Estimating the Probability Density Function of New Fabrics for Fabric Anomaly Detection

Oliver Rippel[1] [a], Maximilian Müller[1], Andreas Münkel[2], Thomas Gries[2] and Dorit Merhof[1] [b]
[1]*Institute of Imaging & Computer Vision, RWTH Aachen University, Aachen, Germany*
[2]*Institut für Textiltechnik, RWTH Aachen University, Aachen, Germany*

Keywords: Anomaly Detection, Quality Control, Fabric Inspection, Transfer Learning, Probability Density Estimation.

Abstract: Image-based quality control aims at detecting anomalies (i.e. defects) in products. Supervised, data driven approaches have greatly improved Anomaly Detection (AD) performance, but suffer from a major drawback: they require large amounts of annotated training data, limiting their economic viability.
In this work, we challenge and overcome this limitation for complex patterned fabrics. Investigating the structure of deep feature representations learned on a large-scale fabric dataset, we find that fabrics form clusters according to their fabric type, whereas anomalies form a cluster on their own. We leverage this clustering behavior to estimate the Probability Density Function (PDF) of new, previously unseen fabrics, in the deep feature representations directly. Using this approach, we outperform supervised and semi-supervised AD approaches trained on new fabrics, requiring only defect-free data for PDF-estimation.

1 INTRODUCTION

The textile industry is one of the biggest industries in the world, producing several million tons of fabric every year. With ever-increasing technological progress, fabric production has become a highly optimized process, leading to low error rates.

Despite their rare occurrence, fabric anomalies still have a strong economic impact, making their detection an essential aspect of fabric production. However, Anomaly Detection (AD) in fabrics is still largely performed by human operators, and the outcome depends on training, skill level and fatigue of the personnel. Even at peak performance, human operators are only capable of detecting 60-80% of defects (Karayiannis et al., 1999; See, 2012), while accounting for at least 10% of total labor costs (Newman and Jain, 1995). Together, this calls for machine vision solutions that are capable of automated defect detection.

With recent advances in Machine Learning, learning-based approaches have seen a strong increase in performance, becoming ever more relevant for automated defect detection. Based on the required degree of supervision, learning-based approaches can be categorized into supervised, semi-supervised and unsupervised algorithms. In the context of AD, these categories are defined as follows (Chandola et al., 2009; Ruff et al., 2020a):

- **supervised:** providing a fully labeled dataset containing both anomalies as well as normal data.
- **semi-supervised:** providing a dataset that contains normal data only.[1]
- **unsupervised:** providing an unlabeled dataset, i.e. a dataset that consists mostly of normal data but may also contain anomalies.

As fabric defects are rare events and expensive to sample, semi-supervised algorithms are most commonly employed in literature. These algorithms have been shown to work for fabrics of low complexity (i.e. unimodal appearance), but show limited performance in fabrics of high complexity (i.e. multimodal appearance) (Mei et al., 2018; Hu et al., 2019).

Supervised approaches have also been successfully applied to fabric defect detection, adapting classification and object detection approaches such as ResNet and YOLO (Zhang et al., 2018; Gao et al.,

[a] https://orcid.org/0000-0002-4556-5094
[b] https://orcid.org/0000-0002-1672-2185

[1]Note that work exists on *general* semi-supervised algorithms that can also make use of partially labeled datasets, but this is not considered further in our work. For details, we refer to (Ruff et al., 2020a).

Rippel, O., Müller, M., Münkel, A., Gries, T. and Merhof, D.
Estimating the Probability Density Function of New Fabrics for Fabric Anomaly Detection.
DOI: 10.5220/0010163604630470
In *Proceedings of the 10th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2021)*, pages 463-470
ISBN: 978-989-758-486-2

2019). However, while progress has been made with respect to detection performance, none of the supervised approaches suit the need of the textile industry for low changeover times. The reason for this is that defects have to be collected and annotated to yield a sufficiently large image basis for every individual fabric, which is a tedious, time-consuming and expensive process.

While algorithms have been proposed to tackle this limitation, research focusses on synthesizing new defective images based on prior knowledge about defect appearances. This knowledge may either be learned implicitly by Generative Adversarial Networks (GANs) (Liu et al., 2019; Rippel et al., 2020b), or explicitly inferred from experts (Han and Yu, 2020).

