Rethinking Pre-Training in Industrial Quality Control

Dilara Yesilbas
Friedrich Alexander University Erlangen-Nuremberg, Germany, dilara.yesilbas@fau.de

Stefan Arnold
Friedrich Alexander University Erlangen-Nuremberg, Germany, stefan.st.arnold@fau.de

Alex Felker
Friedrich Alexander University Erlangen-Nuremberg, Germany, alex.felker@fau.de

Rethinking Pre-Training in Industrial Quality Control

Dilara Yesilbas, Stefan Arnold, and Alex Felker

University of Erlangen-Nuremberg, Institute of Information Systems, Nuremberg, Germany
{dilara.yesilbas, stefan.st.arnold, alex.felker}@fau.de

Abstract. The application of machine learning is of high significance for quality control tasks in the manufacturing industry due to large volumes of machine-generated data. However, labeling data is costly and labor-intensive. In this study, we evaluate the role of manual labeling and the moderating effect of autoencoder-based pre-training in optical quality control using real-world industrial data. We observe that pre-training substantially elevates the classification accuracy for small amounts of labeled data. With increasing amounts of labeled data available during fine-tuning, however, we find diminishing returns, analogous to recent concerns raised in non-industrial applications.

Keywords: Autoencoder, Optical Quality Control, Printed Circuit Boards

1 Introduction

Quality Control (QC) is a critical task in the manufacturing industry. Current systems often still require manual inspection by experts. However, repetitive tasks can be tedious, making human-based QC error-prone. Hence, there is an increasing demand to apply machine learning in QC. As for optical inspection, *Convolutional Neural Networks* (CNNs) are suitable to ensure high quality standards at competitive costs. CNNs are employed in a broad array of applications. Building on the *Transfer Learning* paradigm, publicly available models pre-trained on generic data are typically used and labeled with task-specific data. However, data modalities for specific tasks are often vastly different in industry, fueling the debate on the added value of generic pre-training. Moreover, the high cost of manually labeling data is a major challenge for companies. As a remedy, Masci *et al.* [1] presented an unsupervised alternative in the form of stacked *Convolutional Autoencoders* (CAEs). Being unsupervised, CAEs reduce the amount of labeled data required, making them attractive to companies with large machine-generated data. In this paper, we contribute to the recent debate on pre-training and fine-tuning [2–5] and the cost of manual annotation from the perspective of optical quality control. Using controlled experiments on real-world data, we ablate the role of unsupervised pre-training by comparing its fine-tuned classification accuracy with training from a random initialization. By varying the amount of labeled data by multiple orders of magnitude, we find evidence that manual labeling renders the need for unsupervised pre-training for our task, as recently reported by [4, 5] for the case of supervised pre-training on *ImageNet*.

17th International Conference on Wirtschaftsinformatik
February 2022, Nürnberg, Germany

2 Background

In a series of ablation studies, Erhan *et al.* [2] investigated the role of pre-training and fine-tuning of visual feature hierarchies. More recently, Hendrycks *et al.* [3] showed that pre-training adds robustness to uncertainty estimates for out-of-distribution classifications. In contrast, He *et al.* [4] and Zoph *et al.* [5], experimentally demonstrated that a model trained from-scratch can match or exceed the performance of a model pre-trained with supervision given sufficient training time and labeled data. Instead of training a CNN from-scratch, an initial data representation can first be learned in an unsupervised manner. Most unsupervised methods for pre-training are based on the encoder-decoder paradigm. This family of methods centers on reconstruction, where the input is first transformed into a lower-dimensional space, and then expanded to reconstruct the input data. Zeiler *et al.* [6] demonstrated that deep hierarchies can be formed by stacking multiple autoencoders. This allows unsupervised pre-training to be performed layer by layer, as each layer receives its input from the layer below. For visual features, Masci *et al.* [1] presented a CAE stack that builds on multiple convolution layers to transform images into a high-dimensional feature representation, and then reconstructs the image using strided transposed convolutions. We adopt the CAE for pre-training.

