university of groningen / faculty of science and engineering

Classifying asbestos roofs in the Dutch province of Drenthe using hyperspectral imagery and deep learning

Bachelor's Project Thesis

Nick Ubels, n.s.ubels@student.rug.nl,
Supervisors: Dr. M.A. Wiering & K. Dijkstra

Abstract: Asbestos was used a lot due to its favourable characteristics, but after discovering the health impact of the fibres it was widely banned. Not all asbestos is already gone however, it is still used on various roofs spread across The Netherlands. These roofs need to be remediated before a proposed nationwide ban goes into effect. This study focusses on the province of Drenthe in the north of The Netherlands. Hyperspectral aerial imagery is used to classify buildings as asbestos, clean or suspicious. In an earlier study this has been done using Spectral Angle Mapping, and this is used as a baseline. In this study two deep neural networks, U-Net and DeepLabv3+, are compared to Spectral Angle Mapping to determine if deep learning is a feasible alternative. The networks are trained on three different datasets varying in quality and size. After testing it is found that DeepLabv3+ has better performance on this problem than Spectral Angle Mapping. DeepLabv3+ is able to obtain a mean Intersection over Union of 0.41 versus 0.32 for Spectral Angle Mapping.

1 Introduction

Asbestos was long seen as a miracle material due to its very promising characteristics. Some of these characteristics are resistance to heat and its affordability. This meant that asbestos was widely used for consumer products and construction materials like asbestos roofing. During the end of the 20th century, due to health concerns, asbestos was more and more banned. Asbestos is known to cause a form of cancer called mesothelioma (Kanarek, 2011).

In The Netherlands the government banned the use, re-use, storage, selling, importing and handling of asbestos containing products in July 1993, this was followed by an European ban in 2005. The Dutch ban means that the removal of asbestos containing products is only allowed to be done by specialised companies under strict regulations. The Dutch government is currently working on a full ban on asbestos roofing by 2024. The proposed law has already been approved by the House of Representatives and is currently pending review in the Senate (Rijksoverheid, 2019).

Due to the dangerous nature of asbestos roofing and the planned ban on the material it is important to know which buildings contain asbestos roofing materials. With respect to the ban it is important to have this information to be able to inventorise which buildings still need to be remediated before the deadline. In terms of public safety and environmental protection it is important to have this information in case of emergencies like fires. During a fire asbestos is often spread into the environment. When the presence of asbestos is known clean-up operations could start earlier.

The roofing material itself, often corrugated sheets, contains about 12% asbestos. The rest of the material is cement. In this application the asbestos, almost always chrysotile, is bound in the cement (Harmsma and Mulder, 2006).

In this study hyperspectral aerial imagery is used. This imagery is different from regular aerial imagery. These regular aerial photos only have three bands representing the colours red, green and blue. In some cases an extra infrared band is captured which falls in the visible and near-infrared

portion of the electromagnetic spectrum (between 400 nm and 1400 nm). Aerial imagery consisting of more than three bands is often referred to as multispectral imagery. The hyperspectral imagery used in this study has 420 bands spanning a far wider range on the electromagnetic spectrum. That means that not only the visible and near-infrared light is captured, but also part of the shortwave infrared light. This ranges between 1400 nm and 3000 nm (Schlessinger and Spiro, 1995). As will be explained in section 2.1.2 the sensor used in this study has a spectral range up to 2500 nm.

Hyperspectral data has been used in earlier studies to classify asbestos roofs. In the Karpacz area of Poland Krówczyńska, Wilk, Pabjanek, Zagajewski, and Meuleman (2016) performed a study using the Spectral Angle Mapper (SAM) technique on a small area of the town Karpacz (2 km^2). Krówczyńska et al. (2016) were able to correctly classify 61.54% of the asbestos roofs and 97.98% of the asbestos free roofs. On the other side of Europe in the Aosta Valley in Italy (3263 km^2) Frassy, Candiani, Rusmini, Maianti, Marchesi, Nodari, Via, Albonico, and Gianinetto (2014) also used SAM to detect asbestos roofs. Frassy et al. (2014) were able to correctly classify 43% of the asbestos roofs. However in this study, conducted in a mountainous area, there was a lower spatial resolution caused by not adjusting the altitude above the valley. This led to misclassifications on small buildings. Heuff, Schuurmans, Peters, Vroeijenstijn, and van den Brink (2017) of Royal HaskoningDHV also used SAM to classify asbestos roofs in the Province of Drenthe, in the same area of interest and using the same dataset as this study. They were able to correctly classify 55.8% of the asbestos roofs and 52.4% of the clean roofs.

Deep learning methods rely on multiple layers that are able to learn representations, it can thus be seen as a form of representation learning. In a deep learning method these learned representations vary in terms of abstraction and these different abstraction levels combined make it possible to learn complex representations. For this reason deep learning methods are able to perform well on a wide range of tasks and disciplines and even improving the state of the art in tasks like image recognition (LeCun, Bengio, and Hinton, 2015).

There are at least three main types of methods to discriminate in the case of imagery. The first type, classification, concerns the classification of objects within imagery. Given an image the task at hand is to classify the object visible in the image. The second type, detection, concerns the detection of objects within imagery. Given an image the task is to detect all objects of a given class in that image. The third type, segmentation, concerns the segmentation of different classes of objects within imagery. In this case the task at hand is to discriminate between classes (on pixel level) within an image. Segmentation is the most suitable method for this study because the exact location and shape of buildings are known. An example of an earlier study using segmentation on hyperspectral data is the study by Hu, Huang, Wei, Zhang, and Li (2015) who propose a deep convolutional neural network for classification of materials from hyperspectral aerial imagery. This study is also an example of segmentation.