In this work, we propose an alternative approach: we hypothesize that the deep feature representations learned by a supervised model on a large-scale fabric dataset are discriminative also to new fabric types unseen during training. We analyze the structure of learned deep feature representations using t-distributed Stochastic Neighbor Embedding (t-SNE), and find that fabrics form clusters according to their fabric type, with anomalies forming an additional cluster on their own. Based on this finding, we construct an AD model for new fabrics unseen during training by approximating their Probability Density Function (PDF) in the deep representations, achieving state-of-the-art performance.

1.1 Related Work

In previous work, it has been shown that it is possible to train supervised (Wu et al., 2020) and semi-supervised (Han and Yu, 2020) fabric defect detection methods that can generalize well within a diverse fabric defect dataset. However, Liu et al. (Liu et al., 2019) show poor out-of-the-box performance for supervised fabric defect segmentation applied to new fabrics, and demonstrate that detection performance can be increased by using synthetic defects generated by GANs in addition to normal fabric images for model fine-tuning. Additionally, Rippel et al. (Rippel et al., 2020b) show that supervised defect detection models can also be trained from scratch on new fabrics, again employing defects generated by GANs in combination with normal fabric images as the training dataset. While defect synthesis by means of GANs is also popular for improving performance at general surface inspection tasks (Le et al., 2020), GANs are known to be notoriously difficult to train (Miyato et al., 2018), diminishing the practical applicability of developed approaches.

Weninger et al. (Weninger et al., 2018) demonstrate that fabric defect detection is possible on fabrics unseen during training without relying on defects synthesized by GANs. However, their approach necessitates high-resolution images for float-point detection, increasing computational burden for an eventual machine vision solution. Furthermore, their work utilizes plain-weave fabrics with simple patterns only.

While not directly applied to the AD task, Lee et al. (Lee et al., 2018) show that out-of-distribution (OOD) detection can be achieved by modeling the PDF of input images in learned deep feature representations. This is achieved by linking generative models using Gaussian Discriminant Analysis on deep features to discriminative models trained by the softmax-crossentropy loss. The linkage between deep generative and discriminate models has been applied by Rippel et al. to the industrial AD use case (Rippel et al., 2020a).

Together, this motivated us to construct a transferable fabric anomaly detector by modeling the PDF of new fabrics in deep feature representations learned by training on a large-scale fabric dataset, which is presented in more detail in the following.

2 MODELING THE PDF FOR CROSS-FABRIC ANOMALY DETECTION

We aim to construct a transferable anomaly detector for fabrics by modeling the PDF of new fabrics in deep feature representations. While features learned by Image-Net training have been successfully applied to the industrial AD task in a transfer learning setting (Andrews et al., 2016; Rippel et al., 2020a), the 4-channel dimensionality of our data (cf. Section 3) prevents a straight-forward use of Image-Net features. Therefore, we instead learn domain-specific features from scratch using subsets of our collected dataset by training a supervised, deep anomaly detector. We then extract the deep features before the final mapping to the anomaly score to model the PDF of new fabrics unseen during training, as deeper features have shown increased performance also in the transfer learning AD setting (Andrews et al., 2016; Rippel et al., 2020a).

Rippel et al. (Rippel et al., 2020a) have shown that the individual dimensions of deep feature representations learned by discriminative models are highly correlated. Therefore, the model used to estimate the PDF of new fabrics should be multivariate.

2.1 Modeling Unimodal PDFs

For unimodal data, the PDF can be modeled by using a multivariate Gaussian (Bishop, 2006). A useful anomaly score here is the Mahalanobis distance (Mahalanobis, 1936), which uniquely determines the value of an observation's PDF under the Gaussian.

We estimate the mean of the multivariate Gaussian using Maximum Likelihood (ML) estimation, which corresponds to the empirical mean. For the covariance matrix, we apply shrinkage as proposed by Ledoit et al. (Ledoit et al., 2004). Regularization by means of shrinkage is necessary since the number of observations used for fitting is in the same order of magnitude as the dimensionality of the fitted Gaussian (refer Table 1).

2.2 Modeling Multimodal PDFs

For multimodal data, the PDF can be approximated by fitting a Gaussian Mixture Model (GMM), i.e. a linearly weighted sum of individual Gaussians

$$p(\boldsymbol{x}) = \sum_{i=1}^{K} \phi_i \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \quad (1)$$

with $\sum_{i=1}^{K} \phi_i = 1$. We approximate the parameters of a GMM by using the Expectation Maximization (EM) algorithm (Bishop, 2006).

