

# Automated particle recognition for engine soot nanoparticles

E. Haffner-Staton  | L. Avanzini | A. La Rocca  | S. A. Pfau | A. Cairns

Department of Mechanical, Materials and Manufacturing Engineering, University of Nottingham, University Park, Nottinghamshire, UK

## Correspondence

Dr. Ephraim Haffner-Staton, Department of Mechanical, Materials and Manufacturing Engineering, University of Nottingham, University Park, Nottingham NG7 2RD, UK.

Email:

[ephraim.haffner-staton@nottingham.ac.uk](mailto:ephraim.haffner-staton@nottingham.ac.uk)

## Funding information

Doctoral Training Partnership received by the University of Nottingham in 2016-2017. This grant funded the PhD studies of co-author Sebastian Pfau at the University of Nottingham from 2017-2021.

## Abstract

A pre-trained convolution neural network based on residual error functions (ResNet) was applied to the classification of soot and non-soot carbon nanoparticles in TEM images. Two depths of ResNet, one 18 layers deep and the other 50 layers deep, were trained using training-validation sets of increasing size (containing 100, 400 and 1400 images) and were assessed using an independent test set of 200 images. Network training was optimised in terms of mini-batch size, learning rate and training length. In all tests, ResNet18 and ResNet50 had statistically similar performances, though ResNet18 required only 25–35% of the training time of ResNet50. Training using the 100-, 400- and 1400-image training-validation sets led to classification accuracies of 84%, 88% and 95%, respectively. ResNet18 and ResNet50 were also compared for their ability to categorise soot and non-soot nanoparticles via a fivefold cross-validation experiment using the entire set of 800 images of soot and 800 images of non-soot. Cross-validation was repeated 3 times with different training durations. For all cross-validation experiments, classification accuracy exceeded 91%, with no statistical differences between any of the network trainings. The most efficient network was ResNet18 trained for 5 epochs, which reached 91.2% classification after only 84 s of training on 1600 images. Use of ResNet for classification of 1000 images, the amount suggested for reliable characterisation of soot sample, requires <4 s, compared with >30 min for a skilled operator classifying images manually. Use of convolution neural networks for classification of soot and non-soot nanoparticles in TEM images is highly promising, particularly when manually classified data sets have already been established.

## KEYWORDS

automotive, nanoparticles, neural networks, soot, TEM, vision learning

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Journal of Microscopy* published by John Wiley & Sons Ltd on behalf of Royal Microscopical Society.

## 1 | INTRODUCTION

Incomplete combustion of hydrocarbon fuel in internal combustion engines (ICEs) lead major fraction of exhaust particle matter (PM),<sup>1</sup> accounting for 40% to >90% of the total PM depending on operating conditions.<sup>2</sup> Due to severe negative effects on health and environment, PM has been limited by the European Union emission standards in terms of both particle mass and particle number since 2011.<sup>3</sup> Despite trends towards electrified powertrains and decarbonised fuels, continued improvement and optimisation of ICEs still appear the best short-term solution in terms of affordability, energy security and impact of greenhouse gas emissions.<sup>4</sup> Decarbonisation of haulage/shipping faces difficulties in terms of performance and infrastructure requirements, such that the UK's 'core' NetZero strategy suggests only 13% of surface transport fleet will be decarbonised by 2050.<sup>5</sup> As such, it is likely that engine PM emissions will remain an important target for regulations for the next few decades at least.

Soot emissions from combustion are a major concern for both human health and climate change. Atmospheric carbon particles are considered second only to CO<sub>2</sub> in terms of contribution to the greenhouse effect and have been shown to impact respiratory and cardiovascular health,<sup>6</sup> while also being linked to cancer and impaired neurovascular function.<sup>7</sup> The morphology of soot plays a critical role in determining their effects on human and environmental health. It has been observed that the fractal morphology of soot<sup>8</sup> and its hygroscopic properties<sup>9</sup> influence the depth of inhalation into the lungs, and the surface area of soot nanoparticles has been correlated to their toxicity.<sup>10</sup> The fractal morphology of soot has been seen to influence refractive properties in the atmosphere,<sup>11</sup> influencing light adsorption and contribution to the greenhouse effect. As such, characterisation of soot morphology has been an important topic of research over the last few decades, with a view to more deeply understanding these soot-related phenomena.<sup>12</sup>

Due to the nanoscopic size of soot (50–500 nm),<sup>13,14</sup> transmission electron microscopy is required for imaging. Transmission electron microscopy (TEM) has been widely used to characterise properties such as maximum length, fractal dimension, radius of gyration and primary particle diameter.<sup>13,15–18</sup> However, TEM produces 2D projections while soot nanoparticles are known to be complex and highly irregular 3D structures, leading to significant errors.<sup>17,19–22</sup> Electron tomography (ET) has been used to create 3D reconstructions of soot nanoparticles from a series of 2D TEM images acquired over a wide range of angles (e.g. ±60°).<sup>17,20–25</sup> ET can produce highly accurate 3D models of soot nanoparticles but is significantly slower due to each particle requiring multiple images, and suitable particles being more difficult to locate on a TEM grid.

Work by our research group has aimed to optimise the ET process for soot by implementing automated image acquisition and processing to identify suitable particles over large areas (~40 μm<sup>2</sup>) of the TEM grid.<sup>17,19,23</sup> This procedure also permits rapid measurement of particles in 2D, allowing greatly increased sample sizes and thus increased reliability of TEM-derived measurements.<sup>26</sup> However, due to low contrast of soot against the TEM grid substrate, and high concentrations of non-soot particles in samples, a manual review process was required to separate real soot structures from false positives.<sup>17,23</sup>

Deep learning is a statistical technique, which uses data to classify patterns using neural networks.<sup>27</sup> It can be applied to automate the manual particle detection procedure, classifying nanoparticles in a few seconds, reducing operational time and contributing significantly to the increased throughput of ET for soot. Artificial neural networks (ANNs) are deep learning architectures made of neurons, which receive inputs from other neurons or external sources (e.g. pictures, words), weight them using kernels and compare to a threshold value in order to categorise them. Analysing images is the key to study soot; however, to process an image using a simple artificial neural network would require it to be converted into a 1D vector. This approach does not take into account important features such as pixel arrangements in corners, which helps to differentiate an image from another.<sup>28</sup> As a result, convolutional neural networks (CNNs) are used to analyse and process images by detecting key pattern parameters and are used in software, which classifies images.<sup>29</sup> The machine learning tool detects and segments patterns on the images and learns from them, making the algorithm increasingly accurate as more images are used for training.<sup>30</sup>

