

Inverse rendering for industry inspections

- Comparing traditional and Artificial intelligence methods with inverse rendering to detect defects to components in industry settings.

Invers rendering för industriella inspektioner.

Jonatan Ebenholm

Supervisor : Nithesh Chandler

Examiner : Ehsan Miandji

Upphovsrätt

Detta dokument hålls tillgängligt på Internet - eller dess framtida ersättare - under 25 år från publiceringsdatum under förutsättning att inga extraordinära omständigheter uppstår.

Tillgång till dokumentet innebär tillstånd för var och en att läsa, ladda ner, skriva ut enstaka kopior för enskilt bruk och att använda det oförändrat för ickekommersiell forskning och för undervisning. Överföring av upphovsrätten vid en senare tidpunkt kan inte upphäva detta tillstånd. All annan användning av dokumentet kräver upphovsmannens medgivande. För att garantera äktheten, säkerheten och tillgängligheten finns lösningar av teknisk och administrativ art.

Upphovsmannens ideella rätt innefattar rätt att bli nämnd som upphovsman i den omfattning som god sed kräver vid användning av dokumentet på ovan beskrivna sätt samt skydd mot att dokumentet ändras eller presenteras i sådan form eller i sådant sammanhang som är kränkande för upphovsmannens litterära eller konstnärliga anseende eller egenart.

För ytterligare information om Linköping University Electronic Press se förlagets hemsida <http://www.ep.liu.se/>.

Copyright

The publishers will keep this document online on the Internet - or its possible replacement - for a period of 25 years starting from the date of publication barring exceptional circumstances.

The online availability of the document implies permanent permission for anyone to read, to download, or to print out single copies for his/hers own use and to use it unchanged for non-commercial research and educational purpose. Subsequent transfers of copyright cannot revoke this permission. All other uses of the document are conditional upon the consent of the copyright owner. The publisher has taken technical and administrative measures to assure authenticity, security and accessibility.

According to intellectual property law the author has the right to be mentioned when his/her work is accessed as described above and to be protected against infringement.

For additional information about the Linköping University Electronic Press and its procedures for publication and for assurance of document integrity, please refer to its www home page: <http://www.ep.liu.se/>.

Abstract

The purpose of this thesis is to examine if Spatially varying Bidirectional Reflectance Distribution Functions (SVBRDF) maps generated through inverse rendering can be used in a meaningful way to improve detection of surface defects of a material compared to using static image-space representations. Three research questions were asked, and later answered. These questions are the following:

1. Does representing surfaces in SVBRDF space improve the detectability of small topological surface defects compared to direct image-space analysis?
2. Are there combinations of SVBRDF components (e.g., surface normals, roughness) that increase the accuracy of scratch detection compared to using the full SVBRDF representation or individual components alone?
3. Can geometry-related components of SVBRDF representations, such as surface normals, be reused across different materials with similar physical surface structure for scratch detection?

To answer these questions, both traditional scratch detection, and a proprietary AI anomaly detection system were tested. Synthetic data with ground truth scratch masks was created using Blender to analytically validate the accuracy of the defect detection using F1- and IoU-score, as well as connected-component labeling, for both the SVBRDF maps and image-space inputs.

Real data was also collected, but no defect annotation was made, so the results of the real data were only used with visual inspections which discovered a domain gap with the inverse rendering model.

The results showed that using SVBRDF maps created from inverse rendering did not improve the defect detections compared to image-space inputs when using a single SVBRDF map type, like normal maps, and not for linearly combined maps either. However, results did show that using the generated normal maps during training did improve the AI anomaly detection tool's ability to find defects on different materials with similar physical structure.

Acknowledgments

I would like to thank my supervisor, Nithesh Chandler, for his guidance and feedback throughout this project. I would also like to thank my examiner, Ehsan Miandji, for his valuable insights, as well as the conversations about topics not relating to the project. Furthermore, I would also like to thank Jens Edhammer for the supervision at location, helping me find better and often simpler solutions, and for coming up with the thesis idea in the first place. I would further like to thank Rikard Andersson for the help with setting up the rig which the real data was collected on, as well as Mathilda Larsson for assisting with the proprietary software. And while I have been asked not to mention the company name in the thesis, I would like to thank the company for providing the opportunity to work on this project, and for providing the resources and support needed to complete it.

I would also like to thank my family and friends for their support and encouragement throughout this process. Additionally, I would like to thank anyone else who has contributed to this project in any way, whether it be through providing feedback, resources, or simply moral support.

Finally, I want to thank my wonderful girlfriend Emilin, who made my everyday life much better and manageable during the thesis work, and who has been a constant source of love and support throughout this project.

Contents

Abstract	iii
Acknowledgments	iv
Contents	v
List of Figures	vii
1 Introduction	1
1.1 Background	1
1.2 Motivation	2
1.3 Aim	2
1.4 Research questions	3
1.5 Delimitations	3
2 Theory and Related Work	4
2.1 Theory	4
2.2 Related work	8
3 Methodology	13
3.1 SVBRDF map generation	14
3.2 Image quality metrics	14
3.3 Data	15
3.4 Classification metrics	17
3.5 Anomaly detection through traditional methods	18
3.6 Anomaly detection through deep learning methods	18
3.7 Evaluation Strategy	19
4 Results	22
4.1 Image Quality Metric	22
4.2 SVBRDF maps	23
4.3 Edge detection using traditional methods	26
4.4 AI anomaly detection	27
4.5 Other results	35
5 Discussion	40
5.1 Results	40
5.2 Limitations	43
5.3 The work in a wider context	45
6 Conclusion	46
6.1 Summarization of purpose	46
6.2 Answering the research questions	46

6.3 Further work	47
Bibliography	48
A Graphs from results	50
A.1 Defect detection using traditional methods	50
A.2 Defect detection using AI anomaly detection	56

List of Figures

2.1	Visualization of the hysteresis thresholding step in Canny edge detection, taken from [web:cannyEdge], where any value over maxVal is considered a strong edge, and any value in between minVal and maxVal is considered a weak edge. As seen in the image, the upper line has values that are both weak and strong, but since they are connected to strong edges, they are included in the final edge map. The lower line has no values that are considered strong, so they are excluded from the final edge map.	8
3.1	This figure shows the overall pipeline for the methodology in this thesis. The first step was data collection, where Blender was used for the synthetic data. The second part was Inverse rendering as a preprocessing step. The final step was using the generated SVBRDF maps in anomaly detection, where the results to answer the research questions were gathered.	13
3.2	The rig which was used to collect real data.	16
3.3	Far away image of metal object.	17
4.1	Comparison of material quality metrics for different SVBRDF estimation methods and their corresponding type of map they were used on. Each method horizontal axis represents the different materials that were rendered and used as input data for the NFPLight inverse rendering method. The vertical axis represents the score, or quality, of each material produced using NFPLight inverse rendering when compared to ground truth. Each material has a corresponding line associated with it, which shows the range of quality score for each material under different rotation and translation to the materials UV coordinates.	22
4.2	Heatmap visualization of material quality metrics for different SVBRDF estimation methods. Each method horizontal axis represents if there were defects present in the material (such as scratches), or if the material was defect free. The vertical axis represents the materials that were tested on with NFPLight inverse rendering. The number and corresponding color represents the quality score of each material when compared to ground truth, where higher scores and greener colors indicate better quality.	23
4.3	SVBRDF maps generated by NFPLight from a synthetic Wood63 sample. The near-field (a) and far-field (b) renders are the inputs, and NFPLight estimates the diffuse (c), normal (d), roughness (e), and specular (f) maps. The scratch is encoded mainly in the normal map.	24
4.4	SVBRDF maps generated by NFPLight from a real captured sample, with the same layout as Figure 4.3. The estimated maps are visibly poorer than for the synthetic data, reflecting the domain gap discussed in Section 5.2.	25

4.5	Graph showing the results of using Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection methods without Sobel filtering as a preprocessing step, with the ground truth normal SVBRDF map. Detection rate counts the percentage of scratches with at least 70% coverage from the detection method, F1 score relates to the method explained in the method Section 2.2.4, and lastly IoU refers to another method explained in the method Section 2.2.4. The subject had defects.	26
4.6	Graph showing the results of using Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection methods without Sobel filtering as a preprocessing step, with the normal SVBRDF map created with NFPLight inverse rendering. Detection rate counts the percentage of scratches with at least 70% coverage from the detection method, F1 score relates to the method explained in the method Section 2.2.4, and lastly IoU refers to another method explained in the method Section 2.2.4. The subject had defects.	27
4.7	Graph showing the results of using Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection methods without Sobel filtering as a preprocessing step, with the camera positioned near the subject. Detection rate counts the percentage of scratches with at least 70% coverage from the detection method, F1 score relates to the method explained in the method Section 2.2.4, and lastly IoU refers to another method explained in the method Section 2.2.4. The subject had defects.	27
4.8	Near-field image-space inputs used for training and testing (ground-truth comparison).	28
4.9	Ground truth normal map inputs used for training and testing.	29
4.10	Near-field image-space inputs on the deep-scratch test set.	30
4.11	Near-field image-space inputs on the shallow-scratch test set.	31
4.12	Some of the results from combining different SVBRDF representations heatmaps.	32
4.13	Some of the results from combining different SVBRDF representations heatmaps with less aggressive scratches in the testing data.	33
4.14	Cross-material transferability tests (RQ3): anomaly-detection metrics for near-field image inputs (left column) and normal-map inputs (right column) under three appearance changes, namely a similar material, a visually dissimilar material of the same category, and a diffuse-map swap.	34
4.15	Roughness maps generated by NFPLight from a defective (a) and a non-defective (b) synthetic material.	35
4.16	Result of Otsu thresholding on a normal map created by NFPLight.	36
4.17	Result of Canny edge detection on a normal map created by NFPLight.	37
4.18	Images showing thresholded heatmaps of different inputs into the anomaly detection system, and their intersection with ground truth.	38
4.19	Images showcasing shadows created with thicker subjects, where the bottom of the subject casts a shadow onto the background.	39
A.1	Graph showing the results of using Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection methods without Sobel filtering as a preprocessing step, with the ground truth normal SVBRDF map. Detection rate counts the percentage of scratches with at least 70% coverage from the detection method, F1 score relates to the method explained in the method Section 2.2.4, and lastly IoU refers to another method explained in the method Section 2.2.4. The subject had defects.	50

A.2	Graph showing the results of using Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection methods without Sobel filtering as a preprocessing step, with the camera positioned near from the subject. Detection rate counts the percentage of scratches with at least 70% coverage from the detection method, F1 score relates to the method explained in the method Section 2.2.4, and lastly IoU refers to another method explained in the method Section 2.2.4. The subject had defects.	51
A.3	Graph showing the results of using Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection methods without Sobel filtering as a preprocessing step, with the camera positioned far from the subject. Detection rate counts the percentage of scratches with at least 70% coverage from the detection method, F1 score relates to the method explained in the method Section 2.2.4, and lastly IoU refers to another method explained in the method Section 2.2.4. The subject had defects.	51
A.4	Graph showing the results of using Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection methods without Sobel filtering as a preprocessing step, with the normal SVBRDF map created with NFPLight inverse rendering. Detection rate counts the percentage of scratches with at least 70% coverage from the detection method, F1 score relates to the method explained in the method Section 2.2.4, and lastly IoU refers to another method explained in the method Section 2.2.4. The subject had defects.	51
A.5	Graph showing the results of using Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection methods without Sobel filtering as a preprocessing step, with the ground truth normal SVBRDF map. FP rate refers to the number of false positive pixels recorded as compared to true negatives, Accuracy is the inverse of FP rate. The subject had no defects, so no true positives to record.	52
A.6	Graph showing the results of using Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection methods without Sobel filtering as a preprocessing step, with the camera positioned near from the subject. FP rate refers to the number of false positive pixels recorded as compared to true negatives, Accuracy is the inverse of FP rate. The subject had no defects, so no true positives to record.	52
A.7	Graph showing the results of using Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection methods without Sobel filtering as a preprocessing step, with the camera positioned far from the subject. FP rate refers to the number of false positive pixels recorded as compared to true negatives, Accuracy is the inverse of FP rate. The subject had no defects, so no true positives to record.	52
A.8	Graph showing the results of using Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection methods without Sobel filtering as a preprocessing step, with the normal SVBRDF map created with NFPLight inverse rendering. FP rate refers to the number of false positive pixels recorded as compared to true negatives, Accuracy is the inverse of FP rate. The subject had no defects, so no true positives to record.	53
A.9	Graph showing the results of using Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection methods with Sobel filtering as a preprocessing step, with the camera positioned near from the subject. Detection rate counts the percentage of scratches with at least 70% coverage from the detection method, F1 score relates to the method explained in the method Section 2.2.4, and lastly IoU refers to another method explained in the method Section 2.2.4. The subject had defects.	53

A.10	Graph showing the results of using Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection methods with Sobel filtering as a preprocessing step, with the camera positioned far from the subject. Detection rate counts the percentage of scratches with at least 70% coverage from the detection method, F1 score relates to the method explained in the method Section 2.2.4, and lastly IoU refers to another method explained in the method Section 2.2.4. The subject had defects.	53
A.11	Graph showing the results of using Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection methods with Sobel filtering as a preprocessing step, with the normal SVBRDF map created with NFPLight inverse rendering. Detection rate counts the percentage of scratches with at least 70% coverage from the detection method, F1 score relates to the method explained in the method Section 2.2.4, and lastly IoU refers to another method explained in the method Section 2.2.4. The subject had defects.	54
A.12	Graph showing the results of using Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection methods with Sobel filtering as a preprocessing step, with the camera positioned near from the subject. FP rate refers to the number of false positive pixels recorded as compared to true negatives, Accuracy is the inverse of FP rate. The subject had no defects, so no true positives to record.	54
A.13	Graph showing the results of using Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection methods with Sobel filtering as a preprocessing step, with the camera positioned far from the subject. FP rate refers to the number of false positive pixels recorded as compared to true negatives, Accuracy is the inverse of FP rate. The subject had no defects, so no true positives to record.	54
A.14	Graph showing the results of using Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection methods with Sobel filtering as a preprocessing step, with the normal SVBRDF map created with NFPLight inverse rendering. FP rate refers to the number of false positive pixels recorded as compared to true negatives, Accuracy is the inverse of FP rate. The subject had no defects, so no true positives to record.	55
A.15	Anomaly-detection metrics for each individual input representation on the deep-scratch test set: the diffuse (a), normal (b), roughness (c), and specular (d) SVBRDF maps, and the near-field image-space input (e). Each representation was used for both training and testing.	56
A.16	Anomaly-detection metrics for each individual input representation on the shallow-scratch test set: the diffuse (a), normal (b), roughness (c), and specular (d) SVBRDF maps, and the near-field image-space input (e). Each representation was used for both training and testing.	57
A.17	Two-map combinations of SVBRDF maps on the deep-scratch test set. Each combination was linearly combined and thresholded into a single mask, then compared to the ground-truth scratch mask.	58
A.18	Three- and four-map combinations of SVBRDF maps on the deep-scratch test set. Each combination was linearly combined and thresholded into a single mask, then compared to the ground-truth scratch mask.	59
A.19	Two-map combinations of SVBRDF maps on the shallow-scratch test set. Each combination was linearly combined and thresholded into a single mask, then compared to the ground-truth scratch mask.	60
A.20	Three- and four-map combinations of SVBRDF maps on the shallow-scratch test set. Each combination was linearly combined and thresholded into a single mask, then compared to the ground-truth scratch mask.	61

A decorative element consisting of several thin, vertical black lines of varying heights, located to the left of the chapter title.

1 Introduction

In modern manufacturing, production has shifted from manual craftsmanship to highly automated, high-volume industrial processes. As production lines increase in speed and autonomy, the demand for reliable and consistent quality control becomes critical. One of the most persistent challenges in this context is the visual inspection of surface defects, such as scratches, dents, or irregularities, which may be subtle, appearance-dependent, and highly sensitive to illumination.

Traditional industrial inspection systems largely rely on two-dimensional image-space analysis, using techniques such as edge detection, thresholding, or blob analysis. While effective for detecting gross defects, these approaches often struggle with subtle defects, particularly small surface scratches, whose visibility changes with lighting, viewing direction, or material finish. A shallow scratch on a reflective metal surface, for example, may be nearly invisible under one illumination condition while becoming clearly visible under another. This variability makes purely image-based inspection unreliable.

To address these limitations, this thesis explores inverse rendering as an alternative method for industrial inspection. Rather than operating directly on pixel intensities, inverse rendering aims to decompose an observed image into intrinsic material and geometric properties, such as surface normals and reflectance parameters. By separating geometry from appearance, inverse rendering offers a way to reason about surface structure independently of illumination.

1.1 Background

Inverse rendering is the process of estimating the intrinsic properties of a scene, most notably geometry, material, and lighting, from one or more two-dimensional images. It can be understood as the inverse of forward rendering, where the interaction between light and material is simulated to produce an image. In the inverse setting, the goal is to recover the material representation that best explains the observed appearance [12].

In practice, surface appearance is often modeled using Spatially-Varying Bidirectional Reflectance Distribution Functions (SVBRDFs), which describe how reflectance parameters vary across a surface. Typical SVBRDF representations include maps for albedo, surface normals, and roughness [3, 1]. Some of these components, particularly surface normals, encode geometric information that is largely independent of the underlying material’s color or reflectance.

This decomposition is well suited to industrial inspection. While appearance may vary significantly across materials, coatings, or finishes, the underlying surface geometry of a manufactured component is often tightly controlled. This raises an important question: if geometry can be reliably recovered, can it be used as a stable basis for detecting anomalies even when material appearance changes?

In this thesis, particular focus is placed on small geometric defects such as scratches, which are primarily encoded in surface geometry rather than reflectance.

1.2 Motivation

Recent advances in deep learning have made it possible to estimate SVBRDF representations from a single image [4, 6], removing the need for complex multi-shot capture setups traditionally required by physics-based methods such as photometric stereo [15]. These single-shot neural approaches offer clear practical advantages in industrial environments, where controlled multi-view or multi-illumination capture is often complex and costly.

However, the adoption of such methods for inspection raises several open questions. First, it is unclear whether operating in SVBRDF space actually improves scratch detectability compared to conventional image-space analysis. Second, SVBRDF representations contain multiple components, not all of which may be equally informative for scratch detection. Finally, while material appearance can vary widely between components (e.g., steel, copper, painted steel), the underlying surface geometry may remain identical, suggesting the possibility of reusing geometry-based representations across materials.

Understanding these aspects matters for determining whether inverse rendering is only a visually appealing representation or a useful tool for industrial anomaly detection.

1.3 Aim

The aim of this thesis is to investigate whether SVBRDF-based surface representations provide a more effective and robust basis for detecting small geometric surface defects, specifically scratches, than conventional image-space analysis. In particular, the thesis examines which components of SVBRDF representations are most relevant for anomaly detection, and whether geometry-related components such as surface normal maps can be transferred across different materials with similar physical surface structure.

1.4 Research questions

To fulfill the aim, the following research questions will be answered:

- RQ1** Does representing surfaces in SVBRDF space improve the detectability of small topological surface defects (scratches) compared to direct image-space analysis?
- RQ2** Are there combinations of SVBRDF components (e.g., surface normals, roughness) that increase the accuracy of scratch detection compared to using the full SVBRDF representation or individual components alone?
- RQ3** Can geometry-related components of SVBRDF representations, such as surface normals, be reused across different materials with similar physical surface structure for scratch detection?

1.5 Delimitations

This study is conducted in collaboration with an external company based in Linköping and was delimited to the software ecosystems available within their industrial machine vision portfolio.

The collaboration partner has also asked not to be named in the thesis. Consequently, specific proprietary details regarding internal software architectures, sensitive datasets, or hardware schematics may be generalized or omitted from this report so as not to expose the company's identity.

Due to confidentiality constraints, internal details of the anomaly detection model, such as network architecture, training loss functions, and optimization parameters, cannot be disclosed in this report. The model is therefore treated as a black-box detector with well-defined input and output behavior.

The focus of this thesis is not to evaluate or improve the anomaly detection model itself, but to investigate how different surface representations affect the detector's response. By keeping the detector fixed, the experimental design isolates the impact of representation choice on anomaly detection outcomes.

This study focuses on small geometric surface defects in the form of scratches. Other defect types, such as dents, discoloration, or large-scale deformations, are outside the scope of this work.



2 Theory and Related Work

This chapter covers the theoretical groundwork and current research relevant to the thesis. Since the goal is to determine whether single-shot AI solutions can work for industrial inspection, the physics of light transport and the mathematics behind inverting it must first be understood [12]. Section 2.1 introduces Physically Based Rendering (PBR), Inverse Rendering, and the neural network fundamentals used to solve them. Section 2.2 looks at recent work in Generative Adversarial Networks (GANs) and Neural Radiance Fields (NeRFs).

2.1 Theory

Visual inspection in manufacturing usually relies on standard image processing to spot changes in shape or color. However, as mentioned in the introduction, simple 2D analysis often fails to catch complex surface defects. For example, a scratch on shiny metal might disappear completely under certain lighting. To address this, the analysis must look beyond simple pixel intensity and estimate the intrinsic properties of the surface.

2.1.1 Image Formation and Forward Rendering

Forward rendering describes the process by which a three-dimensional scene is mapped to a two-dimensional image through the interaction of light, geometry, and material properties. This process is formally described by the rendering equation, introduced by Kajiya [7], which expresses outgoing radiance at a surface point as an integral over incoming illumination modulated by surface reflectance.

A direct consequence of the rendering equation is that image appearance is a coupled function of illumination, surface orientation, and material reflectance [7]. As formalized by the BRDF definition of Nicodemus et al.[9], reflectance depends on both incoming and outgoing directions, so the same surface geometry may produce markedly different image intensities under different lighting or viewing conditions. From the inverse perspective, multiple combinations of lighting and material parameters (and, in practice, geometry) can yield visually indistinguishable images, making image-space observations ambiguous [12].

