COMPUTER METHODS IN MATERIALS SCIENCE

2024, vol. 24, no. 3, 33–40
https://doi.org/10.7494/cmms.2024.3.0839

An evaluation of the capabilities of image-based metal component defect recognition with deep learning techniques

Michał P. Wójcik*, Kacper Pawlikowski, Łukasz Madej

AGH University of Krakow, Mickiewicza 30, 30-059 Krakow, Poland.

Abstract

In the era of Industry 4.0, deploying highly specialised machine learning models trained on unique and often scarce datasets is an attractive solution for advancing automated quality control and minimising production costs. Therefore, the main aim of this research is to evaluate the capabilities of three deep learning models (*ResNet-18*, *ResNet-50* and *SE-ResNeXt-101 (32 × 4d)*) in the identification of surface defects in forged products. Leveraging advanced photography techniques, including studio lighting and a shadowless box, high-quality images of complex product surfaces were acquired for the training data set. Given the relatively small size of the image dataset, aggressive data augmentation techniques were introduced during the training and evaluation process to ensure robust model generalisation ability. The results obtained demonstrate the significant impact of data augmentation on model performance, highlighting its importance in training and evaluating deep learning models with limited data. This research also emphasises the need for innovative data pre-processing strategies in an efficient and robust machine learning model delivery to the industrial environment.

Keywords: deep learning, convolutional neural networks, image classification, data augmentation, quality control, surface defect recognition, forging

1. Introduction

In recent years, deep learning has revolutionised the field of computer vision, enabling unprecedented advances in image recognition, object detection, and semantic segmentation (Chai et al., 2021). Convolutional neural networks (CNNs), a cornerstone of this progress, have demonstrated exceptional performance in various tasks by leveraging hierarchical feature extraction and spatial hierarchies (LeCun et al., 2015). The gradient backpropagation algorithm (Rumelhart et al., 1986) plays an important role in these approaches. By efficiently updating the weights of neural networks through gradient descent, backpropagation allows for the training of deep neural models and enables them to learn intricate patterns and features from vast amounts of data (LeCun et al., 2015). However, the success of these approaches relies directly on the large amount of training data involved. To minimise this constraint, an approach of data augmentation or transfer learning can be used.

In the first case, e.g., *AutoAugment* (Cubuk et al., 2019), which is a data augmentation strategy finding algorithm designed to enhance the performance of machine learning models by automatically discovering the optimal augmentation policies, can be used. It leverages a reinforcement learning framework where a controller, typically a recurrent neural network (RNN), generates candidate augmentation policies. These policies, comprising combinations of transformations such as rotations, translations, and colour adjustments, are applied to subsets of the training data to train a child model. The performance of such a child model on a validation set

* Corresponding author: wojcik.michal.2001@gmail.com
ORCID ID's 0009-0004-4279-0124 (M. P. Wójcik), 0000-0002-3990-1661 (K. Pawlikowski), 0000-0003-1032-6963 (Ł. Madej)

serves as a reward signal to iteratively refine the controller using the proximal policy optimisation (PPO) algorithm. This automated process eliminates the need for manual design of augmentation strategies, enabling the discovery of complex and dataset-specific augmentation policies that significantly improve model generalisation and performance and do not hide relevant image features. In particular, *AutoAugment* within the machine learning framework is implemented as ready-to-use, pre-trained augmentation policies that are divided into subpolicies, each consisting of two transformations and applied to images separately.

In the second case, by leveraging pre-trained models on large datasets, transfer learning allows for the adaptation of these models to specific tasks with relatively small amounts of task-specific data (Shin et al., 2016). This approach significantly reduces the computational resources and training time required while still achieving high levels of accuracy. In the context of computer vision, CNNs pre-trained on large image datasets (like, e.g., *ImageNet*; Deng et al., 2009) are fine-tuned on new smaller image datasets, enhancing their ability to perform specific image classification or object detection tasks with high performance and efficiency (Tan Ch. et al., 2018). As the former is particularly valuable for industrial applications, intensive research is carried out in this area.