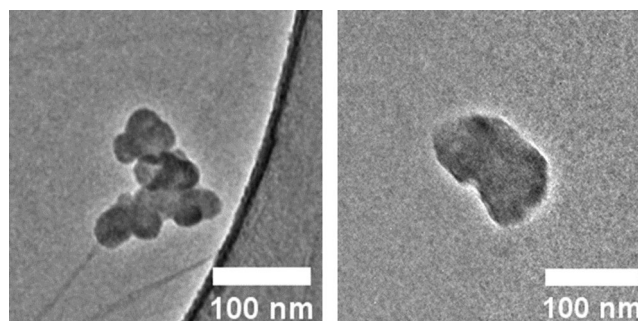
Examples of CNNs in research include segmentation of magnetic resonance images of the brain<sup>31</sup> and identification of ferrography wear particles through image recognition to determine failure modes of metals depending on the shape of wear particles.<sup>32</sup> Additionally, Raghu et al. used pre-trained CNN architectures to classify seizure types.<sup>33</sup> The technique used is called 'transfer learning' and it consists of training a pre-trained architecture,<sup>34</sup> which was originally trained to recognise images of 1000 categories, to recognise images of the topic of interest. ResNet, AlexNet, GoogleNet, VGG and SqueezeNet were the pre-trained CNN architectures investigated.<sup>33</sup> It has been observed that deeper networks can be subject to degradation problems when they converge: as size increases, accuracy gets saturated and degrades rapidly.<sup>35</sup> Hence, an alternative approach to address the problem is formulating an architecture based on residual functions. While stacked convolutional layers learn features, residual learning uses residual errors to study patterns. The pre-trained CNN architecture 'ResNet' does this by subtracting features

which are learned by the input layer, allowing the software to be more easily trainable and to be subject to much less degradation when layers are increased.<sup>36</sup> ResNet is pre-trained to learn 1000 categories of images to be able to classify them. The architecture can be adapted by replacing the final layer before the fully connected layers (called 'fc1000'), with a new layer which only contains categories of interest. Xiang et al. compared ResNet18, ResNet50 and ResNet101 to identify the ideal ResNet architecture to classify two categories of images, with ResNet18 found to be the more accurate despite being the shallower network.<sup>37</sup> Nibali et al. used pre-trained ResNet18 to classify images of Pulmonary nodule,<sup>38</sup> in a study with similar scope to this work.

In this study two depths of ResNet, one 18 layers deep (ResNet18) and one 50 layers deep (ResNet50), were applied to the problem of classifying soot and non-soot carbon nanoparticles TEM images from a sample of soot-in-oil taken from a GDI passenger vehicle. The two networks were first assessed using training-validation sets of increasing size: containing 100, 400 and 1400 images. Though generally an increase in network accuracy with increasing training-validation set size is expected,<sup>39</sup> the rate of this increase is of interest. Due to strong differences in the composition of different soot samples (e.g. gasoline soot-in-oil vs. diesel exhaust soot), it is possible that separate network trainings may be required for different types of soot on a per-sample basis. As such, small training-validation sets may be the only available training resource. Alternatively, similar soot samples (such as those from repeated experiments of the same soot-in-oil) may benefit from training networks on one another's data, so larger training-validation sets are also of interest for a 'best-case' scenario. ResNet18 and ResNet50 were also compared for soot/non-soot classification through a fivefold cross-validation experiment using 1600 images (with a 1:1 ratio of soot and non-soot images). Automated classification of soot and non-soot carbon nanoparticles would not only greatly increase throughput and reliability of soot morphology characterisations, where upwards of 1000 nanoparticles need to be measured,<sup>26</sup> but also permits future studies on the non-soot fraction of nanoparticles in soot-in-oil samples, which at present are not well understood. Modifying the architecture of the pre-trained software and training it to recognise specific classes of interest has the potential to be applied widely in the field of nanoparticles research using TEM.

## 1.1 | Methodology

Nanoparticles were extracted from a sample of soot-laden oil taken from a passenger GDI vehicle, via a solvent

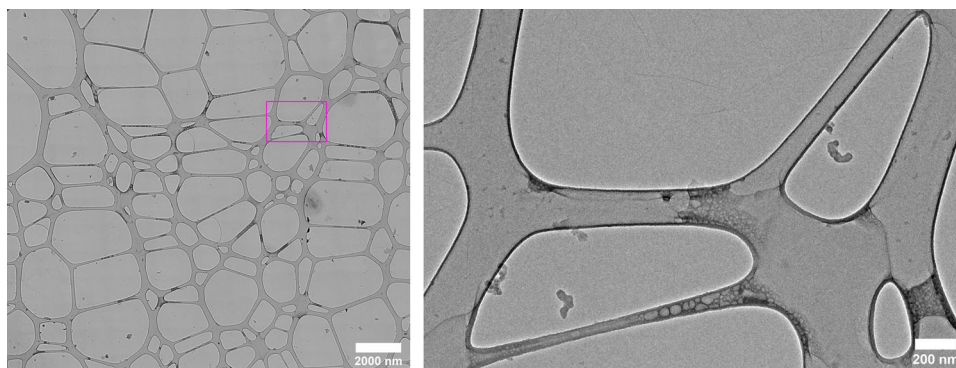


**FIGURE 1** (Left) Typical soot-in-oil nanoparticle imaged at  $\times 10,000$  magnification. Note the clear presence of spherical 'primary particle' sub-units around 30 nm in diameter, and varied grey-levels within the structure due to overlapping of primary particles. (Right) Typical amorphous carbon 'non-soot' nanoparticle from the same soot-in-oil sample. Note lack of primary particle sub-units, less complex and more rounded structure, and more consistent contrast/grey-levels due to more uniform structure

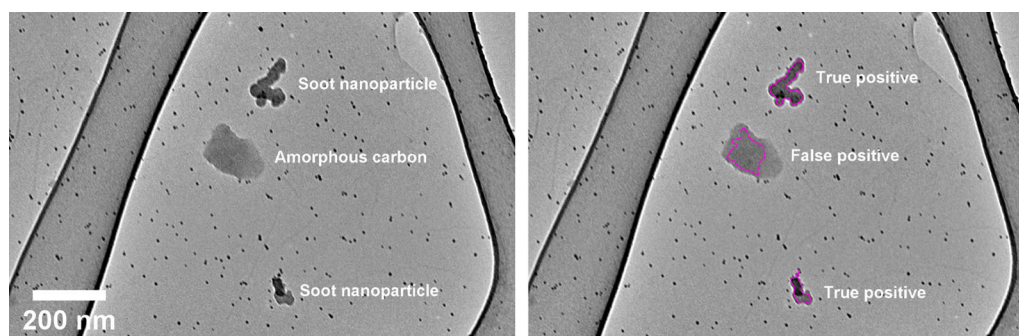
dilution process using heptane and diethyl ether (further details can be found in a previous study).<sup>40</sup> Nanoparticles were prepared for TEM by deposition on 300-mesh copper TEM grids coated with graphene oxide on lacey carbon. Previous study of this sample revealed high concentrations of carbonaceous nanoparticles of similar size and density/contrast soot, but which lack the characteristic fractal-like morphology, primary particle sub-units and graphitic nanostructure of soot. An example of a soot nanoparticle and non-soot amorphous carbon nanoparticle is shown in Figure 1. These non-soot carbon nanoparticles have been referred to as amorphous carbon or sludge-like nanoparticles in previous studies of soot-in-oil,<sup>41</sup> but as yet are not well understood. The input data (i.e. TEM images) for the present work is the same used in the development of the semi-automated procedure,<sup>40</sup> from which a total of 800 images of soot and 800 images of non-soot nanoparticles were collected.

