Elimination of Reflections in Laser Scanner Systems by using Convolutional Neural Networks

O. Alstad [1]   O. Egeland [1]

[1] *Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology (NTNU), N-7491 Trondheim, Norway. E-mail: {from,ola.alstad,olav.egeland}@ntnu.no*

Abstract

This paper presents a machine learning approach for eliminating reflections in line laser scanning of aluminium workpieces to be welded. The elimination of reflections is important to obtain accurate laser scanning of workpiece geometry for highly reflective materials like aluminium. The proposed solution is to use a convolutional neural network (CNN) which is trained to eliminate the reflections. The training of the network is done by simulating the laser line of the scanner in ray-tracing software using aluminium surfaces with appropriate reflection properties. This CNN then recovers the reflected laser line by removing the reflections. The CNN is used with two different camera configurations. In the first configuration one camera and one laser scanner are used. In the second configuration two cameras are used in a stereo arrangement in combination with the line laser. In this case, the planar homography of the laser plane is used to detect possible points on the laser line in a preprocessing step. The high performance of the solution is demonstrated for simulated data.

Keywords: Robotic welding, laser, reflections, filtering, CNN

1 Introduction

In robotic welding it may be required to measure the 3D shape of the workpieces, which can be done with laser line scanners. These 3D measurements can be used to plan the weld path. This technique is challenging to use for robotic welding of aluminium, since aluminium parts that are prepared for welding are highly reflective. Laser scanners will then give strong reflections of the laser line, and it is difficult to determine the geometry of the workpieces.

Laser line scanners are commonly used in a structured light setup for robotic welding, where a line is used as the projected pattern. A comprehensive review is found in (Blais, 2004). A camera with a known geometric relationship to the laser is used to triangulate the measurements of the projection of the line on a two-dimensional cross-section. The scanner is swept along the workpiece to get the full 3D point cloud of the surface.

When the laser line is projected at reflective metals like aluminium, second-order reflections can cause false measurements. These false measurements can be densely clustered and deviate significantly from the actual surface, and reconstruction to 3D can lead to a false geometry which may cause obvious problems when it comes to quality control, seam tracking, and weld path planning (Köhler et al., 2012).

One solution for measuring reflective materials is to apply removable powder to the workpiece before the measurement (Frank Chen and Mumin, 2000), but this approach slows measurement speed and may give reduced accuracy.

In (Trucco et al., 1994) a stationary laser line is projected at reflective objects moving on a conveyor belt, with two cameras mounted at opposite sides of the conveyor belt. The false measurements caused by specular reflections are removed by a geometric constraint,

where true measurements must be consistent for both cameras. A solution with direct calibration and additional consistency methods using two cameras was proposed in (Trucco et al., 1998). A direct illumination constraint ensures that a ray of light can not intersect a surface twice. All ambiguous points are removed when more than one point is observed under this constraint. An observable surface constraint enforces that the surface normals must face towards both of the cameras to be observable. An unobscured-once-viewed constraint states that if a valid point is only seen by one camera, there must be a valid point seen by the other camera that obscures the first point. These constraints were used to filter spurious measurements, but also removed some true points, which gave a ragged appearance of the scan.

Clark et al. (1997) used polarization for disambiguation of the true laser line from erroneous reflections. The projected laser line is linearly polarized and the polarization state of the observed reflections is used to differentiate the true measurements from the false. Their initial experiments showed great promise at removing spurious reflections from an aluminium workpiece.

Li et al. (2019) used a system with two cameras and a line laser, and used a laser plane constraint. The laser stripes were extracted from the images using a structured-light stripes extraction algorithm before the 3D points were calculated with optical triangulation. Erroneous points that did not satisfy the laser plane constraint were removed, which reduced the percentage of erroneous points from 12.61 to 3.05.

In this paper, we present a machine learning approach to the filtering of reflections in laser scanning systems. A simulation pipeline for generating laser scans with spurious reflections on reflective metal parts is presented. A convolutional neural network (CNN) is trained to remove false measurements caused by erroneous reflections. The simulation pipeline and application of the CNN are used with two different camera configurations. The first configuration uses one camera and a line laser. The second configuration uses two cameras in a stereo setup with a line laser. For the stereo camera setup a preprocessing step is proposed, to efficiently combine the views of each of the cameras. The preprocessing step is inspired by the work of Trucco et al. (1994) and Li et al. (2019), while the constraint is learned by the CNN in 2D image space rather than explicitly enforced in 3D coordinates.

The paper is organized as follows. Section 2 presents scanner geometry, reflection models, and the U-net architecture. Section 3 presents the method used for dataset generation and simulation-based training. Section 4 states how the experiments were conducted. Sec-

tion 5 presents the results of a simulation study for the CNN system.

2 Preliminaries

2.1 Laser scanner geometry

A standard laser scanner consists of a camera and a laser that has a fixed geometrical relationship as shown in Figure 1. The laser projects a line that deforms in the view of the camera as it hits an object. The deformed profile of the points allows a direct calculation of the range, which can further be expressed as 3D Cartesian coordinates (Blais, 2004).

Consider a laser scanner set-up with a camera and a line laser. The camera frame is denoted by c and the laser frame is denoted by s. The displacement from the laser frame to the camera frame is given by the rotation matrix (Lynch and Park, 2017)

$$\boldsymbol{R}_{cl} = [\boldsymbol{r}_1, \ \boldsymbol{r}_2, \ \boldsymbol{r}_3] \tag{1}$$

from c to l and the translation vector $\boldsymbol{t}_{cl}$ from c to l in the coordinates of frame c as shown in Figure 1. The scanner projects a laser plane that coincides with the yz plane of the laser frame s.