3 Experiments

We base our methodology on Masci *et al.* [1] but use a residual network as baseline for our CNN and CAE. [1] A similar network architecture is used by [7] in a concurrent research effort. By leveraging skip connections, residual networks provide a good trade-off between performance and number of parameters. In our experiment, we evaluate the unsupervised pre-training and fine-tuning paradigm against from-scratch training for optical quality control. Therefore, we have compiled a balanced real-world dataset of defective and defect-free printed circuit board (PCB) images. For the purpose of optical quality control, the original PCB images are divided into square sections that form our dataset. In our experiment, we divided our dataset into two classes. The first class comprises only defect-free images. In the second class of defective images, we combined five common PCB defect categories into one class. In total, our binary dataset consists of $100,000$ unlabeled auxiliary examples for unsupervised pre-training and further $100,000$ labeled examples for supervised fine-tuning. Both data partitions come from the same production line. We converted the resolution of the grayscale images to shape $3 \times 224 \times 224$. Apart from resizing, we augmented the training examples using flipping and rotation. By splitting the data into training set and a hold-out validation set, we tracked the training progress and adjusted hyperparameters independently for pre-training, fine-tuning, and from-scratch training. Figure 1 depicts our training process.

Pre-Training. To obtain a CAE for pre-training, we detach the fully-connected layer of our residual network and mirror strided transposed convolutional layers analogous to the number of convolution layers. With the objective of extracting the CAE encoder to initialize the convolutional layers of the CNN, we train the CAE stack on large-scale

[1] We identified ResNet-50 on the basis of a preliminary analysis on state-of-the-art networks.

unlabeled PCB images without supervision. Note that the feature map of the CAE stack is high-dimensional. This contrasts with dimensionality reduction, which is typically the main application of CAEs. Without further constraints, a high-dimensional representation enables the CAE to learn the identity function which copies the input onto a feature map. While the identity function simplifies a perfect reconstruction of the input, it prevents the CAE from finding a generalized data representation. To serve as a suitable initialization, a form of regularization must be applied. For regularization of CAEs, sparsity constraints on the representation layer or noisy reconstruction objectives can be utilized. Pooling layers were added in earlier work as an elegant way to enforce sparsity without the need for further regularization. Since residual networks use skip connections instead of pooling layers to add hierarchy, we apply a denoising CAE. At pixel-level, we add random Gaussian noise to each image uniformly, which prevents the CAE from learning a non-trivial latent PCB representation. Using mean squared error as objective function, we pre-train the CAE until convergence using stochastic gradient descent. As we are only interested in the latent representation of the CAE encoder, the CAE decoder and the CAE-generated PCB images can be deleted.

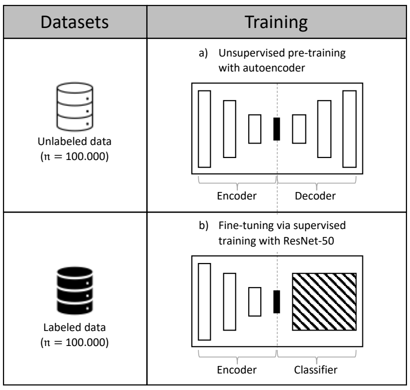

Figure 1. Two-step training process in line with pre-training and fine-tuning paradigm.

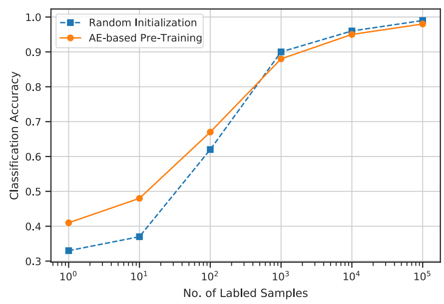

Figure 2. Results: Comparison of classification accuracy for real-world PCB images as a function of the amount of labeled data for autoencoder-based initialization vs. random initialization.