The procedure for rapid soot identification from our previous work was based upon an automated TEM mapping and image processing. The open-source TEM control software SerialEM was used to create large-scale (approximately  $20 \mu\text{m}^2$ ) maps of the surface of the TEM grid via 'montaging'.<sup>42</sup> Montaging works by acquiring a series of overlapping images in a grid-like pattern over a pre-defined area, and cross-correlating these images to stitch them into a single large image (a 'montage', see Figure 2). Several easily identified features of the stitched montage are then aligned to their absolute position in the microscope coordinate system, and a 'montage map' is created. Features of interest in the montage map can then be selected and the microscope is able to automatically locate that feature in the live image feed for subsequent tilt-series acquisition.





**FIGURE 2** Example of a large-scale stitched montage used for mapping (left) and zoomed section showing soot nanoparticles at  $\times 10,000$  magnification (right)



**FIGURE 3** (Left) Sub-section from a montage map showing two soot nanoparticles, one globular amorphous carbon structure, and unclustered gold nanoparticles in the background. (Right) Automated selections from the image processing algorithm, showing two true-positive selections of the soot nanoparticles, and a false-positive selection of the amorphous carbon

To identify features of interest, the montage maps are rapidly processed using an automated image processing algorithm implemented in ImageJ. A series of filtering and thresholding processes, along with constraints on selection size and shape, are applied to automatically identify potential nanoparticles. In practice, however, imaging of soot extracted from engine lubricant oil was prone to false-positives due to the presence of the non-soot amorphous carbon nanoparticles (see Figure 3). Hence, a manual review process by the human operator was required to separate selections of soot nanoparticles and non-soot nanoparticles. ResNet is used in this study as a potential replacement for the manual review process, and its integration into the automated soot characterisation procedure is summarised in Figure 4.

ResNet training and testing computations were carried out on MATLAB using an NVIDIA GeForce GTX 1650 GPU. The architecture of ResNet was modified to replace the fully connected layer with a layer containing pre-classified images of soot and non-soot nanoparticles (i.e. the training-validation set). The image input size to ResNet is  $224 \times 224$  pixels,<sup>33</sup> as a result, both the images used for training and the images used for testing

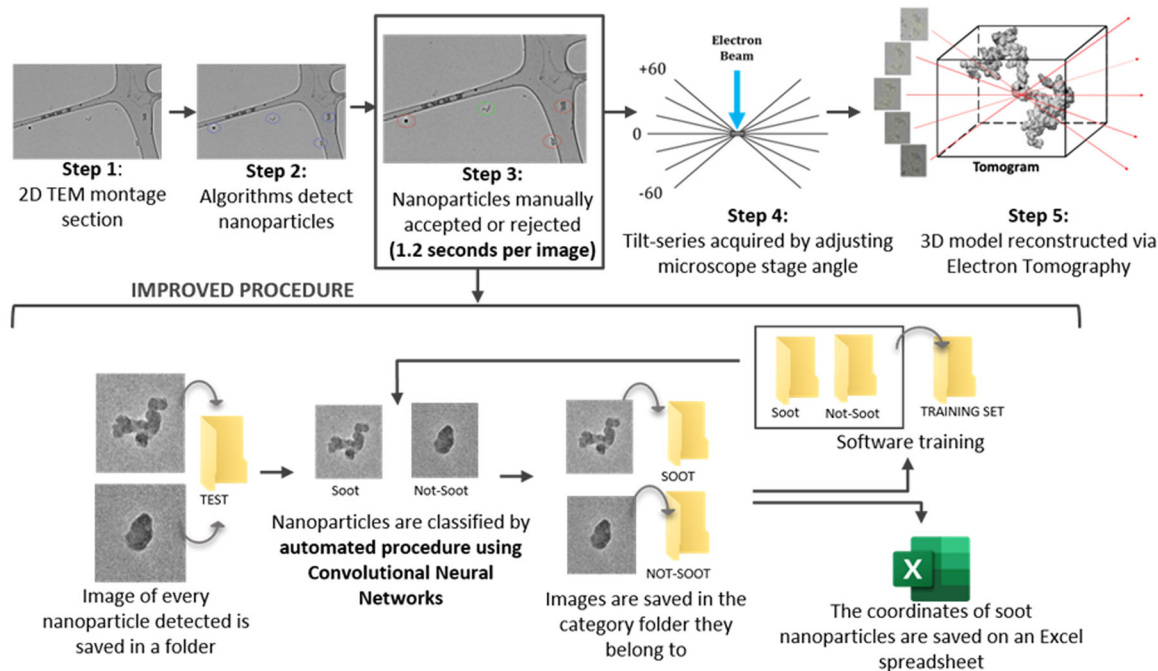
(analysis) must have a squared shape to preserve the aspect ratio once the image sizes are converted. Trained networks were testing by being used to classify an independent test set containing 100 images of soot and 100 images of non-soot nanoparticles. The primary metrics used to determine the quality of test set classification were accuracy and F1 score:

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN}, \quad (1)$$

$$F1\text{score} = \frac{2 \times TP}{2 \times TP + FN + FP}, \quad (2)$$

where the true positive (TP) is the number of soot images correctly classified as soot. The true negative (TN) is the number of non-soot images correctly classified as non-soot. The false positive (FP) is the number of non-soot images incorrectly classified as soot. The false negative (FN) is the number of soot images incorrectly classified as non-soot. Table 1 shows how the values relate to each other.

Accuracy indicates the overall performance in detecting the true positive (TP) and the true negative (TN), i.e.



**FIGURE 4** Current and improved procedure to rapidly identify and reconstruction soot nanoparticles in 3D. Details of electron tomography procedure in steps 4 and 5 can be found in prior publications<sup>19,24</sup>

**TABLE 1** Confusion matrix

Predicted category	Actual category	
	Soot	Non-soot
Soot	True positive	False positive
Non-soot	False negative	True negative

correctly classifying soot and correctly classifying non-soot. Accuracy is equally balanced between TP and TN classifications, so would not indicate if a network is much better at classifying non-soot than soot, or vice versa. The F1 score is the harmonic mean of the precision (the proportion of ‘soot’ classifications that are actually soot nanoparticles) and the recall (the percentage of soot nanoparticles that were correctly classified as ‘soot’), and thus is weighted towards TP selections. In a traditional soot sample characterisation, a higher F1 score would be preferred as correct soot classifications are the primary goal. However, as morphological characterisation of non-soot nanoparticles is also of interest to our research, accuracy is considered as the most important metric.

