

An Ensemble of Fine-Tuned Deep Learning Networks for Wet-Blue Leather Segmentation

by

Masood Aslam,¹ Tariq M. Khan,² Syed Saud Naqvi¹ and Geoff Holmes³

¹Department of Electrical and Computer Engineering, COMSATS University Islamabad, Islamabad Campus, Islamabad 45550, Pakistan

²School of Information Technology, Deakin University, Geelong, VIC 3217, Australia

³NZ Leather and Shoe Research Association (LASRA), Palmerston North 4414, New Zealand

Abstract

As part of industrial quality control in the leather industry, it is important to segment features/defects in wet-blue leather samples. Manual inspection of leather samples is the current norm in industrial settings. To comply with the current industrial standards that advocate large-scale automation, visual inspection based leather processing is imperative. Visual inspection of wet-blue leather features is a challenging problem as the characteristics of these features can take on a variety of shapes and colour variations to constitute various normal and abnormal surface regions. The aim of this work is to automatically segment leather images to detect various features/defects along with the background through visual analysis of the surfaces. To accomplish this, a deep learning-based technique is developed that learns to segment wet-blue leather surface features. On our own curated leather images dataset, the proposed ensemble network performed well, with an F1-Score of 74 percent.

1 Introduction

The leather industry is one of the most important historic industries. The leather is mostly supplied to downstream leather products companies, which utilise it as a raw material to make leather shoes, handbags, luggage, gloves, belts, and sofas, among other things.¹ Inspectors examine the leather by hand, physically inspecting it and marking any defects with chalk. Because manual inspection might lead to fatigue and misidentification, the final judgement must be double-checked and approved by numerous inspectors. As a result, a quick, thorough, and noninvasive leather assessment technique has become necessary.¹

Despite the fact that current manufacturing procedures fulfill high technological standards, quality inspection of leather products has room for improvement in terms of effectiveness and efficiency. The leather materials have a lot of natural defects/features because they are made from animal skin (cowhide, sheepskin, pigskin, and so on). On treated leather surfaces, scars, stains, wrinkles, cuts, and colour variations are common defects, whereas common normal features include creasing and folds that can be of interest due to their frequent occurrence. As a result, developing a systematic procedure for

evaluating leather surface faults/features is crucial for maintaining high-quality product consistency.

Artificial intelligence (AI) is the study of theories, methods, and systems that will allow machines to mimic human intelligence. Neural networks are a first step in this direction, seeking to mimic the capabilities of the human brain. Image processing is an area of AI that focuses on replicating human vision behaviour. While several AI techniques have helped in the replication of simple human vision, tasks like handwritten digit identification, number plate reading, and visual recognition, neural networks have made great progress.² Neural networks have demonstrated their efficacy on a wide range of image processing applications including image enhancement, compression, image segmentation, object recognition and image understanding.³ Image enhancement, compression, image segmentation, object recognition, and image understanding are just a few of the applications for neural networks.²

For visual inspections of leather, non-neural network based AI approaches have already been investigated.^{4,5} Despite their excellent performance, current technologies are still a long way from providing a generic solution for large-scale visual inspection. Furthermore, the full potential of CNN-based (a form of neural network) approaches for visual leather evaluation has yet to be realised. The unavailability of data sets, which is a major hindrance to expansion in this field, is one probable reason. The data from previous studies has not been made publically available for comparison.

The purpose of this research is to develop an automated defect detection system that can segment irregular areas of defects and interesting normal features. To analyse the test dataset and design a robust architecture, different instance segmentation deep learning models are used, such as UNet,⁶ Segnet,⁷ and Fully Convolutional Network (FCN).⁸ The major objective, as set out in this paper, is to propose a systematic ensembling of state-of-the-art CNNs and their adaptation for leather image segmentation.

In this study, we provide a systemically constructed ensemble convolutional neural network for robust leather sample segmentation based on empirical assessments. We also put together a new high-resolution wet-blue leather dataset with two classes containing

*Corresponding author email: tariq_mehmood@comsats.edu.pk

Manuscript received October 5, 2021, accepted publication November 7, 2021

images with abnormal feature types, whereas the third class contains normal leather images with interesting surface features, which occur frequently and therefore can not be neglected. So the classification problem becomes a generic one that is to learn three feature types based on the visual characteristics they present. The images were taken in a controlled environment with a digital camera. Appropriate lighting, distance from the leather surface, range of vision adjustment, and high-definition photos are all required for image capturing with a digital camera equipment.