This study aims to be an extension of the work by Heuff et al. (2017) and in lesser form the work of Krówczyńska et al. (2016) and Frassy et al. (2014) which were all studies using SAM to detect asbestos roofs. As the work by Heuff et al. (2017) concerns the same area of interest a direct comparison will be made to this work.

The research question in this study is: Can Deep Neural Networks outperform Spectral Angle Mapping for the classification of asbestos roofs using hyperspectral data?

In Section 2 the used data, preprocessing and the used networks are explained. The experimental setup of this study will be explained in Section 3. The results of the experiments will be presented in Section 4 and discussed in Section 5. Section 6 wraps everything together and details suggestions for future work.

2 Methods

In this section the methods used in this study will be explained. Section 2.1 will focus on the specifics of the used hyperspectral dataset and the preprocessing, Section 2.2 will explain the SAM method and Section 2.3 will explain the used deep learning methods.

2.1 Data

This section details the region of interest, the used sensor, the flight campaign, the preprocessing steps

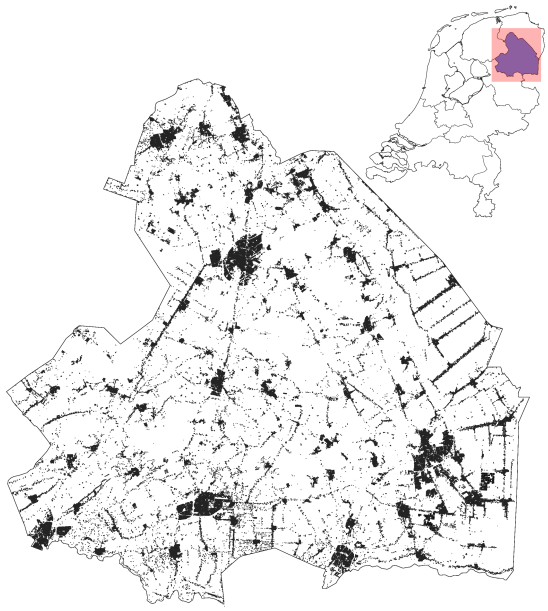

Figure 2.1: Map showing the buildings within the province of Drenthe. Data: province outline by Centraal Bureau voor de Statistiek (2018), buildings by Geodienst Rijksuniversiteit Groningen (2018)

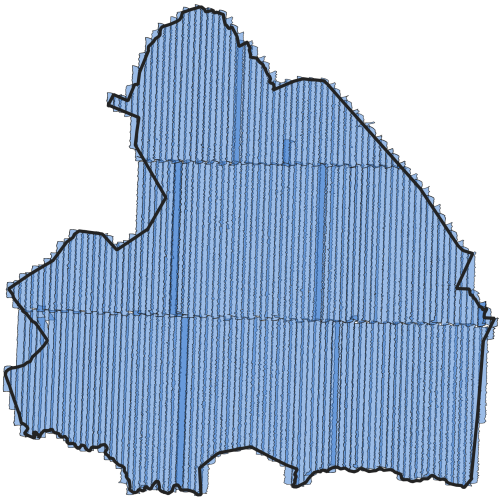

Figure 2.2: Map showing coverage of the hyperspectral data. Darker areas indicate overlap. Data: province outline by Centraal Bureau voor de Statistiek (2018).

and the composition of the training and validation sets.

2.1.1 Region of interest

The region of interest in this study is the province of Drenthe in The Netherlands. This province is located in the north of the country and consists primarily out of arable land and nature. It has a total area of 2680 km^2 and built up area totalling 48 km^2 which is about 1.8% of the province as can be seen in Figure 2.1.

2.1.2 Sensor and flight campaign

Data has been gathered in May 2016 during a flight campaign of 21 days. In total there were 2 test flights and 14 measurement flights operated with a Cessna TU206E. During these 14 measurement flights 240 strips of approximately 20 kilometres long and 1 kilometre wide have been captured from an altitude of 1500 metres. The spatial resolution is 1 m. There is approximately 20% overlap between the strips to prevent gaps caused by movement of the aircraft. The weather conditions during the flight campaign were good, due to lack of clouds. Turbulence however caused a small missing spot of about 8 km^2 (Heuff et al., 2017). A coverage map is included in Figure 2.2.

Inside the plane a Specim AisaFENIX 1K full spectrum hyperspectral camera was mounted. This camera is able to capture 420 bands with a spectral range from 380 nm to 2500 nm (Specim, Spectral Imaging Ltd., 2017). All 420 bands are used in this study. The hyperspectral data used is radiance data. Radiance is the amount of light the sensor is able to pick up from the object being observed. That does mean that atmospheric light is also caught by the sensor, and that the atmosphere will absorb light that will not be able to reach the sensor (Harris Geospatial Solutions, 2013).

2.1.3 Preprocessing

To be able to use the hyperspectral data for deep learning purposes preprocessing was necessary. This preprocessing is primarily extracting the

buildings from the larger hyperspectral strips to be able to train the neural networks. As 1.8% of the total surface area of Drenthe consists of buildings, discarding other areas will speed up processing significantly.