For example, ResNet (residual networks) (He et al., 2016) introduced the concept of residual learning, enabling the training of much deeper networks by mitigating the vanishing gradient problem through identity shortcut connections. Then, ResNeXt (Xie et al., 2017) employed a cardinality dimension, enhancing the model's representational power with multiple parallel paths within each block. DenseNet (densely connected convolutional networks) discussed in (Huang et al., 2017) was developed to maximise information flow between layers via dense connections, where each layer receives the feature maps of all preceding layers as inputs, leading to more efficient parameter usage and improved gradient flow. In 2018, the squeeze-and-excitation networks (SENets) (Hu et al., 2018) introduced a novel channel-wise attention mechanism, adaptively recalibrating channel-wise feature responses, which significantly improved the model's capacity to capture complex dependencies between channels. EfficientNet (Tan M. & Le, 2019) efficiently scaled the network architecture by uniformly scaling all depth, width, and resolution dimensions using a compound scaling method, achieving better performance with fewer parameters. Recently, the vision transformers (ViT) (Dosovitskiy et al., 2021) brought a paradigm shift by leveraging self-attention mechanisms traditionally used in natural language processing and demonstrated that transformers could outperform CNNs on image classification tasks. Lastly, ConvNet (Liu et al., 2022) revisited and modernised the standard CNN architecture by incorporating successful design elements from vision transformers and advanced normalisation techniques.

All of these techniques are more often used by various industries in the area of automated non-destructive testing (NDT), which is crucial for effective quality control in smart manufacturing processes. Comprehensive studies indicated that the integration of CNNs has significantly enhanced the detection and classification of defects in, e.g., metal components. For instance, traditional image processing methods are being supplemented or replaced by advanced deep-learning approaches capable of handling noise, lighting variations, and complex textures (Bhatt et al., 2021; Jia et al., 2024; Niccolai et al., 2021). Some of the research also focuses on the application of imaging modalities like, e.g. microscopy (Tabernik et al., 2020), thermography (Lugin et al., 2023) or X-ray (Yang et al., 2020) in more elaborate identification of defects.

However, all of these approaches focus on architectural improvements and testing various imaging techniques, but none prioritise fast and computationally efficient detection or classification model delivery. This paper aims to address this issue by employing pretrained CNNs in the role of automated feature extractors, eliminating the necessity of hand-crafting classifiers with classic computer vision techniques and the *AutoAugment* generated augmentation policies that are applied to image-based binary classification of defective and non-defective forged components. The concept assumes leveraging transfer learning to enhance the model's performance on a small hand-crafted dataset. Additionally, a comparative analysis will be conducted among three deep CNN models (*ResNet-18*, *ResNet-50* and *SE-ResNeXt-101 (32 × 4d)*) to determine the model architecture complexity required for this task. ResNet-18, a relatively small and simple state-of-the-art model architecture, will be used as an entry point for model selection.

2. Dataset and pre-processing

The custom image dataset containing images of both defective and non-defective forged components was hand-crafted as the starting point of the research. These components are specifically banana-like, with rims on the inner side surface, which is also the most likely to have defects (Fig. 1). The image scene lightning was adjusted to avoid focusing light rays through the concave surface of the component. To achieve that, studio lighting and shadowless box were used. A significant

difference in the illumination of the inner part of the rim of the component was observed. Therefore, each component was placed on the custom 3D-printed stand and photographed in two positions: with a circular recess closer to the bottom (Fig. 1a) and with a circular recess closer to the top (Fig. 1b). By capturing images from both positions, confidence is ensured that any potentially shaded features were not overlooked or missed due to an unfocused upper portion of the photograph. This approach has been applied to both defected and non-defected component image subsets.