The geometry can be described in terms of Plücker coordinates (Pottmann and Wallner, 2001). Consider the Euclidean point $\boldsymbol{x} = [x_1, x_2, x_3]^T$ in the 3D scene, where the coordinates are given in c, and let $\tilde{\boldsymbol{x}} = (\boldsymbol{x}, 1)$ be the corresponding homogeneous representation. Consider a plane with Plücker coordinates $\tilde{\boldsymbol{u}} = (\boldsymbol{u}, u_4)$, where $\boldsymbol{u}$ is the unit normal vector of the plane. Then $\tilde{\boldsymbol{x}}$ is on the plane $\tilde{\boldsymbol{u}}$ whenever $\tilde{\boldsymbol{u}} \cdot \tilde{\boldsymbol{x}} = \boldsymbol{u} \cdot \boldsymbol{x} + u_4 = 0$, and it follows that $u_4 = -\boldsymbol{u} \cdot \boldsymbol{x}$, where $\boldsymbol{x}$ is an arbitrary point on the plane.

The unit normal vector to the laser plane is $\boldsymbol{r}_1$, and, since $\boldsymbol{t}_{cs}$ is a point on the laser plane, it follows that the Plücker coordinates of the laser plane with reference to frame c are given by

$$\tilde{\boldsymbol{u}} = (\boldsymbol{r}_1, -\boldsymbol{t}_{cs} \cdot \boldsymbol{r}_1) \tag{2}$$

Consider a 3D point $\boldsymbol{x}$ in the laser plane $\tilde{\boldsymbol{u}}$ that is imaged to the 2D point $\boldsymbol{s}$ in the normalized image plane (Hartley and Zisserman, 2003). Then $\boldsymbol{s} = \boldsymbol{x}/x_3$ and $u_4 = -\boldsymbol{x} \cdot \boldsymbol{u}$, which gives

$$\boldsymbol{s} \cdot \boldsymbol{u} = \frac{\boldsymbol{x} \cdot \boldsymbol{u}}{x_3} = -\frac{u_4}{x_3} \tag{3}$$

It follows that the 3D point can be calculated from

$$\boldsymbol{x} = x_3 \boldsymbol{s} = -\frac{u_4}{\boldsymbol{s} \cdot \boldsymbol{u}} \boldsymbol{s} = \frac{\boldsymbol{t}_{cs} \cdot \boldsymbol{r}_1}{\boldsymbol{s} \cdot \boldsymbol{r}_1} \boldsymbol{s} \tag{4}$$
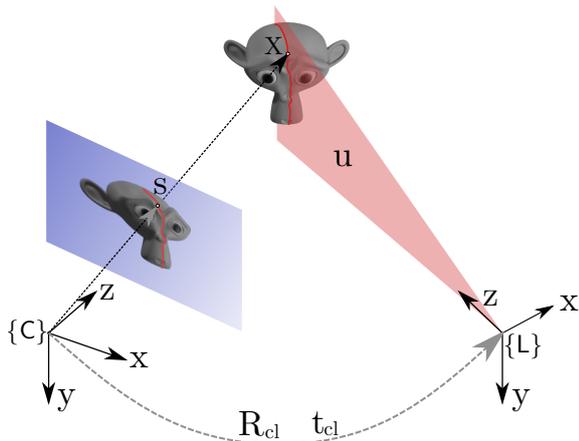
Figure 1: Laser scanner geometry. The camera frame is shown to the left and the laser frame to the right. The normalized image plane is drawn in front of the camera.

2.2 Laser scanner with two cameras

When a laser scanner is used with two cameras, then a 3D point on the laser plane will have to satisfy a planar homography induced by the plane (Hartley and Zisserman, 2003), and this can be used in the preprocessing of the images. In this section, this planar homography is presented. Consider two cameras in a stereo arrangement in combination with a laser scanner. Let the displacement from camera frame 2 to camera frame 1 be given by the rotation matrix $\boldsymbol{R}$ from camera frame 2 to camera frame 1, and the translation $\boldsymbol{t}$ from 2 to 1 in the coordinates of camera frame 2. Consider a point in the scene which is on the laser plane. Let the point have position $\boldsymbol{x}_1$ in camera frame 1, and, accordingly, position

$$\boldsymbol{x}_2 = \boldsymbol{R}\boldsymbol{x}_1 + \boldsymbol{t} \qquad (5)$$

in camera frame 2. The distance from the origin of camera frame 1 to the laser plane $(\boldsymbol{u}, u_4)$ is $d = \boldsymbol{u} \cdot \boldsymbol{x}_1$, where it is assumed that $\boldsymbol{u}$ points away from camera frame 1. It follows that $(\boldsymbol{u} \cdot \boldsymbol{x}_1)/d = 1$, and insertion in (5) gives the relation

$$\boldsymbol{x}_2 = \left(\boldsymbol{R} + \frac{1}{d}\boldsymbol{t}\boldsymbol{u}^T\right) \boldsymbol{x}_1 \qquad (6)$$

for points on the laser plane, where $\boldsymbol{H} = \boldsymbol{R} + \frac{1}{d}\boldsymbol{t}\boldsymbol{u}^T$ is referred to as the planar homography induced by the laser plane.

The pixel coordinates $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ of the point in the scene are given by $\lambda_1\boldsymbol{p}_1 = \boldsymbol{K}_1\boldsymbol{x}_1$ and $\lambda_2\boldsymbol{p}_2 = \boldsymbol{K}_2\boldsymbol{x}_2$ where λ_1 and λ_2 are depth coordinates, and $\boldsymbol{K}_1$ and $\boldsymbol{K}_2$ are the camera parameter matrices. The planar

homography in terms of the pixel coordinates is then

$$\boldsymbol{p}_2 = \boldsymbol{K}_2 \left(\boldsymbol{R} + \frac{1}{d}\boldsymbol{t}\boldsymbol{u}^T\right) \boldsymbol{K}_1^{-1}\boldsymbol{p}_1 \qquad (7)$$

2.3 Reflection models

When a laser scanner is used on reflective surfaces of aluminium parts, there will be significant reflections. This section presents a characterization of such reflections. The laser light will give diffuse, spread, and specular reflections. Figure 2 shows how these reflections scatter in a 2D view, and how these reflections will appear for a sphere (Pharr et al., 2016). A perfectly diffuse surface will scatter light equally in all directions regardless of the incident direction of the light. A spread surface reflects light in a set of reflected directions, which results in the spread surface showing blurry reflections of other objects. A specular surface will reflect light in a single outgoing direction, which makes the material appear like a mirror.