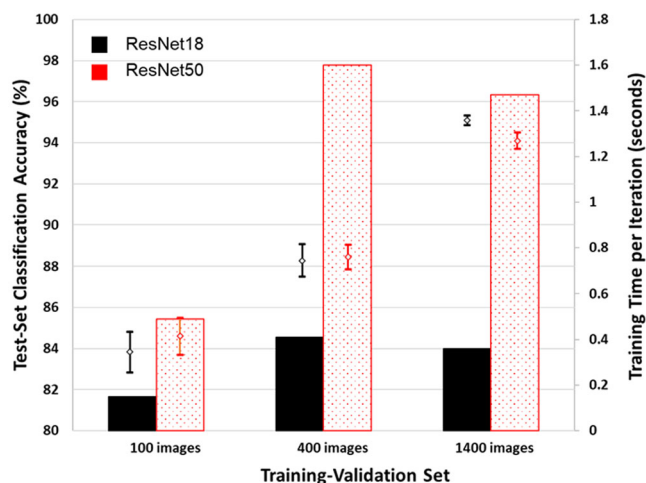
## 2 | RESULTS AND DISCUSSION

First, the performances of ResNet18 and ResNet50 were measured as a function of 3 training-validation set sizes. As described in the methodology, an independent test set of 200 images was separated from the rest of the data for

**TABLE 2** Mini-batch sizes and learning rates used in optimisation of network trainings settings

Mini-batch Size	Learning rate		
	0.01	0.001	0.0001
8	8, 0.01	8, 0.001	8, 0.0001
32	32, 0.01	32, 0.001	32, 0.0001
64	64, 0.01	64, 0.001	64, 0.0001

classification assessments. Three ‘training-validation’ sets were created from the remaining 1400 images, containing 100, 400 and 1400 images each (a 1:1 ratio of soot and non-soot images in all cases). The training-validation split was 70:30 in all cases, the validation frequency was fixed at once per training epoch and the stochastic gradient descent with momentum (SGDM) solver was used. Optimisation of the network training settings was carried out via varying the mini-batch size and the (constant) learning rate. Three mini-batch sizes (8, 32, 64 images) and three constant learning rates (0.01, 0.001, 0.0001) were tested, and at least 3 repeat network trainings were carried out for each combination (see Table 2). These trainings were carried out for an initial 100 training epochs, with the number of training iterations varying depending on the mini-batch size and training-validation set size. Following identification of the optimal training settings, the number of epochs/iterations for network training was also optimised as detailed in the following section.



**FIGURE 5** Optimum accuracies of ResNet18 and ResNet50 networks trained with training-validation sets of increasing size, and training time required in terms of seconds per iteration. Error bars shown are 90% confidence intervals. Means and error bars are calculated from 9 repeated networks trainings in each

Following the tests on the effect of training-validation set sizes, ResNet18 and ResNet50 were ultimately compared for quality on the classification of soot and non-soot nanoparticles via a fivefold cross-validation. The total set of 1600 images was divided into 5 equal ‘folds’ of 320 images, each containing 160 randomly selected images of soot and non-soot. For a single cross-validation experiment, ResNet18 and ResNet50 were trained with this data 5 times, with each fold being used as the validation data once and as training data the other 4 times. At the end of training, the final accuracy measurements were taken by using the network to classify the validation set. The accuracy was then averaged over the 5 trainings to give the final accuracy of the network. For both ResNet18 and ResNet50, training during cross-validation was carried out using the optimal training settings found for the 1400-image training-validation set. The fivefold cross-validation experiment was repeated 3 times, with the networks being trained for 5, 20 and 50 epochs, respectively.

## 2.1 | Performance as a function of training-validation set size

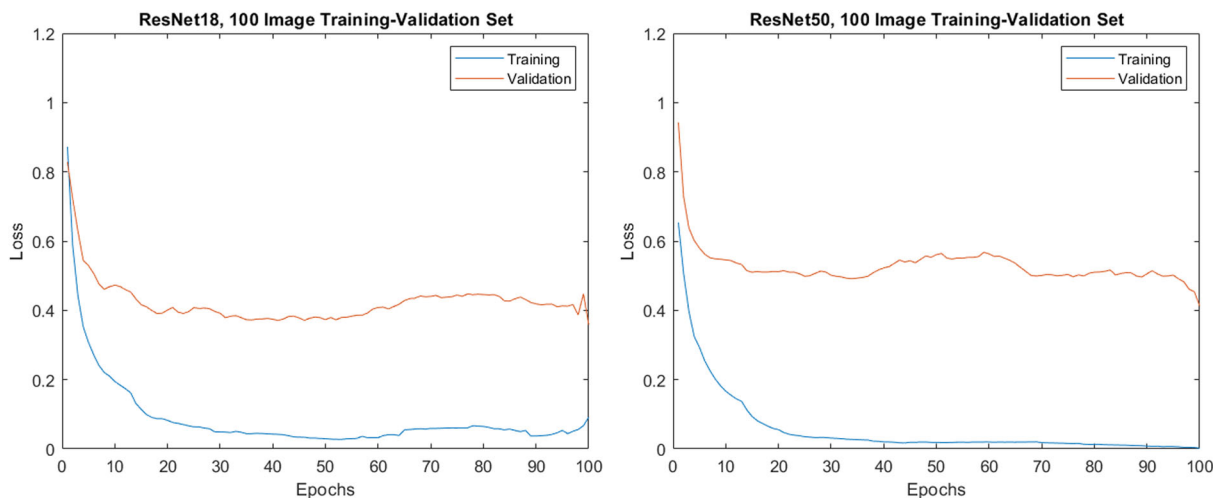
The optimal results from the network trainings using the 3 training-validation sets are shown in Figure 5 and summarised in Table 3.

For the smallest training-validation set, the optimal training settings were a mini-batch size of 8 and a learn rate of 0.0001 for both ResNet18 and ResNet50. For all training-validation set sizes, the combination of a 64-image mini-batch size and a learning rate of 0.0001 produced the lowest accuracies. Previous studies have observed inverse

**TABLE 3** Optimum results of ResNet18 and ResNet50 networks trained with training-validation sets of increasing size

Network	Training-validation images	Mini-batch size	Learn rate	Training epochs	TP	TN	FP	FN	Accuracy ( $\pm 90\%$ CI)	F1 score ( $\pm 90\%$ CI)	Seconds per iteration
ResNet18	100	8	0.0001	200	77.0	90.7	9.3	23.0	83.8% ( $\pm 1.0$ )	82.6 ( $\pm 1.2$ )	0.15
ResNet50	100	8	0.0001	200	78.0	91.2	8.8	22.0	84.6% ( $\pm 0.9$ )	83.5 ( $\pm 0.9$ )	0.49
ResNet18	400	32	0.0001	50	89.7	88.0	12.0	10.3	88.3% ( $\pm 0.8$ )	88.2 ( $\pm 0.8$ )	0.41
ResNet50	400	32	0.0001	50	89.7	87.0	13.0	10.3	88.4% ( $\pm 0.6$ )	88.5 ( $\pm 0.6$ )	1.60
ResNet18	1400	32	0.001	50	92.7	97.5	2.5	7.3	95.1% ( $\pm 0.2$ )	95.0 ( $\pm 0.2$ )	0.36
ResNet50	1400	32	0.001	50	94.2	94.0	6.0	5.8	94.1% ( $\pm 0.4$ )	94.1 ( $\pm 0.4$ )	1.47

Note: Results shown are averages of 9 repeated trainings in all cases.