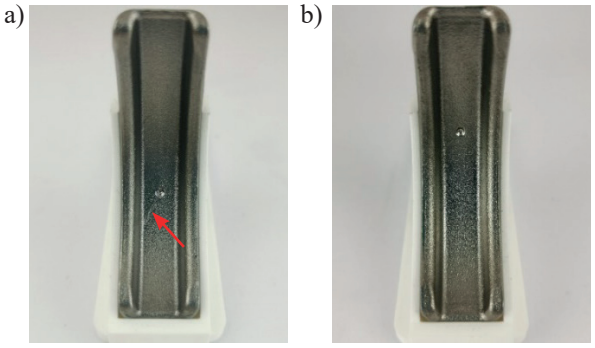

Fig. 1. Example of defective (scratch near recess indicated by red arrow) and non-defective component: a) recess closer to the bottom; b) recess closer to the top

The number of defective components that were available for this research is an order of magnitude greater than the number of non-defective products. The number of images in each set is presented in Table 1. The imbalance in the number of images in the second approach was mitigated using class augmentation techniques. Simultaneously, oversampling with duplication of randomly selected images for the class of non-defective product images and undersampling with a randomly selected minority for the class of defective component images were applied to bring the number of images in both classes closer to each other.

Table 1. The number of images in each class in both approaches, considering class augmentation

Number of images	Defective	Non-defective
Before class augmentation	452	40
After class augmentation	250	250

The *ImageNet* profiled *AutoAugment* generated augmentation policies were randomly applied to each training set image to avoid quick overfitting to a relatively small dataset and ensure maximal generalisation ability. The reason for that is the fact that all CNNs (*ResNet-18*, *ResNet-50* and *SE-ResNeXt-101 (32 × 4d)*) used in training were pre-trained on the *ImageNet* dataset. Exemplary images after augmentation are shown in Figure 2.

To eliminate irrelevant features from a dataset associated with the image white background, an additional mask was applied. The mask takes the form of two black vertical strips, each one-third of the image's width, as seen in Figure 3. It is worth noting that the mask is applied before *AutoAugment* augmentations, so the mask parts are treated as input and might change depending on transformation. The mask was applied during training, validation, and testing operations. The dataset was divided into training, validation, and test sets in a 70 : 10 : 20 ratio.

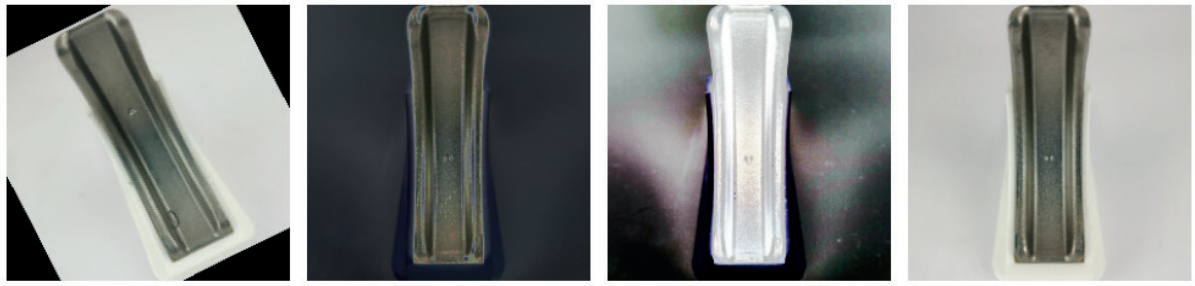

Fig. 2. Exemplary images augmented with *AutoAugment*

Fig. 3. Example images with the applied mask and *AutoAugment*

3. Learning process

Model training was performed in a transfer learning regime after building, preparing, and pre-processing the dataset. This means that CNNs (*ResNet-18*, *ResNet-50*, and *SE-ResNeXt-101 (32 × 4d)*) were used as pre-trained, non-trainable feature extractors. *ResNet-18* (He et al., 2016), a lightweight model with 18 layers, is the most shallow model from the ResNet family. *ResNet-50* (He et al., 2016), a deeper model with 50 layers, provides improved feature extraction capabilities due to its increased depth. *SE-ResNeXt-101 (32 × 4d)* combines the advantages of squeeze-and-excitation blocks (Hu et al., 2018) with the ResNeXt architecture (Xie et al., 2017), offering a robust and highly accurate model with 101 layers and 32 groups of convolutions with a width of 4 channels each.