Most real surfaces exhibit reflection that is a mixture of these types. Most objects we encounter daily are close to diffuse, such as fabric, plastic, and wood. Metals will primarily exhibit a reflection that is between spread and specular depending on the material and roughness of the surface.

2.4 Simulation of reflections using virtual scenes

A virtual scene can be generated with physics-based rendering techniques (Pharr et al., 2016), where a simulated image is generated. This can be used to generate simulated scenes for a laser scanner used on aluminium parts, where the reflections of the laser line are included. This will be used in this paper to train a CNN to distinguish between the true laser line, called the ground truth, and the reflected laser light. In physics-based rendering, the intensity values of the pixels are calculated by tracing the path of light, and by calculating how the light reflects from surfaces in the scene. The most computationally efficient approach is to trace the path of light backwards from the camera. This approach ensures that each path of light contributes to calculating the pixel values of the final image. The backward path tracing method encounters a problem when there is strong indirect light, known as caustics (Pharr et al., 2016). Caustics appear when the light from a light source is reflected from a specular surface and then by a diffuse surface before it arrives at the camera. The reflection of the light from the specular surface is referred to as an indirect light source. The method of backward path tracking relies on knowing the origin of the light sources in the scene. This will

not work for caustics, since the origins of indirect light sources are not known in advance before the rays have been traced. A possible solution is to trace the path of light from the light source through its reflections until it eventually hits the camera. The forward path tracing method is computationally expensive, since most light rays do not hit the camera. Bidirectional path tracing combines the advantage of both methods by tracing paths both from the camera and the light sources, which enables rendering scenes where caustic reflections are of importance.
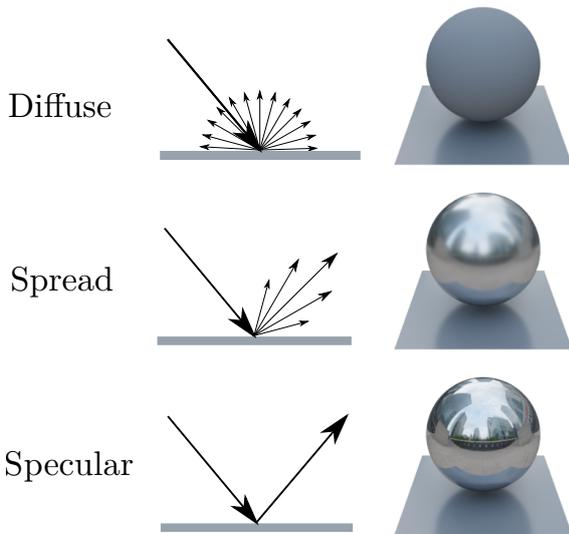
Figure 2: Basic reflection models. Diffuse reflections emit light equally in all directions, spread reflections emit in a set of directions, while specular reflections emit light in a single direction.

2.5 U-Net

In (Ronneberger et al., 2015), a fully convolutional neural network (CNN) for image segmentation is presented. U-Net was originally used for biomedical image segmentation, and has received considerable attention also for other use cases. The U-Net has a network architecture with a contracting path to capture the context of the input image, followed by an expanding path with skip-connections to retain spatial information as shown in Figure 3.

The contracting path follows a typical architecture for CNNs, with a repeated pattern of two convolutions, where each convolution is followed by a rectified linear unit and a max pool operator for downsampling. After each downsampling step, the number of feature channels is doubled. At the deepest layers of the U-Net

network, each spatial position will have a large receptive field and many feature channels to process complex patterns of the input image.

At each step of the expanding path, the feature maps are upscaled and a convolution that halves the number of feature channels is applied. The feature maps are then concatenated with the feature maps with corresponding number of feature maps in the contracting path, which makes up the skip-connections in the network. Two convolutions with rectified linear units are then applied to the concatenated feature maps before the process is repeated at the next step in the expanding path. The skip connections in the expanding path enable the network to maintain the finer-grained spatial information as the feature maps are upscaled.

3 Method

The method proposed in this paper is to train a CNN to eliminate reflections in laser scanning of aluminium parts by training the CNN on simulated scenes. The advantage of this approach is that a large number of aluminium parts with different geometry and reflective properties can be generated in this simulated setup, which gives extensive training data of the CNN. Moreover, the projection of the true laser line on the aluminium parts will be available as the ground truth, which makes it possible to do supervised training (Goodfellow et al., 2016). The simulated training data for the convolutional neural network was generated in the open-source 3D computer graphics software Blender (Roosendaal, 2021).

3.1 Bidirectional path tracer

Simulating the laser produces strong indirect light, which further produces reflections of the scanned part. To get an accurate simulation where caustic reflections are apparent, a bidirectional path tracer as discussed in Section 2.4 was needed. The built-in path tracer in Blender, Cycles, is purely a backward path tracer. The LuxCore renderer was installed as a plug-in, which supported bidirectional path tracing. This solution is used in (Grans and Tingelstad, 2021) where a detailed discussion is included. A comparison of the rendering of the laser line with Cycles and Luxcore is shown in Figure 4, which shows that caustics cause considerable reflections.