The major contributions of the proposed work are:

- a new ensemble method for robust leather defect/feature segmentation,
- a thorough comparative evaluation of the proposed method with benchmark deep learning based segmentation methods,
- introduction of a new high-resolution leather images dataset for stimulating research in the field.

The rest of the paper is organized as follows. Section 2 contains a literature review on machine learning based leather defect inspection. Section 3 describes the proposed method. Section 4 explains the experimental design. In section 5 all the results are presented. Class Activation Maps are explained in section 6. Finally, we conclude our work in section 7.

2 Literature Review

In this section, we review several machine learning based methods proposed in the literature for leather defect segmentation. The task of instance segmentation, which demarcates the image's regions of interest, has received minimal attention from the research community. Lovergine et al.⁹ conducted one of the earliest studies on defect localization and segmentation. A black and white CCD

camera was used to detect and determine the defective areas. The texture orientation features of the leather are reconstructed using a morphological segmentation^{10,11} approach applied to the obtained images. The study has a few qualitative conclusions, but no quantitative methodologies or numerical data with which to evaluate the proposed methodology. Lanzetta and Tantussi¹² provide a laboratory prototype for trimming a leather's external part. Binarization, opening, and laplacian mask approaches are used to locate the trimming path from the leather sample images in order to find the background and troublesome regions. The proposed defect detection system correctly detects the majority of flaws on different types of leather. The surface polish and colour, on the other hand, are still the most essential factors that can influence the inspection's outcome.

A segmentation approach by Liang et. al.¹³ was utilised to automatically capture the leather image and locate defects. The total amount of image data obtained from a single piece of leather was 584, with the tick bite defect being mostly included in the little leather patches. To find the damaged regions, the spotting approach Mask R-CNN (Regional Convolutional Neural Network)¹⁴ was employed. When looking at training data, the average segmentation accuracy was 91.5 percent.

3 The Proposed Method

In this paper, we investigate deep learning architectures for defect segmentation in wet-blue leather. A Nikon Coolpix P300 camera was used to capture the photos. To increase the generalisation of segmentation models on unknown data, training images are augmented to add new image variants. Augmented data is used to train the models. For the inference stage, the best weights of the trained models are saved. The test data and the trained model are used in the inference phase to generate model predictions, which are then used to evaluate the model's performance in the test stage. The generic workflow of the proposed method is shown in Figure 1.

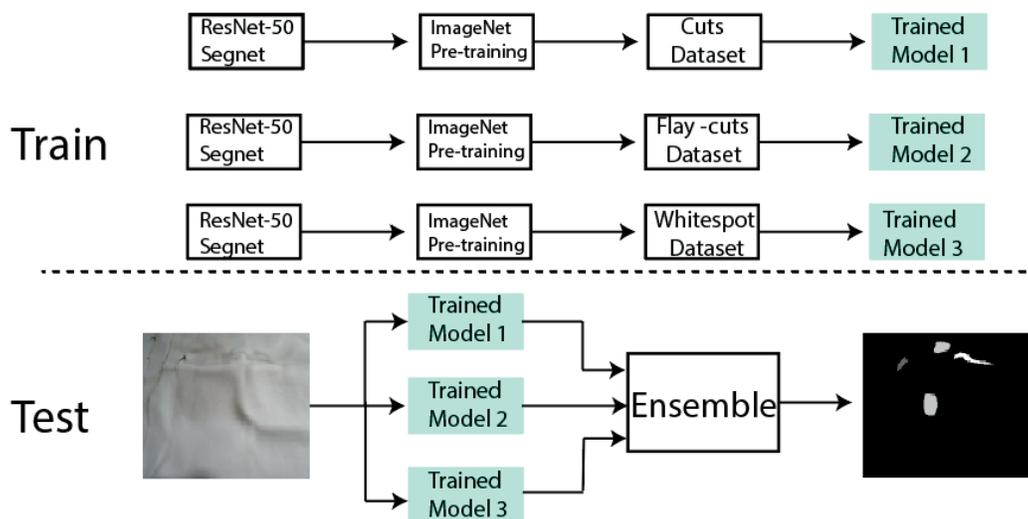


Figure 1. The generic workflow of the proposed method. The training phase is depicted on the top and the inference phase is shown on the bottom.