First of all it was important to make sure that all the different datasets were always in the same projected coordinate system to make sure building polygons line up with the strip outlines and the hyperspectral strips itself. As the hyperspectral strips were captured in WGS 84 / UTM zone 32N (EPSG:32632) it was decided to use this projected coordinate system as the primary projected coordinate system for this study. All other datasets (for example the *Basisregistratie Adressen en Gebouwen (BAG)*, Key Register of Addresses and Buildings in English) have been reprojected from Amersfoort / RD New (EPSG:28992) to WGS 84 / UTM zone 32N (EPSG:32632).

In order to generate sets for training and validation purposes, lists with labelled buildings in ESRI Shapefile format were created (see Section 2.1.4) using building polygons from the BAG. This has one limitation: buildings that consist of several building parts are seen as one building in the BAG. There is no Dutch governmental dataset available that is able to distinct between building parts. That means that in the case of buildings with only one building part roofed with asbestos the entire building is labelled as such (also the asbestos free roofs within this building). An example of such a building is given in Figure 2.3. This building has two parts, a house in the front with an attached shed in the back. The house is roofed with regular roof tiles and the shed is roofed with asbestos.

Using the Rasterio (Gillies et al., 2013–) and Fiona (Gillies et al., 2011–) software packages the building polygons were cut out of the hyperspectral strips. The buildings were masked in such a way that only the building contained data, the surroundings of the building was zeroed out as background class. This was done in order to prevent problems in areas where buildings are close together, residential areas for example. Each side of the image and mask is then padded to the next multiple of 32. This is done because 32 by 32 pixels is the minimum input size for the networks. During validation pixels that are classified as background class and fall within a building polygon are classified as asbestos free. An example of such a cut

Figure 2.3: A combined building in the municipality of Coevorden. Left: hyperspectral data in false RGB. Right: aerial footage by PDOK (2016)

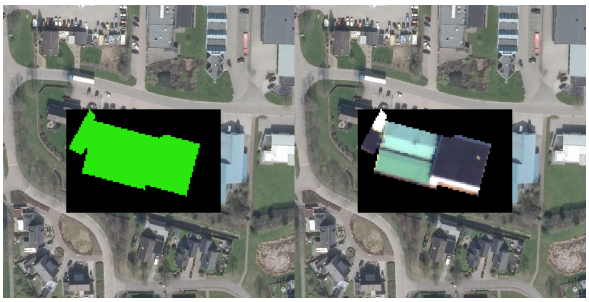

Figure 2.4: Cut out building in Dwingeloo. Left: the cut out mask. Right: the cut out hyperspectral building. Background aerial footage by PDOK (2016)

out building can be found in Figure 2.4. In both the mask and cut out building it is visible that the zeroed out background class (black) covers parts of neighbouring buildings and that only the building itself contains data.

Before the data is used in training or validation it is normalised. This is done by calculating the mean μ and standard deviation σ per band and applying equation 2.1 on each pixel in each band. The resulting normalised pixel on position (i, j) in band b is z_{ij}^b. x is the original pixel on the same position. μ and σ are for the corresponding band b. In all cases the μ and σ are only calculated with training data, for validation the values from the training set are used.

$$z_{ij}^b = \frac{x_{ij}^b - \mu^b}{\sigma^b} \tag{2.1}$$

Data augmentation has been done by randomly cropping training images to a size of 32 by 32 pixels

Figure 2.5: Example of an edited building polygon. Left: original polygon taken from the BAG. Right: edited polygon in which the house is cut off. Background aerial footage by PDOK (2016)

in order to build uniform mini batches. For images with the minimum size of 32 by 32 pixels the entire image is used. Training images are also randomly horizontally and/or vertically flipped with a probability of 50%.

2.1.4 Training and validation sets

For this study four datasets are built, all varying in size. Of those datasets 80% is used for training and 20% for validation. This split is kept the same across the different networks.

For the 96 dataset a list of buildings with asbestos roofs was provided by the Province of Drenthe and these building polygons were edited to remove any building parts which were labelled as asbestos that actually did not contain asbestos. This was done using visual inspection based on aerial photography and an example of such an edit can be found in Figure 2.5. For the 400 and 1000 datasets the buildings with asbestos roofs were randomly sampled out of a list with buildings that have been remediated after the flight campaign. All asbestos free buildings have been randomly sampled from a list with buildings that were determined asbestos free by Heuff et al. (2017), their classification method will be explained in Section 2.2. The test dataset, consisting of 189 buildings, consists of buildings that were inspected by the *Regionale Uitvoeringsdienst Drenthe* in the municipality of Emmen. An overview of the datasets can be found in Table 2.1 and the spectral profiles of the 96, 400 and 1000 set can be found in Appendix A.

2.2 Spectral angle mapping by Heuff et al. (2017)

The spectral angle mapping implementation by Heuff et al. (2017) used as a reference in this research is based on equation 2.2 in which $\theta(x, y)$ is the spectral angle between the pixel x and reference spectrum y for n bands.

$$\theta(x, y) = \cos^{-1}\left(\frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}}\right) \quad (2.2)$$

Heuff et al. (2017) selected 26 reference spectra from buildings that were known to have an asbestos roof. Each pixel is compared to each of these 26 reference spectra. The resulting angles from equation 2.2 are thresholded against 4 rad. If a pixel has at least one of the 26 angles below this threshold it is classified as asbestos.

When less than 10% of the pixels within a building are classified as asbestos the building is marked as asbestos free, between 10% and 50% the building is marked as suspicious and above 50% the building is marked as asbestos roofed (Heuff et al., 2017).