Fine-Tuning. To obtain a CNN for fine-tuning, we add a fully-connected classification layer on top of the CAE encoder. While the convolutional layers are preset with the parameters obtained from unsupervised pre-training, the fully-connected layer is initialized at random. The CNN is then fine-tuned for binary classification of defective and defect-free images of PCBs. To evaluate the role of unsupervised pre-training compared to labor-intensive manual annotation, we sub-sample $\{10^0, 10^1, \ldots, 10^5\}$ labeled PCB images and compare the classification accuracy after fine-tuning.

In Figure 2, we depict the classification accuracy after fine-tuning as a function of the amount of labeled data. By gradually increasing the amount of labeled data up to six orders of magnitude, we measure the value added by labor-intensive manual annotation under moderation of unsupervised pre-training. For comparison, we initialize a residual network of the same topology with random initialization. To ensure reproducibility, its training mimics the fine-tuning of the pre-trained CNN under a static random seed. With increasing amounts of labeled data, we observe diminishing returns of unsupervised pre-training for our case, which is consistent with the findings in [4,5] for supervised pre-training. For low data regimes ($< 10^2$), we find CAE-based pre-training to yield better accuracy, which is, however, trivial and does not meet manufacturing standards. For high data regimes ($\gg 10^3$), from-scratch training consistently outperforms fine-tuning of CAE-based pre-training while achieving operationally reliable accuracies.

4 Discussion

Leveraging stacked autoencoders for unsupervised pre-training has been proposed for industrial applications where manual labeling is tedious but large amounts of unlabeled data exist [7–9]. By conducting controlled experiments on real-world industrial data that exhibit the argued characteristics, we show that the prohibitive costs of manual annotations *cannot* be offset by additional training effort spent on CAE-based pre-training. Although the performance gain of CAE-based pre-training is not marginal for low data regimes, the defect detection accuracy is insufficient for the stringent industrial manufacturing standards. Moreover, the returns of unsupervised pre-training diminish as the data size increases. We therefore recommend economizing on unsupervised CAE-based pre-training in favor of manual labeling and training from-scratch for use cases that require high-quality models. Considering our research roadmap, we plan to extend our study to a multi-class classification where we evaluate the robustness and uncertainty differentiated by defect categories. We hypothesize that CAE-based pre-training may be more beneficial for new production lines with no prior operational experience and rarely labeled data.

References

1. Masci, J., Meier, U., Cireşan, D., Schmidhuber, J.: Stacked convolutional auto-encoders for hierarchical feature extraction. In: International conference on artificial neural networks. pp. 52–59. Springer (2011)
2. Erhan, D., Courville, A., Bengio, Y., Vincent, P.: Why does unsupervised pre-training help deep learning? In: Proceedings of the thirteenth international conference on artificial intelligence and statistics. pp. 201–208. JMLR Workshop and Conference Proceedings (2010)
3. Hendrycks, D., Lee, K., Mazeika, M.: Using pre-training can improve model robustness and uncertainty. In: International Conference on Machine Learning. pp. 2712–2721. PMLR (2019)
4. He, K., Girshick, R., Dollár, P.: Rethinking imagenet pre-training. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4918–4927 (2019)
5. Zoph, B., Ghiasi, G., Lin, T.Y., Cui, Y., Liu, H., Cubuk, E.D., Le, Q.V.: Rethinking pre-training and self-training. arXiv preprint arXiv:2006.06882 (2020)
6. Zeiler, M.D., Krishnan, D., Taylor, G.W., Fergus, R.: Deconvolutional networks. In: 2010 IEEE Computer Society Conference on computer vision and pattern recognition. pp. 2528–2535. IEEE (2010)
7. Xu, H., Huang, Z.: Annotation-free defect detection for glasses based on convolutional auto-encoder with skip connections. Materials Letters 299, 130065 (2021)

8. Bergmann, P., Löwe, S., Fauser, M., Sattlegger, D., Steger, C.: Improving unsupervised defect segmentation by applying structural similarity to autoencoders. arXiv preprint arXiv:1807.02011 (2018)
9. Hu, G., Huang, J., Wang, Q., Li, J., Xu, Z., Huang, X.: Unsupervised fabric defect detection based on a deep convolutional generative adversarial network. Textile Research Journal 90(3-4), 247–270 (2020)