In industrial inspection scenarios, this ambiguity poses a significant challenge. Surface defects such as shallow scratches may be clearly visible under one lighting configuration and nearly invisible under another. Since image-space features encode both intrinsic surface properties and extrinsic lighting effects, defect detection methods operating directly on pixel intensities are highly sensitive to illumination variations [15]. This motivates representations that aim to decouple intrinsic surface properties from illumination.

2.1.2 SVBRDFs as Material Representations

Surface reflectance is commonly modeled using a Bidirectional Reflectance Distribution Function (BRDF), which describes how light is reflected at a surface point as a function of incoming and outgoing directions [9]. For real-world surfaces, reflectance properties vary spatially due to texture, wear, and manufacturing processes, leading to the formulation of Spatially-Varying BRDFs (SVBRDFs).

In practice, SVBRDFs are parameterized using a few spatially varying maps, typically including albedo, surface normals, and roughness [3, 1]. These parameters correspond to physically interpretable surface properties: albedo represents intrinsic diffuse reflectance, normals encode local surface orientation, and roughness describes micro-scale surface irregularities that control specular reflection.

SVBRDF representations provide a material-centric description of surface appearance that can be evaluated under arbitrary illumination conditions. Unlike image-space representations, which entangle lighting and material effects, SVBRDFs explicitly separate different physical aspects of surface appearance [4]. This separation makes SVBRDFs well suited to inspection tasks that require reasoning about intrinsic surface properties rather than appearance under a specific lighting setup.

2.1.3 Geometry vs Appearance

A fundamental distinction in surface modeling is the separation between geometry and appearance. Geometry describes the physical shape of the surface, including both low-frequency form and high-frequency details such as scratches. Appearance, by contrast, describes optical properties such as reflectance, color, and micro-facet behavior [9, 3].

Within an SVBRDF representation, this distinction is reflected in the parameter maps. Surface normal maps primarily encode geometric information, while albedo and roughness capture appearance-related properties that may vary substantially across materials, coatings, or finishes [1]. In industrial manufacturing, components are often produced with identical geometry but finished using different materials or surface treatments, such as bare metal, plated metal, or painted surfaces.

This observation motivates the hypothesis that geometry-related representations are more invariant across materials than appearance-based representations. While albedo and roughness are expected to change with material composition, the underlying surface geometry, and thus the normal map should remain largely consistent for a given manufacturing process. Prior work in inverse rendering and material capture has similarly observed that geometric components are often more stable across varying appearance conditions [12, 4].

2.1.4 The Inverse Rendering Problem

Inverse rendering seeks to recover intrinsic scene properties, such as geometry, material, and lighting, from one or more observed images. Unlike forward rendering, inverse rendering is

ill-posed, as multiple combinations of lighting and material parameters can explain the same image observation [12].

Classical inverse rendering methods address this ambiguity by using multiple images captured under varying illumination or viewing directions, such as photometric stereo, which reliably recovers surface normals given known lighting variations [15]. While effective, such multi-shot approaches require controlled acquisition setups that are often impractical or expensive in industrial production environments.

Recent learning-based methods attempt to overcome these limitations by leveraging data-driven priors learned from large material datasets. Single-shot neural approaches can estimate SVBRDF parameters from a single image but inevitably rely on learned assumptions about typical materials and lighting conditions [4, 6]. As a result, these methods may introduce artifacts or hallucinated details when applied outside their training distribution.

2.1.5 Implications for Anomaly Detection

Anomaly detection in industrial inspection aims to identify deviations from expected surface properties using examples of defect-free components. When operating in image space, anomaly detection methods must implicitly model the combined variability of illumination, material appearance, and geometry. This increases the complexity of the normality model and reduces sensitivity to subtle geometric defects such as scratches [11].

SVBRDF representations offer an alternative by decomposing surface appearance into physically meaningful components. Scratches, which primarily alter surface geometry are expected to manifest as localized deviations in surface normal maps, while appearance-related variations may appear in albedo or roughness maps [1, 4]. This separation enables anomaly detection methods to focus on the most defect-relevant representation components.

SVBRDF-based representations also support more interpretable anomaly responses. Detected deviations can be analyzed in terms of specific surface properties rather than abstract pixel differences, which is valuable in industrial settings where understanding the nature of defects matters [11].

2.1.6 Anomaly Detection as a Representation-Dependent Process

In this work, anomaly detection is treated as a representation-dependent process, following the formulation of semi-supervised anomaly detection where models are trained exclusively on normal data [11]. A fixed internal anomaly detector is used to learn a model of normality from defect-free surface samples and subsequently applied to defective samples to produce pixel-level anomaly scores.

Rather than comparing different anomaly detection algorithms, this thesis focuses on how the choice of input representation affects anomaly detectability. The anomaly detector itself is held constant, and performance differences are attributed solely to differences in the representation space, namely image-space inputs versus SVBRDF-based inputs.

From this perspective, anomaly detection can be viewed as learning a compact manifold of normal surface behavior in a given representation space, i.e., a tightly clustered region where defect-free samples concentrate, with defects corresponding to deviations from this region [11]. If SVBRDF representations reduce nuisance variability, i.e., variations unrelated to surface defects such as illumination changes, the learned normality manifold (the region representing normal surface behavior) is expected to become more compact. This, in turn, increases the separability between normal and defective surface regions. This framing di-

rectly supports the evaluation of SVBRDF representations as a basis for industrial anomaly detection.

This work assumes pixel-wise correspondence between image-space inputs and SVBRDF parameter maps, which is reasonable for planar or approximately planar industrial surfaces captured under fixed viewpoints.

2.1.7 Traditional edge detection methods

The traditional, image-space baselines used in this thesis (Chapter 3) rely on classical edge detection and thresholding. The three algorithms applied are described below.

Canny edge detection

Canny edge detection was first introduced by John F. Canny in 1986 [2]. It is a multi-stage algorithm that detects edges in images by looking for local maxima of the gradient of the image intensity. The algorithm consists of several steps, including noise reduction, gradient calculation, non-maximum suppression, and edge tracking by hysteresis.[5]

1. Noise reduction: The input image is smoothed using a Gaussian filter to reduce noise and improve edge detection accuracy.
2. Finding the intensity gradient of the image: The algorithm computes the gradient magnitude and direction at each pixel using sobel filters.
3. Non-maximum suppression: The algorithm suppresses any pixel that is not a local maximum in the direction of the gradient.
4. Hysteresis thresholding: The algorithm uses two thresholds to identify strong and weak edges. Strong edges are those with gradient magnitudes above the high threshold, while weak edges are those with gradient magnitudes between the low and high thresholds. The algorithm then tracks edges by connecting strong edges and including weak edges that are connected to strong edges. This step is visualized in figure 2.1.

LoG edge detection

LoG edge detection, also commonly referred to as Marr-Hildreth edge detection because David Marr and Ellen Hildreth were the first to introduce it in 1980 [8], is an edge detection method that uses the Laplacian of Gaussian operator to identify edges in an image. The algorithm consists of two main steps: smoothing the image with a Gaussian filter to reduce noise, and then applying the Laplacian operator to detect edges. The Laplacian operator computes the second derivative of the image intensity, which helps to highlight regions of rapid intensity change that correspond to edges. The output of this method is a binary edge map that indicates the locations of detected edges in the input image.

Otsu thresholding

The last method chosen was Otsu thresholding [10], which is a global thresholding technique that automatically determines an optimal threshold value that minimizes the intra-class variance of the thresholded black and white pixels. [13]

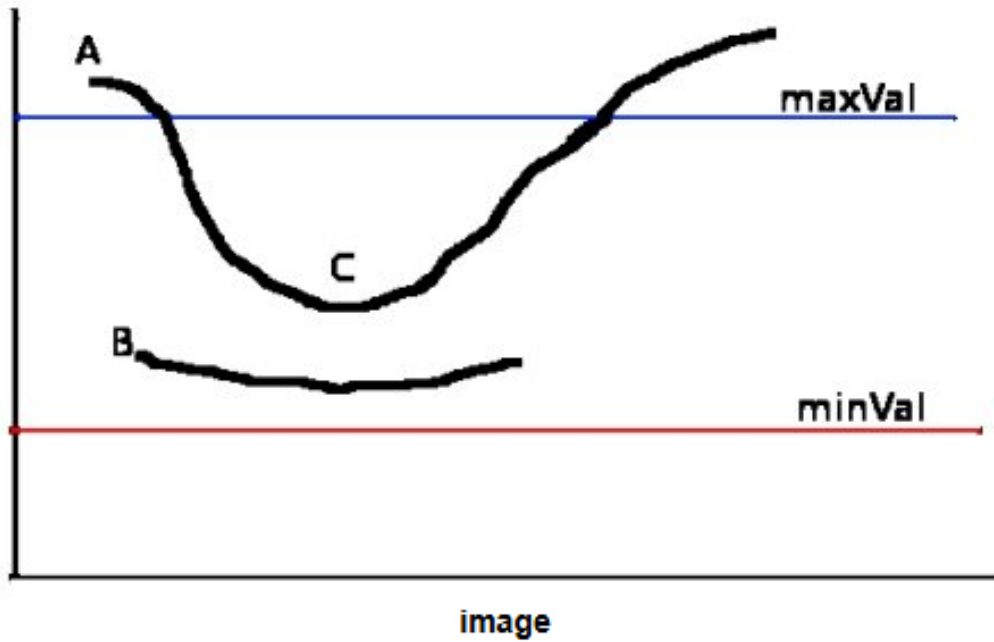


Figure 2.1: Visualization of the hysteresis thresholding step in Canny edge detection, taken from [5], where any value over maxVal is considered a strong edge, and any value in between minVal and maxVal is considered a weak edge. As seen in the image, the upper line has values that are both weak and strong, but since they are connected to strong edges, they are included in the final edge map. The lower line has no values that are considered strong, so they are excluded from the final edge map.

2.2 Related work

Inverse rendering has moved quickly from traditional math-based optimization to data-driven deep learning. This section outlines that progression as well as some validation metrics that are relevant to this thesis.

2.2.1 SVBRDF estimation methods

Estimating Spatially-Varying Bidirectional Reflectance Distribution Functions (SVBRDFs) from images has been an active research topic in computer graphics and vision, motivated by applications in material acquisition, rendering, and surface analysis. Early work focused on physics-based and optimization-driven approaches, often relying on controlled capture setups and multiple observations to recover material parameters with high fidelity.

More recent approaches use data-driven priors learned from large material datasets. Deschaintre et al.[4] introduced a rendering-aware deep network capable of estimating per-pixel albedo, normal, and roughness maps from a single image. By incorporating a differentiable rendering loss, their method enforces consistency between predicted SVBRDF parameters and the observed image under known lighting, producing plausible material decompositions even in the single-shot setting.

Gao et al.[6] extended this line of work by combining learned latent material spaces with explicit inverse rendering optimization at inference time. Their approach supports both single-image and multi-image inputs and produces high-resolution SVBRDF maps that preserve

fine-scale surface detail. Their results show that multi-shot inputs substantially improve generalization under novel lighting conditions compared to single-shot estimation.

More recently, Wang et al.[14] proposed NFPLight, a hybrid capture approach that combines near- and far-field photos with point lighting to increase the amount of reflectance information available per observation. This method achieves high-quality SVBRDF estimation with fewer images than traditional multi-shot setups, bridging the gap between classical photometric techniques and purely single-image methods.

In this thesis, SVBRDF estimation methods are not evaluated as rendering solutions in their own right, but rather as means of generating material representations that serve as inputs to anomaly detection. As such, their relevance is assessed based on how well the resulting SVBRDF components preserve defect-relevant surface information.

2.2.2 Multi-shot vs Single-shot capture

SVBRDF estimation methods can broadly be categorized based on the number of observations required for material capture. Multi-shot methods rely on multiple images of the same surface captured under varying illumination or viewing conditions, while single-shot methods attempt to recover material parameters from a single observation.

Classical multi-shot techniques, such as photometric stereo, exploit known lighting variations to reliably estimate surface normals and, in some cases, reflectance parameters [15]. By observing how surface appearance changes across lighting conditions, these methods significantly reduce the ambiguity inherent in inverse rendering. As a result, multi-shot approaches are often treated as a form of ground truth or reference in material acquisition pipelines.

However, multi-shot capture requires controlled acquisition setups with either moving light sources or multiple synchronized illuminants. In industrial inspection environments, such setups can be costly, mechanically complex, or incompatible with high-throughput production lines. These constraints have motivated increased interest in single-shot SVBRDF estimation, which requires only a single image per surface.

Single-shot methods, including the work of Deschaintre et al. [4], rely heavily on learned priors to compensate for missing information. While these approaches can produce visually plausible SVBRDF maps, their estimates are biased toward the statistics of the training data and may fail to generalize under novel illumination or atypical surface conditions. Gao et al.[6] explicitly demonstrate that single-shot estimates often reproduce input lighting artifacts, whereas multi-shot optimization produces more physically consistent results.

Hybrid approaches, such as NFPLight, attempt to balance these trade-offs by increasing the information content of each capture without requiring extensive multi-shot sequences [14]. These methods suggest that the distinction between single-shot and multi-shot capture is not binary, but rather a continuum of information availability.

In the context of this thesis, the distinction between single-shot and multi-shot SVBRDF estimation is critical not for rendering quality alone, but for understanding how capture strategy affects the robustness and transferability of SVBRDF representations when used for anomaly detection.

2.2.3 Image quality metrics

Image quality metrics provide quantitative measures for assessing the difference between predicted outputs and corresponding ground truth data. Such metrics are widely used in computer vision and image reconstruction tasks to enable objective comparison between

methods and results. In this thesis, these metrics compare the SVBRDF maps generated by NFPLight against the synthetic ground truth, as described in the methodology (Section 3) and reported in Section 4.

Among these, mean angular error is commonly used for evaluating surface normal estimation, while mean absolute error is frequently applied to compare image intensities, such as diffuse and roughness maps.

Mean Angular Error

Mean angular error is a commonly used metric for evaluating the accuracy of estimated surface normals in computer vision tasks such as normal estimation and photometric stereo. It measures the angular deviation between a predicted normal vector and the corresponding ground truth normal vector at each pixel.

The angular error is defined as

$$AngularError = \arccos(\hat{n}_{gt} \cdot \hat{n}_{pred}) \quad (2.1)$$

and the mean angular error is given by

$$MeanAngularError = \frac{1}{N} \sum_{k=1}^N \arccos(\hat{n}_{gt_k} \cdot \hat{n}_{pred_k}). \quad (2.2)$$

Here, N is the total number of pixels ($width \cdot height$), $\hat{n}_{gt}$ denotes the ground truth normal, and $\hat{n}_{pred}$ denotes the predicted normal. The arccosine extracts the angle between the two vectors.

This formulation follows from the relationship between the dot product and the angle between two vectors:

$$a \cdot b = ||a|| ||b|| \cos(\theta) \quad (2.3)$$

which can be rewritten as

$$\frac{a \cdot b}{||a|| ||b||} = \hat{a} \cdot \hat{b} = \cos(\theta), \quad (2.4)$$

where $\hat{a}$ and $\hat{b}$ are normalized vectors. Taking the arccosine of both sides yields equation 2.1.

Mean Absolute Error

Mean absolute error (MAE) is a standard metric used to quantify the difference between predicted and ground truth image intensities. It is widely applied in image reconstruction tasks, including the evaluation of diffuse and roughness maps.

For a given pixel, the absolute error is defined as

$$AbsoluteError = \frac{1}{C} \sum_{c=1}^C |I_{gt}(c) - I_{pred}(c)|, \quad (2.5)$$

where C is the number of channels, and I_{gt} and I_{pred} represent the ground truth and predicted intensities, respectively.

The mean absolute error across an image is then given by

$$MeanAbsoluteError = \frac{1}{CHW} \sum_{c=1}^C \sum_{y=1}^H \sum_{x=1}^W |I_{gt}(x, y, c) - I_{pred}(x, y, c)| \quad (2.6)$$

where H and W denote the image height and width.

2.2.4 Classification metrics

Classification metrics provide quantitative measures for evaluating the performance of binary and multi-class classification tasks. These metrics assess how well predicted labels correspond to ground truth labels, taking into account different types of prediction errors such as false positives and false negatives. In this thesis, they quantify scratch-detection performance for every representation, following the evaluation strategy in Section 3.7 and reported in Section 4.

F1-score

The F1-score is a commonly used classification metric that combines precision and recall into one single measure. It is useful when there is an imbalance between classes or when both false positives and false negatives are important.

Precision is defined as the ratio of true positives to the sum of true positives and false positives:

$$Precision = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \quad (2.7)$$

Recall is defined as the ratio of true positives to the sum of true positives and false negatives:

$$Recall = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (2.8)$$

The F1-score is the harmonic mean of precision and recall:

$$F1\text{-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2.9)$$

A high F1-score indicates a balance between precision and recall, meaning that the classifier achieves both low false positive and low false negative rates.

Intersection over Union

IoU, or Intersection over Union, is a metric used to evaluate the accuracy of object detection and segmentation task. It measures the overlap between the predicted bounding box or segmentation mask and the ground truth. The IoU is calculated as the area of intersection divided by the area of union between the predicted and ground truth regions

$$IoU = \frac{\text{Area of Intersection}}{\text{Area of Union}} \quad (2.10)$$

A higher IoU indicates better agreement between prediction and ground truth, while a lower IoU reflects poor overlap, which may result from misalignment, incorrect scale, or incomplete detection.

Connected-Component Labeling

Connected-component labeling is a technique used in image processing to identify and group connected regions of pixels that share similar properties. In binary images, it is commonly applied to detect distinct objects by assigning a unique label to each connected region.

This method is frequently used in segmentation and object detection tasks to extract individual components from predicted or ground truth masks, enabling further analysis at the object level rather than the pixel level.

3 Methodology

This chapter goes over the methodology that was used to answer the research questions. The pipeline used can be seen in figure 3.1.

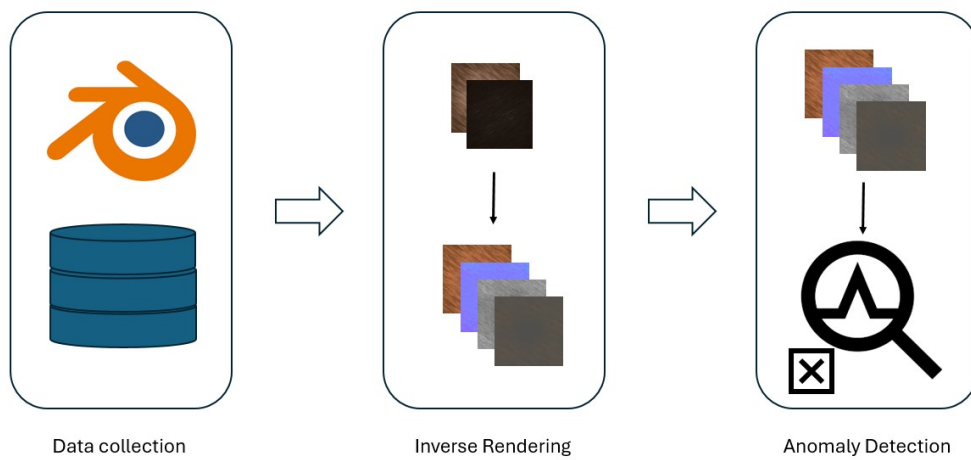


Figure 3.1: This figure shows the overall pipeline for the methodology in this thesis. The first step was data collection, where Blender was used for the synthetic data. The second part was Inverse rendering as a preprocessing step. The final step was using the generated SVBRDF maps in anomaly detection, where the results to answer the research questions were gathered.

3.1 SVBRDF map generation

This section describes the methods used to generate SVBRDF parameter maps that serve as input representations for anomaly detection. The purpose of SVBRDF estimation in this thesis is not photorealistic rendering, but the extraction of physically meaningful, spatially aligned surface representations that preserve scratch-relevant information. To ensure suitability for anomaly detection, the following selection criteria were imposed on SVBRDF estimation methods:

- **Deterministic output:** Given the same input, the method must produce consistent results.
- **Per-pixel correspondence:** Output maps must maintain spatial alignment with the input image.
- **No texture synthesis or tiling:** Methods relying on stochastic texture generation or infinite tiling were excluded.
- **Explicit material decomposition:** The method must produce interpretable SVBRDF components, including at least surface normals and reflectance parameters.

Along with these criteria, since the data was to be collected from scratch, having a model with fewer input images required per generated SVBRDF map was prioritized. Deschaintre et al.'s[4] model would have worked to this end, but the low resolution of 256x256 per image led to using NFPLight instead.

3.1.1 Hybrid Near- and far-field lighting capture, NFPLight

Wang et al.[14] propose NFPLight, a hybrid capture approach that combines near-field and far-field point lighting to increase the information content of each observation. The method uses two images captured under different lighting conditions to estimate SVBRDF parameters without iterative optimization.

NFPLight occupies an intermediate position between classical multi-shot methods and single-image approaches. In this thesis, it is used as a representative example of a low-shot capture strategy that augments information content without requiring extensive capture sequences.

3.2 Image quality metrics

Since the scratch detection task relies on the quality of the input representations, it is important to evaluate the fidelity of the generated SVBRDF maps. To this end, two separate image quality metrics were selected to quantitatively assess the similarity between the generated SVBRDF maps and ground truth maps.

The selected metrics are:

- **MAnGE (Mean Angular Error)**, which measures the average angular difference between the predicted surface normals and the ground truth normals, providing a direct assessment of geometric accuracy (Section 2.2.3).
- **MAE (Mean Absolute Error)**, which measures the average absolute difference between the predicted roughness values and the ground truth roughness values, providing a direct assessment of the accuracy of the roughness estimation (Section 2.2.3).

3.3 Data

The data used in this thesis is a mix of synthetic and real data.

Synthetic data was collected using Blender to get controlled yet varying data with realistic results which can nearly match that of real data.

Real data was collected using a custom built rig, which consists of a camera with two separate lenses with different focal lengths, and an accompanying light acting as a flash.

One important aspect to take into account, which is also why synthetic data is chosen as a starting point, is that the synthetic scratches has ground truth annotated into them. This is not the case for the real data.