Compared to the unimodal setting, modeling multimodal PDFs by means of GMMs introduces an additional hyperparameter, the number of Gaussian mixture components K. We estimate K by using the Bayesian Information Criterion (BIC) proposed by Schwarz et al. (Schwarz et al., 1978). While other metrics such as the Akaike information criterion (AIC) exist, we choose the BIC score for its strong regularization characteristics.

For the multimodal setting, a sensible anomaly score is the negative log-likelihood of x defined as

$$NLL = -log(p(\boldsymbol{x})). \quad (2)$$

We also propose $min(M_i(x))$, i.e. the minimum Mahalanobis distance for all Gaussian Mixture components, as a possible anomaly score to account for large differences in ϕ.

2.3 Learning Deep Feature Representations

In order to learn the deep representations required by our approach, we train a ResNet18 (He et al., 2016) from scratch in a Leave-One-Out (LOO) manner for each fabric present in the dataset, where all fabrics except the one evaluated on are used for training (cf. Figure 2). We employ the sigmoid-crossentropy loss together with the Adam optimizer (Kingma and Ba, 2015), an initial learning rate of 0.001 and a batchsize of 16, training for 15k iterations in total.

To improve robustness of our evaluation w.r.t. the initially available fabric dataset, we performing an additional 5-fold evaluation on each respective dataset used for feature learning stratified for anomaly prevalence, reporting averaged results for our approach.

2.4 Modeling PDF on Held Out Fabric

After having learned the deep feature representations, we apply our PDF modeling strategies to each held out fabric. We make use of two different datasplits to enable fair comparison with both supervised and semi-supervised reference methods that serve as a benchmark. First, we estimate the PDF of the held out fabric using the training set of a **supervised split**, where anomalous data is removed from the training set (see Figure 2). We refer to this setting as "clean", and can use it to compare against supervised baselines as our method assumes that only "normal" data is used for estimating the PDF. Second, we estimate the PDF of the held out fabric using the training set of a **semi-supervised split**, where anomalous data is only present in the test set (cf. Section 1 and Figure 2).

In addition to these two splits, we also estimate the PDF of the held out fabric using the training set of a supervised split where anomalous data remains in the training set (see Figure 2). This setting corresponds to applying our strategy in an **unsupervised** manner, i.e. where unlabeled anomalies are present during PDF estimation.

For each splitting variant, we first model the PDF of the new fabric using the training set and then apply the constructed AD model to the respective test set that is identical to the one used by our baselines. For the GMM setting, models were fit for $K \in \{2,...,19\}$ mixture components, and the best model was selected based on lowest BIC on the training set.

Note that model weights are fixed and the data of held out fabrics used only to parametrize the PDF models. Also, validation sets of held out fabrics are unused by our approach, giving an additional advantage to the reference methods and yielding strong baselines.

To investigate the benefit of our proposed PDF modeling, we also apply the discriminative decision boundary learned on the deep feature representations during training of our deep feature extracting model (henceforth refered to as LOO model) to the held out fabrics.

3 DATASET

The fabric dataset used in this work comprises a total of 20 patterned fabrics. For each fabric, paired front-light RGB and backlight luminance were acquired at 2000 DPI resolution, resulting in a 4 channel image. A defective as well as a defect-free sample image can be found in Figure 1.

In total, the dataset contains 4270 samples across all fabrics, of which 320 are labeled as defective (see Table 1).

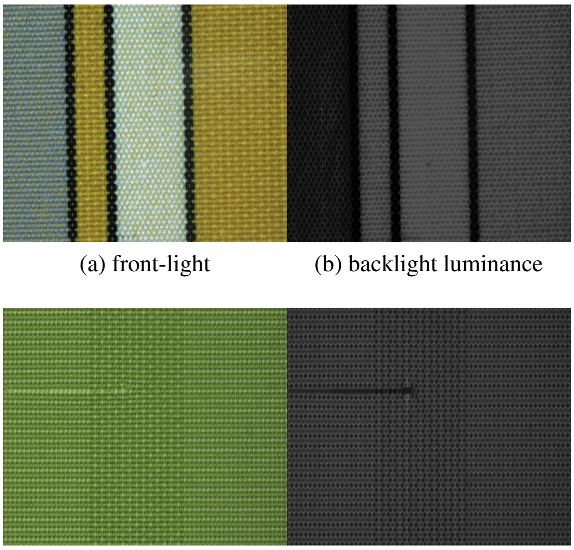
Table 1: Characteristics of the used dataset.