2.3 Classification

For classification using deep convolutional neural networks proposed in this study two existing networks are compared. Both of these networks are implemented in PyTorch (Paszke, Gross, Chintala, Chanan, Yang, DeVito, Lin, Desmaison, Antiga, and Lerer, 2017). U-Net as proposed by Ronneberger, Fischer, and Brox (2015) with some modifications and DeepLabv3+ as proposed by Chen, Zhu, Papandreou, Schroff, and Adam (2018) using ResNet-101 as a network backbone. Both networks are trained on 32 by 32 pixel crops from the full training images and evaluated on the full size images. In the classification using deep neural networks an extra class is used in comparison to the method described in Section 2.2. This extra class represents the masked out background as described in Section 2.1.3. In the case of evaluation the final softmax for both networks is replaced by the maximum of the arguments (argmax) in order to determine the class per pixel.

For both networks an initial learning rate of 0.001 is used with a batch size of 10 buildings. After every 100 epochs the learning rate is halved.

Table 2.1: The four datasets used in this study, their composition, the amount of strips the buildings were sampled out of and roof surfaces in m^2

Dataset	Strips	Asbestos			Clean		
		Build.	Sum (m^2)	Mean ± SD (m^2)	Build.	Sum (m^2)	Mean ± SD (m^2)
96	76	48	29472	601 ± 551	48	51817	1057 ± 1662
400	153	200	103912	519 ± 588	200	18617	84 ± 124
1000	183	500	275593	551 ± 695	500	68612	137 ± 940
Test	11	42	12358	294 ± 507	147	21955	149 ± 206

2.3.1 U-Net

U-net as proposed by Ronneberger et al. (2015) is a fully convolutional network consisting of two paths in a distinctive U shape. One of the paths, the contracting downwards path, consists of two 3×3 convolutions followed by a rectified linear unit and a 2×2 max pooling operation with a stride of 2. The purpose of the max pooling operation is to downsample the input. The amount of feature channels is doubled every downwards step in the first 3×3 convolution. On the expansive upwards path each step is a transposed 2×2 convolution that halves the feature channels. The resulting feature map is concatenated with the feature map from the downwards path on the same depth. This is followed by two times a 3×3 convolution in combination with rectified linear units. The final step is a 1×1 convolution to map the feature vector to the three classes used in this study followed by a softmax (or argmax during evaluation). For training stochastic gradient descent is used with a momentum of 0.9.

The first change made to the original U-net architecture is the padding of the 3×3 convolutions in the contracting downwards path. This was necessary as the original U-net reduces the size of the output image since the convolutions were unpadded. In this study the size and (spatial) resolution of the images has to preserved in order to be able to plot the results back onto the map.

The second change made is the reduction of the depth of the network. The original U-net had a depth of five, in this study this is reduced to three. This is done because the max pooling operation halves the size of the image (originally 32×32). In the case of a depth of five this would mean that, with the four max pooling operations in the network, the images are reduced to 2×2 pixels in the lowest level of the network. To prevent this the depth is reduced to three, resulting in a size of 8×8 in the lowest level.

2.3.2 DeepLabv3+

DeepLabv3+ as proposed by Chen et al. (2018) is a network with an encoder-decoder structure and is an extension of DeepLabv3 proposed in Chen, Papandreou, Schroff, and Adam (2017). In this study the version with ResNet-101 (He, Zhang, Ren, and Sun, 2015) is used as a network backbone. An atrous separable convolution (atrous spatial pyramid pooling) is used within the encoder, the results from this atrous separable convolution are then concatenated and convolved with a 1×1 convolution followed by a batch normalisation and a rectified linear unit. For use in the decoder a bilinear upsampling of 4 is performed.

The decoder takes the low level features determined by ResNet-101 and performs a 1×1 convolution followed by a batch normalisation and rectified linear unit. This is then concatenated with the upsampled results from the encoder. The concatenated result is followed by twice a 3×3 convolution with batch normalisation and rectified linear unit. This is then convolved with a 1×1 convolution to map the feature vector to the three classes used in this study followed by a softmax (or argmax during evaluation). A final bilinear upsampling of 4 is done to recover the original resolution of the image.

The backbone ResNet-101 (He et al., 2015) is a deep residual network consisting of 101 layers. The network is built with building blocks consisting of a 1×1 convolution followed by a rectified linear unit. This is followed by a 3×3 convolution with again a rectified linear unit. After this a 1×1 convolution is performed. Every building block has a residual connection adding the feature map to the feature map at the end of the building block. This is followed

by a rectified linear unit. The exact composition of this network can be found in Table 1 of (He et al., 2015).

For training of DeepLabv3+ the Adam optimiser is used.

2.3.3 Loss functions

Two loss functions are used in this study, cross-entropy loss and Intersection over Union loss. The cross-entropy loss is defined, per pixel, as Equation 2.3 in which p is the prediction, t the truth, c the classes and x the individual pixels.

$$L_{CE}(p,t) = -\sum_{x}\sum_{c} t_c^x \log p_c^x \tag{2.3}$$

The Intersection over Union loss is defined as Equation 2.4 as proposed by Atiqur Rahman and Wang (2016). In the equation p is the prediction, t the true label and c classes. Because not every class exists in each individual image the IoU is calculated over the entire batch.

$$L_{IoU}(p,t) = 1 - \left(\frac{1}{c}\sum_{c}\frac{|t_c \times p_c|}{|t_c + p_c - (t_c \times p_c)|}\right) \tag{2.4}$$

3 Experimental setup

To be able to test the performance of the networks are tested in different configurations resulting in a total of 12 combinations. The split between the data used for training and the data used for validation is that 80% is used for training and 20% is used for validation purposes. The split is randomly generated once and then kept the same over the different trials in order to be able to compare the performance. The exact numbers of training and validation images per dataset can be found in Table 3.1.