3.2 Mesh generation

To automate the process of creating meshes, Cad-Query (Urbaczyk et al., 2021) was used. CadQuery is a python module for building parametric 3D CAD
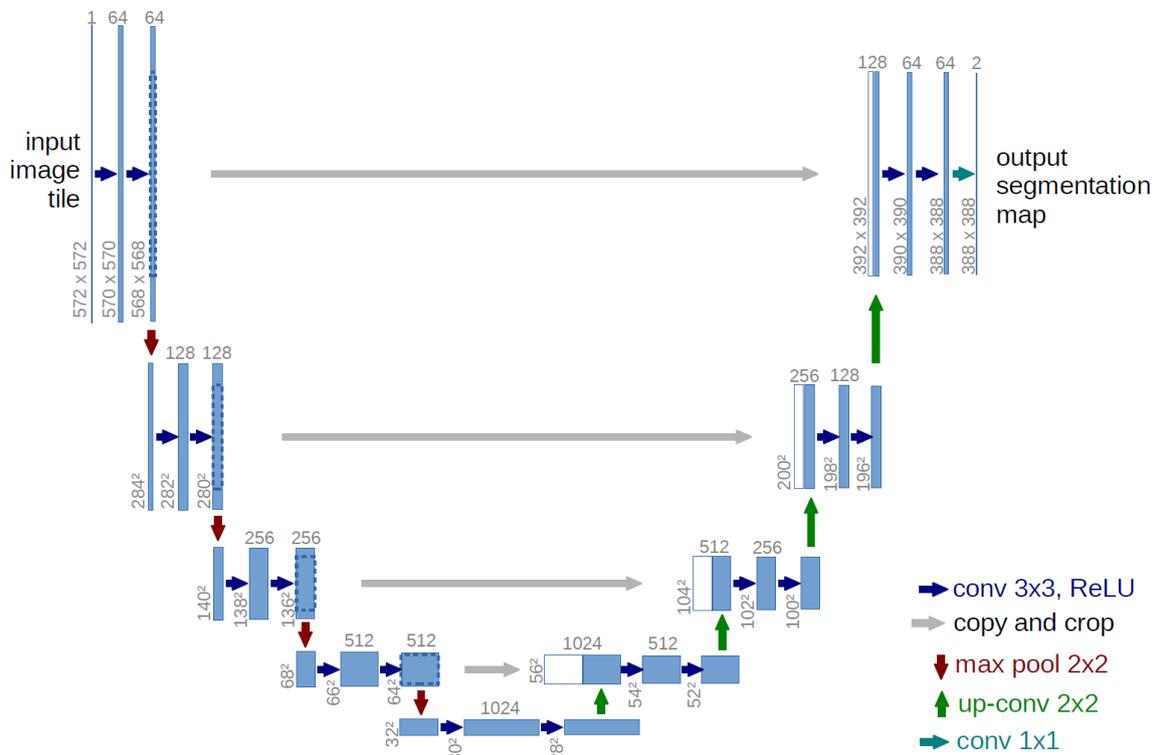
Figure 3: U-Net architecture (Ronneberger et al., 2015). The blue boxes show the feature maps with the number of channels denoted on top. The width and height of the feature maps used in the original paper are shown to the left of each blue box, however this depends on the input image size.
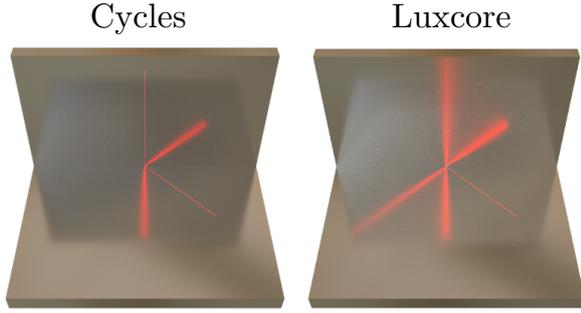
Cycles   Luxcore

Figure 4: Cycles and Luxcore comparison. The Luxcore rendering is able to simulate more realistic reflections.

models. To generate semi-random parts with geometry that resemble industrial parts such as aluminium extrusions, a set of 2D modules were defined as shown in Figure 5. Each module consists of random-length line segments which make up a predefined shape. A corner mesh consisting of five randomly sampled vertical modules, one corner module and five horizontal modules was generated to make up a 2D cross-section. The 2D cross-sections were then extruded at a fixed length to make up a mesh.
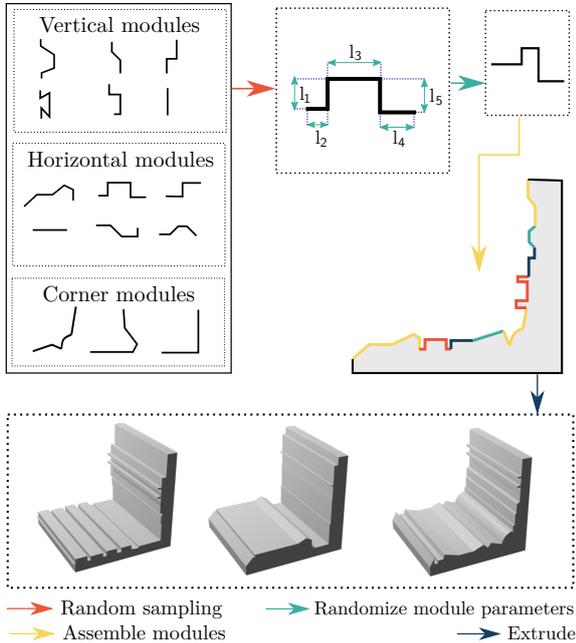
Figure 5: Mesh generation pipeline. A set of predefined modules are randomly sampled and the individual line lengths of the modules are randomized. One corner module, five vertical and five horizontal modules are assembled and extruded to a mesh.