Fabric	images	
	normal	defective
1	470	14
2	242	5
3	148	16
4	229	19
5	227	9
6	530	16
7	388	19
8	159	6
9	118	26
10	78	6
11	35	5
12	112	35
13	201	13
14	64	7
15	305	20
16	45	7
17	389	45
18	55	16
19	42	17
20	113	19
total	3950	320

4 EXPERIMENTS AND RESULTS

In our work, we hypothesize that deep representations learned by a supervised model on a large-scale fabric dataset are discriminative also to new fabric types unseen during training, and propose to achieve this by modeling the PDF of new fabrics in learned representations directly.

To test our hypothesis, we evaluate our approach on every single fabric of the dataset individually, aggregating single fabric performances to generate robust insights. Similarly, we also train state-of-the art supervised and semi-supervised AD algorithms on each fabric to serve as comparison.



(a) front-light (b) backlight luminance

(c) front-light (d) backlight luminance

Figure 1: Representative defect-free (a-b, fabric 8) and defective (c-d, fabric 2) sample images.

Evaluation Details. As AD poses a binary decision problem, we report the Area Under the Receiver Operator Characteristic (ROC) curves as well as the Area Under the Precision Recall (AUPR) curve to evaluate model performance. Note that the AUPR is better suited to report results for skewed/imbalanced datasets such as ours. Further, to improve robustness of evaluations, a 5-fold evaluation is performed for each held out fabric (refer Figur 2 for details).

Supervised Reference Method. As a baseline for supervised AD methods, we train a ResNet18 from scratch on each individual held out fabric on the supervised splits as outlined in Figure 2. Training parameters are identical to those used for feature learning (refer Section 2.3). Model selection was performed based on AUPR achieved on the validation set.

Semi-supervised Reference Method. For the semi-supervised baseline, we train a convolutional autoencoder from scratch on each individual held out fabric, using ResNet18 as encoder and an "inverted" ResNet18 as decoder (i.e. every operation of the encoder should be inverted by the decoder). For the upsampling operations we employ pixel shuffle as introduced by Shi et al. (Shi et al., 2016) to reduce checkerboard artifacts which would be present otherwise. The latent dimension of the bottleneck is set to 32 and yields proper reconstruction of normal images in all fabrics. We train the model using the structured-similarity measure and select

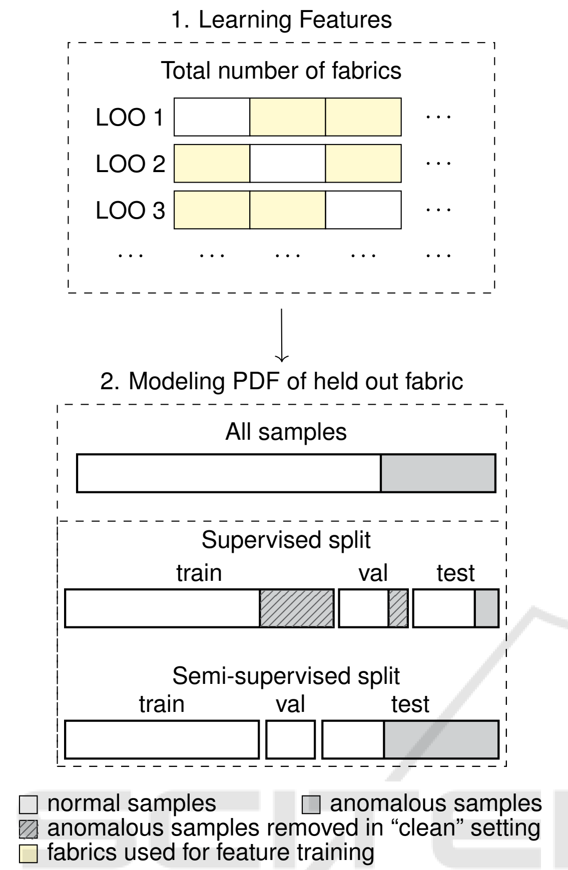


Figure 2: Evaluation pipeline of our approach. In a first step, we learn the deep features required by our approach in a Leave-One-Out (LOO) manner, where a single fabric is held out per run. In the second step, we model the Probability Density Function of the held out fabric in the deep features using three different splits (supervised, supervised – “clean” and semi-supervised).

the model based on the lowest reconstruction loss on the validation set, and training is performed using the Adam optimizer together with an initial learning rate of 0.0005 and a batch-size of 16 for a total of 45k iterations. As autoencoders yield residual images as output, an aggregation is necessary to yield an image-level anomaly score. While the threshold employed for ROC/PR calculation is set on the pixel level, we perform connected component analysis and label a test image as defective only if it contains a connected component at least as big as the smallest anomaly present in the test dataset. Note that by extracting the minimal anomaly size from the test set, knowledge is introduced to the autoencoder approach, increasing complexity of the procedure and giving it an additional advantage.