To be able to compare the performance of the two networks each network is trained with the same hyperparameters given in Table 3.2. U-net is trained with stochastic gradient descent as optimiser and DeepLabv3+ with the Adam optimiser. As mentioned before in section 2.3 the learning rate is halved every 100 epochs. Furthermore early stopping is used as described in Prechelt (2012). Every 100 epochs the mean Intersection over Union is calculated on the validation set, if this value is larger than last iteration training will stop and the previous snapshot of the network will be used for evaluation purposes.

Table 3.1: Amount of training and validation images per dataset

Dataset	Training	Validation
96	78	18
400	320	80
1000	800	200

Table 3.2: Hyperparameters used to train the networks

Parameter	Value
Max Epochs	1000
Batch-size	10
Learning rate	0.001

To assess the performance of the networks the mean Intersection over Union is calculated over the asbestos and clean class. This evaluation is done on the label given to the building using the thresholds as described in Section 2.2. This means that buildings with less than 10% of the pixels classified as asbestos are labelled as clean, between 10% and 50% as suspicious and above 50% the building is marked as asbestos.

Unlike Heuff et al. (2017) the class of suspicious buildings is taken into account in this study. Buildings that are classified as suspicious will be marked as misclassified. A building cannot be suspicious in the real world, it either has an asbestos roof or not.

The mean Intersection over Union is calculated as shown in equation 3.1. In this equation TP_x is the amount of true positives for class x, FP_x the amount of false positives for class x and FN_x represents the amount of false negatives for class x.

$$mIoU = \frac{1}{x}\sum_{x}\frac{TP_x}{TP_x + FP_x + FN_x} \tag{3.1}$$

The percentage of suspicious buildings, $\%_{sus}$ is the percentage of buildings classified as suspicious with respect to the total amount of misclassified buildings. The precision for class x is calculated as shown

in equation 3.2.

$$P_x = \frac{TP_x}{TP_x + FP_x} \times 100 \tag{3.2}$$

The recall for class x is calculated as shown in equation 3.3.

$$R_x = \frac{TP_x}{TP_x + FN_x} \times 100 \tag{3.3}$$

4 Results

The study by Heuff et al. (2017) had a mean Intersection over Union of 0.32 and 84.9% of the incorrectly classified buildings were classified as suspicious. For asbestos roofed buildings there was a precision of 50.0% and a recall of 22.6%. For clean buildings Heuff et al. (2017) had a precision of 51.0% and a recall of 78.1%.

Results of the experiments detailed in Section 3 are given in Table 4.1. What is immediately clear is that the percentage of the erroneous classified buildings that were labelled as suspicious is far lower for U-Net and DeepLabv3+ than SAM by Heuff et al. (2017).

For the dataset containing 96 buildings only DeepLabv3+ trained with cross-entropy loss is able to perform better than SAM. Where SAM obtains an $mIoU$ of 0.32 DeepLabv3+ is able to obtain 0.36. This comes in combination with higher precision and recall scores across both classes. U-Net is able to score a high precision on asbestos when trained with IoU loss, but this comes with a low precision on clean. That means that U-net in this case classifies most buildings as asbestos resulting in the lowest $mIoU$ of 0.14.

U-Net trained with IoU loss is able to perform the best on the dataset consisting of 400 buildings scoring a $mIoU$ of 0.38. However the precision on asbestos is lower than SAM with 31% versus 50%. The higher $mIoU$ is caused by a higher precision on the clean class. DeepLabv3+ is close to SAM on this dataset with a slightly lower $mIoU$ of 0.31 versus 0.32 for SAM caused by a slightly lower precision on the asbestos class.

DeepLabv3+ is the best performing network on the dataset consisting of 1000 buildings scoring a $mIoU$ of 0.41 with both loss functions. However the DeepLabv3+ trained with IoU loss has a worse precision on asbestos than the DeepLabv3+ trained with cross-entropy loss. The DeepLabv3+ trained with cross-entropy loss on the 1000 dataset can thus be seen as the network with the overall best performance.

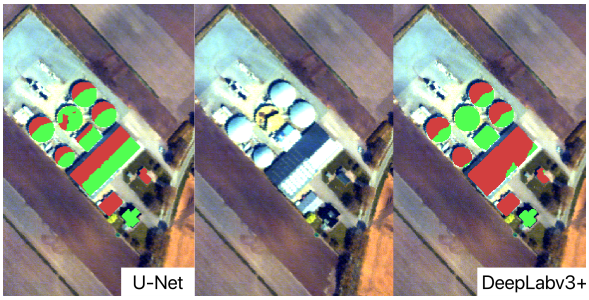

Figure 4.1: Pixelwise segmentation results in different lighting conditions, red is classified as asbestos and green as clean. From left to right: U-Net, raw image and DeepLabv3+

Some example classifications are shown in Figure 4.1 which is based on the networks trained with the 96 dataset and cross-entropy loss. In this example it can be seen that different lighting on parts of a building can lead to different classes due to this lighting difference.

