5 epochs. $k = 5$ implies a 80/20 training/validation split. The mean of the cross validation was saved and used as a metric for stability of the training process and the models. Out of every fold, the model with the best validation accuracy was picked to perform a prediction on the test dataset afterward. From this prediction, the respective confusion matrices were constructed. Table V shows the confusion matrix for $F_S = 176400$ Hz and a mel spectrogram input of 256x256 points. The model performs outstanding except for porosity class 2. This also happened to be the case for the experiments with other sample rates and input sizes. Taking a look at the probability density functions of the respective porosity classes in figure 11, it can be seen that class 2 overlaps with the other classes regarding its porosity percentage. This could make it harder for the model to decisively predict the right porosity class according to its given ground truth label.

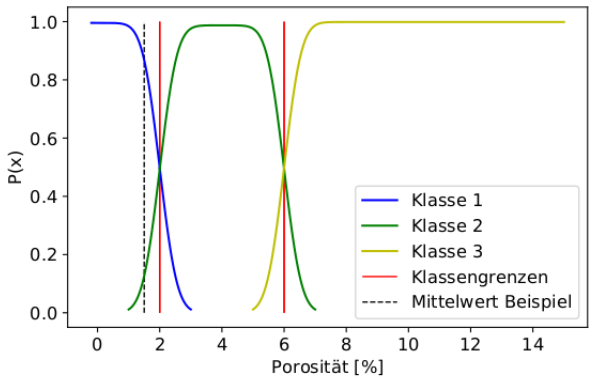

Fig. 11: Probability density functions regarding estimated porosity percentages. The red highlighted lines represent the strict class borders. Taken from [9]

Table II show the results of the 5-fold cross validation after a training of 256x256 mel spectrograms. The models with $F_S = 352800, 705600$ have a high validation accuracy and a good mean, but the prediction fails to impress. The best performing and most stable models are the remaining ones with $F_S = 2822400, 1411200, 176400$ as all metrics are high. $F_S = 176400$ is chosen as the final model, as its prediction and mean are not only high, but mainly training and prediction were faster because of the lower sample rate. Being able to rapidly train and deploy a model is an important advantage. Another interesting observation is that the best validation accuracies do not differ much. This implies that through enough tries/folds, thanks to cross validation, always a top performing model could be found.

TABLE II: Experiment 1: Resulting metrics of a 5-fold cross validation for different sample rates and input mel spectrogram 256x256. F1-Score is for the prediction after training and validation.

F_S, Hz	Best ValAcc, %	Mean, %	F1-Score, %
176400	98,3	86,9	98,5
352800	98,5	85,4	52,5
705600	98,8	89,6	82,1
1411200	98,6	88,1	99
2822400	98,6	85,4	98

Having a look at table III for input size 512x512 (referred to as Experiment 2) and comparing it to table II with input size 256x256 (Experiment 1), some observations can be made. The means of the cross validation are differing more, as well as the validation accuracies. No validation accuracy is as high as the ones from Experiment 1. Maybe the model is already overwhelmed by the larger sizes and needs more filters. The means also differ more, around 10%. From the F1-Scores, no observation or conclusion can be made except that the model with a sampling rate of 176400Hz is unfit. As mel spectrograms with this rate had $1152x128 = 147456$ samples at most, but $512x512 = 262144$ samples, meaning much of the input spectrogram is empty.

Putting further emphasis on model stability, in every fold one or two model outliers are found. These have a validation accuracy differing more than 20% from the best model. As the cross validation has only five folds, this implies not very stable models. Table IV illustrates this.

To overcome this issue, data augmentation can be applied to make the models more robust. Pitch-Shifting, time stretching and amplifying are common augmentation methods in acoustics [24].

TABLE III: Experiment 2: Resulting metrics of a 5-fold cross validation for different sample rates and input mel spectrogram 512x512. F1-Score is for the prediction after training and validation.

F_S, Hz	Best ValAcc, %	Mean, %	F1-Score, %
176400	99,7	74,7	41,4
352800	97	78,9	97,4
705600	93,4	74,3	91,8
1411200	97,1	89,7	82,6
2822400	95,9	89,3	95,3

TABLE IV: 5-fold cross validation for different sample rates and input mel spectrogram 256x256

F_S, Hz / Fold #	1	2	3	4	5	Mean, %
176400	93	98,1	95	50,2	98,3	86,9
352800	98,5	96,7	92,3	94,8	44.9	85,5
705600	86,1	98,1	89,2	98,8	75,7	89,6
1411200	83,6	92,1	69,5	97	98,6	88,1
2822400	98,6	93,9	71,6	66,9	96,4	85,5

To deploy the approach in an industrial environment, the execution time within the application has to be fast. This includes not only the prediction time, but also training time to make the model deployable fast. Table VI shows exemplary execution times with the according preprocessing parameters. The training time is for the whole cross validation process, the prediction for one single sample. One sample is 0,9s long. The training time increases if a higher sample rate is used, taking even as long as 11 days in the case of 512x512 spectrograms, downsampled with 2822400 Hz. This is the only case where real-time prediction is not possible, making deployment unfeasible. The spectrograms with a sample rate of 4 Mhz had sizes of at most 6200x1025, so the one with 2822400Hz has a bit lower, but similar size, leading to this long duration. The other models are suitable for this task.