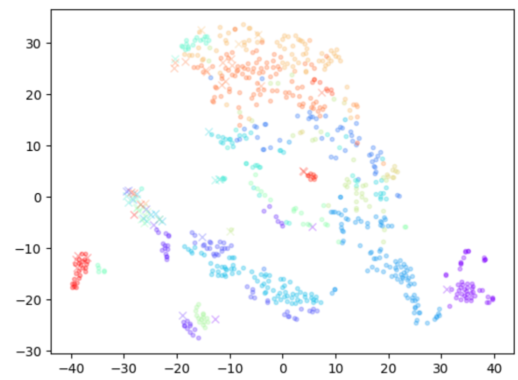


Figure 3: Distribution of deep representations learned by a supervised multi-fabric Anomaly Detection model. Features are extracted from the last layer of a ResNet18 model before the final mapping and visualized by means of t-SNE. Dots denote normal data, whereas crosses denote anomalies. Individual fabrics are color-coded.

Implementation Details. For all approaches, images are resized to a size of 896×896 pixels and training is performed in a patch-wise manner on 224×224 sized patches. We replace Batch Normalization with Instance Normalization (Ulyanov et al., 2016) for all models which was seen to improve performance in every method in preliminary experiments. For supervised methods (i.e. feature learning step of our method and the supervised reference method), patches are cropped around the defect if available (and randomly otherwise), and random oversampling was applied to ensure that 25% of training samples were defective. For semi-supervised methods, patches are cropped randomly. Inference is subsequently performed on whole images, spatially averaging patch-wise generated features and predictions respectively for the supervised methods. For the semi-supervised methods, patch-wise predictions are stitched to form a 896×896-sized reconstructed image as the basis for residual computation. Connected component analysis as described above is applied to the stitched residual image.

Results. Table 2 shows that applying the LOO model to the held out fabrics without any modifications (i.e. applying the decision boundary learned on the large-scale fabric dataset) already performs comparably to the respective reference methods (cf. Table 2). Results also show that additional performance is gained by our proposed PDF modeling of new fabrics in the learned representations. Here, it can be seen that multimodal distributions modeled by GMMs outperform unimodal distributions. Specifically, an AUPR of $86.0 \pm 12.9\%$ (Mean $\pm$ STD) is achieved for GMM modeling and NLL anomaly score

Table 2: Evaluating Anomaly Detection performance. Highest values are boldfaced for Mean, whereas lowest values are boldfaced for STD. Leave-One-Out (LOO) denotes our proposed approach using either unimodal Gaussian (Gaussian) or multimodal Gaussian Mixture Model (GMM) in distance to nearest mixture component (GMM maha) or likelihood mode (GMM NLL). LOO alone denotes applying the initially learned decision boundary to the new fabric. Connected Component AutoEncoder (CCAE) refers to the semi-supervised benchmark, and Classifier to the supervised benchmark. For details regarding data splits we refer to Figure 2. Abbreviations: s = supervised, ss = semi-supervised, us = unsupervised.

split	method	AUPR		AUROC	
		Mean	STD	Mean	STD
	Classifier	68.1	34.6	85.2	22.3
	LOO	65.8	32.0	85.9	16.8
s	LOO Gaussian	69.0	30.6	87.0	15.3
	LOO GMM maha	72.3	21.8	89.8	8.9
	LOO GMM NLL	**73.2**	20.9	**91.2**	**7.4**
	CCAE	78.1	22.7	87.3	14.9
	LOO	80.1	18.1	86.1	11.4
ss	LOO Gaussian	82.7	15.3	87.0	11.1
	LOO GMM maha	84.7	14.3	89.8	8.9
	LOO GMM NLL	**86.0**	**12.9**	**91.4**	**7.3**
	LOO Gaussian	**54.9**	31.0	**82.8**	16.2
us	LOO GMM maha	27.0	13.0	53.6	**7.4**
	LOO GMM NLL	29.6	**12.0**	58.7	11.1

compared to 82.7±15.3% for the unimodal Gaussian in the semi-supervised setting (cf. Table 2). This indicates that the PDF of individual fabrics in the learned representations is indeed multimodal, which is further supported by the clustering tendencies observed in latent space visualization (cf. Figure 3), where two distinct clusters can be observed for the fabric colored in red. Note that both unimodal and multimodal modeling of PDFs outperform the respective state of the art, achieving both higher average AUROC/AUPR values as well as lower standard deviations (cf. Table 2).