A single fully connected trainable layer with two outputs was added to the last layer of each extractor. For *ResNet-18*, the fully-connected layer received 512 inputs and produced 2 outputs. For *ResNet-50* and *SE-ResNeXt-101 (32 × 4d)*, the fully-connected layer received 2048 inputs and produced 2 outputs. The outputs from the fully-connected layer were passed through a softmax activation function to provide the final class probabilities.

Various hyperparameters were tested and finally the best results were provided for 50 epochs, a learning rate of 0.002, and a mini-batch size of 32. The optimisation algorithm employed was ADAM (adaptive moment estimation) to accelerate the learning process and leverage mini-batches. The utilised loss function was the cross-entropy loss function, which is suitable for classification tasks and measures the performance of the model by comparing the predicted probabilities to the actual class labels.

As a result, *ResNet-18* and *ResNet-50* (Fig. 4a and 4b, respectively) hover around the 0.1 threshold, reaching it more quickly. In contrast, *SE-ResNeXt-101 (32 × 4d)* (Fig. 4c) exceeds this threshold, achieving the lowest error in the entire analysis.

As can be seen, the validation error quickly (after a few epochs) reaches very low values and maintains them while the training error decreases slowly. In this case, the training error plot is much more important than the validation error plot, as it indicates the model's generalisation ability. The continuous maintenance of a validation error close to zero confirms the absence of overfitting.