For this exact reason synthetic data will be the starting point since the maps can be compared to one another to see how close the models get to ground truth, before later testing the inputs vs outputs for the anomaly detection methods.

There were two different types of data collected for this thesis.

- Synthetic data, covered in section 3.3.1.
- Real data, covered in section 3.3.2.

3.3.1 Synthetic data collection

The synthetic data was collected in a way that was similar to the examples shown in Wang et al.'s paper [14] but with Blender. The scene was set up to have a plane, a point light, a camera and a box in which everything else was placed inside. After the setup, a python script was created which generated materials using predetermined image textures of realistic materials, and then added scratches using a mask image which were added to the normal map of the material. The script then applied different rotations and translations of the material and its scratches, and then rendered the scene first using the Cycles rendering engine for two photorealistic renders, one near-field image and one far-field image, then a third image was rendered where only the black and white scratch map was visible, which was used as the ground truth for the scratch location. The script loops through this process for a predetermined number of times to create a dataset of synthetic images with varying materials, scratch locations and orientations, and lighting conditions. The resulting dataset consists of aligned sets of near-field images, far-field images, and ground truth scratch masks. The far-field and near-field images were used in the NFPLight inverse rendering method to generate SVBRDF maps, which were then used as input to the anomaly detection method. The ground truth scratch masks were used to evaluate the performance of the anomaly detection methods by comparing the generated anomaly maps to the known scratch locations.

3.3.2 Real data collection

For the real data collection, a custom rig was built and tweaked, which can be seen in figure 3.2. The rig is meant to simulate the data collection shown in Wang et al.'s Github videos¹ [14], but more controlled, as to allow for faster data collection. The rig consists of one camera, one white LED light that is mounted to the camera using a custom 3D printed mount, a retractable pole in which the camera is mounted to, and a box in which the camera, material and light were all enclosed by, with a black cloth covering the box to make sure that no external light or inconsistent light is entering the box during data collection. The retractable pole allows the

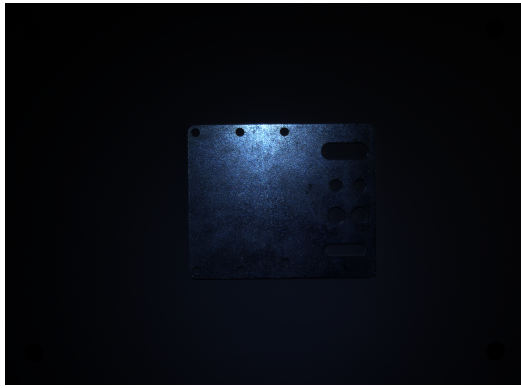
¹Link to Wang et al.'s Github videos



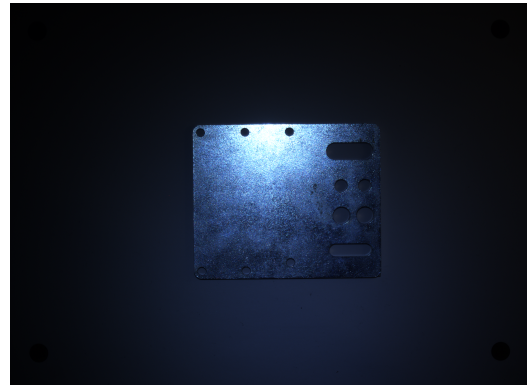
Figure 3.2: The rig which was used to collect real data.

camera to be moved into near-field and far-field positions, which keeps each image aligned and makes sure that the sample overlaps in the near- and far-field images. Two different lenses were chosen, one 12,5mm lens and a 25mm lens, which allows for a larger field of view in the far-field images and a smaller field of view in the near-field images, which is similar to the setup shown in Wang et al.'s videos [14] with their phone. For the near-field image, a distance of 25cm from the subject was chosen, and for the far-field-image, a distance of 50cm from the subject was chosen.

Because some of the materials that were chosen, like metal, were very reflective, there is a tendency for the images to get either blown out at the center of the image where the light is reflecting the most, or to be very dark at the edges of the image where the light is not reflecting as much. To combat this, multi-exposure HDR images were taken, where two images were taken for each near-field and far-field position, one with a short exposure time to capture the details in the bright areas (center of the image), and one with a long exposure time to capture the details in the dark areas (edges of the image). This allows for highly reflective materials to be captured without the loss of details, and therefore data, due to overexposure or underexposure. Some examples of this is seen in figure 3.3



(a) Low exposure time of around 5ms.



(b) Higher exposure time of around 20ms.



(c) High exposure of around 400ms, showing clear dots in the corners of the image used as control points.



(d) Final HDR image of real data.

Figure 3.3: Far away image of metal object.

Finally, four fixed dots were placed around the subject, creating a rectangular shape. The dots were used as control points for *perspective transformation*, allowing the near-field and far-field images to be spatially aligned while also keeping the subject to the exact same position between images even if the camera were to be shifted. This is important for the NFPLight inverse renderer methods, which require spatially aligned inputs.

3.4 Classification metrics

For this thesis, three methods for classification metrics were selected to evaluate the performance and accuracy of the anomaly detection methods. These metrics were:

- F1-score.
- Intersection over Union (IoU).
- Threshold of Connected-Component Labeling, with a detection threshold of 70%.

The generated anomaly maps from the anomaly detection methods will be compared to the ground truth masks of the scratches to evaluate the performance of the methods. The F1-score will be used to evaluate the overall classification performance, while IoU will be used to evaluate the spatial accuracy of the detected anomalies.

However, F1-score and IoU is very unforgiving when detecting small scratches, and a thresholded connected-component labeling was used to see clearly if scratches as a whole were detected or not.

3.5 Anomaly detection through traditional methods

This section describes the traditional, image-space anomaly detection methods used as baselines in this thesis. These methods operate directly on pixel intensities and do not rely on learned representations or material decomposition. Their purpose is to establish a transparent and interpretable reference against which SVBRDF-based representations can be compared. Traditional methods were selected to reflect common practices in industrial machine vision, where simplicity, robustness, and interpretability are often prioritized. All methods described in this section produce pixel-level anomaly maps indicating deviations from expected surface appearance.

The traditional methods described in this thesis serve as image-space baselines and were not expected to capture subtle surface scratches whose visibility depends on illumination. Their role is to provide a point of reference for assessing whether SVBRDF-based representations offer a meaningful advantage over direct image-space analysis.

3.5.1 Edge-based Anomaly detection

Edge-based detection is used as a primary image-space baseline. The underlying assumption is that surface defects such as scratches introduce local intensity discontinuities that can be detected through image gradients.

In this work, three methods for edge-based surface defect detection were considered:

- Canny edge detection, covered in section 2.1.7.
- Laplacian of Gaussian (LoG) edge detection, covered in section 2.1.7.
- Otsu thresholding, covered in section 2.1.7.

In this work, a Canny edge detector is applied to input images to extract high-gradient regions. The resulting edge map is interpreted as a proxy for anomalous surface structure. To reduce sensitivity to noise and minor texture variations, standard non-maximum suppression and hysteresis thresholding are applied as part of the edge detection pipeline.

The output of this method is a binary or weighted edge map that highlights regions of abrupt intensity change. This map is used directly as an anomaly indication or further processed using morphological operations to improve spatial coherence.

3.6 Anomaly detection through deep learning methods

This section describes the deep learning-based anomaly detection approach used in this thesis. Unlike the traditional methods in Section 3.5.1, the deep learning-based detector is not implemented as part of this work but is provided by the industrial collaboration partner.

3.6.1 Anomaly detection paradigm

The internal anomaly detection model follows a semi-supervised anomaly detection paradigm. The model is trained exclusively on defect-free samples (i.e., surfaces without scratches) to learn a representation of normal surface appearance. During inference, the trained model processes unseen samples and produces pixel-level anomaly scores that indicate deviations from the learned normality model.

The detector accepts spatially aligned input representations and produces anomaly maps with the same spatial resolution as the input. This makes it suitable for direct comparison between image-space inputs and SVBRDF-based inputs.

3.6.2 Input representation

The anomaly detector is evaluated using different input representations, including:

- Raw image-space inputs
- Individual SVBRDF parameter maps (e.g., albedo, surface normals, roughness)
- Selected combinations of SVBRDF components (e.g., normals + roughness)

No changes are made to the detector architecture or training procedure between experiments.

3.7 Evaluation Strategy

The evaluation strategy is designed to directly address the research questions by assessing how different surface representations affect the detectability of surface scratches. In line with the theoretical framing of anomaly detection as a representation-dependent process, a fixed anomaly detector is used throughout all experiments. Performance differences are therefore attributed solely to the choice of input representation rather than to differences in detection algorithms.

Evaluation is conducted at the pixel level and focuses on three complementary aspects: (i) representation-level fidelity of SVBRDF components, (ii) scratch detectability across representations, and (iii) transferability of geometry-based representations across materials.

3.7.1 Defect detectability across representations (RQ1 & RQ2)

To address the first two research questions, the anomaly detector is evaluated across a variety of surface representations derived from both image-space and SVBRDF-space inputs.

For each representation, the anomaly detector is fine-tuned exclusively on defect-free samples of either SVBRDF maps or image space inputs to model normal surface behavior. The fine-tuned model is then applied to samples containing known scratches, producing pixel-level anomaly score maps. For RQ1, defect detectability is compared between the baseline image-space inputs and SVBRDF-based representations. SVBRDF inputs include individual parameter maps (albedo, surface normals, roughness, specular).

For RQ2, the focus is on determining whether combinations of SVBRDF components outperform individual components or the full SVBRDF set. To this end, selected combinations (e.g., normals + roughness, normals + albedo) are tested in addition to single-component inputs. These combinations are chosen to isolate complementary information channels, such as geometry-related versus appearance-related cues.

For both RQ1 and RQ2, a separate model is fine-tuned for each individual parameter map. Combined representations are then produced by linearly combining the resulting anomaly heatmaps before thresholding and validation, rather than by stacking the maps into a single multi-channel input.

Synthetic data is used as the initial test environment, since it provides automatic, pixel-accurate ground-truth scratch masks through the rendering pipeline. This allows quantitative evaluation using overlap-based metrics such as precision, recall, and Intersection-over-Union (IoU), which directly measure how well each representation supports accurate localization of small surface scratches. Connected-component analysis is also used, as IoU- and F1-scores can be overly punishing for scratch detection.

Real data would have to be annotated to be able to be used for this evaluation, which is a time-consuming process. Therefore, it is only used for qualitative evaluation, where the generated anomaly maps are visually compared to the input images to assess whether detected anomalies correspond to visible scratches. This qualitative analysis complements the quantitative results from synthetic data and provides insights into real-world applicability. The results will however not be used to directly answer the research questions, but rather to provide a sanity check for the synthetic data results and to identify potential gaps between synthetic and real-world performance.

This evaluation collectively answers whether SVBRDF-based representations improve scratch detectability relative to image-space inputs (RQ1) and whether specific component combinations provide superior anomaly detection performance compared to individual SVBRDF components (RQ2).

To further disentangle the effect of the learned SVBRDF representations from the performance of the anomaly detector, an additional evaluation is conducted using ground-truth SVBRDF maps provided by the synthetic rendering pipeline. In particular, ground-truth normal maps and other parameter maps are directly used as inputs to the anomaly detector, bypassing the inverse rendering step entirely.

This setup serves as an upper-bound reference for representation quality, allowing performance differences to be attributed either to limitations of the inverse rendering method or to intrinsic limitations of the representation itself. If anomaly detection performance improves significantly when using ground-truth maps compared to their estimated counterparts, this indicates that errors introduced during inverse rendering degrade the usefulness of the representation. Conversely, if performance remains similar, it suggests that the representation (e.g., normal maps) may not sufficiently capture the features required for reliable scratch detection, regardless of estimation accuracy. This step is only taken with the base inputs, excluding it from RQ2. This is because there are no defects in the roughness and diffuse maps in ground truth.

3.7.2 Cross-material transferability of Geometry-based representations (RQ3)

To evaluate the transferability of geometry-related SVBRDF components across materials, experiments are conducted where the anomaly detector is trained on defect-free samples of the normal map from one material and tested on surfaces with identical geometry but different material appearance.

Specifically, training is performed using surfaces of a reference material (e.g., uncoated metal), while testing is conducted on geometrically identical surfaces with altered reflectance properties (e.g., plated or painted variants). This procedure is repeated for different input representations, with particular focus on surface normal maps.

Performance degradation under material change is measured using the same detection metrics as in Section 3.7.1. If geometry-based representations are invariant to material appearance, detection performance is expected to degrade less for normal-based inputs than for image-space or appearance-based SVBRDF inputs.

This experiment directly addresses the third research question by quantifying the extent to which geometry-related SVBRDF components can be reused across materials for scratch detection.

4 Results

This section includes the results relevant to answering the research questions of this thesis, which can be found in section 1.4. The results include SVBRDF maps produced with NFPLight inverse rendering, the quality metrics scoring the generated SVBRDF maps, and the scratch detection results from both synthetic data and real data, tested with both traditional methods and the internal anomaly detection tool.

4.1 Image Quality Metric

The following figures 4.1 4.2 are the results from comparing ground truth SVBRDF maps to the generated maps from NFPLight. Specular maps were not compared, as the materials collected did not include them.

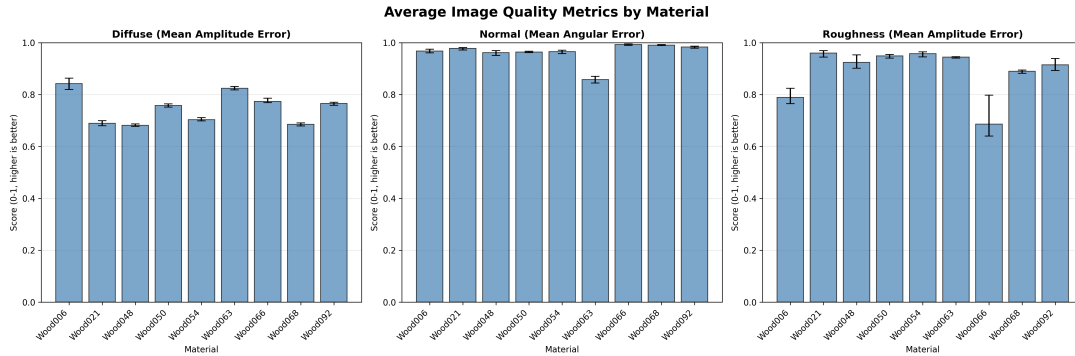


Figure 4.1: Comparison of material quality metrics for different SVBRDF estimation methods and their corresponding type of map they were used on. Each method horizontal axis represents the different materials that were rendered and used as input data for the NFPLight inverse rendering method. The vertical axis represents the score, or quality, of each material produced using NFPLight inverse rendering when compared to ground truth. Each material has a corresponding line associated with it, which shows the range of quality score for each material under different rotation and translation to the materials UV coordinates.

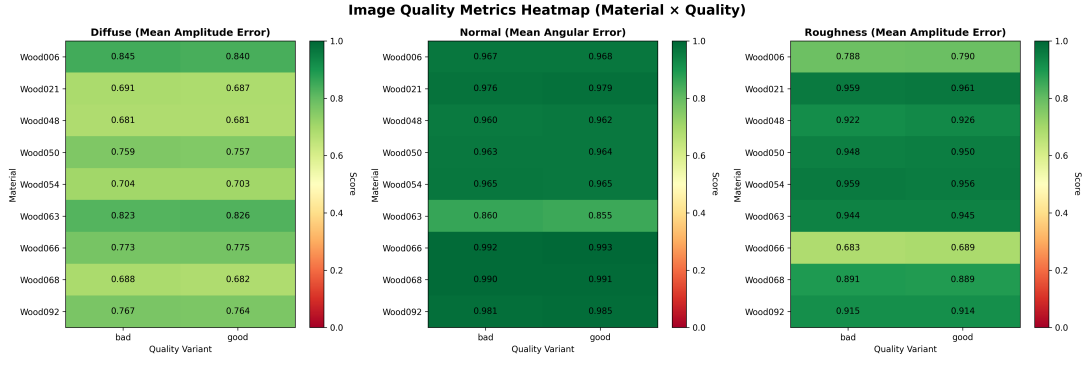


Figure 4.2: Heatmap visualization of material quality metrics for different SVBRDF estimation methods. Each method horizontal axis represents if there were defects present in the material (such as scratches), or if the material was defect free. The vertical axis represents the materials that were tested on with NFPLight inverse rendering. The number and corresponding color represents the quality score of each material when compared to ground truth, where higher scores and greener colors indicate better quality.

4.2 SVBRDF maps

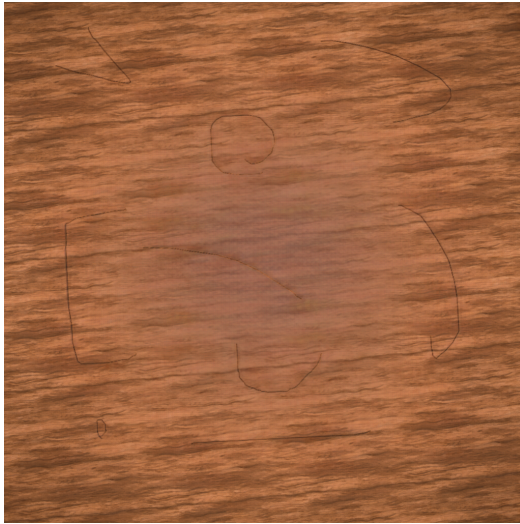
Figure 4.3 shows some of the resulting SVBRDF maps from the Wood63 material.



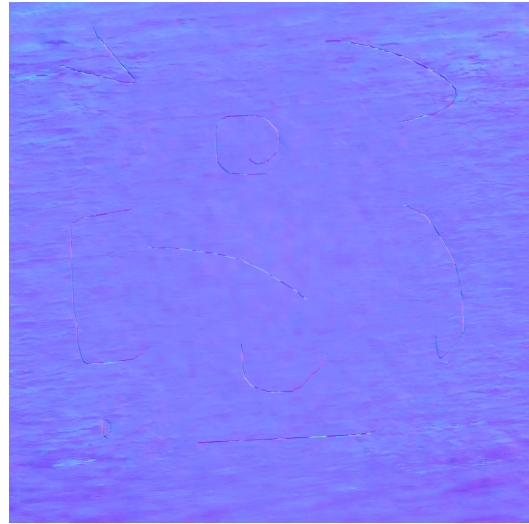
(a) Near-field image.



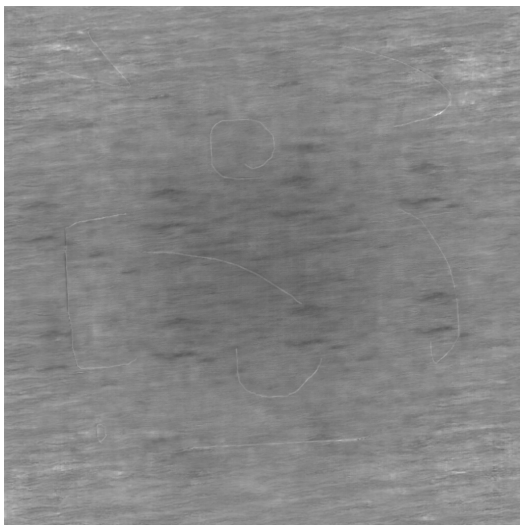
(b) Far-field image.



(c) Diffuse map.



(d) Normal map.



(e) Roughness map.

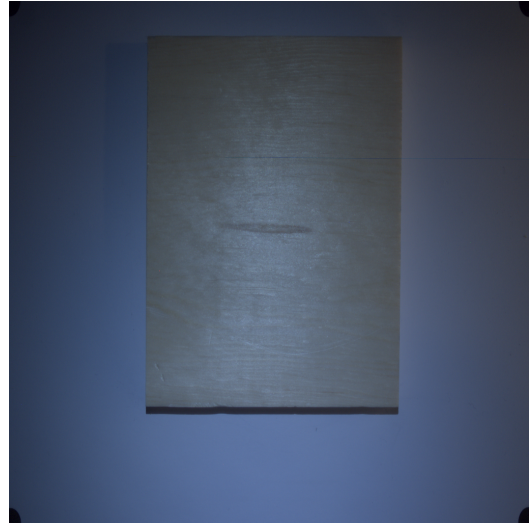


(f) Specular map.

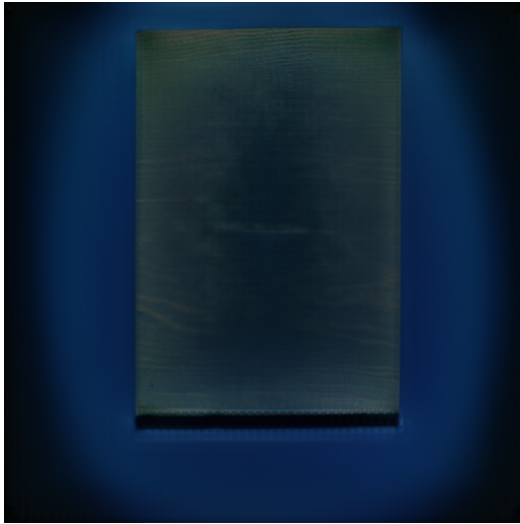
Figure 4.3: SVBRDF maps generated by NFPLight from a synthetic Wood63 sample. The near-field (a) and far-field (b) renders are the inputs, and NFPLight estimates the diffuse (c), normal (d), roughness (e), and specular (f) maps. The scratch is encoded mainly in the normal map.



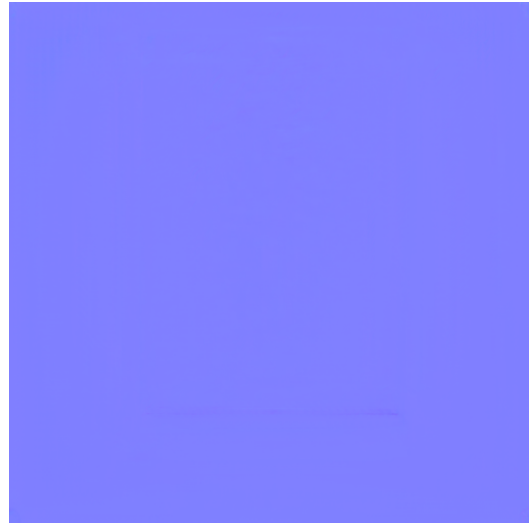
(a) Near-field image.