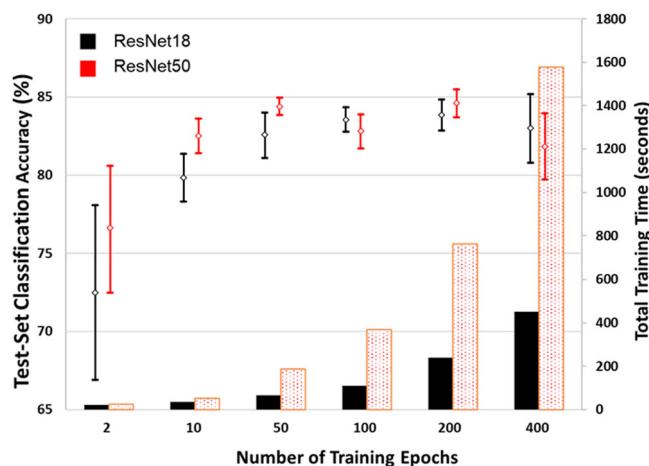


**FIGURE 6** (Left) Training-validation loss profile for ResNet18, trained for 100 epochs on the 100 image training-validation set. (Right) Training-validation loss profile for ResNet50, trained for 100 epochs on the 100 image training-validation set. Note rapid initial decreases in both training and validation loss, followed by period of steady difference between training and validation loss, without noticeable increase in validation loss (i.e. overfitting)

relationships between mini-batch size and learning rate, with larger mini-batch sizes observing improved training when used alongside larger learning rates and vice versa.<sup>43</sup>

Following the training settings optimisation, the training length was optimised by observing the training-validation accuracy and loss values. The aim of training a network like ResNet is generalisation, i.e., producing a network that accurately classifies new data from outside of the training-validation set. Typically, training will continue until a point at which ‘overfitting’ begins to occur, after which the network accuracy would start to decrease. Overfitting occurs when the network begins to ‘memorise’ features of the training images, at the expense of its ability to generalise. Overfitting is seen when improved classification of the training set occurs alongside decreased accuracy when classifying the validation set. One method to avoid overfitting and the associated generalisation errors is to perform early stopping,<sup>44</sup> ceasing training at a point where the validation loss begins to increase above a certain threshold. Training-validation loss curves for ResNet18 and ResNet50 trained on the 100-image set are shown in Figure 6, where some interesting features can be observed.

For the majority of network trainings in this study, the loss curves are similar to those shown in Figure 6. A decrease in both training and validation loss is strong during the first 5–10 epochs of training. After this point, there quickly becomes a deviation between training and validation loss, after which the separation between them remains relatively constant with increasing training duration. Though there are small fluctuations in the validation loss that occur there is generally no obvious onset of overfitting and as such no obvious point at which early stopping

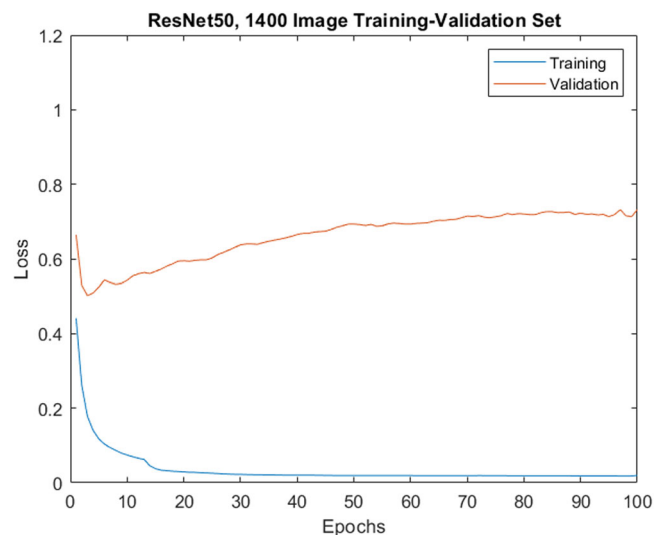


**FIGURE 7** Effect of number of training epochs on the accuracy of networks trained on the 100-image training-validation set. Time required for training also shown

might be applied. To understand the effect of increasing training length on the accuracy of these networks, a series of tests were carried out at different training lengths. Training was repeated on the 100-image set over a range of 2–400 epochs for both ResNet18 and ResNet50, results for which are shown in Figure 7 (where all network trainings were repeated at least 6 times).

For both ResNet18 and ResNet50 the highest average accuracy was observed after training for 200 epochs, at 83.8% and 84.3%, respectively (these results are quoted in Table 3). However, overlapping of the 90% confidence intervals occurred from 50 epochs onwards for ResNet18, and from 10 epochs onwards for ResNet50, showing that there was little benefit of increased training duration





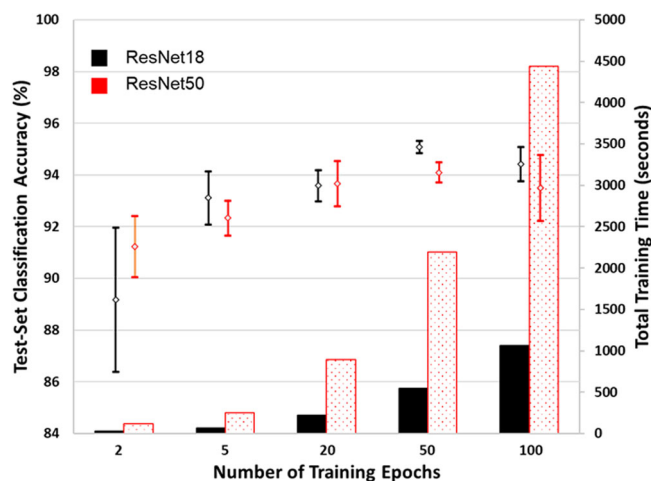
**FIGURE 8** Training-validation loss profile for ResNet50 trained on the 1400 image training-validation set. Note the increase in validation loss occurring from around 2 epochs onwards (overfitting)

within the region of steady training-validation loss difference. For each network, the training time varies approximately linearly with the number of training epochs. ResNet50 is around 3–4 times slower than ResNet18 for trainings from 50 epochs upwards. Even with the limited hardware used in this study, training times are modest.

For the 400-image training-validation set, the optimum training settings shifted to a mini-batch size of 32, and a learning rate of 0.0001. Similar tests were carried out to understand the effect of training length, though were limited to 100 epochs following the observations for the 100-image set that increased training length above 50 epochs did not lead to significant improvements in accuracy. An optimum training length of 50 epochs was observed and resulted in a test-set classification accuracy of 88.3% for ResNet18 (in 162 s) and 88.4% for ResNet50 (in 641 s).