5 Discussion

There were some difficulties that were encountered during this study. In Section 5.1 difficulties with regards to data quality will be discussed. Section 5.2 discusses differences in lighting conditions and Section 5.3 the implications of the suspicious buildings class.

5.1 Data quality

While the hyperspectral data is of excellent quality with a good coverage and almost no cloud coverage the difficulty in terms of data quality lies mostly in the BAG. As explained in section 2.1.3 this Dutch governmental dataset does not discriminate between different building parts. When only a small part of a building has an asbestos roof this could mean that the individual pixels could be classified correctly but the amount of asbestos pixels is not enough to reach the threshold. This will lead to misclassifications as the entire building is then seen as asbestos free or suspicious. A solution could be to look at individual roof surfaces.

Table 4.1: Mean IoU $mIoU$, percentage of error classified as suspicious $\%_{sus}$, precision for asbestos P_a, recall for asbestos R_a, precision for clean P_c and recall for clean R_c on the test dataset

Dataset	Loss func.	Network	Epochs	$mIoU$	$\%_{sus}$	P_a	R_a	P_c	R_c
96	L_{CE}	U-Net	400	0.17	22.6%	88.1%	22.4%	12.9%	79.2%
		DeepLabv3+	200	**0.36**	10.6%	64.3%	**27.8%**	**53.4%**	**83.7%**
	L_{IoU}	U-Net	100	0.14	14.4%	**95.2%**	22.6%	6.8%	83.3%
		DeepLabv3+	200	0.30	14.0%	59.5%	23.1%	43.5%	79.0%
400	L_{CE}	U-Net	200	0.23	47.0%	52.4%	18.5%	34.0%	71.4%
		DeepLabv3+	200	0.31	14.1%	47.6%	22.2%	52.4%	77.8%
	L_{IoU}	U-Net	300	**0.38**	26.5%	31.0%	**25.0%**	**73.5%**	**78.8%**
		DeepLabv3+	100	0.26	15.7%	**61.9%**	22.0%	37.4%	77.5%
1000	L_{CE}	U-Net	200	0.34	40.5%	28.6%	19.7%	66.7%	76.6%
		DeepLabv3+	100	**0.41**	13.0%	**50.0%**	**30.4%**	67.3%	**82.5%**
	L_{IoU}	U-Net	200	0.39	16.7%	28.6%	25.0%	**75.5%**	78.7%
		DeepLabv3+	200	**0.41**	9.2%	40.5%	29.8%	72.8%	81.1%
		SAM		0.32	84.9%	50.0%	22.6%	51.0%	78.1%

Another difficulty related to this is that for the datasets with 400 and 1000 buildings it cannot be guaranteed that the entire building polygon used for the asbestos containing class is 100% asbestos containing. That means that during training the networks are also giving non asbestos containing roofs within buildings that are marked as asbestos containing. Of this building the front part, the house, is roofed with regular roof tiles whereas the barn part is roofed with asbestos roofing material. The full buildings is however marked as asbestos roofed in the training set. Editing the polygons by hand could be a possible solution.

A third difficulty with the data quality is the uncertainty concerning asbestos free buildings. Heuff et al. (2017) did not collect any ground truth data on asbestos free buildings. That means that all asbestos free buildings selected for this study are randomly picked out of the buildings that Heuff et al. (2017) marked as asbestos free. No guarantees can be given that these selected buildings are completely asbestos free. This could be resolved by inspecting more buildings.

5.2 Light

The hyperspectral data has been gathered over multiple days by flying sequential strips as explained in section 2.1.2. This has the effect that the lighting conditions are different for each of the captured strips. That means that, in combination with the fact that the hyperspectral dataset is radiance data, this could lead to extra difficulties.

5.3 Suspicious class

Both SAM in the study by Heuff et al. (2017) and the networks used in this study produce three classes: Building is asbestos roofed, building has a clean roof and building is suspicious. In this study the buildings that are classified as suspicious are seen as buildings that were classified incorrectly. Out of the total errors made by SAM 84.9% of the incorrectly classified buildings were classified as suspicious. In the case of the best performing DeepLabv3+ this is only 13%. This means that DeepLabv3+ marks more buildings incorrectly than SAM. This could be caused by the fact that SAM only looks at one pixel at a time whereas the networks used in this study take surrounding pixels into consideration.

It could be argued that it is advantageous to have more buildings classified as suspicious than erroneously classified buildings. Buildings that are marked as suspicious can be used to decide which buildings to further inspect to determine their final class. With buildings that are erroneously classified this is not possible.

5.4 Asbestos properties

As mentioned in the introduction the asbestos cement used in roofing material contains about 12% asbestos. The remaining 88% is cement, which is also used in other roofing materials. This could explain why the spectra shown in Appendix A are so similar.

5.5 Parameters

Parameters used in this study might not be the optimal parameters. An example of these parameters are the thresholds from Heuff et al. (2017) that were also used in this study.