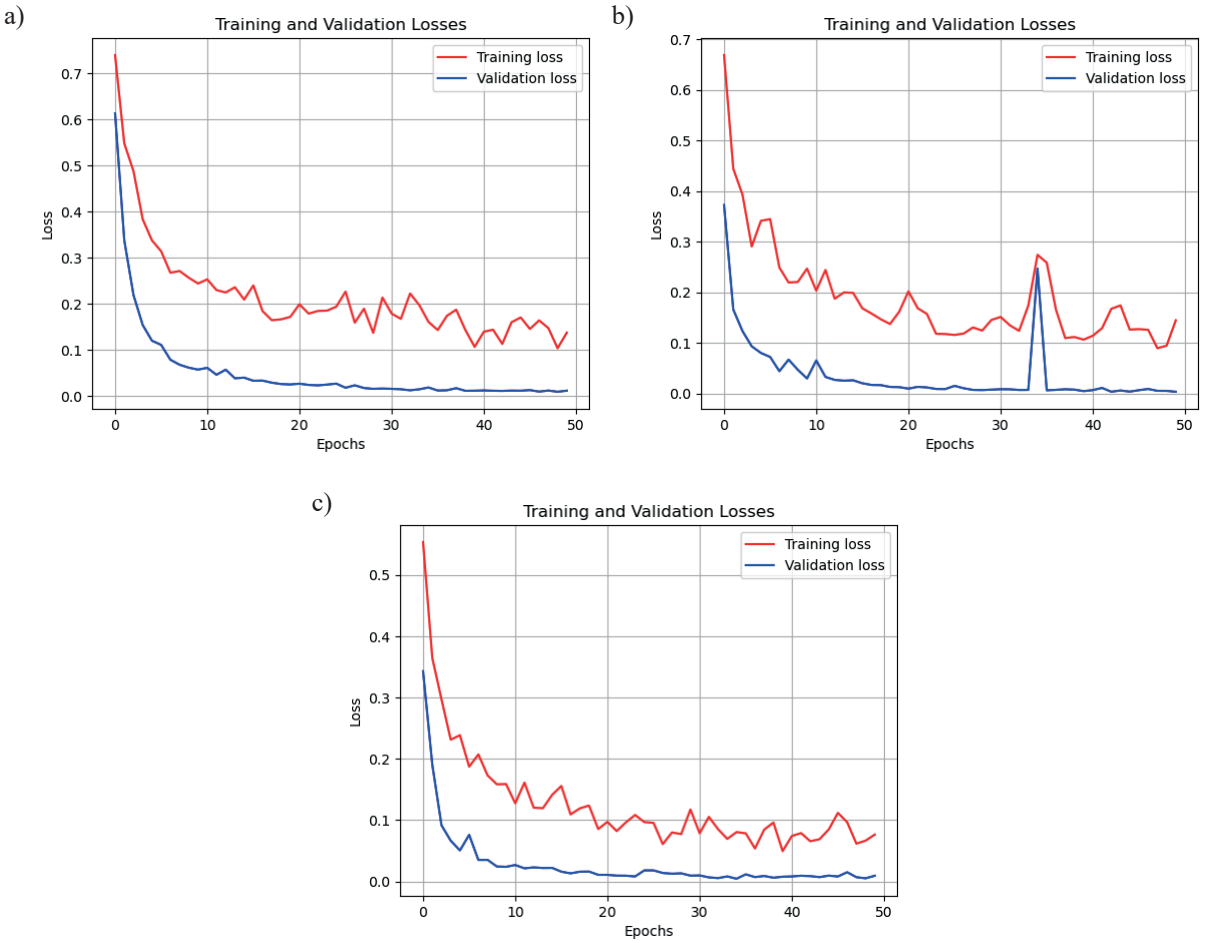

Fig. 4. Loss function values by epoch number: a) *ResNet-18*; b) *ResNet-50*; c) *SE-ResNeXt 101 (32 × 4d)*

4. Model evaluation

To confirm the model generalisation ability, data evaluation was conducted exclusively on the images available in the test set. Metrics considered include accuracy and recall for defective elements. Recall for defective elements is particularly important from the perspective of analysis, as the potential acceptance of a defective product during the quality control process in the industry is significantly more detrimental than rejecting a product without defects.

Additionally, an evaluation was conducted on the test set with applied augmentation to assess the potential generalisation capabilities of the developed models. The reason for this was an exceptionally high performance of all three models on the test set without augmentations – perfect classifier performance was achieved, as shown in Table 2. The advantage of applying the *AutoAugment* augmentation strategy over the testing dataset was that the model was given unknown data additionally impeded by advanced augmentation. Thus, an additional generalisation abilities check might have been performed given the scarce dataset. Moreover, to provide robust results, the investigation was conducted within 50 full iterations over a test set since the implementation of *AutoAugment* augmentation strategy assumes that for each image in a batch, a particular subpolicy (with its probability and magnitude) is chosen randomly. The results of both evaluation approaches (with and without *AutoAugment*) are presented in Table 2, with averaged results obtained from the former approach.

As can be seen above, the apparently perfect classifiers perform differently when tested on an augmented test set. With the *AutoAugment* strategy applied, a slight reduction in accuracy was observed; similarly, recall scores also show a slight decrease. However, this decrease tends to be more pronounced as model complexity increases, which is in contrast to the trend observed within accuracy scores.

Additionally, the results obtained from the augmented evaluation are presented visually in Figure 5 in the form of confusion matrices to display the evaluation details more precisely.