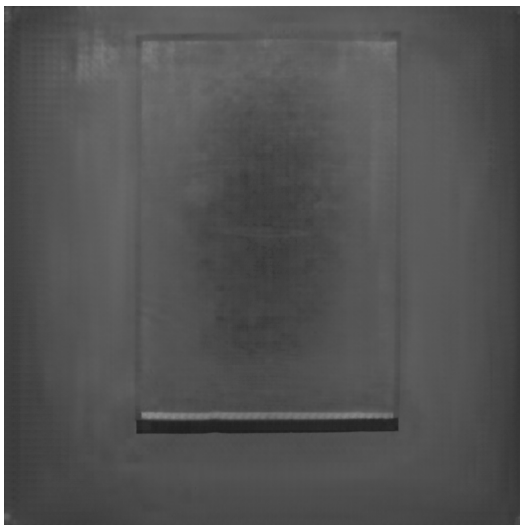
(b) Far-field image.



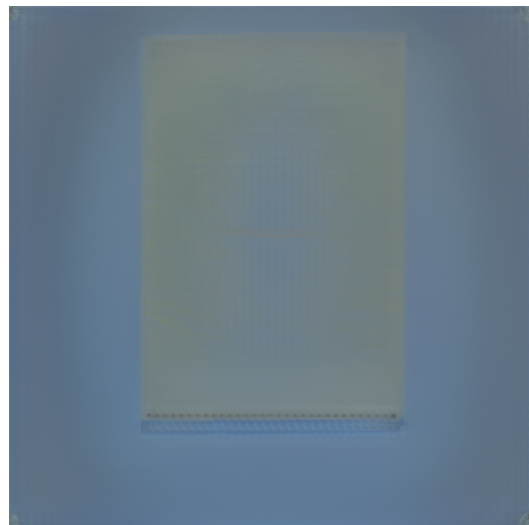
(c) Diffuse map.



(d) Normal map.



(e) Roughness map.



(f) Specular map.

Figure 4.4: SVBRDF maps generated by NFPLight from a real captured sample, with the same layout as Figure 4.3. The estimated maps are visibly poorer than for the synthetic data, reflecting the domain gap discussed in Section 5.2.

4.3 Edge detection using traditional methods

Edge detection with traditional methods include Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection with morphological closing operations to create a bit-mask representing the predicted defects.

All results in this section use the synthetic dataset, where every scratch has a pixel-accurate ground-truth mask. Detection rate is the percentage of scratches covered at least 70% by the prediction, while F1 and IoU measure pixel-level agreement (Sections 2.2.4 and 2.2.4).

4.3.1 Ground truth as input

Figure 4.5 shows the three methods applied to the ground-truth normal maps of defective samples. Canny edge detection clearly outperforms Otsu thresholding and Marr-Hildreth on every metric, and is the only method that reaches a useful detection rate without collapsing its F1 and IoU scores. The corresponding scratch-free results are shown in Appendix A.1.

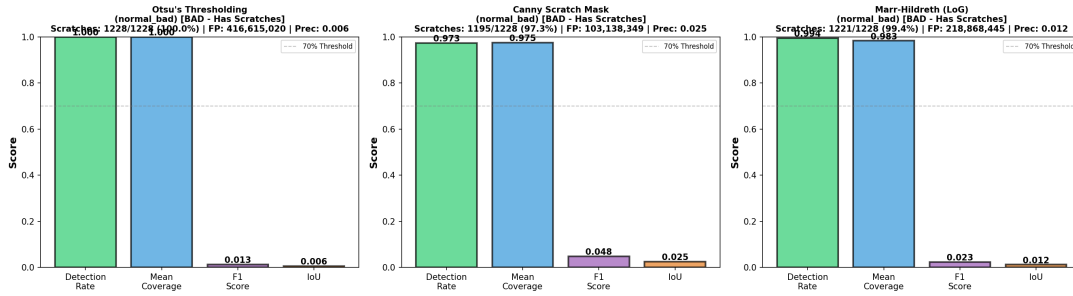


Figure 4.5: Graph showing the results of using Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection methods without Sobel filtering as a preprocessing step, with the ground truth normal SVBRDF map. Detection rate counts the percentage of scratches with at least 70% coverage from the detection method, F1 score relates to the method explained in the method Section 2.2.4, and lastly IoU refers to another method explained in the method Section 2.2.4. The subject had defects.

4.3.2 SVBRDF map inputs

Figure 4.6 shows the same methods applied to the normal maps that NFPLight generates from the near-field and far-field images, again on defective samples. Of the SVBRDF components only the normal map was tested, because the scratches are geometric and are encoded almost entirely in surface orientation rather than in the diffuse, roughness, or specular maps. Canny again performs best, but every method detects fewer scratches on the generated normal maps than on the ground-truth maps above. The scratch-free and Sobel-preprocessed variants are shown in Appendix A.1.

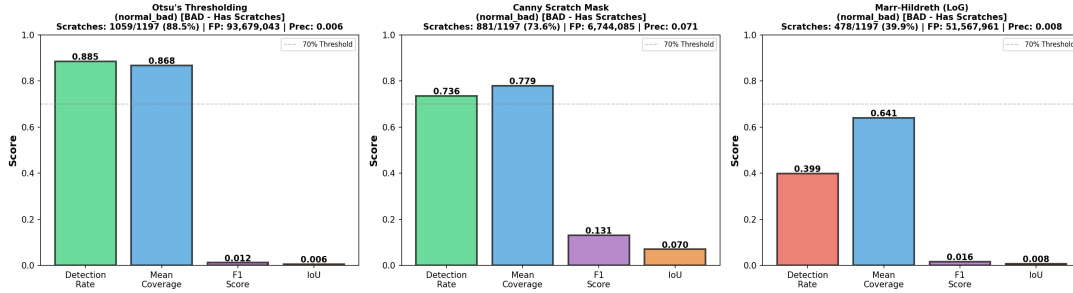


Figure 4.6: Graph showing the results of using Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection methods without Sobel filtering as a preprocessing step, with the normal SVBRDF map created with NFPLight inverse rendering. Detection rate counts the percentage of scratches with at least 70% coverage from the detection method, F1 score relates to the method explained in the method Section 2.2.4, and lastly IoU refers to another method explained in the method Section 2.2.4. The subject had defects.

4.3.3 Direct image-space inputs

Figure 4.7 shows the methods applied directly to the near-field images of defective samples. Detection is markedly better than on the normal maps: Canny again leads and recovers more scratches than any method managed on the normal maps, which indicates that for classical edge detection the raw image-space input carries more usable scratch information than the inverse-rendered normal map, even when that map is taken from ground truth. The far-field results and the scratch-free and Sobel-preprocessed variants are shown in Appendix A.1.

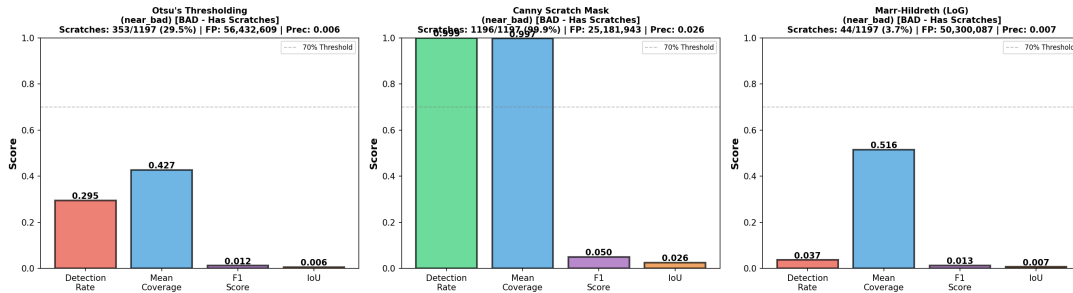


Figure 4.7: Graph showing the results of using Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection methods without Sobel filtering as a preprocessing step, with the camera positioned near the subject. Detection rate counts the percentage of scratches with at least 70% coverage from the detection method, F1 score relates to the method explained in the method Section 2.2.4, and lastly IoU refers to another method explained in the method Section 2.2.4. The subject had defects.

4.4 AI anomaly detection

This section reports the proprietary anomaly detector on the synthetic dataset, using each representation as both the training and the testing input. Detection rate, F1, and IoU are reported as before, and the full set of per-map plots is collected in Appendix A.2.

4.4.1 SVBRDF map inputs

Each SVBRDF component (diffuse, normal, roughness, and specular) was used on its own as the detector input, across two experiments: one with deep, easily visible scratches and one with shallow scratches. None of the individual SVBRDF maps matched the near-field image-space input, and the diffuse map was the strongest of them, which is consistent with it being

the SVBRDF component that most resembles a raw image. The per-map metric plots for both experiments are collected in Appendix A.2, and a qualitative comparison of the thresholded heatmaps is shown in Figure 4.18.

4.4.2 Ground truth as input

To separate the quality of the inverse rendering from the quality of the representation itself, the detector was also trained and tested on ground-truth maps taken directly from the synthetic pipeline. Figures 4.8 and 4.9 show the near-field image-space and ground-truth normal-map results, which act as an upper-bound reference for how much of the performance gap comes from inverse-rendering error rather than from the representation itself.

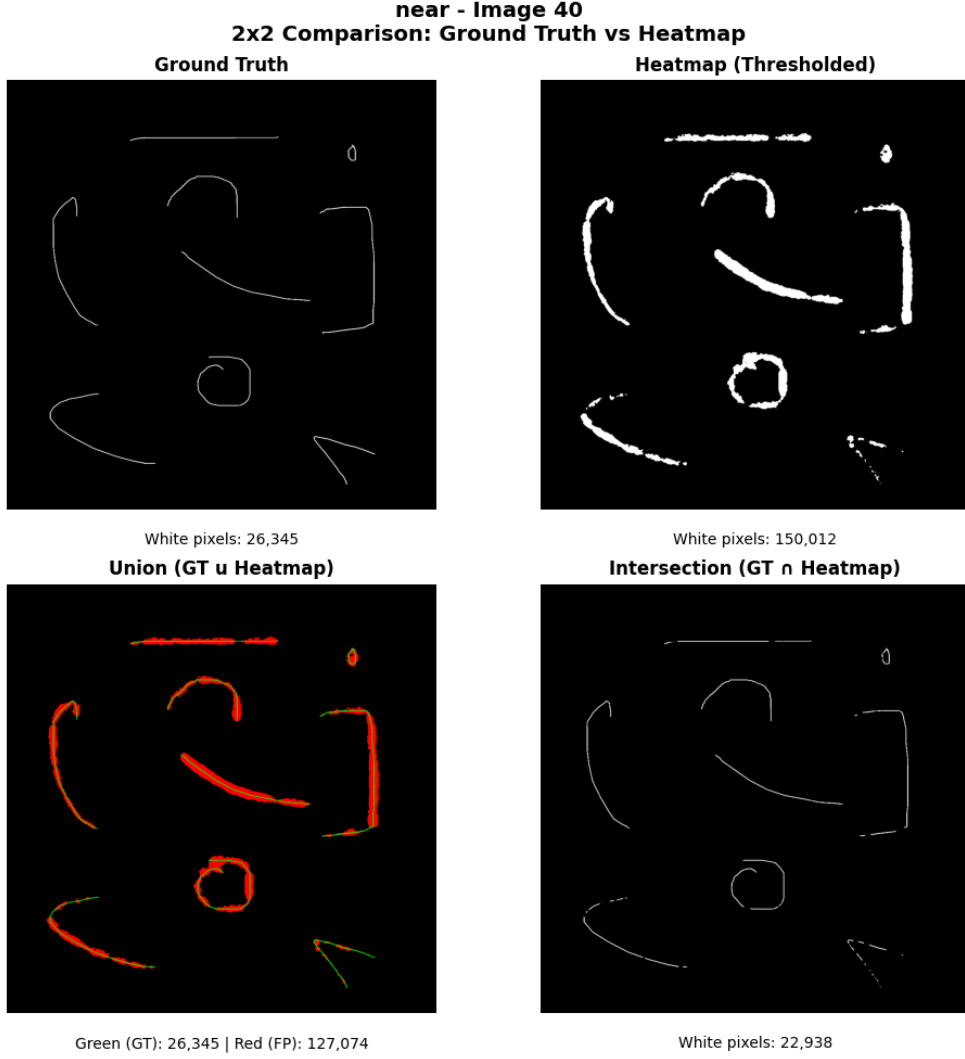


Figure 4.8: Near-field image-space inputs used for training and testing (ground-truth comparison).

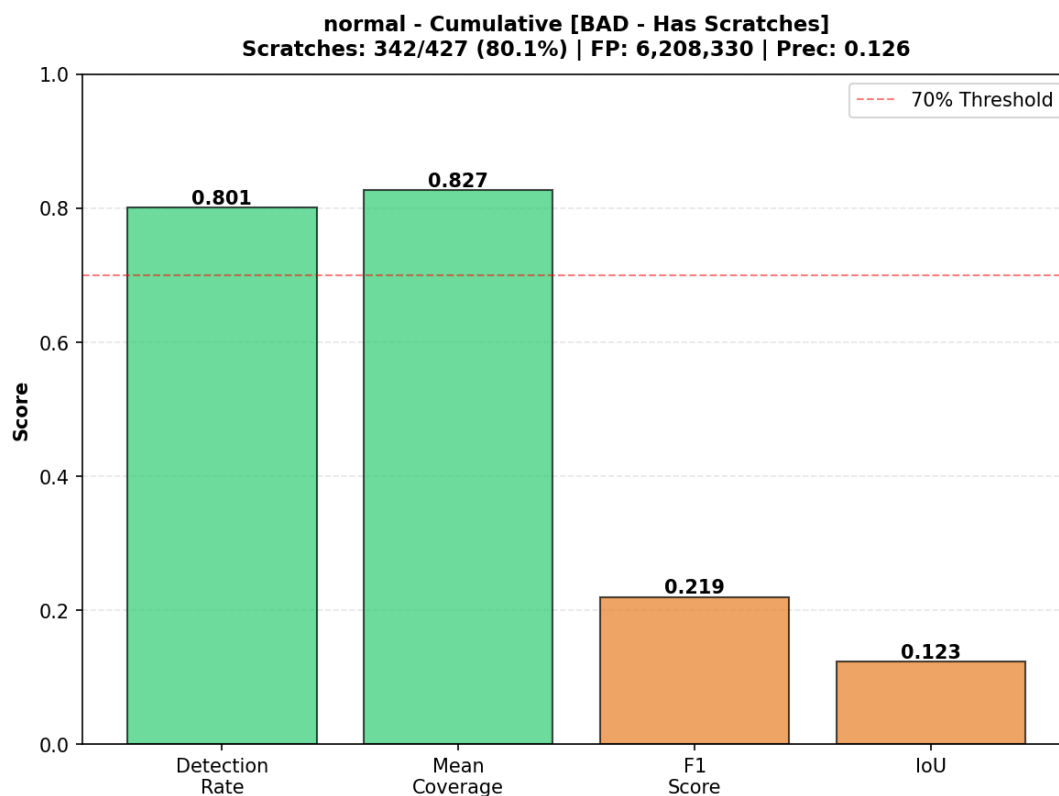


Figure 4.9: Ground truth normal map inputs used for training and testing.

4.4.3 Direct image-space inputs

The near-field image-space input gave the best anomaly-detection performance of any representation, finding essentially every scratch. Figure 4.10 shows this on the deep-scratch test set, and Figure 4.11 on the shallow-scratch set, where the image-space advantage is even clearer.

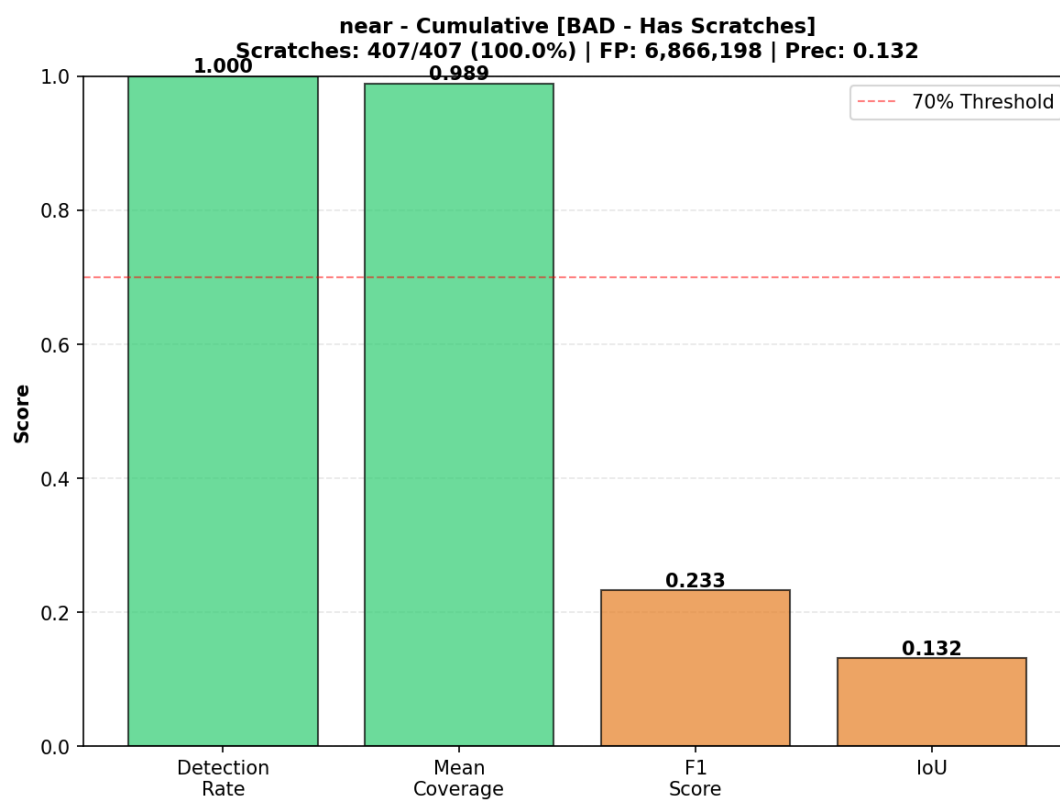


Figure 4.10: Near-field image-space inputs on the deep-scratch test set.

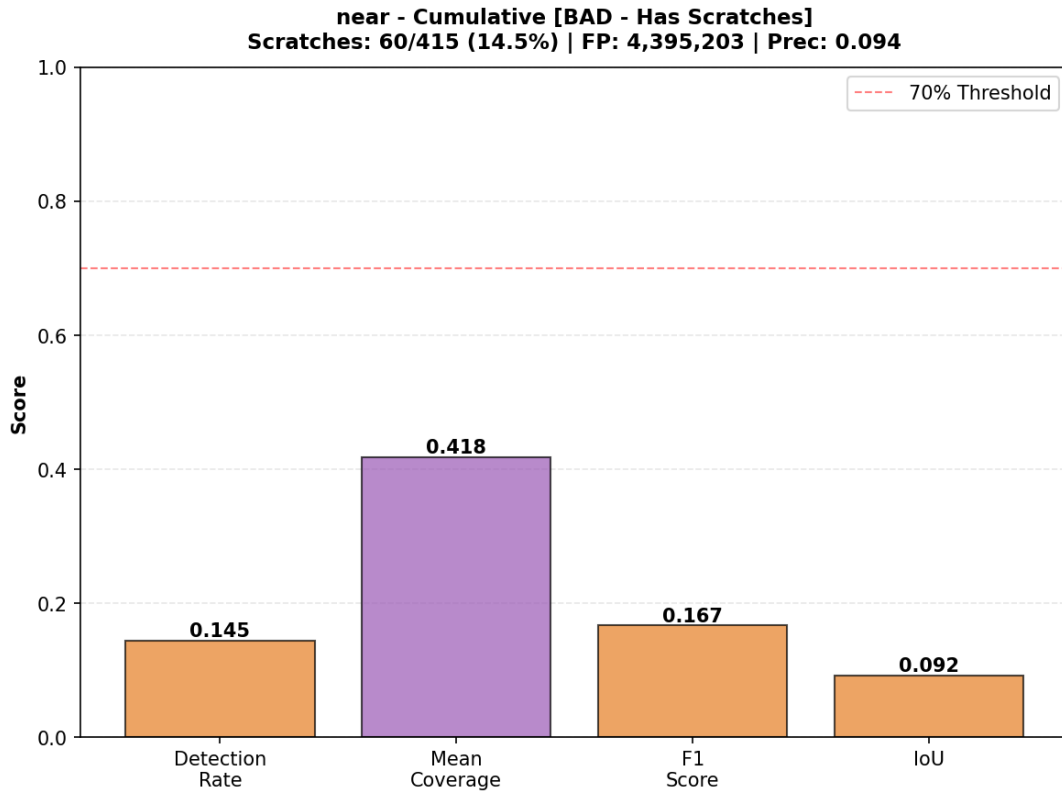
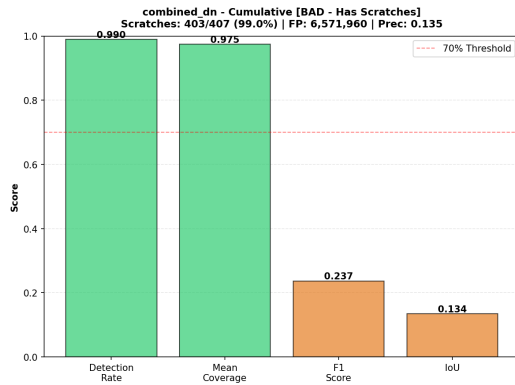


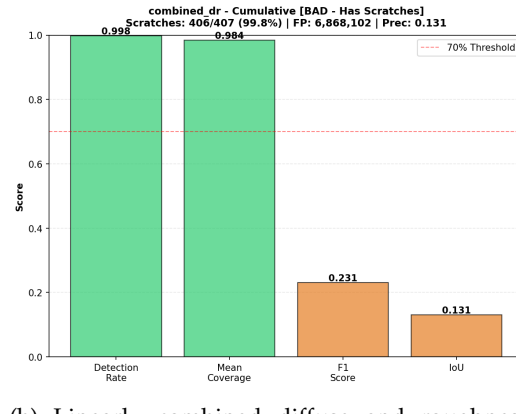
Figure 4.11: Near-field image-space inputs on the shallow-scratch test set.