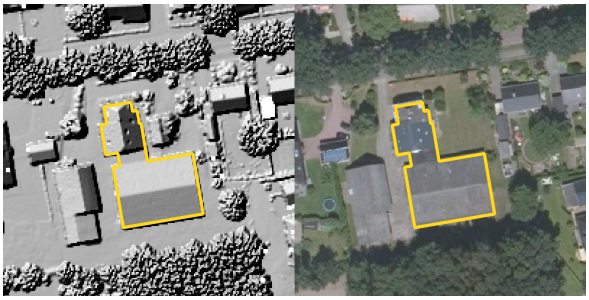

Figure 6.1: A combined building in the municipality of Coevorden. Left: hillshaded digital surface model *Actueel Hoogtebestand Nederland* (PDOK, 2012). Right: aerial footage by PDOK (2016)

6 Conclusion

Looking back at the research question for this study: 'Can Deep Neural Networks outperform Spectral Angle Mapping for the classification of asbestos roofs using hyperspectral data?' we can conclude that DeepLabv3+ with ResNet-101 as a backbone outperforms SAM. DeepLabv3+, trained on the 1000 dataset with cross-entropy loss, was able to score a mean Intersection over Union of 0.41 versus 0.32 for SAM. DeepLabv3+ is able to obtain the same precision as SAM on asbestos and a higher precision on clean roofs. For both the asbestos and clean roofs DeepLabv3+ is able to score a higher recall. It should however be noted that DeepLabv3+ has a lower percentage of the error marked as suspicious than SAM. U-Net was able to obtain a higher mean Intersection over Union than SAM in a few cases. In all these cases the precision on asbestos was lower.

The lower percentage of suspicious buildings in this work in comparison to Heuff et al. (2017) could be explained by the fact that the deep neural networks take the surrounding pixels into account when assessing a pixel. It was also argued in Section 5.3 that having more buildings classified as suspicious could be beneficiary as these can be eligible for extra inspection.

In terms of the data quality several difficulties were identified in Section 5.1, of which most concerned the fact that the *Basisregistratie Adressen en Gebouwen* is unable to distinguish different building parts within a building. Suggestions on how to tackle this will be given in Section 6.1.

Furthermore the material itself, as mentioned in Section 1 consists of about 12% asbestos making it hard to distinguish from other cement rich roofing materials. It should also be noted that the used hyperspectral data is radiance data. This is influenced by external light sources like the sun and different lighting conditions as was explained in Section 5.2. Section 6.1 will explain another data type which might be more suitable for this problem.

In conclusion deep learning, and specifically DeepLabv3+, is outperforming SAM but some improvements could still be made to improve performance which will be detailed in the next section.

6.1 Future work

In future research the data quality of the building polygons used could be improved. A method to do this could be to use a digital surface model like *Actueel Hoogtebestand Nederland (AHN)* to extract different roof surfaces based on calculated roof slopes and slope directions. An example of the digital surface model AHN is given in Figure 6.1, the different parts of the roofs are clearly visible in the hillshaded digital surface model.

By using the extracted roof surfaces during classification the problem of buildings consisting of multiple building parts can be solved. This can be done by classifying and thresholding each roof surface and then classifying the building as asbestos when one of the roof surfaces meets the threshold for asbestos. This can improve the performance on buildings where only a small amount of the total

roof surface is covered with asbestos material.

Another suggestion for future research is to improve the data quality of the sets used for training and validation. As explained in Section 2.1.4 the asbestos building polygons are edited by hand to keep the dataset as pure as possible. This was not done for the 400 and 1000 datasets. By removing the noise of clean building parts from the asbestos building polygons training could improve.

The last suggestion for future research with regards to the used data is to use reflectance data instead of radiance data. As explained in Section 2.1.2 the data used in this study is radiance data which is influenced by atmospheric light. Such influence of light was seen in Figure 4.1. Reflectance data is corrected for this atmospheric light and will be able to capture the properties of the material better. As this study is primarily concerning the distinction of materials this could be of great help.

Another suggestion would be the use of a much simpler network: a multi layer perceptron on small patches (3 by 3 pixels for example). In Dijkstra, van de Loosdrecht, Schomaker, and Wiering (2017) this was a successful approach for hyperspectral images of potato leaves. That study also uses dimensionality reduction which could be interesting for this application, especially due to the high number of spectral bands.

As mentioned in Section 5.5 the parameters used in this study might not be the optimal parameters. A parameter sweep could be attempted to optimise these parameters.

Acknowledgements

Data for this research has been provided by Province of Drenthe, Regionale Uitvoeringsdienst Drenthe and Royal HaskoningDHV. The Research and Innovation Support department of the University of Groningen provided (geo)support, data storage and access to the Peregrine high performance computing cluster.

References

Md Atiqur Rahman and Yang Wang. Optimizing intersection-over-union in deep neural networks for image segmentation. volume 10072, pages 234–244, December 2016. ISBN 978-3-319-50834-4. doi: 10.1007/978-3-319-50835-1_22.

Centraal Bureau voor de Statistiek. Gebiedsindelingen provincie. [Dataset], 2018.

Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *CoRR*, abs/1706.05587, 2017.

Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. *CoRR*, abs/1802.02611, 2018.

Klaas Dijkstra, Jaap van de Loosdrecht, Lambert Schomaker, and Marco Wiering. Hyper-spectral frequency selection for the classification of vegetation diseases. In *Proceedings European Symposium on Artificial Neural Networks*, Bruges (Belgium), April 2017.

Federico Frassy, Gabriele Candiani, Marco Rusmini, Pieralberto Maianti, Andrea Marchesi, Francesco Rota Nodari, Giorgio Dalla Via, Carlo Albonico, and Marco Gianinetto. Mapping asbestos-cement roofing with hyperspectral remote sensing over a large mountain region of the italian western alps. *Sensors (Basel, Switzerland)*, 14(9):15900–15913, 09 2014. doi: 10.3390/s140915900.