3.1 Data Augmentation

In order to train the network with different variations of the input images by artificially generating new images for the training, the data augmentation module was added to our workflow. In order to minimise network overfitting and enhance model generalisation, the data augmentation effect has been practically demonstrated.

The images are first resized to 512×384 before data augmentation. The chosen augmentation techniques employed in our experiments are horizontal flipping, vertical flipping and rotation using a random angle in the range [90, 360]. After the data augmentation process, a total of 960 training images were collected. The training data was split randomly into training, validation sets with a 80:20 ratio.

3.2 Architectures

Pixel-wise masks for each region of interest must be produced to denote the item discovered during the segmentation process. (FCN)⁸ is the first work to train FCN end-to-end for pixel-wise prediction. The key idea of FCN⁸ is to replace the fully connected layer of typical classification neural networks with convolutional layers so that a network output can be a two dimensional heat map, rather than class probability prediction. FCN⁸ implements “skip architecture”, meaning that shallow layer’s outputs are merged to deeper layers so that the network can maintain both local and coarse information. SegNet⁷ consists of an encoder network, a corresponding decoder network followed by a pixel-wise classification layer. Each encoder layer has corresponding decoder layer. The decoder used pooling indices computed in a max pooling step of the corresponding encoder to perform non-linear up-sampling. Since positional or boundary information are lost during the max pooling operations in the encoder network, maintaining positional information for each up-sampling operation in the decoder network is critical for accurate pixel-wise segmentation.

UNet⁶ is used to indicate the precise position of the discovered defect. To summarise, UNet employs a network architecture that supports both downsampling and upsampling. Another name for it is the encoder-decoder structure. The pooling layer is used by the encoder to gradually reduce the spatial dimension of the input data, while the deconvolution layer is used by the decoder to restore the target’s details and the proper spatial dimension. Most of the time, the encoder and decoder will share information directly. This is to help the decoder get the information it needs as quickly as possible. The instance segmentation system employs a four-level network structure. In summary, the encoder is made up of several building blocks, each with two convolutional operation layers and one pooling operation layer. Below the encoder, two convolution layers and a dropout layer are added to serve as a bridge to the decoder. In each block of the decoder design, a maxunpooling layers and two convolution layers follow the decoding method. Surprisingly, after each maxunpooling layer, the processed data is required to perform a depth concatenation with the output produced by the second convolution layer from the associated encoder block. Finally, a convolution layer, a softmax layer, and a pixel classification layer are employed for output prediction.

To ensure that the input and output volumes are always the same, padding is utilised on all convolutional layers.

3.2 Ensembling Convolutional Neural Networks

Ensemble approaches combine several segmentation models, and it has been found that it is possible to obtain greater precision results than a single model. To boost the segmentation performance we trained three models on each class and then combined these modes to form one segmentation model. Instead of training models on complete dataset we trained each model on individual class to make it a binary problem instead of multi-class problem. We trained three UNet⁶ and SegNet⁷ models with ResNet-50 backbone. We fused their results in the inference phase for a given test image.

There have been few papers on the application of deep learning for handling the problem brought about by a small dataset. For small datasets, all these articles use the ensemble technique. When only a limited dataset is employed, over-fitting might occur, resulting in poor segmentation. As a result, the model’s ability to generalize will be limited. To overcome this problem ensemble technique is used for small datasets.¹⁵⁻²¹ In this work, we experimented with various combinations of standard state-of-the-art networks, including UNet,⁶ Segnet⁷ and Fully Connected Network(FCN),⁸ in order to find the ideal collection for segmenting leather defects. The training and testing workflow diagram of the proposed ensemble network is shown in Figure 1. In both related and unrelated image domains, ensemble approaches have previously proven to be the most successful tool. In ensemble techniques, different segmentation models are used, and it has been found that they can achieve higher precision results than a single model. To combine cutting-edge designs, UNet,⁶ SegNet⁷ and the Fully Connected Network (FCN)⁸ were used. Because there are no documented representative ensemble approaches for leather defect segmentation in the literature, ensemble combinations of two network architectures were investigated, and combinations that stand out in terms of learnt representations were chosen for this study. Figure 1 shows three SegNet models, each of which is trained separately on each of the three classes. All of the models are individually trained. Following training, all of the models are integrated to form an ensemble that can be utilised to solve a multi-class segmentation problem.