Table 2. Accuracy and recall of evaluated models depending on the test set preparation approach

Accuracy			
approach	*ResNet-18*	*ResNet-50*	*SE-ResNeXt-101*
Without *AutoAugment*	1.0	1.0	1.0
With *AutoAugment* (avg. of 50)	0.9668	0.9778	0.9860
Recall			
approach	*ResNet-18*	*ResNet-50*	*SE-ResNeXt-101*
Without *AutoAugment*	1.0	1.0	1.0
With *AutoAugment* (avg. of 50)	0.9904	0.9860	0.9832

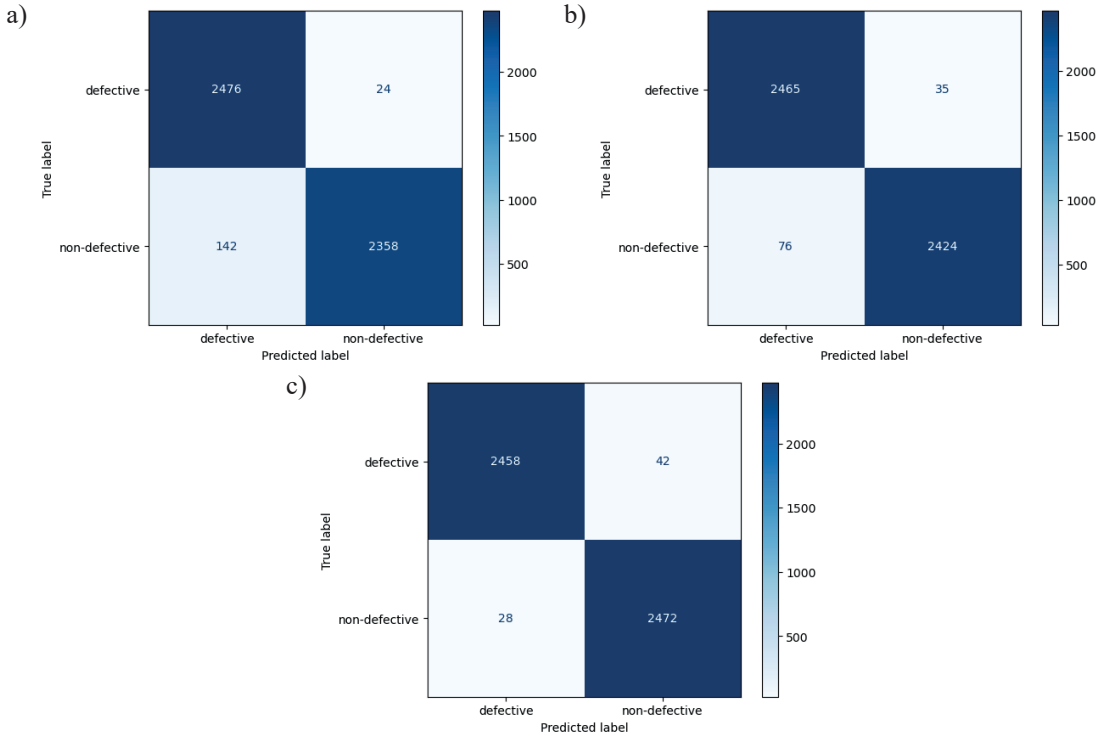

Fig. 5. Confusion matrices obtained in augmented evaluation for: a) *ResNet-18*; b) *ResNet-50*; c) *SE-ResNeXt-101 (32 × 4d)*

The numbers presented in Figure 5 represent predictions summed over 50 full iterations on a test set containing 100 images, resulting in a total of 5000 predictions and ensuring trends and observations described above, providing the broader context of the evaluation process than metrics alone.

5. Conclusion

In this research, the training and evaluation of three deep learning models were performed based on a scarce dataset of images of forged components. Additional augmentations over the dataset were conducted and an additional evaluation was performed, leveraging the *AutoAugment*-generated augmentations to confirm the robustness of the model.

It has to be emphasised that the selection of the best model usually depends on the business and industrial requirements and applications. ResNet-50 can provide the balance between accuracy and recall; however, if maximising one of them is crucial, then choosing one of the other models would be a solution.