3.3 Material generation

Two methods for assigning materials to the meshes were used. The first method assigned materials with physically based rendering (PBR) textures, which consists of images defining color, roughness, normals, and metalness as shown in Figure 6. The second method for material generation mixed Luxcore materials to make materials with different reflective properties. A combination of the predefined metal material and matte material were used with a mix factor m. The metal material's surface properties are defined by the roughness in two perpendicular directions u and v.
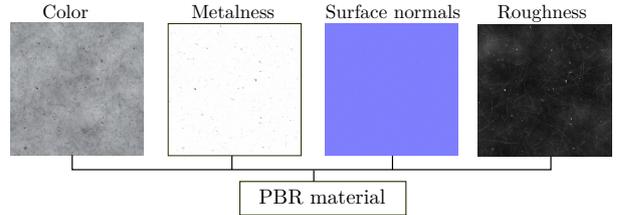
Figure 6: PBR materials require a set of images that defines properties such as color, metalness, surface normals, and microfacet roughness.
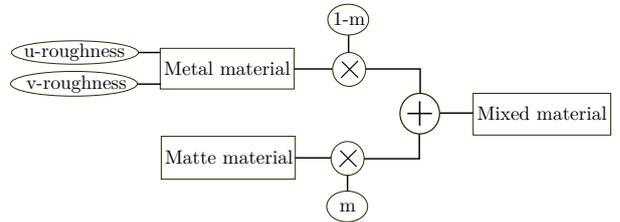
Figure 7: Mixed material generation. A Luxcore metal material is defined by the roughness in the perpendicular directions u and v. The metal material is mixed with a matte material with a mix factor m.

3.4 Laser scanner systems

A laser scanner system with one camera and a laser as shown in Figure 8 was implemented in Blender with the Luxcore plugin. The optical centers of the camera and laser are positioned at a length b apart with an angle θ towards each other.

A second laser scanner system with two cameras in a stereo arrangement and a laser was implemented with one camera on each side of the laser plane as shown in 9. The optical centers of the cameras were positioned with a distance b from the laser, and angled towards the laser at an angle θ.
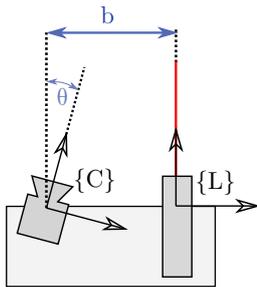
Figure 8: Standard laser scanner
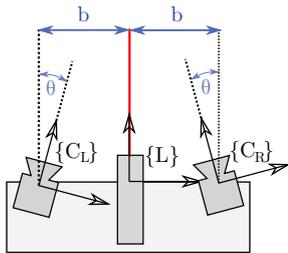
Figure 9: Stereo laser scanner

3.5 Planar homography preprocessing

Consider the stereo laser scanner described in Section 3.4. To enable the U-Net model to process the information from both views, the filtered laser scan images from each camera were combined using a planar homography. The planar homography for pixel coordinates, as given in (Hartley and Zisserman, 2003), from the right to the left camera is

$$\boldsymbol{H} = \boldsymbol{K}_L(\boldsymbol{R} - \boldsymbol{t}\boldsymbol{u}^T/d)\boldsymbol{K}_R^{-1}. \tag{8}$$

where $\boldsymbol{K}_L$ and $\boldsymbol{K}_R$ are the camera matrices of the left and right cameras respectively. $\boldsymbol{R}$ is the rotation matrix and $\boldsymbol{t}$ is the translation vector, both from the left to the right camera. $\boldsymbol{u}$ is the unit normal vector of the laser plane given in the frame of the left camera and d is the distance from the optical center of the left camera to the closest point on the laser plane. Given that the geometrical relationship between the cameras and laser is fixed during the scanning, the planar homography is fixed and can be computed prior to the scanning. The preprocessing pipeline is shown in Figure 10 for the scanning of a reflective part. The method requires a way to filter the laser line from the background as the initial step. The right view is projected with the planar homography as calculated in (8). The filtered left view and projected view are combined to a single image, by assigning the left view to the red channel and the projected view to the green channel of an RGB image, which makes yellow pixels where the two images overlap.

The idea behind using two cameras and the given

pipeline is to use the constraint that the true points must lie on the laser plane at any given time, similar to the constraints used in (Trucco et al., 1994) and (Li et al., 2019). The pixels that are not overlapping in the final preprocessed image in Figure 10 are not consistent with both views, because they do not correspond to the same 3D points. The measurements which are not overlapping are either occluded from one of the views or are not located on the laser plane and are therefore false measurements. Previous papers calculate the 3D points of all measurements before erroneous points are removed. The planar homography enables evaluating the laser plain constraint, line intensity, and shape simultaneously from both views in 2D image space, for example by a convolutional neural network, before they are evaluated to be true or false measurements.

3.6 Evaluation metrics

Three metrics were selected to evaluate the ability of U-Net to filter out the reflections and correctly predict the true laser lines. The first metric was the Srensen-Dice coefficient, which was used as a metric for semantic image segmentation problems in (Zou et al., 2004). The Srensen-Dice coefficient measures the similarity between two sets and is given by

$$DSC = \frac{2\,TP}{FN + FP + 2\,TP} \tag{9}$$

where TP is the number of true positives, FN is the number of false negatives and FP is the number of false positives. True positives are pixels where U-Net correctly predicted the laser line, false positives are incorrect positive pixel predictions and false negatives are incorrect negative pixel predictions.

The second metric measures the ability of the generated output segmentation to find the subpixel center of the line. To generate an estimate of the subpixel center of the line, the center of gravity (CoG) algorithm described in (Fisher and Naidu, 1996) was used. Only measurements that were within a predefined threshold pixels of the actual line were included in the calculation, so that very large outlier errors would not skew the results. The CoG was calculated for each row where the U-Net model had made a prediction in the output segmentation of the laser line. The mean subpixel accuracy was calculated by averaging the error for each of the rows in each predicted segmentation image.