4 Experimental Design

4.1 Dataset

Wet blue leather data was collected at the Fitzherbert Science Centre, Dairy Farm Road, Manawatu-Wanganui, by the Leather and Shoe Research Association of New Zealand (LASRA). Wet blue leather refers to unfinished hides that have been dehaired and chrome tanned to preserve the leather. Because of the chromium tanning chemicals, these skins are commonly referred to as “wet blue”. The images were taken with a Nikon Coolpix P300 camera with a 12MP AF sensor and under ideal lighting conditions. The regions containing the features constitute approximately 5-10 % of the entire

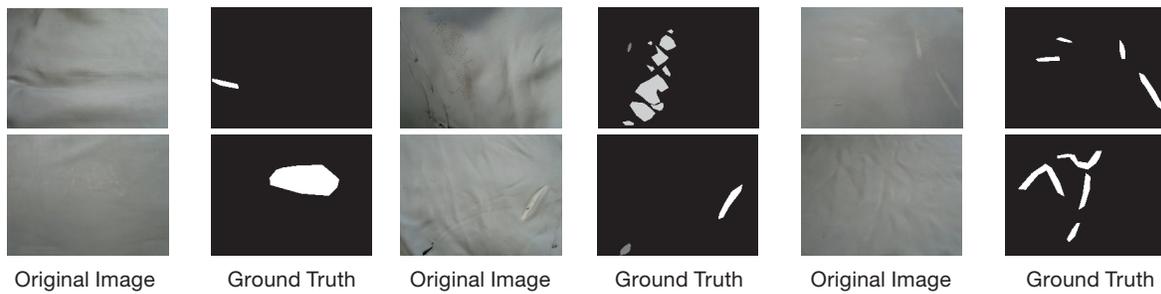


Figure 2. Dataset images with their corresponding ground truth.

image. In the dataset for each original image, the corresponding mask is provided. The mask image is considered to be the ground truth. A few example images from dataset are shown in Figure 2.

4.2 Feature/Defect Types

In this analysis, images having three types of surface features divided into three classes, respectively. The features of the first class have a set of characteristics close to cuts, termed as class A in this work. The class B contain images that include surface features that are in close resemblance to flay-cuts. Finally, the class C comprises of images containing regions that have interesting but normal features. Class C has a combination of surface features having the characteristics of creasing and folds. Although, these are not defects but have visually appealing characteristics, different appearance as compared to plain leather regions and a high frequency of occurrence. The features in class A are usually elliptical in shape and do not carry much textural information, with the exception of at the edges. Class B visual features, on the other hand, have a rich texture and distinct colour tones in general that set them apart. The images from class C are characterized by bright regions which appear either due to leather bends or folds from the pressing of the wet blue on a slamming felt. These regions of high saturation usually have an elongated shape and occur at a high degree of varying scale. The majority of images include

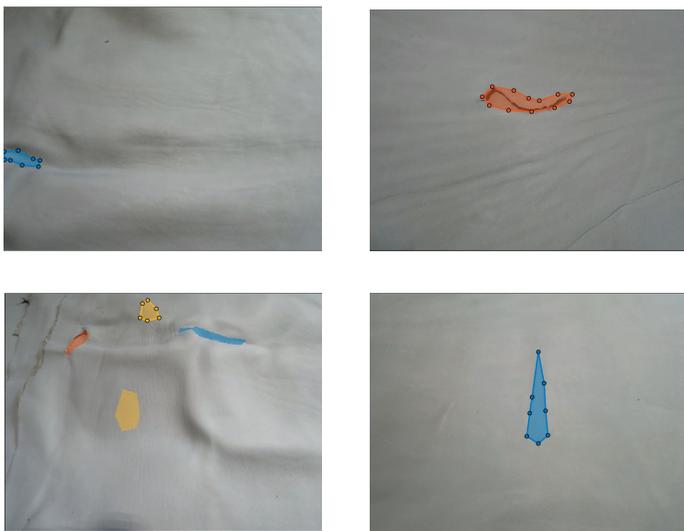


Figure 3. Representative examples of ground truth annotations.