The application of the *AutoAugment* augmentation strategy turns out to not only be useful in the learning process when dealing with scarce hand-crafted datasets but also as a potential generalisation ability indicator when the original test set appears insufficient. However, *AutoAugment* indications must be verified in real-world tests that could ultimately prove right aggressive augmentation strategy. Such proof would result in obtaining a new way of ensuring model robustness before testing under production conditions.

Acknowledgement

The work was undertaken within the framework of fundamental statutory research and Excellence Initiative – Research University.

References

Bhatt, P. M., Malhan, R. K., Rajendran, P., Shah, B. C., Thakar, S., Yoon, Y. J., & Gupta, S. K. (2021). Image-based surface defect detection using deep learning: A review. *Journal of Computing and Information Science in Engineering*, *21*(4), 040801. https://doi.org/10.1115/1.4049535.

Chai, J., Zeng, H., Li, A., & Ngai, E. W. T. (2021). Deep learning in computer vision: A critical review of emerging techniques and application scenarios. *Machine Learning with Applications*, *6*, 100134. https://doi.org/10.1016/j.mlwa.2021.100134.

Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V., & Le, Q. V. (2019). AutoAugment: Learning augmentation strategies from data. In 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE. https://doi.org/10.1109/CVPR.2019.00020.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. https://doi.org/10.1109/CVPR.2009.5206848.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). *An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale*. arXiv. https://doi.org/10.48550/arXiv.2010.11929.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE. https://doi.org/10.1109/CVPR.2016.90.

Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE. https://doi.org/10.1109/CVPR.2018.00745.

Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE. https://doi.org/10.1109/CVPR.2017.243.

Jia, Z., Wang, M., & Zhao, S. (2024). A review of deep learning-based approaches for defect detection in smart manufacturing. *Journal of Optics*, *53*(2), 1345–1351. https://doi.org/10.1007/s12596-023-01340-5.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539.

Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A ConvNet for the 2020s. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE. https://doi.org/10.1109/CVPR52688.2022.01167.

Lugin, S., Müller, D., Finckbohner, M., & Netzelmann, U. (2023). Automated surface defect detection in forged parts by inductively excited thermography and magnetic particle inspection. *Quantitative InfraRed Thermography Journal*, 1–13. https://doi.org/10.1080/17686733.2023.2266901.

Niccolai, A., Caputo, D., Chieco, L., Grimaccia, F., & Mussetta, M. (2021). Machine learning-based detection technique for NDT in industrial manufacturing. *Mathematics*, *9*(11), 1251. https://doi.org/10.3390/math9111251.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536. https://doi.org/10.1038/323533a0.

Shin, H.-Ch., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, Y., 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. https://doi.org/10.1109/tmi.2016.2528162.

Tabernik, D., Šela, S., Skvarč, J., & Skočaj, D. (2020). Segmentation-based deep-learning approach for surface-defect detection. *Journal of Intelligent Manufacturing*, *31*(3), 759–776. https://doi.org/10.1007/s10845-019-01476-x.

Tan, Ch., Sun, F., Kong, T., Zhang, W., Yang, Ch., & Liu, Ch. (2018). A survey on deep transfer learning. In V. Kůrková, Y. Manolopoulos, B. Hammer, L. Iliadis, I. Maglogiannis (Eds.), *Artificial Neural Networks and Machine Learning – ICANN 2018. 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4–7, 2018, Proceedings* (pt. 3, pp. 270–279). Springer Cham. https://doi.org/10.1007/978-3-030-01424-7_27.

Tan, M., & Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning* (vol. 97, pp. 6105–6114). Retrieved from http://proceedings.mlr.press/v97/tan19a.html.

Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE. https://doi.org/10.1109/CVPR.2017.634.

Yang, J., Li, S., Wang, Z., Dong, H., Wang, J., & Tang, S. (2020). Using deep learning to detect defects in manufacturing: A comprehensive survey and current challenges. *Materials*, *13*(24), 5755. https://doi.org/10.3390/ma13245755.