Finally, the third metric is referred to as the outlier fraction. Since large outliers were discarded in the subpixel accuracy metric, this third metric is calculated from the number of outliers for the rows of each segmentation. The same threshold of pixels was used as in second metric. The metric is calculated as a frac-
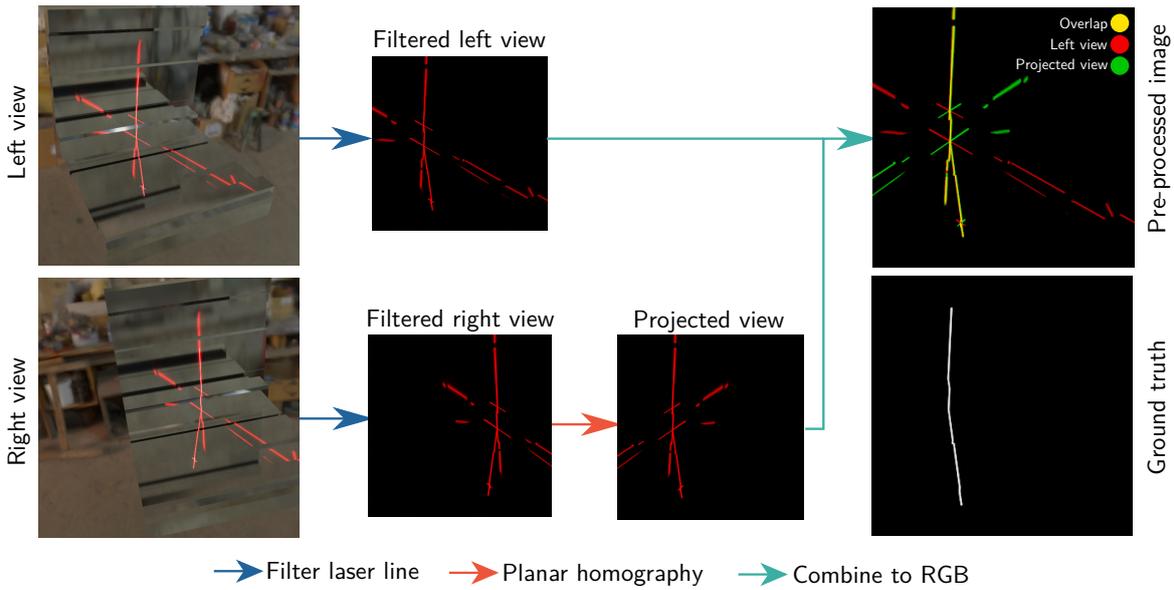
Figure 10: The planar homography preprocessing requires a filtered view of the laser from two cameras. One of the filtered views is then projected using a planar homography, such that when the two views are combined, the true laser line will overlap while many spurious reflections will not.

Figure 11: Types of reflection in test sets. The three test sets consist of sharp specular reflections, blurry reflections, and a variety of reflections produced by PBR materials.

tion, such that the number of outliers of all the rows is divided by the total amount of rows.

4 Experiments

4.1 Dataset generation

Datasets were generated to train, validate and evaluate the performance of the U-Net models. 4300 unique corner meshes were generated according to the method described in Section 3.2. The meshes were loaded into Blender and assigned a material. The material assigned was either a PBR texture picked randomly among 40 metal textures downloaded from (Demes, 2021), or assigned based on the mixed material described in Section 3.3. The training and validation set used a mixture of all types of materials, while the test

sets were split into three based on the type of material. The three test sets consisted of sharp specular reflections, blurry reflections and a variety of reflections generated by PBR materials. An example from each of the test sets is shown in Figure 11. Table 1 summarizes the dataset lengths, and which types of material and parameters were used to generate each dataset. The range for mix factor m, u and v-roughness for the material described in Figure 7 was randomized in $\log_{10}$. The last column in the table states how many percent of each dataset consisted of PBR materials, and how much that was generated from mixed materials. Both scanner configurations in Figure 8 and Figure 9, were set up with baseline $b = 0.2\,\text{m}$ and $\theta = \pi/20$. The datasets were generated by rendering images with resolution 1024 by 1024 pixels with Luxcore renderer, for the two laser scanner systems in Section 3.4. For each scan, an associated ground truth scan was generated by disabling reflections of second-order and higher.

4.2 U-Net model and training

The U-Net model was implemented in Pytorch with architecture as shown in Figure 3. A modification of the original implementation in (Ronneberger et al., 2015) was that instead of the original *valid convolutions*, padding was introduced to have *same convolutions*, which ensured that the output segmentation had the same resolution as the input image. A cross-entropy loss function was used, and it was found that weighting in the loss function was important for sta-

Table 1: Dataset lengths, materials and parameters

Dataset	Dataset length	Mix factor m		u,v-roughness		% PBR materials
		min.	max.	min.	max.	
Training set	3600	0.001	0.003	0.001	0.2	50%
Validation set	400	0.001	0.003	0.001	0.2	50%
Test specular	100	0.0015	0.002	0.001	0.008	0%
Test blurry	100	0.0015	0.002	0.05	0.2	0%
Test PBR	100	-	-	-	-	100%

ble training of the model. This weighting was implemented by weighting incorrect predictions of the laser line nine times higher than incorrect predictions of the background class in the loss function. The U-Net models were trained with a batch size of 2, and all the batches of the training dataset were used in one epoch. Training was run for 10 epochs on the training dataset. The resulting computational time was approximately 8 hours with an RTX 3090 GPU.

4.3 Evalutation

The trained model was used to predict the true laser line segmentation on each of the three test sets, where each of the three test sets has different reflection properties as described in Section 4.1. For each of the test sets, the evaluation metrics described in Section 3.6 were used. The threshold for the mean subpixel accuracy and outlier fraction was chosen to be 5 pixels.

5 Results and discussion

For the evaluation metrics presented in Section 3.6, the numeric results for the single and stereo camera systems are shown in Table 2 for each of the test datasets. Both methods resulted in a Srensen-Dice coefficient of about 0.98 for the three test sets. The mean subpixel error was also roughly the same at about 0.13 pixels error for the three test sets for both methods. The stereo camera method performed better on the outlier fraction with the specular test set, while the single camera method performed better on the blurry test set. The outlier fraction for the PBR test set was similar for both methods.