The training with the spectrograms proved to be underperforming compared to mel spectrograms, being lower around 20% in most metrics. This could be due to the preprocesing done. With $n_mels = 128$, the number of frequency bins mel spectrograms was fixed and rather small. This fits better to the following resizing size-wise. Spectrograms sizes were bigger in general. For example, mel spectrograms with $F_S = 176400$ had a variable size of $1152x128$ at most. The variable size stems from the variable scan speed with some layers taking longer or shorter, resulting in more or less samples. On the contrast, a spectrogram with the same sample rate had $1025x1070$ at most. 1025 stems from the Short Time Fourier Transformation (STFT), which creates a spectrogram with $1+n_fft/2$ rows, i.e. frequency bins. If this size is transformed to 256x256 or 512x512, information naturally gets lost. In further experiments, this should be considered and n_fft set lower or a model used with a larger input and possibly more filters to guarantee that the model extracts the necessary information. Seeing that the highest filter size is 512 from figure 9, additional layers with sizes 1024, 2048, etc. can be added to capture the complex structure of spectrograms better. To confirm this issue, activation maps like in figure 3 can be created. With these images, it can be seen what the models actually learn, i.e. where they place their focus on [31].

TABLE V: Confusion matrix for $F_S = 176400Hz$

	Class 1	Class 2	Class3	F1-score, %
Class 1	244	0	0	98
Class 2	10	192	0	97,5
Class 3	0	0	210	100
F1 macro	98,5%			

TABLE VI: Execution times for chosen experiments. Training time is for a 5-fold cross validation, i.e. 5 trainings + validation. Prediction is for one sample à 0.9s.

Image type	Sample rate	Input size	Training	Prediction
Mel spectrogram	176400	256	9 m	15 ms
Mel spectrogram	176400	512	23 m	15 ms
Mel spectrogram	352800	256	15 m	45 ms
Mel spectrogram	2822400	512	7,5 h	48 ms
Mel spectrogram	2822400	256	9,5 h	111 ms
Spectrogram	2822400	512	11 d	7,5 s
Spectrogram	176400	256	3 h	88 ms

IV. Conclusion

In this work, an approach for detecting porosity defects in L-PBF additive manufacturing technology was introduced. By training CNNs with acoustically recorded data transformed into mel spectrograms, the printed layers were separated by the model in different porosity classes. Sampling rate, input size and image type were varied for the processing. The model performing best on the test data was picked for the final result. By applying k-fold cross validation on the experiments, the robustness of this approach is assured. While the robustness of the models is generally high, there are still some occasional outliers found through cross validation. It was demonstrated that the acoustical approach applied here can reliably detect different porosity classes. A F1-Score of 98,5% and a mean of 86,9% prove this. The prediction of new data can be done in little time, making real-time usage feasible. The proposed pipeline has easily adjustable parts and facilitates comfortable use. The preprocessing has to be only done once, making training faster. The lightweight model can be trained fast and deployed even on embedded hardware. Regarding future work, first and foremost data augmentation should be carried out to further improve the robustness and quality of the model. Pitch-Shifting and amplifying are common methods to achieve this. As a 4 MHz sensor was used to get the dataset [21] used in this work, but not fully utilized as the data was downsampled to guarantee a faster training, the CNN could be extended by adding even more filters such as 1024 or 2048 to it. The experiments conducted have shown that the resizing to 256x256 / 512x512 of big spectrograms have detrimental effects on the training and prediction of the classifier. Because of this, a larger network is needed which can have more and complex filters to mitigate this effect. With more recent networks such as the AST [16] and Transfer Learning, even larger improvements could be made. At last, further defects such as spatter, cracking or balling can be investigated and also be detected by the same network.

References

[1] Fabrizio Fina, Simon Gaisford, and Abdul W Basit. Powder bed fusion: The working process, current applications and opportunities. In *3D Printing of Pharmaceuticals*, pages 81–105. Springer, 2018.

[2] Anton du Plessis, Ina Yadroitsava, and Igor Yadroitsev. Ti6al4v lightweight lattice structures manufactured by laser powder bed fusion for load-bearing applications. *Optics & Laser Technology*, 108:521–528, 2018.

[3] Atheer Awad, Fabrizio Fina, Alvaro Goyanes, Simon Gaisford, and Abdul W Basit. Advances in powder bed fusion 3d printing in drug delivery and healthcare. *Advanced Drug Delivery Reviews*, 2021.

[4] Syed AM Tofail, Elias P Koumoulos, Amit Bandyopadhyay, Susmita Bose, Lisa O'Donoghue, and Costas Charitidis. Additive manufacturing: scientific and technological challenges, market uptake and opportunities. *Materials today*, 21(1):22–37, 2018.