While estimating the PDF by means of GMMs increases AD performance, the method also becomes more sensitive to unlabeled anomalies in the training data and fails in the unsupervised setting (cf. Table 2). Presence of unlabeled anomalies in the dataset used for PDF estimation also affects the unimodal Gaussian negatively, albeit not as strongly as the GMM variants. Regarding the choice of an appropriate anomaly score for multimodal PDFs, NLL outperforms minimum Mahalanobis distance consistently by a small margin (cf. Table 2).

When investigating single fabric performance of the best performing configuration (i.e. GMM based PDF modeling and NLL as anomaly score on semi-supervised splits), it can be seen that within-fabric performance is very robust, whereas performance across fabrics may vary considerably (Figure 4). This variation is stronger for AUPR compared to AUROC.

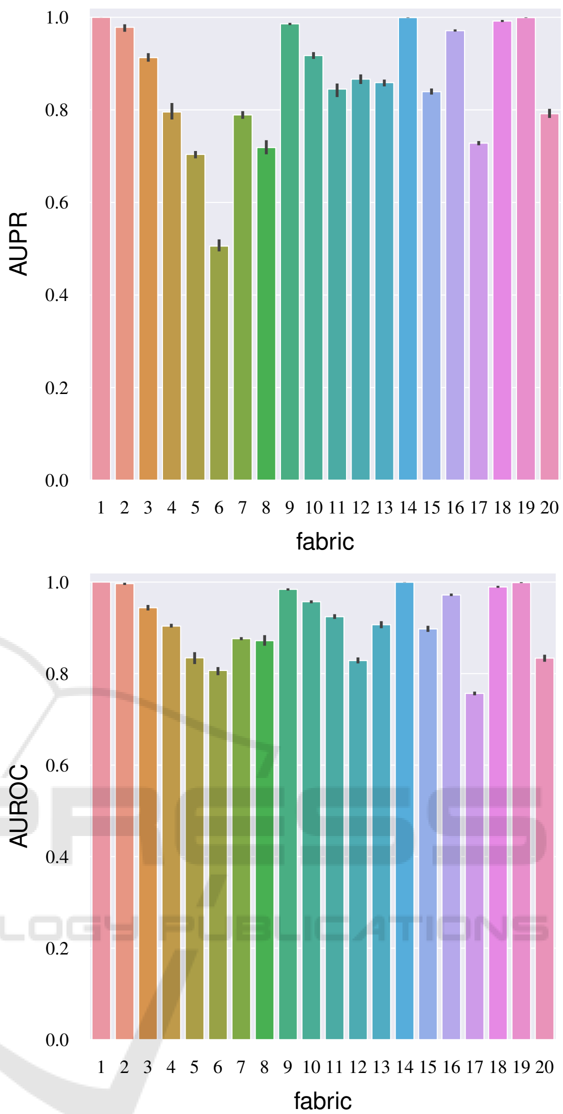


Figure 4: Fabric-level performance results of our LOO GMM NLL approach plotted as means and standard deviation over all 5 folds of semi-supervised dataset splits.

5 DISCUSSION

We have proposed and validated the modeling of PDFs in deep feature representations generated by large-scale dataset training as a method for extending supervised fabric anomaly detectors to previously unseen fabrics, achieving state-of-the-art performance. During evaluation, it was observed that LOO performance on its own (i.e. applying the learned model without modification to new fabrics) was already comparable to prior state of the art. The reason for this may be found in the clustering behavior of anomalies observed in the latent space visualizations, where anomalies form a cluster of their own (cf. Figure 3). Notably, this clustering is learned without any additional loss enforcing it. The poor and extremely varying performance of the supervised reference method, a single fabric classifier, can be explained by the small dataset sizes available for the individual fabrics, causing model overfitting. While positive results are achieved for supervised classifiers in literature, used datasets contain between 1 and 2 order of magnitudes more fabric anomalies than the single fabric datasets used here (Gao et al., 2019).