Blue color represents class C, orange colour represents class A and yellow color represents class B images.

only one feature class, however a small percentage of images have multiple class variations.

4.3 Ground Truth Labelling

Image labeller, a popular MATLAB annotation tool, is used to create the dataset's ground truth annotations. The ability to annotate defects at various scales was a significant problem in the defect labelling process, necessitating the distinction between coarse and fine annotations. This was solved by establishing a balance of coarse and fine annotations. The ground-truth labelled images are shown in Figure 3.

4.4 Experiment configuration

The experiments are conducted using MATLAB 2021a on an Intel(R) Xeon(R) E51620 v2 3.70 GHz processor, RAM 32GB. All the images used in the segmentation task are resized to 512×384 . Concretely, all three FCN,⁸ UNet⁶ and SegNet⁷ architectures are trained with Adam as the optimization algorithm. The initial learning rate is set to 0.001 and the epoch value is 20. Pixel-wise classification layer is used as the last layer in all the algorithms for the segmentation mask output.

4.5 Quantitative measures

When performing classification predictions, four types of outcomes could occur

- **True Positive (TP):** When a defected leather sample is predicted as defected by the model,
- **True Negative (TN):** When a leather sample without any defects/features is predicted as non-defected by the model,
- **False Negative (FN):** When defects/features were present in leather but the model could not recognize them; it is also called as a Type 2 error,
- **False Positive (FP):** When the leather sample was non-defective but the model predicted it as defective (containing features); it is also known as a Type 1 error.

However, TN is not very meaningful in this research because the percentage of defects/features in this dataset is approximately 5-10%. Thus, TN is always high, even when the model does not predict defects/features well. The F1 and IOU scores are more important metrics in standard segmentation evaluation.

The precision of the segmentation model indicates what proportion of positive predictions were deemed correct and is given as

$$\text{Precision} = \frac{TP}{TP + FP} \quad (9)$$

A related measure is recall which measures the proportion of actual positives which were identified correctly

$$\text{Recall} = \frac{TP}{TP + FN} \quad (10)$$

The F1-score is the average of the precision and recall

$$\text{F1 - score} = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (11)$$

Consider the segmentation B_{gt} of a target object with a ground-truth bounding box and a B_{pr} prediction bounding box. Regardless of confidence level, a perfect match is defined as the projected and ground-truth boxes having the same area and location. The IOU is determined by multiplying the amount of overlap (intersection) in the object detection scope between the predicted bounding box B_{pr} and the ground-truth bounding box B_{gt} by the area of their union, which is given in equation

$$J(B_{pr}, B_{gt}) = \text{IOU} = \frac{\text{area}(B_{pr} \cap B_{gt})}{\text{area}(B_{pr} \cup B_{gt})}$$

5 Results

In this section, we begin by presenting a quantitative comparison of state-of-the-art segmentation models. Next, we present qualitative comparison of the employed segmentation models on representative images from the proposed dataset.

5.1 Quantitative Comparison

Table I shows the comparison of precision, recall, F1-score and IoU of state-of-the-art segmentation models on multi-class segmentation. UNet⁶ achieves 0.47 precision, 0.38 recall and 0.42 F1-score and IoU is 0.34. SegNet⁷ achieves 0.75 precision, 0.48 recall and 0.58 F1-score and IoU is 0.42. This shows that shallow networks are able to achieve better AP as compared to the deeper models.

The models' performance on multi-class segmentation had room for improvement, therefore the next step was to train the models on each class separately and evaluate how they performed. As a result, we divided the primary dataset into three datasets, one for each class, and trained the models on these datasets. The comparison of these classes is shown in Table II. Because FCN⁽⁸⁾ was unable to provide sufficient results in the binary class, its results are not included in Table II. The main reason for this is because the FCN⁸ model was over-fitting as the data was further separated into three sections, and it failed to produce any useful results because of small dataset and intra class variation which lead to over-fitting of FCN model.