[5] Haijun Gong. Generation and detection of defects in metallic parts fabricated by selective laser melting and electron beam melting and their effects on mechanical properties. 2013.

[6] Wentai Zhang, Brandon Abranovic, Jacob Hanson-Regalado, Can Koz, Bhavya Duvvuri, Kenji Shimada, Jack Beuth, and Levent Burak Kara. Flaw detection in metal additive manufacturing using deep learned acoustic features. 2020.

[7] Marco Grasso and Bianca Maria Colosimo. Process defects and in situ monitoring methods in metal powder bed fusion: a review. *Measurement Science and Technology*, 28(4):044005, 2017.

[8] Wenjia Wang, Jinqiang Ning, and Steven Y Liang. Prediction of lack-of-fusion porosity in laser powder-bed fusion considering boundary conditions and sensitivity to laser power absorption. *The International Journal of Advanced Manufacturing Technology*, 112(1):61–70, 2021.

[9] N Eschner, L Weiser, B Häfner, and G Lanza. Classification of specimen density in laser powder bed fusion (l-pbf) using in-process structure-borne acoustic process emissions. *Additive Manufacturing*, 34:101324, 2020.

[10] Ikenna A Okaro, Sarini Jayasinghe, Chris Sutcliffe, Kate Black, Paolo Paoletti, and Peter L Green. Automatic fault detection for laser powder-bed fusion using semi-supervised machine learning. *Additive Manufacturing*, 27:42–53, 2019.

[11] Hermann Baumgartl, Josef Tomas, Ricardo Buettner, and Markus Merkel. A deep learning-based model for defect detection in laser-powder bed fusion using in-situ thermographic monitoring. *Progress in Additive Manufacturing*, pages 1–9, 2020.

[12] Erik Westphal and Hermann Seitz. A machine learning method for defect detection and visualization in selective laser sintering based on convolutional neural networks. *Additive Manufacturing*, 41:101965, 2021.

[13] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[14] François Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017.

[15] Eduardo C Nunes. Anomalous sound detection with machine learning: A systematic review. *arXiv preprint arXiv:2102.07820*, 2021.

[16] Yuan Gong, Yu-An Chung, and James Glass. Ast: Audio spectrogram transformer. *arXiv preprint arXiv:2104.01778*, 2021.

[17] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[18] Stefan Kahl, Connor M Wood, Maximilian Eibl, and Holger Klinck. Birdnet: A deep learning solution for avian diversity monitoring. *Ecological Informatics*, 61:101236, 2021.

[19] Shuyang Luo, Xiuquan Ma, Jie Xu, Menglei Li, and Longchao Cao. Deep learning based monitoring of spatter behavior by the acoustic signal in selective laser melting. *Sensors*, 21(21):7179, 2021.

[20] Sergey A Shevchik, Christoph Kenel, Christian Leinenbach, and Kilian Wasmer. Acoustic emission for in situ quality monitoring in additive manufacturing using spectral convolutional neural networks. *Additive Manufacturing*, 21:598–604, 2018.

[21] Niclas Eschner. In-process monitoring of laser powder bed fusion using structure-borne acoustic process emissions to predict porosity ., 2021.

[22] N Eschner, L Weiser, B Häfner, and G Lanza. Development of an acoustic process monitoring system for selective laser melting (slm). In *Proceedings of the 29th Annual International Solid Freeform Fabrication Symposium, Austin, TX, USA*, pages 13–15, 2018.

[23] Brian McFee, Colin Raffel, Dawen Liang, Daniel PW Ellis, Matt McVicar, Eric Battenberg, and Oriol Nieto. librosa: Audio and music signal analysis in python. In *Proceedings of the 14th python in science conference*, volume 8, pages 18–25. Citeseer, 2015.

[24] Pascal Becker, Christian Roth, Arne Roennau, and Ruediger Dillmann. Acoustic anomaly detection in additive manufacturing with long short-term memory neural networks. In *2020 IEEE 7th International Conference on Industrial Engineering and Applications (ICIEA)*, pages 921–926. IEEE, 2020.

[25] Emre Çakir and Tuomas Virtanen. End-to-end polyphonic sound event detection using convolutional recurrent neural networks with learned time-frequency representation input. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7. IEEE, 2018.

[26] Robert Müller, Fabian Ritz, Steffen Illium, and Claudia Linnhoff-Popien. Acoustic anomaly detection for machine sounds based on image transfer learning. *arXiv preprint arXiv:2006.03429*, 2020.

[27] Arun Solanki and Sachin Pandey. Music instrument recognition using deep convolutional neural networks. *International Journal of Information Technology*, pages 1–10, 2019.
[28] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*, 2016.
[29] Payam Refaeilzadeh, Lei Tang, and Huan Liu. Cross-validation. *Encyclopedia of database systems*, 5:532–538, 2009.
[30] Davide Chicco and Giuseppe Jurman. The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation. *BMC genomics*, 21(1):1–13, 2020.
[31] B. Zhou, A. Khosla, Lapedriza. A., A. Oliva, and A. Torralba. Learning Deep Features for Discriminative Localization. *CVPR*, 2016.