For the largest training-validation set of 1400 images, the optimum training settings were a mini-batch size of 32 and a learning rate of 0.001. For this set, the loss data for ResNet18 trainings again showed relatively steady validation loss with increased training duration. However, for ResNet50 the curves showed increasing validation losses occurring from around 2 epochs onwards (see Figure 8). Repeated network trainings were again carried out over a range of lengths, from 2 to 100 epochs. The effect of training length on the accuracy of networks for the 1400-image set is shown in detail in Figure 9, where all network trainings were repeated at least 6 times.

For all trainings except the 2-epoch training with ResNet18, test-set classification accuracy was >90%. Both



**FIGURE 9** Effect of number of training epochs on the accuracy of networks trained on the 1400 image training-validation set. Time required for training also shown

networks observe increasing accuracy up to 50 epochs of training, reaching 95.1% and 94.1% for ResNet18 and ResNet50, respectively, before decreasing slightly with a longer training duration of 100 epochs. There was no major difference in the effect of different training lengths on the accuracy of ResNet18 and ResNet50, despite the differences in the loss curves that were observed. As for the 100-image set, training time varied approximately linearly with the number of training epochs, and ResNet50 was around 3–4 times slower than ResNet18.

As may have been expected, increasing the number of training-validation images led to an increase in the accuracy of both ResNet18 and ResNet50 networks. Accuracy increased from 83–84% for a 100-image set to 94–95% for a 1400-image set. An interesting observation to note was that training with the 100-image set led to networks that were better at classifying the non-soot nanoparticles than the soot nanoparticles. The average number of correctly classified non-soot nanoparticles in the test set (true negatives) was around 15% greater than the number of correctly classified soot nanoparticles (true positives). Networks trained on the larger training-validation sets exhibited approximately equal numbers of true positives and true negatives, reflected in almost identical accuracies and *F1* scores, suggesting similar ability to classify soot and non-soot. As shown in Figure 10, when the number of training epochs was kept constant the time required for training varied approximately linearly with the number of training images. The time required per iteration of training did not vary linearly (see Table 3), as the number of iterations depends on the mini-batch size which changed between the training-validation sets.

Overall, performance of ResNet18 and ResNet50 was almost identical for each of the 3 training-validation sets,



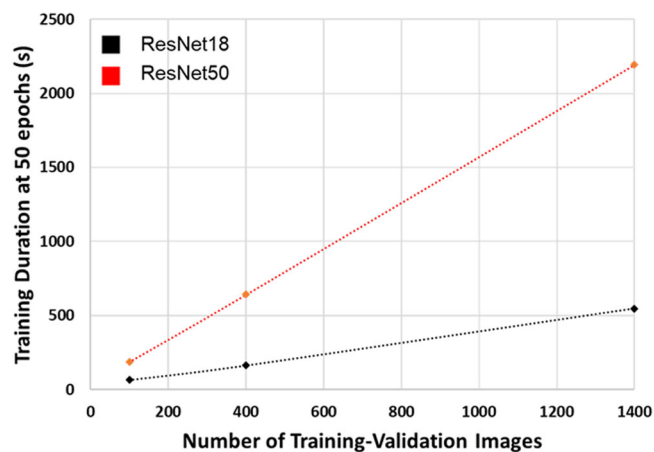


FIGURE 10 Network training time required for 50 epochs of training at the 3 training-validation set sizes

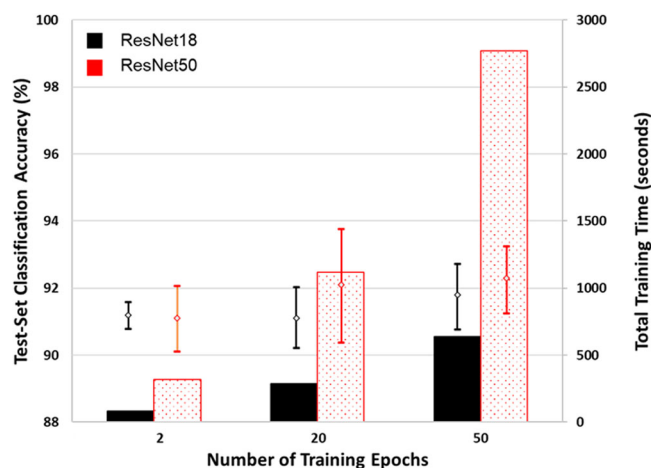


FIGURE 11 Results of fivefold cross-validation experiment at 3 different training lengths

though training of ResNet18 took only 25–35% of the time required for training ResNet50. Additionally, longer training durations were most beneficial for the 100-image set, while training for longer than 5 epochs had only minor benefits for the 1400-image set.

To ultimately compare the quality of ResNet18 and ResNet50 for the problem of soot sample classification, a fivefold cross-validation experiment was carried out. The complete set of 1600 images was randomly divided into 5 ‘folds’ of 320 images, each containing 160 soot and non-soot nanoparticles. Network training was carried out 5 times, with each fold acting as the validation set once and as a training set 4 times. Training was carried out using the mini-batch and learn rate settings derived for the 1400-image training-validation set. The cross-validation experiment was repeated with 3 different trainings lengths: 5, 20 and 50 epochs. Results are shown in Table 4 and summarised in Figure 11.

TABLE 4 Results of fivefold cross-validation experiment at 3 different training lengths

Validation fold	5 epochs		20 epochs		50 epochs	
	ResNet18	ResNet50	ResNet18	ResNet50	ResNet18	ResNet50
1	91.7	93.4	91.9	94.4	91.6	93.4
2	90.0	90.6	90.6	91.6	90.9	91.6
3	91.6	90.2	91.6	90.6	92.8	90.9
4	91.4	91.9	91.9	93.4	92.8	93.1
5	91.3	89.4	89.7	90.3	90.6	92.2
Average accuracy %	91.2	91.1	91.1	92.1	91.8	92.3
Average duration (s)	84	318	285	1119	637	2771

All trainings in the cross-validation experiment showed average accuracies above 91%, with no significant difference observed between ResNet18 and ResNet50. Additionally, the performance after just 5 epochs of training (150 iterations) produced networks of similar quality to 50 epochs of training. Based on these results, the optimum network for the classification of soot and non-soot particles is ResNet18 trained for only 5 epochs, where training required only 84 s for 1600 images. The classification procedure required only 1.7 s for 320 images, which is of similar magnitude to manual classification of a single TEM image by a human operator, highlighting benefits to throughput when using CNNs for classification.