In Table II, we can see the comparison of models on each class and we can see that SegNet performance was better than UNet⁶ in every class. For class A, SegNet was able to achieve F1-Score of 0.86 and IoU of 0.79 while UNet performance was also similar to SegNet. For class B, both models got identical F1-Score and IoU. On class C images, again SegNet outperformed UNet and achieved an F1-Score of 0.7 and an IoU of 0.61. As SegNet performance was better than UNet in each class. For ensemble model we combined all three models of SegNet. The result of our ensemble model which is also the proposed

Table I

Comparison of the segmentation performance of our proposed method with other state-of-the-art segmentation models for the wet-blue leather dataset

Serial #	Model	Precision	Recall	F1-Score	IoU
1	UNet ⁶	0.47	0.38	0.42	0.34
2	SegNet ⁷	0.75	0.48	0.58	0.42
3	FCN ⁸	0.85	0.46	0.60	0.44
4	Proposed	0.85	0.71	0.74	0.69

Table II

Class-wise comparison of the segmentation performance of state-of-the-art segmentation methods

Defect	Model	Backbone	Precision	Recall	F1-Score	IoU
Class A	SegNet ⁷	ResNet-50	0.83	0.90	0.86	0.79
	UNet ⁶	ResNet-50	0.85	0.83	0.84	0.76
Class B	SegNet	ResNet-50	0.88	0.63	0.71	0.63
	UNet	ResNet-50	0.77	0.67	0.71	0.63
Class C	SegNet	ResNet-50	0.83	0.65	0.70	0.61
	UNet	ResNet-50	0.84	0.6	0.66	0.58