Geodienst Rijksuniversiteit Groningen. Basisregistratie adressen en gebouwen. [Dataset], 2018.

Sean Gillies et al. Fiona is OGR's neat, nimble, no-nonsense API. [Software], 2011–. URL `https://github.com/Toblerity/Fiona`.

Sean Gillies et al. Rasterio: geospatial raster I/O for Python programmers. [Software], 2013–. URL `https://github.com/mapbox/rasterio`.

S. Harmsma and H.F.H.M. Mulder. Asbest in kaart, March 2006.

Harris Geospatial Solutions. Radiance vs. reflectance, 2013. URL `https://www.harrisgeospatial.com/Support/Self-Help-Tools/Help-Articles/Help-Articles-Detail/ArtMID/10220/`

ArticleID/19247/3377. [Online; accessed 4-January-2019].

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

Fiona Heuff, Hanneke Schuurmans, Ferdinand Peters, Roeland Vroeijenstijn, and Cors van den Brink. *Handleiding verwachtingskaart asbestdaken Provincie Drenthe*. Royal HaskoningDHV, April 2017.

Wei Hu, Yangyu Huang, Li Wei, Fan Zhang, and Hengchao Li. Deep convolutional neural networks for hyperspectral image classification. 2015:1–12, 2015. doi: 10.1155/2015/258619.

Marty S. Kanarek. Mesothelioma from chrysotile asbestos: Update. 21(9):688–697, 2011. doi: 10.1016/j.annepidem.2011.05.010.

Małgorzata Krówczyńska, Ewa Wilk, Piotr Pabjanek, Bogdan Zagajewski, and Koen Meuleman. Mapping asbestos-cement roofing with the use of apex hyperspectral airborne imagery: Karpacz area, poland –a case study. *Miscellanea Geographica*, 20(1):41–46, 2016. doi: https://doi.org/10.1515/mgrsd-2016-0007.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521:436 EP –, May 2015.

Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. In *NIPS-W*, 2017.

PDOK. Actueel hoogte bestand 2, 0.5m ruw. [Dataset], 2012.

PDOK. Luchtfoto 25cm rgb. [Dataset], 2016.

Lutz Prechelt. Early stopping — but when? In *Lecture Notes in Computer Science*, pages 53–67. Springer Berlin Heidelberg, 2012. doi: 10.1007/978-3-642-35289-8_5.

Rijksoverheid. De belangrijkste asbestregels. https://www.rijksoverheid.nl/onderwerpen/asbest/asbestregels, 2019. [Online; accessed 5-January-2019].

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*, pages 234–241. Springer International Publishing : Cham, 2015. ISBN 978-3-319-24573-7 978-3-319-24574-4. doi: 10.1007/978-3-319-24574-4_28.

Monroe Schlessinger and Irving J. Spiro. *Infrared technology fundamentals*. M. Dekker, New York, 1995. ISBN 0824792599 9780824792596.

Specim, Spectral Imaging Ltd. Specim aisafenix 1k data sheet, 2017. URL http://www.specim.fi/downloads/AisaFenix_1K%20ver1-2017_web.pdf.

A Spectral profiles

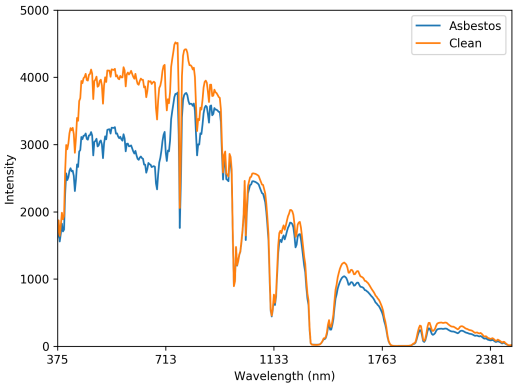

Figure A.1: Spectral profile of asbestos and clean roofs in the 96 dataset

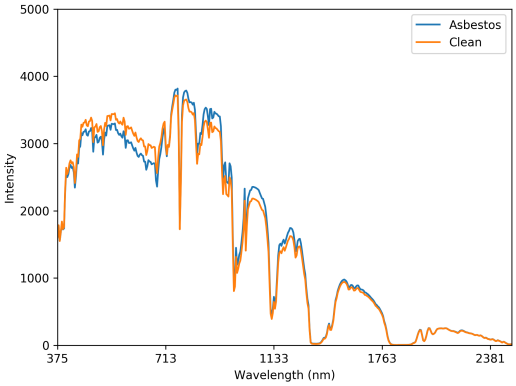

Figure A.2: Spectral profile of asbestos and clean roofs in the 400 dataset

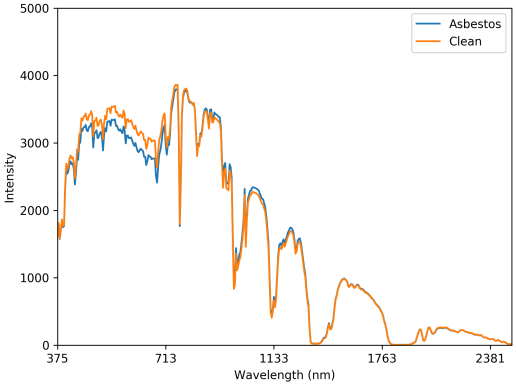

Figure A.3: Spectral profile of asbestos and clean roofs in the 1000 dataset