Overall, due to its faster training speed and similar levels of accuracy, ResNet18 would be considered the better network depth to use for the problem of classifying soot and non-soot nanoparticles, especially in situations where available hardware is limited. Based on the accuracies measured, the use of the convolution neural network ResNet for the problem of soot/non-soot classification is most applicable when a larger amount of training data is available (>400 images). This would equate to a scenario where a sample to be studied is similar to one imaged previously, for example, a repeat measurement or a sample derived from the same engine test, where set of training images would already exist. Applicability to novel samples is limited, as accuracies for a 100-image training set (83–84%) would lead a significant number of incorrectly classified particles being measured and thus significant errors in the morphological characterisation of the sample. Based on the work of Kondo et al., it is suggested that 1000 individual nanoparticles are required to be measured to reliably characterise a sample of soot. Using a manual review process, a highly trained operator would require at least 30 min to categorise images whereas a pre-trained network could complete this almost instantly (<4 s), with as little as 84 s required to train a network to >91% accuracy with a training-validation set of >1000 images. Critically, automated identification significantly reduces the human labour cost and removes the need for a skilled operator, which is a key aspect of a true high-throughput procedure.

### 3 | CONCLUSIONS

This study focused on the development of an automated procedure for soot nanoparticle recognition, for incorporation into a previously developed semi-automated procedure for soot analysis in 2D and 3D. By implementing ‘transfer learning’ on a pre-trained convolutional neural network architecture, the 3D morphological reconstruction procedure of soot nanoparticles is made more efficient by reducing the time of operation. More

specifically, CNNs are used to replace human supervision to classify the nanoparticles detected into ‘soot’ and ‘non-soot’ categories, which can consequently be reconstructed in three dimensions. This work compared two depths of ResNet for the task of soot and non-soot nanoparticle classification and included investigation of the effect of training-validation set size on network accuracy.

- Accuracy of soot/non-soot nanoparticle classification using ResNet increased with increasing number of training images. Training of ResNet with training-validation set sizes of 100, 400 and 1400 images led to classifications accuracies of 83–84%, 88% and 94–95%, respectively.
- For all the training-validation set sizes used in this study, ResNet18 and ResNet50 had almost identical classification accuracies, though ResNet18 required only 25–35% of the time required for training ResNet50.
- Training with a 100-image training-validation set produced networks significantly better at classifying non-soot nanoparticles than soot nanoparticles. Training with larger sets of images produced networks with equivalent ability to categorise soot and non-soot.
- Tests on the effect of training duration on network accuracy showed that optimal training length was short at around 50 epochs, with increased training durations leading to statistically insignificant differences in accuracy.
- Training time varied approximately linearly with number of training images and with number of training epochs. Training time per iteration depends on the network training settings.
- CNNs appear most suitable for soot/non-soot classifications when >400 images that could be used for a training-validation set already exist, for example, a repeat sample study or a sample derived from the same engine testing procedure.
- The fivefold cross-validation experiment showed no significant differences in the accuracy of ResNet18 and ResNet50. No significant difference in accuracy was observed between networks trained for 5, 20 or 50 epochs, showing that optimal network training length was short.
- The optimal result of cross-validation was ResNet18 trained for 5 epochs (on a 1600-image training-validation set). This training required only 84 s and reached 91.2% classification accuracy.
- Classification time using trained networks was negligible, requiring only 1.7 s for 320 images and <4 s for 1000 images. By comparison, classification by a human operator requires up to 2 s per single image, or >30 min to classify 1000 images (the amount suggested to be necessary for reliable morphology characterisations). Classification time is decreased by a factor of

approximately 400–500 when using the automated procedure rather than the manual identification.

- Furthermore, since the training procedure does not need supervision, the software can run during low-demand hours when trained with a large data set of thousands of images.
- The technology proves to be effective when combined with transmission electron microscopy, hence it can be adapted to further research tasks involving nanoparticles' detection and TEM, opening a wide range of opportunities to apply pre-trained convolutional neural networks by adopting the 'transfer learning' approach.

## ACKNOWLEDGEMENTS

This work was supported by the Engineering and Physical Sciences Research Council through the scholarship provided by the EPSRC Thematic Programme in Low-Dimensional Materials and Interfaces for Sebastian Pfau (grant number EP/N50970X/1).

## ORCID

E. Haffner-Staton  <https://orcid.org/0000-0003-0914-795X>

A. La Rocca  <https://orcid.org/0000-0002-0335-415X>

## REFERENCES

1. Mohankumar, S., & Senthilkumar, P. (2017). Particulate matter formation and its control methodologies for diesel engine: A comprehensive review. *Renewable and Sustainable Energy Reviews*, 80, 1227–1238.
2. Tan, P.-Q., Deng, K., & Lu, J.-X. (2004). Analysis of particulate matter composition from a heavy-duty diesel engine. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, 218(11), 1325–1331.
3. Commission Regulation (EU). (2017). 2017/1151 of 1 June 2017 supplementing Regulation (EC) No 715/2007 of the European Parliament and of the Council on type-approval of motor vehicles with respect to emissions from light passenger and commercial vehicles (Euro 5 and Euro 6) and on access to vehicle repair and maintenance information, amending Directive 2007/46/EC of the European Parliament and of the Council, Commission Regulation (EC) No 692/2008 and Commission Regulation (EU) No 1230/2012 and repealing Commission Regulation (EC) No 692/2008. *Official Journal of the European Union*, L175, 1–732.
4. Kalghatgi, G. (2019). Development of fuel/engine systems—the way forward to sustainable transport. *Engineering*, 5(3), 510–518.
5. CCC-Net-Zero. (2019). *The UK's contribution to stopping global warming*. London, UK: Committee on Climate Change.
6. Niranjana, R., & Thakur, A. K. (2017). The toxicological mechanisms of environmental soot (black carbon) and carbon black: Focus on oxidative stress and inflammatory pathways. *Frontiers in Immunology*, 8, 763.
7. Heidari Nejad, S., Takechi, R., Mullins, B. J., Giles, C., Larcombe, A. N., Bertolatti, D., Rumchev, K., Dhaliwal, S., & Mamo, J. (2015). The effect of diesel exhaust exposure on blood–brain barrier integrity and function in a murine model. *Journal of Applied Toxicology*, 35(1), 41–47.
8. Broday, D. M., & Rosenzweig, R. (2011). Deposition of fractal-like soot aggregates in the human respiratory tract. *Journal of Aerosol Science*, 42(6), 372–386.
9. Löndahl, J., Massling, A., Swietlicki, E., Brüner, E. V., Ketzel, M., Pagels, J., & Loft, S. (2009). Experimentally determined human respiratory tract deposition of airborne particles at a busy street. *Environmental Science & Technology*, 43(13), 4659–4664.
10. Oberdörster, G., Celein, R. M., Ferin, J., & Weiss, B. (1995). Association of particulate air pollution and acute mortality: Involvement of ultrafine particles? *Inhalation Toxicology*, 7(1), 111–124.
11. Liu, L., & Mishchenko, M. I. (2007). Scattering and radiative properties of complex soot and soot-containing aggregate particles. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 106(1–3), 262–273.
12. Baldelli, A., Trivanovic, U., Sipkens, T. A., & Rogak, S. N. (2020). On determining soot maturity: A review of the role of microscopy-and spectroscopy-based techniques. *Chemosphere*, 252, 126532.
13. Clague, A. D. H., Donnet, J., Wang, T. K., & Peng, J. C. M. (1999). A comparison of diesel engine soot with carbon black. *Carbon*, 37(10), 1553–1565.
14. Park, K., Kittelson, D. B., & McMurry, P. H. (2004). Structural properties of diesel exhaust particles measured by transmission electron microscopy (TEM): Relationships to particle mass and mobility. *Aerosol Science and Technology*, 38(9), 881–889.
15. Barone, T. L., Storey, J. M. E., Youngquist, A. D., & Szybist, J. P. (2012). An analysis of direct-injection spark-ignition (DISI) soot morphology. *Atmospheric Environment*, 49, 268–274.
16. Dastanpour, R., & Rogak, S. N. (2014). Observations of a correlation between primary particle and aggregate size for soot particles. *Aerosol Science and Technology*, 48(10), 1043–1049.
17. Haffner-Staton, E., La Rocca, A., Cairns, A., McGhee, M., & Fay, M. (2019). Assessing the accuracy of soot nanoparticle morphology measurements using three-dimensional electron tomography. SAE Technical Paper, Report No.: 0148–7191.
18. Koylu, U. O., Faeth, G. M., Farias, T. L., & Carvalho, M. G. (1995). Fractal and projected structure properties of soot aggregates. *Combustion and Flame*, 100(4), 621–633.
19. Haffner-Staton, E., La Rocca, A., & Fay, M. W. (2018). Progress towards a methodology for high throughput 3D reconstruction of soot nanoparticles via electron tomography. *Journal of Microscopy*, 270(3), 272–289.
20. Baldelli, A., Trivanovic, U., & Rogak, S. N. (2019). Electron tomography of soot for validation of 2D image processing and observation of new structural features. *Aerosol Science and Technology*, 53(5), 575–582.
21. Adachi, K., Chung, S. H., Friedrich, H., & Buseck, P. R. (2007). Fractal parameters of individual soot particles determined using electron tomography: Implications for optical properties. *Journal of Geophysical Research: Atmospheres*, 112(D14). <https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2006JD008296>
22. van Poppel, L. H., Friedrich, H., Spinsby, J., Chung, S. H., Seinfeld, J. H., & Buseck, P. R. (2005). Electron tomography