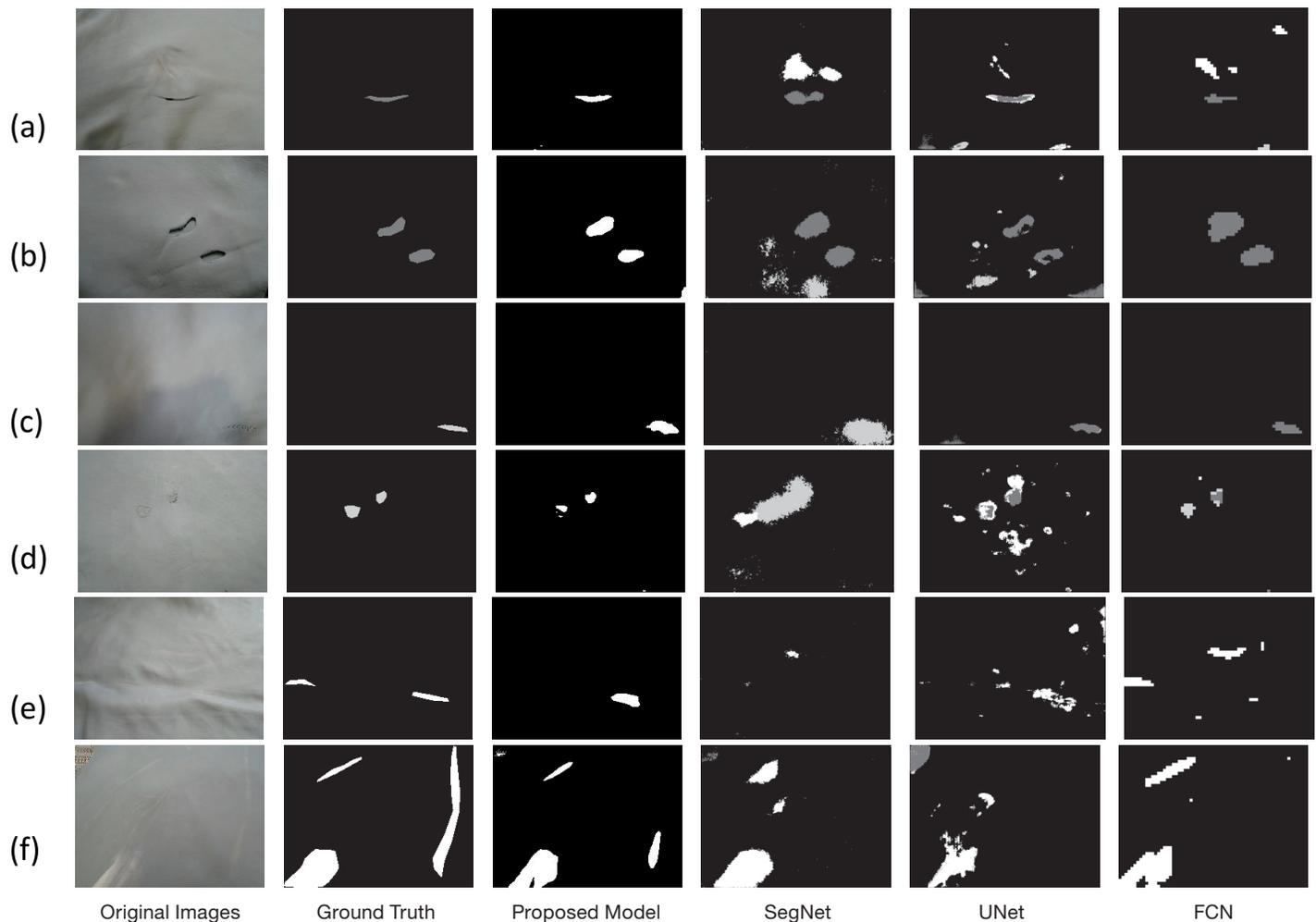


Figure 4. Visual comparison of the state-of-the-art methods for leather defect detection. (a) and (b) rows represent segmentation results of class A, (c) and (d) rows represent segmentation results of class B, (e) and (f) represents segmentation results on class C images.

model is given in Table I. We can see in Table I that proposed model outperformed all other state-of-the-art models.

5.2 Qualitative Comparison

Figure 4 shows the experimental results of different algorithms for defect/feature segmentation. In the first column (a) at number one is the original image that contains class A features, second is the ground truth, third is the result of the proposed method which obtained a close segmentation. The fourth and fifth images contain results of SegNet⁷ and Unet, respectively. Both were able to segment the cut but there were a few false segmentations as well. Similarly, the sixth image contains segmentation results of FCN,⁸ which is similar to both UNet and SegNet.⁷ If we look at the results of column (b) (proposed model) and FCN, both segmented the class A image and do not give false positive results, whereas Unet and SegNet again gave a few false positives as well. In column (c), except for Unet, all the models predicted class B where as Unet, while able to segment the class B region, gave few false positives as well. In column (d), again only the proposed method was able to give the correct segmentation, and again the other models were not able to

give the correct segmentation. Column (e) and (f) contain regions with class C features, which is a very difficult class because of intra class variation. Again, the proposed model performed better than other models but even the proposed model failed to give exact segmentation.

6 Conclusion

Automated visual inspection of leather in an industrial setting has recently received a lot of interest. There have been numerous machine learning algorithms proposed in the past, but convolutional neural network-based approaches remain rare. In this paper, we provide an ensemble convolutional neural network for visual inspection of wet-blue leather. In terms of F1-Score and IoU, our proposed method outperformed three state-of-the-art deep learning-based segmentation algorithms. A F1-Score of 74 percent and IoU of 66.7 percent were achieved by our model. Regardless of its competitive performance, the proposed method would need to be converted for real-time use, which would include thorough video data tweaking. A

significant future direction is the development of a system that can segment leather data in a real-world industrial scenario. Finally, the development of such systems that can characterise different defect types in terms of their properties could lead to artificial intelligence-based automated quality grading of leather samples.

Acknowledgment

This work was supported by NZ Leather and Shoe Research Association (LASRA*), Palmerston North, New Zealand through the Ministry of Business, Innovation and Employment (MBIE) grant number LSRX1801.