Overall, our approach is a simple yet elegant alternative to the defect synthesis based approaches proposed in previous work (Le et al., 2020; Han and Yu, 2020; Rippel et al., 2020b) and requires no training of GANs, which suffer from instable training. Furthermore, compared to above approaches, no additional model fine-tuning/training is required.

However, our work also has limitations. While superior performance was achieved by multimodal modeling of PDFs via GMMs, this approach is very sensitive to anomalies and thus cannot be applied in an unsupervised manner. Instead, it requires a clean, anomaly-free dataset, i.e. a semi-supervised setting. While such a dataset can be easily generated in practice, it would still be interesting to assess performance of more complex, non-parametric PDF estimation algorithms (e.g. (Trentin, 2018)) in the unsupervised setting. We hypothesize the reason for the aforementioned sensitivity to unlabeled anomalies lies in the clustering behavior of anomalies observed in latent embeddings (cf. Figure 3). GMM components will be fit by the EM algorithm to these clusters, and anomalies thus assigned a higher likelihood under our model. As the unimodal Gaussian is more rigid in it's assumptions about normal data distribution, it is less strongly affected by the anomaly clustering, but also yields less performance due to it's inability to reflect the multimodal nature of fabric appearance. If anomalies were to follow a diffuse PDF instead, as is a longstanding assumption in AD (Ruff et al., 2020a), we expect the GMM approach to be more resistant in the unsupervised regime. Apart from our work, said assumption has also been recently challenged by Ruff et al. (Ruff et al., 2020b).

Furthermore, high AUPR values could not be achieved for all fabrics. We give two possible explanations for this: First, undetected anomalies may be present in the datasets, which have been shown to be detrimental to model performance. We will therefore extensively relabel our dataset to eliminate all label noise. An alternative explanation would be that the learned feature representations fail to properly represent some normal data modes for fabrics that are significantly different from the initial dataset, which is supported by low AUPR values co-occurring with high AUROC values (cf. fabric 6 in Figure 4). This is congruent with observations made by Liu et al. (Liu et al., 2019), where their anomaly segmentation model is capable of defect segmentation in new fabrics but yields too many False Positives prior to fine-tuning. Reduction of False Positive Rate could be achieved by including normal data of new fabrics for a model retraining or alternatively fine-tuning, possibly improving results without requiring anomalies. Note that this would increase the complexity of the approach. When viewing the difference in appearance of new fabrics as input domain shifts, performance on new fabrics may be further increased by applying methods targeted at increasing model robustness to input domain shifts (e.g. AugMix as proposed by Hendrycks et al. (Hendrycks et al., 2020)).

6 CONCLUSION

In this work, we proposed the modeling of PDFs in deep representations as a useful transfer learning approach to extend deep fabric AD models to new, previously unseen fabrics. The approach is simple yet elegant and requires only a small dataset of normal images for PDF estimation. Our comparison against semi-supervised and supervised methods demonstrates the efficiency of our approach. We will further extend our approach by incorporating methods that increase robustness to input domain shifts in the initial model training phase. Additionally, we will investigate methods to fine-tuning learned feature representations directly using normal data only to further reduce False Positive Rate.

ACKNOWLEDGEMENTS

This work was supported by the German Federation of Industrial Research Associations (AiF) under the grant number 19811 N.

REFERENCES

Andrews, J., Tanay, T., Morton, E. J., and Griffin, L. D. (2016). Transfer representation-learning for anomaly detection. JMLR, Multidisciplinary Digital Publishing Institute.

Bishop, C. (2006). *Pattern Recognition and Machine Learning*. Springer.

Chandola, V., Banerjee, A., and Kumar, V. (2009). Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):1–58.

Gao, C., Zhou, J., Wong, W. K., and Gao, T. (2019). Woven fabric defect detection based on convolutional neural network for binary classification. In Wong, W. K., editor, *Artificial Intelligence on Fashion and Textiles*, pages 307–313, Cham. Springer International Publishing.

Han, Y.-J. and Yu, H.-J. (2020). Fabric defect detection system using stacked convolutional denoising autoencoders trained with synthetic defect data. *Applied Sciences*, 10(7).

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Hendrycks, D., Mu, N., Cubuk, E. D., Zoph, B., Gilmer, J., and Lakshminarayanan, B. (2020). AugMix: A simple data processing method to improve robustness and uncertainty. *Proceedings of the International Conference on Learning Representations (ICLR)*.