- of nanoparticle clusters: Implications for atmospheric lifetimes and radiative forcing of soot. *Geophysical Research Letters*, 32(L24811). <https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2005GL024461>
23. Haffner-Staton, E., La Rocca, A., Cairns, A., & Fay, M. (2019). Morphological characterization of gasoline soot-in-oil: Development of semi-automated 2D-TEM and comparison with novel high-throughput 3D-TEM. SAE Technical Paper, Report No.: 0148-7191.
  24. Orhan, O., Haffner-Staton, E., La Rocca, A., & Fay, M. (2016). Characterisation of flame-generated soot and soot-in-oil using electron tomography volume reconstructions and comparison with traditional 2D-TEM measurements. *Tribology International*, 104, 272-284.
  25. Adachi, K., Chung, S. H., & Buseck, P. R. (2010). Shapes of soot aerosol particles and implications for their effects on climate. *Journal of Geophysical Research: Atmospheres*, 115.
  26. Kondo, K., Aizawa, T., Kook, S., & Pickett, L. (2013). Uncertainty in sampling and TEM analysis of soot particles in diesel spray flame. SAE Technical Paper, Report No.: 0148-7191.
  27. Marcus, G. (2018). Deep learning: A critical appraisal. arXiv preprint arXiv:180100631.
  28. Gonzalez, R. C., & Woods, R. E. (2018). *Digital image processing*. Pearson.
  29. Albawi, S., Mohammed, T.A., & Al-Zawi, S. (Eds.) (2017). Understanding of a convolutional neural network. In *2017 International Conference on Engineering and Technology (ICET)*. IEEE.
  30. Girshick, R., Donahue, J., Darrell, T., & Malik, J. (Eds.) (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE.
  31. Wachinger, C., Reuter, M., & Klein, T. (2018). DeepNAT: Deep convolutional neural network for segmenting neuroanatomy. *Neuroimage*, 170, 434-445.
  32. Li, Q., Zhao, T., Zhang, L., Sun, W., & Zhao, X. (2017). Ferrography wear particles image recognition based on extreme learning machine. *Journal of Electrical and Computer Engineering*, 2017. <https://www.hindawi.com/journals/jece/2017/3451358/>
  33. Raghu, S., Sriraam, N., Temel, Y., Rao, S. V., & Kubben, P. L. (2020). EEG based multi-class seizure type classification using convolutional neural network and transfer learning. *Neural Networks*, 124, 202-212.
  34. Shin, H.-C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., & Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*, 35(5), 1285-1298.
  35. Wu, Z., Shen, C., & Van Den Hengel, A. (2019). Wider or deeper: Revisiting the ResNet model for visual recognition. *Pattern Recognition*, 90, 119-133.
  36. He, K., Zhang, X., Ren, S., & Sun, J. (Eds.) (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE.
  37. Yu, X., & Wang, S.-H. (2019). Abnormality diagnosis in mammograms by transfer learning based on ResNet18. *Fundamenta Informaticae*, 168(2-4), 219-230.
  38. Nibali, A., He, Z., & Wollersheim, D. (2017). Pulmonary nodule classification with deep residual networks. *International Journal of Computer Assisted Radiology and Surgery*, 12(10), 1799-808.
  39. Luo, C., Li, X., Wang, L., He, J., Li, D., & Zhou, J. (Eds.) (2018). How does the data set affect CNN-based image classification performance? In *5th International Conference on Systems and Informatics (ICSAI)*. IEEE.
  40. Haffner-Staton, E., La Rocca, A., Cairns, A., & Fay, M. (2019). Morphological characterisation of gasoline soot-in-oil: Development of semi-automated 2D-TEM and comparison with novel high-throughput 3D-TEM. SAE Technical Paper, Report No.: 0148-7191.
  41. Uy, D., Ford, M. A., Jayne, D. T., O'Neill, A. E., Haack, L. P., Hangan, J., Jagner, M. J., Sammut, A., & Gangopadhyay, A. K. (2014). Characterization of gasoline soot and comparison to diesel soot: Morphology, chemistry, and wear. *Tribology International*, 80, 198-209.
  42. Mastrorade, D. N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. *Journal of Structural Biology*, 152(1), 36-51.
  43. Smith, S. L., Kindermans, P.-J., Ying, C., & Le, Q. V. (2017). Don't decay the learning rate, increase the batch size. arXiv preprint arXiv:171100489.
  44. Prechelt, L. (1998). Early stopping-but when? In: Orr, G. B., Muller, K. R. (eds) *Neural networks: Tricks of the trade* (pp. 5-69). Springer.

**How to cite this article:** Haffner-Staton, E., Avanzini, L., La Rocca, A., Pfau, S. A., & Cairns, A. (2022). Automated particle recognition for engine soot nanoparticles. *Journal of Microscopy*, 288, 28-39. <https://doi.org/10.1111/jmi.13140>