References

1. M. Aslam, T. M. Khan, S. S. Naqvi, G. Holmes, and R. Naffa, "On the Application of Automated Machine Vision for Leather Defect Inspection and Grading: A Survey," *IEEE Access*, vol. 7, pp. 176065–176086, 2019, doi: 10.1109/ACCESS.2019.2957427.
2. T. Durand, N. Thome, and M. Cord, "WELDON: Weakly Supervised Learning of Deep Convolutional Neural Networks," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 4743–4752.
3. M. Egmont-Petersen, D. de Ridder, and H. Handels, "Image processing with neural networks—a review," *Pattern Recognit.*, vol. 35, no. 10, pp. 2279–2301, 2002, doi: [https://doi.org/10.1016/S0031-3203\(01\)00178-9](https://doi.org/10.1016/S0031-3203(01)00178-9).
4. C. Harris and M. Stephens, "A combined corner and edge detector," in *In Proc. of Fourth Alvey Vision Conference*, 1988, pp. 147–151.
5. K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
6. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9351, pp. 234–241, May 2015, Accessed: Oct. 04, 2021. [Online]. Available: <https://arxiv.org/abs/1505.04597v1>.
7. V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017, doi: 10.1109/TPAMI.2016.2644615.
8. J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Nov. 2014, Accessed: Oct. 04, 2021. [Online]. Available: <https://arxiv.org/abs/1411.4038v2>.
9. F. P. Lovregine, A. Branca, G. Attolico, and A. Distanto, "Leather inspection by oriented texture analysis with a morphological approach," in *Proceedings of International Conference on Image Processing*, 1997, vol. 2, pp. 669–671.
10. C. Giardina and E. Dougherty, "Morphological methods in image and signal processing," *undefined*, 1988.
11. C. K. Lee and S. P. Wong, "A mathematical morphological approach for segmenting heavily noise-corrupted images," *Pattern Recognit.*, vol. 29, no. 8, pp. 1347–1358, 1996, doi: 10.1016/0031-3203(96)86888-9.
12. M. Lanzetta and G. Tantussi, "Design and Development of a Vision Based Leather Trimming Machine," *AMST'02 Adv. Manuf. Syst. Technol.*, pp. 561–568, 2002, doi: 10.1007/978-3-7091-2555-7_64.
13. S. T. Liong, Y. S. Gan, Y.-C. Huang, C.-A. Yuan, and H.-C. Chang, "Automatic Defect Segmentation on Leather with Deep Learning," 2019.
14. K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2017-October, pp. 2980–2988, Dec. 2017, doi: 10.1109/ICCV.2017.322.
15. T. Mauldin, A. H. Ngu, V. Metsis, and M. E. Canby, "Ensemble Deep Learning on Wearables Using Small Datasets," *ACM Trans. Comput. Healthc.*, vol. 2, no. 1, pp. 1–30, 2021, doi: 10.1145/3428666.
16. A. Cazananas-Gordon, E. Parra-Mora, and L. A. D. S. Cruz, "Ensemble Learning Approach to Retinal Thickness Assessment in Optical Coherence Tomography," *IEEE Access*, vol. 9, pp. 67349–67363, 2021, doi: 10.1109/ACCESS.2021.3076427.
17. Y. Chen, D. Li, X. Zhang, J. Jin, and Y. Shen, "Computer aided diagnosis of thyroid nodules based on the devised small-datasets multi-view ensemble learning," *Med. Image Anal.*, vol. 67, p. 101819, 2021, doi: 10.1016/j.media.2020.101819.
18. E. Baykal Kablan, H. Dogan, M. E. Ercin, S. Ersoz, and M. Ekinici, "An ensemble of fine-tuned fully convolutional neural networks for pleural effusion cell nuclei segmentation," *Comput. Electr. Eng.*, vol. 81, p. 106533, 2020, doi: 10.1016/j.compeleceng.2019.106533.
19. B. Savelli, A. Bria, M. Molinara, C. Marrocco, and F. Tortorella, "A multi-context CNN ensemble for small lesion detection," *Artif. Intell. Med.*, vol. 103, no. April 2019, p. 101749, 2020, doi: 10.1016/j.artmed.2019.101749.
20. M. Aslam, T. M. Khan, S. S. Naqvi, G. Holmes, and R. Naffa, "Ensemble Convolutional Neural Networks with Knowledge Transfer for Leather Defect Classification in Industrial Settings," *IEEE Access*, vol. 8, pp. 198600–198614, 2020.
21. M. Aslam, T. M. Khan, S. S. Naqvi, G. Holmes, and R. Naffa, "Learning to Recognize Irregular Features on Leather Surfaces," *J. Am. Leather Chem. Assoc.*, vol. 116, no. 5, 2021.