Inspecting individual image predictions compared with the ground truth revealed several sources of error. One type of error which was apparent in all the segmentation predictions was imperfect edge detection of the scan line as shown in Figure 13. All prediction images had between 1-2 % error in the Srensen-Dice coefficient, even if it was a seemingly perfect prediction otherwise. Another source of error was incorrect predictions of sharp specular reflections made by the U-Net with the standard laser scanner. Sharp specular reflections may appear similar to the true scan line, such that the U-Net model had trouble differentiating between these. The stereo laser scanner was better at differentiating sharp reflections since the preprocessed image often showed that they did not overlap from both views. A comparison of both scanners highlighting this issue is shown in Figure 15. The third source of error that appears in both systems, were strong blurry reflections that overlapped the true scan line. An example is shown in Figure 14, where the true scan line is not distinctly visible due to strong reflections.

6 Conclusion and future work

A machine learning approach has been presented for filtering reflections in laser line scanning of aluminium corner workpieces. A U-Net model was used to filter the spurious reflections, and to create a segmentation of the true line. A pipeline for simulating laser scan measurements on reflective aluminium parts is presented. Randomized corner meshes were generated, given a metal material, and simulated with the laser scanners in Blender. The simulated images were used for training and testing the U-Net model. The CNN was trained with two different camera configurations for the scanner. The first configuration uses a single camera, while the second uses two cameras in a stereo configuration. A preprocessing method was proposed for the stereo laser scanner. Both systems were compared against 3 test sets with distinct types of reflections. Both methods achieved a Srensen-Dice coefficient of over 0.97 across all test sets. For future work, the presented approach will be tested on experimental setups. Using the trained network from simulated data, transfer learning can be used to adapt the network to real data. If a dice score of 0.97 can be achieved on real data, it is deemed to be sufficient for being able to weld aluminium workpieces.

Table 2: Numeric results

Dataset	Srensen-Dice coeff.		Mean subpixel error		Outlier fraction	
	Single camera	Stereo camera	Single camera	Stereo camera	Single camera	Stereo camera
Test specular	0.982	0.986	0.140	0.115	0.202%	0.026%
Test blurry	0.977	0.976	0.134	0.141	0.046%	0.331%
Test PBR	0.979	0.983	0.129	0.123	0.045%	0.038%
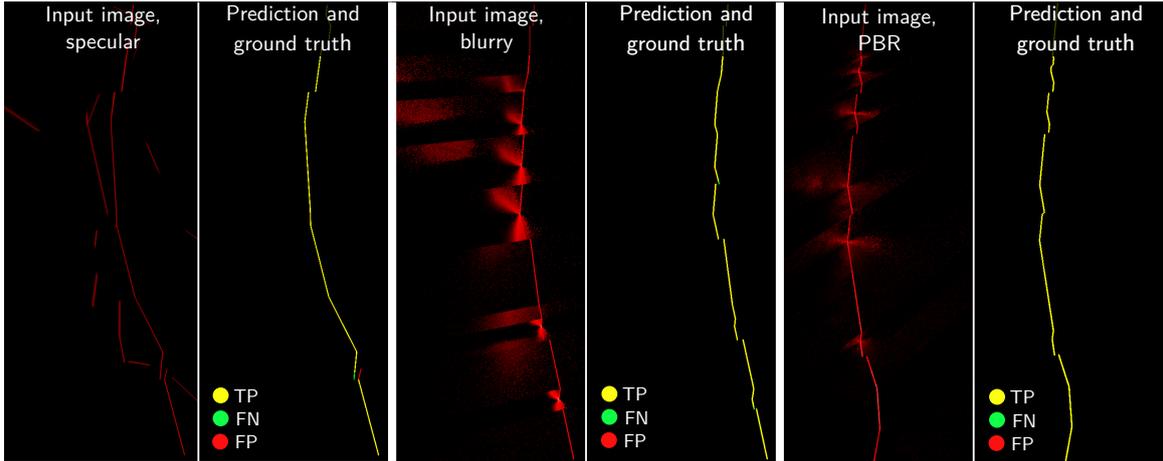
(a) Single camera laser scanner results.
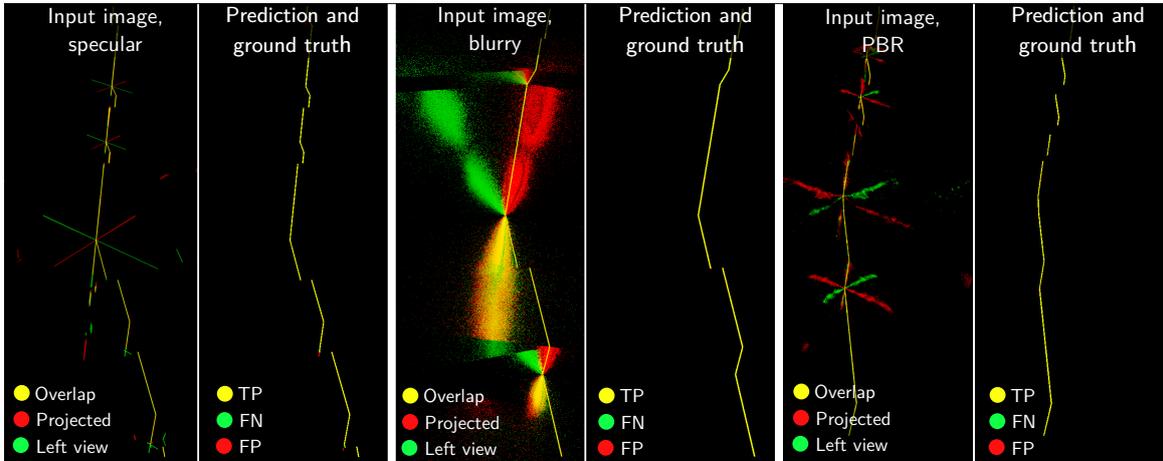
(b) Stereo laser scanner results

Figure 12: U-Net segmentation inputs, predictions, and ground truth. Each of the methods shows three pairs of input and predictions compared with ground truth. For each input-prediction pair, the input image is shown to the left, while the prediction compared with the ground truth is shown to the right. For the prediction-ground truth images, correct predictions are shown in yellow, while false positives are shown in red and false negatives in green.