4.4.4 Combined SVBRDF results

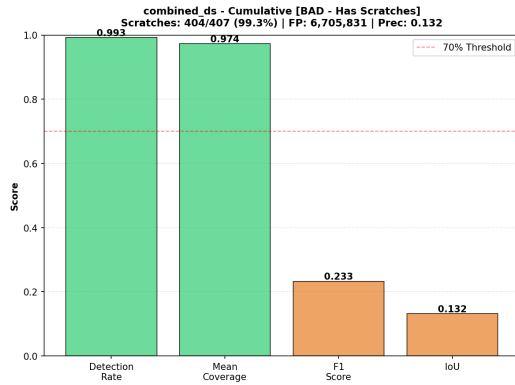
Combining maps means fine-tuning one detector per map and then linearly combining their anomaly heatmaps before thresholding. Figures 4.12 and 4.13 show a selection of these combinations for the deep- and shallow-scratch sets, and the full set is given in Appendix A.2.3. Combining maps occasionally raised the F1 and IoU scores slightly above the near-field image-space input, but never improved the detection rate, which was already near-complete for the image-space input. The clearest effect is on the shallow-scratch set, where combining all four maps gives the best scores of any combination.



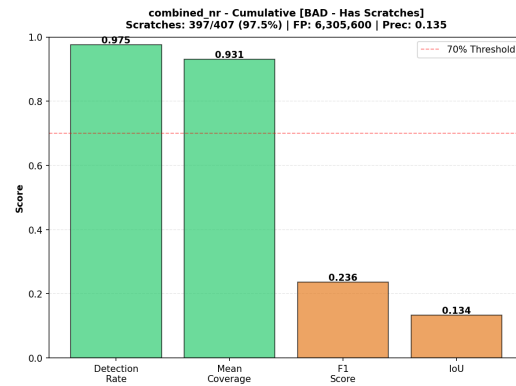
(a) Linearly combined diffuse and normal maps.



(b) Linearly combined diffuse and roughness maps.

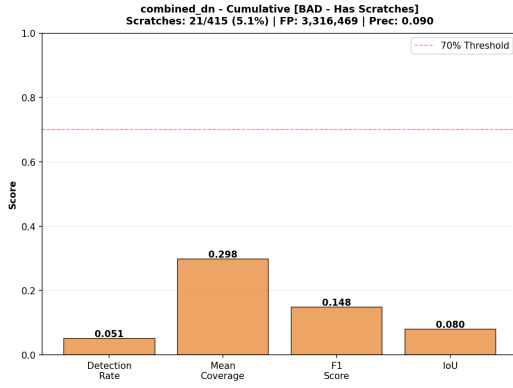


(c) Linearly combined diffuse and specular maps.

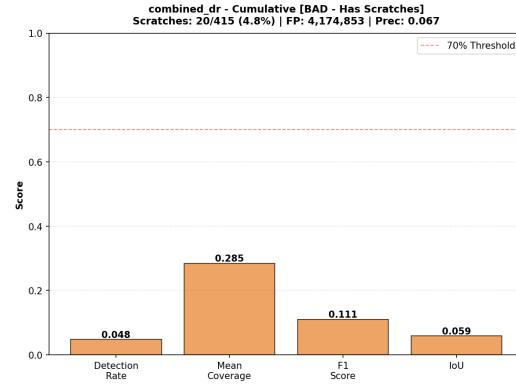


(d) Linearly combined normal and roughness maps.

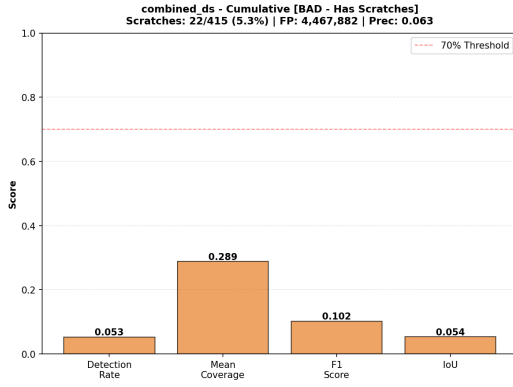
Figure 4.12: Some of the results from combining different SVBRDF representations heatmaps.



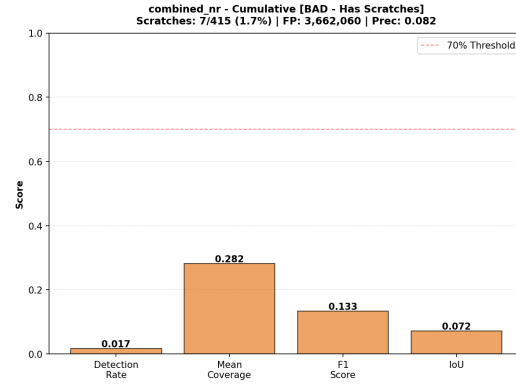
(a) Linearly combined diffuse and normal maps with much less noticeable scratches.



(b) Linearly combined diffuse and roughness maps with much less noticeable scratches.



(c) Linearly combined diffuse and specular maps with much less noticeable scratches.

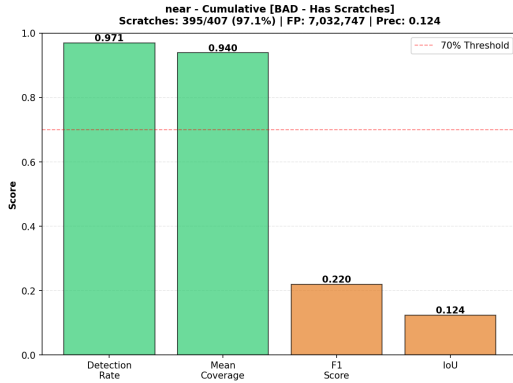


(d) Linearly combined normal and roughness maps with much less noticeable scratches.

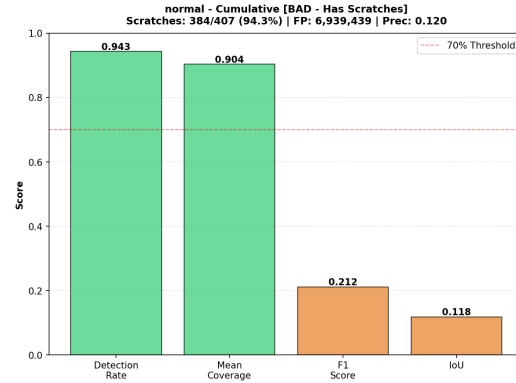
Figure 4.13: Some of the results from combining different SVBRDF representations heatmaps with less aggressive scratches in the testing data.

4.4.5 Detection of material not present during training

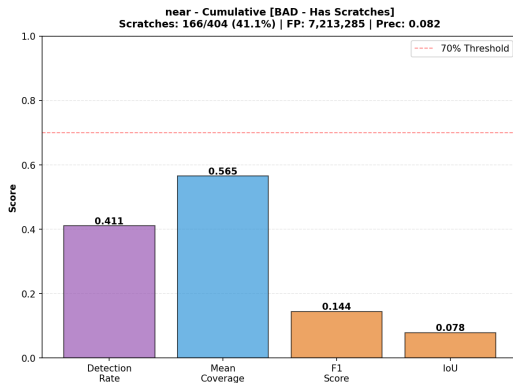
This experiment tests whether a detector trained on one material still finds scratches when the material's appearance changes while its geometry stays the same. Three cases are shown in Figure 4.14: a material almost identical to the training material, a different material from the same category (wood) with a very different appearance, and the training material with only its diffuse map swapped for that of a very different material. In the first case the near-field image and the normal map perform comparably, but as the appearance diverges from the training material the normal-map representation increasingly outperforms the image-space input, reaching roughly 50% higher performance in the diffuse-swap case. This is the main evidence for RQ3 and is analysed in Section 5.1.



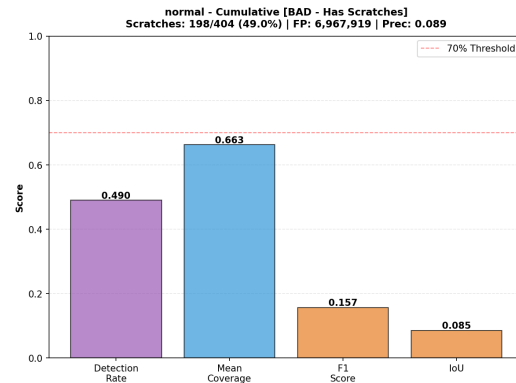
(a) Tests on material similar to the training material.



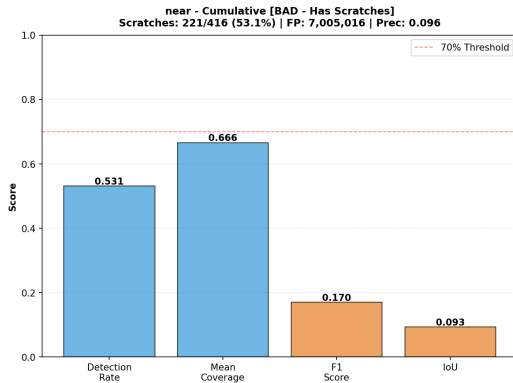
(b) Tests on normal maps with material similar to the training material.



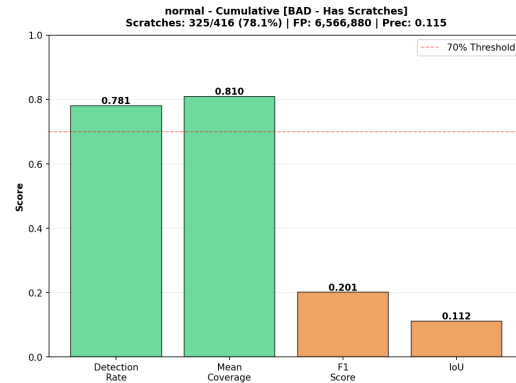
(c) Tests on material which was in a similar category (wood), but very dissimilar visually to the training material.



(d) Tests on normal maps with material which was in a similar category (wood), but very dissimilar visually to the training material.



(e) Tests on material which was the same as the training material, except for the diffuse map, which was changed to a very dissimilar materials diffuse map.

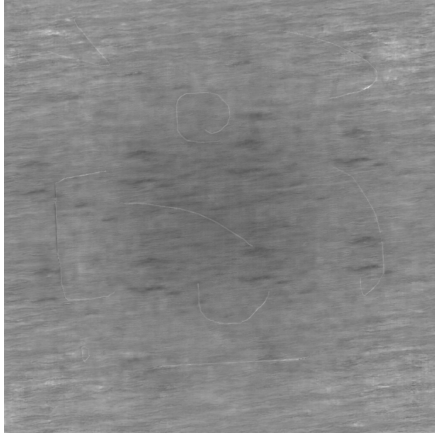


(f) Tests on normal maps with material which was the same as the training material, except for the diffuse map, which was changed to a very dissimilar materials diffuse map.

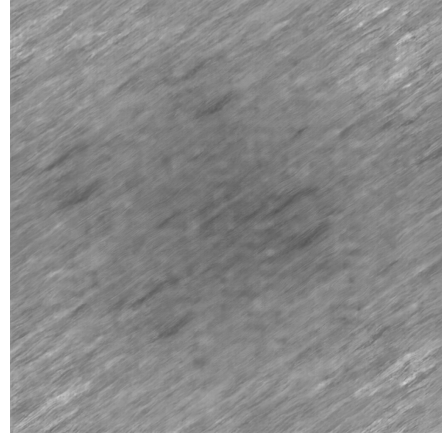
Figure 4.14: Cross-material transferability tests (RQ3): anomaly-detection metrics for near-field image inputs (left column) and normal-map inputs (right column) under three appearance changes, namely a similar material, a visually dissimilar material of the same category, and a diffuse-map swap.

4.5 Other results

This section collects supporting figures referenced from the discussion: example output masks for the traditional methods, the qualitative anomaly-detection heatmaps, a comparison of roughness maps from defective and non-defective samples, and an example of the shadows cast by thicker real subjects.



(a) Roughness map created from defective material.



(b) Roughness map created from non-defective material.

Figure 4.15: Roughness maps generated by NFPLight from a defective (a) and a non-defective (b) synthetic material.

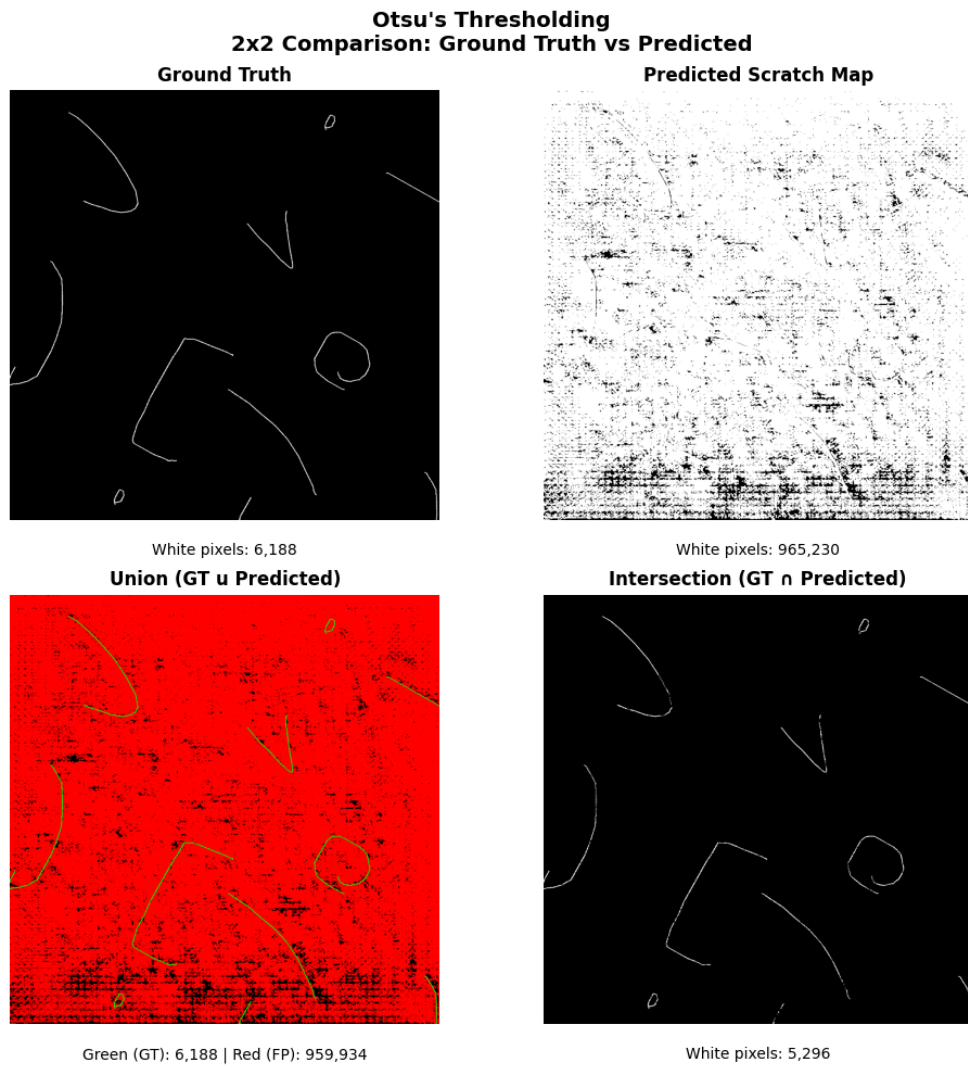


Figure 4.16: Result of Otsu thresholding on a normal map created by NFPLight.

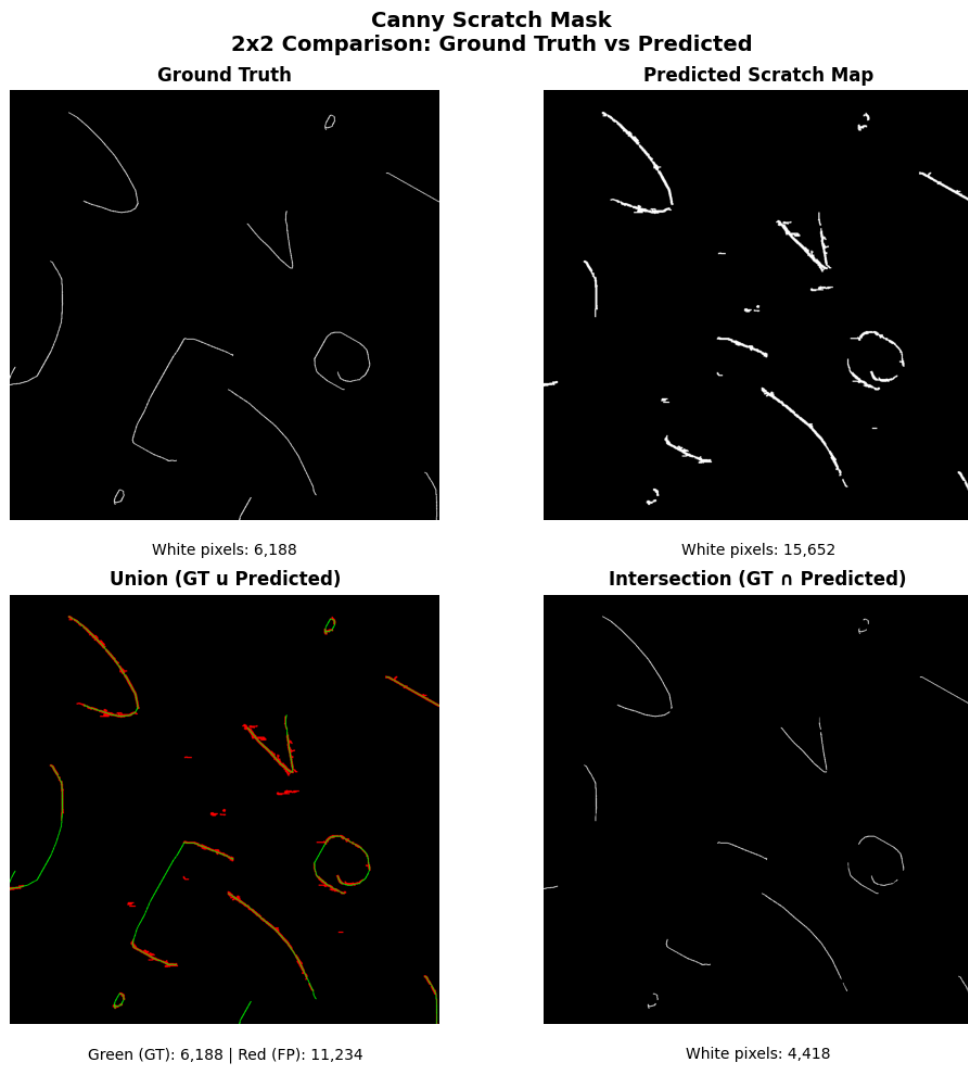
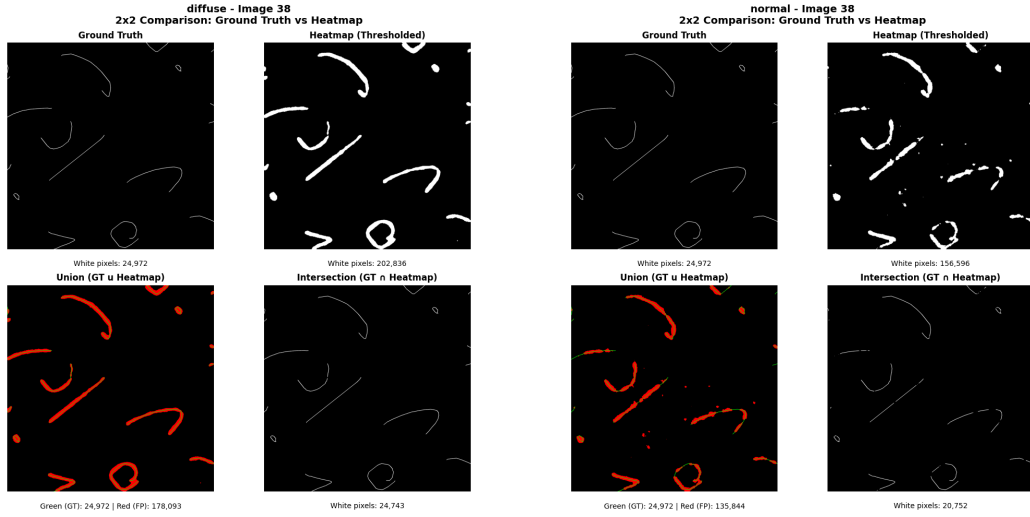
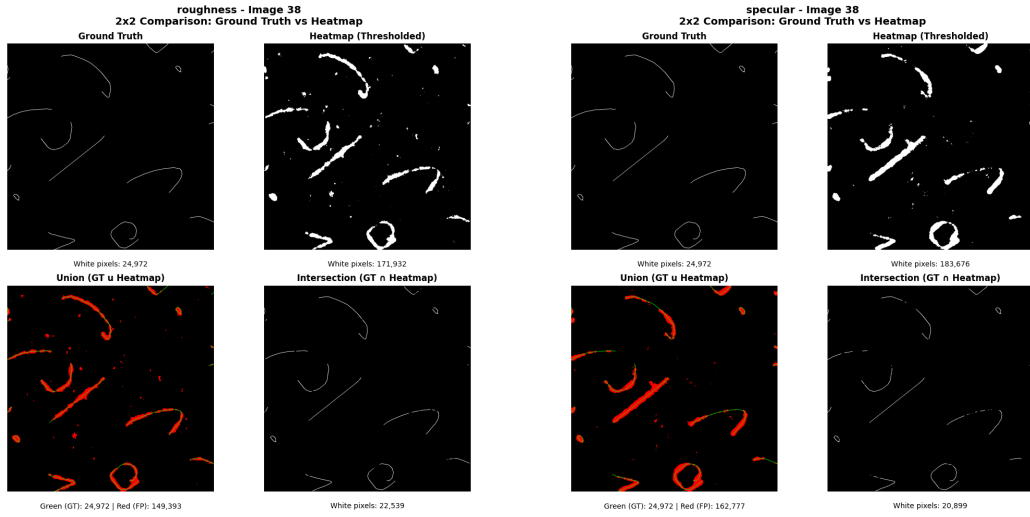


Figure 4.17: Result of Canny edge detection on a normal map created by NFPLight.