Hu, G., Huang, J., Wang, Q., Li, J., Xu, Z., and Huang, X. (2019). Unsupervised fabric defect detection based on a deep convolutional generative adversarial network. *Textile Research Journal*, 0(0):0040517519862880.

Karayiannis, Y. A., Stojanovic, R., Mitropoulos, P., Koulamas, C., Stouraitis, T., Koubias, S., and Papadopoulos, G. (1999). Defect detection and classification on web textile fabric using multiresolution decomposition and neural networks. In *ICECS'99. Proceedings of ICECS '99. 6th IEEE International Conference on Electronics, Circuits and Systems (Cat. No.99EX357)*, volume 2, pages 765–768 vol.2.

Kingma, D. P. and Ba, J. (2015). Adam: A method for stochastic optimization. In Bengio, Y. and LeCun, Y., editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Le, X., Mei, J., Zhang, H., Zhou, B., and Xi, J. (2020). A learning-based approach for surface defect detection using small image datasets. *Neurocomputing*.

Ledoit, O., Wolf, M., et al. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411.

Lee, K., Lee, K., Lee, H., and Shin, J. (2018). A simple unified framework for detecting out-of-distribution samples and adversarial attacks. In *Advances in Neural Information Processing Systems*, pages 7167–7177.

Liu, J., Wang, C., Su, H., Du, B., and Tao, D. (2019). Multistage gan for fabric defect detection. *IEEE Transactions on Image Processing*.

Mahalanobis, P. C. (1936). On the generalized distance in statistics. National Institute of Science of India.

Mei, S., Wang, Y., and Wen, G. (2018). Automatic fabric defect detection with a multi-scale convolutional denoising autoencoder network model. *Sensors*, 18(4).

Miyato, T., Kataoka, T., Koyama, M., and Yoshida, Y. (2018). Spectral normalization for generative adversarial networks. In *International Conference on Learning Representations*.

Newman, T. S. and Jain, A. K. (1995). A survey of automated visual inspection. *Computer vision and image understanding*, 61(2):231–262.

Rippel, O., Mertens, P., and Merhof, D. (2020a). Modeling the distribution of normal data in pre-trained deep features for anomaly detection. *arXiv preprint arXiv:2005.14140*.

Rippel, O., Müller, M., and Merhof, D. (2020b). GAN-based defect synthesis for anomaly detection in fabrics. In *2020 IEEE 25th International Conference on Emerging Technologies and Factory Automation (ETFA)*. IEEE.

Ruff, L., Kauffmann, J. R., Vandermeulen, R. A., Montavon, G., Samek, W., Kloft, M., Dietterich, T. G., and Müller, K.-R. (2020a). A unifying review of deep and shallow anomaly detection. *arXiv preprint arXiv:2009.11732*.

Ruff, L., Vandermeulen, R. A., Franks, B. J., Müller, K.-R., and Kloft, M. (2020b). Rethinking assumptions in deep anomaly detection. *arXiv preprint arXiv:2006.00339*.

Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464.

See, J. E. (2012). Visual inspection: a review of the literature. *Sandia Report SAND2012-8590, Sandia National Laboratories, Albuquerque, New Mexico*.

Shi, W., Caballero, J., Theis, L., Huszar, F., Aitken, A., Ledig, C., and Wang, Z. (2016). Is the deconvolution layer the same as a convolutional layer? *arXiv preprint arXiv:1609.07009*.

Trentin, E. (2018). Soft-constrained neural networks for nonparametric density estimation. *Neural Processing Letters*, 48(2):915–932.

Ulyanov, D., Vedaldi, A., and Lempitsky, V. S. (2016). Instance normalization: The missing ingredient for fast stylization. *CoRR*, abs/1607.08022.

Weninger, L., Kopaczka, M., and Merhof, D. (2018). Defect detection in plain weave fabrics by yarn tracking and fully convolutional networks. In *IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*.

Wu, Y., Zhang, X., and Fang, F. (2020). Automatic fabric defect detection using cascaded mixed feature pyramid with guided localization. *Sensors*, 20(3).

Zhang, H., Zhang, L., Li, P., and Gu, D. (2018). Yarn-dyed fabric defect detection with yolov2 based on deep convolution neural networks. In *2018 IEEE 7th Data Driven Control and Learning Systems Conference (DDCLS)*, pages 170–174.