Acknowledgments
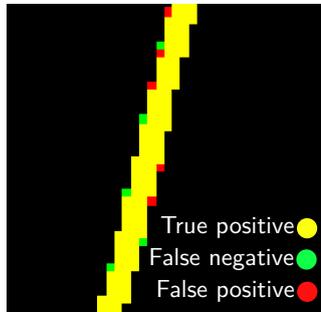
Figure 13: Imperfect edge detection. The U-Net prediction of the laser line is often incorrect at the edges of the laser line.

Standard laser scanner

Compare ground truth and prediction

True positive ●
False negative ●
False positive ●

U-Net prediction

Stereo laser scanner

Overlap ●
Left view ●
Right view ●

Compare ground truth and prediction

True positive ●
False negative ●
False positive ●

U-Net prediction

Figure 15: Specular reflection ambiguity. Sharp specular reflections may have a similar appearance as the true laser line. In certain cases, the single camera scanner incorrectly chooses the false reflection, while the stereo camera scanner chooses the true line since they are overlapping from both cameras.

● True positive
● False negative
● False positive

U-Net prediction

Figure 14: Strong blurry reflections may overlap the true scan line, such that it is not visible among the strong reflections.

References

Blais, F. Review of 20 years of range sensor development. *Journal of electronic imaging*, 2004. 13(1):231–243. doi:10.1117/1.1631921.

Clark, J., Trucco, E., and Wolff, L. B. Using light polarization in laser scanning. *Image and Vision Computing*, 1997. 15(2):107–117. doi:10.1016/S0262-8856(96)01126-2.

Demes, L. Free public domain pbr materials. 2021. URL https://ambientcg.com/.

Fisher, R. and Naidu, D. A comparison of algorithms for subpixel peak detection. In *Image technology*, pages 385–404. Springer, 1996. doi:10.1007/978-3-642-58288-2_15.

Frank Chen, G. M. and Mumin, S. Overview of three-dimensional shape measurement using optical methods. *Optical Engineering*, 2000. 39:10–21. doi:10.1117/1.602438.

Goodfellow, I. J., Bengio, Y., and Courville, A. *Deep Learning.* MIT Press, Cambridge, MA, USA, 2016. http://www.deeplearningbook.org.

Grans, S. and Tingelstad, L. Blazer: Laser scanning simulation using physically based rendering. *arXiv preprint arXiv:2104.05430*, 2021.

Hartley, R. and Zisserman, A. *Multiple View Geometry in Computer Vision.* Cambridge University Press, New York, NY, USA, 2 edition, 2003. doi:10.1017/CBO9780511811685.

Köhler, J., Nöll, T., Reis, G., and Stricker, D. Robust outlier removal from point clouds acquired with structured light. In *Eurographics (Short Papers).* pages 21–24, 2012. doi:10.2312/conf/EG2012/short/021-024.

Li, H., Zhang, X., Zhuang, L., and Yang, Y. Specular surface measurement with laser plane constraint to reduce erroneous points. In *International Conference on Intelligent Robotics and Applications.* Springer, pages 53–63, 2019. doi:10.1007/978-3-030-27541-9_5.

Lynch, K. and Park, F. *Modern Robotics: Mechanics, Planning, and Control.* Cambridge Univeristy Press, 2017.

Pharr, M., Jakob, W., and Humphreys, G. *Physically based rendering: From theory to implementation.* Morgan Kaufmann, 2016. doi:10.1016/C2013-0-15557-2.

Pottmann, H. and Wallner, J. *Computational Line Geometry.* Springer-Verlag, Berlin, 2001. doi:10.1007/978-3-642-04018-4.

Ronneberger, O., Fischer, P., and Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention.* Springer, pages 234–241, 2015. doi:10.1007/978-3-319-24574-4_28.

Roosendaal, T. *Blender - a 3D modelling and rendering package.* Blender Foundation, Stichting Blender Foundation, Amsterdam, 2021. URL http://www.blender.org.

Trucco, E., Fisher, R. B., and Fitzgibbon, A. W. *Direct calibration and data consistency in 3-D laser scanning.* Department of Artificial Intelligence, University of Edinburgh, 1994. doi:10.5244/C.8.48.

Trucco, E., Fisher, R. B., Fitzgibbon, A. W., and Naidu, D. Calibration, data consistency and model acquisition with laser stripers. *International Journal of Computer Integrated Manufacturing*, 1998. 11(4):293–310. doi:10.1080/095119298130642.

Urbaczyk, A., Wright, J., Cowden, D., Solutions, I. T., ZDERYA, H. Y., Boyd, M., Agostini, B., Greminger, M., Buchanan, J., cactrot, huskier, de Len Peque, M. S., Boin, P., Saville, W., Weissinger, B., Ruben, nopria, Osterwood, C., moeb, Kono, A., HLevering, Turner, W., Trho, A., Christoforo, G., just georgeb, Peterson, A., Grunichev, A., Gregg-Smith, A., Bernhard, and Anderson, D. Cadquery/cadquery: Cadquery 2.1. 2021. doi:10.5281/zenodo.4498634.

Zou, K. H., Warfield, S. K., Bharatha, A., Tempany, C. M., Kaus, M. R., Haker, S. J., Wells III, W. M., Jolesz, F. A., and Kikinis, R. Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports. *Academic radiology*, 2004. 11(2):178–189. doi:10.1016/s1076-6332(03)00671-8.