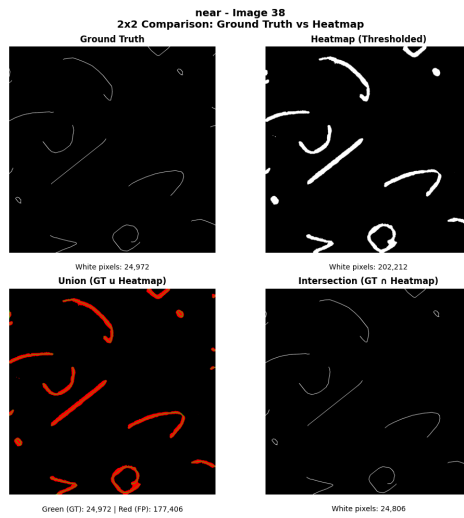
(a) Diffuse map anomaly detection compared to ground truth.

(b) Normal map anomaly detection compared to ground truth.



(c) Roughness map anomaly detection compared to ground truth.

(d) Specular map anomaly detection compared to ground truth.

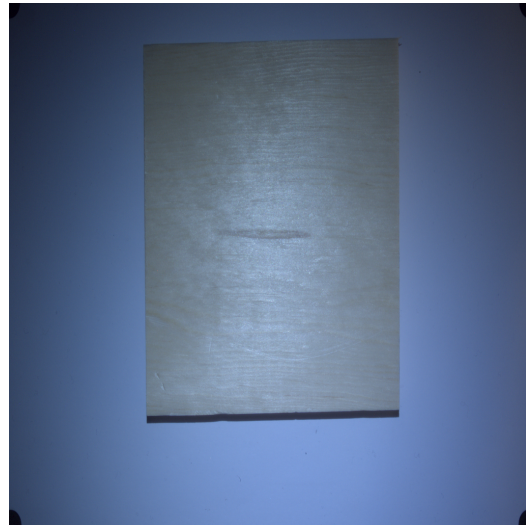


(e) Near-field image anomaly detection compared to ground truth.

Figure 4.18: Images showing thresholded heatmaps of different inputs into the anomaly detection system, and their intersection with ground truth.



(a) Real near-field image with a wood subject.



(b) Real far-field image with a wood subject.

Figure 4.19: Images showcasing shadows created with thicker subjects, where the bottom of the subject casts a shadow onto the background.



5 Discussion

This chapter discusses the results from the work done in this thesis, with the purpose of backing up the conclusions to the research questions asked in section 1.4.

5.1 Results

5.1.1 Image Quality metric

As can be seen in figure 4.1, both normal and roughness maps were mostly identical to their ground truth. Some materials, like wood63 and wood66, did perform worse than other materials for normal maps and roughness respectively. Diffuse maps performed much worse, where not a single resulting diffuse map had a similarity score above 80%.

An important observation is that all SVBRDF maps exhibit low variance across materials, with the largest deviations being approximately 5 percentage points from the mean. This relatively small spread indicates that the inverse rendering process is consistent and does not produce large errors for wood materials. Such consistency is important for anomaly detection, as it ensures that variations in the representation are primarily due to surface properties rather than instability in the reconstruction process.

In figure 4.2, the quality (scratched vs not scratched) of the maps did little to affect the accuracy of the produced maps, as the results vary little. What is important to take into account is that the scratches are a relatively small part of the whole SVBRDF map. Looking at the created roughness map from a scratched material 4.15a vs non-scratched material 4.15b, they are almost exactly the same, other than the tiny scratches. Those scratches are not present in either of the generated maps ground truth, as the defects they experienced were only applied to the normal map of the material. Therefore, even if the scratches were recorded on the diffuse maps, it hardly made a difference in the validation score, since they are such a small part of the total area of the map.

One more noteworthy result is that the normal map comes out almost perfect in nearly all the materials resulting SVBRDF maps. This doesn't come as a surprise, as a flat material doesn't

have much of an angular variance of the surface, meaning that nearly all normal maps are at almost the exact same angle in nearly all pixels.

Assuming the maps were perfect, the results of the edge detection methods could improve, but is not guaranteed. While a perfect remake of the SVBRDF maps would produce the same images under the same lighting condition as the ground truth, this thesis does not aim to re-render the SVBRDF maps, but instead use them for edge detection, which does not require the results to be great for rendering.

Real SVBRDF maps

Figure 4.4 shows an example of real image-space inputs that were used as inputs to create SVBRDF maps, and as can be seen, they are not great representations. As stated in the method (Section 3.7), the results cannot be tested analytically, but through visual inspection of the resulting map, it is easy to conclude that the maps do not represent the overall features of the subject well. The reason may be that the images were not taken under correct parameters, but more likely is that there is a domain gap in the training data for the NFPLight model and the data captured for this thesis. The largest issue, which is discussed later in the discussion of methodology (Section 5.2), might be that the way the data was collected caused a domain gap between NFPLight’s training data and the testing data captured for this thesis.

5.1.2 Edge detection using traditional methods

From the testing performed without applying Sobel filter preprocessing and using Otsu thresholding, Canny edge detection, as well as Marr-Hildreth edge detection, it can be seen that Canny edge detection outperformed the other methods in nearly every metric. For the direct image-space inputs, it was the only method that found more than 30% of the scratches present in the image using a 70% coverage threshold. However, it was outperformed by Otsu thresholding in finding the scratches on the normal maps, but only because Otsu thresholding is oversensitive and fills in most of the image as a positive for defects. This can be seen in figure 4.16. Compare that to Canny edge detection in figure 4.17, which with this particular case had a few false positives, but still performed with much higher accuracy than Otsu thresholding. The accuracy of the methods could already be observed looking at the F1 and IoU score, which Canny edge detection scores the highest on in every metric.

What can be observed comparing the direct image-space inputs to the SVBRDF normal map is that for the near image, it allows detection of scratches with a higher accuracy and more frequently than using the generated normal maps, as well as the ground truth normal maps. This comes from the fact that more scratches were detected using the near-field images as inputs, and with a higher F1- and IoU-score than the normal maps achieved. This also confirms that with standard edge detection algorithms, it appears that, for classical edge detection methods, direct image-space inputs are more effective than normal maps, even when the normal maps are 100% accurate.

Looking towards the scratch free surfaces, it can be observed again that Canny edge detection performs with much higher accuracy for every input, where near-field images had the best performance, with the normal maps as input being close, but still predicting on average about 5% of the image to contain scratches, even though it is scratch free.

Looking at the results which used Sobel filtering as a preprocessing step, it simply worsened the results’ accuracy. As a preprocessing step, it failed to give any real upside, as the increase in detection rate simply came from the model being over-sensitive and labeling the entire image as having scratches.

5.1.3 AI anomaly detection using base SVBRDF map and direct image-space inputs

Using the proprietary anomaly detection tool for anomaly detection, it can be observed that the near-field direct image-space input performed best in both accuracy and frequency of found scratches, as every single one was found, with better F1- and IoU score than any other method. Another way to look at this, is that the raw image was better for performance than using the preprocessing step of inverse rendering. Looking at some examples from figure 4.18, it can be seen that even though the results are clearly overlapping perfectly with the scratches, since the mask created from the heatmap is thicker than the actual ground truth, a lot of false positive pixels are defined. This does not mean that the model performs poorly. Observing the actual shape and positions of the false positive pixels leads to the conclusion that the near-field image is very precise, finding every single scratch that was present and tracing each scratch to near perfection, with some extra thickness.

Differently from what was expected however, is that the best performing SVBRDF map was actually the diffuse map, which scored 401/407 total scratches found, with an F1- and IoU score nearly just as good as using the direct image-space input. This is unexpected since the scratches should not show up in the diffuse map assuming the inverse renderer produced the correct SVBRDF maps, but since the scratches show up in every map in one form or another, it just so happens that the proprietary tool is able to use the diffuse map for better results than any of the other maps.

Looking instead at the surfaces where the scratches are far less noticeable, like in figure 4.13, it can be seen that the near-field image-space input performs best in every measured metric. This in turn suggests that for both very visible and not very visible scratches, image-space solutions are simply better for the proprietary tool to detect scratches for. This might be because the dataset on which the tool's underlying network was trained consisted of image-space inputs, which would also explain why the diffuse maps performed better overall than other SVBRDF maps, as diffuse maps are more similar to image-space than roughness or normal maps.

5.1.4 Combined SVBRDF inputs

Combining the inputs gave very interesting results, where of course they could never outperform detection frequency, since near-field images were 100% accurate in that respect. However, for the deep scratches in Figure 4.12, combining the maps did in some cases end up yielding a slightly higher F1- and IoU-score than simply using the near-field image, while also having a near 100% detection rate of scratches. The small increase in scores is not very impressive though once you take into account the previously discussed thick segmentation that is gathered from thresholding the heatmaps, which means that the performance gain is likely just from the masks trimming their edges a tiny amount when combined.

The surfaces with shallower scratches, shown in Figure 4.13, show the underlying fact however, that the image-space input performs better when the scratches actually become difficult to spot. What is interesting however is that the combination of all maps (diffuse, normal, roughness and specular) actually increased the F1- and IoU-score above any other combination, suggesting that error is decreased as more maps are added together. This is not what is expected of 100% perfectly generated svbrdf maps, since only the normal was affected by the scratches, but since the scratches can be seen in every map after inverse rendering, combining the anomaly detection heatmaps actually grants more precise results as compared to using any other combination.

In terms of Research Question 2, this means that no single map and no smaller combination matched the full set. The accuracy improved steadily from single maps, to pairs, to the full four-map combination, which scored highest of all. The effect is real but small, and it never closed the gap to the near-field image-space input, so it is best read as a minor refinement rather than a practical gain. It may also be partly an artifact of linearly combining and thresholding the heatmaps, which trims the mask edges, rather than a genuine increase in information.

5.1.5 Scratch detection of material not present during training

When testing on a material that is the same as the training data, where color, normals, and roughness differ only slightly due to augmentation, both the near-field image representation and the normal map representation performed comparably (Figures 4.14a and 4.14b). The image-space model achieved marginally higher performance. This indicates that when the test data closely matches the training distribution, both representations are sufficient for detecting scratches.

Repeating the test on a different material from the same category (wood) that looked very different from the training material gave more notable results. The generated normal-map representation began to outperform the image-space model, as seen in Figures 4.14d and 4.14c. The improvements in F1- and IoU-score were relatively small, but the detection rate increased by nearly 20%, with average coverage rising to roughly 66% per scratch from 57%. This suggests that geometry-based representations are less sensitive to changes in visual appearance than image-space inputs.

In the final test, the material was kept geometrically identical to the training material, but its diffuse map was replaced with that of a very different material. This changes the surface's appearance in image space while leaving its geometry, and therefore its normal map, unchanged. Here the normal-map-based model showed a substantial advantage over the image-space model, achieving roughly 50% higher performance, with improved F1- and IoU-scores and about 20% higher mean coverage per detected scratch, as seen in Figures 4.14f and 4.14e.

This result demonstrates that when the appearance of a surface changes, but its geometric structure remains the same, models based on geometry-related representations remain robust, while image-space models degrade. Taken together with the previous experiment, this directly supports Research Question 3. Geometry-related components of SVBRDF representations can generalize across materials with different visual appearance but similar physical surface structure, leading to improved anomaly detection performance compared to image-space methods.

5.2 Limitations

The methods used in this thesis worked towards the purpose of answering the research questions, but there were flaws that could be improved. Firstly, it might have been relevant to test more inverse rendering models. While many were considered, and were also in fact tested to some degree, they were not as thoroughly tested as the ones brought up in this thesis. Additionally, a lot of the models that report great results in their papers happen to not be open-source, which means there were no way to test them at all, unless the authors of said models would provide it for further research.

For real data however, it might have been better to use a model that only needed one input image, since most of the difficulty with the real data collection came from the fact that there needed to be two images that had to line up as closely as possible, while also demanding

very different conditions (near- and far-field positions). The process would have sped up significantly given a single-shot method, and there would be fewer parameters, which would make for fewer opportunities for errors to occur. Also, the real data would preferably be annotated, since that would allow for the real data to be used beyond qualitative analysis and instead be a part of the quantitative results. The data collected also differed in one way from how Wang et al. captured theirs, which was that their data only contained the subject in the images, while the data captured for this thesis did not. All the real data collected had large parts of the background in it, which may be problematic because of the following reasons:

1. The background does not hold any interesting information about the state of the material of the subject.
2. The background takes up space in the resulting data, meaning that the subject has a lower resolution dedicated to it than what is possible.
3. The transition from background to subject creates sharp outlines, and may cast a shadow.

Each reason is considered below, along with why it specifically may have caused issues for the inverse renderer, which in turn gives poor inputs to the AI anomaly detection tool.

1. The proprietary anomaly detection tool might put an overemphasis on the background being the exact same, but depending on the environment, indirect illumination might alter the appearance of the surrounding background. This can lead to the model detecting changes in the background as anomalies, even though they are unrelated to the material itself.
2. The subject having a lower resolution might cause the inverse renderer to miss out on surface features like small scratches or indents, which in turn means that the proprietary anomaly detection tool might not be able to spot the defects on the normal map, as well as possibly not being able to spot the defects on the subjects from static image-space inputs. In both of these cases, a better resolution might have improved the models' ability to account for the defects on the material.
3. The shadows that may be cast is seen in figure 4.19, where the bottom half of the subject can be seen casting a shadow on the background. This became an issue later in anomaly detection, since the inverse renderer correctly identified that there was a steep angular change at the transition from the subject and the background. This change looks a little different for every subject, and the in house anomaly detection system did not learn what separates the good subjects from the bad ones, and most of the time identified the transitions as defects. It is also worth noting that the transitions on the side of the subject also caused these types of normal map steep angular changes, even though there is no shadow cast there, but the shadow did show up in diffuse maps as well as in the specular and roughness maps.

Lastly, it must be pointed out that NFPLight might not have been trained on data that looked like the data created for this thesis, due to the fact that there were too much of the background present. This could have contributed to the aforementioned domain gap, where even though NFPLight may be able to get decent results from their testing, the data collected for this thesis were not able to produce similar results due to its dissimilarity.

F1- and IoU-scores are also of concern, as it was shown previously that even though the scratches are clearly found, the scores are still below 30%. A better way to go about it might be

to do the same method as with the threshold method, where connected-component labeling is applied to the predicted scratch bitmask. If there is an overlap, the scratch is counted as found, and all overlapping pixels are marked as true positive matches, and every other pixel with the same label as the successfully predicted scratch are marked as true negative. This also means that if the bitmask were to be pure white, or some other bogus result, like in figure 4.16, then only a single scratch would be found, and the other ones would be marked as false negative, giving bad results for being overly positive, but also giving fair results for predicting the scratches to be thicker than what ground truth describes.

5.2.1 Replicability

The replicability of this exact thesis depends on which parts that are being replicated. While the traditional methods can easily be reimplemented using OpenCV and Python, to replicate the proprietary anomaly detection result, the actual proprietary anomaly detection tool needs to be used. Custom functions were also written specifically to port the data from the proprietary tool to disk for evaluation. This is not a feature that is kept within the tool, but rather on a custom SDK that was created for this particular project. Therefore, replicating the AI anomaly tool section of the project, which is where the majority of the research questions were answered, is very difficult.

5.3 The work in a wider context

The aim of the work of this thesis could have large impacts both societally and environmentally.

5.3.1 Societal impact

Assuming further work in this field were to find ways to reliably detect scratches on complex surfaces such as glass, manual inspection might be regarded as unnecessary, leading to job displacement in manufacturing sectors. However, this task is a repetitive, possibly straining and mentally taxing, and automating it would lead to less of a bottleneck in manufacturing.

5.3.2 Environmental impact

By being able to detect scratches more accurately using geometry-related components like surface normals, which are less affected by lighting than standard photos, prevention of labeling scratch free subjects as being scratched may lead to fewer good subjects being rejected during manufacturing.

Being able to train a model on a small set of data and reuse the geometrical information across materials also reduces the computational energy required to train new models for every single new material type.



6 Conclusion

This section goes over the answers to the research questions described in section 1.4, based on the results and discussion held in sections 4 and 5.

6.1 Summarization of purpose

The purpose of this thesis work was to determine if SVBRDF-based surface representations, generated using inverse rendering, provide a more effective and robust basis for detecting small surface defects, particularly scratches, than conventional image-space analysis. To this end, the following research questions were asked and answered.

1. Does representing surfaces in SVBRDF space improve the detectability of small topological surface defects (scratches) compared to direct image-space analysis?
2. Are there combinations of SVBRDF components (e.g., surface normals, roughness) that increase the accuracy of scratch detection compared to using the full SVBRDF representation or individual components alone?
3. Can geometry-related components of SVBRDF representations, such as surface normals, be reused across different materials with similar physical surface structure for scratch detection?

6.2 Answering the research questions

The answers to each research question, based on results and discussion, is the following:

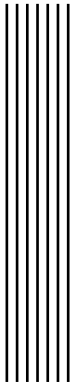
- For traditional methods, where Canny edge detection performed better than all other traditional methods for all forms of input representation, the image-space inputs performed best for all metrics. For the proprietary anomaly detection tool, the image-space inputs again performed better in every aspect. So the detectability of small topological surface defects, specifically scratches, are not improved by using SVBRDF-components by themselves. This was also true when comparing to ground truth normal maps.

- No combination of SVBRDF maps beat the near-field image-space input, which kept a near 100% detection rate for both the deep and the shallow scratches. For the deep scratches the combinations made little difference. For the shallow scratches, which are harder to spot, combining all the maps (diffuse, normal, roughness and specular) did give a higher F1- and IoU-score than any other combination. This happened even though only the normal map was affected by the scratches, because the scratches can be seen in every map after inverse rendering.
- Using normal maps during training did in fact increase the accuracy of the scratch detection compared to using image-space inputs. This proves that using geometry-related components of SVBRDF representations can in fact improve accuracy when detecting scratches on materials with similar physical surface structure.

6.3 Further work

A few ideas for furthering the work done in this thesis are as follows:

- Test the model on materials other than wood, such as metal, glass, and stone. Because these materials differ from wood in reflectance and translucency, they would show whether the cross-material geometry transfer found for RQ3 generalises beyond a single material family.
- Use a more forgiving validation method for accuracy, such as the connected-component approach described in the discussion (Section 5.2). The F1- and IoU-scores used here penalise the small pixel area of thin scratches heavily and likely understate how well the scratches were actually located.
- As newer and better inverse rendering models appear, repeating these experiments with them may yield different conclusions. In particular, a model that reconstructs sharper, more accurate normal maps could narrow the gap between estimated and ground-truth maps seen here, which was a key factor limiting the SVBRDF-based approach.



Bibliography

- [1] Miika Aittala, Tim Weyrich, and Jaakko Lehtinen. “SVBRDF Capture from a Single Image and Sparse Normals”. In: *ACM Transactions on Graphics* 34.6 (2015), 110:1–110:13. DOI: 10.1145/2816795.2818131.
- [2] John Canny. “A Computational Approach To Edge Detection”. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on PAMI-8* (Dec. 1986), pp. 679–698. DOI: 10.1109/TPAMI.1986.4767851.
- [3] Robert L Cook and Kenneth E Torrance. “A reflectance model for computer graphics”. In: *ACM Transactions on Graphics (TOG)* 1.1 (1982), pp. 7–24.
- [4] Valentin Deschaintre, Miika Aittala, Fredo Durand, George Drettakis, and Adrien Bousseau. “Single-image SVBRDF capture with a rendering-aware deep network”. In: *ACM Transactions on Graphics (TOG)* 37.4 (2018), pp. 1–15.
- [5] doxyGen. *Canny Edge Detection*. Last accessed March 12th 2026. 2026. URL: https://docs.opencv.org/4.x/da/d22/tutorial_py_canny.html.
- [6] DUAN GAO, Xiao Li, Yue Dong, Pieter Peers, Kun Xu, and Xin Tong. “Deep inverse rendering for high-resolution SVBRDF estimation from an arbitrary number of images”. In: *ACM Transactions on Graphics (TOG)* 38.4 (July 2019), p. 134.
- [7] James T Kajiya. “The rendering equation”. In: *SIGGRAPH Computer Graphics* 20.4 (1986), pp. 143–150.
- [8] David Marr and Ellen Hildreth. “Theory of edge detection”. In: *Proceedings of the Royal Society of London. Series B. Biological Sciences* 207.1167 (1980), pp. 187–217.
- [9] Fred E. Nicodemus, John C. Richmond, John J. Hsia, Irving W. Ginsberg, and Theodore Limperis. *Geometrical Considerations and Nomenclature for Reflectance*. Tech. rep. NBS Monograph 160. National Bureau of Standards, 1977.
- [10] Nobuyuki Otsu. “A threshold selection method from gray-level histograms”. In: *IEEE Transactions on Systems, Man, and Cybernetics* 9.1 (1979), pp. 62–66.
- [11] Guansong Pang, Chunhua Shen, Longbing Cao, and Anton van den Hengel. “Deep Learning for Anomaly Detection: A Review”. In: *CoRR abs/2007.02500* (2020). arXiv: 2007.02500. URL: <https://arxiv.org/abs/2007.02500>.

- [12] Ravi Ramamoorthi and Pat Hanrahan. "A signal-processing framework for inverse rendering". In: *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. 2001, pp. 117–128.
- [13] Inc. The Mathworks. *Global histogram threshold using Otsu's method*. Last accessed March 13th 2026. 2025. URL: https://se.mathworks.com/help/images/ref/otsuthresh.html#mw_3d0b8fc6-83ae-4190-94c9-1bfa423cbaeb.
- [14] Li Wang, Lianghao Zhang, Fangzhou Gao, Yuzhen Kang, and Jiawan Zhang. "NFP-Light: Deep SVBRDF Estimation via the Combination of Near and Far Field Point Lighting". In: *ACM Trans. Graph.* 43.6 (Oct. 2024). ISSN: 0730-0301. DOI: 10.1145/3687978. URL: <https://doi.org/10.1145/3687978>.
- [15] Robert Woodham. "Photometric method for determining surface orientation from multiple images". In: *Optical Engineering* 19 (Jan. 1980), pp. 139–144.

A

Graphs from results

A.1 Defect detection using traditional methods

A.1.1 No preprocessing

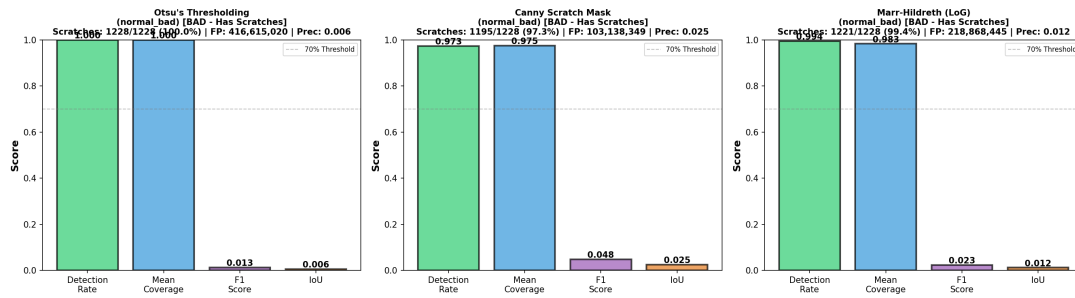


Figure A.1: Graph showing the results of using Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection methods without Sobel filtering as a preprocessing step, with the ground truth normal SVBRDF map. Detection rate counts the percentage of scratches with at least 70% coverage from the detection method, F1 score relates to the method explained in the method Section 2.2.4, and lastly IoU refers to another method explained in the method Section 2.2.4. The subject had defects.

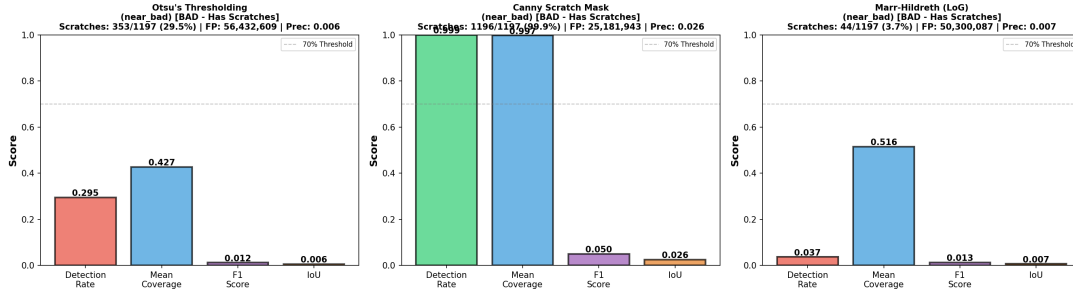


Figure A.2: Graph showing the results of using Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection methods without Sobel filtering as a preprocessing step, with the camera positioned near from the subject. Detection rate counts the percentage of scratches with at least 70% coverage from the detection method, F1 score relates to the method explained in the method Section 2.2.4, and lastly IoU refers to another method explained in the method Section 2.2.4. The subject had defects.

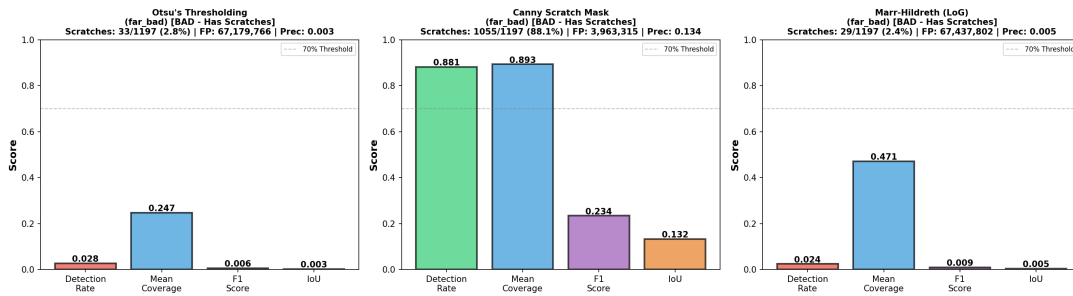


Figure A.3: Graph showing the results of using Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection methods without Sobel filtering as a preprocessing step, with the camera positioned far from the subject. Detection rate counts the percentage of scratches with at least 70% coverage from the detection method, F1 score relates to the method explained in the method Section 2.2.4, and lastly IoU refers to another method explained in the method Section 2.2.4. The subject had defects.

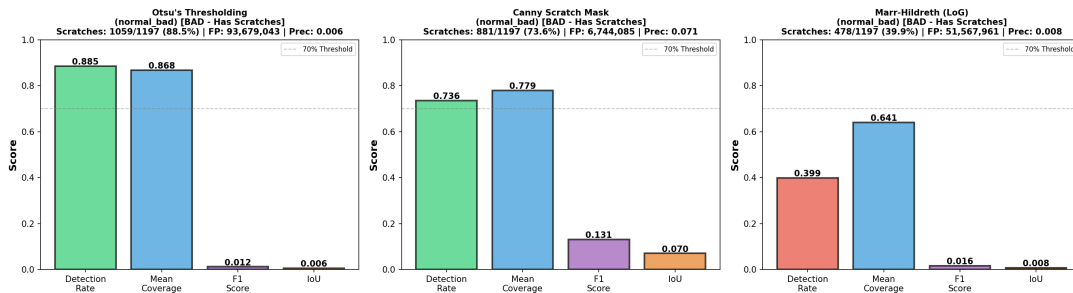


Figure A.4: Graph showing the results of using Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection methods without Sobel filtering as a preprocessing step, with the normal SVBRDF map created with NFPLight inverse rendering. Detection rate counts the percentage of scratches with at least 70% coverage from the detection method, F1 score relates to the method explained in the method Section 2.2.4, and lastly IoU refers to another method explained in the method Section 2.2.4. The subject had defects.

A.1. Defect detection using traditional methods

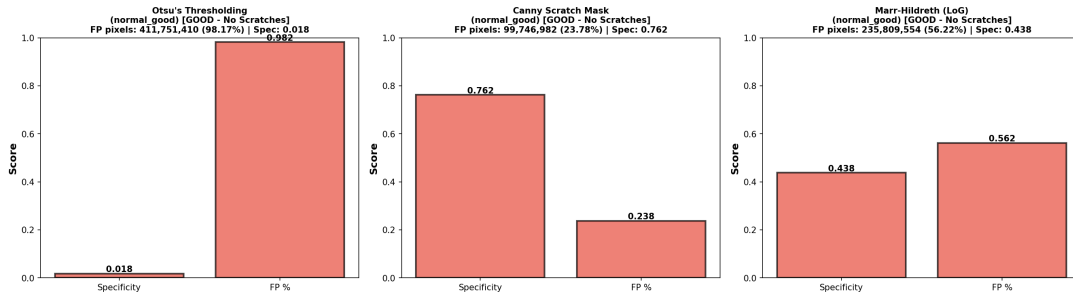


Figure A.5: Graph showing the results of using Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection methods without Sobel filtering as a preprocessing step, with the ground truth normal SVBRDF map. FP rate refers to the number of false positive pixels recorded as compared to true negatives, Accuracy is the inverse of FP rate. The subject had no defects, so no true positives to record.

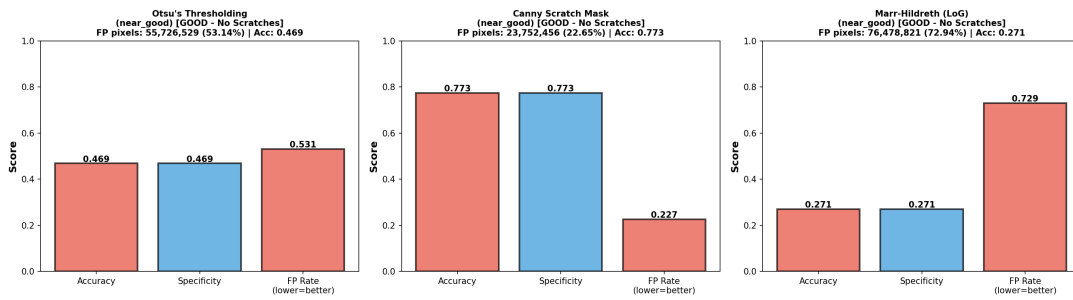


Figure A.6: Graph showing the results of using Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection methods without Sobel filtering as a preprocessing step, with the camera positioned near from the subject. FP rate refers to the number of false positive pixels recorded as compared to true negatives, Accuracy is the inverse of FP rate. The subject had no defects, so no true positives to record.

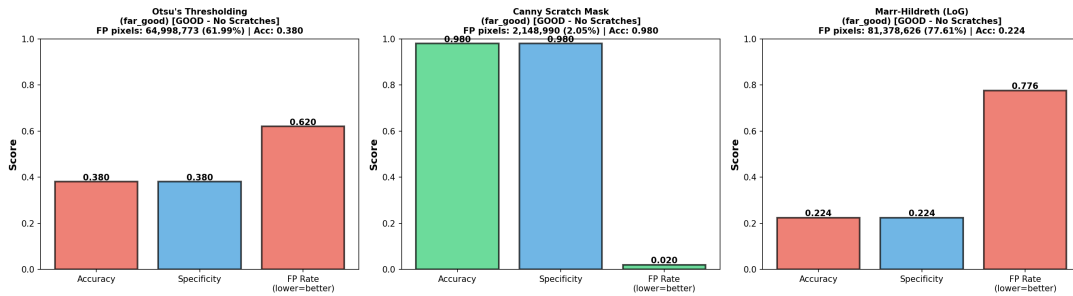


Figure A.7: Graph showing the results of using Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection methods without Sobel filtering as a preprocessing step, with the camera positioned far from the subject. FP rate refers to the number of false positive pixels recorded as compared to true negatives, Accuracy is the inverse of FP rate. The subject had no defects, so no true positives to record.

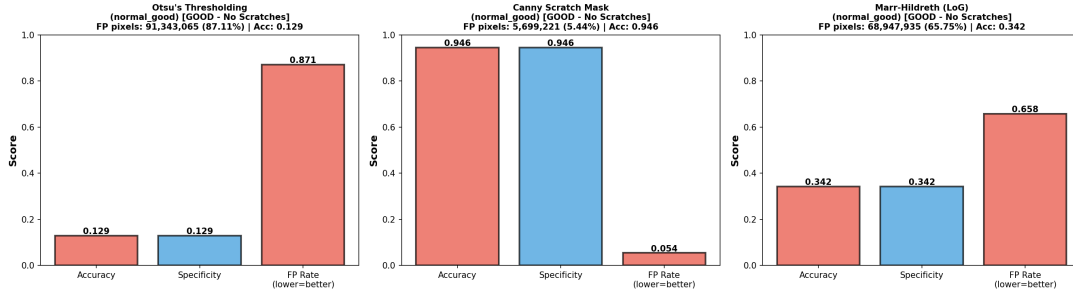


Figure A.8: Graph showing the results of using Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection methods without Sobel filtering as a preprocessing step, with the normal SVBRDF map created with NFPLight inverse rendering. FP rate refers to the number of false positive pixels recorded as compared to true negatives, Accuracy is the inverse of FP rate. The subject had no defects, so no true positives to record.

A.1.2 With preprocessing

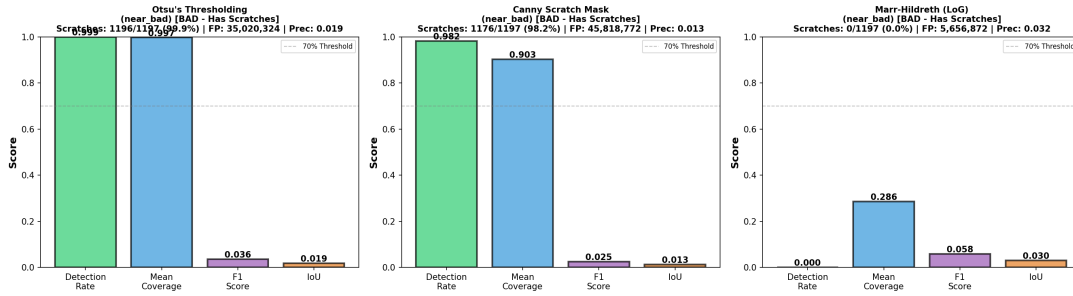


Figure A.9: Graph showing the results of using Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection methods with Sobel filtering as a preprocessing step, with the camera positioned near from the subject. Detection rate counts the percentage of scratches with at least 70% coverage from the detection method, F1 score relates to the method explained in the method Section 2.2.4, and lastly IoU refers to another method explained in the method Section 2.2.4. The subject had defects.

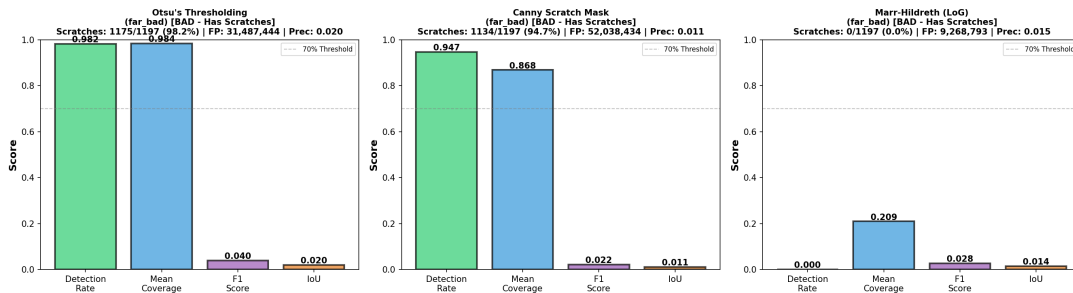


Figure A.10: Graph showing the results of using Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection methods with Sobel filtering as a preprocessing step, with the camera positioned far from the subject. Detection rate counts the percentage of scratches with at least 70% coverage from the detection method, F1 score relates to the method explained in the method Section 2.2.4, and lastly IoU refers to another method explained in the method Section 2.2.4. The subject had defects.

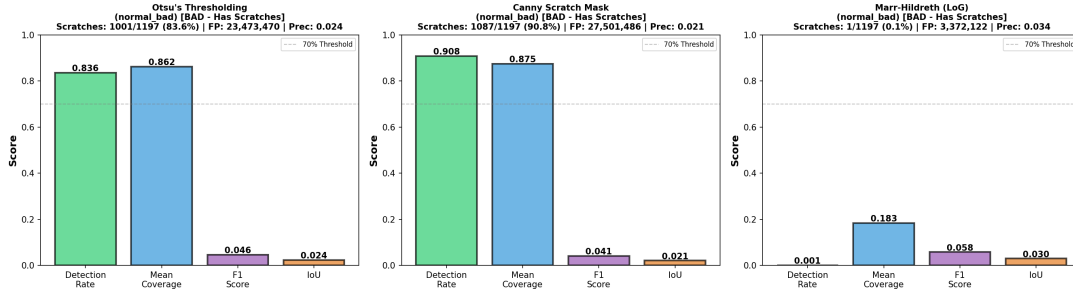


Figure A.11: Graph showing the results of using Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection methods with Sobel filtering as a preprocessing step, with the normal SVBRDF map created with NFPLight inverse rendering. Detection rate counts the percentage of scratches with at least 70% coverage from the detection method, F1 score relates to the method explained in the method Section 2.2.4, and lastly IoU refers to another method explained in the method Section 2.2.4. The subject had defects.

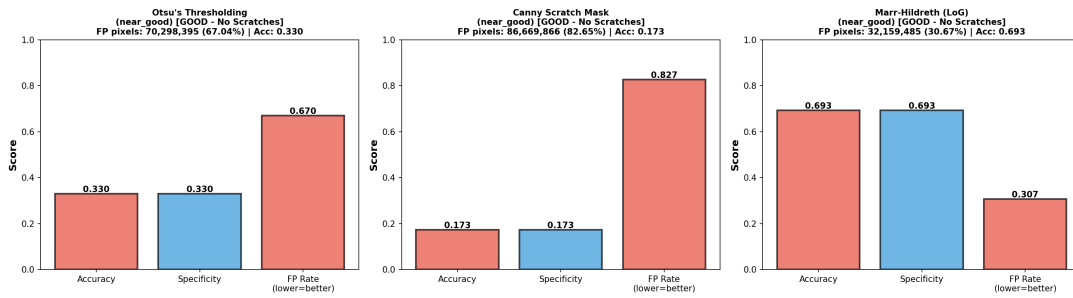


Figure A.12: Graph showing the results of using Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection methods with Sobel filtering as a preprocessing step, with the camera positioned near from the subject. FP rate refers to the number of false positive pixels recorded as compared to true negatives, Accuracy is the inverse of FP rate. The subject had no defects, so no true positives to record.

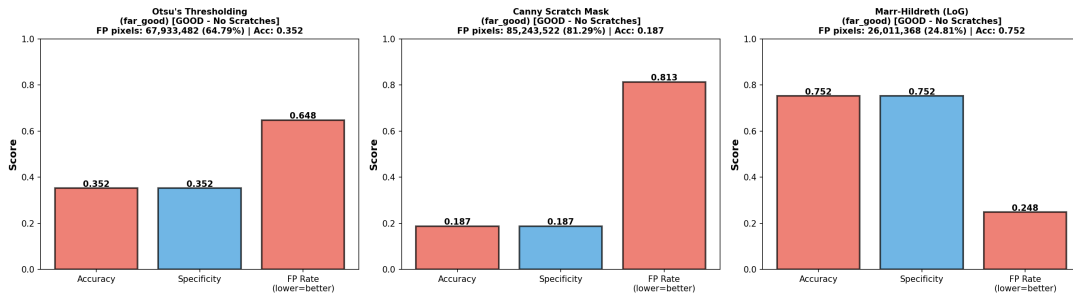


Figure A.13: Graph showing the results of using Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection methods with Sobel filtering as a preprocessing step, with the camera positioned far from the subject. FP rate refers to the number of false positive pixels recorded as compared to true negatives, Accuracy is the inverse of FP rate. The subject had no defects, so no true positives to record.

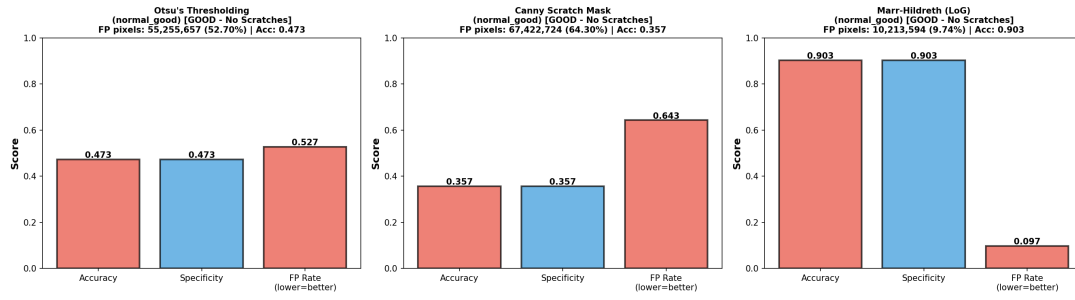
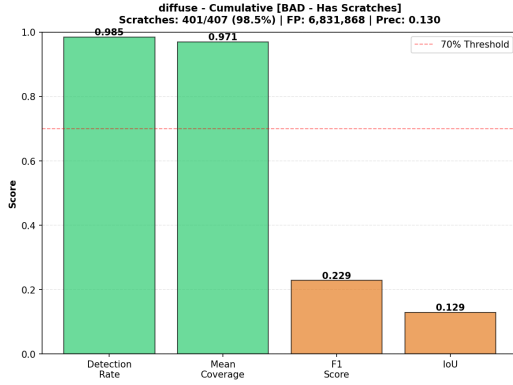


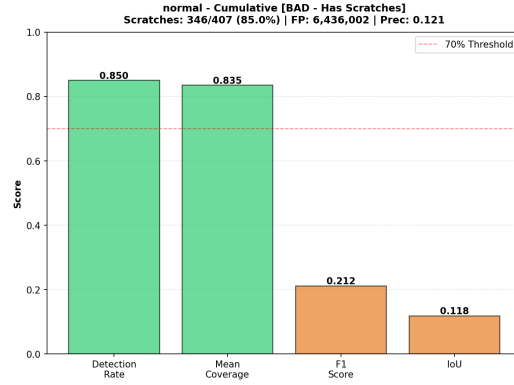
Figure A.14: Graph showing the results of using Otsu thresholding, Canny edge detection, and Marr-Hildreth edge detection methods with Sobel filtering as a preprocessing step, with the normal SVBRDF map created with NFPLight inverse rendering. FP rate refers to the number of false positive pixels recorded as compared to true negatives, Accuracy is the inverse of FP rate. The subject had no defects, so no true positives to record.

A.2 Defect detection using AI anomaly detection

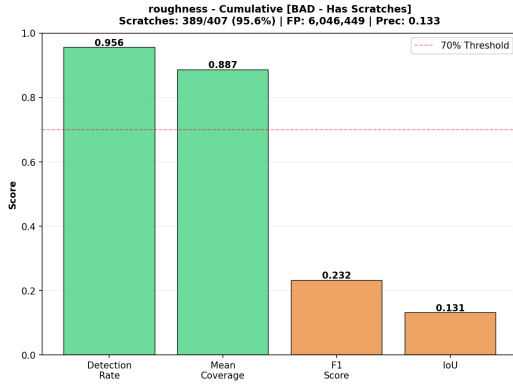
A.2.1 Singular map or image input



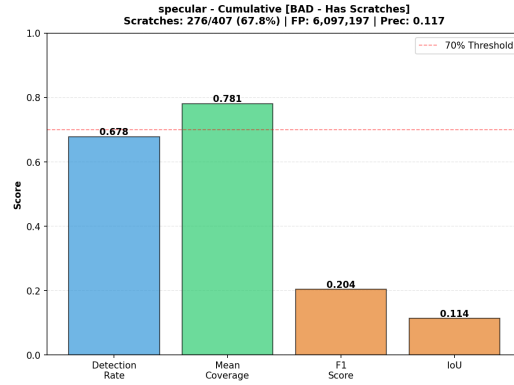
(a) Diffuse SVBRDF map used for training and testing.



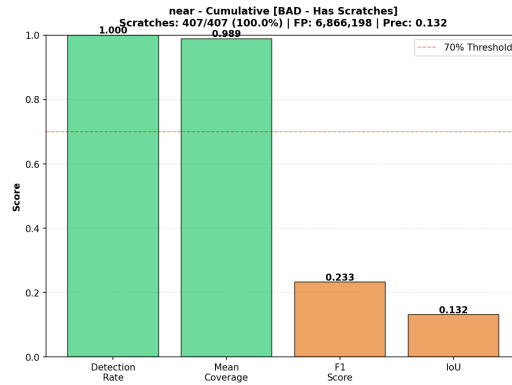
(b) Normal SVBRDF map used for training and testing.



(c) Roughness SVBRDF map used for training and testing.



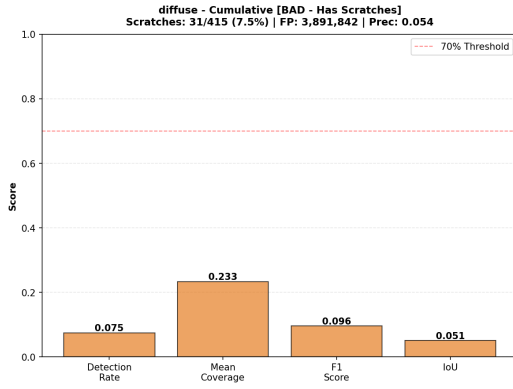
(d) Specular SVBRDF map used for training and testing.



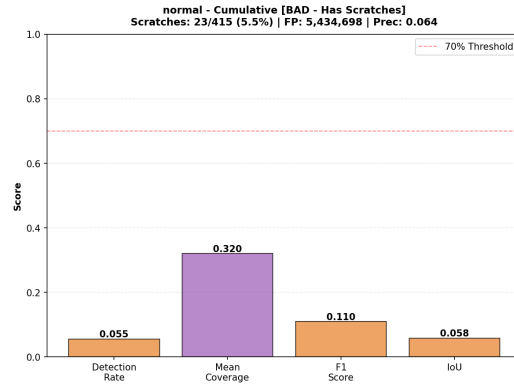
(e) Near-field image-space inputs used for training and testing.

Figure A.15: Anomaly-detection metrics for each individual input representation on the deep-scratch test set: the diffuse (a), normal (b), roughness (c), and specular (d) SVBRDF maps, and the near-field image-space input (e). Each representation was used for both training and testing.

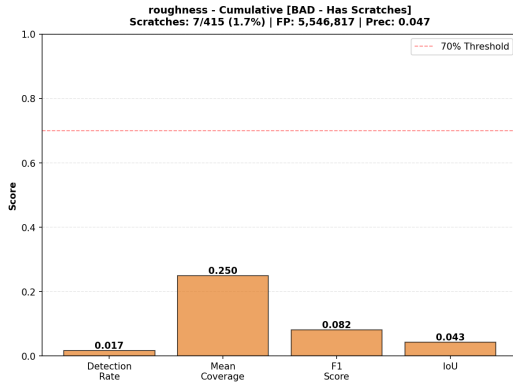
A.2.2 Singular map or image input with shallower scratches



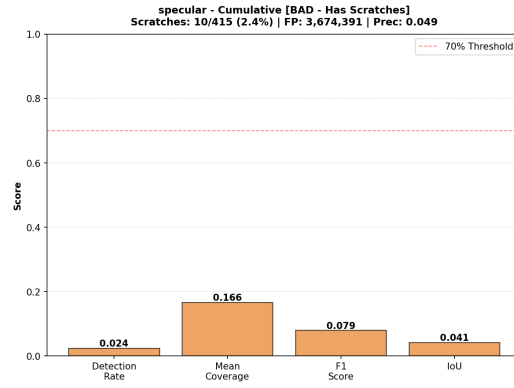
(a) Diffuse SVBRDF map used for training and testing, tested against shallower scratches.



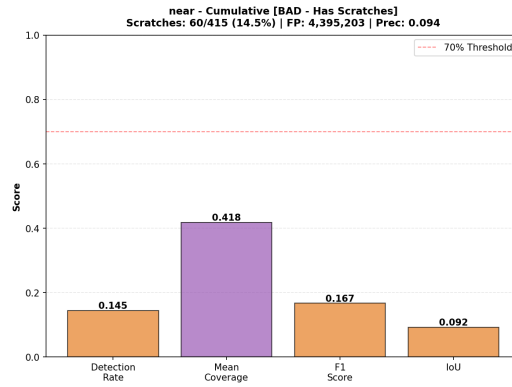
(b) Normal SVBRDF map used for training and testing, tested against shallower scratches.



(c) Roughness SVBRDF map used for training and testing, tested against shallower scratches.



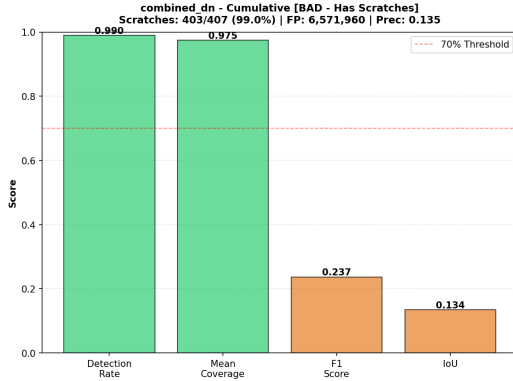
(d) Specular SVBRDF map used for training and testing, tested against shallower scratches.



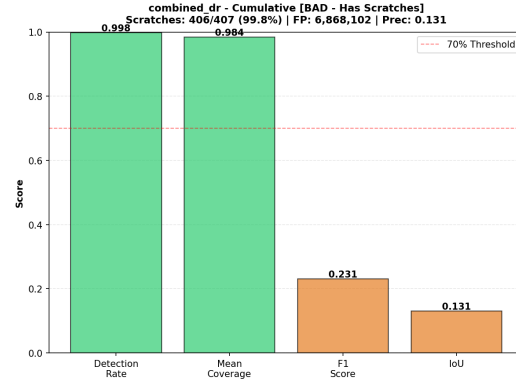
(e) Near-field image-space inputs used for training and testing, tested against shallower scratches.

Figure A.16: Anomaly-detection metrics for each individual input representation on the shallow-scratch test set: the diffuse (a), normal (b), roughness (c), and specular (d) SVBRDF maps, and the near-field image-space input (e). Each representation was used for both training and testing.

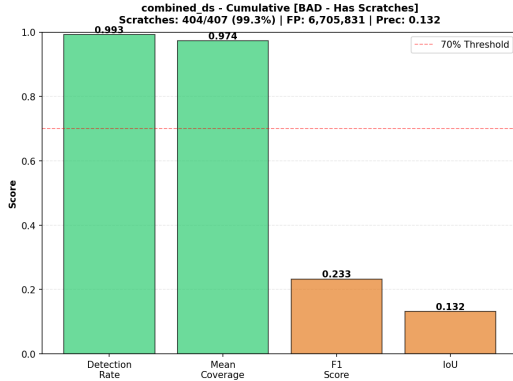
A.2.3 Combining SVBRDF inputs



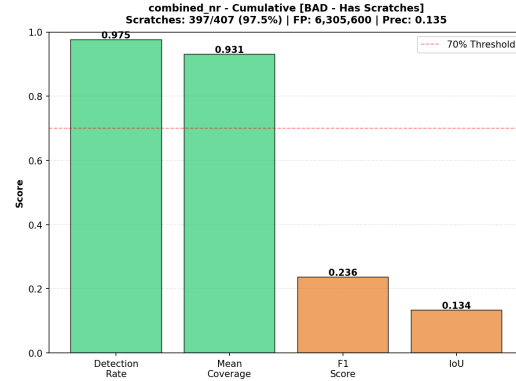
(a) Linearly combined diffuse and normal maps.



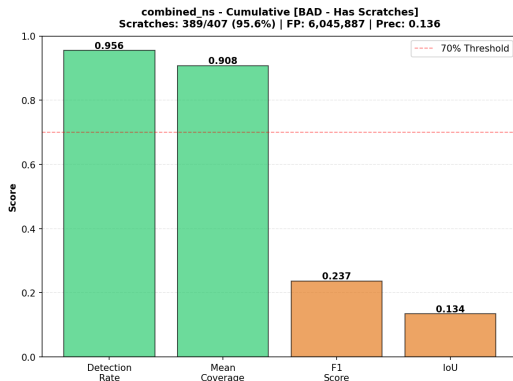
(b) Linearly combined diffuse and roughness maps.



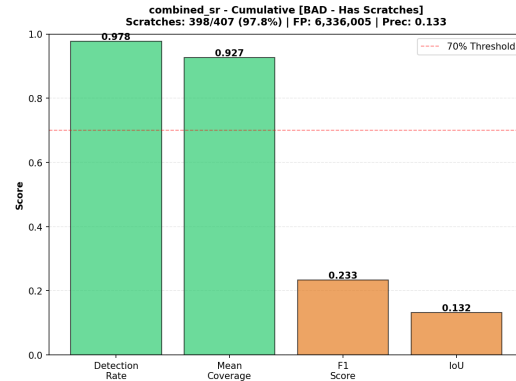
(c) Linearly combined diffuse and specular maps.



(d) Linearly combined normal and roughness maps.

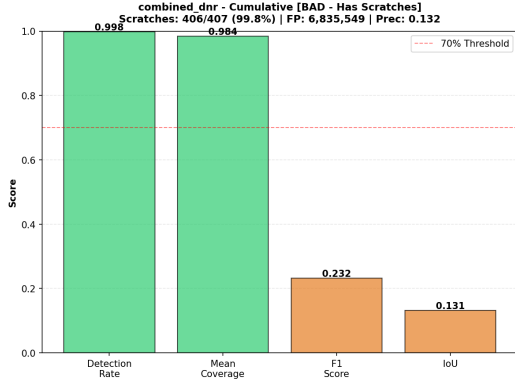


(e) Linearly combined normal and specular maps.

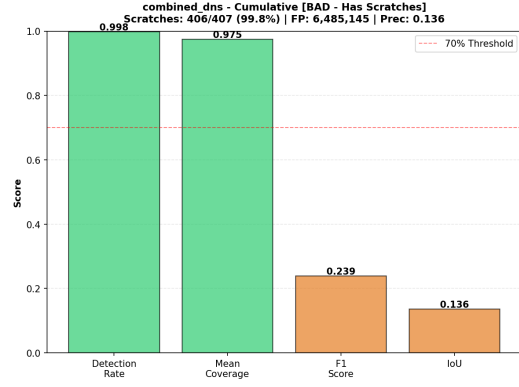


(f) Linearly combined roughness and specular maps.

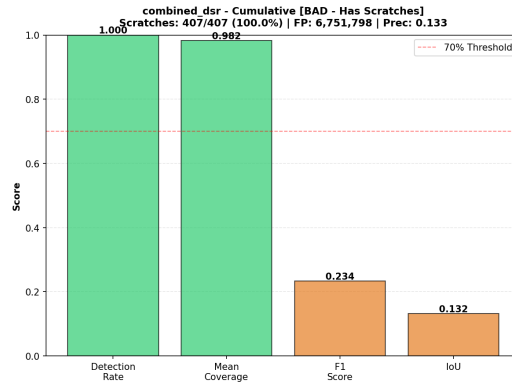
Figure A.17: Two-map combinations of SVBRDF maps on the deep-scratch test set. Each combination was linearly combined and thresholded into a single mask, then compared to the ground-truth scratch mask.



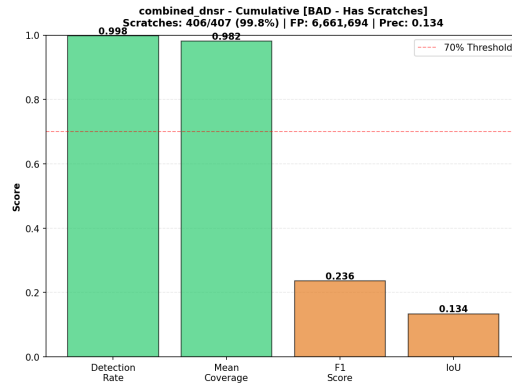
(a) Linearly combined diffuse, normal and roughness maps.



(b) Linearly combined diffuse, normal and specular maps.



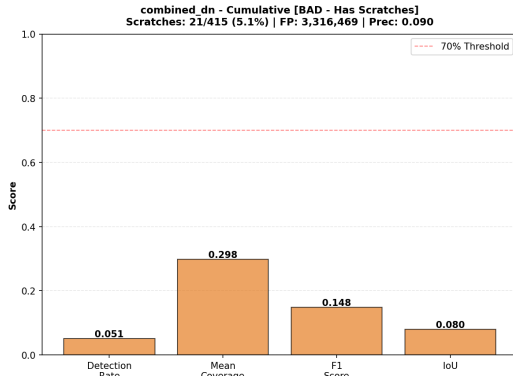
(c) Linearly combined diffuse, roughness and specular maps.



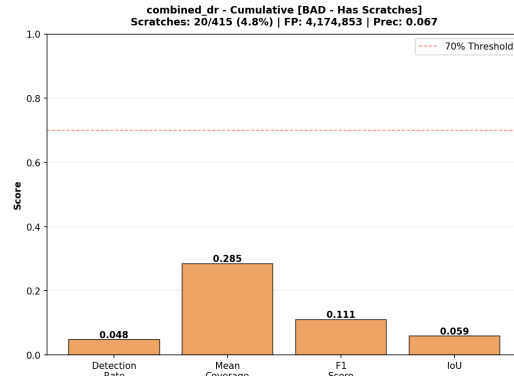
(d) Linearly combined diffuse, normal, roughness and specular maps.

Figure A.18: Three- and four-map combinations of SVBRDF maps on the deep-scratch test set. Each combination was linearly combined and thresholded into a single mask, then compared to the ground-truth scratch mask.

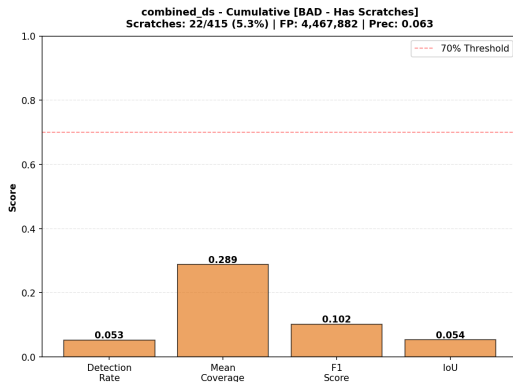
A.2.4 Combining SVBRDF inputs with shallower scratches



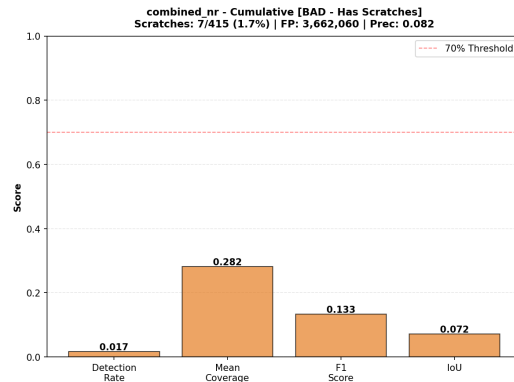
(a) Linearly combined diffuse and normal maps with much less noticeable scratches.



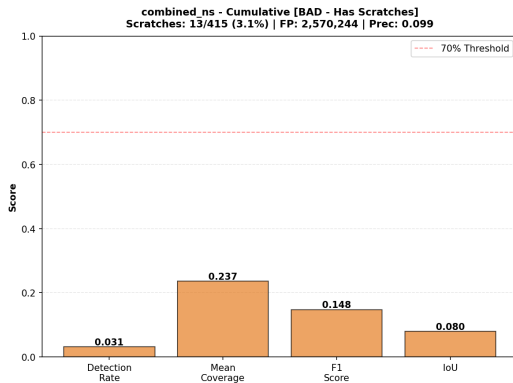
(b) Linearly combined diffuse and roughness maps with much less noticeable scratches.



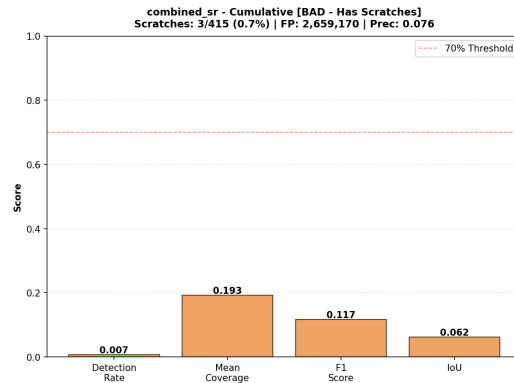
(c) Linearly combined diffuse and specular maps with much less noticeable scratches.



(d) Linearly combined normal and roughness maps with much less noticeable scratches.

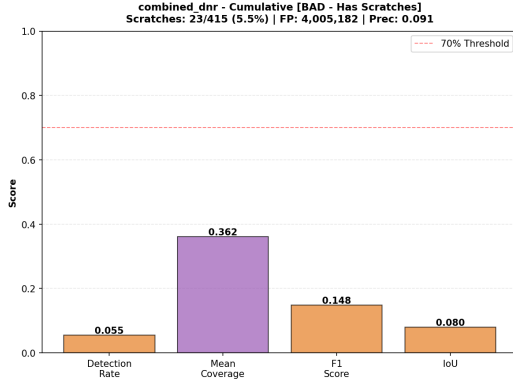


(e) Linearly combined normal and specular maps with much less noticeable scratches.

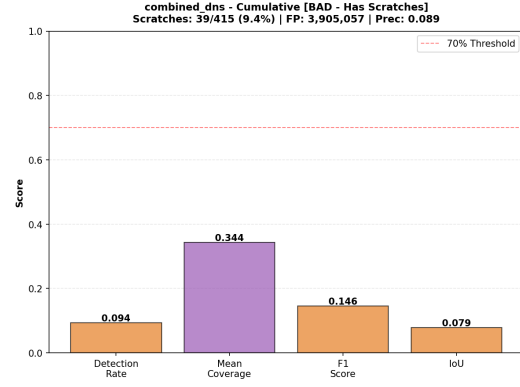


(f) Linearly combined roughness and specular maps with much less noticeable scratches.

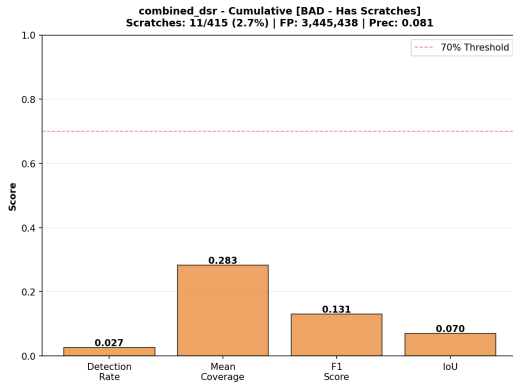
Figure A.19: Two-map combinations of SVBRDF maps on the shallow-scratch test set. Each combination was linearly combined and thresholded into a single mask, then compared to the ground-truth scratch mask.



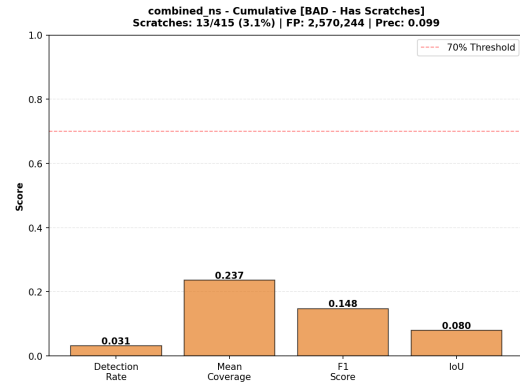
(a) Linearly combined diffuse, normal and roughness maps with much less noticeable scratches.



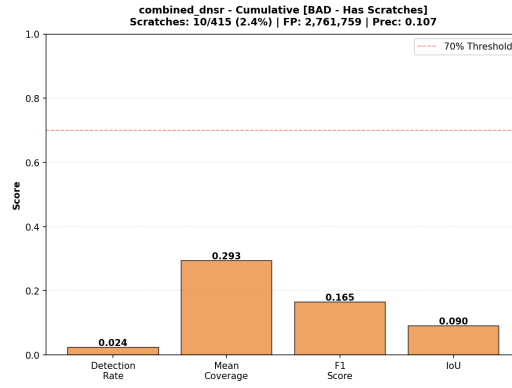
(b) Linearly combined diffuse, normal and specular maps with much less noticeable scratches.



(c) Linearly combined diffuse, roughness and specular maps with much less noticeable scratches.



(d) Linearly combined normal, roughness and specular maps with much less noticeable scratches.



(e) Linearly combined diffuse, normal, roughness and specular maps with much less noticeable scratches.

Figure A.20: Three- and four-map combinations of SVBRDF maps on the shallow-scratch test set. Each combination was linearly combined and thresholded into a single mask, then compared to the ground-truth scratch mask.